Co-Creating Spaces with Local Nonprofits to Expand Civic Data Analysis and Visualization Capacities

https://sites.nd.edu/expanding-nonprofit-data-analysis/

Grace Connors, Rachel Salamone, Matelda Sweis
Lucy Family Institute for Data & Society, University of Notre Dame
Interdisciplinary Traineeship for Socially Responsible and Engaged Data Science
Advised by Mat Sisk and Katie Liu
May 3, 2023

Chapter 1: Introduction, Motivation, Problem Statement

Our world increasingly relies on data-driven decisions. Metrics underlie the work of every sector, and with the recent rise of Environmental, Social, and Governance (ESG) work, many companies have begun to focus on measuring their social and environmental impact in addition to their business successes. Yet, while these for-profit corporations have the funds to stay up-to-date with the latest trends in data science, nonprofit organizations, which almost exclusively focus on social impact, often do not have adequate resources to maintain current data practices. Despite this capacity barrier, they are increasingly expected to prove their successes through data (Bay 2021). Further, many of these organizations lack the capacity to hire data-focused employees or train current employees on data literacy skills. As defined by data scholar Catherine D'Ignazio, data literacy is "the ability to read, work with, analyze, and argue with data as part of a broader process of inquiry into the world" (D'Ignazio 2017). Specifically within the context of non-technical learners in fields such as nonprofits or education, D'Ignazio further argues for *creative* data literacy, which teaches individuals how to read and analyze charts of data, but also to "use that chart to make the world a fairer place" (D'Ignazio 2017). In other terms, data literacy also encompasses the acknowledgement that individuals use data to interact with surrounding social structures (Chiewphasa 2021, *Finding Synergies*). This is directly in line with the mission-driven nature of nonprofit organizations, which seek to engage social structures and serve various populations around the world. Therefore, it is clear that nonprofit organizations are a target audience for data literacy, and thus could benefit from increased skill development in this area.

Yet, because nonprofits are mission-driven rather than profit-driven and seek to promote social justice in various capacities, they often face many challenges to gaining those skills, such as reduced financial resources, staffing limitations, and cultural barriers. McAdoo (2020) notes that nonprofits often struggle to hire those who are data literate as "a direct result of…lack of funding." Moreover, publicly available government grants for nonprofits are often "contingent upon the most vulnerable populations sharing their data," so that nonprofits can prove they were effective in providing assistance (Bay 2021). Nonprofits are then subject to the "data imperative," which is the push towards quantifying outcomes and tracking measurable success. In result, this cultivates a dynamic of "dubious consent," in which nonprofit organizations are required to collect and share data from those they serve, whose vulnerable nature means they may not always be in a position to provide clear consent for that collection (Bay 2021). Furthermore, in this push to quantify and track large amounts of data, these organizations are expected to direct some of their already-limited resources to data collection, maintenance, and cleaning. Instead of focusing directly on the mission of the organization, individuals are diverted from projects to secure grant funding and find ways to manage the nonprofit's data. This phenomenon is also known as the "data drift," as workers could be developing relationships with clients and helping with the day-to-day of the organization, but

instead are confined to often monotonous tasks related to maintaining data (Bay 2021). While perhaps seemingly a consequence of an increased reliance on data, the data imperative and data drift ultimately decrease the level of agency that nonprofits have in their division of labor, which has a material impact on those they set out to serve. A reduction of efficiency follows, as there may be multiple untrained individuals spending time maintaining data rather than serving people directly.

Thus considering these data challenges, in this project, we aim to develop a curriculum to teach core components of data literacy to nonprofits through the Carpentries pedagogical model and "[bridge] the data divide" (Chiewphasa 2021, *Finding Synergies*). The Carpentries model provides an ethical, two-way style of teaching and learning. It can be thought of as "a worldwide community centered around creating open-access curricula and delivering hands-on workshops to increase data literacy and computational skills" (Chiewphasa 2021, *Reflections*). This relatively new pedagogical model was developed by Brent Gorda and Greg Wilson in the late 1990s and consists of short, easily accessible workshops that guide researchers into new topics. With this goal in mind, the Carpentries model therefore assumes that students have no prior knowledge before entering the workshop. Additionally, workshops under this model can be adapted for any domain of research, and the model contains train-the-trainer programs to "equip the future instructors with practical skills to effectively teach students who may come from different backgrounds" (Teal et al. 2015). Leaning on walkthroughs as opposed to lessons geared to experts, these workshops provide a beginner-friendly environment intended to dispel the common concerns around using unfamiliar tools. Moreover, they build a community where researchers and other nonprofit employees can support one another in their learning goals.

In addition, the lessons in the Carpentries pedagogical framework strive to be both informed by the context and crafted to the specific field of the learner, as well as sustainable and scalable. These lessons are often tailored to a particular topic or audience in order to best facilitate the development of relevant skills. Though the amount that participants may learn in a short workshop is limited, the goal is to decrease the time they would spend attempting to learn and build new skills entirely on their own, giving them more time to spend on the functional mission of their organization and potentially, then, increasing their agency (Teal 2015). In a similar way, adaptive learning and sustainability are core components of the Carpentries model. Rather than relying on contractors outside the organization who specialize in data, the Carpentries model aims to equip organizations with applicable skills so they can more efficiently handle their data themselves. In doing so, these skills remain in-house and workers will have the resources necessary to teach future members of the organization (Yoon 2018). Therefore, these workshops alleviate the burden of paying for outsourced skills and provide a level of technical expertise that can be passed on throughout the lifespan of the organization.

This Carpentries model of workshops therefore informed our entire curriculum development process. To tailor the curriculum to the needs of our target audience, we spoke with several local nonprofits in South Bend, Indiana to identify their current data practices and any gaps in their knowledge. The information collected from these ethnographic interviews then shaped the content and design of the curriculum itself, which is presented in easy-to-read, three-part workshops that may be delivered in-person or online and will remain open source for public access on our website.

With the motivation and purpose of our project thus described, we will next outline the ethical implications and justification of our research through the ethics of care framework before covering related work. We will then elaborate on our methodology in the qualitative research process, as well as on the development of the curriculum itself.

Chapter 2: Ethics and Design Statement

Like any project involving data and technology, our workshop process navigates multiple ethical implications. While laws relating to the privacy of sensitive information and guidelines enforced by the Institutional Review Board serve as an important basis for ethical behavior relating to data, the nature and motivations in this project require a deeper commitment to ethics. To frame the ethical considerations of this project, one must first understand data justice, which is the "fairness in the way people are made visible, represented, and treated as a result of their production of digital data" (Taylor 2017). In developing and implementing a pedagogical model for nonprofit data literacy, it is critical to explore the notion of data literacy as social justice, the ethics of the Carpentries model of teaching, and potential ethical issues that may arise to ensure that the project integrates core principles of data justice and the ethics of care.

Feminist philosophers such as Martha Nussbaum often emphasize the importance of increasing agency when defining social justice. By teaching skills that develop data literacy, we attempt to restore some of this agency in our datafied world, creating more positive material consequences for those served by nonprofits and easing the burden of nonprofit administrators. In teaching those who already work for nonprofits, we attempt to directly mitigate the issues created by the data drift and the data imperative by reducing skill-related staffing shortages simply by teaching them new abilities. Further, this fits into the broader theoretical framework of the ethics of care. This framework notes the importance of considering the most vulnerable and disadvantaged in our moral reasoning, which is the exact population that most nonprofits tend to serve. One of the core values undergirding the project is the advancement of social justice and equality for all individuals in society. Teaching data literacy gives individuals more agency and embodies the ethics of care by working towards social justice and tending to the most disadvantaged. Still, while working with nonprofits seems to have a noble underlying motivation to advance social justice, its ultimate success depends

heavily on the teaching strategies and how they are implemented. As Raffaghelli (2021) notes, "whether education can be a response to promote social justice in the peculiar context of datafication is a matter of the pedagogical strategies designed and implemented." She urges for a curriculum that incorporates freedom of choice, participation, and awareness, similar to the Carpentries framework.

The Carpentries model itself incorporates ideas from significant philosophy frameworks, namely the ethics of care, in order to empower the vulnerable through data. Through the Carpentries model, instructors focus on teaching about real-world applications of data literacy that incorporate human data and the impact on vulnerable populations (Bay 2021). The ethics of care framework emphasizes special moral consideration to those who are most vulnerable in society, such as those with disabilities, undocumented individuals, and those facing financial hardship. Given that nonprofits often assist the most vulnerable, an implementation of the Carpentries model can facilitate learning that can improve nonprofit processes and allow workers to accomplish more to assist those who need it. Further, compared to businesses which gain profit through their operations, nonprofits are comparatively vulnerable in that they often rely on donations and grants. Engaging with the Carpentries model therefore provides access to new forms of knowledge and agency.

Community engagement pedagogies, like the Carpentries model, help us to better understand poverty and marginalization beyond numbers. As Bay (2021) notes, research shows that impacts of poverty and marginalization are also "affective, changing the way we interact with others, see ourselves, and understand the world." This further emphasizes the utility of the Carpentries model. The ethics of care calls us to think of ourselves as members of a larger moral community that is deeply integrated. Because we are called to interact directly with students and engage with the organization in the Carpentries framework, we are able to gain a better perspective about who the larger moral community really is. By coming to understand their needs through our relationships and engagements with them (both in class and out of class), we gain a fuller understanding of our society and the impacts of social structures in place.

Complex ecosystems of knowledge are also critical to social justice, which the Carpentries model nourishes. Following in line with a broader theory of human capabilities, Nussbaum believes that a central ethical question is "what is [one] actually able to do and to be?" (Nussbaum 2002). Given the fact that broad access to the Internet has provided a new means to connect and share resources, it would seem that nonprofits have everything they need to learn on their own to be successful and develop their technical skills. While it is possible for nonprofits to learn in this very informal way, a blend of formal and informal education can enhance flexibility, collaboration, and create a more creative and effective learning environment (Lai et al. 2013). This is especially relevant because "simple exposure" to information is often not enough to "[integrate] the information into a mental model" (Becker 2016). In addition, the Carpentries model does not deny that these organizations can function or learn on their own,

but instead recognizes the value of feedback and participation. In fact, informal education and perspectives are "where participation and activism are enacted" (Raffaghelli 2020). Carpentries' two-way learning style promotes this participation, therefore furthering aims of social justice and the ethics of care. Aforementioned informal formative assessments and discussions foster such participation.

In implementing this model of learning to assist nonprofits, it is essential to understand the potential ethical issues involved and the available avenues to addressing them. For instance, without careful consideration, personal biases may influence the perception of organizations, their missions, and the people they try to serve. This could hinder us from providing lesson plans that adequately support equality and justice. However, there are a variety of methods to combat this. As mentioned in the previous section, the Carpentries model emphasizes community engagement and developing partnerships that aim to break down these biases. Additionally, Catherine D'Ignazio and Lauren F. Klein (2020) cite positionality as a powerful concept, stating that it "emphasizes how individuals come to knowledge-making processes from multiple positions, each determined by culture and context." By considering multiple perspectives, we aim to remind both ourselves and the students that all forms of knowledge are situated, and that universal objectivity–even through data–is an unattainable goal. Lastly, we can remind ourselves that the most marginalized individuals served by these nonprofits are the ones who benefit the most from these lessons and deserve extra moral consideration, following an ethics of care perspective.

Another significant ethical issue that may arise comes in the form of dealing with unfamiliar situations and information. More specifically, we may not be familiar with topics involved, organizations we are helping, or how best to go about advising them on ethical practices to use in their specific cases if the need arises. In situations of uncertainty, the use of moral exemplars can be particularly useful. As a starting point, Institutional Review Boards and other legal regulations offer basic guidelines for ethical research. In more complex situations that require more consideration, we consider advice from our project mentors and ideas collected from open discussions and interviews of people who work at the nonprofits. Reflection and formative assessments, which "provide opportunities for learners to practice using their new knowledge in a low-stakes atmosphere" can provide useful feedback (Becker 2016). Fortunately, room for assessment is already deeply interwoven into the Carpentries framework. Therefore, this is not only a helpful model for our students, but also for us as teachers as well; it is important that we take time to reflect on how effectively and ethically we handle aspects of the project. In this way, we can assess how well we are approaching and treating new problems and remain flexible in how we can adapt to them.

The last main ethical issue discussed in this paper stems from the general drawbacks of working with data. Learning about data analysis alone is not enough to constitute data literacy; instructors must remind students that data is inherently reductive and that human lives are too

complex to be represented as a series of numbers or categories. Data often encourages people to make broad generalizations about societal groups, and these broad ways of thinking can overshadow the concept that data is collected from the experiences of real people (Bay 2021). Therefore, it is imperative to integrate discussions on the systems of oppression we seek to target and the groups of people we seek to better understand. We advocate for students to avoid broad generalizations from a strictly quantitative point of view by also incorporating ethnographic interviews and qualitative data. While doing this, we note the limits of advanced data science concepts like artificial intelligence and explain that being overly-trusting of models may ignore the underlying social issues behind them. To help tackle these complex "iceberg" issues–aptly named for their appearance of simplicity but underlying complexity–we can integrate concept maps and lessons rooted in design thinking. In doing so, we call students to reimagine what data is and take a more expansive view so they avoid the trap of using numbers or non-transparent models to over-generalize situations.

When creating a curriculum on data literacy for nonprofit organizations, it is necessary to understand the core goal of data literacy as a mechanism for social justice, the ethics behind the Carpentries model itself, and think through potential ethical issues so that the project properly reflects facets of both data justice and the ethics of care. Of course, a fundamental part of the Carpentries model is to remain flexible and develop the relationships between the teachers and students, and part of this relates to constant communication and collaboration between the Carpentries teachers and the organizations themselves. It is crucial to understand the unique needs of each organization and accept that they may change over time. With the growing popularity of Carpentries model workshops amidst universities and researchers (Teal 2015), we hope to not only provide these nonprofits with the skills needed, but also provide an ethical framework that future workshop trainers may use.

Chapter 3: Related Work
Carpentries Model Implementation

There are two related and noteworthy implementations of the Carpentries model with the goal of developing data literacy in their students: those hosted by the New York Data Carpentries Library Consortium (NYDCLC) and the workshops hosted by Ben Chiewphasa and Anna Moeller at the University of Montana. The NYDCLC is a Carpentries-affiliated "association of academic, public, and school librarians who have come together to form a community of practice focused on building data skills, including accessing, analyzing, using, and visualizing data" (Lindbloom et al. 2021). Workshops were hosted in the state of New York just before COVID-19 shutdowns and were designed to help library professionals with skills such as SQL, web scraping, data cleaning, and data tidying. Its participants were mainly from a variety of academic libraries that were recruited by email. The creators of the program found that there was a continuing need for data literacy. Moreover, participants beforehand reported they "don't

know what we don't know," highlighting the significance of the Carpentries' assumptions of no prior knowledge. Because their findings were not published in an academic journal, it is difficult to assess the effectiveness of the program, though the qualitative interview data collected seems to show that participants enjoyed the program and learned from it.

Next, the workshops created by Ben Chiewphasa and Anna Moeller also serve as a successful example of Carpentries implementation. Though they were initially the first in-person Carpentries workshops on R at the University of Montana, they were later moved online due to COVID. Chiewphasa and Moeller designed these workshops to be "based on established lessons consisting of a number of episodes" (2021, *Reflections*). This program consisted of six workshop episodes in one eight-hour day, with breaks every two hours. They were taught by volunteer instructors who were certified through the Carpentries Instructor Training program. When teaching each lesson, the instructors taught in a team—incorporating interactive activities—as opposed to using a lecture-based lesson plan. The material covered prioritized the practical over the theoretical, aiming to provide its students with applicable skills they could use in their research. One of its goals was to create an inclusive and encouraging learning atmosphere. In emphasizing community-building, the workshop started with icebreakers and had participants work with their neighbors. Aligning with the Carpentries ideals of learning from doing, the workshops involved live coding so that participants could hone skills, work with others even in an online environment, and practice troubleshooting. The last main goal of their workshop was to assess how well these workshops functioned in imparting skills useful for working with R. To do this, feedback and formative assessments were collected from the participants. Because of the COVID-19 shutdown, modifications were made to the timing of the workshops as well as curriculum itself, as the instructors reduced the quantity of materials presented. Despite changes and working in an uncertain environment, learners seemed "excited about data manipulation and visualization," which they were able to learn in subsequent online workshops (Chiewphasa 2021, *Reflections*). This excitement around data science often indicates learning and development of skills, meaning that these workshops were effective at teaching data literacy in the Carpentries framework. The flexibility of moving these workshops to an online environment also speaks to the adaptability of this model, and how it can be improved upon with future iterations.

Critical takeaways from these studies involved feedback from respondents regarding the duration and outline of the workshops as well as important notes for instructors to consider. For example, participants in the NYDCLC workshops emphasized that the workshops must be no longer than half a working day, that they must be learning skills that they would use immediately (including throughout the workshop), and that documentation available to reference after the training concludes is essential. As instructors, we implement this into our own model by keeping workshops short and providing all of our curriculum and documentation to the participants, given that it is open source. The interruption of the University of Montana

workshops by the COVID-19 pandemic also provided valuable insight, as it was a reminder that sometimes unforeseen circumstances can provide obstacles to teaching. Thus, it is essential to remain flexible in order to adapt as necessary. This involves focusing on the basics and remembering that instructors may not always have time to cover all of the course materials they desire. While our curriculum contains a broad range of information, the instructors are able to adapt and are aware that time constraints will likely result in only a small portion being taught.

Courses on Data Science for Nonprofits

There are multiple online courses that currently exist for the purpose of nonprofit education in the use of data science. One such course was created by three Brigham Young University graduate students and is hosted on Udemy, an online learning and teaching marketplace, entitled "Intro to Data Analytics for Nonprofits." The course runs for 3 hours and 15 minutes, costs $19.99, and covers the following topics: what is data, what type of data to collect, how to collect data, Excel basics, data visualization, Tableau basics, and data security (Clinger et al. 2020). With over 2,545 students who have completed the course and a 4.0 overall course rating, this course seems to be effective at accomplishing its goals. This was the most in-depth curriculum specifically created for nonprofit organizations that we were able to source online, which reflects a need for tailored education. The reviews additionally reflected the value of the detailed nature of the curriculum in teaching practical skills within softwares such as Excel and Tableau.

Many other tech companies or universities provide online courses to teach data science skills to a more general audience. Students can take many of these classes on Coursera; some contribute to degrees or certifications, while others simply provide an introduction in sub-topics like data analytics. For instance, "Ask Questions to Make Data-Driven Decisions" is a course created by Google, while "The Data Scientist's Toolbox" is owned by John Hopkins University. Many of these courses are longer than the aforementioned course hosted by Udemy; the Google course takes approximately 21 hours to complete, and the course by Johns Hopkins University takes approximately 18 hours. Courses also tend to handle different content. The course by Google offers insight into problem-solving, decision-making, and organization (*Ask Questions*). On the other hand, the class by Hopkins tackles topics in using R, GitLab, and study design concepts, showing that users with a variety of data-driven needs can find courses to suit them (Leek et al.). However, these examples also highlight the possible overwhelming sensation users may feel from trying to find the right course for them.

Another course through LinkedIn Learning is taught by Barton Paulson, entitled "The Data Science of Nonprofit Organizations." This course runs for 1 hour and 8 minutes, costs $24.99, and specifically covers the following topics: data sources for nonprofits; web analytics for nonprofits; methods for evaluating impact, identifying donors, and assessing need; improving productivity and efficiency; ethical concerns; and careers for data scientists in

nonprofits (Paulson 2018). A final course discussed here hosted on TechSoup, another learning marketplace, is entitled "Data Analytics for Nonprofit Staff 101: Introduction to Data Analytics." It runs for 1 hour, costs $10, and works through the basics of data analytics for nonprofits: why they are necessary, what the components are, and how to evaluate current uses of data analytics (Fung et al.). Seeing as each of these courses are roughly 1 hour, it is evident that they lack in-depth detail on the practical skills related to the topics they cover, again pointing to the need for this type of education for nonprofits. Regardless, they both are attempting to address the specific needs of nonprofit organizations within one coherent body of education, which is still valuable.

These courses already exist, yet many nonprofit organizations still struggle with using data effectively. Therefore, examining the reasons behind this gap was critical before developing our own curriculum. A report by EveryAction and Nonprofit Hub found that the top five factors preventing nonprofits from using data was a lack of time or capacity, a lack of experienced personnel, a lack of one location in which to store data, a lack of tools to analyze data, and a lack of collected data at large (EveryAction and Nonprofit Hub 2013). Many of these topics were reflected in the courses outlined above. In terms of capacity, this might mean that some organizations, especially smaller ones, might not have the resources to pay for multiple staff members to take any of the aforementioned courses, or the time to devote to learning complex skills while completing their usual obligations within or outside the organization. On the note about experienced personnel, the article went further to explain that 40% of participants stated that they had little training in the software the organization was already employing for data collection, reflecting that a need for increased data literacy still exists even within organizations already attempting to use technology (EveryAction and Nonprofit Hub 2013). Finally, the article noted that 97% of participants stated they would be interested in learning more about data, pointing towards the necessity of this type of project (EveryAction and Nonprofit Hub 2013).

Therefore, there is a clear need for increased data literacy within nonprofit organizations and there are gaps that still exist within existing online courses for data science, both of which our project aims to address. With the implementation of Carpentries workshops, nonprofits get to learn in a more interactive environment, one in which participants can discuss with one another, directly ask for clarification from an instructor, and bring to light general use-cases they encounter for which they would like to receive advice. Such an environment is especially helpful when it comes to getting started on hands-on training and troubleshooting technical issues that may be too complex to ask in an online forum. Additionally, many of these online courses are tailored to the average global nonprofit organization, while our project aims to focus more on the needs of our local context in South Bend. Using the topics covered in the courses as a background in the basic needs of organizations and coupling this information with the data needs expressed to us in our community partner meetings, we aim to develop a curriculum that

stands apart from the ones outlined above in both the depth of content covered and importance to our local context.

Chapter 4: Data

While this project is related to data science in that it aims to provide additional resources to build data literacy, it is important to understand the ways in which this is not a traditional data science project. One of the most notable ways that this project diverges from others is that it does not draw on one particular dataset from which to study and make conclusions. The process of curriculum development may not involve the analysis of supposed "raw" data or numbers on a spreadsheet. Because this project aims to also embody the ethics of care and principles of data justice, it is critical to understand that data is far more comprehensive than we commonly believe when we think about data: qualitative data, for example, is just as important as quantitative data. D'Ignazio and Klein (2020) note that "The process of converting qualitative experience into data can be empowering," which is our hope for nonprofits as their workers develop more data literacy skills. Even though this project does not use traditional quantitative data or methods, there are still two areas of data generation in this project: data collected from qualitative interviews of nonprofit workers, and a post-workshop survey to measure the workshop efficacy and engagement levels of the participants.

As mentioned above, one of our primary sources of data in this project was through a series of ethnographic interviews with nonprofit organizations based in South Bend. Ultimately, individuals from nonprofits with whom we confer should be central in the decision-making process. To prepare for these interviews, we wrote out an interview guideline worksheet with a privacy statement we shared at the beginning of the meeting, and then asked specific questions to each nonprofit with which we engaged related to their field of focus. We talked with David Power and Yvonne Chang of the Boys and Girls Club of St. Joseph's County, as well as Jon Schommer, Executive Director of Our Lady of the Road, the organization which also helps facilitate the Motels4Now program. We actively searched out nonprofits in different sectors in order to understand the various needs present in our local nonprofit ecosystem and find any areas of overlap or disjoint between them. Though we wanted perspectives from various nonprofits, we only interviewed local nonprofits in order to maintain the focus of the workshops on providing help to those in the St. Joseph County community, since that was the focus of our project. Typically, these interviews lasted between half an hour to an hour and were conducted online via Zoom. Below are the general questions we asked our nonprofits, divided into categories that we wanted to highlight.

- Overview
  - Tell us about the mission of this organization.

- How does your organization use data? Why is data important for this organization? For example, who are stakeholders/shareholders of the organization? Who looks at the data and what do they need to know from it?
- Data Handling
  - What data is stored? How is that data handled/stored? What software does this organization use, for instance?
  - Who is responsible for managing the data within the organization?
    - Do you have a specific technology-focused staffer or is it handled by someone without much experience?
  - Has the organization faced difficulties in the past with managing data? What are some of those difficulties?
- Skills and Workshop Training
  - What are the skill levels like of people when they first start working with this organization? Is there anything they learn on the job, and if so, how are they trained?
  - What technical skills do you use consistently within the organization? Are there any gaps you perceive where increased education could help?
  - What technical skills do you already know? Ex: using excel or programming
    - What would you say is the average technical understanding of your staff?
  - What does the organization often need help with? If workshops were offered to teach skills related to data science, what skills/softwares/topics would be most helpful for this organization and why?
  - If workshops were offered to teach skills, what is your preferred duration, setting (in-person/online) and amount of workshops offered?
- Security
  - What about security? Consider these risks an organization might have when handling data.
  - Is it all stored in one place? (durability)
  - Are there backups of data? (durability)
  - Can anyone see/alter private data they shouldn't? (integrity)
  - Have you given much thought to each of the following, and if so, how have you ensured that data is handled to meet these properties:
    - Privacy/confidentiality - (Ensuring wrong people don't get access to data they shouldn't)
    - Integrity - (Ensuring data does not contain inconsistent information or errors)
  - Accessibility- (Ensuring it can be accessed when it's needed and within a reasonable amount of time)

- ○ Durability- (Ensuring that the data will not be lost when faced with unexpected software crashes or power outages)
- ○ How can this organization improve upon how they are currently meeting these properties?

After aggregating the notes taken from our interviews, we came to the following conclusions:

1. Many of the nonprofit organizations in our local community do not have the capacity to hire a specific individual with a background in data handling or computer science to be responsible for their data, and instead they rely on staff they already have in-house. They also have individuals handling data constantly who are not aware of positive online behaviors, such as not sending personal data over email. This means that there are many people not well versed in data practices that would benefit from a better education on the basics of data and data security.
2. Again due to the capacity limit, many organizations are using basic platforms to handle their data, such as Google Sheets, and aren't aware of the advanced capabilities these platforms have available for use.
3. The individuals at the organization responsible for acquiring funding are well versed in the mission and vision of the organization, but might not have a handle on the quantitative metrics, and could benefit from advanced understanding of these.
4. Simplicity is critical. The majority of members working at these organizations are not well-versed in data analysis or programming, and have only a surface-level understanding of how their organization's data is managed. These organizations are busy handling the needs of the most vulnerable in our community and do not have the time for a lengthy workshop, nor do they have the time to learn things that are not relevant to their daily lives. Therefore, it is critical to tailor our workshops to these needs to make them most effective.

Based on these observations, we then decided upon the topics for our workshops, of which we designed two, and which will be discussed further in the next chapter.

Secondly, a survey was developed in an effort to collect data after the workshops. This survey aims to collect data regarding the participants' experiences in the workshops, including which workshops they attended, the date(s) that they were a part of the workshops, whether they were instructor-led or if they viewed them asynchronously, the level of difficulty of the workshops, and the level of relevance to their jobs. Additionally, this survey uses a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree" for level of agreement to the following statements:

1. I felt engaged during the initial "lecture" portion of the workshop.
2. I felt engaged during the activity portion of the workshop.

3. I feel as though I have learned something new during my engagement with the workshop.
4. I feel as though I will be able to apply what I have learned to my job.
5. I see myself referencing the learning materials from this workshop in the future.
6. I would recommend this workshop to a colleague or friend.

If the individual engages with multiple workshops, they are prompted to fill out the survey for each workshop they completed. Additionally, there are spaces for open feedback on how the workshops could be improved and general feedback. Respondents also have the option to leave their contact information and their organization's name if they are interested in sharing more of their thoughts on the workshop. Because workshop design must often be iterative to be most effective, we expect and encourage feedback from participants to change and continue to shape these workshops, particularly as the needs of nonprofit organizations change. Ultimately, our focus is on designing the best methods to improve data literacy for nonprofits, so it is essential for us to implement these design thinking strategies.

Chapter 5: Method

As mentioned in Chapter 3, a core question to our project was how to best present these workshops. After much deliberation, we decided upon a hybrid model, where the workshops would be prepared as a rubric to a workshop to be held in-person, but which could also be read and followed along with on one's own timing online. We roughly followed the model of multiple Carpentries workshops posted online, such as the "Building Skill With Practice" workshop on the Carpentries page (Building Skill 2023). These models start by giving an outline of the workshop, the proposed time frame, and the objectives for the lesson. We then added in an introductory step to our workshop to allow participants to connect with one another and think deeper about how the following content relates to their work at their nonprofit, before moving into the content of the workshop. It then moves onto an activity to encourage hands-on experience, before wrapping up with a few questions intended to serve as informal formative assessments.

Admittedly, our backgrounds are in social science and computer science, not understanding the nuances of pedagogy and workshop classroom management. Because of this, a core part of the development of our workshops involved a preliminary step on pedagogical research. While the Carpentries model provides a framework for workshop development, it is important to understand why it is best for these workshops to have a particular flow that involves engagement with others and formative assessment. As Russ Edgarton (2001) famously noted, "learning 'about' things does not enable students to acquire the abilities and understanding they will need for the twenty-first century." He calls for new "pedagogies of engagement." Three key points are critical in pedagogies of engagement: the encouragement of "student-faculty contact, cooperation among students, and active learning,"

the latter of which is often also referred to as "informal cooperative learning groups" (Smith 2005).

After this pedagogical research, we crafted the workshops in line with these practices and the existing Carpentries workshops. We crafted six workshops in a two segment series based on the feedback received from the initial ethnographic interviews. The first in the series is a broad "Data Science Rundown" that does not necessarily involve the development of technical skills, but is intended to build the foundation for more advanced data science skills in the future. This includes three modules: Basic Data Terminology, Data Security, and Data Ethics. The first module on data terminology intends to establish a baseline of understanding; because the Carpentries model assumes no prior knowledge of the topics, we concluded that this would be a valuable workshop to those just starting in data science. The second module from this series focuses explicitly on data security because we found that many individuals within nonprofits may not be as well-versed in ensuring that legally protected data, such as those protected by HIPAA and FERPA, should only be sent in particularly secure means. The third module focuses on data ethics, as we believed that while this topic may not be as directly related to our findings, it would be a disservice to not provide an intersectional lens of data science within our curriculum, particularly because the prioritization of some forms of data over others is a process underlying all things.

The second series revolves around "Measuring Impact" and it includes three modules as well; there are workshops on how to determine what data is "important" for your organization and how to collect it, how to clean data with OpenRefine, and how to analyze and visually represent data. Many of the nonprofit workers with whom we met noted the importance of data from a grant writing perspective, and these modules are intended to more easily facilitate this process. Additionally, OpenRefine was selected as a module because it is a free, open-source software that can be used to clean data in conjunction with Google Sheets and Excel, which many of these organizations are already using to manage their data.

Because the workshops have unfortunately yet to be tested with an audience of nonprofit workers, we do not currently have data from the feedback survey of this process. It is our hope that this work will be continued by students and nonprofits in the future. Additionally, all of the materials developed in this process will remain open source and freely available online. As noted in the previous section, workshop development is often an iterative process and there cannot be clear conclusions drawn about efficacy until these iterations have occurred.

Chapter 6: Societal Impact

Although we were not yet able to present these workshops to nonprofit organizations and test their efficacy in promoting data literacy, this project still serves multiple purposes. First, it is a compilation of the ways in which our datafied world affects nonprofits and provides opportunities for continued research. While there are many research projects studying

nonprofits and their interactions with data, and additional research on the importance of data literacy, few consider both of these topics together. Some suggestions for research moving forward can be placed into two categories: South Bend-specific opportunities and broader nonprofit opportunities. In terms of South Bend, seeing as this project generalized challenges faced by multiple organizations in our community, future research could focus on the data needs of a specific nonprofit and tailor data solutions directly to their purposes as an organization. For example, something that arose in our discussion with Our Lady of the Road and Motels4Now was the challenge of duplicate data, where the local area requires the use of the HMIS data collection system, but the system is very clunky and not customizable, and so staffers collect that same data on their own Google sheets for ease of access. With that need in mind, a research project could be conducted on the efficacy of the HMIS system and a software development project could be done to build a solution to mitigate the duplication problem. In terms of the broader nonprofit sphere, there is much from our project that could lead to further research. For example, a research project could be done on the widespread use of Google Sheets in nonprofit organizations and the variety of tools within that platform of which nonprofits are aware. For example, some of the individuals we talked with did not know of the data modeling capacities of Google Sheets, whereas others used this tool often. Performing a wide-ranging survey of nonprofits across multiple sectors could thus shed some light on knowledge gaps with Google Sheets and inform future workshops such as the ones we have created here. Therefore, this project has had impact both on its own and through its capacity to inform future research projects.

Additionally, this project is intended to provide nonprofit organizations with resources to learn more and improve their data literacy on their own. All materials developed for this project are freely available on our WordPress website linked here. While many of the activities in the curriculum involve discussion with peers, they all can still be viewed and completed asynchronously, using the discussion sections as time for individual reflection. As stated, we may not have current data on the material impact of these workshops, but this can be a part of the continued work of our project and open source information. Ultimately, our hope for these workshops is that they are useful in developing data literacy skills among nonprofit workers, so that they may use these skills in jobs and become more efficient and effective in their role. This increased knowledge would hopefully benefit the clients of nonprofits in that a nonprofit worker may use skills learned from these workshops to write funded grants, innovate and develop new systems of data collection and tracking that fit their needs, and decrease the resource gap between nonprofits and the for-profit sector when it comes to data literacy. Further, the development of more concrete and efficient data literacy tools may enable nonprofit workers to spend more time cultivating relationships with clients, an integral part of nonprofit work, rather than working behind the desk with data. All of these outcomes are

attempted steps towards the promotion of social equality; by empowering nonprofit efficiency, we seek to support their missions as agents of change.

Chapter 7: Conclusion

The goal of this project was to develop a Carpentries-informed curriculum to promote data literacy skills in local South Bend nonprofits based on their needs. Through a series of ethnographic interviews, research on best practices in pedagogy, and research on best practices in industry on data literacy and data security, we developed six workshops within two broad categories: a "Data Science Rundown" and "Measuring Impact." We were unfortunately not able to teach these workshops in real time due to our extensive research on pedagogy and the Carpentries framework, but they will remain freely available online for individuals to browse asynchronously. Additionally, in the future, when the workshops are delivered, there is an opportunity to collect feedback and continue this iterative process of teaching and workshop design through online user surveys. Using the ethics of care philosophical framework, our goal is to support nonprofit organizations and their efforts to enact social change.

Works Cited

Ask Questions to Make Data Driven-Decisions. *Coursera.*
https://www.coursera.org/learn/ask-questions-make-decisions.

Bay, J., and Atherton, R. (2021). Rhetorics of data in nonprofit settings: How community engagement pedagogies can enact social justice. *Computers and Composition,* Volume 61. https://doi.org/10.1016/j.compcom.2021.102656.

Becker, E. (2016). Responding to your learners. *Data Carpentry: Building Communities Teaching Universal Data Literacy*.
https://datacarpentry.org/blog/2016/09/formative-assessment.

"Building Skill with Practice" (2023). *Data Carpentries*.
https://carpentries.github.io/instructor-training/02-practice-learning.html#top.

Chiewphasa B. (2021). Finding Synergies between Data Literacy, Community Engagement, and Librarianship. University of Northern Iowa: Service-Learning Librarian.
https://sllibrarian.uni.edu/articles/202111/finding-synergies-between-data-literacy-community-engagement-and-librarianship.

Chiewphasa B. and Moeller A. K. (2021). Reflections from Transitioning Carpentries Workshops Online. *Journal of eScience Librarianship*;10(4): e1217.
https://doi.org/10.7191/jeslib.2021.1217.

Clinger, E., et al. (2020). Intro to Data Analytics for Nonprofits (Charity Analytics). *Udemy*.
https://www.udemy.com/course/intro-to-data-analytics-for-nonprofits/

D'Ignazio C. (2017). Creative Data Literacy: Bridging the Gap Between the Data-Haves and Data-Have Nots. *Information Design Journal*. https://doi.org/10.1075/idj.23.1.03dig.

D'Ignazio, C., and Klein, L. F. (2020). Data Feminism. *The MIT Press*. Cambridge, MA.
https://doi.org/10.7551/mitpress/11805.001.0001.

Edgerton, R. (2001). Education white paper.
http://www.faculty.umb.edu/john_saltmarsh/resources/resources/Edgerton%20Higher%20Education%20White%20Paper.rtf.

Fung, L., et al., A. Data Analytics for Nonprofit Staff 101: Introduction to Data Analytics. *TechSoup*.
https://techsoup.course.tc/catalog/data-analytics-for-nonprofit-staff-101.

Lai, K.-W., et al. (2013). Blending student technology experiences.

Journal of Computer Assisted Learning, 29: 414-425. https://doi.org/10.1111/jcal.12030.

Leek J., et al. The Data Scientist's Toolbox. *Coursera.*
https://www.coursera.org/learn/data-scientists-tools.

Lindbloom, M., et al. (2021). NYDCLC Focus Group Report. *South*
Central Regional Library Council.
https://github.com/nydclc/surveydata/blob/main/NYDCLC_%20Focus%20Group%20Report.pdf.

McAdoo, M. (2020). Re-envisioning Data for Impact: A Case Study of a Durham Nonprofit
Organization. *University of North Carolina Digital Repository*.
https://doi.org/10.17615/cmyh-0v50.

Nussbaum, M. (2002). Capabilities and Social Justice, *International Studies Review*, Volume 4,
Issue 2, Summer 2002, Pages 123–135, https://doi.org/10.1111/1521-9488.00258.

Poulson, B. (2018). The Data Science of Nonprofit Service Organizations. *LinkedIn Learning*.
https://www.linkedin.com/learning/the-data-science-of-nonprofit-service-organizations-with-barton-poulson.

Raffaghelli, E. J. (2020). Is Data Literacy a Catalyst of Social Justice? A Response from Nine
Data Literacy Initiatives in Higher Education. *Education Sciences*, 10(9), 233.
https://doi.org/10.3390/educsci10090233.

Smith, K., et al. (2005). Pedagogies of Engagement: Classroom-Based Practices. *Journal of*
Engineering Education.
https://www.shsu.edu/academics/cce/documents/Pedagogies_of_engagement_Classroom-based_practices.pdf.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms
globally. *Big Data & Society*. https://doi.org/10.1177/2053951717736335.

Teal, T. K., et al. (2015). Data Carpentry: Workshops to Increase Data Literacy for Researchers.
International Journal of Digital Curation, 10(1).
https://pdfs.semanticscholar.org/47d9/b1f84b88268d3dec36e5ba2774dbeef5f52b.pdf?_ga=2.9597425.513753401.1650301080-1693425776.1650301080.

The State of Data in the Nonprofit Sector. (2013), *EveryAction and Nonprofit Hub*.
https://cdn2.hubspot.net/hubfs/433841/The_State_of_Data_in_The_Nonprofit_Sector.pdf

Yoon, A., Copeland, A. and McNally, P. J. (2018). Empowering Communities with Data: Role of

Data Intermediaries for Communities' Data Utilization. In L. Freund (Ed.), *Proceedings of the Association for Information Science and Technology* (pp. 583– 592.) Hoboken, NJ: Wiley. https://doi.org/10.1002/pra2.2018.14505501063.