COVID-19: THE DATA ABUSE PANDEMIC
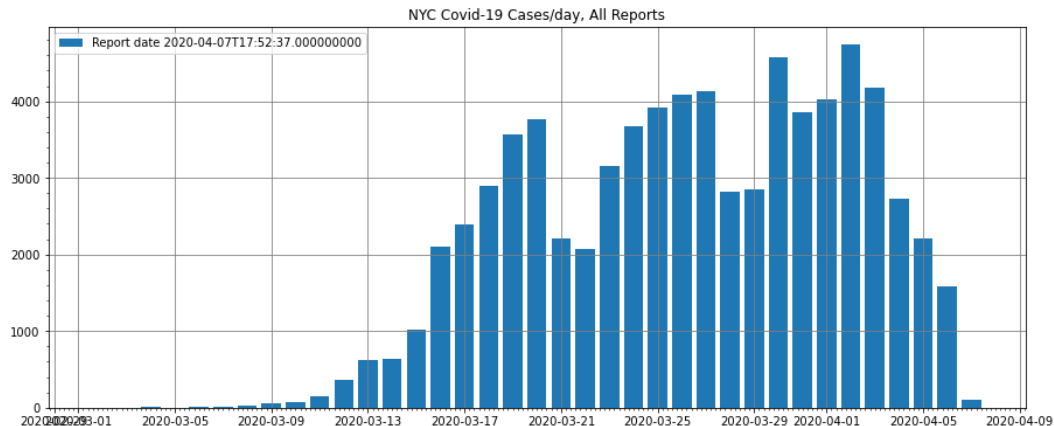
Harvey J. Stein
hjstein@bloomberg.net

Head, Quantitative Risk Analytics
Bloomberg L.P.
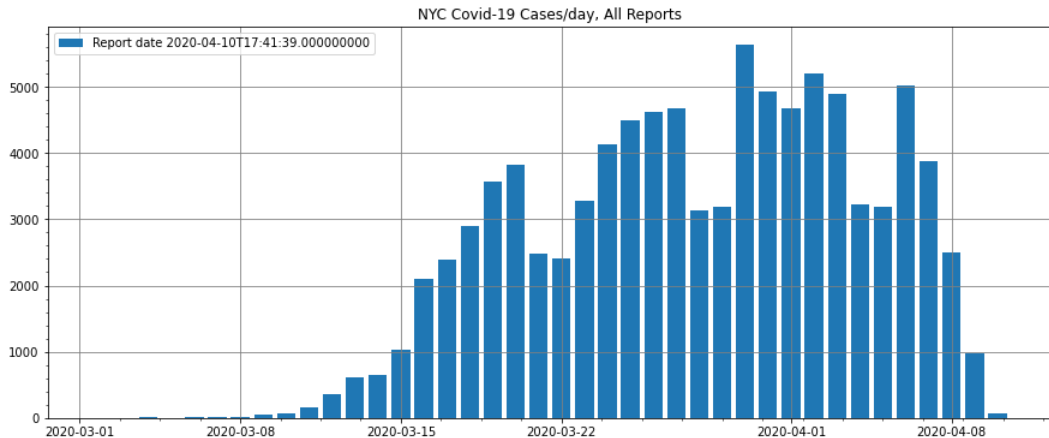
Bloomberg Quant Seminar
May, 2020

Bloomberg

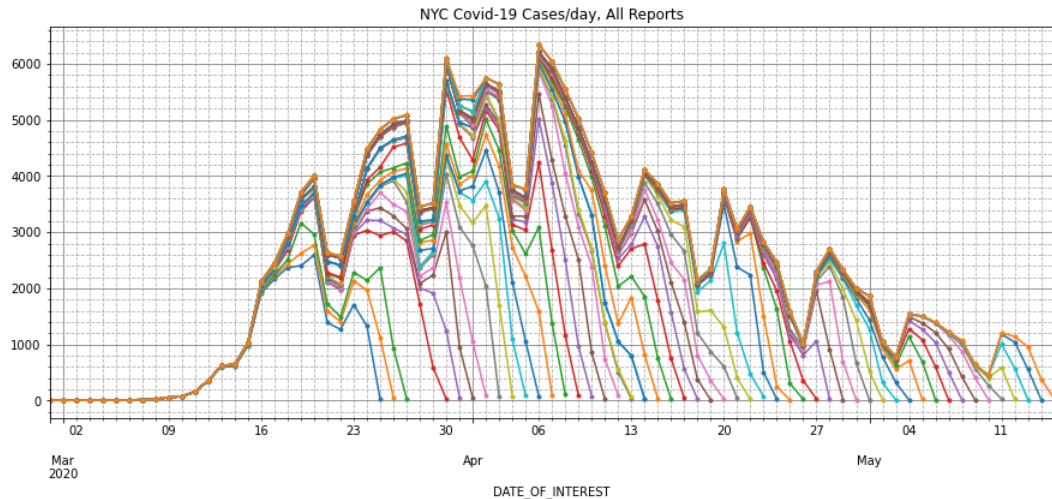# WHAT THEY SAID IN NYC

NYC Covid-19 Cases/day, All Reports

Cases/day are dropping off

THREE DAYS LATER

NYC Covid-19 Cases/day, All Reports

Report date 2020-04-10T17:41:39.000000000

Similar, but not dropping off later!

Bloomberg

WHAT THEY DIDN'T SAY

NYC Covid-19 Cases/day, All Reports

AFTER APPROPRIATE SMOOTHING

NYC Covid-19 Cases/day, 7 Day Avg, All Reports

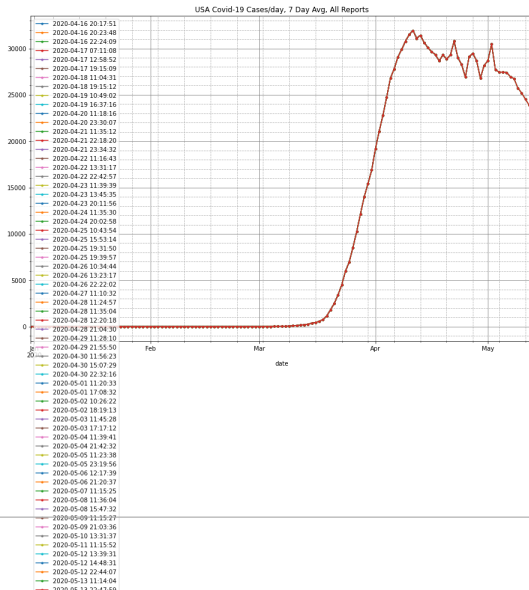Bloomberg
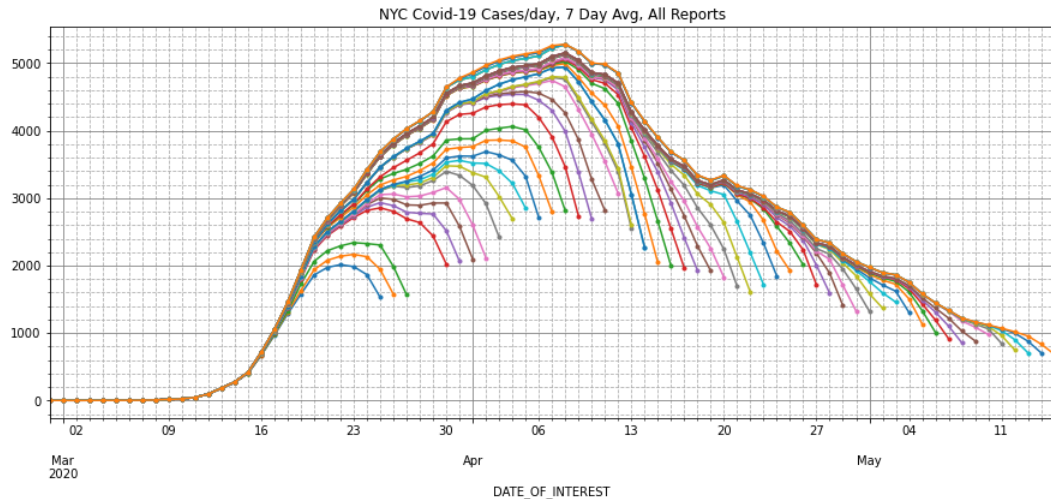
ABUSES

What's going on?
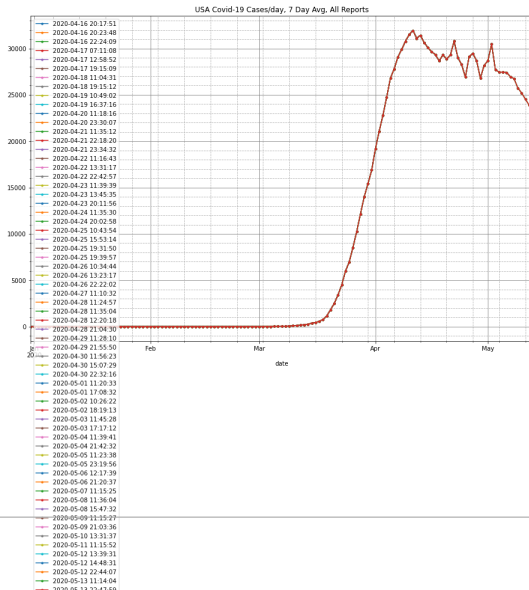- ► Different sites report cases at different times

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!

USA Covid-19 Cases/day, 7 Day Avg, All Reports

USA Covid-19 Cases/day, 7 Day Avg, All Reports

Bloomberg

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
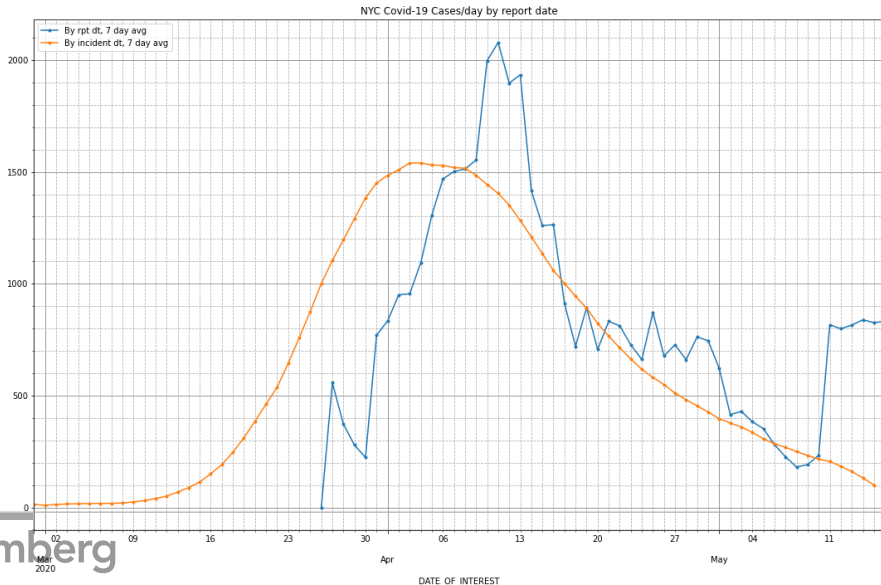
▶ Delays observed peak

▶ Overstates incidents post-peak

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

▶ Our World in Data github 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 report date

IMPACT

NYC Covid-19 Cases/day by report date
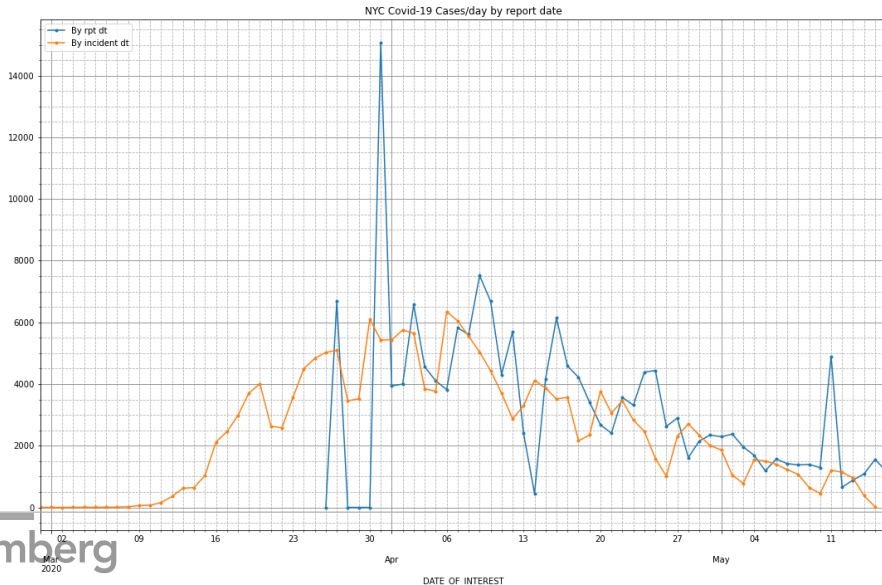
References

[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.

Bloomberg

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.

Bloomberg

REFERENCES

[Ste20c]   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.

[Ste20d]   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.

Bloomberg