DATA311_Project

Parsa

2019-03-07

```
test <- read.csv("Admission_Predict_Ver1.1.csv")
#summary (test)
head(test)
```

```
##   Serial.No. GRE.Score TOEFL.Score University.Rating SOP LOR CGPA Research
## 1          1       337         118                 4 4.5 4.5 9.65        1
## 2          2       324         107                 4 4.0 4.5 8.87        1
## 3          3       316         104                 3 3.0 3.5 8.00        1
## 4          4       322         110                 3 3.5 2.5 8.67        1
## 5          5       314         103                 2 2.0 3.0 8.21        0
## 6          6       330         115                 5 4.5 3.0 9.34        1
##   Chance.of.Admit
## 1            0.92
## 2            0.76
## 3            0.72
## 4            0.80
## 5            0.65
## 6            0.90
```

Logmod Analysis and Plots

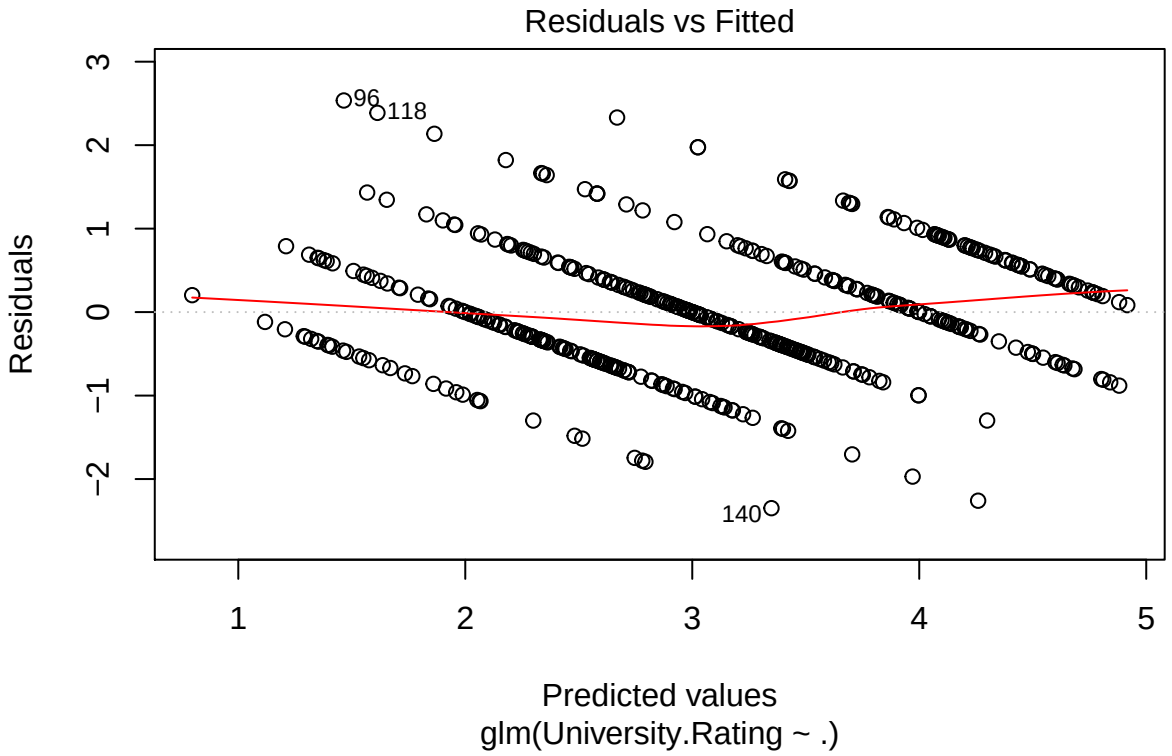
Here's a logmod analysis. No variable selection performed though.

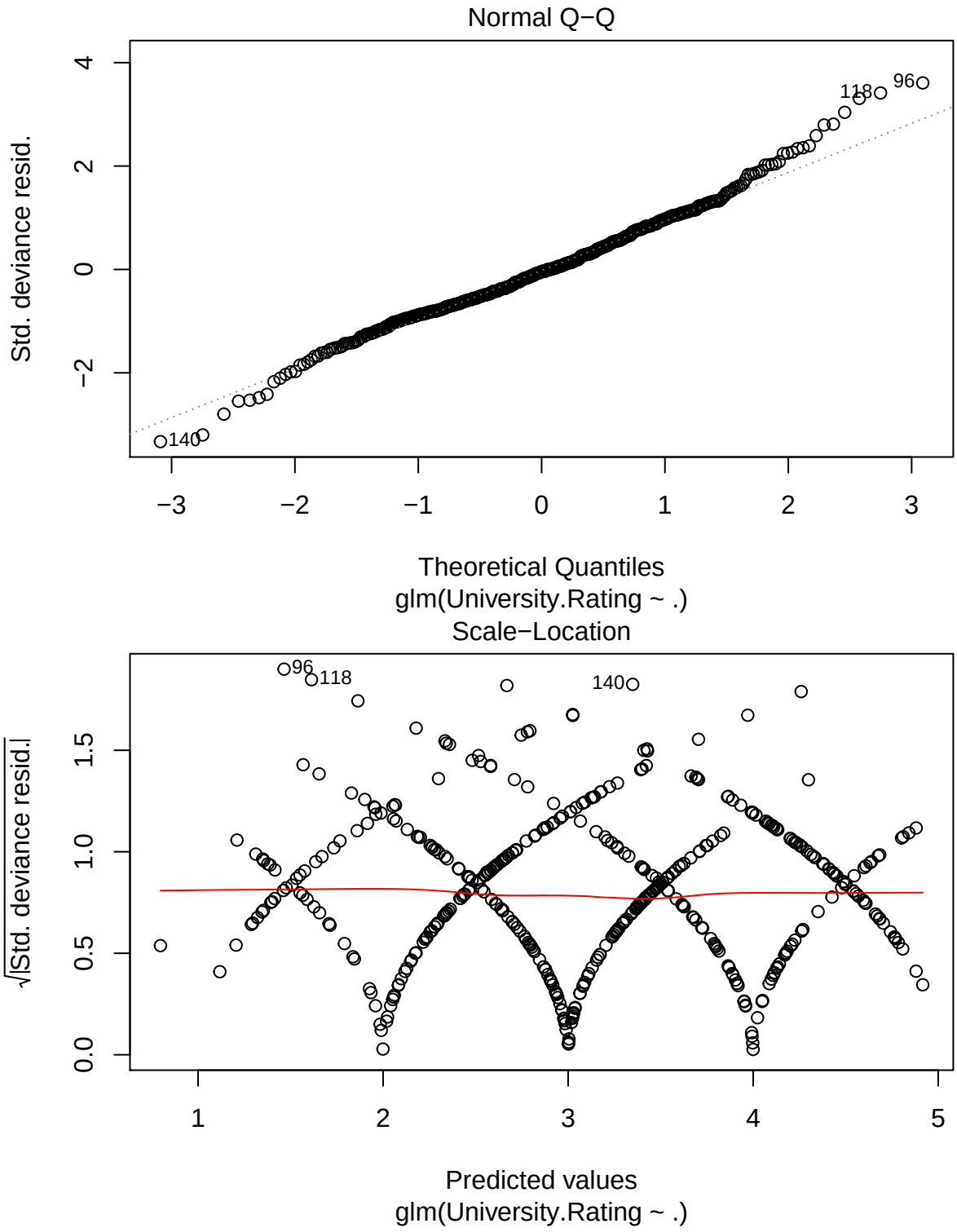
```
attach(test)
University.Rating <- factor(University.Rating)
Research <- factor(Research)
logmod <- glm(University.Rating ~., data=test)
summary(logmod)
```

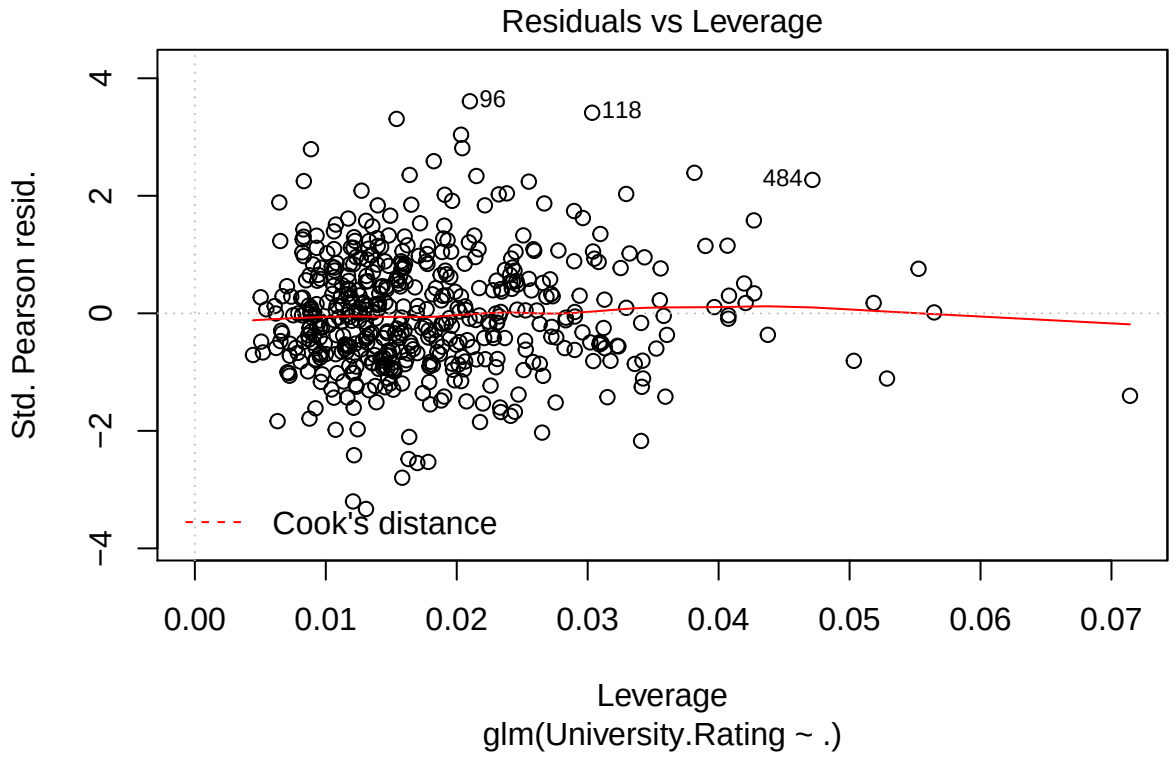
```
##
## Call:
## glm(formula = University.Rating ~ ., data = test)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.34889  -0.46404  -0.02909   0.43638   2.53513
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -5.3520556  1.4229030  -3.761 0.000189 ***
## Serial.No.       0.0001131  0.0002308   0.490 0.624275
## GRE.Score        0.0050723  0.0060361   0.840 0.401135
## TOEFL.Score      0.0184033  0.0104963   1.753 0.080172 .
## SOP              0.4420126  0.0508516   8.692  < 2e-16 ***
## LOR              0.1376178  0.0495241   2.779 0.005665 **
## CGPA             0.2666732  0.1306889   2.041 0.041833 *
## Research         0.0744728  0.0792227   0.940 0.347657
```

```
## Chance.of.Admit  0.7761573  0.5441596   1.426 0.154405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.5042716)
##
##     Null deviance: 652.5  on 499  degrees of freedom
## Residual deviance: 247.6  on 491  degrees of freedom
## AIC: 1087.5
##
## Number of Fisher Scoring iterations: 2
```

```
plot(logmod)
```




Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
glm(University.Rating ~ .)
Scale−Location
√|Std. deviance resid.|
Predicted values
glm(University.Rating ~ .)
96
118
140



Linear Regression and some plots

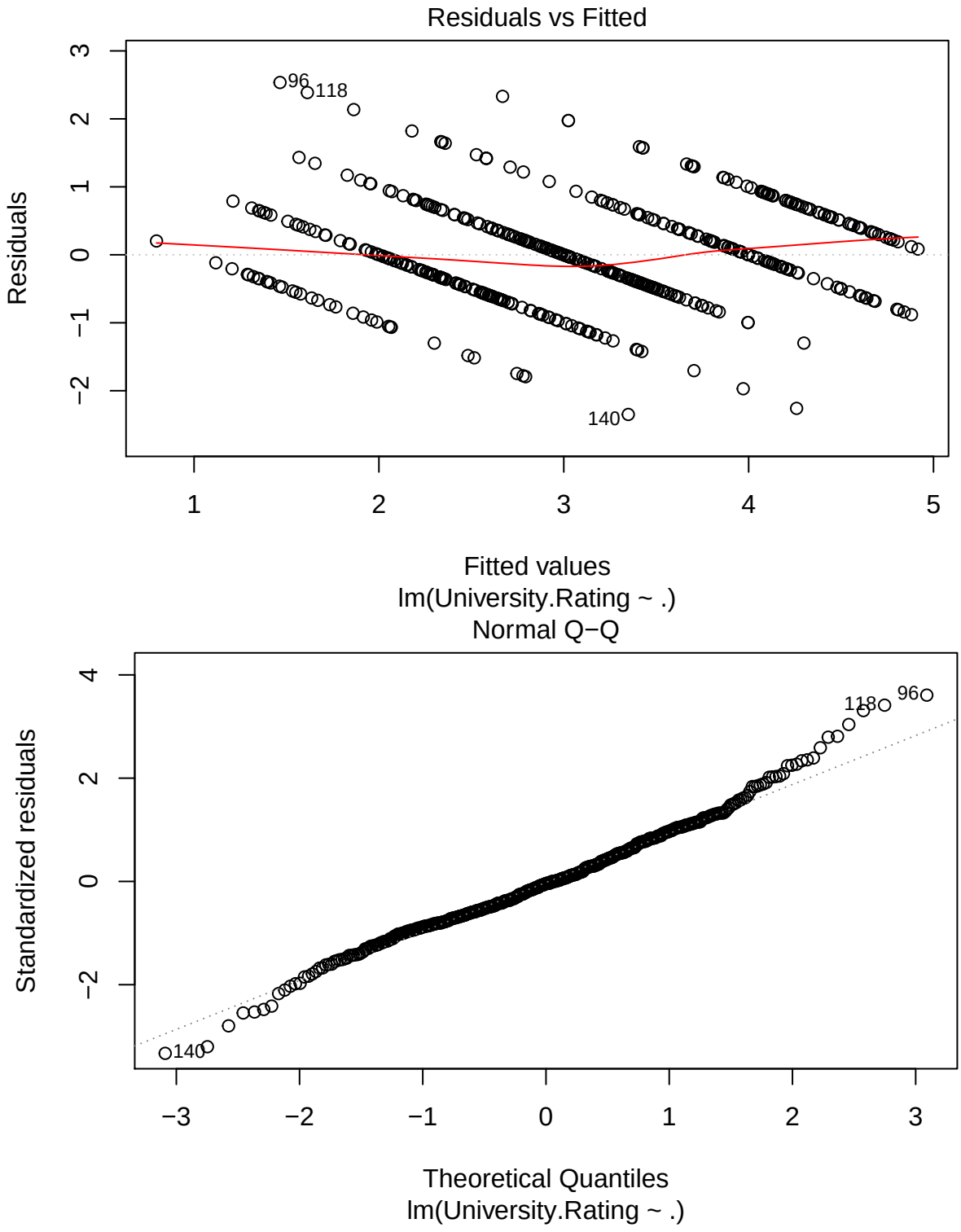
Here's a linear model with a few plots.

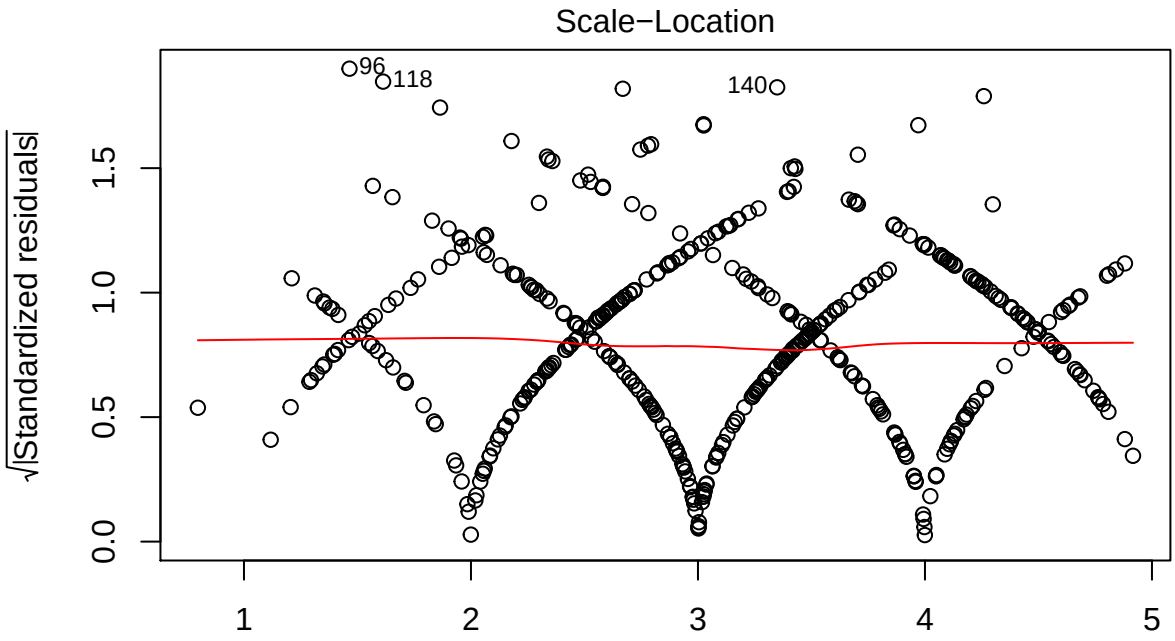
```
linear <- lm(University.Rating ~., data=test)
summary(linear)
```

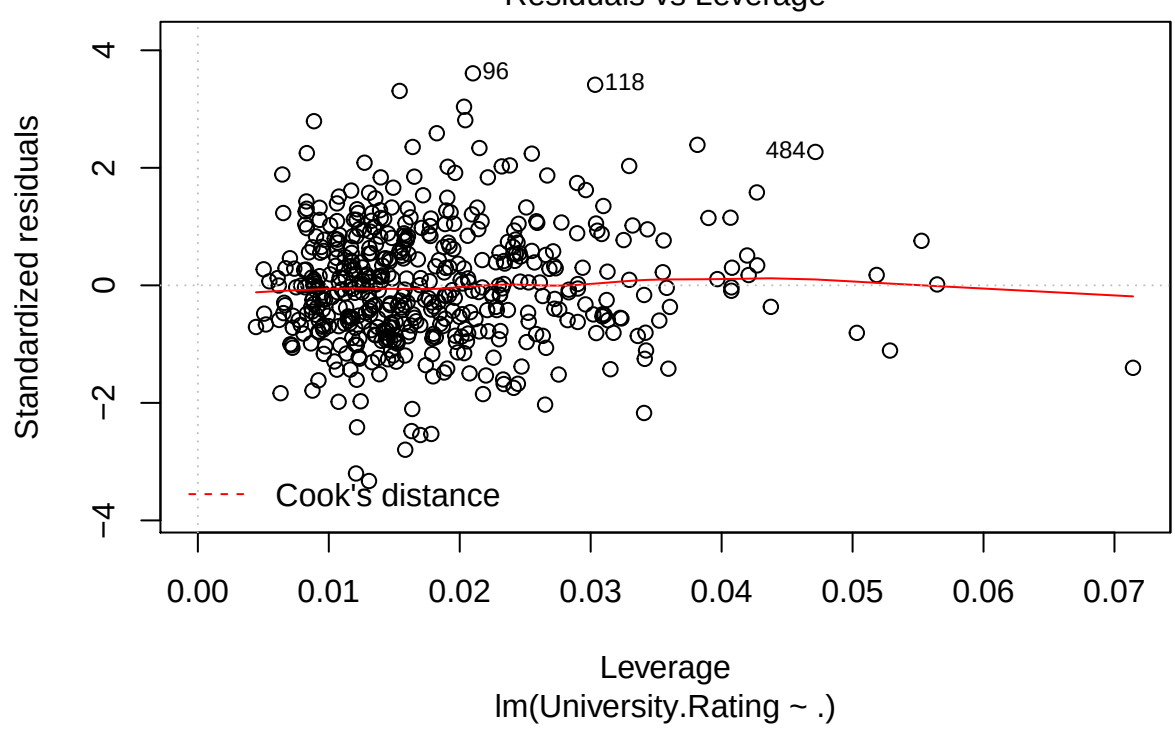
```
##
## Call:
## lm(formula = University.Rating ~ ., data = test)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.34889 -0.46404 -0.02909  0.43638  2.53513
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -5.3520556  1.4229030  -3.761 0.000189 ***
## Serial.No.       0.0001131  0.0002308   0.490 0.624275
## GRE.Score        0.0050723  0.0060361   0.840 0.401135
## TOEFL.Score      0.0184033  0.0104963   1.753 0.080172 .
## SOP              0.4420126  0.0508516   8.692  < 2e-16 ***
## LOR              0.1376178  0.0495241   2.779 0.005665 **
## CGPA             0.2666732  0.1306889   2.041 0.041833 *
## Research         0.0744728  0.0792227   0.940 0.347657
## Chance.of.Admit  0.7761573  0.5441596   1.426 0.154405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.7101 on 491 degrees of freedom
## Multiple R-squared:  0.6205, Adjusted R-squared:  0.6144
## F-statistic: 100.4 on 8 and 491 DF,  p-value: < 2.2e-16
```

```
plot(linear)
```



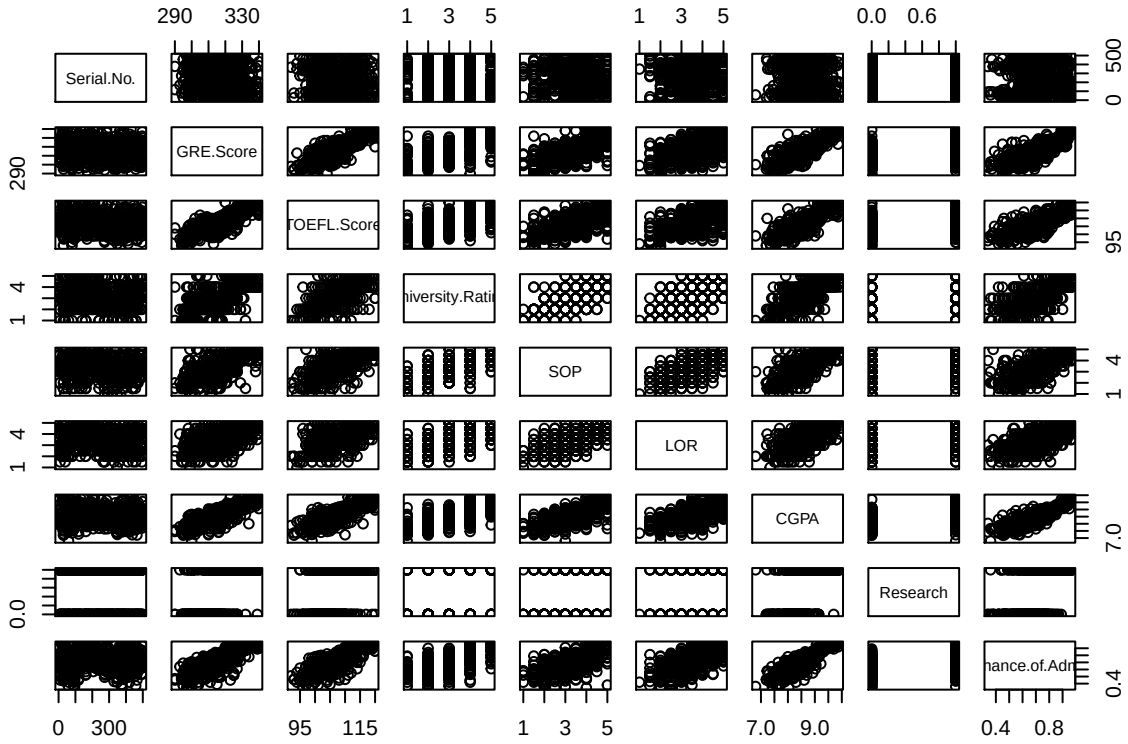




More Plots

Here's a bunch of scatterplots.

```
plot(test)
```



Classification

How about some Classification? Let's try knn

‘‘‘