Data Visualization Report

Catarina Palha (M20190156)
Mafalda Zúquete (M20190257)
Maren Leuthner (M20190134)

January 2020

1 Introduction

The objective of this project was to create a dashboard using data on greenhouse gases to further analyze and highlight world pollution. This was achieved by gathering the emissions per capita for each country in the world in order to understand which countries are the biggest emitters of polluting gases.

With this purpose, multiple visualizations were created to allow the user to get collective insights and perspectives from the same data set. By diversifying the types of visualization and the interactivity that could be realized with the created dashboard, the user should be able to form a sophisticated conclusion from the data set presented.

2 Data Description

The data was gathered from the World Bank Data, which is considered a timely and reliable source of information. First, the data underwent a brief cleaning and merge of several data sets. This was followed by the calculation of the emission and GDP (Gross Domestic Product) values per capita. Thus, the data set had the following variables:

- Country Name
- Year
- CO2 (carbon dioxide) Emissions
- CH4 (methane) Emissions
- N2O (nitrous oxide) Emissions
- GHG (greenhouse gases) Emissions
- GDP

The units of the emissions were in kilo tonnes of CO2 and the GPD in current US dollars. The per capita values were simply obtained by dividing the absolute values of emissions and GDP by the corresponding population values for the relevant year and country.

3 Visualization and Interaction Choices

3.1 Inspiration for Visualization

The main inspiration behind the theme selection were the climate school strikes that had the young generations as protagonists, with the simple demand for scientists to be heard. By exposing the scientific data from multiple angles, the voices could indeed be amplified.

Through the addition of GDP data to the initial data set on emissions, a more validated source of information could be given to the user regarding the reasons behind the emissions, elevating the role of governments regarding the issue of climate change.

The main intention of this framework is to highlight and expose the countries that contributed to air pollution the most. Subsequently, it was decided to use per capita numbers rather than raw data, as it would disregard the size of a country and therefore create a fairer basis for comparison and readability.

3.2 Visual Design

The visual design of the dashboard started with the title at the top of the page, which was followed by the interactivity component parts where the user could engage with the displays and, lastly, the tab navigation bar containing the plots with the chosen selection in the interactive components.

3.3 Interaction

The dashboard was built with three main visualizations to better explore all facets of the data. A consistent visual framework was chosen in order to get a clean overview by switching only the tabs displaying the different visualizations:

- World Map;
- Time Series Plot;
- Bar Plot.

The visualizations explored looked to find visual evidence of the ever growing greenhouse gases emissions and if there were any possible relations between economic power and emission levels.

Familiar widgets were used in the creation of the dashboard for explicit interactivity with the user:

- Dropdown menus;
- Radio buttons.

3.4 Messaging

The interactivity was restricted to the key dimensions regarding the chosen data set: country, year and type of gas. Throughout the dashboard, the user was explicitly guided by the following captions to quickly draw the attention of the user to what is essential in the dashboard:

- Choose one or more countries out of the database;
- Choose a projection: equirectangular or orthographic;
- Choose a scale: linear or logarithmic;
- Choose one or multiple gases out of the database.

A suggested action was already selected, e.g. Portugal was pre-selected as a country to display in the time series data and the bar plots.

Additionally, each visualization had a small caption that clarified which filters were relevant for the specific plot.

4 Reading the Visualization

Scientific data is usually presented in such a way that the average layperson is intimidated. Through visualizations that are easy to understand and by bringing the user into the conversation using interactivity, subjects like climate change and pollution become more accessible to average people, a relevant fact since these people are also affected by this phenomenon and should be able to understand it.

4.1 Data Encoding

Choropleth maps were used for geographic geometry data and were represented with one quantitative attribute for each region. The encoding was made by using the given geometry for the boundaries' marks of each area. The quantitative attribute for choropleth maps was encoded using color [4].

Line charts were ordered by a key attribute and use additionally one quantitative value attribute. The visual effect in a line chart was the usage of point marks with a line connecting all them [2].

Bar charts used a quantitative value, as well as a categorical key attribute. Line marks were used to encode and express the quantitative value attribute along the vertical position, while the separating key attribute could be found in the horizontal position [2].

4.2 Data Filtering

Interactivity components were used in order for the user to be able to engage and specify the preferences regarding the plots that are shown below the navigation bar tabs. This was done by:

- Filtering the year for the world maps and the bar plots;
- Selecting one or more countries to be displayed in the bar plots and time series plots;
- Selecting whether the scale is logarithmic or linear in the time series and bar plots;
- Selecting a gas option to be shown in the world maps and the time series plots.

Plotly already allows the user to grasp detail on demand in the functionality of "zooming in/out" in the graphics, as well as resetting the plot to its original setting prior to the zooming action.

4.3 Reading the World Map

The first visualization that was used was a world map. Since the data set had information on emissions and GDP by country by year, spatial position is an essential element of this data set and, therefore, something the should be represented [4]. By using two maps side-by-side, a clear overview of the emissions and GDP scenario could be obtained. A correlation between emissions per capita and GDP per capita was found with regard to CO2 and total GHG emissions, regardless of the year. When it came to CH4 and H2O emissions, the results are less striking as, overall, there are less emissions of the gases and the countries emit similar values.

Using color, the countries were labelled according to the following scale:

Color	Meaning
Red	High
Yellow	Medium
White	Low
Black	No data available

Both these visualizations could be filtered by year and the projection of the map could also be chosen by the user: either equirectangular or orthographic. However, only the first map could also be filtered by gas.

4.4 Reading the Time Series Data

In order to better explore the evolution of emissions and GDP overtime for each country, it was decided to add line charts to the visualizations.

By including these into the panoply of graphics, a better insight could be gotten regarding trends and correlations, reinforcing the conclusions suggested by the map.

The user could interact with these plots by selecting a country or multiple countries and the type of gas emissions that are to be analysed.

The user would also be able to choose if they wanted to use a logarithmic or linear scale for the graphic [5], and the interval they want to focus on, either by using the slider bar under each plot, or by selecting the buttons that were designed as:

Button	Meaning
YTD	Last year
5Y	Last 5 years
10Y	Last 10 years
ALL	All years available

The use of these line plots highlighted the connection between time, emissions and GDP, and also how economic growth and an increase in emissions tend to work in tandem.

The time series plot served the purpose of highlighting the evolution of the emission levels and economic growth. Similar to the connection that was found between economic power and amounts of GHG emitted, here it could be seen that a positive growth in economy tended to come hand in hand with an increase in emissions. However, unlike with the maps, overall, this tendency was recorded for all gases and not just CO2 and GHG.

By allowing the user to select multiple countries at once, it is also possible to get a comparison between countries and get further insights on how different countries have evolved regarding pollution and economic level.

4.5 Reading the Bar Plot Emissions

Bar plots can be used to lookup and compare values on a scale [2]. This visualization technique was used to separate and order each gas per selected country.

The last visualization in the dashboard was a bar plot exposing the different types of emissions by a country on a selected year, side-by-side. The bar plot in the dashboard showed a line mark and encoded the amounts of each gas emitted by a country as a quantitative value attribute with one spatial position channel.

The user could select the year and the countries they wanted to visualize using this graphic. Also, selecting multiple countries allowed users to get a comparative view for all emissions.

Another interactive component was the ability to choose between a logarithmic scale and a linear one [5].

This visualization showed which type of gas the countries were emitting the most each year and also made it easier to compare countries' emissions.

Unsurprisingly, the highest number of emissions corresponded to CO2; but again, the ability to select multiple countries at once allowed to get a detailed comparison between countries with regard to actual emissions.

5 Color Choice

Choosing color is a significant factor in data visualization [4]. It can be said that a visualization is effective if it works both in black and white and in color, and color blindness of the population should be taken into account. Furthermore, too many colors can harm the effectiveness of the visualization [1].

For these reasons, a sequential color scale was chosen for the maps. The usage of the NASA maps on climate change [3] led as an inspiration to the choice of a color scale going from yellow to read, despite the fact that the color scale is not fully dependent on saturation. However, it was taken into account for the

choice of a color scale since the color scale from the NASA maps is already recognized by the general public. The usage of an alternative scale, such as a red scale, would likely not be well interpreted due to the cultural link between red and danger.

Still, this scale was used on both maps and not only on the map with emissions data. It should be pointed out that the map with GDP was included for comparison with the emissions data. For that reason, it was decided to use the same scale on both maps, as it would make correlations more obvious at a glance.

For the other two visualizations, there was no need to use a color scale. Nonetheless, whenever more than one country was selected, there is a need to differentiating them. Generally, there is no harm in using different hues to differentiate categorical data, with the caveat that once too many colors are added, the plot becomes hard to interpret [4].

Even so, the bar plot and the time series were not meant to show data for all countries at the same time, but to compare data for two, maybe three countries at once. While the user could select as many countries as desired, this would not be advised and would defeat the purpose of the visualization.

6 Technical Aspects

This project was implemented using Plotly with Dash software. Dash was the tool that allowed the dashboard to be fully interactive, but Plotly already had functionalities that promoted user input, such as letting certain lines or bars to be hidden on plots, or returning the values of a section of the graphic when the mouse hovers over it.

Both these packages also allowed for an almost full customization of the plots on the programmer's side, making it easier to implement a vision and deliver it.

However, further styling of the overall dashboard had to be done using CSS, namely positioning the different components of the App where they were wanted or changing the background color of the App.

During the development of the App, GitHub was used in order to optimize group work and avoid mistakes, through the use of branches. Besides its advantages as a tool for group work, the delivery of the project also involved the use of GitHub platform to be able to use the cloud platform Heroku.

7 Final Discussion and Critical Reflection

The objective of the project of displaying several meaningful interactive visualization items using Plotly with Dash software has been achieved. Visualization concepts and techniques explored during this course were used to transform data into interactive visualizations.

However, during the development of this project the main difficulty was found in bringing an idea for a visualization into a practical and meaningful dashboard that would also be interactive and visually pleasing. Not to mention, the technical limitations that always come with working for the first time with new tools and languages, such as Dash, Plotly and CSS.

Another shortcoming of this project was the lack of data after 2012, missing many relevant dates in the climate discussion such as the 2015 Paris agreement and the release of the latest IPCC (Intergovernmental Panel on Climate Change) report, which fostered climate activism. This limited the possibility of reaching practical and relevant up-to-date conclusions to an ever evolving problem.

Still, the data that was used did a good job of portraying the evolution of GHG emissions since 1990, the year that the first IPCC report was published, bringing the climate change issue to light in a way that had not been done before.

Had there been more time, more data and variables could have been gathered to paint a more complex and nuanced picture of the complexity of the issue that is climate change and the role of humanity in it.

References

[1] David Borland and Russell M.Taylor II. Rainbow Color Map (Still) Considered Harmful. *IEEE computer graphics and applications*, 27(2):14–17, 2007.

[2] Tamara Munzner. Rules of Thumb. In *Visualization Analysis and Design*. CRC Press, 2014.

[3] NASA. Climate Time Machine. `https://climate.nasa.gov/interactives/climate-time-machine`. [Online; accessed 15-January-2020].

[4] Claus O. Wilke. *Fundamentals of Data Visualization: a Primer on Making Informative and Compelling Figures*. O'Reilly Media, Inc., 2019.

[5] Naomi Robbins. When Should I Use Logarithmic Scales in My Charts and Graphs? `https://www.forbes.com/sites/naomirobbins/2012/01/19/when-should-i-use-logarithmic-scales-in-my-charts-and-graphs`, 2012. [Online; accessed 15-January-2020].