



## INTRODUCTION

### ABSTRACT

The dataset originates from Kaggle, published by Sobhan Moosavi. The original data has over 7 million observations of accidents across the U.S. with 46 variables, and the Georgia subset includes 169234 observations across 15 variables. In recent years, it appears as if behavior on the roads has become riskier and more unpredictable, which calls for analysis of recent trends in car accidents. Approaching this topic through the lens of the COVID-19 pandemic stems from the result of lockdown and lighter restrictions on obtaining a Driving License in 2020. This study maps and graphs the frequency, severity, and hotspot locations of car accidents before and after the first year of the pandemic.

### VARIABLES

- County* – The GA county where the accident took place.
- Severity* – Rating from 1 to 4 indicating how long traffic was affected.
- Start\_Lat* – The latitude of the area affected by the accident.
- Start\_Lng* – The longitude of the area affected by the accident.
- covid* – Before, during, or after the first year of COVID-19.
- accidents* – Count of accidents by county.
- population* – The total population of each county.

### DATA CLEANING

The original dataset was subset from all U.S. states to only include accidents in Georgia, resulting in 169234 observations. The *date* variable was created from a datetime variable for use in creating the *covid* variable and used for graphing. A count variable, *accidents*, was created from summing the total accidents by county and stored in a separate dataset. The county population was added and used to plot the GA leaflet maps.

## METHODS

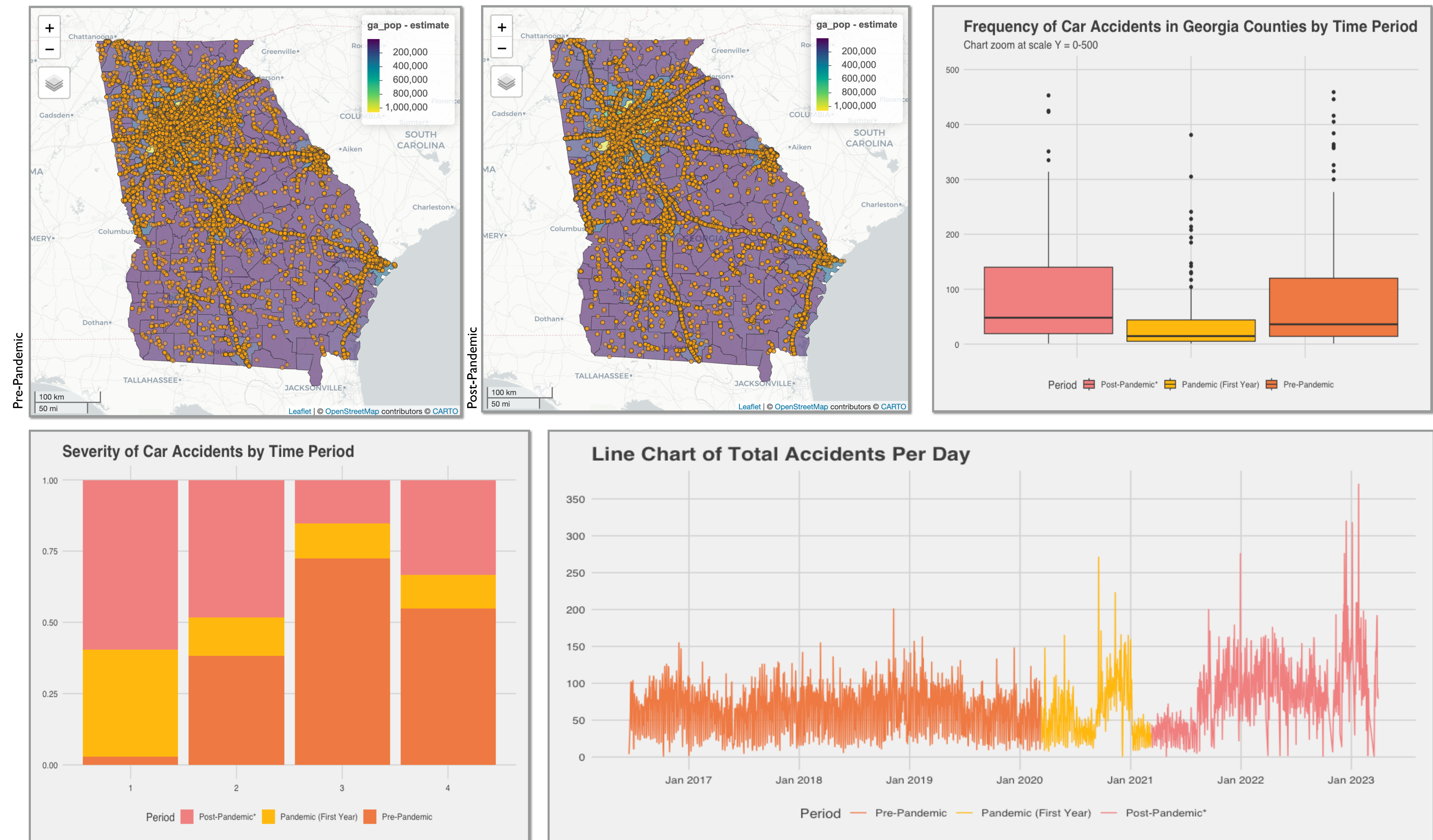
**GGPLOT2** A data visualization package that facilitates the creation of graphs/plots in R. Used to create boxplots, a bar chart, and a time series line chart.

**TIDYCENSUS** A package in R that allows access to U.S. Census Bureau data. Used to join FIPS codes to counties in the dataset so the *population* variable could be added. *Location\_type* and *relative frequency* variables were created using *population*, however both were dropped from the final analysis.

### MAPPING

- Leaflet** is a JavaScript library that facilitates the creation of a variety of map types.
- Mapview** is another mapping package used in conjunction with leaflet to create an interactive map.
- Kernsmooth** is a package that enables the user to add a Kernel Density image overlay on a leaflet map.

## RESULTS



## CONCLUSIONS

**GA Accidents Choropleth Maps** The maps provide an overview of the distribution of car accidents in GA before and after the first year of the COVID-19 pandemic. Visually, there appear to be no significant changes.

**Frequency of Car Accidents Boxplots** The quartiles for the Post-Pandemic boxplot all exceed those in the Pre-Pandemic plot. This suggests that, following the first year of the pandemic, the average accidents per county have steadily increased.

**Severity of Accidents Bar Chart** The bar chart shows that the Pre-Pandemic has the highest proportion of severity ratings 3 and 4, and the Post-Pandemic has the highest proportion of ratings 1 and 2. This indicates that before the first year of the pandemic, there were more accidents that had a greater impact on traffic. This may suggest that accidents have become less inconvenient to traffic, or there may be a greater number of minor accidents than before.

**Line Chart of Total Accidents Per Day** The Pre-Pandemic period saw a consistent average for total accidents per day. The Post-Pandemic period, and during the first year as well, the number of accidents varies significantly more and reaches far higher peaks than before the pandemic. This suggests a general increase in the number of accidents following the first year of the pandemic and potentially a change in overall accident behavior.

**Kennesaw Accident Hotspots Maps** The hotspot locations remain similar in both maps, however the primary hotspot shifts towards the south after the first year of the pandemic. There is a development of a new hotspot towards the northeast of the map which could suggest an increase in accidents. An overall decrease in the lower density regions is to be expected since there is more data available for the Pre-Pandemic period.