

Fundamentals of Data Visualization

Final Project write up

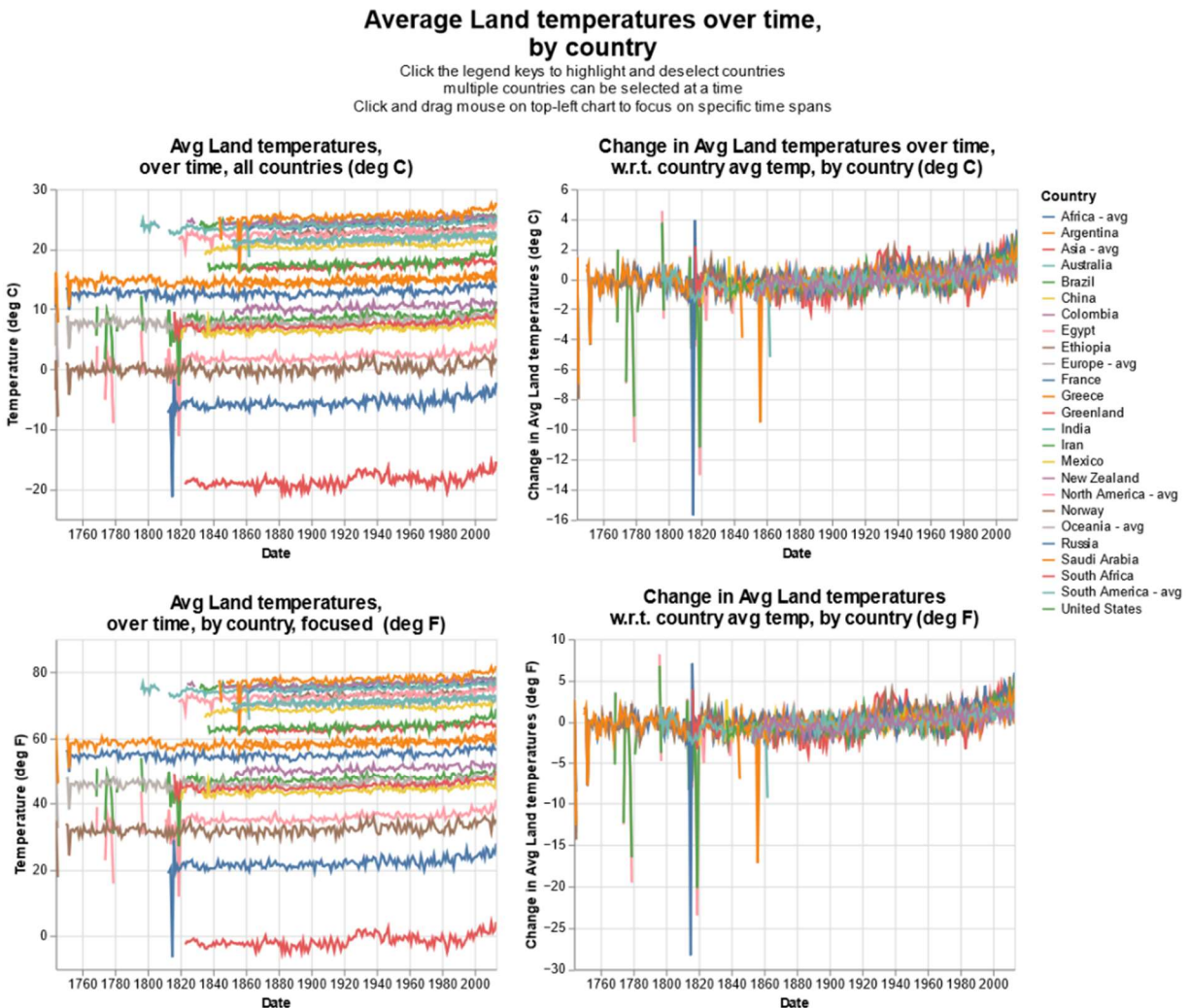
My project is about creating an interactive visualization for data on changes in average land temperature, at different locations on planet earth, in a format that allows users to focus in on different specific locations and time spans while also allowing for comparisons between different locations.

If you wish to engage with the visualization yourself, you can find it here:

https://github.com/KyleRitland/UC-Boulder-Data-Science-/blob/main/vis_eath_temps.html

To have to graphic work, you will need to download the file to your computer, then open the .html file on your browser.

The next picture shows the visualization in its entirety.



I acquired the data for this project from a Kaggle project titled “Climate Change: Earth Surface Temperature Data”, which got its data from the non-profit organization Berkley Earth, who’s founding mission in 2013 was to independently confirm the climate change findings of U.S. federal agencies like NASA and NOAA (National Oceanic and Atmospheric Administration).

The main goal I wanted this visualization to accomplish was to convey that average temperatures have increased by about 2 degrees Celsius or 5 degrees Fahrenheit. This is around the globe, at every location in the data set presented in the visualization.

To accomplish the above goals, I focused on three tasks:

1. Allow users to select one or more countries from a legend, then on one chart, have the visualization highlight the selected countries temperature data and grey out all non-selected data, and on a second chart, show only the selected data and remove all non-selected data while focusing in on the presented data. This would happen simultaneously, where the first chart is in Celsius and the other in Fahrenheit.
2. When users select one or more countries from the legend, the visualization would highlight the selected countries adjusted temperature data and remove all non-selected data while focusing in on the presented data. This would happen on two additional charts simultaneously (separate from the ones mentioned above), where one chart is in Celsius and the other in Fahrenheit.
3. Allow users to control the time span of data presented, so that they can focus in on a smaller set of the data. This would happen on all four charts simultaneously.

I originally only planned on completing tasks two and three, but when multiple countries data were presented on one chart, I decided it would be best to add in task #1, to allow users to see all the data at all times, while being able to focus on the selected data in additional graphs.

At first, I wanted to have users explore only the global average temperature change over time, but after some exploration of that data set, I determined that it was not engaging enough and did not provide a comprehensive enough understanding of how different locations across the planet were experiencing temperate increases. Therefore, I decided to show users the average temperature changes for many different locations across the planet. The original dataset I downloaded contained a spreadsheet of that data by country, so I worked with that. This dataset also had the average temperature changes for every continent except Antarctica, so I included that information as well. Due to the large number of countries listed in the dataset, I decided to include only a small sample of the available countries, with the sample containing enough countries from each continent to give an adequate representation of each continent in the data set.

Key elements:

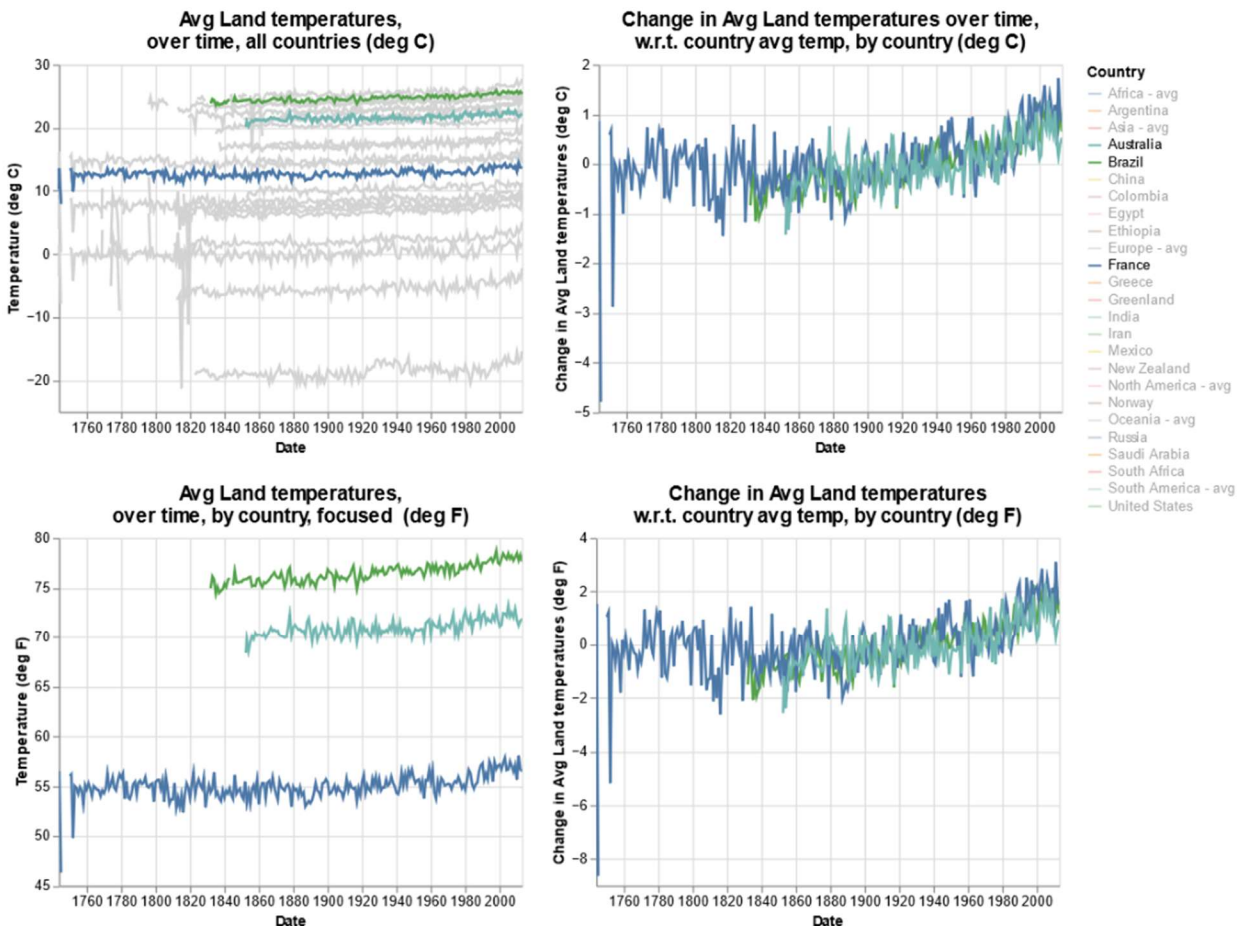
I believe the visualization can be broken down into these key elements:

1. There are four charts, where one always shows all the data available for viewing, and the other three are specialized.
 - a. Chart 1 (top left): Show all countries data, but have unselected data to be greyed out.
 - b. Chart 2 (top right): Show only selected countries data on adjusted average temperatures in degrees Celsius and allow for date range selector to adjust presented data.

- c. Chart 3 (bottom left): Show only selected countries data on average temperatures in degrees Fahrenheit, and allow for date range selector to adjust presented data.
- d. Chart 4 (bottom right): Show only selected countries data on adjusted average temperatures in degrees Fahrenheit, and allow for date range selector to adjust presented data.

I thought it was important to present the data in these ways because this removes the need to keep in mind previously view data while looking at a new chart. It presents the data in multiple unique ways and allows the user to quickly glance from one chart to the other, allowing them to spend more time on observation and analysis and less time jumping back and forth between different plots. In addition, with one of the plots always showing the total set of data, the user can quickly glance between this main chart and all the other more focused charts, so that they can give context to the subset of data they are viewing.

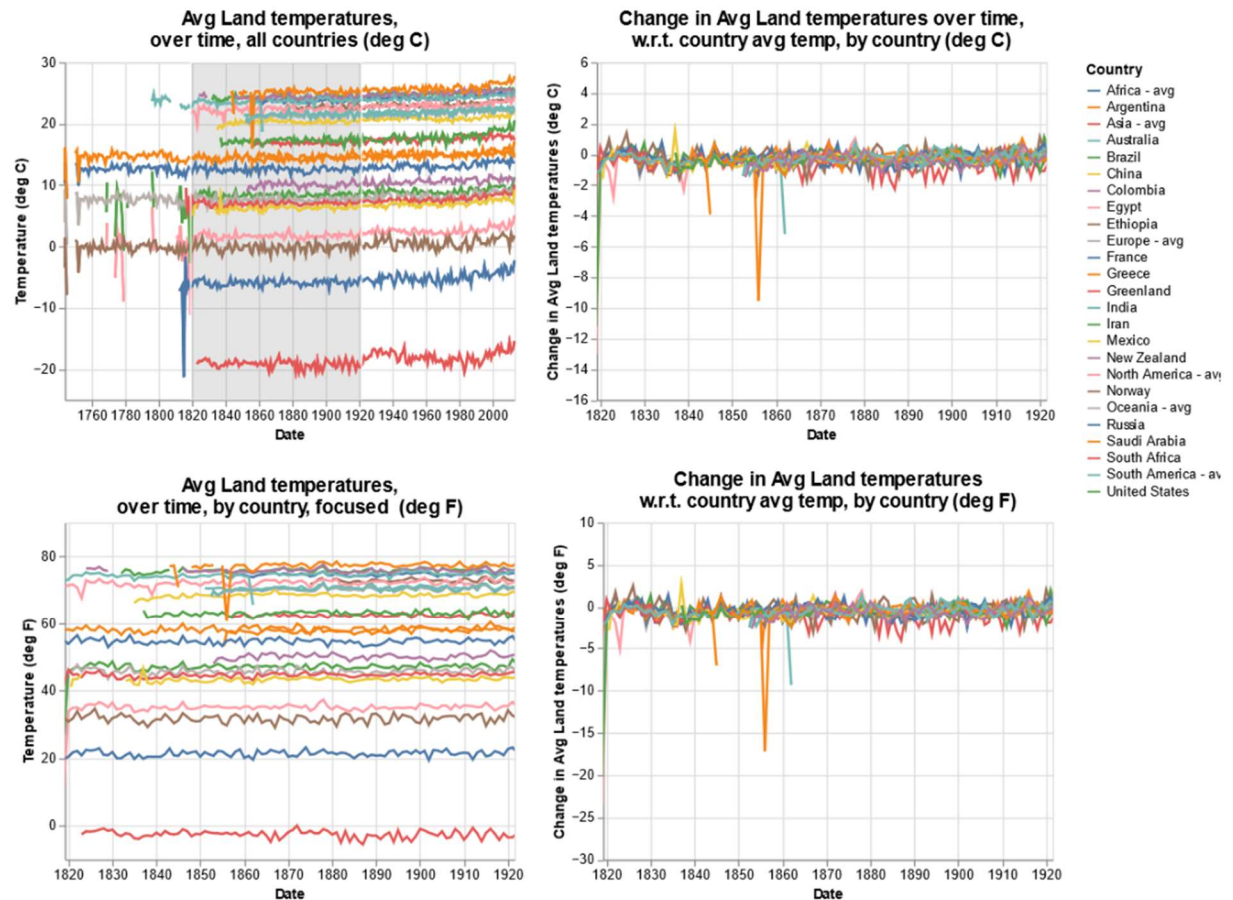
2. The user can select which countries they wish to focus on, and the visualization will present those countries data in specific charts and formats. You can see an example of that below.



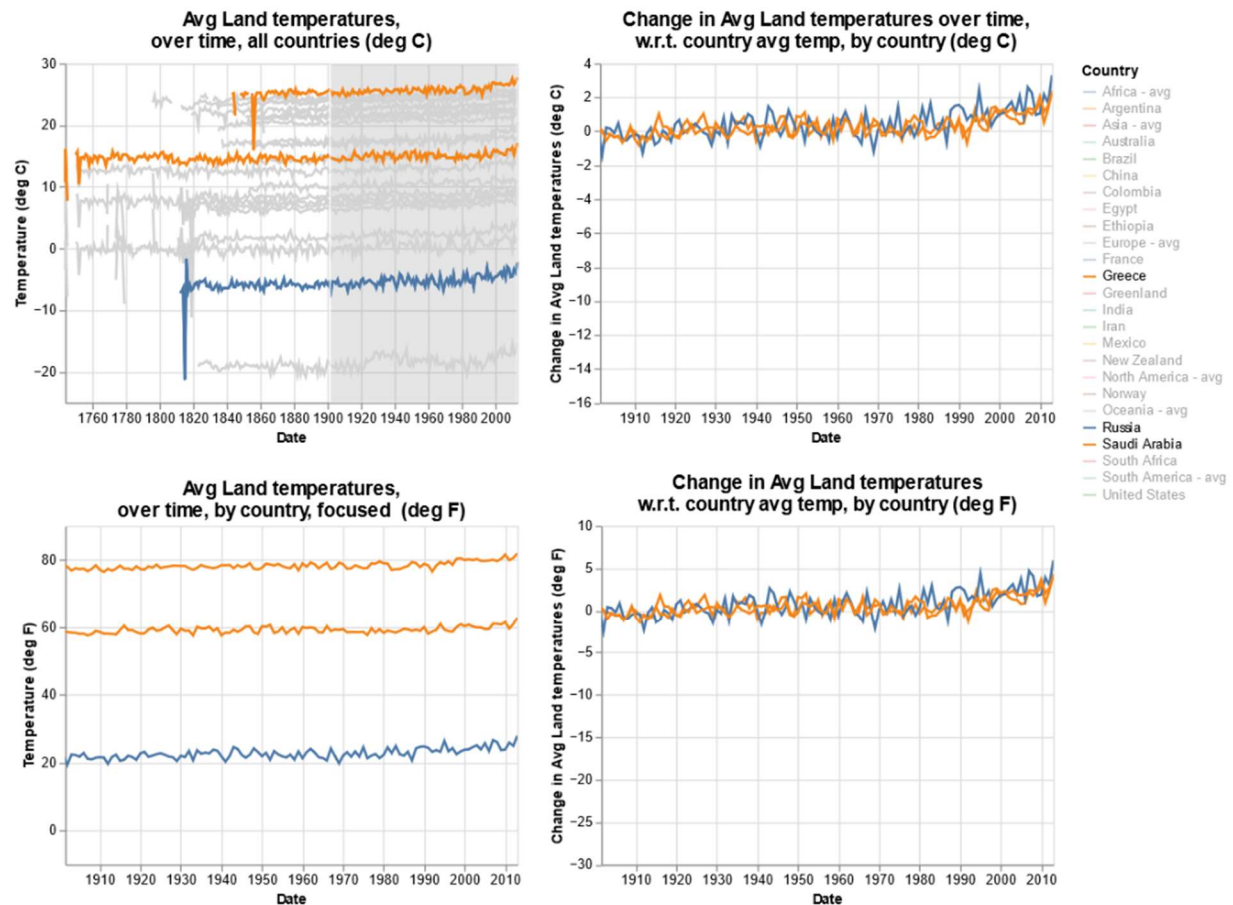
This was a major priority, because there are many different countries and continental averages presented in the visualization, and it is important for the user to be able to focus in on just a few of those options and zoom back out to the total data set. With this functionality, the user can quickly

change perspectives and gain new insights from the data without having to see past a bunch of other data present on the chart.

3. The user can select the time span over which they want to view data, so that if a given countries data spans from 1800 to the present day, they can focus in on the dates between 1900 and 1950. An example is shown below.



To see all features working together, see the next image below.



After looking through the data set originally, I noticed that many countries and continental averages had unexpected peaks and troughs within their data. To be able to focus in on those time spans, it was necessary to incorporate some form of filtering and magnification into the visualization. This filtering, coupled with the main chart always displaying all data over all dates present in the data set, allows the user to focus in on the time span in question while being able to look back at the main chart and see where that span sits in the data set.

The procedure I developed to test this visualization was to present the visualization with a quick tutorial on how to manipulate the graphic and cycle through different countries in the data set, then let the users explore the graphic for five to ten minutes. After that time had passed, I would ask them to share any thoughts, impressions, or realizations that they made while exploring the data in the visualization. With that information, I would check to see if the participants walked away with the understanding that average temperatures for all countries presented in the graphic were increasing and had increased by about the same amount. The people I hoped would help me answer this question were people who were not experts or very knowledgeable about climate change and global warming. I wanted to check that the graphic was illustrative enough for non-experts.

After interviewing three family members on the graphic, the results were that only two out of three participants expressed without solicitation that they saw temperatures increasing for all presented countries, and their temperatures had increased by about the same amount.

The elements of the visualization that worked best were the parts that allowed the viewers to compare one countries data against another, and the charts that showed the net change in a countries average land temperature. From feedback, being able to see that change for each country side-by-side helped bring about that realization.

There were a few elements that detracted from the visualization, and some that were missing that would have helped the views parse the data better. First, the charts that did show the selected countries net change in temperature did not have adjustable y-axis scales, meaning that if the user selected a timespan where the min and max data points were within the bounds of the overall max and min data points, the y-axis scales would not shrink to bring that data into greater focus. This meant that the users had to spend a little more time inspecting the graphs to see the net changes clearly. Second, one viewer said that they would have liked it if there was a 5th graph, on the mimicked the bottom left graph, but showed the temperature in Celsius. This chart would be put side-by-side with the top right graph, and the graph that was present in the top-left corner would be moved up above. This would create a 2 X 3 grid. The viewer in question said that this would help compare temperatures in Celsius in the same way that the two charts in the bottom row allowed for in Fahrenheit.

Altogether, 2 out of 3 indicates a 66% success rate, though with the amount of feedback indicates that the graphic could be much more improved to make the data more visible, and perhaps make the desired insights more easily attainable. With such a small sample size, these are not reliable statistics, and more experimenters would need to be incorporated into the analysis to determine how well the graphic actually does at communicating the information stated above.