Project Milestone 5

Astrid Fuentes, MS

Bellevue University

DSC 540: Data Preparation

Prof. Benjamin Schneider

March 4, 2021

Introduction

Throughout the term we have been building a project consistent of 5 milestones. Each milestone was intended to find, clean, and blend together 3 data sources related to the same topic and then try to answer some relevant questions with it using 5 different visualizations.

Questions and Answers

In Milestone 1, I chose 3 data sources related to Covid-19 and defined some questions I wanted to answer with this data. Let's start by looking back at my original statistical questions and attempt to answer them:

1. Are the number of deaths higher in male patients than female?

   The answer is yes, according to *Figure 1*, there are over 30 thousand more deaths in Males than Females across the United States. *Figures 2* and *3* show the total deaths per gender and state. We can clearly see in Figure 3 that all states have higher total deaths for Males than Females.

Figure 1.

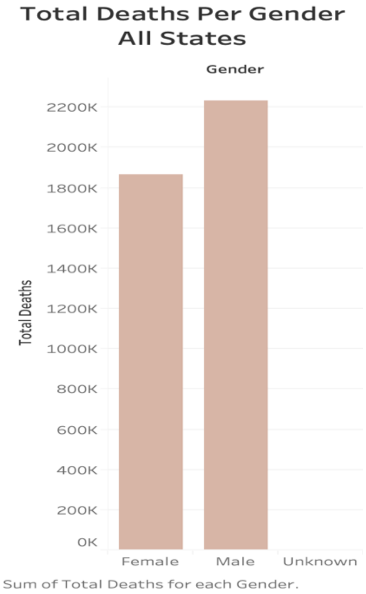


Figure 2.

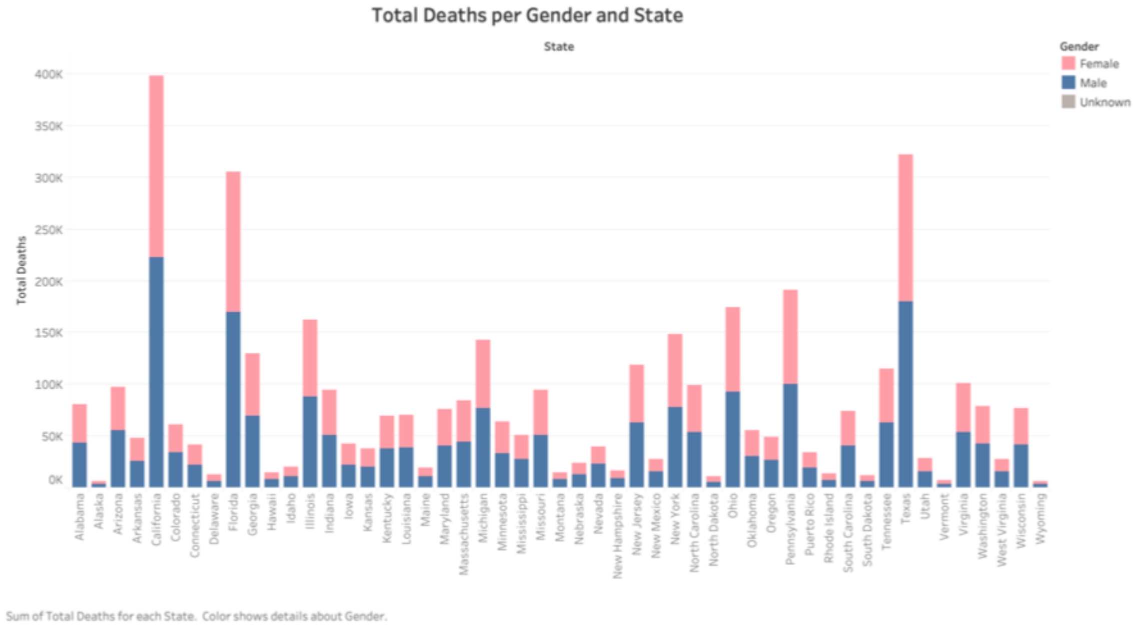
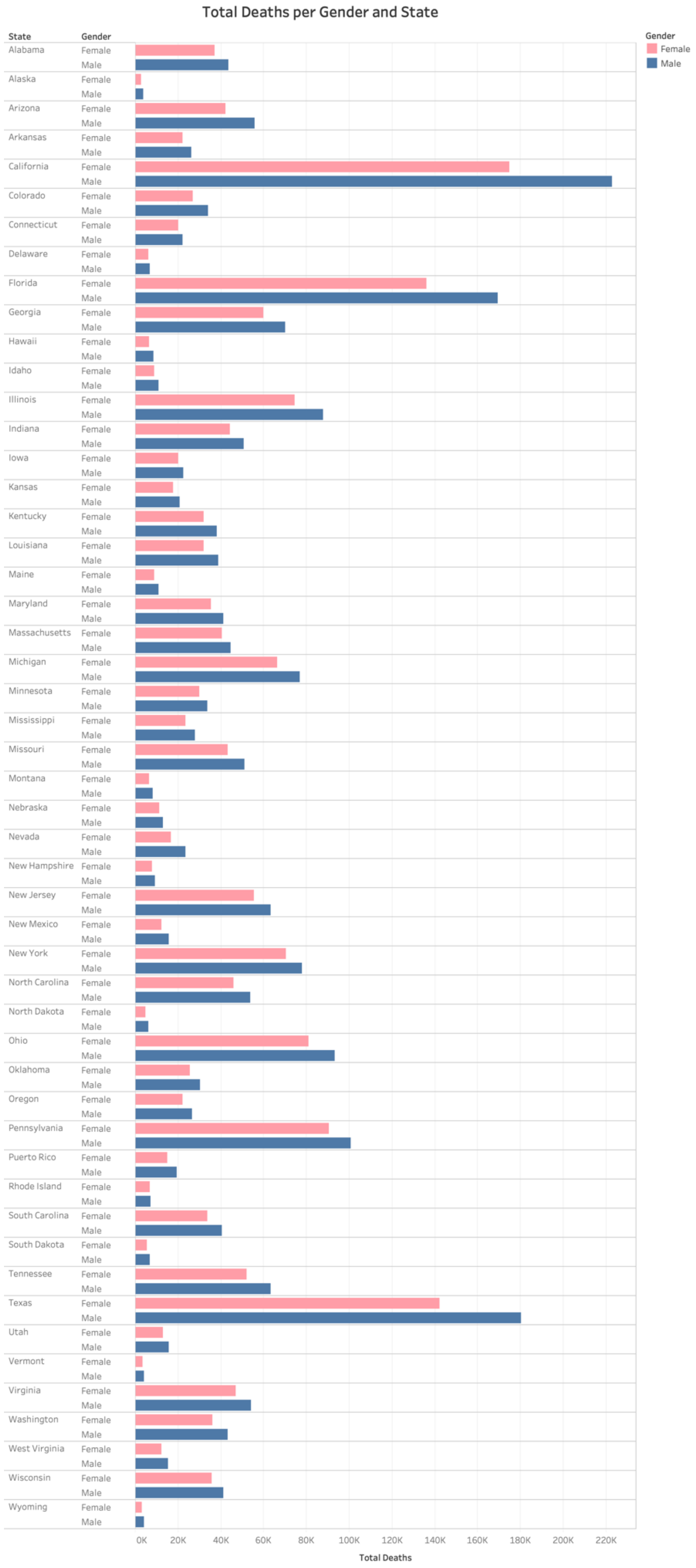

Total Deaths per Gender and State
State
Gender
Female
Male
Unknown
Total Deaths
400K
350K
300K
250K
200K
150K
100K
50K
0K
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Sum of Total Deaths for each State. Color shows details about Gender.

Figure 3.

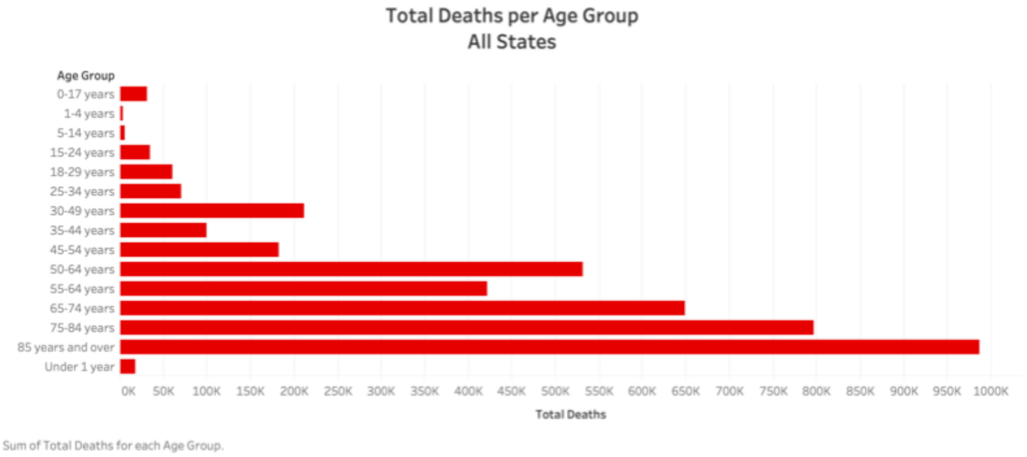


Sum of Total Deaths for each Gender broken down by State. Color shows details about Gender. Details are shown for State. The view is filtered on Gender, which keeps Female and Male.

2. Are seniors' death rate higher than non-senior?

   Yes, *Figure 4* shows how total deaths increases considerably as the age group increases.

   Figure 4.



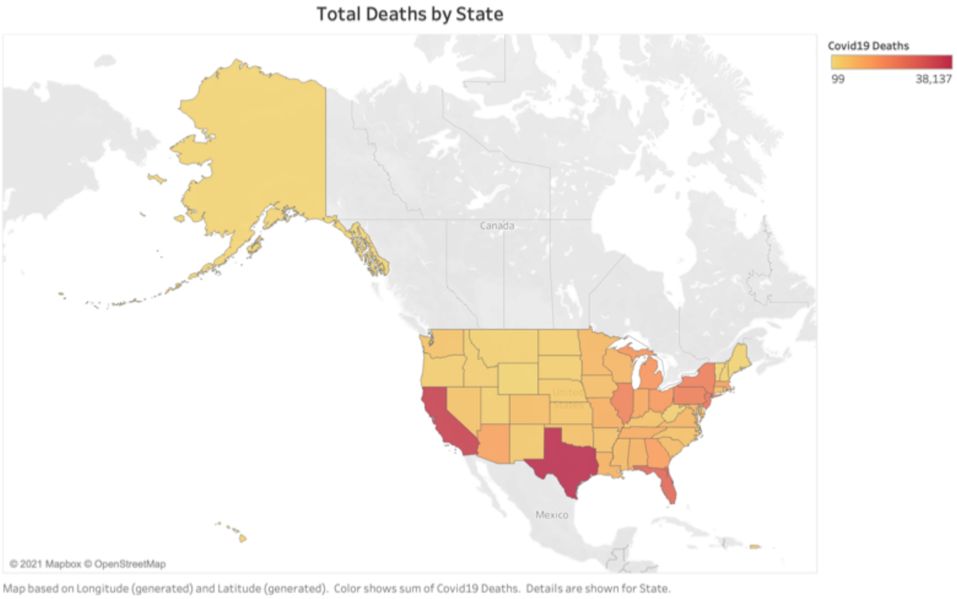
3. Examine the overall trend of the time series and how some of the variables might be contributing to that trend.

   I was unable to answer this question at this time since the dates of report where the same for all records. I will need to obtain dates from another source to be able to create a time series.

4. Are number of deaths higher in certain states during a particular time-period. Possible causes?

   Figure 5 shows that the number of deaths is higher in states like California, Texas, New York, and Florida. Similar to the previous question, I was unable to include the time-period at this time.

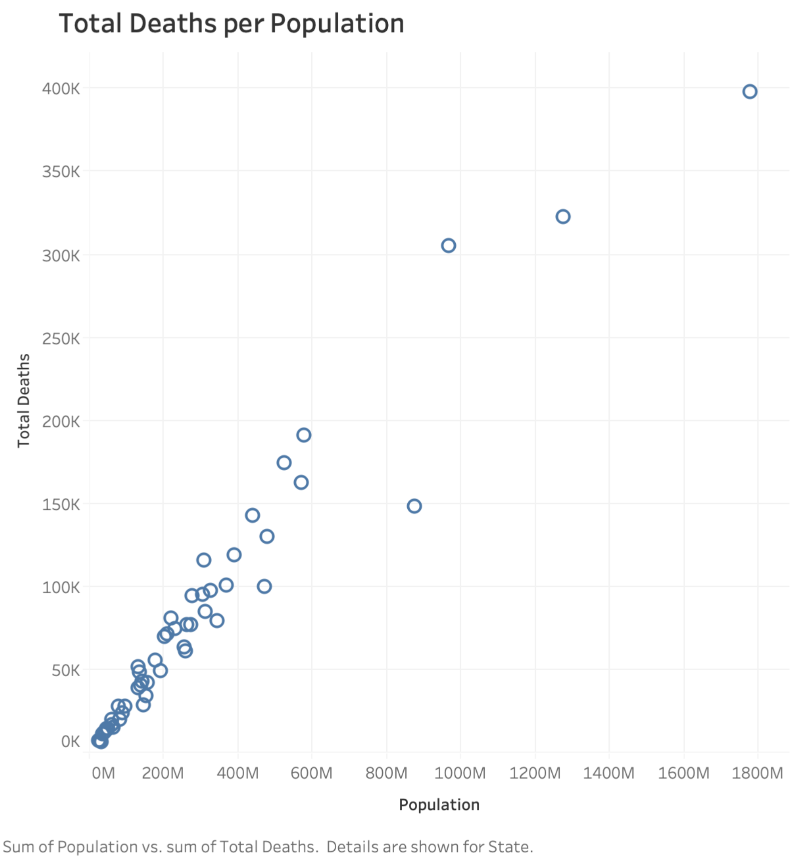
Figure 5.



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Covid19 Deaths. Details are shown for State.

Additionally, *Figure 6* shows that the number of deaths is positively related to the size of the population, as the size of the population increases, the number of deaths by Covid-19 also increases.

Figure 6.



Final Comments

While working on this project I have learned to work with different data sources like flat files, web pages, APIs, etc. I have learned how to clean the data and perform basics checks for duplication, missing information, etc. I have learned to rename columns, replace nulls, and other cleaning and organizing activities.

Furthermore, I learned, not only to connect to a database, but also how to work with databases including creating tables, loading data to them, retrieving data, joining, etc.
Overall, I think this project is very helpful and I plan on keep working on it and improving it as time allows. I know there are things that I could do better but given time constrains I was unable to at this time.

References

Center for Disease Control and Prevention: Provisional COVID-19 Death Counts by Sex, Age, and week. https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-W/vsak-wrfu/data

Worldometer: Coronavirus. https://www.worldometers.info/coronavirus/country/us/

The Covid Tracking Project. https://covidtracking.com/data/api