Team Name: Hidden Dragon
Group: 1
Members: David Tang, Jaya Prakash Reddy Dumpa, Shashank Singh

Dataset: https://ourworldindata.org/coronavirus-data#confirmed-cases. This website has all the data we probably need.

Regarding Dataset (number of records, features):

1. Total Covid-19 deaths:
    a. Features: Country Code, Date, Total Death at a given date
    b. Number of records: 7214 (will increase as time goes by)
2. Daily Covid-19 deaths:
    a. Features: Country Code, Date, Daily confirmed deaths
    b. Number of records: 7214 (will increase as time goes by)
3. Testing done
    a. Features: Country Code, Date, Total number of tests done since outbreak
    b. Number of records: 78 (will increase as time goes by
4. Total tests conducts vs total confirmed cases
    a. Features: Country Code, Data, Total number of confirmed cases
    b. Number of records: 7042 (will increase as time goes by)

We want to determine if there is a correlation between COVID-19 reported cases in a given country and the infection rate between its neighboring countries. For instance, does a country with a high rate of reported COVID-19 cases affect the rate of infection in neighboring countries? Inversely, how does a country's rate of infection get affected based on its neighbors. In order to determine whether there is some correlation, we need to compare it to some baseline. The most sensible baseline would be the ratio of **(the rate of spread)/(country's population density)**. Unfortunately, the data that we have of confirmed cases can be drastically different from the real numbers. The rate of confirmed cases can also vary a lot based on the testing capabilities of each respective country. In order to minimize errors from our assumptions, we should normalize the number of reported cases for each country by some weight. We can say the weight is the ratio of **(positive cases)/(total tests performed)**.

We want to also answer the BONUS question: Is China lying about the number of COVID-19 cases? Their numbers have been flat for a while and we find that hard to believe considering how densely populated the country is. There was also recent news about how 21 million cellphones disappeared in China since the outbreak:

https://www.youtube.com/watch?v=A1F_kWYdqUY
https://www.youtube.com/watch?v=wfDAcuSMSZo

*** If we can find cell phone data for each country, maybe we can include this as part of our analysis.