Xiaotian Huang, DATS 6401, Summer 2020

Progress Report

## Visualization of COVID-19 and SARS

The dataset for COVID-19 and SARS was retrieved from Kaggle.com and JHU GitHub in the form of CSV.

**Data Preprocessing**

First, I changed columns names to “Confirmed”, “Deaths”, and “Recovered” for both datasets to make sure they are consistent with each other. Then, I grouped data by Country for both datasets, so that it would be clearer to see total confirmed cases, recovered cases, and death tolls in each country. I also grouped the data by date to see the trend of virus transmission. In this step, I changed the SARS dataset to time series dataset. I did not change the COVID-19 dataset since a ready-to-use time series dataset is available in JHU GitHub. Lastly, I saved all data to a new CSV file. [Steps described above were done in python]. The next step I’m going to do is to calculate the confirmed rate, recovered rate, and mortality rate and make charts out of those three rates.

**Data Visualization**

For now, I’ve uploaded a ready-to-use dataset to Google Sheet. From the dataset, I created two Geo Charts using Google API to show countries most impacted by the coronavirus. I also imported the data into Tableau to generate graphs illustrating the confirmed trend, mortality trend, and active trend each day.

**Webpage Construction**

I constructed the frame for my webpage in HTML using CSS styling and W3 Schools template. I will then import graphs generated from Tableau and Google API to visualize and analyze the disease transmission trend.