Final

Apurva Hari

August 23, 2016

Consider Anscombe's Quartet, a dataset included with R, designed by a mathematician to demonstrate the dangers in relying only on statistical measurements of a rstdataset

```
#reading the dataset Anscombe
data_set <- anscombe
head(data_set)
```

```
##   x1 x2 x3 x4   y1   y2    y3   y4
## 1 10 10 10  8 8.04 9.14  7.46 6.58
## 2  8  8  8  8 6.95 8.14  6.77 5.76
## 3 13 13 13  8 7.58 8.74 12.74 7.71
## 4  9  9  9  8 8.81 8.77  7.11 8.84
## 5 11 11 11  8 8.33 9.26  7.81 8.47
## 6 14 14 14  8 9.96 8.10  8.84 7.04
```

```
#reshaping the data so that each (x,y) pair are displayed together
mydata=with(anscombe,data.frame(xVal=c(x1,x2,x3,x4), yVal=c(y1,y2,y3,y4), group=gl(4,nrow(anscombe))))

#Finding out the mean and standard deviation of each (x,y) pair
aggregate(.~group,data=mydata,mean)
```

```
##   group xVal     yVal
## 1     1    9 7.500909
## 2     2    9 7.500909
## 3     3    9 7.500000
## 4     4    9 7.500909
```

```
aggregate(.~group,data=mydata,sd)
```

```
##   group     xVal     yVal
## 1     1 3.316625 2.031568
## 2     2 3.316625 2.031657
## 3     3 3.316625 2.030424
## 4     4 3.316625 2.030579
```

```
#Finding out the correlation of each quartet i.e. (x,y) pair
sapply(1:4, function(x) cor(anscombe[, x], anscombe[, x+4]))
```

```
## [1] 0.8164205 0.8162365 0.8162867 0.8165214
```

```
#Applying a linear model on each quartet
lm1 <- lm(y1 ~ x1, data = anscombe)
lm2 <- lm(y2 ~ x2, data = anscombe)
lm3 <- lm(y3 ~ x3, data = anscombe)
```

```r
lm4 <- lm(y4 ~ x4, data = anscombe)

#Finding the summary of each linear model
summary(lm1)
```

```
##
## Call:
## lm(formula = y1 ~ x1, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92127 -0.45577 -0.04136  0.70941  1.83882
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0001     1.1247   2.667  0.02573 *
## x1            0.5001     0.1179   4.241  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6665, Adjusted R-squared:  0.6295
## F-statistic: 17.99 on 1 and 9 DF,  p-value: 0.00217
```

```r
summary(lm2)
```

```
##
## Call:
## lm(formula = y2 ~ x2, data = anscombe)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9009 -0.7609  0.1291  0.9491  1.2691
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.001      1.125   2.667  0.02576 *
## x2             0.500      0.118   4.239  0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6662, Adjusted R-squared:  0.6292
## F-statistic: 17.97 on 1 and 9 DF,  p-value: 0.002179
```

```r
summary(lm3)
```

```
##
## Call:
## lm(formula = y3 ~ x3, data = anscombe)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -1.1586 -0.6146 -0.2303  0.1540  3.2411
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0025     1.1245   2.670  0.02562 *
## x3            0.4997     0.1179   4.239  0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.236 on 9 degrees of freedom
## Multiple R-squared:  0.6663, Adjusted R-squared:  0.6292
## F-statistic: 17.97 on 1 and 9 DF,  p-value: 0.002176
```

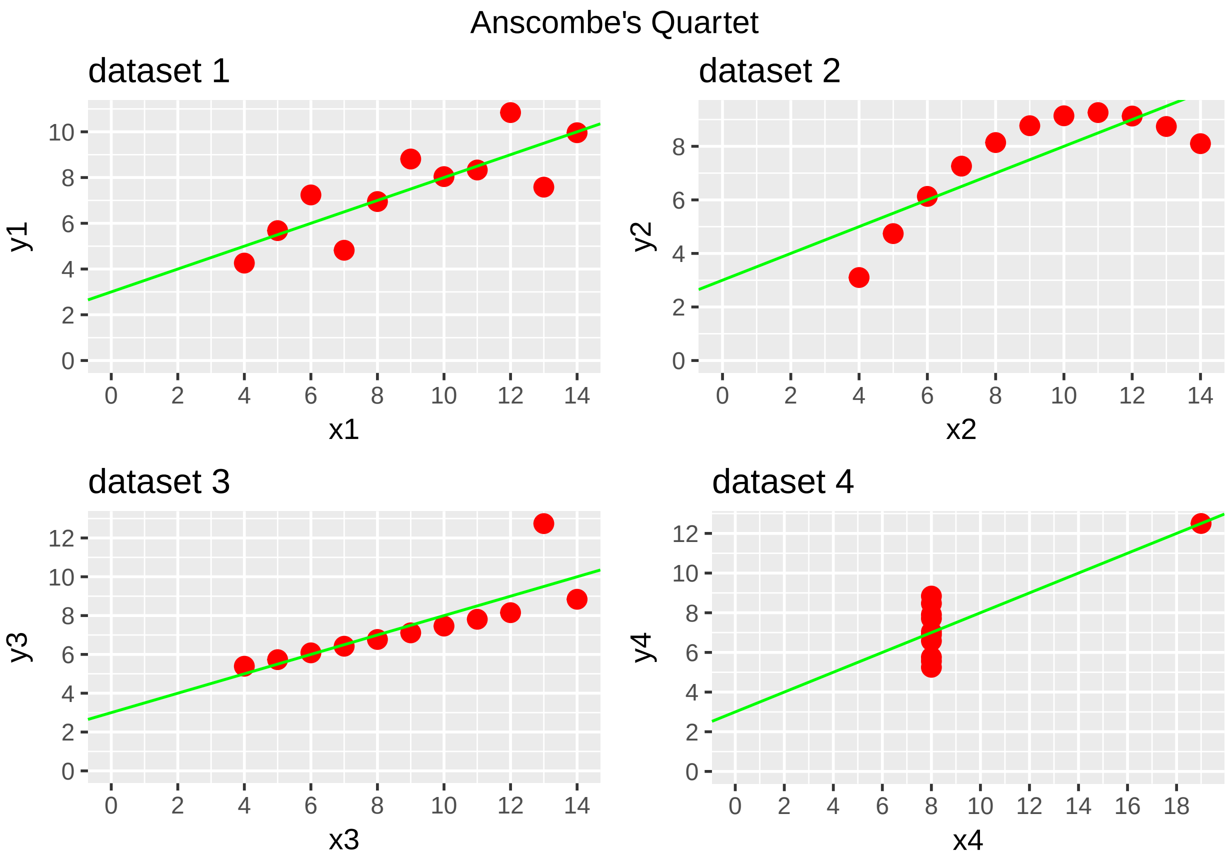
```
summary(lm4)
```

```
##
## Call:
## lm(formula = y4 ~ x4, data = anscombe)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.751 -0.831  0.000  0.809  1.839
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0017     1.1239   2.671  0.02559 *
## x4            0.4999     0.1178   4.243  0.00216 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.236 on 9 degrees of freedom
## Multiple R-squared:  0.6667, Adjusted R-squared:  0.6297
## F-statistic:    18 on 1 and 9 DF,  p-value: 0.002165
```

```
#Plotting each of the (x,y) pairs separately and fitting a line through each quartet
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```
library(gridExtra)

plt1 <- ggplot(anscombe) + geom_point(aes(x1, y1), color = "red", size = 3) + scale_x_continuous(breaks

plt2 <- ggplot(anscombe) + geom_point(aes(x2, y2), color = "red", size = 3) + scale_x_continuous(breaks

plt3 <- ggplot(anscombe) + geom_point(aes(x3, y3), color = "red", size = 3) + scale_x_continuous(breaks

plt4 <- ggplot(anscombe) + geom_point(aes(x4, y4), color = "red", size = 3) + scale_x_continuous(breaks

grid.arrange(plt1, plt2, plt3, plt4, top = "Anscombe's Quartet")
```



data set 1 is clearly linear with some scatter

data set 2 is clearly quadratic

data set 3 has an "outlier"

data set 4 -is not linear at all.

Find a dataset pertinent to your work and/or interests, which will be suitable for analysis by ONE of the following: linear regression

Response -I chose to work on the iris dataset.

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris.The species are Iris setosa, versicolor, and virginica. This is a very well-known dataset and is widely used. I chose this, as it has well established data that will help me understand how to explore data and also fit a linear model for it.

Data Exploration of the Iris dataset.Execute the chosen analysis:

Linear simple or multiple regression: fitting model of all or chosen predictors, maybe some with interactions, maybe using quadratic or log transformations, if appropriate

```
#Checking the dimensionality
dim(iris)
```

```
## [1] 150   5
```

```
#Structure of iris
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
#getting the first 5 rows
iris[1:5,]
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
```

```
#Distribution of every variable
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width   
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100  
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300  
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300  
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199  
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800  
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500  
##        Species  
##  setosa    :50  
##  versicolor:50  
##  virginica :50  
##                 
##                 
## 
```

```
#Frequency
table(iris$Species)
```

```
## 
##     setosa versicolor  virginica 
##         50         50         50
```

```r
#covariance of 2 variables
cov(iris$Petal.Length,iris$Sepal.Length)
```

```
## [1] 1.274315
```

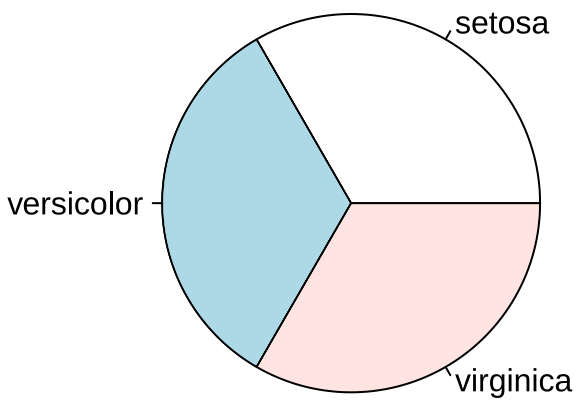
```r
#correlation between 2 variables
cor(iris$Sepal.Length,iris$Petal.Length)
```

```
## [1] 0.8717538
```

```r
#mean of all the flowers
aggregate(iris[,1:4], by=list("Species" = iris$Species), mean)
```

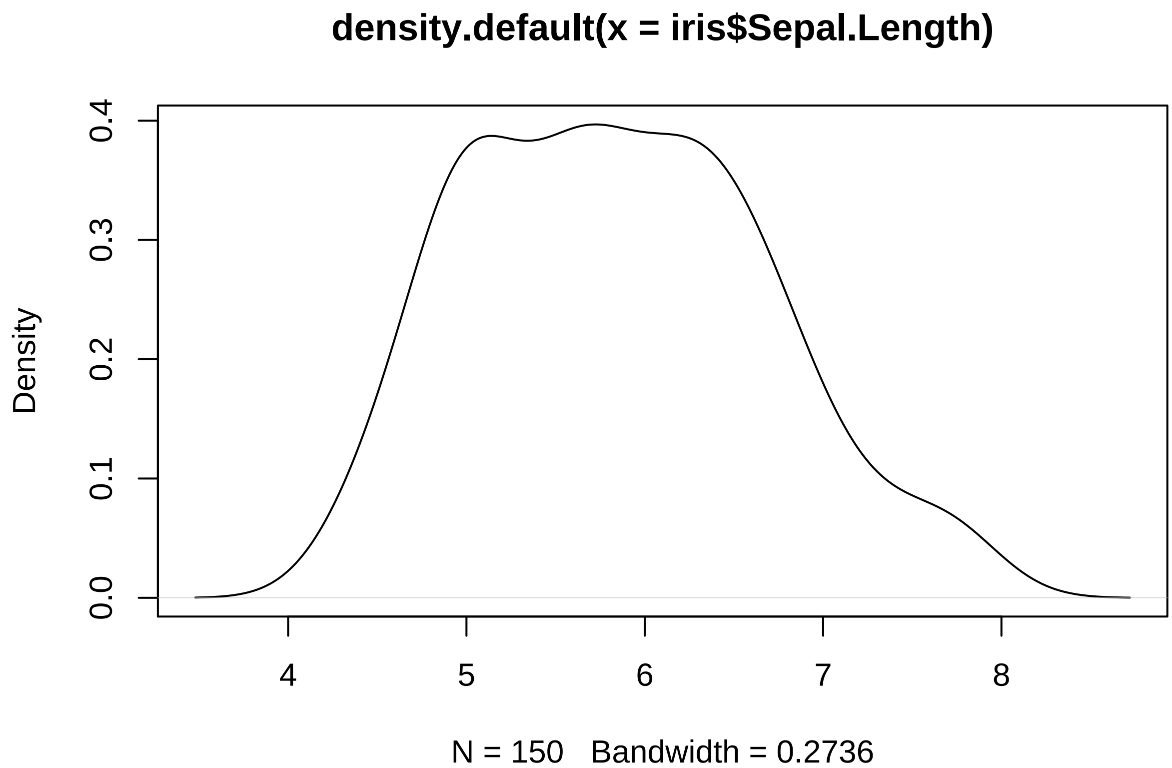
```
##      Species Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1     setosa        5.006       3.428        1.462       0.246
## 2 versicolor        5.936       2.770        4.260       1.326
## 3  virginica        6.588       2.974        5.552       2.026
```

```r
#pie chart
pie(table(iris$Species))
```

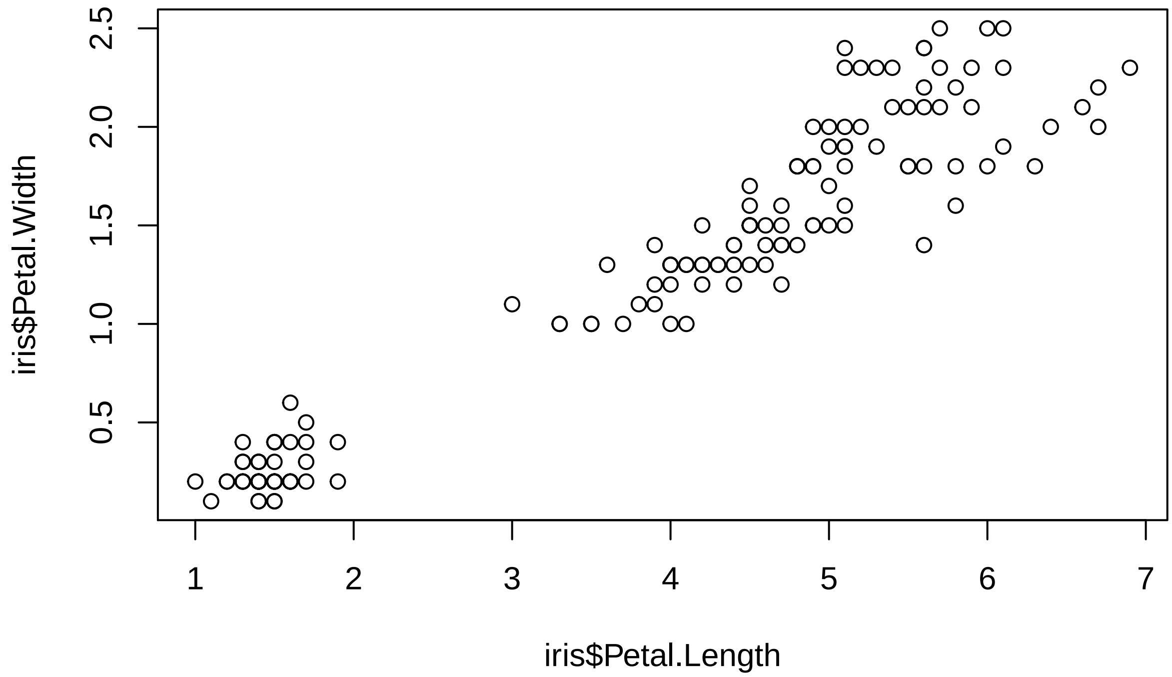


```r
#Histogram
hist(iris$Petal.Length)
```

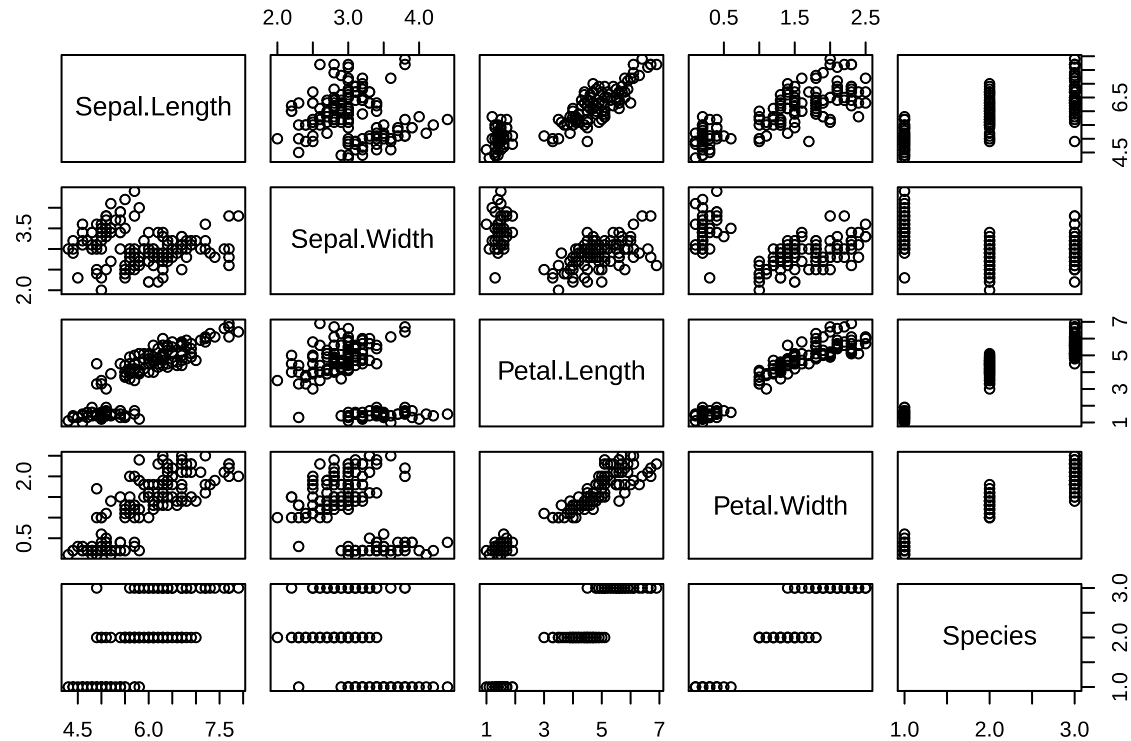

```r
#density
plot(density(iris$Sepal.Length))
```




```r
#Scatter plot
plot(iris$Petal.Length,iris$Petal.Width)
```




```
#Pair Plot
pairs(iris)
```



Execute the chosen analysis: #### Linear simple or multiple regression-fitting model of all or chosen predictors, maybe some with interactions, maybe using quadratic or log transformations, if appropriate. Evaluate your results, and possibly try some and evaluate minor refinements.

```
#Trying a linear regression to predict the petal width
fit <- lm(Petal.Width ~ Petal.Length, data=iris)
class(fit)
```

```
## [1] "lm"
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56515 -0.12358 -0.01898  0.13288  0.64272
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.363076   0.039762  -9.131  4.7e-16 ***
## Petal.Length  0.415755   0.009582  43.387  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.2065 on 148 degrees of freedom
## Multiple R-squared:  0.9271, Adjusted R-squared:  0.9266
## F-statistic:  1882 on 1 and 148 DF,  p-value: < 2.2e-16
```

```r
coefficients(fit) # model coefficients
```

```
##  (Intercept) Petal.Length
##   -0.3630755    0.4157554
```

```r
predict(fit) # fitted predictions
```

```
##          1          2          3          4          5          6
## 0.21898206 0.21898206 0.17740652 0.26055760 0.21898206 0.34370869
##          7          8          9         10         11         12
## 0.21898206 0.26055760 0.21898206 0.26055760 0.26055760 0.30213314
##         13         14         15         16         17         18
## 0.21898206 0.09425544 0.13583098 0.26055760 0.17740652 0.21898206
##         19         20         21         22         23         24
## 0.34370869 0.26055760 0.34370869 0.26055760 0.05267990 0.34370869
##         25         26         27         28         29         30
## 0.42685977 0.30213314 0.30213314 0.26055760 0.21898206 0.30213314
##         31         32         33         34         35         36
## 0.30213314 0.26055760 0.26055760 0.21898206 0.26055760 0.13583098
##         37         38         39         40         41         42
## 0.17740652 0.21898206 0.17740652 0.26055760 0.17740652 0.17740652
##         43         44         45         46         47         48
## 0.17740652 0.30213314 0.42685977 0.21898206 0.30213314 0.21898206
##         49         50         51         52         53         54
## 0.26055760 0.21898206 1.59097494 1.50782385 1.67412602 1.29994614
##         55         56         57         58         59         60
## 1.54939939 1.50782385 1.59097494 1.00891735 1.54939939 1.25837060
##         61         62         63         64         65         66
## 1.09206844 1.38309723 1.29994614 1.59097494 1.13364398 1.46624831
##         67         68         69         70         71         72
## 1.50782385 1.34152169 1.50782385 1.25837060 1.63255048 1.29994614
##         73         74         75         76         77         78
## 1.67412602 1.59097494 1.42467277 1.46624831 1.63255048 1.71570156
##         79         80         81         82         83         84
## 1.50782385 1.09206844 1.21679506 1.17521952 1.25837060 1.75727710
##         85         86         87         88         89         90
## 1.50782385 1.50782385 1.59097494 1.46624831 1.34152169 1.29994614
##         91         92         93         94         95         96
## 1.46624831 1.54939939 1.29994614 1.00891735 1.38309723 1.38309723
##         97         98         99        100        101        102
## 1.38309723 1.42467277 0.88419073 1.34152169 2.13145698 1.75727710
##        103        104        105        106        107        108
## 2.08988144 1.96515481 2.04830589 2.38091023 1.50782385 2.25618360
##        109        110        111        112        113        114
## 2.04830589 2.17303252 1.75727710 1.84042819 1.92357927 1.71570156
##        115        116        117        118        119        120
## 1.75727710 1.84042819 1.92357927 2.42248577 2.50563685 1.71570156
##        121        122        123        124        125        126
```

```
## 2.00673035 1.67412602 2.42248577 1.67412602 2.00673035 2.13145698
##        127        128        129        130        131        132
## 1.63255048 1.67412602 1.96515481 2.04830589 2.17303252 2.29775914
##        133        134        135        136        137        138
## 1.96515481 1.75727710 1.96515481 2.17303252 1.96515481 1.92357927
##        139        140        141        142        143        144
## 1.63255048 1.88200373 1.96515481 1.75727710 1.75727710 2.08988144
##        145        146        147        148        149        150
## 2.00673035 1.79885264 1.71570156 1.79885264 1.88200373 1.75727710
```

```r
predict(fit, newdata=data.frame(Petal.Length=seq(1, 2, by=0.1)))
```

```
##          1          2          3          4          5          6
## 0.05267990 0.09425544 0.13583098 0.17740652 0.21898206 0.26055760
##          7          8          9         10         11
## 0.30213314 0.34370869 0.38528423 0.42685977 0.46843531
```

```r
confint(fit, level=0.95) # Confidence Intervals for model parameters
```

```
##                   2.5 %     97.5 %
## (Intercept)  -0.4416501 -0.2845010
## Petal.Length  0.3968193  0.4346915
```

```r
fitted(fit) # predicted values
```

```
##          1          2          3          4          5          6
## 0.21898206 0.21898206 0.17740652 0.26055760 0.21898206 0.34370869
##          7          8          9         10         11         12
## 0.21898206 0.26055760 0.21898206 0.26055760 0.26055760 0.30213314
##         13         14         15         16         17         18
## 0.21898206 0.09425544 0.13583098 0.26055760 0.17740652 0.21898206
##         19         20         21         22         23         24
## 0.34370869 0.26055760 0.34370869 0.26055760 0.05267990 0.34370869
##         25         26         27         28         29         30
## 0.42685977 0.30213314 0.30213314 0.26055760 0.21898206 0.30213314
##         31         32         33         34         35         36
## 0.30213314 0.26055760 0.26055760 0.21898206 0.26055760 0.13583098
##         37         38         39         40         41         42
## 0.17740652 0.21898206 0.17740652 0.26055760 0.17740652 0.17740652
##         43         44         45         46         47         48
## 0.17740652 0.30213314 0.42685977 0.21898206 0.30213314 0.21898206
##         49         50         51         52         53         54
## 0.26055760 0.21898206 1.59097494 1.50782385 1.67412602 1.29994614
##         55         56         57         58         59         60
## 1.54939939 1.50782385 1.59097494 1.00891735 1.54939939 1.25837060
##         61         62         63         64         65         66
## 1.09206844 1.38309723 1.29994614 1.59097494 1.13364398 1.46624831
##         67         68         69         70         71         72
## 1.50782385 1.34152169 1.50782385 1.25837060 1.63255048 1.29994614
##         73         74         75         76         77         78
## 1.67412602 1.59097494 1.42467277 1.46624831 1.63255048 1.71570156
##         79         80         81         82         83         84
```

```
## 1.50782385 1.09206844 1.21679506 1.17521952 1.25837060 1.75727710
##         85         86         87         88         89         90
## 1.50782385 1.50782385 1.59097494 1.46624831 1.34152169 1.29994614
##         91         92         93         94         95         96
## 1.46624831 1.54939939 1.29994614 1.00891735 1.38309723 1.38309723
##         97         98         99        100        101        102
## 1.38309723 1.42467277 0.88419073 1.34152169 2.13145698 1.75727710
##        103        104        105        106        107        108
## 2.08988144 1.96515481 2.04830589 2.38091023 1.50782385 2.25618360
##        109        110        111        112        113        114
## 2.04830589 2.17303252 1.75727710 1.84042819 1.92357927 1.71570156
##        115        116        117        118        119        120
## 1.75727710 1.84042819 1.92357927 2.42248577 2.50563685 1.71570156
##        121        122        123        124        125        126
## 2.00673035 1.67412602 2.42248577 1.67412602 2.00673035 2.13145698
##        127        128        129        130        131        132
## 1.63255048 1.67412602 1.96515481 2.04830589 2.17303252 2.29775914
##        133        134        135        136        137        138
## 1.96515481 1.75727710 1.96515481 2.17303252 1.96515481 1.92357927
##        139        140        141        142        143        144
## 1.63255048 1.88200373 1.96515481 1.75727710 1.75727710 2.08988144
##        145        146        147        148        149        150
## 2.00673035 1.79885264 1.71570156 1.79885264 1.88200373 1.75727710
```

```r
residuals(fit) # residuals
```

```
##             1             2             3             4             5
## -1.898206e-02 -1.898206e-02  2.259348e-02 -6.055760e-02 -1.898206e-02
##             6             7             8             9            10
##  5.629131e-02  8.101794e-02 -6.055760e-02 -1.898206e-02 -1.605576e-01
##            11            12            13            14            15
## -6.055760e-02 -1.021331e-01 -1.189821e-01  5.744563e-03  6.416902e-02
##            16            17            18            19            20
##  1.394424e-01  2.225935e-01  8.101794e-02 -4.370869e-02  3.944240e-02
##            21            22            23            24            25
## -1.437087e-01  1.394424e-01  1.473201e-01  1.562913e-01 -2.268598e-01
##            26            27            28            29            30
## -1.021331e-01  9.786686e-02 -6.055760e-02 -1.898206e-02 -1.021331e-01
##            31            32            33            34            35
## -1.021331e-01  1.394424e-01 -1.605576e-01 -1.898206e-02 -6.055760e-02
##            36            37            38            39            40
##  6.416902e-02  2.259348e-02 -1.189821e-01  2.259348e-02 -6.055760e-02
##            41            42            43            44            45
##  1.225935e-01  1.225935e-01  2.259348e-02  2.978669e-01 -2.685977e-02
##            46            47            48            49            50
##  8.101794e-02 -1.021331e-01 -1.898206e-02 -6.055760e-02 -1.898206e-02
##            51            52            53            54            55
## -1.909749e-01 -7.823852e-03 -1.741260e-01  5.385591e-05 -4.939939e-02
##            56            57            58            59            60
## -2.078239e-01  9.025064e-03 -8.917353e-03 -2.493994e-01  1.416294e-01
##            61            62            63            64            65
## -9.206844e-02  1.169028e-01 -2.999461e-01 -1.909749e-01  1.663560e-01
##            66            67            68            69            70
## -6.624831e-02 -7.823852e-03 -3.415217e-01 -7.823852e-03 -1.583706e-01
```

```
##            71            72            73            74            75
##  1.674495e-01  5.385591e-05 -1.741260e-01 -3.909749e-01 -1.246728e-01
##            76            77            78            79            80
## -6.624831e-02 -2.325505e-01 -1.570156e-02 -7.823852e-03 -9.206844e-02
##            81            82            83            84            85
## -1.167951e-01 -1.752195e-01 -5.837060e-02 -1.572771e-01 -7.823852e-03
##            86            87            88            89            90
##  9.217615e-02 -9.097494e-02 -1.662483e-01 -4.152169e-02  5.385591e-05
##            91            92            93            94            95
## -2.662483e-01 -1.493994e-01 -9.994614e-02 -8.917353e-03 -8.309723e-02
##            96            97            98            99           100
## -1.830972e-01 -8.309723e-02 -1.246728e-01  2.158093e-01 -4.152169e-02
##           101           102           103           104           105
##  3.685430e-01  1.427229e-01  1.011856e-02 -1.651548e-01  1.516941e-01
##           106           107           108           109           110
## -2.809102e-01  1.921761e-01 -4.561836e-01 -2.483059e-01  3.269675e-01
##           111           112           113           114           115
##  2.427229e-01  5.957181e-02  1.764207e-01  2.842984e-01  6.427229e-01
##           116           117           118           119           120
##  4.595718e-01 -1.235793e-01 -2.224858e-01 -2.056369e-01 -2.157016e-01
##           121           122           123           124           125
##  2.932696e-01  3.258740e-01 -4.224858e-01  1.258740e-01  9.326965e-02
##           126           127           128           129           130
## -3.314570e-01  1.674495e-01  1.258740e-01  1.348452e-01 -4.483059e-01
##           131           132           133           134           135
## -2.730325e-01 -2.977591e-01  2.348452e-01 -2.572771e-01 -5.651548e-01
##           136           137           138           139           140
##  1.269675e-01  4.348452e-01 -1.235793e-01  1.674495e-01  2.179963e-01
##           141           142           143           144           145
##  4.348452e-01  5.427229e-01  1.427229e-01  2.101186e-01  4.932696e-01
##           146           147           148           149           150
##  5.011474e-01  1.842984e-01  2.011474e-01  4.179963e-01  4.272290e-02
```

```r
influence(fit) # regression diagnostics
```

```
## $hat
##           1           2           3           4           5           6
## 0.018641381 0.018641381 0.019678585 0.017647251 0.018641381 0.015788209
##           7           8           9          10          11          12
## 0.018641381 0.017647251 0.018641381 0.017647251 0.017647251 0.016696193
##          13          14          15          16          17          18
## 0.018641381 0.021882212 0.020758861 0.017647251 0.019678585 0.018641381
##          19          20          21          22          23          24
## 0.015788209 0.017647251 0.015788209 0.017647251 0.023048635 0.015788209
##          25          26          27          28          29          30
## 0.014101461 0.016696193 0.016696193 0.017647251 0.018641381 0.016696193
##          31          32          33          34          35          36
## 0.016696193 0.017647251 0.017647251 0.018641381 0.017647251 0.020758861
##          37          38          39          40          41          42
## 0.019678585 0.018641381 0.019678585 0.017647251 0.019678585 0.019678585
##          43          44          45          46          47          48
## 0.019678585 0.016696193 0.014101461 0.018641381 0.016696193 0.018641381
##          49          50          51          52          53          54
## 0.017647251 0.018641381 0.008577749 0.007852395 0.009475395 0.006792794
```

```
##          55          56          57          58          59          60 
## 0.008193536 0.007852395 0.008577749 0.007118427 0.008193536 0.006710093 
##          61          62          63          64          65          66 
## 0.006810023 0.007087415 0.006792794 0.008577749 0.006720431 0.007554329 
##          67          68          69          70          71          72 
## 0.007852395 0.006918568 0.007852395 0.006710093 0.009005035 0.006792794 
##          73          74          75          76          77          78 
## 0.009475395 0.008577749 0.007299335 0.007554329 0.009005035 0.009988828 
##          79          80          81          82          83          84 
## 0.007852395 0.006810023 0.006670466 0.006673912 0.006710093 0.010545335 
##          85          86          87          88          89          90 
## 0.007852395 0.007852395 0.008577749 0.007554329 0.006918568 0.006792794 
##          91          92          93          94          95          96 
## 0.007554329 0.008193536 0.006792794 0.007118427 0.007087415 0.007087415 
##          97          98          99         100         101         102 
## 0.007087415 0.007299335 0.007904083 0.006918568 0.017492187 0.010545335 
##         103         104         105         106         107         108 
## 0.016548021 0.013973965 0.015646929 0.024061718 0.007852395 0.020583123 
##         109         110         111         112         113         114 
## 0.015646929 0.018479426 0.010545335 0.011787567 0.013202092 0.009988828 
##         115         116         117         118         119         120 
## 0.010545335 0.011787567 0.013202092 0.025307396 0.027927972 0.009988828 
##         121         122         123         124         125         126 
## 0.014788910 0.009475395 0.025307396 0.009475395 0.014788910 0.017492187 
##         127         128         129         130         131         132 
## 0.009005035 0.009475395 0.013973965 0.015646929 0.018479426 0.021699581 
##         133         134         135         136         137         138 
## 0.013973965 0.010545335 0.013973965 0.018479426 0.013973965 0.013202092 
##         139         140         141         142         143         144 
## 0.009005035 0.012473293 0.013973965 0.010545335 0.010545335 0.016548021 
##         145         146         147         148         149         150 
## 0.014788910 0.011144914 0.009988828 0.011144914 0.012473293 0.010545335 
## 
## $coefficients
##       (Intercept)  Petal.Length
## 1   -4.980932e-04  9.822838e-05
## 2   -4.980932e-04  9.822838e-05
## 3    6.121378e-04 -1.220040e-04
## 4   -1.537543e-03  2.997800e-04
## 5   -4.980932e-04  9.822838e-05
## 6    1.333944e-03 -2.534987e-04
## 7    2.125927e-03 -4.192517e-04
## 8   -1.537543e-03  2.997800e-04
## 9   -4.980932e-04  9.822838e-05
## 10  -4.076520e-03  7.948129e-04
## 11  -1.537543e-03  2.997800e-04
## 12  -2.506564e-03  4.827341e-04
## 13  -3.122114e-03  6.157084e-04
## 14   1.654980e-04 -3.362005e-05
## 15   1.793521e-03 -3.610055e-04
## 16   3.540409e-03 -6.902857e-04
## 17   6.030850e-03 -1.201997e-03
## 18   2.125927e-03 -4.192517e-04
## 19  -1.035771e-03  1.968349e-04
```

```
## 20   1.001433e-03 -1.952528e-04
## 21  -3.405486e-03  6.471685e-04
## 22   3.540409e-03 -6.902857e-04
## 23   4.371333e-03 -8.956965e-04
## 24   3.703658e-03 -7.038323e-04
## 25  -4.994263e-03  9.207644e-04
## 26  -2.506564e-03  4.827341e-04
## 27   2.401860e-03 -4.625694e-04
## 28  -1.537543e-03  2.997800e-04
## 29  -4.980932e-04  9.822838e-05
## 30  -2.506564e-03  4.827341e-04
## 31  -2.506564e-03  4.827341e-04
## 32   3.540409e-03 -6.902857e-04
## 33  -4.076520e-03  7.948129e-04
## 34  -4.980932e-04  9.822838e-05
## 35  -1.537543e-03  2.997800e-04
## 36   1.793521e-03 -3.610055e-04
## 37   6.121378e-04 -1.220040e-04
## 38  -3.122114e-03  6.157084e-04
## 39   6.121378e-04 -1.220040e-04
## 40  -1.537543e-03  2.997800e-04
## 41   3.321494e-03 -6.620005e-04
## 42   3.321494e-03 -6.620005e-04
## 43   6.121378e-04 -1.220040e-04
## 44   7.310283e-03 -1.407873e-03
## 45  -5.913114e-04  1.090168e-04
## 46   2.125927e-03 -4.192517e-04
## 47  -2.506564e-03  4.827341e-04
## 48  -4.980932e-04  9.822838e-05
## 49  -1.537543e-03  2.997800e-04
## 50  -4.980932e-04  9.822838e-05
## 51   1.844164e-04 -3.907925e-04
## 52  -5.215003e-06 -1.260160e-05
## 53   4.528512e-04 -4.323566e-04
## 54   2.552904e-07  2.826093e-08
## 55   7.372922e-06 -9.032008e-05
## 56  -1.385254e-04 -3.347346e-04
## 57  -8.715123e-06  1.846801e-05
## 58  -9.316711e-05  8.858935e-06
## 59   3.722318e-05 -4.559929e-04
## 60   7.867042e-04  4.360570e-05
## 61  -8.115657e-04  5.150812e-05
## 62   3.637323e-04  1.120763e-04
## 63  -1.421819e-03 -1.573970e-04
## 64   1.844164e-04 -3.907925e-04
## 65   1.330714e-03 -5.699040e-05
## 66  -9.817062e-05 -9.229553e-05
## 67  -5.215003e-06 -1.260160e-05
## 68  -1.340768e-03 -2.533011e-04
## 69  -5.215003e-06 -1.260160e-05
## 70  -8.796960e-04 -4.876008e-05
## 71  -2.985248e-04  3.791908e-04
## 72   2.552904e-07  2.826093e-08
## 73   4.528512e-04 -4.323566e-04
```

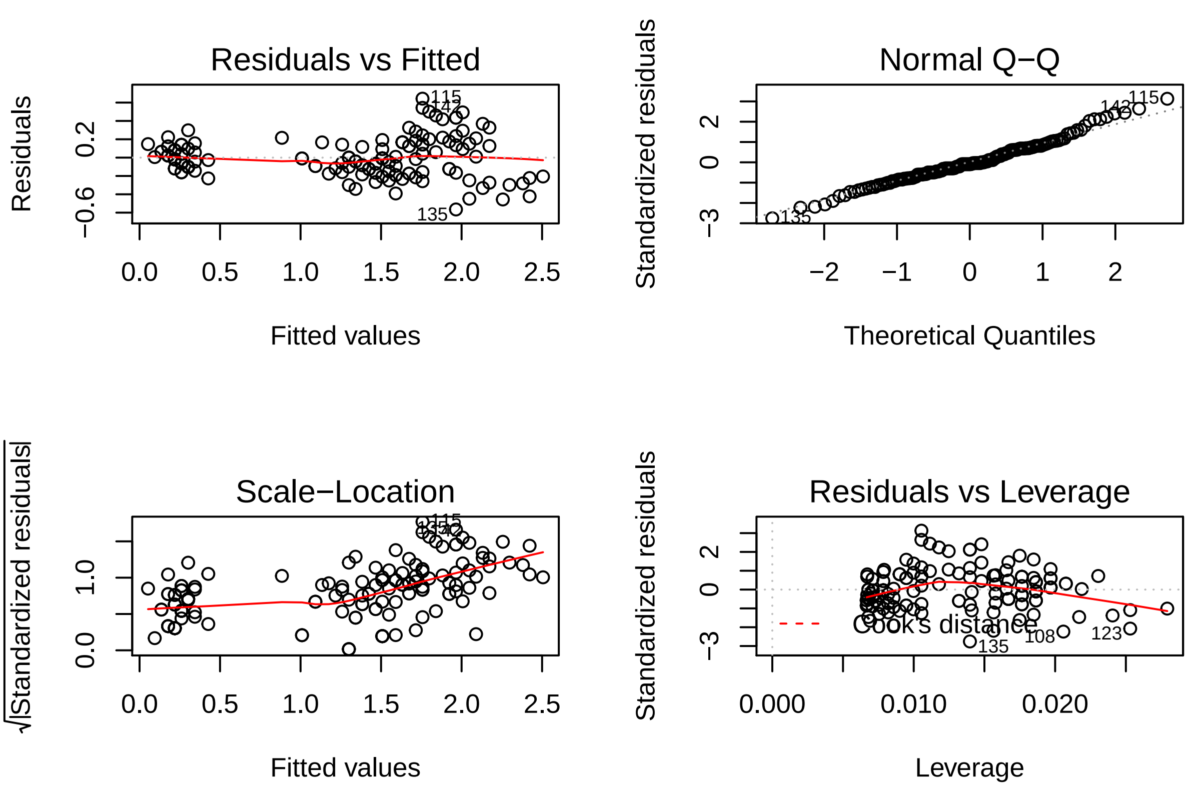
```
## 74   3.775480e-04 -8.000529e-04
## 75  -2.863453e-04 -1.465987e-04
## 76  -9.817062e-05 -9.229553e-05
## 77   4.145851e-04 -5.266124e-04
## 78   5.369260e-05 -4.242305e-05
## 79  -5.215003e-06 -1.260160e-05
## 80  -8.115657e-04  5.150812e-05
## 81  -7.438943e-04 -1.063550e-05
## 82  -1.258783e-03  2.203414e-05
## 83  -3.242293e-04 -1.797149e-05
## 84   6.667709e-04 -4.594092e-04
## 85  -5.215003e-06 -1.260160e-05
## 86   6.144018e-05  1.484649e-04
## 87   8.785065e-05 -1.861622e-04
## 88  -2.463565e-04 -2.316131e-04
## 89  -1.630085e-04 -3.079596e-05
## 90   2.552904e-07  2.826093e-08
## 91  -3.945423e-04 -3.709306e-04
## 92   2.229805e-05 -2.731565e-04
## 93  -4.737695e-04 -5.244681e-05
## 94  -9.316711e-05  8.858935e-06
## 95  -2.585494e-04 -7.966643e-05
## 96  -5.696902e-04 -1.755378e-04
## 97  -2.585494e-04 -7.966643e-05
## 98  -2.863453e-04 -1.465987e-04
## 99   2.784695e-03 -3.551102e-04
## 100 -1.630085e-04 -3.079596e-05
## 101 -4.305777e-03  1.811196e-03
## 102 -6.050689e-04  4.168961e-04
## 103 -1.097769e-04  4.746383e-05
## 104  1.380412e-03 -6.644618e-04
## 105 -1.519507e-03  6.777213e-04
## 106  4.701788e-03 -1.761760e-03
## 107  1.280954e-04  3.095314e-04
## 108  6.477432e-03 -2.549912e-03
## 109  2.487259e-03 -1.109352e-03
## 110 -4.093494e-03  1.680233e-03
## 111 -1.029015e-03  7.089979e-04
## 112 -3.504488e-04  2.001946e-04
## 113 -1.328726e-03  6.707290e-04
## 114 -9.721786e-04  7.681279e-04
## 115 -2.724802e-03  1.877405e-03
## 116 -2.703567e-03  1.544419e-03
## 117  9.307468e-04 -4.698326e-04
## 118  3.913400e-03 -1.446288e-03
## 119  3.969215e-03 -1.431483e-03
## 120  7.376067e-04 -5.827903e-04
## 121 -2.694180e-03  1.244986e-03
## 122 -8.475036e-04  8.091482e-04
## 123  7.431289e-03 -2.746405e-03
## 124 -3.273617e-04  3.125463e-04
## 125 -8.568401e-04  3.959477e-04
## 126  3.872492e-03 -1.628937e-03
## 127 -2.985248e-04  3.791908e-04
```

```
## 128 -3.273617e-04  3.125463e-04
## 129 -1.127075e-03  5.425181e-04
## 130  4.490643e-03 -2.002889e-03
## 131  3.418251e-03 -1.403070e-03
## 132  4.479095e-03 -1.731822e-03
## 133 -1.962904e-03  9.448448e-04
## 134  1.090718e-03 -7.515109e-04
## 135  4.723727e-03 -2.273768e-03
## 136 -1.589579e-03  6.524655e-04
## 137 -3.634562e-03  1.749498e-03
## 138  9.307468e-04 -4.698326e-04
## 139 -2.985248e-04  3.791908e-04
## 140 -1.461981e-03  7.806402e-04
## 141 -3.634562e-03  1.749498e-03
## 142 -2.300855e-03  1.585303e-03
## 143 -6.050689e-04  4.168961e-04
## 144 -2.279589e-03  9.856172e-04
## 145 -4.531519e-03  2.094025e-03
## 146 -2.536058e-03  1.573894e-03
## 147 -6.302215e-04  4.979442e-04
## 148 -1.017907e-03  6.317198e-04
## 149 -2.803271e-03  1.496836e-03
## 150 -1.811223e-04  1.247943e-04
##
## $sigma
##         1         2         3         4         5         6         7
## 0.2071795 0.2071795 0.2071769 0.2071242 0.2071795 0.2071326 0.2070757
##         8         9        10        11        12        13        14
## 0.2071242 0.2071795 0.2067542 0.2071242 0.2070113 0.2069485 0.2071849
##        15        16        17        18        19        20        21
## 0.2071164 0.2068603 0.2063541 0.2070757 0.2071536 0.2071595 0.2068407
##        22        23        24        25        26        27        28
## 0.2068603 0.2068205 0.2067776 0.2063267 0.2070113 0.2070255 0.2071242
##        29        30        31        32        33        34        35
## 0.2071795 0.2070113 0.2070113 0.2068603 0.2067542 0.2071795 0.2071242
##        36        37        38        39        40        41        42
## 0.2071164 0.2071769 0.2069485 0.2071769 0.2071242 0.2069336 0.2069336
##        43        44        45        46        47        48        49
## 0.2071769 0.2056988 0.2071735 0.2070757 0.2070113 0.2071795 0.2071242
##        50        51        52        53        54        55        56
## 0.2071795 0.2065807 0.2071845 0.2066824 0.2071855 0.2071451 0.2064696
##        57        58        59        60        61        62        63
## 0.2071841 0.2071842 0.2061533 0.2068537 0.2070453 0.2069594 0.2056930
##        64        65        66        67        68        69        70
## 0.2065807 0.2067276 0.2071129 0.2071845 0.2052483 0.2071845 0.2067705
##        71        72        73        74        75        76        77
## 0.2067205 0.2071855 0.2066824 0.2046386 0.2069283 0.2071129 0.2062876
##        78        79        80        81        82        83        84
## 0.2071814 0.2071845 0.2070453 0.2069599 0.2066774 0.2071292 0.2067747
##        85        86        87        88        89        90        91
## 0.2071845 0.2070448 0.2070484 0.2067278 0.2071570 0.2071855 0.2060095
##        92        93        94        95        96        97        98
## 0.2068157 0.2070203 0.2071842 0.2070713 0.2066304 0.2070713 0.2069283
##        99       100       101       102       103       104       105
```

```
## 0.2064134 0.2071570 0.2049034 0.2068472 0.2071838 0.2067308 0.2068014
##       106       107       108       109       110       111       112
## 0.2058538 0.2065735 0.2036674 0.2061546 0.2053896 0.2062057 0.2071265
##       113       114       115       116       117       118       119
## 0.2066670 0.2058408 0.2002142 0.2036465 0.2069313 0.2063501 0.2064701
##       120       121       122       123       124       125       126
## 0.2064125 0.2057473 0.2054179 0.2041569 0.2069227 0.2070405 0.2053415
##       127       128       129       130       131       132       133
## 0.2067205 0.2069227 0.2068825 0.2038060 0.2059348 0.2056923 0.2062652
##       134       135       136       137       138       139       140
## 0.2060843 0.2017975 0.2069157 0.2040129 0.2069313 0.2067205 0.2063939
##       141       142       143       144       145       146       147
## 0.2040129 0.2022393 0.2068472 0.2064472 0.2030906 0.2029731 0.2066215
##       148       149       150
## 0.2065127 0.2042602 0.2071552
##
## $wt.res
##             1             2             3             4             5
## -1.898206e-02 -1.898206e-02  2.259348e-02 -6.055760e-02 -1.898206e-02
##             6             7             8             9            10
##  5.629131e-02  8.101794e-02 -6.055760e-02 -1.898206e-02 -1.605576e-01
##            11            12            13            14            15
## -6.055760e-02 -1.021331e-01 -1.189821e-01  5.744563e-03  6.416902e-02
##            16            17            18            19            20
##  1.394424e-01  2.225935e-01  8.101794e-02 -4.370869e-02  3.944240e-02
##            21            22            23            24            25
## -1.437087e-01  1.394424e-01  1.473201e-01  1.562913e-01 -2.268598e-01
##            26            27            28            29            30
## -1.021331e-01  9.786686e-02 -6.055760e-02 -1.898206e-02 -1.021331e-01
##            31            32            33            34            35
## -1.021331e-01  1.394424e-01 -1.605576e-01 -1.898206e-02 -6.055760e-02
##            36            37            38            39            40
##  6.416902e-02  2.259348e-02 -1.189821e-01  2.259348e-02 -6.055760e-02
##            41            42            43            44            45
##  1.225935e-01  1.225935e-01  2.259348e-02  2.978669e-01 -2.685977e-02
##            46            47            48            49            50
##  8.101794e-02 -1.021331e-01 -1.898206e-02 -6.055760e-02 -1.898206e-02
##            51            52            53            54            55
## -1.909749e-01 -7.823852e-03 -1.741260e-01  5.385591e-05 -4.939939e-02
##            56            57            58            59            60
## -2.078239e-01  9.025064e-03 -8.917353e-03 -2.493994e-01  1.416294e-01
##            61            62            63            64            65
## -9.206844e-02  1.169028e-01 -2.999461e-01 -1.909749e-01  1.663560e-01
##            66            67            68            69            70
## -6.624831e-02 -7.823852e-03 -3.415217e-01 -7.823852e-03 -1.583706e-01
##            71            72            73            74            75
##  1.674495e-01  5.385591e-05 -1.741260e-01 -3.909749e-01 -1.246728e-01
##            76            77            78            79            80
## -6.624831e-02 -2.325505e-01 -1.570156e-02 -7.823852e-03 -9.206844e-02
##            81            82            83            84            85
## -1.167951e-01 -1.752195e-01 -5.837060e-02 -1.572771e-01 -7.823852e-03
##            86            87            88            89            90
##  9.217615e-02 -9.097494e-02 -1.662483e-01 -4.152169e-02  5.385591e-05
##            91            92            93            94            95
```

```
## -2.662483e-01 -1.493994e-01 -9.994614e-02 -8.917353e-03 -8.309723e-02
##            96            97            98            99           100
## -1.830972e-01 -8.309723e-02 -1.246728e-01  2.158093e-01 -4.152169e-02
##           101           102           103           104           105
##  3.685430e-01  1.427229e-01  1.011856e-02 -1.651548e-01  1.516941e-01
##           106           107           108           109           110
## -2.809102e-01  1.921761e-01 -4.561836e-01 -2.483059e-01  3.269675e-01
##           111           112           113           114           115
##  2.427229e-01  5.957181e-02  1.764207e-01  2.842984e-01  6.427229e-01
##           116           117           118           119           120
##  4.595718e-01 -1.235793e-01 -2.224858e-01 -2.056369e-01 -2.157016e-01
##           121           122           123           124           125
##  2.932696e-01  3.258740e-01 -4.224858e-01  1.258740e-01  9.326965e-02
##           126           127           128           129           130
## -3.314570e-01  1.674495e-01  1.258740e-01  1.348452e-01 -4.483059e-01
##           131           132           133           134           135
## -2.730325e-01 -2.977591e-01  2.348452e-01 -2.572771e-01 -5.651548e-01
##           136           137           138           139           140
##  1.269675e-01  4.348452e-01 -1.235793e-01  1.674495e-01  2.179963e-01
##           141           142           143           144           145
##  4.348452e-01  5.427229e-01  1.427229e-01  2.101186e-01  4.932696e-01
##           146           147           148           149           150
##  5.011474e-01  1.842984e-01  2.011474e-01  4.179963e-01  4.272290e-02
```

```r
par(mfrow=c(2,2))
plot(fit)
```



```r
#Trying a Multiple Linear Regression
fit2 <- lm(Petal.Width ~ Petal.Length + Sepal.Length + Sepal.Width, data=iris)
summary(fit2) # show results
```

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length + Sepal.Length + Sepal.Width,
##     data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60959 -0.10134 -0.01089  0.09825  0.60685
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.24031    0.17837  -1.347     0.18
## Petal.Length  0.52408    0.02449  21.399  < 2e-16 ***
## Sepal.Length -0.20727    0.04751  -4.363 2.41e-05 ***
## Sepal.Width   0.22283    0.04894   4.553 1.10e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.192 on 146 degrees of freedom
## Multiple R-squared:  0.9379, Adjusted R-squared:  0.9366
## F-statistic: 734.4 on 3 and 146 DF,  p-value: < 2.2e-16
```

```r
#Finding the Interaction Terms and do a linear fit

fit2int <- lm(Petal.Width ~ Petal.Length + Sepal.Length + Sepal.Width + Petal.Length:Sepal.Length, data=

#Analyzing covariance using linear regression
fit3 <- lm(Petal.Width ~ Petal.Length + Sepal.Length + Sepal.Width + Species, data=iris)
summary(fit3)
```

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length + Sepal.Length + Sepal.Width +
##     Species, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59239 -0.08288 -0.01349  0.08773  0.45239
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.47314    0.17659  -2.679  0.00824 **
## Petal.Length       0.24220    0.04884   4.959 1.97e-06 ***
## Sepal.Length      -0.09293    0.04458  -2.084  0.03889 *
## Sepal.Width        0.24220    0.04776   5.072 1.20e-06 ***
## Speciesversicolor  0.64811    0.12314   5.263 5.04e-07 ***
## Speciesvirginica   1.04637    0.16548   6.323 3.03e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1666 on 144 degrees of freedom
## Multiple R-squared:  0.9538, Adjusted R-squared:  0.9522
## F-statistic: 594.9 on 5 and 144 DF,  p-value: < 2.2e-16
```

Evaluate your results, and possibly try some and evaluate minor refinements. Use graphs of the model (Kabacoff, Sec. 8.3.2), assess basic assumptions like , normality, independence, linearity, homoscedasticity; gvlma() in package "gvlma"; look for and possibly deal with outliers, high-leverage points, influential observations.

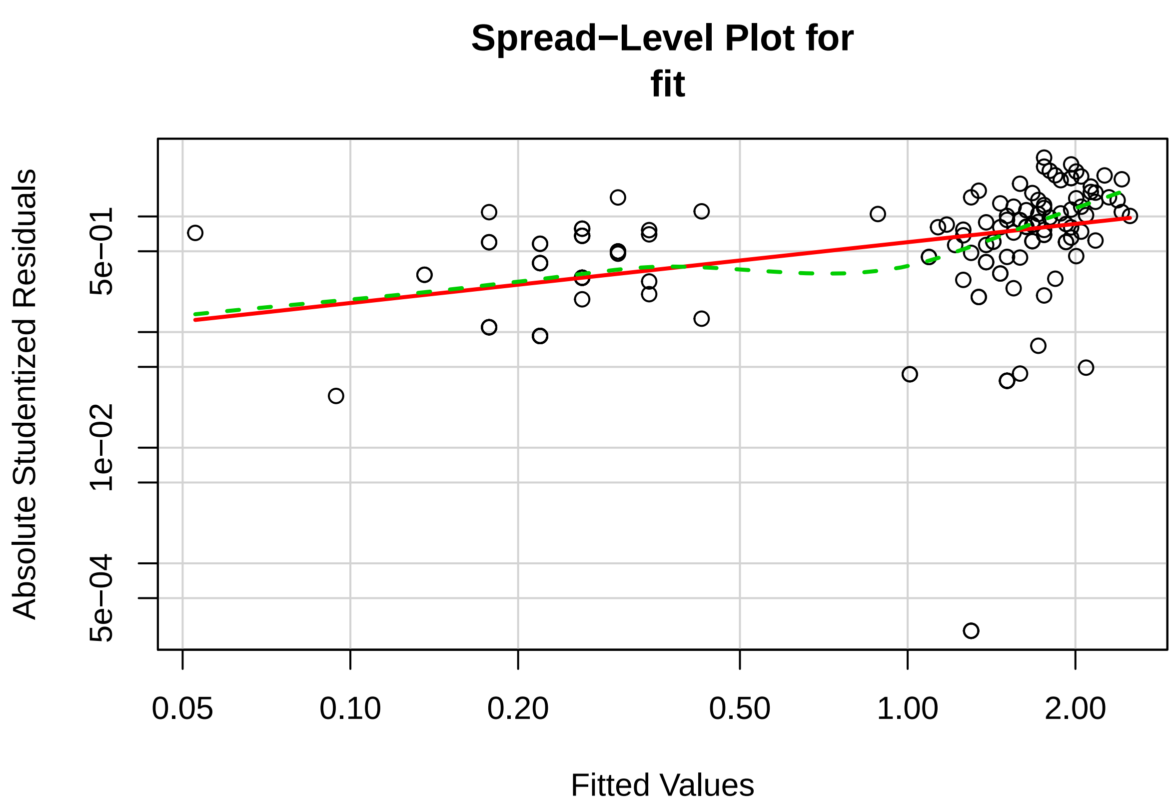
```r
#Evaluating regression using plots
library(car)
```

```
## Warning: package 'car' was built under R version 3.3.3
```

```r
#Checking the Homoscedasticity of the fitted model
ncvTest(fit)
```

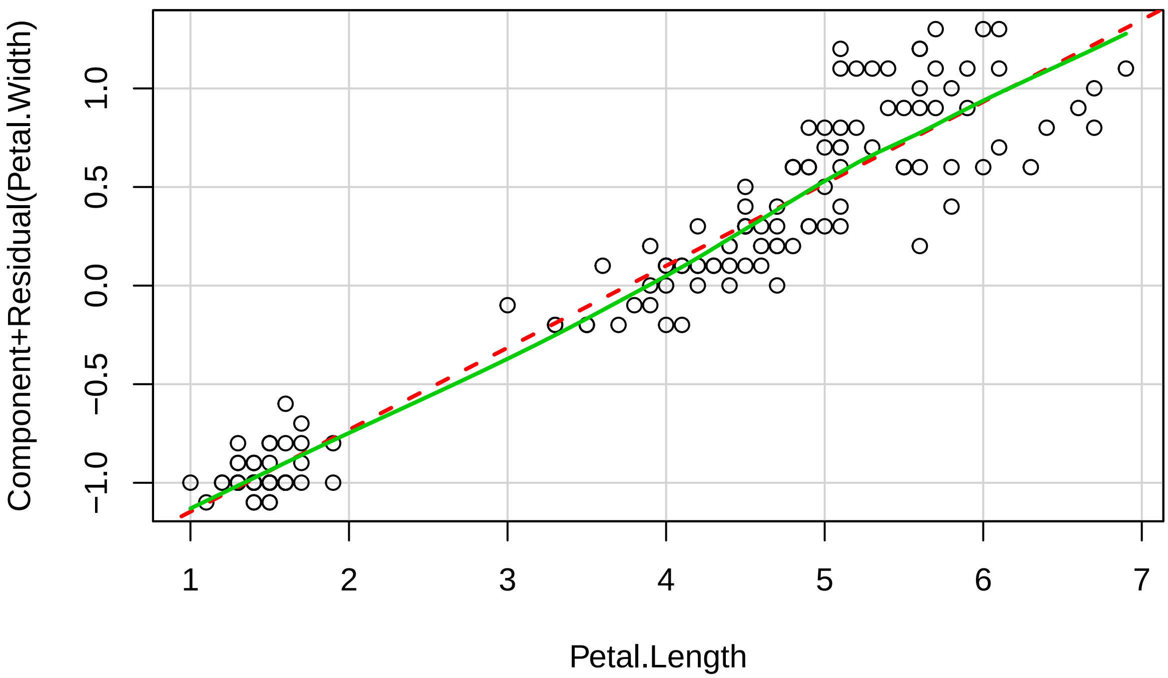
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 37.74426    Df = 1     p = 8.065366e-10
```

```r
spreadLevelPlot(fit) #creating a scatter-plot
```

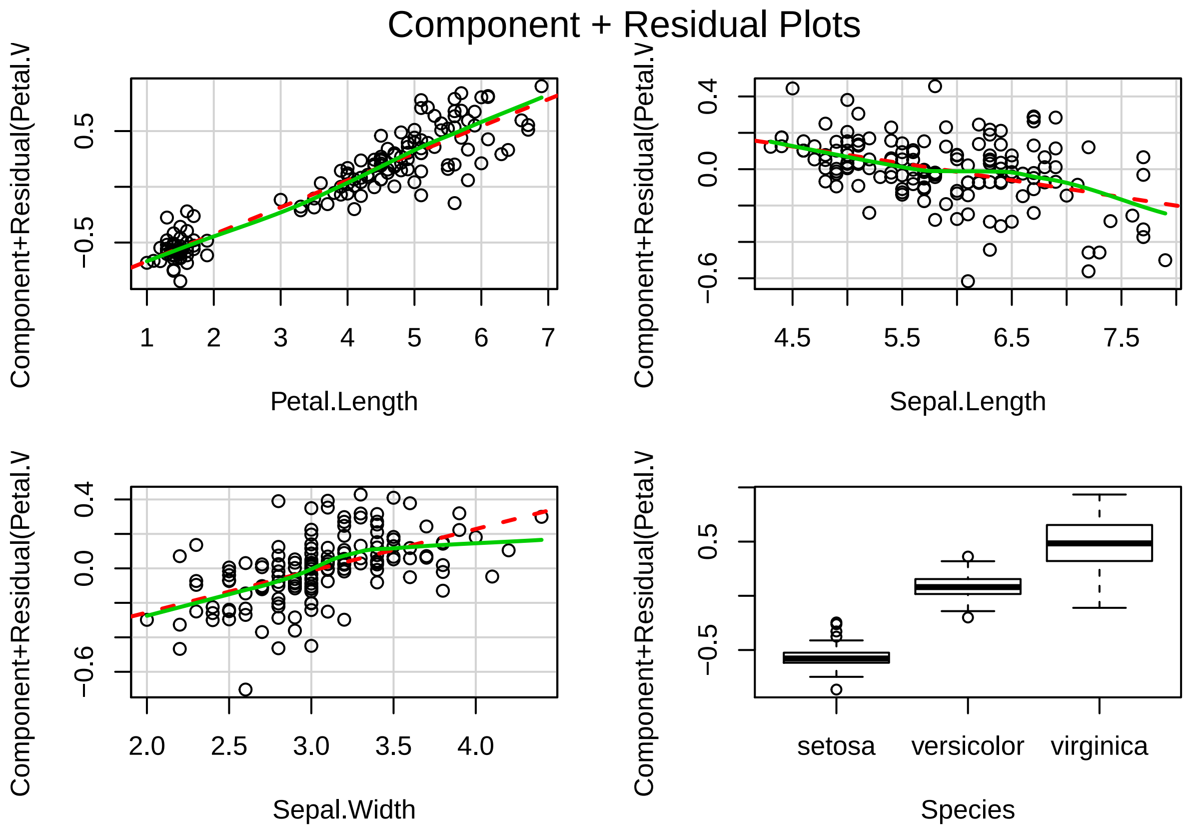


```
##
## Suggested power transformation:  0.4739141
```

```r
#Using crplots to check for linearity
crPlots(fit)
```




```r
crPlots(fit3)
```




```r
#Global validation of the linear model assumption
library(gvlma)

gvlmodel <- gvlma(fit)
summary(gvlmodel)
```

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56515 -0.12358 -0.01898  0.13288  0.64272
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.363076   0.039762  -9.131  4.7e-16 ***
## Petal.Length  0.415755   0.009582  43.387  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2065 on 148 degrees of freedom
## Multiple R-squared:  0.9271, Adjusted R-squared:  0.9266
## F-statistic:  1882 on 1 and 148 DF,  p-value: < 2.2e-16
##
##
```

```
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                       Value   p-value                   Decision
## Global Stat        50.9304 2.308e-10 Assumptions NOT satisfied!
## Skewness            3.2230 7.261e-02    Assumptions acceptable.
## Kurtosis            2.3321 1.267e-01    Assumptions acceptable.
## Link Function       0.1196 7.295e-01    Assumptions acceptable.
## Heteroscedasticity 45.2557 1.729e-11 Assumptions NOT satisfied!
```

```
gvlmodel2 <- gvlma(fit3)
summary(gvlmodel2)
```

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length + Sepal.Length + Sepal.Width +
##     Species, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59239 -0.08288 -0.01349  0.08773  0.45239
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.47314    0.17659  -2.679  0.00824 **
## Petal.Length       0.24220    0.04884   4.959 1.97e-06 ***
## Sepal.Length      -0.09293    0.04458  -2.084  0.03889 *
## Sepal.Width        0.24220    0.04776   5.072 1.20e-06 ***
## Speciesversicolor  0.64811    0.12314   5.263 5.04e-07 ***
## Speciesvirginica   1.04637    0.16548   6.323 3.03e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1666 on 144 degrees of freedom
## Multiple R-squared:  0.9538, Adjusted R-squared:  0.9522
## F-statistic: 594.9 on 5 and 144 DF,  p-value: < 2.2e-16
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit3)
##
##                       Value   p-value                   Decision
## Global Stat        46.7768 1.697e-09 Assumptions NOT satisfied!
## Skewness            0.7055 4.009e-01    Assumptions acceptable.
## Kurtosis            7.5894 5.871e-03 Assumptions NOT satisfied!
## Link Function       3.4730 6.238e-02    Assumptions acceptable.
```

```
## Heteroscedasticity 35.0089 3.282e-09 Assumptions NOT satisfied!
```

```r
#Checking Outliers for a few of the linear models

outlierTest(fit)
```

```
##
## No Studentized residuals with Bonferonni p < 0.05
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferonni p
## 115 3.227238          0.0015412      0.23118
```

```r
outlierTest(fit2)
```

```
##
## No Studentized residuals with Bonferonni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferonni p
## 135 -3.323084          0.0011271      0.16907
```

```r
outlierTest(fit3)
```

```
##      rstudent unadjusted p-value Bonferonni p
## 135 -3.793587         0.00021851     0.032776
```