Introduction

Good afternoon everyone! I bring you greetings from sunny Florida. My name is Kari Jordan and I am the Senior Director of Equity and Assessment for The Carpentries. I'm so happy to be here to address you this afternoon. The title of my keynote is Building Confidence, Achieving Equity: A Call to Action for Open Source Communities. I'd like to start by saying thank you to the csv conference organizers and in particular Danielle Robinson and John Chodacki for all of your help getting me here. I also want to recognize my Carpentries colleagues Serah Njambi Rono and Tracy Teal who are here this week. I am truly blessed to be able to work with these incredible women every day and having you in the audience makes giving this talk a little less mortifying!

Our Journey

Our journey today begins with thoughts on data analytics, coming from someone who frequently feels like an imposter in this space, and with that I'll talk a little bit about imposter syndrome. I'll then share some thoughts on averting the data skills training crisis and explain a bit about The Carpentries and what we do. Then I'll share some personal stories about the role data plays in my life, and what I've learned about building confidence to work with data. We'll talk about mentoring and meaningful equity, and finally end with a call to action. At some point I may burst into song so just be prepared for that. And, hopefully throughout the talk we'll have some fun. So, here we go!

Data Analytics: Thoughts from an Imposter

I remember the first time a colleague introduced me to someone looking to learn about assessment for data science curriculum, and in their introduction they used the word **expert**. I was totally thrown off because at the time I'd only been working in this space for 6 months, and I could **barely** import a csv into R. My research practices were horrible! I had no idea what a workflow was, I'd never even heard of the term reproducibility, and I had been storing my data in multiple formats all over the place!

If you don't believe me, I'll prove it to you. Here is a screenshot of an e-mail I received from a graduate student who wanted to replicate the protocol from my dissertation. My background is mechanical engineering, and my PhD is in engineering education. My dissertation research was an intervention to increase engineering self-efficacy for first year engineering students of color. So in this e-mail, the graduate student wanted to understand how many students were in my treatment group and how many were in my

control group. The student had been e-mailing me back and forth trying to recover my notes and spreadsheets which were nowhere to be found. This was more than two years after I defended and had since completed a post-doc, and was working in assessment for Data Carpentry, a project that stresses the importance of reproducibility in research! Needless to say working in this space, and being introduced as an expert while at the same time being ashamed that I could not help a graduate student reproduce my study, made me feel like a total imposter.

Imposter Syndrome

In my own words, imposter syndrome is the belief that your success is illegitimate, and that at some point, you'll be found out. I was working for a project that develops and teaches lessons on the fundamental data skills needed to conduct research, but my own research practices had been horrible, and any moment now they were going to find me out!

The day I realized I wasn't an imposter in this space was the day I received this note from a colleague. It reads, "thank you for letting me help with the assessment process and driving our community forward with data." You see, what makes an expert isn't that the individual knows everything. Having comprehensive or authoritative knowledge is nothing if you aren't creating an environment where others feel comfortable contributing to the work. For this individual, I had done that. I no longer feel like an imposter. **Plus, I learned how to keep my raw data raw and I learned about version control!**

Data and Society

I'm proud to be in front of a room full of enthusiasts who are passionate about data and its application to society, because even as we sit in this room, decisions are being made on our behalf with all sorts of data. Data analytics helps identify new opportunities, that, in turn, solve problems and help us develop products and services to support the public. What I've learned is that data analytics is not about one person's expertise though. It's about collaboration and community.

From life sciences and banking to manufacturing, healthcare and government, data analytics is solving problems and saving lives, but it's also causing distress and harm because not all of the **experts**, or those with comprehensive knowledge and skills in this space, have a diversity of thought and experiences.

Let's take the United States Transportation Security Administration for example. I fly **a lot** and I cannot tell you how many times I've been pulled aside to have my hair patted down. Thick hair, locs and braids, wigs, and even turbans often trigger alarms for airport security full body scanners. Now don't get me wrong, I appreciate the work our TSA does to keep us safe, but being pulled out of the line in the airport and having someone run their fingers through your hair is extremely frustrating. It makes me wonder if the **experts** designing the technology look anything like me. Who would ever consider me an expert when so many see me as a threat?

Jargon Busting

Throughout this talk I'll refer to several concepts and would like to take the opportunity to define them here. The definitions for these terms were adapted from California State East Bay's Diversity, Leadership, and Employee Wellness site.

When I talk about **accessibility**, I'm referring to program and process design and implementation that offers multiple avenues for access and participation. In other words, **accessibility** is the usability of a product, service, environment, or facility by people with the widest range of capabilities. Four major modification categories are visual, hearing, motor, and cognitive. For example, you may be familiar with accessibility accommodations like closed captioning, or displaying the audio portion of a program as text on the screen. Others include identifying unusual words or jargon busting. Certain conditions make it difficult to understand nonliteral word usage or figurative language. I'll tell you a story about how I learned this the hard way.

A group of Carpentries instructors were putting together a series of community calls with various themes ranging from teaching your first workshop to navigating unpredictable learning environments. Well, I had the bright idea for a call about running workshops with little to no money, and I proposed that we call it *"Ballin' on a budget."* Now this title would have gone over fantastically in the Black community because "ballin" means to have wealth or affluence. However, in other areas across the globe, ballin' is extremely offense. So in our work, what can we do to make our code, our content knowledge, our documentation accessible and the language understandable to all?

Now, let's have a look at diversity. **Diversity** refers to individual differences and group-social differences that can be engaged in the service of learning. Our individual differences can be personality, language, learning preferences, and life experiences. Our group-social differences can be race, ethnicity, class, gender, gender identity,

sexual orientation, sexual identity, country of origin and ability status, as well as cultural, political, religious or other affiliations. Rather than labeling one another and seeing our differences as a threat, we can use our differences to engage in the service of learning, we can see each other as human.

When I talk about equity, it's really creating opportunities for equal access to and participation in programs that are capable of closing participation gaps in our community. For example Angus Maguire adapted this image to illustrate the difference between equality and equity. Equality is about SAMENESS, it promotes justice by giving everyone the *same thing*. But it can only work if everyone starts from the same place. In this example equality only works if everyone is the same height.

Equity is about FAIRNESS, it's about making sure people have access to the *same opportunities*. Sometimes our differences or history can create barriers to participation, so we must FIRST achieve EQUITY before we can enjoy equality.

Lastly, **inclusion** is the active, intentional, and ongoing engagement of diverse people and communities. This increases awareness, content knowledge, and empathic understanding of the ways we interact within (and change) our community. I want to draw your attention to one of my favorite quotes by Verna Myers. Diversity is being invited to the party. Inclusion is being asked to dance. Come on now people. How would you feel. You put your best foot forward, you work hard so that there's no way you don't get asked to the party, and when you get there, you're standing in the corner the whole night because nobody asked you to dance.

You know the song. *I wanna dance with somebody I wanna feel the heat with somebody. Yeah I wanna dance with somebody…with somebody who loves me.*

Inclusion is more than inviting people to the conversation or the workshop or the conference who don't look like you; it's ensuring they're able to interact and contribute in ways that are meaningful to them.

Now that we're all hopefully familiar with the terminology, let's talk about approaches to building confidence to work with data with these considerations in mind.

Training for Data Skills: Averting the Crisis

Training for data skills is more critical now than ever before. In the past decade, we've seen the creation of certification and graduate programs for data science, as well as interactive, self-paced online learning platforms. Today's learners are often learning on

the job and need the flexibility of short, or self-paced learning experiences. Though much of training and learning is happening online, the importance of guided instruction and learner-instructor interaction is still apparent.

Data-heavy research fields require the use of sophisticated tools and computational resources. Unfortunately, many researchers or individuals who have data, but aren't sure what to do with it, are prone to using tools that introduce errors into data, such as automatic date formatting and unexpected sorting results.

There is a better way and there are better tools available today for researchers, including statistical programs such as R or Python, and cloud resources. The problem with these solutions is in training the intended user base on how to leverage them. There is not enough time in their day for formal training programs, and even if there were, there are not enough instructors to support and train large communities. Many communities do not have access to technology, or even the bandwidth to participate in formal training programs.

Low hanging fruit in my opinion is to reach diverse populations through training. Exposing diverse communities to coding and data will expand the number, type, and caliber of projects in our community. So how do we bring people with diverse perspectives to data? How do we empower more people to work with data when the keepers of the conversation don't look like or identify with them? We know that empowering more people to work with data allows us to answer more questions in science and society, but what does it take to attract and retain them? Something The Carpentries is doing to help address these issues is aiming for inclusive pedagogy in our training and resources.

The Carpentries

The Carpentries were formed to teach foundational computation and data science skills through short impactful workshops. The model we follow is to "train the trainer", where we certify volunteer instructors within the communities they are trying to reach. All course materials are open source and in the public domain and free to use and contribute to. In this way we are able to scale beyond our organization and build regional self-sustaining learning and support communities for advancing research.

The Carpentries is a fiscally sponsored project of Community Initiatives, a registered 501(c)3 non-profit based in California (USA). The Carpentries project includes the

Software Carpentry, Data Carpentry, and Library Carpentry communities of Instructors, Trainers, Maintainers, Helpers, and supporters who share a mission to teach foundational computation and data science skills to researchers, technologists, and librarians.

The Carpentries builds global capacity in essential data and computation skills for conducting efficient, open, and reproducible research. We train and foster an active, inclusive, diverse community of learners and instructors that promotes and models the importance of software and data in research. We collaboratively develop openly-available lessons and deliver these lessons using evidence-based teaching practices. We focus on people conducting and supporting research. All Carpentries workshops are community driven and taught globally. I'd like to tell you a bit about each of the lesson programs for The Carpentries.

Software Carpentry

Software Carpentry was founded in 1998 to develop materials and train instructors to teach computing skills to researchers in science, medicine, engineering, and other disciplines. Since 2012, volunteer instructors have run workshops globally, and impacted more than 34,000 researchers. Software Carpentry lessons include three core topics: the Unix shell, version control with Git, and a programming language (R or Python). Curricula for these lessons are available in both English and Spanish.

Data Carpentry

Data Carpentry was founded in 2015. Its curriculum centers around the fundamental data skills needed to conduct reproducible research. Data Carpentry lessons cover several domains (Social Sciences, Ecology, Genomics, Geospatial), and lessons are developed collaboratively. Lesson content includes data organization in spreadsheets, data cleaning with OpenRefine, data management with SQL, and data analysis and visualization in R and Python.

Library Carpentry

Onboarded as a lesson program under The Carpentries in 2018, Library Carpentry develops lessons and teaches workshops for people who work in library or information-related roles. Library Carpentry's goal is to "create an on-ramp to empower this community to use software and data." Additionally, Library Carpentry seeks to train individuals on efficient and effective reproducible data and software practices. Library Carpentry's core curriculum includes an introduction to data, the Unix shell, and OpenRefine.

For the next few minutes I'd like to focus on the pedagogy that has made our approach sustainable for the communities we reach, and what we, the people in this room and those who watch on YouTube later, can do to democratize data, build confidence in ourselves and others, and support our mission and vision to solve problems in a collaborative way, ultimately achieving equity. I'll do this by discussing how The Carpentries teaching practices align with what I've learned in my personal and professional journey, and we'll wrap up with a call to action and a short writing assignment for us all.

My Journey to Data

I took some time to reflect on my life and where data played a significant role. I was born in Detroit, MI in the United States in the 1980s. During a time when my city was at or near the top of unemployment, poverty per capita, and infant mortality, my parents were wed and brought me into this unpredictable world.

During this period my city had gone through horrific disappointment with the reversal of *Milliken v. Bradley*, a federal court decision that would have actively desegregated Detroit and suburban communities, and fixed many of my city's problems.

In the 80s my city became notorious for crime, and was repeatedly dubbed "the arson capital of America", "the murder capital of America", and "the most dangerous city in America".

It wasn't all doom and gloom for my city in the 80s though. The Detroit Tigers won the World Series during this time, and both Nelson Mandela and Pope John Paul II visited my city.

But you know what? I don't remember any of that. I remember growing up in this house. I remember backyard barbeques. I remember slumber parties with my cousin Shawnie who looked just like me. I remember all of my uncles living in our basement at one point or another. I remember making snow angels outside in the winter and jumping through the fire hydrant as water gushed from it into the street in the summer. I also remember learning that opening the fire hydrant was illegal, but I digress...

I remember Christmas lights and Thanksgiving turkeys and the first time my mom added baby carrots to our spaghetti. I remember loving my house, loving my family, and loving my city.

You may have noticed I repeatedly refer to Detroit as "my city". It was then, and though I don't currently live there, it is now because though the plight of my city was gruesome, living there taught me persistence and resilience.

I remember some other things about growing up in Detroit. I remember dropping to the floor as gunshots rang in the New Year, or a random Tuesday.

I remember having our home broken into and all our Christmas gifts stolen.

I remember my brother being arrested because he fit the description of an armed robber.

I remember how difficult it was to plan celebrations with my parents, since they'd been divorced since I was 3.

All of these anecdotes are data points, and if we were to trust this data, a Detroiter like me born in the 80s would, presently live below the poverty line and work in either accommodation and food services.

If we trust the data, a Detroiter like me born in the 80s would presently rely on medicare and other federally funded programs to support herself.

But, there was another plan for me, and that plan begins with these two: Myling and Albert.

My parents taught me to work my tail off to get good grades, I played sports. I sang in the choir. I did everything I needed to do to get into Engineering school. About a 10 hour drive north of Detroit there's a small town in the Upper Peninsula of Michigan named Houghton, and that's where I went for my undergrad and master's. Michigan Technological University was a place I'd never thought I'd end up. I was one of the only women and people of color in most of my classes, but I was determined to do well because I got accepted into engineering school. I was invited to the party.

After earning a bachelor's and master's in mechanical engineering I decided to pursue my PhD, but wanted to address the fact that there weren't enough people who looked like me in the field, so I switched my focus from Mechanical Engineering to Engineering education, and that's how I learned about many of the techniques The Carpentries uses to teach workshops and build community, so let's explore them now.

Approaches to Building Confidence to Work with Data

Building confidence to work with data requires multiple approaches. Combining the power of guided instruction with the flexibility of short, focused learning experiences has the potential to achieve equity among practitioners who want to work with data. By acknowledging and building upon the four pillars of self-efficacy, The Carpentries has been able to develop training and build capacity for communities of practice sharing good data practices. Here's what I mean by self-efficacy.

Albert Bandura's Theory of Self-Efficacy

One of my favorite quotes from Mahatma Gandhi reads, "If I have the belief that I can do it, I shall surely acquire the capacity to do it even if I may not have it at the beginning". Gandhi's quote describes how I think about self-efficacy. Self-efficacy refers to an individual's belief that she is capable of taking action to achieve a certain goal, such as writing a program to solve a problem she's interested in. Self-efficacy is formed by performance outcomes, vicarious experiences, verbal persuasions, and physiological feedback. Increasing self-efficacy can ultimately improve access and equity for individuals who are historical underserved in this space. So I didn't come up with anything I just said. It's all Albert Bandura.

I get this question a lot, what's the difference between self-confidence and self-efficacy? Self-confidence is trusting in your abilities, qualities, or judgements, however, self-confidence refers to a more generalized personality characteristic. Self-efficacy points to your ability to succeed in a specific situation or accomplish a specific task.

Performance Outcomes

When we talk about performance outcomes, it means previous success that leads you to believe you are capable of completing a similar task in the future. Comparing The Carpentries pre and post-workshop scores, we see full point confidence gains in learners being able to write a program to solve a problem in their work, and even their overall confidence in programming. This is after a two day workshop!

Vicarious Experiences

When you see someone else completing a task, you believe you can do the same. This is one of the reasons why I love the Carpentries teaching practices. Rather than displaying a slidedeck of code, instructors deliver content by live coding. These are vicarious experiences This method provides learners the opportunity to practice, and receive continuous feedback about their code. It is important to keep in mind, however,

that feedback is not helpful if you cannot understand it. Live coding facilitates tacit knowledge (i.e., learning by watching *how* instructors do things).

Verbal Persuasion

The best way I can describe verbal persuasions is by showing you a video.

Supportive people in your life such as teachers, family, or mentors help you build self-efficacy. The Carpentries engages in verbal persuasion by way of our mentoring groups. Mentoring groups are small groups of instructors lead by a mentor, normally 4-5 people, and they focus on contributing to lessons, developing confidence and skill in teaching, the logistics of running Carpentries workshops, and tried-and-true events that foster local community building.

Physiological Feedback

Lastly experiencing sensations and emotions can affect your ability to see yourself completing a task. Let's check out this video.

We tend to think of emotional arousal in terms of excitement and joy, like you saw from the video of my father dancing on his birthday, but at times learning environments can create a sense of anxiety if they aren't designed and facilitated with equity, inclusion, and accessibility in mind.

The Carpentries are committed to making participation in our workshops a harassment-free experience for everyone, regardless of level of experience, gender, gender identity and expression, sexual orientation, ability status, personal appearance, body size, race, ethnicity, age, or religion. We establish norms for interaction by having, discussing, and enforcing a Code of Conduct such that our workshops provide open and inclusive learning environments.

Recap: Self-Efficacy

To recap, considering any or all of these factors determining self-efficacy can help us not only build our own self-efficacy and general confidence to work with data, but it can ultimately achieve equity and help each of us feel as if our contributions are welcome and acknowledged.

Call to Action

I've asked us to consider many things this afternoon, and now it's time to bring it home with this thought. Science, scholarship and society is better served by having diverse people, with the skills to use data to address the questions that are important to them. I

encourage us to work together to provide easily accessible resources for people who are unfamiliar with the tools and technology that we've grown to love [AND HATE!].

What if there were greater diversity in the spoken languages we teach and interact?

How can we recognize and appreciate the different cultural norms that exist around data, programming, teaching, and volunteering in different regions?

How can we recognize and value the various types of contributions we see in this space?

How can we work with existing organizations to reach broader communities, rather than building or re-inventing our own networks?

How can we authentically work with broader communities, rather than approach our work with the "we're doing it for them" mentality?

We won't be able to answer these questions or solve these issues over the next two days, but I do want us all to realize that our story and our contributions matter, and if we're going to build confidence in ourselves and the way we interact with one another in this space, we can only go further when we go together.

I wouldn't be a researcher if I didn't collect data, so if you follow the URL on my call to action slide it leads to the link for my slide deck, and also a prompt I'd like you all to respond to, and that is, please share an experience that may have created a barrier to you participating in data open source communities, and how that barrier could have been removed?

Thank you all very much.