**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

**Age Over 65**

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.

**Active Physicians over 100,000**

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.

**Density per County**

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.

**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