DATA VISUALISATION 2 REPORT

**Glen Chen** 30572355 gche0026 – FIT3179 Wednesday 5-7pm Studio – Tutor **Joe Liu**

**URL:** <https://pathfinder4.github.io/FIT3179-Visualisation-2/>

**Words:** [ 966 ]

# Introduction and Background

The infographic produced for this visualisation assignment is focused around the topic of crime, particularly violent crime in the US. The data was chosen from a range of sources, including a compilation of national police jurisdiction and FBI reports. These datasets were chosen to display trends and changes over time and compare different states using a range of long term publicly available datasets.

The design of the graphic outlines a general overview to any non-expert in the field, with the overview statistics on the top half of the graphic being easily understandable without any annotations. In addition, more detailed comparisons are made, through highlighted states chosen to show both the expected violent crime rate of a state and bring a spotlight on certain states which perform as outliers. This is chosen to be used as reference material if required by enthusiasts or specialists wishing to identify relative crime rate between certain states.

The data processing used was fairly complex and involved significant use of the pandas python library to process the large data sets (up to 100s of MB) as can be seen in Appendix A. In addition a range of manual data cleaning methods on excel and python scripting were often used to merge different data sets together. This was necessary to keep the datasets small for quick processing by Vegalite. Additionally, it allows minimal transforming and layering on Vegalite when not required to prevent errors and ensure clean code. If interested, all unprocessed data is publicly listed on the Github, under “data/uprocessedData”.

In addition, all the Vegalite files and graphics are saved as JSON files to be accessed by the HTML scripts. All files are publicly available on the github and open links to be referenced.

# Design Choices

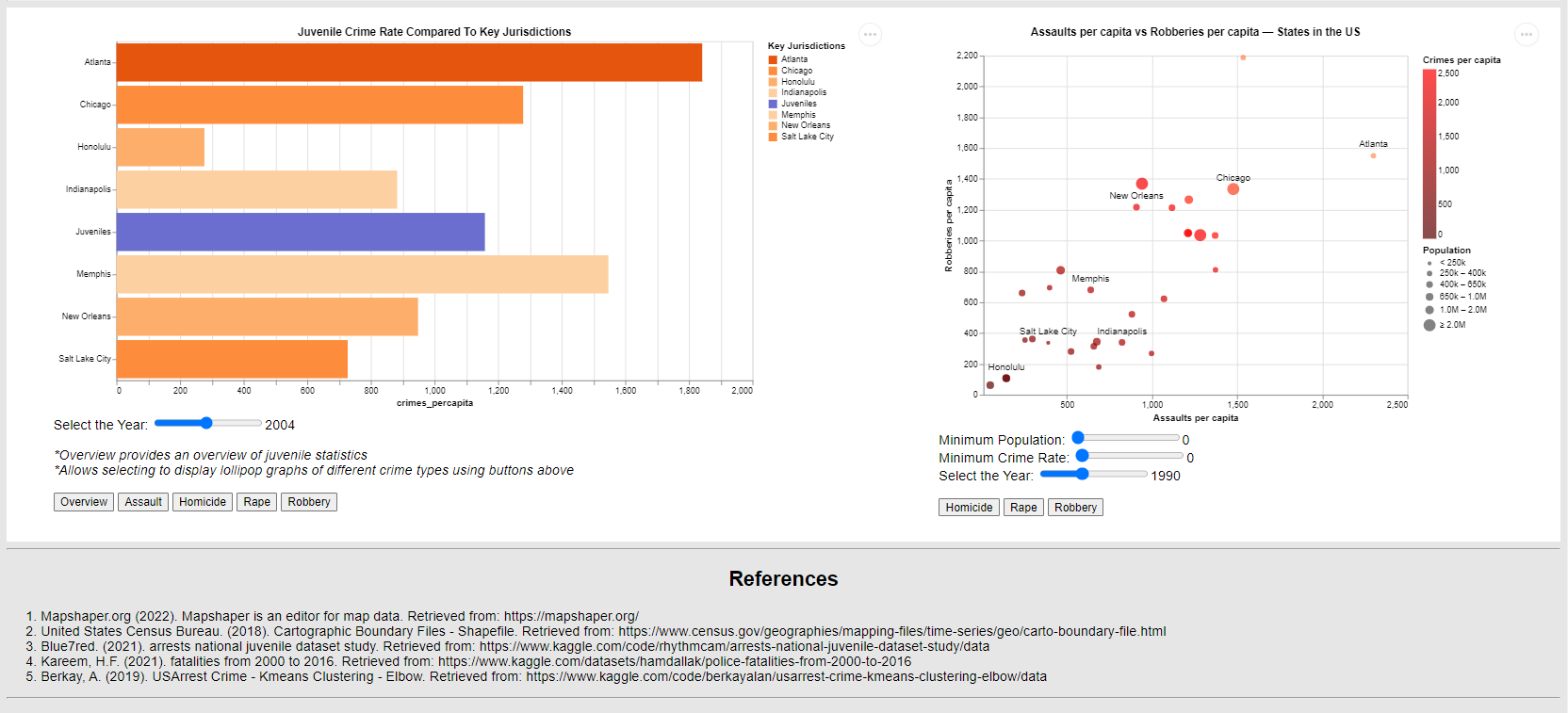
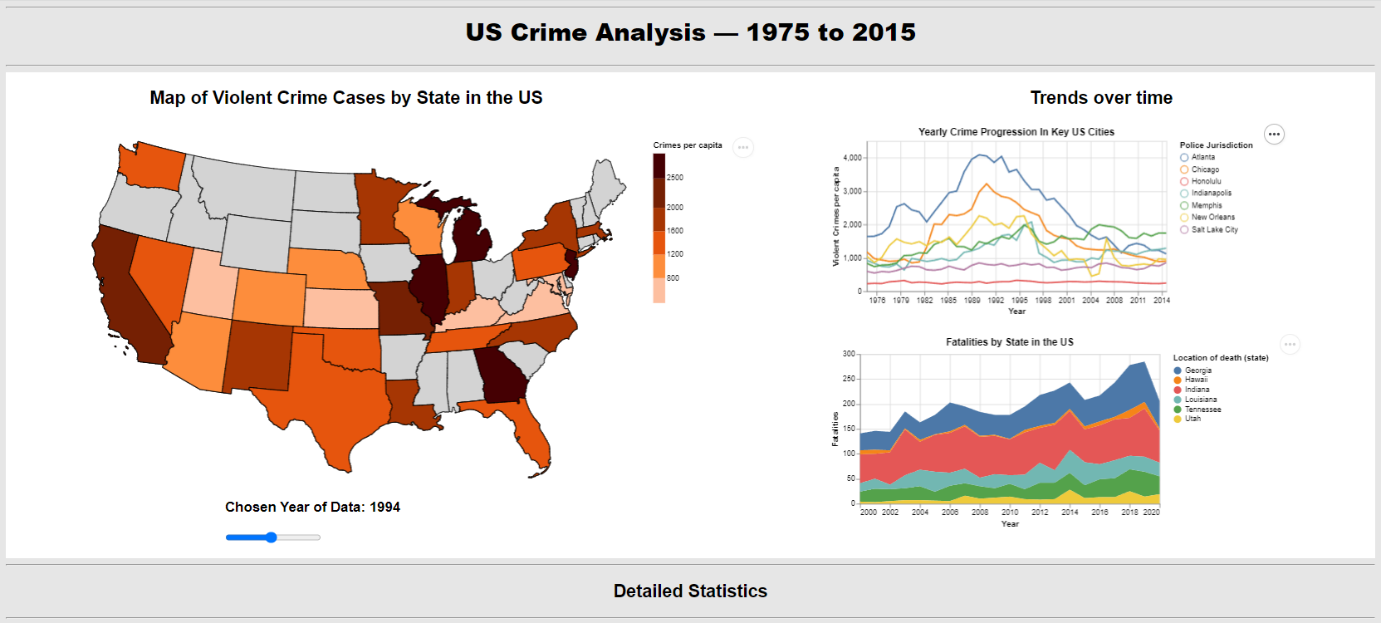


Figure 1 – Simple Overview

The visualisation was chosen to be split into two, with each having its own focus. The top half is designed to give a simple overview of violent crime over the analysed period, while the bottom is to give summarised details on crime types in the different states. This design choice resulted in two distinct rows broken apart by section headers, complemented by two columns both in line with each other across rows.

Each column was kept distinct and congruent with itself by maintaining unique colour palettes. As seen in the left column a largely orange palette was maintained, with a slight change in colour to identify they represent different marks (one categorical and one nominal). Column on the right uses a much more colourful palette, as can be seen in the first row using the same colour scheme to match together with corresponding city and state jurisdictions. The bottom right figure was chosen to be red to orange gradient to maintain the orange colour theme throughout the graphic, without being too colourful and confusing to interpret without any major categorical separation of the states.

The figures were originally to be placed with key jurisdiction graphics all on the right column. However, it was identified that line, area, lollipop and bubble charts all are highly analytically dense graphs and include significantly more white space. Therefore, these graphs can be kept on the right, to keep the map and bar charts, with a lot more colour and filled space to take the main focus of the graphic. This provides a consistent left to right transition on both rows, prevents the graphic from feeling too cluttered and messy. The left column was also made to be slightly wider than the right column to provide further emphasis and draw additional focus to the major figures.

It also ensures a clear story is told, by emphasising the clearer/simpler graphs to the audience for them to understand before moving to more complex graphs that may immediately confuse users. For example, the map on the left is first seen, where users immediately identify the domain of the graphic purely through the first figure. It can then be browsed over to get a general overview and identify patterns, before moving to the right of the map where yearly graphs refer to clear trends over time. Following this, the user can move down to the bottom row of graphs and spend time interacting with the more complex crime types to find their link to other crime types and different state demographics.

The typography used lays a clear hierarchy of the infographic, with a section break format implemented for section headings to even more clearly allocate visual area boundaries. Large headings were used to indicate figure groupings, ie. column grouping headers. Figure titles were used when no column headers were necessary due to only having one figure, making sure we maintain symmetry. The exception was the map figure heading being in the format of a column grouping, however this maintains symmetry and was very beneficial due to this drawing extra focus as the main point of the infographic or the start of the story. Additionally the chart//graph labels, legends and axes were kept in the same format throughout for consistency and to keep all three graphs at a similar importance level.

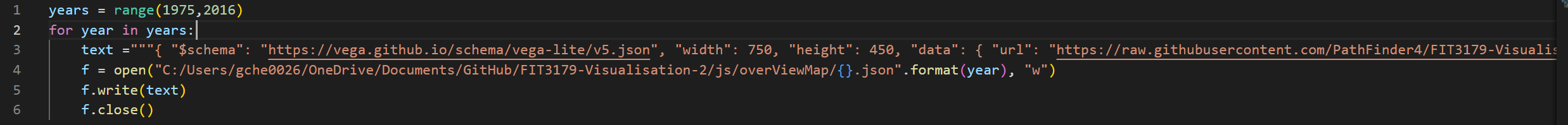
# References

1. Mapshaper.org (2022). Mapshaper is an editor for map data. Retrieved from: <https://mapshaper.org/>
2. United States Census Bureau. (2018). Cartographic Boundary Files - Shapefile. Retrieved from: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>
3. Blue7red. (2021). arrests national juvenile dataset study. Retrieved from: <https://www.kaggle.com/code/rhythmcam/arrests-national-juvenile-dataset-study/data>
4. Kareem, H.F. (2021). fatalities from 2000 to 2016. Retrieved from: <https://www.kaggle.com/datasets/hamdallak/police-fatalities-from-2000-to-2016>
5. Berkay, A. (2019). USArrest Crime - Kmeans Clustering - Elbow. Retrieved from: <https://www.kaggle.com/code/berkayalan/usarrest-crime-kmeans-clustering-elbow/data>

# Appendix A – Data Processing Scripts

### Overview map and yearly line graph data processing





### Fatalities statistics processing



### Juvenile crime statistics

### 

# Appendix B – Github Files and Documentation Links

<https://github.com/PathFinder4/FIT3179-Visualisation-2>

# Appendix C – 5 Design Sheets

### 