Patterns of Bicycle Theft in Vancouver

Ornwipa Thamsuwan

Abstract

This study investigated the patterns of bicycle thefts based on time and location in Vancouver during 2003-2017. Data of crimes came from ‘Vancouver Open Data Catalogue’. Exploratory data analysis was conducted to identify time trends and seasonality of the incidents, and to compare the neighborhoods where incidents occurred. Then, using geographical coordinates as features, the incidents were grouped together as seven clusters and presented on a map. The findings could provide information that aids the investigation of crime.

Motivation

Although bicycle theft may be viewed as minor property crime, it can indeed affect the quality of life of the residents because bicycle is a primary mean of commute for some residents in several cities.

The goal of this project was to unfold any underlying patterns within the time and location of bicycle theft.

The city of Vancouver was used as it had relatively good completeness.

Dataset

Data of crimes in Vancouver (Canada) from 2003 to 2017

Data was downloaded from https://www.kaggle.com/wosaku/crime-in-vancouver.

Dataset contains types of crime, year, month, day, hour, minute, neighbourhood, street block, latitude and longitude of the incidents.

There are 530,652 records from January 2003 to July 2017.

After (or if) dropping NA values, there are 474,015 data points.

Data Preparation and Cleaning

Dataset was completed (no missing records) for the fields: type of crime, year, month, day, and x-y coordinates, which was converted to latitudes and longitudes.

The majority of missing data are the fields: neighbourhood, hour and minute.

Because the dataset was already very large, records with NA was dropped.

Research Questions

1. Is there any underlying pattern of bicycle thefts in relation to neighbourhood and seasonality in 2003-2017 in Vancouver?

2. Can we identify geographical clusters, i.e. street blocks, to group the bicycle theft incidents based on their coordinates (latitudes and longitudes)?

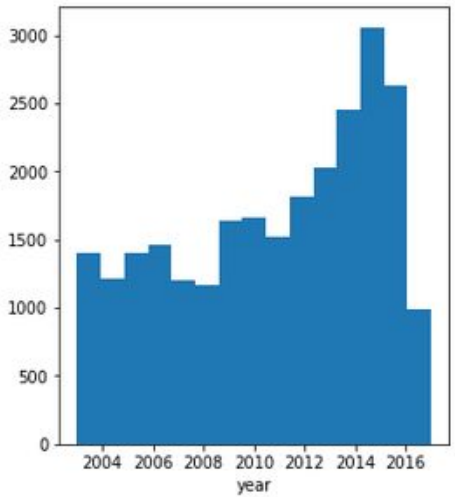
Methods

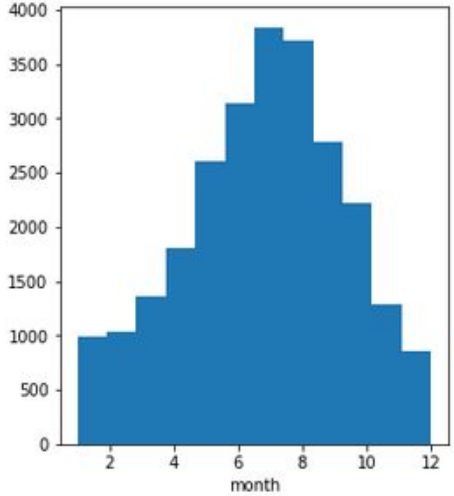
1. For historical patterns…
    - Histogram for patterns in year, month, day and time
    - Bar plots for patterns in neighbourhood

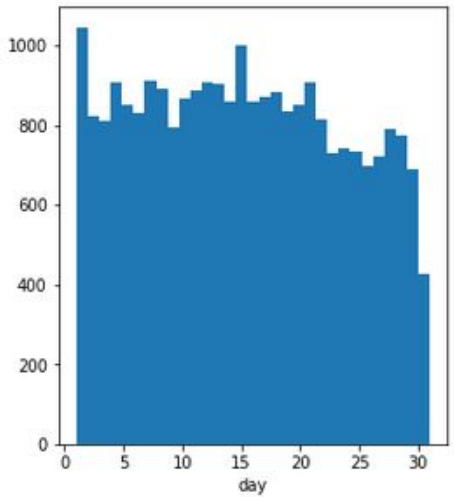
2. For historical hotspots...
    - Mapping for visualization the street blocks of the incidents
    - K-mean clustering with scaled latitude and scaled longitude

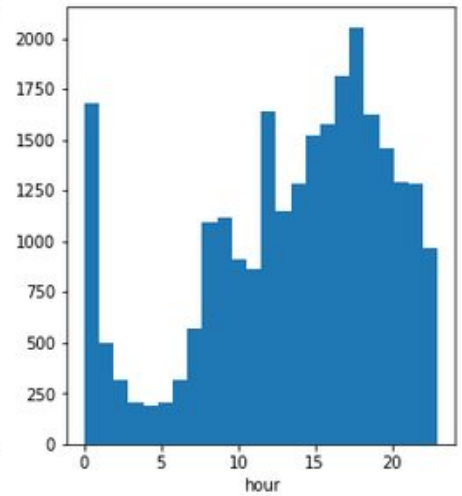
Findings

The number of incidents increased over the years, were particularly high in summer, occurred more on the 1st and 15th day of the month, and more in both afternoon than morning. Note that it is unclear whether the records on midnight could in fact be an unknown time during the night. This study, thus, focused only on daytime bicycle theft.

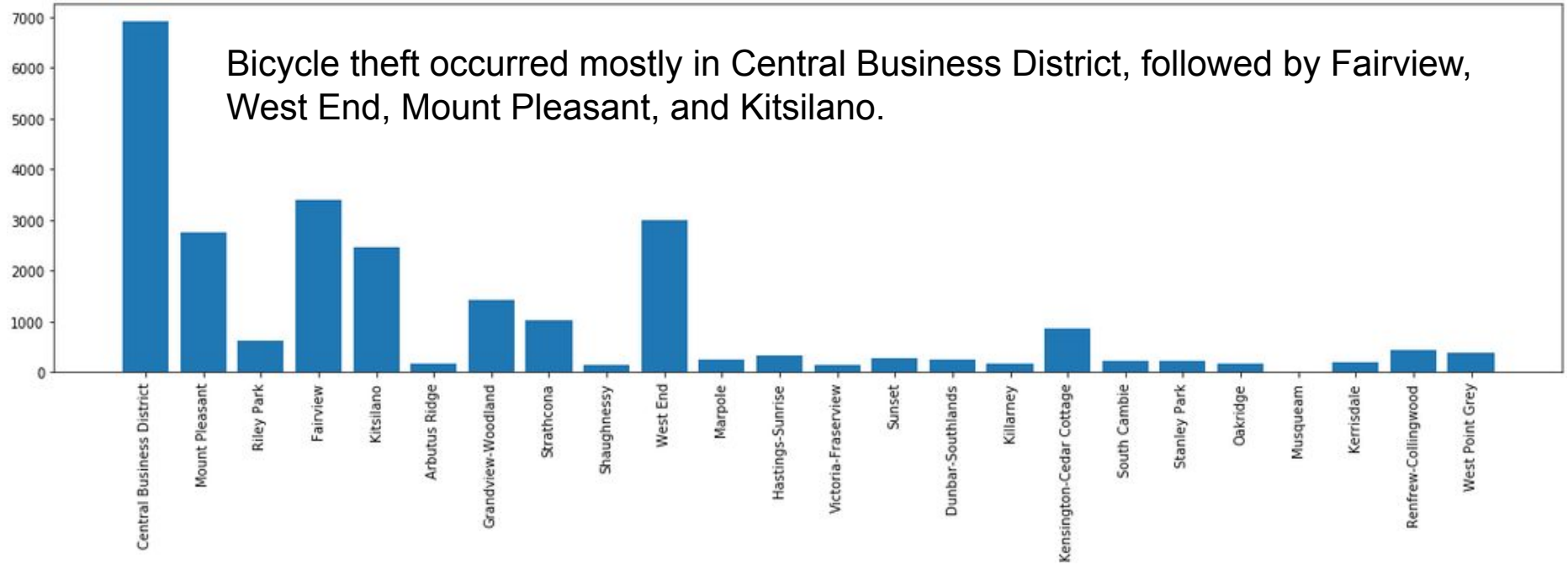








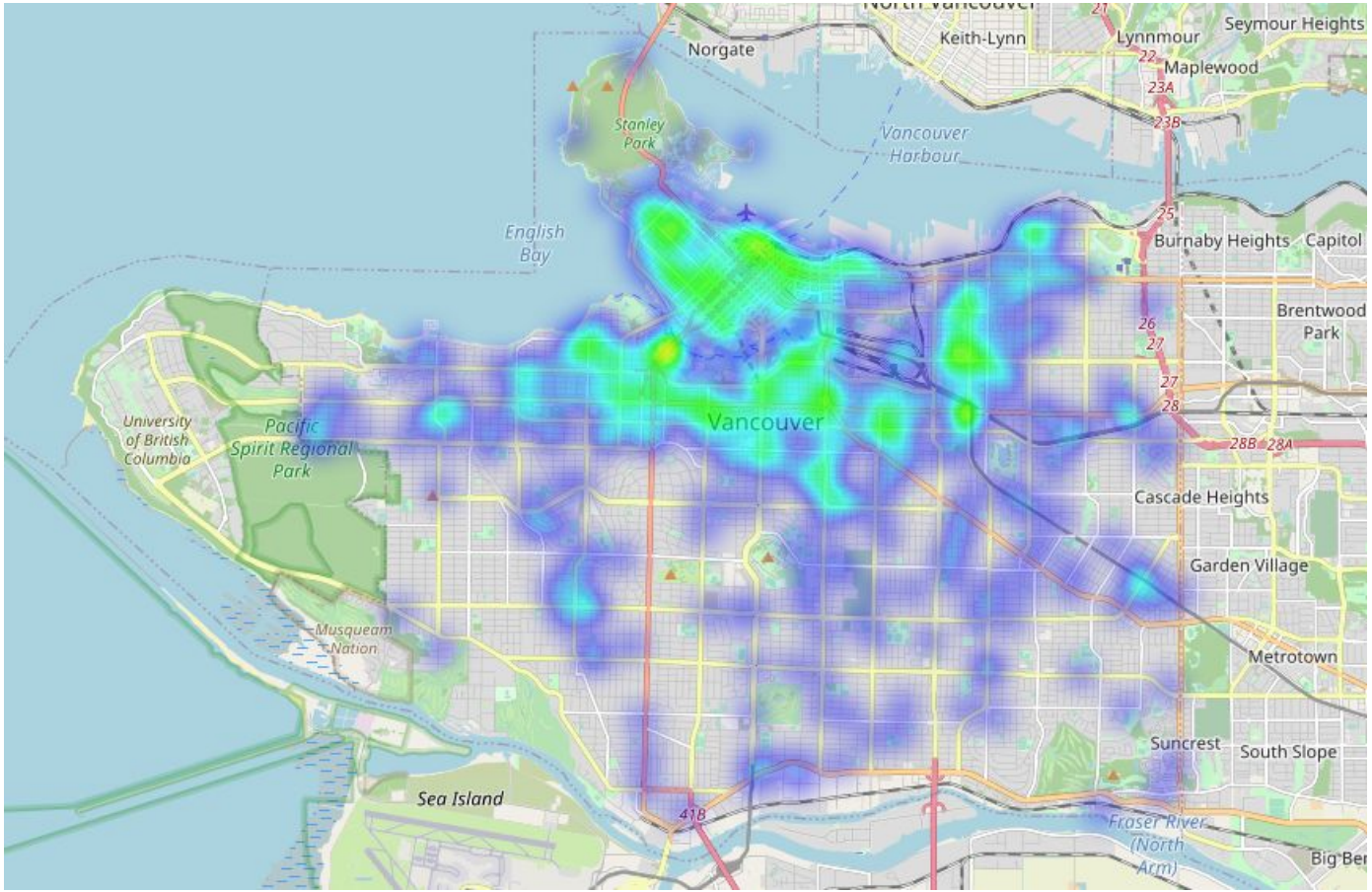
Findings


Bicycle theft occurred mostly in Central Business District, followed by Fairview, West End, Mount Pleasant, and Kitsilano.
7000
6000
5000
4000
3000
2000
1000
0
Central Business District
Mount Pleasant
Riley Park
Fairview
Kitsilano
Arbutus Ridge
Grandview-Woodland
Strathcona
Shaughnessy
West End
Marpole
Hastings-Sunrise
Victoria-Fraserview
Sunset
Dunbar-Southlands
Killarney
Kensington-Cedar Cottage
South Cambie
Stanley Park
Oakridge
Musqueam
Kerrisdale
Renfrew-Collingwood
West Point Grey

Findings

Heatmap showed higher incidents in the same areas as did in the bar plots.

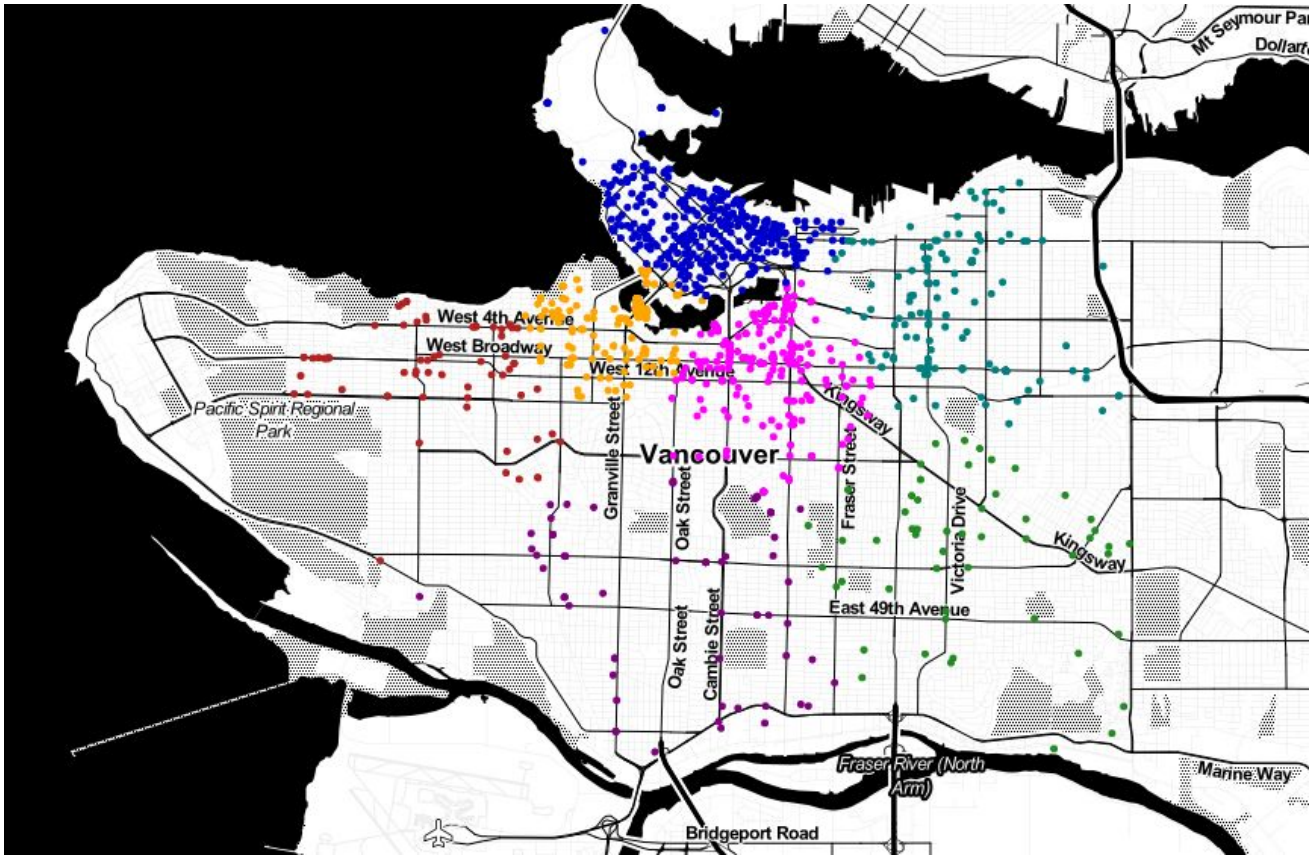
It also appeared there are street blocks of concern.



Findings

Seven clusters were presented to group the area of incident together to help the police decide to allocate their resource, i.e. investigation.

* shown on the map - only 2016



Limitations

Data

The accuracy of the data is unknown but rather to the best knowledge because it was self-reported. The report of thefts is assumed to be retrospective; therefore, the exact time, i.e. hour, is often unknown.

Methods

This analysis was only the discovery of patterns in historical data. If it is used for prediction, cautions need to be made as other interventions of bicycle thefts may be in placed in the future. The choice of k, i.e. the number of cluster, may not be optimized.

Conclusions

1. There were patterns of bicycle thefts in terms of location and time of incidents in 2003-2017.
    - Increasing trend over the years
    - Seasonality with months, days and hours
    - Differences across neighbourhoods

2. We could identify clusters of the bicycle theft incidents based on coordinates.
    - Heatmap suggested certain street blocks and intersections as hotspot.
    - K-mean clustering could group incidents together to further investigate the crime.

Acknowledgements

No other informal analysis was used.

No one provided feedback.

References

Mapping resource: https://www.kaggle.com/alexisbcook/interactive-maps