Reed Foster, Jack Sheridan

Data Wrangling

12/7/2023

**Project Proposal: Covid 19 Impact Analysis**

**1. Introduction:**

The COVID-19 pandemic has had far-reaching effects on societies worldwide. The seriousness of this virus has impacted the way people travel, act, and ultimately think. Within this project, we plan to use data provided by GigaSheet, The New York Times COVID case count and the Google COVID-19 Mobility Reports to gain understanding on different phases of the virus and its impact on various aspects of society within the state of Iowa. From the impacts on people to public health and the food industry, understanding these effects can provide crucial insight into decision making processes in the future.

**2. Data:**

**Data Set 1:**

* Source: Google Mobility Reports
* Link/Citation: <https://www.google.com/covid19/mobility/>

The first data source we are using in this project is from Google’s Community Mobility Reports. The data provided by this website contains anonymized, aggregated location data from mobile devices to illustrate trends in visits to different types of locations across the entire world. The data set includes insight into various categories of places including retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential areas. The reports were originally designed to assist public health officials in making informed decisions to combat the pandemic. However, it is important to note that the Community Mobility Reports are no longer being updated as of 2022-10-15. Although, all historical data remains publicly available for research purposes. This data helps researchers like us provide valuable insights into how mobility patterns have changed in response to the pandemic. We collected the data via a direct CSV download, in which we then cleaned to look at data only from Iowa. Given the immense size of the data file prior to this filter, we thought it would be more relevant and informational to pay attention to the specific effects on the state we are from. The filtered data includes the difference in mobility changes from the baseline by county and sector.

**Data Set 2:**

* Source: New York Times
* <https://www.nytimes.com/interactive/2021/us/iowa-covid-cases.html>

The second dataset used is from the New York Times interactive Covid case count map specifically regarding the state of Iowa. The data provided on this website contains trends and all-time data which has been aggregated for each county within the state. Through interactive and regularly updated data visualizations, it offers comprehensive insight into the COVID-19 situation within the state. This resource provides essential information such as the number of confirmed cases, hospitalizations, and deaths, as well as trends over time. Users can explore data at various geographic levels, from the state level down to county and local regions, enabling them to understand the impact of the pandemic on different areas of Iowa. We plan on scraping data for each county and use the total cases, the number of cases per 100,000, the total deaths, the total deaths per 100,000 and the percentage that is fully vaccinated to involve in our analysis. These numbers are all time numbers meaning they consider every instance since the beginning of Covid. We scraped the data using Power Automate. utilized this data in tandem with the data from the Google Mobility Reports to help us see and provide a better picture of the spread of COVID-19 and its effects on different counties in Iowa.

**Data Set 3:**

* Source: GigaSheet
* Link: <https://gigasheet.com/sample-data/spreadsheet-list-of-all-counties-in-iowacsv>

The third data set we used was from GigaSheet. The data set from this website was downloaded from a direct CSV file link. The data within this file included population data for counties in Iowa. This data was necessary for our analysis so that we could see Covid’s effect on different sized populations.

**Note:**

We initially planned on utilizing economic data from yahoo finance to dig deeper into the economic effects by sector. But we were unable to do so as the majority of our data from other sources lacked a proper date. This disabled us from forming conclusions about economic effects by timeline. So, we instead directed our attention to the effects population size had on various aspects of the pandemic.

**Data Dictionary:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Field Name** | **Source** | **Data Type** | **Description** |
| Country | Google | string | Country listed |
| state | Google | string | State listed |
| county | Google | string | County listed |
| retail\_and\_recreation\_percent\_change\_from\_baseline | Google | float | % change in visits to residential areas compared to baseline |
| grocery\_and\_pharmacy\_percent\_change\_from\_baseline | Google | float | % change in visits to residential areas compared to baseline |
| parks\_percent\_change\_from\_baseline | Google | float | % change in visits to residential areas compared to baseline |
| transit\_stations\_percent\_change\_from\_baseline | Google | float | % change in visits to residential areas compared to baseline |
| workplaces\_percent\_change\_from\_baseline | Google | float | % change in visits to residential areas compared to baseline |
| residential\_percent\_change\_from\_baseline | Google | float | % change in visits to residential areas compared to baseline |
| Total Cases | New York Times | Numeric | Total number of cases based on county |
| Total Deaths | New York Times | Numeric | Total number of deaths based on county |
| Cases per 100,000 | New York Times | Numeric | # of cases per 100,000 residents |
| Deaths per 100,000 | New York Times | Numeric | # of deaths per 100,000 residents |
| Percent fully vaccinated | New York Times | Numeric | % of population fully vaccinated |
| population | GigaSheet | Numeric | # of residents in that county |
| Population\_Category | Engineered by us | string | Counties with less than 50,000 or greater than 50,000 |
| Case\_Death\_Ratio | Engineered by us | float | case/death |

**3. Analysis:**

**Research Questions-**

1. What was the most highly impacted sector in Iowa?
2. How are counties with less than 50,000 residents affected in comparison to those that have greater than 50,000 residents?
3. What are the top 10 worst counties case to death ratio and what were the correlations/effects on mobility patterns?
4. How do the cumulative changes in mobility indicators specifically retail and recreation relate to the total number of COVID-19 cases in the top 10 covid stricken counties?

**Results:**

With the introduction of population data and the deletion of economic data and the concept of time, our project now aims to analyze the different effects COVID had on mobility and medical factors. Specifically, our analysis gives visual, and data driven insight regarding these topics on the effects of COVID-19 across counties in Iowa. The following graphs and charts depict the effects COVID had on various aspects of society/business by looking at mobility trends, as well as medical aspects like the mortality count.

**Question 1-**

A graph showing different colored squares

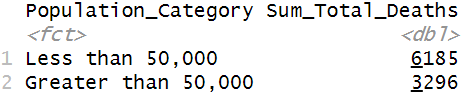
Description automatically generated

Our comprehensive analysis revealed that both parks and transit stations exhibited the most substantial deviations from baseline across all areas examined. The graphical representations incorporated the minimum and maximum values within each category, providing a nuanced depiction of the extent of change throughout the entire pandemic period. The observed trends in mobility metrics align logically with the societal response to the pandemic, wherein individuals actively sought to practice social distancing.

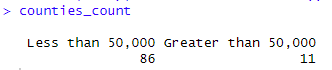
Notably, the data indicates a discernible shift away from crowded transit centers, with a concomitant increase in mobility towards parks. This observed surge in park visitation is a plausible outcome, given the expansive and open nature of these spaces, providing individuals with a venue to adhere to social distancing guidelines more effectively. The prominence of parks as areas with the highest mobility aligns cohesively with the public's inclination to seek open environments perceived as having a lower risk of COVID-19 transmission.

The significance of these mobility patterns extends beyond mere statistical observations, offering valuable insights into the behavioral adaptations of individuals during the pandemic. Understanding these dynamics can inform future urban planning, public health strategies, and crisis management, fostering a more resilient and responsive approach to similar challenges in the future.

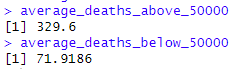
**Question 2-**



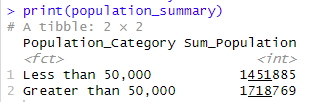
The table above shows the total death count within counties less than and greater than a 50,000-population size. The intent of producing this graph was to see which category produced the greater number of deaths. By seeing this you may think instantly that counties with less than 50,000 residents had it worse, but we thought it would also be important to consider the specific number of counties in each class.



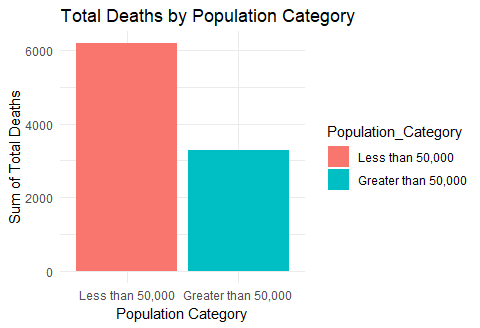
We assumed that because the less than 50,000 categories had more deaths that it was because there are more counties within the data set that are listed below 50,000. So, we wrote some code to provide the total number of counties that fall within each category.As expected, there were substantially more counties in the less than 50,000 grouping. Because of this, we wanted to see the average amount of deaths across each group so that we could get a better picture on the true effect of population on the death count.



As expected, the average amount of deaths per county was larger within counties that have above 50,000 residents. This finding could point to the conclusion that the size of a county does have a correlation with the number of deaths.

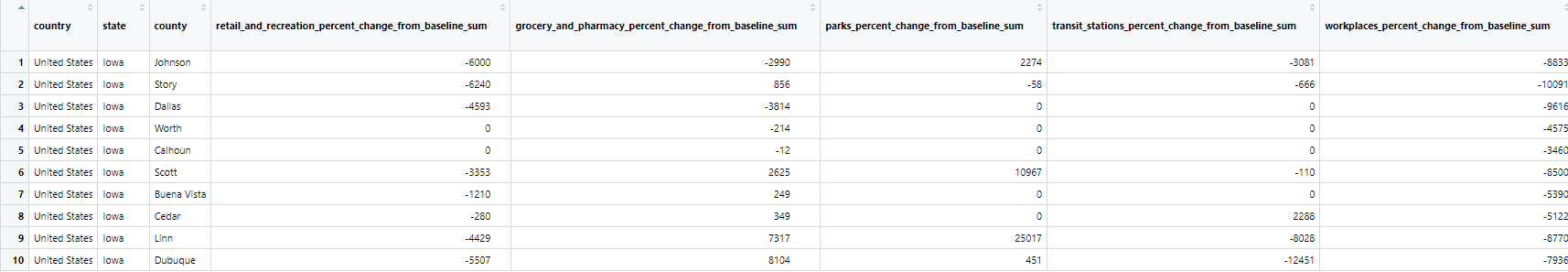


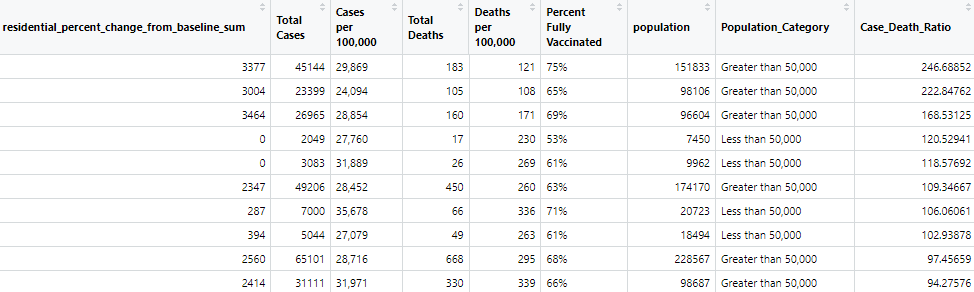
This chart shows the sum of the population for each category. Even though there are way more counties, the total population for counties with greater than 50,000 residents are much larger than the other class. Regardless, the two sums are rather close. So, it is surprising to see that counties with less than 50,000 residents have a higher total death count. The reasoning behind this may be a good thing to investigate in the future as the average population numbers suggest there may be some correlation, but overall numbers pertaining to the sum within both classes are still dodgy.



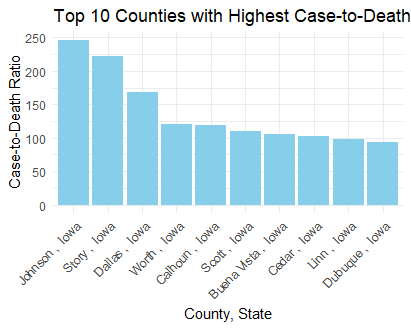
We created this graph to visualize the total number of deaths in each class. Overall, counties with less than 50,000 residents produced more deaths than counties with greater than 50,000 residents. This is rather surprising given that the overall population for counties with more than 50,000 residents is larger. But it is also understandable given the fact that there are way more counties with residents below 50,000. Again, researchers may want to dive deeper to see why the death count was higher for smaller counties when one would normally think that places with higher populations would have more serious consequences.

**Question 3-**





The top 10 worst case to death ratios by county and the associated effects in terms of mobility by sector are shown above in the table. A new column has been calculated to show this total. As expected, the mobility patterns in each of these counties are largely affected in comparison to other counties. This shows correlation between extreme medical effects and mobility tendencies. Also, six out of ten of these counties are those which have populations above 50,000. This provides good insight into why these top 10 counties may have a bad case to death ratio.



This chart visualizes the top 10 counties with the worst case to death ratio, in which Johnson County was the worst. This is not surprising because it is a county that contains some of the largest cities in Iowa.

**Question 4-**

A graph of covid-19 cases

Description automatically generated

We wanted to look further into the effects on mobility of counties that had a high case rate. The analysis of mobility changes in relation to COVID-19 case counts across the top 10 counties revealed noteworthy findings. Contrary to expectations, Polk County, boasting the highest case count, did not exhibit the highest decline in mobility concerning retail and recreation. This observation is intriguing, particularly given Polk County's status as home to Des Moines, the state's largest city. A plausible explanation for this anomaly could be Iowa's comparatively lax approach to COVID regulations, fostering a segment of the population in larger cities, including Des Moines, with a diminished inclination to curtail mobility.

In contrast, Johnson County displayed a substantial reduction in retail and recreation mobility, aligning with its reputation for adopting a more stringent stance toward COVID-19. This correlation suggests that communities with a heightened sense of caution and compliance with pandemic guidelines experienced greater declines in mobility. Notably, Johnson County's business landscape bore the brunt of these measures, witnessing significant challenges during the pandemic as stricter adherence to COVID protocols resulted in a notable loss of customers.

This analysis lends support to the argument that stringent COVID regulations may disproportionately impact small businesses. Going forward, the insights derived from this data can serve as a valuable resource in illustrating the potential adverse effects of stringent pandemic-related measures on local economies, particularly within the context of small businesses.

**Code file specifics:**

All things related to cleaning, merging, and the analysis for questions 2 and 3 are in the file “rcfoster,jsheridan\_merge, cleaning, and analysis for questions 2 and 3.R” and the associated R data file.

All things related to the analysis for questions 1 and 4 are in the file “rcfoster,jsheridan\_Questions 1 and 4 analysis.R” and the associated R data file.

**4. Conclusion:**

In summary, our analysis of COVID-19's impact on Iowa counties revealed significant shifts in mobility patterns, with parks and transit stations as the biggest movers during the pandemic. Counties with populations below 50,000 exhibited higher total death counts, but those exceeding 50,000 showed a higher average death count per county, suggesting a correlation between population size and mortality. The top 10 counties with the worst case to death ratios experienced notable mobility disruptions, predominantly in sectors like parks and transit stations. Surprisingly, stringent COVID-19 regulations in Johnson County correlated with a substantial reduction in retail and recreation mobility, underscoring the potential economic impact of pandemic measures on local businesses. The information provided by this analysis could be useful in the future if the world were to ever experience another pandemic. While this project provides valuable insights, the study is limited by the absence of economic data and time series data. Future research should consider these aspects for a more comprehensive understanding of the pandemic's overall effect on Iowa's counties.