Assignment 1 Spatial Data Handling
Huanye Liu

Data sources

1) The main dataset crime.csv is on the crime incidents in Vancouver, Canada, and it can be downloaded from https://www.kaggle.com/wosaku/crime-in-vancouver/data. The dataset consists of records of crime in Vancouver from 2003 to 2017(to date) with such attributes as crime type, time and neighborhood, projection coordinates X and Y, longitude and altitude. This dataset is used here to aggregate the number of crime incidents for each neighborhood in year 2017(to date).

2) The total number of crime incidents for each neighborhood is so-called the spatially extensive data, and it can be converted to the spatially intensive data, the area density data, by being divided by the area of each neighborhood. I manually entered the neighborhood area information published on the website http://vancouver.ca/news-calendar/areas-of-the-city.aspx, and saved it as vc_area.csv.

3) The third dataset, Vancouver neighborhood boundary shape files including one .dbf file, one .prj file, one .shp file and one .shx file, can be downloaded from http://data.vancouver.ca/datacatalogue/localAreaBoundary.htm. This dataset is used to plot the choropleth map.

Operations

a) Data preprocessing using R

The slight differences between the neighborhood naming system used in crime.csv and the one used in Vancouver neighborhood boundary shape file are resolved using the mutate() and str_replace() functions offered by the R tidyverse library. In addition, a few records with NAs in variable neighborhood are deleted using filter(). The dataset after the cleaning is saved as vc_crime.csv.

b) Select records from year 2017 (to date) and create the point map

The selection can be done by using the Selection Tool function provided by GeoDa after loading vs_crime.csv, and output file is saved as vs_crime_2017.csv. Then the point map can be created using the Points from Table function provided by GeoDa and is saved as vc_crime_2017_point_map.png, which is attached at the end of this report.

c) Aggregate by neighborhood

The aggregation of the number of crime incidents in each neighborhood can be done using the Aggregate function provided by GeoDa after loading vs_crime_2017.csv, and the output file is saved as vc_crime_2017_count.dbf.

d) Merge the crime count with the area by neighborhood

The merging can be done by using the Merge function provided by GeoDa after loading vc_crime_2017_count.dbf, and the output file is saved as vc_crime_2017_area.dbf.

e) Calculate the crime count per hectare

The calculation can be done using the Calculator function provided by GeoDa after loading vc_crime_2017_area.dbf, and the resulting dataset is saved as vc_crime_per_hectare.dbf.

f) Merge the crime count per hectare with boundary files

The merging can be done by using the Merge function provided by GeoDa after loading local_area_boundary.dbf, and the output file is save as vc_crime_density_2017.shp.
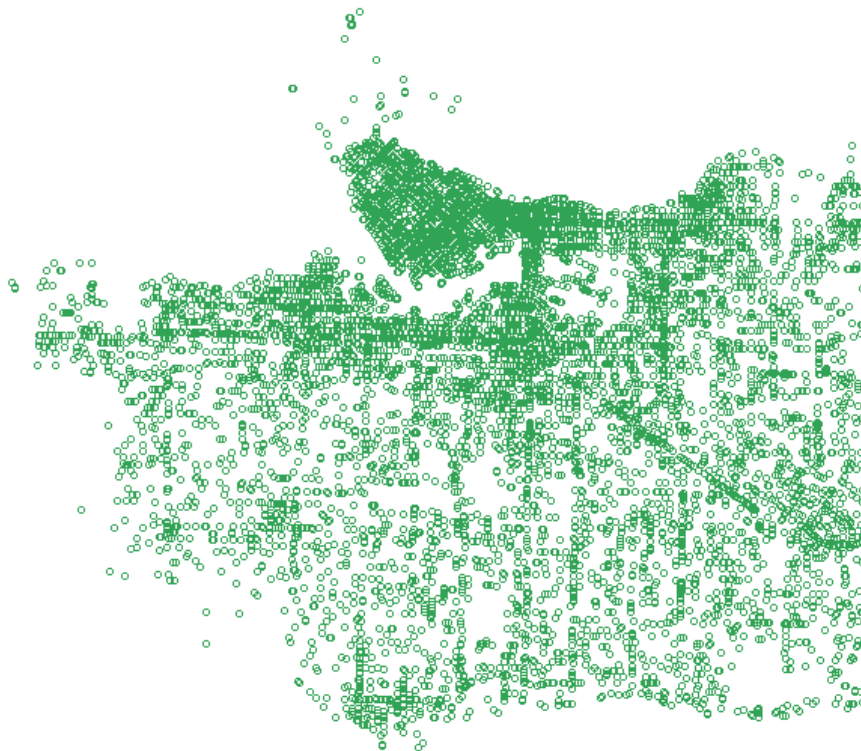
g) Plot the choropleth map

This can be done by using Map function provided by GeoDa after loading vc_crime_density_2017.shp, and the resulting choropleth map is saved as vc_crime_density_2017_chropleth.png.
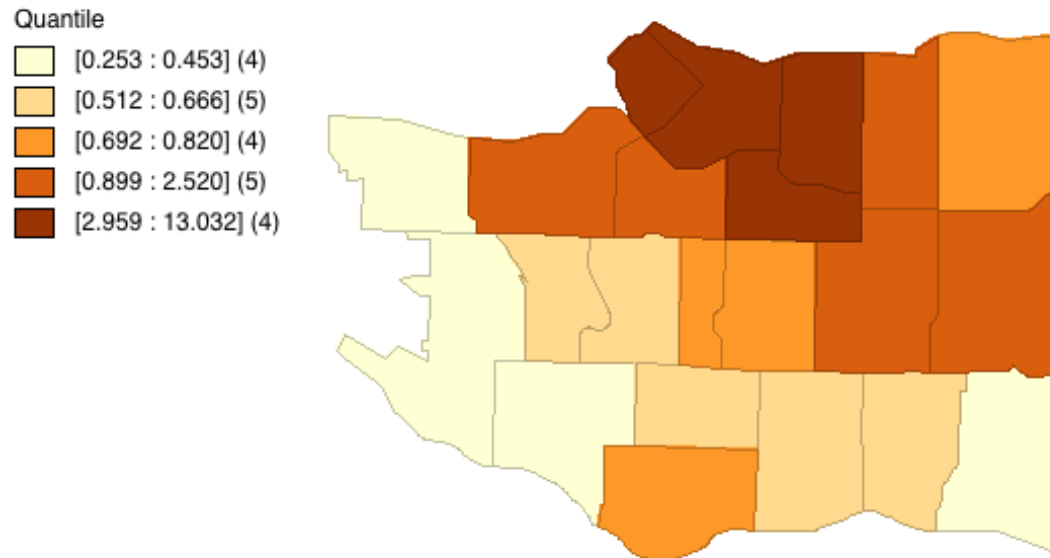
Maps

The point map is shown as below:

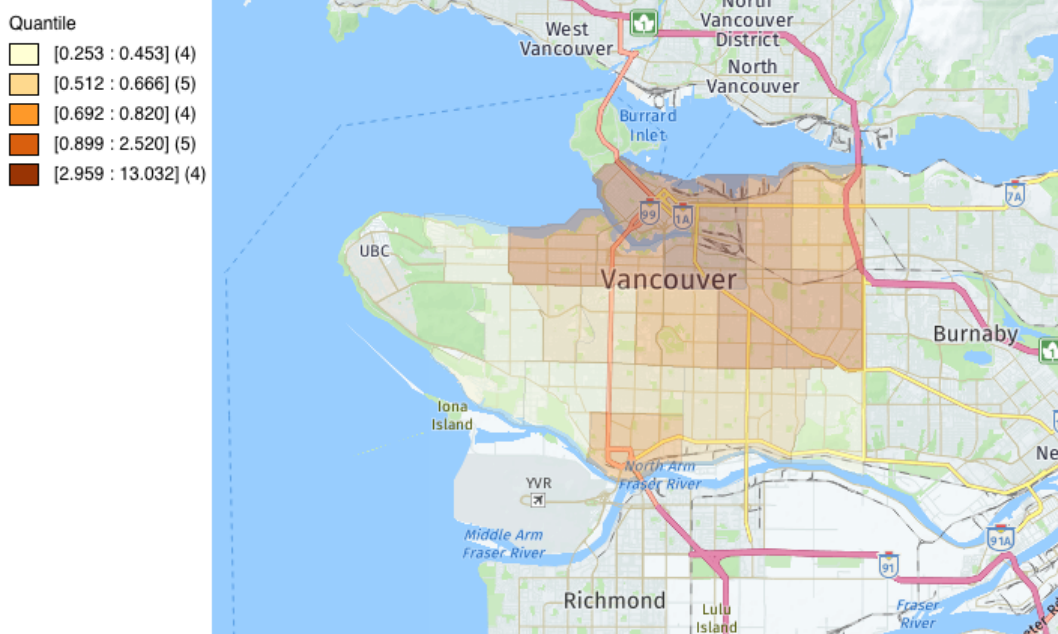

Map - vc_crime_2017
(16205)

The choropleth map is shown as below:

and the choropleth map with base layer is shown as below:

From the choropleth maps, we can see there exists an obvious pattern: the crime count per hectare is higher in the northeast part of Vancouver than the southwest part. One reason may be that those neighborhoods in the northeast are the most populated area with highly developed transportation system, and from the base map we can find that those neighborhoods are also closer to North Vancouver, which may imply higher levels of population mobility and thus higher crime rates.