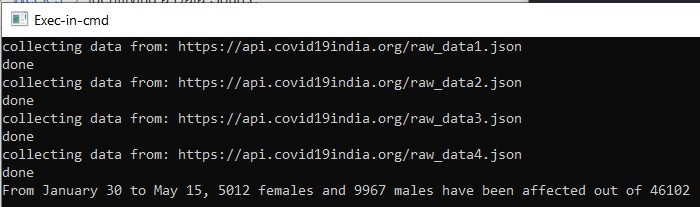
I have examined 2 websites for COVID-19 data. I initially wanted to extract COVID-19 data of my country, Bangladesh but I couldn't find any government website that shows the day to day numbers. They only provide the total number of patients, deceased, etc. So I gave up and started exploring the forum here and found "<https://www.covid19india.org/>" which contains a very well structured data set of COVID-19 in India and used their APIs to do some basic analysis of the data. Their JSON APIs are properly formatted so it was very easy but I wanted to do my own scraping. Then I found "<https://www.worldometers.info/coronavirus/>" which contains COVID-19 daily updated data from 15 February till now. The current pandemic piqued my personal interest in the data so I choose the above 2 websites. Unfortunately, I am not skilled in statistical analysis and mathematics so I just did very simple work.

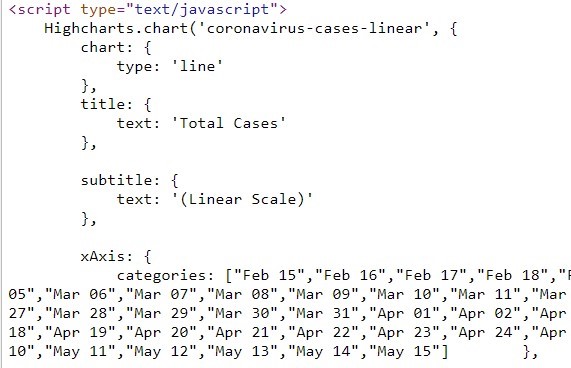
1) The site "<https://www.covid19india.org/>" provides updated COVID-19 data and APIs that make data retrieving easier. Their APIs contain a lot of descriptive data like the daily cases and the gender, location, etc of the patients. But they do not have all the patients' data since they had to collect them personally (without the help of the Indian govt.). I used their APIs to find an approximate number of female and male patients from Jan 30 to May 15. The numbers here are lower than in reality.



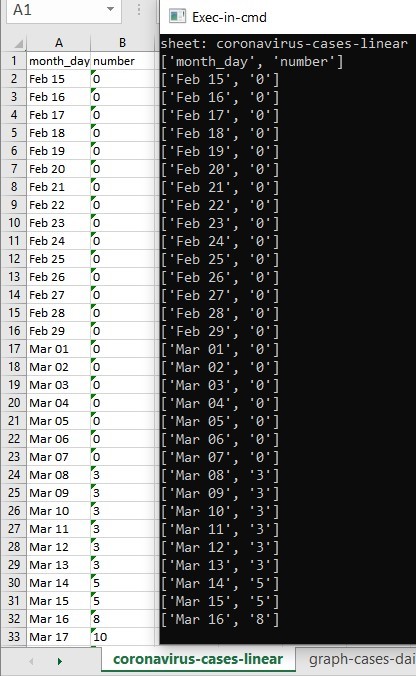
The screen snip below shows the accurate number of confirmed patients in India. The number of patients I've calculated from their APIs is lower.



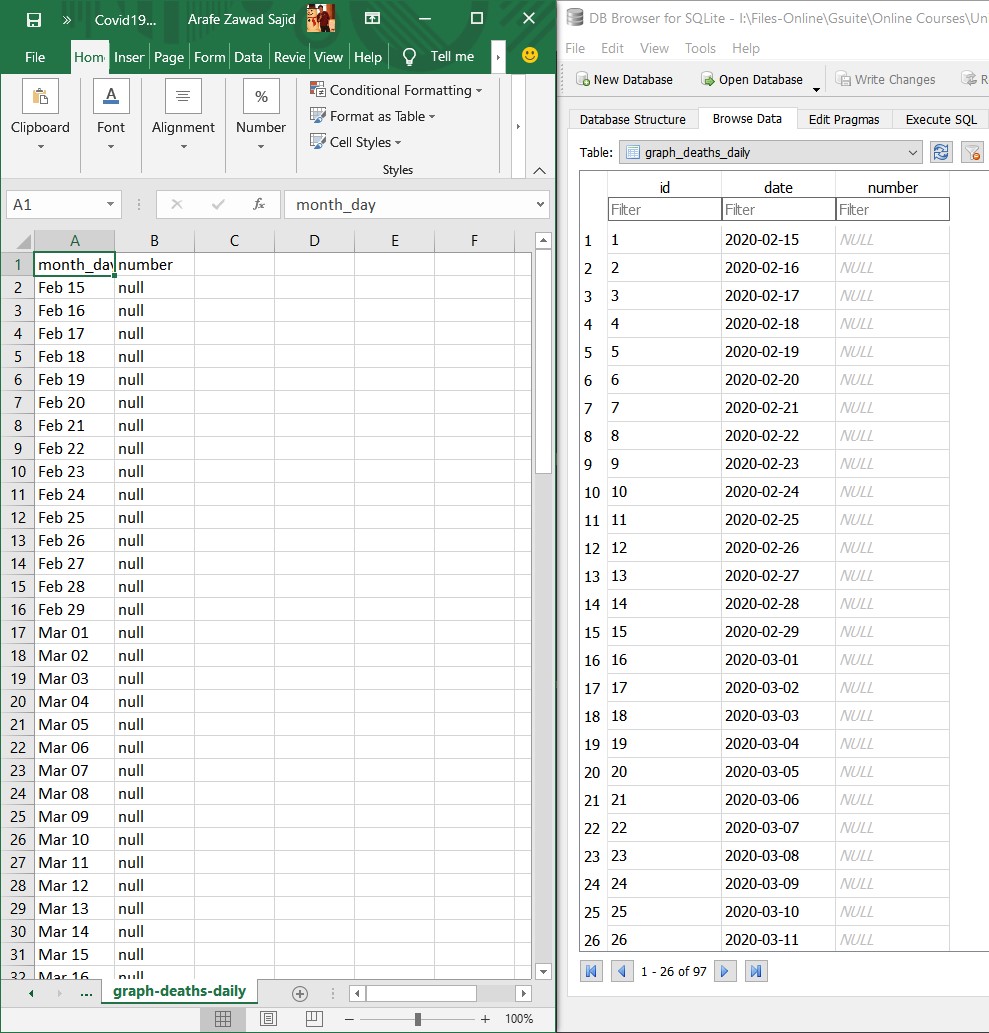
2) The site "<https://www.worldometers.info/coronavirus/country/bangladesh/>" contains the daily COVID-19 data but I could not find any APIs. So I had to crawl my way to get the data. It was tough extracting data from an HTML containing nested javascript. I had to use BeautifulSoup to parse up to the javascript portion and then string manipulate my way to my desired data.



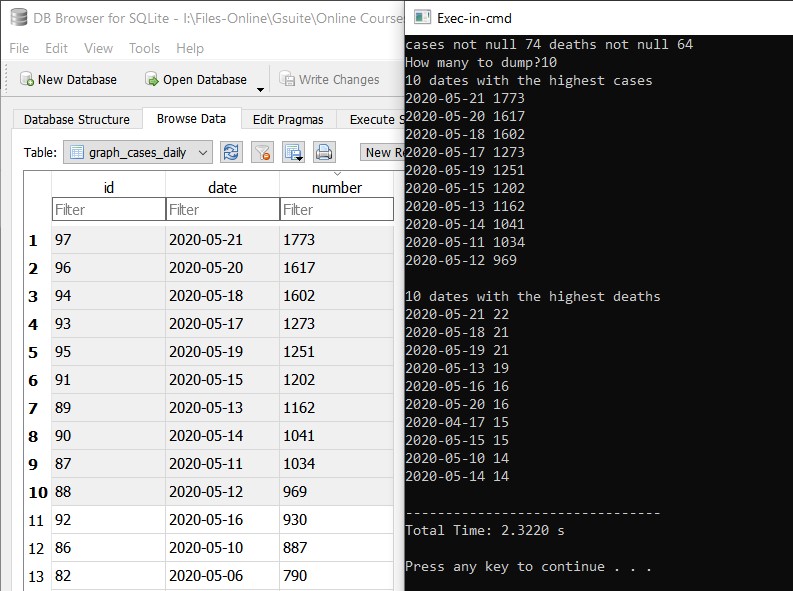
My program scraped the data and saved those in an excel. Then I made another simple program to dump the values from the excel.



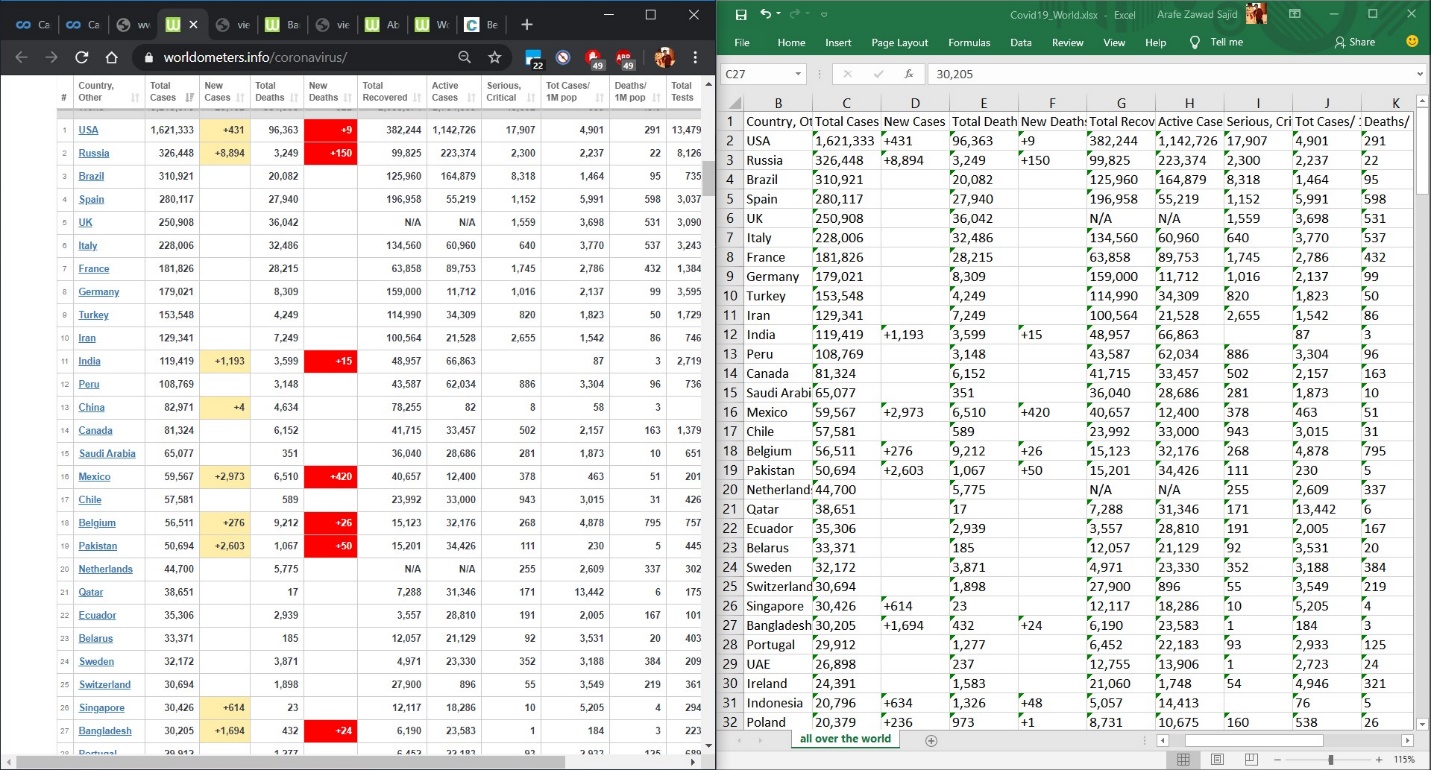
I found this site very interesting as it holds daily data so I am going to focus on this site more, specifically the data they have on my country Bangladesh. In the next step, I clean the data as there are some 'null' strings inside some cells (got the 'null' from the site's HTML) and the dates need to be standardized. I am considering ‘null’ as the absence of data and not as zero. I store the clean data in a database for some basic analysis I will perform in the next step.



Now for a basic analysis of the data. I will just find out the dates when the number of daily cases and the number of daily deaths were the highest in Bangladesh. Dates with the lowest numbers can be found similarly. I will be looking at the daily cases and daily deaths table only since the lastest date does not necessarily hold the highest number. On the other hand, the total number will always be maximum for other tables for the latest date. (I am sorry I am not being able to come up with a better choice of words)



I haven't done any visualization yet. I planned to scrape day to day data of all the countries on this website by spidering through the HTML (saw links of 215 countries) since this is the only website I found that contains daily figures. I have successfully made such a program that can do so but I did not retrieve all 215 countries as it takes a considerable amount of time and leads to creating 215 excel documents. I have also scraped data from this site which shows the current condition of all the countries in the world and put it in an excel.



That's all from me. Would love to have some suggestions on how to visualize the data.