# Housing and transit infrastructure – an exploration of transit-oriented development in the Boston metro area

By Tom Hopper

Massachusetts has made significant investments in transit, particularly in commuter rail development. In order to maximize ridership and get the most use out of the commuter rail system, it is important to create dense housing opportunities around commuter rail stations. However, in many cities and towns across the Commonwealth, there is much resistance to multifamily housing, and in a number of towns with good commuter rail access, little or no housing has been built around these transit investments.

The metro Boston area is also incredibly expensive, and a large contributor to those high housing prices is the lack of supply and inability of our system to respond to increases in demand for housing. Part of this inability to produce sufficient housing is related to the way we make decisions about land use. Every single one of the 351 cities and towns across Massachusetts has independent zoning authority, which means they can allow (or prohibit) any kind of housing they wish. In practice, many communities choose to zone for low-density housing and limit the amount of dense, multifamily housing. This happens for a variety of reasons, but often stems from a desire to exclude certain people or populations from living in the community.

So, in many cases, additional housing development, and specifically denser multifamily housing development and affordable housing development, would provide suburban living opportunities that don’t currently exist. Not only would lower-income households be able to find a home in towns that often provide better opportunities, particularly in terms of schools, but it would also provide the kind of dense, walkable, transit-rich neighborhoods that both millennials and down-sizing baby boomers are growing increasingly fond of.

From a smart growth policy perspective, the new housing production that would help alleviate some of these supply and equity issues should occur in areas that meet sustainability and efficiency goals. However, there has been little work to try to measure how well the Commonwealth has made use of its transit infrastructure to promote denser, transit-oriented housing development. This project will aim to create a methodology for doing that.

## Research Question(s):

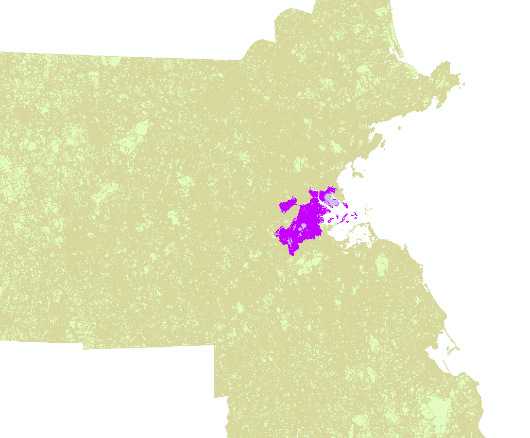
Has dense housing development accompanied transit investments in metro Boston? Where has transit-oriented development taken place, and where are opportunities for more work and development?

## Review of Data:

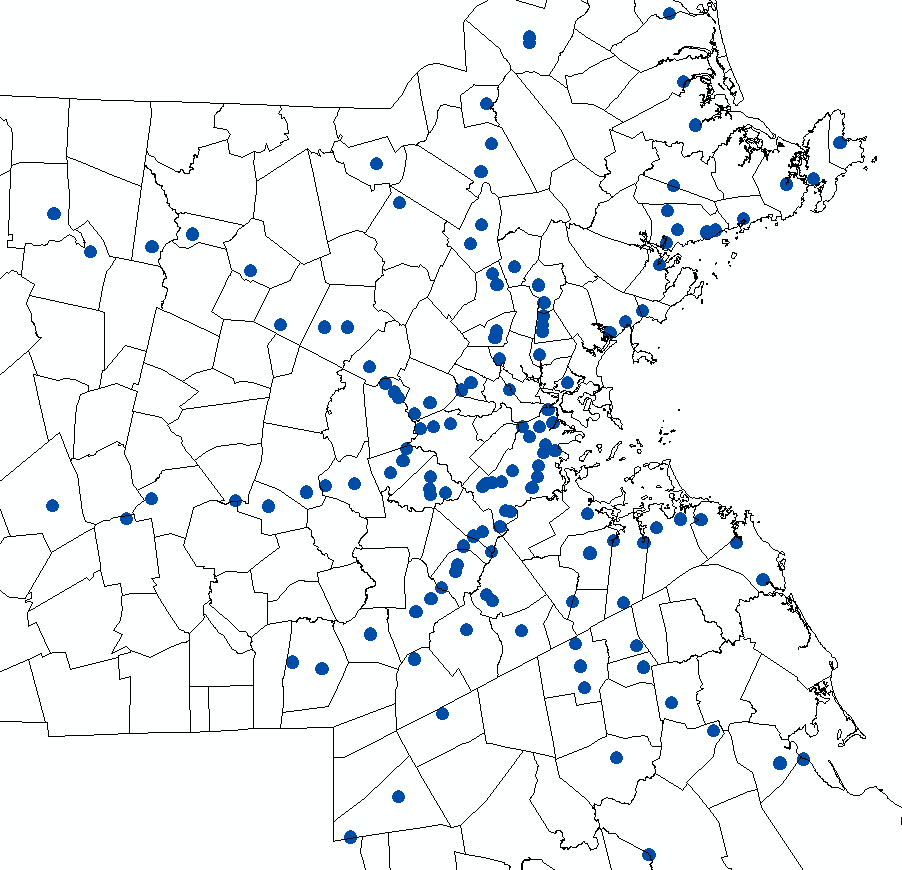
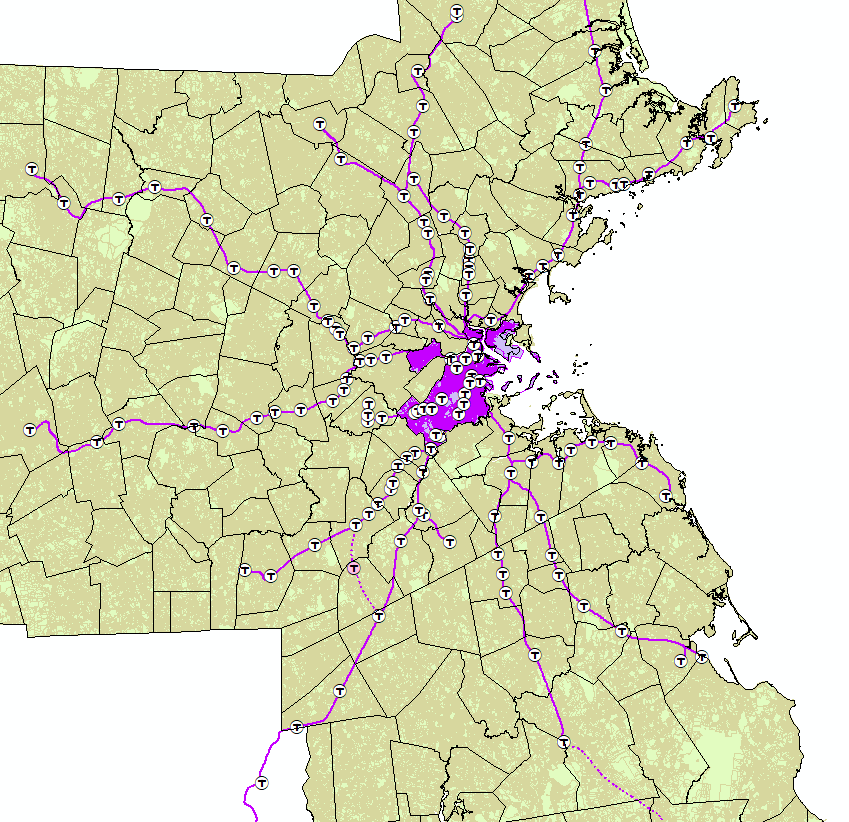
The data used for this project will include:

* Statewide tax assessors parcel maps, curated and distributed by MassGIS (except for city of Boston, which will need to be downloaded separately). These maps provide parcel-level information on use code, number of units, and year built, among other things. This will allow me to see what kinds of buildings, how many units, and when they were built. I will be focusing this level of analysis on the parcels that are within a half mile of a transit station. MassGIS has provided some standardization of land use codes across towns, because every town is able to use and modify whatever use codes they want. This standardization will allow me to use the same selection methodology across geographic boundaries.
* Commuter rail and MBTA layers, also maintained by MassGIS. These layers have line and point data, and I will be focusing on the point data (stations) to create the buffer used to determine station proximity.

## Methods/Walkthrough of GIS processes

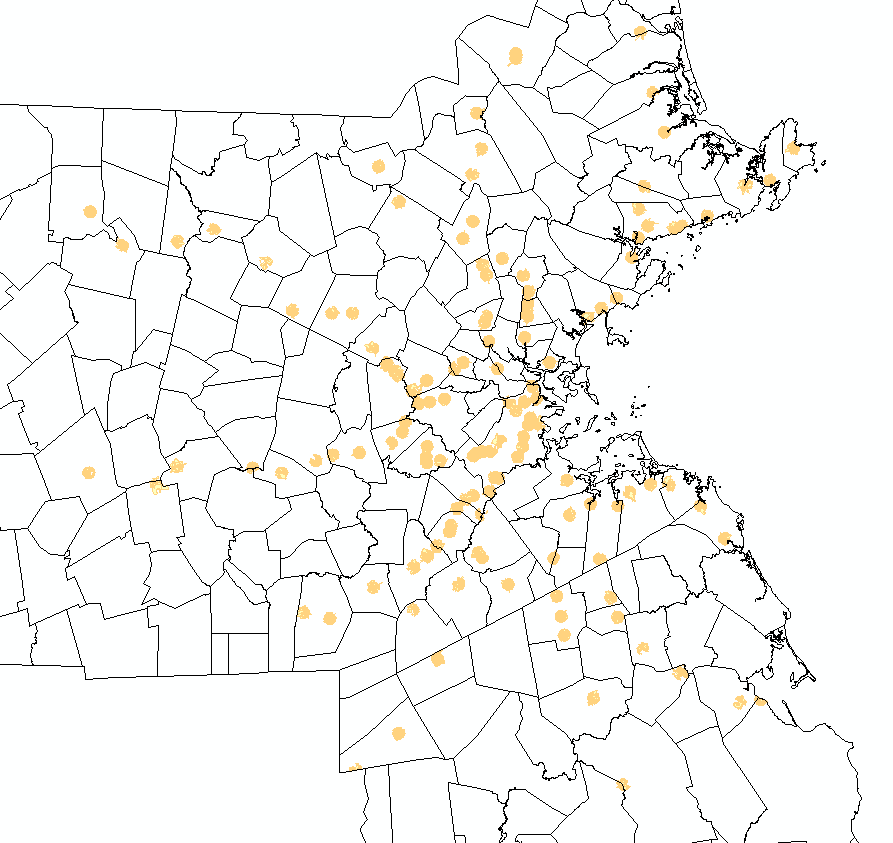


Boston parcels were obtained from the City of Boston, while the rest of the state is compiled and somewhat standardized by MassGIS. The following map shows how detailed these maps are. The data set is very, very large and takes a fair amount of time to render. One of the first steps in my process, therefore, is to make a selection of these parcels relative to the commuter rail stops, which will reduce considerably the number of parcels needed for the analysis.



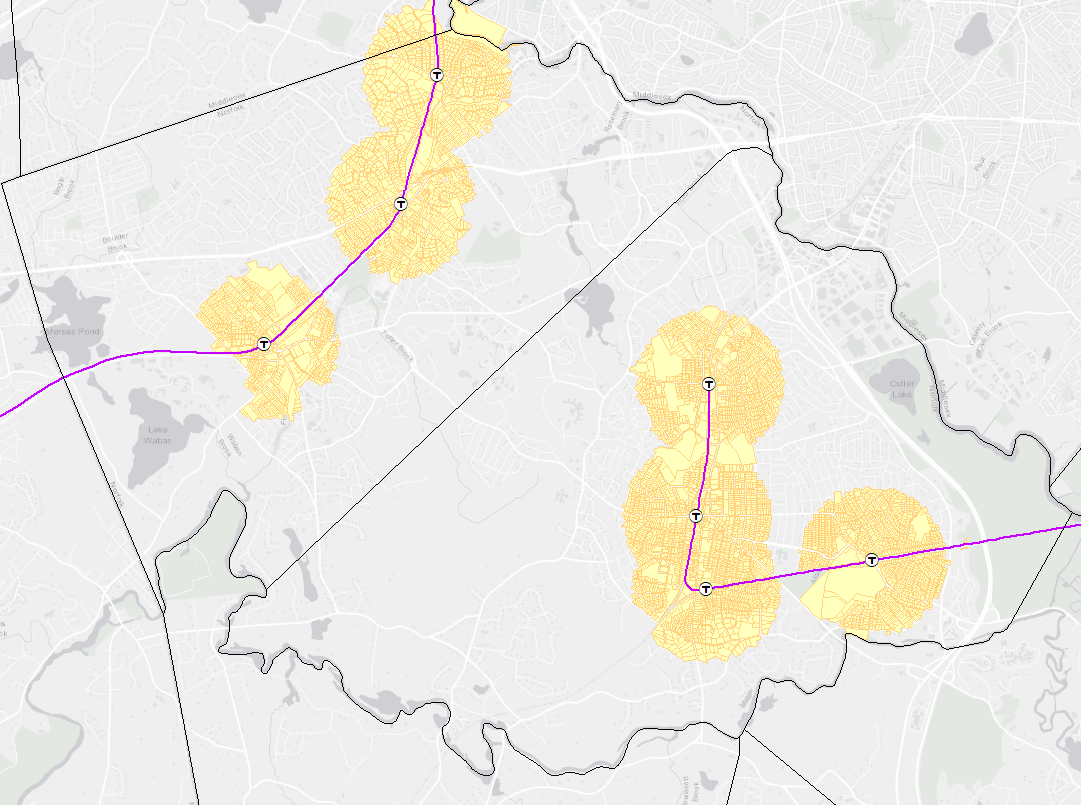
Here’s a look at the parcel maps layered with the commuter rail, along with municipal boundaries. For the purpose of this exercise, only the commuter rail will be analyzed. Future work should include major bus routes and subway, especially when analyzing transit-oriented development in the urban core and other parts of the state that rely on regional bus systems for mobility.

Using the buffer tool, a 0.25 mile buffer was created around each commuter rail stop in the system.

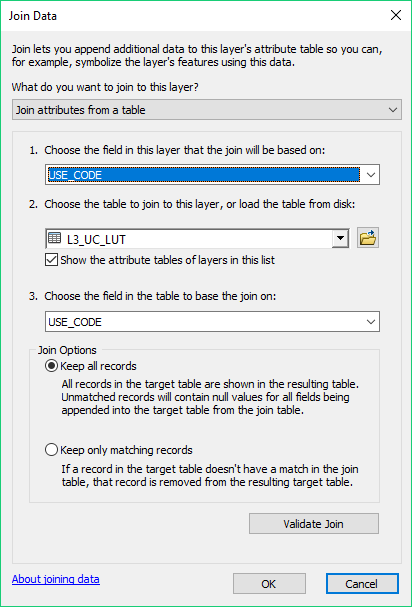


Then, using the select by location method, all parcels with centroids that fall within the buffer layer were selected, and a new layer was created that includes only those parcels. To note, I created buffers and selections on the Boston and non-Boston parcels separately, and then performed a Union to create a single layer from the two subsets. I decided to approach it this way since after the buffer there were far fewer parcels and it took less time to perform the union after creating this subset than it would have if I used the entire statewide parcel data set.

Here is a closer look at how those buffers look around each transit station, with a basemap to provide additional context. I did have to perform some cleanup on these buffers, since some of the parcels were roads that were not associated with real estate and kept the parcels from sticking within the buffer boundaries. After that cleanup, I was able to achieve pretty neat-looking Euclidean buffers.

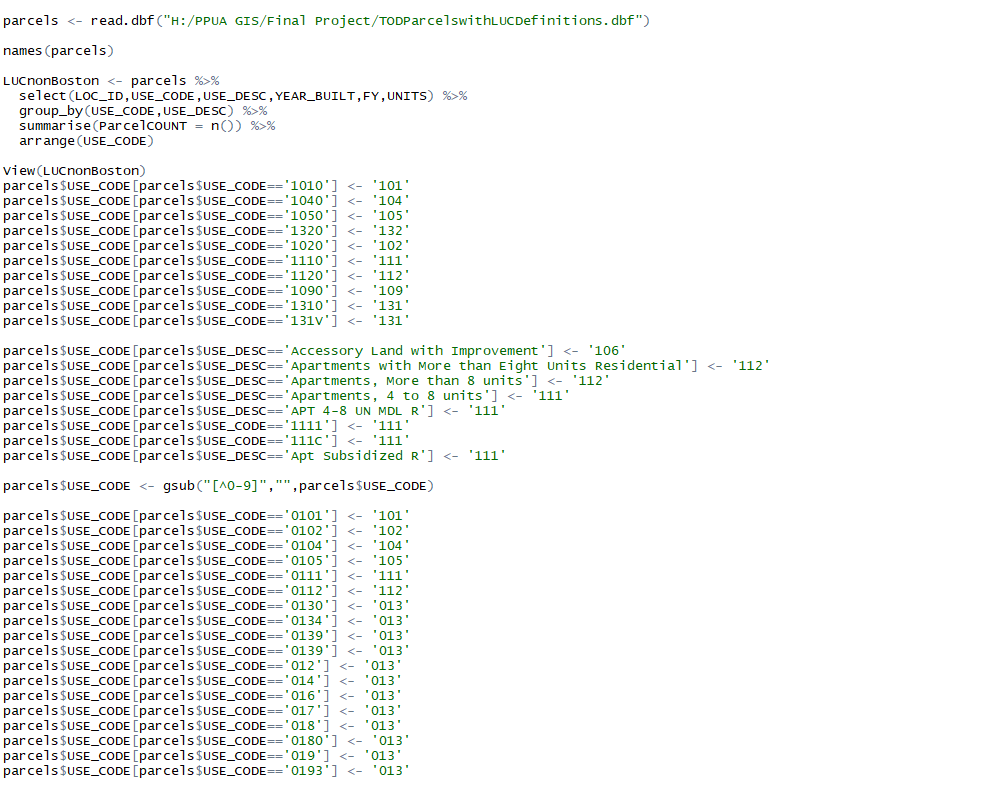


It was at this time that I realized the Boston parcels could not really be used for this project. The records lack a vital piece of information: number of units. So, going forward, I have opted to just work with the non-Boston parcels. In a way, this makes more sense anyway. Since the commuter rail is meant to serve the suburbs, and Boston is fairly well-covered by rapid transit subway lines, I don’t think this compromises the goals of this paper. The aim is to show that a regional investment in transit has not been met with a coordinated housing policy that makes the most of that investment. Since the majority of the stations are in suburbs that produce less housing than Boston, focusing on non-Boston cities and towns will help drive that point home.

With my transit-oriented parcels selected and saved as a shapefile, I was ready to start working with the underlying data. The first step was to join the polygon layer with the underlying data on number of units, year built, use codes, etc, which is stored in a separate dbf file (except for Boston, which had shapefiles with these features within the attribute table already). To do this, I performed a join between the shapefile and the detail table.

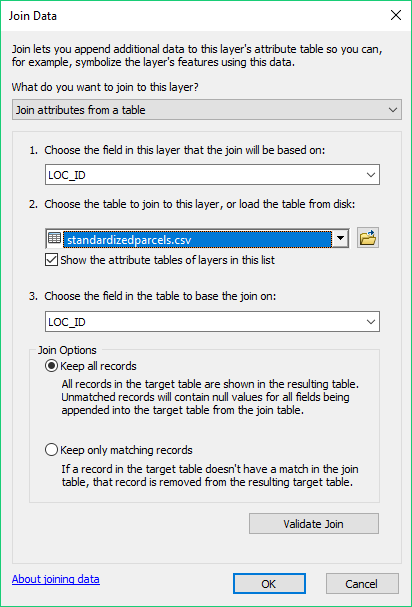
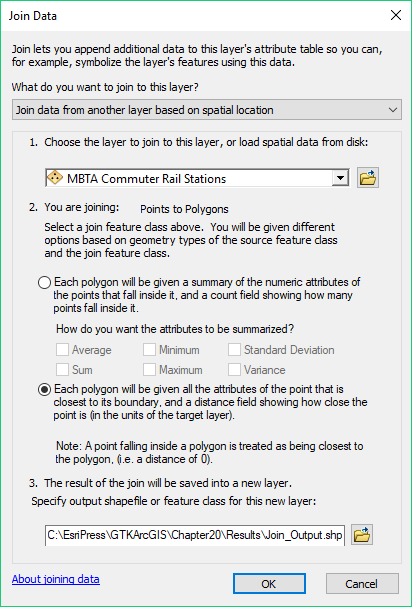
After the join was completed, I exported the attribute table for my transit-oriented parcels layer to a dbf file, so that I could explore and transform that data in more detail. The layer includes 154,679 individual parcels, so I chose to use R for my data analysis since it can fairly easily handle this volume of data.

One of the biggest challenges with this data set comes with assessors use codes. While there are industry standards for these use codes (codes starting with 1 are residential, 3’s are commercial, 4’s are industrial), there is still variation within those conventions from town to town. MassGIS has applied some standardization in its compilation of these disparate assessor’s records, but there are inconsistencies that do remain. So, it is essential to clean up the data a bit, or at least understand the variety of use codes that all reference the same kind of development. For example, if there are use codes 101, 1001, and 100.1, and they all signify a single family house, I need to document and associate those various codes with the same use.

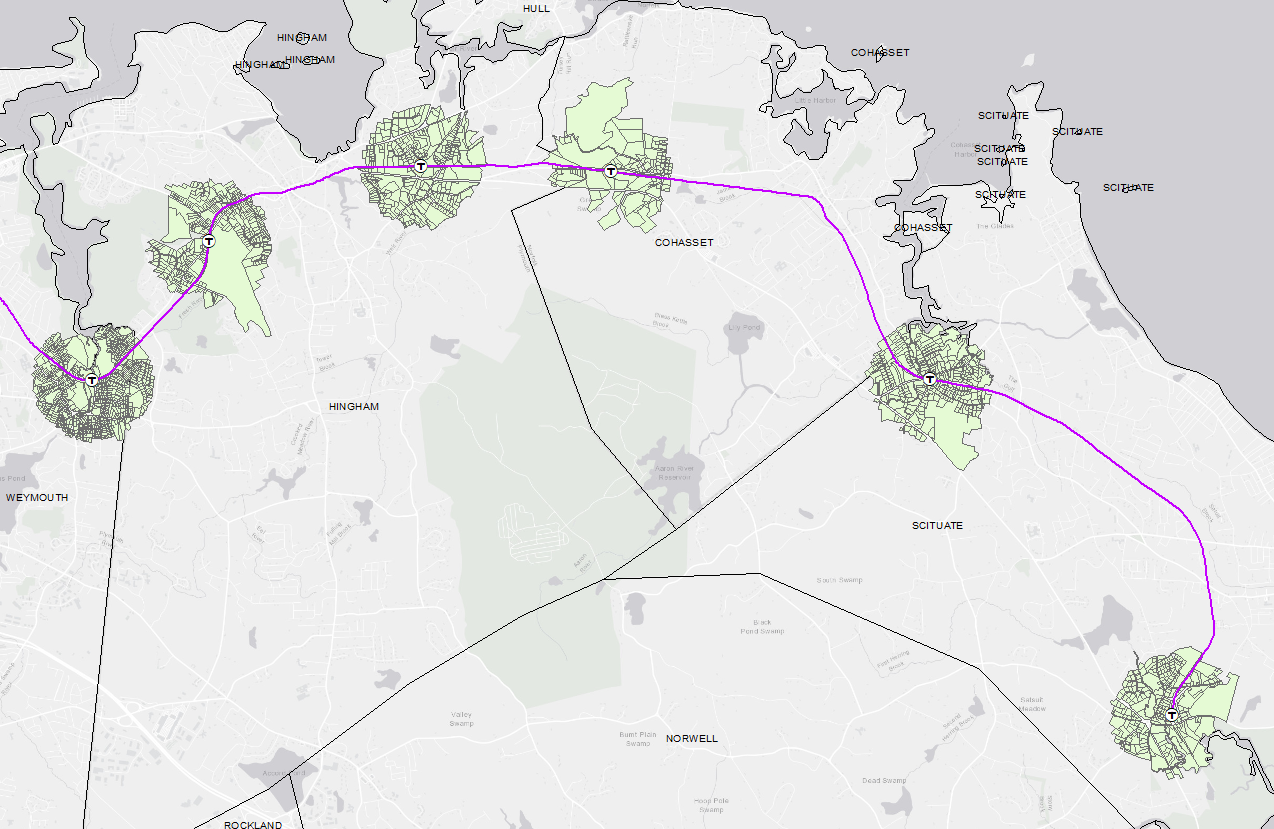
I created a script in R that converts all of the residential parcels with various use codes to a standardized set of residential use codes to make it easier to query and select attributes when I bring the data back into ArcMap (code snippet to the left). This was a bit of a painful process, as I had to find all of the residential parcels and figure out where the variations in codes were coming from. At this stage, I also did some imputation around missing values in the “UNITS” field. For example, where a single family home had a unit count of 0, I’ve changed that to a 1.

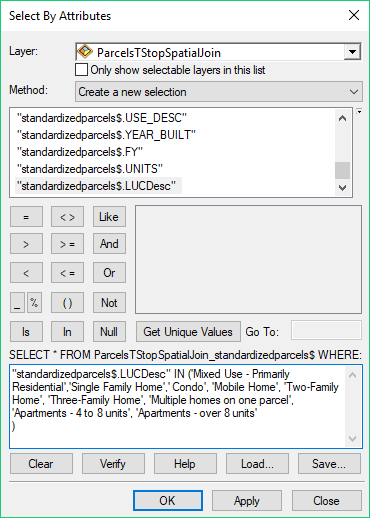
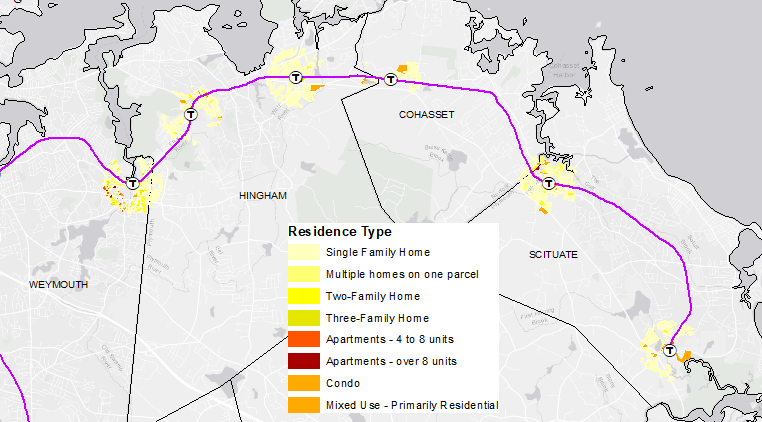
Once these conversions were complete, I wrote the resulting data into a new csv file so that I could upload it ArcMap.

I imported the standardized parcel information into ArcMap and performed a join of those clean parcel numbers to the original shapefile. I can now perform a selection by attributes using the new standardized use codes. I have also performed a spatial join, so that I can encode the parcels themselves with information about the commuter rail station they are closest to. This will allow for easy station-level parcel analysis.

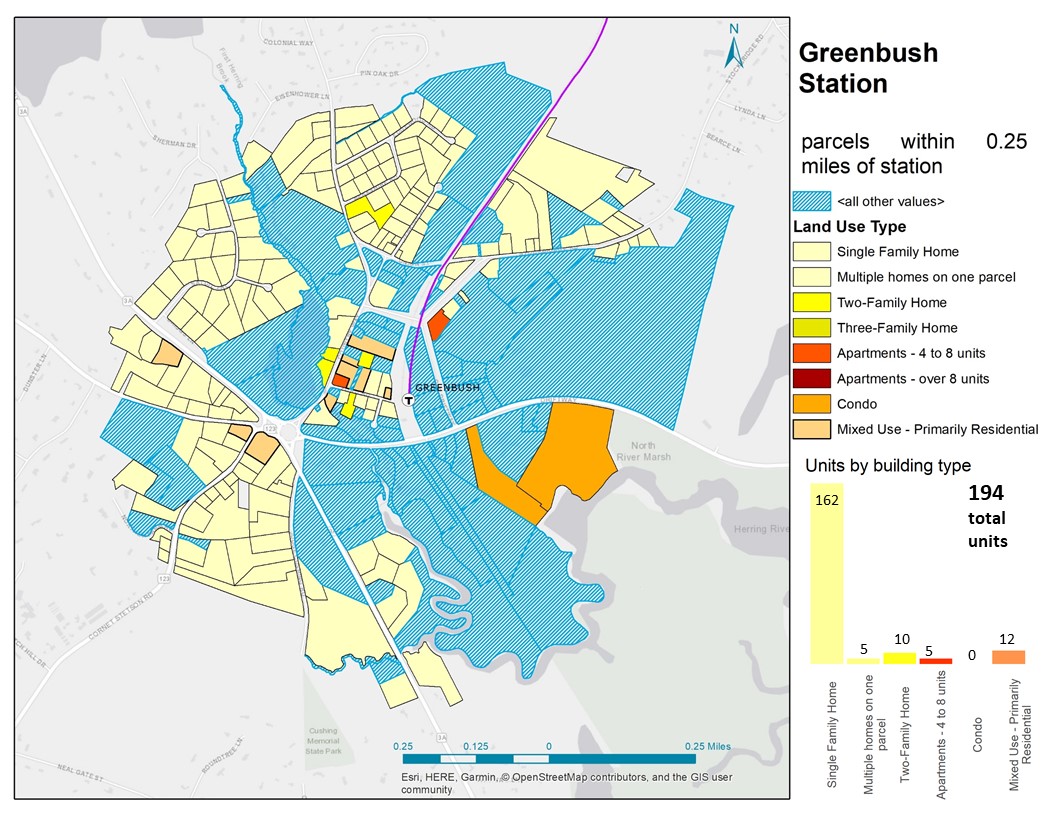
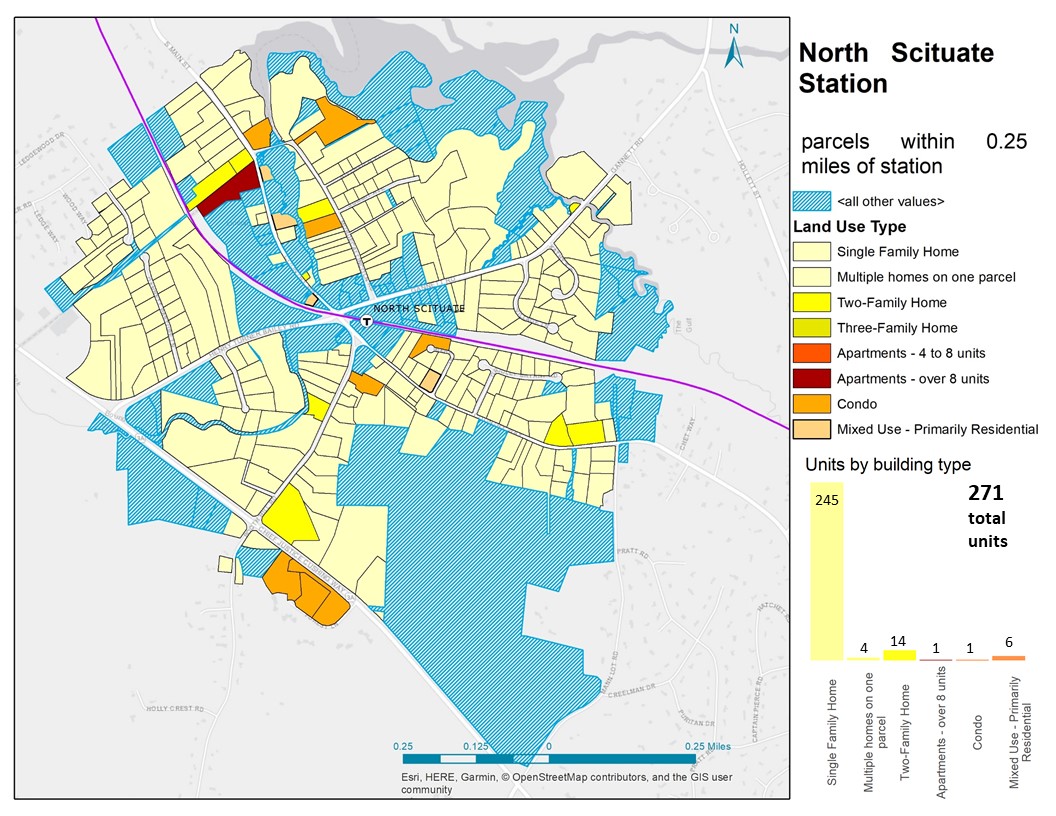
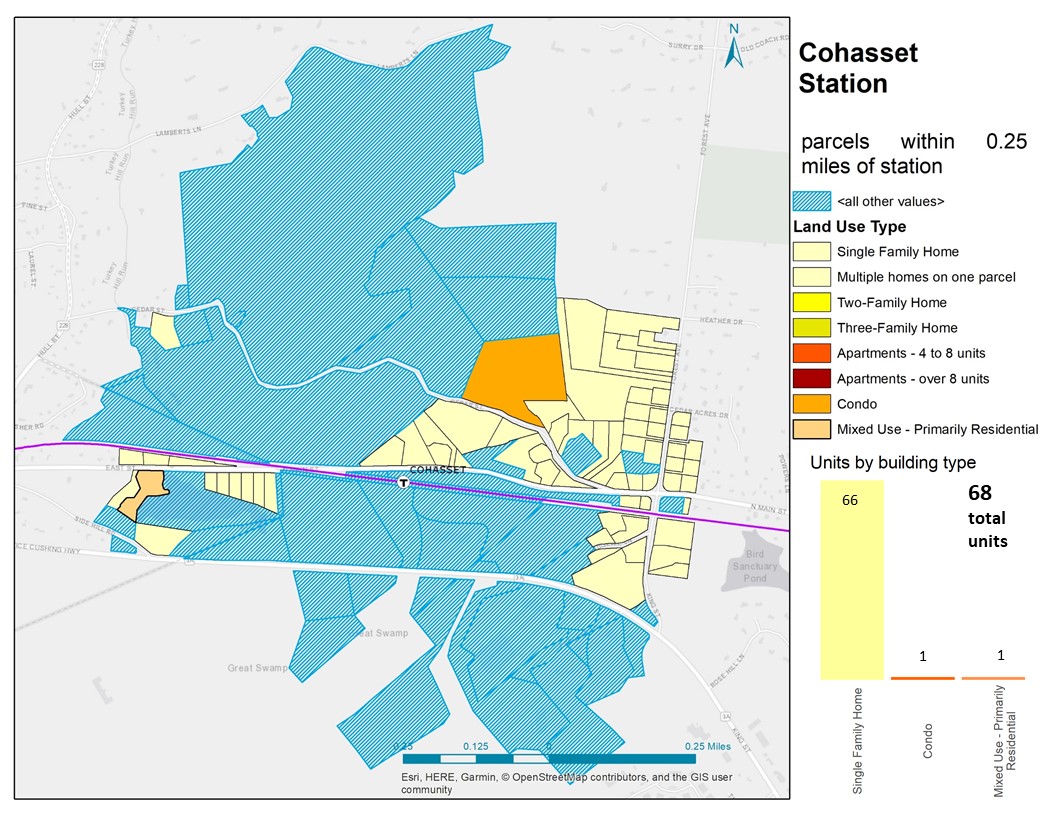
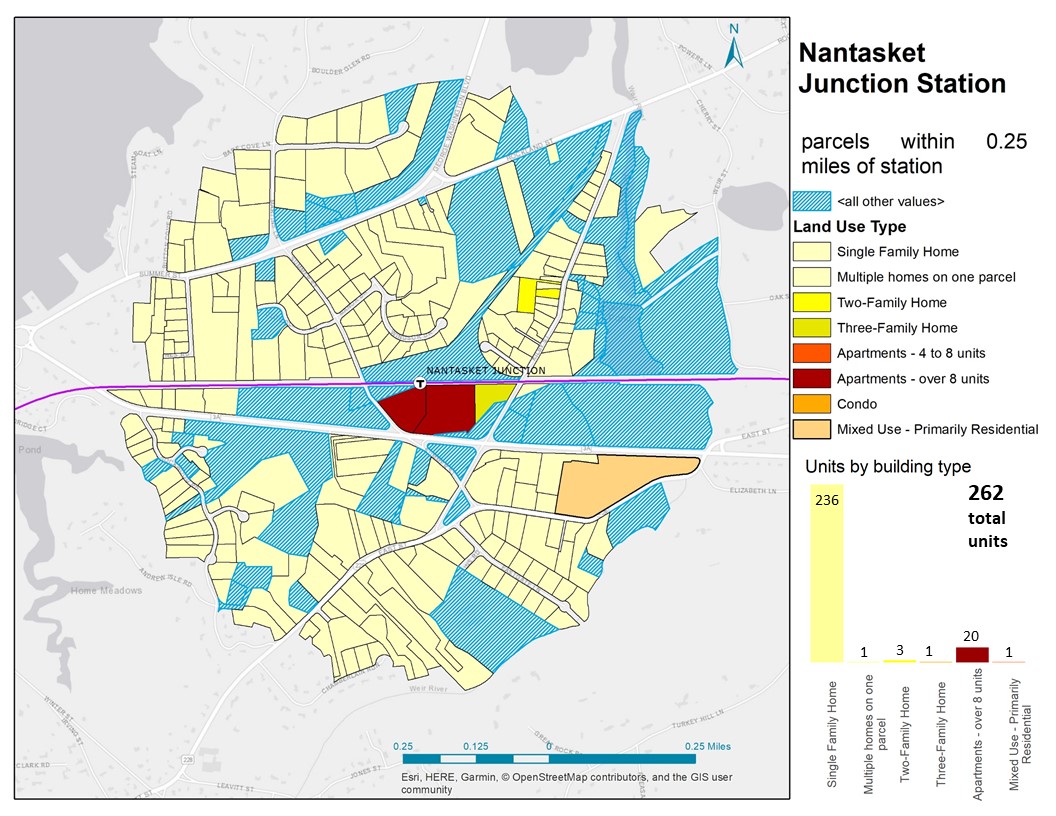
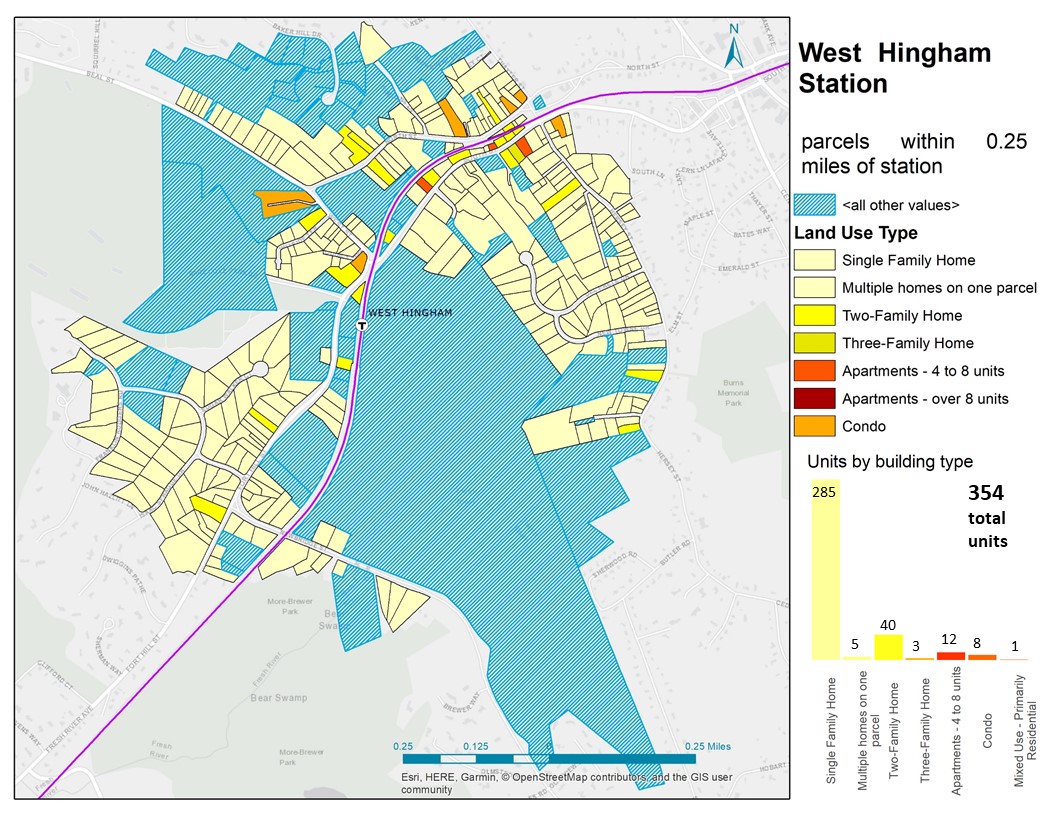
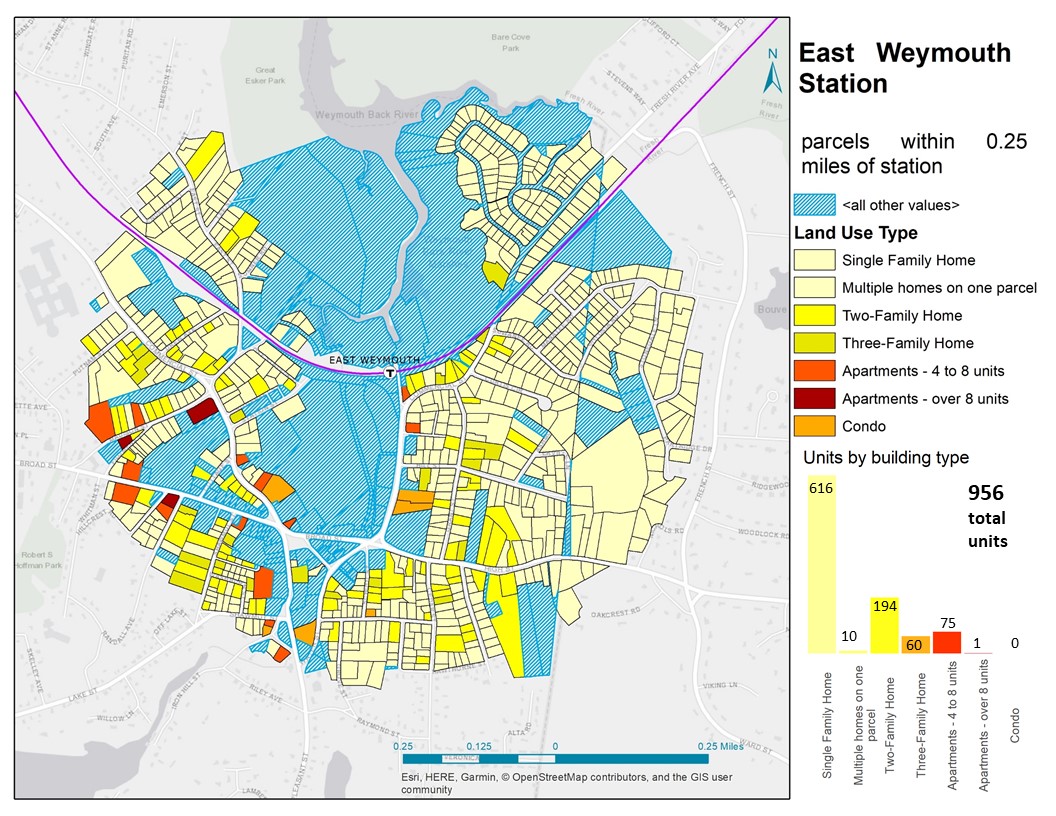
 

The resulting layer is ready to go. Here is a snapshot of a portion of the Greenbush line, which services the south shore region. I will be focusing on the six station areas displayed here to demonstrate the final stages of the analysis.

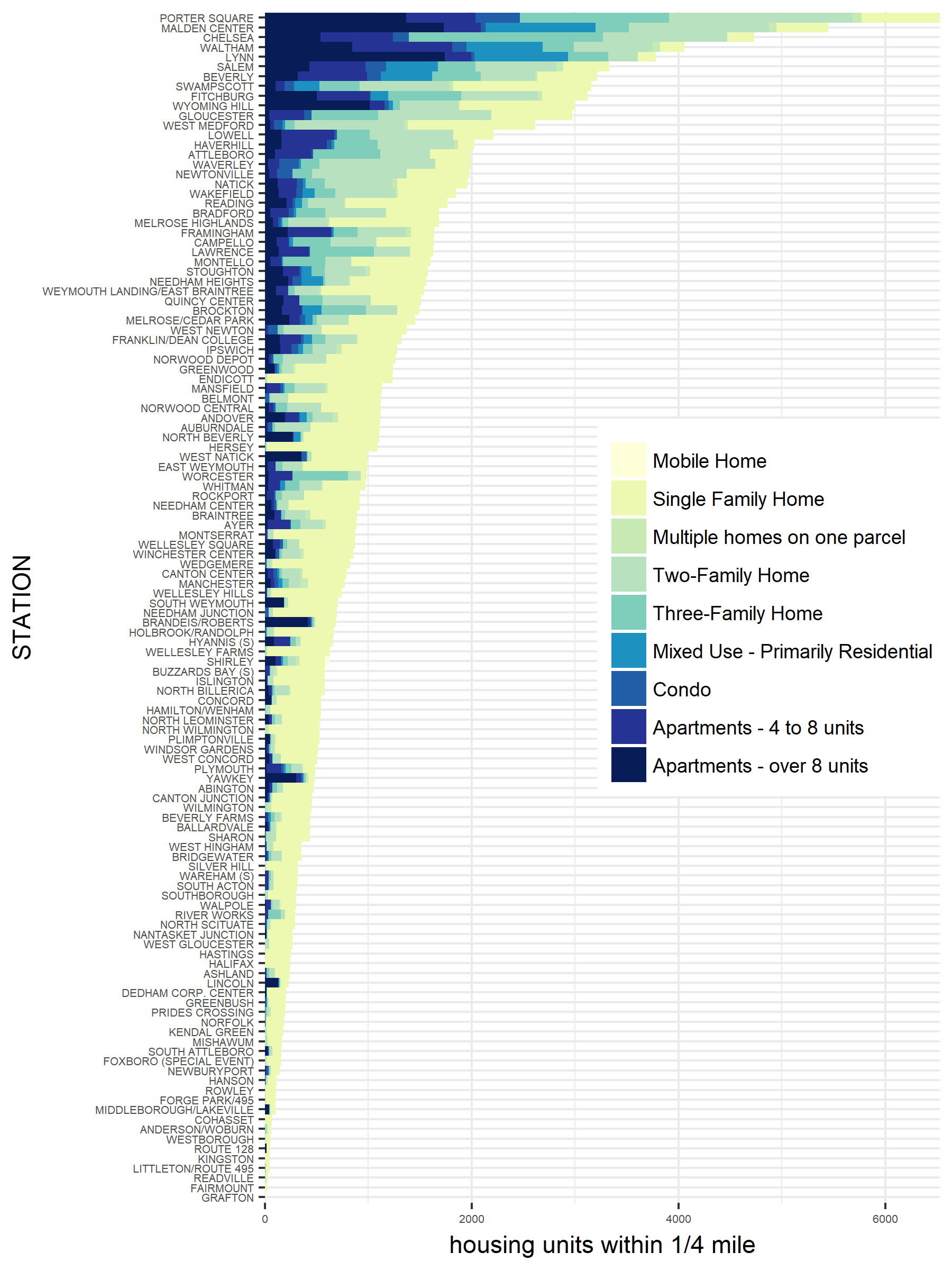


For my final visualizations, I want to be able to display only those parcels that are residential (as opposed to the full set of parcels, which includes commercial, industrial and other kinds of properties). To do this, I am relying on the use code descriptions I set when I was cleaning up the data.

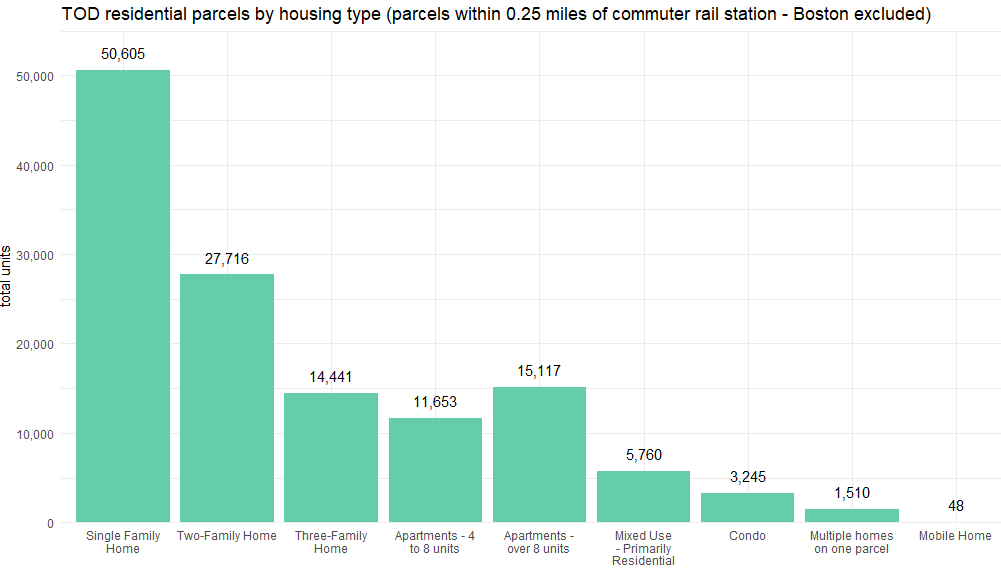
Now to make some layouts that show the results of the analysis and allow for comparisons between stations. Each of the following maps represents a single station area and visualizes the land use codes for the residential parcels. I have also incorporated a simple chart showing the number of housing units by the type of building they are in.



With all of the stations appropriately filtered and recoded, and because I have used a spatial join to associate parcels with their nearest station, a visualization can be made that directly compares number of units by station, with a breakdown by building type. You can immediately see that there is a wide variation in development densities and building types across the commuter rail system.



A count across the entire system is now possible, too. Below is a chart summarizing the total units across the entire non-Boston commuter rail system, broken out by building type. It is immediately apparent that single family homes predominate, and that more dense building types are in much shorter supply in these transit-oriented locations.

****

## Conclusions and future research

It is clear that there is significant variation in the built environment across station areas. This is particularly evident when comparing the area surrounding the Cohasset station with the area surrounding the East Weymouth station. While neither station has an abundance of multifamily housing, which would be the best use of a transit-oriented location, there are significantly more housing units near East Weymouth than there are surrounding Cohasset, which has very little space at all dedicated to housing.

It is also clear that the neighborhoods surrounding many stations have dedicated much of that land to single family housing. This seems to be an enormously inefficient use of these transit-rich locations. It would benefit the commuter rail system as well as the state if these transit-rich locations were to develop housing at a higher density and volume. Multifamily housing, and particularly multifamily rental housing, is in short supply in the suburbs, which can have the effect of excluding lower-income households and those seeking dense, walkable locations. By creating more diverse places in the suburbs around transit infrastructure, the Commonwealth could reduce spatial segregation and create town centers that alleviate some of the pressures in the housing market.

This analysis could provide an initial step in analyzing how the state plans for housing in relation to existing transit infrastructure. The spatial representations made in this paper showing the underutilization of transit-adjacent parcels have the potential to make this lack of coordinated planning apparent. When applied to the whole system, the methodology used here highlights the weak points in the system, and where the state could concentrate efforts to broaden housing opportunities in relation to existing commuter rail investments.

## Data sources:

MassGIS, Standardized “Level 3” Assessors’ Parcels. <https://docs.digital.mass.gov/dataset/massgis-data-standardized-assessors-parcels?_ga=2.163652726.894669162.1534181086-154496739.1511879849>

MassGIS, Trains (and MBTA Commuter Rail). <https://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/trains.html>