Assignment 1 Question 2

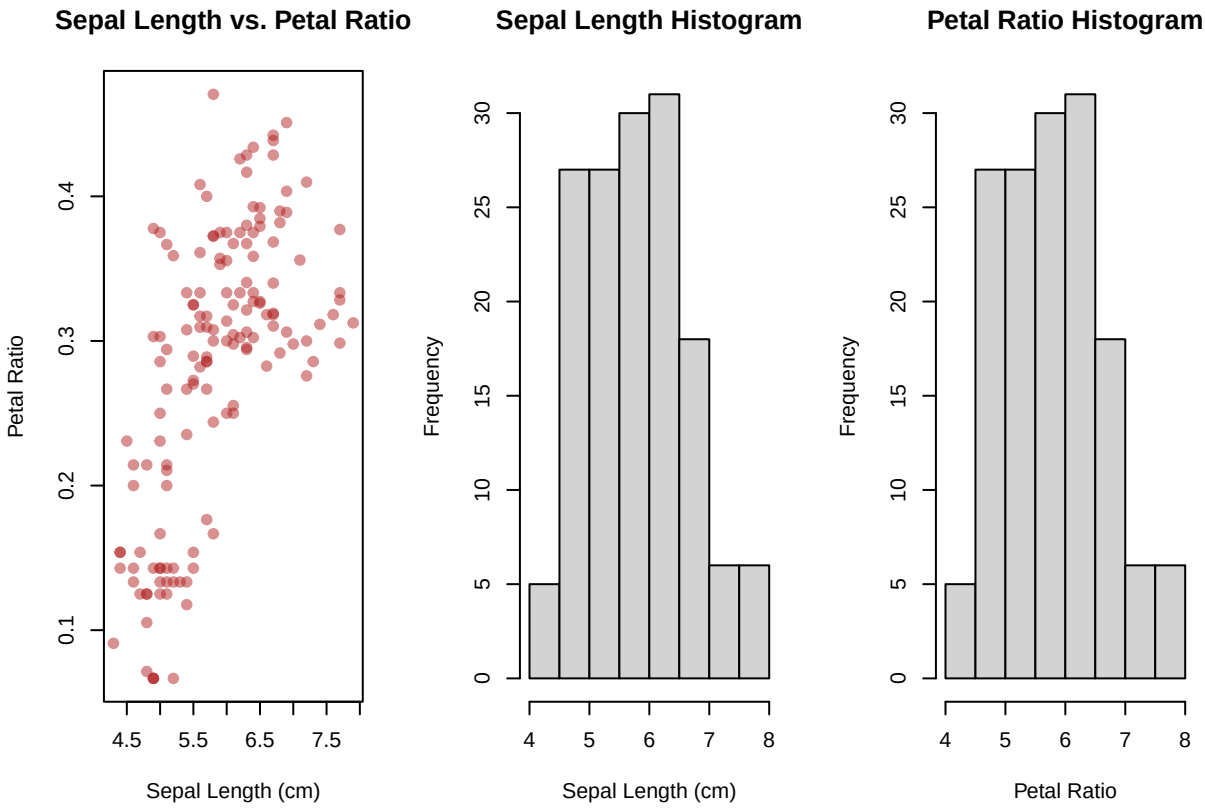
Sheen Thusoo

i. Using the SepalLength and PetalRatio, construct a 1×3 figure containing the two individual histograms and the scatterplot of the two variables.

Power Function provided in assignment.

```
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
}
```

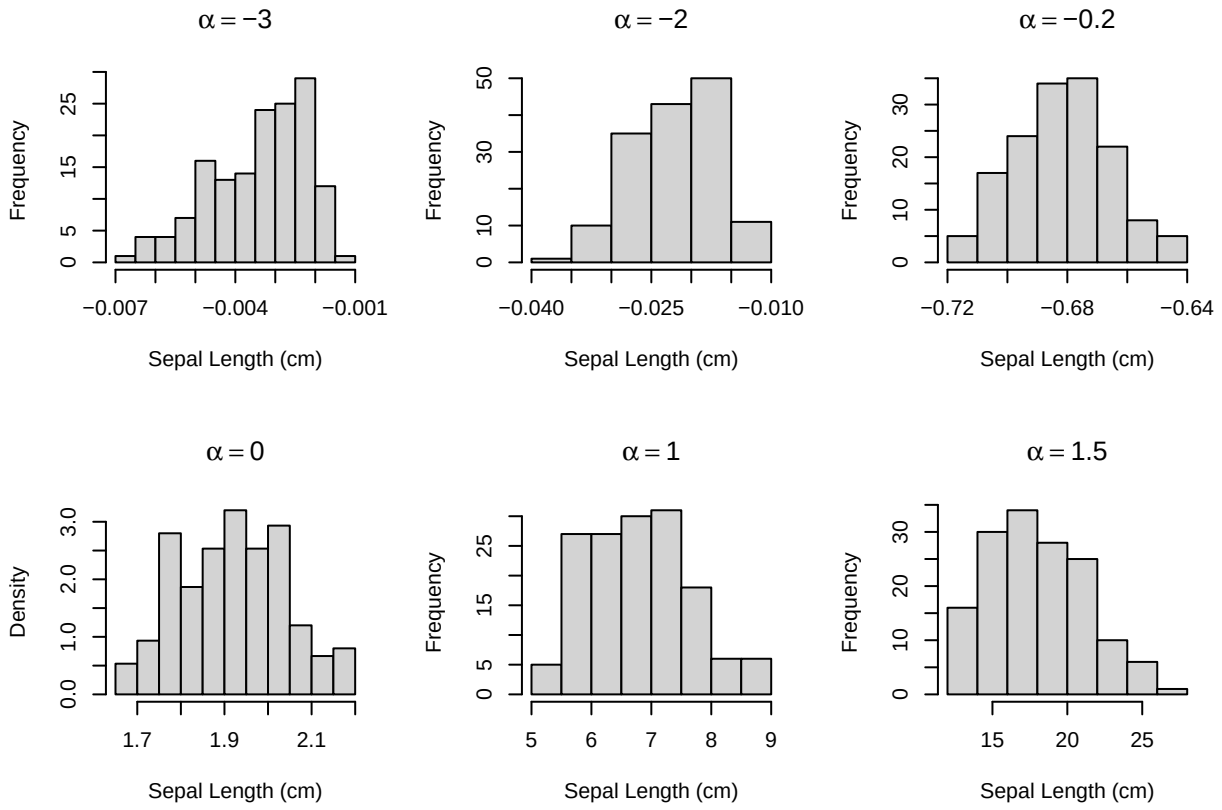
```
iris_data <- read.csv("Iris.csv")
iris_data <- transform(iris_data, PetalRatio = PetalWidth / PetalLength)
par(mfrow=c(1,3))
plot( iris_data$SepalLength, iris_data$PetalRatio, pch = 19,
      col=adjustcolor("firebrick", alpha = 0.5 ),
      main = "Sepal Length vs. Petal Ratio", xlab = "Sepal Length (cm)", ylab = "Petal Ratio" )

hist(iris_data$SepalLength, breaks="FD", main=paste("Sepal Length Histogram"),
     xlab = "Sepal Length (cm)")
hist(iris_data$SepalLength, breaks="FD", main=paste("Petal Ratio Histogram"),
     xlab = "Petal Ratio")
```



ii. What power (the values of α) makes the distribution of SepalLength approximately symmetric? Plot a histogram of the transformed variable.

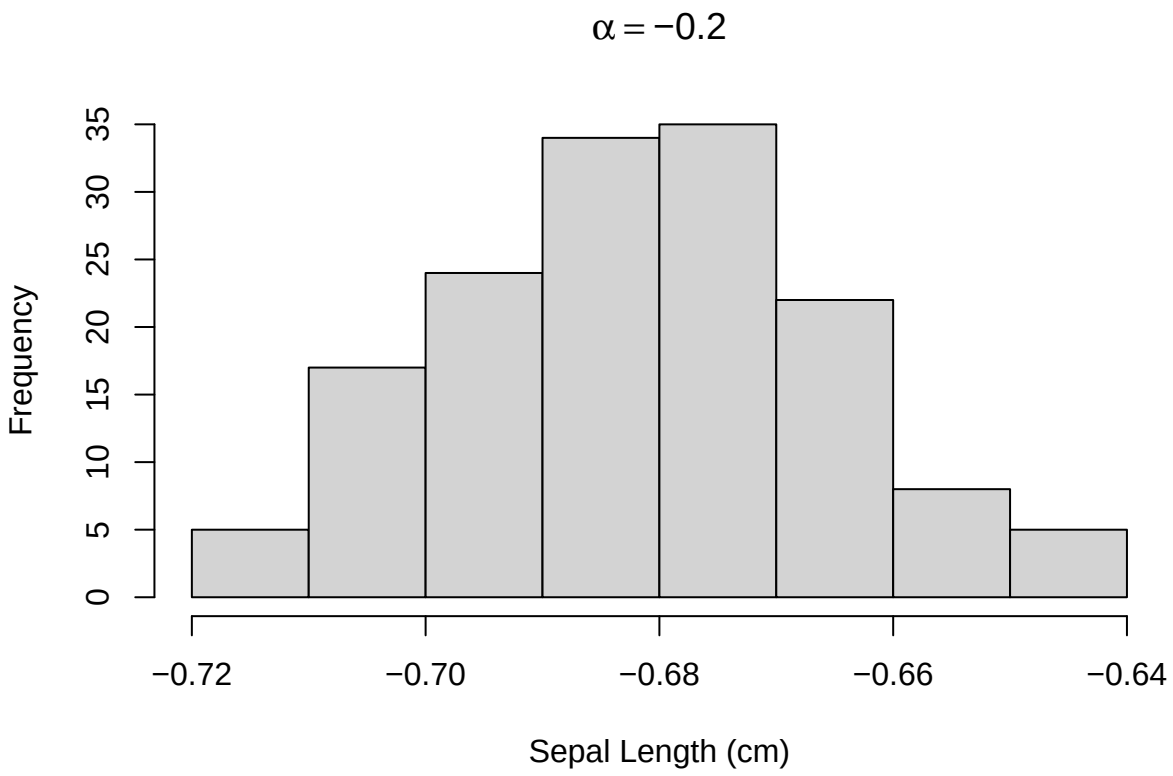
After trying various values of α using a for loop, including (-2, -1.5, -1, -0.2, -0.1, 0, 0.5, 1, 1.5), I found that the value of α that makes the distribution approximately symmetric is -0.2. The values above 1 made the graph left-skewed. I used Bump Rule #1 to find the most appropriate value. Because the bump was on lower values when α was greater than 1, I moved the power lower on the ladder. I tried values like -3, and -2 but noticed that they produced a bump on the higher values, so I needed to move the power up on the ladder. After trying many other values that are shown below, I noticed that a value of -0.2 made the histogram symmetric.

```
par(mfrow=c(2,3))
a = c(-3, -2, -0.2, 0, 1, 1.5)
for (i in 1:length(a) ) hist( powerfun(iris_data$SepalLength + 1, a[i]), breaks="FD",
                              main=bquote(alpha==.(a[i])), xlab = "Sepal Length (cm)", a[i])
```



The histogram for the chosen α value is shown below.

```
a = -0.2
hist( powerfun(iris_data$SepalLength + 1, a), breaks="FD",
                                  main=bquote(alpha==.(a)), xlab = "Sepal Length (cm)", a)
```

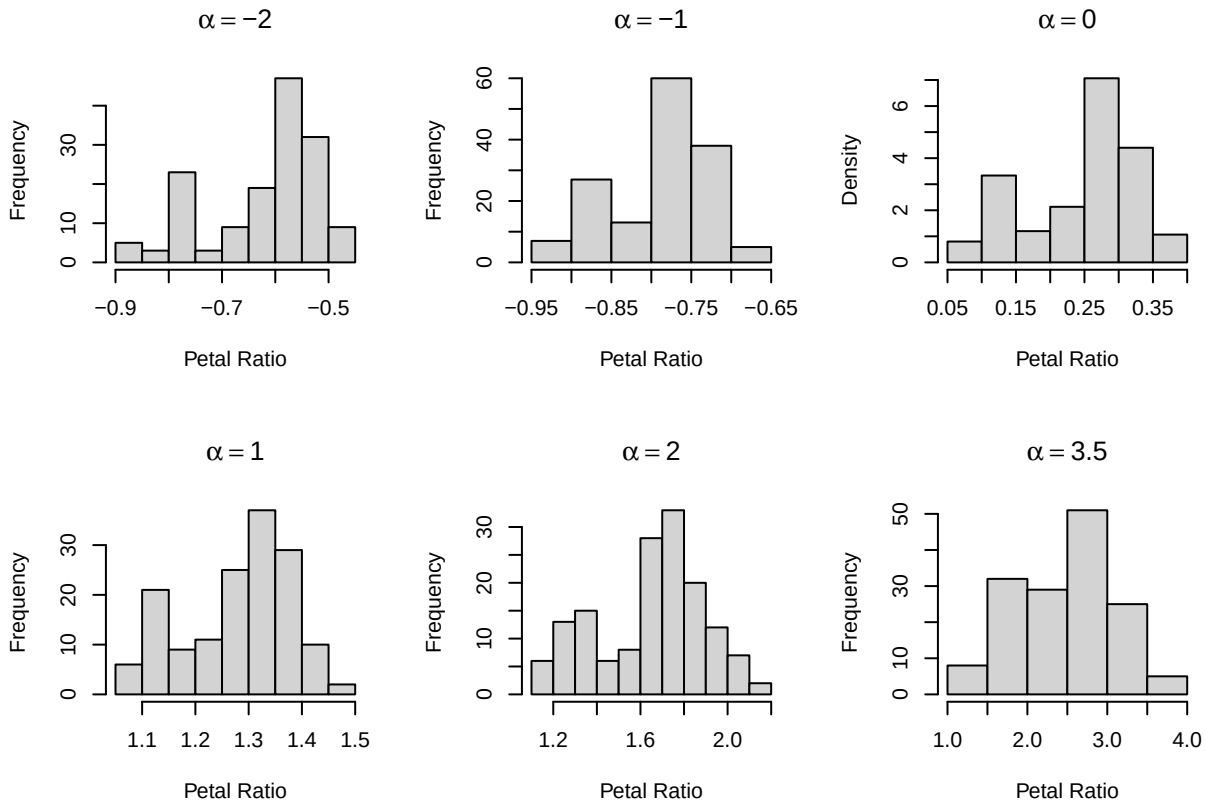


The value $\alpha = -0.2$ makes the distribution of Sepal Length approximately symmetric.

iii. What power (the values of α) makes the distribution of PetalRatio approximately symmetric? Plot a histogram of the transformed variable.

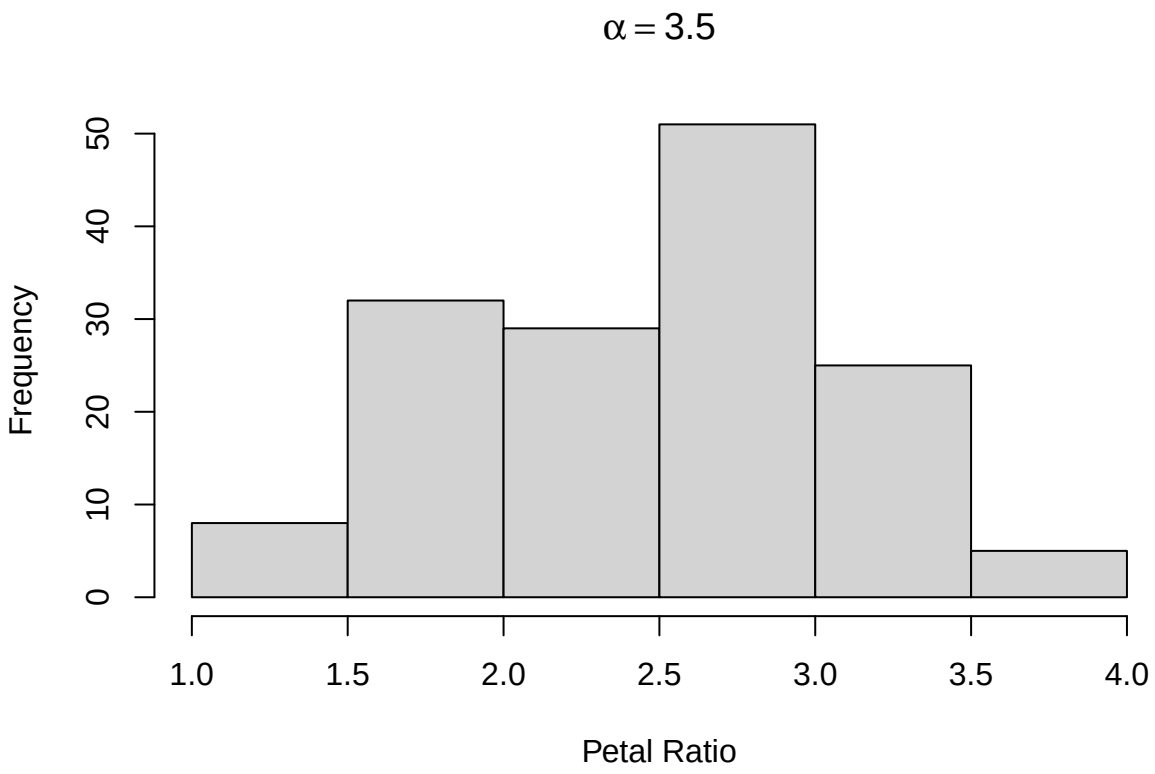
After trying various values of α using a for loop, including (-0.1, 0, 0.5, 0.6, 1, 1.5), I found that the value of α that makes the distribution approximately symmetric is 3.5. I used Bump Rule #1 to find the most appropriate value. Because the bump was on higher values when α was lower than 1, I moved the power higher on the ladder. I tried values like 2.5 and 3 but noticed that they produced a bump on the lower values, so I needed to move the power down on the ladder. After trying many other values that are shown below, I noticed that a value of -0.2 made the histogram symmetric.

```
par(mfrow=c(2,3))
a1 = c(-2, -1, 0, 1, 2, 3.5)
for (i in 1:length(a1) ) hist( powerfun(iris_data$PetalRatio + 1, a1[i]), breaks="FD",
                               xlab = "Petal Ratio", main=bquote(alpha==.(a1[i])), a1[i])
```



The histogram for the chosen α value is shown below.

```
a = 3.5
hist( powerfun(iris_data$PetalRatio + 1, a), breaks="FD",
                                main=bquote(alpha==.(a)), xlab = "Petal Ratio", a)
```

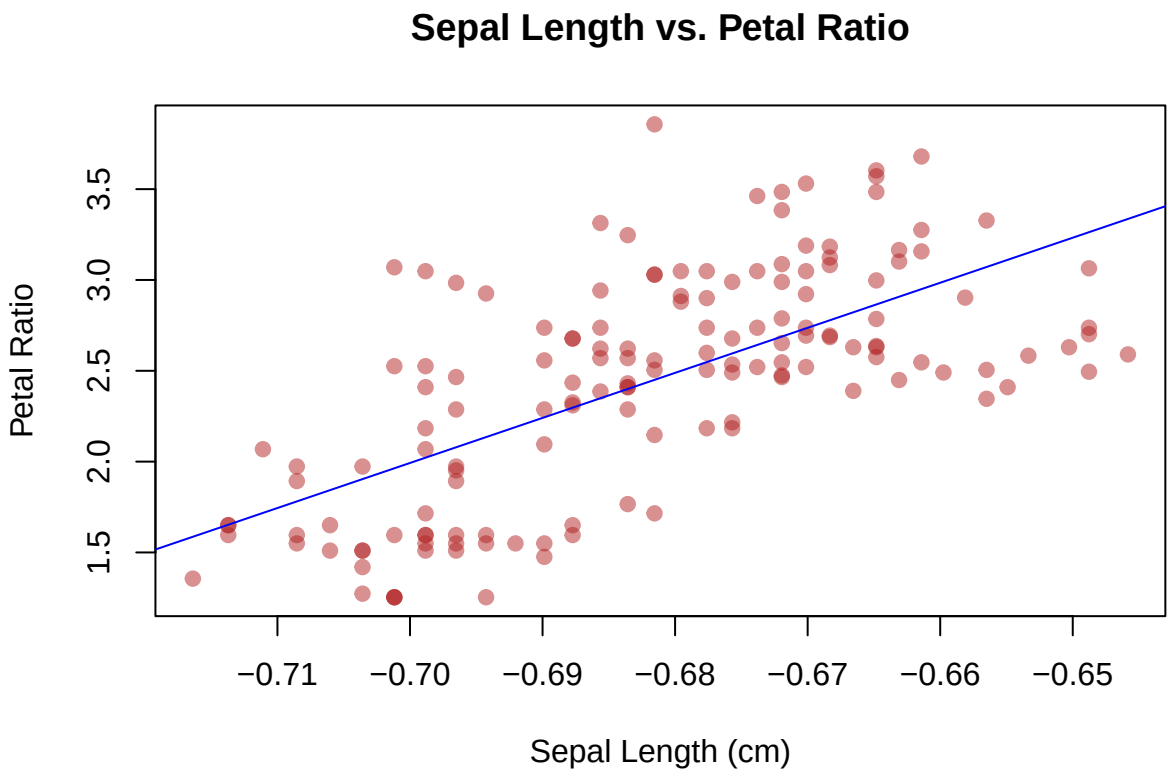


The value $\alpha = 3.5$ makes the distribution of Petal Ratio approximately symmetric.

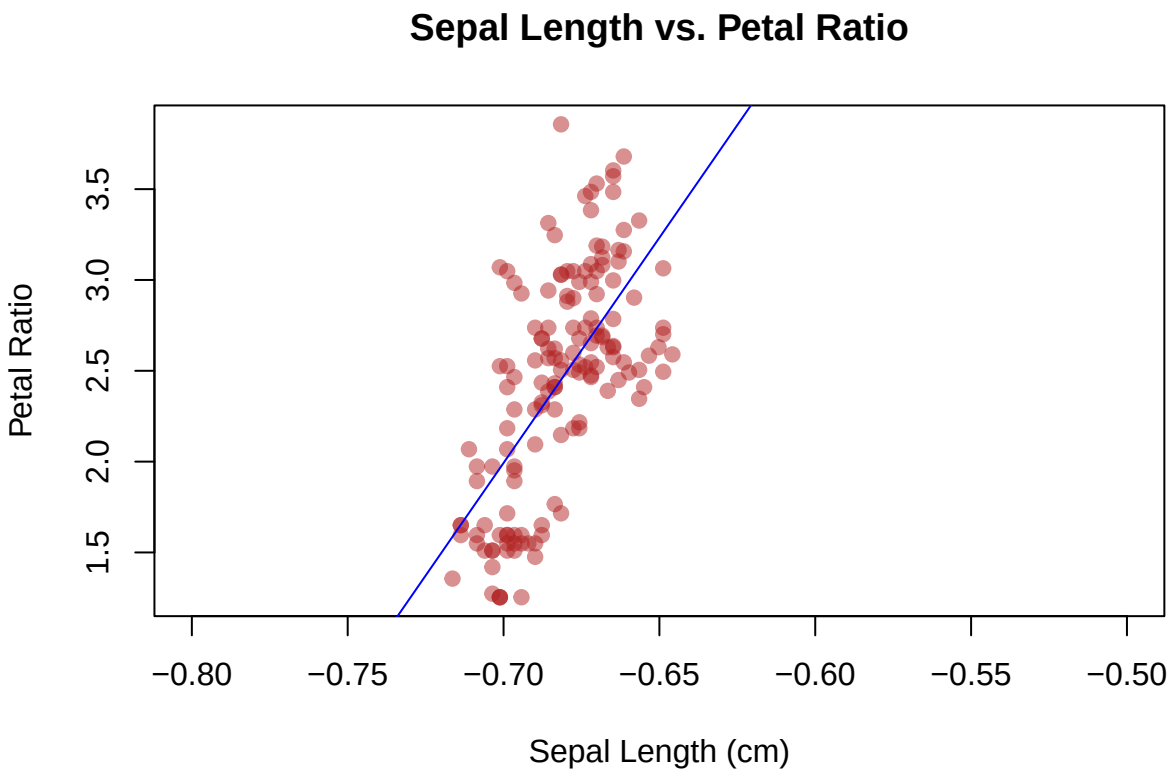
iv. What pair of powers (the values of α) makes the scatter plot of SepalLength and PetalRatio approximately linear? Plot a scatter plot of the transformed variables.

The pairs of values that make the scatterplot approximately linear are $\alpha_x = -0.2$ and $\alpha_y = 3.5$. I used bump rule #2. I found that these values worked best in making the scatterplot similar. When zooming out on the x-axis slightly, the plot looks more linear, as shown in the second figure below.

```
plot( powerfun(iris_data$SepalLength + 1, -0.2), powerfun(iris_data$PetalRatio + 1, 3.5),
      pch = 19, col=adjustcolor("firebrick", alpha = 0.5 ), xlab = "Sepal Length (cm)",
      ylab = "Petal Ratio", main = "Sepal Length vs. Petal Ratio")
abline( lm( I(powerfun(iris_data$PetalRatio + 1, 3.5)) ~
              I(powerfun(iris_data$SepalLength + 1, -0.2))), col="blue" )
```




```
plot( powerfun(iris_data$SepalLength + 1, -0.2), powerfun(iris_data$PetalRatio + 1, 3.5),
      pch = 19, col=adjustcolor("firebrick", alpha = 0.5 ), xlim= c(-0.8, -0.5),
      xlab = "Sepal Length (cm)", ylab = "Petal Ratio", main = "Sepal Length vs. Petal Ratio")
abline( lm( I(powerfun(iris_data$PetalRatio + 1, 3.5)) ~
              I(powerfun(iris_data$SepalLength + 1, -0.2))), col="blue" )
```


Sepal Length vs. Petal Ratio
Petal Ratio
1.5
2.0
2.5
3.0
3.5
−0.80
−0.75
−0.70
−0.65
−0.60
−0.55
−0.50
Sepal Length (cm)