Exploratory Analysis

Fadi

26 January 2018

Data Preperation

Further cleaning was needed to process data better. Observations under Belts, Property Damage, Alcohol, and other variables needed to be converted into binary for easier summary statistics.

This is a breakdown of some proportions

```
## Adding missing grouping variables: `Race`

## # A tibble: 6 x 6
##   Race             Belts `Personal Injury` `Property Damag~   Fatal Alcoh~
##   <chr>            <dbl>             <dbl>            <dbl>   <dbl>  <dbl>
## 1 ASIAN            0.128             0.284            0.288 0.00337 0.0157
## 2 BLACK            0.135             0.256            0.298 0.00397 0.0429
## 3 HISPANIC         0.138             0.236            0.333 0.00296 0.0887
## 4 NATIVE AMERICAN 0.0606             0.121            0.333 0       0
## 5 OTHER            0.143             0.260            0.304 0.00285 0.0257
## 6 WHITE            0.133             0.275            0.327 0.00310 0.0496
```

One key step was to convert the set into distinct observations. Data, in its original format, had multiple observations referring back to the same person; i.e. a person can get pulled over and receive one citations for speeding, one for seatbelt, and one for intoxication, hence, registering more than one observation, **all clocked at the same time**. Data needed to be arranged by date and time, first, and then a secondary data was generated that only included unique observations of Dates & Times. This subset is used to calculate statistics regarding each person, without the redundency of the original.

Another subset was created to analyse violations that strictly involved Accidents.
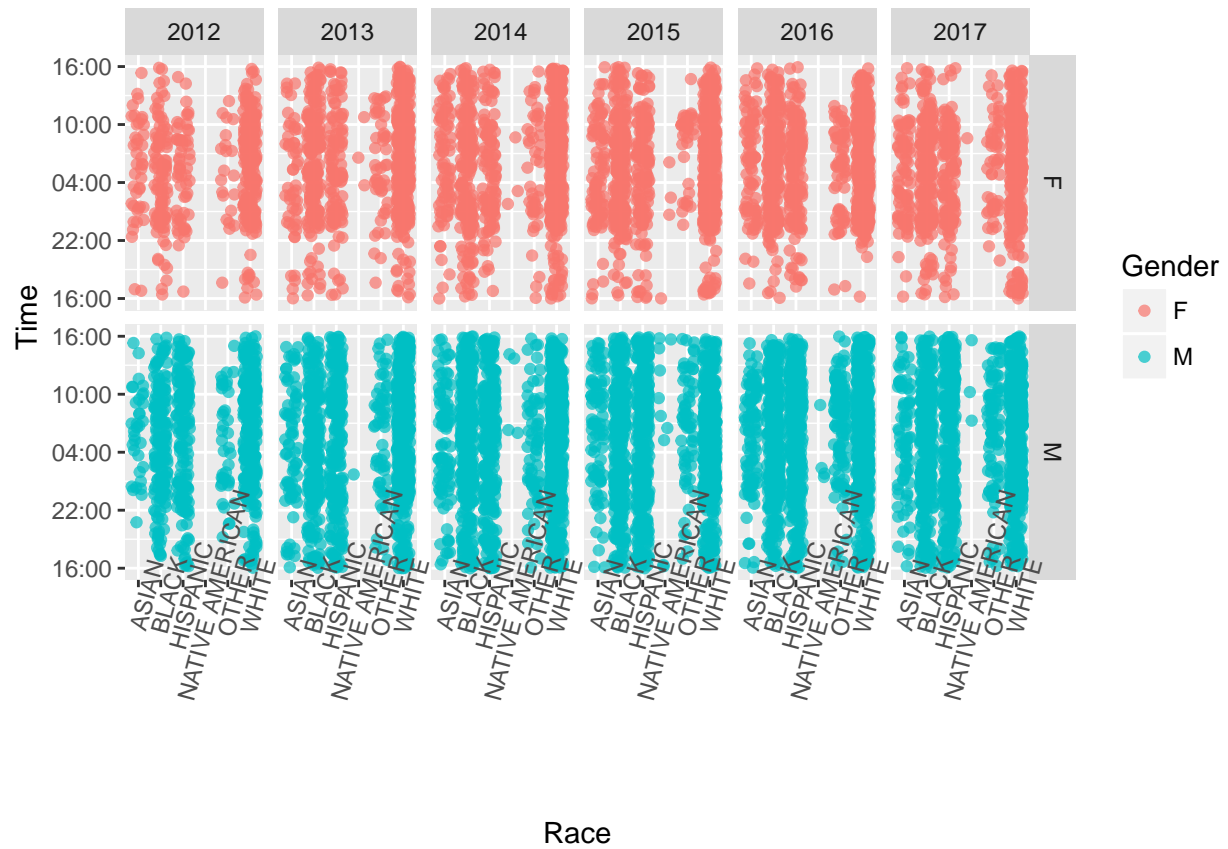
Visualising Data

The following describes the distribution of accidents throughout the day, grouped by car color
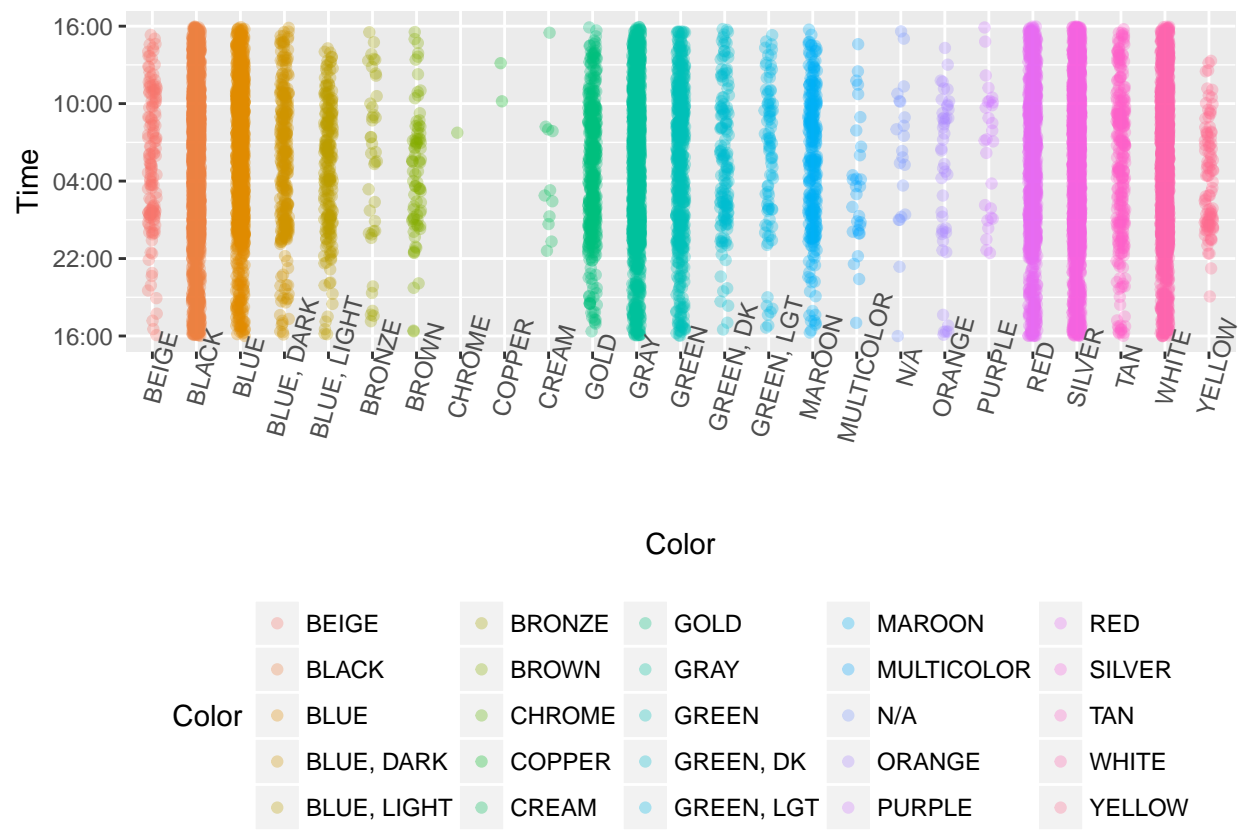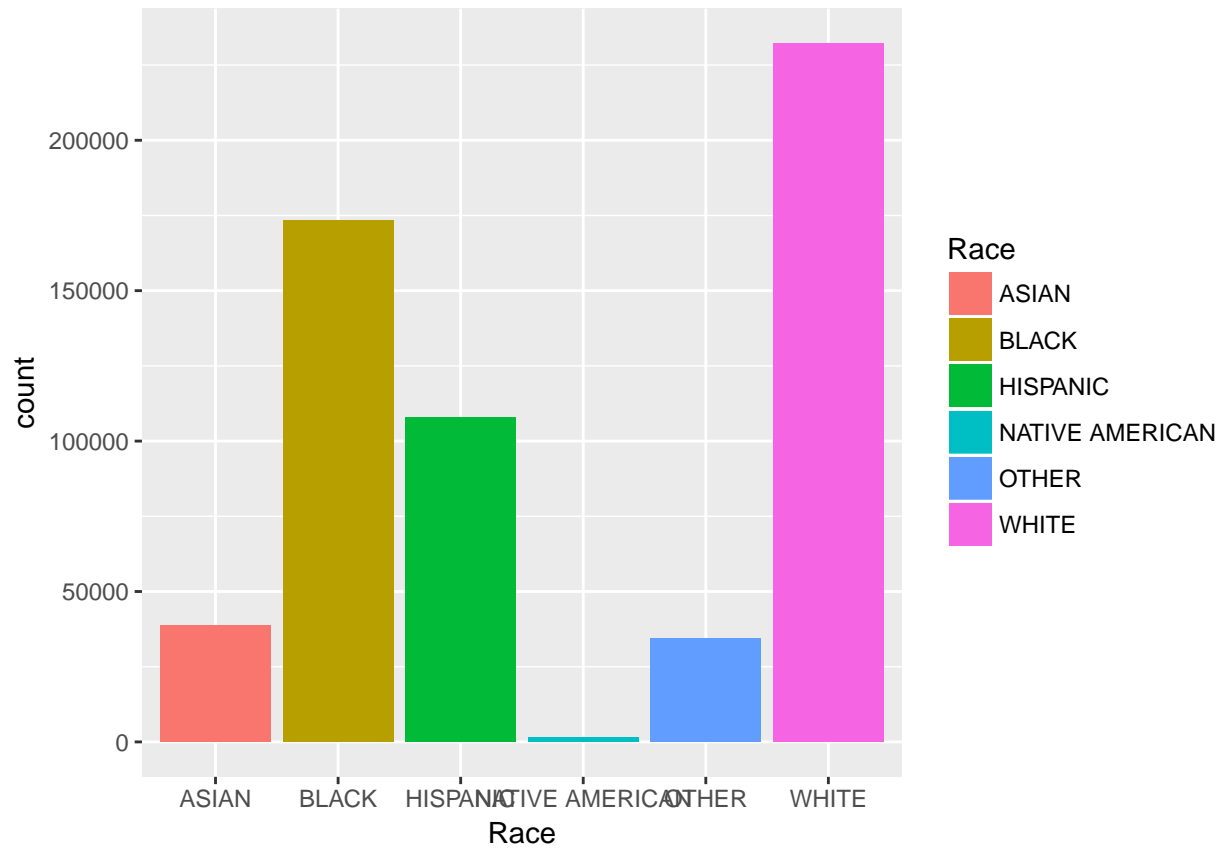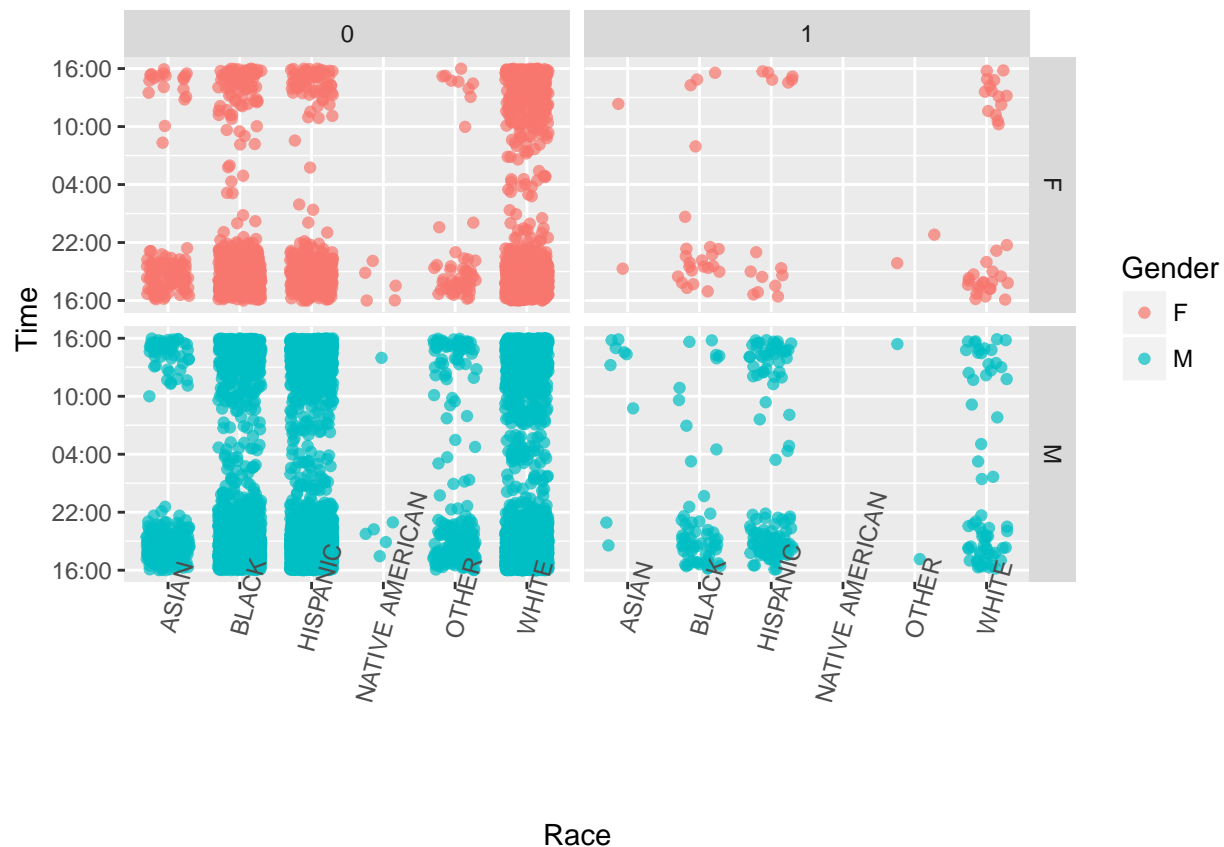
This is distribution of Violations committed by different races

This graphs time of the day against which males and females of different Races are in accidents, across the years

Wearing and not wearing a seatbelt is just as likely when alcohol is not involved. It is a different story drinking is involved. Which kind of is self-explanatory; people who are reckless enough to drive under the influence are less likely to care about a seatbelt.

```
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Montgomery%20County&
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=39.154743,-77.240515&zoom=11&siz
## Warning: Removed 3157 rows containing missing values (geom_point).
```
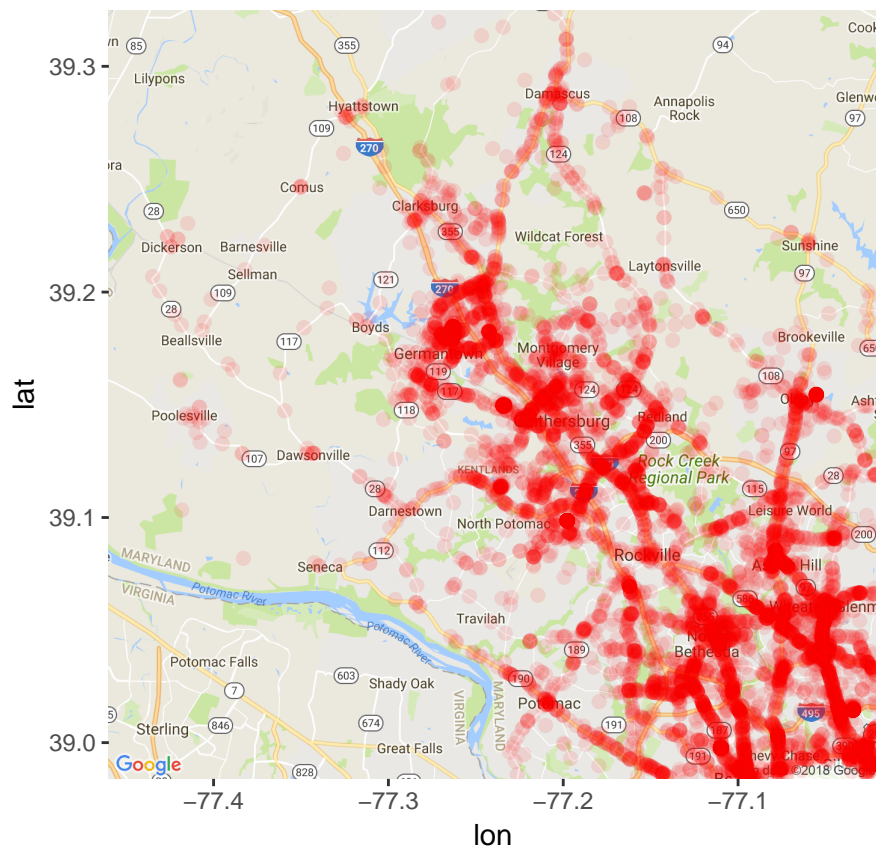
This is a heatmap highlighting the distribution of accidents in Montgomery county.

What's Next

There is still much patttern to be extrapolated. One extra step that needs to be taken, is segregating observations by seasons. What are some trends that could be seen more at winter, than summer time..?