Visualizing a Future: Some thoughts about the library's role in data education

- Thoughts are a work in progress
- Libraries have been integral part of career
- Grew up in Charleston, SC which impacted work
- Acknowledging those in Charleston who impacted work/life
- Libraries / information access as tools for social change
- Did not come into libraries thinking be a data librarian, most are self-taught skills or extracurricular activities
- Govt and info librarian at U of Oregon -- shaped approach to data, realized most people would need more education to take advantage of all that is available to them
- Our Data Society
  - Engaging with data competencies is important for career success and social outcomes for communities. Safiya Noble, Cathy O'Neill, etc.
  - Physical and psychological outcomes for communities
- Google photos -- potential harms that a lack of diverse access can have
- Hinted at larger issues to confront
- Google tagged Black people as gorillas (Twitter thread that shared this)
- Entrenched issues that made this a possibility
- This is one of the problems that comes up when don't have any sizable population of POC in quality control or engineering teams
- "Why aren't there more POC in these teams?"
- Lack of poc in tech, major diversity problem in tech, critically underrepresented in careers
- Richard Newman, *Liberation Technology*
- ProPublica, 2016 investigation -- "Machine Bias"
  - How algorithms run through society and impact systems
  - Found that Black Americans were more likely to be rated as high risk even though that was not the case by a substantial margin
  - Narratives that were being constructed around what it meant to be likely to offend/reoffend while on bail that were stand ins for other factors of people's lives (such as employment, economic level of where they live, etc.)
  - Race is heavily associated with wealth/bias
- These tools are being leveraged in schools as well -- such as learning analytics -- which has led to some positive impacts, but there are negative impacts
  - Learning analytics takes a lot of shortcuts because many details of lives are outside of the classroom and/or from people before them
- Hear about the pipeline problem -- reason why there is such a deficit of underrepresented voices is largely because there just weren't enough people in those communities that had those skills -- this is problematic
- Idea that as a result, there was a problem in the pipeline and initial approaches to solving representation issues there focused on this
- Program that focused on partnership between HBCUs and Google were seen as a key process of "fixing" the pipeline, bringing more STEM/tech curriculum in HBCUs, etc.

- However, not sure if these programs were actually successful and resources that were originally allocated to these programs were reallocated to some other programs
- Pipeline programs -- critical lack of introspection and "critical lack of introspection and a failure of accountability on the part of hiring agents" -- were still very influential. Impacted how thinking about data education.
- Critical rupture in the pipeline has also been impacting how universities have been shaping their curriculum
- U of OR made a big push into a data focused future, like many other universities
- Concerned about readings in the literature -- knowing that university was making change but that minoritized groups often left out of the program-- equity at its center
- Not really any unit on that campus that wasn't exploring data in some way. But a lot of these programs didn't have a support network because there wasn't a support for data education on campus
- Built with equity in mind and within a framework to supply that pipeline. Built a one credit course in RDM and research records management. In conversations with research faculty, they didn't feel comfortable with bringing students into labs or classes bc students didn't have strong background in RDM
  - Translate a skill that they're building, **have visible on their transcripts**
  - Put parameters on it and kept it to STEM students because couldn't target to more than ten programs at a time
  - Very successful, but a lot of people who didn't have opportunity to participate
  - Changed track to focus on non-credit bearing education (to be the biggest provider of that on the campus)
  - Robust education portfolio
  - Recognized that one credit cost money, slowed degree completion
  - What had the greatest currency at the time was the Carpentries, built curriculum around that
  - Built curriculum around industry standard R and Tidyverse for students to have those skills
  - Wanted greater diversity in who was building those programs. Built out a peer training and consultation network that sought to have people from different ethnic, gender, discipline, and degree backgrounds down to lower intimidation for those seeking that work
    - Observed their communities and built responsive curriculum (even built a Tidyverse series in Chinese and started to build one in Spanish)
  - COVID happened -- university pivoted, funding disappeared, they were pulled into only academic departments -- infrastructural loss
  - Could still offer courses because they had the materials, but couldn't offer at the same scale
- At Columbia, also building on pipeline program with different spin. Foundations for Research Computing for grad students and postdocs around these sorts of methods. Very successful. Addressing a pipeline issue.

- Looked at work other institutions were doing and talked to other institutions. WIth private industry building these pipelines, why are we still asking about a lack of diversity in tech careers.
- Is it a question of infrastructure? Pipeline? Culture? All of those?
- Quote: access to education in computer science it's very skewed by income. Many companies don't recruit from women's colleges or HBCUs.
  - The U of OR program wasn't really acknowledging this.
  - Not just quantity but also quality of program for people to participate.
- Resume screeners were screening out HBCUs/Divine Nine because they didn't recognize them
- Are we in libraries in the position to even be advocates when tech culture behaves in this way?
- What tools do we in libraries have to address these issues? Can we change the culture?
- Leverage traditional strengths inside the libraries. Social infrastructure (Eric Klineberg). Shape the way people interact.
- Ruha Benjamin -- "Move slower and empower people." -- questioning success because the emphasis of these programs was not to empower but to share with most amount of people
- Reframe the problem: Eric Liu's laws of power and rules for action:
  - Change the game
  - Change the story
  - Change the equation
- Broad problem is that there is a lack of diversity, not that there aren't enough people with the education for it, but problems in hiring, recruitment, retention, etc. Even as industry are recruiting more, the retention is suffering (ex. Google diversity report)
- Reframe the problem from a pipeline problem to an empowerment problem
  - What are the actual skills we need for people to participate in this?
  - We don't have all the advantages to make impact in the pipeline, but can build out the competencies to prepare students regardless of discipline so that they are able to work in the spaces -- they don't need to be reshaped
  - Change is what we want to do -- move from driving people on the pipeline to actually accessing the skills to move forward
- We are not alone in this -- Noble's The Equity Engine is doing the work.
  - "When Black women are given the runway of time, resources, and networks of power, we often expand opportunity exponentially for others."
- Where do we go from here? What does this mean for libraries? What can we do in libraries to use our skills and position on campus to support equity deserving communities?
- Use our infrastructure and library as a community space to move this forward
- Q: ?? - What impact does the librarians demographic (lack of diversity) impact these inclusive ideas of data education?
  - So much of the language is about pipeline programs to get POC into librarianship. On that level, it has some real impact. Equity deserving groups are pretty underrepresented in specialized data roles and that does play a part.

- - Reaching out to people outside the library for peer roles was also about getting more diverse representation in the program
  - U of OR are continuing to build data specific or digital skill specific programs/internships, making sure very intentional in recruitment, "BIPOC folx strongly encouraged to apply"
  - Don't have a ton of answers with this one, lots of approaches here
- Q: ?? – As you've been developing new programs, have you also been considering any new approaches to marketing your services or outreach on campus to reach a wider group of students?
  - Definitely. It's a continuous trial to build outreach initiatives, starting with basic web presence, direct appeal to personal relationships you have on campus and in your communities. Have done things such as printing out takeaway cards and sticking them in every mailbox in key depts, personal relationships from student core, leveraging colleagues' expertise and notoriety to get publications about them and about services had tremendous impact. As moved into new position, thinking about that a lot. At a really distributed campus, what does it mean to reach key demographics? That was what was really important to them, not just having blanket coverage. More impactful to do that work that way anyway.
- Q: Right at the beginning, said: mentioned that you were self trained as you moved from gov info into data based work. So many of us were self trained. Relationship between self training and rise of programs that are more formalized in training people and how that shifts questions about access and empowerment for communities and communities of students who were training.
  - A: Being self-trained was really important to me and something that I talked about a lot in building out the service. Coming from a background that was initially focused on the humanities and moving into the social sciences. The reason I came into this is because it became relevant in my work as a grad student but then professionally as I started to see a variety of ways for it to be beneficial for me to execute my work. Fortunate that thought had a problem that I could solve, was working in a space that would allow me to access self-education, had a position that gave me the space to explore and do this work. This is not a reality for a lot of people. Appreciated having space/time to do it and have this learning. Semi-structured or semi-formalized experiences. For good or for ill, there is a validation in being a part of this program that gives people this space/time/money to participate in this training -- where that might not be a part of someone's other job or of another job. The more of these that exist, in academic space, attached to a larger structure that the university respects. Helps to translate this discussion.
  - Taught myself the basics of R, but also certified in The Carpentries as an instructor. Having conversations with admin, etc. mentioning that Carpentries certified gave some credibility to be in that space. Self-education component was important because it helped me to appreciate some of the challenges that are being faced by learners in programs or learners who do not have that support network. Framed thinking about what that program was going to be.

- Q: ??building on the idea of instructor demographic and it's effect on student empowerment, do you have suggestions for how those of us who teach visualization workshops can train up folks who can replace us as instructors for marginalized groups in our community?
    - Really challenging thing and really contextual. Have focused historically on trying to bring as many voices into the discussion as possible. Building peer infrastructure works really well. Builds on another generation or community that wasn't tied into your program before. Got a lot of energy and insight from peer instructors. They did a lot of work and were available. Also gave a credential with a university name to help them move into their next project/next level.
    - Looking for ways to get in to communities that can be reached by your program. Important to compensate. Requires time and energy. Equity deserving groups are generally undercompensated anyway. For a lot of budgets that complicates things. Or holding firm and advocating for that.
    - Forming partnerships outside of your dept, libraries, your campus is important as well. A lot of campuses are looking at ways to foster more interactions with the communities that surround their spaces as well. Ex. partnered with local data journalist to teach their first workshop.
- Comment from former colleague at U of OR: Thinking about question about diversity in leadership. Requires a certain amount of active self-awareness and explicitness. Say with your words and actions over time.
- Q: ?? Questions from Emma Slayton (the chosen 1 lol) - What systems for evaluating student learning over longer periods of time have you used, and how do these systems help you to better help people enter and stay in the pipeline long term? Have you noticed big differences in students who learn on their own or in a course
    - With the formal course, had pretty standard methods of evaluation which were basically competency tests. Took theoretical frameworks they were trying to build. Had to actually travel along a research cycle with a dataset from raw to with a DOI/in repository. Was a smaller group and in a contained environment.
    - With non credit bearing, metrics became really difficult to even conceptualize. Even in introduction to R, people were coming in with all different levels of knowledge. How do you devise a metric to see how successful someone is? Started to rely on soft metrics. Looked at the series to see how many people went from point to point in the series. People will come in and come out. Building a pipeline as a complete skill, and if someone went through the entire pipeline felt comfortable that they had the knowledge. Did pre and post tests but not sure how comfortable saying that it's validation.
    - Same thing here with pre and post tests. Very interested in people who attend workshops and participate in community functions after the fact. Need to do more work in engaging with instructional designers for doing assessment for those communities.
- Q: What kinds of differences do you see playing out in different organizations?
    - Do see differences, which is why took a moment to talk about where worked before -- large urban private school, large urban public, large West cost R1, back

into large urban private. There are differences in undergrad and grad communities, but highly contextual on where you're at. At U of OR, focus on undergrads. Data science initiative started in 2016 -- no shortage of grad students who were missing skills that undergrad communities were missing bc no support in their programs.
  ○ Spaces that have undergrad and grad communities -- understanding the narrative where the service is owed between those groups. Navigating that to build the program you want to build is really important. Spaces that say they serve undergrads, spaces that say they only support grad students/researchers because they are the ones doing the research (in their opinion).
  ○ A lot of the skills are transferable.
  ○ Lots of social contexts, political realities, reframe conversations to be successful in what you want to be able to do.
  ○ It's okay to be opinionated about the program you want to deliver. Keep that in mind as you're thinking about what communities you're serving.