

Monash University
FIT5147 Data Exploration and Visualisation
Semester 1, 2022
Data Visualisation Project (DVP)

1. Project Title

Exploration of possible factors related to gun and murder safety issue in the United States

2. Identity

Name: Jiang , SiaoHsuan

Student ID: 33029229

Tutor's Name: Mohit Gupta

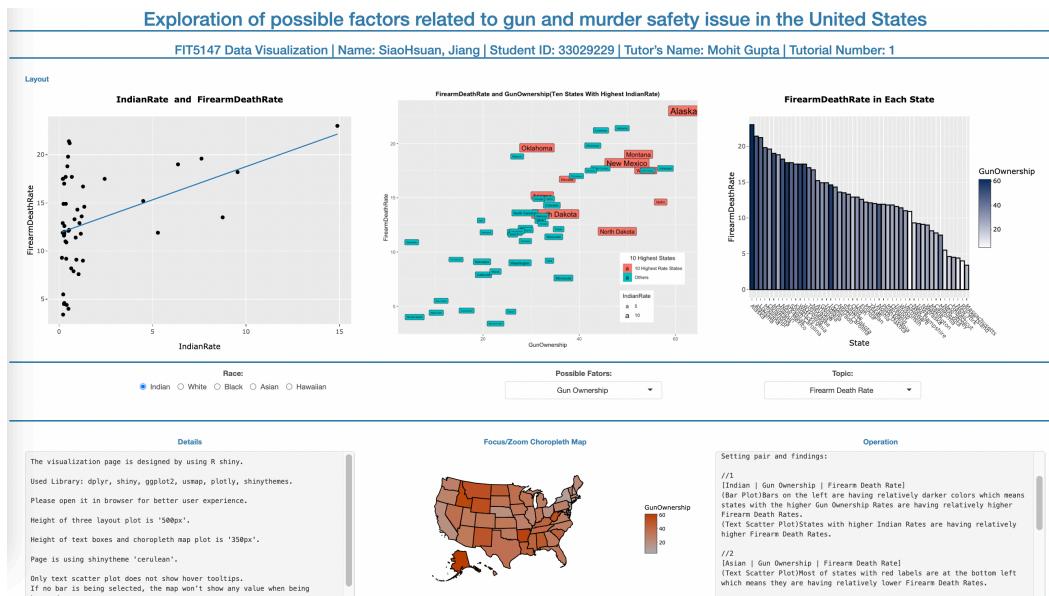
Tutorial Number: 1

3. Introduction

I wondered which state is safer to live in, and I feel safety issue including both gun and murder aspects is worthy of exploring because holding firearms is legal in United States. In this project, I found **Indian Rate** are having positive relationships with **Firearm Death Rate** and **Murders and Nonnegligent Manslaughter Death Rate**. Further, I also found that possible factors like **Gun Ownership**, and **Poverty Rate** are in a significant relationship with **Firearm Death Rate**. In terms of race, **Black** and **White** Rates are having significant relationship with **Murders and Nonnegligent Manslaughter Death Rate** under the evidence of statistic hypothesis test.

With proper visualization on real data, I strongly believe these analytics would help those who are planning to live in the USA and me to acquire better insights and understanding about some implicit risk in each state.

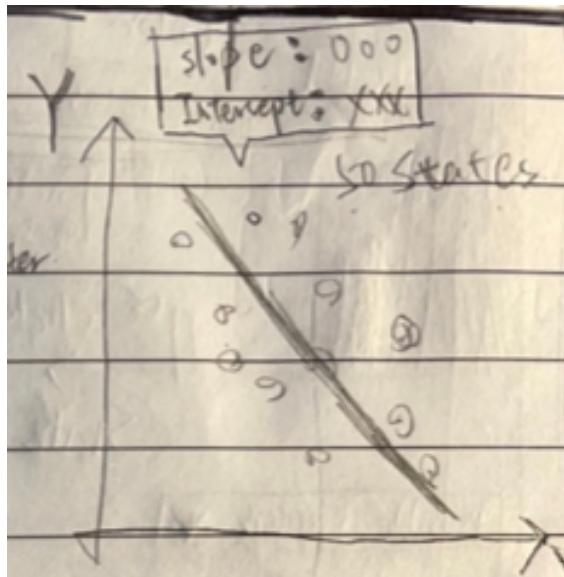
4. Design



Whole Page Screenshot

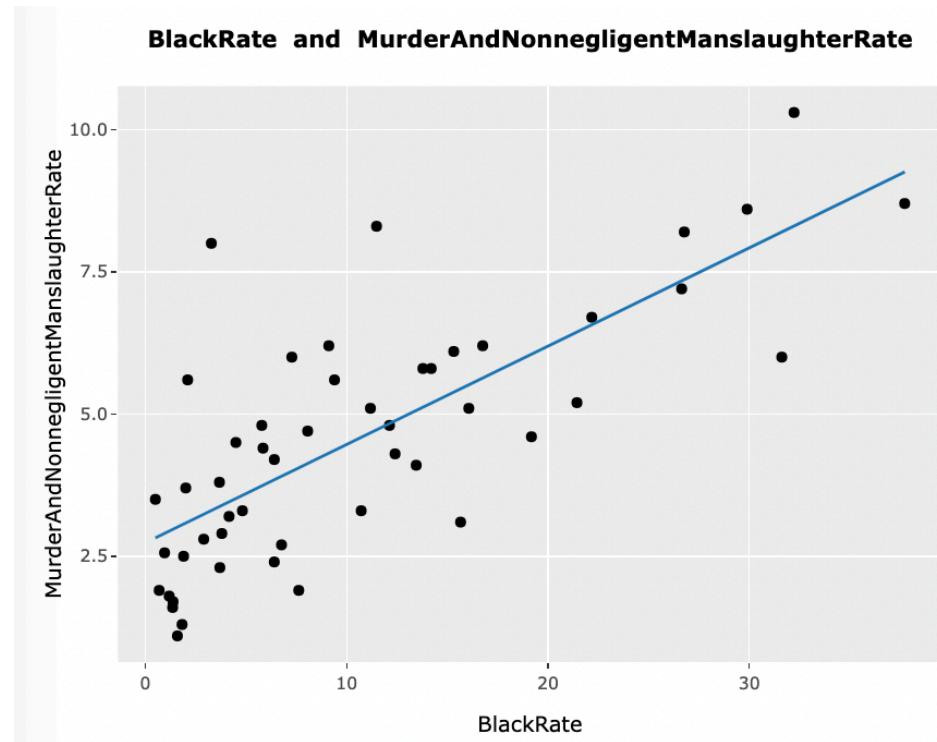
First Chart: Regression Plot

The first chart, scatter plot with regression line, refer to my sheet 2 design.



Regression Plot From Sheet 2

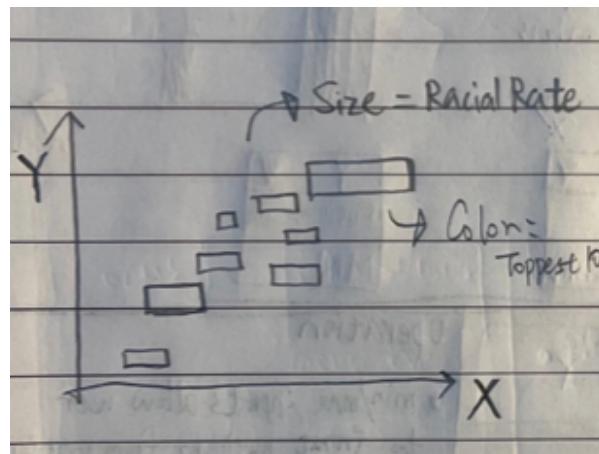
One reason why I select scatter plot is that it is a simple but best way to show relationship between two continuous variables. In addition, adding regression line and statistics results makes the scatter plot more informative and reliable.



Regression Plot From Design Page

Second Chart: Text Scatter Plot

The second chart, scatter plot with text and colored label, refer to my sheet 4 design.

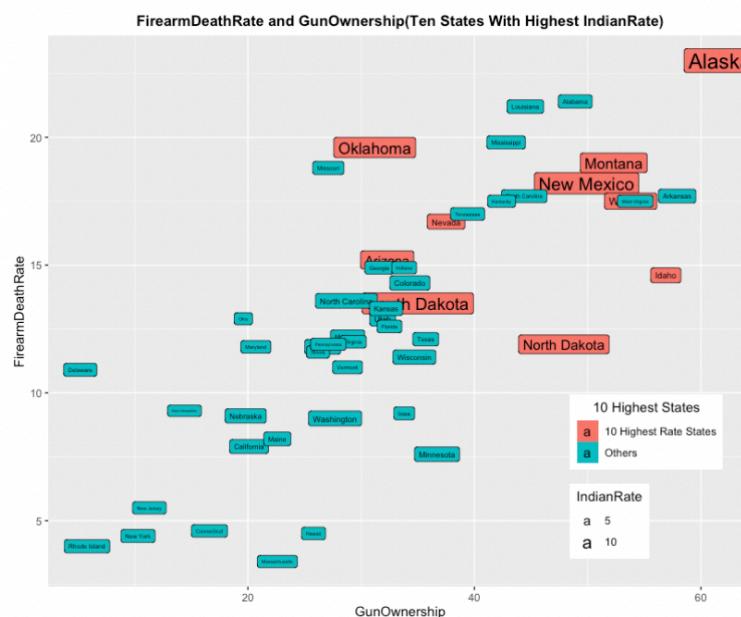


Text Scatter Plot From Sheet 4

As the same reason mentioned above, scatter plot is a good way to show relationship. In this graph, I took the advantage of size and color to add more dimensions on a 2D graph.

- Size wise, I adjusted the size of texts and labels by racial rate.
- Color wise, ten states with highest racial rate would be visualized in a different color.

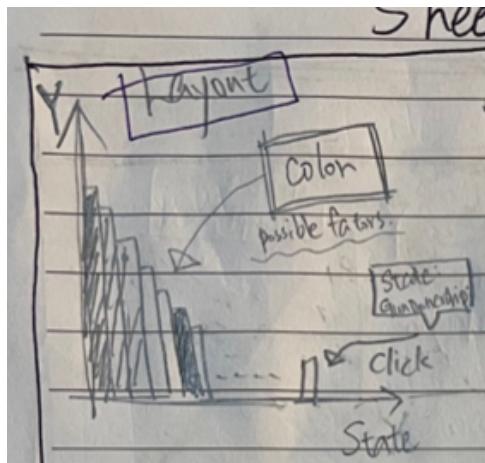
In this way, audience can easily figure out what states are having higher racial rates and whether these states are having higher rates in the possible factor variable at the same time.



Text Scatter Plot from Design Page

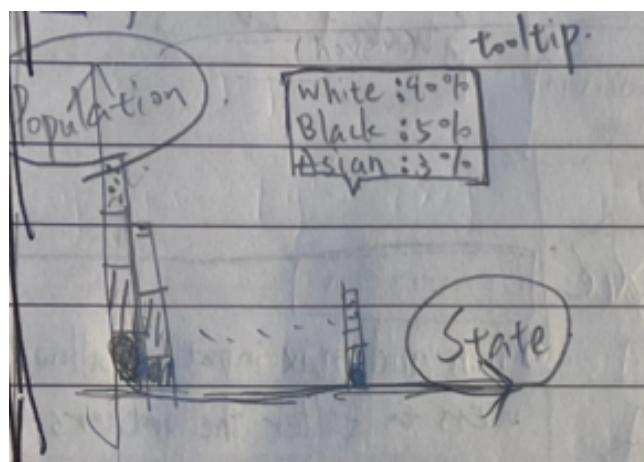
Third Chart: Bar Plot

The third chart, bar chart with continuous palettes, refer to my sheet 3 design.



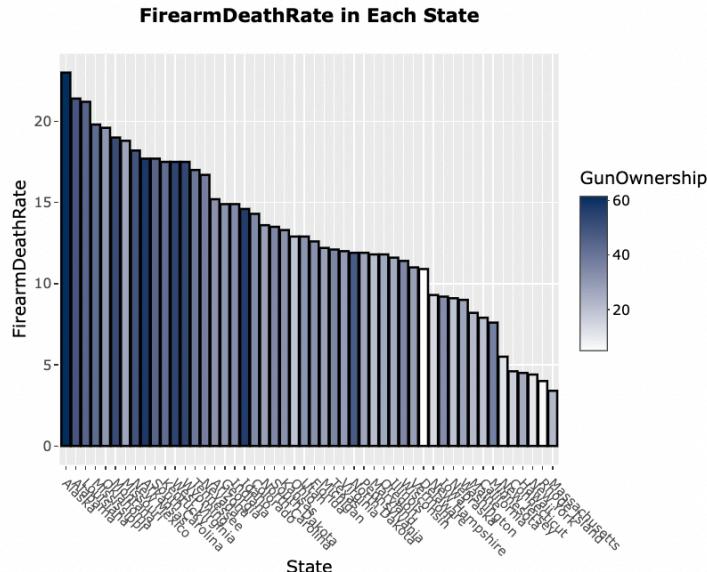
Bar Plot From Sheet 3

Other similar bar plots like piled bar graph in sheet 2 and sheet 4 are also good visualization methods. However, I chose to use bar graph with continuous palette. The main reason I got rid of piled bar graphs is that bars with continuous palette is more informative than piled bar graphs in my project. For example, this piled graph in sheet 2 just show the relative population proportion of different races in each state. It would be hard for audience to compare rates of one specific race among states.



Piled Bar Plot From Sheet 2

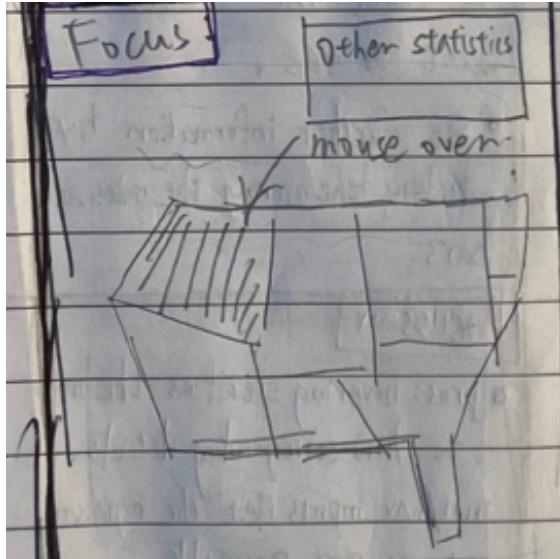
In the chart below, I used color to visualize selected possible factor which is a continuous variable. Since the variable is continuous, utilizing the merit of continuous palettes is a good way. Another reason for using this color palette is that human visual system enables people to easily perceive and compare the relatively higher or lower values through colors. After ordering the colored bars by the value of main interested variable, audience can clearly observe whether states with higher possible factor rate (darker colors) are in a higher death rate.



Bar Plot From Design Page

Focus/Zoom Chart: Choropleth Map

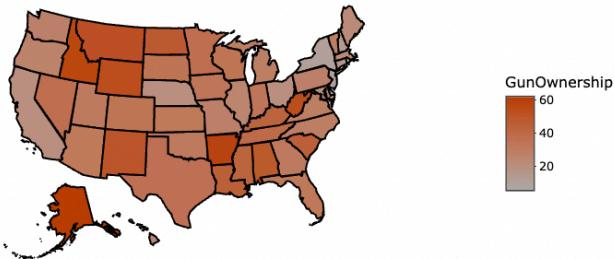
The focus/zoom chart, choropleth map graph with continuous palettes, refer to my sheet 2 design.



Choropleth Map From Sheet 2

My map design shows information of each state in detail when hovering. Audience can not only compare values of a group of joint states but also observe whether there is a tendency between different sides or areas like western America and Eastern America since the color makes it easy for audience to perceive the differences among states.

GunOwnership Choropleth Map



Choropleth Map From Design Page

Manipulation About Variable Selection

As to the manipulation part, radio buttons and select bars are having the similar purpose. They allow users to select one variable in each group to visualize. I placed five radio buttons as one input for five different races and two select bars for two possible factors and two main topics.

Race:

- Indian White Black Asian Hawaiian

Radio Buttons

Topic:

Possible Fators:

Gun Ownership

Gun Ownership

Poverty Rate

Firearm Death Rate

Firearm Death Rate

Murder and Nonnegligent

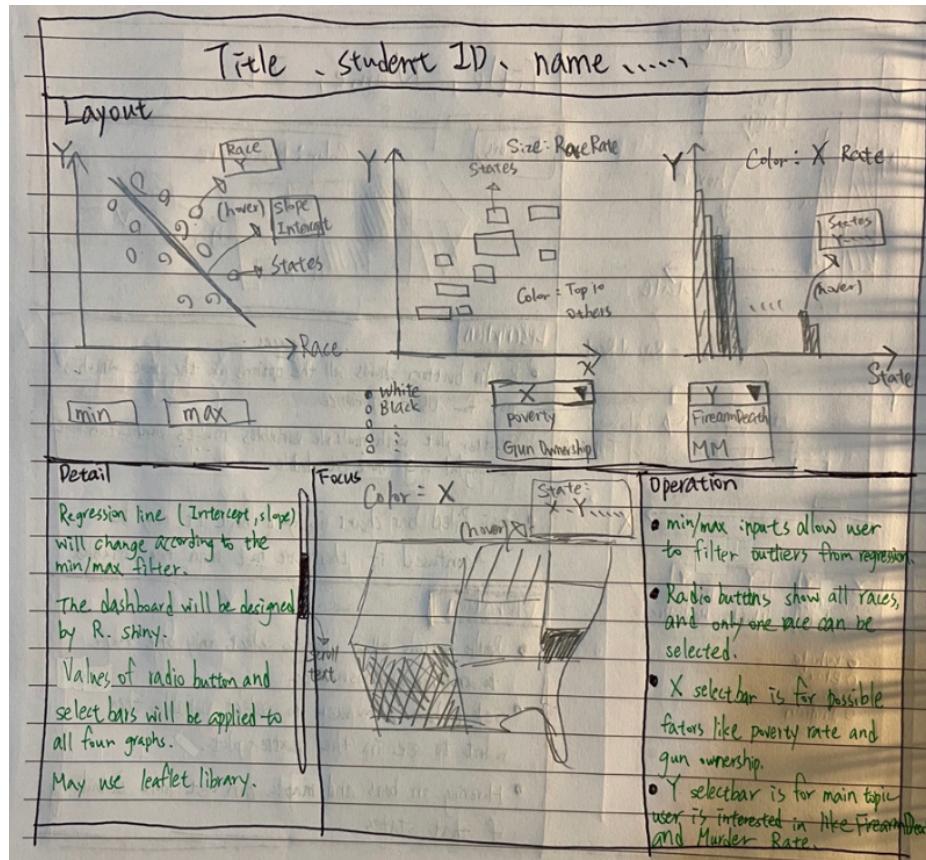
Manslaughter Rate

Two Select Bars

Although three select bars can reach the same goal too, I didn't design as so. The reason is that five races are dependent variables to each other. Different manipulation way makes users to tell the difference between race and the other two groups of variables.

Output Frame

Finally, I chose output style in sheet 2 as my final frame design. Since I decided to place three plots in my layout, I think frame page like sheet 2 is more suitable than the others. Text outputs including detail and operation are placed on each side of the focus/zoom map plot.



Sheet 2 Frame Style

5. Implementation

Libraries used:

```
library(dplyr)
library(shiny)
library(usmap)
library(plotly)
library(ggplot2)
library(shinythemes)
```

Used Libraries

`dplyr` is a powerful library with its own pipe grammar and special verb functions. I used it to do data manipulation.

`shiny` library makes it easier to build interactive apps from R.

`usmap` library provides longitude and latitude data of United States which is used for printing a choropleth map. It builds a plot based on `ggplot2`.

`plotly` library is used to turn a normal `ggplot` into an interactive graph.

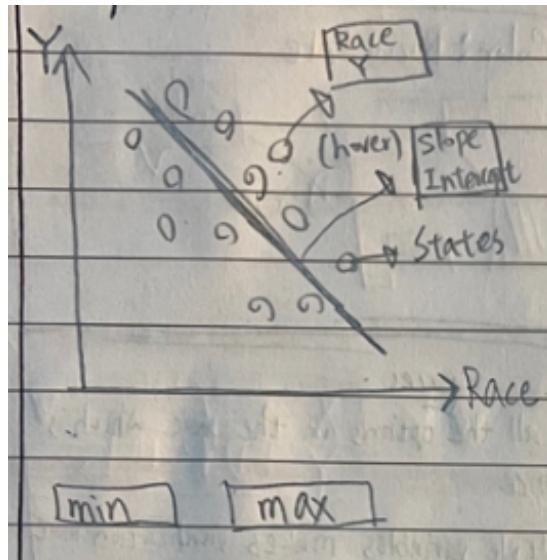
`ggplot2` library allow users to compose a graph by adding independent components like legend boxes or title texts.

`shinythemes` library is used to make the output more stylish and colorful.

Two differences between final design and implementation.

- Remove min and max input boxes.

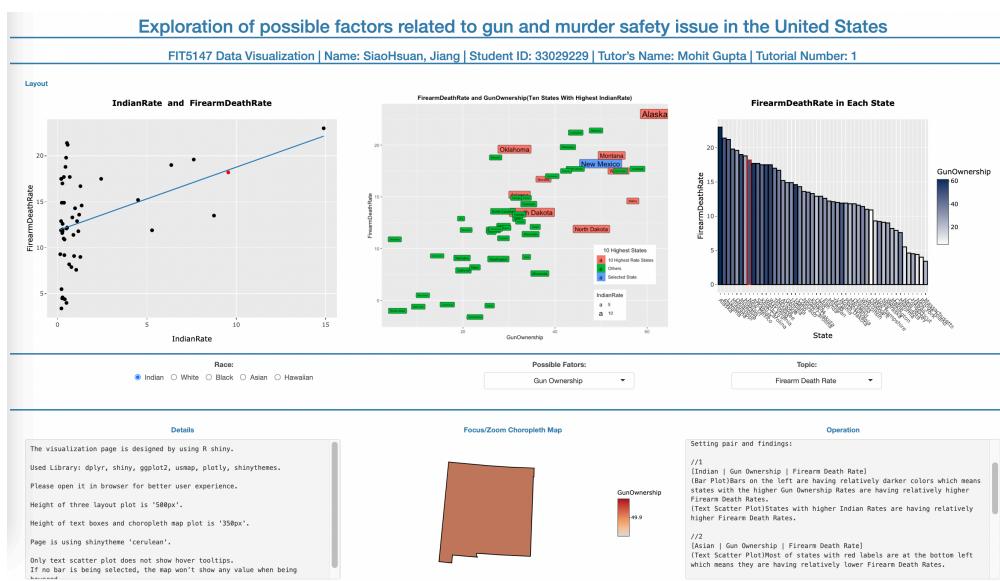
Those two boxes were supposed to let users to filter the outliers of regression plot. On the 5 sheets presentation day, I was recommended to remove them since letting users to manipulate the system and filter the outliers are not what I should concern in this project. Therefore, following the instruction, I removed them from my final design.



Min/Max input boxes in sheet 5

- Plot and plot interaction.

On the final content class day, tutor mentioned that plot and plot interaction is part of the marking criteria. Hence, I came up with some new ideas and added more complexity into my design after seeing the demonstration from tutor in the class.

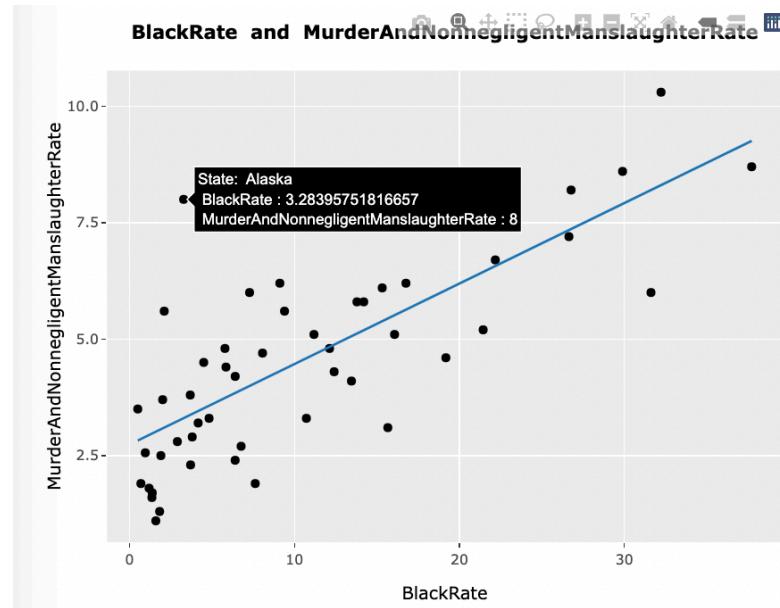


Plot-to-plot Screenshot From Design Page

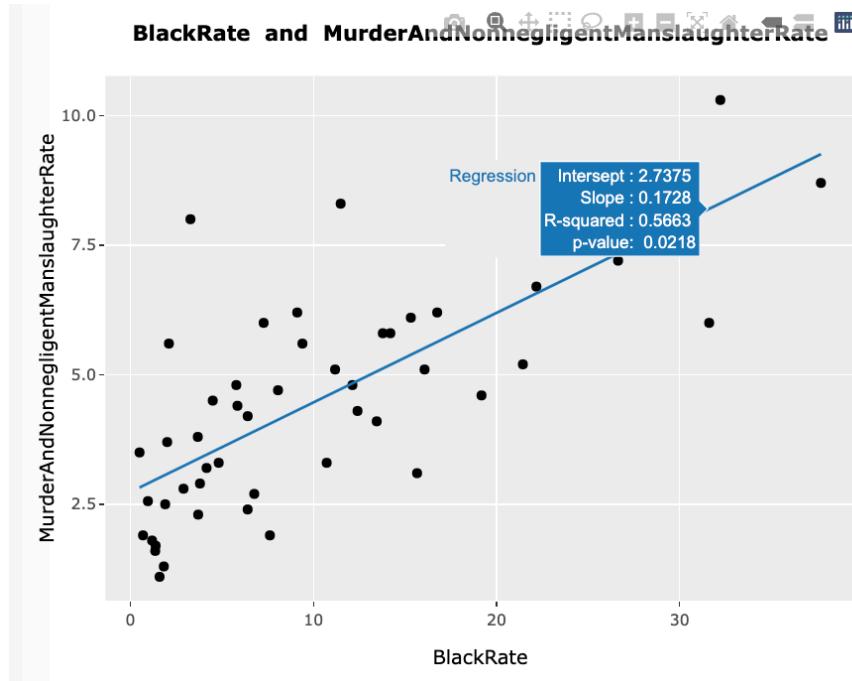
Challenging Part:

- **Regression Plot**

In the regression plot, the regression line wouldn't show up when turn it into interactive plotly plot. Therefore, it massively increased the difficulty to not only visualize a graph but also pop up hover message in the tooltip. I had to combine different methods in different library to make tooltips of both scatter plot points and regression line work perfectly in different text styles.



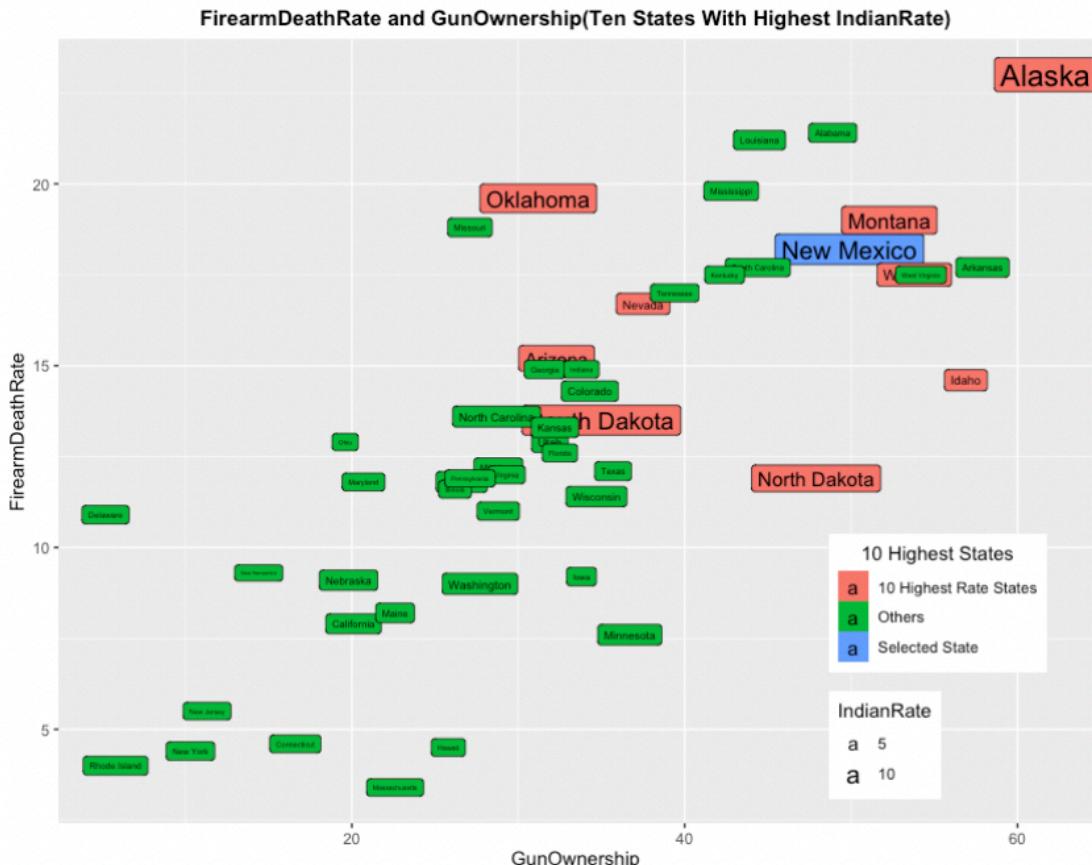
Tooltip Of Point In The Regression Plot



Tooltip Of Line In The Regression Plot

- **Text Scatter Plot**

In terms of the text scatter plot in the middle of the layout, I had to do further data wrangling to get more variables for multidimensional plot. Moreover, because of the plot-to-plot interaction, I had to add one more special signal for the selected state which increase the complexity as well. I will explain more about plot-to-plot interaction in next plot.

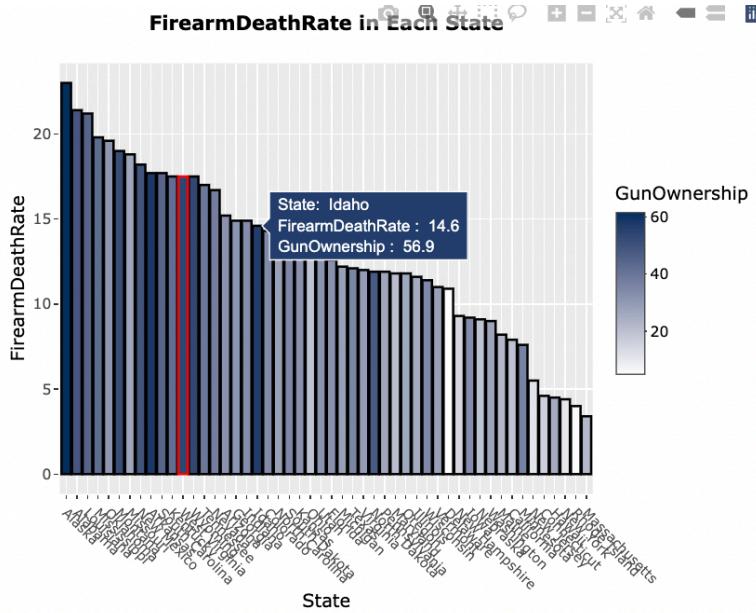


Text Scatter Plot With One More Color (Plot-to-plot interaction)

- **Bar Plot**

The bar plot is the most complex one. It controls all plot-to-plot manipulation. First, getting values from click events is not as direct as D3. Second, after getting the value, I had to spend a lot of time on trial and error to assign values to all the plots including itself. In total, I wrote two kinds of event, clicking and double clicking, for plot-to-plot interaction.

Third, shiny renders events pretty slow. There is an obvious lag when events occur. To solve this problem, I'm using clicking instead of hovering in the plot-to-plot interaction.

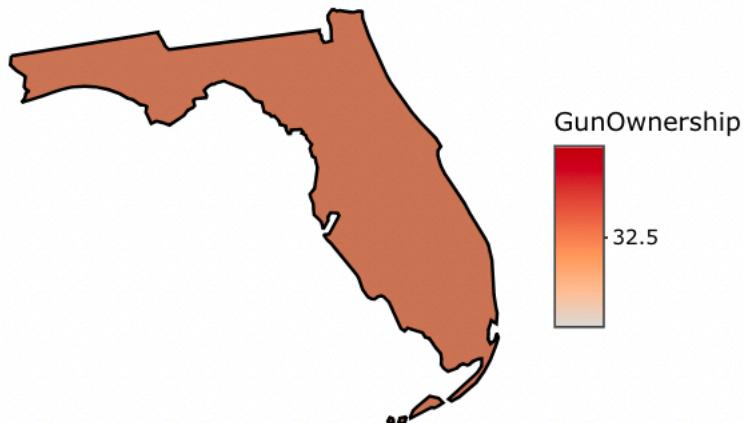


Bar Graph When Click Event Was Triggered (Normal hover works as well.)

- **Map Plot**

About the complexity in map graph, if one bar in the histogram plot is clicked, the map will zoom in to focus on that corresponding state instead of using the original whole America map which increase the complexity.

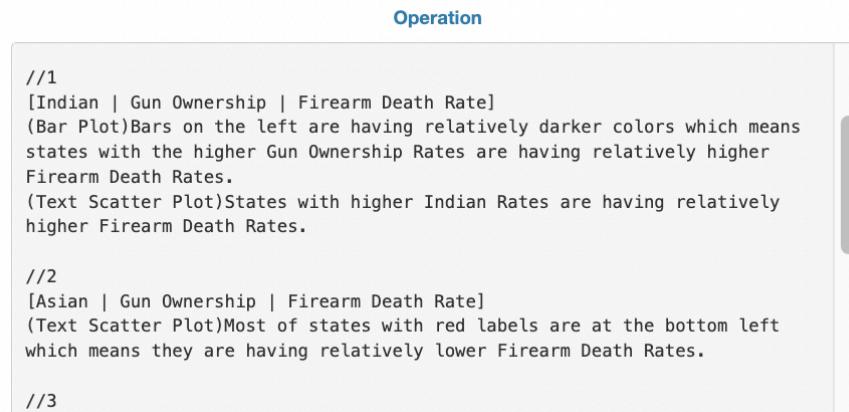
Choropleth Map



State Map When Click Event Was Triggered

- **Text Box**

To make the whole screen look tidier and more controllable, I set the border in every output. However, text box couldn't accommodate all of texts, so I made it scrollable. Making it vertically scrollable was not as easy as I expected in R shiny.



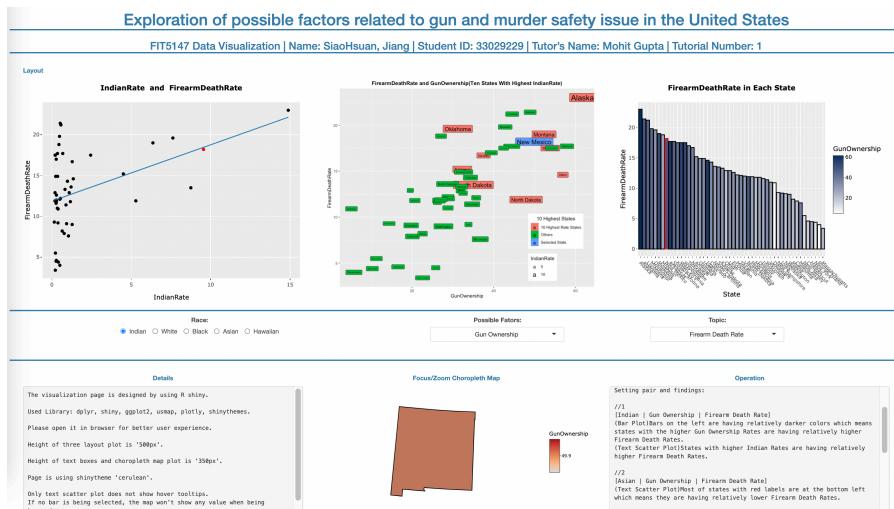
Vertically Scrollable Text Box

Note: All titles in all plots will change to corresponding variable names.

6. User Guide

Interactive Operation:

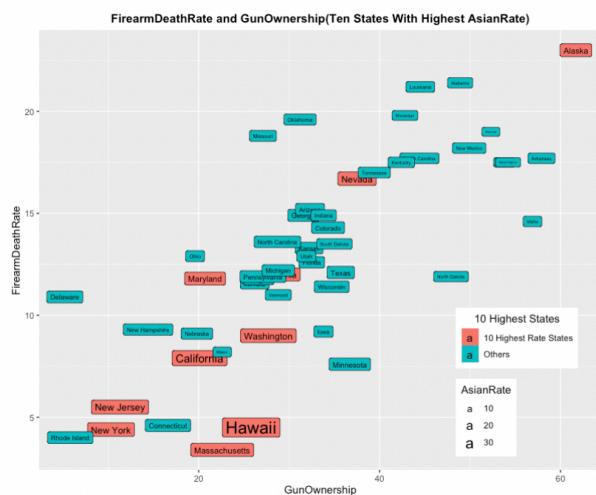
- Once you click on any bar in the bar graph, all 4 graphs will have some differences.
 - The point of the selected state will be highlighted to be red in the regression plot.
 - The label color of the selected state will change in the text scatter plot.
 - The border of the selected bar will turn to red in the bar graph.
 - The map will focus on the selected state, and the tooltip shows information of that state in detail when being hovered.



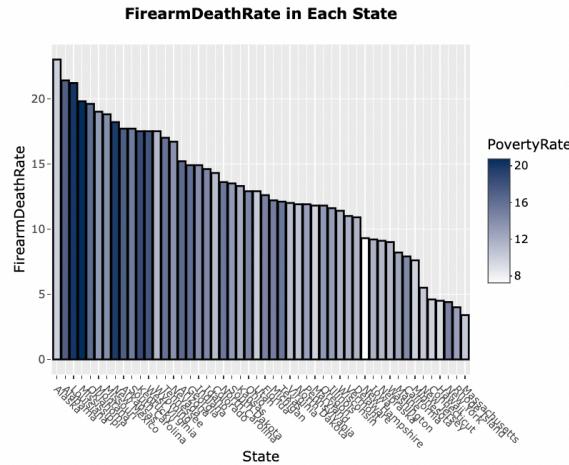
2. Once you double click (or double tap on your track pad) quickly on any place in the bar graph, all 4 graphs will return to the initial status.
3. Only text scatter plot does not show hover tooltips.
4. If no bar is being selected, the map won't show any value when being hovered.

Visualization:

- Step1. Before we do any manipulation, we can see some interesting things from this page. Look at the bar graph, we can clearly observe that bars on the left are having relatively darker colors which means states with the higher **Gun Ownership Rates** are having relatively higher **Firearm Death Rates**.
- Step2. We can observe this from text scatter plot in the middle as well. Moreover, in the text scatter plot, red labels symbolize the top ten states in **Indian Rates**. We can see states with higher **Indian Rates** are having relatively higher **Firearm Death Rates**.
- Step3. Next, if we change the Race from **Indian** to **Asian**, we can see most of states with red labels are at the bottom left which means they are having relatively lower **Firearm Death Rates**.
- Step4. Now, let's change Possible Factors from **Gun Ownership Rates** to **Poverty Rate**, and then observe the bar graph again. We can find the same result with **Gun Ownership Rates** which means the higher **PovertyRate** are having relatively higher **Firearm Death Rates**.



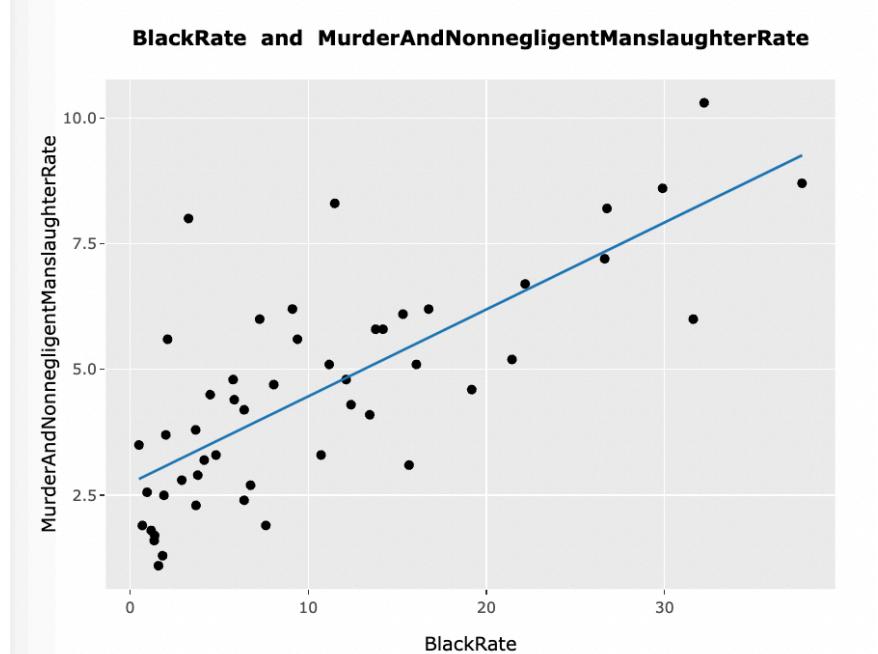
Asian And Firearm Death Rate (Step3.)



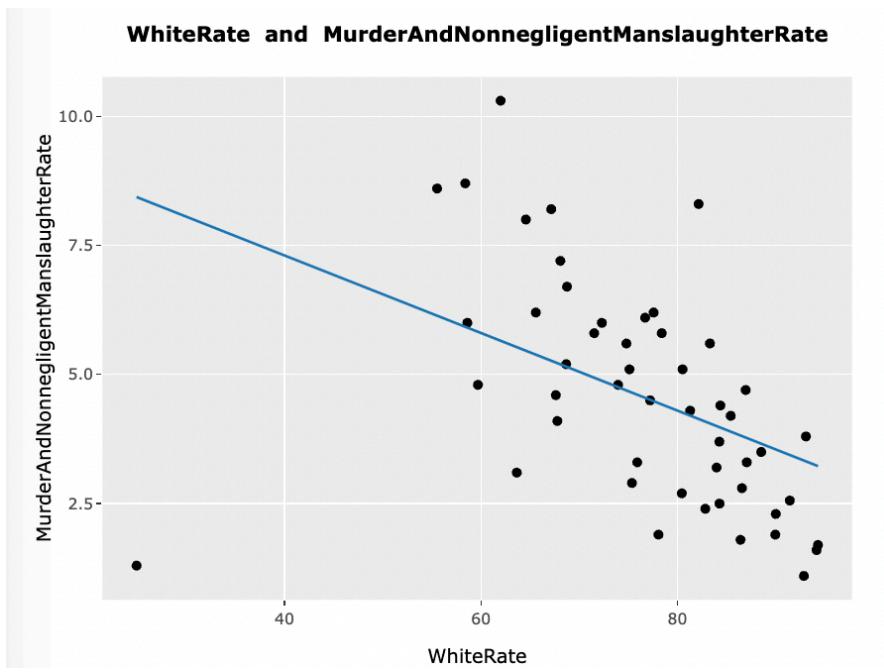
Poverty Rate And Firearm Death Rate (Step4.)

- Step5. After setting race to **Black**, and topic to **MurderAndNonnegligentManslaughterRate**, the regression plot indicates that there is a relationship between **BlackRate** and **MurderAndNonnegligentManslaughterRate**. More theoretically, when hovering on the regression line, we can see the hypothesis of slope = 0 is rejected by the p-value under the assumption of alpha = 0.5. (*p-value - Wikipedia*. En.wikipedia.org. (2022))
- Step6. We can also see the hypothesis is rejected when the Race is **White**.

In Step 5 and Step 6, when we hover on the regression line, we can not only find p-value but also coefficients of regression line and R-square. (*Coefficient of determination - Wikipedia*.
En.wikipedia.org. (2022))



BlackRate And MurderAndNonnegligentManslaughterRate Regression Plot



WhiteRate And MurderAndNonnegligentManslaughterRate Regression Plot

7. Conclusion

In this project, I not only learned how to practically put what we learned in class into practice but also learned that not every kind of races would be associated with **Firearm Deaths**. Like I assumed two main race rates **White** and **Black Rate** would have the strongest relationship with it but apparently not. Instead, two rates are having relationship with **Murder and Nonnegligent Manslaughter Rate** that I hadn't expected. Additionally, other factors like **Gun Ownership** and **Poverty Rate** are having positive relationship with **Firearm Deaths**. Finally, I found states with higher **Indian Rates** are having relatively higher **Firearm Death Rate** and **Gun Ownership Rate** while higher **Asian Rates** are having relatively lower **Firearm Death Rate** and **Gun Ownership Rate**.

If I could do it again, I should have reduced the subjective judgment when doing data exploration so that I could have analyzed the data more neutrally and saved more time on finding useless information.

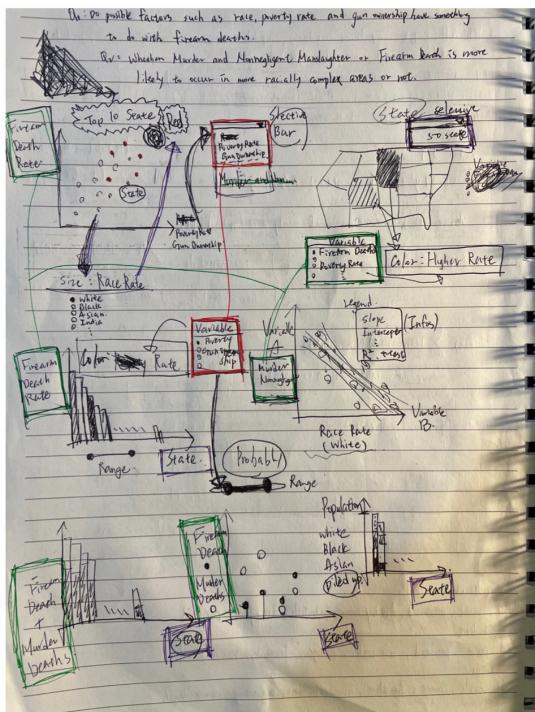
8. Bibliography

p-value - Wikipedia. En.wikipedia.org. (2022). Retrieved 29 May 2022, from <https://en.wikipedia.org/wiki/P-value>.

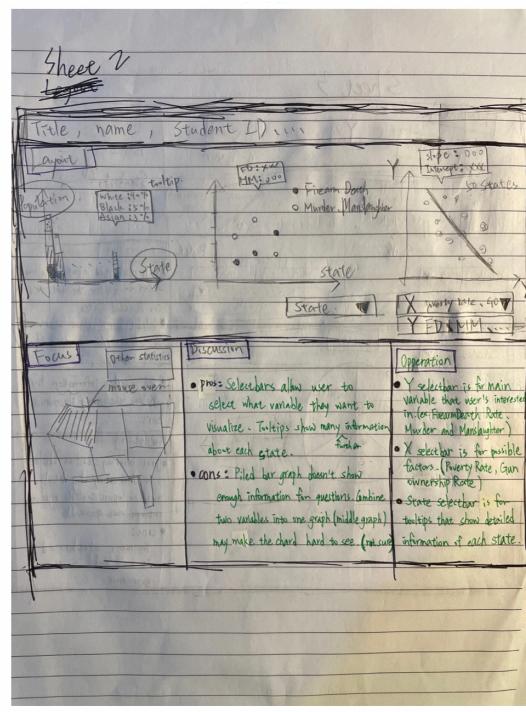
Coefficient of determination - Wikipedia. En.wikipedia.org. (2022). Retrieved 29 May 2022, from https://en.wikipedia.org/wiki/Coefficient_of_determination.

9. Appendix

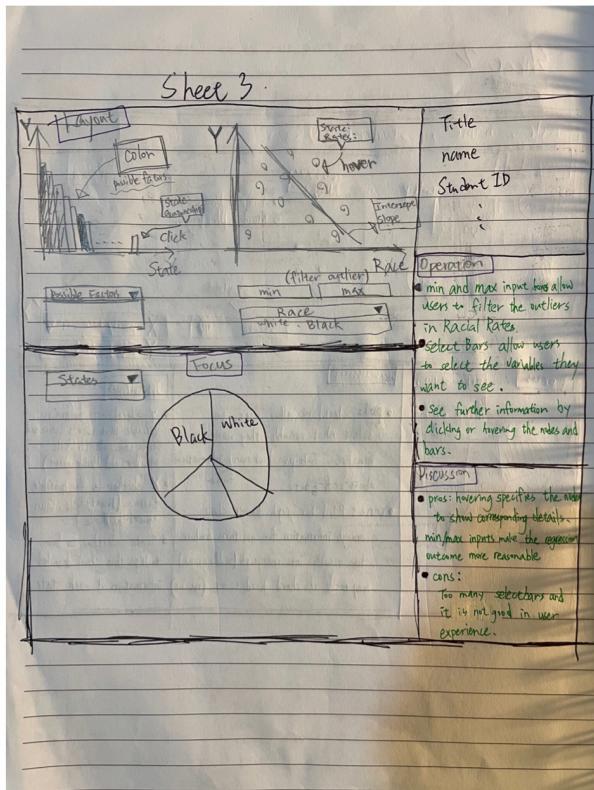
Sheet1



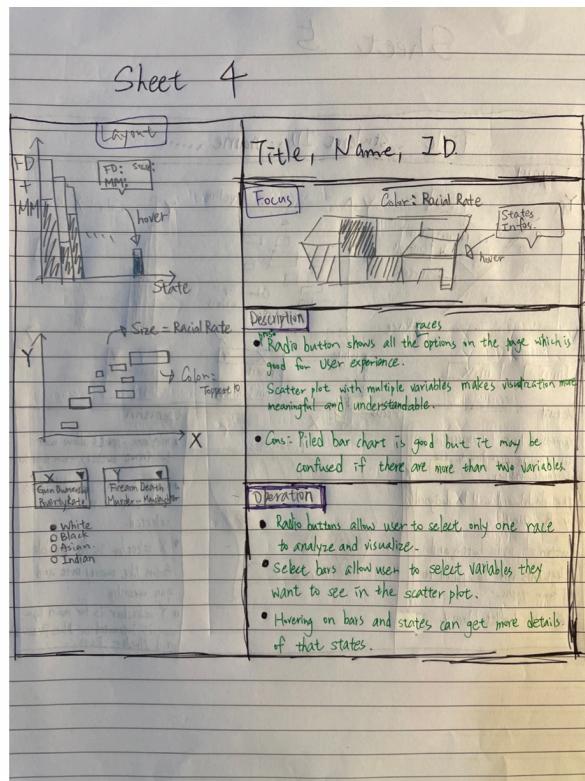
Sheet2



Sheet3



Sheet4



Sheet5

