Sarah Serpas
GIS for Population Science, Spring 2021

Replicating Development/Land Use Typologies for San Antonio/New Braunfels Metro Area

Overview: Creating typologies of geographies can be useful lenses to view other demographic factors and policy implications. Urban planners may primarily examine development patterns on a parcel or census tract level. However, by combining several datasets and completing a k-means cluster analysis, researchers have demonstrated the usefulness of creating a place typology to examine larger development, land use, employment, and housing patterns in a region. Through this project, I propose to replicate a process of creating physical development typologies that was previously used for the NY-NJ-CT Region and the San Francisco Bay Area Region (Montemayor and Calvin, 2015; SPUR 2020). I propose replicating this methodology for the San Antonio-New Braunfels Region, with the possibility of extending to other Metropolitan Areas in the State of Texas. This can be a useful lens to examine, holistically, the types of physical and human environment within a region, and allow policymakers to determine tailored policies for the different physical place types. Further, once the types are created, additional demographic factors can be examined through the lens of the place types.

Geography: For both the New York and Bay Area analyses, a ½ mile by ½ mile grid was chosen as the unit of analysis, effectively rasterizing the regions and requiring all data to be cross-walked into the geography of the grid. The New York Region is 13 thousand square miles with 23 million residents, and the Bay Area Region is roughly 6.9 thousand square miles with 7.1 million residents. The San Antonio New Braunfels Region is 7.4 thousand square miles with a population of 2.4 million. The ½ mile by ½ mile grid should also work for the San Antonio Region, but there may be fewer places of intense density as there are in the Bay Area and NY Region examples. This could lead to a need to revisit the geography of analysis, but an initial analysis at this geography will be completed.

As mentioned above, all data will be crosswalked from their initial geographies (census blocks, parcels, etc) into the ½ mile by ½ mile grid. This will be done based on the rough proportionality of each geography within the grid (i.e. if a block of 100 people is split 40% in one grid and 60% in another, the population will be assigned 40 and 60, respectively). This, admittedly, might oversimplify where the population actually lives within the block, but will still give an approximation that will be useful at the regional scale.

Methods: A k-means cluster analysis will be used to ascertain different “place types” based on the following five factors:

(1) <u>Residential Density:</u> the number of housing units in an area gives an essential datapoint about the development intensity of the area. This does not take into

account the number of people per unit, but does give a valuable view of the residential development intensity in an area.

(2) Employment Density: measuring the number of jobs in an area can determine whether it is a major commercial area or job center.
(3) Intersection Density: the number of road segments that intersect in an area can be a useful proxy for walkability and how accessible and area is to surrounding areas
(4) Land Use Entropy: using an equation described in Montemayor and Calvin, this measure assess how homogenous or heterogenous the land use is within a grid cell, which can be a measure of the "mixed use" nature of the area.
(5) Impervious Surface Cover: to assess developed vs. undeveloped land

Data: The sources and geographies of the data needed for the analysis are listed below

(1) Residential density - block-level, decennial census or block group-level ACS if necessary
(2) Employment density - block-level, LEHD LODES
(3) Intersection Density/Walkability - point-level, approximated by intersection density from TIGER Line
(4) Land use Entropy - parcel-level data with state land use codes from TNRIS Strat Map
(5) Impervious surface Cover - USGS NLCD

Research Questions/Goals:

This k-means cluster analysis will be used to examine the development patterns within the San Antonio-New Braunfels Metropolitan Area, answering the central questions:

- What development types comprise the region?
- How well does the k-means cluster analysis represent the development types within the region? (ground truthing with aerial images and local knowledge)
- Other studies have factored in overlay analysis to more accurately represent the place types, what other factors are needed to more accurately represent the region?

Overall, this analysis will be useful for policy implications and a deeper understanding of the physical development of the San Antonio-New Braunfels Region.

References:

Montemayor, Lucrecia and Calvin, Ellis. Regional Plan Association. 2015. "Identification and Classification of Urban Development Place Types for the New York Metropolitan Region" https://www.researchgate.net/publication/305880202_Identification_and_Classification_of_Urban_Development_Place_Types_for_the_New_York_Metropolitan_Region

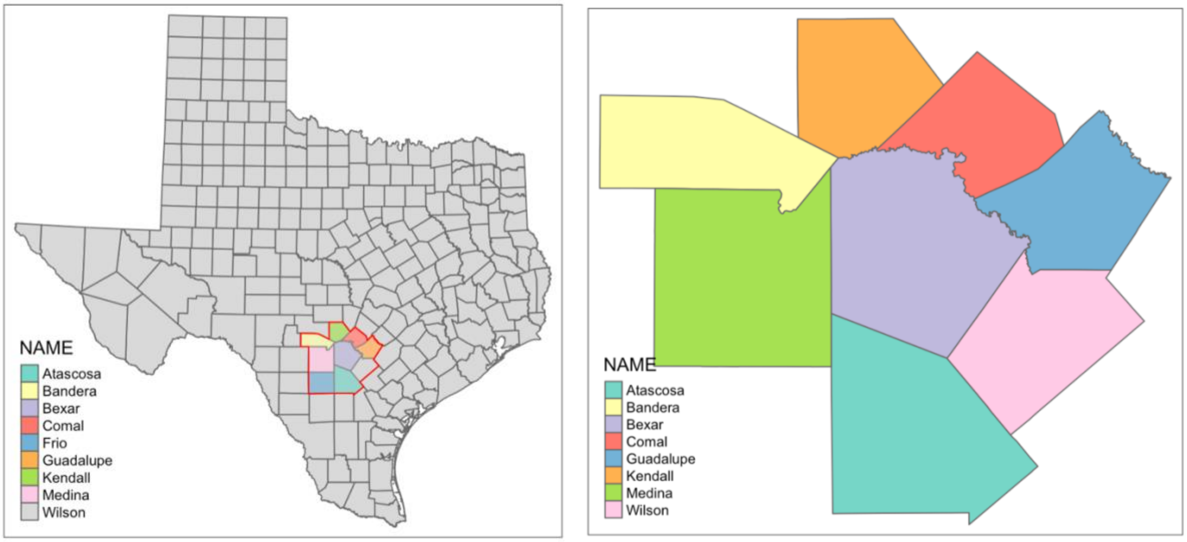
SPUR, Model Places, 2020. https://www.spur.org/sites/default/files/2020-09/spur_aecom_model_places.pdf

RESULTS

Initially, I had hoped to do this same analysis for the four largest metro areas in Texas, but some limitations on computer processing has shifted my focus to only to the San Antonio area.

The San Antonio-New Braunfels Core Based Statistical area, is comprised of 5 counties, shown in the graphics below:

Figure 1: Context and Geography of Interest (Source: TIGRIS)



The analysis will include 4 main data points, that are being crosswalked into a ½ mile by ½ mile grid. The grid is shown in Figure 2 below, with an example aerial view for context of the size in Figure 3. This grid was created in QGIS with Vector Creation > Create Grid.

Figure 2: ½ Mile by ½ Mile Grid

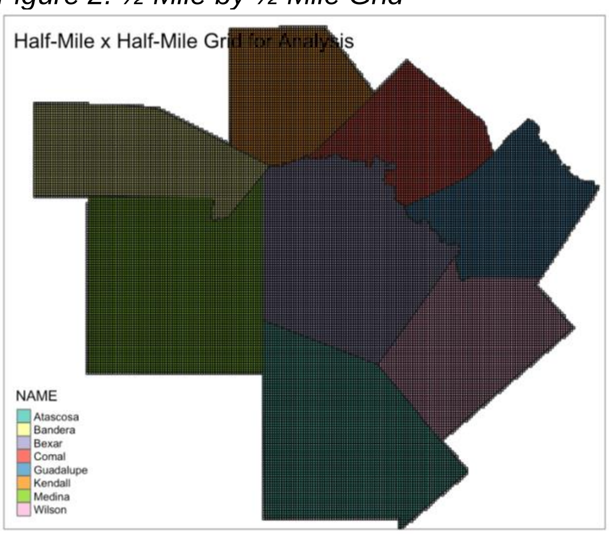
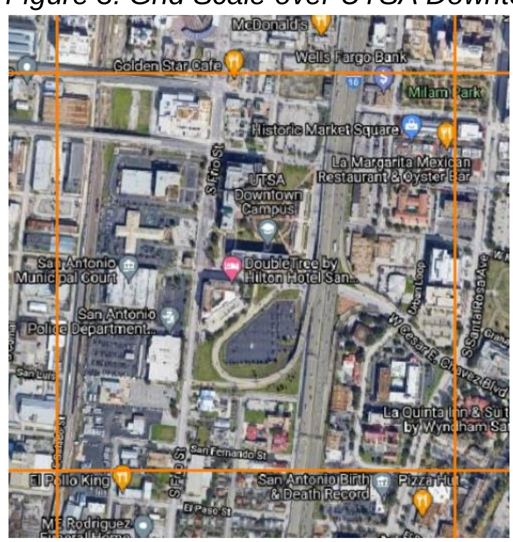


Figure 3: Grid Scale over UTSA Downtown

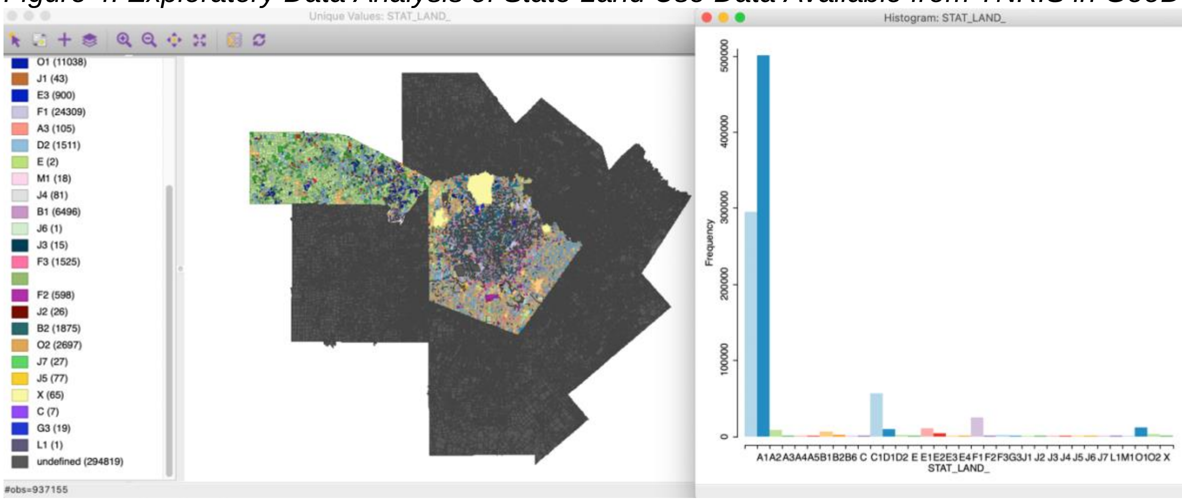


In the example projects from New York and the Bay Area, 5 variables were crosswalked into the grid:

1. Developed Area/Impervious Surface
2. Intersection Density
3. Residential Density
4. Employment Density
5. Land Use Entropy

Through exploratory data analysis, issues were discovered in the Land Use Data available from TNRIS. Exploratory Data Analysis shown in Figure 4 revealed that while there is land use data available for all 8 counties in the San Antonio-New Braunfels MSA, only Kendall and Bexar counties have detailed land use as definted by the State Land Use categories. This was explored with GeoDa, shown below. All of the parcels in grey are solely marked as "undefined." For this reason, this study will not include land use entropy.

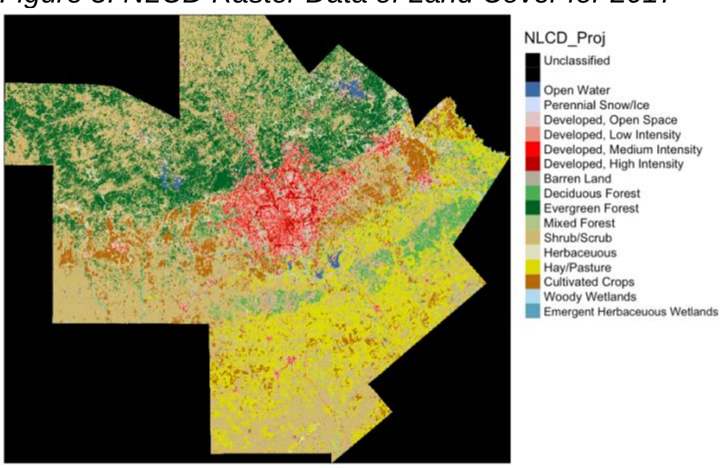
Figure 4: Exploratory Data Analysis of State Land Use Data Available from TNRIS in GeoDa



The other 4 measures will be included, and maps for each are presented below.

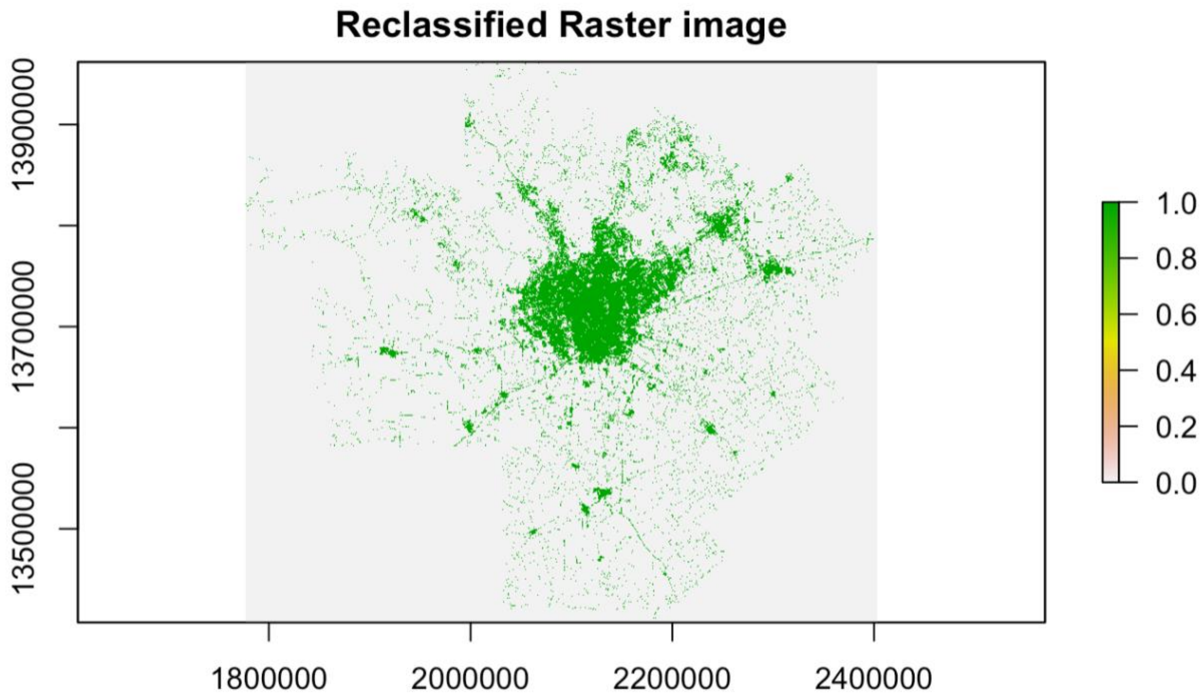
Developed Land is gathered from the NLCD raster data, shown in Figure 5 for the San Antonio-New Braunfels CBSA.

Figure 5: NLCD Raster Data of Land Cover for 2017

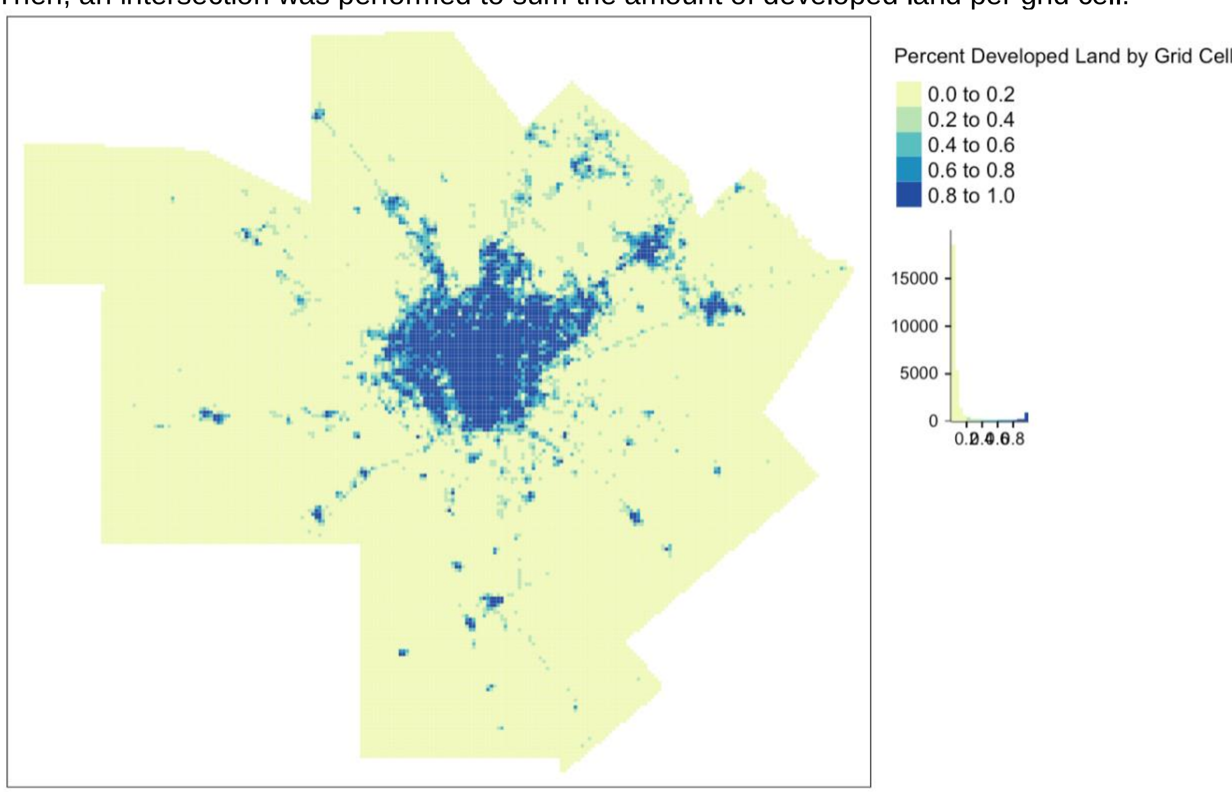


This data reclassified similar to a lab assignment from class, using a simple binary of Developed versus Undeveloped. (4 codes in red in the above raster). This reclassified raster is shown below.

Figure 6: Developed Land

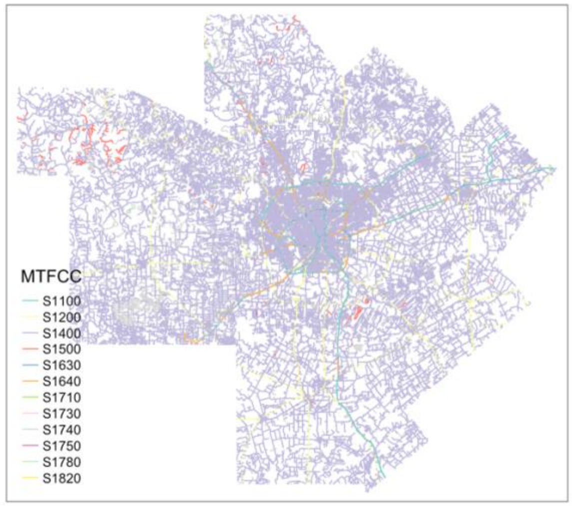


Then, an intersection was performed to sum the amount of developed land per grid cell:

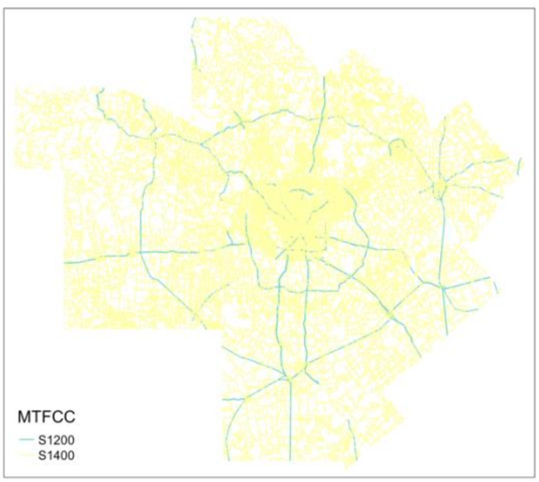


Intersection Density: Using TIGER Line data, the amount of intersections were calculated in each grid cell. This measure is a proxy for walkability.

First, all of the streets in the CBSA were downloaded. All of these roads are shown in black on the left, and symbolized by "functional class" on the right:



The streets were filtered to only select out the functional classes that people might walk on (this removes major freeways and ramps). More about functional classes here: https://www2.census.gov/geo/pdfs/reference/mtfccs2019.pdf



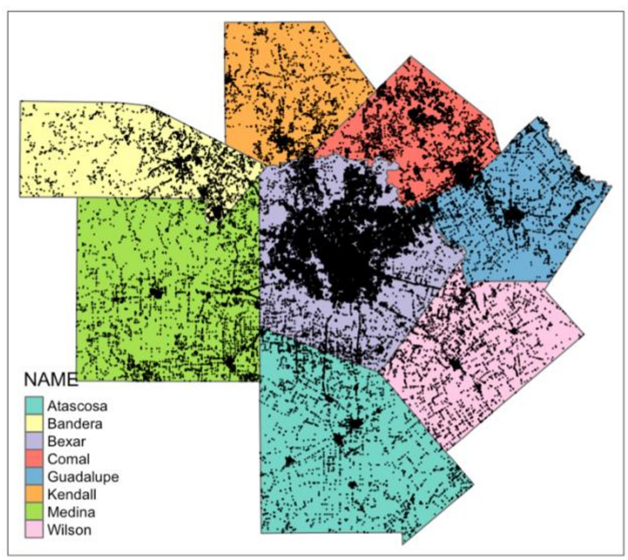
Here we are selecting two classes, shown in the map to the left:

- S1200 Secondary Road (usually has at grade crossings)
- S1400 Local Neighborhood Road, Rural Road, City Street

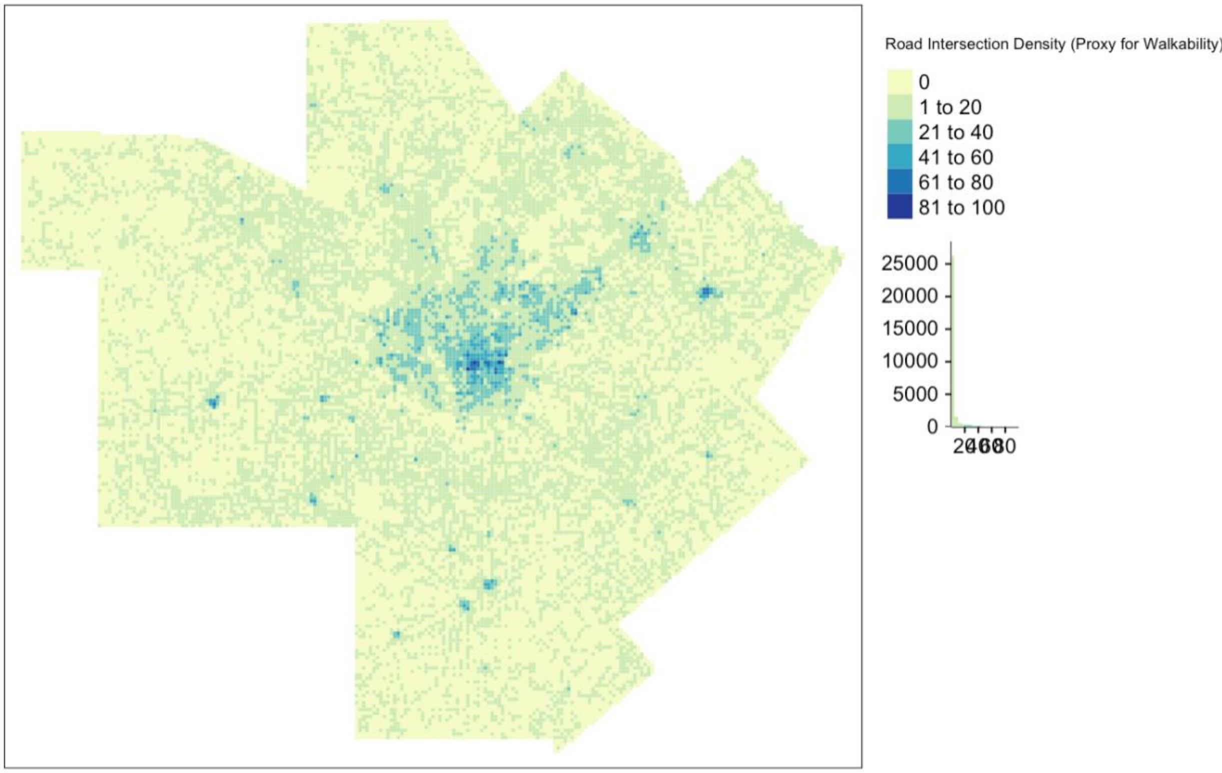
After filtering out the roads, I couldn't do the following QGIS processes in R (not for lack of trying). But in order to select where the streets intersect, I followed these steps:

1. Dissolve all roads on "Name"
2. Multipart to single part
3. Use "Line Intersection" tool with the same shapefile twice
4. Remove duplicate points

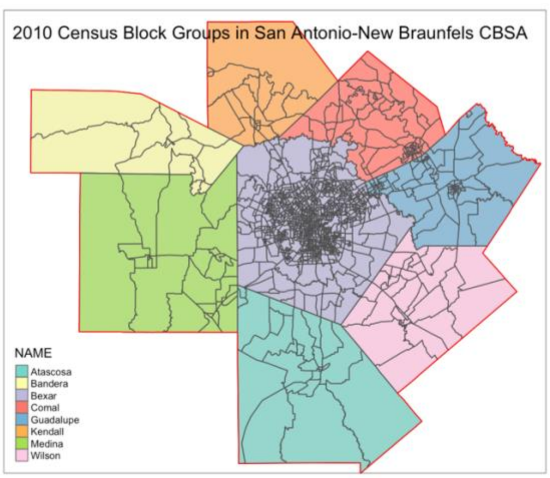
This resulted in the map of intersections to the right.

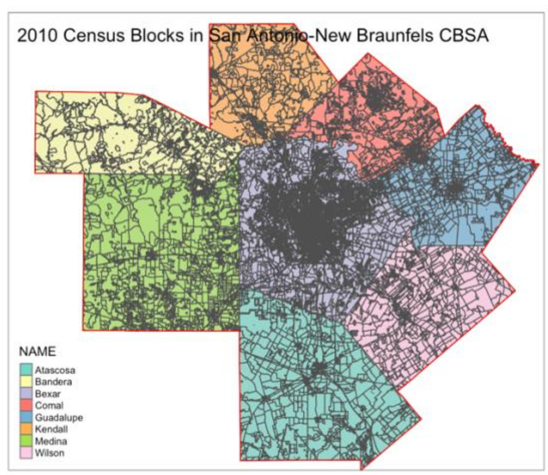


Finally, all of the intersection points were calculated in each grid cell, resulting in the below map:

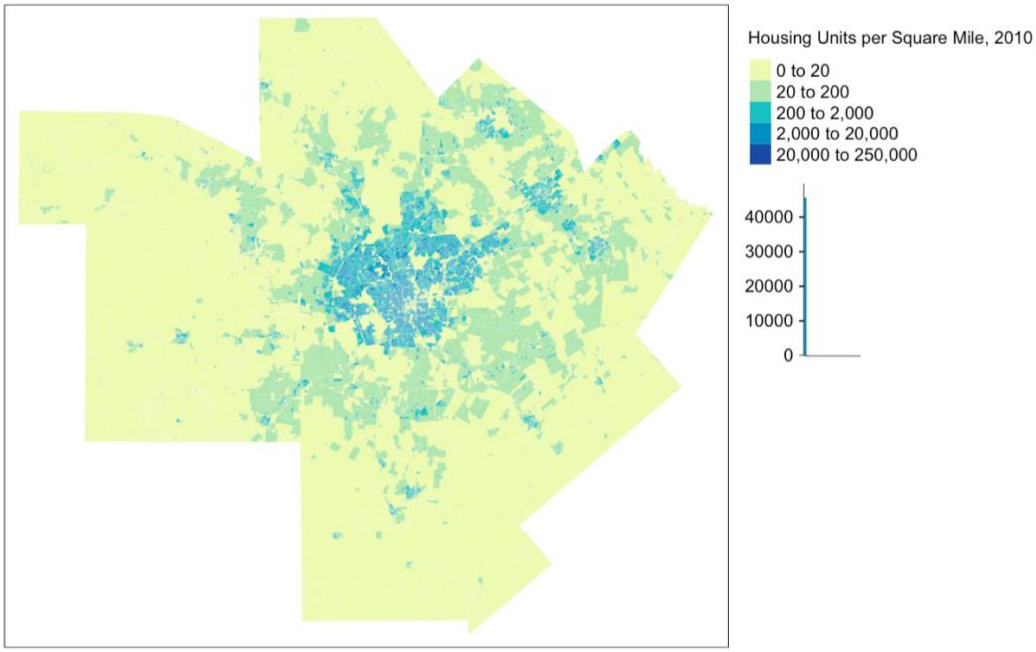


Residential Density: Calculated the housing units per grid cell using Census 2010 Data at the Block Level, and 2019 5 Year ACS Data at the Block Group Level. The census geographies are shown below.

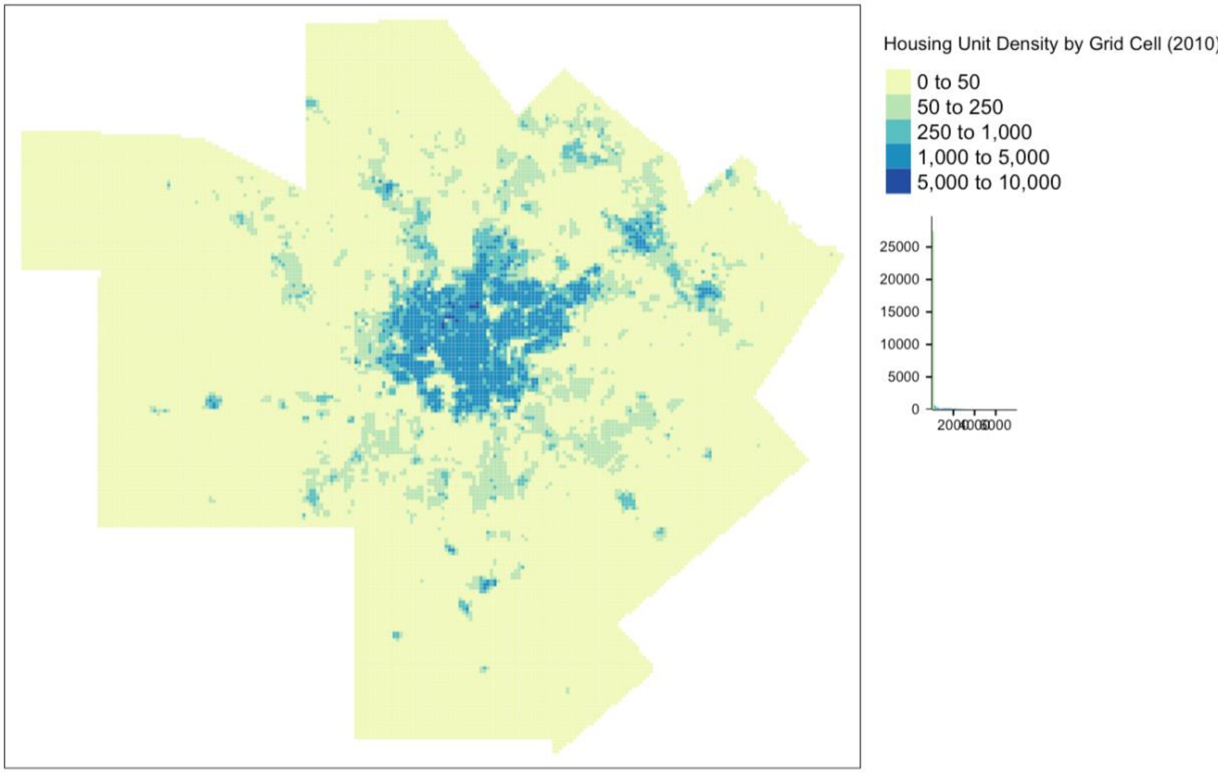




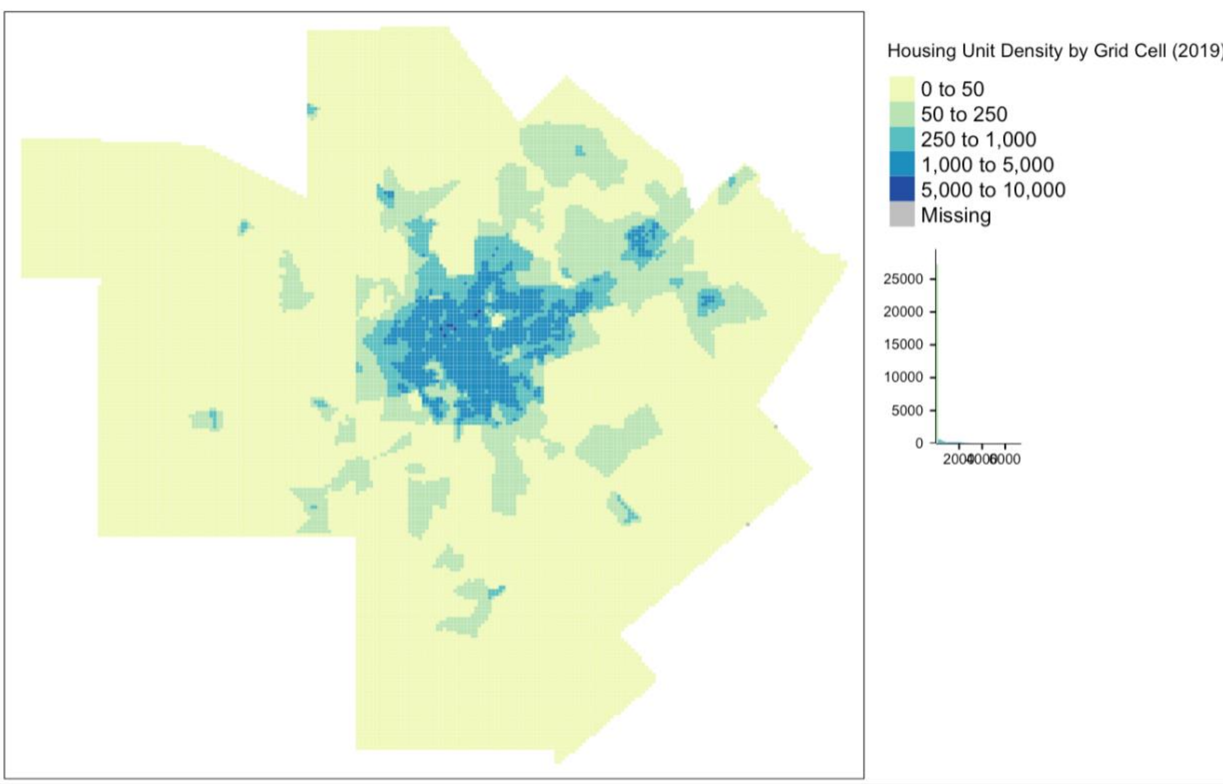
And the values for those geographies (2010 Blocks top, 2019 Block Groups Lower):



Which was translated to the Grid (through a similar process to the apportionment lab in class).
For 2010:

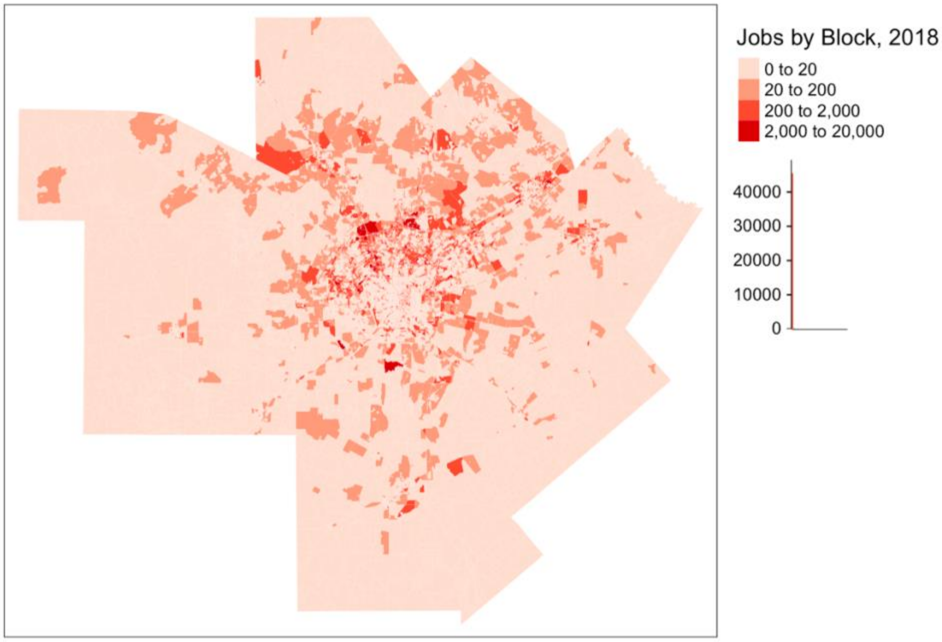

Housing Unit Density by Grid Cell (2010)
0 to 50
50 to 250
250 to 1,000
1,000 to 5,000
5,000 to 10,000
25000
20000
15000
10000
5000
0

And 2019:

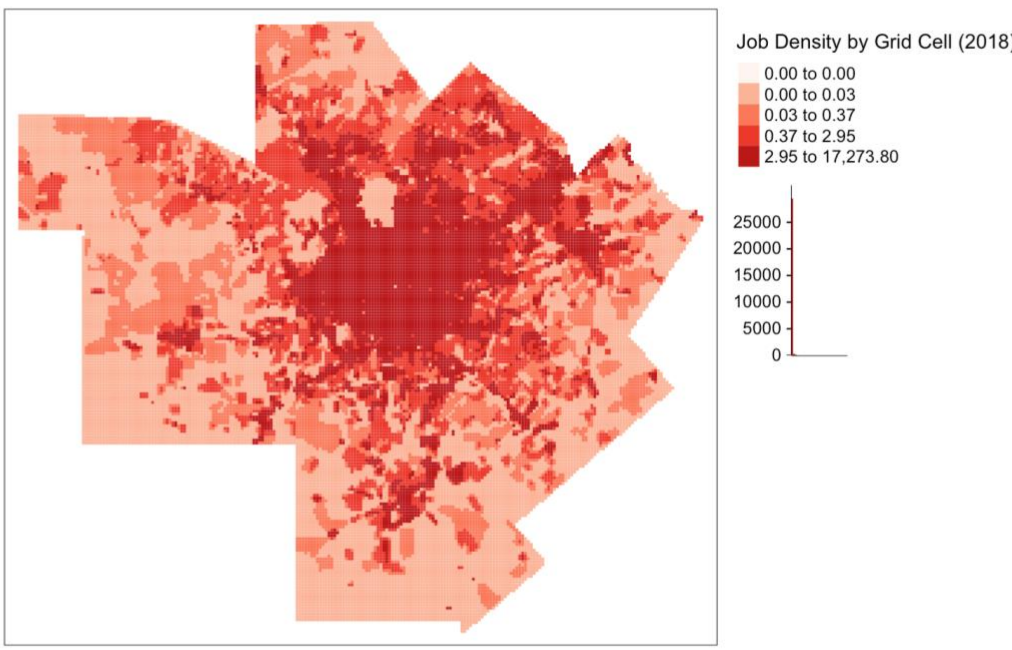

Housing Unit Density by Grid Cell (2019)
0 to 50
50 to 250
250 to 1,000
1,000 to 5,000
5,000 to 10,000
Missing
25000
20000
15000
10000
5000
0

Employment Density: Employment density was downloaded from LEHD for 2018.

First the Jobs data at the block level:



And then at the grid level:

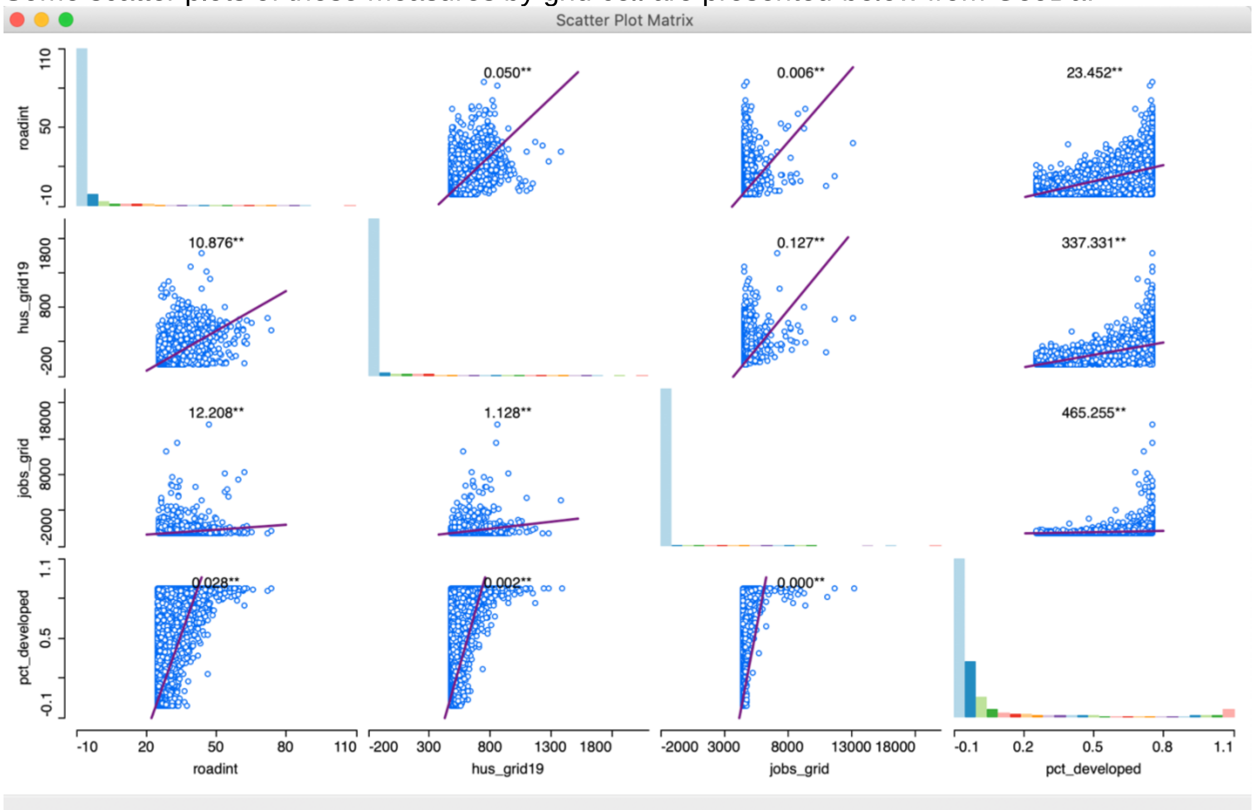


With that, I was then able to rerun the typologies through the KMeans Cluster analysis. Firstly, I put the data into GeoDa and examined some of the plots of the variables of interest. I decided to use counts of the following per grid cell:

- Housing units in 2019 (2019 5 Year ACS crosswalked from Block Group level)
- Jobs in 2018 (LEHD block-level data)

- Road Intersections (calculated in last blog)
- Percent Developed Land (from NLCD Land Cover Data)

Some scatter plots of these measures by grid cell are presented below from GeoDa:

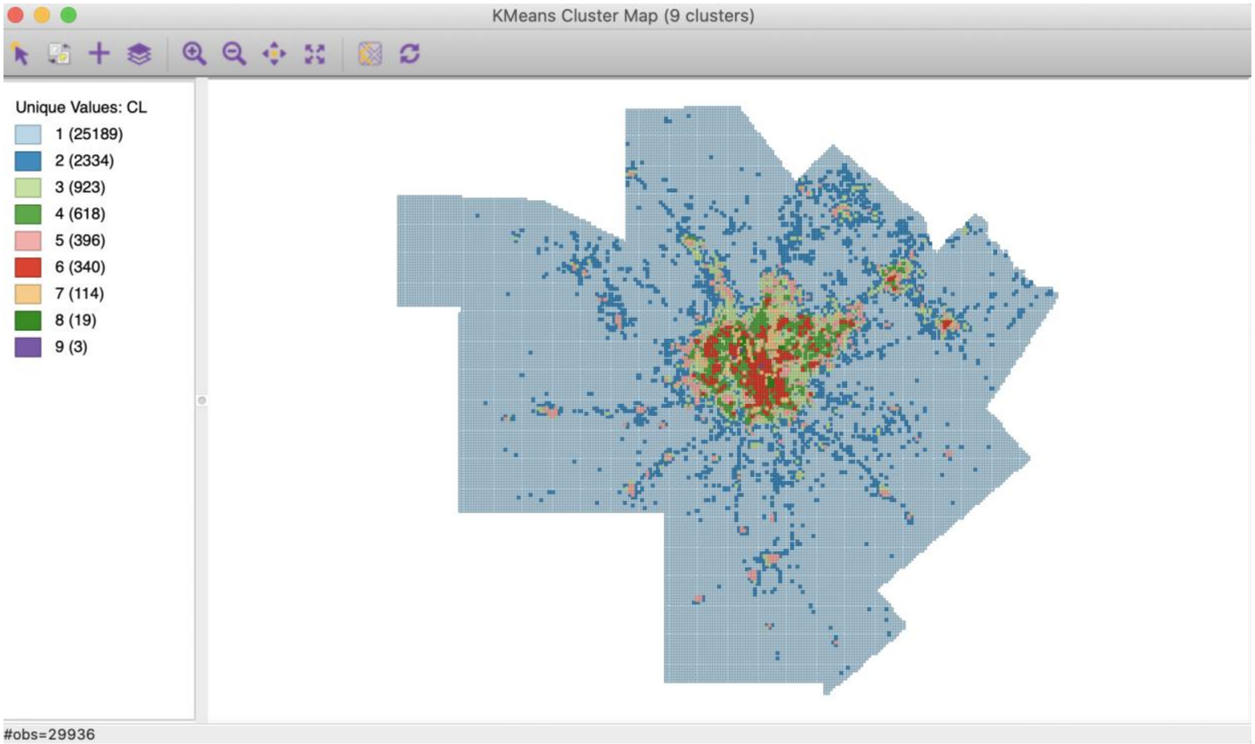


From the histograms in this chart, it is clear there are just a lot of grid cells with low levels of all variables. These are likely the most rural/undeveloped areas. Looking at an aerial view of the region this isn't surprising, since there is mostly green space.

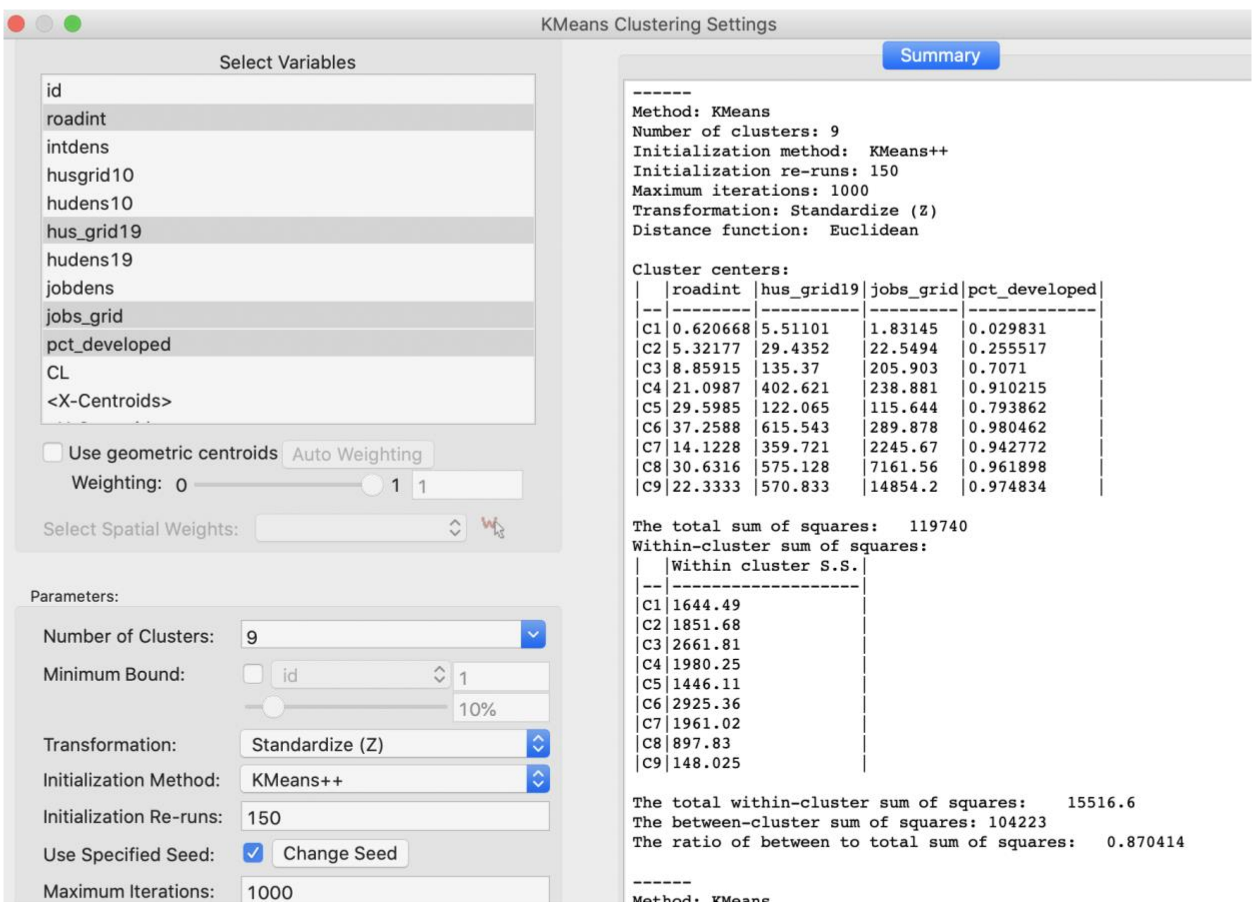
Some of the other scatter plots suggest positive associations. Areas with higher road intersection density are likely to have higher numbers of housing units or jobs, same for areas that are more developed.

For the KMeans analysis here, we will be mostly interested in the areas that have varying degrees of different measures. Are there rural areas that have more of a commercial or residential leaning? What types of areas are walkable (high intersection density) and have high levels of jobs or housing?

The KMeans analysis was conducted in GeoDa. After several iterations, and reading through this guiding blogpost , I landed on 9 categories that seemed like a good starting point for teasing out the different land uses in the area.
The resulting map in GeoDa looked like this:



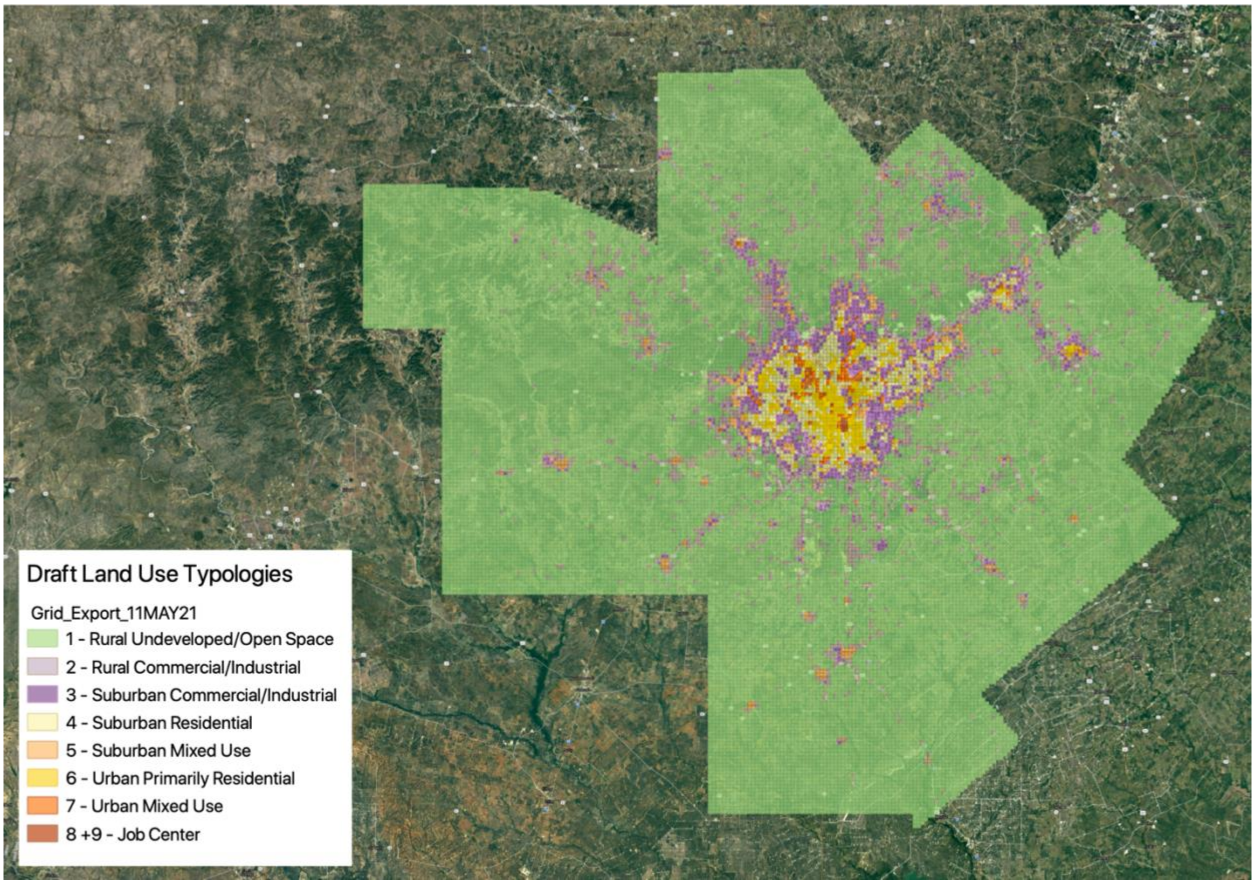
With the details of the analysis here:



From the chart of Cluster Centers (or means for each cluster) we see that on average, grid cells in the C1 cluster have less than one road intersection (i.e. very few roads), 5 housing units, 1 job, and are less than 2% developed. This cluster would represent our most undeveloped areas, and is shown in blue in the initial GeoDa map.

The following clusters have carrying degrees of housing, job, intersection, and development intensity. C2 and C3 still appear rather rural in nature, but have increasing amounts of housing units and jobs. C4, C5, and C6 appear to be more residentially focused than C7, C8, and C9. In order to get a clearer picture of the clusters, I saved the KMeans cluster results and exported the data back into QGis to overlay with Google Satellite Hybrid.

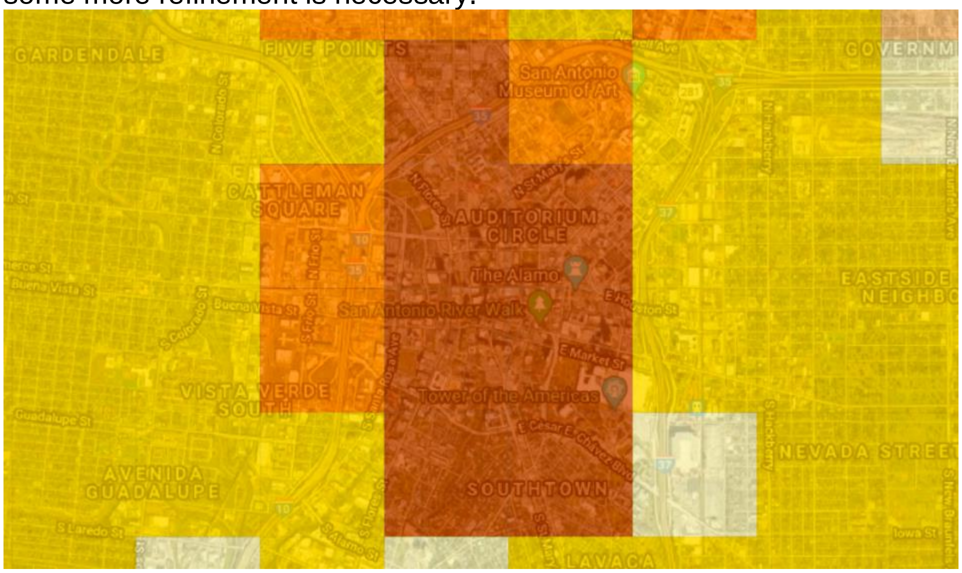
Once doing this, I took some time zooming around the areas, and based on the aerial view underneath the grid cell, and the general aspects of the cluster assigned to each category, I came up with names for 8 separate categories, combining 8 and 9 into one.



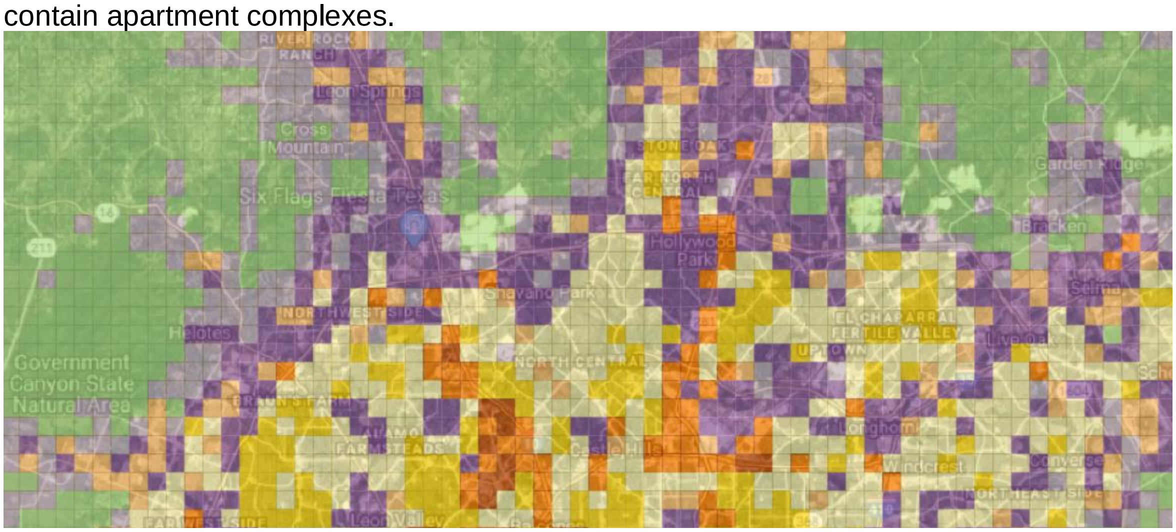
These categories are assigned new colors to correspond to their development type and intensity. Green is primarily rural – all of the grid cells with very little development intensity of any kind. The lighter colors represent lower intensity uses, with darker colors representing higher intensity. Purple is generally used for industrial areas, but here is assigned to areas that may have commercial or industrial uses – since the only variable in the cluster analysis was jobs, we do not have a sense for the type of jobs in those areas. The yellow shades are primarily residential areas, with the orange shades as areas that have more of a mixed use feel – or more of a balance between the number of jobs and housing units. Lastly, the darkest red areas are primarily commercial, and generally are located in downtown, the medical center, over

USAA's campus, and a handful of other spots throughout the city. Clusters 8 and 9 were combined because only 3 cells were assigned to cluster 9 – this seemed to be because they had a much higher level of jobs. This could be due to the way data is reported in the LEHD – sometimes a company headquarters is assigned at one address, and all employees of that company are assigned to that block, even though employees might be located throughout an area. More research would be needed here.

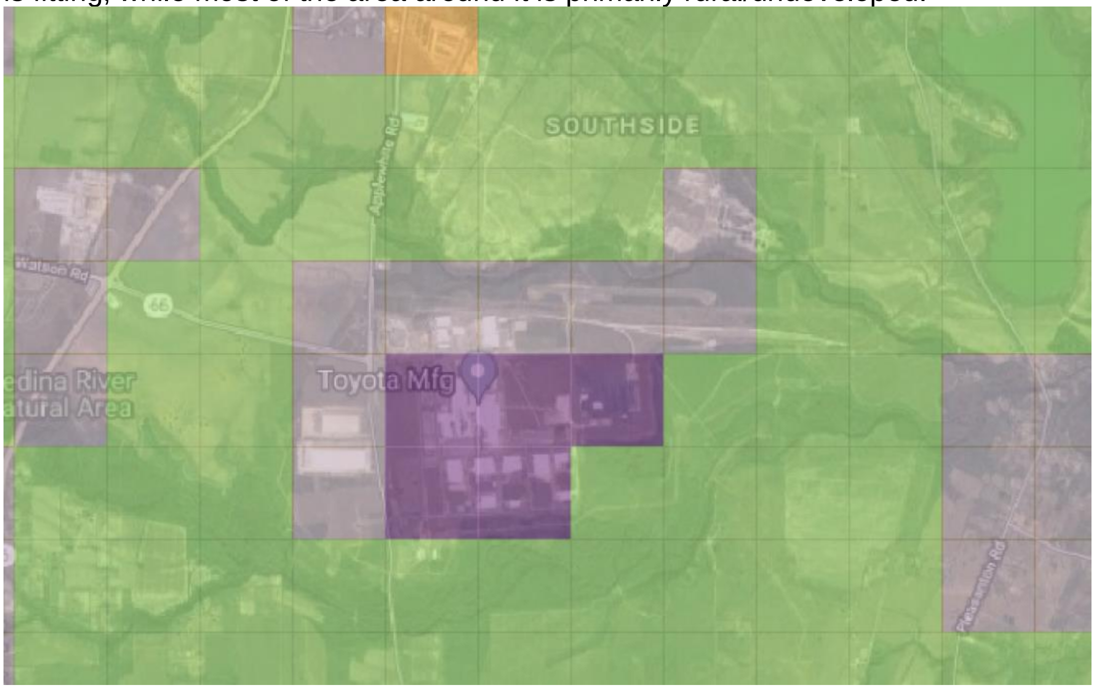
Zooming in on downtown, we can see that the most dense part of the downtown area is classified as a Job Center, while the area near UTSA downtown is "urban mixed use." The Westside and Eastside neighborhoods are categorized as "Urban Primarily Residential" – which seems to be differentiated from other forms of residential development primarily by the increased number of road intersections of older neighborhoods. The grid cell over the Alamodome was flagged as what I categorized as "Suburban Residential" however. Clearly, some more refinement is necessary.



On the north side of town, the shopping areas are classified as "Suburban Mixed use" but it does not distinguish between whether they are commercial or industrial areas. Some areas I classified as "urban mixed use" are also located in north San Antonio, which are likely areas that contain apartment complexes.



South of downtown, the Toyota Plant was classified as "Suburban Commercial/Industrial" which is fitting, while most of the area around it is primarily rural/undeveloped.



A summary table of the total intersections, jobs, and housing units is presented below by Cluster area.

Cluster	Grid Cells	Intersections	Housing Units	Jobs
1	25,189	15,634	138,817	46,132
2	2,334	12,421	68,702	52,630
3	923	8,177	124,946	190,048
4	618	13,039	248,820	147,628
5	396	11,721	48,338	45,795
6	340	12,668	209,285	98,559
7	114	1,610	41,008	256,006
8 + 9	22	649	12,640	180,632
	29,936	75,919	892,556	1,017,432

Next Steps/Refinement:

I think it would be much more interesting to have used the percent of impervious surface from NLCD, rather than the percent of undeveloped land through landcover. The way the categories are defined in the data I used, includes a "developed open space" category, which would include parks or grass in urban areas. I think using impervious surface data instead could have helped differentiate better between, say, Fort Sam Houston and a major industrial area. In the mind of this model, both of those areas have the same level of developed land, but from an impervious surface standpoint the industrial area might go into a separate cluster.