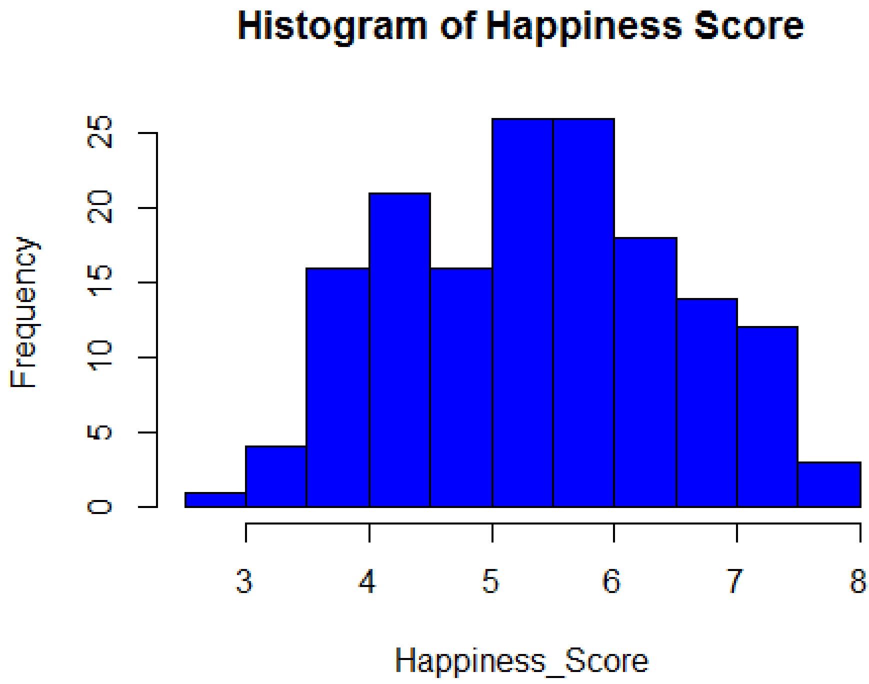
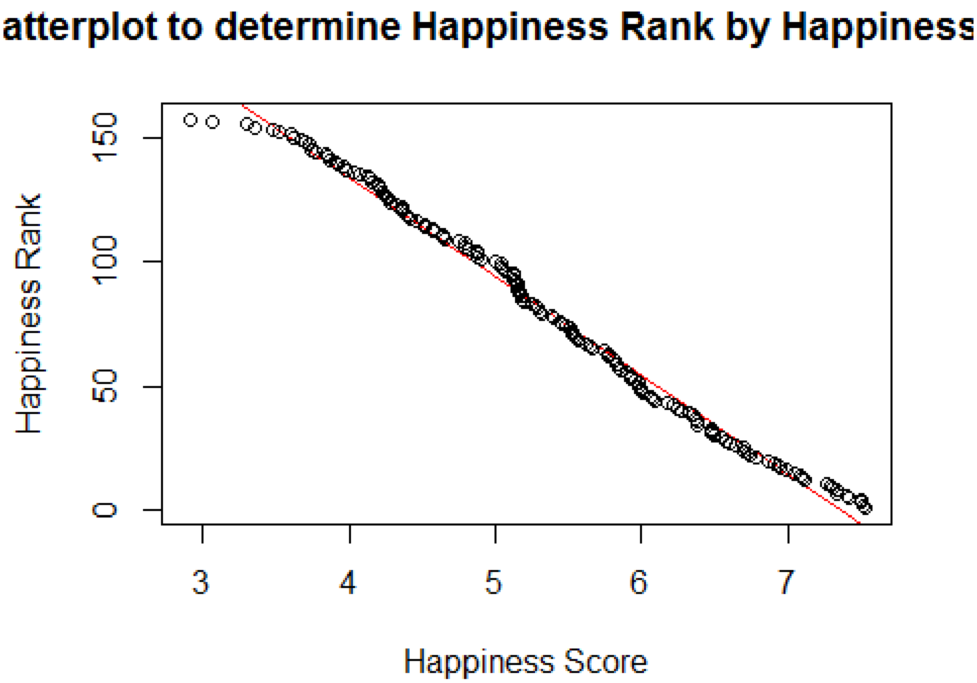
```r
library(readxl)

library(dplyr)

library(lubridate)

library(ggplot2)

library(reshape2)

library(corrplot)

library(Hmisc)

library(caTools)

library(car)

library(gridExtra)

#WorldHappinessData_2016 <- read_excel("C:/Users/Tanmay/Desktop/HBU/ANLY 502
- Analytical Methods I/Project/WorldHappinessData_2016.xlsx")
#View(WorldHappinessData_2016)

setwd("C:/Users/HCL/Documents/Nilay Back
Up/Quicky/Master/harrisburg/Sem1/ANLY 502/Project/")
WorldHappinessData_2016 <- read_excel("data_2016.xlsx", sheet = 1)

str(WorldHappinessData_2016)

## Classes 'tbl_df', 'tbl' and 'data.frame':    157 obs. of  13 variables:
##  $ Country                      : chr  "Denmark" "Switzerland" "Iceland"
"Norway" ...
##  $ Region                       : chr  "Western Europe" "Western Europe"
"Western Europe" "Western Europe" ...
##  $ Happiness_Rank               : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Happiness_Score              : num  7.53 7.51 7.5 7.5 7.41 ...
##  $ Lower Confidence Interval    : num  7.46 7.43 7.33 7.42 7.35 ...
##  $ Upper Confidence Interval    : num  7.59 7.59 7.67 7.58 7.47 ...
##  $ Economy                      : num  1.44 1.53 1.43 1.58 1.41 ...
##  $ Family                       : num  1.16 1.15 1.18 1.13 1.13 ...
##  $ Health_Life_Expectancy       : num  0.795 0.863 0.867 0.796 0.811 ...
##  $ Freedom                      : num  0.579 0.586 0.566 0.596 0.571 ...
##  $ Trust_Government_Corruption: num  0.445 0.412 0.15 0.358 0.41 ...
##  $ Generosity                   : num  0.362 0.281 0.477 0.379 0.255 ...
##  $ Dystopia_Residual            : num  2.74 2.69 2.83 2.66 2.83 ...

# Data clean-up and preparation :
# Remove columns 5 and 6 as they are just CI values
Rev_WorldHappinessData_2016<- WorldHappinessData_2016[, -c(5,6)]
#View(Rev_WorldHappinessData_2016)
```

```
str(Rev_WorldHappinessData_2016)

## Classes 'tbl_df', 'tbl' and 'data.frame':    157 obs. of  11 variables:
##  $ Country                      : chr  "Denmark" "Switzerland" "Iceland"
"Norway" ...
##  $ Region                       : chr  "Western Europe" "Western Europe"
"Western Europe" "Western Europe" ...
##  $ Happiness_Rank               : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Happiness_Score              : num  7.53 7.51 7.5 7.5 7.41 ...
##  $ Economy                      : num  1.44 1.53 1.43 1.58 1.41 ...
##  $ Family                       : num  1.16 1.15 1.18 1.13 1.13 ...
##  $ Health_Life_Expectancy       : num  0.795 0.863 0.867 0.796 0.811 ...
##  $ Freedom                      : num  0.579 0.586 0.566 0.596 0.571 ...
##  $ Trust_Government_Corruption: num  0.445 0.412 0.15 0.358 0.41 ...
##  $ Generosity                   : num  0.362 0.281 0.477 0.379 0.255 ...
##  $ Dystopia_Residual            : num  2.74 2.69 2.83 2.66 2.83 ...

par(mfrow = c(1,1))
### Creating a Histogram for the Happiness_score to depict the Normality of
distribution
hist(Rev_WorldHappinessData_2016$Happiness_Score,xlab = "Happiness_Score",col
= "blue",main = "Histogram of Happiness Score")
```



```
### Scatter plot to determine the relation between Happines Rank and
Happiness Score
plot(Rev_WorldHappinessData_2016$Happiness_Score,Rev_WorldHappinessData_2016$
Happiness_Rank,xlab = "Happiness Score",
     ylab = "Happiness Rank",main ="Scatterplot to determine Happiness Rank
by Happiness Score",
     abline(lm(Rev_WorldHappinessData_2016$Happiness_Rank ~
Rev_WorldHappinessData_2016$Happiness_Score),col = "Red"))
```



```
### Correlation matrix for the Happiness score and other variables
corrdata <- Rev_WorldHappinessData_2016[,-1:-2]
corrdata

## # A tibble: 157 x 9
##    Happiness_Rank Happiness_Score Economy Family Health_Life_Expe~ Freedom
##             <dbl>           <dbl>   <dbl>  <dbl>             <dbl>   <dbl>
##  1             1.            7.53    1.44   1.16             0.795   0.579
##  2             2.            7.51    1.53   1.15             0.863   0.586
##  3             3.            7.50    1.43   1.18             0.867   0.566
##  4             4.            7.50    1.58   1.13             0.796   0.596
##  5             5.            7.41    1.41   1.13             0.811   0.571
##  6             6.            7.40    1.44   1.10             0.828   0.574
##  7             7.            7.34    1.46   1.03             0.812   0.552
##  8             8.            7.33    1.36   1.17             0.831   0.581
##  9             9.            7.31    1.44   1.10             0.851   0.568
## 10            10.            7.29    1.45   1.09             0.831   0.582
## # ... with 147 more rows, and 3 more variables:
## #   Trust_Government_Corruption <dbl>, Generosity <dbl>,
## #   Dystopia_Residual <dbl>
```

```
res <- rcorr(as.matrix(corrdata))
res

##                             Happiness_Rank Happiness_Score Economy Family
## Happiness_Rank                        1.00           -1.00   -0.79  -0.73
## Happiness_Score                      -1.00            1.00    0.79   0.74
## Economy                              -0.79            0.79    1.00   0.67
## Family                               -0.73            0.74    0.67   1.00
## Health_Life_Expectancy               -0.77            0.77    0.84   0.59
## Freedom                              -0.56            0.57    0.36   0.45
## Trust_Government_Corruption          -0.39            0.40    0.29   0.21
## Generosity                           -0.15            0.16   -0.03   0.09
## Dystopia_Residual                    -0.54            0.54    0.07   0.12
##                             Health_Life_Expectancy Freedom
## Happiness_Rank                               -0.77   -0.56
## Happiness_Score                               0.77    0.57
## Economy                                       0.84    0.36
## Family                                        0.59    0.45
## Health_Life_Expectancy                        1.00    0.34
## Freedom                                       0.34    1.00
## Trust_Government_Corruption                   0.25    0.50
## Generosity                                    0.08    0.36
## Dystopia_Residual                             0.10    0.09
##                             Trust_Government_Corruption Generosity
## Happiness_Rank                                    -0.39      -0.15
## Happiness_Score                                    0.40       0.16
## Economy                                            0.29      -0.03
## Family                                             0.21       0.09
## Health_Life_Expectancy                             0.25       0.08
## Freedom                                            0.50       0.36
## Trust_Government_Corruption                        1.00       0.31
## Generosity                                         0.31       1.00
## Dystopia_Residual                                  0.00      -0.13
##                             Dystopia_Residual
## Happiness_Rank                          -0.54
## Happiness_Score                          0.54
## Economy                                  0.07
## Family                                   0.12
## Health_Life_Expectancy                   0.10
## Freedom                                  0.09
## Trust_Government_Corruption              0.00
## Generosity                              -0.13
## Dystopia_Residual                        1.00
##
## n= 157
##
##
## P
##                             Happiness_Rank Happiness_Score Economy Family
## Happiness_Rank                             0.0000          0.0000  0.0000
## Happiness_Score             0.0000                         0.0000  0.0000
## Economy                     0.0000         0.0000                  0.0000
## Family                      0.0000         0.0000          0.0000
## Health_Life_Expectancy      0.0000         0.0000          0.0000  0.0000
## Freedom                     0.0000         0.0000          0.0000  0.0000
## Trust_Government_Corruption 0.0000         0.0000          0.0002  0.0072
## Generosity                  0.0693         0.0498          0.7509  0.2643
## Dystopia_Residual           0.0000         0.0000          0.3931  0.1355
##                             Health_Life_Expectancy Freedom
## Happiness_Rank              0.0000                 0.0000
## Happiness_Score             0.0000                 0.0000
## Economy                     0.0000                 0.0000
## Family                      0.0000                 0.0000
## Health_Life_Expectancy                             0.0000
## Freedom                     0.0000
## Trust_Government_Corruption 0.0016                 0.0000
## Generosity                  0.3442                 0.0000
## Dystopia_Residual           0.2088                 0.2537
##                             Trust_Government_Corruption Generosity
## Happiness_Rank              0.0000                      0.0693
## Happiness_Score             0.0000                      0.0498
## Economy                     0.0002                      0.7509
## Family                      0.0072                      0.2643
## Health_Life_Expectancy      0.0016                      0.3442
## Freedom                     0.0000                      0.0000
## Trust_Government_Corruption                             0.0000
## Generosity                  0.0000
## Dystopia_Residual           0.9712                      0.0968
##                             Dystopia_Residual
## Happiness_Rank              0.0000
## Happiness_Score             0.0000
## Economy                     0.3931
## Family                      0.1355
## Health_Life_Expectancy      0.2088
## Freedom                     0.2537
## Trust_Government_Corruption 0.9712
## Generosity                  0.0968
## Dystopia_Residual
```

```
# Extract the correlation coefficients
res$r
```

```
##                             Happiness_Rank Happiness_Score     Economy
## Happiness_Rank                   1.0000000      -0.9957434 -0.79357712
## Happiness_Score                 -0.9957434       1.0000000  0.79032202
## Economy                         -0.7935771       0.7903220  1.00000000
## Family                          -0.7332763       0.7392516  0.66953969
## Health_Life_Expectancy          -0.7679908       0.7653843  0.83706723
## Freedom                         -0.5571687       0.5668267  0.36228285
## Trust_Government_Corruption     -0.3871016       0.4020322  0.29418478
## Generosity                      -0.1453688       0.1568478 -0.02553066
## Dystopia_Residual               -0.5426158       0.5437376  0.06862402
##                                     Family Health_Life_Expectancy     Freedom
## Happiness_Rank              -0.73327635            -0.76799078 -0.55716871
## Happiness_Score              0.73925158             0.76538433  0.56682667
## Economy                      0.66953969             0.83706723  0.36228285
## Family                       1.00000000             0.58837678  0.45020820
## Health_Life_Expectancy       0.58837678             1.00000000  0.34119929
## Freedom                      0.45020820             0.34119929  1.00000000
## Trust_Government_Corruption  0.21356094             0.24958329  0.50205397
## Generosity                   0.08962885             0.07598731  0.36175133
## Dystopia_Residual            0.11967231             0.10085705  0.09163094
##                             Trust_Government_Corruption  Generosity
## Happiness_Rank                             -0.387101640 -0.14536878
## Happiness_Score                             0.402032245  0.15684780
## Economy                                     0.294184775 -0.02553066
## Family                                      0.213560938  0.08962885
## Health_Life_Expectancy                      0.249583285  0.07598731
## Freedom                                     0.502053974  0.36175133
## Trust_Government_Corruption                 1.000000000  0.30592986
## Generosity                                  0.305929861  1.00000000
## Dystopia_Residual                          -0.002909346 -0.13300141
##                             Dystopia_Residual
## Happiness_Rank                   -0.542615800
## Happiness_Score                   0.543737614
## Economy                           0.068624025
## Family                            0.119672314
## Health_Life_Expectancy            0.100857046
## Freedom                           0.091630941
## Trust_Government_Corruption      -0.002909346
## Generosity                       -0.133001407
## Dystopia_Residual                 1.000000000
```

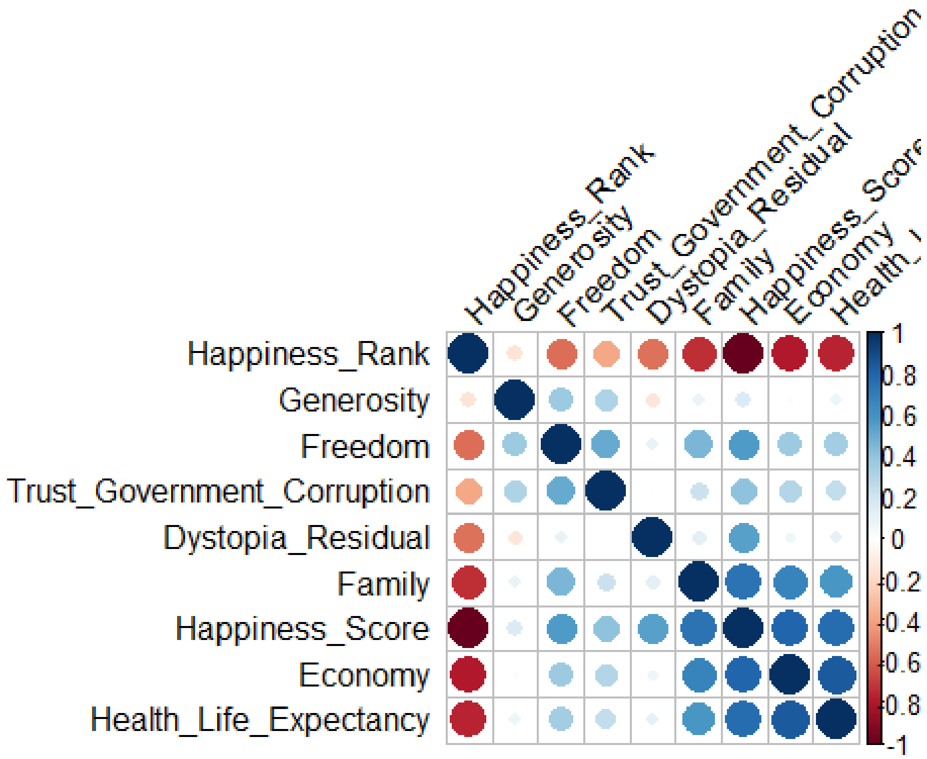
```
# Extract p-values
res$P
```

```
##                             Happiness_Rank Happiness_Score      Economy
## Happiness_Rank                          NA    0.000000e+00 0.000000e+00
## Happiness_Score               0.000000e+00              NA 0.000000e+00
## Economy                       0.000000e+00    0.000000e+00           NA
## Family                        0.000000e+00    0.000000e+00 0.000000e+00
## Health_Life_Expectancy        0.000000e+00    0.000000e+00 0.000000e+00
## Freedom                       3.508305e-14    1.021405e-14 3.123973e-06
## Trust_Government_Corruption   5.492284e-07    1.797843e-07 1.841886e-04
## Generosity                    6.928267e-02    4.979137e-02 7.509449e-01
## Dystopia_Residual             2.122746e-13    1.851852e-13 3.931090e-01
##                                   Family Health_Life_Expectancy
## Happiness_Rank              0.000000e+00           0.000000e+00
## Happiness_Score             0.000000e+00           0.000000e+00
## Economy                     0.000000e+00           0.000000e+00
## Family                                NA           4.440892e-16
## Health_Life_Expectancy      4.440892e-16                     NA
## Freedom                     3.300642e-09           1.225460e-05
## Trust_Government_Corruption 7.240343e-03           1.619787e-03
## Generosity                  2.642850e-01           3.442107e-01
## Dystopia_Residual           1.354692e-01           2.088101e-01
##                                  Freedom Trust_Government_Corruption
## Happiness_Rank              3.508305e-14                5.492284e-07
## Happiness_Score             1.021405e-14                1.797843e-07
## Economy                     3.123973e-06                1.841886e-04
## Family                      3.300642e-09                7.240343e-03
## Health_Life_Expectancy      1.225460e-05                1.619787e-03
## Freedom                               NA                2.100009e-11
## Trust_Government_Corruption 2.100009e-11                          NA
## Generosity                  3.237423e-06                9.759224e-05
## Dystopia_Residual           2.537199e-01                9.711526e-01
##                               Generosity Dystopia_Residual
## Happiness_Rank              6.928267e-02      2.122746e-13
## Happiness_Score             4.979137e-02      1.851852e-13
## Economy                     7.509449e-01      3.931090e-01
## Family                      2.642850e-01      1.354692e-01
## Health_Life_Expectancy      3.442107e-01      2.088101e-01
## Freedom                     3.237423e-06      2.537199e-01
## Trust_Government_Corruption 9.759224e-05      9.711526e-01
## Generosity                            NA      9.679955e-02
## Dystopia_Residual           9.679955e-02                NA
```

```
## Plot Correlation plot
res2<-rcorr(as.matrix(corrdata[,1:8]))
res2
```

```
##                             Happiness_Rank Happiness_Score Economy Family
## Happiness_Rank                        1.00           -1.00   -0.79  -0.73
## Happiness_Score                      -1.00            1.00    0.79   0.74
## Economy                              -0.79            0.79    1.00   0.67
## Family                               -0.73            0.74    0.67   1.00
## Health_Life_Expectancy               -0.77            0.77    0.84   0.59
## Freedom                              -0.56            0.57    0.36   0.45
## Trust_Government_Corruption          -0.39            0.40    0.29   0.21
## Generosity                           -0.15            0.16   -0.03   0.09
##                             Health_Life_Expectancy Freedom
## Happiness_Rank                               -0.77   -0.56
## Happiness_Score                               0.77    0.57
## Economy                                       0.84    0.36
## Family                                        0.59    0.45
## Health_Life_Expectancy                        1.00    0.34
## Freedom                                       0.34    1.00
## Trust_Government_Corruption                   0.25    0.50
## Generosity                                    0.08    0.36
##                             Trust_Government_Corruption Generosity
## Happiness_Rank                                    -0.39      -0.15
## Happiness_Score                                    0.40       0.16
## Economy                                            0.29      -0.03
## Family                                             0.21       0.09
## Health_Life_Expectancy                             0.25       0.08
## Freedom                                            0.50       0.36
## Trust_Government_Corruption                        1.00       0.31
## Generosity                                         0.31       1.00
##
## n= 157
##
##
## P
##                             Happiness_Rank Happiness_Score Economy Family
## Happiness_Rank                             0.0000          0.0000  0.0000
## Happiness_Score             0.0000                         0.0000  0.0000
## Economy                     0.0000         0.0000                  0.0000
## Family                      0.0000         0.0000          0.0000
## Health_Life_Expectancy      0.0000         0.0000          0.0000  0.0000
## Freedom                     0.0000         0.0000          0.0000  0.0000
## Trust_Government_Corruption 0.0000         0.0000          0.0002  0.0072
## Generosity                  0.0693         0.0498          0.7509  0.2643
##                             Health_Life_Expectancy Freedom
## Happiness_Rank              0.0000                 0.0000
## Happiness_Score             0.0000                 0.0000
## Economy                     0.0000                 0.0000
## Family                      0.0000                 0.0000
## Health_Life_Expectancy                             0.0000
## Freedom                     0.0000
## Trust_Government_Corruption 0.0016                 0.0000
## Generosity                  0.3442                 0.0000
##                             Trust_Government_Corruption Generosity
## Happiness_Rank              0.0000                      0.0693
## Happiness_Score             0.0000                      0.0498
## Economy                     0.0002                      0.7509
## Family                      0.0072                      0.2643
## Health_Life_Expectancy      0.0016                      0.3442
## Freedom                     0.0000                      0.0000
## Trust_Government_Corruption                             0.0000
## Generosity                  0.0000
```

```
corrplot(res$r,type = "full",order = "hclust",tl.col = "black", tl.srt = 45)
```



```
## Distribution of Happiness Score by Region
p<- Rev_WorldHappinessData_2016 %>%
  dplyr::mutate(Region=as.factor(Region)) %>%
  dplyr::select(Happiness_Score, Region) %>%
  ggplot2::ggplot(ggplot2::aes(x=Happiness_Score)) +
  ggplot2::geom_density(lty=3) +
  ggplot2::geom_density(ggplot2::aes(fill=Region, colour=Region), alpha=0.55)
+
  ggplot2::xlab("Happiness_Score") +
  ggplot2::ggtitle("Distribution of Happiness_Score by Region") +
  ggplot2::labs(fill="Region", y="Density")

## Warning: package 'bindrcpp' was built under R version 3.4.4

# Display the plot


## Distribution of Economy GDP by Region
q <- Rev_WorldHappinessData_2016 %>%
  dplyr::mutate(Region=as.factor(Region)) %>%
  dplyr::select(Economy, Region) %>%
  ggplot2::ggplot(ggplot2::aes(x=Economy)) +
  ggplot2::geom_density(lty=3) +
  ggplot2::geom_density(ggplot2::aes(fill=Region, colour=Region), alpha=0.55)
+
  ggplot2::xlab("Economy_GDP") +
  ggplot2::ggtitle("Distribution of Economy_GDP by Region") +
  ggplot2::labs(fill="Region", y="Density")
# Display the plot


### Distribution of Family by Region
r <- Rev_WorldHappinessData_2016 %>%
  dplyr::mutate(Region=as.factor(Region)) %>%
  dplyr::select(Family, Region) %>%
  ggplot2::ggplot(ggplot2::aes(x=Family)) +
  ggplot2::geom_density(lty=3) +
  ggplot2::geom_density(ggplot2::aes(fill=Region, colour=Region), alpha=0.55)
+
  ggplot2::xlab("Family") +
  ggplot2::ggtitle("Distribution of Family by Region") +
  ggplot2::labs(fill="Region", y="Density")

# Display the plot


### Distribution of Health_Life_expectancy by Region
s<- Rev_WorldHappinessData_2016 %>%
  dplyr::mutate(Region=as.factor(Region)) %>%
  dplyr::select(Health_Life_Expectancy, Region) %>%
  ggplot2::ggplot(ggplot2::aes(x=Health_Life_Expectancy)) +
  ggplot2::geom_density(lty=3) +
  ggplot2::geom_density(ggplot2::aes(fill=Region, colour=Region), alpha=0.55)
+
  ggplot2::xlab("Health_Life_Expectancy") +
  ggplot2::ggtitle("Distribution of Health_Life_Expectancy by Region") +
  ggplot2::labs(fill="Region", y="Density")

# Display the plot

########################
# train

# Remove Country and Region from data as as they are Categorical values
df = Rev_WorldHappinessData_2016[, -c(1,2)]
str(df)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    157 obs. of  9 variables:
##  $ Happiness_Rank               : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Happiness_Score              : num  7.53 7.51 7.5 7.5 7.41 ...
##  $ Economy                      : num  1.44 1.53 1.43 1.58 1.41 ...
##  $ Family                       : num  1.16 1.15 1.18 1.13 1.13 ...
##  $ Health_Life_Expectancy       : num  0.795 0.863 0.867 0.796 0.811 ...
##  $ Freedom                      : num  0.579 0.586 0.566 0.596 0.571 ...
##  $ Trust_Government_Corruption: num  0.445 0.412 0.15 0.358 0.41 ...
##  $ Generosity                    : num  0.362 0.281 0.477 0.379 0.255 ...
##  $ Dystopia_Residual             : num  2.74 2.69 2.83 2.66 2.83 ...

## Training and Testing the processed data-set:
set.seed(101)

# Randomly Splitting the sample, basically randomly assigns a booleans to a
new column "sample"
sample <- sample.split(df$Happiness_Score, SplitRatio = 0.70) # SplitRatio =
percent of sample==TRUE


# Creating a train data
train = subset(df, sample == TRUE)

# Testing the Data
test = subset(df, sample == FALSE)

#############
# Model selection using Multiple Linear Regression:

# 1. Consider 'Happinesss_score' against 'Economy GDP per Capita' for
Multiple Linear Regression model; given by model1
## Forward selection model
model1 <- lm(Happiness_Score ~ Economy ,train)
summary(model1)

##
## Call:
## lm(formula = Happiness_Score ~ Economy, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85783 -0.46877 -0.03564  0.56025  2.17835
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.2617     0.1610   20.26   <2e-16 ***
## Economy       2.2286     0.1562   14.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7093 on 107 degrees of freedom
## Multiple R-squared:  0.6554, Adjusted R-squared:  0.6522
## F-statistic: 203.5 on 1 and 107 DF,  p-value: < 2.2e-16

#65%

model2 <- lm(Happiness_Score ~ Economy+Family ,train)
summary(model2)
```

```
## 
## Call:
## lm(formula = Happiness_Score ~ Economy + Family, data = train)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.76628 -0.43484  0.04531  0.42332  2.26964 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   2.6228     0.1967  13.337  < 2e-16 ***
## Economy       1.5213     0.2033   7.484 2.22e-11 ***
## Family        1.6289     0.3351   4.861 4.07e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.6444 on 106 degrees of freedom
## Multiple R-squared:  0.7182, Adjusted R-squared:  0.7129 
## F-statistic: 135.1 on 2 and 106 DF,  p-value: < 2.2e-16
```

```
#71%

model3 <- lm(Happiness_Score ~ Economy+Family+ Health_Life_Expectancy ,train)
summary(model3)
```

```
## 
## Call:
## lm(formula = Happiness_Score ~ Economy + Family + Health_Life_Expectancy, 
##     data = train)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.67279 -0.37257  0.03052  0.40805  2.25427 
## 
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)    
## (Intercept)              2.5077     0.1940  12.927  < 2e-16 ***
## Economy                  0.8671     0.2972   2.918  0.00432 ** 
## Family                   1.4993     0.3267   4.589 1.24e-05 ***
## Health_Life_Expectancy   1.5183     0.5178   2.932  0.00413 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.6225 on 105 degrees of freedom
## Multiple R-squared:  0.7395, Adjusted R-squared:  0.7321 
## F-statistic: 99.38 on 3 and 105 DF,  p-value: < 2.2e-16
```

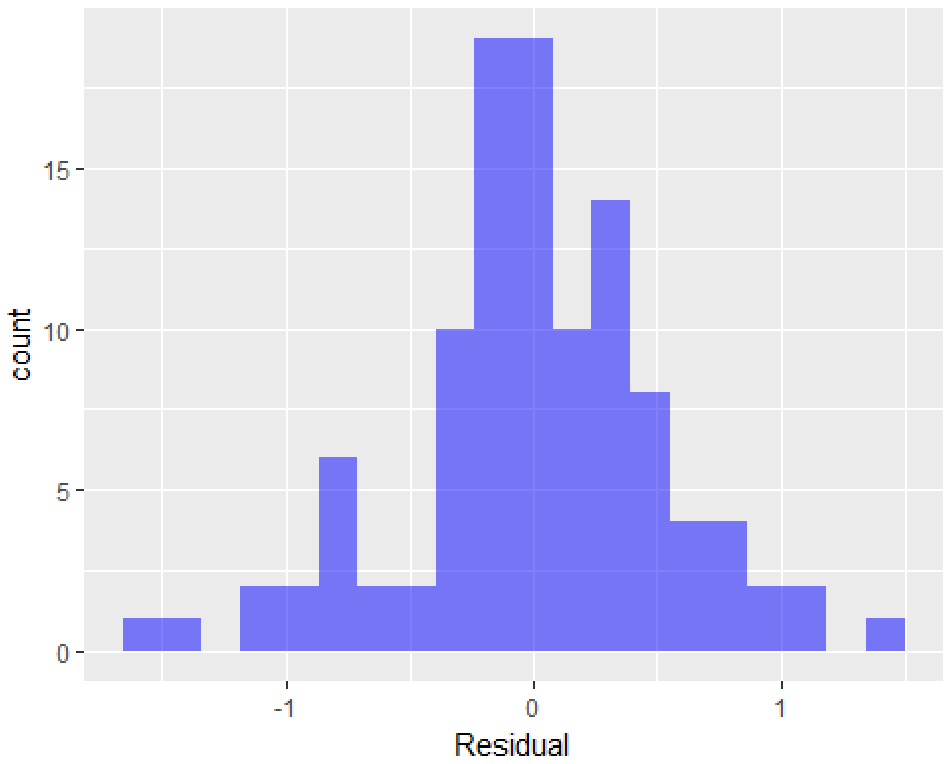
```
#73%

model4 <- lm(Happiness_Score ~ Economy+Family+ Health_Life_Expectancy +
Freedom,train)
summary(model4)

##
## Call:
## lm(formula = Happiness_Score ~ Economy + Family + Health_Life_Expectancy +
##     Freedom, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43885 -0.22300 -0.01371  0.32842  1.51214
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.0850     0.1823  11.438  < 2e-16 ***
## Economy                  0.7234     0.2586   2.797 0.006145 **
## Family                   1.0977     0.2909   3.773 0.000268 ***
## Health_Life_Expectancy   1.5970     0.4488   3.558 0.000564 ***
## Freedom                  2.2734     0.3795   5.990 3.04e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5393 on 104 degrees of freedom
## Multiple R-squared:  0.8063, Adjusted R-squared:  0.7989
## F-statistic: 108.3 on 4 and 104 DF,  p-value: < 2.2e-16

#79

model5 <- lm(Happiness_Score ~ Economy+Family+ Health_Life_Expectancy +
Freedom + Trust_Government_Corruption,train)
summary(model5)

##
## Call:
## lm(formula = Happiness_Score ~ Economy + Family + Health_Life_Expectancy +
##     Freedom + Trust_Government_Corruption, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63274 -0.21810  0.01872  0.29009  1.36488
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2.1256     0.1791  11.869  < 2e-16 ***
## Economy                       0.6800     0.2536   2.681 0.008542 **
## Family                        1.1409     0.2851   4.002 0.000119 ***
## Health_Life_Expectancy        1.5696     0.4391   3.574 0.000536 ***
## Freedom                       1.7247     0.4364   3.953 0.000142 ***
## Trust_Government_Corruption   1.3047     0.5455   2.392 0.018577 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5275 on 103 degrees of freedom
## Multiple R-squared:  0.8165, Adjusted R-squared:  0.8076
## F-statistic: 91.69 on 5 and 103 DF,  p-value: < 2.2e-16
```

```
#80




# Adding geneoristy does not result in higher adjusted R2, so this is our best model



# Grab residuals
res <- residuals(model5)
# Convert to DataFrame for gglpot
res <- as.data.frame(res)
head(res)
```

```
##         res
## 1 0.2651225
## 2 0.1361174
## 3 0.2701727
## 4 0.2441902
## 5 0.3513707
## 6 0.4256380
```

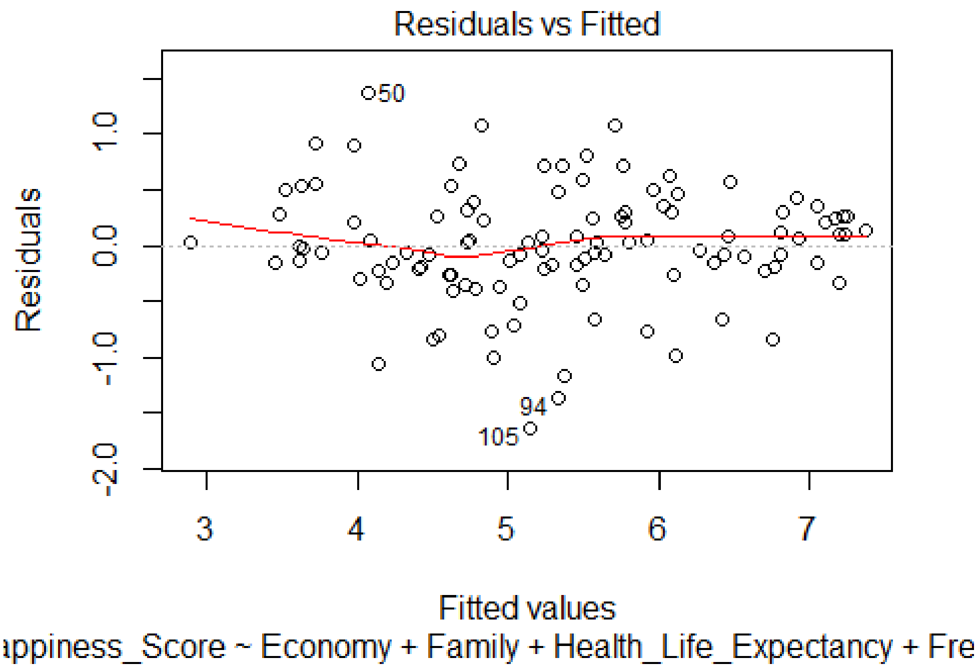
```
##         res
## 1 0.2651225
## 2 0.1361174
## 3 0.2701727
## 4 0.2441902
## 5 0.3513707
## 6 0.4256380
Residual = res
ggplot(Residual,aes(Residual)) +  geom_histogram(fill='blue',alpha=0.5, bins = 20)
```

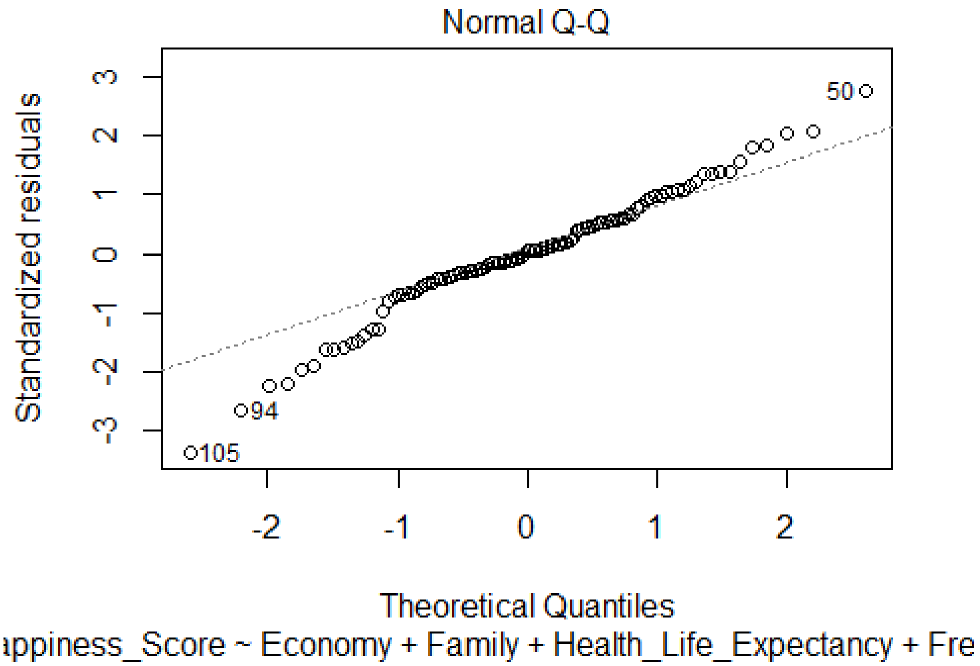
```
## Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.
```

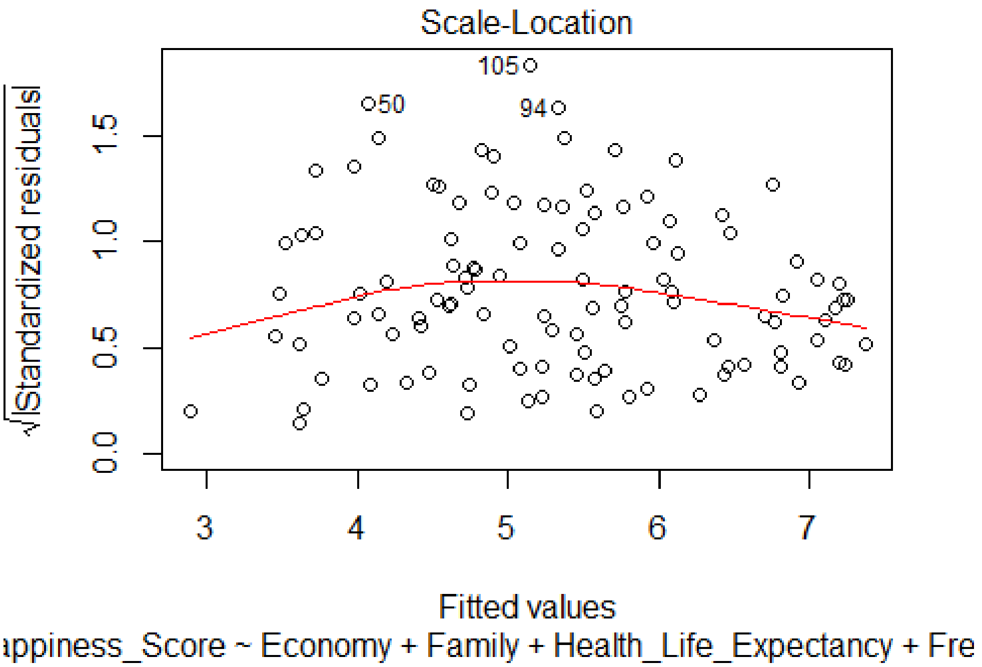


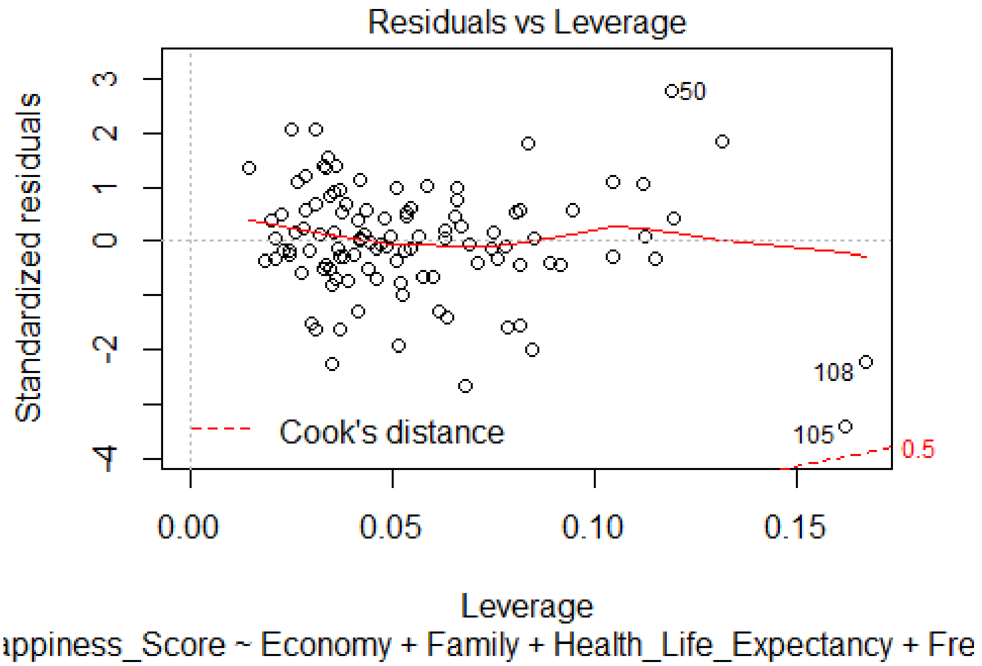
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
plot(model5)
```


Residuals vs Fitted
Residuals
1.0
0.0
-1.0
-2.0
50
94
105
3
4
5
6
7
Fitted values
ıppiness_Score ~ Economy + Family + Health_Life_Expectancy + Fre

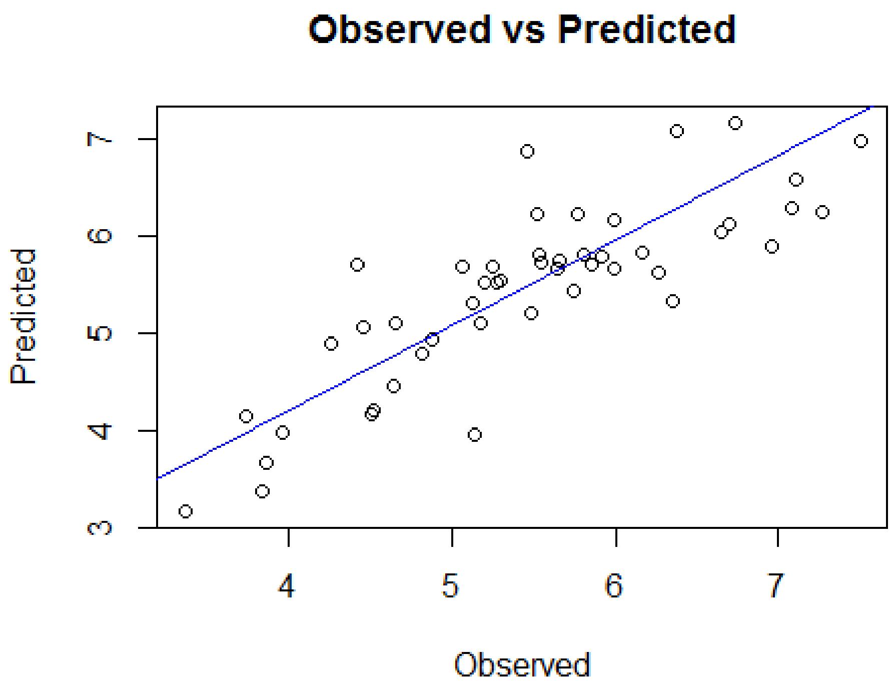

Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
-3
50
94
105
-2
-1
0
1
2
Theoretical Quantiles
ıppiness_Score ~ Economy + Family + Health_Life_Expectancy + Fre



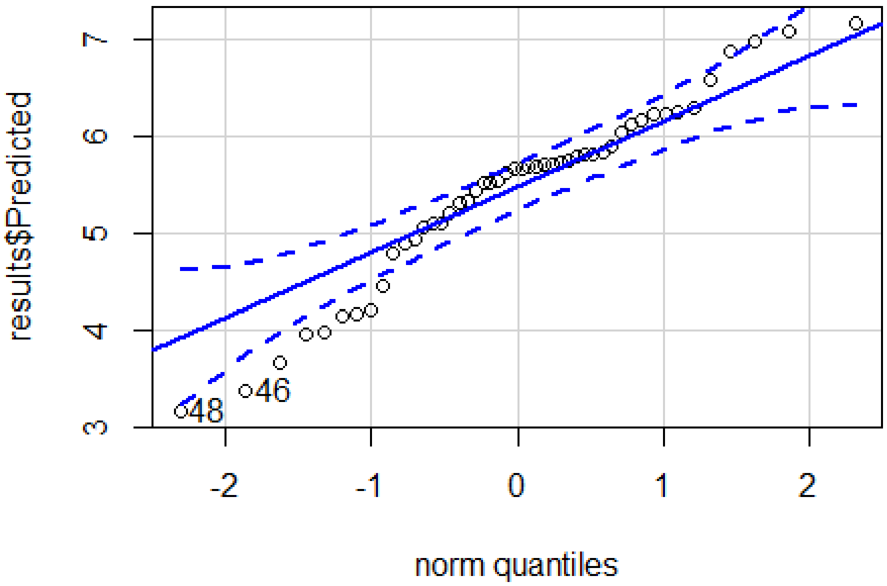


```
G3.predictions <- predict(model5,test)
results <- cbind(G3.predictions,test$Happiness_Score)
colnames(results) <- c('Predicted','Observed')
results <- as.data.frame(results)
```

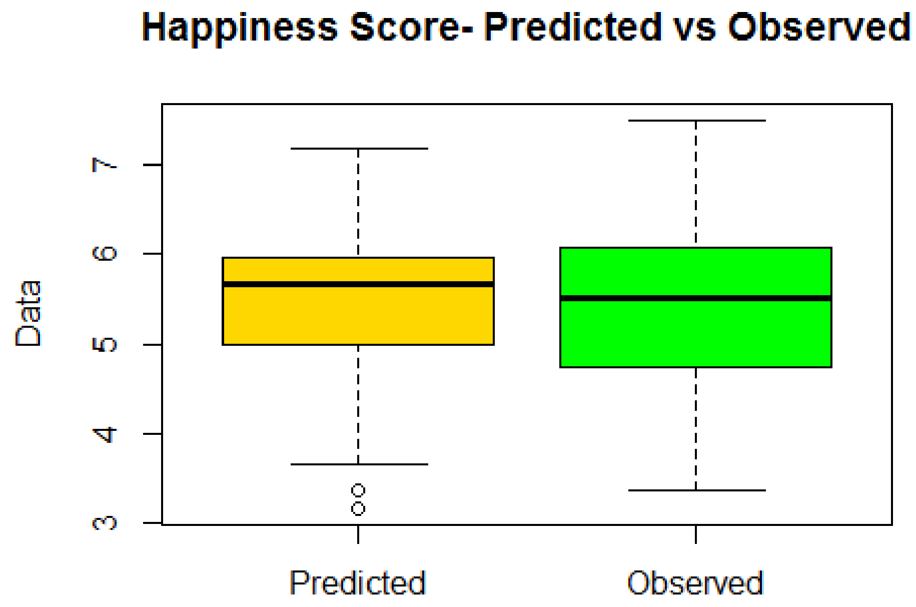
```
plot(results$Observed, results$Predicted, main = "Observed vs Predicted",
xlab = "Observed", ylab = "Predicted")
abline(lm(results$Observed ~ results$Predicted  ), col = "blue")
```



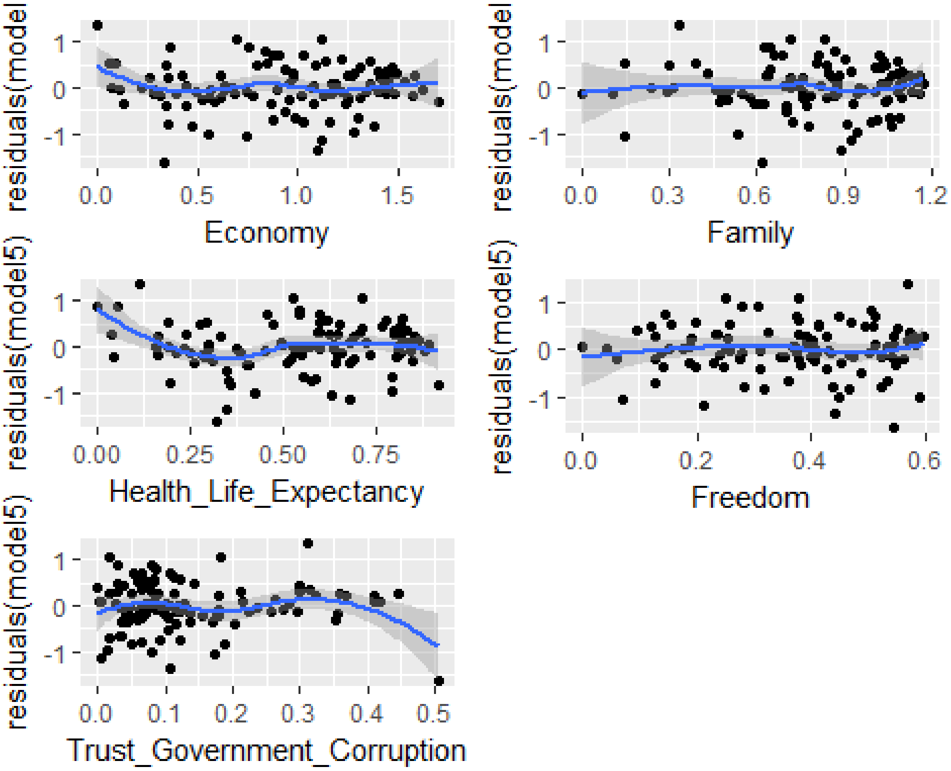
```
## Alternate way of showing the Actuals vs Predicted
##plot(results$Predicted)
##lines(results$Observed)

## Another Option
#plot_internal <- ggplot(results)+aes(x = Observed, y =
Predicted)+geom_point()+
#  +   geom_smooth(method = "lm")
# plot_internal

## if data follows normal distribution or not
qqPlot(results$Predicted, distribution = "norm")
```



```
## [1] 48 46

boxplot(results, main="Happiness Score- Predicted vs Observed",
col=(c("gold", "green")), ylab="Data", names=c("Predicted", "Observed"))
```



```
mse <- mean((results$Observed-results$Predicted)^2)
print(mse)

## [1] 0.3145682

#the result is ok model and there is no overfitting
## Multiple linear regression model for happiness score is - 2.1256 +0.6800
Economy + 1.14 Family + 1.56969  Health_Life_expectancy + 1.7247 Freedon +
1.3047 Trust Government Corruption

#### Predictor vs Residual

plot1 = ggplot(train, aes(Economy, residuals(model5))) + geom_point() +
geom_smooth()
plot2=ggplot(train, aes(Family, residuals(model5))) + geom_point() +
geom_smooth()
plot3=ggplot(train, aes(Health_Life_Expectancy, residuals(model5))) +
geom_point() + geom_smooth()
plot4=ggplot(train, aes(Freedom, residuals(model5))) + geom_point() +
geom_smooth()
plot5=ggplot(train, aes(Trust_Government_Corruption, residuals(model5))) +
geom_point() + geom_smooth()

grid.arrange(plot1,plot2,plot3,plot4,plot5,ncol=2,nrow=3)

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


residuals(model
Economy
residuals(model
Family
residuals(model5)
Health_Life_Expectancy
residuals(model5)
Freedom
residuals(model5)
Trust_Government_Corruption