Notes Analysis

DEREJE GETU DEMISSIE

February 17, 2018

### Notes:

* Changing the class of the variables and recoding is necessary
* Summarising them and see the pattern and If necessary , recode them.
* From the Vehicle data, Number of Occupants can be an interesting
* Drivers BMI and fatal accdients from the vehicle data
* Drivers with previous records (mutate previous records)
* % drinking Drivers per state
* % Speeding Drivers per State
* Fatal Crash Injury specifying color by feading… like from very red to light red

## Data Cleanining Process

* Check for missing data: observations with missing data will be removed for this visualization.
* Converted the date and time columns from strings to their respective classes so that we can compare them as actual dates and times instead of strings. Now that we have a proper date value, we can create another column for days of the week. The data provides us with the vehicle types that are involved in each accident. This is good, but they are split up in each of their own columns depending on the number of vehicles.

## Visualizations

* Count of fatalities by the top five and bottom five states (Histogram)
* What about accidents by days of the week? Histogram by the days of the week (for all states/)
* visualize these accidents on a map via longitude and latitude coordinates from the dataset.
* This might give us a busy visualization thus, making interactive to show for each State will solve the big problem (Hit Maps )
* visualizing by days or visualizing by month might help to clear the busy visualization
* visualizing the data by the time of the day [link](https://nycdatascience.com/blog/student-works/r-visualization/interactive-visualizations-nyc-traffic-accidents-r-shiny/)
* Accident density based on the month of the year.
* Accidents mapped mapped on top of a map of states( make the interactivity works for all)
* Visualization depicting the type of the vehicles involved in the accidents [link](https://www.r-bloggers.com/motor-vehicle-collisions-in-new-york-city-r-shiny-data-visualization/)
* Check [link](https://github.com/pchuck/denver-traffic/blob/master/traffic.md#all-accidents-by-time) has important codes for analysis Neighborhoods with the most accidents

## The 2007-2016 FARS Auxiliary Datasets guide

### Accident File

*ST\_CASE is the unique case identifier for each record* \* Crash Injury Type + Fatal Crash(FARS)

* *Crash Type (A\_CT)*
  + Single Vehicle Crash
  + Two Vechile Crash
  + More Than Two-Vehicle Crash
* *Age(A\_Age)*
  + A\_D0-15
  + A\_D16-24
  + A\_D25-64
  + A\_65PLS
* *Involving a Distracted Driver (A\_DIST)*
  + Involving a Distracted Driver
  + Other Crash
* *Day of Week (A\_DOW)*
  + Weekday
  + Weekend
  + Unknown
* *Time of Day(A\_TOD)*
  + Daytime
  + Nighttime
  + Unknown
* Involving a Drowsy Driver(A\_DROWSY)
  + Involving a Drowsy Driver
  + Other Crash
* Involving a Hit and Run (A\_HR)
  + Yes\_Hit and Run
  + No\_Hit and Run
* Interstate (A\_INTER)
  + Interstate
  + Non Interstate
  + Unknown
* Intersection (A\_INTSEC)
  + Intersection
  + Non-Intersection
  + Unknown
* Involving a Large Truck (A\_LT)
  + Involving a Large Truck
  + Other Crash
* Manner of Collision (A\_MANCOL)
  + Not Collision with Motor Vehicle in Transport
  + Rear-End
  + Head-On
  + Angle
  + Sideswipe
  + Other
  + Unknown
* Involving a Motorcyclek (A\_MC)
  + Motorcycle Involved Crash
  + Other Crash
* Involving a Pedestrian (A\_PED)
  + Involving Pedestrian Involved Crash
  + Other Crash
* Involving a Pedalcyclist (A\_PEDAL)
  + Pedalcyclist Involved Crash
  + Other Crash
* Involving a Police Pursuit (A\_POLPUR)
  + Police Pursuit Involved Crash
  + Other Crash
* Involving a Driver With Positive BAC (A\_POSBAC)
  + Driver With Positive BAC Testing Crash
  + All Drivers With ZERO BAC Testing Crash
  + Unknown BAC Crash
* Involving a Roadway Departure (A\_RD)
  + Yes - Roadway Departure
  + No - Roadway Departure
* NHTSA Region (A\_REGION)
  + 1 = CT, ME, MA, NH, RI, VT
  + 2 = NJ, NY, PA, PR
  + 3 = DE, DC, KY, MD, NC, VA, WV
  + 4 = AL, FL, GA, SC, TN
  + 5 = IL, IN, MI, MN, OH, WI
  + 6 = LA, MS, NM, OK, TX
  + 7 = AR, IA, KS, NE, MO
  + 8 = CO, NV, ND, SD, WY, UT
  + 9 = AZ, CA, HI
  + 10 = AK, ID, MT, OR, WA
* Relationship to the Road (A\_RELRD)
  + On Roadway
  + Off Roadway/Shoulder
  + Off Roadway/Median
  + Off Roadway/Other
  + Off Roadway (no other details)
  + Other/Unknown
* Roadway Function Class(A\_ROADFC)
  + Interstate, principal arterial
  + Freeway and expressway, principal arterial
  + Principal arterial, other
  + Minor arterial
  + Collector
  + Local
  + Unknown
* Involving a Rollover(A\_ROLL)
  + Rollover Involved Crash
  + Other Crash
* Rural/Urban(A\_RU)
  + Rural
  + Urban
  + Unknown
* *Involving Speeding(A\_SPCRA)*
  + Speeding Involved Crash
  + Other Crash

### Vehicle Data

* Vehicle Body Type (A\_BODY)
  + Passenger Car
  + Light Truck - Pickup
  + Light Truck - Utility
  + Light Truck - Van
  + Light Truck - Other
  + Large Truck
  + Motorcycle
  + Bus
  + Other/Unknown
* CDL Status(A\_CDL\_S)
  + Valid
  + Invalid
  + Unknown
  + Not Applicable
* Distracted Driver (A\_DRDIS)
  + Yes-Distracted Driver
  + Other
* Drowsy Driver(A\_DRDRO)
  + Yes-Drowsy Driver
  + Other
* License Compliance(A\_LIC\_C)
  + Valid
  + Invalid
  + Unknown
  + Not Applicable
* License Status(A\_LIC\_S)
  + Valid
  + Invalid
  + Unknown
  + Not Applicable
* Motorcycle License Status(A\_MC\_L\_S)
  + Valid
  + Invalid
  + Unknown
  + Not Applicable
* Speeding Vehicle(A\_SPVEH)
  + Speed Involved
  + No Speed Involved
  + Unknown
  + Not Applicable
* Alcohol Testing(A\_ALCTES)
  + No Alcohol
  + Positive BAC
  + Not Tested
  + Tested, with Unknown Results
  + Unknown if Tested
* Race and Hispanic - Using OMB Guidelines (A\_HRACE)
  + Not Applicable or Not Available
  + Hispanic
  + White Non-Hispanic
  + Black, Non-Hispanic
  + American Indian, Non-Hispanic/Unknown
  + Asian, Non-Hispanic/Unknown
  + Pacific Islander, Non-Hispanic/Unknown
  + Multiple Races, Non-Hispanic/Unknown
  + All Other Non-Hispanic or Race
  + Unknown Race and Unknown Hispanic
* The PERSON Data File The Person data file includes motorist and non-motorist data. It contains the data elements ST\_CASE, STATE, VEH\_NO, and PER\_NO, which are described in the beginning of the Data Element Definitions and Codes section. The Person data file also contains the data elements on the following pages. ST\_CASE, VEH\_NO, and PER\_NO are the unique identifiers for each record. ST\_CASE should be used to merge the Person data file with the Accident data file for a set of all motorists and non-motorists. ST\_CASE and VEH\_NO should be used to merge the Person data file with the Vehicle and Parkwork data files for a set of all motor vehicle occupants. ST\_CASE and PER\_NO should be used to merge the Person data file with non-motorist person-level data files. In the Person data file, motor vehicle occupants are PER\_TYPE = 1, 2, 3, 9. Motor vehicle occupants have assigned vehicle numbers starting with 1. When PER\_TYPE = 3, the occupied vehicle will be found in the PARKWORK data file. Non-motor vehicle occupants are PER\_TYPE = 4, 5, 6, 7, 8, 10 or 19. VEH\_NO = 0 for non-motor vehicle occupants.
* The VEHICLE Data File The Vehicle data file includes in-transport motor vehicle data as well as driver and precrash data. It contains the data elements ST\_CASE, STATE, and VEH\_NO, which are described in the beginning of the Data Element Definitions and Codes section. The Vehicle data file also contains the data elements on the following pages. ST\_CASE and VEH\_NO are the unique identifiers for each record. ST\_CASE should be used to merge the Vehicle data file with the Accident data file. ST\_CASE and VEH\_NO should be used to merge the Vehicle data file with other vehicle-level data files and the Person data file.
* The DISTRACT Data File The Distract data file identifies each driver distraction (as a separate record). It contains the data elements ST\_CASE, STATE, and VEH\_NO, which are described in the beginning of the Data Element Definitions and Codes section. The data file also contains MDRDSTRD which is described below. ST\_CASE, VEH\_NO, and MDRDSTRD are the unique identifiers for each record. ST\_CASE and VEH\_NO should be used to merge the Distract data file with drivers from the Vehicle data file.

# Data Cleaning

* Set the working directory

"C:/Users/Dereje/NSSDS/Testing/Traffic-Fatalities-in-the-US-2016"

## [1] "C:/Users/Dereje/NSSDS/Testing/Traffic-Fatalities-in-the-US-2016"

## Loading Packages

library(shiny)  
library(shinythemes)  
library(readr)  
library(ggplot2)  
library(stringr)  
library(tidyverse)  
library(dplyr)  
library(magrittr)  
library(DT)  
library(tools)

## Importing Data Sets

Distract <- read\_csv("Data/Distract.csv")

## Parsed with column specification:  
## cols(  
## STATE = col\_integer(),  
## ST\_CASE = col\_integer(),  
## VEH\_NO = col\_integer(),  
## MDRDSTRD = col\_integer()  
## )

accident <- read\_csv("Data/accident.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_integer(),  
## TWAY\_ID = col\_character(),  
## TWAY\_ID2 = col\_character(),  
## LATITUDE = col\_double(),  
## LONGITUD = col\_double(),  
## RAIL = col\_character()  
## )

## See spec(...) for full column specifications.

person <- read\_csv("Data/person.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_integer()  
## )  
## See spec(...) for full column specifications.

vehicle <- read\_csv("Data/vehicle.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_integer(),  
## VIN = col\_character(),  
## VIN\_1 = col\_character(),  
## VIN\_2 = col\_character(),  
## VIN\_3 = col\_character(),  
## VIN\_4 = col\_character(),  
## VIN\_5 = col\_character(),  
## VIN\_6 = col\_character(),  
## VIN\_7 = col\_character(),  
## VIN\_8 = col\_character(),  
## VIN\_9 = col\_character(),  
## VIN\_10 = col\_character(),  
## VIN\_11 = col\_character(),  
## VIN\_12 = col\_character(),  
## MCARR\_ID = col\_double(),  
## TRLR1VIN = col\_double(),  
## TRLR2VIN = col\_double(),  
## TRLR3VIN = col\_double()  
## )  
## See spec(...) for full column specifications.

## Warning in rbind(names(probs), probs\_f): number of columns of result is not  
## a multiple of vector length (arg 1)

## Warning: 1640 parsing failures.  
## row # A tibble: 5 x 5 col row col expected actual file expected <int> <chr> <chr> <chr> <chr> actual 1 1389 TRLR1VIN no trailing characters KKVA45256L2 'Data/vehicle.csv' file 2 1389 TRLR2VIN no trailing characters KKVA45286L2 'Data/vehicle.csv' row 3 1394 TRLR1VIN no trailing characters HTNL4841E5G 'Data/vehicle.csv' col 4 1394 TRLR2VIN no trailing characters HTLL4249E5D 'Data/vehicle.csv' expected 5 1516 TRLR1VIN no trailing characters S12E953XYE4 'Data/vehicle.csv'  
## ... ................. ... ...................................................................... ........ ...................................................................... ...... ...................................................................... .... ...................................................................... ... ...................................................................... ... ...................................................................... ........ ......................................................................  
## See problems(...) for more details.

# Merging Vehicle data with Distract data set to capture distraction

vehicle\_distract<-vehicle%>%  
 inner\_join(Distract, by="ST\_CASE","VEH\_NO")

# Nameing columns

acc\_name<-c("STATE", "ST\_CASE", "PERSONS", "COUNTY","CITY","DAY","MONTH","YEAR","DAY\_WEEK",  
"HOUR", "MINUTE","NHS","RUR\_URB", "LATITUDE", "LONGITUD","LGT\_COND", "WEATHER1","WEATHER2", "WEATHER","CF1", "CF2", "CF3", "FATALS", "DRUNK\_DR", "FUNC\_SYS")  
  
per\_name<-c("ST\_CASE","AGE","SEX","RACE","REST\_MIS","DSTATUS","PER\_TYP" ,"REST\_USE", "AIR\_BAG")  
  
vehc\_name<-c("ST\_CASE","UNITTYPE","HIT\_RUN","BODY\_TYP","MAKE","TRAV\_SP","L\_STATE","L\_STATUS","DR\_HGT","DR\_WGT","PREV\_ACC","PREV\_SUS", "PREV\_DWI","PREV\_SPD","PREV\_OTH","SPEEDREL", "DR\_SF1","DR\_SF2","DR\_SF3","DR\_SF4","DR\_DRINK","MDRDSTRD")

# Subsetting the data set with the variables of interest

accident\_2016<-accident %>%   
 select(acc\_name)  
  
vehicle\_2016<-vehicle\_distract %>%  
 select(vehc\_name)  
  
person\_2016<-person %>%  
 select(per\_name)

# merging The data sets in to one master fatality data set

fatality\_2016<-accident\_2016 %>%  
 inner\_join(vehicle\_2016, by="ST\_CASE")  
  
fatality\_2016<-fatality\_2016 %>%  
 inner\_join(person\_2016, by="ST\_CASE")