Wind & Solar Energy Production, Growth, and CO2 Emissions by Country

Jennifer Smith

March 6, 2022

Project Description

This project investigates the wind and solar energy production as well as CO2 emissions of the world's countries over time. It seeks to answer questions such as:

- Which countries produce the most wind/solar energy?
- Which countries have increased their capacity for wind/solar energy the most?
- Which countries emit the most CO2?
- Which countries are producing disproportionate amounts of wind/solar energy compared to their share of the world's CO2 emissions?

Data Sources & Prep

I sourced datasets from Kaggle.com for solar and wind power generation by country. Both were attributed to Wikipedia. Some data was missing from the solar dataset, so I copy/pasted off Wikipedia into a .csv file rather than using the one on Kaggle. I then sourced a table on worldometers.info that contains CO2 emissions by country. Last, I pulled a table containing the populations by country from the same Wikipedia page as the wind dataset.

I liked how the solar data contained both new and total capacity for each year, so before loading into MySQL, I added the "new installed" columns for the wind data and computed those numbers as well. Then I loaded all my datasets but found that the spaces in the column names were problematic, so I edited the column names in excel to eliminate the spaces and loaded them again. I ran several queries, iterating on them as I thought of questions to ask about the data. Once I thought I had all the data that would answer my questions, I exported 6 result tables as .csv files and then uploaded them into Power BI.

Visualizations & Insights

The first visualization (Figure 1) shows the total wind and solar energy production capacity by country as of 2020. I chose a stacked bar because I wanted to show the total capacity but also the mix of solar and wind within the total. I like that it shows the massive difference in the production of China next to even the US, and I also like how you can see which countries are far more invested in one technology than the other (solar dominates in Japan and Vietnam, wind in Canada and Sweden).

I chose the colors because they are soft and muted, and clearly different but harmonious. I also thought the violet felt airy to evoke wind and the yellow is an obvious nod to solar. The SQL query this visualization depicts is Figure 4.

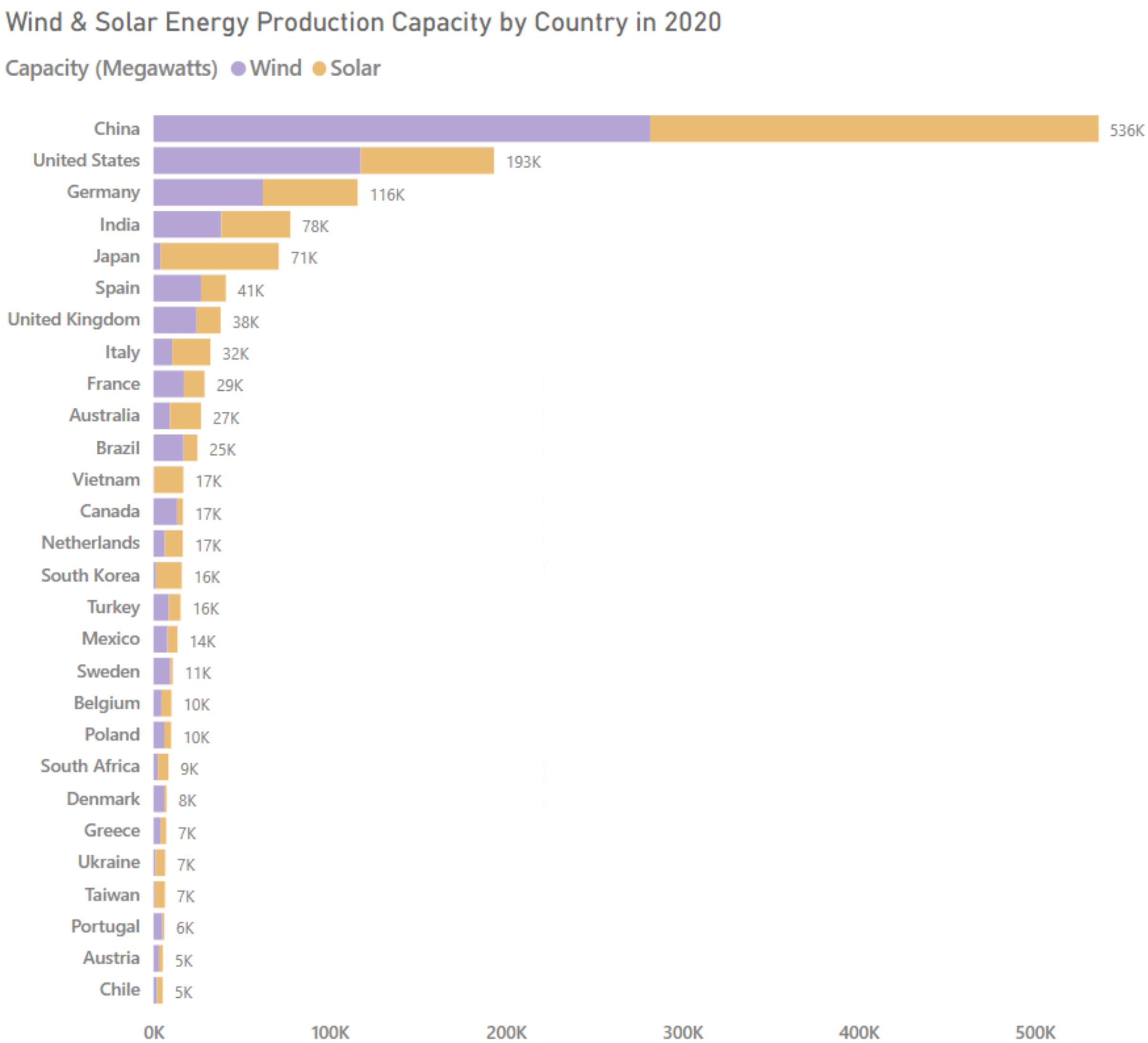

Figure 1

The second visualization (Figure 2) shows the combined growth of solar/wind power by country from 2016 to 2020 and displays the changes both in absolute terms (as Megawatts added) and in percentage growth. When I was exploring the dataset, I noticed an enormous difference between when I displayed the total MW added or the percentage added. I thought this was an interesting story, but it was challenging to depict in one visual because of the drastically different scales. I started with a bar and line graph but felt the jagged lines might be misleading because lines usually imply passage through time, but this data is categorical. I finally found that I could keep the bar/line graph but could format the lines to make them appear as bars. The effect is an overlapping bar chart that shows both the absolute and percentage changes.

I wanted to keep the colors in the same family because they're simply different measures of the same data. I landed on an opaque yellow-leaning blue for the absolute Megawatts because I felt an opaque color was appropriate for the more solid absolute measure, and a transparent, red-leaning blue for the percentage because I thought the transparency reflects the more abstract nature of percentages. I included a line marker on top of the percentage bars to make the tops easier to see and compare, especially with those that don't rise above the opaque bars.

This chart shows not only the enormous addition of Megawatts made by China, but also how small that growth is relative to their existing capacity. It also highlights the investments made by smaller countries like Vietnam, whose addition looks tiny compared to China's in absolute terms but is in reality astronomical relative to their size. The SQL query this visualization depicts is Figure 5.

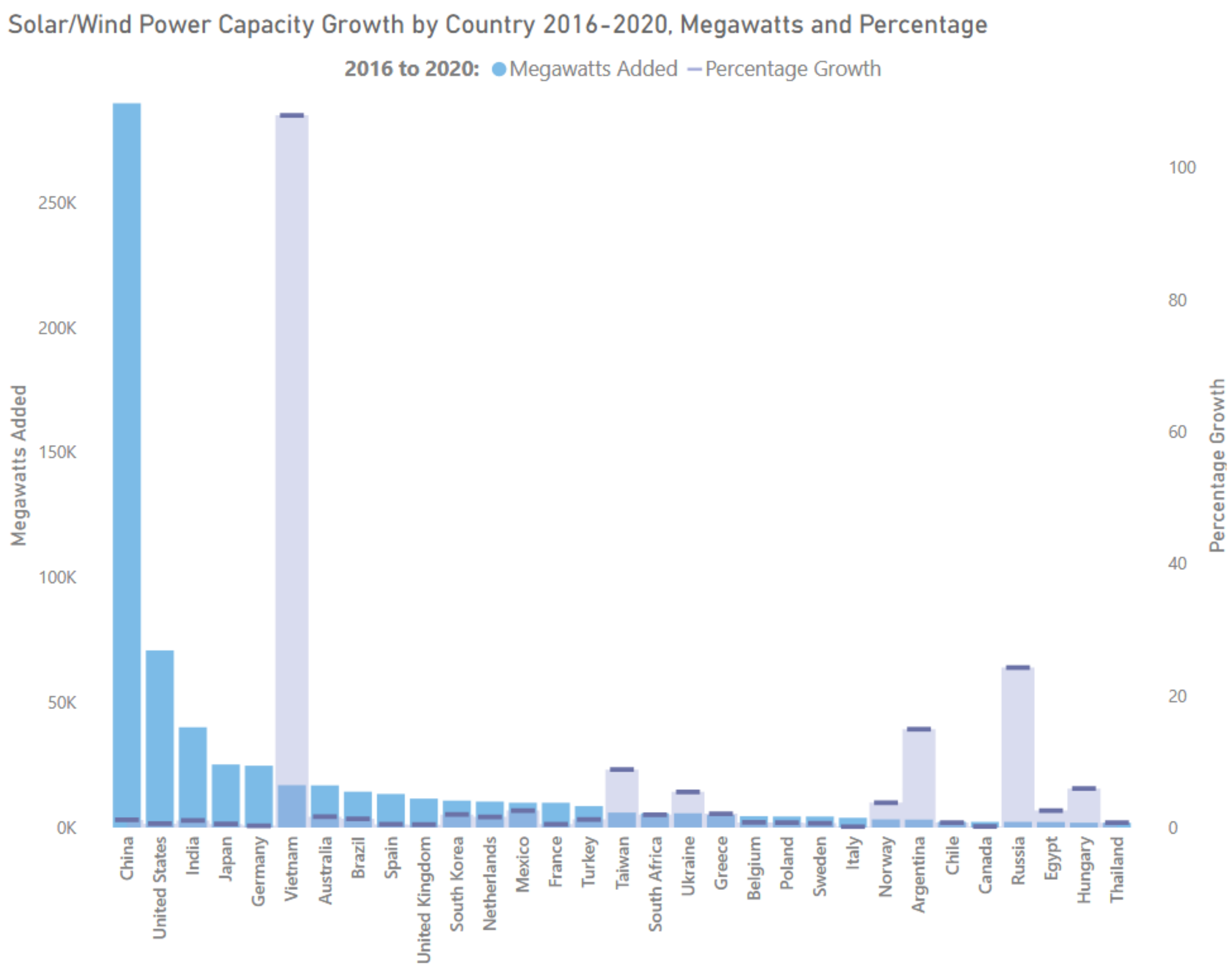

Figure 2

The last chart (Figure 3) shows the share (% of world total) of emissions side by side with the share (% of world total) of solar/wind power produced by each country. This visualization highlights several interesting insights. One is that the top two countries in share of CO2 emissions are also the top two in solar/wind capacity, almost to the exact percentage share. Another is that there are a handful of countries that emit a much higher share of CO2 than their share of solar/wind production (India, Russia, Iran), and several countries that produce a far greater share of solar/wind power than their share of CO2 emissions (Germany, Japan, UK, Italy, Spain).

I kept the same blue for the solar/wind capacity that I had on the previous visual for continuity, then added a grayish brown for the CO2 emissions. I think it works nicely because both colors are still muted and harmonious, but the brown evokes the pollution of CO2, and the blue implies the virtue of clean energy. This visualization pulls from two separate SQL queries (Figures 6-7).

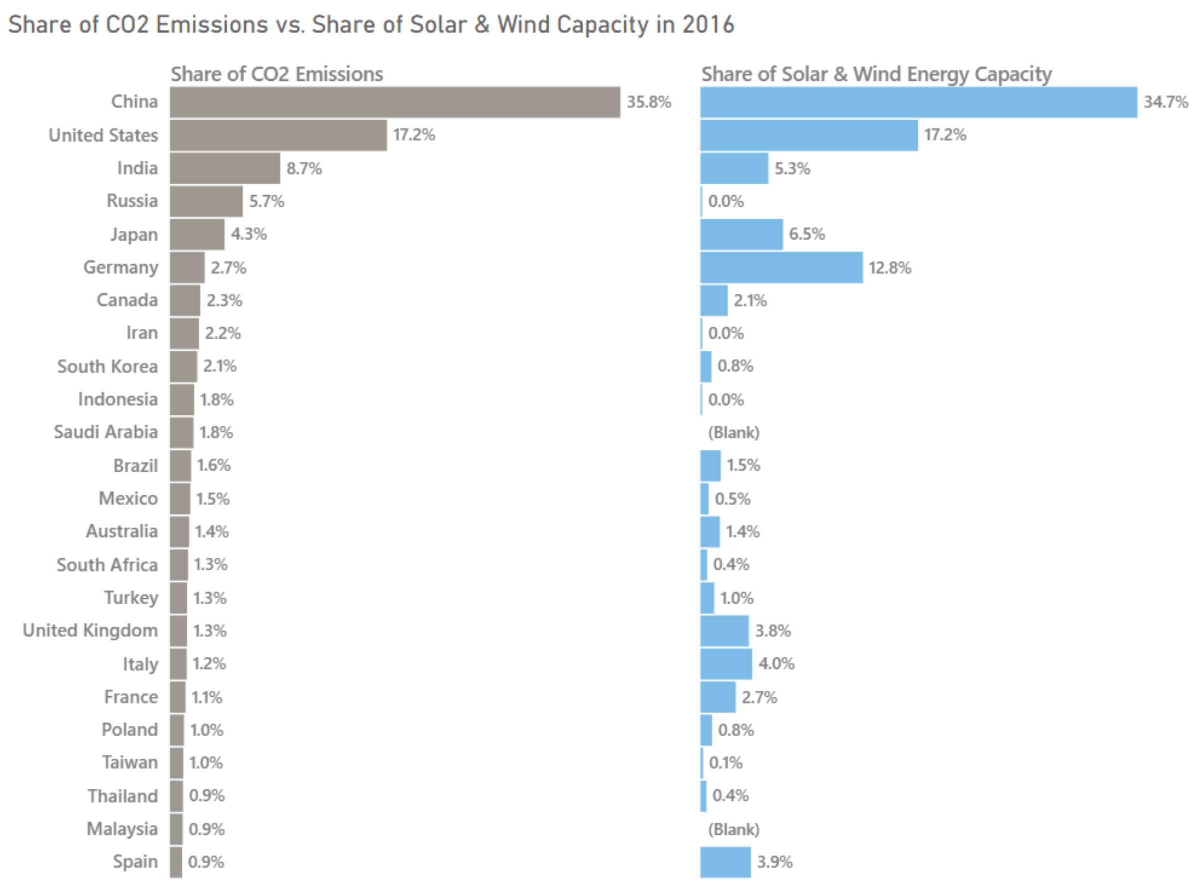

Figure 3

Assignment Questions:

a. How is visualizing data and querying a database similar?

Visualizing data and querying a database are similar in that they both use analytical techniques to make sense of a dataset. In both cases we begin with a pile of data that is essentially meaningless, and we use various techniques to find answers or information that contribute to a story or stories about and within the data.

b. How are they distinct?

They are distinct in that when querying a database, we must know exactly what we want the query to deliver because the very nature of a query is that it asks a specific question. Visualization, on the other hand, *can* sometimes be used to answer a specific question, but it can also be used in a more exploratory capacity. Often trends or relationships in data are difficult if not impossible to spot in the raw data, but when we use visualization to display the shapes of distribution or the slopes of change over time, patterns can become instantly clear.

c. Under what circumstances should you do each?

Querying a database can be used in a couple circumstances. First, it should be used to answer a question that is very narrow and specific, such as "Which customer placed the largest order last month?". Alternatively, it can also be used to gather, organize, and clean a subset of data for further analysis or visualization. For example, if we want to look for trends within the lighting category over the last 18 months, we could query the broader sales database to return all transactions that fit those parameters. The results could then be exported into Excel, Power BI, or another analysis tool.

Visualization has two parts. First, it can be used to explore a large dataset, looking for meaningful patterns or relationships by looking at the distribution, rank/proportion, change over time, relationships between variables, and geographical makeup. Then, once a story about the data is clearly established, visualization can be used to communicate that story. This is when care must be taken to ensure that the visualization is designed to tell the story as clearly as possible.

Queries

```sql
#Total Solar & Wind By Country 2020
SELECT S.Country, S.2020_Total as '2020 Capacity Solar (MW)',
    W.2020_Total as '2020 Capacity Wind (MW)',
    SUM(S.2020_Total + W.2020_Total) as '2020 Capacity Solar & Wind (MW)'
FROM Solar S JOIN Wind W
ON S.Country = W.Country
GROUP BY S.Country, S.2020_Total, W.2020_Total
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	2020 Capacity Solar (MW)	2020 Capacity Wind (MW)	2020 Capacity Solar & Wind (MW)
China	254355	281993	536348
United States	75572	117744	193316
Germany	53783	62184	115967
India	39211	38559	77770
Japan	67000	4206	71206
Spain	14089	27089	41178
United Kingdom	13563	24665	38228
Italy	21600	10839	32439
France	11733	17382	29115
Australia	17627	9457	27084
Brazil	7881	17198	25079
Vietnam	16504	600	17104
Canada	3325	13577	16902
Netherlands	10213	6600	16813
South Korea	14575	1636	16211
Turkey	6668	8832	15500
Mexico	5644	8128	13772
Sweden	1417	9688	11105
Belgium	5646	4692	10338
Poland	3936	6267	10203
South Africa	5990	2636	8626
Denmark	1300	6235	7535
Greece	3247	4113	7360
Ukraine	5360	1402	6762
Taiwan	5817	854	6671

Figure 4

```sql
#5 Year Change in Total Solar & Wind by Country 2016 to 2020
SELECT S.Country, S.2016_Total as '2016 Capacity Solar (MW)',
    W.2016_Total as '2016 Capacity Wind (MW)',
    SUM(S.2016_Total + W.2016_Total) as '2016 Capacity Solar & Wind (MW)',
    S.2020_Total as '2020 Capacity Solar (MW)',
    W.2020_Total as '2020 Capacity Wind (MW)',
    SUM(S.2020_Total + W.2020_Total) as '2020 Capacity Solar & Wind (MW)',
    SUM(S.2020_Total - S.2016_Total + W.2020_Total - W.2016_Total) as '5 Year Change (MW)',
    SUM((S.2020_Total + W.2020_Total)/(S.2016_Total + W.2016_Total)-1) as '5 Year Change (%)'
FROM Solar S JOIN Wind W
ON S.Country = W.Country
GROUP BY S.Country, S.2016_Total, W.2016_Total, S.2017_Total, W.2017_Total, S.2018_Total,
    W.2018_Total, S.2019_Total, W.2019_Total, S.2020_Total, W.2020_Total
ORDER BY 4 desc
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	2016 Capacity Solar (MW)	2016 Capacity Wind (MW)	2016 Capacity Solar & Wind (MW)	2020 Capacity Solar (MW)	2020 Capacity Wind (MW)	2020 Capacity Solar & Wind (MW)	5 Year Change (MW)	5 Year Change (%)
China	78070	168690	246760	254355	281993	536348	289588	1.1736
United States	40300	82183	122483	75572	117744	193316	70833	0.5783
Germany	41220	50019	91239	53783	62184	115967	24728	0.2710
Japan	42750	3234	45984	67000	4206	71206	25222	0.5485
India	9010	28665	37675	39211	38559	77770	40095	1.0642
Italy	19279	9257	28536	21600	10839	32439	3903	0.1368
Spain	4669	23075	27744	14089	27089	41178	13434	0.4842
United Kingdom	11630	15030	26660	13563	24665	38228	11568	0.4339
France	7130	12065	19195	11733	17382	29115	9920	0.5168
Canada	2715	11898	14613	3325	13577	16902	2289	0.1566
Brazil	0	10740	10740	7881	17198	25079	14339	1.3351
Australia	5900	4327	10227	17627	9457	27084	16857	1.6483
Turkey	832	6101	6933	6668	8832	15500	8567	1.2357
Sweden	175	6519	6694	1417	9688	11105	4411	0.6589
Netherlands	2100	4328	6428	10213	6600	16813	10385	1.6156
Denmark	900	5227	6127	1300	6235	7535	1408	0.2298

Figure 5

```sql
#C02 Emissions by Country and population
SELECT P.Country, `CO2`.`CO2_Emissions(tons,2016)`,
    SUM(`CO2`.`CO2_Emissions(tons,2016)`/P.Population_2018)
    as 'CO2 Emissions per Capita', CO2.Share_of_World
FROM CO2 JOIN Population P
ON CO2.Country = P.Country
GROUP BY P.Country, `CO2`.`CO2_Emissions(tons,2016)`, CO2.Share_of_World
ORDER BY 4 desc
;
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	CO2_Emissions(tons,2016)	CO2 Emissions per Capita	Share_of_World
China	10432751400	7.3077	0.2918
United States	5011686600	15.3217	0.1402
India	2533638100	1.8731	0.0709
Russia	1661899300	11.4036	0.0465
Japan	1239592060	9.7451	0.0347
Germany	775752190	9.3324	0.0217
Canada	675918610	18.2313	0.0189
Iran	642560030	7.8552	0.018

Figure 6

```
#5 Year Total Solar & Wind By Country 2016-2020
SELECT S.Country, S.2016_Total as '2016 Capacity Solar (MW)',
    W.2016_Total as '2016 Capacity Wind (MW)',
    SUM(S.2016_Total + W.2016_Total) as '2016 Capacity Solar & Wind (MW)',
    S.2017_Total as '2017 Capacity Solar (MW)', W.2017_Total as '2017 Capacity Wind (MW)',
    SUM(S.2017_Total + W.2017_Total) as '2017 Capacity Solar & Wind (MW)',
    S.2018_Total as '2018 Capacity Solar (MW)', W.2018_Total as '2018 Capacity Wind (MW)',
    SUM(S.2018_Total + W.2018_Total) as '2018 Capacity Solar & Wind (MW)',
    S.2019_Total as '2019 Capacity Solar (MW)', W.2019_Total as '2019 Capacity Wind (MW)',
    SUM(S.2019_Total + W.2019_Total) as '2019 Capacity Solar & Wind (MW)',
    S.2020_Total as '2020 Capacity Solar (MW)', W.2020_Total as '2020 Capacity Wind (MW)',
    SUM(S.2020_Total + W.2020_Total) as '2020 Capacity Solar & Wind (MW)'
FROM Solar S JOIN Wind W
ON S.Country = W.Country
GROUP BY S.Country, S.2016_Total, W.2016_Total, S.2017_Total, W.2017_Total, S.2018_Total,
    W.2018_Total, S.2019_Total, W.2019_Total, S.2020_Total, W.2020_Total
ORDER BY 4 desc
;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	2016 Capacity Solar (MW)	2016 Capacity Wind (MW)	2016 Capacity Solar & Wind (MW)	2020 Capacity Solar (MW)	2020 Capacity Wind (MW)	2020 Capacity Solar & Wind (MW)	5 Year Change (MW)	5 Year Change (%)
China	78070	168690	246760	254355	281993	536348	289588	1.1736
United States	40300	82183	122483	75572	117744	193316	70833	0.5783
Germany	41220	50019	91239	53783	62184	115967	24728	0.2710
Japan	42750	3234	45984	67000	4206	71206	25222	0.5485
India	9010	28665	37675	39211	38559	77770	40095	1.0642
Italy	19279	9257	28536	21600	10839	32439	3903	0.1368
Spain	4669	23075	27744	14089	27089	41178	13434	0.4842
United Kingdom	11630	15030	26660	13563	24665	38228	11568	0.4339
France	7130	12065	19195	11733	17382	29115	9920	0.5168
Canada	2715	11898	14613	3325	13577	16902	2289	0.1566
Brazil	0	10740	10740	7881	17198	25079	14339	1.3351
Australia	5900	4327	10227	17627	9457	27084	16857	1.6483
Turkey	832	6101	6933	6668	8832	15500	8567	1.2357

Figure 7