Jia Ming (Carter) Huang (Student ID: 1003893245)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total
Mark								

Table 1: Marking Table

Due Date: Friday, February 11, 2022, 11:59 PM
Number of Pages: 47

Please note that all the R code used to generate the data used in this analysis is included for your reference in Appendix A.

List of Problems

1 Correlation Between GDPPC and Geographical Location

Since the end of the Second World War, terms such as "Third World" and "Global South" evolved in developed countries to impose an ideology that the world is clearly divided into rich and poor. "Global South" is a recent term that seeks to eliminate some of the derogatory contexts of the term "Third World". This investigation examines the emergence of the term "Global South" and its statistical implications. We must first define key terms that will aid in our investigation.

GDP per capita (GDPPC) is the measure of wealth by dividing the gross domestic product of a country equally among its population. While the purchasing power of US\$1 in Country A may not reflect the purchasing power of US\$1 in Country B, this metric pegs the wealth of nations to the US dollar. Due to the abandonment of Britain's Gold Standard, the Bretton Woods Agreement, and America's role as a weapon's supplier in World War II, the Global Reserve Currency today is the United States Dollar [1], which serves as a reliable benchmark of economic wealth across nations. The term "Global South" implies that nations north of the equator are more wealthy in terms of GDP per capita. In TCdata360 [2], there are 67 high income countries (HICs), of which only 7 are in the Southern Hemisphere.

Before we delve into our statistical analysis, it is important to note that only 13% of the human population lives south of the equator [3], which is a major contributor to the distribution of wealth by geographic location. Intuitively, we recall that developed nations such as Australia and New Zealand are in the Southern hemisphere as well. This begs the question: does GDP per capita increase as latitude increases?

Figures 1, 4, 7, and 10 show the distribution of GDP per capita versus latitude in 1960, 1980, 2000, and 2019, respectively. Notice that the orange dots per the Income Level legend are HICs. If an individual with the belief that the "Global South" is an accurate description of the state of economics of nations, they would appeal to their own confirmation biases that the data indeed proves the dependence of a country's economics on geographic location. As inquisitive data scientists, we extract the R-squared values from the data to see the strength of correlation between latitude and GDP per capita.

Figures 2, 5, 8, and 11 show the correlation between GDP per capita versus latitude in 1960, 1980, 2000, and 2019, respectively. We can see that regardless of time, there is low correlation (< 0.4) between GDP per capita versus latitude, which suggests that the wealth of nations does not depend on latitude. Let us recall that in our earlier analysis, the Southern Hemisphere also hosts wealthy nation's like Australia and New Zealand, and that the concentration of HICs in the South are comparable to that of the North. In the South, $7/67 \approx 10\%$ of HICs are shared by $\approx 13\%$ of the Earth's population. This prompts us to look at the question differently. One hypothesis is that nations further from the equator are wealthier, because harsher climates promote the development of technology [4]. Technology, being a driving force of economic development and superiority is a common denominator between developed countries far from the equator. Could we look at the distance from the equator as an indicator of economic success?

Figures 3, 6, 9, and 12 show the correlation between GDP per capita versus distance from the equator in 1960, 1980, 2000, and 2019, respectively. In Figure 3, we see that the correlation be-

tween distance from the equator and GDP per capita in 1960 is ≈ 0.42, which is the highest linear regression R-squared value in all comparisons. Coupled with high p-values in our regression coefficients, we can conclude that there is not a significant correlation between the geographic location and economic well-being of a country.

Our analysis doesn't stop here. As we analyze the trends over time, from 1960 to 2019, we notice that the correlation between the geographic location and economic well-being of countries decreased significantly. If we fancied the hypothesis that geographically extreme countries have better economic well-being, we can conclude that in 1960, there was an acceptable correlation that countries far from the equator were in fact richer. If we take into consideration the population distribution between north and south, and neglect the data from the southern hemisphere, the term "Global South" may have been sensible. However, as the correlation wanes today, partly due to the rising economies and populations of nation's near the equator such as China, India, and Nigeria, this term is no longer an accurate description of the world's economic affairs. We can see that the "Global North" is a term that includes North America, Europe, Australia and New Zealand, which is more on par with the observation that the correlation between GDP per capita and distance from the equator is stronger than that of absolute latitude.

Therefore, through statistical analysis and an understanding of geopolitical situations that gave rise to terms like "Third World" and "Global South", we realize that the surface-level trends we see in our Latitude versus GDP per capita graphs over time only paint a biased view towards the economic state of the world. There are simply more people, borders, and as a result, conflicts near the equator. The Western World is a direct beneficiary of the consequences of twentieth century conflicts preceded by an era of colonialism. The geographic location of developed nations today is merely a coincidence for their economic well-being. Rather, geographic location may be a catalyst for sociopolitical, cultural, and economic forces that bring about strong economic performance in the modern era. Instead of using the term "Global South", which is quite inaccurate, we should use the term "Developing Nations". Since there is no direct, statistically significant correlation between geography and a country's economic well-being, its preferable to describe a nation by its rate of economic growth. This is a direct, unbiased, value-free, and statistically perfect indicator of economic well-being, as it is a one-to-one comparison between GDP per capita and GDP per capita growth.
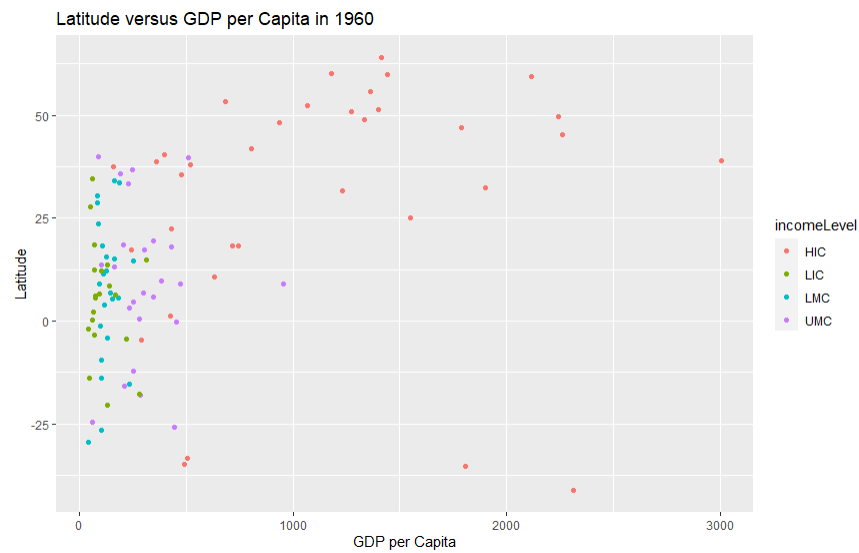
Figure 1: Latitude versus GDP per Capita in 1960

Figure 2: Statistical Correlation by Latitude in 1960

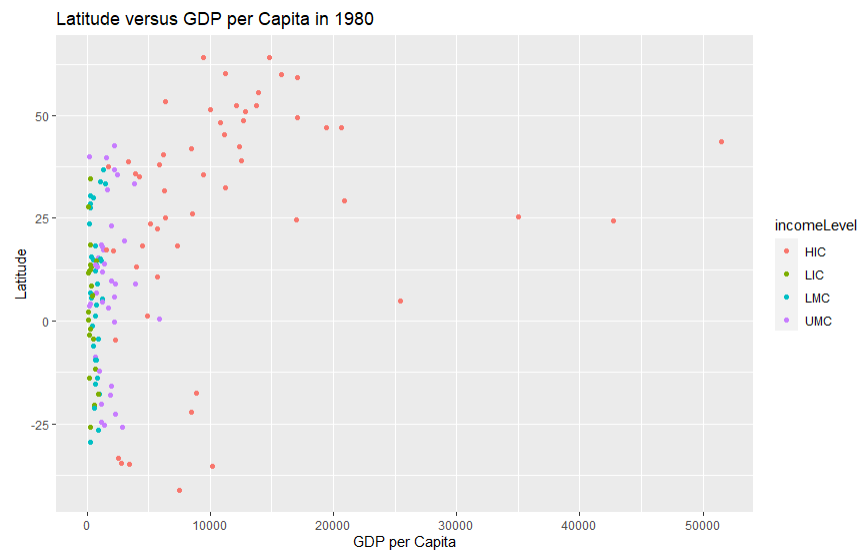Figure 3: Statistical Correlation by Distance From Equator in 1960

Figure 4: Latitude versus GDP per Capita in 1980

```
Call:
lm(formula = gdp ~ geo.lat, data = merged)

Residuals:
   Min    1Q Median    3Q    Max
 -7872  -3784  -1727   1475  42972

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2898.34     668.14   4.338 2.66e-05 ***
geo.lat      129.06      23.86   5.409 2.51e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7004 on 147 degrees of freedom
Multiple R-squared:  0.166,    Adjusted R-squared:  0.1603
F-statistic: 29.26 on 1 and 147 DF,  p-value: 2.513e-07
```

Figure 5: Statistical Correlation by Latitude in 1980

```
Call:
lm(formula = gdp ~ abs(geo.lat), data = merged)

Residuals:
   Min    1Q Median    3Q    Max
 -8410  -2651  -1176   853  42073

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -502.21     979.50  -0.513    0.609
abs(geo.lat)  227.39      34.98   6.500 1.17e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6759 on 147 degrees of freedom
Multiple R-squared:  0.2233,    Adjusted R-squared:  0.218
F-statistic: 42.25 on 1 and 147 DF,  p-value: 1.171e-09
```

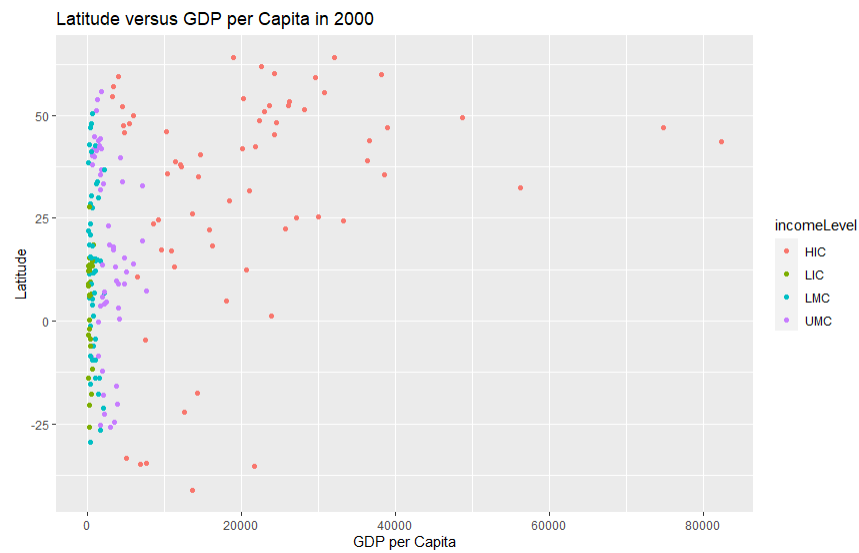Figure 6: Statistical Correlation by Distance From Equator in 1980

Figure 7: Latitude versus GDP per Capita in 2000

```
Call:
lm(formula = gdp ~ geo.lat, data = merged)

Residuals:
   Min    1Q Median    3Q    Max
-13483  -7017  -2991   3235  69575

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4377.13    1078.93   4.057 7.18e-05 ***
geo.lat       192.39      34.51   5.574 8.14e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11960 on 196 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.1368,    Adjusted R-squared:  0.1324
F-statistic: 31.07 on 1 and 196 DF,  p-value: 8.145e-08
```

Figure 8: Statistical Correlation by Latitude in 2000

```
Call:
lm(formula = gdp ~ abs(geo.lat), data = merged)

Residuals:
   Min    1Q Median    3Q    Max
-15844  -6051  -2004   2231  68631

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -381.49    1517.59  -0.251    0.802
abs(geo.lat)   322.78      48.54   6.649 2.86e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11630 on 196 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.1841,    Adjusted R-squared:  0.1799
F-statistic: 44.21 on 1 and 196 DF,  p-value: 2.858e-10
```

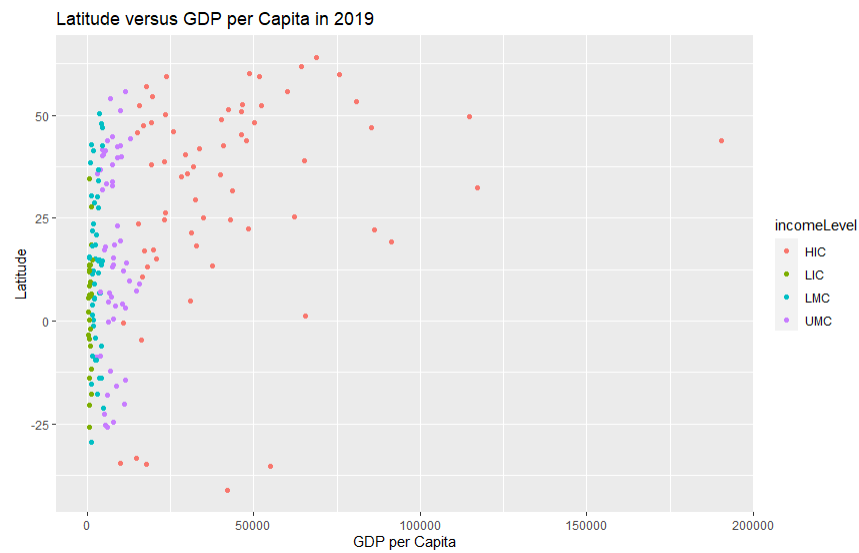Figure 9: Statistical Correlation by Distance From Equator in 2000

Figure 10: Latitude versus GDP per Capita in 2019

Figure 11: Statistical Correlation by Latitude in 2019

Figure 12: Statistical Correlation by Distance From Equator in 2019

2 Fixed Broadband Internet Subscriptions in Landlocked Developing Country's (LLDCs)

In Figure 13, we can see that the Fixed Broadband Internet Subscriptions in Non-LLDCs are higher than in LLDCs between 2017 and 2018. The box plot medians can be used to analyze statistical significance [5]. If we remove the outlier in the LLDC box plot, we can see that its median is lined up with the bottom of the Non-LLDC box plot. This suggests that there is marginally statistical difference between the two, and that it is marginally likely for LLDCs and Non-LLDCs to have statistically significant disparities between the number of Fixed Broadband Internet Subscriptions. However, while both medians are on the lower side of the Fixed Broadband Internet Subscriptions, we can see that in both cases of LLDCs and Non-LLDCs, there is heavy variance towards the higher end of the number of Fixed Broadband Internet Subscriptions, which suggests an internet access inequality around the globe.

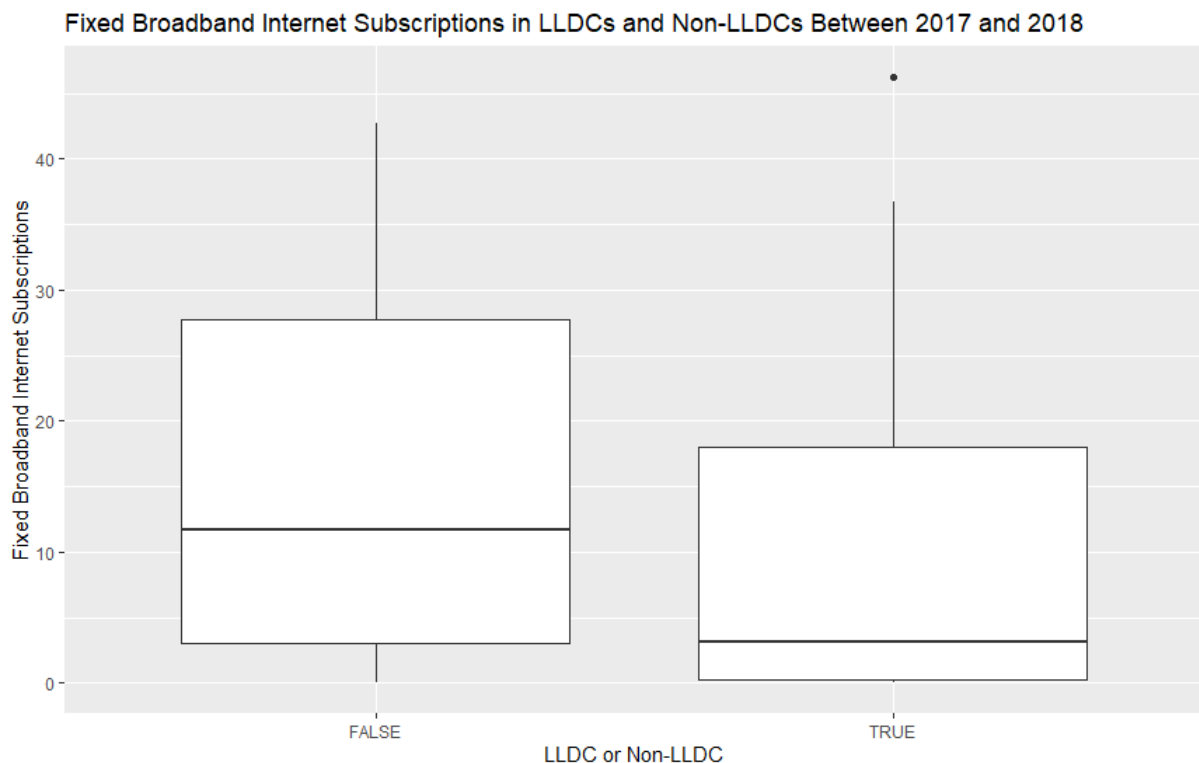Figure 13: Fixed Broadband Internet Subscriptions in LLDCs and Non-LLDCs Between 2017 and 2018

As we analyze the data across time in Figure 14, we see that between 2007 and 2018, both LLDCs and Non-LLDCs experienced significant growth in the number of Fixed Broadband Internet Subscriptions. However, the overall trend seems to suggest that LLDCs are experiencing less growth overall, and the disparity is growing.

While LLDCs are able to gain internet access, they don't have landing sites for the submarine cables that connect the world to the network [6]. Without access to the ocean, LLDCs gain internet access by connecting to landlines in neighbouring nations that have access to the ocean. However, this poses two challenges: LLDCs have to share the same internet infrastructure with other countries, which can disconnect the service due to dispute or necessity [6]. Furthermore, in order for internet to reach landlocked countries, they must build cables across land, which can have difficult terrain. This significantly increases the cost of delivering internet to LLDCs, which raises the cost of internet access for countries already disadvantaged economically due to their geography. Considering that the Internet is responsible for a sizeable portion of the world's economic growth, poor access to Internet is detrimental to LLDCs. Consider that in Yemen, a 5GB movie takes 30 hours to download, while in Taiwan, it takes just 8 minutes [7]. Although Yemen is not an LLDC, the situation highlights how slow internet access can significantly reduce efficiency and in turn, economic growth.

To address the issue of internet cable access in LLDCs, there can be international provisions in place to guarantee access to the ocean. For example, land which internet delivery cables sits on can be considered international territory, and only cables originating from these central delivery lines can serve as a country's sovereign property. This ensures that each country is able to access the Internet fairly, while also maintaining their own infrastructure so as not to do so at the expense of others. This solves the political pressure for nation's to maintain good relations in order to have reliable internet access. Furthermore, in the short term, LLDCs can try to build relations with other Non-LLDCs that they share borders with to provide redundant access to internet in case of disputes. To address the issue of bandwidth, speed, and penetration, new technologies that don't rely on cables can provide sustainable internet to LLDCs. LLDCs can focus on satellite networks that can transmit internet from offshore locations, where they can hook up to the global internet supply. Investing in wireless infrastructure can eliminate the burdens of internet access in LLDCs, so long as wireless technology can be just as reliable as broadband.
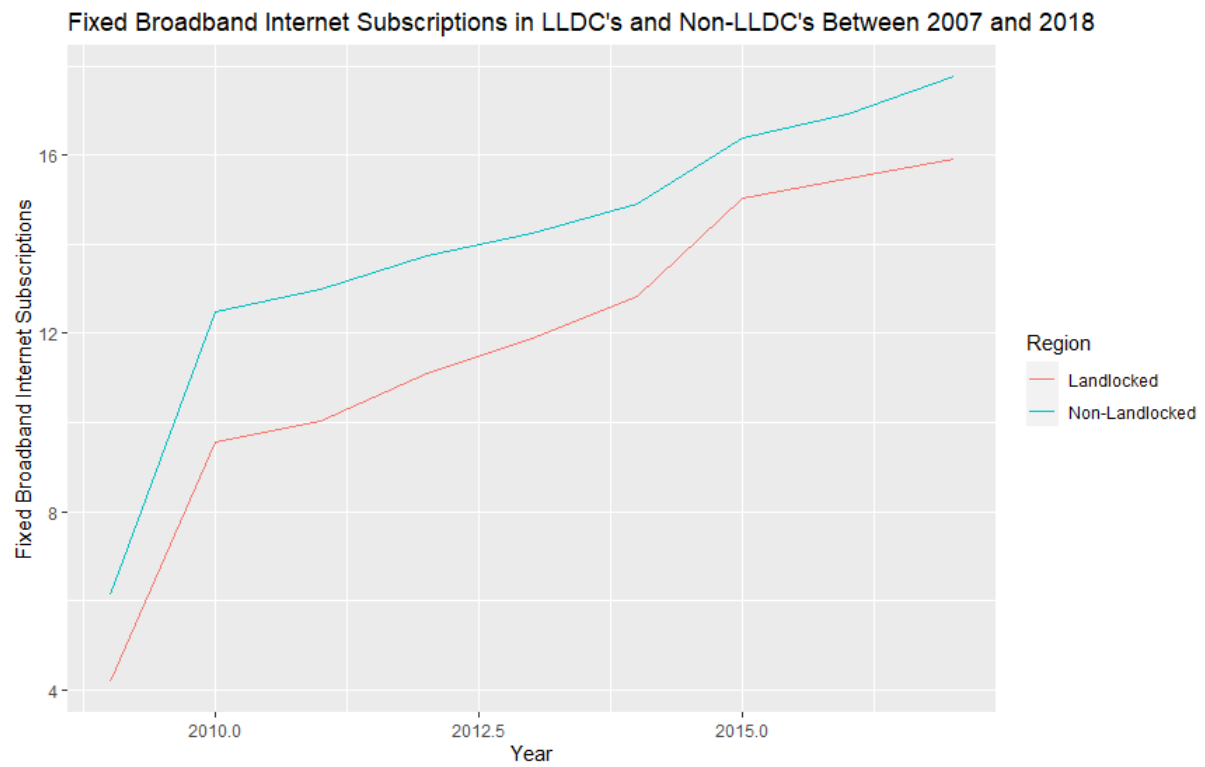
Figure 14: Fixed Broadband Internet Subscriptions in LLDCs and Non-LLDCs Between 2007 and 2018

3 Life Expectancy

Life expectancy has grown significantly since 1960 as a result of technological advancements and the absence of wars of attrition. First, we must discuss how we gather, sanitize, and interpret the life expectancy data. By taking the TCdata360 datasets [2] for life expectancy at birth of males and females from 1960 to 2019, we calculate the cell-by-cell average for each country, each year. Then, we loop through each row and column and take the change in life expectancy for each consecutive year in each country. Finally, we keep track of the maximum and minimum life expectancy at birth, as well as the year that these changes occurred for each nation to gain insight into socioeconomic factors that caused these changes.

Cambodia and Rwanda had the largest increases in life expectancy, respectively. Cambodia's life expectancy at birth increased by 5.72 years from 33.51 to 39.23 between 1981 and 1982. Rwanda's life expectancy at birth increased by 4.45 years from 35.38 to 39.83 between 1996 and 1997.

Not surprisingly, Rwanda and Cambodia had the largest decreases in life expectancy, respectively. Rwanda's life expectancy at birth decreased by 5.02 years from 38.46 to 33.44 between 1989 and 1990. Cambodia's life expectancy at birth decreased by 4.55 years from 32.88 to 28.33 between 1973 and 1974. In fact, in 1993, Rwanda's life expectancy at birth was only 26.20 years old, and in 1977, Cambodia's life expectancy at birth was only 19.34 years old. In comparison, countries like Japan and Iceland had life expectancy at birth in the high 70's during this timeframe.

In Figure 15, we can see the life expectancy at birth of Cambodia and Rwanda. Let us research what happened in these two countries that caused such a massive fluctuation. In Cambodia, the Khmer Rouge began its insurgency in 1968, which overthrew Sihanouk, the head of state of Cambodia. In 1970, North Vietnam and the Khmer Rouge began the Cambodian Civil War. Between 1970 and 1975, between 600000 and 1000000 civil war deaths were recorded, with higher disputed figures as well [8]. Between 1969 and 1973, the US bombed Cambodia, leading to between 30000 and 500000 deaths, which account for approximately 17% of all civil war deaths [8]. Furthermore, between 1975 and 1979, the Cambodian Genocide by the Khmer Rouge killed between 1.5 and 2 million people, which accounted for a quarter of Cambodia's population [9]. Finally, to top it off, the Vietnamese invaded Cambodia in 1979 to destablize the Khmer Rouge regime [8]. Therefore, as civilian deaths were rampant, and the nation was in abject poverty, Cambodia's life expectancy at birth sharply declined between 1970 and 1980.

In the 1300s, the Tutsi people migrated into modern-day Rwanda, which was originally inhabited by the Twa and Hutu peoples. Tutsi's establish a kingdom that lasts over 500 years, until European forces occupy the nation. When Belgium retreated out of Rwanda in the 1950s, the Hutu people issued a manifesto calling for a change in Rwandan power structure, which formed Hutu political parties. Inter-ethnic violence due to historical Tutsi power and oppression lead to civil unrest that extends into the 1980s [10]. In 1990, these ethnic clashes become the prelude to the Rwandan Civil War. At the end of the civil war, in 1994, the Hutu's commit genocide against the Tutsi people, resulting in between 400000 and 800000 deaths in just 100 days [10]. Therefore, the constant ethnic conflicts in the mid-20th century in Rwanda, which sometimes resulted in as many as 20000 deaths in one instance [10], directly contributed to the sharp decline in Rwanda's life expectancy at birth between 1985 and 1994.

Now, we take the gender-averaged mean life expectancy at birth of countries in Sub-Saharan Africa compared to the World, as shown in Figure 16. Life expectancy at birth in 1960 is significantly lower in Sub-Saharan Africa than in the rest of the world. Between 1990 and 2000, it seemed to drop slightly, before increasing at an accelerated rate. Note that the life expectancy at birth of the world includes Sub-Saharan Africa, so you can see a slight dip at the same time Sub-Saharan Africa experienced its plateau. Similarly, we can see that while Sub-Saharan Africa is experiencing significant life expectancy growth rates recently, the World's life expectancy is growing more slowly, despite inheriting the growth rate from Sub-Saharan Africa. This suggests that Sub-Saharan Africa's life expectancy growth rate is significantly higher than the rest of the world.

Research shows that the dip in life expectancy in Sub-Saharan Africa in the 1990s was due to the HIV/AIDS epidemic [11]. In contrast, the recent advancements in vaccinations and medical technology has reduced mortality by protecting against diseases like malaria. The growth in GDP per capita has also been significant, from $600 in 2000 to more than $1500 today, which explains the sharp increase in life expectancy in the 21st century. Furthermore, water quality and food nutrition has increased in Sub-Saharan Africa, allowing the middle-class to prosper and drive up life expectancy [12]. These advancements are partly due to policy changes which open Africa up to investment around the world, like with the Belt and Road Initiative [13].
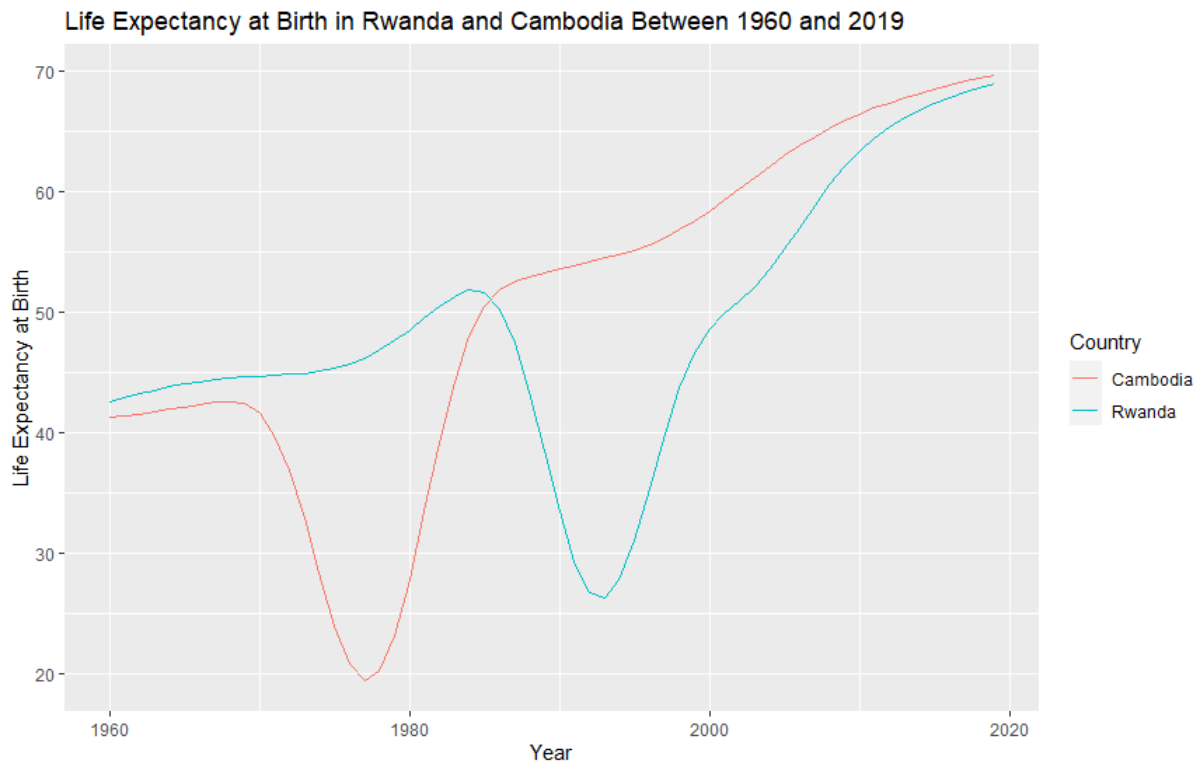

Figure 15: Life Expectancy at Birth in Rwanda and Cambodia Between 1960 and 2019
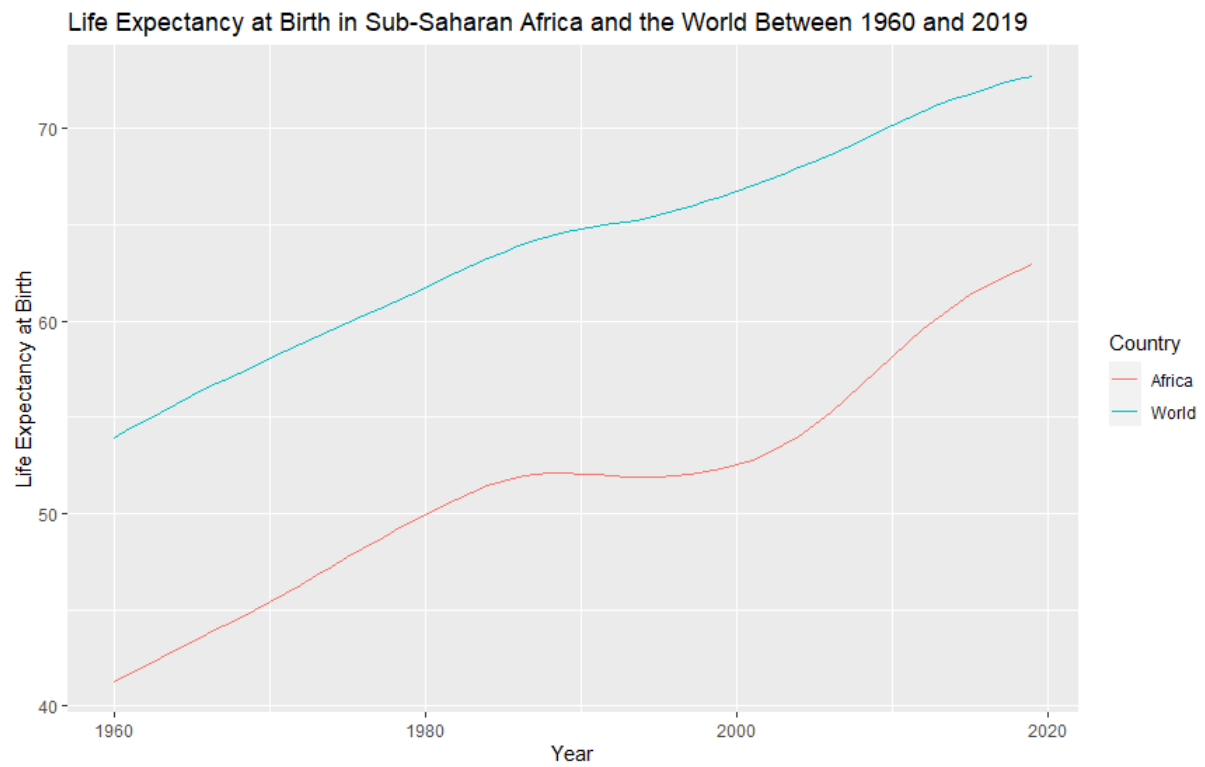
Figure 16: Life Expectancy at Birth in Sub-Saharan Africa and the World Between 1960 and 2019

4   Wealth-Health, Chicken-Egg

In order to extend our analysis of life expectancy rates, let us take a look at Figures 17 and 18, which show Bangladesh, Korea, and Malaysia's life expectancy and GDP per capita, respectively. If we compare the growth of life expectancy at birth, we can see that all three countries exhibit rather uniform growth from 1960 to 2019. On the other hand, the GDP per capita of all three nations are negligibly low between 1960 and 1980, before experiencing sharper growth until 2019. This suggests that nation's get healthier regardless of wealth, and wealth tends to follow after health. In fact, in Bangladesh, we can see that while life expectancy had been steadily growing from 1960 to 2019, the GDP per capita remains low compared to Korea and Malaysia. One hypothesis is that better health and higher life expectancy means more time for an individual to generate wealth. Let us do some research to explore this trend in more detail.

First, we take a look at data during the COVID-19 pandemic, where the standardized deaths due to the virus is correlated with negative GDP growth [14]. This implies that the decrease in health leads to the decrease in GDP, rather than the other way around. The reason is that before the pandemic, the global GDP was rising, and only took a downturn when the pandemic hit. Secondly, we take a look at the Preston Curve in 2012, which shows the relationship between wealth and health weakens as countries become wealthier. In fact, past a certain threshold, we see that higher income means lower life expectancy, and for developed countries, wealth and health are generally insignificantly correlated [15]. In certain countries like Cuba and the Dominican Republic, life expectancy at birth is on par with countries like the US and Slovakia, respectively, despite having a significantly lower GDP per Capita [15]. One explanation for this phenomenon is that health spending in emerging economies, while lower, has a comparatively significant impact than that in a developed nation. This is because the health infrastructure is so basic, that any spending in health will create a sizeable impact. Therefore, it is easier for the health of developing nations to improve without significant economic gains. However, while it seems like health comes before wealth, it is also proper to state that the correlation between the two are statistically insignificant, and there may not be a definitive cause and effect.

In short, a higher standard of health is often accompanied by other factors, such as institutionalization, political stability, and access to clean water and sanitary services. These factors are also direct contributors to economic growth. In developing nations, the impact of health spending is more pronounced than in developed nations, so it is economically prudent for these nations to invest in health. Across a certain threshold, notably that of a developed nation, the relationship between health and wealth becomes insignificant. In emerging economies, we can say that there is some indication that health comes before wealth, but that this relationship depends on other factors which have statistically stronger correlations.
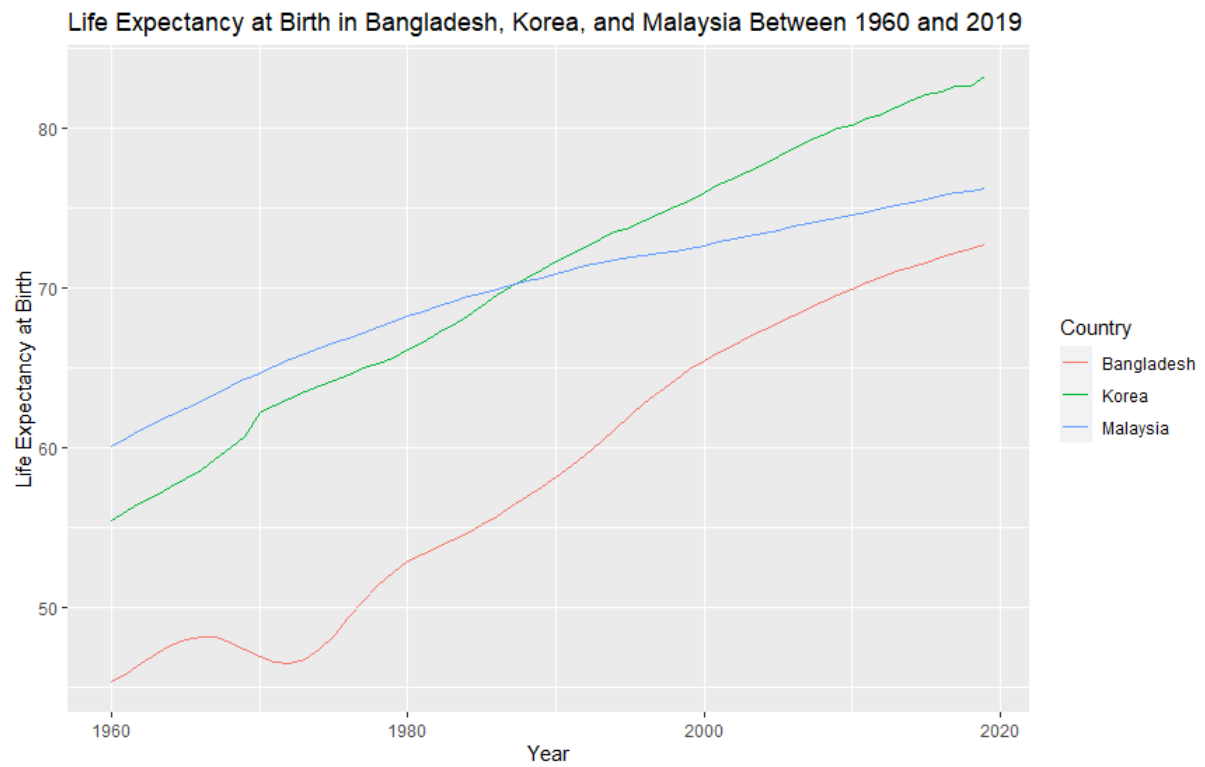
Figure 17: Life Expectancy at Birth in Bangladesh, Korea, and Malaysia Between 1960 and 2019
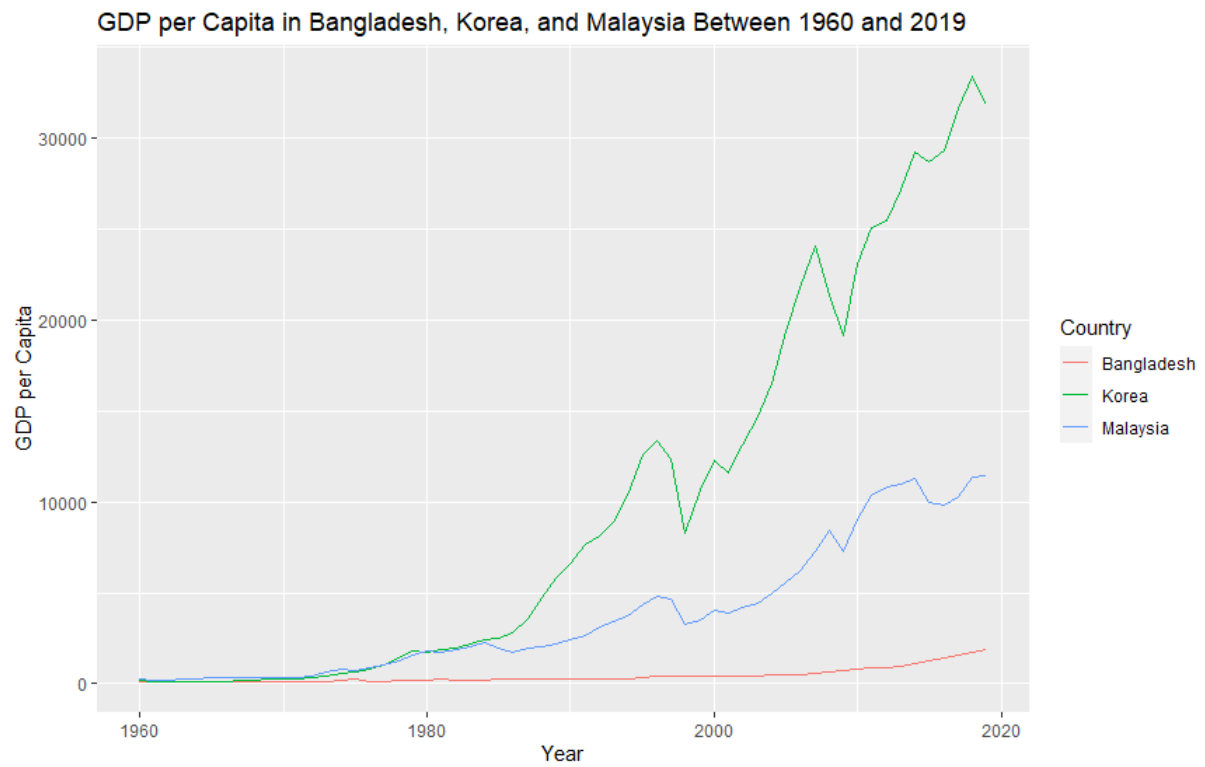
Figure 18: GDP per Capita in Bangladesh, Korea, and Malaysia Between 1960 and 2019

5 Socioeconomic Conditions That Affect a Nation's GII

There are four socioeconomic factors that affect a country's innovation capacity and success. These are political stability, institutions, ease of getting credit, and entrepreneurship as a desirable career choice. In Week 2, class discussions referred to the lack of some of these factors as directly contributing to poverty [16], and in effect, a nation's Global Innovation Index (GII). First, we describe how and why each of these 4 factors influence a country's capacity to innovate.

Political stability is defined by government transparency, predictability, and accountability [17]. The more stable a nation is, the lower the risk to start a business and engage in entrepreneurial activity. Healthy economies are driven by market forces, rather than centralized economic planning, especially for entrepreneurs who are most at risk for policy changes and regulations. Therefore, it is intuitive to associate political stability with a capacity for innovation. We assess this relationship in Figures 19 and 20.

The institutional framework of a nation is also important in assessing the capacity for innovation. Nations with strong institutional frameworks offer many protections for businesses, so that there is predictable and accountable support from public services. Innovation occurs in low-risk settings, so the peace of mind offered by strong institutional frameworks will undoubtedly promote innovation capacity. We assess this relationship in Figures 21 and 22.

Investment is the primary method of raising capital for new businesses [18]. Innovation can only persist if businesses are able to raise capital to run daily business operations. Credit is a way for financial institutions to lend money to business owners while assessing risk. The ease of getting credit is directly related to the ease of raising capital. Therefore, it is imperative for businesses to obtain the economic foundations required to continue operations before innovation can occur. We assess this relationship in Figures 23 and 24.

Entrepreneurship is the fundamental pillar of innovation, because it allows new business owners to bring their ideas into practice. Entrepreneurship as a desirable career choice pushes more individuals to start businesses, rather than to work at corporate settings. This encourages free market competition, as new ideas are being injected into the market, rather than large corporations monopolizing goods and services. As we know, monopolies hinder innovation, because there are no longer disruptive forces to pressure businesses to improve their products [19]. Therefore, entrepreneurship as a desirable career choice will ensure innovation capacity in a nation. We assess this relationship in Figures 25 and 26.

Evidently, Institutions versus GII Index in 2020 shows a significant statistical correlation, with an R-squared value of 0.7617. This suggests that institutional foundations are the primary driver of a nation's innovative capacity and success. Furthermore, if we take a look at Figures 27 and 28, we can see that this same pattern and high correlation exists when we compare Institutions with GDP per Capita across the world. It becomes easy to see that as the Institutions Index increases, so does the wealth of the nation. The conclusion we can draw is that institutionalization builds innovation, and as a result, wealth.

Figure 19: Political Stability versus GII Index in 2015

Figure 20: Statistical Correlation by Political Stability in 2015
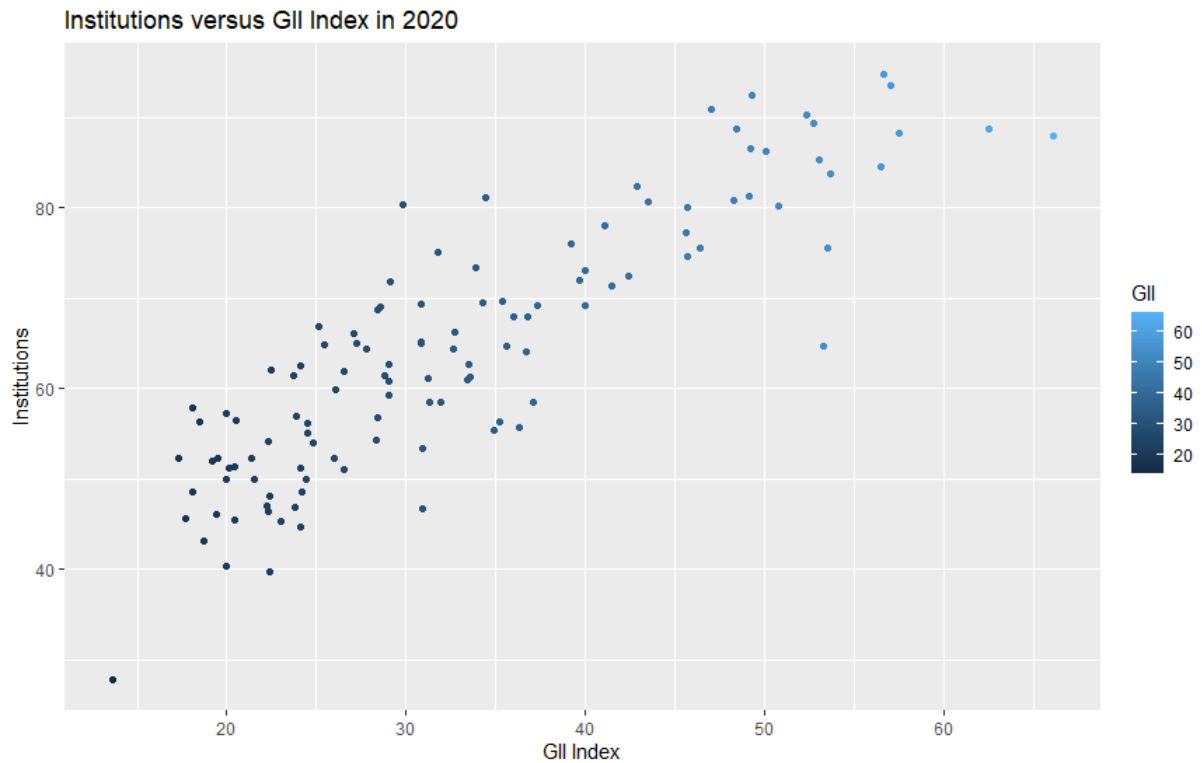
Figure 21: Institutions versus GII Index in 2020

Figure 22: Statistical Correlation by Institutions in 2020

Figure 23: Ease of Getting Credit versus GII Index in 2020

Figure 24: Statistical Correlation by Ease of Getting Credit in 2020

Figure 25: Entrepreneurship as a Desirable Career Choice versus GII Index in 2020

Figure 26: Statistical Correlation by Entrepreneurship as a Desirable Career Choice in 2020
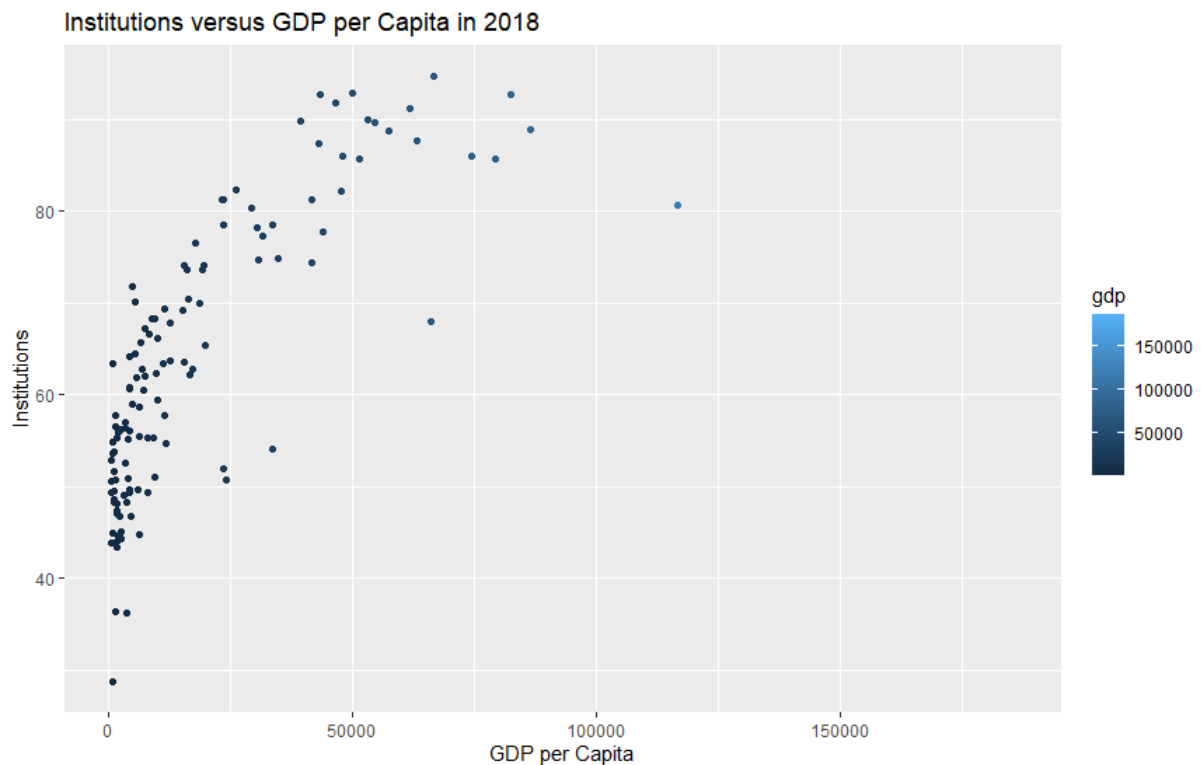
Figure 27: Institutions versus GDP per Capita in 2018

Figure 28: Statistical Correlation by Institutions in 2018
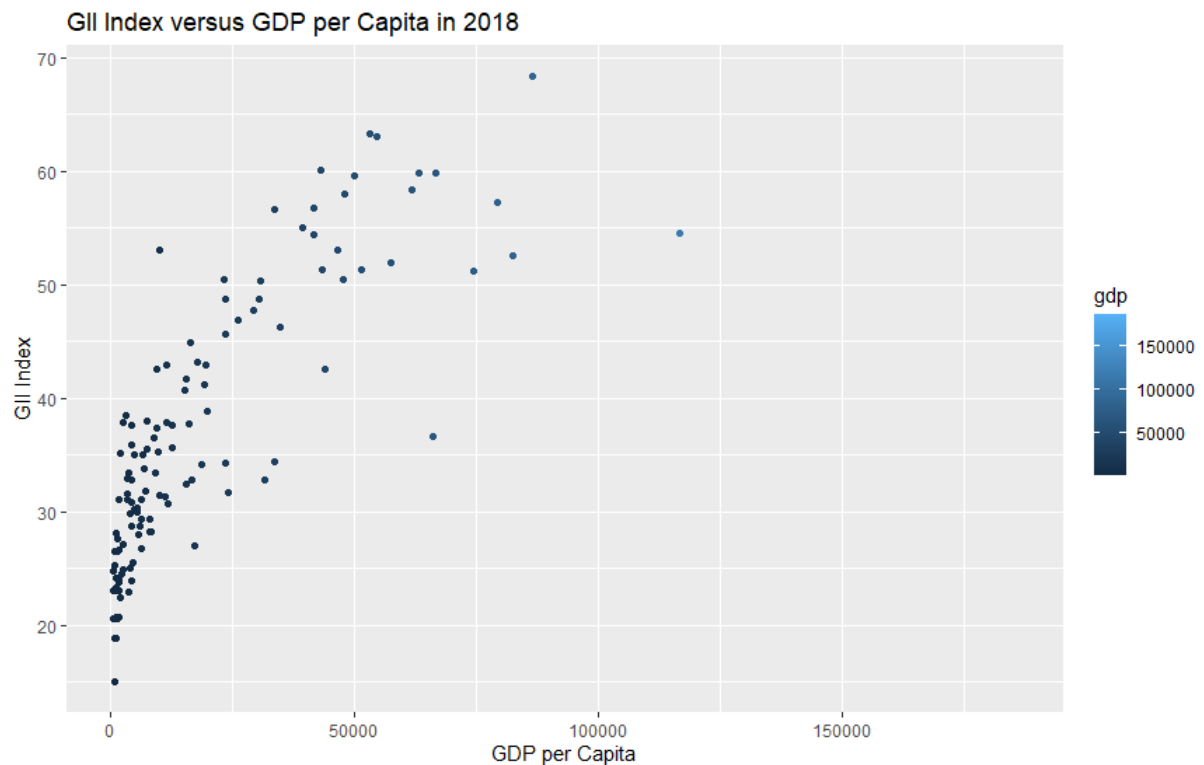
Figure 29: GII Index versus GDP per Capita in 2018

Figure 30: Statistical Correlation by GII Index in 2018

As a remark, we may wish to consider why institution, and not the other three factors? In terms of entrepreneurship, although entrepreneurship does promote innovation, it may also be a fall-back for people who are unable to find a job due to a lack of talent or opportunity. In terms of ease of

getting credit, it is dangerous for banks to lend credit out to anyone, anywhere. This was the primary catalyst for the US stock market crash of 1929 [20]. Finally, political stability is an interesting one. Although the R-squared value is low, it may have been due to the cluster of countries with a low GII index scattered across the political stability spectrum. In fact, the p-value is quite low, which suggests that political stability does promote innovation. However, institutionalization is likely the primary driver of innovation capacity, because it signifies a mature government framework. This framework, understandably, showcases political stability and ease of getting credit in an accountable and predictable way. In fact, entrepreneurs arising from institutionalized countries are often driven and well-educated, rather than becoming entrepreneurs as a necessity for survival, as is often seen in low-income countries. The mingling of these four factors are key indicators for a country's innovative capacity and success; as we can further derive from Figures 29 and 30, where we see how a nation's GII is correlated strongly with it's GDP per capita.

6   Availability of Scientists and Engineers Around the World

Intuitively, the more scientists and engineers there are in a country, the faster its technological development, and in turn, the better its economic well-being. Let us take a look at Figures 31 and 32, which show the distribution of scientists and engineers, and GDP per capita around the world, respectively. Immediately, we can see countries like Canada and India have a high number of scientists and engineers, but a lower GDP per capita than counterparts like the USA and Australia, which have comparatively less scientists and engineers. On the other hand, nations like China have less scientists and engineers, with a strong economy. It is important to note that the availability of scientists and engineers is not an absolute value, but rather, it is relative to the population of each country. This can be interpreted as a "per capita" concentration of scientists and engineers for the purpose of this analysis.
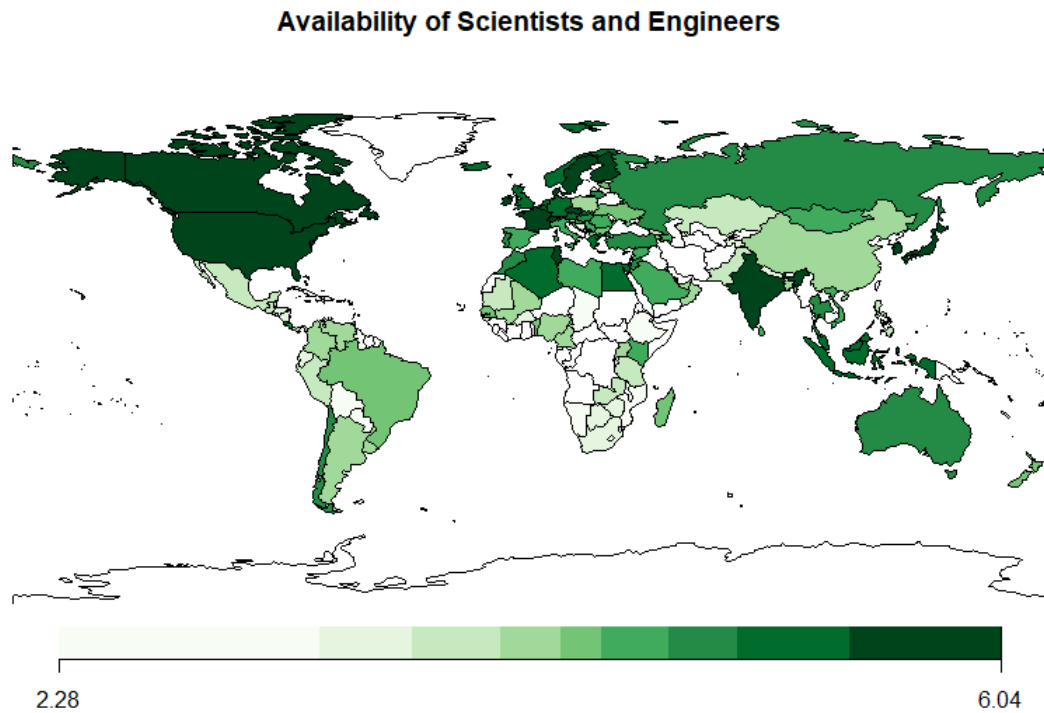
Figure 31: Availability of Scientists and Engineers Around the World Between 2017-2018
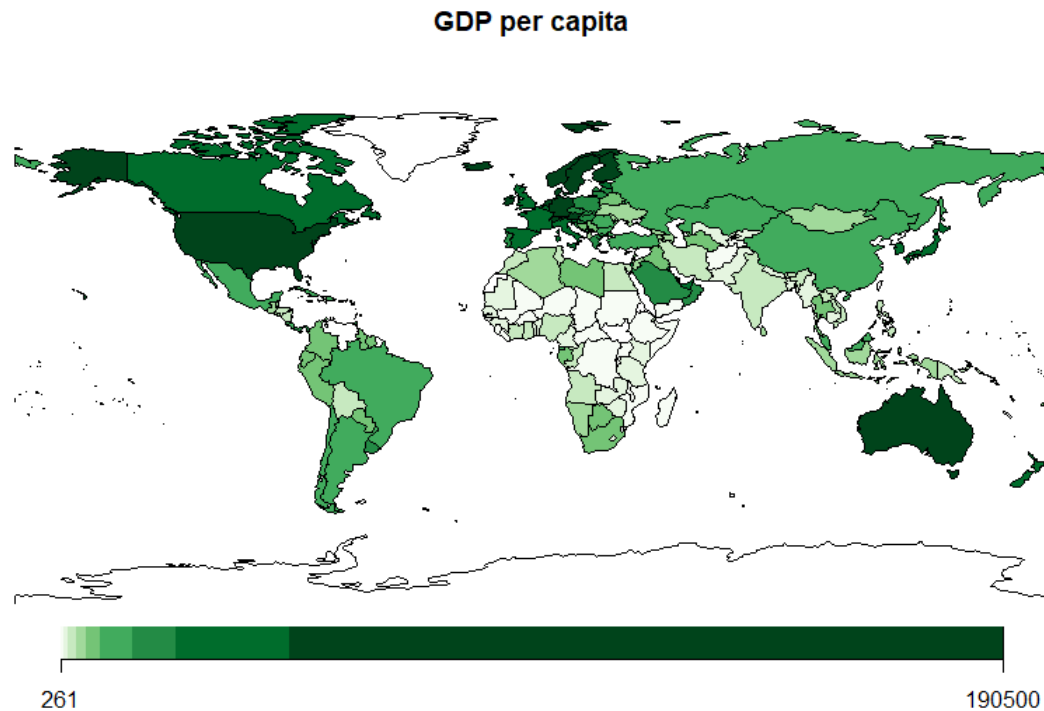
Figure 32: GDP per Capita Around the World in 2018

In order to understand the correlation better, we plot a graph of the Availability of Scientists and Engineers against GDP per Capita in 2018, and run a linear regression model, as shown in Figures 33 and 34, respectively. The R-squared value of 0.3719 shows a low correlation, but is not statistically insignificant. We must understand that a high R-squared value only implies strong prediction, but does not rule out that data is good, and vice versa. Consider the negligible p-value, which suggests that the null hypothesis is false, and that the alternative hypothesis is true [21]. This suggests that there is a significant correlation between the availability of scientists and engineers and GDP per capita of a country. However, as we have mentioned before, there are certain outliers to this correlation.

In general, it seems that more scientists and engineers means a faster modernization process within a country, as the talent required to develop technologies is readily available. Better technologies ensures efficiency in processes, resource gathering, transportation, internet, and other transportation means that allows a nation to create, build, and live. However, in countries like India where manufacturing is the prominent industry [22], the over-saturation of scientists and engineers does not contribute to significant economic growth. On the other hand, in China, where the economy is much more diverse, the relative lack of scientists and engineers does not significantly impede its economic growth. The correlation, therefore, exists because the availability of scientists and engineers remains an indicator of economic well-being, but is not applicable to all sociopolitical

climates and economic conditions, especially in developing countries without mature infrastructure.

Figure 33: Availability of Scientists and Engineers versus GDP per Capita Around the World in 2018

Figure 34: Statistical Correlation by Availability of Scientists and Engineers in 2018

7   Association of Gender Equality and Economic Growth

Does GDP growth vary with gender equality? In order to understand this issue, we pull data from TCdata360's database [2], and find that the most recent gender equality data is from 2019. With this in mind, we plot the correlation between GDP growth and gender equality in 2019, as we can see in Figure 35. In Figure 36, we can see that there is almost no correlation with the data. However, as suggested by the p-value, we can see a line of best fit would suggest strong GDP growth in country's with poor gender equality indices. In fact, the largest GDP growths occurred in countries with a low gender equality index.
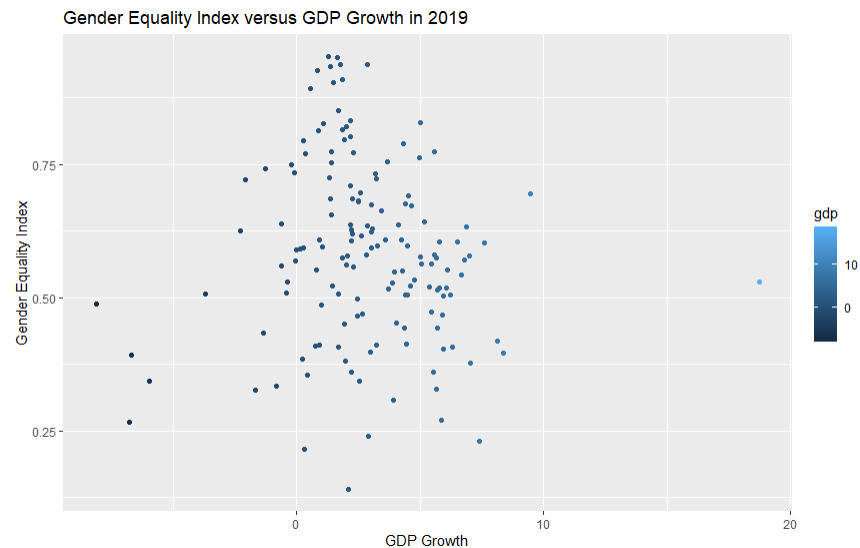
Figure 35: Gender Equality versus GDP Growth in 2019

Figure 36: Statistical Correlation by Gender Equality in 2019

Intuitively, we associate gender equality with the end goal in workplace politics. However, this is often a luxury only developed nations can afford. In order to delve into this problem further, we now plot the correlation between GDP per capita and gender equality in 2019, and analyze its statistical properties, as we see in Figures 37 and 38, respectively. Immediately, we see a cluster or high GDP per capita nations that also exhibit high gender equality indices, as evidenced by a lower p-value and higher R-squared correlation. While the statistical significance for this relationship is yet to exist, it is a stronger relationship than that portrayed by GDP growth. Therefore, we can assume that economies tend to grow more quickly in nation's with a low gender equality index, but developed economies tend to exhibit higher gender equality. This is similar to the chicken and egg problem, where gender equality tends to come after the economy is developed.
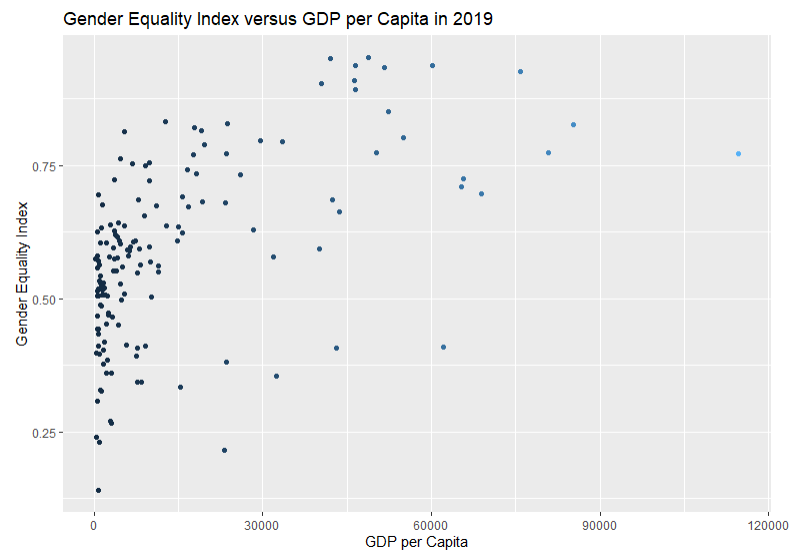
Figure 37: Gender Equality versus GDP per Capita in 2019

Figure 38: Statistical Correlation by Gender Equality in 2019

Regardless of the relationship between gender equality and GDP growth or GDP per capita, the top 5 countries that exhibited gender equality index growth between 2014-2019 were Canada, Singapore, Georgia, Azerbaijan, and Sri Lanka, as we can see in Figure 39.

	Country ISO3	Country Name	Indicator	Subindicator Type	growth
24	CAN	Canada	Gender equality	Scaled to range from 0 (lowest score) to 1 (highest score)	0.1689481537
128	SGP	Singapore	Gender equality	Scaled to range from 0 (lowest score) to 1 (highest score)	0.1564499760
52	GEO	Georgia	Gender equality	Scaled to range from 0 (lowest score) to 1 (highest score)	0.1448461449
9	AZE	Azerbaijan	Gender equality	Scaled to range from 0 (lowest score) to 1 (highest score)	0.1341970942
85	LKA	Sri Lanka	Gender equality	Scaled to range from 0 (lowest score) to 1 (highest score)	0.1340007903

Figure 39: Growth in Gender Equality Index between 2014-2019

8   References

[1] "International Money Transfer: Global Remittance with TranSwap," International Money Transfer — Global Remittance with TranSwap. [Online]. Available: https://www.transwap.com/blog/us-dollar-how-it-became-the-global-currency. [Accessed: 29-Jan-2022].

[2] "Innovation, Technology and Entrepreneurship: A data exploration," Open Trade and Competitiveness Data. [Online]. Available: https://tcdata360.worldbank.org/. [Accessed: 29-Jan-2022].

[3] "Northern Hemisphere," Wikipedia, 04-Jan-2022. [Online]. Available: https://en.wikipedia.org/wiki/Northern_Hemisphere#cite_note-8. [Accessed: 29-Jan-2022].

[4] K. Wilkie, "Why colder countries are richer than Hot Nations," Daily Mail Australia, 01-Jul-2020. [Online]. Available: https://www.dailymail.co.uk/news/article-8477297/Why-colder-countries-richer-hot-nations.html. [Accessed: 29-Jan-2022].

[5] L. Ngo, "How to compare box plots," BioTuring's Blog, 27-Sep-2020. [Online]. Available: https://blog.bioturing.com/2018/05/22/how-to-compare-box-plots/. [Accessed: 29-Jan-2022].

[6] "Internet crossing borders: Boosting the internet in landlocked developing countries," Growing the Internet, 06-Jan-2021. [Online]. Available: https://www.internetsociety.org/resources/doc/2017/lldcreport/. [Accessed: 29-Jan-2022].

[7] P. Fox, "The global digital divide (article)," Khan Academy. [Online]. Available: https://www.khanacademy.org/computing/computers-and-internet/xcae6f4a7ff015e7d:the-internet/xcae6f4a7ff015e7d:the-digital-divide/a/the-global-digital-divide. [Accessed: 29-Jan-2022].

[8] "Cambodian genocide," Wikipedia, 26-Jan-2022. [Online]. Available: https://en.wikipedia.org/wiki/Cambodian_genocide [Accessed: 29-Jan-2022].

[9] "Civil War," Encyclopædia Britannica. [Online]. Available: https://www.britannica.com/place/Cambodia/Civil-war. [Accessed: 29-Jan-2022].

[10] "Rwanda Profile - Timeline," BBC News, 17-Sep-2018. [Online]. Available: https://www.bbc.com/news/world-africa-14093322. [Accessed: 29-Jan-2022].

[11] K. Thelwell, "Rising life expectancy in Africa," The Borgen Project, 24-Apr-2019. [Online]. Available: https://borgenproject.org/rising-life-expectancy-in-africa/. [Accessed: 29-Jan-2022].

[12] R. Ratcliffe, "Middle classes drive up life expectancy in sub-saharan africa," The Guardian, 15-Sep-2018. [Online]. Available: https://www.theguardian.com/global-development/2018/sep/15/middle-classes-drive-up-life-expectancy-sub-saharan-africa-un-human-development-index-11-years. [Accessed: 29-Jan-2022].

[13] Z. Qiang, P. Kusek, V. Steenbergen, and B. Viney, "The road to recovery in Sub-Saharan africa: Capitalizing on transformative opportunities from shifting FDI patterns," World Bank Blogs, 27-May-2021. [Online]. Available: https://blogs.worldbank.org/africacan/road-recovery-sub-saharan-africa-capitalizing-transformative-opportunities-shifting-fdi. [Accessed: 29-Jan-2022].

[14] J. L. Dieleman, T. J. Bollyky, S. Crosby, and S. Kiernan, "Does healthier mean wealthier? measuring countries' economic performance during the pandemic: Think global health," Council on Foreign Relations, 14-Jan-2021. [Online]. Available: https://www.thinkglobalhealth.org/article/does-healthier-mean-wealthier-measuring-countries-economic-performance-during-pandemic. [Accessed: 29-Jan-2022].

[15] "Economic growth and life expectancy – do wealthier countries...," Euromonitor International, 14-Mar-2014. [Online]. Available: https://www.euromonitor.com/article/economic-growth-and-life-expectancy-do-wealthier-countries-live-longer. [Accessed: 29-Jan-2022].

[16] A. Mahmoud, Class Lecture, Topic: "Week 2 - Understanding Poverty" APS420/APS1420, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Jan., 2022.

[17] N. Dutta, R. S. Sobel, and S. Roy, "Entrepreneurship and political risk," Journal of Entrepreneurship and Public Policy, 2013. [Online]. Available: https://ideas.repec.org/a/eme/jepppp/v2y2013i2p130-143.html. [Accessed: 29-Jan-2022].

[18] C. Boyte-White, "How should a company be raising capital?," Investopedia, 03-Nov-2021. [Online]. Available: https://www.investopedia.com/ask/answers/032515/what-are-different-ways-corporations-can-raise-capital.asp. [Accessed: 29-Jan-2022].

[19] M. Riordan, "No monopoly on Innovation," Harvard Business Review, 01-Aug-2014. [Online]. Available: https://hbr.org/2005/12/no-monopoly-on-innovation [Accessed: 29-Jan-2022].

[20] History.com Editors, "Stock market crash of 1929," History.com, 10-May-2010. [Online]. Available: https://www.history.com/topics/great-depression/1929-stock-market-crash. [Accessed: 29-Jan-2022].

[21] S. A. McLeod, "What a p-value tells you about statistical significance.," Simply Psychology, 20-May-2019. [Online]. Available: www.simplypsychology.org/p-value.html. [Accessed: 29-Jan-2022].

[22] "Contribution by some Major Industries," Toppr. [Online]. Available: https://www.toppr.com/guides/fundamentals-of-economics-cma/indian-economy/contribution-by-some-major-industries/. [Accessed: 29-Jan-2022].

9   Appendix A

```
install.packages("tidyverse")
install.packages('abind')

library(abind)
library(data360r)
library(tidyverse)
library(reshape2)
library(rworldmap)
library(ggpmisc)
library(RColorBrewer)
library(gganimate)
library(gridExtra)
library(dplyr)

GDP_percapita <- get_data360(site = "tc",
                              indicator_id = 944,
                              dataset_id = NULL,
                              country_iso3 = NULL,
                              timeframes = c(2019), #comma for list,
                                  colon for range
                              output_type = "wide")

country_data <- get_metadata360(site = "tc",
                                  metadata_type = "countries")

merged <- merge(GDP_percapita, country_data,
                by.x = "Country_ISO3", by.y = "iso3")

names(merged)[names(merged) == '2019'] <- 'gdp'

ggplot(merged, aes(x = gdp, y = geo.lat)) +
  geom_point(aes(colour = incomeLevel)) +
  xlab("GDP per Capita") +
  ylab("Latitude") +
  ggtitle("Latitude versus GDP per Capita in 2019")

model_2019 = lm(gdp~abs(geo.lat), data=merged)

summary(model_2019)

merged %>%
  count(incomeLevel)

stem_avail <- get_data360(site = "tc",
```

```
                              indicator_id = 607,
                              dataset_id = NULL,
                              country_iso3 = NULL,
                              period = c(2017:2018),
                              output_type = "long")


cp <- brewer.pal(9,'Greens')
mapz <- joinCountryData2Map(stem_avail, joinCode = "ISO3",
    nameJoinColumn = "Country_ISO3", nameCountryColumn = "Country_Name")
mapParams <- mapCountryData(mapz, nameColumnToPlot="Observation",
                              addLegend=TRUE, mapTitle="Availability_of_
                                  Scientists_and_Engineers", mapRegion="
                                  World",
                              catMethod="quantiles", numCats=9,
                                  colourPalette=cp,
                              borderCol="black")


cp <- brewer.pal(9,'Greens')
mapz <- joinCountryData2Map(GDP_percapita, joinCode = "ISO3",
    nameJoinColumn = "Country_ISO3", nameCountryColumn = "Country_Name")
mapParams <- mapCountryData(mapz, nameColumnToPlot="2019",
                              addLegend=TRUE, mapTitle="GDP_per_capita",
                                  mapRegion="World",
                              catMethod="quantiles", numCats=9,
                                  colourPalette=cp,
                              borderCol="black")


GDP_percapita <- get_data360(site = "tc",
                              indicator_id = 944,
                              dataset_id = NULL,
                              country_iso3 = NULL,
                              timeframes = c(2018), #comma for list,
                                  colon for range
                              output_type = "wide")

stem_avail_gdp <- merge(GDP_percapita, stem_avail,
                  by.x = "Country_ISO3", by.y = "Country_ISO3")

stem_avail_gdp <- stem_avail_gdp[stem_avail_gdp$Period == '2017-2018',]

names(stem_avail_gdp)[names(stem_avail_gdp) == '2018'] <- 'gdp'

ggplot(stem_avail_gdp, aes(x = gdp, y = Observation)) +
  geom_point(aes(colour = gdp)) +
  xlab("GDP_per_Capita") +
  ylab("Availability_of_Scientists_and_Engineers") +
```

```
  ggtitle("Availability_of_Scientists_and_Engineers_versus_GDP_per_
      Capita_in_2017-2018")

model_stem = lm(gdp~Observation, data=stem_avail_gdp)

summary(model_stem)

gender_equal <- get_data360(site = "tc",
                            indicator_id = 41933,
                            dataset_id = NULL,
                            country_iso3 = NULL,
                            #timeframes = c(2019), #comma for list,
                                colon for range
                            output_type = "wide")

gender_equal_clean <- gender_equal[, c(1, 2, 3, 4, 49)]

GDP_growth <- get_data360(site = "tc",
                          indicator_id = 940,
                          dataset_id = NULL,
                          country_iso3 = NULL,
                          #timeframes = c(2019), #comma for list,
                              colon for range
                          output_type = "wide")

GDP_growth_clean <- GDP_growth[, c(1, 2, 3, 4, 63)]

gender_growth <- merge(GDP_growth_clean, gender_equal_clean,
                       by.x = "Country_ISO3", by.y = "Country_ISO3")

names(gender_growth)[names(gender_growth) == '2019.x'] <- 'gdp'
names(gender_growth)[names(gender_growth) == '2019.y'] <- 'gender'

ggplot(gender_growth, aes(x = gdp, y = gender)) +
  geom_point(aes(colour = gdp)) +
  xlab("GDP_Growth") +
  ylab("Gender_Equality_Index") +
  ggtitle("Gender_Equality_Index_versus_GDP_Growth_in_2019")

model_gender = lm(gdp~gender, data=gender_growth)

summary(model_gender)

GDP_percapita_7 <- get_data360(site = "tc",
                               indicator_id = 944,
                               dataset_id = NULL,
```

```
                                         country_iso3 = NULL,
                                         timeframes = c(2019), #comma for list,
                                             colon for range
                                         output_type = "wide")

gender_capita <- merge(GDP_percapita_7, gender_equal_clean,
                       by.x = "Country_ISO3", by.y = "Country_ISO3")

names(gender_capita)[names(gender_capita) == '2019.x'] <- 'gdp'
names(gender_capita)[names(gender_capita) == '2019.y'] <- 'gender'

ggplot(gender_capita, aes(x = gdp, y = gender)) +
  geom_point(aes(colour = gdp)) +
  xlab("GDP_per_Capita") +
  ylab("Gender_Equality_Index") +
  ggtitle("Gender_Equality_Index_versus_GDP_per_Capita_in_2019")

model_gender_capita = lm(gdp~gender, data=gender_capita)

summary(model_gender_capita)

gender_equal$growth <- (gender_equal$`2019` - gender_equal$`2014`)

gender_equal_growth <- gender_equal[, c(1, 2, 3, 4, 50)]

GII <- get_data360(site = "tc",
                                         indicator_id = 40712,
                                         dataset_id = NULL,
                                         country_iso3 = NULL,
                                         #timeframes = c(2019), #comma for
                                             list, colon for range
                                         output_type = "wide")

GII_2015 <- GII[, c(1, 2, 3, 4, 7)]
GII_2020 <- GII[, c(1, 2, 3, 4, 12)]

political_stability <- get_data360(site = "tc",
                               indicator_id = 40266,
                               dataset_id = NULL,
                               country_iso3 = NULL,
                               #timeframes = c(2019), #comma for list,
                                   colon for range
                               output_type = "wide")

merge_political <- merge(GII_2015, political_stability,
                       by.x = "Country_ISO3", by.y = "Country_ISO3")
```

```r
names(merge_political)[names(merge_political) == '2015.x'] <- 'GII'
names(merge_political)[names(merge_political) == '2015.y'] <- '
   political'

ggplot(merge_political, aes(x = GII, y = political)) +
  geom_point(aes(colour = GII)) +
  xlab("GII Index") +
  ylab("Political Stability") +
  ggtitle("Political Stability versus GII Index in 2015")

model_political = lm(GII~political, data=merge_political)

summary(model_political)

institutions <- get_data360(site = "tc",
                                indicator_id = 40262,
                                dataset_id = NULL,
                                country_iso3 = NULL,
                                #timeframes = c(2019), #comma for
                                    list, colon for range
                                output_type = "wide")

institutions_clean <- institutions[, c(1, 2, 3, 4, 12)]

merge_institutions <- merge(GII_2020, institutions_clean,
                   by.x = "Country ISO3", by.y = "Country ISO3")

names(merge_institutions)[names(merge_institutions) == '2020.x'] <- '
   GII'
names(merge_institutions)[names(merge_institutions) == '2020.y'] <- '
   institutions'

ggplot(merge_institutions, aes(x = GII, y = institutions)) +
  geom_point(aes(colour = GII)) +
  xlab("GII Index") +
  ylab("Institutions") +
  ggtitle("Institutions versus GII Index in 2020")

model_institutions = lm(GII~institutions, data=merge_institutions)

summary(model_institutions)

credit <- get_data360(site = "tc",
                                indicator_id = 40418,
                                dataset_id = NULL,
```

```
                                        country_iso3 = NULL,
                                        #timeframes = c(2019), #comma for
                                            list, colon for range
                                        output_type = "wide")

credit_clean <- credit[, c(1, 2, 3, 4, 12)]

merge_credit <- merge(GII_2020, credit_clean,
                        by.x = "Country_ISO3", by.y = "Country_ISO3")

names(merge_credit)[names(merge_credit) == '2020.x'] <- 'GII'
names(merge_credit)[names(merge_credit) == '2020.y'] <- 'credit'

ggplot(merge_credit, aes(x = GII, y = credit)) +
  geom_point(aes(colour = GII)) +
  xlab("GII_Index") +
  ylab("Credit") +
  ggtitle("Ease_of_Getting_Credit_versus_GII_Index_in_2020")

model_credit = lm(GII~credit, data=merge_credit)

summary(model_credit)

entrepreneurship <- get_data360(site = "tc",
                                        indicator_id = 3111,
                                        dataset_id = NULL,
                                        country_iso3 = NULL,
                                        #timeframes = c(2019), #comma for
                                            list, colon for range
                                        output_type = "wide")

entrepreneurship_clean <- entrepreneurship[, c(1, 2, 3, 4, 22)]

merge_entrepreneurship <- merge(GII_2020, entrepreneurship_clean,
                        by.x = "Country_ISO3", by.y = "Country_ISO3")

names(merge_entrepreneurship)[names(merge_entrepreneurship) == '2020.x'
    ] <- 'GII'
names(merge_entrepreneurship)[names(merge_entrepreneurship) == '2020.y'
    ] <- 'entrepreneurship'

ggplot(merge_entrepreneurship, aes(x = GII, y = entrepreneurship)) +
  geom_point(aes(colour = GII)) +
  xlab("GII_Index") +
  ylab("Entrepreneurship") +
  ggtitle("Entrepreneurship_as_a_Desirable_Career_Choice_versus_GII_
```

```
    Index _in _2020")

model _entrepreneurship = lm(GII~entrepreneurship , data=merge_
    entrepreneurship)

summary(model_entrepreneurship)

institutions _clean _GDP <- institutions [ , c(1, 2, 3, 4, 10)]

merge _institutions _GDP <- merge(GDP_percapita , institutions _clean _GDP,
                        by.x = "Country _ISO3", by.y = "Country _ISO3
                          ")

names(merge_institutions _GDP) [names(merge_institutions _GDP) == '2018.x'
    ] <- 'gdp'
names(merge_institutions _GDP) [names(merge_institutions _GDP) == '2018.y'
    ] <- 'institutions '

ggplot(merge_institutions _GDP, aes(x = gdp, y = institutions )) +
  geom _point(aes(colour = gdp)) +
  xlab("GDP _per _Capita") +
  ylab("Institutions") +
  ggtitle("Institutions _versus _GDP_per _Capita _in _2018")

model _institutions _GDP = lm(gdp~institutions , data=merge_institutions _
    GDP)

summary(model_institutions _GDP)

merge _GII _GDP <- merge(GDP_percapita , GII ,
                            by.x = "Country _ISO3", by.y = "Country _
                              ISO3")

names(merge_GII _GDP) [names(merge_GII _GDP) == '2018.x'] <- 'gdp'
names(merge_GII _GDP) [names(merge_GII _GDP) == '2018.y'] <- 'gii'

ggplot(merge_GII _GDP, aes(x = gdp, y = gii )) +
  geom _point(aes(colour = gdp)) +
  xlab("GDP _per _Capita") +
  ylab("GII _Index") +
  ggtitle("GII _Index _versus _GDP_per _Capita _in _2018")

model _GII _GDP = lm(gdp~gii , data=merge_GII _GDP)

summary(model_GII _GDP)
```

```r
life_expectancy_male <- get_data360(site = "tc",
                                    indicator_id = 837,
                                    dataset_id = NULL,
                                    country_iso3 = NULL,
                                    #timeframes = c(1960:2018), #comma for
                                        list, colon for range
                                    output_type = "wide")

life_expectancy_female <- get_data360(site = "tc",
                                      indicator_id = 836,
                                      dataset_id = NULL,
                                      country_iso3 = NULL,
                                      #timeframes = c(1960:2018), #comma
                                          for list, colon for range
                                      output_type = "wide")

life_expectancy_male <- life_expectancy_male[,-c(1:4) ]
life_expectancy_female <- life_expectancy_female[,-c(1:4) ]

life_expectancy_average <- (life_expectancy_male+life_expectancy_female
    )/2
life_expectancy_average = cbind(life_expectancy_average, Max_Difference
    = 0)
life_expectancy_average = cbind(life_expectancy_average, Max = 0)
life_expectancy_average = cbind(life_expectancy_average, Min_Difference
    = 0)
life_expectancy_average = cbind(life_expectancy_average, Min = 0)

life_expectancy_average <- na.omit(life_expectancy_average)

life_expectancy_average <- data.matrix(life_expectancy_average,
    rownames.force = NA)
for(i in 1:187) {
  life_expectancy_average[i, 61] <- max(diff(life_expectancy_average[i
      ,1:60]))
  life_expectancy_average[i, 62] <- which.max(diff(life_expectancy_
      average[i,1:60])) + 2
  life_expectancy_average[i, 63] <- min(diff(life_expectancy_average[i
      ,1:60]))
  life_expectancy_average[i, 64] <- which.min(diff(life_expectancy_
      average[i,1:60]))+ 2
}

life_expectancy_average <- as.data.frame(life_expectancy_average)

life_expectancy_average <- cbind(a = 0, life_expectancy_average)
```

```r
life_expectancy_average <- cbind(b = 0, life_expectancy_average)

life_expectancy_male <- get_data360(site = "tc",
                                    indicator_id = 837,
                                    dataset_id = NULL,
                                    country_iso3 = NULL,
                                    #timeframes = c(1960:2018), #comma
                                        for list, colon for range
                                    output_type = "wide")

life_expectancy_male <- na.omit(life_expectancy_male)

life_expectancy_average$b = life_expectancy_male$'Country ISO3'
life_expectancy_average$a = life_expectancy_male$'Country Name'
names(life_expectancy_average)[names(life_expectancy_average) == 'b']
    <- 'Country_ISO3'
names(life_expectancy_average)[names(life_expectancy_average) == 'a']
    <- 'Country_Name'

life_expectancy_melt <- life_expectancy_average[,-c(1:2) ]
life_expectancy_melt <- life_expectancy_melt[,-c(61:64) ]
life_expectancy_melt <- as.data.frame(t(as.matrix(life_expectancy_melt)
    ))
life_expectancy_melt <- life_expectancy_melt[c(92,146)]
life_expectancy_melt$Year <- 1960:2019

life_expectancy_melt <- data.frame(x = life_expectancy_melt$Year,
                                   y = c(life_expectancy_melt$V92, life_
                                       expectancy_melt$V146),
                                   Country = c(rep("Cambodia", nrow(life_
                                       expectancy_melt)),
                                            rep("Rwanda", nrow(life_expectancy_
                                                melt)))))

ggplot(life_expectancy_melt, aes(x, y, col = Country)) +
  geom_line() +
  xlab("Year") +
  ylab("Life_Expectancy_at_Birth") +
  ggtitle("Life_Expectancy_at_Birth_in_Rwanda_and_Cambodia_Between_1960
      _and_2019")

life_expectancy_by_region <- merge(life_expectancy_average, country_
    data, by.x = 'Country_ISO3', by.y = 'iso3')
life_expectancy_by_region <- life_expectancy_by_region[,-c(63:70) ]
life_expectancy_by_region <- life_expectancy_by_region[,-c(64:68) ]
life_expectancy_by_region <- life_expectancy_by_region[,-c(1:2) ]
```

```r
life_expectancy_by_region[nrow(life_expectancy_by_region) + 1,] = c(0)

for(i in 1:60) {
  life_expectancy_by_region[188, i] <- mean(life_expectancy_by_region
      [1:187, i])
}

life_expectancy_by_region[188,61] = 'World'
life_expectancy_by_region <- na.omit(life_expectancy_by_region)
life_expectancy_by_region<-life_expectancy_by_region[(life_expectancy_
    by_region$adminRegion =="World" | life_expectancy_by_region$
    adminRegion == "SSA"),]

life_expectancy_by_region[nrow(life_expectancy_by_region) + 1,] = c(0)

for(i in 1:60) {
  life_expectancy_by_region[48, i] <- mean(life_expectancy_by_region
      [1:47, i])
}

life_expectancy_by_region[48,61] = 'SSA'
life_expectancy_by_region <- life_expectancy_by_region[-c(1:46),]

life_expectancy_melt <- as.data.frame(t(as.matrix(life_expectancy_by_
    region)))
life_expectancy_melt$Year <- 1960:2020
life_expectancy_melt[61,1] = 'World'
life_expectancy_melt[61,2] = 'SSA'
names(life_expectancy_melt)[names(life_expectancy_melt) == '188'] <- '
    World'
names(life_expectancy_melt)[names(life_expectancy_melt) == '48'] <- '
    SSA'
life_expectancy_melt <- life_expectancy_melt[-c(61),]

life_expectancy_melt <- data.frame(x = life_expectancy_melt$Year,
                                    y = c(life_expectancy_melt$SSA, life
                                        _expectancy_melt$World),
                                    Country = c(rep("Africa", nrow(life_
                                        expectancy_melt)),
                                              rep("World", nrow(life_
                                                  expectancy_melt))))

life_expectancy_melt$y <- as.numeric(life_expectancy_melt$y)
```

```
ggplot ( life_expectancy_melt , aes (x, y, col = Country )) +
  geom_line () +
  xlab (" Year ") +
  ylab (" Life_Expectancy_at_Birth ") +
  ggtitle (" Life_Expectancy_at_Birth_in_Sub-Saharan_Africa_and_the_World
      _Between_1960_and_2019 ")

life_expectancy_average<-life_expectancy_average [( life_expectancy_
    average$'Country_ISO3' =="BGD" |
                                                 life_expectancy_
                                                     average$'Country
                                                     _ISO3' =="KOR" |
                                                 life_expectancy_
                                                     average$'Country
                                                     _ISO3' =="MYS")
                                                     ,]

life_expectancy_average <- life_expectancy_average [,-c (1:2 ,63:66) ]

life_expectancy_average <- as.data.frame (t(as.matrix ( life_expectancy_
    average )))
names( life_expectancy_average )[names( life_expectancy_average ) == '16']
    <- 'BGD'
names( life_expectancy_average )[names( life_expectancy_average ) == '94']
    <- 'KOR'
names( life_expectancy_average )[names( life_expectancy_average ) == '121']
    <- 'MYS'

life_expectancy_average$Year <- 1960:2019

life_expectancy_average <- data.frame (x = life_expectancy_average$Year ,
                                 y = c( life_expectancy_average$BGD,
                                     life_expectancy_average$KOR, life
                                     _expectancy_average$MYS),
                                 Country = c(rep(" Bangladesh ", nrow(
                                     life_expectancy_average )),
                                         rep(" Korea ", nrow( life_
                                             expectancy_average )),
                                         rep(" Malaysia ", nrow(
                                             life_expectancy_
                                             average ))))

life_expectancy_average$y <- as.numeric ( life_expectancy_average$y)


ggplot ( life_expectancy_average , aes (x, y, col = Country )) +
```

```
  geom_line() +
  xlab("Year") +
  ylab("Life_Expectancy_at_Birth") +
  ggtitle("Life_Expectancy_at_Birth_in_Bangladesh,_Korea,_and_Malaysia_
      Between_1960_and_2019")

GDP_percapita_4 <- get_data360(site = "tc",
                               indicator_id = 944,
                               dataset_id = NULL,
                               country_iso3 = c('BGD', 'KOR', 'MYS'),
                               timeframes = c(1960:2019), #comma for
                                   list, colon for range
                               output_type = "wide")

GDP_percapita_4 <- GDP_percapita_4[,-c(1:4) ]

GDP_percapita_4 <- as.data.frame(t(as.matrix(GDP_percapita_4)))
names(GDP_percapita_4)[names(GDP_percapita_4) == 'V1'] <- 'BGD'
names(GDP_percapita_4)[names(GDP_percapita_4) == 'V2'] <- 'KOR'
names(GDP_percapita_4)[names(GDP_percapita_4) == 'V3'] <- 'MYS'

GDP_percapita_4$Year <- 1960:2019

GDP_percapita_4 <- data.frame(x = GDP_percapita_4$Year,
                              y = c(GDP_percapita_4$BGD, GDP_
                                  percapita_4$KOR, GDP_percapita
                                  _4$MYS),
                              Country = c(rep("Bangladesh",
                                  nrow(GDP_percapita_4)),
                                          rep("Korea", nrow(GDP
                                              _percapita_4)),
                                          rep("Malaysia", nrow(
                                              GDP_percapita_4)))
                                          )

GDP_percapita_4$y <- as.numeric(GDP_percapita_4$y)


ggplot(GDP_percapita_4, aes(x, y, col = Country)) +
  geom_line() +
  xlab("Year") +
  ylab("GDP_per_Capita") +
  ggtitle("GDP_per_Capita_in_Bangladesh,_Korea,_and_Malaysia_Between_
      1960_and_2019")

internet <- get_data360(site = "tc",
```

```
                              indicator_id = 591,
                              dataset_id = NULL,
                              country_iso3 = NULL,
                              #timeframes = c(2017), #comma for list, colon
                                  for range
                              output_type = "wide")

#fetch 2017-2018 data (most recent)
internet_clean <- internet[, c(1, 2, 3, 4, 15)]

landlocked = read.csv("landlocked_dataset.csv")  # read csv file

names(internet_clean)[names(internet_clean) == 'Country_ISO3'] <- 'iso3
   '

landlocked_internet = merge(x=internet_clean,y=landlocked,by="iso3",all
   =TRUE)

landlocked_internet <- landlocked_internet[, -c(4,6,7)]

landlocked_internet <- na.omit(landlocked_internet)

landlocked_internet_group <- landlocked_internet %>%
  group_by(landlocked)

ggplot(landlocked_internet_group, aes(y= `2017-2018`)) +
  geom_boxplot(aes(landlocked)) +
  xlab("LLDC_or_Non-LLDC") +
  ylab("Fixed_Broadband_Internet_Subscriptions") +
  ggtitle("Fixed_Broadband_Internet_Subscriptions_in_LLDCs_and_Non-
      LLDCs_Between_2017_and_2018")

internet <- get_data360(site = "tc",
                              indicator_id = 591,
                              dataset_id = NULL,
                              country_iso3 = NULL,
                              #timeframes = c(2017), #comma for list, colon
                                  for range
                              output_type = "wide")

landlocked = read.csv("landlocked_dataset.csv")  # read csv file

names(internet)[names(internet) == 'Country_ISO3'] <- 'iso3'

landlocked_internet = merge(x=internet,y=landlocked,by="iso3",all=TRUE)
```

```r
landlocked_internet <- landlocked_internet[,-c(4,6,7)]

landlocked_internet <- na.omit(landlocked_internet)
landlocked_internet <- as.data.frame((as.matrix(landlocked_internet)))

landlocked_internet[nrow(landlocked_internet) + 1,] = c(0)
landlocked_internet[nrow(landlocked_internet) + 1,] = c(0)

count_landlocked = 0
count_other = 0
sum_landlocked = 0
sum_other = 0

for(i in 4:12) {
  for(j in 1:102) {
    if(landlocked_internet[j,15] == 'FALSE') {
      sum_other = sum_other + as.numeric(landlocked_internet[j,i])
      count_other = count_other + 1
    }
    else {
      sum_landlocked = sum_landlocked + as.numeric(landlocked_internet[
          j,i])
      count_landlocked = count_landlocked + 1
    }
  }
  landlocked_internet[103, i] = sum_landlocked/count_landlocked
  landlocked_internet[104, i] = sum_other/count_other
  sum_landlocked = 0
  count_landlocked = 0
  sum_other = 0
  count_other = 0
}

landlocked_internet[103,3] <- 'Landlocked'
landlocked_internet[104,3] <- 'Non-Landlocked'

landlocked_internet <- landlocked_internet[,-c(1:2,13:15)]
landlocked_internet <- landlocked_internet[-c(1:102),]

landlocked_internet <- as.data.frame(t(as.matrix(landlocked_internet)))
landlocked_internet$Year <- 2008:2017
landlocked_internet[1,3] <- 2007
names(landlocked_internet)[names(landlocked_internet) == '103'] <- '
    Landlocked'
names(landlocked_internet)[names(landlocked_internet) == '104'] <- 'Non
    -Landlocked'
```

```
landlocked_internet <- landlocked_internet[-c(1),]

landlocked_internet <- data.frame(x = landlocked_internet$Year,
                                  y = c(landlocked_internet$'
                                      Landlocked', landlocked_internet$
                                      'Non-Landlocked'),
                                  Region = c(rep("Landlocked", nrow(
                                      landlocked_internet)),
                                                  rep("Non-Landlocked",
                                                      nrow(landlocked_
                                                      internet)))))

landlocked_internet$y <- as.numeric(landlocked_internet$y)


ggplot(landlocked_internet, aes(x, y, col = Region)) +
  geom_line() +
  xlab("Year") +
  ylab("Fixed_Broadband_Internet_Subscriptions") +
  ggtitle("Fixed_Broadband_Internet_Subscriptions_in_LLDC's_and_Non-
      LLDC's_Between_2007_and_2018")
```