**Data 551: Project**

**By:** Elizaveta Volkova, Chris Donoff, Cristian Hluchy

**Dataset Selection**

Considerable time was spent on the selection of a relevant dataset, as the group felt strongly that the building a presentation would be far more interesting if the information being analyzed was compelling. After applying the constraints that the dataset must have spatial and temporal information, a search of Kaggle with the key word ‘Canada’ yielded a Vancouver Crime dataset that contained all crime information from 2003 to 2017, broken out by time of day, type of crime, location, and neighbourhood. After exploring the data, it was recognized that many questions the group had could not be answered without expanding the data. Subsequently, data from 2016 census was added, with features such as household income, employment, dwelling valuation and household size. From here, the team brainstormed several questions that could be explored in a Tableau Dashboard:

Does the amount (and type of crime) vary by neighborhood?  
How has the amount of crime changed over time (month or year, by neighborhood?)  
Does day of week impact crime?  
Does time of day impact crime?  
Does unemployment rate impact crime?  
Do demographics (household size and neighbourhood size) impact crime?

**Data Cleaning**

In order to join the two datasets, some manipulation was required to ensure the neighbourhoods were similarly named before merging. Once this was complete, thought was given to how the information would be manipulated in Tableau. As a result, several of the columns were aggregated to summarize the demographic parameters of interest. As the team was focusing on neighbourhood for a lot of the analysis, a quick search of the Vancouver open data portal displayed shape files that were easily imported into Tableau, and subsequently joined to the data.

The team also joined the day name and month name to the data in Tableau to make visualizations based on the day of the week the crime took place. At this point, the focus turned to creating a series of visualizations that would answer the aforementioned questions.

**Visualizations**

To answer the first question regarding where the most crimes happen in Vancouver, we generated a density map depicting higher number of crime reports with darker shades of red across the 22 neighbourhoods appearing in the crime.csv file from Kaggle. The neighbourhoods were outlined using the shape files provided from the Vancouver open data portal, and filters were added to enable the selection and focus on neighborhoods of interest. The tableau tooltip functionality was used to provide additional information, including the neighborhood name, total number of crime reports, and breakdown of number of crime reports by type (i.e. Theft from Vehicle, Mischief, etc.) when hovering your mouse over a neighbourhood. One theme from these visualizations revealed that theft from vehicles was consistently the highest proportion of crime reports across neighbourhoods.

To further look at how neighbourhoods differ with respect to the relationship between demographic parameters and the number of crime reports, we produced several scatter plots of interest. The number of crime reports was plotted against the neighbourhood size, and we observed a slightly positive-sloping trend, such that as the density of neighbourhoods increases, more crime reports are reported. This intuitively makes sense, however it could have been that in quieter neighbourhoods criminals feel they can operate unseen, and frequencies could have been mirrored. By linking these results to the breakdown of crime reports by type, we can further identify types of crime that are predominant in smaller sized neighbourhoods. Interestingly, besides theft from vehicles, we do see these less dense neighbourhoods having higher proportions of break-and-enter residential offenses. We also plotted number of crime reports against household size (the average number of occupants in a household). Ignoring the ‘outlier’ of downtown, we still observe a slightly negative trend, such that less crimes are reported in neighbourhoods with larger household sizes. This could be due to a number of factors, such as frequency of activity outdoors (kids playing outside), or even pose as an index of loneliness, where neighbourhoods supporting single occupancy suffer from these isolated individuals performing criminal acts. An interesting null result occurred when looking at the effect of unemployment rate. No detectable pattern was observed, therefore it does not appear as though unemployed individuals with time on their hands get up to nefarious acts.

To answer the aforementioned questions regarding the amount of crime changing over time and time of the day/day of the week impact on the crime amount we generated a separate dashboard with two visualizations: “Crimes by month & day of week” and “Crimes by time of the day”. Both graphs were outlined by type of the crime to see the tendency of crimes happening in different conditions. The tableau tooltip functionality was used to provide additional information, including the filtering by neighbourhood name, hour, month and year of the crimes committed.

Interestingly, the quantity of theft from vehicle crime reports haven’t notably changed between different days and month of the year (with the highest rate of reports at 6pm and lowest one at 5 am) while the quantity of bike thefts has dramatically varied: the number of thefts of bikes more than doubled during summer period. This could be due to the number of factors, such as number of bikes used due to the weather conditions, or even the number of tourists attending the city during summer holidays. The number of thefts of vehicles has also notably varied during different time periods – the highest number has been reported during summer time while the lowest number of thefts of vehicle was reported in February. The lowest number of theft of vehicles has been reported in while the highest rate was noted at 10 pm.

The vehicle collision or pedestrian struck (with fatality) tended to have the lowest rate around all the time periods with highest number of reports at 11am and 7pm. On the contrary, time pattern for break and enter commercial crimes, break and enter residential/other crimes, mischief and theft of the vehicle counted the highest quantity of crimes reported during the nighttime from 10pm to 4am.