World Happiness Report - An exploration of people's happiness levels

Alhad Barbadikar

May 4, 2019

Abstract

When people move across income levels, they are generally happier, materially. Beyond an extent, material goods add little benefit to the well being of an individual. A society is but a simple collection of individuals.Neither one can exist or act in isolation. If a society is to be made happier, it is necessary to find a method to measure happiness. My aim was to explore the dataset & see what influences happiness. Upon doing so,I realized that the variables available could be used for exploration & maybe not for statistical analysis. Here, every country's happiness score was a sum of the other variables measuring it. There were clear differences across regions of the world & I do believe that through this dataset it is possible to understand the needs of people across different regions & help governments do what they should be doing, helping society progress by understanding what makes individuals happier.

1 Introduction

The World Happiness Report is a landmark survey of the state of global happiness that ranks 156 countries by how happy their citizens perceive themselves to be. Every year it focuses on the conditions around the world & connects them to how happiness might be perceived by people living in such a state. While perception may not be an accurate representation of the reality but it does provide a window into an individual's thought. This is further strengthened by the fact that the survey consists of a standard set of core questions that has been translated into major languages of the respective country[1]. The survey is conducted on approximately 1,000 residents in more than 150 countries, using randomly selected, nationally representative samples[2]. I believe that the sample size used in coming up with the estimates of the variables for each country was sufficient to provide convicing evidence of correlations, if not cause-effect relationships. This could help policy makers make important decisions & move away from using GDP or GNP as metrics of social progress & move towards a more inclusive & progressive society by working on increasing happiness of people.

It was the view of Ronald Raegan that, "Governments tend not to solve problems, only rearrange them"[4]. It could be said so in the case of United States, where the economic progress hasn't done much to help with the well-being of people. The problem as rightly pointed by Easterlin is that while richer individuals are in general happier than poorer ones, over time they are going to compare their living conditions to somebody else's. This means that relatively, things remain unchanged[5].This coupled with the disproportionate gains & societal factors like insecurity, loss of social trust, a declining confidence in government & quite rightly the dimnishing returns of income levels makes one wonder if improving on economic indicators has any benefits at all. Measuring a country's well-being using these economic indicators provides an incomplete picture of the conditions of the people. While we have been making great strides across the world in areas of technology, there is also enough evidence that these strides have been accompanied by serious problems of obesity, smoking, diabetes, depression & other ills of modern life[2].

World Happiness Score in 2017
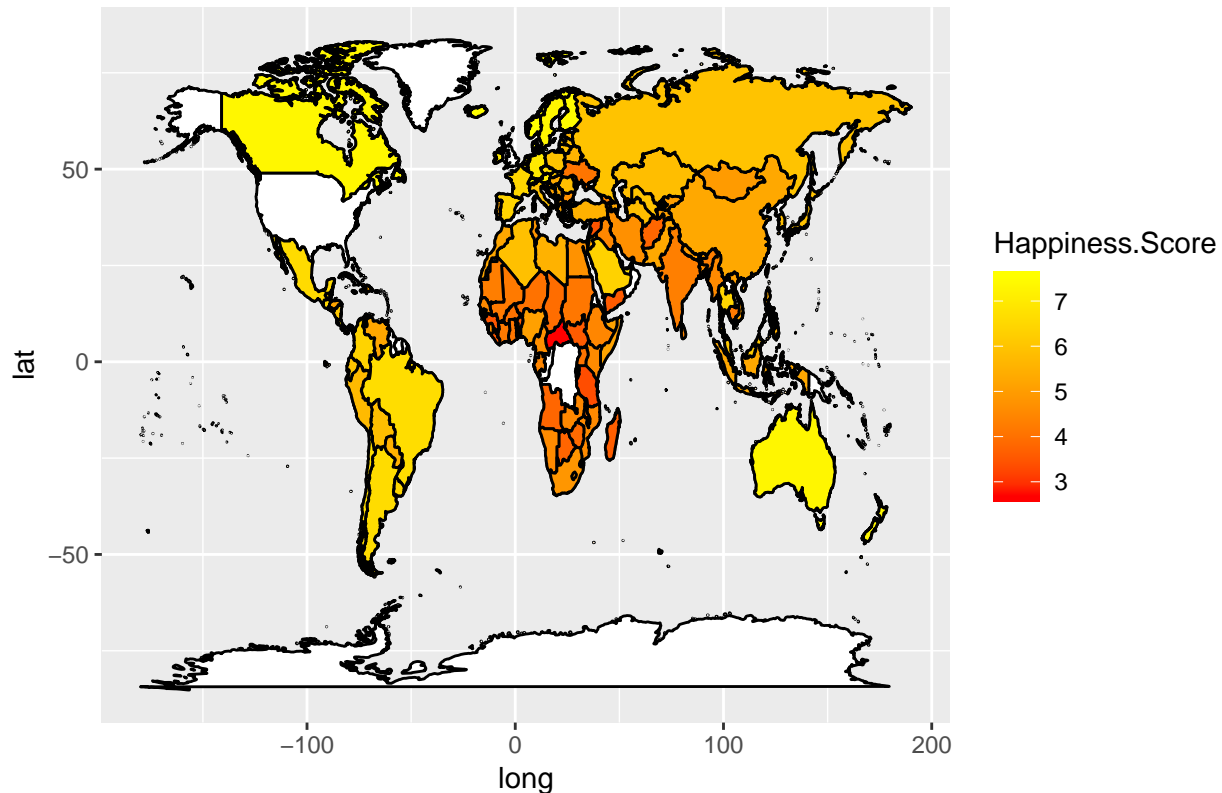
It seems clear from Figure[1] that some Nordic countries & a few European countries are generally happier than the rest of the world. The European economy comprises of more than 740 million people in 50 different countries. It could be that the creation of Euro as an accepted common currency has brought the countries closer through the convenience of a shared currency & has helped in ensuring a stronger cash flow[3]. This thought of reasoning has been criticized & it has been suggested that the happiness of European countries probably comes down to high social welfare spending & less productive hours in these countries. On a PPP basis, the United States is a much better place in comparison to Europe & it simply could be that the happiness of these countries is due to their cultural values & a high-trust society[1].

The findings from the report then cannot be simply put down to policy making & there probably needs more examination as to how that can be achieved. What I did end up doing is understand the contribution of these variables to the happiness scores.

2 Related Work

Happiness has been a subject of much interest since the age of man. It is argued that subjective well-being is a satisfactory empirical approximation to individual utility and that happiness research is able to contribute important insights for economics[2] Bruno S Frey & Alois Stutzer found in their study that the effects flowing directly from the quality of instituitions are often much larger than those that flow through productivity & economic growth[3]. This perhaps is an important result in understanding the overreaching role of governments on indivduals. John Heinwell helped shape the distinction between economy and well-being. His opinion on well-being is closely related to the concept of social progress. He documented cases where social capital saved lives & pointed out how inadequate economic indicators can be. This is very clearly put forth in his example where unemployment lowers the subjective well being by more than the usual measure of economic cost & certainly more than inflation[4]. It could not have been put down in simpler words. A study conducted by Antonella Delle Fave, Ingridar, Teresa Freire, Dianne Vella-Brodrick & Marié P. Wissing on 666 participants

also aimed to examine definitions & experiences of happiness using open-ended questions, which is quite similar to the aim of the World Happiness Report. They made use of the Satisfaction with Life Scale(SWLS) to quantitatively assess the hedonic dimension of happiness. Based on their study, they defined happiness as a condition of psychological balance & harmony[5].

Now that we've established happiness being the measurement of subjective well-being, it becomes important to highlight the differences between cognitive life evaluations (represented by questions asking how happy or satisfied people are with their lives as a whole), and emotional reports[6]. What is important they say for any survey is to understand the importance of the question being asked. They further go on to state that classification of happiness could be done in 3 ways:

1. "How happy are you now?" - a current emotional report
2. "How happy were you yesterday" - as a remembered emotion
3. "How happy are you with your life as a whole these days" - as a form of life evaluation

The survey conducted by the Gallup world Poll used the 3rd question which asks the respondent to evaluate their life. This answer is evaluated on a 11-point ladder scale from 0 to 10, with the bottom rung of a ladder(0) being the worst possible life for them & (10) being the best possible.

Since 2012, a report is published each year exploring aspects of human life & their connection to their current level of happiness[7]. I aimed to explore this dataset & understand how happy people are across the world. From a statistical point of view, I do not believe that the variables used in this survey were measures that reflected the reality of a country. On checking the survey method, it seems that this was a complicated method & was concerned with firstly building a happiness score & then the contribution of each variable to this score. This was an interesting exploration & I do believe that there are differences in perceptions of happiness across the world & different regions need to work on different things.

3 Analysis

I was working on datasets that measured happiness on the Cantril Ladder across the years 2015, 2016 & 2017. Here is the structure of the datasets:

```
## 'data.frame':    158 obs. of  12 variables:
##  $ Country          : Factor w/ 158 levels "Afghanistan",..: 136 59 38 106 25 46 100 135 101 7 ...
##  $ Region           : Factor w/ 10 levels "Australia and New Zealand",..: 10 10 10 10 6 10 10 10 1 1
##  $ Happiness_Rank   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ 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          : num  1.4 1.3 1.33 1.46 1.33 ...
##  $ Family           : num  1.35 1.4 1.36 1.33 1.32 ...
##  $ Health           : num  0.941 0.948 0.875 0.885 0.906 ...
##  $ Freedom          : num  0.666 0.629 0.649 0.67 0.633 ...
##  $ Trust            : 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.frame':    157 obs. of  13 variables:
##  $ Country          : Factor w/ 157 levels "Afghanistan",..: 38 135 58 104 45 26 98 99 7 134 ...
##  $ Region           : Factor w/ 10 levels "Australia and New Zealand",..: 10 10 10 10 10 6 10 1 1 10
##  $ Happiness_Rank   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Happiness_Score  : num  7.53 7.51 7.5 7.5 7.41 ...
##  $ Lower_CI         : num  7.46 7.43 7.33 7.42 7.35 ...
##  $ Upper_CI         : 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           : num  0.795 0.863 0.867 0.796 0.811 ...
```

```
## $ Freedom          : num  0.579 0.586 0.566 0.596 0.571 ...
## $ Trust            : 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.frame':    155 obs. of  12 variables:
## $ Country          : Factor w/ 155 levels "Afghanistan",..: 105 38 58 133 45 99 26 100 132 7 ...
## $ Happiness_Rank   : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Happiness_Score  : num  7.54 7.52 7.5 7.49 7.47 ...
## $ Lower_CI         : num  7.59 7.58 7.62 7.56 7.53 ...
## $ Upper_CI         : num  7.48 7.46 7.39 7.43 7.41 ...
## $ Economy          : num  1.62 1.48 1.48 1.56 1.44 ...
## $ Family           : num  1.53 1.55 1.61 1.52 1.54 ...
## $ Health           : num  0.797 0.793 0.834 0.858 0.809 ...
## $ Freedom          : num  0.635 0.626 0.627 0.62 0.618 ...
## $ Trust            : num  0.362 0.355 0.476 0.291 0.245 ...
## $ Generosity       : num  0.316 0.401 0.154 0.367 0.383 ...
## $ Dystopia_Residual: num  2.28 2.31 2.32 2.28 2.43 ...
```

The columns here as stated before estimate the extent to which each of the six factors - economic production, family support, life expectancy, freedom, absence of corruption & generosity contributes to making the subjective well-being estimates being higher than they are in Dystopia - a hypothetical country that has the world's lowest national averages for each of the respective scores. For my analysis, I removed the columns Standard error from the 2015 dataset, {df_2015}, lower confidence & higher confidence intervals from the 2016 & 2017 datasets respectively {df_2017}.

## Happiness score across Years
Data lines are Happiness scores of countries across years
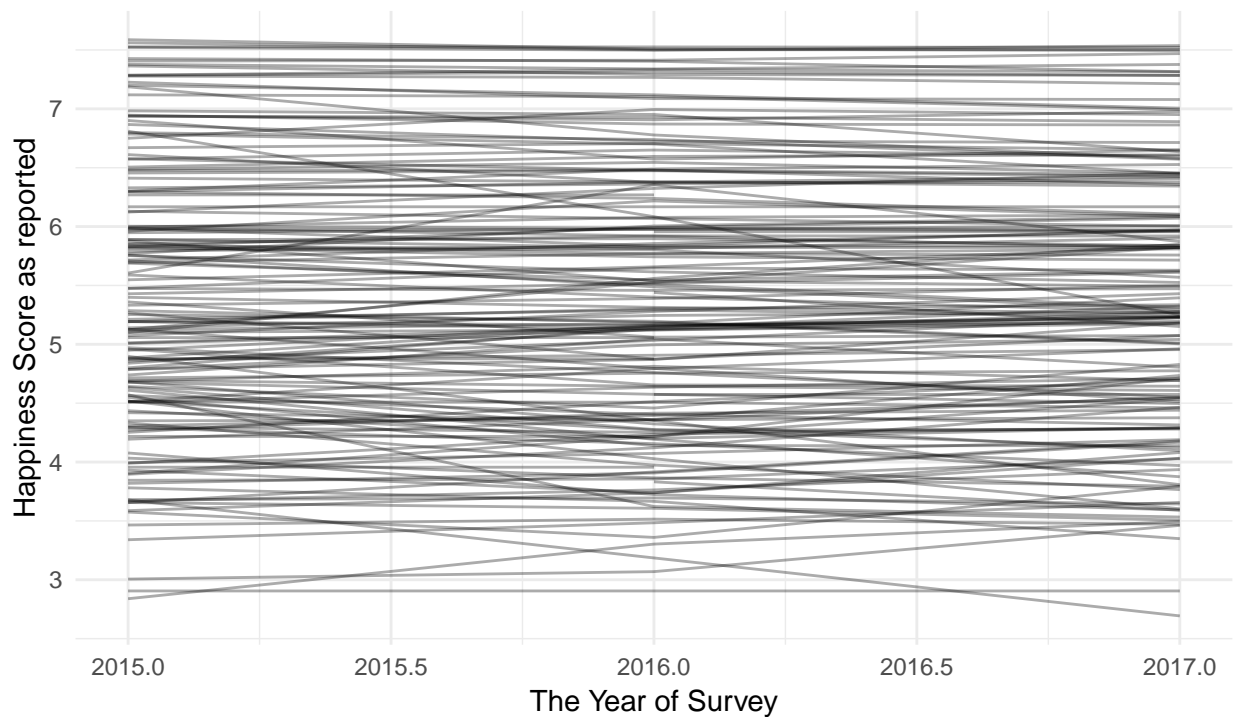
Source: Gallup

What I was trying to understand was how happiness scores have moved across the 3 years for countries. Not

surprisingly, considering the narrow timeframe, there hasn't been much movement of countries. Looking across regions & the mean happiness scores shows that across these 3 years,mean happiness scores haven't changed much.

Means of happiness scores across 10 regions
Region vs Happiness Score

source: mpg
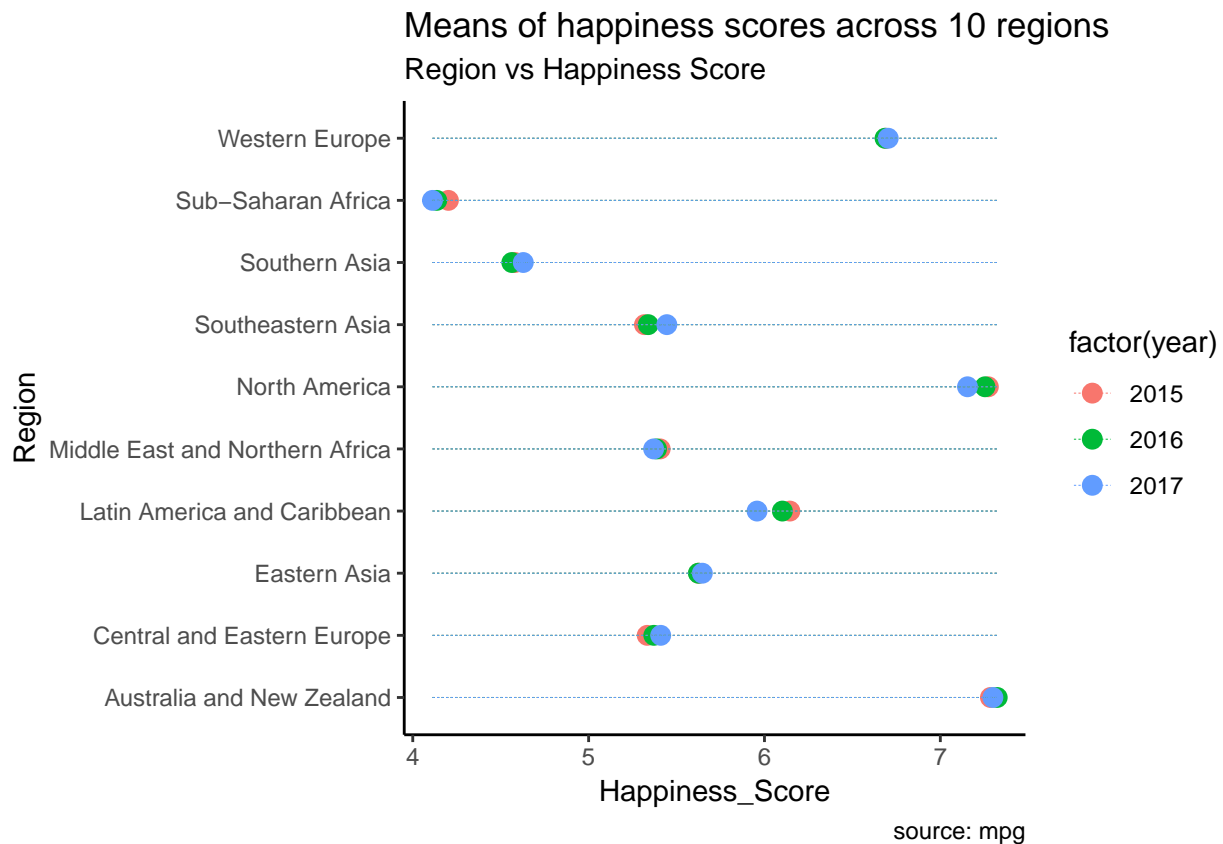
There wasn't a necessity to transform any of the variables. So, I began with my initial idea to simply perform a simple regression. For this, R has a facility to create nested models. I created nested models for each of the countries & ran a simple linear regression of the formula:

$Happiness_Score = $Economy + $Family + $Health + $Freedom + $Trust + $Generosity + $Dystopia_Residual

The nested models for each country gave a R-squared value of 1 but I couldn't make any interpretation of the coefficients which is why I decided to move ahead to a different analysis. Also, there were countries for which a nested linear regression model was a bad fit.

Since I was focused at this point to get the model right across countries, I decided to drop the regions out & proceed.
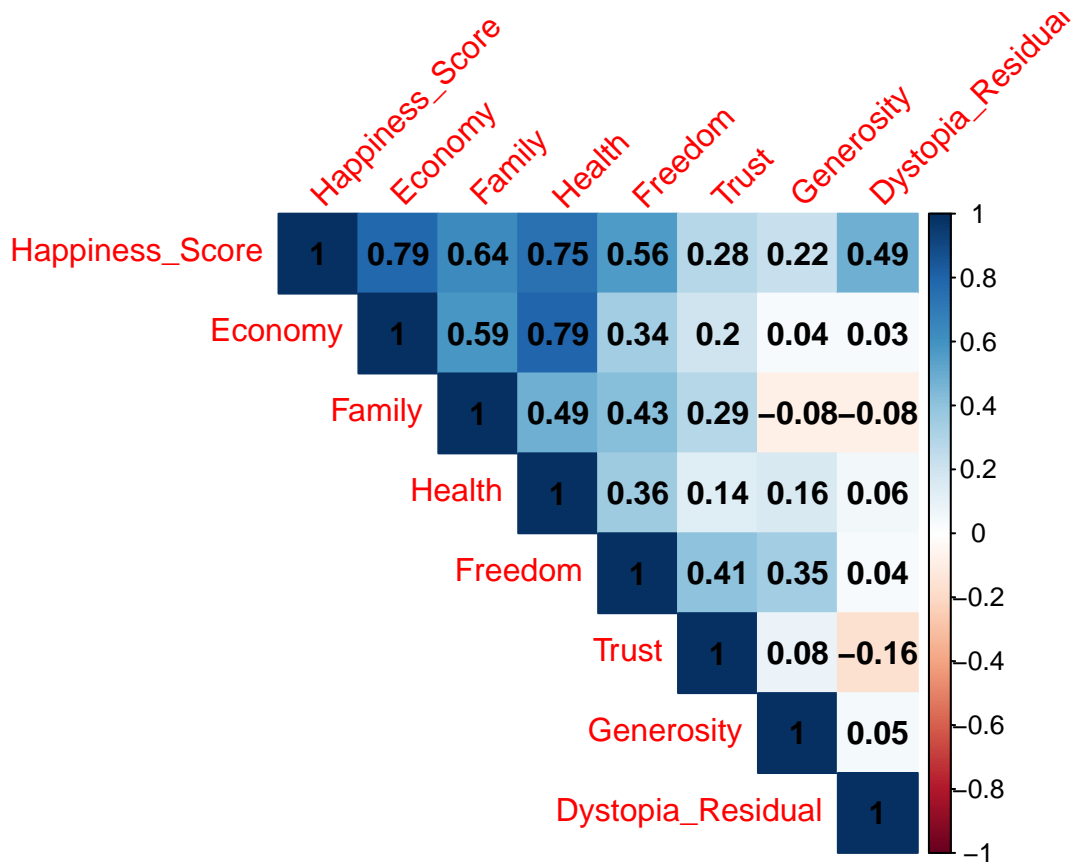
Happiness Score is highly correlated with the economic conditions & life-expectancy more than the rest of the variables. It would make sense that for a country, ensuring conditions for a high life-expectancy & maintaining a stable economy were most correlated.

```
##
## Call:
## lm(formula = Happiness_Score ~ ., data = training_set)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -5.625e-04 -2.322e-04 -7.260e-06  2.559e-04  5.070e-04
##
## Coefficients:
##                     Estimate Std. Error   t value Pr(>|t|)
## (Intercept)        1.525e-04  8.106e-05     1.882   0.0607 .
## Economy            1.000e+00  6.139e-05 16291.251   <2e-16 ***
## Family             1.000e+00  6.156e-05 16244.386   <2e-16 ***
## Health             9.998e-01  1.034e-04  9671.997   <2e-16 ***
## Freedom            1.000e+00  1.274e-04  7852.590   <2e-16 ***
## Trust              1.000e+00  1.257e-04  7953.836   <2e-16 ***
## Generosity         1.000e+00  1.276e-04  7834.715   <2e-16 ***
## Dystopia_Residual  1.000e+00  2.701e-05 37024.462   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002864 on 368 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
```

```
## F-statistic: 8.396e+08 on 7 and 368 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression
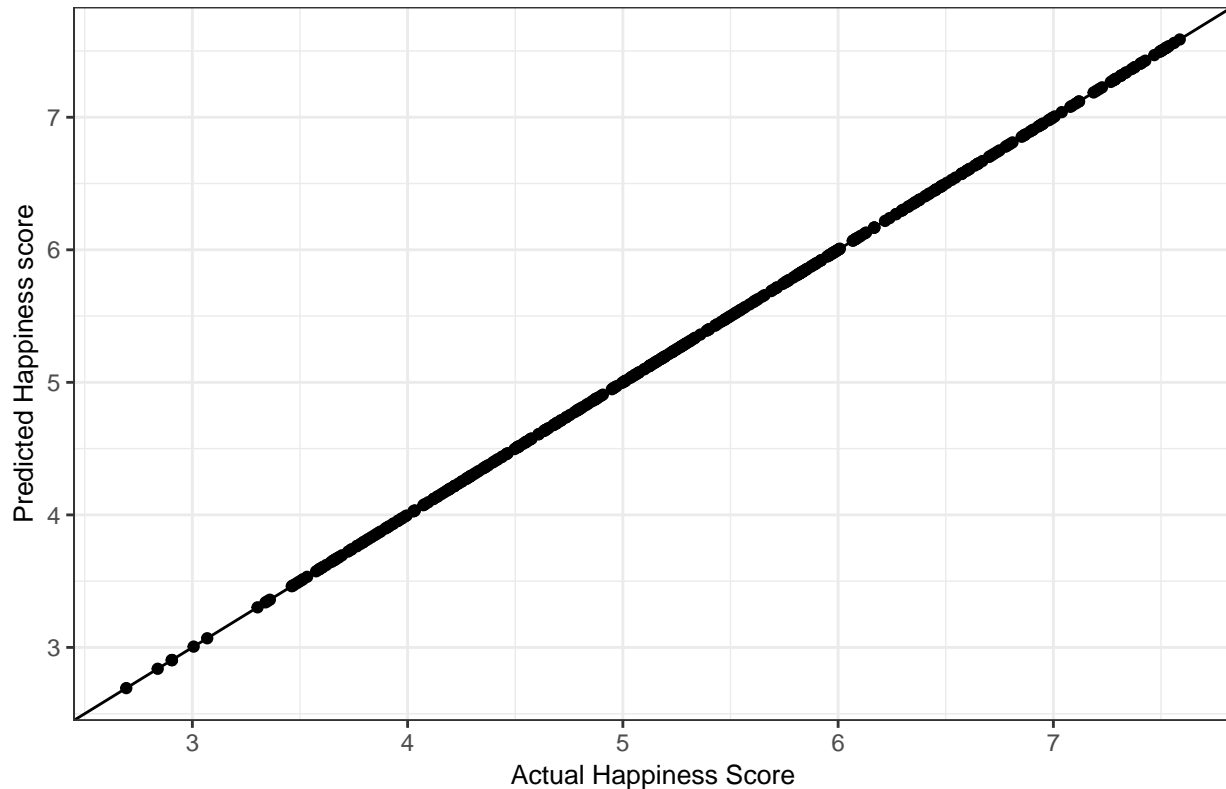
The values fit well. Almost, too well. I tried other regression algorithms in R - Decision Tree, SVM & Random forest regressor. The output of these regressions had similar outputs, all the vaiables were significant at the 95% confidence level.

The equation seems pretty simple: $Happiness\ Score = 1*Economy + 1*Family + 0.98*Health + 1*Freedom + 1*Trust + 1*Generosity + 1*Dystopia\_Residual$

It then occured to me that since happiness scores across years haven't changed much & happiness ranks had stayed constant, I should check for the significance of Regions & Years. On checking the interactions it seemed to me that since there were not significant changes in happiness scores. Using ANOVA, I tried to test then if there was any use of having:

1. Interaction model: Region*Year
2. Additive model: Region + Year
3. Simply Region: Region

```
## Analysis of Variance Table
##
## Response: Happiness_Score
##              Df Sum Sq Mean Sq F value Pr(>F)
## year          2   0.07   0.034  0.0646 0.9374
## Region        9 374.48  41.609 79.2513 <2e-16 ***
## year:Region  18   0.75   0.041  0.0790 1.0000
## Residuals   440 231.01   0.525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
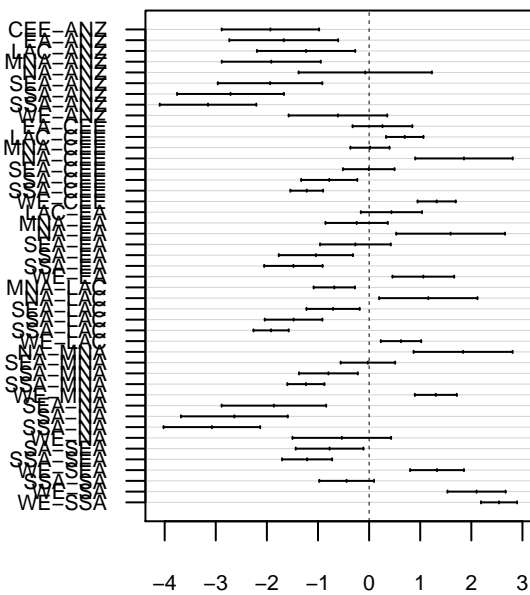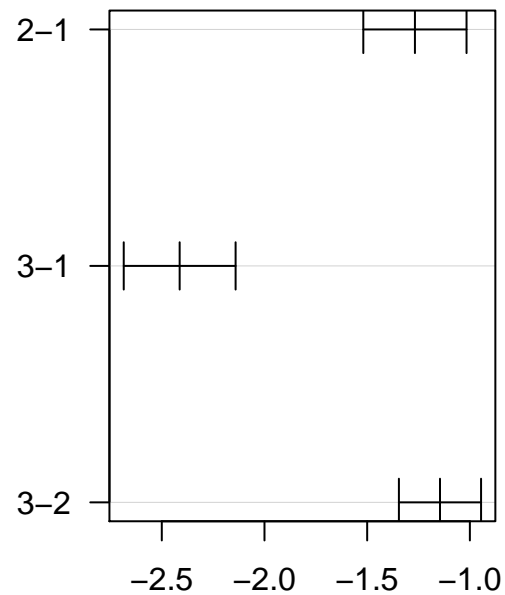
7

```
## Analysis of Variance Table
##
## Response: Happiness_Score
##            Df Sum Sq Mean Sq F value Pr(>F)
## year        2   0.07   0.034   0.067 0.9352
## Region      9 374.48  41.609  82.228 <2e-16 ***
## Residuals 458 231.76   0.506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

95% family−wise confidence leve   **95% family−wise confidence leve**

Differences in mean levels of Region   Differences in mean levels of Region

There is no need for 10 regions, considering that there are quite a few regions with zero differences. So we recluster it to 3 regions only.

```
##
## Call:
## lm(formula = Happiness_Score ~ Economy + Family + Health + Freedom +
##     Trust + Generosity, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.74173 -0.32481 -0.01343  0.36825  1.78524
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0678     0.1312  15.757  < 2e-16 ***
## Economy       1.0598     0.1432   7.400 1.66e-12 ***
## Family        0.7215     0.1345   5.366 1.71e-07 ***
```

Table 1: VIF for the initial regressor model

| regressor	VIF
Economy	3.114395
Family	1.792243
Health	2.823436
Freedom	1.694262
Trust	1.275398
Generosity	1.287416
Dystopia_Residual	1.066395

```
## Health       1.3121    0.2376   5.523 7.72e-08 ***
## Freedom      1.6402    0.2813   5.830 1.55e-08 ***
## Trust        0.2034    0.2707   0.751   0.4532
## Generosity   0.7322    0.3028   2.418   0.0163 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5546 on 275 degrees of freedom
## Multiple R-squared:  0.7743, Adjusted R-squared:  0.7694
## F-statistic: 157.2 on 6 and 275 DF,  p-value: < 2.2e-16

## [1] 0.309762

## [1] 0.2701352

## [1] 0.2594565
```

The Mean Square error is lower with Region interacting with all the variables. Now I wanted to see if I'm working with Panel data.

```
##
##  Hausman Test
##
## data:  Happiness_Score ~ Economy + Family + Health + Freedom + Trust +  ...
## chisq = 4.4575, df = 7, p-value = 0.7258
## alternative hypothesis: one model is inconsistent
```

I believe that just because I couldn't reject the null hypothesis, there is no reason to believe that a random effects model is better. Maybe over a period of few years, it is possible to use this dataset in a fixed-effects model. Plus, the VIF values are less than 5. There would possibly be no problem in using this dataset in the future.

Normal Q–Q Plot

Histogram of residuals(plm.reg
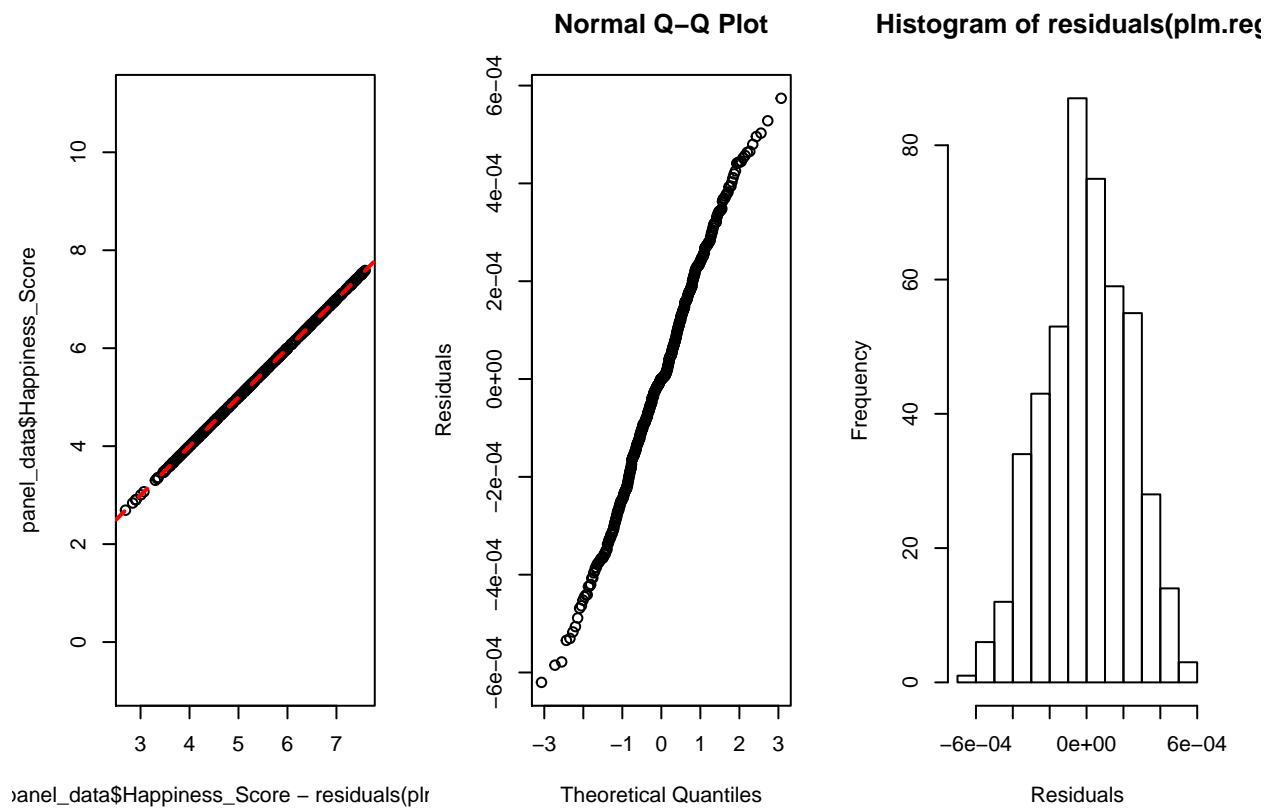
The residuals plots, Q-Q plots & the distribution of the residuals being a normal distribution seems to sugges that this model is fine. A network plot to examine the connection of these variables to Happiness score.
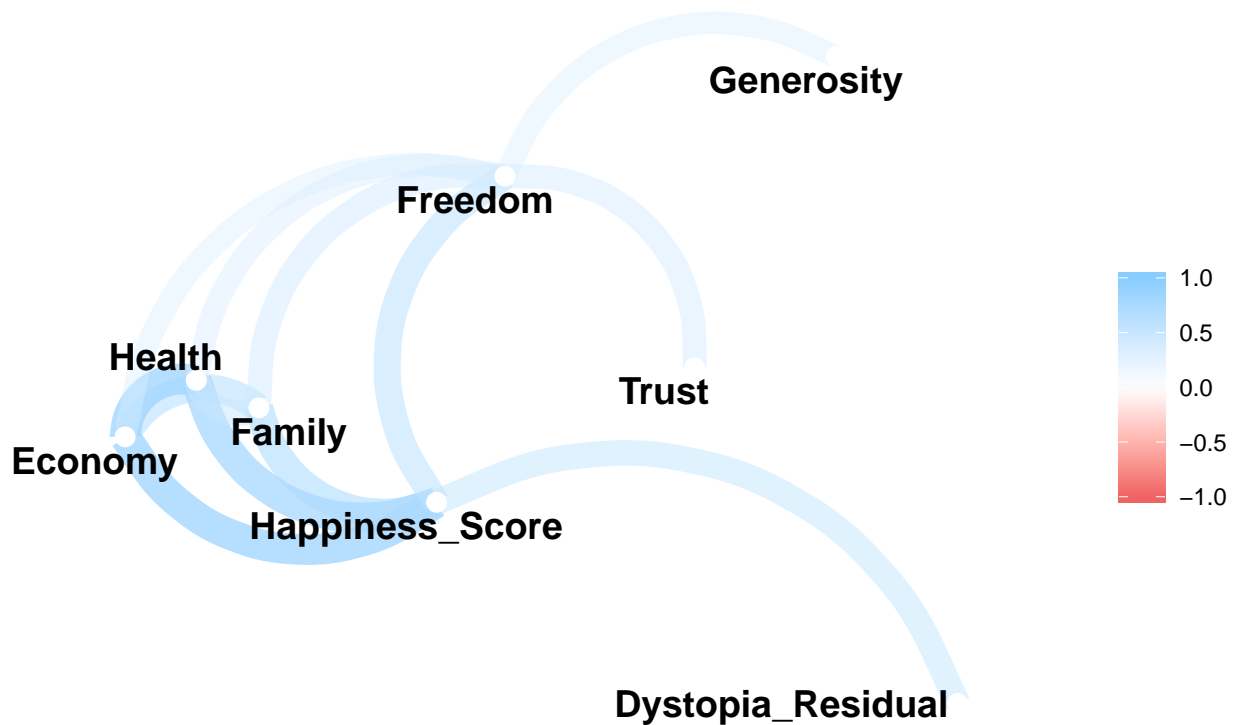
Finally, I used variable selection methods to make sure I have the best model. The olsrr library in R allowed the following:

1. Step all possible
2. Step all subset
3. Forward
4. Backward
5. Step Forward AIC
6. Step Backward AIC
7. Stepwise Both

All the outputs seemed to point out that every variable in use is significant. In conclusion, it seems that the varibale put in use as are accurate in their measurement of happiness score. But we could regroup the world into 3 regions instead of using 10. What would be interesting is to formulate the variables differently & instead of relying on public perception, use the survey results to take perceived happiness scores & then use actual values of a country to estimate the real effects of these variables.

3 References

1. World Happiness Report, 2012
2. World Happiness Report, 2012
3. https://en.wikipedia.org/wiki/Economy_of_Europe, Economy of Europe
4. https://www.goodreads.com/quotes/search?utf8=%E2%9C%93&q=governments+tend+not+to+solve+problems&commit=Search, Ronald Reagan
5. World Happiness Reoport, 2012
6. Kahneman, Diener & Schwarz, eds, (1999)
7. https://worldhappiness.report/download/

4 Citations

1. https://www.econlib.org/archives/2018/02/americans_are_r.html
2. Frey, Bruno, S., and Alois Stutzer. 2002. "What Can Economists Learn from Happiness Research?" Journal of Economic Literature, 40 (2): 402-435.

3.Social Capital,the Economy and Well-Being John Helliwell 2001 4.Social Capital,the Economy and Well-Being John Helliwell 2001

5. Delle Fave, A., Brdar, I., Freire, T. et al. Soc Indic Res (2011) 100: 185. https://doi.org/10.1007/s11205-010-9632-5

5 Appendix

In the appendix, I've included a few other plots & model outputs.
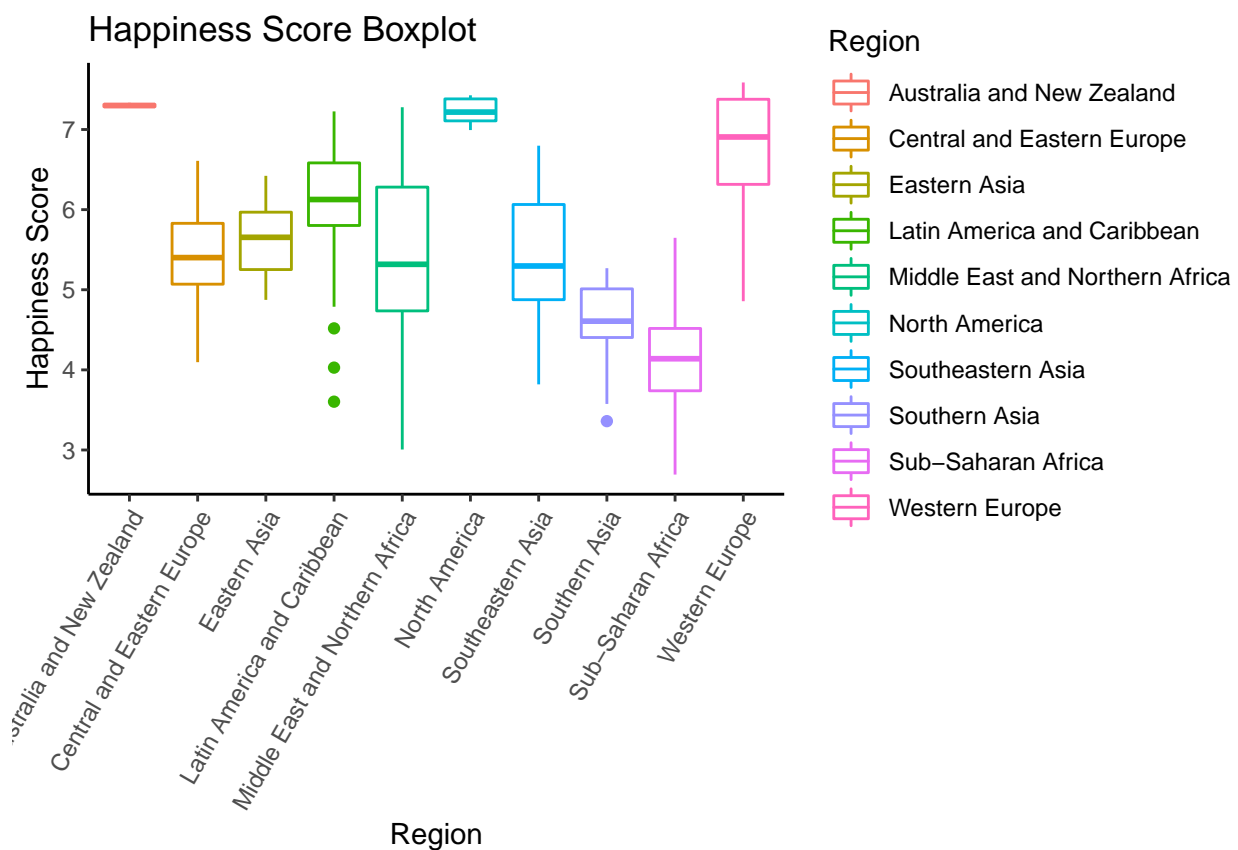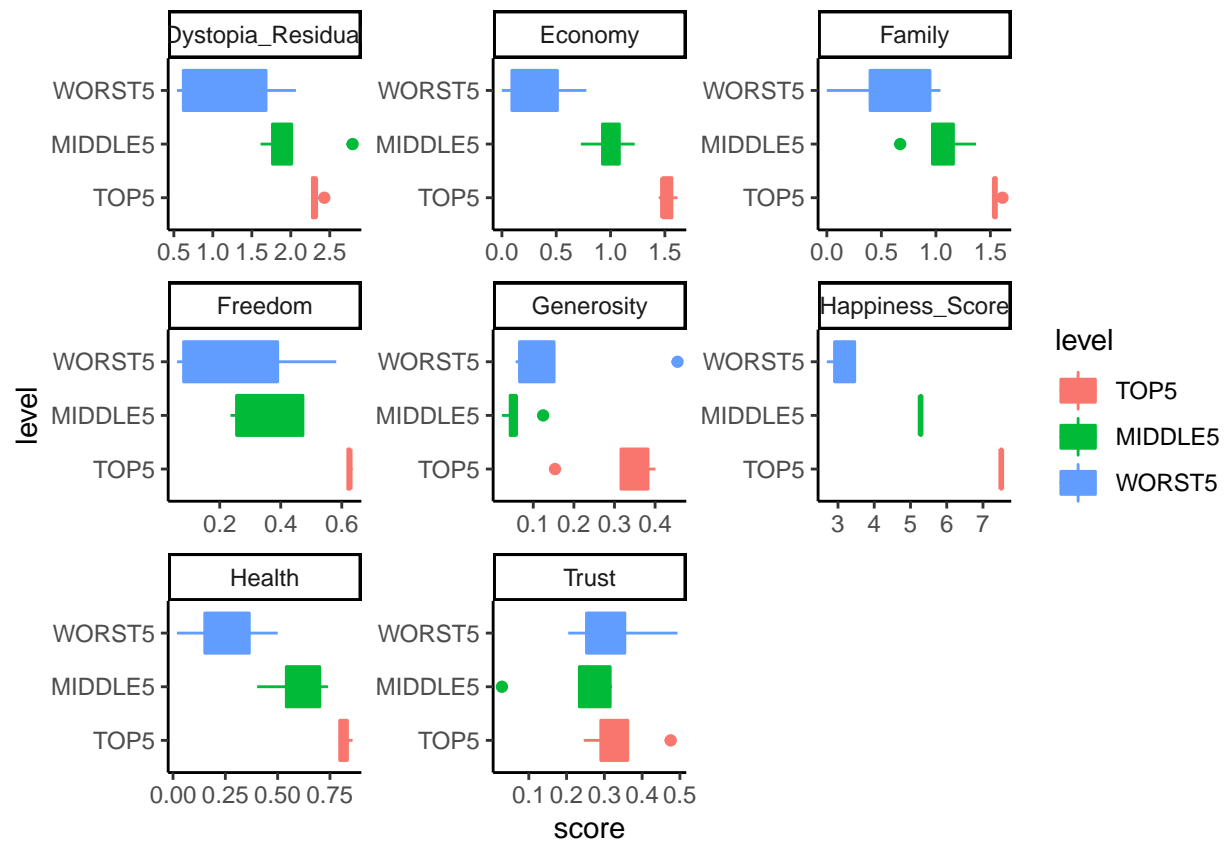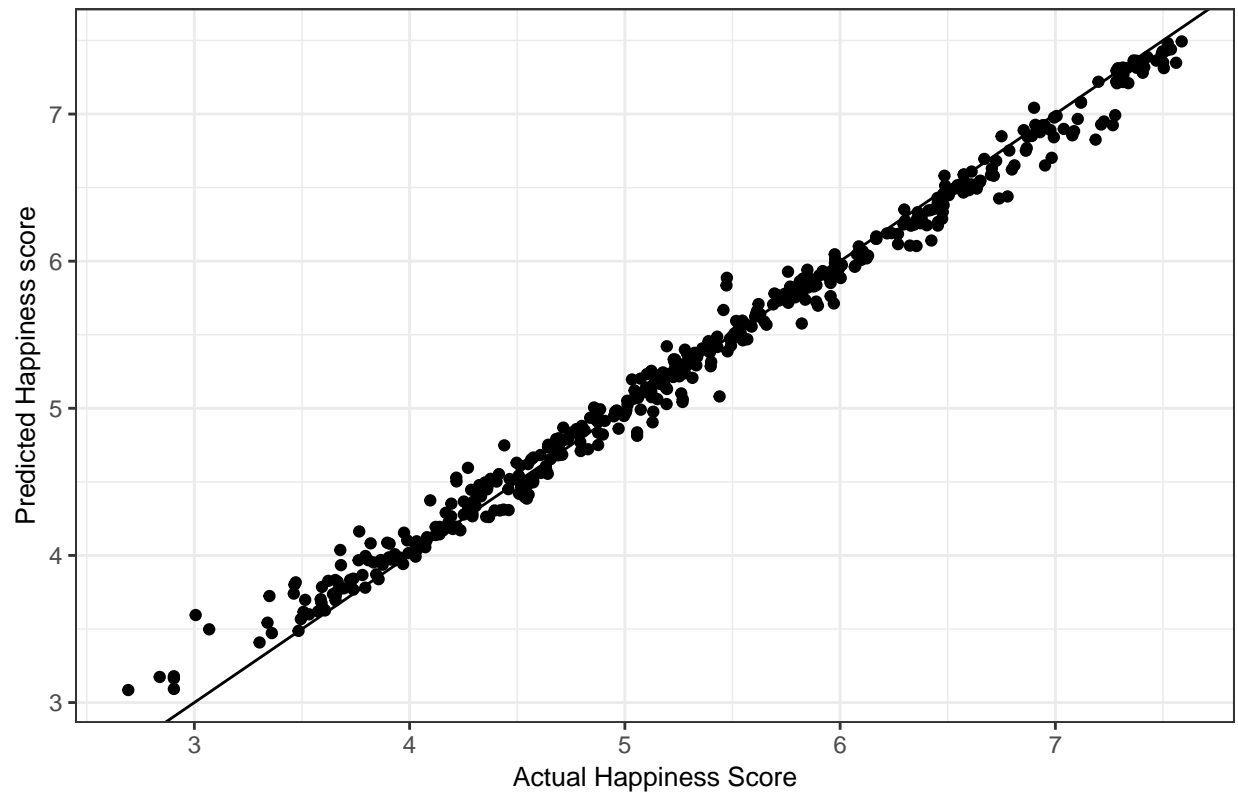
```
##  [1] "Country"          "Region"           "Happiness_Rank"
##  [4] "Happiness_Score"  "Economy"          "Family"
##  [7] "Health"           "Freedom"          "Trust"
## [10] "Generosity"       "Dystopia_Residual" "year"

##
## Call:
## svm(formula = Happiness_Score ~ ., data = dataset, type = "eps-regression",
##     kernel = "radial")
##
##
## Parameters:
##    SVM-Type:  eps-regression
##  SVM-Kernel:  radial
##        cost:  1
##       gamma:  0.1428571
##     epsilon:  0.1
##
##
## Number of Support Vectors:  104
```
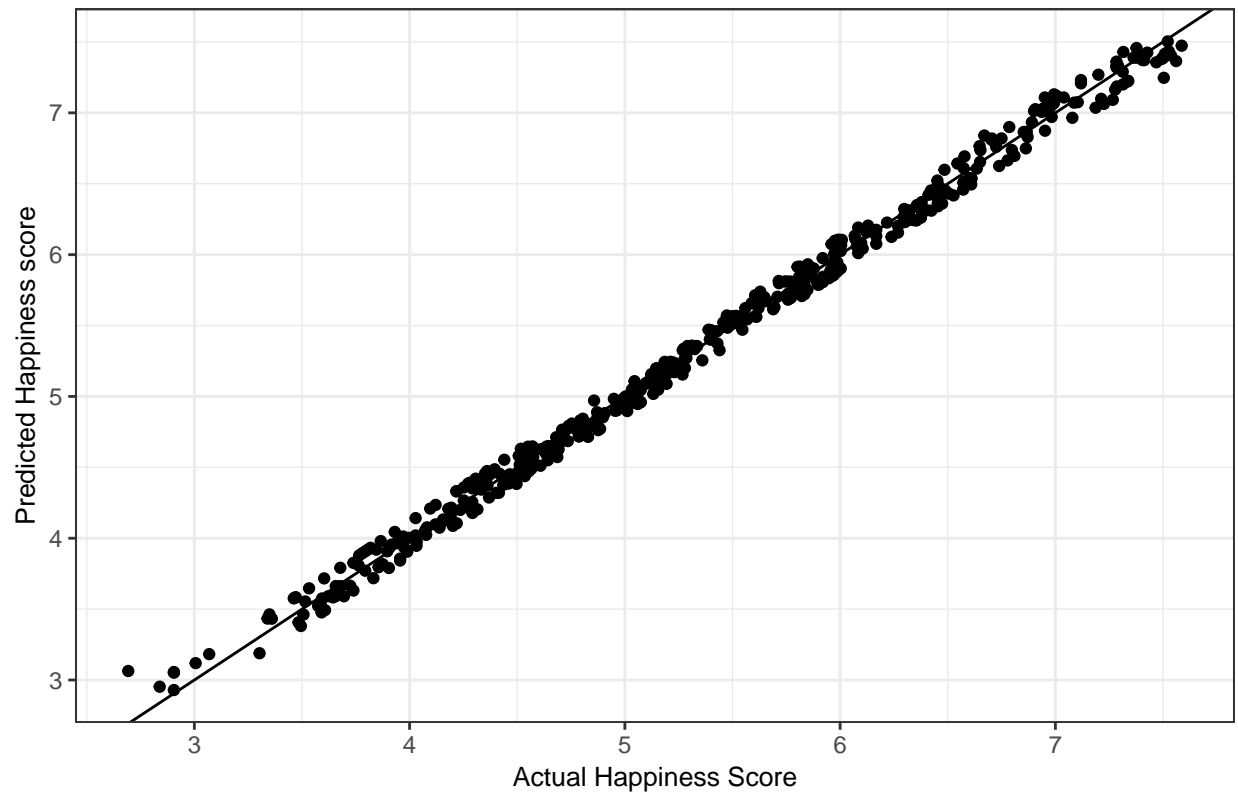
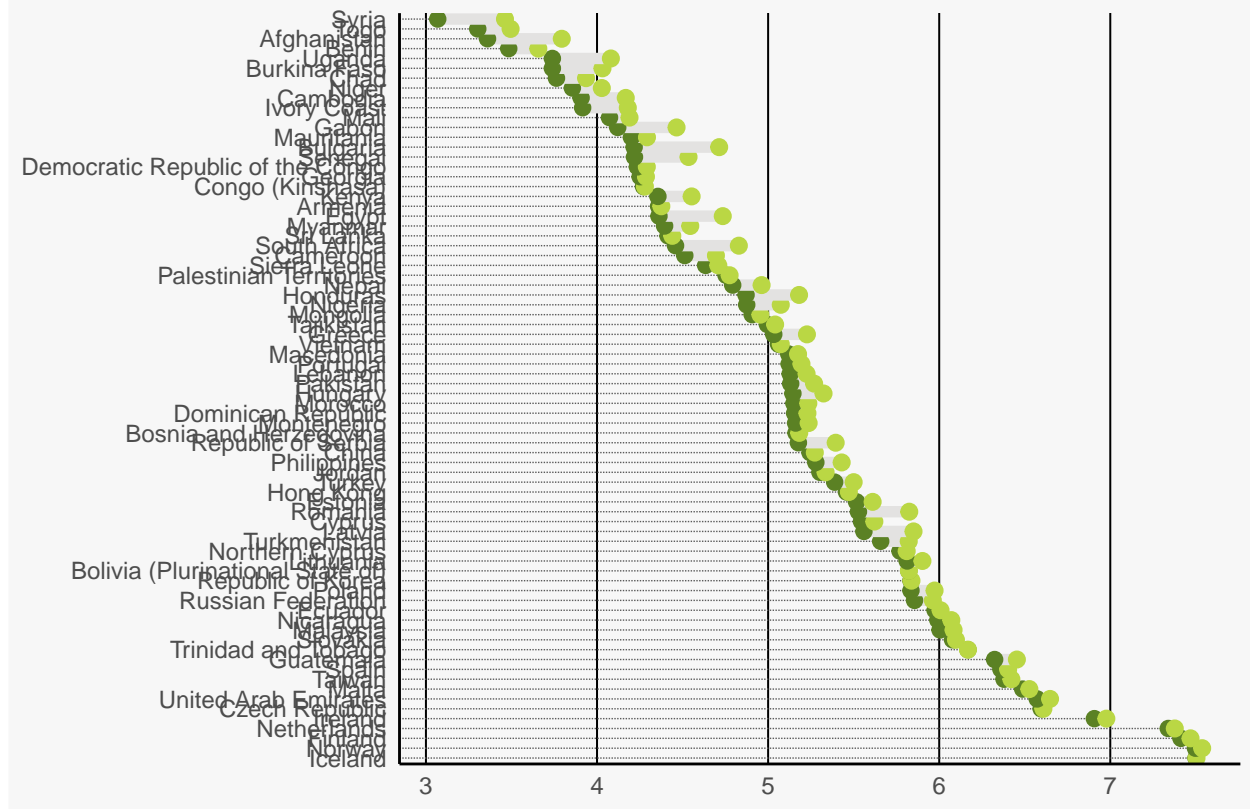**Multiple Linear Regression using recursional partitoning**
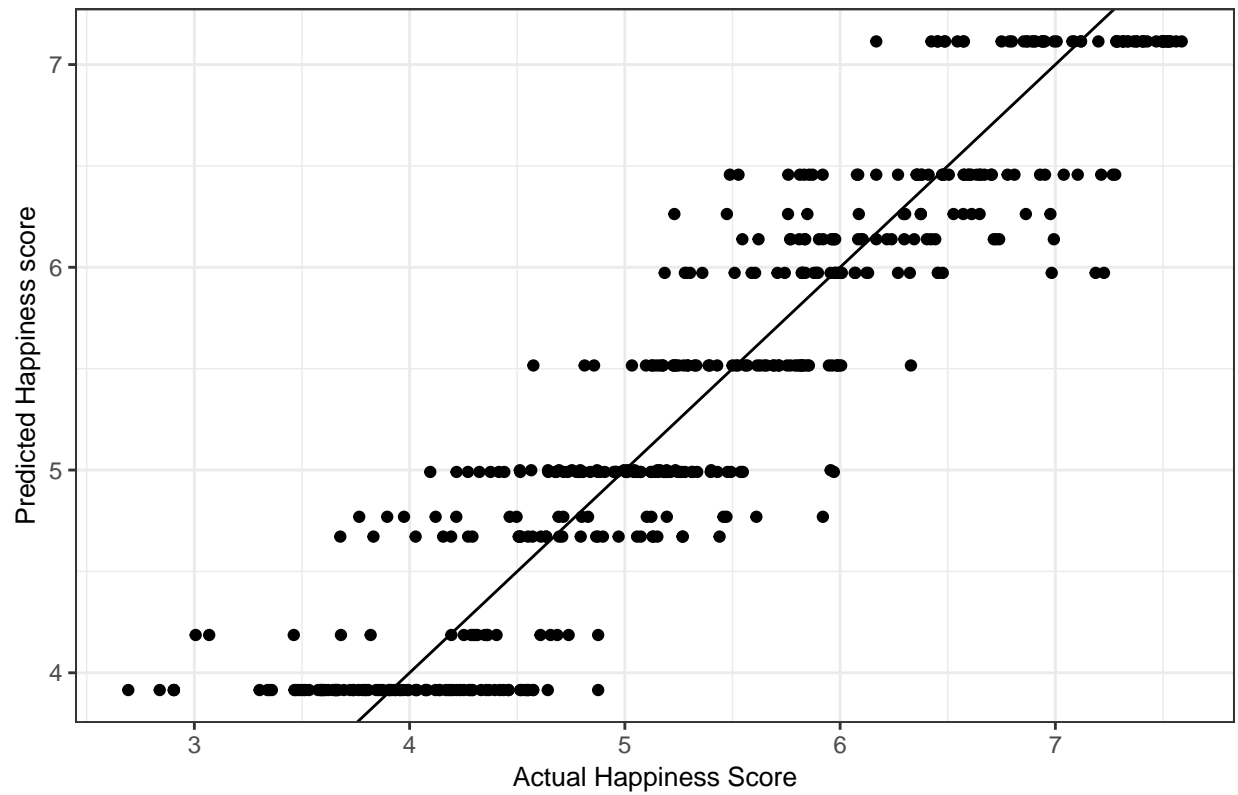
Multiple Linear Regression using Random Forests in R

Multiple Linear Regression using SVM

Country Happiness Scores Increased: 2016 vs 2017

Syria
Afghanistan
Uganda
Burkina Faso
Niger
Cambodia
Ivory Coast
Mali
Gabon
Mauritania
Bulgaria
Senegal
Georgia
Democratic Republic of the Congo
Congo (Kinshasa)
Kenya
Armenia
Myanmar
South Africa
Cameroon
Sierra Leone
Palestinian Territories
Nepal
Honduras
Mongolia
Tajikistan
Vietnam
Macedonia
Lebanon
Pakistan
Hungary
Dominican Republic
Montenegro
Bosnia and Herzegovina
Republic of Serbia
Philippines
Turkey
Hong Kong
Romania
Cyprus
Croatia
Turkmenistan
Northern Cyprus
Latvia
Republic of Korea
Bolivia (Plurinational State of)
Poland
Russian Federation
Nicaragua
Brazil
Trinidad and Tobago
Guatemala
Taiwan
Malta
United Arab Emirates
Czech Republic
Ireland
Netherlands
Norway
Iceland

3   4   5   6   7

Interaction between Region & Year