IAT 355

Final Project Spring 2021
Instructor: Chris Shaw
TA: Aunnoy Mutasim

James Yoo
#301341943

Overview

The right to education is one of the key principles in our life and everyone should have the same opportunity to receive the education they need. Using data visualization, we will explore if there are parental levels of education and parental income can have an influence on a student's performance in school and graduation rate. And if these are factors, identify the patterns and address the issue.

Introduction

Scope

In this data visualization, we are focusing on whether parental level of education and income can influence child performance in school and years to graduate.

Audience

The audience of this project is for the law makers for education such as politicians, ministry and teachers to understand the influence on the student's factor to succeed in school and provide clear visualization for this data.

Domain Questions

1. Do parental level of education have any influence on the child's/student performance
2. Do students with the resource or opportunity to have the test preparation score higher on exams?
3. Do parental income influence student's GPA in high school and college
4. How does parental income and level of education influence student's years to graduate?

Data

For this project, I have used two datasets from Royce Kimmons:

Exam Scores: http://roycekimmons.com/tools/generated_data/exams
Graduation Rate: http://roycekimmons.com/tools/generated_data/graduation_rate

Dimension

- Exams scores
    - math/reading/writing score (quantitative)
    - Parental level of education (nominal)
    - Test preparation course (nominal)

- Graduation Rate
    - Parental income (quantitative)
    - Years to graduate(quantitative)

- High school/college GPA (quantitative)

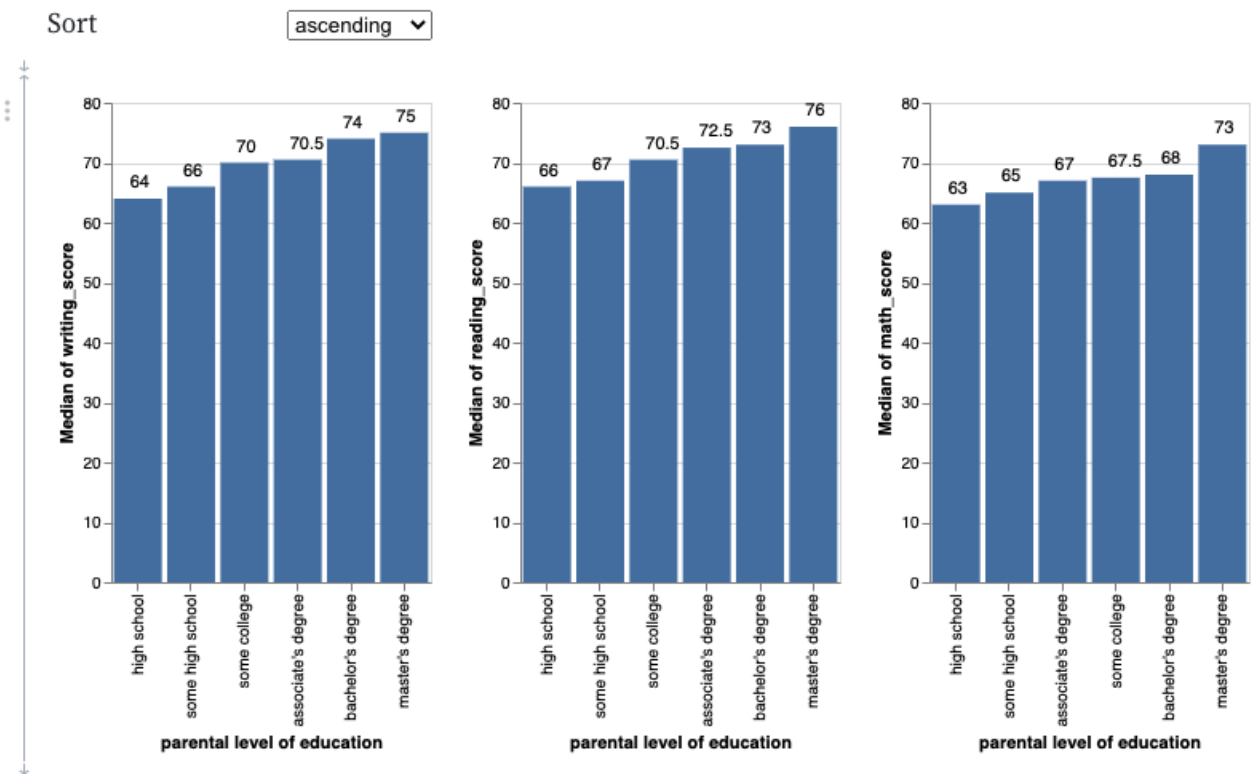
Cleaning

- The two dataset had some dimensions that were not needed in this visualization or to answer the domain questions. The dimensions that were taken out were:
  - Exam score
    - Gender
    - - race/ethnicity
    - Lunch
  - Graduation Rate
    - ACT Composite score
    - SAT total score

Visualization Design

1. Do parental level of education have any influence on the child's/student performance

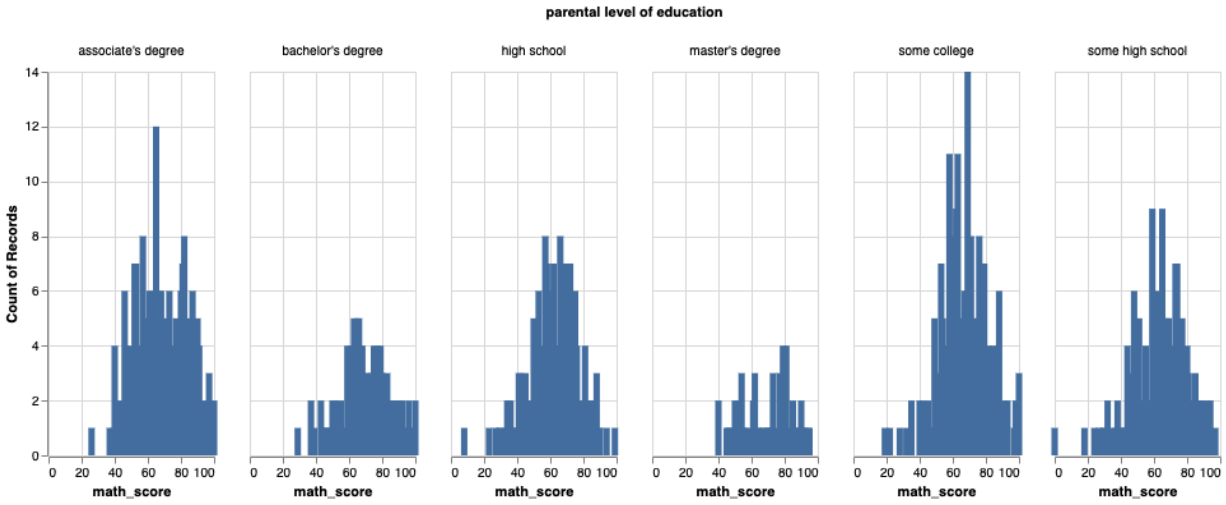
Fig. 1

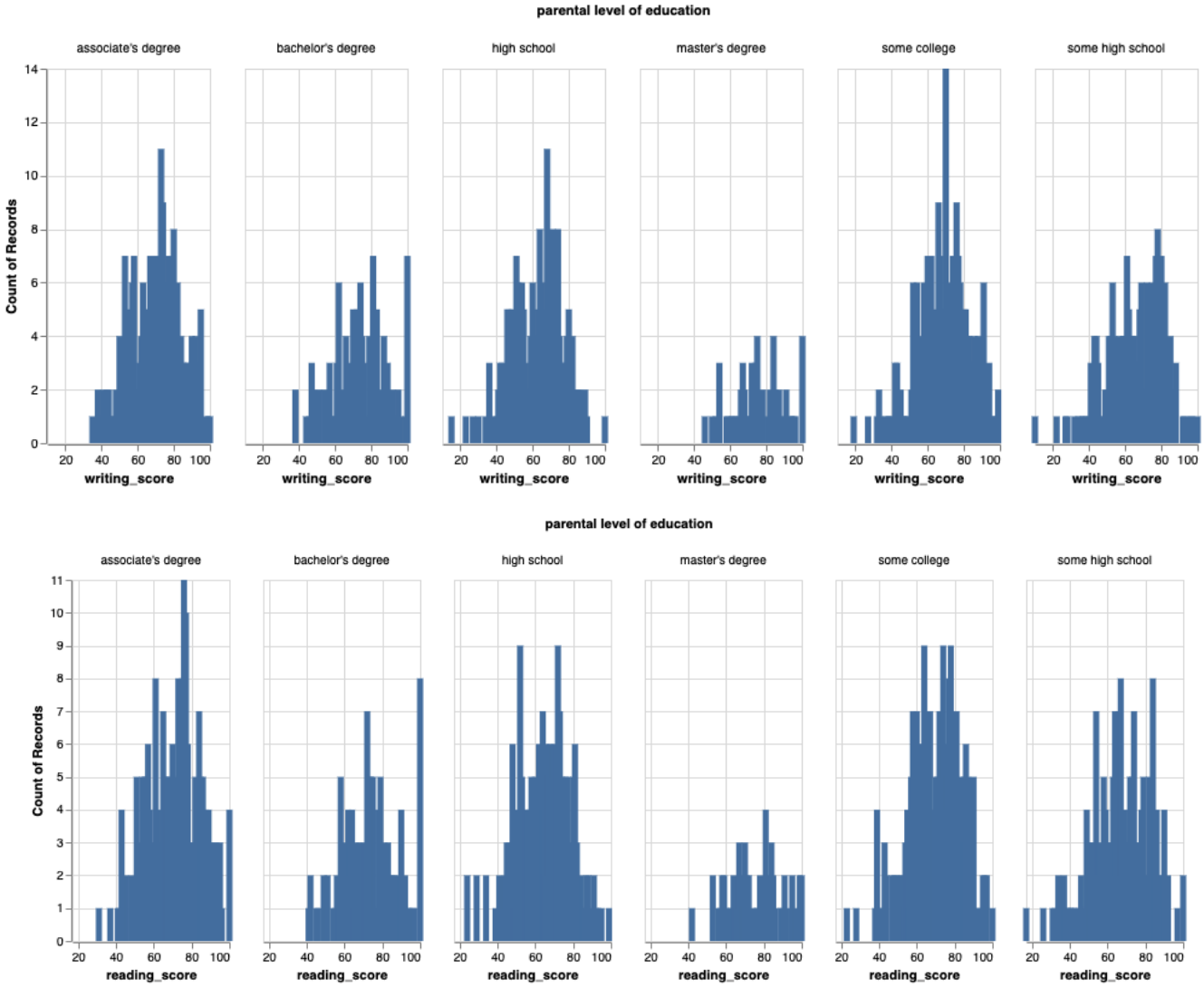


For figure 1, I have used histogram to discover the relationship between parental level of education and the student score on the writing, reading and math test. The histogram shows the median value of individual tests divided into sections of parental level of education. The interaction in this visualization is the ability to sort

histogram from ascending to descending. This allows the user to easily identify the extremum and also sort the data cases.

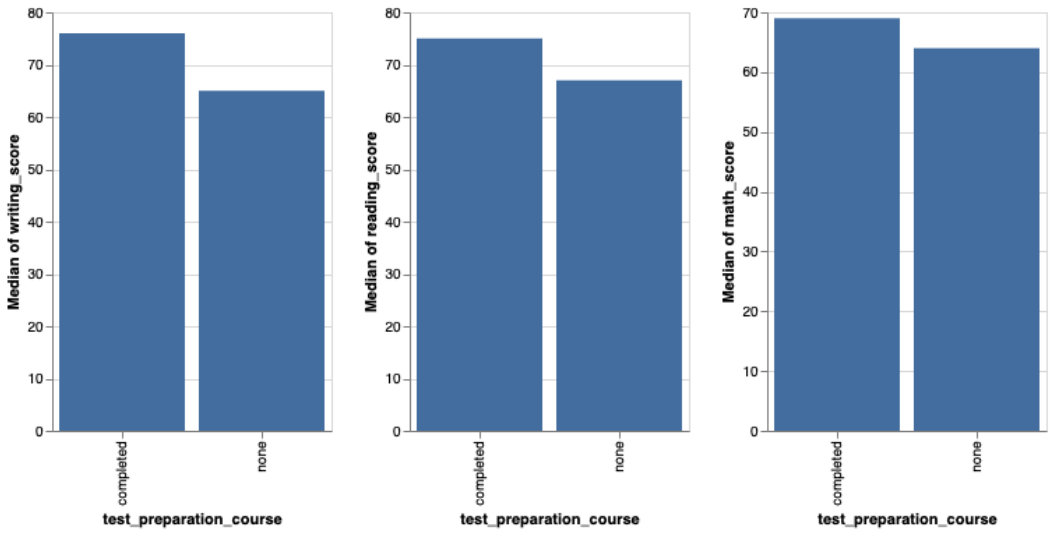
Fig. 2





For figure 2, I have used histogram to display each parental level of education performance on each test using facet technique. It shows subset of the x dimension, which in this case is the parental level of education and

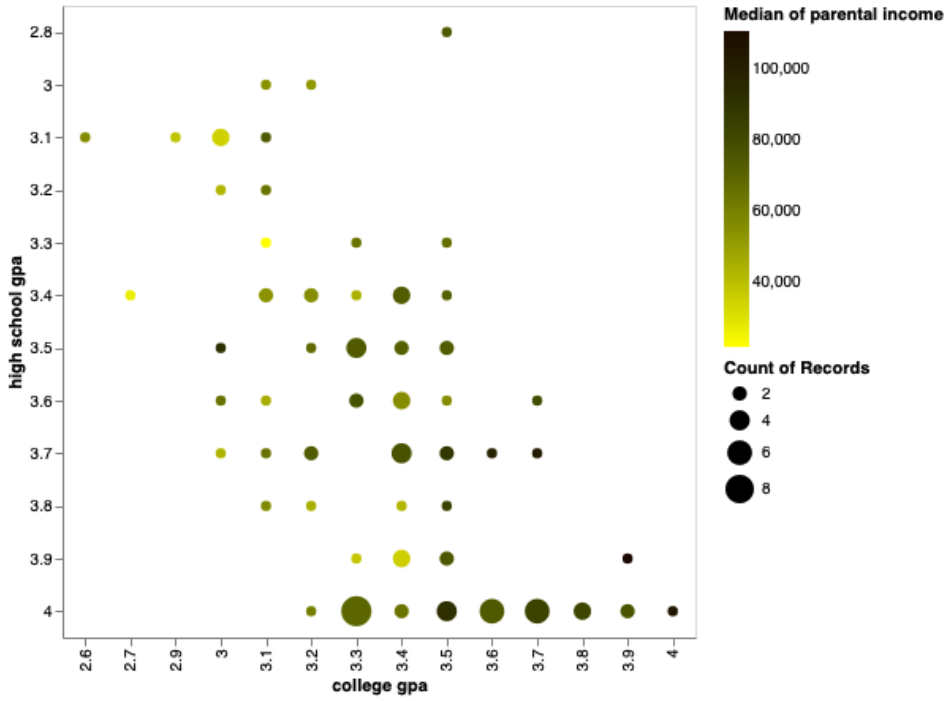
2. Do students with the resource or opportunity to have the test preparation score higher on exams?



This visualization is using histogram to discover whether the test preparation course aided students in getting better marks on the exam. Although test preparation has nothing to do with parental involvement or their background, it may be an indication that the family is wealthy or financially stable enough to afford the student in test preparation course or academy while other households may not be able to afford to do so.

3. Do parental income influence student's GPA in high school and college

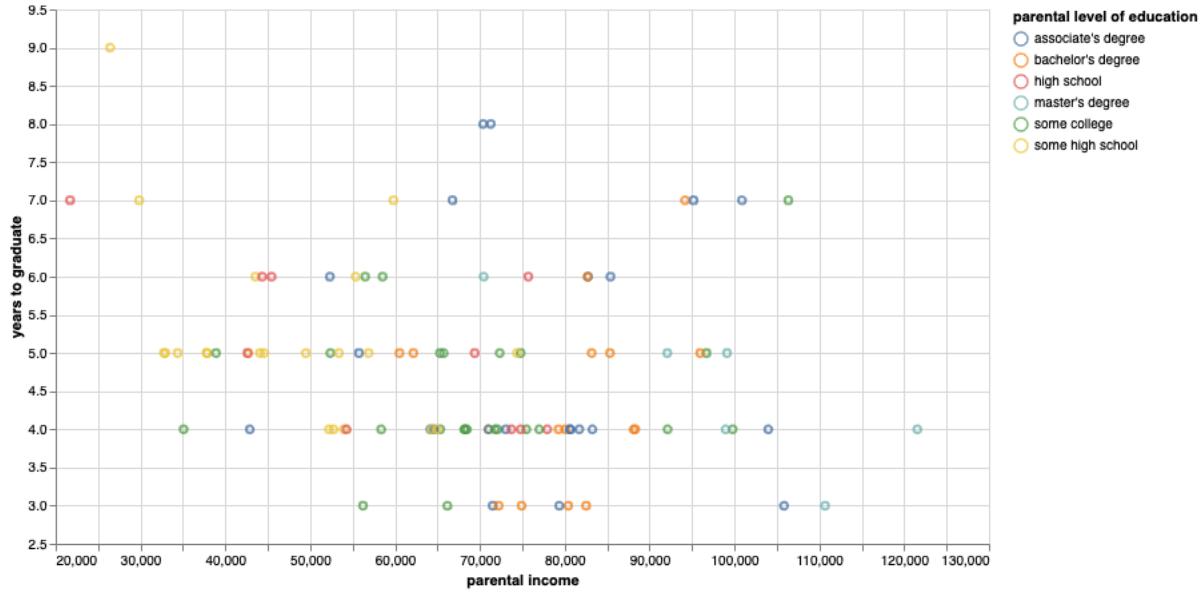
Fig.3



This focuses on the relationship between college and high school GPA with color labelled by median of parental income. We want to see if there are patterns or trends of students with good grades in terms of median parental income.
Using 2D histogram scatterplot, we can see that the darker circles in the histogram means high parental income and as it gets lighter, its less income, and the size of the circle represents the count of records. Although parental income has nothing to do with a student's performance in school, it can relate to the support and the resources they get from which go back to question 2. They can have the resources to get extra help such as test preparation course, academy or after school program to help or not having to work and focus in school work.

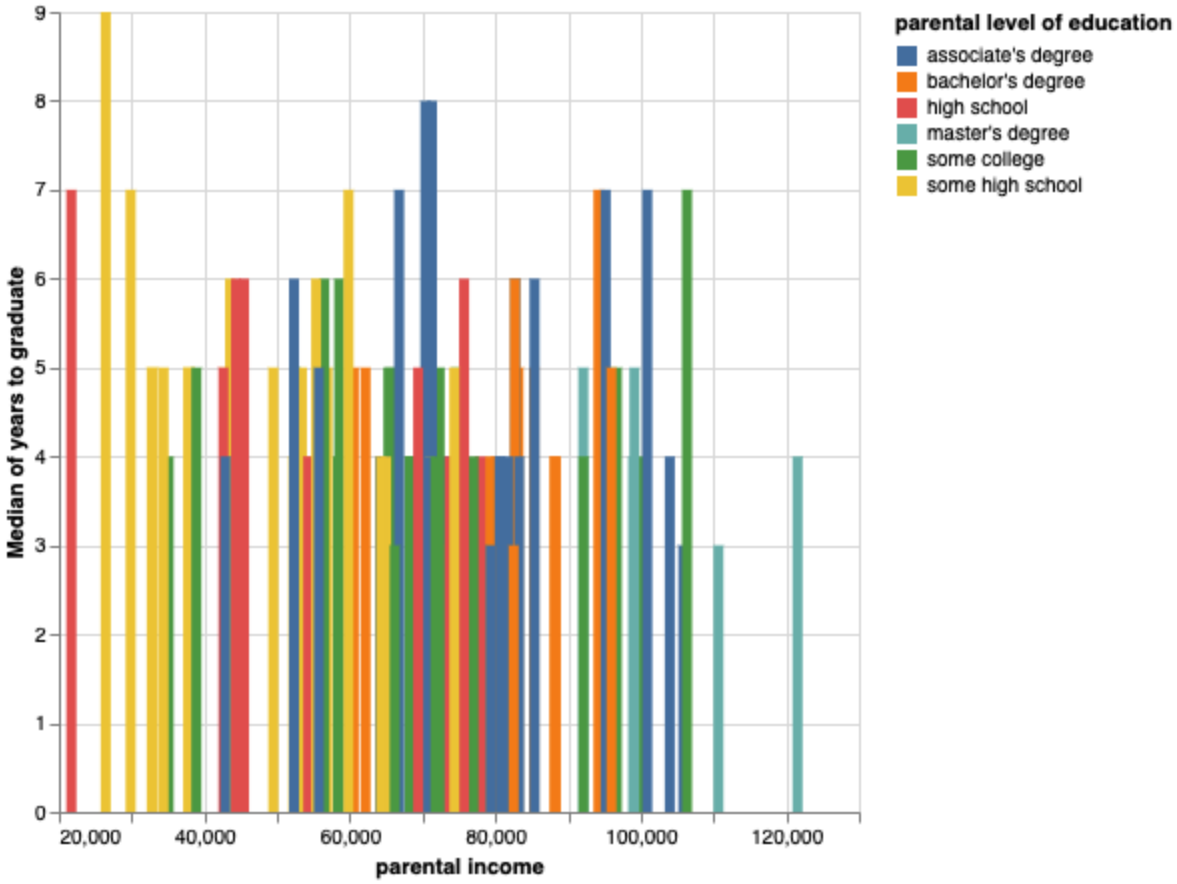
5. How does parental income and level of education influence student's years to graduate?

Fig. 5

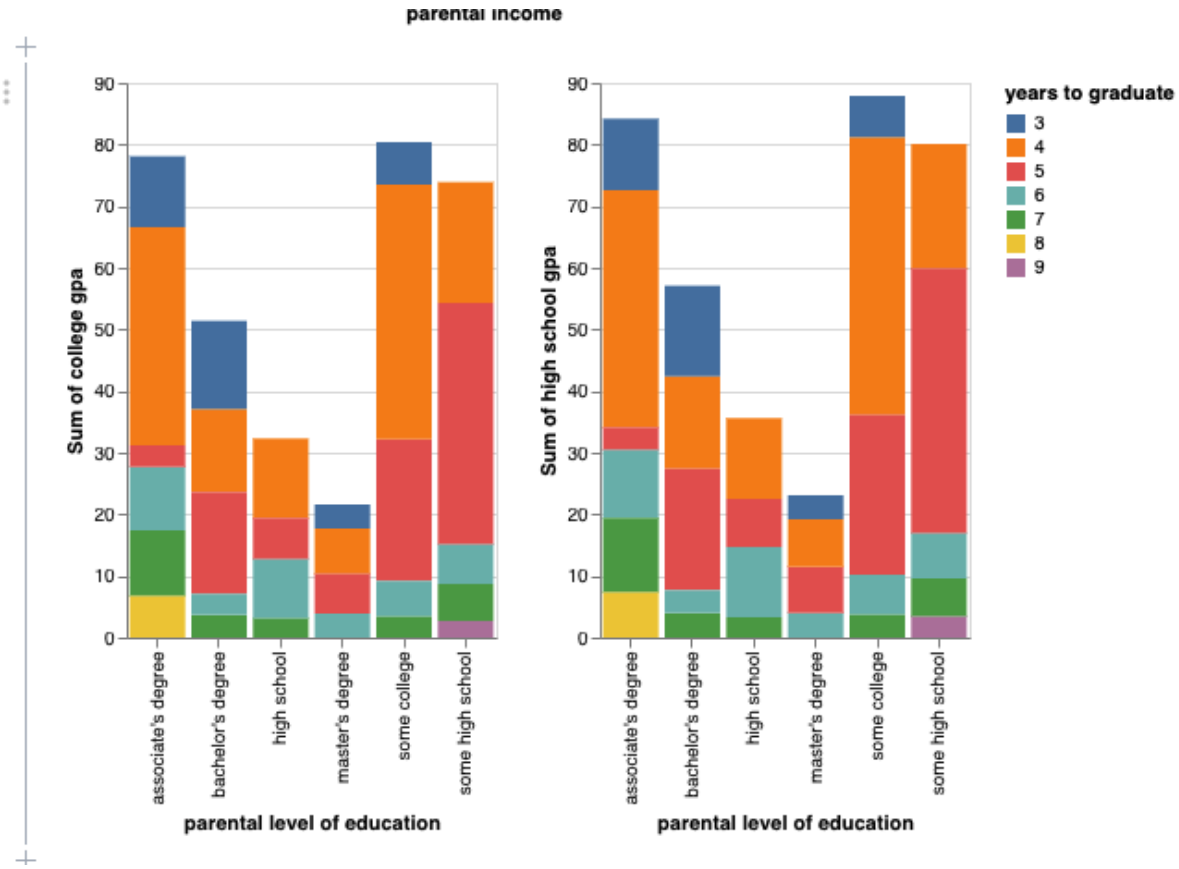


Country_parental_level_of_education null

This data visualization shows the relationship between parental income and years to graduate colored by parental level of education. It allows the user to choose which attribute in the parental level of education to filter out the plots in the graph. Through this data visualization, it allows user to see the populated scattered plots to find about trends and patterns of parental level of education. It was hard to see the pattern when everything was colored but only being able to see one attribute of the plot allows to see the trend and find out the populated area.



This dataset visualization purpose is very similar to the visualization above but this aims to find the median value of people in different incomes and how long it takes them to graduate. The colored bar graph shows easy indications to identify the patterns or populated colour bar to discover what ..



This is a stacked bar to show the proportion between the variable of years to graduate. This displays how many students there are with different parental levels of education and divides students into the color of how many years to graduate. This essentially allows us to identify what is the most common and the least "years to graduate". The interaction was to click on the colored bar and it will show the value but it does not work for both of them.

Further work

My intention was to discover and identify the trends and common patterns in the visualization which worked but not to an extent of what I expected. I wanted to find more datasets that include factors that may influence a student's performance related to parent's but could not find any.

<u>What did not work?</u>

I think the main problem was understanding how to use Vega-Lite and understanding how to use them so the product is the visualization I wanted. There were multiple occasions where I did not know how to make certain visualizations and I used Vega-Lite website examples (which will be in reference) as a guide but there were occasions where I tried the exact but did not work. It may be because I did not understand how to code them properly. The major code problem I faced was working with tooltips and displaying value when clicked on it.

I was forced to remake the Observable multiple times because it would not save which was very frustrating leading me to just publish multiple times whenever there were changes.

Most important struggle I had were writing the report to show the reasoning for the visualization and why or how they work. I have no problem understanding my visual and explaining it to myself the reasoning but wording them on a report was very difficult. I struggled with how to word them.

What would you do differently?
On my next iteration or version, I think my visual was good but if I could make better visualization and have more datas to add, that would be great. Secondly, learn more about data visualization and understand or learn better on how to write better reports. Thirdly, learn more about Vega-Lite coding and other frameworks to implement.

Conclusion
This project has displayed how students can be impacted or influenced in general from the parental level of education and income. Understanding the flaw and pattern in this visualization can help school and education ministry to perhaps make new support for students who are not getting the same resources as other students who are fortunate.

Overall, I learned alot for this project as this was a solo work It was a massive learning experience for me.

Feasibility Pilot
For this feedback on the prototype, I have asked my classmate and friends if they can give me feedback on my prototype and one of the most feedback I received were using too much bar graph and everything seems very similar where x is parental level of education and y being some sort of quantitative value. It was impossible to get rid of exactly that but I decided to explore more datas examples on Vega-Lite website and looked at every example and in my head, I tried to visualize whether certain visuals will work and depending on that, I have made changes.

References
http://roycekimmons.com/tools/generated_data/exams
http://roycekimmons.com/tools/generated_data/graduation_rate
https://vega.github.io/vega-lite/