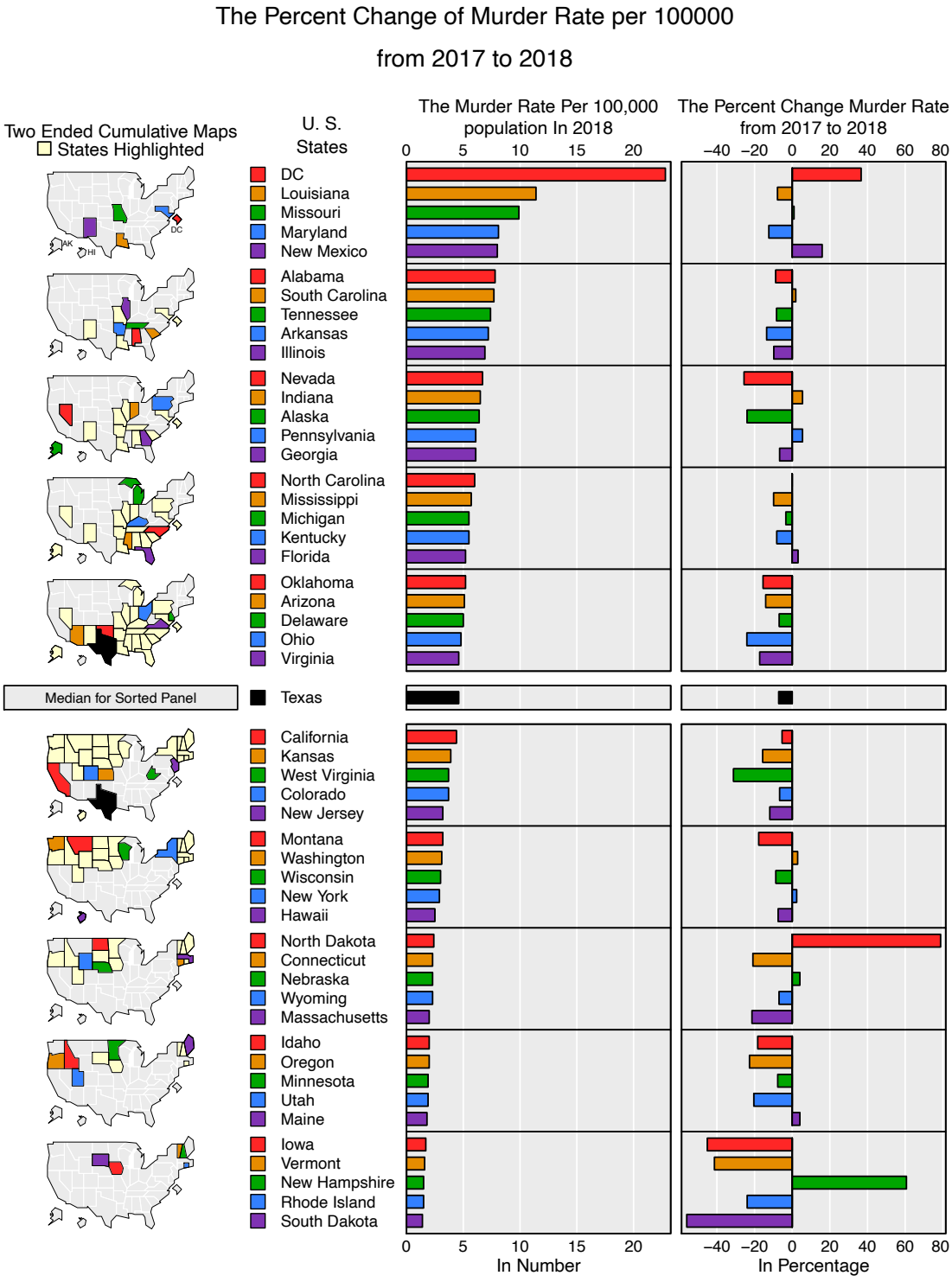
Linked Micromap Visualization Project



1. Introduction:

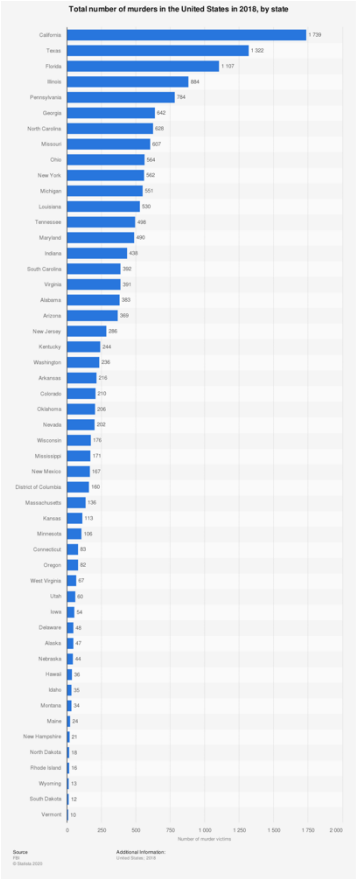


Fig. 1 Bad Graph

Fig. 1 shows the Total Number of Murders in the United States in 2018 per State (Statista, 2019). This bad graph is analyzed through the following 4 aspects:

- **Enable Accurate Comparisons:** According to the graph, the state with the highest total number of Murders is California (Statista, 2019). However, since the numbers of populations in each state are not the same, general people who read the graph can assume that California is the most dangerous state by solely assuming the highest number of murders. **Therefore, new variables should be considered in order to make the comparison between one state to another look fair.** Besides, without map visualization between each state to another, such graph will confuse readers. Map boundaries can help readers focus more on some areas of incidents, like the five states which have the highest murder number and vice versa.

- **Complete the Information:** If we refer to this bad graph, readers only can compare the value between one state and other states in the same year. However, readers sometimes want to know the increment or decrement compared with the year before.
- **Interpretation:** Since this bad graph does not fairly scale the population number in each state, the readers can misunderstand the graph and conclude that the most dangerous state is California. However, if the population numbers in each state are fairly scaled, the graph can give more reliable information to the reader.
- **Attract the Reader:** As we can see, each state on the graph is represented with the same color, which makes it difficult to do the comparison between one state to another. Additionally, less color cannot attract readers very well.

2. Redesign Goals:

The purpose of this project is to reduce above deficiencies in the original bad graph by scaling the population of each state into the same number (per 100,000 population), giving boundary visualization between every five states, and coloring states with different colors.

The new graph should provide more accurate and comparable information between one state to another, deliver more comprehensive information, give a better interpretation for readers, and attract readers more than before.

3. Redesign Process:

The main tool used in redesigning this bad graph is the micromapST package in R. Package "micromapST" provides the users with the ability to quickly create Linked Micromap plots for a collection of geographic areas. Linked Micromaps are visualizations of georeferenced data that link statistical graphics to an organized series of small maps or graphic images. Contained in this package are border group datasets to support creating micromaps for the 50 U.S. states and District of Columbia (51 areas), the U. S. 20 Seer Registries, the 105 counties in the state of Kansas, the 62 counties of New York, the 24 counties of Maryland, the 29 counties of Utah, **the 32 administrative areas in China**, the 218 administrative areas in the UK and Ireland (for testing only), the 25 districts in the city of Seoul South Korea, and the 52 counties on the Africa continent.

3.1. Data Preparation

In order to implement the micromapST function, we should calculate the data scale of 100,000 population of each state and also the data of percent change rate from 2017 to 2018. Afterward, the data should be stored into one dataframe and converted based on how the data will be used and performed in the package micromapST.

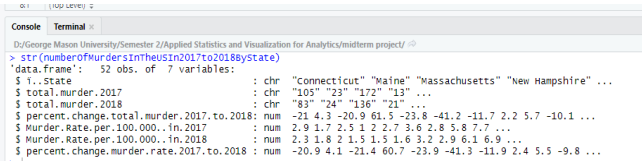


Fig. 2 The Structure of Converted Dataset

3.2. Graph Designing

Use micromapST package to redesign the original bad graph, shown in **Fig. 3.** In detail, the data is represented with map visualization, the murder rate per 100,000 and the percent change of murder rate.

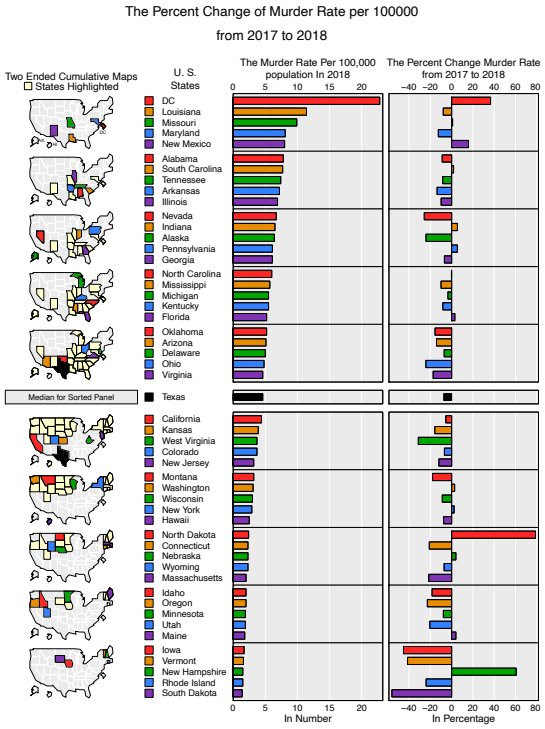
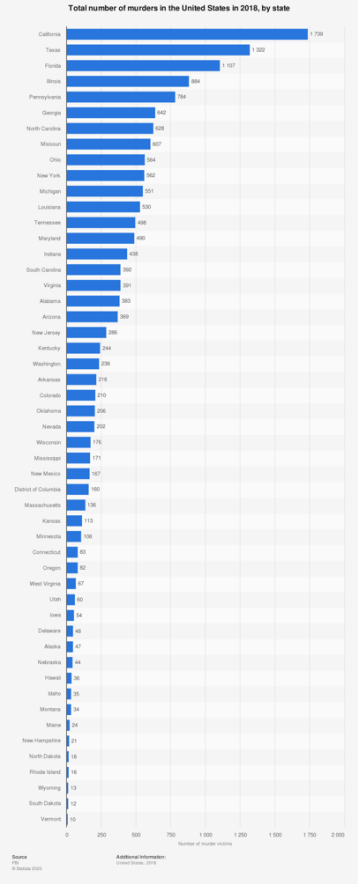
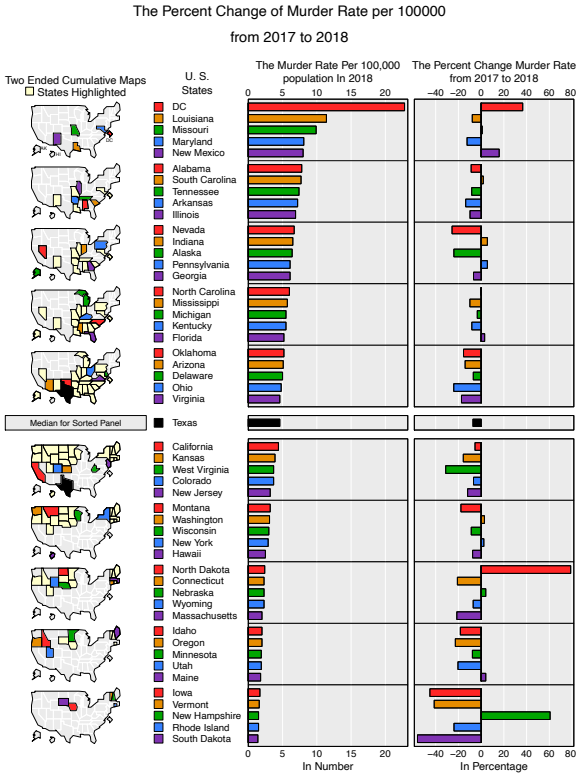


Fig. 3 Redesign Result for Bad Graph 2

4. Redesign Results and Analysis:

Compared with the original graph, the redesigned graph improves all problems mentioned in **section 1**.





Original **Redesigned Graph**

Fig. 4 Comparison between the Original Graph and the Redesigned One

Firstly, the new graph shows that D.C. is the most dangerous area because it has the highest murder rate. However, if we refer to the graph before, California is the most dangerous area because it has the highest number of murders, which will mislead readers.

Secondly, the new graph includes the map visualization. Therefore, readers can see how the pattern expands from one state to another in the geographical level. For example, the south

eastern area is more likely to have high murder rate. Besides, the boundaries can ease the reading when people want to focus only on the five highest dangerous states or vice versa. Also, different colors can help readers make comparison between one state to another more easily.

Thirdly, the redesigned graph visualizes the increment and the decrement percentage of murder in each state. As mentioned before, since the first graph only contains data of 2018, readers will not be able to see more information in the time dimension. For example, we can see that though New Hampshire almost has the lowest murder rate, its murder rate has great increase from 2017 to 2018. There must happened something unnormal in year 2018. Moreover, some states with great decrement of murder rate, like West Virginia, can provide other states with more meaningful experience.

Lastly, it is obvious that the redesigned graph looks much more attractive since it contains the map visualization, more colors, and more information.

Bibliography

Statista. (2019, September). *Total number of murders in the United States in 2018, by state*. Retrieved from Statista: https://www.statista.com/statistics/195331/number-of-murders-in-the-us-by-state/