DATA 607 Assignment 1

Subhalaxmi Rout

Contents

Overview

Basically from the worst driver dataset, we need to find out where is America's worst driver. The given data set shows different states of America and accident percentage based on speed, alcohol, distraction. This data also shows the driver's car insurance and losses incurred by insurance companies for collisions per insured driver.

Below link for the reference. https://fivethirtyeight.com/features/which-state-has-the-worst-drivers/

Load Data into R

```
library(plyr)
library(devtools)
```

```
## Loading required package: usethis
```

```
library(RCurl)

x <-
getURL("https://raw.githubusercontent.com/fivethirtyeight/data/master/bad-drivers/bad-drivers.csv")
worst_driver <-data.frame(read.csv(text=x,  header=FALSE))
dim(worst_driver)
```

```
## [1] 52  8
```

```
head(worst_driver)
```

```
##           V1                                                                 V2
## 1      State Number of drivers involved in fatal collisions per billion miles
## 2    Alabama                                                              18.8
## 3     Alaska                                                              18.1
## 4    Arizona                                                              18.6
## 5   Arkansas                                                              22.4
## 6 California                                                                12
##                                                                    V3
## 1 Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding
## 2                                                                  39
## 3                                                                  41
## 4                                                                  35
## 5                                                                  18
## 6                                                                  35
##                                                                        V4
## 1 Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired
## 2                                                                          30
## 3                                                                          25
## 4                                                                          28
## 5                                                                          26
## 6                                                                          28
##                                                                   V5
## 1 Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted
## 2                                                                  96
## 3                                                                  90
## 4                                                                  84
## 5                                                                  94
## 6                                                                  91
##
## 1 Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accid
## 2
## 3
## 4
## 5
## 6
##                          V7
## 1 Car Insurance Premiums ($)
## 2                    784.55
## 3                   1053.48
## 4                    899.47
## 5                    827.34
## 6                    878.41
##                                                                        V8
## 1 Losses incurred by insurance companies for collisions per insured driver ($)
## 2                                                                 145.08
## 3                                                                 133.93
## 4                                                                 110.35
## 5                                                                 142.39
## 6                                                                 165.63
```

Column names

- States = name of states in USA
- Fatal_collisions_per_billion_mile = Number of drivers involved in fatal collisions per billion mile (National Highway Traffic Safety Administration, 2012)
- Fatal_collision_percent_speeding = Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding(National Highway Traffic Safety Administration, 2009)
- Fatal_collision_percent_alcohol = Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired(National Highway Traffic Safety Administration, 2012)
- Fatal_collision_percent_not_distracted = Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted(National Highway Traffic Safety Administration, 2012)
- Fatal_collision_percent_no_accident = Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accidents(National Highway Traffic Safety Administration, 2012)
- Car_insurence = Car Insurance Premiums(National Association of Insurance Commissioners, 2011)
- Insurance_paid = Losses incurred by insurance companies for collisions per insured driver(National Association of Insurance Commissioners, 2010)

Rename column headers

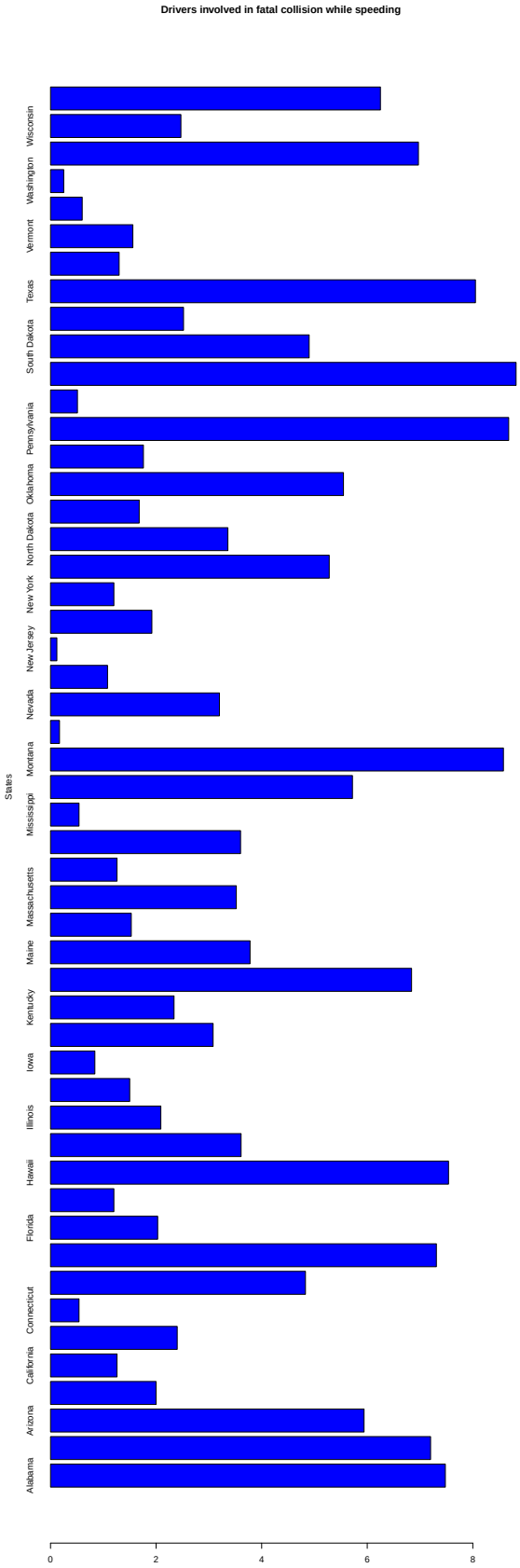
```
colnames(worst_driver) <- c('States','Fatal_collisions_per_billion_mile'
,'Fatal_collision_percent_speeding','Fatal_collision_percent_alcohol'
,'Fatal_collision_percent_not_distracted','Fatal_collision_percent_no_accident'
, 'Car_insurence', 'Insurance_paid')
names(worst_driver)
```

```
## [1] "States"
## [2] "Fatal_collisions_per_billion_mile"
## [3] "Fatal_collision_percent_speeding"
## [4] "Fatal_collision_percent_alcohol"
## [5] "Fatal_collision_percent_not_distracted"
## [6] "Fatal_collision_percent_no_accident"
## [7] "Car_insurence"
## [8] "Insurance_paid"
```

Drivers involved in fatal collision while speeding

```
worst_driver_new  <- as.data.frame(worst_driver[,c(1,2,3)])
colnames(worst_driver_new) <- c("States", "Fatal_collisions_per_billion_mile"
, "Fatal_collision_speeding")
df <- worst_driver_new
df = df[-1,]
View(df)
States <- df$States
miles <- as.integer(df$Fatal_collisions_per_billion_mile)
speed <- (as.integer(df$Fatal_collision_speeding)/100)
collision_speeding <- miles*speed
barplot(collision_speeding,
        names.arg = c("Alabama","Alaska","Arizona","Arkansas","California"
                                            ,"Colorado","Connecticut","Delaware",
"District of Columbia","Florida","Georgia"
```

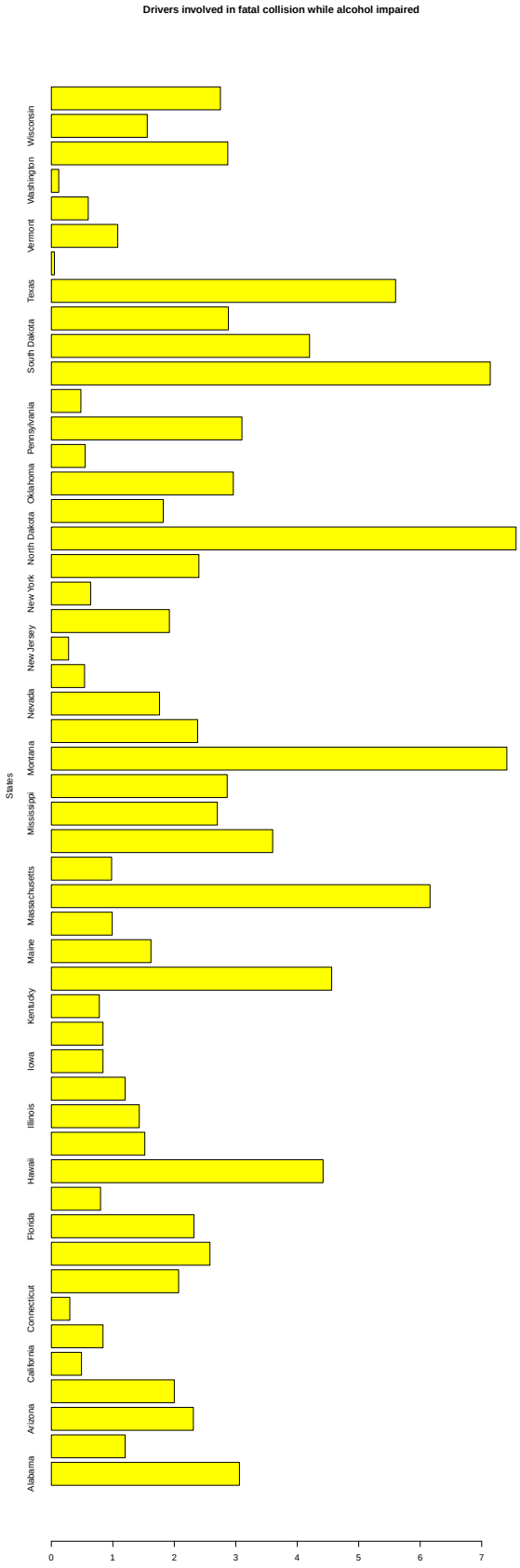
```
,"Hawaii","Idaho","Illinois","Indiana","Iowa"
,"Kansas","Kentucky","Louisiana",
"Maine","Maryland","Massachusetts","Michigan"
,"Minnesota","Mississippi","Missouri"
,"Montana","Nebraska","Nevada",
"New Hampshire","New Jersey","New Mexico","New York","North Carolina","North Dakota","Ohio"
,"Oklahoma","Oregon","Pennsylvania","Rhode Island","South Carolina"
,"South Dakota","Tennessee","Texas","Utah",
"Vermont","Virginia","Washington","West Virginia","Wisconsin","Wyoming"),
ylab = "States",
main = "Drivers involved in fatal collision while speeding",
col = c("blue"),
horiz = TRUE)
```


Drivers involved in fatal collision while speeding
States
Alabama
Arizona
California
Connecticut
Florida
Hawaii
Illinois
Iowa
Kentucky
Maine
Massachusetts
Mississippi
Montana
Nevada
New Jersey
New York
North Dakota
Oklahoma
Pennsylvania
South Dakota
Texas
Vermont
Washington
Wisconsin
0
2
4
6
8

The data available for speeding-related fatalities is from 2009, according to the NHTSA 31% of fatalities occurred while a driver was speeding. In Mississippi, just 15% of traffic fatalities occurred while a driver was speeding, while in Pennsylvania, the share was 50%.

Driver Involved in fatal collision while alcohol impaired

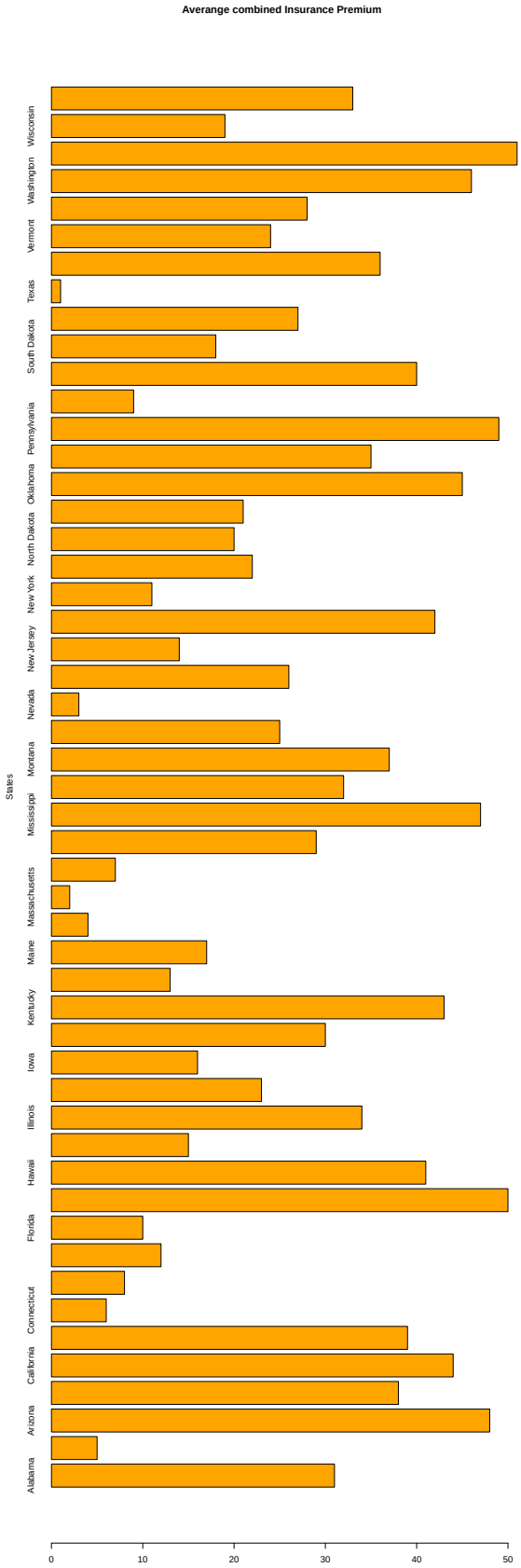
```
worst_driver_new  <- as.data.frame(worst_driver[,c(1,2,4)])
colnames(worst_driver_new) <- c("States", "Fatal_collisions_per_billion_mile"
, "Fatal_collision_alcohol_impaired")
df <- worst_driver_new
df = df[-1,]
View(df)
States <- df$States
miles <- as.integer(df$Fatal_collisions_per_billion_mile)
alcohol_impaired <- (as.integer(df$Fatal_collision_alcohol_impaired)/100)
alcohol_effect <- miles*alcohol_impaired
alcohol_effect
##  [1] 3.06 1.20 2.31 2.00 0.49 0.84 0.30 2.07 2.58 2.32 0.80 4.42 1.52 1.43 1.20
## [16] 0.84 0.84 0.78 4.56 1.62 0.99 6.16 0.98 3.60 2.70 2.86 7.41 2.38 1.76 0.54
## [31] 0.28 1.92 0.64 2.40 7.56 1.82 2.96 0.55 3.10 0.48 7.14 4.20 2.88 5.60 0.05
## [46] 1.08 0.60 0.12 2.87 1.56 2.75
barplot(alcohol_effect, names.arg = c("Alabama","Alaska","Arizona","Arkansas"
,"California","Colorado","Connecticut","Delaware","District of Columbia","Florida"
,"Georgia","Hawaii","Idaho","Illinois","Indiana"
,"Iowa","Kansas","Kentucky","Louisiana","Maine","Maryland","Massachusetts"
,"Michigan","Minnesota","Mississippi","Missouri","Montana","Nebraska","Nevada",
"New Hampshire","New Jersey","New Mexico","New York","North Carolina","North Dakota"
,"Ohio","Oklahoma","Oregon","Pennsylvania","Rhode Island","South Carolina"
,"South Dakota","Tennessee","Texas","Utah",
"Vermont","Virginia","Washington","West Virginia","Wisconsin","Wyoming"),
ylab = "States",
main = "Drivers involved in fatal collision while alcohol impaired",
col = c("yellow"),
horiz = TRUE)
```


Drivers involved in fatal collision while alcohol impaired
States
Alabama
Arizona
California
Connecticut
Florida
Hawaii
Illinois
Iowa
Kentucky
Maine
Massachusetts
Mississippi
Montana
Nevada
New Jersey
New York
North Dakota
Oklahoma
Pennsylvania
South Dakota
Texas
Vermont
Washington
Wisconsin
0
1
2
3
4
5
6
7

Drinking, while driving is irresponsible behavior. 31% percent of fatalities in 2012 occurred while a driver was alcohol-impaired. In Montana, 44% of traffic fatalities that year involved a driver who was alcohol-impaired, but in Utah, it was 16%.

Average combined Insurance Premium

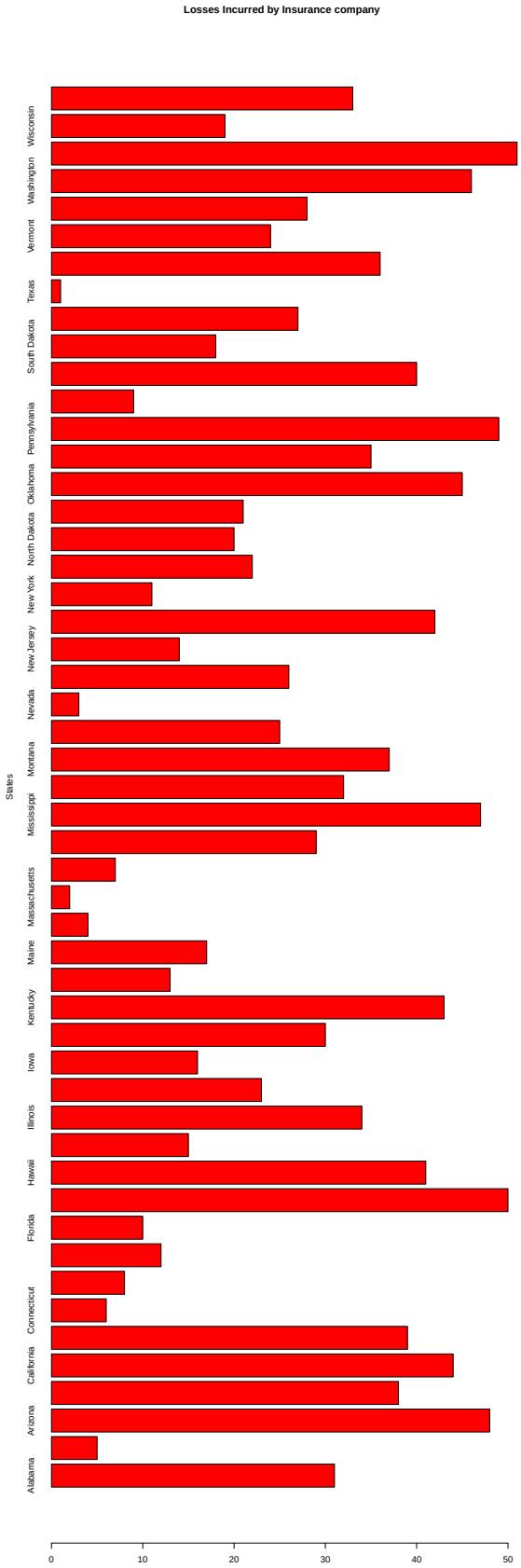
```
worst_driver_new  <- as.data.frame(worst_driver[,c(1,7)])
colnames(worst_driver_new) <- c("States", "Fatal_insurance_premium")
df <- worst_driver_new
df = df[-1,]
View(df)
States <- df$States
car_insurance <- as.integer(df$Fatal_insurance_premium)
barplot(car_insurance, names.arg = c("Alabama","Alaska","Arizona","Arkansas","California"
,"Colorado","Connecticut","Delaware","District of Columbia"
,"Florida","Georgia","Hawaii","Idaho","Illinois"
,"Indiana","Iowa","Kansas","Kentucky","Louisiana","Maine","Maryland"
,"Massachusetts","Michigan","Minnesota","Mississippi","Missouri"
,"Montana","Nebraska","Nevada","New Hampshire","New Jersey","New Mexico"
,"New York","North Carolina","North Dakota","Ohio","Oklahoma","Oregon"
,"Pennsylvania","Rhode Island","South Carolina","South Dakota","Tennessee","Texas","Utah",
"Vermont","Virginia","Washington","West Virginia","Wisconsin","Wyoming"),
ylab = "States",
main = "Averange combined Insurance Premium",
col = c("orange"),
horiz = TRUE)
```


Averange combined Insurance Premium
States
Wisconsin
Washington
Vermont
Texas
South Dakota
Pennsylvania
Oklahoma
North Dakota
New York
New Jersey
Nevada
Montana
Mississippi
Massachusetts
Maine
Kentucky
Iowa
Illinois
Hawaii
Florida
Connecticut
California
Arizona
Alabama
0
10
20
30
40
50

New Jersey found a high-risk driver. The car insurance cost is $1302, which is very expensive comparatively other states. In Idaho, car insurance cost was $641. So, less insurance premium means the low-risk driver and high premium means high-risk drivers.

Losses Incurred by Insurance company

```
worst_driver_new  <- as.data.frame(worst_driver[,c(1,8)])
colnames(worst_driver_new) <- c("States", "Collision_per_insured_driver")
df <- worst_driver_new
View(df)
df = df[-1,]
States <- df$States
collision_per_insured <- as.integer(df$Collision_per_insured_driver)
barplot(car_insurance, names.arg = c("Alabama","Alaska","Arizona","Arkansas"
,"California","Colorado","Connecticut","Delaware","District of Columbia","Florida"
,"Georgia","Hawaii","Idaho","Illinois","Indiana"
,"Iowa","Kansas","Kentucky","Louisiana","Maine","Maryland","Massachusetts"
,"Michigan","Minnesota","Mississippi","Missouri","Montana","Nebraska","Nevada",
"New Hampshire","New Jersey","New Mexico","New York","North Carolina","North Dakota"
,"Ohio","Oklahoma","Oregon","Pennsylvania","Rhode Island"
,"South Carolina","South Dakota","Tennessee","Texas","Utah",
"Vermont","Virginia","Washington","West Virginia","Wisconsin","Wyoming"),
ylab = "States",
main = "Losses Incurred by Insurance company",
col = c("red"),
horiz = TRUE)
```


Losses Incurred by Insurance company
States
Wisconsin
Washington
Vermont
Texas
South Dakota
Pennsylvania
Oklahoma
North Dakota
New York
New Jersey
Nevada
Montana
Mississippi
Massachusetts
Maine
Kentucky
Iowa
Illinois
Hawaii
Florida
Connecticut
California
Arizona
Alabama
0
10
20
30
40
50

Idahoans costing insurers $83 each for collisions in 2010. New Jersians costing insurers $160 for each collision. But Louisiana was $195, the most expensive insurer.

Conclusion

The number of car crashes, even fatal ones, just isn't clear to understand who is and who isn't a bad driver. But we can say that the insurance company thinks that North Carolinians deserve low prices compared to the national average because each of the insured drivers only cost them $127.8 in collision losses in 2010.