Stats 101A project: final first draft

Simran Vatsa

3/15/2018

We began by systematically cleaning the data. For each variable, our first step was consulting the codebook to decide which codes could be converted to NAs. We decided that "not answered" could in every case be classified as NA. However, we were less sure about "Don't know" and "Inapplicable", but in our first iteration of data cleaning, we converted those to NAs too in order to simplify the task at hand and move on to exploratory analysis. We also converted the variables coded to denote "__ or more" to NAs as we felt those might skew our analysis.

```
#data cleanup
happiness_data <- read.table("Happiness.txt", header = TRUE)
head(happiness_data)
```

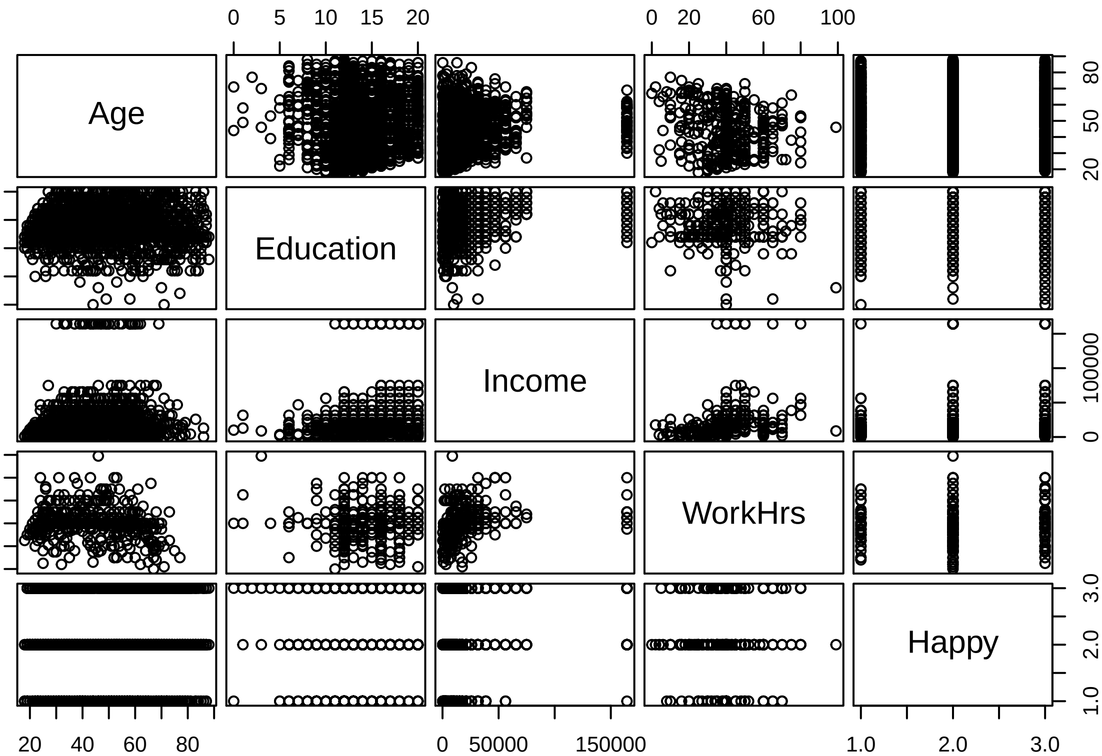
```
##   Household Health OwnHome Instagram Marital Sex Age Children Education
## 3         2      2       0         2       1   1  72        2        16
## 4         4      2       1         0       1   2  43        4        12
## 5         3      1       0         1       1   2  55        2        18
## 6         2      0       1         1       1   2  53        2        14
## 7         3      4       1         0       1   1  50        2        14
## 8         2      2       0         1       1   2  23        3        11
##   JobSat Income WorkHrs Happy
## 3      0      0      -1     1
## 4      0   5265      -1     2
## 5      3    936      15     1
## 6      0      0      -1     1
## 7      0 164382      -1     2
## 8      2   7605      30     1
```

```
happiness_data$Household[happiness_data$Household == 8 | happiness_data$Household == 9] <- NA

happiness_data$Health[happiness_data$Health == 8 | happiness_data$Health == 9 |  happiness_data$Health =
happiness_data$Health[happiness_data$Health == 1] <- 400
happiness_data$Health[happiness_data$Health == 2] <- 300
happiness_data$Health[happiness_data$Health == 3] <- 2
happiness_data$Health[happiness_data$Health == 4] <- 1
happiness_data$Health[happiness_data$Health == 400] <- 4
happiness_data$Health[happiness_data$Health == 300] <- 3

happiness_data$OwnHome[happiness_data$OwnHome == 0 | happiness_data$OwnHome == 8 | happiness_data$OwnHom

happiness_data$Instagram[happiness_data$Instagram == 0 | happiness_data$Instagram == 8 | happiness_data$

happiness_data$Marital[happiness_data$Marital == 9] <- NA

happiness_data$Age[happiness_data$Age == 89 | happiness_data$Age == 98 | happiness_data$Age == 99] <- NA

happiness_data$Children[happiness_data$Children == 8 | happiness_data$Children == 9] <- NA

happiness_data$Education[happiness_data$Education == 97 | happiness_data$Education == 98 | happiness_dat
```

```
happiness_data$JobSat[happiness_data$JobSat == 0 | happiness_data$JobSat == 8 | happiness_data$JobSat ==

happiness_data$Income[happiness_data$Income == 0 | happiness_data$Income == 999998 | happiness_data$Inc

happiness_data$WorkHrs[happiness_data$WorkHrs == -1 | happiness_data$WorkHrs == 998 | happiness_data$Wo

happiness_data$Happy[happiness_data$Happy == 0 | happiness_data$Happy == 8 | happiness_data$Happy == 9]

happiness_data$Happy[happiness_data$Happy == 1] <- 100
happiness_data$Happy[happiness_data$Happy == 3] <- 1
happiness_data$Happy[happiness_data$Happy == 100] <- 3
```

In exploring the data, we began with a scatterplot, which was not at all informative. We then split the variables into numerical and categorical and decided to analyze each group separately. Our numerical variables were Age, Education, Income and WorkHrs; we considered the rest as categorical and converted them into factors in order to build models with them. We put all our numerical variables into a model with Happy, obtaining an unsatisfactory R squared value of 0.017.

```
#exploring data
pairs(happiness_data[, -c(1,2,3,4,5,6,8,10)], gap=0.4,cex.labels=1.5)
```




```
m1 <- lm(happiness_data$Happy ~ happiness_data$Age + happiness_data$Education + happiness_data$Income +

summary(m1)
```

```
##
## Call:
## lm(formula = happiness_data$Happy ~ happiness_data$Age + happiness_data$Education +
##     happiness_data$Income + happiness_data$WorkHrs)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -1.5164 -0.2267 -0.1453  0.7486  0.9651
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.915e+00  2.240e-01   8.546 3.04e-16 ***
## happiness_data$Age        2.136e-03  2.470e-03   0.865    0.388
## happiness_data$Education -2.974e-03  1.052e-02  -0.283    0.778
## happiness_data$Income     1.950e-06  1.412e-06   1.381    0.168
## happiness_data$WorkHrs    4.098e-03  2.552e-03   1.606    0.109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6347 on 384 degrees of freedom
##   (1978 observations deleted due to missingness)
## Multiple R-squared:  0.01724,    Adjusted R-squared:  0.007001
## F-statistic: 1.684 on 4 and 384 DF,  p-value: 0.1529
```

Then, we created a model with all predictors against Happy, with the categorical variables in factor form. We used an inverse response plot on the model, just to see how high of an R squared value we could get to obtain an estimate of what to aim for. This gave us an R squared value of around 0.303, probably from overfitting; the adjusted R squared was 0.2235.

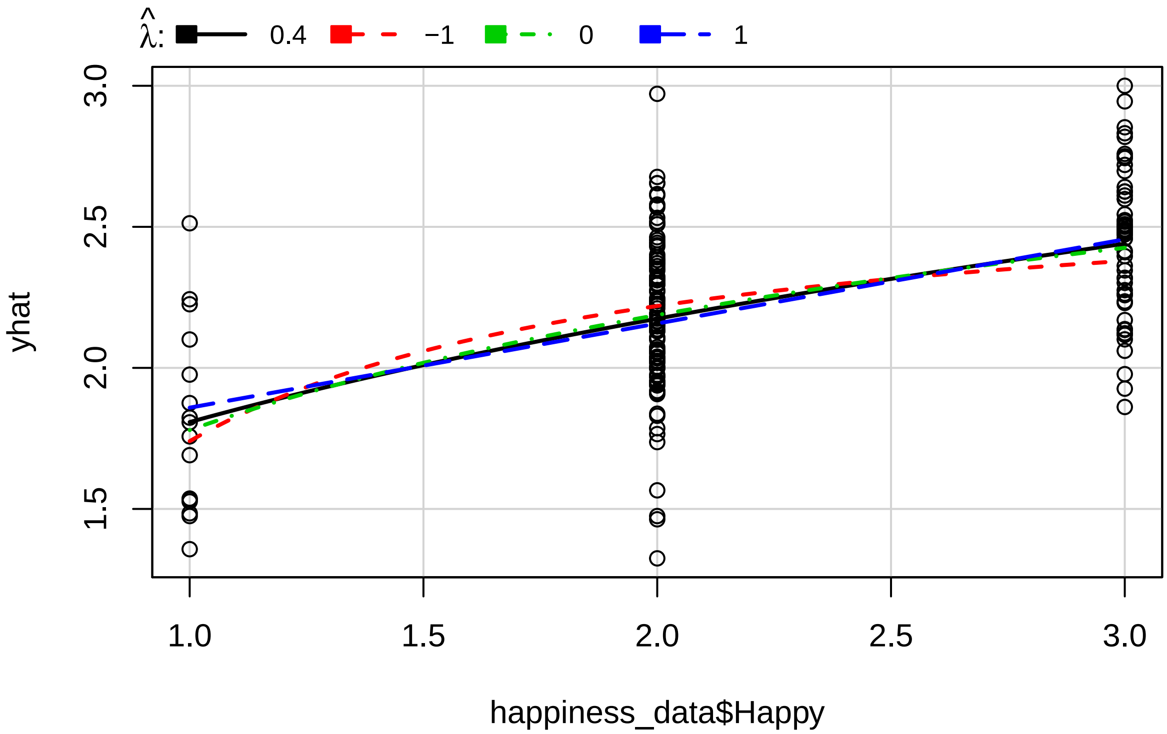
```r
library(alr3)
```

```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 3.4.3
```

```r
m1 <- lm(happiness_data$Happy ~ factor(happiness_data$Household) + factor(happiness_data$OwnHome) + happ
inverse.response.plot(m1,key=TRUE)
```

```
## Warning: 'inverse.response.plot' is deprecated.
## Use 'inverseResponsePlot' instead.
## See help("Deprecated") and help("alr3-deprecated").
```




```
##      lambda      RSS
## 1  0.396029 12.62671
## 2 -1.000000 13.09307
## 3  0.000000 12.66673
## 4  1.000000 12.70891
```

```
m2 <- lm((happiness_data$Happy)^0.396 ~ factor(happiness_data$Household) + factor(happiness_data$OwnHom
summary(m2)
```

```
##
## Call:
## lm(formula = (happiness_data$Happy)^0.396 ~ factor(happiness_data$Household) +
##     factor(happiness_data$OwnHome) + happiness_data$Instagram +
##     factor(happiness_data$Marital) + happiness_data$Children +
##     happiness_data$Education + factor(happiness_data$JobSat) +
##     happiness_data$Income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42922 -0.07632 -0.00066  0.10275  0.28342
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1.437e+00  1.015e-01  14.154  < 2e-16
## factor(happiness_data$Household)2 -4.743e-02  3.287e-02  -1.443 0.151175
## factor(happiness_data$Household)3 -1.058e-02  4.249e-02  -0.249 0.803668
## factor(happiness_data$Household)4 -1.954e-01  8.817e-02  -2.216 0.028286
## factor(happiness_data$Household)5 -8.682e-02  1.472e-01  -0.590 0.556135
## factor(happiness_data$Household)6 -1.814e-01  1.059e-01  -1.712 0.088974
## factor(happiness_data$OwnHome)2    3.689e-02  2.654e-02   1.390 0.166594
## factor(happiness_data$OwnHome)3    8.115e-02  1.483e-01   0.547 0.585169
```

```
## happiness_data$Instagram          -2.169e-02  2.714e-02  -0.799 0.425376
## factor(happiness_data$Marital)2   -2.232e-01  5.854e-02  -3.812 0.000204
## factor(happiness_data$Marital)3   -9.841e-02  3.802e-02  -2.588 0.010638
## factor(happiness_data$Marital)4   -1.541e-01  8.703e-02  -1.771 0.078747
## factor(happiness_data$Marital)5   -9.380e-02  3.764e-02  -2.492 0.013842
## happiness_data$Children           -2.539e-03  9.523e-03  -0.267 0.790147
## happiness_data$Education           5.854e-03  4.913e-03   1.191 0.235436
## factor(happiness_data$JobSat)2    -2.026e-02  3.402e-02  -0.596 0.552445
## factor(happiness_data$JobSat)3    -7.703e-02  3.610e-02  -2.134 0.034568
## factor(happiness_data$JobSat)4    -6.828e-02  6.801e-02  -1.004 0.317088
## factor(happiness_data$JobSat)5    -1.423e-01  5.366e-02  -2.652 0.008917
## factor(happiness_data$JobSat)6    -2.960e-01  7.744e-02  -3.822 0.000197
## factor(happiness_data$JobSat)7    -7.765e-02  1.470e-01  -0.528 0.598234
## happiness_data$Income              6.106e-07  4.021e-07   1.519 0.131076
##
## (Intercept)                       ***
## factor(happiness_data$Household)2
## factor(happiness_data$Household)3
## factor(happiness_data$Household)4 *
## factor(happiness_data$Household)5
## factor(happiness_data$Household)6 .
## factor(happiness_data$OwnHome)2
## factor(happiness_data$OwnHome)3
## happiness_data$Instagram
## factor(happiness_data$Marital)2   ***
## factor(happiness_data$Marital)3   *
## factor(happiness_data$Marital)4   .
## factor(happiness_data$Marital)5   *
## happiness_data$Children
## happiness_data$Education
## factor(happiness_data$JobSat)2
## factor(happiness_data$JobSat)3    *
## factor(happiness_data$JobSat)4
## factor(happiness_data$JobSat)5    **
## factor(happiness_data$JobSat)6    ***
## factor(happiness_data$JobSat)7
## happiness_data$Income
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1406 on 143 degrees of freedom
##   (2202 observations deleted due to missingness)
## Multiple R-squared:  0.3064, Adjusted R-squared:  0.2045
## F-statistic: 3.008 on 21 and 143 DF,  p-value: 5.709e-05
```

At this stage, after getting comfortable with the data, we decided to make some modifications to how we selected codes to turn into NAs; we did not convert 8 (8 or more) in Household or Children, and we converted 0 (Inapplicable) and 8 (Don't know) to 2 (No) in Instagram. With these new conditions, we repeated plotting the full model, carrying out more complete analyses on it this time. Its R squared value was 0.2811, with an adjusted R squared of 0.2069.

```r
#data cleanup
getwd()
```

```
## [1] "/Users/narenakurati/stats-101a-project-download"
```

```r
happiness_data <- read.table("Happiness.txt", header = TRUE)
head(happiness_data)
```

```
##   Household Health OwnHome Instagram Marital Sex Age Children Education
## 3         2      2       0         2       1   1  72        2        16
## 4         4      2       1         0       1   2  43        4        12
## 5         3      1       0         1       1   2  55        2        18
## 6         2      0       1         1       1   2  53        2        14
## 7         3      4       1         0       1   1  50        2        14
## 8         2      2       0         1       1   2  23        3        11
##   JobSat Income WorkHrs Happy
## 3      0      0      -1     1
## 4      0   5265      -1     2
## 5      3    936      15     1
## 6      0      0      -1     1
## 7      0 164382      -1     2
## 8      2   7605      30     1
```

```r
#Household
happiness_data$Household[happiness_data$Household == 9] <- NA

#Health
happiness_data$Health[happiness_data$Health == 8 | happiness_data$Health == 9 |  happiness_data$Health =
happiness_data$Health[happiness_data$Health == 1] <- 400
happiness_data$Health[happiness_data$Health == 2] <- 300
happiness_data$Health[happiness_data$Health == 3] <- 2
happiness_data$Health[happiness_data$Health == 4] <- 1
happiness_data$Health[happiness_data$Health == 400] <- 4
happiness_data$Health[happiness_data$Health == 300] <- 3

#OwnHome
happiness_data$OwnHome[happiness_data$OwnHome == 0 | happiness_data$OwnHome == 8 | happiness_data$OwnHom

#Instagram - Set 'don't know' and 'inapplicable' to 'No'
happiness_data$Instagram[happiness_data$Instagram == 0 | happiness_data$Instagram == 8] <- 2
happiness_data$Instagram[happiness_data$Instagram == 9] <- NA

#Marital
happiness_data$Marital[happiness_data$Marital == 9] <- NA

#Age
happiness_data$Age[happiness_data$Age == 89 | happiness_data$Age == 98 | happiness_data$Age == 99] <- N

#Children
happiness_data$Children[happiness_data$Children == 9] <- NA

#Education
happiness_data$Education[happiness_data$Education == 97 | happiness_data$Education == 98 | happiness_da

#JobSat
happiness_data$JobSat[happiness_data$JobSat == 0 | happiness_data$JobSat == 8 | happiness_data$JobSat =

#Income
happiness_data$Income[happiness_data$Income == 0 | happiness_data$Income == 999998 | happiness_data$Inc
```

```r
#WorkHrs
happiness_data$WorkHrs[happiness_data$WorkHrs == -1 | happiness_data$WorkHrs == 998 | happiness_data$Wo

#Happy
happiness_data$Happy[happiness_data$Happy == 0 | happiness_data$Happy == 8 | happiness_data$Happy == 9]
happiness_data$Happy[happiness_data$Happy == 1] <- 100
happiness_data$Happy[happiness_data$Happy == 3] <- 1
happiness_data$Happy[happiness_data$Happy == 100] <- 3
```

The first conclusion we came to in our model selection process was that WorkHrs could be excluded, as there were 1898 missing values, most of which were -1 (Inapplicable). While we weren't sure whether "Inapplicable" could be considered a missing value per se, there did not seem to be any way around categorizing it so, since we had decided to treat WorkHrs as a numerical variable. Its sheer amount of missing values made it ineligible for model fitting - every attempt to include it resulted in an error being thrown.

```r
#Finding the number of NAs
sum(is.na(happiness_data$WorkHrs))
```

```
## [1] 1898
```

```r
#Insta = 10, Marital = 1
#Household = 1
#Health = 811
#OwnHome = 812
#JobSat = 1612
#WorkHrs = 1898
#Income = 1039
```

We plotted the full model (without WorkHrs). This factored in around 190 observations, as the rest had NAs under some variables. We found the Residuals vs Fitted plot showed a decreasing linear trend, a result of some of the predictor variables being categorical. The standardized residual plot also showed a pattern that skewed the plot much more than it did in the Residuals vs. Fitted plot.

```r
attach(happiness_data)
#Factoring Categorical Variables
JobSat.f <- factor(JobSat)
OwnHome.f <- factor(OwnHome)
Marital.f <- factor(Marital)
Instagram.f <- factor(Instagram)
Health.f <- factor(Health)

attach(happiness_data)
```

```
## The following objects are masked from happiness_data (pos = 3):
##
##     Age, Children, Education, Happy, Health, Household, Income,
##     Instagram, JobSat, Marital, OwnHome, Sex, WorkHrs
```

```r
library(alr3)
#Couldn't include Health as it throws an error
full_model <- lm(Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children + Education + JobSa
summary(full_model)
```

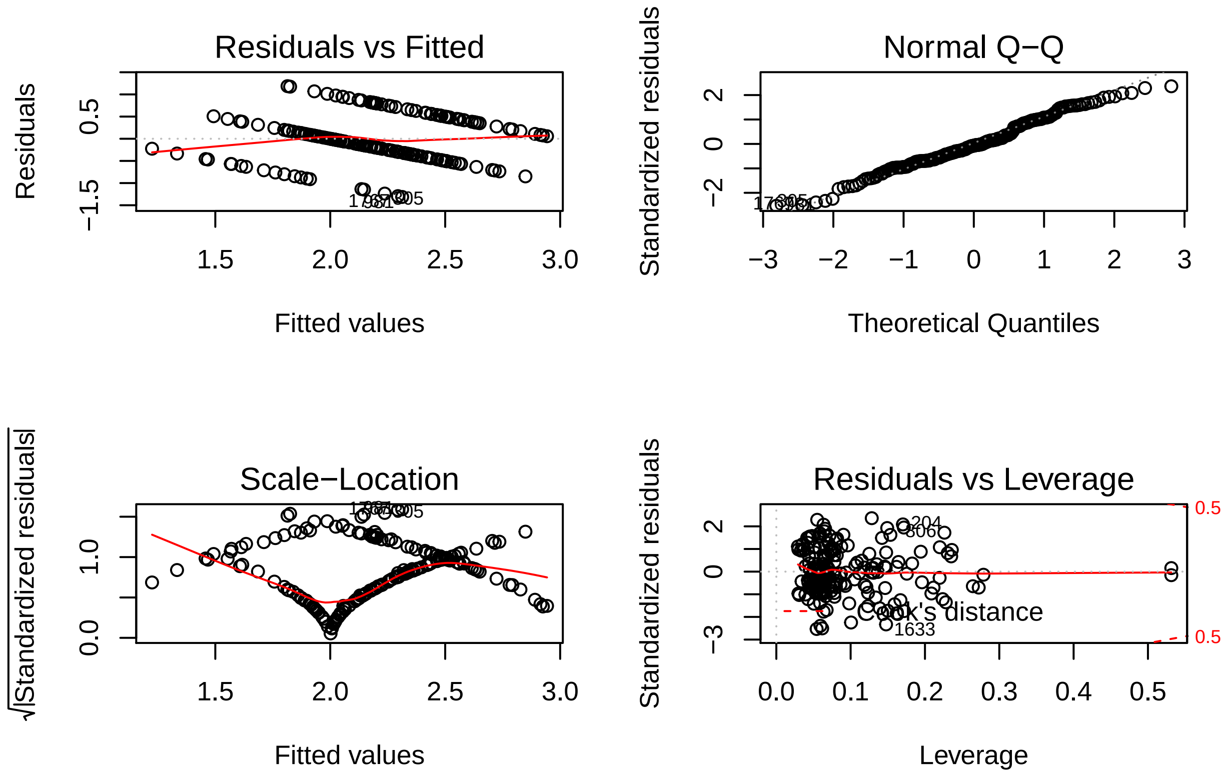
```
##
## Call:
## lm(formula = Happy ~ Household + OwnHome.f + Instagram.f + Marital.f +
```

```
##     Children + Education + JobSat.f + Income + Age + Sex)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31296 -0.34161 -0.03404  0.38300  1.18586
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.766e+00  3.533e-01   7.828 3.79e-13 ***
## Household    -1.355e-01  4.633e-02  -2.925 0.003873 **
## OwnHome.f2    1.446e-02  8.866e-02   0.163 0.870596
## OwnHome.f3    7.315e-01  4.018e-01   1.821 0.070281 .
## Instagram.f2 -6.820e-02  1.004e-01  -0.679 0.498031
## Marital.f2   -6.236e-01  2.100e-01  -2.970 0.003380 **
## Marital.f3   -3.082e-01  1.173e-01  -2.626 0.009363 **
## Marital.f4   -3.157e-01  2.516e-01  -1.255 0.211125
## Marital.f5   -3.743e-01  1.180e-01  -3.173 0.001771 **
## Children     -2.411e-02  3.083e-02  -0.782 0.435233
## Education     2.065e-02  1.623e-02   1.272 0.204876
## JobSat.f2    -1.017e-01  1.145e-01  -0.888 0.375512
## JobSat.f3    -3.232e-01  1.227e-01  -2.635 0.009136 **
## JobSat.f4    -2.273e-01  1.985e-01  -1.145 0.253594
## JobSat.f5    -6.573e-01  1.792e-01  -3.667 0.000321 ***
## JobSat.f6    -9.258e-01  2.631e-01  -3.518 0.000547 ***
## JobSat.f7    -3.099e-01  5.496e-01  -0.564 0.573515
## Income        2.733e-06  1.484e-06   1.841 0.067221 .
## Age          -6.511e-03  3.626e-03  -1.796 0.074208 .
## Sex           9.907e-02  8.440e-02   1.174 0.241943
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5332 on 184 degrees of freedom
##   (2163 observations deleted due to missingness)
## Multiple R-squared:  0.2811, Adjusted R-squared:  0.2069
## F-statistic: 3.787 on 19 and 184 DF,  p-value: 1.092e-06
```

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

```
## Warning: not plotting observations with leverage one:
##   93
```

```
## Warning: not plotting observations with leverage one:
##   93
```

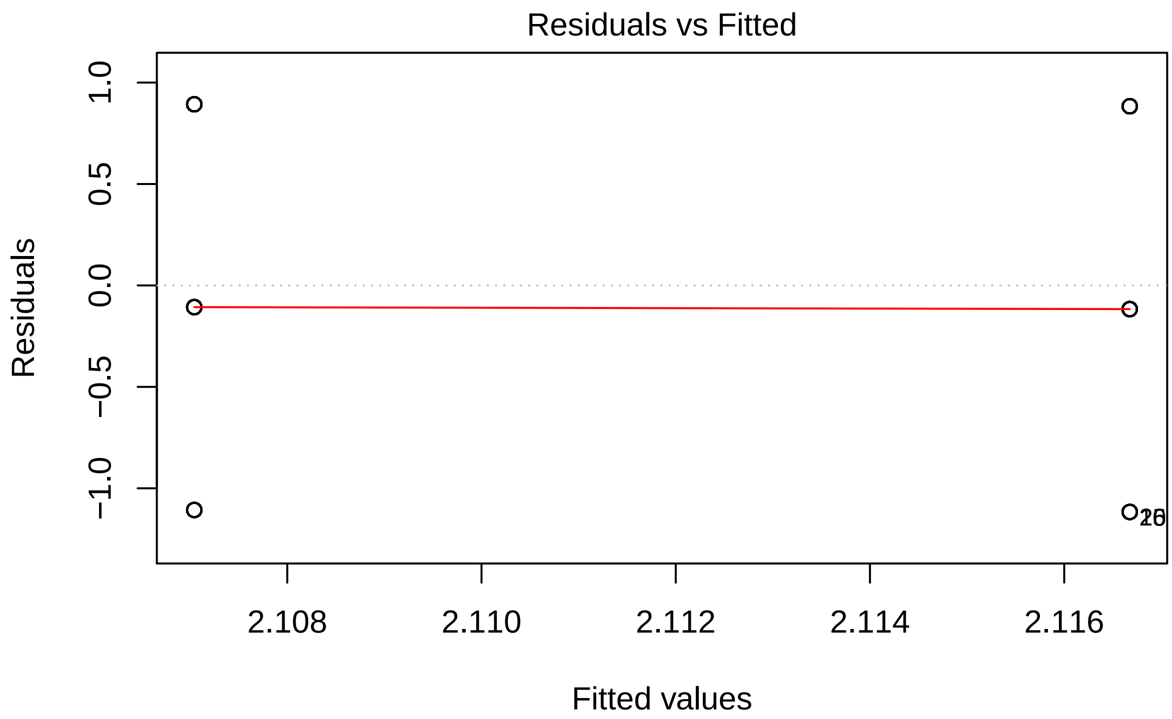


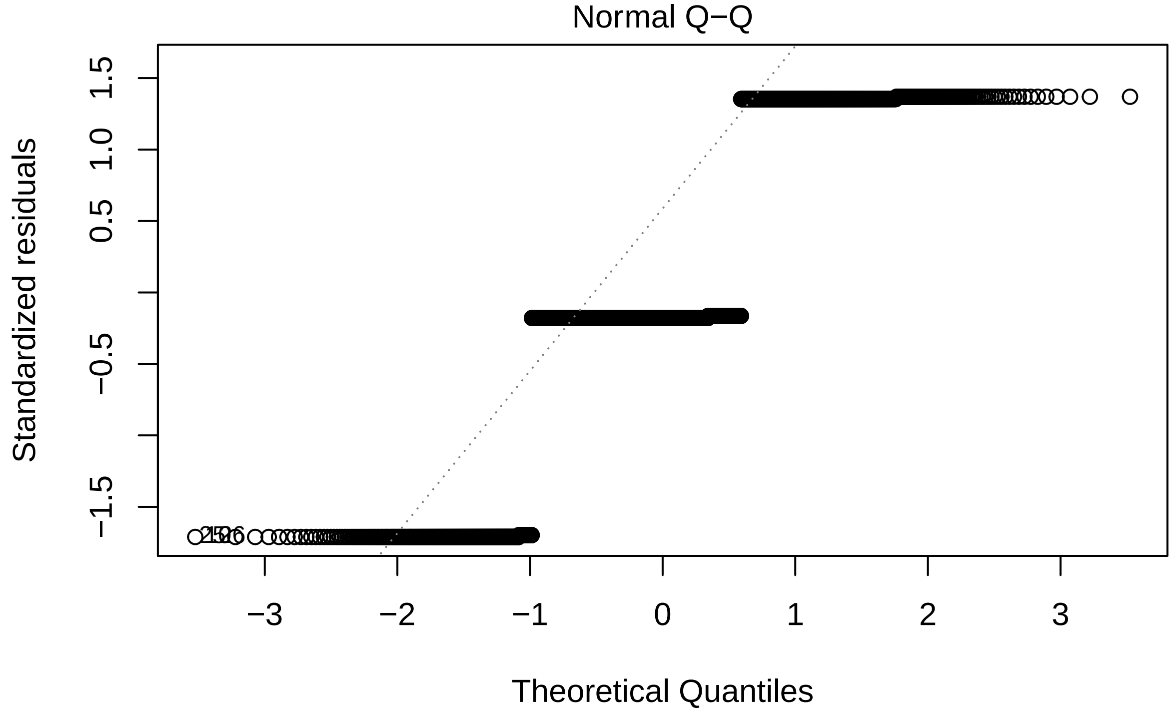
We decided the best way to proceed would be to test each variable's significance individually. We correlated individual variables with Happy, finding that Instagram and Sex did not have statistically significant linear relationships to Happy.

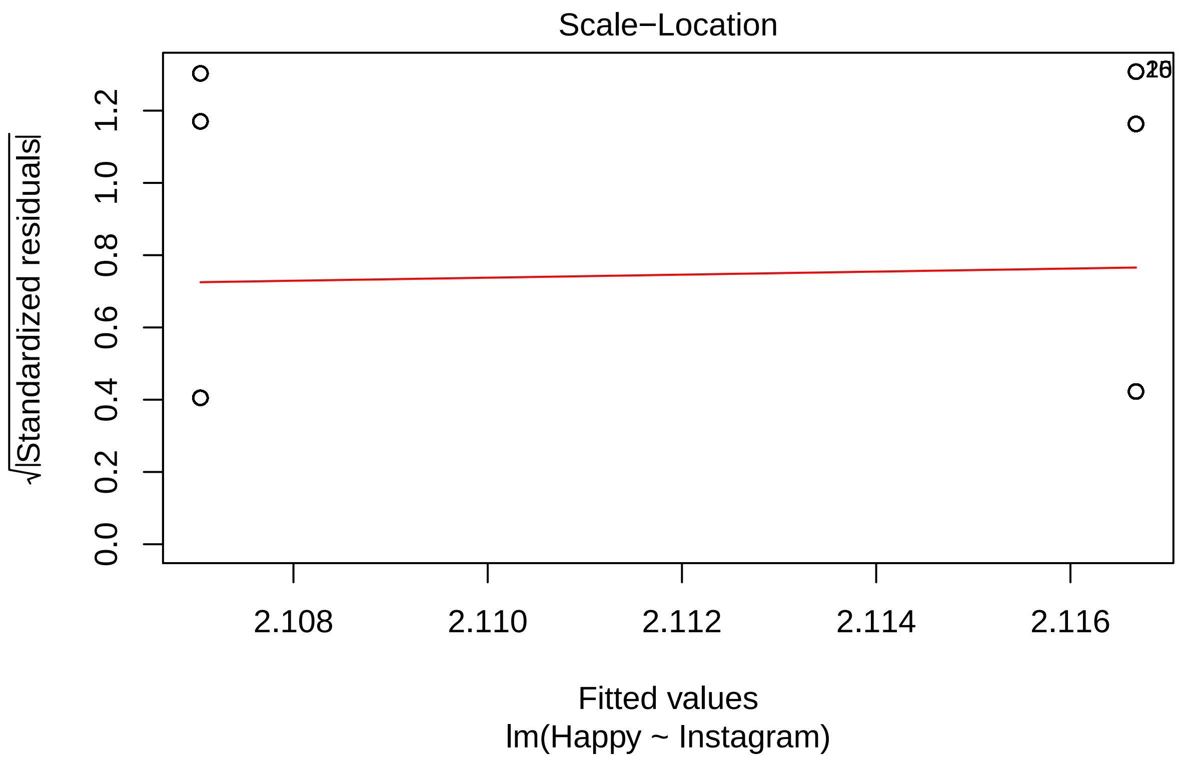
```r
insta <- lm(Happy ~ Instagram)
summary(insta)
```

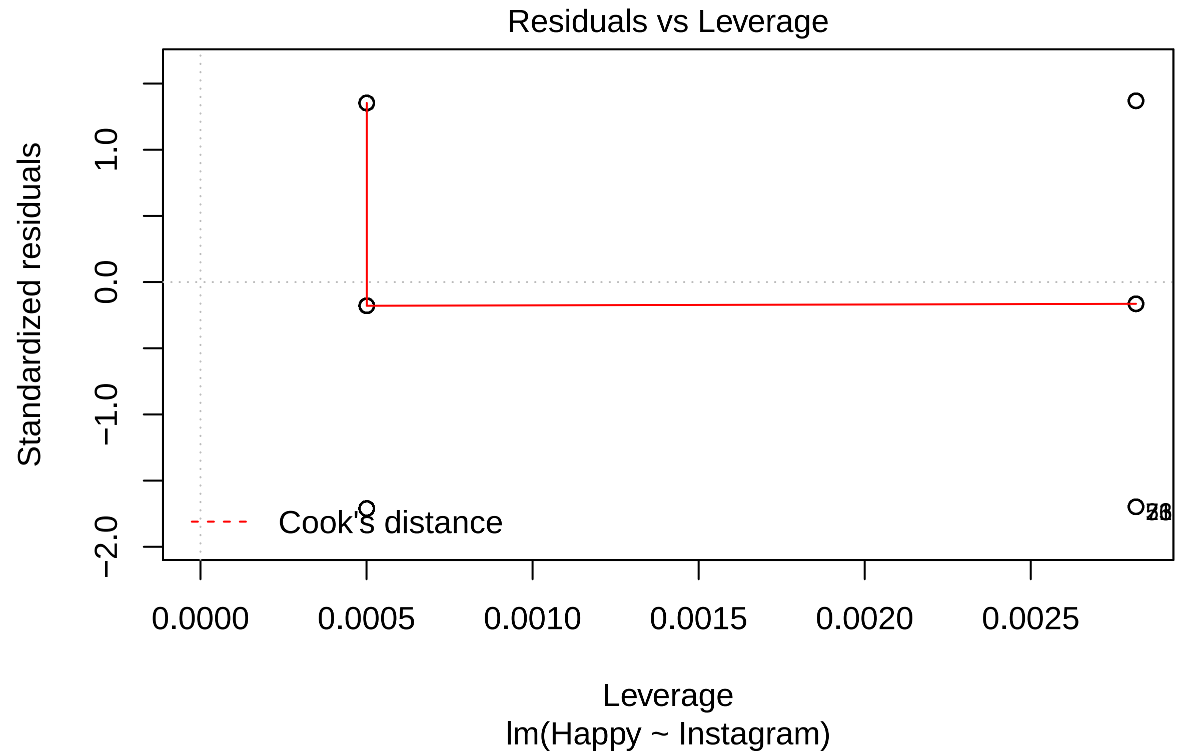
```
##
## Call:
## lm(formula = Happy ~ Instagram)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1167 -0.1167 -0.1167  0.8833  0.8930
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.097409   0.070828  29.613   <2e-16 ***
## Instagram   0.009633   0.037606   0.256    0.798
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6529 on 2350 degrees of freedom
##   (15 observations deleted due to missingness)
## Multiple R-squared:  2.792e-05,  Adjusted R-squared:  -0.0003976
## F-statistic: 0.06561 on 1 and 2350 DF,  p-value: 0.7979
```

```r
plot(insta)
```


Residuals vs Fitted
Residuals
1.0
0.5
0.0
−0.5
−1.0
2.108
2.110
2.112
2.114
2.116
Fitted values
lm(Happy ~ Instagram)


Normal Q−Q
Standardized residuals
1.5
1.0
0.5
−0.5
−1.5
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(Happy ~ Instagram)



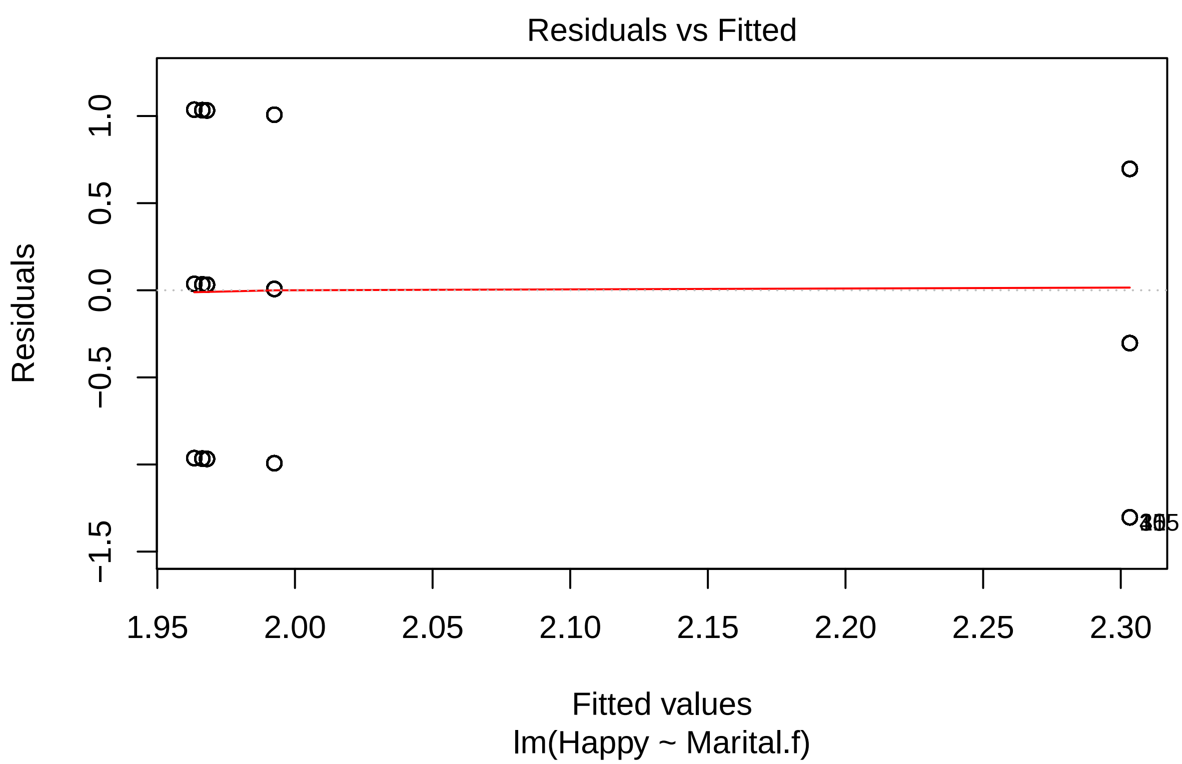


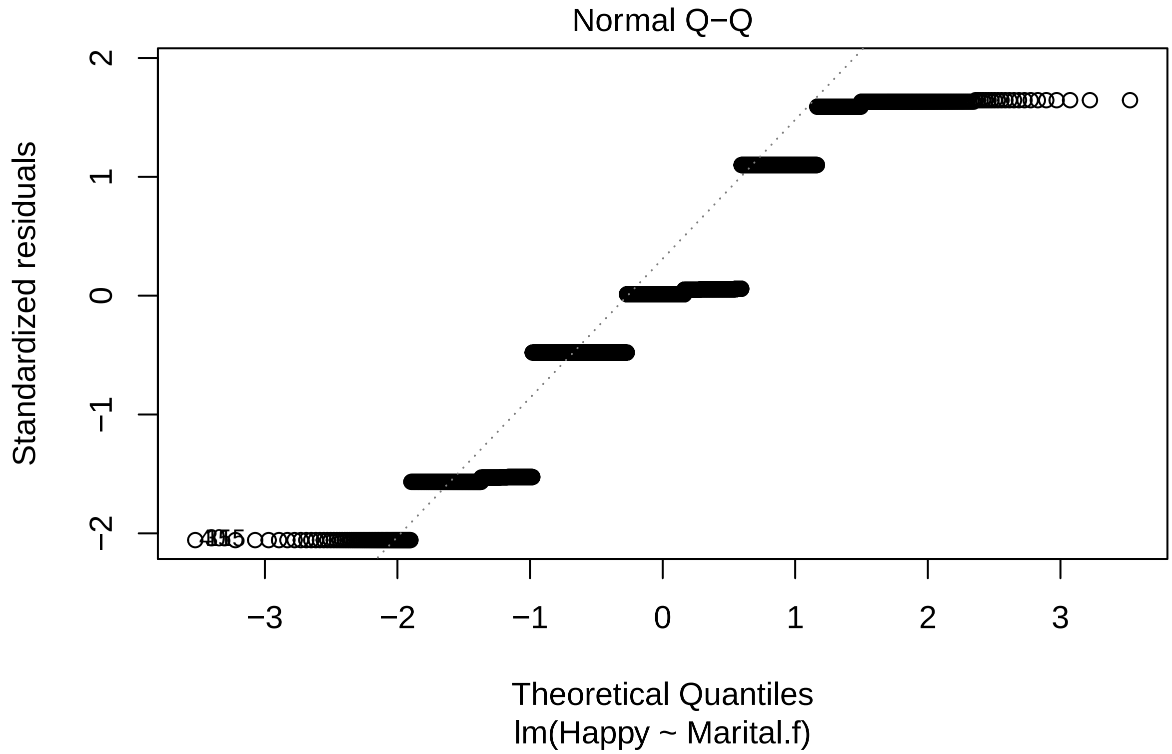

```
#Instagram is insignificant

marital <- lm(Happy ~ Marital.f)
summary(marital)
```

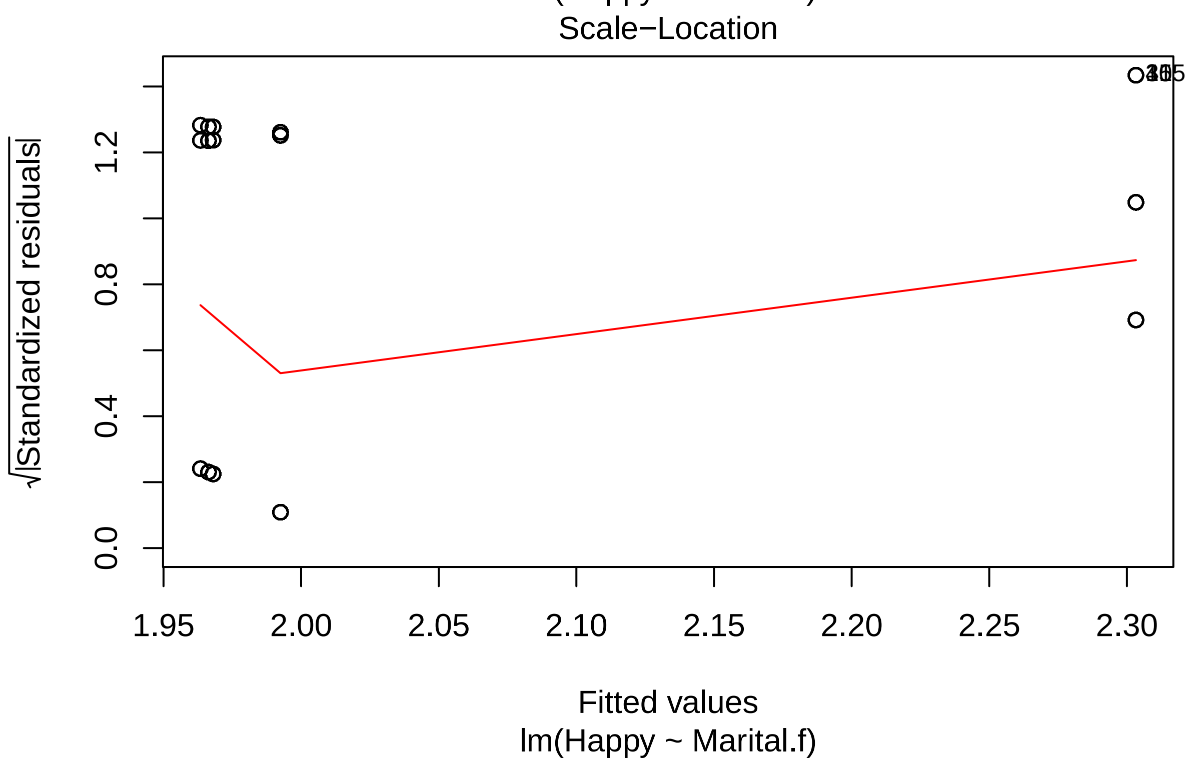
```
##
```

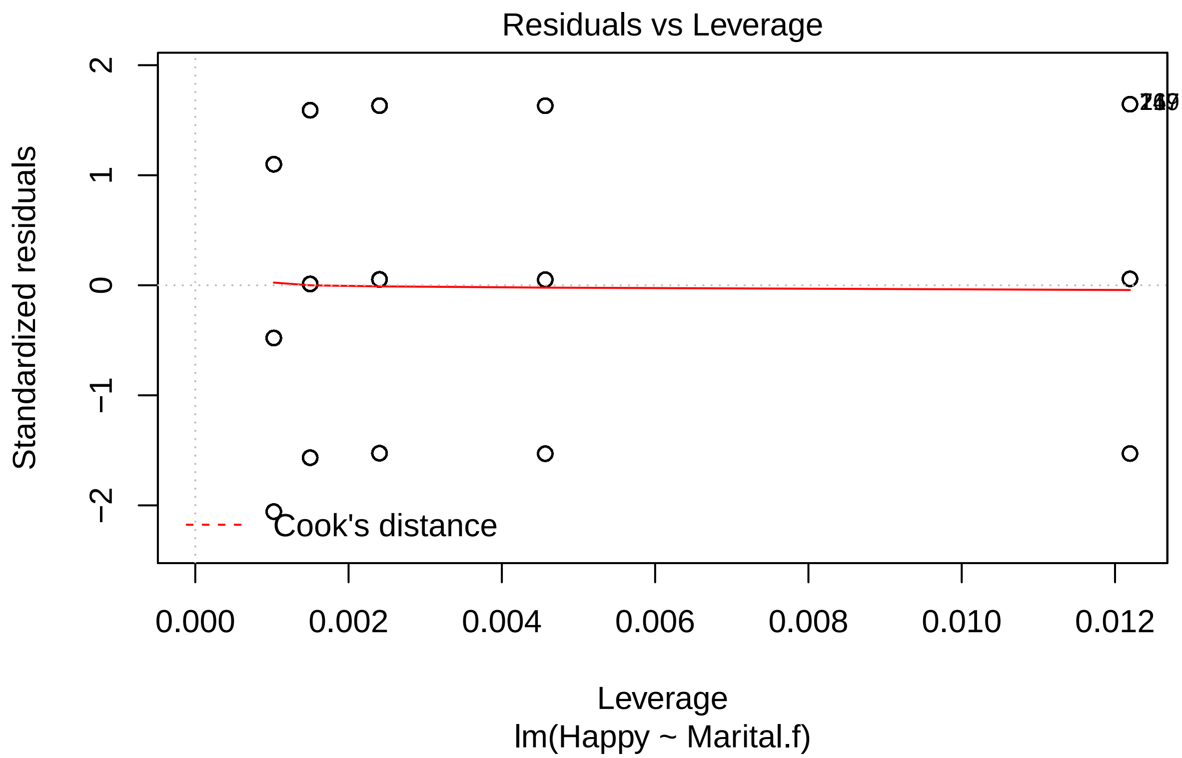
```
## Call:
## lm(formula = Happy ~ Marital.f)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3033 -0.3033  0.0075  0.6967  1.0366
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.30328    0.02029 113.510  < 2e-16 ***
## Marital.f2  -0.33524    0.04740  -7.073 2.00e-12 ***
## Marital.f3  -0.33693    0.03712  -9.077  < 2e-16 ***
## Marital.f4  -0.33986    0.07289  -4.663 3.29e-06 ***
## Marital.f5  -0.31077    0.03185  -9.758  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6339 on 2355 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:   0.06,  Adjusted R-squared:  0.0584
## F-statistic: 37.58 on 4 and 2355 DF,  p-value: < 2.2e-16
```

```r
plot(marital)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Happy ~ Marital.f)

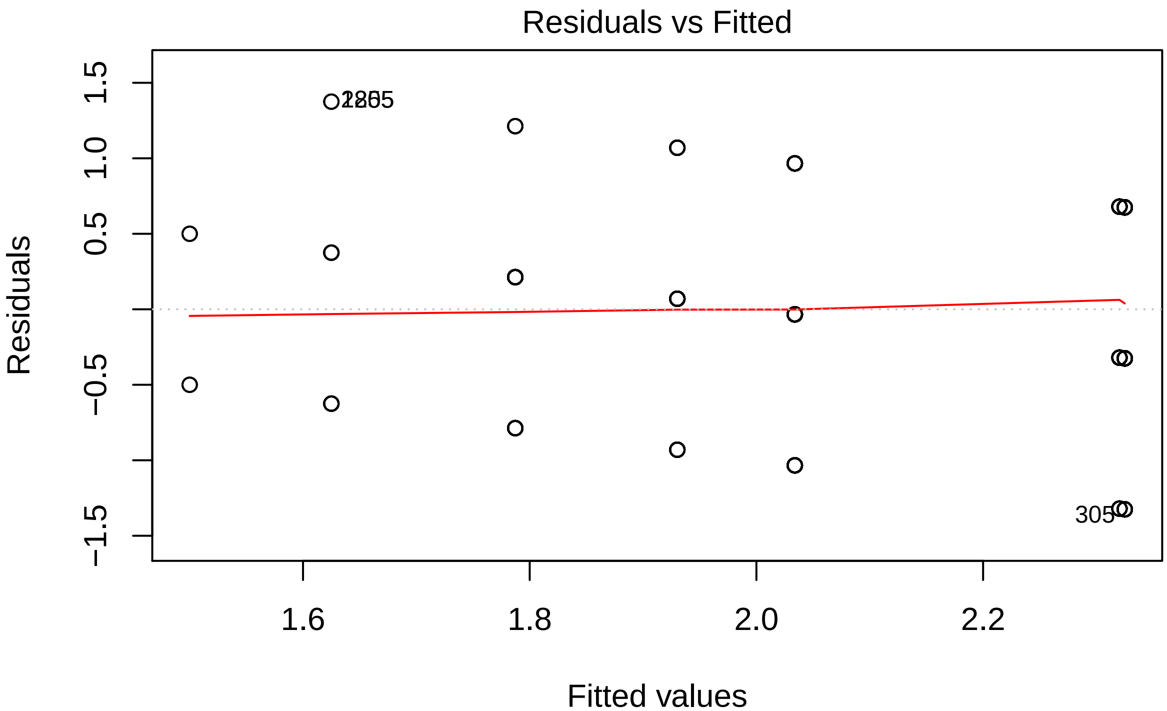

Scale−Location
√|Standardized residuals|
Fitted values
lm(Happy ~ Marital.f)

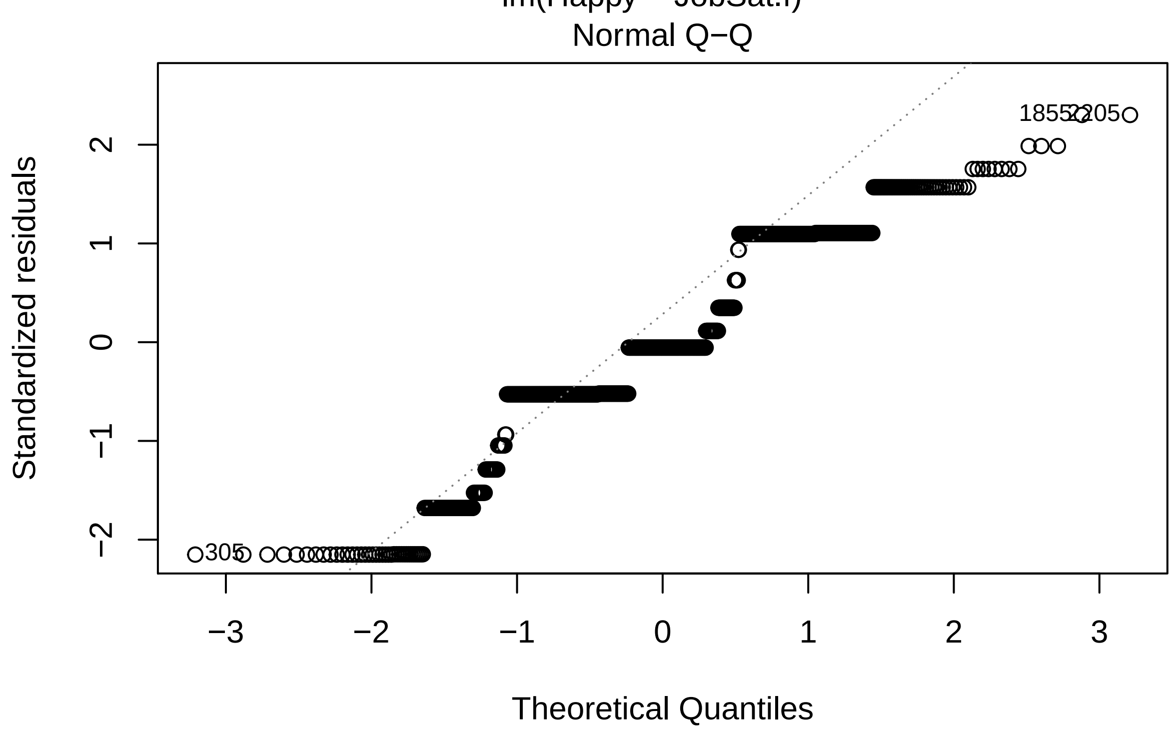


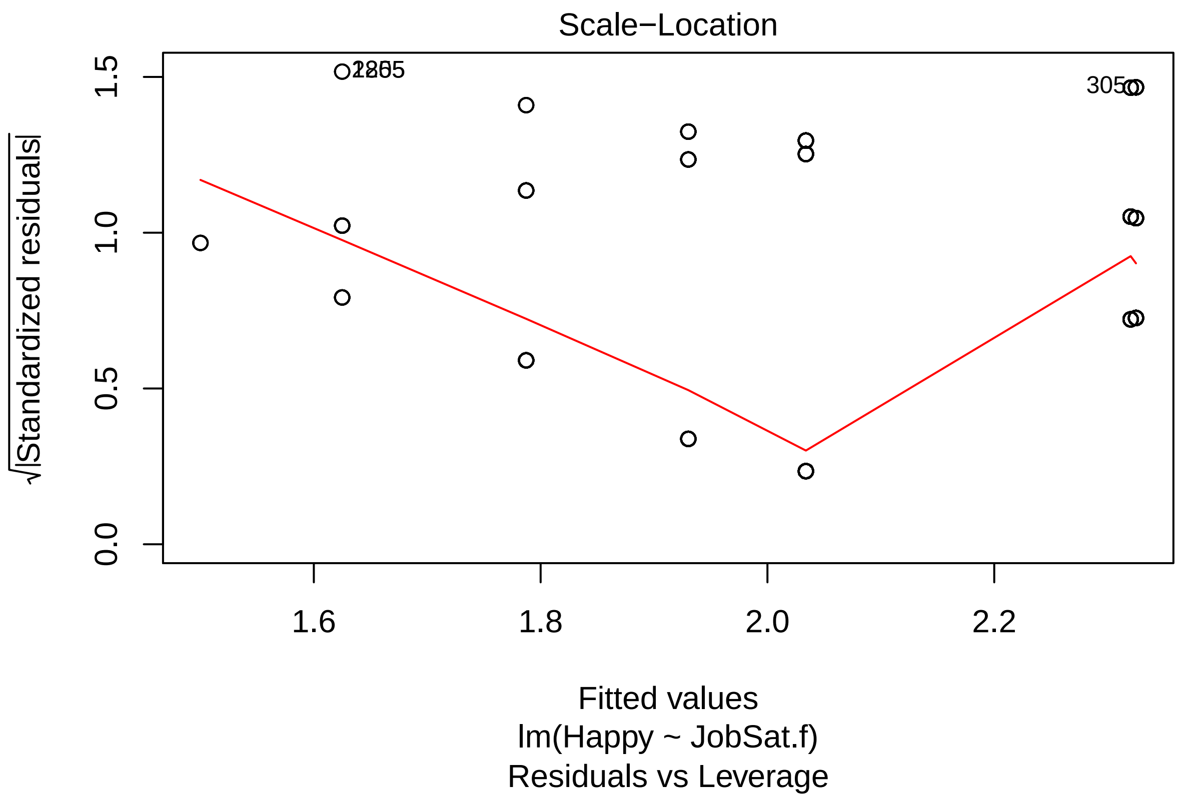

```r
#Marital is significant

Job <- lm(Happy ~ JobSat.f)
summary(Job)
```

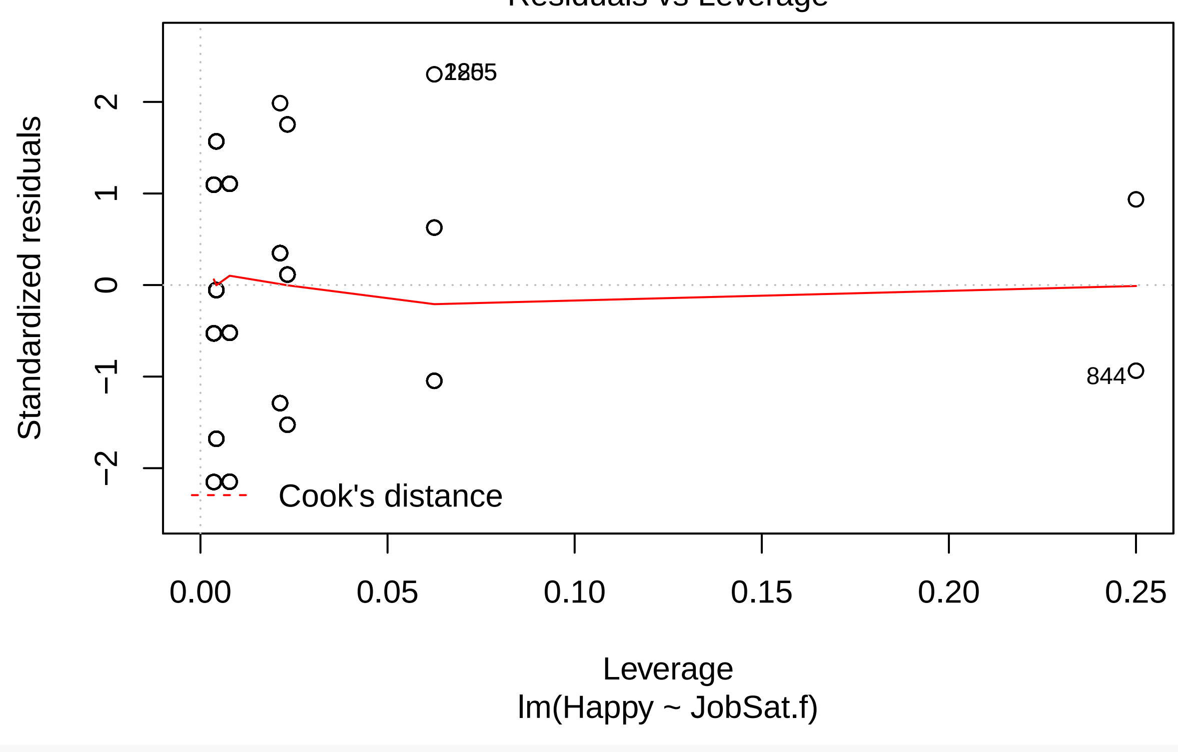
```
##
## Call:
## lm(formula = Happy ~ JobSat.f)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3250 -0.3250 -0.0339  0.6750  1.3750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.320313   0.054541  42.542  < 2e-16 ***
## JobSat.f2    0.004688   0.065838   0.071 0.943260
## JobSat.f3   -0.286414   0.067736  -4.228 2.65e-05 ***
## JobSat.f4   -0.390080   0.108765  -3.586 0.000357 ***
## JobSat.f5   -0.533078   0.105244  -5.065 5.15e-07 ***
## JobSat.f6   -0.695313   0.163624  -4.249 2.41e-05 ***
## JobSat.f7   -0.820313   0.313316  -2.618 0.009020 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6171 on 747 degrees of freedom
##   (1613 observations deleted due to missingness)
## Multiple R-squared:  0.09479,    Adjusted R-squared:  0.08752
## F-statistic: 13.04 on 6 and 747 DF,  p-value: 4.662e-14
```

```r
plot(Job)
```


Residuals vs Fitted
Residuals
1.5
1.0
0.5
−0.5
−1.5
1.6
1.8
2.0
2.2
Fitted values
lm(Happy ~ JobSat.f)
305


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(Happy ~ JobSat.f)
305



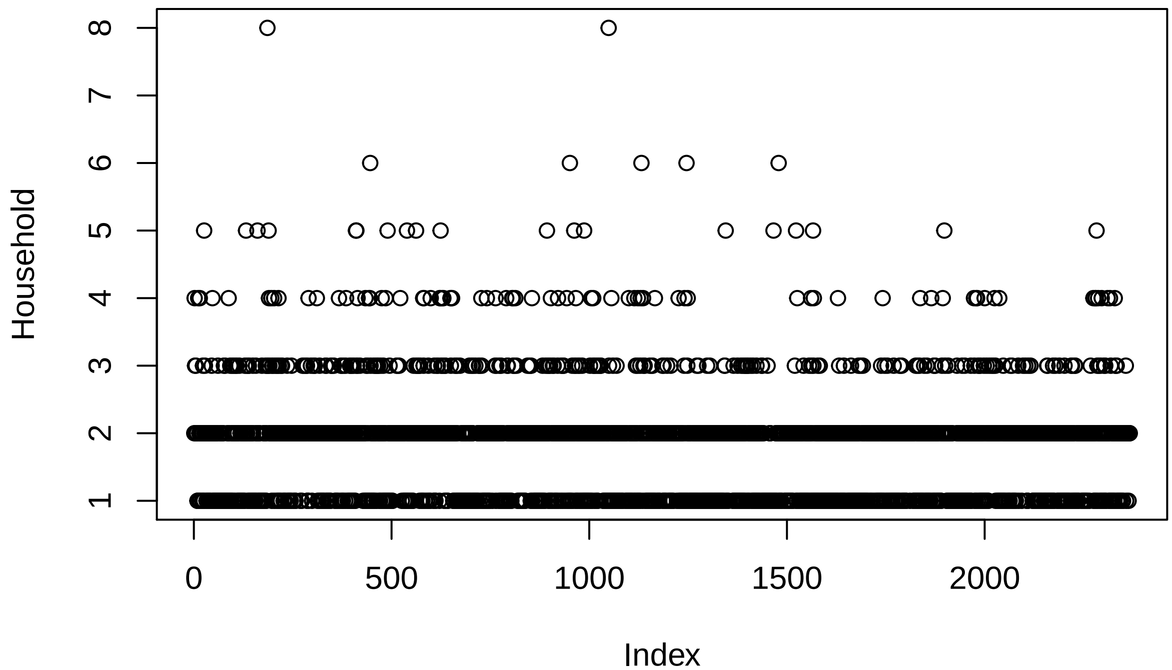



```
#Job is significant

House <- lm(happiness_data$Happy ~ happiness_data$Household)
summary(House)
```

```
##
```

```
## Call:
## lm(formula = happiness_data$Happy ~ happiness_data$Household)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5787 -0.1233 -0.1233  0.8008  0.9526
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.97153    0.03243   60.80  < 2e-16 ***
## happiness_data$Household  0.07590    0.01588    4.78 1.86e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.65 on 2358 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.009596,   Adjusted R-squared:  0.009176
## F-statistic: 22.85 on 1 and 2358 DF,  p-value: 1.862e-06
```

```r
plot(Household)
```

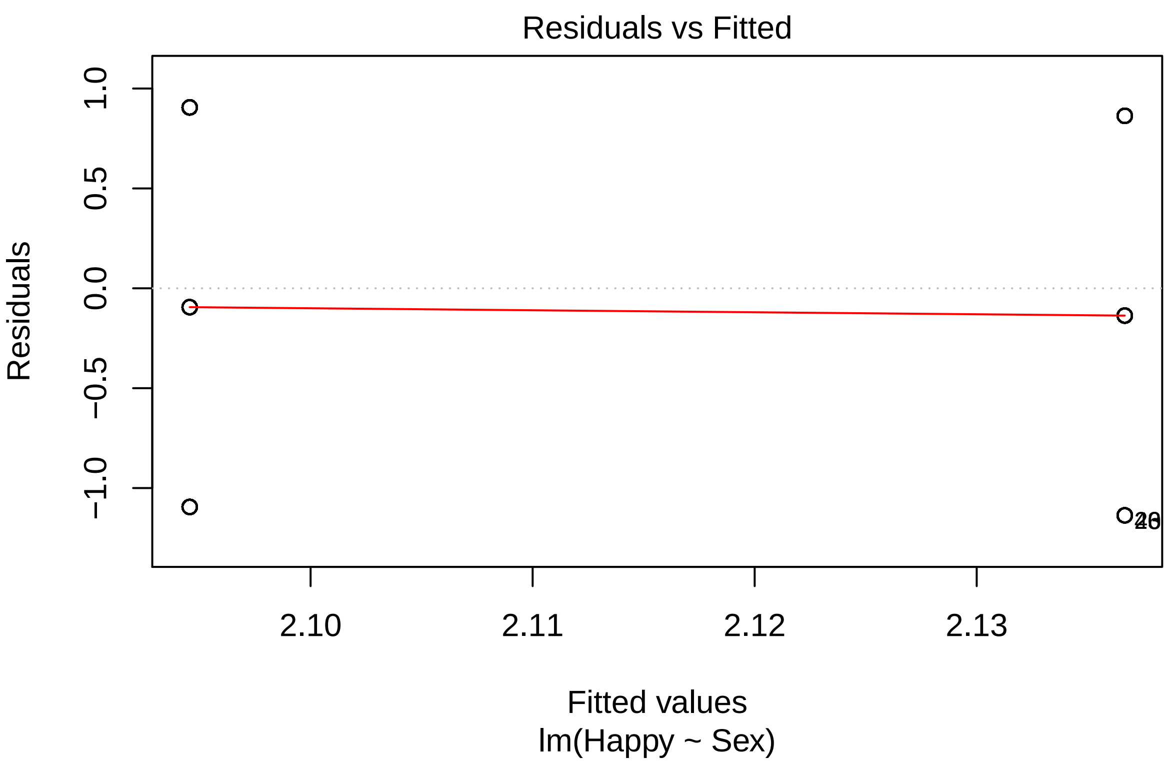


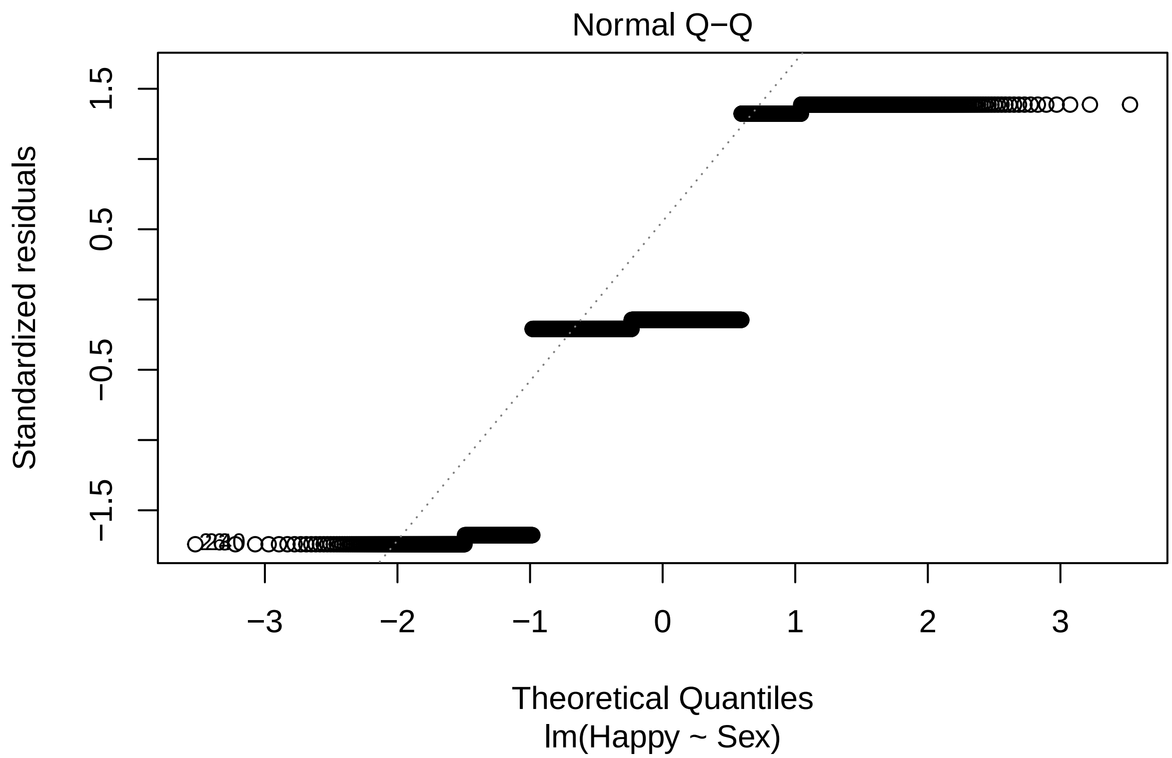

```r
#Household is significant

sex <- lm(Happy ~ Sex)
summary(sex)
```

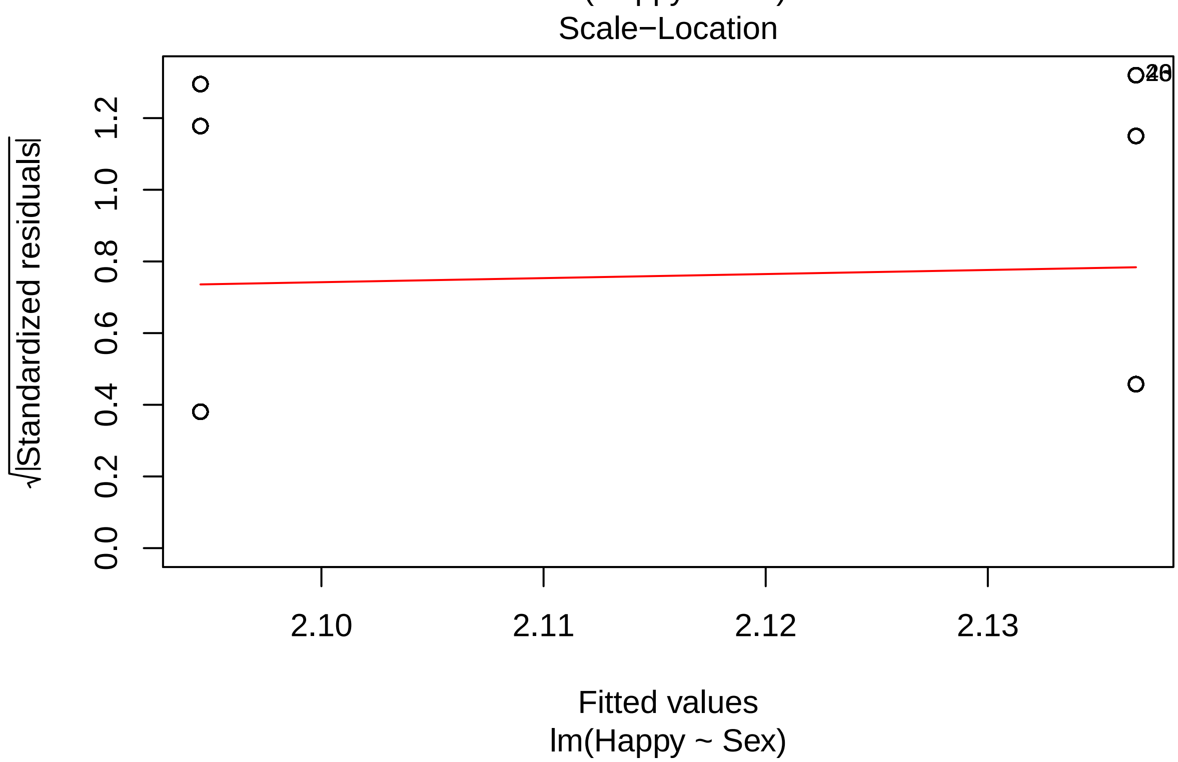
```
##
## Call:
## lm(formula = Happy ~ Sex)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.13667 -0.13667 -0.09455  0.86333  0.90545
##
```

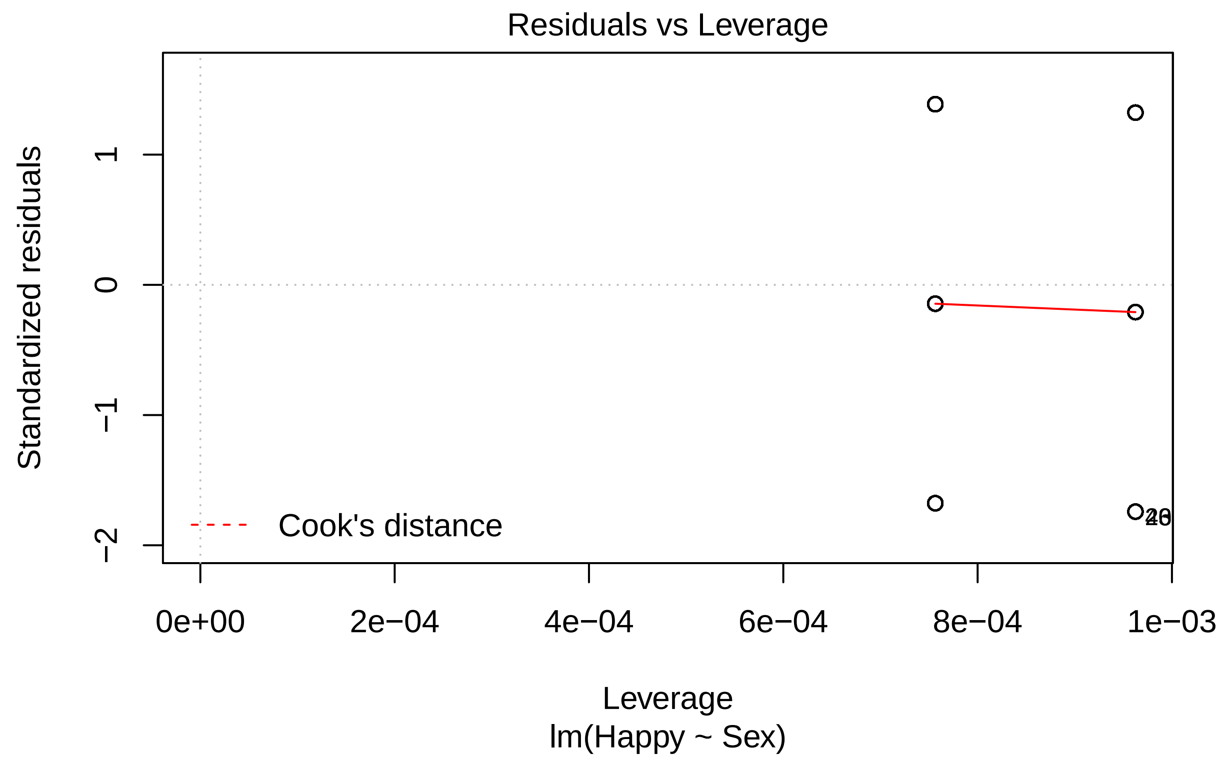
```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.17879    0.04432  49.165   <2e-16 ***
## Sex         -0.04212    0.02707  -1.556     0.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.653 on 2359 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.001025,   Adjusted R-squared:  0.0006015
## F-statistic:  2.42 on 1 and 2359 DF,  p-value: 0.1199
```

```r
plot(sex)
```




Normal Q−Q
Standardized residuals
−1.5
−0.5
0.5
1.5
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(Happy ~ Sex)


Scale−Location
√|Standardized residuals|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
2.10
2.11
2.12
2.13
Fitted values
lm(Happy ~ Sex)




```r
#Sex is insignificant
```

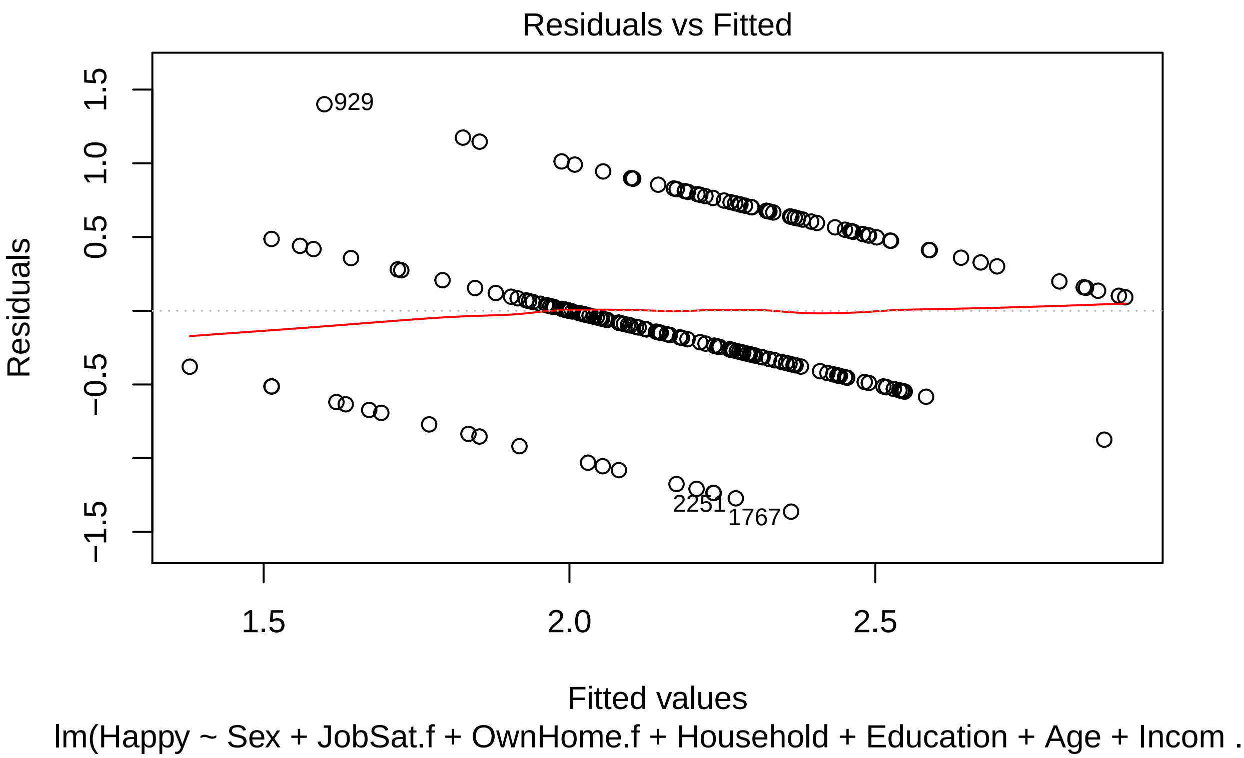
We then used partial F-tests to verify these findings, as well as potentially weed out other variables. To do so, we created models that each excluded one variable and then tested them against our full model. This method found Children and Education to be insignificant in addition to Instagram and Sex. OwnHome, JobSat, Income and Age threw errors in partial F-testing.

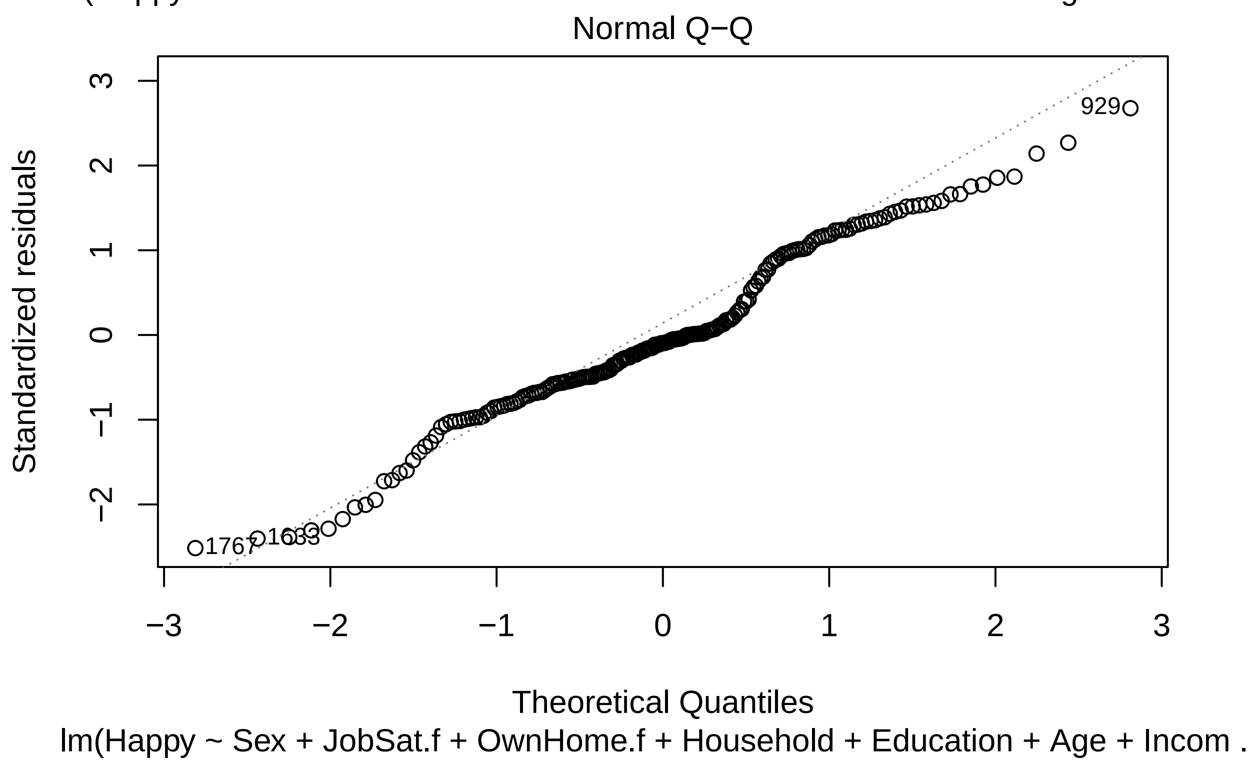
```r
noMarital <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Household + Education + Age + Income + Children + I
anova(full_model, noMarital)
```

```
## Analysis of Variance Table
##
## Model 1: Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children +
##     Education + JobSat.f + Income + Age + Sex
## Model 2: Happy ~ Sex + JobSat.f + OwnHome.f + Household + Education +
##     Age + Income + Children + Instagram.f
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    184 52.304
## 2    188 57.030 -4   -4.7259 4.1563 0.003016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

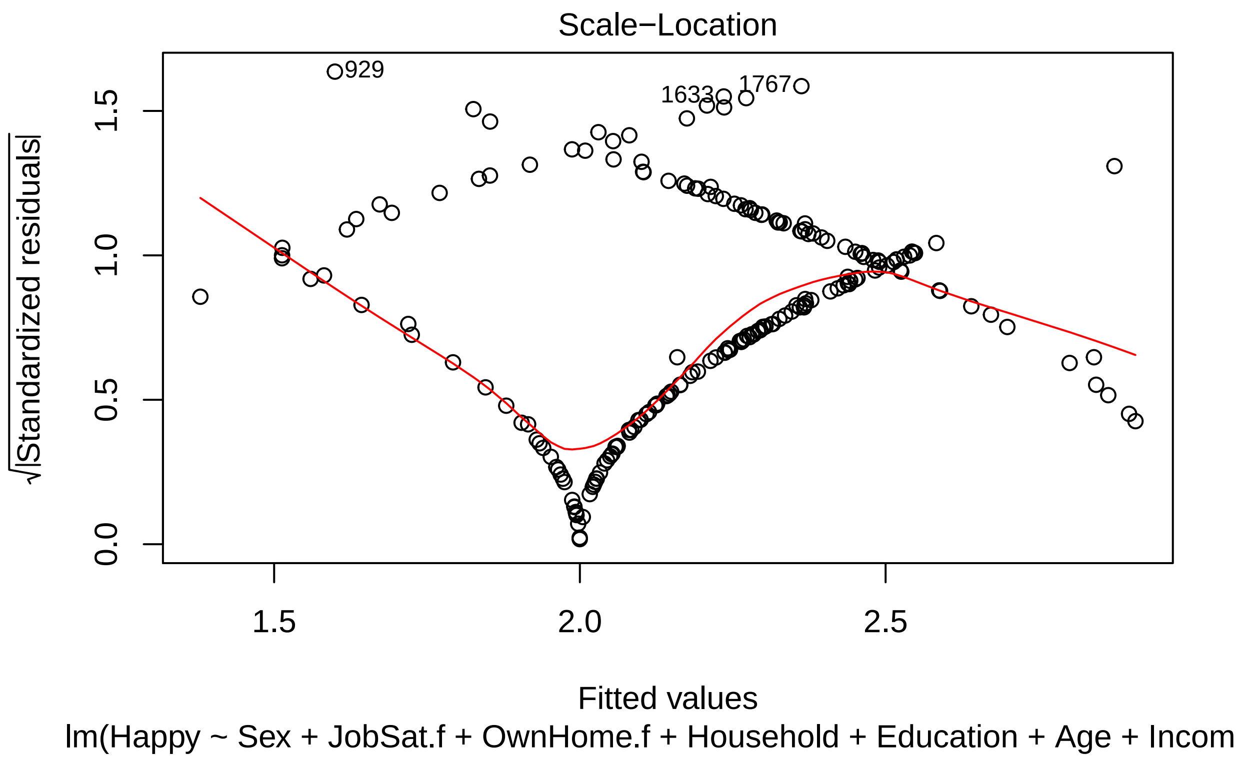
```r
#Marital is Significant
plot(noMarital)
```

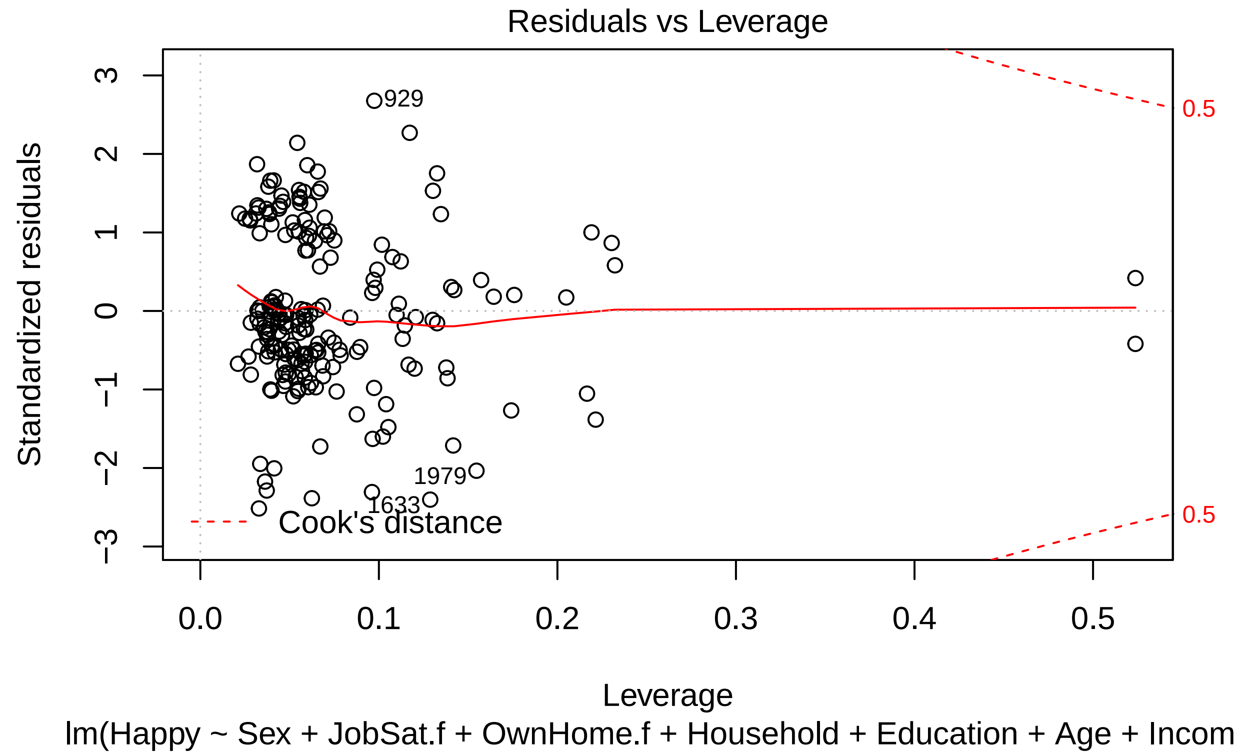
```
## Warning: not plotting observations with leverage one:
##   93
```






```
## Warning: not plotting observations with leverage one:
##   93
```






```
noHousehold <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Marital.f + Education + Age + Income + Children +
anova(full_model, noHousehold)
```

```
## Analysis of Variance Table
##
## Model 1: Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children +
```

```
##     Education + JobSat.f + Income + Age + Sex
## Model 2: Happy ~ Sex + JobSat.f + OwnHome.f + Marital.f + Education +
##     Age + Income + Children + Instagram.f
##   Res.Df    RSS Df Sum of Sq     F   Pr(>F)
## 1    184 52.304
## 2    185 54.737 -1   -2.4327 8.558 0.003873 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Household is significant

noSex <- lm(Happy ~ JobSat.f + OwnHome.f + Household + Marital.f + Education + Age + Income + Children +
anova(full_model, noSex)
```

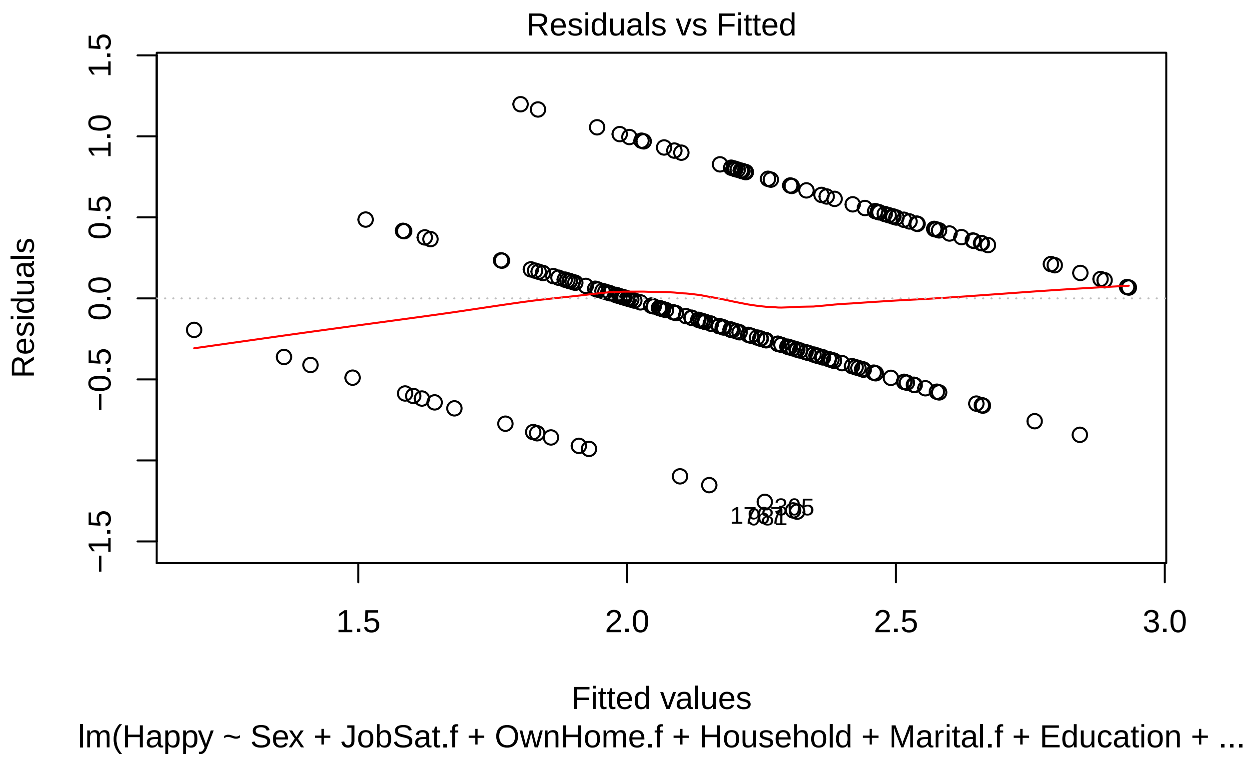
```
## Analysis of Variance Table
##
## Model 1: Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children +
##     Education + JobSat.f + Income + Age + Sex
## Model 2: Happy ~ JobSat.f + OwnHome.f + Household + Marital.f + Education +
##     Age + Income + Children + Instagram.f
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    184 52.304
## 2    185 52.696 -1  -0.39174 1.3781 0.2419
```

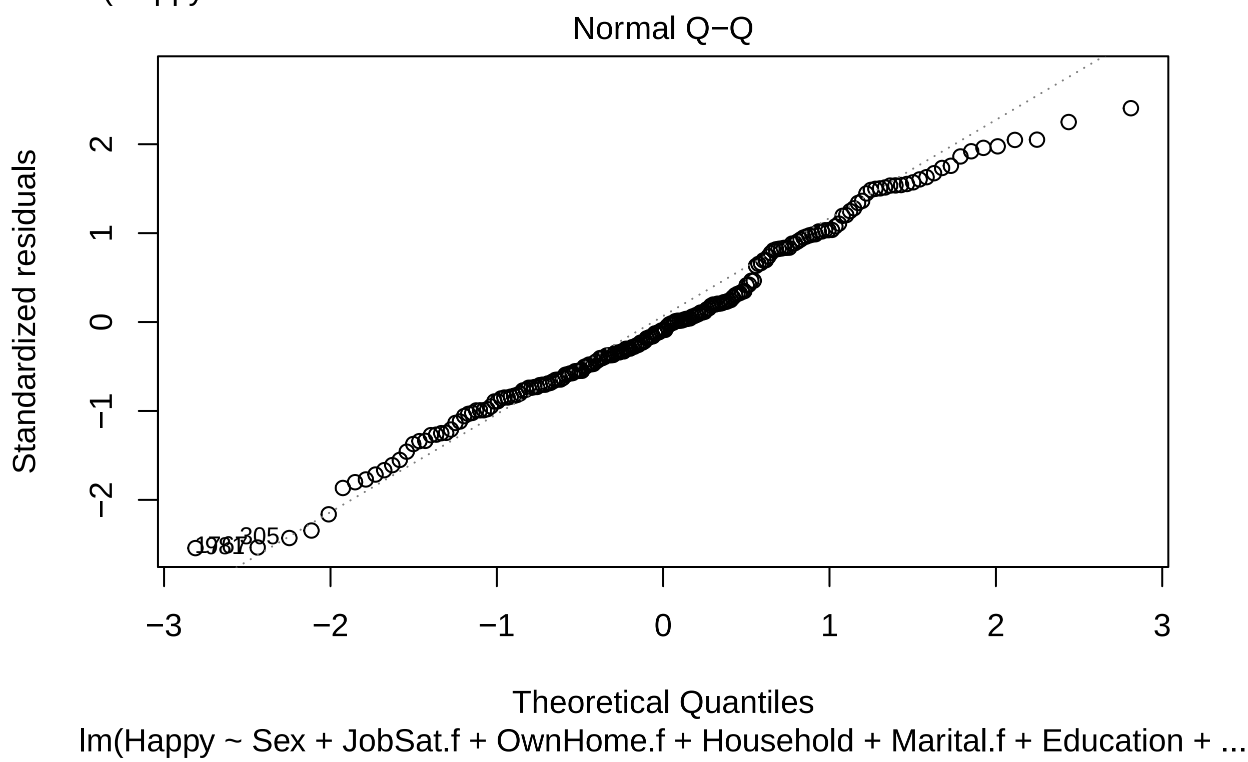
```
#Sex is insignificant

noInstagram <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f + Education + Age + Income
anova(full_model, noInstagram)
```

```
## Analysis of Variance Table
##
## Model 1: Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children +
##     Education + JobSat.f + Income + Age + Sex
## Model 2: Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f +
##     Education + Age + Income + Children
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
## 1    184 52.304
## 2    185 52.435 -1  -0.13103 0.461  0.498
```

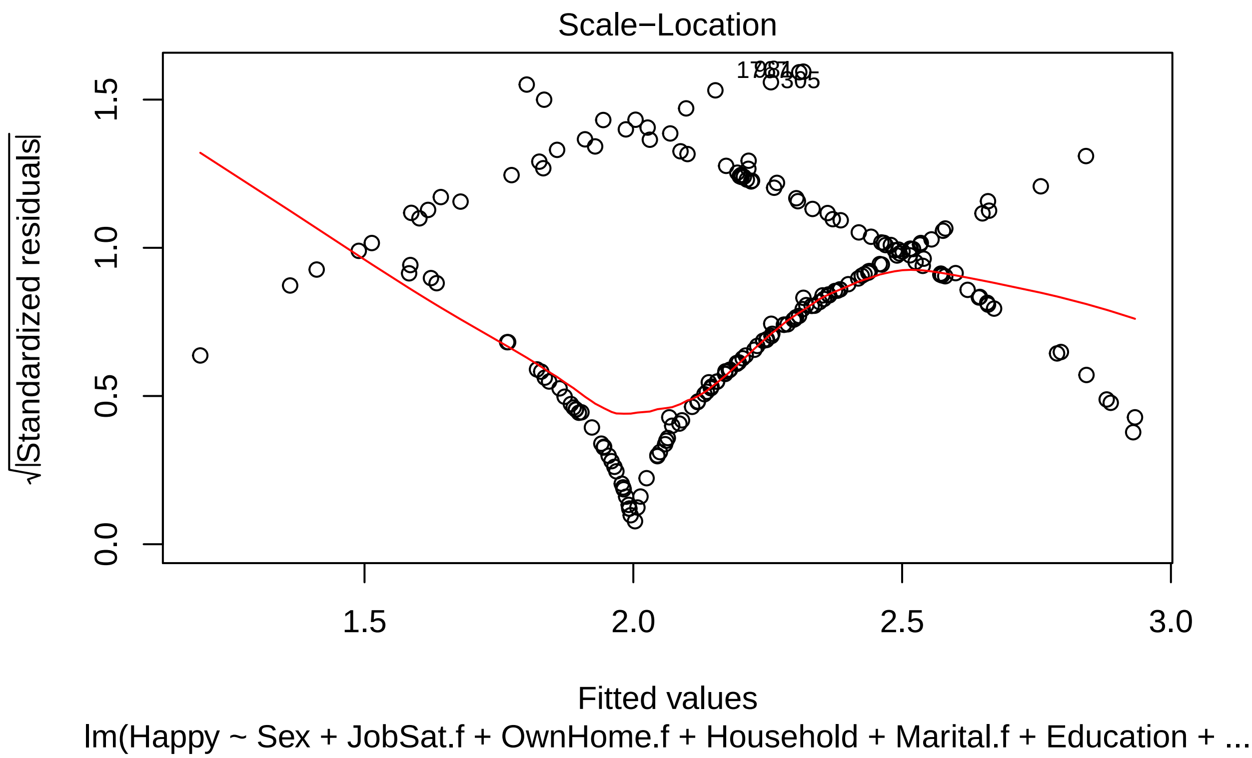
```
#Instagram is insignificant
plot(noInstagram)
```

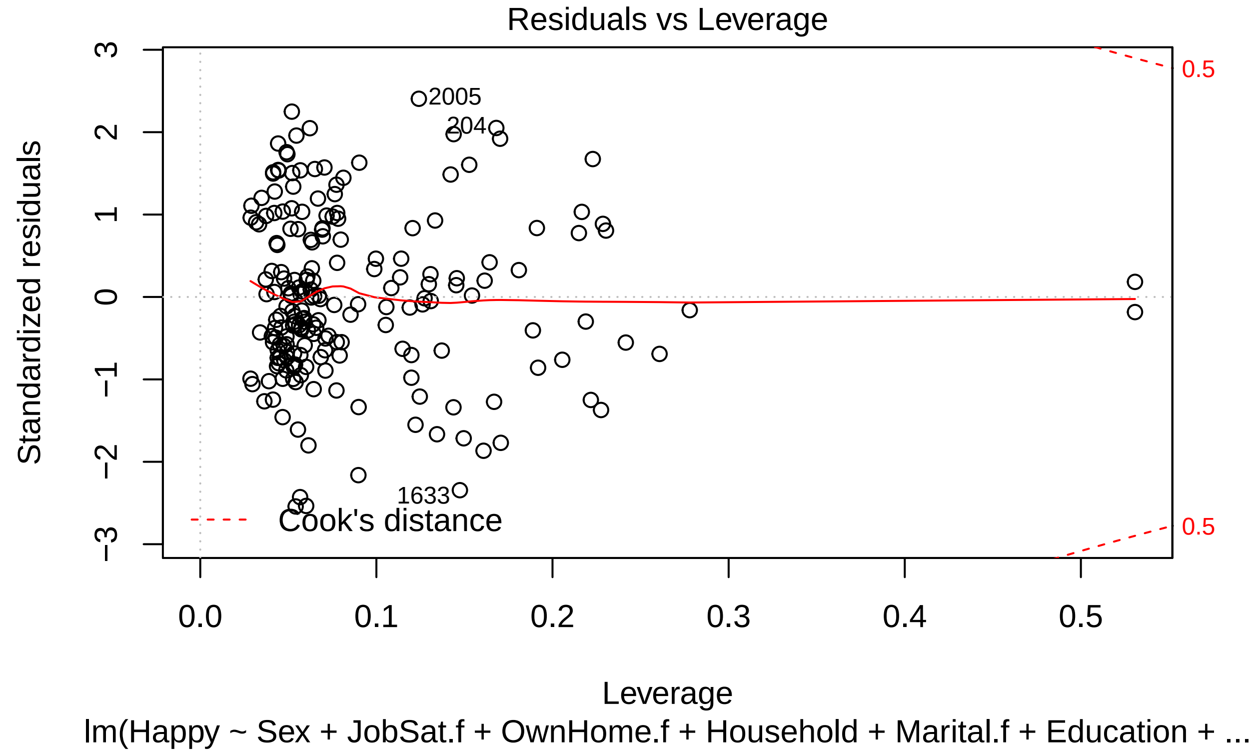
```
## Warning: not plotting observations with leverage one:
##   93
```






```
## Warning: not plotting observations with leverage one:
##   93
```






```
noChildren <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f + Education + Age + Income +
anova(full_model, noChildren)
```

```
## Analysis of Variance Table
##
## Model 1: Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children +
```

```
##     Education + JobSat.f + Income + Age + Sex
## Model 2: Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f +
##     Education + Age + Income + Instagram.f
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    184 52.304
## 2    185 52.478 -1  -0.17382 0.6115 0.4352
```

```
#Children is Insignificant

noEducation <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f + Age + Income + Children +
anova(full_model, noEducation)
```

```
## Analysis of Variance Table
##
## Model 1: Happy ~ Household + OwnHome.f + Instagram.f + Marital.f + Children +
##     Education + JobSat.f + Income + Age + Sex
## Model 2: Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f +
##     Age + Income + Children + Instagram.f
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    184 52.304
## 2    185 52.764 -1  -0.46014 1.6187 0.2049
```

```
#Education is insignificant

#noOwnHome <- lm(Happy ~ Sex + JobSat.f + Household + Marital.f + Education + Age + Income + Children +
#anova(full_model, noOwnHome)
#OwnHome Error

#noJobSat <- lm(Happy ~ Sex + OwnHome.f + Household + Marital.f + Education + Age + Income + Children +
#anova(full_model, noJobSat)
#JobSat Error
names(happiness_data)
```

```
##  [1] "Household" "Health"    "OwnHome"   "Instagram" "Marital"
##  [6] "Sex"       "Age"       "Children"  "Education" "JobSat"
## [11] "Income"    "WorkHrs"   "Happy"
```

```
#noAge <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f + Education  + Income + Childre
#anova(full_model, noAge)
#Age Error

#noIncome <- lm(Happy ~ Sex + JobSat.f + OwnHome.f + Household + Marital.f + Education + Age + Children
#anova(full_model, noIncome)
#Income Error
```

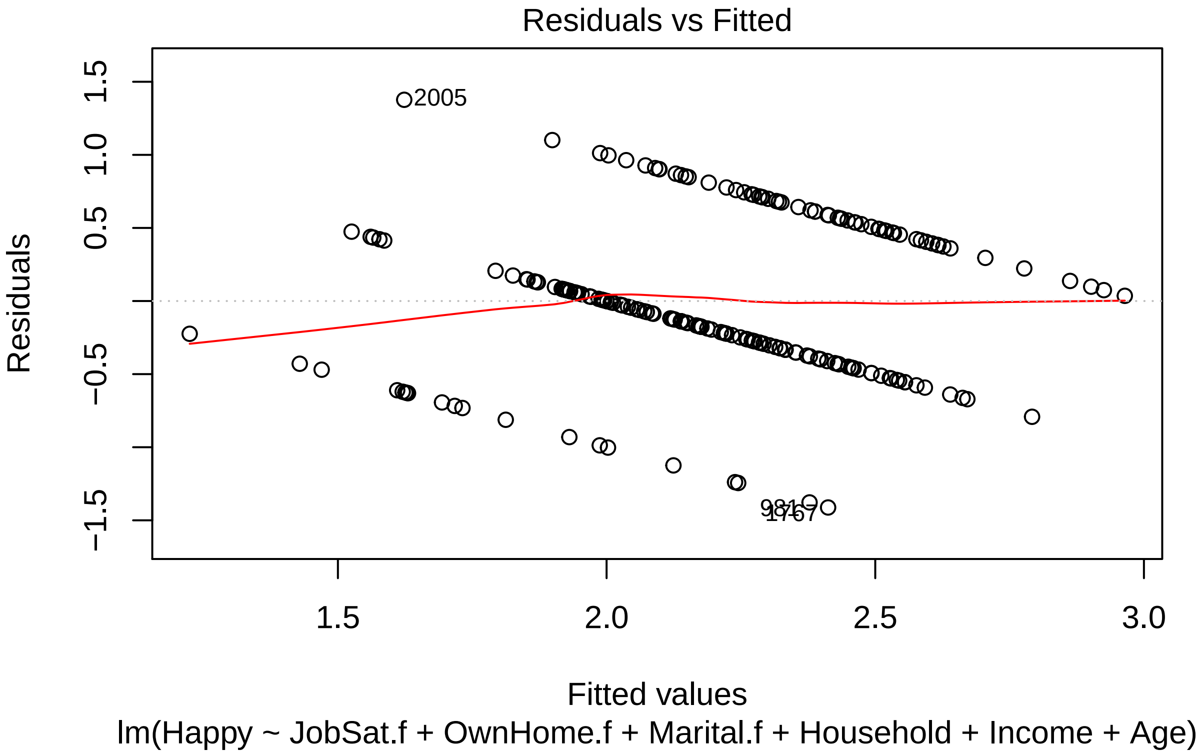
After eliminating the four insignificant variables, we obtained AIC, AICc and BIC values, which were lowest when all six variables were included. Performing forward selection showed OwnHome to be insignificant and performing backward selection showed Age to be insignificant. Since the forward and backward selections were not in agreement, this did not seem like strong enough evidence to exclude the variables to us. We found including all 6 variables gave us the lowest values for each, so we did not choose to omit any variables from the model in this process.

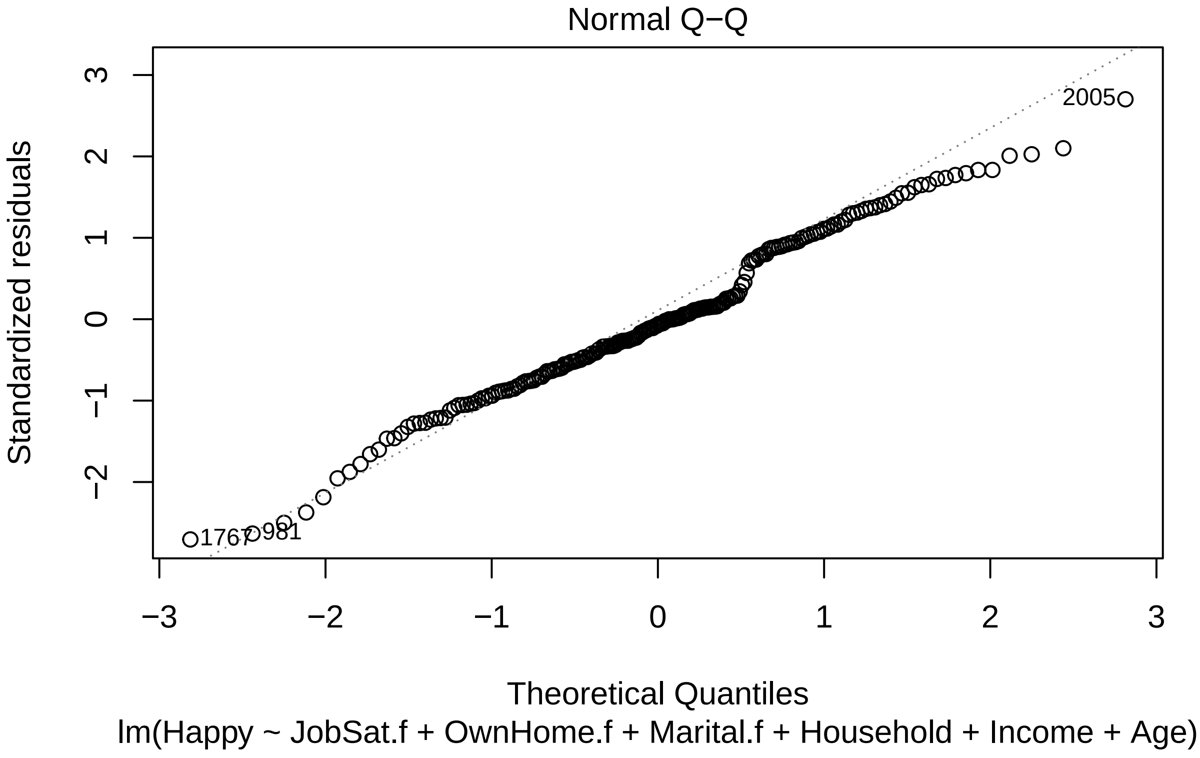
```
#Eliminating education, instagram, children, sex
new_model <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Income + Age)
summary(new_model)
```

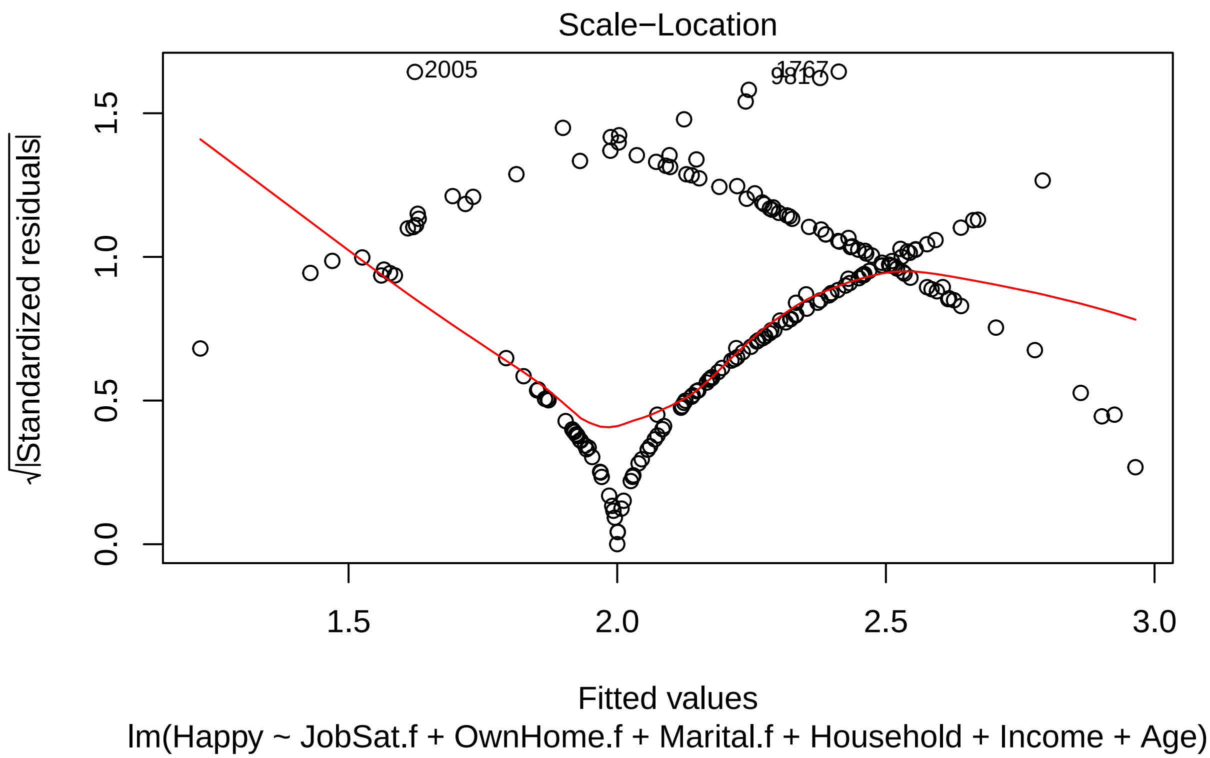
```
##
```

```
## Call:
## lm(formula = Happy ~ JobSat.f + OwnHome.f + Marital.f + Household +
##     Income + Age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41226 -0.33285 -0.03489  0.42291  1.37660
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.121e+00  2.521e-01  12.382  < 2e-16 ***
## JobSat.f2   -7.894e-02  1.132e-01  -0.698 0.486272
## JobSat.f3   -3.008e-01  1.203e-01  -2.501 0.013236 *
## JobSat.f4   -1.885e-01  1.981e-01  -0.952 0.342567
## JobSat.f5   -5.912e-01  1.759e-01  -3.361 0.000941 ***
## JobSat.f6   -8.437e-01  2.613e-01  -3.229 0.001466 **
## JobSat.f7   -3.303e-01  5.483e-01  -0.602 0.547597
## OwnHome.f2   2.905e-03  8.497e-02   0.034 0.972764
## OwnHome.f3   7.351e-01  3.997e-01   1.839 0.067463 .
## Marital.f2  -5.742e-01  2.075e-01  -2.767 0.006218 **
## Marital.f3  -2.642e-01  1.125e-01  -2.348 0.019892 *
## Marital.f4  -2.817e-01  2.516e-01  -1.119 0.264370
## Marital.f5  -3.381e-01  1.051e-01  -3.218 0.001518 **
## Household   -1.370e-01  4.622e-02  -2.965 0.003418 **
## Income       2.964e-06  1.423e-06   2.083 0.038589 *
## Age         -7.862e-03  3.453e-03  -2.277 0.023908 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.535 on 188 degrees of freedom
##   (2163 observations deleted due to missingness)
## Multiple R-squared:  0.2604, Adjusted R-squared:  0.2014
## F-statistic: 4.412 on 15 and 188 DF,  p-value: 4.385e-07
```

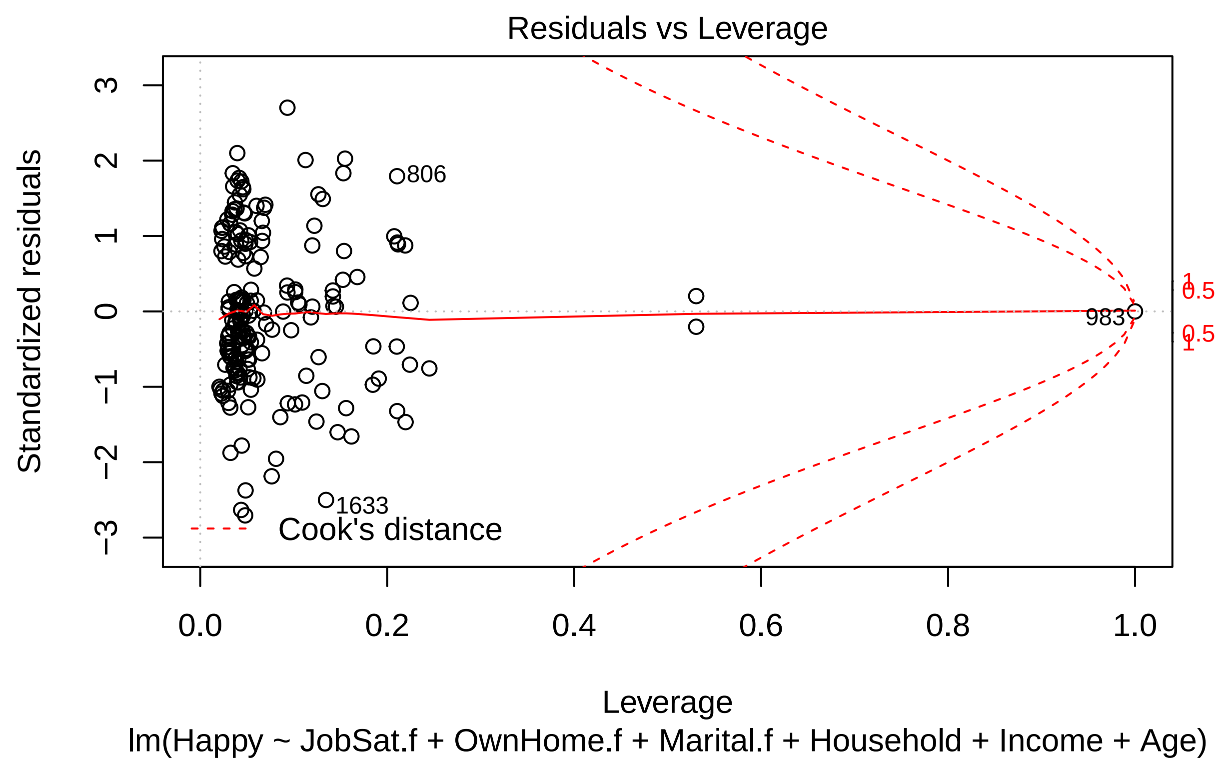
```r
plot(new_model)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Income + Age)
2005
981
1767
1.5
1.0
0.5
−0.5
−1.5
1.5
2.0
2.5
3.0


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Income + Age)
2005
1767
981
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3




```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```r
Rad <- summary(new_model)$adj.r.squared
Rad
```

```
## [1] 0.2013563
```

```r
om1 <- lm(Happy ~ JobSat.f)
om2 <- lm(Happy ~ JobSat.f + OwnHome.f)
om3 <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f)
om4 <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household)
om5 <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Age)
om6 <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Age + Income)
n = length(Happy)


#### subset size = 1 ####
p <- 1
oms1 <- summary(om1)
# AIC
AIC1 <- extractAIC(om1,k=2)[2]
# AICc
AICc1 <- extractAIC(om1,k=2)[2]+2*(p+2)*(p+3)/(n-p-1)
# BIC
BIC1 <- extractAIC(om1,k=log(n))[2]

#### subset size = 2 ####
p <- 2
oms2 <- summary(om2)
# AIC
AIC2 <- extractAIC(om2,k=2)[2]
# AICc
AICc2 <- extractAIC(om2,k=2)[2]+2*(p+2)*(p+3)/(n-p-1)
# BIC
BIC2 <- extractAIC(om2,k=log(n))[2]

#### subset size = 3 ####
p <- 3
oms3 <- summary(om3)
# AIC
AIC3 <- extractAIC(om3,k=2)[2]
# AICc
AICc3 <- extractAIC(om3,k=2)[2]+2*(p+2)*(p+3)/(n-p-1)
# BIC
BIC3 <- extractAIC(om3,k=log(n))[2]

#### subset size = 4 ####
p <- 4
oms4 <- summary(om4)
# AIC
AIC4 <- extractAIC(om4,k=2)[2]
# AICc
AICc4 <- extractAIC(om4,k=2)[2]+2*(p+2)*(p+3)/(n-p-1)
# BIC
BIC4 <- extractAIC(om4,k=log(n))[2]

#### subset size = 5 ####
p <- 5
oms5 <- summary(om5)
```

```r
# AIC
AIC5 <- extractAIC(om5,k=2)[2]
# AICc
AICc5 <- extractAIC(om5,k=2)[2]+2*(p+2)*(p+3)/(n-p-1)
# BIC
BIC5 <- extractAIC(om5,k=log(n))[2]

#### subset size = 6 ####
p <- 6
oms6 <- summary(om6)
# AIC
AIC6 <- extractAIC(om6,k=2)[2]
# AICc
AICc6 <- extractAIC(om6,k=2)[2]+2*(p+2)*(p+3)/(n-p-1)
# BIC
BIC6 <- extractAIC(om6,k=log(n))[2]

## Answer
AIC <- c(AIC1,AIC2,AIC3, AIC4, AIC5, AIC6)
AICc <- c(AICc1, AICc2, AICc3, AICc4, AICc5, AICc6)
BIC <- c(BIC1, BIC2, BIC3, BIC4, BIC5, BIC6)

opmodel <- data.frame(Size=1:6, Radj2= Rad, AIC=AIC, AICc=AICc, BIC=BIC)
opmodel
```

```
##   Size     Radj2       AIC      AICc       BIC
## 1    1 0.2013563 -721.0678 -721.0577 -680.6822
## 2    2 0.2013563 -286.4328 -286.4159 -234.5084
## 3    3 0.2013563 -293.8557 -293.8303 -218.8538
## 4    4 0.2013563 -298.4634 -298.4278 -217.6921
## 5    5 0.2013563 -296.6381 -296.5906 -210.0974
## 6    6 0.2013563 -239.8424 -239.7814 -147.5323
```

```r
#Lowest AIC, AICc, BIC values occur when size = 6. Thus, we are retaining all variables

#Checking Forward Selection

add1(lm(Happy~1), Happy~ JobSat.f + OwnHome.f + Marital.f + Household + Age + Income, test="F")
```

```
## Warning in add1.lm(lm(Happy ~ 1), Happy ~ JobSat.f + OwnHome.f + Marital.f
## + : using the 204/2361 rows from a combined fit

## Single term additions
##
## Model:
## Happy ~ 1
##           Df Sum of Sq    RSS     AIC  F value    Pr(>F)
## <none>                 72.760 -208.31
## JobSat.f   6   10.1623 62.597 -227.00  63.6928 < 2.2e-16 ***
## OwnHome.f  2    0.5933 72.166 -205.99   9.6936 6.417e-05 ***
## Marital.f  4    6.0739 66.686 -218.10  53.6470 < 2.2e-16 ***
## Household  1    0.9268 71.833 -208.93  30.4377 3.824e-08 ***
## Age        1    0.0199 72.740 -206.37   0.6470    0.4213
## Income     1    4.1472 68.613 -218.29 142.5867 < 2.2e-16 ***
## ---
```

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

```r
#Age seems to be insignificant

#Performing another forward selection to see if age is actually insignificant
add1(lm(Happy ~ JobSat.f), Happy~ JobSat.f + OwnHome.f + Marital.f + Household + Age + Income, test="F")
```

```
## Warning in add1.lm(lm(Happy ~ JobSat.f), Happy ~ JobSat.f + OwnHome.f + :
## using the 204/754 rows from a combined fit

## Single term additions
##
## Model:
## Happy ~ JobSat.f
##           Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                 62.597 -227.00
## OwnHome.f  2    0.5230 62.075 -224.72  3.1383  0.043929 *
## Marital.f  4    3.8999 58.698 -232.13 12.3415 1.010e-09 ***
## Household  1    0.8328 61.765 -227.74 10.0581  0.001579 **
## Age        1    0.0335 62.564 -225.11  0.3993  0.527644
## Income     1    2.0339 60.564 -231.74 25.0533 6.963e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Backward Selection to see check the significance of variables
drop1(lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Age + Income), test="F")
```

```
## Single term deletions
##
## Model:
## Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Age +
##     Income
##           Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                 53.815 -239.84
## JobSat.f   6    6.9053 60.721 -227.22  4.0205 0.0008187 ***
## OwnHome.f  2    0.9689 54.784 -240.20  1.6925 0.1868556
## Marital.f  4    4.6015 58.417 -231.10  4.0188 0.0037590 **
## Household  1    2.5168 56.332 -232.52  8.7921 0.0034175 **
## Age        1    1.4842 55.300 -236.29  5.1850 0.0239082 *
## Income     1    1.2422 55.058 -237.19  4.3396 0.0385887 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Ownhome seems to be insignificant
drop1(lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household  + Income), test="F")
```

```
## Single term deletions
##
## Model:
## Happy ~ JobSat.f + OwnHome.f + Marital.f + Household + Income
##           Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                 55.621 -239.72
## JobSat.f   6    6.8734 62.494 -227.72  3.9338 0.0009888 ***
## OwnHome.f  2    0.5914 56.212 -241.54  1.0154 0.3642221
## Marital.f  4    3.7442 59.365 -234.30  3.2144 0.0139487 *
## Household  1    1.7410 57.362 -235.37  5.9786 0.0153868 *
## Income     1    1.0130 56.634 -238.00  3.4785 0.0637063 .
```

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

Our untransformed model thus contained JobSat, OwnHome, Marital and Household in factor form, as well as the numerical variables Income and Age. Its R squared value was 0.2844, and its adjusted R squared value improved to 0.2105.

```r
Household.f <- factor(Household)

Final_model <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household.f + Income + Age)
summary(Final_model)
```

```
##
## Call:
## lm(formula = Happy ~ JobSat.f + OwnHome.f + Marital.f + Household.f +
##     Income + Age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42166 -0.32096 -0.04902  0.44364  1.19129
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.906e+00  2.529e-01  11.493  < 2e-16 ***
## JobSat.f2     -1.066e-01  1.133e-01  -0.942 0.347631
## JobSat.f3     -3.311e-01  1.219e-01  -2.715 0.007250 **
## JobSat.f4     -2.309e-01  2.020e-01  -1.143 0.254575
## JobSat.f5     -6.119e-01  1.768e-01  -3.461 0.000670 ***
## JobSat.f6     -8.561e-01  2.611e-01  -3.279 0.001247 **
## JobSat.f7     -3.018e-01  5.498e-01  -0.549 0.583697
## OwnHome.f2     4.395e-02  8.657e-02   0.508 0.612284
## OwnHome.f3     4.814e-01  4.137e-01   1.164 0.246017
## Marital.f2    -6.542e-01  2.162e-01  -3.025 0.002838 **
## Marital.f3    -2.779e-01  1.284e-01  -2.164 0.031721 *
## Marital.f4    -3.082e-01  2.556e-01  -1.206 0.229406
## Marital.f5    -3.991e-01  1.216e-01  -3.283 0.001230 **
## Household.f2 -1.501e-01  1.124e-01  -1.335 0.183514
## Household.f3 -3.395e-02  1.515e-01  -0.224 0.822903
## Household.f4 -7.714e-01  2.260e-01  -3.414 0.000787 ***
## Household.f5 -5.021e-01  4.074e-01  -1.232 0.219435
## Household.f6 -6.439e-01  3.922e-01  -1.642 0.102308
## Income        2.929e-06  1.435e-06   2.041 0.042641 *
## Age          -5.477e-03  3.593e-03  -1.525 0.129074
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5319 on 184 degrees of freedom
##   (2163 observations deleted due to missingness)
## Multiple R-squared:  0.2844, Adjusted R-squared:  0.2105
## F-statistic:  3.85 on 19 and 184 DF,  p-value: 7.865e-07
```

We used both the Box Cox and inverse response plot methods in transforming our model. Per Box Cox, we raised Age to the power of 0.226. Intuitively, looking at the relationship between Income and Happy, we decided on a logarithmic transformation on Income, as well. The adjusted R squared value reduced a little from this transformation, to 0.2099, and R squared dropped to 0.2838. Taking into account the inverse response plot function's suggestion - raising Happy to the power of -0.1130549 - improved the model a good amount - we obtained a value of 0.3116 for R squared and 0.2405 for adjusted R squared.

```r
#Box Cox transformation
summary(powerTransform(cbind(JobSat.f, OwnHome.f, Marital.f, Household.f, Income, Age)~1))
```

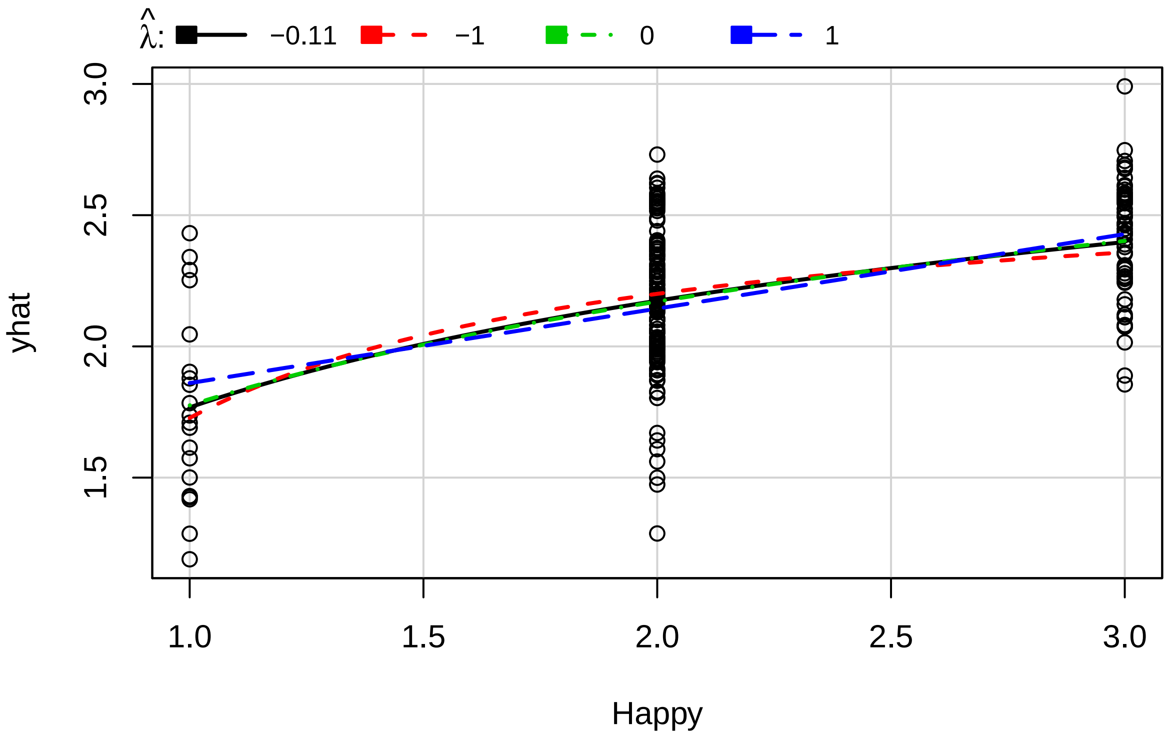
```
## bcPower Transformations to Multinormality
##             Est Power Rounded Pwr Wald Lwr bnd Wald Upr Bnd
## JobSat.f       0.1523        0.00      -0.0976       0.4023
## OwnHome.f     -1.9516       -2.00      -2.6378      -1.2653
## Marital.f      0.2230        0.00      -0.0528       0.4988
## Household.f   -0.2118        0.00      -0.5046       0.0810
## Income         0.2161        0.22       0.1285       0.3037
## Age            0.4522        0.50       0.0432       0.8611
##
## Likelihood ratio tests about transformation parameters
##                                  LRT df         pval
## LR test, lambda = (0 0 0 0 0 0)  69.98314  6 4.121148e-13
## LR test, lambda = (1 1 1 1 1 1) 475.04275  6 0.000000e+00
```

```r
Age_transformed <- Age^0.226

m_new <- lm(Happy ~ JobSat.f + OwnHome.f + Marital.f + Household.f + log(Income) + Age_transformed)
summary(m_new)
```

```
##
## Call:
## lm(formula = Happy ~ JobSat.f + OwnHome.f + Marital.f + Household.f +
##     log(Income) + Age_transformed)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4314 -0.3531 -0.0527  0.4253  1.1449
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.24357    0.78842   4.114 5.86e-05 ***
## JobSat.f2       -0.12561    0.11331  -1.109 0.269066
## JobSat.f3       -0.35845    0.12025  -2.981 0.003264 **
## JobSat.f4       -0.25906    0.20146  -1.286 0.200074
## JobSat.f5       -0.64401    0.17582  -3.663 0.000326 ***
## JobSat.f6       -0.88144    0.26121  -3.374 0.000902 ***
## JobSat.f7       -0.34443    0.55111  -0.625 0.532761
## OwnHome.f2       0.04759    0.08717   0.546 0.585813
## OwnHome.f3       0.47494    0.41535   1.143 0.254337
## Marital.f2      -0.65096    0.21645  -3.008 0.003003 **
## Marital.f3      -0.25990    0.12821  -2.027 0.044099 *
## Marital.f4      -0.29088    0.25699  -1.132 0.259160
## Marital.f5      -0.41117    0.12248  -3.357 0.000957 ***
## Household.f2    -0.15100    0.11243  -1.343 0.180930
## Household.f3    -0.01839    0.15829  -0.116 0.907663
## Household.f4    -0.73994    0.22859  -3.237 0.001433 **
## Household.f5    -0.50310    0.40781  -1.234 0.218904
## Household.f6    -0.65948    0.39219  -1.682 0.094354 .
## log(Income)      0.07454    0.03941   1.891 0.060132 .
## Age_transformed -0.51797    0.29468  -1.758 0.080460 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5322 on 184 degrees of freedom
##   (2163 observations deleted due to missingness)
## Multiple R-squared:  0.2838, Adjusted R-squared:  0.2099
## F-statistic: 3.838 on 19 and 184 DF,  p-value: 8.361e-07
```

```
#Inverse response plot
inverseResponsePlot(m_new, key=TRUE)
```




```
##       lambda      RSS
## 1 -0.1130549 14.45109
## 2 -1.0000000 14.66344
## 3  0.0000000 14.45491
## 4  1.0000000 14.78982
```

```
m_new <- lm(Happy^-0.1130549 ~ JobSat.f + OwnHome.f + Marital.f + Household.f + log(Income) + Age_trans

summary(m_new)
```

```
##
## Call:
## lm(formula = Happy^-0.1130549 ~ JobSat.f + OwnHome.f + Marital.f +
##     Household.f + log(Income) + Age_transformed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.05731 -0.01830  0.00069  0.01517  0.09148
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      0.865687   0.041844  20.688  < 2e-16 ***
## JobSat.f2        0.006675   0.006014   1.110 0.268444    
## JobSat.f3        0.016687   0.006382   2.615 0.009671 ** 
## JobSat.f4        0.013269   0.010692   1.241 0.216188    
## JobSat.f5        0.036356   0.009332   3.896 0.000137 ***
## JobSat.f6        0.052138   0.013864   3.761 0.000227 ***
## JobSat.f7        0.013550   0.029249   0.463 0.643734    
## OwnHome.f2      -0.002581   0.004627  -0.558 0.577581    
## OwnHome.f3      -0.025905   0.022044  -1.175 0.241467    
## Marital.f2       0.043936   0.011488   3.825 0.000179 ***
## Marital.f3       0.016523   0.006805   2.428 0.016136 *  
## Marital.f4       0.013747   0.013639   1.008 0.314823    
## Marital.f5       0.019749   0.006500   3.038 0.002727 ** 
## Household.f2     0.010611   0.005967   1.778 0.077024 .  
## Household.f3     0.004234   0.008401   0.504 0.614863    
## Household.f4     0.049489   0.012132   4.079 6.72e-05 ***
## Household.f5     0.024782   0.021644   1.145 0.253710    
## Household.f6     0.043312   0.020815   2.081 0.038833 *  
## log(Income)     -0.003665   0.002091  -1.752 0.081410 .  
## Age_transformed  0.024886   0.015640   1.591 0.113278    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02824 on 184 degrees of freedom
##   (2163 observations deleted due to missingness)
## Multiple R-squared:  0.3116, Adjusted R-squared:  0.2405
## F-statistic: 4.383 on 19 and 184 DF,  p-value: 4.703e-08
```