Project Report

2024-03-29

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(tidyverse)
```

```
## Warning: package 'forcats' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
```

```
library(showtext)
```

```
## Warning: package 'showtext' was built under R version 4.3.3
```

```
## Loading required package: sysfonts
```

```
## Warning: package 'sysfonts' was built under R version 4.3.3
```

```
## Loading required package: showtextdb
```

```
## Warning: package 'showtextdb' was built under R version 4.3.3
```

```
font_add_google("Open Sans", "Open Sans")
showtext_auto()
```

```
df <- read.csv("data/owid-energy.csv")
```

Question 1

- **Question**: What are the patterns and comparisons in electricity generation among the top 5 countries by energy consumption/electricity generation/GDP and globally in 2023? How do these patterns differ between developed and developing countries or across continents?

We intended to answer the question for year 2023, as 2023 is thought to be the most recent year which have available data for patterns of electricity generation for each country. In this question, we primary needs the data about the share of each type of electricity generation in a country, which is encoded as "x_share_elec" in the table. We want to answer the above question because it can give us insights about what is the different between practices in the top countries in GDP, as well as the difference between low-income countries and high-income countries with regard to this problem.

However, due to the lack of data in 2023, we can only answer the question for the most recent year which have available data, which is 2021. As such, we will choose the data in 2021 in this question.

In the first aspect of the question, we will plot the constitution of energy generation of 5 countries with the highest GDP worldwide. However, due to our lack of information about GDP in the dataset, we have to get 5 countries in the Internet. From our research, 5 countries with the highest GDP are China, the US, India, Japan and Germany.

```
df_2021 = df[df$year == 2021, ]

df_2021_top_5 <- df_2021[df_2021$country == "China" | df_2021$country == "India" | df_2021$country == "
```

From that, we will obtain the important metrics for each country, which are the share of each type of electricity generation.

```
df_2021_top_5_important_metric <- df_2021_top_5[,c("country", "population", "gdp", "biofuel_share_elec"
```

```
df_2021_top_5_important_metric
```

```
##             country population gdp biofuel_share_elec coal_share_elec
## 4349          China 1425893504  NA              2.003          62.932
## 7731        Germany   83408560  NA              8.057          28.253
## 9287          India 1407563904  NA              2.070          74.173
## 10185         Japan  124612528  NA              3.851          32.510
## 20674 United States  336997632  NA              1.307          21.624
##       fossil_share_elec gas_share_elec hydro_share_elec low_carbon_share_elec
## 4349              66.289          3.213           15.323                33.711
## 7731              48.487         16.354            3.377                51.513
## 9287              78.053          3.745            9.356                21.947
## 10185             71.002         35.119            8.256                28.998
## 20674             60.509         38.037            5.936                39.491
##       nuclear_share_elec oil_share_elec other_renewables_share_elec
## 4349               4.803          0.144                       2.003
## 7731              11.873          3.880                       8.098
## 9287               2.563          0.135                       2.070
## 10185              6.387          3.373                       4.166
## 20674             18.742          0.848                       1.746
##       renewables_share_elec solar_share_elec wind_share_elec
## 4349                 28.908           3.854          7.727
## 7731                 39.640           8.474         19.691
```

```
## 9287                19.384           3.986           3.973
## 10185               22.611           9.254           0.935
## 20674               20.750           3.960           9.108
```

However, entries such as "renewables_share_elec" can be overlapping with other types of eletricity generation, so we hypothesize that the electricity generation types which are the most common are: Biofuel, Coal, Gas, Hydro, Nuclear, Oil, Solar and Wind. We test if our assumption is right by add them together.

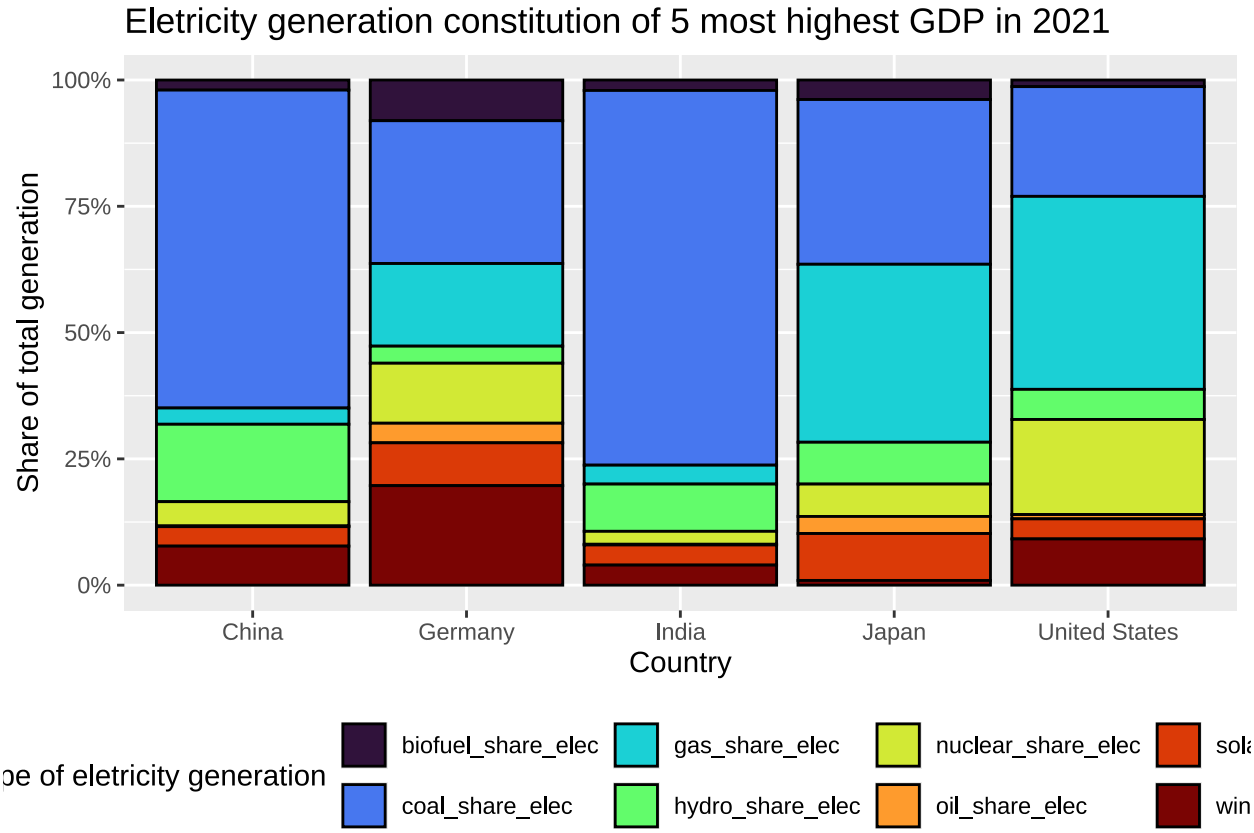
```
df_2021_top_5_important_metric$total = df_2021_top_5_important_metric$biofuel_share_elec + df_2021_top_5
df_2021_top_5_important_metric$gas_share_elec +
df_2021_top_5_important_metric$hydro_share_elec +
df_2021_top_5_important_metric$nuclear_share_elec +
df_2021_top_5_important_metric$oil_share_elec +
df_2021_top_5_important_metric$solar_share_elec +
df_2021_top_5_important_metric$wind_share_elec
```

We see that, the results are approximately 100%, so other kinds of electricity generation is there, but not too much.

```
df_2021_top_5_important_metric <- pivot_longer(df_2021_top_5_important_metric, cols = c("biofuel_share_
```

In this question, to illustrate the constitution of 5 highest GDP countries, we will use the percent stacked bar chart.

```
ggplot(df_2021_top_5_important_metric, aes(fill=eletric_share_type, y=percentage, x=country)) +
  geom_bar(position="fill", stat="identity", color = "black") +
  scale_fill_viridis_d(option = "turbo") +
  theme(legend.position = "bottom") +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Eletricity generation constitution of 5 most highest GDP in 2021", x = "Country", y = "S
```



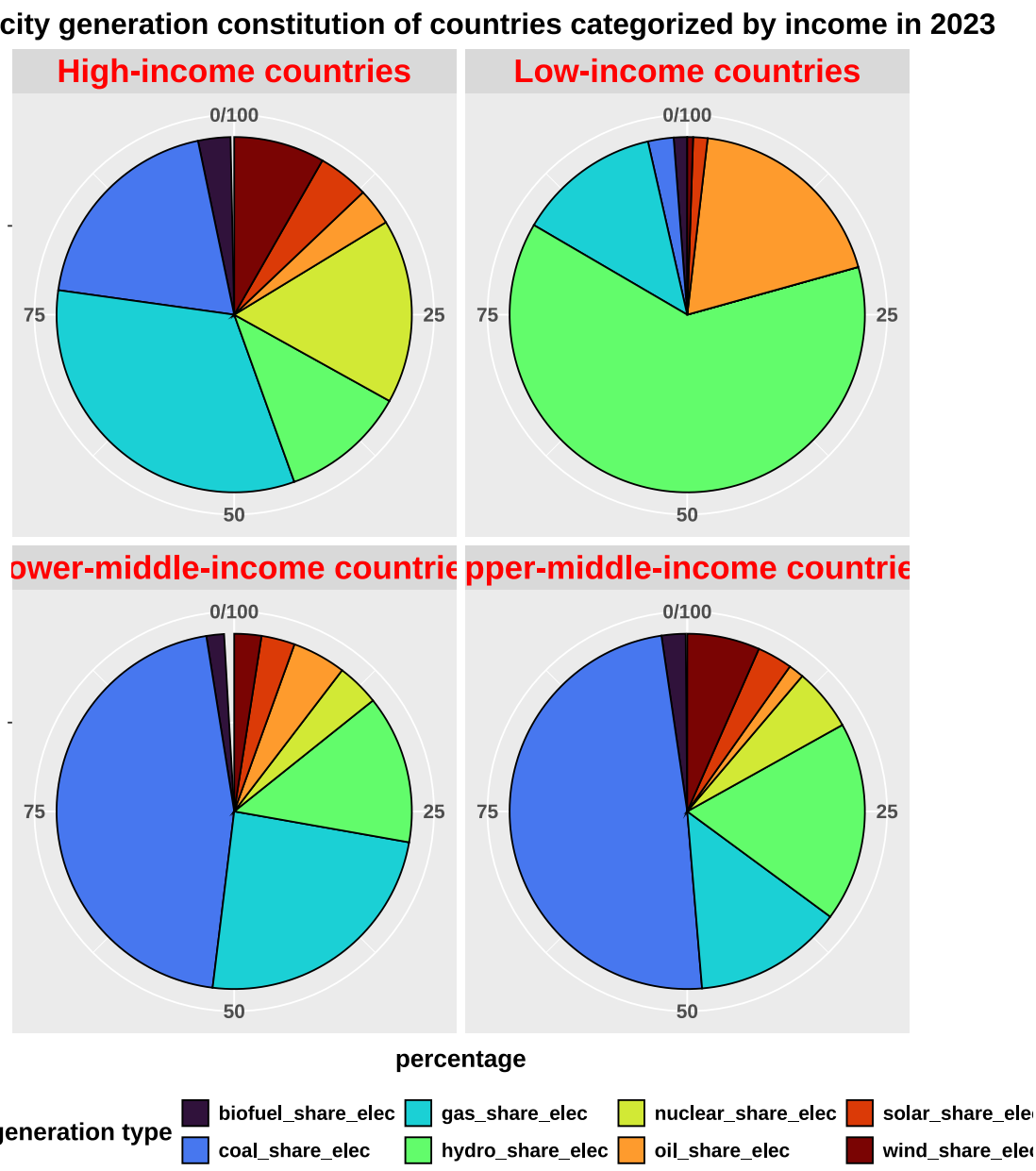
We can see that, there is a wide difference between the constitution of electricity generation in 5 countries above, for example, China and India tends to use more coal as the means of generating electricity, whereas Japan and the US favor gas as the main resource for electricity generation. Overall, we cannot tell a general trend from just the graph above. We need more information.

We suppose that the electrical generation constitution of a country correlates with the development stage of a country. We try to plot 4 pie charts corresponding to: High-income countries, Low-income countries, Lower-middle-income countries, Higher-middle-income countries to see if there is some correlation between 5 countries above and the pie charts generated.

```
df_2021_categorize <- df_2021[df_2021$country == "High-income countries" | df_2021$country == "Low-incor
```

```
df_2021_categorize_important_metric <- pivot_longer(df_2021_categorize, cols = c("biofuel_share_elec",

ggplot(df_2021_categorize_important_metric, aes(x = "", y=percentage, fill=eletric_share_type)) +geom_ba
  scale_fill_viridis_d(option = "turbo") +
  facet_wrap(.~ country,  nrow = 2) +
  theme(legend.position = "bottom", plot.title = element_text(hjust = .5), text=element_text(face="bold
  coord_polar("y", start=0) +
  labs(title = "Eletricity generation constitution of countries categorized by income in 2023", fill =
```



Discussion

Question 2