1.

```
#PROBLEM 1
standardize <- function(data)
{
  D <- ncol(data)
  N <- nrow(data)
  for(j in 1:D)
  {
    mean = mean(data[,j])
    sd = sd(data[,j])
    for(i in 1:N)
    {
      data[i,j] = (data[i,j] - mean)/sd
    }
  }
  return(data)
}
```

2. For the raw data, (w0, w1) = **(-350.737192, 7.717288).**
For the standardized data, (w0, w1) = **(-5.851629e-16, 9.247563e-1)**.

We get $w0^{standardized}$ equals about 0 because standardizing the data makes the mean for the heights and weights column equal to 0. Hence, the y-intercept will equal 0. Furthermore, we get the relationship: $w1^{standardized}$ *sd(weights)/sd(heights) = $w1^{raw\ data}$

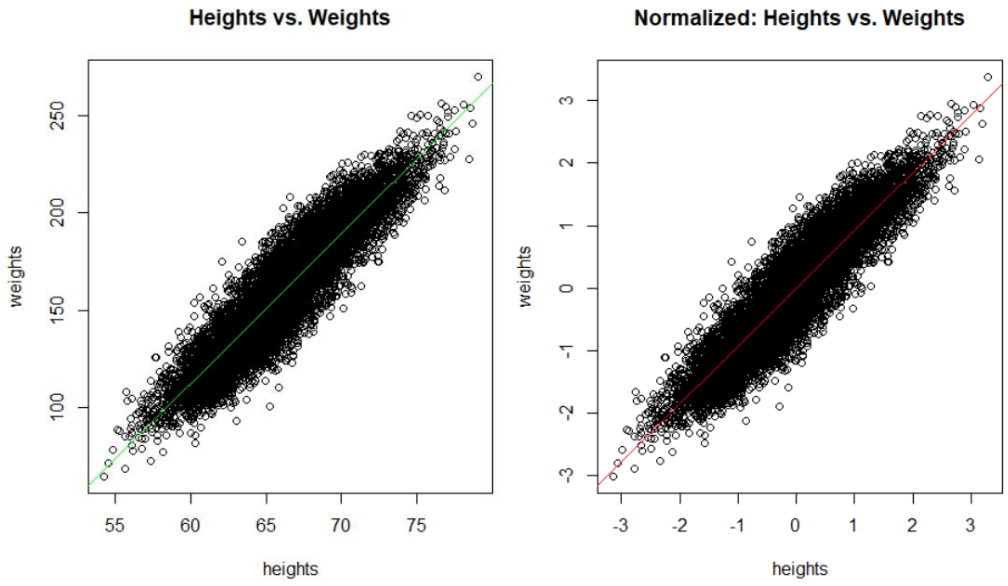
```
#PROBLEM 2
df <- read.csv("height_weight_genders.csv")
xyMatrix <- df[,2:3]
xyMatrix_normal <- standardize(xyMatrix)

findMSE <- function(xVec, yVec)
{
  N = length(xVec)
  w1 = (N*sum(xVec*yVec)) - (sum(xVec)*sum(yVec))
  w1 = w1/((N*sum(xVec^2)) - (sum(xVec))^2)
  w0 = mean(yVec) - w1*mean(xVec)

  return(c(w0, w1))
}

#MSE for regular data
w <- findMSE(xyMatrix[,1], xyMatrix[,2])
#plot for reg data
plot(xyMatrix[,1], xyMatrix[,2], main="Heights vs. Weights", xlab="heights", ylab="weights")
abline(w[1], w[2], col = 'green')

#MSE for normalized data
wNorm <- findMSE(xyMatrix_normal[,1], xyMatrix_normal[,2])
#plot for normalized data
plot(xyMatrix_normal[,1], xyMatrix_normal[,2], main="Normalized: Heights vs. Weights", xlab="heights", ylab="weights")
abline(wNorm[1], wNorm[2], col = 'red')
```


Heights vs. Weights
weights
heights
100
150
200
250
55
60
65
70
75
Normalized: Heights vs. Weights
weights
heights
-3
-2
-1
0
1
2
3