Vancouver Crime Visualization

CPSC 547 Project Update, 2015W1 November 23, 2015

Rex Chang (rexchang@cs.ubc.ca)

1 Project Update

1.1 Data Collection and Cleanup

The crime count data had been acquired from Vancouver Police Department (VPD). Since the October 2015 counts had been released after the proposal submission date, they had been included in the dataset (as opposed to just up to September 2015). Both neighbourhood-specific counts and crime-specific counts had been acquired in their raw forms, which were compiled reports in the PDF format. The PDF files had been semi-manually converted into the CSV format using Tabula (https://github.com/tabulapdf/tabula).

The neighbourhood property incident data had also been acquired from Vancouver Open Catalogue (VOC). Upon closer inspection of the incident data, I found out that while most addresses were in the hundredth form as noted in the proposal, there was a subset of the locations reported in the alternative format of street intersections (for example, BEATTY ST / W GEORGIA ST). This format does not batch-geocode properly like the normal addresses. The Google Maps API can handle intersection formats, but the request will need to be reformatted. In addition, most geocoding services I can find impose usage limits, making it difficult to properly geocode all data available from VOC. Google Maps API allows only 2,500 queries daily, and DataBC geocoder (http://www.data.gov.bc.ca/dbc/geographic/locate/geocoding.page) allows only 1,000 queries per request. More queries can be requested for a price. Given the amount of incidents (close to 20,000 incidents in the year of 2014, for example), it is difficult to process all the data as originally proposed. Since the property incident visualization was originally planned to be implemented last, efforts will be made to focus on the crime count data in the Geographical View first.

1.2 Geographical View

The geographical view requires neighbourhood boundary data to display. The data in KML/SHP form are available and acquired from VOC, and converted to GeoJSON format for implementation. Originally a more detailed geographical map view implementation was explored using Leaflet (with the OpenStreetMap tile layer), with the intention to offer a smoother viewing transition from the Geographical View to the Property Incident View. However, since the incidents are property-specific and not covering all crime types, it makes less sense to build both views on the same map detail as required to display property incidents. It is more efficient for the crime count data to be visualized as a choropleth map. The neighbourhood crime counts had been partially visualized accordingly using D3.js. The current implementation has the crime type and date rate manually coded, and the user selection is to be implemented, as planned in the proposal.

The crime-specific data contain additional crime counts not covered in the neighbourhood-specific data, as noted in the proposal. The additional crime counts can only be properly included in the city crime count visualization. Currently I am exploring whether including these additional information in the Geographical View as proposed is appropriate. The viewer may be confused with additional crime types in the city-specific crime counts.

2 RELATED WORK

There have been many applications of visualizations in the field of criminology, mostly with the focuses on forensics and crime analysis. While the focus of this project is to inform the general public and does not largely overlap those of the related works, numerous visualization techniques had been employed and studied for their appropriate uses in this area, and the analysis and demonstrated uses in those related works are useful in the visualization design in this project.

2.1 ACADEMIC WORKS

M. Andreson conducted case studies specific to the crime statistics in Vancouver [1, 2]. He studied the spatial dimension of automotive theft, breaking and entering, and violent crimes in 1996 to synthesize and assess the two common spatial theories of crime: the social disorganization theory, and the routine activity theory. The level of details required is defined at the Vancouver neighbourhood level, so the granularity of the data analyzed is similar to that of this project. However, the data used were from the VPD CFS (Call-for-service) database, which included specific locations and times of the alleged crime activities. Consequently, the reported counts might not represent the actual criminal activity data. These numbers were further processed as crime rates per census tract, per 1,000 persons, per 1,000 ambient population in the different neighbourhoods, and displayed spatially as choropleth maps for the 3 different crime types. The derived data and the visualizations allowed the author to discuss the validity of the theories in terms of actual demographic and geographical information of the criminal activity.

The TASC (Textual Analysis of Similar Crimes) system [3] utilized natural language processing to derive data from raw investigation reports, and the data were further visualized to support crime investigation analysis. There was a great detail of discussion of the effectiveness of visualization and interaction techniques in the graphical representation of crime data, including focus+context, details-on-demand, and brushing-and-linking. While the TASC system focused on textual visualization, the case studies in this paper suggested the potential usefulness in comparing similarities and differences in crime data, which is what this project hopes to enable the users to do for the Vancouver crime statistics.

In the web-based crime analytics system of newspaper articles [4], similar data mining and machine learning techniques as TASC were employed to assist crime investigations in Sri Lanka. Unlike TASC, the auto analyzer further derived 3 sets of data for visualization: (a) hot spots of crime activity represented as neighbourhood district regions with color encoding indicating the frequency values on a choropleth map; (b) crime comparison of different types of crime as a pie chart; and (c) crime patterns as the time-series plot to represent changes in frequency of crime types. The system allowed user selection of crime types and the years to display in the graphs. The graph and diagram representations made the system simple to use for crime analysis processes used by either law enforcement officers or any other interested users, including the general public.

2.2 OTHER WORKS

The Vancouver Police Department periodically releases crime statistical reports offering additional visualizations to support their summaries. Notably, in the report released in 2014 [5], various bar charts and line charts of city crime counts were employed to illustrate the alleged declining crime rate in Vancouver. VPD also generated “Crime Maps” and made them available online [6], which showed weekly aggregated counts of property incidents on a geographical view of Vancouver neighbourhoods. These crime maps were static graphs and replaced when new weekly data were available.

Thematic maps, such as choropleth maps, were generally employed for crime visualizations for the masses. The San Francisco Crime Map [7] displayed crime counts of different neighbourhood blocks using colors, with different saturation levels indicating different crime frequencies. The data allowed the author to generate interesting findings in the analysis, such as that certain areas did not have higher crime rates as previously perceived, or that certain areas had higher auto theft rates so cars should be parked elsewhere. It is with the same intention that the Vancouver Crime mapping can be used to derive these findings by the general public. While the San Francisco Crime Map was a static graph, the London Crime Mapping [8] offered geographical view of London with zoom-in and zoom-out interactions to enable different levels of details. In addition, each region focused and clicked on generated linked pop-up views of additional comparison info, such as changes in the crime rates since last month, or crime rate trends comparing to the previous year. These interaction techniques appeared to provide a more seamless transition to display comparison data as opposed to a separate comparison view.

Oakland Crimespotting [9] and San Francisco Crimespotting [10], both by Stamen Design, visualized crime statistics in the corresponding regions with much granular controls and details. Crime incidents were marked on the map on the street level with different colors representing different types, and the times of the incidents were included. The system leveraged this detailed level of raw data to give users controls over the time ranges of the day to display the incidents. In addition, a constant overlay of incident count histograms was displayed, acting as a visualization of information itself and also as a control over the date range of display. CrimeReports [11] implemented a similar visualization solution, with emphasis on providing more details for each incident. The additional information could be accessed in the incident pop-up, and the system also allowed filtering of crime incidents based on user selections of date ranges and crime types. While the Vancouver crime data publicly available from VPD are not of the level of detail, the clever uses of visualization in these projects were considered for the design of this project.

3 References

1. Andresen, M. A. "Crime measures and the spatial analysis of criminal activity." British Journal of Criminology 46.2 (2006): 258-285.
2. Andresen, M. A. “A spatial analysis of crime in Vancouver, British Columbia: a synthesis of social disorganization and routine activity theory.” The Canadian Geographer / Le Géographe canadien, 50 (2006): 487–502.
3. Chih-Hao Ku; Nguyen, J.H.; Leroy, G., "TASC - Crime report visualization for investigative analysis: A case study," in Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on , vol., no., pp.466-473, 8-10 Aug. 2012
4. Jayaweera, I.; Sajeewa, C.; Liyanage, S.; Wijewardane, T.; Perera, I.; Wijayasiri, A., "Crime analytics: Analysis of crimes through newspaper articles," in Moratuwa Engineering Research Conference (MERCon), 2015 , vol., no., pp.277-282, 7-8 April 2015
5. Vancouver Police Department. (2014, Feb.). *Vancouver Crime Statistics – 2013 Key Performance Indicators* [Online]. Available: http://vancouver.ca/police/assets/pdf/reports-policies/2013-year-end-kpi-report.pdf. Last accessed 2015-11-22.
6. Vancouver Police Department. *Vancouver Crime Maps* [Online]. Available: http://vancouver.ca/police/CrimeMaps/. Last accessed 2015-11-22.
7. L. Kesteloot. (2003, Oct.). *San Francisco Crime Map* [Online]. Available: http://www.teamten.com/lawrence/projects/sfcrime/. Last accessed 2015-11-22.
8. Metropolitan Police of London. *London Crime Mapping* [Online]. Available: http://maps.met.police.uk/. Last accessed 2015-11-22.

9. Stamen Design. *Oakland Crimespotting* [Online]. Available: http://oakland.crimespotting.org/. Last accessed 2015-11-22.
10. Stamen Design. *San Francisco Crimespotting* [Online]. Available: http://sanfrancisco.crimespotting.org/. Last accessed 2015-11-22.
11. Public Engines. *CrimeReports* [Online]. Available: https://www.crimereports.com/. Last accessed 2015-11-22.