**Show US the Data Podcast Transcripts – April 2022 Episodes 1-6**

**Season 1 Ep 1 Dr. Ray Hart**

Dr. Nancy Potok

Welcome. I'm Nancy Potok, and you’re listening to Show US the Data. Our guest is Dr. Ray Hart, Executive Director of the Council of Great City Schools. Our machine learning tool has identified Dr. Hart as a top researcher using National Assessment of Educational Progress, or NAEP data collected by the National Center for Education Statistics at the US Department of Education. Thanks for joining us, Dr. Hart.

Dr. Hart

Thank you very much for having me. And on behalf of the research team here at the Council, and the Council of Great City Schools, we're happy to be a part of your conversation.

Dr. Potok

Dr. Hart, you've done a lot of work with NAEP data leading to some very important findings. Tell us what was exciting to you about these data? What was the real-world effect your research?

Dr. Hart

Well, first of all, thank you again for having us. And we just finished a report last summer that was titled “Mirrors or Windows: How Well Do Large City Public Schools Overcome the Effects of Poverty and Other Barriers”. And in that report, we looked at comparing large city school districts to the nation or not-large city school districts to the nation-at-large, and whether or not they were having differential impacts on student achievement. Well, we call it torque. We actually controlled for variables such as poverty, race, special education status, and the makeup of districts. And what we learned is that our large urban school districts had significantly more torque than their peers around the country between 2009 and 2019. And so it's exciting for us from the perspective of understanding the difference that our school districts are making, in terms of the educational outputs for kids that they serve throughout the country.

Dr. Potok

Those are exciting findings. Can you elaborate a little more on how your work ultimately makes a difference for the American public?

Dr. Hart

I think for the American public, it sends a message that our large urban school districts in particular, show that for the students that they serve, they are getting far more educational bang for their buck, if you will, then other school districts throughout the nation, which is counterintuitive, if you only look at proficiency. So if you only look at proficiency, it doesn't take into account the students’ that are served level of poverty, that those students served have. But when you do take that into account, you recognize that. To be sure, our school districts have a long way to go. We want to make sure that all of our students have the same performance both in large urban districts as well as our suburban and rural districts. But what it tells us is that those schools and those school systems are actually doing a very good job of educating the kids that walk through their doors.

Dr. Potok

You raise some very important points. Based on your hands on experience, what would be helpful for other researchers to know about these data?

Dr. Hart

I think one of the important things for researchers to know about the data is the real impact of, as we termed it in our report, abject poverty. And one of the things that we looked at was not just poverty, which is measured by free and reduced price lunch traditionally, but whether or not students who were in abject poverty actually have more educational barriers to overcome. And we learned that in fact, they do, when we introduced a new factor from the US Census Bureau that accounted for the school environment. And whether or not the school was in a community that was experiencing real abject poverty. And what we noticed is that it produced very different results for our findings. In fact, what we learned is that those students who are just inside the free reduced price lunch range, but at the top of the range, perform significantly better than students who are really in abject poverty, and the schools that serve them look very different. So accounting for that, going beyond just the free reduced price measure that we've traditionally had in research is an important element of any analysis that you might conduct on this data.

Dr. Potok

Those are wonderful insights. Are there other datasets that would be helpful to access or to link to the NAEP data? Or perhaps there's something you'd like to see improved about these data?

Dr. Hart

I think one of the things that we often ignore in our educational research is the fact that if I'm looking at NAEP, which actually begins in fourth grade, I'm ignoring where students began their educational careers. And so quite often, we look at jurisdictions, we look at states, we look at schools, at the third or the fourth grade level. We make a determination about how well they're educating their students. But I think it would be great if we could link for example, the National Early Childhood Longitudinal Study data to make datasets where we looked at where kids began their educational experience - kindergarten, pre K, and then where they ended up in third or fourth grade. By doing so we would really begin to understand the educational torque that schools have on their kids’ learning experiences, recognizing that not all students begin school and kindergarten or pre K in the same place. And so I think that's an important element that we need to add to our educational research repertoire.

Dr. Potok

Thanks so much for taking the time from your work at the Council of Great City Schools to share these valuable insights with us, Dr. Hart.

Dr. Hart

Thank you very much for having me. It's been a pleasure.

Dr. Potok

Check out the show notes for more information and links to relevant materials. Until next time, this is Nancy Potok with Show US the Data.

**Season 1 Ep 2 Dr. Becca Jablonsky**

Dr. Nancy Potok

Welcome. I'm Nancy Potok, and you're listening to Show US the Data. Our guest is Dr. Becca Jablonsky, Associate Professor and Food Systems Extension Economist at Colorado State University. Our machine learning tool has identified Dr. Jablonsky as a top researcher using the Agricultural Resource Management Survey or ARMS data collected by the National Agricultural Statistics Service at the US Department of Agriculture. Welcome Dr. Jablonsky.

Dr. Jablonsky

Thank you so much for having me.

Dr. Potok

Dr. Jablonsky, you've done a lot of work with ARMS data, leading to some very important findings. Tell us what was exciting to you about these data? What was the real-world effect of your research?

Dr. Jablonsky

Yeah, thanks for asking me that question. A lot of the work I do looks at underserved or underrepresented producers. So I do work with beginning farmers and ranchers or farmers and ranchers that are selling through local food markets, or small and mid scale farmers and ranchers. And the ARMS started asking questions about participation in local food markets in 2008. So we can do a lot of really important work, for example, looking at how these differentiated markets actually impact producer profitability. For example, do farmers selling at farmer's markets actually help to support the next generation of farmers and ranchers? And the results of that kind of work have important implications for the Farm Bill, and the policies that we might pose to support that next generation moving into the future.

Dr. Potok

Those are very interesting findings, can you elaborate a little more on how your work ultimately makes a difference for the American public?

Dr. Jablonsky

For several decades, we've had policies in place that I think have done a really good job in supporting larger scale farmers and ranchers. And what we see is that increasingly, the public is interested in finding opportunities to support diversified food producers, farmers and ranchers to make sure that we're supporting rural economic development, the next generation of farmers and ranchers and small and mid scale operations. And so I think that the work that we're able to do using the ARMS data help us to answer some of those questions to ensure that we have sort of diverse, viable and resilient agricultural economies.

Dr. Potok

Those are great insights into the value of ARMS data. Based on your hands on experience, what would be helpful for other researchers to know about ARMS?

Dr. Jablonsky

The first time I used ARMS, I was actually a visiting scholar at the USDA Economic Research Service. And I think it was really helpful for me to be there in-person. This was probably around 2012. And at that time, there really wasn't a lot of documentation on using ARMS. I was really lucky to have **Anna** Gueorguieva there. And I think she was in the process of writing up the first ARMS manual. So,I highly recommend that if you're interested in using the microdata that you go and actually see if you can do something similar, maybe working alongside some folks that have a lot of experience with these data. Because there's a lot of nuances and questions, though the documentation is now better. I access the data with colleagues. And I will say it took us about a year to get all the access and permissions. But there's just so much more than we're able to do with the microdata that we can't do with the publicly available data, that it has been worth the effort to do it. The richness of the data has just been wonderful. And there are things you have to do every two years as part of a renewal process, and the costs aren't insignificant. So I would just keep those things in mind that if you're thinking about using ARMS, that it's really sort of an upfront investment. It should be something that you're really interested in doing on more of a long term basis.

Dr. Potok

That's very helpful information. Are there other datasets that would be helpful to link to the ARMS data? Or perhaps there's something you'd like to see improved about these data.

Dr. Jablonsky

We do link different datasets with the ARMS. For example, we bring in a Farm to School census variable. And we bring in data from the USDA Ag Marketing Service to really try to understand the food environment. What types of infrastructure are in the community is actually important to understanding the profitability impacts of different things that farms and ranches do. The two things that would actually be most helpful for us is: (1)since ARMS is congressionally mandated to focus on the core agricultural states, we often have small sample sizes, in places where in my case, most local and regional food sales occur other than California. And so you know, if there's a way to expand that, so we get larger sample sizes that would be helpful for the work I do. And then in addition, just like what we've been doing now with the (agricultural) census data, where we're actually looking at survivability of some of these farm operations over time based on some of those key characteristics, if we could do that same thing with the ARMS and really track those operations over time, that would enable us to answer some really important questions.

Dr. Potok

Thanks so much for taking the time from your research at Colorado State University to share these valuable insights with us Dr. Jablonsky. Check out the show notes for more information

and relevant links. Until next time, this is Nancy Potok with Show US the Data. **Season 1 Ep 3**

**Season 1 Episode 3 Dr. Chen Zhen**

Dr. Nancy Potok

Hi. I'm Nancy Potok, and you're listening to Show US the Data. Today our guest is Dr. Chen Zhen. Dr. Zhen is Georgia Athletic Association Professor in Food Choice, Obesity and Health at the University of Georgia. Our machine learning tool has identified Dr. Zhen is a top researcher using retail scanner data made available by the Economic Research Service at the US Department of Agriculture. Welcome Dr. Zhen.

Dr. Chen Zhen

Thanks for having me.

Dr. Potok

Dr. Zhen, you've done a lot of work with retail scanner data, including constructing scanner database panel price indexes. And that's led to some very important findings. Tell us what was exciting to you about this data? What was the real-world effect of your research?

Dr. Zhen

So as economists we know that when prices go up, consumers will substitute for cheaper varieties, go to discount stores or buy larger packages. In our 2019 study, we found that consumers are actually very nimble in adapting to inflation. So the average difference between the CPI and the (scanner data) index accounting for consumer substitution is about three percentage points. So this may sound small, but in the context of inflation, it’s a very meaningful difference. Just last week, the government reported that the January CPI was 7.5% higher than it was a year ago. So we could be looking at 4.5% inflation if we account for consumer substitution. Of course, we only looked at non alcoholic beverages using older data. So we should be cautious in extrapolating our results for the current situation. But economists generally agree that consumers have considerable leeway to mitigate the impact of inflation.

Dr. Potok

Yeah, that scanner data is really helpful. Can you elaborate a little bit more on how your work makes a difference for the American public?

Dr. Zhen

In addition to price indices, and inflation, I also spend a lot of time trying to understand consumer demand for nutrition, and how that demand responds to the price changes. For example, today, there are seven US cities that tax sugar-sweetened beverages at a rate of one penny per ounce or higher. What people may not realize is that it took seven years from when a soda tax was first proposed. And when Berkeley, California became the first city to tax that in 2015, what happened in the interim was a great deal of debate about the potential costs and benefits of taxing soda, with no real world data. So actually, to inform the debate at the time, researchers turned to scanner data to understand how consumers’ soda demand responded to prices. This information is then used to predict what would happen to consumption, tax revenue and obesity. In fact, nearly all studies up to 2015 relied on scanner data.

Dr. Potok

That's really interesting. Based on your hands on experience, what would be helpful for other researchers to know about these data?

Dr. Zhen

Yeah, this is a very good question. So for beginners, my younger self included, it is easy to be awed by the richness of scanner data. But it is also important not to lose sight of the research and policy question. As awesome as scanner data is, it is just a tool. Sometimes it's useful but other times, it's not. There are (consumer) goods that scanner data does not track well. For example, we know that we don't have very good coverage on sales at convenience stores and gas stations. So for items that are consumed on the go, they are less likely to be captured in scanner data. This includes most tobacco products like cigarettes and electronic cigarettes. Whenever you have a researcher embarking on a topic that could potentially involve scanner data, it's very important to learn how these data are collected, and think really hard about whether it's the best data source for the intended research.

Dr. Potok

Those are valuable insights. Are there are other datasets that you've linked to or think would be helpful to access and link to the scanner data? Or perhaps there's something you'd like to see improved about these datasets.

Dr. Zhen

You know, for the last probably six or seven years, the words Big Data have become really buzzwords in both academics and businesses. So people don't realize that the scanner data is really the grandfather of all big data as we know it. It has been existing for over 40 years. As rich as scanner data is, it does not have all the information required for many research topics. For our research, we often had to link scanner data with other datasets such as nutrition facts, label data. These are the labels that you see on a bag or the side of the food packages, and they provide basic nutrition data such as calories, sodium and added sugar. Another more difficult linkage we are attempting as part of the Coleraine competition is to link the 500 in the food code with a few in the USDA Food Code. So the USDA Food Code has information not found on product packages, such as amount of a whole grains or dark green vegetables. This information is necessary to evaluate the overall health information of the purchase. This isn’t easy -it’s an order of magnitude more difficult than the nutrition facts label, because this is a many-to-one match, so it becomes very tricky to define what are the best matches for each Season 1barcode? So for this one, I'm assisting a team, of world class, artificial intelligence, and natural language processing experts to try to be successful.

Dr. Nancy Potok

That's very exciting. I can't wait to see how this turns out. Thanks so much for taking the time from your research at the University of Georgia to share these valuable insights with us Dr. Zhen. Check out the show notes for more information and relevant links. Until next time, this is Nancy Potok with Show US the Data.

**Season 1 Ep 4 Dr. Tiffany Oliver**

Dr. Nancy Potok

Hi, I'm Nancy Potok, and you're listening to Show US the Data. Today our guest is Dr. Tiffany Oliver, Associate Professor and Chair of the Department of Biology at Spelman College. Our machine learning tool has identified Dr. Oliver as a top researcher using the Survey of Earned Doctorates and survey of Doctoral Recipients conducted by the National Center for Science and Engineering Statistics at the National Science Foundation. Welcome Dr. Oliver.

Dr. Oliver

Thank you so much. Good afternoon.

Dr. Nancy Potok

Dr. Oliver, you've done a lot of work with the Survey of Earned Doctorates, leading to some very important findings. Tell us what was exciting to you about this data? What was the real-world effect of your research?

Dr. Oliver

The Survey of Earned Doctorates is a survey that all doctoral students complete prior to graduation. And these data, like you said, they are collected by the National Center for Science and Engineering Statistics, which is supported by the National Science Foundation. And it collects information on their post graduate plans, the types of doctoral degrees that they've received, really just their educational history, demographic characteristics, etc. And so for several years, Spelman College has served as the number one undergraduate institution of origin for Black women that receive a PhD in a STEM discipline. And this means when you look across all institutions in the US, Spelman produces the most women per year that complete a PhD in a STEM discipline. Considering that we're a small liberal arts school with approximately 2,100 students, this is a huge accomplishment. But what does the journey look like for these students once they leave our HBCU (Historically Black Colleges and Universities) as they pursue their PhD in a STEM discipline? Analyzing data from the SED allows us to answer questions pertaining to the student's source of funding for graduate school, information on their student loan debt, the types of jobs that they've pursued after they received their PhD, starting salary, and even the number of dependents in their household. As Black women are sorely underrepresented amongst those that get a PhD in STEM discipline, these data are important, as they can potentially help us identify inequities and create solutions to remove them, hopefully helping us further diversify STEM disciplines.

Dr. Potok

Yeah, that's very impactful. Can you elaborate a little more on how your work makes a difference for the American public?

Dr. Oliver

Sure, I completed this work as a participant in the Leveraging Big Data and Achieve Equity Initiative, which is hosted by the Institute for Capacity Building Excellence and the Coleridge Initiative. And as a participant, I work with four other faculty members from the Atlanta University Center on my group-specific research question, exploring sex and educational debt levels by race among PhD recipients. As a geneticist and a faculty member at an HBCU and Women's College, I specifically was interested in the trajectory of black women PhD recipients. And we found that while there were more Black females than males that received a PhD in 2015, the majority of these PhDs awarded to Black females were in the area of non-STEM disciplines. And this was not the case for White or Asian females, who received more STEM PhDs than non-STEM PhDs. And as it pertains to your question of how these findings make a difference for the American public, we must ask ourselves, what underlies these disparities? How can we reduce these disparities? The answers to these questions are important, as they may help us create more equitable solutions to real world problems by diversifying the STEM workforce.

Dr. Potok

Definitely. So based on your hands-on experience, what would be helpful for other researchers to know about these data?

Dr. Oliver

Just make sure you give yourself enough lead time, when you're going through the process of getting permission (to access the data sets). If you're interested in analyzing the micro level data, it does take a little time for the paperwork to be completed. And if you're answering questions that are specifically looking at minority scholars, if you're interested in populations, specifically, the Native American population, many of those questions actually have, in my experience, quite a bit of missing data, or the numbers are really small, because they are indeed really small, which makes it really difficult because those scholars exist. And it makes it much more difficult to tell their story, using data from this resource that, you know, our country invests a lot of money in.

Dr. Potok

Thanks for those insights. Is there anything you'd like to see improved about these data?

Dr. Oliver

Questions on the race of the graduate school mentor, frequency of interactions with the mentor, some sort of assessment of the quality of the graduate’s relationship with their research mentor. And in general, other questions that assess the research climate. Having these types of questions within the SED or the SDR, maybe something on a Likert scale would be really, really helpful, right? This is a survey that's being collected as students are graduating and receiving their PhD. And just from my perspective, it could provide more information on the quality of these people's experiences. The SED would be an excellent tool to collect these data, and findings may help us better understand how to create more inclusive environments. And this may also help further diversify the discipline.

Dr. Potok

These are very important points. Dr. Oliver, thanks so much for taking the time from your research at Spelman College to share these valuable insights. Check out the show notes for more information and relevant links. Until next time, this is Nancy Potok with Show US the Data.

**Season 1 Ep 5 Dr. Janet Currie**

Dr. Nancy Potok

Hi, I'm Nancy Potok, and you’re listening to Show US the Data. Today our guest is Dr. Janet Currie, the Henry Putnam Professor of Economics and Public Affairs at Princeton University and the co director of Princeton Center for Health and Well Being. She also co directs the Program on Families and Children at the National Bureau of Economic Research. Our machine learning tool has identified Dr. Currie as a top researcher using Vital Statistics data collected by the National Center for Health Statistics at the Centers for Disease Control. Welcome, Dr. Curry.

Dr. Currie

Thank you, Nancy. It's great to be here.

Dr. Potok

Dr. Curry, you've done a lot of work with Vital Statistics data leading to some very important findings on children's health, particularly children in poverty. Tell us what was exciting to you about this data? What was the real-world effect your research had on such a critical topic?

Dr. Currie

Well, the most exciting thing about Vital Statistics data is its comprehensiveness. It includes quite a bit of information about every birth and every death in the United States. So it's one of the few sources of information that we have in the United States that can be used to track the health of every child. Over time. I've been able to use this data to show that the effects of programs such as Medicaid and WIC, the Supplemental Nutrition Program for Women, Infants and Children have had on infant health. I've also used this data to look at trends in child health, and trends in racial gaps in health over time.

Dr. Potok

WIC is such an important program. It's great how you've been able to use Vital Statistics to advance your research. Can you elaborate a little more on how your work ultimately makes a difference for the American public?

Dr. Currie

So, some of our first work using Vital Statistics data studied expansions of the Medicaid program in the 1990s that covered every pregnant woman. We showed that these expansions reduced infant mortality and increased birthweight. This work was part of the evidence base that was supporting the recent expansions of the Medicaid program in the Affordable Care Act. Its subsequent research has followed up on those children who benefited from prenatal Medicaid coverage and has shown that they've grown up to be healthier adults, and that when women in those cohorts give birth to their own children, that their children are healthier as well.

Dr. Potok

That's a great example of evidence building to inform public policy. Based on your hands on experience, what would be helpful for other researchers to know about these data?

Dr. Currie

Well, one important thing for them to know is that the information that is collected changes from time to time. That the CDC puts out a standard birth certificate, and then the states tend to adopt that. So it's important to look and see that any particular variable that you're interested in is the same over some period of time. Another useful thing to keep in mind is that there is missing data. But sometimes it has a useful signal itself. So for example, there are cases in the birth records where you have information about the mother, but there's no information about the father. And if you treat that missing father category as a subcategory, you'll see that those children tend to have worse birth outcomes than other children. So there's information there.

Dr. Potok

That's very insightful advice. Are there other datasets that you think would be helpful to link to the Vital Statistics data? Or is there something you'd like to see improved about these data?

Dr. Currie

Yes, the public use version of the data does not include any geographic identifiers smaller than the state. But you can get permission to use a version of the data that has the county identifiers. And if you have the county identifiers, then you can link to data from the census, for example, about characteristics of the county, such as if it's a poor county or what the median income is in the county. It's also possible to link the individual level Vital Statistics data with individual level records from again, say the census or the Current Population Survey, within a census Research Data Center. And some researchers have done that with very interesting results. One thing that would be useful to know more about in the biostatistics data is how the concept of race is captured, especially on the death certificates. And how that matches with what people's say about their own conception of their race. Because obviously, if you have a death certificate being filled out, it's not being filled out by the respondent. But if you could match to individual data, you could see what people said their own race was at some previous point in time.

Dr. Potok

Yeah, that's a very interesting topic to pursue. Thanks so much for taking the time from your research to share these valuable insights with us Dr. Currie. Check out the show notes for more information and relevant links. Until next time, this is Nancy Potok with Show US the Data.

**Season 1 Episode 6 Dr. Julia Lane: UMETRICS and the Survey of Earned Doctorates**

Dr. Nancy Potok

Hi. I'm Nancy Potok, and you're listening to Show US the Data. Today our guest is Dr. Julia Lane, a well-known NYU economist and researcher and also the co-founder of the Coleridge initiative. Our machine learning tool has identified her as a top researcher using data from the NSF National Center for Science and Engineering statistics in the UMETRICS program. We're very excited to learn more. Welcome, Julia.

Dr. Lane

Thank you, delighted to be here.

Dr. Potok

Julia, you've done a lot of work with the Survey of Earned Doctorates. Tell us what was exciting to you about this data? That is, what was the real-world effect of your research?

Dr. Lane

It really helped answer an important question, How and why does federal research funding shape scientific research? There hasn't been good data to answer that simple question. Which means that the federal government has been flying blind about decisions about where to make those research investments. They're effectively dealing with a black box without data. I wrote a piece for Nature called “Let's Make Science Metrics More Scientific”, and argued that new types of data, combined with the survey data from NCSES (National Center for Science and Engineering Statistics), could answer that question. Our paper showed how to link federal funding to the people who create ideas, who create innovation. And the basic idea was, if you could tie research funding with who is being funded, in which field, and with which results, you unpack that black box. The results showed for the first time which agencies were funding the next generation of scientists, doctoral recipients, in which research field, and what the funding was doing.

Dr. Potok

Wow, that sounds exciting. Maybe you could elaborate for us a little bit more on how your work helps the American taxpayer?

Dr. Lane

Well, the federal government spends over $160 billion a year on research and development. Congress, the President, and a recent Nobel laureate, Paul Romer, have emphasized that the importance of that investment is not just $160 billion. It also is critical to stimulate economic growth, innovation, and high wage jobs. But we didn't have the data necessary to inform how those investments should be made. I think about it in sports - we've seen how data about baseball can transform losing teams into winning teams, businesses who have data out-compete businesses who don't make use of these data. So what we're able to do now is just like in the private sector or in sports, we can help the federal government use their data more effectively, to transform the productivity of our economy. That's a big deal.

Dr. Potok

Yeah, surely. I love that Moneyball example. Tell us, based on your hands on experience, what would be helpful for other researchers to know about these data, particularly if they're considering using them in their own research?

Dr. Lane

Well, they can go to the Survey of Earned Doctorate site at the National Science Foundation. Or they can go to the Institute for Research on Innovation and Science at the University of Michigan. We ran a terrific conference on the value of science in June. And Paul Romer gave a keynote. He's the Nobel laureate. Toby Smith represents the American Association of Universities, and explained how these data enabled us to go beyond anecdotes like, well, if you spend more money on research, you get Tang, which is kind of the classic joke. And there's also a paper on building a data mosaic of the type we're discussing. I should also call out the two papers by Kaye Husbands Fealing. It's on the value of such data to understand diversity and equity and research funding.

Dr. Potok

Great. So are there other datasets beyond those in UMETRICS you think would be helpful to access or maybe to link to the Survey of Earned Doctorates data? Or, based on your hands on work, is there something you'd like to see improved about these data?

Dr. Lane

Well, you know, the wonderful thing about the new types of data that we have in the new technologies is that there are endless possibilities. One of the foci given the concern about diversity and equity is bringing in more minority serving institutions. NSF and Michigan staff are really focused on making that happen. And we're also working with the United Negro College Fund and excellence here to provide richer information about the pathways that black and brown students take as they get research funding. Building links with other users is critical. The Show US the Data Conference highlights how others have used the data so we can learn from their work using the machine learning tools that you referenced at the beginning.

Dr. Potok

Julia, thanks so much for taking the time from your research to share these valuable insights with us. More information and relevant links are in the show notes for our listeners. Until next time, I'm Nancy Potok with Show US the Data.