[00:00:00] It's time for meaningful insight, every researcher's delight. It's data night!

Kasha Ely: Hello, and welcome to Data Night with the Odum Institute. I'm your host, Kasha Ely, and joining me today is Dr. Angela Zoss, Assessment and Data Visualization Analyst at the Duke University Libraries. Later this year, Angela will be teaching a course at the annual Data Matters data science short course series, a week-long event sponsored by the National Consortium for Data Science, RENCI, the Renaissance Computing Institute of UNC Chapel Hill, and Odum. Hi Angela.

Angela Zoss: Hi, it's so great to be here.

Kasha Ely: So great to have you. At Data Matters this year, you'll be teaching Visualization and Data Science using R. Can you give us just a quick overview of what you'll be covering in your course?

Angela Zoss: Absolutely. Visualization in R is something that blends so nicely with other parts of data [00:01:00] analysis, but it's something that sometimes people have trouble picking up the Syntax for or don't quite know what the best use of visualization is in their projects. So, what we'll cover is a really deep dive into one of the most popular packages for visualization in R which is ggplot2. We spend a good amount of time making sure people understand the fundamentals of that package so that they feel really comfortable making a wide variety of visualizations all in the same environment, but static visualizations for publications are only part of the story. And so, we also spend the entire second day going through interactive visualizations, and when I say interactive, I mean things with that you can click on and then you can kind of hover over, and we do that in a variety of packages, but one of the major ones will be covering is shiny, which is a way of creating interactive applications in R

Kasha Ely: That sounds great. Why have you chosen to focus on R specifically?

Angela Zoss: I really like R in general for data science projects. R and [00:02:00] Python share a lot of similar features, but I think R's basis in the statistics community gives it a lot of very tailored applications and packages for data science, and so I find that it's a really nice place to start and the tidy verse especially which is built on top of R gives people a lot of extremely useful functions for data science. And because of that, I think that people learning data science in R actually have a great foundation to apply to other kinds of data science environments, like Python. So, it's one of my favorite tools. I use it a lot. I really like R studio IDE, which is a way of building your R code in a really nice development environment. So, it just feels like a really nice ecosystem for data science work.

Kasha Ely: How do you think data visualization has changed in recent years?

Angela Zoss: Yeah, data visualization has really seen kind of a meteoric rise and public awareness, [00:03:00] I think, over the last - I would say probably 20 years, it's definitely been gaining steam. I think part of that is due to advances in computing. Of course, I think, most people's entry to data visualization is usually through just Microsoft Excel. We see visualizations starting from a very young age now in education, and, and it's become very, very popular and very global. And I think in recent years, we've seen these kinds of ebbs and flows in what we think is a best practice for different kinds of visualizations, how complex we feel like visualizations need to be, how artistic? As visualization becomes a specialty of its own right, we just are starting to see these different pockets of people treating visualizations very differently. Everything from art and artistic kinds of data visualizations to highly scientific and specialized, and then kind of in the middle, these popularized [00:04:01] journalistic, data visualizations, that accompany big news pieces in things like the New York Times or the Washington Post. It's just, we've got such a nice, broad range of visualizations available to us now, both in terms of tools and in terms of examples that we can follow. I think that's been the biggest change maybe in recent years, although I will say on the academic front, maybe the biggest change. I've noticed in recent years is people actually studying how humans perceive visualizations and trying to base our tools and our best practices on human perception and psychology.

Kasha Ely: Can you tell me a little bit more about that? That sounds super interesting.

Angela Zoss: Yeah I think with a lot of different kinds of academic disciplines, you see this interesting historical pattern where the people who advance the field at the beginning are may be operating a little bit more on instinct and their experience [00:05:01] that on science. So, I'm not an expert in other fields, but the I think there are, there are examples that I've heard, at least anecdotally, that sometimes, the biggest names in the beginning are people who have pioneered some theories that then maybe later turned out not to be precisely right, but were based on very good intuitions, and I've seen visualization go through that process. So, in the beginning, visualization as a field maybe really coalesced around some people who had very good artistic instincts and who were really designers and loved visualizations to be pretty and to have certain kinds of aesthetics, and as we've gained an influx of researchers from different fields, we've actually seen people testing different kinds of styles of visualizations to see if they're really effective, to see if they're really well suited for the kinds of tasks that they're being put to, and [00:06:01] to improve on methods based on the results of those kind of psychological studies. So, I think we're doing a better job now of understanding people's visual systems, people's cognitive processing around visual stimuli and trying to build that now, into our into our tools and our design guidelines.

Kasha Ely: Will you be adding this context into the Data Matters course?

Angela Zoss: Absolutely. One of my favorite things to do as I'm teaching a technology, is to contextualize that technology, is to make sure that the computer doesn't own our work, we own our work, and any visualization that you create in R or in any other tool can be customized it. We don't have to stick with what the default layout of a thing is, the default color scheme, the default font. All of this is very adjustable. So, we'll be talking exactly about why we would choose to make some adjustments to our visualizations and how we would [00:07:01] do that and why R is still a really good environment for that and for being able to customize the visualization, but then potentially also what would happen after if you find that there's only so far you can go in your code. What else could you do with your visualization? So, improve this style, so sometimes you'll hit a barrier in R where there's maybe a lot of improvements you can make in R and customize your visualization, but there might be, there might be a point at which you want to improve the style outside of R, and we can talk about when that might happen and the different ways that you might be able to continue to improve your style outside of the R environment.

Kasha Ely: That sounds awesome. Thank you. Are there any current trends in the data visualization world that you're interested in or keeping an eye on?

Angela Zoss: One of my, my hopes for the visualization world seems to be coming true. And that's that our tools will become sophisticated enough to help [00:08:01] people produce visualizations that also perform well on accessibility. So, in web design, interface design, there are more mature tools for making sure that designs work well, even for people with some vision deficiency, and visualization tools in general have not had good success in that area so far, but especially for web-based visualizations, we're seeing increasing accessibility features and then we're also starting to see people develop protocols, and kind of guidelines for best practices for accessibility, even for visualizations that turn up looking just like an image in a paper. So, I think as the community really builds expertise in this area, we're going to have a much better way of accommodating people with different kinds of vision in visualizations, and that's going to be amazing.

Kasha Ely: Thank you so much for [00:09:01] bringing up this topic. It's so important, especially in this field. Can you tell us a little bit about some of those features that are going to help improve the accessibility of specifically data visualizations?

Angela Zoss: Yeah, accessibility is a very, very big topic. And I've been really encouraged that the discussion on visual, on accessibility in visualization and other interfaces is now moving beyond, just complete assumption of people whose visual deficiency makes it basically difficult or impossible to see the image. That's not the only kind of accessibility we need to design for, that we want to, you know, a wide variety of ways. So, the traditional I think accessibility expectation is that someone using a screen reader would be able to get the same experience of that visualization as someone not using a screen reader, but that's only the very tip of the iceberg. There are certainly, and there are certainly people using screen readers for a [00:10:01] variety of reasons, and the more, the more we support that, the better of course. The other types of accessibility modifications that I could see being important for visualizations include things like color contrast, which is important for people with some low vision, and one other interesting component of visualization that I think is under-explored is the visual complexity of a visualization actually has some accessibility implications. So, my understanding is that there are a variety of types of ways people can be influenced by visual complexity, and there are people for whom highly complex charts might be especially difficult to process, people for whom highly complex charts might be a little overstimulating. So, I think there's a lot of areas in accessibility to be explored by the visualization community, and [00:11:01] I think training visualization producers is one way to accomplish gains in this field, but I think by and large the biggest advances come when we develop better standards like the web community has, or we develop better tools - things that can take a visualization and improve the accessibility or tools where, when you're creating the visualization, accessibility is built in. We want to do it by design and not as an afterthought.

Kasha Ely: Absolutely, that's a great Insight. Will you be discussing some of the best practices for accessibility in the class?

Angela Zoss: Yes, I think that's a wonderful topic to include in any kind of training on visualization, and the nice thing about being able to take two full days to focus on how to incorporate visualization into your data science work, is that we can get into some of these very specific topics of where do accessibility concerns show up in different parts of visualization, and how can we accomplish good accessibility in our [00:12:01] in the tools that we have right now?

Kasha Ely: Definitely! That sounds great. When you see data visualization in popular media, what do you notice first?

Angela Zoss: Sure. Well, visualizations in popular media vary quite a bit. It's everything from a map, which is something that people use every day, even just for reference to get from one place to another, maybe not as much during our lockdown, but hopefully again sometime, and then also all the way up to very custom, very specially designed visualizations that pack a real punch both in terms of aesthetic beauty, but also, information density, you know, things that you really have to dig into and that take a little while to understand that, in themselves, offer this kind of exploratory environment. So, I think the things that - I don't know what I notice first. I think that's going to depend a lot on the visualization and what it's trying to get my attention with, but I think what I look for is the argument that a visualization is making. [00:13:01] I think there are choices that every designer makes in a visualization that you choose what to include and what not to include; you choose how to convert a data set into something visual; you choose what data set that you use to make a particular argument or to explore a particular issue. So, I try to take a step back when I'm looking at visualizations and think about the argument that this visualization is either intentionally or unintentionally presenting to me, and so, I like to think of myself as kind of a critical observer, and I can interrogate this visualization as well as use it. It's a tool for me, but I can also challenge the assumptions based in the visualization, and I can make decisions myself about whether I think they made choices that distort the data or choices that support the data.

Kasha Ely: Interesting. Thank you. Is there anything that you see in visualizations a lot? Like, a common thing [00:14:01] that just kind of nails on a chalkboard, kind of thing.

Angela Zoss: Things that our pet peeves for me in visualizations are things that could be slightly better if just a little bit of extra work had been applied or a little extra decision making. So, one of the things that I tell people a lot is that the very simplest thing you can do for some of the very simplest charts is to make a choice about the arrangement of your elements. So, for example, in a bar chart, if you're showing a bar chart, and you have a different bar for every university that you're studying or every field site where you've collected data, and if there's no internal order to those different bars then to me it seems like a no-brainer [00:15:01] to have the bars go from biggest to smallest in your chart, and what that does is that it gives you some things for free. It gives you the highest value for free, gives you the lowest value for free, gives you a ranking of all of the different categories that you're showing in your bar chart. So, there are little things that you can do that are just a single step in whatever tool you're using. You can do this in any tool that creates a bar chart, and it's just that much easier to see, to understand patterns, then if you were to just leave it in alphabetical order or in the order that the data appeared in. So, I always get a little frustrated if it feels like people haven't thought through what arrangement of their data makes sense for this particular data set or what color palette makes sense. Again, accessibility concerns are always frustrating to me, but I think there are little design [00:16:01] choices that everybody can make it very short amounts of time that would be great improvements.

Kasha Ely: Thank you. In your opinion, who should take your course?

Angela Zoss: I like this course for a variety of audiences. I think with a with a long course, you know, you try to make sure that it covers everything you need to know for the kind of core level, but it also doesn't leave anyone out to dry, and so when you have a diverse audience it's hard to do that, but what I have found in my past audiences is that some people come having created visualizations in R, but feeling less confident in their style, in the choices they make around the style visualization. So, I love for those people to come because not only are they going to potentially learn a few new tools for creating visualization, but they're going to learn those concepts, those design guidelines, that really help improve visualizations. Then we also have people who come in with very little experience creating visualizations [00:17:01] in R, and it is hard to get off the ground with certain package that create really beautiful visualizations. They can be a bit opaque, and so having someone there to answer questions immediately, to explain the error messages that are happening, that can really get people past this early hump of, I just want to make this visualization, and I've been searching stack overflow, and I can't get the answer, and I don't understand enough about the fundamentals here to be able to solve this problem. So, I love being able to help people get over that initial hump of how to create a basic visualization. So, I think it's especially nice for those kinds of groups. People who want introductory visualization and people who want introductory design, but then it's also - I think there's a significant portion of the population you come in for the shiny content, and shiny is one of the most complex [00:18:01] parts of R. It's this way of creating an interactive web-based application, using R code. And even though it's written in R, the code you write in shiny is very different from the code you write in a normal R script. And so, I think people who know that they want to get to a certain place with their data science work, know that they want to be communicating to the public, know that they want to be building things like dashboards or shiny apps. It's especially good for the, for people who, who are ready to be building visual content.

Kasha Ely: Nice. So, it sounds like this is just a great course for a whole range of people. So, that's awesome. I'm curious, I think I saw in your bio that you started in communications and moved to information science. Can you tell me what drew you to data vis?

Angela Zoss: Absolutely. So, that's true. I did. I [00:19:01] started with a master's in communication. I was doing my master's in a department that had very strong ties to an information science program, and in particular there is the sweet spot between information science and communication that I really enjoyed, which is human-computer interaction, and the idea of learning about how humans use technology both to communicate and to do work. That's always been really fascinating to me. When I started my doctoral work, I was looking for a place where I would be able to explore these issues a little bit more, and in my program, I found an advisor who was not just interested in the human computer interaction research, but also who was actively engaged in both teaching designing and building tools to support visualization, and so, I became very entrenched [00:20:01] in her lab and her work, and I really found a great passion for it. I think creating visuals to communicate complex data is one of the most exciting things you can do with data. I'm a little biased, but I just think visualization ends up being such a wonderful blend of artistic style and psychological principles and technology work. It just really hit home for me and brought together a lot of the skills that I was really interested in.

Kasha Ely: Awesome. It's always so great talking to people who are so passionate about what they do. So, you've been teaching at Data Matters since 2015. What would you tell people who might be on the fence about attending?

Angela Zoss: I think Data Matters has been one of the highlights of my year as a teacher because the students who come to my courses in Data Matters always come having finally been given the space to delve deeply into [00:21:01] some of these tools and techniques that maybe they're expected to pick up on the fly, and really data science is something that's getting incorporated into a lot of people's positions in small or large parts, and the ability to retrain in data science is very difficult on the fly. It's very difficult to get the space in your job or in your program to learn the fundamentals, and what Data Matters gives people is a really wonderful opportunity to learn from other people in the course, to learn from an instructor, to have courses that build together to create a series of skills that will work nicely with whatever part of data science you're interested in, and especially now that we're able to offer things virtually, I think Data Matters offers so many opportunities for people to really feel like they have a good grounding in data science without having to sign up for a semester-long course [00:22:01] or put their life on hold for a long time. These one week or two day or one day courses, really get people moving in a lot of different areas, but also allow people to transition quickly from the course material to their own projects, and so being able to take what you learned that week and apply it very quickly to your own data, that's the best way to learn, and that's the best way to reinforce those concepts.

Kasha Ely: Well, I'm sold. Sounds amazing. Thank you so much for joining us. This has been really great.

Angela Zoss: Yeah, it's been wonderful to be here, and thanks so much.

Kasha Ely: Definitely. And to everyone listening, thank you for joining us and keep an eye out for more information about Data Matters this year. Until next time, stay safe and well.