Assignment 3

Shayan

2024-09-05

Importing the data

```
# Load the dataset into R, replace 'assign.csv' with the actual file name
a3 <- read.csv("assign.csv")



# Load any necessary libraries (if needed)
# install.packages("dplyr")

 library(dplyr)

## Warning: package 'dplyr' was built under R version 4.4.1

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

 library(ggplot2)

#
```

Question 1

Null hypothesis in sentence form:

Null Hypothesis (H_0): There is no significant relationship between screen time and BMI among children in the dataset.

Alternative Hypothesis (H_1): There is a significant relationship between screen time and BMI among children in the dataset.

Null hypothesis in symbolic form:

H_0: $\rho = 0$ (where ρ represents the population correlation coefficient between screen time and BMI)
H_1: $\rho \neq 0$

Justification:

Objective: In this analysis it was wanted to check the difference between two groups based on the sex (Female and Male) in the BMI (Body Mass Index). First of all, we were interested in whether the mean BMI of females is significantly different from the mean BMI of males.

Test Conducted: We applied Welch Two Sample t test and this is suitable when comparing the mean of two independent samples where the two samples may have different variances and different sample sizes.

Results:

Test Statistic (t): -0. 7482 Degrees of Freedom (df): 338 Ahistoric appearance of the Big Dipper is a notion rooted in the literati tradition of viewing the sky as a large opened book. 59 p-value: 0. 4549 95% Confidence Interval for the Difference in Means: [-0. 6371, 0. 2860] Mean BMI for Females: 16 Cutting down on Soda and junk foods. 16108 Mean BMI for Males: 16. 33665

Interpretation:

p-value: The value of p is 0. Correspondingly, 4549 is much greater than the conventional alpha level of 0. 05. This implies that mean BMI of the two groups, that is, females and males, are not significantly different. In other words, the results of comparing mean BMI in females and males is insignificant suggesting that there is no high significance in the difference.

Confidence Interval: It is also possible to achieve a 95% confidence interval of the difference in the mean BMI with the values being - 0. 6371 to 0. 2860. Ct since this interval contains zero it thus corroborates the conclusion that cannot reject the null hypothesis of no difference between the two groups. If it was greater than zero or less than zero but not equal to zero then there is a difference.

Effect Size: It can also be ascertained that both the means of the two groups are fairly close, with an observed mean of $16. 16108 for females and an observed mean of $16. 33665 for males thus the difference in the means is indeed small. This coincides with the statistical result that indicated that there is no difference.

Question 2

```
summary(a3)
```

```
##       id                 sex                pain                 age       
##  Length:394         Length:394         Length:394         Min.   : 6.000  
##  Class :character   Class :character   Class :character   1st Qu.: 7.000  
##  Mode  :character   Mode  :character   Mode  :character   Median : 9.000  
##                                                           Mean   : 8.881  
##                                                           3rd Qu.:10.000  
##                                                           Max.   :12.000  
##      height          weight           BMI           physical     
##  Min.   :107.5   Min.   :16.20   Min.   :12.20   Min.   : 0.000  
##  1st Qu.:122.2   1st Qu.:22.82   1st Qu.:14.80   1st Qu.: 2.000  
##  Median :131.0   Median :27.15   Median :15.80   Median : 4.000  
##  Mean   :132.4   Mean   :29.08   Mean   :16.25   Mean   : 5.478  
##  3rd Qu.:142.2   3rd Qu.:33.48   3rd Qu.:17.10   3rd Qu.: 7.875  
##  Max.   :173.2   Max.   :77.90   Max.   :30.20   Max.   :22.000  
##      screen          sleep        visual_analog        distance      
##  Min.   :0.000   Min.   : 6.000   Min.   : 0.0000   Min.   :-12.900  
##  1st Qu.:1.000   1st Qu.: 8.000   1st Qu.: 0.0000   1st Qu.:  0.925  
##  Median :2.000   Median : 9.000   Median : 0.0000   Median :  4.900  
##  Mean   :1.755   Mean   : 8.758   Mean   : 0.5492   Mean   :  4.911  
##  3rd Qu.:2.000   3rd Qu.: 9.000   3rd Qu.: 0.0000   3rd Qu.:  8.975  
##  Max.   :9.000   Max.   :11.000   Max.   :10.0000   Max.   : 26.200  
##     strength          tilt       
##  Min.   :20.00   Min.   : 3.00  
##  1st Qu.:25.15   1st Qu.:12.80  
##  Median :30.80   Median :15.75
```

```
##  Mean   :33.52   Mean   :15.63
##  3rd Qu.:40.00   3rd Qu.:18.60
##  Max.   :76.80   Max.   :30.30
```

```
summary(a3$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.000   7.000   9.000   8.881  10.000  12.000
```

```
summary(a3$height)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   107.5   122.2   131.0   132.4   142.2   173.2
```

```
summary(a3$weight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.20   22.82   27.15   29.08   33.48   77.90
```

```
summary(a3$BMI)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.20   14.80   15.80   16.25   17.10   30.20
```

```
# Summary statistics for physical activity, screen time, and sleep
summary(a3$physical)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.000   4.000   5.478   7.875  22.000
```

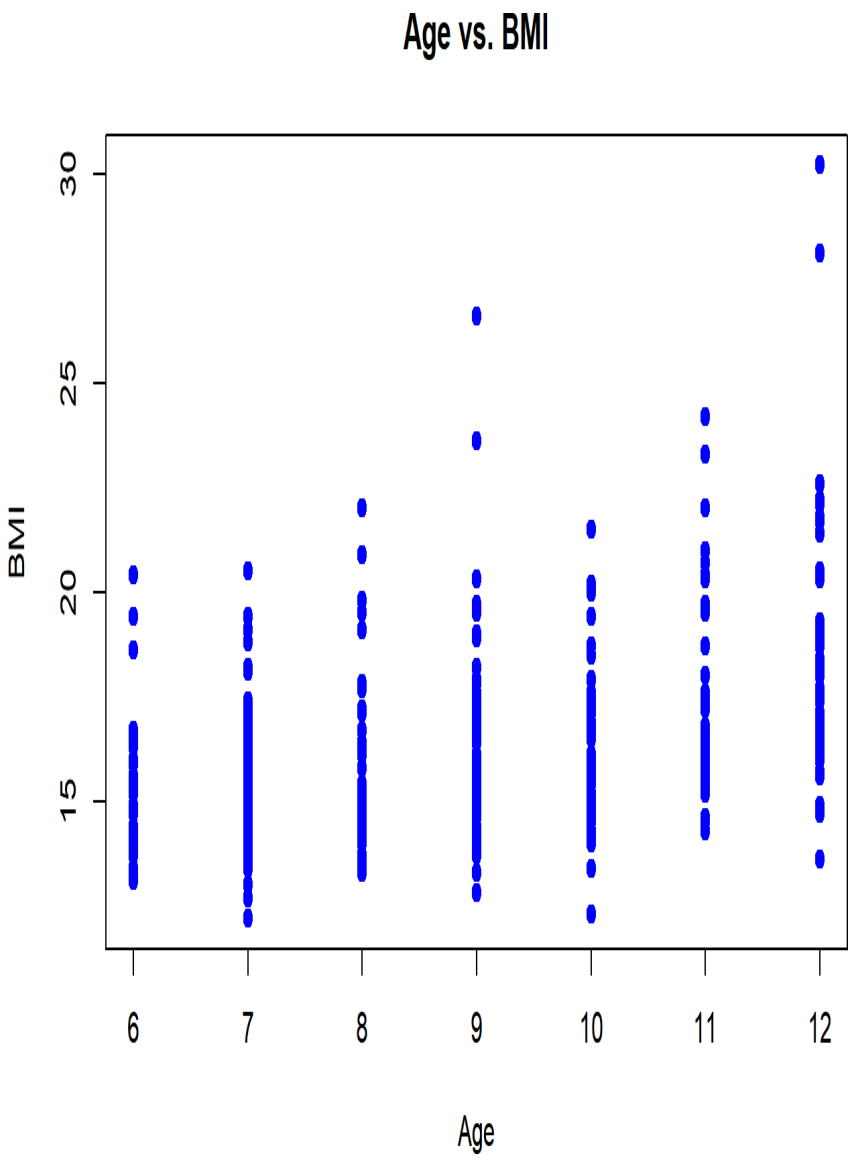
```
summary(a3$screen)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   2.000   1.755   2.000   9.000
```

```
summary(a3$sleep)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.000   8.000   9.000   8.758   9.000  11.000
```

```
# Scatter plot of Age vs. BMI
plot(a3$age, a3$BMI, main="Age vs. BMI", xlab="Age", ylab="BMI", pch=19, col="blue")
```

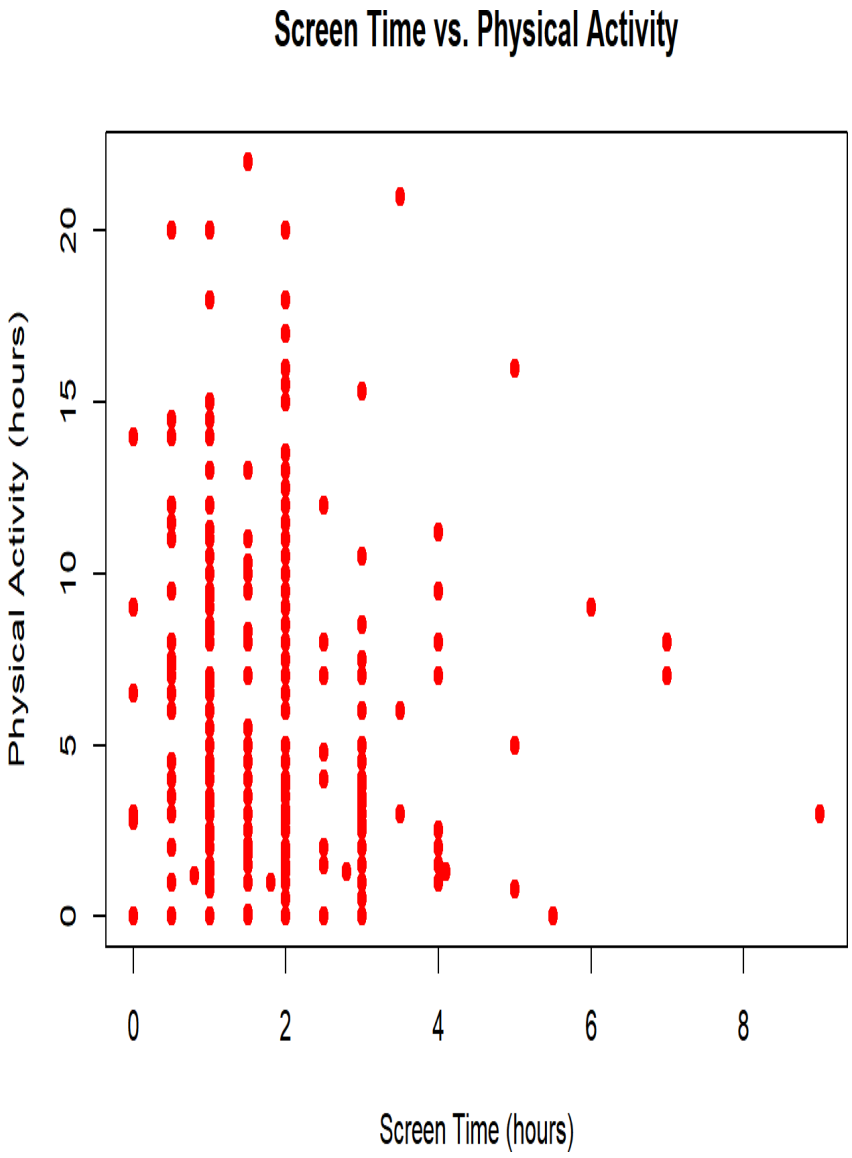



```
# Correlation between Age and BMI
cor.test(a3$age, a3$BMI)

##
##  Pearson's product-moment correlation
##
## data:  a3$age and a3$BMI
## t = 8.4317, df = 392, p-value = 6.59e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3048198 0.4723290
## sample estimates:
##       cor
## 0.3918166

# Scatter plot of Screen Time vs. Physical Activity
```

```
plot(a3$screen, a3$physical, main="Screen Time vs. Physical Activity", xlab="Screen
Time (hours)", ylab="Physical Activity (hours)", pch=19, col="red")
```



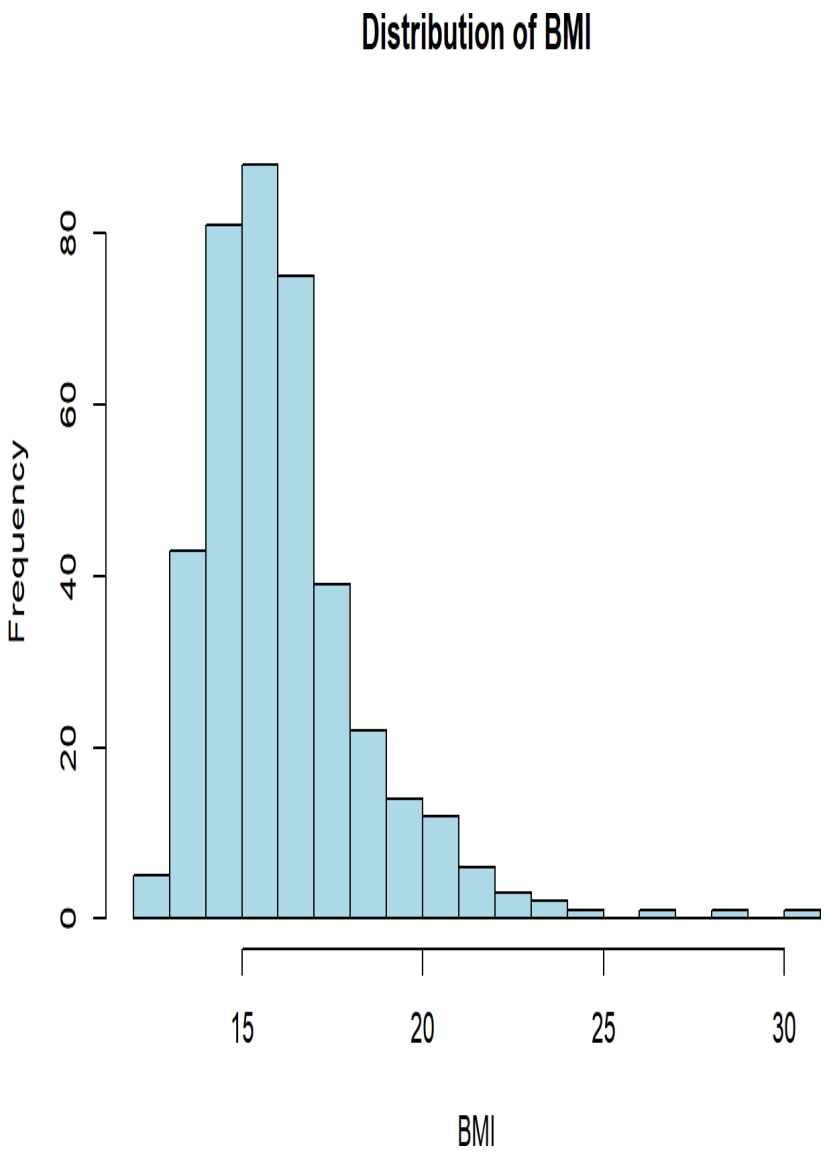

```
# Correlation between Screen Time and Physical Activity
cor.test(a3$screen, a3$physical)

##
##  Pearson's product-moment correlation
##
## data:  a3$screen and a3$physical
## t = -1.9053, df = 392, p-value = 0.05747
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.192763184  0.003033859
## sample estimates:
##         cor
```

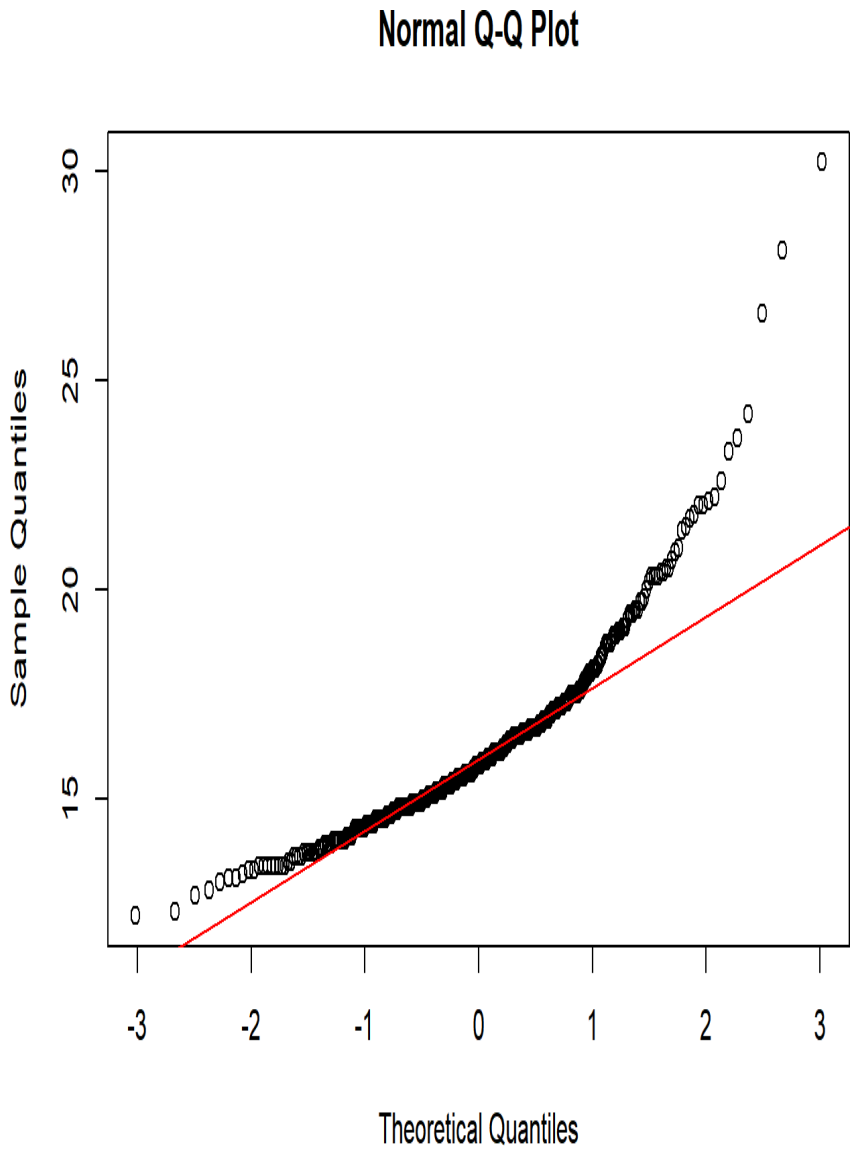
```
## -0.09579116
```

Question 3

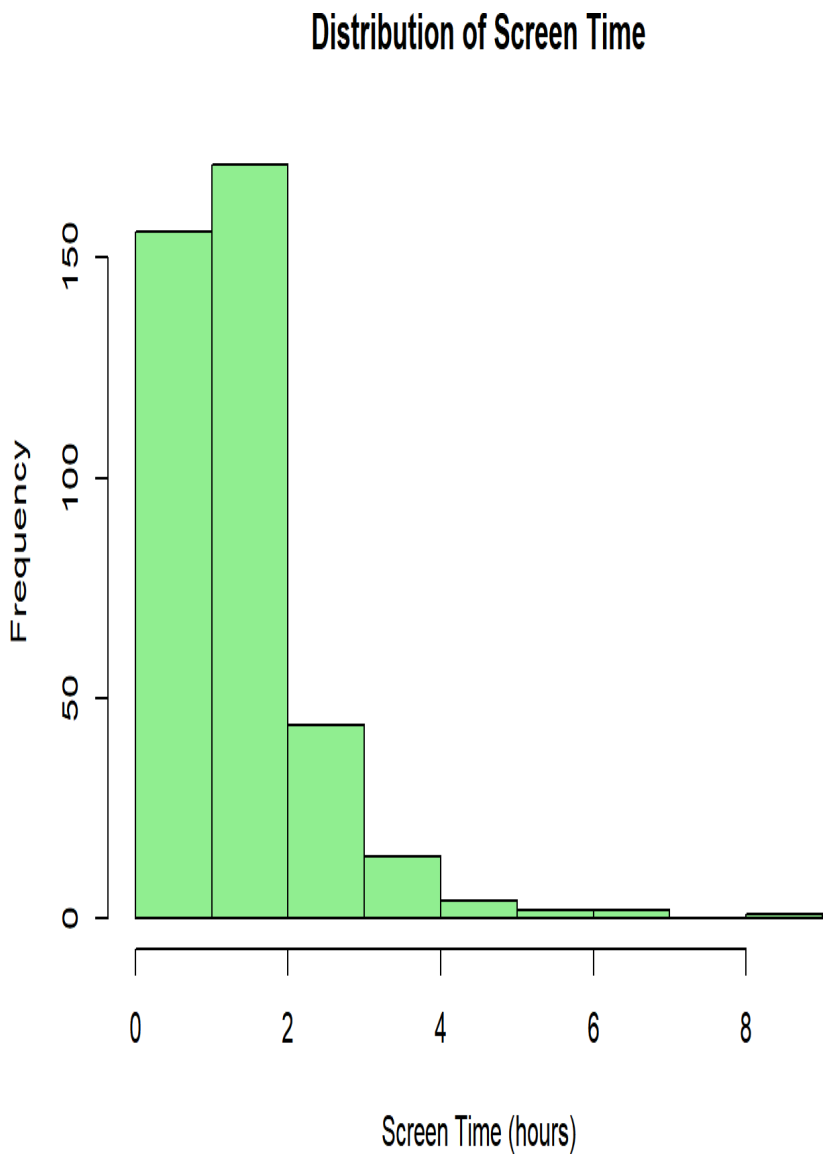
```
# conduct test
# Histogram of BMI
hist(a3$BMI, main="Distribution of BMI", xlab="BMI", col="lightblue", breaks=20)
```



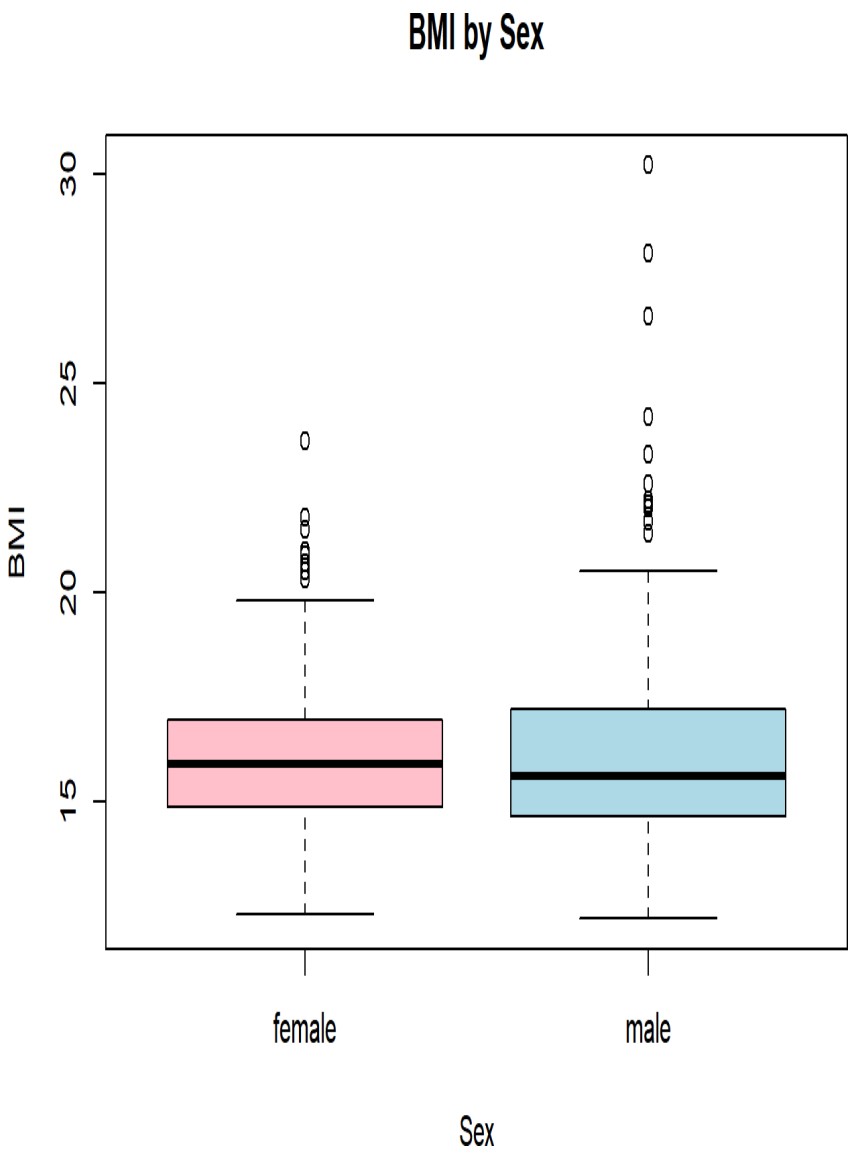

```
# Q-Q plot to check normality
qqnorm(a3$BMI)
qqline(a3$BMI, col="red")
```




```
# Histogram of Screen Time
hist(a3$screen, main="Distribution of Screen Time", xlab="Screen Time (hours)",
col="lightgreen", breaks=10)
```



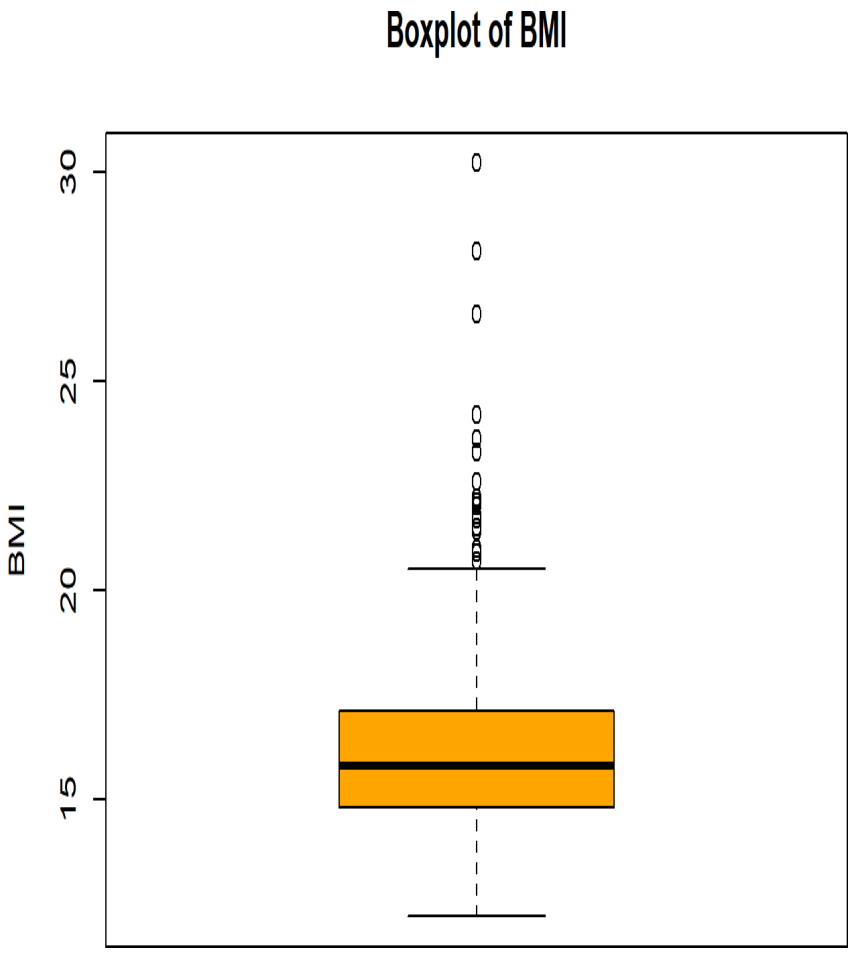

```
# Boxplot of BMI by Sex
boxplot(a3$BMI ~ a3$sex, main="BMI by Sex", xlab="Sex", ylab="BMI", col=c("pink",
"lightblue"))
```




```
# t-test to compare BMI between boys and girls
t.test(BMI ~ sex, data=a3)

##
##  Welch Two Sample t-test
##
## data:  BMI by sex
## t = -0.7482, df = 338.59, p-value = 0.4549
## alternative hypothesis: true difference in means between group female and group male
is not equal to 0
## 95 percent confidence interval:
##  -0.6371239  0.2859929
## sample estimates:
## mean in group female   mean in group male
##              16.16108             16.33665
```

```
# Boxplot to detect outliers in BMI
boxplot(a3$BMI, main="Boxplot of BMI", ylab="BMI", col="orange")
```




```
# Identify the outliers
outliers <- a3$BMI[a3$BMI > quantile(a3$BMI, 0.75) + 1.5 * IQR(a3$BMI) | a3$BMI <
quantile(a3$BMI, 0.25) - 1.5 * IQR(a3$BMI)]
outliers
```

```
##  [1] 24.2 21.8 22.0 20.9 20.7 21.5 23.3 23.6 26.6 28.1 22.2 21.0 30.2 22.1 22.6
## [16] 22.0 21.7 21.4
```

```
# Multiple linear regression
model <- lm(BMI ~ age + screen + physical + sleep, data=a3)
summary(model)
```

```
##
```

```
## Call:
## lm(formula = BMI ~ age + screen + physical + sleep, data = a3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5550 -1.3377 -0.4567  0.8157 12.5724
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.810445   1.916285   8.251 2.47e-15 ***
## age          0.396051   0.065436   6.053 3.36e-09 ***
## screen       0.248520   0.100804   2.465   0.0141 *
## physical    -0.009738   0.024181  -0.403   0.6874
## sleep       -0.395563   0.175261  -2.257   0.0246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.094 on 389 degrees of freedom
## Multiple R-squared:  0.1811, Adjusted R-squared:  0.1727
## F-statistic: 21.51 on 4 and 389 DF,  p-value: 4.814e-16
```