COVID Spread and Geography

Dan Prendergast

Introduction

The COVID-19 Pandemic continues to impact our daily lives. One aspect of the fight against the pandemic is limiting the spread of the virus between countries. In this project, I developed a visualization that shows how new daily cases increase and decrease over time in each country. The objective is to evaluate whether there is a connection between country proximity and outbreaks. In other words, are neighboring countries likely to see an increase in cases when their neighbor has an outbreak? This visualization showed some indications of this phenomenon, but not enough to come to a robust conclusion.

Precondition

1. Learning

Early in the pandemic officials recognized the importance of securing borders to limit the spread of the virus. However, there are multiple routes into a country. Primarily, there are air, ground, and maritime points of entry. Air travel allows the virus to potentially pass directly between countries on opposite ends of the world. I wondered if the spread of COVID was highly correlated with proximity to countries experiencing a wave of new cases. Or perhaps, air travel negates the role of physical proximity to countries with high levels of infection. This project expands my study of COVID from the previous assignment using the COVID dataset available from the website, “Our World in Data” (https://ourworldindata.org/coronavirus).

2. Winnowing

In this project, we are primarily interested in how the virus spreads. The dataset provides several metrics – total cases, total deaths, new cases, vaccination, etc. But to understand the spread of the disease, our team of experts believe new cases per million citizens will provide the best insight into the spread of COVID. Total cases, when shown graphically, will mask the appearance of new cases as the number of total cases becomes large.

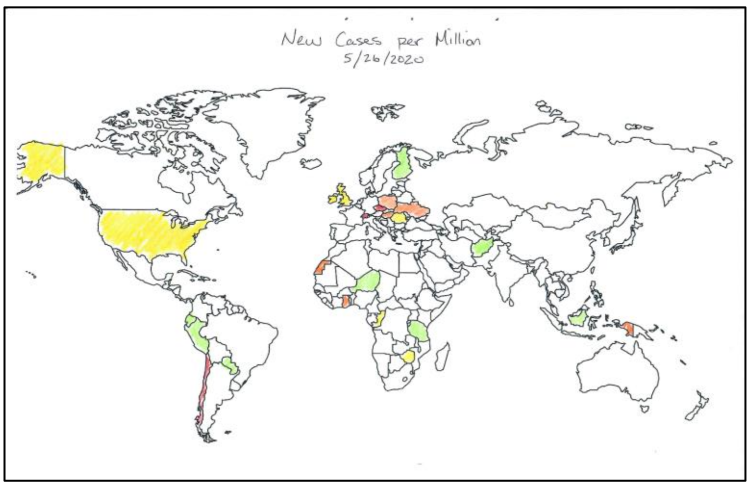
Core

3. Discover

The first task that we hope to accomplish is to visualize the initial spread of COVID to see if physical proximity to high-risk countries correlates to waves of infections in neighboring countries. An appropriate visualization format will be data plotted on a map of the world, and consequently we will be using the COVID statistics combined with spatial data for each country. This analysis is many months after the initiation of the pandemic, and ultimately, the objective is for infectious disease experts to gain an understanding of the initial spread of worldwide pandemics. This understanding might inform a better first response to the next major pandemic.

The second task relates to the spread of COVID after air travel was suspended. International air travel was effectively suspended by early April (Freedman et al., 2020, https://www.washingtonpost.com/graphics/2020/business/coronavirus-airline-industry-collapse/). This raises the question – did the correlation between proximity to high-risk countries and waves in neighboring countries change after international air travel was suspended? Our map of the world visualization should also help us examine this question as well. As vaccines are distributed, I hope we are nearing the end of this phase. However, the question of if, and how, air travel affects the geographic spread of the virus is still important. This understanding may shape policies this year if further waves of infection occur.

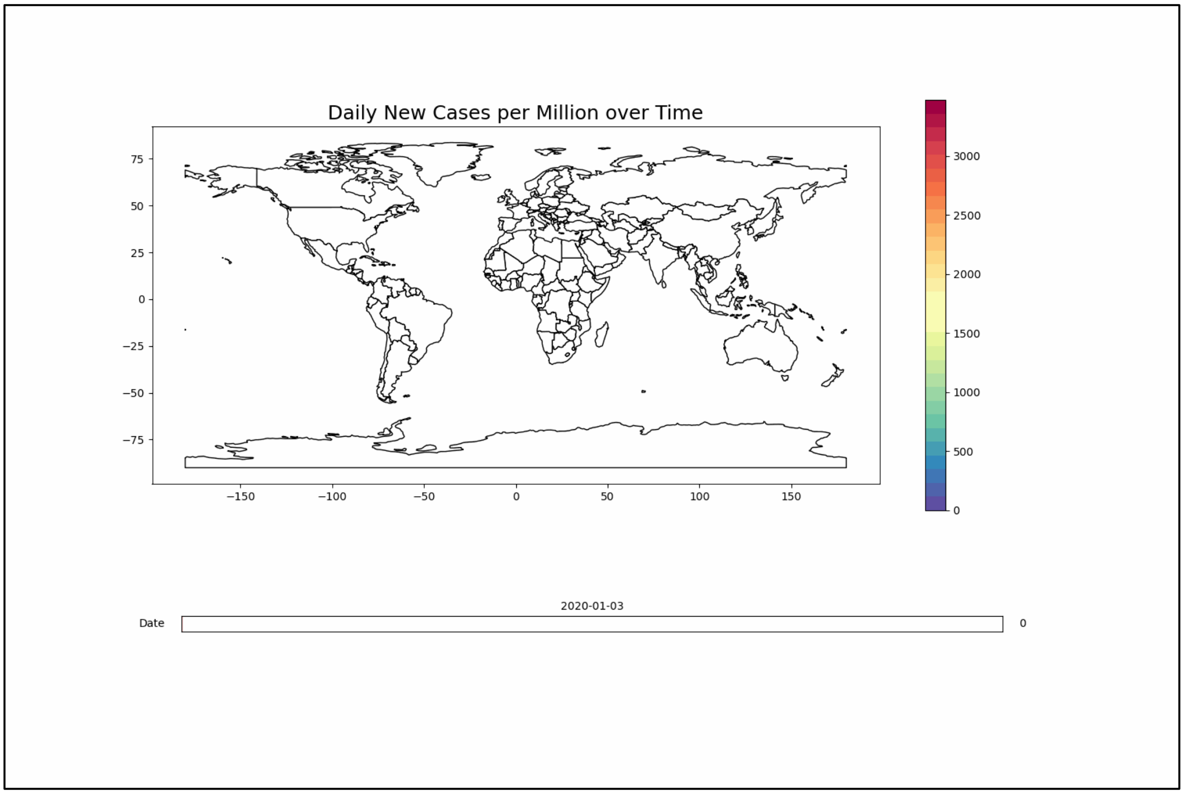
4. Design



The prototype above is a first iteration of the visualization. The first trade-off was how to represent the number of new cases in each country. The designers considered bubbles in the center of each country that grew and shrank as new daily cases increased and decreased. Ultimately, the designers settled on coloring each country according to the number of new daily cases. Bubbles in the center of large countries might not physically touch the border of neighboring countries, which we considered important when illustrating proximity and geographic spread of the virus. A second trade-off was the colormap. We wanted to select a colormap that illustrates a "good" and "bad" condition. The "Spectral" colormap uses blue and green on one end and red on the other. However, the colormap as implemented in the software library needs to be reversed to associate red with a high number of new cases.

5. Implement

I implemented the visualization using the Python libraries GeoPandas and matplotlib. A gif of the first visualization (screenshot below) is in the project folder submitted. Open "covid_spread.gif" to see the animation. The animation should show spread of the virus if proximity to outbreaks in neighboring nations is a factor. If proximity to an outbreak is an important factor, we would expect to see a red bloom in Country A followed by a ripple of warm colors in the neighboring countries in the following days.



6. Deploy

After observing the animation of the new daily cases date, it is difficult to find a trend of outbreaks in a specific country followed by a spread to neighboring countries. There seems to be little indication of this in the early phases as there is almost no discernible pattern to the reports of initial cases (i.e., country goes from white to purple in the gif). This seems intuitive as air travel still permitted people to transmit the virus around the world. After the suspension of air travel in April, there seems to be a large reduction in daily new cases in late April and early May. After April, there appear to be some instances of neighboring countries having correlated outbreaks, particularly in South America, and to a more limited extent in Europe. One explanation is that the virus is being transmitted over ground to neighboring countries.

7. Iterate

The visualization lacks controls to be able to stop, rewind, and replay time periods that require closer analysis. For the second iteration, I will make the slider bar interactive so the viewer can click and drag to select specific days, rewind, and replay portions of the animation.

Analysis

8. Reflect, Part 1:

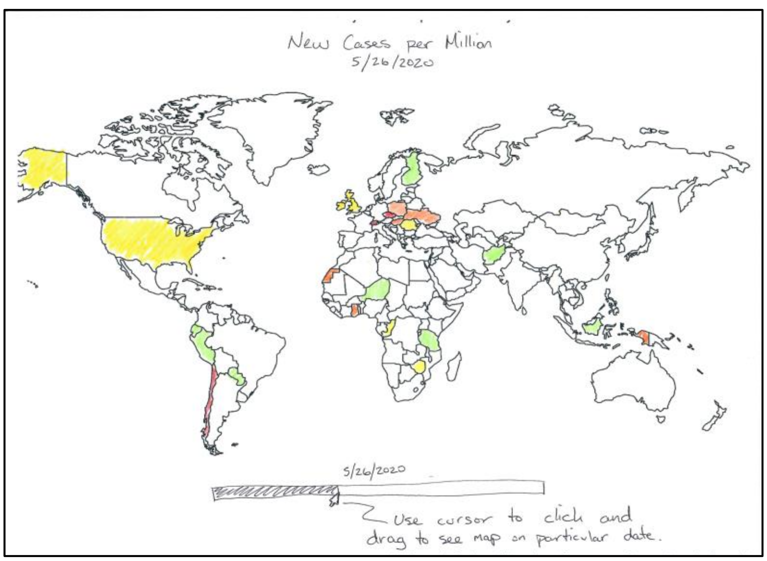
The visualization does provide some evidence of transmission of cases to neighboring countries; however, it is not convincing enough to determine that it is a trend or establish causation. There are many other variables to examine, such as major air routes, porosity of land borders, national policies to handle the pandemic, etc.

9. Reflect, Part 2:

The animation did provide a useful visualization of the rise and fall of new cases in each country. However, I think it needs further details to illustrate transmission pathways. The visualization could integrate major air routes, for example.

Iteration

4b. Design



The prototype drawing does not change much with this iteration. We will add the ability to click-and-drag the date slider at the bottom of the animation. This will allow a more detailed look at a date or range of dates.

5b. Implement

The implementation of this design required adding slider control logic and event handlers to the code. Unfortunately, this implementation broke the “play” feature of the animation, and I was unable to fix it. But the slider allows you to "play" a date range by just clicking and dragging slowly over the dates of interest.

6b. Deploy

The finished visualization is saved as a Jupyter Notebook in the submission zip file. Instructions on how to run and operate the visualization are in the Notebook.

Using the new interface, we can selectively analyze dates of interest. Using the slider bar, I examined the period before and after the suspension of air travel in early April 2020. Before April 10th, there are multiple countries shown in orange and red. After mid-May, the entire globe appears in purple and blue. This is evidence that the suspension of air travel may have helped reduce international transmission.

In mid-June and early-July 2020, we see evidence of cross-border transmission in South America. On several dates we see spikes in new cases in Chile and Brazil, followed by increases in Peru and/or Bolivia. This indicates some support for the hypothesis of transmission by proximity to neighboring outbreaks after the end of air travel.

7b. Iterate

On the next iteration it would be interesting to depict major air routes that can be colored according to routes' departure and arrival nations. Also, higher granularity in large countries (e.g., to the state level in the U.S.) would provide a better illustration of COVID spread.

8b. Reflect, Part 1:

I do not think this iteration provided greater insight into the data, but it made using the data much more efficient by allowing control of the slider. In the previous iteration, I would have to watch the animation several times to record dates/locations of outbreaks. Now, I can simply slide through the dates that interest me.

9b. Reflect, Part 2:

This iteration demonstrated the importance of not breaking features of the previous iteration while adding new features. In this iteration, the auto-play feature of the animation broke. It turned out to be a small issue, but other features could have ruined the experience.

Code References

1. I used this webpage (https://www.kaggle.com/zmcddn/animated-geographical-visualization-with-matplotli) to learn about how to implement map-based data representation. Although I modified this heavily to a more object-oriented implementation and to integrate the interactive slider.