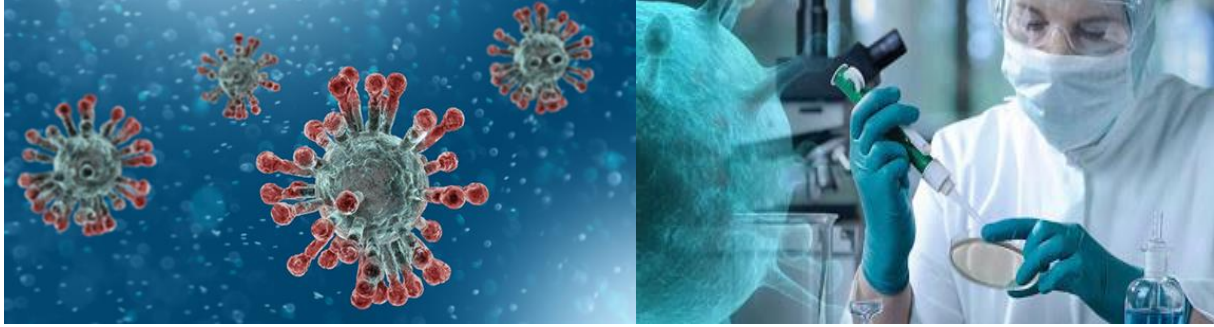# Capstone Project - Biotech Space

Abstract

A well-known biotechnology company is considering opening a new location in the Phoenix Metropolitan area. Their main objective is to research the coronavirus, implement clinical trials at an ideal location, and ultimately find a cure to save lives. Their preferences are in locations where there have been many confirmed coronavirus (COVID-19, SARS-COV-2) cases, therefore geographic data along with demographic attributes would be useful.

Keywords: Arizona Coronavirus Cases, Coronavirus Research, COVID-19, SARS-COV2

By Pam Russell

Description

  State level data was extracted from the CDC database from the National Center for Health Statistics (NCHS) and the Division of Health Informatics and Surveillance (DHIS). These data sets provided up to date statistics in Arizona on confirmed coronavirus cases, cases of mortality, demographic information, and geographic locations.

1) The first data set sample labeled Provisional Death Counts for Coronavirus Disease (COVID-19): Weekly State-Specific Data Updates and the source can be found here: https://data.cdc.gov/resource/pj7m-y5uh

2) The second data set sample labeled NNDSS – Table – 1FF. Severe acute respiratory syndrome-associated coronavirus disease to Shigellosis and the source can be found here: https://data.cdc.gov/resource/x8jf-txib

3) The third data set sample labeled Provisional Death Counts for Coronavirus Disease (COVID-19) includes state level data for deaths related to pneumonia and influenza and the source can be found here: https://data.cdc.gov/resource/hc4f-j6nb

4) The last data set was from the Arizona Department of Health website that contains information on cases in each zip code, and can be found here: https://adhsgis.maps.arcgis.com/apps/opsdashboard/index.html#/84b7f701060641ca8bd9ea0717790906

Methodology

Each data set was collected with the intention of cleaning and wrangling the data to find the appropriate jurisdictions that are going to be examined. Three data sets were examined to enhance the overall knowledge of the area. Several other factors related to the coronavirus were given in the data set to include all variables, and to provide a more comprehensive analysis.

Results

      The first data set gave us some insight on several cases that lead to mortality associated with the coronavirus. In the state of Arizona, the overwhelming percentage of cases with mortality were higher in the White and the Latino group (Shown in Figure 1).
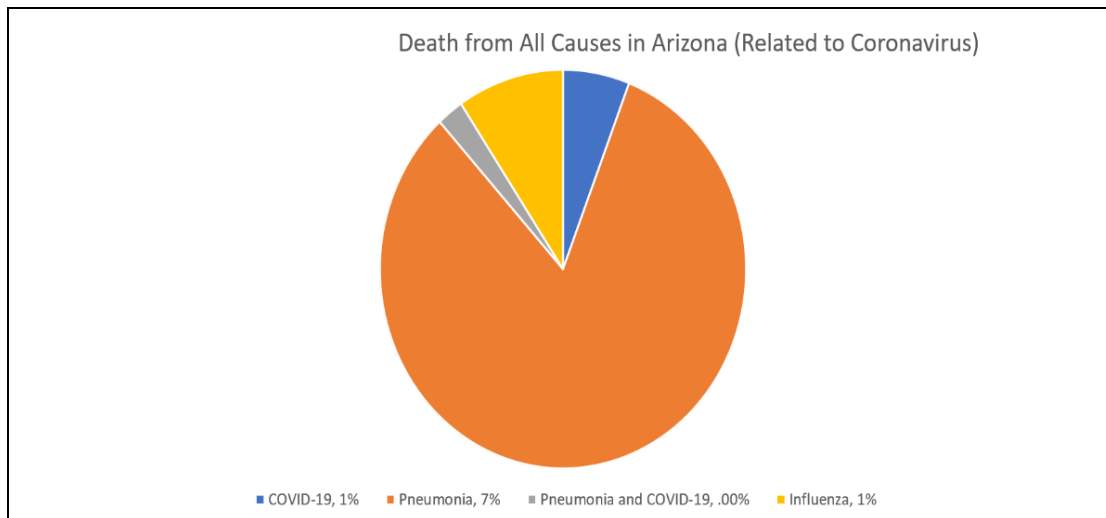
Figure 1.
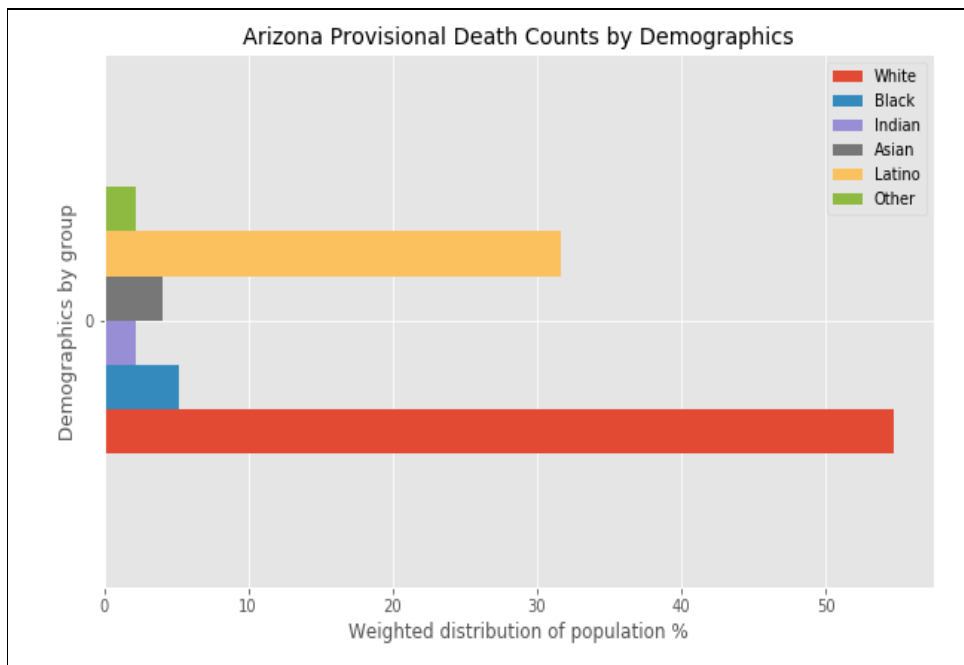
The second data from the CDC connected the coronavirus to enteric bacterial infections associated with the virus. According to the CDC (2019), Shigellosis is an infectious disease caused by a group of bacteria called Shigella. In addition, other researchers have also found bacterial infections linked to the first SARS related coronavirus. As reported by Leung et al. (2003), the coronavirus has been linked to a wide spectrum of diseases in animals, most notably in respiratory and gastrointestinal bacterial infections. This study also found bacterial infections that increased the risks for more severe respiratory issues (Leung et al., 2003). Another study noted by Hulswit et al., (2016), referenced an increased potential for the coronavirus to change tropism which is likely due to the viral spike (S) protein (Hulswit et al., 2016). The information provided on the data set was not as described and did not provide sufficient information that would be necessary for geospatial coordinates. However, when plotted on a histogram, you can see the higher elevation levels on Shigellosis compared to the prior year (Shown in Figure 2). The research that complimented it provided some added benefit.

Figure 2.
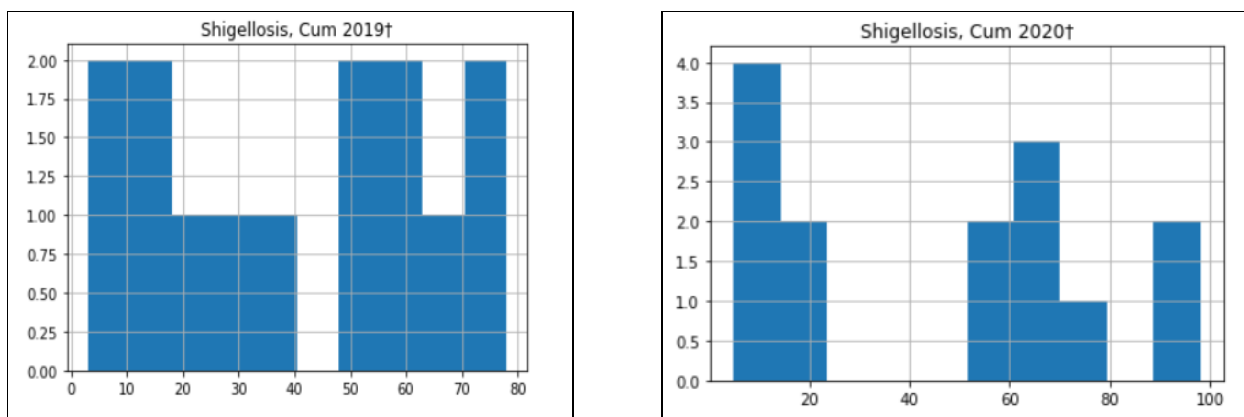
The third data set included all deaths from Arizona but included important factors of consideration that should be included in a data set such as pneumonia, cases that had both pneumonia and covid-19, and influenza. Although the data set did not provide enough information, when calculated to a percentage it provided a more concrete picture (Shown in Figure 3).

Figure 3.

Death from All Causes in Arizona (Related to Coronavirus)

■ COVID-19, 1%   ■ Pneumonia, 7%   ■ Pneumonia and COVID-19, .00%   ■ Influenza, 1%

The last data set provided by the Arizona Department of Health reported on confirmed positive coronavirus cases by zip code, which has been updated rather recently. The data provided key information on zip codes with higher coronavirus confirmed cases (Shown in Figure 4).

Figure 4.
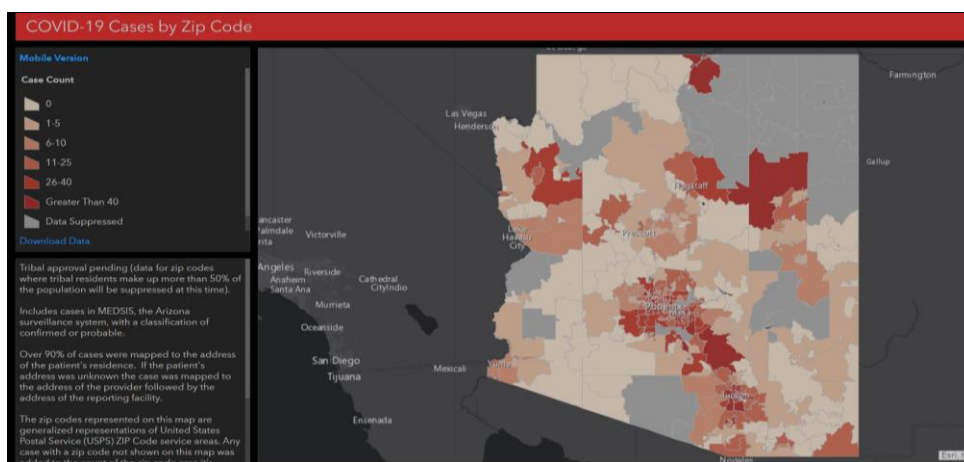
Although this information does show higher cases by zip code, and is rather appealing visually. It lacks for quick clarity on the highest coronavirus confirmed cases. In addition, since it is by zip code, it rules out Foursquare for their API data. In comparison, the histogram graph (Shown in Figure 5) clearly shows the highest zip codes by far being between 85000-85200. This indicates that the best location is going to be the closest to downtown Phoenix.

Figure 5.
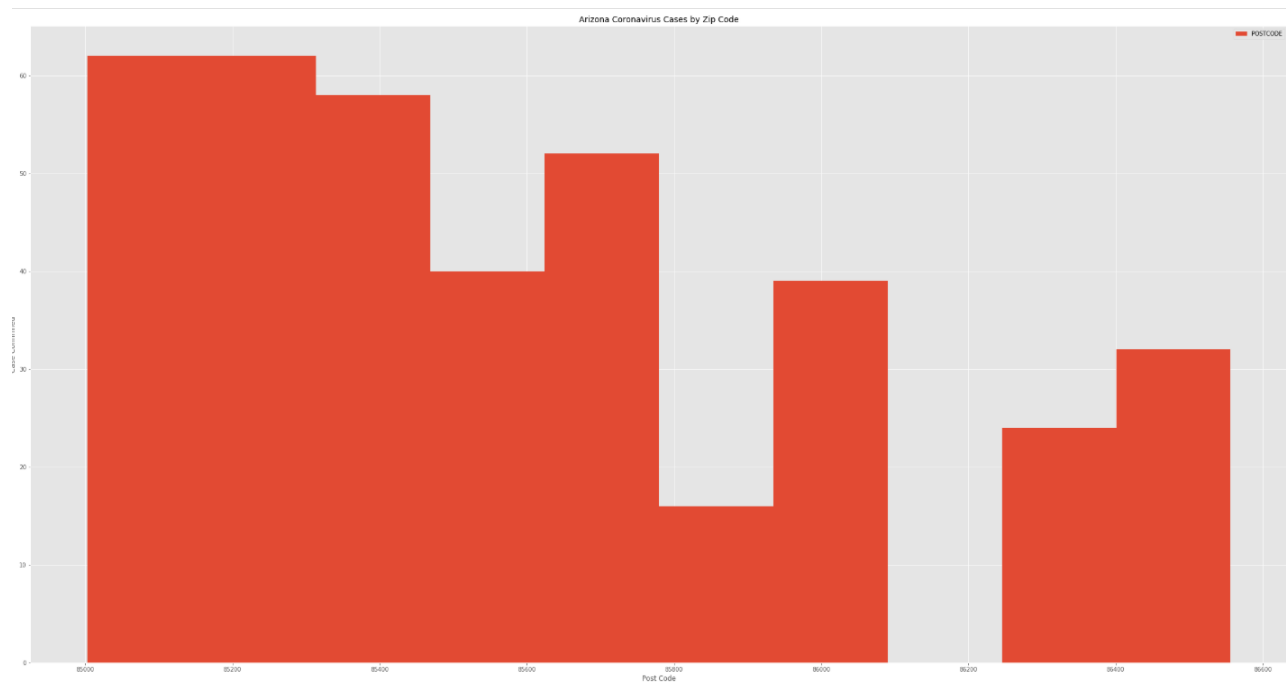
Limitations

    Each data set has delays in reporting, some data is incomplete and does not provide an entire picture at the state level. Therefore, some data may be underestimated. It has also been reported by researchers at ASU that the number of cases should actually be higher when calculating the mathematical modeling that is similar to other places. Another consideration is the level of comorbidities that are associated with the disease. Although this is preliminary data, these factors were not included in any of the data sets. Some of the data sets were also cleaned prior to examining for data analysis.

Discussion

Although each data set provides some insights into the cases surrounding the coronavirus in the Phoenix area, there is still a lot more we need to learn. It has become clear that more research is required by the entire medical community prefer we can assumptions. One observation that can be made is that the coronavirus has affected the general Phoenix metropolitan area. Especially in locations near downtown Phoenix and close suburbs. These are top spots for biotechnology firms that are wanting to make a difference. This would be an important factor to contribute to medical research as changes in the climate will continue to add more infectious diseases and viruses of these kind in the future. Data analysis therefore places a critical role in reporting, analysis and mapping of various amounts of information.

References

Leung, Wai (2003). Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection, *Gastroenterology*. Volume 125, Issue 4, Pages 1011-1017. DOI: 10.1016/j.gastro.2003.08.001. Retrieved from: https://www.gastrojournal.org/article/S0016-5085(03)01215-0/fulltext?referrer=https%3A%2F%2Fwww.gastrojournal.org%2Farticle%2FS0016-5085%2803%2901215-0%2Ffulltext

CDC (2019). Shigella – Shigellosis, Bacteria. Retrieved from: https://www.cdc.gov/shigella/index.html

Sanchez, C.M, Izeta, A, Sanchez-Morgado, J.M, Alonso, S, Sola, I, Balasch, M, Plana-Duran, J, and Enjuanes, L. Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. *J Virol*. 1999; 73: 7607–7618.

Hulswit, R.J.G (2016). Coronavirus Spike Protein and Tropism Changes, *Advances in Virus Research*. Pages 29-57. Published online 2016 Sep 13. doi: 10.1016/bs.aivir.2016.08.004. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7112277/

Arizona Department of Health Services (2019). Confirmed COVID-19 Cases by Zip Code. Retrieved from: https://adhsgis.maps.arcgis.com/apps/opsdashboard/index.html#/84b7f701060641ca8bd9ea0717790906