

# Does Money Make People Attain Freedom?

*Yuta Hayashi*

*1/10/2019*

## Abstract

The goal of this project is to predict the factors that influence human freedom index nationally in different years. The data includes multiple variables that not only describes degree of human rights but also economic freedom. Most of our variables are quantitative, but we also have categorical data for yes or no response to homosexual relationship, year, and countries. We create three models from the data that each has its own meaning to describe expected human freedom index. We compare those models base on F statistic,  $R^2$ , and statistical significance of each variable within a model. All of three data produced statistically significant prediction results with important and interesting interpretation. Specifically this project focuses on the most predictive variables such as trade and same\_sex variables with intriguing interpretations.

## 1) Background and significance

The Human Freedom Index casts a wide net in an attempt to capture as broad a set of freedoms as could be clearly identified and measured. But what is the definition of human freedom? The validity of human freedom scale has been contested among scholars. It is often the case that sociologists and anthropologists interpret freedom as subjective matter and not comparable across individuals. They rather use it as more subjective term like satisfaction to describe the idea of human freedom in the local contexts.[1] Therefore, based on their perspective, there is substantial variability in terms of regions and time. However, economists believe that better economic condition results in more freedom universally and it is not influenced by regions and time.[2] Considering universal factors that contribute to freedom such as the size of government, individual affluence, or freedom to trade is useful since they will give us a way to bring applicability and connectedness for the definition of human freedom in local contexts as well[3]. Furthermore, it is important to analyze what other factors potentially interact with the factors related economic measurement and how they amplify each other to predict the expected human freedom index. This research is very important since analyzing what makes people's freedom leads us to enhance our further happiness level with our quality of life. This project is based on a dataset that is comprised of the results of a survey conducted by the Fraser Institute. On a scale of 0 to 10(this value includes decimal numbers), where 10 represents best indication and 0 represent worst indication, they collected data with human freedom index variables and other variables for 162 countries in 2018. The purpose of this project is to address following questions: 1) we demonstrate what factors influence the expected human freedom by creating multiple regression model 2) we demonstrate whether or not the variables related to economic aspects as interaction terms contribute to forming a better multiple regression model. 3) we interpret each coefficient of a final multiple regression model.

## 2) Methods

### a. Data Collection / Cleaning

Original data comes from one of kaggle's database(<https://www.kaggle.com/gsutters/the-human-freedom-index>). It contains 1458 rows(different countries in different year) and 123 columns(4 response variable: human freedom index and rank, and economic freedom index and rank). In order to attain clean data, we go through the data cleaning process. First of all, we filter all existing variables to 13 variables. We do it, first of all, by finding variables we are interested in, next step is to take a scatter plot of human freedom index versus explanatory variable that are potentially valid to create a multiple regression model. Finally, we rename them to see clearer names. After we do those processes, we modify the dataset to make same\_sex a categorical variable (yes and no) from the original same\_sex quantitative variable. Finally, since there are substantial amount of missing values, we fill them by imputing the average value of each column to missing values. A lot of values in Africa are missing since they failed to collect the data in these regions.

## b. Variable Meanings

All of variables are measured on a 0-10 scale except for `same_sex` and country variables. As we mentioned the larger quantity for all of variables means more freedom. “association” means freedom of association and freedom of assembly. “government” variable measures taxation, government spending, and government investment in and control of enterprises. “same\_sex” variable measures yes and no response to acceptance of individual males to establish same-sex relationships. “money” variable measures the extent to freedom of using money and individual affluence. “trade” variable measures the freedom to trade internationally and measures tariff rates. “legal” variable measures the level of protection of people and their property rights. “rol” variable measures the criminal justice system on such issues as its impartiality, the degree to which improper government influence is present. “ss” variable measures disappearances, conflict, and terrorism. “movement” variable measures the freedom of domestic movement and the freedom to leave the country. “expression” variable measures a broad range of freedom, including that affecting personal expression.

## c. Model

In order to evaluate the data, we use the statistical software R to create a multiple regression model. Our main goal of creating a multiple regression model is heavily on effective and logical model that enables us to attain easy interpretability of explanatory variables. First, we divide explanatory variables into two groups: one group is a set of variables directly correlated to human freedom index. In order to do it, we get two types of correlation coefficients of human freedom index versus each explanatory variable, and economic freedom index versus each explanatory variable to categorize whether an individual is associated with human freedom or economic freedom. We pick variables that have stronger correlation to human freedom index than that of economic freedom index. By doing this process, we attain the explanatory variables: `association`, `same_sex`, `legal`, `rol`, `ss`, `movement`, `expression`. The other group is a set of the variables related to economic aspect as interaction terms: `government`, `money`, and `trade` in the same process we do for the first group. However, we pick variables that have stronger correlation to economic freedom index than that of human freedom index we get Those are also the variables we are interested as they might create insightful interpretations. With those variables, we create a multiple regression model(Model1). It is done by creating a full model in addition to all possible interaction term with three variables related to economic aspects. Then we get individual p value from the summary table in order to determine which factors are crucial to describe the expected value of human freedom index. Note that though a couple of variables related to economic aspects are not enough to describe the model, we leave these variables for two reasons. Firstly, it is because we would like to get the interpretation of interaction terms. Secondly, it is because we make a guess that their statistical significance might improve after we get rid of other variables.

In the end, we get three types of models. The first one is the full model we describe above(Model1). The second one is a model we believe it is the best model(Model2). The third one is a simple linear regression model with trade variable(Model3) since the correlation is the highest among the variables associated to economic aspects. Based on these three models we get F statistic and  $R^2$ . P value for all the models are extremely small because of the big sample size and we could not compare the value each other. Therefore, we get F statistic instead. Note that, in this hypothesis testing, F statistic indicates the statistical difference between two models in ANOVA test.

## 3) Results

From the Table1, Model1 that we throw all the explanatory variables gets F statistic of 1491.879 and an  $R^2$  of 0.9678510. Model2 that we believe the best model gets F statistic of 3011.204 and an  $R^2$  of 0.9466727. Model3 that we only use trade variable as an explanatory variable gets F statistic of 2486.016 and an  $R^2$  of 0.6306459. Based on F statistic, Model2 attains the highest F statistic, which also means that it attains the least p value. For  $R^2$ , Model1 attains the highest  $R^2$  score. However, in theory,  $R^2$  would not decrease when we add additional explanatory variables. There is also little drop from Model1 to Model2 (0.9678510 – 0.9466727), whereas there is substantial drop from Model2 to Model3 (0.9466727 – 0.6306459).

Therefore, we do not have enough evidence to conclude that Model1 has the best model only based on  $R^2$ . Each  $R^2$  tells you the proportion of variation in human freedom index that is explained by the model with given explanatory variables, and the higher  $R^2$  is, the better, though attaining too high  $R^2$  might indicate overfitting(Model1). Therefore, we conclude that Model2 is the best model among three models. Here is the model2.

Model2:

$$\widehat{human.freedom.index} = \beta_0 + \beta_1 * (association) + \beta_2 * (samesex) + \beta_3 * (rol) + \beta_4 * (ss) + \beta_5 * (expression) + \beta_6 * (trade) + \beta_7 * (government) - \beta_8 * (ss * trade)$$

$$\widehat{human.freedom.index} = -0.99 + 0.07 * (association) + 0.31 * (samesex) + 0.24 * (rol) + 0.32 * (ss) + 0.12 * (expression) + 0.43 * (trade) + 0.12 * (government) - 0.02 * (ss * trade)$$

Table 1: The Result of F statistic and  $R^2$

	F	$R^2$
Model1	1491.879	0.9678510
Model2	3011.204	0.9466727
Model3	2486.016	0.6306459

(Intercept): Expected human freedom index in the world of all the time when all of the other variables' scale is 0 with the response no to men's same-sex relationship is  $-0.99$ .

(association): Expected increase in human freedom index in the world of all the time for each additional increase of scale in association, when all of the variables are held constant is  $0.07$ .

(same-sex): Expected difference in human freedom index in the world of all the time for countries that respond yes to men's same-sex relationship, compared to countries that respond no to men's same-sex relationship, when all of other variables are held constant is  $0.31$ .

(rule of law): Expected increase in human freedom index in the world of all the time for each additional increase of scale in rule of law, when all of the variables are held constant is  $0.24$ .

(expression): Expected increase in human freedom index in the world of all the time for each additional increase of scale in expression, when all of the variables are held constant is  $0.12$ .

(government): Expected increase in human freedom index in the world of all the time for each additional increase of scale in government, when all of the variables are held constant is  $0.12$ .

$0.02$ (interaction term): we do not need to interpret it since we already interpret this when we interpret ss and trade coefficients respectively.

(ss(terrorism)): For a given trade scale  $x = a$ , expected increase in human freedom index in the world of all the time for each additional increase of scale in ss, when all of the variables are held constant is  $(0.32 - 0.02a)$ .

(trade): For a given ss scale  $x = b$ , expected increase in human freedom index in the world of all the time for each additional increase of scale in trade, when all of the variables are held constant is  $(0.43 - 0.02b)$ .

#### 4) Discussion/Conclusions

In response to the first question in section2, based on Model2, the final version of our model contains a various indicators including association, same\_sex, rule of law, ss(terrorism indicator), expression, trade, government, and the interaction term between ss and trade. Note that some variables directly indicate human freedom index and others takes economic aspects. There is an interesting finding that trade and government are not only directly correlated to economic index freedom but also human freedom index especially trade variable(see appendix). This means that a lot of economic characteristics of countries directly boost human freedom index of countries. In addition to that, the interaction term between ss and trade tells you that there is evidence that the impact of ss(terrorism) on human freedom index depends on trade. This means that the extra interaction term makes a multiple regression model better to describe the expected human freedom index. We could see further visualization in appendix. Based on the residual versus fitted graph, the linearity and constant variance assumptions appear reasonable. Also based on the qq plot, although there are a couple of outliers at the bottom, the normality assumption also looks reasonable. Therefore, those graphs indicates that this model satisfies the model assumptions.

In conclusion, this model successfully satisfy my initial goal: Does Money Make People Attain Freedom? The answer is yes with the variables related to economic aspects. Although money variable which indicates individual level of affluence is not included in model2, other variables related to economic aspects such government and trade better describes the expected value of human freedom index. However, model3 tells you that with only trade variable, thought it is reasonable model to describe the human freedom index, this lacks the complexity and model2 is much better. Thus, it is not possible to create a model that is better than the model that both includes variables correlated to human freedom index and economic freedom index.

There could be further improvement in the scaling criteria. As we could see in the appendix, scatter plot of association variable is not smooth enough to interpret the coefficient of association variable. When collecting data, it is crucial to set the same measurement. In addition to that, it would be an interesting regression model creation if the sample size is smaller than this dataset since we attain smaller values for p value as we could see in the summary table in appendix. Personal thought is that this makes the data analysis less meaningful and exciting

#### 5) References

[1]

Ng, Y. (1997). *A case for happiness, cardinalism and interpersonal comparability*. Clayton, Vic.: Monash University.<https://doi.org/10.1111/j.1468-0297.1997.tb00087.x>

[2]

Happiness and Economics. (2010) *How the Economy and Institutions Affect Human Well-Being*. Princeton: Princeton University Press.[https://books.google.com/books?id=XD8-H35mspoC&printsec=frontcover&hl=ja&source=gbs\\_ge\\_summary\\_r&cad=0](https://books.google.com/books?id=XD8-H35mspoC&printsec=frontcover&hl=ja&source=gbs_ge_summary_r&cad=0)

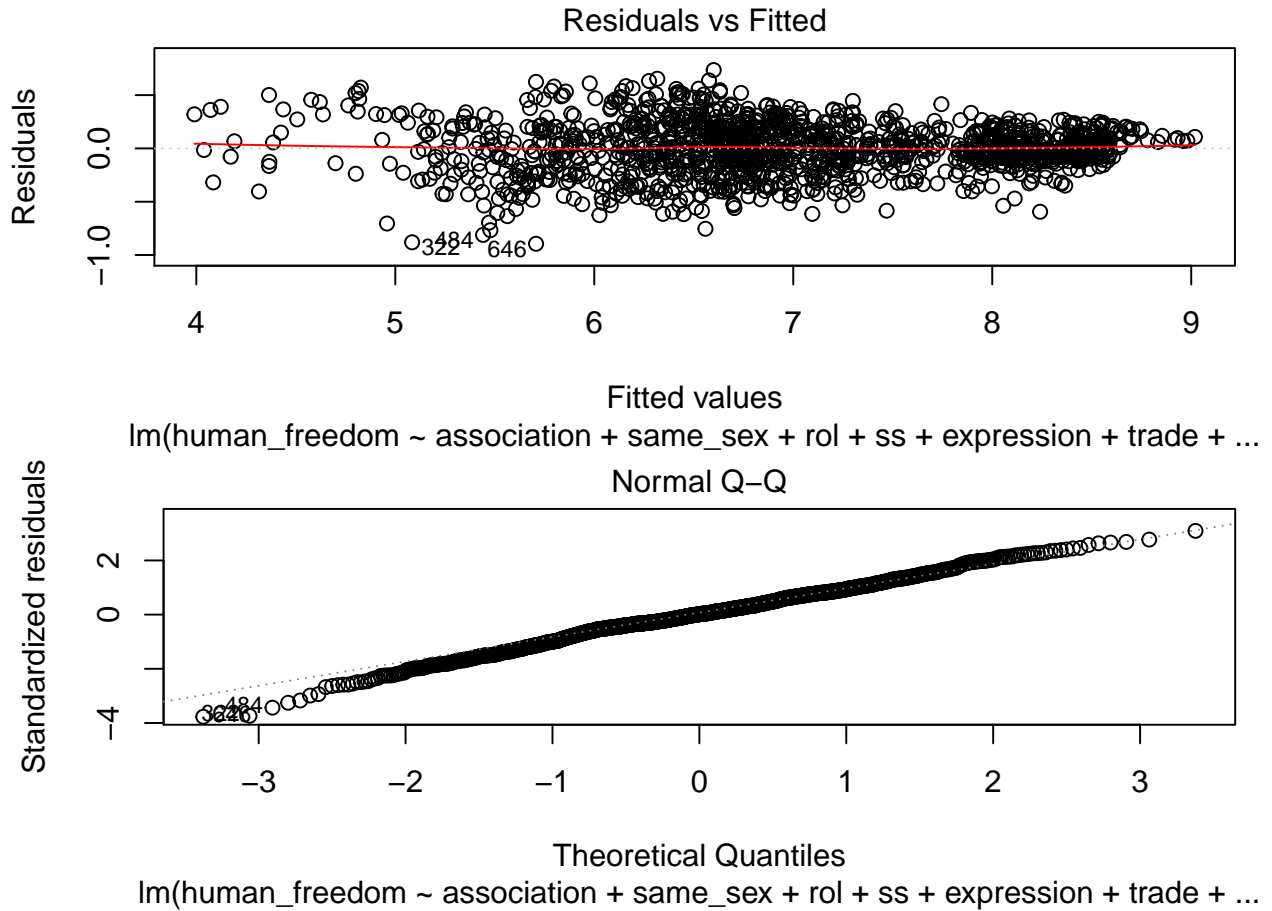
[3]

Ingstad, B., & Whyte, S. R. (2007). *Disability in local and global worlds*. Berkeley: University of California Press.

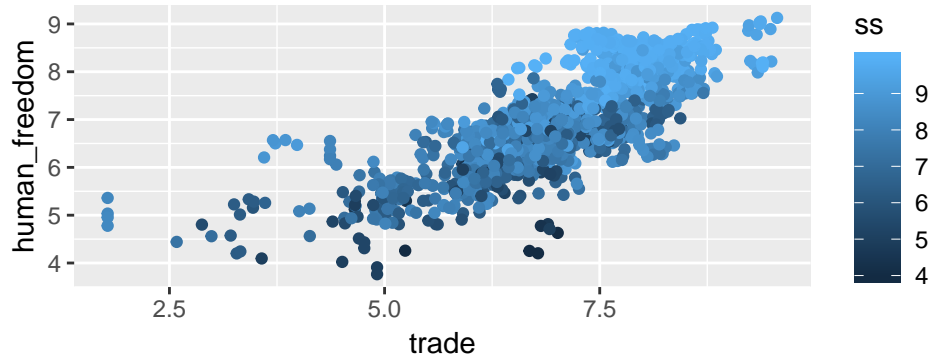
6) Appendix

Table 2: Model2 Summary Table

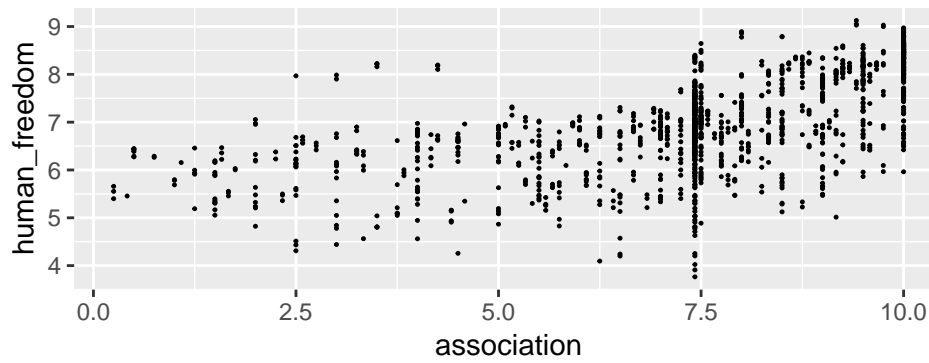
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.9901344	0.2627972	-3.767675	0.0001718
association	0.0714377	0.0042210	16.924192	0.0000000
same_sexYes	0.3087087	0.0155934	19.797406	0.0000000
rol	0.2400908	0.0073742	32.558226	0.0000000
ss	0.3167430	0.0335035	9.454032	0.0000000
expression	0.1174517	0.0067099	17.504324	0.0000000
trade	0.4274873	0.0381213	11.213874	0.0000000
government	0.1188492	0.0056879	20.895271	0.0000000
ss:trade	-0.0225703	0.0048105	-4.691924	0.0000030



Human Freedom Index Versus Trade for Countries With Different Scale of ss(terrorism)



Human Freedom Index Versus Association



7) Data

Original data is from (kaggle's database(<https://www.kaggle.com/gsutters/the-human-freedom-index>)).