STAT605 Final Project

Wenbo Fei, Qin Hao, Enze Wang, Jiantong Wang, Xiaotian Wang

11/20/2019

Data Description - NYU Parking Tickets

42.3M Rows of Parking Ticket Data, Aug 2013-June2017: https://www.kaggle.com/new-york-city/nyc-parking-tickets/

Questions - What would lead to a parking violation?

- Violation lacation: Do specific streets have more violations? - Construct more parking lot.
- Local or foreigner: Are foreigners more easy to make violations? - Add recognizable parking sign
- Do violations happen mainly in specific time in a day? - Change parking rule as time changes
- Find plate(people) who always made a parking violation.
- Plate type, Vehicle body type/brand: Are certain type more likely to violate?
- Does the pattern change by years?

Read Data

```
rawdata2014 = fread("PV2014.csv", select=c(2,3,4,7,8,20,22,25,44,45))
rawdata2014[1,]
```

```
##    Plate ID Registration State Plate Type Vehicle Body Type Vehicle Make
## 1:  GBB9093                 NY        PAS              SUBN         AUDI
##    Violation Time Violation County Street Name Latitude Longitude
## 1:          0752A                      W 175 ST       NA        NA
```

Variable Description

Plate ID: of the violation car. **Registration State**: state issued the plate. **Plate Type**: e.g Commecial vehicle. **Vehicle Body Type**: eg. Suv. **Vehicle Make**: eg. Ford **Violation time**: time in a day the violation happened. **Violation County**: county the violation happened. **Streetname**: street the violation happened. **latitude, longitude**: Where did the violation happen(use to draw on map). There are other variables describe issuer information, vehicle appearance, violation infomation in detail.

Methods

- Plot the violation heatmap to find which street/area has more parking violation
- Use χ^2 test or t-test to see if the certain type of vehicles would have higher possibility of parking violation in certain area.

Computation

The task isn't complex but need a lot of computation and there are data of many years, so we use CHTC, and compute each years parallelly.