Analyzing the Relationship Between National Forest Coverage, Access to Electricity, and GDP Growth Rate

Irti Haq and Charlie Norgaard

Table of Contents

1. Summary of Research Questions and Results

What is the relationship between the change in national forest coverage (% land area) and the change in national GDP?

- Our findings suggest that there tends to be a **negative correlation between national GDP and forest coverage** for most (Less Economically Developed Countries) LEDCs. That is, as national GDP goes up, there tends to be a decrease in forest coverage.
- On the other hand, there appears to be very little correlation between the change in national forest coverage and the change in national GDP.

What is the relationship between the change in national electricity access (% population) and the change in national GDP?

- Our findings suggest that there tends to be a **strong positive correlation between national GDP and the percentage of population with access to electricity**. That is, the national GDP tends to be higher in nations where more citizens have access to electricity.
- On the other hand, we found that there is very little correlation between the change in population access to electricity and the change in GDP.

What is the relationship between the change in national electricity access and the change in national forest coverage?

- Our findings suggest that there tends to be **no correlation between the change in the percentage of the population with access to electricity and the change in national forest coverage**. That is, there is no relationship between electricity access and forest coverage.

2. Motivation

Access to high quality reliable economic data tracking economic growth and development in developing nations remains a major challenge for policy makers and development agencies. Since access to electricity as a ratio of population and deforestation can act as a rough proxy for economic growth and development. Our plan is to show how we can use data of access to electricity and deforestation as a proxy to estimate economic growth and development in countries. Furthemore, by exploring trends pertaining to national urbanization and industrialization, we can better understand the factors that contribute to high or low national GDPs.

3. Datasets

- National GDP (current US$) (from the World Bank)
    - https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
    - This dataset contains the GDP for each country from 1999 - 2019. The GDP is adjusted for inflation and is based on the current US dollar. The data can be downloaded as a CSV file.
- Forest area (% of land area) (from UN SDG Data Hub)
    - **https://data.worldbank.org/indicator/AG.LND.FRST.ZS**
    - This dataset has the forest area as a proportion of total land area for each country from 1999 - 2019, it also has a shape file for each country. We can download the data as a shapefile.
- Access to Electricity (% of Population) (from UN SDG Data Hub)
    - **https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS**
    - This dataset has the Proportion of population with access to electricity for each country from 1999 - 2019, it also has a shape file for each country. We can download the data as a shapefile.
- World Administrative Boundaries (Shapefile)
    - **https://datacatalog.worldbank.org/search/dataset/0038272**
    - This is another shape file that can be used in conjunction with all three datasets

4. Methods

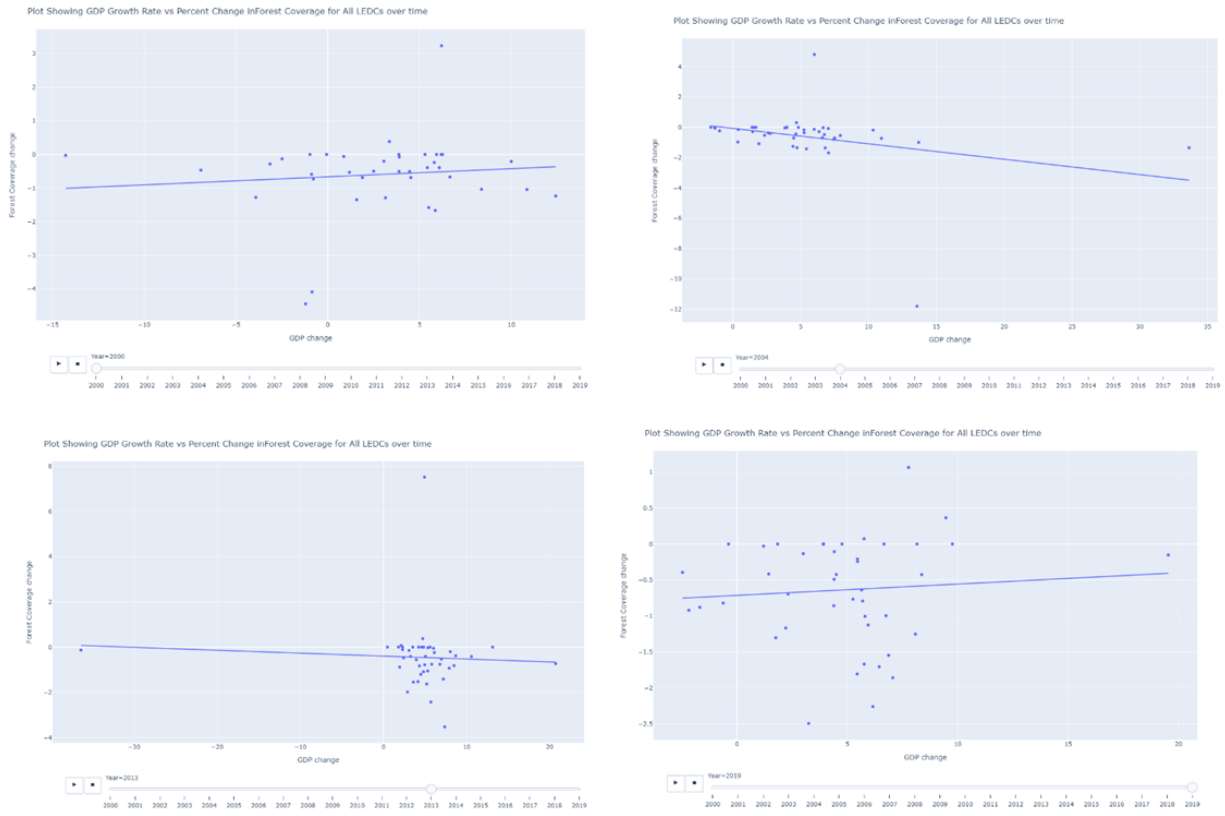
1. We started by filtering all datasets so that they encompassed the shared years (1999-2019), removed all countries that did not fall within the UN's list of underdeveloped countries, and cleaned up the dataset to remove any missing data. In this step, we also found the ISO codes for each country in order to have consistent naming across all three datasets.
2. Then, using Pandas, we calculated variable growth rates for each country by year.
    a. For the GDP dataset, we subtracted the GDP for that year by the GDP for the previous year and divided that by the previous year's GDP.
    b. For the forest coverage dataset, we subtracted the forest coverage for that year by the forest coverage for the previous year and divided that by the previous year's forest coverage.

    c. For the population access to electricity dataset, we subtracted the electricity access percentage for that year by the electricity access percentage for the previous year and divided that by the previous year's electricity access percentage.
3. We then joined the following datasets by country and year:
    a. GDP dataset merged with forest coverage dataset
    b. GDP dataset merged with population access to electricity dataset
    c. Forest coverage dataset merged with population access to electricity dataset
4. Using Plotly, we then created two scatter plots for each of the merged datasets. The plots included sliders to allow us to interactively view changes in the data and relationship over time
    a. For the GDP and Forest Coverage dataset, we plotted GDP on the x-axis and forest coverage on the y-axis. The second plot used the same axes, but instead we plotted the % change for each of the variables.
    b. For the GDP and Electricity Access dataset, we plotted GDP on the x-axis and electricity access on the y-axis. The second plot used the same axes, but instead we plotted the % change for each of the variables.
    c. For the Forest Coverage and Electricity Access dataset, we plotted electricity access on the x-axis and forest coverage on the y-axis. The second plot used the same axes, but instead we plotted the % change for each of the variables.
5. Next, we used the built in plotly function to plot a linear trend line and calculate the R^2 value for every plot
    a. The R^2 tells us how strongly the two variables are correlated and allow us to determine whether the rate of deforestation can be used as a proxy for economic growth
6. We then created an R^2 plot for each of the scatter plots mentioned above, which allowed us visualize the R^2 for each of the measured relationships by country
7. Lastly, we calculated the net change for each of the variables in question for each merged dataset and plotted these values by country on an interactive map using Geopandas and the shapefile

5. Results

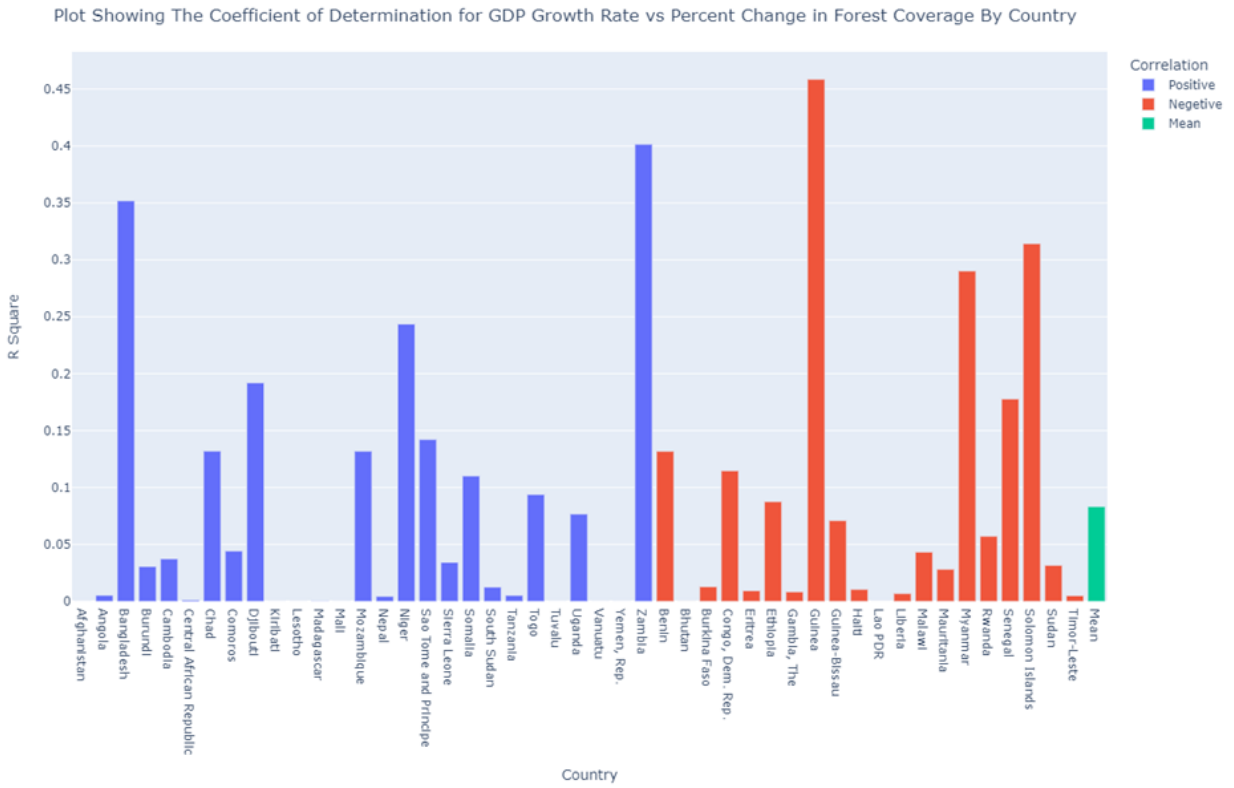
What is the relationship between the change in national forest coverage (% land area) and the change in national GDP?

Initially when trying to answer this question we started by looking at the GDP growth rate versus change in the forecast coverage and we got the following plots.



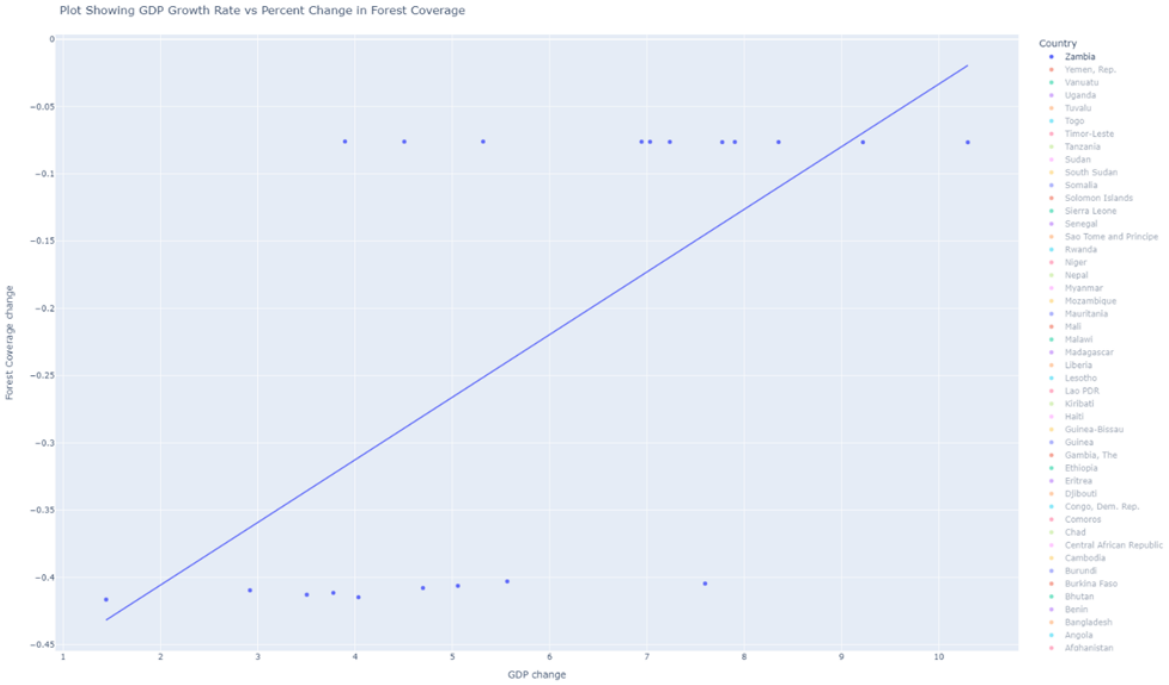
For these plots it looked like when comparing different countries against each other, the relationship between GDP growth and forest coverage seemed completely random. Some years there appeared a very weak positive correlation and other years it appeared like a weak negative correlation or even no correlation at all. Overall, the correlation between the two variables seemed non-existent or random for the most part.

We then decided to look at the relationship between GDP growth and Forest Coverage Change for each country over time. To do this we first started by calculating and plotting the R Squared Correlation Coefficient for each LEDC in the following Bar Chart to get an idea of the overall picture and then using that to examine the relationship more closely for individual countries where there might be a stronger than average correlation. These plots can be seen below.

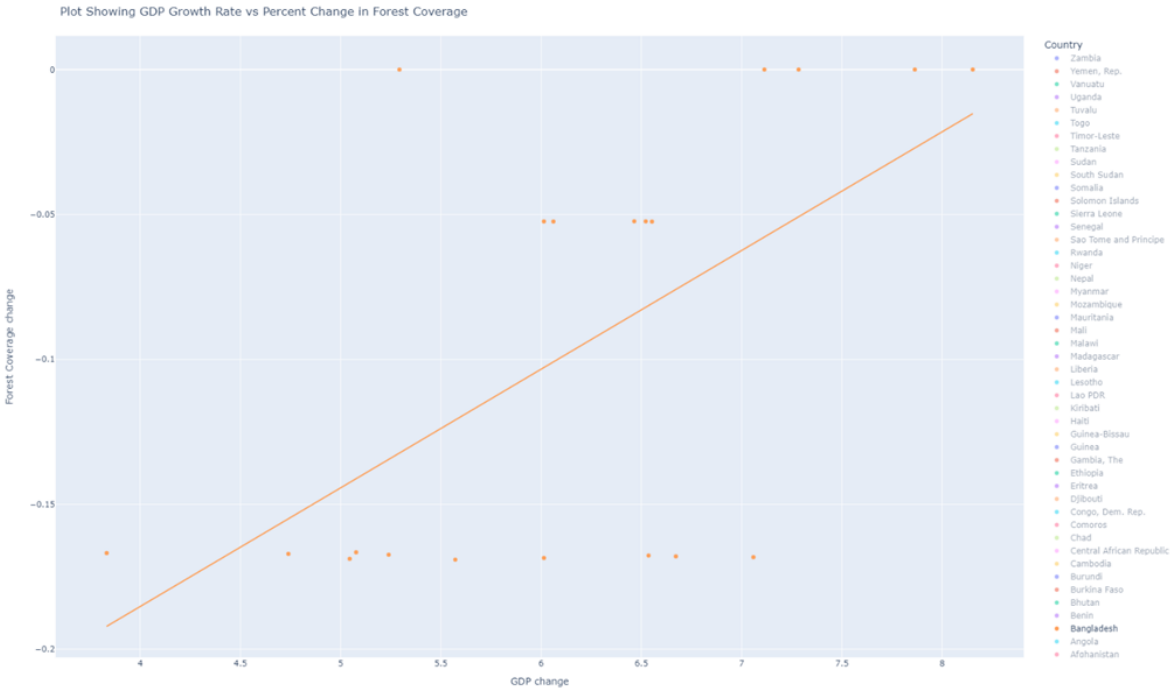


As you can see from this plot is that even for individual countries over time the relationship was extremely weak and almost random with some countries having a positive correlation while others have a negative correlation. On average the R Squared value was very weak at 0.08. And even for countries like Guinea (0.46), Zambia (0.40) and Bangladesh (0.35) who had a relatively medium coefficient the relationship was still extremely weak and sporadic when we looked at the individual plots and looked more like a mathematically anomaly resulting from the way the R Squared value is calculated rather than an indication of any sort of real relationship for Bangladesh and Zambia

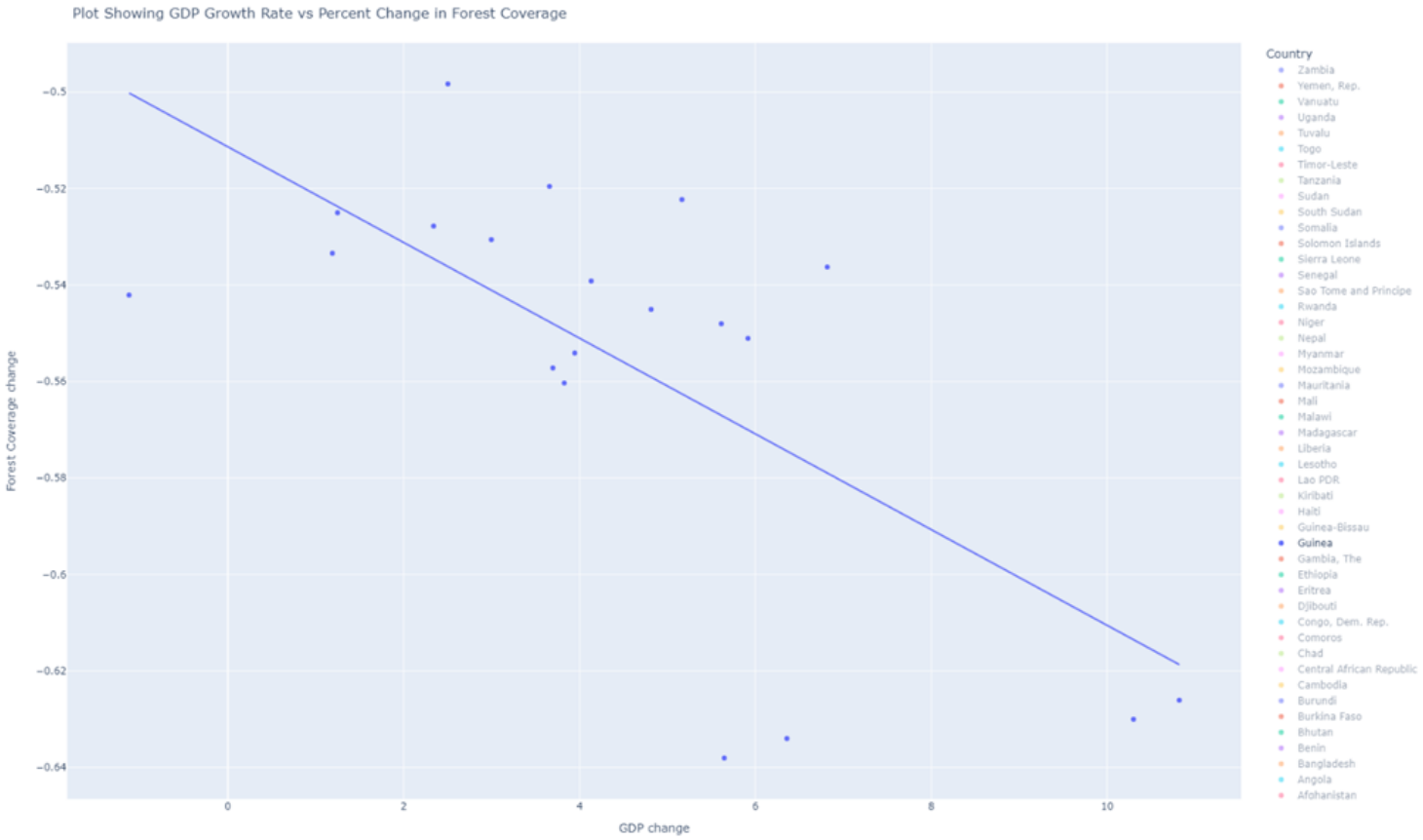
Zambia



Bangladesh



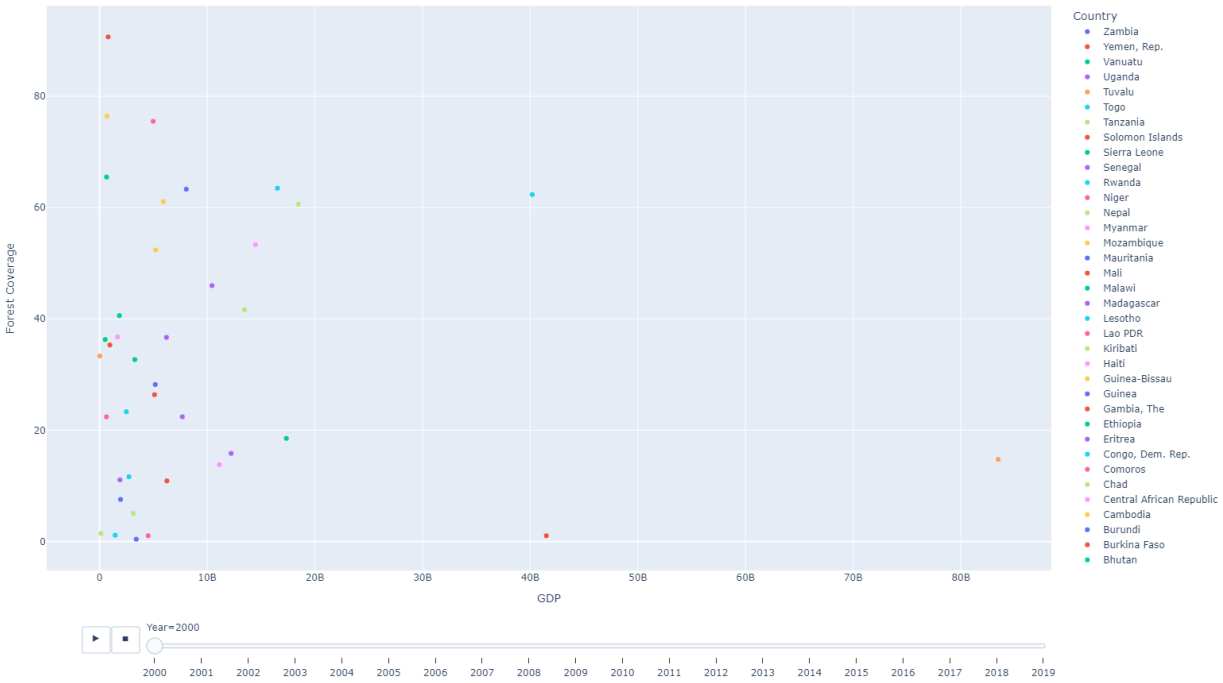
Guinea



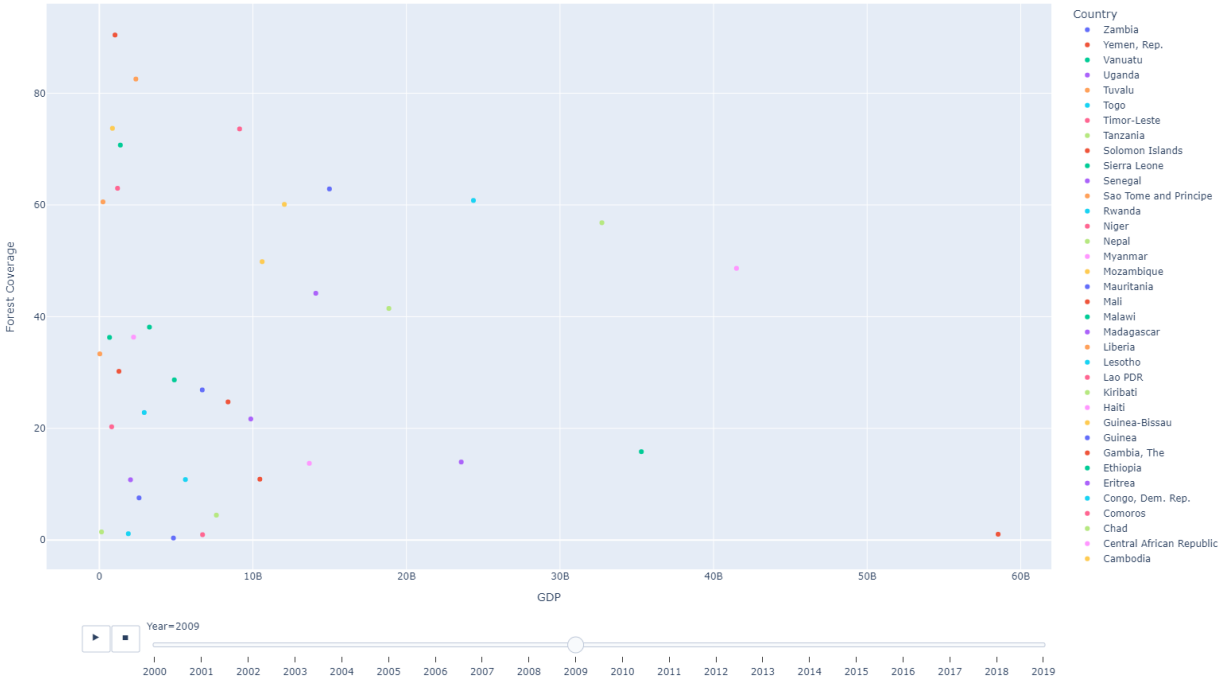
Guinea was the only nation that had somewhat of a plausible relationship

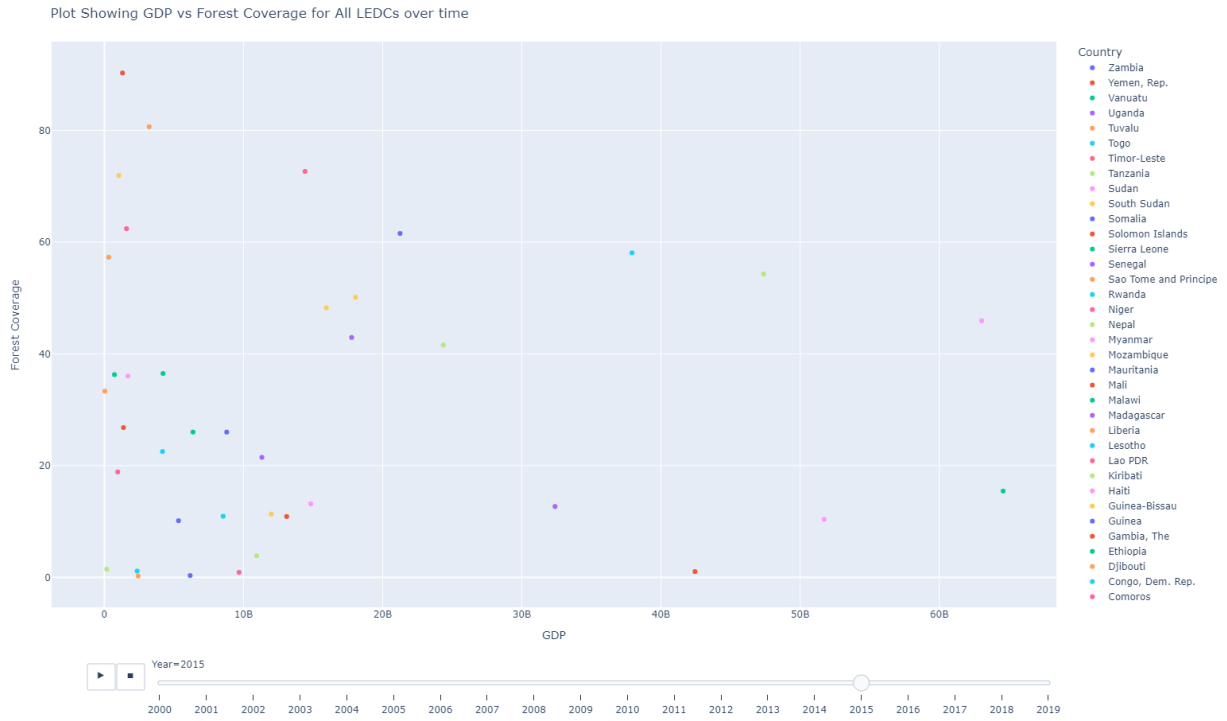
All of this made it seem that there appeared to be no relationship between GDP and Forest coverage however while looking at the Data we notice that when we plotted GDP against Forest coverage for all countries and then moved through the different years using the slider the data points representing individual countries seemed to be moving down to the left. This can be seen in the plot below

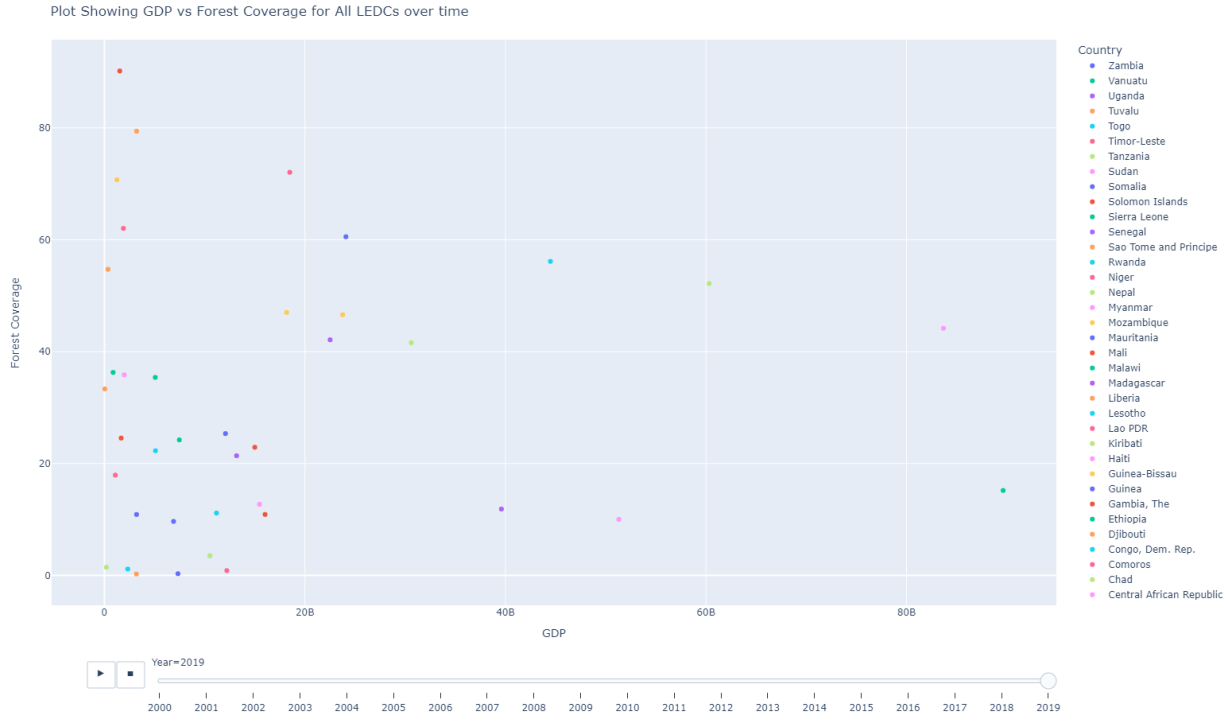
Plot Showing GDP vs Forest Coverage for All LEDCs over time



Plot Showing GDP vs Forest Coverage for All LEDCs over time



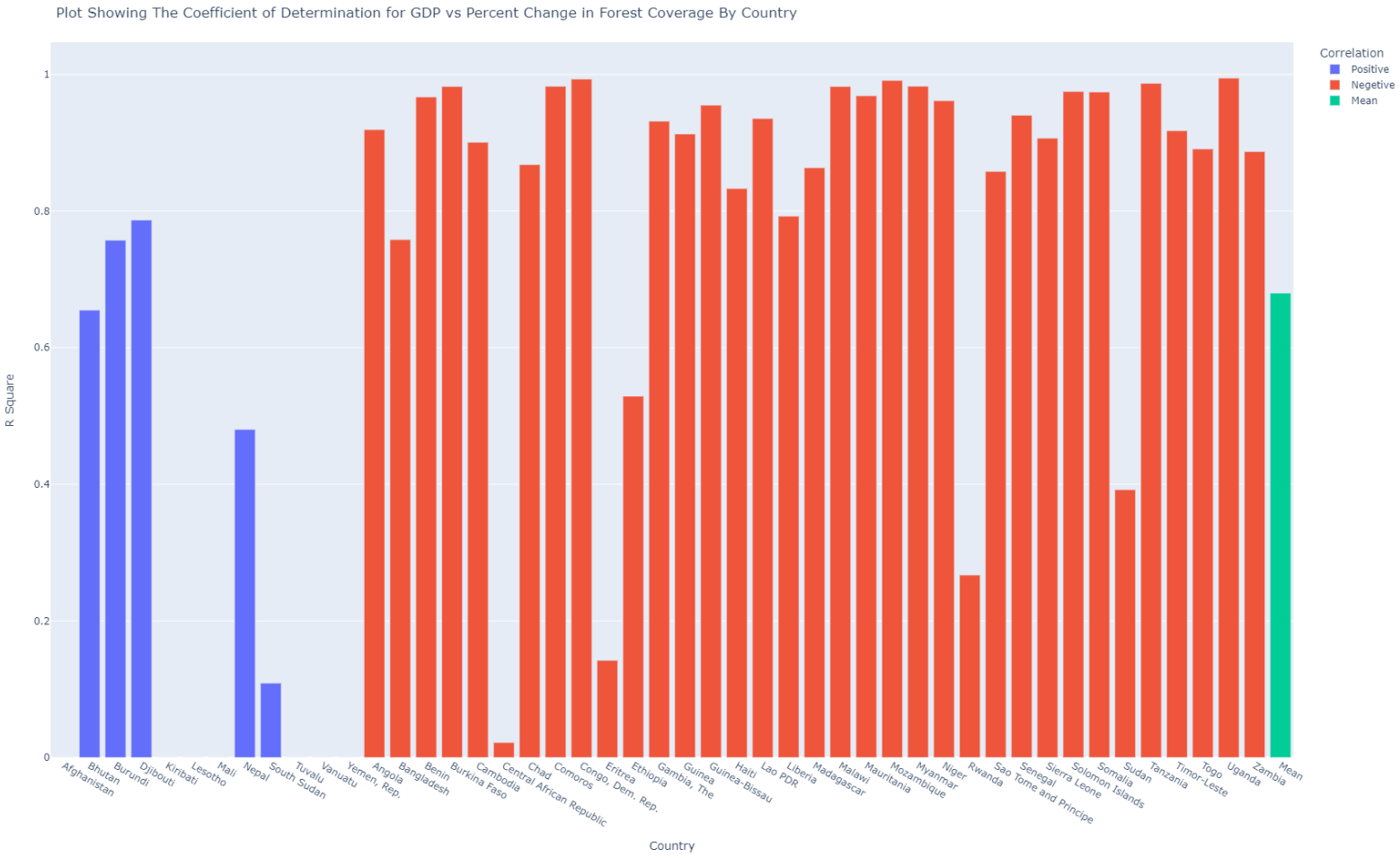




To investigate this further we once again decided to look at the relationship overtime for each individual country and we again calculated and plotted the R2 Value for each country. This time we found a much stronger relationship. As you can see below most countries have a very strong

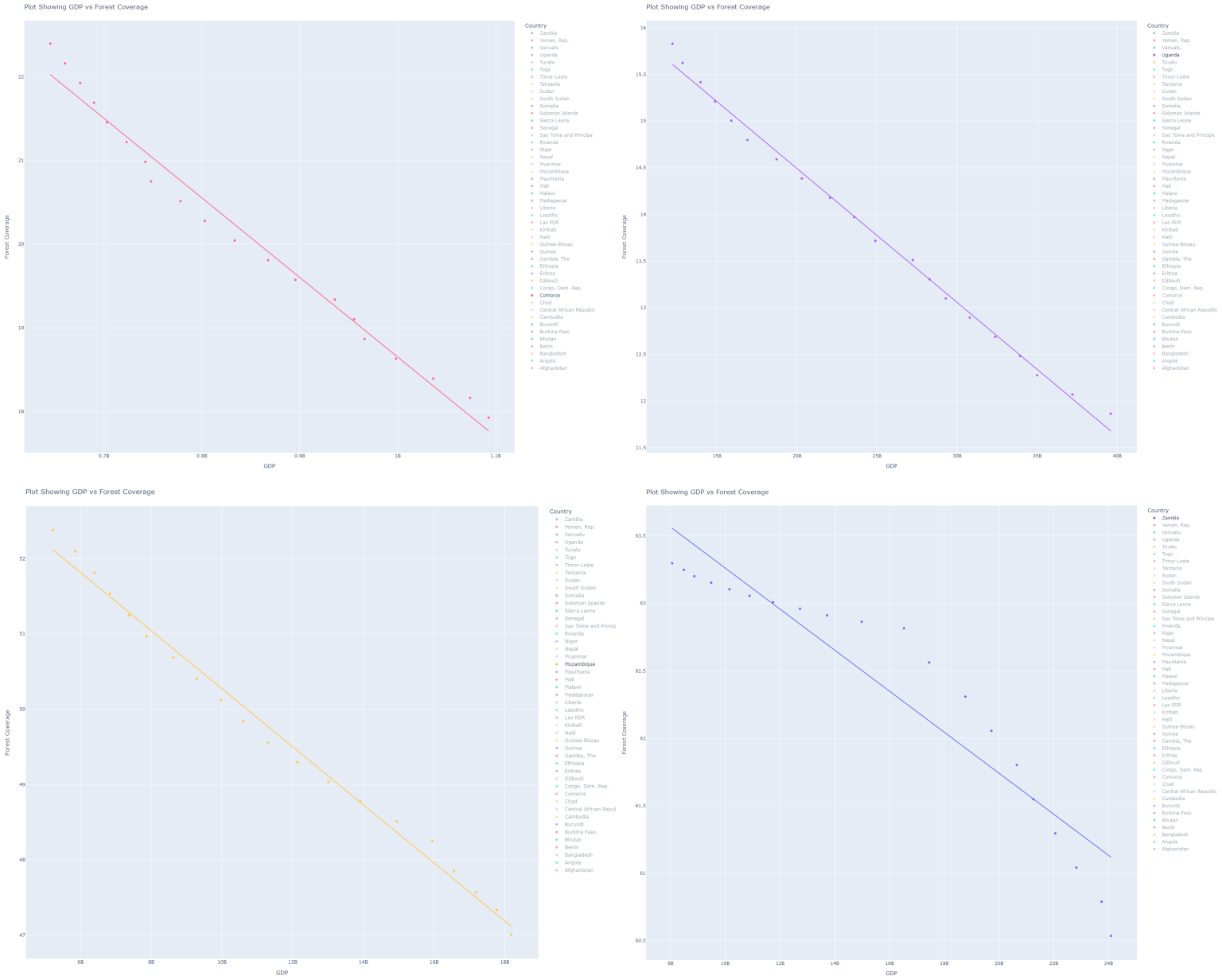
negative correlation between GDP and Forest Coverage. With a majority having a coefficient above 0.8 and an average coefficient of 0.68.

Like before, when we looked at a few individual countries more closely and found plots backing this up, showing a very strong negative correlation between GDP and Forest coverage change



except for a handful of nations. For some countries like Uganda the relationship appears to be near perfect correlation.

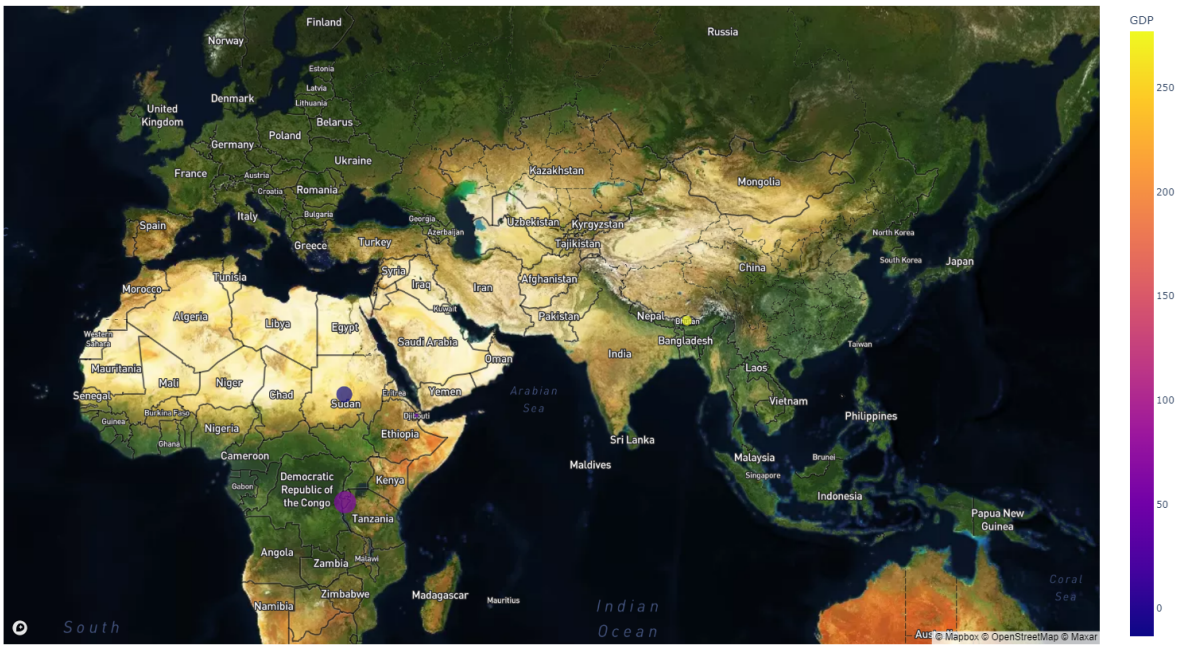
Initially this drastic difference in the strength of the correlation didn't make sense. However, it seems to be the case over time as GDP increased forest cover decreases in most countries however it might be a more cumulative relationship. What all these graphs tell us is that years with High GDP growth aren't also years with high deforestation but rather the effect is more cumulative, as over time a country's GDP and thus its economy grows it's forest coverage starts to shrink over time. It doesn't happen immediately but rather it happens over time.



From Top Left Clockwise: Comoros, Uganda, Mozambique, and Zambia

After finding this relation we decided to test and see if there were any geographic variation in the relationship in the sense that it varied by continent or even if there was sort of a relationship between more arid nations versus more tropical nations. To do this we plotted the Net change in GDP and Net Change in Forest coverage from 1999 – 2019 on top of a satellite map.

Maps Showing Net Change in GDP (%) vs Forest Coverage (%) From 1999 - 2019
Size of Markers is Proportional to Net Change in Forest Coverage

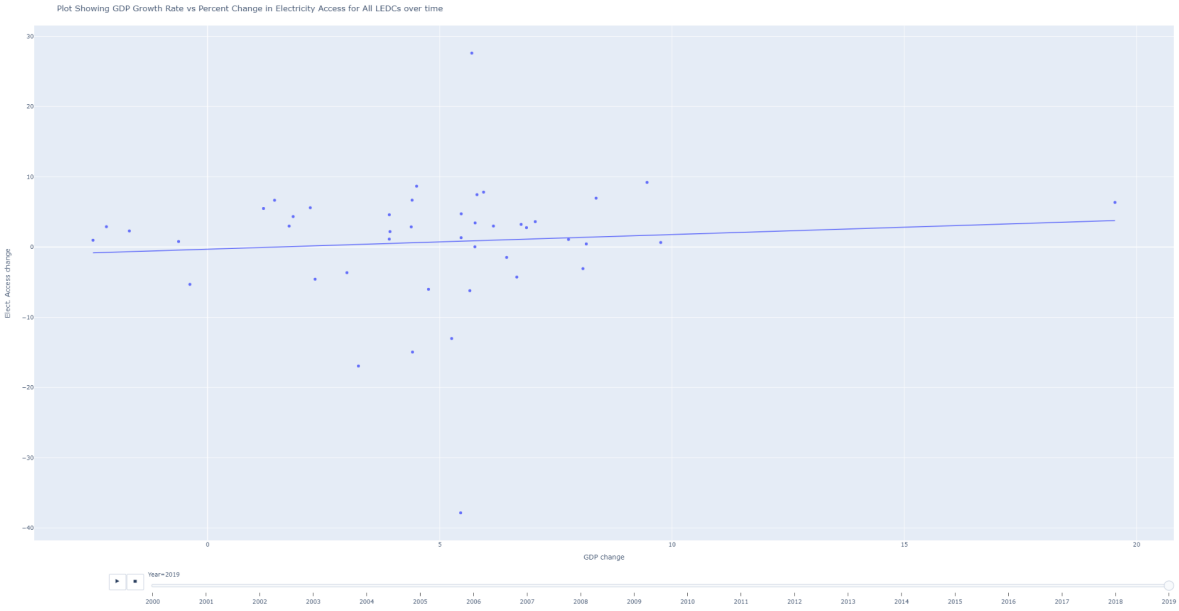


Maps Showing Net Change in GDP (%) vs Deforestation (%) From 1999 - 2019
Size of Markers is Proportional to Rate of Deforestation

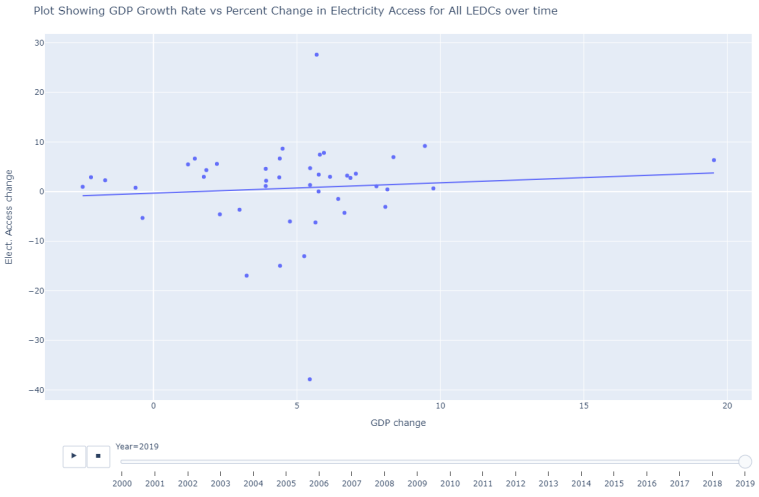
Overall, there doesn't appear to be any sort of geographic variation in this relationship between GDP and forest coverage. There doesn't also appear to be any noticeable differences between more arid nations and more tropical nations, the rates seem similar. The plot also shows that

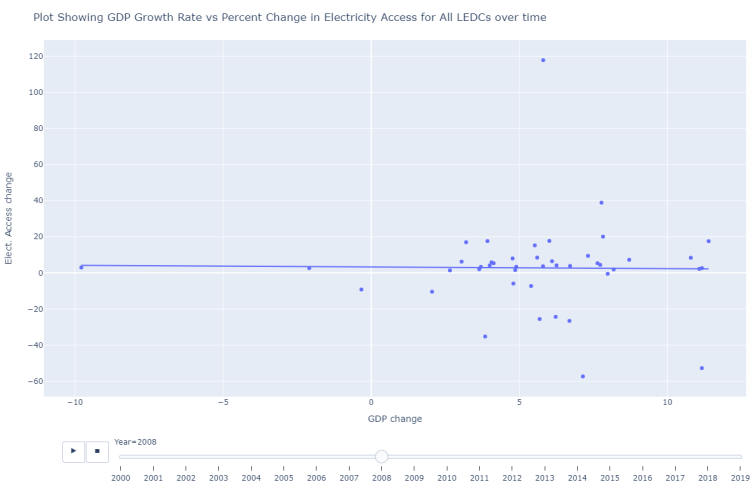
when comparing countries against each other there isn't a correlation between GDP growth from 1999 – 2020 and decrease in forest coverage. In a sense, countries that experienced a large GDP growth aren't also nations that had larger decreases in forest cover relative to other nations that experience low or medium GDP but rather the relationship only exists when we look at an individual country over time.

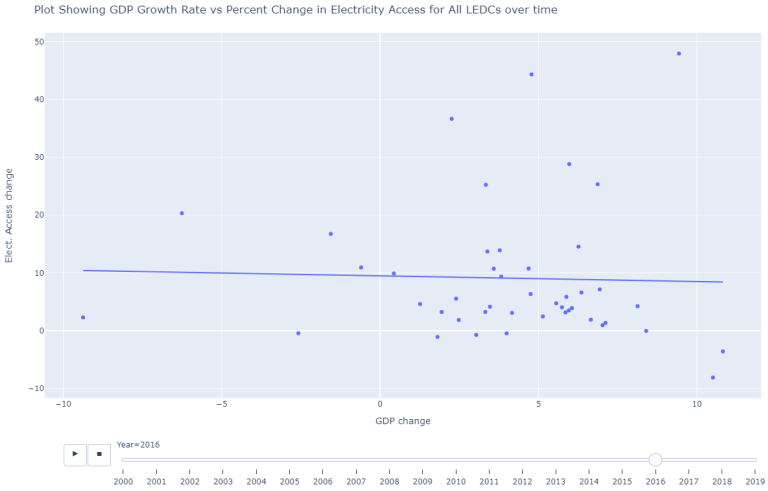
How does increased access to electricity affect the economic growth rate for countries?



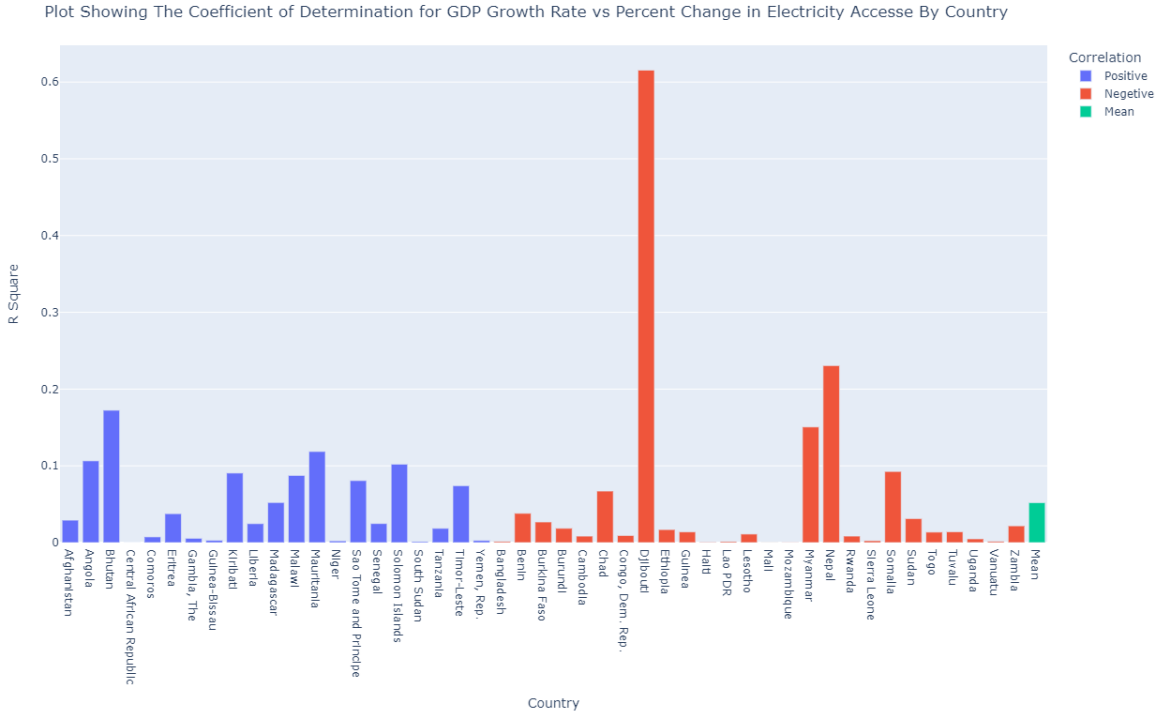
For change in the Electricity access and GDP growth we again noticed similar trends as we did with GDP growth and change in Forest Coverage. When we initially looked at percent change in electricity access against the GDP growth rate we again found a very weak correlation and except for the year 2000 it appeared that that correlation was near zero as you can see from the graphs below.



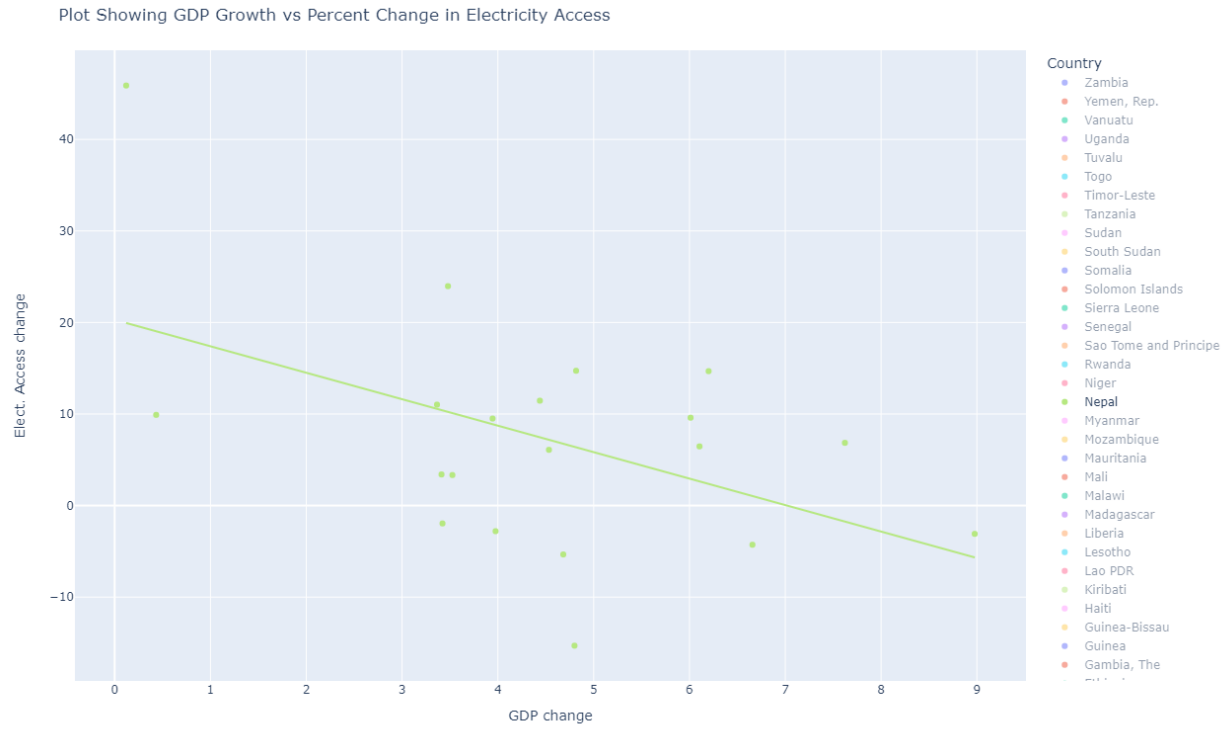




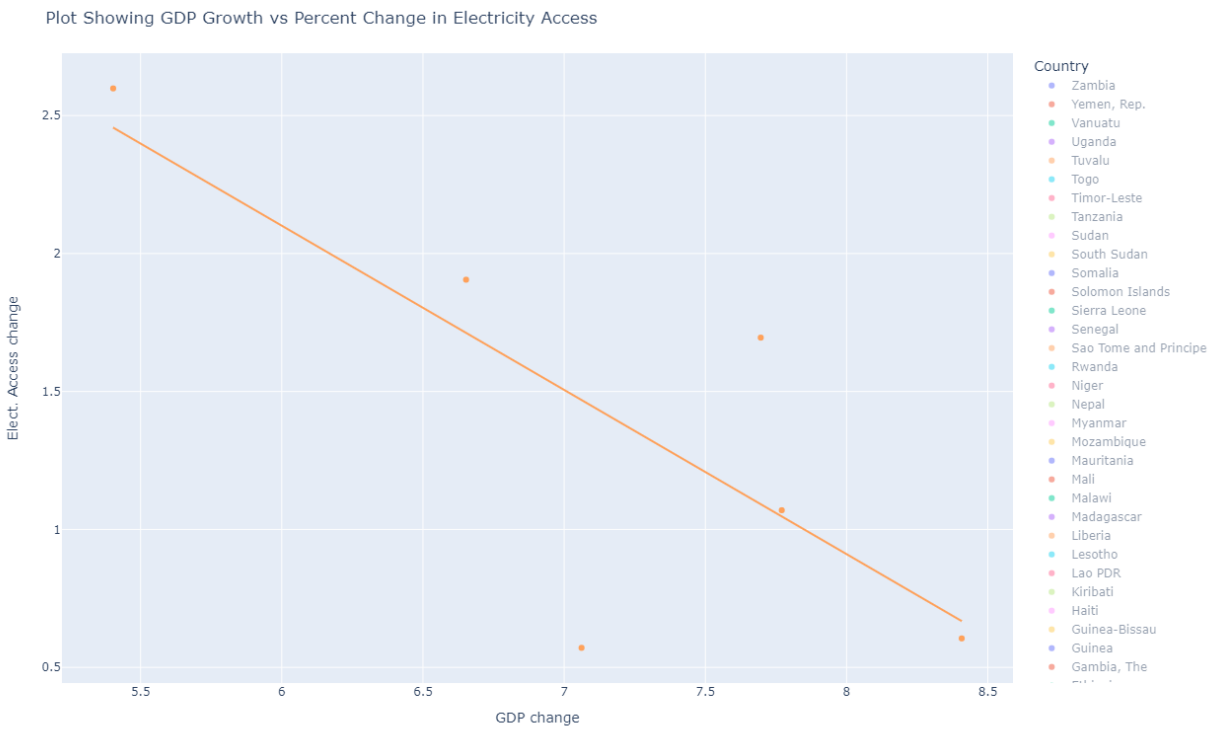
We then again wanted to compare the relationship for each individual nation and so we once again plotted the the R Squared value on a Bar chart for all LEDCs and from these we also found that that the relationship appeared to be very weak we all expect for two having a R Squared value of less than 0.2, indicating a very week, almost non existent relationship between the two variables and again in terms of the direction of the relationship, it also seems to be haphazard with some countries showing a positive correlation and others a negative. But since the correlation is so weak the direction doesn't really matter there is a near zero correlation. This was the case with almost all countries Except two, Djibouti and Nepal.



From the graphs below we can see that while Nepal has a correlation coefficient only slightly above 0.2 the graph does show an indication of a correlation or at least a general trend, it doesn't appear haphazard or random.



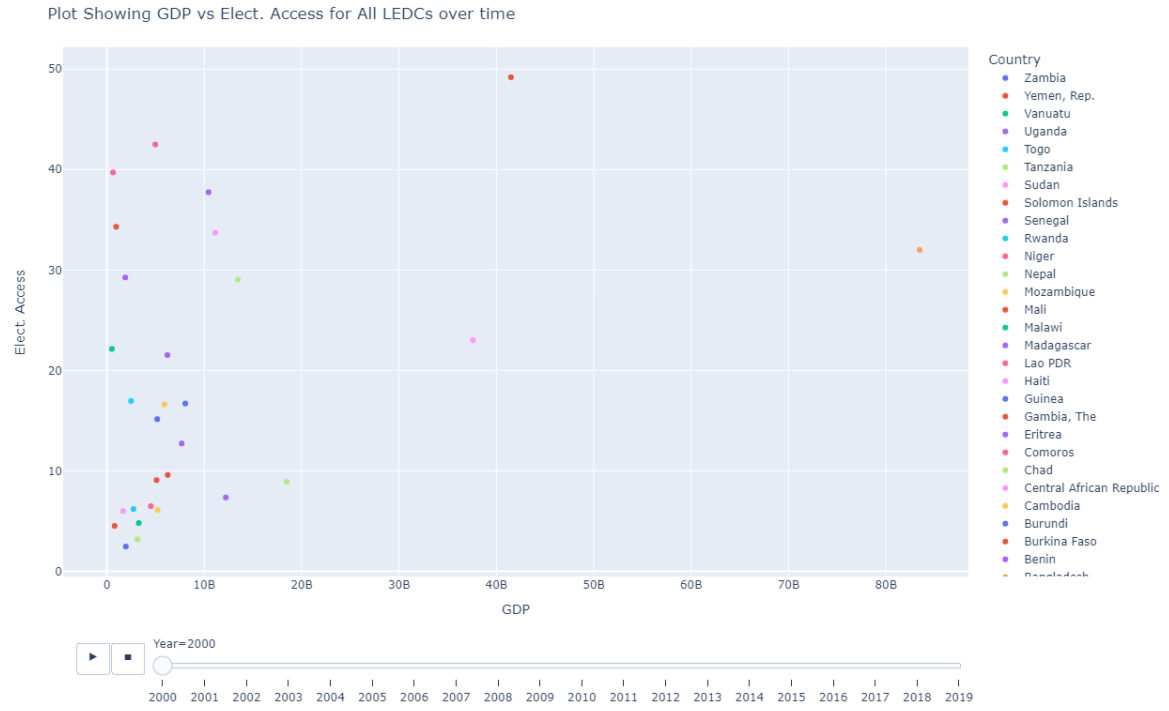
For Djibouti the graph generally shows a fairly strong correlation between GDP growth and Change in electricity access. However this might be a result of limited data points available for Djibouti and might only be a statistician fluke. For Djibouti as you can see there are only 6 data points available compared to around 20 for most countries in the dataset.

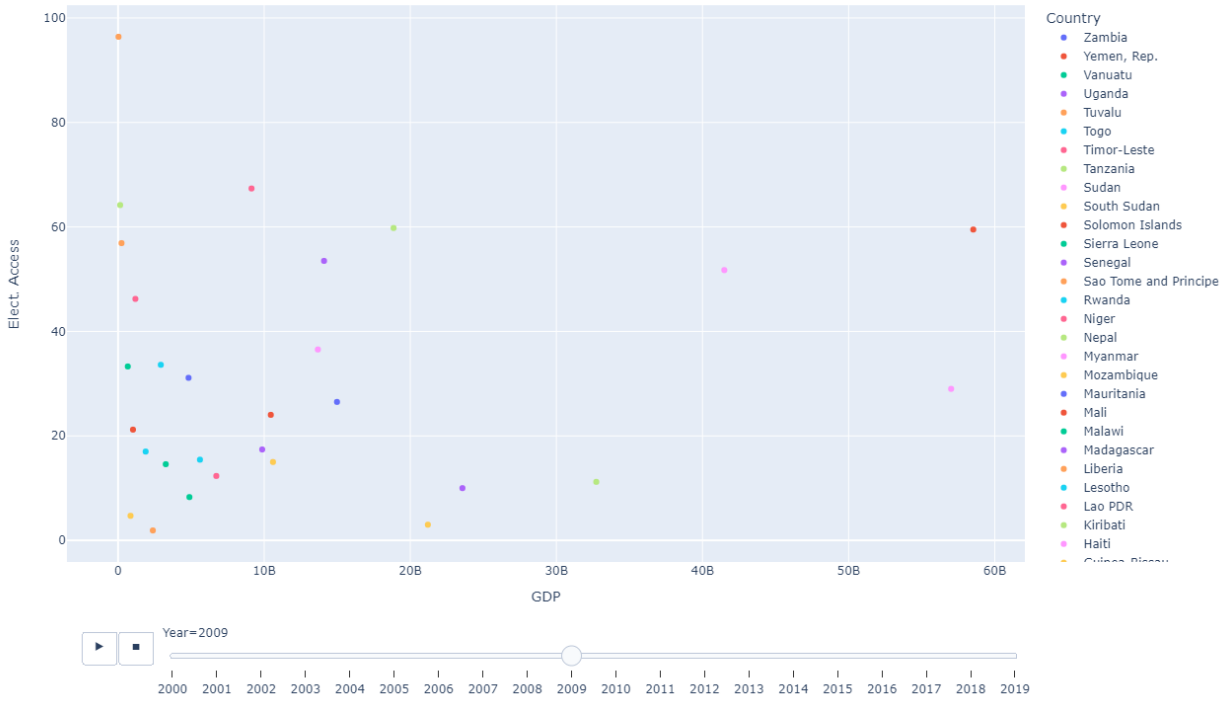


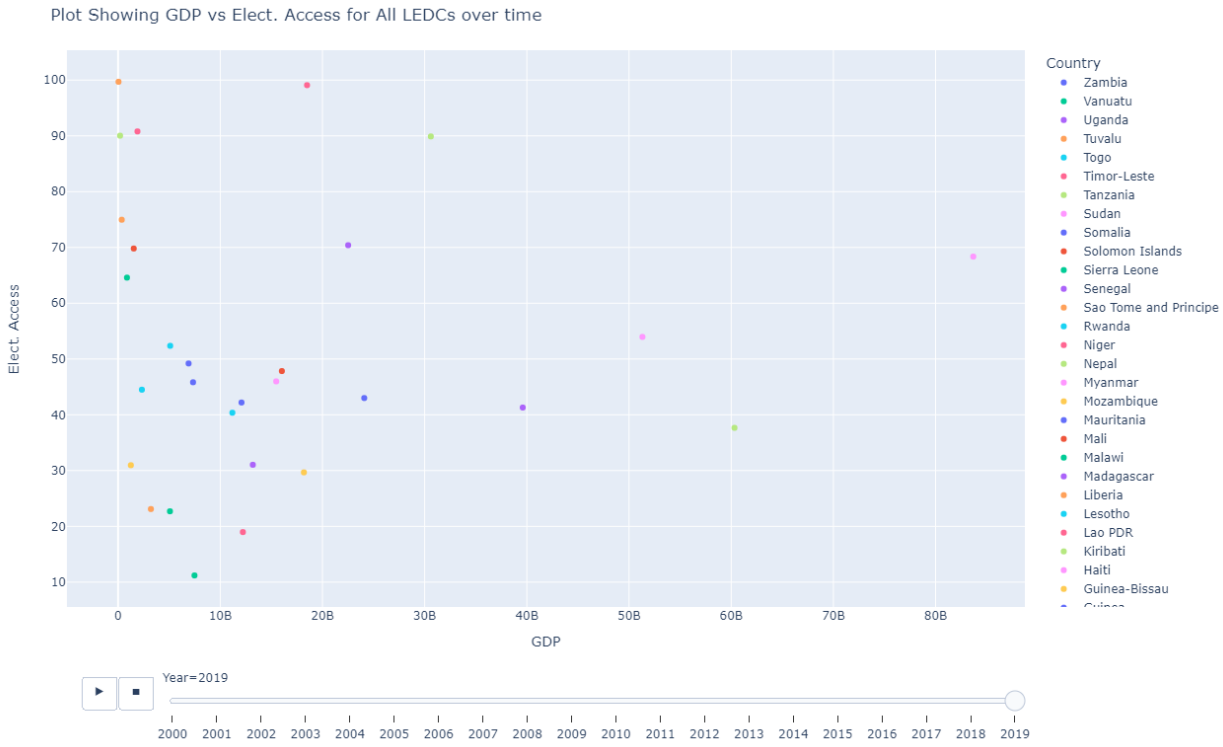
Interestingly enough Djibouti and Nepal both have a negative correlation between GDP growth and change in electricity access which is very counterintuitive as we would assume that as a country's economy grows electricity access would increase. However since we don't see a similar case with most of the other countries, it's fair to assume that Djibouti and Nepal might just represent anomalous cases rather than an indication of some sort of a wider relationship between change in electricity access versus GDP growth.

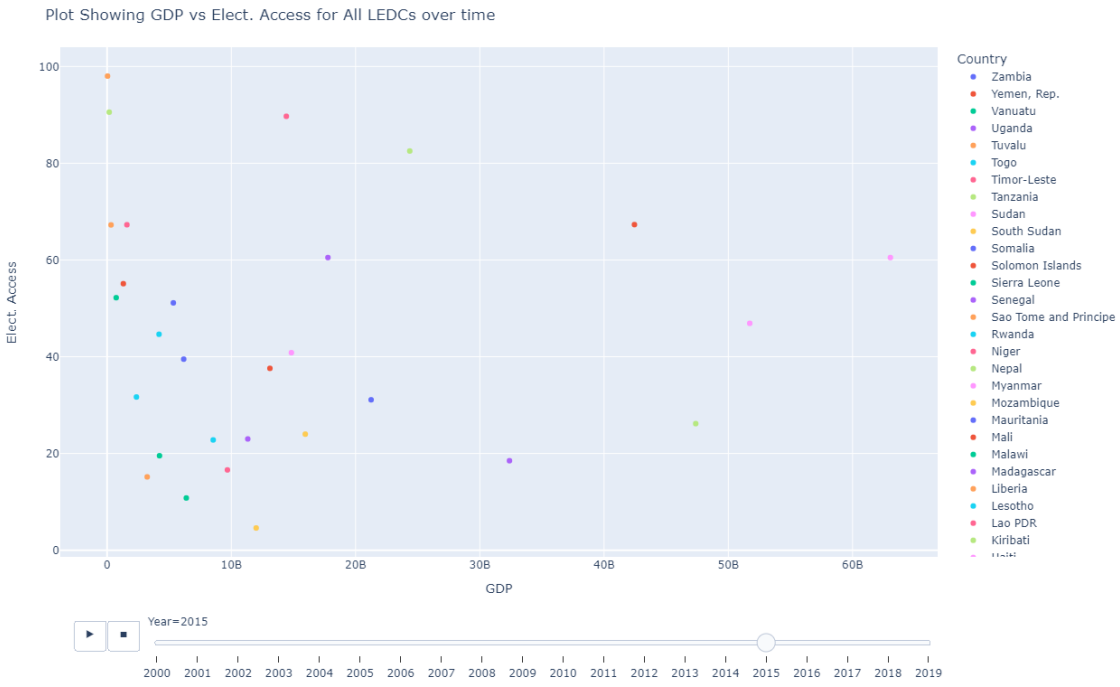
Once we concluded that there doesn't appear to be some sort of a relationship between GDP growth and change in electricity access we again decided to look at just GDP versus electricity access and like we again found that relationship between these two variables are extremely strong and there appears to be a strong positive correlation between GDP and electricity access over time.

Starting again with the GDP versus Electricity access over time for all countries we can see that over time the data points representing countries appear to be moving up and to the left over time. This time there seems to be a larger more drastic movement indicating that 2000 - 2019 there was a pretty drastic increase in GDP and Electricity access in LEDCs.

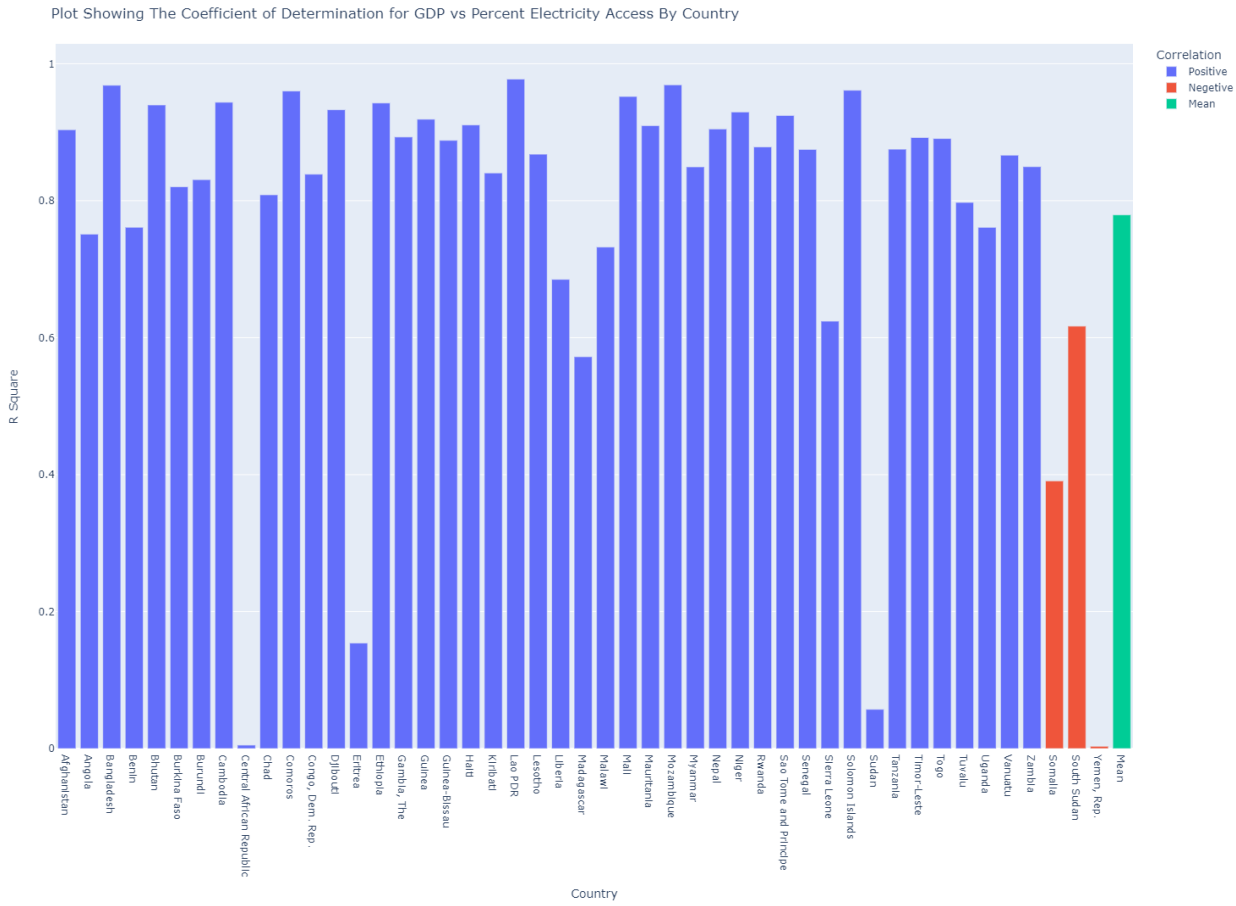

Plot Showing GDP vs Elect. Access for All LEDCs over time
Elect. Access
GDP
0
10
20
30
40
50
0
10B
20B
30B
40B
50B
60B
70B
80B
Country
Zambia
Yemen, Rep.
Vanuatu
Uganda
Togo
Tanzania
Sudan
Solomon Islands
Senegal
Rwanda
Niger
Nepal
Mozambique
Mali
Malawi
Madagascar
Lao PDR
Haiti
Guinea
Gambia, The
Eritrea
Comoros
Chad
Central African Republic
Cambodia
Burundi
Burkina Faso
Benin
Year=2000
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Plot Showing GDP vs Elect. Access for All LEDCs over time
Elect. Access
GDP
0
20
40
60
80
100
0
10B
20B
30B
40B
50B
60B
Country
Zambia
Yemen, Rep.
Vanuatu
Uganda
Tuvalu
Togo
Timor-Leste
Tanzania
Sudan
South Sudan
Solomon Islands
Sierra Leone
Senegal
Sao Tome and Principe
Rwanda
Niger
Nepal
Myanmar
Mozambique
Mauritania
Mali
Malawi
Madagascar
Liberia
Lesotho
Lao PDR
Kiribati
Haiti
Year=2009
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Plot Showing GDP vs Elect. Access for All LEDCs over time
Elect. Access
GDP
Year=2019
Country
Zambia
Vanuatu
Uganda
Tuvalu
Togo
Timor-Leste
Tanzania
Sudan
Somalia
Solomon Islands
Sierra Leone
Senegal
Sao Tome and Principe
Rwanda
Niger
Nepal
Myanmar
Mozambique
Mauritania
Mali
Malawi
Madagascar
Liberia
Lesotho
Lao PDR
Kiribati
Haiti
Guinea-Bissau


Plot Showing GDP vs Elect. Access for All LEDCs over time
Elect. Access
GDP
Year=2015
Country
Zambia
Yemen, Rep.
Vanuatu
Uganda
Tuvalu
Togo
Timor-Leste
Tanzania
Sudan
South Sudan
Somalia
Solomon Islands
Sierra Leone
Senegal
Sao Tome and Principe
Rwanda
Niger
Nepal
Myanmar
Mozambique
Mauritania
Mali
Malawi
Madagascar
Liberia
Lesotho
Lao PDR
Kiribati

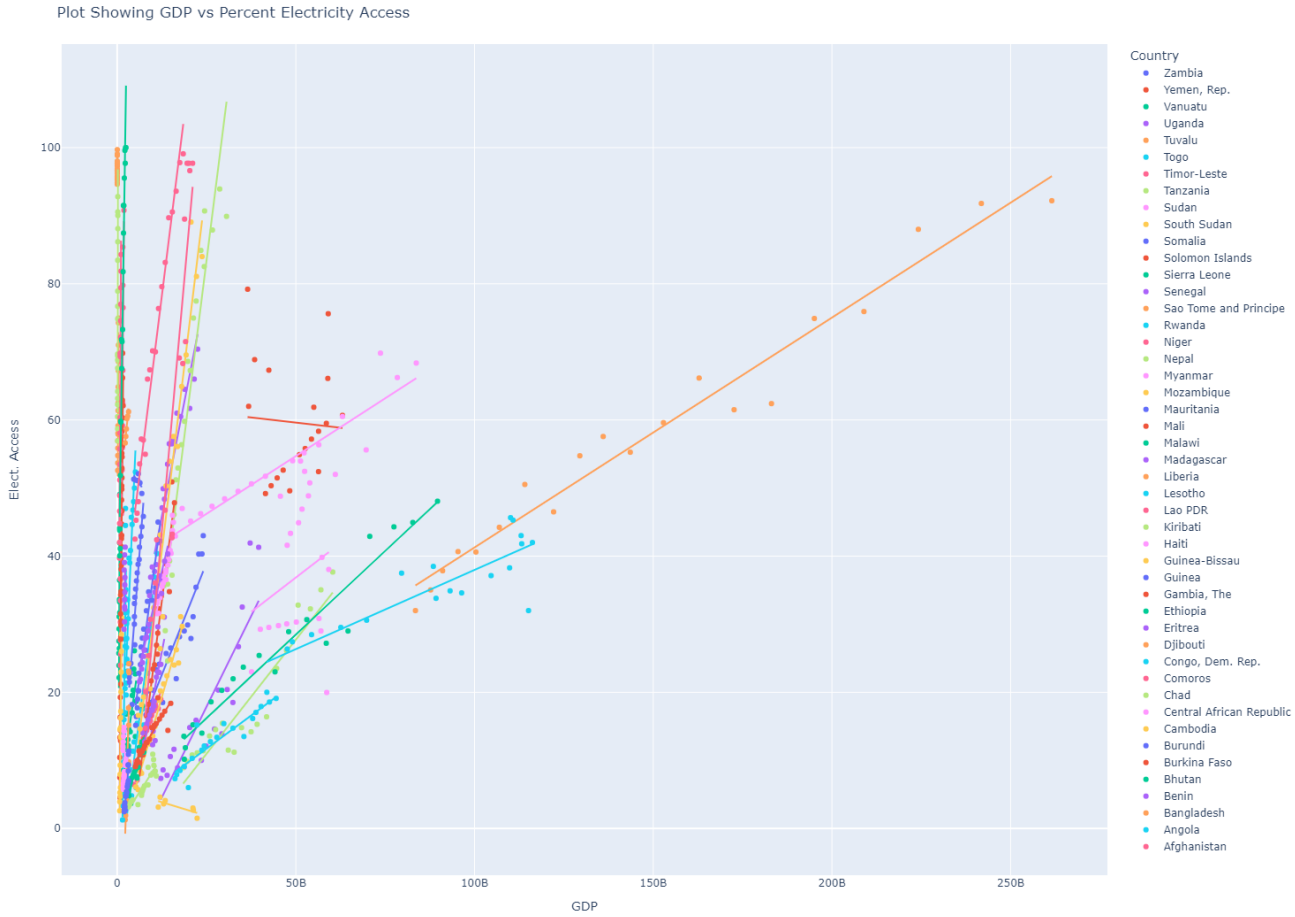
After seeing that there might be a stronger correlation between GDP and electricity access we again decided to calculate the R-Squared value for all LEDCs and plot them on a bar chart.



As you can see from the graph above there is a very positive strong correlation between electricity access and GDP for the vast majority of LEDCs. The correlation between GDP and electricity access seems to be even stronger than between Forest Coverage Change and GDP where the average correlation was 0.68 for GDP and Electricity access is as high as 0.78, indicating a very strong correlation.

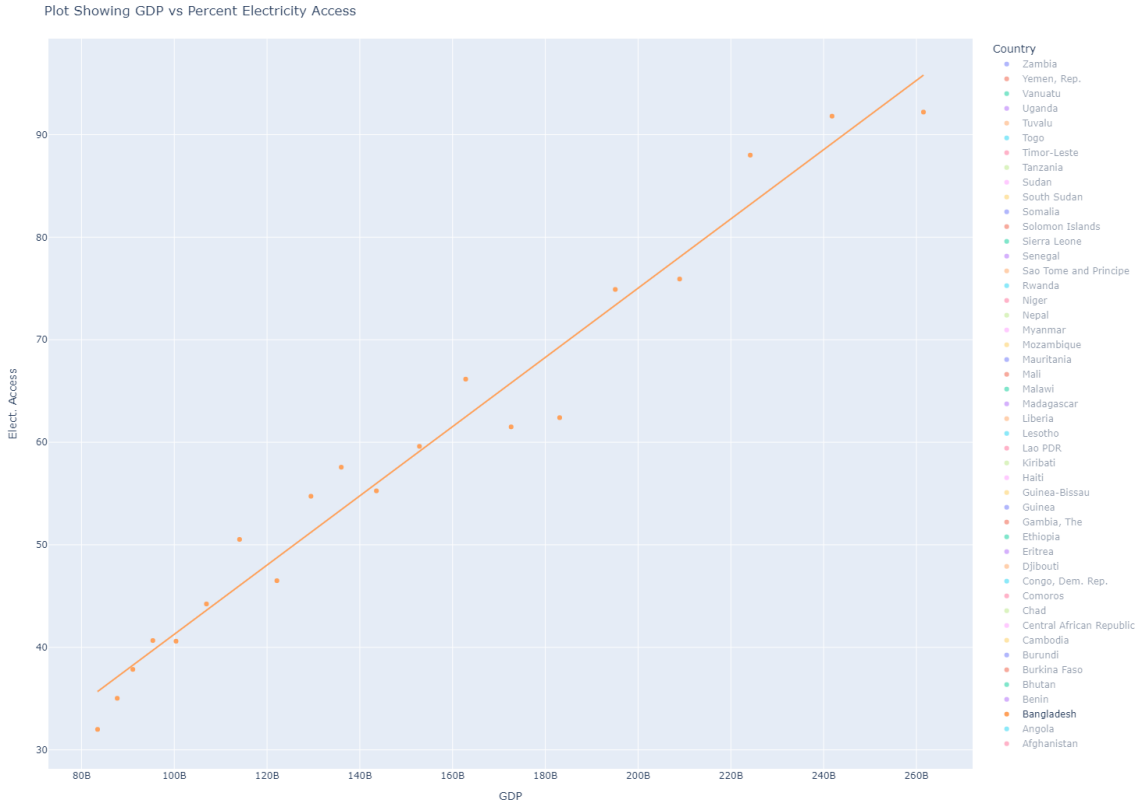
The next graph showing GDP and Percent Electricity access while a bit messy and hard to read, I think still provides useful insight. As you can see from the graph from 2000 to 2019 there was a pretty drastic increase in electricity access in LEDCs. However, interestingly enough, while there is a strong correlation between the two variables, almost all LEDCs experienced a drastic increase in Electricity access irrespective of how much their GDP increased. This can be seen by the cluster of steep trendlines on the left of the chart. While Bangladesh for example saw a near one to one increase in GDP and Electricity access, it was more an exception than the rule. This atleast to me hints at the fact that there might be a spurious relationship between GDP and Electricity access with time acting as an antecedent variable in this case. The reason I think it's time is because we know that most countries' economies grow over time and so it seems does

electricity access. So in this case GDP might just be standing in for time and the real causal might be between GDP and time and electricity access and time.

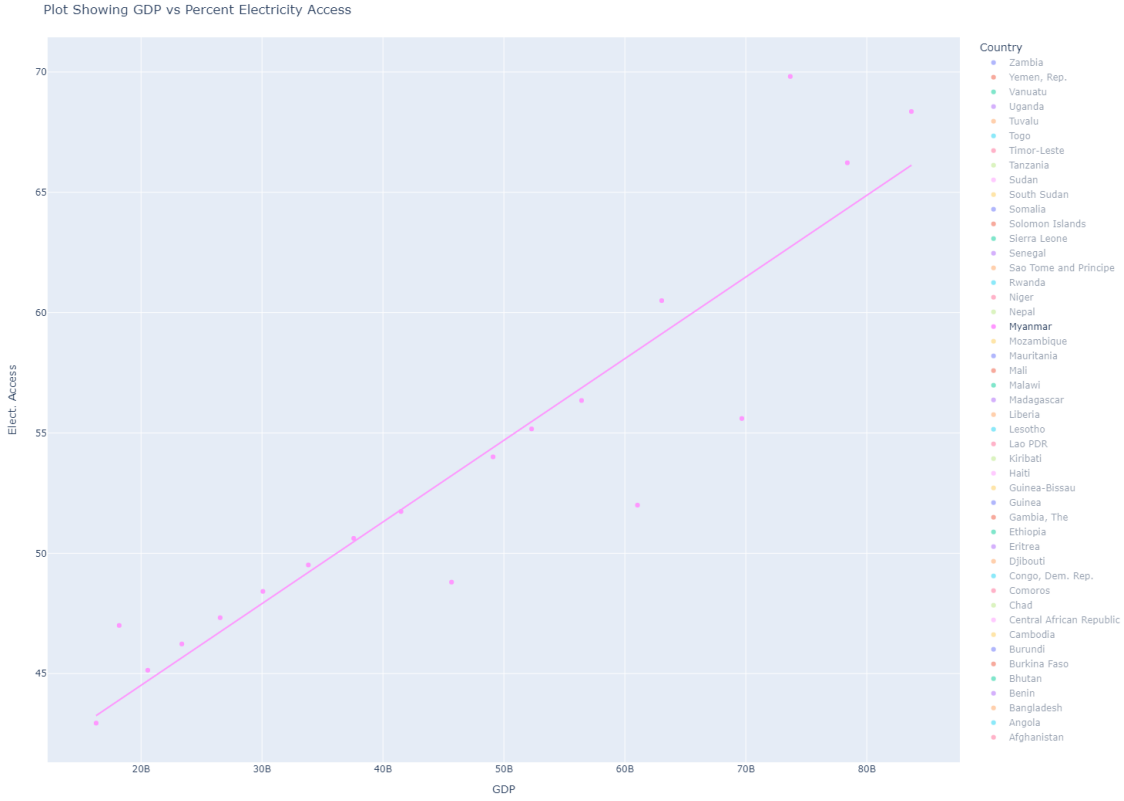


Looking at the individual countries we can see how strong the correlation is between GDP and electricity access. For almost all LEDCs there is a very strong positive correlation.

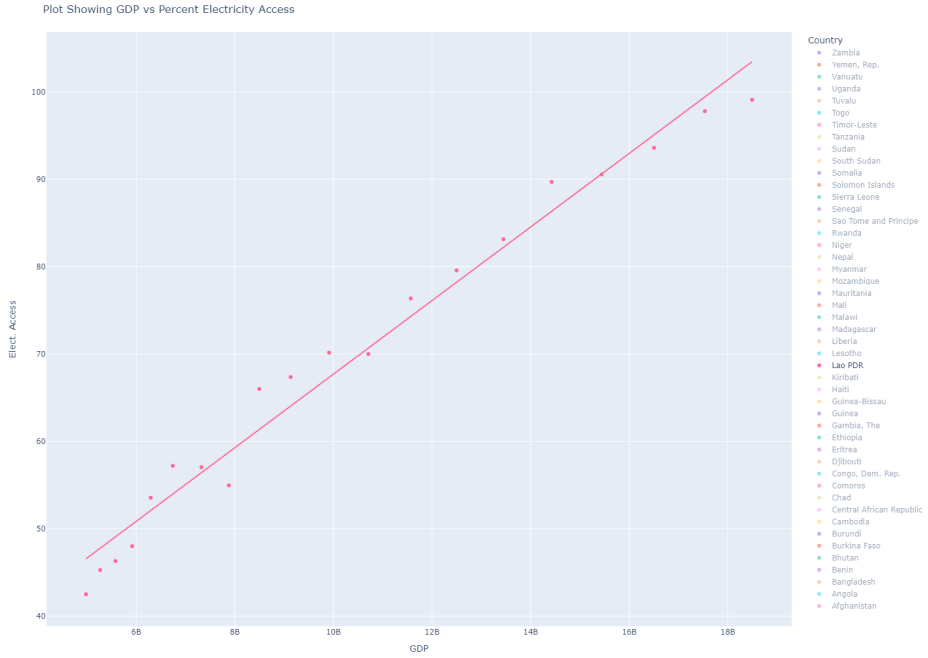
Bangladesh



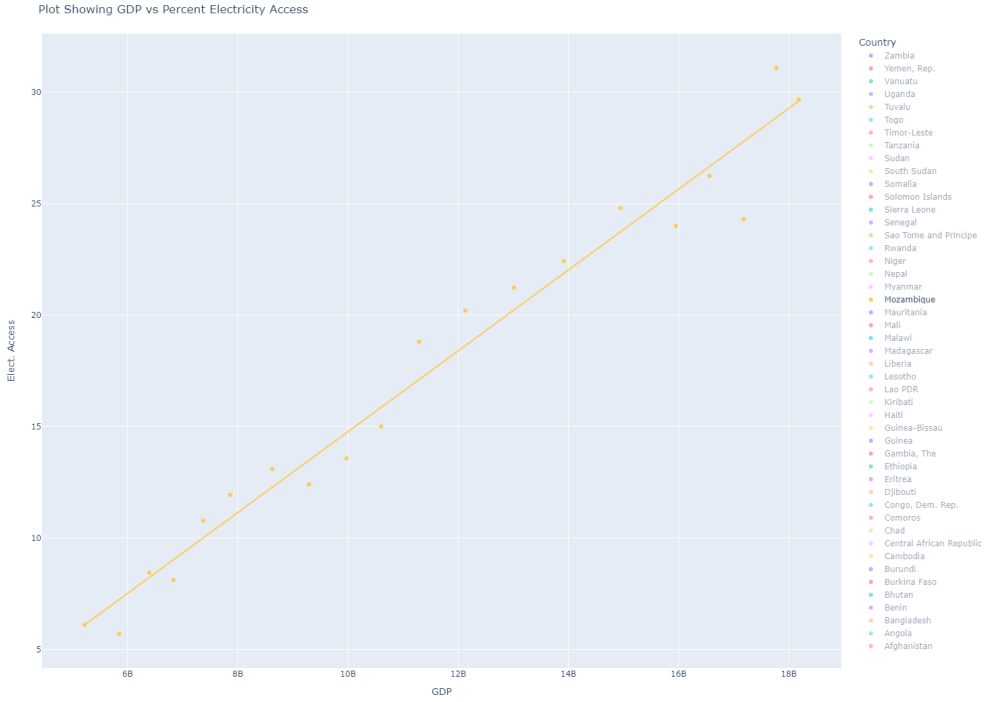
Mayanmar



Laos



Mozambique

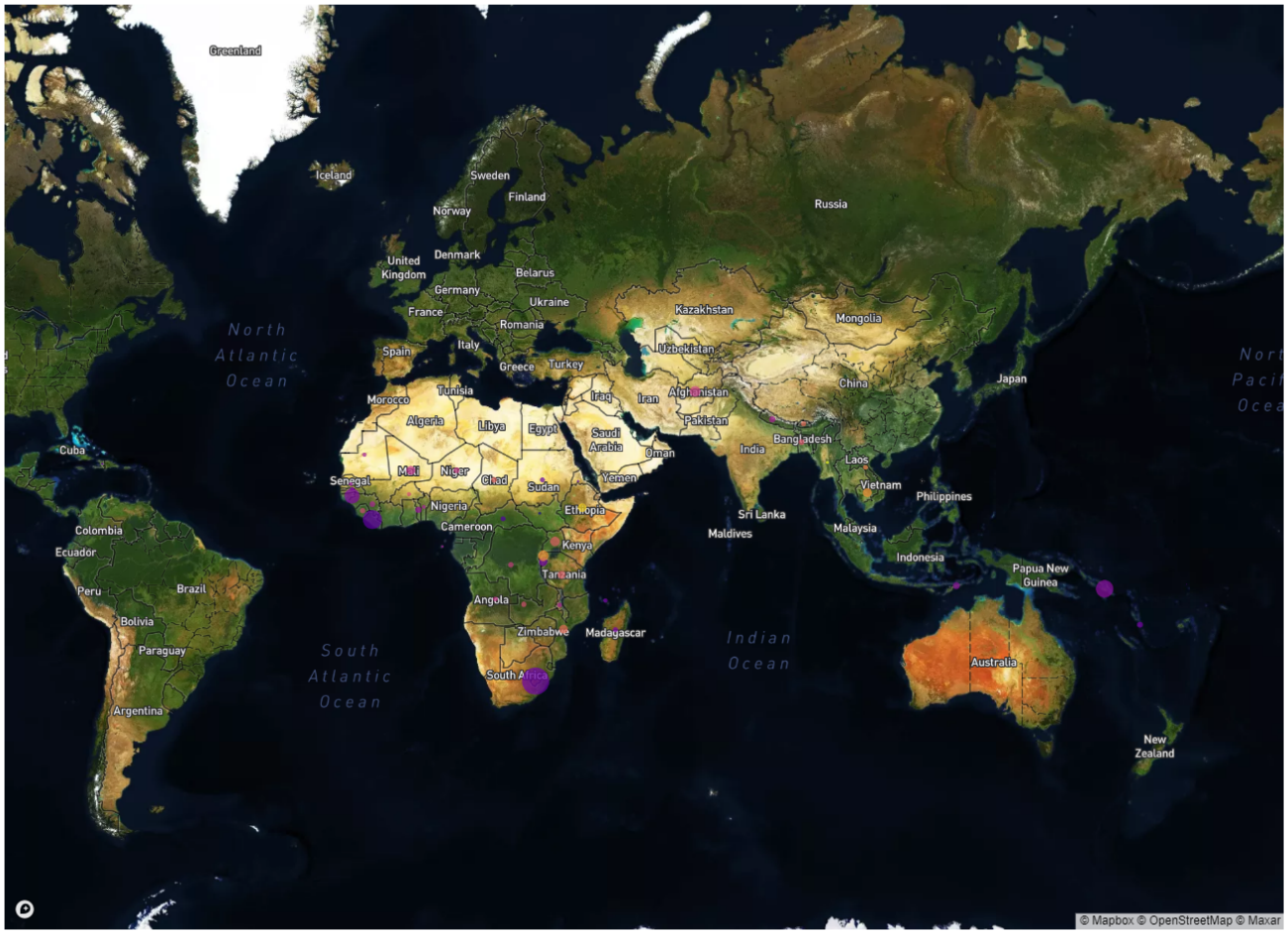


After we were done looking at the relationship for each country we again wanted to explore and see if there are any geographic relationships. From the maps we again see that there doesn't appear to be any significant spatial relationship between net change in electricity access and net change in GDP. But it does look like the African LEDCs seem to have experienced a large net change in electricity access compared to asian LEDCs. This might simply be due to the fact that there are more LEDCs in Africa than asia. But it's also interesting because according to the previous graph we saw that most countries started with a similar level of electricity access in 200 and so this might hint at the fact that there might have been some sort of a large push to increase electricity access in africa from 1999 - 2019. This also supports my hypothesis that there isn't a causal link between GDP and electricity access and the relationship might be spurious with time acting as an antecedent variable and GDP or electricity access just standing in for time

Maps Showing Net Change in GDP (%) vs Decrease in Electricity Access (%) From 1999 - 2019
Size of Markers is Proportional to Change in Electricity Access

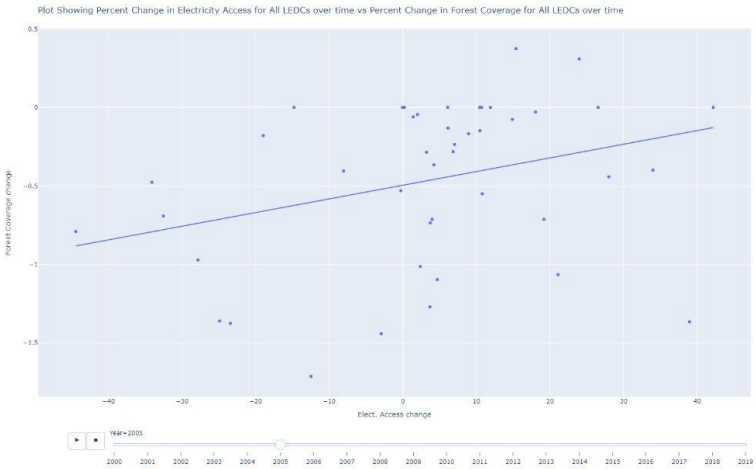
Maps Showing Net Change in GDP (%) vs Change in Electricity Access (%) From 1999 - 2019

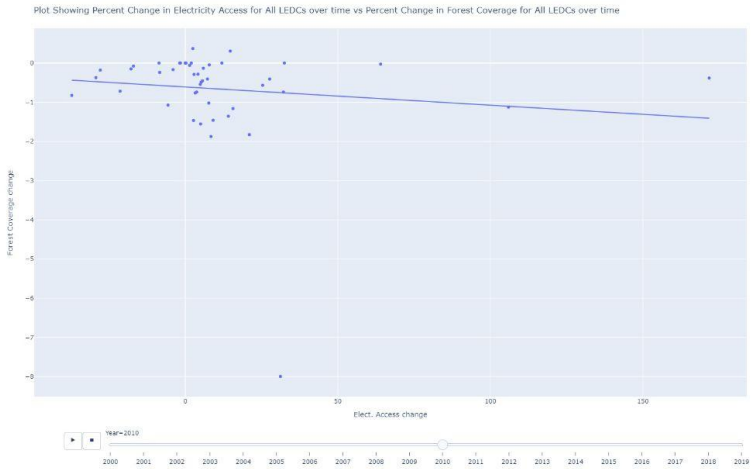
Size of Markers is Proportional to Change in Electricity Access

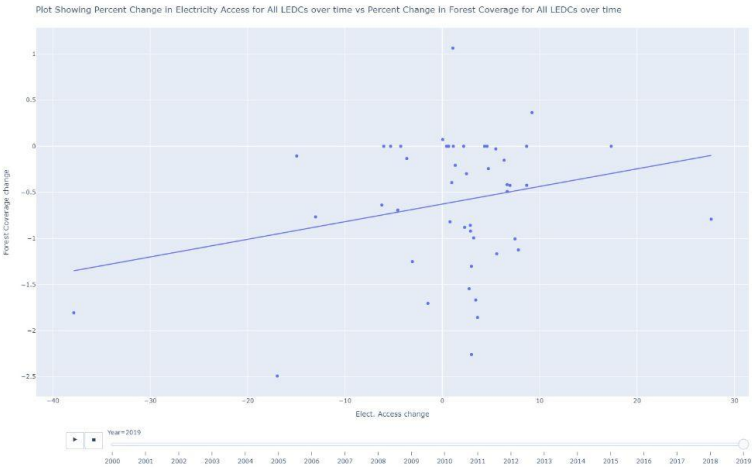


What is the relationship between the change in national electricity access and the change in national forest coverage?

Now that we have looked at GDP and Electricity and GDP and Forest Coverage we thought it would be interesting to look at if there was some sort of relationship between Change in Forest

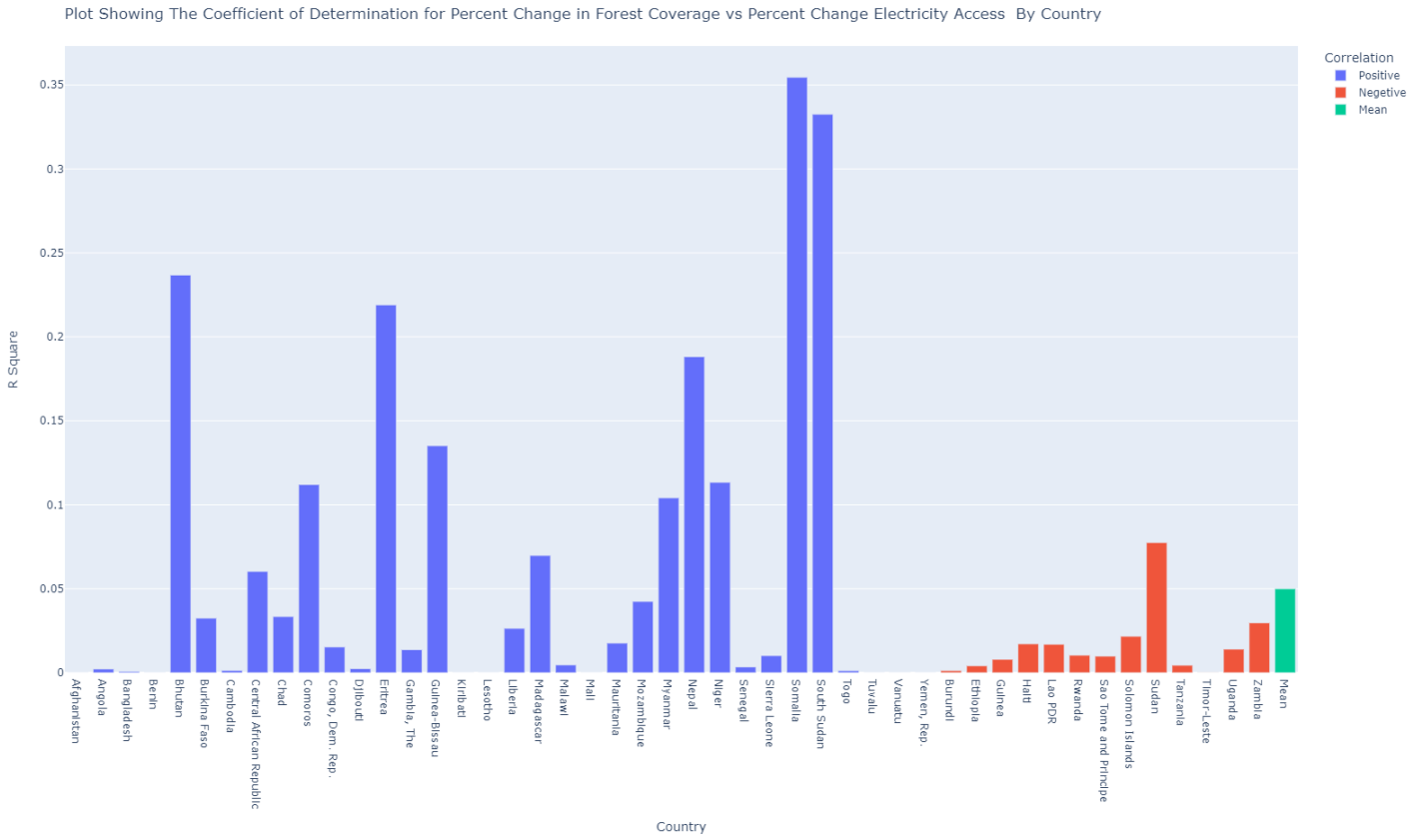




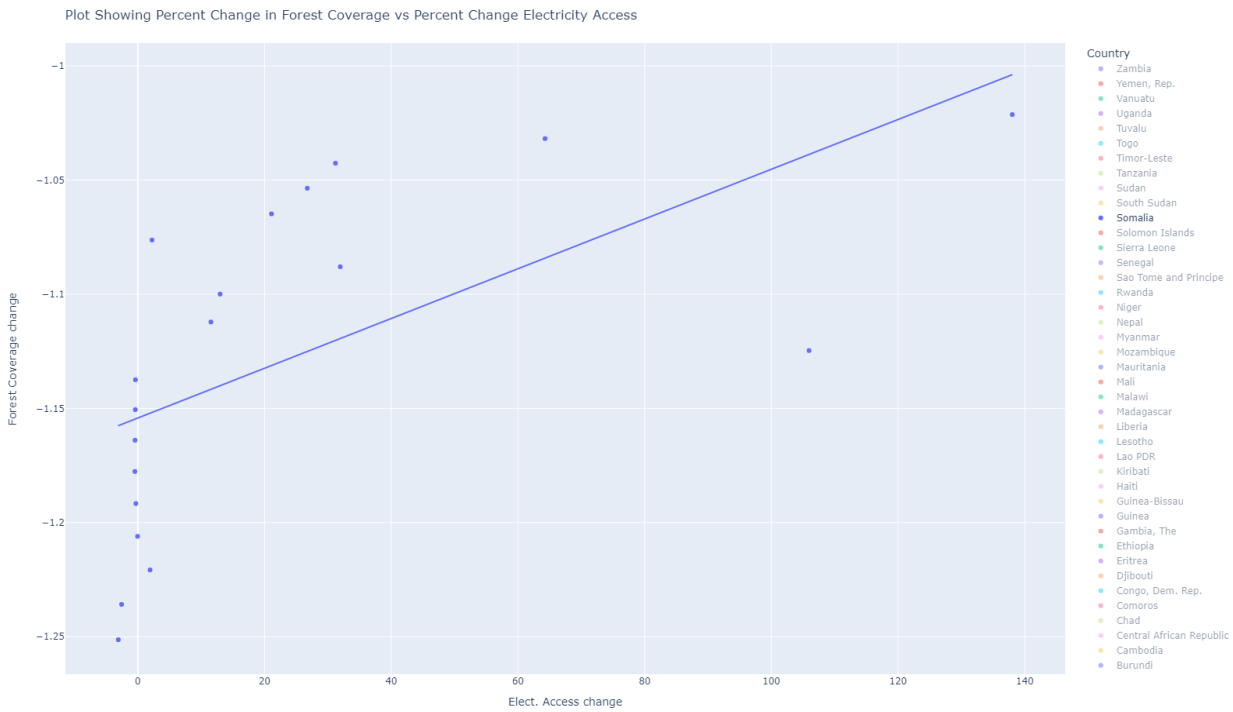


As you can see from the plot above when comparing countries over time the correlation between the two variables is generally weak and often sporadic like the previous two cases for Question 1 and Question 2. When we look at individual countries by looking at the R squared value in the bar chart below the most countries seem to have a positive correlation between the correlation coefficient is so low, on average 0.04, that it isn't anything meaningful as the correlation is nearly zero. This indicates that there is no correlation between Change in Forest Coverage and Change in Electricity access. Even somalia with the strongest correlation has a correlation coefficient of only 0.35 and when we pull up the individual plot for somalia, as can be seen below, we can see that even that seems to be more of a mathematically anomaly resulting from the way the R Squared value is calculated rather than an indication of any sort of real

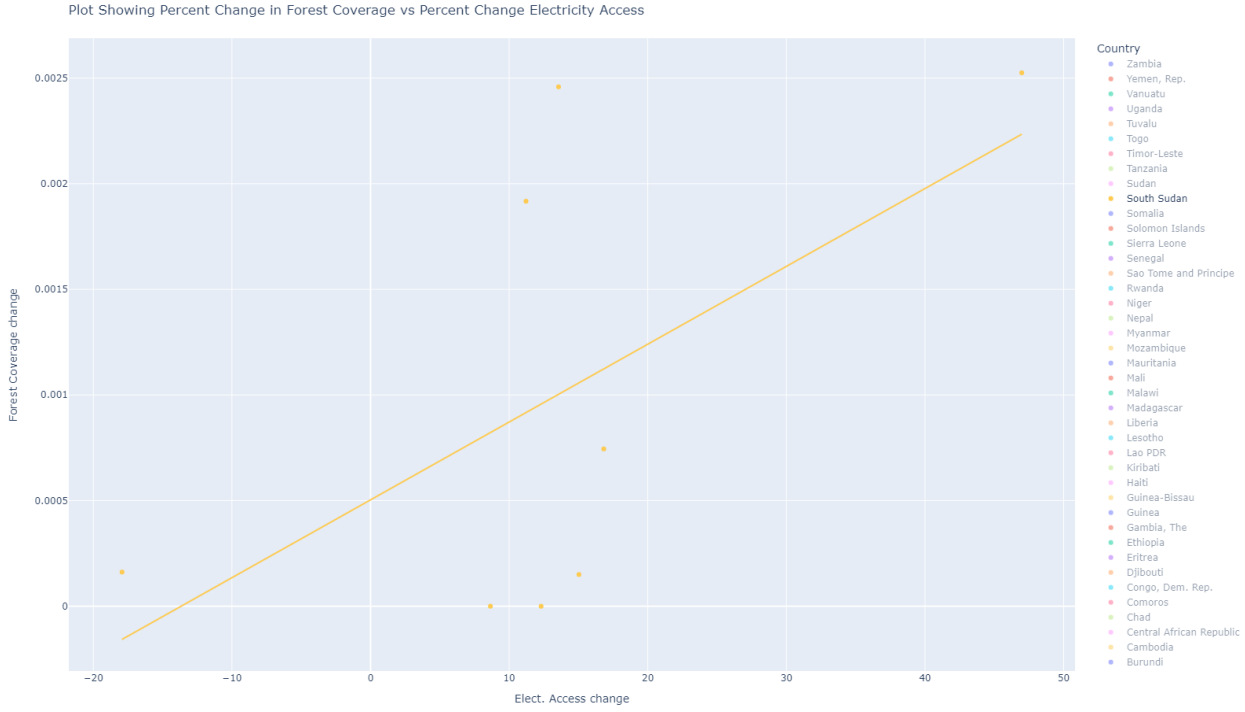
relationship. For South Sudan there are just way too few data points to come up with any sort of conclusion.



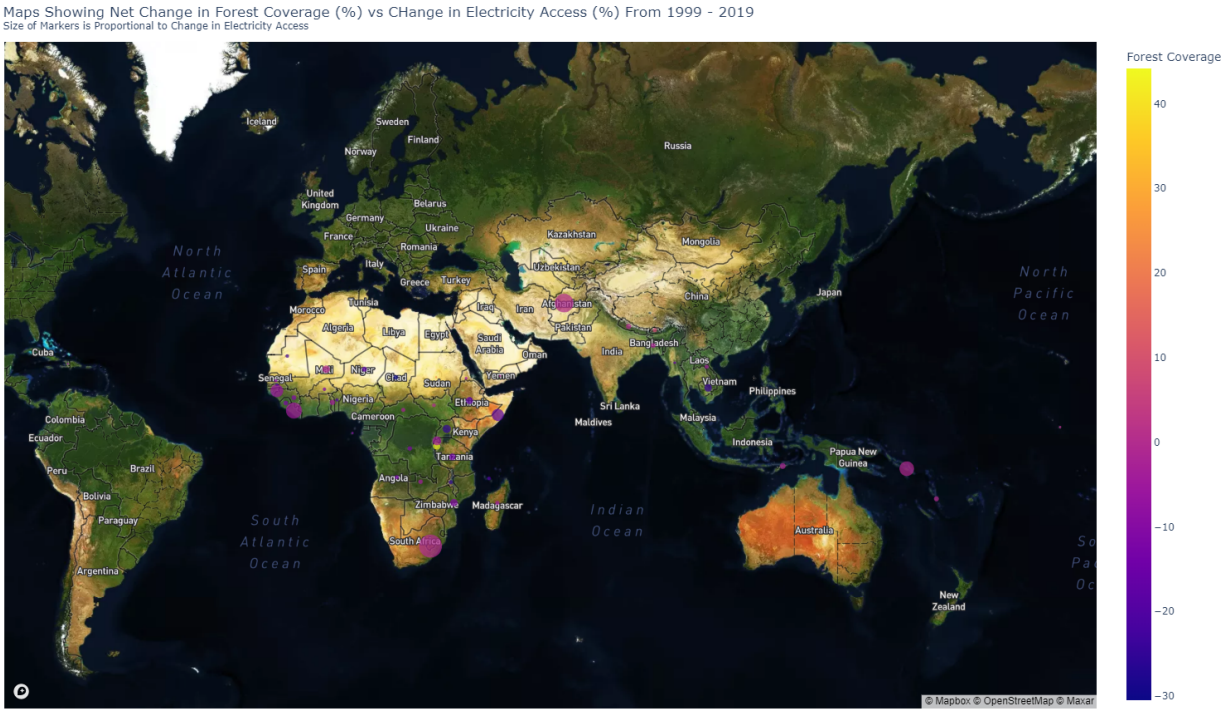
Somalia



South Sudan



After we couldn't find any correlation between these two variables we decided to look at if there are any spatial patterns so again created a map showing the Net Change in Forest Coverage versus Net Change in Electricity Access from 1999 - 2019. However from the map we couldn't see any sort of obvious patterns and relationships and it seemed to be a bit random.



6. Impact and Limitations

Overall, we believe that our report is merely an exploratory analysis, hinting at areas of future research rather than any sort of conclusive findings. The fact is we haven't come close to using the level of statistical analysis or the range of variables needed to conclusive determine whether any of these variables are linked. At the same time while it might seem like a lot, we only used a limited amount of data. We only used data from a handful countries, and we only used 20 years' worth of data, a lot of which is missing. At the same time, we also didn't consider or analyse nearly enough variables or consider multiple dimensions. What we see our report doing is hinting at relationship that could be further explored.

At the same time there is also a very specific reason why we decided to try to find a relationship between forest coverage, electricity access, and GDP. The reason is its very difficult for developing nations to accurately track and measure GDP and it can be a crucial indicator of economic performance and development. A countries GDP can play a major role in determining things like tariff free access to developed market, level of foreign aid, access to developmental loans, amount of Foreign Direct Investment, and many more privileges. These things can play a crucial role in determining the trajectory a country heads. However, GDP is an extreamly difficult indicator to measure for developing nations as a lot economic activity either doesn't get reported or happen in the informal sector. This is where forest coverage and electricity access can come in. Forest coverage and electricity access are two variables that can be estimated using satellite imagery, by either look at the changes in the size of forest or by looking at night-time images looking at light emissions from things like lights bulb. Using satellite images could be a cost-effective way country can get frequent estimation of these two variables and if we can show that they are linked to GDP and GDP growth it could be way for a country to get access to relatively accurate economic data frequently that can be used to guide policy decisions and allow countries to respond to economic downturns more quickly.

However, there are still plenty for risks and biases associated with using out data in that way. For one it could potentially penalize or worse disincentivise countries from protecting forest as countries with greater loss of forest coverage would look like there economies are growing faster. Furthermore, a lot of our initial analysis is based on incomplete data, and while the data is from the World Bank, the fact is there is a lack of high quality and reliable data for developing nations and if these biases mean we might see inaccurate patterns in our data it could potentially misguide both researchers who might use it to figure out areas of further research or even policy makers who might use our findings to guide policy making and this could be problem for developing nations who already lack resources and it mean that potentially valuable resources could be wasted. Finally, GDP isn't a perfect metric for tracking economic development and growth and fails to consider many crucial aspects that could indicate the health of an economy.

First, GDP doesn't take into account the environmental damages of increased economic activity. While things like deforestation could potentially boost economic output in the short run, it could potentially leave a country worse of in the future as could valuable and often time irreplaceable resources. Second GDP doesn't consider wealth disparities. Just because a countries economy is growing doesn't necessarily mean that the average citizen is better off as the increased wealth could be straight to the wealthy and the powerful. This could be harmful because if higher GDP leads to things like reduced foreign aid it could potentially harm average of low-income citizens.

7. Challenge Goals

- *New Library: Plotly*
    - We used the plotly express library to create interactive plots for all of the scatter plots we created. By using plotly, we enabled the user to more closely examine the data by changing certain parameters such as the year and the countries included in the plot. Considering that we are plotting the relationships between many countries and years, this was a particularly valuable feature in our analysis because it enabled the user to more easily study changes in the data over time and for specific countries.
    - We had initially planned to use the Folium and Leafly libraries in conjunction with plotly, but we soon realized that we would be making far more plots than necessary to demonstrate our results. Furthermore, we didn't want to overwhelm our viewers / users with even more plots than necessary.
- *Multiple Datasets*
    - In order to analyze trends in national urbanization and industrialization, we used the following three datasets from the World Bank: National GDP, Forest area (% of land area), and Access to Electricity (% of population). By using multiple datasets, we were able to examine several potential indicators of national urbanization and industrialization trends.

8. Work Plan Evaluation

Reflecting on our project proposal work plan, we definitely deviated away from our initial plans a decent amount, but ultimately performed the same tasks just in a slightly different order. Initially, we planned to start by cleaning all of the data and then performing our analysis techniques on each of the three merged datasets individually. Cleaning and organizing our data definitely took quite a bit longer than we had anticipated as our datasets did not have consistent naming conventions and had different amounts of available data for certain countries and years. Once we determined workarounds for cleaning and merging our data, we moved on to our analysis portion. Given that our datasets were all formatted the same way and we were plotting

and analyzing the same types of relationships, we realized that it would be much more efficient to write more generalized code and functions that could be applied to all three datasets. So, after performing our analysis and generating plots for the GDP and forest coverage dataset, we generalized all of our code by breaking up the various steps of our analysis into functions. In doing so, we were then easily able to input the other two datasets into our functions to perform a consistent analysis across all datasets. Ultimately, our initial work plan suggested a much more inefficient approach to our analysis, but we had minimal issues deviating from our work plan to implement a more efficient approach.

9. Testing

Given that our analysis focused on identifying relationships and trends between several variables, we lacked a conventional means of verifying the accuracy or validity of our results because there were no open source ‘true’ or ‘accepted’ values to check our calculations with. Instead, we recalculated some of the values such as net change and r^2 and compared those values to what was found in our data frames for a few select countries. In particular, we chose to check the results for Afghanistan, Yemen, and Somalia because those countries had more missing data than others and we suspected that our calculations for these countries would be more prone to error. In order to check the validity of our plots, we simply just plotted smaller versions of our plots, only using a few countries at a time, to more easily verify that we were plotting our data correctly and that we could identify any correlations. Also since alot of our plots are interactive we fan filter down and select and individaul country and see if its plotting it right by just comparing it our Dataframes. We where also able to verify all the R Squared values by comparing it the R square values from plotly as plotly calculates it separately in a different way so when they line up we can be confident that it is the right value.

10. Collaboration

In terms of collaboration, we did not use any outside sources except for the doccumation for plotly and the doccumation for the World Bank Api.