

Predicting Covid - 19 Spreads

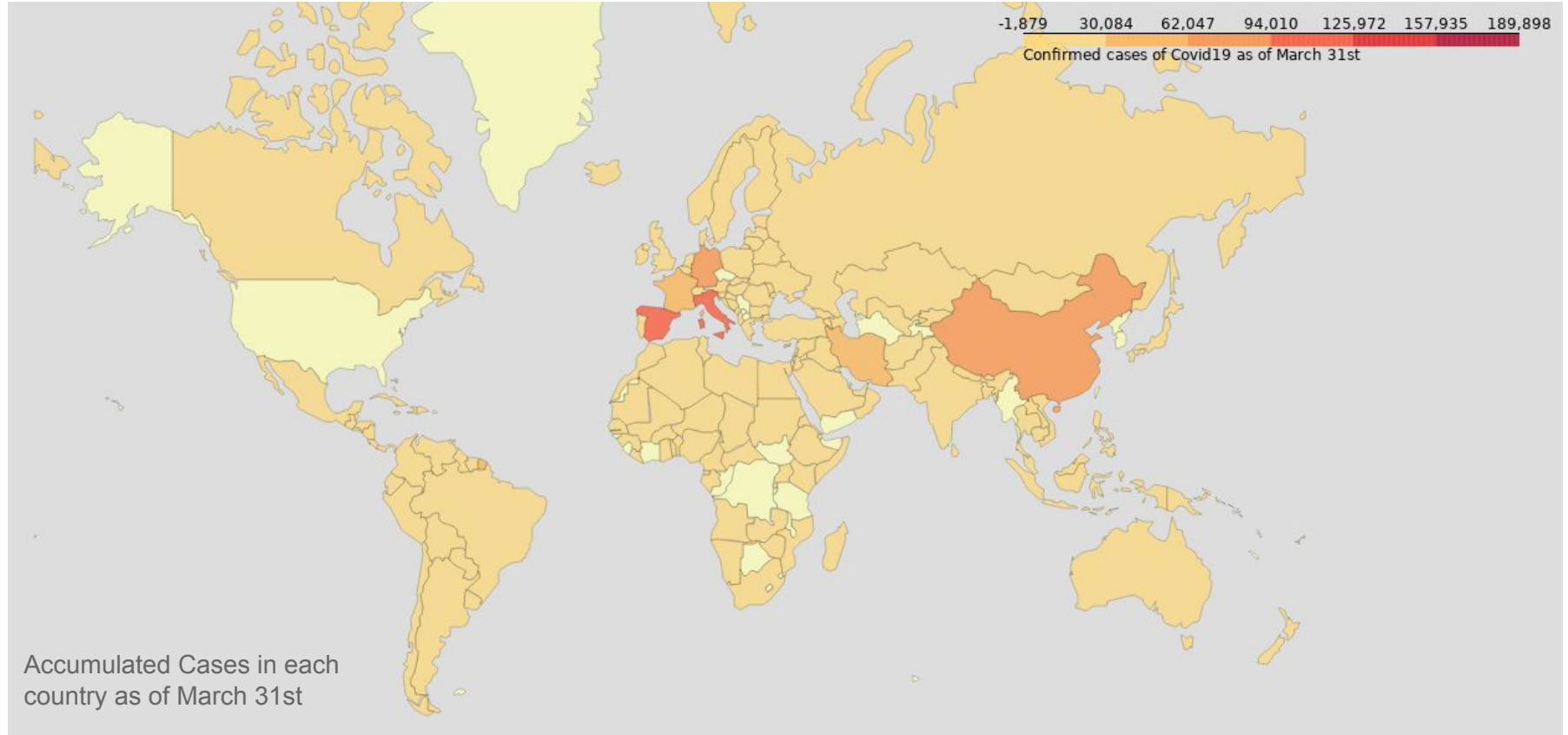
We need to know the trends of this pandemic

-
- The more people are infected person of coronavirus rises, the higher possibility of other person got infected, more people need medical treatment.
 - Patients of coronavirus need 2-3 weeks of hospitalization that tighten limited medical resources.
 - Predicting the spread of virus help managers
 - Medical resource allocation and production
 - Governmental policy makers need this information to estimate necessity of financial support
 - Citizen have interest as well to think how the government policy changes

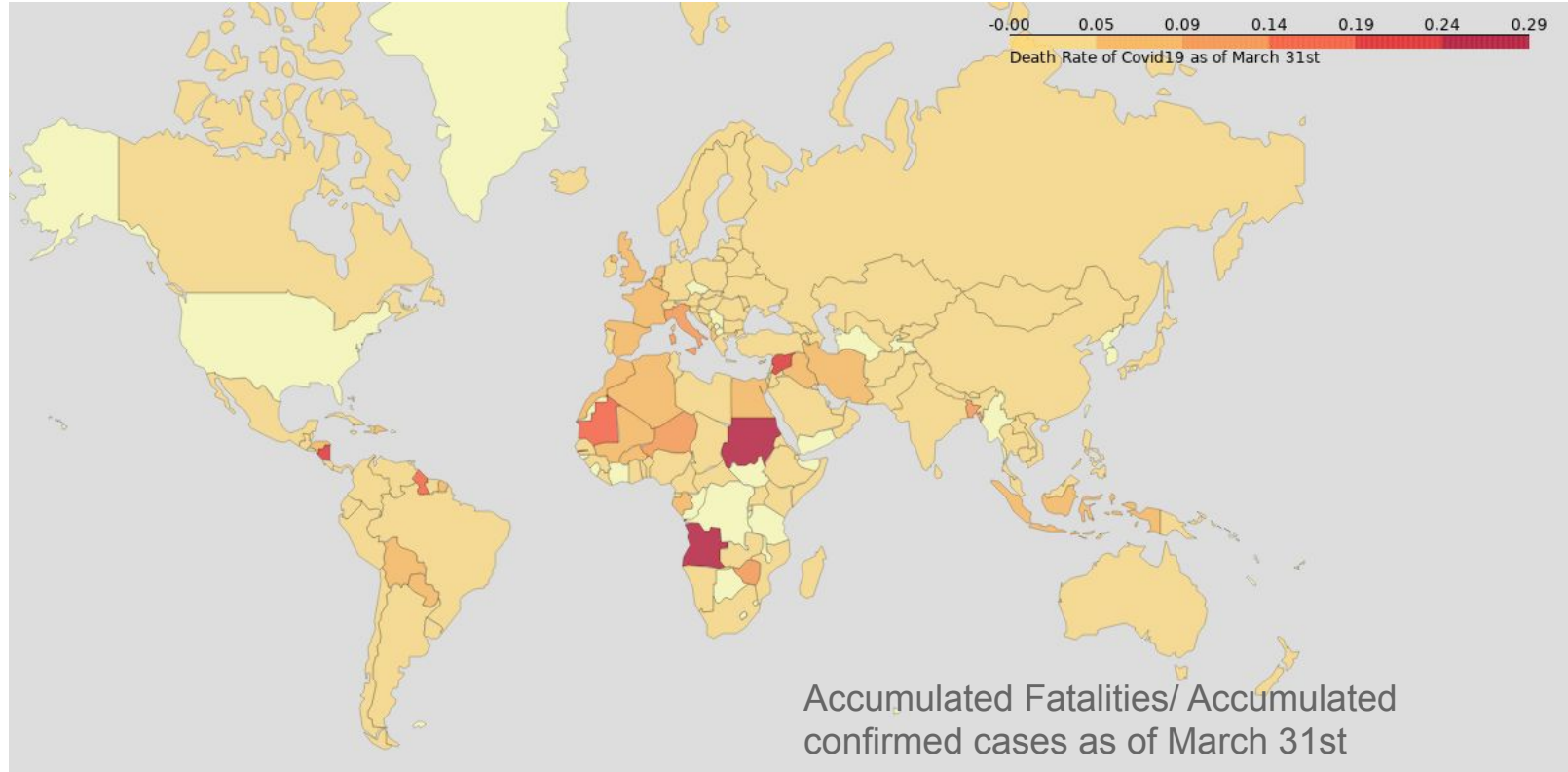
Data acquisition and cleaning

-
- Time-series data of confirmed cases and fatalities by coronavirus per day per country (2020) from Kaggle dataset, 2020 data scraped from [COVID 19 Global Forecasting\(Week2\)](#).
 - Data covers 173 countries from January 22nd to March 31st(70 days)
 - In total, 20580 rows and 6 columns.
 - Cleaned data contains 5 features, Id, date, Country, Confirmed Cases, Fatalities.

European countries and China struggle against the Covid-19



High Mortality Rate in Countries with poor medical resources

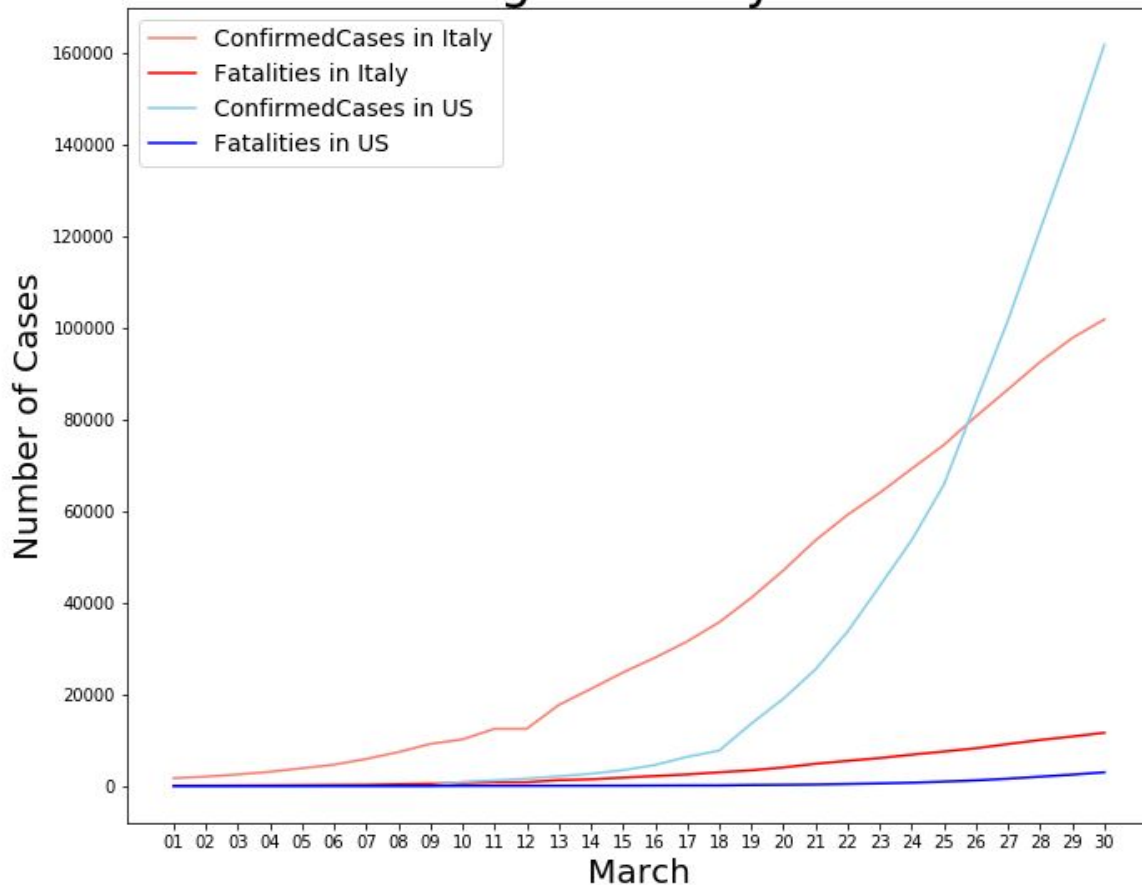


Italy and US is the worst in Fatalities and Accumulated Cases



**Italy and the US
experienced
exponential growth
of Confirmed cases**

Time Series Changes in Italy and US on March

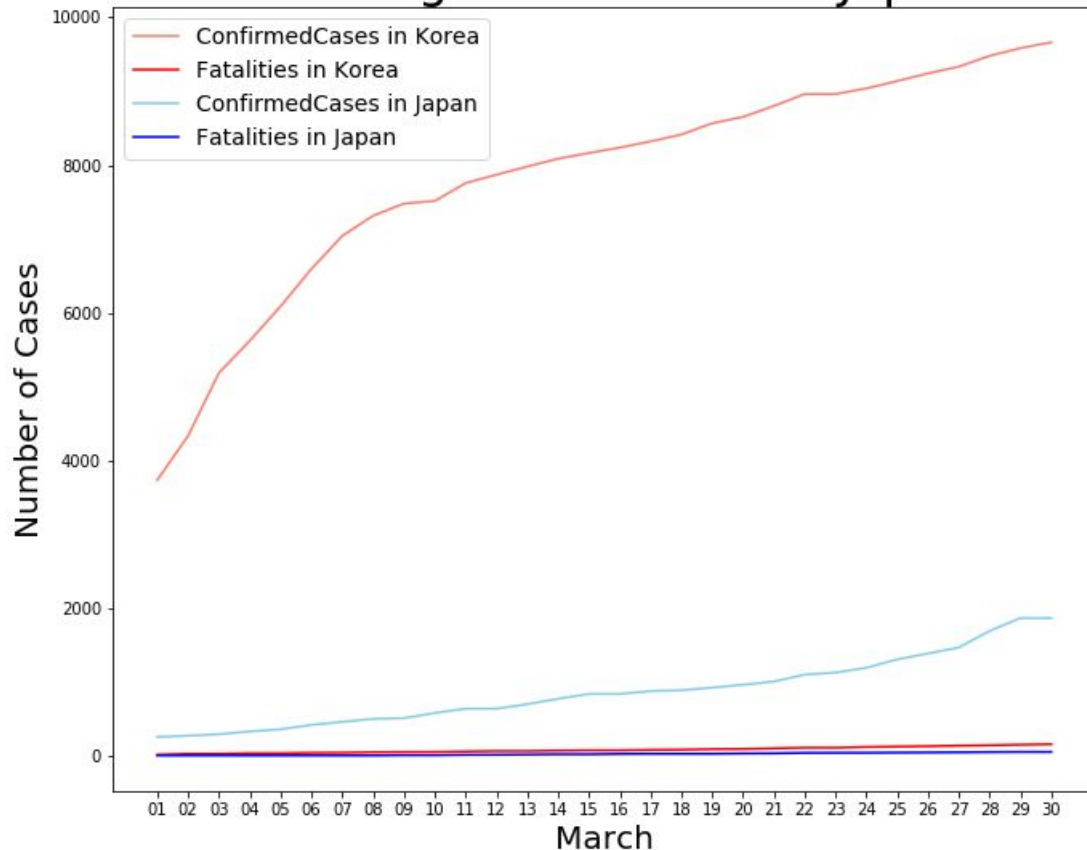


Huge gap in confirmed cases

Fatalities are close, however, Korea succeeded in detecting more patients.

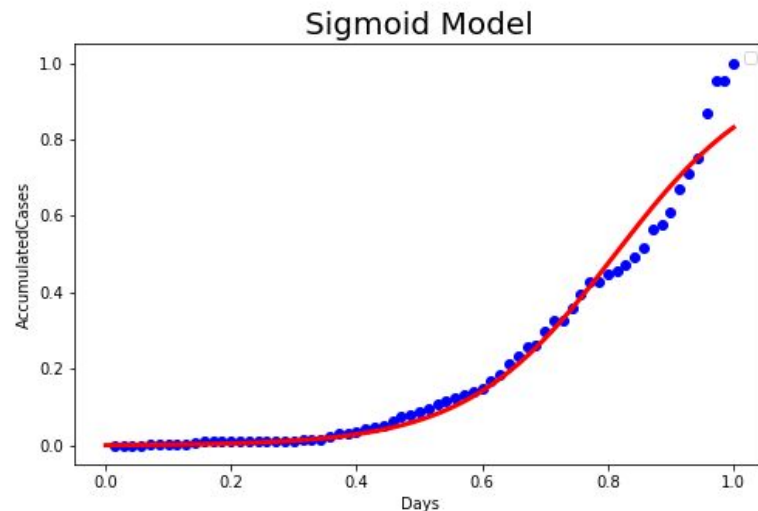
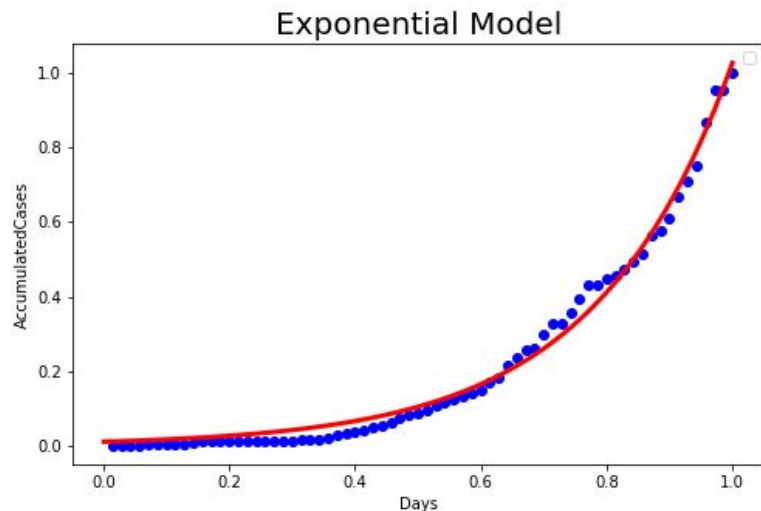
表分ける

Time Series Changes in Korea and Japan on March



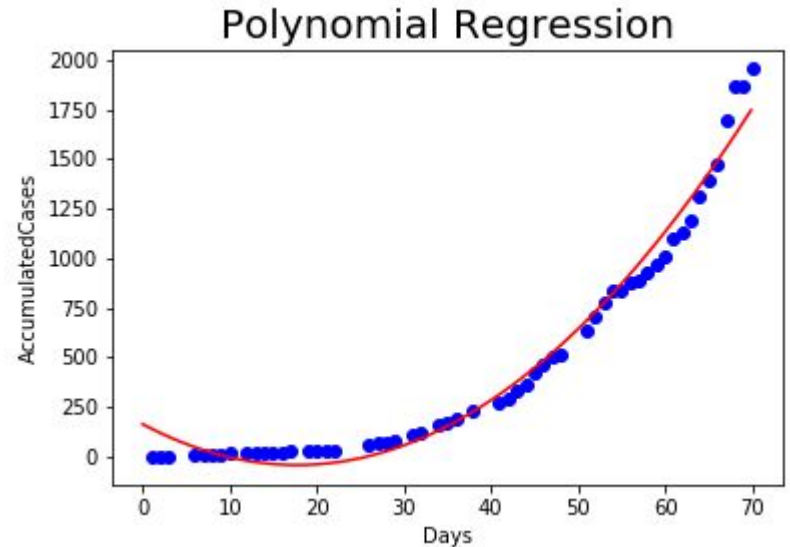
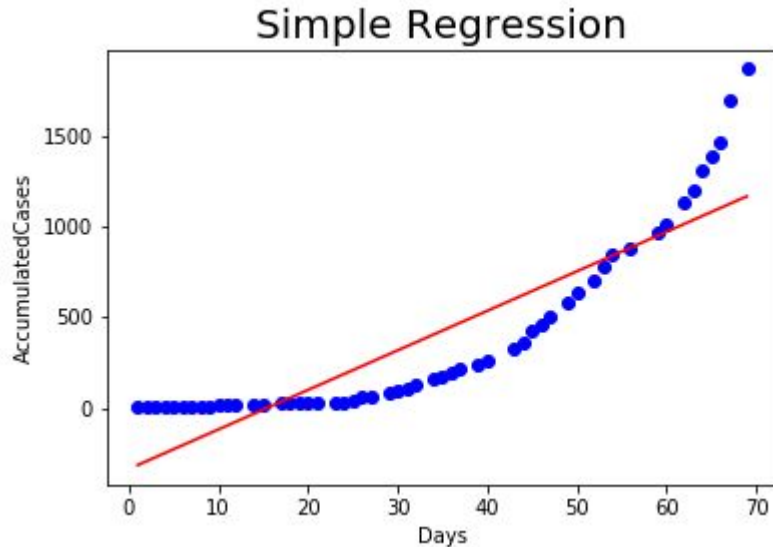
Regression:Non-Linear Regression

Sigmoid Regression best describes the current increase in Japan



Regression models performance:

Sigmoid Regression best describes the current increase in Japan



Model performance for Japanese case

	Exponential Model	Sigmoid Model	Single Regression	Polynomial Regression
Mean absolute error	0.02	0.03	214.14	70.61
Residual sum of squares (MSE)	0.00	0.00	94996.57	11700.58
R2-score	0.99	0.00	0.32	0.96

Model performance for Korean case

	Exponential Model	Sigmoid Model	Single Regression
Mean absolute error	0.16	0.03	1022.42
Residual sum of squares (MSE)	0.03	0.00	1930033.46
R2-score	0.62	0.99	0.84

Conclusion and future directions

- Comparison of estimates of April and actual data
- Use of medical specialized model such as SIR models
- Ideas include:
 - population density, that may affect the speed of spreads
 - GDP per capita for affordance of medical treatment, that might affect fatalities
 - Take medical resources such as the number of hospitals into consideration