Yueying Wang

AIT664-B02

07/20/2021

**An Inquiry of Economic Independence: Part 4**

The topic that I chose for this research project is economic independence as it relates to the level of education that one has completed. The variables that are included are age, marital status, contribution (if married, how much each partner weighs regarding salary), ethnicity, gender, education, and income. The reason that I chose this project is because I have noticed that often blue-collar jobs such as construction bring in larger annual salaries than do white collar jobs (ex. social work or Public Servant) where higher education – as well as more economic investment – is needed. I have always been under the belief that more education always correlated to better salaries as this is what I was taught in school and constantly stressed to about the importance of obtaining higher education. My hypothesis states that if some have a higher degree of education, then that person will earn more money (income) compared to his or her counterparts. I am going to test this hypothesis by using various data sets which will allow me to compare/contrast to see what conclusions I can draw based on the given information. This section, the summary, will set out to review the process embarked upon in this project with each point of commentary under individual subheadings.

Data Visualizations and Analysis

Chart, bar chart, waterfall chart

Description automatically generated

\*Figure 1 – Occupation vs. Income - R-Studio

The first visualization compares Occupation vs. Income and was compiled using R-Studio (R-Studio, 2021). Based on this visualization, we can see that most of the people in the dataset represented two occupations: Engineering and Business/Financial. This is important to note as these industries tend to generate higher incomes compared to others like Farming or Food Serving. Over half of Engineers are making over 100,000 dollars a year while a little over a quarter make more than 50,000 per year. The one missing source is an understanding of how many years in the profession it took to achieve these salaries. Additionally, different types of Engineers make a variety of incomes, Computer vs. Water, but this is not included. Looking at this graph supports my hypothesis as it demonstrates that two fields that require higher learning also show higher income rates. That said, there are two outliers: Legal and Healthcare (together) and Education. In the Legal and Healthcare fields, we should see high income ranges but too few of these professions were included in the data set to accurately draw that conclusion. On the other hand, Education is well-known for low salary compared to level of education although, as with legal and healthcare, I don’t want to make a definitely conclusion as our data set is limited in this field. The audience for this visualization, as will be for all in this visualization report, will be those who are interested in understanding the connection between having a higher degree of education and income. This will apply to a variety of people across academic fields who may be using this data for research or to satisfy personal inquiries. There are no specific needs or requirements that an individual would need in order to understand these visualizations. Specifically, there are no pre-requisite understanding or knowledge that one would need in order to interpret these images. They are straight forward and contain keys that provide the audience with a full understanding/meaning of each diagram. Likewise, there are no requirements other than possessing basic comprehensive abilities.

**Chart

Description automatically generated**

\*Figure 2 – Race vs. Income - R-Studio

This visualization shows Race vs. Income and was compiled using R-Studio. The visualization is quite straightforward as we can see that White/Caucasian individuals are more likely than any other race to make over 100,000 dollars a year. There are many reasons but foremost is that there more white Americans in the United States than any other race therefore they have a higher likelihood. African American and Asians are similar in their tallies as are Hispanic and Mixed Race. Another reason is that less African Americans and Asians were surveyed compared to Whites. This visualization supports my hypothesis based on the fact that white Americans tend to have more access to higher education which in-turn allows them to pursue job fields that make more money and, in-turn, are more successful from a financial standpoint.

Chart, bar chart

Description automatically generated

\*Figure 3 – Education vs. Income - R-Studio

This visualization shows Education vs. Income and was compiled using R-Studio. Based on this bar graph, we can see that the greatest number of entries represented those who have a high level of education which can be inferred those who have a higher degree of education also have a higher income. For In my coding, I included income which sawed that most of the people surveyed had higher education (at least a bachelor’s degree) although their income ranged between middle to upper. These findings align with expectations that would be commonplace based on our fundamental belief that those who are unskilled tend to make the least while those who have skills make more. This visualization supports my hypothesis because clearly the graph shows that people who have higher education also have better income. However, higher income doesn’t mean that they are financial independent. We still need to look at other factors to support my hypothesis.

Chart, line chart

Description automatically generated

\*Figure 4 – Age Distribution - R-Studio

This line graph shows the relationship between number of data entries and age distribution and was produced using R-Studio. Based on this graph, we can see that the highest amount of people surveyed where between 25 to 30 years of age with a consistent decrease past 30. This is important to note based this age category is typically when those with higher education have just graduated and have started to look for employment whereas those who have little to medium education have been in the workplace for 7-10 years. Although not directly related to the hypothesis, this data is important because it provides the viewer with introspective into the age of those in the data set and can lead to further overarching conclusions about the relationship between education and income.

Chart, bar chart

Description automatically generated

\*Figure 5 – location vs. Income - R-Studio

This bar graph shows location vs. Income as it relates to number of entries versus location and was compiled using R-Studio. The graph shows that an overwhelming amount of data entries came from Suburban and Urban locale compared to Rural and Other. Based on this data, it makes sense as it is easier to survey in areas that have higher population density and where most polling takes place. Furthermore, the graph is also showing people who are living in Suburban, and Urban areas have a much higher chance to have a better income. There is a direct connection between higher levels of education in suburban and urban environments compared to rural. This evidence does not directly support my hypothesis but shows that the data taken was from areas promoting higher education.

Chart, bar chart

Description automatically generated

\*Figure 6 – Gender vs. Income - R-Studio

In the following bar graph, we can see gender compared to number of data entries and income which was compiled using R-Studio. Based on this graph, we can see that there is a large disparity between the number of men and women in the data set. I am not sure why there was such a high proportion of men (could be ease of access for the surveyor) but this should definitely be taken into consideration. Based on the males dominating, we can make a direct correlation to why we have so many data entries in the field of Engineering which has been almost complete male dominated. The only factor that would play a negative role would be that statistics show that there is a higher rate of females entering higher education compared to males which could have skewed the data. This graph indirectly supports my hypothesis that links higher education to higher income in that we know fields that are male dominated and these fields make very good money.

Chart

Description automatically generated

\*Figure 7 – Gender vs. Age - R-Studio

This scatterplot graph shows the relation between Gender vs. Age and Income and was created using R-Studio. I have learned the layout of this graph comes from a previous data analytics class that I took (Thomas, 2021). In this graph, we can see that for women their peak in between 35 to 45 years old where the greatest amount made over 600,000 dollars a year. For men, their highest peak in income was between their late 20s to early 40s where the greatest amount is making over 700,000 dollars a year. This data is incredibly relevant because it reflects the income difference between men and women. For my hypothesis, although this graph does not show education, it can be reflected that most of these people are making large sums of money also have a high degree of educational backing.



\*Figure 8 – Age vs. Income and Debt - Tableau

This scatterplot graph shows the relationship between total income vs. total debt and average age and was constructed using Tableau (Tableau, 2021). Although not the most easily readable, we can see that there is a high degree of density toward the bottom of the graph that reflects those in the data set that have a total income less than 1 million dollars. We can also see a relatively substantial clustering around the 2-million-dollar mark which is quite impressive given that a majority of the data set make less than 1 million combined. Likewise, as expected, those who have a lower total income have less total debt but it’s relative to the fact that they do not make as much money so they still may be felt quite burdened. A large amount of debt is found amount student loans and home mortgages. This graph relates to my hypothesis as it shows a relatively high distribution of income, but it does not touch on education. This graph can be of most use when combined with other graphs I have created to support the hypothesis.



\*Figure 9 – Income vs. Mortgage based on Occupation – Tableau- Dual Combination Chart

In this dual combination chart, we can see that income vs. mortgage debt based on occupation. This graph was created using Tableau. The bar graph is interesting because it confirms that those in the Business and Financial industry have the highest incomes while having a mortgage that is high but not excessively based on their profit margin. Another interesting conclusion is the income of food serving employees compared to their total debt. Their debt is very low compared to how much they make which could be a variety a reasons but very difficult to tell. One reason could be because they live at home with family or that they rent rather than buying. The importance of this graph for my hypothesis is that it confirms that those fields that require higher education make more money. It also provides insight into other fields and cross analyzes the total income from each field compared to the amount of mortgage debt. For example, in the legal profession, we know that people cannot enter this profession without higher education and a legal license. However, the high salary they make doesn’t mean that they are financially independent if they have large debt. This situation also applies to the Healthcare profession. On the other hand, we can see that people who are in the business and finance field are doing much better and one of the reasons is the financial knowledge they have gained from school applied to real life.



\*Figure 10 – Gender, Total Income, and Occupation – Tableau -Bubble chart

In this bubble chart, we can see the relationship between Number of the people, Total Income, and Occupation and was constructed using Tableau. This chart shows, as we have noted before, that Engineering is highly dominated by men and has a relatively high-income scale. Similarly, Business and Finance is also male dominated while professions like Education, Farming, and Arts and Media are generally split between male and female workers. Legal Occupations and Healthcare make the most amount of money. As relates to my hypothesis, this chart is important as it supports my argument about professional that require higher education leading to better paying jobs.

In this visualization report, I have produced 10 visualizations that all relate, either directly or indirectly, to supporting my hypothesis. During the process, I noticed that a few of my graphs required editing which reflects back on the importance of constantly checking your work to ensure that your results represent the points that you are trying to prove. I used to variety of graph types as well as charts and used R-Studio and Tableau to assemble them.

Approach to Analysis

The approach that I took in my analysis is that I first set out to find a trustworthy data set that I felt would work well with my developing hypothesis. After I reviewed the data set, I started to implement my findings into different graphical interfaces (visualizations). The two software analysis tools that I used were R-Studio and Tableau. Originally, I wanted to only use R-Studio as my tool for analysis but when I was working on project deliverable 3, I realized that my knowledge of this software was limited. Therefore, I decided to use Tableau for some of the analysis because it provided me with more interesting and better visualizations. Additionally, I was able to dig out more information from the data as this software made it easier than it would have been if I used R-Studio. Reflecting back on my initial approach, I have realized that in research when you set a course of action it does not always work out as seamlessly as one would hope for as was in my case where, when I ran into some issues with using R-Studio, I had to change course and start to use Tableau.

Shifting course to the different variables used in the analysis, my data set presented a multitude of measurables which I narrowed down significantly during the process. The main focus, as can be seen in my above analysis, was to connect job fields to salary and quality of life. I felt that focusing on job fields was most important because humans, like myself, have preconceived notions of how much money (salary) that fields make but these notions where challenged based on my data. This decision was two-fold: firstly, because it supported my hypothesis and secondly, because it was important to bring to the attention of those who may read my research paper that some commonly held notions are being debunked (to a degree).

Effectiveness of Analysis

I felt that my analysis was effective based on the visualizations that I created and subsequent reflection. The basis for my positive determination is that looking at the visualizations you can notice that there is very little skewing as well as definitive data of one or more variable that trend much higher comparatively. The analysis that was provided under each visualization was supportive and detailed showing how different variables advanced my overall hypothesis. Likewise, there was enough data that could be successfully analyzed which is excellent because, from personal experience, I can immediately tell when my data is not working as I cannot draw meaningful conclusions that can be used in support of a hypothesis. I will draw more on my data analysis as it relates to the hypothesis in the next section.

Success of Hypothesis

For this project, my hypothesis states that if some have a higher degree of education, then that person will earn more money (income) compared to his or her counterparts. I supported my hypothesis through my data analysis (visualizations) that clearly showed that those professions that require higher education in general will be reflective in the person’s increased salary. The reason I say “in general” is that there are professions, such as education or social work, where you may have a terminal degree, but it is not reflective in a significant income boost in the public sector. For example, if we look back to Figure 1, Occupation vs. Income, we can see that professions that require higher education such as Engineering do make more money. If we look at Figure 10 - Gender, Total Income, and Occupation – this proves the hypothesis by showing that higher education fields have much larger bubbles meaning that they earn more money each year. This figure also shows that those fields that do not require higher education such as Farming, and Food Service make lesson money on average. Although I chose Figure 1 and 10, all of the subsequent visualizations also play a role in supporting my hypothesis. Based on my data analysis and visualizations, my original hypothesis was supported and is valid.

Lessons Learned from Analysis Project

After completion of this project, there were a few key learning experiences that I would like to reflect on. The first is on developing my hypothesis. In retrospect, I would have clarified “higher education” to meaning obtaining a 4-year degree in a specific academic field. The reasoning behind this action is because there is a slippery slope when you do not clearly define the parameter since holding a Doctorate in a given field does not always equate to earning more salary whereas having a 4-year or master’s degree reflects more dominantly. The second learning experience was in my data collection process. I was lucky to find a solid dataset that included numerous variables to work with but, in hindsight, I would have liked to use multiple datasets that provided a more diversified study. The dataset I chose was predominantly white, at least a college education, and regarding profession, heavily Engineering oriented. The questions that I was left with at the end of the project were one’s such as “What would the data have shown if the study was more ethnically diversified?” or “What if the data had a larger influx of participants who only had a high school diploma or less?” or “What would the data show if most of the participants with higher education where in civil service professions?”. As is the nature of research, the more you dig into a topic, the more questions that arise. The third, and final, learning experience from this project was developing my abilities to using software analysis tools like R-Studio and Tableau. During the project, I actually found it easier to use Tableau and I could create more detailed visualizations than with R-Studio which definitely gave me encouragement for the future. I am going to continue working with R-Studio to improve my abilities so that I can use them in the future.

**Works Cited**

Belon, M. (Ed.). (2017). *r/financialindependence - Official 2020 FI Survey Results*. reddit. https://www.reddit.com/r/financialindependence/comments/m1q8ia/official\_2020\_fi\_survey\_results/.

*Business Intelligence and Analytics Software*. Tableau. https://www.tableau.com/.

*Open source & professional software for data science teams. RStudio*. (2021, May 28). https://www.rstudio.com/.

*Thomas, S (2021): STAT 515: Applied Statistics & Visualization for Analytics (Spring 2021)*