

# COVID-19 Counts in Washington State and Select Other Locations

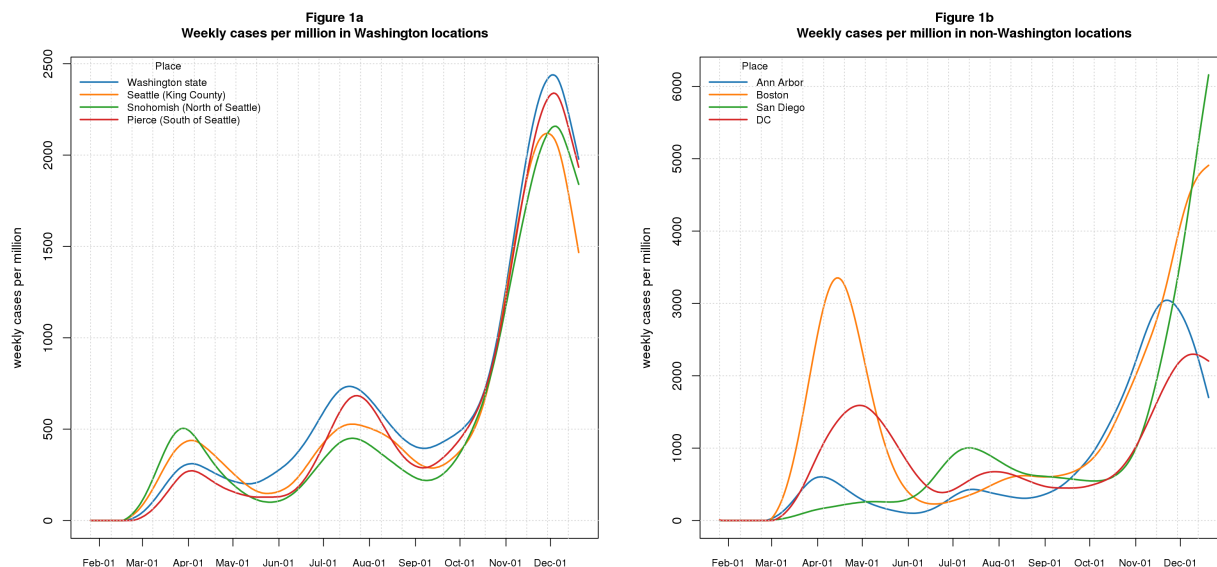
*Nathan (Nat) Goodman*

*December 30, 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.*

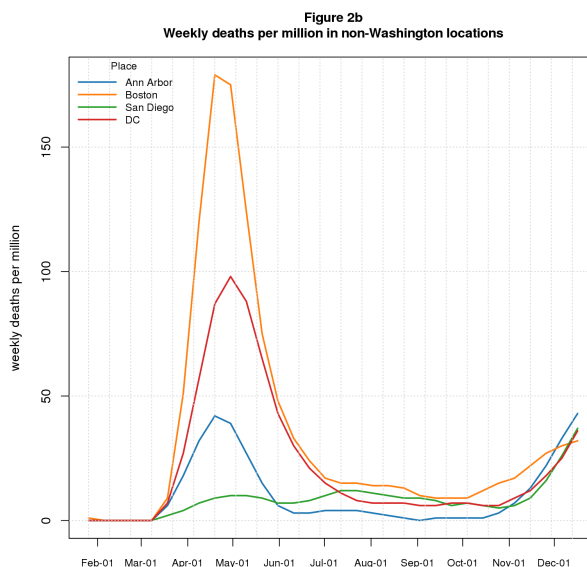
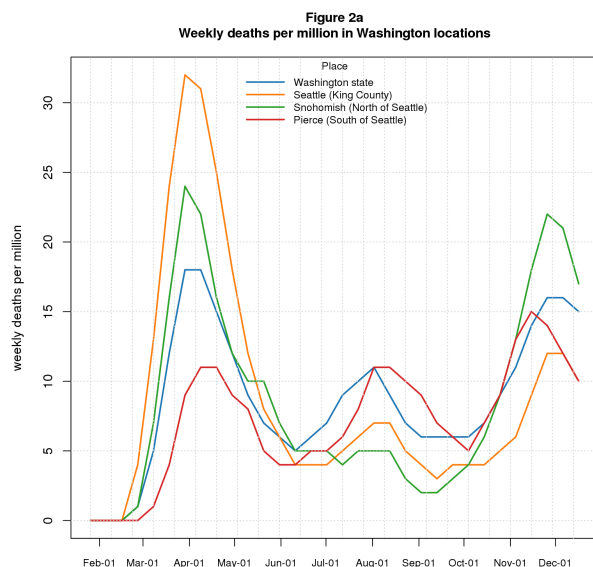
See [below](#) for caveats and details. Starting with this version, I am posting updates on Wednesdays instead of Mondays to accommodate the new [Washington DOH](#) data release schedule.

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 in Boston and San Diego, but declining in Ann Arbor and DC. The decline in Ann Arbor has persisted for four weeks in a row: this very good news may be due to the University of Michigan ending on-site teaching for most students. The decline in DC is slight; we'll know in a few weeks whether it's real.

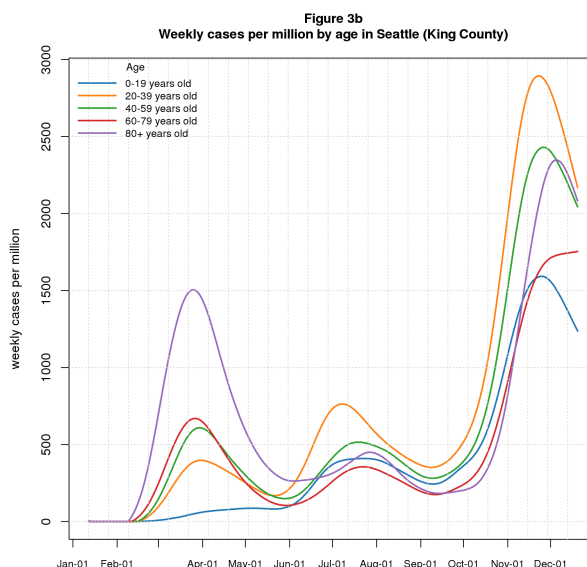
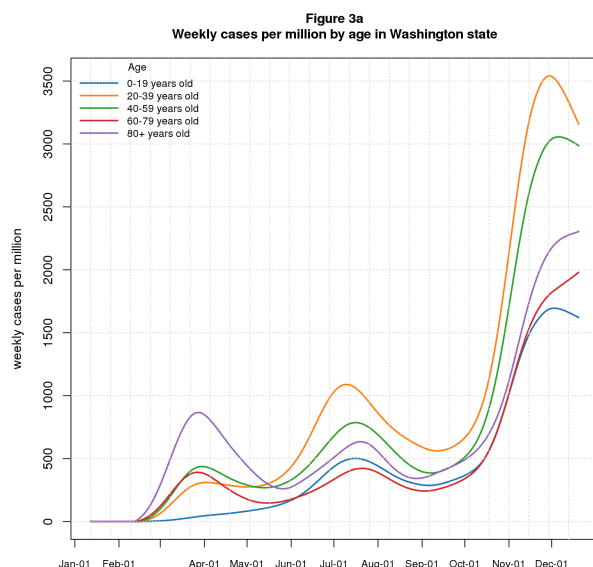
Figures 2a-b show *deaths* per million for the same locations.

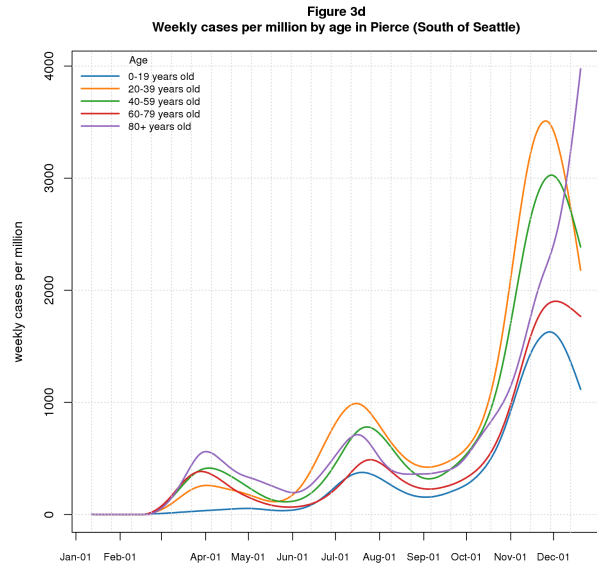
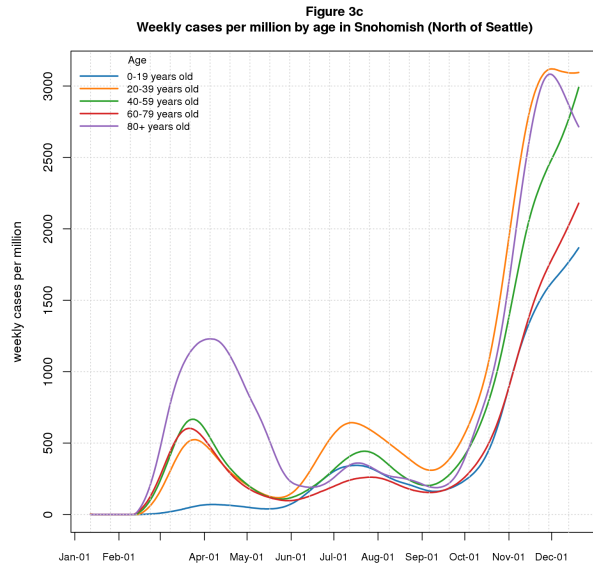


The Washington data (Figure 2a) shows three waves. We may have reached the peak of the third. This is great news heed the [caveats](#) below. The second peak was thankfully lower than the first; the third peak is higher than the second but remains lower than the first. Early peaks dominate the non-Washington 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 note the difference in y-scale: the current Washington rates (10-20 per million) are similar to the non-Washington rates.

The next graphs show the Washington results broken down by age. This data is from Washington State Department of Health (DOH) weekly downloads, described [below](#). An important caveat is that the DOH download systematically undercounts events in recent weeks due to manual curation. I extrapolate data for late time points as discussed [below](#).

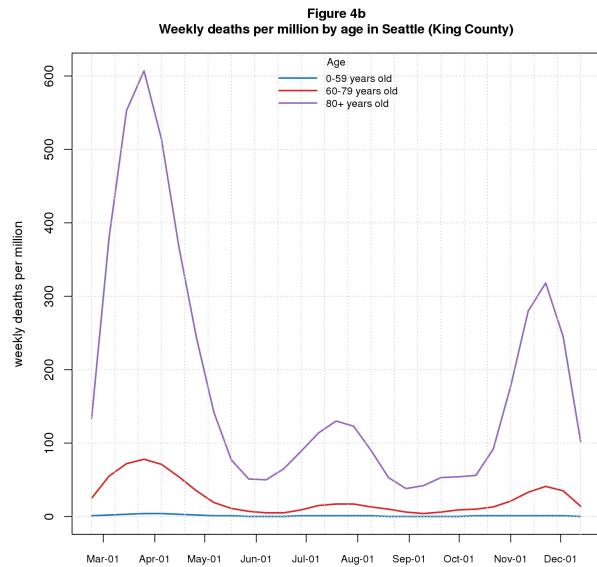
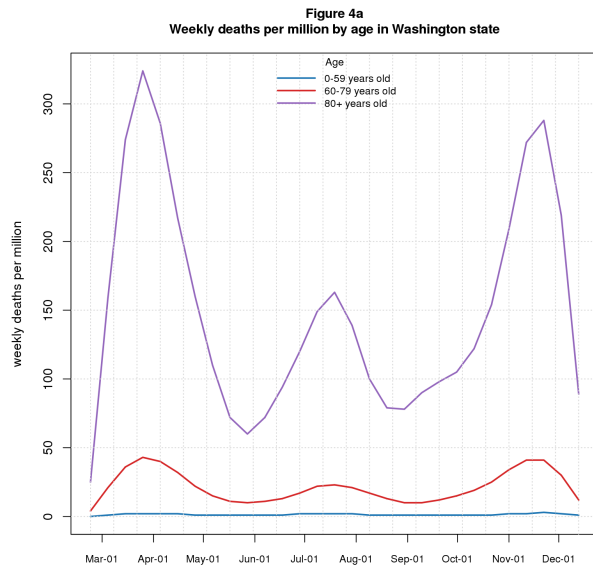
Figures 3a-d are *cases*. The graphs are split into 20-year age ranges starting with 0-19, with a final group for 80+.

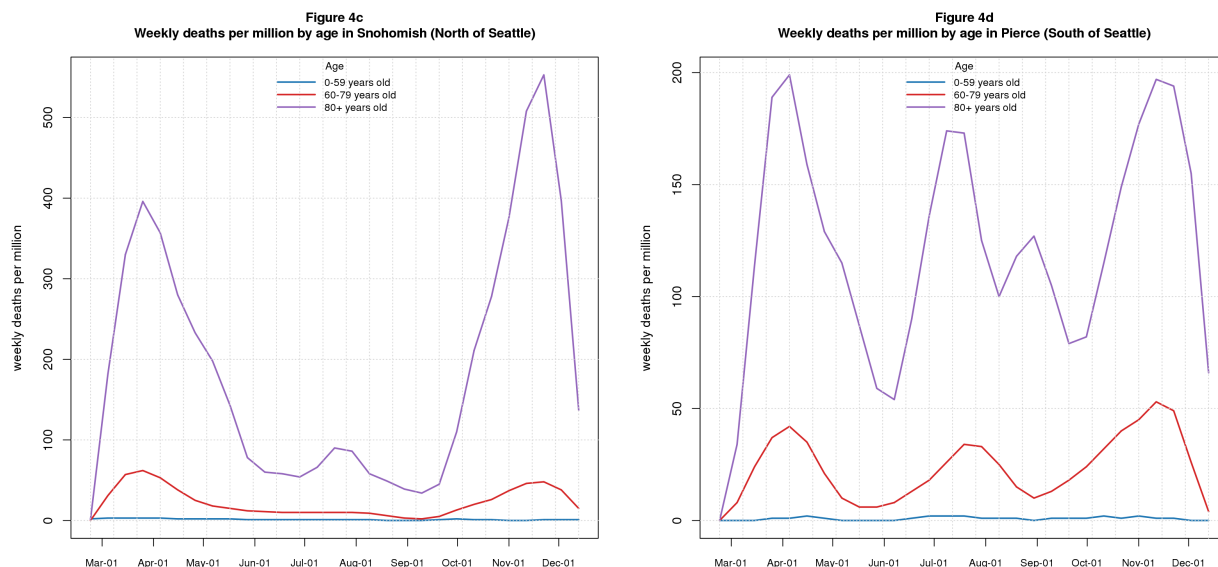




Early on, the pandemic struck older age groups most heavily. Over time, *cases* spread into all age groups, even the young. During the second wave, older groups did better in most locations with young adults (20-39 years) becoming the most affected group. The third wave swept into all age groups with young and middle aged adults (20-39 and 40-59 years) leading the surge.

Figures 4a-d are *deaths*. These graphs aggregate 0-59 into a single group, since the death rate in these ages is near 0.





The shocking devastation of the 80+ age group early in the pandemic jumps off the page. The early death rate for this group in Seattle (King County) reached over 600 per million reflecting the early outbreak at a long term care facility in the area. Statewide, the death rate in the 80+ group shows three waves. Deaths in Snohomish (north of Seattle) for the 80+ group peaked early in the pandemic, then declined and stayed fairly low, but now have climbed back up to nearly the first peak; some of the increase reflects an outbreak in a long term care facility in the county. Deaths in Pierce (south of Seattle) for this group are more variable than elsewhere, possibly reflecting delays in reporting deaths to DOH and political turmoil in the county department responsible for reporting this data. The overall rate seems fairly steady at a level near or slightly below the second and third King county peaks. The very low counts at the latest time points are probably due to reporting lags caused by the the new DOH procedures as discussed [below](#).

## Caveats

1. As of approximately December 20, 2020, [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 the recent reporting change, Washington DOH 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 smoothing spline (R's `smooth.spline`) 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. In previous versions of the document, I used a 3-week rolling mean for this purpose.

## 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 as of December 22, 2020 they switched to Mondays

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, and I turned it off in the December 14 version. In the intervening weeks, I improved the model and have turned it back on.

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](mailto:natg@shore.net).