Take Home Final Part 2

Saumya Jain, Lauren Nelson

I. Measures of Central Tendency

1. Mean

GDP Per Capita (PPP)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
mean <- mean(MyData$GDP_Per_Capita)
print(mean)

## [1] 28531.16
```

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
mean <- mean(MyData$Male_Life_Expectancy)
print(mean)

## [1] 75.18693
```

2. Median

GDP Per Capita (PPP)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
median <- median(MyData$GDP_Per_Capita)
print(median)

## [1] 22639
```

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
median <- median(MyData$Male_Life_Expectancy)
print(median)

## [1] 74.58
```

3. Mode

GDP Per Capita (PPP)

No Mode

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}
mode <- getmode(MyData$Male_Life_Expectancy)
print(mode)

## [1] 79.6
```

II. Measures of Variation

1. Range

GDP Per Capita (PPP)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
range <- range(MyData$GDP_Per_Capita, na.rm = FALSE)
range1 <- (range[2] - range[1])
print(range1)

## [1] 130947
```

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
range <- range(MyData$Male_Life_Expectancy, na.rm = FALSE)
range1 <- (range[2] - range[1])
print(range1)

## [1] 10.6
```

2. Variance

GDP Per Capita (PPP)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
```

```
variance <- var(MyData$GDP_Per_Capita, na.rm = FALSE)
print(variance)

## [1] 474844538
```

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
variance <- var(MyData$Male_Life_Expectancy, na.rm = FALSE)
print(variance)

## [1] 11.01297
```

3. Standard Deviation

GDP Per Capita (PPP)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
sd <- sd(MyData$GDP_Per_Capita, na.rm = FALSE)
print(sd)

## [1] 21790.93
```

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
sd <- sd(MyData$Male_Life_Expectancy, na.rm = FALSE)
print(sd)

## [1] 3.31858
```

III. Correlation GDP Per Capita (PPP) vs. Male Life Expectancy (Years)

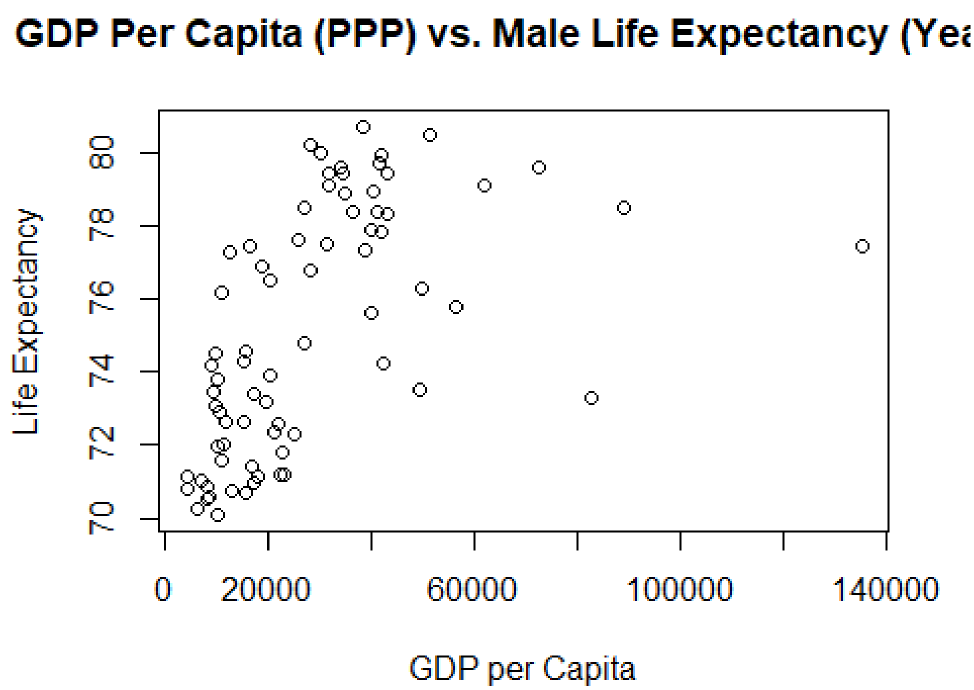
```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
cor.test(Data$GDP_Per_Capita, Data$Male_Life_Expectancy)

##
##  Pearson's product-moment correlation
##
## data:  Data$GDP_Per_Capita and Data$Male_Life_Expectancy
## t = 5.6219, df = 73, p-value = 3.264e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3687105 0.6904932
## sample estimates:
```

```
##       cor
## 0.5496739
```

IV. Scatterplot & Boxplots

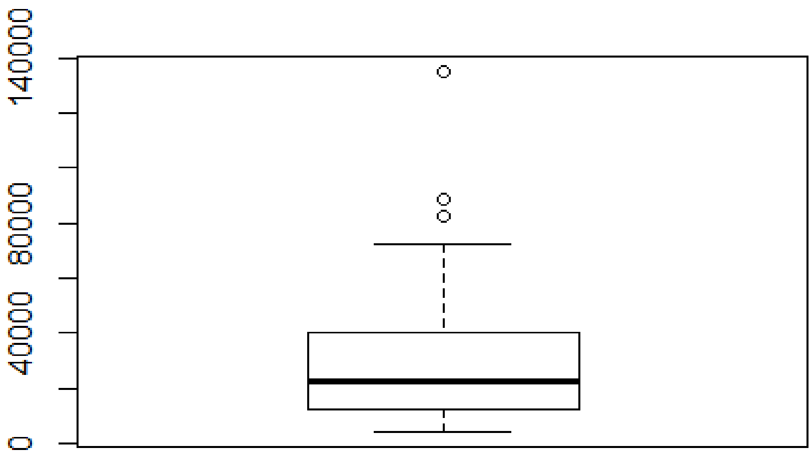
GDP Per Capita (PPP) vs. Male Life Expectancy (Years) Scatterplot

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
plot(Data$GDP_Per_Capita, Data$Male_Life_Expectancy,main="GDP Per Capita (PPP
) vs. Male Life Expectancy (Years)",xlab="GDP per Capita", ylab="Life Expecta
ncy")
```



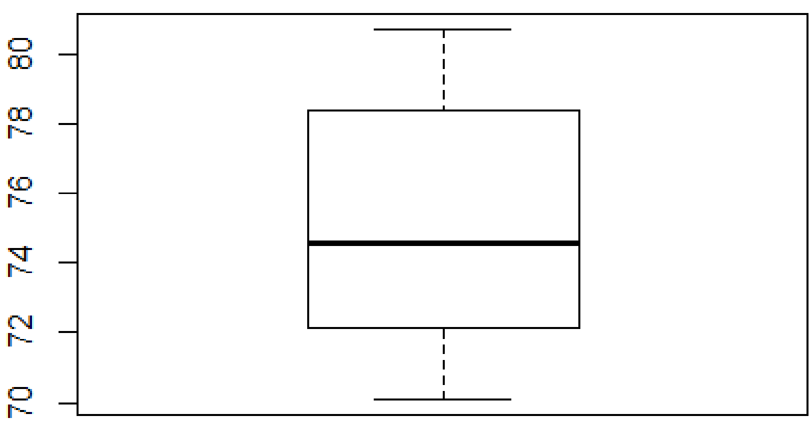
GDP Per Capita (PPP) Boxplot

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
boxplot(Data$GDP_Per_Capita)
```



Male Life Expectancy (Years) Boxplot

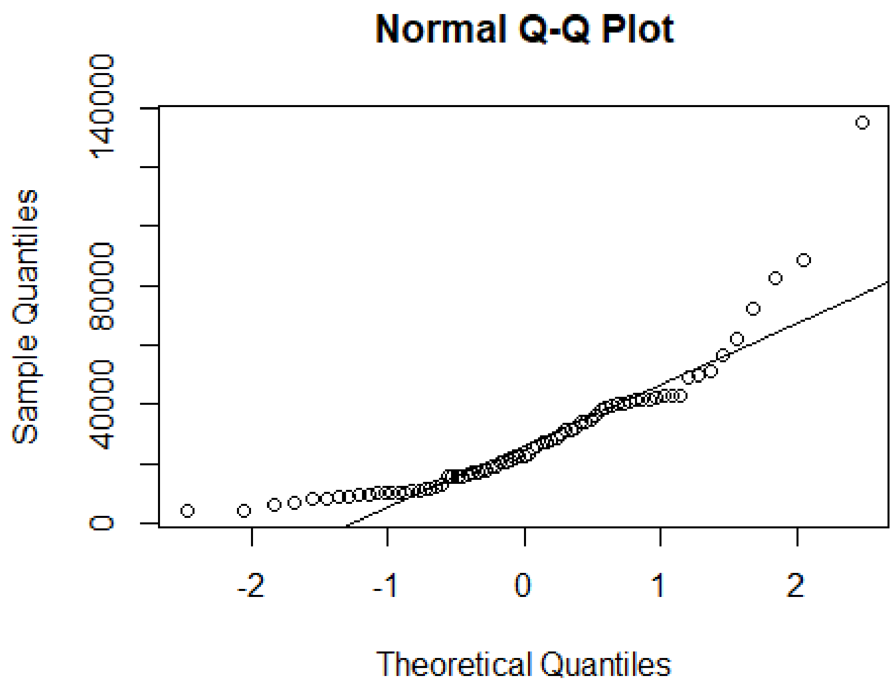
```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
boxplot(Data$Male_Life_Expectancy)
```



V. **Normality Graph with Regression Line**

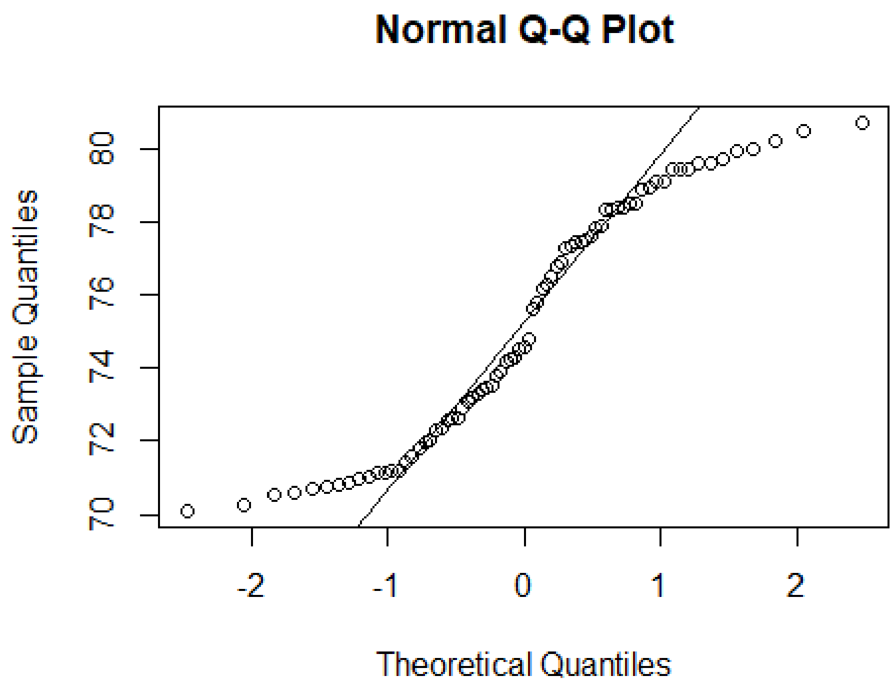
GDP Per Capita (PPP)

```r
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
qqnorm(Data$GDP_Per_Capita)
qqline(Data$GDP_Per_Capita)
```



Male Life Expectancy (Years)

```r
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
qqnorm(Data$Male_Life_Expectancy)
qqline(Data$Male_Life_Expectancy)
```



VI. Outlier Check

GDP Per Capita(PPP)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
FindOutliers <- function(data) {
  lowerq = quantile(data)[2]
  upperq = quantile(data)[3]
  iqr = upperq - lowerq
  extreme.threshold.upper = (iqr * 1.5) + upperq
  extreme.threshold.lower = lowerq - (iqr * 1.5)
  result <- which(data > extreme.threshold.upper | data < extreme.threshold.l
ower)
}

temp <- FindOutliers(Data$GDP_Per_Capita)
testData <- Data[-temp]
```

Outliers: 88848, 135050 (2)

Male Life Expectancy (Years)

```
setwd("C:/Users/Sam/Desktop")
MyData <- read.csv("Data.csv", header = TRUE, sep = ",")
Data <- c(MyData)
FindOutliers <- function(data) {
  lowerq = quantile(data)[2]
  upperq = quantile(data)[3]
  iqr = upperq - lowerq
  extreme.threshold.upper = (iqr * 1.5) + upperq
  extreme.threshold.lower = lowerq - (iqr * 1.5)
  result <- which(data > extreme.threshold.upper | data < extreme.threshold.l
ower)
}

temp <- FindOutliers(Data$Male_Life_Expectancy)
testData <- Data[-temp]
```

Outliers: None