There are 684754 entries of data from 01/01/2012 to 07/04/2015.

For each observations, the information includes

"Date Of Stop ",   "Time Of Stop ", "Agency ", "Sub-Agency ",
"Description ", "Location ", "Latitude ", "Longitude ", "Accident ", "Belts ", "Personal Injury ",
"Property Damage ", "Fatal ", "Commercial License ", "HAZMAT ", "Commercial Vehicle ",
 "Alcohol ", "Work Zone ", "State ", "VehicleType " , "Year ", "Make ", "Model ", "Colour ",
 "Violation Type ", "Charge ", "Article ", "Contributed To Accident ", "Race ", "Gender ",
 "Driver City ", "Driver State ", "DL State ", "Arrest Type ", "Geolocation "

The core features of the observations for the analysis are determined as

Location (Latitude, Longitude), Gender,
Vehicle information (Makes, Car Type, Model),
Violation information (Type of Violation ,Time and Date).

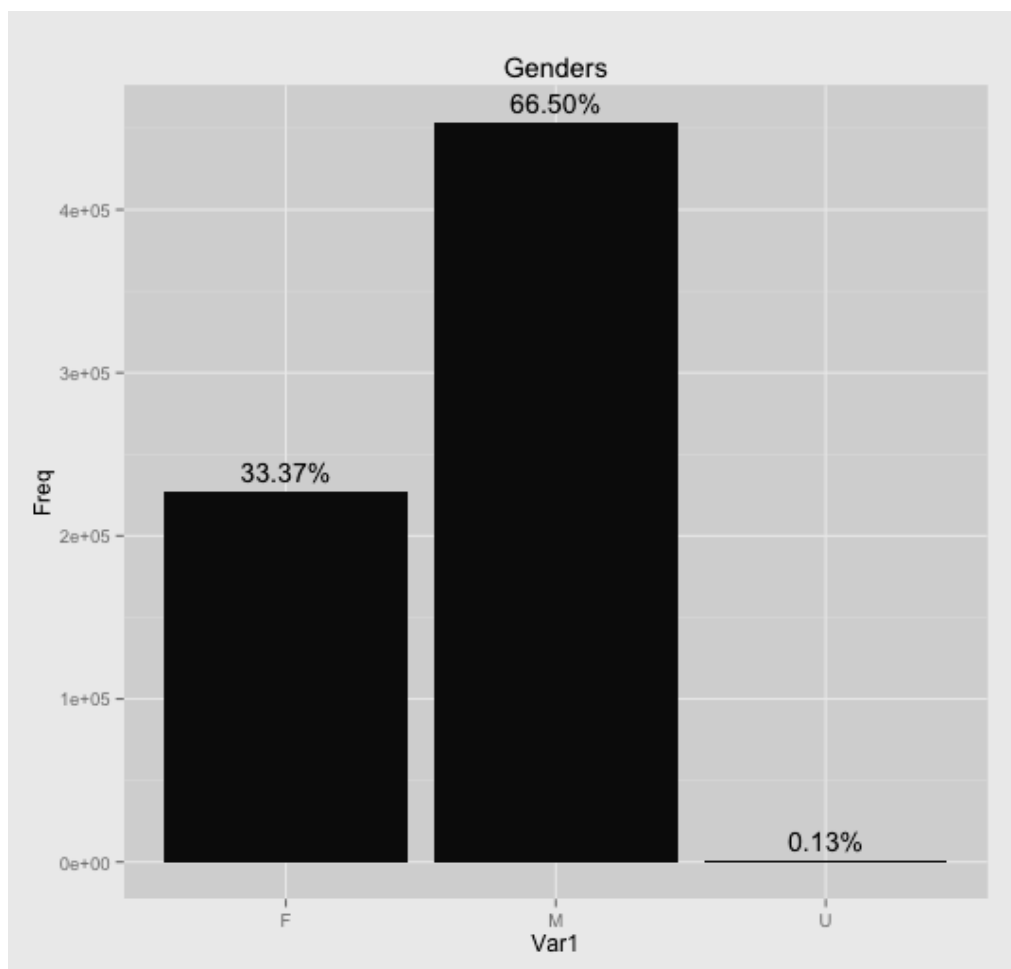**The distribution of the core features are stated below**
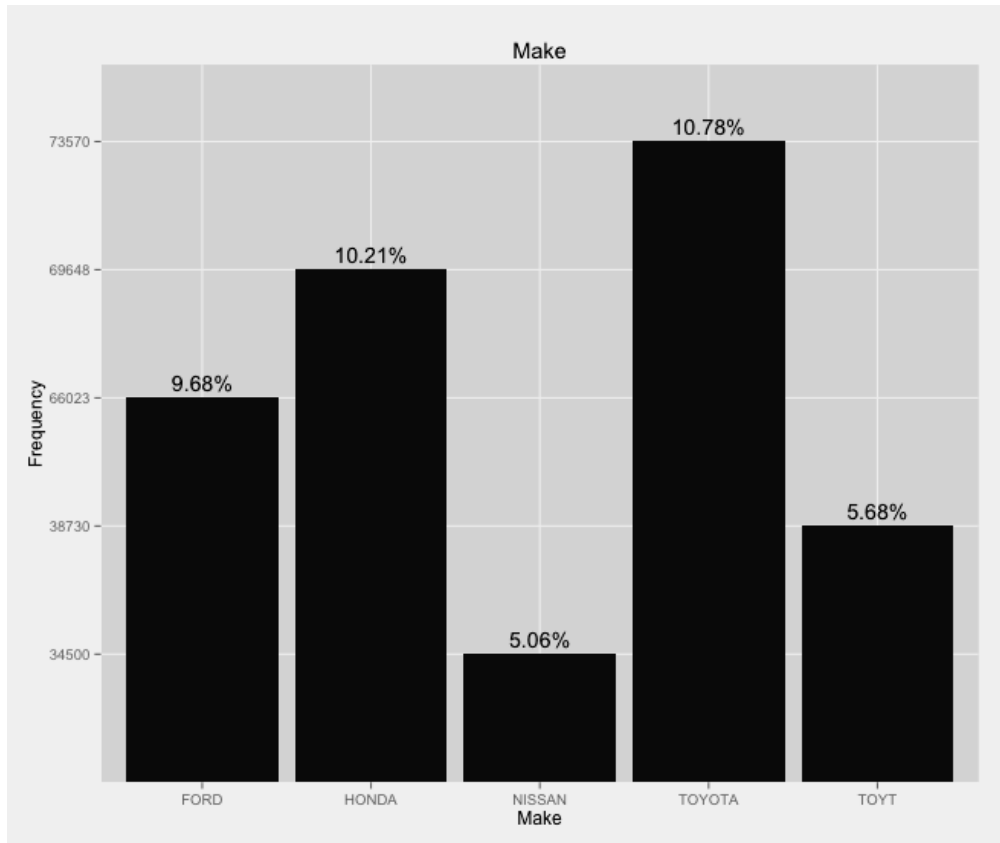

Fig 1. Gender Distribution
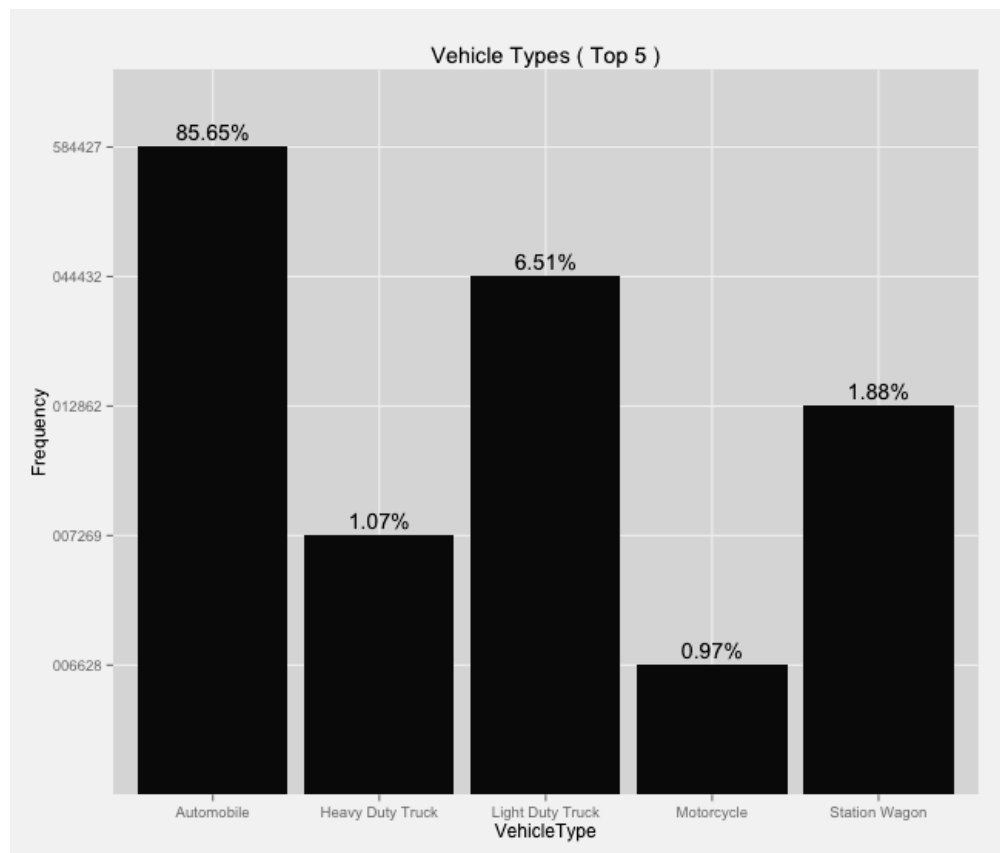
Fig 2. Car Manufacturer Distribution(Top 5)

Fig 3. Vehicle Types Distribution (Top 5)
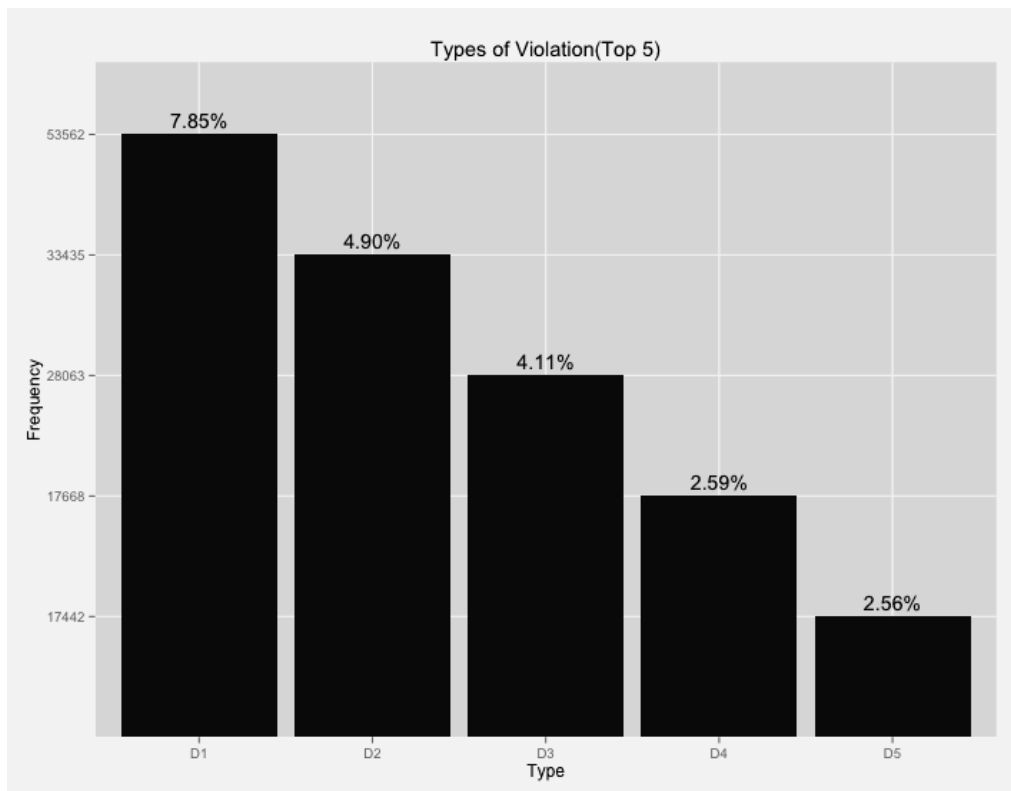
Fig 4. Violation Type Distribution(Top 5)
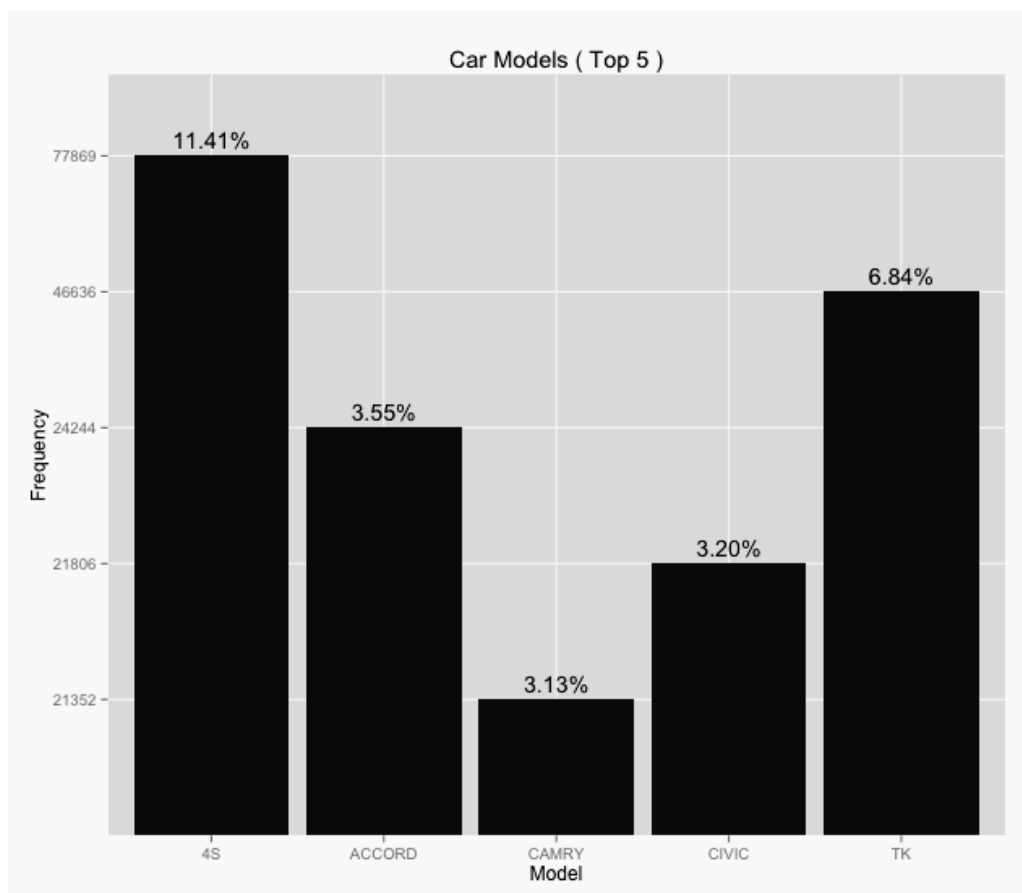

Fig 5. Car Model Distribution (Top 5)

Fig 6. Point Distribution of Locations

Obvious trends in the data from the graphs

1. Most of the people who violated the traffic regulation are male.
2. Most of the cars involved in traffic violation are manufactured by Toyota.
3. Most of the cars involved in traffic violation are automobile.
4. The most common type of traffic violation is "DRIVER FAILURE TO OBEY PROPERLY PLACED TRAFFIC CONTROL DEVICE INSTRUCTIONS"
5. Most of the cars involved in traffic violation are of model 4S.

For the vehicle type distribution, the differences are statistically significant. 85% of the cars are automobile.

Other salient aspects of the data

The dataset includes information on latitude and longitude of the locations of traffic violations. And the description for traffic violations are in text.

Next 5-10 tasks

- Plot a heat map showing date distribution of traffic violations
- Discover more correlations between the places of traffic violations and the actual road map

- Plot various information by month to discover seasonal changes of the characteristics of traffic violations
- Discover more correlations and difference between gender by separating the data by gender.