Tools Used & Download Instructions:
Python 3.10.2: https://www.python.org/
Matplotlib https://matplotlib.org/stable/
Seaborn https://seaborn.pydata.org/installing.html
Plotly https://plotly.github.io/plotly.py-docs/generated/plotly.html
Shapely https://pypi.org/project/Shapely/
NumPy: https://numpy.org/install/
Pandas: https://pandas.pydata.org/getting_started.html
Pyproj: https://pyproj4.github.io/pyproj/stable/

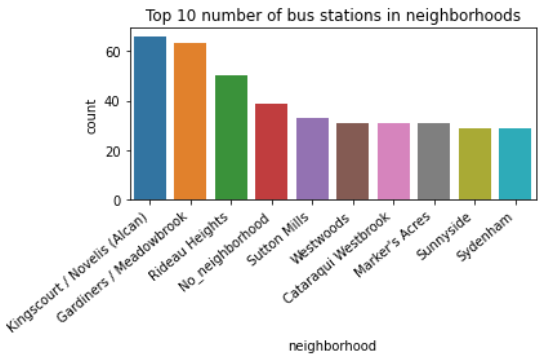
Datasets:

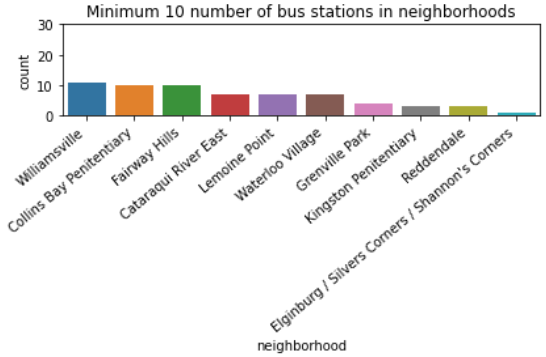
- Provided October 2017 transit usage data
- Neighborhoods: https://opendatakingston.cityofkingston.ca/explore/dataset/neighbourhoods/information/
- Electoral District Census Profiles - Family, Housing, & Mobility: (use it to get populations) https://opendatakingston.cityofkingston.ca/explore/dataset/neighbourhood-census-profiles-family-housing-mobility/table/?disjunctive.community_name
- Transit Bus Stops: https://opendatakingston.cityofkingston.ca/explore/dataset/transit-gtfs-stops/information/
- Transit Bus Routes: https://opendatakingston.cityofkingston.ca/explore/dataset/transit-gtfs-routes/information/?disjunctive.route_short_name
- Driveways (Q3): https://opendatakingston.cityofkingston.ca/explore/dataset/driveways/export/?disjunctive.material
- Road Segments: https://opendatakingston.cityofkingston.ca/explore/dataset/road-segments/information/?disjunctive.gis_class&location=11,44.30764,-76.47755&basemap=72525b

Question 1 : Is the bus service accessible to everyone in Kingston (walk, bike, park & ride to a bus stop)?

We define accessibility as the existing amount of bus stops in a neighbourhood.

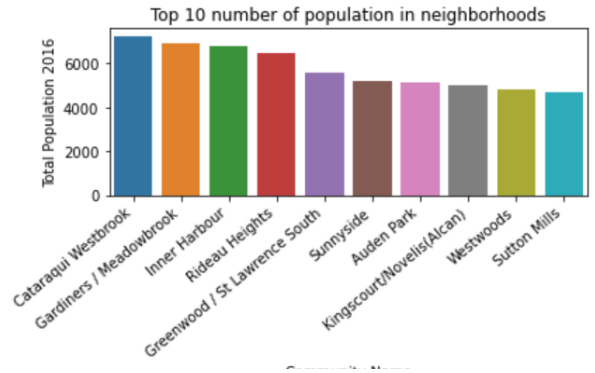
Method
We first used neighborhood.csv dataset and transit-gtfs-stops.csv dataset to find each bus station's neighbourhood. Results are shown below (top and bottom 10 neighbourhoods):

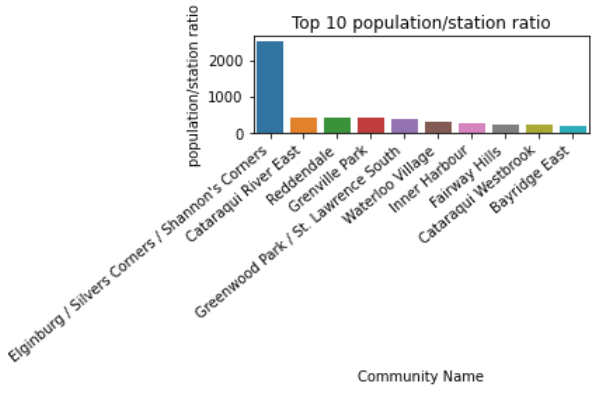




Kingscourt/ Novelis has the largest number of bus stations. There are several neighbourhoods that do not have bus stations, which means that the bus services are not accessible in these neighbourhoods.

The dataset neighbourhood-census-profiles-family-housing-mobility.csv contains the year 2016's population for each neighbourhood. The following bar plots show the most populous neighbourhoods. We can divide the population by the number of stations to find the people/bus station ratio.





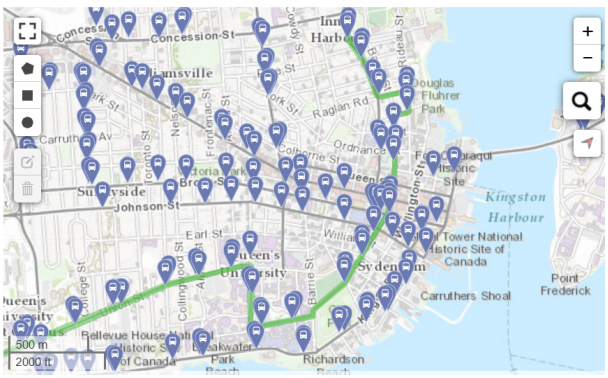
The lower the ratio of people/bus stations, the more accessible bus routes are. For the neighbourhoods Elginburg, Silvers Corners, and Shannon's Corners, we can also say most of the people cannot get access to the bus service, because out of 2500 people, there is a single bus station.

Q1 part 2:
For the analysis above, we are focused on the segments of each neighborhood and discover details precisely according to the bus routes, building information.

For this part of our analysis, we make the assumption that bus stations must be at most 1km away from houses in order to be considered a reasonably accessible (15 minute walk). We use this in conjunction with bus routes and housing information to approximate how accessible bus routes are.

Housing data:
For single-detached homes, the “building type” feature is empty. For apartments, the “building type” feature is either high-density residential or mix-use. There are 58879 houses and 444 apartments in Kingston.

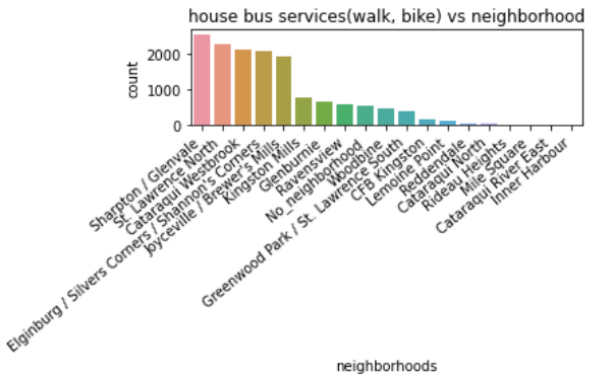
Busing data:
We can source busing routes and stops from the City of Kingston and produce the following maps:

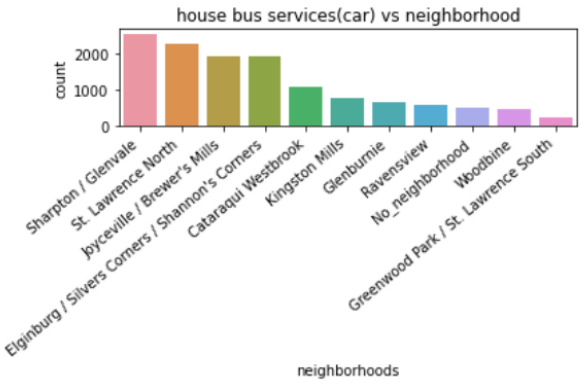


Preprocessing: delete out-of-service routes and use dummy data according to geo_json coordinates to create records for bus routes.

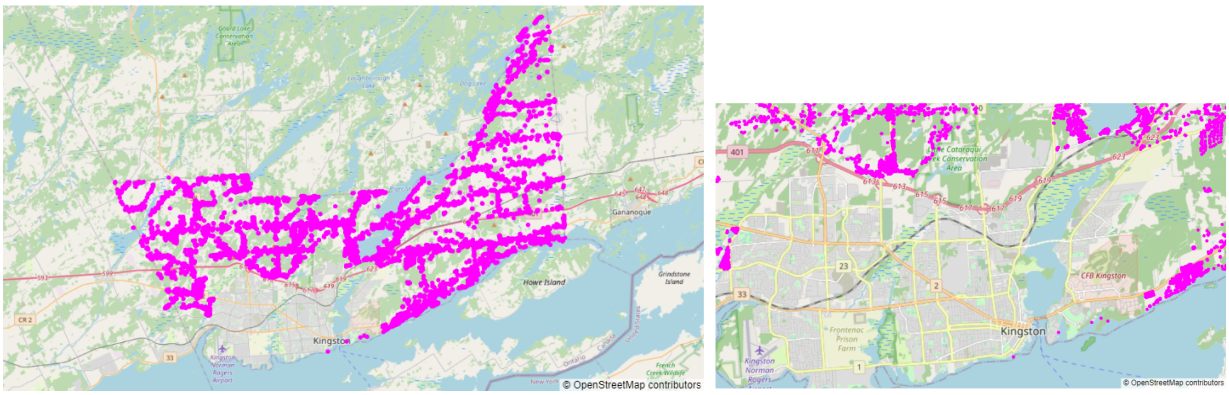
Method:
1. label each coordinate of the bus stop, to find which bus route stops at this station
2. label each bus stop, house/apartment with their respective neighbourhood
3. for each house/apartment, first find all bus stations in the same neighbourhood and calculate the distance between home and bus stations.

There is 1 apartment and 13053 houses that do not have a bus stop within 3km. There is 1 apartment and 14852 houses do not have bus services within 1km. These houses do not meet our definition of reasonably accessible.





Combined with neighborhood data, we can show the number of houses where bus services are not accessible in each neighborhood. Neighbourhoods with no bus stations such as Sharpton and St. Lawerence north have lots of houses with no bus services. These are mapped below:



The pink dots represent houses without reasonable access to bus services. It is clear the locations closer to downtown have greater accessibility.

Summary for question 1:
From **part 1**, we found several neighborhoods have no bus stations, which means bus services are not accessible. 10 neighbourhoods in total did not have a bus station within 3km. Of these are Sharpton, Glenvale, and Ravensview.

We also calculated the population/station ratio; a higher ratio would imply less accessibility.

From **part 2**, we calculated the distance each house was from the nearest bus stop and found that 1 apartment and 14852 houses did not have reasonable access to Kingston's bus system.The vast majority of apartment buildings were close to bus stations.

Question 2: Are there some redundant bus routes that can be cancelled?

Approach
We will define a redundant bus route as a bus route that could be cancelled without reducing the transit system's overall coverage, accessibility, and efficiency.

1. **Determine the most used bus routes** - these are Kingston's most important routes and cannot be deemed redundant.
2. Of the remaining bus routes, consider those that service the same neighbourhoods or cross paths and **determine if routes can be cancelled.**

Data Limitations
While the bus route data is up to date, the transit usage data provided is from 2017 while the population census data is from 2016. Because bus routes have not changed significantly since 2016, COVID-19 data is not relevant (considered special circumstance), and population has not changed significantly, this data is acceptable.

Part 1. Determining the Most Used Bus Routes

We will use the transit data provided for October 2017 and assume this data is representative of typical transit usage.

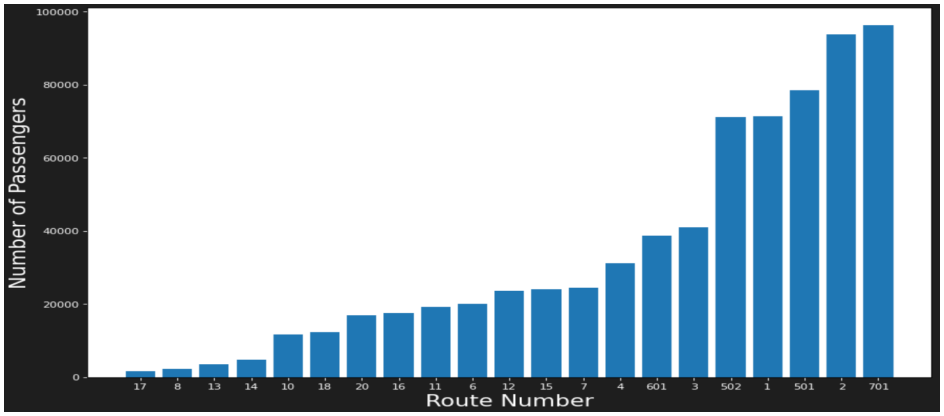


Figure 2.1.1: Most used bus routes in Kingston

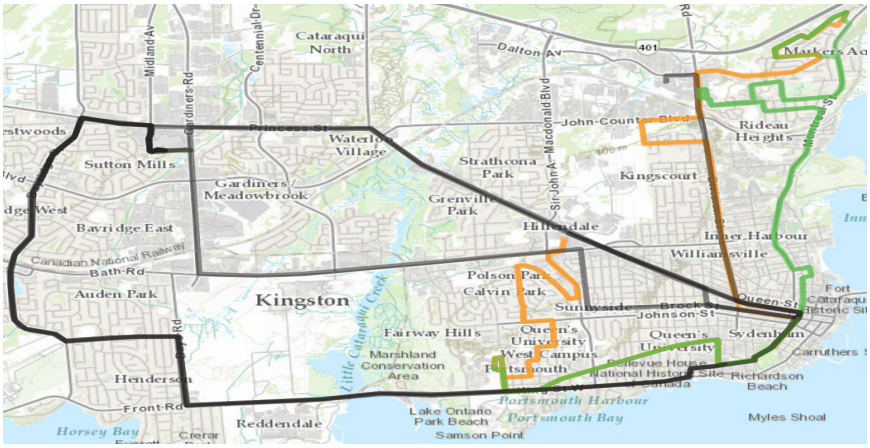


Figure 2.1.2: Grey: 701, Black: 501 & 502, Green: 1, Orange: 2
Bus routes 701, 501, 502, 1, and 2 are by far the most used bus routes.

Route 501 and 502 are responsible for half of the loop each and travel in different directions - If either route were to cease operations, the loop would be too long to be efficient. Therefore, they are not redundant. The same holds true for routes 601 & 602, 701 & 702, and 801 & 802.

Part 2. Determining Redundant Bus Routes

How redundancy is determined:

1. **Frequency of use** (Data source: provided 2017 October data)
   Routes that see a lot of use should not be deemed redundant because even if their area coverage overlaps other routes or if they serve small communities, high-demand implies necessity.

2. **Area coverage** (Data source: Kingston bus map)
   If certain bus routes' stops overlap with other bus routes' stops, it is a factor in deciding whether a route is redundant or not, as overlapping routes may not need to coexist.

3. **Community population** (Data source: Family, housing, mobility census data)
   If multiple routes serve a sparsely populated area, some of these routes could be considered redundant.

Examination: Route 11 - <u>Deemed Redundant</u>

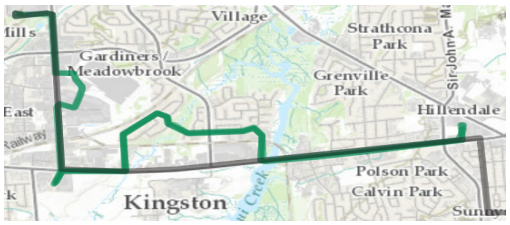


Figure 2.2.1: Grey: 701, Green: 11

Route 11 can be considered redundant because its stops are almost all already covered by Route 701 and it is used infrequently. Route 11 is the 13th most used bus route while Route 701 is the most used route.

Examination: Route 4 - <u>Not Redundant</u>

Figure 2.2.2: Black: 502, Pink: 4 (route partially pictured above)

The only difference between Route 4 and Route 502 is that Route 4 stops more often. They are otherwise identical. However, both travel between the high traffic areas of Cataraqui Centre and downtown Kingston, and both routes are used more than 30 000 times per month.

Redundancy of Certain Route Segments

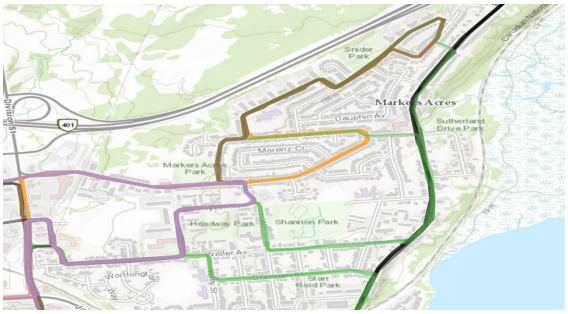


Figure 2.2.3: Bus routes in the Markers Acres Area.

Certain areas of Kingston are overserved while others are underserved. For example, the Markers Acres neighbourhood is served by 8 nearby bus routes. With a population of 2860, the neighbourhood is Kingston's 20th most populous neighbourhood.
Meanwhile, Cararqui Westbrook, Kingston's most populous neighbourhood with 7235 people, has only 4 bus routes with stops nearby, and none within the community.

Therefore, the portions of many bus routes that go into the Markers Acres area can be deemed redundant, and fewer stops in the area would likely improve the efficiency.

Question 3: How many additional bus routes does Kingston need to accommodate all working adults, and what's the net savings on gas?

Question 3 Abstract: The following process has been designed to accommodate some simplifications. In real life, not all routes go downtown, and some routes are obviously more busy than others. we seek to find an *average* capacity per existing bus route in Kingston. Then we simplify so that every 1 additional bus route added will increase the total capacity of the system by that average amount, AND that these routes will get people downtown. We interpret the question as such: How many additional bus routes does Kingston need to accommodate all working adults, and what's the net savings on gas?

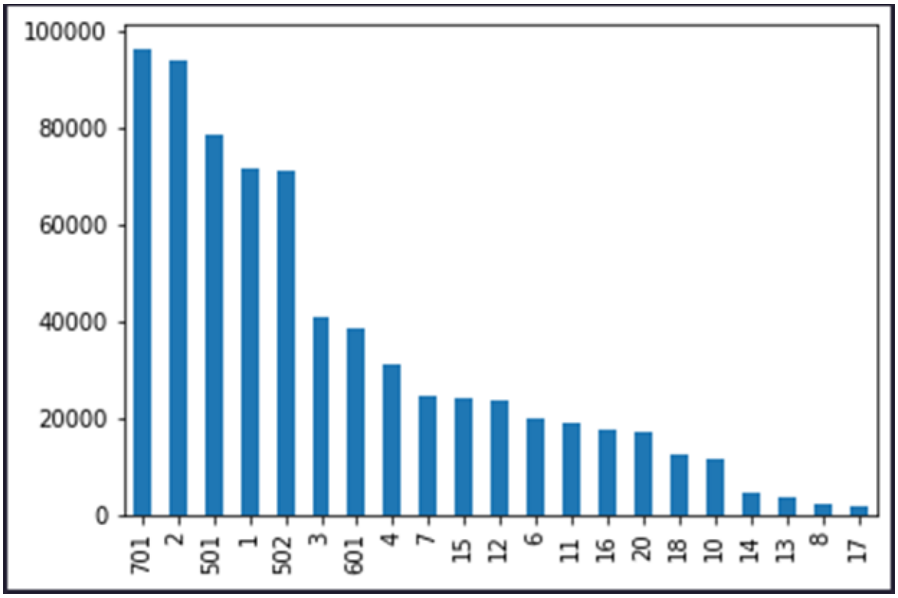
Additional #Bus Routes Needed = Total #Bus Routes Needed - Existing # of Routes
Total #Bus Routes Needed = Required Commutes per month / Bus Route Capacity
Required Commutes per month = # of Working People in Kingston * 2 times a day * 30 days
Bus Route Capacity per Month = Riders per Bus per Month * Average Buses per Route
Net Gas Saved per Month = Total Theoretical Gas Used Now per Month – Total Gas Used by Solution per Month
Total Theoretical Gas Used Now = (2 * Gas used per Car per month * Driveways in Kingston) + (Existing # of Active Buses at Rush Hour * Gas used per Bus per month)
Total Gas Used by Solution per Month = Total #Bus Routes Needed * Average Buses per Route * Gas used per Bus per Month

Required Inputs	Estimated/Calculated Values
Bus Route Capacity per Month *(also Mean rides per route per month)*	33550 (used to justify route 4 as benchmark)
Riders per Bus per Month ~ 7800	~7800
Average Buses per Route	~4.3
Existing # of Active Buses at Rush Hour	~90
# of Working People in Kingston	~92474
Required Commutes per month	~5,548,440
Existing #of Routes	21
Total #Bus Routes Needed	~165
Additional #Bus Routes Needed	~144 (Representing 685% increase in capacity)
Gas used per Car per Month	~40L
Gas used per Bus per Month	~2448L
Total Theoretical Gas Used Now per Month	193,976,360

Total Gas Used by Solution per Month	156,317,040
Net Gas Saved per Month	37,659,320L

Calculation Assumptions:

- Each car represents one working adult. A two-car driveway assumes two working adults per driveway. Thus, each driveway represents two working adults.
- The capacity of the system must accommodate all commuters on buses.



Route	Rides	Estimated #Buses
701	96384	12.35692308
2	93880	12.03589744
501	78520	10.06666667
1	71450	9.16025641
502	71229	9.131923077
3	40953	5.250384615
601	38617	4.950897436
4	31207	4.000897436
7	24508	3.142051282
15	24001	3.077051282
12	23567	3.021410256
6	20116	2.578974359
11	19183	2.459358974
16	17555	2.250641026
20	16939	2.171666667
18	12349	1.583205128
10	11683	1.497820513
14	4777	0.612435897
13	3616	0.463589744
8	2295	0.294230769
17	1710	0.219230769
Total	704539	90.32551282

Figure above plots Route number (x axis) and number of rides (y axis). The chart to the right estimates peak # of buses.

Process to estimate Existing # of Buses:

We could not find data for how many buses exist in the public transit system. So, we estimated how many buses exist through analyzing the schedule for a route with a usage rate close to the mean for all routes (this happened to be route 4):

1. Checking how long it takes to complete the route (Google Maps from terminal to terminal suggests roughly 30 minutes on average, times 2 for the way back) results in estimated 1 hour cycles for a single bus on route 4)…

2. …and scheduled times between buses (15 minutes for rush hours, 30 minutes for others). We are interested in peak capacity and thus using the 15-minute figure. Thus, we estimate route 4 has 4 buses on route (60 minutes / 15 minutes = 4).
3. To extrapolate how many buses are on each of the other routes, we take the ratio of buses to riders per month.

   Python reveals that Route 4 had 31207 riders in one month. 31207 riders / 4 buses = ~7800 riders per bus per month.

   Using this ratio, we estimate a total of 90 buses at peak hours.

To estimate **# of Working people** in Kingston, we feel it is in the spirit of the question to use driveways as a proxy. That is, it is equal to the number of driveways times 2. From the driveways data from opendatakingston, python detects 46237 driveways. For our purposes, this represents an estimate of 92,474. This estimate is reasonable as the population is 136,685 (2017) sourced from Google.

Because we now know we need ~165 bus routes for everyone but only 21 exist, this means theoretically 21/165 of the Required Commutes per Month are currently done by bus. This is validated by our 704,539 figure from the data. The other 4,843,901 would be done by cars (but in real life nowhere near each of these commutes is done in a solo driver car, this is purely theoretical).

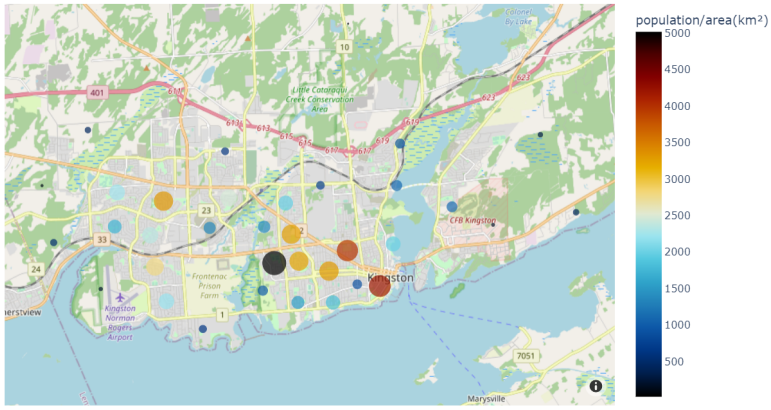
Gas used per Car per Month (and Gas used per Bus per Month) is very roughly estimated as follows:

- Car Average commute: ~10km one way, 20km total daily, 20 times a month -> 400km per month
- Car Average liters/km: 10L/100km ~ 0.1L/km
- Current Car Gas usage: 40L * 4,843,901 = 193,756,040L/Month
- Bus Average commute: 6am to 11pm is 17 hours, @1hr per cycle = 17 cycles, and 16km/cycle (route 4 estimate) = 17 * 16km * 30days = 8160km/month
- Bus Average liters/km: ~30L/100km ~ 0.3L/km
- Current Bus Gas usage: = 2448L * 90 Buses = 220,320L/Month
- Current Car + Gas usage = 587,520L/Month
- Total Gas Used by Solution = **165** * 4.3 * 220320 = 156,317,040L

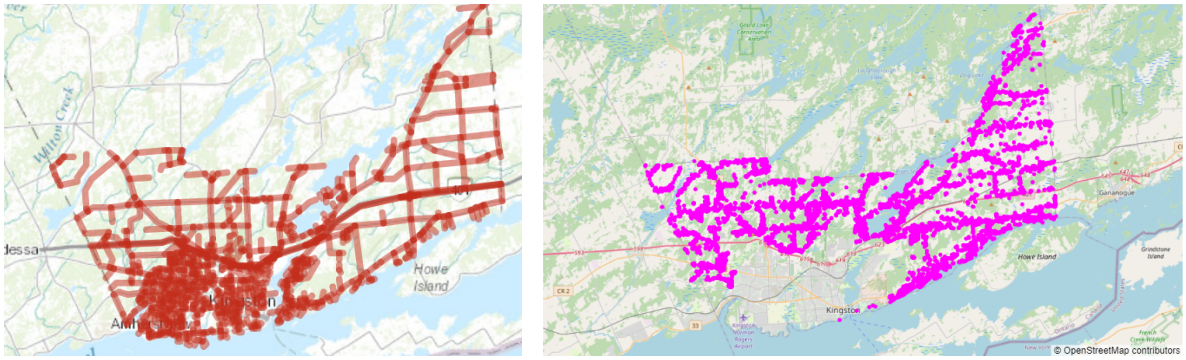
Thus, in a world where everyone is **either on a bus or driving alone** in their car on a commute to work with **average commute distance and fuel efficiency according to our estimates** (10km one way, 0.3L/km and 0.1L/km respectively), Kingston would need 144 more bus routes which would save ~38M litres of gas monthly if everyone were to swap from their cars. Even assuming our estimates are correct, real savings would be much lower because not everyone who doesn't bus to work is driving to work solo.

Question 4: Is there anything else you find interesting in the data that can help with planning the bus routes?

In this question our group focuses on why bus services are not accessible in some neighbourhoods. In question one, we concluded that some communities were less accessible than others. In this section, we examine population, road segments, and income.

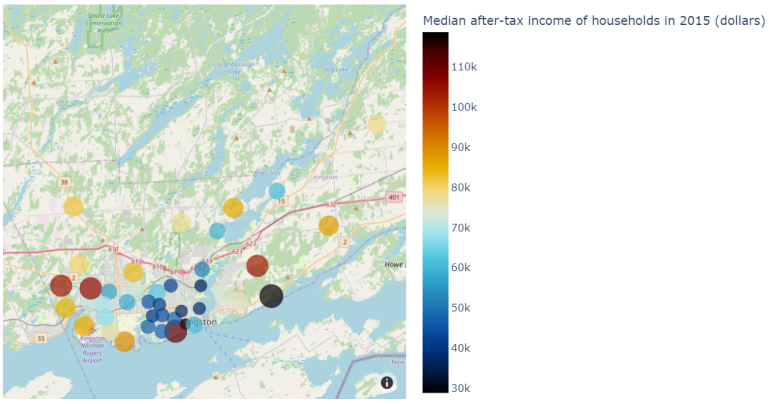


Left: Population by neighbourhood (darker is higher) Right: Neighbourhood zoning
Generally, bus services are more accessible as population density increases.



Left: Road segments Right: Areas without direct bus coverage

The road segment density is much greater than areas without bus coverage. We can observe that areas with low road segment density also have low accessibility for bus service.



Median income by neighbourhood
Central neighbourhoods have lower median income and also have higher bus service accessibility.

Generally speaking, places with high population density, a lot of roads, and high median income tend to have better bus coverage.