More hjstein@gmail.com New Post Design Sign Out

Harvey J. Stein

Essays and commentary from a member of the quantitative community

Tuesday, April 21, 2020

Covid-19 NYC Stats - Not What They Seem

I always say that before developing a model, it's important to look at and understand the data. It's all too easy to jump into modeling and analyses before doing so, and those analyses are likely to be flawed if one ignores the all important step of studying and understanding the data itself.

I'm not immune to this tendency - I fell into that trap in my 4/7 analysis of the NYC Covid-19 data! According to the data at that time (collected from the NYC COVID-19 data website, I suggested that the growth rate in the number of cases had already dropped to zero around 3/30. That wasn't the case.

It's true that NYC posts a full, updated history each day, and they explicitly state that due to "delays in reporting, the most recent data may be incomplete". But the latest data report doesn't allow one to account for the delays in reporting. I had thought the drop-off at the end of the data was due to reporting delays, but it turns out the delays are far larger than that.

How can we tell, if the history isn't posted? Well, it would be nice if the history was easily accessible, but since the data is updated in a git repository, the previous reports are recorded and available. It just takes some work to extract them. So, I forked the above repository and did this work. Here's one result:

Subscribe To

Posts

Comments

Follow by Email

Email address... Submit

Followers

Followers (3)

Follow

Blog Archive

▼ 2020 (2)

▼ April (2)

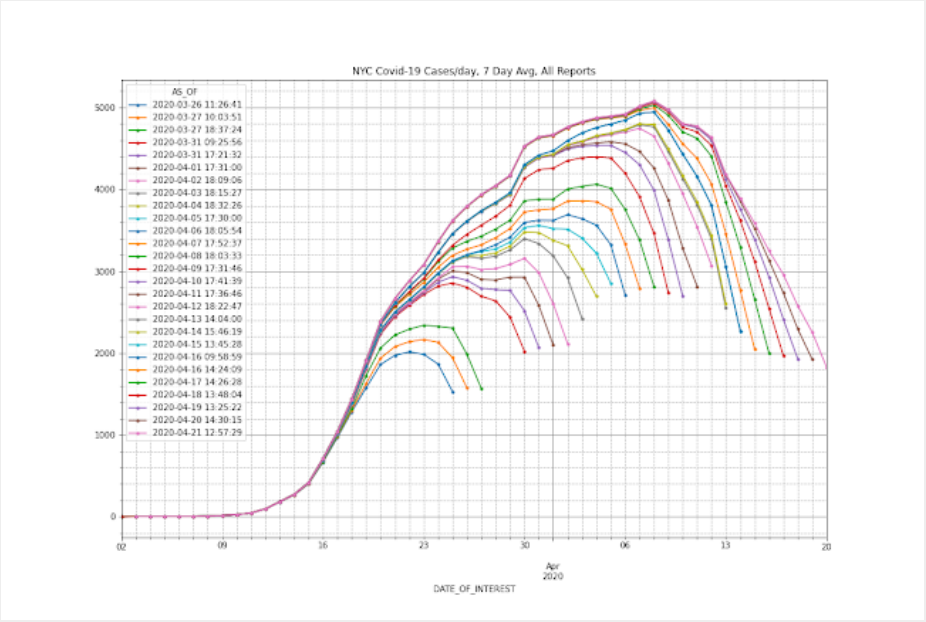
Covid-19 NYC Stats - Not What They Seem

Covid-19 NYC Stats - A Ray of Hope

► 2015 (1)

► 2012 (1)

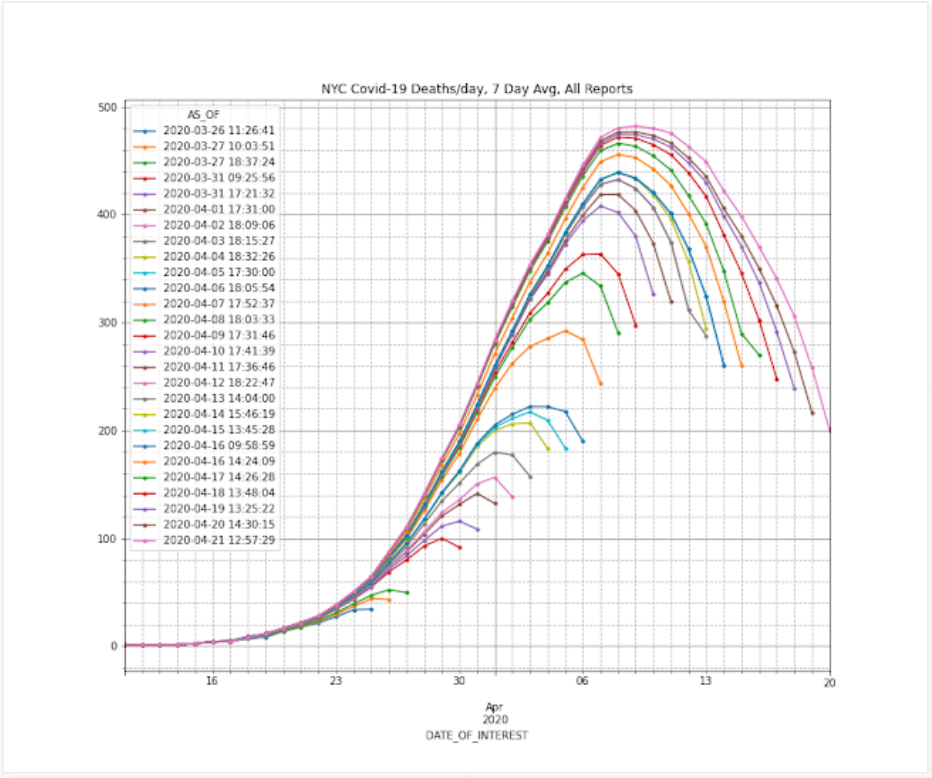
► 2011 (4)



Now you can see that some reports are taking quite long to be incorporated. For example, the 4/15 report updated the number of cases as far back as 3/22! And, while the 4/7 through 4/11 reports flatten out with a growth rate close to zero from 4/1 to 4/6, the 4/12 report changes that, showing the number of cases per day continuing to grow through 4/7!

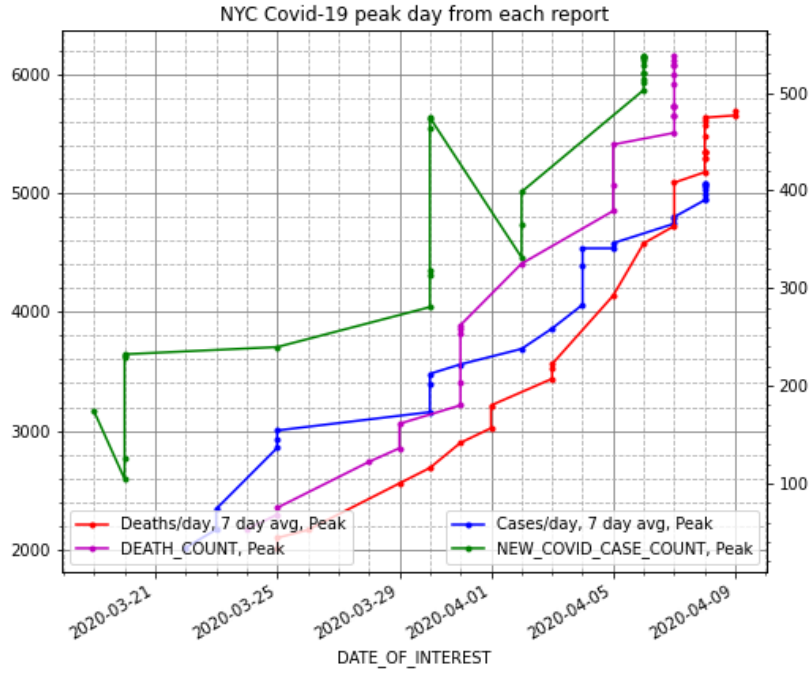
So how do we determine if we're past the peak? If we had the individual reports from each reporter, we could simply look at how the number of cases changes at a given location. That would be ideal. Lacking that, we have to try work around the fact that missing data makes it erroneously appear that the number of cases is dropping.

One possibility is to look at the mortality data instead. The peak in the deaths per day should be about a week or two later than the peak in the number of people infected per day, so the signal will be late, but one would expect this data to be better reported. Here it is (with 7 day averaging):



Even in the mortality data, some reports restate as much as two weeks of the past. So it's also unclear how long we have to wait before we can rely on the numbers.

Another possibility is to analyze the change in the peaks from report to report. The idea is that if we've reached the peak, then all the future reports should show that same peak. So we look at the date of the peak for each report. The longer that date remains the same, the more likely it is that that date was the peak. So let's look at that for the raw data and the 7 day averages for both the infections as well as for the mortalities:



We can see here that the raw cases per day peak has been on 4/5 for about a week, and the mortality rate peak has held fast at 4/6. But between the cyclicity of the data and the missing reports, we might not trust that so much. Looking at the rolling 7 day averages, we get 4/7 and 4/8, but the mortality peak hasn't been maintained for very long.

So, maybe we've already reached the peak, at least in reported cases per day. At least, I hope we did. But without better data, it's hard to say. It's unfortunate that the full data set isn't readily available, so we're left guessing. And, as is usually the case, it's the data...

Posted by Harvey Stein at 5:05 PM

No comments:

Post a Comment

Enter your comment...

Comment as: Harvey Stein (Go Sign out

Publish Preview Notify me

Links to this post

Create a Link

Home Older Post

Subscribe to: Post Comments (Atom)

Simple theme. Theme images by luoman. Powered by Blogger.