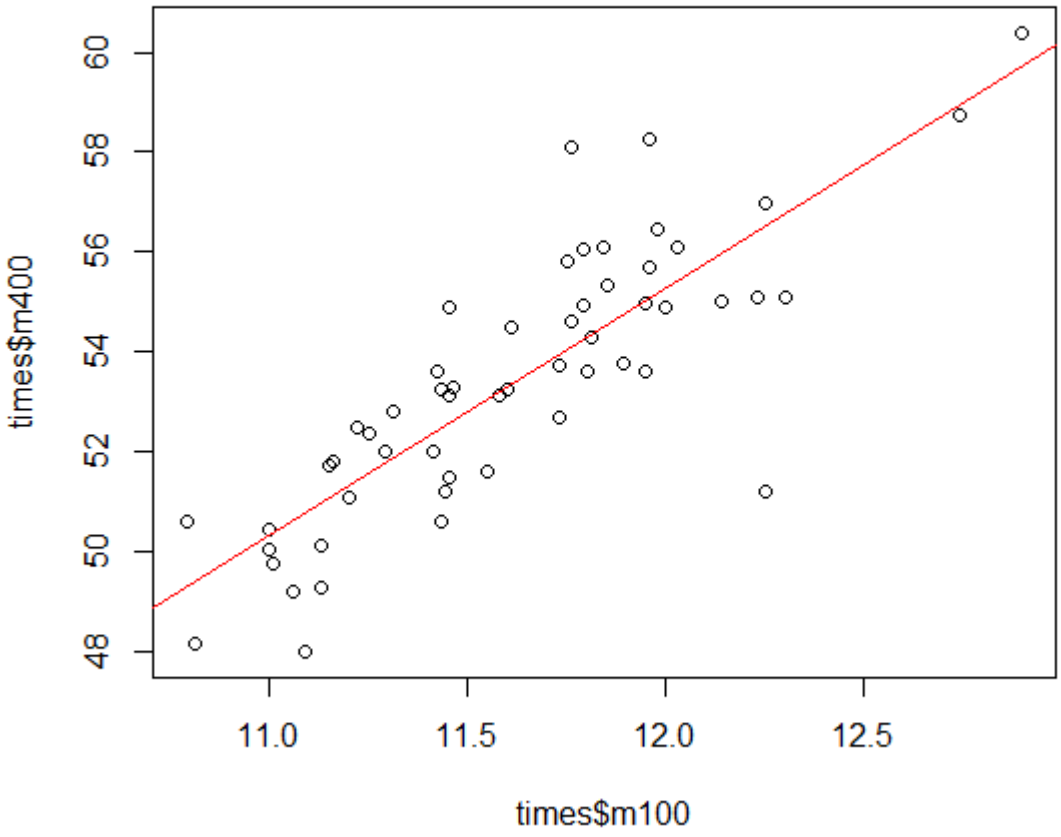
Lab 3

Kyle Salitrik
997543474

After viewing the scatter plots of all of the variables, it seems that there are strong correlations between all of the long distance times and the short distance times. However, using a short runner's speeds to estimate how well they will perform on a long distance run seems to be less accurate as the mean error of the predictor would be large in these cases.

After computing the least square estimators, slope (β0), intercept (β1), and plotting the regression line along with the data, we can see that our estimator fits the data well, however there are still some significant outliers at ≈11.6, ≈11.7, and ≈12.3.



The code and outputs are recorded below.

CODE

```
> times = read.table('record.txt',header=TRUE,sep=' ')
> attach(times)
> plot(times)
> cor(times)
           m100      m200      m400      m800     m1500     m3000
m100  1.0000000 0.9527911 0.8346918 0.7276888 0.7283709 0.7416988
m200  0.9527911 1.0000000 0.8569621 0.7240597 0.6983643 0.7098710
m400  0.8346918 0.8569621 1.0000000 0.8984052 0.7878417 0.7776369
m800  0.7276888 0.7240597 0.8984052 1.0000000 0.9016138 0.8635652
m1500 0.7283709 0.6983643 0.7878417 0.9016138 1.0000000 0.9691690
m3000 0.7416988 0.7098710 0.7776369 0.8635652 0.9691690 1.0000000
> y = m400
> x = m100
> x_bar = mean(x)
> y_bar = mean(y)
> b1 = sum((x-x_bar)*(y-y_bar))/sum((x-x_bar)^2)
> b0 = y_bar - b1*x_bar
> X=as.matrix(cbind(1,m100))
> beta=solve(t(X)%*%X,t(X)%*%y)
> beta
          [,1]
     -4.032628
m100  4.943686
> plot(times$m100,times$m400)
> abline(beta[1],beta[2],col='red')
```