first-sample-code.R

ANTARA GHOSH (2037031)

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'readr' was built under R version 4.0.4
```

```
## Warning: package 'purrr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'stringr' was built under R version 4.0.4
```

```
## Warning: package 'forcats' was built under R version 4.0.4
```

```
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.0.5
```

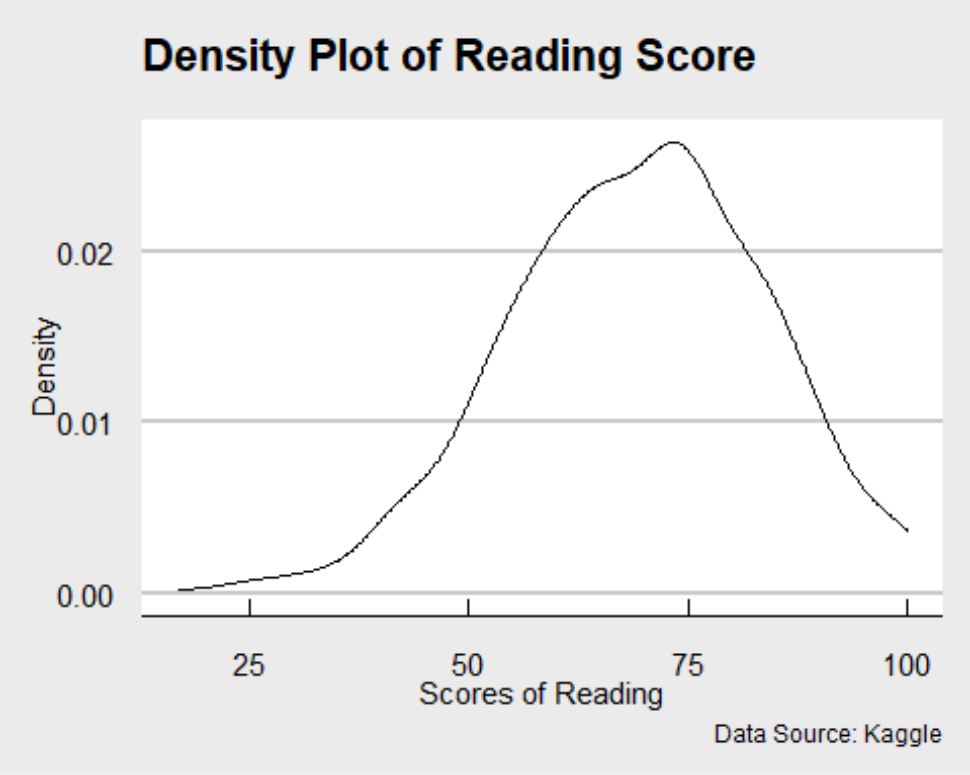
```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.5
```

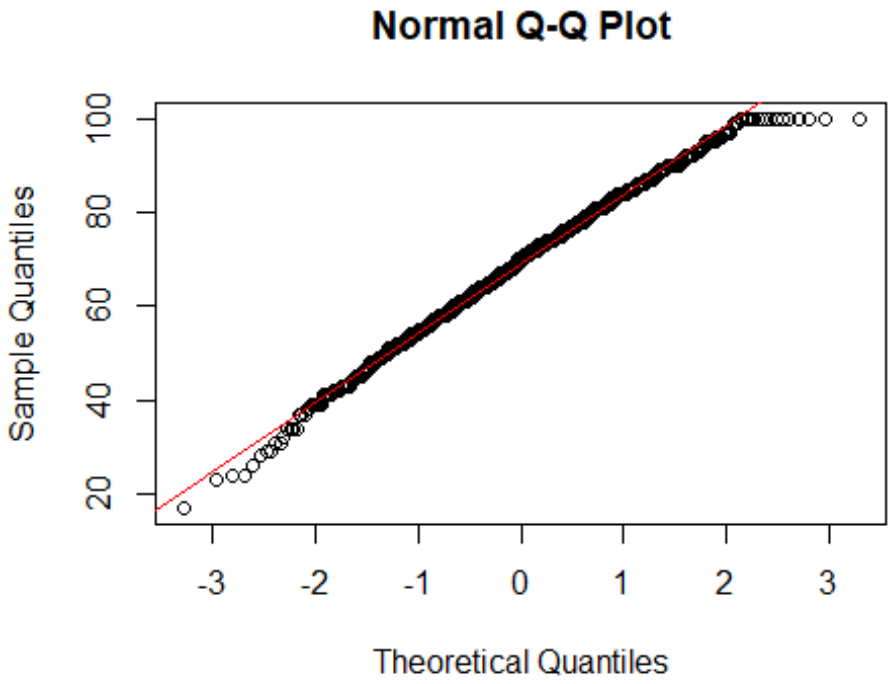
```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

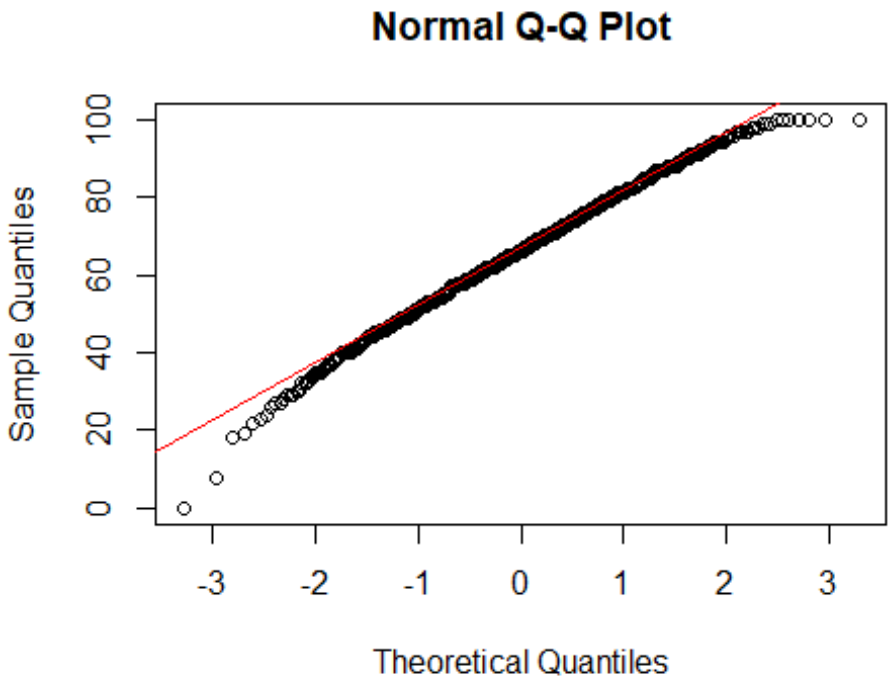
```
## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some

df = read.csv("Data.csv")
# Testing Normality

fig2 = df %>%
  ggplot()+geom_density(aes(x=reading.score))+
  labs(title = 'Density Plot of Reading Score ',
       subtitle = '',
       x= 'Scores of Reading',
       y='Density',
       caption = 'Data Source: Kaggle')+
  theme_economist_white()
fig2
```




```
qqnorm(df$reading.score);qqline(df$reading.score, col = "red")
```




```r
qqnorm(df$math.score);qqline(df$math.score, col = "red")
```




```r
#Data Visualisation
```

```r
#T Test
t.test(df$math.score~df$gender)
```

```
##
##  Welch Two Sample t-test
##
## data:  df$math.score by df$gender
## t = -5.398, df = 997.98, p-value = 8.421e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.947209 -3.242813
## sample estimates:
## mean in group female   mean in group male
##             63.63320             68.72822
```

```r
t.test(df$reading.score~df$gender)
```

```
##
##  Welch Two Sample t-test
##
## data:  df$reading.score by df$gender
## t = 7.9684, df = 996.36, p-value = 4.376e-15
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  5.377941 8.892218
## sample estimates:
## mean in group female   mean in group male
##             72.60811             65.47303
```

```r
#ANOVA Test

anova.test2 = aov(reading.score~gender, data = df)
summary(anova.test2)
```

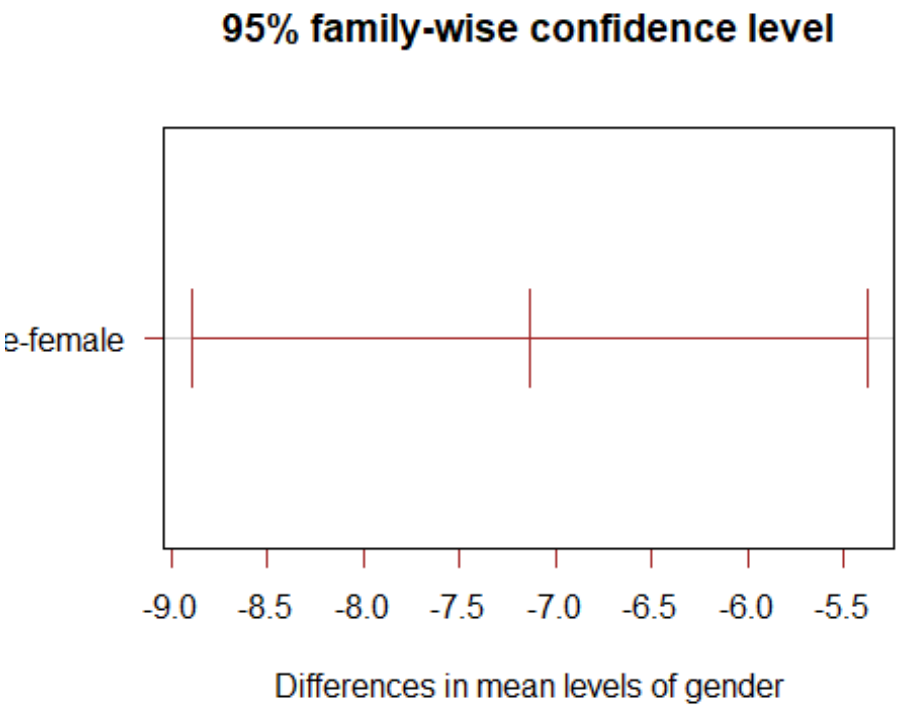
```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## gender        1  12711   12711   63.35 4.68e-15 ***
## Residuals   998 200242     201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova.test3 = aov(writing.score~gender, data = df)
summary(anova.test3)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## gender        1  20931   20931   99.59 <2e-16 ***
## Residuals   998 209746     210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Tukey HSD Test
```

```r
tukey = TukeyHSD(anova.test2)
plot(tukey , las=1 , col="brown")
```




```r
#Mann Whiteny U test


ManU2 = wilcox.test(reading.score~gender, data = df)
ManU2

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  reading.score by gender
## W = 160508, p-value = 5.374e-15
## alternative hypothesis: true location shift is not equal to 0

ManU3 = wilcox.test(writing.score~gender, data = df)
ManU3

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  writing.score by gender
## W = 169957, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0

#Kruskal Test
```

```r
kruskal.test(reading.score~race.ethnicity, data = df)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  reading.score by race.ethnicity
## Kruskal-Wallis chi-squared = 21.354, df = 4, p-value = 0.0002694
```

```r
kruskal.test(writing.score~race.ethnicity, data = df)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  writing.score by race.ethnicity
## Kruskal-Wallis chi-squared = 26.609, df = 4, p-value = 2.385e-05
```