Introduction

Crime happens every day in every city. If it is possible to find a rule when and where a crime may occur, it can help the police to deploy police force. With this rule, the police can get to the scene faster to increase the chances of solving a crime or stop a suspect in the process of committing a crime. Therefore, I choose to analyse the crime data in Vancouver from 2003 to 2017. And the data sources are from Kaggle (Osaku, 2017), City of Vancouver Open Data Portal (2021), and a website records Sunrise Sunset Times of Vancouver in 2010. The intended audience is the police officers from all neighbourhoods. With this visualisation project, they can compare different crime's time characters like hours, months and in a day. This can help them better allocate police force at different time periods.

Design

The first visualisation of my project is a Choropleth map, there are 23 polygons to clearly demarcate of each area, sequential palettes is used to show the number of crimes. The darker the colour, the higher the chance of a crime happening. The map is used to show the crime number of different neighbourhoods. The reason I use Choropleth map is that crime is related to the areas, data are classified and mapped on the map to show differences in quantity and marked as such in the legend (Rhonda. 2022). Colour makes it easier for audience to see the differences between areas (Czapiewski, 2013). They can also find a large amount of information and general patterns in a short time. The colour palette I chose is "YlOrRd", since red colours in the eye are clustered in the area near the centre where the sharpest images are formed (Science Focus).

The second visualisation is 24 donut charts. Compared to pie chart, donut chart is better in comparing a handful of categories immediately and how much they accounted for (Beautiful.AI Team, 2020). For this case, time of a day consists of daytime and night time. At first, I'm thinking about using a map like the first one, and I found some problems in it. Firstly, similar images can bring consistency, but flexible diverging palettes will blur the specific proportions. If the palettes scale adjusted too loosely, 59% daytime crime ratio may have the same colour with 41%, and if the palette scale is too tight, values that differ significantly from 50 percent are not clearly displayed. Secondly, I expected to use stacked bar chart to show the day and night share of crime, however 24 is really a big number, and it is difficult and not realistic to show all 24 bars in one. Although this can not only display the proportion, but also judge the total number of crimes, but it will make it difficult to see the subscript or locate the corresponding neighbourhoods you want to see. Having a neighbourhood with a high crime rate can also make it difficult to see the daytime crime rate in a

neighbourhood with a lower crime rate, therefore it was also deprecated. Finally, I was thinking about only choose some of donuts chart like top 5 daytime crime ratio and the last 5 daytime crime ratio, or make a drop down, these two designs make it hard to show the difference between all neighbourhoods, since the intended audience is the police officers, they need compare their own neighbourhoods with others, so showing all neighbourhoods becomes more important, they can communicate with each other about police deployment to reduce the disparity in the proportion of crime during the day or night, and use this to find solutions to deal with crime. Then turns to colour, I chose orange colour stands for daytime and blue colour stand for night time is because bright colours represent daytime, dark colours represent night time is in line with common sense and people's cognition. I also show tick label for a chart to make it clearer what each colour represents. There are also specific proportions above to make the audience more impressed. Donut charts are sort alphabetically, which allows audience to locate the city they are looking for more quickly.

For the last visualisation, I choose line chart, one dimension is the best way to visualize, line graph is the graph that best represents a trend. Line charts can show change by depicting the change with line segments moving from left to right. As the movement occurs, observe the slope that will move up or down (PPCexpo, 2021). The baseline of two line charts are both zero to avoid misleading to audience.

For all visualisation above, I only have one slider bar for overall interaction, that’s to ensure all data are in the corresponding year, it won't make audience forget to adjust different slider bar to create wrong impression. Year is ordered data, and the position on common scale is the best way to let users interact.

Compared to the original design sheet, I’ve made some changes. After my presentation, the tutor told me that my design is wrong. For each sheet from 2 to 4, I need supply 3 different designs for each question. And I only supply one, so after brainstorming again, I change one of the original design sheets to the current ones.

For the second design sheet, the main design is the same, but I didn’t add population and crime rate into it, since it is hard to find population for all years. According to Statistics Canada (Statistics Canada. 2022), the frequency of census of population is every 5 years. Therefore, I can't get the data for each year.

For the third design sheet, after rethinking I found that the same crime has almost the same trend in timing. Therefore, stacking them together doesn’t affect the trend. Finally, I chose to put crime types as tick label, I can use only

two line charts to show changes in the number of crimes within a day and a year. Visually it brings more information and is easier to understand. The pattern is the same, the only difference is the time dimension.

For the fourth sheet, the original design sheet can show the specific number for each neighbourhood in each year, but it is hard to compare it to different neighbourhoods, so it is also rejected.

Implementation

The library I used are shiny, rgdal, leaflet and plotly. Shiny is an R package that makes it easy to build interactive web applications, it is used to build the website. Rgdal package can read and modify spatial data. In my implementation, I used readOGR() to read OGR vector maps into spatial objects. Leaflet can mapping spital data into a map, it's a package with rich features and easy to use. Plotly is a powerful package capable of making a large number of different graphs including 2d and 3d (Plotly).

During the implementation I encountered some challenging issues. The first challenge is the collection of geographic information. It is easy to find shape files of any county in the United States. When the location is changed to Canada, it turns more difficult. It is easy to find provinces' shape files, but Vancouver is a city. After that, I found a shape file in the Open Data Portal in Vancouver. But the shape file is not complete, Stanley Park is not plotted in the map.

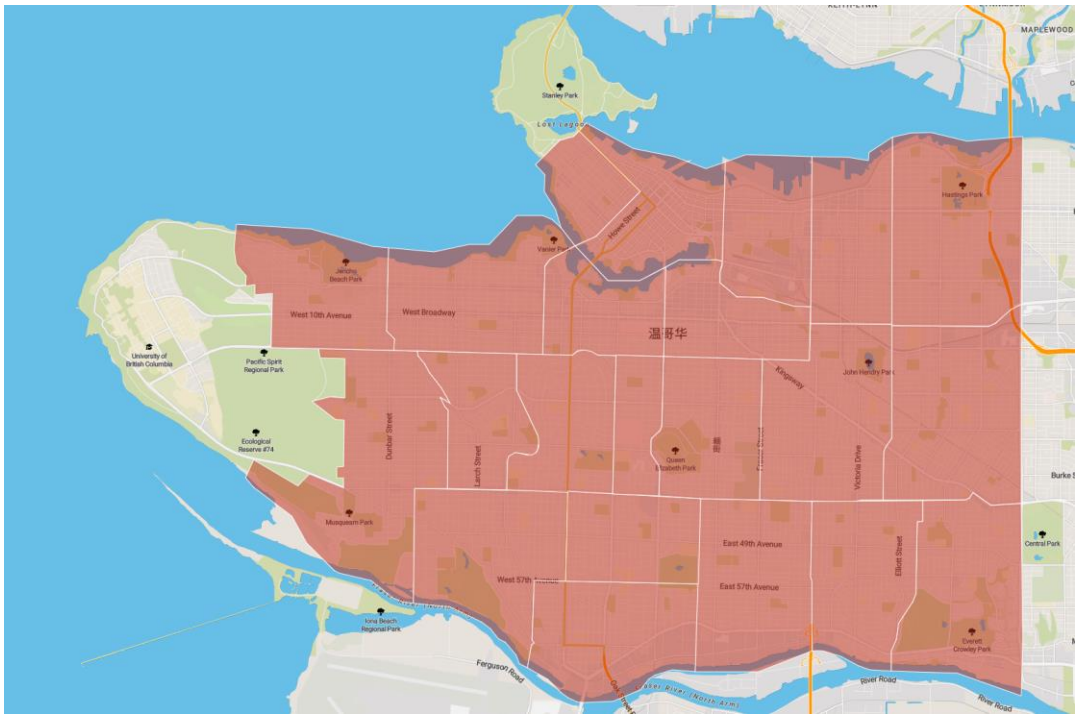

Figure 1 Local area boundary for City of Vancouver in 2019

And then I tried to find another shape file which consists of Stanley Park and

merge them together. After searching, I found a shape file almost suits my needs.

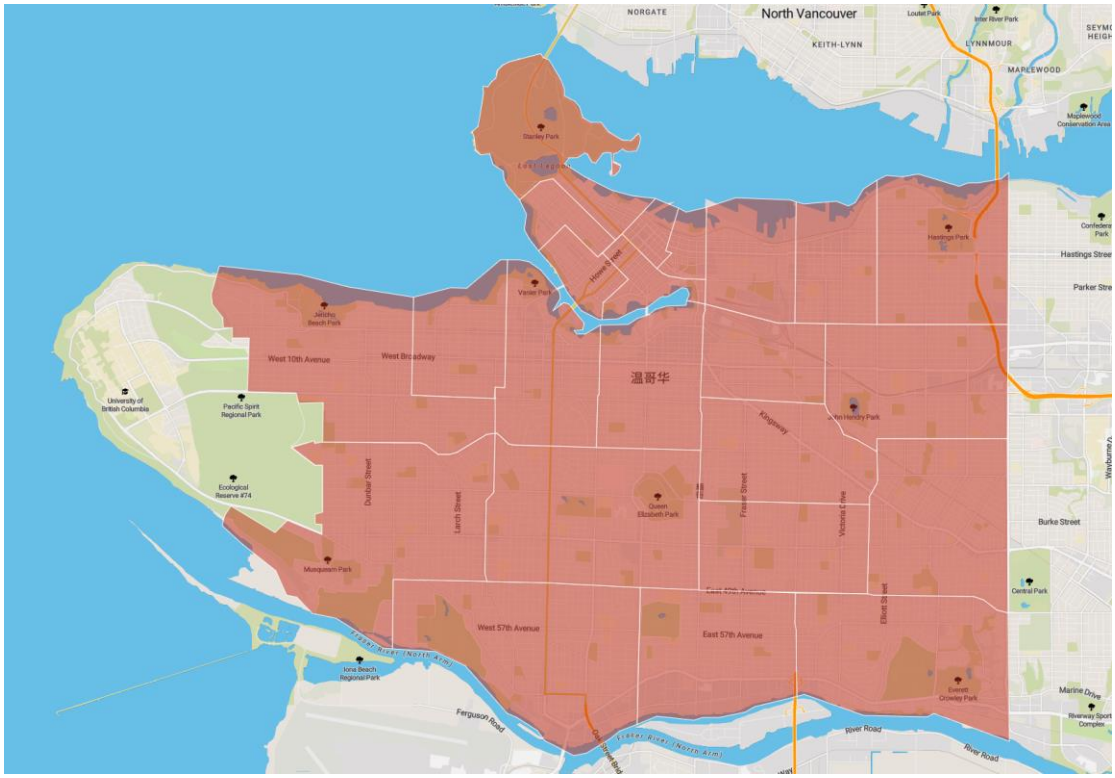

Figure 2 Property use inspection districts, 2021

And then, I use map shaper to edit the shape of Stanley Park, and then delete all unusable shape, finally merge them together.

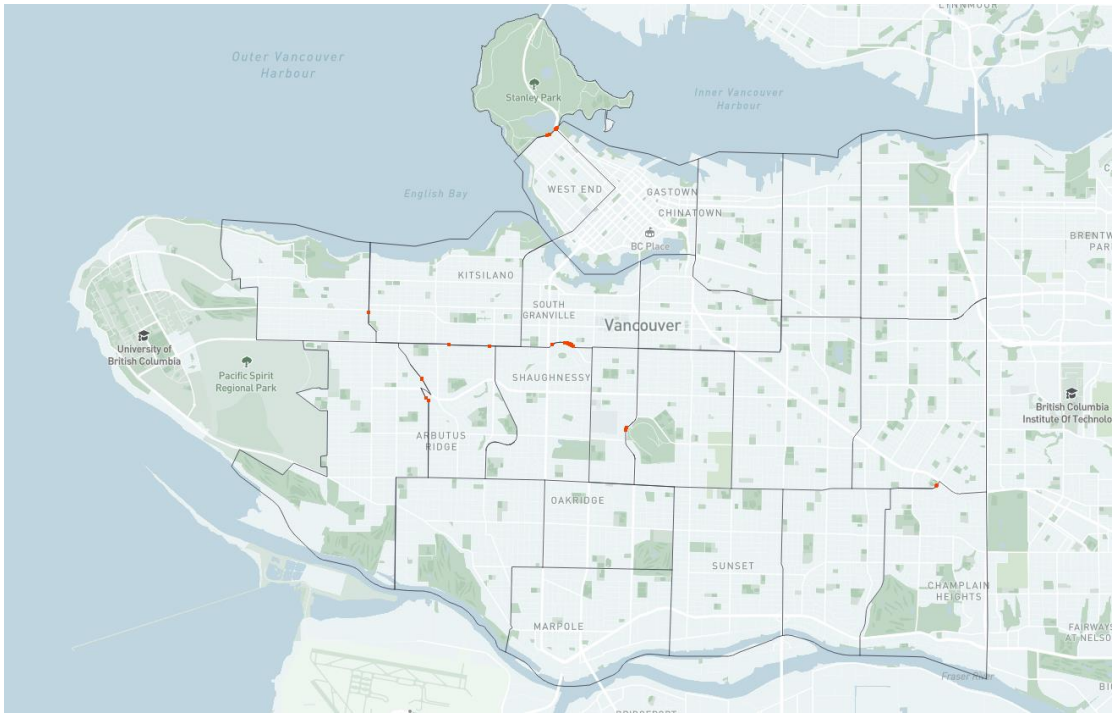

Figure 3 Full Vancouver map edited in map shaper

I also found that if I change the name in shape file, it doesn't correctly display the bound corresponding data in R, so I must keep the original name in the map.

And then is the data wrangling for donut charts, the data used is based on the data exploration project. To compare the sunrise and sunset time to the crime time, I changed all crime year to the same year, but I kept the index of the original dataset.

In [2]:
```
import pandas as pd
a = pd.read_csv('D:/crime_final.csv')
a
```

Out[2]:

Date	DaytimeS	DaytimeE	DaytimeL	index	TYPE	HUNDRED_BLOCK	NEIGHBOURHOOD	Latitude	Longitude
2015-01-01 0:00:00	2015-01-01 08:07:34	2015-01-01 16:25:22	8h 17m 48s	170	Theft from Vehicle	11XX HOWE ST	Central Business District	49.278669	-123.125802
2015-01-01 0:00:00	2015-01-01 08:07:34	2015-01-01 16:25:22	8h 17m 48s	171	Theft from Vehicle	11XX HOWE ST	Central Business District	49.278669	-123.125802
2015-01-01 0:00:00	2015-01-01 08:07:34	2015-01-01 16:25:22	8h 17m 48s	499	Break and Enter Commercial	10XX E GEORGIA ST	Strathcona	49.278202	-123.081422
2015-01-01 0:00:00	2015-01-01 08:07:34	2015-01-01 16:25:22	8h 17m 48s	881	Vehicle Collision or Pedestrian Struck (with I...	NELSON ST / THURLOW ST	West End	49.282485	-123.128052
2015-01-01 0:00:00	2015-01-01 08:07:34	2015-01-01 16:25:22	8h 17m 48s	1114	Vehicle Collision or Pedestrian Struck (with I...	E HASTINGS ST / JACKSON AVE	Strathcona	49.281232	-123.093428
...	...	...	...	...	...	...	...	...	...
2015-12-31 0:00:00	2015-12-31 08:07:38	2015-12-31 16:24:07	8h 16m 29s	455927	Break and Enter Commercial	2XX TERMINAL AVE	Strathcona	49.272497	-123.097541
2015-12-31 0:00:00	2015-12-31 08:07:38	2015-12-31 16:24:07	8h 16m 29s	455977	Break and Enter Residential/Other	31XX W 15TH AVE	Kitsilano	49.258806	-123.175040
2015-12-31 0:00:00	2015-12-31 08:07:38	2015-12-31 16:24:07	8h 16m 29s	456468	Break and Enter Residential/Other	2XX KAMLOOPS ST	Hastings-Sunrise	49.282847	-123.054519
2015-12-31 0:00:00	2015-12-31 08:07:38	2015-12-31 16:24:07	8h 16m 29s	456813	Break and Enter Residential/Other	29XX E 2ND AVE	Hastings-Sunrise	49.268432	-123.041551
2015-12-31 0:00:00	2015-12-31 08:07:38	2015-12-31 16:24:07	8h 16m 29s	457153	Break and Enter Commercial	33XX NORTH ARM AVE	Killarney	49.204849	-123.036004

Figure 4 Data wrangling output for data exploration project

Since the original dataset contains YEAR, MONTH, DAY, HOUR, MINUTE with different column, I can record the YEAR by using the index and column.

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
0	Other Theft	2003	5	12	16.0	15.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47
1	Other Theft	2003	5	7	15.0	20.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47
2	Other Theft	2003	4	23	16.0	40.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47
3	Other Theft	2003	4	20	11.0	15.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47
4	Other Theft	2003	4	12	17.0	45.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47
...	...	...	...	...	...	...	...	...	...	...
474010	Mischief	2017	1	18	14.0	44.0	14XX E HASTINGS ST	Grandview-Woodland	494563.75	5458727.40
474011	Break and Enter Residential/Other	2017	3	3	9.0	16.0	31XX ADANAC ST	Hastings-Sunrise	497265.49	5458296.71
474012	Mischief	2017	5	29	22.0	30.0	14XX E 7TH AVE	Grandview-Woodland	494533.97	5456824.97
474013	Theft from Vehicle	2017	6	5	17.0	0.0	8XX HAMILTON ST	Central Business District	491487.85	5458385.78
474014	Vehicle Collision or Pedestrian Struck (with I...	2017	6	6	17.0	38.0	13XX BLOCK PARK DR	Marpole	490204.00	5451444.00

474015 rows × 12 columns

Figure 5 Original Data frame

```
for indexs, rows in a.iterrows():
    i = int(rows['index'])
    a.loc[indexs,'YEAR'] = df_OAP_deleted.loc[i,'YEAR']
```

Figure 6 Code from anaconda

After that I drop the unused columns and add two new columns in the data frame to count the day and night for each neighbourhood and year. The as_index() function in group_by() is used to let the data frame in the format like SQL. And finally output it.

```
a
```

	YEAR	NEIGHBOURHOOD	light
0	2003.0	Central Business District	night
1	2003.0	Central Business District	night
2	2003.0	Strathcona	night
3	2003.0	West End	night
4	2003.0	Strathcona	night
...	...	...	...
473664	2016.0	Strathcona	night
473665	2016.0	Kitsilano	day
473666	2016.0	Hastings-Sunrise	night
473667	2016.0	Hastings-Sunrise	day
473668	2016.0	Killarney	night

473669 rows × 3 columns

```
for i,r in a.iterrows():
    if r['light'] == 'day':
        a.loc[i,'day'] = 1
    else:
        a.loc[i,'night'] = 1
```

```
b = a.groupby(['NEIGHBOURHOOD','YEAR'],as_index=False)
```

Figure 7 Code from anaconda

The data wrangling for the hours and months is nearly the same in steps, so it is omitted here.

Next part is implementation in R, the implementation of map is done by package leaflet, I used YlOrRd as colour palette, imported the shp file and csv file and then bind them together, and adjusted the transparency. The final visualisation is as follow.

Figure 8 Output from Rstudio

The implementation of donut charts used a for loop to determine whether more crimes are committed during the day or at night. Since at first, I found in each graph, the colour of day and night is not fixed, it is determined by the party with the larger proportion. Therefore, if the daytime crime is fewer, the graph can be normally display, when the daytime is more, I need to use a variable to represent the colour. I also found that the dividing point of two variables is not always directly above the donut charts, so I need to keep the demarcation point directly above by turning a certain angle. I used rotation in plotly to realize it. Now that it is always correct whether the ratio of daytime crime and night time crime. I'm also thing about using the thickness of donut charts to show the number of crime numbers in total, but I found if I did like that, some neighbourhoods with very small number are hard to see, even a line. And normalize the numbers to 0-1 is also not so realistic, due to it might be misleading.

Figure 9 Output from Rstudio

The final visualisation is two line charts. The implementation of these line charts is based on package plotly. I filter the data and records, define x axis with the first 24 and 12 variables for hours and months separately, there are nine different types of crimes in all, and I only choose eight of them is that there are too few samples of the remaining type. The final implementation is as follow.

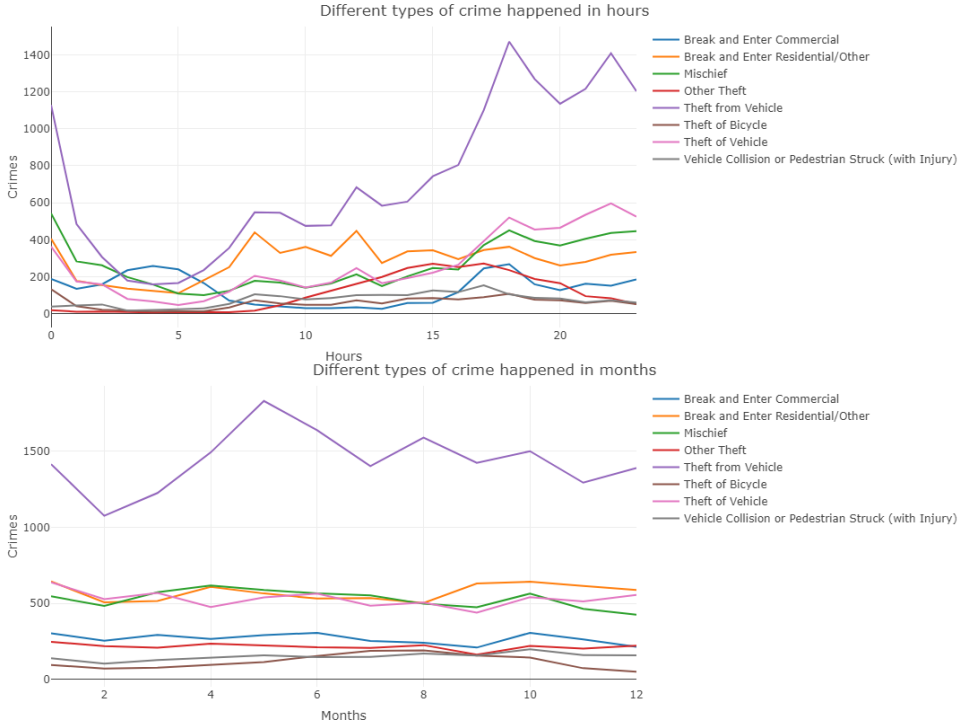

Figure 10 Output from Rstudio

User guide

All three visualizations are controlled by the slider bar at the top, so it is easy for audience to use. Move the slider bar can control the data of different year. As soon as move the slider bar the entire visualization changes.

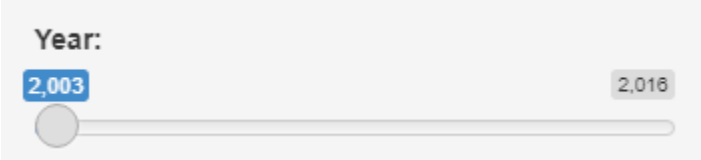

Figure 11 Output of Rstudio

For the first visualisation, click the plus and minus signs to zoom in and out of the map, and mouseover will show the specific number of crimes of that neighbourhoods. And the text on the left are the descriptions and findings from the map.

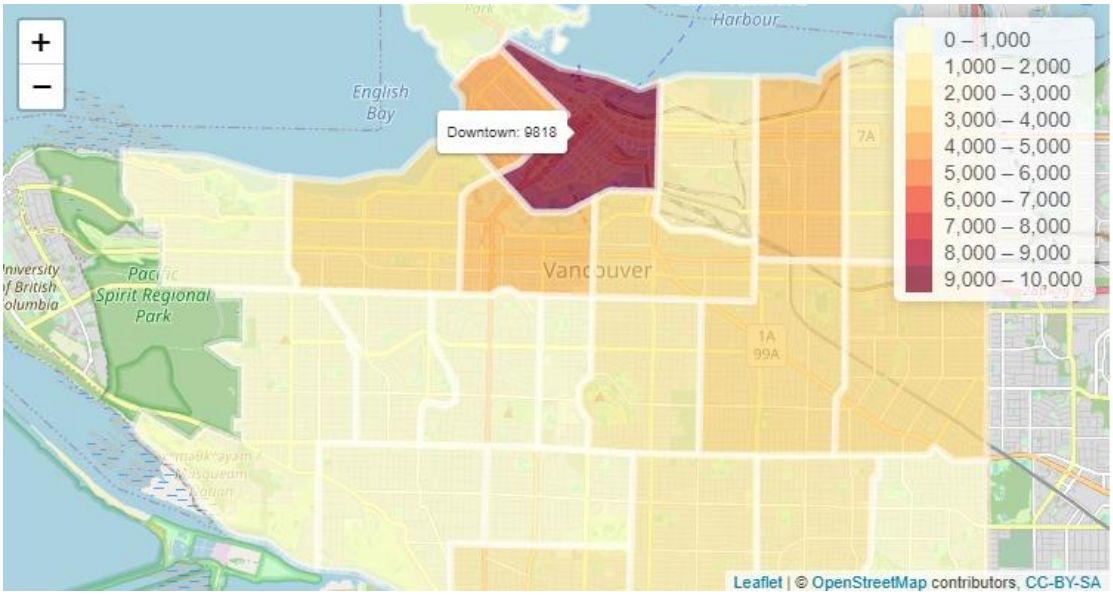

Figure 12 Output of Rstudio

For the second visualisation, a donut chart is set to show the tick label of day and night to reach a consensus with audience that the orange colour represents daytime while blue represents night time. Mouseover any part will display the number and the corresponding information.

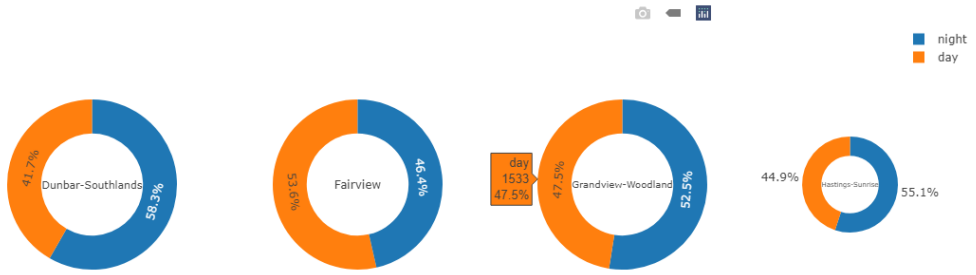

Figure 13 Output of Rstudio

For the last visualisation, mouseover the point in the chart can see the specific value of hours and the number of crimes. Double click the tick label on the right can display the trend selected. The y axis will also automatically change the scale to match the selected crime type. Click the tick label can display or remove the line, click zoom on the toolbar and draw a rectangle on the chart can show the selected part. Click pan can move the graph by holding down the mouse on the graph. Using autoscale or reset axes to restore the chart to the proper proportions, since they are the same in line charts.

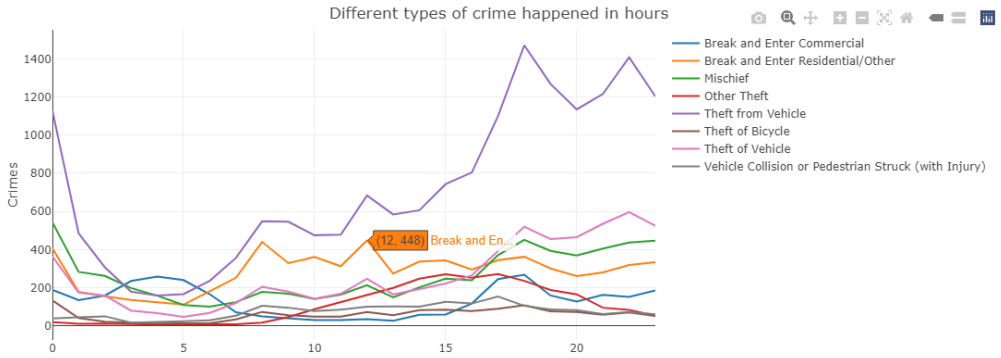

Figure 14 Output of Rstudio

Conclusion

To sum up, from this visualisation project, I find that if just look at the data frame or every single row of the data frame, you can not find anything. Only after visualisation, I can find the rule and answer the question I raised before. In this project, I converted a 500,000 rows data frame and two shape files to three different kinds of charts, which the amount of information contained in it exceeded my expectations. Even seemingly ordinary data can be visualized to discover patterns.

After this visualisation, I understand that making a visualisation is not like write a piece of code. When you want to convert the data, you can use any ways. Whatever the way you choose, the result is the same. When it comes to data

visualisation, every single step will make change on your output, there are also no standard answers. You can choose the style whatever you like, that is what I think is the most novel.

I think for the future, the first thing that could have been done better is the style. For this project, I felt since the intended audience is the police officers, not the public. Thus, the style is not so vital, easy to understand is the most important thing. They need to make study on the visualisation to find something. But when the audience turns to the public, I need to make the view better to attract them watching my visualisation. In this situation, easy to understand and having a good style becomes to the same importance. So my next step is to better improve the style to make the visualisation more charming.

Reference List

Beautiful.AI Team. (2020, October 21). *Battle Of The Charts: Pie Chart Vs. Donut Chart.* Beautiful.ai.
https://www.beautiful.ai/blog/battle-of-the-charts-pie-chart-vs-donut-chart

City of Vancouver. (2019, March 9). *Local area boundary.* City of Vancouver Open Data Portal.
https://opendata.vancouver.ca/explore/dataset/local-area-boundary/information/?disjunctive.name&location=13,49.25083,-123.12335

City of Vancouver. (2021, September 27). *Local area boundary.* City of Vancouver Open Data Portal.
https://opendata.vancouver.ca/explore/dataset/property-use-inspection-districts/information/?location=12,49.25629,-123.12395

Czapiewski, B. (2013, March 11). *Choropleth map in the analysis on the map.* Maps-for-excel.
https://maps-for-excel.com/blog/choropleth-map-in-the-analysis-on-the-map/

Osaku, W. (2017). *Crime in Vancouver.* Kaggle.
https://www.kaggle.com/datasets/wosaku/crime-in-vancouver

Plotly. *Plotly R Open Source Graphing Library.* Plotly.
https://plotly.com/r/

PPCexpo. (2021, November 13). *Best Chart to Show Trends Over Time: Why You Should Use a Line Chart?*
https://ppcexpo.com/blog/best-chart-to-show-trends-over-time

Rhonda. (2022, February 10). *What Is A Choropleth Map? Choropleth Map Advantages And Disadvantages.* Civil Stuff.
https://civilstuff.com/what-is-a-choropleth-map/

Science Focus. *Why are warning signs red?*
https://www.sciencefocus.com/the-human-body/why-are-warning-signs-red/

Statistics Canada. (2022, February 9). *Canada tops G7 growth despite COVID.* Statistics Canada.
https://www150.statcan.gc.ca/n1/daily-quotidien/220209/dq220209a-eng.htm