Updating and Analyzing Mobility Data for COVID-19 Research

Student: Ethan Haller

Mentor: Mandy Wilson

Background

- **The COVID-19 Mobility Surveillance Dashboard was developed Summer 2021 in collaboration with Stanford University**
- **The Dashboard uses graphs and maps along with an adjustable timeline to allow users to visualize how COVID-19 case counts and foot traffic have changed throughout the pandemic**
  - **The COVID-19 case counts come from the COVID-19 Surveillance Dashboard, and the mobility data comes from SafeGraph under their Academic Partnership Program**
  - **The Dashboard covers Restaurants, Retail Stores, Essential Retail, Fitness Centers, and Religious Organizations**
- **Due to interests in other projects, the current range of data is limited to January 2020 - July 2021**

SafeGraph

- **SafeGraph provides places data and visitor patterns data based on mobile phone GPS. The data is updated monthly, and there are currently over 11.8 million POIs covered in the US.**

Project Goals

Primary Goal: Bring the Mobility Surveillance Dashboard up-to-date

Other Goals

- **Schedule jobs to keep the Dashboard up-to-date with new data**
- **Make necessary changes to how data was previously collated to improve the effectiveness of displaying the data**

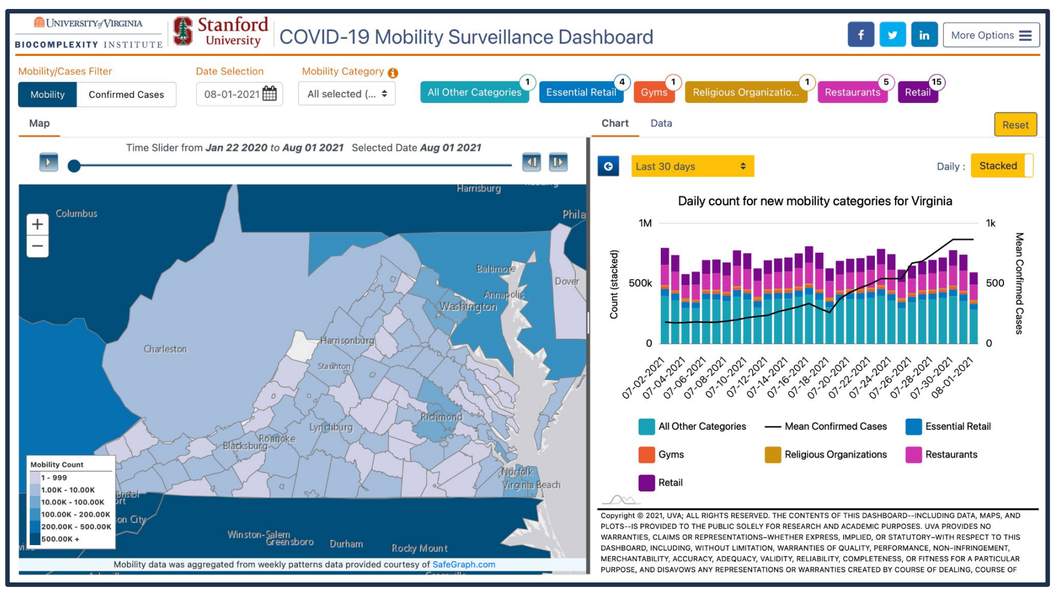

Figure #1. COVID-19 Mobility Surveillance Dashboard

Current Work

- **Identified how data was originally formatted by filtering out POIs in unwanted categories and locations before matching the number of remaining POIs with the value calculated by Stanford**
- **Developed a script using the Python CSV Reader to pull monthly patterns data from Rivanna since August 2021**
- **Filtered out unwanted categories in the data and only kept the following:**
  - **placekey**
  - **date**
  - **num_visits**
  - **region**
  - **poi_cbg**
- **Created a script to load extracted data into a relational database so that it can be used in updating the dashboard**
- **Created a script to extract data from the relational database so that it can be analyzed and used in updating the Dashboard**

Data Field	Description	Value
"placekey"	Individualized "key" for the POI	zzw-222@63f-wt4-btv
location_name	Location name	Carilion Medical Center
street_address	Street address	1906 Belleview Ave SE
city	City	Roanoke
region	State or territory	VA
postal_code	Postal code or Zip code	24014
date_range_start	Start date of data collection	2021-06-01T00:00:00-04:00
date_range_end	End data of data collection	2021-07-01T00:00:00-04:00
raw_visit_counts	Total number of visits	10315
raw_visitor_counts	Total number of visitors	1894
visits_by_day	Visits each day	[389,417,416,392,227,247,407,400,356, ...]
poi_cbg	POI census block group	517700000000
visitor_home_cbgs	Number of visitors from different CBGs	{"511610306001":21,"511610307013":19, ...}
visitor_home_aggregation	Number of visitors from each census tract	{"51770002800":47,"51161030600":43, ...}
visitor_daytime_cbgs	Number of visitors based on work CBG	{"517700028001":93,"517700012001":19, ...}
visitor_country_of_origin	Number of visitors from each country	{"US":1798}
distance_from_home	Mean distance from home travelled by visitors	16528
median_dwell	Meadian minimum dwell time in minutes	122
bucketed_dwell_times	Number of dwell times organized by preset buckets	{"<5":147,"5-10":1303,"11-20":799, ...}
related_same_day_brand	Other brands visited on the same day	{"McDonald's":5,"Walmart":5,"Sheetz":4, ...}
related_same_month_brand	Other brands visited in the same month	{"Walmart":63,"McDonald's":53, ...}
popularity_by_hour	Number of visits by hour over the date range	[1171,1151,1135,1084,1116,1262,2564, ...]
popularity_by_day	Number of visits by day of week over the date range	{"Monday":1570,"Tuesday":2036, ...}
device_type	Number of visitors using Android vs IOS	{"android":926,"ios":967}

Figure #2. SafeGraph Patterns Sample Data

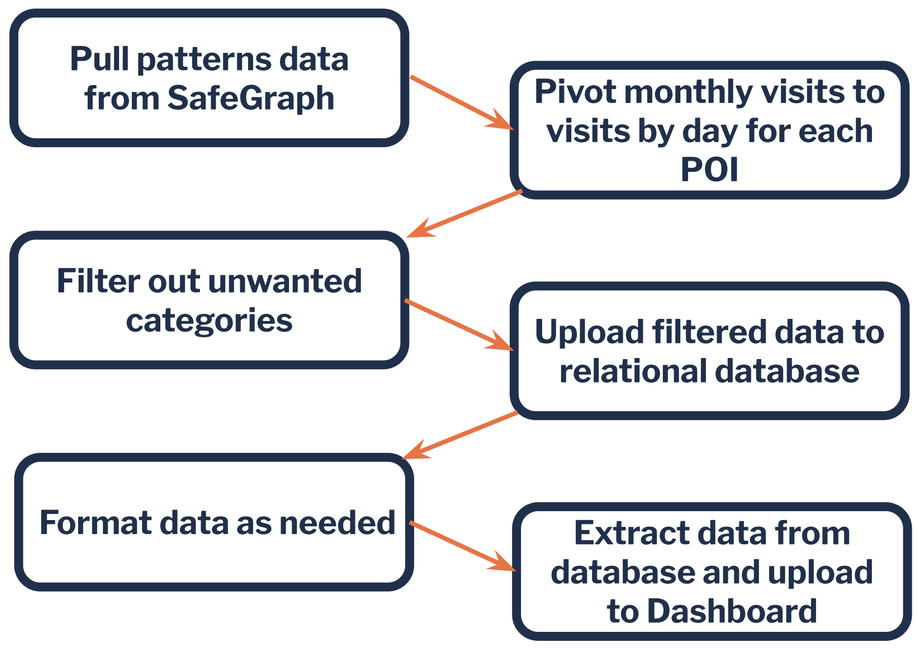

Figure #3. Data collation and formatting process

Challenges Encountered

- **There is no record of how Stanford pulled data from SafeGraph and formatted it to use for the Dashboard**
- **The data format changed over time, so the app has to account for both data formats. GraphQL API was going to be used for data collation, but was unable to account for variations in the data format from SafeGraph**

Memory

- **The data spans over four files for each month, each with millions of rows. These files are a blend of CSV and JSON, so specialized extraction and pivoting of the data was necessary**
- **Filtering data and uploading it into the relational database was originally done with pandas for Python, but was not feasible with the amount of data being used**

Future Work

- **Implement the new and updated data into the Dashboard so there can be a visual representation**
- **Update the Dashboard on a monthly basis by adding new data as it becomes available**
- **Examine the data collected by looking at different categories to see which ones are more significant**
- **Perform an in depth analysis of the data, most likely using AI, to draw meaningful conclusions between foot traffic and COVID-19 cases**

References

- SafeGraph. 2022. SafeGraph Data for Academics. https://www.safegraph.com/academic
- University of Virginia Biocomplexity Institute, & Stanford University. (2021, July). *Covid-19 Mobility Surveillance Dashboard*. nssac.bii.virginia.edu. Retrieved July 25, 2022, from https://nssac.bii.virginia.edu/covid-19/mobilitysurveillance/
- A. S. Peddireddy *et al.*, "From 5Vs to 6Cs: Operationalizing Epidemic Data Management with COVID-19 Surveillance," *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 1380-1387, doi: 10.1109/BigData50022.2020.9378435. (For the COVID-19 Surveillance Dashboard)