Covid-19 in Washington Data Explorer

Christopher WIllett

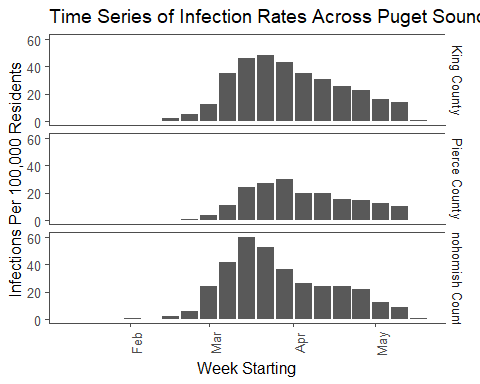
5/26/2020

# The Covid-19 Outbreak in Washington State

All health data comes from the Washington State Department of Health. Populations of counties comes from an unknown source. The outbreak of Covid-19 was initially centered in King County but rapidly spread to Snohomish County and later to Pierce County. Eventually it spread to all counties in the state. As a basic exercise in learning R along with git and Rmarkdown, the following data graphics were put together.

## Infection Rates in Three Puget Sound Counties

The three largest Puget Sound counties are King, Pierce, and Snohomish. The rate of infection is defined to be the number of new cases per 100,000 residents and is aggregated on a weekly basis.

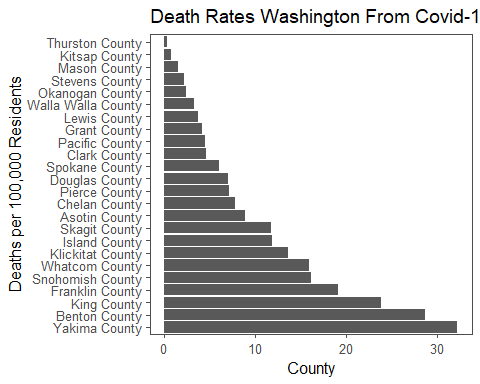


## Analysis of Infection Rates

Several things can be seen from these graphs. First, Pierce County had a distinct infection profile as seen by the flatter curve, less severe overall infection rate, and infection rates. Next, Snohomish County, despite having a lower overall population and more distributed population centers, had a much sharper infection rate curve than King County. Third, the infection rate in Snohomish County decreased more rapidly from peak from King County did, whereas Pierce County had a very slow rate of decrease.

## Death Rates in Washington State

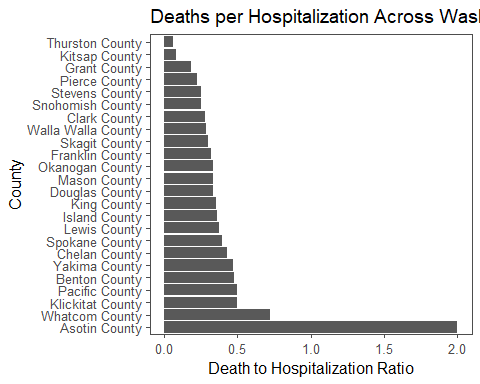
The death rate in each county is the cummulative deaths per 100,000 residents. The data was arranged in to a horizontal bar graph for each county and displayed below. Note that there were some counties that did not report deaths and there were some deaths that were not assigned to a county. These have been removed from the data.

 ## Analysis of Death Rates There are several items of interest that can be seen below. The epicenter of the outbreak, King County, does not have the highest date rate. The first, second, and fourth highest death rates occured in rural counties in the heart of Washington’s agricultural belt. All are adjacent to each other. Determining why the outbreak was more lethal there than in more urbanized county would require more, and disaggregated, data. To make analysis more difficult, an adjacent county, Walla Walla County, ranks near the bottom for death rate.

## Deaths per Hospitalization

One measure of how lethal the Covid-19 disease is could be the ratio of deaths to hospitalizations. It is reasonable to assume that most deaths occured in hospitals rather than outside of medical care. Notable exceptions would be elderly patients in assisted living facilities who died before transport, homeless populations, and populations, such as undocumented immigrants, who chose not to pursue hospital care. Without better disaggregated data from the state it is unclear how many deaths occured outside of hospitals.

Under the assumption that the effect is small, we can examine the death to hospitalization ratio. A ratio of near 1 means all hospitalizations resulted in a death, which would indicate a very lethal disease. A ratio near zero would indicate that hospitalization was effective at preventing deaths.

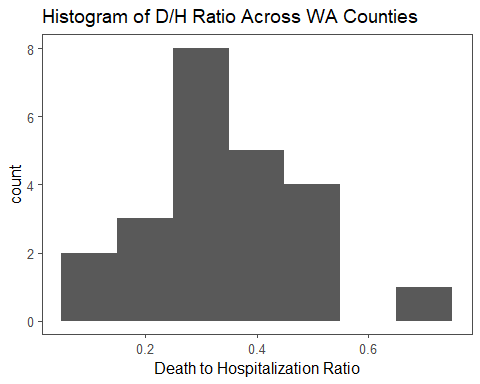


## Analysis of Death to Hospitalization Ratio

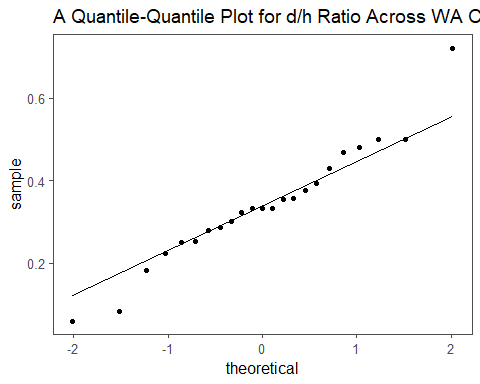
An immediate item of interest is the ratio for Asotin County. The fact that it is 2 indicates that more people died than were hospitalized. However, there were only 2 deaths total and 1 hospitalization, leading to this county being considered an outlier. Whatcom County is also seen as an outlier, although for reasons that are not clear. On the other end, the number of deaths in Thurston (1) and Kistap (2) are so small as to make the statistic irrelevant. The remaining data point show a lethal disease:If a hospitalization is required, the probabilty of death is roughly 30-50%.

## Outlier Analysis of D/H Ratio

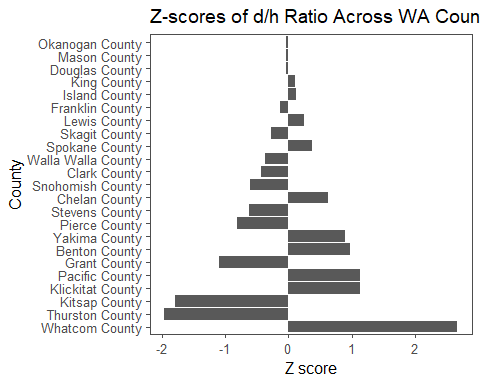
The previous graph shows that Asotin County should be removed from the D/H Ratio analysis, but also reveals a possible problem county from the view point of surviving a hospitalization due to Covid-19, namely Whatcom County. First, we check to see if the D/H ratio distribution across counties in Washington is normal. We can plot a histogram of the distribution with Asotin removed.



This data does show an approximately normal distribution. To further check, we can create a quantile-quantile plot (qqplot) that shows how the actual quantiles match up with the quantiles from a theoretically perfect normal distribution.



This again shows that the d/h ratio data is approximately normal. With that assumption in place, we can compute the z-scores for the d/h ratio data and examine outlines.



## Analysis of Z-scores

In this graph a z-score of 0 indicates that the county meets the statewide average for d/h ratio. A positive z-score indicates that the county is doing worse than the statewide average while a negative score indicates that the county is doing better. An z score whose magnitude (absolute value) is less than one indicates that the county is within about 2/3rds of all other counties while a magnitude of 2 or more indicate that it is outside of about 95% of all other counties.

From the graph we can see that Whatcom county has a z-score nearing 3, making it a severe outlier and indicating a potential problem in the treatement of covid-19 patients who are hospitalized. More data would be required to investigate further. Another item to note is the Thurston and Kitsap counties have very large, negative z-scores. However, both counties have reported very few deaths and, probably, should have been stripped from the analysis.