Bad Drivers Dataset

Sowmya

9/15/2022

## Section 1: Description of the data

The Bad Drivers data set consists of 51 observations of 8 variables. Each observation contains data about road accidents in a US State. The variables include “State”, “Number of drivers involved in fatal collisions per billion miles”, “Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding”, “Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired”, “Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted”, “Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accidents”, “Car Insurance Premiums ($)" and "Losses incurred by insurance companies for collisions per insured driver ($)”

## Section 2: Reading the data into R

data <- read.csv(("https://raw.githubusercontent.com/fivethirtyeight/data/master/bad-drivers/bad-drivers.csv"))  
head(data)

## State Number.of.drivers.involved.in.fatal.collisions.per.billion.miles  
## 1 Alabama 18.8  
## 2 Alaska 18.1  
## 3 Arizona 18.6  
## 4 Arkansas 22.4  
## 5 California 12.0  
## 6 Colorado 13.6  
## Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Were.Speeding  
## 1 39  
## 2 41  
## 3 35  
## 4 18  
## 5 35  
## 6 37  
## Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Were.Alcohol.Impaired  
## 1 30  
## 2 25  
## 3 28  
## 4 26  
## 5 28  
## 6 28  
## Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Were.Not.Distracted  
## 1 96  
## 2 90  
## 3 84  
## 4 94  
## 5 91  
## 6 79  
## Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Had.Not.Been.Involved.In.Any.Previous.Accidents  
## 1 80  
## 2 94  
## 3 96  
## 4 95  
## 5 89  
## 6 95  
## Car.Insurance.Premiums....  
## 1 784.55  
## 2 1053.48  
## 3 899.47  
## 4 827.34  
## 5 878.41  
## 6 835.50  
## Losses.incurred.by.insurance.companies.for.collisions.per.insured.driver....  
## 1 145.08  
## 2 133.93  
## 3 110.35  
## 4 142.39  
## 5 165.63  
## 6 139.91

## Section 3:Clean the data

sum(is.na(data)) #check for NA

## [1] 0

colnames(data) #get columnnames

## [1] "State"   
## [2] "Number.of.drivers.involved.in.fatal.collisions.per.billion.miles"   
## [3] "Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Were.Speeding"   
## [4] "Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Were.Alcohol.Impaired"   
## [5] "Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Were.Not.Distracted"   
## [6] "Percentage.Of.Drivers.Involved.In.Fatal.Collisions.Who.Had.Not.Been.Involved.In.Any.Previous.Accidents"  
## [7] "Car.Insurance.Premiums...."   
## [8] "Losses.incurred.by.insurance.companies.for.collisions.per.insured.driver...."

colnames(data) <-  
 c("State","Fatal\_collisions", "Percentage\_speeding","percentage\_alcohol\_impared", "Percentage\_no\_distracted", "percentage\_with\_no\_prior\_accidents", "Car\_Insurance\_Premiums", "Losses\_to\_insurance\_companies\_per\_driver")   
colnames(data)

## [1] "State"   
## [2] "Fatal\_collisions"   
## [3] "Percentage\_speeding"   
## [4] "percentage\_alcohol\_impared"   
## [5] "Percentage\_no\_distracted"   
## [6] "percentage\_with\_no\_prior\_accidents"   
## [7] "Car\_Insurance\_Premiums"   
## [8] "Losses\_to\_insurance\_companies\_per\_driver"

#change column names

## Section 4: Characteristics of the data

## `r rows = nrow(data)`  
## `r rows`  
##`r columns = ncol(data)`  
##`r columns`  
"This dataframe has `r rows`rows and `r columns` columns. The names of the columns and a brief description of each are in the table below"

## [1] "This dataframe has `r rows`rows and `r columns` columns. The names of the columns and a brief description of each are in the table below"

Section 5: Summary statistics

summary(data)

## State Fatal\_collisions Percentage\_speeding  
## Length:51 Min. : 5.90 Min. :13.00   
## Class :character 1st Qu.:12.75 1st Qu.:23.00   
## Mode :character Median :15.60 Median :34.00   
## Mean :15.79 Mean :31.73   
## 3rd Qu.:18.50 3rd Qu.:38.00   
## Max. :23.90 Max. :54.00   
## percentage\_alcohol\_impared Percentage\_no\_distracted  
## Min. :16.00 Min. : 10.00   
## 1st Qu.:28.00 1st Qu.: 83.00   
## Median :30.00 Median : 88.00   
## Mean :30.69 Mean : 85.92   
## 3rd Qu.:33.00 3rd Qu.: 95.00   
## Max. :44.00 Max. :100.00   
## percentage\_with\_no\_prior\_accidents Car\_Insurance\_Premiums  
## Min. : 76.00 Min. : 642.0   
## 1st Qu.: 83.50 1st Qu.: 768.4   
## Median : 88.00 Median : 859.0   
## Mean : 88.73 Mean : 887.0   
## 3rd Qu.: 95.00 3rd Qu.:1007.9   
## Max. :100.00 Max. :1301.5   
## Losses\_to\_insurance\_companies\_per\_driver  
## Min. : 82.75   
## 1st Qu.:114.64   
## Median :136.05   
## Mean :134.49   
## 3rd Qu.:151.87   
## Max. :194.78