Preliminary Analysis- Traffic Violation in Maryland

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

For each observation, 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

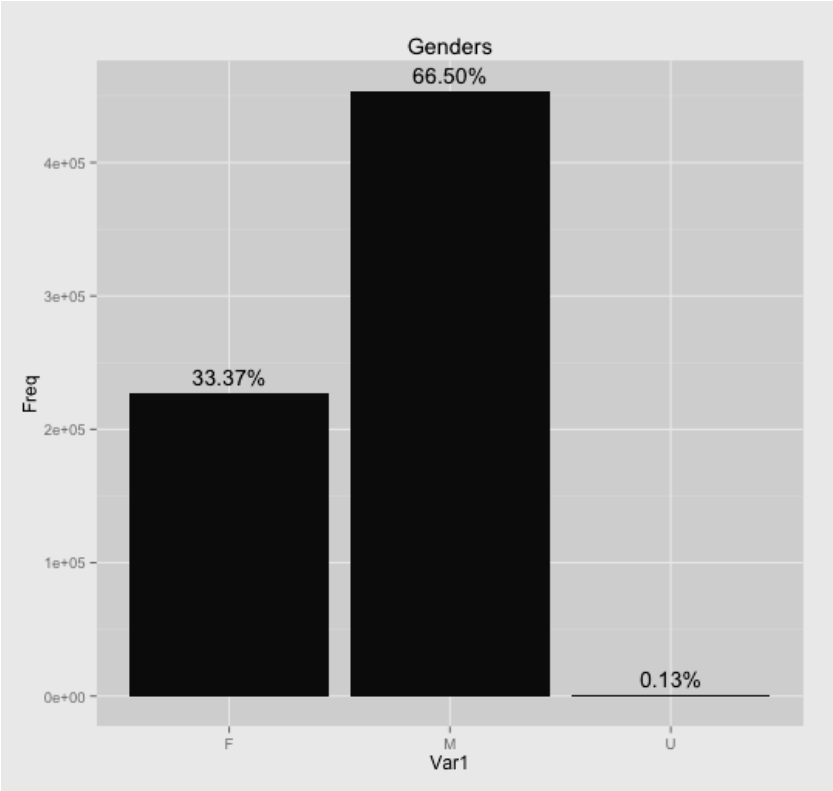


Fig 1. Gender Distribution

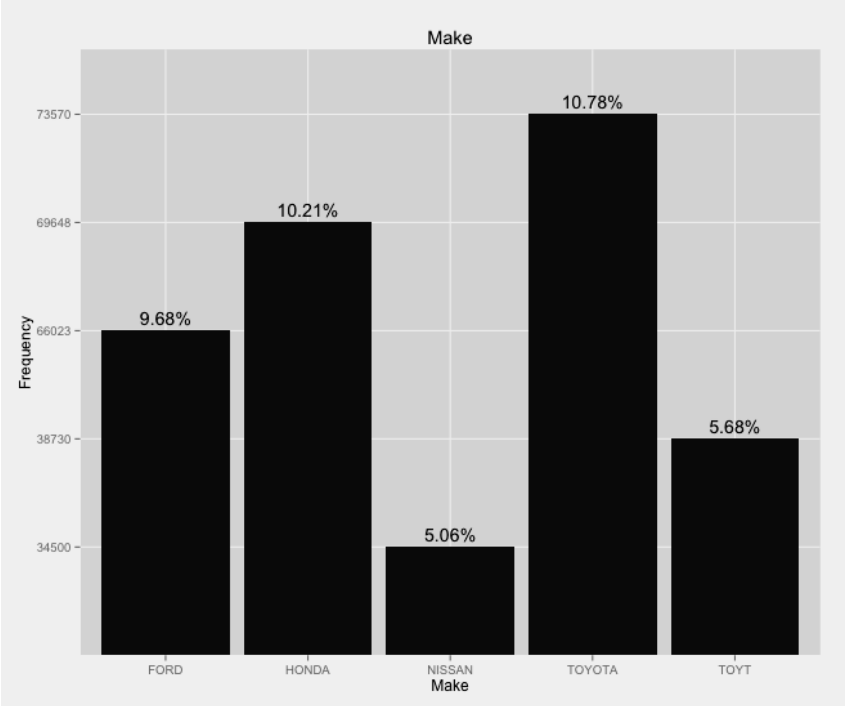


Fig 2. Car Manufacturer Distribution(Top 5)

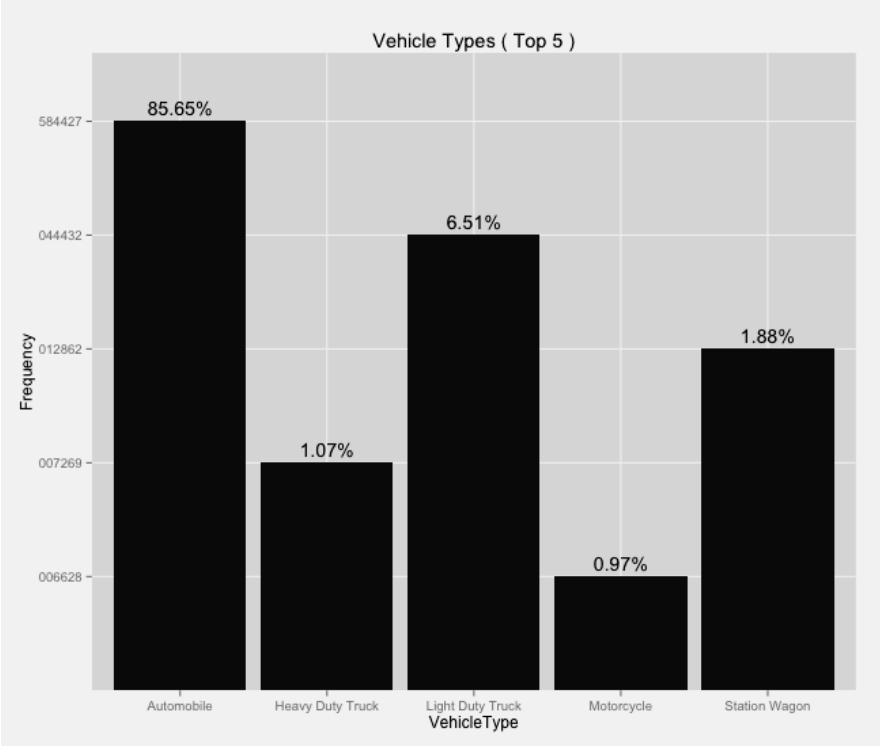


Fig 3. Vehicle Types Distribution (Top 5)

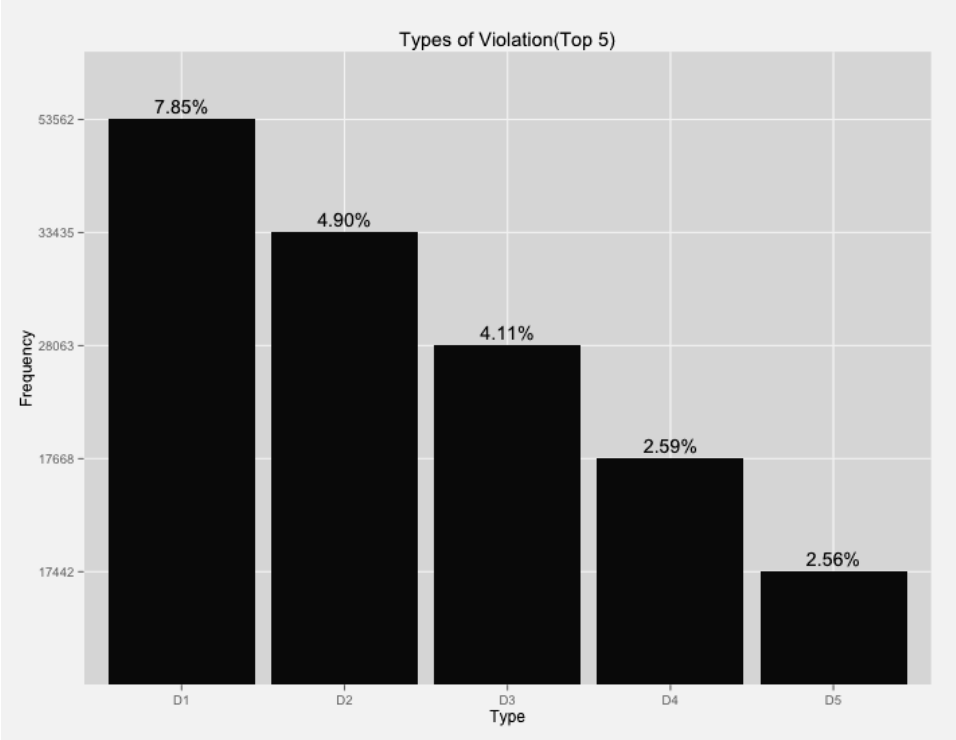


Fig 4. Violation Type Distribution (Top 5)

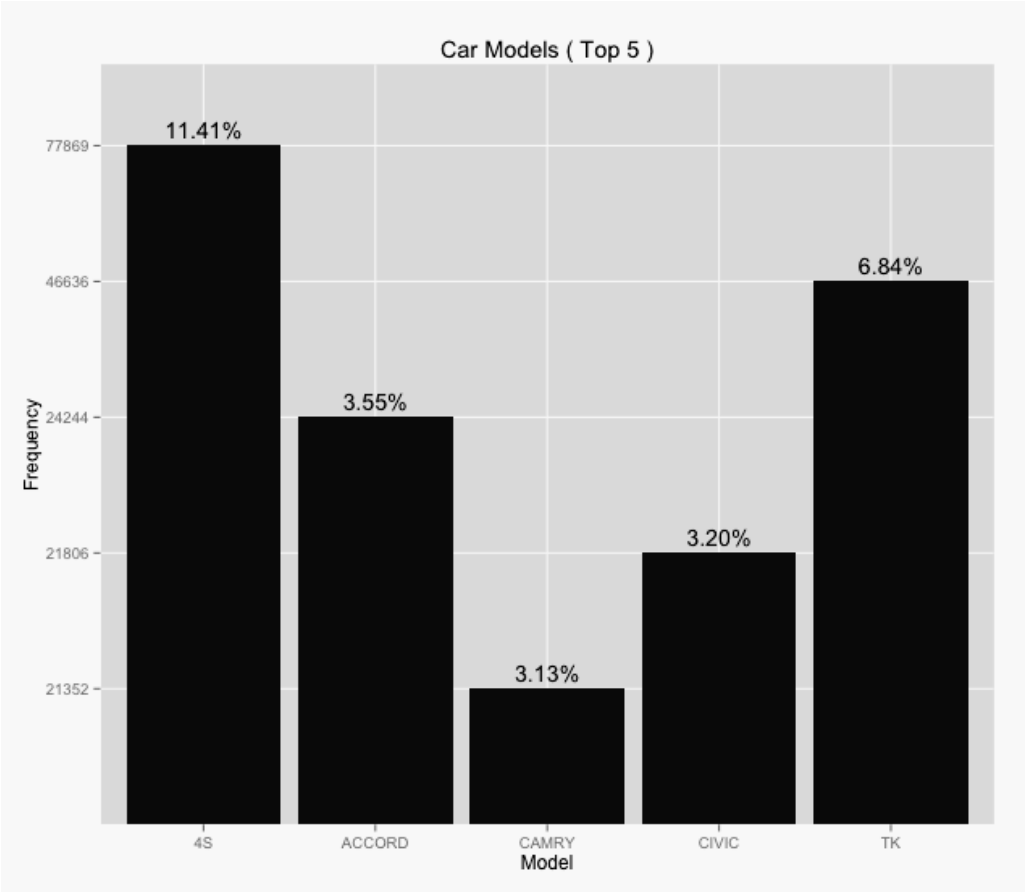


Fig 5. Car Model Distribution (Top 5)

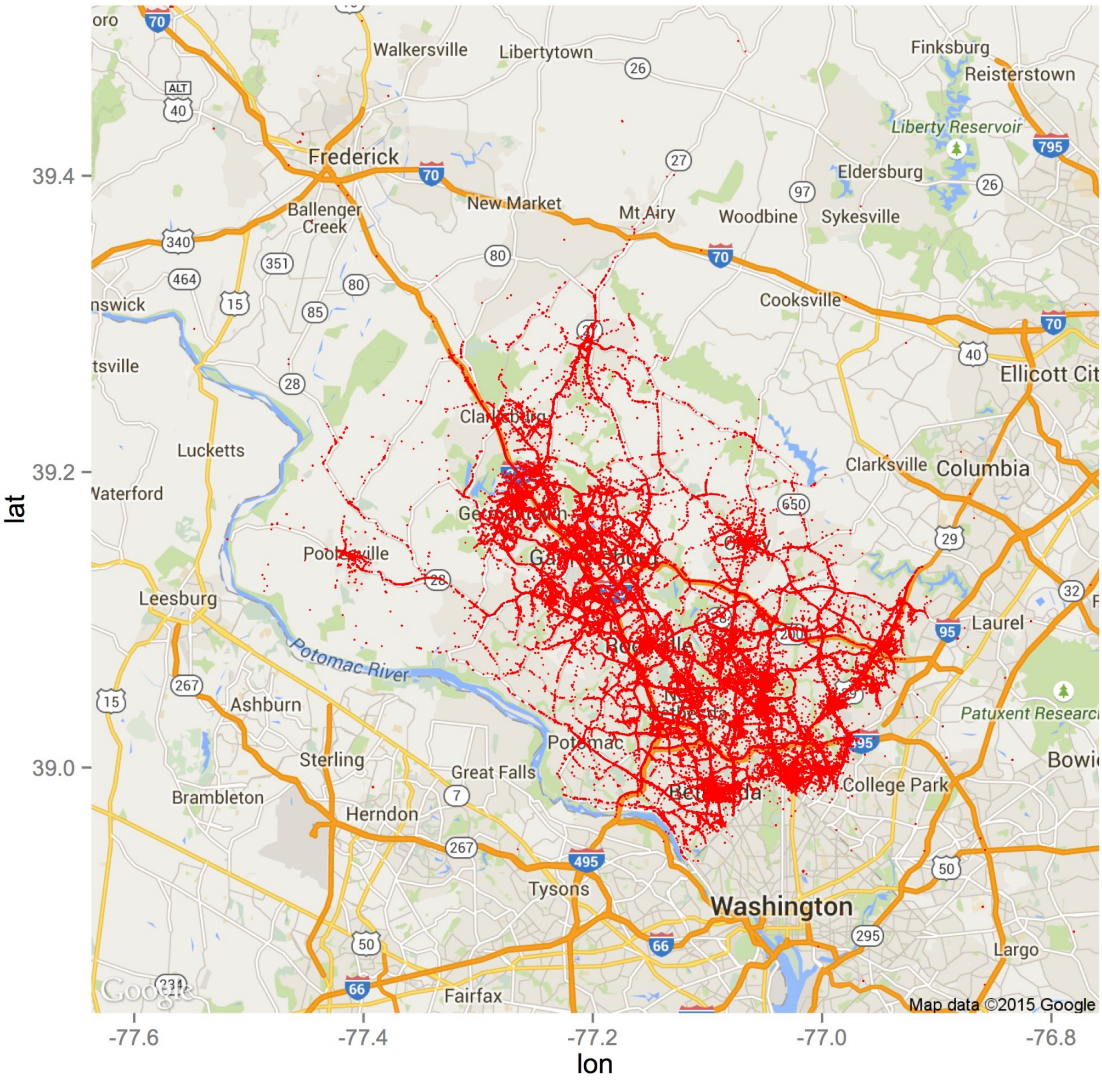


Fig 6. Point Distribution of Locations

Analysis from the graphs above

- Most of them who violated the traffic regulation are **male.**
- Most of the cars involved in traffic violation are manufactured by **Toyota.**
- Most of the cars involved in traffic violation are **automobile**.
- Most of the cars involved in traffic violation are of **model 4S**.
- The main reason for traffic violation is because driver **failed to obey the properly placed traffic control device instructions.**

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

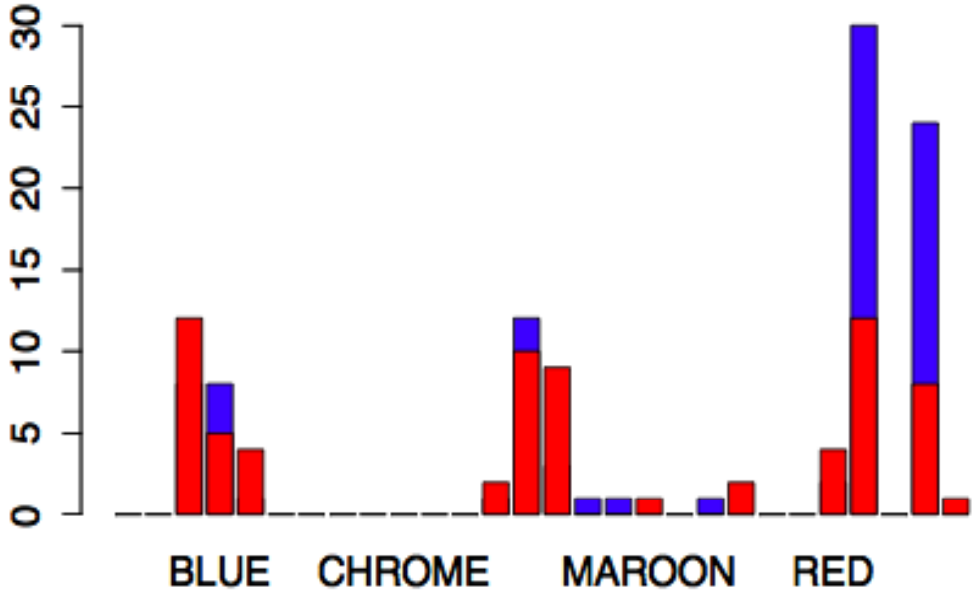
Other salient aspects of the data

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

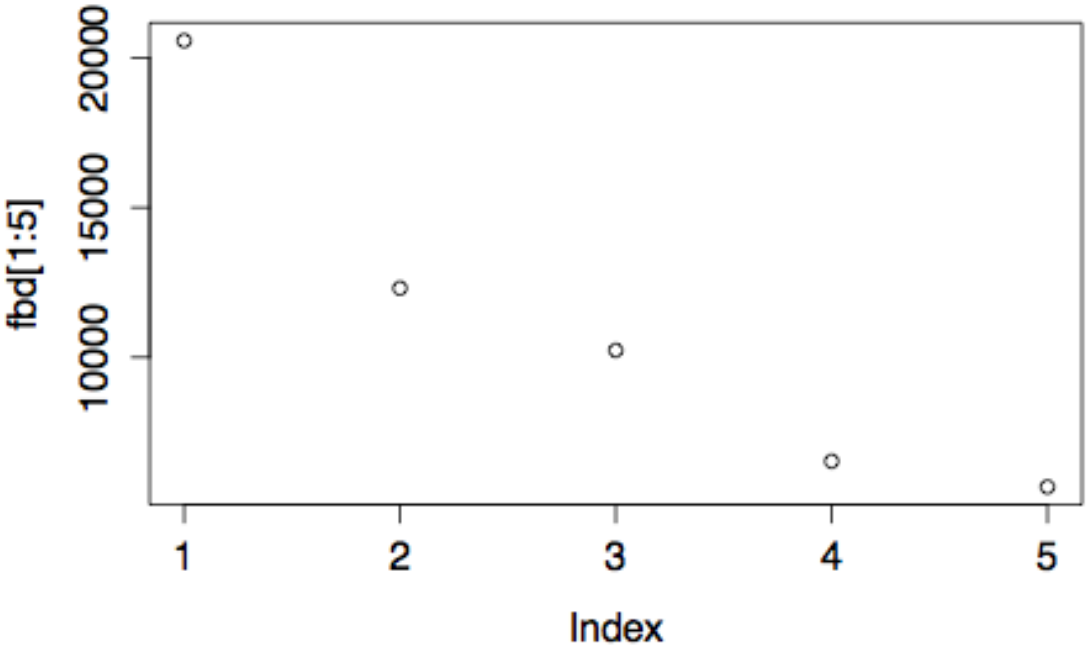
Next 5-10 tasks

- 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

Correlation between gender

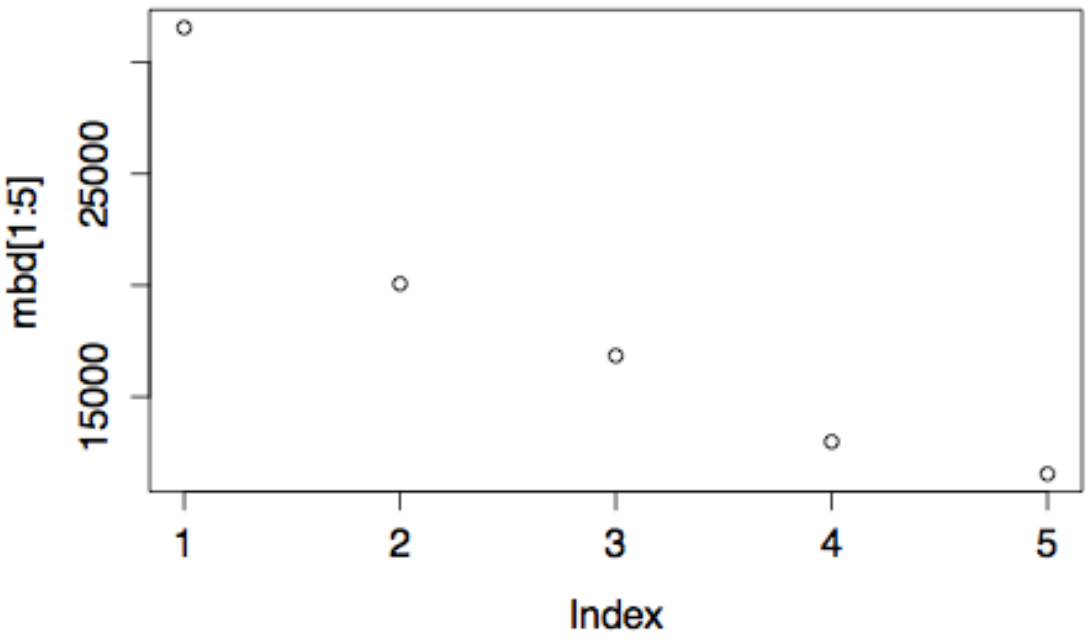


Plot 1
Red= Female, Blue= male who commit fatal



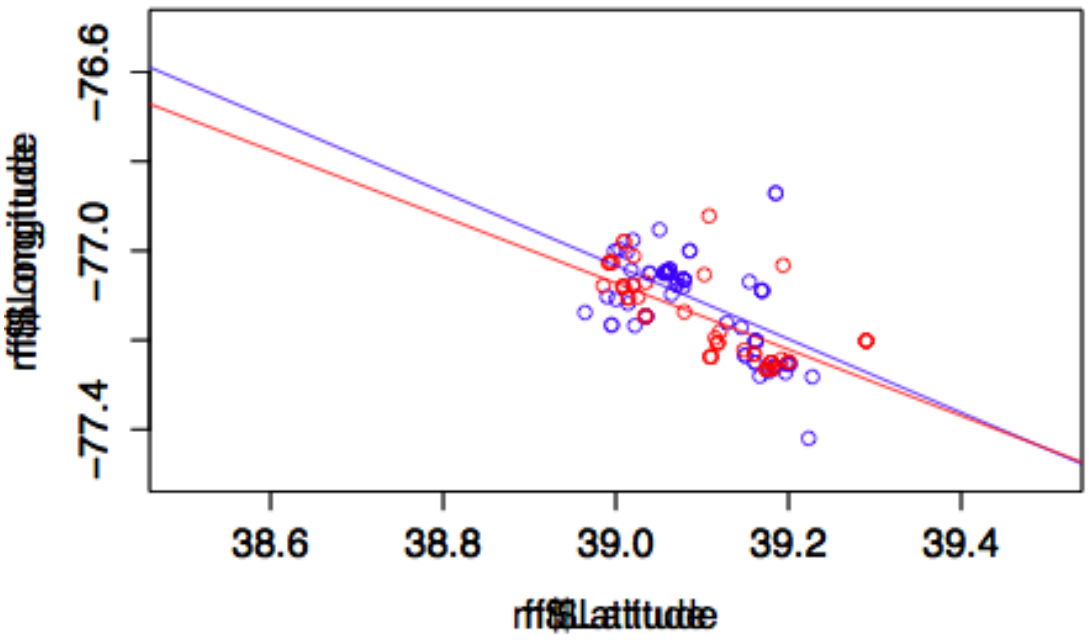
Plot2 means emales who belt=No dirver's top5 description
In plot2, female driver failed because driver **failed to obey the properly placed traffic control device instructions.**

DATA:
FAILURE TO DISPLAY REGISTRATION CARD UPON DEMAND BY POLICE OFFICER
12308
DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGISTRATION
10234
DRIVER FAILURE TO STOP AT STOP SIGN LINE
6533
DISPLAYING EXPIRED REGISTRATION PLATE ISSUED BY ANY STATE
5677

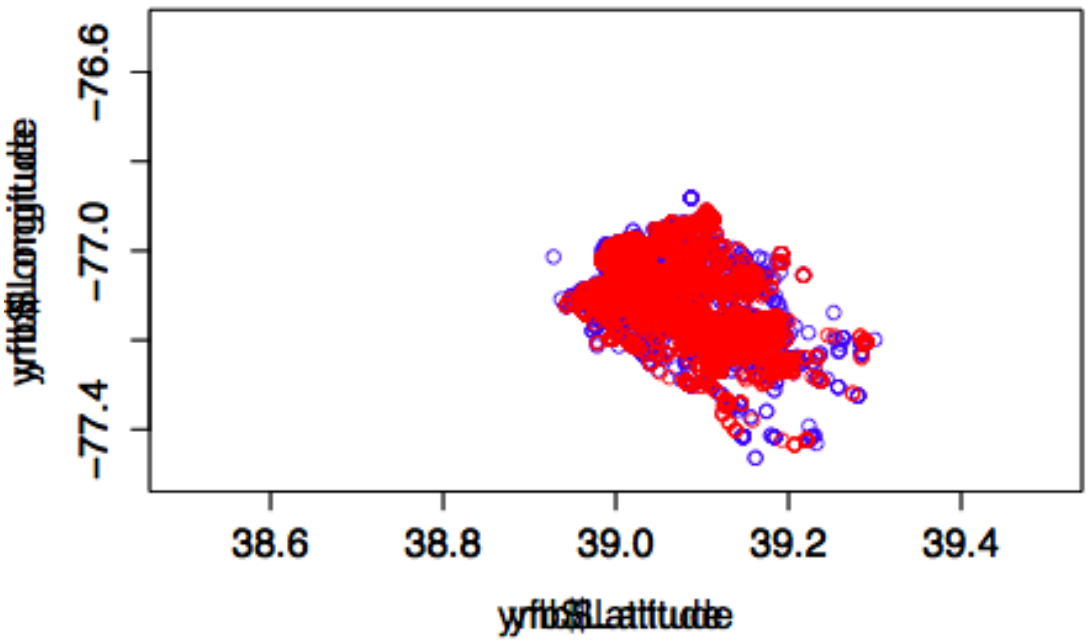


In plot3, male driver failed to obey the properly placed traffic control device instructions.

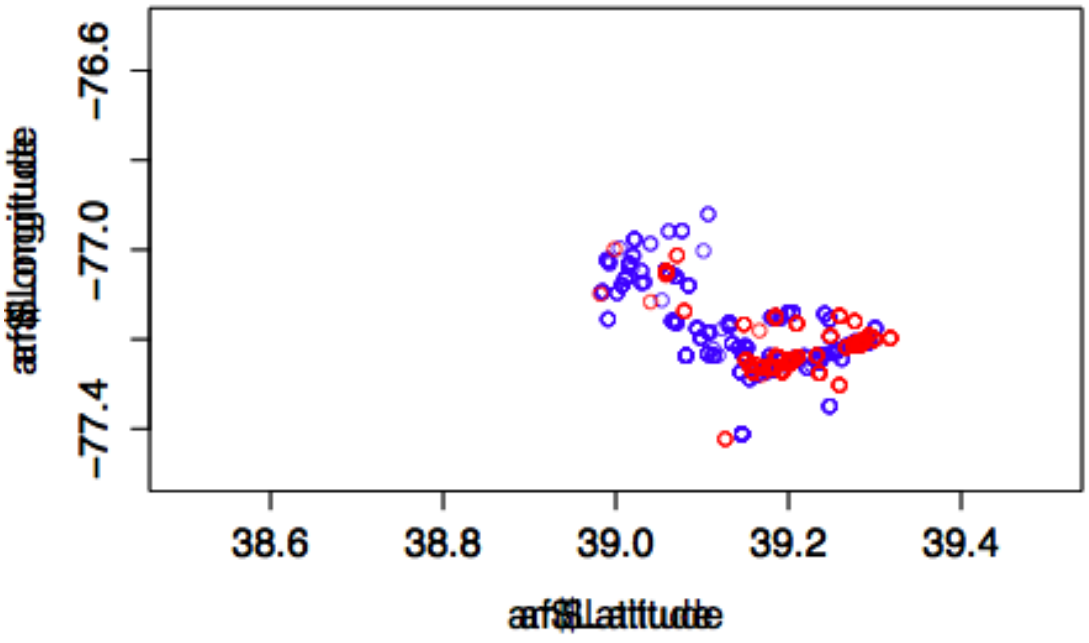
DATA:
DRIVER FAILURE TO OBEY PROPERLY PLACED TRAFFIC CONTROL DEVICE INSTRUCTIONS
31555
FAILURE TO DISPLAY REGISTRATION CARD UPON DEMAND BY POLICE OFFICER
20069
DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGISTRATION
16842
OPERATOR NOT RESTRAINED BY SEATBELT
12986
FAILURE OF INDIVIDUAL DRIVING ON HIGHWAY TO DISPLAY LICENSE TO UNIFORMED POLICE ON DEMAND
11543



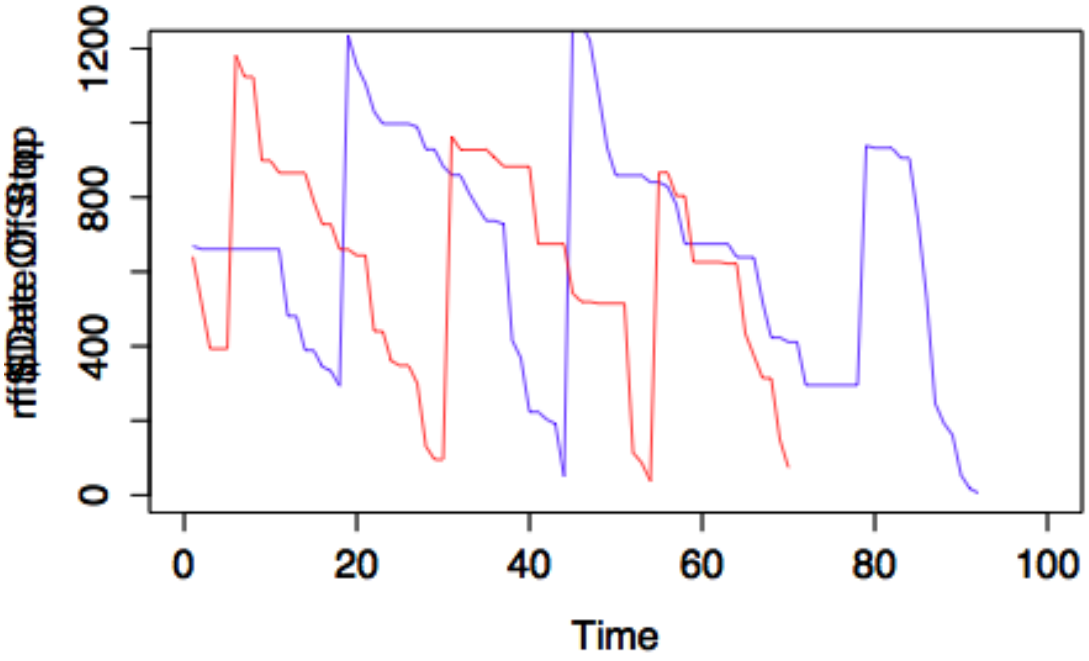
Plot4 means male(blue), female(red), who commit fatal. Plot with regression line



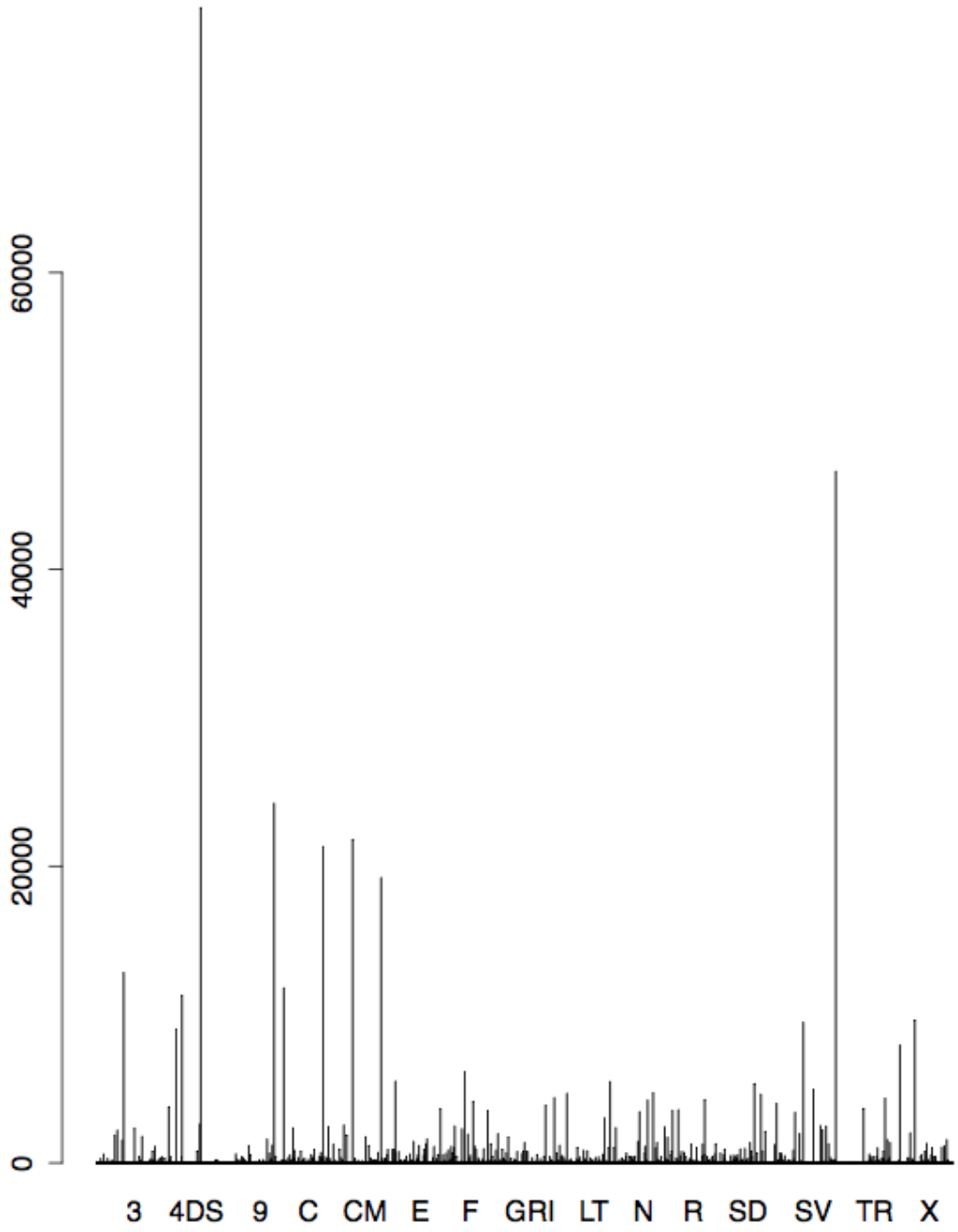
Plot6 means seat belt=Yes data plot



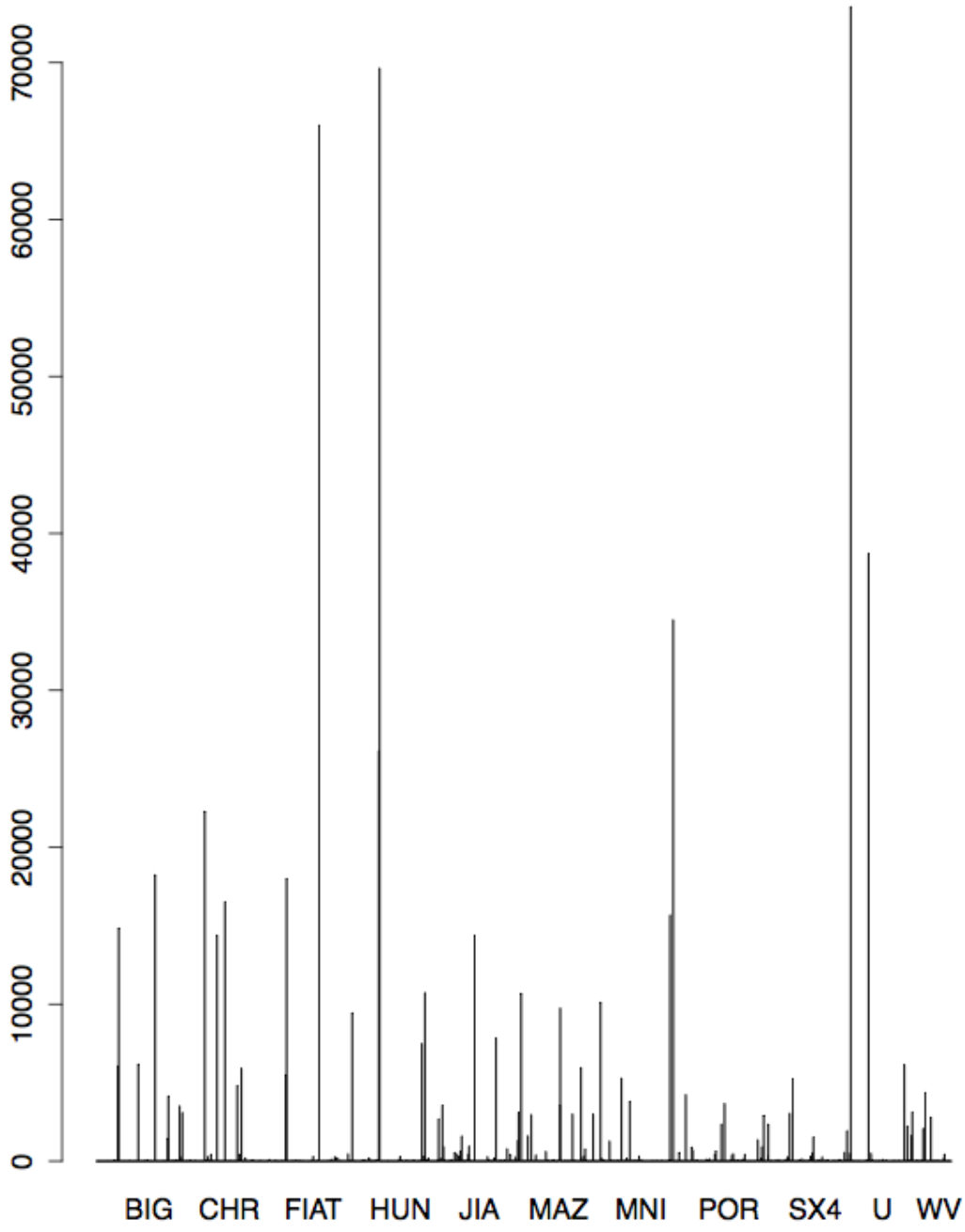
Plot7 means Alcohol = Yes data plot



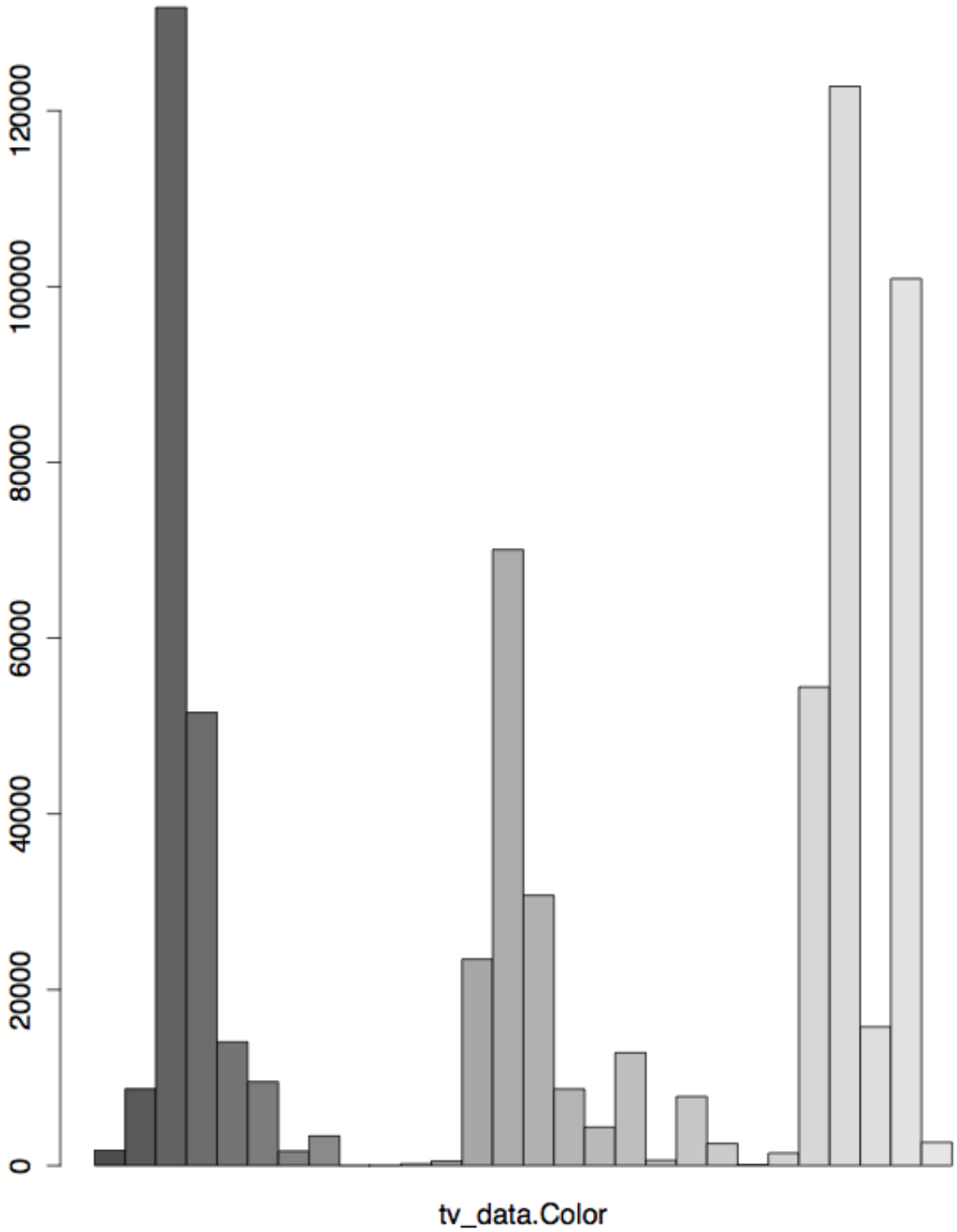
Plot 8 means data of time series in fatal from male, female



The bar graph 'Model_accident' show the particular car model had been related to the accident in Maryland.



‘Maker_accident’, the same reason with above a maker of the car have related to accident of the car model.



‘Colour_Accident’, car colours shows that several colours have higher frequency related with the accident

Done by: Outliers