Relationship between Amount of Financial Aid Received And Socioeconomic Background

Date: 2017/11/24
Github link: https://github.com/wuxiaohua1011/webdataproject1

Group Members:

1. Michael Wu
2. Angie Mejia
3. Pranav Bhasin
4. Sara Wu

Introduction

It is a common belief that people from a lower socio-economical background will have a lower education level. Through this project, we want to find out whether the affluency of a geographical area actually impact students' ability to attain higher education. If so, what kind of conditions, as in geographical, the background of their parents, or others would impact their opportunity the greatest.

Furthermore, we want to explore whether the amount of FAFSA differs for students that come from different socioeconomic backgrounds by comparing dataset from different times. In our project, we categorize socioeconomic backgrounds by the poverty level, average education level, and local unemployment rate for different states.

Related Work

The first related work we looked at for guidance was a report done by the educational testing service. It looked at poverty and the relation to education. This looked at the environmental effects that children in poverty face. It goes into single parent household statistics as well as solutions to the financial stresses through further government intervention. It focused more on younger children's education and finding solutions to underfunded schools. For our research, we looked at those in higher education and how coming from impoverished states influences how much college aid students get.

Secondly, we looked at a report done by the National College Access Network. It's entitled *An Examination of the Relationship between School District FAFSA Completion Rates and District Poverty Levels*. Their research focused on FAFSA completions and finding solutions to making it simpler for students to fill out. This report showed that the

poorer the school district, the less likely the students are to fill out the application for FAFSA. We saw a parallel with this conclusion because through our correlations we saw that even when the poverty rates go up, the amount of grants given stayed the same. This might because students aren't completing the FAFSA application.

Data

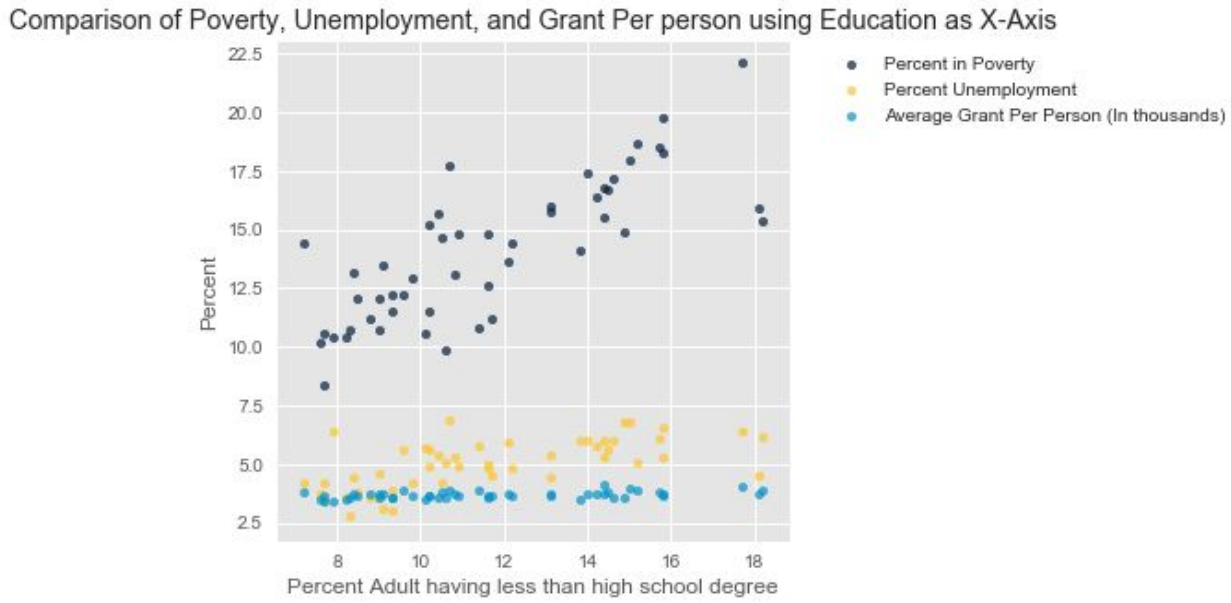
We used the following datasets:

a) FAFSA Dataset
   i) Usage: To find out for a particular location / ethnic group / socio-economic group, how many applied for student loan/grant.
   ii) https://studentaid.ed.gov/sa/data-center
b) United State Department of Agriculture -- Department of Economic Research
   i) Usage: Dataset for Poverty by County
   ii) https://data.ers.usda.gov/reports.aspx?ID=17826
c) United State Department of Agriculture -- Department of Economic Research
   i) Usage: Dataset for Education by County
   ii) https://data.ers.usda.gov/reports.aspx?ID=17829
d) United States Fact Finder
   i) Usage: Downloaded datasets for Population and Unemployment
   ii) https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml

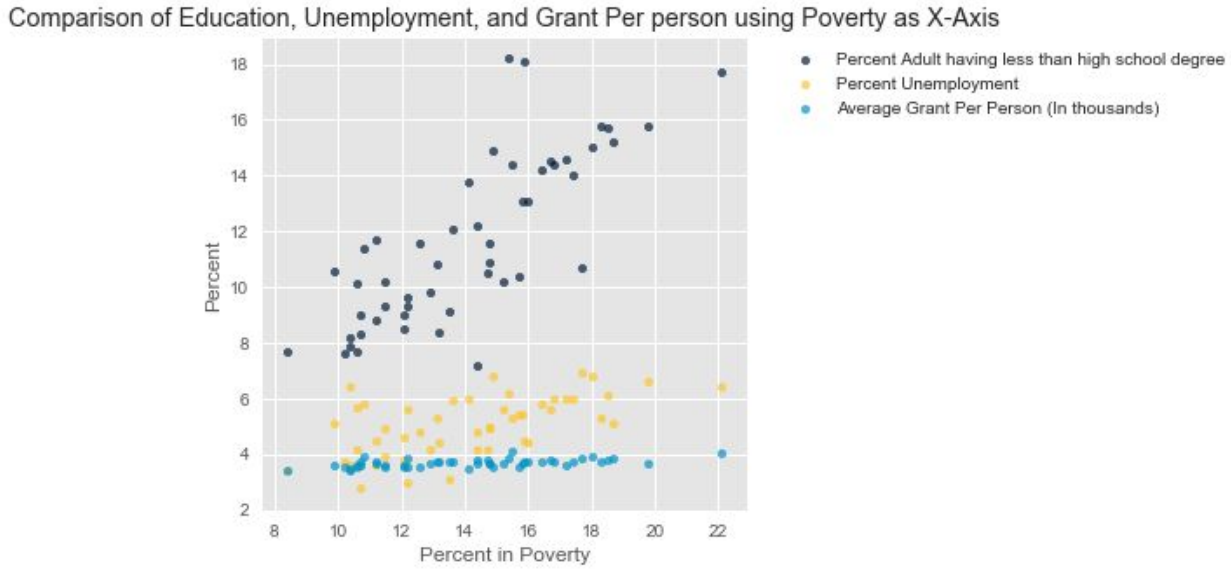
The FAFSA dataset can give us insight in how many students are receiving how much grants from State and Federal level. Also, the Education dataset provided by the United State Department of Agriculture -- Department of Economic Research would give us specific knowledge about how many people are attaining which level of education at which county. Combining both the FAFSA and the Education dataset, we can gain insight in whether having financial aid has any correlation with the highest education level attained in a specific location or the highest degree one can get.

On the other hand, we also want to compare the Education dataset with the Poverty Dataset provided by the US Department of Economic Research. We want to see whether financial aid actually help poorer county/state to in the average education level. Since the Poverty dataset is mostly by percent, we believe it might be helpful to also take a look at the Population dataset, which, if combined with the Poverty dataset, can give us the amount of people at each county/state who are in poverty. And using that number, compared with the Education and Financial Aid dataset, we can clearly see whether Financial Aid is actually effective.

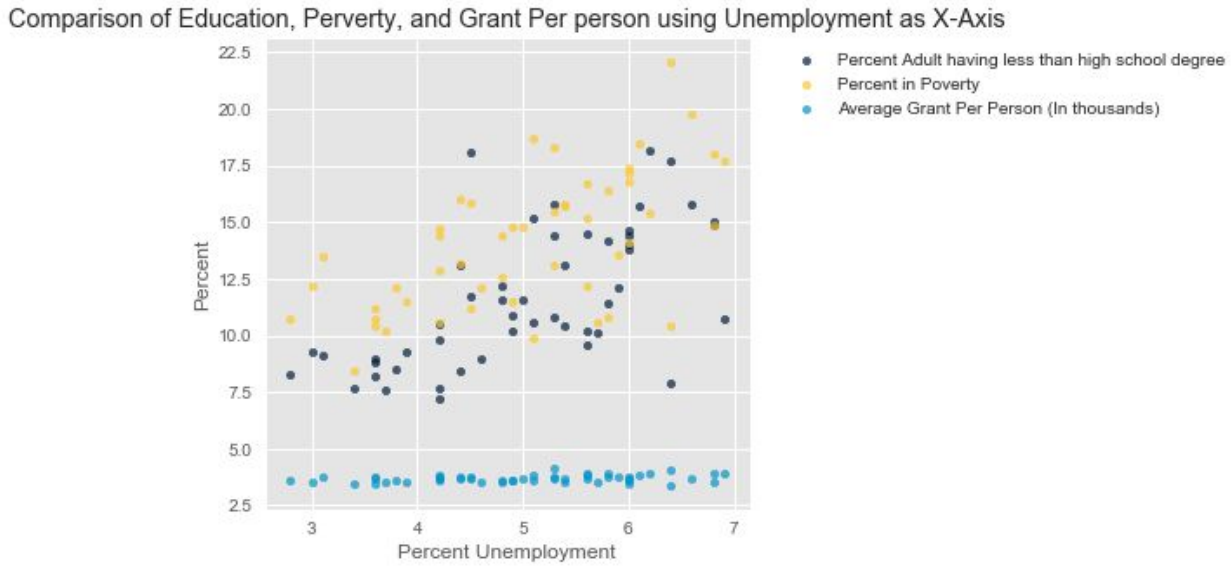
Visualization



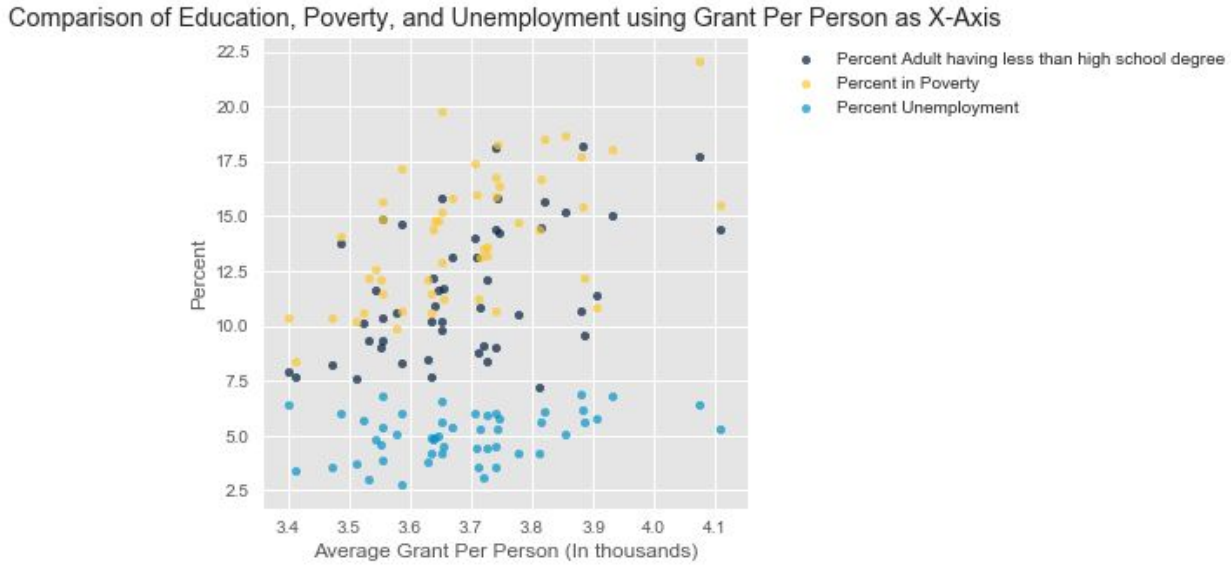
This plot shows that as the percent of adult having less than high school degree increases, the average amount of grant received per person did not change. Although we cannot observe a correlation between education and financial aid, we instead observed a correlation between lower education level and Poverty.



This graph confirms what we seen from the last graph -- there is a positive correlation between education and poverty, but no correlation between amount of financial aid and poverty.



This graph tells us that there is a positive correlation between unemployment, education and poverty, but no correlation between those factors and financial aid. Which confirms what we saw in the previous two graphs.



This graph is more interesting because it tells us something that goes along our intuition. As the amount of financial aid increases, the percent poverty also increase and so does percent of adults having less than high school degree. However, this correlation seems pretty weak because all of the black dots and yellow dots seem concentrate in the 3.5 - 3.7 area rather than spreaded out. And the correlation between unemployment and financial aid is nearly none by eyeballing.

Conclusion

We can see from the first three plots, with unemployment, education, and poverty as x-axis, the trend of the average financial aid per person is nearly a straight line, indicating that as the other three factors increase or decrease, the average financial aid per person does not change significantly for different states.

We think that this is a very interesting result, because counter to our intuition of the poorer one state is, the more financial aid one should receive.

By going one step further, and computed the quantitative correlation between Financial Aid and the other three parameters (please refer to visualization.ipynb), we can conclude that Financial Aid has low correlation with the other three parameters, which are poverty, education, and unemployment rate that we are analyzing in our model. Relate back to our motivation at the beginning of the project, the amount of Financial Aid is roughly the same despite different socioeconomic backgrounds ones from.

Final Thoughts

Since this project is heavily code based, there are several major technical challenges that we had to overcome.

The first is how to read in data correctly. Every dataset has its own way of organizing its values and thus caused problems when reading in. For example, each dataset has its own way of enumerating and indentation of title, and that made us to debug quite a while.

Another technical difficulty would be that we have too much data and how to effectively visually represent it. We found that making one plot that contain all the information is ineffective because there's simply too much information and would make it impossible to understand. Therefore, we used multiple scatter plots to plot the relationship between different parameters instead of making one comprehensive graph that contains everything.

And the technical difficulty we spent the most time on was how to use a certain library. For instance, Panda and Basemap is very new to us and it took us quite a while to get the hang of how to use its functions. One way that we overcame this difficulty is to transfer data all into data science Table since all of us took Data 8 and know how to operate in that environment.

On the other hand. We found that our conclusion agrees with the related work (please refer to our presentation) that we searched up.

The related work indicated that there is a trend that the lower the socioeconomic status, the less likely that the person is going to file FAFSA and get Financial Aid. Therefore, we conjecture that this phenomenon might contribute to our conclusion. If the FAFSA data fails to incorporate all socioeconomic population, then our data might be fluctuated upward and thus got the conclusion that there is low correlation between poverty and financial aid.

Unfortunately, due to the limit of time, we cannot investigate further in this matter in the duration of this project. But, for sure, it would be an interesting aspect of research.

Reference

https://stackoverflow.com/questions/39742305/how-to-use-basemap-python-to-plot-us-with-50-states
http://matplotlib.org/basemap/
https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html
https://matplotlib.org/api/pyplot_api.html
http://data8.org/datascience/tables.html
https://seaborn.pydata.org/
http://www.numpy.org/