

COVID-19 Counts in Washington State and Select Other Locations

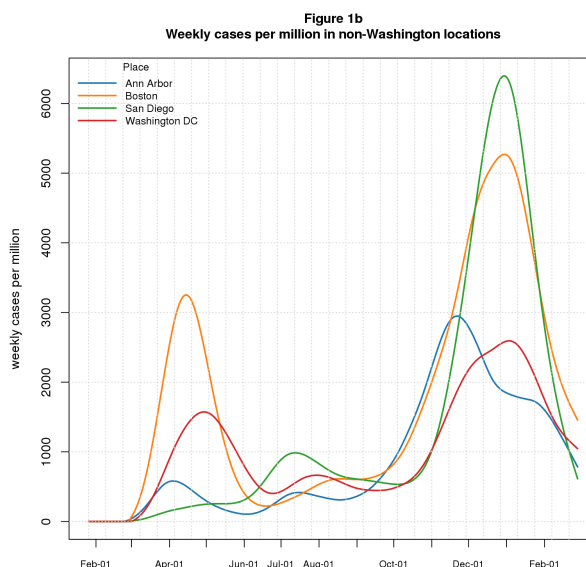
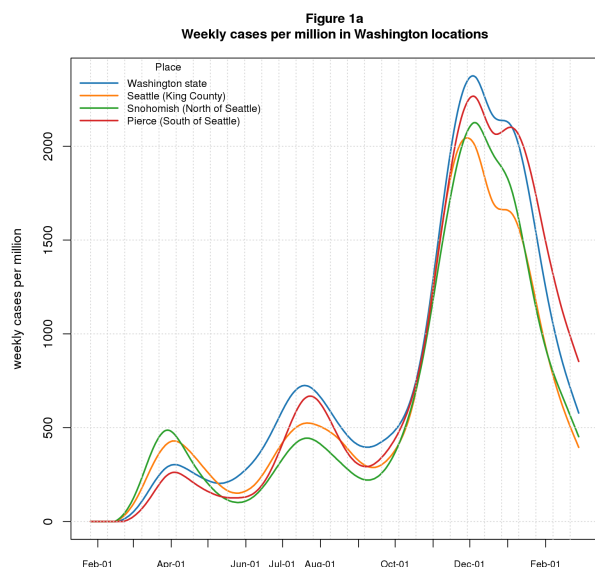
Nathan (Nat) Goodman

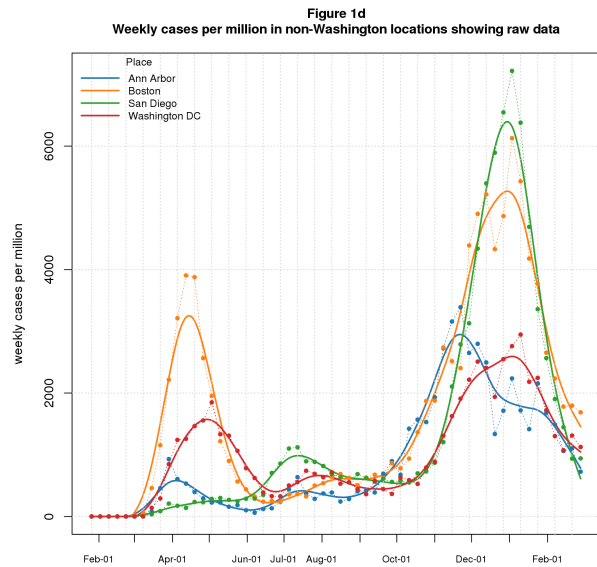
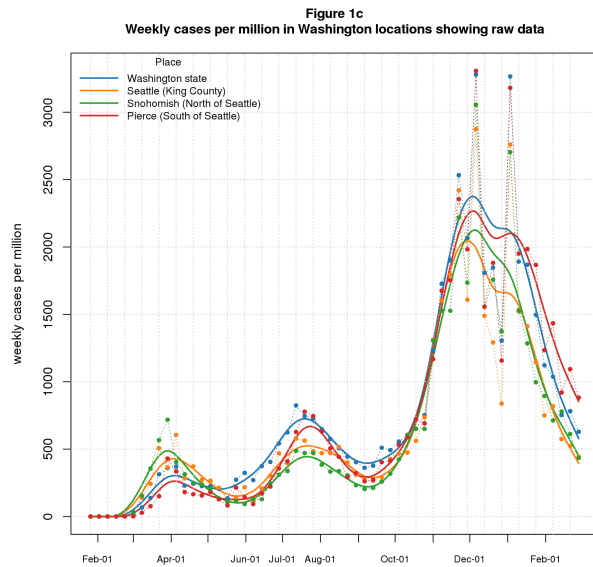
March 10, 2021

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.

Figures 1a-d show *case* counts per million for several Washington and non-Washington locations. 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.

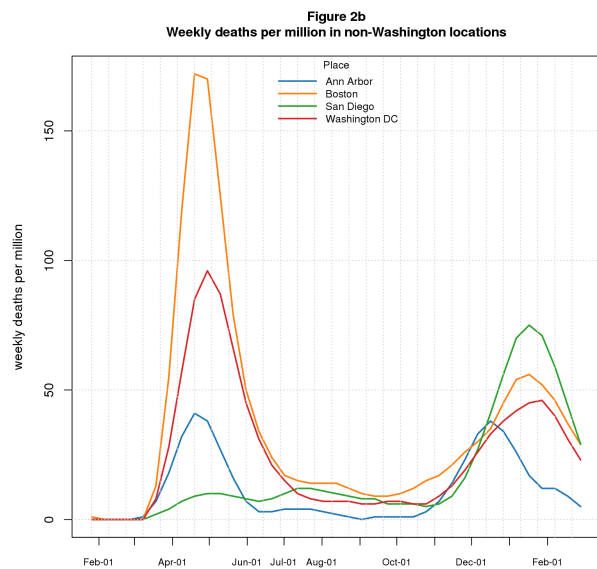
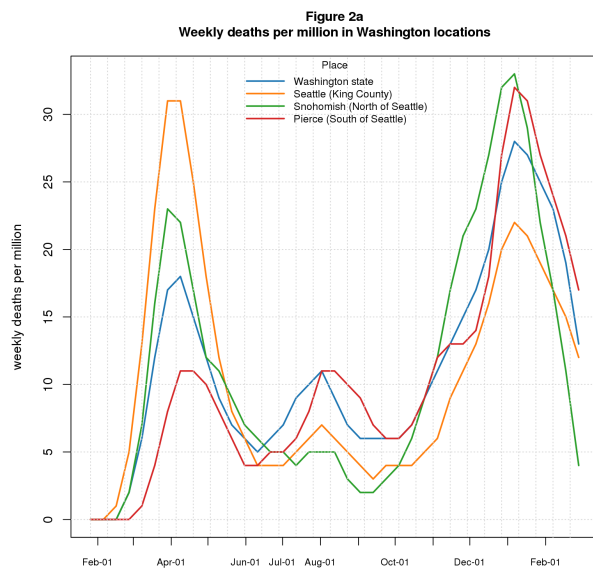
Figures 1a-b (the top row) show smoothed data (see details [below](#)); Figures 1c-d (the bottom row) overlay raw data onto the smoothed. The figures use data from Johns Hopkins Center for Systems Science and Engineering (JHU), described [below](#). When comparing the Washington and non-Washington graphs, please note the difference in y-scale: the highest current Washington rate (just under 900 per million in Pierce) is above the Ann Arbor rate (about 700 per million) and below the rates in the other non-Washington locations (about 900-1700 per million).

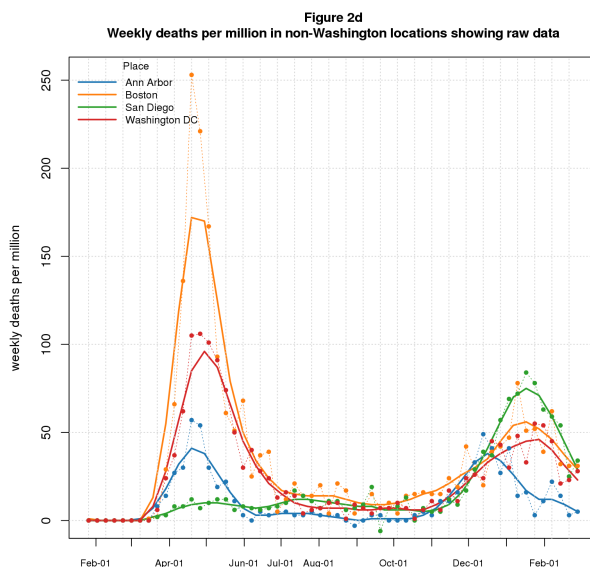
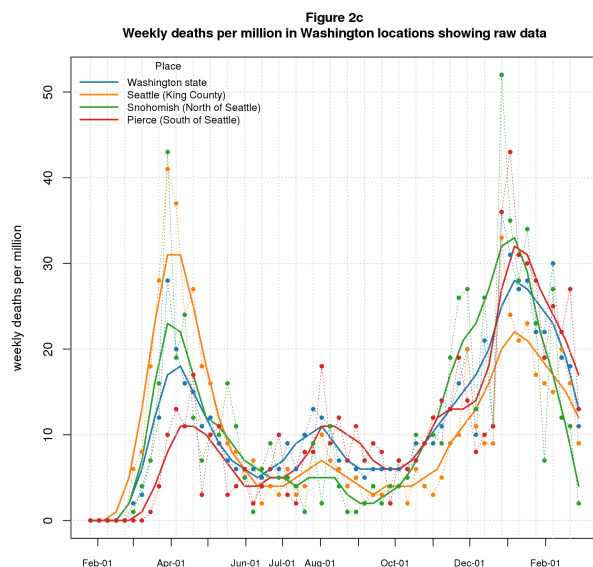




The smoothed graphs for Washington (Figure 1a) show that rates continue to decline; the rates are far below their recent peaks and have finally dropped below the Summer 2020 peaks except in Pierce. The raw data ((Figure 1c) and simple trend analysis (described [below](#)) indicate that the decline is real. The graphs for non-Washington locations (Figures 1b,d) are also falling dramatically. Trend analysis concurs.

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 are 2-13 per million; the non-Washington rates are 5-34 per million.





The smoothed Washington data (Figure 2a) shows three waves. The second peak was thankfully lower than the first; the third wave exceeded the first in all areas except Seattle (King County). The graphs are well down from their recent peaks and are finally down to the level of the second peak in most locations. The raw data (Figure 2c) remains quite variable; trend analysis (described [below](#)) indicates that looking back 8 weeks, the decline is clear.

The smoothed non-Washington data (Figure 2b) shows early peaks in most locations, followed by a long trough, followed by a second wave starting in November. The graphs are well down from their recent peaks and seem to be continuing down. The raw data, though variable, broadly supports this view; trend analysis (described [below](#)) indicates that looking back 8 weeks, the decline is clear.

In most previous versions of the document, I presented Washington results broken down by age using data from Washington State Department of Health (DOH) weekly downloads, described [below](#). I omit this here because the current DOH download has incorrect age data (most entries are zero).

Caveats

1. The term *case* means a person with a detected COVID infection. In some data sources, this includes “confirmed cases”, meaning people with positive molecular COVID tests, as well as “probable cases”. I believe JHU only includes “confirmed cases” based on the name of the file I download.
2. Detected *cases* undercount actual cases by an unknown amount. When testing volume is higher, it’s reasonable to expect the detected count to get closer to the actual count. 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.

Technical Details

1. 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.
2. The trend analysis computes a linear regression (using R's `lm`) over the most recent four, six, or eight weeks of data and reports the computed slope and the p-value for the slope. In essence, this compares the trend to the null hypothesis that the true counts are constant and the observed points are randomly selected from a normal distribution. After looking at trend results across the entire time series, I determined that p-values below 0.1 indicate convincing trends; this cutoff is arbitrary, of course.

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).

Usually, the weekly DOH download reports data by age group: 20-year ranges starting with 0-19, with a final group for 80+. The current download has incorrect age data (most entries are zero) and I omit these results from this version of the document.

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.