Lab 7

Alexander Hernandez

09/15/2022

1) National HighWay System Designation Act

```
speed_file = file.choose()
speed_data = read.table(speed_file, sep='\t', header=TRUE)
```

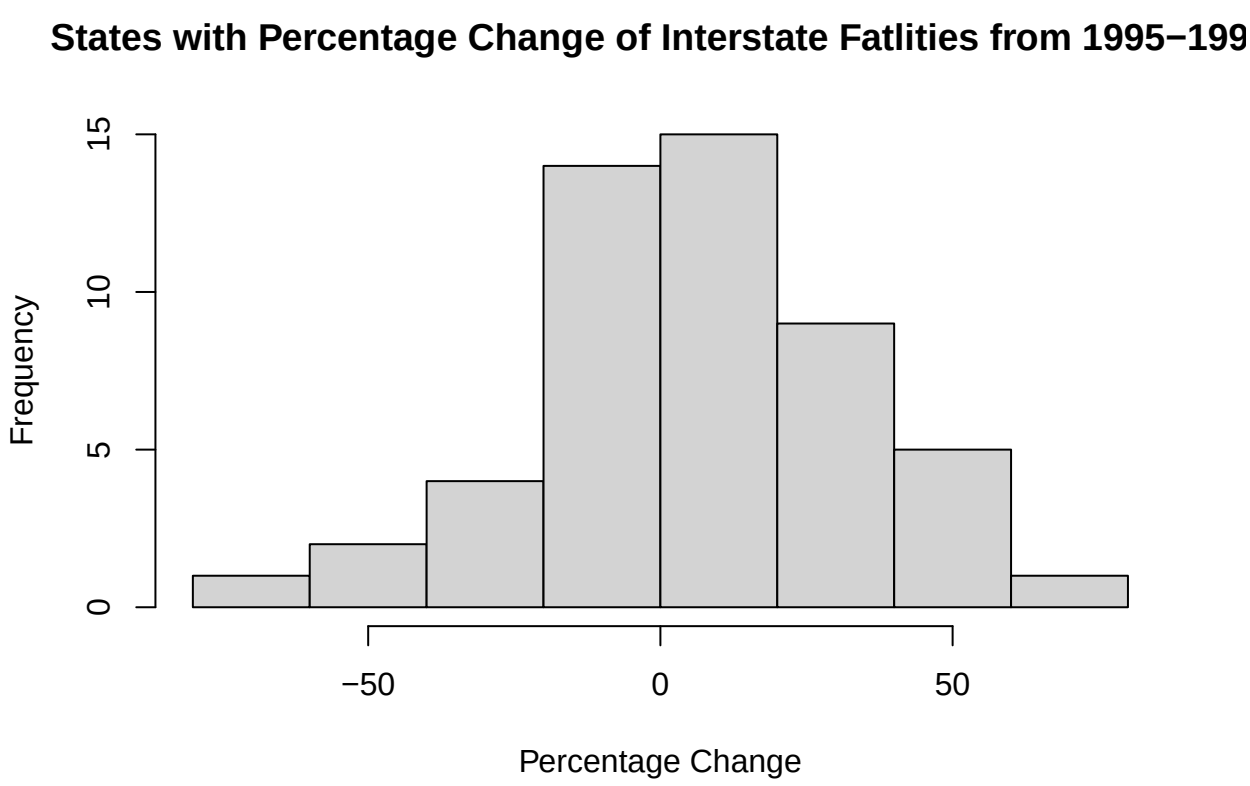
a) Print the first 5 lines of the data

```
speed_data[1:5,]
```

```
##               STATE INCREASe FATALITIESCHANGE
## 1            Alaska       No            -29.0
## 2       Connecticut       No             -4.4
## 3 Dist. of Columbia       No            -80.0
## 4            Hawaii       No            -25.0
## 5           Indiana       No            -13.2
```

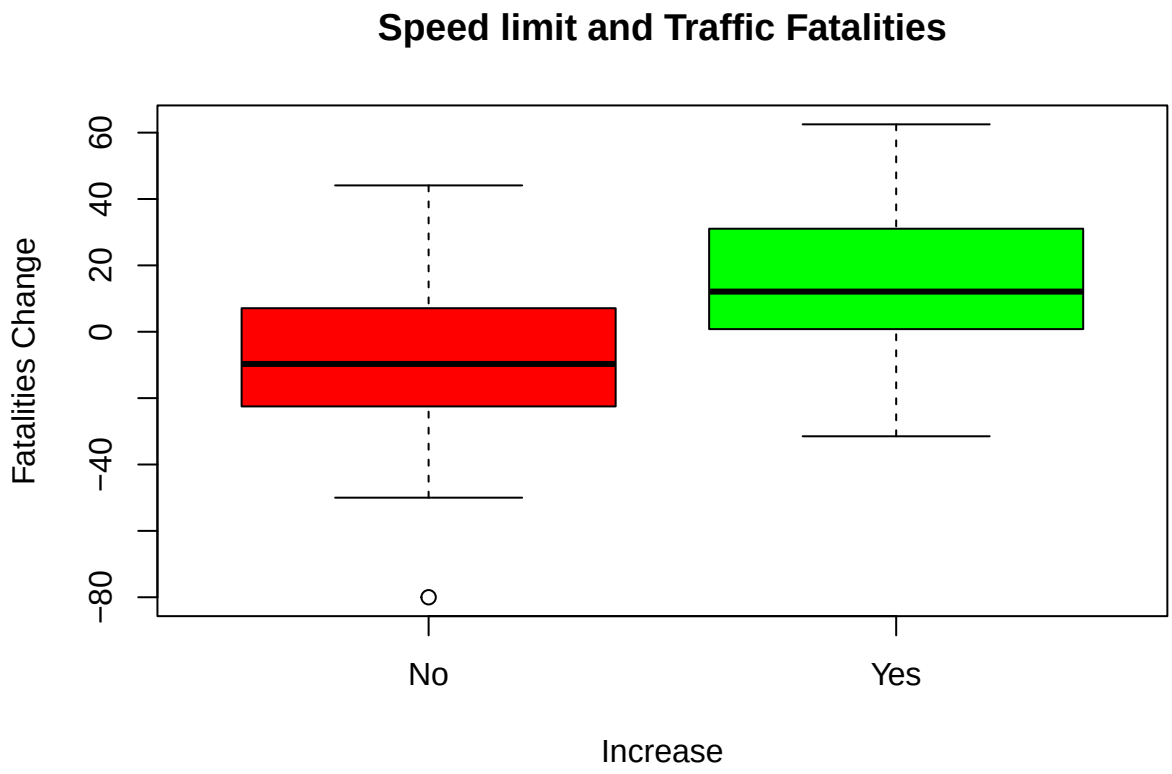
b) Draw the histogram of the %changes fatalities from 95-96

```
hist(speed_data$FATALITIESCHANGE,
     main="States with Percentage Change of Interstate Fatlities from 1995-1996",
     xlab="Percentage Change", ylab="Frequency",
     xlim=range(-80,80))
```



c) Compare speed limit and fatalities with side by side boxplots

```
boxplot(speed_data$FATALITIESCHANGE ~ speed_data$INCREASe,
        col=c("red", "green"),
        main="Speed limit and Traffic Fatalities",
        xlab="Increase", ylab="Fatalities Change")
```



2) Crime Rates for 50 States in 2005

```
# http://datasets.flowingdata.com/crimeRatesByState2005.tsv
```

a) Import the dataset in R and name it 'crime'

```
crime = read.table('http://datasets.flowingdata.com/crimeRatesByState2005.tsv', sep='\t', header=TRUE)
```

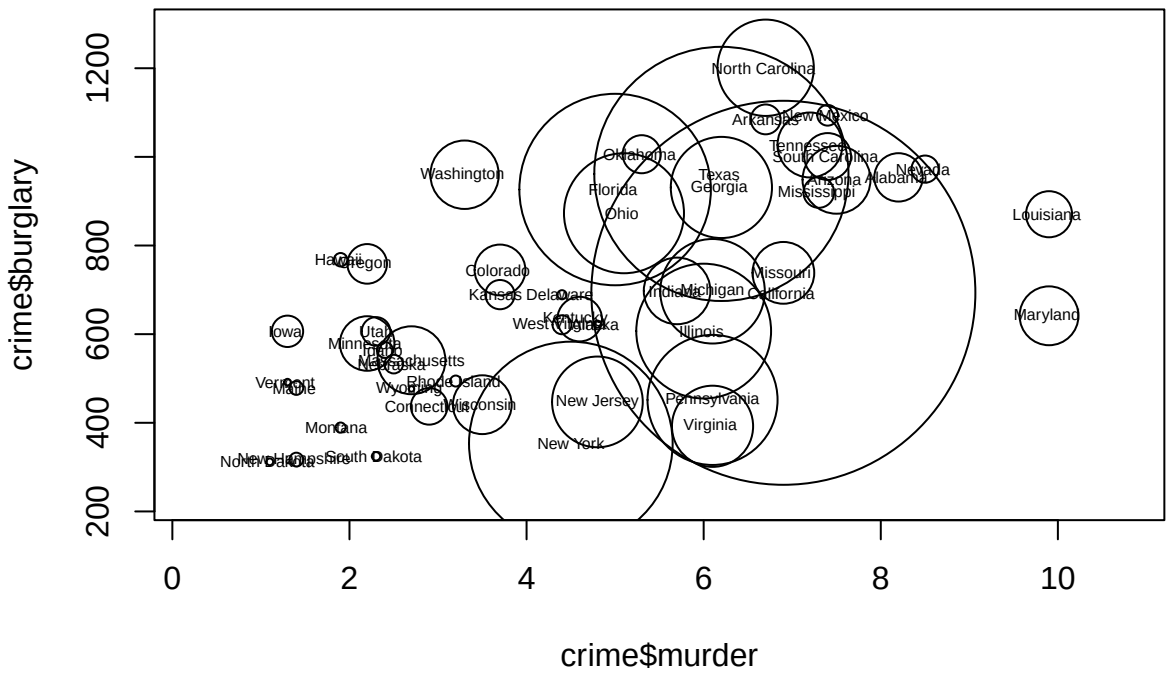
b) How many variables are included in the data

```
ncol(crime)
```

```
## [1] 9
```

c + d) Use code below to draw bubble plots + Add the name of the states using code

```
bubble_plot = symbols(crime$murder, crime$burglary, circles=crime$population)
text(crime$murder, crime$burglary, crime$state, cex=0.5, bg="red")
```



3) Hepatitis

```
#  https://archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/hepatitis.data
```

a) Import the data in R

```
hepatitis = read.table("https://archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/hepatitis.da
```

b) Replace missing values (‘?’) with ‘NA’

```
new_hepatitis = hepatitis
new_hepatitis[new_hepatitis == '?'] = NA
```

c) How many observations contain missing information?

```
sum(is.na(new_hepatitis))
```

```
## [1] 167
```

4) Generate 100 random numbers from a normal distribution with mean 40 and SD 5

```
rnorm(100, mean=40, sd=5)
```

```
##   [1] 41.49080 35.80519 43.14826 44.09761 36.35183 34.84727 34.66580 37.36804
##   [9] 40.34184 40.71316 37.32997 40.28572 37.55897 45.04279 43.83132 33.07987
##  [17] 42.46501 44.63032 57.79478 39.54829 41.31933 39.03895 41.27547 38.16106
##  [25] 43.04461 32.44658 36.36906 43.42939 38.50250 42.92974 35.93219 39.01516
##  [33] 40.05770 36.24910 41.18493 45.13307 42.07047 35.55047 33.09669 36.10549
##  [41] 39.79083 45.77575 52.46373 36.08617 40.23949 29.38989 37.12836 38.01926
##  [49] 40.84117 39.24300 34.52713 43.31179 49.55410 39.44867 38.61474 44.09888
##  [57] 41.00362 36.81348 49.50893 35.10663 37.91504 34.70159 38.03975 34.94875
##  [65] 43.76638 41.47861 46.42036 37.40654 36.88277 36.95302 43.13234 37.94150
##  [73] 41.72921 35.79486 32.71174 45.49792 40.55618 39.65094 43.90598 44.20186
##  [81] 44.89596 41.66720 43.86840 39.75766 39.71557 36.99728 37.96472 41.28260
##  [89] 40.85498 42.41273 38.08200 42.90719 44.57369 47.40352 40.26088 36.55597
##  [97] 41.04387 38.46741 38.13044 47.76598
```

5) Manuscript Data - Import to R without saving and determine dims

```
dim(read.table("http://lib.stat.cmu.edu/jcgs/tu",skip=3, header=TRUE))
```

```
## [1] 136   7
```