Assignment 2: Mapping Parking Violations in NYC

Thomas Brambor

2019-03-10

Parking Violations in NYC

Data

For this assignment, we are going to investigate data on parking violations in NYC.

Parking violations in 2019

NYC Open Data has data on all parking violations issued in NYC since 2014. The updated dataset provided for 2019 currently includes 6.5 million observations (and about 1.2GB of data). To make the assignment manageable, I have reduced it to a subset of tickets issued in 2019 and by Manhattan precincts only, yielding about 284k tickets.

Two support files are also included in the `parking` sub folder:

- the **descriptions of all variables**
- the **dictionary of violation codes**

Police Precincts

A second data source is the shape files of police precincts in NYC.

Exercise

1. Data exploration

Before focusing on the spatial part of the data, let's explore the basic patterns in the data.

a) Violation Code and Fine Amounts

Add the violation code descriptions and fine amounts to the data file. Provide a visual overview of the top 10 most common types of violations (feel free to group them into categories if reasonable). Compare how this ranking differs if we focus on the total amount of revenue generated.

b) Average amount of fine by vehicle

Compare the average amount of fine by vehicle color, vehicle year, and vehicle plate type [Hint: it is sufficient to restrict your attention to commercial (`COM`) and passenger (`PAS`) vehicles]? Briefly describe your findings.

2. Map by Precincts

Read in the shape files for the police precincts and remove all precincts outside of Manhattan.

a) Number of tickets, total fines, and average fines

Provide three maps that show choropleth maps of:

- the total number of tickets
- the total amount of fines
- the average amount of fines

Briefly describe what you learn from these maps in comparison.

b) Types of violations

Group the almost 100 types of ticket violations into a smaller set of 4-6 subgroups (where `other` should be the remainder of violations not included in other groups you defined). [Hint: No need to spend more than 5 minutes thinking about what the right grouping is.]. Provide choropleth maps for each of these subgroups to show where different types of violations are more or less common.

3. Focus on the Upper East

Precinct 19 identifies the Upper East Side. The data currently does not provide latitude and longitude of the violation locations (and I am not sure what these `street_code` variables are for).

a) Ignoring fire hydrants

Restrict your data to parking violations related to fire hydrants (`Violation Code = 40`). Using the variables `Street Name` and `House Number` as well as the knowledge that these addresses are in the Upper East Side of Manhattan, geocode at least 500 addresses. Include a data table of these addresses and the latitude and longitude of these addresses in the output.

b) Interactive Map

Provide an interactive map of the violations you geocoded using `leaflet` . Provide at least three pieces of information on the parking ticket in a popup.

c) Luxury cars and repeat offenders

Using the vehicle `Plate ID` , identify repeat offenders (in the full dataset). Create another variable called `luxury_car` in which you identify luxury car brands using the `Vehicle Make` variable.

Start with the previous map. Distinguish the points by whether the car is a repeat offender and/or luxury car. Add a legend informing the user about the color scheme. Also make sure that the added information about the car type and repeat offender status is now contained in the popup information. Show this map.

d) Cluster

Add marker clustering, so that zooming in will reveal the individual locations but the zoomed out map only shows the clusters. Show the map with clusters.

Submission

Please follow the instructions to submit your homework. The homework is due on Monday, March 25 at 5pm.

Please stay honest!

If you do come across something online that provides part of the analysis / code etc., please no wholesale copying of other ideas. We are trying to evaluate your abilities to visualized data not the ability to do internet searches. Also, this is an individually assigned exercise – please keep your solution to yourself.