**Variable Discussion**

Dependant variable – average cases over 14 days

Independent variables

* Density per county
* Mobility for groceries
* Percent over 65
* Transit scores; public transport
* Climate data
* Active physicians per 100,000

**Active Physicians over 100,000**

**Density per County**

**Transit scores:**transit\_scores - population weighted averages aggregated from town/city level to county. Transit scores - how well a location is served by public transit

**Mobility data:**

The mobility dataset provided by Google demonstrates how visits and length of stay at different places change compared to a baseline. So changes for each day is compared to a baseline value for that day of the week where the baseline is a mediane value for the corresponding day of the week during the 5-week period Jan 3-Feb 6 2020. The 'Residential' category shows a change in duration-the other categories measure a change in total visitors.

Google’s guidance on reading this data recommends to not compare day to day changes and instead, the index is smoother to a rolling 7 day average.

Place categories we shall analyse include:

Grocery and pharmacy: Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.

Transit stations: Mobility trends for places like public transport hubs such as subway, bus, and train stations.

Workplaces: Mobility trends for places of work.

Particularly, the descriptive analysis is to be a plot of changes in visitors to 3 different places relative to a baseline day with the index smoothed to the rolling 7-day average.

**Climate data:**

Average temperature in Jan, 2019 in Fahrenheit

**Stationary Model**

**Time Series Model**

**Stationary Model**

(total covid cases for county) ~ amt that’s 65, active physicians, density of county, transit scores

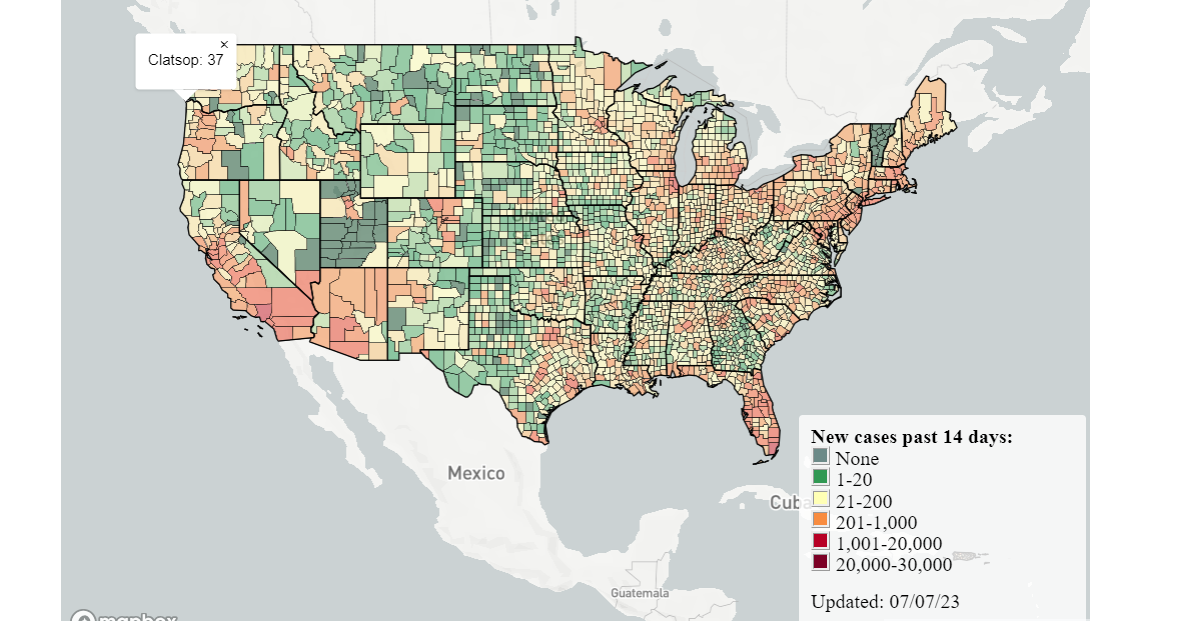
**Time Series model**

Covid cases per date ~ mobility data and climate data

Dissecting the COVID-19 Divide: Why Some US Counties were hit harder than others

After January 20,2020, when Centers for Disease Control and Prevention (CDC) reported the first laboratory-confirmed case of the 2019 Coronavirus, this was promptly announced as a public health emergency on January 31st and became the third-leading cause of death in the US in 2020. Despite the worldwide spread of the pandemic, the effects were profoundly (EVEN? UNEVEN?) in the United States with some areas experience (WHAT RATE) than others. This disparity can be attributed to a combinations of **statistically significant factors** including population density, mobility patterns, age demographics, public transportation scores, climate conditions and healthcare resources.

[INSERT GRAPH: The covid spread]



**Population Density: Urban-Rural Separation**

The density per county, which measures the number of people living in a given area within a county, is an important factor in the spread of infectious diseases like COVID-19. Higher density means more people are concentrated in a smaller area, increasing the likelihood of person-to-person transmission through close contact. By analysing density per county, we can assess how different population concentrations impact the rate of new infections. This information can help identify areas that might need more stringent public health measures to control the spread of the virus.

**[urban spread of people + show most covid cases]**

**Mobility Patterns: Grocery Store Effect**

Lockdowns and stay-at-home orders were implemented in numerous states as a means of slowing the spread of the virus by enforcing physical distance between people. However, these stringent regulations were variable between states and a Columbia University model estimated 54,000 deaths would have been prevented if the enacted restrictions started earlier in a few states.

How effective have these policies been in reducing human movement? What impact did they have on how people work, live and visit? Insights can be drawn from Google’s COVID-19 Community Mobility Reports. Using anonymized data from apps like Google Maps, these reports provide a regularly updated dataset showing how people's movements have changed throughout the pandemic.

Grocery and pharmacy: Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.

Transit stations: Mobility trends for places like public transport hubs such as subway, bus, and train stations.

Workplaces: Mobility trends for places of work.

[insert plot of general mobility over time]

Counties with higher mobility to grocery stores saw increased transmission rates. This is particularly evident in suburban areas where residents travel more frequently for essentials.

[insert top 5 highest mobility trends to grocery stores transmission rates]

The mobility dataset provided by Google demonstrates how visits and length of stay at different places change compared to a baseline. Changes for each day are compared to a baseline value for that day of the week, with the baseline being a median value for the corresponding day of the week during the 5-week period from Jan 3 to Feb 6, 2020. Google recommends not comparing day-to-day changes directly. Instead, they suggest smoothing the index to a rolling 7-day average for more accurate analysis.

**Age Demographics: Vulnerability of Elderly**

The age distribution of a population, particularly the proportion of individuals over 65, is essential in studying COVID-19 spread and impact. Older adults are more susceptible to severe illness and complications from COVID-19. They are also more likely to require hospitalisation and intensive care. By examining the share of the population over 65, we can assess the potential burden on healthcare systems and identify regions where protective measures and vaccination campaigns might need to be prioritised to protect this vulnerable group.

**[Spread of 65+]**

**[table of highest vs lowest 65 + covid cases]**

**Transit Scores: Public transport usage**

Public transportation systems can act as a vector for virus spread, especially in counties with high transit scores. Buses, subways and trains, where social distancing is challenging, can quickly become hotspots. Data shows relationships with extensive public transport networks and higher COVID 19 cases, particularly in the initial phases of the pandemic.

[insert spread of transit scores over America graph]

[insert table with counties w top 5 highest transit scores and their covid cases vs lowest 5]

Transit scores - how well a location is served by public transit

**Climate: The highs and lows of average temperatures**

Research shows that COVID-19 tends to increase as temperature and humidity fall and despite year around activity, COVID-19 had seasonal spikes. It can be noticed with extreme weather conditions …. (DOES COVID INCREASE/DECREASE?)

[insert graph of climate data around America]

With more extreme weather conditions predominantly in (NORTH?SOUTH?) areas, the climate could be a significant factor in the distribution of COVID-19 cases.

**Healthcare Resources: Active Physicians per capita**

The availability of healthcare resources, such as active physicians per 100,000 population, is a critical factor in managing and mitigating the effects of the COVID-19 pandemic. Areas with a higher number of active physicians can provide better medical care, testing, and treatment, which can help control the spread of the virus and reduce mortality rates. By analysing the distribution of active physicians, we can understand the capacity of different regions to respond to the pandemic and highlight areas that may need additional support and resources to effectively manage the healthcare demands caused by COVID-19.

**[spread of physicians + include lowest covid cases]**

**Conclusions:**

The spread of COVID-19 in the United States is a multifaceted issue influenced by an interplay of demographic, behavioral, and environmental factors. High population density, increased mobility, a larger elderly population, extensive public transportation use, climate variations, and healthcare resource availability all contribute to the varied impact seen across different counties. Understanding these factors is crucial for formulating targeted public health responses and preparing for future pandemics. By leveraging this data, policymakers can identify at-risk areas and allocate resources more effectively, ensuring that the most vulnerable populations receive the protection and care they need. This holistic approach is essential for mitigating the impact of COVID-19 and improving the resilience of communities against future public health threats.