Identifying Trends in Renewable Energy Market by Analyzing the Evolution of Energy Finance, Production, and Technologies

Britain Van Bergeyk
Matthew Haddad
Misha Salykova

University Of Calgary DATA 604 Winter 2024: Working with Data at Scale

Introduction

The world is witnessing a monumental shift towards renewable energy as a response to climate change and sustainability concerns [United Nations (n.d.)]. As a result, understanding the dynamics of the renewable energy market becomes crucial. This project aims to analyze the evolution of energy finance, production, and technologies to identify trends and patterns that define the renewable energy landscape globally.

Our group is committed to unraveling the intricate dynamics of energy financing, production, and technologies, with a distinct emphasis on the transformative journey of renewable energy. Our mission is to harness the power of data science to meticulously analyze and interpret the evolving trends, innovations, and challenges shaping the energy landscape (with a focus on renewable energy) over time. By providing actionable insights and empirical evidence, we aspire to empower stakeholders to navigate the complexities of the energy sector, drive informed decision-making, and foster sustainable solutions that propel the world towards a greener and more resilient future.

Individual Datasets

Five datasets were used in the analysis of the energy sector.

1. *Renewable Energy Statistics 2023*, *International Renewable Energy Agency (IRENA)*, *Abu Dhabi & IRENA (2023)*;
2. R*enewable Capacity Statistics 2023, International Renewable Energy Agency (IRENA), Abu Dhabi & IRENA (2023), Renewable Energy Statistics 2023, International Renewable Energy Agency (IRENA), Abu Dhabi*
3. *The Organisation for Economic Co-operation and Development (OECD) and IRENA Public Finance Database, 2023, The international Renewable Energy Agency (IRENA)*
4. *Country Codes and Coordinates,* https://gist.github.com/tadast/8827699
5. *The World Bank DataBank World Development Indicators, The World Bank Group (2024),*

Dataset	Source	License	Format	Key Columns	Dimension (Column*Row)	Description	Responsible Person
Electricity Installed Capacity (MW) by Region, Technology, Year	IRENA	CC BY-SA	CSV	Region/country/area, Technology, Year	4483*25	The dataset provides information on Electricity Installed Capacity by by Region, Technology, Year	Misha
Electricity Production Statistics	IRENA	CC BY-SA	CSV	Country, Technology, Year	97888*4	This dataset contains information on Electricity Production (GWh) by Country, Technology, and Year	Britain
Public Investments (2021 million USD) by Country/area, Technology and Year	IRENA	CC BY-SA	CSV	Country, Technology, Year	4445* 23	This dataset contains information on Public Investments into renewable technology (millions USD) by Country, Technology, and Year	Matthew
World Development Indicators	World Bank Open Data	CC BY 4.0	CSV	Country Name, Country Code, Year, Population	4992*5	This dataset contains Population of Countries by Year	Collective
countries_codes_and_coordinates.csv	gist.github.com	CC BY 4.0	CSV	Country,Alpha-3-code,latitude,longitude	263*6	Dataset contains location and code data for all countries on earth	Matthew

Trends in Renewable Electricity Capacity around the World

By examining changes in installed capacity, the dominant renewable energy technologies, distribution shifts over time, and disparities between various regions, we aim to understand the trends and factors influencing renewable energy deployment globally. We will investigate the growth of renewable electricity capacity worldwide and its implications. Understanding these trends is crucial for policymakers, energy analysts, and stakeholders to assess progress towards sustainable energy goals.

Dataset: The primary dataset for this analysis is sourced from the International Renewable Energy Agency (IRENA). It includes installed renewable electricity capacity (MW) by Region/country/area, Technology, and Year. The dataset is licensed for public use and downloaded in CSV format. Link to Source: https://pxweb.irena.org/pxweb/en/IRENASTAT/IRENASTAT__Power%20Capacity%20and%20Generation/Region_ELECSTAT_2024_H1.px/

The Final Dataset "Renewable Energy Capacity.csv" used in this notebook can be found here: https://github.com/misha-salykova/604project/blob/main/Renewable%20Energy%20Capacity.csv

Data Wrangling Steps:

1. Import libraries

2. Read table into pandas dataframe & clean the table into desired format

3. Establishing a connection to my MySQL database server

4. Load the dataframe into sql database

6. Splitting Table "renew_energy_cap" into separate tables by "Region/country/area" column for readability and to make it more convenient to use sql queries for each of the guiding questions related to either World, Region or Countries. As a result "renew_energy_cap" was split into three tables: world_totals, regions_table, countries_table.

7. Creating sql queries to answer the guiding questions

Fist guiding question we want to answer using this dataset is:

How has the total amount of renewable electricity capacity shifted over time in the world?

To answer this question we used the following query: (Calculate the total renewable energy capacity for all technology types for years 2000-2022 and see how has the total of renewable electricity capacity shifted over time in the world)

```
world_totals_query = """
SELECT
    `Region/country/area`,
    'Total Renewable Energy Capacity' AS `Technology`,
    SUM(`2000`) AS `Total_2000`,
    SUM(`2001`) AS `Total_2001`,
    SUM(`2002`) AS `Total_2002`,
    SUM(`2003`) AS `Total_2003`,
    SUM(`2004`) AS `Total_2004`,
    SUM(`2005`) AS `Total_2005`,
    SUM(`2006`) AS `Total_2006`,
    SUM(`2007`) AS `Total_2007`,
    SUM(`2008`) AS `Total_2008`,
    SUM(`2009`) AS `Total_2009`,
    SUM(`2010`) AS `Total_2010`,
    SUM(`2011`) AS `Total_2011`,
    SUM(`2012`) AS `Total_2012`,
```

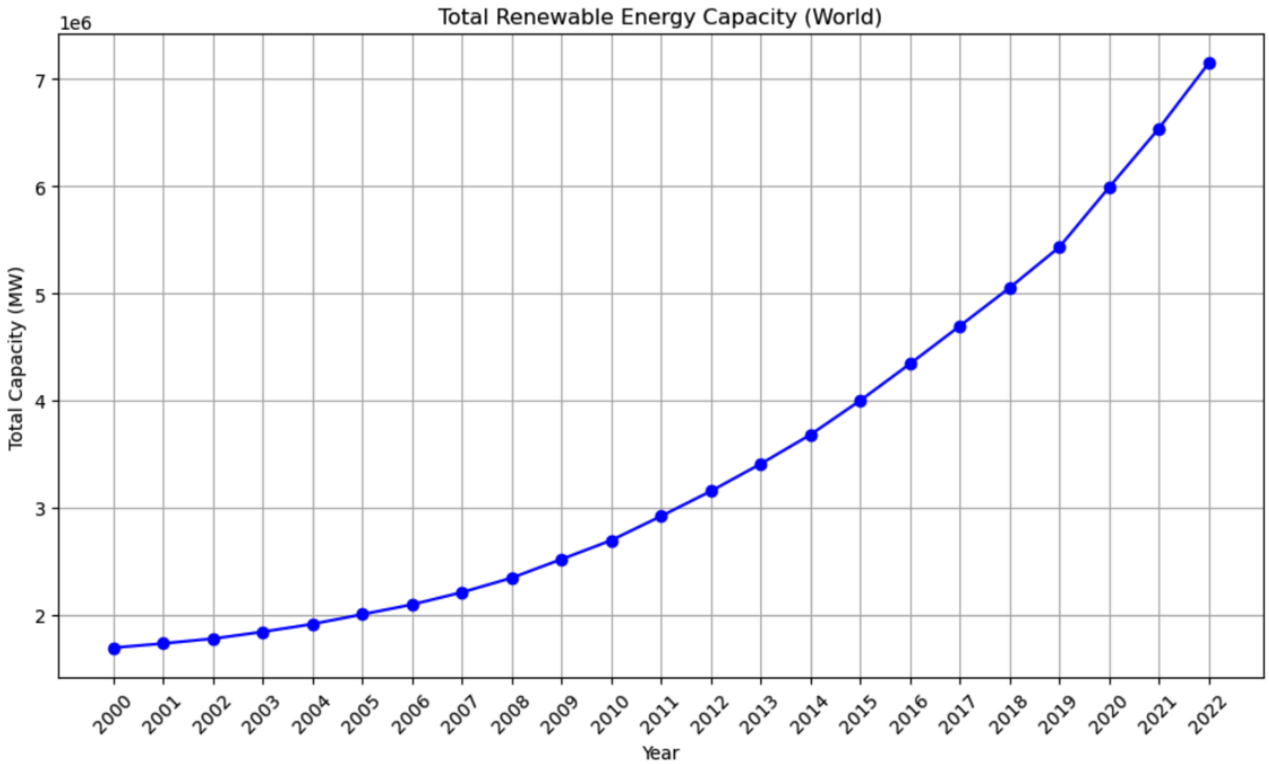
```
    SUM(`2013`) AS `Total_2013`,
    SUM(`2014`) AS `Total_2014`,
    SUM(`2015`) AS `Total_2015`,
    SUM(`2016`) AS `Total_2016`,
    SUM(`2017`) AS `Total_2017`,
    SUM(`2018`) AS `Total_2018`,
    SUM(`2019`) AS `Total_2019`,
    SUM(`2020`) AS `Total_2020`,
    SUM(`2021`) AS `Total_2021`,
    SUM(`2022`) AS `Total_2022`
FROM
    world_totals;
"""

result = pd.read_sql_query(world_totals_query, engine)
result
```

	Region/country/area	Technology	Total_2000	Total_2001	Total_2002	Total_2003	Total_2004	Total_2005	Total_2006	Total_2007	...
0	World	Total Renewable Energy Capacity	1691560.917	1731499.503	1776030.105	1839862.09	1912507.76	2002814.819	2094673.677	2208346.638	...

1 rows × 25 columns

...	Total_2013	Total_2014	Total_2015	Total_2016	Total_2017	Total_2018	Total_2019	Total_2020	Total_2021	Total_2022
...	3407062.588	3679517.566	3998424.851	4341422.973	4694315.512	5051219.452	5431035.426	5989128.519	6539127.084	7148705.973



From the results of this query we can see that there is a significant trend in the Renewable Energy Capacity worldwide over the past two decades:

Consistent Growth: The line graph above indicates a consistent upward trend in Total Renewable Energy Capacity from the year 2000 to 2022. This suggests that there has been a sustained effort globally to increase renewable energy production and infrastructure.

Substantial Increase: The Total Renewable Energy Capacity has significantly increased over this period, as evidenced by the rise from 1,691,560.917 MW in 2000 to 7,148,705.973 MW in 2022. This represents a substantial expansion of renewable energy resources on a global scale.

Quadrupled Capacity: The graph also highlights that Total Renewable Energy Capacity has quadrupled over the past two decades. This indicates a remarkable rate of growth, reflecting the increasing adoption and investment in renewable energy technologies worldwide.

Overall, these observations suggest a positive trend towards greater reliance on renewable energy sources for meeting global energy needs. It signifies a shift towards cleaner and more sustainable energy alternatives, driven by factors such as environmental concerns, technological advancements, and policy incentives.

What technologies contribute the most to the world's renewable electricity capacity and how did the distribution of technologies change over the years?

To answer this question we used the following query to compare the total contribution (in descending order) & distribution of technologies to the world's renewable electricity capacity for years 2000, 2010, 2020. (* Energy capacity totals per each technology are in MW)

```
tech_order_2000_query = """
SELECT `Region/country/area`,`Technology`, `2000`
FROM
    world_totals
ORDER BY `2000` DESC;
"""

result = pd.read_sql_query(tech_order_2000_query, engine)
result
```

	Region/country/area	Technology	2000
0	World	Hydropower	782156.806
1	World	Renewable hydropower	697190.256
2	World	Pumped storage	84966.550
3	World	Bioenergy	28428.425
4	World	Solid biofuels and renewable municipal waste	25509.466
5	World	Other solid biofuels	17657.000
6	World	Wind	16964.244
7	World	Onshore wind energy	16897.294
8	World	Geothermal	8272.950
9	World	Bagasse	4277.739
10	World	Renewable municipal waste	3564.880
11	World	Biogas	2889.805
12	World	Solar	1225.824
13	World	Solar photovoltaic	806.824
14	World	Concentrated solar power	419.000
15	World	Marine	237.750
16	World	Offshore wind energy	66.950
17	World	Liquid biofuels	29.154

```
tech_order_2010_query = """
SELECT `Region/country/area`,`Technology`, `2010`
FROM
    world_totals
ORDER BY `2010` DESC;
"""

result = pd.read_sql_query(tech_order_2010_query, engine)
result
```

	Region/country/area	Technology	2010
0	World	Hydropower	1025719.930
1	World	Renewable hydropower	925667.880
2	World	Wind	181090.194
3	World	Onshore wind energy	178034.844
4	World	Pumped storage	100052.050
5	World	Bioenergy	66099.381
6	World	Solid biofuels and renewable municipal waste	54664.405
7	World	Solar	41577.206
8	World	Solar photovoltaic	40311.386
9	World	Other solid biofuels	37025.000
10	World	Bagasse	10926.679
11	World	Geothermal	9914.276
12	World	Biogas	9634.150
13	World	Renewable municipal waste	6700.103
14	World	Offshore wind energy	3055.350
15	World	Liquid biofuels	1800.826
16	World	Concentrated solar power	1265.820
17	World	Marine	249.595

```
tech_order_2020_query = """
SELECT `Region/country/area`,`Technology`, `2020`
FROM
    world_totals
ORDER BY `2020` DESC;
"""

result = pd.read_sql_query(tech_order_2020_query, engine)
result
```

	Region/country/area	Technology	2020
0	World	Hydropower	1334350.424
1	World	Renewable hydropower	1213171.500
2	World	Wind	733564.569
3	World	Solar	728058.143
4	World	Solar photovoltaic	721545.959
5	World	Onshore wind energy	699195.470
6	World	Bioenergy	135539.406
7	World	Pumped storage	121178.924
8	World	Solid biofuels and renewable municipal waste	111674.695
9	World	Other solid biofuels	74003.000
10	World	Offshore wind energy	34369.099
11	World	Biogas	20676.027
12	World	Bagasse	20378.198
13	World	Renewable municipal waste	17067.295
14	World	Geothermal	14131.551
15	World	Concentrated solar power	6512.184
16	World	Liquid biofuels	3188.684
17	World	Marine	523.391

From the obtained data comparing the total contribution of different technologies to the world's renewable electricity capacity for the years 2000, 2010, and 2020, we can observe the following insights:

Dominant Technologies:

Hydropower: Hydropower consistently emerges as the leading contributor to the world's renewable electricity capacity across all three years, with substantial capacity increments from 2000 to 2020.
Renewable Hydropower: Similarly, the sub-division of hydropower, which is fully renewable hydropower maintains a significant presence, indicating the continued reliance on hydroelectricity for renewable energy generation globally. According to IRENA (2023),this designation emphasizes the sustainability aspect of hydropower generation, highlighting its reliance on naturally replenished water resources and its minimal environmental impact compared to fossil fuels. Renewable hydropower typically excludes controversial or environmentally damaging forms of hydropower, such as large-scale dams with significant ecological consequences or projects that disrupt natural river ecosystems.

Emergence of Wind and Solar Technologies:
Wind: The capacity of wind technologies, both onshore and offshore, shows remarkable growth over the years, indicating the increasing adoption of wind power as a significant contributor to renewable electricity capacity.
Solar: Solar technologies, particularly solar photovoltaic, exhibit exponential growth, reflecting the declining costs and increasing efficiency of solar energy generation. The substantial increase in solar capacity underscores the rising importance of solar power in the renewable energy landscape.

Steady Contribution of Bioenergy:
Bioenergy: Bioenergy, including solid biofuels, biogas, and liquid biofuels, maintains a steady contribution to renewable electricity capacity over the years. While not as dominant as hydropower, wind, or solar, bioenergy technologies continue to play a significant role in the renewable energy mix.

Shift in Distribution:

- The distribution of technologies has undergone a noticeable shift over the years, with increasing contributions from wind and solar technologies at the expense of traditional sources like hydropower.
- While hydropower remains a leading technology, its relative contribution to the total renewable electricity capacity has slightly decreased over time, indicating a diversification of renewable energy sources.
- Conversely, wind and solar technologies have experienced substantial growth in their contributions, reflecting advancements in technology, policy support, and declining costs.

Overall, the data suggests a dynamic and evolving landscape in the world's renewable electricity capacity, with a gradual transition towards a more diversified mix of renewable energy technologies, prominently featuring wind and solar power alongside traditional sources like hydropower and bioenergy.

Which countries have the highest renewable energy capacity in the world in total and per capita?

Let’s start with the Total Renewable Energy Capacity. To answer this question we used the following query:

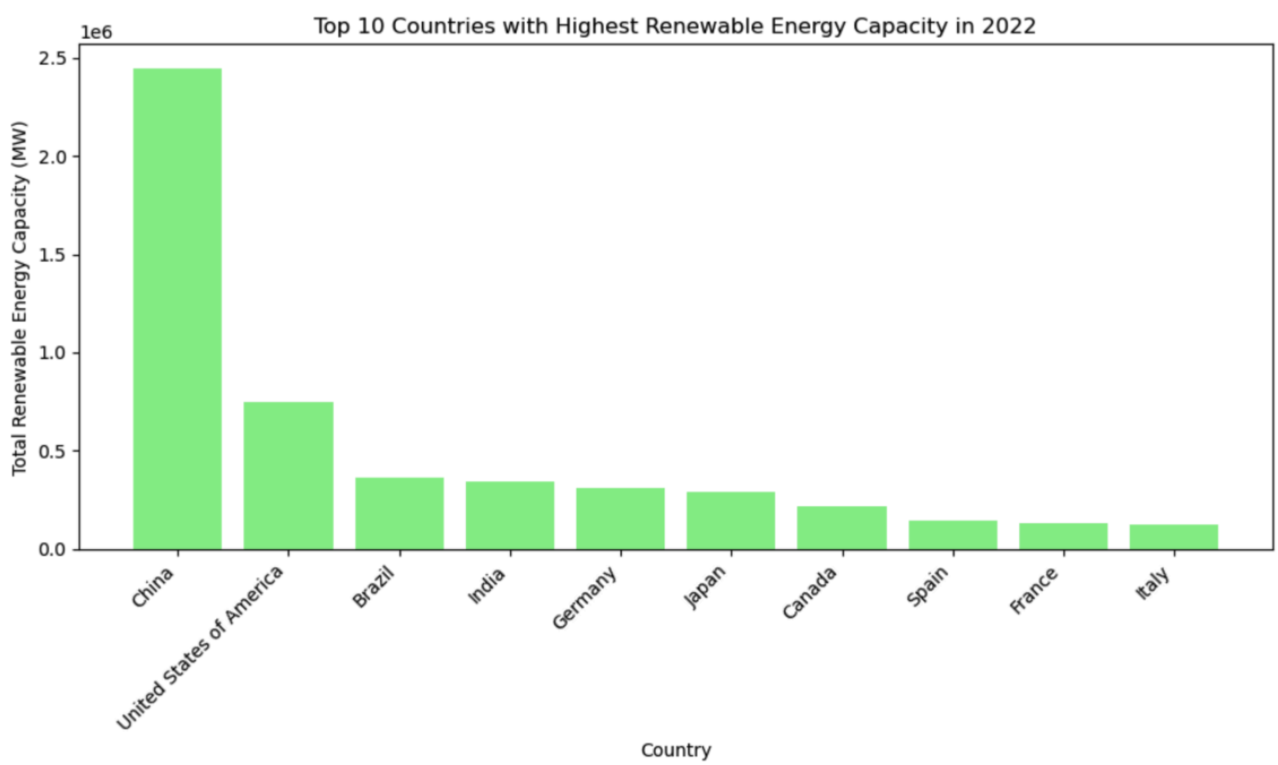
Calculate the total renewable energy capacity for all technology types for the year 2022 for each country. Then find the top 10 countries with the highest renewable energy capacity. (* Energy capacity totals are in MW)

```
top_10_countries = """
SELECT `Region/country/area`, SUM(`2022`) AS total_renewable_capacity_2022
FROM countries_table
GROUP BY `Region/country/area`
ORDER BY total_renewable_capacity_2022 DESC
LIMIT 10;
"""

result = pd.read_sql_query(top_10_countries, engine)
```

```
result
```

	Region/country/area	total_renewable_capacity_2022
0	China	2445334.182
1	United States of America	748479.142
2	Brazil	367225.489
3	India	346252.624
4	Germany	310128.500
5	Japan	291554.500
6	Canada	216473.632
7	Spain	143476.990
8	France	135461.670
9	Italy	127986.395



From the data, we can observe the following insights regarding the countries with the highest renewable energy capacity in the world in total for the year 2022:

1. **China**: China has the highest total renewable energy capacity, with a capacity of 2,445,334.182 MW in 2022. This indicates China's significant investment and development in renewable energy infrastructure.
2. **United States of America**: The USA ranks second in terms of total renewable energy capacity, with a capacity of 748,479.142 MW. Despite being the second-largest, the gap between China and the USA's renewable energy capacity is substantial.
3. **Brazil, India, Germany, Japan**: These countries follow with substantial renewable energy capacities, indicating a global trend towards renewable energy adoption and investment across diverse geographic regions.
4. **Canada, Spain, France, Italy**: These countries complete the top 10 list, showcasing a mix of developed and developing nations committed to expanding their renewable energy capacities.

Now let’s see which countries have the highest renewable energy capacity in the world per capita?

To do this we will join the “`countries_table`” table with the World Bank Data Bank population data. We have exported the previously cleaned table from the World Bank Data Bank population data and joined the tables based on country names. When performing data wrangling, we made sure that the naming conventions across the data sources were consistent.

```
#Per Capita

country_codes = pd.read_csv("CountryCodes.csv")
country_codes.to_sql('country_codes', engine, if_exists="replace",
index=False)

merge_query = """
SELECT country_codes.`Country Code`, `Population`, countries_table.*
FROM countries_table
INNER JOIN country_codes ON countries_table.`Region/country/area` =
country_codes.Country;
"""
merged_country_table = pd.read_sql_query(merge_query, engine)
merged_country_table
```

	Country Code	Population	Region/country/area	Technology	2000	2001	2002	2003	2004	2005	...	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	DZA	30774621	Algeria	Hydropower	276.6	276.6	276.6	276.6	276.6	276.6	...	227.6	227.6	227.6	227.6	228.0	228.0	228.0	208.6	128.9	128.9
1	DZA	30774621	Algeria	Renewable hydropower	276.6	276.6	276.6	276.6	276.6	276.6	...	227.6	227.6	227.6	227.6	228.0	228.0	228.0	208.6	128.9	128.9
2	DZA	30774621	Algeria	Pumped storage	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	DZA	30774621	Algeria	Marine	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	DZA	30774621	Algeria	Wind	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	10.2	10.2	10.2	10.2	10.0	10.0	10.0	10.0	10.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3235	VEN	24427729	Venezuela (Bolivarian Republic of)	Bagasse	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3236	VEN	24427729	Venezuela (Bolivarian Republic of)	Other solid biofuels	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3237	VEN	24427729	Venezuela (Bolivarian Republic of)	Liquid biofuels	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3238	VEN	24427729	Venezuela (Bolivarian Republic of)	Biogas	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3239	VEN	24427729	Venezuela (Bolivarian Republic of)	Geothermal	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

3240 rows × 27 columns

To answer this question "which countries have the highest renewable energy capacity in the world per capita?" we used the following query:

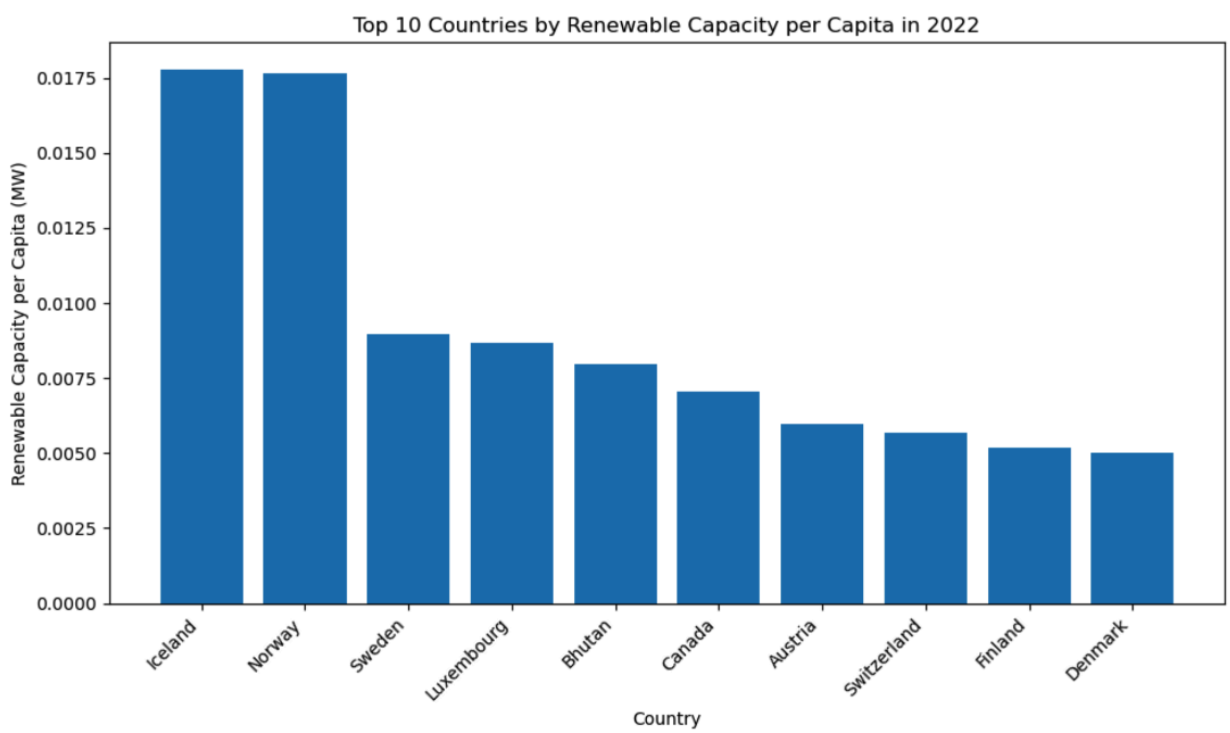
Calculate the total renewable energy capacity for all technology types for the year 2022 for each country and divide it by its population and name the column as `per_capita_renewable_capacity_2022`. Then find the top 10 countries with the highest renewable energy capacity per capita. (* Energy capacity totals are in MW)

```
merged_country_table.to_sql('merged_country_table', engine,
if_exists="replace", index=False)

top_10_countries_per_capita = """
SELECT `Region/country/area`, SUM(`2022`)/`Population` AS
per_capita_renewable_capacity_2022
FROM merged_country_table
GROUP BY `Region/country/area`
ORDER BY per_capita_renewable_capacity_2022 DESC
LIMIT 10;
"""

result = pd.read_sql_query(top_10_countries_per_capita, engine)
result
```

	Region/country/area	per_capita_renewable_capacity_2022
0	Iceland	0.017791
1	Norway	0.017673
2	Sweden	0.008968
3	Luxembourg	0.008664
4	Bhutan	0.007954
5	Canada	0.007055
6	Austria	0.005973
7	Switzerland	0.005676
8	Finland	0.005176
9	Denmark	0.005011



From the results of the query, we can observe the top 10 countries with the highest renewable energy capacity per capita in the world for the year 2022:

1. **Iceland:** Iceland has the highest renewable energy capacity per capita, with a value of 0.017791 MW per capita. This high value is likely due to Iceland's abundant renewable energy resources, particularly geothermal and hydroelectric power. Renewable sources accounted for nearly all of the electricity generated, primarily sourced from hydropower (73%) and geothermal power (27%). [Government of Iceland, n.d.]
2. **Norway:** Norway follows closely behind Iceland, with a per capita renewable energy capacity of 0.017673 MW. Like Iceland, Norway benefits from significant hydropower resources, contributing to its high renewable energy capacity per capita. In Norway, nearly 96% of electricity production relies on renewable hydropower, providing the country's industries with reliable access to affordable and eco-friendly energy. Especially for energy-intensive heavy industries, Norway boasts one of the lowest carbon footprints globally. (Business Norway, 2023)
3. **Sweden:** Sweden ranks third, with a per capita renewable energy capacity of 0.008968 MW. Sweden has made significant investments in renewable energy, particularly in wind and bioenergy, contributing to its relatively high capacity per capita. [IEA Bioenergy, 2021]
4. **Luxembourg:** Luxembourg has a per capita renewable energy capacity of 0.008664 MW, placing it fourth on the list. Despite its small size, this suggests that Luxembourg has been investing in renewable energy projects to reduce its carbon footprint.
5. **Bhutan:** Bhutan, with a per capita renewable energy capacity of 0.007954 MW, ranks fifth.
6. **Canada, Austria, Switzerland, Finland, Denmark:** These countries complete the top 10 list, each with relatively high renewable energy capacity per capita, reflecting their commitment to renewable energy adoption and sustainability.

Overall, these results highlight the importance of both natural resources and government policies in driving renewable energy adoption and achieving high renewable energy capacity per capita.

Comparing the results for countries with highest Total Renewable Energy Capacity and Per Capita Renewable Energy Capacity, we can argue that difference between the top 10 countries by total renewable energy capacity and those with the highest per capita capacity is influenced by factors such as resource abundance, population density, policy decisions, energy consumption patterns, and geopolitical and economic considerations. Countries with abundant resources and smaller populations may rank higher in per capita capacity, while larger nations with high energy demand may lead in total capacity despite lower per capita figures.

How does renewable energy capacity differ in various regions of the world?

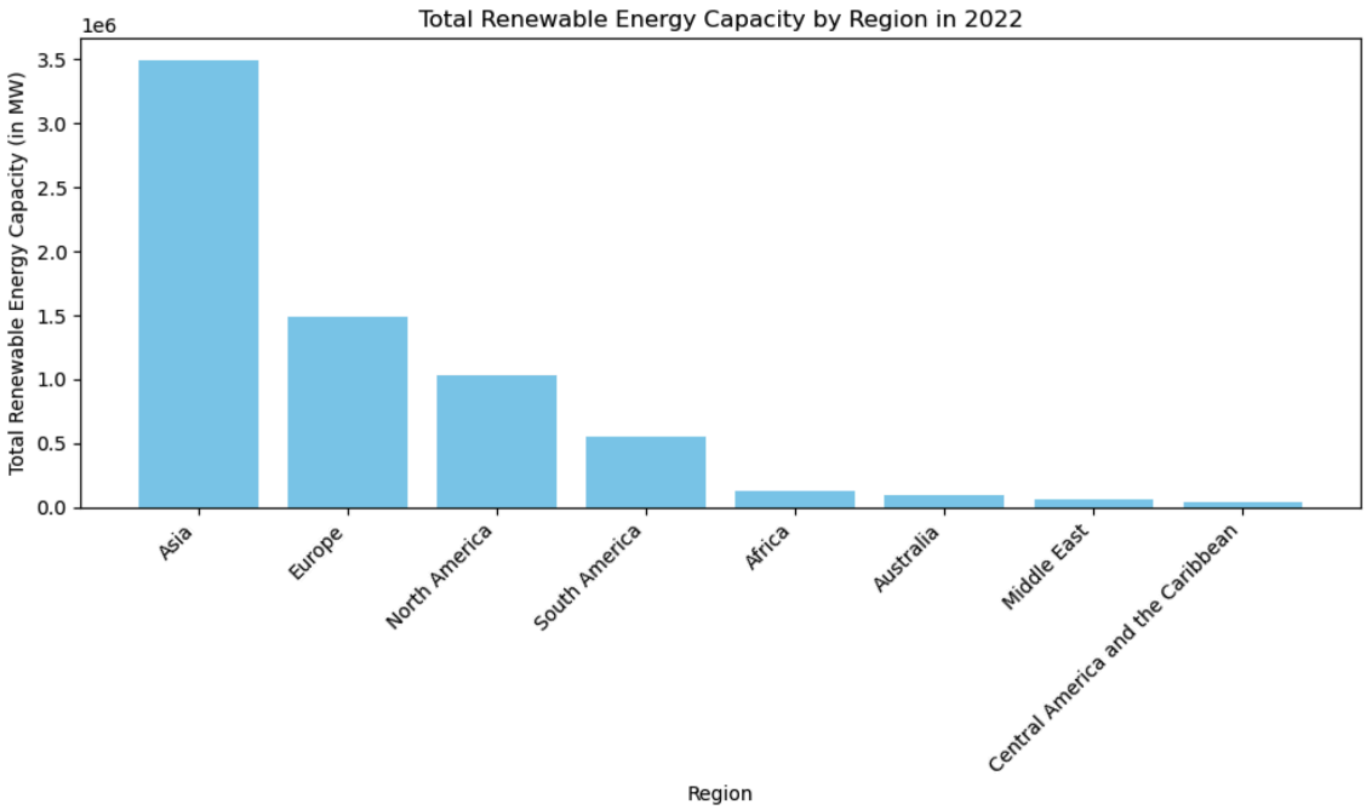
To answer these guiding questions let's create sql queries to:

Calculate the total renewable energy capacity for all technology types for year 2022 for various Regions (Africa, Asia, Central America and the Caribbean, Europe, Middle East, North America, Australia, South America) in the World and sort in descending order

```
Regions_query = """
SELECT `Region/country/area`, SUM(`2022`) AS
`Total_Renewable_Capacity_2022`
```

```
FROM regions_table
WHERE `Region/country/area` IN ('Africa', 'Asia', 'Australia', 'Central
America and the Caribbean', 'Europe', 'Middle East', 'North America',
'South America')
GROUP BY `Region/country/area`
ORDER BY `Total_Renewable_Capacity_2022` DESC;
"""
result = pd.read_sql_query(Regions_query, engine)
result
```

	Region/country/area	Total_Renewable_Capacity_2022
0	Asia	3490600.610
1	Europe	1493116.132
2	North America	1028966.486
3	South America	553234.145
4	Africa	124752.661
5	Australia	99932.980
6	Middle East	62042.454
7	Central America and the Caribbean	38367.894



The differences in renewable energy capacity across various regions are evident from the following insights gathered from the query results:

1. **Asia**: Asia leads with the highest total renewable energy capacity in 2022, indicating significant investment and development in renewable energy infrastructure. This aligns with Asia's status as a rapidly growing economic region with a large population and increasing energy demands. As previously discovered, with China being the highest contributor to the total renewable capacity in the region.
2. **Europe**: Europe follows closely behind Asia in total renewable energy capacity. This reflects Europe's strong commitment to the renewable energy transition, driven by ambitious policy frameworks, technological advancements, and widespread public support for sustainability initiatives.
3. **North America**: North America ranks third in total renewable energy capacity, with substantial contributions from countries like the United States and Canada. While North America has vast renewable energy potential, the distribution of capacity varies across different states and provinces due to diverse geographical and regulatory factors.
4. **South America**: South America emerges as a significant player in renewable energy, with notable contributions to total capacity. Countries like Brazil and Chile have made substantial investments in renewable energy projects, particularly in hydropower, wind, and solar energy. More on this will be covered in our investments section of the report.
5. **Africa**: Africa's renewable energy capacity is comparatively lower than other regions, highlighting the need for increased investment and development in the continent's renewable energy sector. Despite abundant renewable resources, challenges such as financing, infrastructure, and political instability could pose barriers to growth.
6. **Australia**: Australia's renewable energy capacity is comparatively lower than other regions, despite its vast renewable energy potential.

7. **Middle East**: The Middle East exhibits modest renewable energy capacity. While the region has historically relied on fossil fuels, there is growing interest in diversifying energy sources to mitigate environmental impacts and enhance energy security [Maguire, 2023].
8. **Central America and the Caribbean**: Central America and the Caribbean show lowest renewable energy capacity compared to all other regions.

These insights highlight the diverse landscape of renewable energy capacity worldwide, influenced by factors such as geographic location, resource availability, policy frameworks, economic conditions, and technological advancements.

Trends in Electricity Production around the World

After exploring the electricity capacity of renewables we will look at the Electricity Production Table from IRENA. This table considers production from not only renewables but other technologies as well such as oil and gas, coal, nuclear, etc. To gain certain insights based on populations (per capita statistics) the first job was to join this table with the World Bank Data Bank population data. This was performed in python using pandas. The reason we used pandas was because the tables required a significant amount of wrangling to allow them to be merged since the naming conventions across the data sources were not consistent.

Once we had the tables joined we began asking questions about the data, the first question we sought to answer was "On an annual basis, which technology produces the most electricity globally?". To answer this question we used the following query:

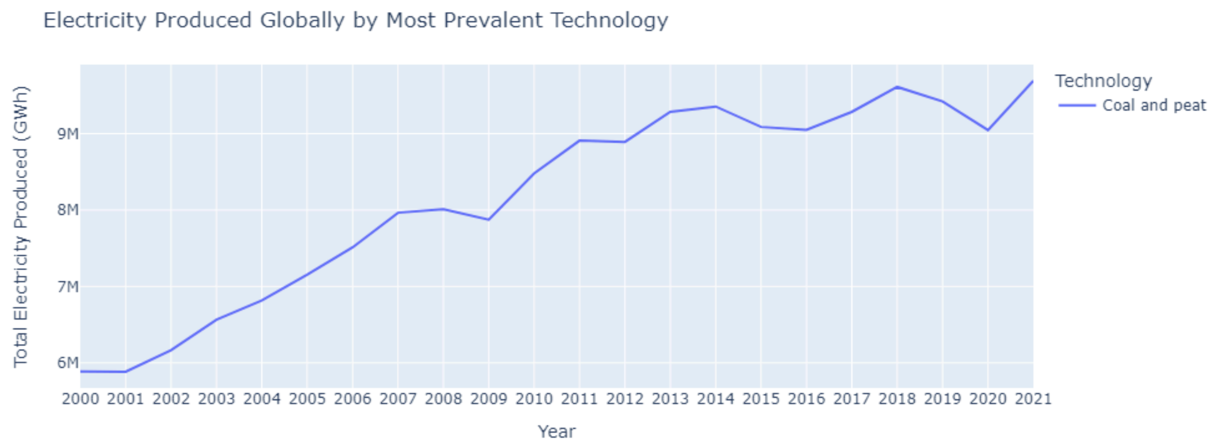
```
SELECT Time, Technology, SUM(`Electricity Production (GWh)`)
            AS `Total Electricity Production`
            FROM elec_and_pop
            WHERE NOT Time = 2022
            GROUP BY Technology, Time
```

```
mostproducerdf = firstquerydf.groupby("Time").max("Total Electricity
Production").merge(firstquerydf, how = "inner")
```

	Time	Technology	Total Electricity Production
0	2000	Biogas	13486.52
1	2001	Biogas	14963.91
2	2002	Biogas	16776.78
3	2003	Biogas	19328.11
4	2004	Biogas	20402.91
...	...	...	...
413	2017	Solid biofuels	336114.26
414	2018	Solid biofuels	349084.18
415	2019	Solid biofuels	369578.39
416	2020	Solid biofuels	373784.38
417	2021	Solid biofuels	404179.46

418 rows × 3 columns

Using the table above along with pandas we found the Technology that produced the most electricity in each year from 2000-2021 and determined that it was Coal and Peat that produced the most for the entire duration of the data. We can see a visualization of this production below:



It is heartening to see that in spite of constant growth in production by coal and peat, the growth rate has tapered to a relatively flat slope. This is quite desirable from a climate perspective since coal and peat burning produces some of the most CO2 per kWh of electricity produced [Quaschning, 2022].

The second question we were interested in from this dataset was "Which countries have the highest gross and per capita electricity production in 2021?". We decided to approach this question with the following query:

```
SELECT Country, Time, SUM(`Electricity Production (GWh)`) AS Total,
                   `Population, total [SP.POP.TOTL]` AS Population,
                   (SUM(`Electricity Production (GWh)`) / `Population, total
[SP.POP.TOTL]`) AS `Per Capita Total`
                   FROM elec_and_pop
                   WHERE Time = 2021
                   GROUP BY Country, Time,`Population, total [SP.POP.TOTL]`
                   ORDER BY `Per capita Total` DESC
```

The output of this query was the following table:

	Country	Time	Total	Population	Per Capita Total
0	Iceland	2021	19614.20	372520.0	0.052653
1	Norway	2021	157950.43	5408320.0	0.029205
2	Bahrain	2021	33376.80	1463265.0	0.022810
3	Qatar	2021	51918.14	2688235.0	0.019313
4	Kuwait	2021	80783.08	4250114.0	0.019007
...	...	...	...	...	...
202	Niger (the)	2021	536.14	25252722.0	0.000021
203	Chad	2021	335.54	17179740.0	0.000020
204	Guinea-Bissau	2021	24.78	2060721.0	0.000012
205	Somalia	2021	0.00	17065581.0	0.000000
206	South Sudan	2021	0.00	10748272.0	0.000000

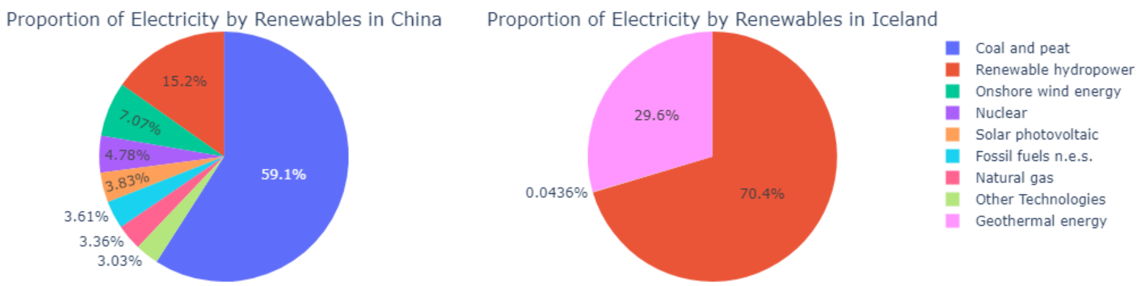
207 rows × 5 columns

To answer the question at hand we looked for the maximum value in the Total and Per Capita Total columns respectively. This gave us the result that China had the highest total production in 2021 and Iceland had highest per capita production. China having the largest gross production is very intuitive since it has the largest population. Iceland having the largest per capita production is directly in line with the observation in the previous section about Iceland having the largest per capita capacity.

The third question of this section relied on question two, it asks: “For those countries what proportion of their electricity production is provided by which technologies in 2021?”. The resulting query is shown below:

```
SELECT Country, Technology, `Electricity Production (GWh)`
                FROM elec_and_pop
                WHERE Time = 2021 AND Country IN ('China', 'Iceland');
```

This query gives us the production breakdown of both countries in 2021. We can visualize the results of this using plotly specifically with two pie charts:



China still relies heavily on Coal and Peat, however 30% of their electricity is produced by clean technologies. Iceland on the other hand generates the vast majority of their electricity using renewable technologies, specifically renewable hydropower and geothermal energy.

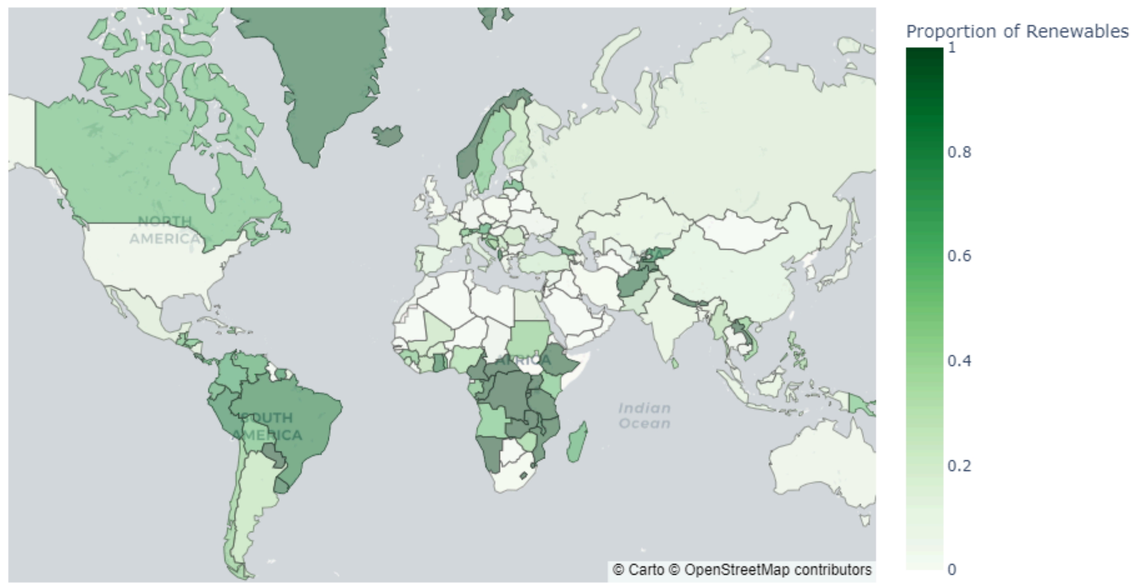
The fourth query we chose to run on this table was to answer the following question: "What proportion of the electricity production is provided by renewables vs. non renewables per country by year?". We first had to break the technologies into two sets, renewable and non-renewable, then we parsed the table into two subsets one with renewable technologies, the other with non-renewables. In these subsets we summed all the production, then we joined these two subsets into one table. This process can be seen in the code block below:

```
SELECT renew.Country, renew.`Country Code`, renew.Time, renew.`Total
Renewable Energy`, nonrenew.`Total Non-Renewable Energy`,
                  (renew.`Total Renewable Energy`/(nonrenew.`Total
Non-Renewable Energy`+renew.`Total Renewable Energy`)) AS `Proportion of
Renewables`,
                  (nonrenew.`Total Non-Renewable Energy`/(nonrenew.`Total
Non-Renewable Energy`+renew.`Total Renewable Energy`)) AS `Proportion of
Non-Renewables`
                  FROM (SELECT Country, `Country Code`, Time,
SUM(`Electricity Production (GWh)`) AS `Total Renewable Energy`
                  FROM elec_and_pop
                  WHERE Technology IN ('Solar photovoltaic',
                                        'Solar thermal energy',
                                        'Onshore wind energy',
                                        'Offshore wind energy',
                                        'Renewable hydropower',
                                        'Mixed Hydro Plants',
```

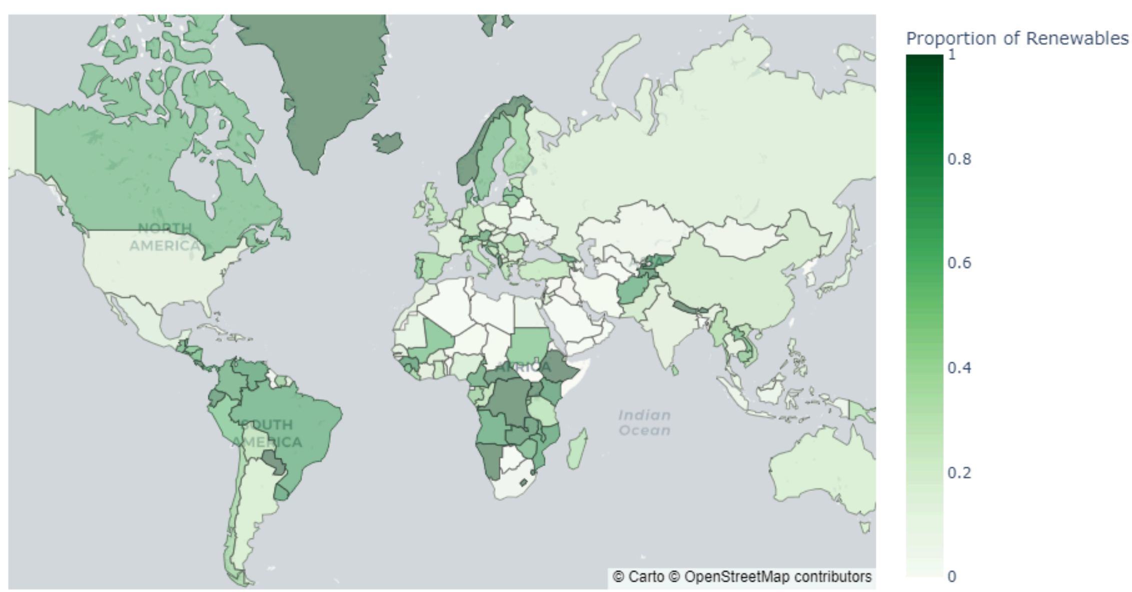
```
                                            'Pumped storage',
                                            'Marine energy',
                                            'Solid biofuels',
                                            'Renewable municipal waste',
                                            'Liquid biofuels',
                                            'Biogas',
                                            'Geothermal energy')
                                            AND NOT Time = 2022
                  GROUP BY Country, Time) AS renew JOIN
                  (SELECT Country, `Country Code`, Time, SUM(`Electricity
Production (GWh)`) AS `Total Non-Renewable Energy`
                  FROM elec_and_pop
                  WHERE Technology IN ('Coal and peat',
                                            'Oil',
                                            'Natural gas',
                                            'Fossil fuels n.e.s.',
                                            'Nuclear',
                                            'Other non-renewable energy')
                                            AND NOT Time = 2022
                  GROUP BY Country, Time) AS nonrenew
                  ON renew.`Country Code` = nonrenew.`Country Code` AND
renew.Time = nonrenew.Time
```

To answer the question we look at a Choropleth map throughout all the years (for the purposes of this report we will display only the years 2000 and 2021):

Proportion of Electricity Generation Which is Renewable in 2000



Proportion of Electricity Generation Which is Renewable in 2021



As we can see by comparing the two maps, there was a lot of change in the last 20 years. Central Eruope went from having sparse renewable energy to having a large chunk of its energy be renewable. China, which we saw earlier was the largest producer in 2021, has increased their renewable electricity production in the last 20 years. We can also see that Canada has increased the proportion of renewable electricity.

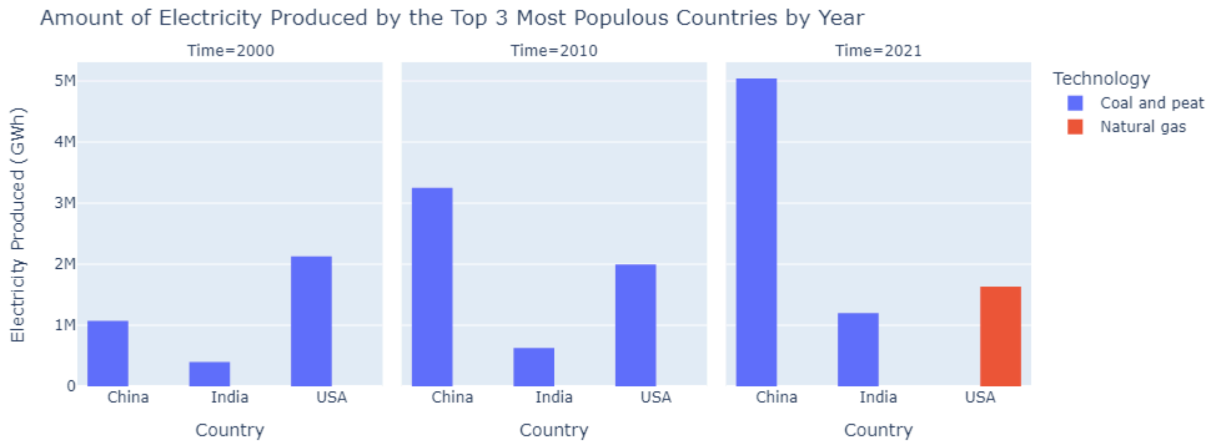
The final question of this section was “Which energy technology is most prevalent in the top 3 most populous countries in 2000, 2010, and 2021?”.
To answer this we first needed to determine which countries were the most populous in these 3 years:

```
SELECT Country, Time, `Population, total [SP.POP.TOTL]` AS Population
                  FROM elec_and_pop
                  WHERE Time IN (2000,2010,2021)
                  GROUP BY Country, Time
                  ORDER BY Population DESC
                  LIMIT 9
```

Next we use these three countries to determine which technology was most common in each countries and year.

```
SELECT prevenergy.Country,prevenergy.Time,techs.Technology,
prevenergy.`Prevalent Tech (GWh)`
                  FROM (SELECT Country, Time, Technology, MAX(`Electricity
Production (GWh)`) AS `Prevalent Tech (GWh)`
                  FROM elec_and_pop
                  WHERE Time IN (2000, 2010, 2021) AND Country  IN
('China', 'India', 'United States of America (the)')
                  GROUP BY Country, Time) prevenergy
                  JOIN (SELECT Country, Time, Technology, `Electricity
Production (GWh)`
                  FROM elec_and_pop
                  WHERE Time IN (2000, 2010, 2021) AND Country IN ('China',
'India', 'United States of America (the)')
                  GROUP BY Country, Time, Technology) techs
                  ON prevenergy.Country = techs.Country AND
prevenergy.`Prevalent Tech (GWh)` = techs.`Electricity Production (GWh)`
```

To visualize this query we use three bar graphs, one for each year. The color represents which technology was most prevalent in each country and the height is the amount of electricity produced by that technology.



We can see from the above figure that Coal and Peat based electricity increased in China and India and remains the largest technology in these countries. In the US, however, we see that Coal and Peatwas on the decline between 2000 and 2010, then sometime between 2010 and 2021 natural gas overtook Coal as the dominant technology in the United States. This is a good change since according to the Energy Information Administration natural gas produces about half the amount of CO2 per million Btu [US EIA, 2023].

Analyzing Public Investment Into Renewables by Country

Finally we examine public investment of renewables from the IRENA database. Like the other tables this dataset has investment from renewables and nonrenewable technologies. To answer some location based queries, this dataset was combined with a coordinates table that has the latitude and longitude for each country. This was joined by country code, as for future work, country code is much cleaner to work with. Furthermore, like the precious datasets, pandas was used to wrangle the dataset, removing instructional text from the IRENA database, renaming columns, and sectioning off renewable and nonrenewable data. The wrangling was done as such:

```
file = pd.read_csv("PublicInvestmentEnergy2.csv")

# change column 2 to Year
file = file.rename(columns={'Unnamed: 1': 'Year'})
file = file.rename(columns={'Unnamed: 0': 'Countries'})
df = file[:-2]
df.ffill(inplace=True)
df["Year"] = pd.to_datetime(df["Year"], format='%Y')

grouped = df.groupby(['Countries', 'Year']).sum()
sumdf = pd.DataFrame(grouped.sum(axis=1))
sumdf = sumdf.rename(columns={0 : "Public Investments (2020 million USD)"})
```

```
# Making the Technologies into their own field
tech = list(grouped)
yucky_tech = ['Coal and peat', 'Oil', 'Natural gas', 'Non-renewable municipal
waste', 'Fossil fuels n.e.s.']
cleantech = [i for i in tech if i not in yucky_tech]
```

Figure : Data Wrangling for Public Investment Dataset

After the wrangling, five guiding questions were queried using SQL. To do so, three tables were made:

- A table containing all technologies, the year, the country, and the investment.

```
# For ALL Technologies
melted_tech = pd.melt(df, id_vars=["Countries", "Year"], value_vars=tech,
var_name='Technology', value_name='Investment')
melted_tech.head()
```

- A table containing all nonrenewable technologies, the year, the country, and the investment.

```
# For NONRENEWABLE Technologies
melted_yucky_tech = pd.melt(df, id_vars=["Countries", "Year"],
value_vars=yucky_tech, var_name='Technology', value_name='Investment')
melted_yucky_tech.head()
```

- A table containing all renewable technologies, the year, the country, and the investment.

```
# For RENEWABLE Technologies
melted_cleantech = pd.melt(df, id_vars=["Countries", "Year"],
value_vars=cleantech, var_name='Clean_Technology', value_name='Investment')
melted_cleantech.head()
```

1) **How do renewable versus nonrenewable technologies compare in total investment?**

```
sql = """
        SELECT Countries, SUM(Investment) as Total_Investment
        FROM tech_table
        GROUP BY Countries
        ORDER BY Total_Investment DESC
        ;

    """
```

```
total_investment = pd.read_sql_query(sql, engine)
total_investment

total_investment.to_sql("total_investment",engine,if_exists="replace",
index=False)
total_investment = pd.read_sql_table("total_investment", engine)
total_investment.iloc[10:21]
```

To answer this question, we can sum investment over the entire period of examination and list the top countries in order. To better visualize the magnitude of difference between each country a bar graph was made.

	Countries	Total_Investment
0	Brazil	84963.57
1	Russian Federation (the)	34478.19
2	India	30332.02
3	Pakistan	27076.29
4	Indonesia	20343.43
5	United Kingdom of Great Britain and Northern I...	19267.91
6	Viet Nam	19069.11
7	South Africa	14207.46
8	Egypt	13738.13
9	Bangladesh	12701.45

Table: List of the top 10 countries by investment by their total investment in all energy sector related technologies from 2000-2021

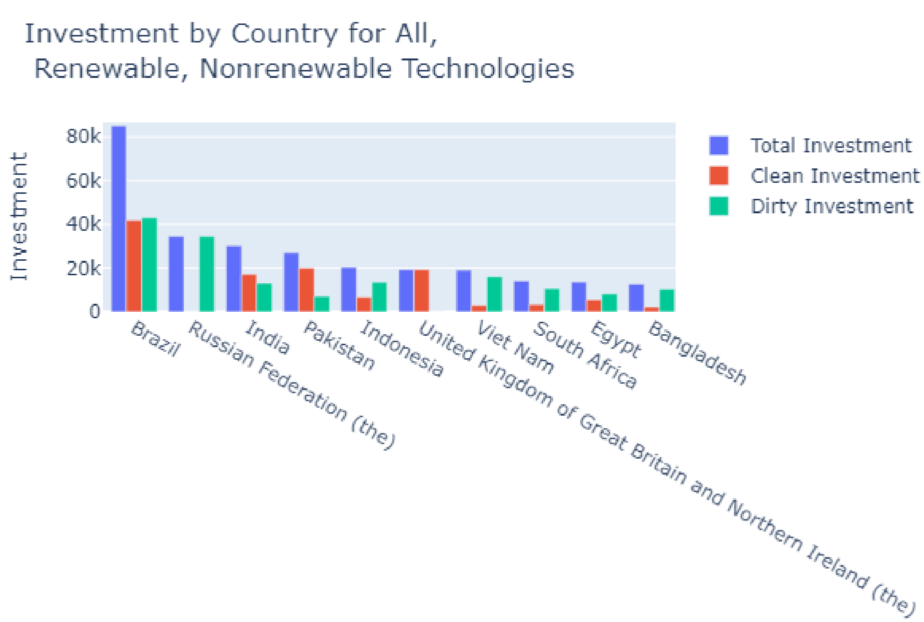


Figure : Top 10 countries by investment by their total, renewable, and nonrenewable investment in all energy sector related technologies from 2000-2021

This data shows us that Brazil, Russia and India are the top three investors in energy investment in the past ~20 years. Where over these years the investment into renewable and nonrenewable resources has stayed roughly equal with nonrenewable investment narrowly exceeds renewable investment.

```
sql = """
        SELECT Countries, SUM(Investment) as Total_Investment
        FROM tech_table
        GROUP BY Countries
        ORDER BY Total_Investment DESC
        ;

    """
```

```
sql2 = """
        SELECT Countries, SUM(Investment) as Total_Investment_Renewable
        FROM clean_tech_table
        GROUP BY Countries
        ORDER BY Total_Investment_Renewable DESC
        ;

    """
```

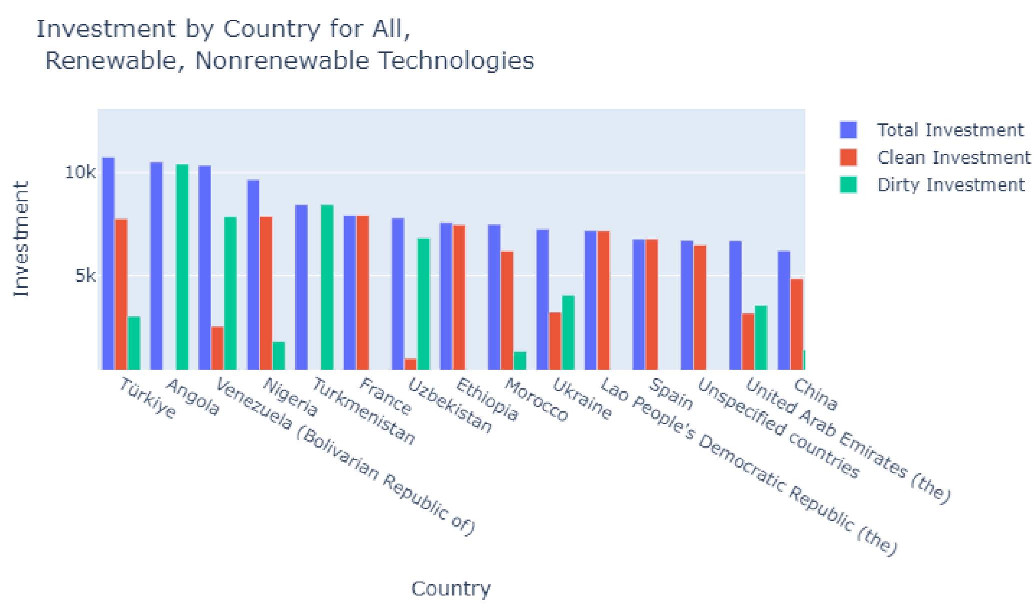
```
sql3 = """
        SELECT Countries, SUM(Investment) as Total_Investment_Nonrenewable
        FROM yucky_tech_table
        GROUP BY Countries
        ORDER BY Total_Investment_Nonrenewable DESC
        ;

    """
```

```
joined_tech =
pd.concat([total_investment,clean_investment['Total_Investment_Renewable'],
dirty_investment['Total_Investment_Nonrenewable']], axis=1)
```

To examine if this trend of funding continued through to a further subset of countries, the query was extended to the next ten countries. These are countries that rank 10-20 in total investment over the 2000-2022 period.

	Countries	Total_Investment	Total_Investment_Renewable	Total_Investment_Nonrenewable
10	Türkiye	10741.54	6721.26	7849.59
11	Angola	10522.91	6473.78	7161.68
12	Venezuela (Bolivarian Republic of)	10355.20	6194.38	6825.73
13	Nigeria	9646.33	5545.19	4030.76
14	Turkmenistan	8434.03	5281.17	3536.97
15	France	7933.84	4839.91	3130.47
16	Uzbekistan	7784.96	4827.48	2997.94
17	Ethiopia	7570.52	4806.47	2253.48
18	Morocco	7481.87	4548.60	2059.75
19	Ukraine	7236.81	4392.99	1990.41
20	Lao People's Democratic Republic (the)	7166.01	4220.45	1979.88



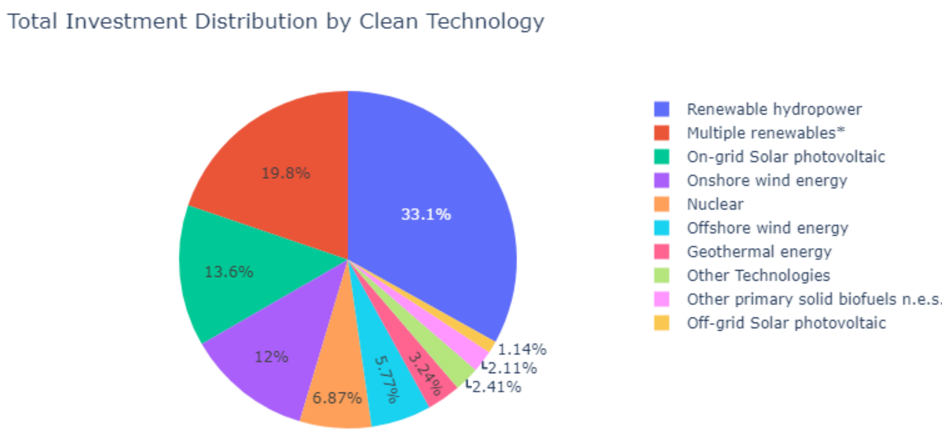
Looking at the next ten countries that invested into technology, we see more diversity in the technology invested in. There are more countries investing more into renewable technology than nonrenewable, which is in contrast to that of the aforementioned top ten countries.

Which renewable technology receives the most funding? How has funding for renewable technologies changed over time?

```
sql4 = """
        SELECT Clean_Technology, SUM(Investment) as Total_Investment
        FROM clean_tech_table
        GROUP BY Clean_Technology
        ORDER BY Total_Investment DESC;
    """
```

	Clean_Technology	Total_Investment
0	Renewable hydropower	109728.73
1	Multiple renewables*	65615.98
2	On-grid Solar photovoltaic	45009.21
3	Onshore wind energy	39912.84
4	Nuclear	22779.36
5	Offshore wind energy	19141.30
6	Geothermal energy	10744.91
7	Other primary solid biofuels n.e.s.	7013.64
8	Off-grid Solar photovoltaic	3778.41
9	Liquid biofuels	2686.90
10	Concentrated solar power	2684.49
11	Renewable municipal waste	1756.89
12	Biogas	520.68
13	Pumped storage	187.78
14	Solar thermal energy	133.32
15	Marine energy	16.00

To help visualize this, a pie chart was made:

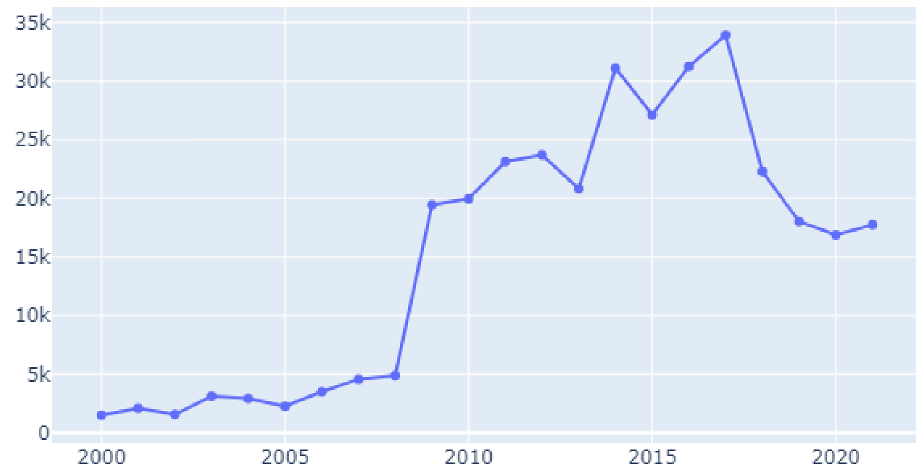


We can see that by a large margin the most invested in technology was renewable hydropower. This is most probably because hydropower offers low cost, flexible and secondary positive effects such as flood control[energy.gov, 2023]. We can also see that countries tend to invest in a diverse set of technologies, “Multiple renewables”, as this would most likely be the safest investment, as is any diverse portfolio.

```
sql5 = """
        SELECT Year, SUM(Investment) as Total_Investment
        FROM clean_tech_table
        GROUP BY Year
        ORDER BY Year ASC;
    """
```

	Year	Total_Investment
0	2000-01-01	1500.24
1	2001-01-01	2075.83
2	2002-01-01	1558.31
3	2003-01-01	3114.01
4	2004-01-01	2915.40
5	2005-01-01	2252.08
6	2006-01-01	3499.79
7	2007-01-01	4579.61
8	2008-01-01	4863.41
9	2009-01-01	19443.45
10	2010-01-01	19962.44
11	2011-01-01	23121.32
12	2012-01-01	23697.22
13	2013-01-01	20810.84
14	2014-01-01	31116.55
15	2015-01-01	27121.38
16	2016-01-01	31238.69
17	2017-01-01	33925.58
18	2018-01-01	22284.24
19	2019-01-01	18018.06
20	2020-01-01	16880.37
21	2021-01-01	17731.62

Total Investment Distribution by Clean Technology Per Year



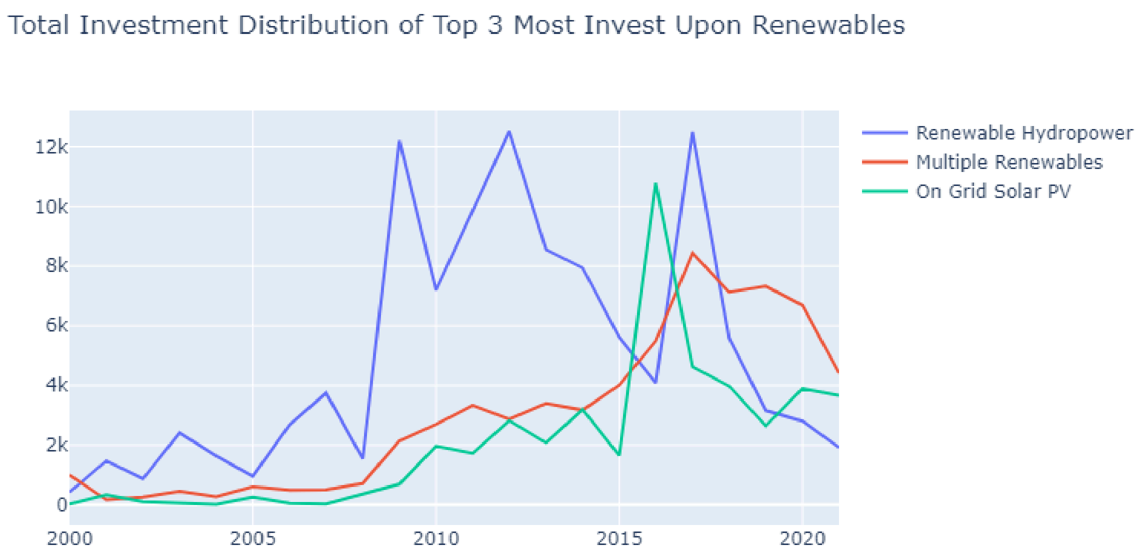
The steady trend upwards from 2008 to 2009 is most probably caused by large “Green Stimulus Programmes”, as discussed by Laszlo Varro [Laszlo Varro, 2023]. The decrease we see after 2017 could be partially from the USA pulling out of the Paris climate agreement. The multitude of factors contributing to the downward trend in investment are complex and extensive. Without a comprehensive model to analyze these influences, addressing them adequately is beyond the scope of this project.

2) **How has the funding for the top renewable technologies changed over time?**

```
sql1 = """
        SELECT Clean_technology
        FROM funding_change
        GROUP BY Clean_technology, Total_Investment
        ORDER BY Total_Investment DESC
        LIMIT 3
    """
```

	Clean_technology
0	Renewable hydropower
1	Multiple renewables*
2	On-grid Solar photovoltaic

```
sqlll = """
    SELECT Year,
    SUM(CASE WHEN Clean_Technology = 'Renewable hydropower' THEN Investment
ELSE 0 END) AS Renewable_Hydropower_Investment,
    SUM(CASE WHEN Clean_Technology = 'Multiple renewables*' THEN Investment
ELSE 0 END) AS Multiple_Renewables_Investment,
    SUM(CASE WHEN Clean_Technology = 'On-grid Solar photovoltaic' THEN
Investment ELSE 0 END) AS On_Grid_Solar_PV_Investment
    FROM clean_tech_table
    GROUP BY  Year
    ORDER BY Countries;
"""
```



From this data we see that Renewable hydropower has consistently garnered more funding compared to multiple renewables and on-grid solar PV, showing its established position as a primary source of renewable energy investment. This investment pattern may stem from hydropower's reliability, scalability, and long-term viability, which often make it a compelling choice for investors seeking sustainable energy solutions. However, in 2018 we start to see that the investment for both multiple renewables and on-grid solar PV overcomes the investment of renewable hydropower. This indicates that countries may have reached a potential limit on the hydro infrastructure that they can sustainably build.

3) **Which countries are investing the most and least into renewable and nonrenewable technology?**

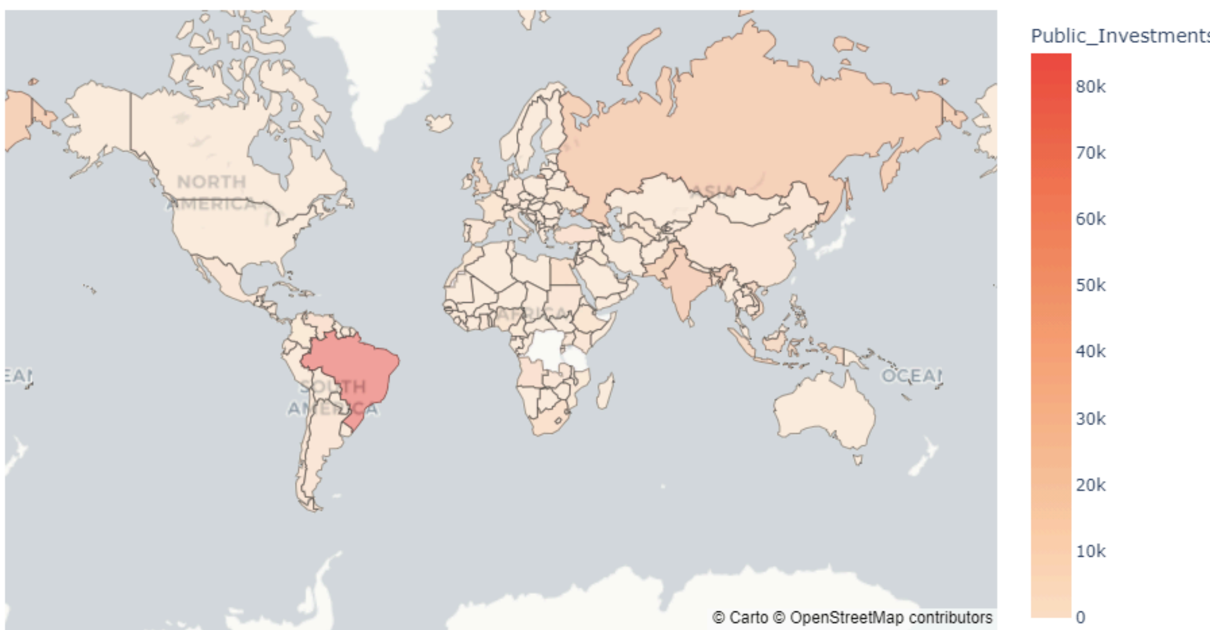
```
sqllll = """
```

```
        SELECT Countries, Total_Investment
    FROM (SELECT Countries, SUM(Investment) AS Total_Investment
        FROM clean_tech_table
        GROUP BY Countries) AS CountryInvestments
    ORDER BY Total_Investment ASC
    LIMIT 1;
"""
sqlllll = """
        SELECT Countries, Total_Investment
    FROM (SELECT Countries, SUM(Investment) AS Total_Investment
        FROM yucky_tech_table
        GROUP BY Countries) AS CountryInvestments
    ORDER BY Total_Investment ASC
    LIMIT 1;
"""
sqlllll = """
        SELECT Countries, Total_Investment
    FROM (SELECT Countries, SUM(Investment) AS Total_Investment
        FROM yucky_tech_table
        GROUP BY Countries) AS CountryInvestments
    ORDER BY Total_Investment DESC
    LIMIT 1;
"""
sqlllll = """
        SELECT Countries, Total_Investment
    FROM (SELECT Countries, SUM(Investment) AS Total_Investment
        FROM yucky_tech_table
        GROUP BY Countries) AS CountryInvestments
    ORDER BY Total_Investment DESC
    LIMIT 1;
"""
```

	Technology Type	Country	Total Investment(Millions USD)
MOST	RENEWABLE	Brazil	41913.13
MOST	NONRENEWABLE	Brazil	43050.44
LEAST	RENEWABLE	Anguilla	0.05
LEAST	NONRENEWABLE	Anguilla	0.0

Here we see that Brazil has the most investment in renewable and nonrenewable technologies, with Anguilla investing the least in both as well. It should be noted that there were multiple countries that had $0 investment into nonrenewable resources and Anguilla was set here as a name for the table. The other categories were unique.

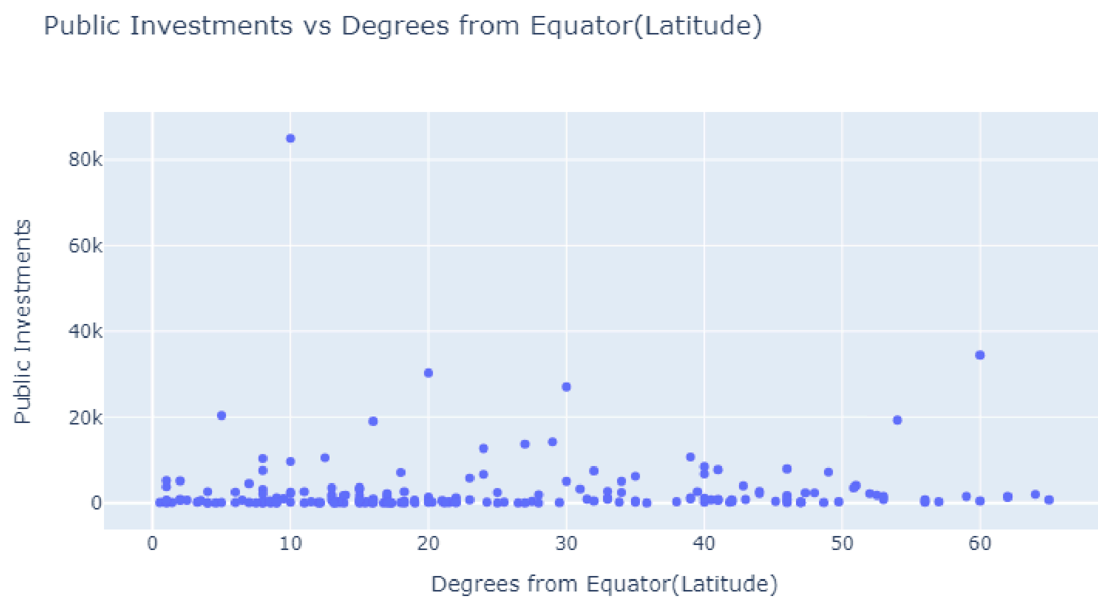
4) **How does technology funding depend on the location of the country?**



From the heat map we can see no discernible trend in the location of the country and the amount invested into energy production technologies. We observe a discernible trend between a country's population size and its investment level. Generally, countries with larger populations tend to exhibit darker shades on the map, indicating higher investment levels. While this correlation isn't flawless – as evidenced by the USA having a larger population than Brazil yet different investment levels – it's reasonable to expect that nations with a substantial share of the world's population would allocate more resources towards investment initiatives.

```
country_coords_sql['Absolute_Latitude'] = country_coords_sql['Latitude'].abs()

fig = go.Figure(data=go.Scatter(x=country_coords_sql['Absolute_Latitude'],
y=country_coords_sql['Public_Investments'],mode='markers'))
fig.update_layout(title='Public Investments vs Degrees from Equator(Latitude)',
                  xaxis_title='Degrees from Equator(Latitude)',
                  yaxis_title='Public Investments')
fig.show()
```



From this plot we can see no discernible trend, suggesting that the latitude of a country has no significant impact on its investment level. Regardless of a country's distance from the equator, its investment remains consistent compared to others.

Data Exploration

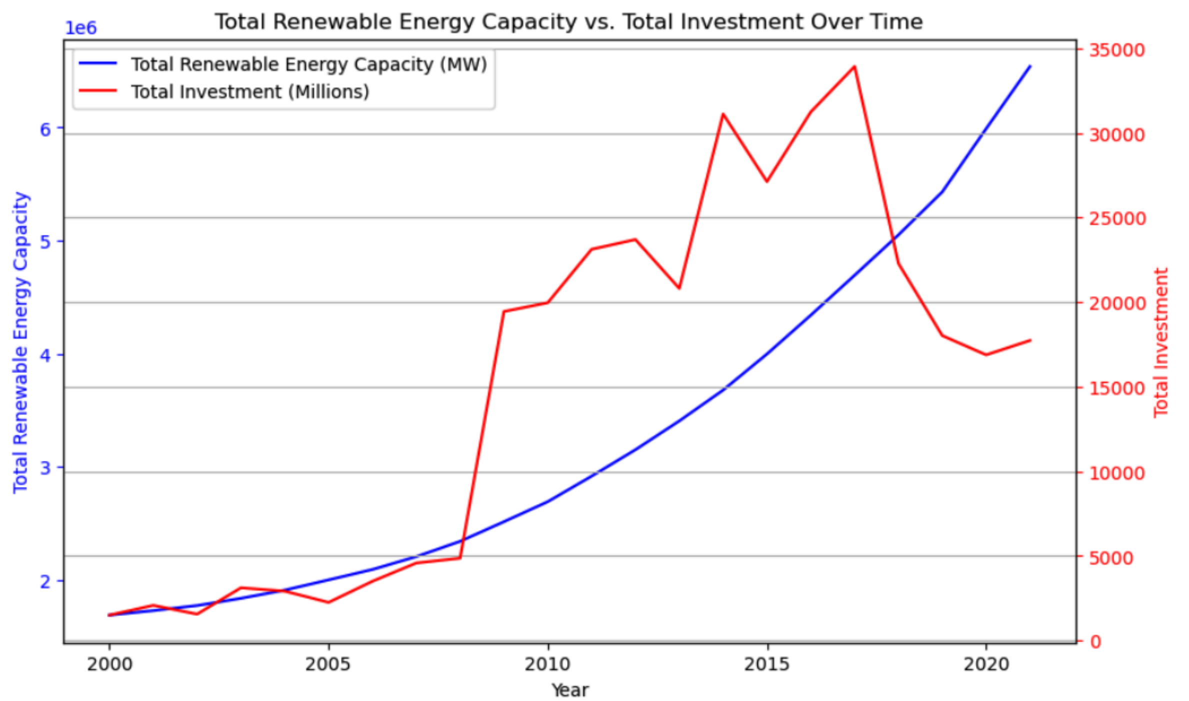
What is the correlation between total investment in renewable energy and changes in renewable energy capacity ?

To answer this question we joined the Renewable Capacity and Investments tables and see if there is a correlation between the two. We looked at world totals for both capacity and investment for years 2000-2021

```
join_cap_inv = """
SELECT capacity_world_totals.Year,
capacity_world_totals.Total_Renewable_Energy_Capacity,
investment_world_totals.Total_Investment
FROM capacity_world_totals
INNER JOIN investment_world_totals
ON capacity_world_totals.Year = investment_world_totals.Year;

"""
result = pd.read_sql_query(join_cap_inv, engine)
result
```

	Year	Total_Renewable_Energy_Capacity	Total_Investment
0	2000	1691560.917	1500.24
1	2001	1731499.503	2075.83
2	2002	1776030.105	1558.31
3	2003	1839862.090	3114.01
4	2004	1912507.760	2915.40
5	2005	2002814.819	2252.08
6	2006	2094673.677	3499.79
7	2007	2208346.638	4579.61
8	2008	2343337.860	4863.41
9	2009	2516985.683	19443.45
10	2010	2693789.075	19962.44
11	2011	2919240.911	23121.32
12	2012	3152876.231	23697.22
13	2013	3407062.588	20810.84
14	2014	3679517.566	31116.55
15	2015	3998424.851	27121.38
16	2016	4341422.973	31238.69
17	2017	4694315.512	33925.58
18	2018	5051219.452	22284.24
19	2019	5431035.426	18018.06
20	2020	5989128.519	16880.37
21	2021	6539127.084	17731.62



From the query result table and the “Total Renewable Energy Capacity vs. Total Investment Over Time” graph, we can observe the following insights:

1. **Trend of Renewable Energy Capacity**: Over the years, there is a clear trend of increasing renewable energy capacity. The values in the "Total_Renewable_Energy_Capacity" column show a consistent upward trajectory from 2000 to 2021, indicating a continuous growth in renewable energy infrastructure.
2. **Variation in Total Investment**: The "Total_Investment" column displays varying levels of investment in renewable energy across different years. While some years witness relatively low investments, others show higher investment amounts. This variation suggests fluctuations in funding and financial commitment to renewable energy projects over time. There has been a significant increase in the total investment in 2009, however it did not reflect on the Total Capacity.
3. **Potential Correlation**: We have calculated the correlation coefficient between total investment and renewable energy capacity, and it is 0.650448. A correlation coefficient of 0.6504 indicates a moderate positive linear relationship between total investment in renewable energy and changes in renewable energy capacity. Since the correlation coefficient is positive, it indicates that as total investment in renewable energy increases, there tends to be an increase in renewable energy capacity. The magnitude of 0.6504 suggests that the relationship is moderately strong. In other words, there is a noticeable tendency for total investment and changes in capacity to move together, but it's not a perfect relationship. However, it's important to note that correlation does not imply causation. While there's a correlation between these two variables, it doesn't necessarily mean that changes in investment directly cause changes in capacity or vice versa. Other factors could also influence both variables.

4. **Impact of External Factors**: External factors such as policy changes, technological advancements, and market dynamics likely influence both investment decisions and capacity growth. Analyzing these factors alongside investment and capacity data could offer deeper insights into the drivers behind renewable energy development.

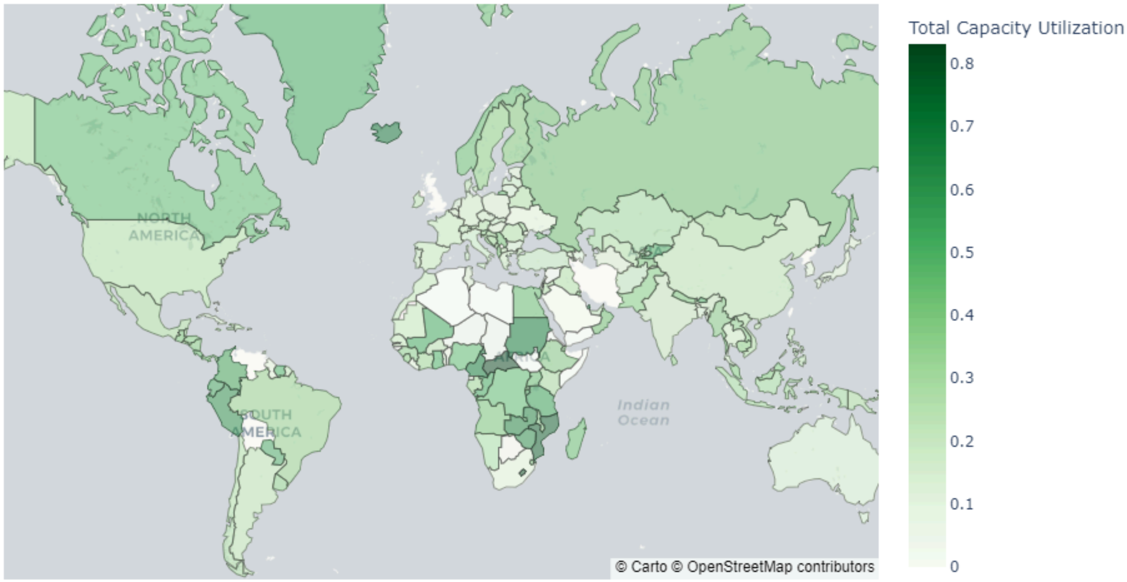
Overall, the data highlights the dynamic nature of renewable energy growth, influenced by various factors including investment trends, technological progress, and policy environments. Further analysis, including correlation calculations and consideration of external factors, would be valuable for a comprehensive understanding of the relationship between investment and capacity changes.

What proportion of the renewable energy capacity is utilized by countries in 2021?

We seek to answer this question by joining the renewable capacity table with the electricity production table. We can look at the amount of electricity utilized by taking the ratio between the total electricity produced and the total electricity capacity. To do this we need to ensure that the units make sense, since capacity is in units of MW it is a measure of power meaning that to relate this to GWh (which is a unit of energy) we need to consider the time period of interest. Since we are looking at the utilization over the span of a year we will divide the capacity by 8760 which is the number of hours in a year. We will also divide it by 1000 to convert MW to GW. One potential issue with this conversion is that this assumes that the electricity plants have 100% uptime, so in reality the statistic we are computing is an upper bound on the utilization, it measures the utilization if there was no downtime at production facilities.

```
SELECT pop.Country, pop.`Country Code`, (SUM(pop.`ELECTRICITY PRODUCTION
(GWh)`)/(SUM(cap.`Renewable Capacity (MW)`)/1000))/8760 AS `Total Capacity
Utilization`
                   FROM elec_and_pop_new AS pop LEFT JOIN elec_cap AS cap
                   ON pop.Country = cap.`Region/country/area` AND pop.Time =
cap.Year AND pop.Technology = cap.` Technology`
                   WHERE YEAR = '2021'
                   GROUP BY pop.Country
                   ORDER BY pop.Country DESC
```

The figure above shows the results of the query in the form of a Choropleth map with the greener the color the higher



the optimal utilization of electricity capacity. We can see that with 100% uptime Canada is only using approximately half of its renewable capacity for active generation. It is interesting to observe how large parts of Central Africa seem to have quite high optimal utilization. This is likely due to their relatively low total renewable capacity that we saw in an earlier section of this paper. If they have low renewable capacity to begin with then it makes sense that they would use as much of its capacity as possible.

We notice that the major populous countries such as China, India and the USA have fairly low optimal utilization at around 20%, this is likely due to their already existing reliance on fossil fuels as evidenced from the production section above. It can be quite costly to transition existing infrastructure to renewable energy, costs to retrofit buildings, and deactivation of existing plants are all significant considerations [UNCTAD, 2023].

Which countries have the most efficient use of funding via production?

```
sqljoin1 = """
SELECT
    cc.Countries,
    cc.Public_Investments,
    cc.Alpha_3_Code,
    poptable1.`Per Capita Total`,
    poptable1.`GWh_Total`,
    poptable1.`Population`,
    (cc.Public_Investments / poptable1.`Per Capita Total`) AS Efficiency
```

```sql
FROM
    cc
JOIN
    poptable1 ON cc.Alpha_3_Code = poptable1.`Country Code`
GROUP BY
    cc.Countries
ORDER BY
    Efficiency DESC
LIMIT 10
```

	Countries	Public_Investments	Alpha_3_Code	Per Capita Total	GWh_Total	Population	GWh_Dollar
0	Trinidad and Tobago	0.43	TTO	0.006075	9267.96	1.525663e+06	21553.395349
1	United States of America	238.68	USA	0.013175	4374676.40	3.320316e+08	18328.625775
2	Algeria	11.16	DZA	0.001921	84856.11	4.417797e+07	7603.594086
3	Australia	37.38	AUS	0.010339	265555.00	2.568541e+07	7104.200107
4	French Polynesia	0.25	PYF	0.001589	483.12	3.040320e+05	1932.480000
5	Canada	440.57	CAN	0.016820	642974.00	3.822650e+07	1459.413941
6	China	6203.85	CHN	0.006043	8534250.00	1.412360e+09	1375.637709
7	Libya	42.58	LBY	0.005140	34621.00	6.735277e+06	813.081259
8	Iran	483.41	IRN	0.004090	359622.34	8.792343e+07	743.928218
9	Malta	4.57	MLT	0.004272	2215.06	5.185360e+05	484.695842

The joined tables allow us to make a connection between the population and energy capacity with the public investment by each country. Without the join, this would be impossible to compare.

The purpose of the join was to study the Investment Utility (IU) in units of GWh/million USD, of a country's spending on energy technologies. The IU represents the efficiency of public investments in terms of electricity production per capita. Higher values indicate cheaper purchasing price of energy by country. Countries like Trinidad and Tobago and the USA have relatively IU values, suggesting effective allocation of resources in their energy sectors. Even in the top 10 most efficient countries, we see upwards of 44x the IU values between the #10 and #1 spot.

	Countries	Public_Investments	Alpha_3_Code	Per Capita Total	GWh_Total	Population	GWh_Dollar
0	China	6203.85	CHN	0.006043	8534250.00	1.412360e+09	1375.637709
1	India	30332.02	IND	0.001153	1622611.33	1.407564e+09	53.494997
2	United States of America	238.68	USA	0.013175	4374676.40	3.320316e+08	18328.625775
3	Indonesia	20343.43	IDN	0.001054	288466.32	2.737532e+08	14.179827
4	Pakistan	27076.29	PAK	0.000655	151526.59	2.314021e+08	5.596283
5	Brazil	84963.57	BRA	0.003062	656187.08	2.143262e+08	7.723158
6	Nigeria	9646.33	NGA	0.000172	36633.79	2.134013e+08	3.797692
7	Bangladesh	12701.45	BGD	0.000467	79019.90	1.693563e+08	6.221329
8	Russian Federation	34478.19	RUS	0.007077	1019983.92	1.441305e+08	29.583453
9	Mexico	5730.69	MEX	0.003068	388767.04	1.267051e+08	67.839482

When arranging the list by population, it's evident that China and particularly the USA stand out as significant outliers, providing energy at a considerably lower cost relative to their larger populations. This disparity is even more pronounced in the case of the USA, where the substantial investment further accentuates its outlier status.

```
sqljoin2 = """
SELECT
    cc.Countries,
    cc.Public_Investments,
    cc.Alpha_3_Code,
    poptable1.`Per Capita Total`,
    poptable1.`GWh_Total`,
    poptable1.`Population`,
    (cc.Public_Investments / poptable1.`Per Capita Total`) AS Efficiency
FROM
    cc
JOIN
    poptable1 ON cc.Alpha_3_Code = poptable1.`Country Code`
GROUP BY
    cc.Countries
ORDER BY
    Efficiency ASC
LIMIT 10
"""
```

	Countries	Public_Investments	Alpha_3_Code	Per Capita Total	GWh_Total	Population	GWh_Dollar
0	South Sudan	31.41	SSD	0.000000	0.00	10748272.0	0.000000
1	Somalia	171.69	SOM	0.000000	0.00	17065581.0	0.000000
2	Tuvalu	56.34	TUV	0.000696	7.80	11204.0	0.138445
3	Liberia	675.07	LBR	0.000039	201.00	5193416.0	0.297747
4	Solomon Islands	281.80	SLB	0.000122	86.28	707851.0	0.306175
5	Chad	705.52	TCD	0.000020	335.54	17179740.0	0.475592
6	Guinea-Bissau	49.25	GNB	0.000012	24.78	2060721.0	0.503147
7	Tonga	139.07	TON	0.000691	73.28	106017.0	0.526929
8	Kiribati	54.50	KIR	0.000239	30.82	128874.0	0.565505
9	Niger	933.31	NER	0.000021	536.14	25252722.0	0.574450

It's noteworthy that both South Sudan and Somalia show investments in energy but lack GWh_Total output. Several factors could explain this phenomenon: either the data for these countries is too small to accurately reflect their energy production, there may have been floating-point errors during the table join process, or perchance, these nations have encountered challenges in translating public investment into tangible energy infrastructure.

```
sqljoin3 = """
SELECT
    Countries, row_number AS World_Rank
FROM
    (
        SELECT
            ROW_NUMBER() OVER (ORDER BY cc.Countries) AS row_number,
            cc.Countries,
            cc.Public_Investments,
            cc.Alpha_3_Code,
            poptable1.`Per Capita Total`,
            poptable1.`GWh_Total`,
            poptable1.`Population`,
            (cc.Public_Investments / poptable1.`Per Capita Total`) AS
Efficiency
        FROM
            cc
        JOIN
            poptable1 ON cc.Alpha_3_Code = poptable1.`Country Code`
        GROUP BY
            cc.Countries
        ORDER BY
            Efficiency DESC
```

```
    ) AS numbered_rows
WHERE
    Alpha_3_Code IN ('USA','GBR','CHN','BRA');
“”””
```

	Countries	World_Rank	GWh_Dollar
0	United States of America	2	18328.625775
1	China	7	1375.637709
2	United Kingdom of Great Britain and Northern I...	82	15.949628
3	Brazil	105	7.723158

For further investigation, we see that the United Kingdom, renowned as a significant global player, ranks relatively low in the world ranking, charging its citizens a comparatively higher amount per GWh per dollar. Similarly, Brazil, previously identified as having substantial funding in both renewable and nonrenewable resources, still imposes a significant cost burden on each citizen. It is possible that the large amount of public investment is still being paid for by the citizens and in the future we will see this investment pay off.

Discussion

Britain: Throughout the course of this project I have gained a significant amount of experience and knowledge in the analysis of large datasets. First and foremost, I went from little to no experience working with SQL to being able to comfortably write queries using aggregates, joins and boolean expressions. This experience was mostly gained during the individual component of the project where I had to first join two tables together, then working with the joined table I wrote 6 queries that sought out a variety of insights from the single table. Another major milestone in my learning throughout this project was how to more effectively clean the data using pandas and python. The biggest wrangling process for me was cleaning the names of Countries for the purpose of joining the world bank data with the IRENA data.

Matthew: The biggest lesson I've learned through doing this project was navigating the complexities of joins between tables. Creating a primary key for my dataset to match with other data sets, particularly when integrating longitude and latitude data, and population and energy data proved to be a challenge as fields from different tables needed cleaning or altering. From this experience, opting for a manual approach to changing country names might have proven more straightforward than scripting country code matching.
Exploring alternative methods for joining tables through country names instead of relying solely on Alpha 3 codes could have potentially streamlined the workflow and expedited the process. Furthermore, normalizing all technologies across datasets to serve as a primary key for joins would have allowed for deeper analysis and provided a more comprehensive understanding of the role of each technology over the past two decades. In the

future, adding data comparing GDP to public investment and comparing to precious guiding questions, would add significance to the rise and fall of investment in energy technologies over the studied 20 year period.

Overall, the data provided by the IRENA database presented a significant challenge due to its varying formats across different technology and country tables. Inconsistencies, such as the inclusion of continent data in some tables but not others, and adding different technologies/removing others between datasets, added complexity to the process of comparison and data integration. In retrospect choosing different, more raw datasets to analyze, may have provided a smoother experience.

Misha: Through this project, I've gained a deeper understanding of the critical role data cleaning and wrangling play in preparing datasets for analysis, particularly before uploading them into an SQL database. Recognizing the significance of understanding data types within tables has been pivotal for effectively joining datasets, ensuring compatibility and accuracy in analyses. I've come to appreciate the importance of expanding my proficiency in SQL, enabling me to rely less on pandas and leverage the full capabilities of SQL for data manipulation and querying. In the project, the understanding of join keys was crucial for deriving insights that were only possible by combining data from multiple tables. By identifying and utilizing the appropriate join keys, we were able to establish relationships between different datasets and merge them effectively. This allowed us to perform comprehensive analyses that spanned multiple dimensions, such as correlating renewable energy capacity with energy investment.

Something to improve in future projects is implementing a standard data workflow. Our data was not transformed in a consistent way and the group members exported data from the database in different shapes. This led to a number of challenges when joining such as requiring reshaping, data type verification, and fixing join columns to match correctly. If we set out a proper data workflow and performed the same wrangling tasks throughout the entire process it would have resulted in a much smoother analysis.

One way to extend this project would be to include climate data in our analysis as a large part of the discourse on renewable energy products is related to the corresponding CO2 emissions that are produced by fossil fuels and the like. In the future a goal would be to keep parts of our analysis here but to add further questions that could be answered using climate data, for example we could ask “What are the carbon emissions like in countries that invest in renewable energy?”, or “Is there a negative correlation between renewable electricity production/investment and GHG emissions?”

Conclusion

The goal of our project was to, using data on renewable energy, explore the state of the renewable energy landscape as well as observe the evolution over the past 20 years of this field. We did this by looking at electricity production, renewable electricity capacity, as well as renewable electricity investment around the world from 2000 to 2021. We gained many insights throughout this report, some of the major ones are:

1. There has been consistent growth and substantial increase in total renewable energy capacity worldwide over the past two decades.
2. We've also identified dominant technologies, observed shifts in distribution, and examined differences in renewable energy capacity across various regions.

3. Renewable energy production is on the rise across the globe, and Coal and Peat usage, while still increasing, is increasing at a much slower rate in recent years.
4. In the USA, electricity produced by burning Coal and Peat has been on the decline, and has seen natural gas overtake it as the largest producer of electricity in the country.
5. Renewable hydropower has been the most publicly invested renewable technology across the globe over the past 20 years.
6. The latitudinal location of a country has no impact on the amount said country will invest into its energy sector.

Upon joining the tables we were able to answer three more questions that could not be clarified by the tables on their own:
1. What is the correlation between renewable electricity capacity and renewable electricity investment
2. What proportion of the renewable energy capacity is utilized by countries in 2021?
3. Which countries have the most efficient use of funding via production?

Overall, our project contributes to a deeper understanding of the renewable energy landscape and provides valuable insights into its trajectory, highlighting opportunities for further research to accelerate the transition towards a sustainable energy future.

References

1. United Nations. (n.d.). *Renewable Energy.* Retrieved from https://www.un.org/en/climatechange/raising-ambition/renewable-energy#:~:text=Cheap%20electricity%20from%20renewable%20sources,helping%20to%20mitigate%20climate%20change
2. IRENA (2023, February). *The changing role of hydropower: Challenges and opportunities* [PDF]. International Renewable Energy Agency. https://www.irena.org/Publications/2023/Feb/The-changing-role-of-hydropower-Challenges-and-opportunities
3. Government of Iceland. (n.d.). *Energy*. Retrieved from https://www.government.is/topics/business-and-industry/energy/#:~:text=Renewable%20energy%20provided%20almost%20100,supplier%20of%20electricity%20in%20Iceland.
4. Business Norway (2023, March 15). *How Norway produces hydropower with a minimal carbon footprint*. Retrieved from https://businessnorway.com/articles/how-norway-produces-hydropower-with-a-minimal-carbon-footprint
5. IEA Bioenergy. (2021, October). *Implementation of bioenergy in Sweden – 2021 update, Country Reports.* Retrieved from https://www.ieabioenergy.com/wp-content/uploads/2021/11/CountryReport2021_Sweden_final.pdf

6. Maguire, G. (2023, April 6). *Middle East starts cleaning up its power act with renewable push.* Reuters. https://www.reuters.com/business/energy/middle-east-starts-cleaning-up-its-power-act-with-renewable-push-maguire-2023-04-05/
7. Quaschning, V. (2022, November) *Specific Carbon Dioxide Emissions of Various Fuels.* Retrieved From https://www.volker-quaschning.de/datserv/CO2-spez/index_e.php
8. U.S. Energy Information Administration (Sept 2023) *Carbon Dioxide Emission Coefficients*; Retrieved from: [https://www.eia.gov/environment/emissions/co2_vol_mass.php]
9. UN Conference on Trade and Development (2023) *The costs of achieving the SDGs: Energy Transition;* Retrieved From: [https://unctad.org/sdg-costing/energy-transition]
10. Office of ENERGY EFFICIENCY & RENEWABLE ENERGY(2023), *Benefits of Hydropower,* Retrieved From:[https://www.energy.gov/eere/water/benefits-hydropower#:~:text=Hydropower%20provides%20low%2Dcost%20electricity,complements%20other%20renewable%20energy%20sources]
11. IEA, *Green stimulus after the 2007 crisis,* Retrieved From: [https://www.iea.org/articles/green-stimulus-after-the-2008-crisis]