HW2

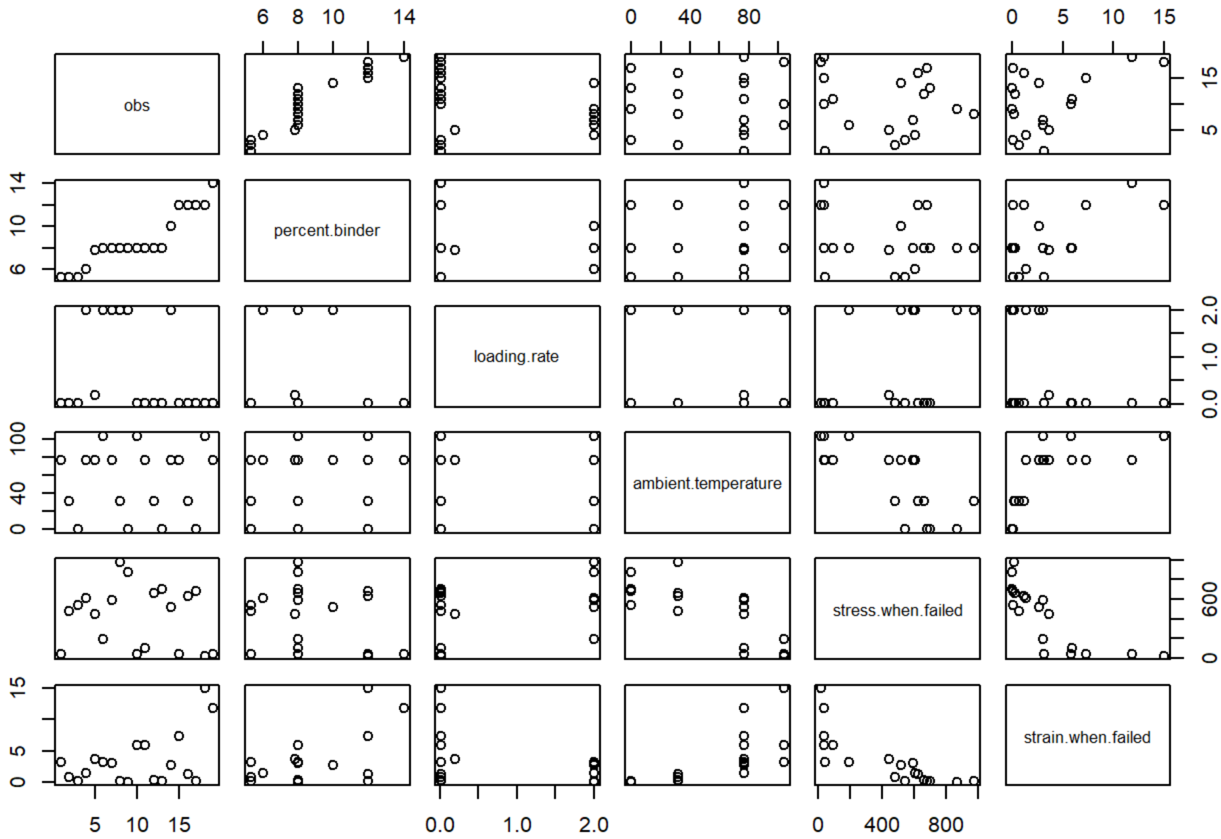
8.3

Here is my data to question 8, part 3,

```
link<-("C:/Users/ASEARLY/Downloads/CompanionAsset_9780128230435_DataSets (3)/Data Tables 4th e .t
s<-read.table(link,col.names=c("obs","percent binder","loading rate","ambient temperature","stres
head(s)
```

```
  obs percent.binder loading.rate ambient.temperature stress.when.failed
1   1            5.3         0.02                  77                 42
2   2            5.3         0.02                  32                481
3   3            5.3         0.02                   0                543
4   4            6.0         2.00                  77                609
5   5            7.8         0.20                  77                444
6   6            8.0         2.00                 104                194
  strain.when.failed
1               3.20
2               0.73
3               0.16
4               1.44
5               3.68
6               3.11
```

```
n<-nrow(s)
plot(s)
```

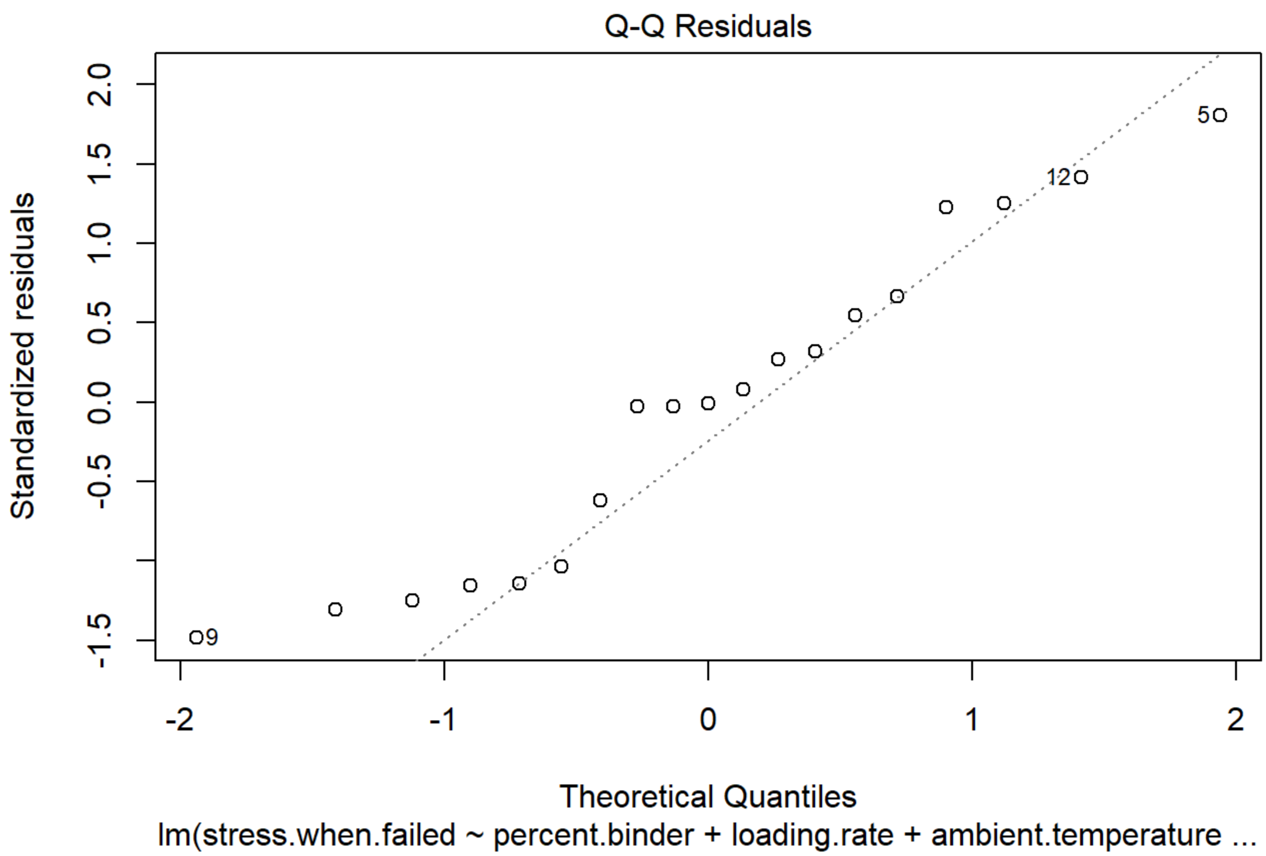



```r
lm_out <- lm(stress.when.failed ~ percent.binder + loading.rate + ambient.temperature, data = s)
summary(lm_out)
```

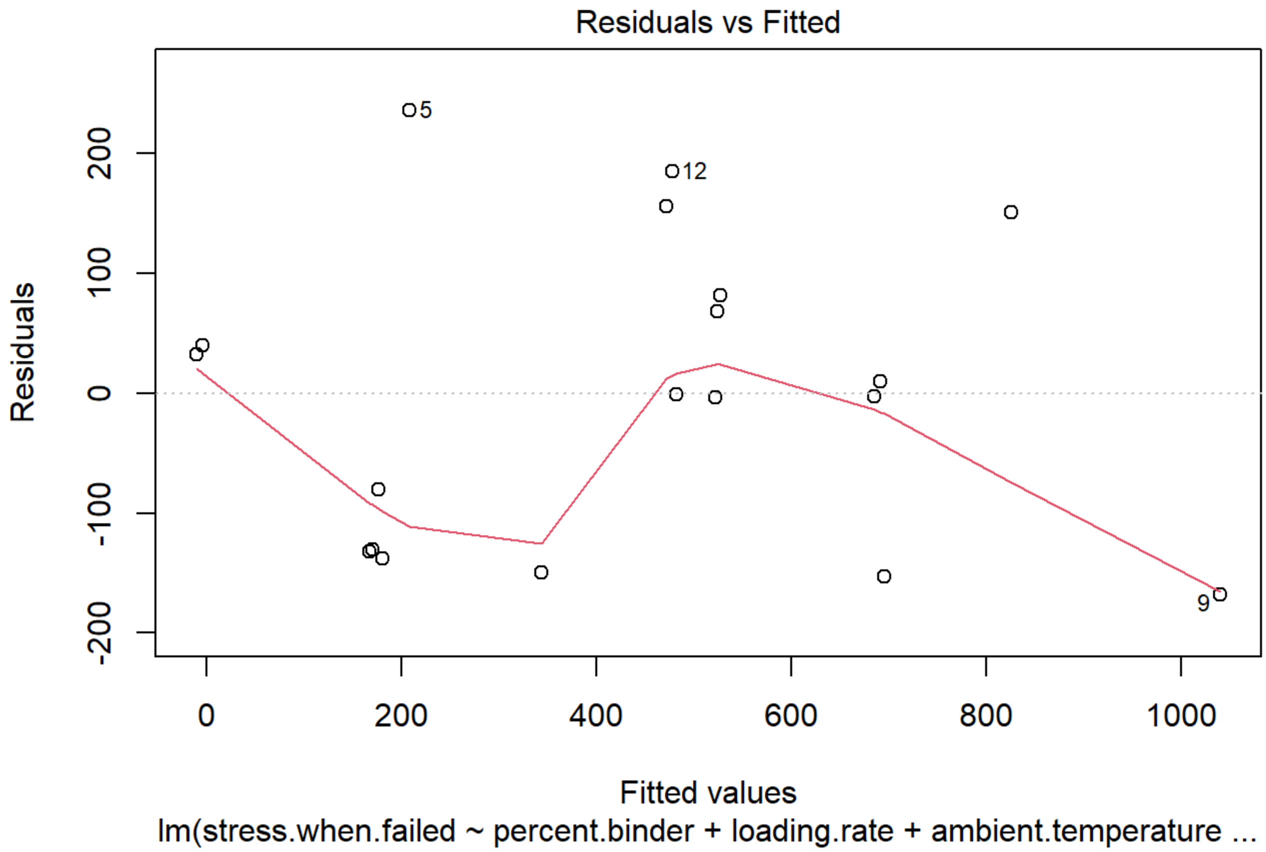
```
Call:
lm(formula = stress.when.failed ~ percent.binder + loading.rate + 
    ambient.temperature, data = s)

Residuals:
     Min       1Q   Median       3Q      Max 
-168.380 -131.124   -0.743   74.773  235.765 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         700.6180   125.8722   5.566 5.40e-05 ***
percent.binder       -1.5257    13.0242  -0.117 0.908302    
loading.rate        175.9839    35.6550   4.936 0.000179 ***
ambient.temperature  -6.6971     0.8847  -7.570 1.69e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 137.9 on 15 degrees of freedom
```

```
Multiple R-squared:  0.8376,    Adjusted R-squared:  0.8051
F-statistic: 25.79 on 3 and 15 DF,  p-value: 3.599e-06
```

```
plot(lm_out,which=2)
```




```
plot(lm_out,which=1)
```



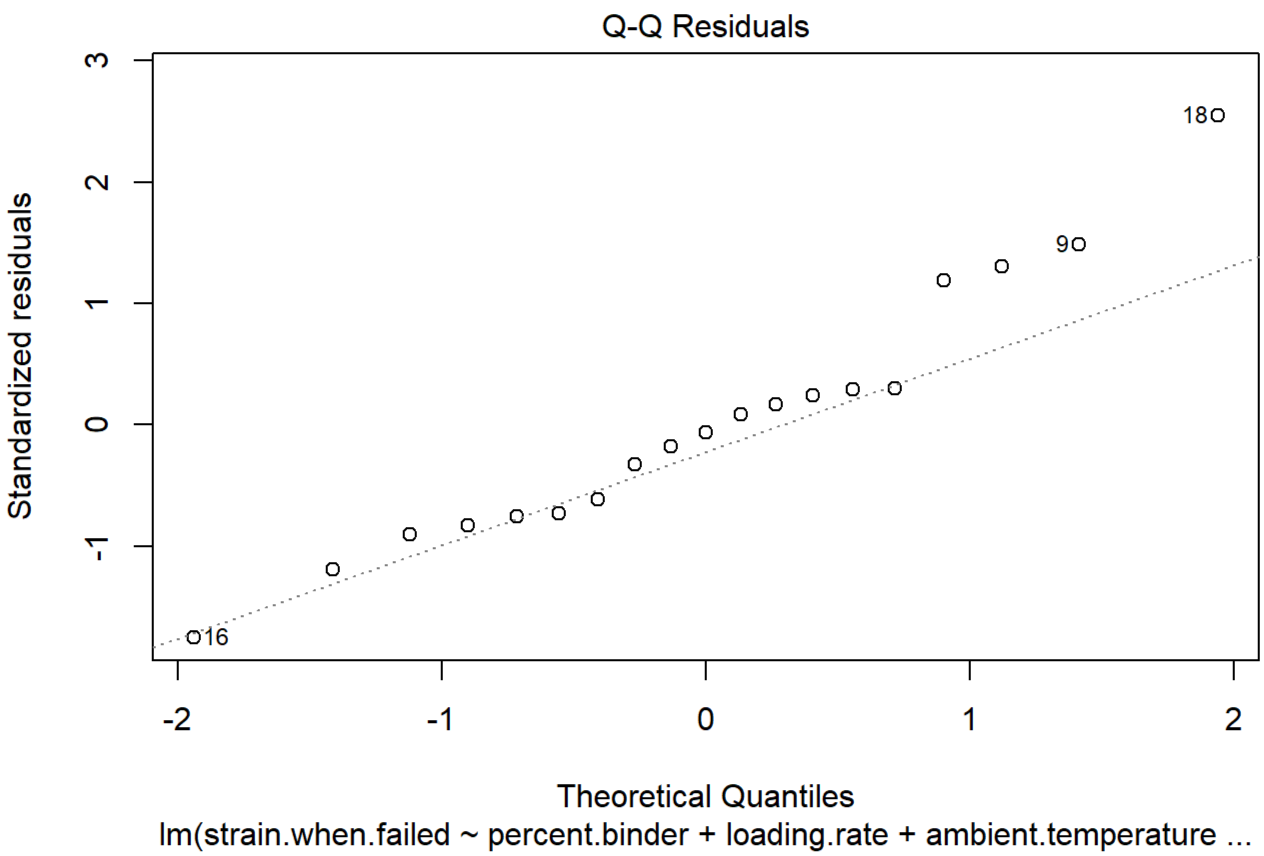
Interpretation: When interpreting the Q-Q plot for the data in relation to stress, the graph is lightly-tailed, making it not quite normal. Addiotionally, based on the residual plot, there is a fairly consistent variance throughout the data as well. Additionally, since the p-value is much lower than 0.05, we should reject the null hypothesis that the means for all of the variables equals 0. Finally, the R-squared value of 0.8051 shows a strong and positive correlation between at least two of the variables.

```
lm_out <- lm(strain.when.failed ~ percent.binder + loading.rate + ambient.temperature, data = s)
summary(lm_out)
```

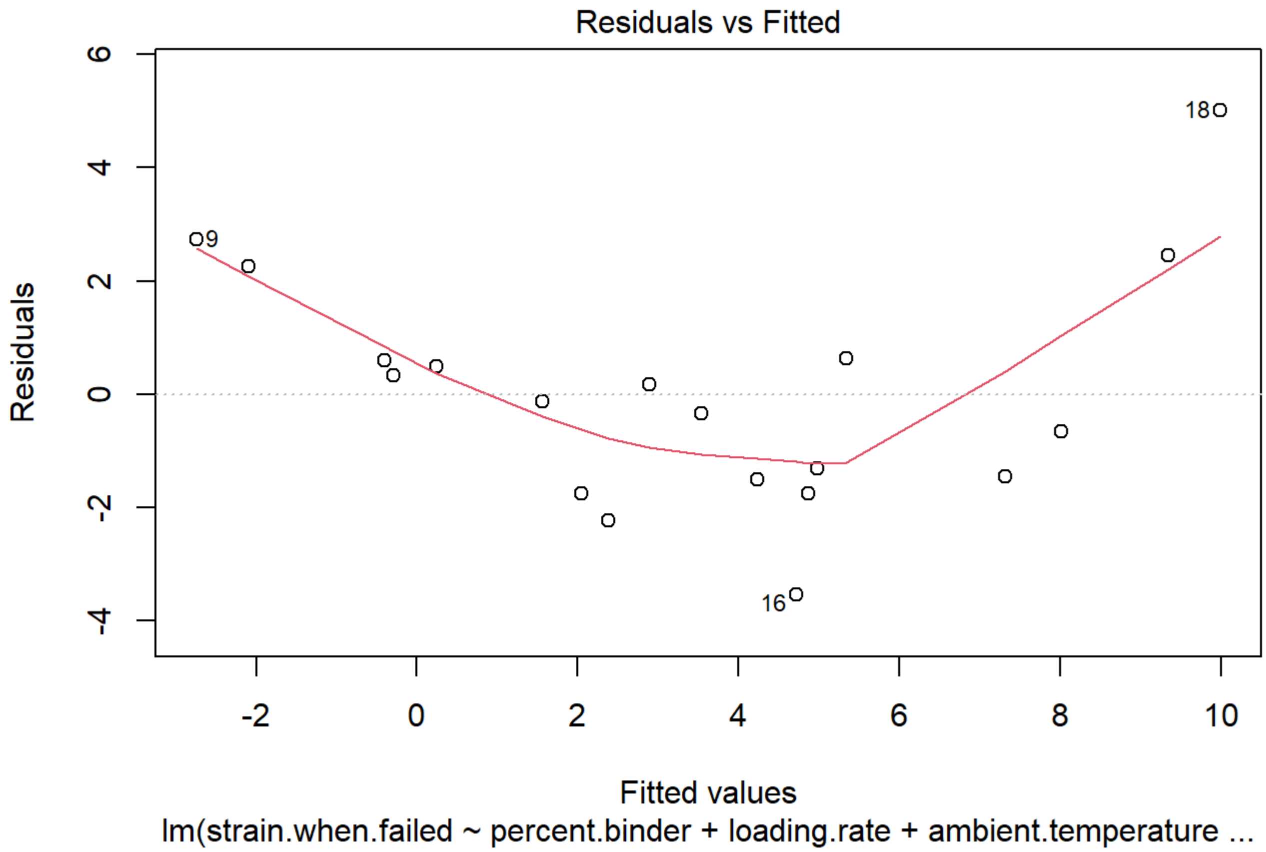
```
Call:
lm(formula = strain.when.failed ~ percent.binder + loading.rate + 
    ambient.temperature, data = s)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5466 -1.4827 -0.1190  0.6097  5.0135 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -5.61130    2.04575  -2.743 0.015100 *  
percent.binder       0.66754    0.21168   3.154 0.006558 ** 
loading.rate        -1.23535    0.57949  -2.132 0.049966 *  
ambient.temperature  0.07319    0.01438   5.090 0.000133 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.241 on 15 degrees of freedom
Multiple R-squared:  0.7601,	Adjusted R-squared:  0.7121
F-statistic: 15.84 on 3 and 15 DF,  p-value: 6.447e-05
```

```
plot(lm_out,which=2)
```




```
plot(lm_out,which=1)
```



Interpretation: The Q-Q plot for the data in relation to strain is lightly-tailed, making the data not quite normal. Based on the residual plot, the variance is not very consistent for the data. Additionally, since the p-value is much lower than 0.05, we should reject the null hypothesis that the means for all of the variables equals 0. Finally, the R-squared value of 0.8051 shows a strong and positive correlation between at least two of the variables. Interpretation:

8.5

Here is my answer to question 8.5

```
link<-("C:/Users/ASEARLY/Downloads/CompanionAsset_9780128230435_DataSets (5)/Data Tables 4th edit:
yt<-read.table(link,col.names=c("dbh","height","age","grav","weight"))
```

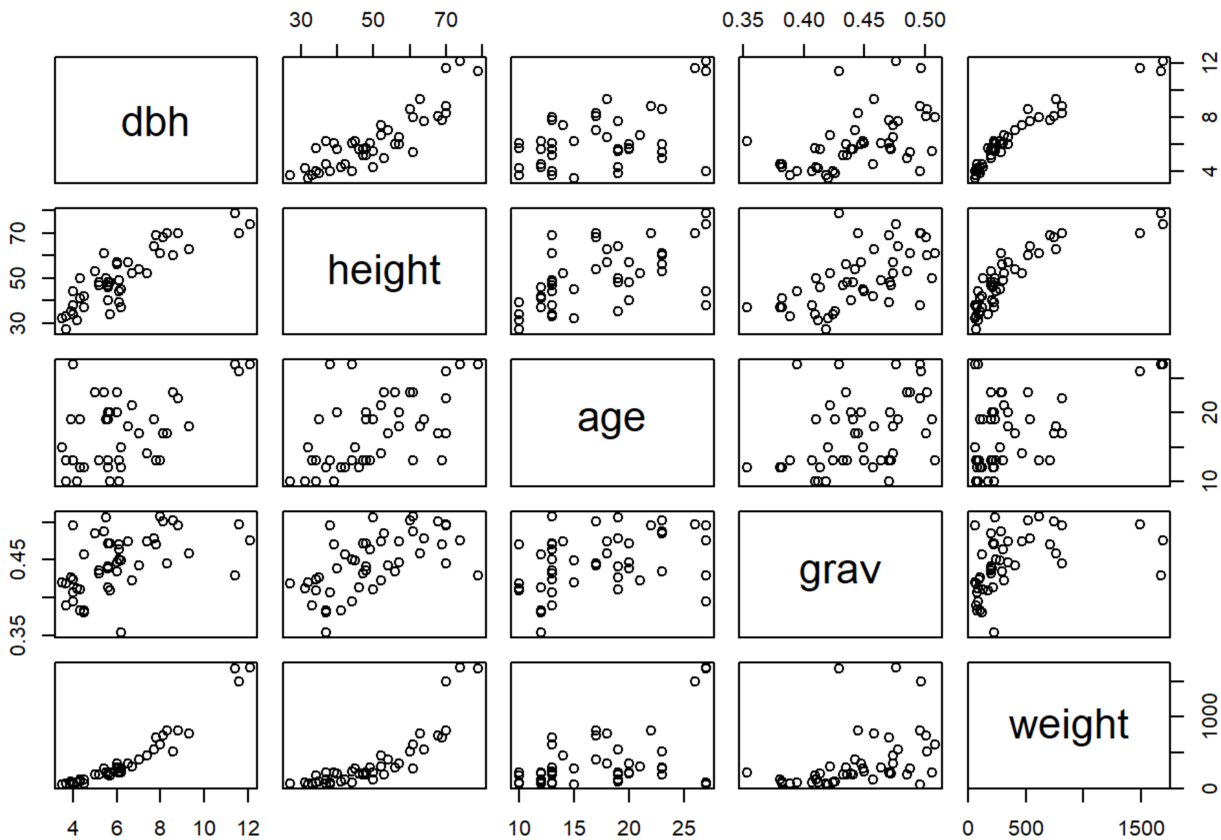
```
Warning in read.table(link, col.names = c("dbh", "height", "age", "grav", :
header and 'col.names' are of different lengths
```

```
head(yt)
```

```
  dbh height age  grav weight
1 5.7     34  10 0.409    174
2 8.1     68  17 0.501    745
3 8.3     70  17 0.445    814
```

```
4  7.0      54  17 0.442    408
5  6.2      37  12 0.353    226
6 11.4      79  27 0.429   1675
```

```
n<-nrow(yt)
plot(yt)
```

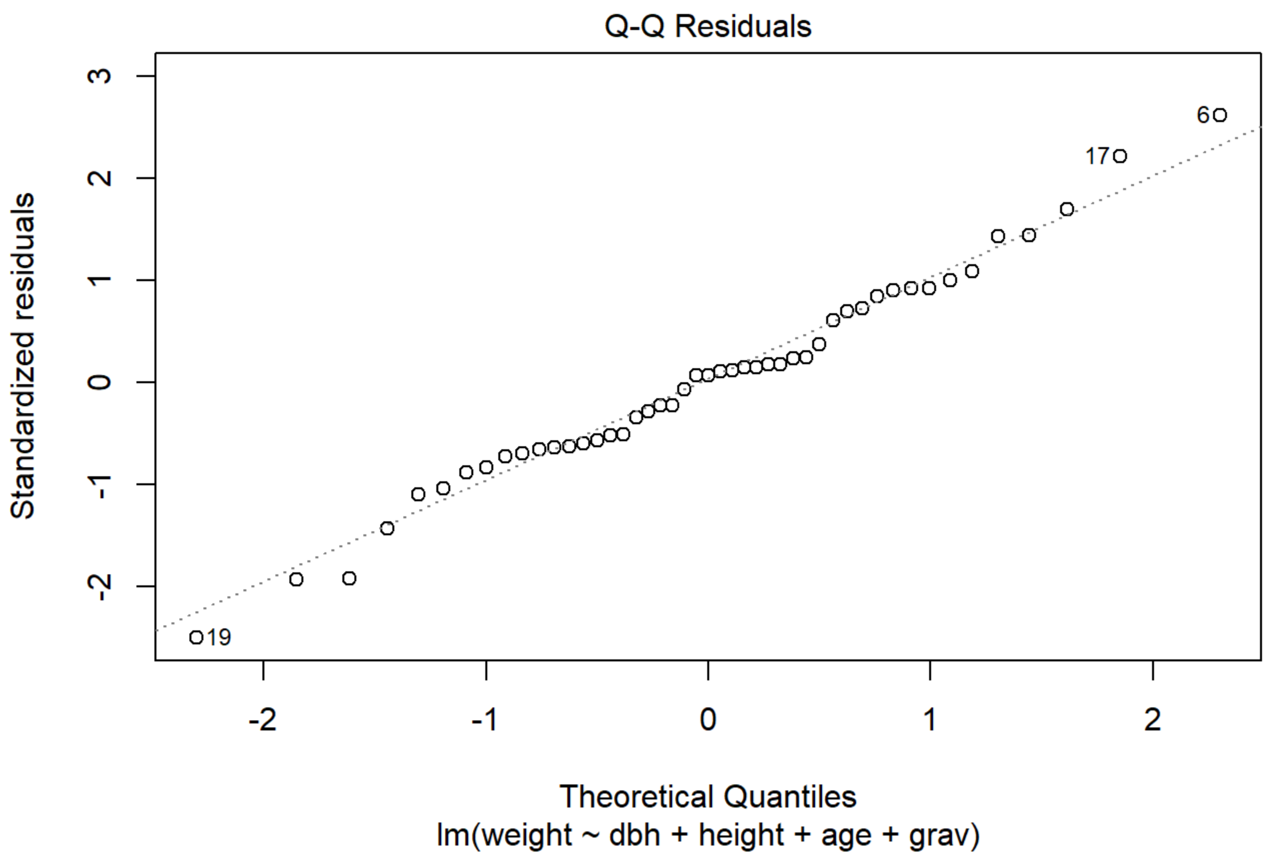



```
lm_out <- lm(weight ~ dbh + height + age + grav, data = yt)
summary(lm_out)
```

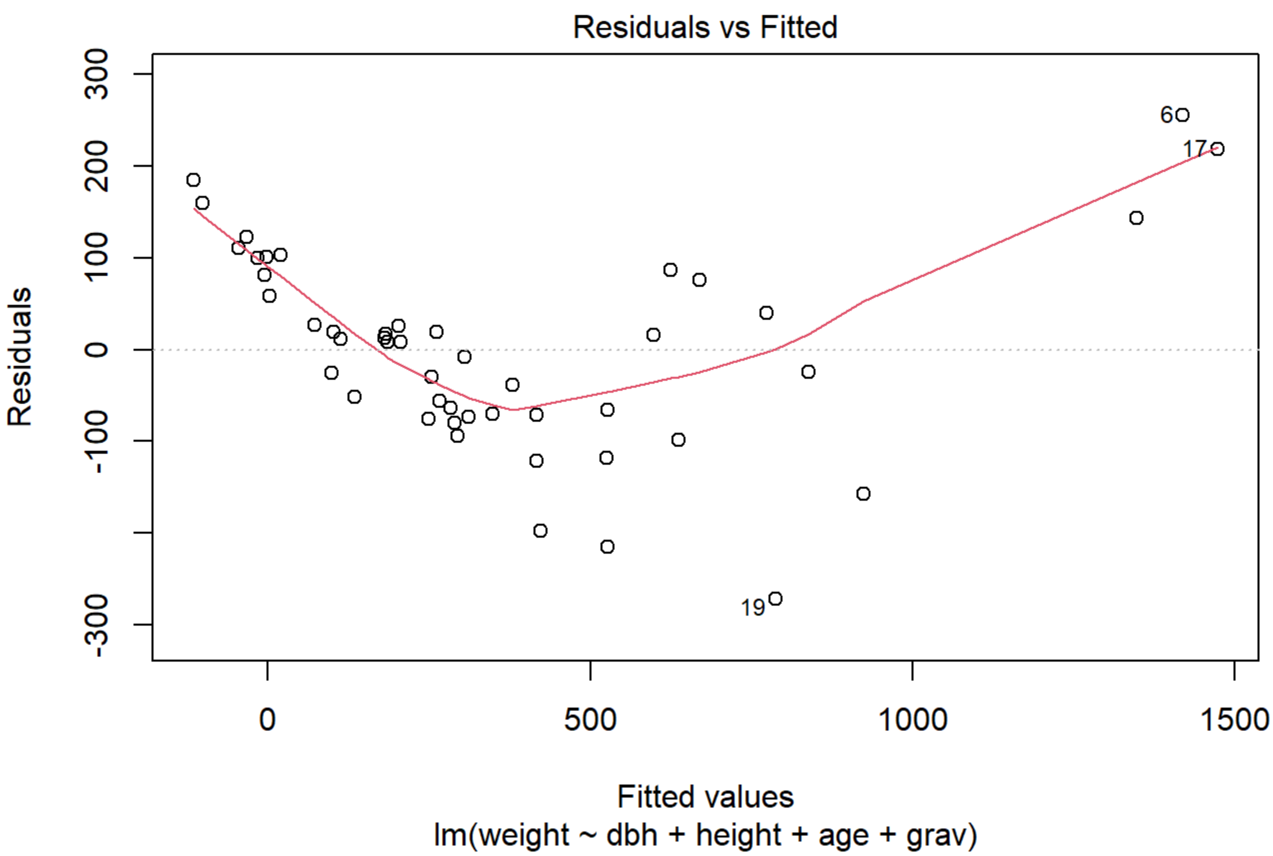
```
Call:
lm(formula = weight ~ dbh + height + age + grav, data = yt)

Residuals:
     Min       1Q   Median       3Q      Max 
-272.190  -70.956    8.219   78.582  255.427 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -379.248    206.691  -1.835   0.0736 .  
dbh           170.220     16.238  10.483 2.69e-13 ***
height          1.900      3.017   0.630   0.5322    
```

```
age              8.146      4.020   2.026   0.0491 *
grav         -1192.868    548.927  -2.173   0.0355 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 115.1 on 42 degrees of freedom
Multiple R-squared:  0.922, Adjusted R-squared:  0.9145
F-statistic: 124.1 on 4 and 42 DF,  p-value: < 2.2e-16
```

```
plot(lm_out,which=2)
```




```
plot(lm_out,which=1)
```



a

The Q-Q plot shows that the data appears to be normal, however the residual plot shows that the variance of the data is not constant, meaning the equation may not actually be very useful. The model is not quite adequate due the the variance not being constant.

b

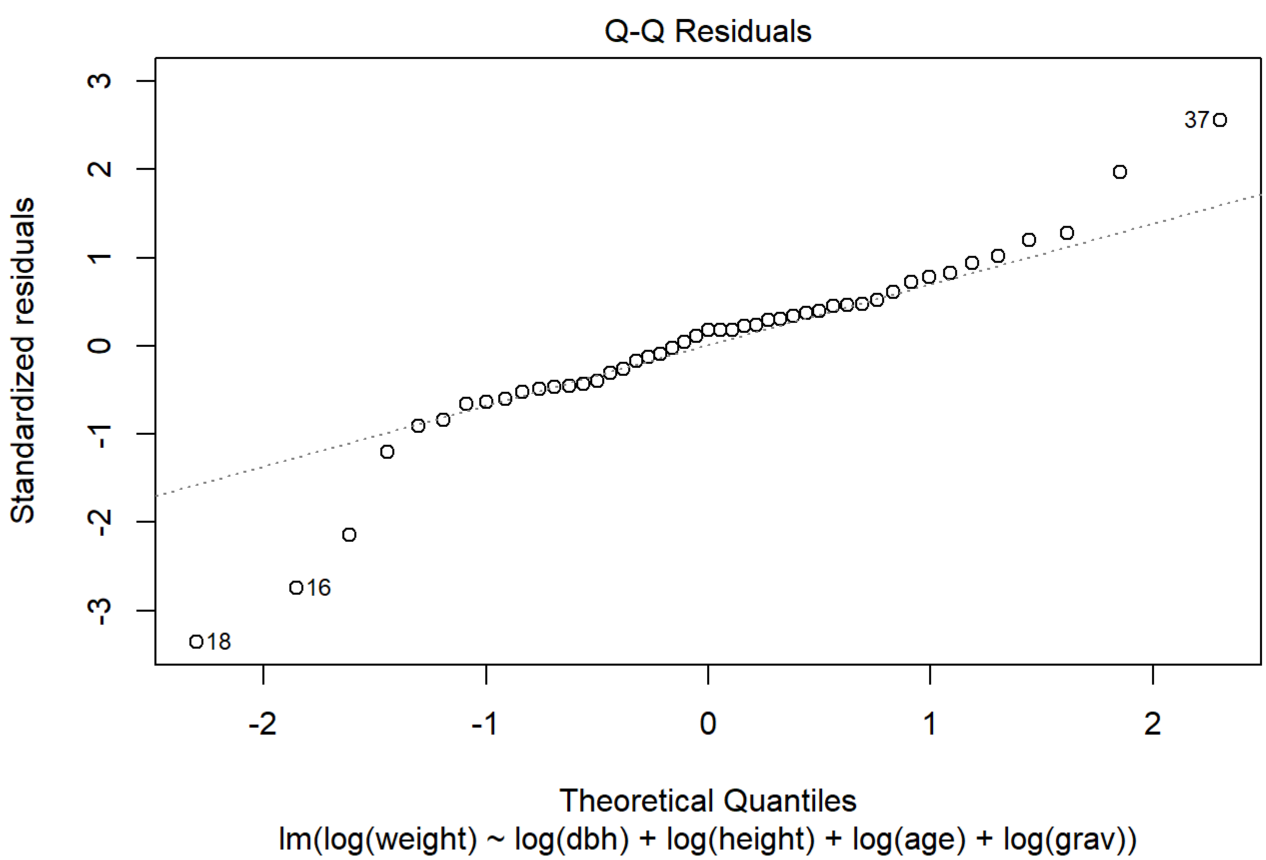
The Q-Q plot was very strong, suggesting normal data, however the variance was not constant throughout, meaning that a log transformation should be used on the data

```
pt<-lm(log(weight) ~ log(dbh) + log(height) + log(age) + log(grav), data = yt)
summary(pt)
```

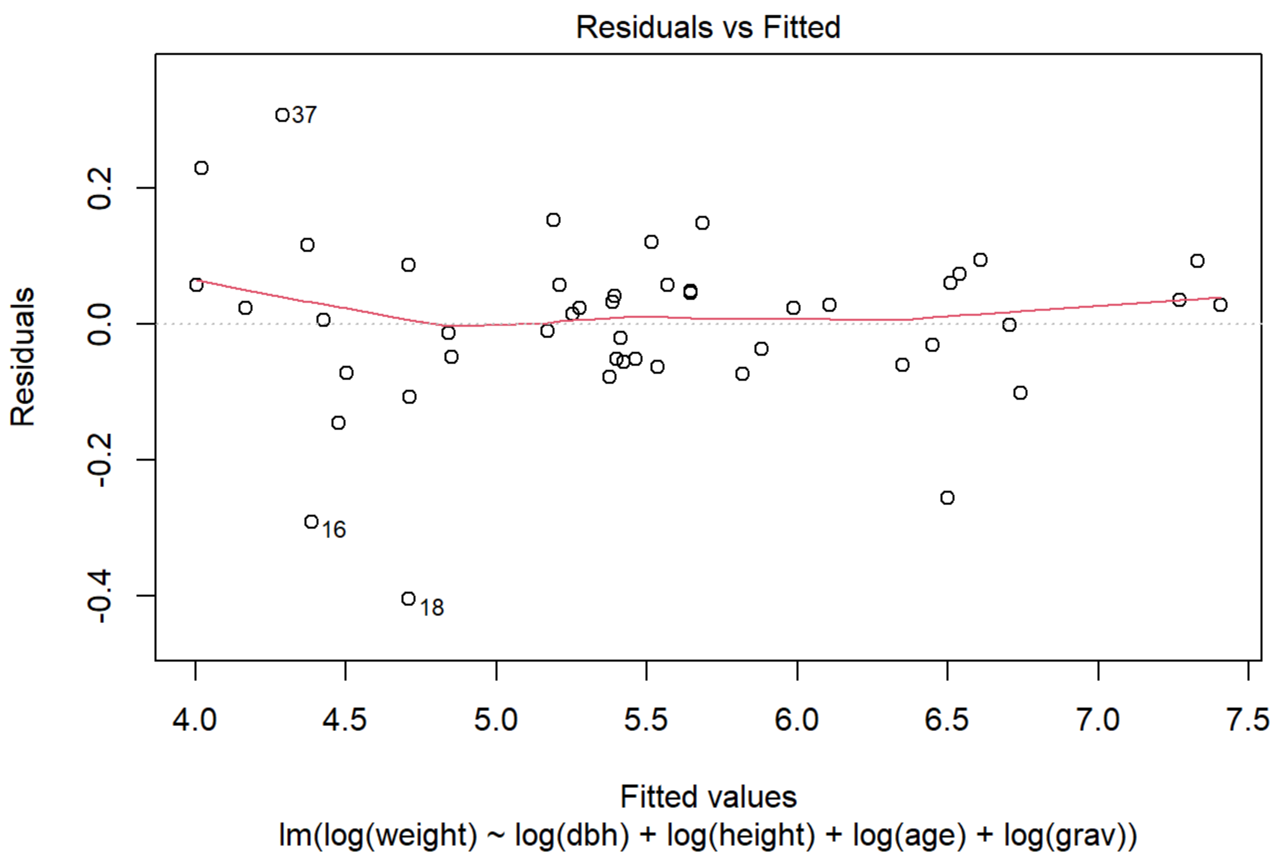
```
Call:
lm(formula = log(weight) ~ log(dbh) + log(height) + log(age) +
    log(grav), data = yt)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40546 -0.05483  0.02230  0.05642  0.30638
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.5582     0.5653  -2.757  0.00861 ** 
log(dbh)      2.1448     0.1191  18.007  < 2e-16 ***
log(height)   0.9778     0.1696   5.765 8.64e-07 ***
log(age)     -0.1551     0.0805  -1.927  0.06083 .  
log(grav)     0.1077     0.2675   0.403  0.68910    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1264 on 42 degrees of freedom
Multiple R-squared:  0.982, Adjusted R-squared:  0.9803 
F-statistic: 572.9 on 4 and 42 DF,  p-value: < 2.2e-16
```

```r
plot(pt,which=2)
```




```r
plot(pt,which=1)
```



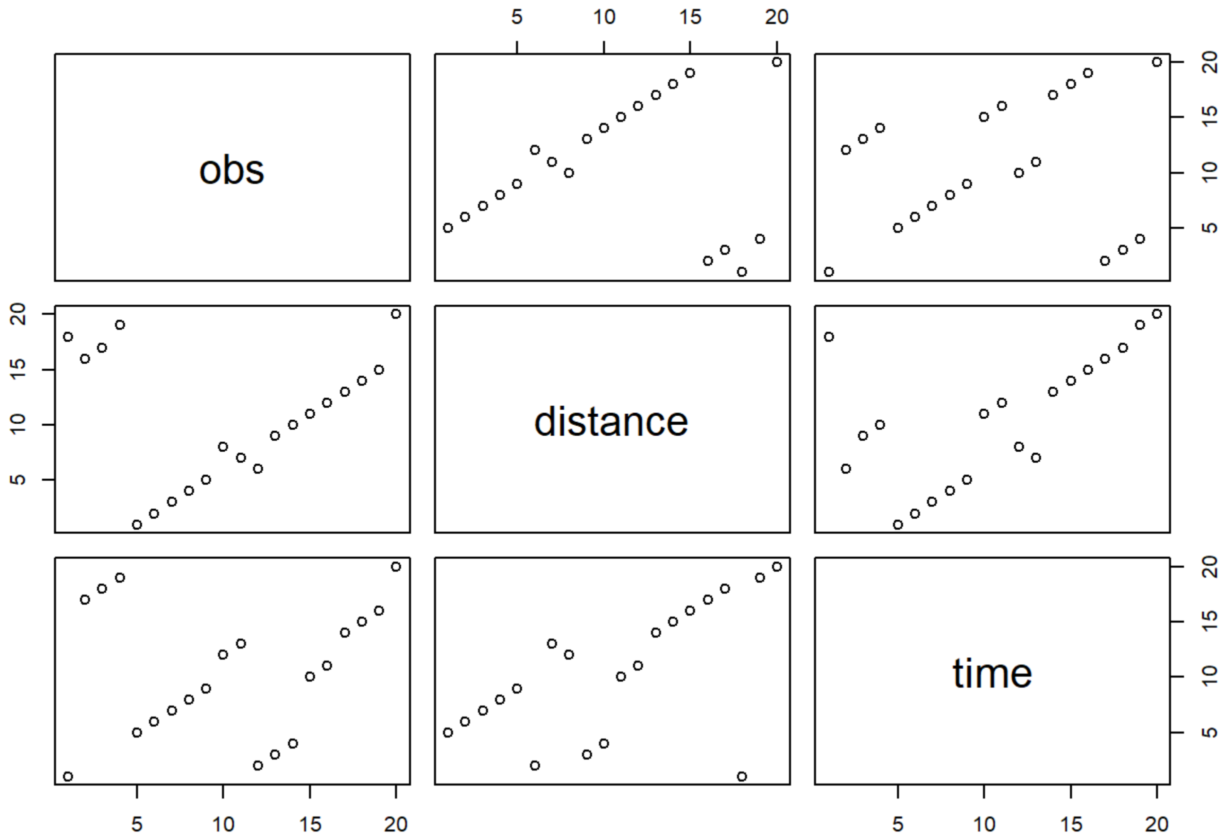
8.7

a.

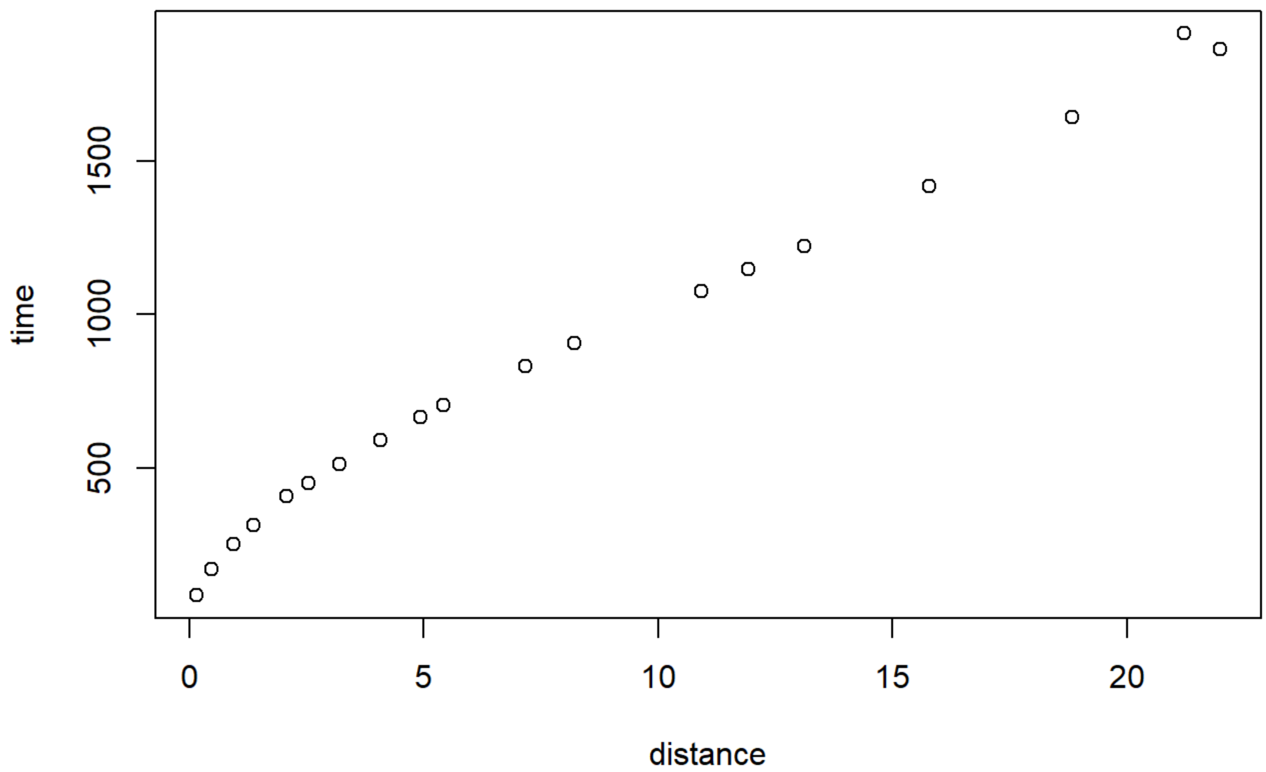
```
link<-("C:/Users/ASEARLY/Downloads/CompanionAsset_9780128230435_DataSets (5)/Data Tables 4th edit:
dbi<-read.table(link,col.names=c("obs","distance","time"))
head(dbi)
```

```
  obs distance time
1 obs distance time
2   1       85 0.15
3   2      169 0.48
4   3      251 0.95
5   4      315 1.37
6   5      408 2.08
```

```
n<-nrow(dbi)
plot(dbi)
```



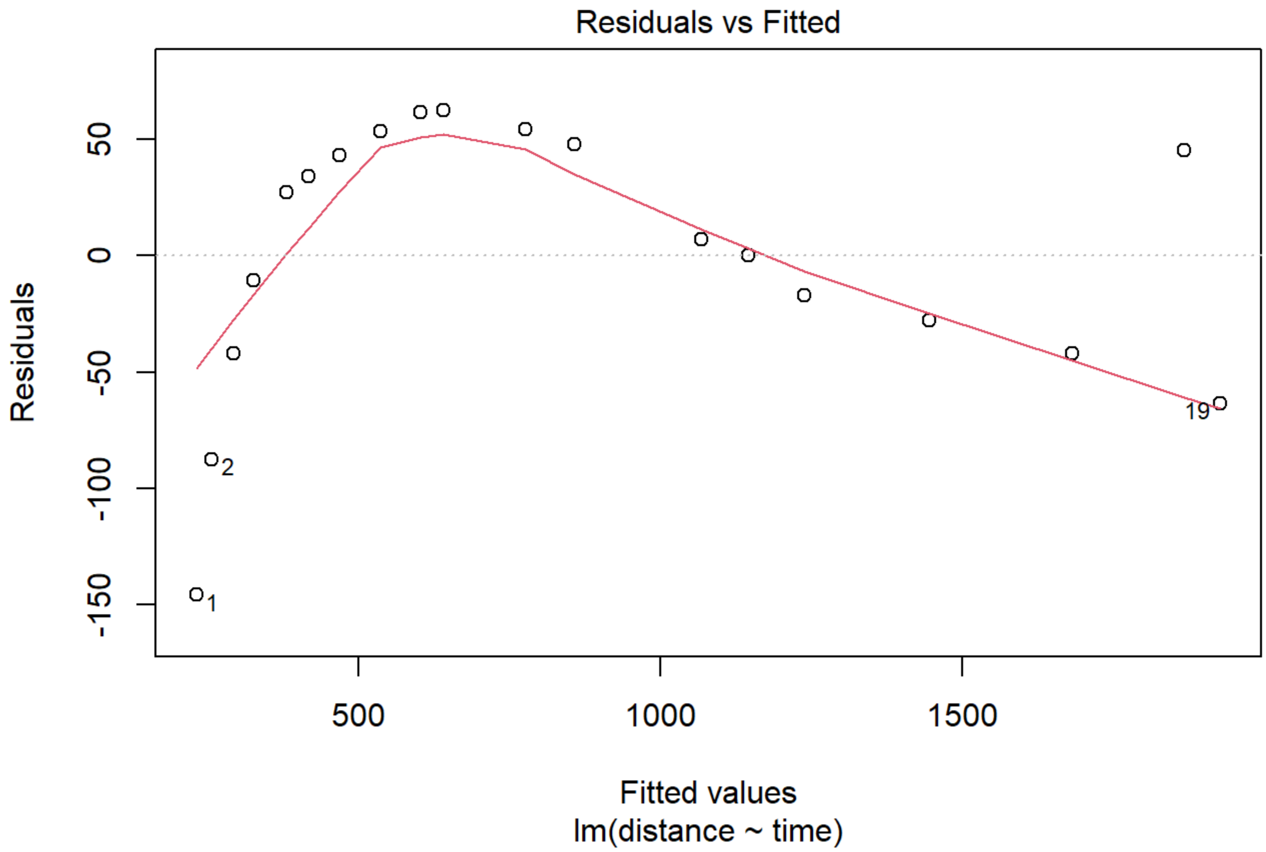

```
distance<-c(85,169,251,315,408,450,511,590,664,703,831,906,1075,1146,1222,1418,1641,1914,1864)
time<-c(.15,.48,.95,1.37,2.08,2.53,3.2,4.08,4.93,5.42,7.17,8.22,10.92,11.92,13.12,15.78,18.83,21.
plot(distance~time, ylab="time", xlab="distance")
```




```
lm1<-lm(distance~time)
summary(lm1)
```

```
Call:
lm(formula = distance ~ time)

Residuals:
     Min       1Q   Median       3Q      Max 
-145.933  -34.798    6.973   46.620   62.458 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  219.275     20.780   10.55    7e-09 ***
time          77.725      1.938   40.10   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59.1 on 17 degrees of freedom
Multiple R-squared:  0.9895,	Adjusted R-squared:  0.9889 
F-statistic:  1608 on 1 and 17 DF,  p-value: < 2.2e-16
```

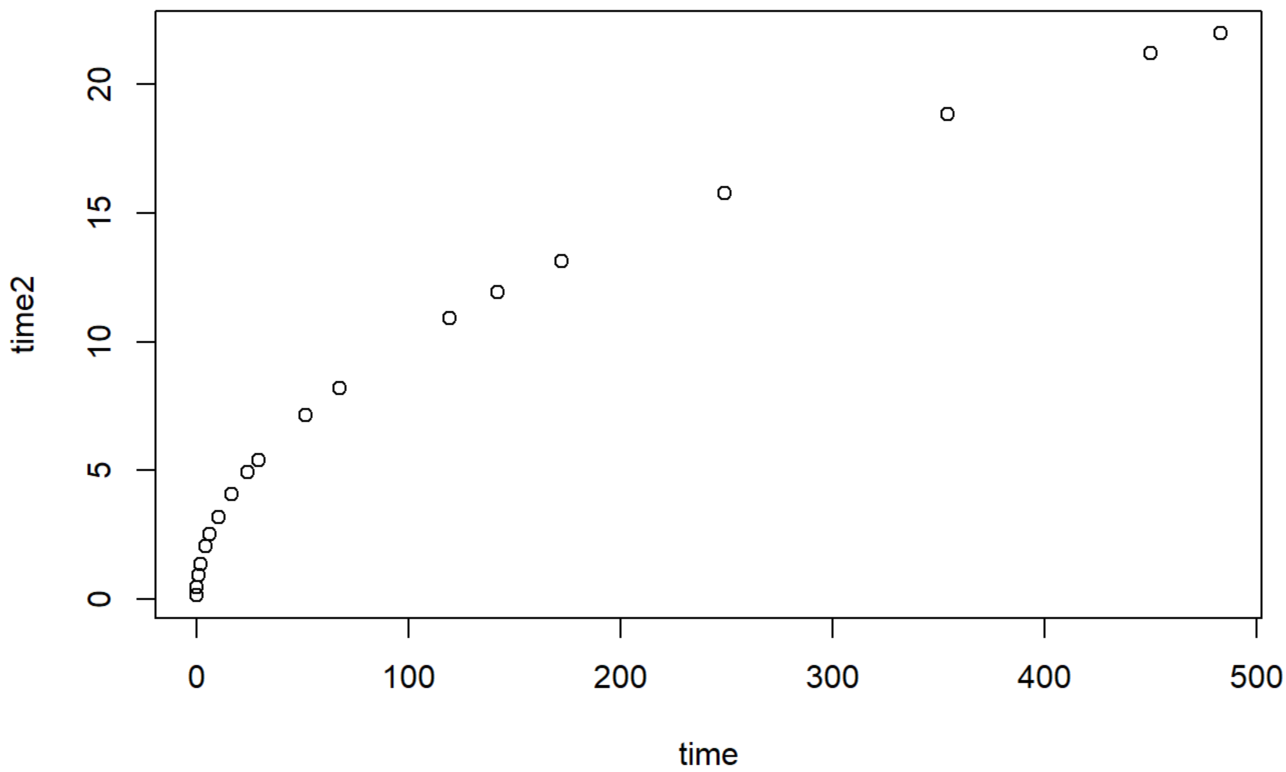
```
plot(lm1,which=1)
```



This residual plot suggests that the variances are not constant and that a linear regression model does not fit the data.

b.

```
time<-c(.15,.48,.95,1.37,2.08,2.53,3.2,4.08,4.93,5.42,7.17,8.22,10.92,11.92,13.12,15.78,18.83,21.
time2 <- time^2
plot(time~time2, ylab="time2", xlab="time")
```




```r
summary(lm1)
```

```
Call:
lm(formula = distance ~ time)

Residuals:
     Min       1Q   Median       3Q      Max
-145.933  -34.798    6.973   46.620   62.458

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  219.275     20.780   10.55    7e-09 ***
time          77.725      1.938   40.10   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59.1 on 17 degrees of freedom
Multiple R-squared:  0.9895,    Adjusted R-squared:  0.9889
F-statistic:  1608 on 1 and 17 DF,  p-value: < 2.2e-16
```

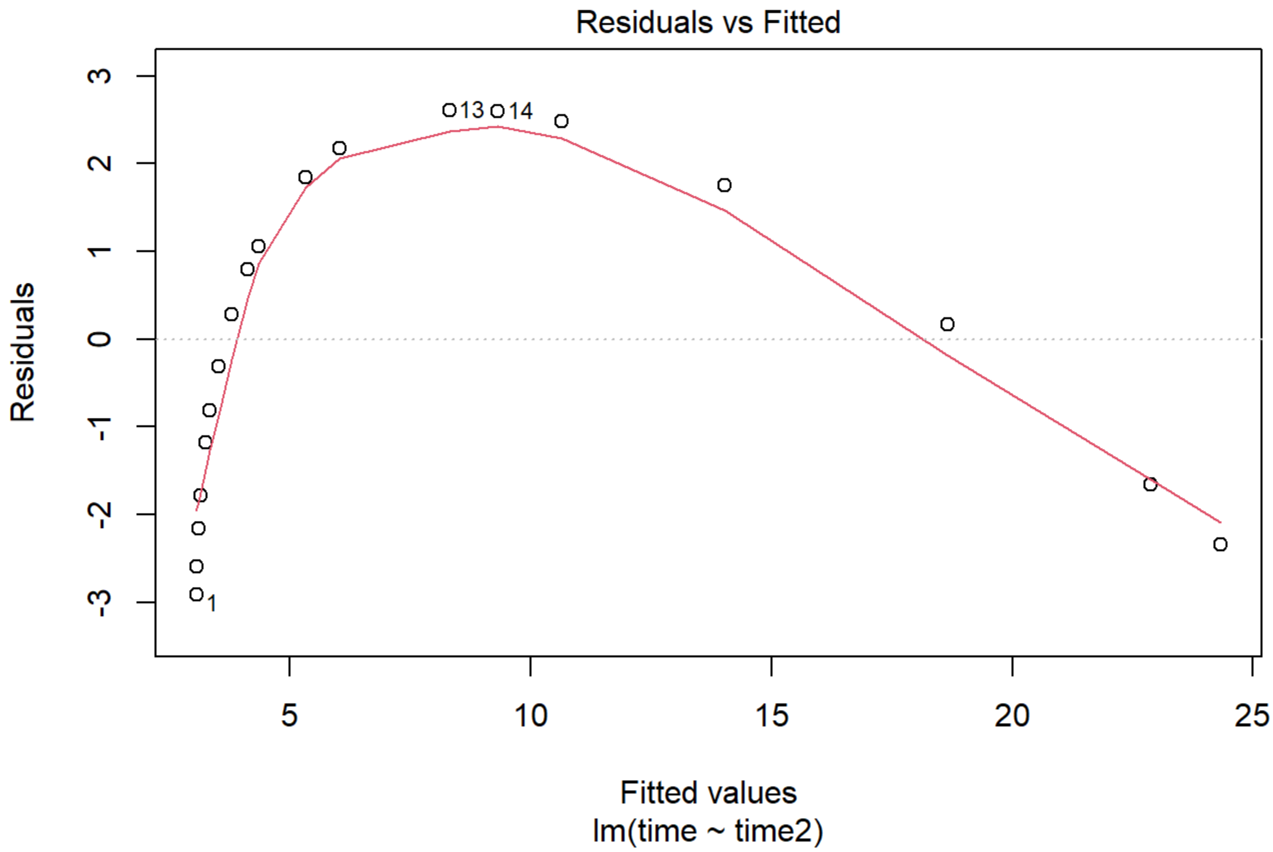
Based on the plot and the R-squared being 0.92, there appears to be a strong, linear relationship between time and time^2.

###c.

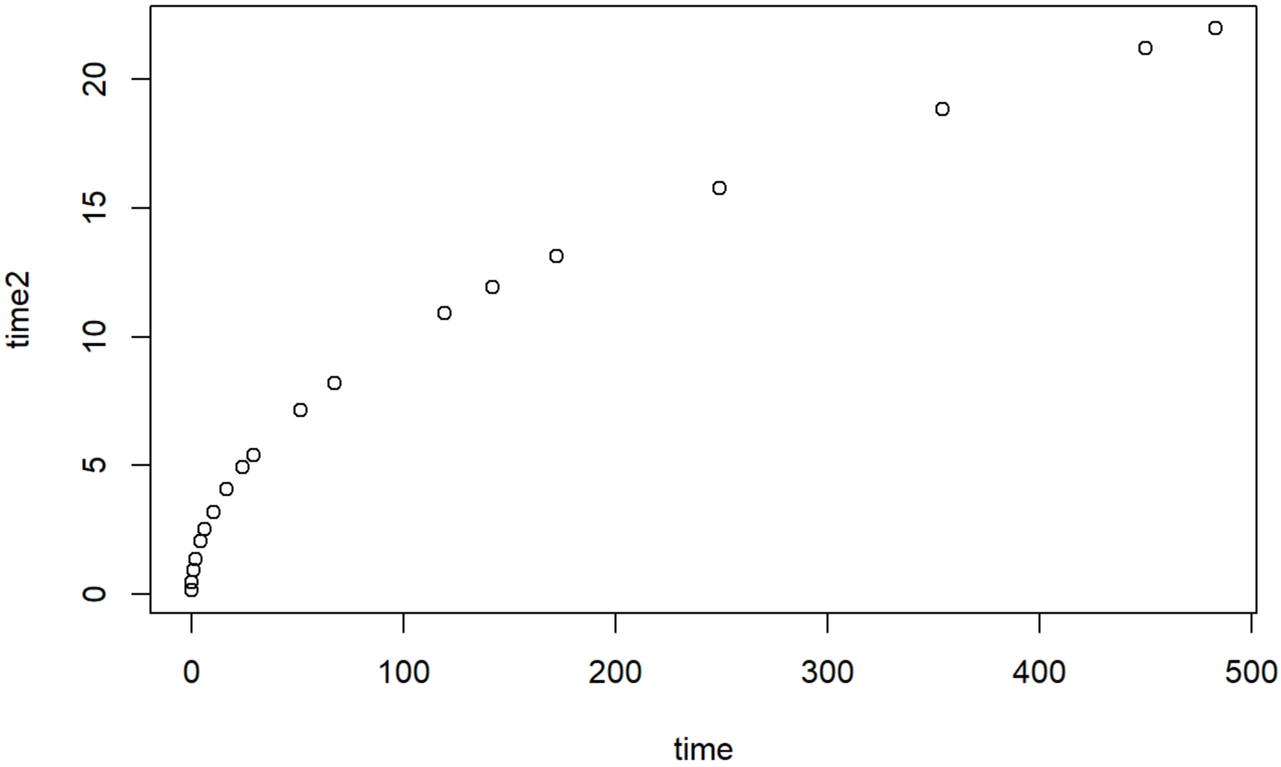
```
lm1<-lm(time~time2)
summary(lm1)
```

```
Call:
lm(formula = time ~ time2)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.9174 -1.7193  0.1621  1.7992  2.6066 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.066429   0.566260   5.415 4.64e-05 ***
time2       0.044001   0.002956  14.883 3.50e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.974 on 17 degrees of freedom
Multiple R-squared:  0.9287,    Adjusted R-squared:  0.9245 
F-statistic: 221.5 on 1 and 17 DF,  p-value: 3.504e-11
```

```
plot(lm1,which=1)
```




```r
plot(time~time2, ylab="time2", xlab="time")
```



There is clearly a pattern in this residual plot, meaning that a linear model would not be the best fit for this data. There is clearly a pattern in this residuals graph,