Is GDP a Good Indicator of Quality of Life?

Final Project Group 82

Charles Gutcho, Minh Nguyen, Aimee Moreno
Chloe Suarez, Oscar Lam, Walton Bullard

Introduction

Understanding quality of life is important in the backdrop of a globalized economic ecosystem rapidly experiencing growth and changes. It is a valid topic of research as we are concerned about people's share of the benefits from economic growth. The extent of this globalization-driven economic improvement (which we assume is comparable with the nation's real GDP value) should translate to improved quality of life, as we expect countries would invest some of its economic gains to improve their civilians' quality of life. GDP is the value indicating a country's aggregated value-added economic performance during a time period. The WHO defines "Quality of Life as an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns." Given the WHO's 2017 global development data on 103 countries with recorded GDP values, each group member would select an indicator/variable that reflects an aspect of life quality from the WHO database, and perform analyses to assess citizens' quality of life with respect to the nations' GDP values. Ultimately, our aim is to explore potential relationships between a nation's GDP value and its population's quality of life, as indirectly examined through related variables. Thus, our research question is: "Is GDP a Good Indicator of Quality of Life?".

Methodology and Limitations

We first decided to use data from the countries with a reported GDP value in the dataset, and then removed the countries that had missing values for each of our selected variables. We removed missing/NA data because we saw no constructive purpose of having missing or unmatched data (which becomes important as some of us plotted two variables together in search for a statistical relationship). We all put GDP as our explanatory variable, with the purpose of seeing whether economic performance can influence our selected variables' values. Finally, we used ggplot to create visual representations of our data as aides for our individual analyses.

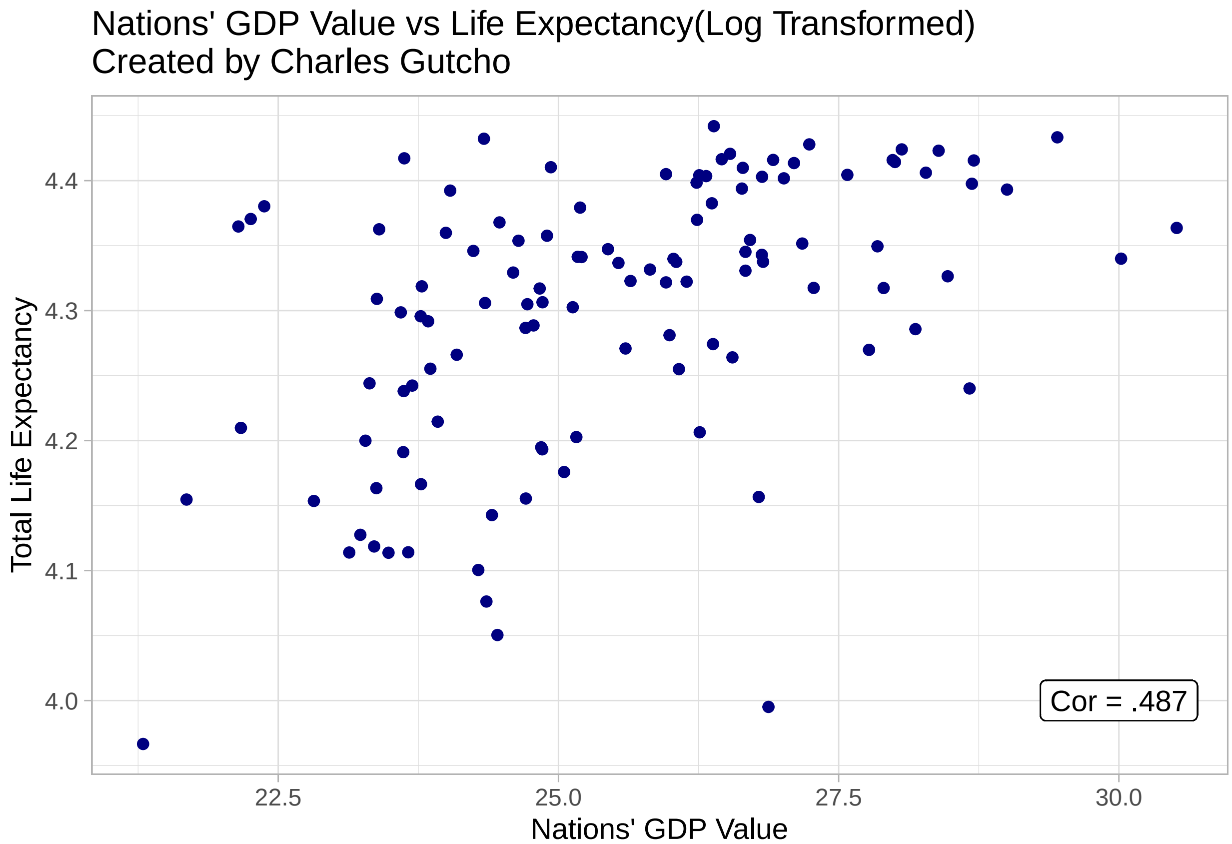
The main drawback of removing unavailable data is that this step reduced our pool of available information to use. For example, in Minh's analysis between nations' GDP values and number of telephone subscriptions, she should have had 121 pairs of data, but after filtering and omitting nonfunctional values, she only has 103 pairs of values left. Moreover, the available pairs of values would vary per group member - each person has a different number of countries with missing data for their chosen variable. This limitation would mean that all of our data points will not homogeneously represent every country (ie, Chloe doesn't have a pair of data for Korea but Aimee does). It could affect the validity of our conclusion, but we believe only to a small extent as we each did not remove too many countries with missing data.

Additionally, some members chose to systematically log our data (for variables that made sense) because it makes our statistical analysis on the linearity of our data more valid. "Logarithmically transforming variables in a regression model is a very common way to handle situations when it seems like there is a non-linear relationship between the independent and dependent variables. Using the logarithm of one or more variables instead of the un-logged form makes the effective relationship non-linear, while still preserving the linear model." Thus, log transformation allows us to compare our two variables in a more linear manner, especially when some of our variables do not illustrate a clear linear relationship between each other. These aspects are especially important as all of us are analyzing very large GDP raw data values. Log transformation makes sense for our analysis, for example between GDP and telephone subscriptions, the data are bunched up at the lower values with a few noticeable outliers at larger intervals.

However, mathematically, we realized that we could only do log transformations for values larger than 0. For example, this became an issue for Oscar because he has negative inflow values (negative inflows happen in real life but unfortunately, he could not include these values for his data analysis). Thus, some of our inputted data for log transformation was even more reduced, and insufficient data can affect the reliability of our conclusions.

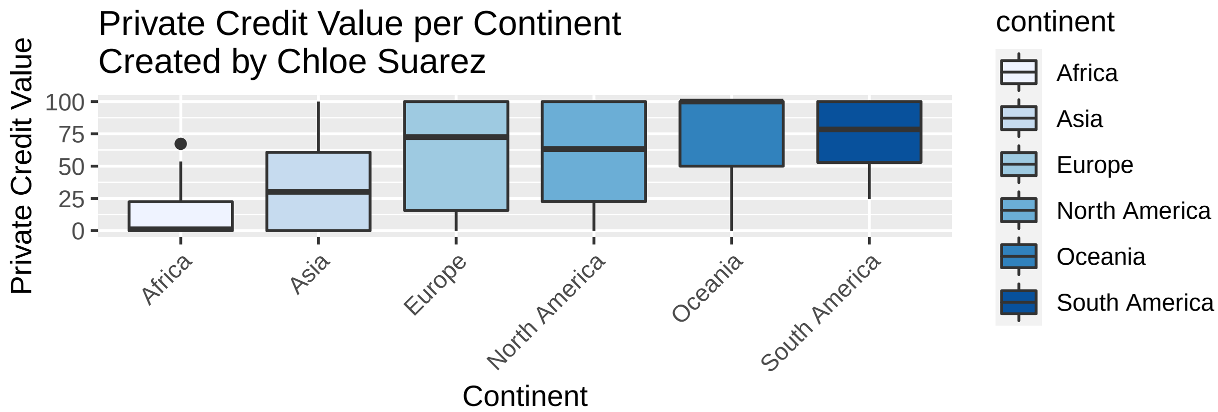
Exploratory Data Analysis/ Description of Data

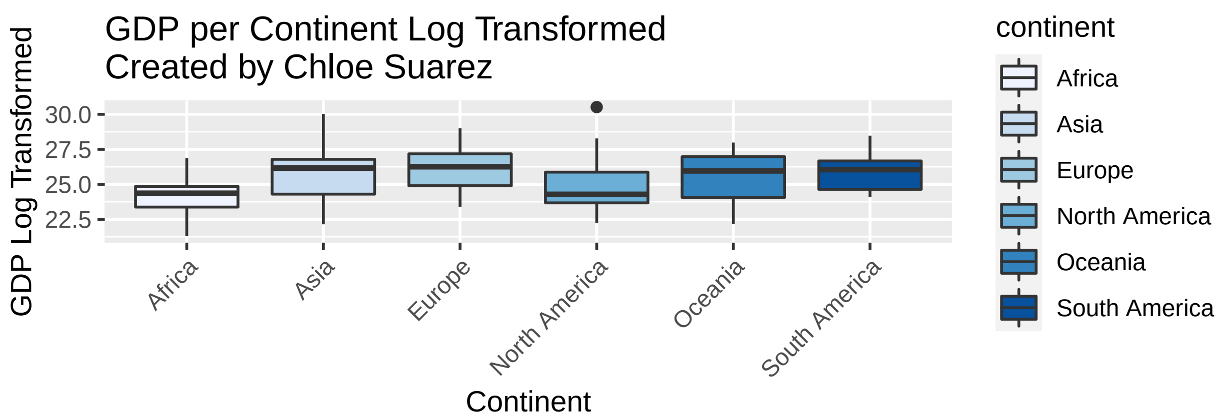
We chose to systematically log our data (for variables that made sense) because it makes our statistical analysis on the linearity of our data more valid. "Logarithmically transforming variables in a regression model is a very common way to handle situations when it seems like there is a non-linear relationship between the independent and dependent variables. Using the logarithm of one or more variables instead of the un-logged form makes the effective relationship non-linear, while still preserving the linear model." Thus, log transformation allows us to compare our two variables in a more linear manner, especially when some of our variables do not illustrate a clear linear relationship between each other. These aspects are especially important as all of us are analyzing very large GDP raw data values. Log transformation makes sense for our analysis, for example between GDP and telephone subscriptions, the data are bunched up at the lower values with a few noticeable outliers at larger intervals.



GDP vs Life Expectancy

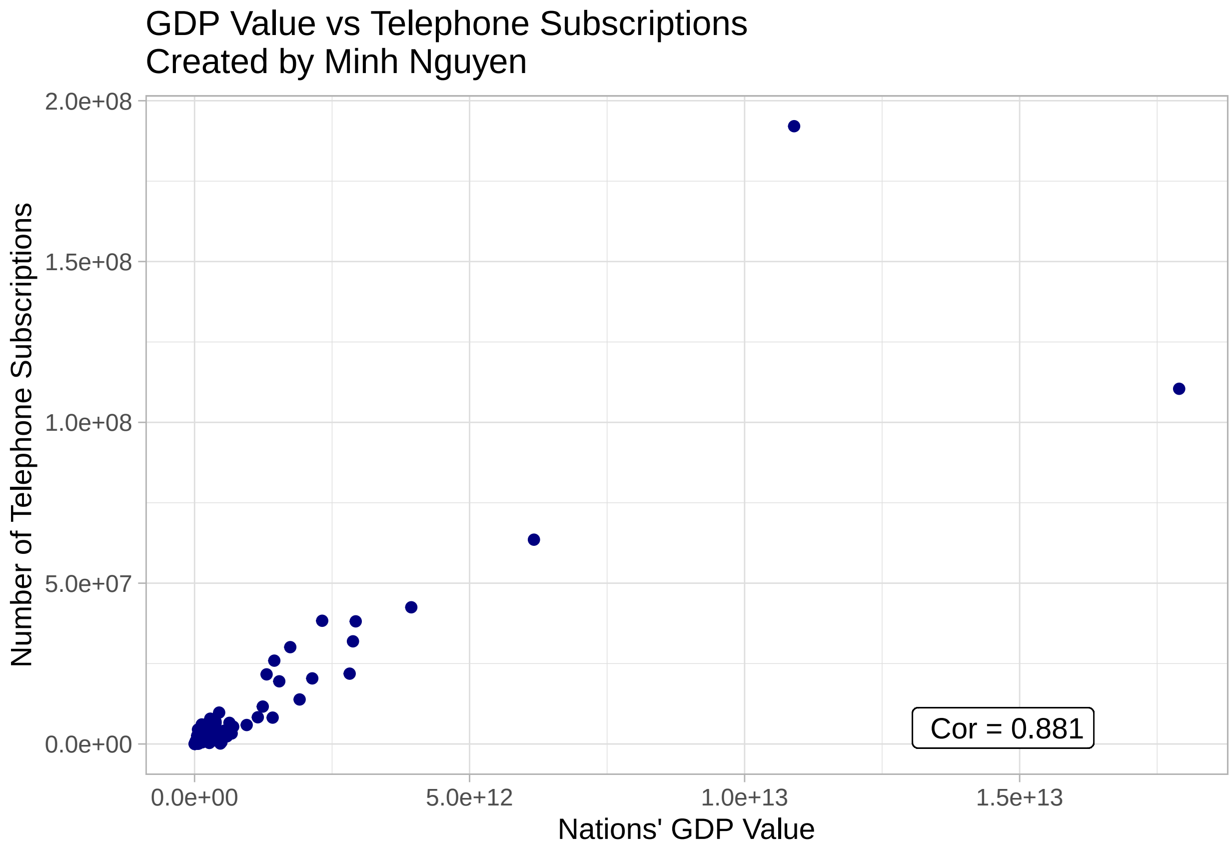
When it comes to determining how to measure quality of life, I thought the life expectancy of a country's citizens would be a good indicator of such a concept. In order to do this, I used the GDP variable as well as the life expectancy total variable, which encompasses the life expectancy of all citizens of a country regardless of gender, to create a scatter plot. The first scatter plot I created put GDP on the x-axis and life expectancy on the y-axis, but this didn't really work at first as a result of GDP being measured in billions of dollars and life expectancy only having a max of about 85 years old. This resulted in the scales of both the x and y axes being completely different and a graph that was very difficult to understand. To solve this problem, I logged both variables and plotted them against each other again and this logged graph was much easier to understand and make sense of. The logged graph illustrated a much clearer picture about the relationship between life expectancy and GDP which seemed much closer to being linear. When running the cor() function on these variables, I get an r-value of .488 which tells me that there is a significant relationship between the two variables. With this being said, we can make a general assumption stating that GDP in some cases is a good indicator of life expectancy within a country.





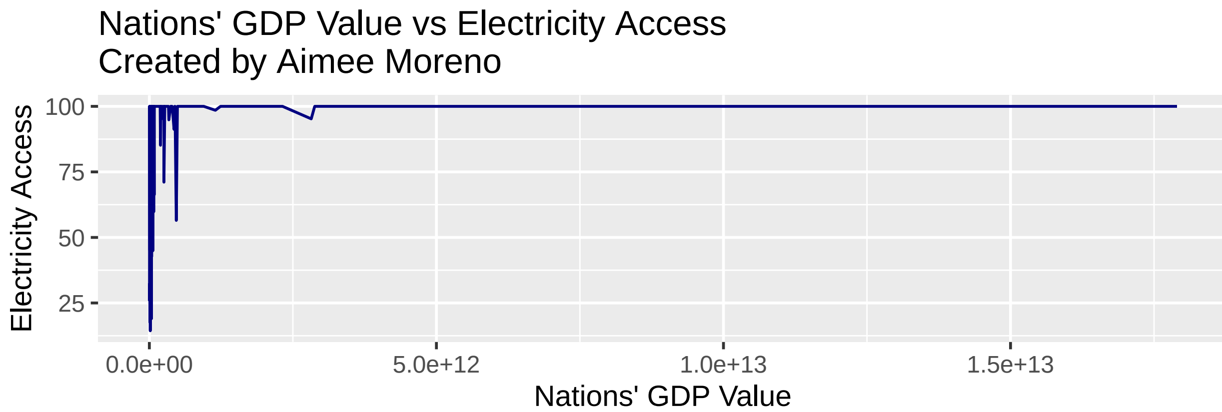
GDP vs Private Credit

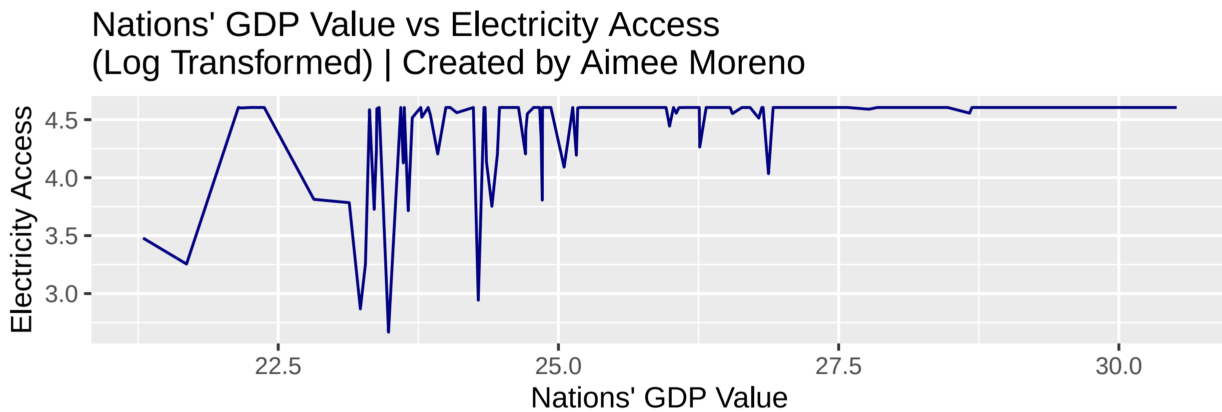
Prior to answering the overarching question, “Is GDP a good indicator of quality of life?”, we thought that it would be sensible to also examine the variable private credit. As a percentage of the adult population, private credit is defined as the number of people or businesses listed by a private credit agency with current information on repayment history, unpaid debts, or credit outstanding. One can imagine that if the majority of a population has a poor repayment history, for example, it may be exceedingly more difficult to borrow from the government, pay off pre existing debts in the short term, etc. The general speculation is that when the bulk of a population has a low private credit percentage proportionally, the overall quality of life may be lower on average than a country with a much higher private credit value and, on top of that, will have a less prosperous economy (measured by GDP). With the private credit values being percentages and the GDP values being exponential, it was ultimately most effective to create a boxplot per continent of both private credit and GDP log-transformed.



GDP vs Telephone Subcriptiopns by Minh Nguyen

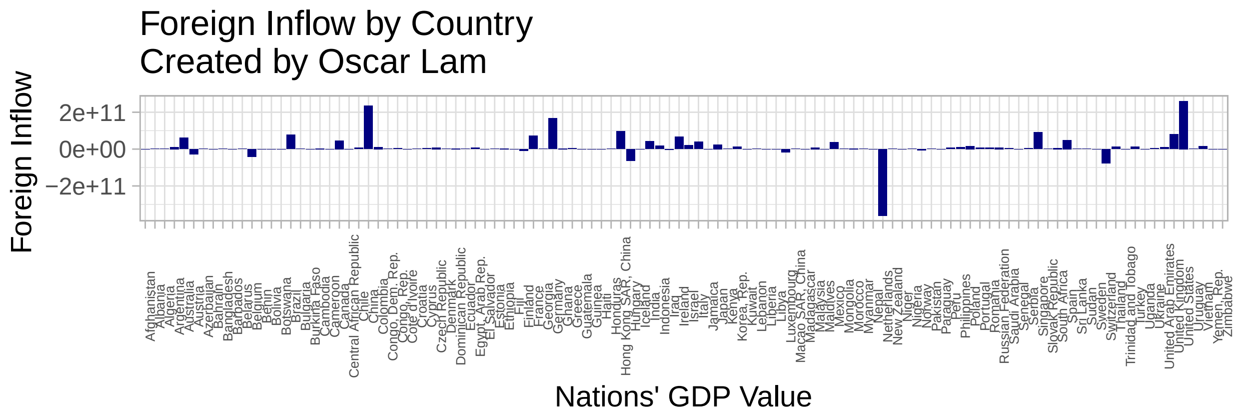
The telephone data records every available country's number of "active number of analogue fixed telephone lines" and any registered telephone communication subscriptions that use nationalized telecommunication services, including "fixed public payphones". Civilians' wide access to telecommunication services is a promising indicator of a nation's economic performance because it implies a working telecommunication infrastructure. Thus, a nation's GDP value can be indicative of its economic development and also its national infrastructure to accommodate telecommunication networks. A regression analysis is performed to examine the (statistical) relationship strength between countries' GDP values and their number of telephone subscriptions. The response value (as expressed on my y-axis) is each available country's number of telephone subscriptions, and it is regressed on the explanatory value of the country's corresponding GDP value (on the x-axis). Upon plotting the points into a scatter plot, I saw that the two variables show a promising trend between each other. The correlation coefficient is 0.88, reflecting a significant and positive relationship between the examined variables; thus, a country's GDP value could be a reliable predictor of the nation's number of telecommunication subscriptions. According to macroeconomic empirical research/studies, telecommunication penetration and implementation has a dual effect on and from economic growth: its quality requires proper economic and technical investment, and in turn, it improves labor efficiency and business operations to drive economic growth. These aspects are related to the nation's economic performance and reflected by its GDP value. Landline establishments require functional national infrastructure - a long-term and expensive government investment. Nations with a high GDP can afford this public infrastructure and also benefit from it. An enhanced medium of communication and information sharing corresponds to increases in cellular/telecommunication subscription rates and is driven (or also positively impacts) GDP. The World Bank data and the analysis I've done align and corroborate the macroeconomic research about the two variables discussed above.

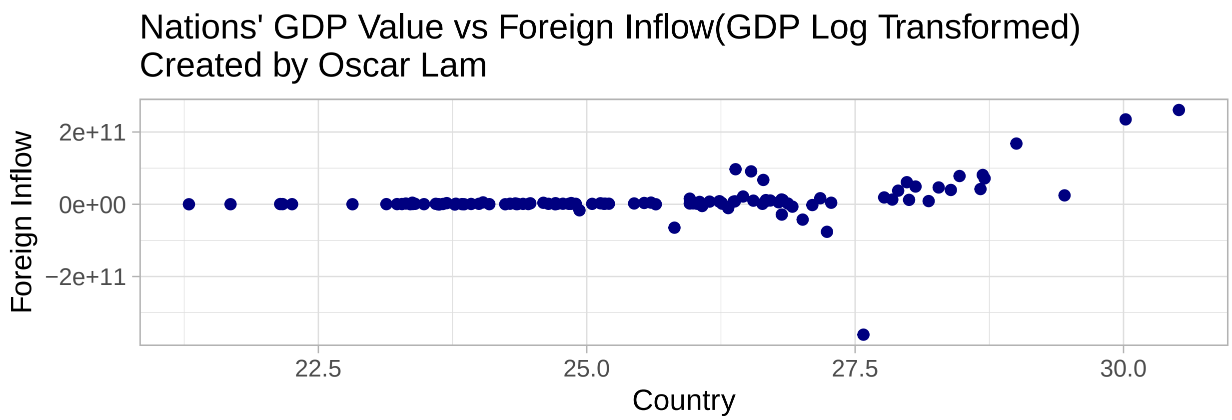




GDP vs Electricity Access by Aimee Moreno

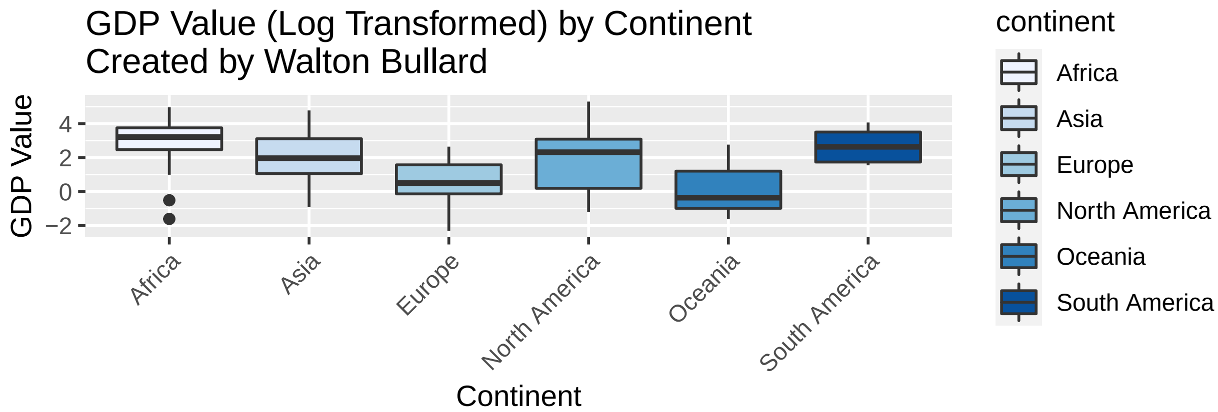
The line graphs compare the nations' GDP values (x-axis) to each of their electricity access rates, with one of the diagrams having its values log-transformed. By looking at the graphs together, one can see that the one to the left has its data much more grouped up than the other. There is a vertical line at the beginning of the diagram indicating nations with the lowest GDPs can obtain all rates of electricity access. However, low access remains within countries with minimal GDP and does not reappear with higher GDP levels. Complete electricity access becomes significantly constant as GDP increases,as does with the log-transformed graph (right). It appears to have more variance in electricity access and GDP, making our data clearer. Because access now tends to fluctuate throughout nations' gross domestic product, one concludes that GDP does not determine a population's electricity access.

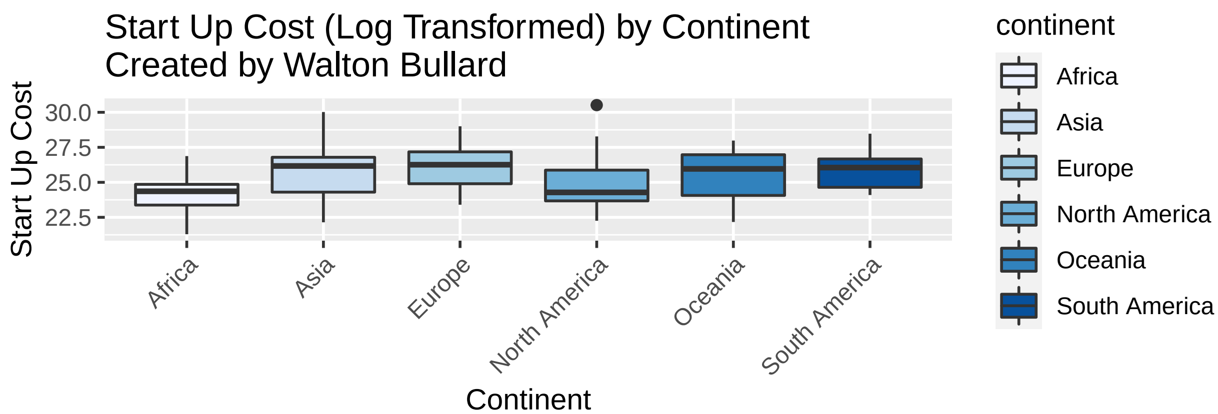




GDP vs Foreign Inflow by Oscar Lam

Foreign inflow, or foreign direct investments refers to the net inflows of foreign direct equity investments in a country's economy. FDI is a huge factor in a country's economic growth, which hence leads to an increase of overall quality of life. When we look at the graph representing foreign inflow by each country, we see that there is an uneven distribution with a few outliers. After applying a log transformation to the GDP data, we were able to see a scatterplot with rising trend as log transformed GDP grew larger. This shows that countries with a relatively small GDP has smaller foreign inflows, while countries with a big GDP with have a significantly larger foreign inflows. The correlation coefficient computed with the GDP and their foreign inflows is 0.627, which shows a low positive correlation between the two. The correlation is weak because there are several outliers such as Netherlands with has a foreign inflow of -361000000000 (Lowest foreign inflow).We are unable to fit a regression line to the graph since the regression is computed with non-log transformed data and after log transforming the data, some of the values turned into N/A (countries with negative foreign inflow).

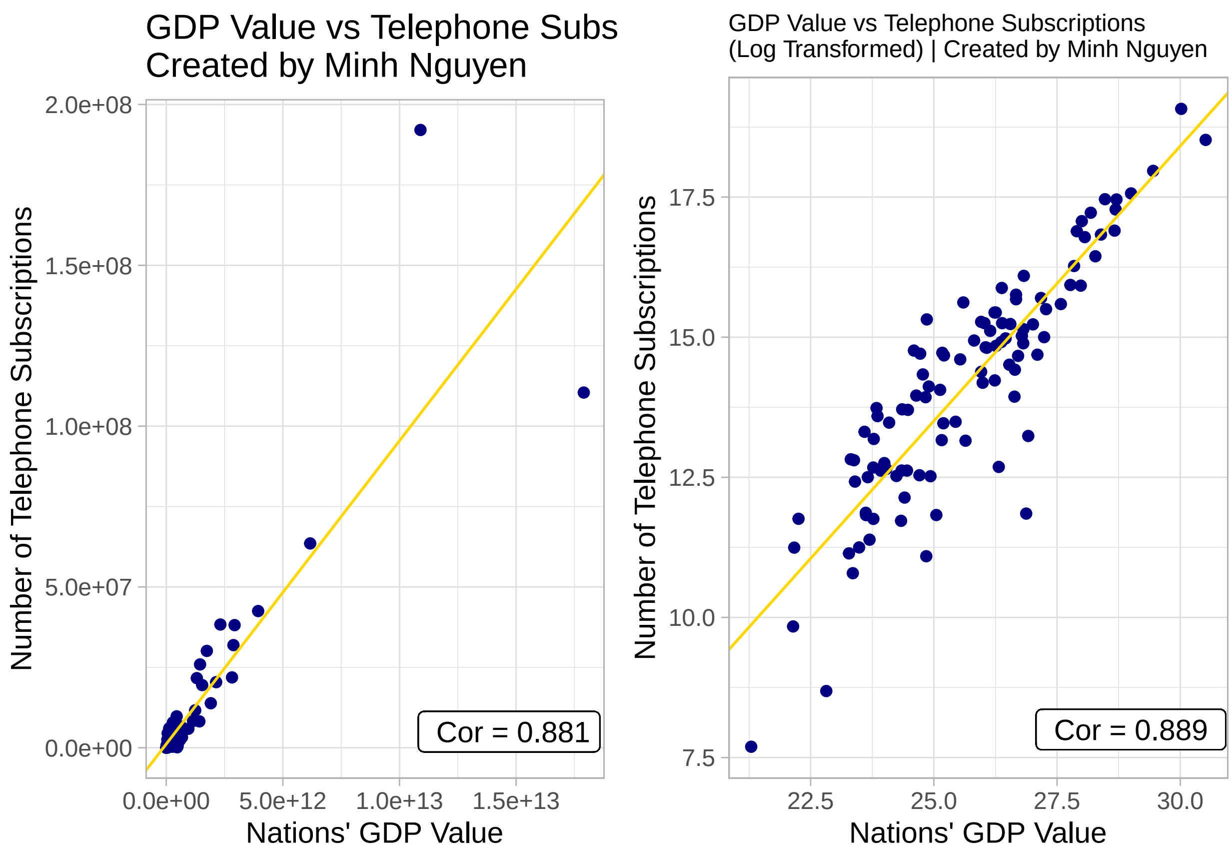




GDP vs Startup Cost by Walton Bullard

As the Gross Domestic Product (GDP) is dependent on the amount of finished goods that a country is able to produce within a specific time period, it becomes dependent on the number of companies that are then able to provide the process of turning raw materials into finished goods. With that in mind, the ability to start companies, including those that produce the finished goods, becomes a variable that should be taken into account when it comes down to distinguishing what factors within a country allows for it to have a higher GDP, especially considering the fact that the lower the start-up cost, then the more start-ups can be created and therefore more finished goods can be produced. Therefore, it can be predicted that as a country has a lower start-up cost rate (i.e. more startups can be created due to having a lower cost), the higher their GDP will be. A regression model is performed in order to find the relation between the GDP and the start-up cost of a country as well as the strength of the relationship between them. From this, the correlation coefficient is calculated to be determined as r = -0.139. With a correlation coefficient of -0.139, it can be concluded that there is a weak negative relationship between the start-up cost and the GDP, meaning that to some extent, as the start-up cost decreases, the GDP of a country will increase. After performing a log transformation, the correlation coefficient is -0.360, once again supporting the fact that the relationship between the start-up cost of a country and the GDP is a weak negative relationship. This can also be seen in the graph above where continents with lower start-up cost on average would have a relatively higher GDP than countries with a higher start-up cost. From this, it can be concluded that though the start-up cost within a country does not have a strong impact on GDP of a country, it does have an effect on the GDP through the fact that it is common for continents with lower start-up cost to have a higher GDP.

Data Analysis




```
##
## Call:
## lm(formula = log_total_telephone ~ log_gdpM, data = info)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4903 -0.4892  0.0959  0.6339  1.9517
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11.0201     1.2801  -8.609    9e-14 ***
## log_gdpM      0.9810     0.0497  19.738   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9357 on 103 degrees of freedom
## Multiple R-squared:  0.7909, Adjusted R-squared:  0.7889
## F-statistic: 389.6 on 1 and 103 DF,  p-value: < 2.2e-16

## [1] "Residual Standard Error using sd() value from r"

## [1] 0.9311852

## [1] "Residual Standard Error using rms to calculate sd value"
```

```
## [1] 0.9267404
```

As taught in class, the Residual Standard Error aims to measure the quality of our linear regression fit; because this regression is a best-fit linear model, it inevitably contains a degree of error. This error needs to be accounted as it shows the degree of failure in perfectly predicting the response variable (nations' log-transformed telephone subscription values) based on the predictor (nations' log-transformed GDP values), we are interested in looking at the deviation between the actual log-transformed telephone values from the true regression line's value. From what we researched, the "degree of freedom" reflects "the number of data points that went into the estimation of the parameters used after taking into account these parameters." In this analysis section, we have XX data points and two parameters (intercept and slope).

The Residual Standard Error (as given by R Studio) is 0.9357 and as we calculated manually using the shortcut for RMSE for regression, we got a close approximation of 0.9267; the two values are very similar despite R's method around "losing degrees of freedom" required a division by (n-2), so we can say that the RMS error of regression is about 0.93 log-transformed telephone subscription values. The Residual Standard Error values inform us the vertical spread about our constructed regression line; the error is the vertical distance between a point and its predicted value based on the regression line equation. Hence, this method informs us that the logged telephone numbers are about 0.93 values above/below the regression line, and we can say that there is accuracy in the regression prediction by 6.56%.

Conclusion

In an attempt to answer the question, "Is GDP a good indicator of quality of life?" we analyzed the following variables: life expectancy, private credit, telephone subscriptions, electricity access, foreign inflow, and start-up cost. While the definition of "quality of life" differs from person to person, it becomes much easier to quantify and illustrate with the aid of the data provided. Moreover, when paired with a variable such as GDP, we can gauge the overall success of a country's economy while cross-analyzing it with the specified variables. Through each variables' correlation with GDP, we could analyze the strength of their relationships. Telephone subscriptions had the highest correlation at 0.88, indicating a strong relation. To further explore such a link, the p-value was calculated, being approximately equal to 0 (R-studio says it is 2.2e-16). Given the value, we can say a relationship between the predictor and the response is present. The RSE estimates the standard deviation of the response from the population regression line. The RSE is 0.93 telephone subscription units, while the mean value for the response is 14.18139, indicating a percentage error of roughly 6.57%. With all this being said, this project has taught us that GDP can, in some cases, be a promising indicator for determining a nation's quality of life.

References

"Analysis of the Impact of Information Communication Technology on Economic Growth: Empirical Evidence from Asian Countries." Journal of Asian Business and Economic Studies, https://www.emerald.com/insight/content/doi/10.1108/JABES-07-2020-0082/full/html.

Digital Commons - Bucknell University. https://digitalcommons.bucknell.edu/cgi/viewcontent.cgi?article=1003&context=honors_theses.

Linear Regression Models with Logarithmic Transformations. https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf.

Quick Guide: Interpreting Simple Linear Model Output in R, https://feliperego.github.io/blog/2015/10/23/Interpreting-Model-Output-In-R.