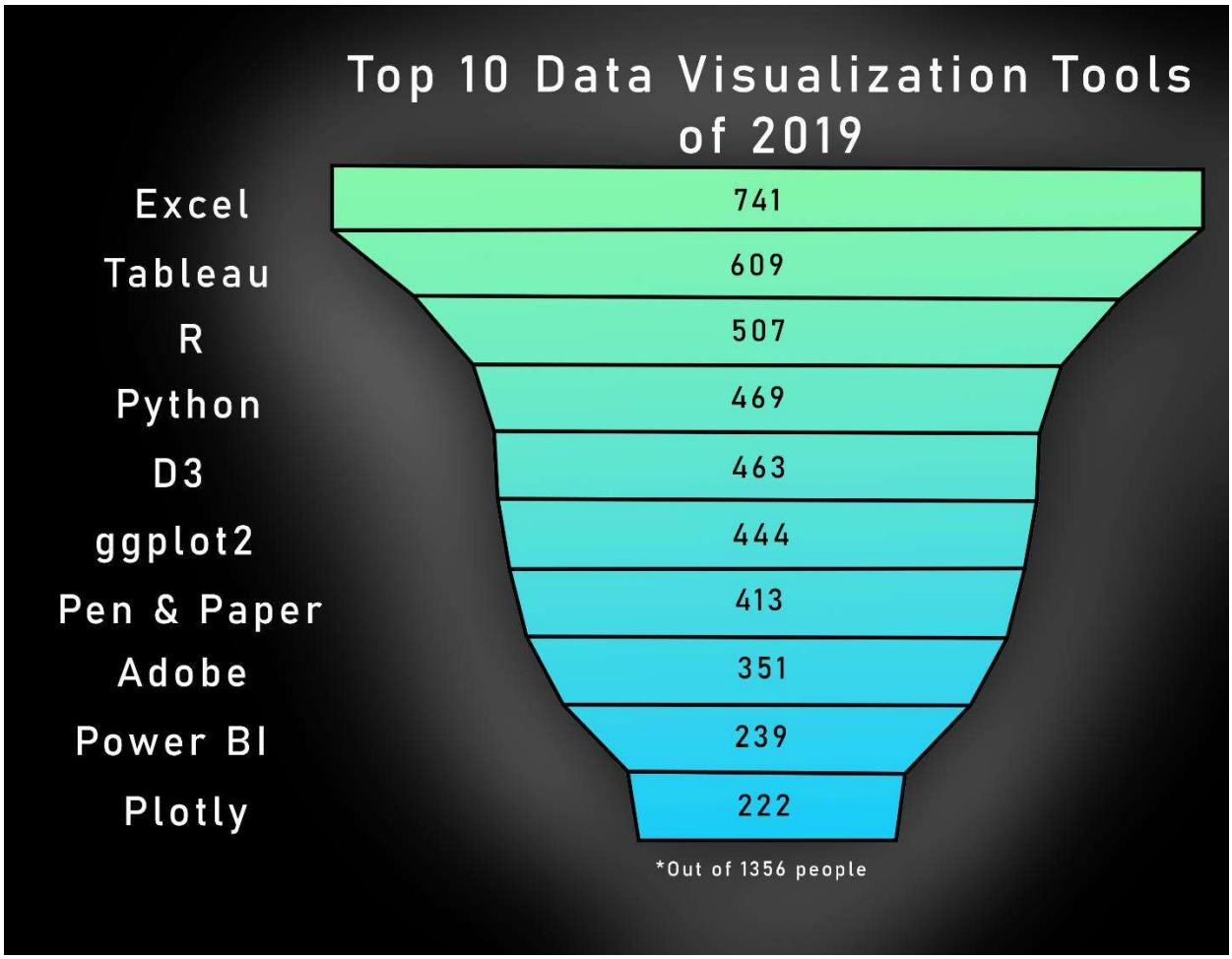
Preparing For Your Data Visualization Career



Fantastic Four

Name (full name)	Purdue Email address
Alex Silalahi	asilalah@purdue.edu
Liam McGuire	mcguir53@purdue.edu
Oliver Zink	ozink@purdue.edu
Gopika Nair	nair99@purdue.edu

Project URL:

https://sites.google.com/view/cgt-270-fantastic-four/home

Table of Contents

Introduction....4
Background....4
Questions....4
Problem Statement....4
Methodology....5
Results....6
Discussion (What's the story?) and Conclusion....9
References....11
Appendix A – Resources Used....12
Datasets....12
Tools used....12
Appendix B – Project Web Page....12
Appendix C – Percent Contribution....13
Group Contributions....13
Individual Contributions....13
Appendix D – Individual Contributions....14
Team Member #1: Alex Silalahi....15
Team Member #2: Liam McGuire....16
Team Member #3: Oliver Zink....17
Team Member #4: Gopika Nair....18
Appendix E - Diversity Statement....19
Appendix F – Team Consensus....20
Team Consensus....20

Preparing For Your Data Visualization Career

Introduction

In this report, we will explore the various skills, assets, and experiences future Data Visualizers should consider before going into the field. Specifically, we will be looking to see if there is a correlation between someone's degree and how much they make and/or a correlation between someone's degree and job position, top 10 most common used tools by Data Visualizers, and whether being self-taught vs. going to college affects their salary. Through this, we will determine the best course of action to become a successful Data Visualizer.

Background

In the dataset we selected, data_repub 2019, we were given access to survey results from data visualizers who are already in the industry. In this survey, individuals were asked about their salary, level of education, degrees acquired, identity, and career-specific questions that highlighted their personal experience in the data visualization field. We chose this dataset because, as college students, we were drawn to the idea of having access to peoples' real-world experiences as they navigated a career they worked to get into. We felt that the survey responses were informative of what is expected from a data visualizer and could best be put to use by spreading the information to those who need it. In this case, we believe our audience is future data visualizers that are looking to find a successful career and are in the process of/need direction on how to accomplish that.

Questions

The main question we are addressing is what future Data Visualizers need to know to prepare for their career.

- Is there a correlation between someone's degree and how much they make?
- Is there a correlation between someone's degree and job position?
- What are the most common tools used by Data Visualizers in the field?
- Is going to college more beneficial salary-wise than being self-taught? Is there a big difference or not?

Our audience consists of people who want to go into Data Visualization and are curious to know what skills, assets, and experiences they might need to succeed in this field. The problem we are addressing is giving people the information they need to not only prepare for their future career, but to prepare themselves for competitiveness in said field. Though schools are implementing more diverse classes such as Data Visualization and Data Literacy, providing them with additional information from Data Visualizers in the field will help further prepare them on what is expected.

Problem Statement

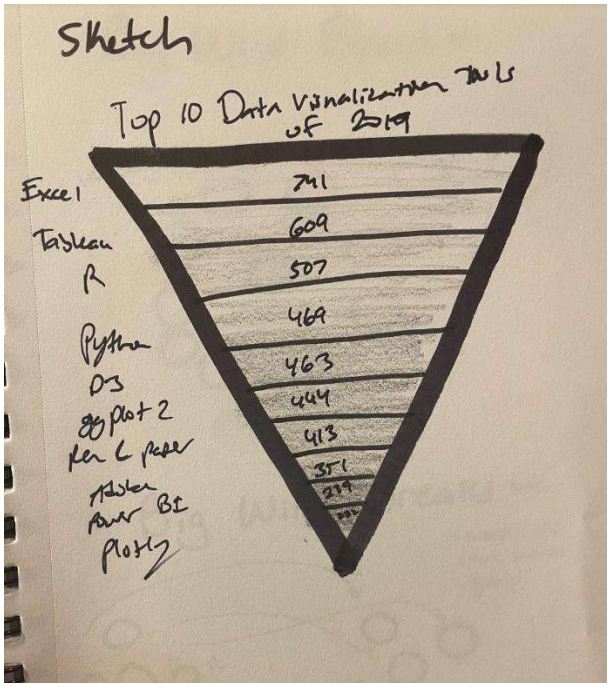
The reason this is significant is because certain individuals will want to learn more about their field and how much they learn, along with being competitive to reach these positions. It's important to visualize data for people who want to know more about their desired field. What we did differently in comparison

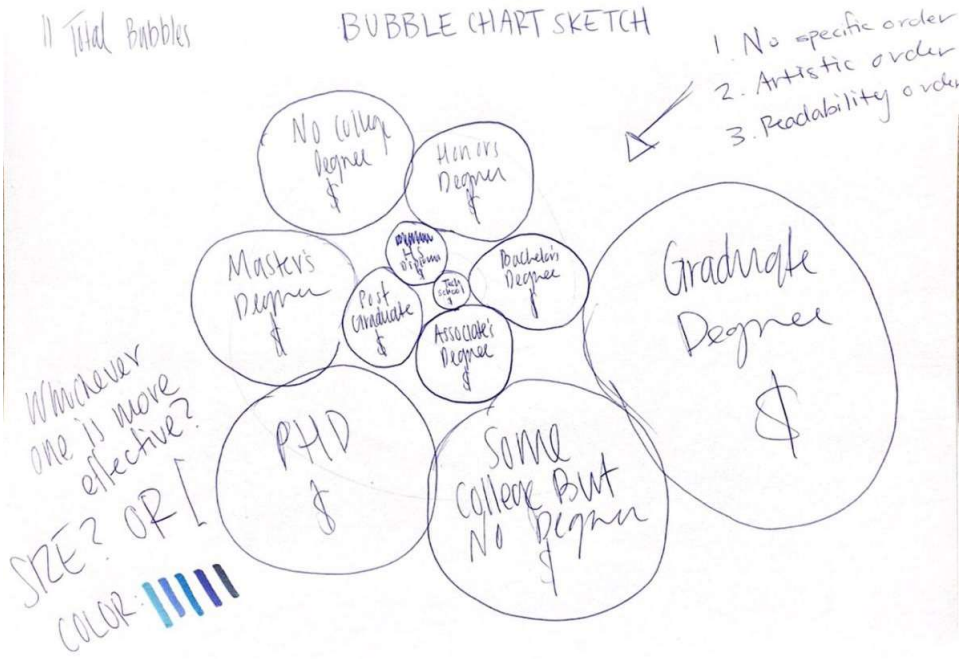
to other groups is found in the questions we set out to answer. Some groups were interested in finding out how different job positions affected an individual's experience with the data visualization process, while other groups focused more on demographics in the data visualization field and how a specific group of people were affected. Our team, on the other hand, wanted to explore how *future* data visualizers would be affected by our findings.

Methodology

When we were first given the chance to pick our dataset, we were overwhelmed by the options laid before us. Before we decided on "data_repub_2019", we went through each sheet and looked for data that intrigued us. We eventually decided on data_repub_2019 because we had a lot of questions surrounding the survey data we were presented with. We were specifically drawn to the 2019 data because of its accurate representation of the Data Visualization field prior to Covid-19. The data dealt with topics like degree of education, salary, job title, and many other things; all of which were variables that could be affected in the workforce after the pandemic. So, after we had chosen the dataset, we worked to come up with an overarching question and 4 sub-questions that each person from the team would set out to answer through a visualization. Parsing the data proved to be more difficult than we assumed. Just by overlooking the data, we realized that a lot of survey responses were sent in through text-box responses instead of multiple-choice. This affected our parsing stage because we each had to go through around 1360 responses and manually group responses that were similar. While this step is pertinent to gauging a better understanding of our data, it was also a necessary step in terms of program readability. This makes sense considering each step in the data visualization process affects the next and those before it since it's iterative. In this case, thoroughly parsing our data made the visualization stage much smoother. For example, at this point in the semester, we have learned that Tableau is extremely helpful for creating visuals but comes at the cost of cleaning your data beforehand; it took rewriting multiple strings for Tableau to register what we were trying to do. When we got to the mining stage, we knew we wanted to divide and conquer our main question into 4 sub-questions. This would allow each of us to explore a question that intrigued us and create a visual that answered it. Much like our experience with the parsing stage, filtering such a large dataset was quite tedious. While this process looked different for all of us, we each saw a decrease in viable responses after filtering through a total of around 1360 survey responses. A lot of the filtering was standard procedure like disregarding empty cells and changing lower case titles to uppercase. Then there were more specific clean ups like having to group synonymous responses together from questions that allowed participants to type their answers with no restraints. Lastly, there was the filtering of extraneous responses that didn't directly relate to a sub-question we were trying to answer. After we filtered our data, we sketched what we wanted our visuals to look like. This worked as a guide during the visualization process as it can be tricky memorizing what you want your visual to look like while simultaneously trying to manipulate a software program. After we all created our first round of visualizations, we critiqued each other, received critiques on presentation day, then revised our visuals. The feedback we received was critical and valuable as it helped us refine any obscurities that would steer from our main focus: creating visualizations that inform future Data Visualizers of what they need to know to prepare for their career.

Example sketch:





Results

Choose one of your team's "BEST" visualization and insert it here. This visualization should be the best representation of the team's effort. Provide a figure caption. If the team would like to include additional visualizations, add them to this section. Make sure each visualization has a figure caption AND includes the name of the person who created the visualization. Also make sure the appropriate page layout (portrait or landscape) is used as well as the appropriate chart type and layout (see best practices for visualization data and data visualization checklist).

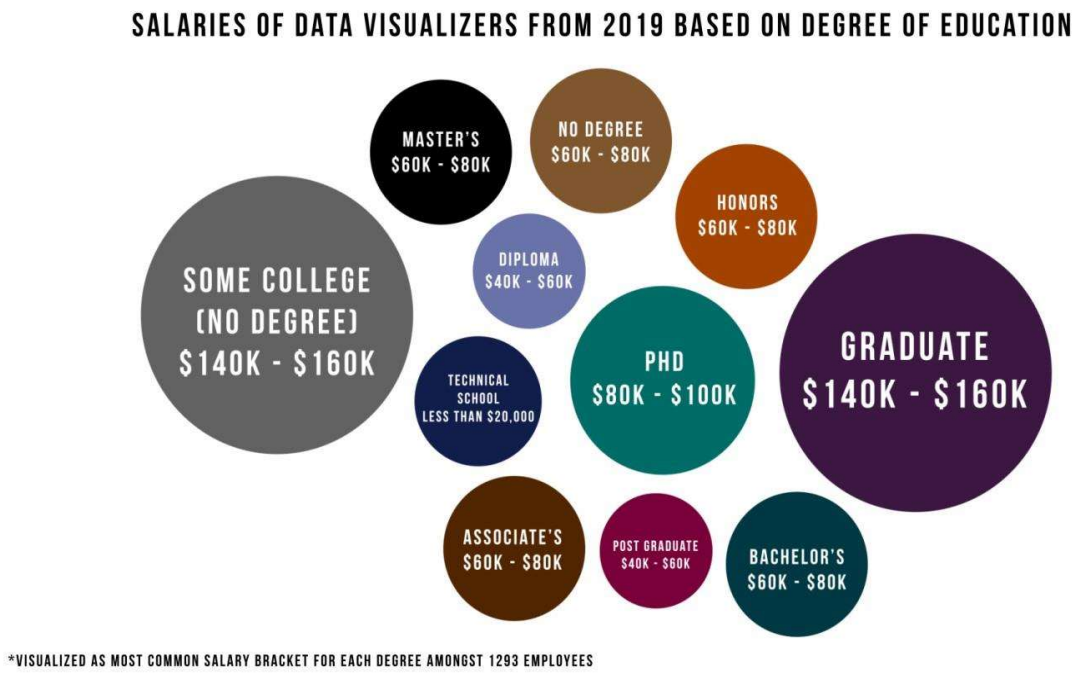


Figure 1 (Gopika N). This bubble graph depicts the relationship between a Data Visualizer's degree of education and how much they were earning as of 2019. While this data was collected from a survey that 1360 employees took, not everyone replied with what degree(s) they hold. Responses that listed their salary but not their degree were disregarded as to not affect the rest of the data.

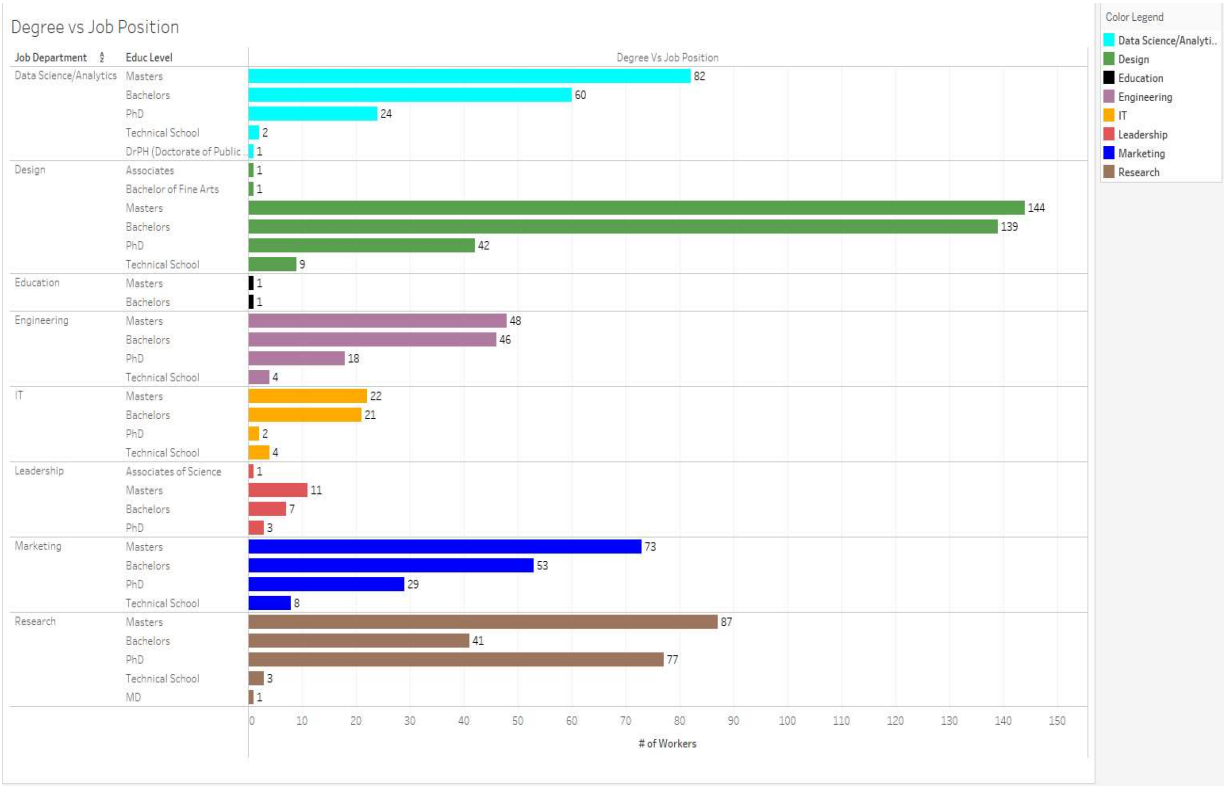


Figure 2 (Liam). This bar graph depicts a comparison between a worker's degree and job position, with a sample size of 150 workers. We wanted to answer the question of whether the better a worker's degree the more job opportunities and high-end positions they'll be eligible for.

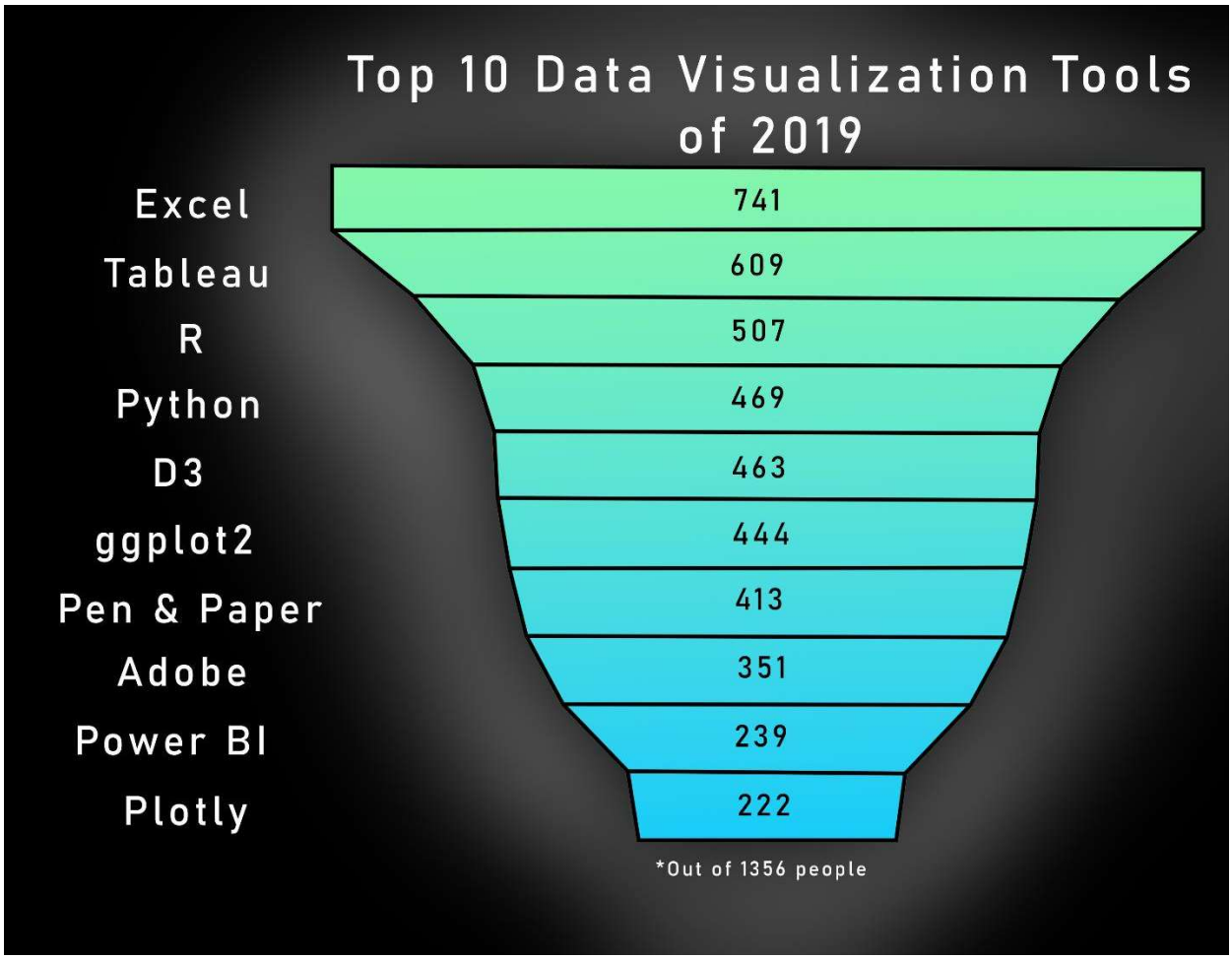


Figure 3 (Alex S). This funnel graph depicts the top 10 data visualization tools based on data collected from data_repub_2019. 1356 people in the industry were surveyed and listed the tools they use in their day-to-day work, and it was compiled to visually represent how many out of the 1356 data visualizers use these tools. For instance, 741 out of the 1356 people surveyed use Excel, while only 222 out of the 1356 use Plotly.

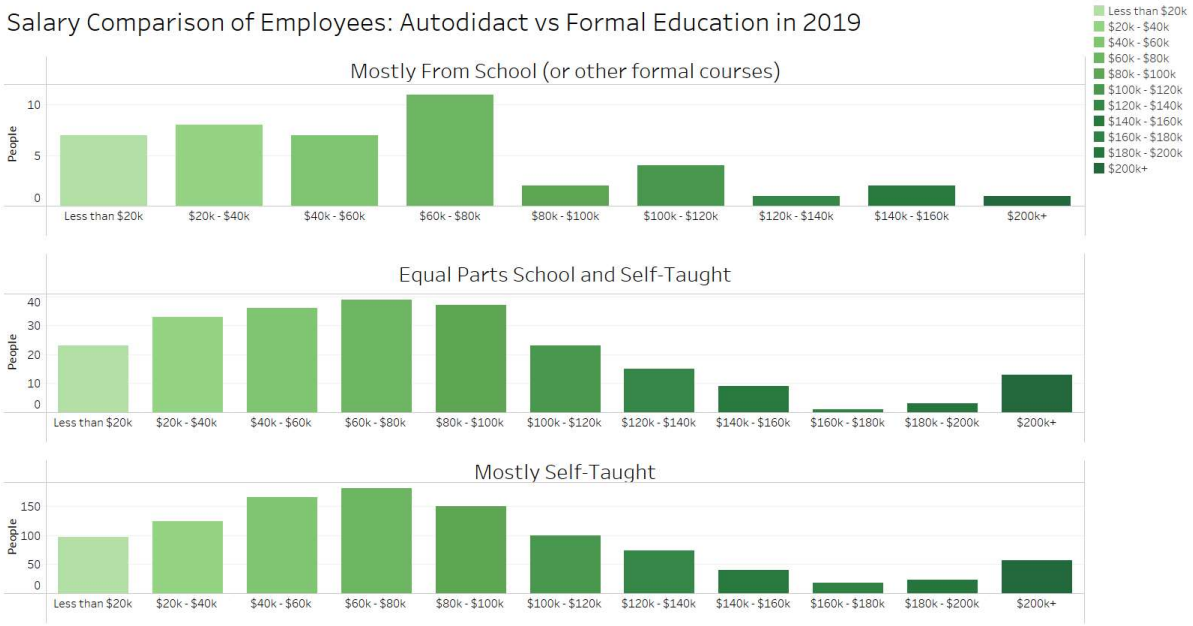


Figure 4 (Oliver). For my visualization I wanted to make a comparison-based bar graph to show the differences between the styles of education. Showing various sources of self-taught workers being paid just as much if not more than formally educated workers.

Discussion (What's the story?) and Conclusion

Discuss your results (the figures in the Results section). Do your visualizations address the problem stated in the Problem Statement Section? Explain. What insights did the team uncover? What recommendations can be made based on these insights?

In "Data Science for Undergraduates: Opportunities and Options" (2018), the three academies (sciences, engineering, and medicine) claimed that "Data science is emerging as a field that is revolutionizing science and industries alike", and that "Work across nearly all domains is becoming more data driven" (1). 4 years later, the field has only continued to evolve. As a result, universities have been implementing various programs to prepare students for the variety of roles they can play in the field such as providing online classes and training programs to practice important skillsets, majors, minors, and even certificates in Data Visualization that can be accompanied by another major. Even at Purdue University, Data Visualization is required for every CGT major. This ensures that everyone receives an introduction to the importance of accurately visualizing complex datasets, processing data, creating effective visualizations, and learning how to use these new-found skills in a variety of different fields.

Though practicing and utilizing these skills is strongly suggested for those preparing to enter the field of Data Visualization, what's also important is knowing what the field is like. What are Data Visualizers' experiences in the field? What skills and assets will future Data Visualizers need? Should they go to college, if so, is there a big difference between pursuing a bachelor's degree versus a master's degree? Therefore, it's not only vital to practice the skills needed for the industry, but to also learn more about said industry from people already in the field. That way, future Data Visualizers can learn what they need to become successful in the field and determine the best course of action — which is our main focus.

Our visualizations address the problem of future data visualizers not knowing how to prepare for their desired job or what to expect. In figure 1, for someone planning to enter the Data Visualization field,

having access to a visualization that details which degree can maximize your salary is beneficial to your future-planning process. Our audience will find a similar use with figure 2 as it lists the most common job positions of a data visualizer based on their degree of education. While figures 1 and 2 emphasize how influential obtaining a specific degree can be on your salary or job title, figure 4 provides insight on if a degree is really *necessary* in terms of the salaries non-degree holders receive. These 3 figures ultimately connect with figure 3 where Data Visualization tools are talked about. With figure 3, future data visualizers can learn about the tools they will be expected to know once they enter the industry. This is especially important considering, as stated previously, high schools and colleges are increasingly offering courses that teach students these skills. All in all, our visualizations work to inform our audience and prevent them from being left in the dark about how to plan for their future.

With all our 4 visualizations, we came across several insights. With figure 1, we surprisingly found that there is no strong correlation between the degree one pursues and overall success. For instance, employees with only *some* college experiences were making more than those with a PhD degree. It's clear from the 5 pink bubbles that the most common salary for someone in this field was $60k - $80k. Therefore, while there is a universal understanding that the higher degree of education one acquires, the more job prospects and income they are met with, this idea is challenged through these results. With figure 2, the most common job position that people got with a bachelors and masters was Data Science/Analytics, while for PhD, more people went into research. However, more people with a Master's got into Data Science/Analytics than those with a Bachelor's. Therefore, this connects to figure 1 since a Master's got a lot of jobs and PhD wasn't making the most. Next, with figure 3, Excel was the most common visualization tool in 2019, with 741 out of the 1356 people surveyed using it, while Tableau comes in second with 609 out of the 1356 people surveyed using it. These top two tools are what we are using in our class right now. Giving us access to these tools in school that are already used in the industry allows students to practice and enhance their skills with such tools, preparing them for their career. In addition, despite how technology continues to evolve and continues to be integrated in everyday life, pen and paper was still one of the top 10 tools used for data visualization, revealing how pen & paper and sketching is still important. Last but not least, with figure 4, the majority of the answers in the survey were mostly self-taught, and all 3 bar charts were slightly skewed to the right. Therefore, there is no direct correlation between having an education or being fully self-taught.

As a result of our findings, we came up with a few recommendations to give to the audience. One recommendation is that, like any career you go into, passion and interest should always be the priority above salary. It's evident from the results of figure 1 that you have the potential to succeed regardless of what degree you acquire, so you might as well go into a field you truly care about. From figure 2, we also recommend that if you're interested in learning more about the many job opportunities tha come with getting a degree, you should research more and talk to your advisor about your career plan. Moreover, when it comes to the tools from figure 3, if you're interested in going into Data Visualization or creating visuals in the future, it's recommended to either practice using those tools yourself or taking classes where you can utilize said tools and apply those skills. Lastly, when it comes to a self-taught vs. formal education, it is recommended to focus on what you *can* get from whatever form of education you choose, and making the most of it rather than focusing on where the education will lead you.

When it comes to our findings, it's also important to note the potential weaknesses since no finding is perfect. Our potential weaknesses stemmed from the assumptions we made. We assumed that because the people surveyed were free to type their responses, they were being honest. Secondly, our data is from 2019, so we assumed that these findings still applied after the height of the Covid-19 Pandemic.

References

If references are listed, make sure they are cited in the body of the document. See Purdue Online Writing Lab for how to cite and list full citations. Improperly cited work will be treated as plagiarism and handled accordingly.

National Academies of Sciences, Engineering, and Medicine. 2018. Data Science for Undergraduates:

> Opportunities and Options. Washington, DC: The National Academies Press. https://doi.org/10.17226/25104.

Appendix A – Resources Used

Datasets

List the name of the data set provided and a description of the additional data set acquired.

Datasets must be available on the project web page.

All data must be accessible to the public. No paid data sources allowed.

Tools used

List all tools used in the project and a brief description (see the *examples* below); update accordingly.

Tool/Application	**Description**
Excel	Data cleaning
Tableau	Data visualization
Adobe Photoshop	Data visualization
Google Sites	Web development
Premiere Pro	5-minute video
Google Sheets	To put dataset on Google Sites, converted Excel to Google Sheets

Appendix B – Project Web Page

The project web page will be an extension of the final report. You will be allowed to add content to the project web page up to last day of classes. The project web page should contain *(at a minimum)* the following sections:

About The team

List each team member, provide a short bio (150 words or less) for each team member, Provide photo (headshot only) dress appropriately.

The Hackathon Challenge

Describe the team's focus/goal related to the challenge, Who's the audience? What assumptions are made?

Methodology

Describe the team's data visualization workflow and process.

Deliverables

5-minute video (1 pt deduction for each minute over if over 5:00:00 minutes), Hackathon Report, Team agreement (signed by all team members)

Results

This the team's time to shine! Visualizations created by the team that support the team's solution to the challenge, Visualizations must be relevant to the question(s) the team is answering in regards to the visualization challenge.

Conclusions

What insights are presented? What recommendations did the team make?

Appendix C – Percent Contribution

Group Contributions

In this section list the tasks that were completed by all team members for example: contributed to the data visualization process, brain stormed topic ideas, served as rotating team leader, contributed content to the short story (summary), contributed content to the 5-minute video, reading the final deliverable before submission,

Individual Contributions

In the table below list each team member's full name, their contribution (body of work) and their % of the work completed. The total must add up to 100%.

Team Member	**Contribution**	**Contribution**
Gopika Nair	*Created slides for 5-mintue video, wrote the script for 5-minute video, served as rotational team leader twice, wrote the methodology & background in the documentation, and created Visualization 1 (bubble chart).*	*30%*
Alex Silalahi	*Created and organized website, slides, and hackathon report. With documentation, wrote out the introduction, questions, and story and conclusion. Also created Visualization 3. Served as rotational team leader twice but was the overall organizer for tasks. Acquired addition article for the story.*	*40%*
Liam McGuire	*Created Visualization 3, created the document of team's group script*	*5%*
Oliver Zink	*The 5-minute Video and Visualization 4, and the diversity statement.*	*25%*
	Total contributions must equal 100%	100%

Appendix D – Individual Contributions

In this appendix each team member must contribute a one-page document relating the team's topic/data. The one-page document must contain: (1) a description of the problem, (2) a comparison to the team's findings with insights related to the hackathon data (3) a visualization to support items (1) and (2).

Each person should create their individual page **(1-page only)** and make it available to the designated team member who will upload the final document.

This will be viewed and assessed as part of each person's individual contribution.

Leave this page as is.

Start adding individual page content on the next page.

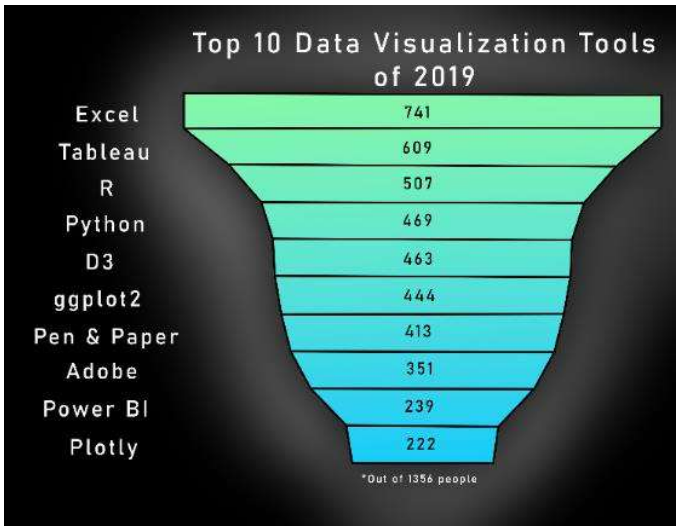
REMOVE any blank pages before submitting.

Team Member #1: Alex Silalahi

Group Topic: What Future Data Visualizers need to know to Prepare for their Career
Your Topic/Question: What are the most common tools used by Data Visualizers in the field?

Describe the diversity YOU bring to the group (150 words or less):
The diversity I bring is that I am double majoring in Game Development and Design and Creative Writing, so I love combining both the technical and creative side of things. I am also familiar with using web starters like Weebly and WordPress for creating a quick website. In addition, I know how to use Adobe Illustrator and Adobe Photoshop too.

Include your story and visualization below (**do not go over one page**). Single spaced, 11-pt font, Times New Roman.



I decided to focus on the most common tools used by Data Visualizers in 2019 because I was curious to not only see if we in CGT 270 are also using the same tools, but also to see what other visualization tools are used in the field. Since we are focusing on preparing future Data Visualizers, I felt that one of the most important pieces of information is knowing what tools future Data Visualizers should be familiar with. Therefore, I gathered all the visualization tools that the people surveyed listed and put them in a funnel graph to depict the top 10 visualization tools. I wasn't surprised that Excel and Tableau were at the top, with more than half of the visualizers surveyed using them. After all, these are the tools that we are learning to use in CGT 270! But I was also surprised by some of the other tools that I've never heard of before, but reached the top 10, such as D3 and ggplot2. Therefore, based on my findings, giving us access to tools in school that are already used in the industry allows students to practice and enhance their skills with such tools, preparing them for their career. The visualization also depicts that, despite how technology continues to evolve and become integrated into our day to day lives, pen & paper is still one of the top 10 tools for data visualizations, showing how the sketching process is still applicable today.

Besides working on visualization #3, I was also in charge of creating the website and the slides for our presentation. I also contributed to this documentation by finding an extra article and working on introduction, questions, and the discussion and conclusion portion. Though I did serve as rotational leader twice, I was the overall organizer for tasks and making sure we were completing everything on time and coming to our weekly meetings outside of class.

Team Member #2: Liam McGuire

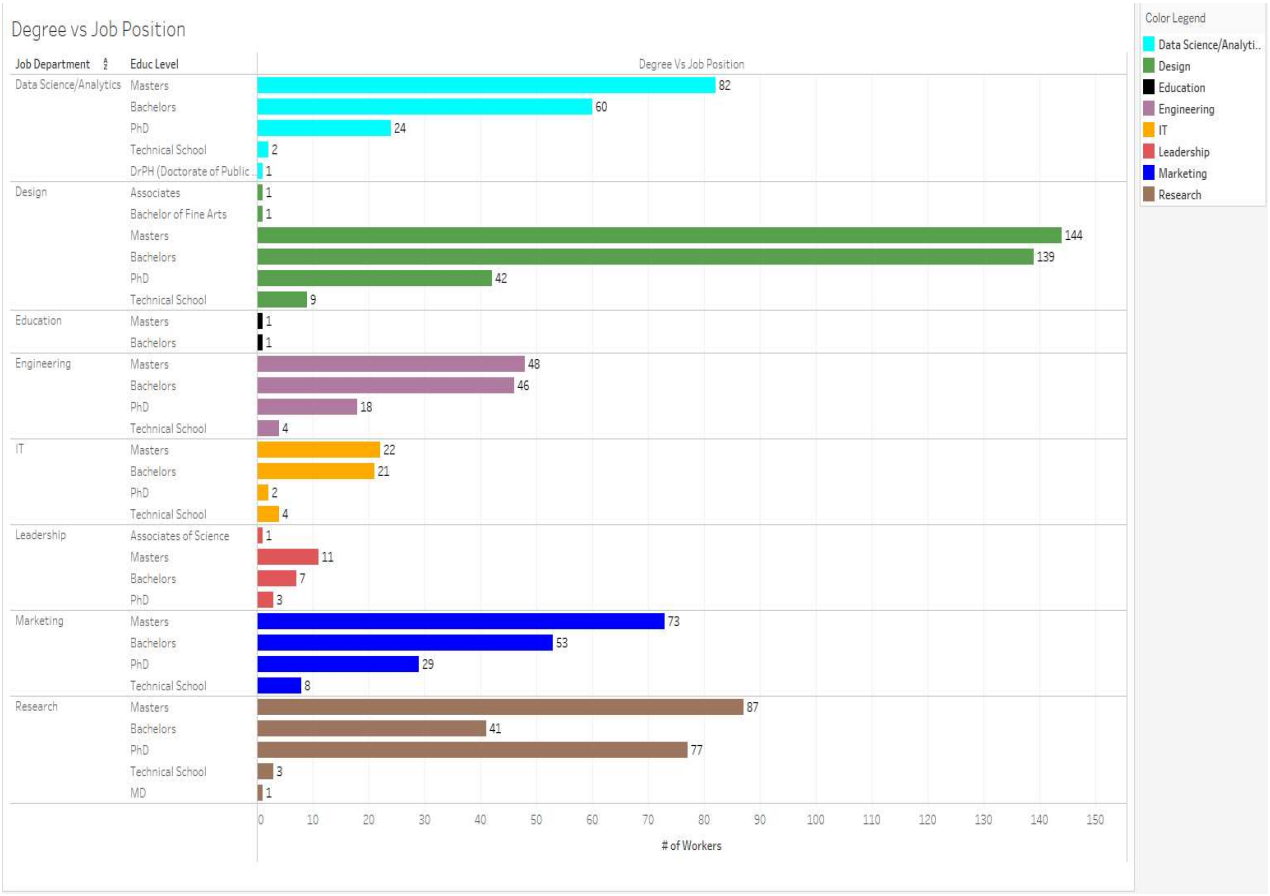
Group Topic: How to Prepare for a Successful Data Visualization Career
Your Topic/Question: Does a person's degree have a heavy factor in determining their job positions and job opportunities?
Describe the diversity YOU bring to the group (150 words or less):

I am familiar with some computer programs such as Adobe Photoshop, Microsoft Excel, and Tableau Desktop. I bring a diversity of experience. I'm from Indiana.

Include your story and visualization below (**do not go over one page**). Single spaced, 11-pt font, Times New Roman.

I decided to go over the topic of how certain degrees can lead to more job opportunities.

My data shows that Bachelors, Masters, and PhD degrees tend to open up more opportunities for certain job opportunities and job positions. The more time you put into pursuing a degree and the better it is, the more job prospects you will have in the future.

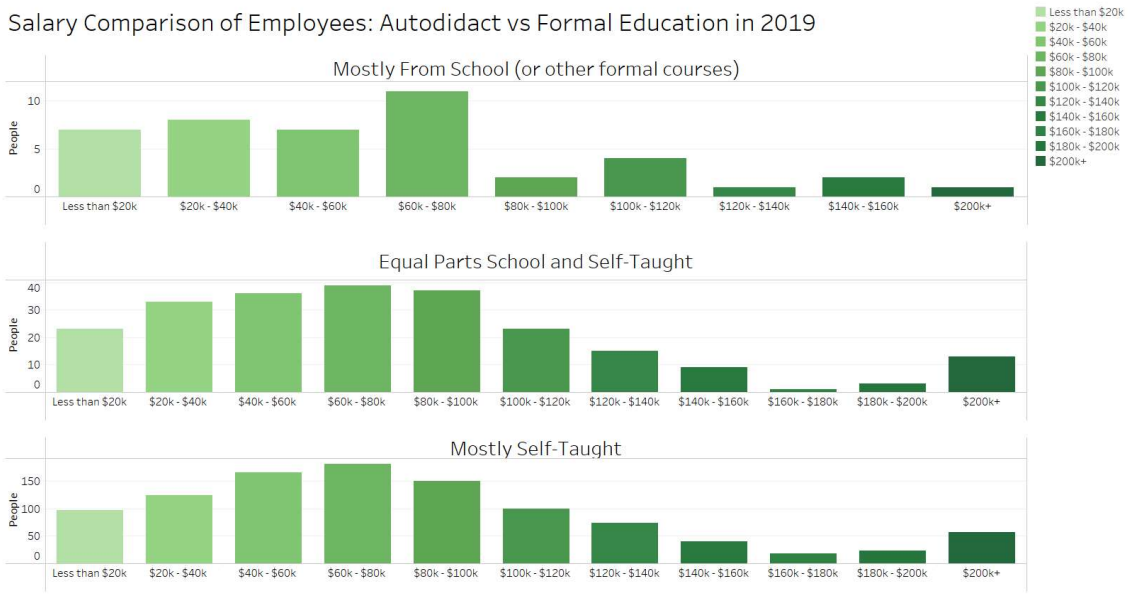


Team Member #3: Oliver Zink

Group Topic: How to Prepare for a Successful Data Visualization Career
Your Topic/Question: Is going to college more beneficial salary-wise than being self-taught?
Describe the diversity YOU bring to the group (150 words or less):

I am very familiar with Digital Visual Design and have a lot of experience with photoshop and premiere pro.
I also have ADHD.

Include your story and visualization below (**do not go over one page**). Single spaced, 11-pt font, Times New Roman.



I decided to focus on the comparison of salaries between the style of education people acquire from self-taught to formal education (with the middle ground being both).

My data shows that while mostly formal education would be the natural go to. It doesn't effect the actual outcome of your salary vs making sure the education is of actual quality. Showing a much more preferred swing to self-taught vs full formal educations.

Team Member #4: Gopika Nair

Group Topic: What Future Data Visualizers Need to Know to Prepare for Their Career

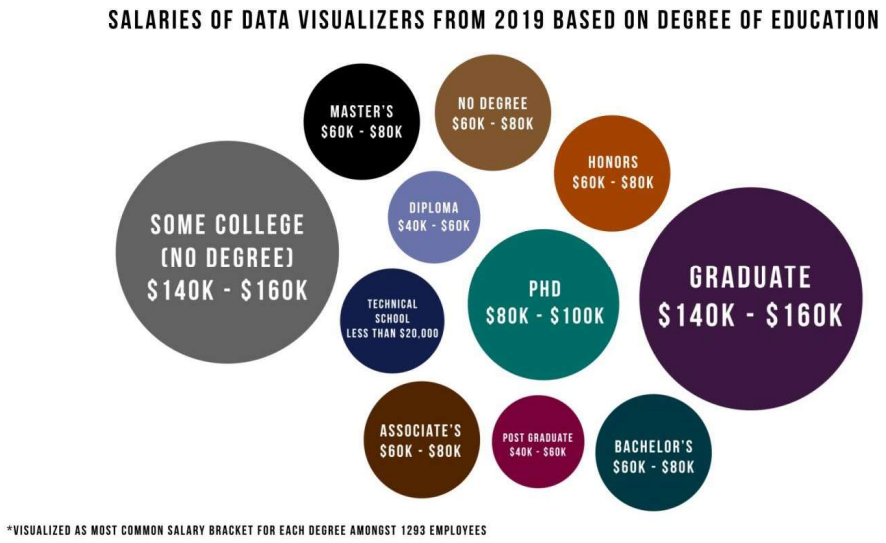
Your Topic/Question: Is there a relationship between someone's degree of learning and how much they make/what job position they find themselves in?

Describe the diversity YOU bring to the group (150 words or less): As an Animation and Visual Effects major, I bring my creative and analytical side to the group. I have prior knowledge in Adobe programs and have recently learned 3D modelling. After keeping up with an online portfolio for most of my classes, I have also learned to navigate website tools like Weebly.

Include your story and visualization below (**do not go over one page**). Single spaced, 11-pt font, Times New Roman.

I wanted to explore the relationship between a data visualizer's degree of learning and the salary they earn because I felt as though my results had the potential to debunk an age-old question. We are taught from a young age that success comes from education, but is that true? With that philosophy, if correlation were to equal causation, then the more education you receive, the more success you should gain. However, based on my results from data visualizers in the industry during 2019, the aforementioned claim is not true. From my results, I learned that those with a graduate degree earned the most amongst data visualizers ($140-160k). However, that's not the interesting part. The surprising part is that those with some college experience but no official degree were earning the same as those with a graduate degree. This begged the question of if pursuing higher education was really worth it? I think another interesting takeaway was that there were varying degree holders earning from the same salary bracket. For example, the most common salary bracket for someone in the Data Visualization field is $60k-80k. The degrees in this category ranged from no college degree (including no courses taken) all the way to Master's degree holders. If employees were earning the same or even double regardless of if they had graduate experience or not, then maybe degree of education isn't a variable that ensures success for this field.

Other than creating my visual, I created some of the slides for our 5-minute video, wrote the script for 5-minute video, served as rotational team leader twice, and wrote the methodology & background in the documentation.



Appendix E - Diversity Statement

Some of the most enlightening outcomes are generated by diverse teams working together to solve complex problems. What does diversity mean and why is it important? Merriam-Webster defines diversity as: 1) the quality or state of having many different forms, types, ideas, etc., 2) the state of having people who are different races or who have different cultures in a group or organization. When solving complex problems having adequate representation is important. In the context of the hackathon, diversity could mean (but is not limited to): varied perspectives, varied points of view, different academic majors represented, different academic levels (Freshmen, Sophomore, Junior, Seniors) on the team, different ethnicities (state this professionally). Having a diverse team from different backgrounds can boot engagement and productivity and make us smarter (read short article: "How diversity actually makes us smarter").

In the space below, provide a statement describing the group's diverse make up and how the diversity of the group contributed to the outcomes of the team's deliverables for the hackathon. Every team member must contribute to the development of the diversity statement.

The members that make up Team Fantastic Four have a diverse group of majors including Animation and Visual Effect Composition, Creative Writing, Design/General Communication, and Game Development and Design. With one person being online, it added a different view from another campus about the whole project. Oliver has been in sports for all middle school and high school, including swimming and tennis. Alex co-created the badminton club in her high school and is familiar with coding and websites. Liam, being born and raised in Indiana, gives him the standard American values. Gopika had helped produce a film that had won an award at a state-wide event in 2019. The different skills and knowledge gave us a nice workflow where everyone could hold their own space and work hard.

Appendix F – Team Consensus

Team Consensus

I have read and approve of the content as a representation of the team's work and my contribution.

Oliver Zink	*Oliver Zink*	4/24/2022
Print Team Member Full Name	Signature	Date

Alex Silalahi	*Alex Silalahi*	4/24/2022
Print Team Member Full Name	Signature	Date

Gopika Nair	*Gopika Nair*	4/24/2022
Print Team Member Full Name	Signature	Date

Liam McGuire	*Liam McGuire*	4/24/2022
Print Team Member Full Name	Signature	Date

Save this document as:

HackathonTeamName_CGT270Spring2022_FinalReport.pdf