Sarah Spicer  
UEP 232 – Intro to GIS  
Semester Project  
9 December 2008

Project Description

My project examines the use, efficiency, and possible improvement opportunities for public transit in the Boston metro area. Spatial analysis enabled me to develop a ranking system by census block group. This ranking quantified relationships between my six component maps. As expected, the maps also show a strong relationship between population density and public transit use, as well as between median income and public transit. The final map shows block groups ranked by a combination of population density, commute time, median income, and transit accessibility. Combining these factors in GIS using a ranking system enabled me to concisely map these relationships on my final map, showing a well rounded picture of Boston’s transit needs.

Data Sources

All of my data came from the Mass GIS website, under the census section (http://www.mass.gov/mgis/census2000.htm). This data dates from 2000, when the census was taken, and the shape files are in the Massachusetts State Plane (NAD83 Meters) coordinate system. The data I used is organized by block group (http://www.mass.gov/mgis/cen2000_blockgroups.htm). In specific, my project used the following layers:
- **CENSUS2000BLOCKGROUPS_POLY**
  This layer formed the base of my study, with all my data linked to the blockgroups therein.
- **TOWNS_POLY**
  I created a selection from this layer that included only the metro Boston municipalities which I wanted to include in the study. I also included the whole data layer separately to provide a reference point for the selected communities, so that they weren’t floating on their own.
- **MBTABUSSTOPS_PT**
  This contains the point information for bus stops. I used this layer to perform network analysis.
- **MBTA_NODE**
  This contains the point information of rapid transit (T) stops. I used this layer also to perform network analysis.
- **CEN2K_BG_TRNS_COM_MEANS**
  This data set gives the number of people using each type of transit on their daily commutes, including public transit, walking, biking, driving, etc. Since I was just focused on the use of public transportation, I compared the public transit field with all other fields added together.
- **CEN2K_BG_TRNS_COM_TIME**
  This gives information on the number of people whose commutes take less than 5, 5-10, 10-15, etc, minutes. In order to classify these, I decided that "long commutes" meant commutes longer than 45 minutes. I had first tried this using 30 minutes and up and decided that there were too many people whose commutes took longer than 30 minutes, so this wasn’t enough of an extreme.
- **CEN2K_CT_MED_INC_FAM**
  This contains the median household income per block group. There is a field containing median income per household (INC_MED_HS), which is what I mapped.
• NEWENGLAND_POLY
  Political boundaries of New England, used to make a locus map.
• MBTA_ARC
  Used to make the map of the T system.

I also considered including layers displaying transit routes, boardings by station (from CTPS), and water body files such as HYDRO25K_POLY, but I found that these layers made my maps too cluttered. I believe that the boardings by station data, indicating the frequency/amount of transit use in specific neighborhoods, was covered by the “percent of commutes by public transit” map instead. My intent was to keep the maps as simple as possible to make sure it was easy for the viewer to understand my project’s focus.

List of major steps in data preparation and analysis

The first step was to define the project area and build a base map. I selected Boston and municipalities surrounding it. I retained other municipalities around the selected ones to give context. I then added the census block groups layer and created a selection of only the block groups within the selected towns. I also added the T stops and bus stops layers and selected only those stops falling within my study area. Once I had this layer, I was ready to start my analysis. I joined the household income, commute transit mode, and commute transit time to this layer using the LOGRECNO field. I created the following additional fields in the table:

- Long_trans: Sum of all commutes by public transit longer than 45 minutes (adding the Pub_GT60, Pub_45_59, etc.)
- Long_Other: Sum of all commutes not but public transit longer than 45 minutes.
- LongTransP: Percentage of transit commutes longer than 45 minutes.
- LongOtherP: Percentage of non-transit commutes longer than 45 minutes.
- Rank_PopDn: This is where the rankings begin. I used a process of selection to divide these fields into 5 sections, with divisions based on the quantile classification of the data. For this field, for example, I selected block groups where the population density was 0-11 residents per dry acre and used field calculator to set the selection =1. Then the next quantile division was calculated = 2, and so on, to rank all the population densities on a scale from 1-5.
- Rank_Income: used the same process to rank the median income (the INC_MED_HS field).
- Rank_LongT: used the same process to rank the percentage of long commutes by public transit.
- Rank_LongO: ranking of the percentage of long commutes not by public transit.
- Rank_Dist: This ranking, for the distance from transit stops map, worked differently.
With my 500-meter buffer representing walking distance, I determined that areas should be considered either within walking distance or outside of walking distance, gaining a ranking of 0 or 1. It was difficult to ascertain what selection criteria should be used, however, as I discuss further below. I ended up selecting block groups that contained neither T stops nor Bus stops and giving them a rank of 1; block groups that contained either were given a 0. However, I was not entirely satisfied with this, so I selected by hand the block groups that were clearly not covered by the 500-meter buffer. I realize this is not very precise, but I did not want to neglect block groups that had, say, 1 transit stop within them, but were otherwise completely devoid of transit. Please see more on this on page 5.

- Rank_Sum: This field contains the sum of all the rankings, added up using field calculator. It is this field that was used to make the final map.

From this attribute table, I was able to make most of my maps. First, I mapped population density by dividing the total_pop field by the dry_acres field. This was a good first step to spatially understand Boston’s neighborhoods a bit better and see where most of the potential transit riders were living. I also had a hunch that this map would be the basis for a pattern: higher density areas showing greater use of transit; greater density of transit stops; lower income; and shorter commutes. Next I mapped median income, making a choropleth based on the INC_MED_HS field. As expected, this map looks like the exact reverse of the population density map, with higher density areas being the poorer areas. Poorer, denser areas should surely receive the best transport infrastructure since there are more people there, period; likely more people there are also without cars because of either parking or income constraints.

Next, I created the “Commuters Using Public Transit” map. This was created by dividing the number of workers using public transit by the number of total commuting workers. Again, this map matched the pattern I expected, with more transit users living in denser, poorer neighborhoods. However, I decided that, though important to spatially define where more transit users live, this map was limited in use because of the question of cause and effect. Should areas with low transit use receive less transit support, since there are apparently fewer...
users there? Or should areas with low transit use receive additional transit to try to encourage transit use? In other words would more people use transit if more transit was available? Or are they demographically inclined not to use transit, period? Since my study cannot answer these questions, this map is for display purposes only and was not used in the transit-need rankings of the block groups.

Next, I examined the breakdown of long commutes. As mentioned above, I defined a “long commute” as one longer than 45 minutes, and added up the number of commuters per block group with commutes longer than 45 minutes, those taking public transit and those using other means. I then mapped the percentage of both of these modes out of the total number of commuters.

My last small map, which shows the walking distance around transit stops, was made using network analyst. To do this, I followed the tutorial on network analyst and just applied the steps to my own project. I actually did this twice, once for the T stops, and once for the bus stops. I then also changed the setting from 800m to 500m, because I felt this measurement better reflected how far a commuter was willing to walk on a daily basis.

The last, “Results” map was a basic choropleth of the rankings sum.

**Difficulties Encountered**

My main difficulty was the fact that my information was available only in census block group form. I would have liked the ability to perform a more detailed analysis using a smaller unit of measure. The larger block group measure was particularly difficult when trying to rank areas using the 500-meter buffer around transit stops. If I selected all the block groups that the buffers intersected, I felt that I was selecting too many of the block groups were selected. If I selected ones that contained all of the buffer, there were far too few selected. So, I ended up essentially ignoring the 500-meter buffer in my analysis, which was a big disappointment. I instead selected block groups that contained T stops and bus stops. I then de-selected block groups that were not significantly covered by the 500-meter radius. I was much more comfortable with this selection than those I had come up with before; but I also feel that this was a much less data-driven and precise method. I did consider trying to use a census block layer to find which blocks (rather than block groups) fell within the 500-meter buffer; however, I was unable to come up with a concise method for incorporating this information with all the other data I had, which was block-group based. I thought maybe I could devise a formula by which I could join the blocks with the block groups (with a certain number of blocks falling into each block group) and then determining area, and then determining what percentage of the block group’s area was covered by the 500-meter buffer. Alternatively, I may have been able to determine the area of each segment of the 500-meter buffer within each respective block group; however, as might be obvious, this would be a
highly laborious task and would result in a level of detail mismatched with the rest of my data. So, I instead selected blocks where there were transit stops, but were not covered more than 50% by the 500-meter radius. I wish there had been a more precise way to select these blocks, but I believe my end product is more reflective of reality than if I had relied only on selecting groups that contained transit stops. This addition didn’t produce a huge difference, but I felt it was a more accurate result. Also, I wanted to make sure the buffer/walking distance was part of the analysis.

This is a screen shot of all the blocks that were selected (those that contained no transit stops and those that were not covered in the majority by the buffer zone.)

I am also uncertain whether it was wise to use meters versus feet. I used meters because all my data was in meters by default. After I did the network analyst operation, I wondered if American viewers might have a more difficult time relating to walking distance in meters/kilometers versus miles. I personally don’t mind, but if I were to redo the network analyst, I might redo it in feet. (Just as a note, the network analyst took about an hour to complete given the number of points that had to be computed and buffered, so switching back
and forth on units or buffer size was cumbersome. I did change the buffer from 800m to 500m, and it took another hour. This was longer than expected!

Lastly, since the last census was in 2000, my data is quite outdated. I would really like to redo this project when the 2010 census data becomes available, to see how transit use has changed in Boston over those 10 years.

**Concluding Thoughts**

I think that my approach worked well other than the difficulty I had properly selecting block groups based on transit walking distance. I think that combining the 6 factors made for a well-rounded analysis- more demographic data giving a better picture of Boston’s neighborhoods and their needs.

The pattern that emerged was one that I expected. The poorer, denser areas ended up with the highest scores, showing that although sustainable suburbs are also a priority, we still need to pay the most attention to the inner city/urban areas.

I know that the MBTA and EOT have their own studies and maps that would make mine look really simple by comparison, but I'm quite pleased with mine for a first try. As a non-driver I wanted to take a closer look at the transit system I use regularly, to appreciate what good design was already there and to see if I could recommend improvements. I believe my final map, in a very basic sense, might be interesting to MBTA / EOT officials as an “outsider’s view” of the transit system, and I hope to complete increasingly detailed studies in a similar vein in the future.