Jaiden Atterbury
jatter41@uw.edu
Section AA
02/21/2024

A5: Ethical and Deceptive Visualization in Observable + Vega-Lite

Link to assignment specification:

https://docs.google.com/document/d/140T_au5LSdlFIVJ0iNcTLf1YyXQFf8XJ_LdmVRNdzRI/edit

Overview of Process:

To start off this assignment, my first course of action was to find a data set that would allow me multiple avenues to create deceptive visualizations. Due to my interests in epidemiology, my initial plan was to find a health related data set on topics such as life expectancy, mortality rates, disease prevalence, etc. However, I wasn't able to find any data sets that I felt would be nice to work with (except for one that I used for additional visualizations, see "Bonus Visualizations" section for more details). With limited inspiration in mind, I turned my focus to finding a topic that has modern societal relevance, and in this search I found a topic that has high stakes as well as modern relevance; that topic being climate change and CO2 emissions.

In INFO 201, I got first-hand experience working with a climate data set, so I thought this assignment could be a "full circle moment" and allow me to creatively extend on something that I've had experience working with. Although I've had experience working with CO2 data before, the data set I ended up choosing was one I've never worked with in any other assignment or project. The data set that I chose comes from a popular and reputable source/website called Our World In Data, which is a "scientific online publication that focuses on large global problems." In particular, the data set comes from this article, which sought to answer questions such as: "How much CO2 does the world emit?" and "Which countries emit the most CO2?" The data set is updated regularly and includes data on CO2 emissions (annual, per capita, cumulative and consumption-based), other greenhouse gasses, energy mix, and other relevant metrics. The data set itself can be found and downloaded on Github through the following link.

As explained above, the data set contains many attributes/columns (80 to be exact). A codebook explaining the meaning of each variable can be found here. Specifically, for this assignment, my main focus was on the variable co2, which is explained in the code book as representing, "Annual CO2 emissions - Annual total emissions of carbon dioxide (CO2), excluding land-use change, measured in million tonnes." The reason

why my focus was on this specific variable, was because my analysis and visualizations were not focused on breaking down the different types of emission and their causes, instead their focus was on total emissions as a whole. Namely, the question that my visualizations intended to answer was: "**For the top ten countries leading the world's decarbonization mission, what has their trend been in total CO2 emissions since the beginning of the 21st century?**" Although the visualizations focal point is the top ten countries leading the world's decarbonization mission, as explained in this Climate Trade article, the United States was also included as an option to visualize, as I felt that readers would also be interested in how their carbon footprint has changed in the 21st century, even though they aren't listed in the aforementioned article. As I knew that these countries would have decreasing, constant, or mildly increasing carbon footprints over time, one way that I could deceive the audience into believing otherwise is through the use of bar charts that plots a country's CO2 emissions over time, but with a twist. One visualization would plot a country's total CO2 emissions in a current year for all of the years in the 21st century that are represented in this data set, and the other would do the same thing, **but,** instead of plotting the raw total CO2 emissions for a country in a given year, it would instead plot the *cumulative* CO2 emissions since the start of the 21st century up until the given year. This deceptive visualization would thus give the illusion that the carbon footprint of these countries is getting drastically worse. In both visualizations, the bars are colored by population, which allows the users to see a "per capita" value without actually encoding that value.

Ethical Description (Labeled as Visualization 1):

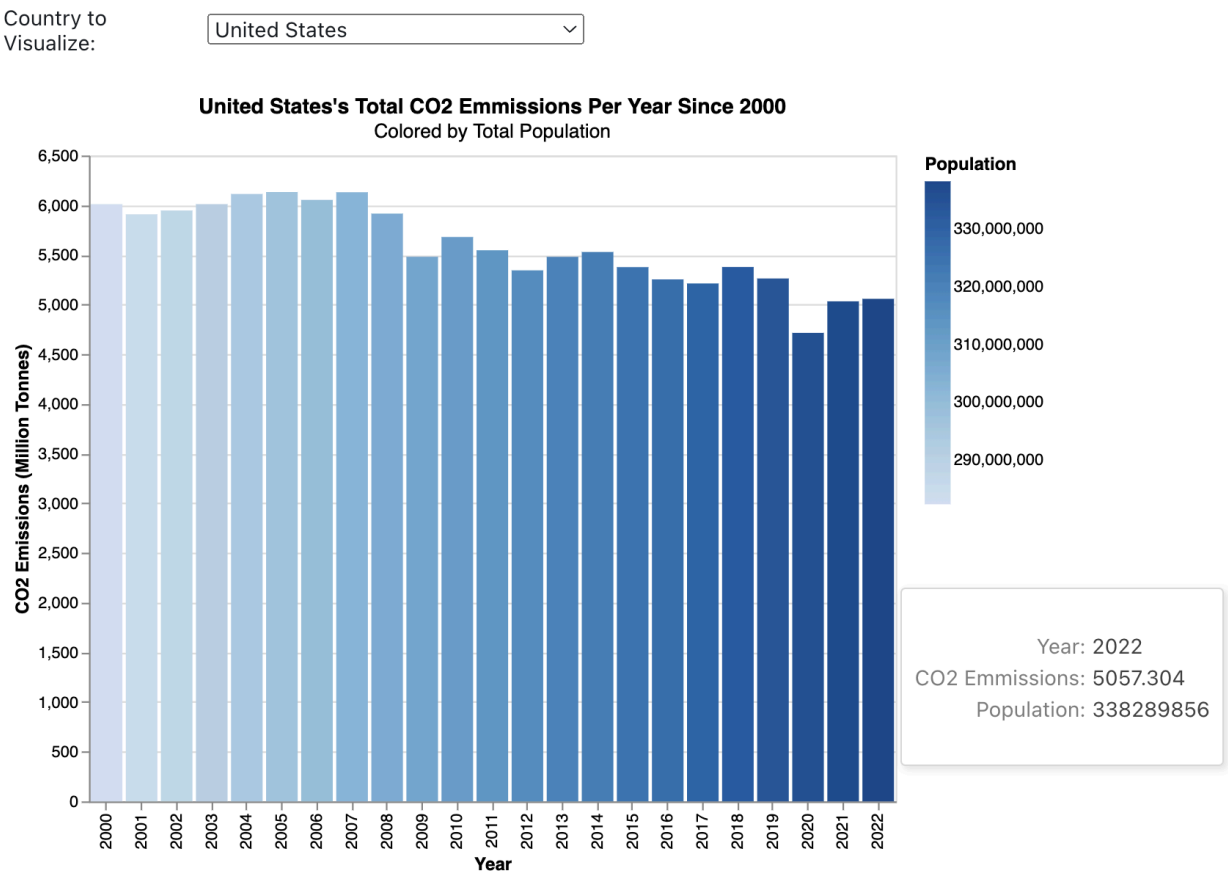
Now that I have explained my general process and rationale for the assignment, I will display and explain my ethical visualization, which can be found through this link. A screenshot of this visualization is provided after this description.

For the ethical visualization of the CO2 emissions data, I decided to create a bar chart that, for a user selected country (which is changed through the drop down menu above the visualization), plots the total amount of CO2 emissions from the given country across all of the years in the 21st century (which was 2000-2022) colored by population. This bar chart was used in order to visualize the trend in CO2 emission while also allowing for the population to be encoded by color, which a line graph wouldn't be able to do. Each bar represents a different year, its height corresponds to the given country's total CO2 emissions for that year, and its shade of blue corresponds to the given country's population in that given year. Users can get a breakdown of the aforementioned information stored in each bar by hovering over a bar of their choice.

This ethical visualization allows the user to visualize the trend of CO2 emissions over the entirety of the 21st century, for one of eleven possible countries, ten of which are

determined to minimize their carbon footprint. As can be seen from this visualization, for the United States, since the year 2000, despite continual growth in its population, the total amount of CO2 emissions has steadily decreased, with only a few years of increase. This trend is similar among all of the countries that the user can select, with only a few deviating from this pattern. Thus, to answer the proposed question, the user would conclude that, for the countries who are committed to reducing their carbon footprint, most of these countries have seen either a decrease in their overall carbon emissions, or minimal gains in their overall carbon emissions since the start of the 21st century. This conclusion is what the user would hopefully expect to occur given the rise of climate crisis awareness in the 21st century.

As mentioned in the opening paragraph of this section, here is the corresponding visualization, as well as the link to the Observable notebook it is contained in (in case it was overlooked).



Link to Observable notebook: https://observablehq.com/d/c675a1cee09a5c04

Deceptive Description (Labeled as Visualization 2):

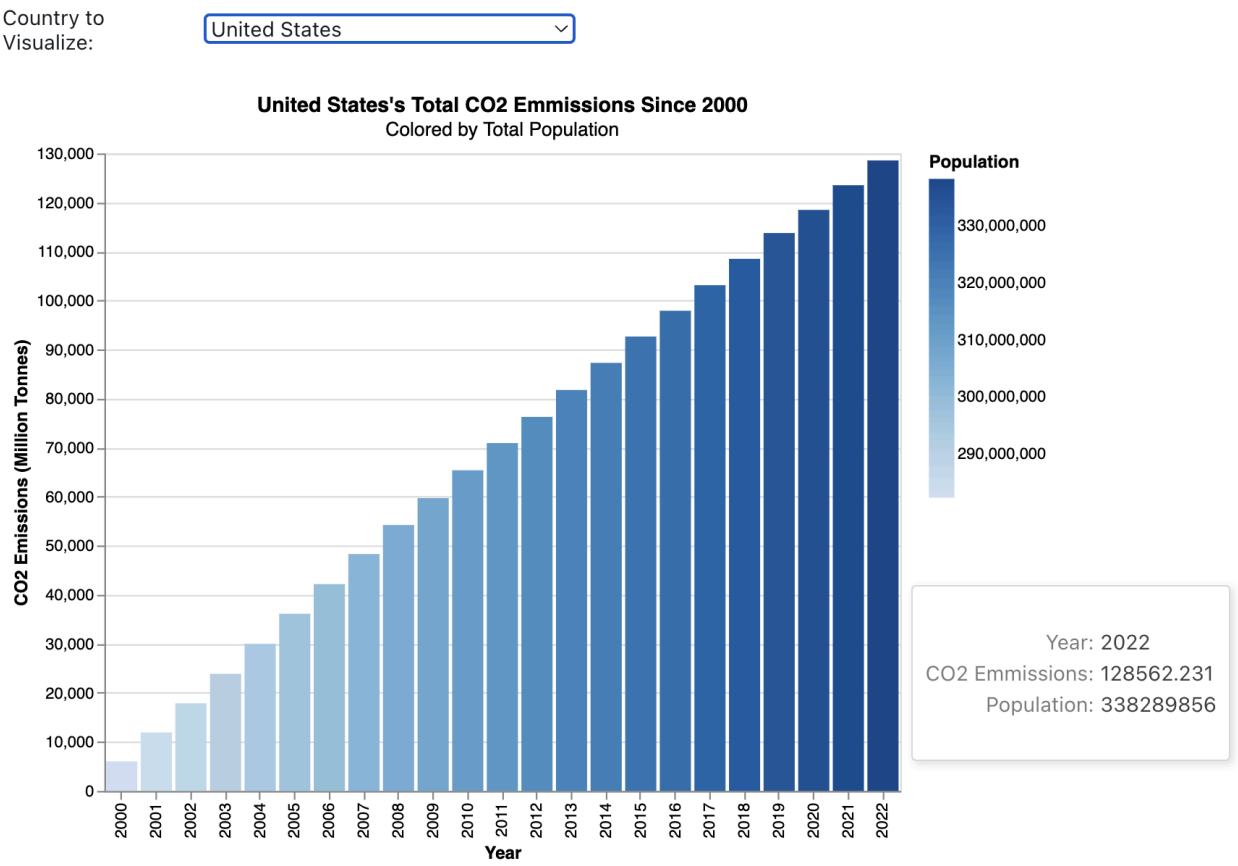
Now that I have explained what was occurring in the ethical visualization, I will now display and explain my deceptive visualization, which can be found through this link. A screenshot of this visualization is provided after this description.

For the deceptive visualization of the CO2 emissions data, I decided to create another bar chart that, for a user selected country (which is changed through the drop down menu above the visualization), plots the cumulative amount of CO2 emissions from the given country across all of the years in the 21st century (which was 2000-2022) colored by population. In particular, each bar represents a different year, its height corresponds to the given country's total *cumulative* CO2 emissions up to and through that year, and its shade of blue corresponds to the given country's population in the given year. Users can get a breakdown of the aforementioned information stored in each bar by hovering over a bar of choice.

This bar chart differs from the ethical visualization in the fact that it plots cumulative amounts instead of by year amounts. This bar chart was used in order to deceive the viewers by masking an aggregate measure as a "trend," thus leading viewers to inaccurately depict how the CO2 emissions of a selected country have changed over time. This deception is masked in the fact that the y-axis label says "CO2 Emissions (Million Tonnes)" without mentioning that this CO2 emissions value is actually cumulative.

This deceptive visualization prevents the user from being able to visualize the trend of CO2 emissions over the entirety of the 21st century, for one of eleven possible countries. Instead, this visualization pushes the user into a false narrative due to the way the axes were labeled. As can be seen from this visualization, users will notice that, for the United States, since the year 2000, along with its constant population growth, the total amount of CO2 emissions has drastically increased in each year. This trend is exactly the same for every other country that the user can visualize. Thus, to answer the proposed question, the user would conclude that, for the countries who are committed to reducing their carbon footprint, since the start of the 21st century, all of these countries have seen drastic growths in their CO2 emissions. This conclusion is not good as it leads users to believe that these countries are lying about their intentions, or current mitigation techniques are not working when they actually are (on average).

As mentioned in the opening paragraph of this section, here is the corresponding visualization, as well as the link to the Observable notebook it is contained in (in case it was overlooked).



Link to Observable notebook: https://observablehq.com/d/3060cd6f7aa7db04

Learning Reflection:

Before this assignment, as was shown in some of the visualizations I made in the section activities, I struggled with coming up with ideas for deceptive visualizations. This was mainly due to the fact that I had never purposely tried to deceive my audience when making visualizations. Thus, when I was forced to make something deceptive I couldn't shift gears and stray away from my normal mindset. However, once I was able to work with a data set that was meaningful to me, I was able to shift my focus and create something that was "correct" but would lead the readers into a wrong conclusion. Therefore, in terms of deceptive visualizations, this assignment allowed me to learn how to formulate a question and make visualizations that answer that question in both the correct, and incorrect fashion. Through the creation of these visualizations, I have come to the conclusion that the most effective way to make a visualization deceptive, without being entirely incorrect, is to omit important details from the axis labels. For example, my deceptive visualization wouldn't lead readers to draw the wrong conclusions if the y-axis label included all of the relevant information. Through the use of this assignment

and the accompanying slides, I now know how to identify and prevent deceptive visualizations from being deployed in settings where a large audience will be reached. Lastly, on a more technical note, this assignment allowed me to refine my Vega-Lite skills, as in order to create the cumulative bar chart, I had to learn how to use the vl.window() function to create an “aggregate window.”

Bonus Visualizations:

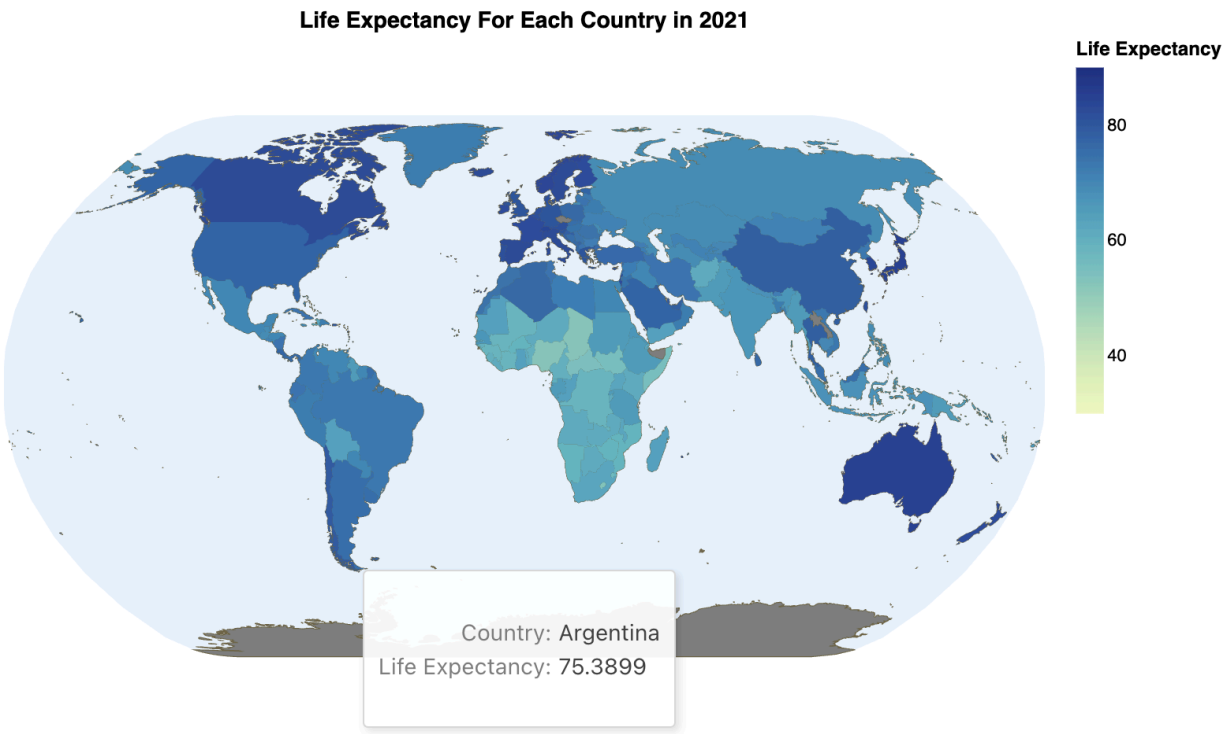
As was mentioned in the “Overview of Process” section, I made additional visualizations on the topic of life expectancy. I am including these here since I am proud of their outcome and don’t want them to be created for no reason.

The following visualizations also come from a dataset from Our World In Data. In particular, this dataset focuses on the “period life expectancy” of countries around the world from 1950-2021. The authors define period life expectancy as a metric that summarizes death rates across all age groups in one particular year. Specifically, they explain that, for a given year, period life expectancy represents “the remaining average lifespan for a hypothetical group of people, if they experienced the same age-specific death rates throughout the rest of their lives as the age-specific death rates seen in that particular year.” This data set can be found by following this link. Modifications/cleaning of this data set were made in the statistical programming language called R, in the IDE RStudio. The question these visualizations intended to answer was: “**Currently, around the globe, which countries have the highest life expectancies? Which countries have the lowest**?”

Ethical Visualization:

The ethical visualization is a choropleth map that displays the countries of the world, with the fill color determined by each country’s life expectancy value in 2021. Users will notice that most countries have similar life expectancy values, but countries in more developed continents, such as Europe, have higher life expectancies than countries in less developed continents, like Africa.

Below is the ethical visualization and a link to its corresponding Observable notebook.

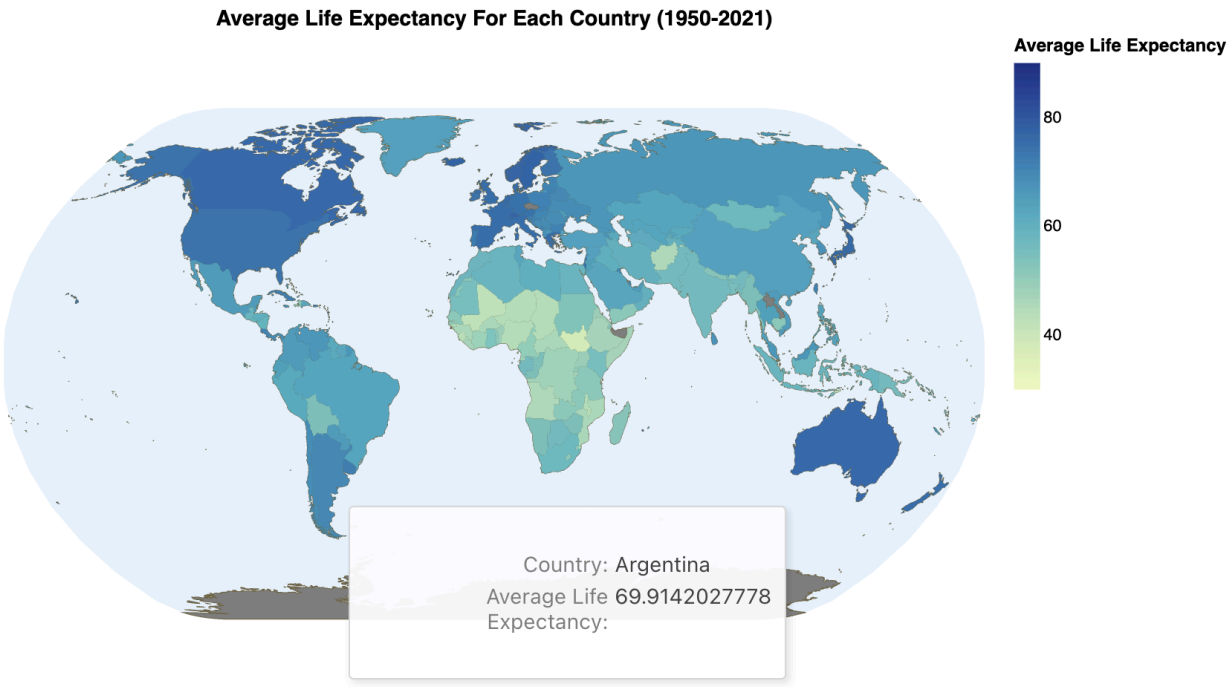


Link to Observable notebook: https://observablehq.com/d/43de5f50ecff8835

Deceptive Visualization:

The deceptive visualization is also a choropleth map that displays the countries of the world, with the fill color determined by each country's **average** life expectancy value across the years 1950-2021. This visualization is deceptive because it seeks to answer the question of how life expectancy currently is, with an aggregate measure that takes into account data from nearly a century ago. Again, users will notice that countries in more developed continents, such as Europe, have higher life expectancies than countries in less developed continents, like Africa. However, users will perceive these differences to be much bigger than they truly are in current times.

Below is the deceptive visualization and a link to its corresponding Observable notebook.



Link to Observable notebook: https://observablehq.com/d/aea13bacbefa3dbf