MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS

Data Visualization Project Report

Presented by: Diogo Hipólito, m20210633
Guzel Bayazitova, m20210699
Jessica Routzahn, m20210987
Mohamed Ali Felfel, m20211322

Course: Data Visualization 2022

Submission 17th April 2022

Paper Contents

Additional Information

GitHub Repository: The group's GitHub repository can be found here: https://github.com/DiogoNovaIms/data_vis

App Location: A link to the dashboard/app through Heroku that has been developed can be found here: https://average-temperature.herokuapp.com

***Please note that on April 12, GitHub suffered a cyber security attack impacting GitHub's ability to upload and connect withHeroku and Travis-CI. This problem was not solved before the April 17 deadline. ***

Abbreviations

NOAA	National Oceanic and Atmospheric Administration
GDP	Gross Domestic Product
GHG	Greenhouse Gas
CO2	Carbon Dioxide

FIGURES

Figure 1: Static Version of Animated Choropleth Map
Worldwide Average Temperature Evolution from 1900 to 2010

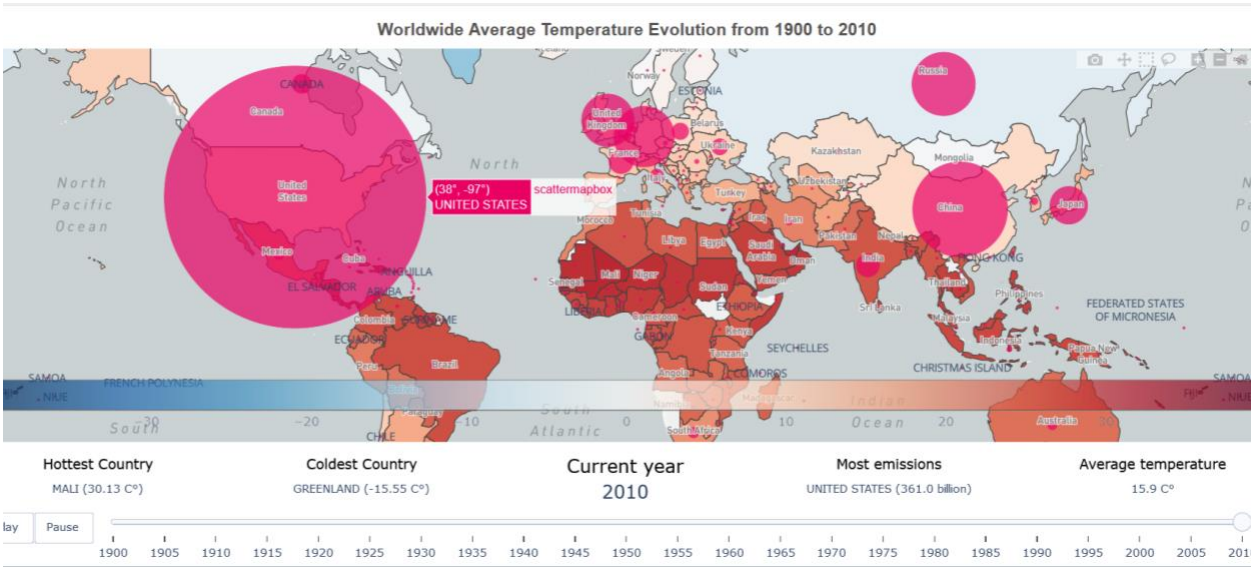


Figure 2: Static Version Correlation Scatterplot
GHG and Increase in Average Temperature

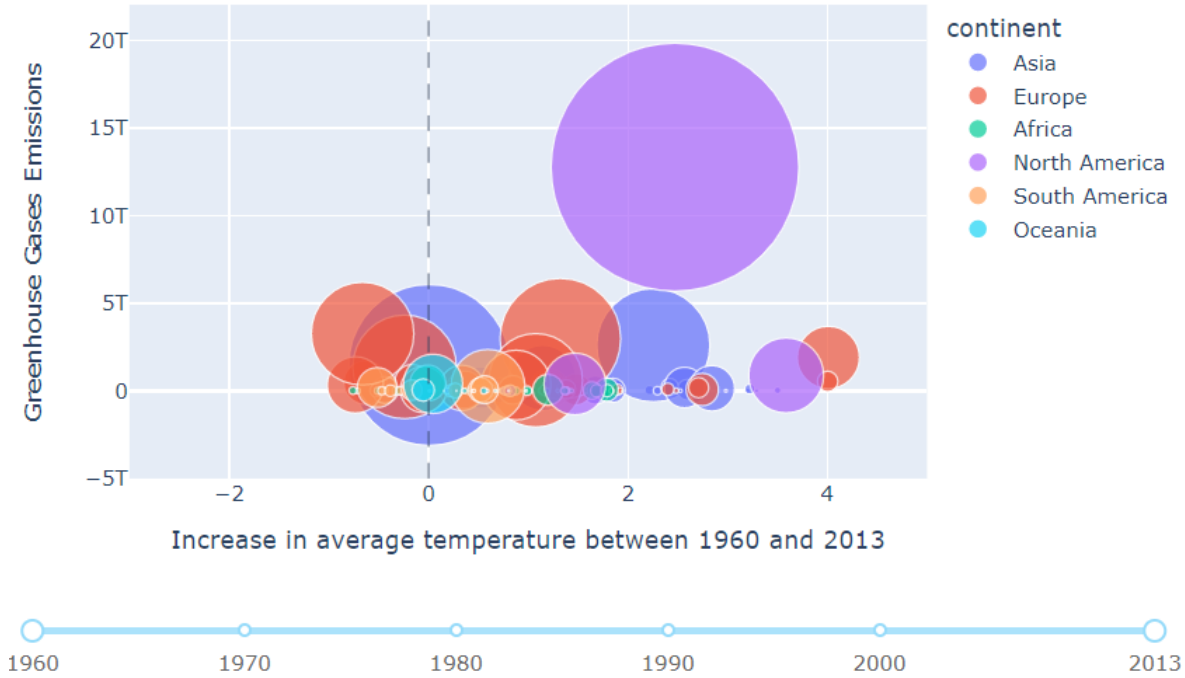
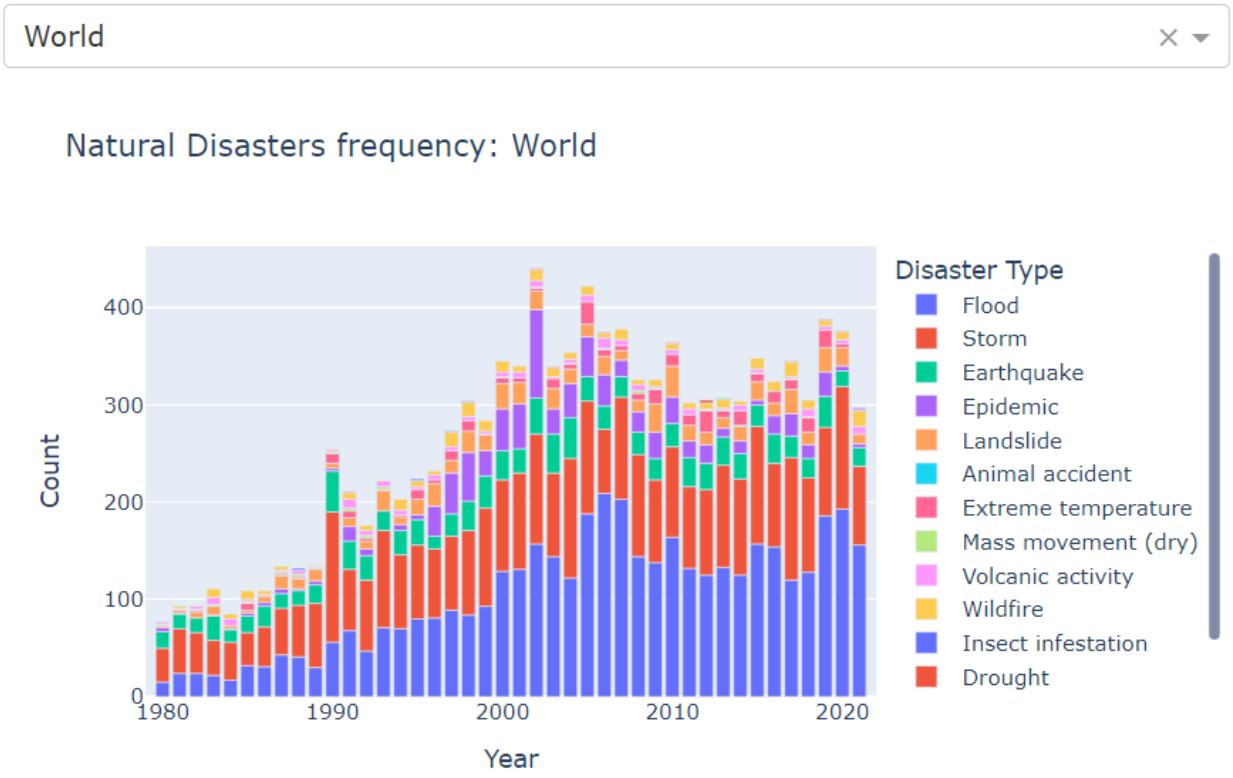


Figure 3: Static Version of Stacked Bar Chart
Disasters by Country and Slider with Year



TABLES

Table 1: Global Surface Temperature Data Description

VARIABLE	TYPE	DESCRIPTION
dt	object	Date-time of when temperature was collected
Average Temperature	float64	Temperature value
Temperature Uncertainty	float64	Temperature uncertainty value
Country	object	Country name

Table 2: GHG Emission Data Description

VARIABLE	TYPE	DESCRIPTION
Country	object	Country name
Year	int64	Year emissions were captured
GHG	float64	Amount of emissions in tons

Table 3: Natural Disaster Data Description

VARIABLE	TYPE	DESCRIPTION
Year	int64	The year the disaster took place
Disaster Group	object	One unique value - Only Natural Disaster
Disaster Subgroup	object	Subgroup of disaster - example: Hydrological or Meteorological
Disaster Type	object	Type of disaster such as earthquake or flood
Disaster Subtype	object	Subtype of disaster such as storm, tropical cyclone or convection storm
Country	object	Country of natural disaster occurrence
Region	object	Region of natural disaster occurrence
Continent	object	Continent of natural disaster occurrence

Table 4: GDP Data Description

VARIABLE	TYPE	DESCRIPTION
Country Name	object	Country name
Code	object	Three letter country code
1960	float64	GDP in the year 1960
1961	float64	GDP in the year 1961
1962	float64	GDP in the year 1962
…..	…..	…..
…..	…..	…..
2020	float64	GDP in the year 2020

1. Introduction

Accepted by virtually the entire international community, climate change and global warming are often central topics of discussion. This project draws its inspiration from this very real and pertinent issue affecting the international community for centuries to come. Research from the United States National Oceanic and Atmospheric Administration's (NOAA) climate division categorizes and catalogs global temperature increases[1]. However, looking at global temperature increases only provides a piece of the global warming story. Why are global temperatures rising, are some places getting hotter than others, and who is being affected by these temperature increases? In order to address these questions additional data besides global temperatures was needed.

The burning of fossil fuels emits billions of tons of carbon and other GHGs into the atmosphere each year inadvertently harming the global planet[2]. Continued growth in GHG emissions will lead to an atmospheric concentration not experienced in tens to hundreds of millions of years. Despite reductions in recent years the globe is not yet at a rate that would limit global average temperature change to well below 3.6°F (2°C), or pre-industrial levels. As the world continues to burn fossil fuels global surface temperature will continue to increase causing changes to global environments. Characteristics of extreme events such as hurricanes, extreme flooding events, drought, and wildfires may even increase in frequency.

This project looked at global temperature data from 1700 to 2021, global greenhouse gas (GHG) emissions, 120 years of natural disaster data, and gross domestic product (GDP) data in order to find answers to some of the above questions and display them in a meaningful way through visualizations. The idea behind including these additional data sources was to show not only that global temperatures are rising, but that emissions are rising as well.

2. Dataset Descriptions

For this project we started data exploration with global temperature data. The global surface temperature data[3] has been compiled on Kaggle from Berkeley Earth, which is affiliated with the Lawrence Berkeley National Laboratory. This data set combines 1.6 billion temperature reports from 16 pre-existing archives. A brief description of the variables in this dataset is provided in **table 1**. This spatial data provides the average temperature for each country in the dataset for every year from 1900 to 2010.

The next step in the group's data journey was to find and include possible explanations for the increase in global surface temperature. While emissions alone are not the signal cause for global climate change they are a leading human-caused increase in climate change, especially carbon dioxide (CO2) through the burning of fossil fuels. The CO2 and GHG emissions dataset[4], also collected from Kaggle, contains emission data of greenhouse gasses (GHG) for different countries from the year 1750 to 2019. A brief description of the variables in this dataset are provided in **table 2**.

As we know global climate change has many measurable impacts on human society and the environment affecting the world's population more each day. Understanding the parameters created by the climate crisis is important for deciding the measures society must take against it. The group wanted to show not only that temperatures have been rising, but these rising temperatures are impacting the globe through natural climate related events and disasters. A

dataset originally published from Kaggle was added to the larger database for this project. The natural disasters dataset[5] contains natural disaster data by country from 1900 to 2021. A brief description of the variables in this dataset are provided in **table 3**.

The last dataset that was incorporated into this project was gross domestic product (GDP) data[6] originally gathered from the World Bank. This dataset contained the GDP of over 200 countries for all years between 1960 and 2020. A brief description of the variables in this dataset are provided in **table 4**. This last dataset was added and incorporated into the project because the group wanted to explore the relationship between the richest nations and the highest emitters of GHGs thereby their overall effect on the changes in global temperature rise.

2.1. Data Encodings

For this project data encodings were used for each visualization. Starting with the first visualization the animated choropleth map shows the temperature quantitative attribute encoded as color over global regions and countries delimited as area markers[7]. The region shapes were provided by the original dataset. For the color map the group chose to represent the temperature scale from blue to red with blue representing decreases in average temperature and red representing increases in temperature.

For the correlation scatterplot two quantitative value variables, temperature and emissions, were encoded using both vertical and horizontal spatial position channels. The mark type is a point. Each point represents a country with horizontal and vertical spatial position encoding the primary quantativities attributes of average temperature and GHG emissions. The color channel is used for the categorical region attribute and the size channel for quantitative GDP attribute. This scatterplot is effective at providing the viewer with a characterization of the relationship between emissions and temperature for each country. With this encoding viewers can easily make informed judgments about the highest emitters. This plot and encodings used were particularly useful in pointing to outliers. In this case the outliers are not really outliers at all but the focal point of the plot.

For the stacked bar chart color and length were used to distinguish between each disaster type. Each component of the bar is separately stacked so that the full bar height shows the value of for the combination of disasters[8]. The viewer or user is able to easily distinguish between the heights of the lowest bar component, the full combined bar, and compare against other bars across the time scale.

3. Visualizations

Once all of the data was collected and compiled for the project the group needed to create visualizations in order to explain what is happening with the data. For this project all files and work was performed in GitHub and executed through Heroku. The dashboard created for this project contains many interactions available to the user. Interactions are important as they allow the user to see many combinations of the sample data and visual encodings as opposed to status views that only show one aspect of the dataset.

3.1. Layout

The layout for the visualization in this project was designed by the group, it was not based on layouts available in the Plotly Dash Gallery. Html components were used to divide areas in the dashboard. In order to set the stage and present the overarching storyline viewers are presented with an animated map when they first open the dashboard. Then they are suggested to scroll down and observe two further visualizations, which can be in turn updated via year range slider and country drop-down menu.

3.2. Visualization and Interaction Choices

As described above the first visualization is an animated map showing the increases to global temperature and bubbles representing greenhouse gas emission over time. A static image of this visualization can be seen in **figure 1**. An animated map was chosen over a static map in order to better represent the change in both temperature and emissions over an expansive amount of time. The idea was to help people better track changes as opposed to having the audience click through various points in time[9]. There are many interactions available to the audience through this visualization. The map can be zoomed and adjusted to the desired area. The user is able to hover over each country and view temperature information about each country. The user is also able to play and pause the visualization at different time periods using the slider at the bottom of the map. While the visualization is running the bottom interface displays the hottest country and the coldest country, as well as the country with the highest level of GHG emissions and its average temperature in degC at that time period. This adjusts as time moves forward.

Additionally this dashboard includes a correlation plot between GHG and the increase in average surface temperature with GDP as the element size. This plot can be seen in **figure 2**. It is showing that the total increase to average temperature over time is related to emissions. This plot shows that the highest emitters, the United States and China, are not experiencing the greatest number of average temperature increases. Users are able to interact with this graph by toggling a specific decade, 1960 through 2013, with the year range slider at the bottom of the correlation plot.

The dashboard also contains a stacked bar chart of natural disasters. The stacked bar chart contains the total number of natural disasters by type per given year from 1980 through 2020. A static image of this can be seen in **figure 3**. The user is able to interact with this visualization in several ways. First the user can hover their cursor over a given year for additional information about the number of disasters by type for that year. The user is also able to select region specific data to display using a dropdown. The disaster data options are by world or global data, by region, and by country.

4. Discussion

This project and the visualizations created in it have accurately displayed and encapsulated the fact that the global planet is indeed warming. We then further offer the user or viewer to interact with the visualizations finding different combinations and scenarios for relationships between global surface temperatures, global emissions, global natural disasters, and global GDP. Do the richest nations emit more GHGs than other nations having a disproportionate effect of global

climate change? This project cannot begin to infer this based on the available data we have, however the richest countries do not seem to be experiencing the effects of climate change at the same rate as other nations in Africa and Southeast Asia while emitting more than these regions as well.

4.1. Limitations

One possible limitation of this project is the type of data used to generate visualizations. Temperature is essential to measuring and describing global climate trends, however the type of temperature data has varying effects on human life. For this project the group only used surface temperature data. Temperature data from varying sources such as air and sea temperatures could lead to more dynamic research on natural disaster and climate related events giving a more comprehensive look at how temperature affects humans and ecosystems.

4.2. Future Work

Future work for this project could take many shapes and forms. The group for this project pulled data from four different sources containing completely different data. Additional and future work could be explored in many ways for each of these datasets. For example, the natural disaster dataset could be explored in many ways including pulling information about the nominal costs of each disaster, the GDP impact of the disaster years or decades after it occurs, impacts on the population for each occurrence of disaster, and much more. The effects of climate warming are possibly limitless. Future work and analysis should aim to explore all the possibilities.

References

[1] Rebecca Lindsey, & Luann Dahlman. (2021, March 15). *Climate Change: Global Temperature*. Https://Www.Climate.Gov/News-Features/Understanding-Climate/Climate-Change-Global-Temperature.

[2] Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., DeAngelo, B., Doherty, S., Hayhoe, K., Horton, R., Kossin, J. P., Taylor, P. C., Waple, A. M., & Yohe, C. P. (2017). *Executive summary. Climate Science Special Report: Fourth National Climate Assessment, Volume I*. https://doi.org/10.7930/J0DJ5CTG

[3] Kaggle. (n.d.-b). *Climate Change: Earth Surface Temperature Data*. Accessed on 04/01/2022.Https://Www.Kaggle.Com/Datasets/Berkeleyearth/Climate-Change-Earth-Surface-Temperature-Data?Select=GlobalLandTemperaturesByCity.Csv.

[4] Kaggle. (n.d.-c). *CO2 and GHG emission data*. Accessed on 04/01/2022. Https://Www.Kaggle.Com/Datasets/Srikantsahu/Co2-and-Ghg-Emission-Data.

[5] Kaggle. (n.d.-a). All Natural Disasters 1900-2021/ EOSDIS. Accessed on 04/01/2022. *Https://Www.Kaggle.Com/Datasets/Brsdincer/All-Natural-Disasters-19002021-Eosdis*.

[6] Kaggle. (n.d.-d). World GDP(GDP, GDP per capita, and annual growths). Accessed on 04/01/2022.*Https://Www.Kaggle.Com/Datasets/Zgrcemta/World-Gdpgdp-Gdp-per-Capita-and-Annual-Growths?Select=gdp_per_capita.Csv*.

[7] Munzner, & Tamara. (n.d.). *Visualization Analysis & Design Tamara Munzner A K Peters Visualization Series Illustrations by Eamonn Maguire*.
.[pg. 181]

[8] Munzner, & Tamara. (n.d.). *Visualization Analysis & Design Tamara Munzner A K Peters Visualization Series Illustrations by Eamonn Maguire*. [pg. 147]

[9] Munzner, & Tamara. (n.d.). *Visualization Analysis & Design Tamara Munzner A K Peters Visualization Series Illustrations by Eamonn Maguire*. [pg. 132-133]