A-4 Exploratory Visual Analysis

- Abhinav Duvvuri

1. Introduction

The goal is to analyze the World Development Indicators dataset and generate visualizations to explore the data and formulate inquiries. This dataset holds a lot of indicators, so I first looked through them to find interesting indicators using which I came up with a question. Then, I started extracting, filtering and cleaning the data with Python, a programming language widely used in data processing. Finally, I used Tableau, a software which can produce interactive data visualizations, to build the visualizations used in this analysis. This assignment is broken down into four sections: an Introduction, a Data Profile, a Question Exploration, a Conclusion and a Reference.

2. Data Profile

The dataset "World Development Indicators" compiled by the World Bank is an internationally comparable high-quality dataset focusing on relevant global development indicators. The database has over 1,400 time series indicators for 217 economies with data from 1960(for many indicators) until 2022. The dataset has **392883 rows and 67 columns**.

The dataset can be downloaded as a zip file of an excel sheet which is **76.1MB** in size. The Excel sheet has 6 sheets named Data, Country, Series, Country-Series, Series-Time and Footnote. The Data tab has the actual data of interest for this analysis. The columns have Country Name, Country Code, Indicator Name and Indicator Code which are **qualitative** and values of each indicator for the years **1960-2022**(each as a separate column) which are **quantitative**. These indicators cover 17 broad topics.

While the data is very well-compiled, there are a lot of null values. Many indicators only **have data once every few years**, which is expected since some indicators are only calculated periodically. Also, indicators **before 1990** seem to have **a lot more null values** than more recent indicators.

Another issue I noticed is that some **indicator measurements were stopped** after a few years while a **few others were renamed** as different indicators and continued. The user must carefully read the documentation supplied to understand these changes and process the data effectively. Though the documentation supplied was comprehensive, I **would have appreciated in-line documentation** to make it easier to read.

3. Question Exploration

3.1 Initial Thoughts and Inspiration

The number of indications in this dataset overwhelmed me the first time I downloaded and accessed it. However, after some investigation and reading the documentation, I gained a general understanding of what the dataset held. I was able to appreciate the importance of having such vast compilations of data and how it helps gain valuable insights at a global scale. I also appreciate the fact that it is open to everyone.

Initially, I tried to group different indicators into categories to find topics. But there were too many indicators to do this by hand, so I quickly loaded the data into a Jupyter Notebook and tried looking at the various indicators sorted by categories. After some inspection, one particularly interesting indicator I found was "Access to electricity (% of the population)". This took me back to my school days in India where I had learnt that many rural areas did not have access to electricity. I wanted to find the current relevance of that fact. While

researching this indicator in the dataset, I found that it was also an indicator used to measure the "Affordable and Clean Energy" goal of the "2030 Agenda for Sustainable Development Goals (SDG)" initiative. So, after reading about the SDGs, I changed my question to be "**Has the world made significant strides towards achieving the goal of affordable and clean energy outlined in the Sustainable Development Goals?**" While answering this question, I also answered other interesting and relevant questions that arose during the process.

3.2 Analysis

The sustainability development goal "Affordable and Clean Energy" has three main subparts. In this analysis, however, we will investigate the first two parts (to make it a little less complicated) of the goal which are to ensure **universal access** to affordable, reliable and modern energy services and to increase substantially the **share of renewable energy** in the global energy mix.

I first analyzed each of these subparts separately and when put together, I was able to answer my initial question.

3.2.1 Have we achieved universal access to electricity?

I first wanted to look at how people in different countries had access to electricity. I plotted Fig-1 which shows the percentage of the population with access to electricity in different countries in 2020 sorted from lowest to highest. Though the plot was a little congested, it was easy to infer that **we have not yet achieved universal access to electricity**. Further, majority of African population doesn't have universal access to electricity. To visualize this better, I plotted a choropleth (Fig-2) of the world map with a lighter color signifying less access to electricity. It was clear now that even though most countries had over 90% percentage of people with access**,** many **Africans in the Sub-Saharan region had little access to electricity.** The same was true with a few South Asian countries. This choropleth gave me the hint that I should be dividing the world into larger regions to analyze better.

3.2.1.1 Analyzing the Progress towards Access of Electricity over the years

My next question was: **How have these regions progressed over the years?** I was more interested in Sub-Saharan Africa and South Asian regions as these are the last regions to get access to electricity and once, they do, this sub goal can be fulfilled. Fig 3 shows how the 'Access to electricity (% of the population)' indicator varied from 2000 to 2020. I chose this period since it is the most relevant. The year 2021 was not chosen since it did not have values for all countries.

From Fig 3, although the Sub-Saharan African region has made satisfactory progress (25.65% - 48.48% over 20 years), **more than half the population didn't have access to electricity even in 2020**. This shows that most of the efforts of electrification must be put in place here to increase access to electricity overall.

The next two subsections are a small detour from our final goal. However, they provide meaningful insights into the results inferred from Fig-3 and answer my initial rural-urban electricity gap inquisition.

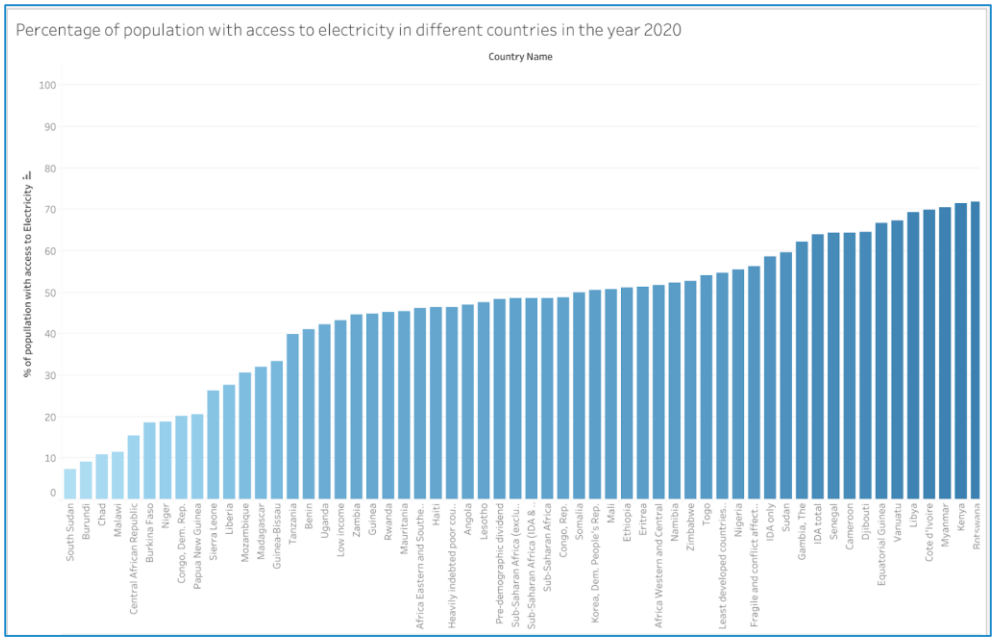

Figure 1: % population with access to electricity in different countries

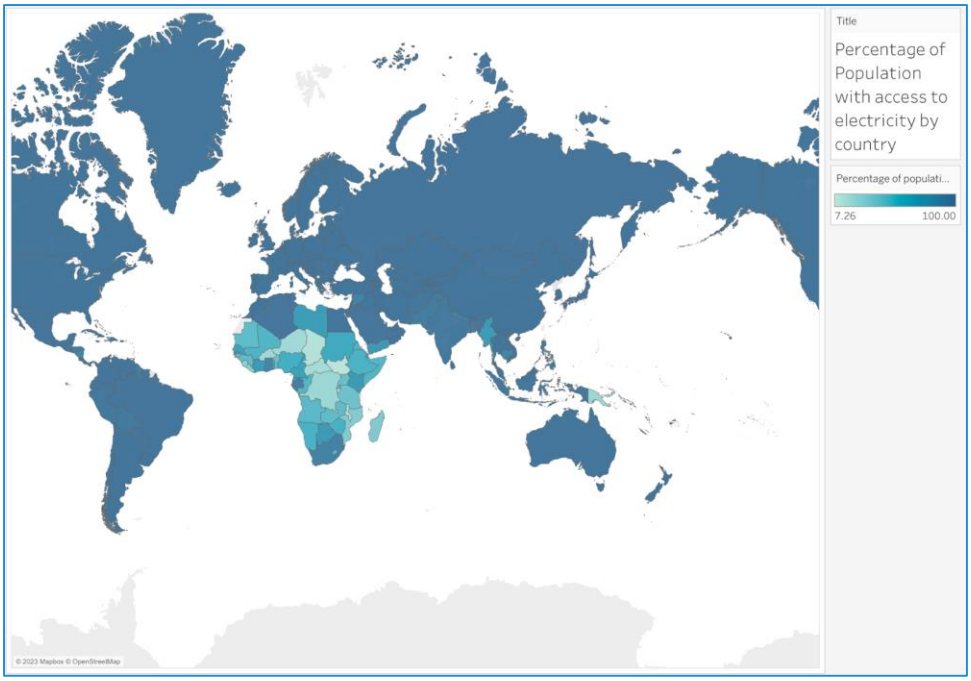

Figure 2: Choropleth map showing %of population with access to electricity

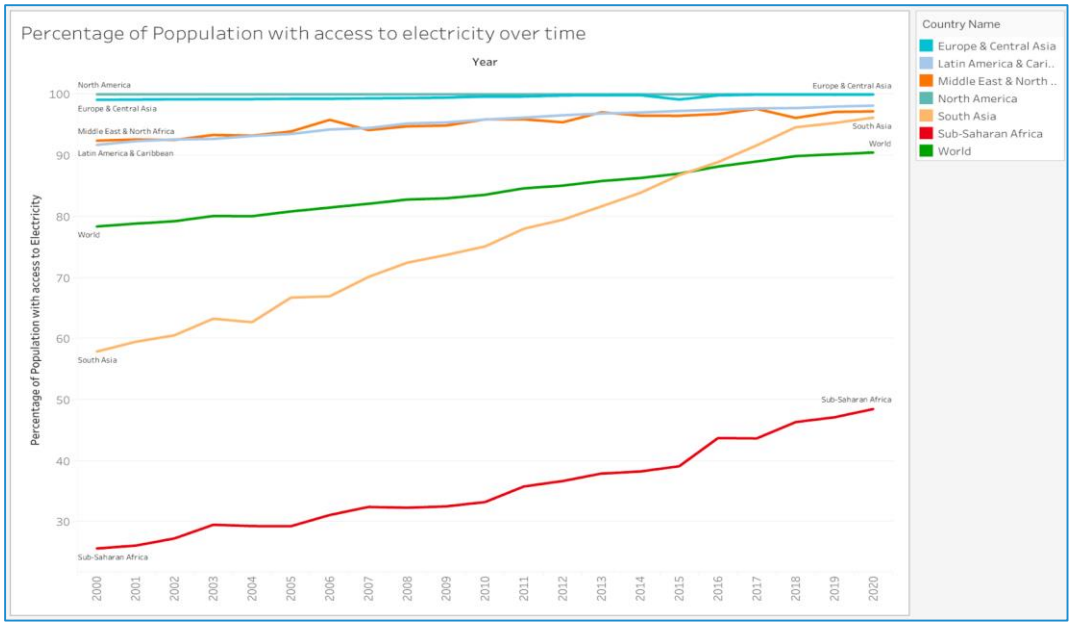

Figure 3: % of population with access to electricity from 2000-2020

3.2.1.2 Hypothesis – India's quick progress in electrification is the key for South Asia's progress

An interesting observation I made from Fig-3 was that South Asia made the highest progress from 57.89% in 2000 to 96.19% by 2020. I hypothesized that this was due to India's significant electrification process in the last two decades. To prove this, I made two plots. The first one shows how access to electricity increased in India and South Asia from 2000 to 2020(Fig-4). We can clearly see that the **as India's line grows over time, South Asia's line grows along with it** as if mimicking it. The pie-chart (Fig-5) shows the percentage of Indian population in South Asia emphasizing that due to the **high population percentage**, **the increase in access to electricity in India drives the increase in the overall region**. The above two figures together prove my hypothesis.

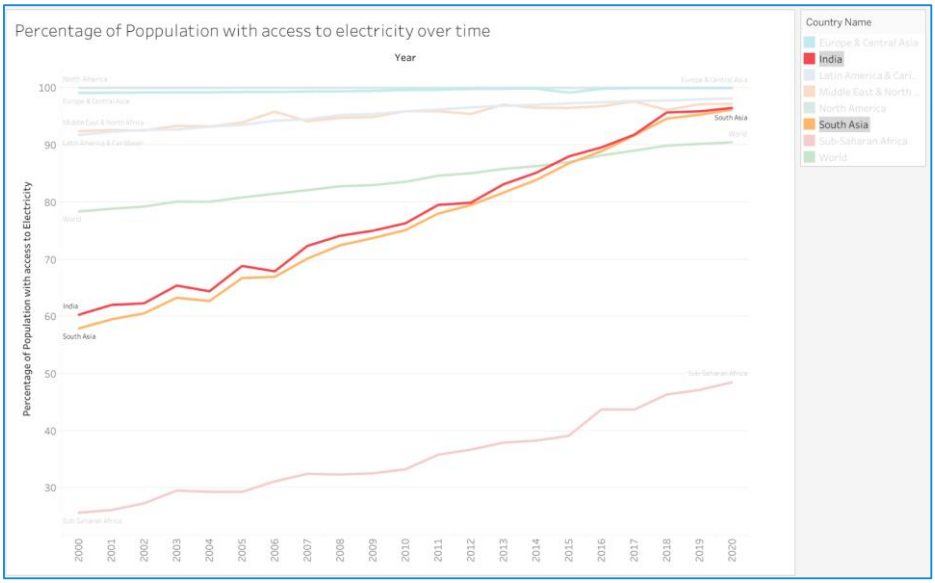

Figure 4: % of population with access to electricity from 2000-2020

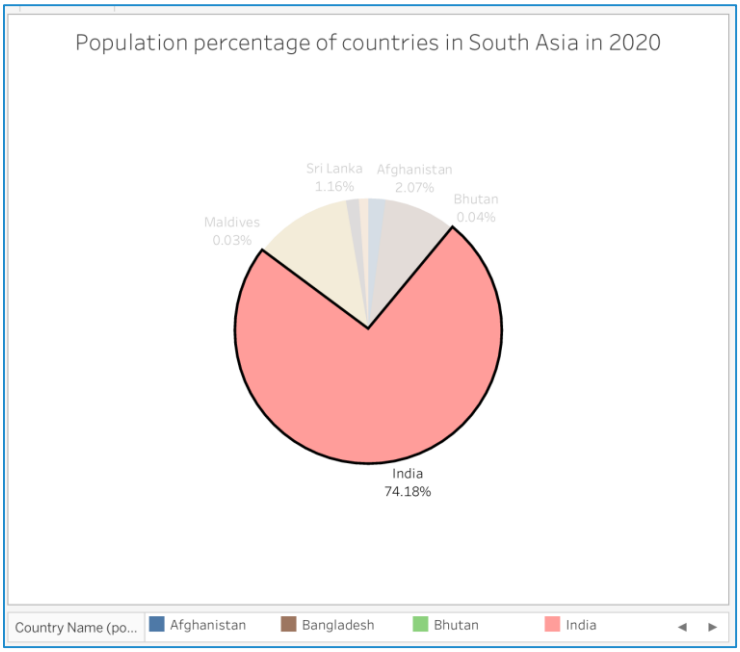

Figure 5: Population % of countries in South Asia in 2020

3.2.1.3 Urban-Rural gap in access to electricity

The next question I had was the split between the urban and rural populations without access to electricity. I used a bar chart to visualize the same for the year 2020. From Fig 6, we can see that almost **17% of the world's rural areas did not have access to electricity**. On the contrary, **only 2.7% of the world's urban population did not have electricity**. This highlights a **significant urban-rural gap**. Most of North America, Europe and Central Asia had electricity in both areas. Latin America, South Asia, North Africa and the Middle East had a high gap between rural and urban populations. However, it was **Sub-Saharan Africa** which had over **71%** of its **rural population** without electricity and **21.56%** in **Urban areas**.

3.2.2 Renewable energy

To understand how much of the energy consumption was renewable, I plotted Fig 7 showing the percentage of renewable energy consumed between 2001-2020. While the **world average was around 19.7%**, the **Sub-Saharan Africa** region had the least access to electricity actually **had the highest percentage of renewable energy consumption**. Also, **in Europe, the Middle East and North America** where almost everyone has access to electricity, most of it comes from **non-renewable sources**. However, one positive thing to note is that **the curve is moving upwards for almost all regions after 2018** meaning most countries are trying to use more renewable sources for electricity.

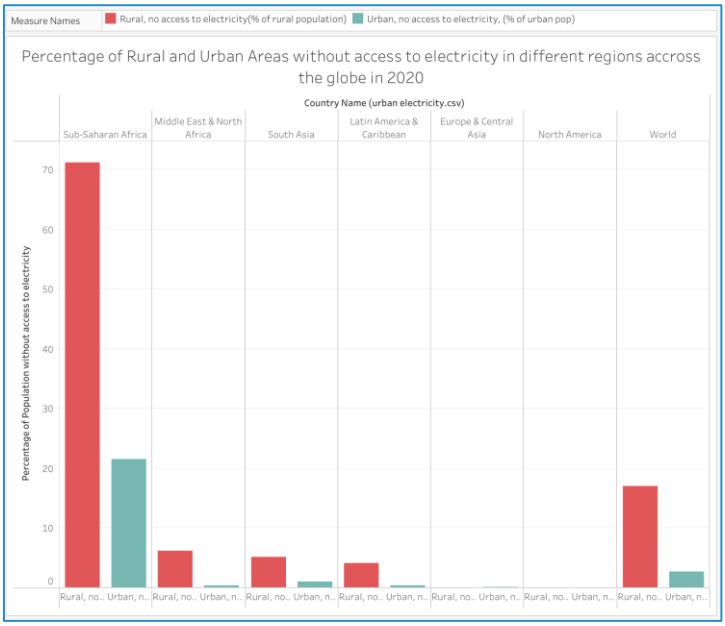

Figure 6: % of rural and urban areas without access to electricity in 2020

3.2.3 Combining access to Electricity and Renewable Energy consumption

One question I had after seeing this graph was: Why did the consumption of renewable energy decrease in South Asia from around 50% in 2000 to 35% in 2018 and **if it was correlated in any way with the increased number of people having access to electricity?** Is it true for other regions also? Further, I wanted to understand the positive shift (Fig-7) many regions had after 2018. Finally, I wanted to answer the main question of How the world progressing towards the Affordable and Clean Energy Goal?

To answer these questions, I made Fig-8, my main visualization. It is a plot between the percentage of the population with access to electricity and the Percentage of renewable energy consumed **over time**. I have used even year values as the plot points from 2000 to 2020 to improve excellence and understandability while still preserving integrity. I felt this style of plotting promotes less paper to ink ratio while improving cognition. The plot has interesting results. Firstly, **there is certainly a correlation between access to electricity and renewable energy consumption**. **North America, Europe and Central Asia,** where almost 100% of the population have access to electricity, consume one of the **lowest percentages of renewable energy**. Thie correlation is further fortified by the fact that when **South Asia** saw its **highest increase in the percentage of the population with access to electricity,** there was a visible **decrease in the percentage of renewable energy**. This shows that most of the access was given through renewable sources. East Asia and the Pacific regions also have a similar story. Further, in **Sub-Saharan Africa**, there was a **decrease in the share of renewable energy** until around 2014 which further emphasizes the point.

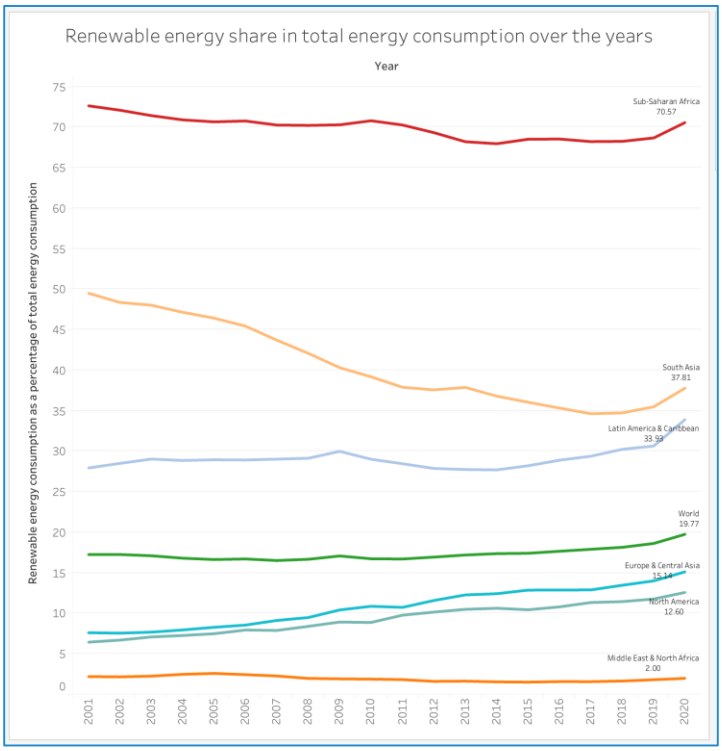

Figure 7: % of Renewable energy as a % of total energy consumption over time

To answer the final question, the two main indicators which I needed were "access to electricity" and "share of renewable energy" used. That is exactly Fig-8. To meet the goal, we need to Maximize these two indicators which can be reduced to moving to the top right of this plot. It is intuitively clear from the plot that **all regions have started moving towards the top right.** Many of them have even taken a turn in the opposite direction to move this way. This gives us **a positive sign** that all the regions and hence the **world are putting efforts into making energy both available to everyone as well as clean**. That said **the increase is not uniform and equally distributed** across different regions.

4. Conclusion

To conclude, **yes, the world has started making significant strides towards achieving the goal of affordable and clean energy outlined in the SDGs**. This is **particularly true from 2018** though some regions started ahead in time. The global **share of renewable energy** in the power mix has increased from **16.7% consumption in 2010 to 19.77% in 2020.** In addition, the percentage of people with access to electricity **has increased by nearly 12% since 2010**. That said, significant efforts must be put in to achieve this goal by 2030. Further, inferring from Fig-8, **all the regions have changed course and are moving towards more accessible and renewable energy which will likely yield a quicker growth in the coming years.**

Some other notable conclusions drawn in other parts of this analysis include the **rural-urban gap** in access to electricity and the necessity to put in **more efforts in Sub-Saharan Africa region to increase access to electricity**. I also explored how India's quick progress in electrification was the key for South Asia's rapid increase in percentage of people with access to electricity.

While formulating this assignment's question, I had to consider several indicators, reformulate my question and backtrack multiple times. Additionally, I also tried animations in tableau and played around with several other

story lines. I came up with a few other questions through this process which I presented in this text. I used python for data-processing. The libraries in python along with the class lecture notes and Jupyter notebooks made it easier to navigate my way through syntax. For the plots, I used tableau since the plots were intuitive and easy to make. However, I am still skeptical if tableau will handle more complex multivariate plots. But that is an exploration for another day! I thoroughly enjoyed the process of this exploration.

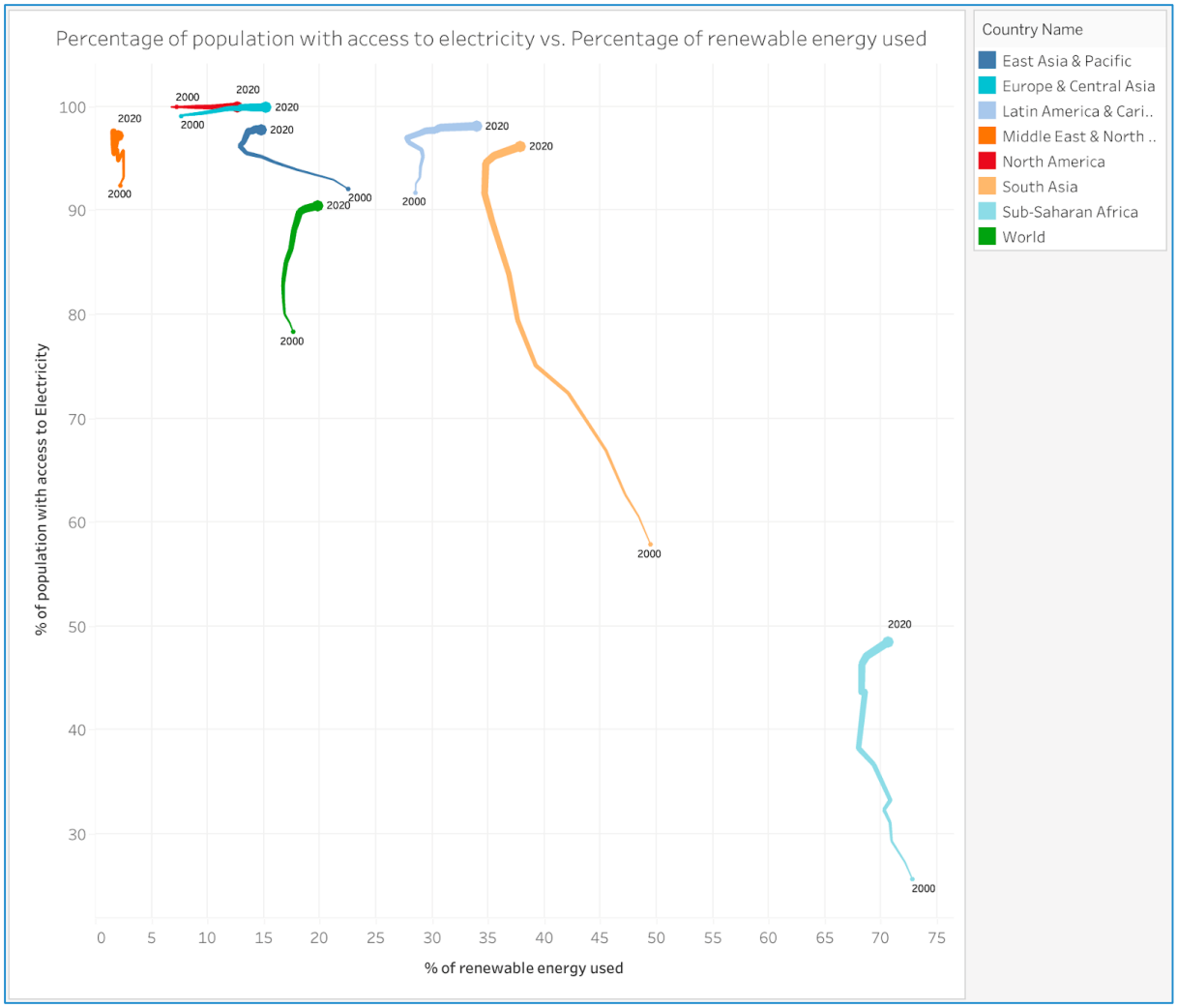

Fig 8: % of population with access to electricity vs. Percentage of renewable energy used over time

5. References

1. World Development Indicators: https://datatopics.worldbank.org/world-development-indicators/
2. Global indicator framework for the Sustainable Development Goals: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202023%20refinement_Eng.pdf
3. Tracking SDGs: https://trackingsdg7.esmap.org
4. Tableau Help: https://help.tableau.com/current/pro/desktop/en-us/viewparts_marks_marktypes.html
5. Lecture slides for Data 511: Data visualization for data scientists.
6. Visualizing Data by Ben Fry. Copyright 2008 Ben Fry, 978-0-596-51455-6.