Data Visualisation

Visualising Crime Data for St Louis City, Missouri, USA

Raghav Mehta
School of Computing (MSc in Data Analytics)
National College of Ireland,
Dublin, Ireland
x17155151@student.ncirl.ie

1 Introduction

This project uses the crime data of St. Louis City, Missouri, USA (https://www.stlouis-mo.gov/data/crime.cfm). The data downloaded is from the period January 2017 to June 2019 and every row pertains to a unique crime incident reported to the police. Two additional datasets which have been used are the neighborhood names for the 88 neighborhoods in St. Louis City and data from FBI's Uniform Crime Reporting (UCR) program. The UCR program helps to categorize the data in 2 parts – violent & property related crimes (Part 1) and less serious crimes (Part 2). In total there are 26 sub-categories in total (8 for Part 1 and 18 for Part 2). Population of each of the 88 neighborhoods has also been scrapped from Wikipedia. Hate crime statistics are also downloaded from the same website as the crime data. These hate crime statistics are coerced into one file through R studio.

The crime data (20 individual CSV files month-wise from January 2017 to June 2019), neighborhood data, population data and the UCR data have been merged and cleansed in R studio. Month and Year have been extracted from the '*CodedMonth'* column. A field has been generated called *'Crime per capita'* by simply dividing each offence count by the population of the neighborhood. Summing the values of this field for each neighborhood for any given year gives us the crime per capita of that neighborhood for the given year.

Additional preprocessing includes converting the crime location coordinates from XY format to latitude longitude. This has also been performed in R studio. The XY coordinates given use the State Plane North American Datum 1983 (NAD83) format which is not readable by Tableau. These coordinates are given in US feet and are first converted to meters. Subsequently using the R studio package *'rgdal'* , the coordinates are converted to the WGS 84 geocoding format. St Louis falls under ESRI code 102296, which is essential to reverse the mapping projections from NAD83 to WGS 84 format.

All visualizations are created in Tableau 2018.1, with the exception of the pie and sunburst chart which are created in Microsoft Excel. R studio v1.1.383 has been used for scrapping, preprocessing and cleaning the data.

An online interactive visualization has been created using the Tableau tool. It is published online and available at the following link:

https://public.tableau.com/profile/raghav.mehta#!/vizhome/StLouisCrimeData/Dashboard1

2 Visualisations

2.1 Visualisation 1

The first visualisation is a simple *bar chart* where the number of crimes has been plotted for each month only for the years 2017 and 2018 by applying a filter on the years field. This has been done as our data is up to June 2019 and would show incorrect analysis due to missing statistics for the months of July to December. The total number of crimes for every month are also labelled as well as the intensity of the colour blue also depicts the same. This redundancy mapping (Ward et. al, 2015) of colour and length of bar chart is used as memorability leads to an effective visualization (Borkin et. al, 2013). The bar chart has been used as we are looking to compare the categories of a single variable and the number of categories is not very high. It is clearly visible that the crime statistics are higher in the summer months.

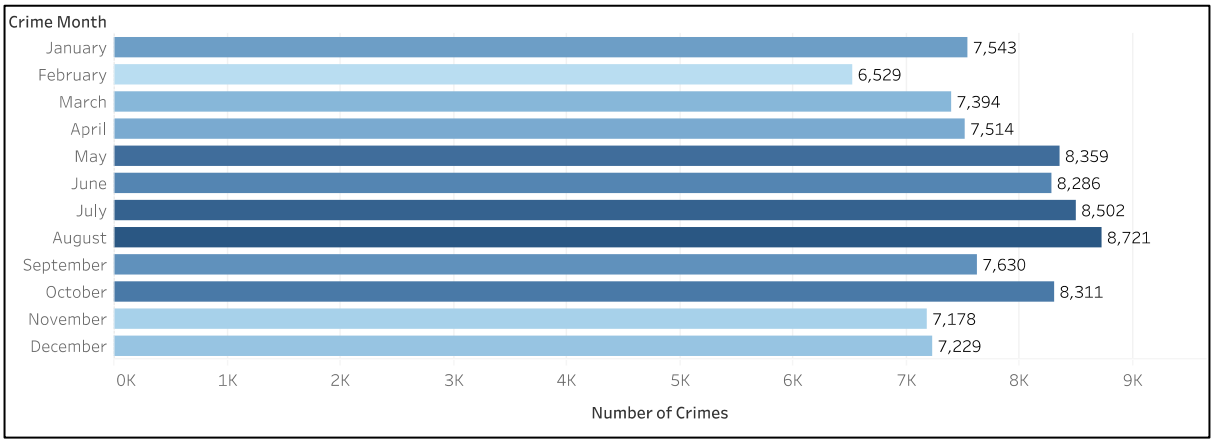

Fig 1. Crime Distribution By Month

2.2 Visualisation 2

This visualisation is a *geographical map* of St. Louis City with the crime locations mapped. The crimes are only mapped for the neighborhoods with the maximum crime over the period of 30 months by applying a filter in Tableau. This visualisation uses the latitude and longitude coordinates which had been transformed in R studio. The crimes are colour coded as per the neighborhood to easily be able to differentiate between the neighborhoods. We see that in 2 cases, there are 2 adjoining neighborhoods with maximum crime reported. It is also noteworthy that the population in these neighborhoods is greater than the other regions of St. Louis City, as denoted by the grayscale on the map. This helps to understand that crime in these districts within St. Louis City is high and one should be wary in these areas. Since this is the only method to map geocoded data, hence this symbol map was used in Tableau to plot the data.

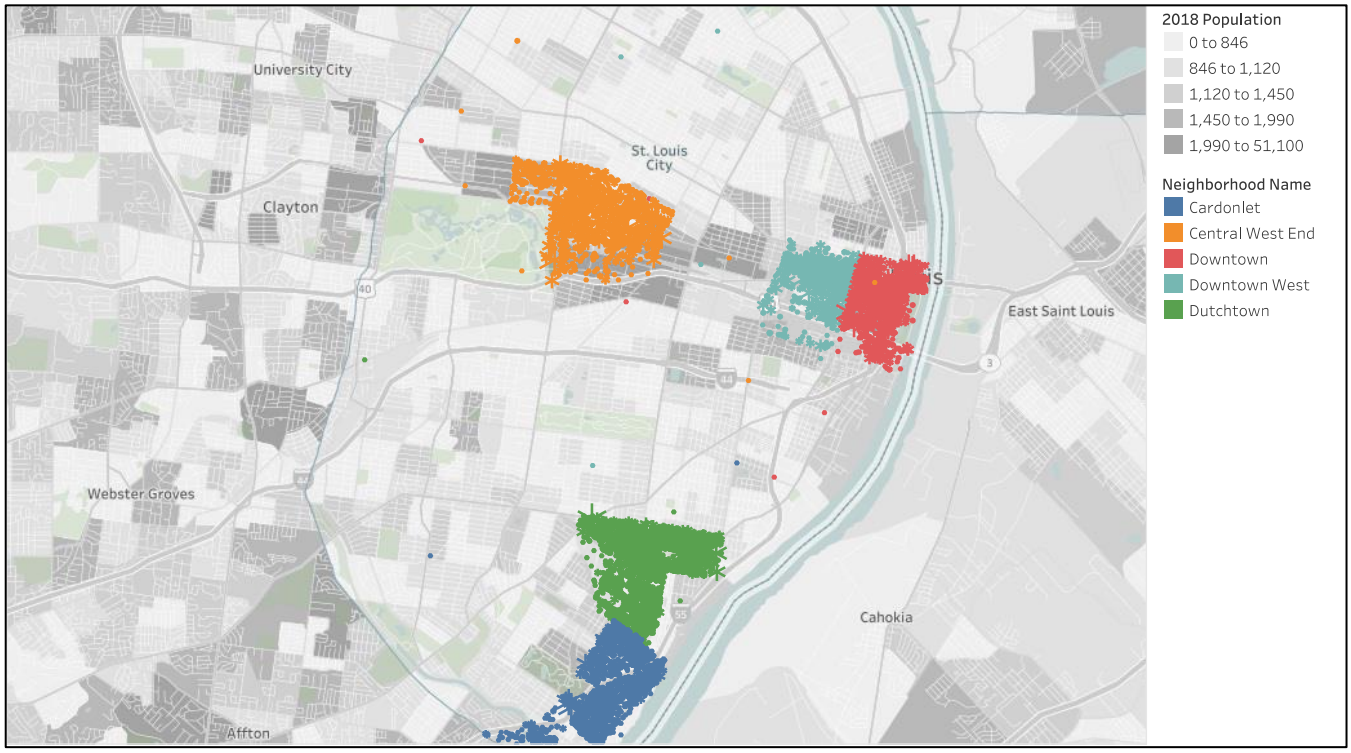

Fig 2. Top 5 Neighborhoods With Maximum Crime Reported

2.3 Visualisation 3

This visualisation is known as *Bump Chart.* Instead of using absolute numbers, this chart type helps in easily understanding trends in data by ranking the categories. The top most category is ranked 1 while the bottom most is ranked 10. This visualisation is created in Tableau using the dual axis function on a calculated field called *Rank* (which ranks the UCR crime type on the basis of the number of crimes). Only the top 10 most occurring crime types out of the 26 have been filtered out. Each colour represents a certain crime type and shows the occurrence rank of the crime through the years 2017 to 2019. We infer from this visualisation that Larceny-theft is the still the most occurring crime throughout the period. Drug abuse violations have gone up from 2017 to 2018 as compared to robbery and similarly more Motor Theft crimes have been reported than Burglary from 2018 to 2019. This simple ranking helps to understand whether a particular crime type has decreased as opposed to others and helps the police department to retrospect on their actions or change in laws which has resulted in a positive or negative manner.

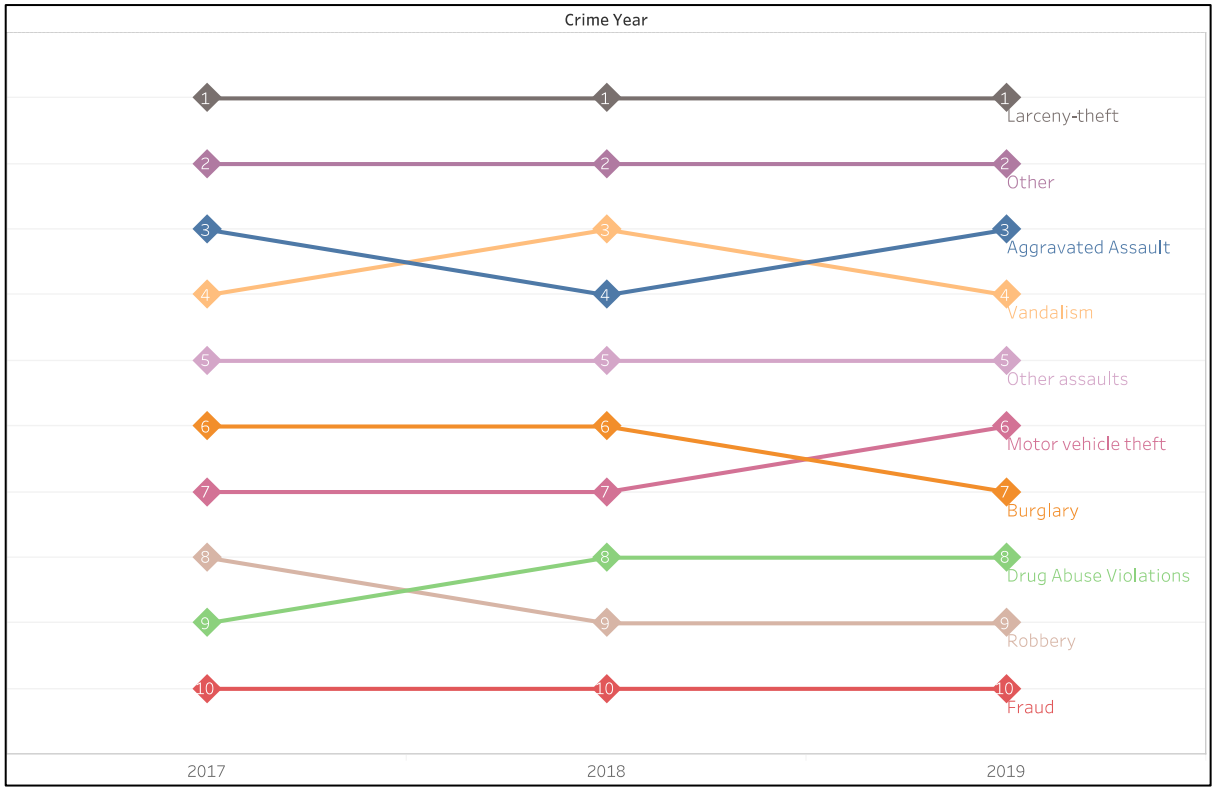

Fig 3. Ranking Of Top 10 Crime Types (2017 - 2019)

2.4 Visualisation 4

The following visualisation is a *Bubble chart* which uses three variables – the UCR crime type, the crime category and number of crimes committed in each category. The size of the bubbles represent the number of crimes and the bubbles are colour coded red and orange based on the UCR crime type. Since crimes are heinous acts therefore the colours used are red and orange, red depicting the more violent crimes whereas orange depicting less serious and white collar crimes. As we can see that there is a strong likelihood of violent crimes occurring in St. Louis City and the uncategorised '*Other'* crimes are also high in occurrence. However overall the red bubbles, though only 8 as opposed to 26 orange ones, are large and make a clear impact on simply viewing the visualisation. Hence theft, burglary, assault and vandalism are more common and the police need to step up to counter it.

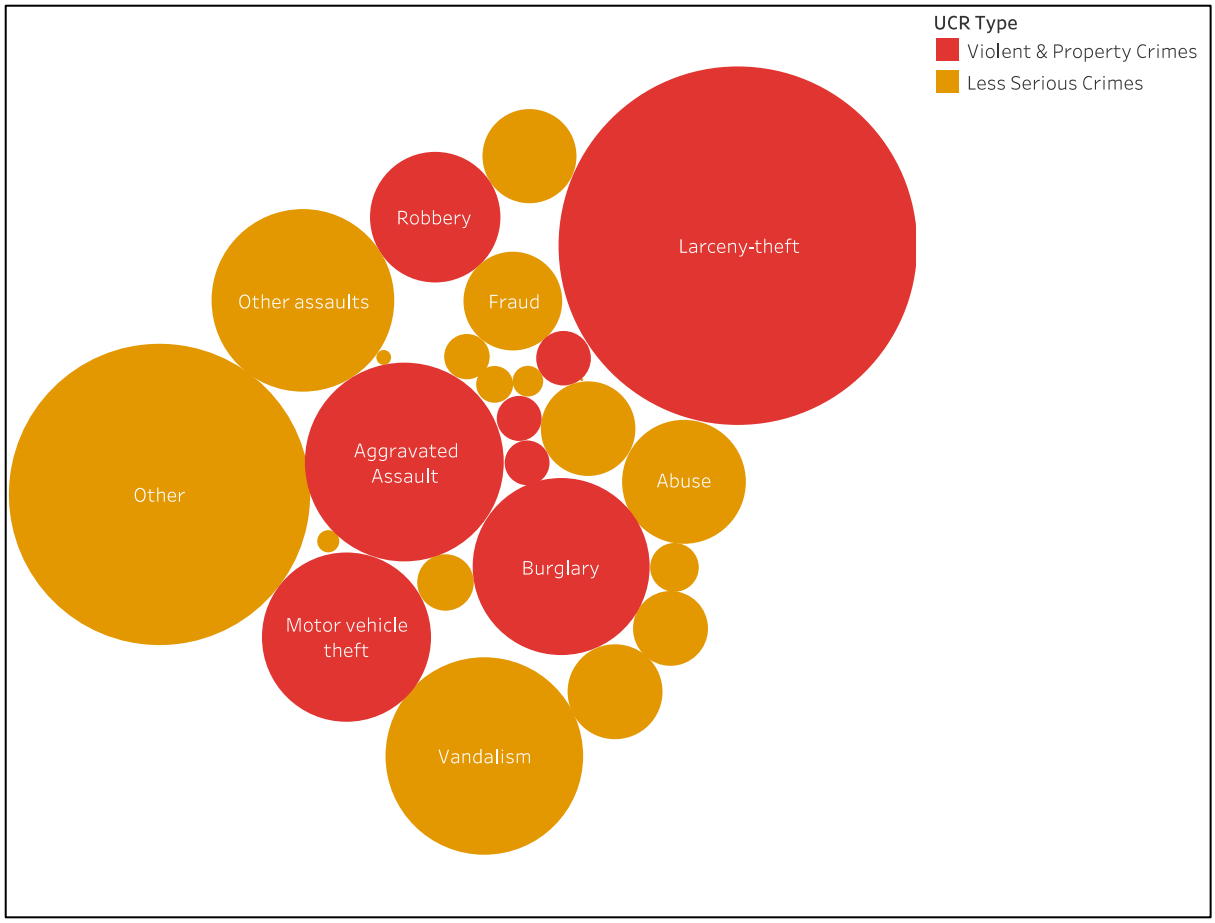

Fig 4. UCR Crime Types

2.5 Visualisation 5

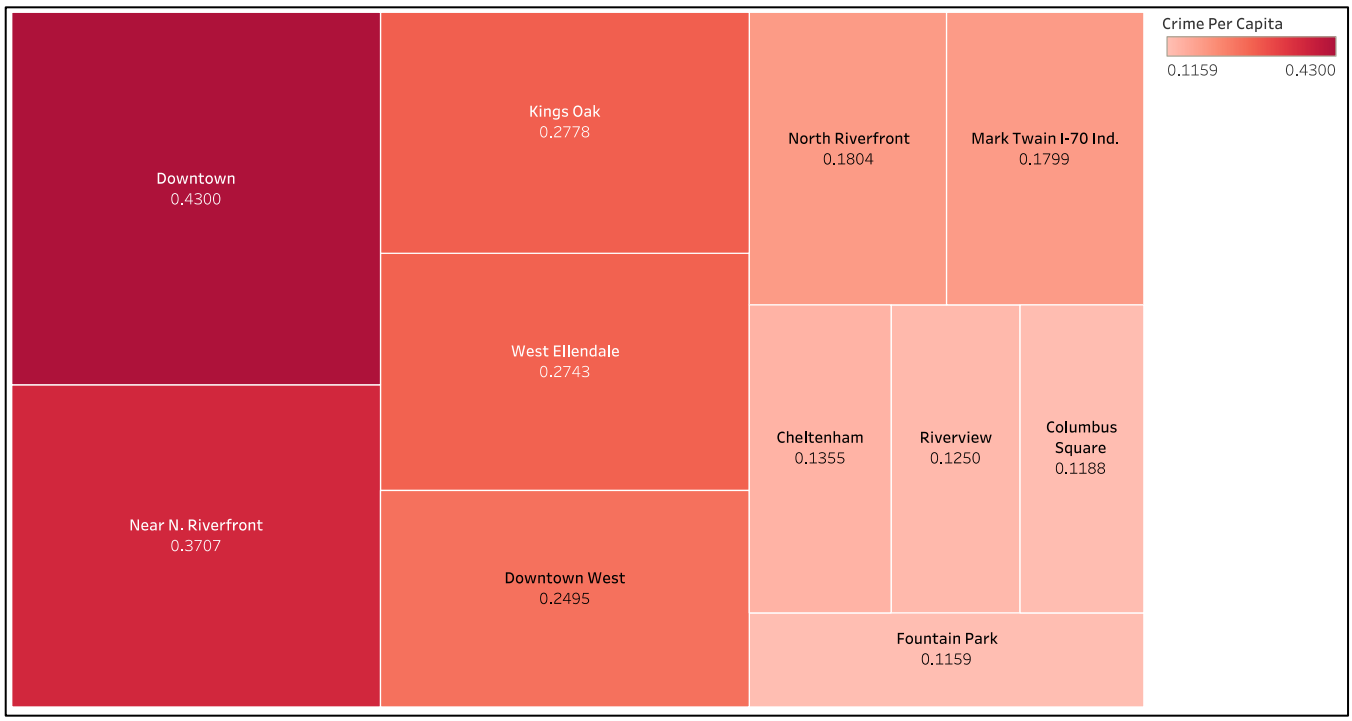

Fig 5. Crime Per Capita In The 10 Most Unsafe Neighborhoods For 2019

This visualization is a *tree map*. While tree maps are used for hierarchical mapping using multivariate data, it is also effective in showing the composition of a single variable. This visualization again reinforces redundant mappings to show the composition of one variable. For this visualization, the crime per capita of every neighbourhood is calculated by simply dividing the each record (count : 1) by the neighbourhood population and aggregating this value for every neighbourhood for the year 2019. These calculations have been done in R studio itself. The colour scheme used is the red spectrum as it is universally used to denote danger. The size of an individual tile and the intensity of the colour, both denote the crime per capita. Since there are 88 neighbourhoods in St. Louis City, a filter has been applied in Tableau to highlight only the top 10 most unsafe areas by crime per capita.

2.6 Visualisation 6

The following visualisation is a *Word Cloud* symbolising the most common crime descriptions in St. Louis City. The data points for this visualisation covers all crimes for the past 30 months. This visualisation gives a simple and clear story with just one glance as to which are the most often committed crimes in the city. The data used for this visualization is just the crime description for all the records. Using different colours while has no apparent coding, is just used to make the visualisation aesthetic and easy to understand. These are the cornerstones for good design (Rams, 2012) and makes the visualisation accessible on the whole as it is unobtrusive as well (Kirk, 2016).

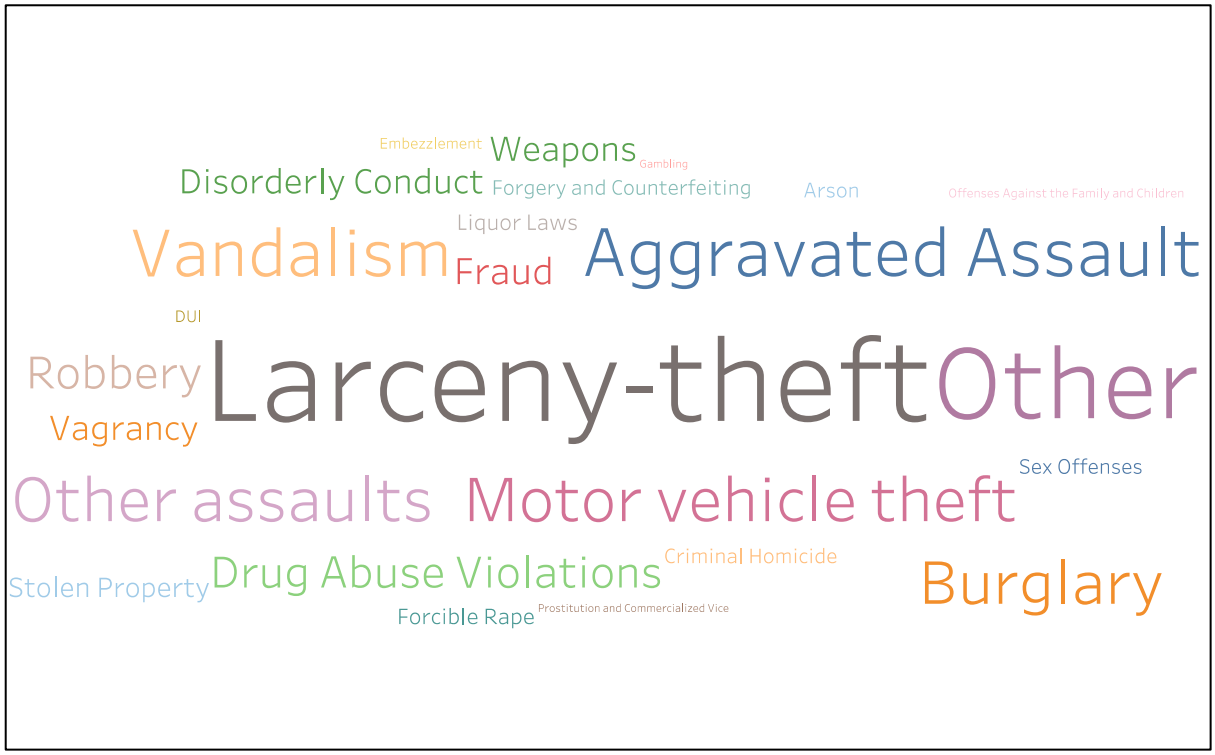

Fig 6. Word Cloud: Most Frequent Crime

2.7 Visualisation 7

This visualization is *sunburst chart* and is created using Microsoft Excel. The preprocessing for this data, by selecting the required data points was done in R studio and exported to a CSV. It depicts the breakdown of the two Uniform Crime Reporting categories of the FBI and the subcategories amongst them. Hence a hierarchical mapping seemed to be a suitable choice for the same. The visualisation is only for the 5 neighbourhoods of St. Louis City with the maximum crime. The colour scheme is a dark and light shade of dull orange to depict the more and less crime types. As evident from the visualisation,

there is no clear trend in the type of crimes being committed in these neighbourhoods and there is an equal likely chance of any crime occurring. However 42% of the crimes are of serious nature, which is an alarmingly high number.

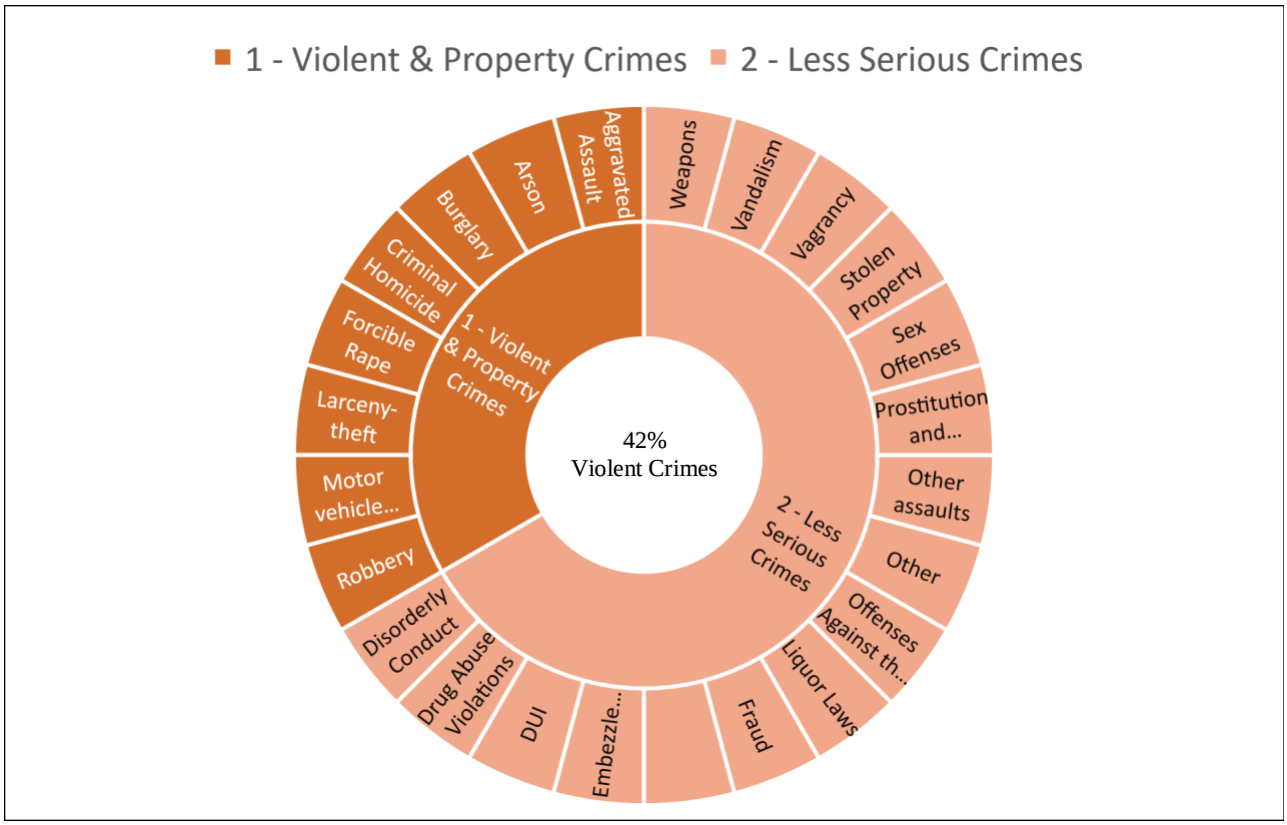

Fig 7. Crime Types and Crime For 5 Most Violent Neighborhoods

2.8 Visualisation 8

Race Desc	Location Desc	
White	Restaurant	6
	Residence/Home	3
	Highway/Road/Alley/Str..	3
	Church/Synagogue/Tem..	1
	Bar/Night Club	1
Unknown	Residence/Home	1
	Parking Lot/Garage	1
	Highway/Road/Alley/Str..	1
	Drug Store/Dr. Office/Ho..	1
Black	Highway/Road/Alley/Str..	4
	Air/Bus/Train Terminal	1
Asian	Air/Bus/Train Terminal	1

Fig 8. Hate Crime Statistics By Victim's Race & Location

The next visualisation is just a simple *tabular chart*. The data used for this visualisation is hate crime statistics for St. Louis City from 2017 to date. The table shows the number of hate crimes reported for

each of the race and the location of occurrence. Deep blue green colour represents highlights the most crimes committed in a specific location for each race. It is evident from the simple table that majority of the victims belong to the *White* race. This is contrasting from the hate crime statistics when we compare it with the entire nation. The visualisation also shows that the common places where such hate crimes occur are Highways/Roads/Alleys/Streets and Restaurants. While throughout the country hate crimes are targeted towards *Black* race, it shows that St. Louis may be a black dominated city. Hence the occurrence of hate crimes against *White* race is greater. But simply comparing the hate crime numbers vis-à-vis the other types of crimes, one can interpret that it is not a frequently occurring crime nor a serious issue.

2.9 Visualisation 9

This visualisation is a *pie chart* which shows the composition of the hate crime types in particular. Since the angle of the pie represents the share of the total, it is a simple yet effective visualisation to show composition when we are not dealing with absolute numbers. It is elegant with minimal design and hence considered a good design (Rams, 2012). The colour scheme is bright red to lighter shades depicting the seriousness of crime. Aggravated assault is considered most violent as it could be life threatening. While destruction/damage/vandalism is the lightest shade as it may not cause any physical injury or mental trauma to the victim. A staggering 62% of the hate crimes resulted in violence of the nature of assault, meaning that the victim has been physically attacked in these cases. This chart has been created in Microsoft Excel.

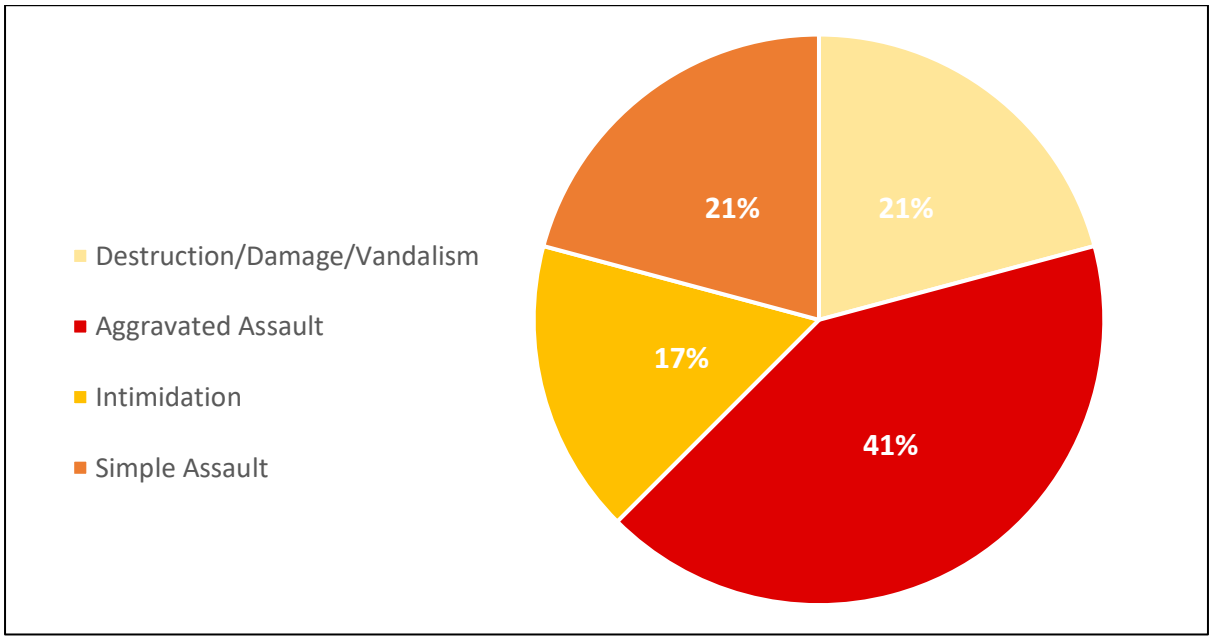

Fig 9. Hate Crime Offence Distribution

2.10 Visualisation 10

This visualisation is a *bar chart* which has been segregated by 2 categories. On the bottom axis we have the year of occurrence of hate crime whereas on the top we have the race description of the victim. The bar charts are colour coded to depict the race. It is easily understandable to see the trends amongst a single race (same colour bar charts) by viewing the length of the bars as well as seeing the difference amongst the hate crimes between different races. It can be seen that the number of hate crimes decreased to only 2 cases in the year of 2018. However there has been a decrease from 2017 to 2019. Since the data is only up to June 2019, one may be unsure as to hate crimes have risen from 2017 but can be

definitive that there has been a surge from 2018. *Asians* are the least targeted race to hate crime in St. Louis City which may be due to lesser Asian immigrants or lack of reporting these crimes to the police. Using semantically resonant colours for Race if done on skin colour would be ethically ungrounded and hence universal colours have been used to denote so.

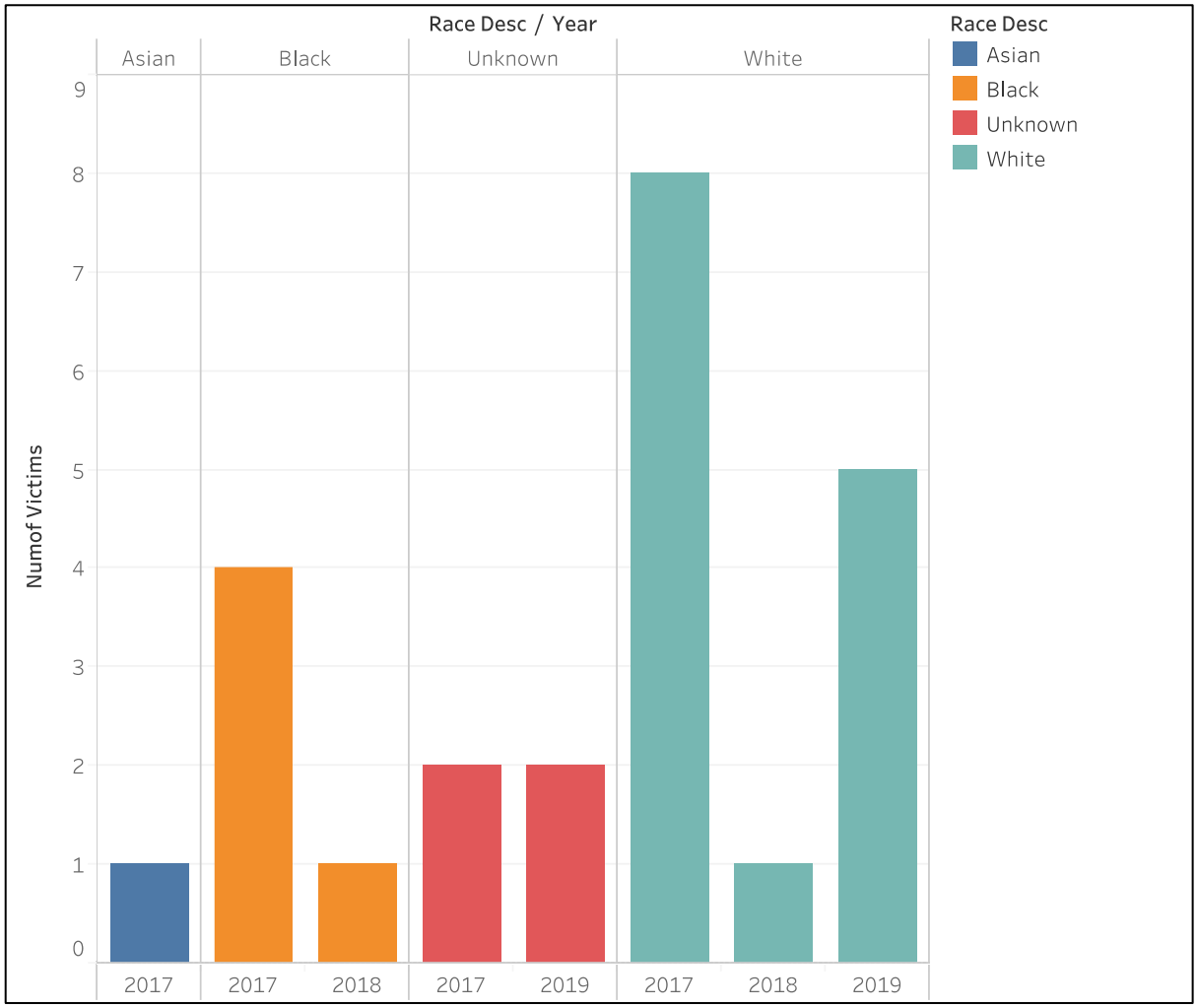

Fig 10. Hate Crime By Race Per Year

3 Summary

The report and visualisations attached give an insight into the crime occurrence across St. Louis City over the last 3 years. Certain visualisations are multivariate while some are univariate but reinforce analysis that is trying to be portrayed. Colour (hue and intensity), length and size are the most common pre-attentive attributes implied in the report. They add for non-complex visualisations which require lesser time to comprehend and easy to understand for the stakeholders. Some trends which are visible is that the densely populated regions have far greater crime rates (per capita) as opposed to sparsely populated regions. This may also be an indicator of the income range in those areas. Violent crimes are very common in St. Louis City, with theft, assault and burglary being the most common ones. Hate crimes, while insignificant in number, do occur but mainly against the White race, indicating that this city may have a considerable number of Black or other non-White races. Intelligence can also be drawn on the most unsafe neighbourhoods of the city, them being closely clustered and having all types of crime occurring in abundance. Thus these visualisations help the police department by facilitating the understanding of crimes in St. Louis City.

REFERENCES

Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A. and Pfister, H., 2013. What makes a visualization memorable?. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), pp.2306-2315.

Ingraham, C., 2015. 'The ugly truth about hate crimes — in 5 charts and maps', *The Washington Post*, 18 June 2015 [online] Available at: https://www.washingtonpost.com/news/wonk/wp/2015/06/18/5-charts-show-the-stubborn-persistence-of-american-hate-crime/ [Accessed 2 Aug. 2019].

Kirk, A., 2016. *Data visualisation: a handbook for data driven design.* Sage.

Rams, D., 2012. Ten principles for good design. *bit. ly/QUk4gY, retrieved Sept*.

Ward, M.O., Grinstein, G. and Keim, D., 2015. *Interactive data visualization: foundations, techniques, and applications.* AK Peters/CRC Press.