Project part 1:

How data was collected:

Data from individual countries was obtained from the individual country statistics pages from worldometer.com. For each countries page, the worldometer URL is <https://www.worldometers.info/coronavirus/country/> \*country name\* /. This allowed for us to create generate a full url for each country from an input country name, however it is sensitive to spelling (ex for the US it needs to be US not USA or United States) so we generate a static dict for the country names we want to scrape from. Then to scrape data, we used requests library to pull the page contents as an HTML and the Beautiful Soup python library to parse that data. We wanted the daily deaths and total deaths for each date. This data is stored in a highchart format in the HTML file, so it could be scraped by searching a few consecutive unique code sequences. for example, the dates column was "Deaths per Day<br>", "categories:", '[' , and closing with , ']'. After the data was extracted like this, it was put into a Pandas dataframe, the dates were reformatted, and extra columns were added with data normalized to 100k population ( country populations extracted from worldometer as well). This dataframe is then exported as either a pandas dataframe or dict. A separate function can combine dataframes into a nested dict json export

The json file is organized in a nested dict structure where:

{ 'country' : {'date': {1: date1, …}, 'death rate': {1: deaths , …},'death c': {1: dr , …} }, 'country2' : { …}}