Data scraping was accomplished using the Beautiful Soup library. Right clicking on a data point on the website and choosing ‘inspect’ enables us to see the element tags and scroll through the nested HTML data structure of a table on a website. We can use the tag name and attributes to call up this data in our code. Furthermore, knowing the structure allows us to loop through multiple data points associated with country names which are stored in the contents of the tags. A country name could then be used as input for a function that would retrieve data specific to it.

From the ‘Worldometer’ website, once the data is separated by country there are 13 data points: Total Cases, New Cases, Total Deaths, New Deaths, Total Recovered, New Recovered, Active Cases, Serious Critical, Cases per Million, Deaths per Million, Total Tests, Tests per Million, and Population. For this function, only Population, Total Deaths, Deaths per Million, and New Deaths are stored. Then New Deaths per Million is calculated and stored from New Deaths and Population. The data is sorted in to a multi-dimensional dictionary - organized first by country and then data type - and then saved to a json file, titled ‘COVID\_\*CurrectDate\*.json’, in the current file location.