Speaker 1:
This is *ACM Bytecast*, a podcast series from the Association for Computing Machinery, the world's largest educational and scientific computing society. We talk to researchers, practitioners and innovators who are at the intersection of computing research and practice. They share their experiences, the lessons they've learned and their own visions for the future of computing.

Speaker 1:
Today's guest is a master storyteller. He's the force behind transforming your spreadsheets of data into works of art that speak to you. His work in data visualization is considered pioneering. Jeff Heer is a professor of computer science and engineering at the University of Washington where he directs the interactive data lab, and he is also chief experience officer at Trifacta, a company he co-founded. He has won several awards and has been on the MIT Technology Review's Innovators Under 35 list. Jeff, welcome to *ACM Bytecast*.

Jeff Heer:
Thank you for having me.

Speaker 1:
I'd like to lead with a simple question that I ask all of my guests. If you could please introduce yourself and talk about what you currently do and give us a little bit of insight into what drew you into this field of work.

Jeff Heer:
Sure. So as you mentioned, I'm a professor at the University of Washington and then also an entrepreneur, having founded a company. And generally I'm interested in how we help people make sense of and more successfully work with data. And so I draw on a background in human computer interaction, data visualization and a larger field of interactive data analysis.

Jeff Heer:
And so we both run experiments and studies to understand how people work with and perceive data and then build various systems. So that has included visualization tools such as D3, it's written by my former student Mike Bostock and the Vega and Vega-Lite tools for data visualization as well, among with a number of other systems.

Jeff Heer:
Let's see, well, in terms of background, thinking it way back, I was intrigued by computers as a child originally playing games and just messing around with applications on my parents'

Commodore and later Amiga computers. And I was interested enough that when I applied to colleges, I selected computing as a primary interest. And so I was fortunate enough to be accepted into the EECS major, that's electrical engineering and computer science at UC Berkeley.

Jeff Heer:
And I didn't have any programming experience prior to college. But when I started taking comp science courses, my interest really magnified. Programming was something I came to really enjoy. At the same time, I also had a longstanding interest in psychology and cognitive science, which ended up being my unofficial minor at Berkeley. And so this naturally led for me the human computer interaction as a field that ritually combines these two interests.

Jeff Heer:
So while I was at Berkeley, I took my first HCI course with Professor James Landay and during one of our discussion sections, my TA brought up the topic of information visualization, which is using interactive graphics to present and explore data. And he demoed a technique that was developed at Xerox PARC called the hyperbolic tree.

Jeff Heer:
And the idea was to take a branching hierarchy, whether it's a file system, textonomy, something else. And then position the nodes and edges of that tree within a hyperbolic geometric space, which like many tree structures actually expands exponentially. Then you can project it back into the Euclidean space to create a visualization. So after experience the elegance of this idea, but really the mesmerizing feel of whipping through large information spaces, I was hooked. And that began my journey as a data visualization researcher.

Jeff Heer:
And so then beyond my experiences at Berkeley, the biggest catalyst for my career was first an undergraduate internship I had at Xerox PARC where I got to work in the user interface group alongside a number of HCI pioneers who mixed both a computer science and psychology background. Prior to graduate school, I actually returned to PARC for a year as a member of research staff.

Jeff Heer:
And during that time I began earnestly working on visualization projects, many of which that would inform a lot of the work I would do in the years to come. And funny enough, some of those projects there even involve studies that showed why alternative browsers might be

preferable to the hyperbolic tree. So we've been arguing against the technique I first fell in love with that had captured my imagination.

Jeff Heer:
So we have to be good scientists and discard our pet projects when the evidence mounts against them. And so these experiences at PARC and at Berkeley laid the foundation for my work in graduate school and then everything beyond. And so what ultimately sustained my interest in the area of data visualization in particular was really the fun of being able to draw on such a richly interdisciplinary field.

Jeff Heer:
So that includes computer science, statistics, psychology and design, and then to be able to apply that to create new systems and techniques that demonstrably help people better understand their data. And so I've been exploring variations on that theme ever since.

Speaker 1:
Wonderful. Jeff, is that common? Do you find that most researchers who work in this field have that sort of interdisciplinary experience or interest?

Jeff Heer:
Yes, I would say so. So certainly in visualization and of course in human computer interaction more broadly, it's usually a conglomeration of people with various backgrounds, both in terms of interdisciplinary teams but also interdisciplinary people, whether it's someone with backgrounds in both art and technology or empirical studies of behavior, whether that's coming from psychology or qualitative research methods.

Jeff Heer:
The intersection of these things, I think is one of the most enjoyable and dynamic fun tensions that helps drive the field forward. I'd also note that one other aspect, certainly within my own research group, I think it really attracts people who, while they have deep intellectual academic research interests, are also very oriented to the world at large. They want to build tools or techniques that people out in the world will use and benefit from. So I find that a lot of the students who apply to my group self select because they're really excited to really build things that others can pick up and use in addition to making fundamental research contributions.

Speaker 1:

Got it. Would you say that from the time that you got into this field to where we are today, the interest as well as the demand for better visualization, the interest in data and using it to make crucial decisions has really been magnified?

Jeff Heer:
I think without a doubt. Certainly when I started, visualization already had a long history prior to computer scientists, for instance, statisticians and others, certainly mapmakers, photographers, others done really groundbreaking work in advancing visual forms of information presentation. But around the time I graduated with my PhD and switched to becoming a professor, I started as an assistant professor at Stanford was really when terms like big data and data science entered the mainstream, and so the amount of kind of general interests, business interests, et cetera, in data systems, including data visualization really took off.

Jeff Heer:
And actually kind of funny I remember I had applied to multiple faculty positions back in 2008 and the two word, and one of the places I visited but ultimately didn't go to at the time was the University of Washington. And later after a number of years at Stanford, I was back on the job market considering other places ended up interviewing at the University of Washington and one of my colleagues there looked at me and he's like, "You're back." And he's all, "Since we last spoke, the world has really turned your way."

Jeff Heer:
So he's just pointing out this stuff that maybe they thought was interesting, but they didn't know how important it was, the importance had become undeniable. And so that's, I think, been both a blessing and then also I think in some ways a curse as we as a society are still trying to figure out how to best use data productively, reliably and in a way that protects individual privacy and liberties as well.

Speaker 1:
Absolutely. I mean, you touch upon so many different aspects that I'd probably get into more during our discussion. But you mentioned big data, right? I mean, today we collect data about everything, every trip to the grocery store. I think my car is probably collecting millions of data points about my driving and the route and the air quality, et cetera. So how does volume of data really play into the field of data visualizations? Is that a challenge? Are there specific issues with representation of this data or the accuracy of it?

Jeff Heer:

Yes, absolutely. I think as data sets get larger and more diverse, it brings new and interesting challenges. Maybe not new in the sense that no one's ever thought about them before, but certainly more pressing an issue of scale. But I think I would start by actually pointing out that regardless of scale, I think some of the core things that make data visualization important are there regardless of whether you're talking about small, medium or large data, however you choose to create thresholds to define those.

Jeff Heer:
And so the goal of data visualization, one way of framing it is helping people use vision to think. So that is use our long evolved perceptual system to understand patterns, trends and outliers in data, and these use graphical displays and input techniques as a mean to develop and further investigate domain specific questions. And so the greater availability of data, which I think is an important component of this so-called big data movement just becomes more pressing, we need these abilities given that variety of data sets.

Jeff Heer:
And so I think that given the number of times that graphic has influenced a business decision, it's probably too numerous to count. I mean, like how many millions of Excel charts have been included in presentations to decision-makers? I mean, your visualization is I think quite important and quite prolific. But as a result of this, and this relates to big data as well, I think my favorite examples tend to involve cases where data visualization helps us identify questions that help decision makers be more skeptical.

Jeff Heer:
It's not just about taking this book of data and using it to drive decisions. I think we really need to think about what the data can and cannot support. So just as one example, some years ago, I once tried to export my social graph from Facebook and I visualize it in various ways. And initially it all look quite reasonable to me, I could identify recognizable clusters of friends and so forth.

Jeff Heer:
But then I digitalize it in one particular way as an adjacency matrix with rows and columns ordered by join dates really showing the sort of checkerboard pattern of all the links between people. And in this particular form, I discovered that a large proportion of newcomers were missing any connections between each other, which means people who joined maybe even months ago, for some reason, had no friendship connection in this dataset.

Jeff Heer:
And that just didn't seem right. It turns out this was a missing data problem where a silent query limit was enforcement left out up to 20% of the edges in the graph. So any conclusions or models built from this data would have suffered from a garbage in garbage out problem. And so I think one thing that comes up with big data is actually even harder data quality challenges.

Jeff Heer:
So given that list, so look at some of the different ways big data affects analysis and particularly visualization. So we can roughly group it into three kinds of challenges. So one would be that we have lots of records. So as I've imagined, we have a database table, this table has lots of rows, maybe millions, billions, or more rows.

Jeff Heer:
And this is an interesting scalability problem. It's one that the database field and the visualization field have both worked on in recent years. But we can largely handle this, not that there aren't challenges that remained. But using scalable database systems and smart indexing approaches, you can actually approach kind of real-time interactive querying over summaries of billions of records of data.

Jeff Heer:
So that's something the field I think has successfully contributed in recent years. A more difficult challenge is big data in the sense of having lots of attributes. If you think about it as a data table having thousands or hundreds of thousands or more columns, and this is harder in just that the algorithms involved are much more expensive, harder to get quick results for analysis.

Jeff Heer:
And the right thing to do is often task dependent. Do I need to include all of these variables in my analysis? Can I combine them in some way? Can I leave some out? It's just one example. There's lots of work on dimensionality reduction algorithms that try and take high dimensional data and then project it down to 2D or 3D representations in ways that still preserve important aspects of the structure of the data in order for us to visualize it.

Jeff Heer:
So still lots of work there. And this also brushes up against visualization techniques for machine learning, which I think we'll talk about later in the interview. And then the third challenge is the great variety of data sources. So big data, isn't just, oh, I have lots of data that all has a

homogeneous structure, but rather it's many different types of observations from different devices, maybe from different people, collected for different reasons.

Jeff Heer:
And so if I want to meaningfully combine those data, because I think it would be useful for my analysis, this requires significant effort just transforming that data to clean it up and combine it in an accurate way. And I think this also raises potential pitfall or danger with respect to opportunistic analysis and that many datasets may have been designed and collected for a specific purpose and then people want to analyze that data for purposes for which it was never originally intended.

Jeff Heer:
And so to do that without fooling ourselves or running a foul of any kind of maybe invisible data quality issues, we really need to know how that data was collected, what might have been omitted, overlooked, what biases might be underlying the collection of this data. And so if we want to be able to draw causal conclusions, for example, we need to know about a control for confounders.

Jeff Heer:
And I think that's actually one of the larger challenges of so-called big data and that it's often opportunistic in nature and that it wasn't a scientific collection exercise that was designed for a specific analysis, rather we're trying to repurpose data and we have to be very careful about how we go about doing that.

Speaker 1:
I think it is excellent points across the board. But one question I had was as researchers or as practitioners, what can we do to protect from these kind of sort of pitfalls, right? So for example, are there tools that we build that can actually have some intelligence that will help us steer away from making these sort of huge glaring errors? I mean, data cleanliness of course, is a huge issue, but what other things are you maybe thinking about and say, hey, I'm going to build these stools, which will really help protect against some of these issues.

Jeff Heer:
It's an interesting question. It's one that the information visualization community is actively thinking about these days. So just to give you a couple of examples, in our own work, we often think about perceptual models of visualization. So when someone looks at a visualization, what

do they take away from it? Oftentimes that's operationalized terms of what value comparisons or patterns can you rapidly and accurately decode seeing the graphic.

Jeff Heer:
And so we've built for example, digitalization recommender systems that allow people to communicate the data that they're interested in. And then the system might help produce visualizations that on average for a large number of people are maybe more likely to be interpreted accurately. So that's one kind of more intelligent control where it's not just let me give you unbridled power to create visualizations.

Jeff Heer:
Maybe the tools starts to bring a bit of their own design opinion, ideally backed by perceptual science and helping guide you through the space of choices. But that's already at the point where you've already committed to visualizing particular data variables, perhaps transformed a certain way. And higher level problems in the space is the process of analysis itself.

Jeff Heer:
So for example, there were researchers asking questions like how do we identify and potentially mitigate cognitive biases that came up in people's exploration processes, whether that's is there appropriate coverage of the data or are people overlooking potential relationships of interests? And so to do this some of the strategies people are thinking about is, well, can I look not just at the data itself and potential statistical analysis of that data, how do I log the various interactions that people are doing so I learn something about people's processes and what they have and haven't looked at?

Jeff Heer:
Building a model of the user, not just of the data in a way to try and make recommendations. And I think some of the most exciting ideas in this space are being currently looked at. I know, for example, my colleague, Jessica Holman who's now a professor at Northwestern has been thinking deeply about these issues as well, in terms of, well, how do we safeguard processes of exploratory data analysis?

Jeff Heer:
I might look at hundreds of different charts and think I come away with some interesting findings, but when you're looking at that many things, the odds that you'll see something spurious just due to chance goes up and up and up and up. And so are there ways to bring ideas

from statistical methodology more directly into some of these interactive exploration tools and help safeguard against false discovery.

Speaker 1:
Got it. And I think, I mean, when you talk about biases, I mean, I'd like to go back to our previous conversation you were talking about having some sort of ML in use of ML and data visualization. What is the scene in that space? How is it that you're using artificial intelligence in your research?

Jeff Heer:
Sure. Well, like almost every other part of computer science ML was making a splash in the database community. And so I know at least three areas that have recently gained prominence in our field. And so one hot topic these days, and I think both in visualization and beyond is the use of a visualization interaction to aid machine learning interpretability, right?

Jeff Heer:
So either you're understanding why a model make the predictions it does, how well can we compare and predict model behavior? I think this is particularly critical for production deployment situations where you're going to put out a model that might be affecting hundreds, if not millions or more people, and these aspects of quality control and testing of models, extremely important.

Jeff Heer:
And so within the visualization field in particular, I think initially it started with important but relatively straightforward representations, like showing the architecture and activation pattern within a neural network which is sort of a direct representation, but one that was often hard to interpret what that really meant in terms of the output. And that's also included high dimensional visualization techniques I alluded to earlier.

Jeff Heer:
So for example, a lot of machine learning techniques learn a latent vector space where points in that space, whether they're words or images or whatever else they represent, their similarity or distance within that high dimensional geometric space has important meaning, right? So more synonymous words might be related or things such as genderization or parallel structural boards might be reflected as linear structures in this space.

Jeff Heer:

And so visualization of these vector spaces has been another kind of common, popular approach to trying to make sense of what are the internal representations that ML techniques are learning. This is something that we've worked on in our group, for example, trying to map human, meaningful concepts within the space and then seeing how they compare. And then also seeing how they change, for example, if you change the number of dimensions that algorithm's allowed to use for learning these spaces, how does it affect the representations that are then formed?

Jeff Heer:
Of course, many, many others are looking at these techniques as well. And so these efforts are pushing forward, really trying to help better elucidate why ML techniques make the predictions they do ideally in ways that people can begin to understand. I'd say there's still a really long way to go but I'm hopeful these efforts will ultimately influence how we go about designing and engineering ML systems.

Jeff Heer:
Not just slapping a visualization on them after the fact, but understanding how these tools can be part of a design process for end user facing systems that use ML to really, again, improve the quality control and help reduce bias and many of the other problems we've seen ML systems are deployed at scale and make mistakes. So that's one area, important one.

Jeff Heer:
Another area where machine learning is being picked up is using machine learning as a method to help generate visualizations. So this includes a wide variety of projects, including visualization recommender systems, so using ML techniques to help reason through the design space of visualizations and then help recommend potential charts given a task. That includes natural language interfaces to be able to describe either data features or correlations or tasks you'd like to see and then having the system create a corresponding representation that's responsive to that query.

Jeff Heer:
And then of course, the underlying use of ML algorithms similar to classical data mining techniques to try detect and highlight patterns of potential interests within a dataset. And that might be pre-trained ML algorithms or mixed initiative systems where people are providing examples to train up a system rapidly for the types of patterns they want to discover, then letting the algorithm loose on the data and then kind of visualizing the corresponding highlights that come back.

Jeff Heer:
And then the third area, which I find pretty fun to think about is so using ML and particularly computer vision methods to try and automatically interpret charts or to try and simulate in some way, human perception. So just one example, my former postdoc Jorge Poco led a project on reverse engineering visualizations. And this takes a bitmap image as input and then tries to produce this output, the actual visualization program that could produce that image.

Jeff Heer:
So if visualization is typically concerned with taking data and going to image, the corresponding inverse problem and a borrowing from [inaudible 00:21:43] is to start with that visualization image and can the computer tell you back what were the encodings applied, perhaps even begin to recover some of the underlying data that's been visualized. And then lastly, others in the same topic have started looking at using ML techniques to partially evaluate the effectiveness of visualizations.

Jeff Heer:
And I think this by and large is still work very much in an early stage, but you can imagine potentially trying to run user studies, but on a neural network rather than actual people as ways to maybe initially test some different visualization ideas, maybe in terms of low level perception before them moving onto more costly human subject experiments.

Speaker 1:
Got it. I mean, I think so many areas that you touched upon certainly seems like an extremely rich space to work in. But I'd love to go back to the point that you were talking about about interactive data analysis. I know that's an area of interest as well as deep work that you've spent time on. What is it that drove your particular interest in this area, Jeff? And was the area of work even popular when you started?

Jeff Heer:
Yeah. I mean, so broadly speaking, the task of interactive data analysis has been around for centuries because any data analysis requires human intervention, whether it's in the design of the data to collect in terms of what questions you're trying to answer, the choice of models, choice of graphics, et cetera. Though obviously in a modern context, the rise of computers brought a fundamental sea change to how we go about interactive data analysis and what we can do.

Jeff Heer:

And for me, me being interested in that topic it was a natural expanding of my research horizons over time. And so I started off with this deep interest in data visualization. In fact, when I started at Stanford, my group was called the Stanford Visualization Group, but as our interests expanded and I moved to UDaB, I even renamed as the interactive data lab, because what we did, while largely focused on visualization, kind of grew beyond just visualization itself.

Jeff Heer:
So just working on visualizations alone, one can't help but run headlong into many other challenges that you associate with data analysis. So one notable among these is data wrangling, which is of course the process of cleaning, preparing and profiling a dataset in order to understand its shape and structure and then make it actionable for further downstream analysis. For example, trying to create better tools that make it easier for anyone to properly clean, format and prepare their data.

Jeff Heer:
I was the PhD topic of my former student, Sean Kandel, we met along with Joe Hellerstein, Sean and I we founded the company Trifacta, which we're commercializing Sean's thesis. And so that was just kind of one step, from visualization to considering data wrangling and data transformation. And of course there are many other steps in an analysis pipeline. And so more recently we've become quite interested in considering this larger lifecycle of end to end data analysis and in particular, what are the myriad decisions that people make throughout this process of analysis?

Jeff Heer:
So just as one example of work in this space, my student Yong Liu, and we're in collaboration with Alex Kadell and Tim Althoff, we've been investigating ways to account for all of these different decisions throughout an analysis process, something that many other researchers are also quite interested in, particularly in the face of replication crises and a variety of scientific disciplines.

Jeff Heer:
And so in our case, we ran interview studies with a variety of analysts to better understand how they do and don't make decisions, how they choose to include certain results and not others and their research reports. And based on that, we developed recently a tool called Boba for authoring and visualizing what are called multiverse analyses.

Jeff Heer:

So rather than evaluate just a single set of analytic decisions, such as one choice of variables, one way to handle outliers, one choice of model machinery, instead multiverse analyses seek to enumerate what are all the all priority reasonable specifications. So what are all the decisions that given some theoretical and methodological background, you're not sure how to choose among them, but all seem like valid analysis decisions, actually specify them all and then evaluate them all in parallel.

Jeff Heer:
And so in our tool aids with specifying this combinatorial space of decisions and then visualizing and performing inference on the results. So for example, I'm not just seeing what all the different effect sizes might've been coming out of that, but also helping to identify which of these analytic decisions like choice of a covariate, choice of a particular model, machinery or parameter choice, to which of these decisions are the final results most sensitive? And so the goal of this and some other recent projects in my lab is sort of promote more comprehensive, robust, and transparent analysis results.

Speaker 1:
Got it. You mentioned Trifacta. And I know that you wear two primary hats, maybe three, you're a researcher, you're a teacher and you're an entrepreneur. How do those worlds blend? How did you decide from being more in the academic space to jump headlong into industry and start a company?

Jeff Heer:
So it happened rather naturally, or so I think. So I was university professor, pre-tenure, so working quite hard and focused on initially building up the research group, having my students become successful, et cetera. But along the way an important component of that is not just publishing papers, but we've produced a lot of open source software.

Jeff Heer:
So I'd say one of my other hats among the others you listed is being an open source developer and maintainer, and that's true to this day. But in particular, when it came to Sean's thesis, our work on a system called Data Wrangler, this had interesting algorithmic components, but it was primarily an interactive UI.

Jeff Heer:
I mean, unlike some of our other projects which resulted in new toolkits or APIs, this felt like the kind of project that really would benefit from having the infrastructure of a company behind it,

both to make it robust, releasable, the training, et cetera, it just seemed like a better fit in terms of bringing that technology to the broader world.

Jeff Heer:
It also aligned with Sean's interests. So Sean was finishing his PhD and he was oriented towards industry as opposed to academia, and our collaborator, Joe Hellerstein was also keen to start a company. And so they kind of convinced me that not just being an advisor, but to really jump in with both feet with them and help get the company up and running.

Jeff Heer:
And so I can certainly tell you it was quite exhausting having these multiple hats to wear, particularly with the start of the company. One thing that ended up working well for me was that from the beginning sort of my part-time status was baked in as part of the DNA of the company. And so I was working quite hard for both the university and the company, but it was also understood from the beginning that there were certain limits.

Jeff Heer:
And so as any entrepreneur starting out, you make many different mistakes, but I think one thing that we got right as a set of co-founders was really clearly laying out our expectations of each other from day one and being really well synced. And so that then as both good times and bad times come upon us in the weeks and months and years to follow, we had that shared understanding. And I think that allowed us to continue to be great collaborators throughout the years.

Speaker 1:
Excellent. And do you find that... so I know a lot of people actually find that the blend between academia and industry can be quite challenging. I mean, there is intent, of course. Everybody understands the value with those sorts of interactions, but it's not always easy to put into practice. What do you find are maybe techniques that help you to just ride along both these worlds in a way that is effective? And do you find that one sort of helps fuel the other part of your work?

Jeff Heer:
Yes, absolutely. So, I mean, I think it fuels the work in more than one way, interesting. So I think when I first founded the company with Joe and Sean, it was actually really refreshing to just have a change of context and then a change of expectation, the type of work that you have to do, right? So in many projects, you do a proof of concept and research, you make a prototype, you show that it works, you evaluate it, you publish the results.

Jeff Heer:
You're proud of what you do, but it's not necessarily production ready for real-world use. And there's also sort of a bias in terms of novelty, sometimes in terms of cleverness, et cetera, as opposed to simplicity or effectiveness to solve the core problem. I mean, obviously for it to be research, it's not enough for it to be useful, it has to be new. And I think that that makes sense.

Jeff Heer:
But when you switch to the industrial context, you can be really focused on what works. Whether or not it's new, that creates new pressures, of course, in terms of things being usable, correct, not too buggy, et cetera. But I found that the change of emphasis and thought was really refreshing and surprising. You work long enough in one strain of work you start to tire that as well.

Jeff Heer:
So I also then, when I switch back to more focus on academic work, I was sort of newly reinvigorated and refreshed by that because it was this change. So just psychologically having these different realms in which to do this work, I found particularly rewarding. And of course the content fueled different insights as well.

Jeff Heer:
So this was certainly true even before we started the company, just being at Stanford, having a lot of exposure to Silicon Valley and what people in industry were thinking helped me navigate choices of research projects and that there were many projects that might seem interesting, but which ones seemed like they might be more impactful, were meeting a need that was being demonstrated out in the real world.

Jeff Heer:
And then going into the company sort of feet first only further reinforces that. So I do think there is at least a high potential for a virtuous cycle between research and practice. It's useful for them not to be too tightly coupled, but I think the important flow of information between the two is vital to both of them succeeding. And certainly at an individual level I find that more invigorating and interesting and a great source for new ideas on what to work on next.

Speaker 1:
And does it require an inordinate amount of strength? It sounds like a lot of work.

Jeff Heer:

It's certainly a lot of work, but I think it's also... it's about finding joy in the work. And so, I mean, not that you have the expectations that everything is going to be fun all the time, far from it. But nevertheless, the intellectual challenge and the interestingness of what you're working on, but just as importantly, the people you're working on it with, whether that's been my students and collaborators in research, my co-founders and collaborators in the company.

Jeff Heer:
Having those environments that you really enjoy being a part of, I think is part of what allows you to go, you're trying to tackle these problems with the requisite amount of energy. For what it's worth though, I also always want to get at least eight hours of sleep if not more per night. And so I think having a healthy and balanced life actually helps you do these things more effectively too.

Speaker 1:
Awesome. Thank you. What would you say would be your advice to somebody maybe early in career Jeff, if they wanted to get into this field? I mean, a lot of us end up going into college and thinking about computer science as an overall sort of major that we want to focus on. But how would you say that at what point does the decision say, hey, I want to get into this field of work, into data visualization, what is it that I need to identify, maybe an ability in myself to say I think I might be good at this?

Jeff Heer:
One thing I'd recommend is if you have that opportunity, take a variety of courses. So for me I really enjoyed my undergraduate years taking a variety of courses in computer science, right? So everything from data visualization, they didn't have a class, but it was part of human computer interaction. So I was exposed to it there, but also in classes on AI, computer vision, database systems, any of these topics can be quite interesting.

Jeff Heer:
But also taking classes outside your major. So, as I mentioned earlier, the various cognitive science courses I took were some of my favorites. When I went to graduate school, actually I took half of my coursework, not in the CS department, which was great, but also at the school of information at Berkeley, where I got exposed to many other approaches and methodologies that have been really informative and influential for me.

Jeff Heer:

So the first is just having I think a broad based education that allows you to figure out what you're passionate about, but also allows you to build an intellectual foundation that's broad enough that even if it's not until years later, all of a sudden there will be just that amazing connection between things that you didn't realize at the time, but then it becomes central to some later projects. So that's one.

Jeff Heer:
And the case of data visualization specifically, should your interests draw you that way, there's a really rich online resources and community. So on social media and Twitter and others, there's many active data visualization researchers and practitioners sharing ideas, sharing designs, sharing process. There's lots of great visualization tools out there, including many tutorials and free curriculum out there.

Jeff Heer:
So lots of things to not only become acquainted with the field, but really start to build up your skills and deeper understanding. And from that point, I recommend just diving in, starting to do the work of data visualization. What are the questions or topics that are most important to you? Maybe starting even in your local area, what kind of government or other services data might you collect, begin to create analyses and visualizations around that?

Jeff Heer:
I think it's a discipline that really rewards simultaneously learning concepts and fundamentals, but really putting them into practice through hands-on exercises. And so that's how I'd recommend to get started. And again, I find it to be a very welcoming community. I hope that's true for everyone else as well really getting involved, whether it's in terms of the research community or the practice community. I think there's a lot of meetups, et cetera, in various cities, lots of ways to start to connect.

Speaker 1:
Excellent. I mean, I think you bring up a very good point, experimentation, starting small, trying out, using the open source tools that might be available would be a great way to at least understand what does this entail and get a flavor of the field. So thank you. Thank you for that excellent advice. I'd love to know, what do you do outside of work? What are your interests?

Jeff Heer:
Well, at the moment I'm actually not in Seattle where I teach at the university, I'm actually living in Berlin, Germany. So we came here for sabbatical last year and due to the Coronavirus

pandemic, ended up staying put for the time being. And so while the overall reasons for having to stay are obviously quite unfortunate, nevertheless, we're trying to take advantage of being in a different culture, myself, my wife and my family, we're all learning... perfecting our German and really kind of enjoying the interesting contrast between the culture here and in the United States.

Jeff Heer:
I also have two small children. So I think my number one and most time consuming hobby is spending time with them and also just sort of being fascinated at how they make sense of the world and watching them learn. I also have to admit though that, as I mentioned before, programming is something that I came to really enjoy. So as boring as it may be, actually programming and particularly supporting a number of our open source projects is also one of my biggest hobbies. Unfortunately, not my only one, but I would remiss and I'd be quite dishonest if I didn't include it among some of the top line responses here.

Speaker 1:
Got it. Any chances of Oktoberfest happening this year in Germany?

Jeff Heer:
No. Interesting enough, it happens in September anyway, so it's come and gone. And I'm in Berlin, not Bavaria. So it's slightly different, but no, I definitely saw when the Oktoberfest beers came through the supermarket and are largely gone now unfortunately.

Speaker 1:
Well, I hope you get to enjoy all the other lovely things that Germany has to offer. We'd love to close the interview with our final byte. What is it that you are most excited about in this field of data visualization over the next five years Jeff?

Jeff Heer:
I think so much. I think there's really interesting questions that people are starting to address that are, I think, long-standing core visualization questions. So that includes not just how people decode everyday visualizations like bar charts or line charts, they're really starting to look more deeply at how we understand uncertainty and representations of uncertainty, and how do we reason about that, and corresponding perceptual studies that are continuing to get richer in terms of expanding our understanding of what people do and don't perceive in graphics.

Jeff Heer:

So that's long-standing, but I'm excited about the progress that a number of my colleagues are making the space. But I also think what's important for the field of data visualization, certainly over the next five years, if not longer, is thinking beyond visualization. So not just taking data, mapping it to an image, interacting with that image, et cetera, I think visualization alone is not enough.

Jeff Heer:
So one reason for this is that there are people who have different capabilities in terms of their sensory and physical abilities. So some people have visual impairments, other people have other issues that affect how they interface with computer systems. And I think data visualization needs to do a better job of thinking beyond the purely visual, whether that's thinking about tactics, Sonic or other modalities, or even generating meaningful text summaries of what is otherwise visual content.

Jeff Heer:
So accessibility, I think, is an important area that visualization researchers care about but I think as a community, we haven't done anywhere near enough. I think it requires a lot more attention. The other topics that I already mentioned. So I'm kind of voting with my feet in these cases. I hope to work on accessibility in the months to come, but we're also already working on end to end analysis.

Jeff Heer:
So not just looking at visualization, but looking at it as one component in this larger process of data analysis and making that larger process kind of the overarching phenomena of our study and what our tool seeks to support, whether that's better and of cognitive support for processes of exploration with data, or as I mentioned before, better tools for scaffolding and analysis and helping promote robust and transparent results.

Jeff Heer:
And then finally, we also touched upon machine learning. And so I think visualization researchers and practitioners have an important role to play in a more human centered approach to machine learning system design, evaluation, and deployment. So both helping both the creators of these systems and the users of this system have a better understanding of how they work, but most importantly, I think making sure that the systems and the deployments are accountable and responsible for the decisions and recommendations that they make.

Speaker 1:

Thank you, Jeff. This has been such a rich and detailed conversation. We really appreciate the fact that you took so much time to expand the idea of data visualization for us and our listeners. Thank you for taking the time to speak to us at *ACM Bytecast*.

Jeff Heer:
Thank you so much for having me.

Speaker 1:
ACM Bytecast is a production of the Association for Computing Machinery Practitioners Board. To learn more about ACM and its activities, visit acm.org. For more information about this and other episodes, please visit our website at acm.org/bytecast. That's acm.org/ B-Y-T-E-C-A-S-T.