Nathan Ewing

Analysis of COVID-19 in Denver, Colorado

GEOG 4580 Introduction to Geographic Databases

**Project Introduction**

The coronavirus outbreak, first identified by the Wuhan Municipal Health Commission in December of 2019, has since extended to every corner of the globe, killing more than 1.6 million and sickening more than 71 million in a matter of months [2]. According to CNN’s live coronavirus tracking dashboard, the United States alone has reported over 16 million confirmed cases, earning condemnation as a global hotspot.

Although the coronavirus has devastated families across the nation, in the United States, minority groups have disproportionately suffered from health impacts and socioeconomic factors. Health factors that contribute to greater risk from coronavirus, such as diabetes, heart disease, and obesity are more rampant in minority communities, reducing the ability of certain populations to cope with the impact of the pandemic [3].

The goal of this project is to analyze how socioeconomic factors impact positive cases and rates of Covid-19 infection and will serve as a case study to identify if lapse in health care coverage exist for low-income, underserved communities in Colorado. As a new resident of the state of Colorado, this project will provide invaluable insight into the challenges faced by large swaths of the community. In addition, this project will enable me to identify the areas and neighborhoods most heavily impacted by the pandemic.

**To conduct this investigation, I seek to answer:**

* Counties with the highest number of confirmed cases?
* Counties with the largest land mass?
* Which counties have the largest percentage of black residents?
* How do Income Levels Impact COVID-19 Positivity Rates?
* What are the most popular stores in Denver?
* What is the relationship between store type, city, and percent of black resident deaths in the area?

While this personal project serves as a method to familiarize myself with the area, extensions of this project can be employed by the State to administer and expand health care coverage. Such analysis can be used to identify areas disproportionately impacted by disease, and policy can be devised to help mitigate the issues.

**Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data & Link  (Click for Link) | Source | # of Rows | Datum | Unit of Analysis |
|  |  |  |  |  |
| [Colorado COVID-19 Positive Cases of Infection by County](https://hub.arcgis.com/datasets/CDPHE::colorado-covid-19-positive-cases-and-rates-of-infection-by-county-of-identification) | ArcGIS Hub | 67 |  | Vector |
| [Colorado Census Data](https://www.census.gov/quickfacts/CO) | United States Census | -- |  | -- |
| [Food Stores in Denver](https://data.colorado.gov/Business/Food-Stores-in-Denver/hysf-mrke) | City and County of Denver | 588 |  | Vector Point |
| [2020 USA Median Household Income](https://learngis2.maps.arcgis.com/home/item.html?id=20a60423d37c49ba9253526859ba93e1) | ESRI Living Atlas | 62,439 |  | Shape\_Area |
| Data Description |  |  |  |  |

[Colorado COVID-19 Positive Cases of Infection by County](https://hub.arcgis.com/datasets/CDPHE::colorado-covid-19-positive-cases-and-rates-of-infection-by-county-of-identification)

This dataset is published by the Colorado Department of Public Health and Environment and contains the number of COVID-19 positive cases by county, county rate of infection per 100,000 persons, death data by county, statewide COVID-19 prevalence data and associated statewide COVID-19 related statistics.

[Colorado Census Data](https://www.census.gov/quickfacts/CO)

The census data is used to understand population size and demographic features related to each county in the state of Colorado.

[Food Stores in Denver](https://data.colorado.gov/Business/Food-Stores-in-Denver/hysf-mrke)

Data indicates store type, name, location address, etc., and will be used to determine socioeconomic implications, as shopping standards and methods vary across socioeconomic fronts. For example, different people shop at 711 as compared to higher-end convenience stores such as Russell’s.

[2020 USA Median Household Income](https://learngis2.maps.arcgis.com/home/item.html?id=20a60423d37c49ba9253526859ba93e1) *(\*\*\*Not used in SQL. Only used on ArcGIS mapping*)

This layer shows the median household income in the U.S. in 2020 in a multiscale map by country, state, county, ZIP Code, tract, and block group.

**Analysis**

**Data Collection & Cleaning**

After the data was collected through the sources outlined in the Data section, they were cleaned and processed through Python. Very little cleaning and pre-processing was needed, as the sources previously prepped the data for usage, however I did check the validity and identified any null values. Null values were identified and removed or imputed with the average.

The data was imported to PostgreSQL using the PostGIS shapefile import manager.

**Diagram

Description automatically generatedDatabase Management**

Using the depicted relational database, I joined the tables based on shared geoms and county names, enabling a smooth flow of information and queries to be formed utilizing each table. This ER Diagram represents the tables used within SQL queries. As the dataset [2020 USA Median Household Income](https://learngis2.maps.arcgis.com/home/item.html?id=20a60423d37c49ba9253526859ba93e1) was only used in ArcGIS Pro, it was not included within the diagram above.

**Results**

**Counties with the highest number of confirmed cases?**

To start the analysis, I wanted to quickly look at the breakdown in cases between counties. To identify the counties with the highest number of confirmed cases, I ran a query on the [Colorado COVID-19 Positive Cases of Infection by County](https://hub.arcgis.com/datasets/CDPHE::colorado-covid-19-positive-cases-and-rates-of-infection-by-county-of-identification) dataset, selecting the county, population, positivity rate, and total number of cases. Denver is identified as the area with the largest number of confirmed cases, followed by Adams county and Arapahoe county. With a large percentage of the state’s population residing in these areas, these results are unsurprising, yet provide a starting point for deeper analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **COUNTY** | **COUNTY POP** | **COUNTY RATE** | **TOTAL POSITIVE** |
| Denver | 729239 | 2827.74 | 20621 |
| Adams | 517885 | 3134.67 | 16234 |
| Arapahoe | 656822 | 2207.9 | 14502 |

**Map

Description automatically generated**

**Counties with the largest land mass?**

|  |  |
| --- | --- |
| **COUNTY** | **SQ KILOMETERS** |
| Las Animas | 13095.01 |
| Moffat | 12572.77 |
| Weld | 11061.1 |
| Mesa | 9000.459 |
| Rio Blanco | 8800.926 |

Next, I wanted to affirm that the number of positive cases related to population and not land mass. To conduct this analysis, I ran a SQL query using ST\_Area to identify the counties with the largest land mass and limited the results to the top 10. I had to offset the results by 3, as the first 3 rows provided descriptive statistics such as averages. The table below shows the top 5 counties with the largest land mass.

SELECT county,

ST\_Area(geom)/100 as sq\_kilometers

FROM public.colorado\_covid

ORDER BY ST\_Area(geom) DESC

LIMIT 10 OFFSET 3;

**Which counties have the largest percentage of black residents?**

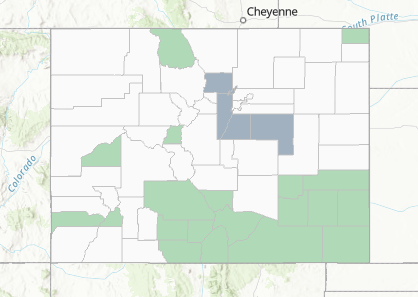
Now that it is confirmed that population has a greater impact on positive cases than land mass, I wanted to look at the breakdown in population demographics. As the project introduction pointed out, minority populations have been disproportionately impacted by the pandemic. The following query identifies the counties that have the largest percentage of black residents.

|  |  |  |  |
| --- | --- | --- | --- |
| **COUNTY** | **COUNTY POP** | **BLACK POP** | **% POP Black** |
| "Adams" | 452396 | 10706 | 42 |
| "El Paso" | 615884 | 33670 | 18 |
| "Arapahoe" | 573116 | 37428 | 15 |
| "Denver" | 620908 | 61649 | 10 |

Again, Denver, Arapahoe, and Adams are present in the list, demonstrating that these areas not only have the largest number of positive cases, but also that they have the largest percentage of black residents.

**How do Income Levels Impact COVID-19 Positivity Rates?**

Graphical user interface, application

Description automatically generated

Demonstrated in the visual above, the counties which we identified as areas with the largest number of confirmed cases, Denver, Adams, and Arapahoe, are also the areas with the highest median household income. Indicating that financial factors may not have a substantial impact on positivity rates.

However, when we take a more granular approach and look at particular neighborhoods within the Denver area, we can see that there are a number of low income pockets contained within the city. Although our initial observation was that income does not have a substantial impact on positve cases, this visualization provides an alternative view and demonstrates that finances, in fact, may play a role.

**Map

Description automatically generatedMap of Neighborhoods within Denver, CO**

Graphical user interface, application

Description automatically generated

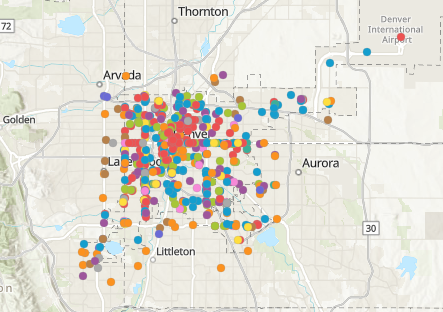
**What are the most popular stores in Denver?**

|  |  |  |
| --- | --- | --- |
| **Store** | **Type** | **Amount** |
| "7-Eleven" | "Convenience Store" | 46 |
| "7-Eleven" | "Convenience Store With Gas" | 30 |
| "King Soopers" | "Supermarket" | 29 |
| "Family Dollar" | "Dollar Store" | 17 |
| "Safeway" | "Supermarket" | 15 |

For a deeper look at financial background and COVID-19 positivity rates, we will look at the specific type of stores most prevalent throughout the city. This will give us an idea on purchase habits, income levels, and price points. Low-income families are more reliant on 7-Eleven’s and dollar stores, while more affluent families can stretch their income and frequent more traditional supermarkets and upscale convenience stops.

My analysis demonstrates that 7-Elevens dominant the area, signaling that lower-end, convenient products make up most purchases in the area. Family Dollar, an all-purpose dollar store, also accounts for 17 stores within the area, further suggesting this theme.

In addition, the visualization provided depicts the store locations. Using the vector points, I graphed h the specific locations of the stores within the dataset. Encoded in Red are the 7-Eleven’s, and we can see that they are heavily concentrated centrally within the city of Denver. From this, we can infer that central residents are lower-income and rely heavily on convenient purchases, while the outskirts and suburbs are more affluent families that purchase goods at supermarkets and higher-end stores.



Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generatedGraphical user interface, text, application

Description automatically generatedMap

Description automatically generated

**What is the relationship between store type, city, and percent of black resident deaths in the area?**

With an understanding that the city of Denver has the largest population, boasts a large percentage of black residents, and contains pockets of low-income areas, I wanted to look at the percentage of black resident deaths and the corresponding common stores in the area. This couples financial, demographic, and purchase habit information, utilizing data from each table involved in the analysis. Using the ST\_Within function to find the store locations within the shape areas provided by the Colorado COVID-19 county database enabled me to identify the common stores in the area. Next, I specified the state name and county name from the census data, accomplishing a utilization of each of the three tables and successfully producing a table with common stores, the type, corresponding city, number of county deaths, and percentage of black resident deaths in the area.

The results were striking, as the percentage of black deaths was much lower than I hypothesized. With black residents accounting for roughly 10% of the population in Denver proper, I expected the percentage of black resident deaths to exceed this amount. However, the results displayed the opposite effect, indicating that black residents in the area may not be at a substantially greater risk.

I do want to point out, however, that the analysis was very limited and direct linkages from income status to deaths rates were not included within this data set. Rather, I was only able to incorporate income levels via an analysis on area and common stores. In the future, and for a more application-based analysis, I would use data that painted a clearer picture on individual family income levels and individual area death rates.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STORE | TYPE | CITY | AMOUNT | COUNTY DEATHS | % Black DEATHS |
| "7-Eleven" | "Convenience Store" | "Denver" | 46 | 452 | 0.007332 |
| "7-Eleven" | "Convenience Store With Gas" | "Denver" | 30 | 452 | 0.007332 |
| "Family Dollar" | "Dollar Store" | "Denver" | 17 | 452 | 0.007332 |
| "Conoco" | "Convenience Store With Gas" | "Denver" | 13 | 452 | 0.007332 |
| "King Soopers" | "Supermarket" | "Denver" | 13 | 452 | 0.007332 |
| "Russell's" | "Convenience Store" | "Denver" | 13 | 452 | 0.007332 |
| "Walgreen" | "Pharmacy" | "Denver" | 13 | 452 | 0.007332 |
| "Diamond Shamrock" | "Convenience Store With Gas" | "Denver" | 10 | 452 | 0.007332 |
| "Safeway" | "Supermarket" | "Denver" | 10 | 452 | 0.007332 |
| "Dollar Tree" | "Dollar Store" | "Denver" | 9 | 452 | 0.007332 |

**Conclusion**

From the analysis, income levels impact consumption habits, as evidenced through the common stores, such as convenience stores and dollar stores, centered around lower-income areas, however it was inconclusive that financial factors and demographic factors heavily influenced COVID-19 death rates.

We can see that death rate and area population are heavily correlated, as demonstrated through the identification of Denver, Arapahoe, and Adams county as the areas with the highest number of deaths, and we can see that this is a stronger factor in COVID-19 death rates than income levels. Income levels in these areas are higher than in other areas within Colorado, yet the positivity rates and deaths rates are also greater.

While the analysis affirms the commonly shared thoughts surrounding the contraction of the virus, the analysis does not provide any additional insight. However, as this was a personal project aimed at better understanding the area, this result and affirmation was sufficient.

Changes for Future Analysis

For future research and a more accurate depiction on the linkage between socioeconomic factors and COVID-19 positivity rates, I would add additional datasets and take a more refined approach to identifying income levels in particular neighborhoods. Doing so would enable me to dive further into the differences between neighborhoods in Denver and provided a clearer understanding on how financial factors impact COVID-19 positivity and death rates. I would also look more in-depth at deaths by location. Similarly, this will provide a more granular approach to analyzing the disparities in area and financial factors on health.

Project Reflection

This project affirmed my interest in GIS applications and provided great insight into real-world applications. The exposure to new technologies, such as ArcPRO Online, were invaluable in the creation of this project, as I was able to quickly create visualization through the additional, cleanings, and processing of layers. I was able to handle the SQL queries well and even created a query that combined each of the three tables, using the geographic ST\_Within function to find certain vector points within an area. This was gratifying and a great boost of confidence. While I am pleased with the visualizations and queries, I know the analysis and overall project could be improved with better, more in-depth data sources. Choosing the right data and supplementing these resources is critical for accurate analysis. This was the weakest part of my project, as I was limited in scope and application. Overall, I am disappointed in the outcome of this project due to the incomplete or insufficient data used in the analysis.