

Team members:

Nick Abcarius, Andrey Yakymovych, Lia Freed-Doerr, Sabarno Islam, Zibo Zhang, Joshua Winter, Emma Baier, Kwan Tien Ooi, Yi-Hsuan Cheng

Write-up:

For over a year, global travel has been at a standstill due to public safety protocols. Now, as COVID-19 cases have begun to decrease, the world is opening back up and people are developing their summer travel plans. We wish to enable individuals to make a smart choice for their travel destination so that the spread of the virus continues to decline.

The objective of our project is to provide a visualization of data to show COVID-19 infection rates for multiple times in a region. The end result will include a map with regions. Users can use a slider to choose the time which would color the map with infection rates. Users can also select regions to show the trend line in infections.

This project is intended to help users make efficient and wise day-to-day decisions about their personal measures of protection against COVID-19 by providing them with information about COVID-19 infection rates. The model implemented in our project aims to make it easier for people to protect themselves and others from Covid-19 through the presentation of data that represents a snapshot of the dynamic conditions of the pandemic at a given time. By keeping users informed about the current number of COVID-19 infections in their population, our project should allow for users to solve problems they have with making informed decisions about their own safety measures, as well as problems they have with estimating how safe it is to go back to "normal" activities, such as social interaction and travel, as COVID-19 numbers rise and drop.

Since we are trying to predict the proportion of individuals in a given region (a continuous variable) based upon a labeled dataset, this will be a regression problem. We are using this covid19-open-data from Google Cloud, linked in the following: <https://console.cloud.google.com/marketplace/product/bigquery-public-datasets/covid19-open-data>. This resource has a wealth of information, with a lot of country data and data by state available for a number of new cases, number of recovered cases, age range info, economic info, comorbidity info, among many, many other factors, totalling up to over 10 million entries and almost 700 attributes. Problems include the vast amount of null data as a result of different countries and/or states reporting different data and, subsequently, the amount of data cleaning that will be required. However, we have some valuable data that we will use in our predictions, likely including the amount of new confirmed cases per age group, population data, and potentially, the more recent vaccination data.