COVID-19 Counts in Washington State and Select Other Locations

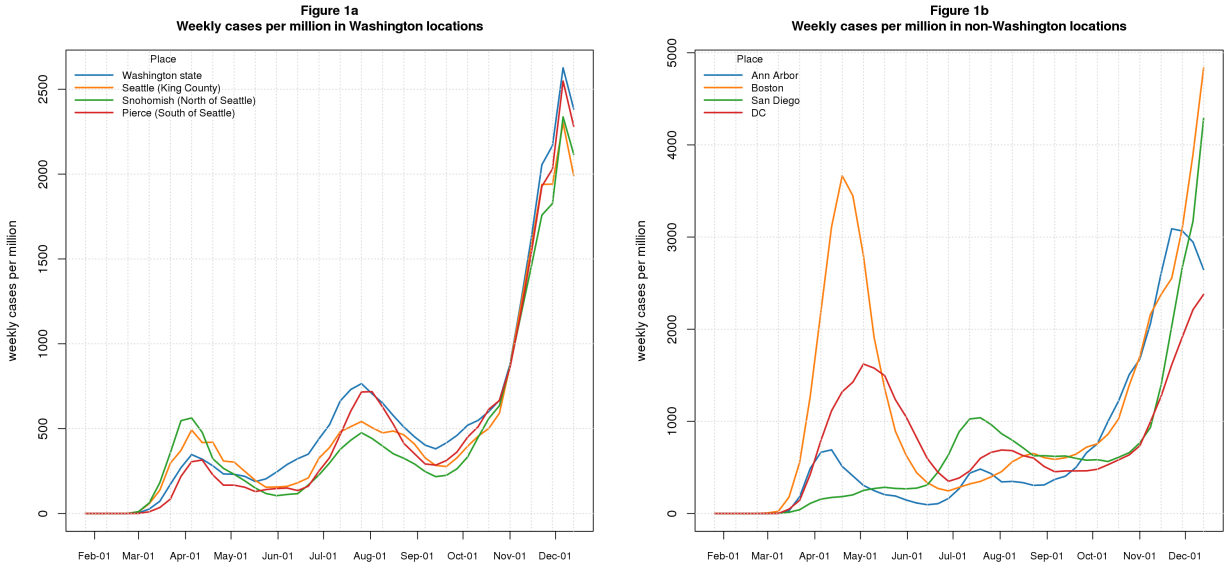
Nathan (Nat) Goodman

December 21, 2020

Counts of cases and deaths are key metrics of COVID-19 prevalence and burden, and are the basis for model-based estimates and predictions of these statistics. I present here graphs showing these metrics over time in Washington state and a few other USA locations of interest to me. I hope to update the graphs weekly.

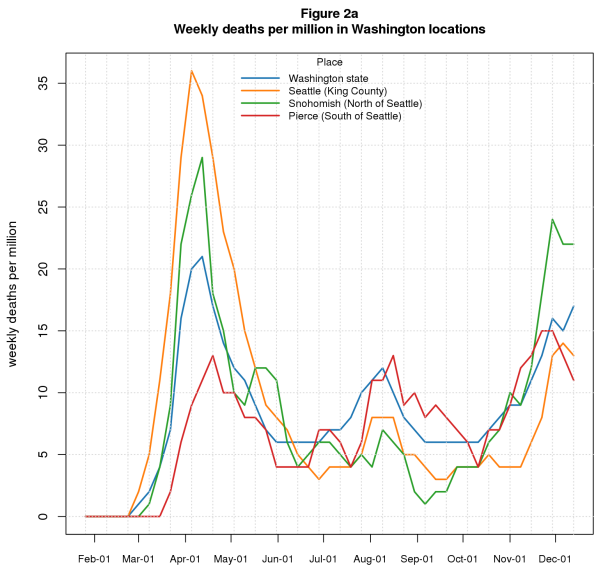
See below for caveats and details. This update does not include Washington state data broken down by ages (Figures 3-5 in previous versions), because the Washington DOH data download was not available in time.

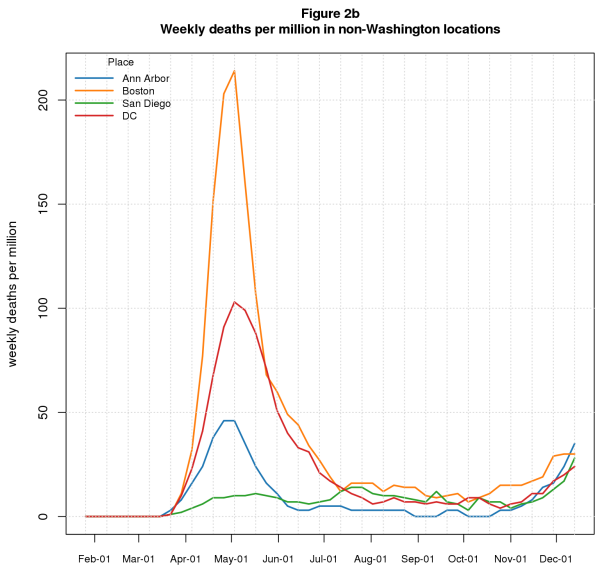
Figure 1a shows *case* counts per million for Washington state as a whole, the Seattle area where I live, and the adjacent counties to the north and south (Snohomish and Pierce, resp.). Figure 1b shows *case* counts for Ann Arbor, Boston, San Diego, and Washington DC. The figures use data from Johns Hopkins Center for Systems Science and Engineering (JHU), described below.



In Washington state (Figure 1a), cases seem to have hit their peak after a prolonged increase and are heading down (but see caveats below). In non-Washington locations (Figure 1b), cases are increasing everywhere except Ann Arbor. The decline in Ann Arbor has persisted for three weeks in a row and may even be accelerating. This very good news may be due to the University of Michigan ending on-site teaching for most students.

Figures 2a-b show *deaths* per million for the same locations.





The Washington data (Figure 2a) shows three waves. The trend is continuing up for the state as a whole (the blue line) but is flat or declining for the specific locations shown here (but heed the caveats below). The second peak was thankfully lower than the first; the third peak is higher than the second but remains lower than the first. Early peaks dominate the non-Washington data (Figure 2b) making it hard to see details later on, but the rates are heading up. There's a hint of a flattening (or, at least, slower increase) in Boston. When comparing Figures 2a and 2b, please heed the difference in y-scale: the current Washington rates (10-20 per million) are similar to the non-Washington rates.

Caveats

1. As of this reporting period, Washington State Department of Health (DOH) changed its criteria for reporting cases and deaths resulting in reduced counts. Some of the reduction is due to lags in updating the database for the new procedure. It's unclear when the database will catch up and whether the final updated counts will revert to the levels reported in the previous database. Other data sources, including JHU, ultimately rely on DOH for their Washington state data, errors in the DOH data affect them all.
2. The term *case* means a person with a detected COVID infection. Until this week, the Washington state data limited this to "confirmed cases", meaning people with positive molecular COVID tests but going forward they plan to separate out "probable cases". Other states already do this, but the data source I use here only includes "confirmed cases" (or so I believe based on the name of the file I download).
3. Detected *cases* undercount actual cases by an unknown amount. As testing volume increases over time, it's reasonable to expect the detected count to get closer to the actual count. Some of the increase in *cases* we see in the data is due to this artifact. Modelers attempt to correct for this. I don't include any such corrections here.
4. The same issues apply to *deaths* to a lesser extent, except perhaps early in the pandemic.
5. The geographic granularity in the underlying data is *state* or *county*. I refer to locations by city names reasoning that readers are more likely to know "Seattle" or "Ann Arbor" than "King" or "Washtenaw".
6. The date granularity in the graphs is weekly. The underlying JHU data is daily; I sum the data by week before graphing.
7. I smooth the graphs using a 3-week rolling mean for visual appeal. This is especially important for the *deaths* graphs where the counts are so low that unsmoothed week-to-week variation makes the graphs

hard to read.

Data Sources

Washington State Department of Health (DOH)

DOH provides three COVID data streams.

1. Washington Disease Reporting System (WDRS) provides daily "hot off the presses" results for use by public health officials, health care providers, and qualified researchers. It is not available to the general public, including yours truly.
2. COVID-19 Data Dashboard provides a web graphical user interface to summary data from WDRS for the general public. (At least, I think the data is from WDRS - they don't actually say).
3. Weekly data downloads (available from the Data Dashboard web page) of data curated by DOH staff. The curation corrects errors in the daily feed, such as, duplicate reports, multiple test results for the same incident (e.g., initial and confirmation tests for the same individual), incorrect reporting dates, incorrect county assignments (e.g., when an individual crosses county lines to get tested).

In past, DOH updated the weekly data on Sundays but they now plan to release the updates on Mondays. As of Tuesday December 22, the update scheduled for December 21 had not yet been released and I decided to omit the DOH data from this version of the document.

The weekly downloads lag behind the daily feed causing data for the last few weeks to be incomplete. In versions of this document prior to December 14, I corrected for this undercount by extrapolating data from the preceding six weeks using a linear model (R's `lm`). Performance of the extrapolation became erratic in the most recent data, and I turned it off in the December 14 version. The point is moot in the present version since the DOH data is not available.

The weekly DOH download reports data by age group: 20-year ranges starting with 0-19, with a final group for 80+.

The DOH download includes data on hospital admissions in addition to *cases* and *deaths*, although I don't show this data here.

Johns Hopkins Center for Systems Science and Engineering (JHU)

JHU CSSE has created an impressive portal for COVID data and analysis. They provide their data to the public through a GitHub repository. The data I use is from the csse_covid_19_data/csse_covid_19_time_series directory: time_series_covid19_confirmed_US.csv for *cases* and time_series_covid19_deaths_US.csv for *deaths*.

JHU updates the data daily. I usually download the data on Mondays to align with the DOH weekly data drops.

Other Data Sources

I use two other COVID data sources in my project although not in this document.

- New York Times COVID Repository. The file I download is us-counties.csv. Like Washington DOH and JHU, NYT has county-level data. Unlike these, it includes "probable" as well as "confirmed" cases and deaths; I see no way to separate the two categories.
- COVID Tracking Project. This project reports a wide range of interesting statistics (negative test counts, for example), but I only use the *case* and *death* data. It does not provide county-level data so is not useful for the non-Washington locations I show. The file I download is https://covidtracking.com/

data/download/washington-history.csv. I use this only as a check on the state-level Washington data from the other sources.

The population data used for the per capita calculations is from Census Reporter. The file connecting Census Reporter *geoids* to counties is the Census Bureau Gazetteer.

Comments Please!

Please post comments on Twitter or Facebook, or contact me by email natg@shore.net.