Modelling the COVID-19 epidemic

Jack Wu *jack@jackwu.us* *http://covid.jackwu.us/*
March 2021

Abstract
Based on current compartment model for covid-19, I developed this new model, combined with the stochastic process, using (1) mathematics analytic methods (2) numerical differential solvers and curve fitting (3) computer simulations to better understand the public covid-19 data, caliber parameters for each county and region, therefore to explore the reason and forecast the trends for the coronavirus. For more details, please visit this project's website: http://covid.jackwu.us/

Introduction

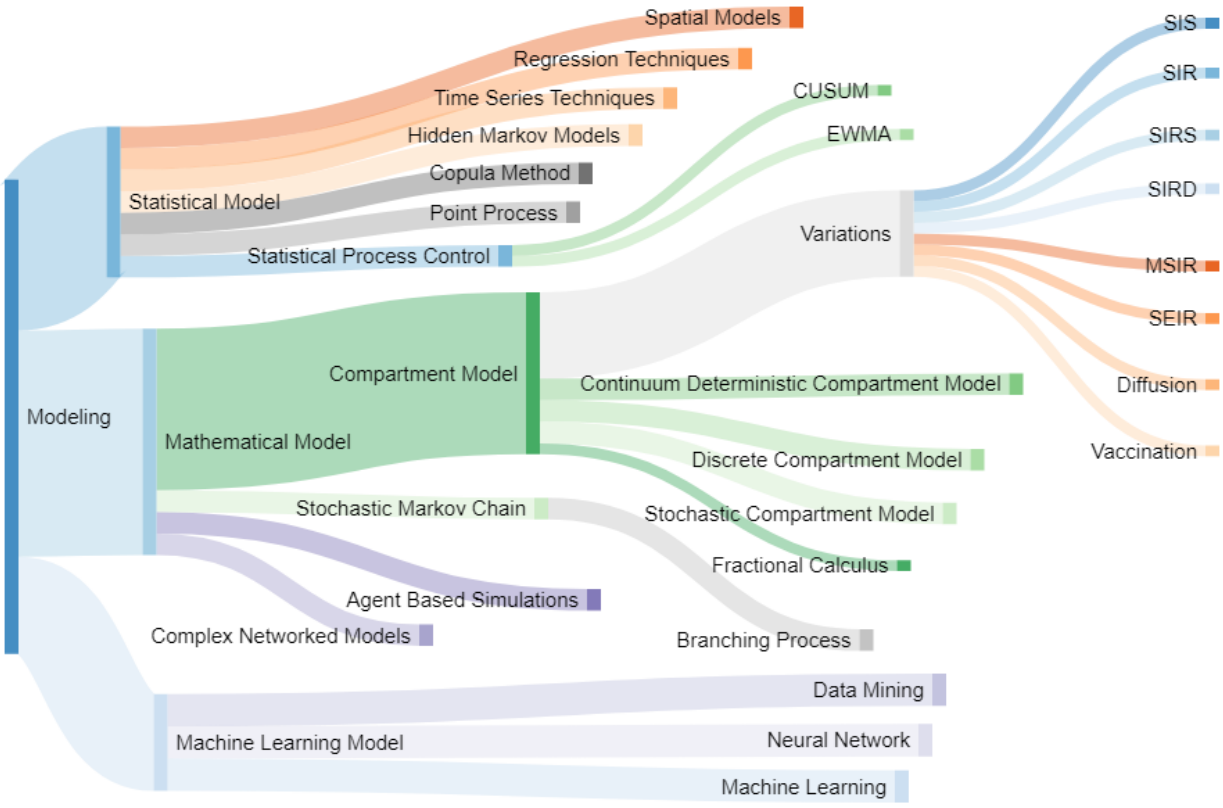
Since the outbreak of covid-19 last December(Wu 2020[1]), there have been over 6 million confirmed cases and 388 thousands of deaths(CSSE,2020[2]). Costs over $5 trillion and affects the life of every human being on earth in such a short time, just 5 months[3]. Scientists around the world have contributed 23,000 research papers on it (Google Scholar Search). Among them, mathematical modeling is used to predict the spread of the virus for policymakers to better allocate resources and plan interventions. SDS, 2020[4] summaries these approaches to SAR, Ebola, MERS and SARS-CoV-2 (i.e. Covid-19).

There are two kinds of approaches for the modelling: statistics modelling and compartment modelling. Statistical models are usually based on the historical data from earlier outbreak, to construct the statistical model, then to predict the spread of virus in the US and other countries. The Institute for Health Metrics and Evaluation (IHME) outlined the spread epidemic curve in China and Italy, and applied it to all the states of America and most other countries[5][6].

The other popular model is the compartmental model, which is based on conservation law. The rate of change equals the difference between the influx to and outflux from a compartment, which leads to the establishment of the differential equations. By solving and simulating the equations, we can understand how public interventions affect the epidemic, predict how the virus spreads, and know what the minimum vaccines needed in a given population[7].

Literature review

Governmens, research organizations, universities around the world have spent a huge amount of resources to develop models to understand and predict the spread of COVID-19 since pandemic outburst late 2019. Scientists from different fields, from infectious diseases and medicine biology to mathematics and computer science, have developed and researched various models. This has led to what is likely the fastest rate of scientific research in history. There are also quite a lot of review articles about models such as reviewing all the methods in compartment models. However, I haven't seen a thorough overview about all the methods across different disciplines including infectious disease medicine, pharmacokinetics, mathematics, computer simulation, artificial intelligence and machine learning. Therefore here I present and discuss the basics of these approaches about their concepts, principles, implementations, applications and limitations as the starting point of my research.



1) Statistical Model

a) Regression model

i) Linear regression model

The trend line is a typical linear regression model. It minimizes the sum of the squared deviations from the data to get the trend of the spread of COVID. My home state Delaware government website[10] provides daily trend lines for Current Hospitalizations, New Hospital Admissions (Confirmed and Suspected with COVID-19), Percent of persons testing positive, New Positive Cases, Tests Performed ……, this is the easiest and most widely seen statistical method used in numerical Covid-19 websites.

ii) Cyclic regression model

This more advanced model describes the seasonal behavior of diseases, and their tendencies to be more prominent during specific times during a cycle. Serfling 1963[9] first developed statistical analysis of excess pneumonia-influenza deaths based on the seasonal pattern. He combined a linear term describing secular trend with sine and cosine terms describing seasonal cyclic change to form an equation of the type for $\widehat{Y}$, the expected mean value to the total deaths in 4 weeks:

$$\widehat{Y} = u + bt + \sum a_i cos\theta + \sum b_i sin\theta \qquad [1]$$

in which θis a linear function of time t. Serfling[9] used a least square estimation of the parameters in equation (1) and derived this cyclic regression model for the pneumonia and influenza deaths:

$$\widehat{Y} = 300.5 + 2.1t + 97.6\,cos(2\pi t/13 - 2.67) \quad [2]$$

Nowadays CDC uses more square terms $(t^2)$to account long-term changes due to nonlinear factors such as population growth

iii) Double integration method for determining the period of a long term cyclic trend:

$$\widehat{Y} = \overline{Y} + Asin(\frac{2\pi t}{T}) \quad [3]$$

where $\widehat{Y}$is the observed value, $\overline{Y}$the mean value, A the amplitude, T the period, t the time. Let's integrate $\widehat{Y} - \overline{Y}$twice,

$$\int_0^T \int_0^T (\widehat{Y} - \overline{Y})dtdt = -\frac{AT^2}{4\pi^2} sin(\frac{2\pi t}{T}) \quad [4]$$

Therefore we can find out the pandemic period by subtracting the observed value with their mean value

Regression models, both linear and nonlinear, are very popular among epidemiologists for the prediction and surveillance of outbreaks of new emerging epidemics.

b) Time series analysis on AutoRegressive Integrated Moving Average Model (ARIMA) and seasonal ARIMA[11] [12]

Assume y(t) denotes a stationary stochastic process at time t with mean value μ,
the backward shift operator z^{-1} is the defined as $z^{-k}y(t) = y(t - k)$,
the differencing operator of order d: Δ^d is defined as $\Delta^d \equiv (1 - z^{-1})^d$,
the autoregressive operator $A(z^{-1}) = 1 + a_1 z^{-1} + \ldots\ldots + a_m z^{-m}$
the moving-average operator $B(z^{-1}) = 1 + b_1 z^{-1} + \ldots\ldots + b_n z^{-n}$
the residual (noise) at time t is e(t) which cannot be predicted from previous measurements. Then the ARIMA model is modelled by the equation:

$$A(z^{-1})[\Delta^d y(t) - \mu] = B(z^{-1})e(t)$$

[5]

Special cases:

1) For d=0 and m=0, it's the moving average model
2) For d=1 and m=n=0, it's the random walk with drift
3) For $\Delta_S^d \equiv (1 - z^{-1})^d (1 - z^{-k})^S$where k is the length of seasonal cycle and S is the degree of seasonal differencing: then it becomes the seasonal ARIMA model

Some parts of the time series are used as a training set, and the remaining data is used as a validation set. The goodness-of-fit model is used for forecasting disease evolution.

c) Statistical process control methods

i) Cumulative sum charts (CUSUM)

CUSUM[13] is a sequential analysis technique most commonly used technique for the detection of disease outbreaks. Assume at time t_i (i=1,2,...,n), the number of infected cases is $y(t_i)$, then CUSUM is:

$CUSUM(i) = \sum_{j=1}^{i} (y(t_i) - k)$ or in a recursive form as

$$CUSUM(0) = 0$$
$$CUSUM(i) = max(0, CUSUM(i-1) + y(t_i) - k),\ i \geq 0 \quad [6]$$

where k is a reference value to the difference between to the in-control and the out-of-control mean. The threshold h is 3 times the standard deviation from mean value of in-control observations

If $CUSUM(i) < h$, then the process is “in-control”.

When t=t_i $CUSUM(i) > h$, then the process is “out-of-control”.

We can calculate k

$$k = -\ log\frac{f(\theta_1|y(t_i))}{f(\theta_0|y(t_i))} \quad [7]$$

Here $f(\theta_0)$ and $f(\theta_1)$ are the probability functions of the in-control and out-of-control processes with parameters θ_0 and θ_1 respectively. They can be estimated using data from the past. For Poisson distributions, formula [7] is simplified as:

$$k = \frac{\mu_1 - \mu_0}{log(\mu_1) - log(\mu_0)} \quad [8]$$

ii) Exponentially weighted moving average (EWMA) [14] using the following recursive statistical estimator,

$$z(t_0) \equiv z(0) = 0$$
$$z(t_i) = (1-\gamma)z(t_{i-1}) + \gamma y(t_i)\ for\ i \geq 1 \quad [9]$$

where the constant smoothing coefficient$\gamma$is the degree of weighting decrease between 0 and 1. It is a “forgetting” factor for weighing the significance of past values. $y(t_i)$ is the value at time t_i,$z(t_i)$ is the value of the EMA at time t_i. THe recursive solution is

$$z(t_i) = \gamma[y(t_i) + (1-\gamma)y(t_{i-1}) + \ldots + (1-\gamma)^k y(t_{i-k})] + (1-\gamma)^{k+1} z(t_{i-k-1}) \quad [10]$$

d) Hidden Markov Models (HMM) - statistical correlation in time series

HMM is a statistical Markov model for the Markov process system[15]. HMM are known for their applications to many areas including infectious disease. When epidemic outbreaks like COVID-19, we can observe some possible indicators of the disease, but we cannot monitor and record explicitly the characteristics of the disease. HMM are then exploited under these limitations. We can use them to forecast the evolution of COVID-19 by monitoring the number of reported cases.

e) Spatial models - monitor, identify and forecast disease outbreaks in different locations

f) Copula methods

2) Mathematical Model

a) Compartment model

Compartment model[16] has a long history, as originated by Kermack and McKendrick in 1927.[17] It has been used widely in epidemics research and other areas, like pharmacokinetics and pharmacodynamics to study the movement of drugs through the

body and the body's biological response to drugs. It's a prime example of the application of conservation laws from physics and differential equations from mathematics to medical research.

i) Variations on model

(1) SIS model $S \leftrightarrow I$

This is the 2 compartments model, Susceptible and Infectious.

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma I$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

where S(t) is the susceptible population, I(t) is the infected population, β is the COVID-19 transmission rate, and γ is the recovery rate. N is the total population, and is a constant:

$$S(t) + I(t) = N, \frac{dS(t)}{dt} + \frac{dI(t)}{dt} = \frac{dN}{dt} = 0$$

Assume initial condition $I(0)=I_0$, The solution to the system of equations is a logistic function:

$$I(t) = \frac{(\beta-\gamma)N}{\beta+[(\beta-\gamma)(N/I_0)-\beta]e^{(\gamma-\beta)t}}$$

Introducing the basic reproduction number $R_0 = \frac{\beta}{\gamma}$, if

$$R_0 \leq 1 \Rightarrow \lim_{t\to\infty} I(t) = 0$$

$$R_0 > 1 \Rightarrow \lim_{t\to\infty} I(t) = (1 - R^{-1})N = I_\infty$$

I(t) can be rewritten as

$$I(t) = \frac{I_\infty}{1+(I_\infty/I_0-1)e^{(\gamma-\beta)t}}$$

In the special case assuming $\gamma = 0$, the SIS model degenerates to a SI model with a simple exponential logistic growth function.

(2) SIR model $S \to I \to R$

For covid-19, the most widely used compartmental model is the 3-compartment model, also known as the SIR model. S stands for the stock of the Susceptible population, I is the stock of Infected, and R is the stock of the Removed population (including both recovery and death). The conservation law requires that:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

(3) SIRS model $S \to \overleftarrow{I} \to R$

A recovered individual may become susceptible again after a period of time 1/α. Adding this possibility to the SIR model changes it into the SIRS model:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \alpha R$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I - \alpha R$$

When $dI/dt > 0$, $dS/dt < 0$the disease becomes an epidemic

(4) SIRD model $S \to I \to (R\ D)$

A new compartment, D, stands for the deceased population, and μ the mortality rate for the disease. Adding this compartment to the SIR model changes it into the SIRD model:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \mu D$$

$$\frac{dR}{dt} = \gamma I$$

$$\frac{dD}{dt} = \mu I$$

(5) MSIR model $M \to S \to I \to R$

A new compartment M represents those with passive immunity, and Λis the birth rate. Adding this compartment to the SIR model changes it into the MSIR model.

$$\frac{dM}{dt} = \Lambda - \delta M - \mu M$$

$$\frac{dS}{dt} = \delta M - \frac{\beta SI}{N} - \mu S$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \mu I$$

$$\frac{dR}{dt} = \gamma I - \mu R$$

(6) SEIR model $S \to E \to I \to R$

The Incubation period is very significant for COVID-19 in particular. Patients could be infected but not yet infectious. We need to add an *Exposed* compartment between Susceptible and Infectious, which represents those patients. Assuming the incubation period is a^{-1} and birth rate and death rate are the same at Nμ, the system of equations below is derived. Adding the exposed compartment to the SIR model changes it into the SEIR model.

$$\frac{dS}{dt} = \mu N - \mu S - \frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - (\mu + a)E$$

$$\frac{dI}{dt} = aE - (\gamma + \mu)I$$

$$\frac{dR}{dt} = \gamma I - \mu R$$

Now the basic reproduction number R_0 becomes $R_0=\frac{a}{\mu+a}\frac{\beta}{\mu+\gamma}$, if

$R_0 \leq 1 \Rightarrow \lim_{t\to\infty}(SEIT) = Disease\ Free\ Equilibrium$

$R_0 > 1 \Rightarrow \lim_{t\to\infty}(SEIT) = Endemic\ Equilibrium$

(7) SEIS model

(8) MSEIR model

(9) MSEIRS model

(10) Diffusion model

Infectious disease spreads not only over time, but also diffuses across space. In order to study this effect, we can add a diffusion term into the equations:

$$\partial t\ S = D_s\nabla^2 S - \beta SI/N$$

$$\partial t\ I = D_I\nabla^2 I + \beta SI/N - \gamma I$$

$$\partial t\ R = D_R\nabla^2 R + \gamma I$$

Where D_s, D_I, and D_R are diffusion constants.

(11) Vaccination model

Let V be the vaccinated population, P be the newborn vaccinated population, and ρthe adult vaccinated rate

$dS/dt = \mu N(1-P) - \mu S - \rho S - \beta SI/N$

$dI/dt = \beta SI/N - (\mu + \gamma)I$

$dV/dt = \mu NP + \rho S - \mu V$

Eradication condition becomes

$P \geq 1 - (1 + \rho/\mu)/R_0$

ii) Continuous model vs Discrete model

(1) Continuous model

In the real world, infected subjects are discrete individuals, and therefore continuum models just describe the coarse grained dynamics of the epidemics in the population. Therefore, we can use ODE/PDE (ordinary or partial differential equations), which are powerful mathematical tools for the evolution of the disease as a function of parameters, such as age, quarantine, etc.

(2) Discrete model

However, the spread of COVID-19 is very complicated, discrete and stochastic. We have to use a more realistic model to understand more details. For discrete time equivalent, rewrite the SIR model as:

$S_{t+1} = S_t - \beta S_t I_t/N_t$

$$I_{t+1} = I_t + \beta S_t I_t / N_t - \gamma I_t$$
$$R_{t+1} = R_t + \gamma I_t$$

iii) Deterministic model vs Stochastic model

(1) Deterministic model

(2) Stochastic model

The discrete model above remains deterministic: for given values of the rates β and γ, dynamics will be fixed. It is fairly straightforward to convert this discrete model into a stochastic one by adding appropriate probability distributions to the model. There are at least 3 types of such distributions which will be useful to consider

(a) Binomial distribution

A deterministic model per susceptible rate $\beta I_t / N_t$, where the probability for an individual to move from S to I at time t is $p(S \to I)_t = 1 - e^{-\beta I_t / N_t}$

(b) Poisson distribution

Assume the imported rate of new infectious cases is ϵ

$$I_{t+1} = I_t + \beta S_t I_t / N_t - \gamma I_t + \epsilon$$

(c) Multinomial distribution

iv) Fractional Calculus

Normal integer calculus fails in model fitting to the real world. Rates of movement from one compartment to another aren't always proportional to the compartment population. The more advanced fractional calculus has been widely used in science and has rewritten physics, chemistry, engineering, pharmacokinetics, and infectious disease research. Fractional calculus involves replacing the traditional calculus $\frac{d^n}{dt^n}$(where n must be an integer) to $D_t^\alpha$where α can be a complex number:

$$Re(\alpha) > 0 \Rightarrow D_t^\alpha f(t) = \frac{d^\alpha f(t)}{dt^\alpha}$$

$$Re(\alpha) < 0 \Rightarrow D_t^\alpha f(t) = \frac{1}{\Gamma(-\alpha)} \int^t (t - \tau)^{-\alpha-1} f(\tau) d\tau$$

For example, $\frac{d^{\frac{1}{2}}}{dt^{\frac{1}{2}}} t = \frac{2}{\sqrt{\pi}} t^{\frac{1}{2}}$, still

$$\frac{d}{dt} t = \frac{d^{\frac{1}{2}}}{dt^{\frac{1}{2}}} \left(\frac{d^{\frac{1}{2}}}{dt^{\frac{1}{2}}} t \right) = \frac{d^{\frac{1}{2}}}{dt^{\frac{1}{2}}} \left(\frac{2}{\sqrt{\pi}} t^{\frac{1}{2}} \right) = 1$$

Then the SIR model changes to:

$$D_t^\alpha S(t) = -\beta SI/N$$
$$D_t^\alpha I(t) = \beta SI/N - \gamma I$$

$$D_t^{\alpha} R(t) = \gamma I$$

b) Agent Based Simulations

c) Complex Networked Models

3) Machine Learning Model

a) Data Mining

b) Machine Learning

1. Stochastic Branching Model

Hellewell et al 2020[8] first developed a stochastic branching model, parameterised to the COVID-19 outbreak, and used the model to quantify the potential effectiveness of contact tracing and isolation of cases at controlling COVID-19.

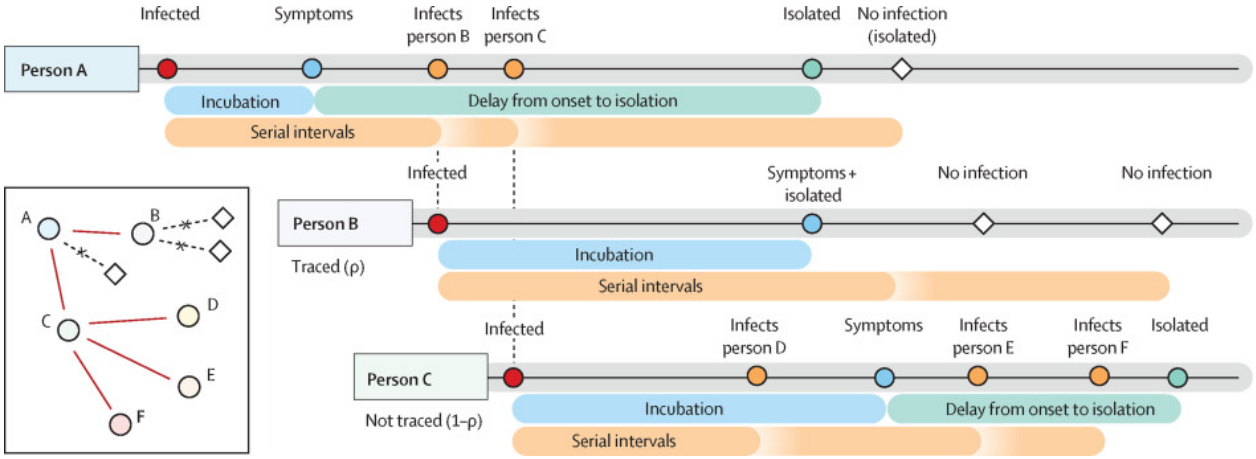


Fig 1: Stochastic Branching Model

Their key findings are (1) higher R0 with a higher percentage of contacts that had to be traced; (2) The probability of controlling an outbreak decreased with initial cases count increased, which in turn increases R0, and causes more transmission before symptom onset. They drew the conclusion that highly effective contact tracing and case isolation is enough to control COVID-19. There are many strong assumptions, such as (1) isolation reduces spread completely which is not necessarily true; (2) outbreak will stop within 12 to 16 weeks from the initial case due to no new infections, which also is not true.

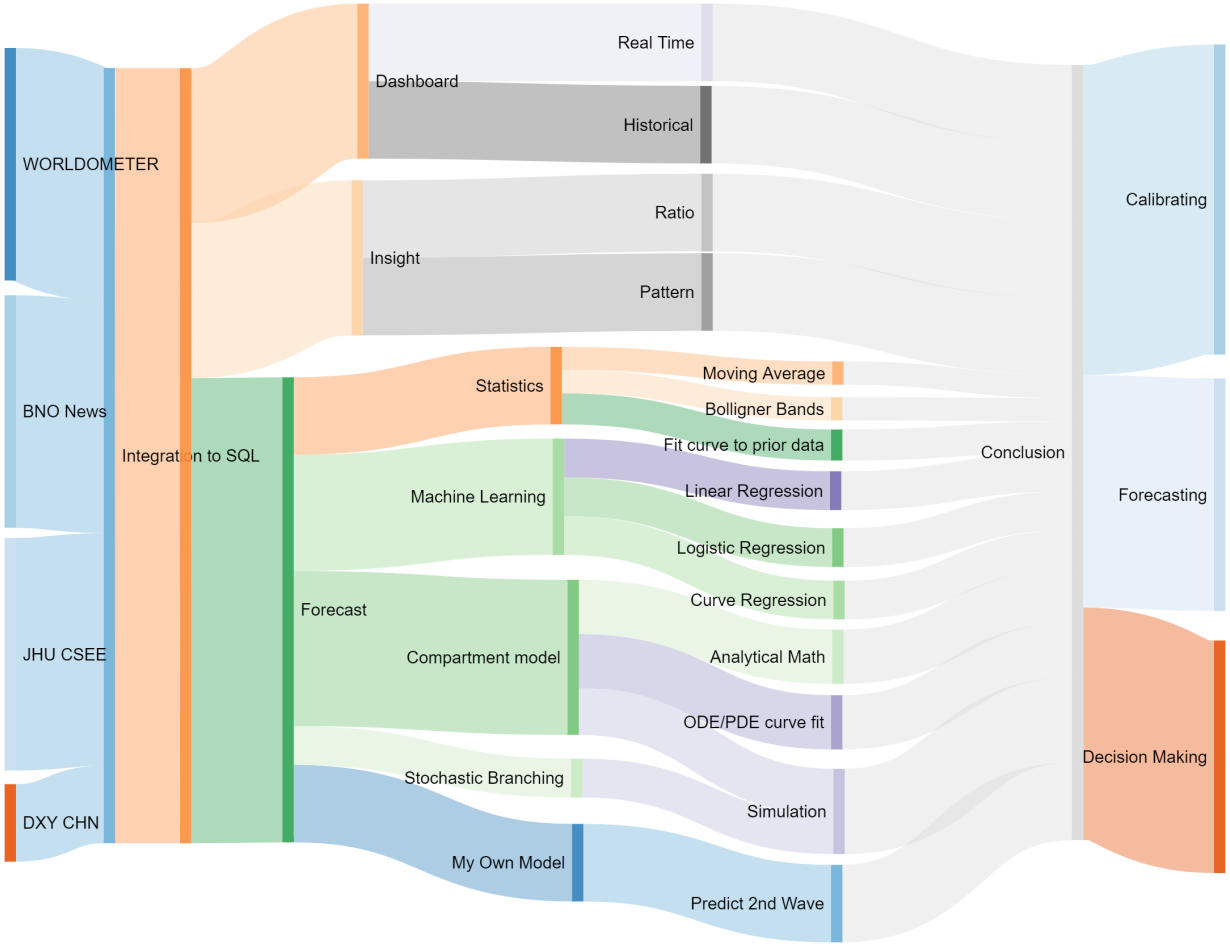
My work

In my research, I use a basic SIR model and calibrate it to match the real observed trends of COVID in various countries and regions. By solving for the parameters of COVID spread, we can use those parameters to analyze what caused the differing trends of coronavirus in different countries, and forecast what will occur in the future by extending our model.

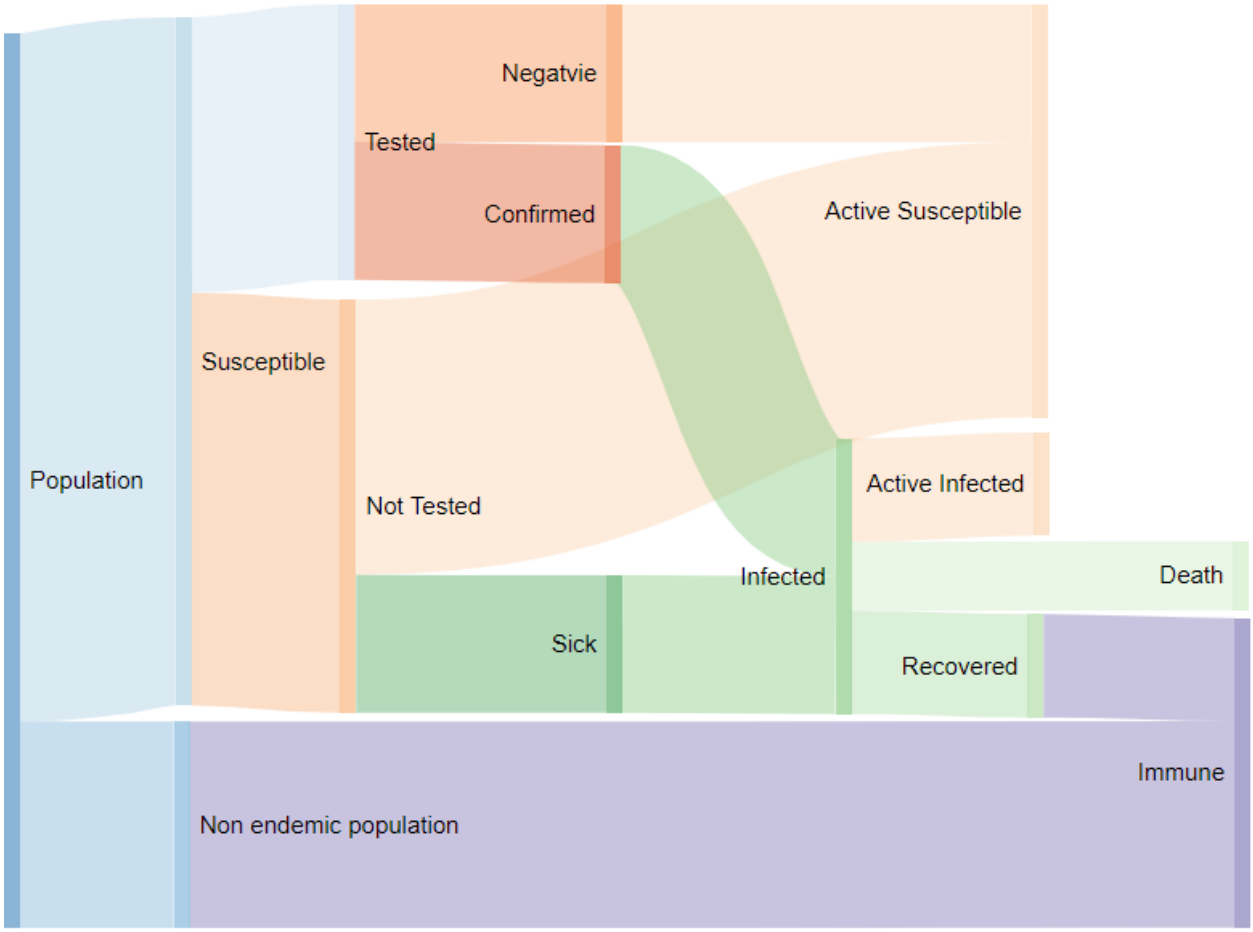
Main

1. My New model

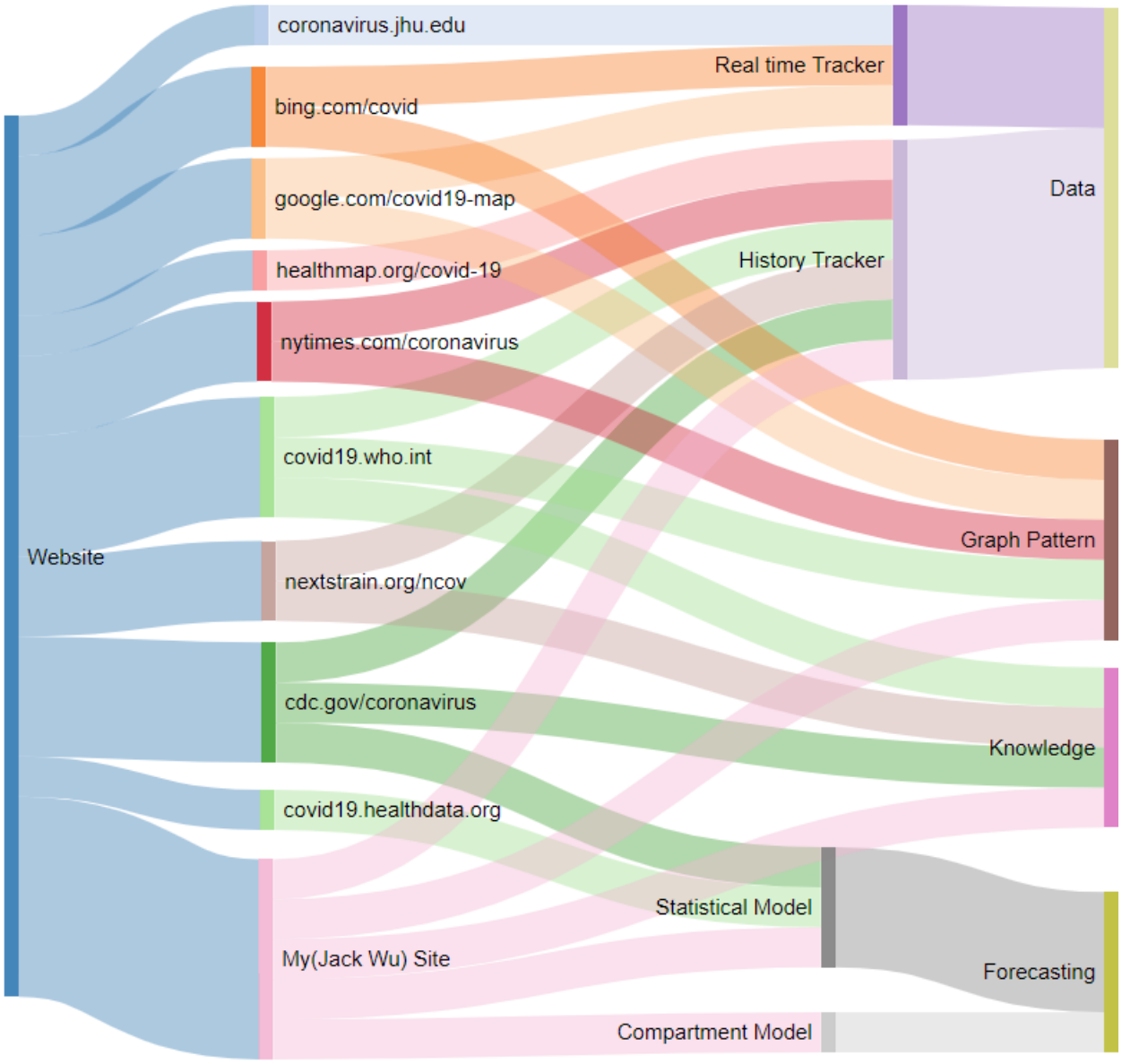
My project workflow, from data collection to models, methods, findings, to conclusions, and applications as the following figure:

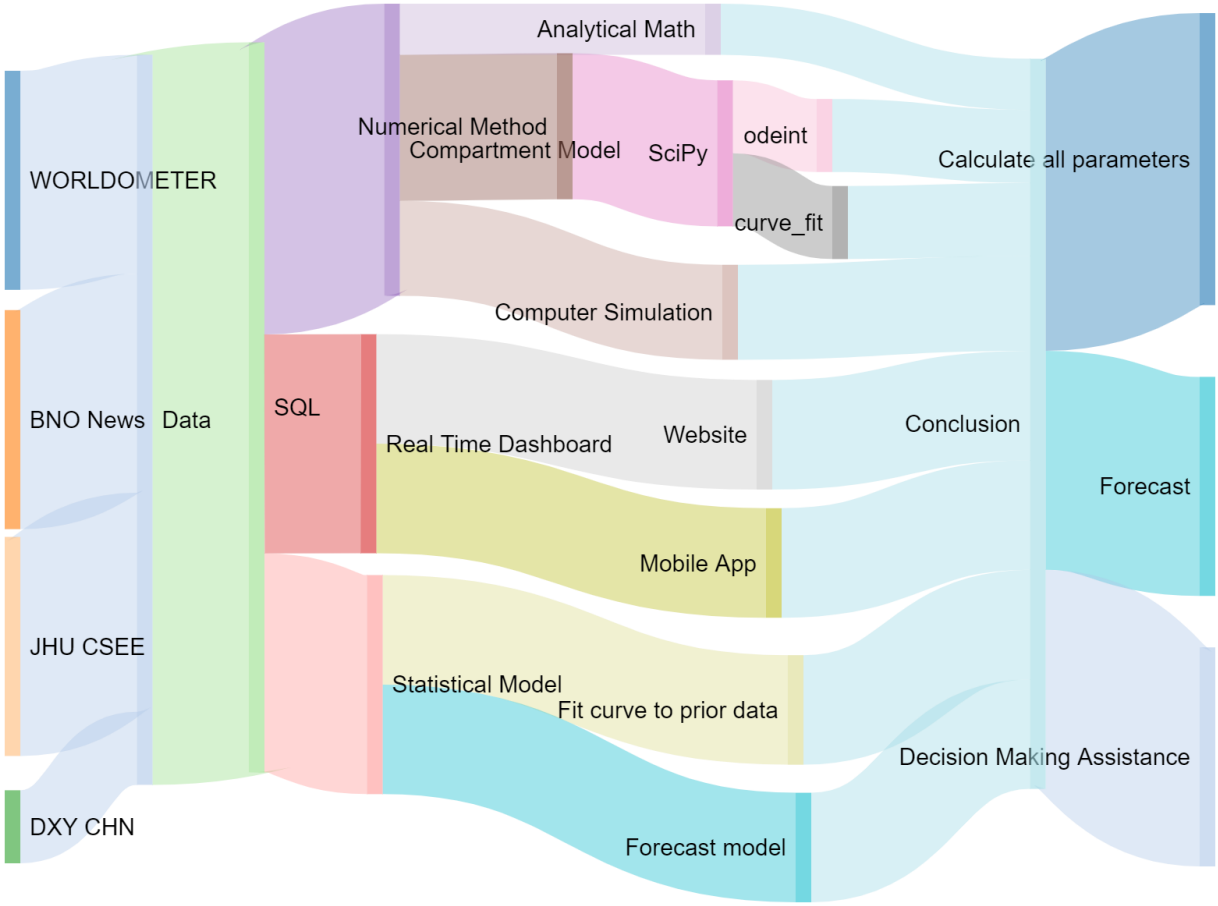


My compartment model:



I have published all the analysis to my website: http://jackwu.us , which is the only website that provides real time compartment modeling for every country/state currently available:


Website
coronavirus.jhu.edu
bing.com/covid
google.com/covid19-map
healthmap.org/covid-19
nytimes.com/coronavirus
covid19.who.int
nextstrain.org/ncov
cdc.gov/coronavirus
covid19.healthdata.org
My(Jack Wu) Site
Real time Tracker
History Tracker
Statistical Model
Compartment Model
Data
Graph Pattern
Knowledge
Forecasting



2. Mathematics

This paper uses the classic SIR model to model COVID-19, with three compartments: Susceptible, Infected, and Removed. The susceptible compartment includes people who do not have the disease, but are capable of catching it. The infected compartment includes those who actively have the disease. The Removed compartment includes both people who have Recovered, and are thus immune due to immunologies, and are dead, and are Removed since they cannot contribute to the spread of the disease anymore.

There are two ways for individuals to change compartments; they can move from Susceptible to Infected through transmission (catch the virus), or move from Infected to Removed through recovery (recover or die from the virus). We make the same assumptions about virus spread as in previous SIR models: that transmission is jointly proportional to the populations of the susceptible and infected compartments, and that recovery is jointly proportional to the population of the infected compartment. This is represented by the following set of differential equations:

$$D_t^{\alpha} S(t) = -\beta SI/N$$

$$D_t^{\alpha} I(t) = \beta SI/N - \gamma I$$

$$D_t^{\alpha} R(t) = \gamma I$$

One of the most straightforward ways to forecast a pandemic is to use known data to create a best fit model. This model has certain parameters that need to be calculated by a best-fit model.

There are many different models that can be used for a best fit. The simplest model is a linear regression, or a trendline. A linear regression with equation mx + b can be fit to some given data. A linear trendline can be useful for predicting numbers over a short period of time, but is ineffective for long periods of time because the overall data trend is not linear.

Our model is an expanded version of a simple linear regression. To better model the entire trajectory of the pandemic, we add two additional terms to our model equation. Using these two terms, we are able to model: a purely exponential growth, a logistic growth, a "bell" curve, and a "second wave" curve, as seen in the below picture.

Here we analyze the SIR model and solve its differential equations analytically.

Using the expressions for dS/dt and dI/dt, we derive

dI/dS = -1 + α/βS

$$\int_{I0}^{I} dI = -\int_{S0}^{S} dS + (\alpha/\beta) \int_{S0}^{S} 1/S \, ds$$

$I - I_0 = (S_0 - S) + (\alpha/\beta) * (\ln S - \ln S_0)$

3. Data collection, extract, transform, load, orchestration, normalization

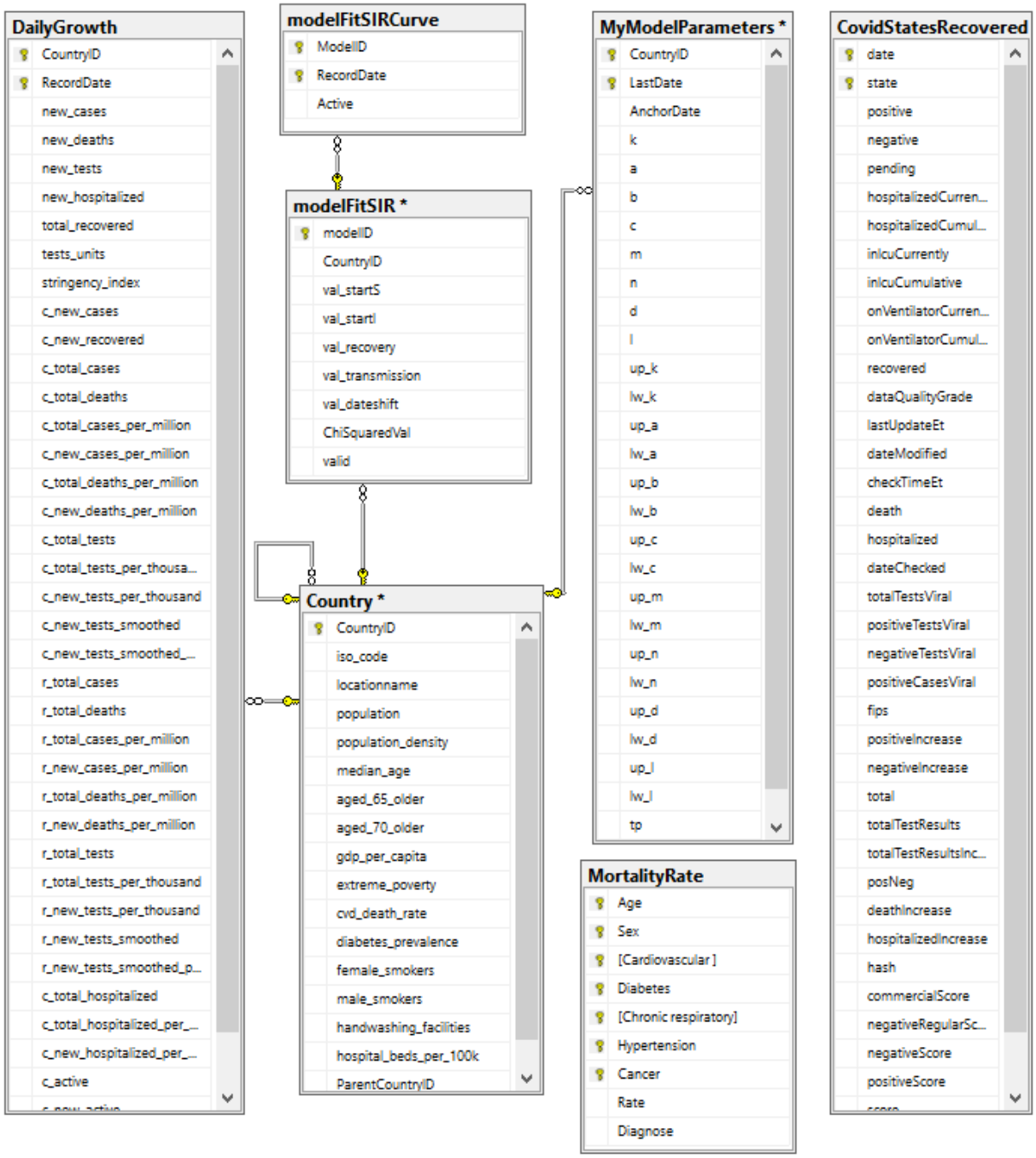
The data used in this study came from primarily four sources. The first resource is from the publication *Our World in Data*[7],(https://ourworldindata.org/coronavirus-source-data) which had data concerning confirmed cases, deaths, tests, and specifics about each country; however, it did not have a count for the number of those who had recovered from COVID, making it impossible to determine the number of people infected at any one time. To fill in this data for our model, we used a second source, from *The Humanitarian Data Exchange, (https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases)* to get the number of recovered individuals and then calculate infectives. The third source I used in this project is JHU CSSE data (https://github.com/CSSEGISandData/COVID-19) which are aggregated from tens of different sources for US states data. The fourth source is https://ncov.dxy.cn/ncovh5/view/pneumonia, which contains China Covid 19 data.

The reasons for collecting multiple resources include the presence of data for more areas/countries/states in the world, data for different time intervals, and the inclusion of more fields and attributes, for my

research needs. However, because the data is coming from four different sources, there were several issues as following:

1. Different file formats: mostly, they are 5 types of data formats:
    a. .csv (comma separated values file)
    b. .xls or .xlsx (Microsoft Excel file)
    c. Html file from web page
    d. Web service/API interface files
        i. Xml file
        ii. JSON
2. Redundant data with repetition of data
3. Missing data
    a. The Humanitarian Data Exchange Data lacked any records from the United States of America, so the number infected at any time could not be calculated for the US.
4. Discrepancies that harmed the accuracy of the data and caused erroneous results.
    a. The recovered count (from the Humanitarian Data Exchange) exceeded the total cases count (from Our World in Data) for the country of New Zealand, causing a negative number of active infectives.

In order to solve these issues, I used Microsoft SSIS (SQL Server Integration Services) tool for ETL (extract, transform, load) and orchestration. I built a SQL relational database to store and process the data. I established all primary keys for each table, and foreign keys for referential integrity. Here is the collected data that was loaded into a MS SQL database.



Total data size is 208 Mb. All the processed data, have been posted to my web site at:

http://www.jackwu.us/Data

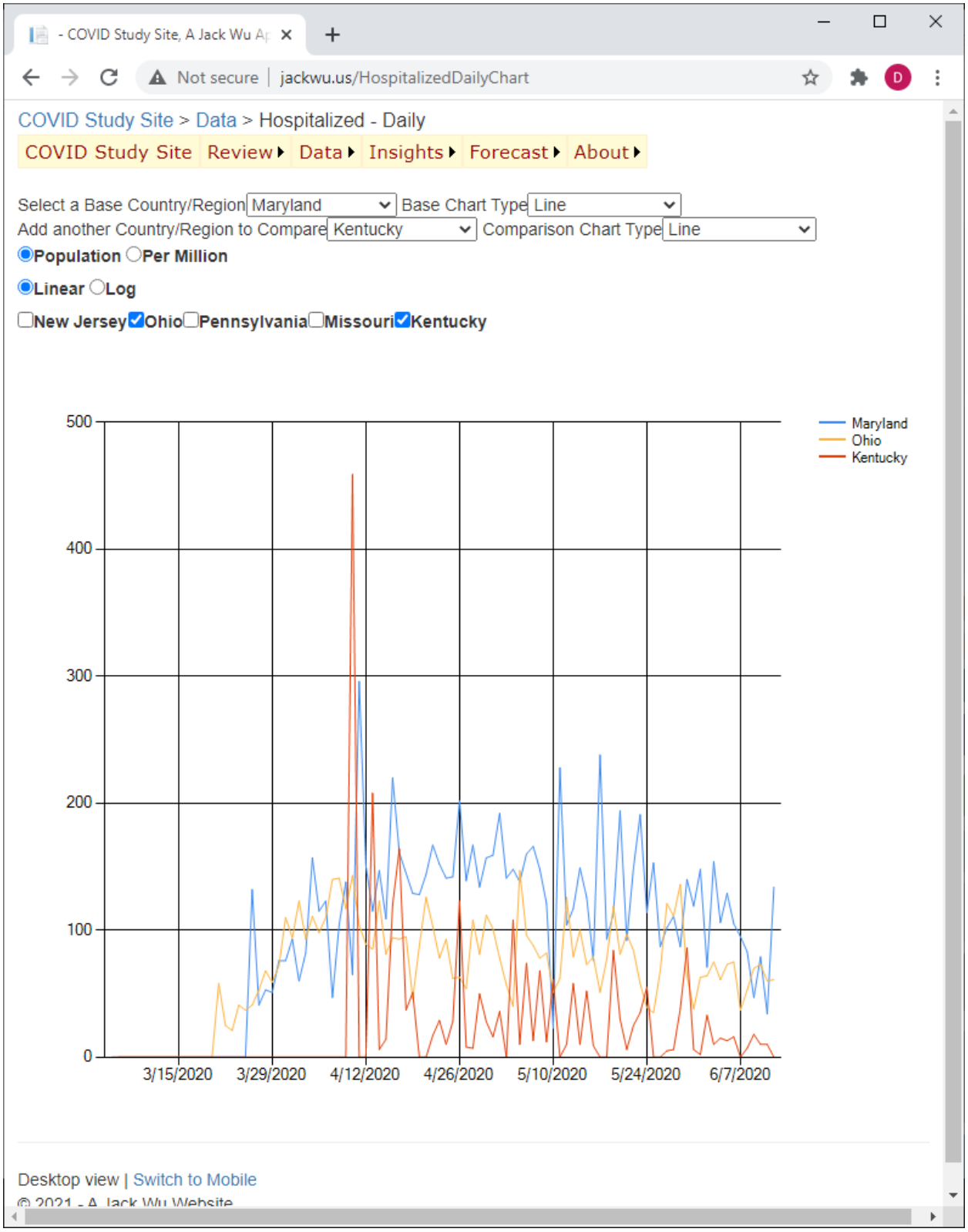
Both table views and graph views are available. For example, There are 280 countrie/regions data in table view at

http://www.jackwu.us/TableView

which can browse to 14 pages, and each country is selectable to view daily records - here's a screenshot:


- COVID Study Site, A Jack Wu Ap
Not secure | jackwu.us/TableView
COVID Study Site > Data > Table View
COVID Study Site
Review
Data
Insights
Forecast
About
Latest Data
Location | Total Cases | Total Deaths | Total Recoveries | Active | Total Tests
Select Afghanistan 13036 235 1259 11542
Select Albania 1076 33 851 192
Select Algeria 8997 630 5422 2945
Select Andorra 763 51 684 28
Select Angola 73 4 18 51
Select Anguilla 3 0
Select Antigua and Barbuda 25 3 19 3
Select Argentina 14689 508 4788 9393 107327
Select Armenia 8216 113 3297 4806
Select Aruba 101 3
Select Australia 7150 103 104 6943 1070290
Select Austria 16543 668 15347 528 415224
Select Azerbaijan 4759 56 3125 1578
Select Bahamas 101 11 48 42
Select Bahrain 10052 15 5700 4337 249666
Select Bangladesh 40321 559 9015 30747 275659
Select Barbados 92 7 76 9
Select Belarus 39858 219 17390 22249 245960
Select Belgium 57849 9388 15682 32779 658513
Select Belize 18 2 16 0
12345678910...
History for Australia
Date | Total Cases | Total Deaths | Total Recoveries | Active | Total Tests
5/29/2020 7150 103 104 6943 1070290
5/28/2020 7139 103 104 6932 1070290
5/27/2020 7133 102 104 6927 1041167
5/26/2020 7118 102 104 6912 991069
5/25/2020 7109 102 104 6903 947046
5/24/2020 7106 102 104 6900 927641
5/23/2020 7095 101 104 6890 895108
5/22/2020 7081 100 104 6877 873528
5/21/2020 7079 100 104 6875 840530
5/20/2020 7068 99 104 6865 814413
5/19/2020 7060 99 104 6857 788716
5/18/2020 7045 98 104 6843 764880
5/17/2020 7036 98 104 6834 744972
5/16/2020 7019 98 104 6817 718498
5/15/2020 6989 98 104 6787 686662
12345678910

There are 12 different graph views of raw data, each graphing a different statistic. Here's a screenshot for Accumulated Hospitalized count, which compares between an unlimited number of states and countries.



Following is the flow chart for the processing of data:

4. Calibration

There are 5 parameters needed to produce an SIR model and graph. They are: (1) the starting value of S, or susceptibles, (2) the starting value of I, or infectives, (3) the coefficient for the rate of transmission, β, (4) the coefficient for the rate of recovery, α, and (5) the "time shift", or the real-world date at which the SIR simulation should begin. These values were calibrated for many countries, with the goal of finding a best-fit. The best fit was defined to be the set of parameters that minimizes the sum of the squares of errors (differences between our mode simulation and real observed data).

Python and various libraries such as scipy, numpy, pandas, and matplotlib were used to simulate an infection using the SIR model, read information from SQL, find the best-fit parameters, and display data using a graph. The best fit was calculated using the scipy.optimize.curve_fit function.

The initial guesses were calculated in certain groups. Data from countries were split into one of four categories:

-Growth, a continuous increase in cases
-Logistic, a growth followed by a relatively flat plateau
-Bell, a growth followed by a drop, similar to a graph created by an SIR model
-Second Wave, a growth followed by a drop, and then another growth

The best fit model for the countries in each type was calculated using a similar procedure. First, we manually fit the curve for a country that has an ideal graph; smooth and easy to fit using simple experimentation and manual tinkering. Once a satisfactory initial guess is created for the base country, that initial guess is copied over to all other countries of the same type, and then scaled based on the peak active count in each country compared to the peak active count in the base country. All of the initial guesses for each country is run through the scipy.optimize.curve_fit function to find the local best fit.

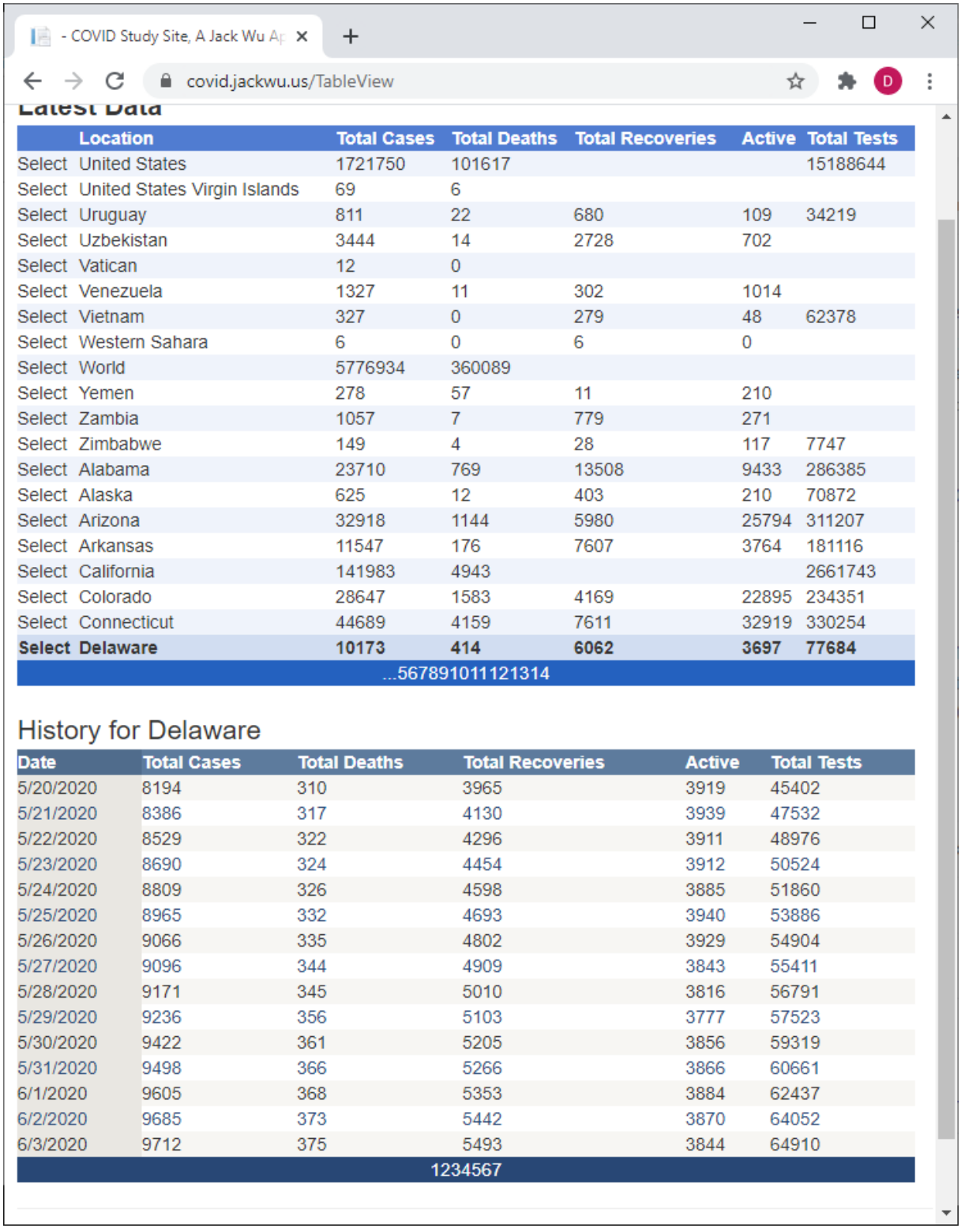
5. Data fitting and numerical results

5.1 Tracking Daily Records

5.1.1 Table view

I have extracted, transformed and loaded COVID-19 data using web services from different sites. The challenges associated with the process of ETL include: (1) The management of huge amounts of data, with my database size growing to gigabyte size. I have learned partition techniques to speed up my database performance. (2) The data are from different sources, with different structures and different schemas. Certain sources, like John Hopkins University, have even changed their own data schema (JHU has changed 4 times since I loaded their data). So far,

I have collected COVID-19 data for all countries and also all states in the United States. A table view can be seen by visiting my web site at https://covid.jackwu.us/TableView. Below is a screenshot for Delaware.



Latest Data

	Location	Total Cases	Total Deaths	Total Recoveries	Active	Total Tests
Select	United States	1721750	101617			15188644
Select	United States Virgin Islands	69	6			
Select	Uruguay	811	22	680	109	34219
Select	Uzbekistan	3444	14	2728	702	
Select	Vatican	12	0			
Select	Venezuela	1327	11	302	1014	
Select	Vietnam	327	0	279	48	62378
Select	Western Sahara	6	0	6	0	
Select	World	5776934	360089			
Select	Yemen	278	57	11	210	
Select	Zambia	1057	7	779	271	
Select	Zimbabwe	149	4	28	117	7747
Select	Alabama	23710	769	13508	9433	286385
Select	Alaska	625	12	403	210	70872
Select	Arizona	32918	1144	5980	25794	311207
Select	Arkansas	11547	176	7607	3764	181116
Select	California	141983	4943			2661743
Select	Colorado	28647	1583	4169	22895	234351
Select	Connecticut	44689	4159	7611	32919	330254
Select	**Delaware**	**10173**	**414**	**6062**	**3697**	**77684**

...567891011121314

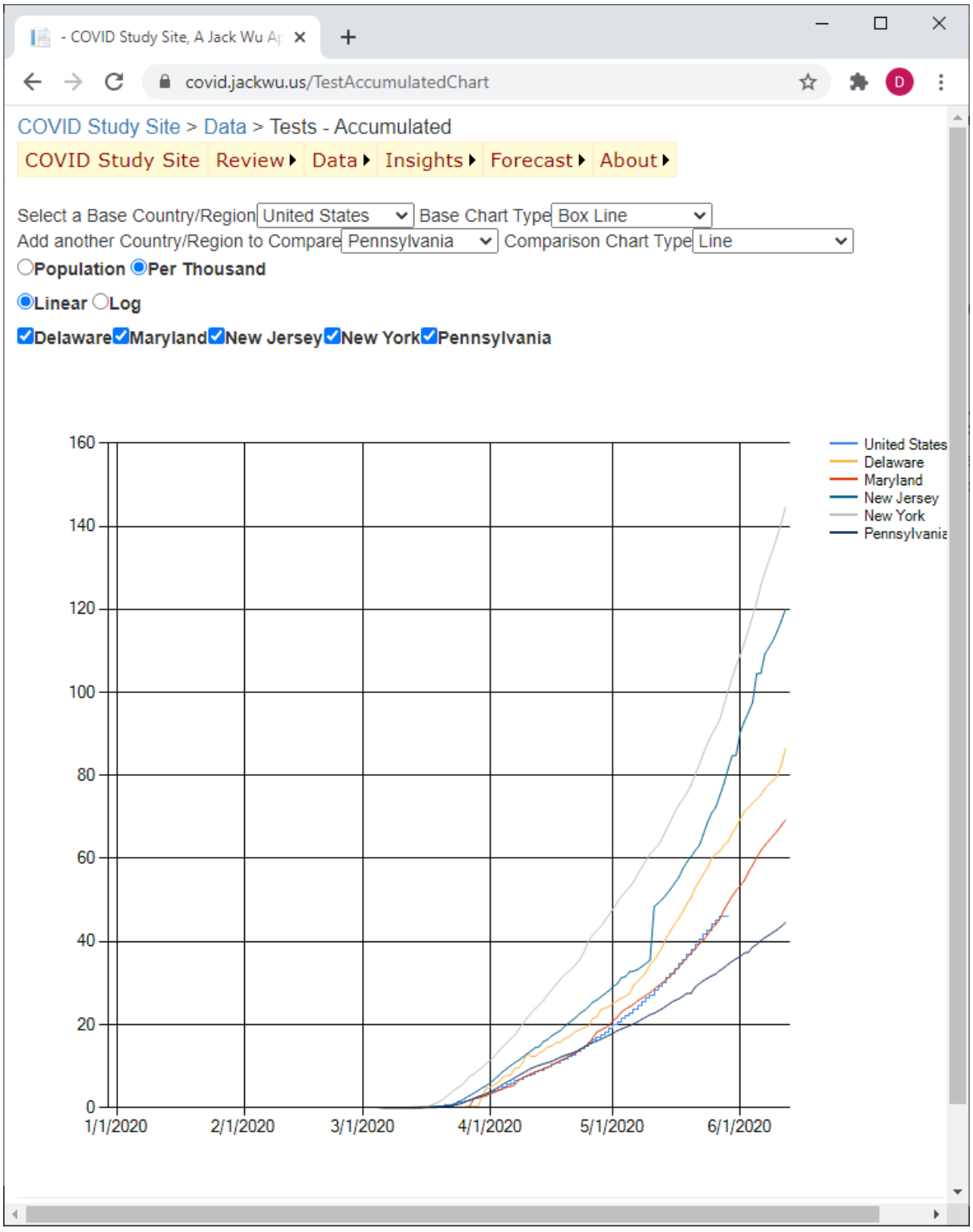
History for Delaware

Date	Total Cases	Total Deaths	Total Recoveries	Active	Total Tests
5/20/2020	8194	310	3965	3919	45402
5/21/2020	8386	317	4130	3939	47532
5/22/2020	8529	322	4296	3911	48976
5/23/2020	8690	324	4454	3912	50524
5/24/2020	8809	326	4598	3885	51860
5/25/2020	8965	332	4693	3940	53886
5/26/2020	9066	335	4802	3929	54904
5/27/2020	9096	344	4909	3843	55411
5/28/2020	9171	345	5010	3816	56791
5/29/2020	9236	356	5103	3777	57523
5/30/2020	9422	361	5205	3856	59319
5/31/2020	9498	366	5266	3866	60661
6/1/2020	9605	368	5353	3884	62437
6/2/2020	9685	373	5442	3870	64052
6/3/2020	9712	375	5493	3844	64910

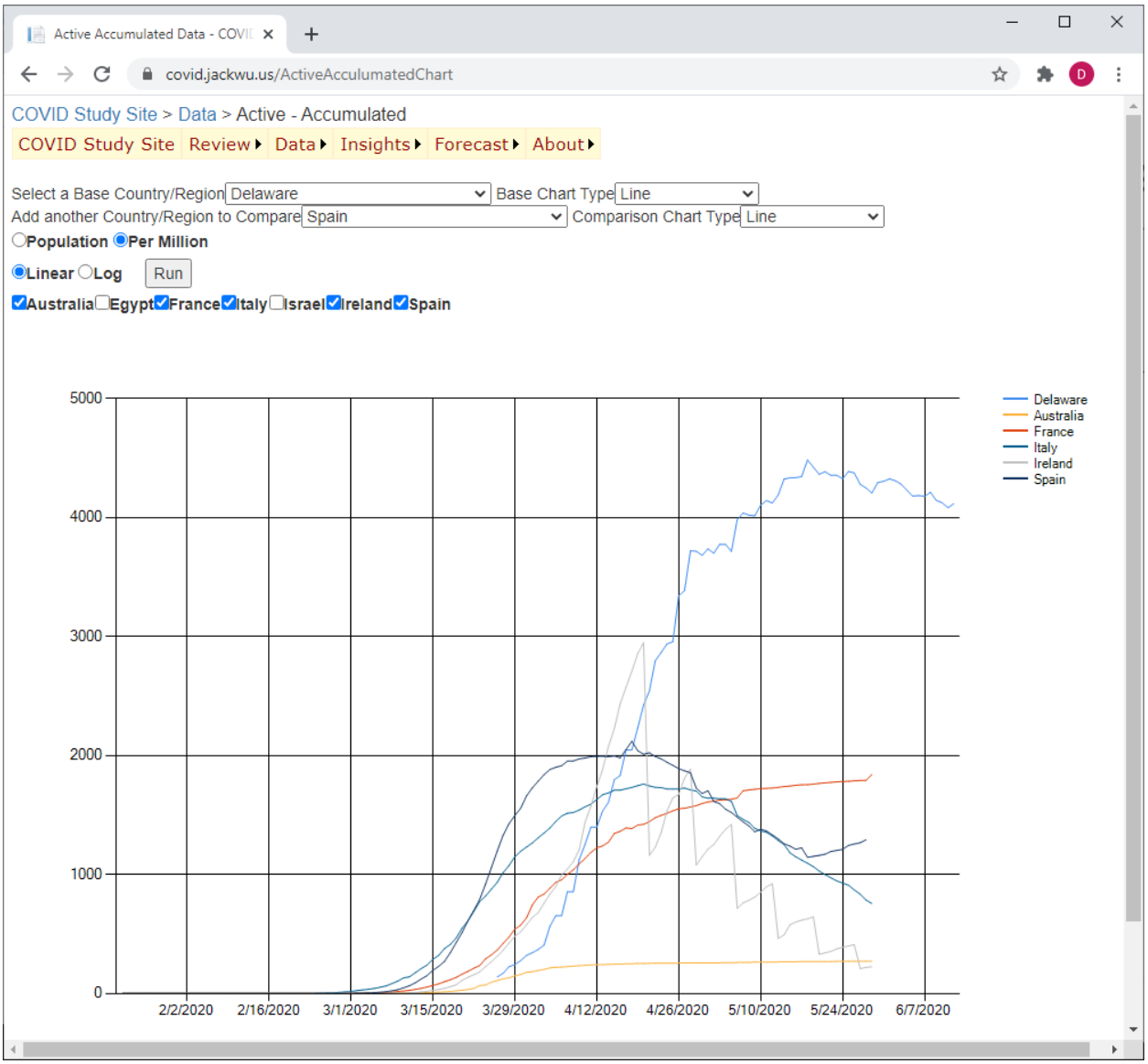
1234567

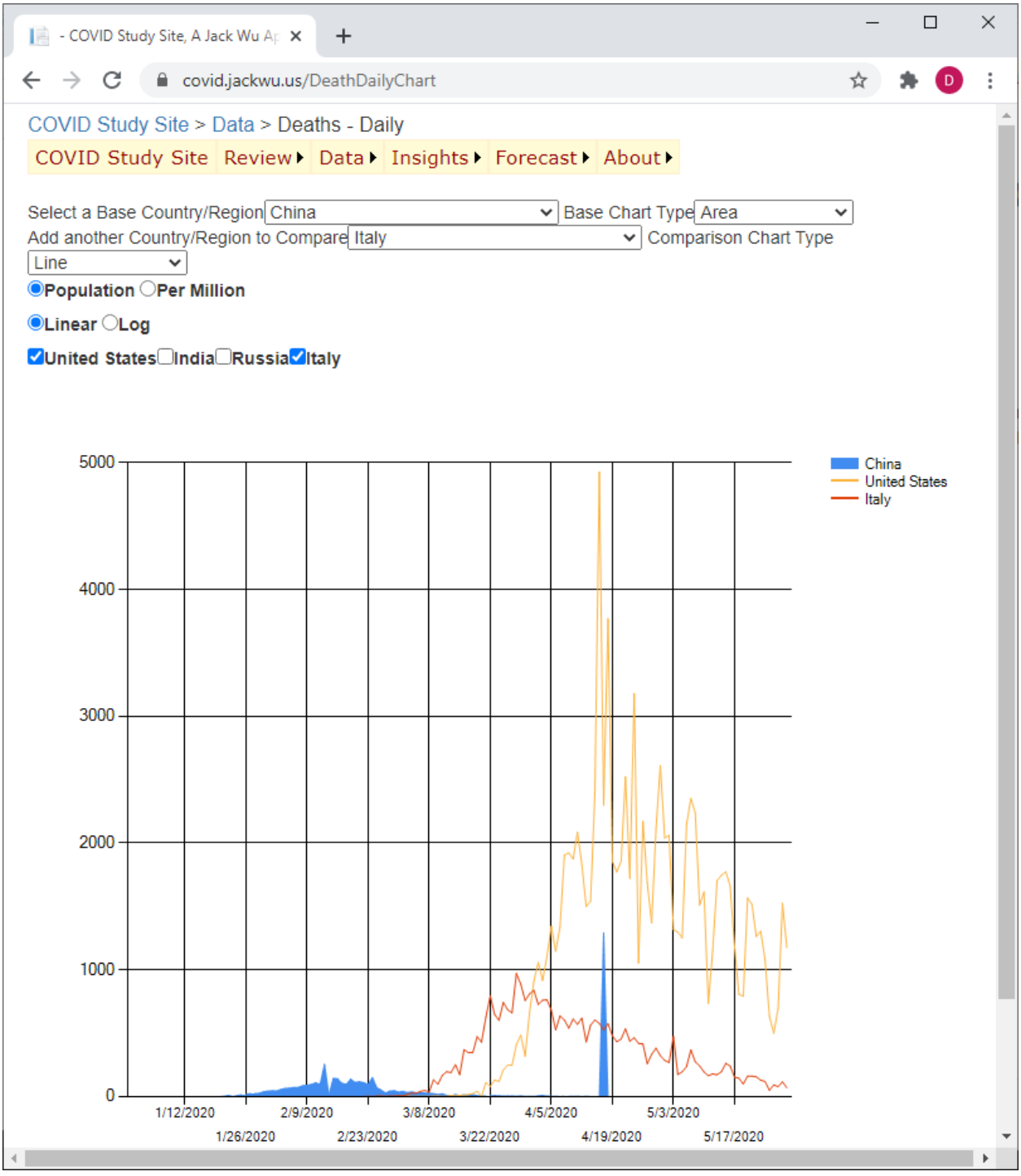
5.1.2 Graph view

Real time tracking all data related to COVID-19 in all countries and states in graph view (more than 10 webpages similar to https://covid.jackwu.us/ActiveDailyChart):



My web site is probably the only web site on the internet which can provide unlimited comparison between any country and state together on the graph. My web site is also probably the only website can compare all of the following: (1) test cases daily (2) test accumulated (3) confirmed cases daily (4) confirmed cases accumulated (5) recovered daily (6) recovered accumulated (7) death - daily (8) death accumulated (9) active infected daily (10) active infected accumulated (11) hospitalized daily (12) hospitalized accumulated, across the world, on total population, or per million, and either in linear or logarithmic y-axis scale, using horizontal lines, vertical lines, box lines, dots, dashes, columns, bars(horizontal columns), stacked columns and areas.

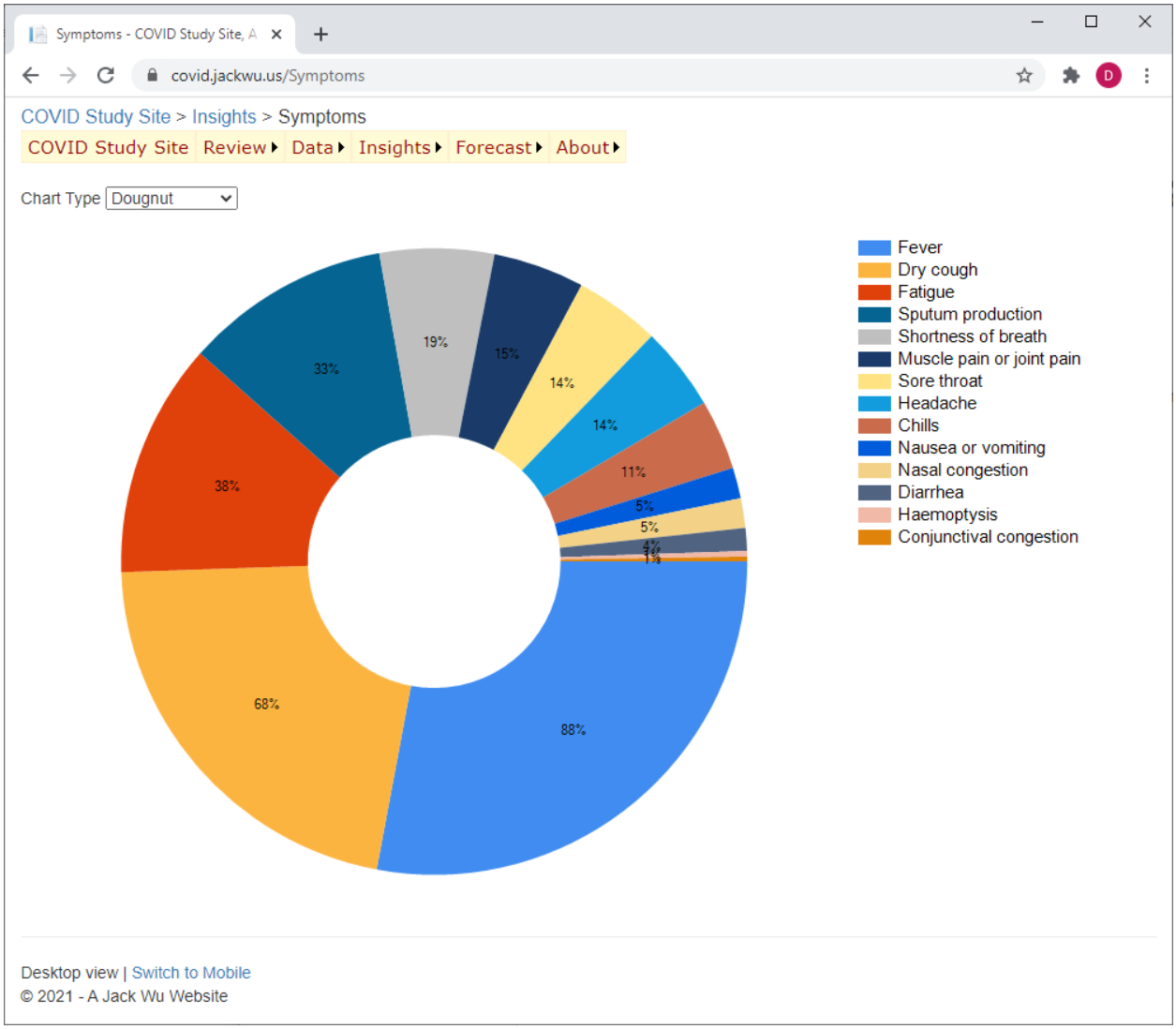




5.2 Insights from Data

5.2.1 Relation between Infected and Symptoms

By aggregating all the cases with symptoms, I found out that Fever and dry cough are the most common symptoms for COVID-19. https://covid.jackwu.us/Symptoms

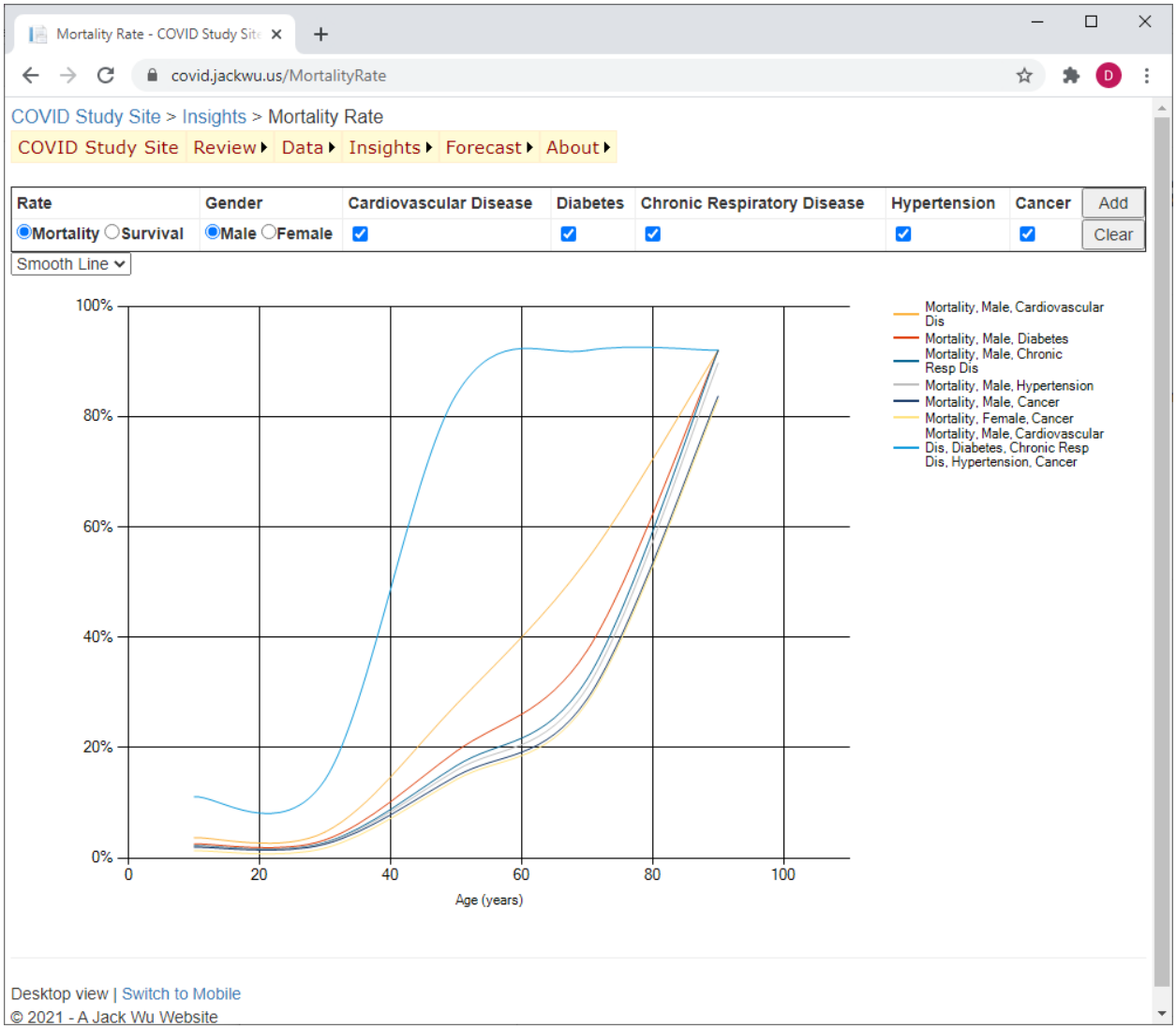


5.2.2 Mortality Rates between Existing Diseases, Age, and Gender

I collected data for Cardiovascular Disease, Diabetes, Chronic Respiratory Disease, Hypertension and Cancer. https://covid.jackwu.us/MortalityRate I found:

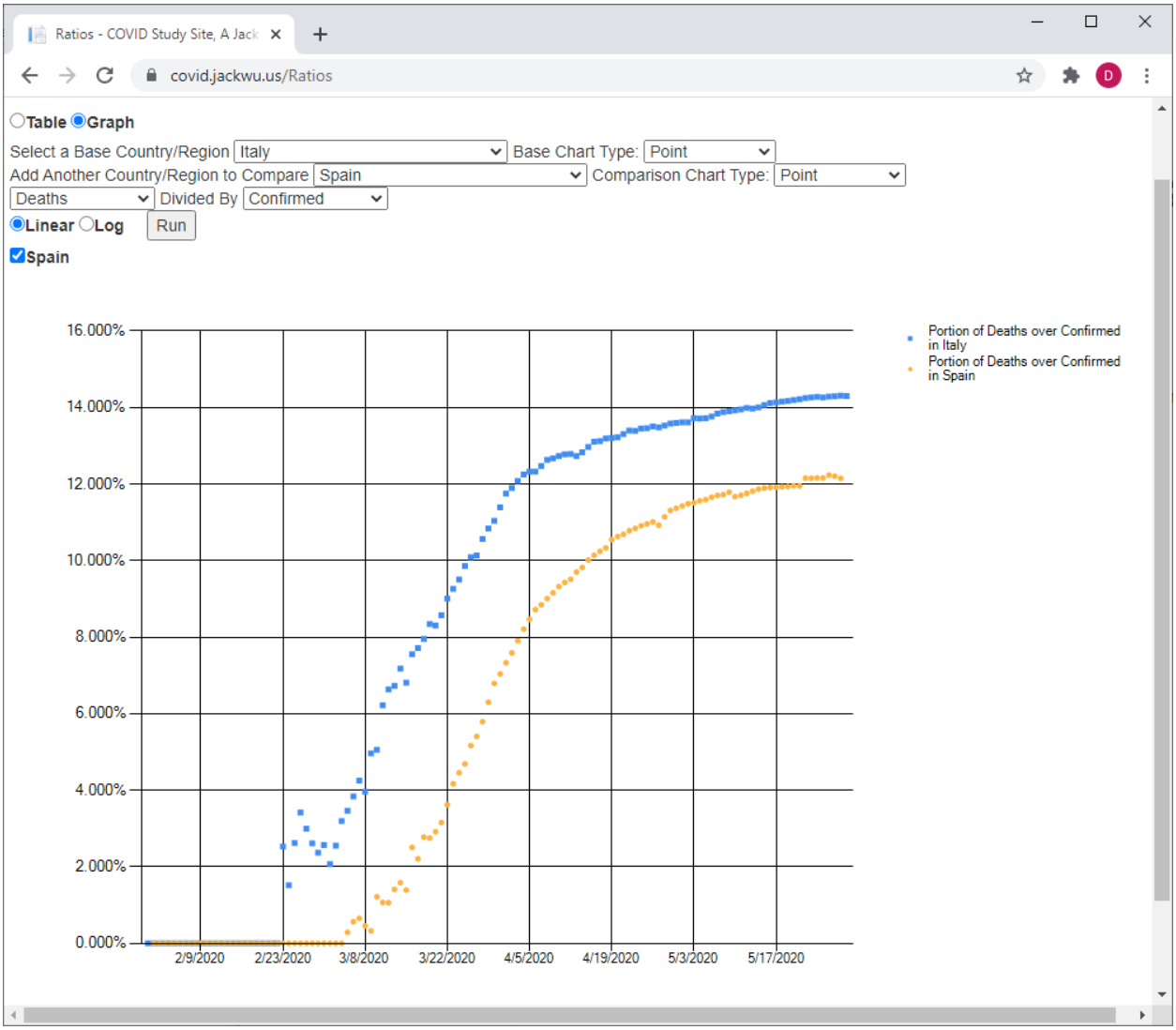
(1) For single existing condition, the mortality rates grow linearly along the age, from 0% to 95%

(2) For multiple existing conditions, the mortality rates jump to 95% at the age of 45 years old

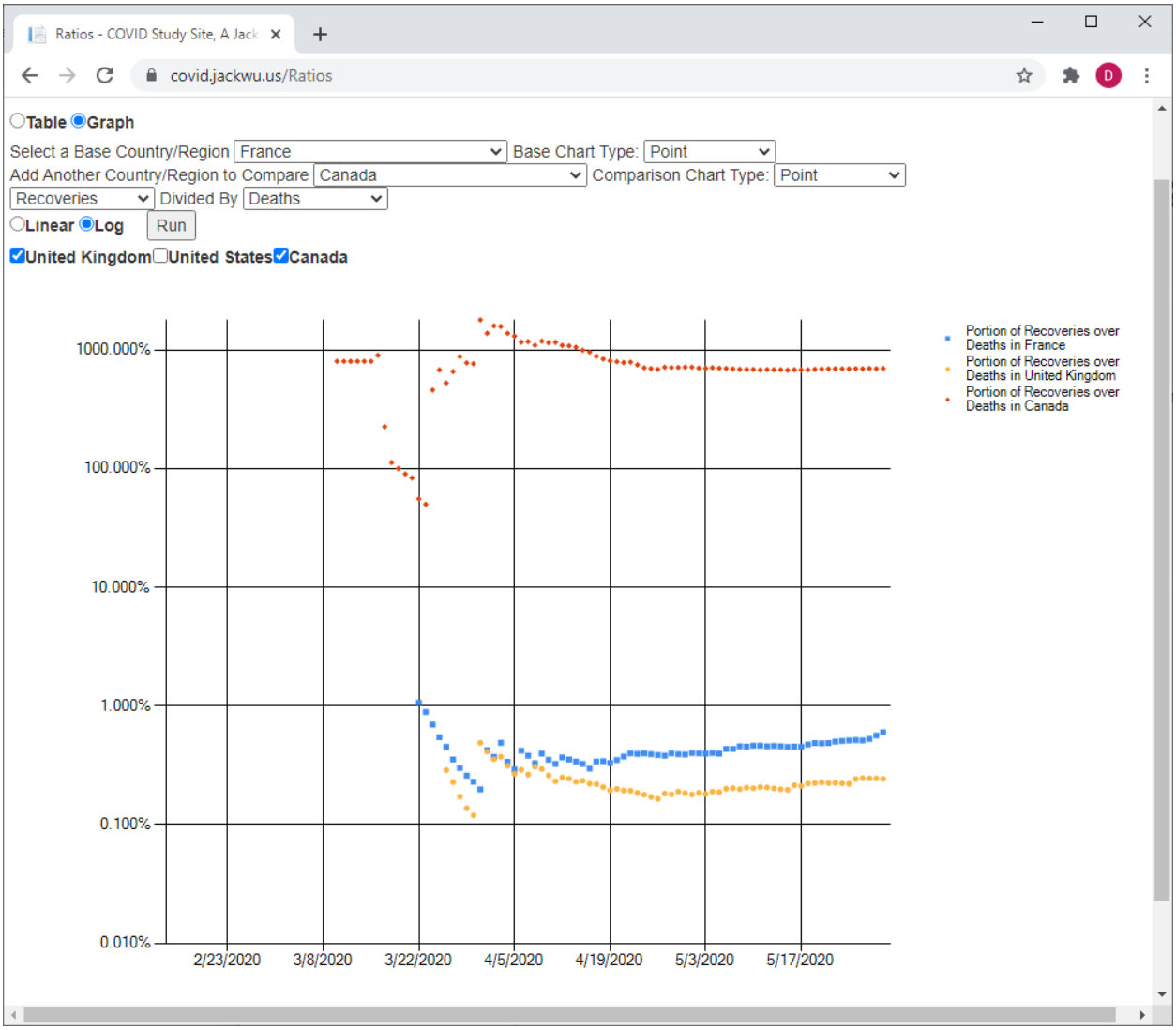


5.2.3 All ratios comparison between all countries and states

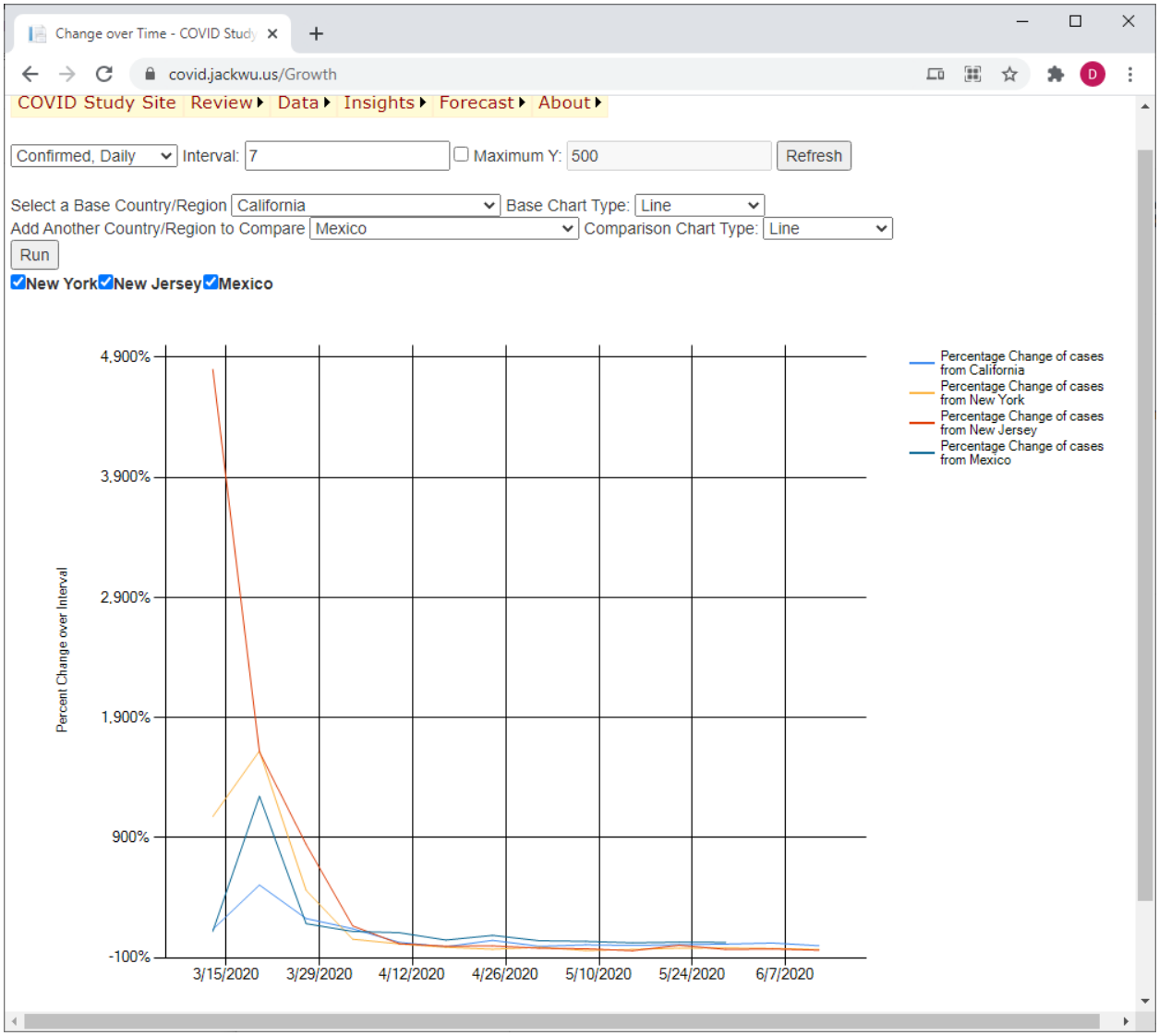
This unique feature from my website provides the ratio comparison between any country and state in table view or graph view: https://covid.jackwu.us/Ratios The following is a deaths over confirmed cases comparison between Italy and Spain:



The following compares Recovery population over Death population for France, UK and Canada:



5.2.4 Dynamic (change over time) comparison of ratios between any countries and states https://covid.jackwu.us/Growth



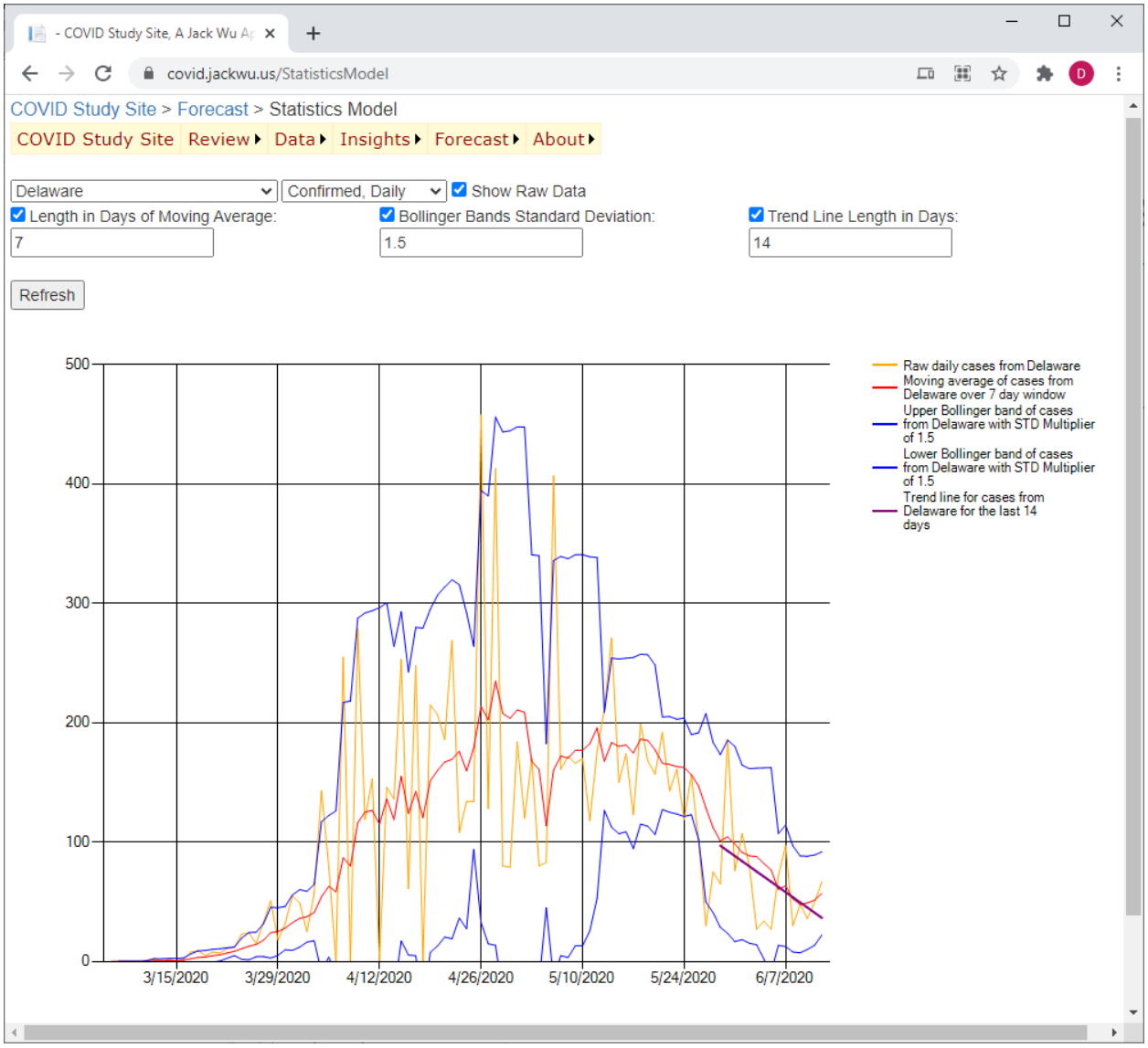
5.3 Forecasting from Model

5.3.1 Statistical model

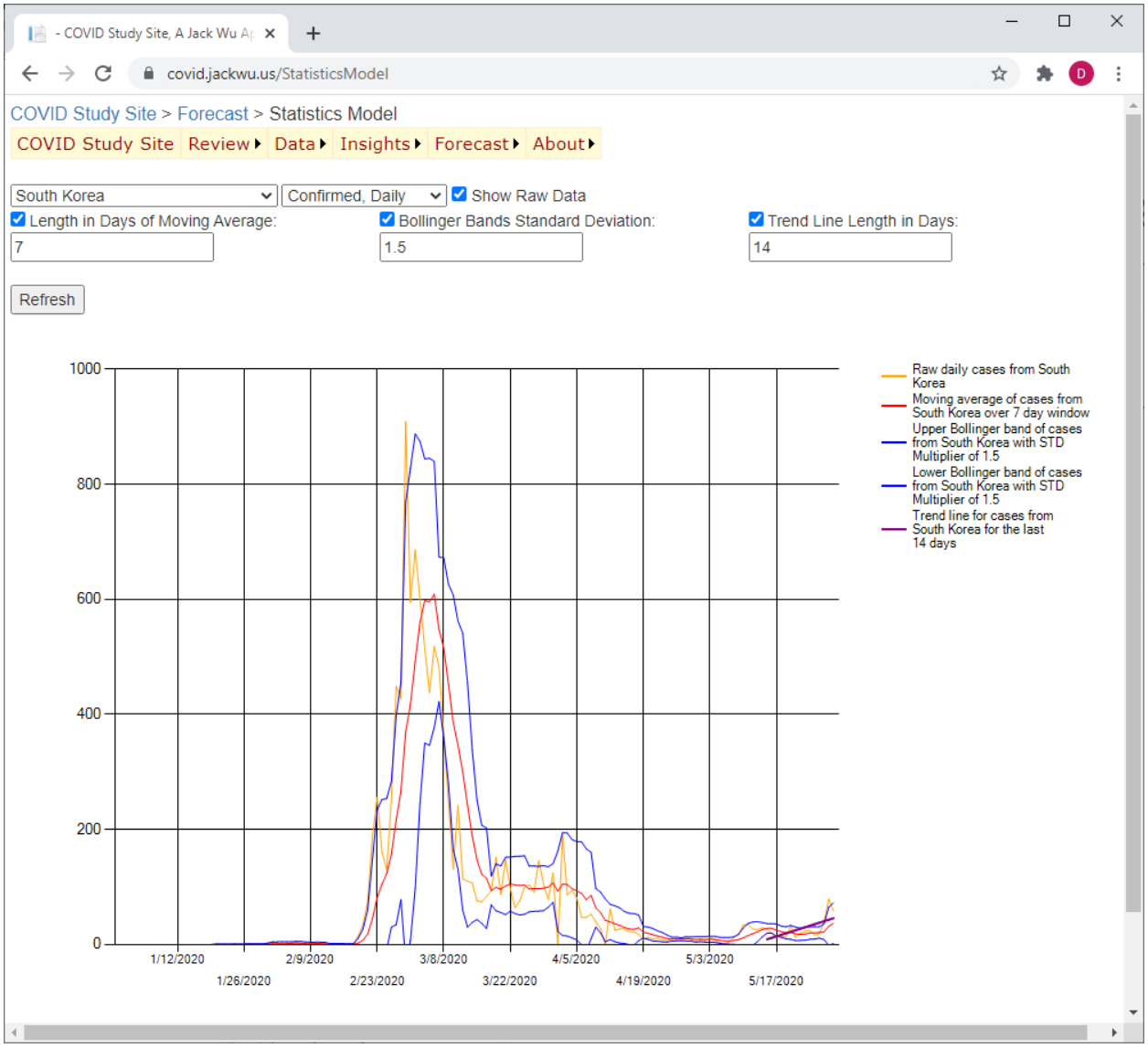
On this page https://covid.jackwu.us/StatisticsModel I provided three statistical calculations for all countries and states

1) Moving average with the length in days adjustable
2) Bollinger Bands calculated from an inputted value for number of standard deviations
3) Trend line for any length of days

The following is for Delaware:



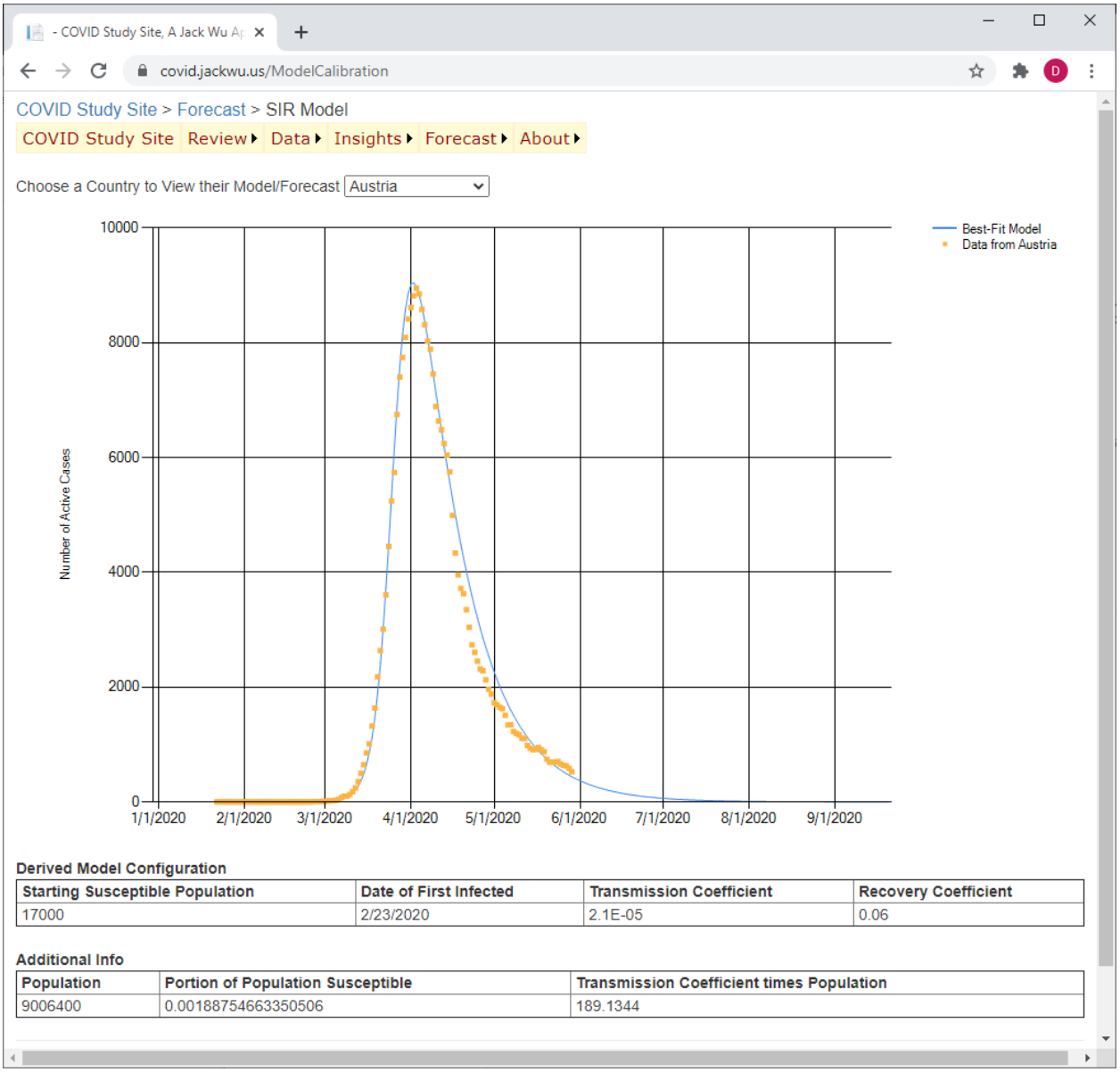
For South Korea:



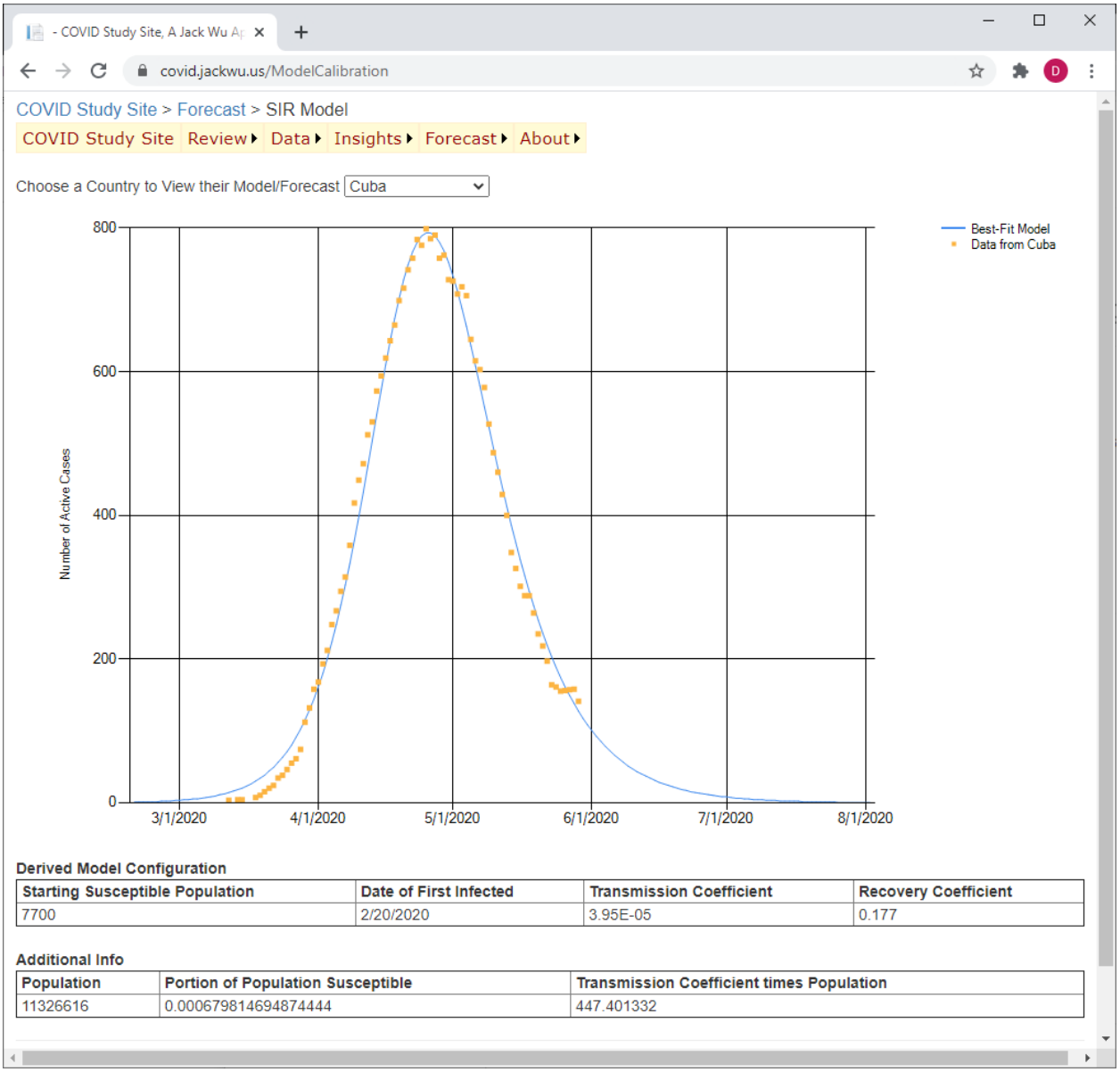
5.3.2 Compartment model

My work involves curve fitting to calculate compartment model parameters. My work is different from other people's in that other researchers target one country or one region, and manually adjust and calculate those SIR model parameters. By contrast, my work is universal, and by choosing any country or state data, my model will automatically, and immediately do the curve fitting and calculate the parameters for you. For example, the following curve fittings are for certain countries: dots represent real data, and the curve represents the best-fit model calculation. All the parameters are calculated in the tables below the graph:

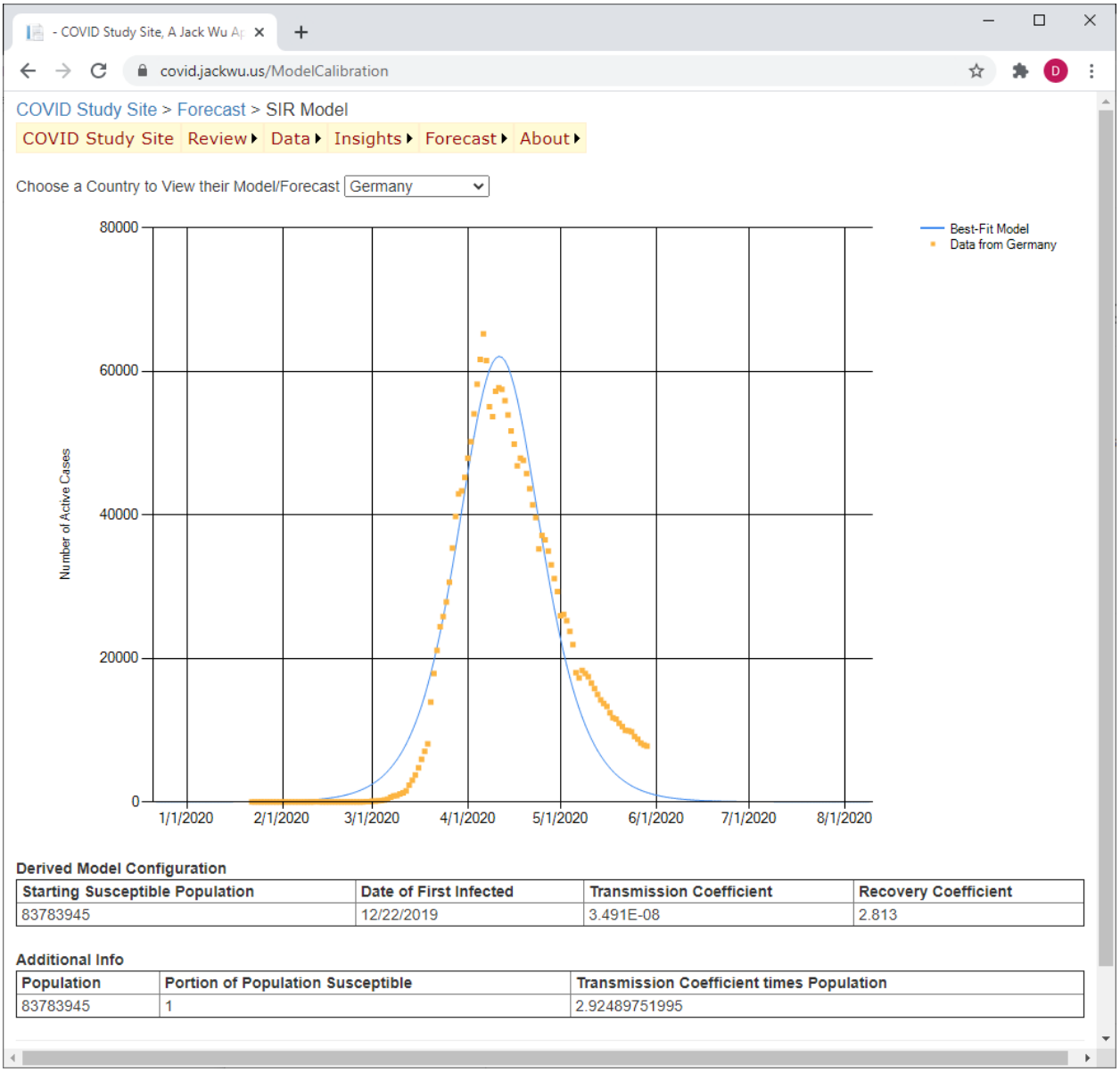
For Austria:


- COVID Study Site, A Jack Wu Ap
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Austria
Best-Fit Model
Data from Austria
Number of Active Cases
10000
8000
6000
4000
2000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
9/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
17000
2/23/2020
2.1E-05
0.06
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
9006400
0.00188754663350506
189.1344

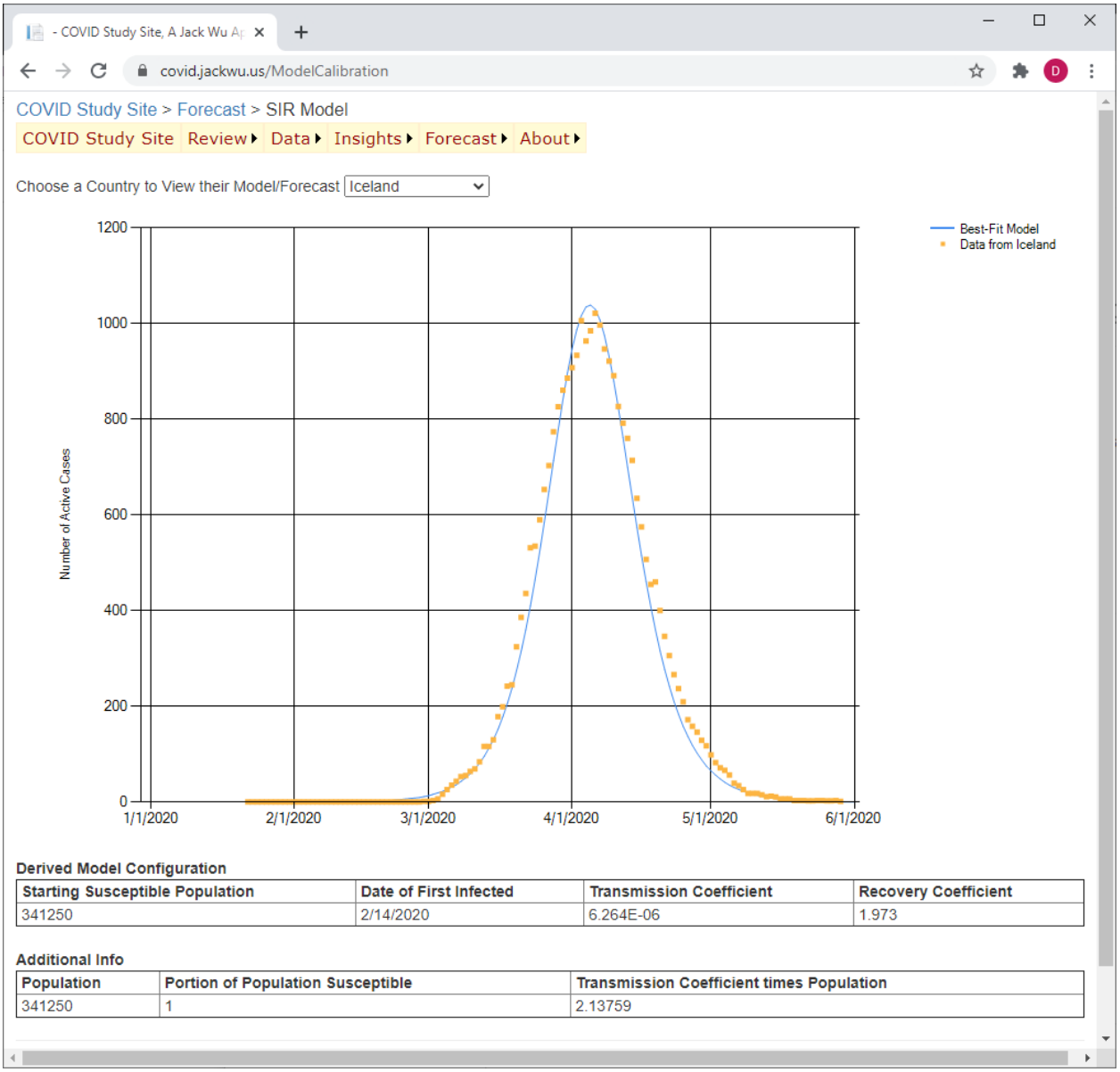
For Cuba:


- COVID Study Site, A Jack Wu Ap
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Cuba
Best-Fit Model
Data from Cuba
Number of Active Cases
800
600
400
200
0
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
7700
2/20/2020
3.95E-05
0.177
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
11326616
0.000679814694874444
447.401332

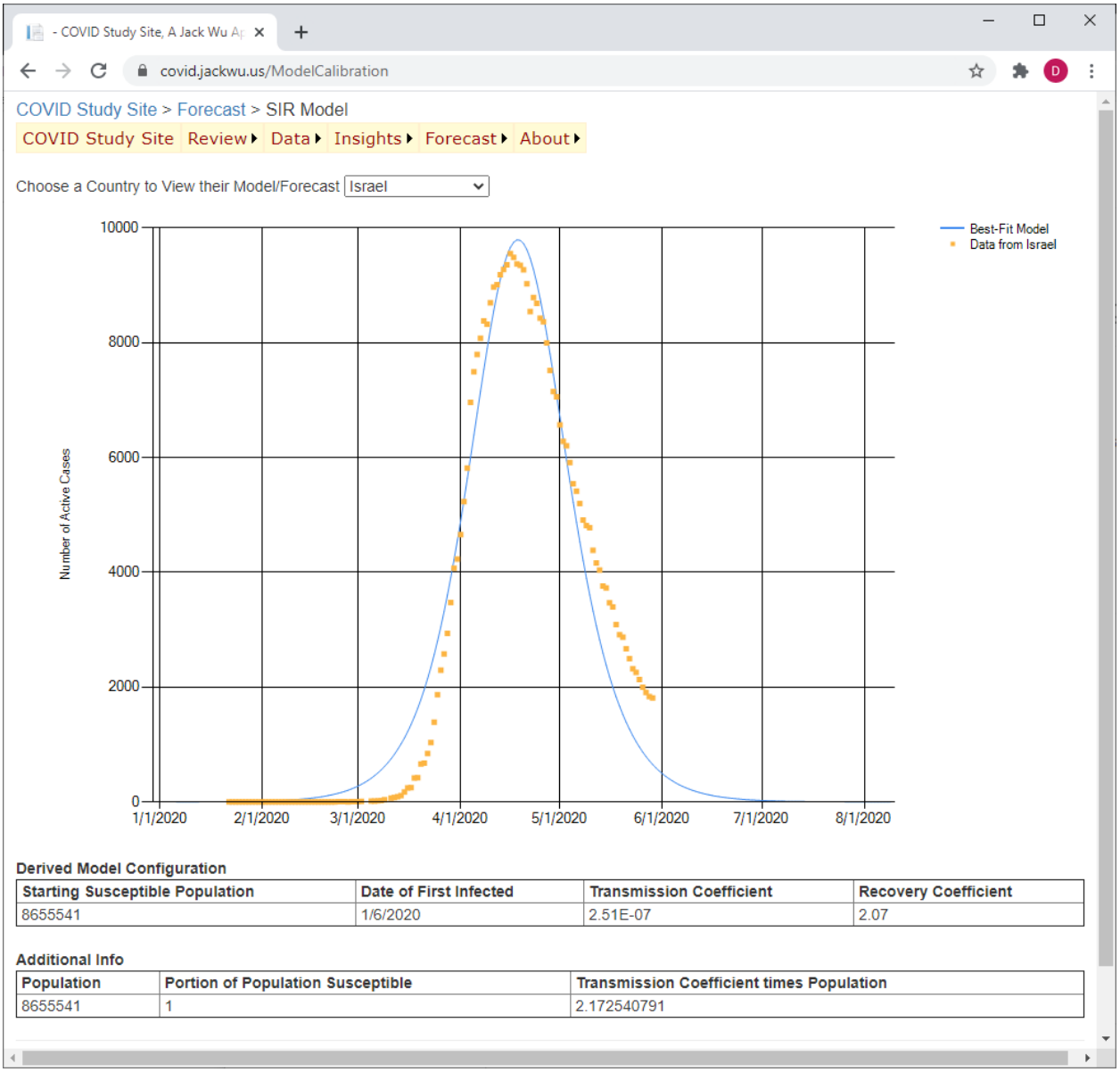
For Germany:


- COVID Study Site, A Jack Wu Ap
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Germany
Best-Fit Model
Data from Germany
Number of Active Cases
80000
60000
40000
20000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
83783945
12/22/2019
3.491E-08
2.813
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
83783945
1
2.92489751995

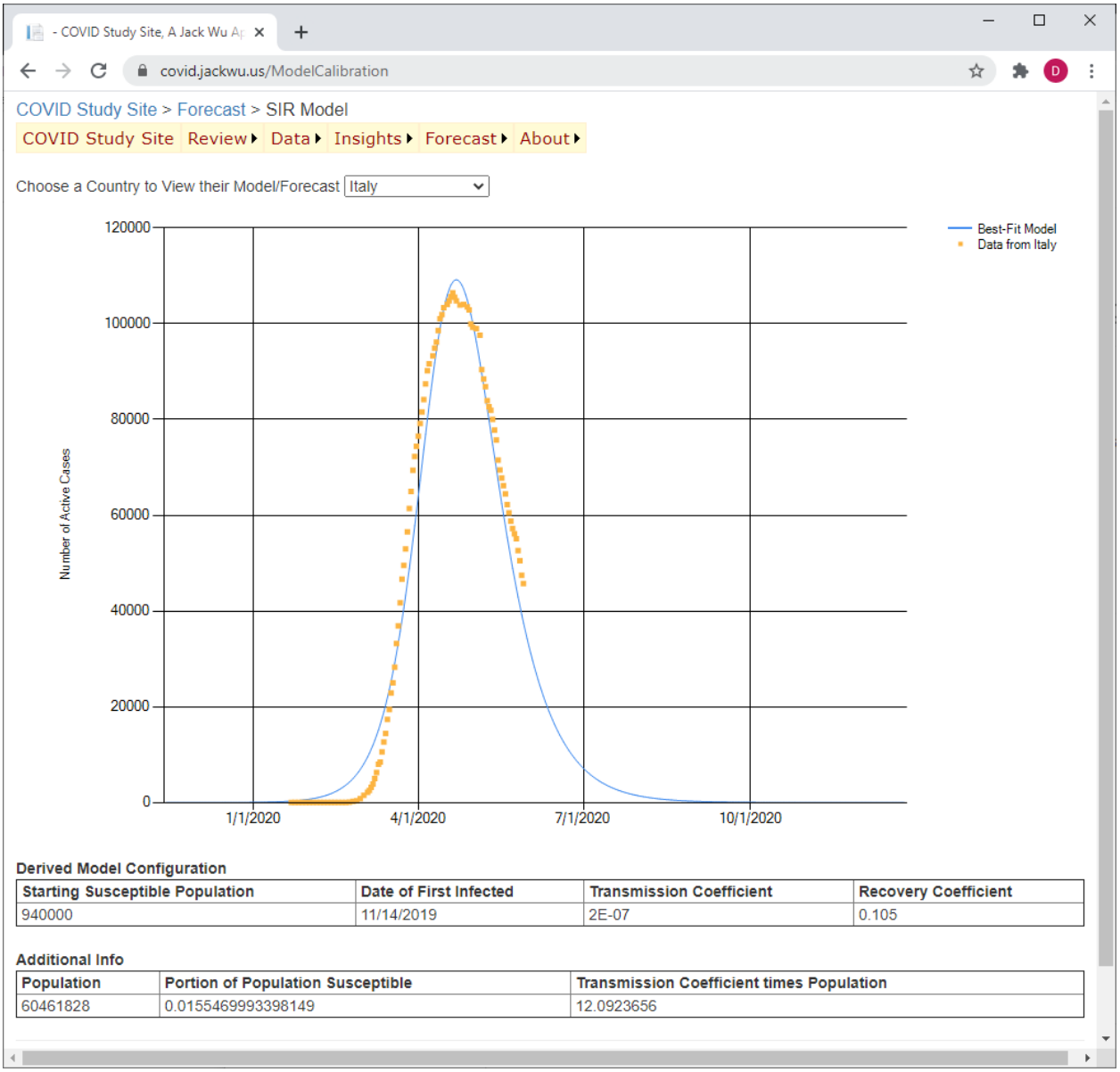
For Iceland:


- COVID Study Site, A Jack Wu A
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Iceland
Best-Fit Model
Data from Iceland
Number of Active Cases
1200
1000
800
600
400
200
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
341250
2/14/2020
6.264E-06
1.973
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
341250
1
2.13759

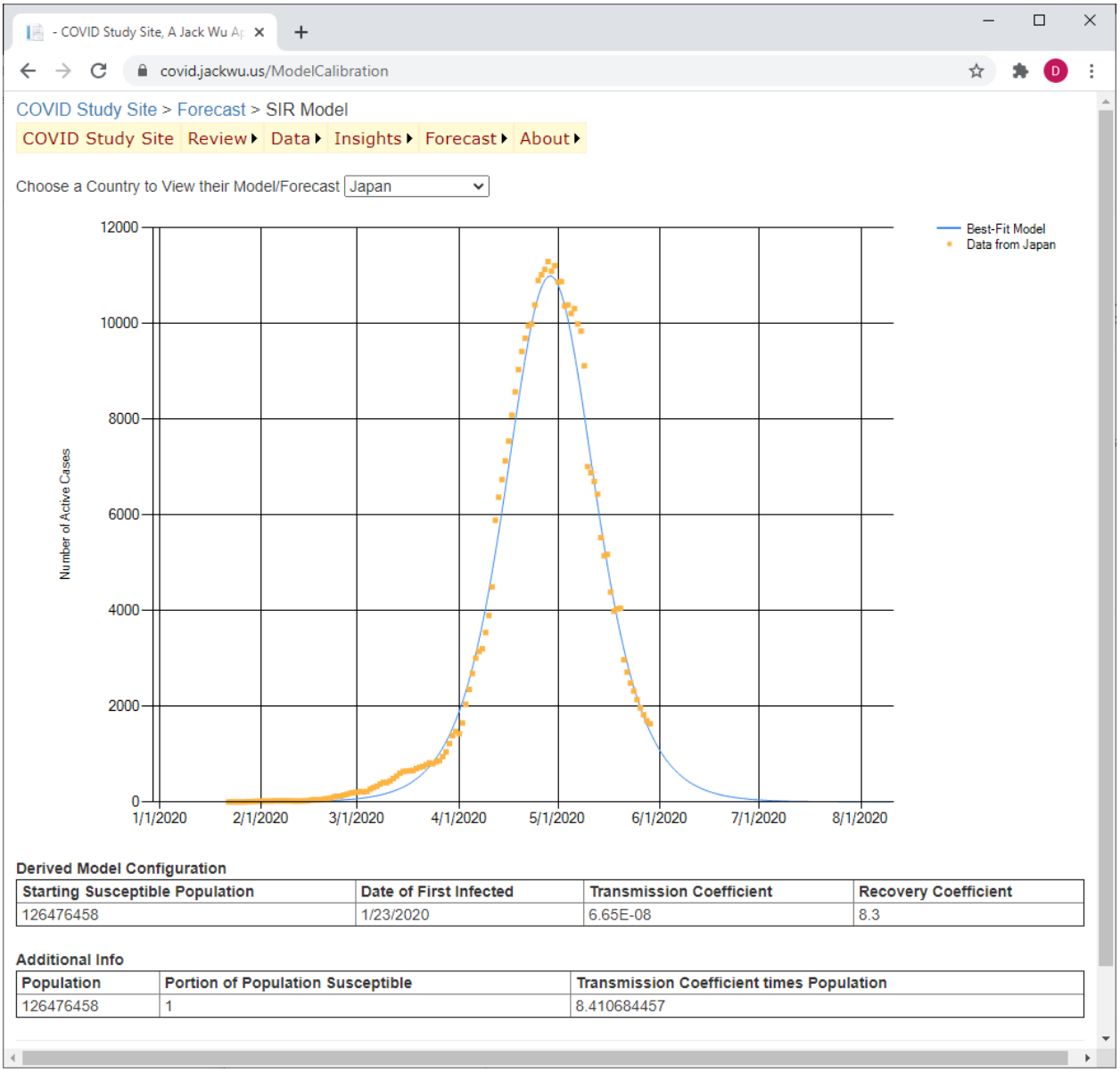
For Israel:


- COVID Study Site, A Jack Wu Ap
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Israel
Best-Fit Model
Data from Israel
Number of Active Cases
10000
8000
6000
4000
2000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
8655541
1/6/2020
2.51E-07
2.07
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
8655541
1
2.172540791

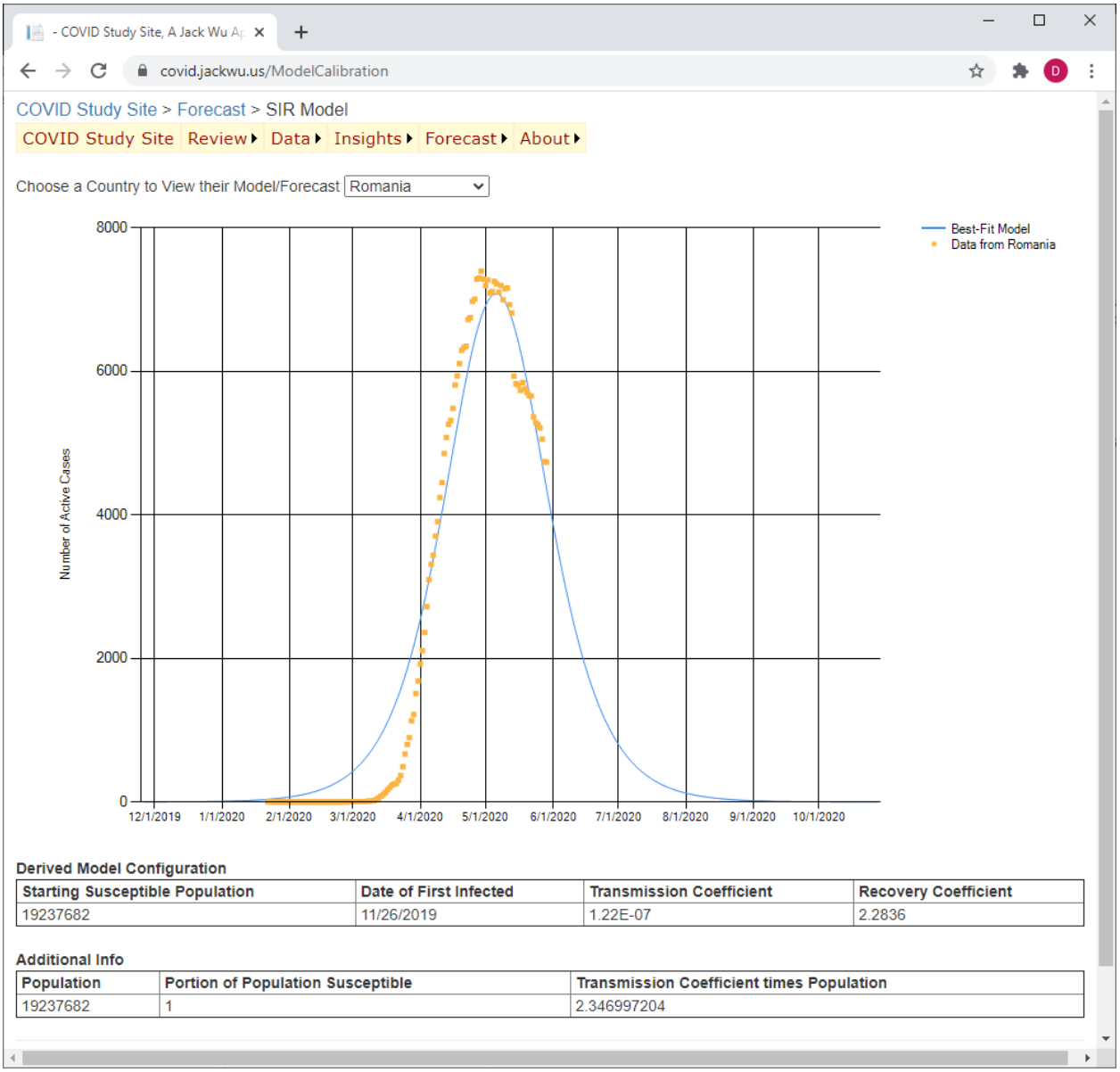
For Italy:


- COVID Study Site, A Jack Wu Ap
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Italy
Best-Fit Model
Data from Italy
Number of Active Cases
120000
100000
80000
60000
40000
20000
0
1/1/2020
4/1/2020
7/1/2020
10/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
940000
11/14/2019
2E-07
0.105
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
60461828
0.0155469993398149
12.0923656

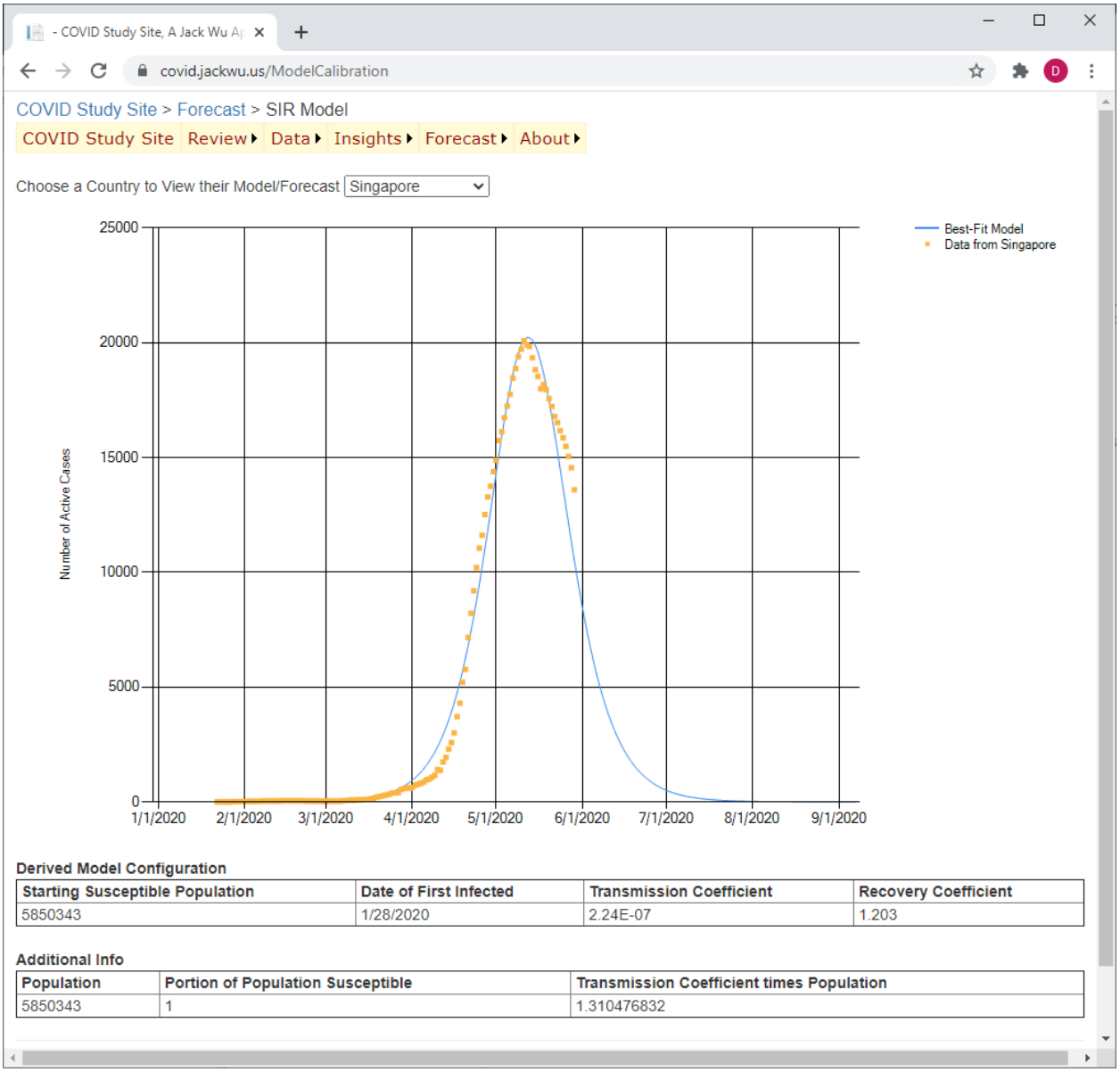
For Japan:


- COVID Study Site, A Jack Wu Ap
covid.jackwu.us/ModelCalibration
COVID Study Site > Forecast > SIR Model
COVID Study Site
Review
Data
Insights
Forecast
About
Choose a Country to View their Model/Forecast
Japan
Best-Fit Model
Data from Japan
Number of Active Cases
12000
10000
8000
6000
4000
2000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
Derived Model Configuration
Starting Susceptible Population
Date of First Infected
Transmission Coefficient
Recovery Coefficient
126476458
1/23/2020
6.65E-08
8.3
Additional Info
Population
Portion of Population Susceptible
Transmission Coefficient times Population
126476458
1
8.410684457

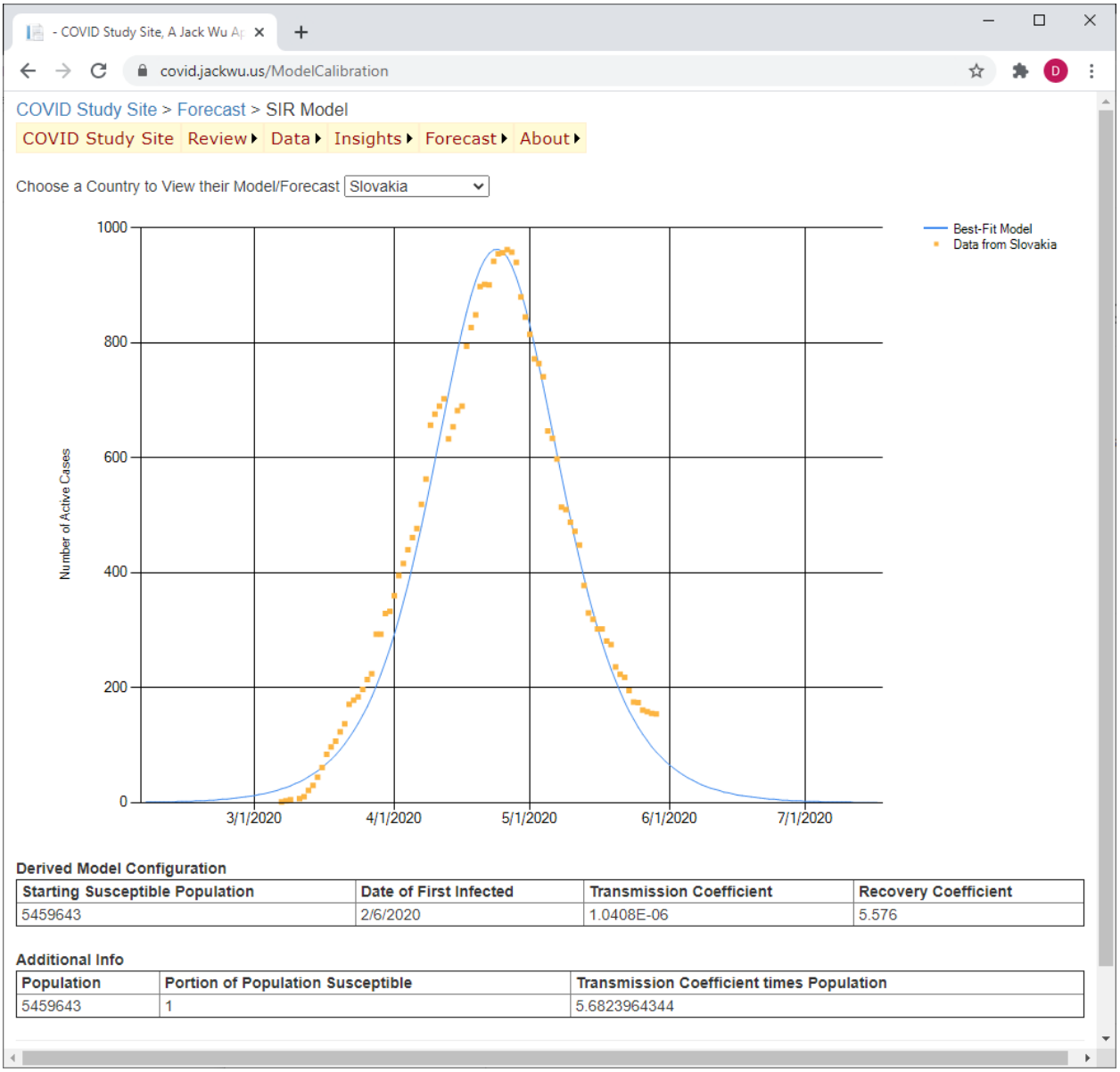
For Romania:



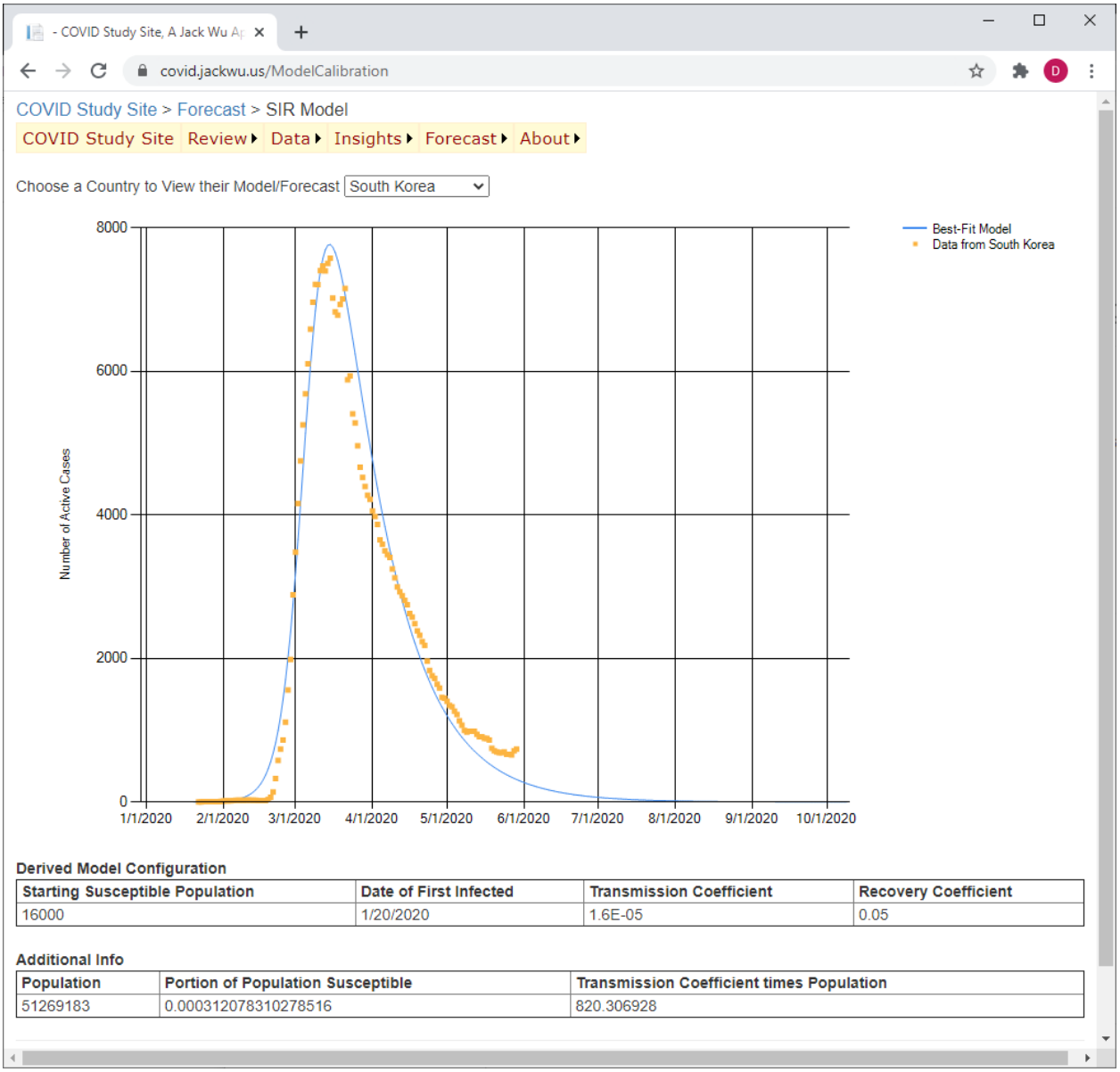
For Singapre:



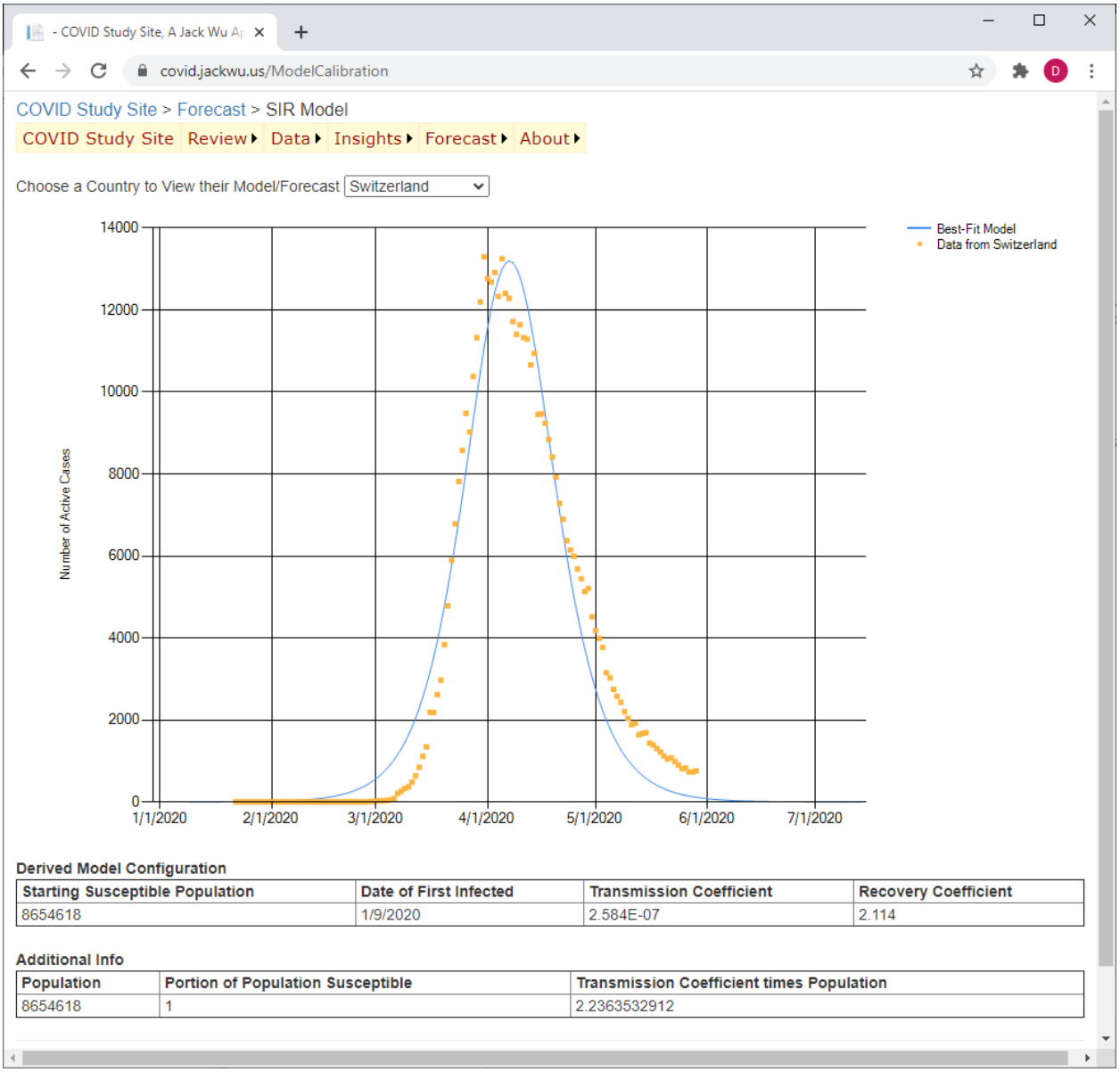
For Slovakia:



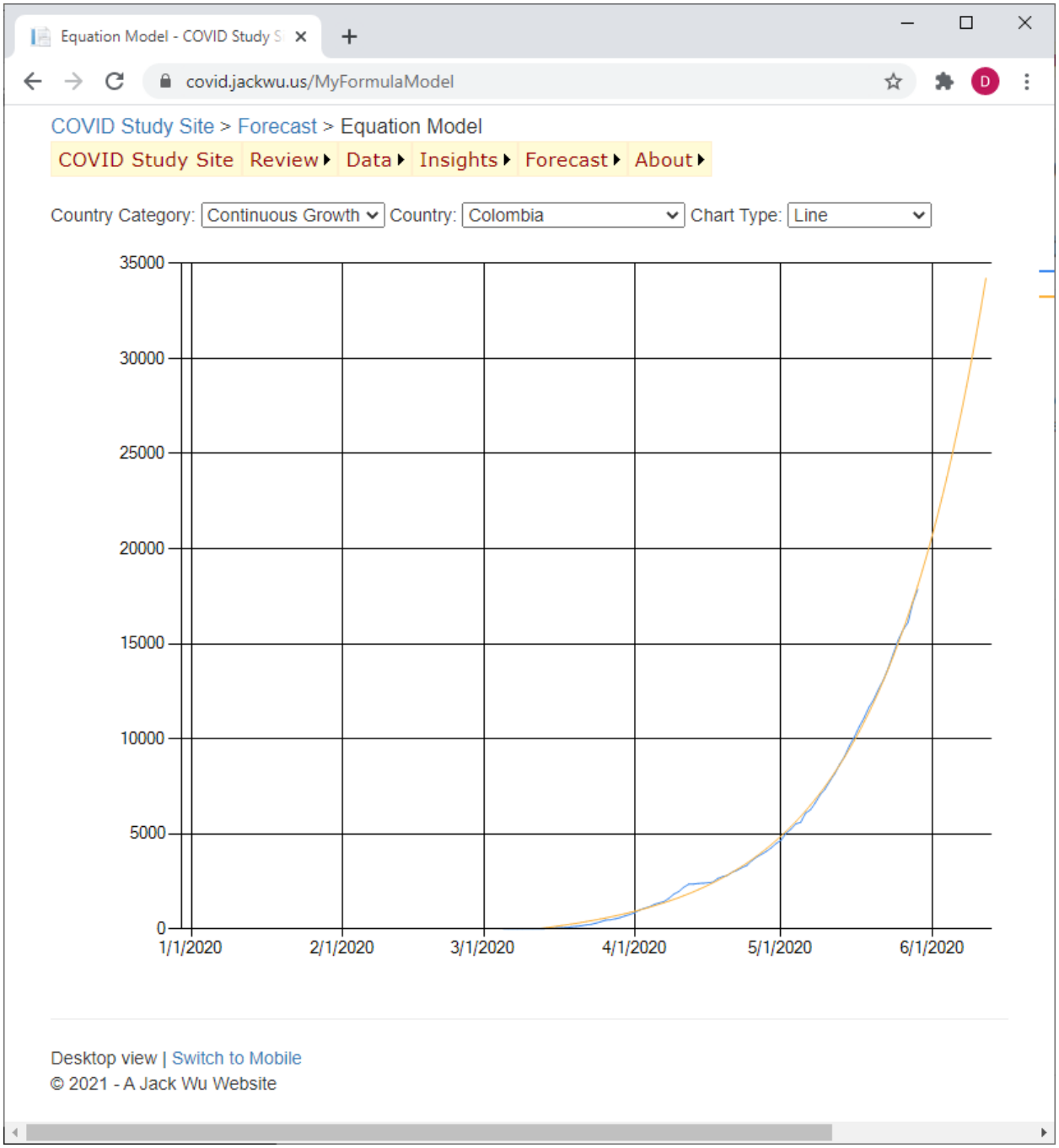
For South Korea:



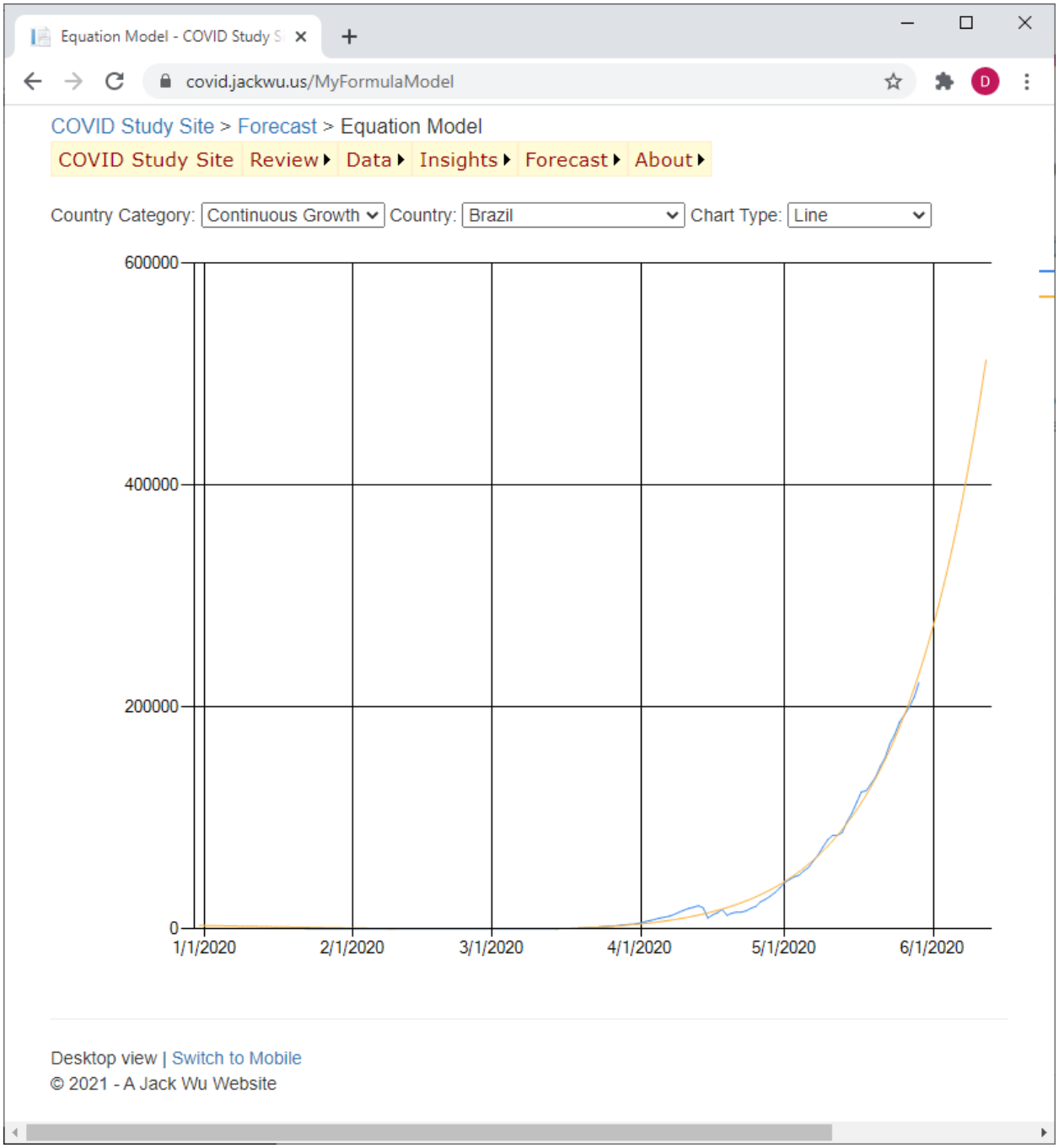
For Switzerland:



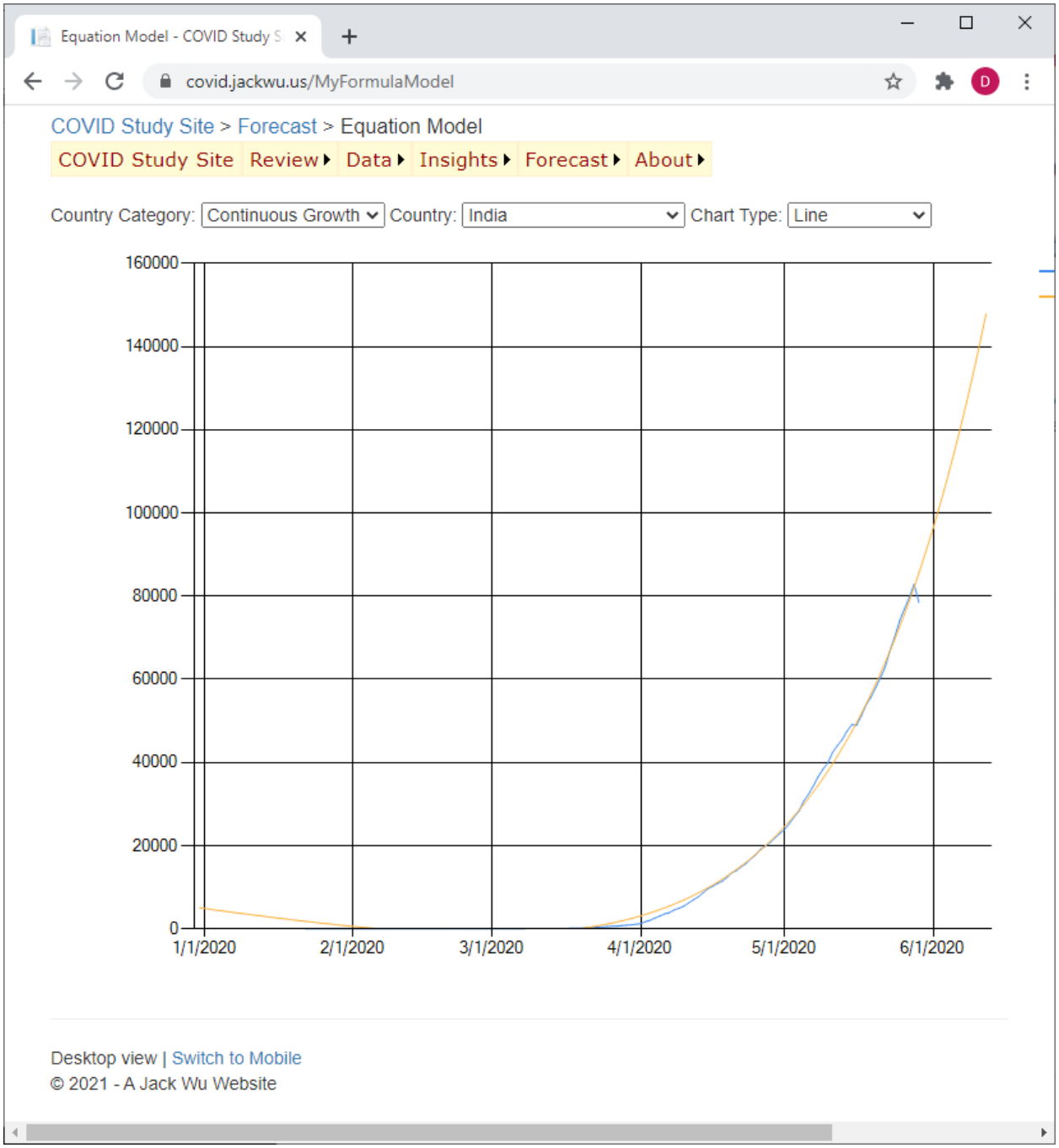
For Columbia:


Equation Model - COVID Study S
covid.jackwu.us/MyFormulaModel
COVID Study Site > Forecast > Equation Model
COVID Study Site
Review
Data
Insights
Forecast
About
Country Category: Continuous Growth
Country: Colombia
Chart Type: Line
35000
30000
25000
20000
15000
10000
5000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
Desktop view | Switch to Mobile
© 2021 - A Jack Wu Website

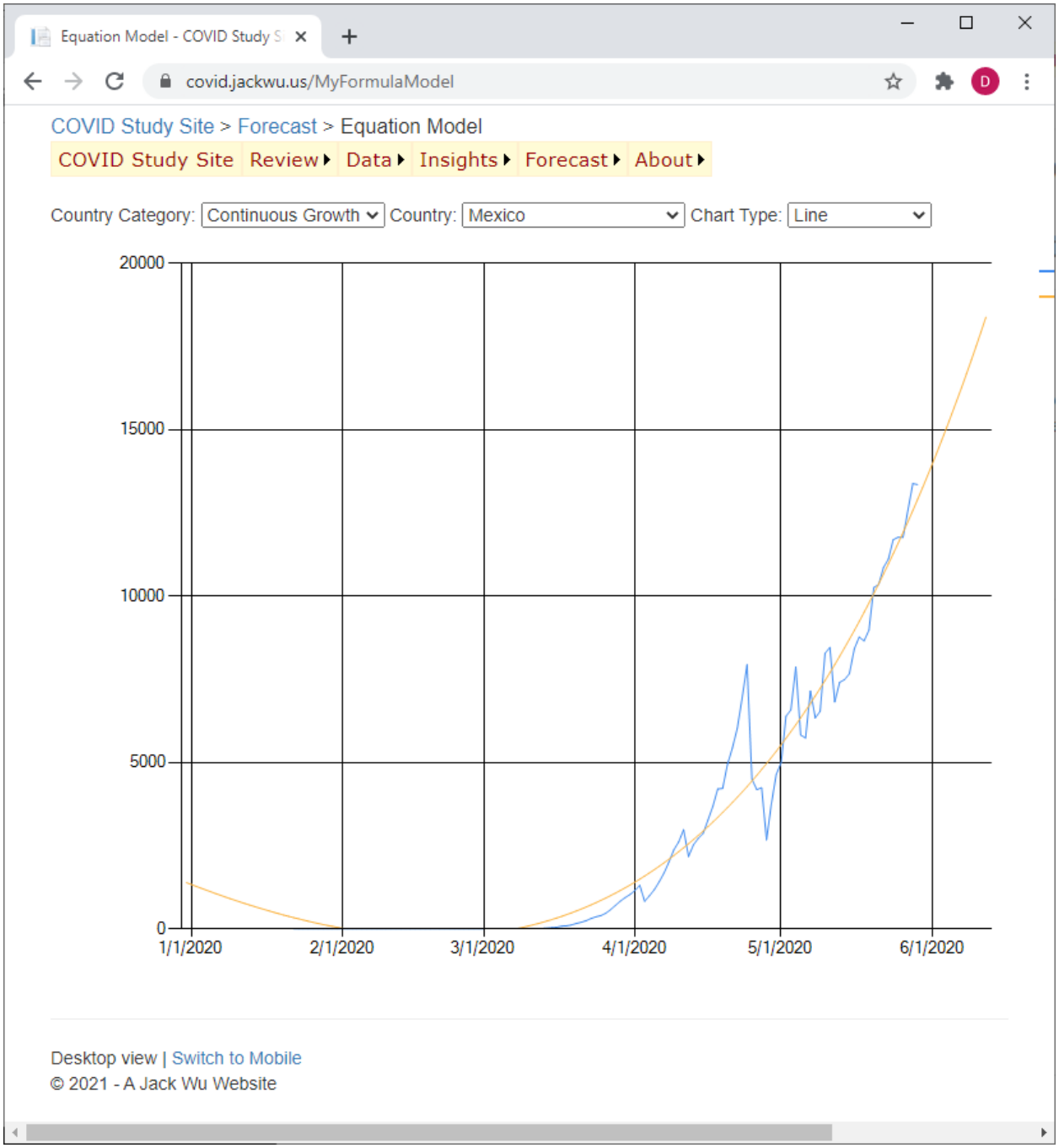
For Brazil:


Equation Model - COVID Study S
covid.jackwu.us/MyFormulaModel
COVID Study Site > Forecast > Equation Model
COVID Study Site
Review
Data
Insights
Forecast
About
Country Category: Continuous Growth
Country: Brazil
Chart Type: Line
600000
400000
200000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
Desktop view | Switch to Mobile
© 2021 - A Jack Wu Website

For India:


Equation Model - COVID Study S
covid.jackwu.us/MyFormulaModel
COVID Study Site > Forecast > Equation Model
COVID Study Site
Review
Data
Insights
Forecast
About
Country Category: Continuous Growth
Country: India
Chart Type: Line
160000
140000
120000
100000
80000
60000
40000
20000
0
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
Desktop view | Switch to Mobile
© 2021 - A Jack Wu Website

For Mexico:



Clearly, my model fits the real data very well, and it give the prediction for forecasting. Unlike most other researches they developed one model for one country, I developed one model to fit all the countries

6. Results

There were three main outcomes for curve-fitting each country: either a success, in which the curve_fit function found a best fit to the data, a bad fit, in which the curve_fit settled on a fit that was obviously not a best fit (usually a simple linear regression), or a runtime error, caused by the curve_fit function surpassing the allotted 'iteration" count for finding the local minimum. In the case of the bad fit and run-time error, this was usually due to a poor initial guess, as the model became divergent instead of convergent at that point. In the case of the runtime error, the curve_fit function found no minimum, while with the bad fit, the result would usually be a simple linear regression over the entire graph, ignoring the two exponential terms. This made the predicted curve very obviously inaccurate with respect to the data.

Another issue is present with specifically the "Bell" chart type. The SIR Model and most countries have an overall trajectory that rises, reaches a peak, falls, and then settles to 0. However, a perfect bell is very hard to recreate with the chosen formula; instead of creating a bell, starting at and settling at 0, the model had to use the first "hump" in a second wave graph to produce the bell. This resulted in a predicted exponential upward trajectory for practically every bell graph that came out with a good fit, and many countries had bad linear regression fits.

Efficient Testing:

Near the beginning of the COVID 19 pandemic, China was tasked with testing the population of Wuhan, 11 million, in merely 10 days. Through standard testing, by testing one person at a time, 11 million tests would be required. However, COVID tests are capable of testing more than one person at a time. Testee samples can be combined, and if any one sample is positive, the entire test will output positive. While this functionality at first seems insignificant, it has great power in that if it outputs negative, it is able to confirm negative test results for multiple people with a single test. In this appendix, we analyze and optimize this testing method.

In order to minimize the required tests, we need to maximize the verdicts made (number of people confirmed positive or negative) per test. For tests with a single blood sample, the verdicts per test is obviously 1. For tests with multiple blood samples, we can only reach verdicts if the result comes back negative, confirming everyone in the test as negative; a positive result would be ambiguous. We can quantify the number of verdicts per test with a mathematical equation. Let b be the number of blood samples per test, and f be the frequency of a positive case. The chance we will get a negative test result, confirming all testees as negative, is the probability that everyone in the test really is negative, or $(1-f)^b$. If the test result is negative, the number of verdicts made is equal to b. Thus, the estimated number of verdicts per test is equal to $V(b) = (1-f)^b * b$. We want to maximize the number of verdicts per test, as that maximizes the amount of information we gain from each test. To find the b value that will maximize the verdicts per test, we calculate its derivative and set it to 0:*

$V'(b) = \ln(1-f) * b * (1-f)^b + (1-f)^b = 0$
$(\ln(1-f) * b + 1)(1-f)^b = 0$
$b = -1 / \ln(1-f)$

With a hypothetical frequency of 10% positive cases, we derive a value of b = 9.49, which means that V(9) or V(10) gives us the most optimal verdicts per test setup. V(9) and V(10) for this case are equal, which means that both b=9 and b=10 will work equally well (however, b = 9 is more consistent and less sensitive to fluctuations, so in practice, it would be preferred). Assuming b = 10 for simplicity and a population of 10000 people with a positive frequency of 10%, we can expect out of 100 tests, we can expect ~34.9 tests to come up negative, confirmining 349 people as negative. This is much better than using one test per person, which would only be able to confirm 100 people with the used 100 tests.

As a hypothetical, let us assume 35 tests appeared negative, allowing us to confirm 350 people as negative and leaving 650 people uncertain. Here, we can repeat the previous process on the smaller population. However, because some confirmed negatives have been removed from the population, the positive frequency has changed, and thus we must recalculate our b value. Within our hypothetical, and letting P represent the uncertain population, the new frequency is equal to $F_{new} = F_{old} * P_{old} / P_{new} =$ *15.4%. Our new b value is thus 5.98, which is rounded to 6. V(6) = 2.2, so we can expect to confirm 2.2 people per test.*

This process repeats until our f value gets high enough and our maximum b value gets low enough that the maximum value for b is under 1. This means that direct individual testing is now more efficient than group testing, and thus all remaining uncertain individuals should be directly tested.

This process is extremely effective when the positive frequency is very low, such as at the beginning of an outbreak. With a value of f = 0.001 (0.1%), V(1000) reaches a value of 368, meaning that the first round of tests are over 360 times as efficient. A self-created simulation estimates that the total number of tests needed to confirm 10000 people is under 200 tests, resulting in the average test being 50 times as efficient. With a lower rate of f = 0.0001, (0.01%), which equates to about 33,000 people for the US (the number of confirmed cases around the first half of April), the first round of tests are more than 3600 times as efficient.

Of course, there are plenty of practical issues in implementing such a theoretical conclusion. Using thousands of blood samples for a single test would be both a challenge for logistics and test centers, as well as for the medical community to be able to reliably detect very small amounts of virus. The effect of false negatives and false positives would be greatly amplified. Additionally, if the true number of positive cases was significantly higher than expected, there could be many more positive test results than anticipated and the effectiveness of these tests would be dramatically limited. However, China was still able to utilize this strategy using about 30 blood samples per test.

** You could reach a verdict from a positive test result by using process of elimination; for example, if 5 blood samples test positive, and 4 of those samples are later confirmed negative, then the last sample must be positive; however, this minor and situational optimization is excluded for simplicity.*

Conclusion and Discussion

My work on COVID-19 modeling has completed the following:

1) A thorough review for almost all models currently used in COVID-19 studies, from Statical models to mathematical models, from branching process to compartment models, from spatial modeling to global surveillance, from continuous differential equations to discrete mathematics, from stochastic process and the Markov chain to fractional calculus.
2) The first and only website available on the Internet that can provide unlimited comparison between almost all parameters for an unlimited number of countries and states.
3) The first and only website can do Statistical modeling and SIR modeling for any country and state. Usually a scientific paper published only for one or two countries or regions on SIR modeling.
4) From the pattern of the evolution of the COVID-19, I predicted growth. I also found the relations between the growth of the cases with the counties.
5) I created a ratio comparison between any country and state, which is not seen at anywhere else
6) I found for multiple existing conditions, the mortality rates jump to 95% at the age of 45 years old

Ethics Declaration: Competing Interest

The author declares that he has no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The Author also has no affiliations with any institution. Unlike most such research funded by industry and government, the author does not have any sponsors, and pays all expenses out of his own pocket. Therefore, this is purely evidence based research, not message driven strategy publication planning for promoting theories, methods, protocols, and medications.

Discussion

The most obvious improvement to this study is to use a more complicated model. Very complex models with many compartments have been made, such as the SIDARTHE model[6], which has eight compartments. The next step is to design a model that incorporates testing and diagnosed individuals, as the data we have observed and based our model off of is not the true number of infectives,but instead the confirmed cases of those who have been infected.

Future Work

The graph below shows the researcher's own goals and process towards COVID-19 research.

Additional data sources that could be used include WorldOmeter, BNO, JHU, and DXY. COVID-19 numbers from these websites will be loaded into an MSSQL database. The collected data is used for three primary methods. The first is numerical analysis, which includes analytical math / statistics, fitting a country's data versus time to a curve generated by a compartment mode, and computer simulation. The

second method is a Real Time Dashboard that publicly displays the collected data in ways that will make the data easier to interpret, primarily graphs. The third method is to use the data to fit a statistical model, and use data from prior outbreaks such as SARS to predict the trajectory of COVID. When combined together, the results of these research methods will allow scientists to calculate the parameters of regions regarding their infection, recovery, and death rates; forecast how the disease will grow or diminish in the future; and help politicians and strategists make crucial decisions regarding the COVID-19 response.

Researchers have used two different approaches for modelling: statistical modelling and compartmental modelling. Statistical models are constructed using observed data from the past, including both data from the current epidemic and data from previous epidemics. Using the known trajectories of past diseases, statistical models are then extended to determine the most likely result for the current outbreak. The Institute for Health Metrics and Evaluation (IHME) has used this method to outline the epidemic curve in China, Italy, all US states, and many other countries[4][5].

The other popular model is the compartmental model, which comes from the field of epidemiology itself. The compartmental model keeps track of a population, placing each individual into one compartment, or state. Over time, individuals move from one state to another, causing the number of individuals in each compartment to change. The rate of change equals the difference between the influx and outflux of the compartment, which leads to the establishment of a set of differential equations. By solving and/or simulating these equations, we can predict the spread of the virus, understand how outside forces (such as public intervention or lockdown) affect the epidemic, and know what measures are necessary to end the epidemic in a given population[6].

References

[1] Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020;323(13):1239–1242. doi:10.1001/jama.2020.2648

[2] CSSE (Center for Systems Science and Engineering) COVID-19 Dashboard Johns Hopkins. University (2020) Available at https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

[3] WHO. Coronavirus Disease 2019 (COVID-19): Situation Report 76 (WHO, 2020).

[4] SDS (System Dynamics Society) COVID-19 Resource Page (2020) Available at https://www.systemdynamics.org/covid-19

[5] IHME(Institute for Health Metrics and Evaluation) (2020) Available at https://covid19.healthdata.org/

[6] Los Alamos National Laboratory (2020) COVID-19 Confirmed and Forecasted Case Data Available at https://covid-19.bsvgateway.org

[7] Giordano, G., Blanchini, F., Bruno, R. et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med (2020). https://doi.org/10.1038/s41591-020-0883-7

[8] Hellewell J Abbott S Gimma A Bosse NI Jarvis C
Feasibility of controlling 2019-nCoV outbreaks by isolation of cases and contacts.
medRxiv. 2020; DOI:10.1101/2020.02.08.20021162

[9] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963; 78:494-506; PMID:19316455; http://dx.doi.org/10.2307/4591848

[10] State of Delaware -- My Healthy Community
https://myhealthycommunity.dhss.delaware.gov/locations/state

[11] Wikipedia Autoregressive integrated moving average
https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

[12] Constantinos I. Siettos & Lucia Russo (2013) Mathematical modeling of infectious disease dynamics, Virulence, 4:4, 295-306, DOI: 10.4161/viru.24041

[13] Wikipedia CUSUM https://en.wikipedia.org/wiki/CUSUM

[14] Wikipedia Exponential Moving Average
https://en.wikipedia.org/wiki/Moving_average#Exponential_moving_average

[15] Wikipedia Hidden Markov model https://en.wikipedia.org/wiki/Hidden_Markov_model

[16] Wikipedia Compartmental models in epidemiology
https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

[17] Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A. 115 (772): 700–721. Bibcode:1927RSPSA.115..700K. doi:10.1098/rspa.1927.0118.