R Final Project : Breast Cancer Classification :: Cancer Cases

Utpal Mishra - 20207425

25 December 2020

Import Libraries

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 3.6.3
```

Import Data

```
library(readxl)

data <- read.csv("E:/UCD/Lectures/Semester 1/Data Programming with R/Final Project/CancerCases.csv")
data
```

```
##          Cancer   Cases
## 1        Breast 2261419
## 2          Lung 2206771
## 3    Colorectum 1931590
## 4      Prostate 1414259
## 5       Stomach 1089103
## 6         Liver  905677
## 7  Cervix uteri  604127
## 8 Other cancers 8879843
```

Data Analysis

```
summary(data)
```

```
##           Cancer      Cases
##  Breast       :1   Min.   : 604127
##  Cervix uteri :1   1st Qu.:1043246
##  Colorectum   :1   Median :1672924
##  Liver        :1   Mean   :2411599
##  Lung         :1   3rd Qu.:2220433
##  Other cancers:1   Max.   :8879843
##  (Other)      :2
```

```
describe(data)
```

```
##         vars n      mean         sd    median   trimmed       mad    min
## Cancer*    1 8       4.5       2.45       4.5       4.5      2.97      1
## Cases      2 8 2411598.6 2683588.65 1672924.5 2411598.6 869037.85 604127
##              max   range skew kurtosis       se
## Cancer*        8       7 0.00    -1.65      0.87
## Cases    8879843 8275716 1.67     1.27 948791.87
```

Boxplot

Frequency plot for all the cancer cases being witnessed and as can be seen, breast cancer occupies the second spot.

```
library(ggplot2) #using ggplot2 to plot the frequency plot for the cancer cases
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
p <- ggplot(data, aes(x = Cancer, y = Cases, fill = Cases), main = "Frequency plot for the Cancer Cases", xlab = "Frequency"
, ylab = "Cancer") +  geom_bar(stat="identity") + theme_minimal()
#ggplot(data = data, aes(x = Cancer, y = Cases, fill = Cancer)) + geom_bar(stat="identity") + guides(fill=FALSE)
p
```