World_Happiness_report

Raghavendran Shankar

19 March 2017

Loading dataset and libraries

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(ggrepel)
```

```
## Warning: package 'ggrepel' was built under R version 3.3.2
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.3.2
```

```
library(GGally)
```

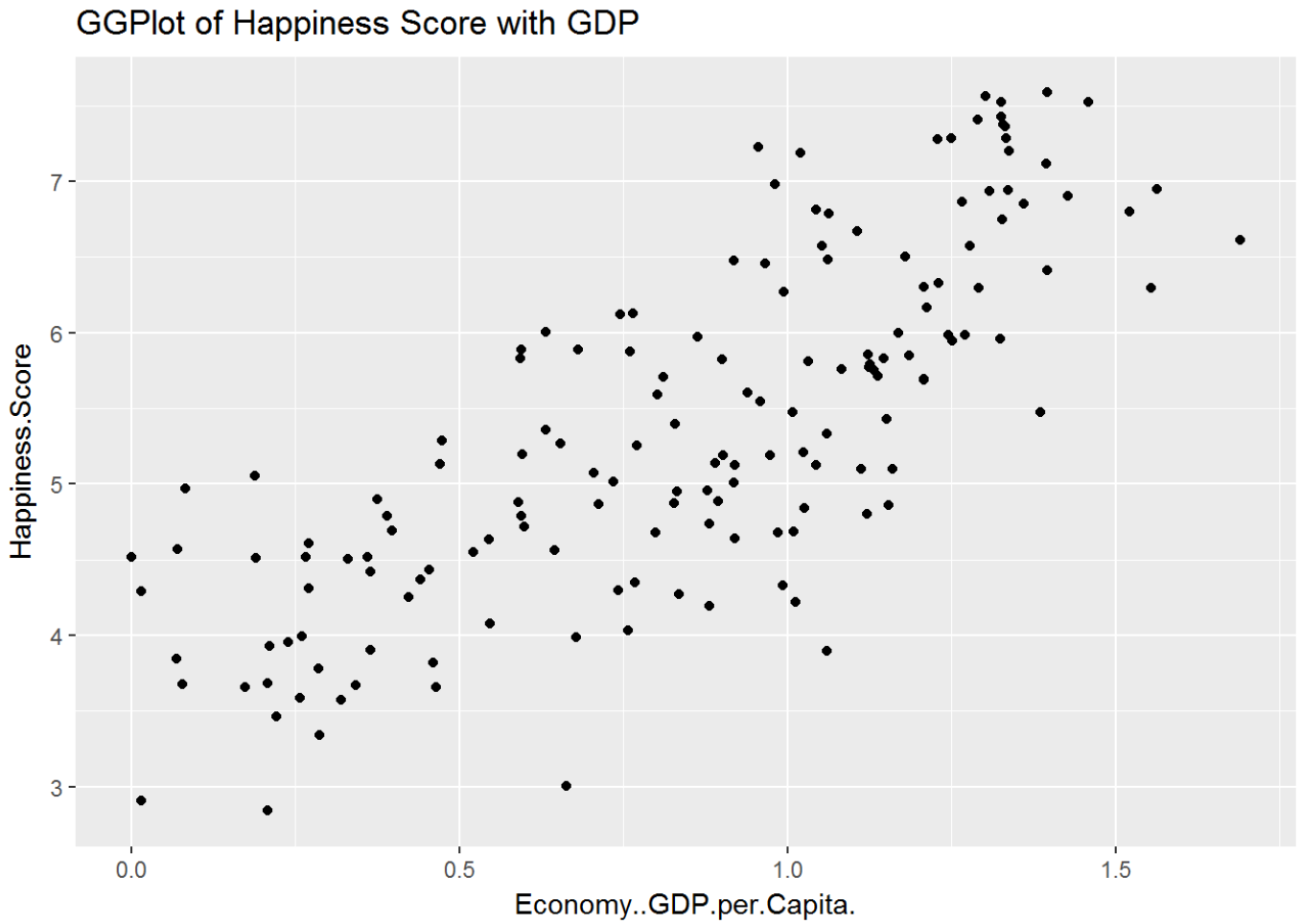
```
## Warning: package 'GGally' was built under R version 3.3.2
```

```
happy_2015 <- read.csv('2015.csv')
str(head(happy_2015))
```

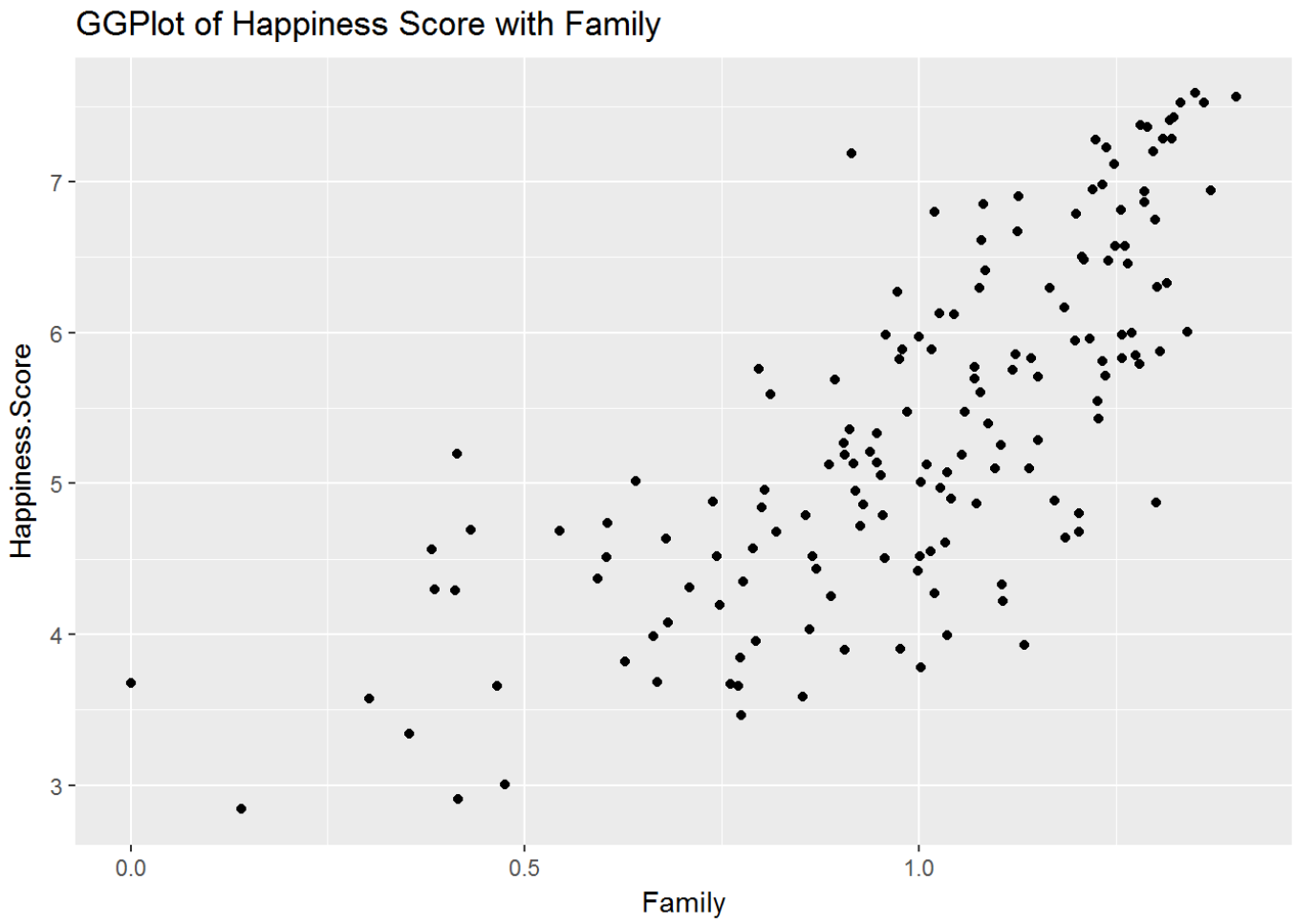
```
## 'data.frame':    6 obs. of  12 variables:
##  $ Country                      : Factor w/ 158 levels "Afghanistan",..: 136 59
38 106 25 46
##  $ Region                       : Factor w/ 10 levels "Australia and New Zealan
d",..: 10 10 10 10 6 10
##  $ Happiness.Rank               : int  1 2 3 4 5 6
##  $ Happiness.Score              : num  7.59 7.56 7.53 7.52 7.43 ...
##  $ Standard.Error               : num  0.0341 0.0488 0.0333 0.0388 0.0355 ...
##  $ Economy..GDP.per.Capita.     : num  1.4 1.3 1.33 1.46 1.33 ...
##  $ Family                       : num  1.35 1.4 1.36 1.33 1.32 ...
##  $ Health..Life.Expectancy.     : num  0.941 0.948 0.875 0.885 0.906 ...
##  $ Freedom                      : num  0.666 0.629 0.649 0.67 0.633 ...
##  $ Trust..Government.Corruption.: num  0.42 0.141 0.484 0.365 0.33 ...
##  $ Generosity                   : num  0.297 0.436 0.341 0.347 0.458 ...
##  $ Dystopia.Residual            : num  2.52 2.7 2.49 2.47 2.45 ...
```

Data Visualization with ggplots and geom points

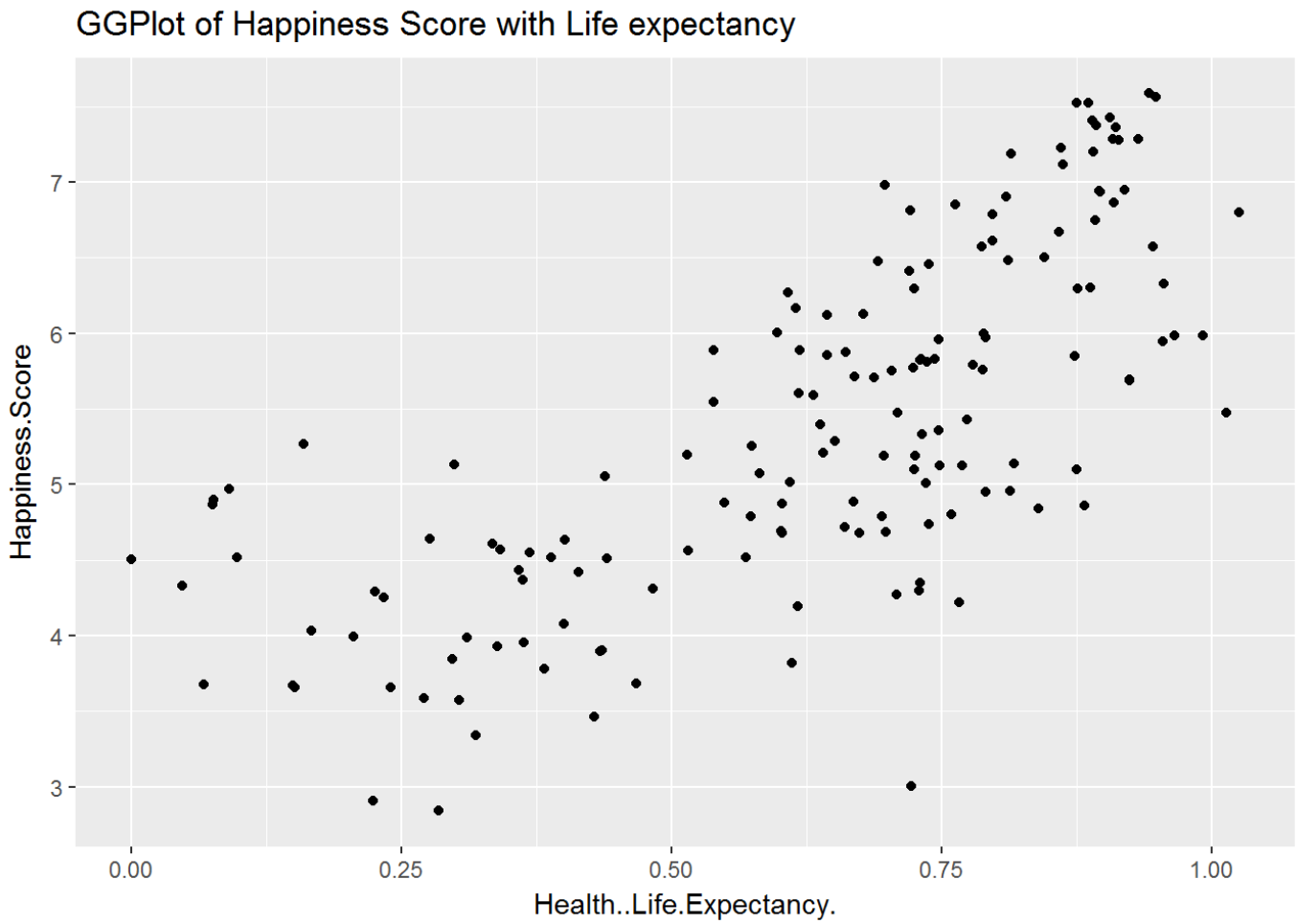
```
# ggplot for various features of Country across Happiness score
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Economy..GDP.per.Capita.))+geo
m_point()+labs(title = "GGPlot of Happiness Score with GDP")
```



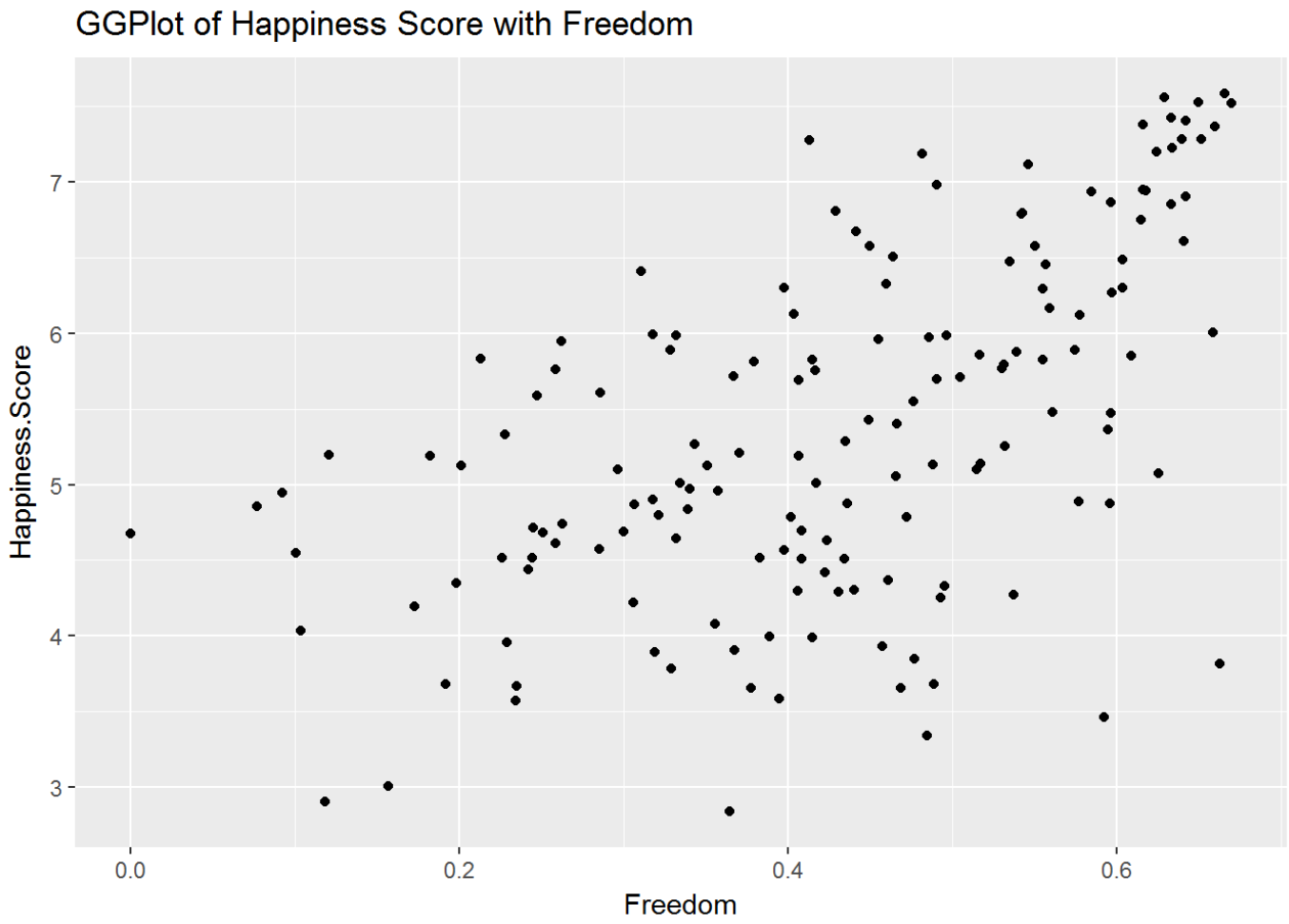

```
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Family))+geom_point()+labs(ti
tle = "GGPlot of Happiness Score with Family")
```



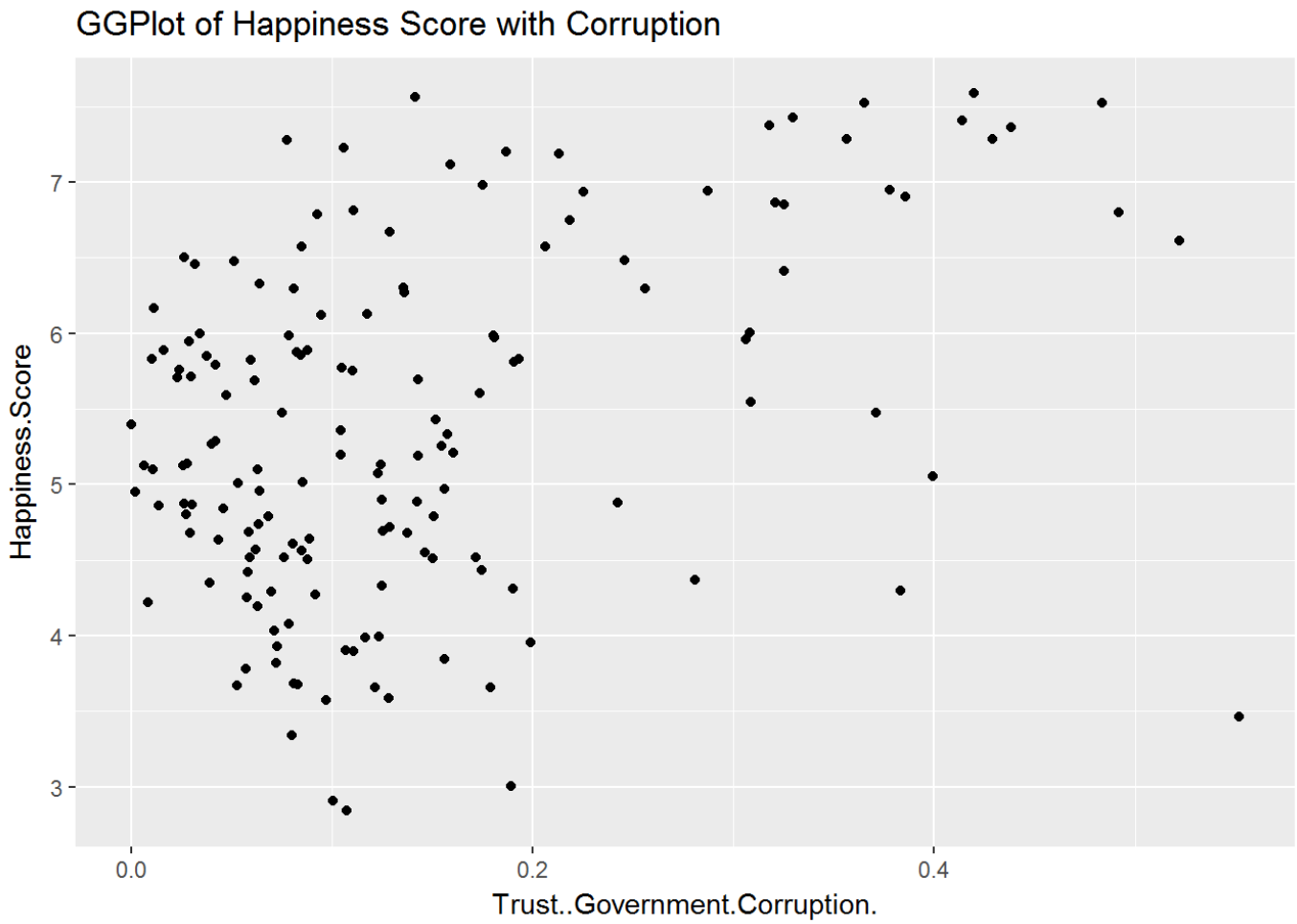

```
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Health..Life.Expectancy.))+geo
m_point()+labs(title = "GGPlot of Happiness Score with Life expectancy")
```



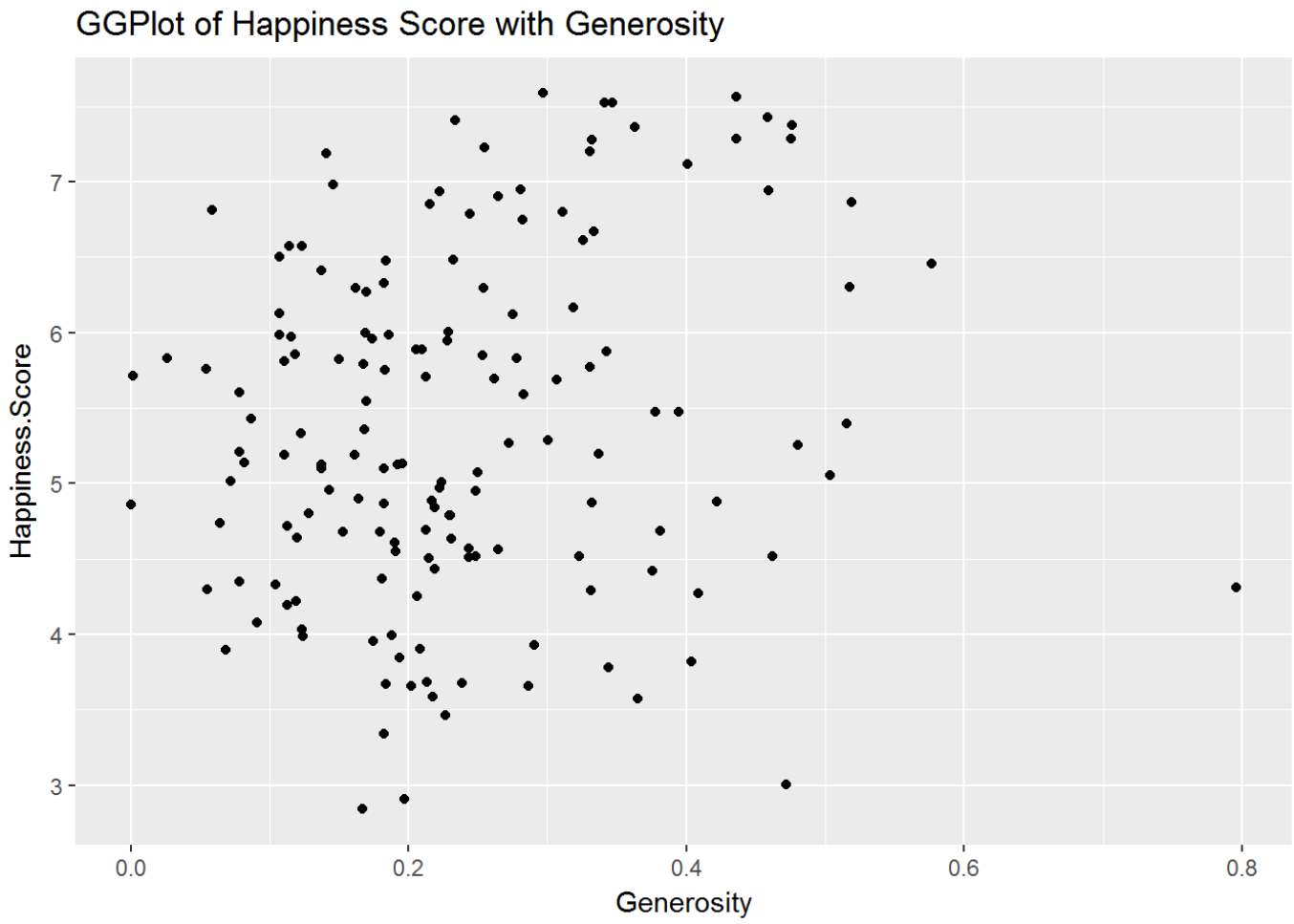

```
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Freedom))+geom_point()+labs(t
itle = "GGPlot of Happiness Score with Freedom")
```



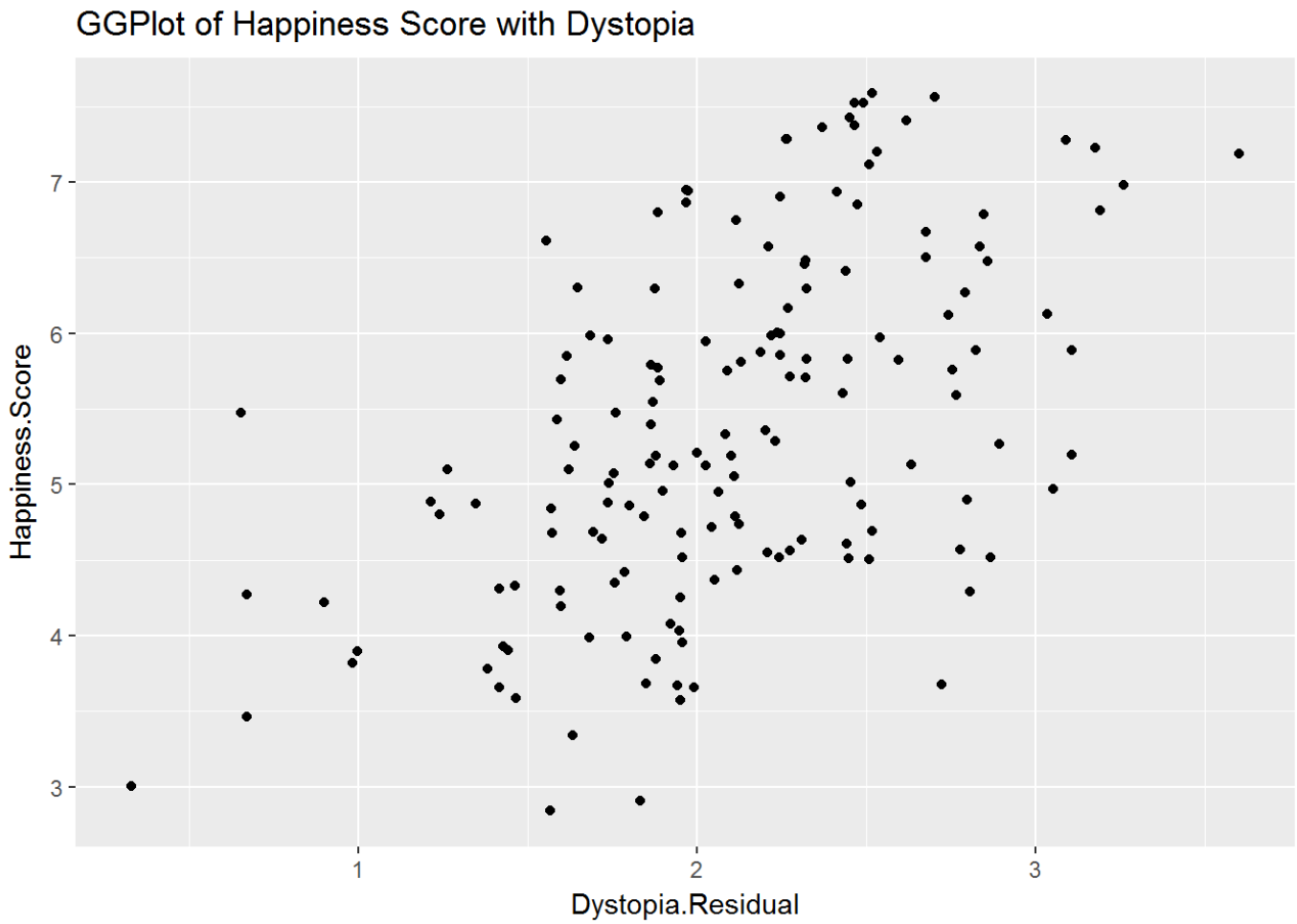

```
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Trust..Government.Corruption.)
)+geom_point()+labs(title = "GGPlot of Happiness Score with Corruption")
```




```
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Generosity))+geom_point()+lab
s(title = "GGPlot of Happiness Score with Generosity")
```



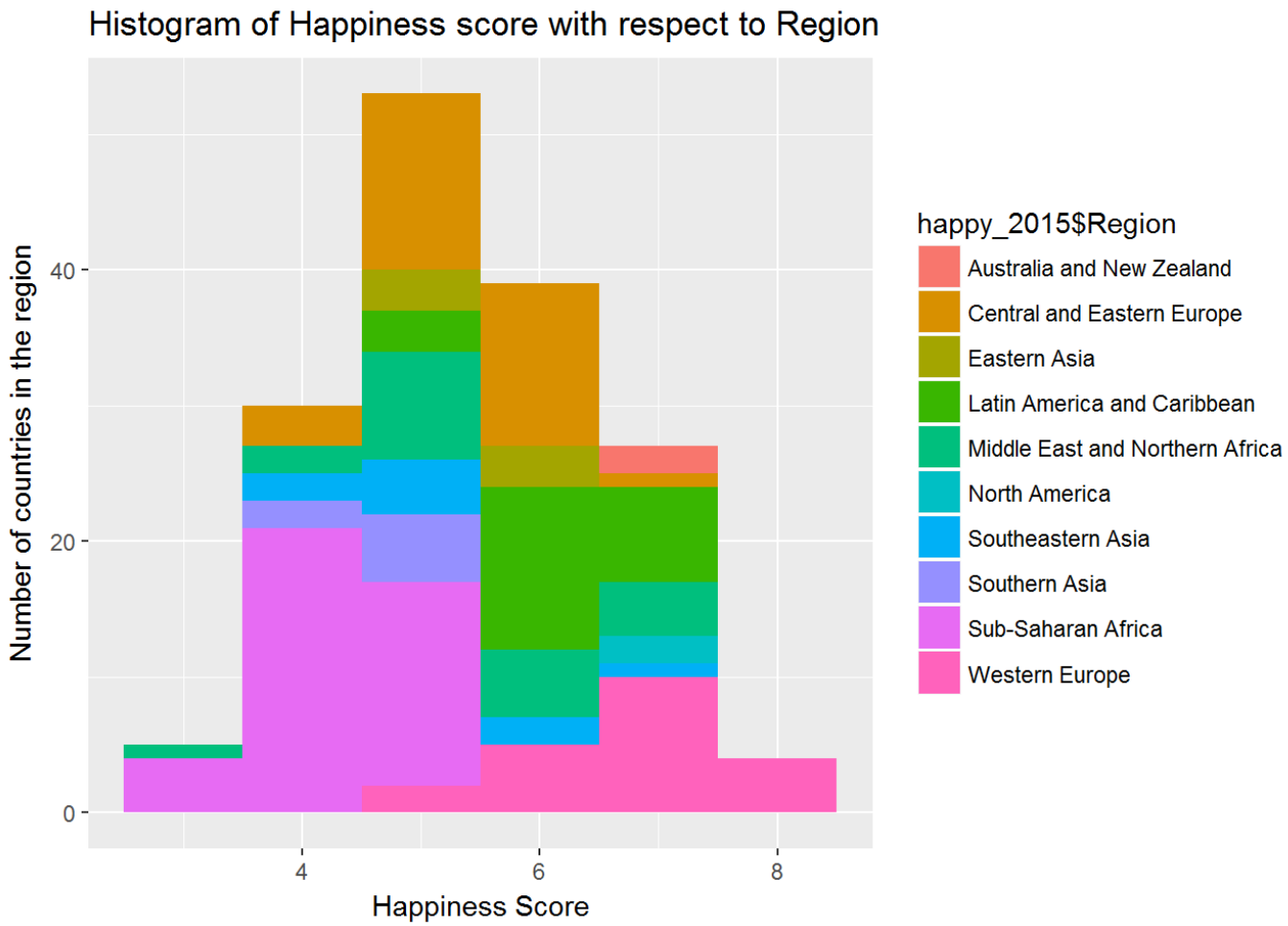

```
ggplot(data = happy_2015,aes(y = Happiness.Score,x = Dystopia.Residual))+geom_poin
t()+labs(title = "GGPlot of Happiness Score with Dystopia")
```



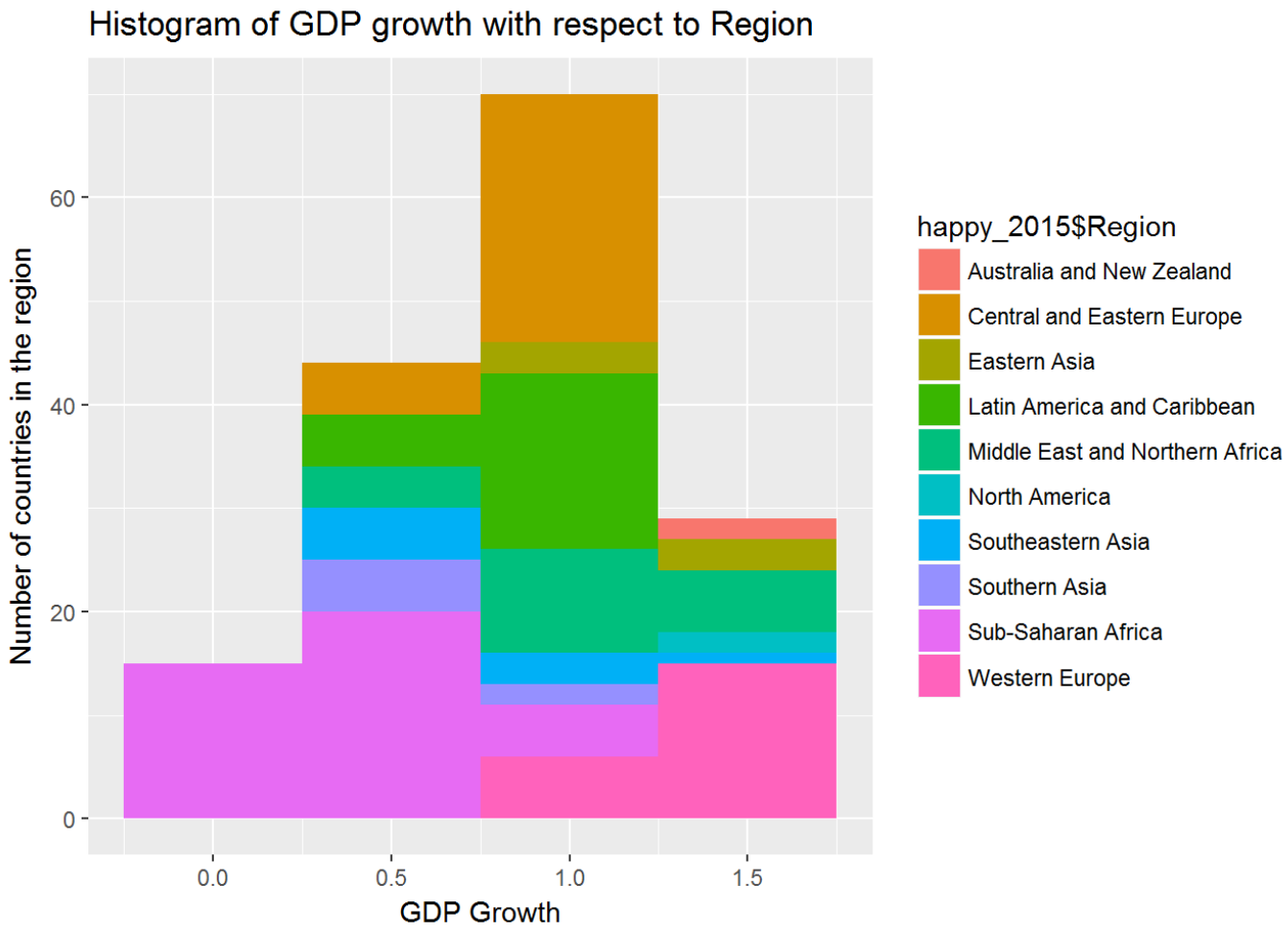

```
# Inference: There is a strong positive increase in trend between features like GDP growth, Life expectancy, Freedom and Family with Happiness score.
```

Data Visualization with ggplots and geom histogram

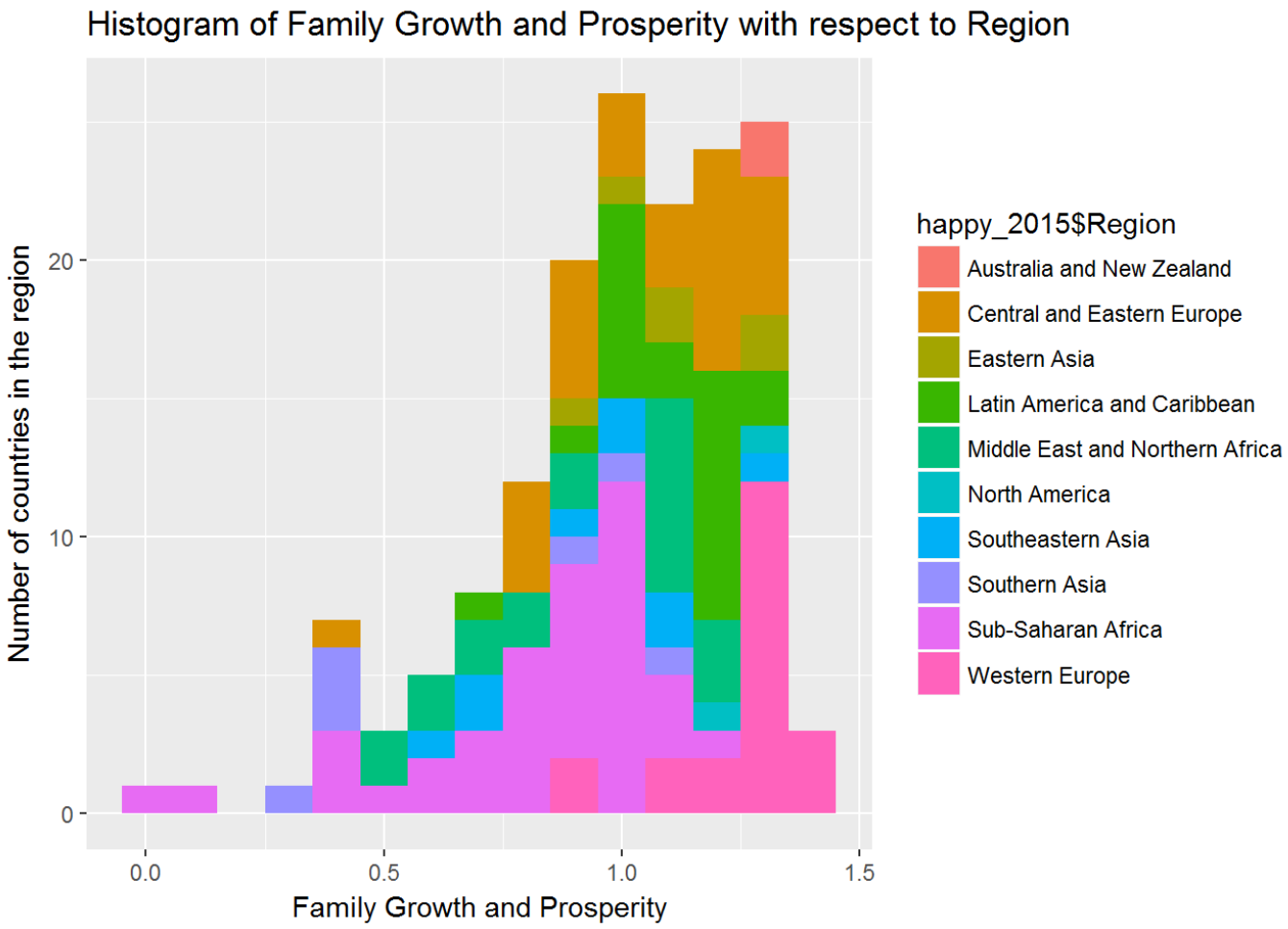
```
ggplot(melt(as.numeric(happy_2015$Happiness.Score)),aes(x=as.numeric(happy_2015$Happiness.Score),fill= happy_2015$Region)) + geom_histogram(position="stack",binwidth = 1) + labs(title = "Histogram of Happiness score with respect to Region") + xlab("Happiness Score") +ylab("Number of countries in the region")
```



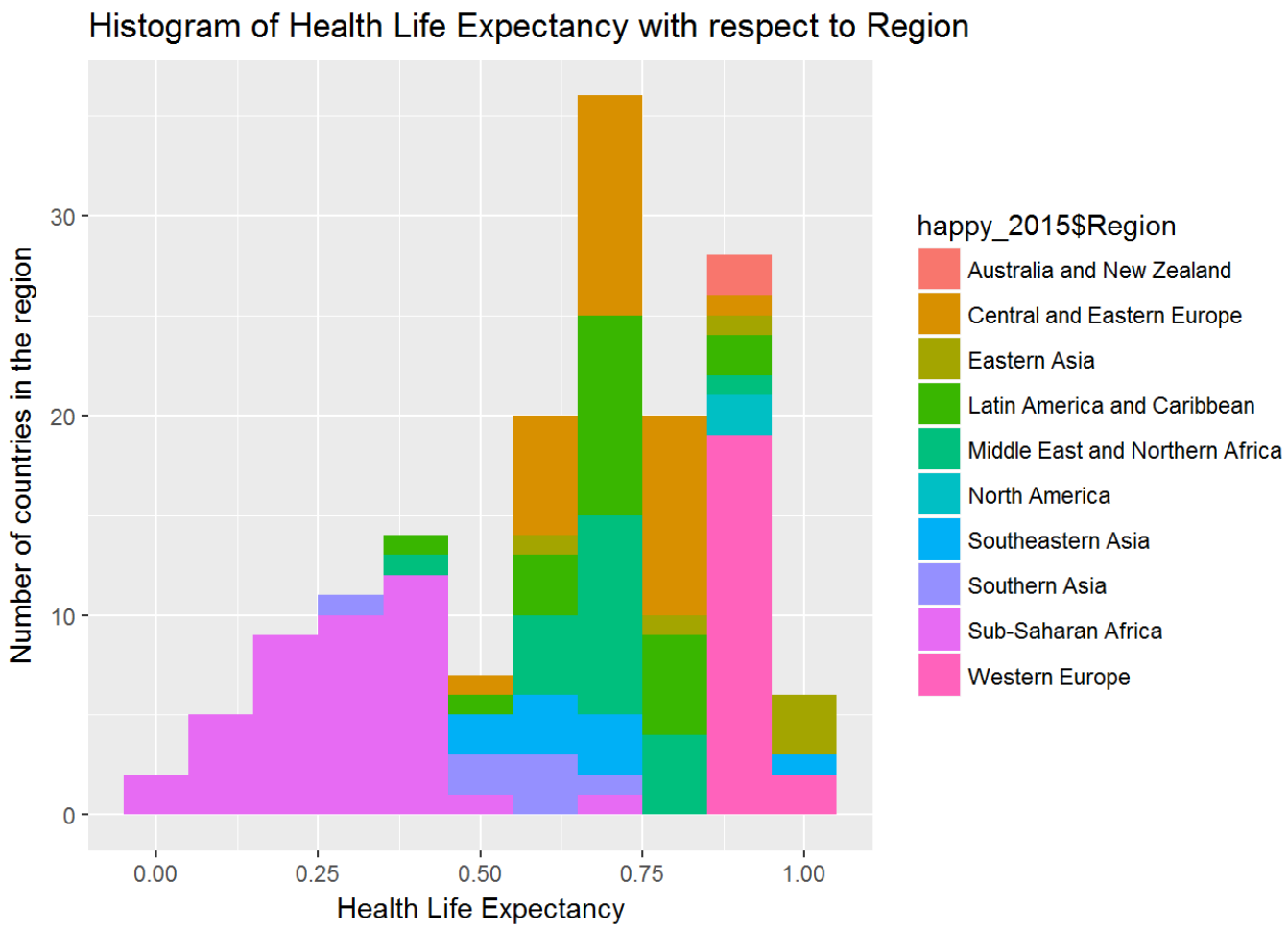

```
# Happiness score is more in Western Europe, Australia and North America and less i
n Middle East and Sub Saharan Africa

ggplot(melt(as.numeric(happy_2015$Economy..GDP.per.Capita.)),aes(x=as.numeric(happy
_2015$Economy..GDP.per.Capita.),fill= happy_2015$Region)) + geom_histogram(position
="stack",binwidth = 0.5) + labs(title = "Histogram of GDP growth with respect to Re
gion") + xlab("GDP Growth") +ylab("Number of countries in the region")
```



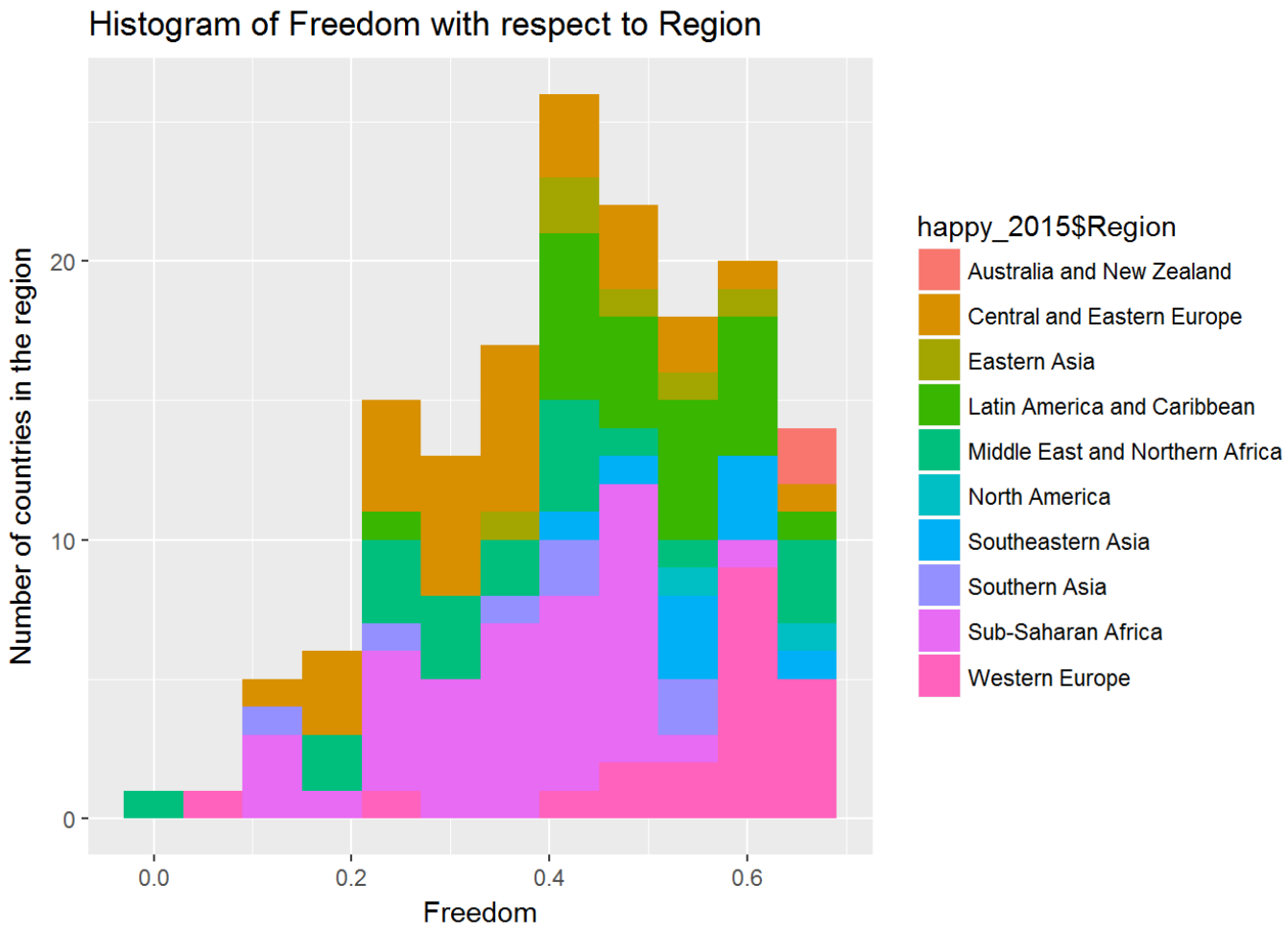

```
# GDP Growth is more in Western Europe and North America, Australia and NewZealand
and less in some parts of Latin America, southern Asia and very less in Sub-Saharan
Africa

ggplot(melt(as.numeric(happy_2015$Family)),aes(x=as.numeric(happy_2015$Family),fill
= happy_2015$Region)) + geom_histogram(position="stack",binwidth = 0.1) + labs(titl
e = "Histogram of Family Growth and Prosperity with respect to Region") + xlab("Fam
ily Growth and Prosperity") +ylab("Number of countries in the region")
```



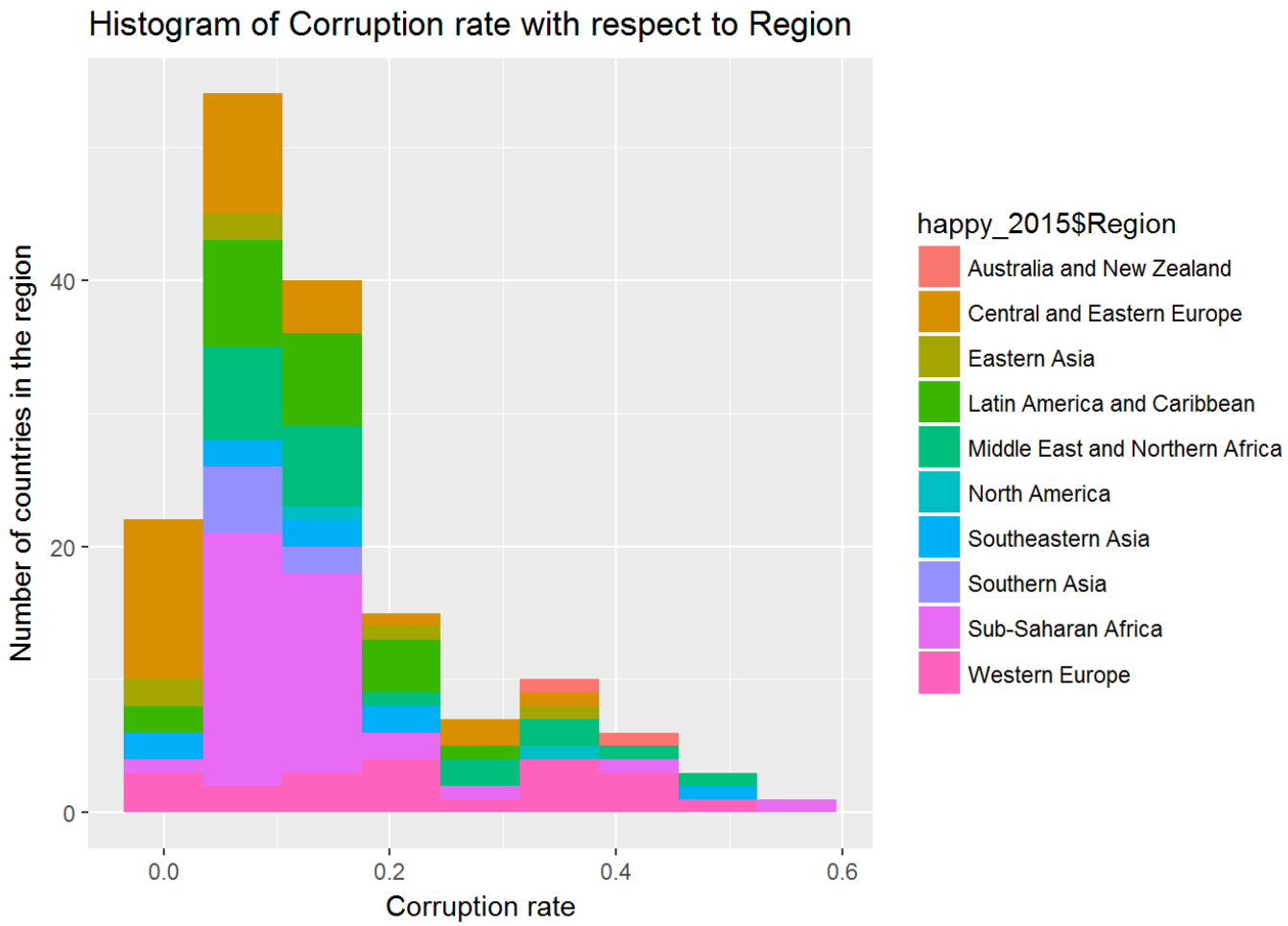

```
# Family growth and prosperity is seen most in Western Europe, North America, some parts of SouthEastern Asia and Australia and NewZealand followed by major parts in Latin America, Middle East and Central Europe.

ggplot(melt(as.numeric(happy_2015$Health..Life.Expectancy.)),aes(x=as.numeric(happy_2015$Health..Life.Expectancy.),fill= happy_2015$Region)) + geom_histogram(position="stack",binwidth = 0.1) + labs(title = "Histogram of Health Life Expectancy with respect to Region") + xlab("Health Life Expectancy") +ylab("Number of countries in the region")
```



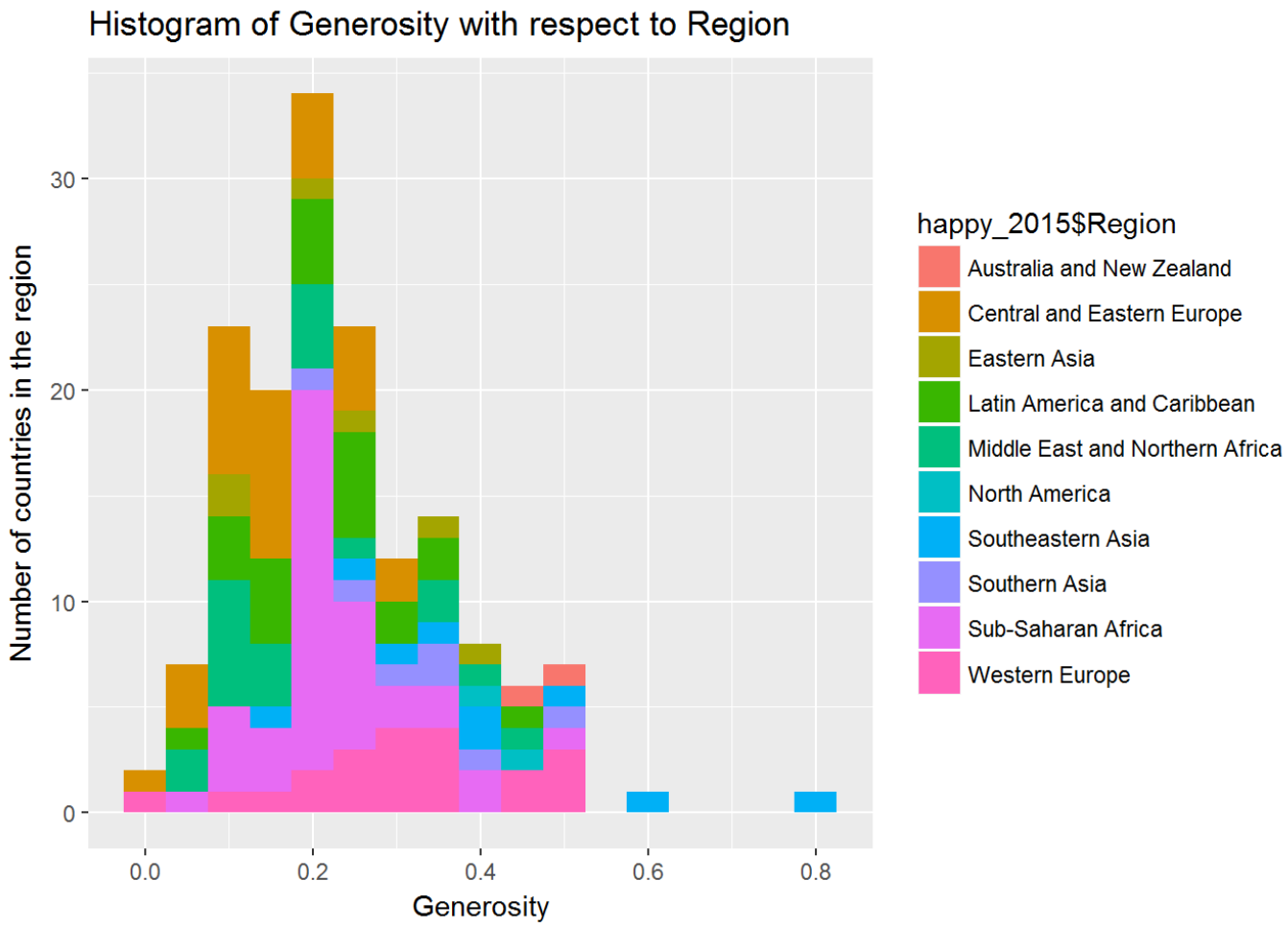

```
# Health Life Expectancy is the best in Western Europe, some parts of Eastern Asia, Australia and North America followed by Central and Eastern Europe. The least health expectancy rate lies in Sub-Sharan Africa region

ggplot(melt(as.numeric(happy_2015$Freedom)),aes(x=as.numeric(happy_2015$Freedom),fill= happy_2015$Region)) + geom_histogram(position="stack",binwidth = 0.06) + labs(title = "Histogram of Freedom with respect to Region") + xlab("Freedom") +ylab("Number of countries in the region")
```



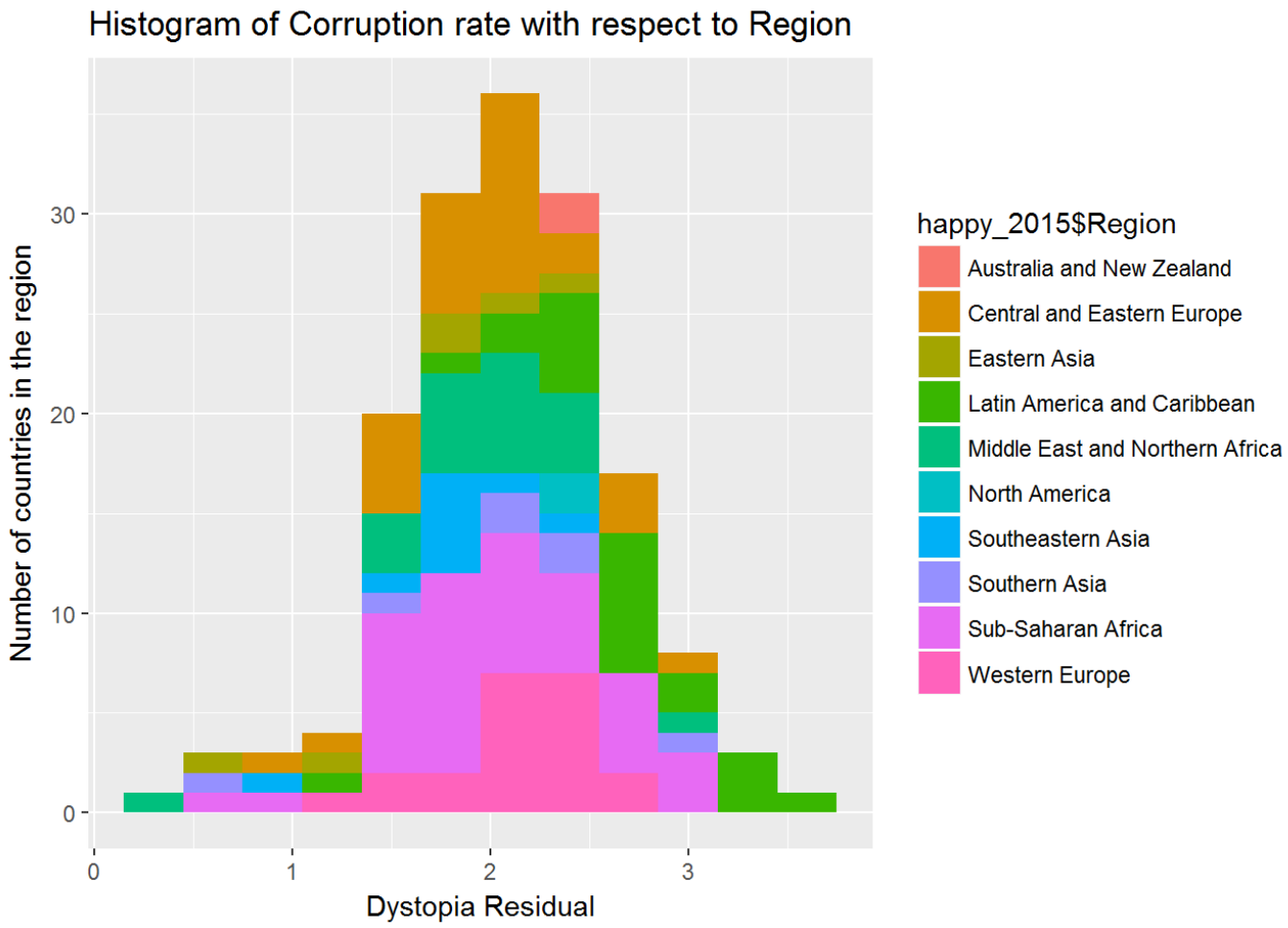

```
#Freedom is more prevalent in Western Europe, North America, South Eastern Asia and Australia followed by major parts of Latin America. The countries which does not contribute towards freedom are Iraq and Greece.

ggplot(melt(as.numeric(happy_2015$Trust..Government.Corruption.)),aes(x=as.numeric(happy_2015$Trust..Government.Corruption.),fill= happy_2015$Region)) + geom_histogram(position="stack",binwidth = 0.07) + labs(title = "Histogram of Corruption rate with respect to Region") + xlab("Corruption rate") +ylab("Number of countries in the region")
```




```
ggplot(melt(as.numeric(happy_2015$Generosity)),aes(x=as.numeric(happy_2015$Generosity),fill= happy_2015$Region)) + geom_histogram(position="stack",binwidth = 0.05) + labs(title = "Histogram of Generosity with respect to Region") + xlab("Generosity") +ylab("Number of countries in the region")
```

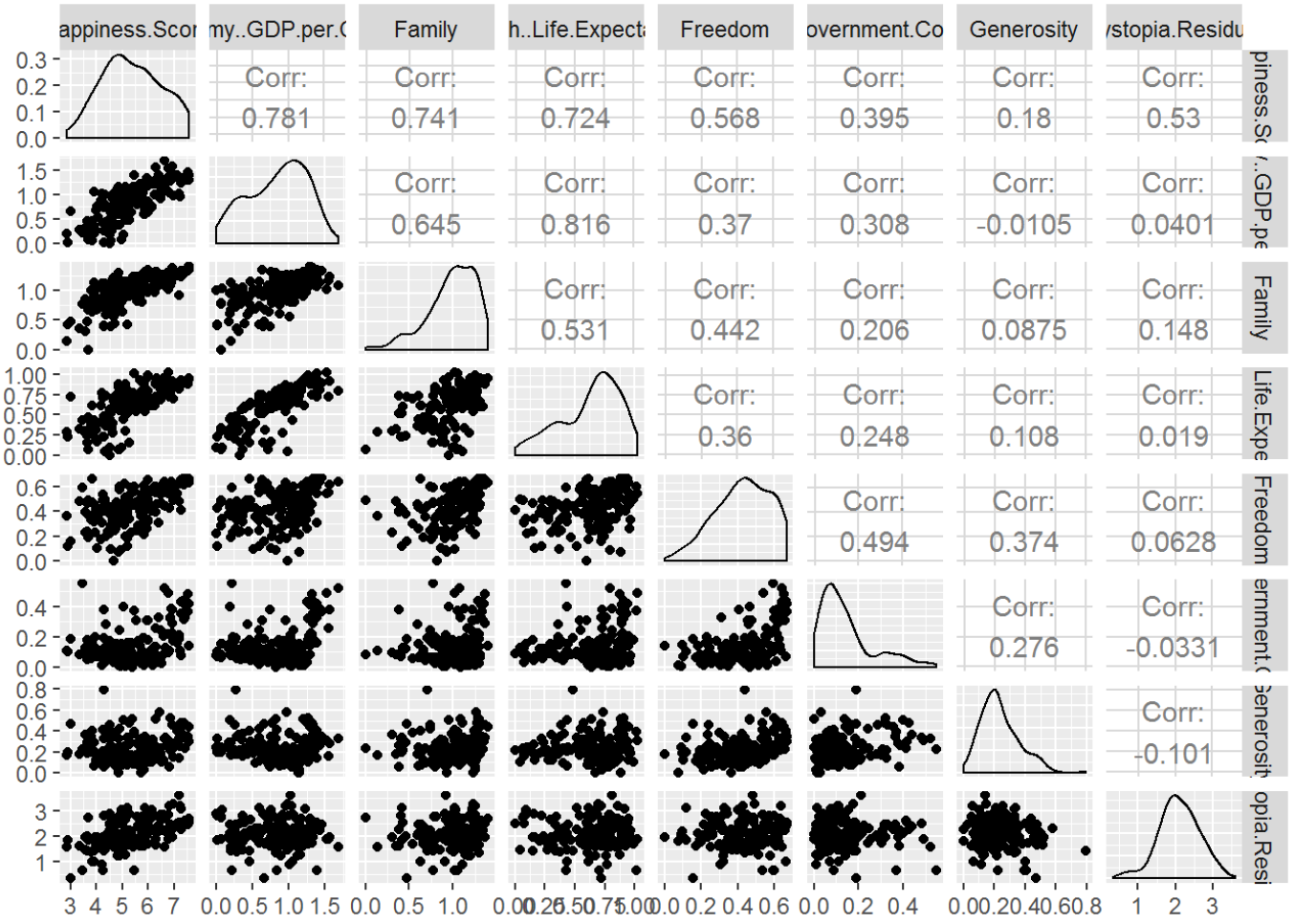



```
ggplot(melt(as.numeric(happy_2015$Dystopia.Residual)),aes(x=as.numeric(happy_2015$D
ystopia.Residual),fill= happy_2015$Region)) + geom_histogram(position="stack",binwi
dth = 0.3) + labs(title = "Histogram of Corruption rate with respect to Region") +
xlab("Dystopia Residual") +ylab("Number of countries in the region")
```



Data Visualization and Correlation

```
ggpairs(data = happy_2015,columns = c(4,6:12))
```




```
# From the correlation, it is seen that Happiness score strongly depends on GDP, Fa
mily and Life expectancy of which all the three are directly peopotional to the hap
piness score. The weak corrleations are Government corruption and Generosity.
```

Linear Regression Model for Happiness Score on 2015 dataset

```
happy_part_2015 <- happy_2015[c(-1,-2,-3,-5)]
lm_model <- lm(formula = Happiness.Score ~.,data = happy_part_2015)
summary(lm_model)
```

```
##
## Call:
## lm(formula = Happiness.Score ~ ., data = happy_part_2015)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -5.701e-04 -2.224e-04 -2.580e-06  2.471e-04  5.054e-04
##
## Coefficients:
##                                Estimate Std. Error   t value Pr(>|t|)
## (Intercept)                   6.405e-05  1.245e-04     0.514    0.608
## Economy..GDP.per.Capita.      1.000e+00  1.129e-04  8855.750   <2e-16 ***
## Family                        1.000e+00  1.153e-04  8675.863   <2e-16 ***
## Health..Life.Expectancy.      9.999e-01  1.619e-04  6175.103   <2e-16 ***
## Freedom                       9.997e-01  1.976e-04  5059.468   <2e-16 ***
## Trust..Government.Corruption. 9.999e-01  2.236e-04  4470.866   <2e-16 ***
## Generosity                    1.000e+00  2.018e-04  4956.272   <2e-16 ***
## Dystopia.Residual             1.000e+00  4.166e-05 24003.904   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002821 on 150 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 3.695e+08 on 7 and 150 DF,  p-value: < 2.2e-16
```

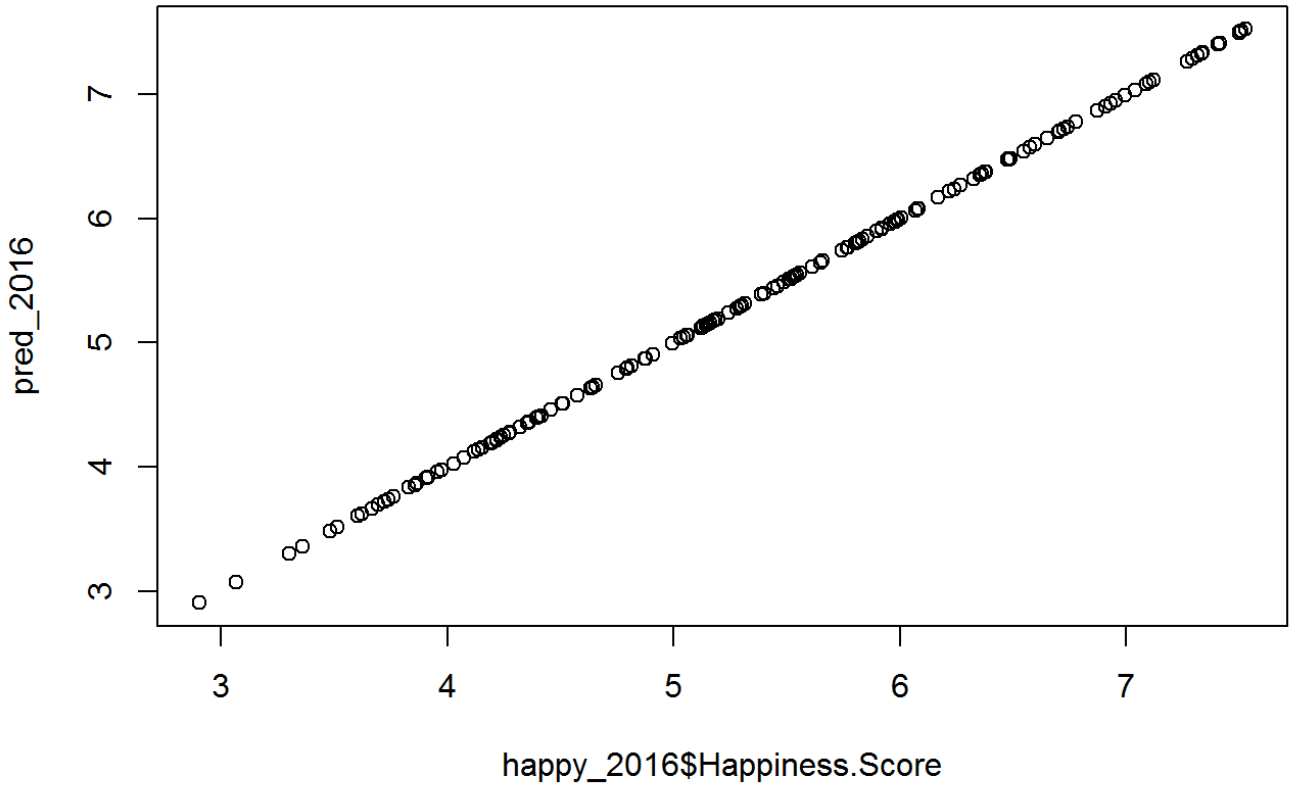
```
# The summary model has Adjusted r-squared value of 1 which is 100% accurate model.
All the predictors are significant with response variable(Happiness score).
```

Prediction of Linear model on 2016 dataset

```
happy_2016 <- read.csv('2016.csv')
pred_2016 <- predict(object = lm_model,newdata = happy_2016[7:13])# Predict Happin
ess score for 2016 data
pred_2016
```

```
##        1        2        3        4        5        6        7        8
## 7.525575 7.508631 7.501392 7.497571 7.413462 7.404020 7.339146 7.334414
##        9       10       11       12       13       14       15       16
## 7.312625 7.291362 7.267043 7.118973 7.103724 7.086692 7.038771 6.994472
##       17       18       19       20       21       22       23       24
## 6.952123 6.929438 6.906649 6.871171 6.778218 6.738688 6.725360 6.704849
##       25       26       27       28       29       30       31       32
## 6.701442 6.650174 6.596324 6.573326 6.544578 6.487913 6.480944 6.477614
##       33       34       35       36       37       38       39       40
## 6.473668 6.379070 6.379506 6.374535 6.361068 6.355004 6.323644 6.269306
##       41       42       43       44       45       46       47       48
## 6.239316 6.218179 6.167775 6.084522 6.078573 6.068420 6.005092 5.991910
##       49       50       51       52       53       54       55       56
## 5.987019 5.977583 5.976410 5.955672 5.921028 5.918656 5.897355 5.856056
##       57       58       59       60       61       62       63       64
## 5.834972 5.835176 5.822115 5.813147 5.802362 5.770523 5.767677 5.742930
##       65       66       67       68       69       70       71       72
## 5.657745 5.647805 5.615434 5.559620 5.545782 5.537958 5.528442 5.516632
##       73       74       75       76       77       78       79       80
## 5.509925 5.487965 5.458010 5.440316 5.401526 5.389520 5.313593 5.303273
##       81       82       83       84       85       86       87       88
## 5.291542 5.278960 5.245241 5.195607 5.184684 5.176889 5.163341 5.161080
##       89       90       91       92       93       94       95       96
## 5.154916 5.151255 5.145375 5.132439 5.129505 5.122898 5.120786 5.060693
##       97       98       99      100      101      102      103      104
## 5.057305 5.045466 5.032928 4.995744 4.906892 4.876084 4.875440 4.870639
##      105      106      107      108      109      110      111      112
## 4.813262 4.795360 4.792725 4.753781 4.655258 4.643129 4.634739 4.575006
##      113      114      115      116      117      118      119      120
## 4.573994 4.512629 4.508035 4.458650 4.415440 4.404439 4.395321 4.361737
##      121      122      123      124      125      126      127      128
## 4.359637 4.355713 4.324521 4.276001 4.272030 4.252344 4.235733 4.218886
##      129      130      131      132      133      134      135      136
## 4.216628 4.201317 4.192591 4.155947 4.138806 4.121544 4.073173 4.027851
##      137      138      139      140      141      142      143      144
## 3.973882 3.955697 3.915714 3.906973 3.866597 3.856156 3.832092 3.762912
##      145      146      147      148      149      150      151      152
## 3.739223 3.739354 3.723644 3.694896 3.665787 3.621743 3.607285 3.514969
##      153      154      155      156      157
## 3.483686 3.359868 3.302584 3.069181 2.904631
```

```
plot(happy_2016$Happiness.Score,pred_2016) # Compares the actual 2016 data with the predicted data
```




```
#Conclusion: The Multiple linear regression model works perfect on 2015 data and is
able to predict the happiness score of 2016 data.
```