

COVID-19 Counts in Washington State and Select Other Locations

Nathan (*Nat*) Goodman

February 3, 2021

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 update the graphs and this write-up weekly. Previous versions are [here](#).

See [below](#) for caveats and details. I originally posted updates on Mondays but have switched to Wednesdays to accommodate the recently changed [Washington DOH](#) data release schedule.

Figures 1a-d show *case* counts per million for several Washington and non-Washington locations. Figures 1a-b (the top row) shows smoothed data (see details [below](#)); Figures 1c-d (the bottom row) overlay raw data onto the smoothed. The Washington locations are the entire state, the Seattle area where I live, and the adjacent counties to the north and south (Snohomish and Pierce, resp.). The non-Washington locations are Ann Arbor, Boston, San Diego, and Washington DC.

The figures use data from Johns Hopkins Center for Systems Science and Engineering (JHU), described [below](#). When comparing the Washington and non-Washington graphs, please note the difference in y-scale: the current Washington rates (1000-3500 per million) are similar to the rates in Ann Arbor and Washington DC, and well below the rates in Boston and San Diego.

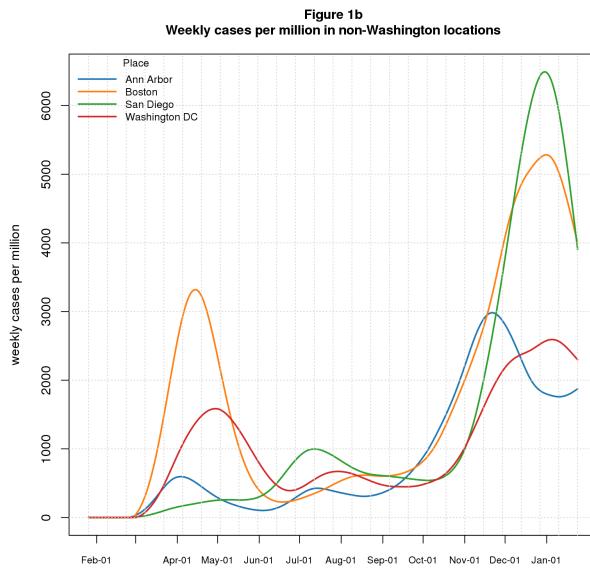
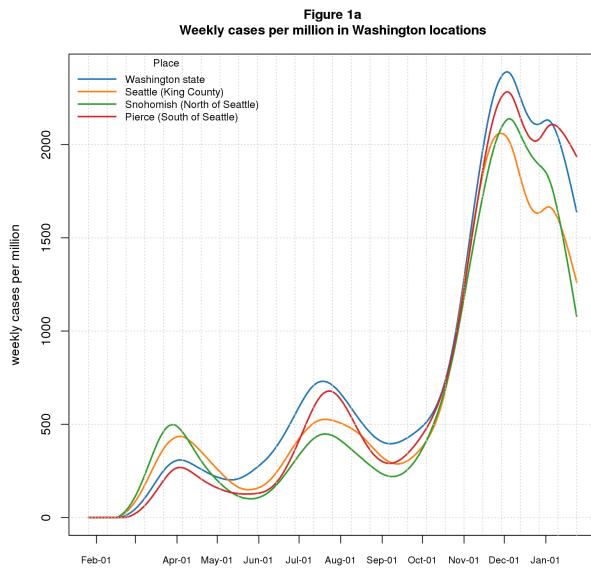


Figure 1c
Weekly cases per million in Washington locations showing raw data

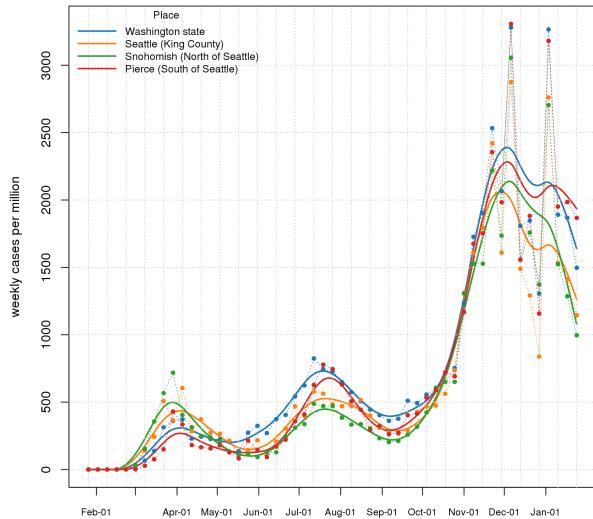
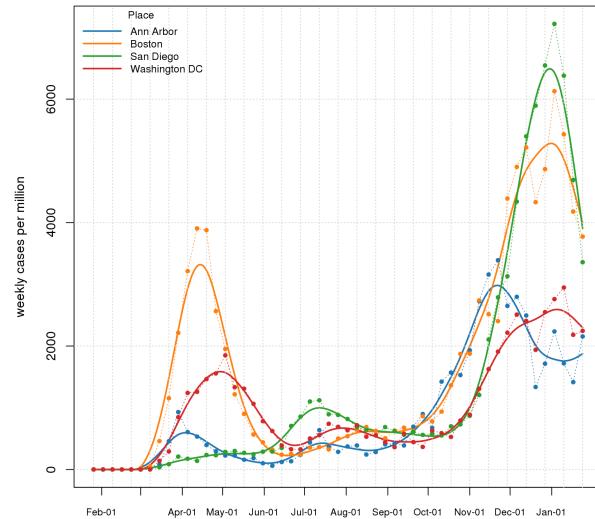


Figure 1d
Weekly cases per million in non-Washington locations showing raw data



From the smoothed data (Figure 1a) it looks like cases in Washington state have hit their peak and are heading down. The raw data (Figure 1c), though ragged, suggests the downward trend may be real.

In non-Washington locations, the smoothed data (Figure 1b) shows cases below their recent peaks everywhere, and downward trends everywhere except Ann Arbor. The raw data (Figure 1d) suggests the reduction is real. The upward bend in Ann Arbor may be an artifact of the smoothing.

Figures 2a-d show *deaths* per million for the same locations. When comparing the Washington and non-Washington graphs, again please note the difference in y-scale: the current Washington rates (20-50 per million) are similar to the non-Washington rates.

Figure 2a
Weekly deaths per million in Washington locations

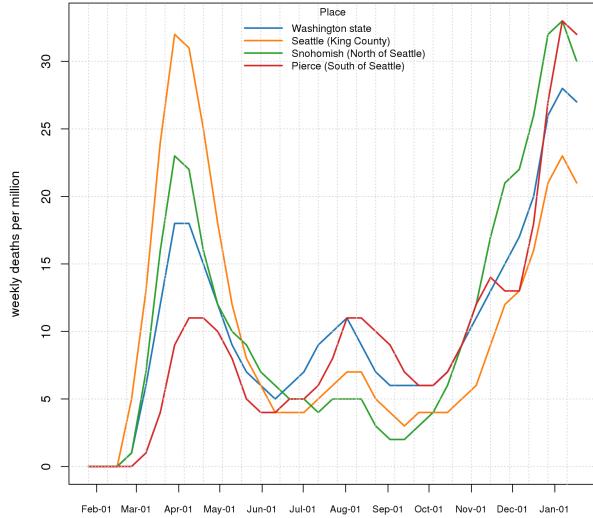


Figure 2b
Weekly deaths per million in non-Washington locations

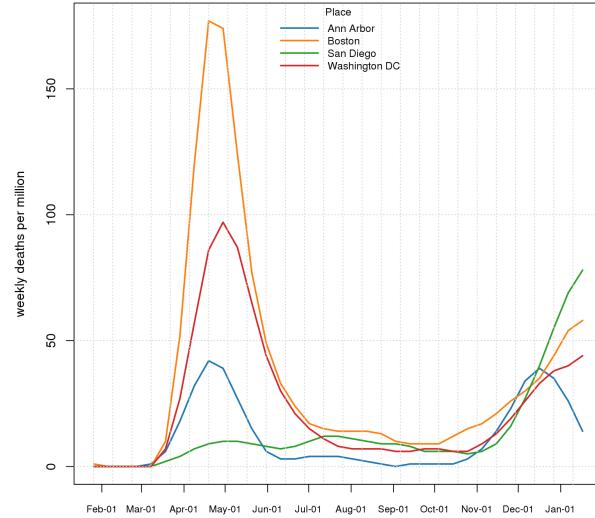


Figure 2c
Weekly deaths per million in Washington locations showing raw data

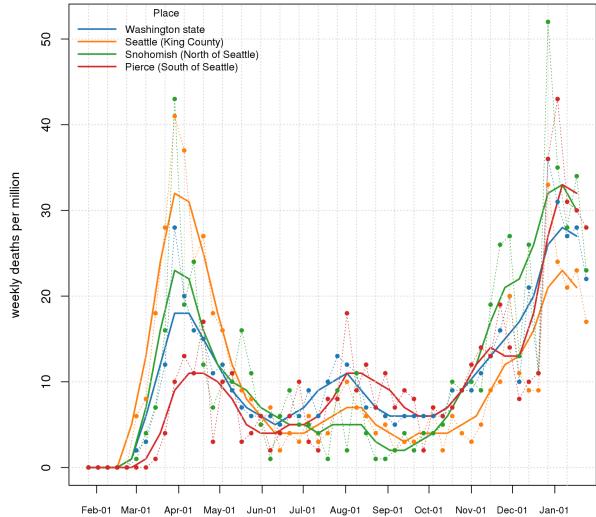
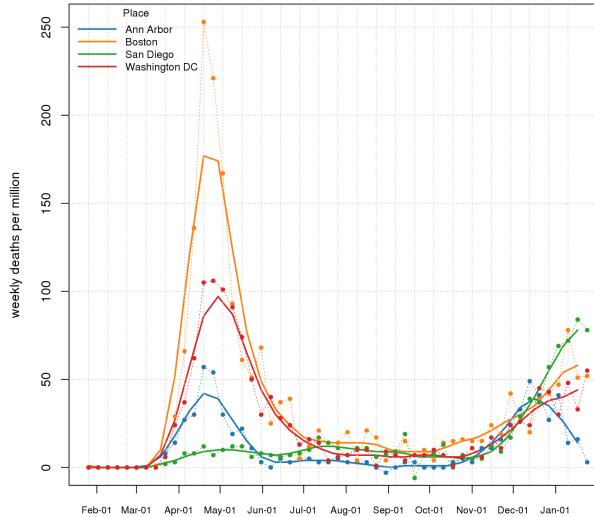


Figure 2d
Weekly deaths per million in non-Washington locations showing raw data



The smoothed Washington data (Figure 2a) shows three waves. The second peak was thankfully lower than the first; the third wave is already higher than the first in all areas except Seattle (King County). The graphs are heading down now, but it's too soon to know if the trend is real. The non-Washington data (Figure 2b) shows early peaks in most location, followed by a long trough, followed by a second wave starting in November. Rates continue to rise in Boston, San Diego, and Washington DC, but are clearly declining in Ann Arbor.

The raw data for Washington (Figure 2c) is very ragged. The smoothed curves undershoot the early peak. Data from mid-November to early January was extremely variable. Data for the last few weeks is lower than the recent peaks, offering hope that the downward trend is real. For non-Washington locations, the raw data (Figure 2d) is better behaved and supports the smoothed curves.

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](#). Further, the current data release (version 21-01-31) is completely missing the final two week: it should have data through the week of January 24 but stops at January 10. Further, this release has anomalously high counts for early dates, e.g, 10,331 cases statewide for the week of January 12, 2020, which was before the pandemic began! I truncate the data to avoid these problem dates.

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

Figure 3a
Weekly cases per million by age in Washington state

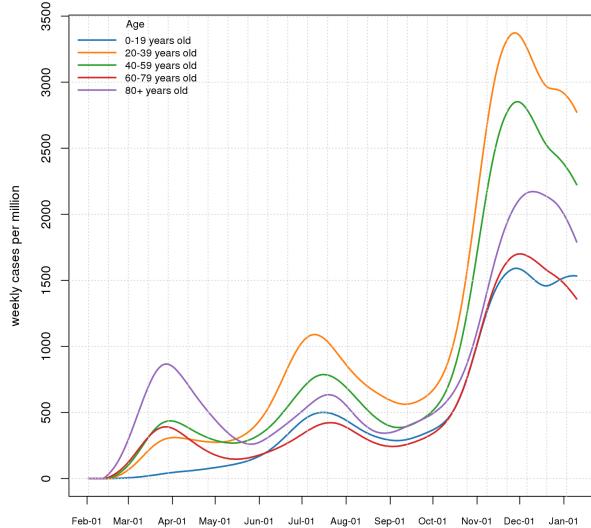


Figure 3b
Weekly cases per million by age in Seattle (King County)

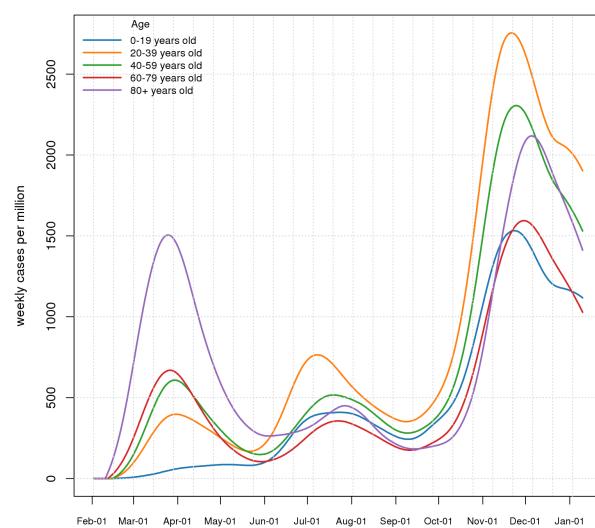


Figure 3c
Weekly cases per million by age in Snohomish (North of Seattle)

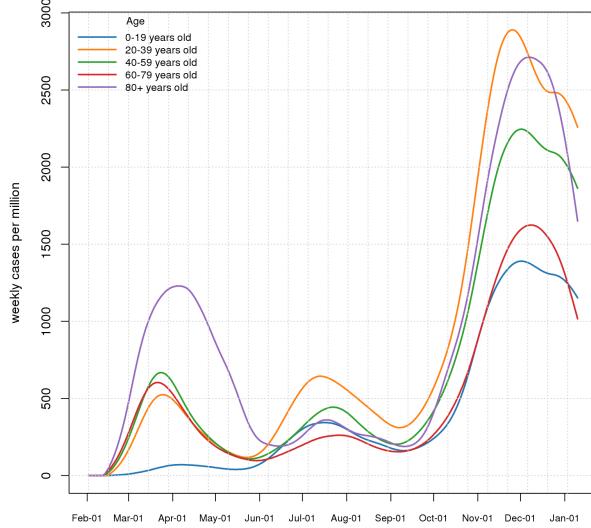
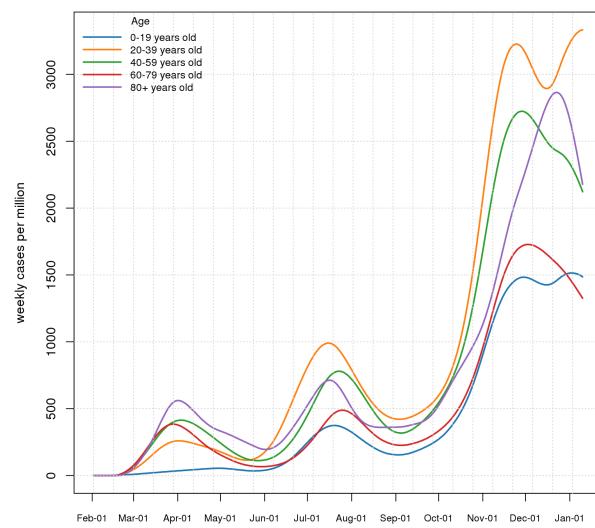


Figure 3d
Weekly cases per million by age in Pierce (South of Seattle)



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. As the wave has grown, the oldest people (80+) are again strongly affected.

Cases seem to be heading down now, but bear in the mind the variability of recent Washington data illustrated in Figure 1c.

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

Figure 4a
Weekly deaths per million by age in Washington state

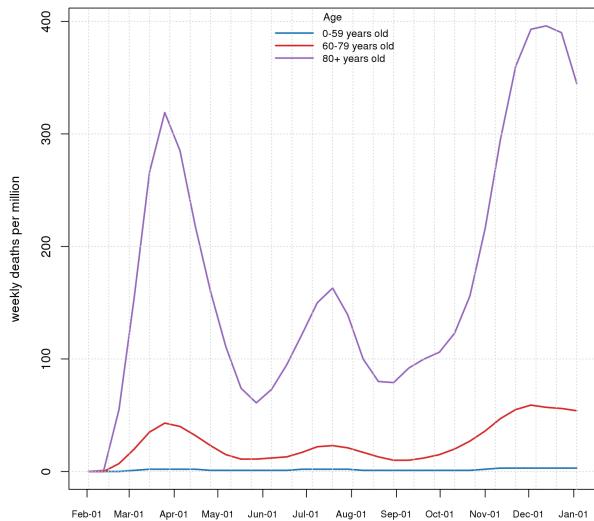


Figure 4b
Weekly deaths per million by age in Seattle (King County)

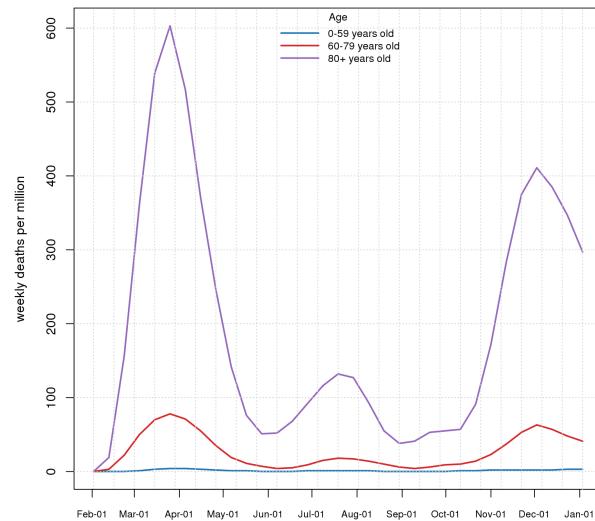


Figure 4c
Weekly deaths per million by age in Snohomish (North of Seattle)

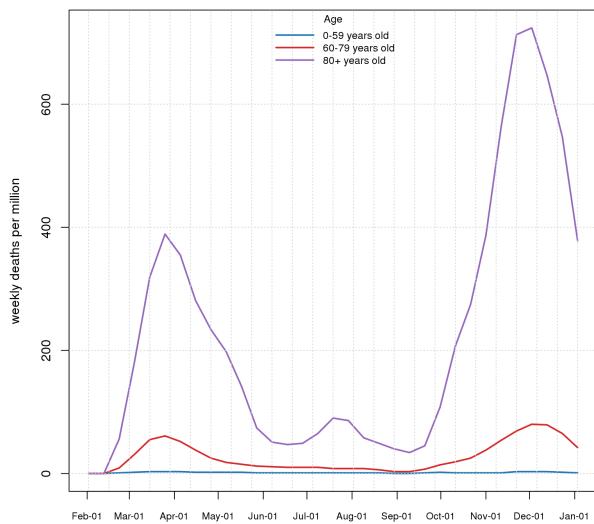
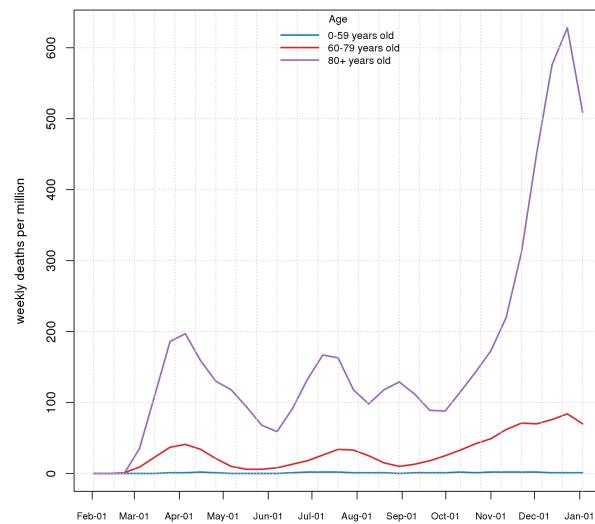


Figure 4d
Weekly deaths per million by age in Pierce (South of Seattle)



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 above the early Seattle 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 stayed fairly low until mid-November but now has now surged beyond the level of the early Seattle peak.

The decline at the latest time points in all locations may be due to reporting lags or other data problems, including the variability shown in Figure 2c. With cases so high in the 80+ age group (as shown in Figures 3a-d above), increasing deaths seem sadly inevitable.

Caveats

1. 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).
2. 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.
3. The same issues apply to *deaths* to a lesser extent, except perhaps early in the pandemic.
4. 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”.
5. The date granularity in the graphs is weekly. The underlying JHU data is daily; I sum the data by week before graphing.
6. I truncate the data to the last full week prior to the week reported here. Thus, data for the January 13 update includes counts through the first week of January, ending January 9.
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 versions of the document prior to December 30, 2020, 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. I attempt to correct this undercount through a linear extrapolation function (using R’s `lm`). I have tweaked the extrapolation repeatedly, even turning it off for a few weeks. The current version using a linear model that combines date and recentness effects.

The current data release (version 21-01-31) has additional problems. It is completely missing the final two weeks: it should have data through the week of January 24 but stops at January 10. Further, this release has anomalously high counts for early dates, e.g, 10,331 cases statewide for the week of January 12, 2020, which was before the pandemic began! The combination of missing data for January 2021 and excess data for January 2020 makes me wonder whether this may be a “Y2.1K” problem; perhaps data that belongs in 2021

is instead being reported with 2020 dates. I will investigate further and possibly adjust the data in future versions of this document.

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 download the data the same day as the DOH data (now Tuesdays) for operational convenience.

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.