Data Wrangling Group 9 Project Update for 16Mar2023

So far, our group was able to separate the COVID-19 data. We separated by cases (active cases, confirmed cases, deaths, and hospitalized) and vaccines for single dates over a period of time. These were separated into data frames, therefore the csv files remain the same, however when the script is run, they can be accessed as their individual dataset/frame.

For the air quality data, we are using CHPC services as the data set is very large. The .csv files were copied over to that server however when trying to install with pip3, a read only file system error occurs. Therefore we had to create a virtual environment where we were able to install all the dependencies required but faced an error of the kernel crashing while trying to read the .csv file (converting it into a dataframe). After multiple tries we noticed that the kernel automatically stopped after 2-3 minutes and we were not able to proceed further to successfully convert the .csv file into a dataframe. Next steps would include troubleshooting this issue, which may require reaching out to CHPC personnel.