

Section 1: Basic info

Project title: Crime in Vancouver

Team Info: Bolin Wang (i3l2b), Michaux Sun (m5j2b), Kerry Zhou (q2k2b)

Section 2: Design Rationale

Components:

selectable / filterable data:

1. year
 - a. array: [start year, end year]
 - b. radio button for “all year”, horizontal slider with a widget for range selection
2. neighborhood
 - a. array: [name 1, name 2]
 - b. selected by clicking on the choropleth map
3. crime type
 - a. string
 - b. filtered by the segmented donut embraced the choropleth map, can only select one crime at a time
4. Day / Night
 - a. filtered by the toggle

widgets:

1. vertical slider: year bar
 - a. default year will be set to “all year”
 - b. user scroll along the vertical direction to select different years, the corresponding data for the selected year will reflect on the choropleth map
2. toggle: switching between day and night mode

Analysis for Innovative View:

Innovative view: choropleth map + donut chart + pie chart

1. choropleth map

Derived data:

We've converted attribute X and Y of utm value to lat/long and replaced them with two new attributes
lat: [49.201201, 49.313349], long: [-123.224021, -123.023393]

Mark:

- interlocking area encode neighborhoods
- point encode crime case

Channel:

- spatial region (identity) encode the geometric position of each neighborhood(categorical)
- color saturation (magnitude) encode severe level of different crime type number in total (quantitative, more number of crime cases, more saturate color) across neighborhoods
- spatial position (magnitude) encode crime cases with different lat/long (quantitative)

- a. **zoom out**: 24 neighborhoods (located by geographical location)

- i. no crime selected:

Task & interaction: A user can get a general view of the severe level (more number of crime cases, more severe) of all crime types across neighborhoods in all years (2015-2017).

The severe level is quantitative values displaying [NumPerNeighbourhood_all_yrs: 147,

67975] in different **saturation**s of color gray by default. The use of saturation will enable users to get a general idea of which area experiences the most crime cases and which area has less.

ii. select crime:

Task & interaction: If a user wants to see the severe level of a specific crime type across neighborhoods, they can achieve that by clicking on the arc of the crime type on the donut chart. Then the color on the map for neighborhoods will respond to the crime type corresponding to the crime color with different saturation. Again, the use of saturation will enable users to get a general idea of which area experiences the most crime cases of the selected crime type and which area has less.

iii. select neighborhood:

Task & interaction: When zoomed out, a user can select up to **three** neighborhoods to display the corresponding information in the line chart. Since the line chart displays the total number of crimes in years (2015-2017 by default), the user can compare the trends of these 3 neighborhoods easily. The limitation of 3 is to avoid too much line crosses also helps the user narrow down their selection (i.e., to limit their choice of neighborhoods that they would like to live in).

b. **zoom in:**

i. don't select a crime:

Task & interaction: If a user would like to see the detailed crime distribution within a specific neighborhood, they can zoom in by placing the mouse on that neighborhood, then scroll up by mouse or fingers. The crime cases with all types will show up within that neighborhood. This gives the user a general view of which specific region in that chosen neighborhood has more crime types, also a pattern of the distribution (i.e., a circular distribution may indicate there is a playground or parking lot nearby, then the user can choose to avoid that area if they have expensive cars).

ii. select a crime:

Task & interaction: After zoomed in, a user can choose **one** crime type at a time to further analyze the distribution of that particular type within a chosen neighborhood. For example, if a user drives an expensive car, they might want to avoid living in an area that has an intensive “vehicle collision” crime type.

2. donut chart

Derived Data:

a. Aggregated crime types

Aggregated crime type	Sub-crime type(s)
Theft	Theft from Vehicle, Other Theft, Theft of Vehicle, Theft of Bicycle
Mischief	(no sub-crime type here)
Break and Enter	Break and Enter Commercial, Break and Enter Residential/Other
Vehicle Collision or Pedestrian Struck	Vehicle Collision or Pedestrian Struck (with Injury), Vehicle Collision or Pedestrian Struck (with Fatality)

b. The percentage of each aggregated crime type to all crimes

Marks:

Interlocking area encode four general crime types

Channels:

- Angle encodes the number of a specific crime type
- Color hue (identity) encodes different types of aggregated crime type (categorical).

Task:

A user can see a rough part-to-whole relationship by the length of each segment to the whole circle. By selecting a segment in the donut chart, the application will filter the data with this aggregated crime type. For example, by selecting the Theft arc, a user can see the **distribution** of **Theft** in Oakridge from 2017-2021 from the choropleth, the **trend** of **Theft** in Oakridge from 2017-2021 from the line chart, and the **ranking** and **percentage** of **Theft** in Oakridge from 2017-2021 from the bar chart.

3. pie chart

Derived data:

The percentage of the sub-crime types to the corresponding aggregated crime type

Marks:

interlocking area encode sub crime type to the aggregated crime type

Channels:

Angle (2D area varies, magnitude) encodes the percentage (quantitative) of the sub crime type to the aggregated crime type. Each pie represents a sub-crime of an aggregated crime type. The area of each pie shows the occupation of this sub-crime to the aggregated crime type (e.g. 20 cases Break and Enter Commercial of a particular neighborhood in a specific year range, 80 cases Break and Enter Residential/Other of a particular neighborhood in a specific year range, then the pie chart to the Break and Enter donut arc shows two pies, one counts for 20% and the other one counts for 80%)

Task:

By selecting a slice in each pie chart, the application will filter the data with this particular sub crime type. For example, by selecting the slice **Theft from Vehicle** of the pie chart near the Theft arc, a user can see the **distribution** of **Theft from Vehicle** in Oakridge from 2017-2021 from the choropleth, the **trend** of **Theft from Vehicle** in Oakridge from 2017-2021 from the line chart, and the **ranking** and **percentage** of **Theft from Vehicle** in Oakridge from 2017-2021 from the bar chart.

Analysis for Other Views

4. line chart (x-axis: all the year, y-axis: number of cases of a crime / all types of crime)

Derived data: daily total crimes for a certain year range

Task: Line chart can easily convey the trend of the total crimes among certain. Therefore, when the user selects at most three neighborhoods in the choropleth map, he can compare the selected neighborhoods, comparing with each neighborhood and see whether the crimes are rising or not.

- a. By default, the entire Vancouver area will be displayed, therefore we see the trend of all crimes throughout the entire Vancouver area (1 line).
- b. If we select two neighborhoods in the map, we see the trends of those two neighborhoods (i.e. two lines) and if we click one of the crime types, the line chart will only show the trend of that specific crime type.

5. bar chart

Derived data: the total crimes for a selected data and get the counts each crime types in order to compute the percentage

Task: Bar chart uses 1d length which is on a common scale and all the bars are aligned and ordered by descending order, which makes users easier to see the most frequent crime types in the most recently clicked neighborhood.

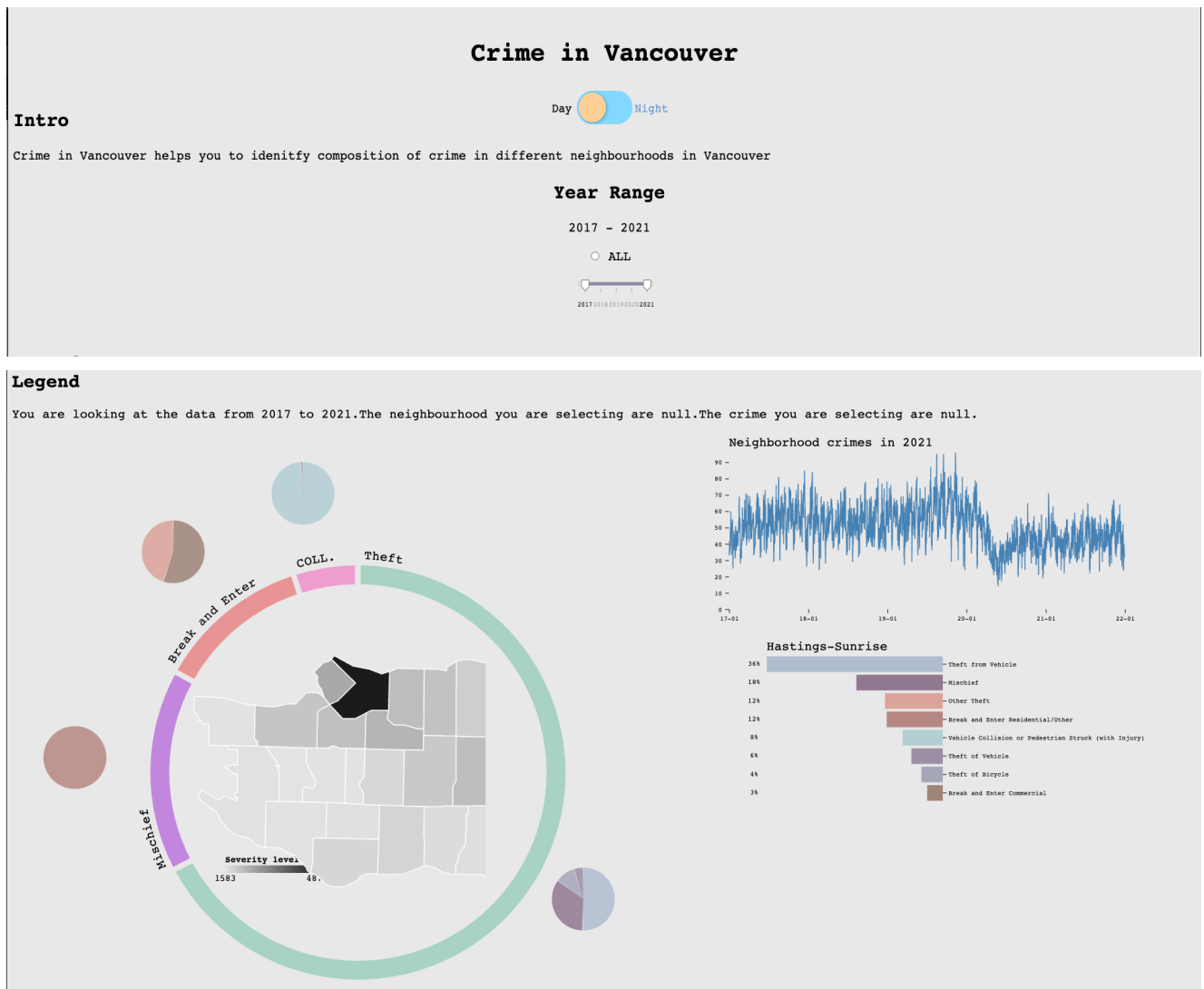
- a. when a neighborhood is selected: display and rank the percentage of all types of crime of the latest selected neighborhood
- b. when a neighborhood is not selected: it keeps the latest selected neighborhoods

Section 3: Changes Since Proposal

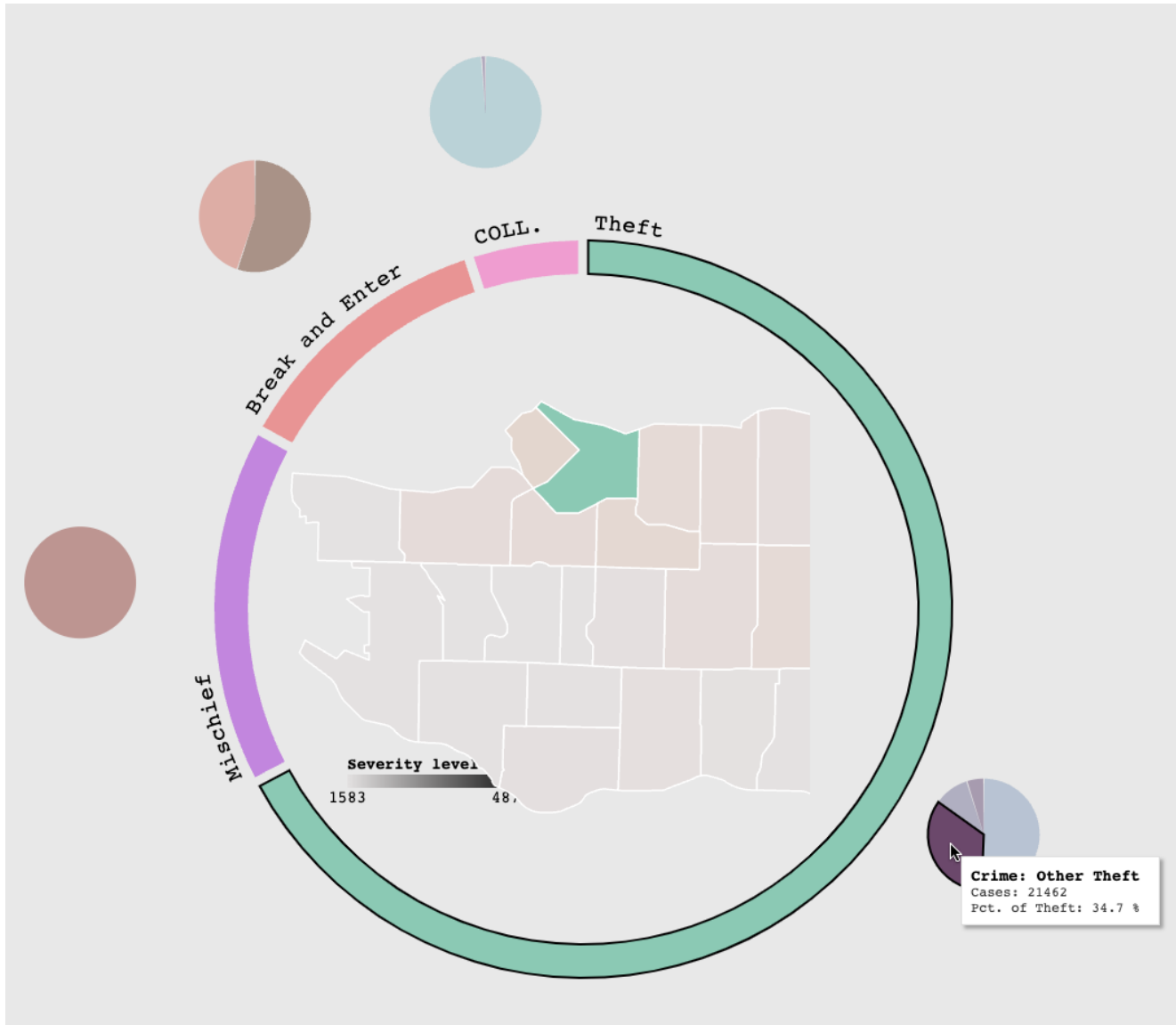
Our vision and scope did not change since the proposal. However, what we have modified so far is that instead of directly showing all crime types, now we aggregate those nine crime types into four big categories (theft, mischief, break & enter, vehicle collision & pedestrian strike). The reason behind this is that first we want to minimize the arc sections in the donut chart (since more arcs, more colors, less effective to perceive). Moreover, a user won't necessarily check a detailed theft type for example theft of bicycle or theft of vehicle at the very beginning, he/she is more likely to see the general total theft cases and then go inside the big category to figure out the detail sub theft cases. From that perspective, we want to display a parent-child relationship for crime types in our innovative view.

Section 4: Interface screenshots

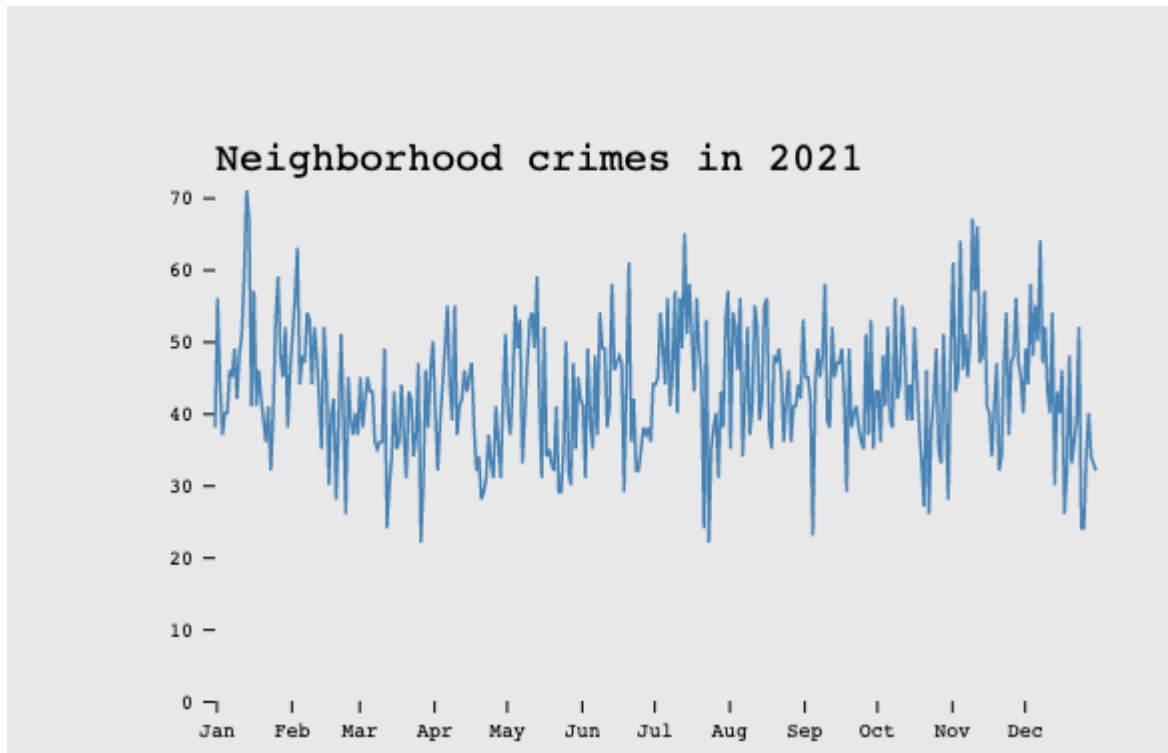
Overview:



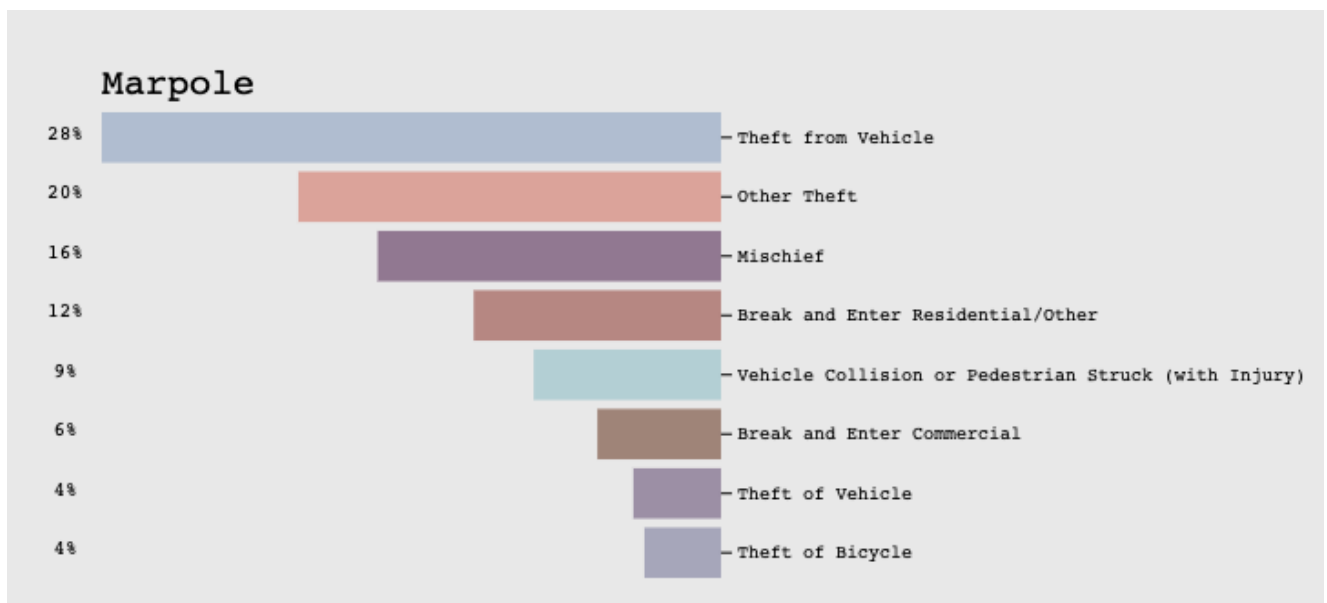
Innovative view:



Line Chart



Bar Chart



Section 5: Link to Data source

Link

latested updated on Mar 23, 2022:

[link for source from 2015 - 2020](#) + [link for source in 2021](#) -> [preprocessed final dataset link](#)

Section 6: Data preprocessing

Data Preprocessing Steps:

1. Combine two datasets, shrink the size of dataset from 2015 - 2021 to 2017 - 2021
2. Removed rows whose NEIGHBORHOOD is missing values (NaN)
3. Removed rows whose X and Y is 0.0 (which is obviously out of Vancouver region)
4. **[Transform: lat, long]** Convert X and Y, which are utm value to lat(latitude) and long (longitude)
5. **[Derived: TIME (day/night)]** Generate column TIME by the daytime (6:00-18:00, value: 'day') and nighttime (18:00-next day 6:00, value: night)
6. Removed the useless columns HUNDRED_BLOCK, X, Y, MINUTE
7. **[Aggregation: NumPerNeighbourhood_all_yrs]** A grouping by neighborhoods across 5 years will be generated to calculate the total number of crimes within each neighborhood.
8. **[Aggregation: NumPerNeighbourhood_[2017, 2021]]** A grouping by neighborhood and a certain year will be generated to calculate the total number of crimes within each neighborhood and in that particular year.

Section 7: Work breakdown, schedule / progress

PS: We highlight the extra work with **yellow highlight**

Task & Description		EST. comp. date	Actual comp. date	EST. time	Actual time	asgmt.
Write up documentation (M1)	Design overall structure Sketch charts Breakdown work	Mar 11	Mar 11	6 hours for everyone	6 hours for everyone	BW, MS, KZ
Set up project structure	Init project hierarchy and components' scripts Init git branch	Mar 13	Mar 11	0.5 hours for everyone	0.5 hours for everyone	BW, MS, KZ
Data Preprocessing	Adjusting dataset through the implementation	Mar 13	Mar 13	2 hours	5 hours	BW
Add ReadMe	add readme to the repo	Mar 13	Mar 15	1 hour	1 hour	KZ
Implement day & night toggle	Create initial static view of the day and night toggle, together with the text label	Mar 14	Mar 19	8.5 hours	8.5hour	KZ
	Filter data by day time and night time day time: (by default) 6 am - 6 pm night time: 6 pm - 6 am (next day)	Mar 15	Mar 19	6.5 hours	6.5hour	KZ
Implement	Create initial static	Mar 16	Mar 18	6 hours	3 hours	BW

year bar slider	view of “2015 - 2021 vertical scrolling bar”					
	Create initial static view of “all year” button	Mar 17	Mar 18	4.5 hours	2 hours	BW
	Add interaction to the bar and button i.e. when mouse click on, select the all years / a specific year	Mar 18	Mar 18	8 hours	6 hours	BW
Implement donut chart	Create initial static view of donut chart	Mar 18	Mar 22	2 hours	4 hours	BW
	Add interaction to the donut chart: select an aggregated crime type and pass out	Mar 18	Mar 22	4 hours	4 hours	BW
	Add typer effect to the name of the crime	Mar 18	Mar 22	3 hours	3 hours	BW
	Have the donut chart data listen to the year bar	Apr 1	Apr 1	1 hour	TBD	BW
Implement icon filter (each icon represents a type of crime)	Create initial static view of the icon filter (has been removed)	Mar 19		4.5 hours	4 hours	BW
	Add interaction to the icons: when mouse click on, select the crime type (has been removed)	Mar 19		6 hours	5 hours	BW
Implement piechart	Create initial static view of the pic charts	Mar 29	Mar 29	3 hours	3 hours	BW
	Add interaction to the pie chart: selecting a pie, selecting a sub crime type	Mar 29	Mar 29	3 hours	3 hours	BW
	Have the pie chart data listen to the year bar	Apr 1	Apr 1	1 hour	TBD	BW
Implement choropleth of zoom-out view	Create an initial static view with legend	Mar 25	Mar 20	8 hours	6 hours	MS
	Have the data items listen to the day & night toggle	Mar 25	Mar 25	4 hours	2 hours	MS
	Have the data items listen to the year bar	Mar 25	TBD	4 hours	TBD	MS

	(all year and a specific year) By default: all year					
	Link color hue of each region to the selected crime type	Mar 25	Mar 24	6 hours	2 hours	MS
	Link saturation of each region to the number of cases of a specific crime type	Mar 25	Mar 20	6 hours	3 hour	MS
	Add interaction: - when mouse hover on, show the tooltip of crime type introduction - allow user to select upto 3 neighborhoods and dispatch neighborhood list to other charts	Mar 25	Mar 25	6 hours	8 hours	MS
Implement choropleth of zoom-in view	Create an initial static view of the zoom-in view, show crime cases on a zoomed in neighborhood	Mar 25	Mar 24	8 hours	10 hours	MS
	Link the view changes to the mouse event - click to zoom into a neighborhood (changed) - scroll up/down to a specific neighborhood - prevent user click on region on zoom in - fix choropleth position upon zoom in	Mar 25	Mar 25	4 hours	12 hours	MS
	Have the cases related to this neighborhood listen to the year bar (a specific year) By default: 2021	Mar 25	TBD	5 hours	TBD	MS
	Have the cases related to selected neighborhood listen to the crime type filter	Mar 25	Mar 26	4 hours	5 hour	MS
Merge choropleth map into	positioning choropleth to the center of donut chart,	Mar 26	Mar 26	2 hour	1 hour	MS, BW, KZ

composite map	data linkage mering					
Implement line chart	Create an initial static view of the line chart	Mar 26	Mar 26	4.5 hours	5 hours	KZ
	Have x-axis listen to the the year bar (all year / a specific year), by default all year	Mar 26	Mar.29	4 hours	4 hours	KZ
	Have data items listen to the filtered crime type By default: none	Mar 26	Mar.29	4 hours	4 hours	KZ
	Have data items listen to the selected region(s) By default: all the regions	April 3		4.5 hours		
Implement bar chart	Create an initial static view of the bar chart	Mar 26	Mar 26	4.5 hours	4.5 hours	KZ
	Have data listen to year range	Mar 26	Mar 26	4 hours	4 hours	KZ
	Have data listen to the selected region	April 3				
Add legend	Add legend / tips to the webpage	April 5		4 hours		
	make text label interactive with the parameters (year, crime, neighborhood, day&night) in the chart	April 5		4 hours	TBD	BW / MS / KZ
Wrap up	Polish UI and harmonize colors	April 5		4 hours		
	Refactor and clean code	April 5		4 hours		BW: 6 hours MS: KZ
Total: Approximate 53 hours/person	BW	currently: 52.5(exclude the hours for the icon filter) hours				
	MS	currently: 49 hours				
	KZ	currently: 42 hours				

Appendix:

Updated data details:

Attr. name	Attr. type	Cardinality for categorical attr.	Range for quantitative attr.
CASE_ID	categorical	163178	
TYPE	categorical	9	
YEAR	categorical	5	
MONTH	categorical	12	
DAY	categorical	31	
HOUR	categorical	24	
TIME	categorical	2	
NEIGHBOURHOOD	categorical	24	
lat	quantitative		[49.201201, 49.313349]
long	quantitative		[-123.224021, -123.023393]
NumPerNeighbourhood_all_yrs	quantitative		[84, 48725]
NumPerNeighbourhood_2017	quantitative		[16, 9962]
NumPerNeighbourhood_2018	quantitative		[17, 10852]
NumPerNeighbourhood_2019	quantitative		[12, 12368]
NumPerNeighbourhood_2020	quantitative		[23, 7707]
NumPerNeighbourhood_2021	quantitative		[16, 7836]