Our dataset is data from the coronavirus outbreak. We have data on how many confirmed cases, people recovered, and people died are in various cities, counties, states, and countries, around the world. We have access to updates to this dataset every 24 hours online and we got the dataset from Kaggle.

Here are our questions we would like to investigate:

1. **For the entire world, can we predict how many total people will have the virus the next day?**
   1. **Can we scale this down to specific regions or states?**
2. **For certain countries can we graph the percentage of people infected over time?**
   1. **For the U.S. we will sum over all the regions, states, cities we have to get the total data for the United States and compare that with other countries.**
3. **For selected countries with distinct average temperatures, is there a correlation between the number infected and the average temperature?**
4. **For selected countries with distinct GDP, is there a correlation between the number infected and the GDP?**

We would like to do some linear or exponential regressions on this data.