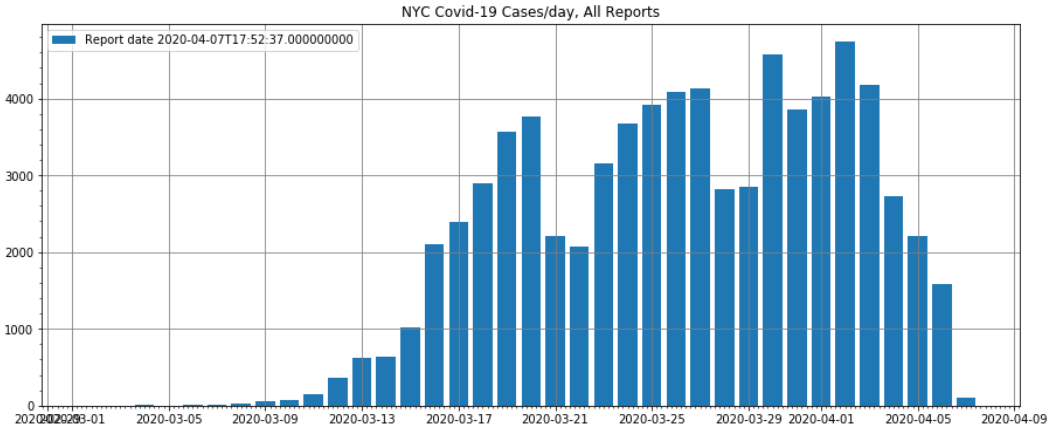
COVID-19: THE DATA ABUSE PANDEMIC

Harvey J. Stein
hjstein@bloomberg.net

Head, Quantitative Risk Analytics
Bloomberg L.P.

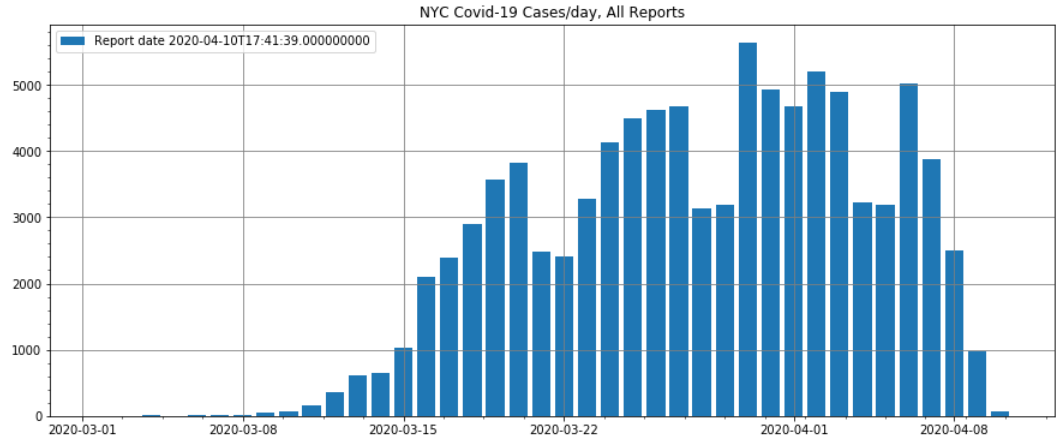
Bloomberg Quant Seminar
May, 2020
DRAFT SLIDES

WHAT THEY SAID IN NYC



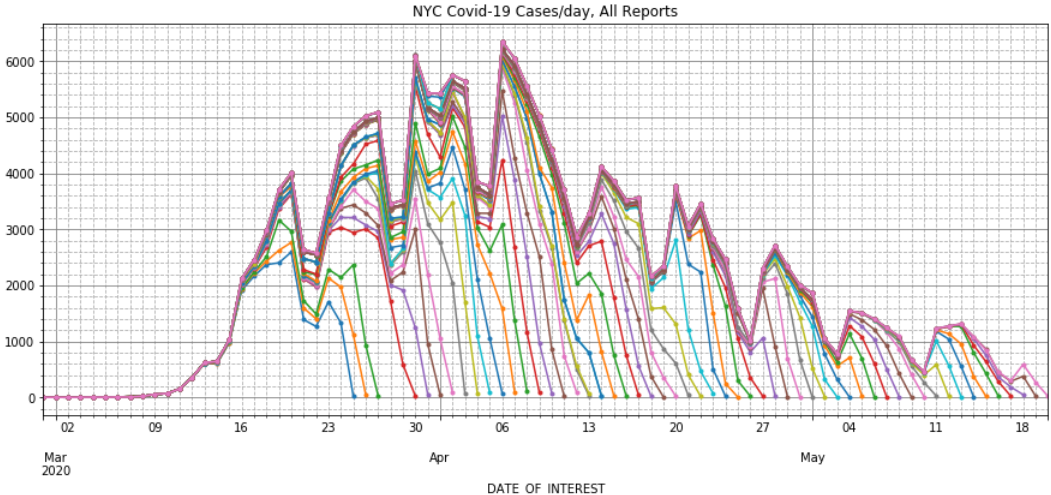
Cases/day are dropping off

THREE DAYS LATER

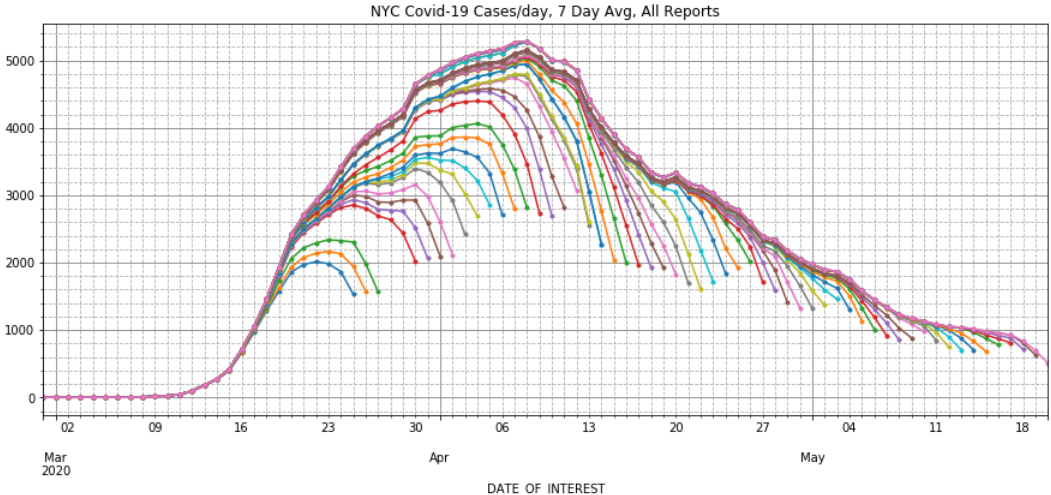


Similar, but dropping off later!

WHAT THEY DIDN'T SAY



AFTER APPROPRIATE SMOOTHING



ABUSES

What's going on?

- Different sites report cases at different times

Impact:

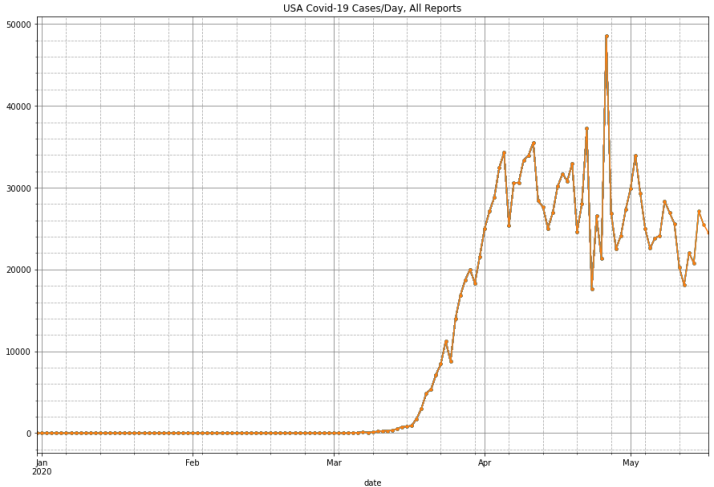
- Reporting delays can account for 30% changes as much as 3 weeks later
- Latest counts for each date are reported – history is harder to extract
- Low weekend counts obscure trends
- Reporting aggregates instead of per site data makes analysis difficult

But this is the good case!

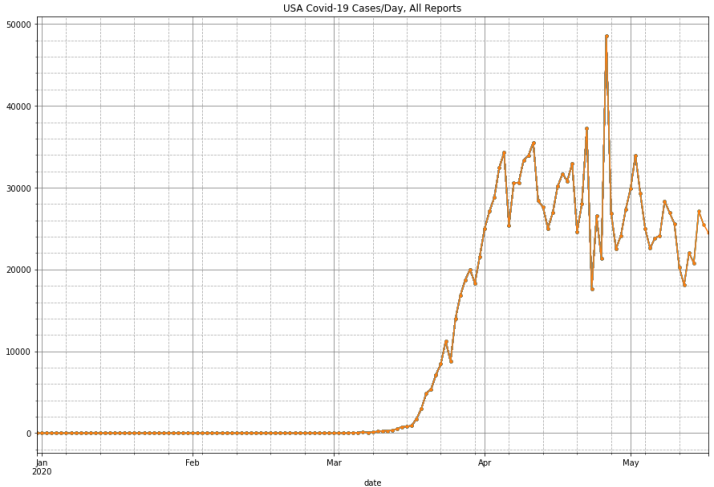
See:

- *COVID-19 NYC Stats – A Ray Of Hope* [Ste20d]
- *COVID-19 NYC Stats – Not What They Seem* [Ste20e]

WHAT THEY SAID FOR THE USA



WHAT THEY DIDN'T SAY



ABUSES

What's going on?

- Recording by report date instead of by incident date!

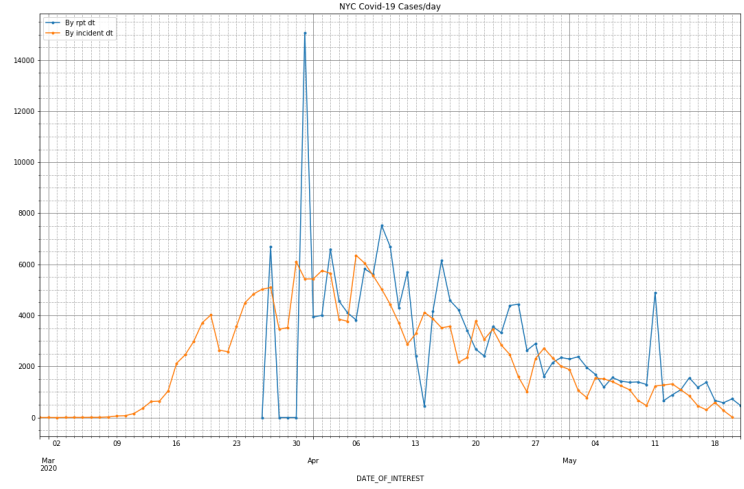
Impact:

- No record of when incidents actually occurred
- Misstates ramp up to peak
- Delays observed peak
- Overstates incidents post-peak
- News reports of yesterday's numbers are deceptive

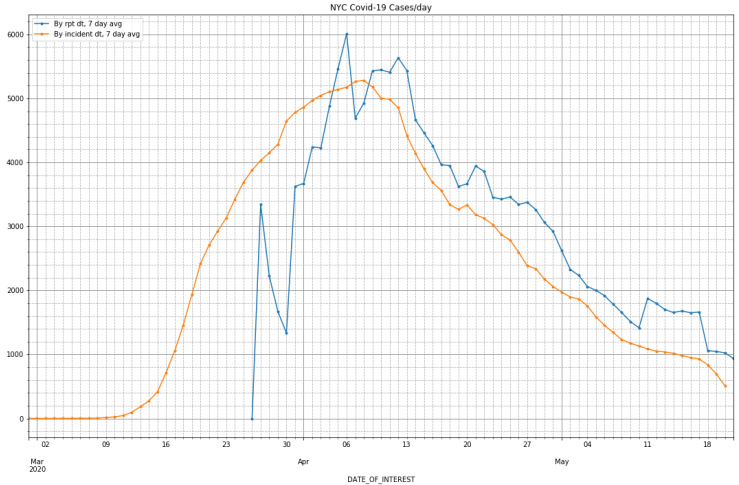
Recording by report date is rampant – all major sources do it:

- Our World in Data repository (data from the European CDC)
- Johns Hopkins University COVID-19 repository
- New York Times COVID-19 repository

IMPACT


NYC Covid-19 Cases/day
By rpt dt
By incident dt
14000
12000
10000
8000
6000
4000
2000
0
02
09
16
23
30
06
13
20
27
04
11
18
Mar
2020
Apr
May
DATE_OF_INTEREST

IMPACT


NYC Covid-19 Cases/day
By rpt dt, 7 day avg
By incident dt, 7 day avg
6000
5000
4000
3000
2000
1000
0
02
09
16
23
30
06
13
20
27
04
11
18
Mar
2020
Apr
May
DATE_OF_INTEREST

SUMMARY

What now?

- Know your data!
- Be aware that recent reports are inaccurate – either missing data or misrepresenting counts
- Account for this in your modeling
- Ask data sources to collect by incident date
- Ask data sources to make the report history more readily available
- Ask data sources to publish on a per site basis instead of in aggregate
- Investigate actual meaning of data – classification of COVID-19 hospitalizations, deaths due to COVID-19, ...

See COVID-19 Data Collection – Garbage In, Garbage Out [Ste20c])

References

[OWI20] OWID. *Data on COVID-19 (coronavirus) confirmed cases, deaths, and tests, All countries, Updated daily by Our World in Data*. 2020. URL: https://github.com/owid/covid-19-data.

[Ste20a] Harvey Stein. *Analysis of NYC COVID-19 infection rate*. Fork of https://github.com/nychealth/coronavirus-data. 2020. URL: https://github.com/hjstein/coronavirus-data.

[Ste20b] Harvey Stein. *Analysis of Our World of Data's COVID-19 dataset*. Fork of https://github.com/owid/covid-19-data. 2020. URL: https://github.com/hjstein/covid-19-data.

REFERENCES

[Ste20c] Harvey Stein. *COVID-19 Data Collection – Garbage In, Garbage Out*. Blog, Harvey J. Stein, Essays and commentary from a member of the quantitative community. Apr. 2020. URL: http://hjstein.blogspot.com/2020/05/covid-19-data-collection-garbage-in_33.html.

[Ste20d] Harvey Stein. *COVID-19 NYC Stats – A Ray Of Hope*. Blog, Harvey J. Stein, Essays and commentary from a member of the quantitative community. Apr. 2020. URL: https://hjstein.blogspot.com/2020/04/covid-19-nyc-stats-ray-of-hope.html.

[Ste20e] Harvey Stein. *COVID-19 NYC Stats – Not What They Seem*. Blog, Harvey J. Stein, Essays and commentary from a member of the quantitative community. Apr. 2020. URL: https://hjstein.blogspot.com/2020/04/covid-19-nyc-stats-not-what-they-seem.html.

REFERENCES

[Joh20] Johns Hopkins University. *Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE*. 2020. URL: https://github.com/CSSEGISandData/COVID-19.

[New20a] New York City. *NYC Health Coronavirus Data*. 2020. URL: https://github.com/nychealth/coronavirus-data.

[New20b] New York Times. *An ongoing repository of data on coronavirus cases and deaths in the U.S.*. 2020. URL: https://github.com/nytimes/covid-19-data.