Individual Coursework Submission Form

Specialist Masters Programme

Surname: Chawla	**First Name: Sarthak**
MSc in: Business Analytics	**Student ID number: 220042213**
Module Code: SMM635	
Module Title: Data Visualisation	
Lecturer: Simone Santoni	**Submission Date: 6th December,2022**

Declaration:

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

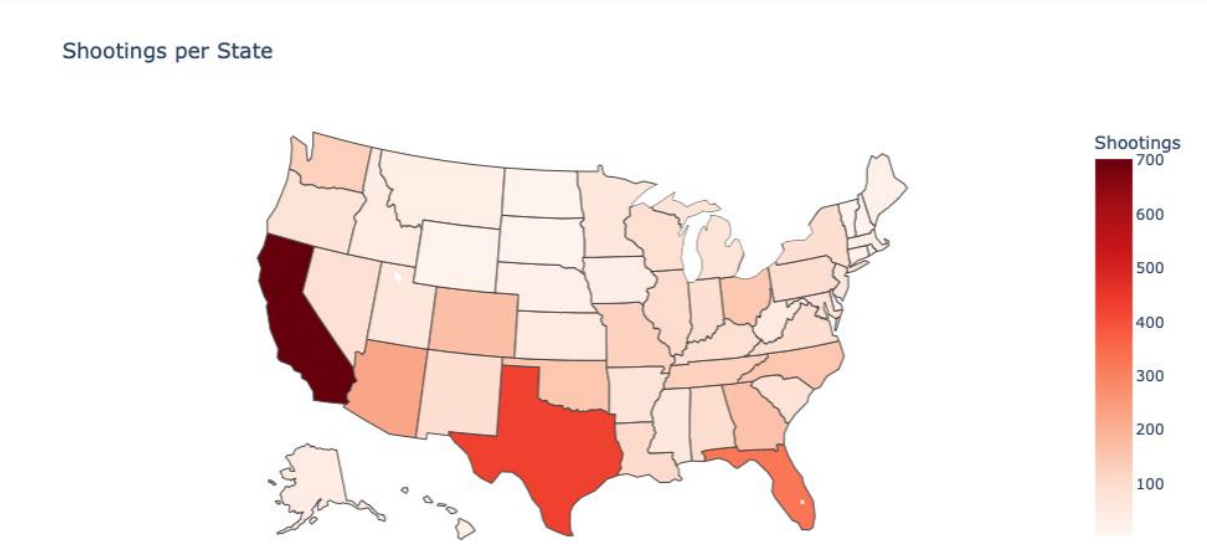
Deduction for Late Submission: | **Final Mark:** **%**

SMM635
Data Visualisation

Final course project submission

Visualisation #1

Which states are shooting hotspots?



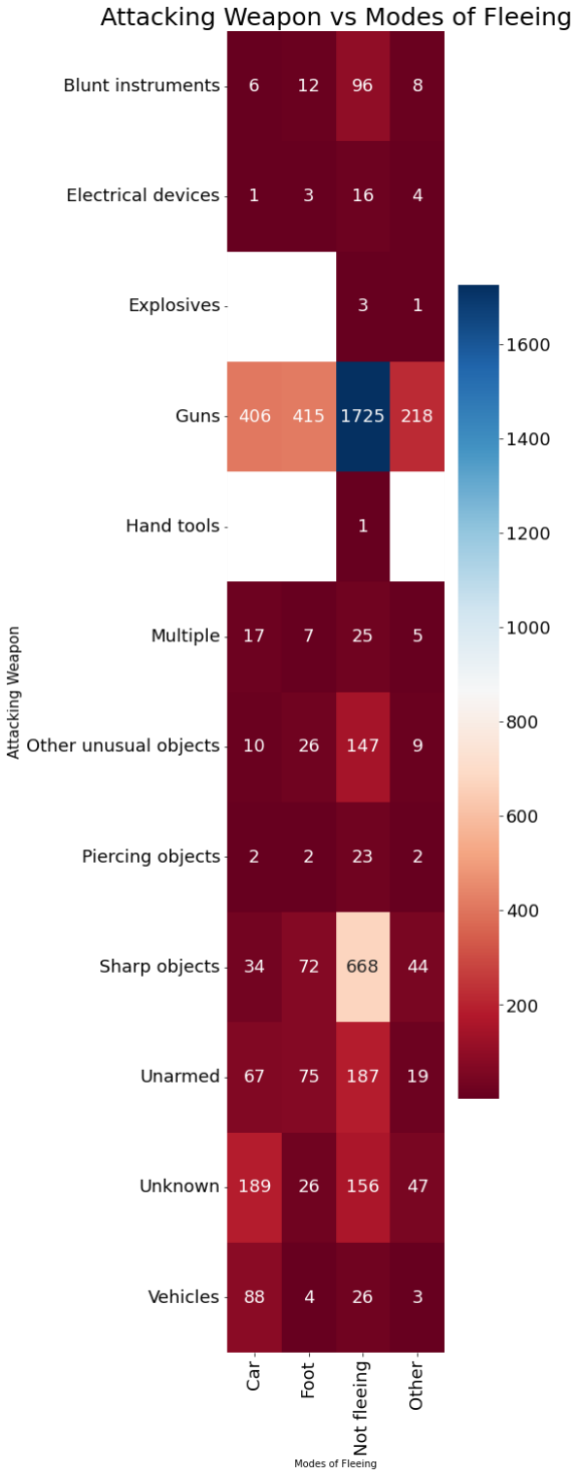
A choropleth plot was used to represent the various states that were determined to be shooting hotspots in the United States. The legend was used to depict the range of counts and their relationship to colour. Each state block shows the count of shootings in that state. Spectrum colour coding has been implemented.

Instead of being soiled into separate bars, which made it difficult for users/clients to understand, choropleth was one of the most appropriate visualisations for mapping the various shooting hotspots in the United States. The chart is also useful in that it can show the various zones of the country that were heavily affected, such as the west, south, and south-east, which would not have been possible with any other graph. The colour scheme is designed in such a way that the more shootings in a specific region, the denser the colour block. Spectral Colour was designed with readability in mind, as shades of red are among the easiest to identify and correlate.

The main takeaway from the chart is that California, Texas, and Florida have the highest number of shootouts in the country, with California having the most.

Visualisation #2

How did the victims try to flee during the shootout, and what weapon were they carrying?



A heat-map has been created to represent the composition of victims who attempted to flee during the shootout and their respective mode of escape, as well as the weapon they were carrying. The two axes have been marked with “Modes of fleeing” and “Attacking Weapon” respectively. The legend has been used to depict the spectrum of counts and their relation to colour. Each cell represents the number of victims who attempted to flee during the shootout using that specific mode of escape and the weapon they were carrying.

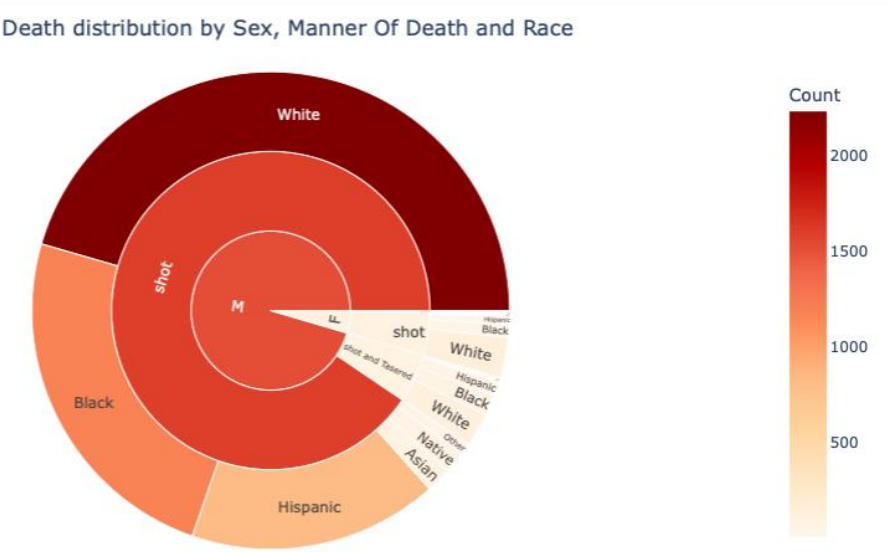
Instead of being soiled into separate bars, which makes it difficult to understand for users/clients, a heat-map was one of the most appropriate visualisations for mapping this type of distribution. The denser the colour block, the more victims there are in a specific block. Furthermore, the count has been assigned to each cell for a better understanding of the data's statistical inference.

The main insights are as follows:

1. Majority of victims were carrying a gun and did not attempt to flee.
2. Sharp objects were the second most carried weapons by victims, and they did not attempt to flee either.
3. The most common mode of escape was by car, with majority of them carrying guns.
4. The majority of victims who carried a gun attempted to flee on foot.

Visualisation #3

What is the gender and race death distribution of victims, and how did they die?



The Sunburst plot was designed to depict the pattern of death distribution of victims across two sexes and their respective races, considering how they died. The chart is divided into three sections based on gender, which are then extended to another circle that explains the manner of death, and finally to the outer circle, which represents the race of each victim who was shot. The colour scheme is designed in such a way that the denser the colour block is, the higher the count for a specific region.

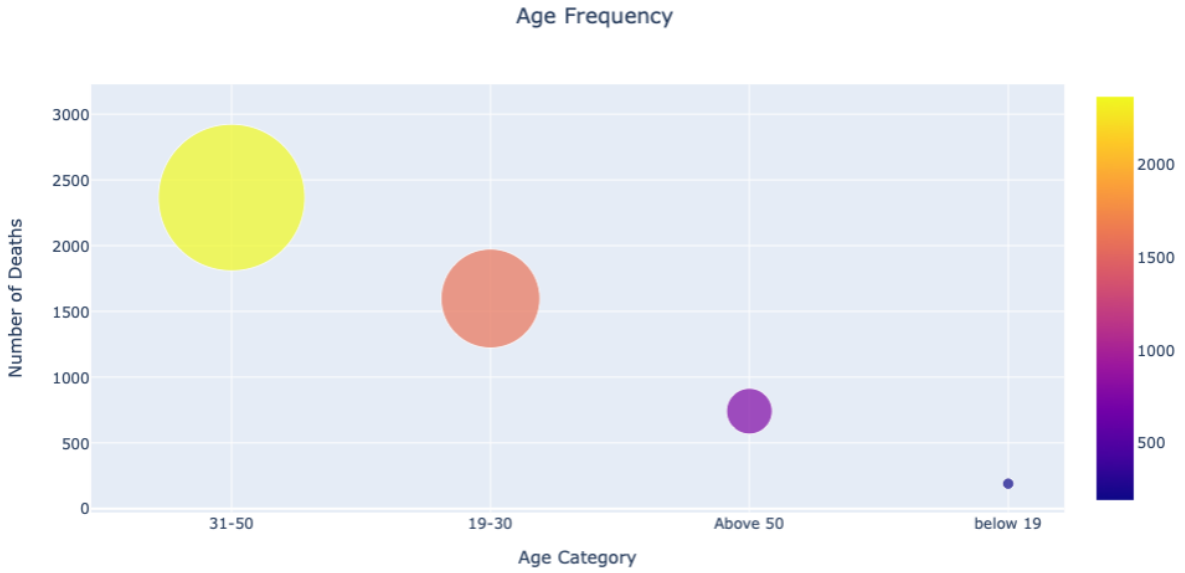
The Sunburst Plot depicts hierarchy as a series of concentric rings, with each ring representing a level in the hierarchy. Focusing on a ring segment provides a sense of the segment's part-to-whole relationship with respect to its parent ring segment.

The main insights we collect from the visualisation are as follows:

1. Males had a much higher death rate than females among the total population of victims.
2. The fact that none of the female victims were tasered and all were shot to death was an interesting takeaway.
3. White men and women were shot the most. This is explained by the fact that the majority of the victims were also white.
4. Hispanic women were the ones who were killed the least of all victim population.
5. There were no female native, Asian, or other victims among those killed in the shootings.

Visualisation #4

How old were the majority of the victims?



To depict the composition of deaths among victims of various ages, a Scatter Plot with X-axis = "Age Category" and Y-axis = "Number of Deaths" was created. The graph is composed of four bubbles, each representing a different age group. Each age group has been assigned a colour code. The larger the bubble, the more deaths occurred in that age group.

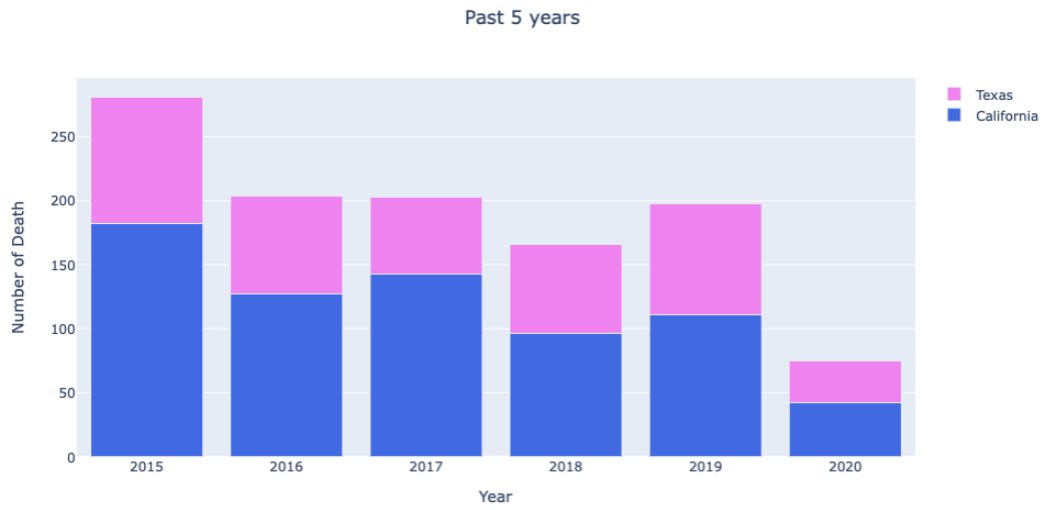
Scatter Plot was used because it reports the value of individual data points as well as the pattern among the complete data. The scatter plot allowed for the representation of victim deaths within and across age groups in a single plot, simplifying the process of analysis and user interaction.

The main insights we collect from the visualisation are as follows:

1. Most victims were between the ages of 31 and 50.
2. People under the age of 19 died the least, which is understandable given their small population size in the dataset.
3. People over the age of 50 made up a relatively small proportion of the victims.

Visualisation #5

What are the statistics on shootings in the two hotspots over the last five years?



A stacked bar chart has been created, with the X-axis representing "Years" and the Y-axis representing "Number of Deaths." California and Texas are the two most popular shooting locations. Legend has been defined to categorise both the states with different colours. The plot's height has been adjusted to improve the visual accuracy of the plot.

The stacked bar chart is extremely useful for representing a large amount of information in a small space, making it easy for users/clients to understand. It is also effective at visualizing multiple data series (categorical) on the same axis and stacked into parent categories.

The main insights we collect from the visualisation are as follows:

1. When compared to 2015, the number of shootouts in both cities falls dramatically in 2020.
2. In both cities, shootouts increased dramatically in 2019 when compared to the previous year.
3. In comparison to Texas, California has a higher composition of shootouts over the years.
4. Despite the fact that there was not much difference in the number of shootouts between 2016 and 2017, the number of shootouts in California increased dramatically in 2017, while those in Texas decreased.