“It Only Gets Better from Here:”

Evaluating the Effectiveness of GDP as a Predictor for Quality of Life

Justin Puleio

Data Science

Professor McCabe

December 18 2023

Introduction

The interconnectedness of society in an increasingly globalized world is a force which is hard to understate. Modern advancements have created an ever-tightening link between members of a society and the nations that govern them, creating distinct and apparent influences on how people live. The analysis laid out in this paper seeks to analyze the potential existence of these connections, namely how large-scale factors such as GDP often may influence the lives of individuals within a society through means such as their quality of life. The question this paper seeks to discover is thus outlined: is GDP linked to quality of life improvements? The significance of this question is apparent: if true, this relationship can allow for applications which can aid in predicting or evaluating the state of nations or regions globally. Predictive models could serve to identify struggling nations or highlight ones which might soon be, giving governments and other actors the information to effectively and successfully target specific sectors to avoid or improve undesirable conditions. Having substantial potential as a tool for improving security and stability on a global scale, the fundamental understanding of how these two variables move is key in deciphering an ever-complex and ever-globalizing world.

The approach taken during the research of this relationship was centralized around discovery. Extracted and analyzed from data compiled by the World Bank, this approach observed two variables, GDP and quality of life, and evaluated how they could be compared and combined into a predictive model. From there, these models were further evaluated on a country-to-country scale, compared and analyzed together to attempt to further prove or disprove the existence of some kind of relationship between variables. Ultimately, the results of this research found these predictive models to be somewhat viable, leading to further predictions to outline the potential for future applications consistent with the findings of this research.

Background

The existence of a relationship between GDP and quality of life is not necessarily a new concept. Various studies and observations exist which attempt to equate the two together or discover some kind of relationship, each to a varying degree of success. Despite the ample amounts of research, the viability of these models often vary due to the muddled nature of the variables studied. For instance, the quantification of "life quality" is fairly difficult, with no true metric in how it is defined. The question of how to quantify variables such as leisure in a study poses problems due to the difference in preferences or ability; some may choose to be leisurely but be able to work while others may do so because they are unable to for durations they would desire, and others still may be leisurely but be unable to work at all (Dynan and Sheiner 2018). Assigned values on key factors of study such as leisure are difficult at best and completely inaccurate at others; it is hard to quantify and separate personal and bounded choices which have no standardized metric. As a consequence, conclusions in such an analysis must be taken with a grain of salt. Similarly, the expectations of this research, while they ultimately seek to establish the existence of a relationship between GDP and quality of life, also serve as a test which seeks to evaluate their viability as a means of observing their relationship and predicting their trajectories. The results of this research attempts to create a model which is expected to replicate results consistent with works such as that of Proto et al., namely that "[quality of life] strongly increases with GDP…but the relation becomes much less steep…[and] flattens" after a specific GDP threshold (Proto et al. 2013). The strong increase with GDP refers to a massive rise in quality of life during initial increases from a low GDP. Under this assumption, countries experiencing industrialization or substantial economic growth, such as developing countries with smaller economies, would observe substantial increases in quality of life, leading to a steeper relationship. Oppositely, countries who already have or are in the process of industrialization are going to have comparatively shallow relationships; the benefits associated with increased GDP will reach that critical threshold in which quality of life will have reached a relatively high level and begun to flatten. It would thus be expected that any potential graph measuring GDP against quality of life would reflect this in the form of a curved arc. Along with various other methods of determining viability, this result is one of the primary criteria in which the prediction model will be evaluated.

Data and Approach

Data was evaluated from the World Bank Data set provided by the World Bank. The data set, which is a 44-variable CSV data frame, evaluates an expansive list of various world indicators across 216 countries over the period of 1980-2020. Indicators include, but are not limited to: access to electricity as a percentage of the population, HIV rate as a percentage of the population, GDP, and GDP growth rates, as well as various other useful classifications such as region and latitude/longitude. While there are various variables available for extraction in the data set, not all were used in this research (see Appendix A for a complete list of utilized variables and their definitions). Moreover, additional variables were also created and analyzed based on existing variables in the set. To evaluate the relationship between quality of life and GDP, a metric measuring quality of life had to first be established. Variables were chosen based on quantity of workable entries and their contribution to the reduction of the prediction error. Entries were deemed “workable” based on their ability to provide enough data entries for a wide range of countries. Variables such as EG.ELC.ACCS.ZS largely provided complete and expansive data across multiple countries, regions, and years; variables such as SL.TLF.0714.ZS did not, and were thus were not considered to be incorporated as a factor into the “quality of life” metric. Additionally, variables had to be compatible in the creation of a basic "quality of life" score. This was determined by evaluating the unit of variables: variables such as EG.ELC.ACCS.ZS and SE.ADT.LITR.ZS both referred to percentages of the population, and were thus compatible, whereas variables such as EN.ATM.CO2E.PC, which refers to a weight in metric tons, would not be compatible for calculating a basic score. Eventually, four variables were settled on in determining the "quality of life" of a particular country: SE.ADT.LITR.ZS, EG.ELC.ACCS.ZS, IT.NET.USER.ZS, and SH.DYN.AIDS.ZS (See Appendix A for more information). From there, life.quality was assigned to each entry as a separate variable, calculated by adding the percentages of each variable and dividing by four to result in a "quality of life" score based on a scale of 100, with 100 being the highest score and best quality of life. This was concluded on the basis that countries with higher qualities of life would see more of their population contain high percentages of each of the variables, with a life.quality score of 100 representing 100% of the population maintaining all four variables. It should be noted however that life.quality considers the inversion of SH.DYN.AIDS.ZS, notated as non.HIV.population, in order to ensure compatibility with other variables. Of all the countries which could receive a life.quality score, mean life.quality was roughly 46.30, with a range between 5.64 and 72.74.

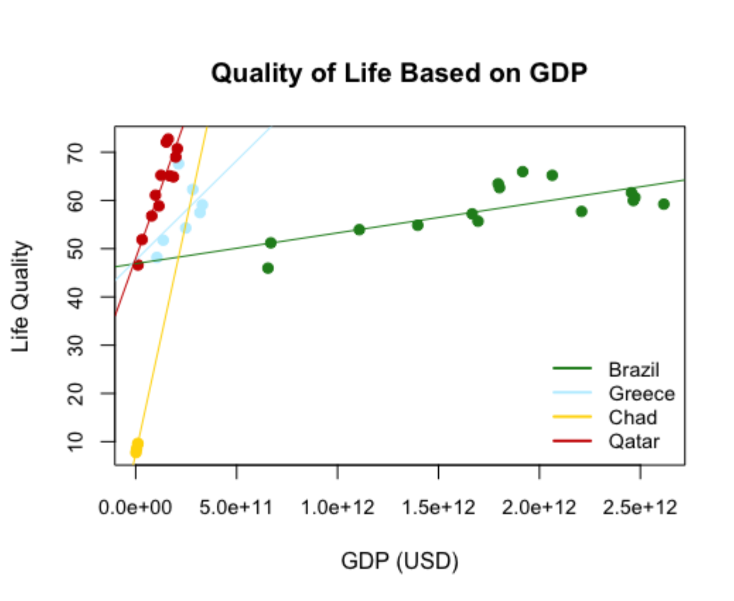
The established life.quality score then created a basis for further evaluation. The original file, wdidata20, was subset into wdidata20.quality, containing only complete entries which contained all four variables necessary to calculate life.quality score. This was then further subset into the countries and countries2 datasets, created to store means and other general information in which only one entry per country was required. These datasets were used in combination with each other to calculate and store additional variables such as prediction.error. The variable prediction.error utilized the wdidata20.quality dataset to extract all life.quality scores and

NY.GDP.MKTP.CD measurements from a particular country and calculate a linear regression from which a prediction error could be calculated using the root-means-square error (RMSE) method. Linear regressions were unique to individual countries in order to establish unique prediction errors per country in which the life.quality regression model (LQRM) could be compared. Further refining such as the removal of countries with two or less entries, which eliminated perfect linear regressions with a “false zero” prediction error, helped to improve the evaluation of the model and its viability (See Appendix B for comparison).

Results

The results of this analysis reinforced the notion that GDP is linked to quality of life to some degree. While countries produced varying prediction errors, they remained between the range of 0.06 and 5.24 with a mean prediction error of 1.92, or roughly 4% of the mean life quality. Figure 1 below examines four sample countries and their respective regression models:

Fig. 1: Quality of Life Based on GDP



Regression models of four countries. Countries were selected based on prediction error, economy size, and number of entries to evaluate LQRM across differing factors.

Brazil has a large economy and multiple entries with an intermediate prediction error.

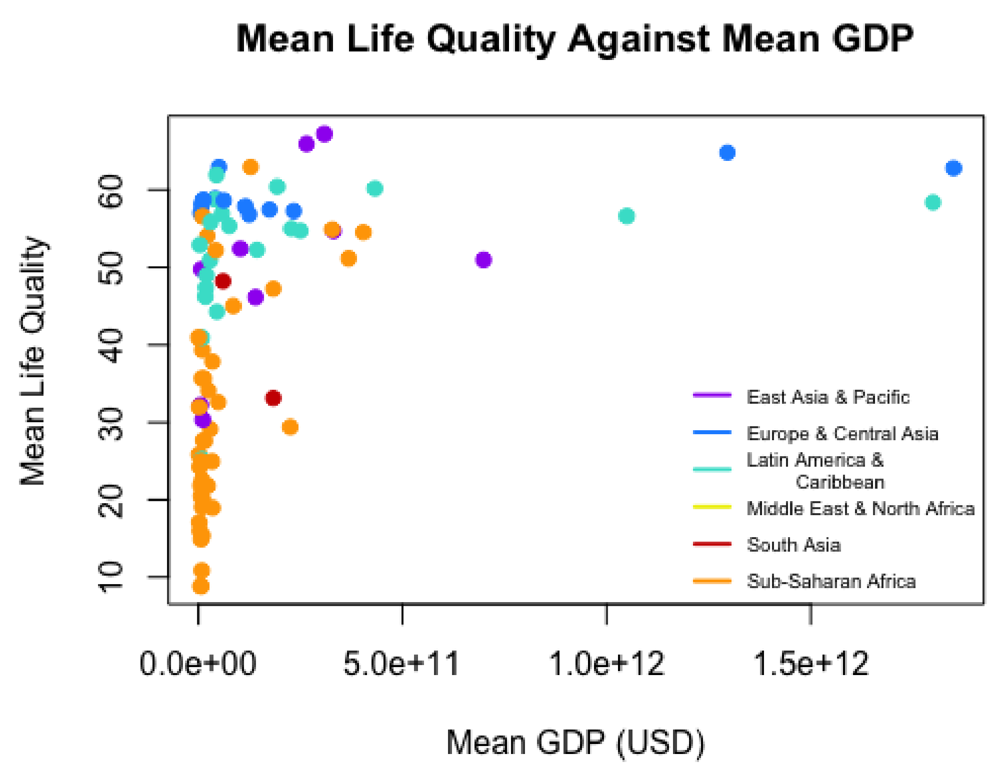
Greece has an intermediate economy and intermediate entries with a high prediction error.

Chad has a small economy and the fewest entries with the lowest prediction error.

Qatar has an intermediate economy and intermediate entries with an intermediate prediction error.

The figure observes the expected positive relationship between GDP and quality of life. As expected, countries such as Chad have steep slopes, highlighting the substantial increases in quality of life that GDP increased would bring. Likewise, intermediate economies such as Qatar and Greece show less steep slopes, showing a "leveling off" consistent with the expectations established by Proto et al. Finally, large economies such as Brazil observe the shallowest slopes, with GDP changes only incrementally increasing quality of life over time. This trend is further supported when graphing mean.lq against mean.gdp, as outlined in Figure 2:

Fig. 2: Mean Life Quality Against Mean GDP



Scatter plot of every country by their mean.lq and mean.gdp. Countries are split by region, denoted by color.

Figure 2 further supports the arguments made by Proto et al. As predicted, life.quality projects a correlation which appears to follow a curved arc, with substantial rises in mean life quality as mean GDP increases. Developing regions such as Sub-Saharan Africa, which contains countries with primarily small economies, are shown to follow a trend which would favor a steep positive regression. Oppositely, other regions such as Latin America and the Caribbean have approached the "leveling off" point, reflected in a cluster of points in the top left of Figure 2 that seems to observe a shallower, but somewhat steep, curve. Prediction error also seemed to be consistent across regions, as noted by Table 1 below:

Tab. 1: Prediction Error Significance Across Regions

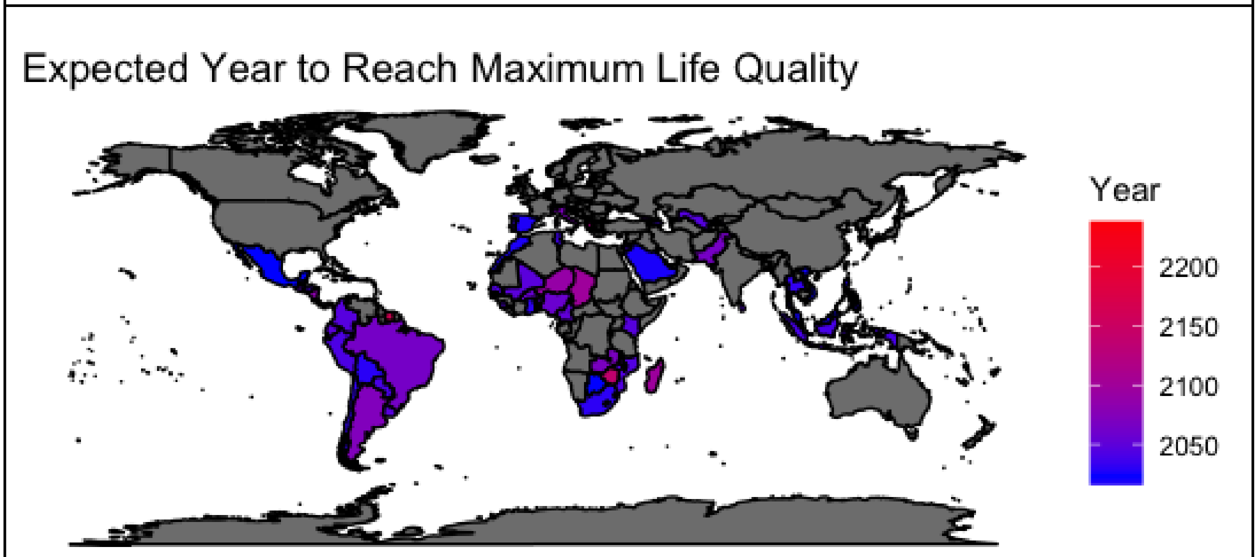
Region	Mean Prediction Error	Mean Life Quality	% of Regional Mean
East Asia / Pacific	1.786572	49.95335	3.5
Europe / Central Asia	3.055684	59.21333	5.2
Latin America / Caribbean	1.475912	52.14982	2.8
Middle East / North Africa	2.538839	52.42813	4.8
South Asia	0.8897491	40.65976	2.2
Sub-Saharan Africa	1.702891	26.55969	6.4

Table displaying significance of prediction error as a percentage of mean of life quality per region.

All regions show a prediction error of around 5% of their regional mean life quality, suggesting a fair amount of accuracy across regions. Further evaluations were made to determine whether the number of entries per country affected prediction error, though it was determined that this had minimal effect (See Appendix C for further details).

The establishment of the relationship between quality of life and GDP allows for some interesting predictions. For example, the creation of a LQRM for individual countries allows for individual predictions based on GDP trends. To elaborate, individual LQRMs can be used to predict how a country might develop over time given GDP trends, including how long it might take for it to reach a certain life.quality score. Figure 3 proposes this possibility using the linear, LQRM model to determine the year in which select countries will reach a life.quality score of 100:

Fig. 3: Expected Year to Reach Maximum Life Quality



Map showing select countries and the expected year for them to reach a life.quality score of 100. Countries projected to reach the maximum life.quality score closer to 2023 are depicted more blue, whereas further projections are depicted closer to red. Any country which was projected to already have hit their maximum life.quality score are green. Countries with no projection remain gray.

Predictions such as these can serve to provide interesting graphics showcasing regional development as well as useful insight on the trajectory of specific countries. These can be used to assess the health of regions or a particular country, and show specific areas which might need to be targeted globally by world-players in order to improve global conditions. While utilizing varying models may affect accuracy, these predictions provide a wealth of basic information which can be utilized in a wide array of ways.

Conclusion

It would be incorrect to assume that such evaluations did not come without error, however. Various limitations serve to create "holes" in the data in which accuracy cannot be tested. For example, in order to compile the life.quality score, many entries had to be removed from consideration due to their "incompleteness:" i.e. entries which did not contain all four variables. This removed valuable data, causing entire countries and regions such as North America to be removed from analysis. Furthermore, while the LQRM was somewhat accurate and consistent with the hypotheses, it is limited as a long-term predictor; the linear nature of the model means it cannot accommodate predictions consistent with the "leveling off" which was found to occur as GDP increases beyond a specific threshold. More advanced predictions, which could project a curved, nonlinear regression would likely be able to remedy this issue, though further testing and analysis would be required to evaluate any such models.

The role of GDP in a country's quality of life cannot be understated. As concluded by the observations made, GDP has some positive relationship with quality of life, in which increases to GDP often see a simultaneous rise in quality of life. As showcased by the research outlined in this paper, the LQRM produced is consistent with the expectations and hypotheses previously outlined. This affirmation is significant in the possibilities it allows for; the discovery of a relationship between these two variables provides a better understanding of an increasingly complex world. These complex societal networks, knowing their relationship with other factors, can thus be better measured, observed, and manipulated, creating a plethora of opportunities and advantages which can be utilized by further study of this subject.

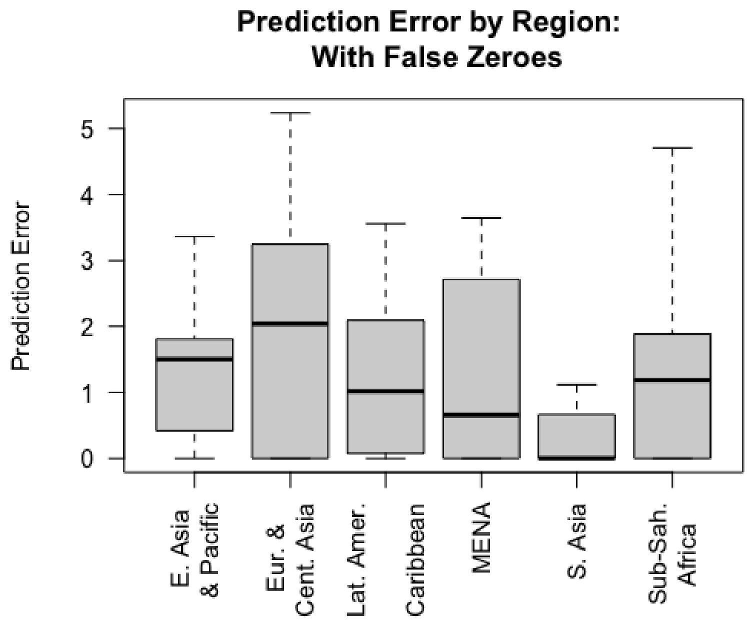
Appendix A

Variables Utilized in Research

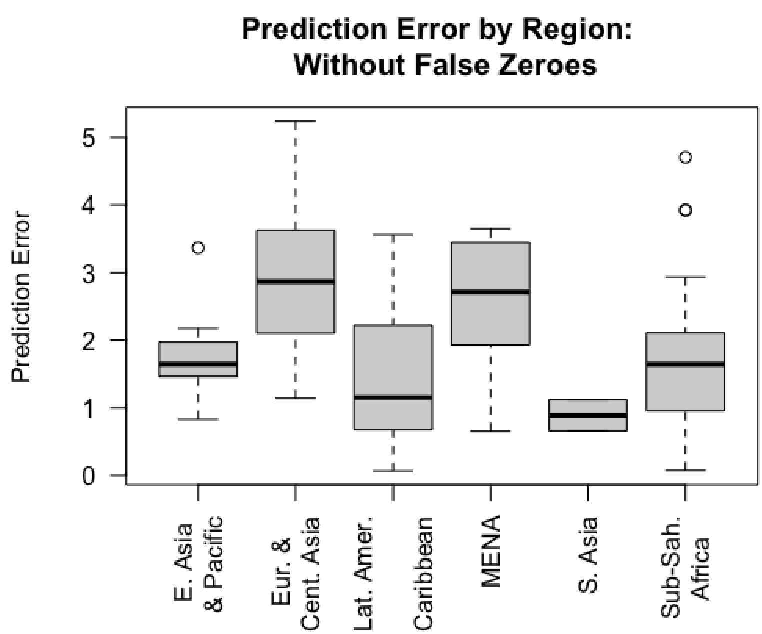
Name	Dataset Location	Variable Type	Original WDI Variable	Description
country	wdidata20.quality, countries, countries2	grouping	Yes	Country
year	wdidata20.quality	grouping	Yes	Year of WDI Indicator
region	wdidata20.quality, countries, countries2	grouping	Yes	Region group of country: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, South Asia, Sub-Saharan Africa
SE.ADT.LITR.ZS	wdidata20.quality	grouping	Yes	Literacy rate, adult total (% of people ages 15 and above)
EG.ELC.ACCS.ZS	wdidata20.quality	grouping	Yes	Access to electricity
IT.NET.USER.ZS	wdidata20.quality	grouping	Yes	Individuals using the Internet (% of population)
SH.DYN.AIDS.ZS	wdidata20.quality	grouping	Yes	Prevalence of HIV, total (% of population ages 15-49)
NY.GDP.MKTP.CD	wdidata20.quality	treatment	Yes	GDP current US$
NY.GDP.MKTP.KD.ZG	wdidata20.quality	treatment	Yes	GDP annual growth (%)
non.HIV.population	wdidata20.quality	grouping	No	Non-HIV-affected population, % (inverse of SH.DYN.AIDS.ZS)
life.quality	wdidata20.quality	outcome	No	Life quality metric (scale of 0-100)
prediction.error	countries, countries2	outcome	No	Prediction error of linear regression, per country

false.zero	countries, countries2	grouping	No	Prediction error of 0 based on having two entries or less, per country (0= no error, 1= false zero)
mean.lq	countries2	outcome	No	Mean life quality, per country (scale of 0-100)
mean.gdp	countries2	treatment	No	Mean GDP, per country (scale of 0-100)
entries	countries2	grouping	No	Frequency of entries in wdidata20.quality, per country
most.recent.year	countries2	grouping	No	Most recent entry year in wdidata20.quality
years.to.100	countries2	outcome	No	Years until life.quality metric reaches 100
expected.100.year	countries2	outcome	No	Expected year life.quality metric reaches 100

Appendix B

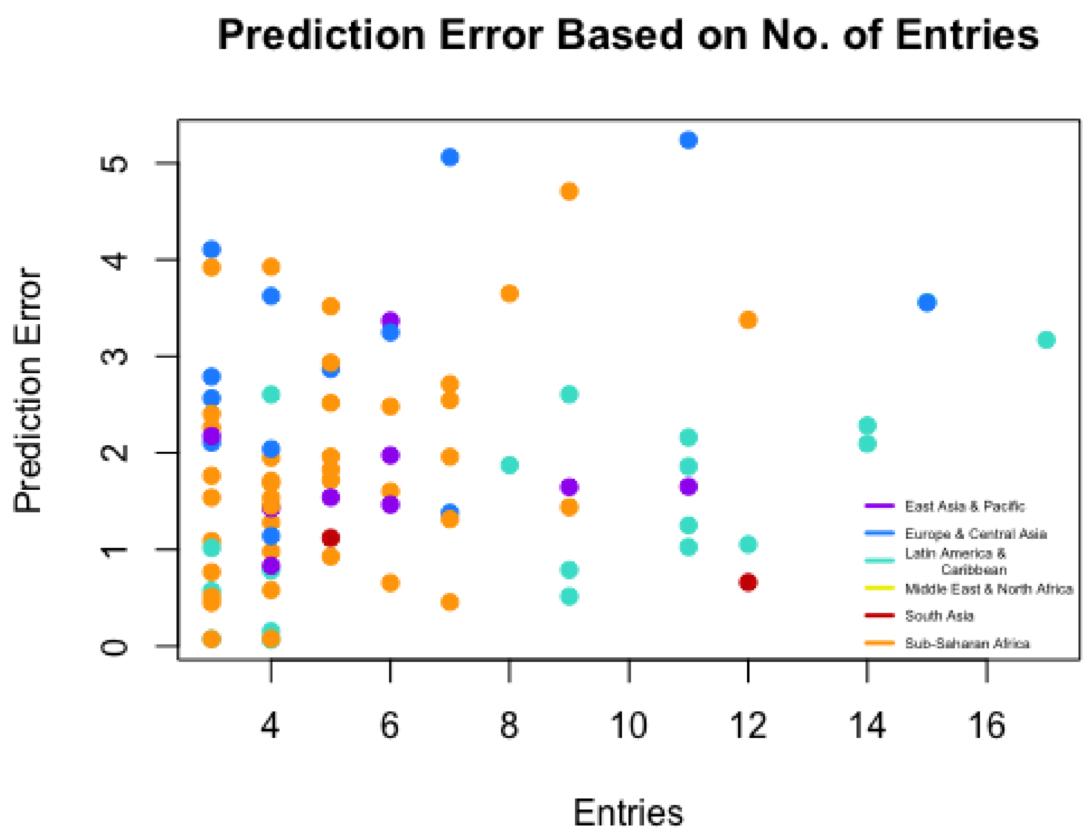


False zeroes significantly drag prediction errors closer to zero per region, falsely suggesting a more accurate model.



The removal of false zeroes allows for a more accurate prediction error which can better assess the viability of the model.

Appendix C



Observing the graph, no correlation seems to exist between the number of entries a country has and their prediction error. This is further supported by the low correlation of the two variables, which is roughly 0.257. No correlation seems to exist on a regional basis.

Works Cited

Dynan, Karen, and Louise Scheiner. "GDP as a Measure of Economic Well-Being." *Hutchins Center at Brookings,* Center on Fiscal and Monetary Policy at Brookings, Aug. 2018, www.brookings.edu/wp-content/uploads/2018/08/WP43-8.23.18.pdf.

Proto Eugenio, et al. "A Reassessment of the Relationship Between GDP and Life Satisfaction." PLoS ONE 8(11): e79358. 2013, https://doi.org/10.1371/journal.pone.0079358.

"World Bank Data." World Bank, 2020. https://www.dropbox.com/scl/fo/f185czaubsejosyyfhl9e/h/WorldBank?dl=0&subfolder_nav_tracking=1.