Data608_Final project Proposal

Mengqin Cai

10/18/2020

DATA 608 Final Project Proposal

As the largest populated city in New York and also in US, New York City has more than 8 million population. When the city keep growing, road safety is a critical issue for people who live in this crowded city.Therefore, it is very important to look at the past collision history data and see what we can learn from the data to help better prevent or avoid collisions in the future.

Data Sources

The dataset I used is from the New York City open data.https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95. The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC.

Sample of the Data

```
data<-read.csv("https://raw.githubusercontent.com/DaisyCai2019/NewData/master/2020%20Motor_Vehicle_colli
head(data)
```

```
##   CRASH.DATE CRASH.TIME  BOROUGH ZIP.CODE LATITUDE LONGITUDE
## 1  10/9/2020      22:26               NA 40.82874 -73.83578
## 2  10/8/2020       8:00               NA 40.60728 -74.14512
## 3  10/6/2020      13:30               NA 40.78905 -73.94861
## 4  10/6/2020      21:00               NA 40.58396 -73.92530
## 5  10/5/2020      17:00 BROOKLYN    11233 40.67279 -73.92223
## 6  10/4/2020       9:00               NA 40.82011 -73.90820
##                  LOCATION                   ON.STREET.NAME
## 1  (40.828743, -73.83578) CROSS BRONX EXPY
## 2  (40.607277, -74.14512)
## 3  (40.789055, -73.94861) LEXINGTON AVENUE
## 4    (40.58396, -73.9253) BELT PARKWAY
## 5  (40.67279, -73.922226) PROSPECT PLACE
## 6 (40.820114, -73.908195) EAST 158 STREET
##   CROSS.STREET.NAME                        OFF.STREET.NAME
## 1
## 2                 680       SOUTH GANNON AVENUE
## 3
## 4
## 5      RALPH AVENUE
## 6
##   NUMBER.OF.PERSONS.INJURED NUMBER.OF.PERSONS.KILLED
## 1                         1                        0
```

```
## 2                         0                         0
## 3                         0                         0
## 4                         0                         0
## 5                         2                         0
## 6                         0                         0
##   NUMBER.OF.PEDESTRIANS.INJURED NUMBER.OF.PEDESTRIANS.KILLED
## 1                             0                            0
## 2                             0                            0
## 3                             0                            0
## 4                             0                            0
## 5                             0                            0
## 6                             0                            0
##   NUMBER.OF.CYCLIST.INJURED NUMBER.OF.CYCLIST.KILLED
## 1                         0                       0
## 2                         0                       0
## 3                         0                       0
## 4                         0                       0
## 5                         0                       0
## 6                         0                       0
##   NUMBER.OF.MOTORIST.INJURED NUMBER.OF.MOTORIST.KILLED
## 1                          1                         0
## 2                          0                         0
## 3                          0                         0
## 4                          0                         0
## 5                          2                         0
## 6                          0                         0
##   CONTRIBUTING.FACTOR.VEHICLE.1 CONTRIBUTING.FACTOR.VEHICLE.2
## 1                   Unspecified                   Unspecified
## 2                         Glare                   Unspecified
## 3           Passing Too Closely                   Unspecified
## 4                   Unspecified
## 5                   Unspecified                   Unspecified
## 6                   Unspecified                   Unspecified
##   CONTRIBUTING.FACTOR.VEHICLE.3 CONTRIBUTING.FACTOR.VEHICLE.4
## 1
## 2
## 3
## 4
## 5
## 6
##   CONTRIBUTING.FACTOR.VEHICLE.5 COLLISION_ID VEHICLE.TYPE.CODE.1
## 1                                    4356497               Sedan
## 2                                    4355988                Dump
## 3                                    4355248           Box Truck
## 4                                    4355340               Sedan
## 5                                    4355008               Sedan
## 6                                    4354622               Sedan
##   VEHICLE.TYPE.CODE.2 VEHICLE.TYPE.CODE.3 VEHICLE.TYPE.CODE.4
## 1               Sedan
## 2                Dump
## 3                 Bus
## 4               Sedan
## 5               Sedan
## 6               Sedan
```

```
##   VEHICLE.TYPE.CODE.5
## 1
## 2
## 3
## 4
## 5
## 6
```

The motor vehicle collision database includes the date and time, location (as borough, street names, zip code and latitude and longitude coordinates), injuries and fatalities, vehicle number and types, and related factors for all 1.7 Million collisions in New York City from 2016 until now.

Analysis and Visualization

Because the dataset in NYC open data contains 1.7m collision cases, some of the analysis will only focus on the years of 2020, instead of from 2016 to 2019.

- I will do some preliminary analysis to see the overall picture of the five boroughs' accident rate changed by years.
- I will also create a shiny app to show the viewers different collision rates according to different years, boroughs or Vehicle types.
- The intensity of the Heatmap will show the viewers which area has the highest Vehicle collision rate, so they can be more careful when they past these areas.
- The data include the Collision Date, so I will create a graph to show the peak day of a week in terms of number of collisions