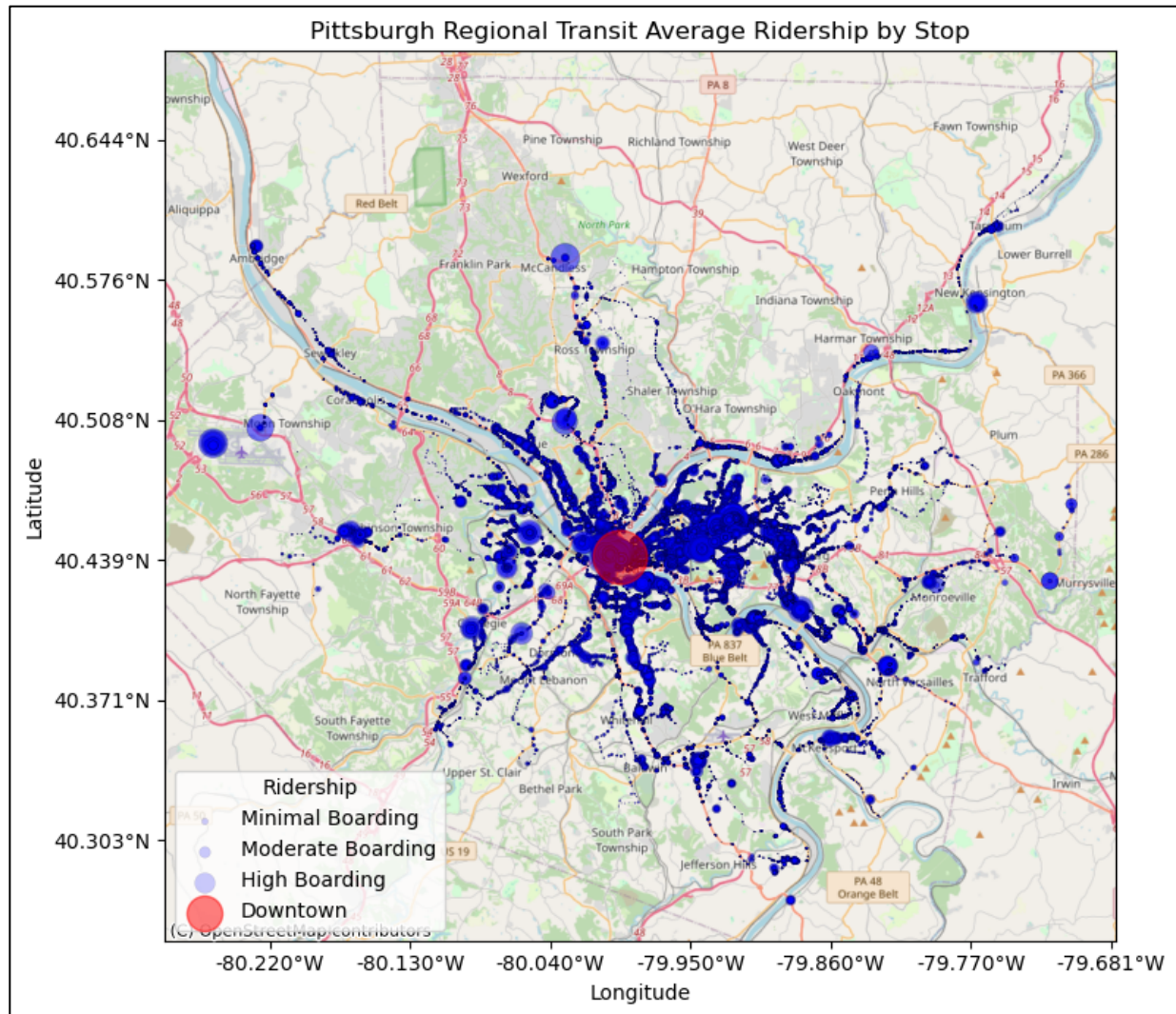


### Public Transit Usage in Pittsburgh, PA



For my visualization, I used the Pittsburgh Regional Transit Monthly Updating Stop Usage dataset from the Western Pennsylvania Regional Data Center. I chose to analyze ridership by looking at the average number of people who used a bus stop by either getting on or off a bus there. I also investigated other details related to ridership, like route direction, service day, and frequented stops. I took the ridership data and visualized it using a scatterplot with different size points to represent how much ridership occurred at each stop. I also chose to add an icon showing the area where downtown Pittsburgh is, to see where most ridership is happening in relation to the city's main urban center. The legend indicates the size points that symbolize minimal, moderate, and high boarding, and the axis labels are for longitude and latitude.



## Findings

- Based on the visualization results, it appears that most ridership is concentrated in one area in the center of the figure.
- After analyzing the figure, I observed that most of the ridership centers around the area marked as downtown. This led me to conclude that high amounts of ridership are more common closer to the main urban area of Pittsburgh. Living or working in an urban area likely requires more use of public transit as less parking space is available.
- The data visualization for most frequented stops also reflects this finding, with four of the top five bus stops being located in downtown Pittsburgh.
- I also noticed that it seems like fewer people are using bus stops and riding the bus in areas further away from the main city center.
- I also identified the top fifteen most frequented routes, with route P1 having the highest daily average riders.
- A substantial portion of these top fifteen routes go through downtown Pittsburgh, further verifying my findings that most ridership occurs close to the city center.
- I also compared riders getting on and off these frequented routes and found that usually the amount of people getting onto routes is similar to the amount of people getting off.
- When looking at service day, I found that most ridership occurs on weekdays. Although routes like the O1 and P3 are very frequented, they do not operate on the weekends, which is shown in the visualization.
- When examining the direction of ridership on routes, I found that most ridership occurs on buses that travel in both directions, rather than just inbound or outbound.

## Data and Methods

The visualizations use the Pittsburgh Regional Transit Bus Stop Usage dataset, which includes information about stops, the locations of stops, details about stops, routes, directions of routes, service days, and the number of people getting on and off at any given stop. To produce my visualizations, I first read in the data from Pittsburgh Regional Transit and learned more about it by checking the data types and dimensions. Then, I dealt with missing values by removing nulls from the data. I also checked for outliers in the avg\_ons and avg\_offs columns to ensure I used accurate ridership values.

To produce the map visualization, I first converted the data frame to a GeoDataFrame using bus stops' longitude and latitude values. I also made a GeoDataFrame marker for the approximate location of downtown Pittsburgh. From there, I made a scatterplot of the bus stops with different marker sizes depending on the average amount of riders getting on and off at that stop. I also plotted the downtown marker and layered these points over a basemap. To produce accurate x and y ticks, I transformed the longitude and latitude values to display as they would on a map. Finally, I added a legend and labels to help describe what the points represented.

For my other visualizations, I first found the top fifteen most frequented routes by grouping by route name and aggregating by average total riders. From there I was able to visualize average ons and offs for those routes as well as look at the service days these routes ran. I analyzed the direction of routes in relation to ridership by grouping by direction and then aggregating by average total riders. Finally, I examined frequented stops by seeing which stops had the highest ridership.

## Significance

The ridership visualization is important to the topic of public transit usage because it could help public transit staff and route planners easily identify where most people are using bus stops and riding buses. This information helps inform where routes are needed and if they are adequate for the amount of people using them. Making effective public transit design decisions enables more people to use public transit, benefiting both the company and its users. Increased public transit usage has been shown to save people money, cause less harm to the environment, and reduce traffic congestion.

## GitHub Link:

[https://github.com/fon215/INFSCI520\\_Project.git](https://github.com/fon215/INFSCI520_Project.git)