Reem Saleh

Documentation for R Coal Power Visualization Project

In order to clean the datasets, visualize the data, and report on findings, I used R and R Studio. All of the data used was given to me in the form of Excel files, which I initially imported as a CSV file in R, cleaned, processed them, and eventually saved the files into new MS Excel Files labeled Processed Data for the respective years 2017 and 2023. I used resources including the Global Energy Monitor website and its wiki pages to understand the meaning behind the columns and rows of the data, and I also used the documentation notes that were included in the Excel files. The libraries I used in R were used to conduct data wrangling and analysis as well as create visualizations, some of which were interactive. I did some manual cleaning of the data in MS Excel by making sure the columns were formatted correctly and column names were the same for both datasets. I also defined what columns to keep in order to optimize the dataset and get rid of unnecessary columns, most of which were blank and had no data. I chose my visualizations based on the type of data I had. The capacity MW data was numeric, but the country names were categorical, so the best graph that would fit that type of data is a bar graph. Most of my visualizations were bar graphs. However, in the future I do intend to explore other visualizations and explore other parts of the data more in depth.