Appendix

Explor data

Top countries:

```
##       Country        Region Happiness.Rank Happiness.Score
## 1 Switzerland Western Europe              1           7.587
## 2     Iceland Western Europe              2           7.561

##       Country        Region Happiness.Rank Happiness.Score
## 1     Denmark Western Europe              1           7.526
## 2 Switzerland Western Europe              2           7.509

##   Country        Region Happiness.Rank Happiness.Score
## 1  Norway Western Europe              1           7.537
## 2 Denmark Western Europe              2           7.522

##   Country        Region Happiness.Rank Happiness.Score
## 1 Finland Western Europe              1           7.632
## 2  Norway Western Europe              2           7.594

##   Country        Region Happiness.Rank Happiness.Score
## 1 Finland Western Europe              1           7.769
## 2 Denmark Western Europe              2           7.600

##   Country        Region Happiness.Rank Happiness.Score
## 1 Finland Western Europe              1          7.8087
## 2 Denmark Western Europe              2          7.6456

##   Country        Region Happiness.Rank Happiness.Score
## 1 Finland Western Europe              1           7.842
## 2 Denmark Western Europe              2           7.620

##   Country        Region Happiness.Rank Happiness.Score
## 1 Finland Western Europe              1           7.821
## 2 Denmark Western Europe              2           7.636
```

Finland is the happiest country from 2018 until now.

Bottom countries:

```
##       Country                    Region Happiness.Rank Happiness.Score
## 157 Burundi Sub-Saharan Africa                       5           2.905
## 158    Togo Sub-Saharan Africa                       5           2.839

##       Country                        Region Happiness.Rank Happiness.Score
## 156   Syria Middle East and North Africa               156           3.069
## 157 Burundi            Sub-Saharan Africa               157           2.905

##                      Country             Region Happiness.Rank Happiness.Score
## 154                  Burundi Sub-Saharan Africa            154           2.905
## 155 Central African Republic Sub-Saharan Africa            155           2.693
```

```
##                         Country                 Region Happiness.Rank Happiness.Score
## 155 Central African Republic Sub-Saharan Africa                155           3.083
## 156                  Burundi Sub-Saharan Africa                156           2.905
##                         Country                 Region Happiness.Rank Happiness.Score
## 155 Central African Republic Sub-Saharan Africa                155           3.083
## 156              South Sudan Sub-Saharan Africa                156           2.853
##         Country             Region Happiness.Rank Happiness.Score
## 152 South Sudan Sub-Saharan Africa            152          2.8166
## 153 Afghanistan         South Asia            153          2.5669
##         Country             Region Happiness.Rank Happiness.Score
## 148    Zimbabwe Sub-Saharan Africa            148           3.145
## 149 Afghanistan         South Asia            149           2.523
##         Country                       Region Happiness.Rank Happiness.Score
## 145     Lebanon Middle East and North Africa            145           2.955
## 146 Afghanistan                   South Asia            146           2.404
```
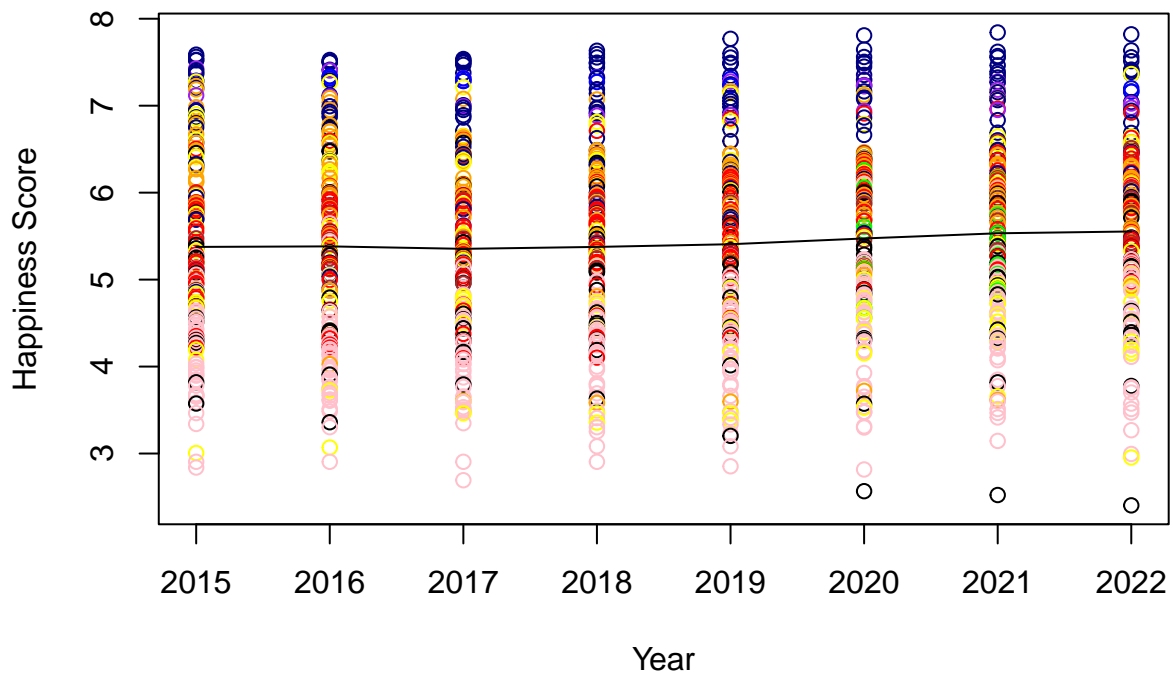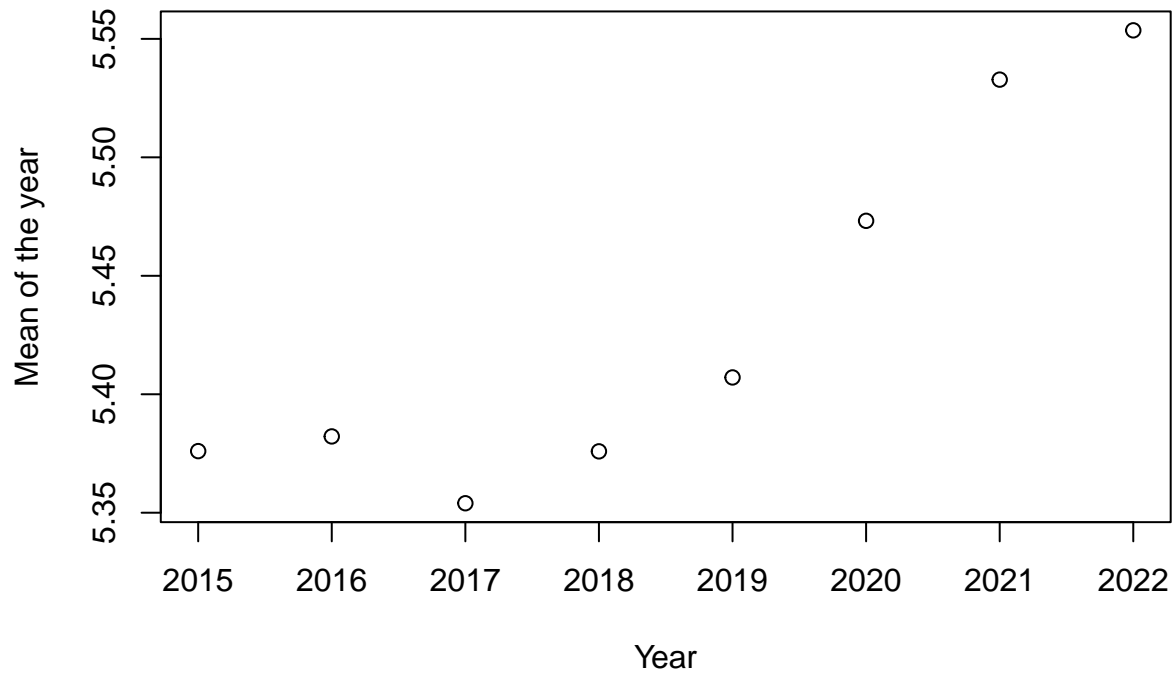
Burundi was in the saddest two countries from 2015 to 2018. From 2020 to now Afghanistan is the saddest country.


Happiness Scores across years

Happiness mean

We see there are no significant difference between the years except small non-linear increasing.

```
##  [1] "Australia and New Zealand"           "Central and Eastern Europe"
##  [3] "Commonwealth of Independent States"  "East Asia"
##  [5] "Latin America and Caribbean"         "Middle East and North Africa"
##  [7] "North America and ANZ"               "South Asia"
##  [9] "Sub-Saharan Africa"                  "Western Europe"
```

Happiness Scores across Regions
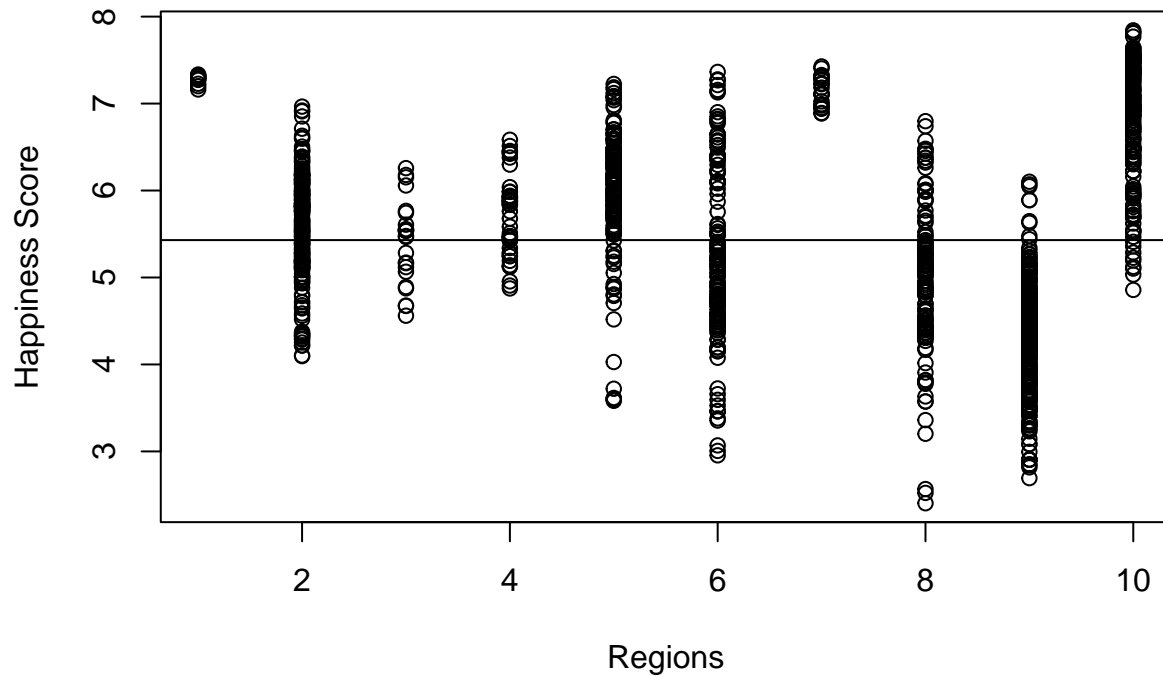
Australia, New Zealand, North America, and ANZ always had a high happiness score. Sub-Saharan Africa vary from average countries to sad countries. Opposite to Sub-Saharan Africa, Western Europe vary from average countries to happy countries.

Regions will be added to the model to test if their effect is significant.

```
## [1] "Central and Eastern Europe"      "Commonwealth of Independent States"
## [3] "East Asia"                        "Latin America and Caribbean"
## [5] "Middle East and North Africa"     "North America and ANZ"
## [7] "South Asia"                        "Sub-Saharan Africa"
## [9] "Western Europe"
```
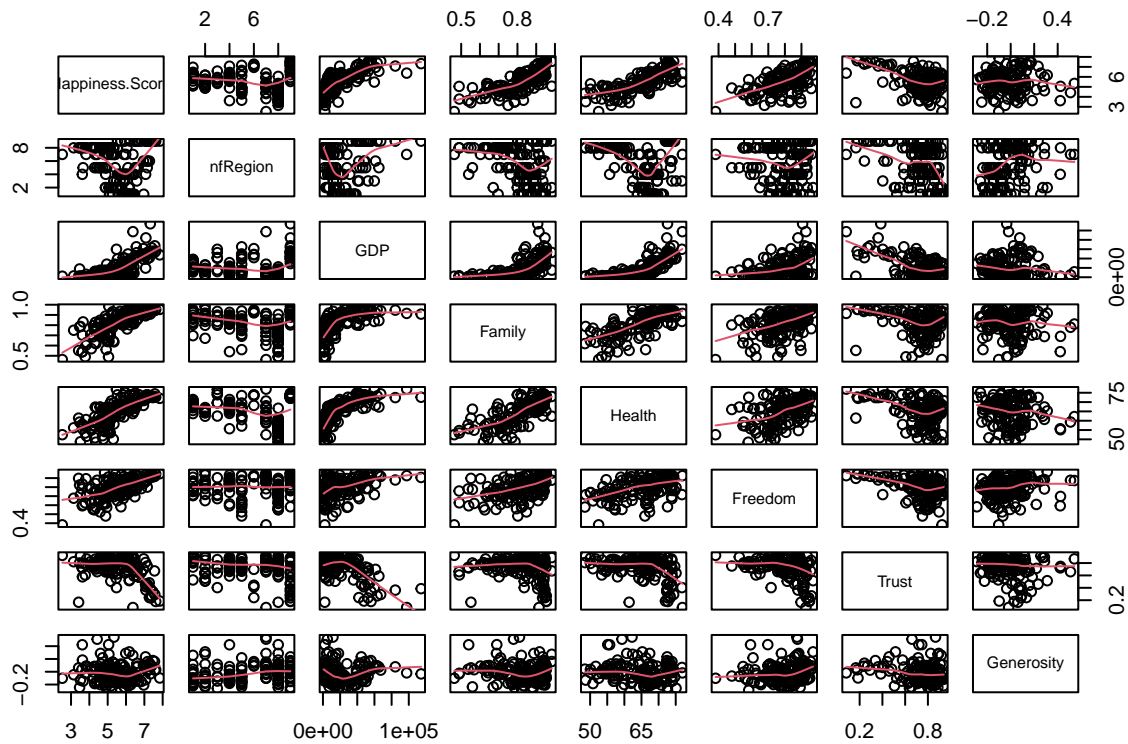
GDP need to be transformed by logged (Economy variable).
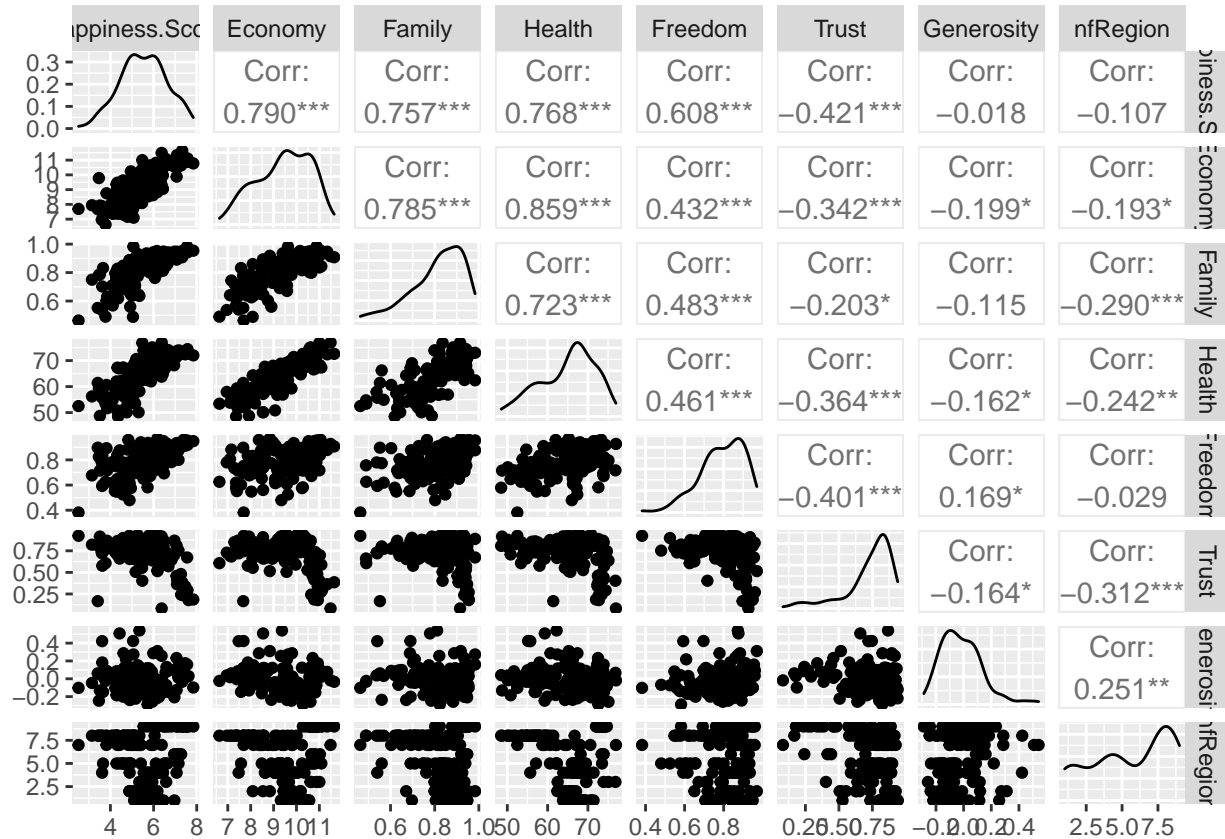
The high correlations are: Economy:Health (0.859), Economy:Family (0.785), Family:Health (0.723).

Modeling

```
##     Economy      Family      Health     Freedom      Trust Generosity   nfRegion
##    5.165619    3.063508    4.270720    1.586418    1.596079   1.209740   1.392907
```

Economy have the highest VIF (5.1656), so we will delete it.

```
## $coefficients
##                  Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) -2.55826899 0.76117040 -3.3609675 9.979689e-04
## Family        3.60817144 0.61144136  5.9010916 2.532619e-08
## Health        0.05984863 0.01080485  5.5390526 1.429924e-07
## Freedom       1.95549466 0.50908277  3.8412116 1.839147e-04
## Trust        -0.59161132 0.32121639 -1.8417844 6.759296e-02
## Generosity    0.14124324 0.32975160  0.4283322 6.690582e-01
## nfRegion      0.02602237 0.01977424  1.3159736 1.903034e-01
##
## $adj.r.squared
## [1] 0.7302522

##     Family      Health     Freedom      Trust Generosity   nfRegion
##    2.347598    2.539512    1.583570    1.576725   1.174107   1.376531
```
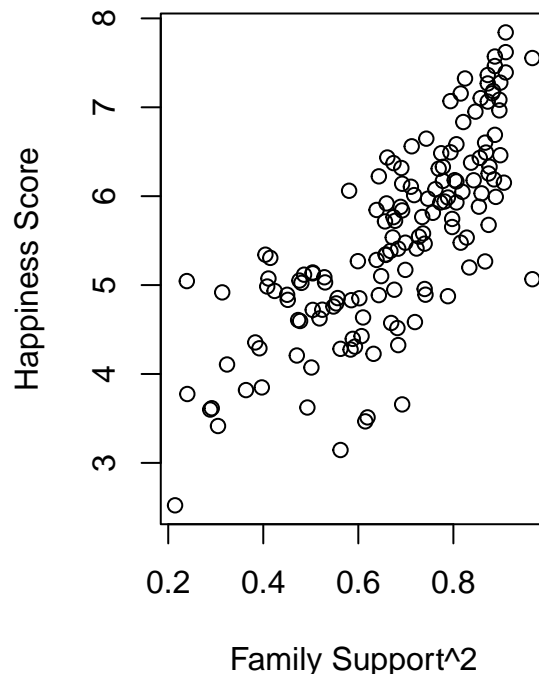
VIF values reasonable now.

Family has a curved plot against Happiness Scores,

Happiness Scores vs Family
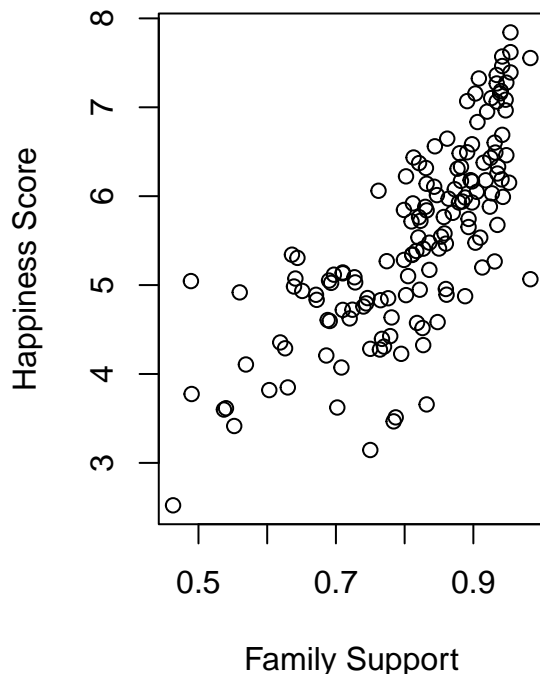
Happiness Scores vs Family^2

after squaring Family, it give us more linear fit.

backward selection:

```
## $coefficients
##                  Estimate Std. Error    t value      Pr(>|t|)
```

```
## (Intercept) -1.22234905 0.79189494 -1.5435748 1.249170e-01
## I(Family^2)  2.49362080 0.39910555  6.2480233 4.556896e-09
## Health        0.05759131 0.01072591  5.3693657 3.147896e-07
## Freedom       1.95264959 0.50155618  3.8931822 1.516926e-04
## Trust        -0.51930264 0.31647920 -1.6408745 1.030366e-01
## Generosity    0.12644323 0.32584820  0.3880434 6.985650e-01
## nfRegion      0.02815384 0.01957451  1.4382910 1.525518e-01
##
## $adj.r.squared
## [1] 0.7365324

## $coefficients
##               Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) -1.20438303 0.78818891 -1.528039 1.287119e-01
## I(Family^2)  2.49179465 0.39789078  6.262509 4.176596e-09
## Health        0.05681212 0.01050493  5.408139 2.608912e-07
## Freedom       1.99688595 0.48697626  4.100582 6.884045e-05
## Trust        -0.53161799 0.31394726 -1.693335 9.256893e-02
## nfRegion      0.02921128 0.01932624  1.511482 1.328723e-01
##
## $adj.r.squared
## [1] 0.7380974

## $coefficients
##               Estimate  Std. Error     t value      Pr(>|t|)
## I(Family^2)  2.68475501 0.379065013  7.0825714 5.779352e-11
## Health        0.04508280 0.007204745  6.2573755 4.225517e-09
## Freedom       1.69890286 0.448288686  3.7897518 2.210358e-04
## Trust        -0.88081431 0.216259980 -4.0729418 7.633422e-05
## nfRegion      0.01517556 0.017082493  0.8883693 3.758237e-01
##
## $adj.r.squared
## [1] 0.9904006

## $coefficients
##               Estimate Std. Error   t value      Pr(>|t|)
## I(Family^2)  2.56237255 0.35288942  7.261120 2.160964e-11
## Health        0.04671889 0.00696032  6.712175 4.024866e-10
## Freedom       1.79735646 0.43405788  4.140822 5.849402e-05
## Trust        -0.90475890 0.21441772 -4.219609 4.294328e-05
##
## $adj.r.squared
## [1] 0.9904146
```
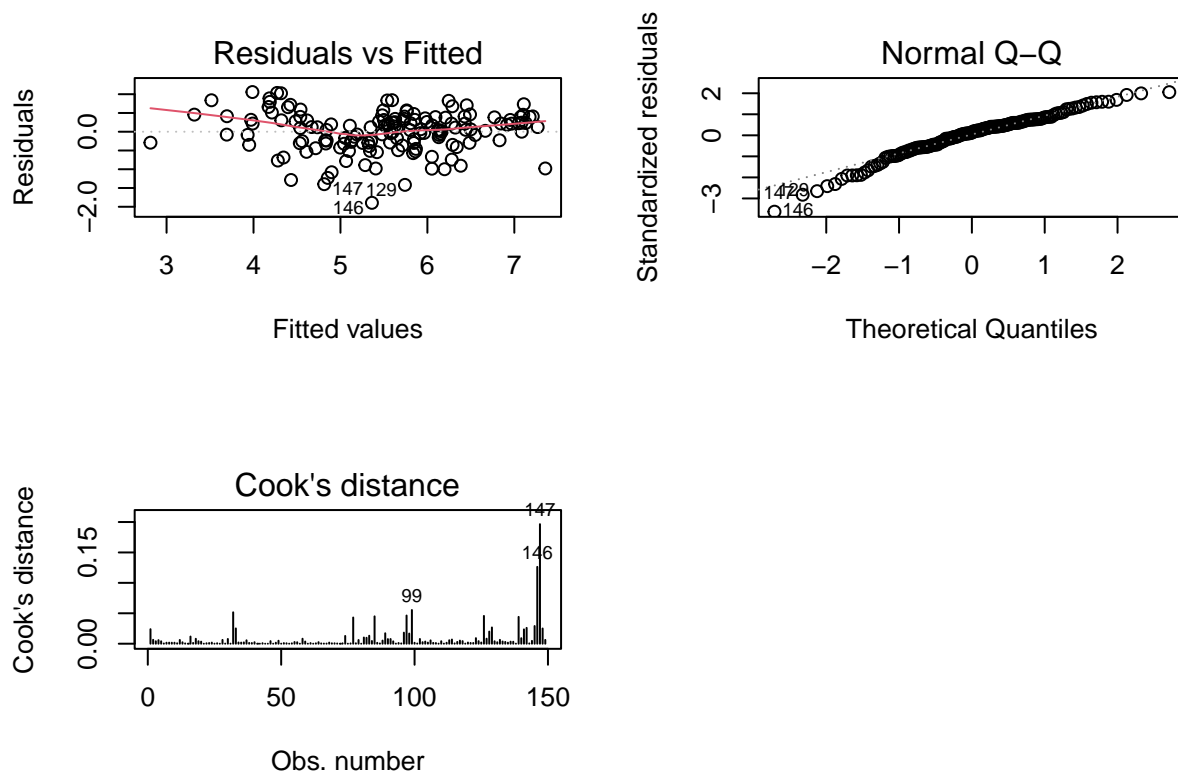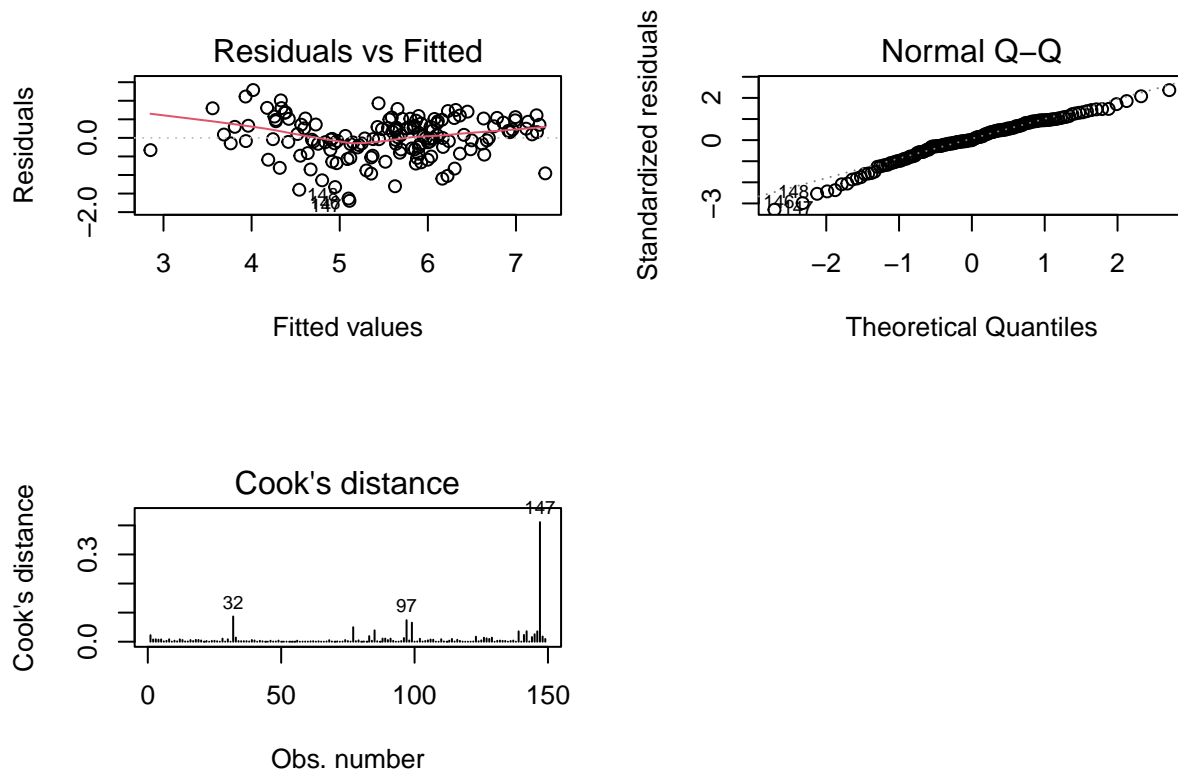
For model0:


```
##
##  Shapiro-Wilk normality test
##
## data:  model0$residuals
## W = 0.96799, p-value = 0.001502

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 5.199572, Df = 1, p = 0.022592
```

For modelf:


```
##
##  Shapiro-Wilk normality test
##
## data:  modelf$residuals
## W = 0.9746, p-value = 0.007287

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 5.526284, Df = 1, p = 0.018733

## Analysis of Variance Table
##
## Model 1: Happiness.Score ~ (Country + Region + Happiness.Rank + Economy +
##     Family + Health + Freedom + Trust + Generosity + nfRegion) -
##     Happiness.Rank - Country - Region
## Model 2: Happiness.Score ~ I(Family^2) + Health + Freedom + Trust - 1
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    141 41.374
## 2    145 44.140 -4   -2.7658 2.3564 0.05657 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
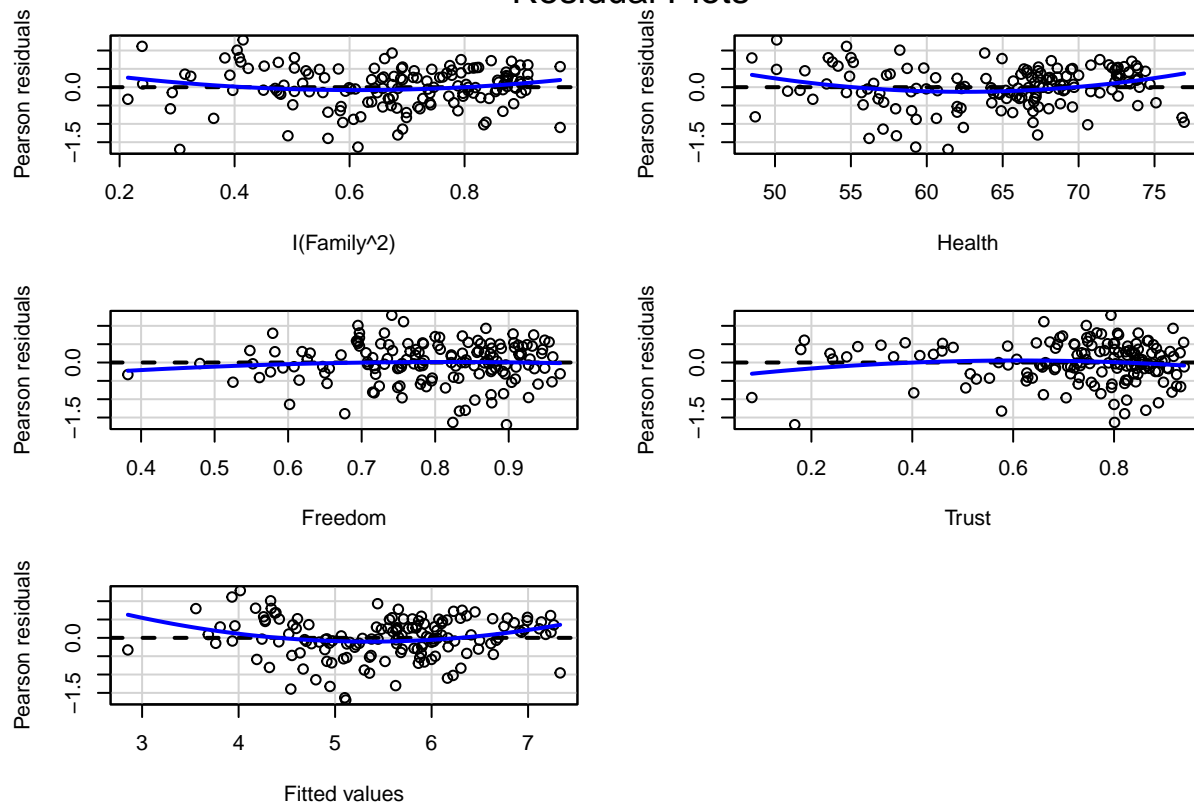
anova fail to reject **modelf** different than **model0**, but **modelf** have fewer variables and all of them are significant, and higher adj.R^2 (0.9904).
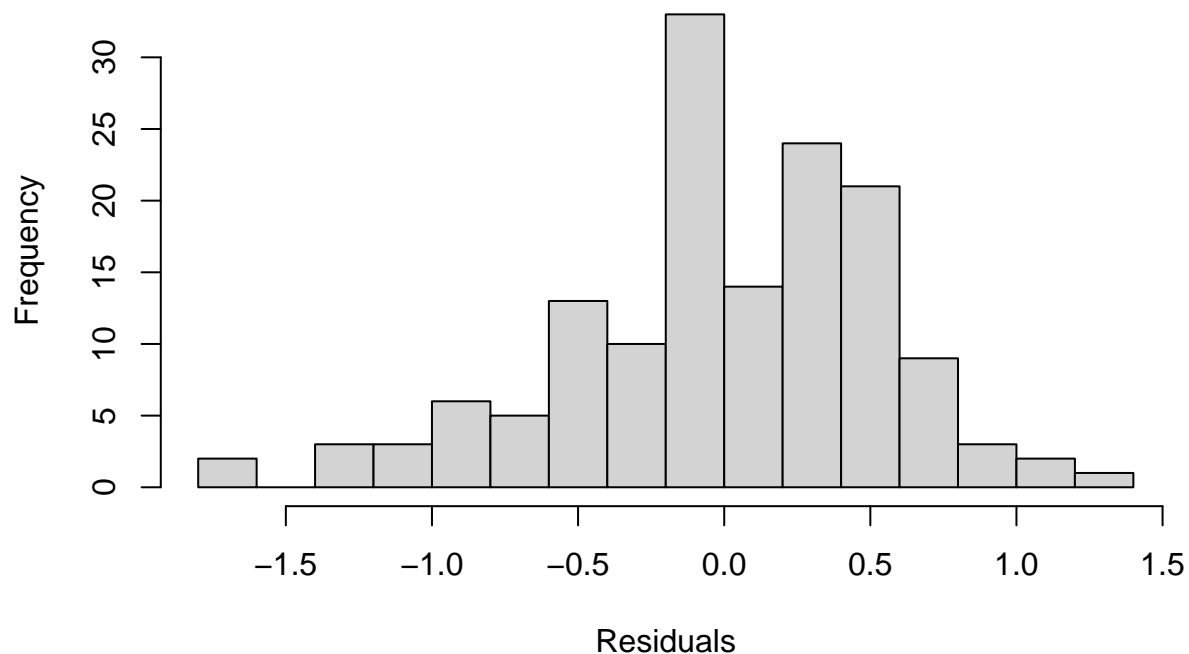
The model rejected by Shapiro test for normality and by ncvTest, but it still better than other models.

9

Residual Plots


```
##              Test stat Pr(>|Test stat|)
## I(Family^2)    2.1420          0.033872 *
## Health         1.4614          0.146096
## Freedom        0.6453          0.519779
## Trust         -0.6865          0.493516
## Tukey test     2.7610          0.005762 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Histogram of model's residuals

residual plots, and the histogram of residuals give acceptable graphs.

The Final model:

Happiness.Score = 2.5624×I(Family^2) + 0.0467×Health + 1.7974×Freedom - 0.9048×Trust

Logistic

AIC for backward selection models:

```
## $coefficients
##                  Estimate Std. Error    z value      Pr(>|z|)
## (Intercept) -48.73754564  9.8995978 -4.9231844 8.514714e-07
## Economy       0.28952772  0.5838647  0.4958816 6.199780e-01
## Family       17.87566209  6.0489108  2.9551869 3.124796e-03
## Health        0.30986660  0.1169902  2.6486539 8.081304e-03
## Freedom      10.76609513  3.6148374  2.9783069 2.898456e-03
## Trust         2.57296672  2.4147095  1.0655388 2.866322e-01
## Generosity   -2.80202105  2.3829734 -1.1758507 2.396545e-01
## nfRegion     -0.04551269  0.1304758 -0.3488210 7.272237e-01
##
## $aic
## [1] 89.43756

## $coefficients
```

```
##                   Estimate Std. Error    z value      Pr(>|z|)
## (Intercept) -48.8788808  9.7667393 -5.0046263 5.597050e-07
## Economy        0.2529835  0.5705227  0.4434241 6.574590e-01
## Family        18.3951069  5.8989073  3.1183923 1.818406e-03
## Health         0.3087508  0.1151683  2.6808669 7.343171e-03
## Freedom       10.6138808  3.5403187  2.9980015 2.717563e-03
## Trust          2.6566456  2.4113532  1.1017240 2.705817e-01
## Generosity    -2.8423349  2.3607389 -1.2040022 2.285887e-01
##
## $aic
## [1] 87.5586

## $coefficients
##                   Estimate Std. Error   z value      Pr(>|z|)
## (Intercept) -49.1918501  9.8853243 -4.976251 6.482773e-07
## Family        19.3035211  5.5560009  3.474355 5.120825e-04
## Health         0.3393704  0.0949768  3.573193 3.526547e-04
## Freedom       10.5531846  3.5356482  2.984795 2.837686e-03
## Trust          2.6333556  2.4012467  1.096662 2.727893e-01
## Generosity    -2.9521495  2.3721643 -1.244496 2.133170e-01
##
## $aic
## [1] 85.75755

## $coefficients
##                   Estimate Std. Error   z value      Pr(>|z|)
## (Intercept) -44.4886290 8.43734065 -5.272826 1.343389e-07
## Family        19.0413157 5.46676356  3.483106 4.956325e-04
## Health         0.3151952 0.09266464  3.401461 6.702673e-04
## Freedom        9.5472433 3.33775767  2.860376 4.231385e-03
## Generosity    -2.5489485 2.31143455 -1.102756 2.701331e-01
##
## $aic
## [1] 84.92341

## $coefficients
##                   Estimate Std. Error   z value      Pr(>|z|)
## (Intercept) -44.3548603 8.40431071 -5.277632 1.308638e-07
## Family        18.3825787 5.37690288  3.418804 6.289694e-04
## Health         0.3361003 0.09451098  3.556203 3.762528e-04
## Freedom        8.4725532 3.14252208  2.696100 7.015664e-03
##
## $aic
## [1] 84.22116
```
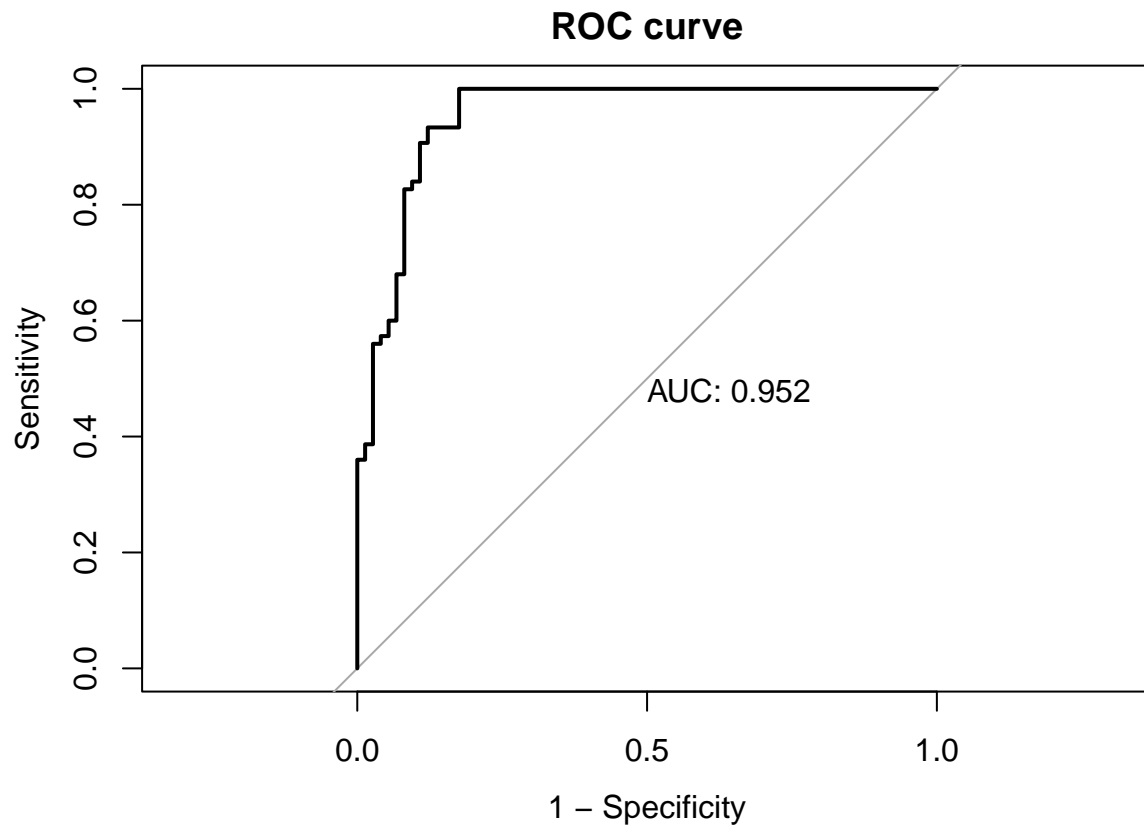
ROC curve

From the ROC plot curve and the high value of AUC equal to 0.952, we conclude that the Lmodel4 have a good fit.

The final logistic model: **log(odds) = -44.3549 + 18.3826×Family + 0.3361×Health + 8.4726×Freedom**

confusion matrix

```
##
## Lmodel4.pred  0   1
##          No  64   5
##          Yes 10  70
```

```
## [1] "Accuracy = 0.8993"
```

```
## [1] "Sensitivity = 0.8649"
```

```
## [1] "Specificity = 0.9333"
```

All Accuracy, Sensitivity, and Specificity have big value which tell the predict is good.