LIVABILITY OF NEIGHBORHOODS IN THE CITY OF AUSTIN

Yining Wang

GEO 386G

5/3/2018

Introduction

With the rapid growing population in the city of Austin, locating the neighborhood with a better livability becomes a concern, especially for the newcomers. This project is a suitability study that aims to identify and sort neighborhoods in the city of Austin by their livability. Livability is defined based on four criteria in this project, namely, risk of flooding, frequency of crime, walking distance to park, and vegetation coverage. Each of the individual criteria will be studied and internally scaled, before being weighted overlaid to generate a synthesized map showing livability levels in different parts of the city of Austin.

Data Collection & Processing

DEM data is the key input for risk of flooding analysis, and is downloaded from the USGS's EarthExplorer website https://earthexplorer.usgs.gov/ (Fig. 1). The DEM is collected by the ASTER satellite in 2011, and the resolution is 1 arc-second (30 m).

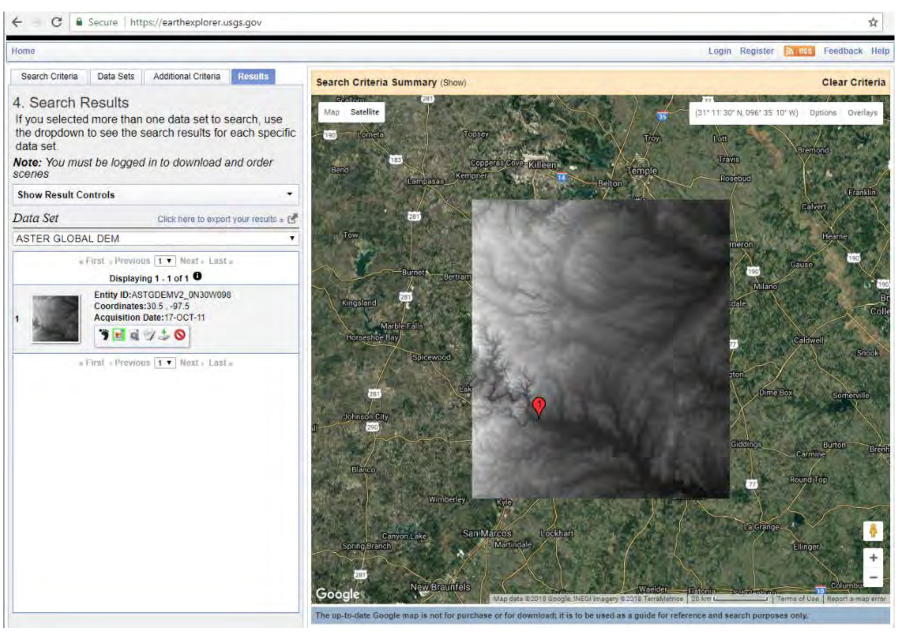


Figure 1. *Downloading Austin DEM data from the EarthExplorer website.*

Satellite images in separated spectral bands are downloaded from USGS's GloVis website https://glovis.usgs.gov/ (Fig. 2) for the purpose of calculating NDVI. Since the data is collected on April 23, 2018, when the trees and grasses are back to life in Austin, the data would be representative and informative on the vegetation density in Austin.

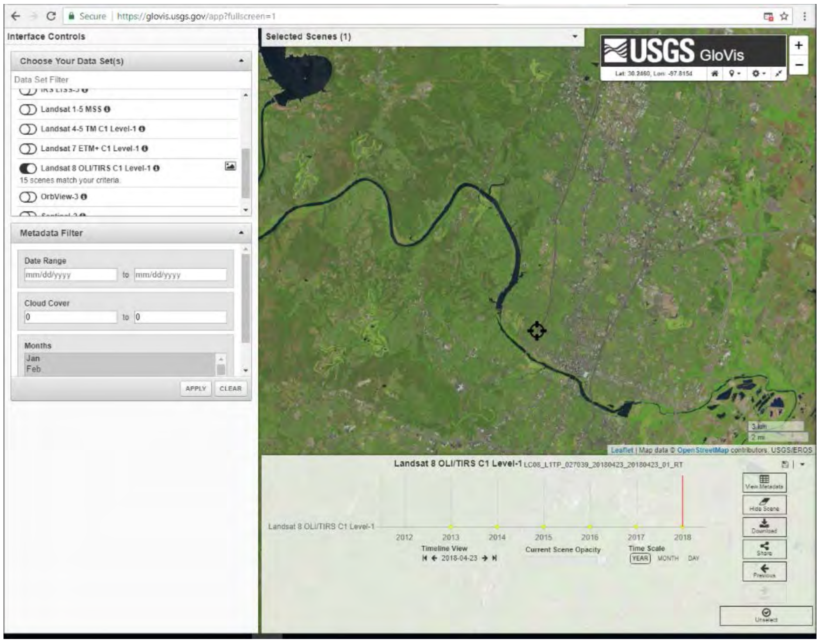


Figure 2. *Downloading Landsat 8 spectral bands from the GloVis website.*

The Landsat 8 satellite images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. The resolution of Band 8 (panchromatic) is 15 meters. The band assignments and their corresponding wavelengths are given by Table 1:

Table 1. *Spectral bands of Landsat 8 images and their corresponding wavelength.*

Landsat 8	**Bands**	**Wavelength (micrometers)**	**Resolution (meters)**
Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS)	Band 1 - Ultra Blue (coastal/aerosol)	0.435 - 0.451	30
	Band 2 - Blue	0.452 - 0.512	30
	Band 3 - Green	0.533 - 0.590	30
	Band 4 - Red	0.636 - 0.673	30
	Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
	Band 6 - Shortwave Infrared (SWIR) 1	1.566 - 1.651	30
	Band 7 - Shortwave Infrared (SWIR) 2	2.107 - 2.294	30
	Band 8 - Panchromatic	0.503 - 0.676	15
	Band 9 - Cirrus	1.363 - 1.384	30
	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

* TIRS bands are acquired at 100 meter resolution, but are resampled to 30 meter in delivered data product.

Crime data is collected from the City of Austin’s public data website https://data.austintexas.gov/Public-Safety/Annual-Crime-Dataset-2015/spbg-9v94 (Fig. 3). This data set is a spreadsheet listing all individual crime incident recorded in the year of 2015. Each crime incident is reported with a location and zip code. The zip code could be used to georeferenced the data, but some data processing is required before it being imported into ArcMap.

Figure 3. *Downloading the 2015 annual crime dataset from the City of Austin's public data website.*

Since the goal is to classify each postal zone by crime incident density, unique values of zip codes are extracted from the spreadsheet by using the *advanced filter* function. Counts of crime incidents are made by using the *countif* function in excel. The result is shown in Table 2.

Table 2. *Count of crime incidents associated with zip codes in the year of 2015*

O	P	Q	R
Zip Code	Crime Count in 2015	Zip Code	Crime Count in 2015
78613	390	78732	1
78617	276	78733	1
78652	13	78735	298
78653	27	78736	71
78660	114	78737	4
78701	2103	78739	185
78702	1668	78741	2973
78703	738	78742	30
78704	2571	78744	1921
78705	1346	78745	2422
78712	3	78746	812
78717	237	78747	275
78719	186	78748	1536
78721	482	78749	845
78722	332	78750	419
78723	2124	78751	927
78724	562	78752	1178
78725	45	78753	3472
78726	237	78754	428
78727	539	78756	352
78728	13	78757	1186
78729	493	78758	2563
78730	47	78759	1403
78731	566		

Shapefiles of roads, postal areas with zip codes, parks and open spaces, census tracts with population are collected from the CAPCOG Regional Open Data website https://regional-open-data-capcog.opendata.arcgis.com/. All the data imported and produced in this project are projected to the NAD83 UTM Zone 14N coordinate system.

Analysis

The ArcMap Modelbuilder function is used to track and automate workflow (Fig. 4). Most of the steps implemented in this project are shown in the model, except some data processing steps such as select by attribute and some network analysis steps such as finding the closest facility. However, steps not mapped on the modelbuilder will be explained in text.

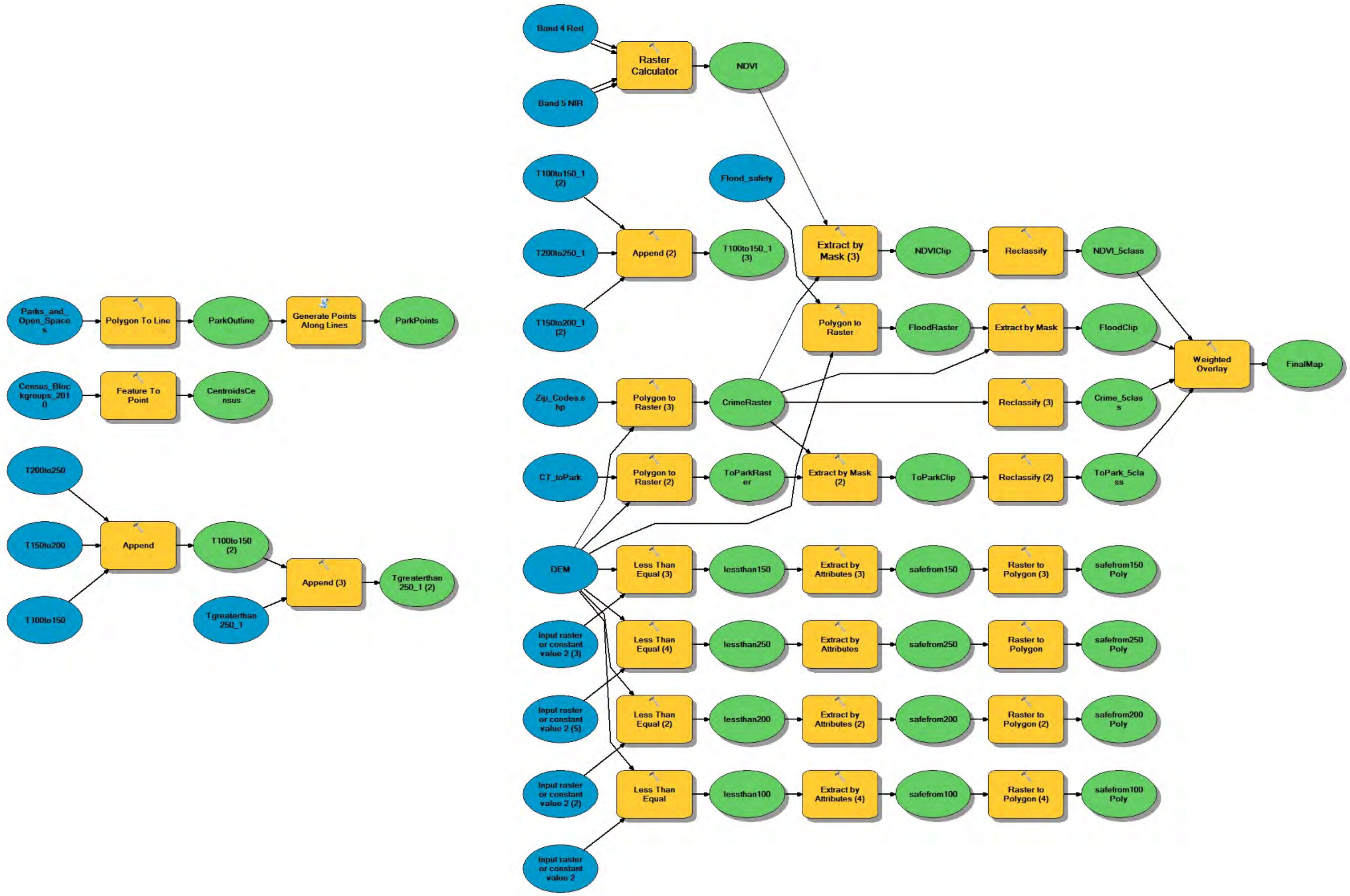


Figure 4. *Model built for this project. Yellow boxes represent tools while blue and green circles are input and output datasets.*

Safety from Flooding

In order to rank the safety of an area from flooding, such as during hurricane events, the Digital Elevation Model (DEM) is used together with the *Less Than Equal* tool. The vertical unit of the DEM data is meters, so the thresholds are also set in meters. Four threshold values, namely 100 m, 150 m, 200 m, and 250 m are used to represent four levels of flooding events. The resulting binary raster files all contain two values, 0 and 1. By using Extract by Attributes, values of 0 are extracted and the resulting raster files represent areas safe from each level of flooding event. These files are then converted into vector files by using the *Raster to Polygon* tool, before being appended to a single vector file using the *append* tool (Fig. 5). The result is a vector file with four polygons each with a unique attribute value indicating level of flooding.

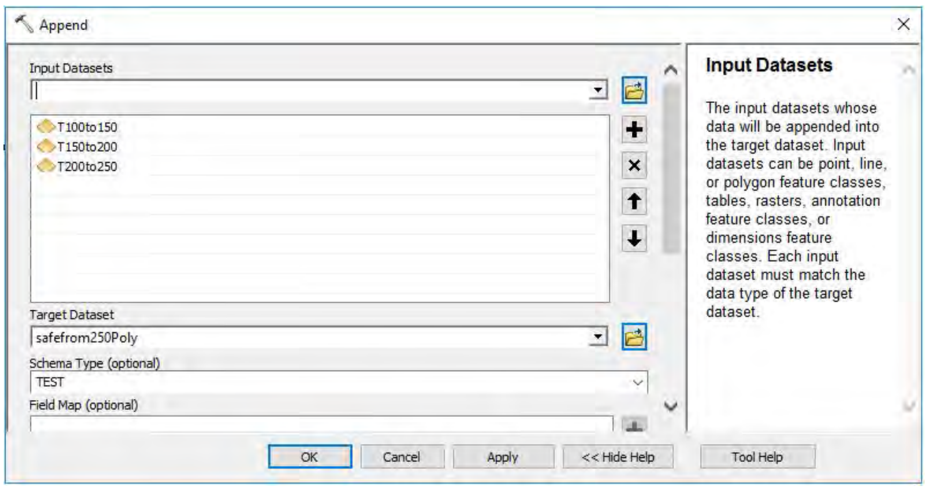


Figure 5*. Appending so that all vector files are combined to one.*

Crime Rate

The crime incident count is associated with zip code in the excel file, and to geospatially express the crime rate, the crime count is manually put into the attribute table of the zip code feature class in ArcMap. A new field containing calculated area of each postal area is added to the attribute table. This allows calculation of crime density by dividing crime counts by the area of the associated postal area. The individual postal areas are then symbolized with graduated colors representing crime density in the year of 2015.

Walking Distance to Parks and Open Spaces

Only walkable networks are considered in this criteria, i.e., expressway, freeway, and service roads are not taken into account. Therefore, in the attribute table of the feature class representing all roads in the City of Austin, road types of “FWY”, “FR”, “EXPY”, “HWY”, “RAMP”, and “SVRD” are deleted from the feature class using the editing tool. Figure 6 shows a comparison between the roads feature class before and after deleting the non-walkable portions.

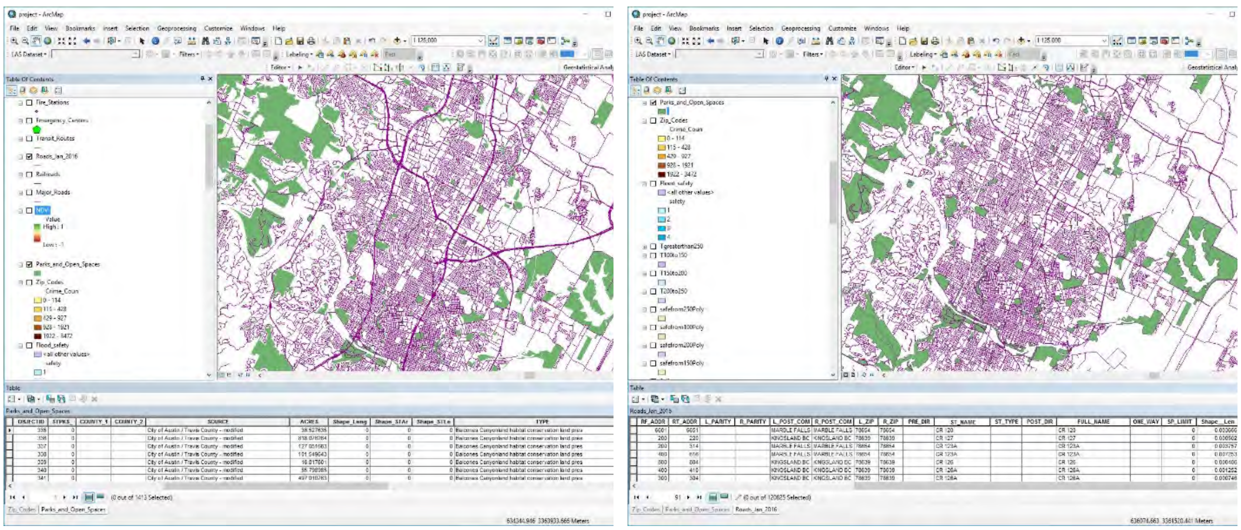


***Figure 6**. Before (left) and after (right) removing the non-walkable roads. Note all the highways and freeways (e.g., IH35, Mopac expressway) are removed. Purple lines represent roads and green polygons are parks and open spaces.*

Network analysis is the ideal extension to use in finding the least distance path from each census tract to the nearest park. However, the input could only be point features. Point features representing Parks’ “entrance” are created by first extracting parks’ outlines using *polygon to line* function, followed by the *generate points along lines* function. When generating the points representing park outlines, a distance of 100 m is used so that each point is spaced 100 m apart from each other for each outline. The result is shown in Figure 7.

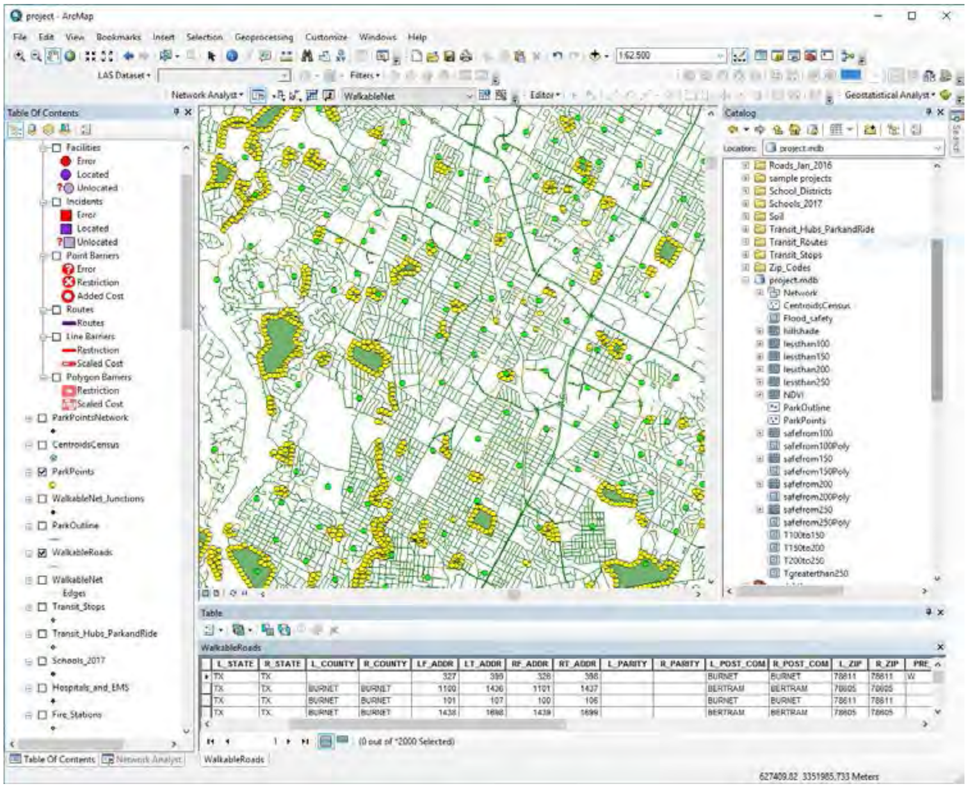


***Figure 7**. Outlines of park (green polygons) represented by point features (yellow disks) spaced 100 m apart. Also shown are census tract centroids (green disks) and walkable road network (green lines).*

A network could then be created using the walkable road feature. In the attribute table, length of each road is calculated and time to walk through each road is estimated by dividing the distance by an average walking pace of 1.4 m/s, i.e., 84 m/min. The time is thus expressed in minutes. Routes from census tract centroids to the closest park could be then derived by using the *closest facility* function in network analysis. The points representing park boundaries are loaded as facilities, and census tract centroids are loaded as incidents. The network used is the walkable roads that is already built. Note that the routes calculated in this case is not the Euclidean distance, but is based on walkable roads solely, so that it gives a better representation of real-world scenarios when a resident wants to walk to the closest park or any other green spaces. Figure 8 shows the routes derived from executing the *closest facility* function. Each green dot is the centroid of its associated census tract, with green polygons representing parks, and yellow lines representing routes to the closest park.

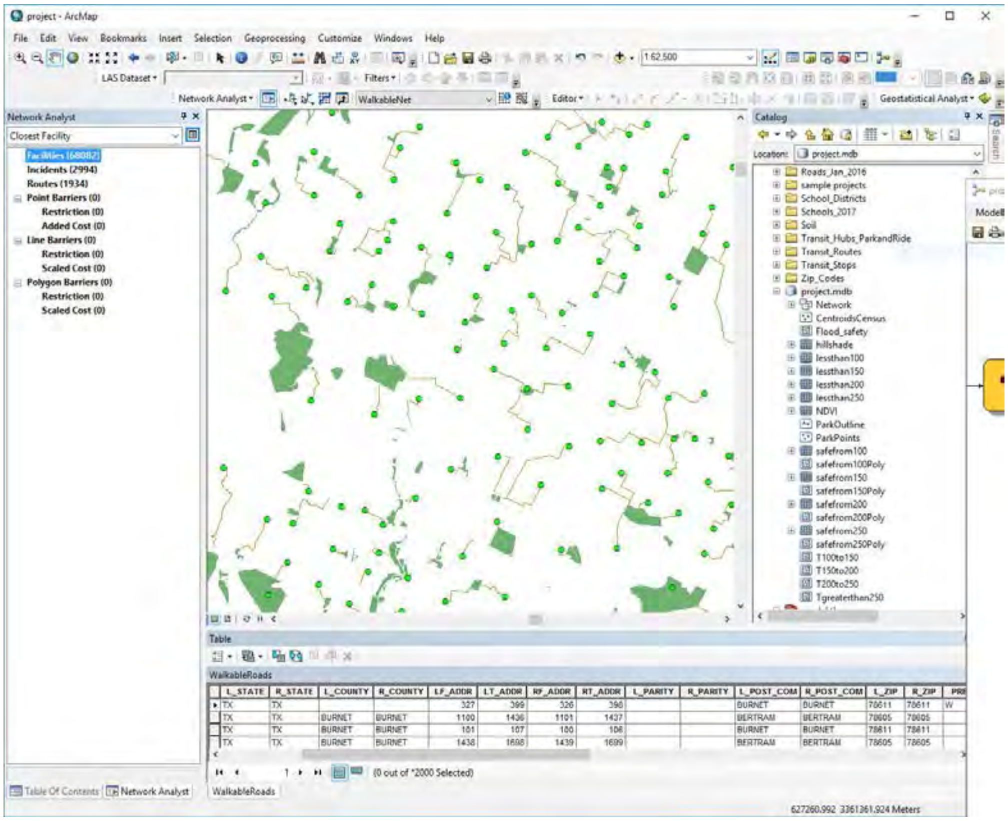


Figure 8*. Result of closest facility analysis. Green disks, green polygons, and yellow lines represent census tract centroid, parks, and shortest routes respectively.*

To associate the proximity of a census tract to a park, each route has to be linked to the centroid and census tract it is associated. This is done by setting the name field of incidents, i.e., centroids as CTID, which is also the unique ID that links between centroids and census tract polygons. Figure 9 shows a selected census tract, shortest route to the nearest park, and centroid of that census tract, all with the same CTID = 147.

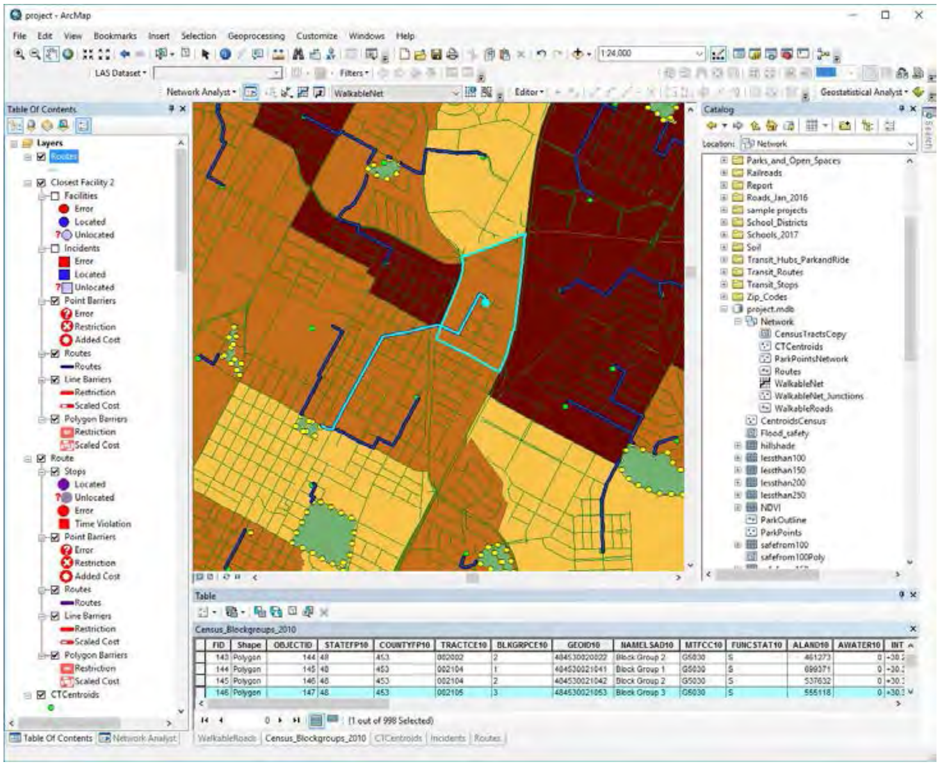


Figure 9*. The bright turquoise highlights associated census tract centroid, census tract polygon and shortest path to the closest park. Different shades of polygons represent census tracts symbolized in terms of population density. Green polygons are parks outlined by yellow dots.*

The census tracts could then be symbolized to represent easiness to access parks and green spaces based on walking distance. This is achieved by creating a new field, in the census tract attribute table, representing the length of routes, which is calculated using the *calculate geometry* function. The quantile method is used so that in each class there are same number of counts. 5 levels of proximity to parks are shown in Figure 10, with darker red representing census tracts with worse access to parks and open spaces, while lighter color represents the opposite. Green polygons are parks, and green circles represent census tract centroids.

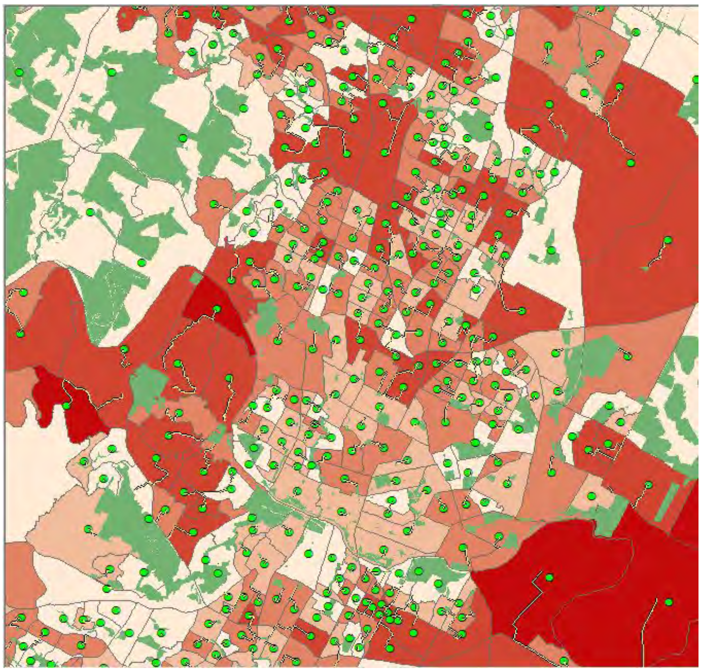


__Figure 10__. Census tracts in graduated color indicating walking distance to the closest park. Darker red represents longer walking distance. Routes are shown as green lines. Also shown are parks (green polygons) and census tract centroids (bright green disks).

Vegetation Coverage

Normalized Difference Vegetation Index (NDVI) is used as an indicator of vegetation density. The idea behind is that pigment in plant leaves, chlorophyll, strongly absorbs visible light (0.4-0.7 um) for use in photosynthesis, but on the other hand, strongly reflects near-infrared light (0.7-1.1 um). Therefore, the higher the difference between reflected near-infrared (NIR) and red, the denser the vegetation.

For NDVI calculation, Bands 4 (Red) and Band 5 (NIR) are used as inputs into the simple equation:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

Calculation is done by the *Raster Calculator* function in ArcMap, with the exact equation input being:

Float (NIR – Red) / Float (NIR + Red)

The Float function is used so that the output is the full span of decimals between the lower and upper bound of -1 (least vegetation coverage) and 1 (most vegetation coverage).

Weighted Overlay

To synthesize all the data created so far, the data need to be overlaid upon each other, and they need to be raster datasets rather than vector. Therefore, zip codes with crime density, safety from flooding, census tracts with distance to closest park are converted to raster files by using the *polygon to raster* tool. Then all the raster files, including the NDVI, are extracted by mask, with the mask being the zip codes polygon, because it is the feature with the smallest spatial extent. These clipped raster files are then reclassified so they each have five classes, except for flooding safety which is already sorted into four classes originally. The last step is to weighted overlay the raster files (Fig. 11).

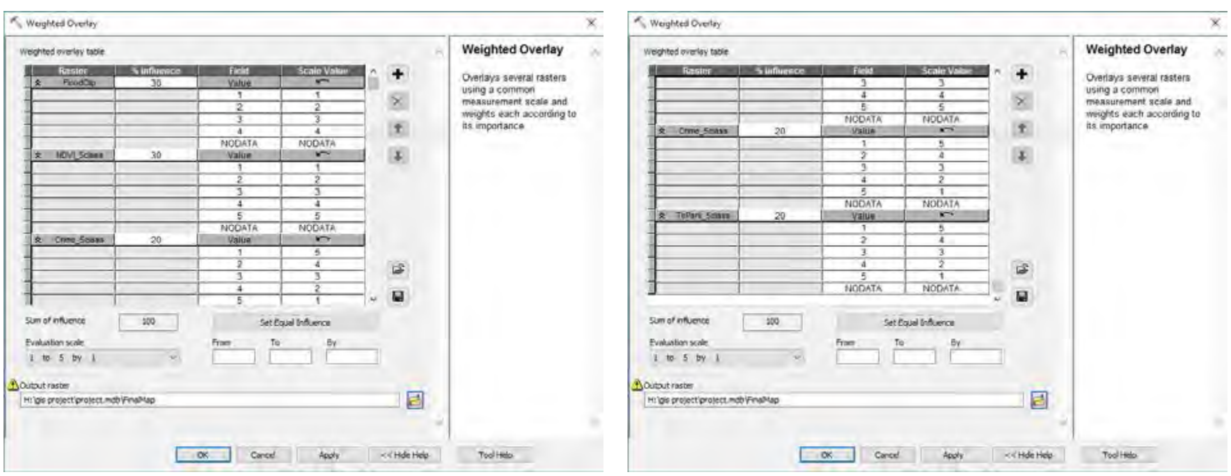


Figure 11*. Using the Weighted Overlay tool to get synthesized scale based on the four criteria. Note the scale value for safety from flood and NDVI values are in the same order as their initial field value, but the crime density and distance to park are ordered backwards. The evaluation scale is from 1 to 5 by 1 with 5 representing highest livability. The % influence is assigned to each layer indivicually.*

The scale value of flood safety and NDVI are with the same order as their original values, since we want higher scales to represent better livability. The crime density and distance to park rasters are scaled backward, because in these cases lower raster values represent more desirable living conditions. The four criteria are given different influence in the overlaying process. Safety from flooding and NDVI influence the process by 30% individually and crime density and walking distance to parks are given 20% of importance.

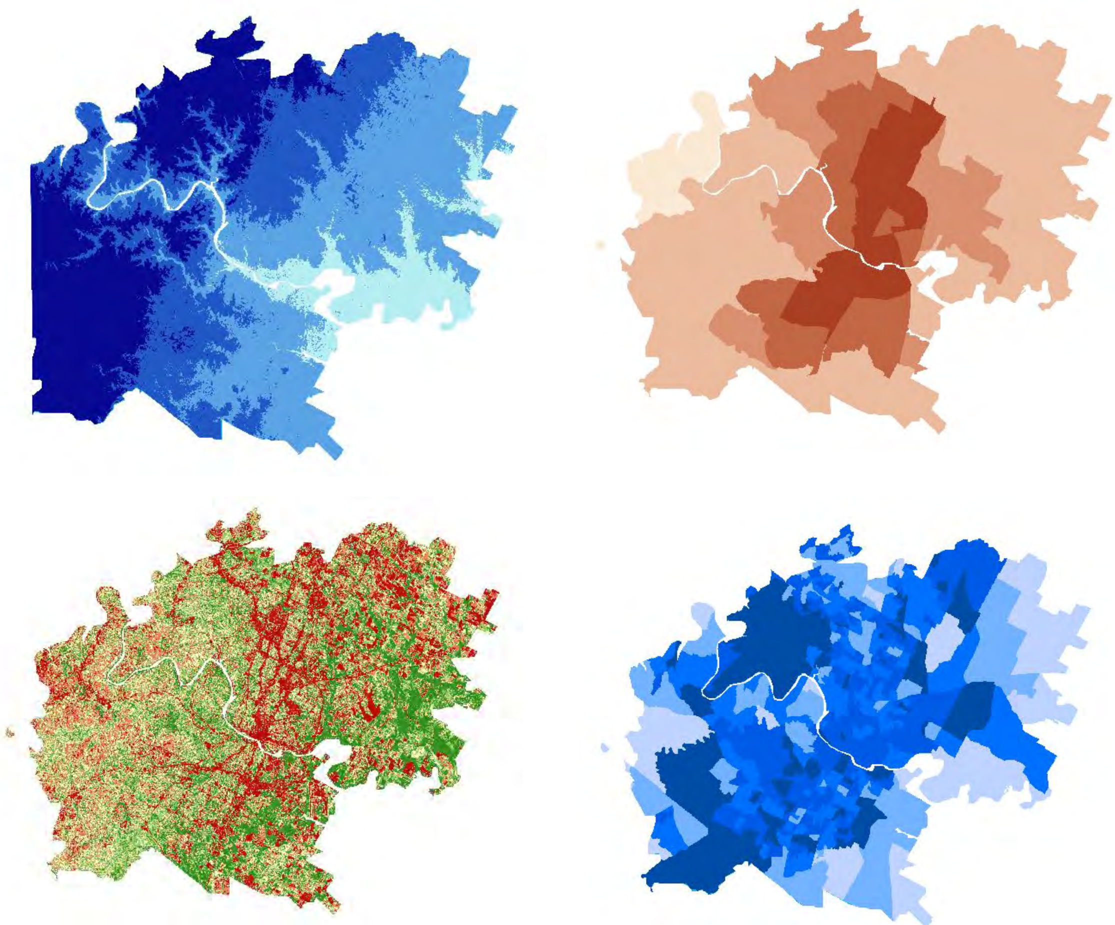


Figure 12. *The final maps of the four criteria, clipped to the city outline. Safety from flooding is shown on the top left with darker color indicating safer regions from flooding. Crime rate is shown on the top right with darker color indicating higher density of crime incidents. NDVI is shown on the bottom left, with greener area representing higher vegetation density. Distance to park is shown on the bottom right, with darker color representing shorter walking distance to the closest park.*

The distribution of population density and livability follows the geographic signature of Balcones Fault (Fig. 13 & Fig. 14). Population is much denser along the fault strike of northeast-southwest, rather than spreading east-westwards. Livability is higher on the west side of the city overall, potentially due to better vegetation coverage on the west side, and higher elevation and thus lower risk of flooding. The population density ideally should some degree of positive correlation with livability, and to first order they do correlate, but only for area with denser population (Fig. 14). For example, at the intersection between US183 and Loop 1 in northern

Austin, there is a patch of red on the livability map, which corresponds to an enclave of low population density. Another patch at the intersection of US183 and the Colorado River has a very low livability (a scale of 1), and on the population map, this area does show a low population density. However, the first order correlation does not work for areas on some of the suburban regions, especially those on the west side. The population density on the western suburbs are very low, but the corresponding livability is high. This is predominantly due to lack of consideration of distance to downtown. Table 3 lists count of pixels representing various scales of livability. The counts show a considerably good spread that none of the classes predominate the whole map. However, most of the mapping area is classified as livability = 2, 3, or 4.

Table 3*. Livability sorted in five classes and counts of pixels in each class.*

Livability	Count of pixels	% Coverage
1	6671	0.35
2	395429	20.69
3	1020380	53.41
4	474306	24.83
5	13751	0.72

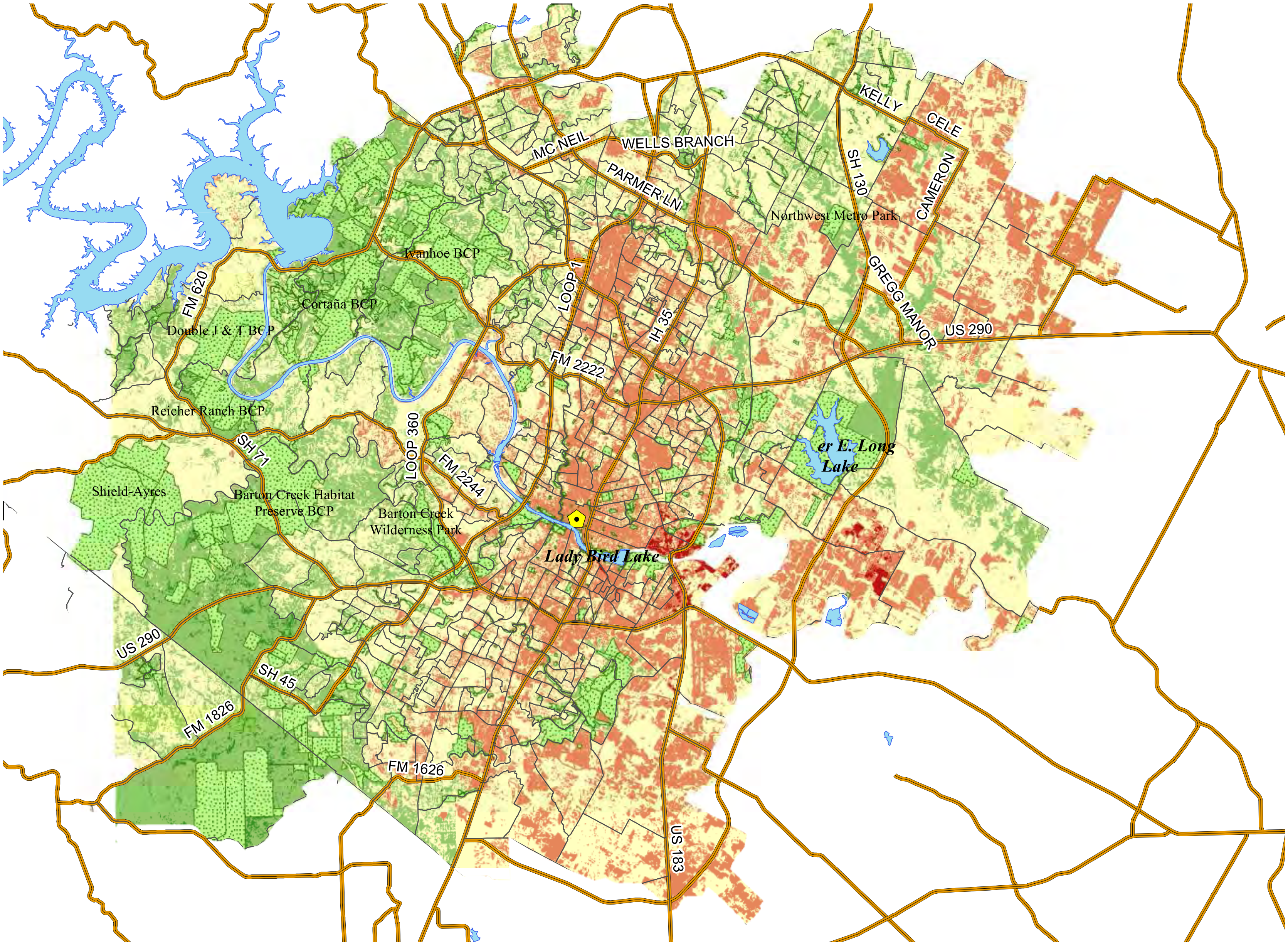


Figure 13. *Final map of livability analysis.*

Limitation and Outlook

Crime rate is only limited to the 2015 dataset, rather than being an averaged value through years. This could introduce certain bias to the analysis. Furthermore, the crime count is normalized by area of each postal area to represent the crime density, but population density is another factor influencing the total number of crime incident, which is not normalized in this analysis.

In the network analysis, the input incident feature is set to centroids of each census tract. Therefore, the shortest routes derived are not necessarily the shortest for every part of the census tract. It is a only first order approximation and is meant to represent the average distance.

Distance to downtown (not only Euclidean distance, but also network distance based on access to drivable roads) is another important factor that will be considered in real-world scenarios when people choose a neighborhood to live in. This should be taken into account in future analysis. Another approach in testing the livability data is to compare the derived livability with population density and unit density data. Further work could be done by exporting and measuring areas with higher livability, and comparing them with population density, on a census tract basis. A scattered plot would also help visualize whether there is a linear correlation or not. However, this is better done after importing the criteria of distance to downtown. That way, regions in the suburbs, especially on the west side of the city would get a more realistic livability rating, which is not as high as the current one.

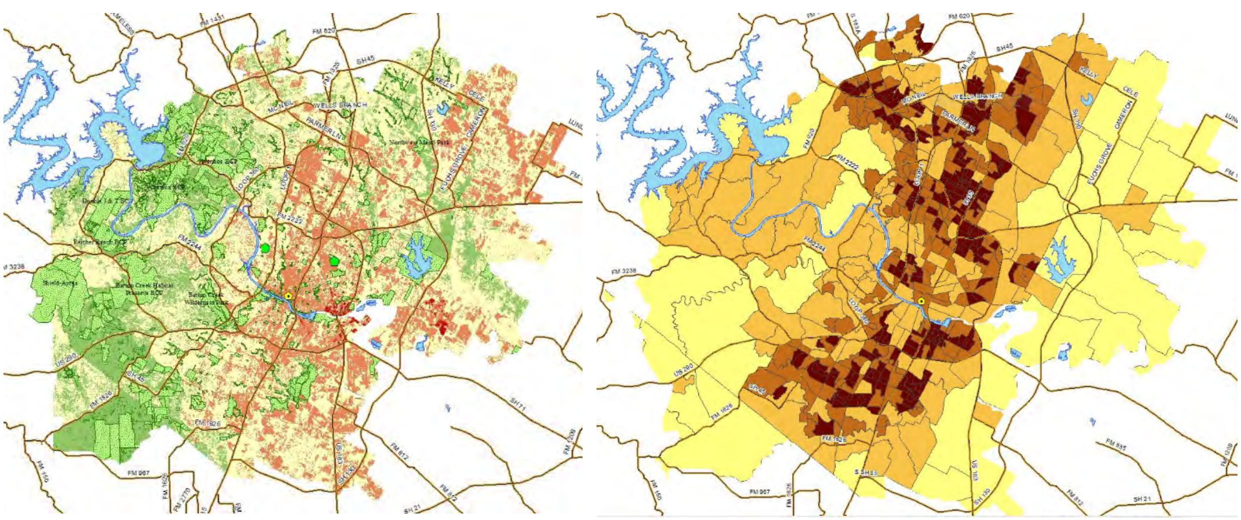


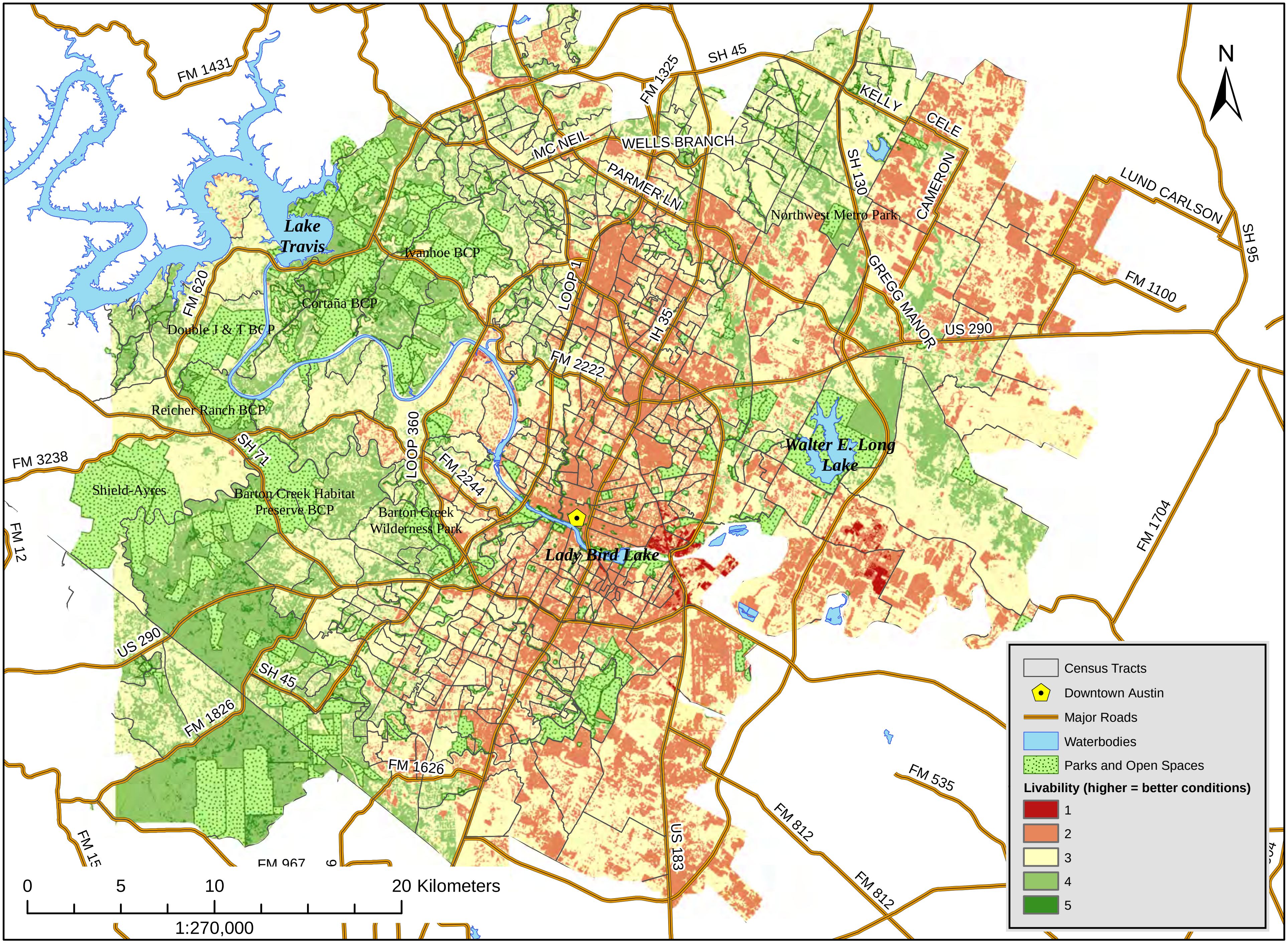
***Figure 14**. Correlating livability with census tract population density (right). Darker color represents higher population density.*

Reference

Barsi, J.A.; Lee, K.; Kvaran, G.; Markham, B.L.; Pedelty, J.A. The Spectral Response of the Landsat-8 Operational Land Imager. Remote Sens. 2014, 6, 10232-10251. doi:10.3390/rs61010232

Livability of Neighborhoods in the City of Austin, TX

Based on risk of flooding, crime rate, accessibility to green space, and vegetation coverage



Projected coordinate system: NAD83 UTM Zone 14N. Made by Yining Wang (yw9594) 4/29/2018