# Analysis of Using GDP as the Indicator of a Country's Sustainability Practices

Final Project Group 39: Ayush Pradip, Jay Chiang, Aaliyah Pondor, Christian Whang, Yuzhen Yu

## Introduction

Our world faces a number of complex issues which our society aims to address. After looking at the data as a group, we were drawn to the social, political, and economic implications of our findings and the discussion of development theory. Sustainable development is an issue which has become increasingly relevant over the past decade due to global trends in exploitation of resources and violations of civil and political rights. The Sustainable Development Goals (SDGs), was set by the United Nations as an international standard of what sustainable development should be ideally. For the purpose of this project, we have decided to look into goals 3,5,8,9 and 14 (see below).

Reference to Sustainable Development Goals (SDGs):

#3: Good Health and Well-Being

#5: Gender Equality

#8: Decent work and Economic Growth

#9: Industry Innovation and Infrastructure

#14: Life Below Water

As GDP is commonly used as a metric to assess a country's level of economic development, we decided to focus on how a country's GDP serves as an indication of it's sustainable practices from a social, technological and ecological perspective?

Our first investigative topic aims to establish whether there is a relationship between GDP and parliamentary positions held by women. This refers to goal 5 of the SDGs (Gender Equality). We were probed to explore this issue in greater depth due to the assumption that countries with greater economic development, indicated by GDP, also have more developed social and civil freedoms pertaining to women's rights and gender equality. As an extension to this question we wanted to look at individual countries and compare the strength of the correlation between GDP and percentage of parliamentary positions held by women to better gauge what factors might influence these results.

Our second investigative topic aims to determine whether there is a relationship between GDP and life expectancy of the total population. This pertains to the third SDG given by the UN, Good Health and Well-being. We assume that countries with higher economic development have the infrastructure and resources to have better healthcare systems, correlating to a higher life expectancy for the total population.

In our third investigative question, we wanted to determine the connection between how threatened fish are with respect to GDP. Due to the devastating effects of climate change, fish populations are rapidly depleting, thus presenting an environmental threat. This issue has been aggravated by the problem of overfishing in many countries. China is one of the world's leading economies and has the largest fishing industry too. Does economic growth have a correlation with the level of threat to fishing populations? Can we say that the path to economic growth might not always be sustainable?

In our fourth investigative question, we compare the GDP of a country to the number of active analogue fixed telephone lines. This falls under the United Nations SDG goal number nine, industry innovation and infrastructure. This statistical comparison gives us insight into how GDP and technological advancement are associated and whether an economically developed country is also technologically developed and has technology that is accessible.

Lastly, for our fifth and final investigative question, we compare the distribution of the world's GDP to the distribution of foreign investments. More specifically, we qualify foreign investments using the variables foreign inflow and outflow. This falls under number 8 of the SDG goal, which describes economic growth. We argue that high foreign inflow and outflow are characteristics of an economy with growth potential. Therefore, the statistical comparison between the distribution of GDP and the two factors may indicate a relationship between the two or at the very least, give insight to how the global economy functions and grows. While there are a number of cultural nuances that may make our understanding of these questions limited, we still find it important to continue asking these questions to better understand global trends and how to go about fixing these issues.

## Summary Statistics (GDP)

## Variable Description

GDP: Gross domestic product (GDP) in constant 2010 U.S. dollars

Fish threatened: Number of fish species classified by the IUCN as endangered, vulnerable, rare, indeterminate, out of danger, or insufficiently known

Women in parliament: Percentage of parliamentary seats in a single or lower chamber held by women

Foreign inflow: Net inflows of direct investment equity flows in the reporting economy including the sum of equity capital, reinvestment of earnings, and other capital

Foreign outflow: Net outflows of direct investment equity flows in the reporting economy including the sum of equity capital, reinvestment of earnings, and other capital
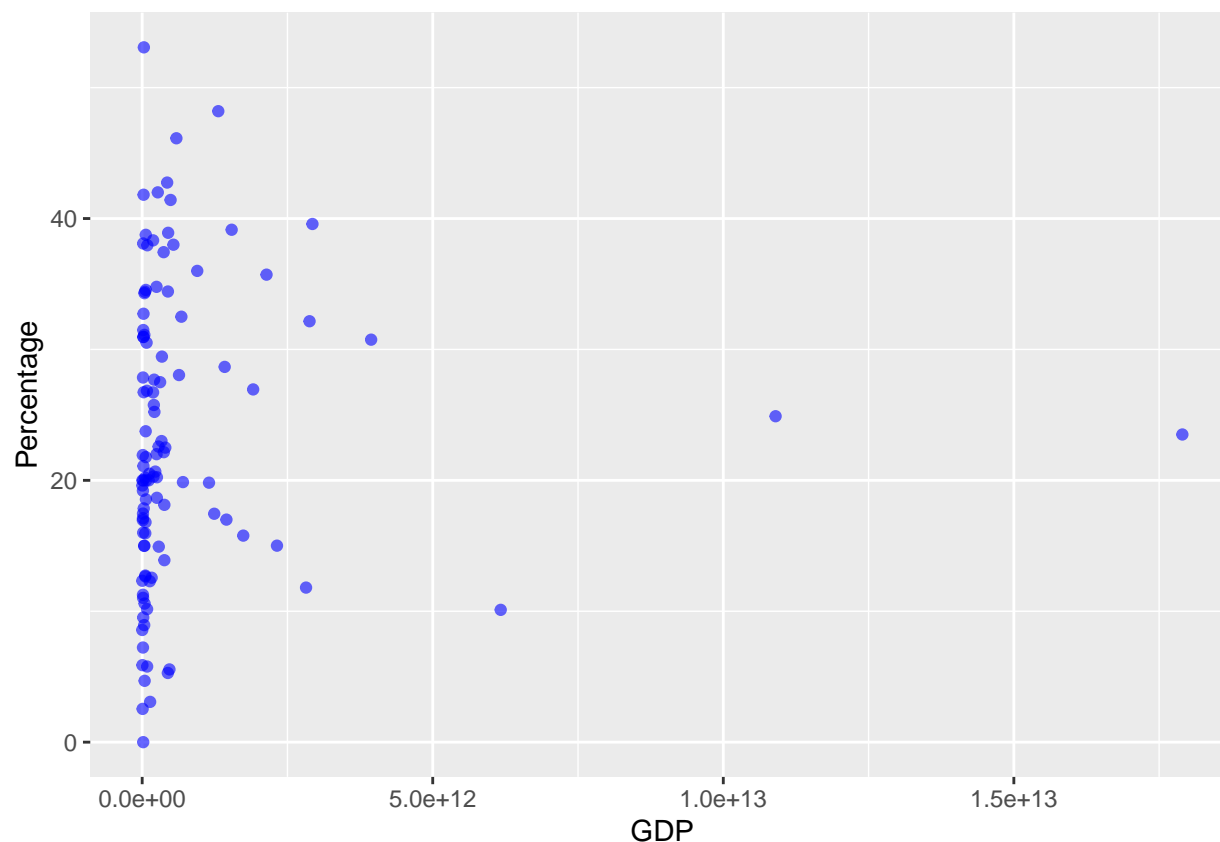
Telephone: Number of of active number of analogue fixed telephone lines, voice-over-IP (VoIP) subscriptions, fixed wireless local loop (WLL) subscriptions, ISDN voice-channel equivalents and fixed public payphones

Life expectancy total: Life expectancy at birth, total (years)

# Exploratory Data Analysis

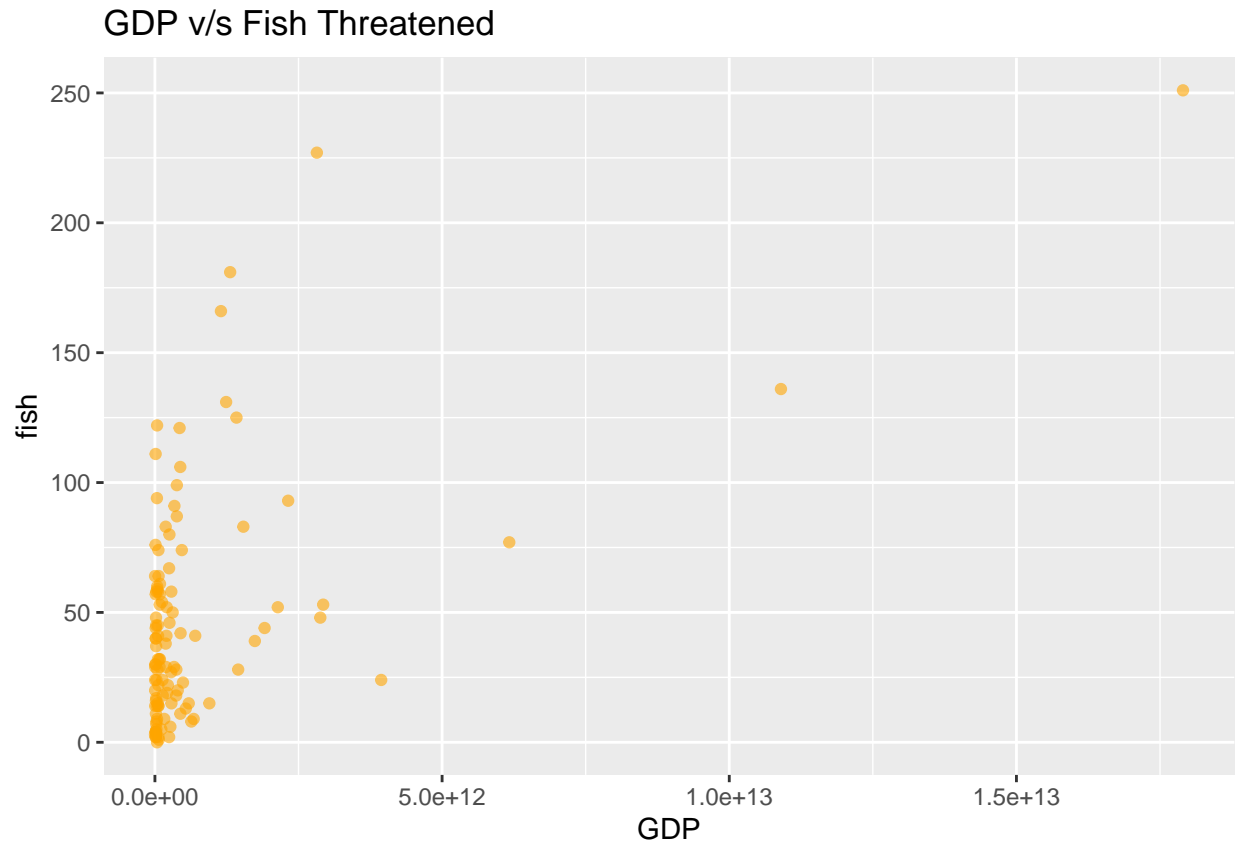#Christian Whang - GDP v Percentage of Parliamentary Positions Held by Women

```
## Warning: Removed 11 rows containing missing values (geom_point).
```



When the GDP of all of the countries is graphed in a scatter plot in relation to the ratio of parliamentary positions held by women, it appears that the scatter plot shows weak to no correlation between the two variables. While we made the assumption that more economically developed countries also have more civil and social liberties pertaining to gender equity, it appears that this is untrue as countries with high GDP show a similar range of percentages of women in parliament to countries with low GDPs. For example, a country with a GDP around $3*10^{12}$ can have less than 20 percent of parliamentary positions held by women, and a country with a much smaller GDP can have over 50 percent of parliamentary positions held by women. It is possible that this is due to a large number of cultural factors that make it harder for women to gain political seats in the parliament. This trend shows that a country's economic status does not serve as an indication of the percentage of parliamentary positions held by women.

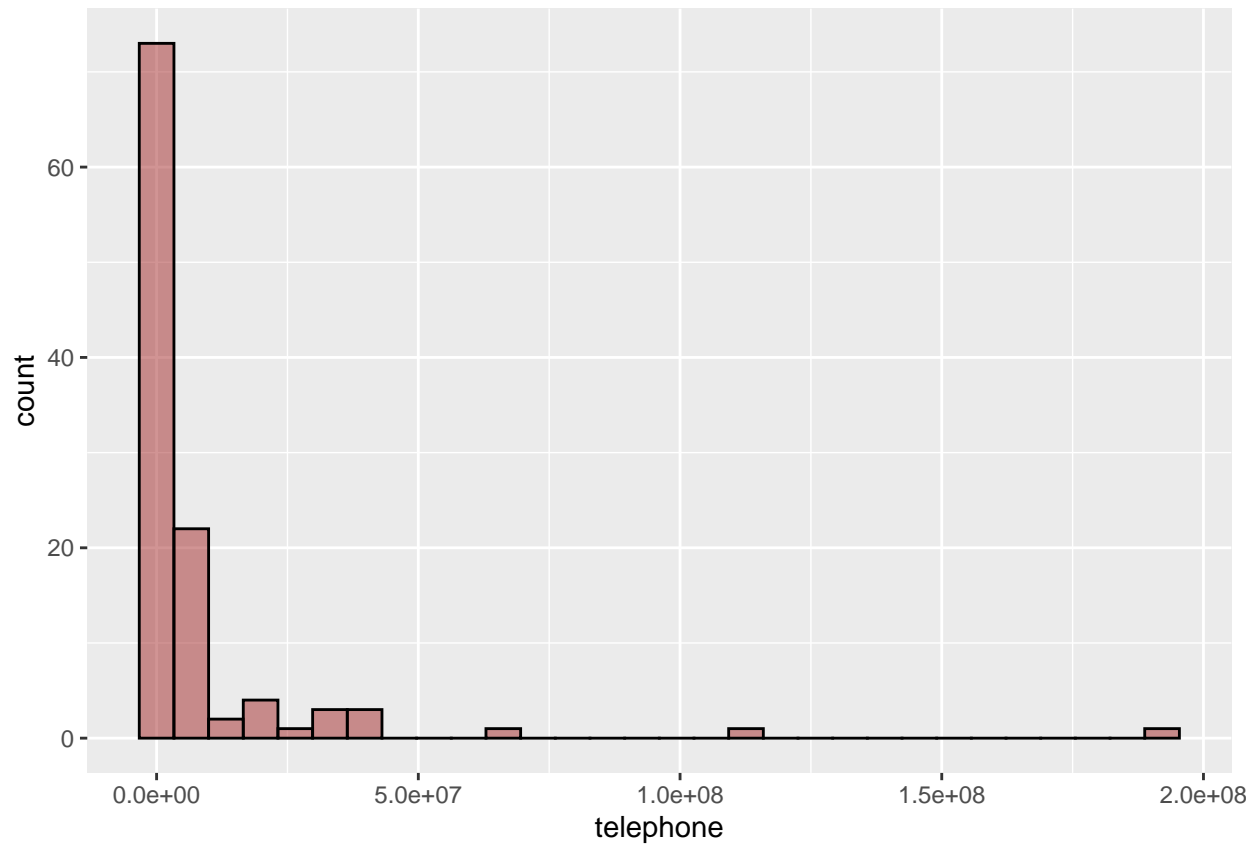#Aaliyah Pondor - GDP v Fish threatened

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
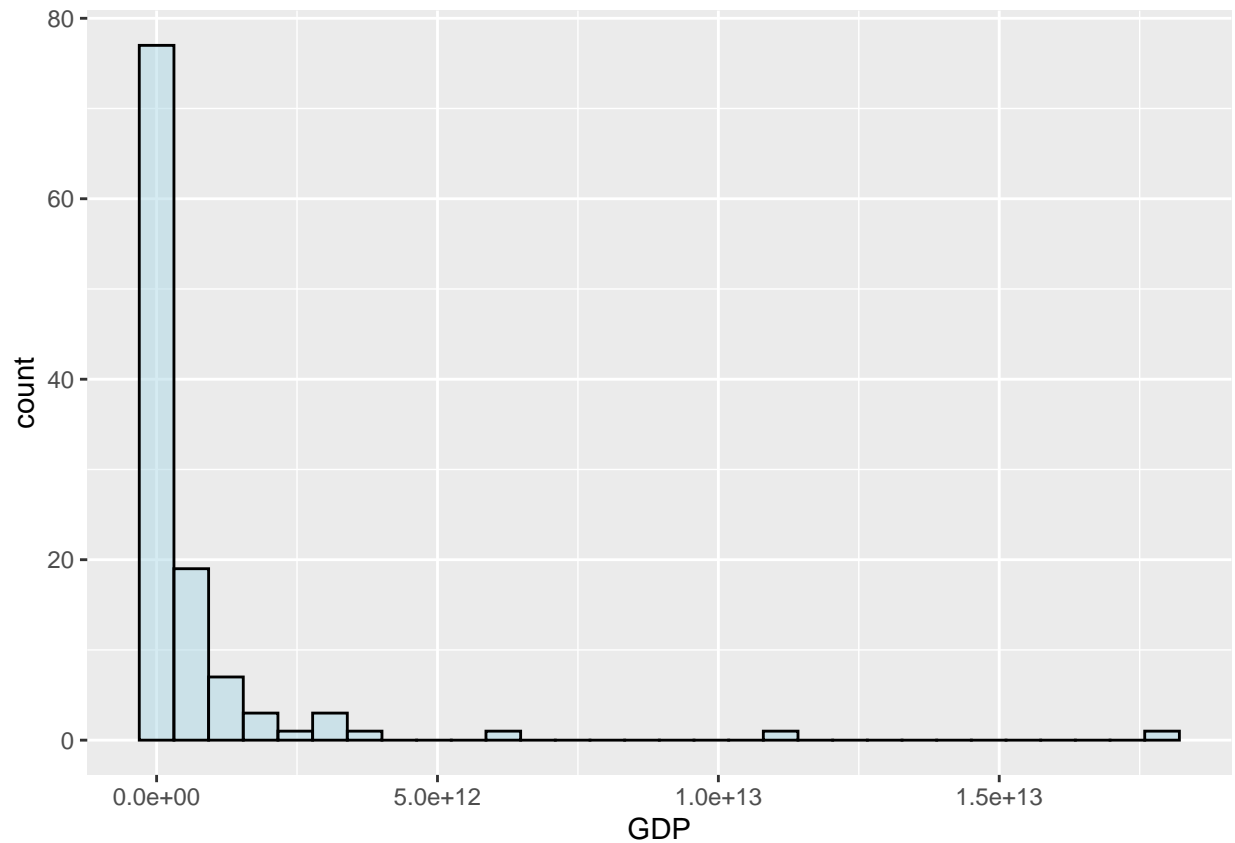
## GDP v/s Fish Threatened



We plotted a scatter plot where GDP is on the x-axis and fish threatened are on the y-axis. Given the large data set, we set a cut-off point for the countries based on their level of GDP. Although this is a limitation in our EDA, it was necessary to show the weak positive relationship between the amount of fish threatened and the level of GDP. There appears to be a cluster of data points on the left side of the scatter plot which indicates that countries with a lower GDP do also partake in overfishing. This is not a sustainable practice. Economic dependence on the fishing industry will negatively impact global environmental protection. We can see from the scatter plot that there are a few countries which appear to be outliers. Despite being economically advanced, they do not have a high level of fish threatened. This could be because the country is land-locked or because of environmental protection acts which help build a sustainable green economy. Possible confounding variables which could influence the correlation in this case include the level of economic dependence of fishing, whether the country is land-locked or not, as well as the type of fish species present in different parts of the world.

#Ayush Pradip - GDP v Telephone

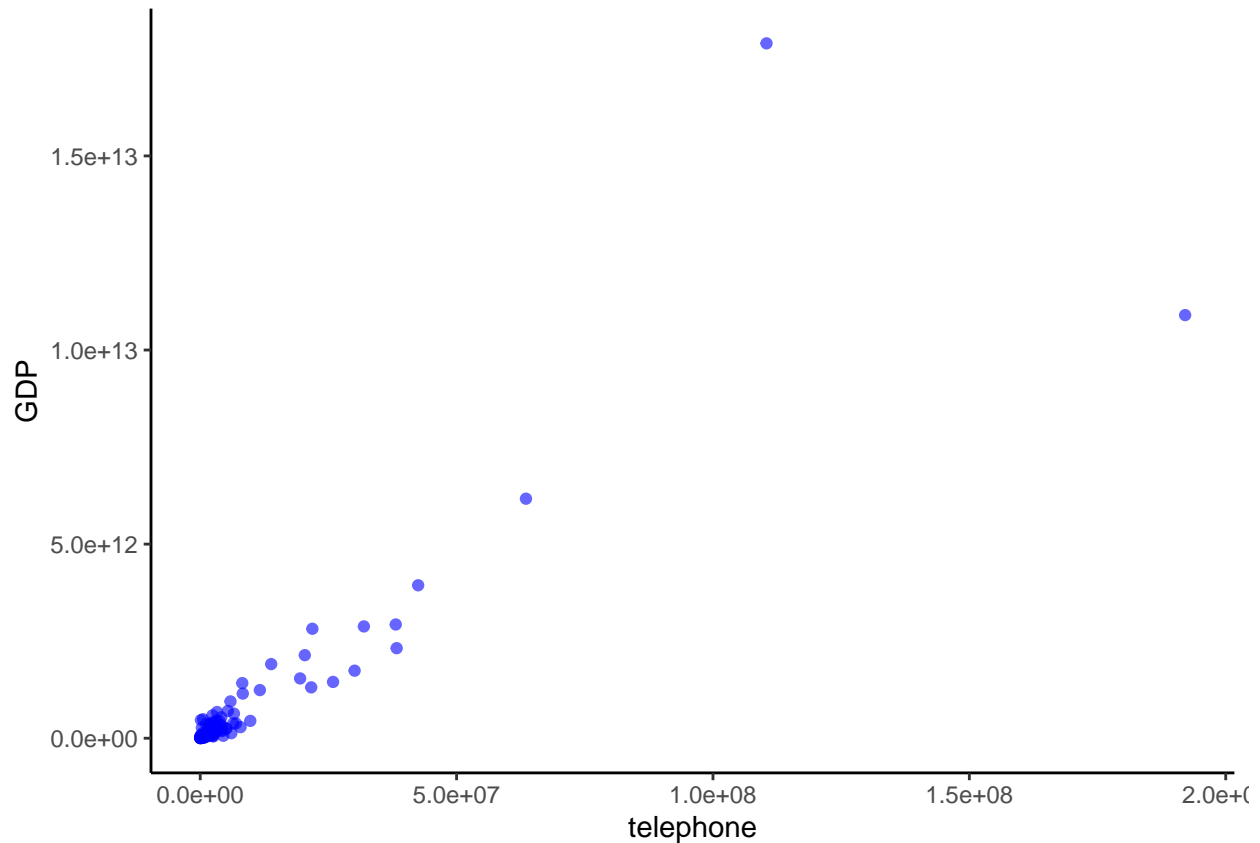## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

This part of the EDA aims to observe 2 different variables: number of telephones and GDP of countries. The histograms for the distributions of countries' GDP and number of telephones are displayed above. Both distributions are right-skewed, unimodal, and have a very similar shape. This led me to believe that as the number of telephones in a country increases, the GDP also increases.

```
## [1] 0.8822214
```

```
## Warning: Removed 13 rows containing missing values (geom_point).
```

I had to create another graph in order to confirm my belief of an association between the GDP and number of telephones in a country. The figures above is a scatter plot of all the countries with GDP as the y-axis and total number of telephones as the x-axis. Using the dataset as a whole, there appears to be a correlation between GDP and telephones. After calculating, we found the correlation to be 0.882, which shows a strongly positive correlation.
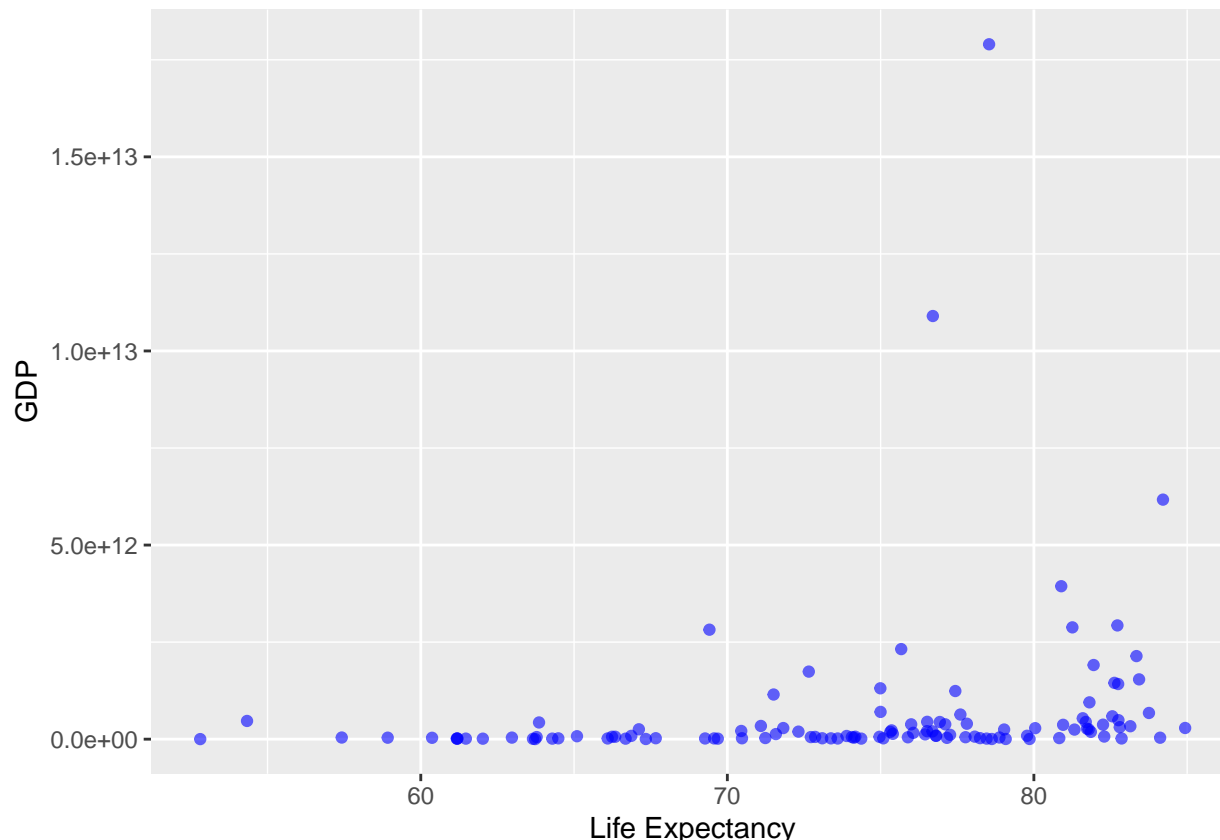
There are many possible reasons why there is a moderate correlation between telephones and GDP. Telephones are considered a normal good, so as income increases, demand for telephones and total GDP both increase. As income tends to increase, more households are able to afford telephones. Telephones also contribute to a better developed economic infrastructure. Telephones make businesses much more efficient because individuals don't need to travel back and forth for business operations & transactions. In developing countries, telephones are primarily used by high-traffic professional users. A higher usage of telephones leads to increases in productivity in businesses by saving time and money, which then leads to a growth in GDP. Additionally, the sales of telephones directly contribute to the GDP. In fact, the mobile industry contributed approximately 5% of the global GDP in 2020. The production and sale of a telephone is able to increase both the number of telephones available and the GDP.

Although the scatter plot shows a moderate linear relationship, a confounding factor that influences this correlation is the technological advancements in the software and hardware industries. These affect the features and quality of phones, thereby affecting the demand and total number of telephones sold. They also affect the consumption and business investments that are part of these industries, which affects the GDP as well. Apart from this, another confounding factor is americans' spending habits. If consumers change how much they're willing to spend, then this affects how likely they are to purchase telephones as well. The change in consumption also

directly contributes to a change in GDP. The population of a country is a very important confounding factor because a higher population tends to increase a country's GDP. It also results in a higher number of telephones because there are more people using telephones. Even after considering confounding factors, the scatter plot presented useful information and we were able to draw interesting findings from it.
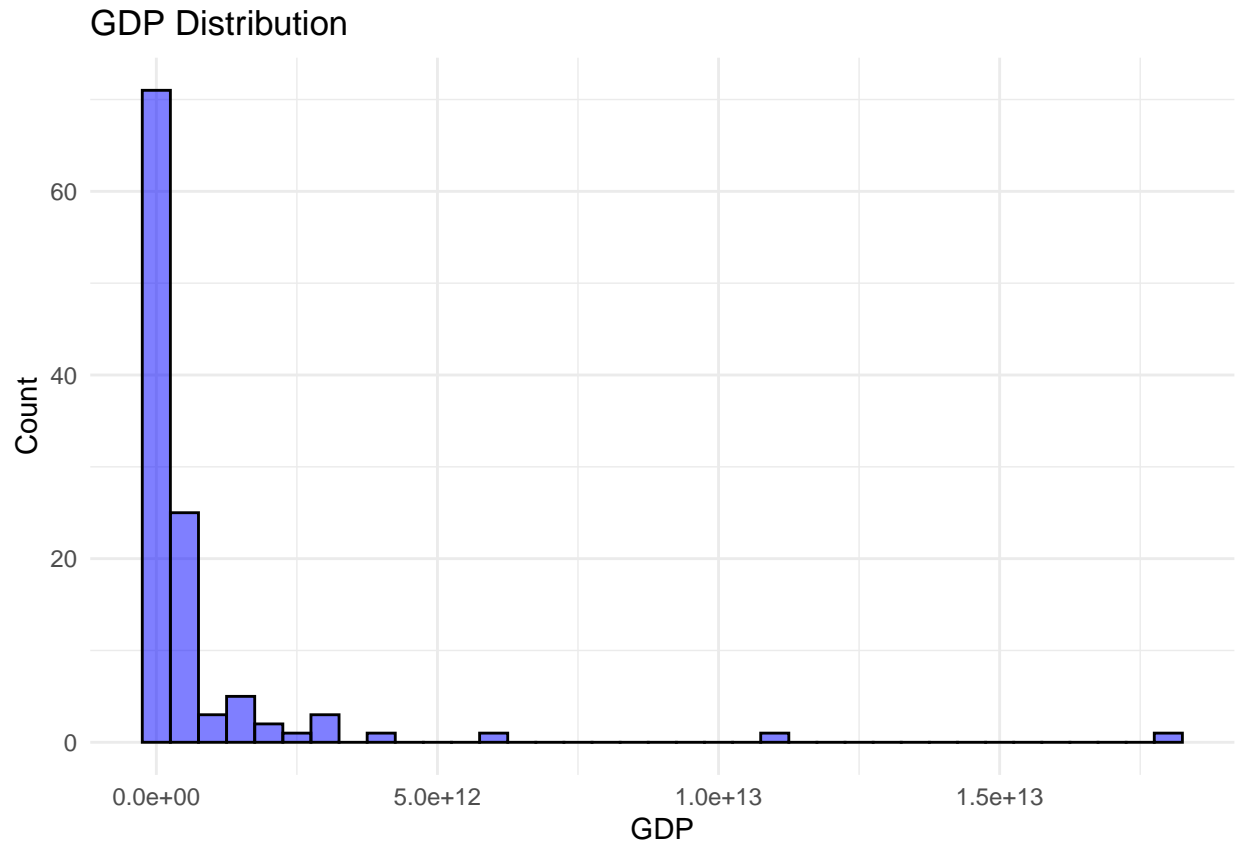
#Yuzhen Yu - GDP v Life expectancy total

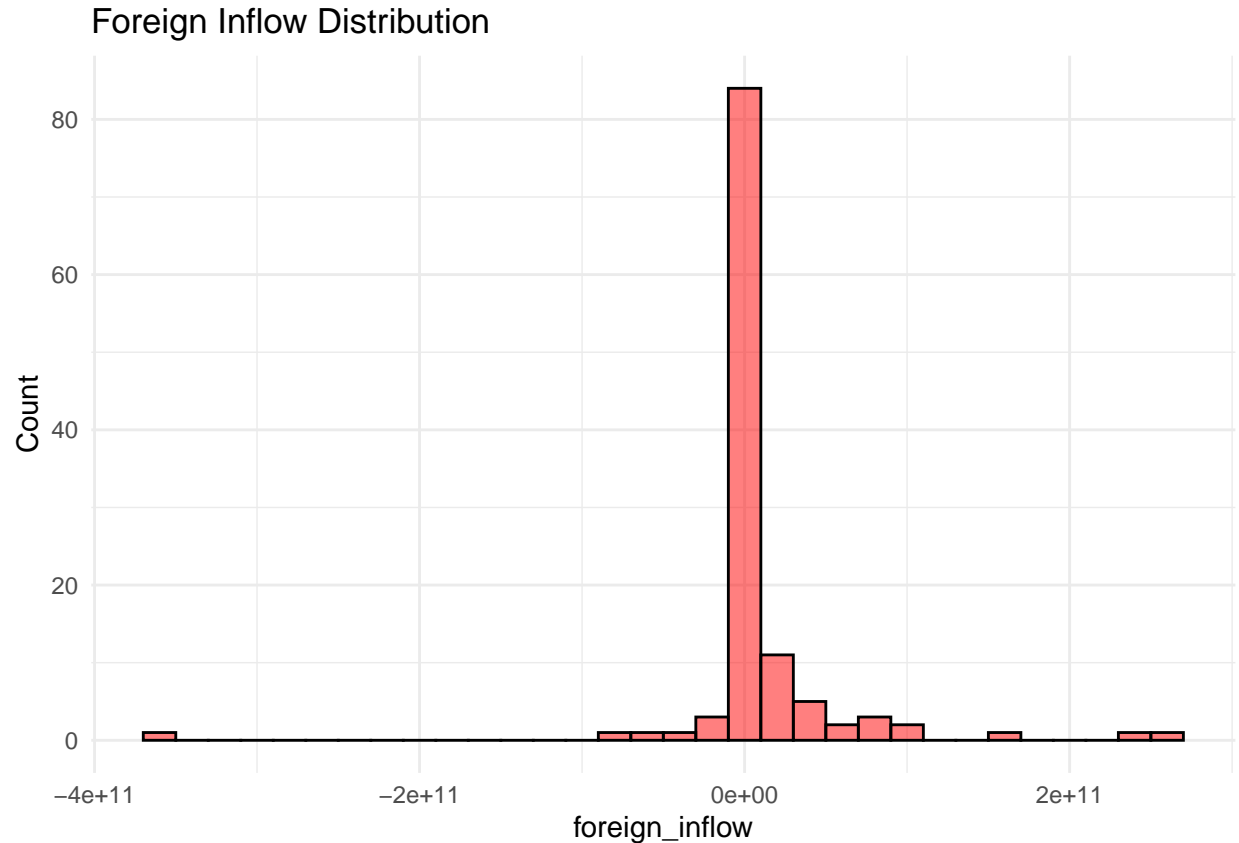## Warning: Removed 6 rows containing missing values (geom_point).



The country's total life expectancy is considered to be related to GDP in the traditional sense. It is generally believed that the higher the GDP of a country, the longer the life expectancy of its citizens. When we are studying the correlation between total life expectancy and GDP, we examine human's total life expectancy rates in relation to their country's GDP. We plotted the total life expectancy as an independent variable and the GDP in 2018 as a dependent variable. When all the countries are mixed together, we get the data points scattered randomly near the x-axis. According to the figure above, we can observe that the graph is almost a linear sector and does not fit in a linear regression model. The correlation between national total life expectancy and GDP is concentrated on the right side of the regression line, and the correlation coefficient is 0.1969699. Therefore, we can conclude that there is no linear correlation between these two variables. The total life expectancy of a citizen does not reflect the relationship with the country's GDP.

#Jay Chiang - GDP v Investments

## GDP Distribution



```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## 1.770e+09 2.663e+10 1.003e+11 6.987e+11 4.220e+11 1.790e+13        6
```
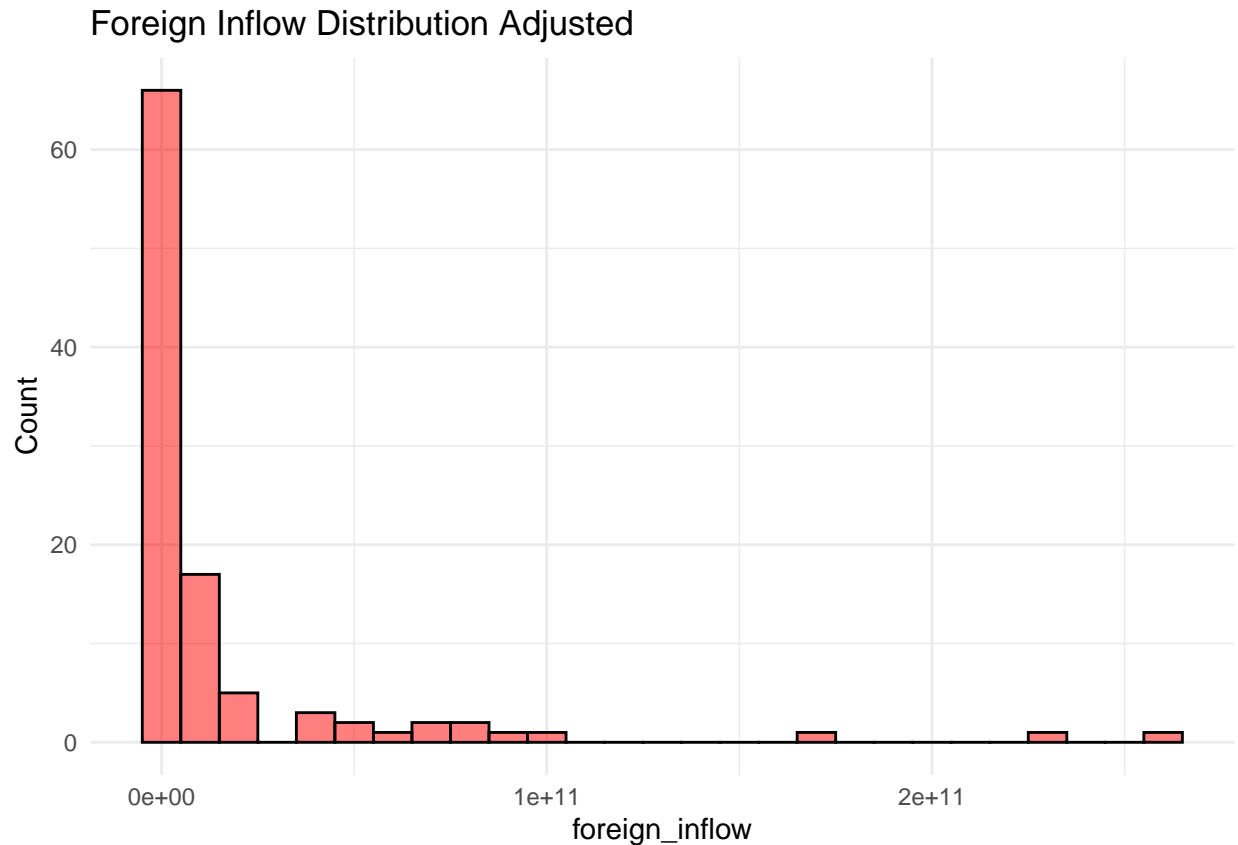
The figure titled "GDP Distribution" shows the distribution of the GDP of every country in the data set. The distribution is unimodal and has its minimum at 1.770e+09 USD, median at 1.003e+11 USD, and maximum at 1.790e+13. The mean GDP is higher than the 3rd quartile, which not only shows that the distribution is heavily right-skewed, but also that a few countries have GDPs that are significantly higher than the rest. Simply by looking at the histogram, roughly half of the countries are in the first bin, which means that they have a GDP below 5e+11 USD. On the other end of the spectrum, the countries with the highest GDPs are distributed scarcely and far away from the cluster. It gives insight to the world economy as it shows that generally speaking, most countries in the world have comparable GDPs and are clustered in the relatively low or moderate side. Certain countries with powerful economies and very high GDPs, such as the US and China, are outliers and pull the mean away from the median.

## Foreign Inflow Distribution



```
##       Min.    1st Qu.    Median      Mean    3rd Qu.      Max.      NA's
## -3.610e+11  3.528e+08  1.466e+09  9.875e+09  8.325e+09  2.610e+11         3
```

```
## [1] "number of countries with foreign inflow < 0:  14"
```

The figure above shows the distribution of the foreign inflow. Foreign inflow is defined as the "Net inflows of direct investment equity flows in the reporting economy." This statistic essentially measures how much investment is put into the economy by foreign entities. We expect the distribution of foreign inflow to be similar to that of GDP. This is because we believe that most countries would have a relatively low or moderate amount of foreign investment while a few "superstars" would have a lot of foreign inflow. Looking at the figure titled "Foreign Outflow Distribution", which has a minimum of -3.610e+11 USD, median of 1.466e+9 USD, and a maximum of 2.610e+11 USD. Its mean, 9.875e+9 USD, is higher than the 3rd quartile, which means that similar to GDP, the distribution of foreign inflow is also right-skewed and unimodal. However, they are different because the latter has negative values. A negative foreign inflow means that the value of disinvestments exceeds that of investments. We believe that countries under crisis (e.g. political instability, war, etc.) would have negative foreign inflows since the risk would push investors away. There are 14 countries that fit this profile, and if we can adjust the histogram by excluding them.

## Foreign Inflow Distribution Adjusted



After adjusting the histogram, it is quite evident that the shape of the distributions of foreign inflow and GDP are very similar. Both are unimodal, strongly right-skewed, and have a long right tail with a few outliers. This validates our expectations. On a global scale, the strength of economies (qualified by GDP) and the amount of foreign direct investment (qualified by foriegn inflow) are distributed similarly. While it is tempting to conclude that weaker economies tend to receive less foreign investment while stronger economies receive more, we do not know that for sure. The reason is due to the nature of histograms. For example, although we see a distant outlier at the maximum for both graphs, we don't know if the two reference the same country. Therefore, we can't say that a strong economy correlates with more forign investment. Similarly, the cluster of low and moderate values may not consist of the same countries either, which means that we cannot establish association between the two statistics. All we can conclude is that on a global scale, there is a cluster for relatively weak and moderate economies while strong economies separate themselves by existing as distant outliers, as indicated by both GDP and foreign inflow.

## Data Analysis

In the EDA, we created graphs, broadly analyzed the plots, and came up with reasons that we believe explained why there was an association between 2 different variables. In the data analysis section, we intend on using regression analysis and hypothesis testing to further explain some of our earlier findings and observations.

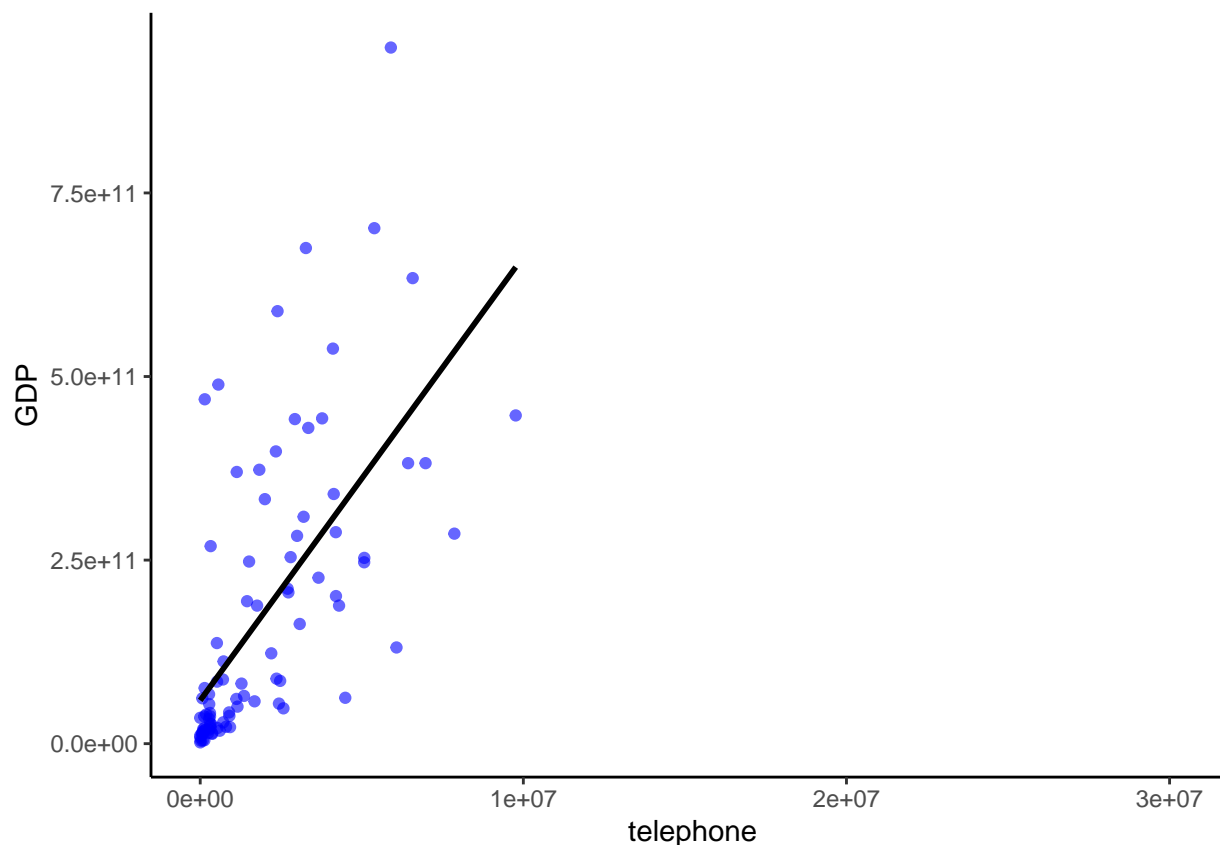#Further Data Analysis - GDP v Telephone

When looking closely into the scatter plot between telephones and GDP, we noticed primarily two influential points: China and the United States. China is an outlier because it has the

highest population in the world so there is a much higher number of people who have telephone lines. China also has a relatively low GDP per capita, which is why its' GDP: telephone ratio is significantly lower than that typically present in other countries. In contrast, the U.S. has one of the highest GDP per capita in the world while also being the country with the highest total GDP. Since the population of the U.S. is much lower than other countries with similar levels of GDP, its GDP:telephone ratio is much higher than that of other countries. Infrastructure was established over a century ago in the U.S. and telephones were adopted by Americans much earlier than the rest of the world. Even as the population grew, many households already had telephones and didn't need additional ones. The growth in the number of telephones couldn't keep up with the growth in GDP in the U.S. These reasons are why the U.S. is an obvious outlier.

```
## [1] 0.6515349
```

```
## Warning: Removed 13 rows containing non-finite values (stat_smooth).
```
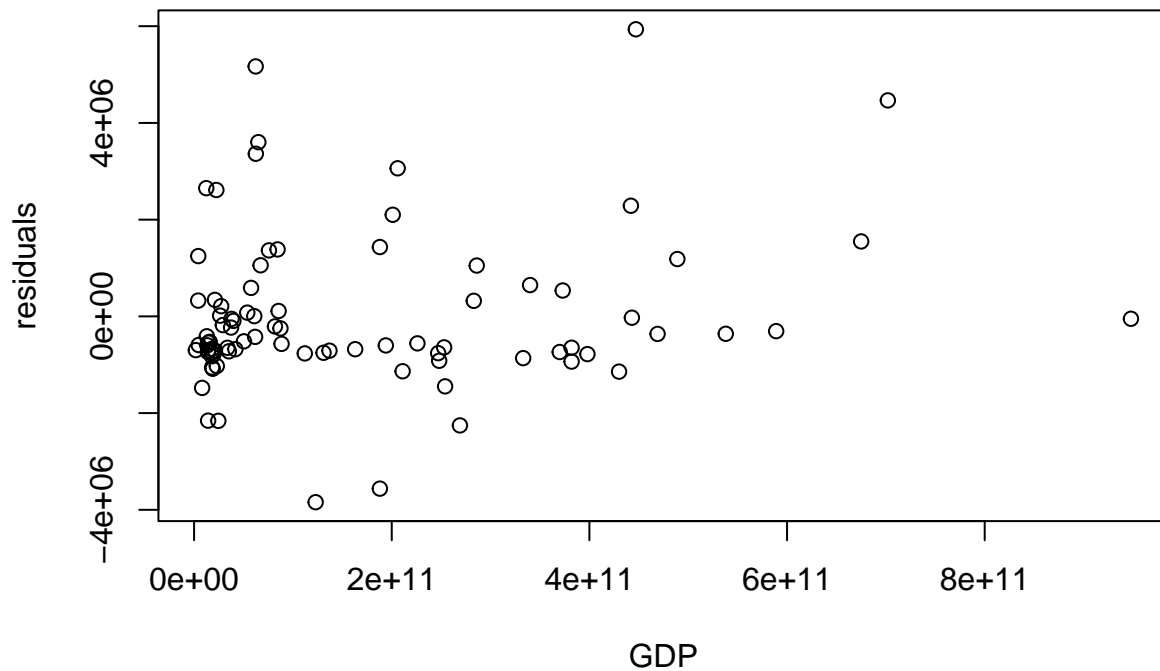
```
## Warning: Removed 13 rows containing missing values (geom_point).
```



The figure above is the scatter plot after removing all outliers. Removing outliers decreased the correlation between GDP and telephones to 0.652, which is still a moderately positive correlation. To learn more about the correlation between GDP and number of telephones, linear regression analysis was conducted on the scatterplot while excluding outliers. The scatter plot was found to be heteroscedastic so we couldn't use the normal approximation within vertical slices.

However, as seen in the graph above, a equation for the regression line was created for the scatter plot: y = 5.895* 10^10 + 6.043* 10^10*x, with x = number of telephones and y = GDP. We can even use the regression line to make predictions and gauge a country's forecasted GDP.

```
##
## Call:
## lm(formula = GDP ~ telephone)
##
## Coefficients:
## (Intercept)    telephone
##    5.895e+10    6.043e+04
```



To measure the accuracy of the linear model, we created a residual plot. As seen in the figure above, there appears to be no trend in the residual plot, which indicates that the linear model provides a decent fit to the data. Additionally, $R^2 = 0.42$, which means the regression model explains 42% of the variability in the GDP that's attributed to the number of telephones.

Although the correlation doesn't mean causation, based on the regression model and updated scatter plot, the data supports our initial belief that the number of telephones in a country is associated with its GDP.

#Further Data Analysis - GDP v Percentage of Parliamentary Positions Held by Women

```
##              country_name female_pop_perc parliament_women     p-value
## 14                Bolivia         49.78340         53.07692  0.67524427
## 101                Sweden         49.94578         46.13181  0.33871991
## 81           South Africa         50.69415         42.74809  0.13821081
## 25                Finland         50.72076         42.00000  0.13737907
## 89   United Arab Emirates         30.63669         22.50000  0.11886698
```

13

```
## 77            Norway       49.52463       41.42012 0.11377479
## 85           Senegal       51.27733       41.81818 0.07280322
## 51           Iceland       49.81171       38.09524 0.06267875
## 26          Ethiopia       49.97889       38.75686 0.05981453
## 11           Belgium       50.59332       38.00000 0.05015731
## 72       New Zealand       50.83777       38.33333 0.04667772
```

Based on the results of EDA, we found no association between GDP and the percentage of women in the parliament. This suggests that a country's economic status does not reflect its social progress, especially in terms of gender equality. In other words, an economically advanced country is just as likely to have gender inequality (as measured by the proportion of women in the legislature), as opposed to our predictions. We decide to investigate this issue further by comparing the two variables. We believe that in a country with no gender bias, the proportion of women in the legislature would be equal to the proportion of women in the entire population. To determine the extent of gender bias in the legislative system of all countries, we will do a hypothesis test for each country using the two variables to calculate the p-value.

Null Hypothesis - there is no bias in the parliamentary system (% of women in population = % of women in parliament)

Alternate Hypothesis - (% of women in parliament < % of women in population)

Assumptions: No laws restricting women to be in parliament (not taking into account differences in legal systems/culture). Assuming that regardless of the legislative system, the number of seats in the legislature is equivalent to the number of seats in parliament.We are assuming that each country has a parliamentary size of at least 50 and will use 50 as our sample size for the hypothesis test. 50 was chosen as sample size to be a conservative estimate for the p-values because the vast majority of countries have a legislature size greater than 50 according to Wikipedia.

First, we reorganized the main dataframe to extract the country names, percentage of females in total population, and percentage of females in parliament.

Then, we made an vector of p-values for each country using pbinom, where: q = percentage of women in parliament * 50 (i.e. the number of women in parliament if the country had a parliament size of 50), size = 50, prob = percentage of women in total population

Lastly, we append the vector to our table and sort the p-value column in ascending order.

Using an alpha of 0.05, we fail to reject the null hypothesis for 11 countries when size = 50. In other words, there are 11 countries that do not show gender bias in their legislative system. On the other hand, we reject the null and accept the alternate hypothesis for 109 countries, which means that those countries exhibit gender bias in their parliaments since the number of seats taken by women is lower than what it "should be."

This analysis is significant because it proves that the difference in percentage of women in population and percentage of women in parliament cannot be explained by chance for most countries in the world. In other words, it further proves our first point that economic status is not an indicator of gender equality and that economically advanced countries fall prey to gender bias. It is important to note that this is the result of using 50 as our sample size. Countries like China and the UK have legislature sizes of >1000, and most countries have theirs higher than 50. If we used a more aggressive yet accurate estimate of parliament size, such as 100, the p-values would definitely decrease, and the number of countries that do not show gender bias would decrease

even more. Interestingly, Bolivia is the only country in the world that has more women in parliament than in the population by proportion. On the other end of the spectrum, Yemen has 0 seats taken by women. Overall, we conclude that there seems to be systemic gender inequality in most of the world, and economic progress alone does not eliminate it.

#Conclusion

Sustainable development goals cannot be placed in a hierarchy - they are all important factors when looking at social and economic development. As a group, we asked one overlying question: What are some of the most important issues countries face? To explore what issues to focus on, we analyzed the association of different variables with GDP by using scatter plots and histograms. Based on our analysis, we narrowed down our exploration to find that technological progress and gender equality are two issues which have a relationship with GDP worth focusing on.

For our 1st detailed analysis, we looked at the correlation between the number of telephones and GDP of a country. We found a moderately positive correlation between the 2 variables after accounting for outliers. After creating a linear regression model and a residual plot, we were able to confirm that there was in fact an association between the number of telephones and GDP of a country.

For our 2nd detailed analysis, we wanted to identify whether there was systematic gender inequality in the parliament across countries in the world. We decided to conduct a hypothesis test that compared the proportion of women in the population of a country to the proportion of women in the parliament of that country. After conducting the hypothesis test, the results allowed us to come to the conclusion that the difference between the proportion of women in the population and in the parliament was not just due to chance. Although there may be confounding factors influencing the result, we are quite confident that gender inequality in the parliament is present in countries around the world.

Overall, we were able to narrow down our list of issues that countries face and draw interesting findings. We can confidently say that some issues countries should continue to focus on improving are gender inequality in the parliament and technological advancements in areas such as increasing availability of telephones.