

# **Police Killings in the United States**

**IAT 814 Project Report**

**Simon Fraser University**

**By: Saptarshi Dutta Gupta (301400590)  
Sameer Pasha (301406535)**

## I. Overview

According to the Guardian, the US police kill more in days than any other country in years. In 2019, 1,004 people were shot and killed by police according to *The Washington Post*, whereas the "Mapping Police Violence" project counted 1,098 killed <sup>[1]</sup>. Police killings are violent in nature which disrupt the normal functioning of the society and have considerable effects on mental health, neighbourhood peace, life and politics. The death of George Floyd, Sam Dubose among others have led to violent protests in the past that have attracted substantial attention to the racialized character of police violence against the civilians. <sup>[2]</sup>

Police killings have been associated with Ethnicity solely in the past. Our motivation for the project is to create a interactive visualization dashboard that will enable law enforcement officials to delve deeper into this subject and explore the different factors like Age, Gender, Signs of Mental Illness, Location apart from Ethnicity that lead to the killings and how these different factors relate to each other.

## II. Introduction

According to statistics, 1 in every 1000 black men can expect to be killed by the police. On an average, the probability of being killed by police is about 1 in 2,000 for men and 1 in 33,000 in women. The age group between 20-35 years are at the highest risk. For young men of color, police use of force is among the leading causes of death. <sup>[2]</sup>

It is not mandatory for the police departments to report when an officer kills a civilian. Therefore the data collected by the FBI is not reliable <sup>[3]</sup>. Newspapers like Washington Post and the Guardian keep an active record of the police killings taking place everyday across every province in the United States along with Personal Identifiable Information (PII) for every victim. We have used these publicly available datasets for our visualization. Our dashboard will enable reconfirming the statistics, bias and hypothesis related to police killings described above and create an interactive data story about the fatal police shootings that will help the law enforcement officials investigate how the different factors connect to each other and help in formulating relevant and more effective laws to handle the situations better and improve the overall operations.

We have used different visualization charts like line charts, choropleth maps, bar charts and parallel coordinate plots to visualize the effect of the several factors like Age, Gender, Threat Level on the police killings and how factors like Mental Illness correlate with factors like possession of weapons or threat level. We have kept in mind the principles of effective data visualization during the design of the graphs and selection of the colors to identify key insights easily.

### III. Questions we are trying to answer

Police killings have been majorly associated with ethnicity in the past and in recent times, voices have been raised against police brutality in the form of protests such as Black Lives Matter. Our Goal is to find out if there are factors other than ethnicity that impact these police killings.

We are trying to answer the following questions:

- How are the police killings spread across various states?
- Is there an impact created by the purchase of weapons such as guns?
- Is there any particular gender/age group that is more prone to these killings?
- Is mental health an important factor while determining these killings?

We aim to answer the above questions by building an interactive dashboard incorporating the visual principles of design. This dashboard will serve the law officials as a tool to identify potential causes and police brutalities and potentially create awareness programs and educate officials to reduce such killings from happening in the future.

### IV. Dataset used

#### 1. Washington Post Police Shootings Data <sup>[4]</sup>

The Washington Post Police Shootings Data contains data for every fatal police shooting in the United States from January 1, 2015. The daily started tracking details from each of the killings including the race, age, gender of the deceased as well as the circumstances in which the shooting takes place example whether the victim was fleeing, the type of weapon they possessed (if any), if they had any signs of mental illness, the threat level of the victim etc. The main columns in the database that were being used in the analysis is described below:

Name of the column	Description	Type of the variable
name	Name of the victim	String; nominal variable
date	The date of the shooting in yyyy-mm-dd format	String; ordinal variable
manner_of_death	Indicates how the victim died. Either <b>shot</b> or <b>shot and tasered</b>	String; Categorical Variable
armed	indicates that the victim was armed with some sort of weapon that a police officer believed could inflict harm. Either <b>undetermined</b> , <b>unknown</b> or <b>unarmed</b>	String; Categorical Variable
age	The age of the victim	Integer; Quantitative Variable

gender	The gender of the victim	String; Categorical Variable
race	The race of the victim	String; Categorical Variable
city	The municipality where the fatal shooting took place	String; Categorical Variable
state	Two letter postal code abbreviation	String; Categorical Variable
Signs of mental illness	Indicator if the victim had a history of mental health issues	String; Categorical Variable
threat_level	Indicator of the victims threat level	String; Categorical Variable
flee	Indicator that shows how the victim was moving away from the officers	String; Categorical Variable

**Raw Data dimensions:** the original dataset had 6139 rows and 17 columns but we ended up using 12 columns as mentioned above.

## 2. Gun Killings Database [5]

The data comes from the FBI's National Instant Criminal Background Check System. The NICS is used to determine whether a prospective buyer is eligible to buy firearms or explosives. Gun shops call into this system to ensure that each customer does not have a criminal record or isn't otherwise ineligible to make a purchase. The data has been supplemented with state level data from [census.gov](#).

Name of the column	Description	Type of the variable
month	The date containing the year and month information of the purchase	String
totals	Total number of weapons purchased for a given month	Integer

**Raw data dimensions:** The original dataset had 12485 rows and 27 columns, we ended up using only two of the columns. The month column was used to fetch details pertaining to a particular year and month. Whereas, the totals column had details about all different types of guns purchased for that given month of a year.

## V. Data Transformation and Aggregation

### a) Washington Post Data

**Transformation:** In the washington post dataset, we had to convert the **gender**, **race** and **state** from the abbreviated form to a unified form. For instance, in the gender column, M was changed to male and F was changed to Female. In the race column, A/Asian/Pacific islander all mean Asian. Moreover, the age was split into age bins of 0-25, 25-35, 35-45, 45-55, 55-65, 65-75 and 75-85.

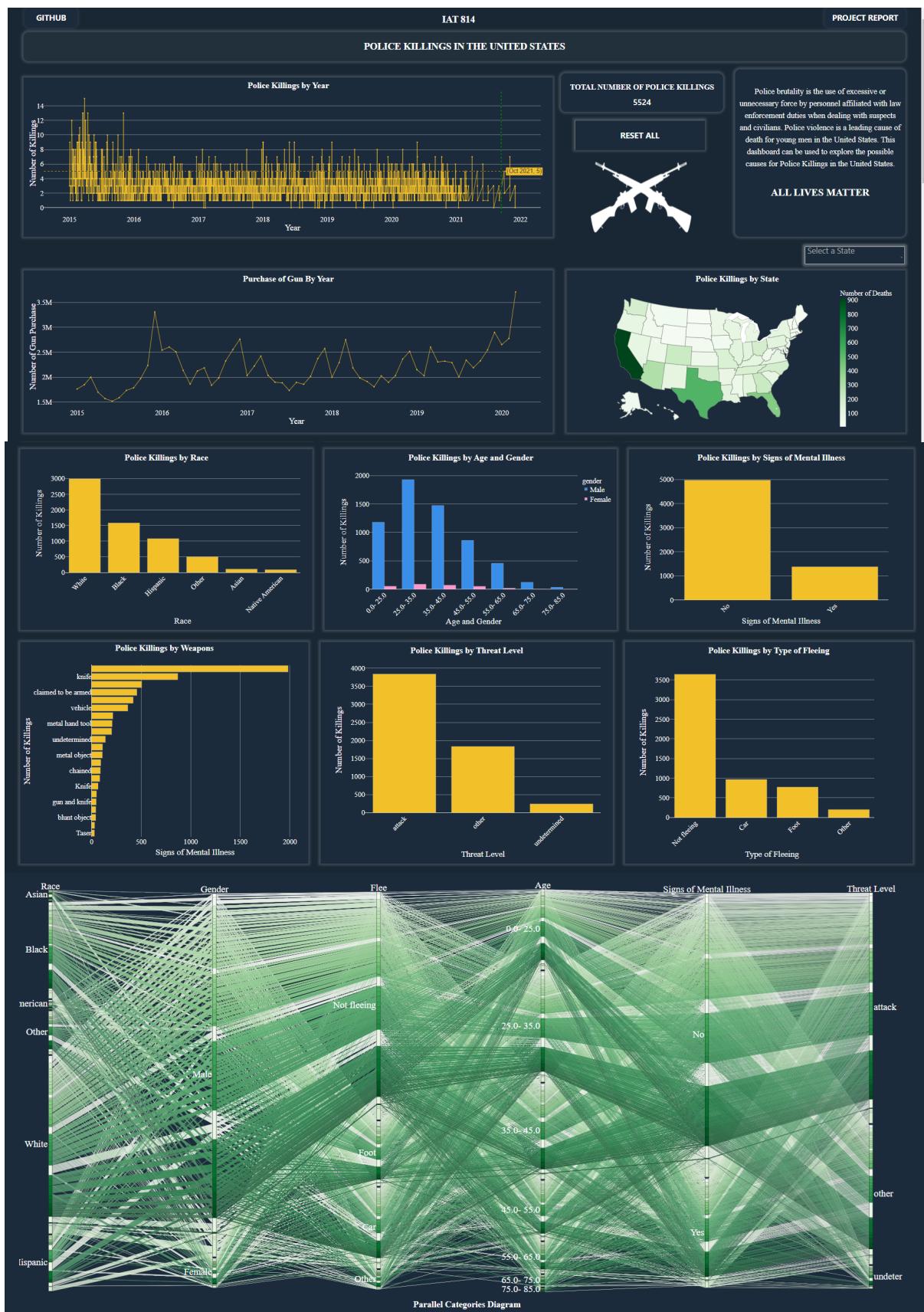
**Aggregation:** In order to obtain killings pertaining to a particular category, selections can be made by clicking(for bar charts and choropleth maps), making a bounding box(for line graphs) or selecting from a dropdown(for choropleth maps) . On making a selection, the entire dataset is filtered to have values pertaining to the selected attribute. For instance, if we select Asians, the dataset is filtered by the selection to only include rows corresponding to Asians. Once the dataset is filtered, it gets aggregated based on race, age, gender, signs of mental illness, weapons possessed, threat level and type of fleeing. For consecutive selections, the data frame is further filtered based on the multi-level selections and again aggregated with the same attributes. Moreover, for the killings by weapons data, in case there are less than 25 records for any weapon, it gets grouped into a column called others.

### b) FBI Gun Purchase data

**Transformation:** The month column contains both the year and month information. This was transformed to a date type string in the format (YYYY-MM-DD) to match the first dataset.

## VI. Visualization Design

Visualization dashboards help to easily visualize the data, filter on demand and helps in better decision making. Therefore, we have designed our app in the form of a visualization dashboard using Plotly Dash that enables quick analysis, helps identify patterns and makes the overall story easy to understand. All the components except the Parallel categories plot are interactive in our dashboard and enable the user to see the change in the police killings on drilling down a particular factor. Following is an overview of our Police Killings dashboard:



**Fig 1. Overview of dashboard**

The various components used in the dashboard are:

a) Line charts

- i) In order to visualize the Police Killings with respect to time, a line chart has been used to show the data from 2015-2021. Since we have a datetime variable and quantitative attribute, a line chart can be used to visualize the trends easily. On hovering, the month and year along with the count of the number of deaths is shown.
- ii) Similarly for visualizing the purchase of guns with respect to the years 2015-2021 a line chart has been used.

Note that both the line charts have been encoded with an yellow shade and points are being used in addition to lines for better exploration of the data when the user hovers over the points.

b) Indicator figure for the total number of killings

To indicate the total number of police killings according to the user selection, an indicator figure has been used to display the count of the deaths which gets updated on selection of the other graphs.

c) Choropleth map

Washington posts police killings data contained the state in which the killings have taken place. Therefore, for visualizing state wise statistics a choropleth map encoded using color saturation representing the number of deaths has been used. The initials of the State and the count of the deaths can be seen on hovering over the respective States.

- i) Multiselect dropdown list: We also have a dropdown list for the various states in case the user wants to visualize the deaths for more than one state at the same time.

d) Bar graphs

In order to visualize the Police Killings for factors such as Race, Gender, Age, Mental Illness, Weapons, Threat Level and Types of Fleeing we are using Bar charts. Since for each of these factors there is one categorical attribute and one numerical attribute bar charts can best visualize the killings.

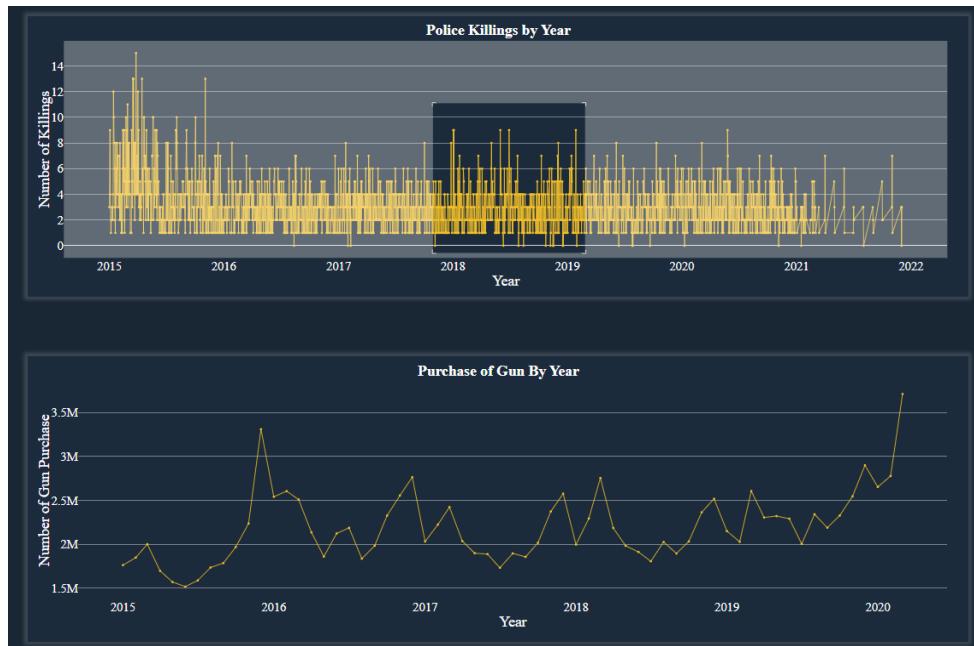
- i) Police Killings by Race: Bar chart showing the police killings for the various categories of race including White, Black, Hispanic, Asian, Native American and Others. The bar charts have been encoded using the similar yellow color as the line charts.
- ii) Police Killings by Age and Gender: A grouped bar chart that visualizes the Police Killings with respect to both categories Age and Gender. In order to differentiate between the two gender categories, blue and pink color encoding has been used.
- iii) Police Killings by Signs of Mental Illness: Bar chart showing the number of police killings based on whether the victim has signs of mental illness(Yes/No).
- iv) Police Killings by Weapons: Horizontal bar chart that shows how the kind of weapons that the victim possessed affect killings.

- v) Police Killings by Threat Level: Bar chart showing the number of police killings with respect to the threat level of the victim (categories: attack/others/undetermined)
  - vi) Police Killings by Type of Fleeing: Bar chart for showing the number of police killings with respect to the victim state of fleeing (categories: Not fleeing/Car/Foot/Other)
- e) **Parallel categories plot** [5]
- Since we have a multidimensional categorical dataset, a parallel categories plot can effectively visualize patterns from the dataset quickly. Each variable is represented using a rectangle and the height of the rectangles reflect the relative frequency of occurrence of the corresponding values. Therefore we have shown how the categories like Race, Gender, Age, Mental Illness, Threat Level and Type of Fleeing are related to each other using the parallel categories plot. Further we have used the same color encoding as that of the choropleth map that reflects the state in which the particular killing took place.

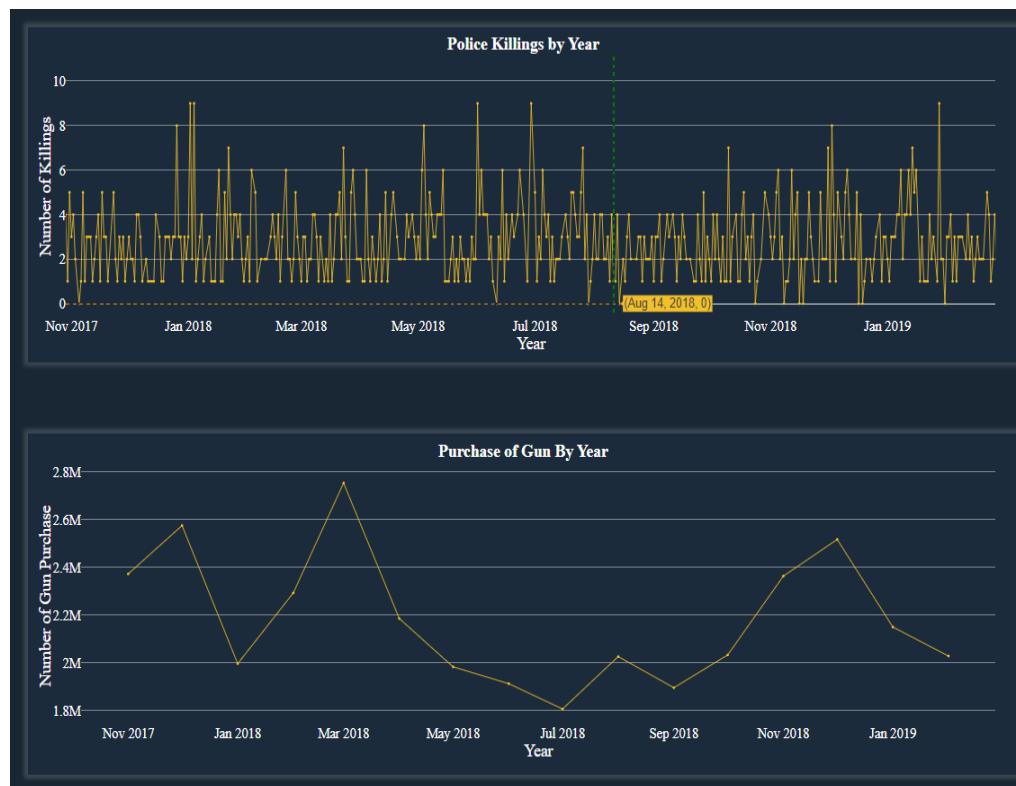
## VII. Usage of Interactions

Our dashboard makes extensive use of brushing and filtering to see the change in police killings for all other factors when a particular factor is highlighted.

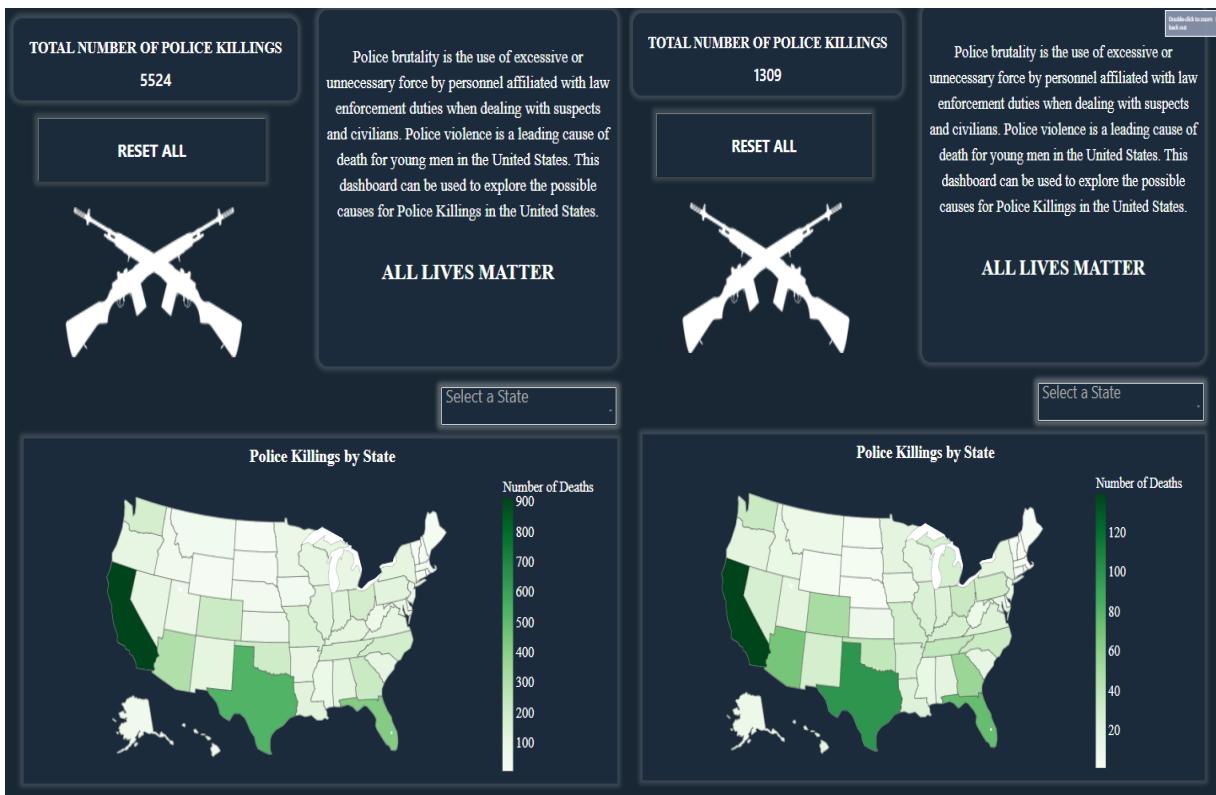
- a) **Line chart**  
We have made use of a time scrubber operation in both the line charts i.e. *Police Killings By Year* and *Purchase of Guns by Year*. We can use the mouse to make a bounding box around the years for which we want to visualize the data. According to the selection made, all other graphs including the indicator figure, bar graphs, choropleth map and parallel categories diagram will change. The Purchase of Guns By Year will also be shown only for the time period selected on the Police Killings By Year line chart. A similar bounding box can be made for the Purchase of Guns by Year line chart which will change the timeline for the Police Killings by Year line chart. Figure 2, 3 and 4 explains the interactions:



**Fig 2. Filter by year by selecting a bounding box in the line chart**

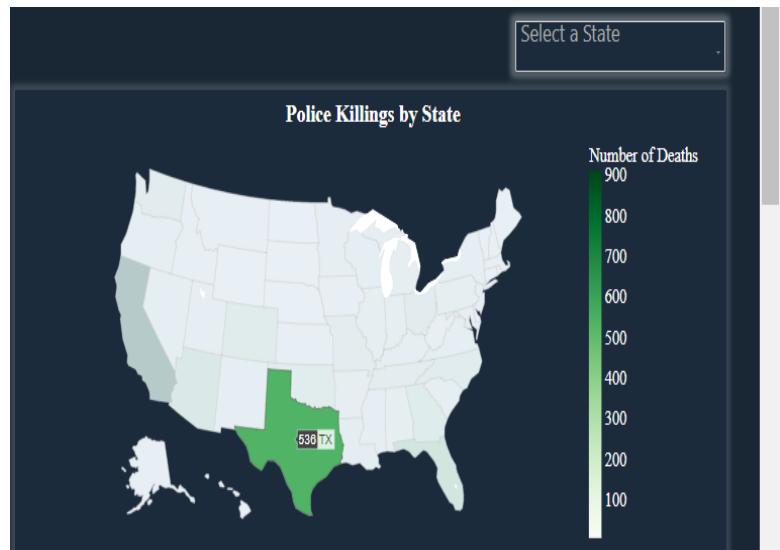


**Fig 3. Change in the Line charts after the bounding box is made**



**Fig 4. Change in the indicator figure before(left) and after(right) the bounding box selection**

- b) **Choropleth map:** There are two types of interactions for the choropleth map. The user can either select a single state by clicking on a particular state or if the user wants to select a region,a multiselect dropdown menu (with autocomplete feature) can be used to select more than one state at the same time. Based on the selection, the bar charts, line charts, indicator diagram and parallel categories diagram will be updated. In the figure on the right (Fig 5), Texas has been selected.



**Fig 5. Filter by State by Clicking on a State**

Based on the selection, all the other graphs are updated. The updated bar chart on selecting Texas is shown below (Fig 6):

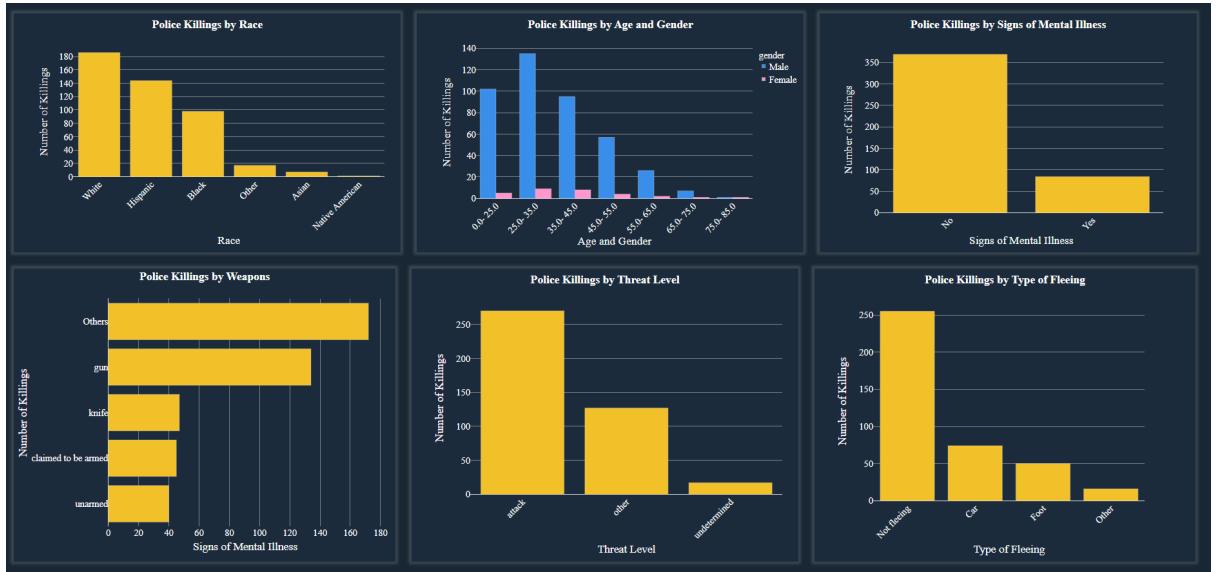


Fig 6. Updated bar chart on filtering by state (Texas) in Fig 5

Similarly, if the user wants to see only the Police Killings for the west coast state, California, Oregon and Washington can be selected from the drop down menu and the corresponding states will be highlighted in the choropleth map. As in the earlier cases, the bar charts, indicator figure, line charts and parallel coordinate figures will be updated.

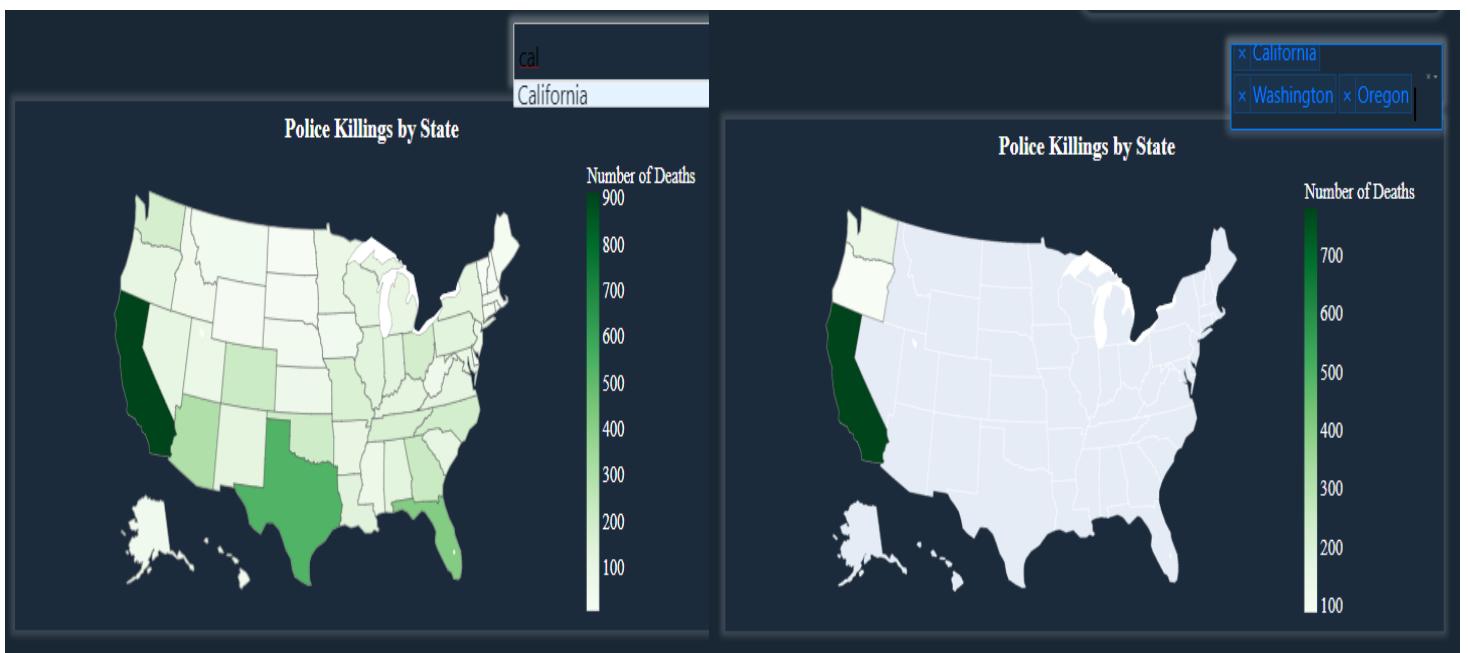
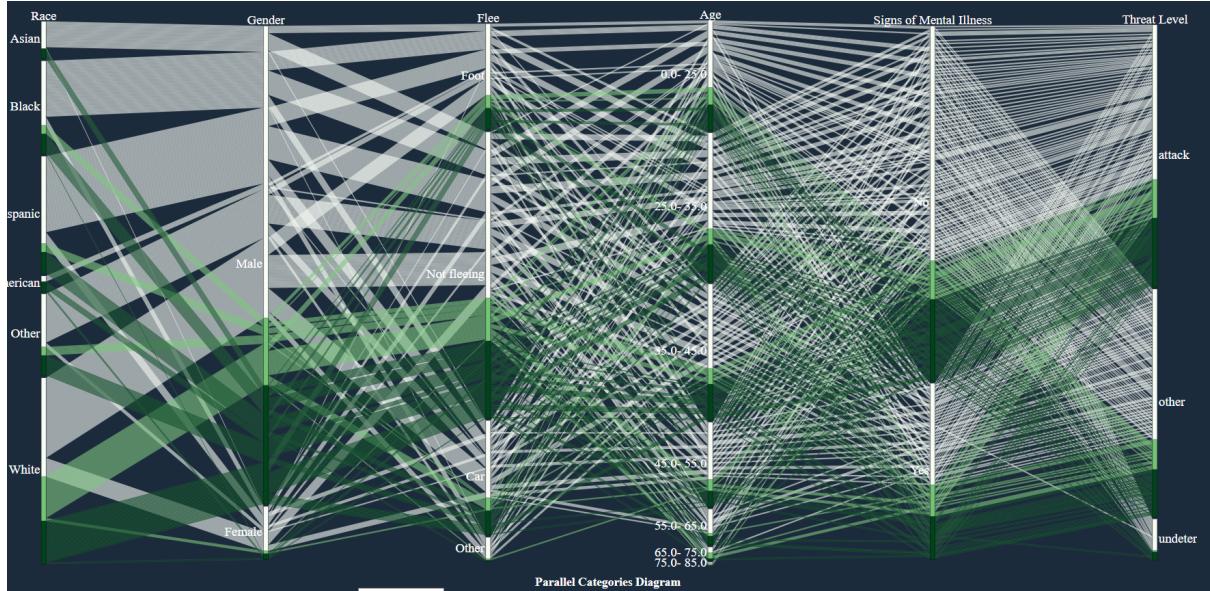


Fig 7. (left) Selection of the states from the dropdown menu (right) The choropleth map after the selections are made

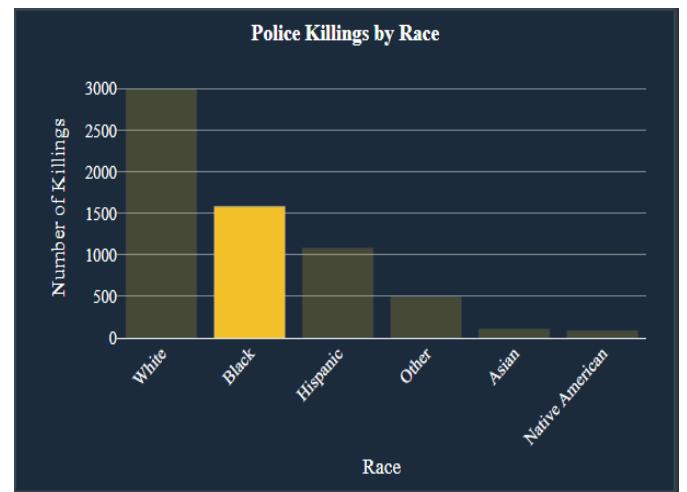
On selecting the states from the dropdown menu, the parallel categories diagram gets updated with the same color coding as that of the selected states. Therefore, the parallel categories visualizes state wise information along with Race. Gender, Age, Signs of Mental Illness, Flee and Threat Level.



**Fig 8. Updated parallel categories diagram on selecting the provinces in Fig 7**

- c) **Bar Charts:** For each of the bar charts we can select a particular bar to drill down into a particular category. On selecting a particular bar, the selection gets highlighted while the opacity of the other bars reduces. Following are the interactions available for each of the bar charts:

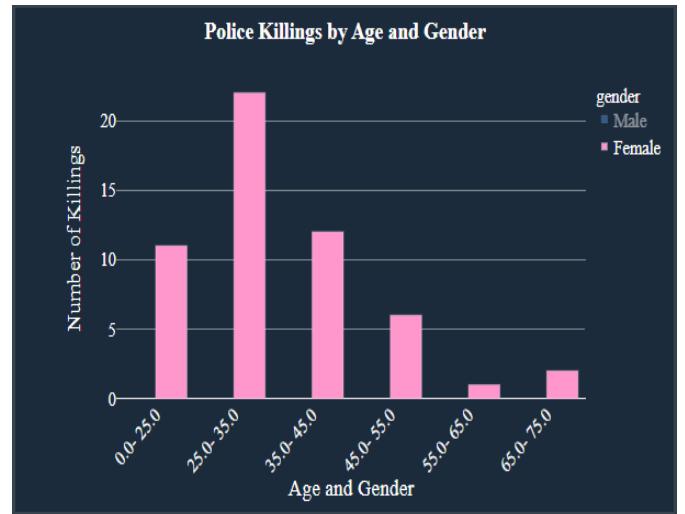
- Filter by Race:** On clicking on a particular race (for eg. Black) the selection is highlighted and the other graphs are updated (Fig 9)
- Filter by Age and Gender:** We can select a particular gender on the legend to visualize the Killings only for one particular gender (Fig 10). The user can also make a selection from the grouped bar chart.



**Fig 9. Selecting a particular race by clicking on the bar**

- Filter by Signs of Mental Illness**
- Filter by Type of Weapons**

- v) [Filter by Threat Level](#)  
vi) [Filter by Type of Fleeing](#)



**Fig 10. Filter by Gender on clicking on the Male tab in the Legend**

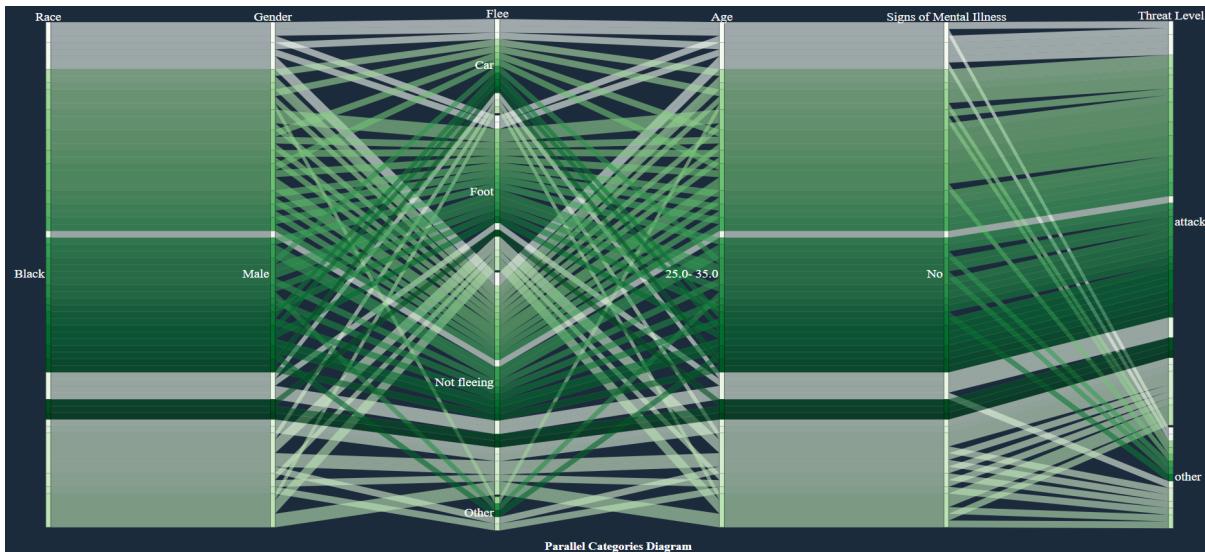
The user can further filter on multiple factors like Race=Black, Age and Gender=Male (25-35) and Signs of Mental Illness=No and Weapon=Gun. Based on the selection the other graphs are updated. Below is an example of the scenario (Fig 11 and 12)



**Fig 11. Filter on multiple bar charts. Selections made: Race=Black, Age Gender=Male&25-35, Signs of Mental Illness=No, Weapons=gun**



**Figs 12 (above and below). Change in the rest of the graphs on selection in Fig 11**



## VIII. Design Challenges

Although the dashboard supports querying a variety of factors simultaneously to see how these factors could affect killings, there are a few interactions that aren't supported in the current iteration of the dashboard. The ability to make multiple selections on any graph or map is currently not supported by plotly which is why it is not possible to compare different entities belonging to the same category. Moreover, the selections are not highlighted in the parallel coordinates graph and the size of the hover label could not be increased as it is a limitation of the library and there is an open issue [here](#).

## **IX. Future Work**

In the future, we plan on adding detailed level wise statistics for province, state and city. We'd also like to find how factors such as education level, marital status, employment history and economic status of the victim affect the police killings. We'd also like to support a region level view of the states, for instance, let's say that we'd want to see the number of police killings in the west coast. Currently this is not supported directly but the user can manually select all the states that belong to the west coast in order to select multiple states. However, we can group multiple states based on the region and have that as an option in order to query based on the region type, ie. west coast, east coast etc. In order to make multiple selections for any given category, we can try to add a bounding box that can make multiple selections and show results based on these multi-selections.

## **X. Conclusion**

We have utilized the principles of visualization design to build an interactive dashboard that can help identify the factors that impact police killings. This dashboard lets the user explore and find how factors such as age, gender, location, signs of mental illness, possession of weapons impact these killings. The dashboard can be used by law enforcement officers and policy makers so they can create awareness programmes and organize sessions to educate police officers to potentially avoid such incidents in the future.

## **XI. References**

- [1] [Use of deadly police force in US](#)
- [2] [Risk of being killed by the police](#)
- [3] [How can we enhance police accountability](#)
- [4] [Washington Post data](#)
- [5] [FBI Gun data](#)
- [6] [Parallel Categories Diagram](#)
- [7] [Plotly documentation](#)
- [8] [London Accidents dashboard](#)
- [9] [Plotly Community Forum](#)
- [10] [Dashboard example](#)