Global CO2 and Energy Usage

CS171 Process Book

Anna Zhong and Rachael Smith

Overview and Motivation

As mentioned in our proposal, we want to explore the correlations between human activities and the volume of global CO2 emissions over the years. Our original intention is to span over various indicators, such as population, GDP, energy usage (fossil fuels and alternative), energy production and deforestation (by forest area), in order to seek a comprehensive understanding of the various human factors contributing to carbon emissions that in turn might contribute to global warming. However, we want to clarify that this project is not about the scientific aspect of climate change, although it would definitely be useful to show why people should care about this visualization either through one scientific data set (temperature of sea level) or an introductory page on our final website that places our visualization in context.

Questions

Since we are interested in exploring how humans might have contributed to global warming, questions that we want to explore include:

- At what rate did our population growth, or deforestation, or use of oil and other fossil fuels translate to growing amounts of CO2 in the atmosphere? To visualize the pace and scale with which some of these factors are growing, we will be using colors, circles and a timescale, as detailed later in the process book.
- What does this look like today, and in looking at today's situation, how do we expect the data to look in the near future?
- What are some factors affecting the rate of CO2 emissions in the atmosphere, other than GDP and population? Are there correlations between the amount of renewable energy being used and a slowing of the growth of CO2 emissions? For this question, it would be useful to visualize not only the total volume of CO2 emissions every year, but also the rate of change from the previous.
- How might one country compare to another country (perhaps its neighbor, or major trade partner, or a country in the same income group)? To answer this, we could allow for two countries to be selected and compare their details next to each other.

Hopefully, by trying to answer these questions, we visualize some interesting findings other than the expected upwards trend.

Data (source, scraping, cleanup)

We got our data from the World Bank. Their enormous dataset can be accessed at http://data.worldbank.org/ , and it provides free data about developing countries across the globe. For our project we made use of several sets which together provided GDP, Population, CO2 emissions, Total Energy Use, % Alternative and Nuclear energy, and % Renewable Combustibles for most countries for the years 1960-2014. For latitude and longitude data we used a separate dataset accessible here.

The world bank data sets were available as xml files. Due to our familiarity with using json files with d3, our first step in the datascraping process was to use a free online service to convert the xml files to json. A mistake we made early on in our project was to assume that because our files would be in json format data cleanup would be a rather straightforward, reasonably quick process. It was not. Although data cleanup was somewhat of a tear-wrenching struggle, we *were* fortunate that once we figured out a method to cleanup one dataset it was VERY easy to do the rest, as all of the World Bank datasets were structurally isometric.

Our (semi) raw dataset contained an object for each country, for each year between 1960 and 2014. Each object contains a value for country name, country ID, the specific parameter for the dataset (for instance, CO2 emissions) One of the biggest challenges was finding a way to condense each of the 53 entries for each country, for each year, into a single object for each country. This involved iterating through our *very* raw large dataset several times, and it was challenging to find a method efficient enough that it did not 'brick' our browser.

In our data cleanup we condense and combine several different World Bank Datasets into two datasets for visualization: one dataset for our "worldVis" view contains an object for each country which contains an array of values representing GDP values for each year 1960-2014, and similar arrays for population and CO2 emissions. Our second dataset is structurally similar, but contains arrays with values for Total Energy Usage, %Nuclear and Alternative Energy, and % Energy from Renewable Combustibles. The latter set is used for our "countryVis" view.

The second major setback in data clean up was a horrific bug-- we had *all* of the population, gdp, and co2 arrays in our CO2 dataset pointing to the *same* 'physical' array, without realizing it. It took us over an hour to diagnose this bug, and it left us feeling dejected and as if we deserved to re-take cs50. However, we have persevered, and we have come out of the Great Data Bug Fiasco of 2015, stronger and better. We (and our beautifully cleaned data) are now prepared to tackle whatever the CS171 final project throws at us with a renewed vigour and refreshed fighting spirit. (We <3 Data!)

Project Objectives (as also written in the Project Proposal)
Ideally, our visualization will consist of three "chapters". We consider it essential that we implement the first "chapter," we are hopeful that we will implement the second "chapter" and it would be outstanding if we could accomplish the third.

The first chapter tells the story of the history of global carbon emissions. Our hope is to educate users on how emissions have changed over time, and how these changes have correlated with economic growth, population growth and deforestation. In this chapter, we are mainly interested in seeing exponential growth in all of these factors within the past twenty – thirty years, although

we are aware that this doesn't necessarily prove any causation between the factors.

The second chapter tells the story of global carbon emissions today. Here, we hope to answer questions about how the income of each country, as well as its geographical region, affects its current emissions. We intend to look at the relative amounts of total emissions per country, as well as per capita emissions. Users will be able to visualize both the geographic distribution of emissions today, as well as the economic distribution of emissions.

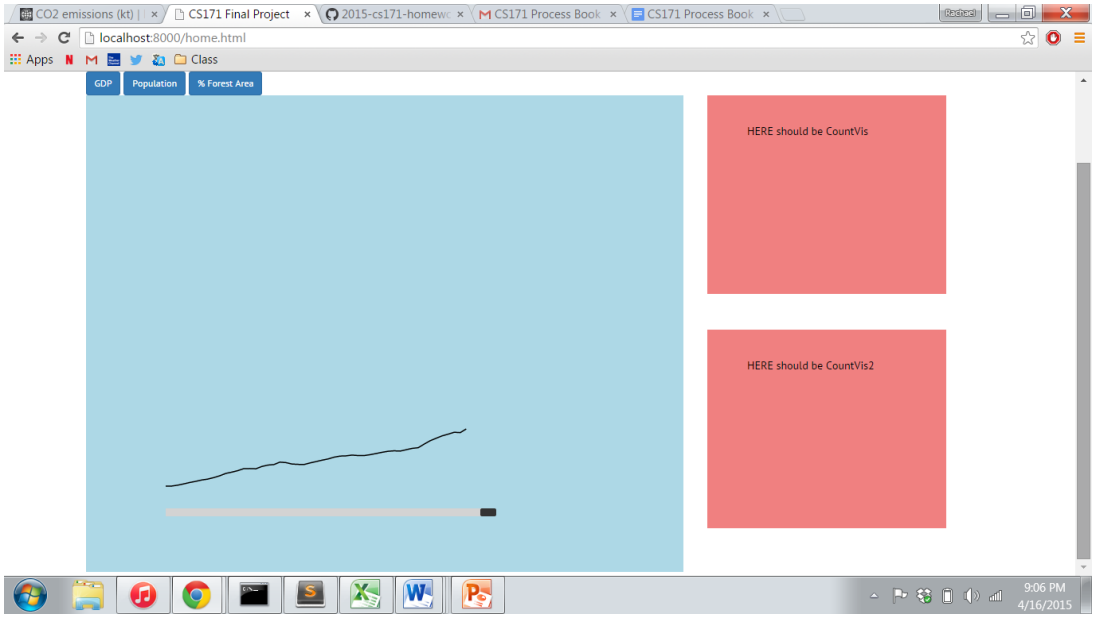
The third chapter will tell story of carbon emissions in the USA today. For instance, we think it would be interesting to visualize the breakdown of emissions by industry. More research will be necessary to determine the precise nature of this chapter, as this will depend on what (if any) relevant data is available.

The ideology is that each chapter will progress to an increasingly more specific set of data to visualize. Each previous chapter serves to contextualize the restricted focus of the subsequent chapters.

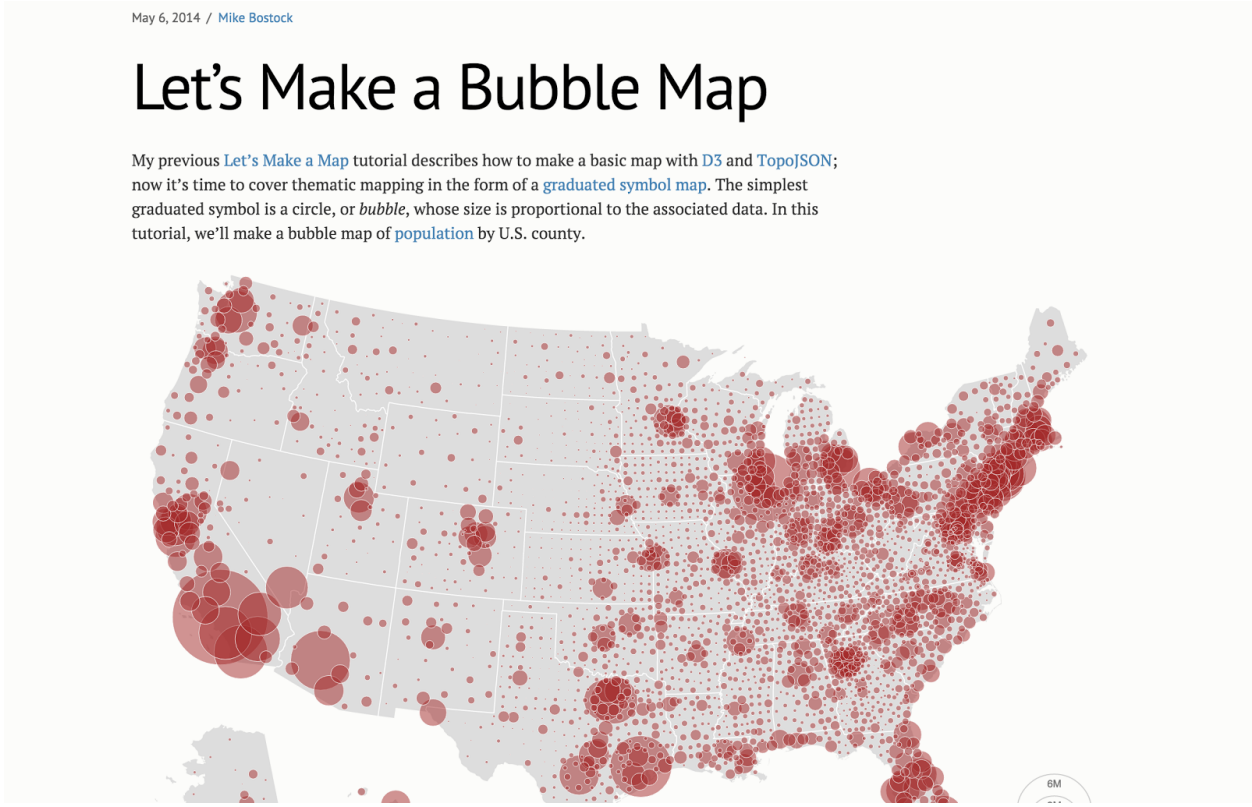
Exploratory Data Analysis (initial vs final visualizations)
We drew some inspiration from our problem sets, and we intend to link different visualizations together to make them interactive. We are still currently in the early stages of our implementation, so we are still learning from our initial (current) visualizations. As per our project proposal, we will aim for three "chapters", although this might change depending on adjustments we make during our implementation. Problems and improvements will be detailed in the design evolution and evaluation sections below.

Design Evolution and Implementation
The step by step process of our implementation is as follows.

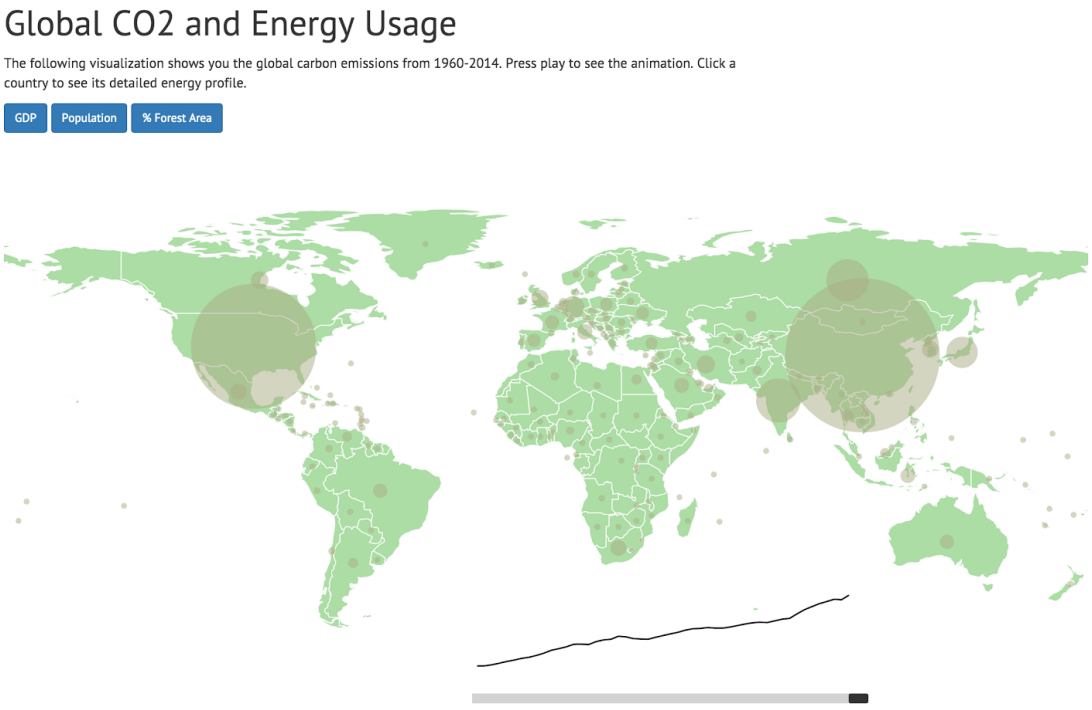
1. Structuring the layout of our first "chapter". As per our current plan, this chapter is the most complex of our visualizations, and here we have borrowed from Homework 3.



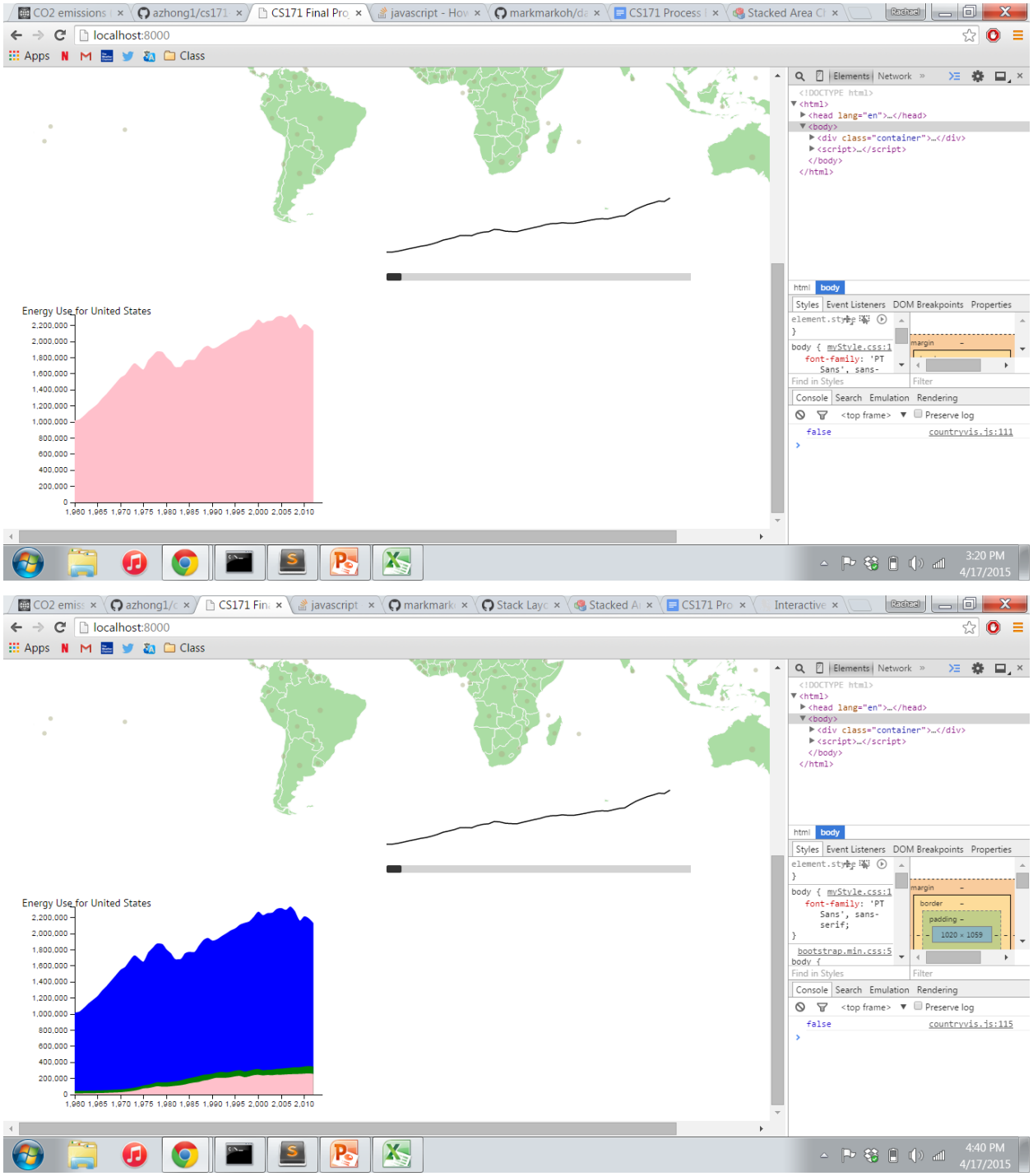
2. Now to create the main map view as well as our "bubbles" of CO2! We took inspiration from Mike Bostock's work on "bubble maps", as seen in this image. We thought that the bubbles were especially fitting for our visualization because we the circular symbol can be seen as analogous to a cloud of CO2 gas rising from a country. These bubbles will also change in size with time.



3. Datamaps. We are using a Javascript library built with d3 called datamaps. It allows us to generate chloropleth maps and bubbles, which is exactly what we need. Having said that, it took a lot of manipulation of our numerous data sets (as detailed in the 'Data' section above) to get it working correctly. The image here shows our visualization after we have our bubbles working, with the slider currently at 2012. Already, we have a sense that the visualization might encounter problems with a couple of countries dominating the data set, namely the United States and China, as we saw in the node visualizations of our second homework. For now, the bubbles look ok, but we think they would have to be complemented with tooltips on hover in order to extract details about the smaller countries. The tooltips will be added after the first milestone.

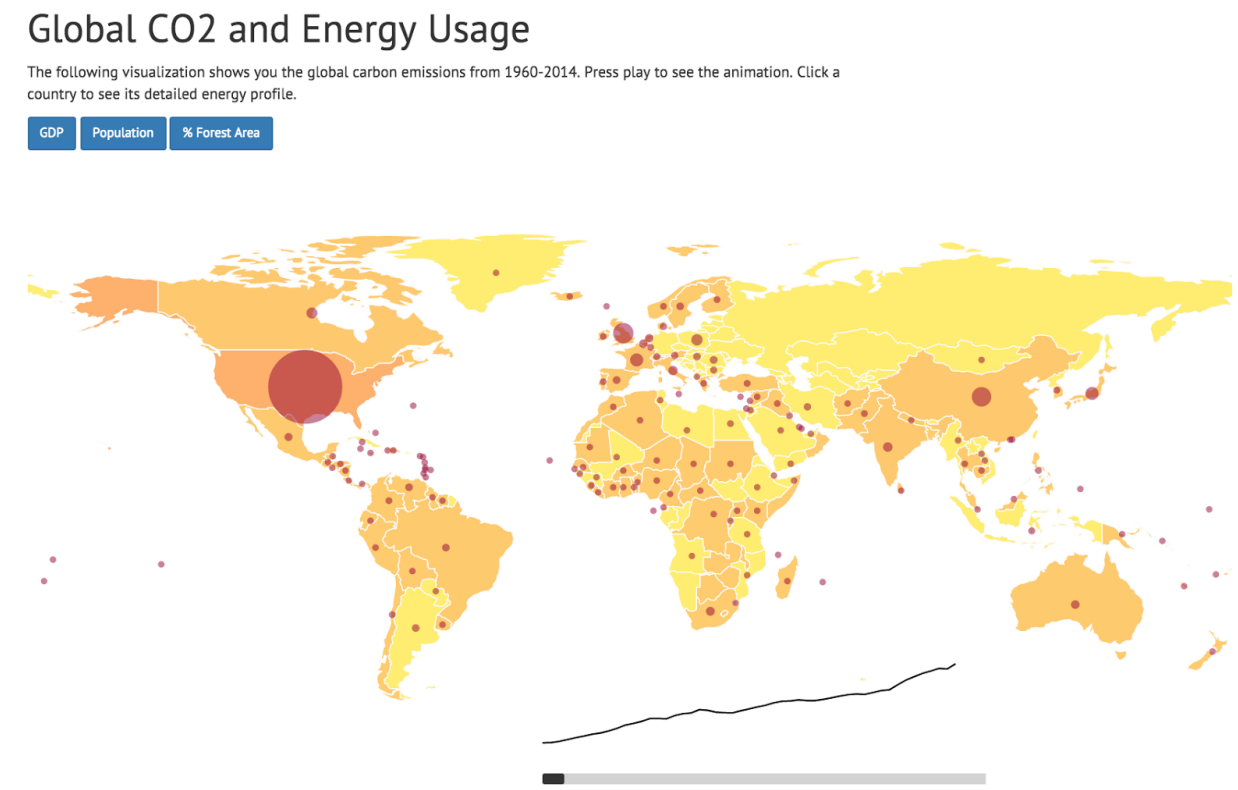


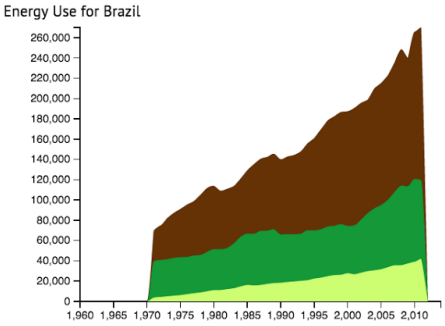
4. The interactive area chart. The following two images show a separate area graph when any country is clicked. In the second image, we have included all three of our data sets as planned. The blue area represents total energy usage, the pink area represents the % of total energy that is alternative or nuclear, and the green area shows the % of total energy in other combustible renewables. As we were creating this part, we realized improvements that might be made after the first milestone. These include having the area chart as a bar chart with individual bars for each year. They will be grouped closely, but more importantly, this will allow the user to hover over a certain bar and see that exact value of energy used that year on tooltip.



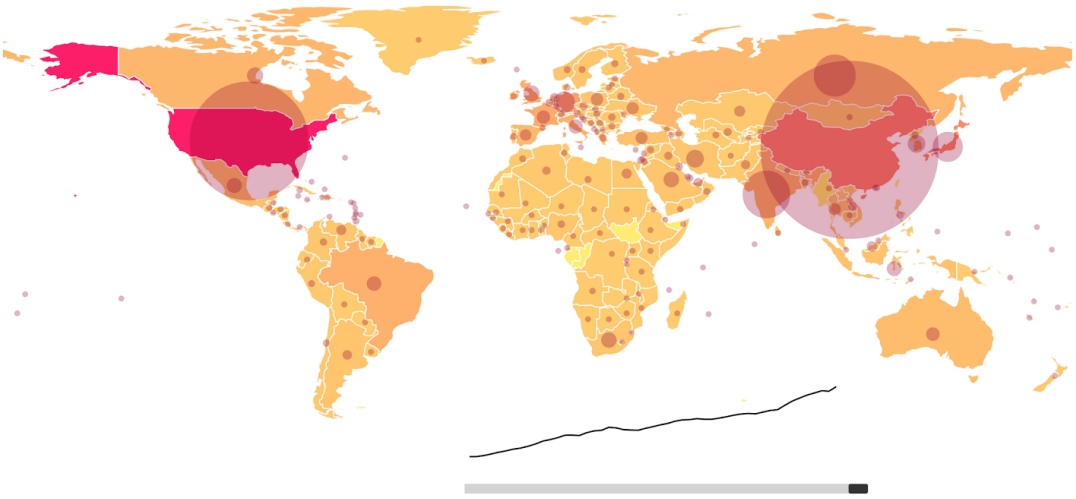
5. Now for the fun coloring part! The next two images show the “heat map” for GDP, for which we are using a yellow-orange-coral color scheme for now. The last two images show the heat map for population, for which we are using a blue-purple color scheme. For the population heat map, this color scheme does not appear to be expressive enough, so we will be re-evaluating the gradient for after the milestone. A particularly noticeable finding is the rate of China’s growth in

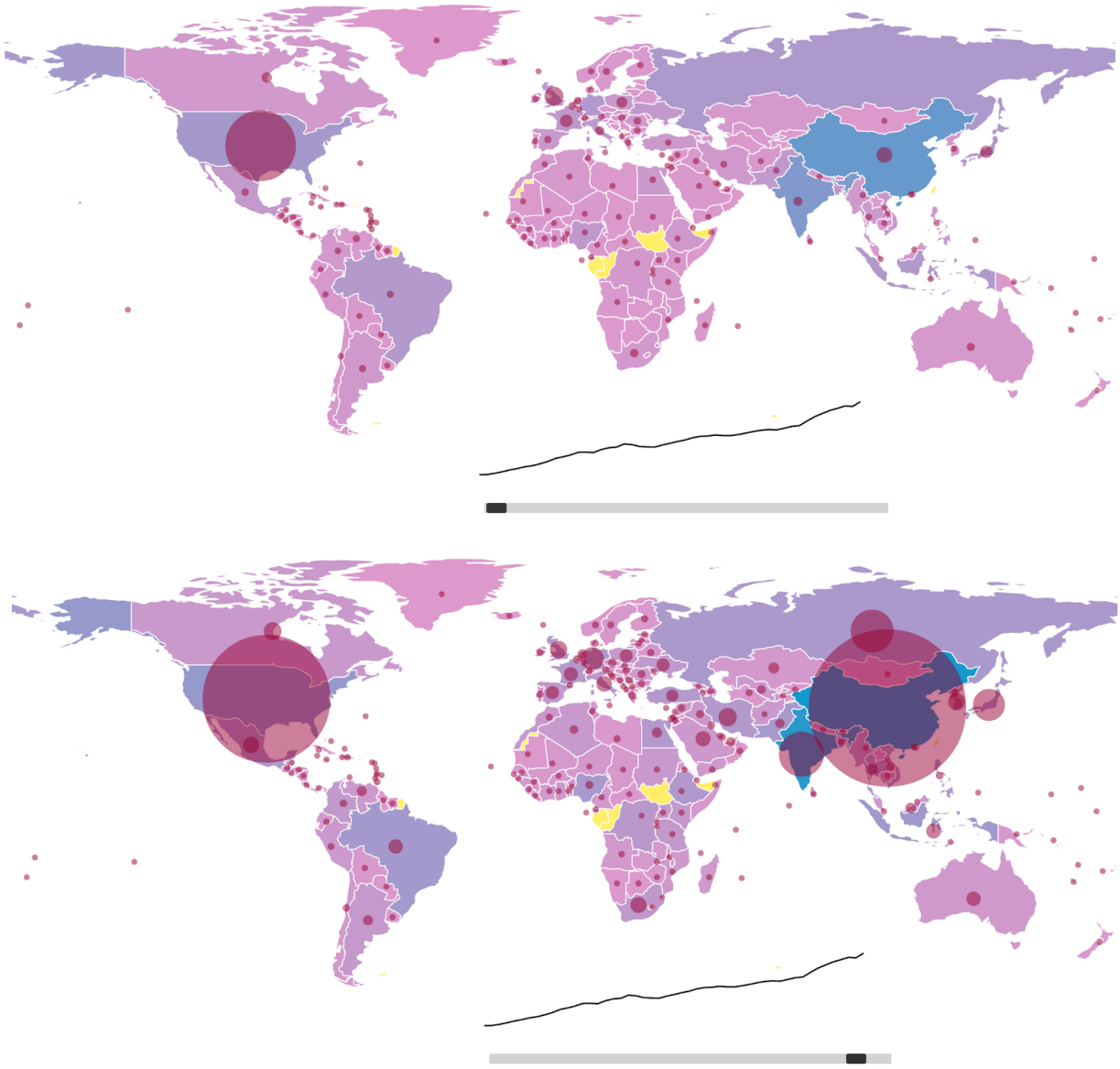
terms of GDP, population and CO2 emissions in the last 10 years. This would be a good thing to explore using a rate of change graph that we can implement after the milestone.





GDP | Population | % Forest Area





Evaluations (for milestone one)
Setting the foundations for project took a lot longer than expected, but we think we're in good shape to finish it off. We hope to focus on the more nuanced questions listed at the beginning of this process book after our milestone.