***Practical Significance* | Episode 4: Data Science – What's the Buzz?**

Donna LaLonde: Welcome everyone to *Practical Significance* where Ron and Donna get to have great conversations with wonderful people. And the April *Practical Significance* episode is no exception. We're joined by **Kathy Ensor** and **Mine Çetinkaya-Rundel** to talk about what's going on in their lives and in data science. And I'll start out with a question for both of you since we are celebrating Mathematics and Statistics Awareness Month. Could you tell us about a project or projects that you are working on that you believe will have a big impact. And Kathy, I'll start with you and then we'll turn it over to Mine.

Kathy: Thank you for including me on this fabulous podcast. You know, like many of us across the statistics profession, I have been involved in several COVID projects over the past year, and I think one of the most important ones is looking at wastewater for the city of Houston as a surveillance tool. Not all for the SARS-CoV-2 virus, but for other outbreaks that might occur in the future. I am working with all these fabulous scientists in the Houston area on the statistics component and the data science component, because it really is: measurements come in, answers come out and we have about 30 minutes to turn that around.

And so, developing that system for properly accounting for all the uncertainty and these measurements, each measurement is extremely expensive in terms of money and personnel. And we end up with about maybe 2000 measurements each week. And so, pulling all of that together with all the dynamics from a statistics perspective has been challenging, but it's also been very rewarding because we do see when those answers come out, they impact public policy immediately. So, we have that leading health expert in Houston, anxiously awaiting our results as well as the mayor and people in the surrounding areas so that they can properly help our community fight this pandemic.

And I know many statisticians across the board have been contributing in so many ways. As we march through this rather odd couple of years, and it's just so important and we may be growing a little weary at this point, but we are seeing the light at the end of the tunnel. I believe I do want to point to one other effort that I have an ongoing in my research, and that is the area of quantitative finances. I do lead the Center for Computational Finance and Economic Systems at Rice. This is an area where statisticians and data scientists have transformed how financial markets work and that will continue to be the case. And so, I also think this is very important because what we try to do with our center is to educate these quantitative scientists beyond their algorithms. And I think that's an important thing. We need to understand the impact of the algorithms that we're developing. And with that, I'll turn it over to Mine.

Mine: Well, thank you for having me as well. This is a great question and makes me think, am I doing anything that will have a big impact? Right now. it's DataFest season for us. So, I am working on organizing DataFest this year. And one of the things that I'm both nervous and excited about is that obviously just like everything we're running these events virtually and we're thinking alot about how to give students a sense of community in this virtual environment. And I'm excited that something positive might come out of that, especially because in the past, one of the ways it seemed like other schools and institutions could, host their own DataFest would be that they'd need to have the budget for the physical space and what not.

And that's not always necessarily true for every school. And I'm wondering if we can learn something from this that might help supplement the DataFest efforts in the future as well. So, trying to see the bright side of that. Another project that I'm working on that I'm really excited about is a new textbook with Jo Hardin, as my coauthor, called *Introduction to Modern Statistics*. And we've really been trying to think about how we bring a little more of the hands-on data and analysis and more modern approaches, like computational and simulation-based approaches to introductory statistics and trying to do it in a way that's in line with all the other open-ended textbooks we have, which students seem to find very approachable.

And so, it's been really fun working on that project with her, and we're kind of hopefully finishing up our first edition for publication. So that's another thing that I hope will be helpful. This is going to be freely available, also on the web. And then also we do have print copies that are available at the cost of printing, and we've been crunching some numbers in terms of how much money students may have saved using these textbooks as part of their courses. It's really, really nice to see the sum of that amount. And just knowing that we're able to provide introductory textbooks for free to new learners.

Ron: Thank you, Mine. And thanks, Kathy. Mine, we'll circle back to more about the intro course and DataFest a little bit later in this podcast, but I'm going to swing over to Kathy again for a minute. We read everywhere these days, AI, machine learning, data science, et cetera, there's a lot of buzz. So, what Donna and I wanted to ask you today, Kathy is as president-elect of the ASA, what do you want the public to know about the role of statistics in these various endeavors?

Kathy: Sure. That's such a great question, Ron and such an important one as well. So statistics is rock solid one of the foundations of data science, AI, and machine learning. And, when I sit through presentations on these topics and I see a perturbation of principal components, elevated a century later, but all the work that we developed over these many years in our profession and contributed to the world are now being put in day-to-day use. And statisticians continue to develop at the cutting edge of this broad field of AI, machine learning, and data science. And so, we each enter this arena and in different ways, but ultimately there is a methodology behind every answer that comes forward and statisticians think hard about those methodologies and how they might be used, but also what are some of the pitfalls that might happen with their use.

And so, we'll continue to do that as a field, continue to innovate on this area, this broad area. So, we've seen AI, machine learning, and data science evolve. I remember, maybe a decade back watching some leading statisticians who had entered into the field of self-driving cars, which we know will be with our society in the next decade. Those will become a reality. They're already becoming a reality. And so, the algorithm at the base was the Hidden Markov model. And I thought, wow, I think it hits the Markov models all the time. That's where I innovate. And so that's really picking out these key strategies and statistics and then building on them. And it's just so important. This foundational work that will drive much of our world in the future.

Ron: I really appreciate that Kathy and self-driving cars are an exciting thing. I operate a poorly driven car. That may be another thing about the self-driving technology is that it will help those poorly driven cars be driven better. So, you'll have cars that keep you in your lane and keep you from rear ending the person in front of you. So, so it's, so it's a win-win, right?

Donna: Yeah. So, thanks for that. I know I'm definitely looking forward to that. Not for you, Ron, but for me, because you know that I hate to drive. So, I'm going to switch gears a little bit and come back to the education theme and Mine, I know you as a faculty member at the University of Edinburgh and Duke University and as a professional educator at R Studio. In fact, this morning, I was actually listening to the recording of your presentation on providing feedback at scale from the R Studio Global 2021. So, I'll give a shout out to that presentation. I wonder if you could talk with us a little bit about how you're thinking about the content of the intro stats course or the intro data science course has evolved.

Mine: Yeah, that's something I spend so many of my waking and sleeping hours thinking about, I suppose. And, and I think things have evolved. So, I feel like I've been teaching for about 10 years or so at this point. And I will say that one of the ways my thinking has evolved about approaching the introductory classroom is I used to think a lot about there are the tools in front of us. How do we make the best of the existing tools? How do we make it approachable for our students, like computational tools say, let's say we're teaching art. And I think one of the ways my thinking has reshaped a bit, especially with my work with R Studio is that sometimes as educators, I feel like we don't have to contort our teaching into what the tools are. It's also possible to make the tools what we need as well.

And this is not some revolutionary thinking. People have been creating tools for introductory statistics education for decades and decades at this point. But some of those tools perhaps used to be those that would be just used in the introductory classroom and not necessarily grow with the students. But I feel like especially within our community, there is a really welcoming attitude for new learners and a lot of desire from lots of packaged developers, but also infrastructure developers to make things a little bit easier. The first time you get an error, the first time you must debug something to make things a little easier for students and making easier actually ends up making it approachable for the whole community as well.

So, I feel like one of the ways my thinking has evolved is what the best way would be to introduce a particular tool or technique to a student. And if what's existing out there isn't good enough - can we develop something that would actually fit better into this paradigm? And as I said, I think maybe for things like an R Package, this might really be some little update, like updating the documentation to make it more approachable. And that's something I feel like you don't have to be a developer to do. You could be an educator who is making this sort of contribution to existing R Packages, for example, but even thinking beyond that and thinking about computational infrastructures and just asking people like system administrators, can you make this easier for my students? “I know this is the system that you have set up, but it is simply not working for me.”

And people don't always say “yes” to these things, but I think at least starting to have those conversations can really change what the experience of students is like, which then brings me to the other point of like how things have evolved.

I think it is a given at this point that computation should be a part of introductory statistics, as much as it is a part of introductory data science. That's part of the GAISE recommendations, for example. So, it's also what a lot of us have been doing in our classes for years and years. But just saying that you want to teach computing, doesn't necessarily mean that's going to go well, if the tooling that you're using isn't designed necessarily to be used by new learners who are not only working with data for the first time, but just writing any piece of code for the first time in their lives. Another way my thinking about what's important in an intro course has evolved is I feel like when I was first starting to teach, we used to think a lot about what might be the hardest thing for students to learn.

And let's try to spend the most amount of time on that. And that thought process generally resulted in spending more time on the mathematical approaches for doing inference. So mathematical models for doing inference because they do tend to be a little bit harder for students to wrap their head around. The central limit theorem, for example, is not necessarily an easy thing to learn. And it looks like a histogram should be an easier thing to learn. That's probably true, but I feel like that doesn't mean we shouldn't spend time in the classroom teaching students how to do exploratory data analysis. So simply because the most basic exploratory data analysis kind of tools like histogram summary statistics are easy to wrap one's head around doesn't necessarily mean when somebody gets a data set in front of them, they know how to start exploring it. So, I think more and more educators have been trying to make room for more meaningful exploratory data analysis in the intro course.

Not the standard, “let's draw a box plot by hand,” but thinking how we might approach this analysis to begin with. Might we want to transform the data a little bit to give us something a bit more meaningful to model? And that's exciting because I think it's hard to get excited about a data frame somebody has handed you, but I think it's a lot easier to get excited about a data frame somebody has handed you and then you've put your own touch on it to make it perhaps a bit more useful for the type of analysis you want to run. So, in an intro course we always talk about, we want to get students excited about working with data. So, I feel like we must give them the data to get excited about.

And I'm really happy to see that I'm not alone in of some of these thoughts that many other educators are tackling this as well. And the one last thing I'll mention is not only are people working on this, but it seems like more and more. I'm seeing folks openly sharing their teaching resources as well. When we first started open intro as an open-source project, I think 11 years ago at this point, I don't think there were as many source introductory stats materials out there that's being generated by lots of folks and putting out on the web. And it seems like that trend is changing. And that's a really nice thing to see as well.

Ron: Mine, thank you so much for that. I was thinking about my long connection with the introductory statistics course. I taught it for the first time in the fall of 1978. And over the years, it has been really exciting to see the changes, the evolutions that have taken place, but it seems like that evolution as you just described, it is accelerating and that is super exciting. So, thanks for that. Kathy, I'm going to turn to you. And switch gears back to data science again and ask you as you think about it from your leadership perspective in the ASA, but also from your leadership perspective as an academic. What's most exciting to you about data science and what keeps you up at night? What worries you a bit?

Kathy: What a great question. And before I answer that, I just have to applaud Mine for these efforts. It's just how I try to teach it. She's so right. That sometimes the tools don't allow us that innovation at the teaching as easily as we would like it to be. So, I am super excited about data science and the contributions that statistics makes to this ever-changing the world. So, we are living in a data-driven world. Our world will continue to be a data-driven world. And what changes is the ability to process information quickly and to get credible answers to information that's flowing through.

I think the kinds of problems that data science are addressing are broadening, as we see these methodologies and this type of thinking – even our social and urban space, as well as. The more scientific space has, for example, the self-driving cars, getting us to Mars, you know, all of these things that we as data scientists and statisticians are contributing to. But I think what keeps me up at night is maybe a little bit of the overuse and the ability to get answers and then maybe not stepping back and understanding that maybe those answers aren't exactly right. And as credible data scientists, that is our role. If we don't bring in the issues related to the tools that we're developing and that are being used, that other people aren't going to understand it. And so, that's a little bit of what keeps me up at night.

You know, we have these extremely powerful tools that we're building and that are exponentiate in terms of their ability to find answers. But as we know, even if we go back to just those very simple, basic statistical methodologies, we know that there's always assumptions. There's always issues that might change. The answers are changed the way that we view the answers that, that our methodologies provide for us. And I guess I worry that sometimes we don't take enough time with that end game, you know, so it's data to knowledge, it's data to action, but we want to make sure that we're acting on credible information that any algorithm brings to us.

Donna: Kathy, thanks so much for that. I have to say that one of the aspects of the communications that the ASA does, and members of our community contribute to is to remind us about the importance of data ethics. And so, it makes me really proud actually, that that keeps you up at night because I know that that means that the ASA we'll continue to focus on that and play a leadership role in making sure that there's accountability. Thank you for your leadership. And for that answer, I'm going to swing back to Mine and talk a little bit more about ASA DataFest. We are about to begin ASA DataFest season for 2021. I wonder if you could tell us a little bit more about ASA DataFest events and why you commit so much time and energy to bringing the event to the huge number of participants. Since I've been a part of the ASA, I've just seen the number of participants grow and that's so exciting.

Mine: Yeah. It is such a fun event. ASA DataFest in a nutshell is we call it a celebration of data. It is a data competition, but it's meant to be a friendly competition that takes place over one weekend. And during this time, we give teams a surprise dataset – a big and complex data set. So, I don't mean big – just in the sense of row, lots of rows, but more importantly, a complex real data set. One they probably wouldn't come across otherwise. It's usually provided to us by one of our data partners and it's not an otherwise publicly available data set. And we give them some rough guidance and a vague challenge to get them to get creative. And one of the things I love about this competition is that we have the competition aspect to get and keep students motivated.

It's hard to otherwise. You know, definitely it's a full weekend; however, much food you might be given at the venue and we have prizes in different categories like "best data visualization," "best use of outside data," and "best insight." And, I have to say that one of those categories "best use of outside data" is so near and dear to my heart because it really gets students to be creative about. Okay, I have this data set in front of me. What other information or pieces of data could I find out in the world and bring it together to tell a data story with this? So, the fact that we have these three categories, I think also makes the event very approachable to students from a variety of levels. First year undergraduates don't feel intimidated as much to participate because if this were just about the "best model," I think, they very well might feel intimidated because their peers who are say fourth year students who've done a lot of modeling courses would inevitably do better than them.

But here there's a lot of room for creativity, which I think is really nice. Usually, we hold DataFest in person, which means we feed the students. They're staying up late at night. We have consultants that come in and chat with them. And these consultants are both academics or graduate students or faculty, but also data professionals that are local to the university where the event is happening. This time, we're going to do things virtually, but those interactions both with others in the university and data professionals outside of the university are so valuable for our students. They really strike really nice conversations with them. And I hope that we're going to be able to keep that going. We tried this last year. Some of the DataFest events did take place last year, but this season 1 year ago was an even more uncertain time for everyone.

Many folks have also canceled their event. But I think this time we are all the organizers. I think we have over 30 institutions and 30 events running with even more institutions participating. And we are thinking a lot about how to bring back the sense of community to the virtual environment. And I'm cautiously optimistic that we're going to be able to pull it off. The reason why I've put in so much effort into this? Well, it's myself and Rob Gould, who was at UCLA and the founder of DataFest, are kind of the co-leads. One of my tasks often ends up being the first person to crunch the data that we're given to wrangle it a bit, to play around with it a bit to see is this the sort of thing that we can build a challenge around.

And I personally really enjoy that activity myself. So that's a selfish reason why I continue to be involved. But beyond that, I have seen over the years, the impact that it has had on our students. So, for example, not only do students find the event enjoyable, but we have DataFest alums who are always wanting to come back and participate again the next year. If they've graduated and you reach out to them and say, "Hey, would you like to be a consultant or a judge?" They will fly out and participate. They will ask their employers. Is this something you can support for me? So, I can go back and do it because they clearly have gotten a lot out of it. Many students will tell us that this becomes a really nice experience, not just from a learning perspective, but also a good story to tell, say at their interviews.

I think they're often asked questions like, "Tell us about a time you were crunched for time and you had to work with a team to solve a problem." And it's like the definition of DataFest. So, it gives them that story to tell. And I also like that we take things lightly. It is DataFest, not StatFest. It's also not a hackathon. It's about celebrating data. I personally find it nice that there are lots of these like hackathon efforts out there that have like this let's save the world kind of undertone. And I think that's a really nice motivation, but I think it's hard to imagine that you can really make an impact. So, such a big impact on such a big question over one weekend of like pulling a couple, all-nighters, a real change doesn't come like that.

But with DataFest, we don't try to over promise, but we tell them during these 48 hours, you will learn a lot. And they certainly do. I'll say what all they think as a faculty, DataFest has been incredibly informative for revealing gaps in our curriculum. One year we had a time series type, data set, or at least a data set that would really be helpful for students to know something about how to do time series analysis. Time series notoriously tends to be not something as well covered in undergraduate statistics, curricular, similarly, spatial data analysis. So, these are the sorts of things that you probably kind of know as a faculty, but in an event like this and seeing the students struggle and you knowing if they had just known about XYZ, they would have such an easier time and we're fully capable of teaching them that; we just hadn't necessarily reviewed our curriculum with that in mind can be really informative.

Another thing is it's a really nice opportunity for students to get to know other faculty in the department as well. I always try to get, you know, maybe pester is the right word, all the faculty in my department to sign up as consultants, because some of them don't necessarily see undergraduates until they are a third- or fourth-year students, because maybe they tend to teach more graduate courses - advanced courses, and it becomes an opportunity for them to get to know these students early on in their careers for the students to get to know them. So, I think all that ends up being quite fun. I'll end with an anecdote. My first involvement with DataFest, as I said, I was a PhD student on my last year at UCLA when Rob started this event and my involvement was getting In and Out burgers at midnight, that was my job, midnight snack.

I don't think I had a whole lot more to do with it, but I went and got the burgers. People who know me know that I don't mind doing things late at night. So that was my task. And then I lingered and I talked to the students, ended up being serving as a consultant and like helping them solve some problems they were working with. And then the next year after I graduated and started my position at Duke, I reached out to Rob and said, I wanted to do this here too. Can we do it together? And I'm really thankful to my department at Duke who supported this tiny event. We were in with just 23 students that year, that we did concurrently with the UCLA DataFest. And since then, it has been growing. I checked the numbers before today. And I think we're expecting about 2,500 students to go through it this year, even in a pandemic year. So, I think that's great success and it's wonderful to see it grow. And I just saw that we even have an institution participating from Turkey which is where I'm from. So that's really, really nice to see.

Ron: And your passion for this really shows through. And it's very much appreciated, I guess, in Durham, you're out of luck on getting In and Out burgers, but otherwise you were able to recreate the event there just fine. So, as I was thinking about this podcast and, and knowing that we were going to ask you both a question about the future. So, thinking about that famous quote, something along the lines of "Prediction is very difficult, especially if it's about the future." I thought I'd look up to see who said it. I vaguely remembered that it was a physicist. And I thought you'd be interested to know that in addition to seeing Niels Bohr, it was also credited to both Mark Twain and Yogi Berra.

And of course, it was because everything Yogi Berra is alleged to have said that half the things he said, he didn't say, but I'm going to ask you now – if maybe not to predict the future, perhaps to envision it or hope for it in terms of what you think statistics and data science would be like in 10 years. And again, not so much a prediction as maybe a dream or hope for the future of statistics in 10 years. And Mine, let me start with you.

Mine: Thinking about the of future of statistics and data science and where they will be – here's my answer: I hope they'll be sitting in a tree “K I S S I N G.” This is what I came up with. Um, I think there's so many commonalities in what I think of statistics and what I think of data science. And my hope is that these fields evolve together. And there's also computing and mathematics there. And lots of humanities. Like we were talking about data ethics earlier, you know, us statisticians are not, probably not the best trained people to talk about ethics alone. So, there's a lot of space in their room for kind of bringing all these disciplines together. And I just hope that as things evolve, things have evolved together. I think sometimes with the idea of data science and I'll speak mostly from an academic perspective here, there's like this turf war stuff that happens a little bit with different existing departments and disciplines.

And I understand entirely why, because funding is limited only so many people will get to have a piece of a limited pie, but I do hope that that sort of stuff keeps promoting kind of growth, but not necessarily make it difficult for people to innovate. I hope that we can put a positive spin on all of that. I hope that we can broaden the definition of a statistic. I remembered a quote from Rafael Irizarry - he had said in a blog post, “The term data scientists therefore became useful for making the distinction between someone, with experience analyzing data in all its messy glory versus someone that can prove an estimate is asymptotically normal.” And I personally consider myself a statistician and I can prove some things asymptotically normal, but I wouldn't say that's what I specialize in.

I think I'm more specialized in trying to analyze data in all its messy glory. So, I think as statisticians, we have actually always had to do this sort of thing as part of our data analysis, but maybe it wasn't always, at least in my view, so well-represented in our curricula when we are training others. And I hope that that can change a little bit. And if data science, the existence of data science ends up being the catalyst for that, that's fantastic. I hope how we approach statistics; education changes a little bit, and it has been changing. And I hope that continues such that we don't think of competing as something students will just have to figure out at home doesn't require airtime and similarly working with messy data. And I hope that that is not just true at the undergraduate level, but at the graduate level as well.

And the other thing I will say is that (I'll put faculty member with a teaching focus hat on here) is that actually a lot of teaching faculty have been doing a lot of this new teaching of how do we incorporate data science into the introductory classroom? And I feel like our upper-level classes have not changed as drastically. And I hope that that effort that all the faculty are putting into innovating in the introductory classroom is well-recognized; sometimes departments tend to not recognize these. I hope data science becomes a catalyst for that. And I hope that we can keep that momentum to really think about about our curriculum as a whole, as opposed to just the introductory classroom.

Ron: Thanks so much. Kathy what's on your mind about this.

Kathy First of all, I just have to celebrate everything that Mine says. It just resonates so strongly with my perspective as well. And, and I do teach time series. So that is one of my contributions to the next generation iin trying to bring in the messy data aspects as well as the importance of some of the theorems. We don't necessarily prove them in the class that I have. But what I want to sort of move a little bit more toward the innovation and technology side of where I think statistics and data science will be in 10 years or what I hope to see happen. So we've been living with these innovations for the last 20 years and so many tremendous scientific advances have been found in personalized medicine and the knowledge that's being gained from a neuroscience and understanding the gut microbiome, for example.

So, these are all scientific technology questions where statistics and data science are really moving the boundaries with these new methodologies that are being brought forward by our community. But as I sit and watch this and I do, in addition to my own work, I’m surrounded by very active research statisticians and data scientists. So, it really is drinking from a fire hose to try to keep up with everything. But what we begin to see is maybe it's time to think more deeply about some of that robustness and sensitivity analysis that really cameinto regression in the 1980s. For example, as we started to understand the influence that one observation might have. And, so I think some of those automated, you know, going beyond prediction and more toward estimation and inference. Not, not just focusing on predicting from the data that we have.

And so, I think the development of some of those tools or that type of thinking is beginning to enter, I would say, the machine learning community that drives much of the research. It's a statistics innovation paradigm and the patient community. I put that in the "machine learning big bucket." So, as I do think from a technology perspective that we're going to be seeing some of our great thinkers in this arena coming up with new tools for us to implement on a day-to-day basis. So, then there's the other piece, you know, not all data is big. And I do worry that some of our curriculum has so focused on this exciting data science world that we forget that not all data is big. Like for example, with my wastewater effort over the summer, that this is not a big data problem, this is an intricate modeling problem.

And so, we want to maintain that component of our curriculum and maybe bring a little bit more resurgence in experiments of design and sampling as some of those topics are not so celebrated in our graduate curriculum one of these days. So I do hope that those are some things that come back in. But I am fully on the page of training people, both in the foundations and in the implementation. But at the end of the day, what we're really interested is how can we make an impact on big questions in society, science, and engineering. And, I hope that statisticians and data scientists, I know that they will, but that is what I hope for the future.

Donna: Well, Mine and Kathy, thanks for actually making me optimistic about the future. Sometimes over the past year, it's been a little hard to be optimistic. So, it's nice that you can provide a future focused and positive outlook. Well, to conclude our conversation, I'm going to bring us back to celebrating Mathematics and Statistics Awareness Month. And I know this is a difficult ask, but I wonder if you would share the names of one or two individuals who have influenced your work and who you would like us to celebrate during April's Mathematics and Statistics Awareness Month. And Mine, I'll start with you. And then Kathy, have you chime in as well.

Mine: So this is really a hard question to answer. I knew that this question was coming, and we'll just have to pick two names out of a hat. So, one of the names that I'll give in terms of people who have, influenced my work and how I approach my work is Rob Gould, who I've been working on DataFest with. And he was kind of a mentor for me when I was in grad school. And I think one of the things that I really like about his approach to education is he's truly a risk taker. I don't know if it's something about being from Southern California, or something, but just this kind of trying out new things and doing it with an ease, I think is really nice. I hope to have that sort of confidence every once in a while. I think it's really helpful because it's hard to kind of innovate, especially with something like Intro to Stats where things are changing, but then things aren't changing all that much from year to year. So, I think that for me has been like a really nice way of thinking about what projects I'll be working on, what can be impactful, what feels like taking what level of risk, and whether I can pull that off or not.

So, kudos to him. And, I really enjoyed working with him on DataFest. And I really liked that we're still able to build on that. Another person I'll mention is Jenny Bryan, who I'm lucky to work with as part of my R Studio work. But also, I feel like has had so much influence on how many of us approach computing, especially with R and D. I feel her work has influenced mine. She has a really good style for articulating things and does not shy away from almost over articulating them. But if you read things that she's written, like her books, manuals or documentation, there's a particular style where you really start reading and you think, well, if I decide to take this on, and then I struggle, I think I'll be able to find the answer here.

And I think given that kind of assurance to the reader is difficult. So, in addition to her like technical work, I think the way she articulates some of these ideas in a way that's both very approachable, whatever level you're at, but also with enough detail, even if you are a kind of a super user is really nice.

And I feel like that sort of approach has shown up in my work in the form of writing more and more of the nitty gritty of teaching computing. So, I feel like we would say these words of like, I use these tools and then I set it up this way. Then you do this in your classroom, then your students will have an easy time. I feel like I've said these sorts of words in conference talks and what not before but thinking about how helpful certain documentation has been for me, for other aspects of computing, I've been writing the kind of detailed things about how one can use a particular tool for teaching. And I've had positive feedback on that. And I must give credit to Jenny's style.

Donna: Those are definitely two people were worthy of shout outs. Kathy?

Kathy: Donna, thank you so much. What a great question. And you know, there's just so many people throughout my life that have had just such a positive impact on my growth as a statistician and in my growth as a human being. And I hope that I am able to, to share and pass that along to the next generation. That's one of my goals every morning when I wake up, but I'd like to maybe talk about two specific individuals and they come from the early part of my career because you know, I'm old enough that when I started as a statistician, I really was one of the very, very, few women in academia and our profession. And so, you know, we'd look back on those days and we, we don't really necessarily remember all those struggles. And because today the statistics profession is just such a rich and diverse community and, and we should strive to even make it richer and more diverse always, but to people that, that just early on made a huge impact on my ability to contribute the way that I do. But first, my advisor, Joe Newton, he was just simply an exceptional advisor.

And, we stay in close touch, regular touch. He's been advising me throughout my life. And I consider he and his wife very, very close friends. But on the statistics side, Joe is just an exceptional statistician coming from the time series arena. But he also understands how statistics really bridges with computer science and mathematics and how our foundations really truly provide opportunities for all the innovations that we've seen over these many decades that statistics has contributed to, but he also understood that statisticians make excellent leaders and he himself went on to become a Dean of Arts and Sciences at Texas A & M and really led the growth of science at Texas A & M for many, many years. And, and even now there's a statistician as Dean – Val Johnson, who is continuing the tradition. So, Joe was one and then the other is my colleague, Jim Thompson. So, Jim Thompson was the founder of the statistics department at Rice. He and David Scott joined together to create this fabulous new department. And I was actually the first hire of that department. And so, you know, Jim and all my colleagues, but let me focus just on Jim.

Jim really encouraged my interest in growing my statistical abilities to contribute to society. So, not just to contribute to advancing the boundaries of academia but stepping outside of those parameters to bring science to the everyday world. So, kind of the term "civic scientist," and really pushing the idea of a civic scientist and Jim himself, led many efforts in that regard. Many global efforts in that regard. And, I'm sure that I won't measure up to some of the contributions that he made, but just by celebrating and embracing my interest in that direction, I do on a regular basis, use my statistical talents and knowledge to contribute to different aspects of science and engineering for the greater good. And I think many of us do, but without the support structure that allows us to entertain those bigger issues, it just wouldn't happen. So, I'll land on, on Joe and Jim, and really thank them and acknowledged their contributions to the field of statistics, but even more importantly to society in general.

Donna: Well, thank you to Kathy and Mine for a wonderful conversation. And now as is tradition, I will turn it over to my colleague, Ron for "Ron's Top 10."

Ron: Thank you, Donna. Oh, we've had a great discussion today and we see that there's a lot of things for statisticians to think about and a lot of things for statisticians to worry about. And among those things is ensuring that the data we work with is high quality.

And as a public service and putting the practical in *Practical Significance*, I'm going to alert our listeners today to the Top 10 signs that your data may be rubbish:

#10: You were told it was collected randomly, but randomly was in scare quotes.
#9: The data was recorded in crayon.
#8: The data collection was by four middle schoolers on bikes who were paid by the number of responses.
#7: Data cleaning consisted of removing all the numbers.
#6: The researcher didn't like every third data point is the number six, six, six.
#5: No one seems to know why the data was collected in the first place.
#4: Survey instructions included good things will happen to you. The more often you select options C.
#3: Participation was limited only to people who still use Windows 98.
#2: Entry was by a monkey typing for an infinite amount of time.
And the #1 your data may be rubbish: Questionnaire design was done by Four Seasons Total Surveys.

Donna: Thanks very much, Ron. And thank all of you for joining us for *Practical Significance*. We look forward to talking with you again next month.