IT UNIVERSITY OF CPH

Data Visualisation Exam Project

May 16th, 2024

Abel Szemler
absz@itu.dk

Tobias Heine Zschiegner Michelsen
tomi@itu.dk

This report is presented for the course:
Data Visualisation and Data-driven Decision Making (Spring 2024)
4th semester of BSc, Data Science
IT University of Copenhagen
Denmark

1 Introduction

This report outlines the journey from the initial brainstorming of interesting concepts to the development of a fully-fledged data visualization. The goal is to present the group's thought process and to highlight the underlying steps involved. The report will be structured as follows:

1. Brainstorming and Theme Selection
2. Data Collection, cleaning and exploration
3. Intended Message with the Visualization
4. Design Rationale
5. Potential Ideas for Further Research
6. Visualisations

2 Brainstorming and Theme Selection

To begin with, the group investigated potential ideas and themes that were interesting to us personally. We ended up with a broad spectrum of themes, namely: Wealth, Climate, Diversity, Social Equality, War Spendings, Social Media, and Sport. Furthermore, we also looked into data types and agreed on working with time series data and geographical data.

This initially led us to the idea of working with the climate theme. We wanted to see how pollution hotspots evolved over time. Additionally, we wanted to introduce the theme 'Wealth' to investigate whether there exists a strong relationship between the two. Due to the nature of pollution and wealth being very broad themes, we chose to narrow it down. Instead of investigating pollution as a whole, we focused on CO_2 emissions. Similarly, 'Wealth' was boiled down to Gross Domestic Product (GDP). It is important to mention that these two metrics are not representative of the themes as a whole. There are many more metrics which could be relevant, but for the scope of this project, we concluded that it would be best to keep it on the simpler side.

3 Data Collection, cleaning and exploration

The data used in this project originates from four datasets that were ultimately merged into one. The first dataset contains the yearly CO_2 emissions from fossil fuels for every country [1]. In this dataset, there is a large disparity in the times when emissions were first recorded. To illustrate, Austria has yearly CO_2 records from 1807 and onwards, whereas Angola only has records from 1950 till now. Additionally, the data has missing values as well as records stating zero emissions for certain entries. For example, the data suggests that Saudi Arabia had zero emissions in the years 1951 and 1952. These fluctuations or potential inaccuracies in the data are difficult to account for and have a large impact on the visual interpretation. After creating a map based on annual data, certain countries would start to "blink" when going from colored to neutral, which most likely is not due to the countries' emissions rapidly decreasing or increasing. To counter this we filled up the dataset, either with zeroes or the value in the previous year. By doing this, we altered the dataset, while aiming to maintain a truthful representation.

The second dataset contains data on every countries' population between 10.000 BCE and 2021 [2]. These values were used in conjunction with the yearly CO_2 emissions to estimate the CO_2 emission per capita for every country.

The third dataset contains data on every countries' share of global CO_2 emission, also from fossil fuel sources, represented as a percentage [3].

Lastly, the fourth dataset contained every countries' GDP in billion USD from 1960 to 2022 [4]. Territories and regions such as "Middle East", "European Union (27)" and "Low-income" were filtered out, as we were only interested in countries. Furthermore, the time span was filtered to only include values between the year 1870 and 2021.

4 Intended Message with the Visualization

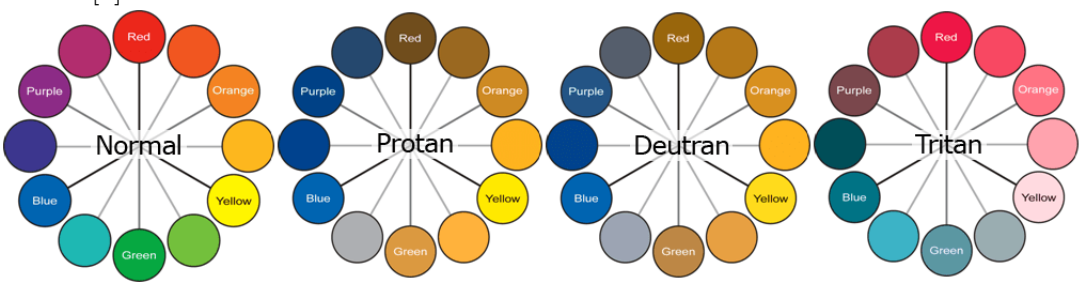
Our intention was to create an interactive tool to explore the changes in CO_2 emission of countries throughout history. The goal is to allow people to use this dashboard to potentially answer some questions or discover something interesting about our environment and the challenges we are facing. We include two maps that use different measures of emissions, one looks at how much CO_2 a country as a whole emits, while the other map takes into consideration the different population sizes as well. These two maps allow a viewer to compare countries from different perspectives and may allow some to arrive at interesting conclusions.

The message, other than raising awareness about the increased level of CO_2 in our air and its possible consequences, is the importance of making one's own decisions. Starkly different conclusions can be made about a topic if we approach it from a different viewpoint, so it is important to have an open mind and to take multiple perspectives into consideration.

5 Design Rationale

After deciding on the topic and what type of data we were using, maps were an obvious choice to represent the changing levels of CO_2 emissions in countries over a given time frame. We thought they looked fascinating as an animated graph, and that they emphasize important aspects of the data through color, like the overall increasing tendency and the differences between countries. We wanted to complement the maps with related graphs and arrived at a treemap showing a share of global CO_2 emissions per country over the same time-frame as the map. This treemap works in unison with the maps and can be used to make some additional comparison. Finally, we added a line graph showing the changing levels of GDP and annual emission of countries since 1960, which provides some insight into the connection of GDP and emission.

For the maps we chose a green-red diverging color palette, green to symbolize the lower emission thus more environmentally friendly state of countries, and red for the opposite. However, we would like to note that a green-red palette is not colorblind friendly and for a more inclusive design we would choose a blue-red palette, although for now we would like to keep it as is due to what each color represents.[6]



Choosing a center point and finding an appropriate transformation of the values, was crucial in making the maps more vibrant with a higher variety in color. Both annual emission of a country and its per-capita variant had a high variety meaning that most countries were neutrally colored while very few had a bright red color. To even out the disparity between countries we took the square root of the shown values which decreased the range drastically. Next center points had to be chosen. for the per-capita map we based it on the Paris Agreement, which has a goal to limit the temperature increase to 1.5 degrees Celsius above pre-industrial levels. To do this, now each person should have an average 2.3 tons of CO_2 emitted annually [5]. So we used this value as a center point for the color range. For the annual emission, this is not so easy, as countries come in all different sizes and there is not one metric that fits them all, so we averaged every registered CO_2 value and divided by the number of observations and used that as a center point. As for the treemap and line graph we used the same Tableau color package to represent the countries with the same color on both graphs.

6 Potential Ideas for Further Research

Further research could use more specific location data in connection with emission, creating a heat map with increased intensity at the local area where the emission happens. This could show a more detailed map with intensely colored hubs, where most emission happens. It could also be interesting to investigate how CO_2 emissions are outsourced vs insourced.

7 Visualisations

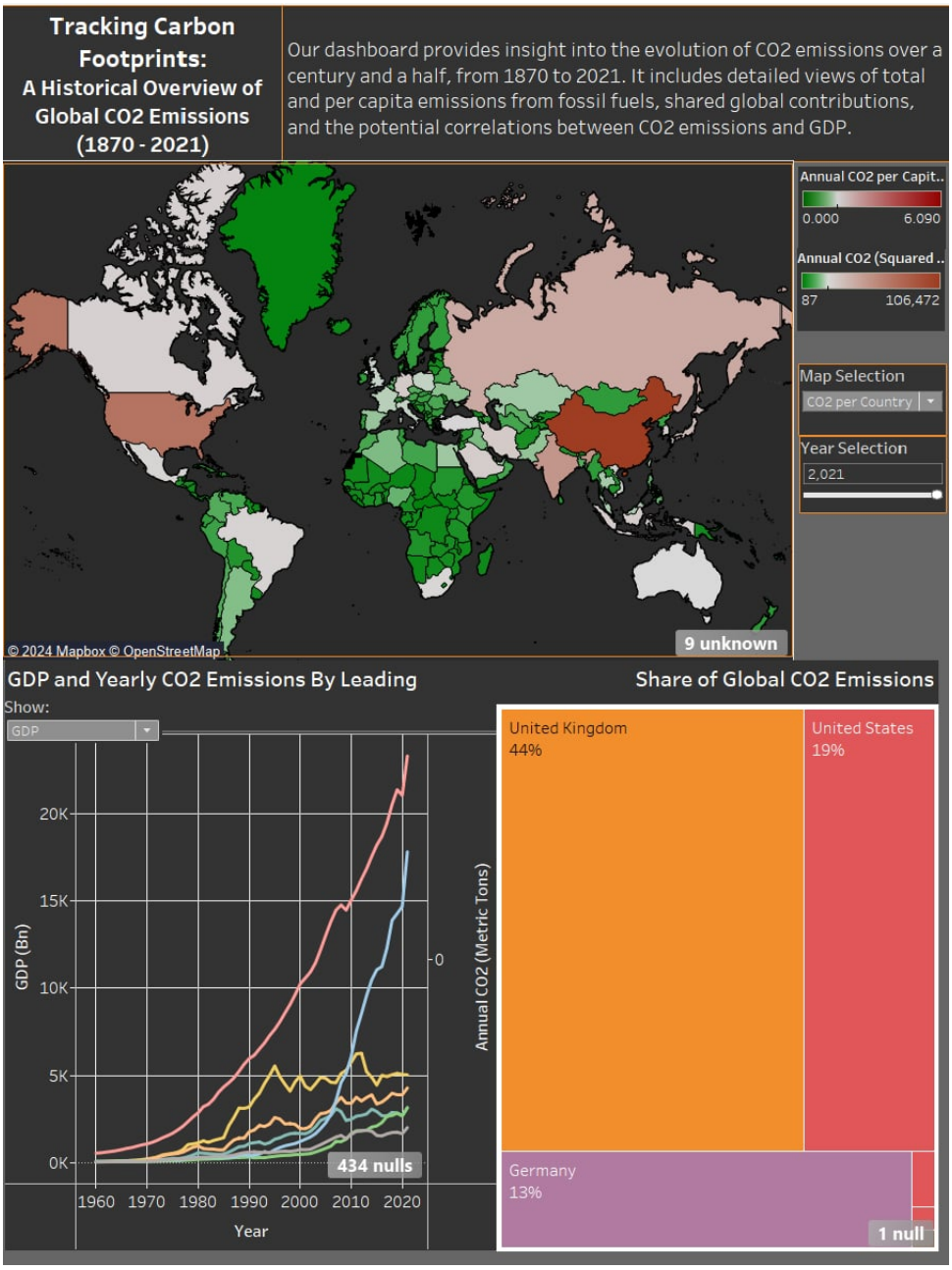


Figure 1: The complete dashboard features a map with a dropdown for toggling between CO2 per-capita and per country views, a line graph with a dropdown to select between leading countries' GDP or yearly CO2 emissions, and a treemap that displays the share of global CO_2 emissions, all integrated with a year selection slider that also updates the map. All 3 visualisations

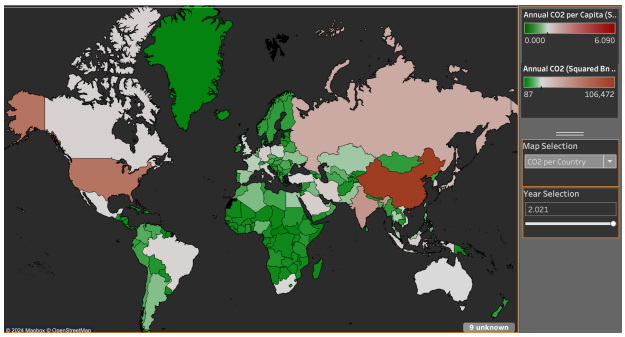


Figure 2: Map showing the yearly CO_2 emissions per country.

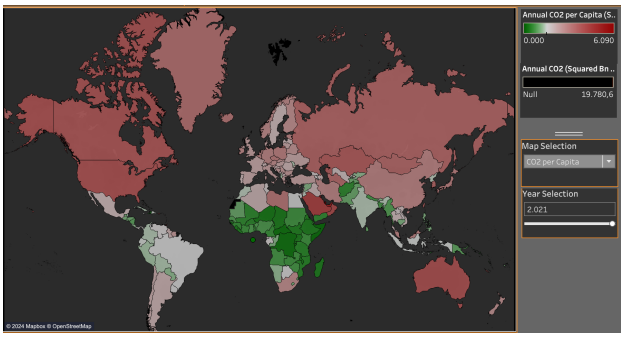


Figure 3: Map showing the yearly CO_2 emissions per Capita for every country.

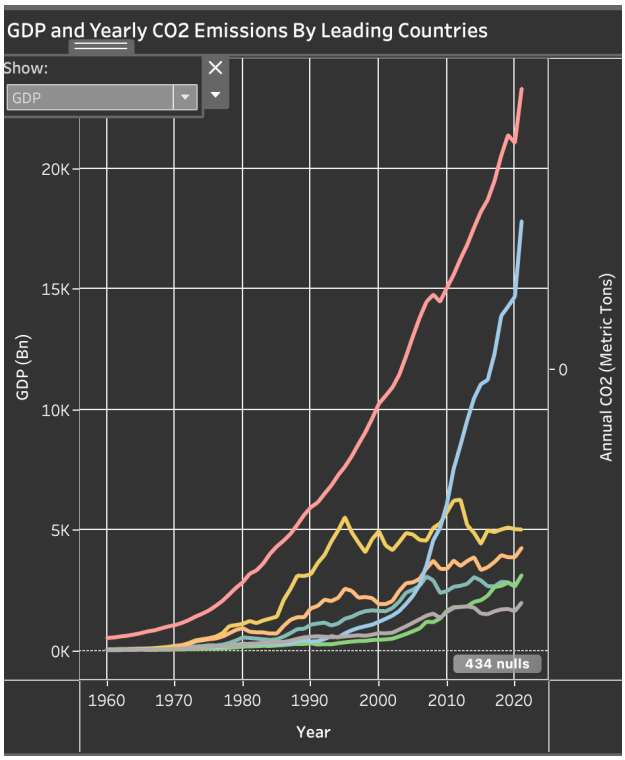


Figure 4: Line Graph with dropdown to toggle between GDP and the Yearly CO_2 emissions by Leading Countries (U.S.A, India, China, Japan, Germany, U.K , Canada)

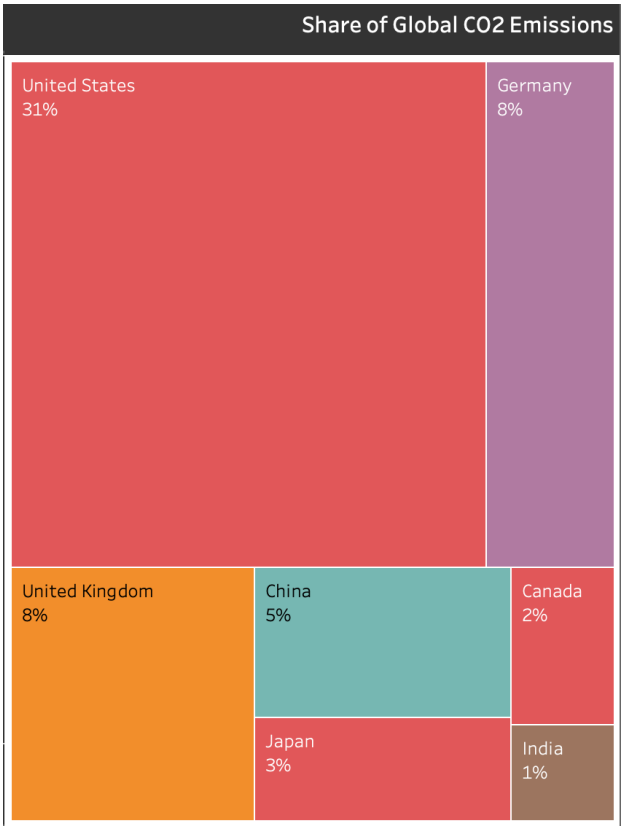


Figure 5: Treemap showing share of global CO_2 Emissions for Leading Countries (U.S.A, India, China, Japan, Germany, U.K , Canada). The Year selection slider updates the map accordingly.

References

[1] "CO_2 emissions," Our World In Data, [Online]. Available: https://ourworldindata.org/co2-emissions. [Accessed: May 14, 2024].

[2] "Population, 10,000 BCE to 2021," Our World In Data, [Online]. Available: https://ourworldindata.org/grapher/population. [Accessed: May 14, 2024].

[3] "Share of global cumulative CO emissions" Our World In Data, [Online]. Available: https://ourworldindata.org/grapher/share-of-cumulative-co2. [Accessed: May 14, 2024].

[4] "GDP (current USD)," World Bank national accounts data, and OECD National Accounts data files, [Online]. Available: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD. [Accessed: May 14, 2024].

[5] "Carbon emissions of richest 1% set to be 30 times the 1.5°C limit in 2030," Kai Tabacek; Annie Thériault, [Online]. Available: https://www.oxfam.org/en/press-releases/carbon-emissions-richest-1-set-be-30-times-15degc-limit-2030. [Accessed: May 14, 2024].

[6] "How To Use Color Blind Friendly Palettes in Your Design," Rachel Cravit, [Online]. Available: https://venngage.com/blog/color-blind-friendly-palette/. [Accessed: May 14, 2024].