

Analysis of United States Covid-19 case number trends during late January 2021

Chris Briggs
DSC 680 Summer 2021

Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data.

The data is related to public health. Specifically, I will be looking at the daily case numbers of Covid-19 by state in the United States during the period April 1, 2020 to June 1, 2021.

References:

- <https://www.worldometers.info/coronavirus/>
 - The data source, providing daily covid-19 cases by state
- <https://coronavirus.jhu.edu/map.html>
 - Johns Hopkins covid-19 data resource, as a check on worldometers
- https://en.wikipedia.org/wiki/Logistic_function
 - logistic functions
- <https://www.maa.org/book/export/html/115630>
 - logistical growth models
- <https://www.sciencedirect.com/science/article/abs/pii/S1476945X08000822>
 - on logistic growth with non-constant carrying capacity
- https://phe.rockefeller.edu/next1000/PDF_FILES/cc5.pdf
 - an extension to the logistic growth model, allowing for non-constant carrying capacity (here referred to as the "limit").
- <https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus>
 - information on the multiple-wave nature of the covid-19 pandemic
- <https://www.nature.com/articles/s41598-021-85875-2>
 - information on the multiple-wave nature of pandemics in general
- <http://www.sam.math.ethz.ch/~grsam/SS20/NAII/resources/slides/ODE-Lecture4.pdf>
 - numerical solutions to ordinary differential equations
- <https://slate.com/technology/2021/02/why-are-covid-cases-falling.html>
 - media acknowledgment of the subject phenomenon, and speculation as to the cause

Which Data?

What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

The data is published by worldometers.info. I am interested in the number of new cases per day by state. The data is spread across URLs, one per state, of the form e.g. <https://www.worldometers.info/coronavirus/usa/alabama/> for Alabama. I will programmatically retrieve the data from the pages' html code.

Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.

It is my sense that there was an anomalous drop in cases during the last two weeks of January 2021. By anomalous, I mean that previous trends in the rise and fall of cases per state were upset widely, and that this turn is significant enough to demand an explanation besides chance. Understanding the course of the pandemic, including uncovering unexpected turns in the case numbers, is an important component for determining public health policy during the remainder of this pandemic and during future potential pandemics.

What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

I will fit models to the daily case numbers. I can fit a model by state and then examine the departure from each of these individual models per state during late January 2021, or I can determine a single model of uniform best fit and then measure the departure from this model during late January 2021. A single model would require phase shifting to account for different timelines of outbreaks in different states. A single model would also require amplitude shifting to account for different population sizes. Essentially, the single model would incorporate these as parameters.

Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

I do not anticipate any challenges acquiring the data, although if worldometers.info refuses requests from script modalities, then acquiring the data may take some work. I anticipate the greatest challenges to lie in determining what sort of model or models will be appropriate for the daily case numbers. Working from theory, the daily cases should probably follow a logistic growth curve. Because the virus appears to be affected by seasonal conditions, the underlying carrying capacity of the curve will vary by season, which explains the characteristic "multiple

wave" phenomenon. Modeling is further complicated by the unequal application of public health measures (e.g. lockdowns). Limitation: if an anomalous effect is discovered, the study will **not** establish the cause for such an anomaly.

Concluding Remarks

Tie it all together. Think of this section as your final report's abstract.

The Covid-19 pandemic wreaked havoc on public health, in addition to devastating world economies as a result of reactive measures. There has been much discussion and speculation trying to determine what course the pandemic will take next, as this knowledge is important to determining public health policy. In particular, public health restrictions (e.g. lockdowns) may be prudent in advance of cases increasing, but are damaging and should not be done needlessly. An important part of prediction is understanding the clues we can find in past data. Around late January 2021, there appeared to be a sharp drop-off in case numbers all around the United States. This study will examine this drop-off in the context of previous trends by state, and quantify the extent to which this drop-off was anomalous.