Cody Coponiti

Data Management for GIS

Final Project

The Effect of Proximity to Metrorail Stations on Housing Cost in Fairfax Virginia

Abstract:

When looking for a place to live the cost of residency is typically the most influential variable in that decision making process. However other factors such as school districts, distance from work, and access to public transportation play a role as well. These secondary factors are so commonly known by the real estate industry that they actually influence the cost. We see this effect in metro centers particularly where a home of the same or comparable size is sometimes two to three times more expensive than a home in a more rural environment. In major metro centers like New York City, Chicago, and the DMV (DC, Maryland, Virginia) that access to public transportation is crucial to daily life of many of its residents because using a personal vehicle as your primary mode of transportation is difficult or inconvenient (traffic, lack of parking, higher fuel cost). This causes those looking to live in metro areas to highly prioritize homes with close access to

public transportation, particularly trains/subways stations. In this paper we look at Fairfax County Virginia housing cost broken down by zip code and relate that to the zip code's proximity/housing of metrorail stations and can confirm that zip codes with metrorail stations in or with in a half a mile of them have higher percentages of high value homes.

Introduction:

Is there a correlation between housing cost and proximity to metrorail stations in Fairfax, VA? I decided to choose this topic because I moved to Fairfax County VA six years ago and have lived in four different residencies at very different quality but for similar prices. This got me wondering why my first two residences (which were both tiny and kind of rough) were the same value as my more recent places that have gotten progressively nicer and larger as I have moved. What it was (and what brought me to this question) was that my first two places were within walking distance to a metrorail station.

Besides the fact that I live in Fairfax, VA, and my personal connection to it, I choose to use Fairfax over D.C. itself or another city like New York or Chicago was because Fairfax has both urban areas and rural areas. This allowed for clearer contrasting of housing cost between in urban zip codes closer to metro stops and

rural zip codes. I also chose Fairfax because they have and regularly update spatial, demographic, and economic data for the whole county, something I had a hard time finding for D.C..

I think this research will be helpful for those who live in or near metro area because people in these areas tend to move more frequently. Understanding that something as simple as moving one mile over to a zip code without a metrorail station can either save you a decent chunk of change or allow you to spend the same amount and get that extra room/bathroom/walk in closet that they really want. I also think the timing of this research is important too, in a post covid world where a lot of jobs in metropolitan areas have gone to remote or hybrid positions needing to have access to public transport to get to work has become less necessary or moot.

Research Context/Background:

The Environmental and Energy Study Institute has also run a study on the impact of public transportation on real estate values in 2013. In their research they found that there is a direct connection between the cost of homes and their proximity to a well-connected public transit system. In their findings they also state that homes with good access to public transit values stay more consistent with normal economic trends even in times of economic recession, when home values typically crash (EESI 2013).

The National Association of Realtors (NAR) also found evidence that homes with easy access to public transportation leads to a hefty increase in property value. In an article published by the NAR they coined the term “transit premium” because of the factor that access a good metro system affects real estate prices. (NAR 2012)

The American Association of State Highway and Transportation Officials determined in a joint study they conducted with the American Public Transportation Associations determined that residences located with in half a mile of public transit services had higher home values then those further away (AASHTO 2019). This study was the basis for my decision to use 800m (roughly half a mile) as my buffer in this project.

Materials and Methods:

For this project I started my search for relevant data on Data.gov which houses a huge amount of data sets from different levels of government across the United States. Originally, I was going to focus on obtaining data sets focused on Washington D.C.’s metrorail stations and housing values, but unfortunately D.C. did not have much in the way of spatial data, which simply would not do when my project focuses on proximity. However, amongst the D.C. data sets there was a spatial data set on metrorail stations posted by Fairfax County, VA (the search pulled this data set because Fairfax and D.C. are serviced by the same Metrorail system, the WMTA).

From there I was able to find that Fairfax County had its own page on arcgis.com where they stored over 120 different spatial data layers and tables. This was perfect for my research and after looking though several different spatial data sets and reviewing their metadata I was able to identify about 7 data sets that I thought might be useful to my project. I did however cut that

number down to three data sets that I actually utilized (Metrorail Stations, Supervisory Districts, and Zip Codes). These three data sets gave me the perfect base for my spatial data analysis, but I was still lacking my demographic and economic data.

To find this information I went to Fairfax County (not to be confused with Fairfax City) website to find my demographic and economic data. I was able to find the perfect data set for my needs, it had the population and housing demographics as well as the housing value data in one table organized by zip code, community, and block group as a CSV file. I utilized the version using zip codes because I did not have a good data set from the Fairfax arcgis data that paired well with community or block group.

I then downloaded my shapefiles from the Fairfax arcgis page and stored the unzipped data in their own folders inside a folder I made specifically for this project. When uploading the shapefiles to PGAdmin I had to manually assign the SRID from their metadata (which I had to pull from uploading them in to ArcPro). The Zip Codes (table "zip_codes") and Supervisor District (table "supervisor_districts") files had a SRID of 4326 and the Metrorail Stations file had a SRID of 4269 (table "metrorail_stations). To assure that all my spatial data would cooperate with each other as I was running spatial queries, I had to transform the metrorail_stations SRID to match the other two spatial data tables using the following query.

```
SELECT ST_AsText(ST_Transform(geom,4326))
FROM metrorail_stations
```

Next, I uploaded my non spatial data to PGAdmin and created the non-spatial table "pop_zip". After uploading all my data into my PGAdmin database I cleaned up the data tables by deleting the excess columns from my tables that contained information that I would not be

using in my project, making them easier to look at and use. I then needed to make my non spatial data useable in spatial queries, I did so by running a join between pop_zip and zip_codes to create a new spatial table zip_merge (Then deleting pop_zip table to avoid confusion for future queries).

```
--Join spatial and non spatial data to form usable data
create table zip_merge as
select *
from zip_code
inner join pop_zip using (zipcode)
```

I also ran a join to create a new table that would join my demographic data to my supervisor district data as well.

```
create table zip_pop as
select  population,
zm.zipcode, zm.zipcity, sd.district, zm.geom
from zip_merge as zm
join supervisor_districts as sd
on ST_Contains (sd.geom, ST_Centroid (zm.geom))
where population > 0
group by  zm.zipcity, zm.zipcode, sd.district, zm.population, zm.geom
order by zm.population DESC
```

I then ran a spatial query to identify the more urban areas of Fairfax by calculating the population density of the supervisor districts, I went with the supervisory districts because there were a few zip codes in this area that are used only for federal government offices and military facilities that do not have residences or skewed population densities. With this query I was able to that our five urban districts were “Providence”, Braddock”, “Mason”, “Franconia”, and “Hunter Mill”

```sql
-- population desnity of Supervisory Disct (person/km2)
Select
sd.district,
sum (zm.population) / (ST_Area (sd.geom) / 1000000.0) AS popn_per_sqkm
from zip_merge as zm
join supervisor_districts as sd
on ST_Overlaps (zm.geom, sd.geom)
group by sd.district, sd.geom
order by popn_per_sqkm DESC
```

Data Output | Messages | Geometry Viewer | Notifications

	district character varying (12)	popn_per_sqkm double precision
1	PROVIDENCE	59354707954099.43
2	BRADDOCK	44894258788195.4
3	MASON	43336725703651.75
4	FRANCONIA	42752959341217.83
5	HUNTER MILL	28304698888399.34

I then ran a spatial join between supervisor_districts and zip_codes to identify that the five urban districts are composed of 23 of Fairfax's 79 zip codes and created a table district_zips.

	district character varying (12)	zipcode numeric
1	MASON	22311
2	BRADDOCK	22032
3	FRANCONIA	22150
4	FRANCONIA	22315
5	BRADDOCK	22151
6	MASON	22206
7	MASON	22304
8	FRANCONIA	22303
9	FRANCONIA	22310
10	MASON	22302
11	PROVIDENCE	22027
12	MASON	22041

Total rows: 23 of 23 Query complete 0

12	MASON	22041
13	MASON	22203
14	BRADDOCK	22035
15	BRADDOCK	22015
16	MASON	22204
17	FRANCONIA	22306
18	MASON	22044
19	BRADDOCK	22030
20	MASON	22312
21	PROVIDENCE	22031
22	BRADDOCK	22003
23	MASON	22042

Total rows: 23 of 23 Query complete 0

I then ran an 800m buffer (based of the ½ mile claim made by AASHTO) of the metrorail stations to identify which zip codes/districts that the metrorail stations were within.

```sql
-- Station/ zip join
select s.name, z.zipcode, z.district
from metrorail_stations as s
join zip_pop as z
On St_Dwithin,z.geom, s.geom, 800)
group by s.name, z.zipcode, z.district
```

name character varying	zipcode numeric	district character varying
Mclean	22102	Dransvill
Wiehle-Reston East	22102	Dransvill
Tysons Corner	22102	Dransvill
Greensboro	22102	Dransvill
Franconia-Springfield	22150	Franconia
Van Dorn Street	22310	Franconia
Innovation Center	20171	Hunter Mill
Reston Town Center	20190	Hunter Mill
Dunn Loring - Merrifield	22180	Hunter Mill
Spring Hill	22182	Hunter Mill
Virginia Square - GMU	22203	Mason
Vienna/Fairfax GMU	22031	Providence

This allowed me to see that 12 metrorail stations were located in 9 zip codes, with one of the zip codes “22102” being in a district that we determined as not urban. This is because the majority of this district is rural or suburban except for the zip code that contains the city of Mclean.

I then ran a spatial join between zip_merge and supervisor_districts to determine the precent of high value homes (homes with a value of between $500 thousand and $1million)

```
--500k to 1 m
select
100.0 * sum (value_mid)/ sum (households)as midvalue_pct,
zm.zipcode, zm.zipcity, sd.district
from supervisor_districts as sd
join zip_merge as zm
on ST_Intersects (sd.geom, zm.geom)
where households > 0
group by  zm.zipcity, zm.zipcode, sd.district
order by midvalue_pct DESC
```

```
select
100.0 * sum (value_mid)/ sum (households)as midvalue_pct,
zm.zipcode, zm.zipcity, sd.district, zm.population
from supervisor_districts as sd
join zip_merge as zm
on ST_Contains (sd.geom, ST_Centroid (zm.geom))
where households > 0
group by  zm.zipcity, zm.zipcode, sd.district, zm.population
order by midvalue_pct DESC
```

And then cleaned the data by running it with ST_Centroid query to make sure there was no duplication of data. I then ran the same queries for very high value homes ($1m+ value) and those considered low value by Fairfax County (under 500K in value.

```
select
100.0 * sum (value_high)/ sum (households)as highvalue_pct,
zm.zipcode, zm.zipcity, sd.district, zm.population
from supervisor_districts as sd
join zip_merge as zm
on ST_Contains (sd.geom, ST_Centroid (zm.geom))
where households > 0
group by  zm.zipcity, zm.zipcode, sd.district, zm.population
order by highvalue_pct DESC
```

```
Create Table lV_homes as
select
100.0 * sum (value_low)/sum (households) as lowvalue_pct,
zm.zipcode, zm.zipcity, sd.district, zm.population
from supervisor_districts as sd
join zip_merge as zm
on ST_Contains (sd.geom, ST_Centroid (zm.geom))
where households > 0
group by  zm.zipcity, zm.zipcode, sd.district, zm.population
order by lowvalue_pct DESC
```

From the tables created thus far I was set up to complete my analysis of my data sets.

Results:

For my results I ran a simple SQL query for on my previously created tables hv_home_pop, mv_homes, and lv_homes to get the percentages of homes in each value range for our 8 zip codes (removing 22203 because it was a non-residential zip code).

Percentage of homes over $1million by zip codes within ½ a mile of a metrorail station:

```
Select highvalue_pct, zipcode
from hv_home_pop
where zipcode = 22102
or zipcode =22150
or zipcode= 22310
or zipcode = 20171
or zipcode =20190
or zipcode = 22180
or zipcode =22182
or zipcode =22031
Order By zipcode DESC
```

	highvalue_pct numeric	zipcode numeric
1	3.5664915675974567	22310
2	38.396999309051424	22182
3	21.65062174087445	22180
4	1.6003507618108077	22150
5	25.8146109209939	22102
6	5.690794776720941	22031
7	2.1828103683492497	20190
8	6.412644083152967	20171

Percentage of homes between $500K and $1million by zip codes within ½ a mile of a metrorail station:

```
Select midvalue_pct, zipcode
from mv_homes
where zipcode = 22102
or zipcode =22150
or zipcode= 22310
or zipcode = 20171
or zipcode =20190
or zipcode = 22180
or zipcode =22182
or zipcode =22031
Order By zipcode desc
```

	midvalue_pct numeric	zipcode numeric
1	64.21527969772372	22310
2	44.181225940183595	22182
3	51.042920176494185	22180
4	55.19017866929738	22150
5	8.272578485344443	22102
6	42.0683540222473	22031
7	25.80263756252842	20190
8	48.58811093773594	20171

Percentage of homes between $500K and $1million by zip codes within ½ a mile of a metrorail station:

```
Select lowvalue_pct, zipcode
from lv_homes
where zipcode = 22102
or zipcode =22150
or zipcode= 22310
or zipcode = 20171
or zipcode =20190
or zipcode = 22180
or zipcode =22182
or zipcode =22031
Order By zipcode DESC
```

	lowvalue_pct numeric	zipcode numeric
1	23.066998433324116	22310
2	0.4639226137597473	22182
3	7.109907741676695	22180
4	11.750520662062918	22150
5	28.939145960422557	22102
6	13.348379816217959	22031
7	34.07912687585266	20190
8	11.909196154426938	20171

Discussion:

From our results we can validate our original hypothesis that zip codes within ½ mile of metrorail station will have high percentages of high value homes. In looking at our results table we see that all of our zip codes have a higher percentage of $500K+ homes (combining the percentages of our first two results tables) than low value homes (our final results table). The only zip code that had a comparable high vs low value home percentage was 22102 which is the Dansville district, which was the only district that had metrorail stations that we determined as a non-urban district. The other two zip codes that stood out were 22180 and 22182 which both had 70%+ high value homes.

For this project I did have some limitation on the scope of my research and the spatial parameters I would have liked to use. As far as scope I would like to be able to compare these results to other metro areas like D.C. proper, Chicago, and New York. I think having these other cites as comparisons would strengthen the statistical merit of my results. I would also have rather been able to preform this research using a more uniform geographic unit than zip codes like

census tracts, but unfortunately for this project I did not have solid demographic and economic data (which was a nonspatial CSV) that was organized by census blocks/tracts. My only two options being zip codes or communities (which I found to be even more vague).

Conclusion:

After running our test, we can confirm that residences that are closer to metrorail stations have a tendency to have high values. In the case of Fairfax County, VA the majority of houses within a ½ mile of metrorail stations will have a value of over $500k. This information will be beneficial to house hunters, real estate agents, and investors alike. With this information house hunters are now armed with the knowledge that moving as little as a half mile over to a zip code without a metro station that they can avoid the "transit premium" that comes with living near a metro stop.

For realtors this information can be important in both your attempts at selling and the decision on which residences to take on as clients. If a realtor knows a home is near a metro station, they should not be afraid to have a higher selling price than they would for a similar size home that is not close to public transportation. And since realtors frequently work on a commissioned based system it would not be a bad idea to take on homes that are closer to metro access that tend to cost more. And for investors and land developers this information can be important in the decision when developing new land near metro station that focusing on residential is a smart investment.

Bibliography:

The American Association of State Highway and Transportation Officials (AASHTO). "Study Links Real Estate Values to Public Transportation Access." AASHTO Journal, October

18, 2019. https://aashtojournal.org/2019/10/18/study-links-real-estate-values-to-public-transportation-access/.

Environmental and Energy Study Institute (EESI). “The Impact of Public Transportation on Real Estate Values.” EESI. Environmental and Energy Study Institute, March 21, 2013. https://www.eesi.org/briefings/view/the-impact-of-public-transportation-on-real-estate-values.

“Metrorail Stations.” Fairfax County GIS & Mapping Services Open Data Site. Fairfax County . Accessed May 6, 2023. https://data-fairfaxcountygis.opendata.arcgis.com/datasets/Fairfaxcountygis::metrorail-stations/explore?location=38.856465%2C-77.192750%2C12.17.

National Association of Realtors (NAR). “Public Transportation Boosts Property Values.” www.nar.realtor, June 16, 2014. https://www.nar.realtor/articles/public-transportation-boosts-property-values.

“Site Settings.” Fairfax County Virginia. Fairfax County , 2020. https://www.fairfaxcounty.gov/demographics/interactive-map-zip-code.

“Supervisor Districts.” Fairfax County GIS & Mapping Services Open Data Site. Farifax County . Accessed May 6, 2023. https://data-fairfaxcountygis.opendata.arcgis.com/datasets/Fairfaxcountygis::supervisor-districts/explore.

“ZIP Codes.” Fairfax County GIS & Mapping Services Open Data Site. FairFax County. Accessed May 6, 2023. https://data-fairfaxcountygis.opendata.arcgis.com/datasets/Fairfaxcountygis::zip-codes/explore?location=38.837081%2C-77.287400%2C10.89.