



MONASH
University

MONASH
BUSINESS
SCHOOL

Department of
Econometrics &
Business Statistics

☎ (03) 9905 2478
✉ BusEco-Econometrics@monash.edu

ABN: 12 377 614 012

World Happiness Analysis: Evidence from the World Happiness Report

(Elvis) Zhixiang Yang

EBS Honours Student

Yiqi Wang

Master of BA Student

Xintong You

Master of BA Student

Report for
Group 07 ETC5513

27 May 2022



Contents

1	Introduction	3
2	Research Questions	3
3	Data	3
3.1	Description	3
3.2	Pre-processing.	4
4	Exploratory Data Analysis (World Happiness from 2015 to 2022)	4
4.1	The development of the World Happiness	4
4.2	Research questions	4
4.3	The trends in happiness 2015-2022	5
4.4	The relationship between economic situation and health status with the happiness in 2015-2022	6
5	Exploratory data analysis (Pandemic Influences)	8
5.1	Introduction	8
5.2	What are the countries which ranks top 10 in happiness score since the COVID-19 outbreak?	8
5.3	The distribution of the top 10 countries on the world map in 2021	9
5.4	Relation between Happiness score and other indicators in 2021	9
6	Modelling	11
6.1	What is the most important variable to explain the happiness score differences across different countries and years?	11
6.2	Multivariate Linear Model Analysis.	13
6.3	Endogeneity and Sample Selection Bias	14
6.4	Residual Diagnostic for the Regression Model	16
7	Conclusion	19
8	Acknowledgement	19
9	Appendix	21

1 Introduction

Rojas and Vittersø (2010) stated that many factors might contribute to the happiness score. They also mentioned that the conceptions of happiness differ not only across people, but also across the culture and region etc. If we want to improve the happiness, we cannot stimulate all relevant variables. Thus, an analysis about the world happiness report will help to figure out several common key variables that influence the happiness of human beings. Then, we can try to improve the at a worldwide level, especially after the pandemic.

Apart from this, Helliwell et al. (2021) stated that the global pandemic has significantly influenced the happiness of people in different regions. Considering the pandemic, we will also analyse the impacts of COVID-19 to the happiness score for top-ranking countries and the variable relationships after the pandemic happened.

In the modelling part, we will provide a detailed analysis of the data fits in various models. After obtaining two most important variables, we will add another two new relevant variables to build up a multivariate linear regression model with data from 2016 to 2020.

However, it is common to have heteroskedasticity problem in the model. Besides, some variables are not collected after year 2020, leading the model to be not perfect. Critical analysis will be conducted in the end, supported by residual diagnostic.

2 Research Questions

The report will be divided into different sections to explore following research questions:

- The influences of COVID-19 on the world happiness score and the correlations between happiness in 2021.
- The changes of happiness between 2015 and 2022 in different regions.
- The impact of economic situation and health status on happiness.
- The important variable in explaining happiness scores via different models and discover their marginal effects in a linear model.

3 Data

3.1 Description

Our data comes from Kaggle, which is a report of the World Happiness. In the eight datasets, the variables explaining the happiness score including Economy(measured in GDP per capita), Health status, Freedom, Generosity are consistent each year. Nevertheless, after year 2020, few new variables

have been added and there are few changes with the existing variables. The happiness score is ranked from the highest to the lowest one and gained via an anonymous worldwide survey across different countries around the world.

Apart from our existing data, we will also join two new datasets in the modelling part gained from the World Bank Data due to the potential issues in our existing dataset. More information will be discussed in the multivariate linear model part.

3.2 Pre-processing.

The data gained from Kaggle is not clean. In this report, we tried to combine all the historical data with common factors, which are Economy, Health, Freedom, Generosity each year and drop the NA values. However, for specific yearly analysis after 2020, we decide to keep all the new variable in explaining the happiness score.

In addition, two variable related to Consumer Price Index and Populations from the World Bank data will be joined across different countries in each year when we trying to build up our multivariate linear model.

4 Exploratory Data Analysis (World Happiness from 2015 to 2022)

In this part, we will investigate the trends of happiness score represented by region between 2015 and 2022. After this, we will analyse the impacts of annual health and economic status on happiness score.

Combined with our research questions, we consolidate all the datasets for each year between 2015 to 2022 into a new dataset, and then tease out the variables to use, which are year, region, happiness score, economy, and health.

4.1 The development of the World Happiness

Helliwell and Wang (2012) stated that with the continuous progress of human society, the happiness score has become an important indicator for measuring the living standards across regions. In real world analysis, the happiness score is related to many factors, it will also affect nearly all aspects of human activities. Thus, we will start with two research questions to explore the happiness score of the world.

4.2 Research questions

- How will happiness trends change between 2015 to 2022 in different regions?
- What is the relationship for both economic situation and health status to the happiness score?

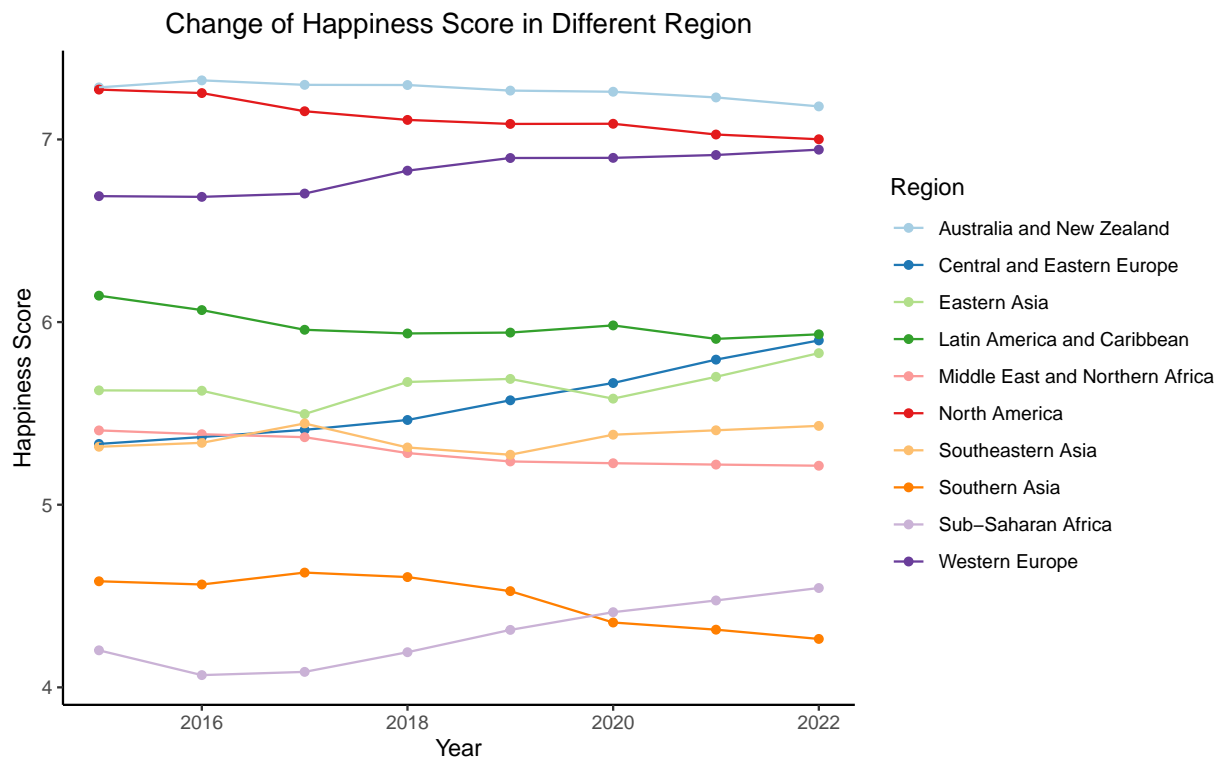


Figure 1: *Change of Happiness Score in Different Region*

4.3 The trends in happiness 2015-2022

From the Figure 1, we can see that these trends in different regions can be divided into three different levels of happiness score.

We can see from the plot that both South Asia and Sub-Saharan Africa remained at a low level for the past few years. While Southern Asia countries started declining in 2017, Sub-Saharan Africa's happiness began to increase year by year after reaching its trough in 2016. In addition, the declining trend of happiness for people in Southeastern Asia may due to the population booming in these years.

Moreover, the three regions had a relatively high scores are: Australia and New Zealand, North America, and Western Europe. We can see people living in these developed area feels happier than others, which may be due to the good social welfare, less stress and better working environment. Nevertheless, an interesting point to notice is that even there are many modern countries in Eastern Asia, people didn't really feel happy in the past few years. This may due to the living condition and working stresses are high from these Eastern Asian countries.

Then, the remaining regions are all centered in middle level. Among these regions, Central and Eastern Europe is the only region have an obvious increasing trend over the years, while the rest of the region is in a state of slightly fluctuating but generally stable trends.

Overall, we conclude that for those regions with relatively better economic situation and social welfare would generally lead to a higher happiness scores with less fluctuations than those regions with poor economic situation.

4.4 The relationship between economic situation and health status with the happiness in 2015-2022

People's understanding of happiness is inseparable from their own living conditions. In this section, we will also explore the relationship between happiness and economic and health status.

From the figure 2, we can find that there is a positive correlation between economic status, health status and happiness score from 2015 to 2022, which is make sense that a better the economic status and health status will also make people happier.

There is an overall trend from 2015 to 2022 that the correlation of health and happiness is increasing over the years. Nevertheless, the correlation of economic situation with happiness score is decreasing. From 2017 to 2018, the average influences of economic on happiness score increased slightly, while the influence of health on happiness score decreased slightly during this period.

On the other hand, from 2019 to 2022, the influence of the economic status on the happiness score is getting lower while the influence of the health on the happiness score has increased significantly especially after 2020, which maybe due to the pandemic that people care more about their health status.

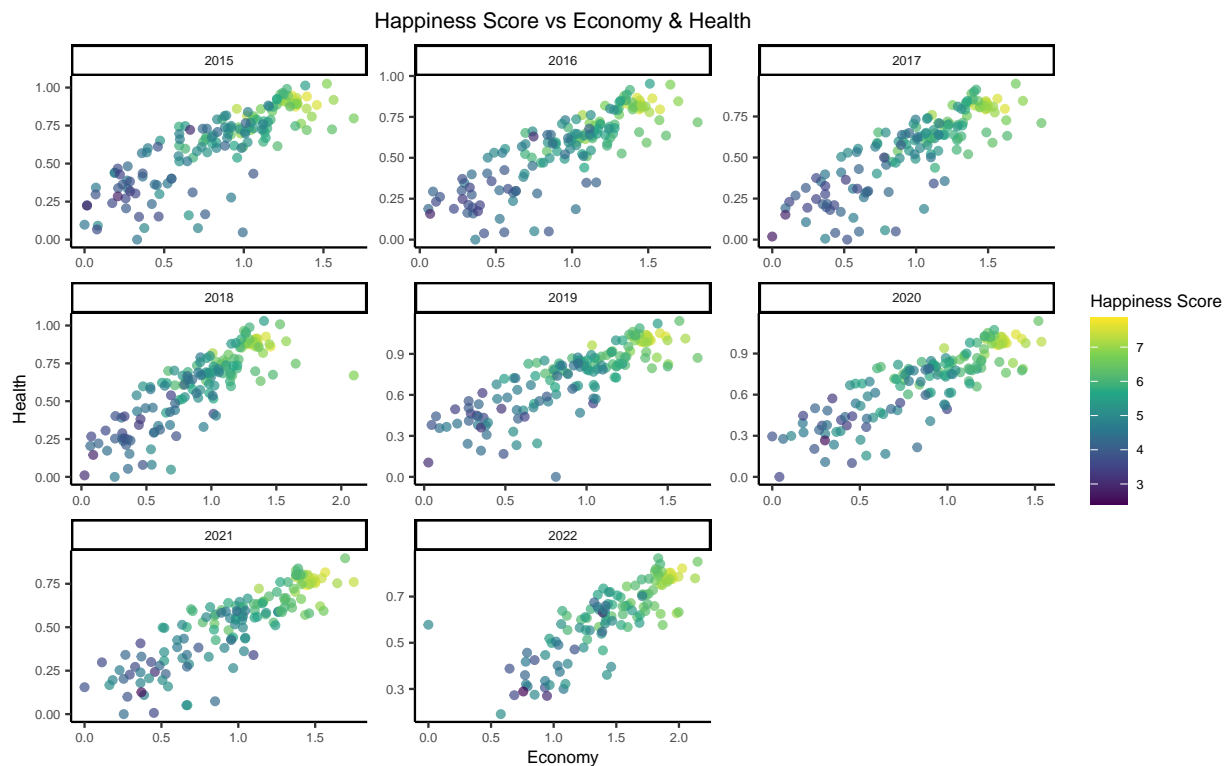


Figure 2: Happiness Score vs Economy & Health

Year	Intercept	Economy	Health
2015	3.250	1.616	1.203
2016	2.980	1.516	1.662
2017	2.968	1.504	1.632
2018	3.085	1.511	1.565
2019	2.892	1.359	1.775
2020	3.117	1.305	1.791
2021	3.298	1.316	1.854
2022	2.370	1.406	2.085

Table 1: *Simple linear relationship between economic status and health status with happiness score each year: 3-digits*

We also conducted an simple regression analysis. From the Table 1, we can see the impact of economic situation on the slope decreased very slowly from 1.616 to 1.406 over the years. On the contrary, the influence of health on happiness score changed a lot with the slope increased from 1.2 to 2.08. Through this simple linear regression analysis, we can also see that people gradually focus on their health status nowadays even though the pandemic has been controlled.

5 Exploratory data analysis (Pandemic Influences)

5.1 Introduction

According to Helliwell et al. (2021), economic situation, people's health and freedom have been severely affected across different countries since the outbreak of COVID-19 in 2020, which will directly impact the world.

In the following research, the top 10 happiest countries in the world and their regions will be explored. Then, factors associated with the happiness score will be explored by analyzing the relation between happiness score and six selected indicators in 2021.

5.2 What are the countries which ranks top 10 in happiness score since the COVID-19 outbreak?

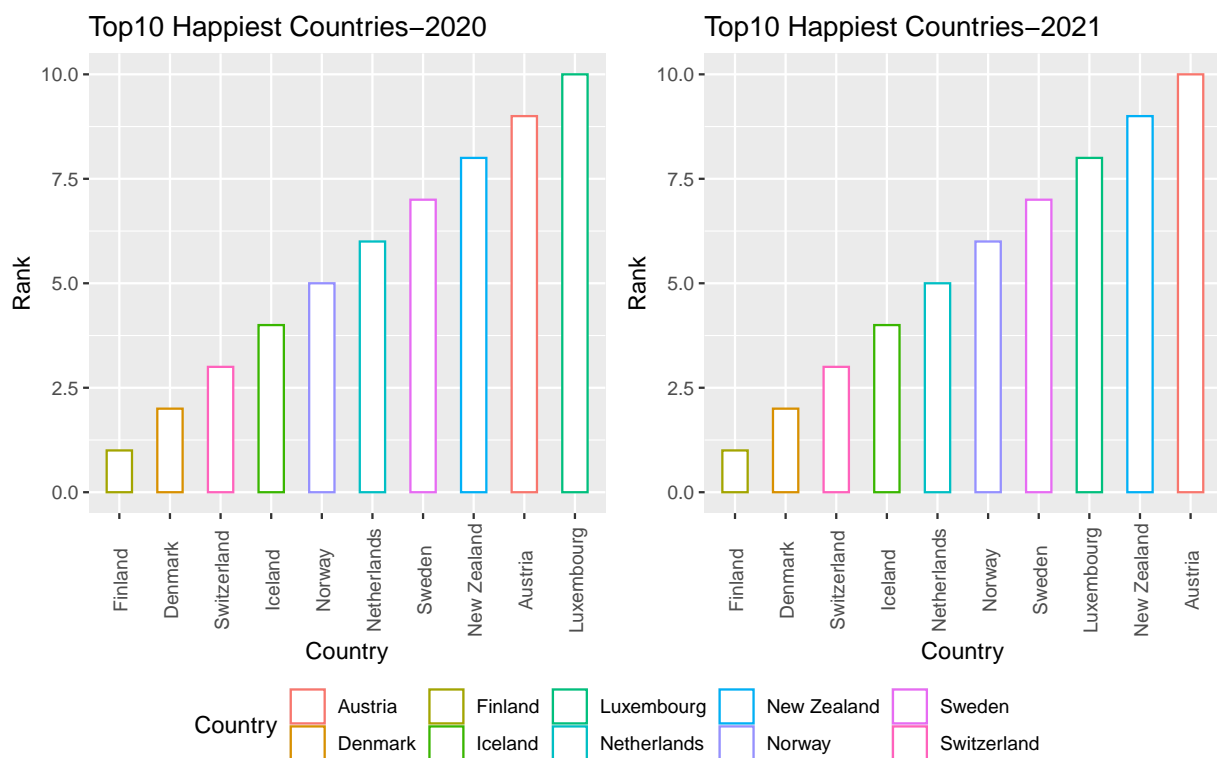


Figure 3: The top 10 countries in happiness score since COVID-19

In the Figure 3, the top 10 countries in the World Happiness have not changed since 2020 despite the impact of COVID-19, while their ranks have changed slightly.

In addition, Finland has been the happiest country for two consecutive years in 2020 and 2021. We believe this is mainly due to the fact that Finland has a well structured social welfare and health care system (Lappi-Seppälä, 2006).

The distribution of top 10 countries on the world map in 2021

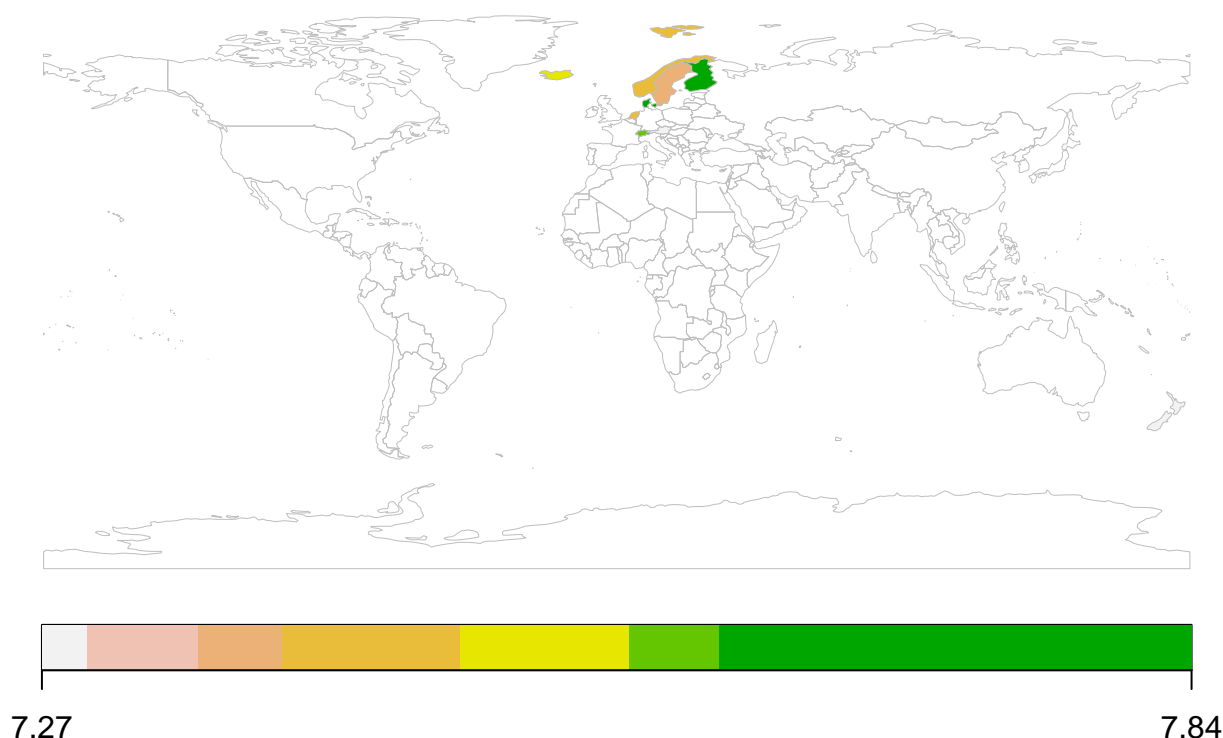


Figure 4: *The distribution of top 10 countries on the world map in 2021*

5.3 The distribution of the top 10 countries on the world map in 2021

According to the Figure 4, these countries are mainly northern and western European countries. Obviously, they are all developed countries which have a technologically advanced infrastructure and their economy is highly developed.

So besides high economy, what are the other indicators that can affect happiness score?

5.4 Relation between Happiness score and other indicators in 2021

The 6 linear graphs in Figure 5 demonstrate the relationships between the happiness and 6 attributes of the countries in 2021. R-squared (R^2) represents the proportion of the variance for happiness that's explained by an independent variable in the regression model. For example, in the graph on the upper left, its R-squared is 0.624, indicating that there are 62.4% of their happiness scores can be explained by their GDP.

Therefore, we can see that both social support and health life expectancy explain happiness in a relatively high proportion which is 57.3% and 59% respectively. While freedom, generosity and trust in government corruption are not good in explaining happiness score.

To summarize, in this part we found the world's top 10 happiest countries are mainly concentrated in Northern and Western Europe in 2020 and 2021, which have high economic level and GDP. From the

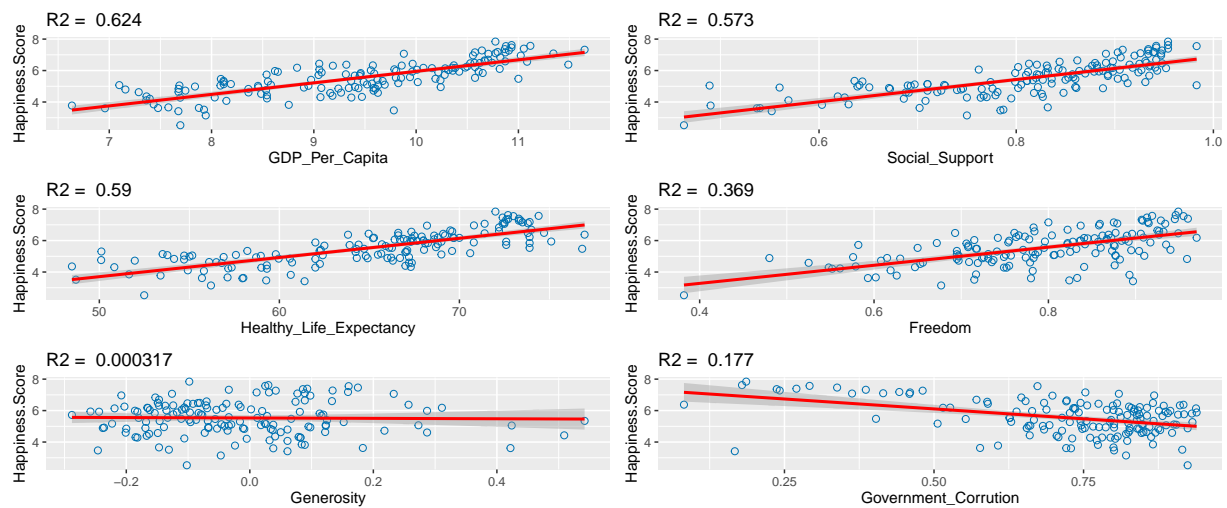


Figure 5: Relation between Happiness score and 6 indicators in 2021

data of the world, the variables highly related to happiness are GDP, social support and health life expectancy.

6 Modelling

6.1 What is the most important variable to explain the happiness score differences across different countries and years?

In this part, we will conduct few statistical analysis based on various models. Our main target is to find the important variable in explaining happiness scores via different models and conduct an analysis based on their marginal effects in a multivariate linear model.

Since there are many different variables across each year in our data, we only keep the common variables (Economy, Health, Generosity and Freedom) for these years to conduct our analysis. Before exploring how each factor will contribute to the happiness score, we first need to select a model that can represents our data well by running several tests on a few common models, listed in Table2:

<Possible Models >	<Model description >
Multivariate Linear Model	Simple Linear regression with multiple variables
Support Vector Machine Model	Use multiple learning algorithms (resampling and tree) to give us better results.
Decision Tree Model	Binary tree model have control statement.
Random Forest Model	Use multiple learning algorithms (resampling and tree) to give us better results.

Table 2: *Model Description of our Possible Models*

We have divided the historical data (from 2015 to 2022) into two separated sets, a training test and a test set with different split ratio (see Figure 6; Figure 12 to 15 in appendix). The test set will be used to examine which model has the best goodness-of-fit after building up the model with the training set.

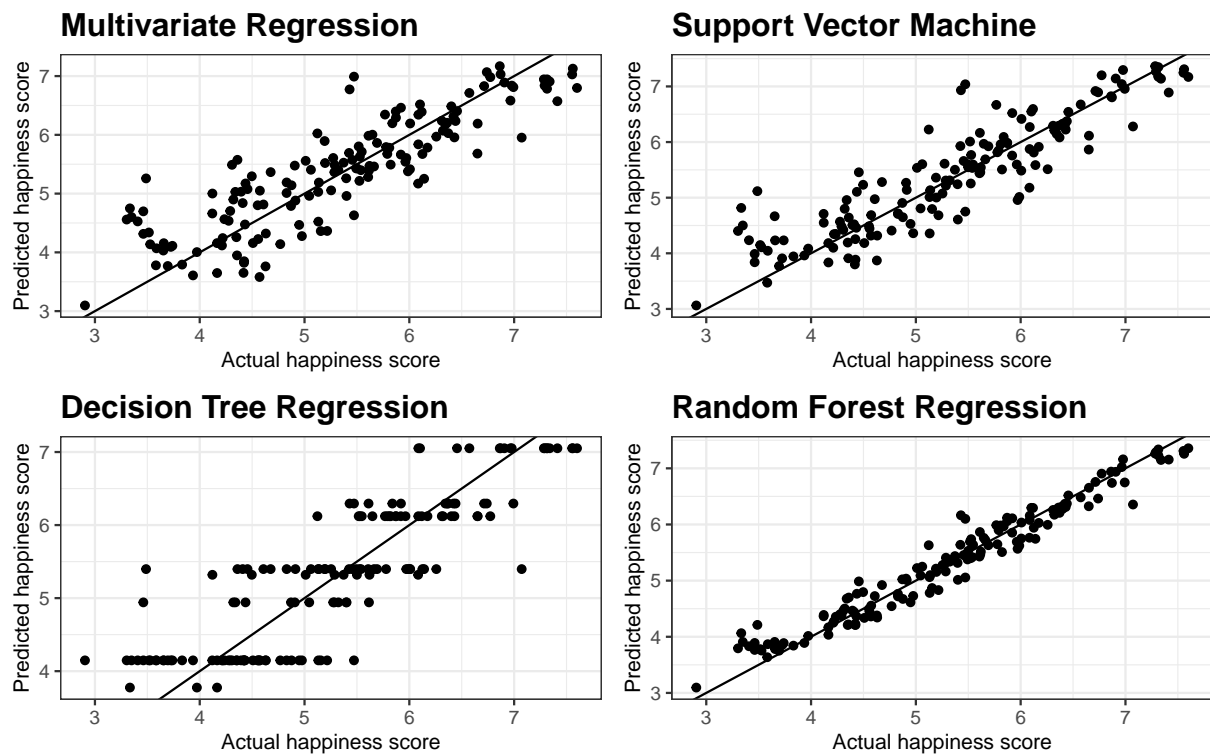


Figure 6: Model Training Split Ratio is 0.8

6.1.1 Random Forest regression

Random Forest Model has a variable selecting system (via bootstrapping) to decide the most significant tree and is able to reduce overwriting compared to the decision tree. With that said, the random forest is a strong modeling technique and more robust than other methods (Lieberman, 2017). We can see from the plot that this model has captured the data very well in the past few years based on various training sets (see Figure 12 to 15).

By checking the loading for each variable in the RF model, we get the order of importance in Table 3.

Table 3: Variable importance for Random Forest model

	IncNodePurity
Economy	348.6690
Health	327.0846
Freedom	169.3299
Generosity	97.7974

Table 4: Linear regression model for happiness scores without new data

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.4374	0.0753	32.3704	0
Health	1.2232	0.1392	8.7871	0
Economy	1.4543	0.0842	17.2820	0
Freedom	1.3722	0.1096	12.5222	0
Generosity	1.1702	0.1397	8.3785	0

Based on the result in Table 3, the most important variable is Economy, followed by Health, Freedom and Generosity being the last.

6.2 Multivariate Linear Model Analysis.

The RF model helps us to determine which variables to choose in the multivariate linear regression model. The advantage of using multivariate linear regression is that it can allow us to analyse the relationship between different variables in a statistical coherent way (Voxco, 2022), such as, marginal effects and percentage changes.

If using a classic linear model with all four variables,

$$\text{Happiness score} = \text{Economy} + \text{Health} + \text{Generosity} + \text{Freedom}$$

They have similar impacts on explaining the happiness score and all of them are significant in Table 4. Thus, it is hard to tease the importance of each variable out with this model.

As a result, we will drop the least insignificant two and add another two new variables, Consumer Price Index and the population size of each country, gained from the World Bank data from 2016 to 2022 (World Bank, 2022). After matching the data for each country and drop the NA values, we can use the new dataset to construct a new Linear model. Nevertheless, due to the lack of data in 2021 and 2022, we can only use the data up to 2020.

New Multivariate Linear Model with a natural log of score :

$$\begin{aligned} \log(\text{score}) = & -4.6000 - 0.0008 \text{ cpi} + 0.2591 \log(\text{economy}) \\ & -0.0026 \log(\text{population}) + 0.0114 \log(\text{health}) + 0.0032 \text{ year} \end{aligned}$$

Table 5: *Multivariate Linear regression model for the log data with R-squared is 0.6076*

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.5999	11.5632	-0.3978	0.6909
cpi	-0.0008	0.0002	-4.2994	0.0000
log_eco	0.2591	0.0094	27.5487	0.0000
log_population	-0.0026	0.0036	-0.7072	0.4797
log_health	0.0114	0.0042	2.7419	0.0063
year	0.0032	0.0057	0.5508	0.5820

We can see in Table 5 that 60.76% of the variation in $\log(score)$ can be explained by the model. It is noticeable that the economy status(GDP per capita) has the largest influence on the happiness scores than other variables.

Besides, in econometric contexts, coefficients in the log-log model can be interpreted as the percentage change in dependent variable when there is a one percentage change increase in the regressor (Benoit, 2011).

For example:

$$\begin{aligned}
 \frac{\Delta \log(\text{Happiness Score})}{\Delta \log(\text{Economy})} &\approx \frac{\Delta \text{Happiness Score}}{\Delta \text{Economy}} \frac{\text{Economy}}{\text{Happiness Score}} \\
 &= \frac{\% \Delta \text{Happiness Score}}{\% \Delta \text{Economy}} \\
 &= \text{ME}(\text{Coefficient}) \\
 &= 0.2591
 \end{aligned}$$

Here, we can interpret that a 1% increase in Economy(measured in GDP per capita) will increase the happiness score by 0.2591%, keeping all other regressors constant. Similarly for the rest variables, a 1% percent increase in health will increase happiness by 0.0114%. On the contrary, we can see that an increase in population and CPI have negative impacts, which are -0.0026% for Population and -0.08 for CPI respectively.

6.3 Endogeneity and Sample Selection Bias

This model may have an endogenous problem, which is caused by omitting variables, as we only include 5 relevant variables in the model. There are some latent variables that affect the happiness score but are not included in this model such as education level and culture backgrounds.

Apparently, our model also has some bias in sample selection. In our model, data for 34 countries are not included, which are colored in red in Figure 7. These missing countries are either sparsely

Country included in our dataset

Red: Countries are not Included; Blue: Countries are included

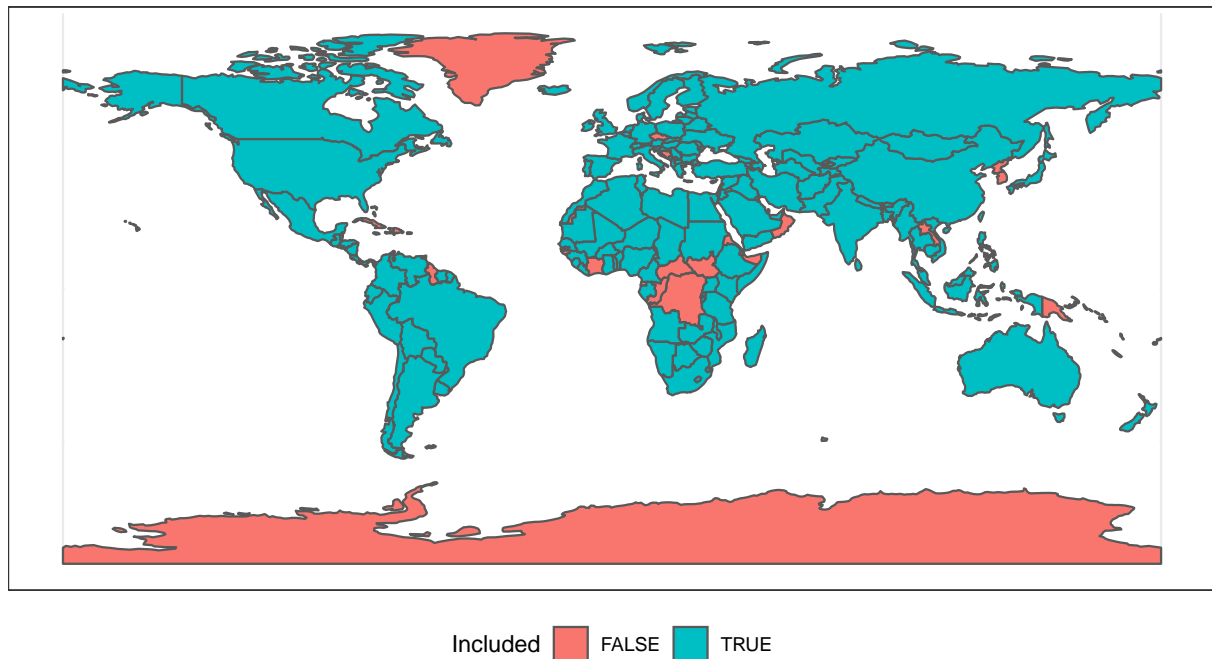


Figure 7: Colour the country that have included in Worldmap

populated or still in development, so the model is not revealing all information due to the sample selection bias. To solve this, sample selection models such as Heckman model or Tobit can be considered in Econometrics areas.

6.4 Residual Diagnostic for the Regression Model

From the Residual and Fitted plot in Figure 9, we can see it has a non-constant variance across the fitted value, which indicates the presence of Heteroskedasticity. Moreover, the error distribution (see Figure 8) and the Q-Q plot Figure (see both Figure 10 and Figure 11) also suggest the density of our model is somewhat to a normal distribution but influenced by outliers.

In conclusion, our model did a good job in explaining the relationships between happiness score and our selected aggressors even though there are still some limitations.

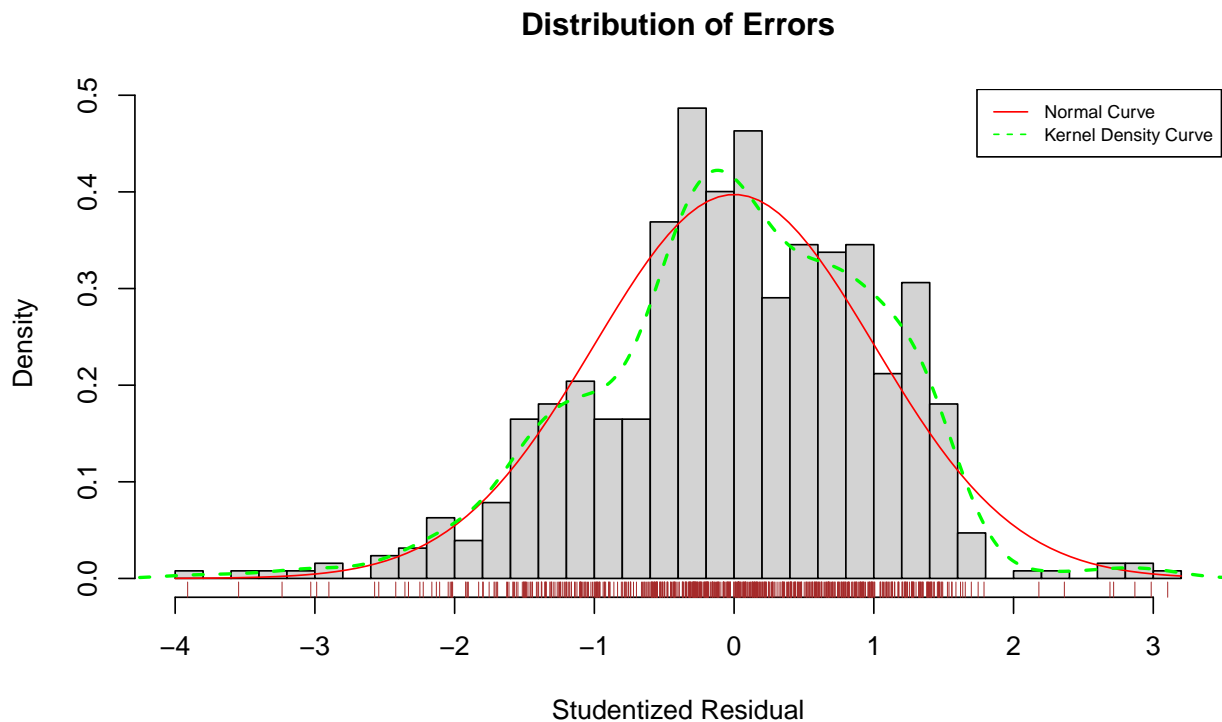


Figure 8: *Distribution of the residual term*

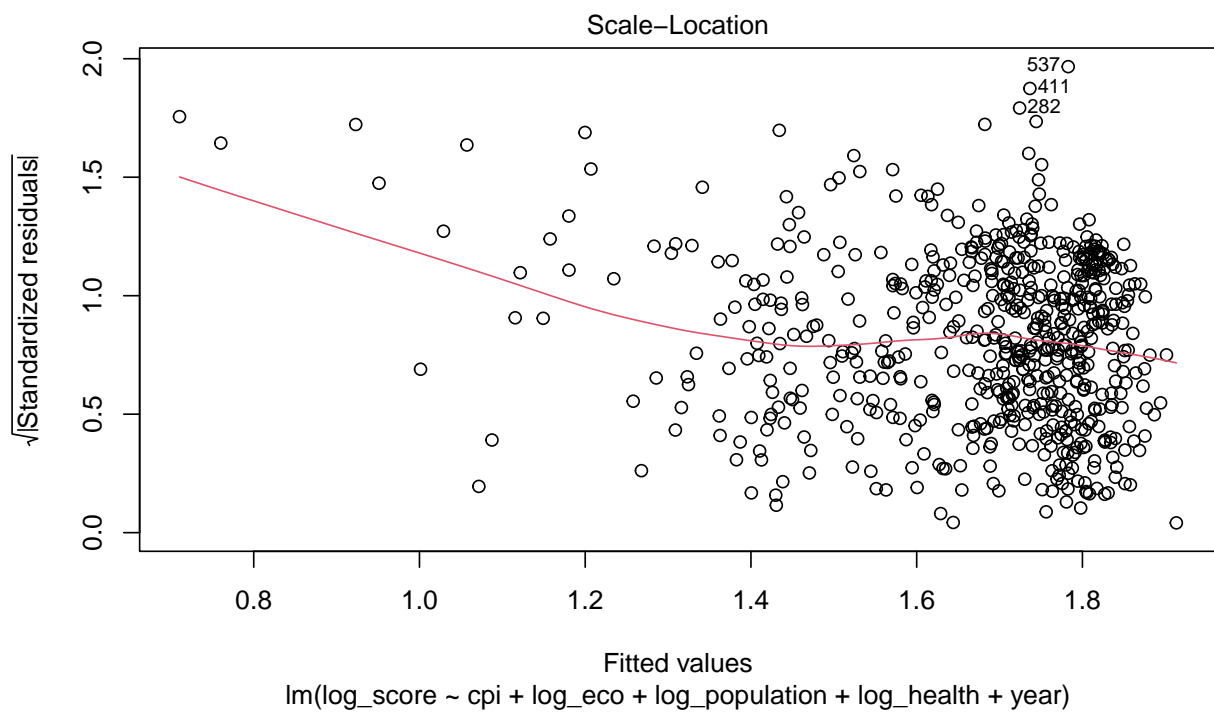


Figure 9: Residual vs Fitted value plot (with standardised residuals)

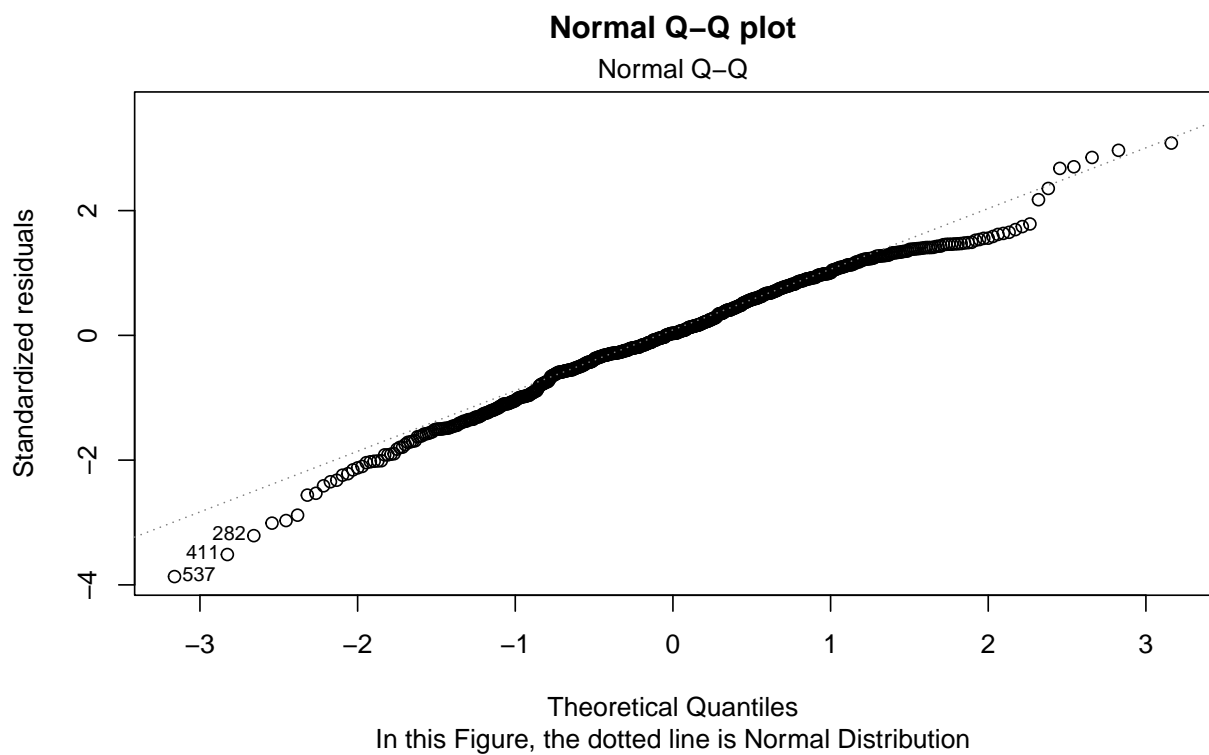


Figure 10: Q-Q plot fitted model residual density vs normal distribution residual density

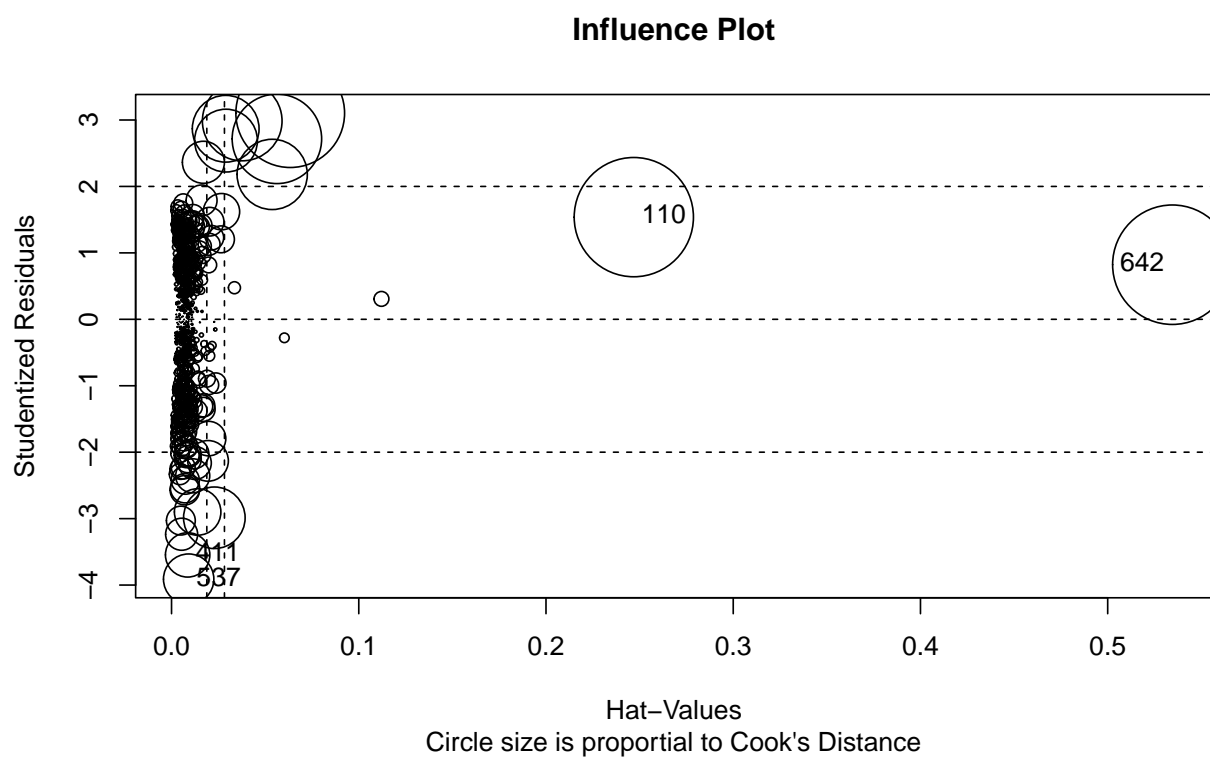


Figure 11: *Influence Plot, the bigger circle means outliers*

7 Conclusion

According to above the analysis, we can conclude that the Happiness Scores are closely related to the economic activities and health status. In those developed countries with better living and health conditions, people generally feel happier than those people living in rural areas.

On the other hand, we find that the happiness scores for people living in Sub-Saharan and Southeast Asia are relatively low, which is mainly due to the poor economic development and harsh living environment (Helliwell et al., 2021). In terms of the happiness scores at region level, Europe is the most livable continent with majority of countries higher than the average level. Even during the pandemic period, people in European countries still feel happier than other regions.

Apart from this, it is evident that during the COVID-19 period, Health, Economy and Social Support have a close relationship with the happiness score. The correlation between health and happiness is much stronger than before.

Through modelling, over years from 2015 to 2022, two most important variables in explaining the happiness scores across different countries are Economy (measured in GDP per capita) and Health. Based on the results in the final multivariate linear model, Economy is the most significant and essential factor.

Furthermore, using the advantage of linear model, both Economy and Health status have positive marginal effect on the happiness score, while CPI and Population have negative impacts.

Nevertheless, we still need to deal with the presence of heteroskedasticity, the endogeneity bias and possible sample selection bias to improve the accuracy and usefulness of our model. Further studies can be done to minimizing these limitations and come up with a better modelling for global macroeconomic data.

8 Acknowledgement

We shall never forget the hard works done by our lecturer **Patricia** on every Tuesday nights from Week 1 to Week 12.

We will also remember our tutors **Naveen** and **Fan**. They performed hard works on our tutorials and gave us nice recommendations during the presentation session.

Moreover, we'd like to thank all **EBS scholars** from **Monash University** for their excellent monash package produced such a nice report template.

Without any of their help, our works could not get finished so smoothly.

Overall, I'd also like to thank those package developers who fight for the open source software, which made us easier to plot so many beautiful plots.

Here are the lists of packages we used:

- package tidyverse (Wickham et al., [2019](#))
- package GGally (Schloerke et al., [2021](#))
- package ggplot2 (Wickham, [2016](#))
- package readr (Wickham, Hester, & Bryan, [2022](#))
- package dplyr (Wickham et al., [2022](#))
- package gridExtra (Auguie, [2017](#))
- package RColorBrewer (Neuwirth, [2014](#))
- package gganimate (Pedersen & Robinson, [2020](#))
- package repr (Angerer, Kluyver, & Schulz, [2022](#))
- package neuralnet (Fritsch, Guenther, & Wright, [2019](#))
- package ggpubr (Kassambara, [2020](#))
- package caTools (Tuszynski, [2021](#))
- package rworldmap (South, [2011](#))
- package rnaturalearth (South, [2017](#))
- package sf (Pebesma, [2018](#))
- package car (Fox & Weisberg, [2019](#))
- package kableExtra (Zhu, [2021](#))
- package here (Müller, [2020](#))
- package lubridate (Grolemund & Wickham, [2011](#))
- package cowplot (Wilke, [2020](#))
- package janitor (Firke, [2021](#))
- package maps (code by Richard A. Becker, version by Ray Brownrigg. Enhancements by Thomas P Minka, & Deckmyn., [2021](#))
- package plotly (Sievert, [2020](#))
- package wbstats (Piburn, [2020](#))
- package lemon (Edwards, [2020](#))
- package sp (Bivand, Pebesma, & Gomez-Rubio, [2013](#))

9 Appendix

We here take the Economy Situation (measured in GDP per capita) as an example:

$$\frac{\partial \log(\text{Happiness Score})}{\partial \log(\text{Economy})} = \text{ME}(\text{Coefficient})$$

From (1) we can say, when other variables remain constant

$$\Delta \log(\text{Happiness Score}) = \text{ME}(\text{Coefficient}) \times \Delta \log(\text{Economy})$$

By using the infinite approaching approximation of log function, we can know

$$\begin{aligned} \log(x_0 + \Delta x) - \log(x_0) &\approx \log(x_0) + \log'(x_0)\Delta x - \log(x_0) \\ &= \frac{\Delta x}{x_0} = \text{percentage change in } x \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\Delta \log(\text{Happiness Score})}{\Delta \log(\text{Economy})} &\approx \frac{\Delta \text{Happiness Score}}{\Delta \text{Economy}} \frac{\text{Economy}}{\text{Happiness Score}} \\ &= \frac{\% \Delta \text{Happiness Score}}{\% \Delta \text{Economy}} \\ &= \text{ME}(\text{Coefficient}) \end{aligned}$$

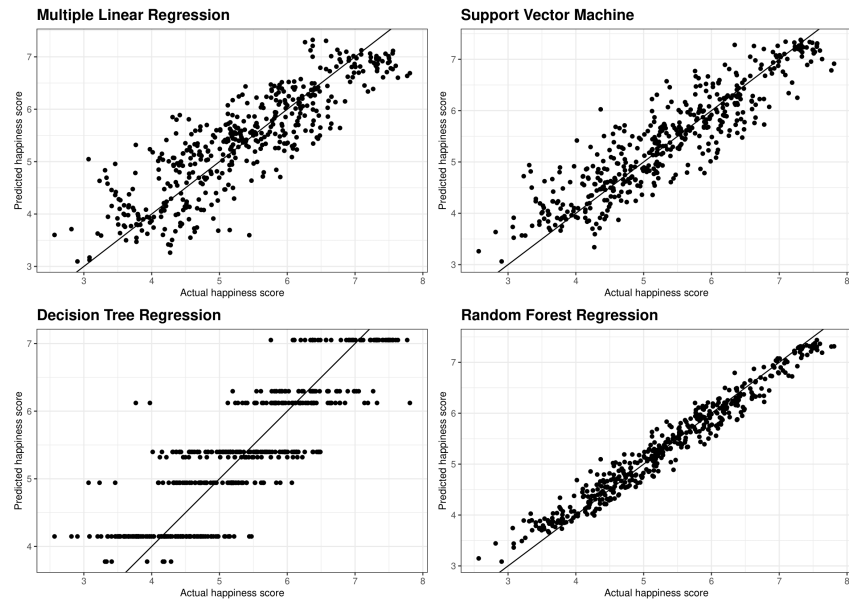


Figure 12: Models' performance under the training test split ratio = 0.4

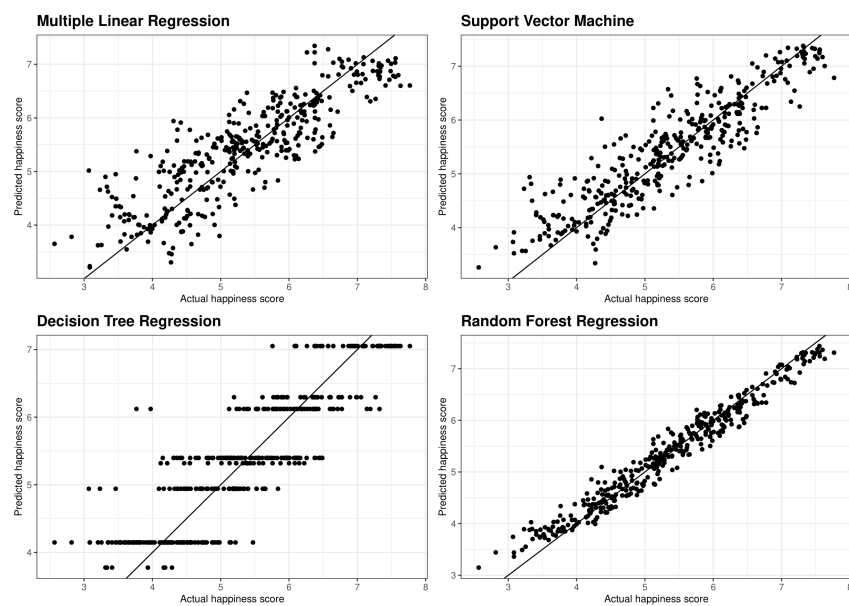


Figure 13: Models' performance under the training test split ratio = 0.5

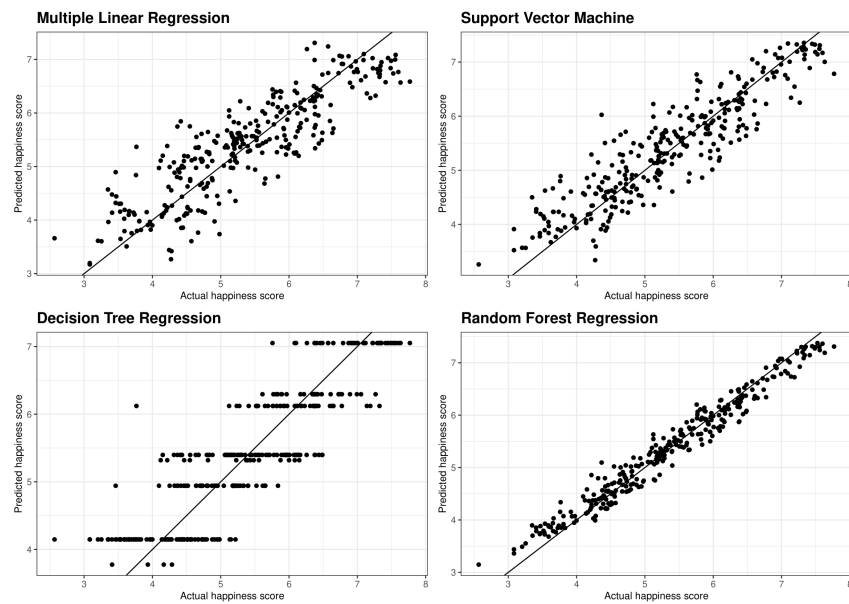


Figure 14: *Models' performance under the training test split ratio = 0.6*

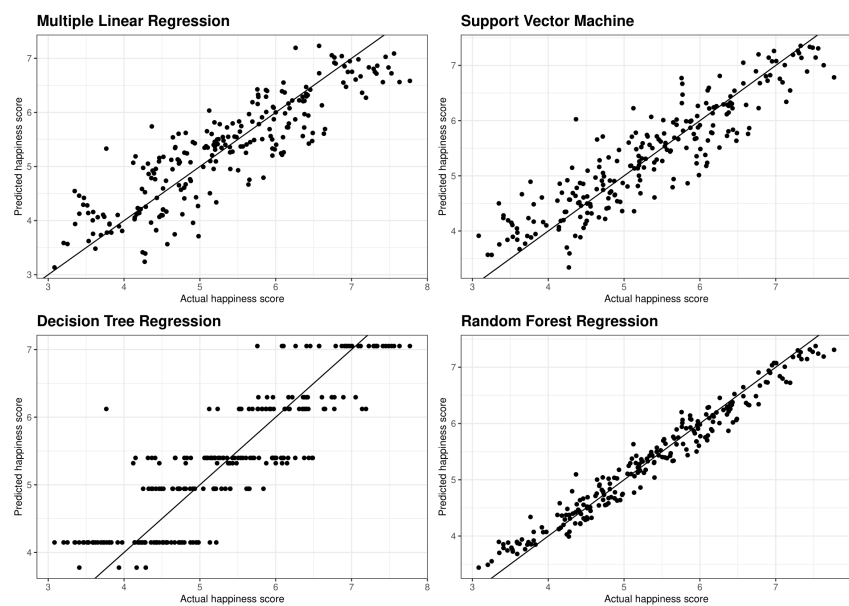


Figure 15: *Models' performance under the training test split ratio = 0.7*

References

- Angerer, P, Kluyver, T, & Schulz, J. (2022). *Repr: Serializable representations* [R package version 1.1.4].
<https://CRAN.R-project.org/package=repr>
- Auguie, B. (2017). *Gridextra: Miscellaneous functions for "grid" graphics* [R package version 2.3].
<https://CRAN.R-project.org/package=gridExtra>
- Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, **22**(1), 23–36.
- Bivand, RS, Pebesma, E, & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, second edition*. Springer, NY. <https://asdar-book.org/>
- code by Richard A. Becker, OS, version by Ray Brownrigg. Enhancements by Thomas P Minka, ARWR, & Deckmyn., A. (2021). *Maps: Draw geographical maps* [R package version 3.4.0].
<https://CRAN.R-project.org/package=maps>
- Edwards, SM. (2020). *Lemon: Freshing up your 'ggplot2' plots* [R package version 0.4.5]. <https://CRAN.R-project.org/package=lemon>
- Firke, S. (2021). *Janitor: Simple tools for examining and cleaning dirty data* [R package version 2.1.0].
<https://CRAN.R-project.org/package=janitor>
- Fox, J, & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Fritsch, S, Guenther, F, & Wright, MN. (2019). *Neuralnet: Training of neural networks* [R package version 1.44.2]. <https://CRAN.R-project.org/package=neuralnet>
- Grolemund, G, & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, **40**(3), 1–25.
- Helliwell, JF, Layard, R, Sachs, JD, & Neve, JED. (2021). World happiness report 2021.
- Helliwell, JF, & Wang, S. (2012). The state of world happiness. *World happiness report*, 10–57.
- Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots* [R package version 0.4.0].
<https://CRAN.R-project.org/package=ggpubr>
- Lappi-Seppälä, T. (2006). Finland: A model of tolerance. *Comparative youth justice*, 177–195.
- Liberian, N. (2017). Decision trees and random forests. *Towards Data Science*.
- Müller, K. (2020). *Here: A simpler way to find your files* [R package version 1.0.1]. <https://CRAN.R-project.org/package=here>
- Neuwirth, E. (2014). *Rcolorbrewer: Colorbrewer palettes* [R package version 1.1-2]. <https://CRAN.R-project.org/package=RColorBrewer>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, **10**(1), 439–446.

- Pedersen, TL, & Robinson, D. (2020). *Gganimate: A grammar of animated graphics* [R package version 1.0.7]. <https://CRAN.R-project.org/package=gganimate>
- Piburn, J. (2020). *Wbstats: Programmatic access to the world bank api*. Oak Ridge National Laboratory. Oak Ridge, Tennessee. <https://doi.org/10.11578/dc.20171025.1827>
- Rojas, M, & Vittersø, J. (2010). Conceptual referent for happiness: Cross-country comparisons. *Journal of Social Research & Policy*, 1(2), 103.
- Schloerke, B, Cook, D, Larmarange, J, Briatte, F, Marbach, M, Thoen, E, Elberg, A, & Crowley, J. (2021). *Ggally: Extension to 'ggplot2'* [R package version 2.1.2]. <https://CRAN.R-project.org/package=GGally>
- Sievert, C. (2020). *Interactive web-based data visualization with r, plotly, and shiny*. Chapman; Hall/CRC. <https://plotly-r.com>
- South, A. (2011). Rworldmap: A new r package for mapping global data. *The R Journal*, 3(1), 35–43.
- South, A. (2017). *Rnaturalearth: World map data from natural earth* [R package version 0.1.0]. <https://CRAN.R-project.org/package=rnaturalearth>
- Tuszynski, J. (2021). *Catools: Tools: Moving window statistics, gif, base64, roc auc, etc* [R package version 1.18.2]. <https://CRAN.R-project.org/package=caTools>
- Voxco. (2022). *Multivariate regression: Definition, example and steps*.
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H, Averick, M, Bryan, J, Chang, W, McGowan, LD, François, R, Golemund, G, Hayes, A, Henry, L, Hester, J, Kuhn, M, Pedersen, TL, Miller, E, Bache, SM, Müller, K, Ooms, J, Robinson, D, Seidel, DP, Spinu, V, ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Wickham, H, François, R, Henry, L, & Müller, K. (2022). *Dplyr: A grammar of data manipulation* [R package version 1.0.8]. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H, Hester, J, & Bryan, J. (2022). *Readr: Read rectangular text data* [R package version 2.1.2]. <https://CRAN.R-project.org/package=readr>
- Wilke, CO. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'* [R package version 1.1.1]. <https://CRAN.R-project.org/package=cowplot>
- World Bank. (2022). *World bank data*.
- Zhu, H. (2021). *Kableextra: Construct complex table with 'kable' and pipe syntax* [R package version 1.3.4]. <https://CRAN.R-project.org/package=kableExtra>