Analysis on Happiness

Ruiyi Feng

12/06/2022

Abstract

The topic of the report is the world happiness level. Nowadays, people are focusing more and more on their mental health situations. Therefore, it is a good idea to use some elements to measure the average happiness level of countries around the world. In this report, I use multilevel model to find the correlation between happiness level of different countries and see which of the element make largest contribution to the happiness score.

Introduction

Happiness score is one of the methods to measure the feelings of populations around the world. There are many perspectives we need to consider. Specific explanations are included in the following table.

Column Name	Explanation
Country	Name of the country
Region	Region which the country belongs to
Happiness Rank	Rank of the country based on happiness score
Happiness Score	The score measuring happiness level of population in the country

Lower Confidence Interval	Lower confidence interval of happiness score in the country
Upper Confidence Interval	Higher confidence interval of happiness score in the country
Economy GDP per Capita	Level of Gross Domestic Product contributing to happiness score
Family	How family situation contributes to happiness score
Health Life Expectancy	Level of health life expectancy contributing to happiness score
Freedom	Level of freedom contributing to happiness score
Trust Government Corruption	Level of trust government corruption contributing to happiness score
Generosity	Level of generosity contributing to happiness score
Dystopia Residual	Level of dystopia residual contributing to happiness score

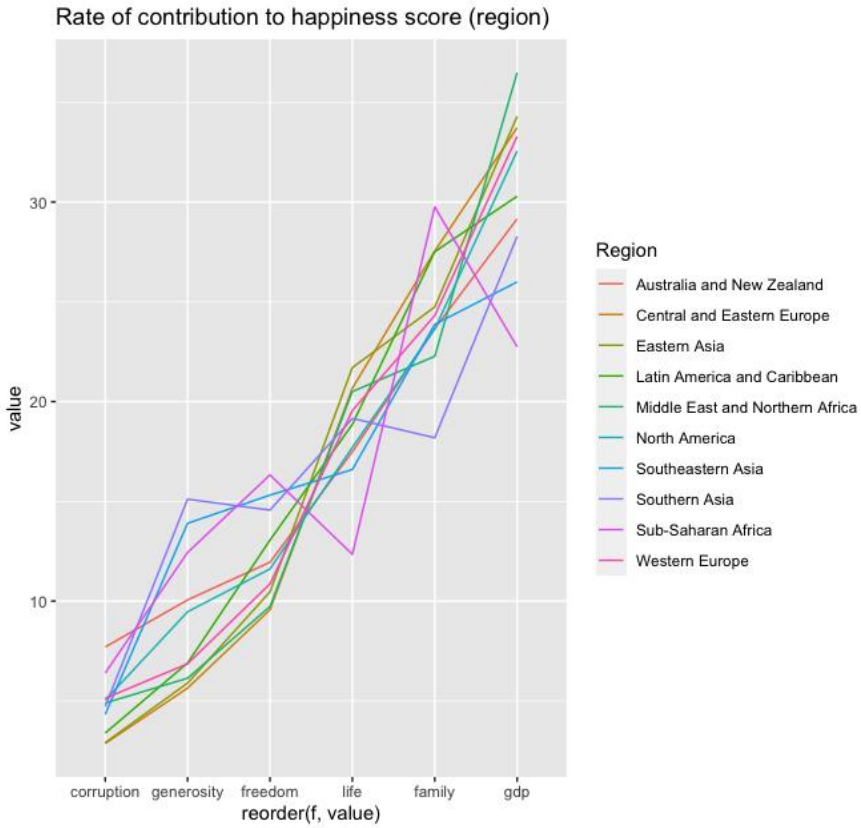
Method

Data processing

The dataset is from Kaggle and can be used directly since it is cleaned. The cleaned data has 157 observations of 13 variables. I choose some of the variables to use in my analysis. In addition, in order to check the correlation among the variables, another data frame is made, which only keeps numeric data.

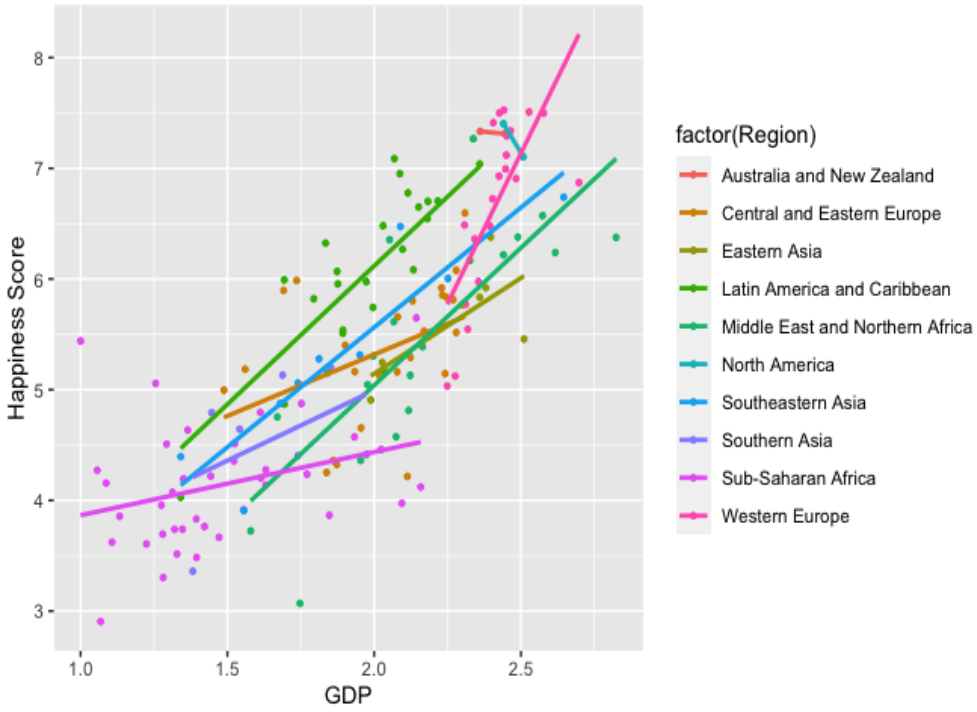
EDA

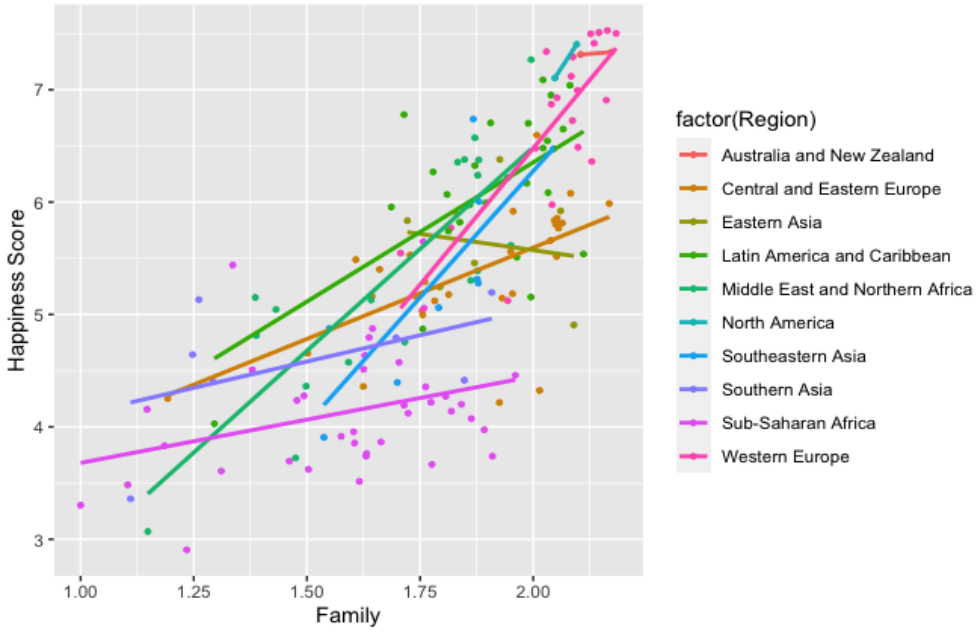
The following plot shows how the happiness score in different regions are affected by the six factors.



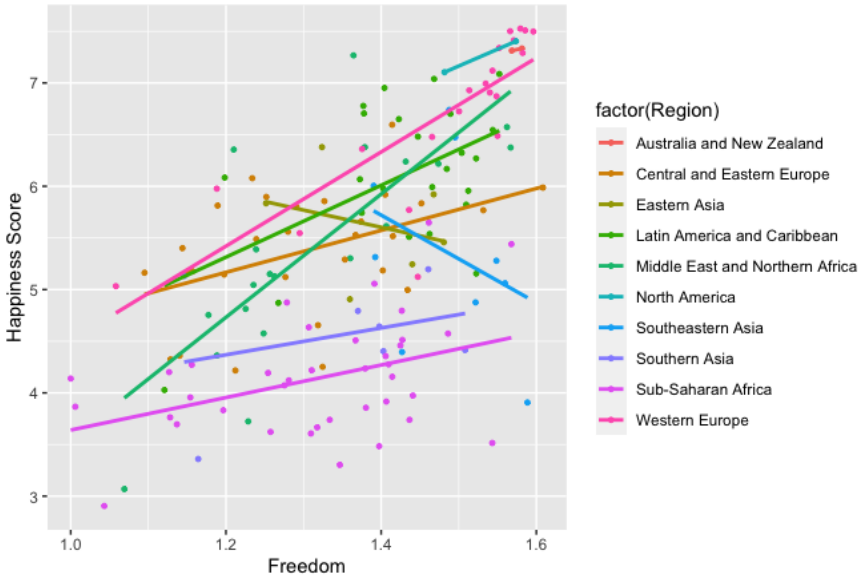
In general, contribution of six factors is similar for most of the regions.

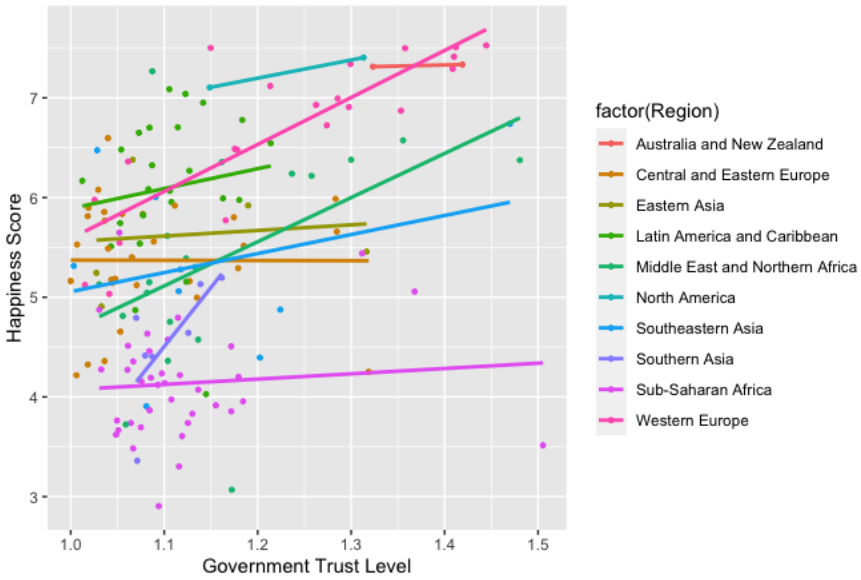
The following plots show specifically on how each factor influences the happiness score based on regions.



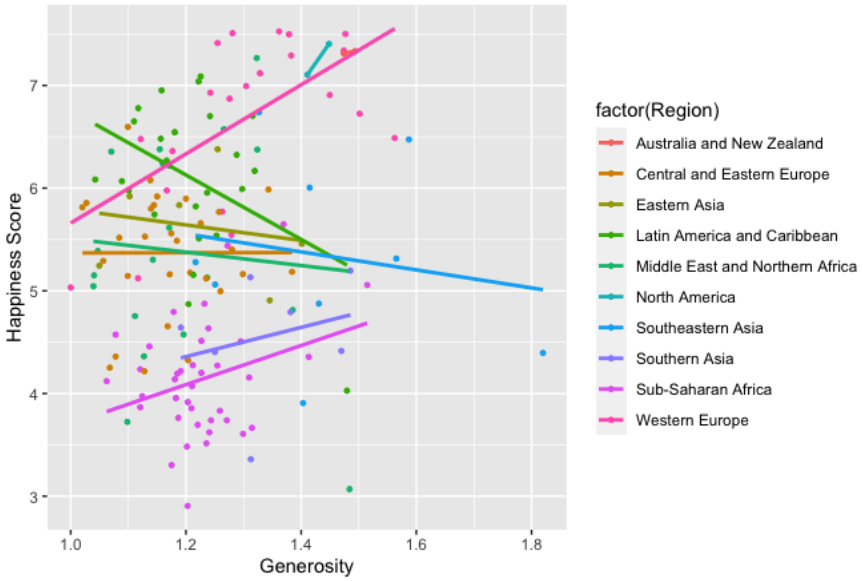


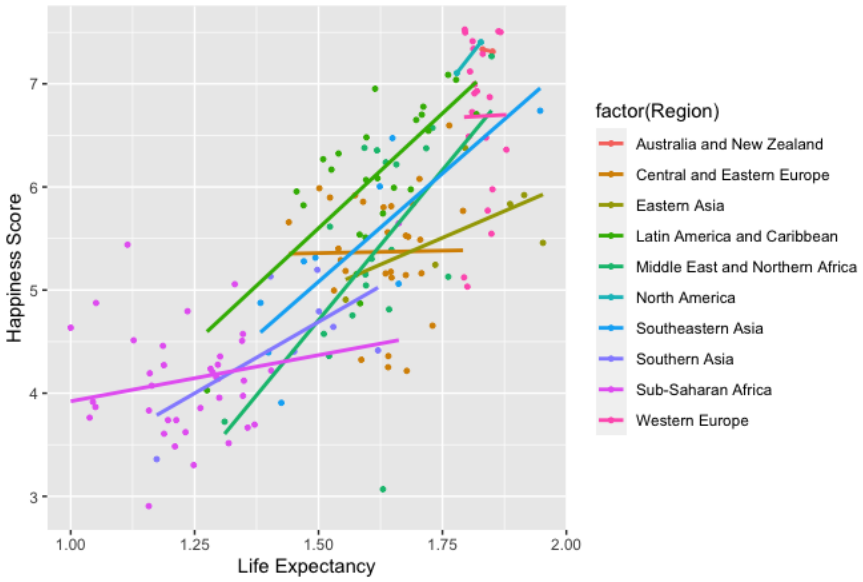
For most regions, GDP and family situation have positive effects on the happiness score, which means that higher GDP land better family lead to higher happiness level on average. Although slopes vary, most of them are upwards.





For the two factors freedom and government trust level, situations are quite similar as the two factors above. For the factor of government trust level, all slopes are going upwards. This means that for all regions in the study, the more people trust their government, the happier they are.





The first plot on this page shows a different conclusion for the factor generosity compared to the rest factors. This factor has negative effects on happiness score for more than half of the regions.

The second plot shows that life expectancy has similar situation as the factor government trust level.

Model fitting

Since regions are divided into groups, here I use the multilevel model to fit the dataset.

The function for the model is

```
model <-
lmer(Happiness.Score~1+Generosity+Trust..Government.Corruption.+Freedom+Health..Life.Expectancy.+Family+Economy..GDP.per.Capita.+(1+Generosity|Region)+(1+Trust..Government.Corruption.|Region)+(1+Freedom|Region)+(1+Health..Life.Expectancy.|Region)+(1+Family|Region)+(1+Economy..GDP.per.Capita.|Region),data = happiness)
```

Result

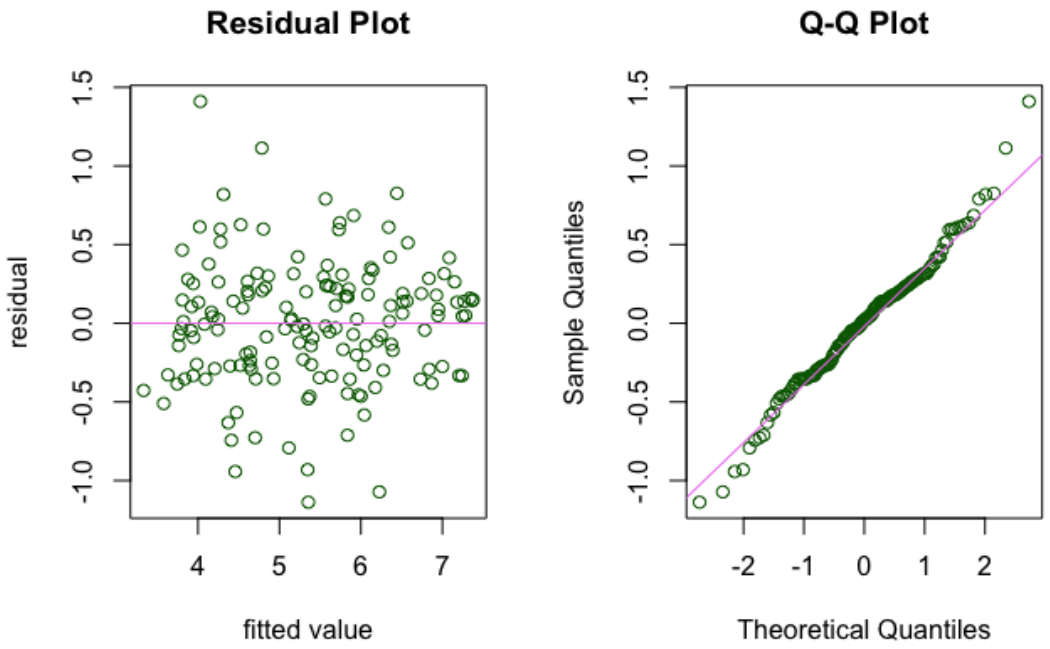
Model Coefficients

```
Fixed effects:
                             Estimate Std. Error       df t value Pr(>|t|)
(Intercept)                    2.3525     0.3558   5.2560   6.611 0.000981 ***
Generosity                     0.2713     0.5464   5.2504   0.497 0.639559
Trust..Government.Corruption.  0.7861     0.5394   6.6483   1.457 0.190604
Freedom                        0.9131     0.3446 105.3311   2.650 0.009294 **
Health..Life.Expectancy.       0.9445     0.4663   6.9684   2.025 0.082645 .
Family                         1.2232     0.3543   5.9845   3.453 0.013643 *
Economy..GDP.per.Capita.       0.9911     0.2292  11.6386   4.325 0.001059 **
```

```
Correlation of Fixed Effects:
            (Intr) Gnrsty T..G.C Freedm H..L.E Family
Generosity  -0.450
Trst..Gv.C.  0.074 -0.084
Freedom     -0.100 -0.124 -0.273
Hlth..Lf.E. -0.333 -0.035 -0.043  0.004
Family      -0.586 -0.069  0.044 -0.142 -0.047
Ec..GDP..C.  0.026  0.123 -0.112 -0.099 -0.530 -0.216
```

Happiness score = 2.35+0.27*(Generosity)+0.79*(Trust Government Corruption) +0.91*(Freedom)+0.94*(Life Expectancy)+1.22*(Family)+0.99(GDP)

Model Validation



The residual plot shows that the average residual value is close to zero. For the Q-Q Plot, most of the points are either very close to the line or on the line. Therefore, it makes sense.

Discussion

Except for the multilevel model, I also tried some other models including linear model and logistic regression model. However, since the factors are not definitely independent with each other, finally I chose the multilevel model. One potential problem is that the dataset does not contain a very large quantity of

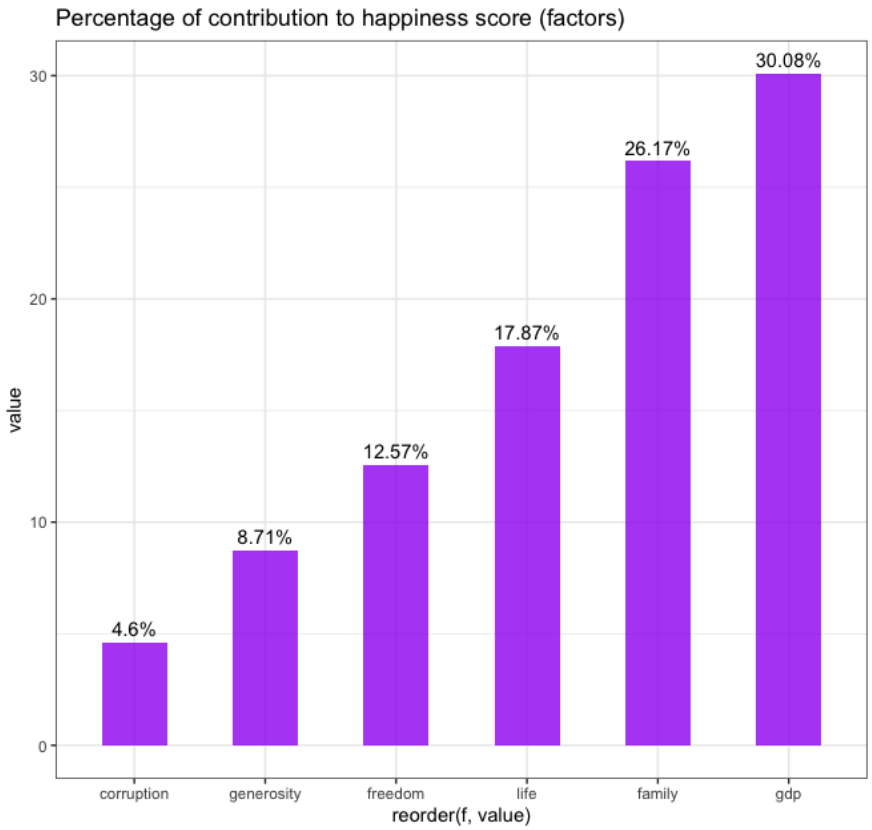
observations. If there are more observations, the model and results might be more accurate and convincing.

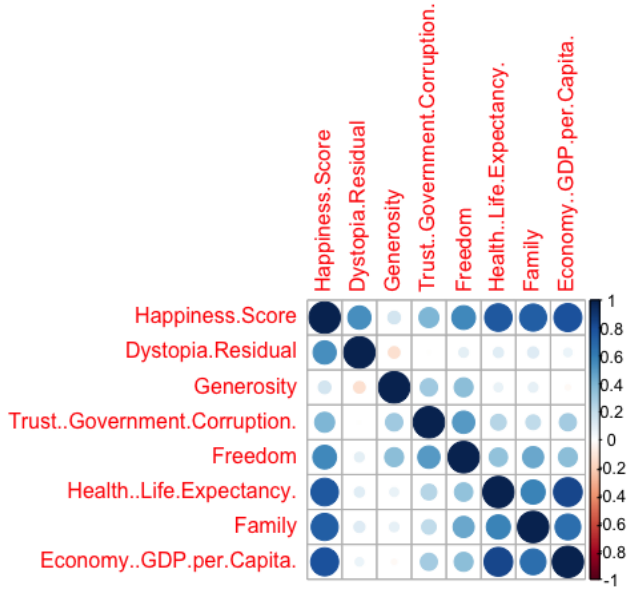
Appendix

Figure 1 is a brief plot showing the average happiness score by regions.

Figure 2 shows the percentage that each factor has on the happiness score.

Figure 3 shows the correlation among the variables and the outcome (happiness score).


Percentage of contribution to happiness score (factors)
value
reorder(f, value)
corruption
generosity
freedom
life
family
gdp
4.6%
8.71%
12.57%
17.87%
26.17%
30.08%
0
10
20
30


Happiness.Score
Dystopia.Residual
Generosity
Trust..Government.Corruption.
Freedom
Health..Life.Expectancy.
Family
Economy..GDP.per.Capita.
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1