Q2

Aaron Krishnapillai

1/26/2023

a)

```
directory <- "C:/Users/Aaron/OneDrive/Documents/TERM 3A/STAT 341/A1/data"
dirsep <-"/"
filename <- paste(directory, "Iris.csv", sep=dirsep)
data <- read.csv(filename, header=TRUE)
dim(data)
```

```
## [1] 150   5
```

We have 150 rows and 5 columns

b)

```
knitr::kable(table(data$Species), col.names = c('Species', 'Freq'))
```

Species	Freq
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

c)

```
unique(data[data$SepalWidth == min(data$SepalWidth),]$Species)
```

```
## [1] "Iris-versicolor"
```

The iris with the smallest sepal width is Iris-versicolor

```
unique(data[data$SepalWidth == max(data$SepalWidth),]$Species)
```

```
## [1] "Iris-setosa"
```

The iris with the largest sepal width is Iris-setosa d)

```
knitr::kable(sapply(unique(data$Species),
                    function(x) mean(data[data$Species == x,]$SepalLength)),
             col.names = c( 'Average'))
```

	Average
Iris-setosa	5.006
Iris-versicolor	5.936
Iris-virginica	6.588

e)

i.)

```
data$PetalRatio <- data$PetalWidth/data$PetalLength
unique(data[data$PetalRatio == min(data$PetalRatio),])
```

```
##    SepalLength SepalWidth PetalLength PetalWidth     Species PetalRatio
## 10         4.9        3.1         1.5        0.1 Iris-setosa 0.06666667
## 33         5.2        4.1         1.5        0.1 Iris-setosa 0.06666667
```

The iris with the smallest Petal Ratio is Iris-setosa

```
data$Species[data$PetalRatio == max(data$PetalRatio)]
```

```
## [1] "Iris-virginica"
```

The iris with the largest Petal Ratio is Iris-virginica

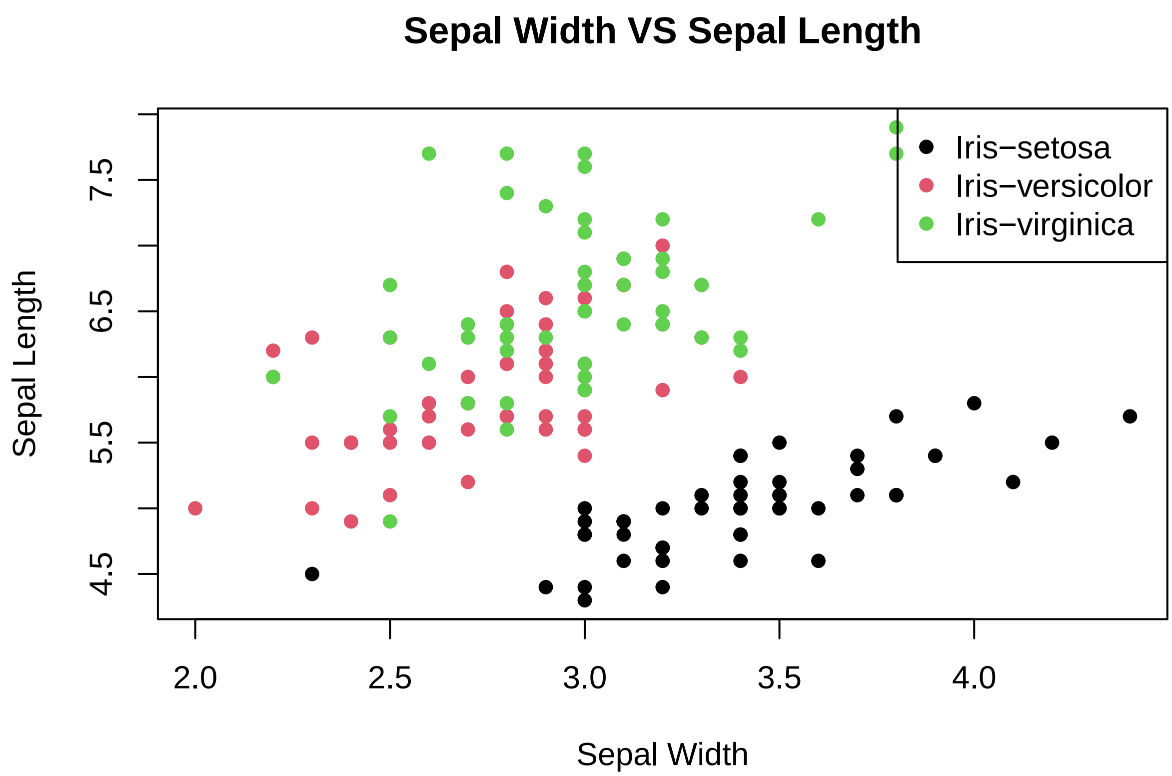
ii.)

```
knitr::kable(table(data$Species[data$PetalRatio > 0.3])
             /table(data$Species),
             col.names = c('Species', 'Proportion over 0.3'))
```

Species	Proportion over 0.3
Iris-setosa	0.04
Iris-versicolor	0.66
Iris-virginica	0.86

f)

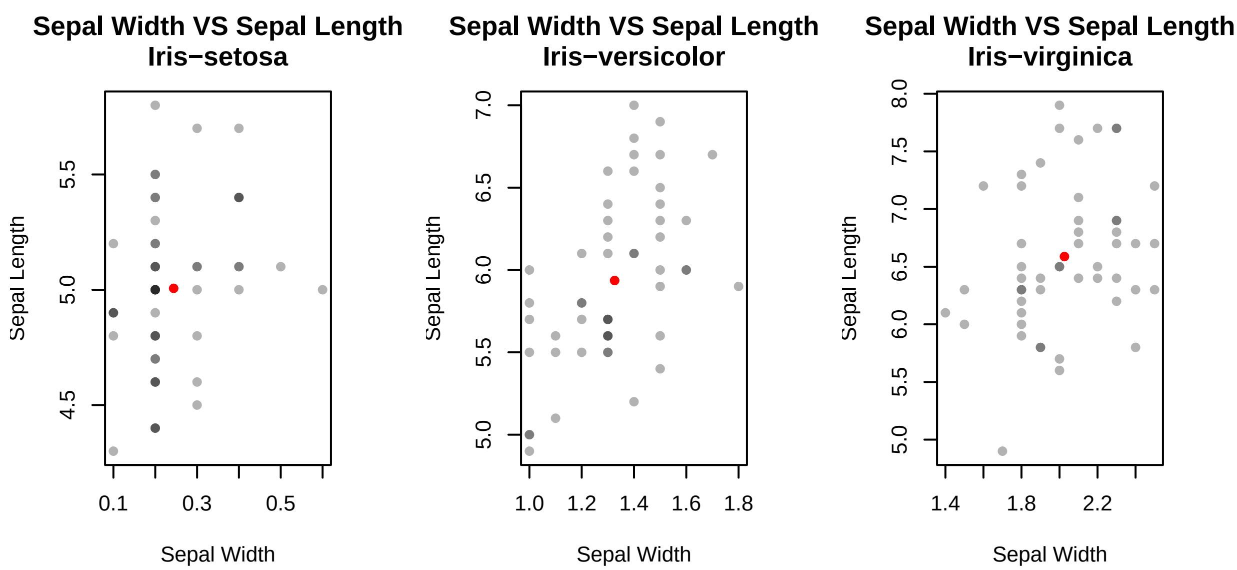
```r
plot(data$SepalWidth, data$SepalLength,
     col = factor(data$Species),
     pch = 16,
     main = "Sepal Width VS Sepal Length",
     xlab = "Sepal Width",
     ylab = 'Sepal Length')

legend("topright",
       legend = levels(factor(data$Species)),
       pch = 16,
       col = factor(levels(factor(data$Species))))
```



We can see that all of the species have are following a positive linear relationship. As the sepal width increases, on average so does the sepal lenth for each species. The rate of increase seems to be different for each species. Iris-Virginica seems to grow at a faster rate than the others as most points are have a higher Sepal length while also maintaining a lower sepal width than the others.

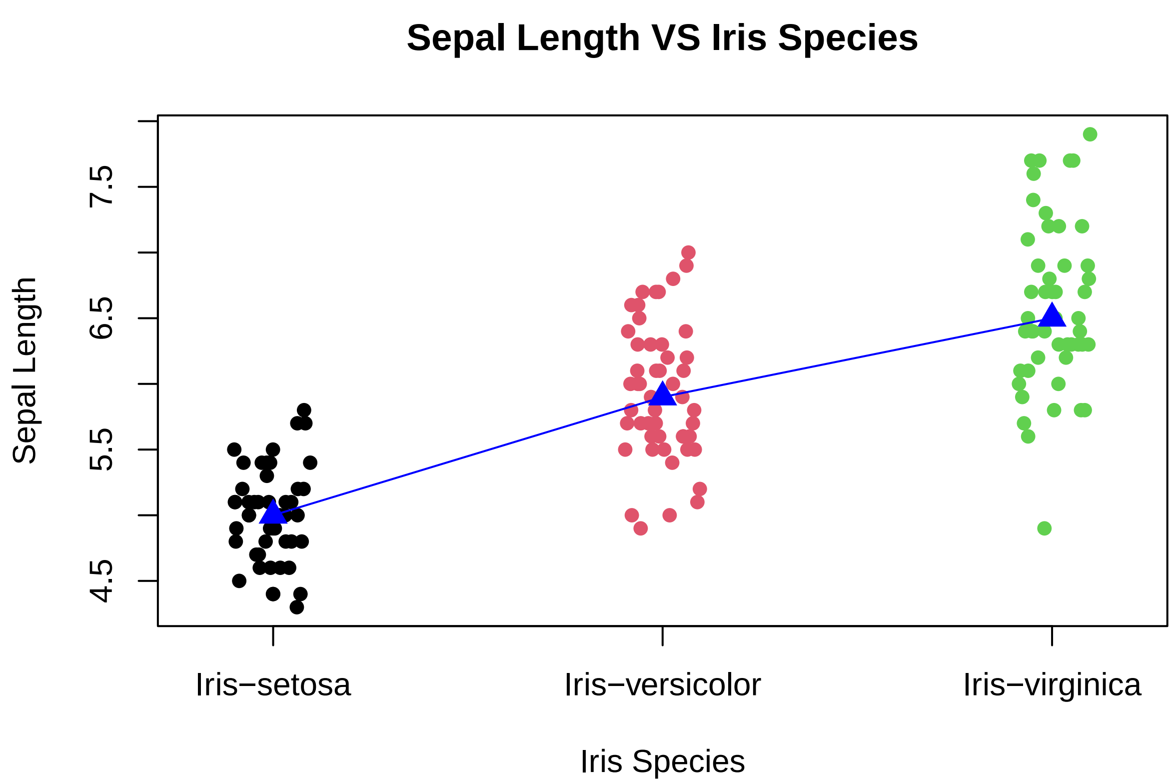
g) par

```
par(mfrow=c(1,3), mar= c(4,3.75,4,3.75))
for (i in levels(factor(data$Species))){
  plot( data$PetalWidth[data$Species == i],
       data$SepalLength[data$Species == i],
       main = paste("Sepal Width VS Sepal Length", i, sep="\n"),
       xlab = "Sepal Width",
       ylab = "Sepal Length",
       col=adjustcolor("black", alpha = 0.3 ),
       pch = 16)
  points(mean(data$PetalWidth[data$Species == i]),
        mean(data$SepalLength[data$Species == i]),
         col = 'red', pch = 16, cex.main=0.5)

}
```



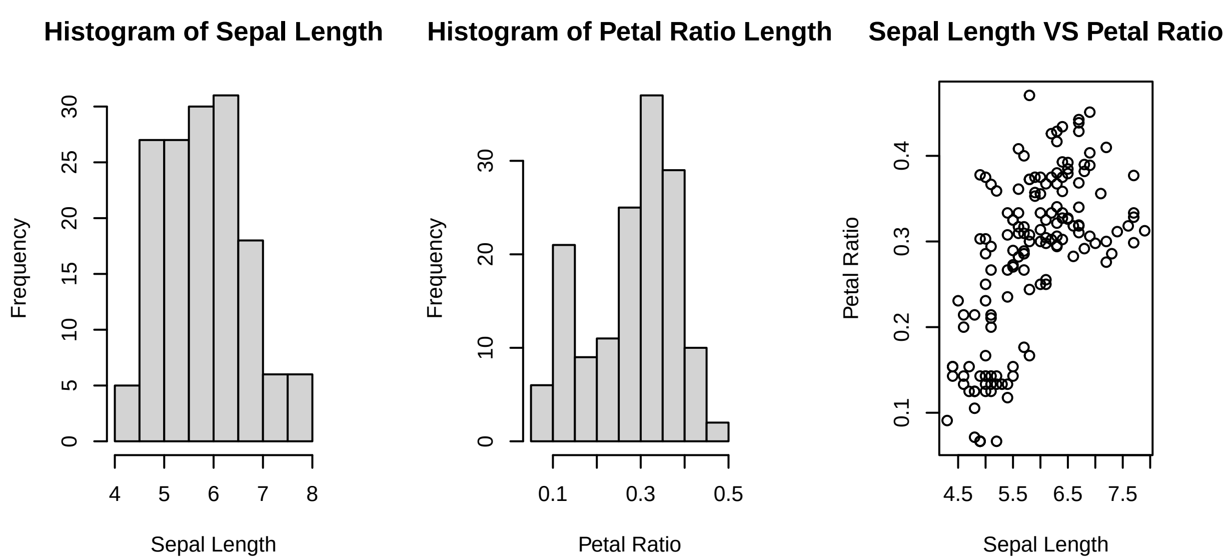
h)

```
medians <- sapply(levels(factor(data$Species)),
                  function(i) median(data$SepalLength[data$Species==i]))

stripchart(data$SepalLength~data$Species,
           vertical = TRUE,
           method = 'jitter',
           col  = factor(levels(factor(data$Species))),
           pch = 16,
           main = "Sepal Length VS Iris Species",
           xlab = 'Iris Species',
           ylab = 'Sepal Length'
           )
points(medians, col = 'blue', pch = 17, cex = 1.5)
lines(medians, col = 'blue')
```



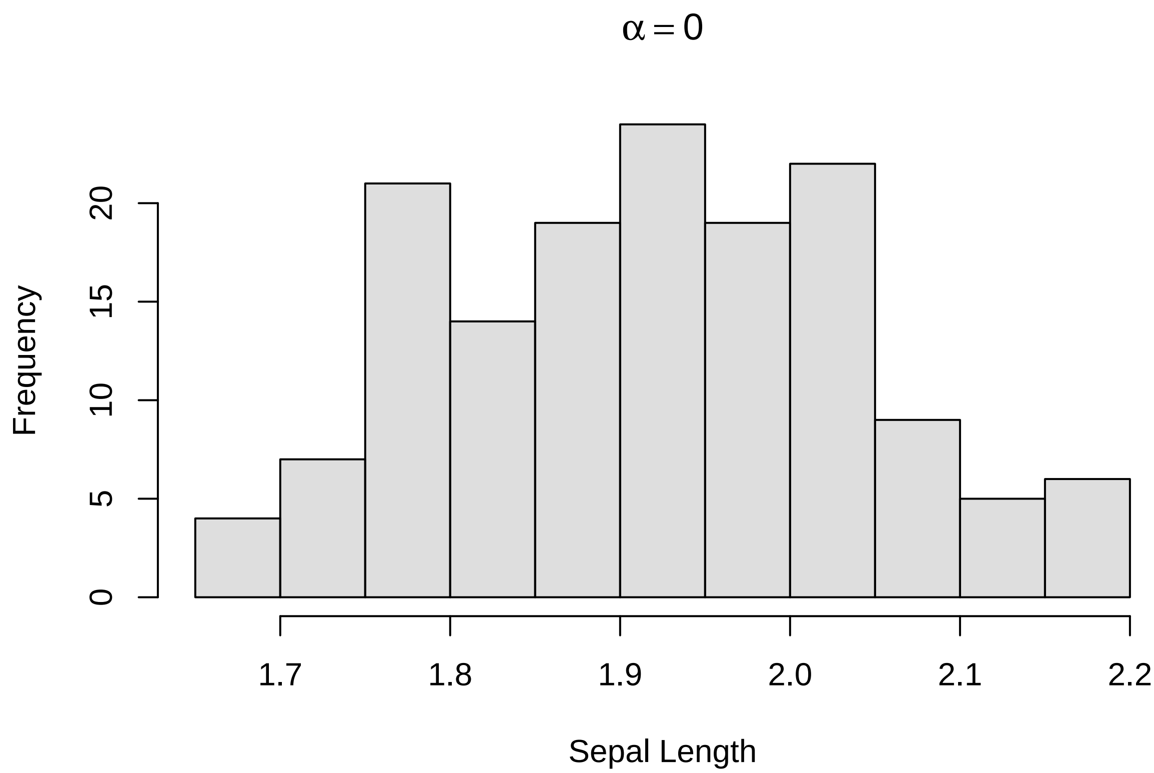
i.)

```
powerfun <- function(x, alpha) {
  if(sum(x <= 0) > 1) stop("x must be positive")
  if (alpha == 0)
    log(x)
  else if (alpha > 0) {
    x^alpha
  } else -x^alpha
}

par(mfrow=c(1,3), mar= c(4,4,4,4))
hist(data$SepalLength, main = 'Histogram of Sepal Length',
     xlab = 'Sepal Length', ylab = 'Frequency')
hist(data$PetalRatio, main = 'Histogram of Petal Ratio Length',
     xlab = 'Petal Ratio', ylab = 'Frequency')
plot(data$SepalLength, data$PetalRatio, main = 'Sepal Length VS Petal Ratio',
     xlab = 'Sepal Length', ylab = 'Petal Ratio')
```



ii.)

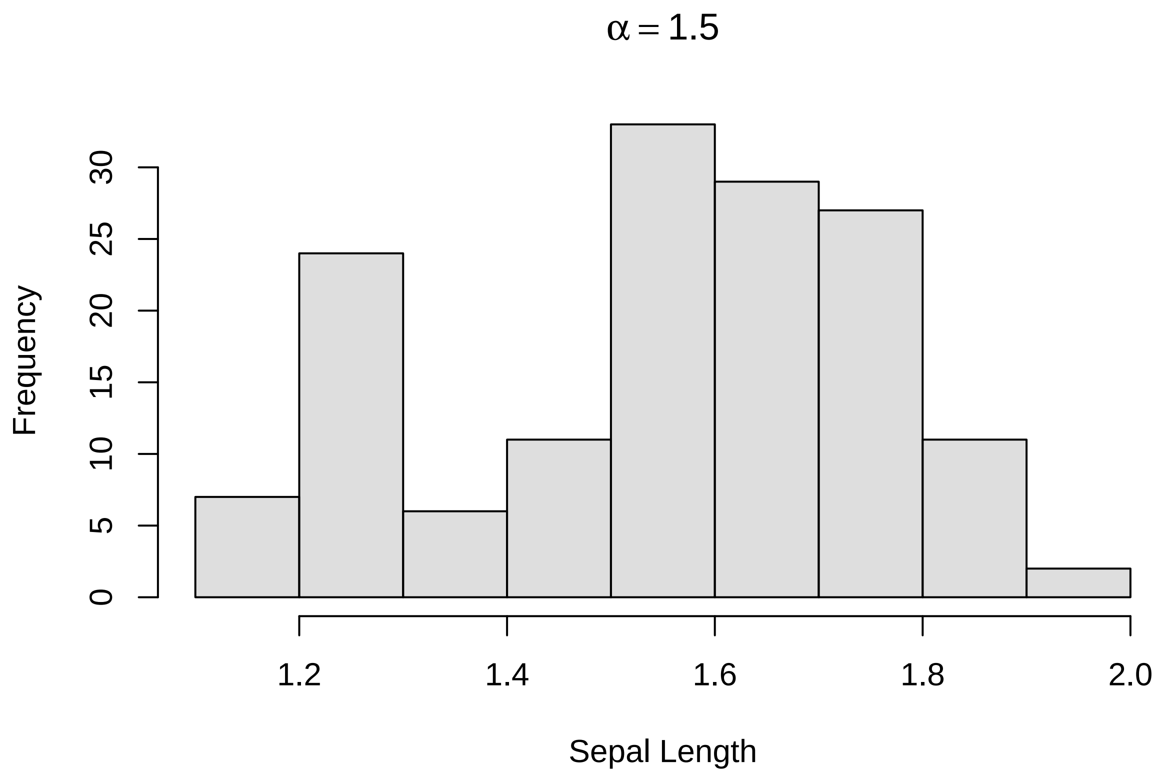
```
hist( powerfun(data$SepalLength + 1, 0),
      col=adjustcolor("grey", alpha = 0.5),
      main= bquote(alpha == .(0)), xlab = 'Sepal Length')
```



The value of α that make the distribution symmetric is 0

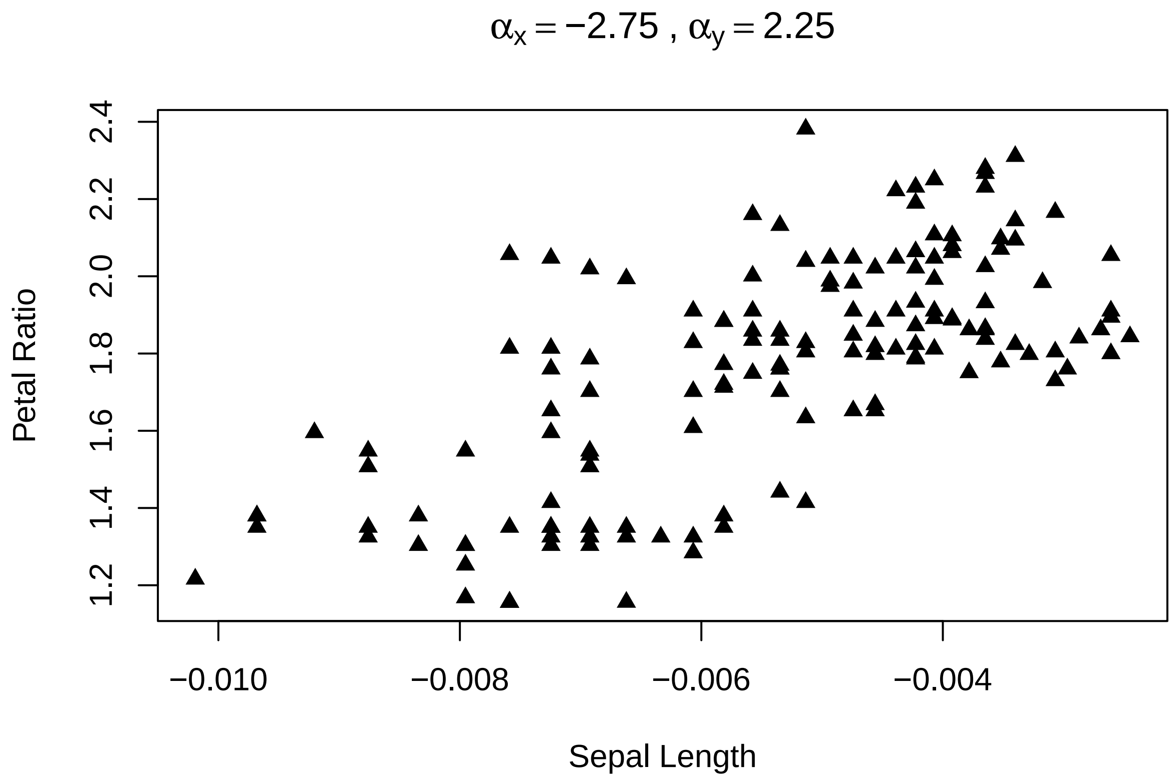
iii)

```
hist( powerfun(data$PetalRatio + 1, 1.75),
      col=adjustcolor("grey", alpha = 0.5),
      main= bquote(alpha == .(1.5)), xlab = 'Sepal Length')
```



iv)

```
par(mfrow = c(1,1))
plot(powerfun(data$SepalLength + 1, -2.75),
     powerfun(data$PetalRatio + 1, 2.25),
     pch = 17, xlab = 'Sepal Length', ylab = 'Petal Ratio',
     main = bquote(alpha[x] == .(-2.75) ~ "," ~ alpha[y] == .(2.25))
     )
```



The pairs of powers that make the scatter plot approximately linear is is $\alpha_{sepalLenth} = -2.75$ and $\alpha_{petalRatin} = 2.25$