Homework 6

Devin Etcitty

4/2/2017

STAT4205

dce2108

```
library(alr4)
library(ggplot2)
library(tidyverse)
```

Problem 7.6

7.6.1 - 7.6.4

stopping dataframe

heteroscedastic homosecdastic

What are theset erms?

```
dim(stopping)
```

```
## [1] 62  2
```

```
names(stopping)
```

```
## [1] "Speed"    "Distance"
```

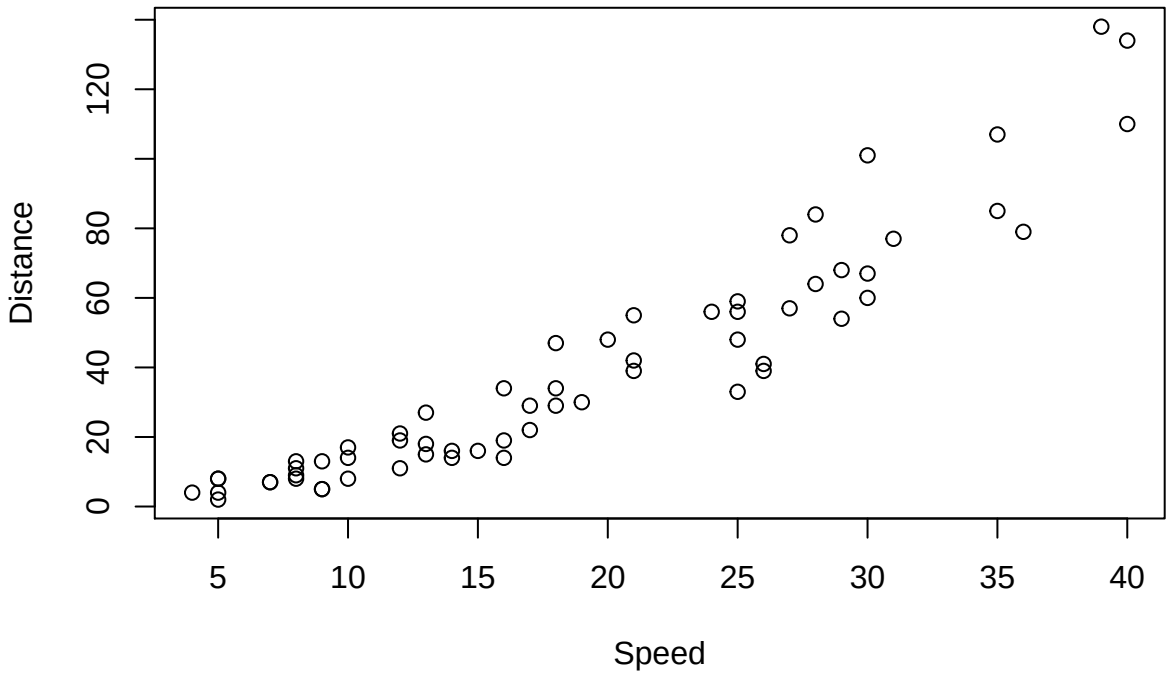
```
sapply(stopping, class)
```

```
##     Speed  Distance
## "integer" "integer"
```

```
head(stopping)
```

```
##   Speed Distance
## 1     4        4
## 2     5        2
## 3     5        4
## 4     5        8
## 5     5        8
## 6     7        7
```

```
plot(stopping)
```




```
qd <- lm(Distance ~ Speed + I(Speed^2),data=stopping)
summary(qd)
```

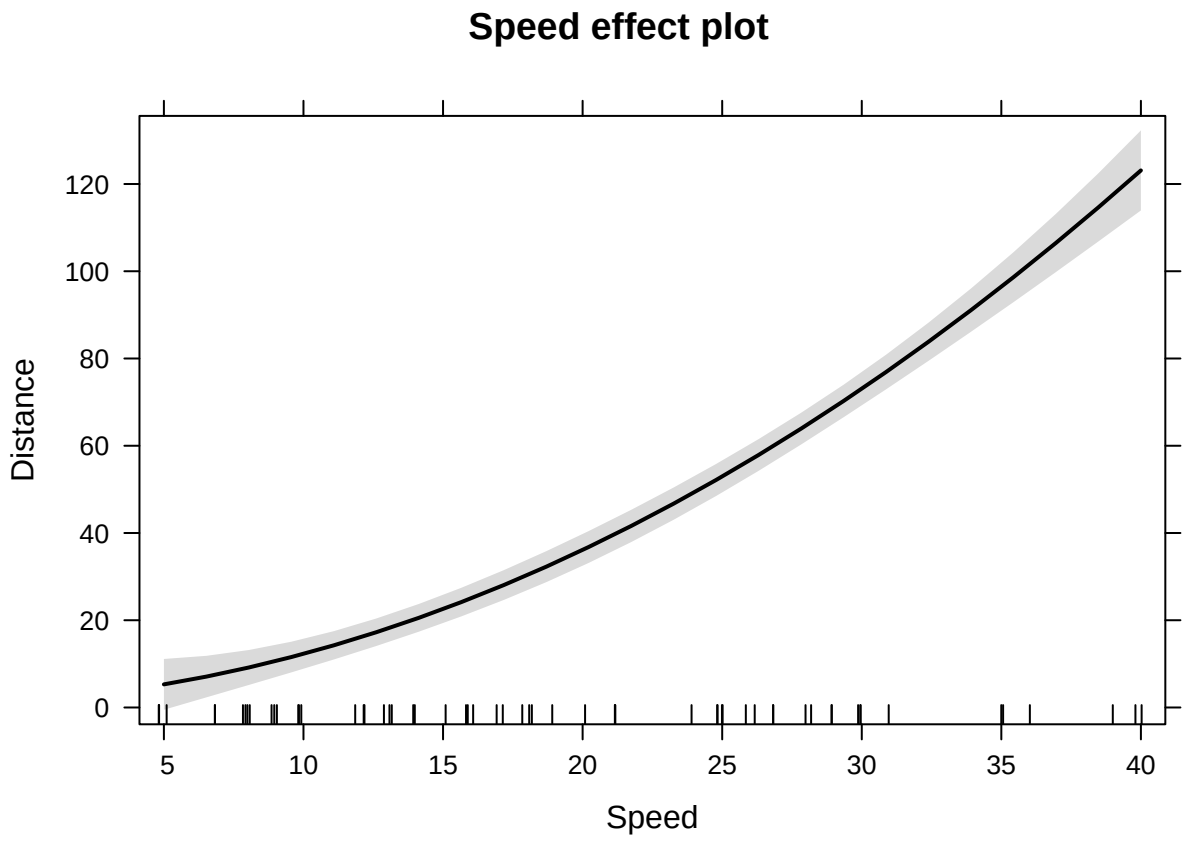
```
##
## Call:
## lm(formula = Distance ~ Speed + I(Speed^2), data = stopping)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5192  -5.4527  -0.5519   3.8442  27.9373
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.58036    5.10266   0.310    0.758
## Speed        0.41607    0.55641   0.748    0.458
## I(Speed^2)   0.06556    0.01303   5.033 4.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.927 on 59 degrees of freedom
## Multiple R-squared:  0.9144, Adjusted R-squared:  0.9115
## F-statistic: 315.3 on 2 and 59 DF,  p-value: < 2.2e-16
```

```
p1 <- lm(Distance ~ poly(Speed, 2, raw=TRUE), data = stopping)
summary(p1)
```

```
##
## Call:
## lm(formula = Distance ~ poly(Speed, 2, raw = TRUE), data = stopping)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5192  -5.4527  -0.5519   3.8442  27.9373
```

```
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.58036    5.10266   0.310    0.758
## poly(Speed, 2, raw = TRUE)1  0.41607    0.55641   0.748    0.458
## poly(Speed, 2, raw = TRUE)2  0.06556    0.01303   5.033 4.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.927 on 59 degrees of freedom
## Multiple R-squared:  0.9144, Adjusted R-squared:  0.9115
## F-statistic: 315.3 on 2 and 59 DF,  p-value: < 2.2e-16
```

```
print(plot(Effect("Speed", p1)))
```

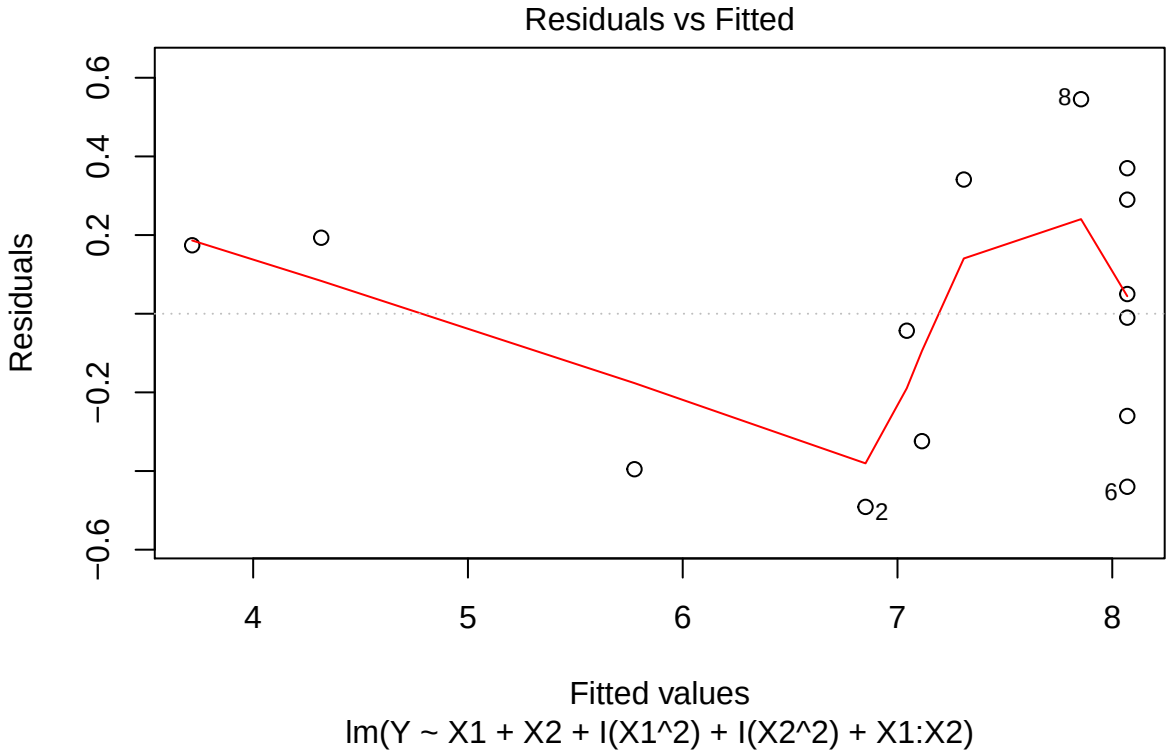


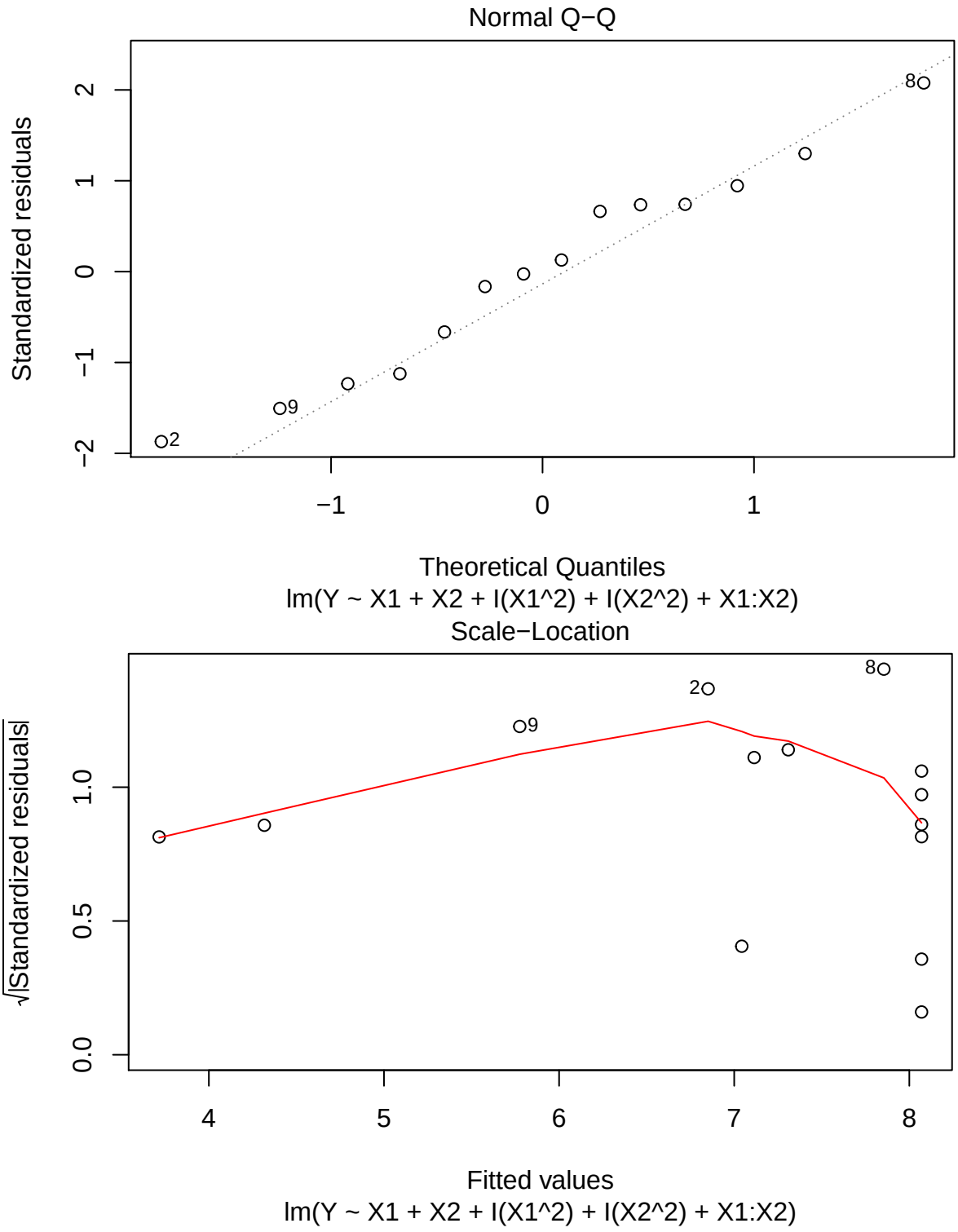

```
mod2 <- lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) +  X1:X2, data=cakes)
summary(mod2)
```

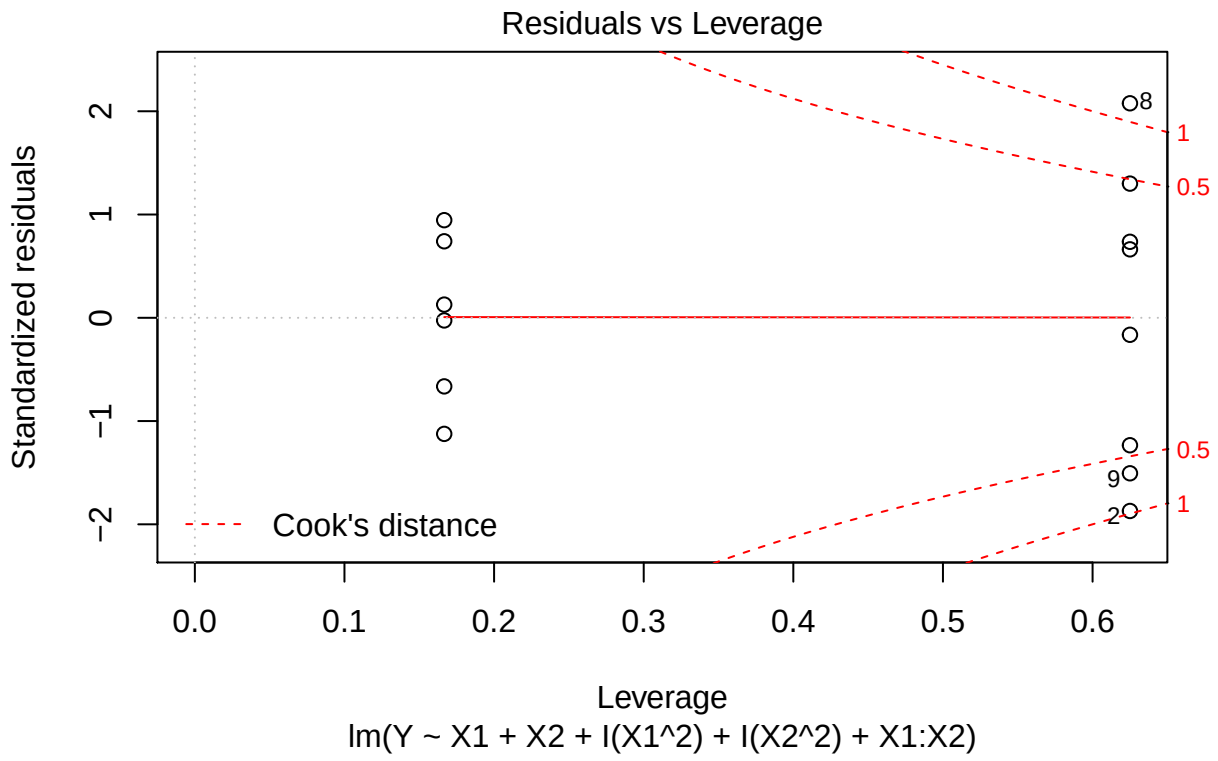
```
##
## Call:
## lm(formula = Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2, data = cakes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4912 -0.3080  0.0200  0.2658  0.5454
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -2.204e+03  2.416e+02  -9.125 1.67e-05 ***
## X1           2.592e+01  4.659e+00   5.563 0.000533 ***
## X2           9.918e+00  1.167e+00   8.502 2.81e-05 ***
## I(X1^2)     -1.569e-01  3.945e-02  -3.977 0.004079 **
## I(X2^2)     -1.195e-02  1.578e-03  -7.574 6.46e-05 ***
## X1:X2       -4.163e-02  1.072e-02  -3.883 0.004654 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4288 on 8 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9167
## F-statistic:  29.6 on 5 and 8 DF,  p-value: 5.864e-05
```

```
plot(mod2)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2)
Scale−Location
√|Standardized residuals|
Fitted values
lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2)




```
vcov(mod2)
```

```
##               (Intercept)            X1            X2       I(X1^2)
## (Intercept) 58361.2541243 -6.670524e+02 -2.668210e+02  2.488704e+00
## X1           -667.0524374  2.170546e+01  1.642140e+00 -1.089175e-01
## X2           -266.8209749  1.642140e+00  1.360860e+00 -3.351307e-03
## I(X1^2)         2.4887042 -1.089175e-01 -3.351307e-03  1.555964e-03
## I(X2^2)         0.3106805 -3.351307e-04 -1.742679e-03  4.787581e-06
## X1:X2           1.4075488 -4.021568e-02 -4.021568e-03  7.495493e-17
##                   I(X2^2)         X1:X2
## (Intercept)  3.106805e-01  1.407549e+00
## X1          -3.351307e-04 -4.021568e-02
## X2          -1.742679e-03 -4.021568e-03
## I(X1^2)      4.787581e-06  7.495493e-17
## I(X2^2)      2.489542e-06 -3.702888e-18
## X1:X2       -3.702888e-18  1.149019e-04
```

```
sqrt(diag(vcov(mod2)))
```

```
##  (Intercept)           X1           X2      I(X1^2)      I(X2^2)
## 2.415807e+02 4.658911e+00 1.166559e+00 3.944571e-02 1.577828e-03
##        X1:X2
## 1.071923e-02
```

```
summary(mod2)$coef
```

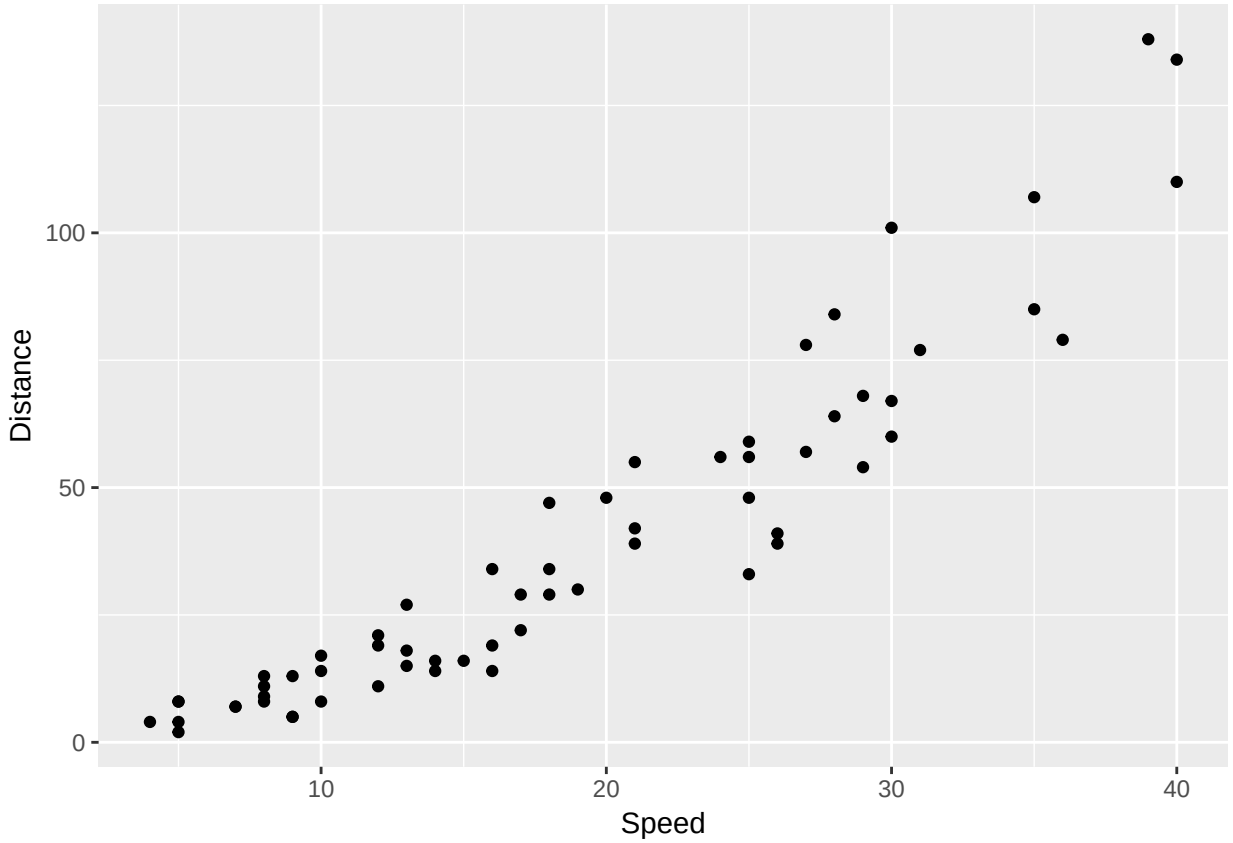
```
##                Estimate   Std. Error   t value     Pr(>|t|)
## (Intercept) -2204.484987 2.415807e+02 -9.125251 1.673889e-05
## X1             25.917558 4.658911e+00  5.563007 5.327892e-04
## X2              9.918267 1.166559e+00  8.502156 2.810176e-05
## I(X1^2)        -0.156875 3.944571e-02 -3.976985 4.078860e-03
## I(X2^2)        -0.011950 1.577828e-03 -7.573701 6.462492e-05
```

```
## X1:X2          -0.041625 1.071923e-02 -3.883207 4.653652e-03
```

```
z <- with(cakes, polym(X1, X2, degree=2, raw=TRUE))
#summary(z)
model.matrix(~ z)
```

```
##    (Intercept)     z1.0    z2.0     z0.1     z1.1     z0.2
## 1            1 33.00000 1089.00 340.0000 11220.00 115600.0
## 2            1 37.00000 1369.00 340.0000 12580.00 115600.0
## 3            1 33.00000 1089.00 360.0000 11880.00 129600.0
## 4            1 37.00000 1369.00 360.0000 13320.00 129600.0
## 5            1 35.00000 1225.00 350.0000 12250.00 122500.0
## 6            1 35.00000 1225.00 350.0000 12250.00 122500.0
## 7            1 35.00000 1225.00 350.0000 12250.00 122500.0
## 8            1 37.82843 1430.99 350.0000 13239.95 122500.0
## 9            1 32.17157 1035.01 350.0000 11260.05 122500.0
## 10           1 35.00000 1225.00 364.1421 12744.97 132599.5
## 11           1 35.00000 1225.00 335.8579 11755.03 112800.5
## 12           1 35.00000 1225.00 350.0000 12250.00 122500.0
## 13           1 35.00000 1225.00 350.0000 12250.00 122500.0
## 14           1 35.00000 1225.00 350.0000 12250.00 122500.0
## attr(,"assign")
## [1] 0 1 1 1 1 1
```

```
ggplot(stopping, aes(Speed, Distance)) +
  geom_point()
```



7.6.1

This graph supports a quadratic regression model because the graph has a curve and less linear.

7.6.2

7.6.3

7.6.4

```
head(stopping)
```

```
##   Speed Distance
## 1     4        4
## 2     5        2
## 3     5        4
## 4     5        8
## 5     5        8
## 6     7        7
```

```
summary(stopping)
```

```
##      Speed          Distance
##  Min.   : 4.00   Min.   :  2.00
##  1st Qu.:10.00   1st Qu.: 13.25
##  Median :17.50   Median : 29.50
##  Mean   :18.92   Mean   : 39.31
##  3rd Qu.:26.75   3rd Qu.: 56.75
##  Max.   :40.00   Max.   :138.00
```

Problem 7.8

jevons dataframe

```
dim(jevons)
```

```
## [1] 5 6
```

```
names(jevons)
```

```
## [1] "Age"    "n"      "Weight" "SD"     "Min"    "Max"
```

```
sapply(jevons, class)
```

```
##       Age         n    Weight        SD       Min       Max
## "integer" "integer" "numeric" "numeric" "numeric" "numeric"
```

```
head(jevons)
```

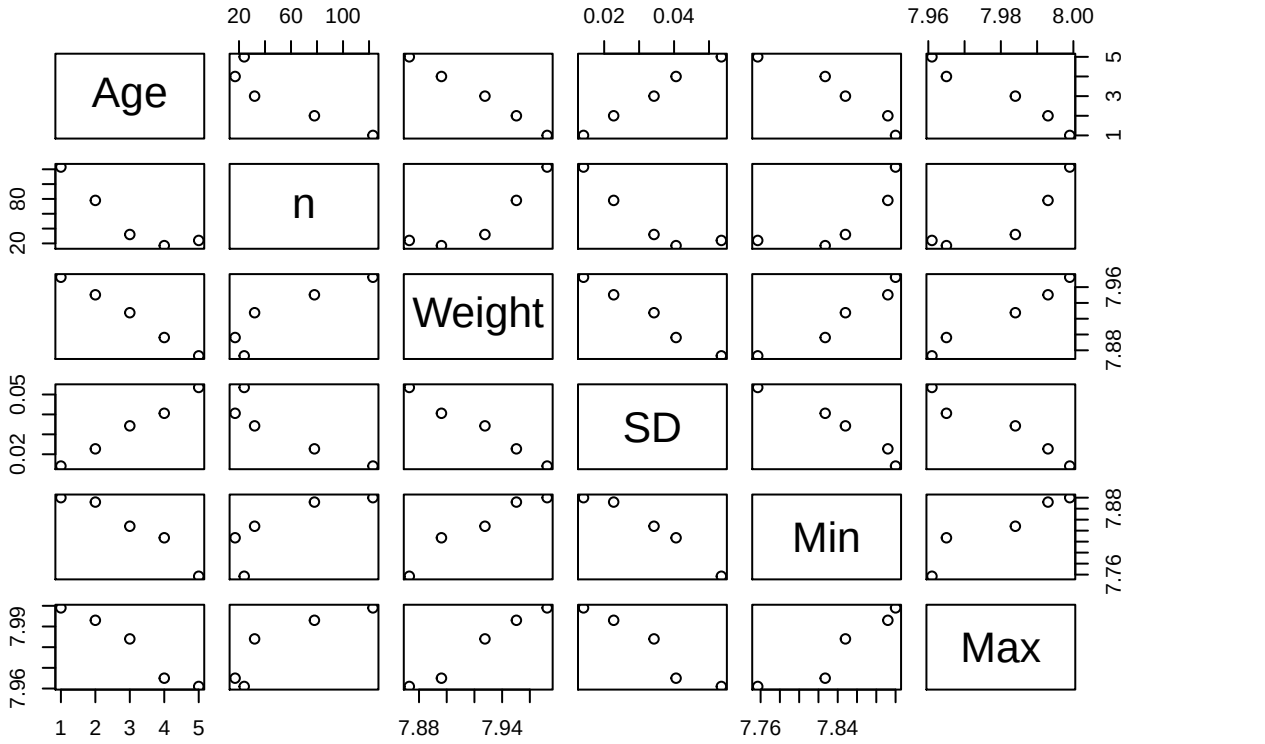
```
##   Age   n Weight      SD   Min   Max
## 1   1 123 7.9725 0.01409 7.900 7.999
## 2   2  78 7.9503 0.02272 7.892 7.993
## 3   3  32 7.9276 0.03426 7.848 7.984
## 4   4  17 7.8962 0.04057 7.827 7.965
## 5   5  24 7.8730 0.05353 7.757 7.961
```

```
summary(jevons)
```

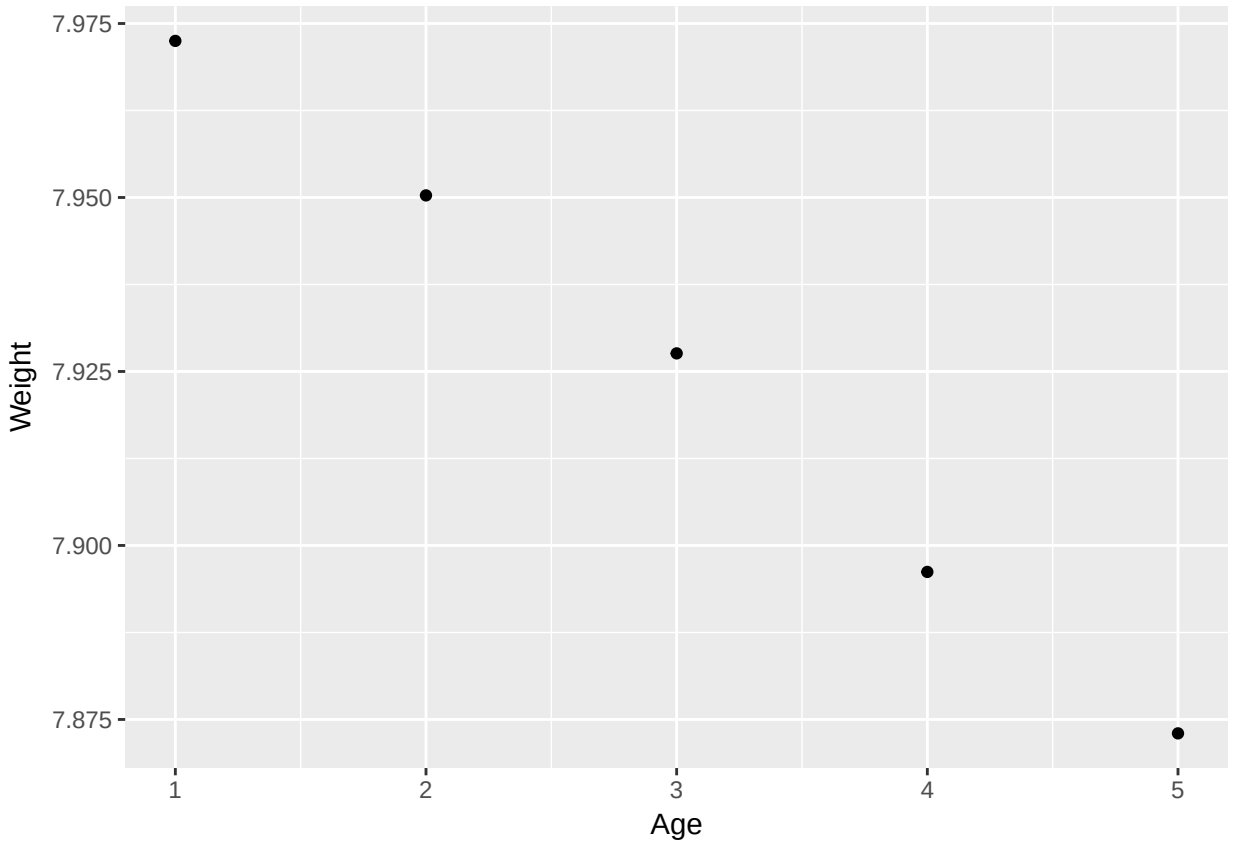
```
##       Age          n             Weight            SD
##  Min.   :1   Min.   : 17.0   Min.   :7.873   Min.   :0.01409
##  1st Qu.:2   1st Qu.: 24.0   1st Qu.:7.896   1st Qu.:0.02272
##  Median :3   Median : 32.0   Median :7.928   Median :0.03426
```

```
##  Mean   :3   Mean   : 54.8   Mean   :7.924   Mean   :0.03303
##  3rd Qu.:4   3rd Qu.: 78.0   3rd Qu.:7.950   3rd Qu.:0.04057
##  Max.   :5   Max.   :123.0   Max.   :7.973   Max.   :0.05353
##        Min              Max
##  Min.   :7.757   Min.   :7.961
##  1st Qu.:7.827   1st Qu.:7.965
##  Median :7.848   Median :7.984
##  Mean   :7.845   Mean   :7.980
##  3rd Qu.:7.892   3rd Qu.:7.993
##  Max.   :7.900   Max.   :7.999
```

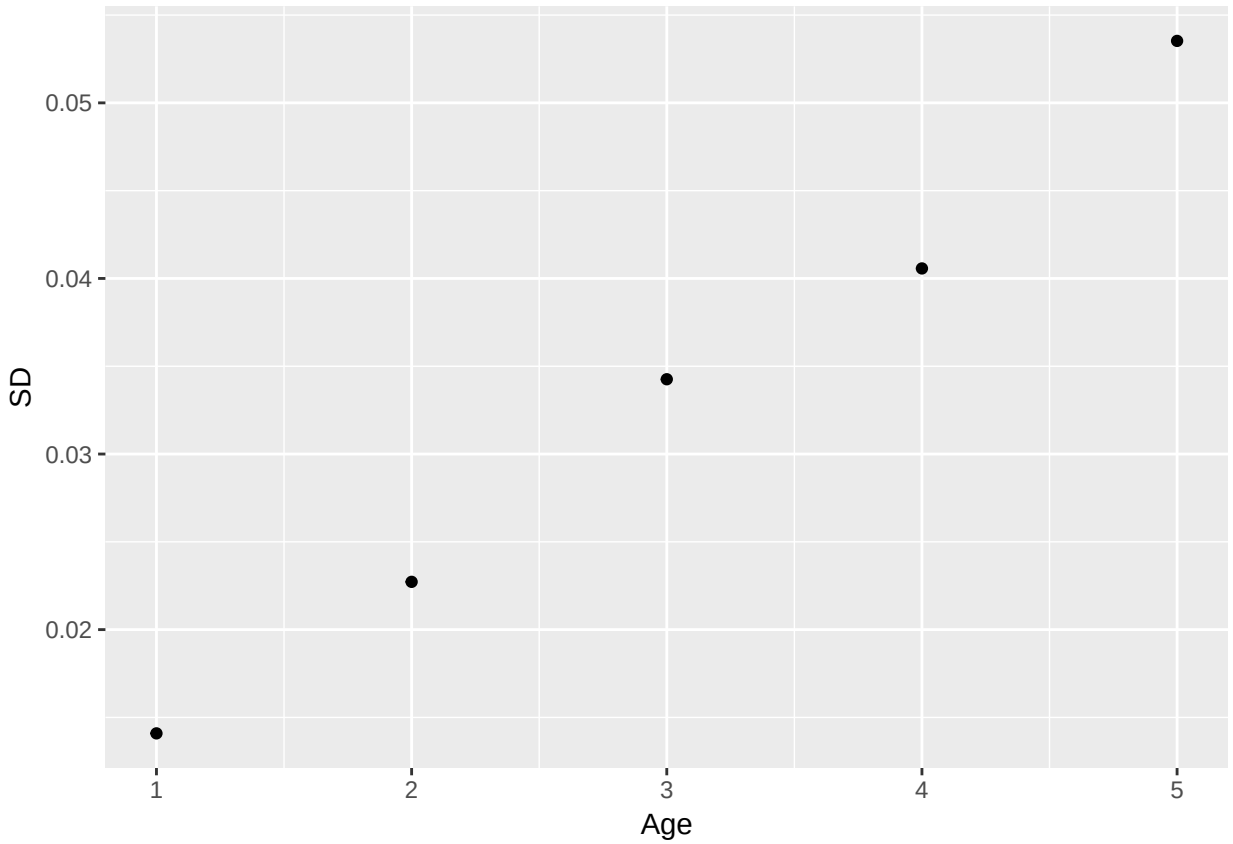
```
plot(jevons)
```




```
ggplot(jevons, aes(Age, Weight)) +
  geom_point()
```

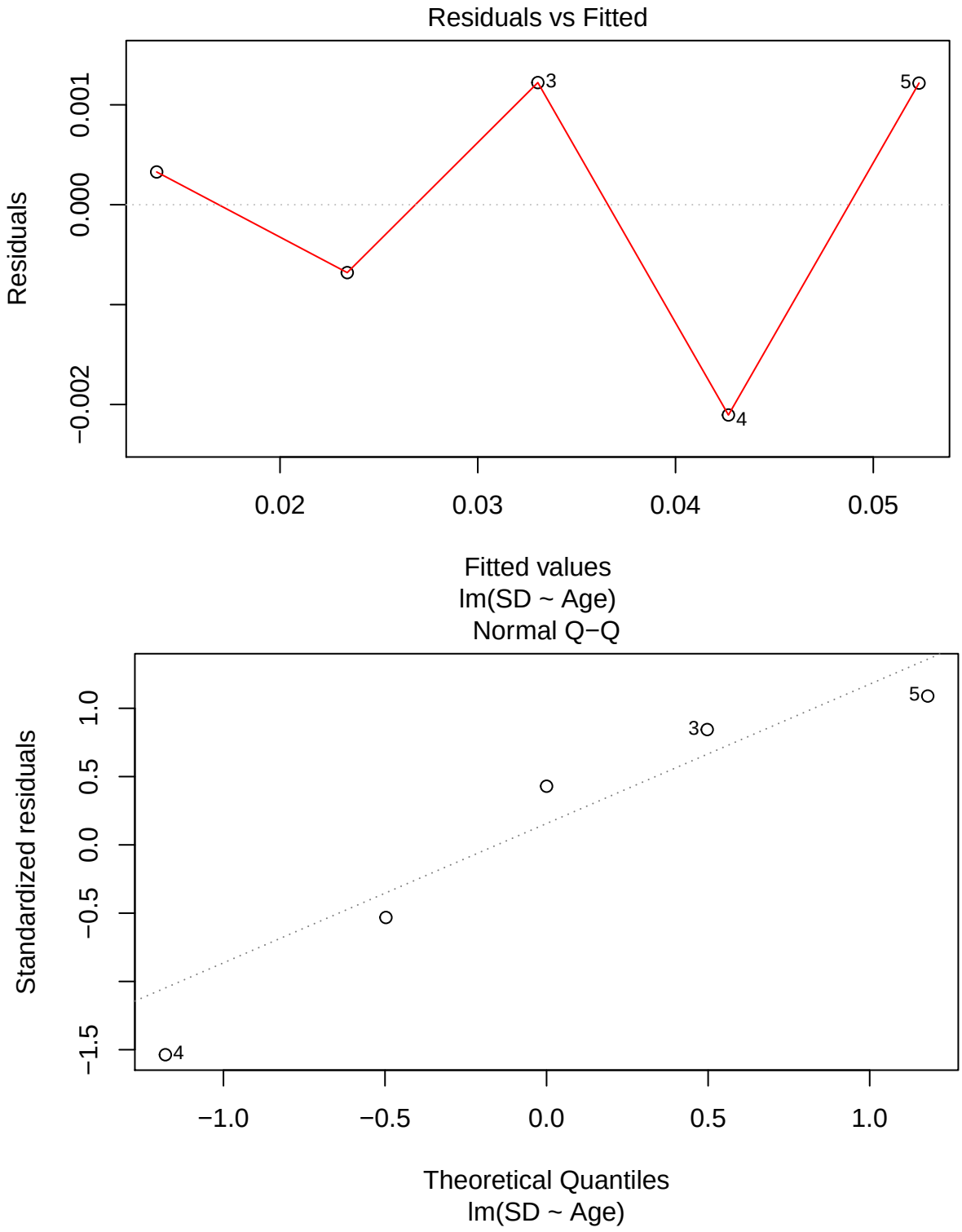


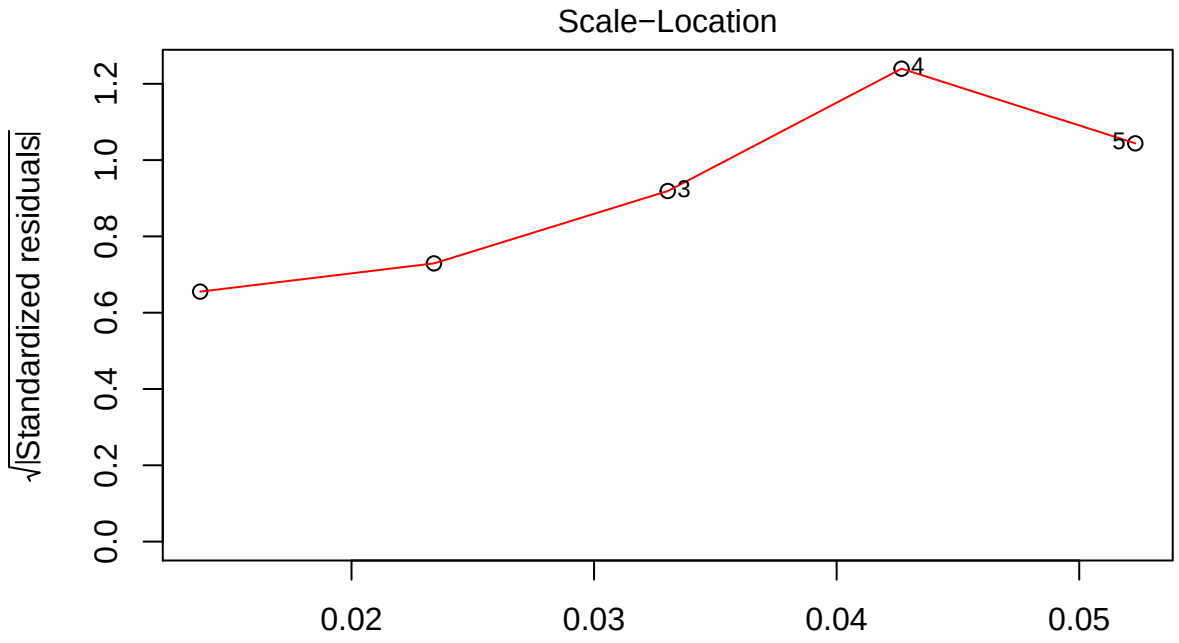

```
ggplot(jevons, aes(Age, SD)) +
  geom_point()
```

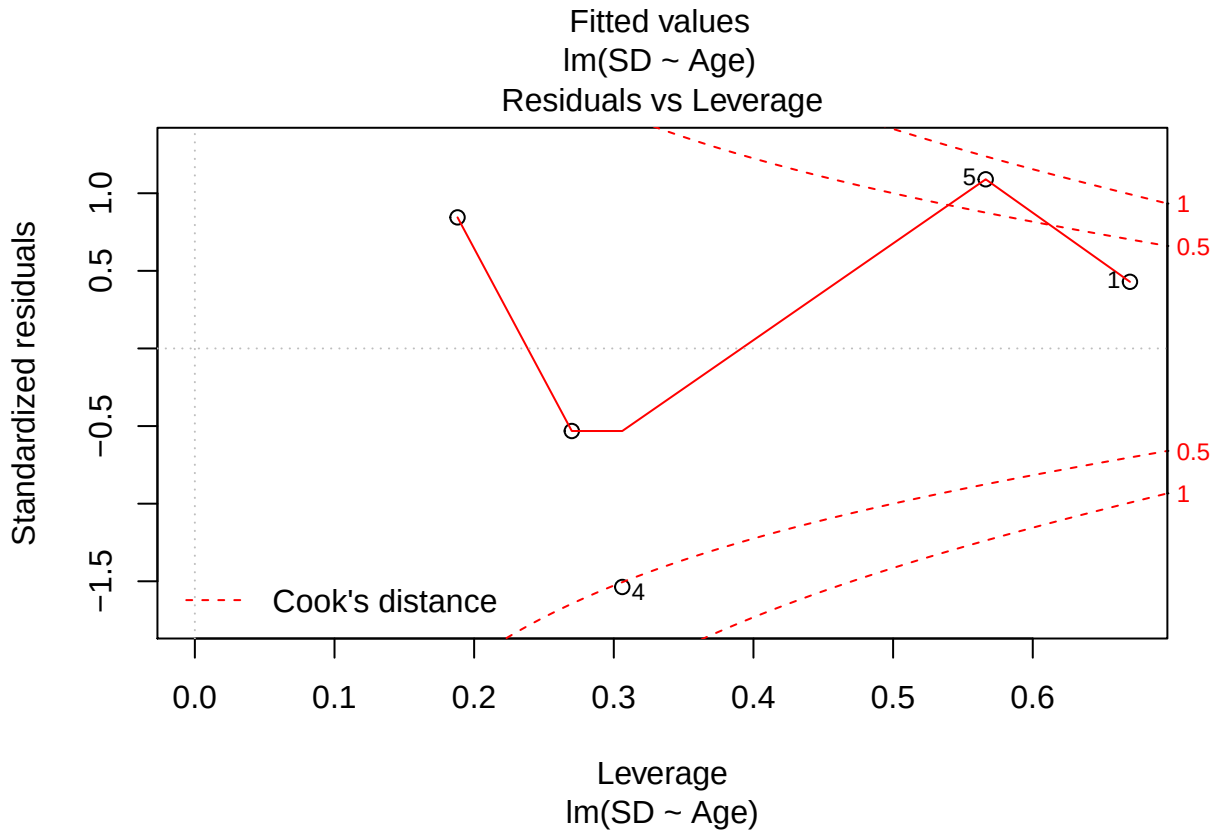



```
model.1 <- lm(SD ~ Age, data=jevons)
wts <- (nrow(jevons) * fitted(lm(abs(residuals(model.1)) ~ jevons$SD))^2) / (jevons$SD)^2
model2 <- lm(SD ~ Age, data=jevons, weights=wts)

plot(model2)
```


Residuals vs Fitted
Residuals
0.001
0.000
−0.002
3
5
4
0.02
0.03
0.04
0.05
Fitted values
lm(SD ~ Age)
Normal Q−Q
Standardized residuals
1.0
0.5
0.0
−0.5
−1.5
5
3
4
−1.0
−0.5
0.0
0.5
1.0
Theoretical Quantiles
lm(SD ~ Age)






```
summary(model2)
```

```
##
## Call:
## lm(formula = SD ~ Age, data = jevons, weights = wts)
##
```

```
## Weighted Residuals:
##          1          2          3          4          5
##  2.967e-05 -5.463e-05  9.147e-05 -1.539e-04  8.628e-05
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0041245  0.0014901   2.768 0.069687 .
## Age         0.0096376  0.0004796  20.095 0.000269 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0001202 on 3 degrees of freedom
## Multiple R-squared:  0.9926, Adjusted R-squared:  0.9902
## F-statistic: 403.8 on 1 and 3 DF,  p-value: 0.0002694
```

```
summary(jevons)
```

```
##       Age          n             Weight           SD
##  Min.   :1   Min.   : 17.0   Min.   :7.873   Min.   :0.01409
##  1st Qu.:2   1st Qu.: 24.0   1st Qu.:7.896   1st Qu.:0.02272
##  Median :3   Median : 32.0   Median :7.928   Median :0.03426
##  Mean   :3   Mean   : 54.8   Mean   :7.924   Mean   :0.03303
##  3rd Qu.:4   3rd Qu.: 78.0   3rd Qu.:7.950   3rd Qu.:0.04057
##  Max.   :5   Max.   :123.0   Max.   :7.973   Max.   :0.05353
##       Min             Max
##  Min.   :7.757   Min.   :7.961
##  1st Qu.:7.827   1st Qu.:7.965
##  Median :7.848   Median :7.984
##  Mean   :7.845   Mean   :7.980
##  3rd Qu.:7.892   3rd Qu.:7.993
##  Max.   :7.900   Max.   :7.999
```

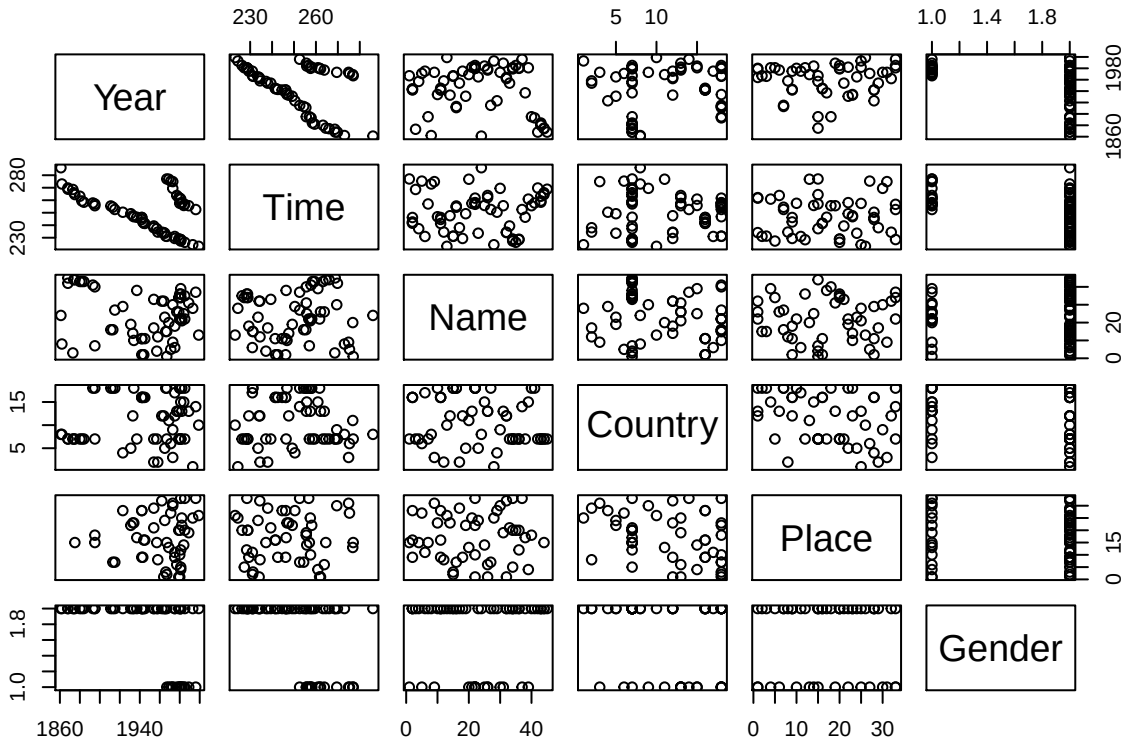
Problem 7.12

mile dataframe

```
head(mile)
```

```
##   Year  Time            Name Country Place Gender
## 1 1861 286.0     N.S. Greene     IRL  <NA>   Male
## 2 1862 273.0   George Farran     IRL  <NA>   Male
## 3 1868 269.8 Walter Chinnery     GBR  <NA>   Male
## 4 1868 268.8   William Gibbs     GBR  <NA>   Male
## 5 1873 268.6  Charles Gunton     GBR  <NA>   Male
## 6 1874 266.0    Walter Slade     GBR  <NA>   Male
```

```
plot(mile)
```

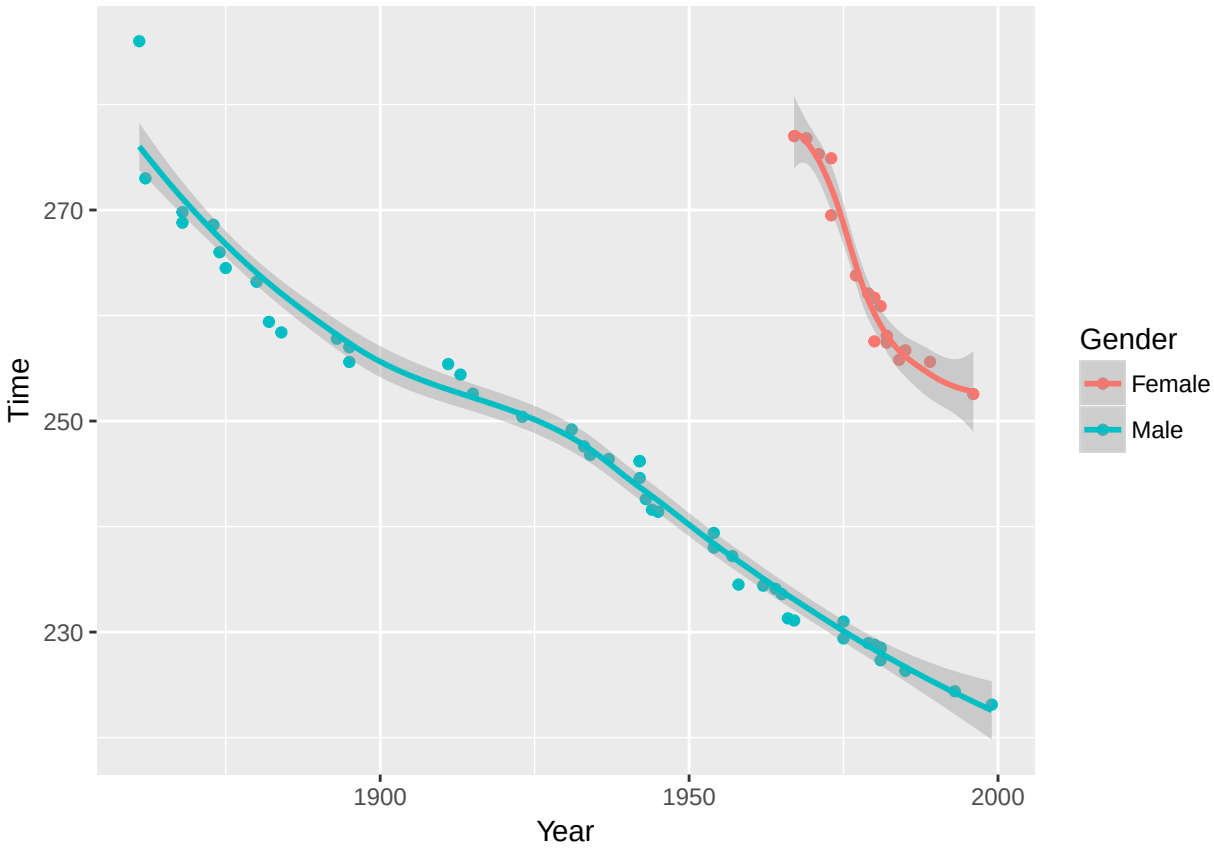



```
summary(mile)
```

```
##       Year          Time                    Name       Country
##  Min.   :1861   Min.   :223.1   Mary Decker   : 4   GBR    :19
##  1st Qu.:1917   1st Qu.:235.2   Arne Andersson: 3   USA    :12
##  Median :1960   Median :252.6   Gunder Hagg   : 3   SWE    : 6
##  Mean   :1945   Mean   :250.3   Sebastian Coe : 3   NZL    : 4
##  3rd Qu.:1980   3rd Qu.:261.5   Walter George : 3   ROM    : 4
##  Max.   :1999   Max.   :286.0   Jim Ryun      : 2   AUS    : 2
##                                 (Other)       :44   (Other):15
##        Place       Gender
##  London   : 4   Female:16
##  Auckland : 3   Male  :46
##  Goteborg : 3
##  Oslo     : 3
##  Stockholm: 3
##  (Other)  :35
##  NA's     :11
```

```
ggplot(mile, aes(Year, Time, color=Gender)) +
  geom_point() +
  geom_smooth(method=)
```

```
## `geom_smooth()` using method = 'loess'
```



Problem 8.1

baeskel dataframe

```
dim(baeskel)
```

```
## [1] 12  2
```

```
names(baeskel)
```

```
## [1] "Sulfur"  "Tension"
```

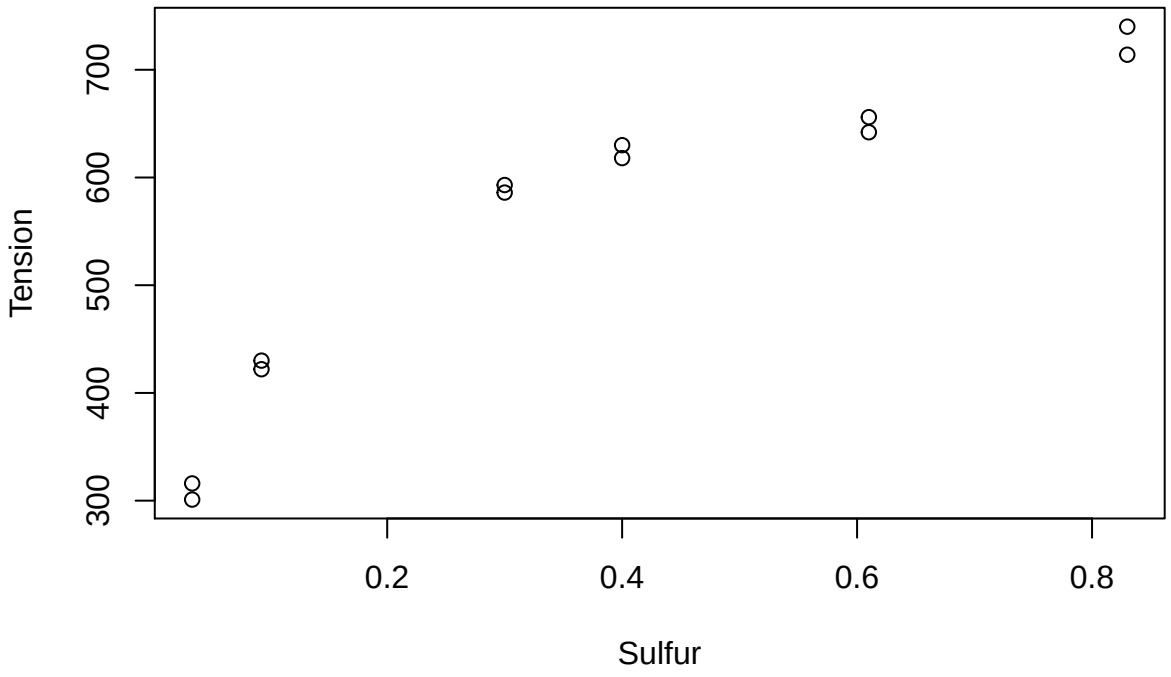
```
sapply(baeskel, class)
```

```
##    Sulfur   Tension
## "numeric" "integer"
```

```
head(baeskel)
```

```
##   Sulfur Tension
## 1  0.034     301
## 2  0.034     316
## 3  0.093     430
## 4  0.093     422
## 5  0.300     593
## 6  0.300     586
```

```
plot(baeskel)
```



Problem 8.2

stopping dataframe

```
dim(stopping)
```

```
## [1] 62  2
```

```
names(stopping)
```

```
## [1] "Speed"    "Distance"
```

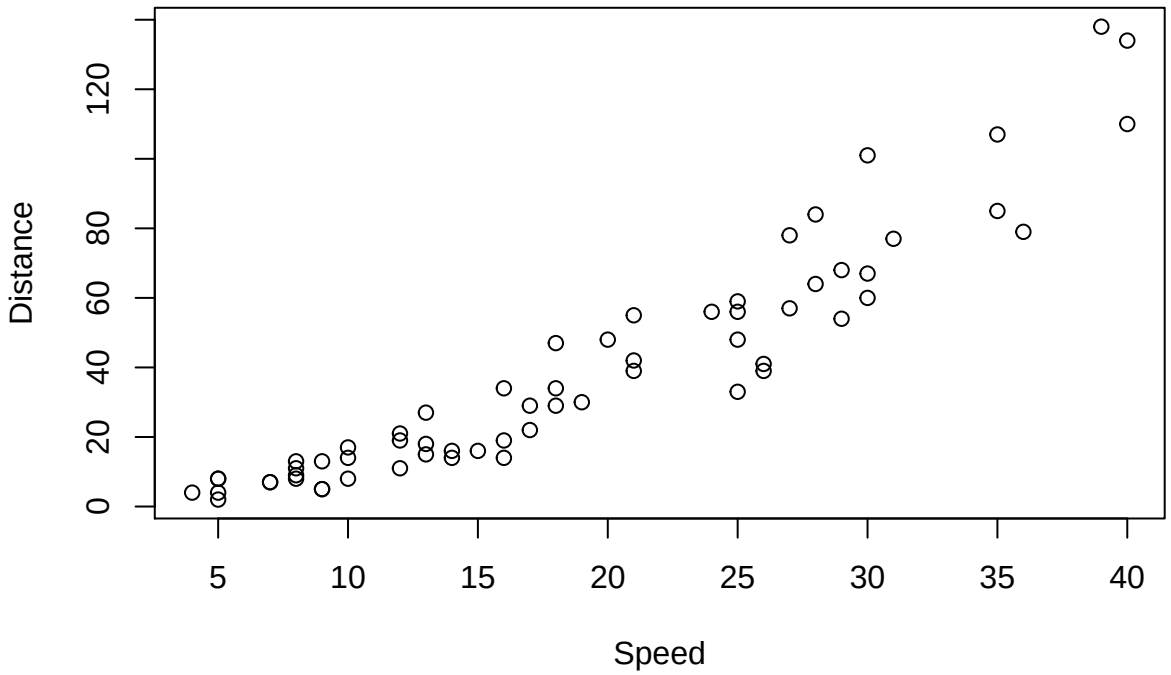
```
sapply(stopping, class)
```

```
##     Speed  Distance
## "integer" "integer"
```

```
head(stopping)
```

```
##   Speed Distance
## 1     4        4
## 2     5        2
## 3     5        4
## 4     5        8
## 5     5        8
## 6     7        7
```

```
plot(stopping)
```




```
summary(stopping)
```

```
##      Speed           Distance
##  Min.   : 4.00   Min.   :  2.00
##  1st Qu.:10.00   1st Qu.: 13.25
##  Median :17.50   Median : 29.50
##  Mean   :18.92   Mean   : 39.31
##  3rd Qu.:26.75   3rd Qu.: 56.75
##  Max.   :40.00   Max.   :138.00
```

Problem 8.3

water dataframe

```
dim(water)
```

```
## [1] 43  8
```

```
names(water)
```

```
## [1] "Year"    "APMAM"   "APSAB"   "APSLAKE" "OPBPC"   "OPRC"    "OPSLAKE"
## [8] "BSAAM"
```

```
sapply(water, class)
```

```
##      Year     APMAM     APSAB   APSLAKE     OPBPC      OPRC   OPSLAKE
## "integer" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##     BSAAM
## "integer"
```

```
head(water)
```

```
##   Year APMAM APSAB APSLAKE OPBPC  OPRC OPSLAKE  BSAAM
## 1 1948  9.13  3.58    3.91  4.10  7.43    6.47  54235
## 2 1949  5.28  4.82    5.20  7.55 11.11   10.26  67567
## 3 1950  4.20  3.77    3.67  9.52 12.20   11.35  66161
```

```
## 4 1951  4.60  4.46    3.93 11.14 15.15   11.13  68094
## 5 1952  7.15  4.99    4.88 16.34 20.05   22.81 107080
## 6 1953  9.70  5.65    4.91  8.88  8.15    7.41  67594
```

```
plot(water)
```