Docket #0640 Date Filed: 10/18/2023

UNITED STATES BANKRUPTCY COURT
DISTRICT OF MASSACHUSETTS - CENTRAL DIVISION

IN THE MATTER OF:	Case #14-40987
TELEXFREE, LLC	
	Worcester, Massachusetts
	Wednesday, October 11, 2023
Debtor.	10:28 a.m.
DARR,	Adv. Proc. 16-4006
Plaintiff,	
v.	
ARGUETA, ET AL.,	
Defendants.	
DARR,	Adv. Proc. 16-4007
Plaintiff,	
v.	
ALECCI, ET AL.,	
Defendants.	

TRANSCRIPT OF HEARING ON:
ADV. PROC. 16-04006:
#442 MOTION OF DOMESTIC AND INTERNATIONAL CLASS REPRESENTATIVES TO EXCLUDE TESTIMONY OF DR. CAMERON E. FREER AS INADMISSABLE UNDER DAUBERT;
#444 MOTION OF DOMESTIC AND INTERNATIONAL CLASS REPRESENTATIVES FOR SUMMARY JUDGMENT;
#470 MOTION OF DOMESTIC AND INTERNATIONAL CLASS REPRESENTATIVES TO STRIKE THE AFFIDAVITS OF JEAN LOUIS SORONDO AND STEPHEN B. DARR PURSUANT TO FRCP 37(C)(1) AND PRECLUDE THEIR TESTIMONY, OR ALTERNATIVELY, TO RESCHEDULE THE DAUBERT HEARING, AMEND THE DEADLINE FOR EXPERT DISCLOSURE, AND INCREASE THE BUDGET OF CLASS DEFENDANTS' FOR LEGAL FEES AND EXPERT COSTS;
#593 MOTION OF DOMESTIC AND INTERNATIONAL CLASS REPRESENTATIVES FOR SUMMARY JUDGMENT;
#456 MOTION OF PLAINTIFF TO PARTIALLY STRIKE PORTIONS OF THE AFFIDAVIT OF FRANTZ BALAN UNDER FEDERAL RULES OF EVIDENCE 602 AND 702;

#482 MOTION OF PLAINTIFF TO EXCLUDE NEWLY DISCLOSED DENNIS EXHIBITS AND RELATED TESTIMONY;
ADV. PROC. 16-04007:
#605 MOTION OF PLAINTIFF TO PARTIALLY STRIKE PORTIONS OF THE AFFIDAVIT OF FRANTZ BALAN UNDER FEDERAL RULES OF EVIDENCE 602 AND 702;
#630 MOTION OF PLAINTIFF TO EXCLUDE NEWLY DISCLOSED DENNIS EXHIBITS AND RELATED TESTIMONY;
#591 MOTION OF DOMESTIC & INTERNATIONAL CLASS REPRESENTATIVES TO EXCLUDE TESTIMONY OF DR. CAMERON E. FREER AS INADMISSIBLE UNDER DAUBERT;
#619 MOTION OF DOMESTIC AND INTERNATIONAL CLASS REPRESENTATIVES TO STRIKE THE AFFIDAVITS OF JEAN LOUIS SORONDO AND STEPHEN B. DARR PURSUANT TO FRCP 37(C)(1) AND PRECLUDE THEIR TESTIMONY, OR ALTERNATIVELY, TO RESCHEDULE THE DAUBERT HEARING, AMEND THE DEADLINE FOR EXPERT DISCLOSURE, AND INCREASE THE BUDGET OF CLASS DEFENDANTS' FOR LEGAL FEES AND EXPERT COSTS;

BEFORE THE HONORABLE ELIZABETH D. KATZ

APPEARANCES:

For Stephen Darr, Trustee:	DANIEL J. LYNE, ESQ. (Via Zoom) ANDREW G. LIZOTTE, ESQ. ALEXANDRA M. PAPAS, ESQ. Murphy & King, P.C. 28 State Street Suite 3101 Boston, MA 02109
For Frantz Balan & Sandro Paulo Freitas & Marco Puzzarini:	ILYAS RONA, ESQ. MICHAEL J. DURAN, ESQ. Milligan Rona Duran & King LLC 28 State Street Suite 802 Boston, MA 02109

Electronic Sound Recording Operator: ALBERTO BARRERA

Proceedings Recorded by Electronic Sound Recording
Transcript Produced by Certified Transcription Service
eScribers, LLC
7227 N. 16th Street, Suite #207
Phoenix, AZ 85020
800-257-0885

I N D E X

WITNESSES:	DIRECT	CROSS	REDIRECT	RECROSS	VOIR DIRE
For The Plaintiff:					
CAMERON FREER					
(By Mr. Lyne)	25	133			

EXHIBITS:	DESCRIPTION	I.D.	EVID.
For the Trustee:			
1	Dr. Freer report		129
2	Dr. Freer report		129
3 - 33	Various documents		131
34	Article		132

(At 10:28 a.m.)

THE CLERK: Please be seated. Case number 16-4006, Darr v. Argueta, and case number 16-4007, Darr v. Alecci.

In each of these matters, we're here for hearing on the motion of the domestic and international class representatives to exclude testimony of Dr. Cameron E. Freer as inadmissible under Daubert; a hearing on the motion of plaintiff to exclude newly disclosed Dennis exhibits and related testimony; a hearing on the motion of the plaintiff to partially strike portions of the affidavit of Frantz Balan under Federal Rules of Evidence 602 and 702; a hearing on the motion of domestic and international class representatives to strike the affidavits of Jean Louis Sorondo and Stephen Darr, pursuant to Federal Rules of Civil Procedure 37(c)(1), and preclude their testimony, or alternatively, to reschedule the Daubert hearing, amend the deadline for expert disclosure, and increase the budget of class defendants for legal fees and expert costs; and lastly, a hearing on the motion of domestic and international class representatives for summary judgment.

For the record, beginning with Mr. Lizotte on my right, could I ask counsel present in the courtroom to please identify themselves as well as who they represent.

MR. LIZOTTE: Andrew Lizotte for the trustee, Your Honor.

MS. PAPAS: Alexandra Papas for the trustee.

MR. RONA: Good morning, Your Honor. Ilyas Rona on behalf of the class defendants.

MR. DURAN: Good morning, Your Honor. Michael Duran on behalf of the class defendants.

THE CLERK: Thank you. Counsel on Zoom, if you could just confirm that you can hear us and see us properly. You can do that just by identifying yourself as well as who you represent.

MR. LYNE: This is Daniel Lyne on behalf of the trustee. I can see the Court, and I can see Mr. Rona's table. I don't see -- I think I see the witness box. I do not see trustee counsel table.

THE CLERK: Okay. Mr. Lyne, what you're actually looking at, it's not intended to be Mr. Rona's table. You're looking at the podium.

MR. LYNE: Oh.

THE CLERK: So when somebody is standing at the podium, that's who you will see. And on the screen that has the court seal, that's where any exhibits that are being shown electronically will appear as well.

MR. LYNE: Got it.

THE COURT: Good morning, everyone. So I was anticipating proceeding with the Daubert motion and specifically just whether to admit the testimony of Dr. Freer. So we had these other pleadings come through with regards to

Mr. Sorondo and Mr. Darr. If possible, I sort of would like to put that off until I get a handle on the first part of this. Are the parties agreeable to having that be how we proceed?

Attorney Rona?

MR. RONA: Yes, Your Honor. I thought that Your Honor had sort of stated or hinted that that's what was going to happen, and we're fine with that. As I understand it, there was a motion filed yesterday, and there's the motion involving the Frantz Balan affidavit.

THE COURT: Right.

MR. RONA: I think those still need to be addressed for purposes of today.

MR. LYNE: And the trustee is amenable as well.

THE COURT: Okay. So Attorney Rona, you do think I need to figure out whether to strike portions of the affidavit of Mr. Balan today as part of this hearing, or you don't?

MR. RONA: Well --

THE COURT: I'm misunderstanding.

MR. RONA: -- I guess it's really their motion, so I'm trying to understand the purpose of their motion.

THE COURT: Right.

MR. RONA: But I'll tell you what our intention is. Our intention is to present the affidavit of Frantz Balan, because that is a document that Josh Dennis actually did

review as part of his expert work. And experts can review affidavits. They can rely on hearsay. So to me, that wasn't controversial. And it, along with everything else Mr. Dennis has reviewed, helped inform his opinions.

As I understand it, the motion is to try to excise portions of the affidavit, and either to create a situation where, I guess, Mr. Dennis could not have relied on it, or just to keep it out of Your Honor's consideration. I'm not clear exactly what the purpose is. But to the extent they're going to make arguments as to Mr. Dennis' testimony, I do think that better to deal with those now so that I can understand what I'm going to be questioning Mr. Dennis about.

THE COURT: I see.

Attorney Lyne, are you the counsel I should be directing --

MR. LYNE: I am. The motion lays out exactly, I think, it's only portions of Mr. Balan's testimony in which he purports to suggest he is speaking for the larger population of the Telex participants.

I have no problem with him testifying through affidavit with respect to his personal experience. I would not want Mr. Dennis to rely upon his more global statements with respect to what all people did because there's no foundation for that. It's not a hearsay issue. It's that Mr. Balan is incompetent to testify with respect to what other

participants did. That's outside his ken.

And so I'm fine with the Balan test -- excuse me. I'm fine with the Balan affidavit coming in, as redacted, in my motion to partially strike. That I have no problem with. And I think I've been pretty precise with respect to which portions of the Balan affidavit I believe he is incompetent to testify to. That said, if that is allowed, then Mr. Dennis is free to refer to the Balan affidavit, as redacted, in any way, size, shape, or form he chooses.

THE COURT: Thank you.

MR. RONA: Your Honor, may I just address the competency part? I didn't cover that in my --

THE COURT: Go ahead.

MR. RONA: Just very briefly, that Mr. Balan worked with a team of people. And so his statements are based on his personal knowledge of his interactions with the team.

I think what Mr. Lyne is worried about is not in play here, because I don't think Mr. Dennis says: Mr. Balan says something; therefore it applies to everybody and therefore that's my opinion. It's simply one of many data points.

So again, I don't see the need to excise any portions of the affidavit. I think it's pretty clear that he is speaking from personal knowledge, as a participant who had a team of people working with him, and was aware of how all of the members of his team had the same issues with getting

credits, with recruiting new participants.

MR. LYNE: Your Honor, if I may?

THE COURT: I'm not inclined to strike portions of an affidavit. It seems to me the issues that the trustee has brought up are really amenable to cross-examination. And in highlighting that, if it's your belief that Mr. Dennis is relying on information that's not correct, then that's available for cross-examination. And if I sustain the objection, maybe the defendants need to put Mr. Balan on the stand. I don't know what's going to happen.

But I think right now I'm not going to make a ruling on that motion. Part of the issue for me is, until I really hear what the expert has to say, I'm sort of behind the eight ball. You're all years ahead of me, to be perfectly frank on that. Obviously, you already had an entire hearing on experts years ago.

So I'm just not going to rule on it right now and ask that the parties bring it back to my attention when the time comes, when Mr. Dennis is testifying, and we can deal with it then.

And the same thing with all the other pleadings that are on for today. I'd like to just hold them all in abeyance, except for the two motions that are requesting that I exclude the testimony of Dr. Freer. And as to those motions -- and obviously there's just one in each adversary, so it's really

the same motion -- because the trustee ultimately has the burden of proof on this admissibility -- normally, whoever brings the motion would put on the witness. But I assume the parties realize I expect the trustee to put on Dr. Freer because it is their burden. And I just wanted to put that --

Attorney Lyne, that was your expectation?

MR. LYNE: Correct.

THE COURT: So the schedule for the Court will be, we happen to have one matter at 11:30 that will not take more than five minutes, that I'll have the parties actually stay exactly where they are. We're going to take it. The parties are going to be on the phone, I think.

Mr. Reynolds, right?

THE CLERK: Correct.

THE COURT: But outside of that, the Court is planning on having the hearing today and tomorrow go from 10 to 1, take a lunch from 1 to 2, and then we'll go from 2 to 4. If, at any time, between 10 and 1 and 2 and 4, somebody needs a break, obviously, just let me know. But I wanted to let the parties know that was the plan of the Court. And we have no other matters that I'm aware of for tomorrow. So it would just be us.

With that, Attorney Lyne, would you be prepared to -- hold on one second.

Mr. Reynolds?

THE COURT: Okay. So let's proceed. For purposes of the docket -- I mean, of the record, it's docket number 442 on case number 14-40987. And it's docket number 591 on case number 14-40987, is my understanding.

THE CLERK: Judge, if I could --

THE COURT: No, I'm sorry. I'm saying -- I'm sorry; that's the underlying case number. Let me just correct myself there. So --

THE CLERK: In case number 16-4006 --

THE COURT: Thank you.

THE CLERK: -- this is pleading number 442. And in case number 16-4007, it is pleading number 591.

THE COURT: Thank you, Mr. Reynolds.

So Attorney Lyne, I will permit a brief opening, if you're prepared to do that, and then I'll let Attorney Rona give his brief opening, or Attorney Duran, whoever is going to.

MR. LYNE: Well, just as a practical matter, I just want to bring you the Court's attention, the trustee did file a motion to exclude some newly-disclosed exhibits and any related testimony that might follow from those newly-disclosed exhibits. And I'm not exactly sure how you want to deal with that issue.

THE COURT: So are those exhibits, Mr. Rona or Mr. Duran, anything you're intending to use during your cross-

examination of Dr. Freer?

MR. RONA: Some of them, Your Honor, Yes. And I just want to just give a little bit of background. The parties exchanged exhibit lists on, I believe, the 2nd, or thereabouts, and the list that was, I think, jointly filed said that the parties could add supplemental exhibits by 5 p.m. on Monday.

And both sides have actually done that. Both sides -- the trustee has prepared a PowerPoint which they have labeled as an exhibit. And ordinarily I'd be fine with that. But now we have data that's the trustee's data, and Freer's analysis of it, that we've prepared slices of or compilations of under 1006. That's what most of the paperwork in this case is, is attempts to take a massive pool of data and show you little pieces of it.

And so we found additional pieces that we were going to use both with Mr. Dennis and with the cross of Dr. Freer. And so I don't see any reason why those can't be used. We've disclosed them; they have copies. And again, this is the trustee's own data.

THE COURT: So it's essentially a chalk or some sort of a -- it's not really a chalk, if you're looking to introduce it, but it's a honing in on specific pieces of evidence that had already been provided?

MR. RONA: That's correct, Your Honor.

THE COURT: Attorney Lyne?

MR. LYNE: I don't think that's actually correct. And just by way of background, the parties had been in the process of submitting proposed exhibits starting on September 27th. The pre-trial memo was filed on October 2nd. None of these new exhibits were identified or disclosed. We received a dump of a substantial number of new tables.

I'll ask, if possible, just to bring up Exhibit 9, if it's possible to project that. That's a hundred pages of new tables which had never been disclosed. If you could simply scroll through that, Ms. Papas, and gives the Court a sense of the extent of which there appears to be a whole new set of tables, up to twenty-seven fields per table.

It is almost -- well, it's not almost impossible; it is entirely impossible for me to absorb what is being proposed to be used on cross or on direct with Mr. Dennis. The point here is that these are not simply excerpts from prior tables. There has been synthesis here, clearly done by Mr. Dennis, and it is the product of what must be at least weeks, if not months, of additional work by him, identified, none of which appears in any of his exhibits to his report, none of which was disclosed prior to his deposition.

And to put this in front of us at 5 p.m., two days before trial is about -- it's actually thirty-six hours before trial is about to start on this issue -- I think is grossly

prejudicial and unfair. I never received any heads up with respect to this. And there is, it appears, from the short amount of time that I've been able to look at it, they are really attempting to introduce information which relates to low rep ID and nontriangular credit transactions between two individual defendants, none of which is relevant in this particular Daubert hearing.

And so it's really, I think, grossly inappropriate for counsel for the class defendants, who have provided me with something that I cannot synthesize, which I could not talk to my expert about, and which clearly is the product of not simply excerpts from tables, but compilations in ways that I cannot understand, and which clearly address issues which are not part of this case, as we discussed last Friday, the issue of low rep ID, the issue of credit transactions. That doesn't go to the validity of Dr. Freer's aggregate and methodology.

This is an entirely different issue and can be addressed at some later point in time. But as to Dr. Freer, the issue before the Court is under 702, frankly, as amended, effective December of this year, is his methodology the product of a peer-reviewed process, and does that process provide reasonably accurate results, and are the results more likely than not to be helpful to the Court in the determination of its decision on the admissibility of his

methodology and report.

This is something entirely different. Let's not mix apples and oranges and certainly don't put me in a position where I have no idea what's going on and how Mr. Rona intends to use this on cross with my client who has had no opportunity to look at this or even talk to me about this.

THE COURT: Thank you, Attorney Lyne.

MR. RONA: Your Honor, I just wanted to respond quickly, because I think I was accused of acting in bad faith. The exhibits that are before Your Honor were prepared on Monday. It wasn't weeks in coming. I had a conversation with Mr. Dennis. We talked about ideas for how to approach some of the issues based on all of the briefing in this case. And the tables are in the exact same format that Mr. Dennis has prepared appended to his exhibit.

And I don't know -- it's unclear to me how I'm supposed to cross Dr. Freer with what Your Honor will see in their presentation. If you printed all the paper in TelexFree, it would be a pile taller than the Burj Khalifa. And I'm not clear how I can cross-examine Mr. Freer with the Burj Khalifa.

I think what I have to do, and what I've done, is I've disclosed in advance slices of that data that I would like to put in front of Dr. Freer to question his methodology and his assumptions. And he is free to have looked at it

yesterday, he can look at it on the stand, and he can answer the questions to the best of his ability.

THE COURT: I do see that the pre-trial indicates that the parties are permitted to present demonstratives not later than Monday, October 9th at 5 o'clock. So that appears to be what was done here.

It is not clear to me how this document's going to be used. And the way I've understood the case so far, there appears to be two pieces of it, which is Dr. Freer's ability to figure out user accounts and aggregate them. And then I think, initially, the defendants thought that Dr. Freer was also then figuring out net winners and net losers. But now it's been clarified that Mr. Sorondo and Mr. Darr took the second step.

And to the extent these demonstrative wind up going to that second step, it certainly is not an issue for me today. But if they do go towards Dr. Freer's cross-examination, I will have to reconsider. But my ruling right now is I need you to object during the hearing, Attorney Lyne, so I really understand how you may be prejudiced by this, because right now I'm not clear that you are.

It appears to be a demonstrative. It appears the parties agreed they could be provided a couple of days ago, and they were provided a couple of days ago. So without prejudice to the trustee bringing an objection during trial,

right now I'm not going to exclude these exhibits. And with that, I would like a brief opening statement, please.

MR. LYNE: Thank you, Your Honor. The issue in front of the Court is simple, whether Dr. Freer, a highly credentialed scientist, and specifically in the data science field, has had the ability to determine whether: one, the data in the TelexFree dataset is of sufficient quality to permit him to proceed with an aggregation methodology; two, what aggregation methodology is appropriate under these particular circumstances; three, whether the results that he obtained as a result of the work he has done is sufficiently reliable and it should be permitted to be allowed by the Court to be entered into evidence. Anything else is not for the Court's determination at this point in time.

As it happens, Dr. Freer, an MIT research scientist in the Center for Probabilistic Computing, is uniquely positioned to have done the work that he did in this case. He is a preeminent data scientist in the field of probabilistic computing. He will explain, in the context of his presentation, what probabilistic computing is, how it differs from deterministic computing that was be used by Huron.

He will describe in detail the steps he took to prepare the data for a probabilistic computing engine of his choice, what probabilistic engine he used, and whether or not that is a peer-reviewed methodology that is generally accepted

in the data science world, and after that, how he then took the aggregate output from that probabilistic step and then used additional, well-accepted data science tools to then cluster those results in a way that provided the Court -- will provide the Court and does provide the Court with a series of clusters assigned, each of which is assigned to one individual and only one individual.

He was provided with a net equity calculation table, which he ran against the clusters, and that provides a sum, an end result with respect to the cluster net winner category.

But the Freer methodology is, in fact, agnostic to any net equity table that the Court wishes to apply. What the Court determines with respect to that is reserved for another day. But under 702, the only issue is, is Dr. Freer qualified, did he use appropriate peer-reviewed processing, was the data of sufficient quality to permit him to use those processes, and as a result, is the aggregate output, the cluster analysis, sufficiently reliable that the Court should deem it to be helpful to its resolution of the issues in this case.

And I would submit to you that there is no one perhaps better suited to participate and direct this probabilistic exercise than Dr. Cameron E. Freer, a preeminent data scientist.

THE COURT: Thank you. And defense, do you wish to

give a brief opening?

MR. RONA: Yes, Your Honor. Thank you. May I proceed from here, or would Your Honor --

THE COURT: Whichever you want.

MR. RONA: Thank you, Your Honor.

THE CLERK: Counsel, if you wouldn't mind standing at the podium. If that's going to be a difficulty, that's fine. But if it's not --

MR. RONA: That's fine. That's fine.

One moment, Your Honor.

Thank you, Your Honor. Daubert actually has two components to it, Your Honor. There's not just the component that Mr. Lyne explained, but there's also the question of whether what the expert is doing is relevant to the task at hand.

So even the best method, the most gold plated methodology can't answer the wrong question. And so what's important here is to answer the correct question, and that is sometimes called fit. And so many of the arguments that we've raised have to do with the lack of fit of Dr. Freer's analysis.

And Your Honor has mentioned that you see net equity as being separate and apart from what we're talking about today, but we have several reasons why we don't believe that that's the case.

One, we have a stipulation in this case that separated the class action into a phase one and phase two. And phase one said that it will deal with the admissibility of the expert opinion of Borelian Corporation -- that's Dr. Freer -- and will include, among other things, the following: the integrity and reliability of the sig (ph.) data -- I'm going to get to that in a second -- the reasonableness of the assumptions made by Borelian in computing the amount of net winnings, including but not limited to inclusion, exclusion of transfers of credits, and assumptions made with respect to the treatment of triangular transactions.

So the issue is that that was what this was all about. This is about understanding who is a net winner. Dr. Freer's output, which is two massive pieces of data -- one is actually more massive than the other, but the less massive one is simply a list of cluster IDs -- I'll get to what that is in a second -- and the net equity alleged by the trustee for that cluster.

So you can't know who's a net winner without knowing net equity. And so the trustee has to both identify who the net winner is and allege what their net winnings are. And as things stand, as I understand now the state of the record, they don't have an expert to talk about net equity. And according to the stipulation, that was supposed to be Borelian Corporation.

And the other part of Dr. Freer's lack of fit is that he says I'm not identifying participants. You're going to hear Dr. Freer say I'm just doing clusters. The problem is that's not helpful to Your Honor. Your Honor doesn't need to know about clusters. Your Honor needs to know who's a net winner, who's the participant tied to that account. And that's disclosed, both in prior filings in this Court, but also on the second page of Dr. Freer's report. He says, my task, as I understand it, is to aggregate accounts to a participant.

So the fact that he doesn't even attempt to do either the identification of a participant or the computation of net winnings makes there be a lack of fit in this case.

Turning now to his methodology, as I said before, data reliability is an issue in this case. And we have no quibbles with Dr. Freer's credentials. He is certainly preeminent in his field. The problem is, as a result of his preeminence, he has worked on very common types of cases where you have data that's actually validated, you have data that's actually reliable, you have data that has some what is called ground truth to it. And all you're doing is linking records. So when two hospitals merge, for example, that's a challenge that the industry will have, and Dr. Freer deals with those challenges.

TelexFree is something completely different. This

data is not validated. I think everyone's going to admit in this case that there was no quality control to the data. And he did no separate analysis to try to understand how reliable this data is. It's not a question of is it or is it not reliable; it's a question of how reliable is it. If you can't tell what how reliable it is, then how reliable are your conclusions? And he doesn't address those data problems.

And I will note that that, under 702, it's not just, again, your methodology is great; it's does it reflect a reliable application of the principles and methods to the facts of the case? So while what Dr. Freer did would have been fabulous if he was, for example, trying to match up lawyers and their BBOs, if the lists got separated, where that data is pretty clean and valid, no one is going to be using my BBO, and I'm the only one in Massachusetts -- I've checked -- that has my name, so it's pretty easy to match. But what if lawyers have been using fake BBOs for years? That would be a problem.

Similarly, Dr. Freer makes assumptions about the name in this case. The name you're going to hear repeatedly is sort of the thing that everything pivots around. It's the center of the universe in TelexFree. The problem is Dr. Freer gave the name a special role, says it has a special meaning, but can't really articulate how he came to that, how he tested it, how he measures it, and really does the same -- makes the

same mistake that Mr. Martin made the first time around.

Your Honor may recall that Tim Martin thought that the name was the most important thing and everything had to match based on the name. So to say, well, okay, maybe the name is not that important but it's still special, is really a distinction without a difference in this case.

Turning to the actual methodology, it's a ten-step methodology. Your Honor's going to hear about it. And each step, in isolation, you can say that's reasonable, that's reasonable, that's reasonable. But when you add them together, there are evident problems. And they're evident in the slices of data that Dr. Freer presents. They're evident in slices of data appended to Mr. Dennis' report and in other slices that Your Honor will see.

Specifically, there's the issue of transitive closure, which is, if A is similar to B, and B is similar to C, and C is similar to D, we're going to put them all in the same thing, even if A and D are different. Even if your own methodology said A and D don't belong together, we're going to stick them together.

And the problem is that that has happened. And Dr. Freer doesn't even quantify how often it happens. And so that's an issue. And it's an issue because, when you put two accounts that don't belong together, you affect the net equity.

Similarly, there's still the issue of under-aggregation, which is you separate accounts. So I may have some accounts over here, I may have some accounts over there. The problem is that one set might be positive, one set might be negative, and therefore each one is creating a false picture of my true net equity.

So at the end of the day, the issue is not specifically whether Mr. Freer chose peer-reviewed methodologies. The issue is whether he reliably applied those methodologies to the facts of this case, with the data that we have, without any attempt to do any additional analysis or validation.

THE COURT: Thank you.

MR. RONA: Thank you, Your Honor.

THE COURT: Attorney Lyne, are you ready to call Dr. Freer?

I am. The trustee calls Dr. Cameron Freer to the stand.

THE COURT: Go ahead, sir, and stand up -- I mean, stay standing and raise your right hand, and Mr. Reynolds will swear you in.

DR. CAMERON FREER SWORN

THE CLERK: Thank you. You can be seated. If you could, just for the record, state your name and your business address, please.

Cameron Eric Freer. My business address is 955 Massachusetts Avenue, number 256, Cambridge, Massachusetts 02139.

THE CLERK: Thank you.

MR. LYNE: I'm ready to proceed, Your Honor.

THE COURT: Go right ahead.

MR. LYNE: Thank you.

DIRECT EXAMINATION

BY MR. LYNE:

Q. Dr. Freer, are you currently employed?

A. Yes, I am.

Q. And where is that?

A. At MIT, the Massachusetts Institute of Technology.

Q. And what's your role at MIT?

A. I'm a research scientist in the Probabilistic Computing Project within the Department of Brain and Cognitive Sciences.

Q. And when you talk, sort of, generally, about the Center for Probabilistic Computing, can you just give us, sort of, a general sense of the types of things that you do as a research scientist at MIT's Center for Probabilistic Computing?

A. Yes. So it's a research center that it's in the Department of Brain and Cognitive Sciences, but it's more of a computer science research group with some connections to neuroscience and cognitive science. And there's a range of

research and engineering that goes on, from theoretical to quite practical. And I engage in research on my own, with graduate students, with other collaborators, write papers, advise graduate students, do some teaching.

Q. And I just want to hearken back to Mr. Rona's opening, where he said that you worked with ground truth sets or so-called validating sets. Is that true? Is that the sum total of your experience, working with ground truth sets?

A. No. Quite often there's datasets for which you don't have any labeled training data. This is what's known in machine learning as unsupervised learning, where you have techniques that need to work automatically on data that's not labeled at all. This is a large part of machine learning and artificial intelligence and -- and one of the large components of datasets that I work with.

Q. And Doctor, Mr. Rona also suggested to the Court, in his opening statement, that name is what he referred to as the center of the universe in your methodology; is that true?

A. No. Name plays a role, like many other fields. It is indeed true that name plays a somewhat different role from some of the other fields in a way that I'll describe. But it's one component among many.

Q. And with respect to transitive closure, was the sum total of your transitive closure work in this case simply saying if A equals B, and B equals C, then A equals C?

A. No. So first of all, transitive closure, it's -- it's a concept, and every cluster is transitively closed. You can't have a cluster that doesn't have the property that if A is in the cluster with B, and B is in the cluster with C, that A is not in the cluster with C. That's just a property of clusters. Also, it happens that there's quite a lot that goes into, as I'll describe later, I think, involving the dynamic clustering step, carefully balancing how and when to use transitive closure.

Q. Okay. So let me just -- I wanted to get that out, but let me go back.

MR. LYNE: And if you can display, Ms. Papas, Exhibit 3, which is Dr. Freer's CV.

BY MR. LYNE:

Q. I'm directing your attention to what is Exhibit 3 in the exhibit book.

THE COURT: Hold on. Hold on one second. We need to fix the -- okay.

MR. LYNE: All set?

THE WITNESS: Should I use this so I can see it more clearly?

MR. LYNE: Sure. And this is Exhibit 3 in the exhibit book. I assume all counsel and the Court has a copy, so if they want to look at it directly --

THE COURT: Yes.

MR. LYNE: -- please do so.
BY MR. LYNE:
Q. Sir, directing your attention to your curriculum vitae, is this the most current version of your curriculum vitae?
A. Yes.
Q. All right. And how old is it at this point in time?
A. I believe it's from about two years ago. There's a number of papers since then that haven't been included, but I have not prepared a more recent CV yet.
Q. All right. In terms of experience in the industry, do you see the reference to Borelian Corporation, Remine, Gamalon?
A. I do.
Q. On the first page?
A. Yes, I do.
Q. All right. And could you provide the Court with, sort of, a general description with respect to those three companies that you have worked with?
A. Yes. So Borelian Corporation is my consulting firm through which I engage in data science research, data science consulting, and also some scientific research. So at University of Toronto, at MIT, before my most current appointment, and at Boston University, it's been more research or teaching appointments.

And also this is the vehicle through which I was engaged as chief scientist of Remine. Remine is a real estate data analytics firm, and as it was starting, I was brought in to develop some of their initial machine learning models, analyze datasets that they needed linked, and compare datasets as they were deciding what to license. And I also helped recruit and build their initial data science team.

Gamalon Labs is a company that began as a contractor to Darrpa (ph.) for research and probabilistic programing languages, which is closely related to what I work on at the MIT Probabilistic Computing Project. So for a number of years, I was a research scientist at this, essentially, research lab startup.

Q. All right. And if you turn to the second page, under research interests, it states that your interests are interactions of randomness and computing, including the foundations of probabilistic computing, efficient samplers, and testing methods for probabilistic inference and the mathematics of random structures. When you talk about testing methods for probabilistic inference, can you describe for the Court what that means?

A. Yes. So this relates to questions and statistics where if you -- for example, let's say you have some probability distribution of interest, and you have some data points, and you want to know if they were probable -- if it is plausible

or likely that they were sampled from the distribution in question, there's always some amount of uncertainty because you could see some things that are unlucky. But this corner of statistical inference is involved in determining when -- when a collection of datasets likely came from a given distribution or identifying if two datasets likely came from the same distribution as each other, questions like this. So there's theoretical questions involved, but it has many practical applications as well.

Q. And will you be addressing this whole issue of probabilistic inference in your presentation once we get to it?

A. Yes. I'll be discussing what it means to have a statistically valid sample, an appropriate sample size, and things of that nature.

Q. Great. And there is a lengthy list of publications. And are you an author on each of the named publications?

A. I am.

Q. All right. And were these peer-reviewed publications?

A. Yes, peer-reviewed, according to the standards in a range of fields, including mathematics, computer science, statistics, I believe, a couple in physics.

Q. All right. And if you turn to Exhibits 11, 13 and 15, would these be examples of some of your publications that

would be relevant in your engagement here?
A. Yes. Each of these three is by me, and they're relevant in different ways.
Q. All right. You have given a number of workshops. If you look at page 9, for example –
THE COURT: Of which exhibit, Attorney Lyne?
MR. LYNE: I'm sorry.
THE COURT: Page 9 on what --
MR. LYNE: Exhibit 3.
THE COURT: Of Exhibit 3? Okay.
MR. LYNE: Yes.
BY MR. LYNE:
Q. Referring to the lower half, where you say one of your workshop categories is Bayesian statistics, machine learning and probabilistic programing, would these be examples of your work in the probabilistic computing fields that would be relevant to your engagement here?
A. Yes. Yeah, this includes a number of workshops, some of which I co-organized. Many of them I was an invited participant, invited speaker.
Q. All right. And there are a number of teaching positions which you have held, which appear starting on page 11, correct?
A. Yes, that's right.
Q. And continuing onto page 12?

A. Yes.
Q. Both in the University of Hawaii at Manoa --
A. Correct.
Q. -- MIT and Harvard University, correct?
A. Yes.
MR. LYNE: Your Honor, I would tender this individual as an expert in the field of probabilistic record linkage. I'm not sure if there's an objection to examining him at this point in time, but I'm tendering him in the field.
THE COURT: Any objection to the qualifications?
MR. RONA: No, Your Honor.
THE COURT: Go right ahead. DIRECT EXAMINATION
BY MR. LYNE:
Q. All right. I'd like to start with a presentation. Did you prepare a presentation for the Court today?
A. I did.
Q. And is that contained in Exhibit 33 of your exhibit book?
A. I believe so. One second. Yes.
Q. And I'd like to start on the second slide with the scope of your engagement. Were you engaged to perform any services in connection with the trustee's litigation with participants in the TelexFree scheme?
A. Yes. Beginning in July of 2021, I was engaged to analyze

the relevant datasets and determine whether it was possible to build an accurate and reliable model of record linkage using this data, and then having determined that it was, I was engaged to design the model, implement it, run it, obtain the cluster results, and also to assess the error rate of the method.

Q. And when you say "analyze whether it's possible to build", did you look at, in connection with the assessment, as to the data quality of the TelexFree dataset and whether it was appropriate to use for this probabilistic process?

A. Yes. So a large part of the analysis involved looking at the data, assessing its quality, determining which fields were relevant. And the conclusion that I came to was that it was eminently useful for this, and more than sufficient, despite the fact that, of course, it does contain all sorts of name variations. You know, there are ways in which it's somewhat messier than some datasets, and yet it's more than reliable to carry out the analysis in question.

Q. All right. And you'll proceed and view with the Court exactly what you did to come up with that assessment later?

A. Yes.

Q. Okay. I want to go to the next slide and just sort of get a general sense from you, what is record linkage in the context of data science?

A. So record linkage involves taking either two datasets or one dataset -- it's one dataset in this case -- and linking records -- linking one or more records to one another based on not just their similarity but whether, in the assessment of the method, they belong to the same entity or not. And in this case, the question is one of participants, those persons, and so the question is how to link records together if, and only if, they're assessed to belong to the same person.

Q. And are there different types of record linkage protocols which are used by data scientists to link records to, in this particular case, individuals?

A. Yes. Broadly speaking, the field refers to deterministic record linkage and probabilistic record linkage methods.

Q. And when you talk about deterministic, what is a deterministic methodology? What is the process?

A. The idea -- so in both cases, at some level, one needs to consider the match -- a potential match between every record and every other record in question. And in deterministic record linkage, the idea is that you get a yes or no answer for every such pair. And this could be on the basis of a particular set of fields matching exactly, or to an appropriate degree. It could also be based on some minimum number of fields, for example. But in any case, the methodology comes to a yes-or-no decision for every pair in question. That's sort of the essence of what makes it

deterministic.

Q. What about probabilistic?

A. So probabilistic record linkage, in contrast, it's assessing the degree of match for every pair. And this is based on the relative closeness. It can take into account degrees of partial match within a field. It can also take into account some fields more strongly than others. And there's, in fact, methods within probabilistic record linkage to automatically determine the appropriate and reliable relative weights of the fields using the data itself, which is especially helpful in cases where you don't have any answers in advance and you need the data to speak for itself.

Of course, this can be computationally more difficult to implement, but the payoff, as I've described in the bottom bullet point of each of these, is that, while deterministic record linkage is often simpler and faster and easier to implement, it often works best when you have data in a uniformly high quality, perhaps with some unique identifier already, whereas probabilistic record linkage, while it can be more difficult to design and implement, the payoff is that, when there's more variability in the fields, it can still achieve high quality results.

Q. To use Dr. -- excuse me -- to use Mr. Rona's term of ground truth sets, does probabilistic computing require a ground truth set?

A. No. As I described earlier, there's a whole field of machine learning called unsupervised learning, which consists in methods that work even when you don't necessarily have labeled training data.

Q. As a result of doing a preparatory review, did you actually develop an aggregation methodology?

A. I did.

MR. LYNE: All right. Can I have the next slide, please?

BY MR. LYNE:

Q. Here's a, sort of, simplified flowchart. And if you could just generally walk the Court -- we'll look at each of these in detail later, but generally walk the Court and orient the Court to what the methodology is and what the sequencing is in terms of how your methodology proceeded from start to finish.

A. Sure. So broadly speaking, one needs to begin with input data. I can tell you what that consisted of in this case and how I assessed its quality and so on. And then once one has the input data, there's a number of steps that need to be taken to prepare it for the key step of matching certain records to certain other records. And so this is what I've illustrated here in steps one through five.

So some of these steps involve cleaning or otherwise working with the data to get it in a form where higher quality

matches are possible than with the raw data. These are standard techniques within record linkage. And the other half of -- of one through five, roughly, involved reducing the overall number of comparisons.

If you have two records that are a very clear match in advance, one doesn't need to do something computationally more extensive. Likewise, if you have two records that are very clearly not a match, one doesn't have to compare those. And this is essential with large datasets because, as I'll describe later, it's totally infeasible, even with modern computing, to compare every record to every other in a time-consuming way. So it's important to winnow them in advance. So then having --

Q. What's the next step after that, after the data has been prepared and cleaned?

A. Yes. So after the data has been prepared, in terms of cleaning and reducing the number of comparisons, then there's the key step, which in this case is a probabilistic record linkage technique known as Fellegi-Sunter with expectation maximization.

So I'll get into the details later, but expectation maximization is what allows this to operate in an unsupervised way from the data itself. Fellegi-Sunter is a technique that's been key in record linkage for many decades and is still at the core of many of the main techniques. And so this

is what determines a probability of a match for every pair. It is not yet deciding yes or no, but it assesses, for every pair that's worth considering, the degree of match.

Q. All right. And we'll go through that in some detail later. What are the next steps?

A. So steps seven through nine involve actually forming the clusters of user accounts, one per participant, from that match probability information. So this is the spot that transitive closure comes in, as we were discussing earlier. And the idea is to --

Q. Well, can I just stop you here --

A. Yesh.

Q. -- for one moment? Once you have a cluster, does the cluster identify the individual accounts that are contained within those clusters?

A. Yes. So the cluster consists of a set -- each cluster is a set of user accounts, and each user account has contact information. That's what it mainly consists of, name, email address, postal address, phone number, and so on. And within a cluster, that information -- by the whole point of the clustering, that information is broadly similar within a cluster, and that identifies the participant.

So you might have multiple name variants, maybe some include a middle name, or not, some may have typos, but it's still referring to the same person. Likewise, there might be

two totally different addresses. Maybe they have two homes; maybe they moved. That still is identifying the person. It's giving additional information than a single address. But that identifies them. And similarly with the phone number and the rest.

Q. All right. And once you've completed the clusters, and you had a cluster assigned to individuals, and for each individual, all of their individual account records, what step, if any, did you then take?

A. Yeah. So at the request of the trustee, I then took those clusters, which had not used the net equity in any way to that point; it was just the contact information that went into that. But I then took the net equity numbers that I was provided by the trustee, one per account, and I simply tabulated them per cluster.

So every cluster consists of one or more accounts. I added up the net equity for each of those accounts, and that's what I called the total net equity per cluster. And when that quantity was positive, I called it a net winner. When that quantity was negative, I called it a net loser.

And -- but that's simply adding up those numbers. If you had given me different numbers, it would have taken -- or it would still take just a few seconds to run that tabulation. And it was input before that step.

Q. All right. And I assume those results are then

passed on to the trustee?
A. Yes. So I passed on those results, which are summarized in the fine print in purple at the bottom. So roughly 1.5 million clusters of 11 million user accounts, which is saying that 1.5 million participants are identified from that set of 11 million user accounts. And of those, approximately 80,000 were net winners. And so you can see there were many more net losers, the bulk of the 1.5 million. And it's between one and two billion dollars that was gained by the net winners and was lost by the net losers.

Q. All right. Now, generally speaking, the TelexFree dataset, would this be considered to be a large dataset for purposes of data linkage?
A. Yes, it's very large from the standard of data linkage. I mean, we're obviously used to computers doing millions of calculations, and indeed, for some purposes, a million, or even eleven million, is not a large number, but it's very large from the standpoint of record linkage.

This is essentially because, as we were discussing, in principle, you have to compare every record to every other. And as I said, there's techniques to winnow that before doing so, but if you didn't, you're talking about the square of the number, right? Eleven million compared to eleven million, that's on the order of tens of trillions that have to be compared. And that's completely infeasible. And so that's

why one needs to reduce it further. But that's why this is also among the largest types of aggregation tasks that one performs in data linkage.

And just as a visual sense of how large the initial dataset is, which doesn't even get into the complexity of matching, like I was saying, if you printed each record on one page of ordinary copy paper, that would be taller than the tallest buildings on earth, as we discussed.

Q. All right.

THE COURT: Hold on one second, Attorney Lyne. Hold on one second.

(Pause)

THE COURT: So everyone, just stay where you are. We're going to take -- I mentioned we have a very brief 11:30 hearing. And Mr. Reynolds needs to see if we have the parties available, and we'll quickly interrupt and take that case.

(Off the record at 11:26 a.m.)

(On the record at 11:30 a.m.)

THE COURT: Okay. We'll go back to the hearing Darr v. Alecci and Darr v. Argueta.

And Attorney Lyne, remember to unmute yourself if you didn't already. And please continue.

MR. LYNE: I have, and I'm ready to go.

Can I have the next slide, please?

RESUMED DIRECT EXAMINATION

BY MR. LYNE:

Q. I just want to ask you some questions with respect to the TelexFree dataset. Could you describe for the Court what was provided to you by the trustee or Huron?

A. Yes. So the key records that were provided to me were in four tables. Each of them had a Portuguese name, Representante, Fatura, Transferencia, and Bonus, and these referred, respectively, to account table, invoice table, transfer table, and bonus table.

The account table is what contains the information about the individual accounts, the name, email, phone number, all of the fields that I was referring to earlier, while the other three tables are entirely about transactions, invoices and bonuses, as they say.

And I believe those went into the net equity calculations that I was provided. But after examining them, I determined that they did not provide information about the identity of participants. And so the remainder of my focus was on Representante, the account table.

Q. And can you describe for the Court what the Representante table included?

A. Yes. So --

MR. LYNE: But let me just note, Your Honor, the Representante table has been provided to you as Exhibit 4 on a flash drive. It's simply way too large to do a hard copy.

There's eleven million records here.
BY MR. LYNE:
Q. So continue on the Representante table.
A. Yes. So --
Q. What was in there?
A. So you can think of these tables as having rows and columns. So there's seventeen million, approximately, rows in this. Each one is a separate user account. This is the complete set of user accounts. And forty-four fields; these are the columns. And so we'll get into them later, I believe.
Q. Let me stop you right there. The table key for the columns, I believe, is Exhibit 5 in your exhibit book. Could you just confirm that?
A. Yes, that looks correct.
Q. Okay. All right. So seventeen million user account records, forty-four fields, fields identified by the table key as Exhibit 5. Did you do anything to reduce the set size of the Representante table before you proceeded?
A. Yes. So at the trustee's direction, I focused my attention on the approximately eleven million user accounts, as I mentioned, on the previous slide. These are the ones that were attributable to the TelexFree scheme and had at least one paid invoice.
So there are some other records in that same Representante table that were determined to be part of the

related but different Ympactus, Y-M-P-A-C-T-U-S, scheme that took place primarily in Brazil. That was about four million records of the total, and there were about two million for which there was no paid invoice for the account. And I believe that, not only are they not relevant to the aggregation, because they have no impact on the net equity, but also that was how the Ympactus determination was made based on the currency. And so I was told to -- to filter to the remaining eleven million that were attributable to the --

MR. LYNE: I am frozen.

A. -- scheme in question.

MR. LYNE: I'm sorry, I froze, so I apologize

THE WITNESS: We see you now.

BY MR. LYNE:

Q. Yes. Could you also identify any other tables that were provided to you in connection with your engagement?

A. Yes. So there was an additional table that we called Rep ID Summary, that was provided to me by Huron, and that contained the net equity calculation per user account that I referred to earlier, the -- the number, the dollar amount per user account. And so, as I described earlier, that played a role only in step ten, in adding up those quantities once the clusters were determined.

MR. LYNE: And for the record, the Rep ID table is Exhibit 6, also a flash drive, again, because the actual

physical tables are simply unwieldy.
BY MR. LYNE:
Q. Now, you had talked a bit about the use of a Fellegi-Sunter model. I'd like to have you sort of walk generally -- have the Court walk generally --
MR. LYNE: Strike that. Let me start again.
BY MR. LYNE:
Q. I'd like to have you just walk the Court through roughly, generally, what the Fellegi-Sunter model is. And I believe, if you go to the next slide, there's some information.
A. Yes. So the Fellegi-Sunter method, it's named after the two authors of a 1969 paper, which is one of the foundational papers in the field of probabilistic record linkage. And even today, this model still underlies most of the standard approaches in probabilistic record linkage for large datasets. Obviously, computing power has increased vastly since then, and so there's various extensions, including the one that I used. And -- but it still is -- is at the core of this.
And as I described about probabilistic record linkage in general, the task is to somehow account for the varying degrees of agreement across multiple fields. And -- and so FS provides -- I'll often referred to it as FS as an abbreviation. It often provides -- or it provides a systematic way of taking into account varying degrees of

agreement on these different fields.

And as I said, the output is a match probability for every pair of records under consideration. So close to a hundred percent means the model assesses this pair of records to be a very likely match. Close to zero percent indicates that it considers them unlikely to be a match.

And then one still -- if one is trying to cluster the records, one still needs to perform a clustering step after this. But the output of FS itself is these match probabilities.

Q. If we turn to the next slide, and this is, admittedly, a very truncated example, but could you walk the Courts generally through how you actually do a pair match analysis?

A. Yes. So the idea is to look at all of the fields, all of the columns under consideration, and assess, for that pair, the degree to which it's a match, and then somehow combine that information to obtain a match probability for the pair on the whole.

So in this example, you can see there's some variation in the name, possibly a middle name, or some -- some last name component that is in one and not the other. The email addresses are identical, character for character. The telephone numbers differ substantially, as do the street address and, presumably, the zip, though there are several

digits in common. It could be a transposition error or something.

And so the input to Fellegi-Sunter is information that tabulates the varying degrees of agreement, that the name is close but not exact, the email is an exact match, telephone quite different, and so on. And then it uses a set of parameters that describe the relative weights of the field to figure out how to combine that information, from the records themselves, about how much that particular pair agrees. And the output of -- of the methodology then is the match probability using those parameters and the actual data.

Q. All right. And you said you used Fellegi-Sunter with something called expectation maximization.

MR. LYNE: If you go to the next slide.

BY MR. LYNE:

Q. Can you describe for the Court what is the expectation maximization? Where does it come from? What were its origins? And how do you use it in probabilistic computing?

A. Yes. So a couple decades after the introduction of FS itself, computational power and statistical sophistication had gotten to a point where William Winkler, who was at the time at the U.S. Bureau of the Census, and where he's remained for quite a while, determined how to use this standard technique from computer science, known as expectation maximization, to

determine these relative weights that I was describing from the dataset itself, which is what makes it possible to perform in an unsupervised way on very large datasets.

And so the idea is that, in addition to comparing the contents of each record or each pair of records, the dataset on the whole informs the choice of which fields are more or less important in order to determine distinctions between and match between pairs of records. And because this is an unsupervised method, this is what allows it to be performed without having gold standard labeled data, which is often the case --

Q. Would that be the same term --

A. -- with these large datasets.

Q. Let me just stop you here. Would that be the same term that Mr. -- or similar or an analog to the term Mr. Rona used when he said ground truth sets?

A. Yes. So ground truth data refers to data that you know the answer for already. And so sometimes you have a dataset where some amount is labeled in that way. Also, for testing purposes, you might start with a dataset that you know the answer for and use that to evaluate the method itself.

And in fact, that is what's happened with these techniques. Of course, after somebody has an idea that this might work in an unsupervised way, you of course have to test it on datasets which you know the answer to assess its

reliability. And that's been the case since -- since the introduction of the method. And it's been, of course, tested on increasingly large and sophisticated datasets since.

So shortly after its introduction, Winkler pioneered its use in the 1990 U.S. census. The idea is that some records do refer to the same person, but the dataset is somewhat messy, right? There's all sorts of variations and typos and things. And the U.K. Ministry of Justice has used an extension of -- of FSEM, which I'll use to refer to Fellegi-Sunter with expectation maximization. So quite recently, the U.K. Ministry of Justice used it to link records and criminal court data. And this is generally still one of the -- the most reliable methods for working with large datasets which you don't know the answer in advance.

Q. There's a reference to the Amerado (ph.) paper at the final bullet.

MR. LYNE: I'll just point out to the Court that that's Exhibit 12 in the plaintiff's exhibit book.

BY MR. LYNE:

Q. I'll just use Amerado as an example, but did the paper find -- can you sort of describe what Amerado did to test the accuracy of the FSEM output, unsupervised learning output, against a dataset of some other gold standard or ground truth dataset? Can you just give a sense to the Court as to how that worked?

A. Yeah. So like I described, this is one of many papers in the area which proposes some -- some additions, some improvements and enhancements, extensions of the FSEM method, and then seeks to validate it by testing it on datasets for which we know the answer. And so it does this and finds -- finds its particular extensions to be highly reliable.

And also -- well, maybe two additional points. One, they describe in this paper how the error rate can be assessed from the data itself, in an unsupervised way, and evaluate the accuracy of these error rates that they -- that they calculate in an unsupervised way, but again, against datasets which they know the answer.

And also this is the specific variant. It's very close to this, what I used in my methodology. So the specific extension of FSEM quite closely follows this in the Amerado paper.

MR. LYNE: All right. If I can have the next slide, please. I just want to orient the Court of where we're about to go.

BY MR. LYNE:

Q. So I'd like to spend some time talking about all of the steps leading up to FSEM. So if we could start with the next slide, can you describe, did you do anything to determine the data quality of the overall dataset that you were using in Representante, and describe for the Court what it is that you

did, or why is it important -- let's start with why is it important to assess data quality before starting this process?
A. Yes. So it's important to know before you start what kind of data you're working with, both to determine if it's feasible and appropriate to perform record linkage on the dataset, but also to determine which parts of it to use.

It's entirely possible that you have a dataset in which some fields are quite high quality and others are less so, and the process will be improved by first winnowing out those columns that are not particularly informative or might even confuse the -- the automated methodology, and leave the higher quality ones there.

And another thing that one needs to keep in mind is that there are opportunities for boosting the quality via cleaning and so on. And so, if there are small inconsistencies, those can often be sort of ironed out. And so it's keeping an eye on what's going to be suitable for -- for later on in the methodology.

THE COURT: Unfortunately, Attorney Lyne just went frozen for a moment.

Attorney Papas or Lizotte, do you have the ability to text him to see if he can reconnect?

(Pause)

THE COURT: So it wasn't Attorney Lyne; it was the Zoom that disconnected. But we're getting it back right now.

(Pause)

MR. LYNE: I apologize.

THE COURT: It wasn't you.

MR. LYNE: Can you hear me?

THE COURT: Yeah. It wasn't you, Attorney Lyne. The Zoom call dropped in the courtroom.

MR. LYNE: Oh.

THE COURT: So I think we are --

UNIDENTIFIED SPEAKER: Don't blame us.

THE COURT: -- all set.

Go ahead, Attorney Lyne.

BY MR. LYNE:

Q. Okay. You talked generally -- if you could turn to slide 11 and walk the Court through what this slide represents, that'd be helpful.

A. Sure. It might be helpful for me to first explain some things from 10, because the last thing I described was merely why it's important and not what I did.

Q. Okay. All right.

A. So I don't need to read through all of it, but just in general terms, one needs to figure out the meaning of the fields to determine which are even relevant, because if you have a high quality column that's not relevant to identifying a participant, then it's not relevant.

And then for each field, determine the quality, which can

be done in systematic ways, involving assessing how often data is present, also looking at statistically valid samples. And then there's also questions of how the data columns relate to one another. Are there fields that provide different kinds of complementary information. Email address, phone number, and name are all different and provide a balanced picture rather than tons of different phone numbers or tons of -- of the same thing.

So -- so I said there were forty-four fields in -- in this table of user account records, and I examined each of them, according to what I just said, to determine what they contained and if they were relevant, in what way, and to assess their quality, and in particular, for each of them --

Q. What did you do in that process?

A. Yeah. So like I mentioned a minute ago, there's two general kinds of things that you can do to a column. One is you can perform a systematic test on the computer that looks at all of the data, all, approximately, eleven million values. So that's easy to do for things like how often it's empty, but also more sophisticated things like how often does it -- you know, the phone number contain only digits and the kinds of other symbols that you expect in a phone number or things like that.

And then there's also stuff that you can determine using statistically valid, randomly selected values. And so this is

helpful for things for which human judgment is required. There's not any systematic test that tells you if something is a name or not, but you can select a statistically valid sample and examine it and determine roughly what percentage appear to be names.

And this is important in this dataset because a lot of the fields, even the ones that are quite high quality, you know, they're not perfect. There are typos, there are variations. And so this is a way -- it's a semi-check, but it's also a key way to get a sense of the dataset in ways that are harder to automate, but still doing it in a statistically valid way.

Q. So if you look at, for example, Rep Nom, the name field, could you identify for the Court what it is that you looked at? How did you address the quality issues in that field?

A. Right. So there's -- there's three points that I've indicated here. One is asking how often it's empty. That's a minuscule portion in this case, 0.002 percent. Then there's a kind of character check that I ran on the entire list of eleven million. I believe it was something like does it have five or more characters and all of its characters are among alpha -- alphabetical things that are typically found in a name. I'll note this is actually an undercount in that some of these names are names of businesses, and it's much more

reasonable to have a number, for example, in a business name. But nevertheless, ninety-seven percent passed that character check.

And then as an example of what I meant by random sampling, I took a statistically valid sample from this and visually examined it to determine to what extent it appeared to be a name. So occasionally, though not all that often, it looks like random keyboard matching, right? This is a relatively small percentage of the total. That's a different issue than, say, an occasional variation in whether you include your middle name, for example. And so using that methodology, I determined that approximately ninety-five percent of those values appear to be a name.

Q. We'll get to your random sample, but could you do the same analysis or describe what you did with, for example, next field, Rep Email.

A. Yes. So Rep Email, I did analogous steps. So it's empty close to one percent at the time, which is more than the .002, but still quite a small amount. With email, it's possible to run even more systematic checks because there's standards governing parts of the internet. That's what this RFC number is referring to. This is a document --

Q. And let me stop you there just for the record. The RFC standard is Exhibit 26 in the plaintiff's exhibit book.

A. Yes, I believe this is from the internet engineering task

force, and it describes how email addresses are meant to be formatted, the idea that you have an at symbol, that you shouldn't have spaces in them, but certain other characters are allowed. And so this is the kind of thing that you can quickly run accurately on the full dataset to determine how many pass the standard.

It's a different question from whether that's a currently active email address, of course, but it tells you if it's formatted correctly. And so 98.2 percent passed that standard.

And then there's various other things you can do as a sanity check. Well, over a large majority, in fact, close to seventy-two percent are at Gmail or Hotmail alone. And of course, there's many other domains, but this generally passes the smell test or what you expect email addresses from a number of years ago to look like.

Q. And the same, for example, would be true with respect to the cell phone field?

A. Yes, the cell phone is, of course, a different kind of systematic check. And -- and also, like I said, you keep an eye on how you're going to clean it later. And so if there's variation of whether parentheses or hyphens or things are included, that doesn't really matter much for this stage because you can filter them out later, leaving the numerical content. And so the validation criteria for this involved, I

believe, excluding other symbols, but just looking, is it numerical, plus parentheses, plus hyphen, the kinds of things that are occasionally in phone numbers.

Q. So this chart is color coded. Could you describe for the Court what the color codes are intended to represent?

A. Yes. So blue is Rep ID. This is the unique record identifier. There's no content to speak of in terms of the participant in this number itself, except that it's the unique ID used to refer to the record. I believe that it is in the order in which the accounts were created, but otherwise it provides no information.

THE COURT: This is just chronological, starting with number one?

THE WITNESS: I believe it's chronological. It's not starting with number one, and some numbers are skipped. But -- but yeah, it's a number that goes well into the millions.

THE COURT: Okay. But it's just the number and that's all it is?

THE WITNESS: Yeah, it's just a number. So like a six or seven digit number, typically.

THE COURT: Okay.

THE WITNESS: And we'll see examples of that later.

A. The green fields are the ones for which, upon analyzing what they contained and the quality of their contents, I

determined were high quality and suitable for inclusion in my methodology.

The yellow ones I determined to be of intermediate quality, typically because they were either often empty or if not empty, didn't always contain what they were labelled as. But nevertheless, when they were present, they were more helpful to include than not.

So for example, the date of birth, it's only present about twenty percent of the time, but it is always a date in the appropriate format, and it's a wide range of dates. It's not like everyone always put in January 1st of 1900 or anything like that, and it is useful when present.

And for a lot of these, the way that I determined whether it seemed suitable involved also the connection with other fields, like I mentioned, not just whether it's complementary, but if you look at records that have the same name and the same email, do they tend to have the same date of birth? And the answer is yes, when present, though it's often not there.

But if you ask about phone number, which is much more often present, you know, if you look at records with the same name and the same email, the phone number tends to be the same. Not always, but in a way that seems eminently consistent with the fields representing what -- what they're supposed to be, what they're stated to be, and what -- what I've determined that they're likely to contain.

And then the red fields are ones that I excluded from later consideration for essentially three reasons that are indicated in that legend. So it might have been high quality but not at all related to the participant's identity. So some of these were sort of internal bookkeeping on the server of various sorts or just not related to the participant's identity.

The second reason could be that, it may have been decent quality in itself, but the linkage value was substantially subsumed by higher quality fields. So you'll notice, for address, I include the street name, the house number and the analog of Zip code, depending on what country, but -- but not, for example, the city or state or neighborhood. There's a bunch of other fields.

And so some of those were reasonably high quality, but they're redundant and generally lower quality than the ones that I did include. And this is a sort of standard principle in data science and record linkage also that, when you have substantial overlap, you sometimes get better results by removing some of the redundancy and sort of focusing on more diverse complementary fields.

And then the third reason is if they're mostly empty. In fact, some of these were literally entirely empty. And so of course, that provides no value.

Q. You mentioned that you did some statistical sampling

as well.

MR. LYNE: If I could get to the next page, please.

BY MR. LYNE:

Q. In your reply report, there is an Exhibit 2.1, and it says four hundred randomly sampled values for all twelve chosen data fields. Can you describe to the Court what you did here in terms of how did you how did you pick these? What were you looking for?

A. Right. So the idea is, in a large dataset, especially one as large as this with eleven million records, you have to be systematic in your approach throughout. And one of the key aspects in assessing data quality, a way to be systematic, is to ensure that you're looking at a random sample.

And it turns out that, even with an enormous dataset, the size of the random sample doesn't itself have to be enormous. And as we can discuss later, 400 is actually plenty to get a fairly accurate sense of what the data contains, so long as you're selecting those in a random way.

And so I did this in a -- in a random but -- but reproducible way also. There's ways of random sampling in ways that you can bring up that same random sample for later consideration.

Q. All right. And when you looked at the four hundred randomly sampled values across twelve fields, did you come to any conclusions?

MR. LYNE: And perhaps, if we could, and maybe it will be easier, Ms. Papas, can you actually bring up example 2.1 in a larger format?

Okay. It's still a little hard to read, but it may be easier.

BY MR. LYNE:

Q. If you could just describe what is it that you were generally looking for in these 400 randomly sampled account records?

A. Right. So for the 400 records, you can look at it field by field, and maybe we'll talk about that in a minute. But from looking at it holistically across all twelve fields, you can get a sense of this consistency that I was talking about earlier, and actually, for all the fields, not just one at a time.

So you can get a sense of whether, you know, the different address components tend to hang together and seem broadly consistent with being all part of the same address, you know, the street number appears in -- in street number, the street name and the Zip code, right? Not all of them obviously are easy to inspect by hand, but you can get a general sense of this and see how it hangs together.

And you also see, for example, that the name and email, sometimes people's names are components of their email address, not always, but often. And you see that kind of

thing happening in exactly the way that you'd expect for accurately --

Q. All right. Can I ask you to please turn to the next line, and that would maybe give some color to what you just said.

A. Yeah. So the next slide shows just the name field and email field from that same collection of 400 records, except, whereas the original set on the previous slide was, I believe, sorted by Rep ID, or possibly just in the random order they were sampled, here I've sorted these -- I've sorted the name alphabetically just as the string appears. So typically it's sorted by first character, first name, but you still see patterns emerge.

And the email address I've actually sorted starting from the back character, the final character, which is a little unusual, but draws out in this case how the patterns and the domain names come out.

THE COURT: Can you make that bigger? I can't quite see that yet.

MR. LYNE: Yeah. Can we actually pull up 2.3 and 2.4 as PDFs so that we can scroll through it side by side? That would be helpful.

BY MR. LYNE:

Q. So Dr. Freer, if we scroll down, let's start with the name field in your 400 random samples. Did you determine

whether or not they appeared to be actual names?
A. Yes. So as I described, it seemed that on the order of ninety-five percent were actual names. And you see some variation, of course, in capitalization. Some begin with a capital letter, some don't. Some are in all caps. There's variation in whether people include middle names or more than two tokens, two words in total, or middle initials. And occasionally you see instances of what I was describing as keyboard mashing, where there's sort of random characters, but it's relatively infrequent and broadly consistent with the earlier estimates.

You also see that this is quite consistent with it being essentially never empty, as was analyzed. Also because of the sorting alphabetically, you can see these numerical ones at the beginning, which are presumably not actual names, but again, consistent with what you would expect for figures like ninety-five percent out of four hundred.

And indeed that's the point of a statistically valid sample of an appropriate size that, with relatively small error bars, any phenomenon you see here is broadly representative of how frequent that phenomenon is overall in the dataset.

Q. Could you just walk the Court through the Exhibit 2.4 to your reply, which I think is the email?
A. Yes. So if you look at the very top, you see something

like four blank rows, which is, again, exactly what you would expect if approximately one percent were empty. Those get sorted to the top here. And then you see a small number that don't appear to be properly formatted email addresses, those numbers. And then essentially everything else appears to be a properly formatted email address, which is, again, different from whether there's a typo or it's a currently active account.

And because of how it's sorted by domain, you can also get a sense of this 163.com domain, which I believe is some Chinese email service. And then you see quite a lot for Gmail, quite a lot for Hotmail, some from Yahoo in Brazil.

And -- and again, the email addresses, they generally look like email addresses. Of course, people can and do use random or complicated looking strings, but often it contains name components or some other cute phrase in a way that seems quite typical of email addresses.

Q. So as a result of the data quality review which you performed on the TelexFree dataset, did you come to a conclusion as to whether or not the dataset was of sufficient quality to proceed with your data linkage analysis?

A. Yes. So the conclusion that I came to was that it was of generally high quality, at least after restricting to the fields in question, and that those fields, individually and especially in combination, were more than sufficient to carry

out the linkage analysis.

And it's important to note that they're not perfect, right? This is entirely consistent with these various small inconsistencies and -- and occasional, though infrequent larger ones. But -- but this is more than enough to get started with the cleaning process and then perform the linkage.

And of course, some amount of quality you can tell from the end result. And I think you'll see at the end that often these clusters appear to be considerably higher quality than the input data, which is again a phenomenon that happens via the cleaning and the process actually bringing more structure to it than -- than initially.

Q. You had talked about the statistically valid sample set.

MR. LYNE: Can I have the next slide, please?

BY MR. LYNE:

Q. Could you provide the Court with a bit of a tutorial with respect to what does statistical sampling mean? What is its purpose?

A. Yes. So as I said before, essentially, the purpose of statistical sampling is to enable one to draw valid conclusions about a large dataset, often called the population, from looking at a much smaller dataset, often called a sample.

And in order to do that, one needs to pay attention to how much confidence one needs to have about it. There's always some possibility that, due to bad luck, you're looking at an unrepresentative part, even when you randomly sample. And the point is to choose a sample size that's large enough that that chance of bad luck is actually quite minimized. And it also depends on the degree of -- not just this degree of confidence, but also the margin of error. And so this is standard and statistical sampling methodology for -- for things like surveys and polls and also throughout statistical sampling.

Q. Is there a subdiscipline in the field of statistics referred to as inferential sampling or inferential statistics?

A. I believe so. But this is -- I mean, this is an even more general concept, I'd say, that just when assessing appropriate -- assessing appropriate sample size sort of occurs throughout.

MR. LYNE: Okay. Can I have the next slide, please?

BY MR. LYNE:

Q. And I want to just ask you some questions with respect to what makes a sample statistically invalid.

A. Right. So one of the key ways that you can have an invalid sample is by engaging in what's called selection bias. So this could be because you've sampled the data points, not in a random way, perhaps by looking for a particular pattern

and taking the ones that match that, or choosing them by hand in some way. Or it could even be subtler if you use randomness but it's not fair in some ways.

But -- but one of the key ways that you can get selection bias is if you're aiming to support a desired conclusion. And it's important to note that you can have selection bias even if every single point in the sample does in fact come from the original dataset.

So I've drawn an illustration here. So the population, the large dataset, the full dataset on the left, I mean there's an ellipsis dot, small dot, dot, dot, indicating this is not all 1,005 points. But imagine that you have 1,000 red squares and five blue squares, it's possible, again in the way that I was talking about bad luck that you're trying to minimize, to have a sample with four blue and one red square out of the total. But it's extraordinarily unlikely if this were happening using valid random sampling. And instead, a sample of which you have four blue and one red out of the total is highly unlikely to occur.

And if it were determined in an invalid way using selection bias, then the main point is that you cannot conclude anything about the larger dataset on the basis of that, even though every single point, every square does come from the population. So the sample is, in that case, not representative, and you don't get to do the thing that you

would like to do of drawing conclusions about the whole population on the basis of the sample

MR. LYNE: All right. Next slide, please.

BY MR. LYNE:

Q. So how do you how avoid statistically invalid sampling? What's the process?

A. I would say the two key aspects are to do it in a random way, using fair randomness in your method that you're selecting them, and then also to select them in a suitable number, in a large enough size that that the sample is likely to be representative.

And this is what I was talking about with sample size before. And while any subset of 5 of these 1,005 is possible, it's overwhelmingly likely in this case that, if you were doing it in a valid way, your sample of size 5 would have all red squares. And indeed then you might be able to correctly conclude that the overwhelming majority of the total are red.

MR. LYNE: Can I have the next slide, please?

BY MR. LYNE:

Q. And maybe you could talk about how is it that you calculate how many examples you need to pull before the sample set is statistically sufficient, such that you can draw some general set-wide characteristics from the sample set.

A. Yes. So -- so this chart is showing, on the x axis at the bottom, the population size. So this is the total number

of data points. And then the number of examples is, as we're calling them, from the -- the sample, that's on the y axis, vertically on the left.

And what you see when you -- when you use the standard statistical formulas -- and in this case I've actually run this using Qualtrics the -- the survey firm's sample size calculator. It's an easy-to-use tool that you can try out yourself even. You put in --

MR. LYNE: Let me stop you here. For the record, that is Exhibit 29, a static version of it is Exhibit 29 to the trustee's exhibit book.

A. So -- so this is a generic sample size calculator. Sometimes you can make do with even more specific numbers if you have additional structure in the data. But in a case like this, where we don't know much about the dataset to begin with, using a general purpose tool like this is appropriate.

And what you find is that, for small populations, the ideal sample size -- well, so first of all, you have to specify the confidence and the margin of error, like I said. So the confidence is about, sort of, excluding these instances of bad luck where you see something that's unrepresentative just by dumb luck. And the margin of error is about how accurately you want to be able to estimate whatever quantity it is that you're trying to infer from the small sample about the larger population.

And so the two features I would point out of this is, when the population size is small, the ideal sample size is actually close to the total. If you have just ten data points in total, you have to look at most of them to be confident. If you have a hundred data points, you still have to look at more than half of them.

But it quickly plateaus, or nearly plateaus, and what you find is that between three hundred and four hundred samples, for this stated ninety-five percent confidence and five percent margin of error, which of course you can vary, but between three hundred and four hundred is not only sufficient for getting a fairly accurate idea about what's going on in a population of size ten thousand, but it barely increases as the sample size increases. And so ten million, or even beyond a hundred million, still four hundred samples gives you a pretty clear idea of what's happening so long as they're chosen randomly.

And indeed, this is borne out by what we saw earlier, right? The .8 percent of email addresses that we expected to be to be empty, indeed it is for -- so that would suggest, if the sample size is large enough, you should see .8 percent plus or minus a small amount. And indeed, four out of four hundred is extremely close to .8, as you would expect.

THE COURT: Can I just interrupt for one second? I just want to make sure that I understand what you did, Doctor,

in terms of the 400 records.

THE WITNESS: Yeah.

THE COURT: Did you take 400 records and for all of those 400 records look at all of the different fields that you thought you were going to use? Or is it 400 records the name field, a different random 400 records the email field? So can you explain what exactly are we talking about here?

THE WITNESS: Yes. So this exhibit was -- or this Exhibit 2.1 and 2.3 and 2.4 that we just talked about, that was the former of what you said. It's 400 records overall and then zooming in on the fields for each.

But that's not necessary to the extent that you're interested in assessing field quality in isolation. So I was interested in both, as I said, and for purposes of looking at how the fields hang together, how they relate to one another, it is important that it be 400 records in their entirety or for these twelve fields.

But in order to assess which of the forty-four to use in the first place, I was initially looking at it column by column. So you know, if I determined that one of the columns was overwhelmingly empty, or overwhelmingly not what it stated to be, that's sufficient without looking at the whole context already to exclude that. But then it's important to look at the twelve that I narrowed it down to jointly together.

THE COURT: And the 400 records, are they picked by a

computer or they're picked by you?

THE WITNESS: They're picked by a computer. And like I mentioned earlier, it's not just random, but in a reproducible way. There's a concept in -- in computer science known as pseudo random number generators. The idea is they're -- they're more than random enough for essentially any purpose, including cryptography and really adversarial techniques that are trying to -- to determine differences between that and pure randomness.

But because of how they're chosen in a systematic way with what's called a random seed, it's reproducible. And so one standard technique is to fix the random seed, like 101, for example, you fix that number, and then every time you run that code on that computer, or any similarly set up computer, it will get exactly the same 400 records. And so that's what I did here. That's what I did throughout -- throughout the examples in my report and the reply.

THE COURT: Okay. I just didn't want to get too far in case I was misunderstanding. Thank you.

Attorney Lyne, go right ahead.

MR. LYNE: Sure. If we could get the next slide.

BY MR. LYNE:

Q. And I'd ask you to start by talking generally about the next step, which is cleaning the data in anticipation of doing your deeper calculations.

A. Sure. So the purpose of cleaning, in the context of record linkage, is to generally boost the potential quality of matches without losing much, if any, useful information. And so some general techniques that you might want to apply are, assuming that the case, lowercase versus capitals, is not particularly significant, or as in this situation, where there's considerable variation which doesn't seem to be that meaningful, you might convert everything to lowercase. You're losing a tiny bit of information, but you're also gaining the ability to match cases where people inputted the same name in two slightly different ways.

THE COURT: And this is just the software that's doing this. You're not doing this by hand.

THE WITNESS: Exactly. Yeah. Everything with eleven million records essentially has to be.

THE COURT: Right.

A. So -- right, and this is a very standard technique to convert to lowercase. Another very standard data science technique is to get rid of redundant spaces. You have to be a little bit careful because a space between name components can be informative, but it's totally fair and helpful to take, you know, three internal spaces and collapse that to one, take leading and trailing spaces and trim those. So that -- that I did for all the fields.

Q. And is that a standard tool or practice that's used

in data science?
A. Yes. So all these transformations that I just described, and I'm about to, are totally standard and -- and indeed recommended within data science, especially for purposes of record linkage because, as I said, it has the potential to boost the quality of the record linkage while losing little, if any, of the content.
MR. LYNE: All right. Next slide, please.
A. So there's also additional cleaning steps that I took depending on the field. So while it's almost always permissible and indeed recommended to collapse multiple internal spaces, you might want to get rid of all spaces for an email address because, as I said, the email standard actually does not allow spaces. And yet if somebody had an email address that was entirely valid except for a space they accidentally typed at the beginning, the end, in the middle, you can boost the quality of the linkage by simply removing that entirely.
And also, as I mentioned earlier, phone numbers, there's a variety of punctuation that people use in phone numbers. But if you enter a number just seven digits versus three, hyphen, four digits, if that refers to the same number, one would like to match that. And so the idea for phone numbers is to remove all nondigit characters.
And then also there's some variety in whether people

include area code, country code, things like that. And some phone numbers are only six digits, maybe some a bit longer than seven, other than the area and country code. And there's standards for this that I -- that I confirmed with. But generally for -- for matching purposes, it's therefore sufficient to look at the last seven digits, as I did, so -- the last five to seven digits. And so that allows for phone numbers to match, even if they varied in whether they included the error code, for example.

And similarly with the analog of Social Security numbers, though there one has to potentially leave alphabetical characters. In the U.S., Social Security numbers are all numerical, but in some other countries there are some letters but seldom weird characters or punctuation.

Q. So you've gone through the cleaning process, standardized the data fields in a way that you deemed appropriate and consistent with data science standards. Could you talk a little bit next -- and if I could have the next slide -- about what you're doing with that data?

A. Yes. So as I mentioned at the beginning, the preparatory steps, before determining match probabilities, generally fall into two categories. One involves cleaning and otherwise improving the quality of the data. The other involve reducing the number of pairs of records that even need to be considered by the computer in a more laborious way.

And so just to give you a sense of -- of what's going to happen later, once we've winnowed these records, either by -- or once we've winnowed the pairs of records to the ones under consideration, we still need to consider for every pair whether or not that pair of records is a match.

And so in this example set of records, we have the twelve fields under consideration, color coded as before. And comparing every record to every other record, this is essentially a round robin tournament. And so for four records, there's six pairwise comparisons. The first needs to be compared with the second, third, and fourth. The second also needs to be compared with the third and fourth, and then the third and the fourth need to be compared.

Six is not that much bigger than four, but when you're starting with eleven million, the number under consideration is enormous, in the trillions, like I was saying. And so there's essentially two techniques that I used to reduce this number of comparisons.

One technique involves quickly saying yes to pairs that are identical enough that they're a clear match, and then the other involves quickly saying no to pairs that don't deserve further consideration. And the result --

MR. LYNE: Can we have the next slide, please?

A. Yeah. So -- so the goal of -- of these next two steps is to reduce the number of pairs. So deterministic grouping,

this is a way to get to a quick accurate yes after cleaning. So here we have four records post-cleaning. And if you look at the eight highest quality of the fields -- so as I said, the green are high quality and the yellow are intermediate quality, often not present -- it's a standard technique to perform deterministic grouping early on in the process to reduce the number of pairs.

And this is by reducing the total number of records, not by removing them, but rather clumping them together. So previously, previous to this step, every record was an actual user account, or after step two, it was an actual cleaned user account. So still eleven million of them, just in clean form.

After this step, there's 4.3 million of what I call account groupings, each of which corresponds to one or more user accounts. And the rule --

Q. And those account groupings that you collapsed down to 4.3 million, were they exact -- the ones that you actually collapsed, those were exact matches across all eight fields?

A. Yes. So that's -- that's the rule. So if you look at the eight green fields for a pair of records, if they are identical after cleaning, then it's abundantly clear that they ought to match, right? There's no question, even if the date of birth is present in one and not in another, if you agree on name, email, mobile, telephone, street, house number, Zip, and login character for character -- with one caveat on login, I

should say in a second -- then that's overwhelming evidence of a match. And so by combining them, that's totally fair and in fact is a standard technique. And then we reduce the number of comparisons because we only need to compare between account groupings later.

So two points. One, we're not throwing away any of the information on the yellow. We're just not using it for the exact match. And so there will be zero or more password, zero or more Social Security number analogs, and zero or more dates of birth per account grouping. And they just all appear together as a set for -- for later use in the process.

The other caveat is about login name. So login name is actually unique across the database, as you would expect. You don't want to have two different accounts with the same login name, and this login name and password is how people access the accounts. And so login names are never going to match exactly.

On the other hand, they're essentially always present and quite useful because, as you'll see later, many people did things like their name, followed by one, their name followed by two, and so on. And so that's quite useful information to use. And by truncating it, so -- so I analyzed the patterns. More often this kind of pattern of incrementing by numbers was at the end. Looking at the first four digit -- the first four characters of it provided a nice balance between frequently

matching while still preserving information for the match. And so the login that's being matched here is the first four characters of the login.

So you can view this as a -- this truncation as a kind of cleaning step, but it's different in flavor from the others. And it's important to do before the deterministic grouping. Otherwise, literally nothing will match.

Q. Okay. So you're down to 4.36 million account groupings. What step then followed? I mean --

A. So --

Q. -- what did you do with that information?

A. So --

MR. LYNE: Let's go to the next slide.

A. The next step is the -- the other side of this process that I described, where you give rapid no answers, rather than yes answers, to pairs of records that are clearly not worthy of more detailed consideration.

And so in order to reduce the number of pairwise comparisons there's, again, this very standard technique in record linkage, usually called blocking, sometimes called indexing. There's a bunch of different terminology, but blocking is most standard and what I use.

And so the idea is if you didn't have blocking -- let's say the task is to find puzzle pieces that agree in shape and color. So here we have a bunch of pieces, varying shapes,

varying colors. If you wanted to identify all pairs that matched in this way, you'd think you'd have to look through every pair one at a time. And that's a pretty huge number, even with just these several dozen. And this is what creates the trillions of comparisons with millions of records.

The idea of blocking is to see if there's techniques that are -- that can be run much more efficiently than a full comparison, but such that only if it passes that task is it worthy of further consideration. So if you're trying to match on color and shape, let's say color is a lot more efficient, certainly intuitively, as humans, it is. This is something you might actually do with puzzles. You could first filter based on color and say, okay, I still need to compare every pair, in principle, but for the orange pieces I only need to compare them to every other orange piece. For the blue ones of each shade, I only need to compare them to the other blue ones of that shade. Or maybe you have a more flexible rule which says, I can't quite determine the shades of blue; I'm going to compare all of the blue-ish ones to all of the blue-ish ones.

All of these are legitimate and standard ways of blocking. And so with these four orange pieces, there's six pairwise comparisons, and then there's some number for the blues and so on. There still can be quite a lot, but it's vastly, vastly less than everything pairwise.

Q. If you go through that process --

MR. LYNE: Could I have the next slide, please?

BY MR. LYNE:

Q. What was the result of your blocking analysis?

A. Right. So -- so this chart is showing 9.5 trillion potential pairs. This is already after the deterministic step. The number would have been much larger for the eleven million. This is only just for the 4.3 million, but there's 9.5 trillion combinations, pair -- pairwise combinations of the 4.3 million.

And after running the blocking rules -- I can describe more what I mean by blocking in a moment in this -- not just the puzzle analogy, but the result was 117 million pairs. Now, that still is a large number, but that's now in the realm of operations that, you know, a computer could potentially do. In this case, you know, a large computer with many cores, and a lot of memory, and -- and taking many hours. But this is in the realm of something that's feasible and possible today, whereas 9.5 trillion is a total nonstarter.

Q. Okay. Well, let's talk about what the rules were for blocking. Next slide, please.

A. So there's a number of ways to -- to do blocking. And in the puzzle analogy, I said that you could group together by color, or you could also group together by nearly the same color, give all of the blue-ish pieces a shot against all of

the other blue-ish pieces. And that more flexible approach, it can be more computationally time consuming, but this is appropriate and even more inclusive and fair in cases where there's some variation in the data.

And so the key ingredient to determine when things are close enough on a given field is what's called fuzzy matching. So the idea is if you have some text, a string of characters, and you want to assess how close two of them are to one another, for example, a name and another potential -- or two names, you want to know if they're likely the same name or not. So fuzzy matching, it's typically measured with a score expressed in a percentage where 100 percent indicates total agreement. Like, exactly the same string and zero indicates total disagreement in some form.

And if you look at these examples, so this is using one of the standard string-matching functions, fuzzy -- fuzzy matching for strings that I used. So the first row to Xiomara A Martinez, it's identical character for character. And so the result is 100. For the second row, there's two versions of what you might pronounce as Alice Addams, but in the second one, there's a second D. This doesn't matter whether it's -- whether it's a typo an actually different name.

In any case, this string-matching function, the fuzzy matching says they're are ninety-six percent match. So there's a calculation that goes into it. There's a range

of -- of possible fuzzy matching functions and this is one of the industry standard ones that I used. It's the same function for each of these four rows.

In the third example, Jacqueline D. Fuller (ph.) versus Jacqueline Fuller. They're still quite close, but there's more character differences. There's the D, there's the period, there's an extra space, and the fuzzy matching function says these are a ninety-two percent match. And then the -- the fourth row, they're clearly very different names, but like many long strings, they contain some characters in common.

And so the comparison result is not all that close to zero. It's forty-five in this case, and that's what you typically find if you have long enough strings. A middling result in the thirties, or forties, or fifties, might be indicative of not at all a good match, but it's not literally total disagreement.

Q. All right.

MR. LYNE: Can we have the next slide, please.

BY MR. LYNE:

Q. Maybe if you could walk the Court through this example. I think it would be helpful.

A. So the example here, so this is real data taken from North Carolina voter records, I believe. This is an example that. It's used in a number of surveys, in the literature of

record linkage, and which I use as a running example in the tutorial part of my initial expert report. And so in these seven records, you see a name, a street address, age, sex, and so on.

And the idea is that, without blocking, you'd have to do all twenty-one pairwise comparisons. And with blocking with a particular rule I'll describe in a second, there's only three pairs that need consideration. And this makes intuitive sense, right? If you look at the initial seven, the Dominic Shaw (ph.), one is Jr., one a Sr. They have different ages, but they're worth -- it's worth looking at the details. Similarly, the two Xiomara Martinez, there's different addresses, but the same age, whether or not they're the same, it's worth looking into that.

Virginia Mullinax there is no other record that's remotely worth comparing with, and Jacqueline D. Fuller versus Jacqueline Fuller, the Age is off by two. The address is different. It may or may not be the same, but again, it's worth looking into. But also quite clearly, none of the other eighteen pairs are remotely possible matches. And so the particular blocking rule that I use in this example is using two of the fields.

So blocking can use one or more fields and it's flexible. And in the choice of how you use the fuzzy matching, there's standards in data science for determining which kinds of rules

are appropriate and what thresholds. And that's the sort of techniques that I used in the actual data set.

In this one, just for illustrative purposes, it's with the same string-matching function you saw in the fuzzy matching slide before. And the rule is if either name is above an eighty percent match, or street address is above an eighty percent match in either case, include the pair of records. And indeed --

Q. All right. And if you look at your report, the actual specific rules that you use in blocking would be contained in paragraphs 144 through 168 of your report?

A. That's right. Yeah. So that lays out the specific rules that I use, the specific thresholds, analogous to eighty that I chose, and the reasons why, including documentation of -- of again, statistically valid samples that I looked at to see how those rules would play out on actual records.

Q. Was there a role, a specific role, of name or using name in your blocking step?

MR. LYNE: And if I can have the next slide, please.

THE WITNESS: Yeah. So name did play a special role in blocking. And the kind of challenge here is that it's often the case that you have pairs of records that agree on a lot of fields other than name, but disagree on name.

So one common circumstance could be husband and wife, or other family, or household members. They might have the

same address. Sometimes couples share an email address even, right? And so it's possible to have agreement on many fields other than name disagree substantially on name, and they're clearly different persons. You could also have a business and a person at the same address, but one is the business name, the other is a person. Those are my understanding is to be considered different for purposes of this aggregation, the business and the person, and in these cases, name distinguishes them.

And the reason why is it sounds almost too obvious to state when -- when you spell it out, but we use names to identify people. And so this is a standard challenge in -- in record linkage. I call it a challenge because if you simply clean the data, select the fields, and turn the crank of probabilistic record linkage, without in some way telling the method what kind of entity you're looking for you may not get the result you want. And what I mean by entity here, is let's say you have a database of property records.

This is actually the case in my work at Remine, the real estate startup. And you're trying to identify when two properties are the same. There might be a name field corresponding to the owner or the renter, but what you're interested in doing is matching records that refer to the same address. Or maybe there's a parcel ID, and even if you have two records with the same name, maybe somebody owns more than

one property. That is not evidence that they're the same address. And if they have very different addresses, they're different properties. Even with the same owner.

The reverse situation applies here. If you have two people with very different names but at the same address, still they're to be considered different. But the method, if you don't tell it, has no way of knowing if you're trying to identify persons or trying to identify properties. And so it's important to -- to do what's called disambiguation in -- in the literature.

So the -- the solution to this challenge, which is a very standard solution to the standard challenge, is to disambiguate records that are otherwise similar using name. And I accomplish this in the blocking step, which is again, a standard way to do it before you have to do all the match comparisons. And so the specific rule that's part of -- of my blocking rules is that if the names on two records are very different, that pair will not make it into the blocking set. And therefore it will not be considered for pairwise match. And this is both necessary and appropriate. If you didn't do it, you would actually not get the result that you're looking for. But I should also note that --

MR. LYNE: Can I have the next slide, please?

BY MR. LYNE:

Q. I'm just wondering, is this role of name and

blocking, is this something to which the data science recognizes as a somewhat standard rule in using in a blocking analysis?

A. Yes, it's standard in -- in all the aspects that I was saying. So first of all, this -- this point that I was saying is it sounds almost silly when one says it, but names are used to identify people. You know, personal names are the way that we distinguish between people that are otherwise similar. For example, when I was identifying myself for the record at the beginning, I stated my name, right? This is a very normal and natural thing, but this is understood in this technical sense in the data science literature is in this quote of Kristian (ph.) at the top.

The second quote from a paper by Benjelloun and others is speaking to this point that the purpose matters. The application matters. If you're trying to identify customers as -- as they describe, that's quite different from trying to match products. Even when the fields are -- are similar in -- in the respect of databases. And this final point, which is actually by Winkler of the Census Bureau and the FSEM method, there's -- this doesn't all come out in the quote itself, but if you look at the context, he's actually talking about a situation performing blocking based on name and -- and a very analogous way to how I am. And he explains in similar terms the reason why and -- and also how name provides more

distinguishing power between individuals than the other fields.

Q. All right. So you've gone through the fuzzy matching step, you've disambiguated, you've reduced the blocking set down to something which is manageable. What happens next?

MR. LYNE: And if we could have the next slide.

THE WITNESS: All right. So there's -- there's one more preparatory step before you can actually run FSEM, and that is to go from the -- the data values themselves, the name which is now cleaned and merged in account groupings, and filtered, and blocking, but still a name string.

And assuming the street address and -- and the other fields to go from that to a degree of agreement for each pair of records on each column and that vector of degrees of agreement is then the numerical input to FSEM. And so I'll note that despite this important role that name has to play in blocking, and indeed is crucial for the method to work, still, name is just a field like any other pass at this point. And all the fields play a role in FSEM. And indeed, as I said earlier, with respect to this EM method that makes it unsupervised, it's the EM method that determines the relative weights of the fields. How much of a role name should play once it's done disambiguating.

And so there's an example here just to show what I

mean by -- by turning it into numerical degrees of agreement. So there's -- there's two parts of this. First, you again use fuzzy matching. So fuzzy matching plays a role in blocking, but also in this rubric calculation. This is a collection of vectors and so fuzzy matching for each -- so this is a single pair at the top.

For this pair, we want to create a single row of degree of agreement. So first we perform fuzzy matching using an appropriate matching function for each column. It varies somewhat by column in the way that I described because the meanings are different and they have to be fuzzily matched in a different way. And then having determined -- though in this example, for simplicity, I'm just using the same matching function across all the fields like before. So you see that the name agrees ninety-five percent, street address, thirty-two percent, and so on.

And then a standard second half of such a step is to assign these values. They're called gamma, in the FSEM literature, degrees of map. So one quite standard way is to have one encoding indicated here by three for perfect match. Perfect match can and should count for more than a near perfect match in many cases. A good match in this case above ninety-two percent, an okay match above seventy-five percent. And then zero for anything else.

Now, you can think of this as partial credit, but

again, it's FSEM, the EM step that actually determines how much each of these points are worth. And also, it's important to track missing information separately. So if you look at birth, it's New York State for one of them and missing for the other. Missing, this should be treated differently than total disagreement because there's often missing values in a way that really should be considered something more like, you know, like NA, like there's no -- no comment as opposed to a mismatch. And FSEM will systematically take this input and accordingly treat missing data differently from total disagreement. So the end result in that bottom row is simply these numbers from applying that that gamma rubric.

MR. LYNE: Your Honor, we're just about to turn to the FSEM model, which would be step six. Do you want to take a break for lunch now, or what's your pleasure? You can go or stop?

THE COURT: Well, let me just see what the next slide looks like.

MR. LYNE: Sure.

THE COURT: Okay.

MR. LYNE: No problem.

THE COURT: So let me just ask the doctor a question on that, if you can go to the slide we were just looking at. Not this one, the previous one.

BY THE COURT:

Q. So now, I mean, I assume maybe you're getting there, but so now you have all these at the bottom of that slide. The names, you know, has a value of two street address zero. What number do you need to say that Dominic Shaw Jr. and Sr. either are or are not the same person?

A. Yes. So that's exactly what -- what the next step does, yeah.

Q. That's what we're getting to get to next. Okay.

THE COURT: And that's going to take a long time, Attorney Lyne, is that what you're thinking?

THE WITNESS: I think it's really short.

MR. LYNE: Well, if you are more broadly speaking, two more steps. FSEM, and then there's a series of steps that have to happen to cluster. And there are some slides that explain what clustering is and dynamic clustering.

THE COURT: All right. I'll take your lead. If this seems like it would be a good natural break, we're close enough to 1 o'clock to take a lunch.

MR. LYNE: We can go at 1. I'm happy either way.

THE COURT: Okay. Let's keep going.

MR. LYNE: Stop at 1?

THE COURT: Let's go to 1 o'clock. Yeah.

MR. LYNE: All right. So go to the next step. It's step six, the implementation of the FSEM model. And if you could just give me the next one. Thank you.

BY MR. LYNE:

Q. Could you describe for me what is it that FSEM does?

A. Yeah. So this is exactly what -- what you're asking. This is the step in which essentially the key step in which you go from, for every pair of records, a vector of degrees of agreement to an assessment of the match probability. A number between zero and 100 percent. And this is not yet determining the clusters, but it is saying for every pair of records that has made it this far the degree of match.

And so FS is the sort of general recipe by which these are combined. And as I said earlier, EM is the way that allows FS to determine which fields should count more than others, and even more specifically, which fields with which gamma values should count for more than others in light of the entire data set, and what provides matching value -- what provides sort of matching power.

And so the -- the transformation that happens is from one row per pair of accounts with those numbers to just appending an additional column with the match probability, as is indicated here. So we had 117 million pairs of account groupings going into this step with these numerical scores. And what we have out is the same 117 million pairs that were worthy of consideration. But now with a match probability.

Q. If you go to the next slide, just sort of walk the Court through the waterfall example. This is not an exact --

I will say this is not an example for telling, it's just an illustrative example.

A. Yeah, so this is an illustrative example, but using the same software implementation of FSEM that I did. This is from Robin Linacre This is the software which was used in that Ministry of Justice example last year also. So the idea is that for the pair of records, so if you look right under where it says first name on the right, there's Robn with an I versus Robyn with a Y.

So that's not a perfect match, but it's a pretty good match. And what the method has assessed is that it's a surprisingly good match in the context of a typical pair that made it into blocking. It's enough of a better match than typical that it's going to assign some points in its favor. The green positive amount.

Then for the surname, this is an exact match and that exact match is worth even more positive points is even more surprising compared to a typical pair of records. And then the post code you see there, that's an exact match. That's worth even more points as an exact match because in some sense it's either rare or the EMF is able determined that in the context of how often the record pairs tend to match on the different fields, this is even more likely to indicate that the pair of records is a match. Then for gender, female and male do not match. And so you would expect this to lower the

score. It doesn't lower it by that large an amount, essentially because in this case FSEM must have determined that that kind of mismatch occurs frequently enough in the data that it's not too surprising.

And then what you get out is a final score and you're adding and subtracting numbers for the score. This corresponds to multiplying or dividing probabilities. And so on the right-hand side, this shows the probability in this case ninety-four percent chance of a match. And so this is --

Q. Did the FSEM model deuce all of this weighting represented by this waterfall chart automatically as part of the process?

A. Yes. So essentially what happens is, is the EM algorithm guidance, the FS formula through determining the appropriate weight of each column with gamma value. And once it's determined the appropriate weights in light of what impact that has on all record pairs in blocking, it then runs that one final time for every pair. It does so having determined the weights from EM. Then there's one final pass of FS which runs essentially this waterfall chart with those final numbers for every pair. And so the output is a match probability for every pair.

Q. All right. And does that then complete the FSEM process, that output?

A. Yes, that completes FSEM Cropper said. At this point, we

have a match probability for every pair of records that made it into blocking, but we still haven't determined how to cluster them in total.

Q. Okay.

MR. LYNE: And can I have the next slide, please.

BY MR. LYNE:

Q. Now, just to orient we are at which steps right now?

A. So we just finished step six. And steps seven through nine are the steps in which I formed the clusters from those match probabilities.

Q. Okay. And can you describe for me generally, maybe next slide what it is that you do?

A. Yeah. So there's -- there's a very standard way to go from match probabilities to clusters, which is known as threshold and with transitive closure, which I'll describe in a moment. And so this is -- it's a standard. It's also a very general-purpose technique. And there's lots of -- of different ways of using this concept. And so a lot of the details in steps seven and eight involve the particular application of this.

But the general idea is, so the task is to create clusters such that every record appears in one and only one cluster, and it's a way of partitioning, of dividing up the whole record set. And the inputs that one has are these

pairwise match probabilities. And the intended meaning is that when the match probability is high, you want to increase the chance -- or rather when the match probability is high, it should be the case that that pair of records occurs in the same cluster.

And when the match probability is low, it should be the case that that pair of records is contained in two different clusters. And there's tension between these -- these various numbers, right? It might be that if it were all 100 or zero, then it would potentially be easy, right? You would just grouped together the stuff that the pairs that are 100 and everything else that's forced by those constraints. And then group nothing else together.

But when there's intermediate numbers, the most basic thing you could do is choose a threshold. And indeed this -- this is a standard way to do it. In fact, what -- what I did was somewhat more elaborate, as I'll describe. But -- but just to illustrate the basic concept, if you pick a particular threshold, then you say that pairs of records whose match probability exceeds that threshold get joined together in a cluster and anything else that's forced by that constraint. That's the transitive closure part. You still have to form a cluster. And anything that is not required by that, that ends up in a different cluster.

So just to walk through an example, let's say we have

four records, A, B ,C ,D, the representing account groupings at this point in the methodology. And let's say we have a threshold of eighty percent and what's shown here on the left, we have pairwise match probability of ninety percent between A and B, ninety-five percent between A and C, sixty percent between C and D, and then none of the other dotted edges are present, indicating that those pairs did not make it into the blocking set.

Okay. So we first identify which edges are above the eighty percent threshold in that -- in this case it's the two green ones, ninety and ninety-five, and not the dotted red one. So then we keep the edges that are above the threshold in the third frame, and then finally we form the clusters that are consistent with the edges.

Now, this is what I was calling forced by those constraints, and this is where transitive closure comes in. It's simply not possible to have a clustering by any method whatsoever, to have a set of clusters such that A and B are in the same cluster, A and C are in the same cluster, but B and C are not, right? If you're in a cluster of something else, that's what it means to -- to have such a partition, a set of clusters.

And so by force, by these constraints, what I mean is that B and C are added to the same cluster as each other. D is not required to be in the same cluster as any of the

others, and so it ends up in its own cluster. And so B and C are forced into that cluster by transitive closure. But of course as the threshold changes, depending on what threshold is chosen, there's different things that are -- that are forced in.

And so it's important to choose the threshold in a way that's consistent with the data and the objectives of minimizing error. But -- but any -- any method whatsoever is going to result in transitive closed clusters because clusters have this property.

MR. LYNE: Your Honor, we're just about to go, and just so you know, cluster slides relative to transcript closure and move on then to dynamic clustering. Do you want to stop here or do you want to keep going?

THE COURT: Yeah, we can stop here. I actually had a question, and sometimes some of these things get bore out by cross-examination or later.

BY THE COURT:

Q. But while I'm having it on the brain, when the previous slide -- when you were talking about the EM -- yeah, right there. So are you the one deciding how important first name, surname, postcode and gender are, or does the EM direct that?

A. The latter, EM directs that. And this is what I was talking about as letting the data speak for itself. So

it's -- it's an iterative algorithm. It uses the data to inform an initial -- an initial assignment of the relative weights of the fields and the gamma values. And then it says, what would that look like on the whole dataset? And then it uses that answer to hone in on accurate values.

So there's on the order of a dozen or so iterations, each of which takes a few minutes or a bit longer. Looking at the impact of each of these decisions and the parameters get honed in on more and more precisely as it continues to see which parameters best fit the data.

The -- the intuition for this, which I believe one of the exhibits called the intuition behind expectation maximization, the idea, it sounds a little counterintuitive at first that you could do this entirely from the data without having anything labeled. But the general principle is that in essentially every real data set agreement on some fields is indicative of agreement on other fields.

And conversely, essentially because while many of the records are not perfect matches, you tend to either largely agree on most things or largely disagree. And it's not entirely true. You know, addresses, cities, appear more frequently colliding than the names and so on, but it's different for different fields in a way that allows EM to sort of see the impact of considering which fields are most informative to the process and in that way uses this -- this

fact about data sets, that partial agreement tends to be indicative of more agreement to -- to figure it out.

THE COURT: Thank you. That really -- now I get the slide so let's take a lunch and we'll come back and 2 o'clock.

THE BAILIFF: All rise.

(Off the record at 12:55 p.m.)

(On the record at 2:07 p.m.)

THE BAILIFF: Please be seated.

THE COURT: Good afternoon. Dr. Freer, if you want to come back to the witness stand, please. And you're still under oath. And attorney Lyne, go right ahead.

MR. LYNE: Thank you, Your Honor.

BY MR. LYNE:

Q. Before we continue on clustering, I just want to go back to step six, which was slide thirty-one, for one moment, and that's the waterfall based on match weights and using the match weights to generate a match probability between pairwise sets. And the question I have is at any point in the FSEM algorithm, does any name get any intrinsically higher weight than any other field?

A. No. As I described earlier, it's the EM aspect of FSEM that determines the relative weights of the fields automatically from the data. And once the data gets to this point, name is treated as a field just like any other with the data speaking for itself in the automated EM way.

Q. At any point in this process, do you put your finger on the scale in any way with respect to assigning match weights to individual pairing, pairwise sets.

A. No. Every pair of records is treated in the same way by the method.

Q. All right. We were talking about thresholding for clustering purposes, and we have talked about joining edges based on the threshold of some percent, in this case, eighty percent.

MR. LYNE: Go to the next slide, please.

BY MR. LYNE:

Q. I'd like to have you sort of give a visual example of how the clustering process works.

A. Yes. So the example I've drawn here, so in this analogy, the idea is that apples are meant to be clustered with other apples and oranges with other oranges. That's the goal. But there's a different match probability for some pairs compared to others, which I've illustrated here via nearness physically. So the idea is that the closest match probability pairwise is between the two apples at the top and there's considerably smaller match probability to the apple at the bottom from either of those two. But nevertheless, the goal in this example is to have one cluster consisting of all, and only the apples, and another cluster consisting of the one orange on its own.

MR. LYNE: So if you go to the next slide, please.

THE WITNESS: Right. So the way that I'm going to visualize clusters for this analogy is via a circle, like I did in the previous example. And here the diameter of the circle is going to represent the fixed threshold in question. The idea is because distance in this example is standing in for the match probability, Having a fixed threshold is -- is more or less summarized by having circles of -- of equal diameter. And so fixed threshold means the same diameter for -- for all of them. So on the next slide --

MR. LYNE: Next slide.

THE WITNESS: So on the next line, here's one possible scenario. You might have a cluster that includes all four. This is a fixed threshold because there's only one cluster. All the clusters are the same size in this example, and this is what's sometimes called over clustering. The cluster is more inclusive than is desirable because there's an apple and orange in the same cluster. But the goal was for them to be in different clusters. I can show you a different scenario on the next one.

MR. LYNE: Okay.

THE WITNESS: So here -- so here I have three clusters again, all the same size. And here we have what's called under clustering. So the orange is indeed separate from all of the apples, but the apples are forced into two

separate clusters because of this fixed threshold. So this corresponds to the threshold of the distance are inversely related, right? This is a higher threshold which results in more discerning clusters. And so only the two apples that are closest together with the highest match probability make it in, the ones that have lower match probabilities. Pairwise end up in different clusters. So this is also not desirable. But as I illustrate on the next slide --

BY MR. LYNE:

Q. What is the goal?

A. Yeah. Well, so the goal, the -- the ultimate goal that is to have the apples in one cluster on their own and the orange and another cluster on its own. But these goals are in tension, especially in the context of a larger data set. I mean, I've shown this just for these -- for these four records, but in fact, there's even more complex constraints that come from the way that other records are close to some of these and separate from others. And it's inevitable in any large realistic data set to achieve the goal, only imperfectly. And that's expected and normal that you'll have some amount of over clustering and some amount of under clustering. But still it's important to balance them. Ideally, both are as low as possible, but there's this tension where when you try harder to achieve less over clustering, you inevitably get more under clustering and vice versa.

But there is -- it's not totally zero sum. And so this is showing for fixed thresholds. You try to minimize both sides of this tradeoff, but they have to be in balance. But what I'll show in a moment with dynamic clustering is how you can sometimes do better than either of these by not using a fixed threshold.

Q. So could you describe what dynamic clustering is?

A. Yeah. So dynamic clustering --

Q. In the next line.

A. Yeah, the next line. Yeah. So dynamic clustering, which is sometimes called heterogeneous clustering. The idea is that in some datasets, like with TelexFree, you have some parts of the data set that are more tightly coupled, more tightly bound to each other than in other portions. So you might have one participant who very carefully put in their data exactly the same way each time. We're very close to that.

And then you might have other participants who, for any number of reasons, ended up with more typos, more variations. And ideally, we'd like to accommodate both. We would like to be able to have a rather tight fixed -- well, a rather tight threshold for the portions where there was less variation and then a looser threshold, a more accommodating threshold for the portions of the data set where there was more variability.

And if you're not able to do this, you can still balance

the over and under clustering, but inevitably you're going to do worse in one scenario or the other. And so it's standard in the literature to try to do better than any fixed clustering by -- by choosing a variable threshold in a way that's dictated by the sort of location within -- within the dataset for the -- the portions that are tightly coupled, use a different threshold from the portion that are -- are more loosely coupled. And so I'll get into what --

Q. And did you set a variety of thresholds to accommodate the concept that certain portions of the set would be more tightly coupled, others less tightly coupled?

A. Yes. So this I achieve using dynamic clustering. And so that's described in step eight. Step seven is preparatory for it. The idea is the ingredients that go into dynamic clustering are pre computed cluster sets for each fixed threshold for a wide range of thresholds, some that are quite high, some that are quite low. And then that allows the dynamic clustering method to choose between them in a systematic way, depending on which portion of the data set. So when it finds itself in a section that's better served by a high threshold that chooses that, and when it finds itself in a portion of the dataset that's better served by a lower threshold that chooses that.

So -- so in this step seven, I create 149 sets of clusters, one for each of 149 fixed thresholds. And the idea

is to choose these not directly in terms of match probabilities, but in terms of percentiles of them, so as to achieve broad coverage among the tighter and the looser ones. And especially to even give fine granularity at the extremes. There's interesting differences between 99th percentile and 99.1 percentile. And so that's where these numbers at the bottom row come from.

Q. And so as a result, maybe the next slide would show, at least in the graphic form, gross form, what that would accomplish?

A. Yes. So as I alluded to before, the ideal situation would be what shown on the right where you do better than the over clustered or under clustered scenario by allowing the method to automatically choose a more inclusive threshold for the apples and more exclusive one for the orange.

Q. And is this something that is a acknowledged data science tool?

A. Yes. So this is --

Q. This concept of dynamic clustering?

A. Yes. So this concept of dynamic clustering, also known as heterogeneous clustering is standard in the data science literature. And there's a number of standard ways to perform it. A lot of them are framed in terms of graph theory, an area that I have quite a bit of experience in, especially as it relates to clustering.

And the idea is to essentially develop a notion of over clustering and under clustering that's appropriate to the data set in terms of the -- the weighted graph. So the ingredient to that thresholding with transitive closure slide was a graph as a collection of nodes with weighted edges between them, ninety-five percent, ninety percent and so on. That's input. The dynamic clustering method and the output is clusters in a way that's chosen so as to balance things as on the previous slide. So --

Q. And once you went through that -- sorry. Go ahead.

A. Oh, and so in this particular case, there's -- there's a single parameter that controls the choice of which fixed threshold -- well, so for any, for any dynamic clustering parameter, you're going to have a different fixed threshold chosen in some regions from others. And that one slide, or that one parameter itself can be optimized so as to achieve the desired balance of over and under clustering across the entire dataset as a whole.

Q. And if you look at the graph, could you explain to the to the Court what's that graph represents?

A. Yes. So the, the X axis on the bottom, the minimum over clustering average, that's the technical name for the specific dynamic clustering parameter that I chose is appropriate here. Again, an analog of standard notions in the graph theory literature, and for each of quite a few values of that

parameter, I computed the average over clustering and the average under clustering across the entire data set.

So this is not random sampling. This is actually looking at the full results among the 4.3 million account groupings and assessing what the balance is. And I determined that a parameter value of .60 provides an acceptable tradeoff between the over and under clustering. So inevitably, as in any real large dataset, there will be some amount of clustering, some amount of under clustering. But this achieves an appropriate balance. And actually all of the points here are already a better balance than if I were forced to choose a fixed threshold.

Q. All right. Now, once you've gone through that process, what was the next step?

A. Well, so the result of this process, having determined that parameter is then a single collection of clusters of the account groupings. So we're now to the point that we have a single set of clusters in mind. But these clusters still consist of account groupings rather than user accounts. And so in step nine, I need to reverse the process of step three. This is quite straightforward. This is just a grouping, based on the lookup table that described what happened in step three.

And so here we have that example account grouping which corresponded in -- in the example to three individual user

accounts. So simply expanded that. You need to look at the original account data because the -- the three yellow fields have merely sets at this point. And so to know which corresponds with which.

But all that account information is there and the rule is simply if two account groupings are in the same cluster as each other, then all their constituent user accounts are in the same cluster. And if two account groupings are in different clusters, those get expanded to different clusters of user accounts. So this is exactly what you would expect, to just undo step three.

Q. And at the end of that process, what do you actually have?

A. So at that point, we have a single set of clusters for all eleven million records that are relevant. So at that point, the clustering analysis is itself done.

Q. All right. And then finally, what did you do with that set of clusters once you had gotten through step nine?

A. So then at that point, at the trustee's request, I took the net equity numbers, one per user account that I been provided, and simply tabulated them for each cluster of user accounts. So each user account corresponds to a single participant. It already identifies it in terms of the contact information, but then you simply join, you append the column of the net equity values that -- that I was given and the

total net equity, as I called it for a cluster, is simply the sum of the net equity across each of the user accounts in its cluster. And so this is some dollar amount when it's positive that's declared a net winner. When it's negative, that's declared a net loser.

Q. And if you turn to the next slide would this be a summary by tier?

A. Yes. Yes.

Q. If you could just explain to the Court we're looking at.

A. Yes. So this is showing for that single clustering that resulted in step nine and then using the assignment of net winner and net loser that results from step ten, just stratifying this by tier the by the net gains of the participant zero to 100, $100 to $1,000, and so on. And so looking at this, you get a nice picture overall of what happened with user accounts and what the -- both in terms of the clustering process and information about how the participants behaved.

And there's quite a lot of things that you can notice that are interesting about this, which both pass the smell test and tell you something interesting about the scheme. And if you look at the user account column number three, that's highest in the $10,000 range or so but those clusters tend to have more accounts per cluster and as a result there's fewer

such clusters.

But if you look at the dollar amount, that peaks even higher because even more money changed hands among the largest participants, even though there were fewer of them. And so all of these things sort of check out in the way that you might expect for such a scheme, but it also gives you some interesting quantitative information about it.

Q. And just to -- I've asked you this before, I want to ask you again. At step ten, is the clustering in -- all the way step one through step nine, is it agnostic to the net equity form or net equity table that you applied in step ten?

A. Yes. So it's agnostic to it. It literally doesn't make contact with the -- the net equity inputs. The net equity only comes in at step ten. And even at that point, it's merely adding those numbers. So if you were to give me other numbers, it would be a matter of seconds for the computer to determine the new totals. And but -- but among as you asked the steps, one through nine do not use the net equity in any way either.

Q. Was there any way for you to not simply look at this chart and determine that it appeared just to pass the smell test, but actually try to attempt to determine what the error rate was in your output?

A. Yes. So as I mentioned at the beginning, the Namrata (ph.) paper, whose method is very close to what I performed,

makes a point which is very interesting and quite useful in practice here, which is that not only is the assignment of clusters unsupervised, but so is error rate calculations. So this may sound a little counterintuitive at first, but the method can actually assess its own accuracy, both in terms of false positives and false negatives using the internal data that it's produced.

Essentially, it has much more information available to it than just the final clustering. It has all of the match scores of all of the pairs of records, the match probabilities of the -- of the pairs of records it considered. And so inevitably, as one is balancing over and under clustering, as the method is doing it in this automated way, it will make decisions that it knows based on the match probabilities are suboptimal for one pair of records or another, even though it's for the greater good of having a -- an overall highest possible quality clustering. And so using that match probability information internally, it's able to assess the error rate, both false positives and false negatives.

Q. And could you just give us a quick tutorial on how that false positive, false negative linkage process works and maybe start with the next slide?

A. Yeah. So there's -- there's four notions to describe which -- for which I'll use two of the examples from earlier. So a true positive is a linkage in -- in the assigned

clustering that ought to be there according to the truth of the situation. So here we have a cluster that has all four records and there's three pairwise comparisons among the apples. Those are all correct. And so each of those is a true positive if we go to the next slide. So --

MR. LYNE: The next slide.

A. So false positives, sometimes called type one errors, those are incorrect linkages. There are linkages that are asserted but that are not actually correct. So in the same example, the -- so you'll recall there's six comparisons among four objects, the other three, so we said there's three true positives, the other three are between that orange and one of the other apples. Each of those is a false positive.

BY MR. LYNE:

Q. Okay. And what about negative linkages?

A. Right. So just as -- yeah, so just as false positives were best explained, using that over clustered example, false negatives are best explained using an under clustered example. So here we have the fixed thresholding that was under clustered. So the orange is correctly in a cluster on its own, but the apples are split up. So a true negative is an asserted non-link which is correctly, described as a non-link. So the orange is correctly separated by clusters from all three of the apples. So all three of those are true negatives.

And finally the fourth notion, false negatives, also known as type two errors, those are missed linkages. Those are things that should have been linked but which were not by the cluster. So that's two of the remaining three arrows, of course, the apples on the top. That's a true positive still. But the -- each of the apples at the top to the apple at the bottom, each of those are false negatives, they're missed linkages that should have been there.

Q. All right. How does one calculate an error rate under these circumstances?

A. Yeah. So two of the key error rate calculations and the two that were mentioned in that Namrata quote are what are known as false discovery rate and false negative rate. Those are a measure of false positives and false negatives respectively. And I've written here the formulas in terms of those four notions of true positive, false positive, true negative, and false negative. And from those flow two other quantities, describe a very simple formula as precision is just one minus the FDR, the false discovery rate. And recall is one minus the FNR. So the false discovery rate and false negative rate, those are error rates. You want to have them as low as possible. Also their intention with one another for the same reason over and under clustering. If you try hard to minimize one, the other may come up.

Precision and recall being their opposites respectively.

Those you want to have as high as possible. And those two are in tension for the same reason. And so precision is, roughly speaking, the proportion of matches found that we're correct, and recall is the proportion of correct matches that were found. They're kind of dual notions to one another.

And again, the first is -- is something that's high when there's not many false positives. The second is something that's high when there's not many false negatives. And then finally, it's standard in the literature to combine the two of them, the F1 accuracy notion.

So accuracy is a technical term and here it means this -- this F1 combination of them. This takes into account both false positives and false negatives. This you want to have high, like precision and recall, close to 100 percent and it can't be close at 100 percent unless both precision and recall are high. So this is taking sort of holistic view of both false negatives and false positives and false negatives.

Q. And did you do the calculation for the error rates in the prior aggregation output?

A. Yes. So I assess this using the appropriate analog of Namrata calculations, which again are using the model's own data to assess itself. And using those, I found the false discovery rate to be approximately two percent. So that's a measure of how frequently their match is found that were incorrect. The false negative rate less than one half

percent, which is again, the proportion of correct matches that were not found. So precision, one minus that is ninety-eight percent. Recall one minus the false negative rate, 99.6 percent. And these get combined into an F1 accuracy of .988.

And I'll note that not only does that number seem quite high, as is correct, like it's a very good accuracy score, but in fact, the literature justifies this. And so this front end would quote a very good -- this is actually a study that examines different record matching software to assess which are sufficiently high quality for use in data science. And they consider ninety percent, which this considerably exceeds to be a very good score for such software. And this is in cases where they -- they know the answer and they're comparing it to that.

Q. Could we look at some example output clusters? Did you actually pull some clusters and look at them to determine that they appeared to be populated the way that one would expect given a .998988 error rate?

A. Yes. So it's important to have some -- some grounding in what the method actually produced. And so I'll walk through a bunch of -- of these examples. So these both confirm and are consistent with these assessed error rates. And they're also a nice opportunity to sort of point out some interesting features, including what I mentioned before, that often the clusters are actually considerably higher quality than the

input data. It's able these methods are designed to work in the presence of noisy and messy data and nevertheless achieve excellent results. So there's two kinds of --

MR. LYNE: Can I stop you here and say for the Court's benefit, all of these tables that we're about to look at are Exhibits 1.1 in the plaintiff exhibit book, trustee's exhibit book.

THE COURT: Okay.

BY MR. LYNE:

Q. So you said you pull some randomly for examples of size ten, when you say you randomly selected them, just to remind the Court, what did you do?

A. Yes. So these are selected randomly in this reproducible way using fixed seed for the random number generator. I believe 101 is -- is a standard starting point. And so you fix the seed and then say, give me four reproducible but randomly chosen clusters in this case, among those that are exactly size ten. So this is one way of slicing it. So these are not representative of the dataset on the whole, the representative of the clusters of size ten. But that's nice also to -- to fix that, you can slice and dice it in lots of ways and by restricting to a particular slices of it, you get an even more consistent view of what clusters of that size look like. And there's quite a few of them, right? There's -- not only are there many clusters of size five to

twenty, for example, but of exactly size ten. There still are nearly 36,000 clusters because the dataset is that large.

MR. LYNE: And so next slide, please.

BY MR. LYNE:

Q. Could you just identify this cluster and whatever observations you wish to the advisor?

A. Yeah. So you look at this and this is totally raw user account data. This is not cleaned or in any way processed, other than them having been grouped together by the entire process. And so what you see here is that essentially, all of the first few fields are just identical across the ten of them.

The -- the name, the email, the cell phone, the address components, the login is of course, not, because as we said, these are unique. No -- every record has a different login from everything else. And here you also see a nice example of the kind of pattern that I was describing. Even though they're different as they have to be in this case, it seems like additional evidence, right, the method wasn't looking at how these count up in that sequence. But looking at this, that seems to me additional evidence even that these were created by the same person.

The rep CPF, which is a Social Security analog, Those are not the same as each other, though they agree a fair amount and that was taken into account by the method. They can't all

be correct if this is the same participant. And so this is presumably an example where the data was imperfect and nevertheless a high-quality cluster resulted. Also --

Q. Can I stop -- can I stop you here and just ask, is it possible looking at this randomly selected cluster to identify who the participants?

A. Yes. I mean, if you wanted to know who the person is, you look at the content. And so there's this name, Alexandra Rodriguez. There's email, there's cell, and phone number. The fact that there's two different numbers among the cell phone and the phone number that doesn't make it not identified, that's simply additional information with which one can contact them. And it's also not complete. And nevertheless, it identifies them. It doesn't have the date of birth. This person presumably has an actual date of birth, but nevertheless, this identifies them pretty clearly.

Q. All right. Let's look at the second cluster you randomly pulled.

A. All right. So the second one, we again see that the name agrees on all of them. One of the advantages of -- of this being statistically valid sample is, you know if you look at several of these, I mean, there's only four here, but the randomly chosen, you click at more of them and by seeing how often the name matches exactly, that gives you an accurate sense of how often the name matches exactly in the full

dataset because these are selected randomly. The email you see, there's actually a slight variation at the bottom. G-A-M-A-I-L, presumably a typo in Gmail. In any case, it doesn't match exactly.

And so this is something -- it wouldn't have been cleaned up because it still has no spaces and so on. But the fuzzy matching that's part of the rubric would have at this point notice that they're quite similar and so there would have been partial but not full credit between those email addresses and nevertheless the method matches them together rep and varies. One has CPF Plaza while the others have 111.

The rep login, this again seems an extremely clear pattern, and this is the -- exactly the thing that the truncation of the first four characters was intended. So in this case the truncation would have been A-L-E-X, which matches on the nose in every case. You see that the password and the date of birth occurred just for the first record and not for the others. Nevertheless, they matched. And this is an example also of what I meant at the beginning by helpful when present, though not always present.

In this case it didn't help to link them because it wasn't present for more than one. But it's entirely possible that this date of birth is accurate for this person, in which case it's additional information. But again, the -- the person is quite clearly identified by -- by the information

there.

Also, this is confirmation, as we were saying, of the -- the error rates, right? This is -- it's fairly straightforward to see false positives in this, which there don't appear to be any. And indeed in two samples of size ten, you wouldn't expect to see any typically, given that two percent error rate.

Q. All right. And we turn to the third cluster?

A. Sure. The third one, again, very similar in some respects, but a few -- a few differences. While the name always is identical, except in the last one where the -- there's a missing H and the email, it's agreeing on the first nine, but not the tenth.

Those could be indications that perhaps this is a false positive linkage between that and the others or not. It's hard to tell without the full context. But you look at the other data, the cell and phone, presumably that's not the actual cell or phone. Nevertheless, it provides linkage value. Most people did not put in all zeros. And so this is actually evidence that it's the same person, even though it's presumably not the correct number. Rep N, There's a space in the bottom one that's not in the others and the reps up, the zip code analog varies a little bit.

Again, the login appears very clearly counting up additional evidence. The CPF, they can't all be right if it's

the same participant. This also has an example of a hyphen that would have been removed. And -- and again, some of the fields are missing the password and the date of birth. And nevertheless, this seems extremely high quality.

Q. All right. And going to the fourth?

A. The fourth. I'm not sure there's much to point out that hasn't been already said. The -- the name and email agree exactly as do the cell and phone, which are identical to each other in this case. Maybe it's somebody who only has a cell phone. The login names are again counting up. CPF includes two different values, presumably not both right. But again, it works despite that.

Q. You also look at other clusters besides the four randomly selected clusters of set ten?

A. Yes, a number of them. One I included in the original report, the one whose name was primarily some variant of Maria Nieves, chosen as the very largest cluster by net equity. So there the idea is not that this is representative. Presumably there's all sorts of features that occur with higher or lower prevalence in the highest net equity clusters. Also, this is paying attention to the net equity and how they're selected. So for all these reasons, it's not as representative. But there's interesting things one can determine.

Also, presumably the Court is interested in -- in the high net equity ones and how they look. So here I included

just as sort of additional flavor, the tenth largest, the twentieth largest, and the thirtieth largest. These typically include a large number of accounts because there's such high dollar amounts. And so here on this slide, all you see is a summary, though the full records are included and we can look at them if you'd like.

Q. Well, just when you talk about the count on the right-hand column, because I just want to make sure the Court understands what she's looking at.

A. Right. So the idea is in order to summarize, -- so this this cluster has 444 user accounts and over $2 million in total net equity. And so in order to summarize these 444 rows on fewer than, say, forty pages of information, I've taken the first five fields in question, the name email, both phone numbers, and the main address component.

And again, no cleaning, just looking at exact matches on those five. And this gives you some sense of the variation. So seventy of those 444 rows exactly match that top row on -- on those five fields, then forty-one of them exactly match the next one, and so on. And so those exact matches, I mean, they may or may not match on the other fields, though already this is extremely strong evidence that it's the -- the same person in question.

But then you do -- it draws out some of the variety. And so you see some name variance. There's the presence of Cesar,

or not, Julio (ph.) versus Jolio (ph.). The emails have some variety. There's missing phone numbers and so on. But generally speaking, there's two or maybe three variants that occur throughout this. There's actually four pages. Even with the summary, the most of them near the end have a count of just one despite agreeing in most of the fields, just not in exactly the same pattern still.

So I think this is a nice example of -- of the more loosely clustered thing that I was talking about, where the others could actually tolerate quite a high threshold and still be as high quality as they are. For this to -- to work as well as it does, it had to be choosing the method, systematically chose a more inclusive, lower threshold so as to incorporate this wider variety in values.

Q. And did you do the same thing for the twentieth and third largest clusters?

A. Yes. So same method, the twentieth largest, there's only seventy-eight records despite still nearly $2 million in net equity here. And so here again, there's some name variation. You see multiple middle names or secondary last names being included, or not, or being abbreviated. One of these has -- or actually, I guess, two of the records on one row here have what looks like some phone number or some other sequence of digits place in the email. Nevertheless, they match. There's two different email variants and so on. And the thirtieth

largest --

Q. Let me just stop you here. Is it fair to say that by looking at the -- this particular cluster, the twentieth largest cluster, seventy-eight rows, that it does appear to be a cluster of a individual named Melba A Santos [sic] Estevez (ph.)?

A. Santana Estevez.

Q. Femaleofnaturalmama@hotmail.com?

A. Yes, agreed. I mean, presumably that's their email and not the 2600 number. And again, it seems pretty clearly the name of an individual, even though they're not all spelled exactly the same, and in a similarly, slight variation in the address and so on.

Q. By pulling these examples, were you able to determine, at least in gross, whether the output appeared to be consistent with the error rate of .988 that you had calculated previously?

A. Yes. This seems eminently consistent with those numbers. And you can get a more quantitative sense even from some of the larger clusters. You know, in the case of Maria Nieves, you might expect with two percent false positive rate to actually see some that -- that are perhaps false positives among the 400 odd.

And indeed, there's a few that that do seem more questionable, entirely consistent with that two percent rate.

And -- and also like I said, with the -- with the random samples, you know, it would be troublesome and perhaps inconsistent with those assessed error rates if you started seeing quite a few that seemed out of place in those randomly selected clusters. And indeed all of them seem nearly perfect again, despite the input quality not being --

Q. Sorry, finish your thought.

A. Oh, I was saying, again, despite the input quality being far from perfect and having all manner of inconsistencies, nevertheless, the cluster quality is extremely high.

Q. And just for the record, the Maria Nieves discussion and the Maria Nieves cluster is part of your reply report?

A. That was part of the initial expert report.

Q. The initial one. Okay. So in conclusion, were you able to determine whether the input data was a sufficient quality to use for this engagement?

A. Yes, the input data was of more than sufficient quality.

Q. And were there are a number of -- sufficient number of high-quality fields that allowed you to proceed here with your engagement?

A. Yes, the fields that I selected in blue, and green, and yellow, were more than sufficient and provided useful variety and complementary aspects.

Q. And the steps one through five, the cleaning step and blocking steps, were any of those steps that you performed

inconsistent with generally accepted data science practice?
A. No, they were all quite standard data science techniques. Both suggested, and in fact, crucial in some cases.
Q. And the output of that process was it sufficient for you to proceed with using the validity center with the expectation maximization model to provide metrics for all pairwise combinations?
A. Yes.
Q. All right. And using those match weights, were you able to then provide a cluster which you considered -- or clusters that which you consider to be accurate to within two percent?
A. Yes. And I think that's borne out by both the error rate calculations and the examination of these clusters.
Q. In your opinion, does the Freer aggregation methodology fairly aggregate and link user accounts into clusters in the Telexfree database?
A. Yes, I believe it does.
Q. Consistent with accepted data science practice?
A. Yes.
Q. All right.
MR. LYNE: Your Honor, I would offer both reports by Dr. Freer, which are included as Exhibits 1 and 2 in the Freer exhibit notebook. I will note that in the final plaintiff's exhibit list, there was no objection raised by the class

defendants, and therefore, I believe they should be entered as exhibits at this point in time.

THE COURT: Any objection?

MR. RONA: No, Your Honor. My understanding is that in a Daubert hearing the reports go in for review.

THE COURT: Yes, I will review them. Thank you.

TRUSTEE'S EXHIBIT 1 WAS ADMITTED INTO EVIDENCE

TRUSTEE'S EXHIBIT 2 WAS ADMITTED INTO EVIDENCE

MR. LYNE: All right. And I would now move for the remainder of the exhibits in the plaintiff's exhibit book to be put in, although I will state that there was three objections, 4, 6, and 8, which are the account table, the field description table, and the summary table. The objection raised was subject to review. I believe counsel has all of those. So if the counsel intends to withdraw the objection, I think it would be helpful.

THE COURT: Well, here's the thing, Attorney Lyne, I'm not going to read the Journal of Biomedical Informatics, A Transparent and Transportable Methodology for Evaluating Data Linkage Software, for example, this article. You didn't point to it; I'm not reading it. So you can admit all the evidence you want. But if Dr. Freer didn't talk about it, I'm not looking at it. So is there any objection to any of the exhibits that I have?

MR. RONA: I think Mr. Lyne was referring to --

they've submitted thumb drives with what I understand to be the Burj Khalifa Telexfree data, and we have no objection to that. That's what this case is about. I don't know if I objected to the articles.

THE COURT: I thought you were -- I thought he was speaking -- I thought you were talking about specific exhibits that I have in front of me. Maybe I misunderstood.

THE WITNESS: Exhibit 8 is Frante Boyd (ph.). I don't know if that's what you meant, Mr. Lyne.

THE COURT: Yeah. I just happened to open a tab, and I'm letting you know that, to the extent you want all this to come in, that's fine, but --

MR. LYNE: To answer you, for what it's worth, but to answer the Court's question, the only other objection was to P25, which is a Wikipedia page talking about Portuguese names, on relevance and hearsay grounds. But other than that, there were no objections raised to any of the material contained in the plaintiff's exhibit book. Whether the Court chooses to look at any of those, just for background purposes, is certainly up to the Court.

I'm just simply asking the Court to put all of that into -- to admit the book, as submitted, as exhibits in support of Dr. Freer's methodology and output. And to the extent the Court wishes to look at it, or not wishes to look at a given an article, you would at least simply have them --

THE COURT: Right.

MR. LYNE: -- and have a sense of the support underlying the methodological process generally.

THE COURT: Yes, Attorney.

MR. RONA: Excuse me, Your Honor, really quickly. I don't think we have an objection to 8. And it wasn't touched on by Dr. Freer, so I'm not sure that really matters.

And then my objection to the Wikipedia article is just Wikipedia is dynamic and questionably sourced at times. And so it wasn't clear to me what use it was going to be given, but it looks like Dr. Freer wasn't asked about it. So again, I think it's immaterial, but I'll withdraw the objection.

THE COURT: Okay. So all of the exhibits that I have that appear to end with number 33 are admitted for purposes of this hearing.

TRUSTEE'S EXHIBITS 3 to 33 WERE ADMITTED INTO EVIDENCE

MR. LYNE: Thank you. With that, Your Honor, the direct examination is complete.

THE COURT: Oh, there was some sort of an addendum that was provided during the break.

THE CLERK: Does this exhibit have a number?

MS. PAPAS: We didn't number it, but it would be 34. I just sent it to them last night.

MR. LYNE: All right. It's another article. The

quote happened to have been on one of Dr. Freer's presentation slides and that provides the resource.

THE COURT: So it says Swoosh, a Generic Approach To Entity Resolution. It will be included in the materials.

TRUSTEE'S EXHIBIT 34 WAS ADMITTED INTO EVIDENCE

MR. LYNE: Thank you.

THE COURT: Cross-examination.

MR. RONA: Thank you. Thank you, Your Honor. If I could have a moment to try to reposition myself.

THE COURT: Absolutely.

MR. RONA: Or can I proceed from here?

THE COURT: So Attorney Lyne, if counsel questions from where they're standing right now, are you able to see them on Zoom?

Attorney Lyne?

MR. LYNE: Yes, Your Honor. It was on mute. I apologize.

THE COURT: Can you see them or do you need them to be at the podium?

MR. LYNE: Well, I cannot see the heads, but I can certainly see torsos.

THE COURT: Okay.

MR. RONA: I mean, I can move to the podium, but I would just need a few minutes, Your Honor.

MR. LYNE: You know what? I'm fine. Maybe Mr. Rona,

if he sat, I can at least see him. But it's really up to you, Your Honor.

THE COURT: Why don't I have you stay where you are, Attorney Rona. And if you would like to sit, you may sit, if you like to stand. I can see that Attorney Lyne can see just up to the top of your tie, basically.

MR. RONA: Okay. Well, with Your Honor's permission, I will sit then.

THE COURT: Yes.

MR. RONA: Okay. Thank you.

CROSS-EXAMINATION

BY MR. RONA:

Q. Good afternoon, Dr. Freer, I just want to before I move off of one of the last points you covered, those ten account clusters that you took at random, that we just saw some of them. Do you know if those are net winner or net loser accounts?

A. I don't -- I didn't check.

Q. Okay. You don't know. So you wouldn't have any basis to disagree with me if I said they were net loser accounts?

A. I wouldn't agree either. I'd have to see.

Q. Okay. You'd have to see. Do you know, incidentally, what the average number of accounts for a net loser is as opposed to a net winner?

A. I believe that's in that table of my report, of which you saw one third on the summary slide. But I don't have the number in front of me.

Q. Well, do you know that according to Mr. Dennis' calculation, the average net loser has only six accounts?

A. That seems plausible. I don't know if I've seen that number specifically.

Q. And the average net winner has more than twenty accounts, the average net winner has more than twenty accounts. You agree with that?

A. Again, I'd have to see the number, but it seems consistent.

Q. So by choosing ten accounts, you picked a number closer to the average net loser than the average net winner. Do you agree with that?

A. Closer? I mean, geometric averages are sometimes more relevant than arithmetic, so it seems in between. I also looked at many others there's included in here size forty, for example. You can look at any number.

Q. Well, my question was about the ten that you showed to the Court today. You would agree ten is closer to the six average for net losers than it is to the net winner average that's in the twenties?

A. Again, arithmetically, though I think geometrically in terms of multiplication is probably more relevant. It's

slightly closer to multiplicative also.

Q. And you didn't choose those because they were closer to the net winner -- net loser average; is that right?

A. I chose ten because it fit on a slide.

Q. Okay.

A. Forty fits on a page. So the examples here, it's the most that you can show on a page, and that was the most that I could comfortable with it on the slide, you could easily run up to sixty or twenty.

Q. Okay. And during your testimony, you mentioned that money was changing hands between participants. Do you recall testifying to that?

A. Today?

Q. Today.

A. Not specifically.

Q. Okay. You don't remember saying that money was changing hands between participants?

A. Do you mean physically as opposed to through Telexfree payments?

Q. Well, I'm just --do you remember talking about Telexfree participants exchanging money?

A. My understanding is that that happened in some cases. I don't know too many details on that.

Q. Okay. You don't know that any money changed hands between Telexfree participants based on your analysis. Is

that correct?
A. My analysis didn't involve it. My background of the case, my -- my understanding at the beginning of the case is that there was some money that changed hands directly. So money that changed hands through payments to Telexfree indirectly, some with credits. I don't know many details of any of that, but I'm generally aware.
Q. Okay. But just to be clear that you don't have any actual knowledge of how much money changed which hands?
A. Actual knowledge, I mean, I've been given the tables and I systematically determine that they were not relevant for the cluster analysis, and so I didn't use them further. But I don't know what you mean by actual knowledge.
Q. Well, is that part of your opinion?
A. My opinion is that the three tables that I did not use are not relevant to determining the identities of the participants, and hence I didn't use them.
Q. All right. Well, let's talk about the scope of your assignment.
MR. RONA: I could just have -- I'd like to put Dr. Freer's report, its Defense Exhibit 5, up on the screen.
THE COURT: Who are you asking Attorney Rona?
MR. RONA: Oh, I was -- how do I take control of the video thing?
THE WITNESS: Well, plugging HDMI or something.

MR. RONA: I thought Mr. Reynolds had to flip a switch. But if that's not the case --

MR. REYNOLDS: No Mr. Reynolds told you the other day.

MR. RONA: Okay. Your Honor, if we could just have a maybe a quick recess while I figure this out.

THE COURT: We'll see if IT can come too. Sure.

MR. RONA: Okay. Thank you, Your Honor.

THE BAILIFF: All right.

(Off the record at 2:59 p.m.)

(On the record at 3:27 p.m.)

THE COURT: Please be seated. Attorney Rona, go right ahead.

MR. RONA: Thank you, Your Honor. And I apologize about the technical difficulties.

RESUMED CROSS-EXAMINATION

BY MR. RONA:

Q. Dr. Freer, you put your scope of assignment in your report; is that right?

A. Yes.

Q. Okay. And if we look at your report, this has been marked as Exhibit 1, but it's also Defense Exhibit 5, as well. But if I go to page two of your report, you have a paragraph under your scope of engagement; is that right?

A. Two paragraphs.

Q. Okay. Two paragraphs. And you say that specifically -- well, you say Borelian Corporation, and that's you. Is that correct?
A. Yes.
Q. Okay. Has been engaged by Stephen Darr, the Trustee of Telexfree to develop and implement an aggregation methodology for grouping the user accounts associated with each participant. Do you see that?
A. I do.
Q. Okay. And so do you agree that the scope of your assignment was to identify a participant?
A. Identify a participant?
Q. Well, to identify a participant, that's an alleged net winner.
A. Not a participant. The millions of them.
Q. Okay. Well, do you agree that that -- the scope of your assignment included identifying each participant?
A. Yes.
Q. Okay. You were deposed in this case; is that right?
A. Correct.
Q. Do you recall in your deposition saying that this -- the purpose of your assignment was not to -- was simply to identify a cluster, not to link the cluster to any particular participant? Do you recall testifying to that effect?
A. I recall you saying something to the effect of clustering

individuals or clustering accounts to participants in a way that seemed, to me, inconsistent with the way it works, as there's no predefined list of participants.

Q. Right.

A. It emerges from the data.

Q. Okay. Right. Correct. There is no predefined list, correct?

A. I agree there's no predefined list. And I believe that the part that you're referring to was where I was explaining that.

Q. And so it's your clustering that's going to form the list; is that right?

A. Correct.

Q. And we'll get to some examples. But you don't actually identify if there's variation in the name or even different names. Who within that variation is the participant? Isn't that right?

A. I don't understand the premise of the question. If you have, for example, a name with first name, last name, and another with first name, middle name, last name. I don't understand the sense in which that doesn't still identify the participant.

Q. Okay. Well, let's go to Maria Nieves, who's identified in your report. Is that right? You have her user data tabled in your report; is that right?

A. I believe there's an appendix for which most of the names were some variation number and others. Yes.

Q. Okay. So if we go to page eighty of your report, which is Exhibit 3, you have a cluster --

A. I'm not sure where we're at here.

Q. -- with foster ID 132785. Oh, let me know when you're there.

A. I'm not sure I have it in this binder.

Q. Do you have it on the screen; or no?

A. I can kind of see it, yes.

Q. Okay. So it's cluster 132785. Do you see that?

A. Correct.

Q. Okay. And if I go to -- well, let me stay on the first page. If you look at row ten, I'll try to zoom in a little bit. Do you see a Maria Nieves?

A. Row ten? Yes.

Q. Okay. And now this clustering you refer to in your report as Maria Teresa Milagros Nieves (ph.). Isn't that right?

A. I don't think so. I think I'm careful to refer to it by the cluster ID. That's why I was hesitating a moment ago when you asked if it identified her.

Q. Okay. Well, we'll go -- I'll come back to that in a second. Row ten has a name that has only two tokens, as you call it, correct?

A. Correct.
Q. And it has a different email than at least the majority of the other emails on this page. There's obviously many pages to this exhibit, but at least on this page has a different email address, correct?
A. I believe so though I can't see the full page at the moment, but certainly from the portion displayed.
Q. Okay. And it has a different physical address?
A. Again, yes. For the portion displayed.
Q. Okay. And so what I'm trying to understand is who have you identified as a participant in this case? Is it Maria Teresa Milagros Nieves, let's say, in row nine? Or have you identified Maria Nieves with a different email, different street address, in row ten?
A. The premise of your question suggests that they're different, which they may not be. To the extent that the clusters are accurate, the information should be for the same person. There may be occasional instances of over clustering, but to the -- and indeed, the method error rate suggests there will be occasional such instances. But to the extent that they're accurate there, the premise doesn't even make sense.
Q. Okay. Well, let's go to the next page. Row fifty-seven, there's a Maria Lares (ph.). Do you see that?
A. I do.
Q. Okay. Different email address than the email

address that appears in the majority of other rows. Do you agree with that?

A. For the portion shown, yes.

Q. Okay. And a different street address?

A. Yes.

Q. Okay. So is Maria Lares the person that we're going to assign to this cluster, or is it Maria Nieves, or is it Maria Teresa Nieves? Or is it Maria Teresa Milagros Nieves? How do you know who's the person? The one and only participant to which we're assigning this cluster?

A. Again, there's generally two possibilities. There may be some degree of over clustering, as you would expect, in a large cluster. There also may be variation in name. I would expect that when you're asking is it Maria Teresa Milagros Nieves or Maria Teresa Nieves, the question is, yes, it's two variations of the same name. If it happens that Maria Lares is an instance of over clustering, which again, I couldn't say without looking at more data, but it's possible then that would be an instance of the kind of error in the technical sense that one expects to find in any reasonable and reliable clustering. And so -- but the premise of your question doesn't make sense to the extent that, as in the vast majority of cases, you're asking about variations of the same person. Is it this person or that same person? Yes.

Q. Well, with Maria Lares, you would be willing to

concede that that would be an example of the error of over clustering?
A Not without looking at more information, but it certainly seems more plausible than between, for example, row fifty-two and sixty-one.
Q. And but you would not say looking at this data alone, that you can tell whether the Maria Nieves in row fifty-eight, who has the Maria, the M Maria, Mariashousecleaning@gmail.com, email and the Pleasant Street address is the same or different than Maria Teresa Nieves with the TTMilagros@gmail.com email, who resides on Brackett Road?
A. That one seems intermediate, and I could imagine either way being possible that either its name and -- or not name variation in this case, but some email and address variation. Maybe the person had more than one address over the two years in question. Or maybe it's an instance of over clustering, so that one is intermediate and more difficult to tell just from looking at this.
Q. Okay. You can't tell is the correct way to state it?
A. I can't tell from what you've shown me here, and maybe that in the context of -- it wasn't my charge to investigate every one of the numerous instances of over clustering and under clustering that are surely present as the error rate being nonzero suggests. And so I didn't investigate every

single one. In this case, it might be possible with further investigation, but I certainly couldn't tell you for that intermediate case from what I see here.

Q. Okay. Well, I might agree with you. It's not possible to investigate every case, but did you investigate any case?

A. I didn't investigate this case.

Q. Okay. Did you investigate any other cases of suspected over clustering or under clustering?

A. There were many cases that I looked at in the course of evaluating the method, where in some general sense, I did the kind of thing you're describing here of seeing, are there some that seem pretty clearly over clustering errors? Are there some that seem plausible or not? And you investigate a bit more and then you find in the light of the full cluster or other things in the data set, something that explains why, counterintuitively, it ought to belong there. There's -- there's many instances of that that I looked at in the course of convincing myself that the error rate was broadly consistent with the data. But for any specific case, I couldn't tell you without a computer in front of me.

Q. Well, you don't put in your report, am I right, any cases where you discuss whether particular results are in error or not?

A. Correct. In the report, I don't summarize this except to

say that it's broadly consistent with the error rate.

Q. Okay. And do you have any explanation for how Maria Lares appears in a cluster that is described as being a Maria Teresa Milagros Nieves cluster?

A. I don't understand the second half of the question. You're the one describing it as such.

Q. Okay.

A. Do you mean that is described by you?

Q. Well, you refer to this cluster in your report; do you not?

A. I believe in the second last section or so.

Q. Right. So let me -- and I apologize, I'm using an unfamiliar computer, so everything's going to take me slightly longer to get to where I want to go.

A. Paragraph 217 on page seventy-five.

Q. All right. Thank you, Doctor. All right. Almost there. So you were describing in paragraph 217 the diversity of values --

A. Yes.

Q. -- in a given column. And you describe Maria Teresa Milagros Nieves or Maria Nieves. And then you mentioned that there are many variations.

A. Yes.

Q. But you don't mention anybody with the last name Lares, correct?

A. Not in that list there. No.
Q. Okay. And of these variations, you can't be sure whether any of these variations, for example, Mar Ives (ph.) is a different person or not than any of the other names that are listed?
A. What do you mean by ensure?
Q. You can't be sure that Mar Ives is either the same person or a different person from Maria Teresa Milagros Nieves, correct?
A. In any particular case, I'm not 100 percent sure.
Q. Okay.
A. My expectation is that the vast majority of these, especially when weighted by the number of rows are correct.
Q. Okay. And incidentally, your application of the FSEM model is what assigns the closeness of the match of two names, correct?
A. Not -- well of the names specifically is in the rubric the step before FSEM. FSEM is taking into account all the fields at once. But yes.
Q. Okay. And you don't do any rereview based on any other data to determine whether pairwise matches or the weights accorded to matches have any -- make sense in the real world?
A. I'm not sure what you mean by any other data, but the kinds of clusters I was looking at and showed near the end

earlier today are an instance of looking at the results and -- and confirming to what extent they make sense. I certainly looked at the output and didn't blindly submit it.

Q. Okay. So for example, you don't do any independent research into whether how common or prevalent a name like Maria Nieves is. Is that right?

A. No, that was not my charge.

Q. Okay. And so maybe that name might be unfamiliar to me and seem pretty unique and specific, but that could be a very common name in a different national origin or ethnic community. Is that right?

A. I don't know that would it --

MR. LYNE: Objection. I think it calls for speculation. There's no evidence to support the hypothesis.

THE COURT: Can you rephrase the question?

MR. RONA: Sure.

BY MR. RONA:

Q. You haven't done any research on the prevalence of names based in any type of community, correct?

A. Not explicitly, but as a human who's lived in a particular country, I have enough sense to know that X, G, H, T, is much less likely to be a name, at least in the US, than John Smith.

Q. Okay.

A. And that was part of what went into my determination of

the portion that seemed to be a name. That was my judgment as not an expert in names, but as a person generally familiar with how names are used.

Q. Okay. But you don't -- you didn't check to see how common Maria Nieves is; for example?

A. Common where?

Q. Well, let's say in Massachusetts, how many are there in --

A. No, that's not something I specifically looked at.

Q. Okay. So you've a couple of times said that some somebody can go through a cluster and pick out contact information and can decide -- I don't want to -- I want to understand what your testimony is, that somebody else could decide who's going to be sued. Is that what's going to happen?

A. I don't know too many details about what will happen after me, but certainly I'm not the one suing.

Q. Okay. And you're not the one who's going to do that secondary analysis. Correct?

A. What secondary analysis? I thought you said suing a second ago.

Q. Well let's take Maria Nieves, to your understanding, who's responsible for looking at the cluster and deciding who is the person who's going to be pursued in litigation?

A. Again, the premise doesn't make sense to me. To the

extent the cluster is accurate, it refers to one individual. I grant that there's multiple addresses. I believe in most cases the multiple addresses refer to multiple addresses of that participant. My understanding is that the trustee, for example, would determine whether to mail something to all those addresses at once, or to take the most recent one, or to mail to whichever one they want. That I'm not part of. To the extent the clusters are accurate, it is one individual anyway, and so the premise doesn't make sense.

Q. Okay. But you haven't yourself, it's not part of your opinion to specifically identify by contact information, which is the starting point for the trustee's pursuit and litigation?

A. I've not told the trustee where to start From the data that I gave. I gave as much contact information as I had/

Q. Okay.

A. The full set of it.

Q. So you have not actually identified the specific 81,000 or so net people that your analysis suggests are net winners. You have not pinpointed any one of them.

A. I disagree with that. I've identified them via the contact information. In the vast majority of cases, there isn't even a question about what you're talking about. For example, those clusters of size ten, what you're asking does need to make sense in terms of the variation, because there is

no variation. When there is variation, I provided the full set of addresses.

My understanding is that as in, for example, a credit report, is a credit report invalid because it contains more than one address for that person. That's not my understanding. To the extent that you provided -- or the dataset has a cell phone and another phone number, they're often the same. Perhaps because someone only had a cell phone when they're two different numbers that doesn't unidentified them. That's simply more contact information.

Again, the cluster is not perfect, as no one expects them to be for any large realistic data set, but I believe they're more than sufficient accuracy and in general they'll provide different variants of the contact information for the one individual that identify.

Q. Well, so is there a step after your analysis that's necessary in order to state with any amount of confidence the name of the net winners? Meaning have you provided prepared the list of net winners or is that another step?

A. I've prepared the list of net winners.

Q. You have prepared them?

A. Yes, that is the output of my methodology.

Q. Okay. And so I'm going to go back to Ms. Nieves and ask you if you can identify who is the net winner in your aggregation. And I apologize, there's a lot to get through

here. Who is the net winner in -- I can do this faster. Okay. Who is the net winner in this aggregation?
A. The person with those 471 rows of contact information.
Q. Okay. And that would include Maria Lares?
A. That is a row among them.
Q. Okay.
A. It may be an error, as I said.
Q. And it would also include the Maria Nieves with the different email address. Is that right?
A. Different? I mean, there's more than two email addresses, I would assume. Well, maybe -- maybe there's two.
Q. Well, row 160 is a different one. MariaNieves@mariashousecleaning.
A. Again, I don't grant the premise that that's a different person, that it's identifying. If it's not an instance of over clustering, then it's simply a second email address for the same person. In which case, I guess, I would say yes, it does include that. But again, that depends on the answer to that subsidiary question.
Q. Okay. And is Maria Lares -- is Maria Lares identified as a net winner by you? Row fifty-seven.
A. Again, you're asking is if it's a separate person, it's either an error or the same person.
Q. Okay.
A. If it's the same person, then yes, if it's an error, then

No.

Q. Well, did you go back and look through your analysis and identify errors?

A. Yes. That I talked about for quite a few minutes at the end.

Q. No, I mean, not calculate them based on probabilities. I'm saying did you go back and specifically identify rows of data that are an error, whether over clustering or under clustering?

A. Yes. As I explained to the extent that it seemed necessary to confirm to me that those numbers were plausibly accurate. And I determined that, yes, the -- the stated over clustering and under clustering amounts, the precision recall estimates were broadly consistent with what I saw, looking for actual instances of false positives and false negatives of the data.

Q. Okay. Do you disclose somewhere with respect to this cluster, which ones, which rows are the error? And if there are additional rows not included that were under clustered, do you disclose those rows?

A. Not in the report, the ways we've been discussing, some are obviously more plausibly such than others.

Q. Well, have you done anything to prevent Rita Lares from receiving a summons based on the fact she's in your cluster?

A. I'm not sure I understand the process that will happen after me well enough to know who will receive a summons.

Q. Okay. That's not part of your concern, is that right?

A. My concern is to produce the full set of contact information such that the trustee can issue summons as appropriate.

Q. All right. Let me try and move on. I'm putting it in front of you, this is in Defense Exhibit 5. It's Sub–Exhibit T. This is it. I'm going to try to make this legible. This contains --

MR. LYNE: I'm sorry to interrupt, but you said to Exhibit 5, Sub–Exhibit T, we still on the report?

MR. RONA: No, I've gone to Defense Exhibit 5. And this is an exhibit to my affidavit, which was filed in this case, ECF443. And are you familiar with Benjamin Argueta?

A. To some degree?

Q. Who is Benjamin Argueta?

A. I believe they may have been named in some earlier filings in this case.

Q. Okay.

A. And I recognize the name from one of the larger clusters by Net Equity.

Q. And the cluster ID for Benjamin Argueta is 2522900. Is that right?

A. If that's the Ben Argueta you're referring to? Conceivably, there's other people with that name, in which case they might have other cluster IDs. But that's the cluster you seem to be showing me at the top here.

Q. Okay. And I want to look at whether Benjamin Argueta has over clustering and under clustering. And to do that, I'm going to first just make sure you and I agree on terms. Under clustering would be if Benjamin Argueta had an account and for whatever reason it didn't appear in his cluster. Would you agree with that?

A. No. What you're calling his cluster is presuming some connection of one of the clusters to him and one not, I would say under clustering is when you have two clusters in the methodology that belong to the same participant, in which case it belongs to both of them. It's not that it belongs to, it's not that one belongs to Ben Argueta and the other doesn't. The whole point of under clustering is that it belongs to both.

Q. They both would belong to him.

A. They both belong to him.

Q. Okay. And it would be important to make sure that that doesn't happen. Do you agree with that? To the extent possible.

A. To the extent possible. But again, some percentage of that is expected and unavoidable.

Q. And are you familiar or do you have any awareness of the impact on someone's net equity if there's under clustering?
A. In a specific case, what do you mean by that?
Q. In any case.
A. In the -- the sum of the net equity changes when the list of accounts changes.
Q. Okay. And if the if the person is a net winner, they obviously have positive net equity. Do you agree with that?
A. Obviously. I mean, that's the definition as I understand it.
Q. Yeah. And if they're under aggregated to accounts that have a negative value, then in a sense their net equity is overstated by the under clustering; isn't that right?
A. No. Once again, if there's under clustering, they both belong to them. And so there isn't the one person you get to refer to in one case and not with the other cluster. They're both their cluster.
Q. Okay. Well, let's look at Benjamin Argueta a little bit. The first two rows have a rep login of Maia, Saul, and Maia Saul One. Do you see that?
A. I do.
Q. Okay. And I'm going to -- I apologize because I don't have the all the page numbers here. I'm going to try to

do this by eye.

THE COURT: I know we lost some time Attorney Rona, but I like to stick what I -- to what I tell people. So if we're going to end at 4, would this be a good and natural break because it sounds like you have a series of questions right now.

MR. RONA: Yeah. Okay. I think it's fine to break now, Your Honor. And that might give me a chance to recover from the technical setback. Thank you.

THE COURT: So we're going to recess tomorrow morning at 10 o'clock and see everybody then. Thank you.

(End at 3:54 PM)

* * * * * * * * * * *

I certify that the foregoing is a true and accurate transcript from the digitally sound recorded record of the proceedings.



/s/ Sharona Shapiro October 18, 2023

AAERT Certified Electronic Transcriber (CET-492) Date

ESCRIBERS LLC
eScribers
7227 N. 16th Street, Suite #207
Phoenix, AZ 85020
973-406-2250
e-mail operations@escribers.net

$

$1,000 (1)
111:15
$10,000 (1)
111:24
$100 (1)
111:15
$2 (2)
124:11;125:18

[

[sic] (1)
126:5

A

abbreviated (1)
125:21
abbreviation (1)
45:24
abeyance (1)
9:22
ability (5)
16:2,9;17:6;51:21;
73:10
able (12)
14:3;68:16;69:23;
94:21;105:21,25;
113:18;118:1;126:14;
127:15;128:10;
132:13
above (6)
85:6,6;90:22,23;
98:9,12
Absolutely (1)
132:10
absorb (1)
13:15
abundantly (1)
77:21
acceptable (1)
109:6
accepted (3)
17:25;128:1,19
access (1)
78:15
accidentally (1)
74:16
accommodate (2)
105:20;106:10
accommodating (1)
105:23
accomplish (2)
87:14;107:10
accorded (1)
146:22
according (5)
20:24;30:21;53:11;
114:1;134:4
accordingly (1)
91:10
account (47)
21:6;35:5,7;38:17;
39:8,14;42:8,10,19;
43:8,15;44:4,19,21;
45:21,25;53:10;61:8;
64:8;77:11,12,14,16;
78:4,10;79:8;89:11;
93:20;98:1;109:4,17,
19,24;110:2,5,6,8,20,
22;111:23;116:12;
119:8,25;129:12;
133:15;146:18;154:9
accounts (43)
16:10;21:9;23:24;
24:2,3,3;38:7,14,17;
39:16,17;40:4,6;
42:11;43:9,20;57:10;
77:15;78:14,16;
93:18;109:19;110:1,
7,10,22;111:2,17,25;
124:3,11;128:16;
133:17,21,24;134:5,9,
10,13;138:7;139:1;
155:7,13
accuracy (8)
49:22;50:10;113:5;
116:10,11;117:4,6;
150:13
accurate (14)
14:23;33:2;60:17;
70:12;77:1;100:5;
120:24;121:23;
128:11;141:17,21;
149:1,8;152:12
accurately (3)
56:5;62:2;69:23
accused (1)
15:9
achieve (7)
35:22;104:19,24;
106:12;107:3;108:16;
118:2
achieves (1)
109:9
acknowledged (1)
107:16
across (10)
45:22;60:24;61:12;
77:18;78:13;90:14;
108:17;109:2;111:2;
119:11
acting (1)
15:9
action (1)
20:2
active (2)
56:8;64:7
actual (17)
23:7;44:25;47:11;
63:1,3,15;77:10,11;
85:2,10,16;120:15;
122:18;136:9,10,13;
152:15
actually (53)
5:13;6:25;10:10;
12:8;13:2,24;19:11;
20:15;21:19,20;36:6;
38:6;46:13;54:24;
60:16;61:2,14;62:14,
20;65:12;66:6;69:5;
70:3;74:14;77:17;
78:13;80:12;82:22;
86:19;87:21;88:20,
22;89:9;91:1;99:15;
109:3,10;110:12;
112:22;113:5;114:9;
117:8,16,20,25;121:2;
122:20;125:4,10,22;
126:22;139:15;
149:18
add (2)
12:6;23:10
Addams (1)
82:20
added (2)
39:17;98:24
addendum (1)
131:20
adding (4)
39:21;44:22;95:6;
112:15
addition (1)
48:4
additional (17)
12:16;13:20;18:3;
24:11;39:3;44:17;
50:7;69:14;74:9;
93:19;119:19,21;
120:12;121:24;
122:25;124:1;152:19
additions (1)
50:2
address (47)
8:11;14:13;22:7;
24:25;25:1;38:19,19;
39:3;46:25;53:5;
54:15;56:8;59:11;
61:17,18,25;62:14;
64:6;74:13,15;84:3,
17;85:6;86:1,1,5,24;
87:2,5;89:13;90:15;
92:3;119:13;124:15;
126:13;141:5,8,14,25;
142:1,4;143:10,14,15;
150:5;151:9,16
addressed (2)
6:12;14:19
addresses (19)
39:1;46:23;56:1,15;
64:4,13,14,17;70:19;
84:13;87:2;100:21;
121:9;149:2,3,3,6;
150:2;151:11
addressing (1)
30:10
admissibility (3)
10:2;14:25;20:3
admit (4)
5:24;22:1;129:21;
130:22
ADMITTED (5)
129:7,8;131:15,17;
132:5
admittedly (1)
46:12
advance (5)
15:23;35:12;37:6,
12;49:14
advantages (1)
120:20
adversarial (1)
72:7
adversary (1)
9:25
advise (1)
26:4
advisor (1)
119:6
affect (1)
23:24
affidavit (12)
4:10;6:10,16,24;
7:6,21;8:3,6,8,22;9:4;
153:15
affidavits (2)
4:13;7:2
afternoon (2)
101:9;133:13
again (50)
8:21;12:19;22:9;
44:25;45:6;50:11;
63:16;64:1,6,13;
65:11;67:13;79:19;
84:18;85:15;87:14;
90:2;91:1;103:23;
108:24;112:9;116:6,
21;117:1;120:19;
121:12,24;122:9,24;
123:2,10,11;124:16;
125:19;126:10;127:6,
8;131:12;134:11,24;
141:9;142:11,17;
148:25;150:11;
151:14,18,22;154:24;
155:16
against (4)
18:9;49:23;50:11;
81:25
age (3)
84:3,13,17
ages (1)
84:10
aggregate (6)
14:16;16:10;18:2,
17;21:9;128:16
aggregated (1)
155:13
aggregation (12)
17:8,9;24:2;36:6;
41:2;44:6;86:7;
116:19;128:15;138:6;
150:25;151:2
agnostic (3)
18:11;112:10,12
ago (8)
9:16;16:23,24;28:8;
53:15;56:16;140:21;
148:21
agree (19)
77:23;79:24;85:22;
100:20;119:24;123:7;
133:22;134:10,15,21;
138:10,16;139:8;
142:2;144:4;154:7,
10,22;155:9
agreeable (1)
6:3
agreed (2)
16:23;126:9
agreeing (2)
122:12;125:6
agreement (14)
45:22;46:1;47:4;
82:13;86:2;89:14,16;
90:1,8;93:6;100:16,
17;101:1,2
agrees (3)
47:9;90:15;120:20
ahead (10)
8:13;9:14;24:19;
25:6;32:12;52:11;
72:20;101:11;108:10;
137:13
aiming (1)
67:5
Alecci (2)
4:3;41:20
A-L-E-X (1)
121:15
Alexandra (2)
4:25;120:8
algorithm (3)
95:13;100:1;101:19
Alice (1)
82:20
allege (1)
20:21
alleged (2)
20:17;138:13
allow (1)
74:14
allowed (4)
8:7;17:12;56:4;
127:19
allowing (1)
107:13
allows (6)
37:22;48:9;75:7;
93:12;100:23;106:17
alluded (1)
107:11

almost (6)
13:14,14;74:10;
86:10;88:6;145:16
alone (2)
56:13;143:7
along (1)
7:3
alpha (1)
54:23
alphabetical (2)
54:23;75:11
alphabetically (2)
62:11;63:14
alternatively (1)
4:15
although (1)
129:11
always (11)
30:2;58:5,9,11,22;
61:25;66:3;74:10;
78:18;121:20;122:11
amenable (2)
6:14;9:5
amend (1)
4:16
amended (1)
14:20
Amerado (4)
49:15,20,21;50:15
among (14)
20:5;26:22;41:2;
54:22;107:3;109:4;
112:3,17;114:3,10;
118:17;120:10;
126:23;151:5
amount (18)
14:3;20:8;30:2;
44:20;48:19;55:19;
65:8;70:22;94:15;
95:1;104:21,21;
109:8,9;111:3;112:2;
119:24;150:17
amounts (2)
124:4;152:13
analog (7)
48:15;59:12;75:10;
108:24;116:20;
119:23;122:23
analogous (3)
55:17;85:13;88:24
analogs (1)
78:9
analogy (4)
81:13,23;102:14;
103:3
analysis (22)
12:12;18:18;19:21;
22:3;24:11;33:11,18;
46:14;55:15;64:21;
65:1;81:4;88:3;
110:16;135:25;136:2,
12;148:19,20;149:19;
150:16;152:2
analytics (1)
29:3
analyze (3)
29:4;32:25;33:7
analyzed (2)
63:13;78:22
analyzing (1)
57:24
Andrew (1)
4:23
anticipating (1)
5:23
anticipation (1)
72:24
apart (1)
19:23
apologize (7)
44:12;52:2;132:17;
137:14;145:12;
150:25;155:24
appear (12)
5:20;31:22;54:4;
55:13;64:4;65:10;
78:10;100:21;122:5;
126:4;131:15;154:9
appeared (5)
55:6;63:1;112:21;
117:17;126:15
appears (14)
13:12,21;14:2;16:5,
9,22,22;61:19;62:11;
64:5;96:23;122:24;
142:1;145:3
append (1)
110:24
appended (2)
15:15;23:13
appending (1)
93:18
appendix (1)
140:1
apple (3)
102:21;103:18;
115:6
apples (16)
15:3;102:15,16,20,
24;103:25,25;104:4,
12;107:15;114:4,13,
21,24;115:5,6
application (4)
22:10;88:16;96:21;
146:14
applications (1)
30:9
applied (2)
24:9;112:11
applies (2)
8:19;87:4
apply (2)
18:12;73:4
applying (1)
91:12
appointment (1)
28:24
appointments (1)
28:25
approach (4)
15:12;60:11;82:1;
132:3
approaches (1)
45:16
appropriate (24)
17:9;18:15;30:14;
33:10;34:22;35:9;
51:5;58:10;63:19;
66:16,16;69:16;
75:17;82:3;85:1;
87:20;90:9;95:14,16;
108:2,23;109:9;
116:20;153:7
approximately (7)
40:6;43:7,20;53:18;
55:12;64:2;116:23
area (4)
50:2;75:1,3;107:24
Argueta (10)
4:3;41:20;153:16,
18,24;154:1,6,8,16;
155:20
arguments (2)
7:10;19:19
arithmetic (1)
134:17
arithmetically (1)
134:24
around (2)
22:21;23:1
arrows (1)
115:4
article (4)
129:20;130:25;
131:8,25
articles (1)
130:4
articulate (1)
22:24
artificial (1)
26:14
aspect (1)
101:21
aspects (4)
60:12;68:7;88:4;
127:23
asserted (2)
114:9,22
assess (12)
33:5;46:16;48:25;
51:2;53:13;71:18;
82:8;113:5,18;
116:20,22;117:9
assessed (6)
34:8;36:19;50:8;
94:11;117:22;127:3
assesses (2)
38:2;46:4
assessing (8)
33:12;35:4;53:1;
60:12;66:15,16;
71:13;109:5
assessment (4)
33:8,20;34:4;93:6
assign (3)
90:18;94:14;142:7
assigned (4)
18:6,6;39:7;113:25
assigning (2)
102:2;142:10
assignment (8)
100:2;111:12;
113:2;136:19;137:18;
138:11,17,22
assigns (1)
146:15
associated (1)
138:7
assume (5)
10:3;27:23;39:25;
92:1;151:11
assuming (2)
73:5;89:13
assumptions (4)
15:25;20:8,10;
22:19
attempt (3)
21:11;24:11;112:22
attempting (1)
14:4
attempts (1)
12:14
attention (7)
9:18;11:19;27:15;
28:3;43:20;66:1;
123:21
Attorney (32)
6:5,15;7:14;10:6,
23;11:14,15,16;13:1;
15:7;16:19;24:15;
31:6;41:10,21;51:19,
21,24;52:5,11;72:20;
92:10;101:11;129:17;
131:4;132:12,15;
133:4,5;136:22;
137:12;156:2
attributable (2)
43:22;44:9
author (1)
30:17
authors (1)
45:13
automate (1)
54:11
automated (3)
51:11;101:25;
113:13
automatically (5)
26:12;35:9;95:11;
101:23;107:14
available (3)
9:8;41:16;113:8
Avenue (1)
25:2
average (12)
108:22;109:1,2;
133:24;134:5,8,9,14,
14,22,22;135:3
averages (1)
134:16
avoid (1)
68:5
aware (3)
8:24;10:21;136:7
awareness (1)
155:1
away (1)
78:6
axis (3)
68:24;69:2;108:21

B

back (13)
9:18;26:5;27:11;
41:19;51:25;62:15;
101:4,10,15;140:23;
150:23;152:2,7
background (4)
12:3;13:3;130:19;
136:2
bad (5)
15:9;66:3,6;67:14;
69:21
BAILIFF (3)
101:5,8;137:9
Balan (12)
4:10;6:10,17,24;
7:25;8:2,3,6,8,14,18;
9:9
balance (9)
78:25;104:22;
105:3,25;108:8,17;
109:5,10,11
balanced (1)
53:6
balancing (2)
27:8;113:12
Balan's (1)
7:17
ball (1)
9:14
barely (1)
70:13
bars (1)
63:20
based (18)
8:15;15:13;23:4;
34:3,22;35:5;44:8;
80:13;88:23;101:16;
102:8;109:21;113:14;
135:25;146:20;
147:19;152:6,24
basic (2)
97:14,18

basically (1)
133:6
basis (4)
34:20;67:22;68:2;
133:20
Bayesian (1)
31:14
BBO (1)
22:15
BBOs (2)
22:13,17
began (1)
29:8
begin (3)
36:17;63:4;69:15
beginning (9)
4:20;32:25;63:15;
74:16;75:20;88:10;
112:24;121:19;136:3
behalf (3)
5:2,4,9
behaved (1)
111:19
behind (2)
9:13;100:12
belief (1)
9:6
belong (9)
23:19,24;34:5,8;
144:17;154:14,19,20;
155:17
belongs (4)
154:15,15,16,17
Ben (2)
154:1,16
benefit (1)
118:5
Benjamin (6)
153:16,18,24;
154:5,8;155:20
Benjelloun (1)
88:14
besides (1)
123:13
best (6)
16:2;19:16;35:17;
100:10;114:17,18
better (10)
7:11;18:22;59:19;
94:13;105:5;106:3,
20,22;107:12;109:11
beyond (1)
70:14
bias (4)
66:23;67:5,6,21
bigger (2)
62:18;76:14
billion (1)
40:9
binder (1)
140:8
Biomedical (1)
129:18
birth (10)
58:8,17;77:23;
78:10;91:4;120:15,
15;121:17,23;123:3
bit (13)
12:3;45:3;65:18;
73:9,20;75:2,18;
100:7;107:24;122:23;
140:15;144:14;
155:21
blame (1)
52:9
blank (1)
64:1
blindly (1)
147:3
blocking (32)
79:20,22,23;80:6,
22;81:4,11,12,21,22;
84:5,6,21,23;85:10,
18,21;87:14,17,18;
88:1,2,23;89:5,12,18;
90:3;94:13;95:17;
96:2;98:8;127:25
blue (8)
57:6;67:13,15,18;
80:15,16,18;127:21
blue-ish (4)
80:19,20;81:25;
82:1
blues (1)
80:24
Bonus (2)
42:7,9
bonuses (1)
42:14
book (12)
27:16,23;32:19;
43:12;49:18;55:24;
69:11;118:6,7;
129:10;130:18,22
bookkeeping (1)
59:5
boost (3)
73:2;74:6,17
boosting (1)
51:14
bore (1)
99:16
Borelian (6)
20:4,8,24;28:12,20;
138:2
borne (2)
70:18;128:13
Boston (1)
28:24
both (32)
12:8,8,17;20:20;
21:7;32:2;34:16;51:4;
71:14;87:20;104:23;
105:2,20;111:17,21;
113:5,19;116:12,15,
16;117:21;123:11;
124:14;128:3,13,22;
154:15,18,19,20;
155:16,19
bottom (11)
35:14;40:3;68:25;
91:11;92:2;102:22;
107:7;108:21;115:7;
121:2;122:22
bound (1)
105:14
box (1)
5:11
Boyd (1)
130:8
Brackett (1)
143:11
Brain (3)
25:16,23;99:19
Brazil (2)
44:2;64:12
break (6)
10:19;91:15;92:17;
131:21;156:5,7
brief (5)
11:14,16;17:2;19:1;
41:14
briefing (1)
15:13
briefly (1)
8:14
bring (5)
9:18;11:19;13:8;
60:21;61:2
bringing (2)
16:25;65:12
brings (1)
10:3
broad (1)
107:3
Broadly (10)
34:12;36:17;38:21;
61:18;63:10,20;
92:12;144:19;145:1;
152:14
brought (2)
9:5;29:3
budget (1)
4:17
build (3)
29:7;33:2,8
buildings (1)
41:8
bulk (1)
40:8
bullet (2)
35:15;49:16
bunch (4)
59:14;79:21,25;
117:21
burden (2)
10:2,5
Bureau (2)
47:23;88:20
Burj (3)
15:19,21;130:2
business (6)
24:24;25:1;55:1;
86:4,5,8
businesses (1)
54:25

C

calculate (4)
50:10;68:21;115:9;
152:6
calculated (1)
126:17
calculation (6)
18:8;44:19;82:25;
90:4;116:18;134:5
calculations (7)
40:16;42:15;72:25;
113:3;115:11;116:21;
128:14
calculator (2)
69:7,12
call (5)
24:15;52:6;77:13;
86:13;140:25
called (23)
19:19;21:20;36:2;
39:18,19,20;44:17;
47:13;65:23,25;
66:23;72:11;79:20,
20;82:6;87:9;90:18;
100:12;103:16,24;
105:11;111:1;114:7
calling (3)
69:2;98:15;154:11
calls (2)
24:17;147:13
Cambridge (1)
25:2
came (5)
22:24;30:5,6;33:13;
64:22
Cameron (5)
4:6;18:23;24:17,22;
25:1
can (130)
5:6,7,10,10;7:1,2,
11;9:19;14:18;15:20;
16:1,1;23:9;24:23;
25:18;27:12,20;
29:20;35:5,6,13,19,
21;36:8,18;38:11;
40:7;41:24;42:20;
43:6;46:20;47:16;
49:21,24;50:8,17,23;
51:16,22;52:4,25;
53:16,17,24;54:3;
56:4,11,24;60:6,16,
21;61:2,10,13,16,21;
62:3,18,20,21;63:14;
64:9,14;65:8,16;
66:18,22;67:4,6;
68:18,22;69:7,13;
70:10,24;71:6;73:20;
74:17;76:23;79:4;
80:7,24;81:11;82:2;
83:19;84:23;85:19;
87:23;89:9;90:21,25;
91:15,23;92:19;96:5,
12;99:15;103:19;
105:4,25;108:16;
111:20;113:5;118:4,
21;120:4,4,13;
123:23;124:5;126:19;
129:21;132:11,18,20,
23;133:1,5,5;134:19;
135:7;137:7;140:10;
143:7;147:15;148:11,
12;150:24;151:1;
153:6
capital (1)
63:5
capitalization (1)
63:4
capitals (1)
73:5
caps (1)
63:5
careful (2)
73:20;140:20
carefully (2)
27:8;105:15
Carolina (1)
83:24
carry (2)
33:18;64:25
Case (85)
4:2,3;11:3,3,7,9,12;
12:13;14:14;15:13;
16:8;17:17;18:20;
19:25;20:1;21:13,15;
22:2,11,20;23:6;
24:10;26:24;34:2,6,
11,23;36:18;37:18;
41:16;48:11;49:1;
54:19;62:16;67:24;
68:14;69:5,14;72:19;
73:5;81:16;82:23;
83:13;85:7,22;86:19;
90:22;95:2,9;97:4,7;
98:10;102:8;108:11;
118:17;119:18;121:3,
15,16,21,24;123:9;
126:20;130:3;136:3,
3;137:2;138:19;
141:11;143:14;144:1,
3,5,6,7,20;146:10;
151:17;153:16,20;
154:3,14;155:4,5,18
cases (16)
21:18;34:16;35:11;
73:10;82:3;86:8;
90:22;117:13;128:3;
135:22;142:23;144:8,

10,23;149:3,22
categories (2)
31:14;75:22
category (1)
18:10
caveat (2)
77:25;78:12
cell (11)
56:18,19;119:13;
120:9,10;122:17,18;
123:8,9;150:7,8
Census (3)
47:23;49:5;88:20
Center (7)
17:16;22:22;25:18,
20,22;26:18;128:5
certain (4)
36:21,22;56:3;
106:10
certainly (11)
15:3;16:16;21:16;
80:11;130:20;132:21;
141:7;143:3;144:2;
147:2;148:17
Cesar (1)
124:25
chalk (2)
12:21,22
challenge (6)
21:22;85:21;86:12,
13;87:11,12
challenges (1)
21:24
chance (4)
66:6;95:9;97:3;
156:8
changed (5)
112:3;135:24;
136:4,5,9
changes (3)
99:3;155:6,7
changing (2)
135:11,17
character (12)
46:23,23;54:20;
55:2;62:12,15,15;
77:25,25;82:18,18;
83:6
characteristics (1)
68:23
characters (12)
54:22,22;56:3;63:9;
74:24;75:12,14;
78:25;79:3;82:7;
83:10;121:14
charge (2)
143:22;147:7
chart (6)
57:4;68:24;81:5;
95:11,20;112:21
check (7)
54:20;55:3;56:12,
20;112:5;133:18;
148:4
checked (1)
22:15
checks (1)
55:20
chief (1)
29:2
Chinese (1)
64:11
choice (4)
17:24;48:6;84:24;
108:12
choose (8)
66:5;97:15;99:6;
106:18;107:1,14;
109:11;135:2
chooses (4)
8:9;106:21,23;
130:18
choosing (4)
67:1;106:4;125:12;
134:13
chose (5)
24:8;85:14;108:23;
125:13;135:4
chosen (9)
60:6;70:17;72:10;
99:4;108:8,15;
118:17;120:23;
123:17
chronological (2)
57:12,14
circle (2)
103:3,5
circles (1)
103:8
circumstance (1)
85:24
circumstances (2)
17:10;115:10
cities (1)
100:21
city (1)
59:13
Civil (1)
4:14
clarified (1)
16:13
class (9)
4:5,12,17,19;5:2,4;
14:9;20:2;128:25
clean (4)
22:14;56:21;77:12;
86:14
cleaned (5)
37:15;77:11;89:11;
119:8;121:5
cleaning (15)
36:24;37:17;51:14;
65:6,12;72:24;73:1;
74:9;75:15,22;77:1,
21;79:5;124:16;
127:24
clear (12)
7:9;8:22;15:20;
16:7,21;37:5;70:16;
76:20;77:21;121:12;
131:10;136:8
clearly (14)
13:18;14:11,13;
27:21;37:8;79:16;
83:9;84:19;86:4;
120:16;121:25;
122:24;126:10;
144:13
CLERK (12)
4:2;5:5,13,17;
10:14;11:5,9,11;19:6;
24:23;25:4;131:22
click (1)
120:23
client (1)
15:5
close (18)
46:3,5;47:5;50:13;
55:18;56:12;70:3,23;
82:6,8;83:5,12;92:17;
104:17;105:16;
112:25;116:14,15
closed (2)
27:2;99:9
closely (2)
29:10;50:15
closeness (2)
35:5;146:15
closer (5)
134:14,16,21;
135:1,2
closest (2)
102:19;104:5
closure (12)
23:16;26:23,24;
27:1,9;38:9;96:16;
97:22;98:16;99:2,13;
108:4
clumping (1)
77:9
cluster (94)
18:4,10,18;20:16,
18;27:2,3,4,4,5;33:5;
38:13,14,16,16,20,22;
39:7,15,16,18;46:7;
92:14;96:3,24;97:5,
21,23,24;98:19,19,20,
24,25;99:1,2,12;
102:23,24;103:13,15,
17,18;104:12,13;
106:15;110:6,8,21;
111:1,3,25;114:2,20;
115:4;119:5;120:3,5,
17;122:8;123:17;
124:11;126:3,4,5;
127:10,12;128:10;
136:12;138:23,23;
140:4,11,21;142:7,10,
13;144:15;145:3,4,9;
148:11,23;149:1;
150:11;152:18,25;
153:24;154:3,4,10,11;
155:18,19
clustered (8)
102:15;107:13,13;
114:17,18,20;125:9;
152:20
clustering (79)
27:8;38:21;46:8;
92:15,15;98:17;
99:13;101:14;102:7,
13;103:16,24;104:21,
22,24,25;105:4,7,8,
10,11;106:1,4,12,15,
18;107:19,20,21,25;
108:2,2,7,13,17,22,
23;109:1,2,7,8,9;
110:16;111:11,18;
112:9;113:9,12,17;
114:1;115:23;138:25;
139:1,11;140:17;
141:18;142:12,17,21;
143:2,16,23,24;144:9,
9,13;151:16;152:9,9,
13,13;154:6,6,8,13,
17;155:3,15,16
clusters (67)
18:6,9;21:3,5;27:6;
38:7,15;39:6,11;40:4;
44:23;65:10;93:8;
96:10,15,23;97:8;
98:13,18,22;99:9,9;
103:3,15,19,23;104:1,
4,7;106:25;108:7;
109:16,18,18;110:9,9,
14,18;111:24;112:1;
113:3;114:23;117:15,
16,25;118:17,20,23,
25;119:2;123:13,14,
20;125:16;126:20;
127:5;128:11,14,17;
133:15;141:17;
146:25;149:8,24;
153:22;154:12,13
code (9)
59:12;61:20;72:14;
75:1,1,3,9;94:19;
122:23
coded (2)
57:4;76:7
codes (1)
57:5
Cognitive (3)
25:16,23,25
collaborators (1)
26:3
collapse (2)
73:22;74:11
collapsed (2)
77:16,18
collection (5)
30:5;62:7;90:4;
108:5;109:16
colliding (1)
100:22
color (10)
57:4,5;62:4;76:7;
79:25;80:10,10,13;
81:24,25
colors (1)
80:1
column (13)
52:23;53:16;71:19,
20;89:15;90:9,10;
93:19;95:15;110:24;
111:23;124:8;145:20
columns (7)
43:7,10,12;46:16;
51:10;53:3;71:20
combination (2)
64:25;116:12
combinations (3)
81:9,9;128:7
combine (3)
46:17;47:8;116:9
combined (2)
93:11;117:4
combining (1)
78:2
comfortable (1)
135:8
coming (2)
8:3;15:11
comment (1)
91:8
common (8)
21:18;47:1;83:11;
85:24;147:5,10;
148:5,6
community (2)
147:11,19
companies (1)
28:19
company (1)
29:8
compare (9)
29:5;37:8,11;40:20;
78:4;80:13,15,16,19
compared (7)
40:23,25;76:11,12,
13;94:18;102:17
comparing (4)
48:4;76:8;84:16;
117:13
comparison (2)
80:8;83:12
comparisons (12)
37:4,17;76:10,18;
78:4;79:19;80:5,23;
84:6;87:16;114:3,10
competency (1)
8:12
compilations (2)
12:12;14:12
complementary (4)

53:5;58:15;59:21;
127:23
complete (4)
43:9;95:23;120:13;
131:19
completed (1)
39:6
completely (2)
21:25;40:25
completes (1)
95:25
complex (1)
104:16
complexity (1)
41:5
complicated (1)
64:15
component (4)
19:12;26:22;46:22;
124:15
components (7)
19:12;26:14;61:17,
24;64:16;73:20;
119:14
computation (1)
21:12
computational (1)
47:21
computationally (3)
35:13;37:6;82:2
computed (2)
106:15;109:1
computer (15)
25:24;30:22;47:25;
53:17;72:1,2,4,14,14;
75:25;81:15,16;
112:16;144:21;
145:13
computers (1)
40:15
Computing (17)
17:16,19,20,21,23;
20:8;25:15,18,21;
29:11,16,17;31:16;
35:24;37:11;45:17;
47:19
concede (1)
143:1
Conceivably (1)
154:2
concept (8)
27:2;66:15;72:4;
96:19;97:18;106:10;
107:19,20
concern (2)
153:3,5
conclude (2)
67:22;68:17
conclusion (5)
33:13;64:20,22;
67:5;127:14
conclusions (4)
22:7;60:25;65:23;
68:1
confidence (6)
66:2,8;69:19,20;
70:9;150:17
confident (1)
70:4
confirm (4)
5:6;43:13;117:21;
152:11
confirmation (1)
122:2
confirmed (1)
75:4
confirming (1)
147:2
confuse (1)
51:11
connection (5)
32:23;33:8;44:16;
58:14;154:12
connections (1)
25:24
consider (4)
34:17;76:4;117:11;
128:11
considerable (1)
73:7
considerably (4)
65:10;102:21;
117:11,25
consideration (13)
7:8;46:3,16;59:2;
60:22;76:4,7,15,22;
79:17;80:9;84:8;
93:23
considered (8)
40:12;75:24;86:7;
87:6,19;91:7;113:11;
128:10
considering (2)
38:3;100:24
considers (1)
46:6
consist (1)
109:19
consisted (1)
36:18
consistency (1)
61:13
consistent (19)
58:23;61:18;63:10,
12,16;65:3;75:17;
98:14;99:7;117:22;
118:23;126:16,18,25;
128:19;134:12;
144:20;145:1;152:14
consisting (2)
102:23,24
consists (4)
36:2;38:16,18;
39:16
constituent (1)
110:7
constraint (1)
97:21
constraints (4)
97:12;98:16,23;
104:16
consulting (2)
28:20,22
consuming (2)
37:12;82:2
contact (13)
38:17;39:12;
110:23;112:13;
120:13;148:11;
149:11,15,22;150:10,
14;151:3;153:5
contain (5)
33:15;53:21;58:5,
25;83:10
contained (8)
32:18;38:15;44:19;
53:12;57:25;85:11;
97:7;130:17
contains (5)
42:10;60:17;64:15;
150:4;153:11
content (4)
56:25;57:7;74:7;
120:8
contents (2)
48:5;57:25
context (10)
17:19;33:25;71:22;
73:1;88:22;94:12,22;
104:14;122:16;
143:22
continue (3)
41:22;43:3;101:14
continues (1)
100:9
continuing (1)
31:25
contractor (1)
29:8
contrast (1)
35:3
control (2)
22:2;136:23
controls (1)
108:12
controversial (1)
7:3
conversation (1)
15:11
conversely (1)
100:18
convert (2)
73:8,18
convincing (1)
144:19
co-organized (1)
31:19
copies (1)
12:19
copy (3)
27:23;41:7;42:25
core (2)
37:25;45:19
cores (1)
81:16
corner (1)
30:3
Corporation (5)
20:4,25;28:12,20;
138:2
correctly (5)
56:9;68:16;114:20,
22,23
corresponded (1)
109:25
corresponding (1)
86:22
corresponds (5)
77:14;95:7;104:2;
110:4,22
costs (1)
4:18
counsel (10)
4:21;5:5,12;7:14;
14:9;19:6;27:23;
129:14,15;132:12
count (6)
90:21;93:12,14;
119:20;124:7;125:5
counterintuitive (2)
100:13;113:4
counterintuitively (1)
144:17
counting (2)
122:24;123:10
countries (1)
75:13
country (4)
59:12;75:1,3;
147:21
couple (5)
16:23,24;30:23;
47:20;148:10
coupled (5)
105:13;106:6,8,11,
11
couples (1)
86:1
course (18)
33:15;35:13;48:23,
24;49:2;56:8,14,19;
59:24;63:4;64:14;
65:8;70:10;99:3;
115:5;119:14;144:10,
18
Court (142)
5:10,19,22;6:11,15,
19,22;7:13;8:10,13;
9:3;10:8,8,15,15,20;
11:1,6,10,13,24;
12:21;13:1,11;14:20,
24;15:7;16:3;17:4,12;
18:4,5,5,12,13,18,25;
19:4;21:7;24:13,15,
19;25:6;26:16;27:17,
23,25;28:17;29:21;
31:6,8,10;32:10,12,
15;33:20;36:12,13,
14;41:10,13,19;42:3,
20;45:5,8;47:16;
49:11,17,24;50:18,25;
51:19,24;52:3,5,8,10,
14;54:14;57:5,12,18,
22;60:6;62:18;63:23;
65:18;70:24;71:3,25;
72:18;73:12,16;
83:21;91:17,20,22,25;
92:9,16,20,22;93:25;
99:15,18;101:3,9;
108:20;111:9;118:8,
12;123:24;124:8;
129:3,6,17;130:5,10,
18,20,21,24;131:1,4,
14,20;132:3,7,10,12,
18,22;133:3,9;
134:21;136:22;137:7,
12;147:15;156:2,10
courtroom (2)
4:21;52:6
Courts (1)
46:13
Court's (4)
11:19;17:13;118:5;
130:14
cover (1)
8:12
coverage (1)
107:3
covered (1)
133:14
CPF (4)
119:23;121:11;
122:25;123:10
crank (1)
86:14
create (4)
7:6;90:7;96:22;
106:24
created (2)
57:10;119:22
creates (1)
80:4
creating (1)
24:5
credentialed (1)
17:5
credentials (1)
21:16
credit (6)
14:5,15;90:25;
121:9;150:3,4
credits (3)
9:1;20:10;136:6
criminal (1)
49:11

criteria (1)
56:25
Cropper (1)
95:25
cross (4)
12:17;13:16;15:5,17
cross- (2)
11:25;16:17
cross-examination (6)
9:5,8;99:17;132:7;133:11;137:16
cross-examine (1)
15:20
crucial (2)
89:18;128:3
cryptography (1)
72:7
currency (1)
44:8
current (2)
28:4,23
currently (3)
25:10;56:7;64:7
curriculum (2)
28:3,4
customers (1)
88:16
cute (1)
64:16
CV (2)
27:13;28:10

D

Daniel (1)
5:9
Darr (8)
4:3,3,13;6:1;16:13;41:19,20;138:5
Darrpa (1)
29:9
data (145)
8:20;12:11,11,14,20;15:23;17:5,6,18,23;18:1,3,16,24;20:6,14;21:15,19,19,20;22:1,2,4,7,14;23:12,13;24:10;26:10,12;28:21,21;29:2,7,24;33:3,9,12,25;34:10;35:10,12,17;36:4,18,20,25;37:1,14,16,23;40:13,14;41:3;47:11;48:10,17,17;49:12;50:9,24;51:2,4;53:1,3,18;59:18;60:6,12,17;64:18,21;65:11;66:24;69:1,14;70:3,5;72:24;73:18;74:1,4;75:16,17,19,23;82:4;83:23;84:25;85:2;86:14;88:1,12;89:10;91:10;93:15;95:4;99:7,25;100:1,10,14,16;101:1,23,23,25;104:14,19;105:13,16,24;106:19;107:16,21;108:2;109:2;110:2;113:6;116:22;117:10;118:1,2;119:8;120:2;122:17;127:15,17;128:1,2,19;129:19;130:2;139:5,25;142:18;143:6;144:16,20;146:21,24;149:14;150:12;152:8,16
database (3)
78:13;86:18;128:17
databases (1)
88:19
dataset (42)
17:7;33:9;34:2,2;40:12,12;41:5;42:3;48:2,5,18,20;49:6,23,24;50:24;51:6,7;54:6,10;56:5;60:9,14;63:22;64:19,20;65:23,24;67:8,10,10,22;69:15;100:4;106:6,22;108:18;109:8;118:19;119:2;121:1;150:7
datasets (19)
26:9,15;29:5,5;30:5,6;33:1,17;34:1;37:9;45:16;48:3,13,25;49:3,13;50:4,11;105:12
date (9)
58:8,9,17;77:22;120:14,15;121:17,23;123:3
dates (2)
58:10;78:9
Daubert (6)
4:7,16;5:23;14:7;19:11;129:5
day (3)
18:14;24:7;137:4
days (3)
13:23;16:23,24
deadline (1)
4:16
deal (4)
7:11;9:19;11:22;20:3
deals (1)
21:23
decades (2)
37:24;47:20
December (1)
14:21
decent (1)
59:8
decide (2)
148:12,14
deciding (4)
29:6;38:2;99:21;148:23
decision (2)
14:25;34:24
decisions (2)
100:8;113:14
declared (2)
111:4,5
deem (1)
18:19
deemed (1)
75:16
deeper (1)
72:25
defendants (8)
4:17;5:2,4;9:9;14:6,9;16:11;129:1
defense (5)
18:25;136:21;137:22;153:9,14
definition (1)
155:11
degree (11)
34:22;35:4;38:3;46:17;66:7,7;89:14;90:8;93:9;142:12;153:17
degrees (8)
35:6;45:22,25;47:4;89:15;90:1,19;93:5
demonstrative (2)
16:15,22
demonstratives (1)
16:4
Dennis (15)
4:8;6:25;7:3,7,12,22;8:7,18;9:6,19;12:17;13:16,18;15:12,14
Dennis' (3)
7:10;23:13;134:4
Department (2)
25:16,23
depending (4)
59:12;74:10;99:3;106:19
depends (2)
66:7;151:18
deposed (1)
138:19
deposition (2)
13:22;138:21
describe (28)
17:22;26:21;27:7;29:20;37:10;42:3,20;47:7,16;49:21;50:8,23,25;55:15;57:4;60:6;61:7;81:11;84:7;88:17;93:2;96:12,16;97:17;105:7;113:23;115:18;145:20
described (16)
35:14;36:1;44:21;45:20;50:1;52:17;63:2;74:2;79:15;90:10;101:21;106:13;109:22;114:22;145:3,8
describes (1)
56:1
describing (6)
48:1;63:8;119:17;144:12;145:6,17
description (2)
28:18;129:13
deserve (1)
76:21
design (2)
33:4;35:20
designed (1)
118:1
desirable (2)
103:17;104:7
desired (2)
67:5;108:17
despite (7)
33:14;89:17;123:12;125:6,18;127:6,8
detail (3)
17:22;36:13;38:4
detailed (1)
79:17
details (6)
37:21;84:11;96:20;135:23;136:6;148:16
determination (4)
14:25;17:14;44:7;147:25
determine (30)
17:6;33:1;35:9;48:1,7;50:23;51:4,6;52:22,25;53:11,24;54:4;55:6;56:5;62:25;72:8;80:18;82:5;93:12;112:17,21,22;117:16;123:23;126:15;127:15;136:11;146:21;149:5
determined (21)
33:3;42:16;43:25;44:23;47:24;55:12;58:1,3,13,25;67:20;71:20;90:12;94:21;95:2,16,18;96:2;109:5,15;152:12
determines (5)
18:13;38:1;89:22;91:1;101:22
determining (7)
30:4;33:12;75:21;84:25;93:7;95:14;136:16
deterministic (11)
17:21;34:12,14,15,18;35:1,15;76:25;77:6;79:6;81:6
deuce (1)
95:10
develop (4)
29:4;36:6;108:1;138:6
diameter (3)
103:4,9,9
dice (1)
118:21
dictated (1)
106:5
differ (1)
46:24
difference (1)
23:6
differences (4)
72:8;83:6;107:5;122:10
different (81)
14:18;15:2;21:25;23:18;26:20;31:3;34:9;39:1,22;44:1;46:1;47:6;53:4,6,7;55:9;56:7,19;61:17;64:6;71:4,6;73:11;78:14;79:5,21;82:22;83:9;84:10,12,18;86:4,7;87:2,3,5,6,18;88:17;90:11,12;94:23;96:19;97:7,24;99:4;100:23,23;102:17;103:19,19;104:7;106:7;108:14;110:9,9;117:9;119:15,18;120:10;123:11;125:25;139:16;141:2,5,8,13,13,16,25;142:4;143:10;146:4,8;147:10;150:9,14;151:9,10,12,14
differently (2)
91:5,10
differs (1)
17:20
difficult (3)
35:13,20;143:17
difficulties (1)
137:15
difficulty (1)
19:7
digit (2)
57:21;78:24
digits (8)
47:1;53:21;74:21,22;75:2,6,7;125:24
direct (7)
13:16;18:22;25:8;32:12;41:25;99:22;131:19

directing (3)
7:15;27:15;28:3
direction (1)
43:19
directly (3)
27:24;107:1;136:4
directs (1)
99:24
disagree (5)
85:23;86:3;100:20;
133:20;149:21
disagreement (4)
82:14;83:17;91:6,
11
disambiguate (1)
87:13
disambiguated (1)
89:4
disambiguating (1)
89:24
disambiguation (1)
87:9
discerning (1)
104:4
disclose (2)
152:17,20
disclosed (7)
4:8;12:19;13:6,10,
22;15:23;21:7
disclosure (1)
4:16
disconnected (1)
51:25
discovery (4)
115:13,19,20;
116:23
discuss (2)
60:16;144:23
discussed (2)
14:14;41:8
discussing (4)
30:13;38:9;40:19;
152:21
discussion (1)
127:11
display (1)
27:12
displayed (2)
141:7,9
distance (2)
103:6;104:2
distinction (1)
23:6
distinctions (1)
48:7
distinguish (1)
88:8
distinguishes (1)
86:9
distinguishing (1)
89:1
distribution (4)
29:24;30:1,6,7
diverse (1)
59:21
diversity (1)
145:17
dividing (2)
95:7;96:24
docket (3)
11:2,2,3
Doctor (4)
26:16;70:25;91:22;
145:16
document (2)
6:25;55:22
documentation (1)
85:14
document's (1)
16:7
dollar (4)
44:20;111:3;112:2;
124:4
dollars (1)
40:9
domain (3)
62:17;64:9,10
domains (1)
56:14
domestic (3)
4:5,12,18
Dominic (2)
84:9;92:4
done (11)
12:8;13:18;15:22;
16:6;17:11,17;53:1;
89:24;110:16;147:18;
152:23
dot (4)
67:11,11,11,11
dotted (2)
98:6,11
down (5)
62:24;71:24;77:16;
79:8;89:5
dozen (2)
80:4;100:6
Dr (47)
4:6;5:24;9:24;10:4;
12:1,17;14:16,19;
15:17,24;16:9,11,17;
17:4,15;18:14,23;
19:20;20:4,13;21:1,3,
8,16,23;22:11,19,22;
23:12,21;24:15,17,22;
25:10;27:13;35:23;
62:24;101:9;128:23;
129:22;130:23;131:7,
11;132:1;133:13;
136:20;137:18
draw (2)
65:22;68:22
drawing (1)
68:1
drawn (2)
67:9;102:14
draws (2)
62:16;124:24
drive (2)
42:25;44:25
drives (1)
130:1
dropped (1)
52:6
dual (1)
116:5
due (1)
66:3
dumb (1)
69:22
dump (1)
13:7
Duran (4)
5:3,3;11:16,25
during (5)
11:25;16:19,25;
131:21;135:10
dynamic (16)
27:7;92:15;99:13;
105:4,7,8,10;106:12,
14,18;107:19,20;
108:7,13,23;131:9

E

earlier (16)
36:1;38:9;42:12;
44:20,21;61:14;
63:11;70:18;72:3;
74:19;89:21;93:11;
101:21;113:24;147:1;
153:19
early (1)
77:6
earth (1)
41:8
easier (3)
35:16;61:2,5
easily (1)
135:8
easy (4)
22:16;53:19;61:21;
97:10
easy-to-use (1)
69:7
ECF443 (1)
153:16
edges (6)
98:6,9,12,14;102:7;
108:5
effect (2)
138:24,25
effective (1)
14:21
efficient (2)
29:17;80:10
efficiently (1)
80:7
eight (7)
9:13;77:3,18,20;
96:20;106:13;117:3
eighteen (1)
84:20
eighty (7)
85:6,7,13;98:3,10;
102:8;140:3
either (19)
7:6;21:11;34:1;
58:4;76:2;85:5,7;
92:5,19;94:21;
100:19;102:22;105:5;
112:19;133:22;
143:12,13;146:7;
151:23
elaborate (1)
97:17
electronically (1)
5:20
eleven (14)
40:17,23,23;43:1,
20;44:9;53:18;54:21;
60:10;73:14;76:15;
77:12;81:7;110:15
ellipsis (1)
67:11
else (10)
7:3;17:13;64:5;
90:24;97:12,13,21;
98:20;119:16;148:13
EM (12)
89:21,22;91:1;
93:11;95:13,19;
99:20,22,24;100:23;
101:21,25
email (52)
38:18;42:11;46:22;
47:5;53:5;55:16,17,
19;56:1,8,15;58:17,
21;61:23,24;62:7,14;
63:24;64:4,6,11,13,
14,17;70:19;71:6;
74:13,13,15;77:24;
86:1;119:13;120:9;
121:1,9;122:12;
123:7;124:14;125:24,
25;126:9;141:2,5,13,
25,25;143:9,11,14;
151:9,10,16
emails (2)
125:1;141:3
emerge (1)
62:13
emerges (1)
139:5
EMF (1)
94:21
eminently (3)
33:14;58:22;126:18
employed (1)
25:10
empty (11)
53:19;54:18;55:17;
58:4,5;59:22,23;
63:13;64:2;70:20;
71:21
enable (1)
65:22
encoding (1)
90:20
end (16)
18:10;24:7;65:9,9;
74:16;78:24;91:11;
104:7;110:12;117:7;
125:5;131:15;146:25;
152:5;156:4,12
ended (1)
105:19
ends (2)
97:23;99:1
engage (2)
26:2;28:21
engaged (5)
29:1;32:22,25;33:4;
138:5
engagement (7)
31:1,17;32:22;
44:16;127:16,20;
137:24
engaging (1)
66:23
engine (2)
17:23,24
engineering (2)
26:1;55:25
enhancements (1)
50:3
enormous (3)
60:14,15;76:16
enough (13)
65:5;66:5;68:10;
70:21;72:6;76:20;
82:6;83:14;92:18;
94:13;95:3;147:21;
153:2
ensure (2)
60:13;146:6
enter (1)
74:21
entered (2)
17:13;129:1
entire (6)
9:15;54:20;93:15;
108:18;109:2;119:9
entirely (13)
13:15;14:18;15:2;
42:13;51:7;59:23;
65:3;74:15,18;
100:14,21;121:22;
126:25
entirety (1)
71:16
entity (4)
34:5;86:16,17;
132:4
equal (1)

103:8
equals (3)
26:25,25,25
equity (35)
18:8,12;19:22;
20:17,20,23;23:25;
24:6;39:11,13,17,18;
42:15;44:6,19;
110:20,25;111:1,2;
112:11,11,13,13,18;
123:17,20,21,25;
124:12;125:19;
153:23;155:2,6,9,14
Eric (1)
25:1
error (37)
33:5;47:1;50:8,10;
63:20;66:8;69:19,22;
70:10;75:9;99:8;
112:22;113:3,19;
115:9,11,21;116:18;
117:18,22;122:3,7;
126:16;127:3;128:13;
141:19;142:19;143:1,
24;144:19,24;145:1;
151:7,23,25;152:8,18
errors (4)
114:7;115:2;
144:13;152:3
especially (8)
35:11;60:9;64:25;
74:4;104:14;107:4,
24;146:13
essence (1)
34:25
essential (1)
37:9
essentially (21)
12:21;29:12;40:19;
59:2;63:13;64:5;
65:21;72:6;73:15;
76:9,17;78:18;93:4;
95:2,13,20;100:16,18;
108:1;113:8;119:10
estate (2)
29:2;86:20
Estevez (2)
126:5,7
estimate (1)
69:23
estimates (2)
63:11;152:14
ethnic (1)
147:10
evaluate (2)
48:21;50:9
Evaluating (2)
129:19;144:11
even (55)
15:6;19:16;21:11;
23:18,18,22;36:3;
37:10;40:17;41:5;
45:14;51:10;52:22;
54:7;55:20;60:14;
66:4,14;67:2,6,23;
69:8,13;70:14;75:8,
24;77:22;80:4;82:3;
86:1,24;87:3;88:18;
93:13;94:17,17,20,23;
104:16;107:4;112:2,
3,4,14;113:15;
118:23;119:17,21;
122:20;125:4;126:11,
19;139:15;141:21;
149:23
everybody (2)
8:19;156:11
everyone (3)
5:22;41:13;58:11
everyone's (1)
22:1
everything's (1)
145:13
Evidence (16)
4:11;12:24;17:13;
78:1;87:1;119:19,21;
122:20,25;124:22;
129:7,8,21;131:17;
132:5;147:14
evident (3)
23:11,11,12
exact (13)
15:14;47:5,5;77:17,
18;78:8;93:25;94:16,
17,19,20;124:16,20
exactly (28)
7:9,16;10:11;11:22;
33:20;34:21;62:1;
64:1;71:7;72:15;
73:14;78:17;82:13;
92:6;93:3;105:16;
110:10;118:18;119:1;
120:24,25;121:4,13;
123:8;124:18,19;
125:7;126:12
examination (7)
12:1;16:18;25:8;
32:12;41:25;128:14;
131:19
examine (1)
54:4
examined (2)
53:10;55:6
examines (1)
117:9
examining (2)
32:8;42:16
example (66)
21:22;22:12;29:23;
31:5;34:23;46:12,20;
49:20;54:13;55:1,4,
11,15;56:17;58:8;
59:13;61:2,23;72:13;
75:9;76:6;82:9;83:4,
22,23,24;84:1,21;
88:9;89:25;90:13;
93:25;94:1,2,3,6;
97:25;102:12,14,23;
103:4,6,15;109:24,25;
114:10,17,18;117:15;
119:1,16;120:2;
121:19;123:1;125:8;
129:20;134:19;
139:19;143:1,4;
146:3;147:4;148:5;
149:5,24;150:3
examples (13)
30:25;31:15;57:23;
68:21;69:1;72:17;
82:15;113:24;117:21;
118:10;126:14;135:6;
139:14
exceeds (2)
97:20;117:11
excellent (1)
118:3
except (6)
9:23;57:8;62:7;
74:15;122:11;144:25
excerpts (2)
13:17;14:12
exchanged (1)
12:4
exchanging (1)
135:21
excise (2)
7:5;8:21
exclude (6)
4:6,8;9:23;11:20;
17:1;71:23
excluded (1)
59:1
excluding (2)
57:1;69:20
exclusion (1)
20:9
exclusive (1)
107:15
excuse (3)
8:2;35:23;131:5
exercise (1)
18:23
exhibit (51)
12:4,10;13:8;15:15;
27:12,15,16,22,23;
31:6,9,10;32:18,18;
42:24;43:12,12,17;
44:25;49:18,18;
55:24,24;60:4;63:23;
69:10,10,11;71:8,9;
118:6,7;128:24,25;
129:7,8,10;130:8,18;
131:22;132:5;136:21;
137:22,22;140:4;
141:4;153:9,10,13,14,
15
exhibits (22)
4:8;5:19;11:20,22,
24;12:6;13:4,6,21;
15:10;17:1;30:24;
100:12;118:6;128:23;
129:2,10,24;130:6,22;
131:14,17
expanded (2)
110:1,9
expect (16)
10:4;53:22;56:15;
62:1;63:16;64:2;
70:23;78:13;94:25;
110:10;112:6;117:18;
122:6;126:21;142:12,
14
expectation (10)
10:6;37:19,21;
47:13,17,25;49:10;
100:12;128:6;146:12
expected (3)
70:19;104:20;
154:25
expects (2)
142:20;150:11
experience (4)
7:21;26:8;28:11;
107:24
expert (12)
4:16,18;7:1;9:13;
14:11;19:14;20:4,23;
32:7;84:2;127:13;
148:2
experts (2)
7:1;9:15
explain (6)
17:19;52:16;71:7;
92:15;108:19;111:9
explained (4)
19:13;114:17,18;
152:10
explaining (1)
139:9
explains (2)
88:24;144:16
explanation (1)
145:2
explicitly (1)
147:20
expressed (1)
82:12
extension (2)
49:8;50:15
extensions (3)
45:18;50:3,6
extensive (1)
37:7
extent (17)
7:9;13:12;16:15;
55:6;71:12;130:11,
24;141:16,20;142:22;
147:2;149:1,8;150:6;
152:10;154:22,24
extra (1)
83:7
extraordinarily (1)
67:16
extremely (5)
70:23;121:12;
123:4;124:22;127:10
extremes (1)
107:4
eye (3)
51:16;56:21;156:1

F

F1 (3)
116:10,12;117:4
fabulous (1)
22:12
fact (16)
18:11;21:11;33:15;
35:8;48:22;56:12;
59:23;67:7;78:3;
97:16;101:1;104:16;
117:7;120:10;128:3;
152:24
facts (2)
22:11;24:10
fair (7)
67:3;68:8;73:21;
78:2;82:3;119:24;
126:2
fairly (4)
60:17;70:12;122:3;
128:16
faith (1)
15:9
fake (1)
22:17
fall (1)
75:21
false (38)
24:5;113:6,6,19,19,
21,21;114:7,13,16,17;
115:1,7,13,13,14,14,
16,17,19,20,20;116:7,
8,13,13,17,17,17,22,
25;117:3;122:4,14;
126:21,22;152:15,15
familiar (3)
148:2;153:16;155:1
family (1)
85:25
far (4)
16:8;72:18;93:9;
127:9
faster (2)
35:16;151:1
Fatura (1)
42:7
favor (1)
94:14
FDR (1)
115:19
feasible (2)
51:5;81:18
features (3)

70:1;117:24;123:19
Federal (2)
4:11,14
fees (1)
4:17
Fellegi-Sunter (8)
37:19,23;45:4,9,12;
47:3,12;49:9
female (1)
94:24
Femaleofnaturalmama@hotmailcom (1)
126:8
few (11)
39:23;100:7;
108:25;118:24;
119:11;122:10,10;
126:24;127:4;132:24;
152:4
fewer (3)
111:25;112:4;
124:13
field (31)
17:6,18;21:17;32:7,
9;34:12;35:6;36:1;
45:14;47:7;52:25;
54:14,16;55:16;
56:18;61:10,11;62:6,
7,25;66:12;71:6,6,13;
74:10;82:6;86:21;
89:19;101:20,24;
129:13
fields (77)
13:13;26:19,21;
30:22;31:16;33:12;
34:21,23;35:7,10,21;
42:12;43:9,16,16;
45:22;46:1,15;48:6;
51:8;52:22;53:4,9;
54:7;57:24;58:15,23;
59:1,10,14,21;60:6,
24;61:12,14;64:24,
24;71:4,11,15,17;
73:24;75:16;76:7;
77:3,18,20;84:22,23;
85:23;86:2,14;88:18;
89:2,14,20,23;90:14;
93:12,13;94:23;
100:3,16,17,23,24;
101:22;110:2;119:11;
123:3;124:14,19,21;
125:6;127:19,21;
146:19
fifties (1)
83:15
fifty- (1)
141:22
fifty-eight (1)
143:8
fifty-seven (1)
151:21
fifty-two (1)
143:4
figure (6)
6:16;16:10;47:8;
52:21;101:2;137:6
figures (1)
63:16
figuring (1)
16:12
file (1)
11:19
filed (4)
6:9;12:5;13:5;
153:15
filings (2)
21:7;153:20
filter (3)
44:8;56:24;80:12
filtered (1)
89:12
final (9)
49:16;62:15;88:19;
95:5,18,19,20;113:9;
128:24
finally (4)
98:13;110:17;
115:1;116:9
find (7)
49:21;69:17;70:8;
79:24;83:14;142:20;
144:15
finds (4)
50:5,6;106:20,21
fine (12)
6:8;8:2,3;12:10;
19:7,9,9;40:3;107:4;
130:12;132:25;156:7
finger (1)
102:1
finish (2)
36:16;127:7
finished (1)
96:9
firm (2)
28:20;29:3
firm's (1)
69:6
first (35)
6:2;23:1;27:1;
28:15;51:9;52:16;
62:12,12;69:18;
71:19;76:10;78:24,
24;79:2;80:12;82:17;
88:5;90:2,8;94:8;
98:9;99:21;100:13;
113:4;116:6;119:11;
121:14,17;122:12;
124:14;139:19,20;
140:14;154:7;155:21
fit (6)
19:19,20;21:1,13;
100:10;135:4
fits (1)
135:6
five (12)
10:10;36:23;37:3;
54:22;67:13;70:9;
75:7;118:25;124:14,
17,19;127:24
fix (5)
27:18;72:12,13;
118:16,21
fixed (16)
103:5,7,9,14;104:1;
105:2,5,21;106:3,15,
25;108:12,14;109:11;
114:19;118:14
flash (2)
42:25;44:25
flavor (2)
79:5;124:1
flexible (3)
80:17;82:1;84:23
flip (1)
137:1
flow (1)
115:17
flowchart (1)
36:11
FNR (1)
115:20
focus (1)
42:18
focused (1)
43:19
focusing (1)
59:20
follow (1)
11:21
followed (3)
78:20,20;79:9
following (1)
20:5
follows (1)
50:15
force (2)
56:1;98:23
forced (7)
97:12,21;98:15;
99:2,5;103:25;109:11
form (10)
8:9;36:25;77:12;
82:14;97:22;98:13;
107:9,9;112:11;
139:11
format (3)
15:14;58:10;61:3
formatted (4)
56:2,9;64:4,6
formed (1)
96:10
former (1)
71:10
forming (1)
38:6
formula (2)
95:14;115:18
formulas (2)
69:5;115:15
forties (1)
83:15
forty (3)
124:13;134:18;
135:6
forty-five (1)
83:13
forty-four (4)
43:9,16;53:9;71:18
forty-one (1)
124:19
foster (1)
140:6
found (7)
12:16;54:23;116:3,
5,22,24;117:2
foundation (1)
7:24
foundational (1)
45:13
foundations (1)
29:17
four (34)
42:6;44:2;60:5,23;
63:17;64:1;67:15,18;
70:8,11,15,22,22;
74:22;76:9,14;77:2;
78:24,24;79:2;80:22;
83:3;98:1;103:14;
104:15;113:23;114:2,
11;115:16;118:16;
120:22;121:14;
123:13;125:4
fourth (7)
76:11,12,13;83:9;
115:1;123:5,6
frame (1)
98:13
framed (1)
107:23
frank (1)
9:14
frankly (1)
14:20
Frante (1)
130:8
Frantz (3)
4:10;6:10,24
free (2)
8:8;15:25
Freer (40)
4:6;5:24;9:24;10:4;
12:1,17;14:19;15:17,
20,24;16:11;17:4,15;
18:11,14,23;20:5;
21:3,23;22:11,19,22;
23:12,22;24:8,16,17,
22;25:1,10;62:24;
101:9;128:15,23,23;
129:22;131:7,11;
133:13;137:18
Freer's (13)
12:11;14:16;16:9,
17;19:20;20:14;21:1,
8,16;27:13;130:23;
132:1;136:21
frequent (1)
63:21
frequently (4)
78:25;95:3;100:22;
116:24
Friday (1)
14:14
front (8)
13:23;15:24;17:3;
117:7;130:7;134:3;
144:21;153:9
froze (1)
44:12
frozen (2)
44:10;51:20
FS (8)
45:22,23;46:9;
47:20;93:10,12;
95:14,19
FSEM (26)
49:9,22;50:3,15,22;
88:20;89:9,16,20;
90:18;91:1,9,14;
92:13,24;93:2;94:4;
95:2,10,23,25;101:18,
21;146:15,18,18
full (13)
56:5;67:10;80:7;
109:4;120:25;121:9;
122:16;124:5;141:6;
144:15;149:17;150:1;
153:5
Fuller (4)
83:4,5;84:16,17
function (6)
82:23;83:3,8;85:4;
90:9,14
functions (2)
82:16;83:1
further (5)
41:1;76:22;80:9;
136:12;144:1
fuzzily (1)
90:11
fuzzy (15)
82:6,11,16,16,23;
83:1,7;84:24;85:4;
89:3;90:3,3,5,8;121:6

G

gained (1)
40:9
gaining (1)
73:9
gains (1)
111:14
G-A-M-A-I-L (1)
121:3
Gamalon (2)

28:13;29:8
gamma (5)
90:18;91:12;93:14;
95:15;100:3
gave (3)
22:23;149:15,15
gender (2)
94:24;99:22
general (16)
25:19;28:18;33:24;
45:21;52:21;53:16;
61:22;66:15;68:23;
69:16;73:4;93:10;
96:22;100:15;144:11;
150:13
generally (27)
17:25;25:17;36:12,
13;40:11;45:5,5,9;
46:13;49:12;52:13;
56:14;59:16;61:8;
64:13,23;72:23;73:2;
75:5,21;96:12;125:3;
128:1;131:3;136:7;
142:11;148:2
general-purpose (1)
96:18
generate (1)
101:17
generator (1)
118:14
generators (1)
72:5
generic (2)
69:12;132:3
geometric (1)
134:16
geometrically (1)
134:24
gets (1)
101:23
given (11)
30:5;31:4;39:22;
82:6;110:25;117:18;
122:6;130:25;131:11;
136:10;145:20
gives (5)
13:11;70:15;112:6;
120:24;124:17
giving (1)
39:3
global (1)
7:22
Gmail (3)
56:13;64:12;121:3
goal (8)
76:24;102:16,22;
103:18;104:10,11,11,
19
goals (1)
104:13
goes (4)
26:1;27:6;57:16;
82:25
gold (3)
19:16;48:10;49:23
Good (15)
5:1,3,22;83:16;
90:22;92:17;94:10,
12;101:9;113:16;
117:6,8,12;133:13;
156:4
governing (1)
55:21
graduate (2)
26:3,4
grant (2)
149:2;151:14
granularity (1)
107:4
graph (6)
107:23;108:3,4,19,
20,24
graphic (1)
107:9
great (2)
22:9;30:16
greater (1)
113:16
green (6)
57:24;77:4,20;
94:15;98:11;127:21
gross (2)
107:9;126:15
grossly (2)
13:25;14:8
ground (8)
21:21;26:6,8;35:24,
25;48:16,17;49:24
grounding (1)
117:19
grounds (1)
130:16
group (4)
25:24;81:23,24;
97:13
grouped (2)
97:11;119:9
grouping (7)
76:25;77:6;78:10;
79:6;109:21,24;138:7
groupings (12)
77:14,16;78:5;79:9;
89:11;93:21;98:1;
109:4,17,19;110:6,8
guess (4)
6:20;7:7;125:22;
151:17
guidance (1)
95:14

H

had/ (1)
149:15
half (6)
31:13;37:2;70:6;
90:17;116:25;145:5
hand (6)
19:15;24:20;61:21;
67:1;73:13;78:18
handle (1)
6:2
hands (7)
112:3;135:11,17,
24;136:4,5,9
hang (2)
61:17;71:15
hangs (1)
61:22
happen (9)
6:8;9:10;10:9;76:2;
92:14;148:15,16;
153:1;154:22
happened (7)
23:21;48:22;
109:22;111:17;
130:10;132:1;135:22
happening (3)
62:1;67:17;70:16
happens (8)
17:15;23:22;27:6;
65:11;89:6;93:17;
95:13;142:16
happy (1)
92:19
hard (4)
42:25;61:4;115:23;
122:16
harder (2)
54:11;104:24
Harvard (1)
32:4
Hawaii (1)
32:2
HDMI (1)
136:25
heads (2)
14:1;132:20
hear (6)
5:6;9:13;21:3;
22:20;23:8;52:4
hearing (15)
4:4,7,9,11,16,18;
6:17;9:15;10:16;14:7;
16:19;41:15,19;
129:5;131:16
hearken (1)
26:5
hearsay (3)
7:2,24;130:16
held (1)
31:22
help (1)
121:21
helped (2)
7:4;29:6
helpful (13)
14:24;18:19;21:4;
35:11;52:15,16;54:1;
58:7;62:22;73:21;
83:22;121:19;129:16
hence (1)
136:17
Here's (3)
36:11;103:12;
129:17
hesitating (1)
140:21
heterogeneous (2)
105:11;107:21
high (27)
35:18,22;51:8;
52:23;54:7;58:1;59:3,
15;64:23;77:4;97:2,3;
106:17,21;116:1,6,8,
14,16;117:6,10;123:4,
25;124:3;125:10,11;
127:10
higher (9)
36:25;51:11;59:10;
65:10;101:19;104:3;
112:3;117:25;123:19
highest (5)
77:3;104:5;111:24;
113:16;123:20
highlighting (1)
9:6
highly (3)
17:4;50:6;67:19
high-quality (2)
120:3;127:19
hinted (1)
6:7
hold (6)
9:22;10:24;27:17,
17;41:10,10
holistic (1)
116:16
holistically (1)
61:12
homes (1)
39:1
hone (1)
100:5
honed (1)
100:8
honing (1)
12:23
Honor (45)
4:24;5:1,3;6:6,7;
8:11;9:2;12:2,25;
15:8,10,17;17:3;19:2,
3,5,10,11,12,22;21:4,
4,5;23:2,14;24:14;
25:5;32:6,11;42:23;
91:13;99:11;101:12;
128:22;129:4;131:5,
18;132:8,16,24;
133:2;137:5,8,14;
156:8
Honor's (3)
7:8;23:8;133:7
hospitals (1)
21:22
Hotmail (2)
56:13;64:12
hours (2)
13:24;81:17
house (2)
59:11;77:24
household (1)
85:25
huge (1)
80:3
human (2)
54:1;147:20
humans (1)
80:11
hundred (13)
13:9;46:4;60:5,23;
63:17;70:5,8,8,11,11,
15,15,23
Huron (3)
17:21;42:4;44:18
husband (1)
85:24
hyphen (3)
57:2;74:22;123:1
hyphens (1)
56:22
hypothesis (1)
147:14

I

ID (11)
14:5,15;44:18,24;
57:6,9;62:9;86:24;
140:6,21;153:24
idea (30)
15:4;34:16,19;
38:10;46:15;48:4,23;
49:5;56:2;60:9;70:12,
16;72:5;74:23;79:23;
80:6;82:7;84:5;94:6;
96:22;100:13;102:15,
19;103:6;105:11;
106:14,25;108:1;
123:18;124:10
ideal (3)
69:18;70:2;107:11
Ideally (2)
104:23;105:20
ideas (1)
15:12
identical (7)
46:23;76:20;77:21;
82:18;119:11;122:11;
123:8
identification (1)
21:12
identified (13)
13:6,20;40:5;43:16;
120:12;121:25;
139:24;140:22;

141:11,13;149:18,21; 151:21
identifier (2) 35:18;57:7
identifies (5) 38:22;39:4;110:23; 120:14,16
identify (26) 4:22;20:20;38:14; 44:15;54:14;80:1; 86:12,20;87:8,8;88:7, 16;98:9;119:5;120:6; 138:11,12,13,23; 139:15,21;149:11; 150:15,24;152:3,8
identifying (8) 5:7;21:2;30:6;39:2; 52:23;88:9;138:17; 151:15
identities (1) 136:16
identity (3) 42:17;59:4,7
IDs (2) 20:16;154:3
illustrate (2) 97:18;104:8
illustrated (2) 36:23;102:18
illustration (1) 67:9
illustrative (3) 85:3;94:2,3
Ilyas (1) 5:1
imagine (2) 67:12;143:12
immaterial (1) 131:12
impact (5) 44:6;95:16;100:8, 24;155:2
imperfect (1) 120:2
imperfectly (1) 104:20
implement (5) 33:4;35:14,17,20; 138:6
implementation (2) 92:24;94:4
important (23) 19:18;23:3,5;37:12; 48:7;51:1,2,3;52:18; 54:6;65:2;67:6;71:16, 23;79:6;87:9;89:17; 91:2;99:6,21;104:22; 117:19;154:21
impossible (2) 13:14,15
improved (1) 51:9
improvements (1) 50:3
improving (1) 75:23
inadmissible (1) 4:7
inappropriate (1) 14:8
incidentally (2) 133:24;146:14
inclined (1) 9:3
include (13) 20:5;38:24;55:11; 58:7;59:11,17;63:6; 75:1;85:7;124:3; 151:4,8,18
included (13) 28:9;42:21;56:23; 75:8;123:15,25; 124:5;125:21;128:23; 132:4;134:18;138:17; 152:19
includes (3) 31:18;103:13; 123:10
including (7) 20:9;29:16;30:22; 45:18;72:7;85:14; 117:24
inclusion (2) 20:9;58:1
inclusive (4) 82:3;103:17; 107:14;125:13
incompetent (2) 7:25;8:6
inconsistencies (3) 51:15;65:4;127:9
inconsistent (3) 127:3;128:1;139:2
incorporate (1) 125:14
incorrect (2) 114:8;116:25
increase (2) 4:17;97:2
increased (1) 45:17
increases (2) 70:13,14
increasingly (1) 49:3
incrementing (1) 78:23
indeed (18) 26:20;40:16;63:18; 68:16;70:18,20,22; 74:3,11;85:8;89:18, 20;97:15;103:24; 122:5;126:24;127:5; 141:19
independent (1) 147:4
indexing (1) 79:21
indicate (1) 94:23
indicated (4) 54:18;59:3;90:20; 93:20
indicates (4) 16:3;46:5;82:12,13
indicating (2) 67:11;98:7
indications (1) 122:14
indicative (3) 83:16;100:17;101:2
indirectly (1) 136:6
individual (15) 14:6;18:6,7;32:6; 38:14;39:8,8;42:11; 102:3;109:25;126:5, 11;149:1,8;150:15
individually (1) 64:24
individuals (4) 34:11;39:7;89:1; 139:1
industry (3) 21:23;28:11;83:2
inevitable (1) 104:18
inevitably (4) 104:25;106:1; 109:7;113:12
infeasible (2) 37:10;40:25
infer (1) 69:24
inference (4) 29:18,20;30:4,11
inferential (2) 66:13,13
inform (2) 7:4;100:2
Informatics (1) 129:18
information (43) 9:7;14:4;38:8,18, 20,21;39:3,12;42:10, 17;45:11;46:18;47:3, 8;53:5;57:11;73:3,9; 78:7,21;79:1,11;91:3; 110:5,24;111:18; 112:7;113:8,18; 120:12;121:24,25; 124:13;141:17;143:3; 148:12;149:11,15,22; 150:10,14;151:3; 153:6
informative (3) 51:10;73:21;100:25
informs (1) 48:6
infrequent (2) 63:10;65:4
ingredient (2) 82:5;108:3
ingredients (1) 106:14
initial (9) 29:4,7;41:4;84:2,9; 100:2,2;127:13,14
initially (3) 16:11;65:13;71:19
initials (1) 63:7
input (13) 36:17,20;39:24; 47:3;65:11;89:16; 91:9;108:6;118:1; 127:6,8,15,17
inputs (2) 96:25;112:13
inputted (1) 73:10
inspect (1) 61:21
instance (5) 142:17,19;143:16; 147:1;151:15
instances (7) 63:8;69:20;141:18, 20;143:23;144:18; 152:15
instead (1) 67:17
Institute (1) 25:13
integrity (1) 20:6
intelligence (1) 26:14
intended (4) 5:14;57:5;97:1; 121:14
intending (1) 11:25
intends (2) 15:4;129:15
intention (3) 6:23,24;115:22
interactions (2) 8:16;29:16
interest (1) 29:24
interested (4) 71:13,14;86:23; 123:24
interesting (7) 107:5;111:21,22; 112:7;113:1;117:23; 123:23
interests (2) 29:15,15
intermediate (6) 58:3;77:4;97:14; 143:12,17;144:3
internal (4) 59:5;73:22;74:12; 113:6
internally (1) 113:18
international (3) 4:5,12,19
internet (2) 55:21,25
interrupt (3) 41:16;70:24;153:12
into (38) 17:13;20:2;27:7; 35:5,7;37:21;39:13; 41:5;42:15;43:10; 45:25;57:16;75:22; 82:25;84:14,19; 87:18;90:1;93:21; 94:13;96:2;98:7;99:2; 103:25;106:8,14; 116:12;117:4;119:25; 128:16;129:7,8; 130:22;131:17;132:5; 146:18;147:5,25
intrinsically (1) 101:19
introduce (2) 12:23;14:4
introduction (3) 47:20;49:2,4
intuition (2) 100:11,12
intuitive (1) 84:8
intuitively (1) 80:11
invalid (5) 66:21,23;67:20; 68:5;150:4
inversely (1) 104:2
investigate (7) 143:22,25;144:5,5, 7,8,14
investigation (1) 144:2
invited (2) 31:19,20
invoice (3) 42:8;43:23;44:4
invoices (1) 42:13
involve (5) 36:24;38:6;75:23; 96:20;136:2
involved (6) 30:4,8;33:11;37:3; 56:25;58:14
involves (4) 34:1;75:22;76:19, 21
involving (3)

6:10;27:7;53:1
ironed (1)
51:16
isolation (2)
23:9;71:13
issue (22)
7:24;9:12;11:23;
13:25;14:15,15,18,20;
16:16;17:3;18:14;
20:12;21:15;23:15,
23,23;24:1,7,9;30:10;
55:10;153:6
issues (6)
8:25;9:4;14:13;
15:13;18:19;54:15
iterations (1)
100:6
iterative (1)
100:1
Ives (2)
146:3,7

J

Jacqueline (4)
83:4,5;84:16,17
January (1)
58:11
Jean (1)
4:13
John (1)
147:23
join (1)
110:24
joined (1)
97:20
joining (1)
102:7
jointly (2)
12:5;71:24
Jolio (1)
125:1
Josh (1)
6:25
Journal (1)
129:18
Jr (2)
84:10;92:4
Judge (1)
11:5
judgment (3)
4:19;54:1;148:1
Julio (1)
125:1
July (1)
32:25
Justice (3)
49:8,11;94:6
justifies (1)
117:7

K

keep (6)
7:8;51:13;56:20;
92:20;98:12;99:14
keeping (1)
51:16
ken (1)
8:1
key (14)
36:21;37:18,24;
42:5;43:11,16;54:10;
60:11;66:22;67:4;
68:7;82:5;93:4;
115:11
keyboard (2)
55:8;63:9
Khalifa (3)
15:19,21;130:2
kind (15)
51:4;54:20;56:4,19;
61:25;78:23;79:4;
85:21;86:16;95:3;
116:5;119:17;140:10;
142:19;144:12
kinds (7)
53:4,16,21;57:2;
84:25;118:3;146:25
knowing (2)
20:19;87:7
knowledge (5)
8:16,23;136:9,10,
13
known (8)
26:10;37:19;47:25;
72:5;96:15;107:20;
115:2,13
knows (1)
113:14
Kristian (1)
88:12

L

lab (1)
29:13
labeled (7)
12:10;26:10,13;
36:4;48:10,19;100:15
labelled (1)
58:5
laborious (1)
75:25
Labs (1)
29:8
lack (3)
19:20;21:1,13
languages (1)
29:10
Lares (10)
141:23;142:6,16,
25;145:3,25;151:4,20,
20;152:23
large (32)
26:13,14;33:11;
37:9;40:12,14,17,18;
41:4;42:25;45:16;
48:3,13;49:3,13;
56:12;60:9,10;65:23;
66:5;67:10;68:10;
70:21;81:14,16;95:1;
104:19;109:8;119:2;
124:3;142:13;150:12
largely (2)
100:19,20
larger (9)
7:18;61:3;65:5;
67:22;69:25;81:7;
104:14;126:20;
153:22
largest (10)
41:2;112:3;123:17;
124:1,2,2;125:16,17;
126:1,4
last (14)
14:14;46:21;52:17;
75:6,7;94:6;122:11;
125:20;131:24;
133:14;139:19,20;
145:11,24
lastly (1)
4:18
later (21)
14:19;16:5;27:7;
33:21;36:13;37:10,
21;38:5;43:10;51:17;
56:21,24;57:23;59:2;
60:16,21;76:2;78:5,
11,19;99:17
latter (1)
99:24
lawyers (2)
22:13,17
lays (2)
7:16;85:12
lead (1)
92:16
leading (2)
50:22;73:23
learning (8)
26:11,11,13;29:4;
31:15;36:2,2;49:22
least (10)
13:19;43:23;64:23;
107:9;126:15;130:25;
133:1;141:2,4;147:22
leave (2)
51:11;75:11
leaving (1)
56:24
left (3)
67:10;69:3;98:3
legal (1)
4:17
legend (1)
59:3
legible (1)
153:11
legitimate (1)
80:21
lengthy (1)
30:16
less (10)
20:15;48:7;51:8;
80:25;103:8;104:24;
105:22;106:11;
116:25;147:22
letter (1)
63:5
letters (1)
75:13
letting (2)
99:25;130:11
level (1)
34:16
license (1)
29:6
light (3)
93:14;95:16;144:15
likely (11)
14:24;30:1,5,6;
46:5;58:25;68:10,14;
82:10;94:23;147:22
Likewise (2)
37:7;38:25
limited (1)
20:9
Linacre (1)
94:5
line (4)
62:4;103:12;105:9,
10
link (6)
34:7,10;49:11;
121:21;128:16;
138:23
linkage (41)
32:7;33:2,24;34:1,
9,13,13,19;35:3,8,16,
19;37:2,19,24;40:13,
14,18;41:3;45:14,16,
20;51:5;59:9,18;
64:21;65:1,7;73:2;
74:5,6,17;79:20;84:1;
86:13,15;113:21,25;
122:15,18;129:20
linkages (5)
114:8,8,15;115:2,8
linked (2)
29:5;115:3
linking (3)
21:21;34:2,3
list (13)
12:5;20:16;30:16;
54:20;128:25;139:3,
7,8,12;146:1;150:19,
20;155:6
listed (1)
146:5
lists (2)
12:4;22:13
literally (4)
59:23;79:7;83:16;
112:12
literature (9)
83:25;87:10;88:12;
90:19;106:3;107:22;
108:25;116:9;117:7
litigation (3)
32:23;148:24;
149:13
little (12)
12:3,15;61:4;62:15;
73:20;74:6;75:18;
100:13;113:4;122:23;
140:15;155:20
lived (1)
147:20
Lizotte (4)
4:20,23,23;51:21
location (1)
106:5
login (15)
77:25,25;78:12,12,
14,15,16;79:2,3;
119:14,15;121:12;
122:24;123:10;
155:21
long (5)
60:17;70:16;83:10,
14;92:9
longer (3)
75:2;100:7;145:14
look (56)
14:3;15:6;16:1;
27:24;31:5;33:8;
36:12;46:15;54:13;
56:16;58:16,20;
61:10;63:25;64:14;
70:4,5;71:4,23;75:6;
77:2,19;80:2;82:15;
84:9;85:9;88:22;91:3;
94:7;100:4;108:19;
110:1;111:23;112:2,
20;117:15,16;118:5,
24;119:7;120:8,17,
21;122:16;123:13,25;
124:5;130:19,24,24;
134:19;137:21;
140:14;152:2;154:5;
155:20
looked (9)
15:25;54:15;60:23;
85:15;134:18;144:10,
18;147:3;148:9
looking (43)
5:14,15;12:22;
33:11;53:2;57:1;60:8,
13;61:8,12;64:15;
65:24;66:3,25;71:14,
19,22;78:24;84:11,14,
19;86:16;87:21;
91:23;100:7;109:3;
111:9,16;119:19,20;

120:5;124:9,16;
126:3;129:23;142:18;
143:3,6,18;146:25;
147:1;148:23;152:14
looks (6)
43:14;53:17;55:8;
91:18;125:23;131:11
lookup (1)
109:22
loosely (2)
106:8;125:9
looser (2)
105:23;107:3
loser (9)
39:20;111:5,13;
133:17,20,25;134:5,
14;135:3
losers (4)
16:12;40:8,10;
134:22
losing (3)
73:3,9;74:6
lost (2)
40:10;156:2
lot (13)
27:6;54:6;58:13;
64:11,12;80:10,24;
81:17;85:23;96:19;
107:23;111:20;
150:25
lots (2)
96:18;118:21
Louis (1)
4:13
low (6)
14:5,15;97:6;
104:23;106:17;
115:22
lower (8)
31:13;59:16;94:25;
95:1;104:6;106:22;
123:19;125:13
lowercase (3)
73:5,8,18
luck (5)
66:3,6;67:14;69:21,
22
lunch (4)
10:17;91:15;92:18;
101:4
Lyne (141)
5:9,9,13,16,21;
6:14;7:14,16;8:17;
9:2;10:6,7,23;11:14,
18;13:1,2;15:7;16:19;
17:3;19:13;24:15;
25:5,7,9;27:12,14,19,
22;28:1,2;31:6,7,9,11,
12;32:6,13;36:8,10;
41:10,21,23;42:1,23;
43:2;44:10,12,14,24;
45:2,6,7;47:14,15;
49:17,19;50:17,20;
51:19,24;52:2,4,5,7,
11,12;60:2,3;61:1,6;
62:20,23;65:16,17;
66:18,19;68:3,4,18,
19;69:9;72:20,21,22;
74:8;76:23;79:13;
81:2,3;83:19,20;
85:19;87:23,24;89:7;
91:13,19,21;92:10,12,
19,21,23;93:1;96:5,6;
99:11;101:11,12,13;
102:10,11;103:1,11,
21;104:9;114:6,14;
118:4,9;119:3,4;
128:22;129:9,17,25;
130:9,13;131:2,18,25;
132:6,12,15,16,20,25;
133:5;147:13;153:12

M

machine (5)
26:11,13;29:4;
31:14;36:2
Maia (2)
155:21,22
mail (2)
149:5,7
main (3)
37:25;67:21;124:15
mainly (1)
38:18
majority (7)
56:12;68:17;141:3;
142:1,22;146:12;
149:22
makes (9)
21:13;22:19,25;
34:25;48:2;66:21;
84:8;89:21;113:1
male (1)
94:25
manageable (1)
89:5
manner (1)
127:9
Manoa (1)
32:2
many (32)
8:20;19:19;26:19,
22;30:8;31:19;37:24,
25;40:7;50:1;56:6,14;
68:21;78:19;81:16,
17;83:10;86:2;90:22;
100:18;116:7,8;
118:25;134:18;
135:23;136:6;141:4;
144:10,18;145:22;
148:7,16
map (1)
90:19
Mar (2)
146:3,7
margin (4)
66:8;69:19,22;
70:10
Maria (34)
123:16;126:20;
127:11,12;139:23;
140:15,18;141:12,13,
23;142:6,7,8,8,14,15,
16,25;143:7,8,8,10;
145:2,3,20,21;146:8;
147:6;148:5,22;
151:4,8,20,20
MariaNieves@mariashousecleaning (1)
151:13
Mariashousecleaning@gmailcom (1)
143:9
marked (1)
137:22
Martin (2)
23:1,2
Martinez (2)
82:18;84:12
mashing (1)
63:9
Massachusetts (5)
22:15;25:2,2,13;
148:7
massive (4)
12:14;20:14,15,15
match (98)
22:12,16;23:4;
34:17,17;35:4,6;37:5,
8;38:1,3,8;46:2,5,6,9,
13,17,18;47:5,10;
48:8;67:1;73:10;
74:23;75:8,21;76:5,
20;77:22;78:2,8,16;
79:1,7;80:9;82:24;
83:8,16;85:6,7;87:15,
19;88:18;90:20,21,22,
22,23;93:6,9,19,23;
94:10,11,12,13,16,17,
19,20,22,24,25;95:9,
21;96:1,11,15;97:1,2,
3,6,19;98:4;101:16,
17,17;102:2,17,19,21;
103:7;104:5,6;107:1;
113:9,10,14,17;
116:24;121:4;124:18,
19,21;125:24;128:9;
146:15
matched (4)
79:2;80:2;90:11;
121:18
matches (16)
37:1;73:3;77:18;
84:20;100:19;116:3,
4;117:1;120:24,25;
121:10,16;124:16,20;
146:21,22
matching (26)
34:21;36:21;41:6;
55:8;75:5;79:1;82:6,
11,17,24;83:1,7;
84:24;85:5;86:23;
89:4;90:3,3,5,8,9,13;
93:15,16;117:9;121:7
material (1)
130:17
materials (1)
132:4
mathematics (2)
29:19;30:22
matter (5)
10:9;11:18;56:23;
82:21;112:16
matters (5)
4:4;10:21;88:15,16;
131:7
maximization (8)
37:20,22;47:13,17,
25;49:10;100:12;
128:6
may (24)
8:11;9:2;16:20;
19:2;23:2;24:2,3;
38:24;59:8;61:4;
84:18,18;86:16;
113:4;115:24;124:21,
21;133:4;141:16,18;
142:11,13;151:7;
153:19
maybe (30)
9:9;23:4;38:23;
39:1,2;50:7;61:1,11;
62:4;68:20;75:2;
80:17;83:21;86:24,
25;92:1;96:12;107:8;
113:22;123:9;125:3;
130:7;132:25;137:6;
143:15,16,21;147:8;
151:11,11
mean (29)
11:2;24:19;40:15;
65:19;66:14;67:10;
79:9;81:12;86:17;
90:1;92:1;98:23;
104:15;120:7,22;
124:20;126:9;132:23;
134:16;135:18;
136:10,13;145:8;
146:6,24;151:10;
152:6;155:4,11
meaning (4)
22:23;52:21;97:1;
150:18
meaningful (1)
73:8
meanings (1)
90:11
means (6)
29:21;30:13;46:4;
98:21;103:9;116:11
meant (5)
55:4;56:1;102:15;
121:19;130:9
measure (2)
115:14;116:24
measured (1)
82:11
measures (1)
22:25
Melba (1)
126:5
members (2)
8:25;85:25
memo (1)
13:5
memory (1)
81:17
mention (1)
145:24
mentioned (14)
19:22;41:14;43:21;
53:15;58:15;59:25;
72:3;74:19;75:20;
112:24;115:12;
117:24;135:10;
145:21
merely (3)
52:17;110:3;112:15
merge (1)
21:22
merged (1)
89:11
messier (1)
33:17
messy (2)
49:6;118:2
method (33)
19:16;33:6;34:5;
45:12;48:9,21;49:2;
50:3;68:8;86:16;87:6;
88:20;89:18,21,22;
94:11;98:17;99:8;
102:5;106:18;107:14;
108:7;112:25;113:5,
13;117:20;119:19,25;
121:10;125:12,17;
141:19;144:11
methodological (1)
131:3
methodologies (2)
24:9,10
methodology (34)
14:17,21;15:1,24;
17:8,9,25;18:11;
19:17;21:14;22:9;
23:7,8,19;26:18;
34:15,24;36:6,14,15;
47:10;50:14;51:11,
18;55:12;58:2;66:9;
98:2;128:16;129:19;
130:23;138:7;150:22;
154:14
methods (8)
22:10;29:18,20;
34:13;35:8;36:3;
49:13;118:1

metrics (1)
128:6
Michael (1)
5:3
middle (8)
38:24;46:21;55:11;
63:6,7;74:16;125:20;
139:20
middling (1)
83:14
might (30)
11:21;24:4,4;38:23,
25;48:20,24;51:10;
52:16;59:3;68:16;
73:4,8;74:12;80:12;
82:20;83:15;85:25;
86:21;97:9;103:13;
105:15,18;112:6;
126:21;144:1,4;
147:8;154:3;156:8
Milagros (7)
140:18;141:12;
142:8,14;145:4,21;
146:8
million (37)
40:4,4,5,6,8,16,17,
23,23;43:1,7,15,20;
44:2,3,9;53:18;54:21;
60:10;70:14,15;
73:15;76:15;77:12,
13,17;79:8;81:8,8,10,
13;93:20,22;109:4;
110:15;124:11;
125:18
millions (4)
40:15;57:17;80:5;
138:15
mind (3)
19:6;51:13;109:18
minimize (3)
67:15;105:2;115:24
minimized (1)
66:6
minimizing (1)
99:8
minimum (2)
34:22;108:21
Ministry (3)
49:8,11;94:6
minus (5)
70:22;115:19,20;
117:2,3
minuscule (1)
54:19
minute (2)
53:15;61:11
minutes (4)
10:10;100:7;
132:24;152:4
mismatch (2)
91:9;95:3
missed (2)
115:2,7
missing (8)
91:3,4,5,6,10;
122:12;123:3;125:2
mistake (1)
23:1
misunderstanding (2)
6:19;72:19
misunderstood (1)
130:7
MIT (6)
17:15;25:13,14;
28:23;29:11;32:4
MIT's (1)
25:20
mix (1)
15:2
mobile (1)
77:24
model (11)
33:2,4;45:4,9,15;
46:4;91:14;92:24;
95:10;128:6;146:15
models (1)
29:4
model's (1)
116:21
modern (1)
37:10
moment (10)
19:10;38:13;51:20;
81:12;96:17;101:15;
105:4;132:9;140:21;
141:7
Monday (3)
12:7;15:11;16:5
money (8)
112:3;135:11,16,
21,24;136:4,5,9
months (1)
13:20
more (112)
7:22;10:9;14:23;
20:15;25:23;27:20;
28:10,24;33:14,17;
34:3;35:7,13,20,21;
37:6;39:16;40:7;48:6;
53:20;54:22,25;
55:18,20;58:6,19;
59:20;63:6;64:25;
65:5,12;66:15;69:13;
70:6;72:6;75:25;
77:14;78:8,9,9,23;
79:17;80:7,10,17;
81:12;82:1,2,3;83:6;
84:23;86:25;88:25;
89:9;90:21;91:7;
92:12,13;93:12,13,14;
94:17,17,20,23;97:17;
100:9,9,21;101:2;
103:8,17;104:4,16,25;
105:13,13,19,19,23,
24;106:7,11;107:14,
15;111:25;112:3;
113:8;118:23;120:23;
121:22;125:8,13;
126:19,24;127:17,22;
134:8,9,16,25;142:18;
143:3,4,15,17;144:15;
150:4,10,13;151:10;
152:22
morning (4)
5:1,3,22;156:10
most (20)
12:13;19:16;23:3;
28:4,23;45:15;49:12;
70:4;79:22;97:14;
100:20,24;122:19;
125:5,6;135:7,7;
140:1;149:2,6
mostly (1)
59:22
motion (17)
4:5,7,9,12,18;5:23;
6:9,9,20,21;7:5,16;
8:4;9:12;10:1,3;11:20
motions (2)
9:23,24
move (5)
99:13;129:9;
132:23;133:14;153:8
moved (1)
39:2
much (18)
47:9;54:25;56:23;
58:19;65:24;66:2;
69:15;73:3;76:14;
80:7;81:7;89:23;91:2;
113:8;123:6;136:9;
147:22;149:15
Mullinax (1)
84:15
multiple (7)
38:23;45:22;74:11;
125:20;149:2,3,3
multiplication (1)
134:25
multiplicative (1)
135:1
multiplying (1)
95:7
must (2)
13:19;95:2
mute (1)
132:16
myself (4)
11:7;88:9;132:9;
144:19

N

NA (1)
91:8
name (114)
22:16,19,20,23;
23:3,4,5;24:24;26:17,
19,20;33:15;38:18,23,
24;42:6,11;46:21,21,
21;47:4;53:6;54:3,13,
24;55:1,7,11,13;
58:16,21;59:11;
61:20,23;62:6,10,12,
25;64:16;71:5;73:10,
20;77:24;78:12,12,15,
15,20,20;82:9,10,22;
84:3;85:5,17,18,20,
23,23;86:3,3,5,8,21,
25;87:13,25;88:10,23,
25;89:10,12,17,19,23;
90:15;94:8;99:22;
101:19,24;108:22;
119:13;120:8,19,24,
25;122:10;123:7,16;
124:14,25;125:19;
126:11;139:15,19,19,
19,20,20,20;140:24;
142:13,16;143:13,13;
145:24;147:5,8,10,22;
148:1;150:18;153:22;
154:2
named (4)
30:17;45:12;126:5;
153:19
names (31)
54:5,25,25;61:24;
62:17;63:1,3,6,15;
78:16;82:10;83:9;
86:11;87:5,17;88:6,7;
92:3;100:22;123:10;
125:20,20;130:15;
139:16;140:1;146:4,
16,17;147:19;148:2,3
Namrata (3)
112:24;115:12;
116:21
narrowed (1)
71:24
national (1)
147:10
natural (3)
88:11;92:17;156:4
nature (1)
30:15
near (3)
90:21;125:5;146:25
nearly (5)
70:7;81:24;119:2;
125:18;127:5
nearness (1)
102:18
necessarily (1)
36:3
necessary (4)
71:12;87:20;
150:17;152:11
need (27)
6:12,16;8:21;9:9;
16:19;21:4;26:12;
27:17;35:12;36:20;
37:6;52:20;68:21;
75:24;76:4,13;78:4;
80:13,14,16;84:8;
92:4;109:20;110:1;
132:18,24;149:25
needed (1)
29:5
needs (13)
10:18;21:5;34:16;
36:17;41:1,15;46:8;
51:13;52:21;66:1,2;
76:10,12
negative (13)
24:5;39:20;111:4;
113:21;114:15,21;
115:13,17,17,21;
116:25;117:3;155:14
negatives (12)
113:6,19;114:18,
25;115:1,7,14;116:8,
13,17,17;152:15
neighborhood (1)
59:13
net (81)
16:12,12;18:8,10,
12;19:22;20:8,13,17,
19,20,21,21,23;21:5,
12;23:24;24:6;39:11,
13,17,18,19,20;40:7,
7,9,10;42:15;44:6,19;
110:20,25;111:1,2,4,
5,12,13,14;112:10,11,
13,13,18;123:17,20,
21,25;124:12;125:18;
133:16,16,20,24,25;
134:5,8,9,14,14,22,
22;135:3,3;138:14;
149:19,19;150:18,19,
20,24;151:1,2,21;
153:23;155:2,6,8,9,14
neuroscience (1)
25:25
nevertheless (13)
55:2;58:6;102:22;
118:2;120:3,14,16;
121:10,18;122:18;
123:4;125:24;127:10
new (7)
9:1;13:6,7,9,12;
91:4;112:17
newly (1)
4:8
newly-disclosed (2)
11:20,21
next (60)
33:23;36:8;37:14;
38:5;41:24;45:10;
46:11;47:14;50:17,
23;55:16;60:2;62:3,6;
65:16;66:18;68:3,18;
72:21,24;74:8;75:18,
18;76:23,24;79:13,
14;81:2,21;83:19;
85:19;87:23;89:6,7;

91:17;92:6,8,23,25; 93:24;96:5,13; 102:10;103:1,10,11, 12,20;104:8;105:9, 10;107:8;109:14; 111:6;113:22;114:5, 6;119:3;124:20; 141:22
nice (6)
78:25;111:16; 117:23;118:20; 119:16;125:8
Nieves (25)
123:17;126:20; 127:11,12;139:23; 140:15,18;141:12,13; 142:7,8,8,15,15; 143:7,10;145:4,21,21; 146:9;147:6;148:5, 22;150:23;151:8
night (1)
131:24
nine (9)
38:6;96:10;109:20; 110:18;111:12; 112:10,18;122:13; 141:12
ninety (4)
98:4,11;108:6; 117:11
ninety- (1)
117:2
ninety-five (8)
55:12;63:3,17;70:9; 90:15;98:5,11;108:6
ninety-four (1)
95:9
ninety-seven (1)
55:2
ninety-six (1)
82:24
ninety-two (2)
83:8;90:23
nodes (1)
108:5
noisy (1)
118:2
Nom (1)
54:13
nondigit (1)
74:24
None (6)
13:5,20,21;14:6; 84:19;98:6
non-link (2)
114:22,22
nonstarter (1)
81:19
nontriangular (1)
14:5
nonzero (1)
143:25
normal (2)
88:10;104:20
normally (1)
10:2
North (1)
83:24
nose (1)
121:16
note (9)
22:8;42:23;54:24; 65:2;67:6;87:22; 89:17;117:5;128:24
notebook (1)
128:24
notice (3)
59:10;111:20;121:8
notion (3)
108:1;115:1;116:10
notions (4)
108:24;113:23; 115:16;116:5
number (98)
4:2,3;11:2,3,3,4,7,9, 11,12,12;13:7;25:2; 28:9;29:11;31:4,18, 21;34:23;36:20;37:4, 17;38:19;39:4;40:17, 23;42:11;44:20;53:5, 21,22;55:1,21;56:16; 57:8,13,15,16,18,20, 21;58:19,21;59:11; 61:19,19;64:3;68:10, 25;69:1;72:5,13; 74:21,22;75:24; 76:15,18,25;77:7,8, 24;78:3,9;79:18;80:3, 23;81:7,14,22;83:25; 92:4;93:6;105:19; 107:22;111:23;117:5; 118:14;120:9,11; 122:21;123:15;124:3; 125:23;126:10; 127:18,18;131:15,22, 23;133:24;134:3,7,11, 13,19;140:2;146:13; 150:7
numbers (34)
39:13,21,22;46:24; 53:7;57:3,15;64:5; 69:13;74:19,20,23; 75:2,8,10,12;78:23; 91:12;93:18;95:6,20; 97:9,14;107:6; 110:20;112:15,16; 120:10;124:15;125:2; 126:18;150:9;152:11; 155:25
numerical (7)
56:24;57:2;63:14; 75:13;89:16;90:1; 93:21
numerous (1)
143:23

O

oath (1)
101:11
object (1)
16:19
objected (1)
130:4
objection (15)
9:9;16:25;32:8,10; 128:25;129:3,13,15, 23;130:2,14;131:6,8, 13;147:13
objections (2)
129:12;130:17
objectives (1)
99:7
objects (1)
114:11
observations (1)
119:6
obtain (2)
33:4;46:18
obtained (1)
17:10
obvious (1)
86:10
Obviously (10)
9:15,25;10:19; 40:15;45:17;61:21; 141:3;152:22;155:9, 11
occasional (4)
55:10;65:4;141:18, 20
occasionally (3)
55:7;57:3;63:8
occur (3)
67:19;123:19;125:4
occurred (1)
121:17
occurs (3)
66:17;95:3;97:4
o'clock (5)
16:5;92:18,22; 101:4;156:11
October (2)
13:5;16:5
odd (1)
126:23
off (6)
6:2;41:17;84:17; 101:6;133:14;137:10
offer (1)
128:22
often (30)
23:22;26:9;35:16, 17;45:23,24;48:10; 51:16;53:1,19,20; 54:18;55:7;58:4,18, 20;61:25;64:15;65:9, 23,24;77:5;78:23; 85:22;91:6;94:22; 117:24;120:24,25; 150:8
old (1)
28:7
once (16)
30:11;36:19;38:13; 39:6;44:22;76:2,3; 89:24;95:15;101:23; 108:10;109:13; 110:18;146:19;149:6; 155:16
one (196)
8:20;9:25;10:9,24; 17:6;18:6,7,21;19:10; 20:1,2,3,14,15;22:14, 15;24:4,4,5;26:14,22; 27:17;31:13;32:20; 34:2,2,3,3,6,16;36:17, 19,23;37:3,6,8;38:7, 13;39:14,16;40:8; 41:1,2,6,10,11;43:8, 23;45:13,18;46:7,7,8, 22;49:12;50:1,7; 51:13;52:21;53:4,16; 54:18;55:18;57:13, 15;60:10,11;61:14; 64:2;65:22;66:1,2,22; 67:4,15,18;70:24; 71:15,20;72:12; 73:22;74:22;75:11, 22;76:19;77:14,23, 25;78:6,20;80:3;82:8, 15,21;83:1;84:10,10, 23;85:3,24;86:5;87:1; 88:6;89:8;90:19,20; 91:4,24,24;92:25; 93:17;95:18,19; 96:23,23,25;98:12; 99:21;100:11;101:15; 102:23,24;103:12,14, 20;104:12;105:15; 106:2,25;107:15; 108:15,16;110:20; 112:10,18;113:12,15; 114:7,12;115:9,19,20, 22,24;116:5,25;117:2, 3,17;118:18;120:13, 19,20;121:11,22; 122:9,11,22;123:15, 16,23;124:20;125:6, 21,22;127:14,24; 132:1;133:14;134:2; 142:9,20;143:12,15, 17,23;144:1;145:6; 148:17,18;149:1,6,7, 8,20;150:5,11,14; 151:12;153:22; 154:12,12,16;155:17, 18,22
ones (22)
43:21;51:12;54:7; 57:24;58:3;59:1,16; 63:14;65:5;67:1;76:3; 77:17;80:15,17,19,20; 83:2;98:11;104:6; 107:3;123:25;152:18
only (34)
7:17;18:7,14;22:15; 34:8;44:5,22;53:21; 58:8;70:11;75:2;78:4; 80:8,14,16;81:8;84:7; 96:23;102:24;103:14; 104:4,19;112:14; 113:2;117:5;118:25; 120:22;123:9;125:17; 130:14;134:5;140:24; 142:9;150:8
onto (1)
31:25
open (1)
130:10
opening (6)
11:14,16;17:2;19:1; 26:6,17
operate (1)
37:22
operations (1)
81:15
opinion (6)
8:20;20:4;128:15; 136:14,15;149:11
opinions (1)
7:4
opportunities (1)
51:14
opportunity (2)
15:5;117:23
opposed (3)
91:8;133:25;135:18
opposites (1)
115:25
optimized (1)
108:16
orange (11)
80:14,15,22; 102:25;103:18,24; 104:13;107:15; 114:12,20,23
oranges (3)
15:3;102:16,16
order (12)
40:24;48:7;57:10; 62:9;63:2;66:1;71:18; 79:18;100:6;124:10, 12;150:17
ordinarily (1)
12:10
ordinary (1)
41:7
orient (3)
36:13;50:18;96:7
origin (1)
147:10
original (4)
62:8;67:8;110:2;

123:15
origins (1)
47:18
others (19)
35:7;51:8;79:5;
88:14;93:13,14;99:1;
102:18;104:18;
106:11;108:15;
121:11,18;122:15,22;
125:10;134:18;140:2;
152:22
otherwise (6)
36:24;57:10;75:22;
79:7;87:13;88:8
ought (3)
77:22;114:1;144:17
out (40)
6:16;7:8,16;16:10,
12;27:10;33:18;47:8;
49:17;51:9,16;52:21;
56:24;60:14;62:16,
17;63:17;65:1;67:16,
18;69:7;70:1,18,22;
85:12,16;86:11;
88:21;93:22;95:5;
99:16;101:2;112:5;
117:23;123:6;124:24;
127:4;128:13;137:6;
148:11
output (19)
18:2,17;20:14;46:2,
9;47:10;49:22,23;
95:21,24;108:7;
112:23;116:19;
117:15;126:15;128:4;
130:23;147:3;150:22
outside (2)
8:1;10:15
over (30)
24:3,3;56:12;
103:16;104:21,24;
106:1;107:13;108:1,
17,21;109:1,7;
113:12;114:17;
115:23;124:11;
141:18;142:12,17;
143:1,15,16,23;144:9,
13;151:16;152:8,12;
154:6
overall (6)
37:4;50:24;63:21;
71:10;111:16;113:16
overlap (1)
59:19
overstated (1)
155:15
overwhelming (2)
68:17;78:1
overwhelmingly (3)
68:14;71:21,21
own (10)
12:20;23:18;26:2;
99:1;102:25;104:12,
13;113:5;114:21;
116:21
owner (2)
86:22;87:3
owns (1)
86:25

P

P25 (1)
130:15
page (21)
21:8;28:15;29:14;
31:5,8,22,25;41:7;
60:2;130:15;135:6,7;
137:23;140:3,14;
141:3,4,6,22;145:15;
155:25
pages (4)
13:9;124:13;125:4;
141:4
paid (2)
43:23;44:4
pair (38)
34:20,24;35:4;38:1,
3;46:3,4,13,16,18;
47:9;48:5;76:4,5;
77:20;80:3,14;81:9;
85:7;87:18;89:14;
90:6,7;93:5,8,18;94:7,
12,18,24;95:18,21,22;
96:1;97:4,7;102:4;
113:15
pairing (1)
102:3
pairs (24)
48:8;75:24;76:3,19,
21,25;77:7;79:16;
80:1;81:6,13;84:8,20;
85:22;93:20,22;
94:22;95:17;97:11,
19;98:7;102:17;
113:10,11
pairwise (16)
76:10;79:18;80:23,
25;81:9;84:6;87:19;
97:1;98:4;101:17;
102:3,20;104:6;
114:3;128:7;146:21
Papas (7)
4:25,25;13:11;
27:12;51:21;61:2;
131:23
paper (9)
15:18;41:7;45:13;
49:15,21;50:8,16;
88:14;112:25
papers (4)
26:3;28:9;45:14;
50:1
paperwork (1)
12:13
paragraph (3)
137:23;145:15,17
paragraphs (3)
85:11;137:25;138:1
parameter (7)
108:12,14,16,23;
109:1,6,16
parameters (4)
47:7,11;100:8,10
parcel (1)
86:24
parentheses (2)
56:22;57:2
part (25)
6:2,17;7:1;8:12;
9:12;14:14;21:1;
26:13;33:11;43:25;
61:18;66:4;84:2;
87:16;95:11;97:22;
121:7;127:12,13;
136:14;139:9;147:25;
149:7,10;153:3
partial (4)
35:6;90:25;101:1;
121:9
partially (2)
4:10;8:4
participant (27)
8:23;21:6,10,12;
31:20;38:7,22;52:24;
57:8;105:15;110:23;
111:15;120:1;123:1;
138:8,11,12,13,15,17,
24;139:17,22;141:11;
142:10;149:4;154:14
participants (19)
7:19;8:1;9:1;21:2;
32:24;34:6;40:5;
42:18;105:18;111:19;
112:4;120:6;135:11,
17,21,25;136:17;
139:1,3
participant's (2)
59:4,6
participate (1)
18:22
particular (19)
14:7;17:9;34:11,21;
47:9;50:6;53:13;
66:25;84:7,21;96:20;
97:18;108:11;118:22;
126:3;138:23;144:23;
146:10;147:21
particularly (2)
51:10;73:6
parties (12)
6:3;9:18;10:4,10,
11,20;12:3,6;13:3;
16:4,23;41:15
partition (1)
98:21
partitioning (1)
96:24
parts (4)
51:6;55:21;90:2;
105:13
pass (5)
56:6;89:19;95:19;
111:21;112:21
passed (4)
40:1,2;55:2;56:9
passes (2)
56:14;80:8
password (4)
78:8,15;121:16;
123:3
pattern (5)
66:25;78:23;
119:17;121:13;125:7
patterns (3)
62:13,16;78:22
Pause (3)
41:12;51:23;52:1
pay (1)
66:1
paying (1)
123:21
payments (2)
135:19;136:5
payoff (2)
35:14,20
PDFs (1)
62:21
peaks (1)
112:2
peer-reviewed (6)
14:22;17:25;18:15;
24:8;30:19,21
people (18)
7:23;8:15,24;63:6;
64:14;73:10;74:20,
25;78:15,19;86:12;
87:5;88:7,8;122:19;
149:19;154:2;156:3
people's (1)
61:24
per (11)
13:13;38:7;39:14,
15,18;44:19,20;
78:10;93:18;110:20;
111:25
percent (48)
46:4,5;54:19;55:2,
13,18;56:9,13;58:9;
63:3,17;64:2;70:9,10,
19,21;82:12,24;83:8;
85:6,7;90:15,16,23,
23;93:7;95:9;98:3,4,
5,5,10;102:8,9;108:6,
6;116:14,15,23;117:1,
3,4,11;122:7;126:21,
25;128:12;146:10
percentage (4)
54:4;55:9;82:12;
154:24
percentile (2)
107:5,6
percentiles (1)
107:2
perfect (10)
54:8;65:2;90:20,21,
22;94:10;100:19;
127:5,9;150:11
perfectly (1)
9:14
perform (9)
32:22;46:8;48:2;
51:5;53:17;65:6;77:6;
90:8;107:22
performed (4)
48:9;64:19;112:25;
127:25
performing (1)
88:23
performs (1)
41:3
perhaps (8)
18:22;35:18;61:1;
66:25;122:14;126:22;
127:2;150:8
period (1)
83:7
permissible (1)
74:11
permission (1)
133:7
permit (3)
11:14;17:7;18:16
permitted (2)
16:4;17:12
person (36)
34:8;38:25;39:2;
49:6;86:5,6,8;92:5;
119:22;120:7,15;
121:23,25;122:20;
124:22;141:18;142:6,
9,23,24,24;143:15;
146:4,8,8;148:2,24;
150:5;151:3,15,17,22,
23,25;155:8,17
personal (4)
7:21;8:16,23;88:7
persons (3)
34:6;86:4;87:8
ph (14)
20:6;29:9;49:15;
83:4;84:10;88:13;
112:25;125:1,1;
126:6;130:8;140:18;
141:23;146:3
phase (3)
20:2,2,3
phenomenon (3)
63:20,21;65:11
phone (32)
10:12;38:19;39:4;
42:11;53:5,7,21,22;
56:18,19;57:3;58:19,
21;74:19,20,23;75:2,
7;119:13;120:9,11,

11;122:17,18;123:8,10;124:14;125:2,23;150:7,7,8
phrase (1)
64:16
physical (2)
45:1;141:8
physically (2)
102:19;135:18
physics (1)
30:23
pick (3)
60:7;97:18;148:11
picked (4)
71:25;72:1,2;134:13
picture (3)
24:6;53:6;111:16
piece (1)
80:15
pieces (11)
12:15,16,23;16:9;20:14;79:24,25;80:14,22;81:25;82:1
pile (1)
15:19
pinpointed (1)
149:20
pioneered (1)
49:4
pivots (1)
22:21
place (4)
44:2;71:19;125:24;127:4
plaintiff (3)
4:8,9;118:6
plaintiff's (5)
49:18;55:24;128:24;129:10;130:18
plan (1)
10:20
planning (1)
10:16
plateaus (2)
70:7,7
plated (1)
19:16
plausible (4)
29:25;134:6;143:4;144:14
plausibly (2)
152:11,22
play (6)
8:17;85:16,20;89:17,20,23
played (1)
44:21
plays (3)
26:19,20;90:3
Plaza (1)
121:11
pleading (2)
11:11,12
pleadings (2)
5:25;9:21
Pleasant (1)
143:9
Please (29)
4:2,21;17:2;24:25;28:1;36:9;41:22,24;50:18;60:2;62:3;65:16;66:18;68:3,18;74:8;76:23;81:2,21;83:19;85:19;87:23;96:5;101:8,10;102:10;103:1;119:3;137:12
pleasure (1)
91:15
plenty (1)
60:16
plugging (1)
136:25
plus (3)
57:2,2;70:22
pm (7)
12:7;13:23;101:6,7;137:10,11;156:12
podium (5)
5:15,18;19:7;132:19,23
point (40)
13:16;14:19;17:14;28:7;32:9;35:15;38:20;39:12;47:22;49:17;63:18;66:5;67:7,21,23;70:1;88:5,15,19;89:19;95:25;98:2;101:18,24;102:1;109:17;110:3,14,16,19;112:14;113:1;117:23;118:15;121:7;123:6;129:2,20;149:12;154:17
points (16)
8:20;29:24;50:7;54:17;66:24;67:12;69:1;70:3,5;78:6;91:2;94:14,17,20;109:10;133:14
polls (1)
66:10
pool (1)
12:14
populated (1)
117:17
population (9)
7:18;65:24;67:9,24;68:2,25;69:25;70:2,13
populations (1)
69:17
portion (8)
54:19;106:7,19,22;141:7,9;142:3;148:1
portions (12)
4:10;6:16;7:6,17;8:6,21;9:3;105:14,22,24;106:6,10
Portuguese (2)
42:6;130:15
position (1)
15:3
positioned (1)
17:17
positions (1)
31:22
positive (15)
24:4;39:19;94:15,17;111:3;113:21,25;114:5,13;115:5,16,16;122:15;126:21;155:9
positives (12)
113:6,19;114:7,12,16;115:14;116:7,13,17;122:4;126:22;152:15
possibilities (1)
142:11
possibility (1)
66:3
possible (29)
6:1;13:8,9;33:1,7;37:1;48:2;51:7;55:19;67:13;68:13;81:18;83:1;84:20;86:2;98:17;103:13;104:23;113:17;115:22;116:1;120:5;121:22;142:18;143:13;144:1,5;154:23,24
possibly (2)
46:21;62:9
post (1)
94:19
postal (1)
38:19
post-cleaning (1)
77:2
postcode (1)
99:22
potential (5)
34:17;73:2;74:5;81:6;82:9
potentially (3)
75:11;81:15;97:10
power (4)
45:17;47:21;89:1;93:16
PowerPoint (1)
12:9
practical (3)
11:18;26:2;30:9
practice (4)
73:25;113:2;128:1,19
pre (1)
106:15
precise (1)
8:5
precisely (1)
100:9
precision (7)
115:18,25;116:2,14,15;117:2;152:13
preclude (1)
4:15
predefined (3)
139:3,6,8
preeminence (1)
21:18
preeminent (3)
17:18;18:23;21:17
prejudice (1)
16:25
prejudiced (1)
16:20
prejudicial (1)
14:1
premise (7)
139:18;141:15,21;142:21;148:25;149:9;151:14
preparatory (4)
36:5;75:20;89:9;106:13
prepare (3)
17:23;32:15;36:21
prepared (12)
10:23;11:15;12:9,12;15:10,15;28:10;37:15,16;150:18,20,21
presence (2)
118:2;124:25
present (17)
4:21;6:24;16:4;53:2;58:6,8,12,18,20;77:5,23;78:18;98:7;121:20,20,22;143:24
presentation (6)
15:18;17:20;30:11;32:15,15;132:1
presents (1)
23:12
preserving (1)
79:1
presumably (11)
46:25;63:15;120:2,15;121:3;122:17,21;123:11,18,24;126:9
presuming (1)
154:11
pre-trial (2)
13:5;16:3
pretty (11)
8:5,22;22:14,16;70:16;80:3;94:10;120:16;126:10;144:13;147:9
prevalence (2)
123:20;147:18
prevalent (1)
147:5
prevent (1)
152:23
previous (7)
43:21;62:8;77:10;91:24;99:20;103:4;108:8
previously (2)
77:10;126:17
primarily (2)
44:2;123:16
principle (4)
40:20;59:17;80:14;100:15
principles (1)
22:10
print (1)
40:3
printed (2)
15:18;41:6
prior (4)
13:17,22;21:7;116:19
Probabilistic (32)
17:16,18,20,23,24;18:2,23;25:15,18,20;29:9,11,17,18,20;30:11;31:15,16;32:7;33:10;34:13;35:2,3,8,19,24;37:18;45:14,16,20;47:18;86:15
probabilities (11)
46:10;75:21;95:7;96:11,15;97:1;104:6;107:2;113:10,14;152:7
probability (24)
29:23;38:1,8;46:2,18;47:11;93:6,19,23;95:8,21;96:1;97:2,3,6,20;98:4;101:17;102:17,19,21;103:7;104:5;113:18
probable (1)
29:25
probably (1)
134:25
problem (9)
7:20;8:4;21:3,17;22:18,22;23:21;24:4;91:21
problems (2)
22:7;23:11
Procedure (1)
4:14
proceed (10)
6:4;11:1;17:8;19:3;25:5;33:19;64:21;127:19;128:5;132:11
proceeded (2)

36:15;43:18
proceeding (1)
5:23
process (31)
13:4;14:22,22;
33:10;34:15;51:2,9;
53:14;65:6,12;68:6;
75:15;77:6;78:11;
79:14;81:1;95:12,24;
100:25;102:1,13;
109:14,15,20;110:12;
111:18;113:21;
119:10;128:4;131:3;
153:1
processed (1)
119:8
processes (1)
18:17
processing (1)
18:15
produce (1)
153:5
produced (2)
113:7;117:20
product (3)
13:19;14:11,22
products (1)
88:18
programing (2)
29:9;31:15
project (3)
13:9;25:16;29:11
pronounce (1)
82:20
proof (1)
10:2
properly (3)
5:6;64:4,6
properties (3)
86:21;87:3,8
property (5)
27:3,5;86:18;87:1;
99:10
proportion (3)
116:3,4;117:1
proposed (2)
13:4,15
proposes (1)
50:2
protocols (1)
34:10
provide (11)
14:23;18:5,5;28:17;
42:17;53:4,6;65:18;
128:6,10;150:13
provided (20)
12:24;14:9;16:23,
24;18:4,8;39:14;42:4,
5,16,24;44:16,18;
78:25;110:21;127:22;
131:21;150:1,6,18
provides (12)
18:9;45:23,24,24;
57:11;59:24;88:25;
93:15,16;109:6;
122:18;132:2
pseudo (1)
72:5
publications (4)
30:16,17,20,25
pull (4)
62:20;68:21;
117:16;118:10
pulled (1)
120:18
pulling (1)
126:14
punctuation (2)
74:20;75:14
pure (1)
72:9
purple (1)
40:3
purports (1)
7:18
purpose (9)
6:21;7:9;65:20,21;
69:16;72:7;73:1;
88:15;138:22
purposes (13)
6:13;11:1;40:13,16;
48:20;71:14;74:4;
75:5;85:3;86:7;102:7;
130:19;131:15
pursuant (1)
4:14
pursued (1)
148:24
pursuit (1)
149:12
put (20)
6:2;9:9;10:3,4,5;
13:23;15:3,24;23:17,
23;58:11;69:8;102:1;
105:15;122:19;
129:11;130:21;
136:20;137:18;
144:22
putting (1)
153:8
puzzle (3)
79:24;81:13,23
puzzles (1)
80:12

Q

qualifications (1)
32:10
qualified (1)
18:15
quality (51)
17:7;18:16;22:2;
33:9,12;35:18,22;
36:19,25;50:24;51:2,
8,12,14;52:23,25;
53:13;54:7,15;57:25;
58:1,4;59:3,9,10,15,
16;60:12;64:18,21,
23;65:8,10;71:13;
73:2;74:6,17;75:23;
77:3,4,5;113:17;
117:10,25;123:4;
125:11;127:6,8,10,16,
17
Qualtrics (1)
69:6
quantify (1)
23:22
quantitative (2)
112:7;126:19
quantities (2)
44:22;115:18
quantity (3)
39:19,20;69:23
questionable (1)
126:25
questionably (1)
131:9
quibbles (1)
21:16
quick (3)
77:1;113:20;137:6
quickly (7)
15:9;41:16;56:5;
70:7;76:19,21;131:5
quite (39)
26:2,9;27:6;47:6,
24;49:10;50:15;51:8;
54:7;55:19;62:18;
63:12;64:11,12,17;
66:6;78:19,21;80:18,
24;83:5;84:19;88:17;
90:19;106:16,17;
107:24;108:25;
109:21;111:20;113:1;
117:5;118:24;121:8,
25;125:10;127:4;
128:2;152:4
quote (6)
88:12,14,21;
115:12;117:8;132:1

R

raise (1)
24:20
raised (4)
19:20;128:25;
129:14;130:17
ran (2)
18:9;54:20
random (27)
29:19;55:4,8,14;
60:13,15,18,19,20,21;
62:9,25;63:9;64:15;
66:25;67:17;68:7;
71:6;72:3,5,6,11,12;
109:3;118:14;127:1;
133:15
randomly (16)
53:25;60:5,24;61:8;
66:4;70:17;118:10,
11,13,17;120:5,18,23;
121:1;123:14;127:4
randomness (4)
29:16;67:3;68:8;
72:9
range (6)
25:25;30:21;58:10;
82:25;106:16;111:24
rapid (1)
79:15
rare (1)
94:21
rate (25)
33:5;50:8;112:23;
113:3,19;115:9,11,13,
13,19,20,21;116:23,
25;117:3,18;122:7;
126:16,21,25;128:13;
141:19;143:24;
144:19;145:1
rates (6)
50:10;115:21;
116:18;117:22;122:3;
127:3
rather (7)
53:6;77:9;79:15;
97:3;105:21,21;
109:19
raw (2)
37:1;119:7
read (3)
52:20;61:4;129:18
reading (1)
129:21
ready (3)
24:15;25:5;41:23
real (6)
29:2;83:23;86:19;
100:16;109:7;146:22
realistic (2)
104:19;150:12
realize (1)
10:4
really (19)
6:20;9:5,12,25;
12:22;14:4,8;16:20;
22:24,25;23:5;56:23;
72:7;91:7;92:11;
101:3;131:5,7;133:1
realm (2)
81:14,18
reason (8)
12:18;59:8,22;
86:10;88:25;115:23;
116:2;154:9
reasonable (5)
23:9,10,10;55:1;
142:20
reasonableness (1)
20:7
reasonably (2)
14:23;59:15
reasons (5)
19:24;59:2;85:14;
105:19;123:22
recall (13)
23:2;114:10;
115:19,25;116:4,14,
15;117:3;135:11;
138:21,24,25;152:13
receive (1)
153:2
received (2)
13:6;14:1
receiving (1)
152:24
recent (2)
28:10;149:6
recently (1)
49:10
recess (2)
137:6;156:10
recipe (1)
93:10
recognize (1)
153:22
recognizes (1)
88:2
recommended (2)
74:4,11
reconnect (1)
51:22
reconsider (1)
16:18
record (62)
4:20;11:2;20:22;
24:24;32:7;33:2,24;
34:1,9,13,13,17,18,
19;35:3,8,16,19;37:2,
11,18,24;40:18,20;
41:6,17,18;44:24;
45:14,16,20;48:5;
51:5;55:23;57:6,9;
59:18;69:9;73:2;74:5,
6;76:8,8;77:10;79:20;
84:1,15;86:13,15;
88:9;94:22;95:17;
96:23,25;101:6,7;
117:9;119:15;121:17;
127:11;137:10,11
records (84)
21:21;34:3,3,7,10;
36:22,22;37:5,7;39:8;
42:5;43:1,16,24;44:3;
46:3,4,8;47:8;48:5,8;
49:5,11;53:10;58:16,
20;60:10;61:9,10;
62:7;71:1,3,4,5,6,10,
16,25;72:15;73:15;
75:24;76:2,3,5,6,10;
77:2,8,20;79:16;80:5;
83:24;84:3;85:8,16,

22;86:18,23,25;87:13,17;89:15;93:5,8;94:7,18,24;96:1;97:4,7,19;98:1;100:19;102:4;104:16,17;110:15;113:10,11,15;114:3;124:5;125:18,22
recover (1)
156:8
recruit (1)
29:6
recruiting (1)
9:1
red (7)
59:1;67:12,15,18;68:16,17;98:11
redacted (2)
8:3,8
reduce (7)
41:1;43:17;76:17,25;77:7;78:3;79:18
reduced (1)
89:4
reducing (4)
37:3,17;75:23;77:8
redundancy (1)
59:20
redundant (2)
59:16;73:19
refer (10)
8:8;49:6,9;57:9;86:23;140:17,20;145:9;149:3;155:18
reference (2)
28:12;49:15
referred (5)
26:17;42:8;44:20;45:23;66:13
Referring (7)
31:13;38:25;42:12;55:22;129:25;139:9;154:1
refers (4)
34:12;48:17;74:22;149:1
reflect (1)
22:9
regards (1)
5:25
regions (1)
108:15
relate (2)
53:3;71:15
related (7)
4:9;11:21;29:10;44:1;59:4,6;104:3
relates (3)
14:4;29:22;107:25
relative (8)
35:5,10;47:7;48:1;89:22;99:12;100:2;101:22
relatively (3)
55:9;63:10,19
relevance (1)
130:16
relevant (17)
14:6;19:14;31:1,2,17;33:1,13;44:5;52:22,23,24;53:12;110:15;134:17,25;136:11,16
reliability (3)
20:6;21:15;49:1
reliable (15)
17:11;18:18;21:20;22:3,5,5,6,6,10;33:2,17;35:9;49:13;50:6;142:20
reliably (1)
24:9
relied (1)
7:7
rely (2)
7:2,22
relying (1)
9:7
remainder (2)
42:18;129:10
remained (1)
47:23
remaining (2)
44:9;115:4
remember (3)
41:21;135:16,20
remind (1)
118:12
Remine (4)
28:12;29:2,2;86:19
remotely (2)
84:16,20
remove (1)
74:24
removed (1)
123:2
removing (3)
59:20;74:17;77:9
renter (1)
86:22
rep (14)
14:5,15;44:17,24;54:13;55:16,17;57:6;62:9;119:23;121:10,12;122:21;155:21
repeatedly (1)
22:20
rephrase (1)
147:15
reply (4)
60:4;63:24;72:17;127:12
report (28)
13:21;15:1;21:8;23:13;60:4;72:17;84:2;85:9,11;123:16;127:12,13;134:1;136:21;137:19,21,23;139:24,25;140:3,18;144:22,25;145:9;150:4,4;152:21;153:13
reports (2)
128:22;129:5
reposition (1)
132:9
represent (4)
4:22;5:8;57:5;103:5
Representante (8)
42:7,19,21,24;43:3,18,25;50:25
representative (7)
63:21;67:25;68:11;118:19,20;123:18,22
representatives (3)
4:6,12,19
represented (1)
95:11
representing (2)
58:23;98:1
represents (2)
52:15;108:20
reproducible (5)
60:20;72:4,11;118:13,16
reps (1)
122:22
request (2)
39:10;110:19
requesting (1)
9:23
require (1)
35:24
required (3)
54:1;97:23;98:25
rereview (1)
146:20
reschedule (1)
4:15
research (16)
17:15;25:15,20,22,24;26:1,2;28:21,22,24;29:9,12,13,15;147:5,18
reserved (1)
18:13
resides (1)
143:11
resolution (2)
18:19;132:4
resource (1)
132:2
respect (18)
7:21,23,25;8:5;14:2;18:10,13;20:10;26:23;28:18;42:2;56:18;65:19;66:21;88:19;89:21;102:2;152:17
respectively (3)
42:8;115:15,25
respects (1)
122:10
respond (1)
15:8
responsible (1)
148:23
rest (1)
39:5
restricting (2)
64:23;118:22
result (20)
17:11;18:10,17;21:17;36:5;64:18;65:9;76:22;81:4,13;82:19;83:12,15;86:17;87:21;91:11;99:9;107:8;109:15;111:25
resulted (2)
111:12;120:3
results (15)
14:23,23;17:10;18:4;33:5;35:22;39:25;40:2;59:19;104:3;109:4;111:13;118:3;144:23;147:1
RESUMED (2)
41:25;137:16
reverse (2)
87:4;109:20
review (7)
7:1,1;36:5;64:18;129:5,6,14
reviewed (1)
7:4
Reynolds (8)
10:13,25;11:13;24:20;41:15;137:1,3,3
RFC (2)
55:21,24
rid (2)
73:19;74:12
right (123)
4:21;6:11,22;9:11,17;10:13;16:18,21;17:1;24:20;25:6;28:7,11,17;29:14;30:19,24;31:4,21,24;32:12,14;33:19;36:8;38:4;39:6,25;40:11,23;41:9;43:11,15;47:12;49:7;50:17;51:25;52:19;54:17;55:8;60:9,23;61:10,20;62:3;65:3;66:22;68:3;70:19;72:20;73:16,17;74:8;77:22;81:5;83:18;84:9;85:9,12;86:2;88:10;89:3,8;92:16,23;94:7,8;95:23;96:7;97:9,10;98:20;99:21;101:11;102:6;103:2;104:3;107:12;109:13;110:17;114:16;115:9;118:24;119:19;120:17,19;122:3,8,25;123:5,11;124:10;128:9,21;129:9;131:1,25;132:13;135:3;136:18;137:9,13,19,24;138:19;139:4,6,12,17,24,25;140:19;144:22;145:12,16,16;147:6,11;151:9;153:4,8,25;155:15;156:6
right-hand (2)
95:8;124:8
rise (1)
101:5
Rita (1)
152:23
Road (1)
143:11
robin (2)
76:9;94:5
Robn (1)
94:8
Robyn (1)
94:9
Rodriguez (1)
120:9
role (13)
22:23;25:14;26:19,20;44:22;85:17,17,20;87:25;89:17,20,23;90:3
Rona (49)
5:1,1;6:5,6,12,15,18,20,23;8:11,14;11:15,24;12:2,25;15:4,8;19:2,5,9;24:14;26:16;32:11;48:15;129:4,25;131:5;132:8,11,23,25;133:4,7,10,12;136:20,22,23;137:1,5,8,12,14,17;147:16,17;153:14;156:2,7
Rona's (4)
5:10,14;26:5;35:23
roughly (5)
37:3;40:3;45:9;54:4;116:2
round (1)
76:9
row (20)
82:17,19;83:9;90:7;91:11;93:18;107:7;124:18;125:22;140:14,16,24;141:12,14,22;143:4,7;151:5,

12,21
rows (15)
43:6,7;64:1;83:3;
124:12,18;126:4;
142:1;146:13;151:3;
152:8,18,19,20;
155:21
rubric (4)
90:4;91:12;121:7;
146:17
rule (10)
9:17;77:15,19;
80:17;84:7,21;85:5;
87:16;88:2;110:5
Rules (9)
4:11,14;81:11,20;
84:25;85:10,12,16;
87:17
ruling (2)
9:11;16:18
run (9)
33:4;39:23;55:20;
56:5;69:5;72:13;80:7;
89:9;135:8
running (2)
81:11;84:1
runs (2)
95:17,20

S

same (88)
8:25;9:21;10:1;
15:14;22:25;23:1,18;
30:7;34:5,8;38:25;
43:24;48:12,14;49:6;
53:7;55:15;56:17;
58:16,17,17,20,21,22;
60:21;61:18;62:7;
72:15;73:10;74:22;
78:14;81:24;82:10,
13;83:2;84:13,13,18;
85:4;86:1,5,21,23,25;
87:1,3,5;90:13;92:5;
93:22;94:4;97:5;
98:19,19,24,25;102:4;
103:9,15,18,23;
105:16;110:6,8;
114:9;115:23;116:2;
119:22,24;120:1;
122:20;123:1;124:22;
125:7,15,17;126:12;
141:17;142:16,23,24;
143:10;146:7;150:8;
151:17,23,25;154:14
sample (35)
30:14,14;54:3;55:5,
14;60:13,15,21;
63:19;65:14,25;66:4,
5,16,21,23;67:7,15,
18,24;68:2,10,12,15,
21,23;69:2,6,12,18,
24;70:2,14,21;120:21
sampled (6)
30:1;60:5,24;61:8;
62:10;66:24
samplers (1)
29:17
samples (7)
53:2;62:25;70:8,15;
85:15;122:5;127:2
sampling (11)
55:5;59:25;60:20;
65:19,22;66:9,11,13;
67:17;68:6;109:3
sanity (1)
56:12
Santana (1)
126:7
Santos (1)
126:5
sat (1)
133:1
Saul (2)
155:21,22
saw (5)
70:18;85:4;133:15;
134:2;152:14
saying (16)
11:6;26:25;40:4;
41:6;76:16,19,21;
88:5,5;93:8;122:2;
127:8;135:16;138:21,
25;152:7
scale (1)
102:2
scenario (4)
103:13,20;106:2;
107:13
schedule (1)
10:8
scheme (6)
32:24;43:22;44:1,
11;111:22;112:6
science (26)
17:5;18:1,3;25:24,
25;28:21,21;29:7;
30:22;33:25;47:25;
59:18;72:4;73:18;
74:1,4;75:17;84:25;
88:1,12;107:17,21;
117:10;128:1,2,19
Sciences (2)
25:16,23
scientific (1)
28:22
scientist (8)
17:5,15,18;18:24;
25:15,20;29:2,12
scientists (1)
34:10
scope (6)
32:22;136:18;
137:18,24;138:10,16
score (6)
82:11;95:1,5,6;
117:6,12
scores (2)
93:21;113:10
screen (3)
5:18;136:21;140:9
scroll (3)
13:11;62:21,24
seal (1)
5:19
seated (4)
4:2;24:23;101:8;
137:12
second (31)
10:24;16:14,16;
20:7,17;21:8;27:17;
29:14;32:20,21;
41:10,11;59:8;70:24;
76:11,11;78:1;82:19,
20,21;84:7;88:14;
90:17;116:7;120:17,
19;140:24;145:5,11;
148:21;151:16
secondary (3)
125:20;148:19,20
seconds (2)
39:23;112:16
section (2)
106:20;145:11
Security (4)
75:10,12;78:9;
119:23
seed (4)
72:11,12;118:14,16
seeing (3)
120:23;127:4;
144:12
seeks (1)
50:4
seem (9)
61:17;73:7;117:5;
126:24;127:5;144:13,
14;147:9;154:4
seemed (6)
58:14;63:2;127:4;
139:2;148:1;152:10
seems (15)
9:4;58:22;64:16;
92:17;119:18,21;
121:12;123:4;126:10,
18;134:6,11,17;143:4,
12
seldom (1)
75:14
select (3)
54:3;68:9;86:14
selected (9)
53:25;118:11,13;
120:5;121:1;123:14,
21;127:5,21
selecting (2)
60:18;68:9
selection (4)
66:23;67:4,6,21
semi-check (1)
54:9
sense (31)
13:11;25:19;33:24;
41:4;49:24;54:10;
60:17;61:13,16,22;
64:10;76:1;84:9;
88:11;94:20;120:25;
124:17;126:19;131:2;
139:21;141:21;
142:20,22;144:11;
146:22;147:2,21;
148:25;149:9,25;
155:14
sent (1)
131:24
separate (8)
19:23;22:3;24:2;
43:8;103:24;104:1,
18;151:22
separated (3)
20:2;22:13;114:23
separately (1)
91:3
September (1)
13:4
sequence (2)
119:20;125:23
sequencing (1)
36:14
series (3)
18:5;92:13;156:5
served (2)
106:20,22
server (1)
59:5
service (1)
64:11
services (1)
32:23
set (47)
13:12;24:4,4;27:19;
34:21;35:25;38:16,
17;40:5;43:9,17;47:6;
52:10;62:8;65:15;
68:22,23;72:14;76:6;
78:11;85:2;87:18;
89:5;93:15;96:25;
98:8,18,21;100:16;
104:14,19;105:13,24;
106:9,10,19;108:3;
109:2,18;110:14,18;
123:14;144:16;
149:17;150:2,12;
153:5
setback (1)
156:9
sets (11)
26:6,7,8;35:24;
48:16;101:1,18;
102:3;106:15,24;
110:3
set-wide (1)
68:23
seven (13)
38:6;57:21;74:21;
75:3,6,7;84:3,9;96:9,
20;106:13,24;141:23
seventeen (2)
43:7,15
seventy (1)
124:18
seventy-eight (2)
125:18;126:4
seventy-five (2)
90:23;145:15
seventy-two (1)
56:13
several (4)
19:24;46:25;80:4;
120:22
sex (1)
84:3
shade (2)
80:16,17
shades (1)
80:18
shape (3)
8:9;79:24;80:10
shapes (1)
79:25
share (1)
86:1
Shaw (2)
84:10;92:4
short (2)
14:2;92:11
shortly (1)
49:4
shot (1)
81:25
show (6)
12:14;89:25;
103:19;105:4;107:8;
135:7
showed (2)
134:20;146:25
showing (5)
68:24;81:5;105:2;
111:11;154:4
shown (6)
5:19;98:3;104:15;
107:12;142:3;143:21
shows (2)
62:6;95:8
side (4)
62:21,21;79:14;
95:8
sides (3)
12:8,9;105:2
sig (1)
20:6
significant (1)
73:6
silly (1)
88:6

similar (11)
23:16,16,17;38:21;
48:15;87:13;88:8,18,
24;121:8;122:9
similarity (1)
34:4
Similarly (7)
22:19;24:1;39:4;
72:14;75:10;84:12;
126:12
simple (2)
17:4;115:18
simpler (1)
35:16
simplicity (1)
90:13
simplified (1)
36:11
simply (26)
8:20;13:10,17;
14:12;20:16;26:24;
39:14,21;42:25;45:1;
74:17;86:13;91:11;
98:17;110:1,6,21,24;
111:1;112:20;120:12;
130:21,25;138:22;
150:10;151:16
single (12)
39:3;67:7,23;90:5,
7;108:12;109:16,18;
110:14,22;111:11;
144:1
sit (3)
133:4,4,8
situation (6)
7:6;73:6;87:4;
88:23;107:11;114:2
six (12)
57:21;75:2;76:10,
14;80:22;91:14;
92:24;96:9;101:15;
114:10;134:5,21
sixty (2)
98:5;135:9
sixty-one (1)
143:5
size (30)
8:9;30:14;43:17;
60:15;63:19;66:5,16;
68:10,12,15,25;69:6,
12,18;70:2,2,13,14,
21;103:15,23;118:11,
18,20,23,25;119:1;
122:5;134:18;149:24
skipped (1)
57:15
slice (1)
118:21
slices (6)
12:12;15:23;23:12,
13,14;118:22
slicing (1)
118:18
slide (57)
32:21;33:23;36:8;
41:24;43:21;45:10;
46:11;47:14;50:17,
23;52:14,14;62:6,8;
65:16;66:18;68:3,18;
72:21;74:8;75:19;
76:23;79:13;81:2,21;
83:19;85:5,19;87:23;
89:7;91:17,23;92:2;
93:24;96:5,13;99:20;
101:4,15;102:10;
103:1,10,11;104:8;
107:8;108:4,9,15;
111:6;113:22;114:5,
6;119:3;124:4;134:2;
135:4,8
slides (3)
92:14;99:12;132:2
slight (2)
121:2;126:12
slightly (3)
73:11;135:1;145:13
small (11)
51:15;55:9,19;
63:19;64:3;65:3;
67:11;69:17,24;70:2,
22
smaller (2)
65:24;102:21
smell (3)
56:15;111:21;
112:21
Smith (1)
147:23
so-called (1)
26:7
Social (4)
75:10,12;78:9;
119:23
software (6)
73:12;94:4,5;117:9,
12;129:20
solution (2)
87:11,12
somebody (8)
5:17;10:18;48:23;
74:14;86:25;123:9;
148:11,13
somehow (2)
45:21;46:17
someone (1)
150:8
someone's (1)
155:2
sometimes (13)
19:19;48:18;59:19;
61:24;69:13;79:20;
86:1;99:16;103:16;
105:5,11;114:7;
134:16
somewhat (6)
26:20;33:16;49:6;
88:2;90:10;97:17
somewhere (1)
152:17
sophisticated (2)
49:3;53:20
sophistication (1)
47:21
Sorondo (3)
4:13;6:1;16:13
sorry (7)
11:6,6;31:7;44:12;
108:10;127:7;153:12
sort (32)
6:1,7;9:13;12:21;
22:21;25:17,18;
28:18;33:23;34:25;
36:11;45:4;49:21;
51:16;59:5,17,20;
63:9;66:16;69:20;
85:1;93:10,16,24;
100:23;102:12;106:5;
112:5;116:16;117:23;
124:1;131:20
sorted (7)
62:9,10,10,12,14;
64:3,9
sorting (1)
63:14
sorts (4)
33:15;49:7;59:6;
123:19
sound (1)
113:4
sounds (4)
86:10;88:6;100:13;
156:5
sourced (1)
131:9
space (4)
73:20;74:15;83:7;
122:21
spaces (8)
56:3;73:19,22,23;
74:12,12,14;121:6
speak (3)
35:12;57:7;99:25
speaker (2)
31:20;52:9
speaking (11)
7:18;8:23;34:12;
36:17;40:11;88:15;
92:12;101:25;116:3;
125:3;130:6
special (4)
22:23,23;23:5;
85:20
specific (15)
12:23;50:13,14;
69:13;85:10,12,13,17;
87:16;108:22;130:6;
144:20;147:9;149:18;
155:4
specifically (12)
5:24;17:5;23:15;
24:8;93:13;134:7;
135:15;138:2;146:17;
148:9;149:11;152:7
specify (1)
69:19
speculation (1)
147:14
spell (1)
86:11
spelled (1)
126:11
spend (1)
50:21
split (1)
114:21
spot (1)
38:8
square (3)
40:22;67:15,23
squares (3)
67:13,13;68:16
Sr (2)
84:10;92:4
stage (1)
56:23
stand (7)
9:10;16:1;20:22;
24:18,19;101:10;
133:5
standard (43)
37:2;40:14;45:15;
47:24;48:10;49:23;
55:24;56:6,10;59:17;
66:9;69:4;72:12;
73:17,18,25;74:3,13;
77:5;78:3;79:19,22;
80:21;82:16;83:2;
86:12;87:12,12,15;
88:2,4;90:17,19;
96:14,17;97:16;
106:2;107:21,22;
108:24;116:9;118:15;
128:2
standardized (1)
75:16
standards (5)
30:21;55:20;75:4,
17;84:25
standing (5)
5:17;19:6;24:20;
103:6;132:13
standpoint (1)
40:18
start (13)
13:25;32:14,21;
36:15;45:6;48:20;
50:22;51:1,3;62:24;
72:23;113:22;149:14
started (2)
65:6;127:3
starting (10)
13:4;29:3;31:22;
51:2;57:12,15;62:14;
76:15;118:15;149:12
startup (2)
29:13;86:20
state (8)
20:22;24:24;59:13;
86:11;91:4;129:11;
143:19;150:17
stated (6)
6:7;58:24;70:9;
71:21;88:10;152:12
statement (2)
17:2;26:17
statements (2)
7:22;8:15
states (1)
29:15
static (1)
69:10
statistical (8)
30:4;47:21;59:25;
65:19,22;66:9,10;
69:5
statistically (13)
30:14;53:2,25;54:3,
11;55:5;63:18;65:14;
66:21;68:5,22;85:15;
120:21
statistics (5)
29:22;30:23;31:14;
66:12,13
stay (5)
10:10;24:20;41:13;
133:3;140:13
step (55)
16:14,16;18:2;23:9;
27:8;36:21;37:14,18;
39:9,24;44:22;46:8;
72:24;77:10,11,13;
79:5,9,14;81:7;85:18;
87:14;89:4,9;90:17;
91:1,14;92:6,23,24;
93:4,4,21;96:9;
101:15;106:13,13,24;
109:14,20,20,22;
110:11,18;111:12,13;
112:9,10,10,11,14;
127:24;146:18;
150:16,19
Stephen (2)
4:13;138:5
steps (21)
17:22;36:20,23,24;
38:5,6;50:22;55:17;
74:9;75:21;76:24;
92:13,13;96:7,9,10,
20;112:18;127:24,25,
25
stick (2)
23:20;156:3
still (43)
6:12;23:5;24:1;
35:21;37:25;38:25;

39:2,23;45:15,19;
46:7,8;49:12;54:11;
55:19;61:4;62:12;
70:5,15;76:4;77:12;
79:1;80:13,24;81:14;
83:5;87:6;89:12,18;
96:2;97:22;101:10;
104:22;105:25;
109:18;115:5;119:1;
121:6;125:7,11,18;
139:21;153:13
stipulation (2)
20:1,24
stop (13)
38:11;43:11;48:14;
55:23;69:9;91:16;
92:21;99:14,15;
118:4;120:4,4;126:2
straightforward (2)
109:21;122:4
stratifying (1)
111:14
street (14)
46:24;59:11;61:19,
19,20;77:24;84:3;
85:6;89:13;90:15;
92:3;141:14;142:4;
143:9
strike (6)
4:10,13;6:16;8:4;
9:3;45:6
string (4)
62:11;82:7,13;
89:12
string-matching (3)
82:16,23;85:4
strings (4)
64:15;82:17;83:10,
14
strong (1)
124:22
strongly (1)
35:7
structure (2)
65:12;69:14
structures (1)
29:19
students (2)
26:3,4
study (1)
117:8
stuff (2)
53:24;97:11
Sub– (1)
153:9
subdiscipline (1)
66:12
Sub–Exhibit (1)
153:13
subject (1)
129:14
submit (2)
18:21;147:3
submitted (2)
130:1,22
submitting (1)
13:4
suboptimal (1)
113:15
subset (1)
68:13
subsidiary (1)
151:19
substantial (2)
13:7;59:19
substantially (3)
46:24;59:9;86:3
subsumed (1)
59:10
subtler (1)
67:2
subtracting (1)
95:6
sued (1)
148:14
sufficient (15)
17:7;18:16;33:14;
64:20,25;68:22;
70:11;71:22;75:6;
127:15,17,18,22;
128:4;150:13
sufficiently (3)
17:11;18:18;117:10
suggest (2)
7:18;70:20
suggested (2)
26:16;128:3
suggests (4)
141:15,19;143:25;
149:19
suing (2)
148:17,20
suitable (4)
51:17;58:1,14;68:9
suited (1)
18:22
sum (6)
18:9;26:7,23;105:1;
111:2;155:6
summarize (3)
124:10,12;144:25
summarized (2)
40:2;103:8
summary (7)
4:19;44:18;111:7;
124:5;125:5;129:13;
134:2
summons (3)
152:24;153:2,6
supplemental (1)
12:6
support (4)
67:5;130:23;131:2;
147:14
supposed (3)
15:17;20:24;58:24
sure (24)
11:22;27:22;32:8;
36:17;52:16;70:25;
72:21;73:1;91:19;
122:9;123:6;124:8;
131:7;137:7;140:5,8;
146:2,7,10,24;147:16;
153:1;154:7,21
surely (1)
143:24
surname (2)
94:16;99:22
surprising (2)
94:18;95:4
surprisingly (1)
94:12
survey (1)
69:6
surveys (2)
66:10;83:25
suspected (1)
144:9
sustain (1)
9:8
swear (1)
24:21
switch (1)
137:2
Swoosh (1)
132:3
SWORN (1)
24:22
symbol (1)
56:2
symbols (2)
53:22;57:1
synthesis (1)
13:18
synthesize (1)
14:10
systematic (10)
45:25;53:1,17;54:2;
55:20;56:20;60:11,
12;72:10;106:19
systematically (3)
91:9;125:13;136:11

T

tab (1)
130:10
table (28)
5:10,12,14;13:13;
18:8,12;42:8,8,9,9,10,
19,21,24;43:3,11,16,
18,25;44:17,24;
53:10;109:22;112:11;
129:12,13,13;134:1
tabled (1)
139:25
tables (14)
13:7,10,13,17;
14:12;15:14;42:6,13;
43:6;44:15;45:1;
118:5;136:10,15
tabulated (2)
39:15;110:21
tabulates (1)
47:4
tabulation (1)
39:23
talk (13)
14:11;15:6;20:23;
25:17;29:19;34:14;
61:11;68:20;75:18;
81:20;124:7;129:22;
136:18
talked (7)
15:12;45:3;52:13;
65:14;71:9;102:7;
152:4
talking (17)
19:23;40:22;50:21;
61:13;67:14;68:12;
71:7;72:23;88:22;
99:20,25;102:6;
125:9;130:6,15;
135:20;149:23
taller (2)
15:19;41:7
tallest (1)
41:8
task (7)
19:14;21:9;45:21;
55:25;79:24;80:8;
96:22
tasks (1)
41:2
teaching (3)
26:4;28:25;31:21
team (5)
8:15,16,24,25;29:7
technical (6)
88:11;108:22;
116:11;137:15;
142:19;156:9
technique (11)
37:19,23;47:24;
72:12;73:17,19;
76:19;77:5;78:3;
79:19;96:18
techniques (11)
26:12;37:2,25;
40:21;48:23;72:8;
73:4;76:17;80:6;85:2;
128:2
Technology (1)
25:13
telephone (3)
46:24;47:5;77:24
Telex (1)
7:19
TelexFree (18)
15:19;17:7;21:25;
22:22;32:24;33:9;
40:11;42:3;43:22;
64:19;105:12;128:17;
130:2;135:18,21,25;
136:5;138:6
telling (2)
86:15;94:1
tells (2)
54:2;56:8
ten (25)
44:22;70:3,13,14;
111:13;112:9,11,14;
118:11,18,20;119:1,
11;122:6;123:14;
133:14;134:13,20,21;
135:4;140:14,16,24;
141:14;149:24
tend (5)
58:17;61:17;94:22;
100:19;111:24
tender (1)
32:6
tendering (1)
32:9
tends (2)
58:21;101:1
tens (1)
40:24
tension (4)
97:8;104:14,23;
116:2
ten-step (1)
23:7
tenth (2)
122:13;124:1
Teresa (10)
140:18;141:12;
142:8,8,14,15;143:10;
145:4,20;146:8
term (5)
35:23;48:12,15,15;
116:11
terminology (1)
79:21
terms (19)
28:11;36:15;37:16;
52:21;57:7;60:7;71:1;
88:24;107:1,2,23;
108:3;110:23;111:17;
113:5;115:15;134:25;
149:25;154:8
test (8)
8:2;48:24;49:22;
53:17;54:2;56:15;
111:22;112:22
tested (2)
22:24;49:2
testify (2)
7:25;8:7
testifying (4)
7:20;9:19;135:12;
138:24
testimony (10)
4:6,9,15;5:24;7:10,
17;9:24;11:21;

135:10;148:13
testing (4)
29:18,19;48:19;
50:4
that'd (1)
52:15
theoretical (2)
26:1;30:8
theory (2)
107:23;108:24
thereabouts (1)
12:5
therefore (6)
8:19,19;24:5;75:5;
87:19;129:1
thinking (1)
92:10
third (10)
59:22;76:11,12,13;
83:4;98:13;122:8,9;
125:16;134:2
thirties (1)
83:15
thirtieth (2)
124:2;125:25
thirty- (1)
90:15
thirty-one (1)
101:15
thirty-six (1)
13:24
though (18)
46:25;55:7;58:18;
65:4;67:23;75:11;
90:12;112:4;113:15;
119:17,24;121:20;
122:20;124:5,21;
126:11;134:24;141:6
thought (10)
6:6;16:11;23:2;
71:5;127:7;130:5,5,6;
137:1;148:20
thousand (1)
70:13
three (29)
17:10;28:18;31:2;
42:13;54:17;59:2;
70:8,11;73:22;74:21;
84:7;90:20;103:22;
109:20,23,25;110:2,
11;111:23;114:3,11,
11,12,24,24;115:4;
125:3;129:11;136:15
threshold (33)
96:16;97:15,19,20;
98:3,10,12;99:3,3,6;
102:8;103:5,7,9,14;
104:1,2,3;105:6,22,
23,23;106:4,7,16,21,
23;107:14;108:13,14;
109:12;125:10,13
thresholding (3)
102:6;108:4;114:19
thresholds (6)
85:1,13;105:2;
106:9,16,25
throughout (6)
60:11;66:10,17;
72:16,16;125:4
throwing (1)
78:6
thumb (1)
130:1
tie (1)
133:6
tied (1)
21:6
tier (2)
111:7,14
tight (2)
105:21,21
tighter (1)
107:3
tightly (5)
105:13,14;106:6,
11,11
Tim (1)
23:2
time- (1)
37:11
times (2)
131:9;148:10
tiny (1)
73:9
today (13)
6:13,17;9:22;10:16;
16:17;19:24;32:16;
45:15;81:18;134:21;
135:13,14;147:1
together (19)
23:11,19,20,24;
34:7;61:17,22;71:15,
24;77:9;78:11;81:23,
24;97:11,13,20;
104:5;119:9;121:10
tokens (2)
63:7;140:24
told (3)
44:8;137:3;149:14
tolerate (1)
125:10
tomorrow (3)
10:16,21;156:10
tons (2)
53:7,7
took (10)
16:13;17:22;18:1;
39:10,13;44:2;55:5;
74:9;110:19;133:15
tool (4)
69:7,16;73:25;
107:17
tools (1)
18:3
top (10)
63:25;64:3;88:13;
90:6;102:20;115:5,6;
124:18;133:6;154:4
Toronto (1)
28:23
torsos (1)
132:21
total (22)
26:8,24;39:18;44:3;
55:9;63:7;67:16,19;
68:17,25;70:3,4;77:8;
81:19;82:12,14;
83:17;91:5,10;96:3;
111:1;124:12
totally (7)
37:10;39:1;73:21;
74:3;78:2;105:1;
119:7
totals (1)
112:17
touched (1)
131:6
tournament (1)
76:9
towards (1)
16:17
track (1)
91:3
tradeoff (2)
105:3;109:6
trailing (1)
73:23
training (2)
26:10;36:4
transactions (4)
14:5,15;20:11;
42:13
transcript (1)
99:12
transfer (1)
42:9
Transferencia (1)
42:7
transfers (1)
20:10
transformation (1)
93:17
transformations (1)
74:2
transitive (12)
23:15;26:23,24;
27:1,9;38:9;96:16;
97:22;98:16;99:2,9;
108:4
transitively (1)
27:2
Transparent (1)
129:19
Transportable (1)
129:19
transposition (1)
47:1
treat (1)
91:10
treated (3)
91:5;101:24;102:4
treatment (1)
20:11
trial (3)
13:24,25;16:25
triangular (1)
20:11
trillion (3)
81:5,9,19
trillions (3)
40:24;76:16;80:5
trim (1)
73:23
troublesome (1)
127:2
true (14)
24:6;26:7,18,20;
56:17;100:21;113:25;
114:5,11,21,24;115:5,
16,16
truncated (1)
46:12
truncating (1)
78:22
truncation (3)
79:4;121:14,15
trustee (22)
4:23,25;5:10,12;
6:14;9:4;10:1,4;
11:19;12:9;16:25;
20:17,20;24:17;
39:10,14;40:1;42:4;
138:5;149:4,14;153:6
trustee's (12)
12:11,20;32:23;
43:19;69:11;110:19;
118:6;129:7,8;
131:17;132:5;149:12
truth (9)
21:21;26:6,8;35:24,
25;48:16,17;49:24;
114:1
try (13)
7:5;22:3;69:7;
104:24;105:2;106:3;
112:22;115:23;132:9;
140:14;153:8,10;
155:25
trying (13)
6:21;22:12;46:7;
67:14;69:24;72:8;
80:9;86:20;87:7,8;
88:16,17;141:10
TTMilagros@gmailcom (1)
143:11
turn (9)
29:14;30:24;46:11;
52:13;62:3;86:14;
91:13;111:6;122:8
Turning (3)
21:14;23:7;90:1
turns (1)
60:14
tutorial (3)
65:18;84:2;113:20
twelve (6)
60:5,24;61:12;
71:17,24;76:6
twenties (1)
134:23
twentieth (4)
124:2;125:15,17;
126:3
twenty (5)
58:9;119:1;134:8,9;
135:9
twenty-one (1)
84:6
twenty-seven (1)
13:13
two (89)
9:23;13:23;14:5;
16:9;17:8;19:11;20:2,
14;21:22;23:23;28:8;
30:6;34:1;37:5,7;
39:1,1;40:9;44:3;
45:13;50:7;53:15;
63:7,7;68:7;70:1;
73:11;75:22;76:17,
24;77:11;78:6,14,21;
82:8,9,19;84:12,17,
22;86:20,25;87:4,17;
90:2,16;92:3,13;97:7;
98:10;102:20,22;
103:25;104:4;110:6,
8;113:24;115:2,4,11,
12,17;116:1,9,23;
118:3;120:10;122:5,
6;123:11;125:3,22,
25;126:21,25;128:11;
137:23,25;138:1;
140:24;142:11,15;
143:15;146:15;150:9;
151:10,11;154:13;
155:21
type (3)
114:7;115:2;147:19
typed (1)
74:16
types (4)
21:18;25:19;34:9;
41:2
typical (4)
64:17;94:12,14,18
typically (8)
54:23;57:21;58:4;
62:11;82:11;83:14;
122:6;124:2
typo (3)
64:7;82:22;121:3
typos (4)
38:24;49:7;54:8;
105:19

U

UK (2)
49:8,10
ultimate (1)
104:11
ultimately (1)
10:1
unavoidable (1)
154:25
uncertainty (1)
30:2
unclear (1)
15:16
under (44)
4:7,11;12:13;14:20;17:9;18:14;22:8;29:15;46:3,16;76:3,7,15;94:7;101:11;103:24;104:21,25;106:1;107:13;108:2,17;109:2,7,9;113:12;114:18,19;115:10,23;137:24;143:24;144:9;152:9,13,19;154:6,8,13,17;155:2,13,15,16
under- (1)
24:1
undercount (1)
54:24
underlies (1)
45:15
underlying (2)
11:7;131:3
understands (1)
124:9
understood (2)
16:8;88:11
undo (1)
110:11
unfair (1)
14:1
unfamiliar (2)
145:13;147:8
Unfortunately (1)
51:19
UNIDENTIFIED (2)
52:9;150:9
uniformly (1)
35:18
unique (6)
35:18;57:6,8;78:13;119:15;147:9
uniquely (1)
17:16
universe (2)
22:22;26:18
University (4)
28:23,24;32:2,4
unless (1)
116:15
unlikely (3)
46:6;67:16,19
unlucky (1)
30:3
unmute (1)
41:21
unrepresentative (2)
66:4;69:21
unsupervised (11)
26:11;36:2;37:22;48:3,9,24;49:22;50:9,11;89:22;113:3
unusual (1)
62:16
unwieldy (1)
45:1
up (33)
9:5;13:8,13;14:1;16:15;22:12;24:19;33:20;39:17,21;44:22;50:22;60:21;61:2;62:20;72:14;96:24;97:24;99:1;104:7;105:19;114:21;115:24;119:20;121:6;122:22,24;123:10;130:20;133:1,6;135:9;136:21
upon (2)
7:22;57:24
use (44)
11:25;12:17;15:5;18:15,16;27:8,20;33:10;35:23,23;45:3;47:18,24;48:21;49:5,9,20;51:6;64:14;67:2;69:4;71:5,18;74:20;78:11,22;79:22;84:1,21,23,24;85:10,13;86:11;90:2;106:6;112:18;113:24;117:10;127:16;131:10;136:12,15,17
used (25)
12:18;13:16;16:8;17:21,24;18:3;34:10;39:11;40:15;45:19;47:12;48:16;49:8,11;50:14;57:9;73:25;76:17;82:17;83:2,25;85:2;88:6;94:5;148:3
useful (7)
33:14;58:12;73:3;78:19,21;113:1;127:22
user (31)
16:10;38:7,17,17;40:4,6;43:8,9,15,20;44:19,21;53:10;77:11,11,15;109:19,25;110:7,10,20,21,22;111:2,17,23;119:7;124:11;128:16;138:7;139:24
uses (4)
47:6;100:1,5,25
using (38)
22:14,17;33:2;35:10;47:11;50:24;53:24;55:11;67:17,20;68:8;69:6,16;78:7;82:15;84:21;85:17;87:13;88:2;90:8,13;94:3;96:19;101:16;105:5;106:12;111:12;113:6,17;114:17,18;116:20,21,22;118:14;128:5,9;145:12
usually (1)
79:20

V

valid (15)
22:14;30:14;53:2,25;54:3,12;55:5;63:18;65:14,22;67:17;68:15;74:15;85:15;120:21
validate (1)
50:4
validated (2)
21:19;22:1
validating (1)
26:7
validation (2)
24:12;56:25
validity (2)
14:16;128:5
value (8)
59:9,24;92:3;93:15;95:15;109:6;122:19;155:14
values (16)
53:18,25;55:13;60:5,24;89:10;90:18;91:6;93:14;100:3,5;108:25;110:25;123:11;125:14;145:18
variability (2)
35:21;105:24
variable (1)
106:4
variance (1)
124:25
variant (2)
50:13;123:16
variants (4)
38:23;125:3,25;150:14
variation (21)
46:20;55:10;56:22;63:4,6;73:7;82:4;105:22;121:2;124:17;125:19;126:12;139:15,16;140:2;142:13;143:14,14;149:25;150:1,1
variations (9)
33:16;49:7;54:9;105:19;142:16,23;145:22;146:2,3
varied (1)
75:8
varies (3)
90:9;121:10;122:23
variety (7)
74:20,25;106:9;124:24;125:2,14;127:22
various (5)
45:18;56:11;59:6;65:3;97:8
vary (1)
70:10
varying (5)
45:21,25;47:4;79:25;80:1
vast (3)
142:22;146:12;149:22
vastly (3)
45:17;80:25,25
vector (2)
89:15;93:5
vectors (1)
90:5
vehicle (1)
29:1
versa (1)
104:25
version (2)
28:4;69:10
versions (1)
82:19
versus (6)
73:5;74:21;83:4;84:16;94:8;125:1
vertically (1)
69:3
via (5)
51:14;65:11;102:18;103:3;149:21
vice (1)
104:25
video (1)
136:24
view (4)
33:19;79:4;116:16;118:23
Virginia (1)
84:15
visual (2)
41:4;102:12
visualize (1)
103:3
visually (1)
55:6
vitae (2)
28:4,5
voter (1)
83:24

W

walk (12)
36:12,13;45:4,5,8;46:12;52:14;63:23;83:21;93:24;97:25;117:20
waterfall (4)
93:25;95:11,20;101:16
way (77)
8:8;13:3;16:8;18:4;26:21;37:12,23;39:11;42:25;45:25;48:3,19,24;50:9,11;53:12;54:9,10,12;58:13,22;60:12,18,20;62:1;64:16;66:25;67:2,13,20;68:8,15;72:4,10;75:16,25;77:1;80:2;86:15;87:7,15;88:7,24;90:10,12,19;91:6;92:19;93:11;96:14,24;97:16;99:6;100:23,25;101:25;102:2,4;103:2;104:17;105:16;106:4,19;108:8;112:5,10,19,20;113:13;117:17;118:14,18;119:8;139:1,2;143:13,19
ways (17)
14:12;31:3;33:16;53:1;54:10;60:20,21;66:22;67:3,4;73:11;80:21;81:22;96:19;107:22;118:22;152:21
weeks (2)
13:19;15:11
weight (2)
95:15;101:19
weighted (3)
108:3,5;146:13
weighting (1)
95:10
weights (13)
35:10;47:7;48:1;89:23;95:16,19;100:3;101:16,17,22;102:3;128:9;146:22
weird (1)
75:14
well-accepted (1)
18:3
what's (22)
9:10;15:4;19:17;25:14;26:10;37:14;48:22;51:17;66:23;

68:6;70:12,16;72:11;
76:1;82:6;87:9;91:15;
98:3;103:16,23;
108:20;148:14
whatsoever (2)
98:18;99:8
whereas (3)
35:19;62:8;81:19
Whichever (2)
19:4;149:7
whole (13)
13:12;30:10;36:1;
38:20;46:19;48:6;
68:1;71:22;96:25;
100:4;108:18;118:19;
154:17
who's (10)
20:19;21:5,6;
139:23;142:9;147:20;
148:14,18,23,24
whose (3)
97:19;112:25;
123:16
wide (2)
58:10;106:16
wider (1)
125:14
wife (1)
85:24
Wikipedia (3)
130:15;131:8,9
William (1)
47:22
willing (1)
142:25
wind (1)
16:15
Winkler (3)
47:22;49:4;88:20
winner (21)
18:10;20:13,19,21;
21:6;39:19;111:4,13;
133:16,25;134:8,9,14,
22;135:3;138:14;
150:24;151:1,2,21;
155:8
winners (7)
16:12;40:7,9;
149:20;150:18,19,20
winnings (3)
20:9,21;21:13
winnow (2)
37:12;40:21
winnowed (2)
76:2,3
winnowing (1)
51:9
wish (2)
18:25;119:6
wishes (3)
18:12;130:24,24
withdraw (2)
129:15;131:12
within (12)
25:16;35:6,8;37:2;
38:15,19,21;74:4;
106:5,5;128:11;
139:16
without (14)
16:24;20:19;23:6;
24:11;48:10;71:22;
73:3;84:5;86:15;
100:14;122:16;
142:18;143:3;144:21
witness (20)
5:11;10:3;27:20;
44:13;57:14,20,23;
71:2,8;72:2;73:14;
85:20;89:8;92:11;
101:10;103:2,12,22;
130:8;136:25
wondering (1)
87:25
words (1)
63:7
work (15)
7:1;13:20;17:11,17;
26:12,15,24;29:10;
31:16;36:3;48:24;
86:19;89:18;118:1;
125:11
worked (5)
8:14;21:18;26:6;
28:19;49:25
working (5)
8:24;26:8;36:25;
49:13;51:4
works (5)
35:17;102:13;
113:21;123:12;139:2
workshop (1)
31:14
workshops (2)
31:4,18
world (2)
18:1;146:23
worried (1)
8:17
worse (1)
106:2
worth (10)
38:3;84:11,11,14,
16,19;91:2;94:17,20;
130:13
worthy (3)
79:16;80:9;93:23
write (1)
26:3
written (1)
115:15
wrong (1)
19:17

X

Xiomara (2)
82:17;84:12

Y

Yahoo (1)
64:12
year (2)
14:21;94:6
years (7)
9:14,16;22:17;28:8;
29:12;56:16;143:15
yellow (5)
58:3;77:4;78:7;
110:2;127:22
Yesh (1)
38:12
yes-or-no (1)
34:24
yesterday (2)
6:9;16:1
Ympactus (2)
44:1,7
Y-M-P-A-C-T-U-S (1)
44:1
York (1)
91:4

Z

zero (12)
46:5;78:8,8,9;
82:13;83:13;90:24;
92:3;93:7;97:9;105:1;
111:15
zeros (1)
122:19
zip (5)
46:25;59:12;61:20;
77:24;122:23
Zoom (5)
5:5;51:25;52:6;
132:14;140:14
zooming (1)
71:11

0

0.002 (1)
54:19
002 (1)
55:18
02139 (1)
25:3

1

1 (10)
10:17,17,18;92:18,
19,21,22;128:23;
129:7;137:22
1,000 (1)
67:12
1,005 (2)
67:12;68:13
1.1 (1)
118:6
1.5 (3)
40:3,5,8
10 (4)
10:16,18;52:17;
156:11
10:28 (1)
4:1
100 (9)
82:12,19;93:7;97:9,
11;111:15;116:14,15;
146:10
1006 (1)
12:13
101 (2)
72:12;118:15
11 (5)
30:24;31:23;40:4,6;
52:14
11:26 (1)
41:17
11:30 (3)
10:9;41:14,18
111 (1)
121:11
117 (3)
81:13;93:20,22
12 (2)
31:25;49:18
12:55 (1)
101:6
13 (1)
30:24
132785 (2)
140:6,11
144 (1)
85:11
14-40987 (2)
11:3,4
149 (2)
106:24,25
15 (1)
30:25
160 (1)
151:12
163com (1)
64:10
16-4006 (2)
4:2;11:9
16-4007 (2)
4:3;11:12
168 (1)
85:11
1900 (1)
58:11
1969 (1)
45:13
1990 (1)
49:5
1st (1)
58:11

2

2 (6)
10:17,17,18;101:4;
128:23;129:8
2.1 (3)
60:4;61:3;71:9
2.3 (2)
62:20;71:9
2.4 (3)
62:20;63:24;71:9
2:07 (1)
101:7
2:59 (1)
137:10
2021 (1)
32:25
217 (2)
145:15,17
2522900 (1)
153:24
256 (1)
25:2
26 (1)
55:24
2600 (1)
126:10
27th (1)
13:5
29 (2)
69:10,10
2nd (2)
12:4;13:5

3

3 (7)
27:13,15,22;31:9,
10;131:17;140:4
3:27 (1)
137:11
3:54 (1)
156:12
33 (3)
32:18;131:15,17
34 (2)
131:23;132:5
36,000 (1)
119:2
37c1 (1)
4:14

4

4 (5)
10:17,18;42:24;
129:12;156:4
4.3 (5)
77:13,17;81:8,10;
109:4
4.36 (1)
79:8

400 (15)
60:16;61:8,10;62:7,
25;71:1,3,4,5,6,10,16,
25;72:15;126:23
442 (2)
11:2,11
444 (3)
124:11,12,18
471 (1)
151:3

5

5 (12)
12:6;13:23;16:5;
43:12,17;68:13,15;
136:21;137:22;153:9,
13,14
591 (2)
11:3,12

6

6 (2)
44:25;129:12
60 (1)
109:6
602 (1)
4:11

7

702 (4)
4:11;14:20;18:14;
22:8

8

8 (6)
70:19,21,23;
129:12;130:8;131:6
80,000 (1)
40:6
81,000 (1)
149:19

9

9 (3)
13:8;31:5,8
9.5 (3)
81:5,9,19
955 (1)
25:1
98.2 (1)
56:9
988 (2)
117:4;126:16
99.1 (1)
107:6
99.6 (1)
117:3
998988 (1)
117:18
99th (1)
107:5
9th (1)
16:5