

Education's Effect on Pay

Distribution of Pay Grade By Level of Education



Team name

Name (full name)	Purdue Email address
Lauren Behnke	behnkel@purdue.edu
David Luo	luo354@purdue.edu
Kyle Choi	choi687@purdue.edu

Project URL: <https://2clap98884296.wordpress.com>

Table of Contents

Introduction	3
Background	3
Questions	3
Problem Statement	3
Methodology	3
Results	4
Discussion (What's the story?) and Conclusion	4
References	5
Appendix A – Resources Used	5
Datasets	5
Tools used	5
Appendix B – Project Web Page	6
Appendix C – Percent Contribution	7
Group Contributions	7
Individual Contributions	7
Appendix D – Individual Contributions	8
Team Member #1: Kyle Choi	9
Team Member #2: Lauren Benhke	10
Team Member #3: David Luo	11
Appendix E - Diversity Statement	12
Appendix F – Team Consensus	13
Team Consensus	13

Education's Effect on Pay

Introduction

Team Clap aims to examine the relationship between the level of education and pay grade within the data visualization field. It is generally accepted knowledge that a higher level of education results in higher pay and we were curious if this applied to the data visualization field.

Background

The data used was the State of The Data Visualization Industry, an annual survey taken by people within the data visualization industry to obtain information such as pay and also more specific information year to year. We chose the 2019 survey because we wanted to avoid any effects of COVID-19 on the data. In addition, the 2019 survey had relatively complete fields for what we wanted to analyze.

Questions

Our primary question was to see if it is worth pursuing higher education in the data visualization field. Our primary audience is students looking to go into data visualization. In addition, this may also be useful to those in the industry that may be looking into higher education. This analysis should help these groups in their decision on if they want to pursue higher education.

Problem Statement

By focusing on education, we hope to give important insight to people looking to higher education in data visualization to facilitate better decision-making. Our team will look not only at the number of people from each level of education but also normalize the data to better view the differences within each education level.

Methodology

After acquiring the dataset, our team first parsed and cleaned the necessary data. From the 2019 data much of the education level data was messy and from free text entries. A python script was used to scrape certain phrases from the free text entry into smaller categories. After the data was cleaned, several exploratory visualizations were created to refine our primary question. We then chose the visualizations that best tackled the question and refined them into our final visualization.

Results

Distribution of Pay Grade By Level of Education



Figure caption: A stacked bar chart showing the 4 levels of education that were filtered from the data set. Each category was normalized so each category totaled to 100%. *Because of how the data was scraped it is possible that some students are included in each category. For example, a PhD student may be listed as a PhD.

Discussion (What's the story?) and Conclusion

We can notice that the other category does not have as many of the highest pay grades as the other three categories. However, it seems no matter the level of education data visualizers make a respectable amount of money as in all four levels, more than 40% of visualizers make more than 80k/year. One surprising finding is that master's holders seem to make slightly less money than both bachelor's and PhD holders. This could lead to the conclusion that while it may be worth it to pursue a PhD it may not be worth it to only pursue a master's degree. What is undoubtedly true is that it is worth completing some level of

higher education in order to potentially obtain a higher salary. In order to make more definitive claims, the data needs to be more strictly cleaned or better collected.

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.

Hovey, P. (2009, March 20). 3D Data Visualization—Dewey Calendar [Review of 3D Data Visualization—Dewey Calendar]. Pehr Hovey.
<http://pehrhovey.net/blog/2009/03/3d-data-visualization-dewey-calendar/>

R.bhabuk, U. (2013, April 8). R graph gallery: RG#38: Stacked bar chart (number and percent). R Graph Gallery. <http://rgraphgallery.blogspot.com/2013/04/rg-stacked-bar-chart-number-and-percent.html>

Appendix A – Resources Used

Datasets

2021SOTI_SurveyDataForPublic

*This and filtered data is available via GitHub on our website.

Tools used

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

Tool/Application	Description
Python/Pandas/Jupyter notebook	Data cleaning/Exploratory analysis
Excel	Data viewing
Tableau	Data visualization
Wordpress	Web 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

What was your team's main goal/question?

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

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, and reading the final deliverable before submission.

Individual Contributions

Table 1 shows an example of what a team contributions table might look like.

Table 1 Example Team Contribution Table.

Team Member	Contribution	Contribution
<i>Example Team Member 1</i>	<i>Developed the project web page, acquired additional data for the project</i>	<i>25%</i>
<i>Example Team Member 2</i>	<i>Responsible for gathering written contributions from the team and combining them into a cohesive story, data wrangling (parsing, filtering) .</i>	<i>20%</i>
<i>Example Team Member 3</i>	<i>Videographer for the 5-minute video (recording and editing)</i>	<i>15%</i>
Total		100%

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
<i>Kyle Choi</i>	<i>Created one visualization, created the slideshow, and edited and cleaned the video</i>	<i>33.33%</i>
<i>Lauren Behnke</i>	<i>Created two visualizations, created website</i>	<i>33.33%</i>
<i>David Luo</i>	<i>Created two visualizations, three exploratory graphs, and cleaned data</i>	<i>33.34%</i>
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.

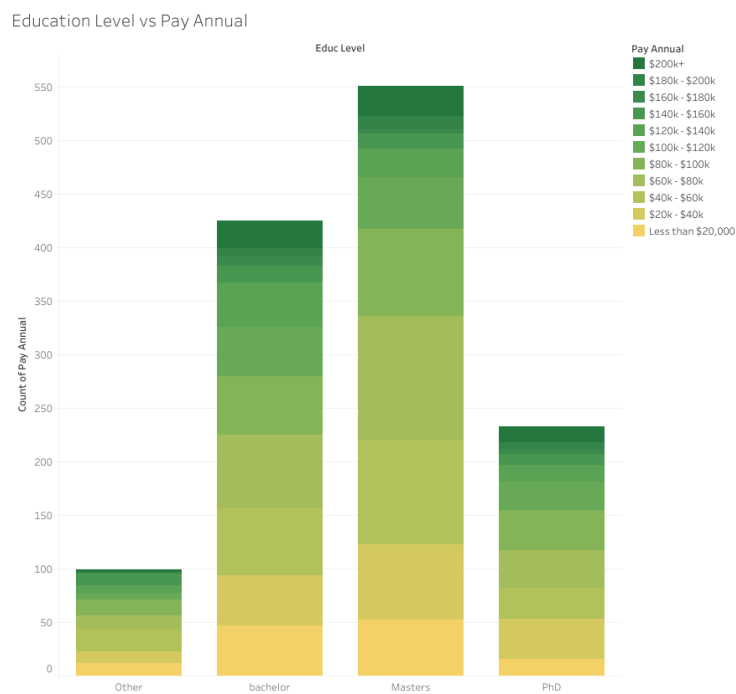
Team Member #1: Kyle Choi

Group Topic: Factors that go into pay grade, mainly education

Your Topic/Question: How does education affect annual pay?

Describe the diversity YOU bring to the group (150 words or less): The diversity that I bring to the group is my cultural identity. As I was born and raised in Korea, and went to a Korean elementary and middle school, quite the adjustment was necessary when coming to America. I provided a worldwide perspective with my extensive traveling experience, and made sure to keep in mind countries that may be underrepresented and overlooked.

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



My visualization shows the counts of different education levels compared to annual pay. This was the main focal point of the group, as our main topic was focused on how education level, among other factors, can affect annual pay. At first glance, it may seem as if bachelor's and master's make the most money. It is important, however, to note that these are merely counts of education levels, not how much a certain degree makes. The pay distribution of different degrees is not the main focus point of the graph, so it is very likely that simply more people in the data visualization industry come from a background of master's or bachelor's degrees.

My main contribution comes in the form of presentation, as I created the majority of the slides, as well as recorded and edited the presentation video. As there was a lot of unwanted background noises such as static, door closing, and computer notifications in the raw footage, I compressed the audio in order to suppress the sound of static. For the louder noises, such as a roommate walking in, or a computer notification, I separated the high frequencies from the mid and low frequencies, and equalized the bands to cut out the noise.

Team Member #2: Lauren Behnke

Group Topic: We looked at analyzing the relationship between education level and pay grade within the data visualization field.

Your Topic/Question: How does where one lives in the world affect pay?

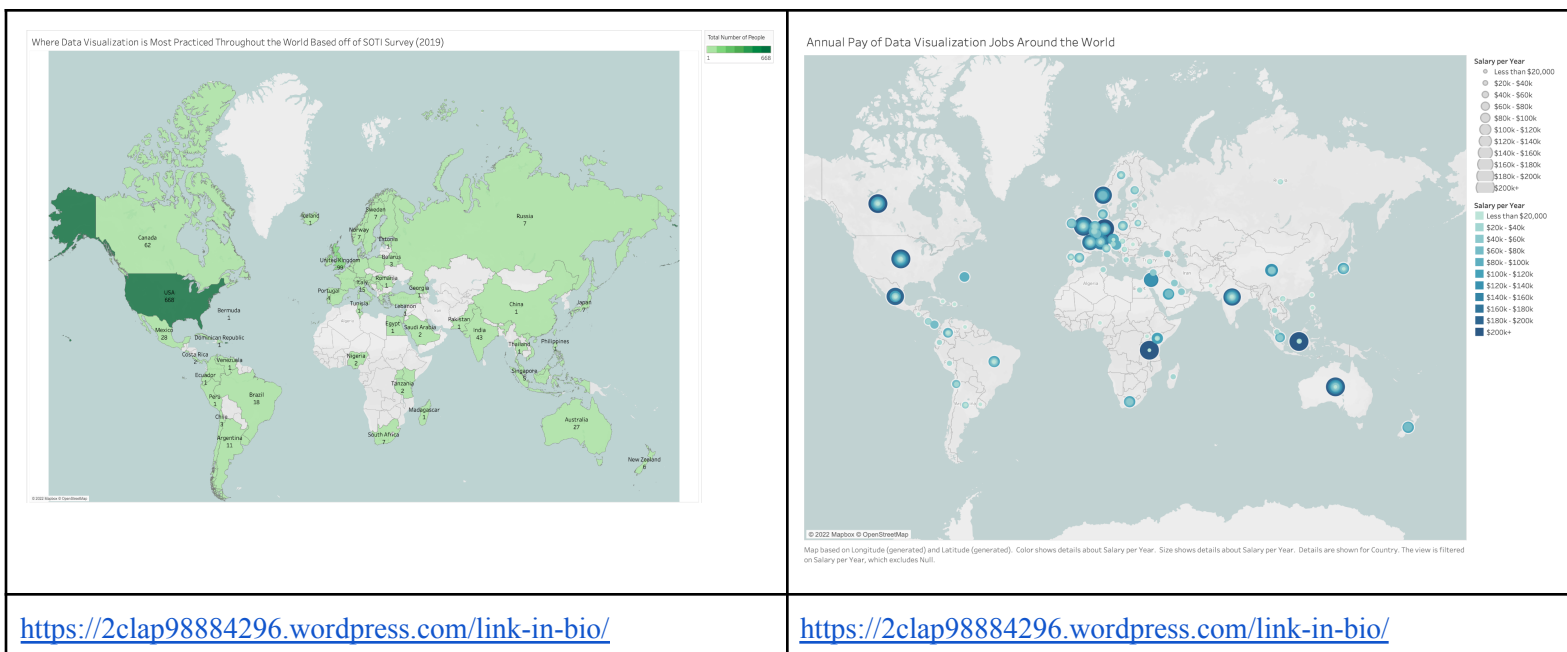
Describe the diversity YOU bring to the group (150 words or less):

The diversity I bring to this group is being a woman studying a major in technology which I believe is important when bringing new outlooks and ideas to the group. I believe that I am very good at listening to others and working to communicate ideas. My experience with using Wordpress has helped me succeed with this project to create our website.

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

For my contribution to this project, I created our team's website along with two visualizations that looked at how many people had data visualization jobs in each country (based on who took the survey) and another one looking at salary based on location. For my first visualization, I used Tableau to create a general map that pinpoints how many people have jobs working in the data visualization field in each country. From the map, we can see that the United States is the leading country with the number of people working in data visualization. Most of the other countries in the world fell behind. For instance, I thought it was important to note that some more developed countries fell behind the United States. An example being Australia only had 27 people compared to the United States' 668 people.

For my second visualization, I created another map that shows the different salary ranges people were making in each of those countries. I believe that this was effective because it shows the highest salaries one could make living in that country. I noticed some shocking results when using this map. I saw that countries such as Tanzania and Indonesia had people making \$200,000 a year or more. This was surprising because not many people were working in the data vis field in these countries. They each had less than 4 people but some of those people were a part of the highest salary bracket.



*links to see larger versions of visualizations on website

Team Member #3: David Luo

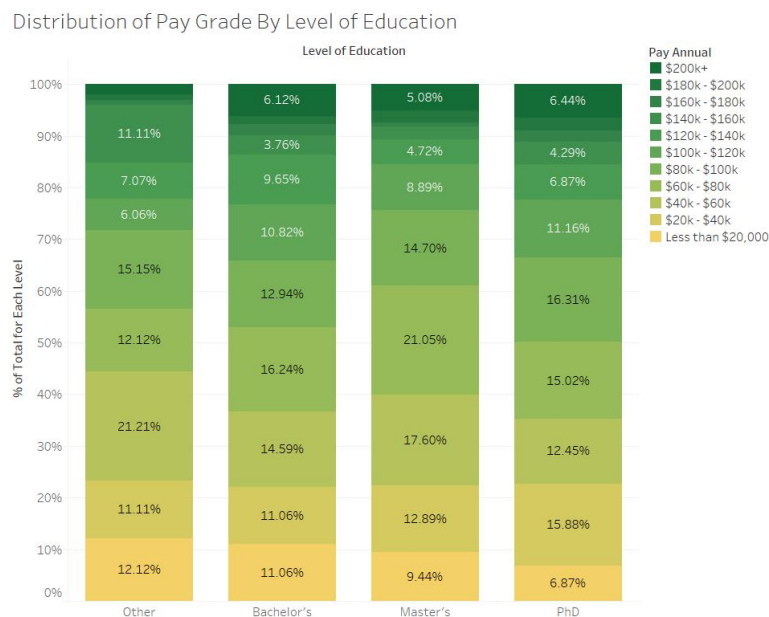
Group Topic: Effects on Pay

Your Topic/Question: Exploratory Analysis. How does pay compare based on education level?

Describe the diversity YOU bring to the group (150 words or less):

I bring a unique cultural background being an Asian American from the midwestern and western United States. In addition, I bring experience from both a computer science and a statistical background thanks to my education in data science. My previous experiences with hackathons also provided organizational support in the form of version control tools.

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



After performing exhaustive exploratory analysis and data cleaning, I decided the data subset with a good sample size and few outliers involved annual pay and level of education. Annual pay was used instead of hourly pay because freelancers caused some outliers within that field. Education level was chosen because fields like experience did not have sufficient sample sizes in their higher categories. All four categories of education had sample sizes greater than fifty after cleaning.

One interesting pattern to look at is the downward trend of the less than \$20,000 block as education rises. This is an interesting pattern with several potential explanations. People with higher degrees may be less likely to take part time, freelance, or internship work. People with higher degrees with part time or freelance work may also demand to be paid significantly more, bringing them out of that pay range.

Another interesting thing is that the three degrees are much more similar than one might initially think. The difference within each block is typically no more than a few percent. This brings into question how education actually affects pay in the data visualization industry and deserves more research.

However, because of how the data was cleaned, there is a possibility that the lower pay grades have been inflated in the different degrees due to including current students. In the future it may be wise to filter out people not in full time work.

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 boost 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 makeup 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.

Team 2Clap consists of members each from different majors. Lauren Behnke is a data visualization major, Kyle Choi is a computer science major, and David Luo is a data science major. Lauren Behnke is studying data visualization and management. She has experience using Wordpress which her skills helped to develop the website. She has also attended a conference for UNICEF in Washington D.C. to advocate for children's rights. David Luo provides a more statistical perspective from his background in data science. In addition, his experiences from previous hackathons provided organizational support with version control tools like GitHub. Kyle Choi provided a cultural point of view from his childhood/adolescence, as he was raised in Korea. Furthermore, his computer science background provided a more technical aspect to the hackathon. Each of our team members are unique in our own ways and are able to bring different perspectives when interacting with one another. However, we each have an understanding and desire to excel in everything we do, including working with an open mind to absorb as much technical and cultural information as we can.

\

Appendix F – Team Consensus

Team Consensus

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

David Luo



4/29/2022

Print Team Member Full Name

Signature

Date

Lauren Behnke



4/29/2022

Print Team Member Full Name

Signature

Date

Kyle Choi



4/29/2022

Print Team Member Full Name

Signature

Date

Save this document as:

HackathonTeamName_CGT270Spring2022_FinalReport.pdf