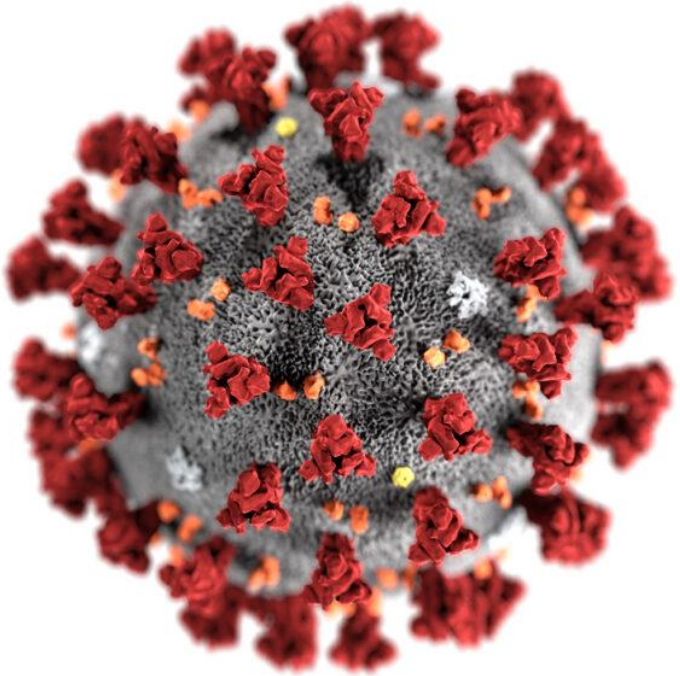
COVID-19 Predictions

David Moste, Sadia Perveen, Vanita Thompson

The Problem

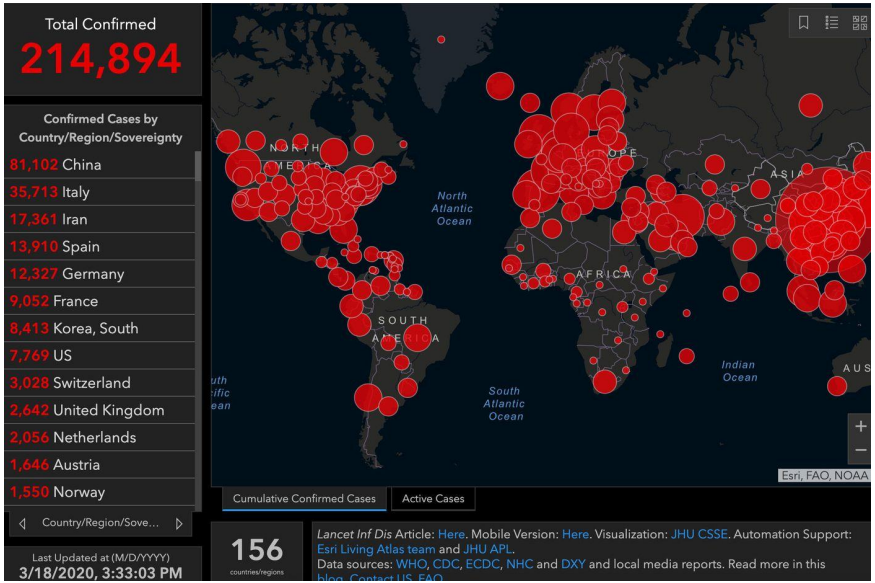
We have proposed an analytical model to predict future COVID-19 cases using global COVID-19 data. The proposed model will use previous confirmed cases, deaths reported, recovered cases and daily cases.



The Data

The data we used for this analysis was obtained from the Johns Hopkins COVID-19 dataset.

This is an extensive dataset that contains information about confirmed cases, deaths, recovered patients, and much more from many countries and provinces around the world.



Our Methods

Our methods can be broken up into the following steps:

1. Data cleaning and preparation
2. Exploratory data analysis
3. Model Building

Data Cleaning and Preparation

Data preparation for this dataset was a little intense. After bringing all the data into a workbook, we pulled out all of the dates from the data and summed all of our data for each country (cases, deaths, and recoveries). We then determined the daily increase for each category and for each country. To finish our preparation, we copied this process for each province listed in the data.

Exploratory Data Analysis

To start our analysis, we built a couple of data tables displaying the following: Number of Confirmed Cases, Number of Deaths, Number of Recoveries, Number of Active Cases, and Mortality Rate. Each table was accompanied by a cmap that highlighted extreme situations. We followed this by making a couple charts (bar and pie) to show the most infected countries and provinces. We finished our EDA by graphing COVID-19 deaths, cases, and recoveries over time.

Model Building

To see if we could predict future confirmed cases ,we built two models: our first model was a linear regression and our second model was an SVM.

Using the sci-kit learn package, we built a linear regression model after splitting our data into training and test sets (model can be seen on next slide). For the first 400 days after January 1, 2020 the model appears to have pretty good agreement with the data.

Using the same package, we also built a polynomial SVM regression model to fit our data. It ended up producing a nearly identical model to our linear regression.

Results

We ended up with two models that predict future cases of COVID-19 around the world. Both models indicate a continued increase in cases, which could be refined with more data.

Any increase is unlikely to meet the expectations of our model as people become vaccinated and transmission slows over time.