November 6, 2017
CPSC 547 Project
CrimeVis: a tool for crime visualization in Vancouver

ALEXANDRA KIM, kimalser@cs.ubc.ca
AMON GE, amon@cs.ubc.ca

1 INTRODUCTION

The Vancouver Police Department (VPD) releases crime reports on a weekly basis. In order to discover various patterns in crime distribution and location, it is important to visualize this data for analysis by human experts or regular users who are interested in discovering trends. Both visualizations available at the time are limited in expressive power and granularity, limiting the analysis that could be done. We propose a new tool, CrimeVis, that shows a more detailed analysis of trends in the data, as well as providing flexible granularity for visualizing the data.

2 DATA

Our dataset comes from the city of Vancouver's open data catalogue, provided by the Vancouver Police Department [1].

As of Oct 28, the `.csv` file has 541273 rows ($\tilde{5}$0MB), where each row represents a single instance of crime. Statistics are not exact as the data is updated weekly on Sunday mornings. Categorical and numerical attributes are outlined in tables 1 and 2 respectively.

There is a temporal component to the data: when a crime is investigated, the time of the crime is reported up to minute precision. The spatial coordinates are given in UTM Zone 10.

At first sight, there can be some missing pieces of data with either empty values or zeros. This data is missing due for the privacy reasons, and mostly comes from crimes of the types "Offence Against a Person" and "Homicide". Neighbourhood, location and hour/minute attributes are missing for those types. As well, "X NK_LOC ST" is a default location value used for incidents with unknown location, and will need to be filtered out of the data.

3 PREVIOUS WORK

To the best of our knowledge, only two visualizations exist for crime data in the city of Vancouver. The first vis, GeoDash, is offered by the VPD [2]. The tool provides basic functionality of viewing the geographic distribution of crimes. It is only possible, however, to see locations of crimes occurred in the past week. Crime locations are denoted with icons, where the icon specifies the type of crime. The drawbacks of this vis implementation are many:

(1) it is only possible to see the data for the past week, meaning it is outright impossible to examine trends at a greater time scale;
(2) there is heavy occlusion due to bulky crime icons on the map;
(3) there is no aggregation of data;
(4) doing any kind of detailed analysis would require a lot of manual labour, since trends in the data are not shown and data is represented by individual points on the map.

The second visualization is provided by Rex Chang, 2015 [3]. It improves over the VPD visualization in a number of ways, notably enabling users to choose time period of interest and using color to encode the number of crimes per region, providing some aggregated results to assist a user with analysis. However:

(1) it is not possible to view data at a finer granularity than months;
(2) it is very difficult to compare geographical visualizations across different time periods, since only one map view can be shown at a time;
(3) the two plots showing crime rates changing over time aggregate all crime types and either show only one neighbourhood alone or sum over all neighbourhoods, losing the ability to compare trends between crime type and region over time.

In our project, we will improve on these two visualizations. With CrimeVis, we address two use cases that are difficult or impossible with the current visualizations.

4 USAGE SCENARIOS

4.1 Comparing geographical trends between two times

Given the dataset, it is natural for a user to explore trends in crime rates over time, particularly to compare two time periods with each other to find out if crimes rates have changed. What if the user is at the same time interested in seeing the geographical distribution of crimes? It is possible that as the city develops, geographical distribution of crimes may also change. Unfortunately, with the existing tools, it is not possible to do such kind of analysis.

4.2 Observing trends over time

Of course, users might also be interested in seeing trends over a continuous period of time. How did crime dynamics change over the past five years in my neighbourhood? What is the safest neighbourhood in town? What should I look out for most when I go to neighbourhood X? We anticipate that the questions of this kind would be as well of interest when exploring the crimes in the city. Temporal aspect of data would matter the most here and it would be interesting to see the dynamics of crime rates in different neighbourhoods over a period of time and how the rates and crime types compare to each other.

5 PROPOSED SOLUTION

5.1 Comparing temporal and geographical trends

To address the first scenario, consider first a heatmap that aggregates crime over a specified time period (see Fig. 1). The heatmap uses a sequential color gradient to represent crime *intensity*.

Name	Description	# of Categories	Example
TYPE	Type of crime	11	Mischief
HUNDRED_BLOCK	Offset location of crime	21321 + 1 blank	85XX STANLEY PARK DR
NEIGHBOURHOOD	Vancouver neighbourhood	24 + 1 blank	Mount Pleasant

Table 1. Categorical attributes

Name	Description	Min value	Max value	Median value
YEAR	Year of crime occurred	2003	2017	-
MONTH	Month of crime occurred	1	12	-
DAY	Day of crime occurred	1	31	-
HOUR	Hour of crime occurred	0	23	-
MINUTE	Minute of crime occurred	0	59	-
X	UTM Zone 10 X coordinate of the crime	0	511303	491505.5
Y	UTM Zone 10 Y coordinate of the crime	0	5512579	5456845.31

Table 2. Numerical attributes. Note that XY coordinates are (0,0) for crimes where location has been removed for privacy reasons.

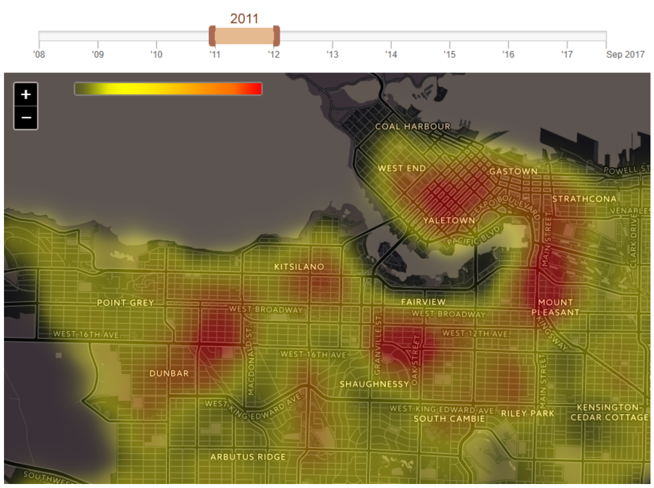

Fig. 1. Mockup. A heatmap of crimes during the year of 2011.

To derive the intensity of a crime at each geographical coordinate, each incidence of crime during the specified time period contributes some intensity around its location, to account for the inaccuracy of the data (the VPD offset geographical coordinates to desensitize the data), as well as to spread the influence of a single crime to the region surrounding it.

Ideas for mapping a single crime incidence to intensity are:

(1) a constant increase within some radius of the crime's location, accounting for the data's offset
(2) a gaussian centered around the crime's location
(3) a gaussian centered around the crime's location with some radial cutoff, and/or cutoff by neighbourhood boundaries

The specific time and time period of interest can be changed by dragging the time slider. The length of time to aggregate crimes over is reflected by the width of the selected time period in the slider (see Fig. 2).

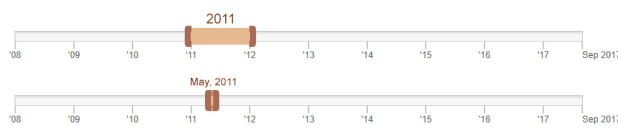

Fig. 2. Mockup. The time slider allows the user to examine crimes at a specific time, over a specific time range. The top slider aggregates crimes over the year of 2011, while the bottom slider only aggregates crimes in May of 2011.

The problem with this heatmap is that crime *intensity*, our derived value, is not an absolute quantity, since one particular crime spreads its influence spatially. Rather than displaying these derived heatmaps directly, we focus on using these heatmaps to make binary comparisons between two time periods. For instance, we could compare the crime heatmap for the years 2011 and 2015, and use the two derived intensities to calculate the percentage change in crime from 2011 to 2015 (see Fig. 3). A diverging color gradient is used instead to represent increases and decreases in crime at a particular geographical location. This not only addresses the relative nature of the derived crime intensities, but also normalizes these values. It is not unexpected to find high crime rates downtown, for instance, since there is a higher population density downtown. By representing the *percentage change* in crime at each location, we essentially normalize the absolute crime intensity at each location at the current time period by the crime intensities at the earlier time period to which we compare to.

To summarize, our heatmap shows the percentage increase / decrease in crime between two time periods. Interactive sliders allow flexibility in exploring the data: one for adjusting the granularity of time periods (e.g. week(s), month(s), etc.) and the other for finding the two periods to compare.

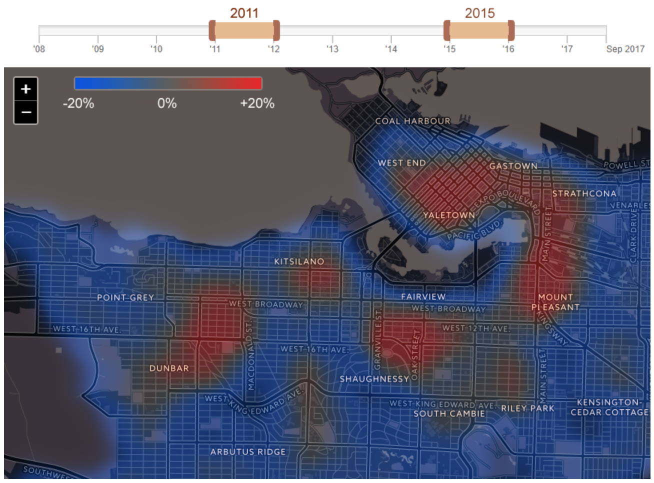

Fig. 3. Mockup. Comparison heatmap between crimes in the year 2011 to the year 2015. Red colors represent an increase in crime, while blues represent a decrease in crime.

This binary comparison heatmap can be extended to other temporal aggregations that don't directly translate on the time slider, such as comparing overall day / night trends or summer / winter trends. Further extensions could involve filtering by specific crime type with checkboxes (rather than aggregating over all crimes). In general, we find this solution to be a useful visualization that presents more granular geographical data to make binary comparisons across flexible times.

5.2 Observing trends over time

It is not very practical to show geographical data for multiple time frames. Animations were considered, but in the end we opted for an approach that could represent aggregated trends on one static screen. We dismiss geographical information and aggregate the number of crimes per region per type of crime in a dotplot (see Fig. 4). Here, blocks of time lie on the x-axis; neighbourhoods of the city lie on the y-axis; and the colour channel encodes crime type. Each dot represents a fixed number of crime occurrences, so that the number of dots in each cell reflects the number of crimes of a particular type committed in the corresponding region.

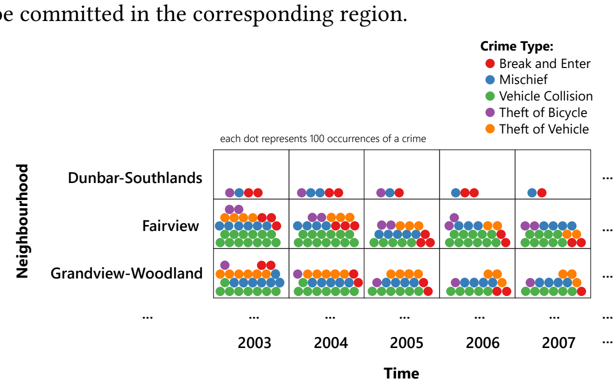

Fig. 4. Mockup. Dotplot of crimes over time, where each dot represents one hundred occurrences of a crime. Neighbourhoods are segregated on the y-axis, and time periods on the x-axis. Colour encodes the type of crime.

This visualization is meant to supplement the weaknesses of the heatmap proposed in Section 5.1 - namely, observing trends over time and across multiple crime types, as well as address limitations in existing work. Similar to the time sliders in the previous section, we allow the user to select the time period of interest; our dotplot should automatically quantize the horizontal axis into a suitable number of bins (e.g. each bin represents minutes / hours / days / weeks / months / years). This requires a balance between keeping the time granularity fine-grained while ensuring a dense enough dotplot in bins to maximize screen space and be useful in pattern recognition. Alternatively, as an extension, the user can manually adjust the time granularity within reason.

This dotplot uses density as a channel to encode the crime count, which is easily scalable by dynamically altering the number of crimes a single dot represents, as well as the size of each dot. This way, we can always maximize the use of screen space. Since we have 11 crime types, colour is a suitable encoding at this scale. Filtering by crime type is again a useful extension; hovering over a particular coloured dot can fade dots of other colours to grey, for instance (see Fig. 5).

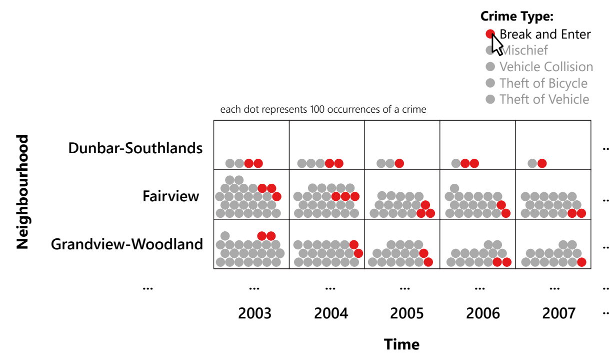

Fig. 5. Mockup. Hovering over the Break and Enter dot fades other colours to grey.

6 IMPLEMENTATION

CrimeVis will be designed as a web tool, available for use in modern browsers. Therefore, we are going to use such web technologies, as HTML, CSS and Javascript and their libraries. The visualization itself is going to heavily rely on Javascript's D3 library [4]. The use of open map solutions, such as Google Maps[5] or Leaflet[6], is as well possible.

7 MILESTONES

To ensure a timely implementation of the project we created a schedule with main milestones and implementation dates breakdown. Please refer to Table 3 on the following page.

Since both of the team members possess symmetric skillsets, the hours in the table are combined and two of us are expected to work an equal amount of hours on each task.

REFERENCES

[1] Open Data. Data catalogue: Crime. http://data.vancouver.ca/datacatalogue/crime-data.htm. Accessed: 2017-11-06.

Table 3. Schedule and milestones

Task	Hours	Deadline	Description
Pitch	10	October 17	Explore datasets, come up with idea, make slides, rehearse
Project meeting	4	October 26	Brainstorm ideas, questions, explore existing vis, meet with Tamara
Proposal	14	November 6	Brainstorm visualizations, write proposal
Implementation	70	December 10	Implementing the vis tool
- d3 basics	10	November 8	- learning how to use d3
- heatmap	25	November 21	- implementing heatmap showing %-change (section 5.1)
- time plot	25	December 5	- implementing timeplot showing trends over time(section 5.2)
- web design	10	December 10	- putting the two plots together, working on UI
Peer reviews 1	6	November 21	Prepare slides and mockup
Peer reviews 2	4	December 6	Prepare slides and demo
Final presentation	8	December 12	Make slides, rehearse
Final paper	24	December 15	Final detailed project writeup

[2] Vancouver Police Department. Geodash. http://geodash.vpd.ca/. Accessed: 2017-11-04.
[3] Rex Chang. Vancouver crime map visualization. http://rexchang.com/vancouver-crimemap/. Accessed: 2017-11-04.
[4] Mike Bostock. D3.js - data-driven documents. https://d3js.org/. Accessed: 2017-11-06.
[5] Google Maps. Google maps javascript api | google developers. https://developers.google.com/maps/documentation/javascript/. Accessed: 2017-11-05.
[6] Vladimir Agafonkin. Leaflet - a javascript library for interactive maps. http://leafletjs.com/. Accessed: 2017-11-05.