Andy Freed
GEOG 588
GIS Project

Research Question:

The question I'm investigating is admittedly self-centered. My wife and I are starting to look for homes in Portland, and I'd like to expand on the "this sounds like a nice neighborhood" method of narrowing down search locations. Access to GIS software and data will allow me to examine a number of factors that ordinarily might not be considered by a typical real estate agent and might otherwise label us as "picky." Ultimately, the research question is this:

"Where should we look for homes in Portland?"

Tools like RMLS search are already available for the public, and they provide a wealth of information, but exclude a number of attributes that we consider essential to our permanent residence. As such, some unique data sources will need to be created for this project. These include coverages for brewpubs, theaters, grocers and restaurants. We'd like to look for a home in places within easy walking distance to a local, grocers, food, public transportation, parks, schools, etc. The selection of data and sources will be discussed later.

Additionally, I'd like to investigate ways to use results from this analysis to compare with homes for sale on sites like RMLS.com. This may be too far beyond the scope of the project, but I'm curious if there is any way to narrow down by zip code + 4 or neighborhood to improve searching. If nothing else, it will at least give us some locations to do some footwork.

Data Sources

A bulk of the data for analysis came from the RLIS Lite data available from Metro such as Freeways, Major Arterials, Streets, Taxlots, City Fill, Schools, Soils, Parks, Transit, etc.

Some data sources were created for the purpose of this exercise. To create these data sources I've utilized online databases such as Qwest's Dex service (http://www.dexonline.com) for grocers and theaters, and Beer Advocate's beerfly database for pubs & beer bars. Tabular lists of these data were created and arranged into CSV files then imported into ArcGIS using the address matching tools provided in the software.

Data Collection

Luckily most of the data used in isolating potential home sites is already available through the RLIS dataset at PSU. To simplify, I may need to combine some coverages and clip the extent to make it more efficient for running analysis. By using the CITY_FILL shape of just Portland, it was easy to clip much of the RLIS data to just the City of Portland boundary, one of our first requirements. This also made for much easier analysis later on by excluding extraneous info.

I requested some crime info from Corporate GIS at the City of Portland by contacting the person listed in the meta-data on the website, but never received any response. This information would be nice to have, and I may try again to obtain it.

There was a fair amount of data to collect, but using the streets info from the RLIS data, it was easy to create a geocoded datasets for things like pubs. In many cases, I was able to create a CSV file with place names, street address, city, state and zip code that were then fairly easily imported into a shapefile. The same process was used in creating coverages for theaters, grocers and restaurants, though the later was taken from Verizon's SuperPages because it allowed for a larger set of results to be displayed at a single time. There were 1900 restaurants in Portland, and 5% of which couldn't be easily geocoded.

Analysis

One of the major challenges to doing site analysis on such a large dataset is the amount of time it takes. Some simple clip operations based on the city boundary allowed me to create new coverages from results as analysis proceeds to cut down on computer time. It will be especially useful to start with major factors like cutting out non-matching data – such as tax lots that are way out of price range. The downside to doing this is that you loose that data for future analysis.

My wife and I compiled a list of ideal conditions for site location which we prioritized and set up distances we'd be willing to travel to each location. We also listed items we wanted to avoid and determined the distance from these items. Ultimately I've broken the analysis into three major parts:

- Finding appropriate Taxlots
- Finding features to avoid (freeways, etc)
- Finding features that attract (pubs, grocers, schools, restaurants, parks)

Taxlots

Finding the appropriate taxlots was quite simple but required some time to run queries. Using Windows Remote Desktop Client allowed me to start queries from home so they could run while I wasn't using the computer.

Starting with the Metro taxlot shapefile, I queried for:
[SITECITY] LIKE 'Portland*'
[LANDUSE] = 'SFR'
[TOTALVAL] > 175000 AND [TOTALVAL] < 225000
[YEARBUILT] < 1950
[AREA] > 4400
[BLDGSQFT] > 1500

I also ran a separate query for taxlots that had been recently sold:

[SALEDATE] > '200201' AND [SALEPRICE] >175000 AND [SALEPRICE] < 225000 AND [YEARBUILT] < 1950 AND [BLDGSQFT] > 1500 AND [LANDUSE] = 'SFR' AND [AREA] > 4400

This gave us just the taxlots that met what we felt to be a good baseline to start looking with. The values were the result of negotiations between interested parties.

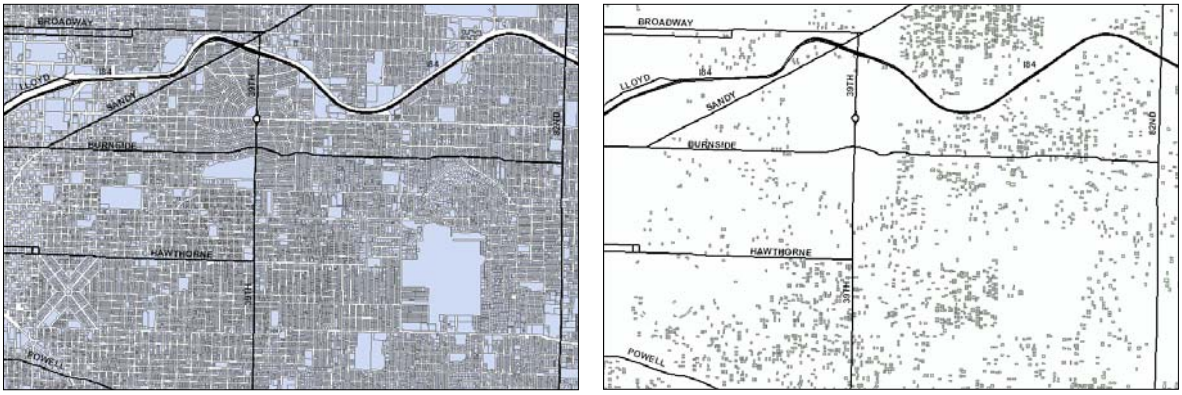

Fig 1: Taxlots before and after query

Avoid

Finding features to avoid was also fairly easy since the data was readily available. The restrictions were based on some information we gained from research (distance from freeway for air quality) and others from personal preference (distance from major arterials). The resulting shapefiles were merged to create a massive blob of “avoid” areas. (Fig 2)

Fig 2: Avoid coverage

The features that we included in the avoid coverage were:

FEMA 100 year flood plains
Hydric Soils (susceptible to liquefaction)
Interstate/Freeways (with 1000’ buffer)
Major Arterials (with 200’ buffer)
Bus lines (with 200’ buffer)
Train tracks (with 200’ buffer)

Some buffer distances were selected on whim, others on known issues. The 200 ft. buffer from arterials, bus lines, and train tracks were for noise and safety issues. The 1000 ft. buffer around freeways was based on the results of several reports describing the increased incidence of repertory disease other health problems that seem to correlate with nearness to freeways.

The hydric soil data was used in place of earthquake hazard, as hydric soils tend to liquefy during an earthquake, increasing the potential damage caused by an earthquake. I was not unable to obtain crime statistics, which would be a welcome addition to the avoid blob.

Attract

The attractions analysis was more challenging in both data collection and in combining the data in a useful manner. The final attractions (or ideal conditions) are as follows:

Grocery within 0.5 miles
Park or open space within 0.25 miles
Bus line within 0.25 miles
Restaurant within 0.5 miles
Good school within 1 mile
Pub within 0.5 miles
Conditional: Theater within 1 mile

This part of the analysis required the most work, since these datasets were generally not available. The Bus line, Schools, and Parks were available through RLIS. However, we wanted to only consider public schools that were rated Strong or better by the Portland Public School District (See Appendix). It was quite simple to join school ratings with the RLIS data set, and the result was a 1 mile buffer around successful public schools. Ironically, in the period of time between creating the dataset and performing analysis, 3+ schools have been marked for closure. Talk about real life lesson in data accuracy.

Creating datasets for Grocers, Restaurants, Pubs and Theaters was done in a similar fashion using an online phone directory and pasting results into a text file. The text file was then massaged (using macros or manually) to convert it into a CSV file that could be imported into ArcGIS. ArcGIS's address geocoding tool was then used to match street addresses and zip codes to the RLIS streets data. The smaller datasets like theaters, pubs and grocers had a high level of success in geocoding. The restaurants coverage only matched 95% of the 1800 entries, which seemed good enough of a sample for this project.

Finally, all of these features were intersected in pairs (You need ArcEditor to do more than two at a time) until a final coverage was created of all the attractive locations. This was then dissolved to create a simple polygon for the final analysis.

An alternate attract analysis was created that excluded theaters. The importance of this amenity waned after actually going to a mega-cinema. Besides, we go to movies so infrequently that driving wouldn't be an issue. The removal of theater restrictions clearly increased the areas available.

Fig 3: Final Attract regions

Fig 4: Alternate Attract regions

Final Analysis

Once all three separate pieces were completed, it was time to complete the final analysis. This is where I encountered the first technical problem. ArcGIS includes a CLIP tool, but doesn't include a tool that does the opposite of clip. I remember using a tool called ERASE in the past, but it was clearly missing from ArcMap. Some scouring of discussion boards on the ESRI site led me to XTool Pro; a plug-in for ArcGIS that adds ERASE functionality (among other things). Luckily, it has a 30 day free trial, so I was able to use Erase to exclude the Avoid areas from the Attract areas.

Before the Analysis (Fig. 5) you get a pretty good idea about what will be affected. Using XTool Pro's Erase function, I was able to erase the sections of attract that were also in the avoid area. Then, taking these results, I was able to find taxlots in those final desirable areas using Select by Location:

- Select by Location
- Select Features From taxlots
- That are contained by Attract

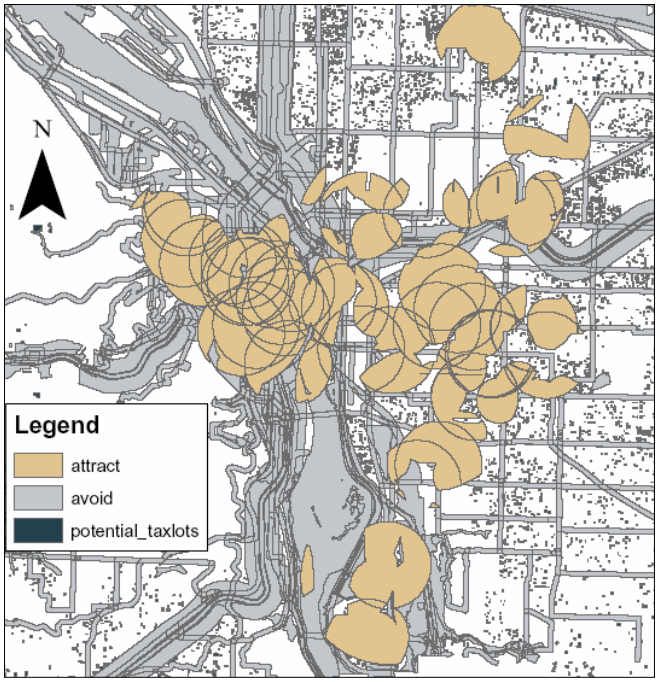

Fig 5: pre-analysis

Results

The resulting taxlots are ideal according to our list of requirements because they are meet all our attract requirements, avoid the restrictions we set, and are within the taxlot limits we set. But the taxlots aren't necessarily what we're looking for. We're just looking for areas where we want to look for a home. Those odd-shaped polygons may not help when watching for houses the come on the market, so we need some easier way to identify them. If only the city was broken down into smaller areas with some unifying label… Oh – neighborhoods!

The last part is to take the ideal taxlots and determine which neighborhoods they fall in. RLIS contains a neighborhood set, so a simple query gives us the following map.

There are a few odd results that do require some consideration though. Both the Northwest District and Goose Hollow seem to only have one applicable taxlot, otherwise, the results aren't that much different than we expected. The lack of property on the west side of the river was likely due to the lack of theaters in Portland city limits on that side of the river.

Since theaters aren't *as* big of a draw to us, I decided to exclude it from a second, alternate analysis that provided slightly different results.

As you can see, there are quite a few addition neighborhoods, especially in North Portland and Southwest Portland.

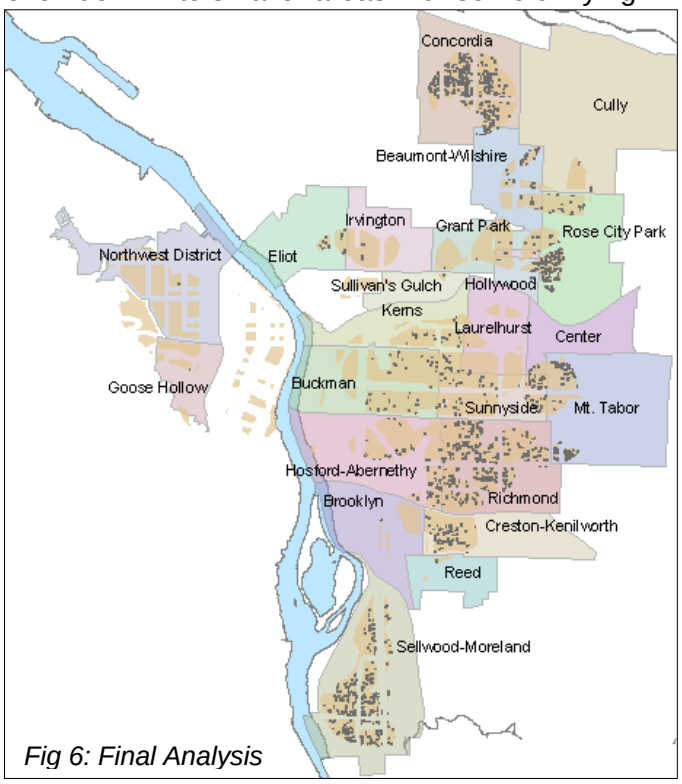

Fig 6: Final Analysis

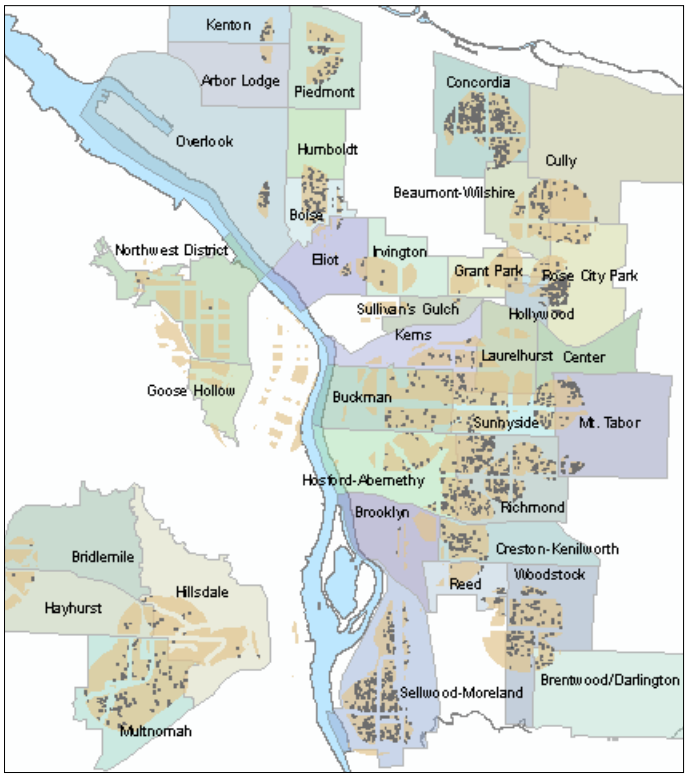

Fig 7: Alternate Analysis

Discussion of Results

This project has shown some of the issues inherent to GIS. It has provided us with a way to identify ideal locations throughout the city, but I'm quite hesitant to share the results because I don't want people to go and buy up all the good homes in the areas I've targeted.

There are some problems with my methods, such as the discrepancy between assessed value and market value. It might be worth consulting with someone who knows about the difference to see if there is a way to estimate market value based on assessed value or other easily identifiable factors.

Lastly, the results of the analysis confirmed some of what we already expected to be true, but also revealed a number of other locations in the city that we had not considered. I would like to explore the results again once I obtain crime statistics; homicide can really affect the attractiveness (or livability) of a location.

Appendix

Freeway Health Info:

http://airmap.unh.edu/assessment/pdf/021021-UltrafineParticleStudies.pdf
http://www.sierraclub.org/sprawl/report04_highwayhealth/report.pdf

School Report Card:

http://www.pps.k12.or.us/news-c/ode_rptcards/PPS1J.pdf
Or http://freed.dyndns.org/gis/ideal_pdx/schoolrating.txt

Portland Pubs & Beer Bars:

http://www.beeradvocate.com/beerfly/city/16/
Or
http://freed.dyndns.org/gis/ideal_pdx/pubs.txt
(Most of these were “ground truthed” by the author)

Restaurants:

http://freed.dyndns.org/gis/ideal_pdx/restaurants.txt
Source: Verizon Superpages

Theaters:

http://freed.dyndns.org/gis/ideal_pdx/theaters_pdx.txt
Source: Qwest Dex and Fandango.com