*Applicable datasets are made available by their respective providers pursuant to the terms set forth on the applicable sites on which they are made available.*

***Licensed to the Apache Software Foundation (ASF) under one or more contributor license agreements.  See the NOTICE file distributed with this work for additional information regarding copyright ownership.  The ASF licenses this file to you under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License.  You may obtain a copy of the License at***

[***http://www.apache.org/licenses/LICENSE-2.0***](http://www.apache.org/licenses/LICENSE-2.0)***. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  See the License for the specific language governing permissions and limitations under the License.***

**READ ME COVID Case Projection Model**

**Development of Projection Model & References:**

The python notebooks pertaining to the projection of confirmed cases and hospitalizations follow the naming convention as shown below:

New\_code\_run\_SVI\_<day (of month) of run>\_<month of run>\_<State Code for State in US>.ipynb

For example, the notebook New\_code\_run\_SVI\_13\_5\_CT.ipynb runs the model with actual data observed up to the 13th of May, and generates three datasets for the State of Connecticut:

* Confirmed cumulative case projections (for three different social distancing scenarios), along with their upper and lower bounds.
* Confirmed daily cases projections (for three different social distancing scenarios), along with their upper and lower bounds.
* Projected hospitalizations (cumulative and active/effective) with their upper and lower bounds.

Separate notebooks have been created for each State’s projections, owing to the long runtime for the model. The State to be run can be changed by simply initializing the value of ‘i’ in the notebook. For example: i=4 refers to California, i=31 refers to New Jersey, and i=34 refers to New York. Additionally, the model leverages the Social Vulnerability Index data to differentiate the default distribution of each State, depending on how vulnerable the State is reported to be.

This projection model can very well be replicated at the County Level by reading the data at the county level instead of the State Level (lockdown dates, County population and Social Vulnerability Index values will need to be initialized separately).

The notebook can be run on Python (>3.6) and leverages the following packages:

* Theano (>=1.0.4)
* Pymc3 (>=3.7)
* Matplotlib (>=3.2)
* Datetime (>=4.3)
* Pandas (>=1.0.3)
* Numpy (>=1.18.4)

It’s important to note that development and support for Theano has ended, and is only used here as a backend for pymc3. The newer version, pymc4 was considered initially, as it runs with Tensorflow rather than Theano. However, pymc4 is still at a pre-release phase, and some parts of it may not be as stable as required.

*Applicable datasets are made available by their respective providers pursuant to the terms set forth on the applicable sites on which they are made available.*

The datasets used in the Python code can be downloaded at the county level from the following sources:

* <https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_confirmed_usafacts.csv>
* <https://svi.cdc.gov/data-and-tools-download.html>

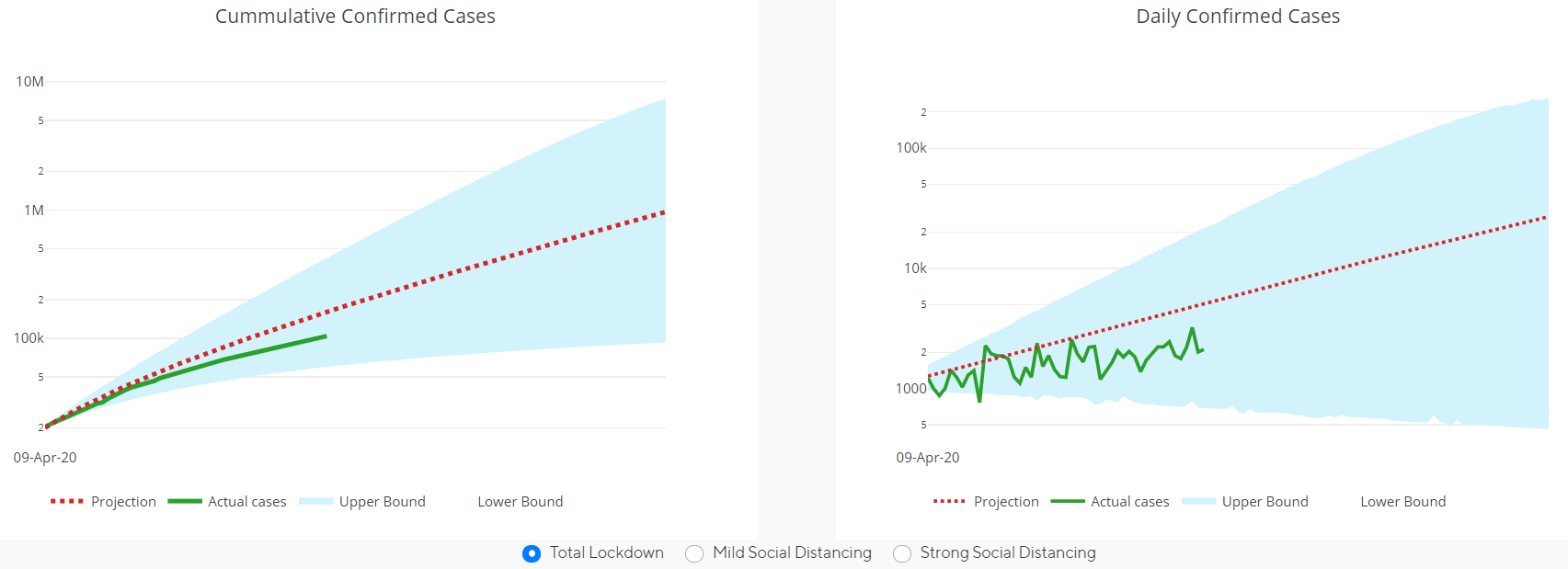
(No direct download link. Go to Data Download section and select Year – 2018, Geography – United States, Geography Type – Counties, File Type – CSV. Documentation for the same can be found here: <https://svi.cdc.gov/Documents/Data/2018_SVI_Data/SVI2018Documentation.pdf>)

* <https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html>
* <https://www.aljazeera.com/news/2020/03/emergencies-closures-states-handling-coronavirus-200317213356419.html>

(Lockdown dates need to be updated based on the changes on this site)

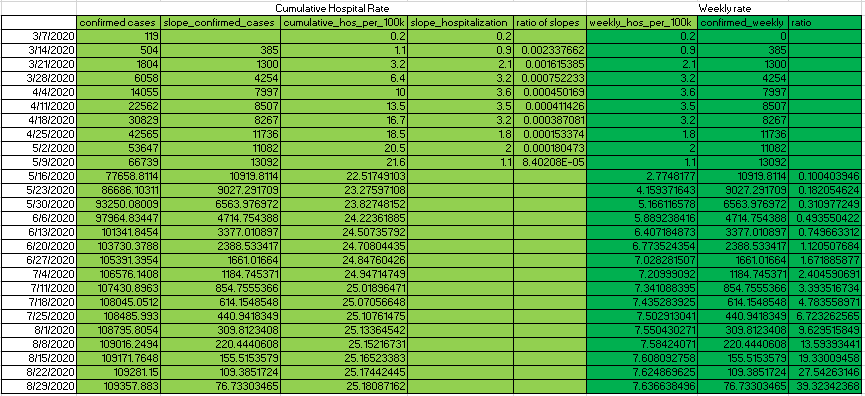
*Applicable datasets are made available by their respective providers pursuant to the terms set forth on the applicable sites on which they are made available.*

The projection approach here is a Bayesian MCMC Sampling algorithm that projects the infection rate and recovery rate over time to be fed into a conventional Epidemiological SIR (Susceptible-Infected-Recovered) Model for the Simulation. The different social distancing scenarios correspond to the behavior of the infection rate in these respective scenarios. Thus, the model projection essentially looks at a what-if scenario for the projected confirmed cases if certain lockdown measures are to be lifted at the time of projection.



The simulation is regularly validated against the actual reported cases, and for 5 out of 6 States, the actual cases fall within a 95% confidence interval for the models projections.

The hospitalization projections are calculated by creating a relationship between the first order derivatives of the actual positive cases and the hospitalization rate reported by the CDC for each state on a weekly basis (which can be found for each state on <https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html>). Refer to the image below for further information.



Here, the column ‘cumulative\_hos\_per\_100k’ refers to the weekly cumulative hospitalization rates per 100,000 population for a State. We refer to values published by the CDC up until the 9th of May (for hospitalization projections for the week of 16th May onwards). In order to infer a hospitalization rate for the next few weeks, we look at calculating a ratio of first order derivatives in terms of hospitalizations and confirmed cases for the reported period. In other words, we look at the weekly change in reported hospitalization rates till the 9th of May, as well as the weekly change in the reported confirmed cases till the 9th of May. We then calculate the ratio of change in weekly hospitalizations to the change in weekly confirmed cases. The last value of the reported ratio is then used to show the expected increase/decrease in number of hospitalizations as a factor of the projected confirmed cases, based on which, we can calculate projected hospitalization rates over the next few weeks.

This relationship is then extended through the projection timeline of the confirmed cases.

**Cautions when utilizing this Projection Model:**

* The accuracy and confidence limits of the projections developed by this COVID Case Projection Model depend on the use of the most up to date, publicly available data associated with the area being evaluated. As information about positive case rates for a particular region and social distancing measures in place may change over time, projections will change.
* This COVID Case Projection Model has not been endorsed, assessed or cleared for use by any public health or regulatory authorities for its proposed use.