The Values of Professional Experience in Data Visualization

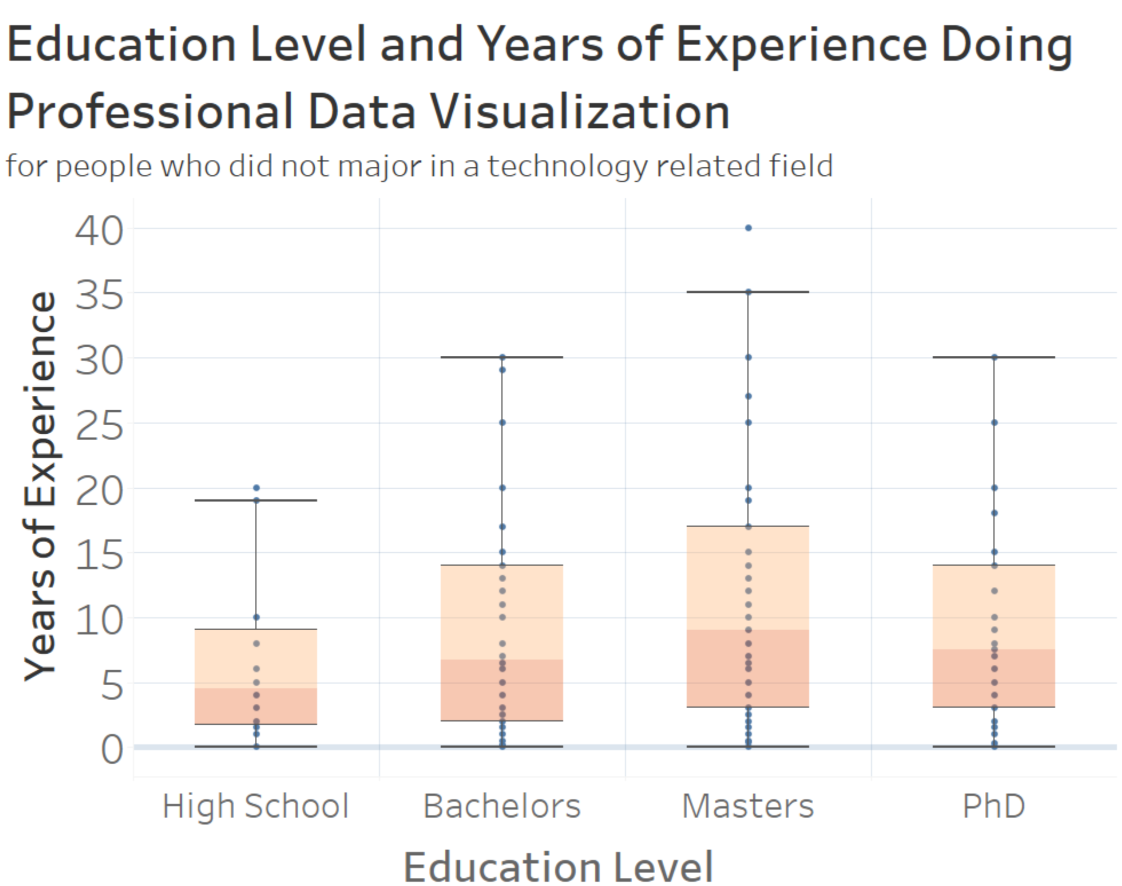

Figure 1. The visualization, created by Rae Fu, depicts the spread of years of experience depending on their education level. The median, third quartile, and fourth quartile all show the same pattern of increasing years of experience as education level increases, with the exception of Ph.D.

Team name

Name (full name)	Purdue Email address
Gabby Willard	gwillard@purdue.edu
Rae Fu	fu345@purdue.edu
Ace Glover	gglover@purdue.edu

Project URL: https://teamrocketviz.weebly.com/

Table of Contents

The Values of Professional Experience in Data Visualization

Introduction

There are many people in the world who currently have jobs that involve data visualization, however many of those people did not major in that field of study as an undergraduate. An example taken from phData describes that "Most current leaders in data analytics do not have majors in Business Analytics," depicting that people do not necessarily end up in a career that matches their major (Christensen, 2022). We wanted to know how people are currently doing with jobs involving data visualization without majoring in technology-related fields.

Background

Our team chose to focus on the benefits of having a professional data visualization job. We wanted to focus on collecting data that was captured in a year that was fairly recent but did not have data alteration results due to being affected by the COVID-19 pandemic. With this in mind, we decided to explore the Data Visualization Society's 2021 State of the Industry Survey Data. Inside the survey data, there is a spreadsheet called "data_repub_2019" which asks the survey participants a wide variety of questions about the data visualization field. Question topics from this spreadsheet that we chose to zone in on include the education level of data visualization workers, the confidence level in data visualization skills/tools that data visualization workers rate themselves with, and the annual pay compared to the years of experience for data visualization workers.

Questions

Is being exposed to data visualization important to non-technology degree people for their careers? Is there a correlation between education level and years of experience? Does the data visualization professionals survey reveal a correlation of annual pay to job experience to education levels and majors?

This discussion is targeted toward non-technology degree people seeking a new job. This is especially towards college students in non-technology majors so they might be better prepared when faced with these unexpected job tasks. Since our group was interested in data visualization professionals who did not obtain a college degree within the technology/computer science field, we wanted to see how common this situation was in the workplace and how certain aspects of their data visualization careers correlate.

Problem Statement

The problem we are trying to solve is the issue that people assume that their future job will only involve their area of study, however, that is often not the case. People are not familiar with jobs, and in this case, we are looking at ones that involve data visualization. People should learn about data visualization as it could be crucial towards their future job even if their area of study does not involve technology. We want people to see how data visualization skills are necessary in a variety of careers.

Methodology

We followed the data visualization process of acquiring, parsing, mining, filtering, representing, critiquing, and refining the visualizations that each of us made for the Hackathon.

The data is structured, and it is a combination of text and numbers. The survey is primary data. The data is stored within a CSV Excel file the data is also stored in a spreadsheet and is accessible through datavisualizationsociety.org.

EducUndergradMajor with parameters of categories and technology-related majors. The categories would be science, liberal arts, business, engineering, mathematics, visual, and performing arts. Average time spent each week on data visualization, data engineering, data science, designing, and data preparation with parameters of 0 to 50, 0-40, 0-60, 0-60, 0-110 hours a week. EducLevel is the education level with high school, bachelor's, master's, and Ph.D. YearsDVExperience is the years of experience doing professional data visualization, with 0-40 years. PayAnnual are ranges of annual pay, from less than $20k to $80k-$100k. Skills hold me much more than my tools do 1-5 is a rating of limited skills. Whether they were hired to do data visualization for most of their projects, with either hired to do data visualization or data visualization is only part of their job.

The dependent variables are annual pay, time spent on each data-related activity, hired to do data visualizations, and skills. The education undergraduate major and education level are independent variables.

Our audience is college students who are not aware of the importance of data visualization. We want to show them the importance of these skill sets.

The data is recorded by datavisualizationsociety.org annually through an online survey. There are only a few years worth of data so it would be beneficial if there were more records in order to show trends.

We decided to focus on how people are currently doing in jobs involving data visualization without majoring in technology-related fields.

The data is raw and is in a format that allows the creation of new categories since it is in an editable spreadsheet.

The 2019 Repub dataset covers current jobs that involve data visualization and the education of those with these jobs. The time frame is 2019, the surveyed people had these jobs in 2019. The data is for data visualization jobs.

The fields are EducUndergradMajor (string), average time spent each week on data visualization, data engineering, data science, designing, and data preparation (float), EducLevel (string), YearsDVExperience (integer), PayAnnual (string), Skills hold me much more than my tools (integer), Whether they were hired to do data visualization for most of their projects (string).

We assume that the respondents are random and have the same understanding of what a question is asking them, like they consider the same activities under data science versus data prep.

Basic descriptors for each data component are Education Undergraduate Major: strings; average time spent each week on data visualization: min: 0.0, max: 50.0, average: 10.533; average time spent each week on data engineering: min: 0.0, max: 40.0, average: 5.14; average time spent each week on data science: min: 0.0, max: 60.0, average: 5.44; average time spent each week on designing: min: 0.0, max: 60.0, average: 6.52; average time spent each week on data preparation: min: 0.0, max: 110.0, average: 7.56; Education Level: Strings; YearsDVExperience: min: 0, max: 40, average: 5.44; Annual Pay: Strings; Skills hold me much more than my tools: min: 1, max: 5, average: 3.21; Whether they were hired to do data visualization for most of their projects: strings.

The average time spent on data-related activities is similar. Education undergraduate major is nominal, average time spent each week on data visualization, data engineering, data science, designing, and data preparation are continuous. Education level is ordinal while years of experience is continuous. Annual pay is ordinal since it was shown as ranges as strings and skills are nominal because although they were labeled as 1 through 5, they had different meanings for each number which were not in a ranked order. Whether they were hired to do data visualization for most of their projects is nominal.

For the Repub 2019 dataset, it was stored all at one time after collecting all of the surveys. The distribution of the data has many values, more than 1000 data points and is dense.

After deciding on the questions (Is being exposed to data visualization important to non-technology degree people for their careers? Is there a correlation between education level and years of experience? Does the data visualization professionals survey reveal a correlation of annual pay to job experience to education levels and majors?)

We filtered the dataset so that only those with majors that were not technology-related would be left.

For each specific question, we deleted the irrelevant variables and null values before creating necessary categories and then sketching out a visualization.

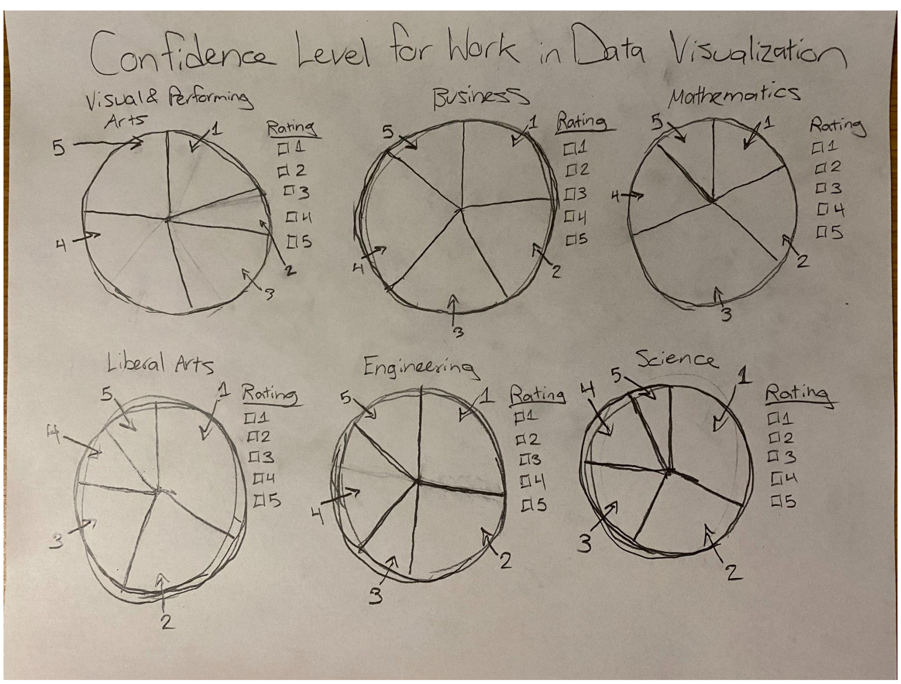

Figure 2. This sketch, made by Gabby Willard, shows the original concept for a visualization displaying how confident people working in data visualization are about their data visualization skills.

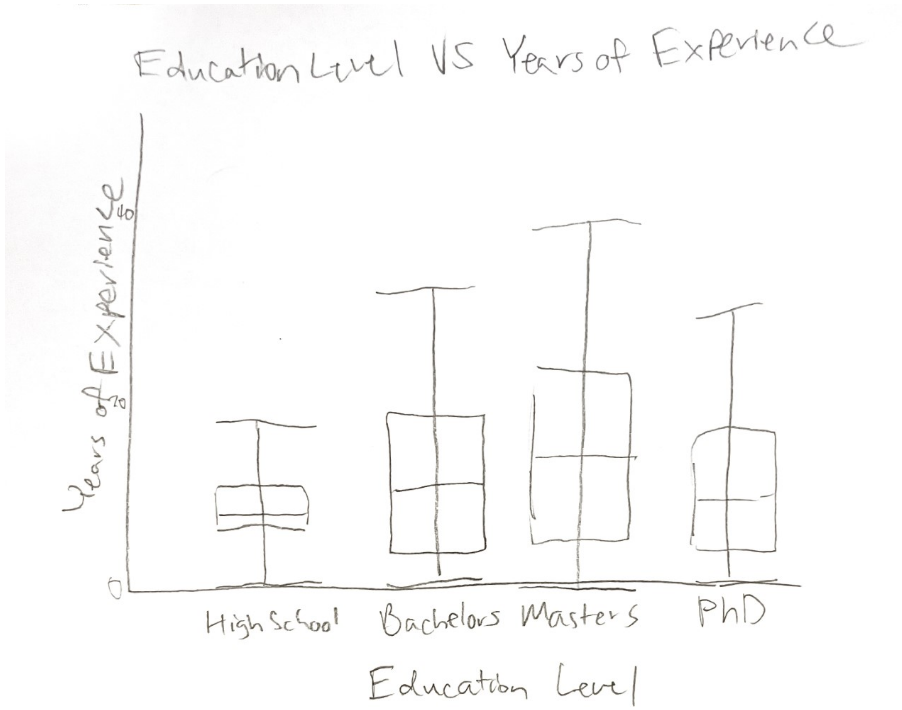

Figure 3. This sketch, created by Rae Fu, depicts the spread of education level with years of experience.

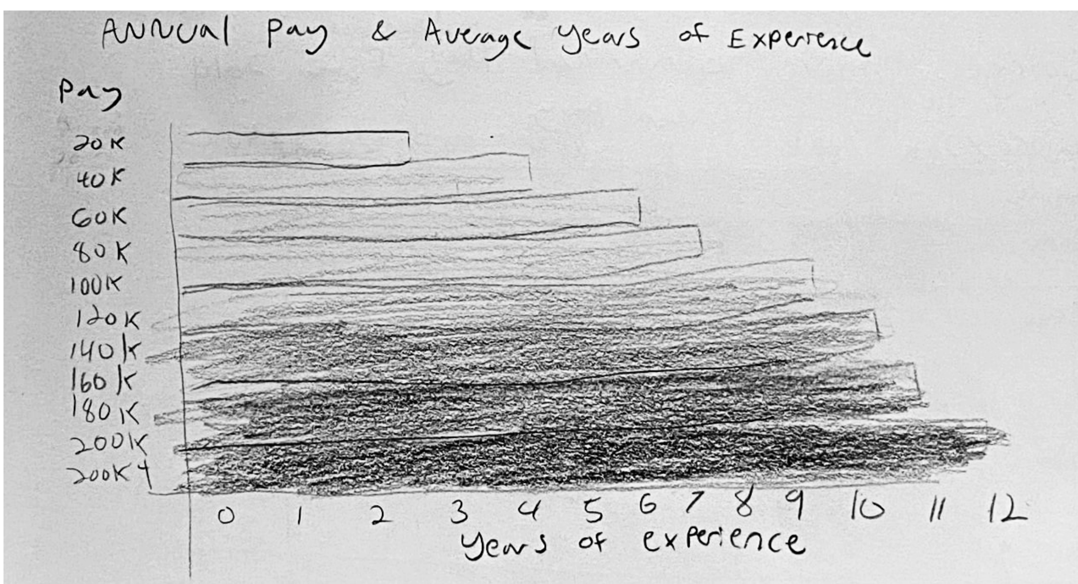

Figure 4. This sketch, created by Ace Glover, shows the average years of experience doing professional data visualization compared to annual pay.

Results

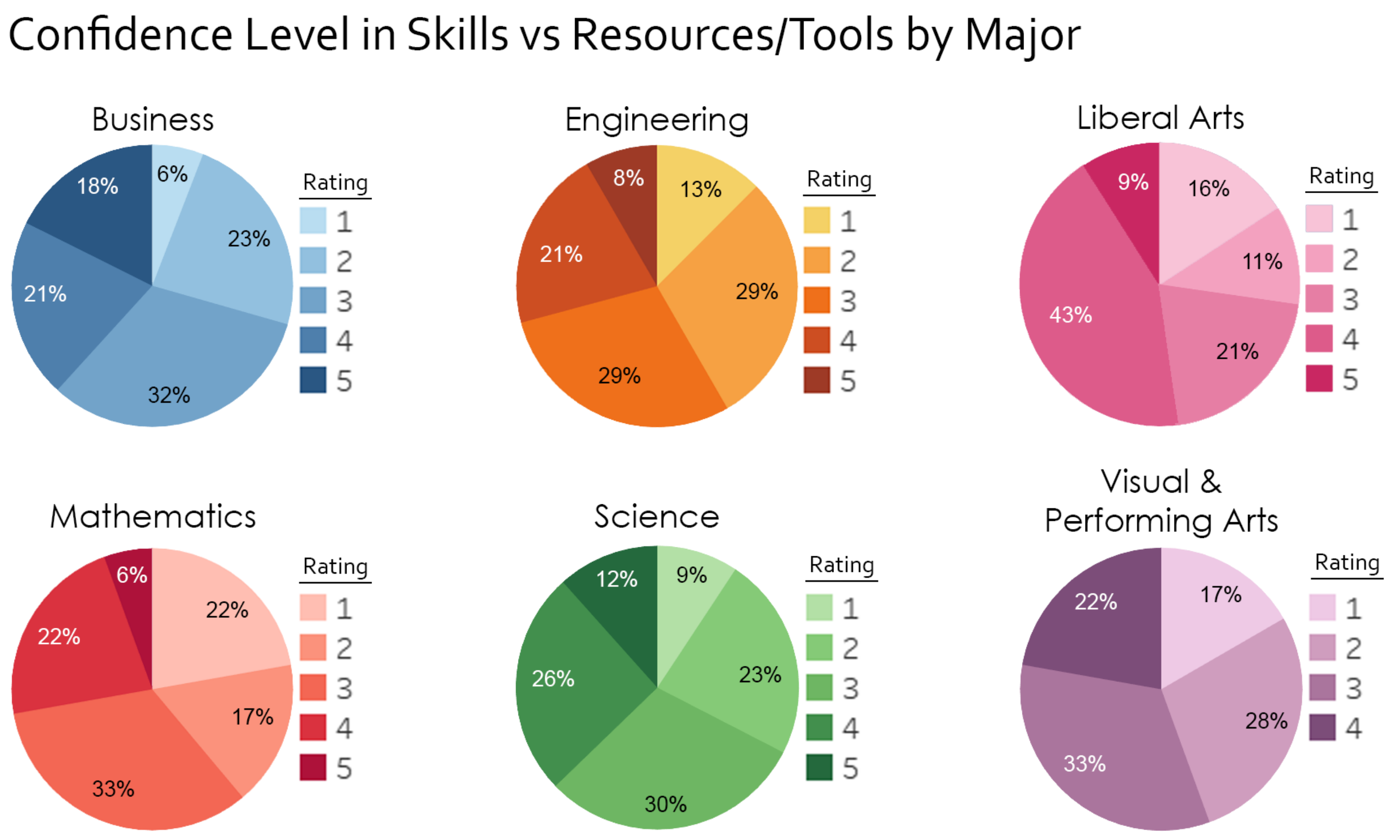

Figure 5. This visualization, made by Gabby Willard, shows how data visualization employees, based on their college major, rate their own data visualization job performance.

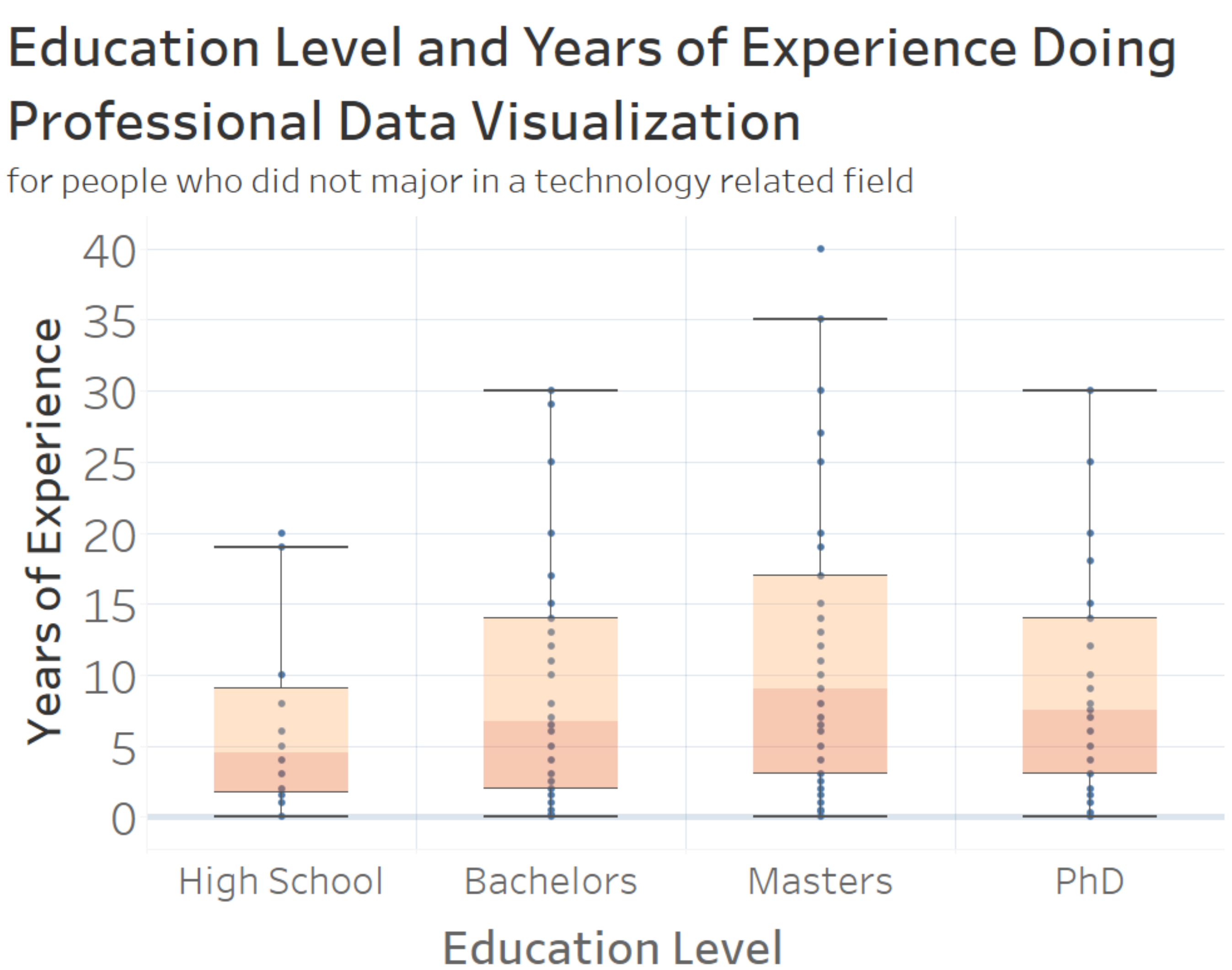

Figure 6. The visualization, made by Rae Fu, depicts the spread in years of experience depending on their education level. The median, third quartile, and fourth quartile all show the same pattern of increasing years of experience as education level increases, with the exception of PhD.

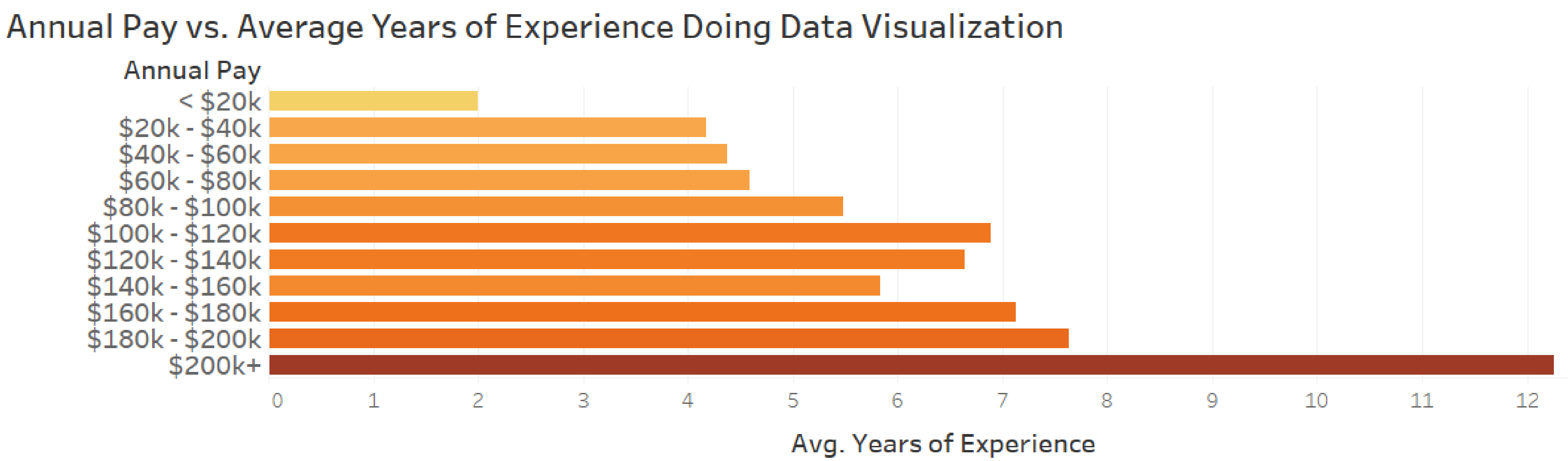

Figure 7. This visualization, created by Ace Glover, shows average years of experience doing data visualization compared to annual pay.

Discussion (What's the story?) and Conclusion

Figures 5 through 7 are the results that the team came up with.

To find out how confident data visualization workers are with their data visualization skills and tools, results from the Data Visualization Society's 2021 State of the Industry Survey Data in the "data_repub_2019" spreadsheet was used. The results that were used to create figure 5 include all of the respondents who said that they were hired to do data visualization, the ratings of their skills/tools, and their majors. All of the majors data was filtered into categories of non-technology majors. With the results displayed in figure 5, it appears that business majors are the most confident in the tools that they have for data visualization but are not at all confident in their skills, having a percentage of 18%. The mathematics majors had a greater percentage than the other majors (of 22%), that needed more data visualization tools but were confident in their skills. Looking at the business majors only, the results show that even though they may have the greatest percentage (of 18%) out of the other majors for confidence in their tools, the majority of the business majors answered that their tools and skills balance equally with that percentage being 32%. For engineering, the majority vote was tied (both being 29%) for being neutral and being somewhat confident in skills but needing a few more tools. With liberal arts, the majority vote of 43% was being somewhat confident in tools and needed a few more skills. Mathematics had the majority vote of 33% for being neutral with skills and tools. Science had the majority vote of 30% being neutral. Lastly, visual and performing arts had the majority vote of 33% being somewhat confident in tools but need a few more skills. An interesting observation was that none of the visual and performing arts majors answered that they were confident in their tools but needed more skills. Four out of the six graphs had a majority vote of being in the neutral stance on confidence in their skills and tools equally.

Figure 6 depicting the education level and the spread of years of experience, supports the idea that any education level would have a similar chance of having a job that involves data visualizations. This is due to a weak correlation in an increase of years of experience with an increase in education level. PhD did not match the other education levels in having the median, third quartile, and fourth quartile, making the argument that there is no difference between education levels. This supports the idea that many jobs include data visualization tasks, and that these jobs are not limited to specific education levels since the average years of experience doing professional data visualizations is close to the same, which means that they began professional jobs at around the same time, a certain number of years ago. Under these conditions, starting with data visualizations at the same time requires the jobs involving data visualization to be available for all education levels. With education level not being a deciding factor in whether a job involves data visualizations, the visualization shows that students of any education level have a possibility of working a job that involves data visualization in the future.

Figure 7 shows the annual pay and compares it to the years of experience doing data visualization. By showing this graph it lets the viewer know that people generally receive higher amounts of annual pay when they have been doing data visualization for longer. Our target group was non-technology majors, so this group of people likely would not have been taught how to do data visualization. If these people had been doing data visualization in college, they may be able to get more experience doing data visualization while still attending school. It is beneficial to start earlier in order to get a headstart when going into future careers as people with more experience get paid more. From the graph it shows that people generally need to be doing data visualization for at least three years in order to get into higher pay range values. When people receive higher annual pay and have been doing data visualization longer, it shows how important it is to learn these skills for careers.

The visualizations show that people of all majors and education levels require data visualization skills for their jobs. Recommendations would be for students to take a data visualization class to learn the skills that are likely necessary for their future job.

References

Christensen, A. (2022, February 3). *What's the best undergraduate major for a career in data analytics?* phData. Retrieved April 28, 2022, from https://www.phdata.io/blog/best-degree-for-data-analytics/?msclkid=ce645b9ec72311ec8cd46f46e50e53f8

Makulec, A. (2017). Data Visualization Society. Retrieved April 28, 2022, from https://www.datavisualizationsociety.org/

Appendix A – Resources Used

Datasets

The data that we used came from the Data Visualization Society's 2021 State of the Industry Survey Data which was provided to us on Brightspace. The specific data set that we used inside of the Survey data is a spreadsheet called "data_repub_2019."

Tools used

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

Tool/Application	**Description**
Google Docs	Hackathon Report
Google Slides	Presentation Slides
Microsoft Excel	Data Filtering
Tableau	Data Visualization
Adobe Photoshop	Data Visualization Editing
After Effects	Video Editing
Weebly	Website Development

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 the 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 is 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

As a team, we gathered data and decided what topics we were going to work on for our team results and each of our individual contributions. We came up with a story, each wrote parts of the report, rotated team leaders, scripted our own speaking parts in the video, recorded our lines for the video, and made slides to display during the video.

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
Gabby Willard	*Created a visualization to display for the team results about the confidence level that data visualization professionals have in their skills and tools. Wrote the background, questions, part of the methodology, and part of the discussion and conclusion. Made sketches to include in the methodology with captions. Created an individual visualization displaying the non-technology majors of people working in data visualization with a story. Wrote the diversity statement. Followed the data visualization process for all of the visualizations that I made. Put tons of time into mining and filtering many rows of the data for different topics of the report.*	*33 ⅓ %*
Rae Fu	*Filtered out the technology-related majors and created the visualizations on Education Level and Time Spent on Data Related Activities Each Week. Composited the video and created the website. Wrote the problem statement, part of the methodology, discussion and conclusion, and introduction.*	*33 ⅓ %*
Ace Glover	*Helped to narrow down topics to research for the data set we chose. Filtered the data out that I needed in order to create my visualizations after the technology field was filtered out. While using the data visualization process I created the visualizations comparing annual pay to college majors as well as comparing annual pay to years of experience. I helped critique the other group members' visuals and helped create the presentation slides. After helping with the slides I recorded a portion of the audio for our video.*	*33 ⅓ %*
	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: Gabby Willard

Group Topic: The Values of Professional Experience in Data Visualization
Your Topic/Question: Excluding all technology/computer science college majors, what college graduate degrees make up the 2019 Data Visualization workforce?
Describe the diversity YOU bring to the group (150 words or less): I bring an artistic perspective to creating and editing visualizations using my skills in Adobe Photoshop, Illustrator, and Tableau. I have experience with coding languages such as Java, Python, R, HTML, and CSS. I am a member of Women in Technology at Purdue. I am in a learning community called the "Data Mine." For leadership, I was a member of the Academic Honor Society in high school and I was an active team member for multiple sports teams.

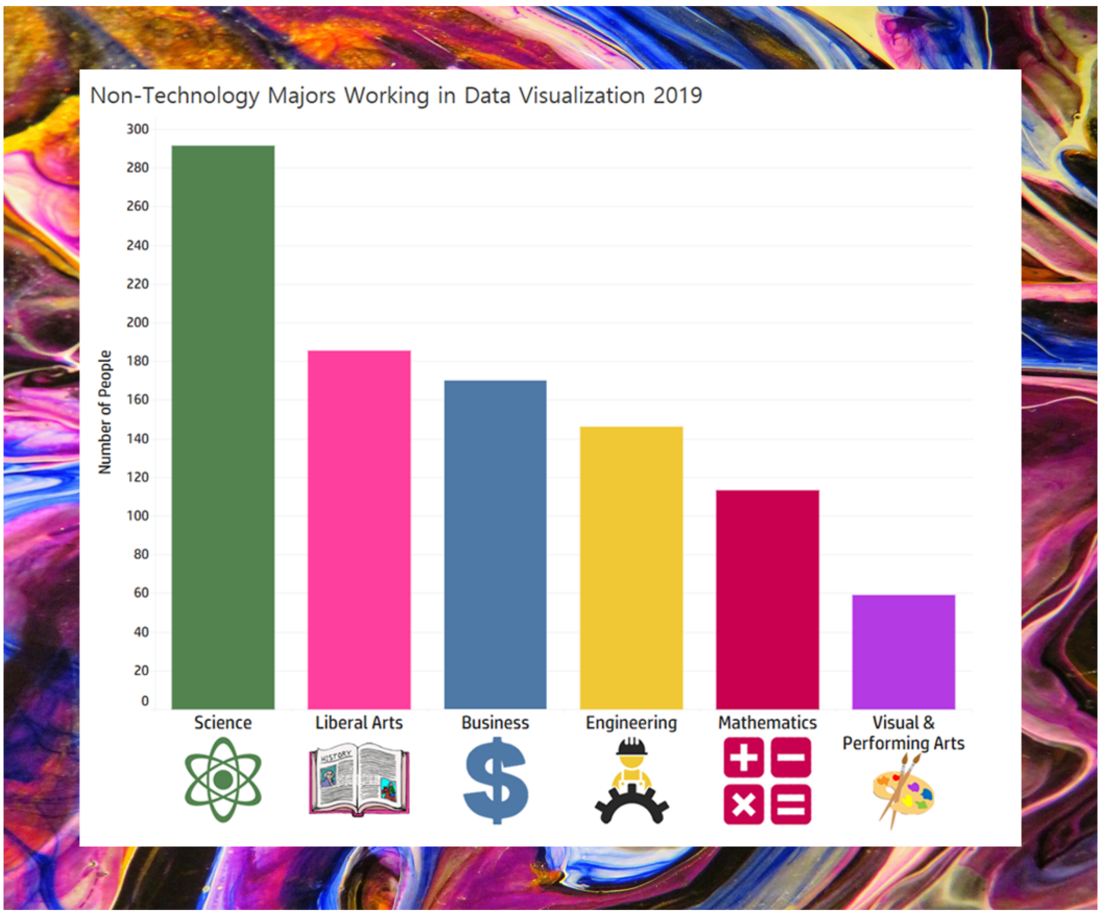

Figure 8. This visualization shows the number of people who graduated in non-technology majors who work in data visualization in 2019.

I was curious to see the breakdown of how many people were inside each of the non-technology majors that I had used in figure 5 used for the team's results. I was surprised that the biggest non-technology major category that worked in data visualization was science and that visual and performing arts had the least amount of people. To make the visualization look more appealing, I gave each major a color along with a symbol to spot the categories easier.

Team Member #2: Rae Fu

Group Topic: The Values of Professional Experience in Data Visualization
Your Topic/Question: Breakdown of Time Spent on Data Related Activities Each Week
Describe the diversity YOU bring to the group (150 words or less): The diversity I bring to the group is that since I've gotten two regional scholastic art gold key awards and I also took many computer science courses in high school, I bring a perspective that merges the visuals with the technical. In addition, I lived in Canada for nine years, which gives me a different perspective on intrinsic values.

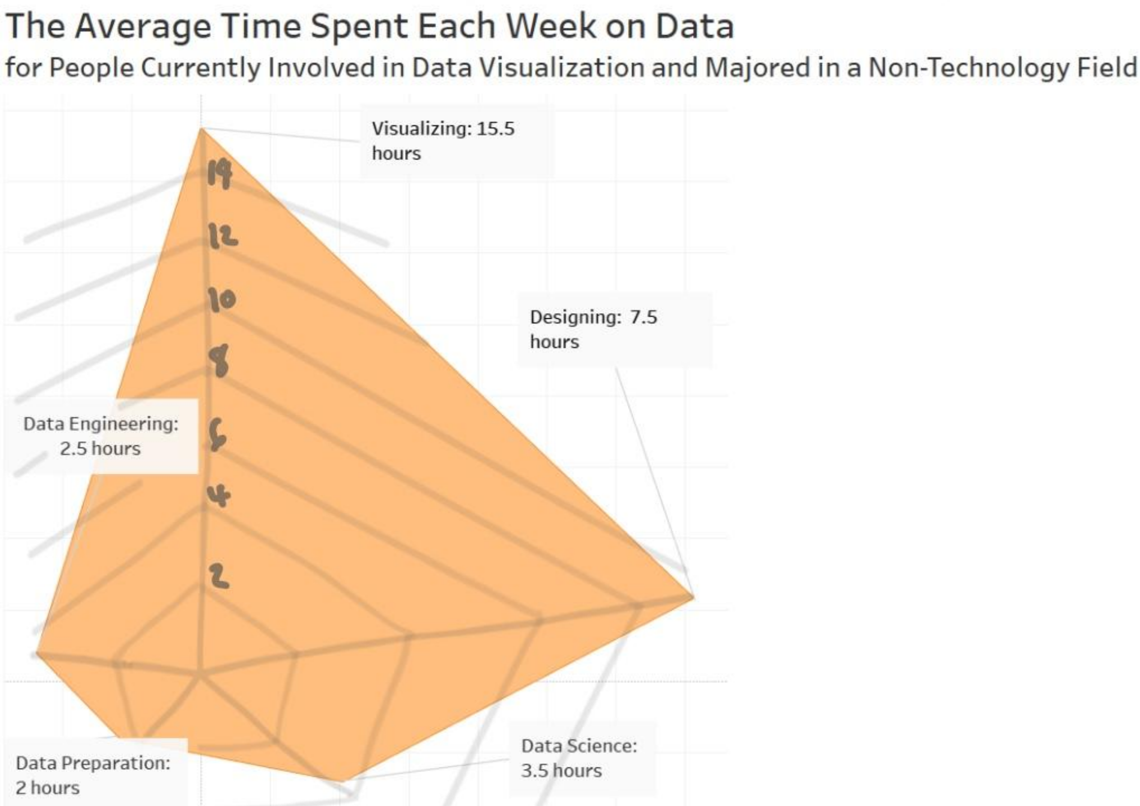

Figure 9. This radar graph shows the breakdown of the average time spent each week on data. they spend an average of 31 hours a week in total on these activities and most of it was spent visualizing.

From the team's findings earlier, we discovered that inside the non-technology-related majors, there is a large percentage of the user group who believe they need more data visualization skills to improve their job performance. This problem is whether they need to improve their data visualization skills to improve their job performance. This graph answers the problem by using quantitative data and comparing it to other data activities. It shows a breakdown of the time spent doing the data-related activities each week. The activity with the average most time spent each week is visualizing. With the assumption that greater time spent on activities correlates to needing improvement and the method of being more efficient would be through expanding their skillset, the users should focus on getting better at their data visualization skills.

Team Member #3: Ace Glover

Group Topic: The Values of Professional Experience in Data Visualization
Your Topic/Question: Comparing the 5-10 most popular college majors and annual pay.
Describe the diversity YOU bring to the group (150 words or less):
The diversity that I can bring to the group is from living in three different states and from having a background in art. By living in different states, I was able to gain a different perspective of different schools along with different people. I graduated with an Associates of Arts degree from Waubonsee Community College which is in Illinois. Due to this I know a lot about art and design, but since I am continuing my education by learning about Animation, I can provide skill sets from that as well.

College Majors vs Annual Pay from 2019

Annual Pay

College Major	< $20k	$20k - $40k	$40k - $60k	$60k - $80k	$80k - $100k	$100k - $120k	$120k - $140k	$140k - $160k	$160k - $180k	$180k - $200k	$200k+
Math	3	6	7	15	6	6	9	2	3	3	4
Biological	5	12	9	14	5	3	3	2		1	5
Economics	4	3	13	11	7	5	4	1		3	4
Business	1	4	4	11	7	1	5	5	2	2	1
Art & Design	3	3	4	9	8	1	4	3	2		1
Psychology		4	10	8	12	8	4		1	2	4
English		2	2	4	2	3	1		1		2
Engineering	4	4	5	3	7	5	4			1	3

Figure 10. This graph shows the annual pay compared to the 5-10 most common college majors.

I wanted to look into annual pay further from the team visual which related annual pay to the average years of experience. I looked at this topic to see if a person's college major affected their annual pay. I found that most people fall into a pay range of $60k-80k and that there are fewer people in majors like English and Engineering. The problem I tried to solve with this data is to see if there were majors that were paid less than other majors. I found that Psychology and Biology majors have a larger range in pay values and that out of the most popular majors from our data set, Biological majors generally pay less. As you can see from the graph, there are a large number of Biology majors in the $20k-40k range opposed to other majors.

Appendix E - Diversity Statement

Team Rocket is made up of three people who are diverse in their fields of studies, skills that they bring to contribute to the team, backgrounds, cultures, and outlooks on the world. Our team members grew up in different places such as Indiana, Illinois, Toronto (Canada), and Delaware. We come from different fields of study at Purdue University and this includes Gabby Willard with Computer Graphics Technology (in Data Visualization), Rae Fu with animation and visual effects, and Ace Glover with Animation. With the different fields of study, each one of us brings a variety of different skill sets to the table such as knowledge of Adobe Photoshop, Adobe After Effects, Data Compiling, Tableau, years of sketching experience, programming, and geometric modeling. As a team that unites unique individuals together, we come together with lots of strengths in different areas to distribute into our hackathon as a whole. Each one of us brings brand-new perspectives and outlooks to collaborate ideas effectively.

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.

Gabby Willard	Gabby Willard	4/28/2022
Print Team Member Full Name	Signature	Date
Ace Glover	Ace Glover	4/28/2022
Print Team Member Full Name	Signature	Date
Rae Fu	Rae Fu	4/28/2022
Print Team Member Full Name	Signature	Date