CS 4804 – Final Project

Process Book

Alysha Creelman, Connor Ehrensperger, Audrey Mongillo

February 2024

01

Overview and Motivation

Understanding Climate Change

The goal of this project is to provide an exploratory interface that allows user's to see how the emission of greenhouse gases in the United States and the rest of the world has changed over the last decade. When greenhouse gases are emitted into our atmosphere, they trap the sun's heat, leading to global warming and climate change. We feel that it is important for our readers to understand where the highest concentrations of these gases have been coming from.

All members of our team have an interest in climate change that motivated us to create this visualization. Audrey completed a GPS freshman year related to climate change that has sparked her interest. Alysha did her humanities seminar on "climate fiction." Exploring the literature related to climate change made her interested in exploring data visualization in this field.

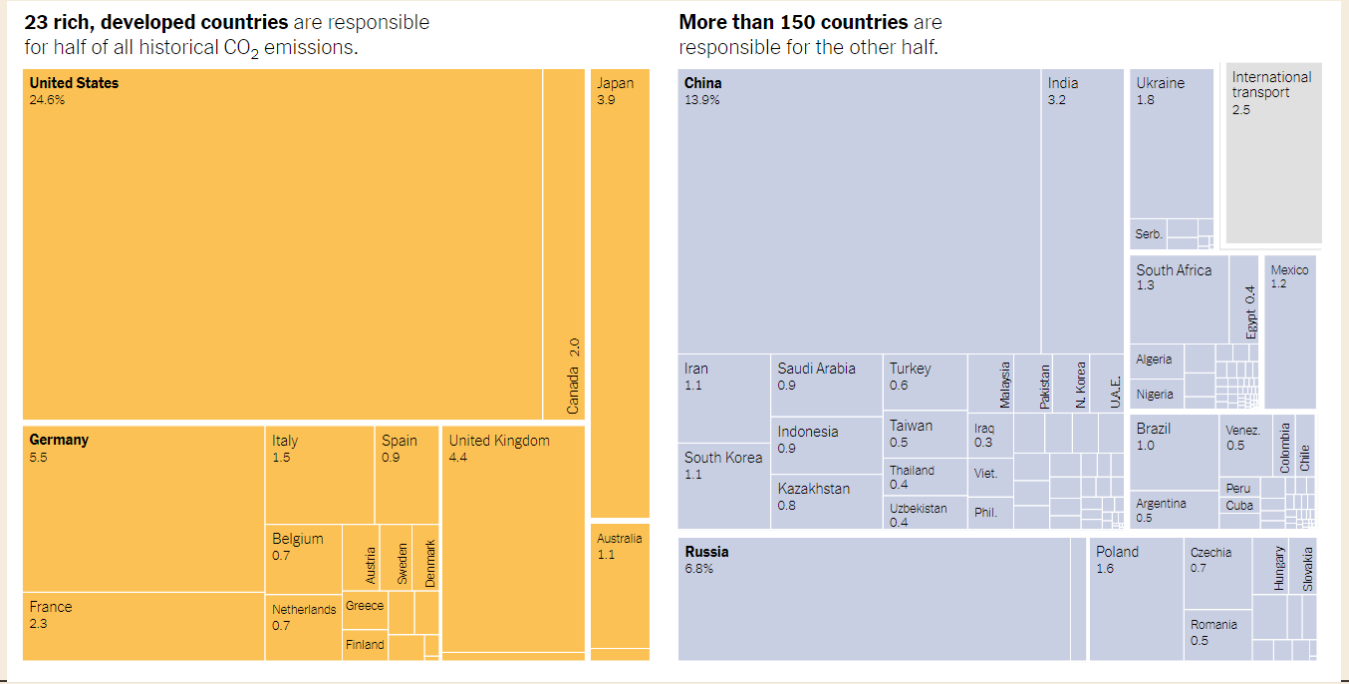
02

Related Work

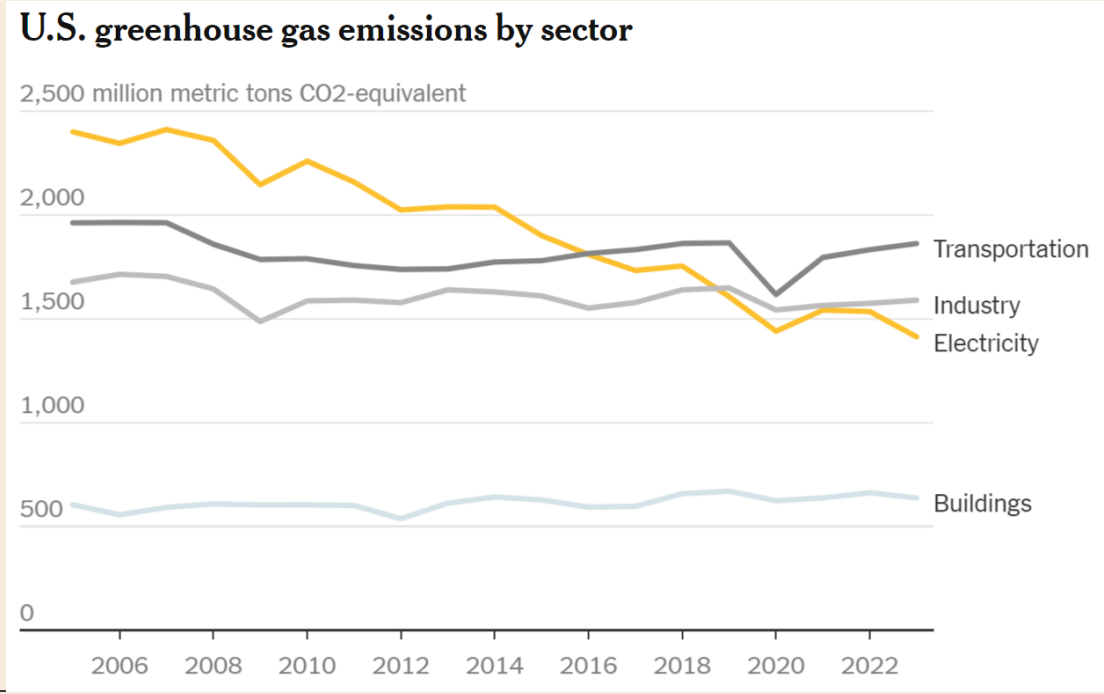
New York Times

The New York times has been doing great work related to climate change visualization. With professionals such as Nadja Popovich and Blacki Migliozzi who are committed to teaching the public about climate change through engaging visuals, the New York Times has built up an impressive collection of articles and graphics. Teach About Climate Change With 30 Graphs From The New York Times is an article published by Michael Gonchar on January 31, 2024 featuring some of the New York Times' top climate change data visualizations. These visualizations have inspired our work and gave us many ideas about how we can visualize such data.

This visualization, which can be found inside the New York Times article Who Has The Most Historical Responsibility for Climate Change, gave the team the idea to search for a second dataset that would allow us to explore emissions by country. We ended up using a tree map as one of our visualizations as well.



This visualization, which can be found inside the New York Times article U.S. Carbon Emissions Fell in 2023 as Coal Use Tumbled to New Lows, showed the team how effective simple line graphs can be in showing trends over time. A similar graph was created for showing emissions by country over time.



03

Questions

Original Questions

- What is the breakdown of greenhouse gas emissions in the United States?
- How have greenhouse gas emissions changed in the last decade in the US?
- What states have the highest emissions?

While we wanted to keep only looking at data from 2010-2022 for consistency, we decided that we would like to broaden the scope of this project to put the United States in context with the rest of the world. As a result, we developed additional questions.

Additional Questions

- How do greenhouse gas emissions in the United States compare with the rest of the World?
- Which 5 countries produce the most greenhouse gases?
- Do other countries follow the same trend as the United States in steadily decreasing greenhouse gas emissions over the last decade?
- What percent of global gas emissions can be attributed to these top 5 countries?

04

Data

EPA Dataset

Link to site with data: https://ghgdata.epa.gov/ghgp/main.do#/facility/

We first began our search by looking for datasets that would provide a breakdown of greenhouse gas emissions within the United States. After looking at some data sets, we were able to find the U.S. Environmental Protection Agency's data on greenhouse gas emissions reported by large facilities across the United States through the Data Is Plural Weekly Newsletter. These facilities are required to report annual data about their emissions to the EPA as part of the Greenhouse Gas Reporting Program (GHGRP).

While this data only records the greenhouse gas emissions from these large facilities, we still feel that it is a good way for the user to explore the breakdown of emissions across the United States.

EPA Dataset- Cleanup

- Data from the EPA had to be downloaded as multiple separate CSV files (one for each type of gas) and spanned multiple sheets within each file
- To clean the EPA data, we wrote a simple Python script that combined multiple Excel pages into one large CSV file using the pandas library. After this we just had to remove repeated header rows, which was easy enough to do by hand.

Our World in Data

Link to repository: https://github.com/owid/co2-data

When we decided to expand the scope of our project, we started looking for datasets that would break down greenhouse gas emissions around the world. We found the Our World in Data Github repository containing a dataset on CO2 and greenhouse gas emissions. This dataset is built upon a number of other datasets from the Energy Institute, U.S. Energy Information Administration, Global Carbon Project, and more.

Our World in Data- Cleanup

When first downloaded, this data contains 48,059 rows and 79 columns. In an effort to make the data easier to digest, all rows for years outside the range of 2010-2022 were removed using Excel. Next, the data was made into a table where the team was able to filter on country to see the different countries represented in this dataset. Because the dataset contains rows for each continent, the world, and a few other grouped areas, any rows that did not represent one country were removed to ensure that no country was represented twice.

Creating Multiple Smaller Datasets

When working with this data, our team ended up creating a few smaller datasets that were used in the creation of the graphs. For example, the dataset used to create the tree map only has 6 different country types; China, United States, India, Russia, Japan, and All Other Countries. The sum of greenhouse gas emissions for all other countries was calculated in Excel.

country	year	total_ghg
China	2010-2020	123167.1
United States	2010-2020	63386.64
India	2010-2020	33193.49
Russia	2010-2020	18664.76
Japan	2010-2020	13237.03
All Other Countries	2010-2020	250553.1

05

Exploratory Data Analysis

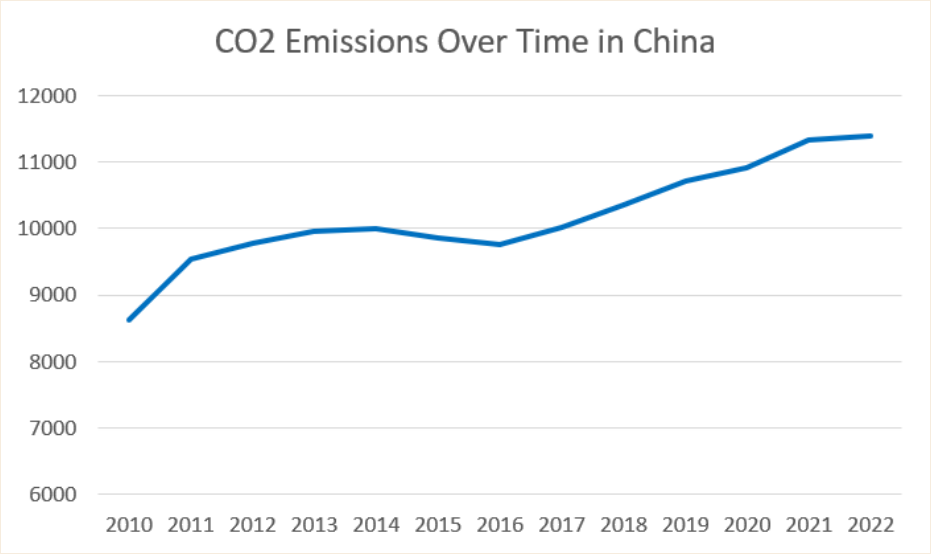
Our World in Data- Exploratory Analysis

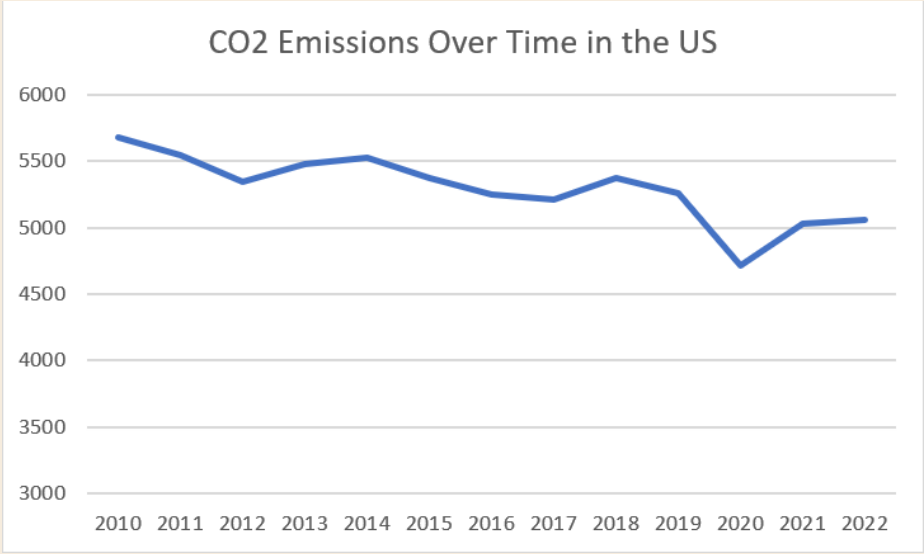
Due to the size of this dataset, the team decided to do some exploratory analysis of the data in Excel. Since we were interested in learning which countries are the top contributors to global emissions, we decided to filter the data on the year 2022 and sort based on highest number of CO2 emissions. As it turns out, China, United States, India, Russia, and Japan has the top 5 highest emissions in the world for all years between 2010 and 2022. A dataset with these 5 countries and their data was created. While we ended up switching over to visualizing combined GHGs only for simplicity (which does not have data for 2021-2022), the top countries stayed the same.

A	B	C
country	year	co2
China	2022	11396.777
United States	2022	5057.304
India	2022	2829.644
Russia	2022	1652.177
Japan	2022	1053.798

Our World in Data- Exploratory Analysis

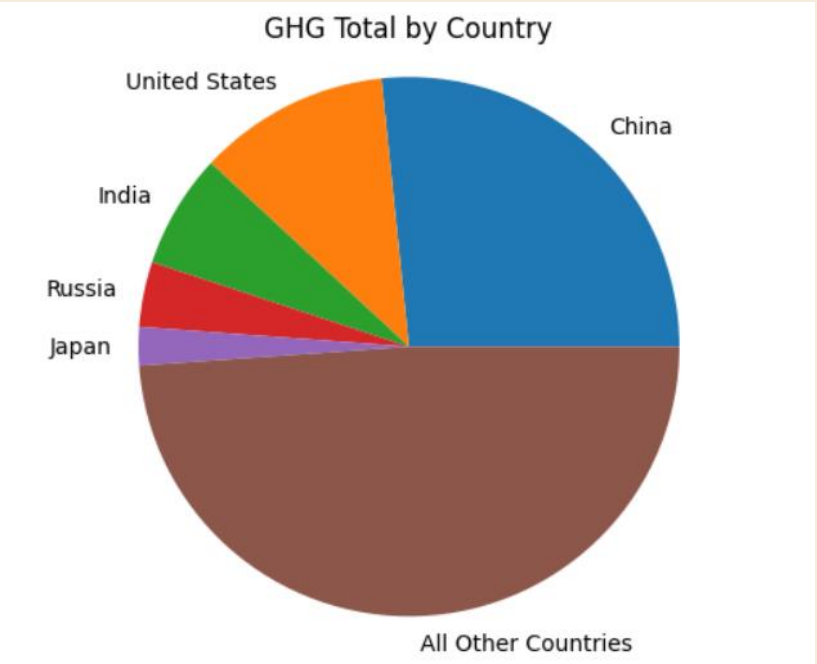
In order to see if there are differences in the trend of emissions in each of the top 5 countries, the team created a line graph for the top 2: China and the United States. After seeing that these trend lines are clearly different, we decided that a line graph showing all 5 trends on one graph would be interesting and informative.





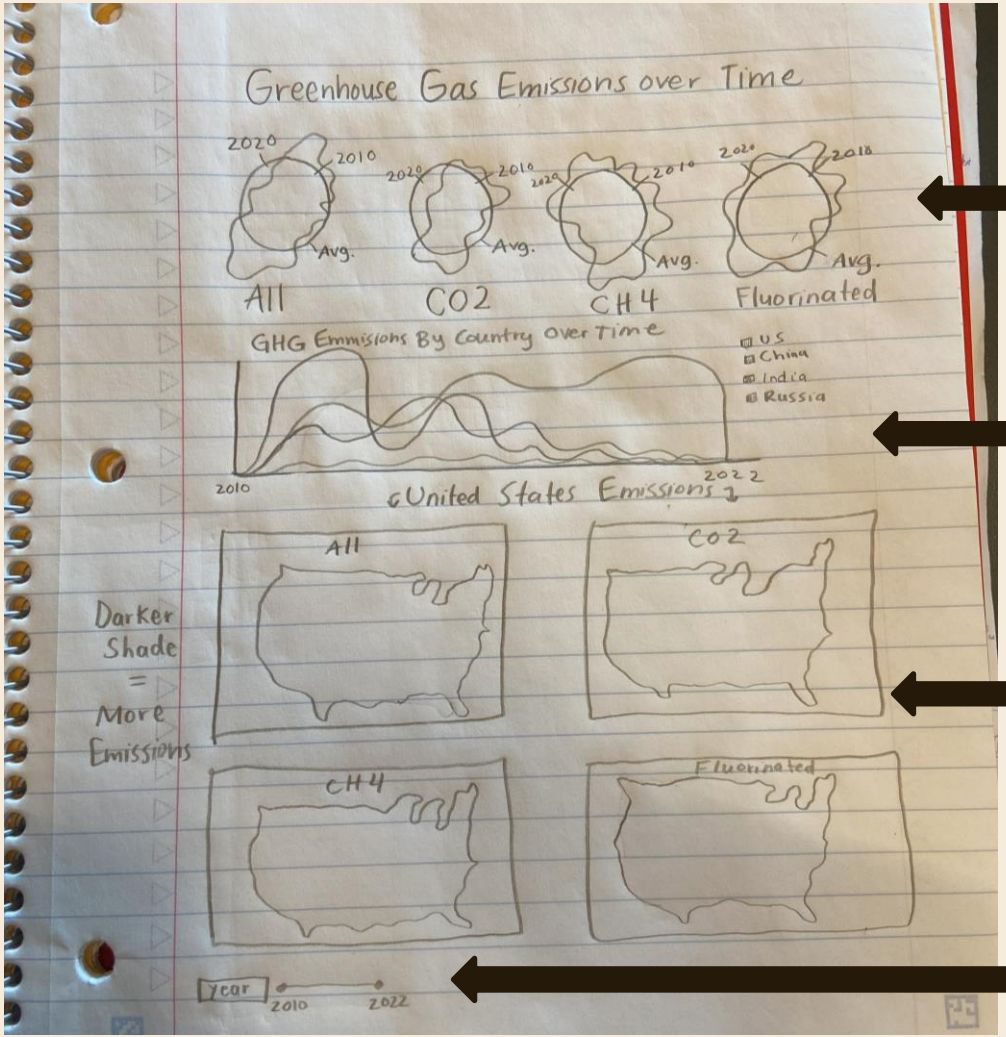
Exploratory Analysis in Python

- Python was also used to perform some exploratory data analysis of the country data
- Creating line graphs and pie charts helped uncover something interesting: The top 5 countries with highest emissions account for nearly 50% of global emissions



06

Design Evolution



Radial Plots to show emissions over time

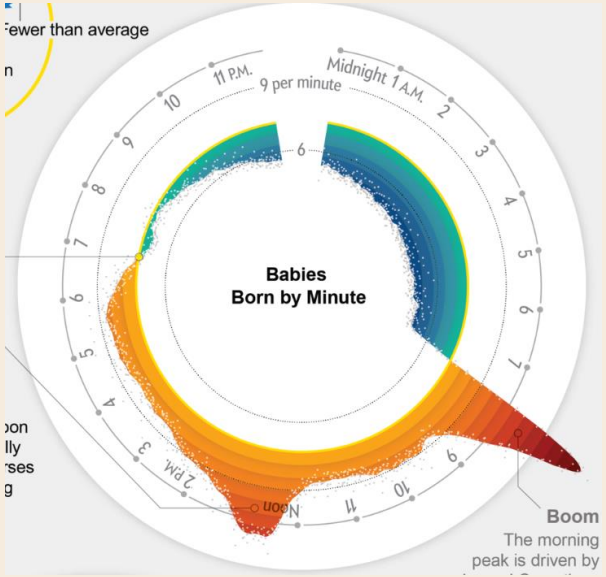
Area graph of emissions over time by country

Maps of concentration of each gas type in the US

Allow the user to filter on year 2010-2022

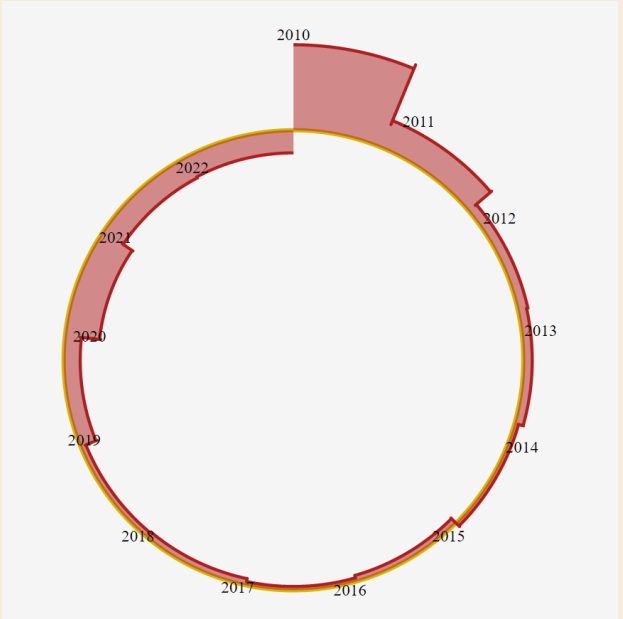
Issues with Radial Plots

After our first attempt at implementing a radial plot, we realized an issue: Unlike our inspiration graph, we do not have data in small enough increments to create a nice, smooth graph. Since we only have data for each year, there is not a way for us to make our lines smooth without misleading the reader. This is difficult to read, so we decided against the radial plots.



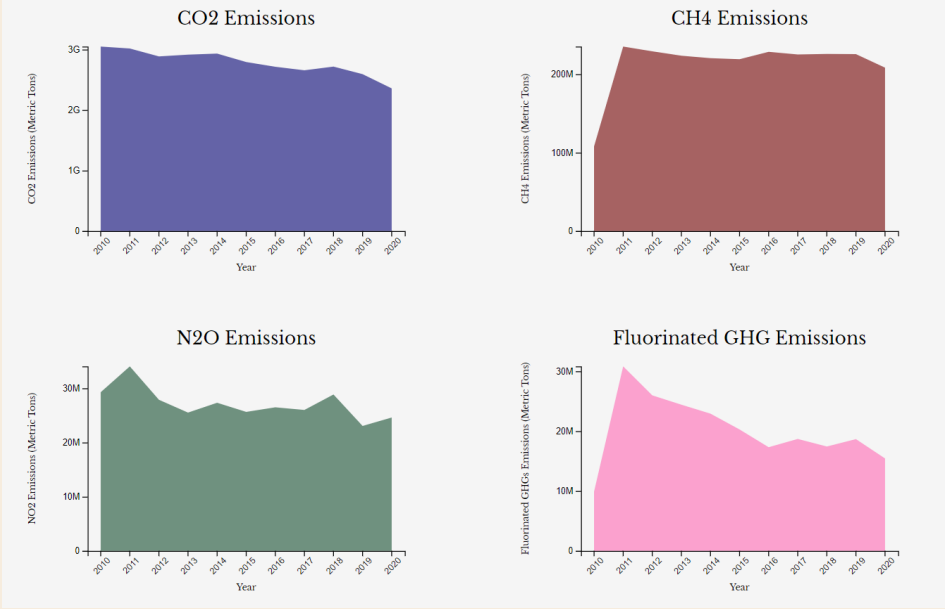
Inspiration from one of Alysha's weekly reflections

Initial implementation with our data



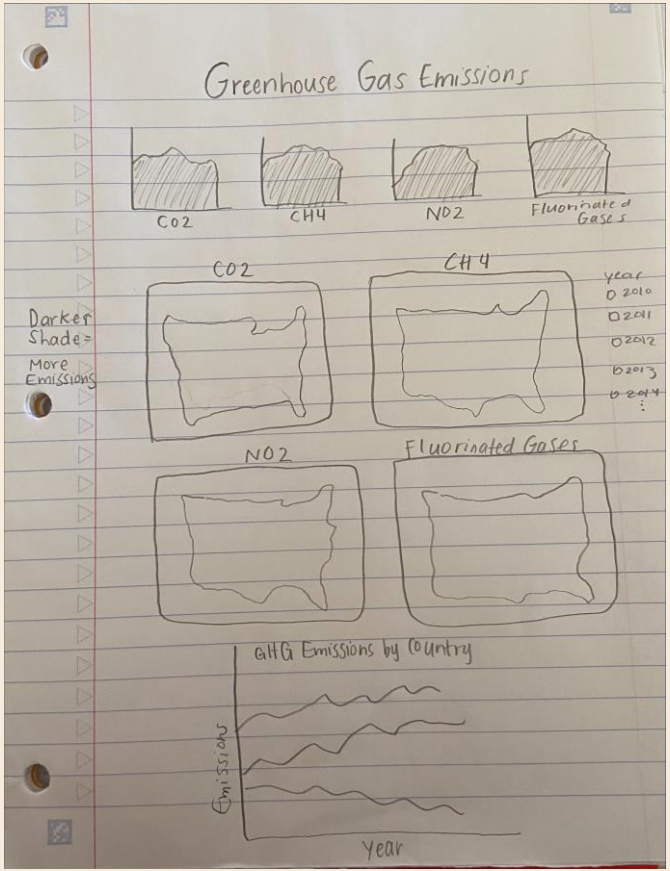
Switching to Area Charts

- Instead of using radial plots for each of the individual greenhouse gases, we decided to switch to area charts. We thought the data as a whole would be better visualized in this form.



Switching to Country Line Chart

- After switching to small multiples of area charts to show the emissions of each gas over time in the United States, we decided that it would be more engaging for the reader if we used a different type of chart to now show changes in country emissions over time. We decided on a line chart based on inspiration from the NY Times.



Tree Map Addition

- After visualizing the emissions of the top 5 contributing countries to GHG emissions compared to all other countries in Python, we wanted to add another map to help the user understand just how much more some of these countries are emitting than other countries. We decided to add a tree map so that the user could see the visualization of the data through rectangles. We believe this was a good addition to help better convey our story.

Other Deviations

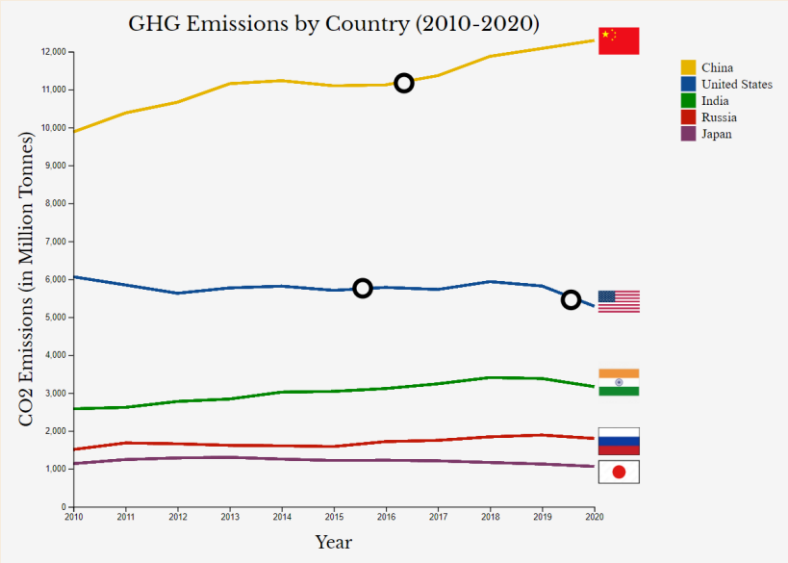
- Changing the order of visualizations to start broad (countries) and then zoom in (United States specifically)
- Adding additonal interactivity to better engage the reader and tell and story
- Adding flags to the country visualizations
- Using all country GHG emission totals instead of just CO2 emissions to better align with the overall story
- Cutting data to only use 2010-2020 instead of 2010-2022 due to missing data in the Our World in Data country dataset

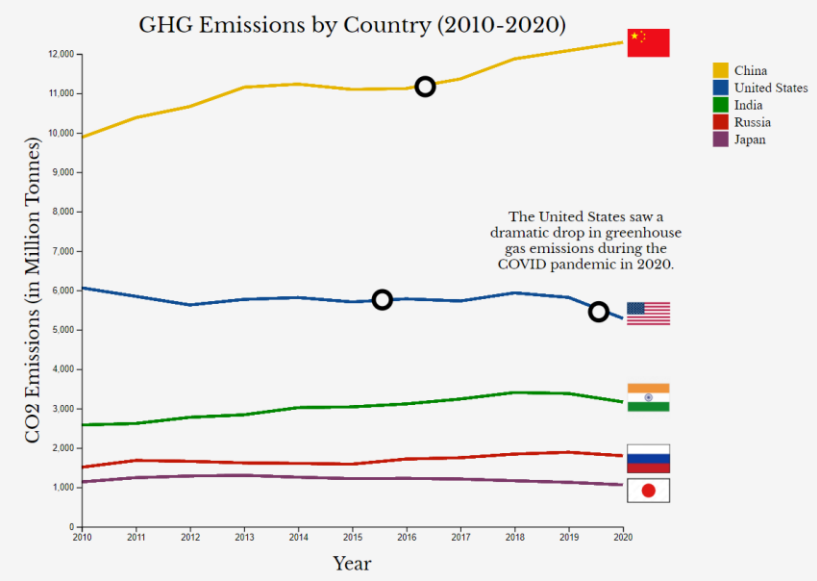
07

Implementation

Visualization 1

Intent: Our first visualization shows the GHG emissions from the top 5 contributing countries. This visual is used to show the user just how much greenhouse gases some countries are emitting

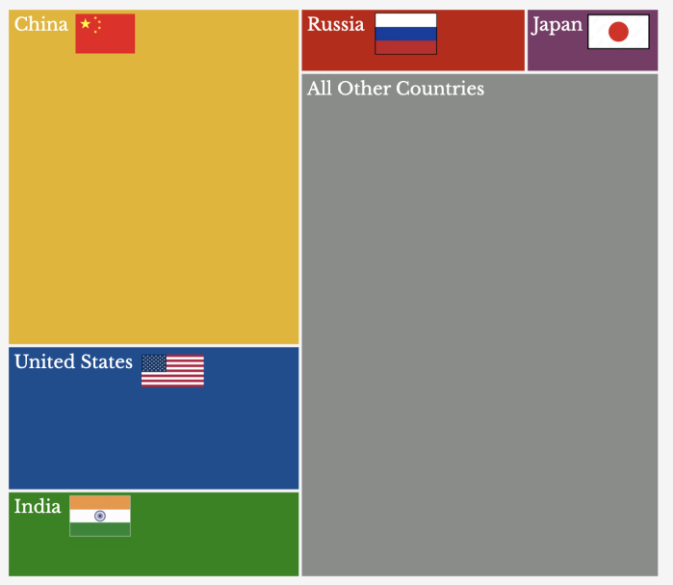


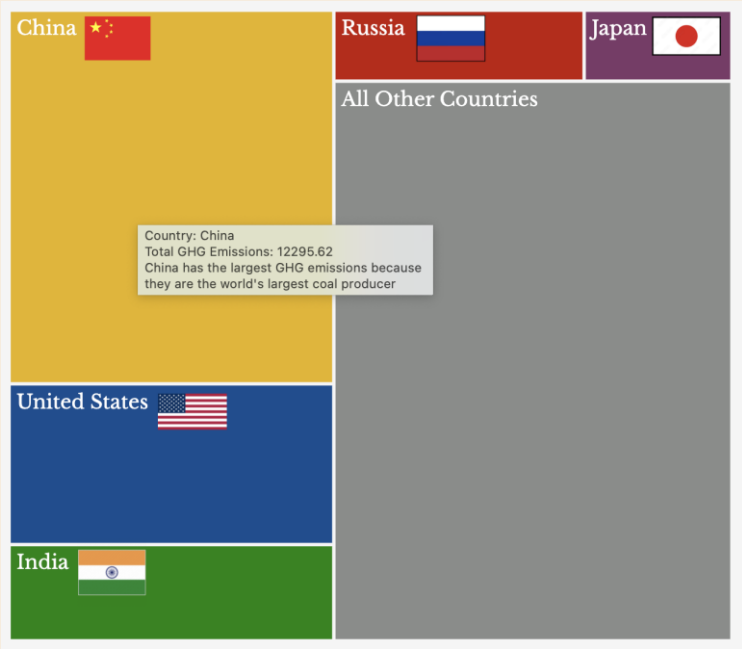


Functionality: On this visualization you can hover over the circles to read more information about that spot on the graph.

Visualization 2

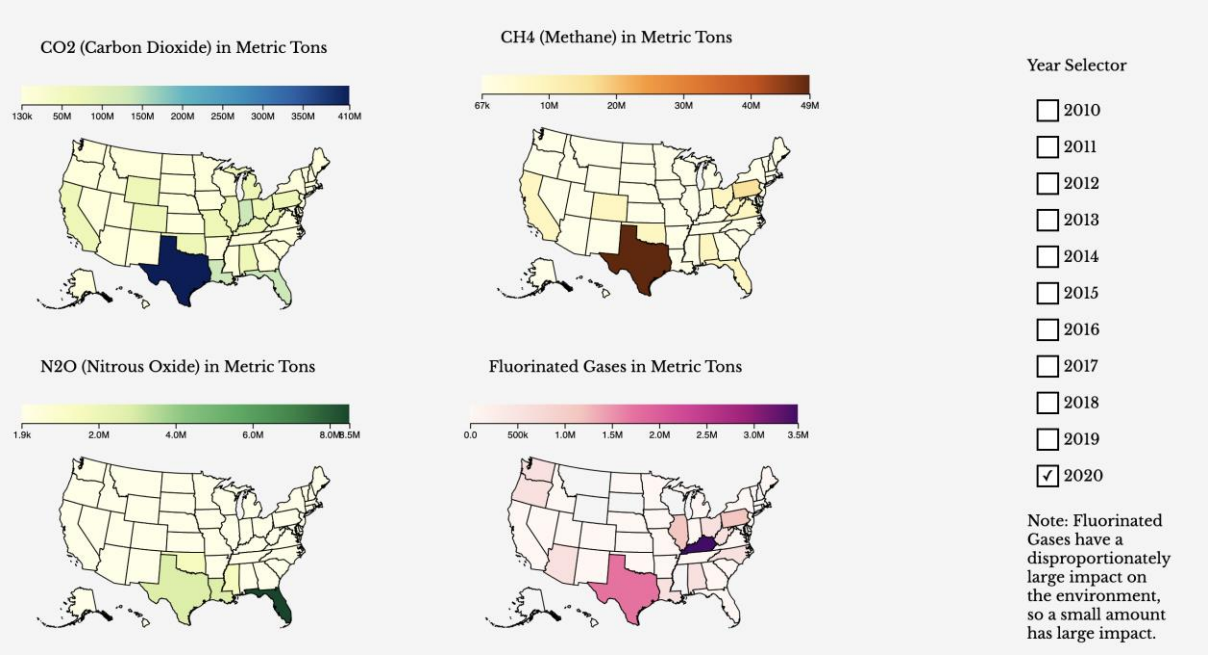
Intent: Our second visualization better shows the comparison between the emissions of the top 5 contributors. This is to help the reader visualize just how much greenhouse gases some countries are emitting.





Functionality: On this visualization you can hover over each of the squares to learn more about that country and their emissions.

Visualization 3

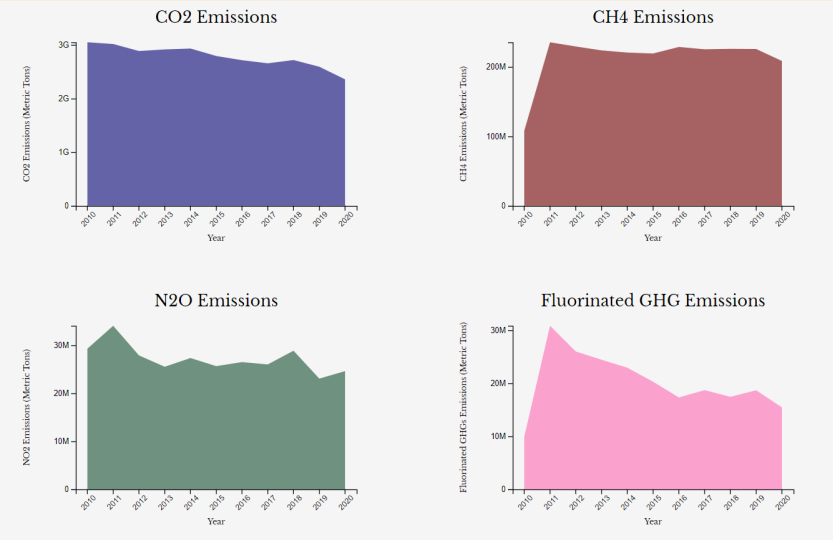


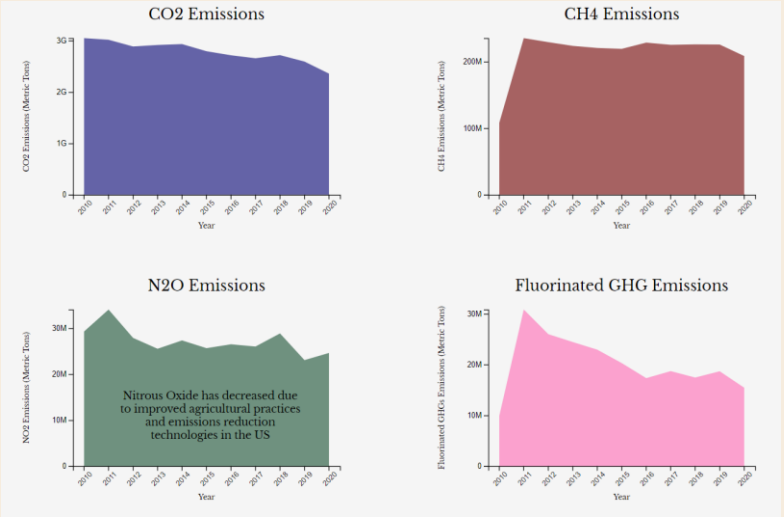
Intent: Our third visualization shows where in the US most greenhouse gases have been emitted over the past decade. This helps the reader get a sense of where US greenhouse gas emissions are coming from.

Functionality: On this visualization you can select one or many years to view their data on the graphs.

Visualization 4

Intent: Our final visualization shows how the emissions of each of the individual greenhouse gases have changed over the past decade. This shows the reader which gases are the most prevalent and how those values are changing.





Functionality: On this visualization you can mouse over each of the graphs to learn more about that specific greenhouse gas.

08

Evaluation

What Did We Learn?

We learned a lot about the data from creating our visualizations. Firstly, we did not realize just how much China contributes to greenhouse gas emissions. We thought it was very shocking to learn that they emit almost as much as every other country does combined. Secondly, we found it shocking to see just how much emissions come from Texas in the US. We never knew that Texas was such a big contributor to GHG emissions. Finally, we learned that greenhouse gas emissions have actually gone down over the past decade. We assumed since there is such a large concern for reducing emissions, that they would still be increasing or staying the same, but did not realize how much has already been done to reduce emissions in the US.

For the most part, our visualizations answer the questions we set out to explore at the beginning of this project. We were able to learn about greenhouse gas emissions over the past decade and learn about emissions from around the world and in the US specifically.

What We Learned About Creating Visualizations

- It is important that the text you include complements your data and does not overwhelm the reader by trying to explain the visualization and its functionality in depth.
- Sometimes simple visualizations are most effective in visualizing data for the public. The New York Times is a perfect example of the effectiveness of clean, straightforward visuals.
- Interactivity helps engage the reader! Allowing the reader to hover over a graph and learn more helps tell the story more naturally and keeps the reader looking at the visualization for longer.
- The aesthetic of the overall web page is important. This includes fonts, length of text, colors of visualizations, and so much more.
- Small multiples can be a better way to show multiple areas of data at once rather than using a filter and relying on the reader to go through each filter to see all the data.

Future Improvements

Our visualizations work well. They accomplish our goal and the questions we sought out to answer. Some future improvements we would like to make to them include: reworking some of the tooltips within the graph spaces to be more obvious to the user. We feel as though they sometimes get lost or go unnoticed. We also would like to add more data to our visualizations by exploring a larger time period. Looking at one decade was useful, but if we were to broaden our scope to 2 or 3 decades we might find much more interesting data and be able to make more interesting visualizations.

Check out our web page!

https://audreymongillo.github.io/final/