STAT511 HW3

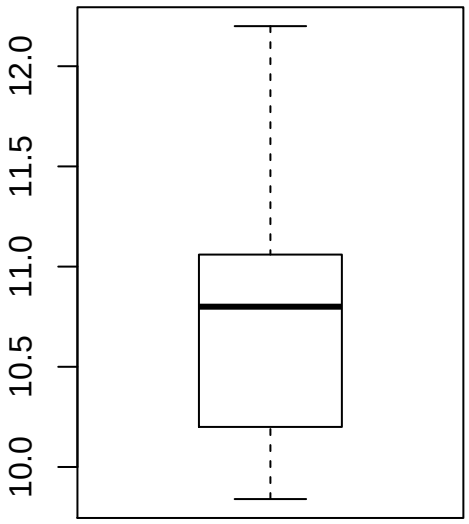
Ben Straub

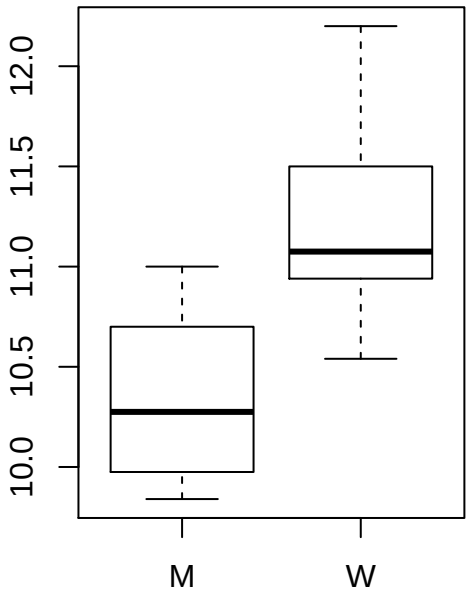
September 26, 2015

(4) Olympic 100m Gold Medal Times

Exploratory Data Analysis

```
'data.frame':   42 obs. of  3 variables:
 $ year    : int  1900 1904 1908 1912 1920 1924 1928 1932 1936 1948 ...
 $ goldtime: num  11 11 10.8 10.8 10.8 10.6 10.8 10.3 10.3 10.3 ...
 $ gender  : Factor w/ 2 levels "M","W": 1 1 1 1 1 1 1 1 1 1 ...
```





Six Number Summary seperated by gender

```
      year         goldtime      gender
 Min.   :1900   Min.   : 9.840   M:24
 1st Qu.:1927   1st Qu.: 9.982   W: 0
 Median :1958   Median :10.275
 Mean   :1954   Mean   :10.318
 3rd Qu.:1981   3rd Qu.:10.650
 Max.   :2004   Max.   :11.000
```

```
      year         goldtime     gender
 Min.   :1928   Min.   :10.54   M: 0
 1st Qu.:1953   1st Qu.:10.95   W:18
 Median :1970   Median :11.07
 Mean   :1969   Mean   :11.23
```

```
 3rd Qu.:1987   3rd Qu.:11.50
 Max.   :2004   Max.   :12.20
```

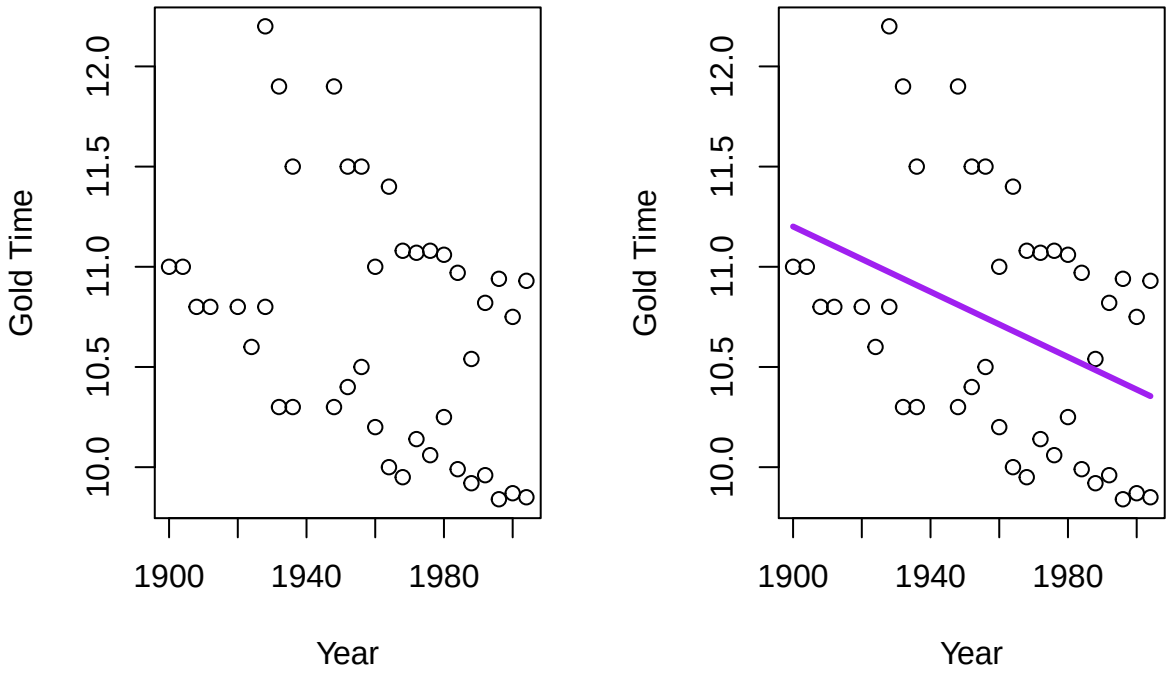
Summary of Model

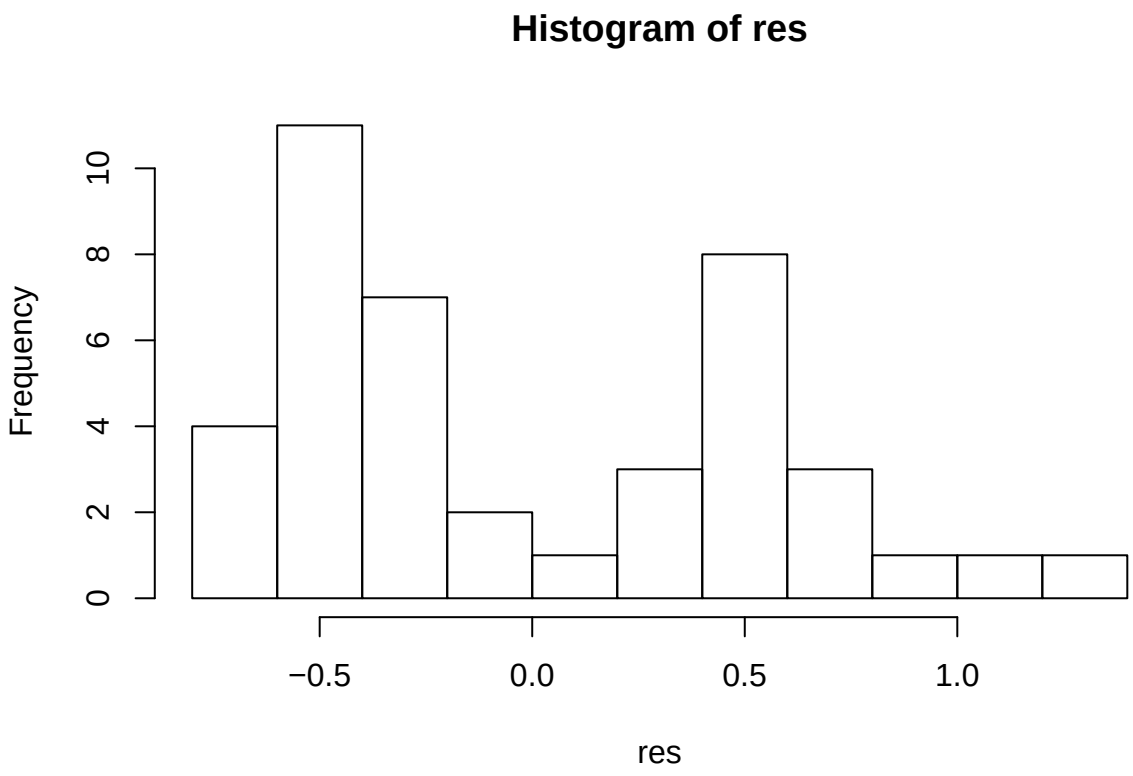
```
Call:
lm(formula = goldtime ~ year, data = oly)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6976 -0.5089 -0.2196  0.4869  1.2269

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.663985   5.865382   4.546 4.97e-05 ***
year        -0.008138   0.002991  -2.721   0.0096 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5679 on 40 degrees of freedom
Multiple R-squared:  0.1561,    Adjusted R-squared:  0.135
F-statistic: 7.401 on 1 and 40 DF,  p-value: 0.0096
```

Coefficients of Model

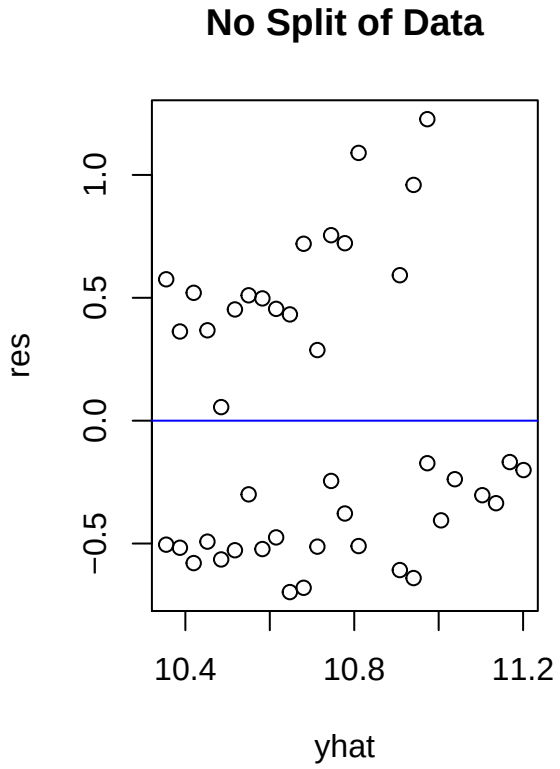
```
 (Intercept)         year
26.663984630 -0.008138418
```

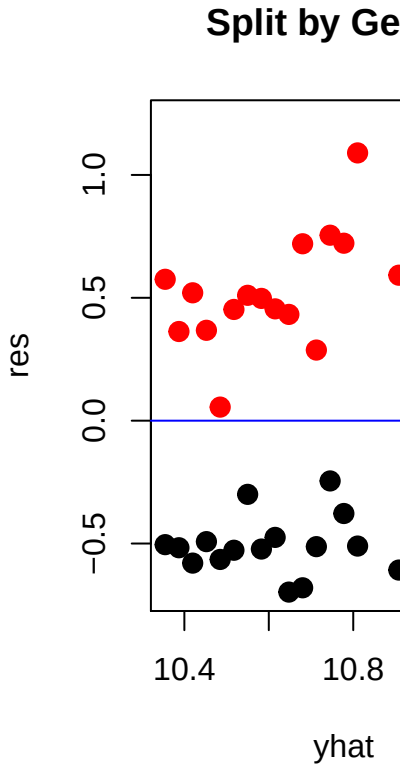
Year and Goldtime

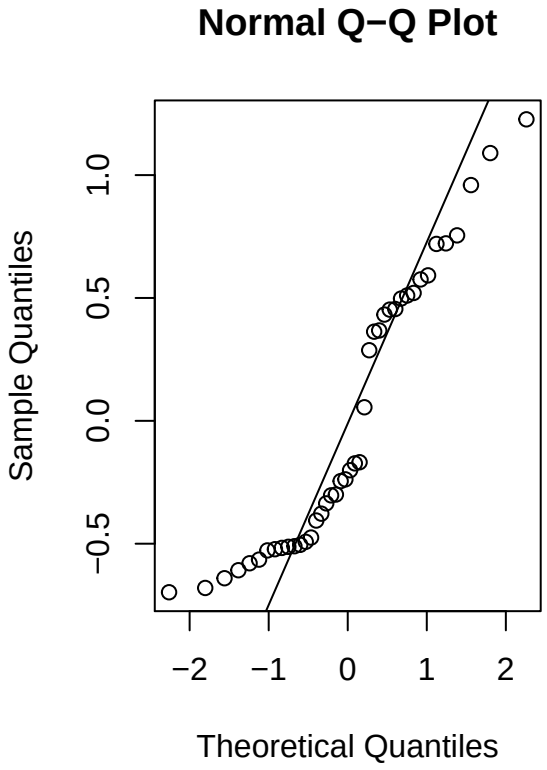




OBSERVATION: **The histogram of the residuals has a bimodal distribution, which indicates we have two diffegoldtime groups.**







The curgoldtime model is not a good fit for the data. We can see that the Histogram of the Residuals has a bimodal distributions, which indicates two distinct groups in the data. The Residuals plotted against yhats indicates heteroscadacity. If we color code based on gender we can see an obvious split of the data into two distinct groups.

The gold time for each race goes down by 0.011 for each Olympic Year.

NEW MODEL with Year and Gender As an Interaction Effect

```
Call:
lm(formula = goldtime ~ year * factor(gender), data = oly)

Residuals:
     Min       1Q   Median       3Q      Max
-0.37579 -0.05460  0.00738  0.08276  0.32234

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          31.826453   2.128910  14.950  < 2e-16 ***
year                 -0.011006   0.001089 -10.104 2.56e-12 ***
factor(gender)W      12.520596   4.076141   3.072  0.00392 **
year:factor(gender)W -0.005817   0.002074  -2.804  0.00791 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1707 on 38 degrees of freedom
Multiple R-squared:  0.9275,    Adjusted R-squared:  0.9218
F-statistic: 162.1 on 3 and 38 DF,  p-value: < 2.2e-16

         (Intercept)                 year      factor(gender)W
        31.826452523         -0.011005562          12.520596237
year:factor(gender)W
        -0.005816509
```