Crime Analysis – City of Toronto

IAT 814 Final Report

Submitted to: Dr. Lyn Bartram
Simon Fraser University

By: Chithra Bhat
ID: 301346287

ABSTRACT

Toronto, the provincial capital of Ontario, is ranked as the fourth safest major city in the world, and the safest major city in North America. However, in 2018, Toronto had the highest homicide rate among major Canadian cities and was often on national headlines for fluctuations in violent crime.

Crimes are a common social problem affecting the quality of life and the economic growth of a country. With the increase in number of crimes each year, law enforcement agencies are demanding new approaches to protect their communities. Crime analysis concentrates on a set of systematic approaches to provide timely and useful information on crime patterns and trends. Effective crime analysis influences all areas and operations of a police department by refining and disseminating valuable data which can be used to improve the operations and administration of police departments and to augment patrol activities in crime prevention.

The focus of this project is to perform an in-depth analysis of the major types of crimes that occurred in the city, observe the trend over the years, retrieve valuable information from crime data using exploratory data analysis and determine how various attributes, such as education, income and other attributes contribute to specific crimes in order to help police department make better decisions.

I. INTRODUCTION

Toronto is ranked as the safest major city in North America and the fourth safest major city in the world, as cited in Wikipedia[1]. It consists of 140 officially recognized neighborhoods. As is the case with any big city, some neighborhoods are considered to be less safe than others. Several reasons are attributed to higher crime such as lower income, unemployment rate, lower literacy and access to education, among other reasons.

An analysis of crime and neighborhood data within Toronto will provide us with a good understanding of how many of these assumptions are true and to what degree. It might additionally reveal hidden patterns, trends between some factors and major crime that would not be obvious at the outset. The problem may be articulated as – finding major crime trends in Toronto's neighborhoods, identifying potential factors related to major crime, and using these factors to help police department make better decision.

II. DATASET DESCRIPTION

The datasets used in this project are taken from Toronto Police Service Public Safety Data Portal[2] and Toronto Open Data Catalogue[3]. These datasets contain about 166853 records from 2014-2018. The following are the datasets used in this project:

- **MCI 2014 to 2018:** The Major Crime Indicators (MCI) dataset includes all MCI occurrences by reported date and related offenses. The time frame of this dataset is from 2014 to 2018. The MCI categories include - Assault, Break and Enter, Auto Theft, Robbery and Theft Over (Excludes Sexual Assaults). The location of crime

[1] https://en.wikipedia.org/wiki/Crime_in_Toronto
[2] http://data.torontopolice.on.ca/search?collection=Dataset
[3] https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/

occurrences have been offset to the nearest road intersection node to protect the privacy of parties involved in the occurrence. Hence, all the locations are considered as approximate locations. The dataset consists of 27 columns. It has been simplified to just 7 columns for analysis.

- **Homicide:** This dataset includes all Homicides from 2004 to 2018. Divisional association to each homicide is based on current Police Divisional Boundary. This dataset[4] has about 353 records from 2014 to 2018.

- **Wellbeing Toronto – Education:** The education dataset[5] is provided by Toronto Public Library, Toronto District School Board, Toronto Catholic District School Board and Mothercraft. The data for each neighbourhood are based on the mathematical aggregation of smaller sub-areas (in this case Census Tracts) that when combined, define the entire neighbourhood. I have cleaned this dataset to get the percentage of total population without any education certificate for each neighbourhood.

- **Neighbourhood Profiles 2016:** City of Toronto Neighbourhood Profiles use the Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each neighbourhood. The data is sourced from a number of Census tables released by Statistics Canada. The dataset consists of 2383 records and 146 columns. I have cleaned this dataset to extract unemployment rate, individuals with more than 80k income and individuals with less than 5k income for all 140 neighbourhoods.

- **Neighbourhoods:** This is a shapefile consisting of boundaries of City of Toronto neighbourhoods.

[4] http://data.torontopolice.on.ca/datasets/homicide-1
[5] https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#dae097ba-5a76-b817-e724-362611c10317

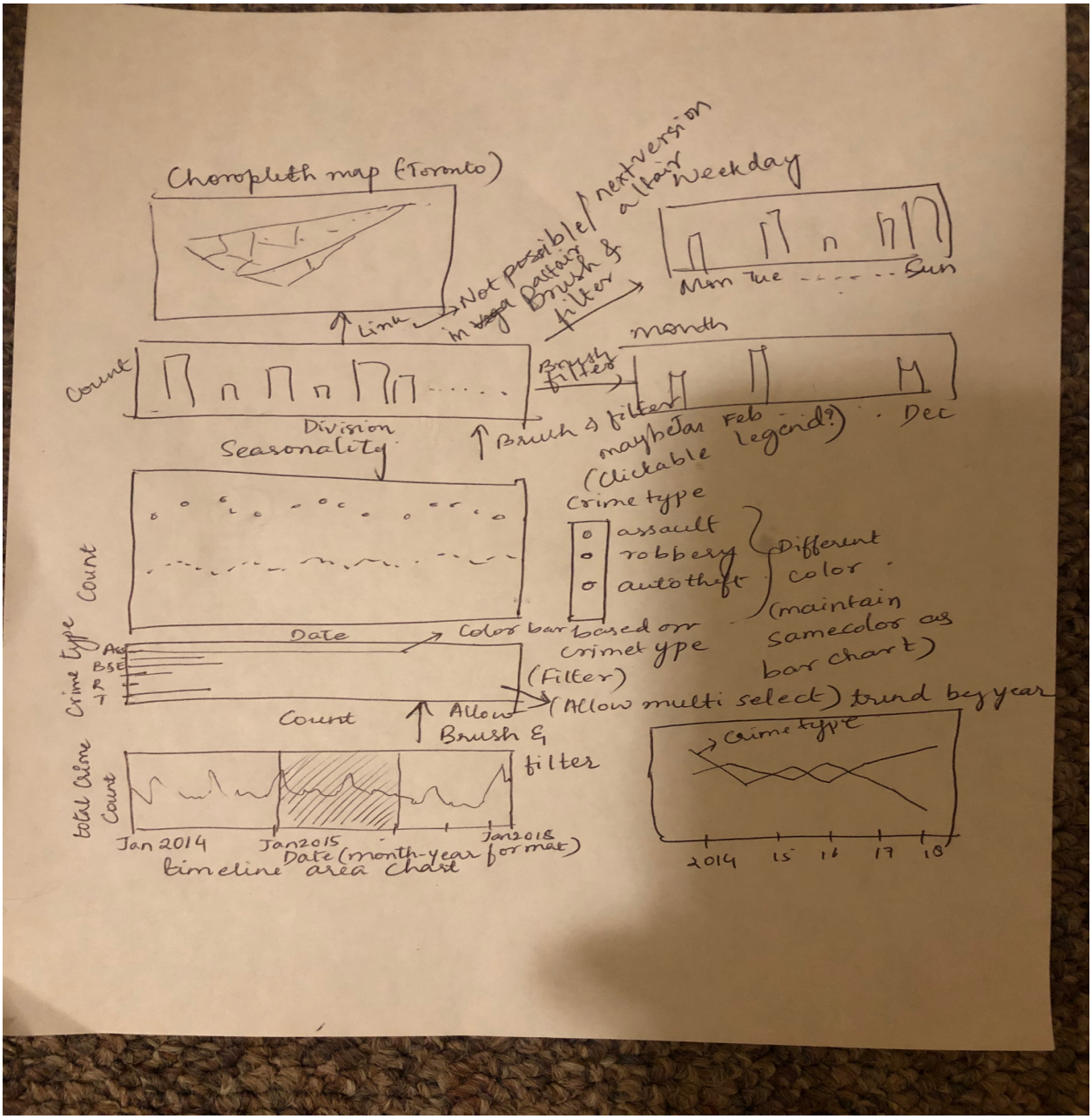
III. DATA PREPROCESSING

The steps involved in data cleaning and preprocessing are described below:

1. Both MCI and Homicide datasets had records starting from the year 2000 to 2018, but were reported only after 2014. In order to maintain consistency, I filtered the data that occurred and were reported starting from 2014 to 2018.
2. The education dataset column names were very long with spaces in between. Renamed the columns to meaningful names. Computed percentage of population with no education certificate from the data.
3. The income dataset had too much information. I had to pick the rows that I was interested in, did data wrangling and converted it to desired format to join with Neighbourhood data.
4. Filled missing values with appropriate values.
5. Added new columns to extract day of the week and month from the date column.
6. The combined dataset was huge with too many and it slowed down the process of loading my viz. I dropped all the unnecessary columns, computed new values and added additional information to the tooltip.
7. To plot the choropleth map, converted latitude and longitude information to POLYGON type.

IV. DESIGN

After feedback from Professor, I had to change how I showcased my output. So, below is the design for the project.



(From bottom to top)The initial idea was to have a timeline indicating number of crimes committed per day. The user will be able to brush to select the time period that he/she is interested in. Then, they can also click on any Crime Type(multi-select), Divisions(also multi-select) to see crime statistics by month, year , weekday based on other filters. Also, I wanted the choropleth map to respond to Crime Type selection. But, due to limitations in Altair(We cannot filter choropleth map based on dropdown selection or brush. It will be released in the next version), the choropleth map now responds to ipywidgets dropdown and that's why we have to use jupyter notebook to run the project. However, all the other

charts respond to brushing and selection event and the chart can be saved as .html or .png or .svg using the three ellipsis to the top-right of the chart.

V. RUNNING THE APPLICATION

In order to run the project files, install dependencies of the project by entering the following in a command line:

```
$ pip3 install –r requirements.txt
```

Then, launch jupyter notebook with the below command:

```
$ jupyter notebook ––NotebookApp.iopub_data_rate_limit=10000000000
```

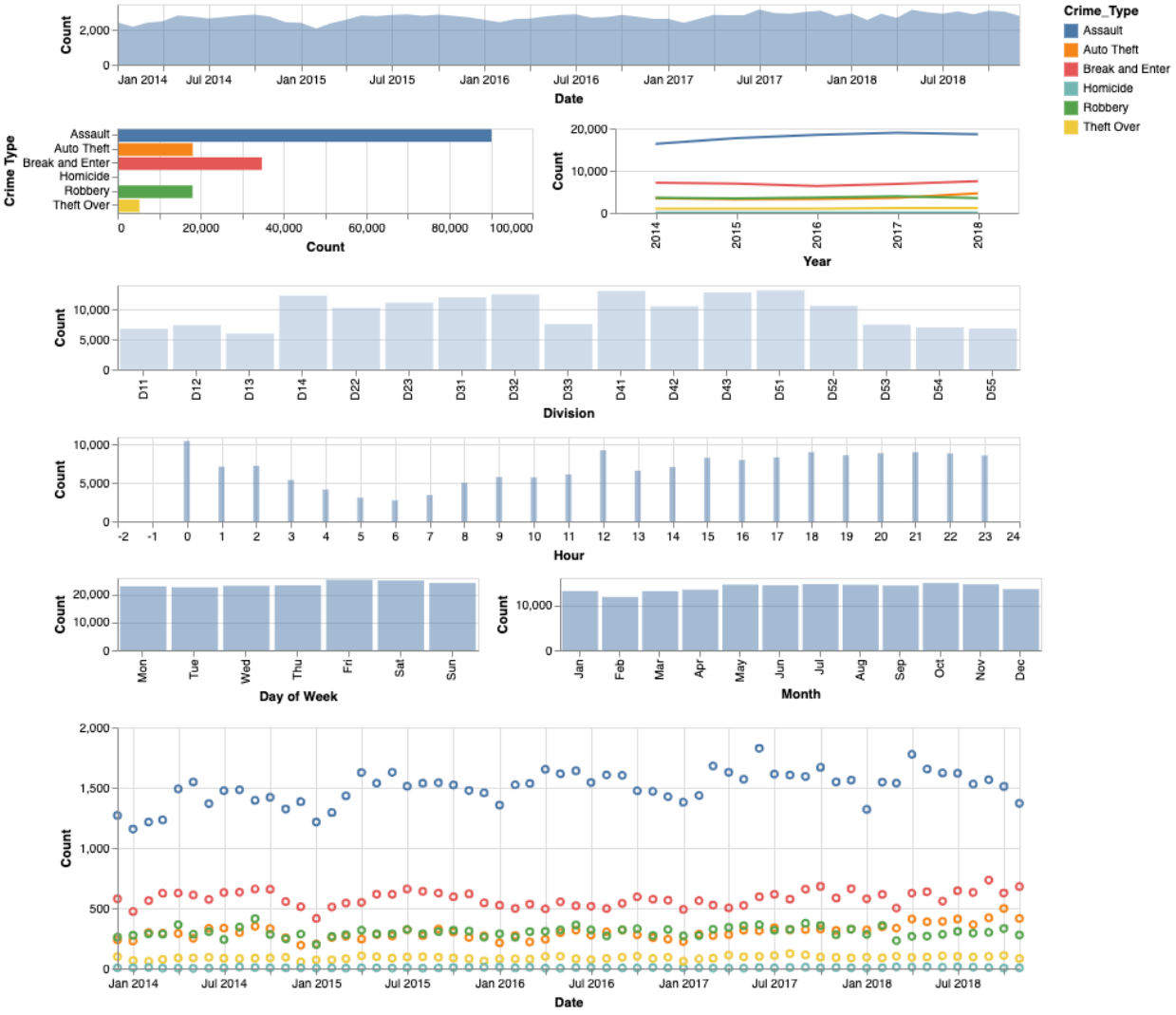
Use `Shift + Enter` to run the cells.

VI. RESULTS

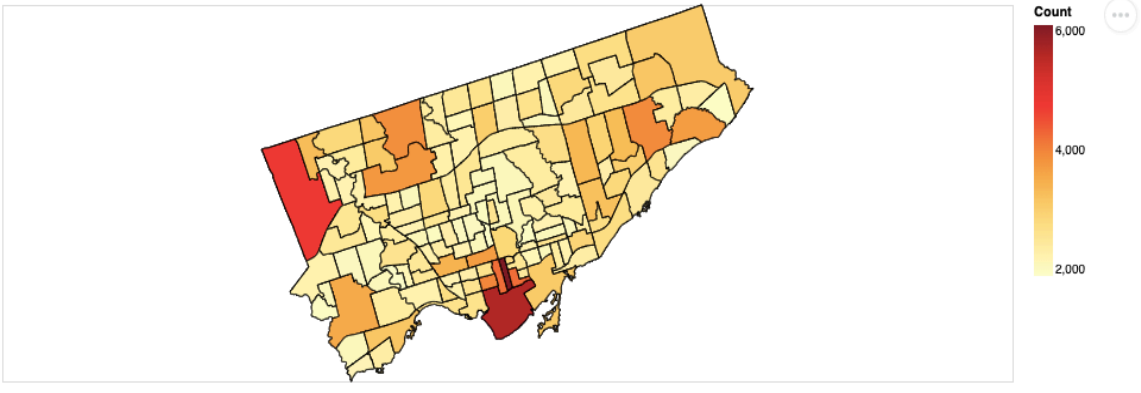
This project was developed using Altair, Geopandas, IPywidgets and Pandas. Looking at the viz below, we can answer a few questions. From the first chart it is clear that the number of crimes per year is increasing.

A summarized initial visualization of all the major crimes in Toronto

1. What is the most prominent major crime indicator?
    - Assault is the most prevalent form of major crime in Toronto followed by Break and Enter.
2. Which division has witnessed the highest crime? Which division is the safest?
    - D51 followed by D41 has witnessed highest crime. While, D13 and D11 has witnessed least crime.
3. More crimes are committed on weekends than on weekdays.
4. Thanks to harsh Canadian winters, the city is most safe around December and February. One can also see some seasonal pattern in the crime trend. They tend to increase around May to August and then gradually decrease.
5. What types of crime are most frequent at each hour?
    - All crimes seem to happen mostly after 12pm until midnight. Break and Enter happens in the afternoon, and all other theft after 8 PM.
6. Most dangerous neighbourhoods?
    - Waterfront and Church-Yong Corridor seems to be the most dangerous neighbourhoods.



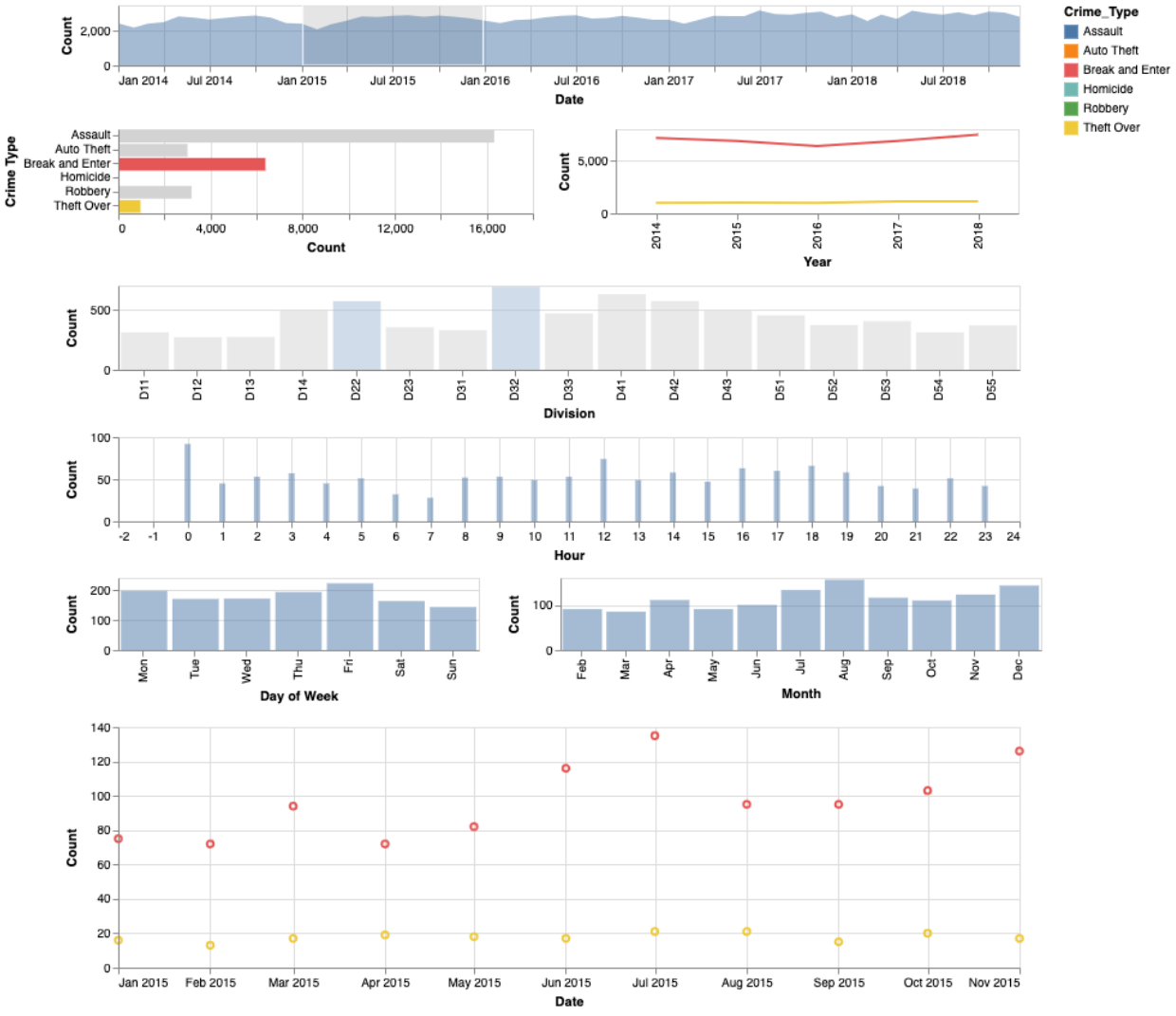
CrimeType All



VII. BRUSHING AND LINKING AND COLORS

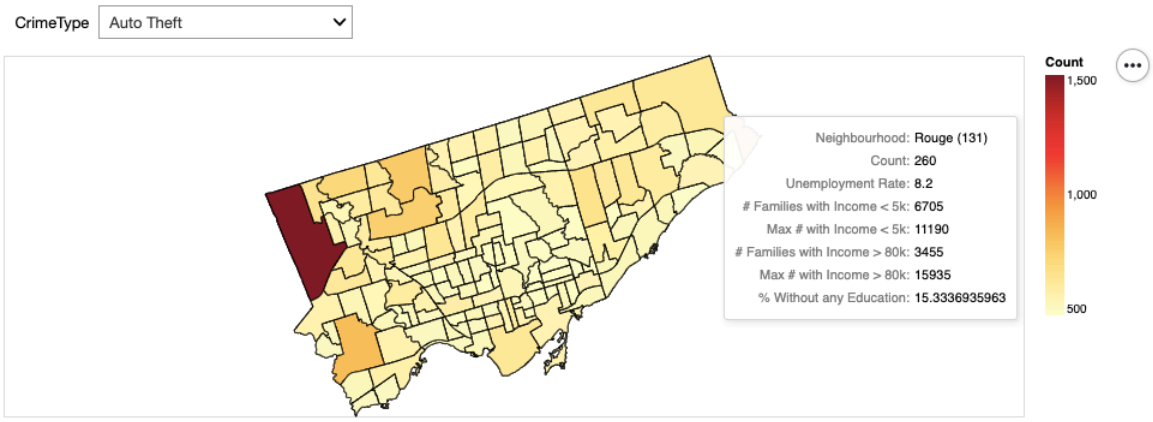
One of the important function of this project is brushing and filtering. When the user drags their cursor on the area chart(timeline), all the charts except the choropleth map, are filtered. The user can now also select a particular crime type or multiple crime types to filter it furthermore. He can also filter it further by division and all the charts below will change accordingly.

Furthermore, each crime type(categorical) is given a unique color and this color is not shared by any other crime type. Throughout the visualization, the selected bars are highlighted using their original color and the rest are colored grey(to mean they are not selected). The chart below shows multiple selection with multiple transform filters.



In the above fig, the first three rows are clickable(selectable) and the last three are non-clickable. The last three displays the count with all the filters.

In the choropleth map, the user can change the crime type to any specific crime type to see safest and most dangerous neighbourhoods. A bright red color indicates danger and cool shade like light yellow indicates that the crime rate is comparatively less. Furthermore, the user can also use his cursor to on a neighbourhood to get the employment and education statistics. For example, I can select Auto Theft and we can see that it happens mostly in one neighbourhood i.e., West Humber – Clairville (Maybe don't park your car there ☺).



Some observations from employment and income dataset are below:

1. Waterfront has the least percentage of uneducated people. Even though, assault is more common here, crimes like Robbery and Homicide are not popular.
2. Most crimes happen around those neighbourhood where percentage of uneducated in the surrounding neighbourhood is also higher.
3. Downtown area has maximum people who make more than 80K and also maximum people who make less than 5K CAD PA. This could be the reason why Break and Enter is most common here. Also, income inequality could be one of the factors affecting crime in these areas.
4. Homicides are more common in the areas where people are less educated.

VIII. DISCUSSION, FEEDBACK AND VISUALIZATION PRINCIPLES

For the design of this project, I have followed few visualization principles such as careful selection of colors, size of the data, choice of visualization idioms etc. which will be discussed below.

1. **The choice of visualization idioms**

   When I first started this project, I had multiple bar charts and they were executed in multiple cells. It was not very user friendly and the user had to switch between the cells to make sense of the visualization. This would mean the user had to remember what he saw in one cell and relate that to the output in the other cell. This made visualization not very user friendly. This point was made clear to me by my Professor and I have put all charts into one cell to avoid switchbacks. Now, it is lot easier to compare the values.

2. **The choice of colors**

   No two crime types have the same color and the default bar chart colors are not same as crime type colors. This way the user do not have to wonder what the color mean every time.

3. **Reducing complexity**

   Initial code was really slow because I was trying to merge neighbourhood data with income , education and plot it on a choropleth map with multiple layers. This took too much time to execute and sometimes the notebook would crash. To avoid that, I have changed the code to make the data lot smaller, added tooltip wherever required and even in terms of code written to get the chart is lot smaller compared to the original code.

IX. FEASIBILITY STUDY, LIMITATIONS AND FUTURE WORK

I had all my classmates, Professor, TA, three friends and myself for feasibility test. The participants in this study are all graduate students with knowledge about visualizations. I received lot of feedback initially to make this project better and some really nice comments. Most of them liked the choropleth map where it highlights West Humber for Auto Theft. A

few who saw TPS analytics dashboard mentioned that this project was lot easier to get insights and you don't have to scroll a lot and it is not dense. Few participants spoke highly of the seasonality pattern in the crime which was not evident from the existing solution.

The participants also pointed out some flaws in the project. Initial visualization was not great. As I mentioned before, user had to switch between the cells, remember the output and compare it with the charts in the new cell. I have tried my best to incorporate all their feedbacks in this new visualization.

I also received a feedback from my roommate who said " I wish you could collapse the code part and display just the charts in the browser". Well, because in Altair we cannot use dropdown value to filter data for choropleth map, this is currently not do-able. This is likely to be updated in the next version. This can be part of future work where we can integrate dropdown value to build a choropleth map and then we can save the chart, that way we can avoid seeing the code.

X. CONCLUSION

I have presented Crime Analysis for the City of Toronto and presented the findings from feasibility study. I have also discussed about the design principle and visualization principles behind this project. The purpose of this visualization is to help police department and law makers identify crime patterns and trends to make better decisions.

XI. REFERENCES

1. Brantingham, P.J. Crime diversity. Criminology 2016, 54, 553–586. [CrossRef]
2. Charron, M. Neighbourhood Characteristics and the Distribution of Crime in Toronto, Ontario: Analysis on Youth Crime; Statistics Canada and Canadian Centre for Justice Statistics: Ottawa, ON, Canada, 2011; No. 85–561.
3. Lersch, M.K.; Hart, C.T. Space, Time, and Crime; Caroline Academic Press: Durham, NC, USA, 2011.
4. https://altair-viz.github.io/