Praveen Subramani
Athena: praveens
December 7, 2008
11.188

11.188 Final Project: Placing a New Upscale Restaurant in Suffolk County, MA

Abstract

According to an August 2008 New York Times article, over 60% of new restaurants, delis, and food shops close within three years of opening[1]. This sobering statistic is well known by experienced restaurateurs, many of whom agree that location is overwhelmingly the most important factor for success in the restaurant industry. This 11.188 final projected involved the use of ArcGIS software in conjunction with 2000 US Census data, road and public transportation data, land use data from MassGIS, and other demographics statistics from Suffolk county to inform a selection of an upscale Boston-area restaurant targeted at prosperous, working young people of ages 24-39.

Criteria & Methodology

As with any thorough site suitability analysis, the first and perhaps most challenging step was the establishment of relevant criteria for placement of a new Boston area restaurant. I decided to base my site selection on the following factors:

First, I chose Suffolk County, MA – composed of the cities of Boston, Chelsea, Revere, and Winthrop – as the basis for locating the restaurant. (**Figure 1)**

Subsequently, I sought an area with a high percentage of the target 24-39 year-old age group. This value was calculated from the P008 category of the US 2000 Census Summary File 3 (hereupon referred to as SF3), specifically as number of people in this age group divided by total population for whom age data was known, multiplied by 100 to give a percentage. The mean value for percentage of 24 to 39 year olds as a fraction of total population for the county was 28.8% with a standard deviation of 10.45% so I defined a high percentage of the age group as more than a standard deviation above the mean, so I extracted block groups with more than 39.25% of the target age group as "high" (28.8% +10.45% = 39.25%). (**Figure 2)**

Next, I examined land use data which I obtained through the MassGIS website in the form of a shape file with 37-use divisions for the entire state. I extracted the relevant data for Suffolk county and decided that the restaurant should be placed in a commercial zone that was in proximity to residential zones to facilitate access from homes. A 250 meter buffer was created around residential zones and then intersected with commercially zoned tracts to obtain the appropriate land use polygons. (**Figure 3)**

I then investigated areas of above-average population density to ensure that the restaurant is placed in a highly populated area so that a large customer base is available. I obtained this metric by dividing the total population of each block group (also obtained from the P008 category of SF3) by each block group's area in square kilometers, which I calculated using the "calculate geometry" tool in the attribute table. The mean value was 8800 people per square kilometer, so I selected block groups with density field greater than 8800 to obtain groups with an above-average population density. (**Figure 4)**

The next factor I examined was employment rate, which was motivated by the fact that unemployed people are unlikely to patronize an upscale restaurant. I obtained data for employment through the P043 category of SF3, and calculated it as the number of employed civilians as a fraction of total civilians multiplied by 100 to give a percentage (I chose to exclude armed forces and other non-civilians to simplify the calculation). The mean employment percentage for Suffolk County was 92.6%, with a standard deviation of 7.3%. I originally planned to define "high employment rates" as greater than one standard deviation above the mean, but this definition essentially eliminated any block groups that had less than 100% employment, which was a far too strict cutoff. Therefore, after testing several cutoffs, I settled on looking for "above-average" employment, so block groups with 92.6% or higher employment rate were selected. (**Figure 5)**

---

[1] Maynard, Micheline. "Love Food? Think Twice Before Jumping In." New York Times 26 Aug. 2008. 7 Dec. 2008 < http://www.nytimes.com/2008/08/27/dining/27fail.html>

I then explored median household income to ensure that the new restaurant would be located in a high-income region since the analysis focuses on the placement of an upscale restaurant geared at patrons with high disposable income. Median household income was obtained through the P053001 field of SF3, and had a mean value of $41,424 with a standard deviation of $8976.5 for Suffolk County, so I extracted block groups with high household income, defined as greater than one standard deviation above the mean ($41424 + 8976.5 = $50400.5). (**Figure 6)**

Next, I addressed accessibility concerns for the new restaurant by looking at major roads in Suffolk County (obtained through the majmhda1.shp file in the course data locker) as well as MBTA subway and train stations, obtained through the MIT Geodata Repository. After extracting the relevant information for Suffolk County, I created a buffer of 250 meters around major roads and a similar buffer of 250 meters around MBTA stations using the buffer tool through the proximity toolkit of ArcToolbox. I then made a union of the resulting polygons to yield sites easily accessible through road and public transportation networks. (**Figure 7)**

The next step involved intersecting the polygons resulting from the above criteria analysis using the intersect tool from ArcToolbox and dissolving the resulting polygons to obtain some continuous potential sites for the restaurant. I calculated the area for each of the 26 resulting polygons using the "calculate geometry" tool in the attribute table of the layer. However, several sites were quite small and I decided on a minimum cutoff of 2500 square feet for the area of the plot after doing some research on urban restaurant sizes. After extracting sites meeting the minimum area requirement, I was left with 18 potential sites for the new restaurant, which I mapped over Boston-area planning districts (obtained through the MIT GeoData Repository) to get an idea of the neighborhoods that were best suited for the restaurant. Unsurprisingly, 15 of the 18 sites were concentrated in a box centered upon downtown/central Boston, with the most potential sites in the South End. (**Map 1)**

The final step of the analysis was to examine a raster of existing restaurants in the state of MA that I obtained through the 11.521 class data locker. This raster proved extremely useful as it provided the number of existing restaurants per grid cell which I mapped in conjunction with my final potential sites from the suitability analysis to allow the viewer to see where restaurants already existed in relation to the ideal sites. Predictably, nearly all of the ideal sites had already been capitalized upon by other restaurants, since these sites are the results of so many criteria for a successful restaurant. (**Map 2**)

Conclusions

The final maps of the potential sites overlain upon the Boston planning districts and the existing restaurant raster provide an informative view of several possible sites for a new upscale restaurant in Suffolk County. This analysis represents a significant first step in assessing several objective factors involved with placing the restaurant, and the presence of several existing restaurants in the final sites affirms the validity of the analysis since there are already successful restaurants operating in these regions. Of course, location is only one of a huge variety of factors associated with opening and maintaining a thriving restaurant, but this analysis leads me to conclude that GIS software in conjunction with census, transportation, land use, and other demographic data can indeed be used to effectively inform placement of a restaurant, and could contribute to the success of a new restaurant.

Using this data, a restaurateur can begin to examine areas of desired restaurant concentration for placement of a new restaurant. Placing a restaurant in a site where there are already ten existing restaurants is probably not the best idea, for example, because the new restaurant would face intense competition from its neighbors. However, several areas of greater Boston such as the North End (Boston's 'Little Italy') do indeed have a large concentration of restaurants in a small area, and many of these restaurants prove highly successful due to the theme and character of the neighborhood. Thus it is probably in a restaurant-owner's best interest to place the restaurant in a site where there are some existing restaurants to attract customers there in the first place, but not so many that a new establishment would be ignored or out-competed by peer restaurants.

Praveen Subramani
Athena: praveens

December 7, 2008
11.188

Major Challenges & Limitations

One of the biggest challenges with conducting this analysis for the final project was acquiring all of the necessary data to conduct a thorough and meaningful analysis. Unlike our lab and homework exercises in which most of the relevant data was provided to us and easily accessible through the course locker, I had to acquire data for this project from a variety of different sources and then combine them in a common data frame and account for projection differences. I ended up obtaining data from the US 2000 Census SF3 for MA state, the MIT Geodata Repository, the MassGIS website, the 11.520 course locker, and even some class data for 11.521. Finding all the data for the analysis was highly challenging and combining relevant information involved the use of MS Access and several joins in ArcGIS, as well as hefty use of the "intersect" tool through the ArcToolbox.

Furthermore, this project gave me hands-on experience with conducting an entire site suitability analysis from start to finish. The site suitability analysis conducted through Homework 2 and 3 for the Cambridge Senior Center was challenging even when all the necessary criteria were specified for us in the assignment and the data was provided through the course locker. For this final project, I had to determine my own definitions and cutoffs for all my criteria, piece together data from different sources and in different file formats, and explore new tools in ArcGIS to perform the analysis. Determining the cutoffs and definitions for criteria such as "high population density" or "high percentage of 24-39 year olds" was particularly challenging because it involved looking at a variety of statistical measures such as the mean and standard deviation for the data, and then looking at several cutoffs based on these measures to get final polygons for each criterion. If the cutoffs were too strict, the final results did not yield a sufficient number of suitable sites. Alternatively, lax cutoffs resulted in too many final sites that did not sufficiently represent the selected criteria, so a great deal of careful balancing as well as trial-and-error was required to get a reasonable number of final sites.

The biggest limitation of the final analysis is probably the age of the data. Most of the data, particularly demographic information from the census, is from 1999 and 2000. Restaurants come and go very quickly and a huge number of current factors contribute to their success, so an analysis that is based on 8 or 9 year-old data has limited relevance to placing a restaurant today. However, I used the most current and relevant data I could find, and the final results still hold some worth in current times.

New GIS Tools Used and Techniques Learned

The most useful new tool I encountered was the "Census Data" tool of the MIT GeoData repository toolbar. The Census Data tool allows a user to choose specific fields from Summary Files 1 and 3 from the US 2000 Census for any counties in MA and create an output shape file containing only specified fields and subject characteristics. This tool proved incredibly useful to me because after establishing a number of criteria, I realized that many of them would only be available through census data. I initially thought I might have to download the entire SF3 shape file for the whole state which would be very large and slow to deal with, and then sift through all the fields using the PDF manual for SF3 to find the ones I was interested in and then manually extract them in MS Access. Fortunately, the Census Data tool of the MIT GeoData repository allowed me to generate a shape file just for my target area of interest (Suffolk County) with the attributes I selected, so I only had to deal with a file that contained exactly the census information I was concerned with.

I also gained some good experience with utilizing the MassGIS website to obtain shape file data and associated AVL files to give pre-defined classification and legend schemes. We worked a little with this in class, but often the files were provided for us or we were given a link to the relevant webpage. This final project taught me about data acquisition and how to effectively manage data acquired through a variety of sources.

Praveen Subramani
Athena: praveens

December 7, 2008
11.188

Figures

Note: The following figures are **not** intended to be used as complete or professional maps. Consequently, they do not individually feature essential map elements such as titles, legends, source listings, etc. They are simply meant to be shown as an explanatory accompaniment to the Criteria & Methodology section so the reader can get a taste of the different stages of the suitability analysis. The two final, professional maps that show the final sites overlain on Boston planning districts and on the existing restaurants raster are provided at the end.

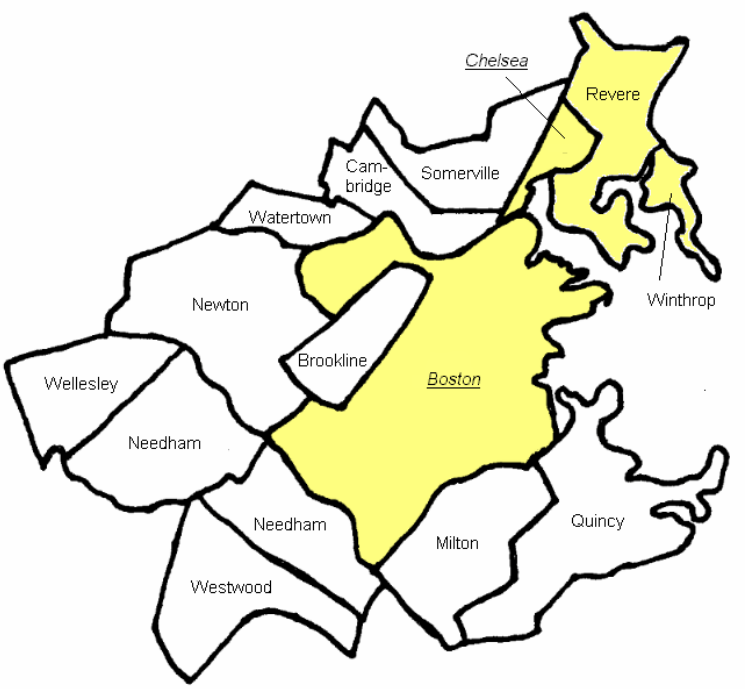


Figure 1: Suffolk County (shaded in yellow) and surrounding cities.
Source: http://www.mass.gov/agr/mosquito/suffolk.gif

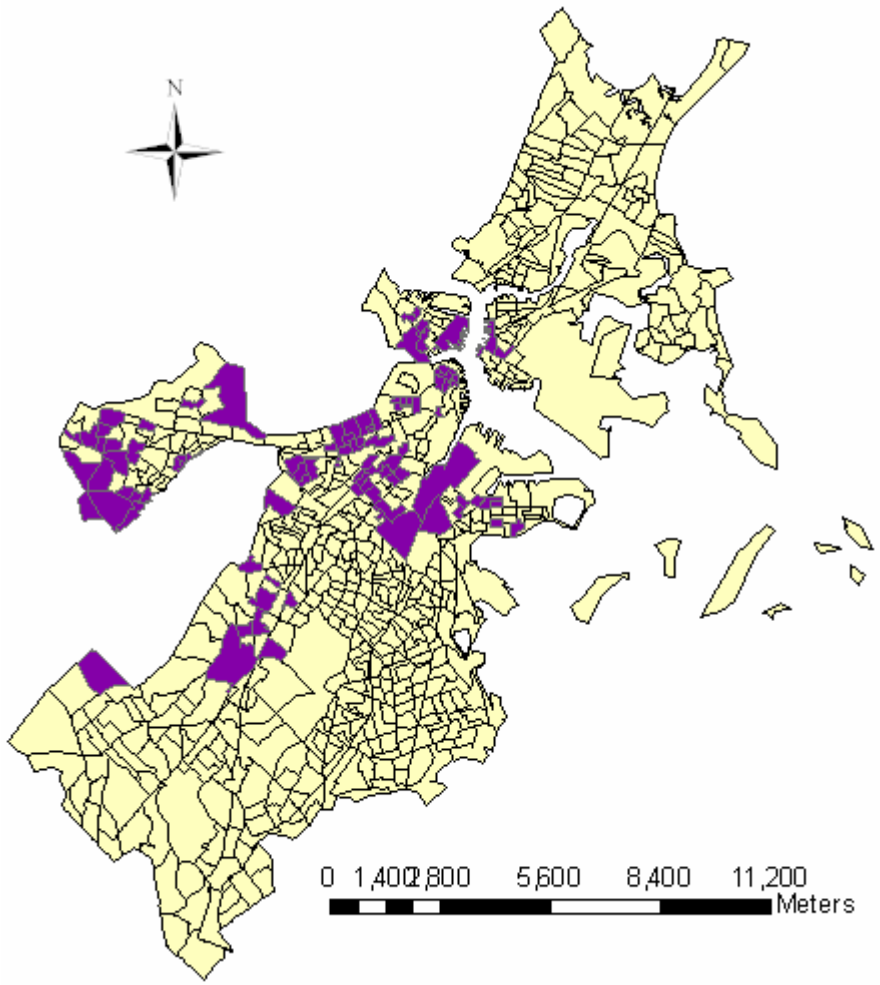


Figure 2: Block groups in Suffolk County with a high percentage of 24-39 year old age group (shaded in purple). High is defined as more than one standard deviation above the mean. Mean value for percentage of 24 to 39 year olds as a fraction of total population is 28.8% and a standard deviation is 10.45% so we define a high percentage of young people as more than a standard deviation above the mean = (28.8% +10.45%) = 39.25%

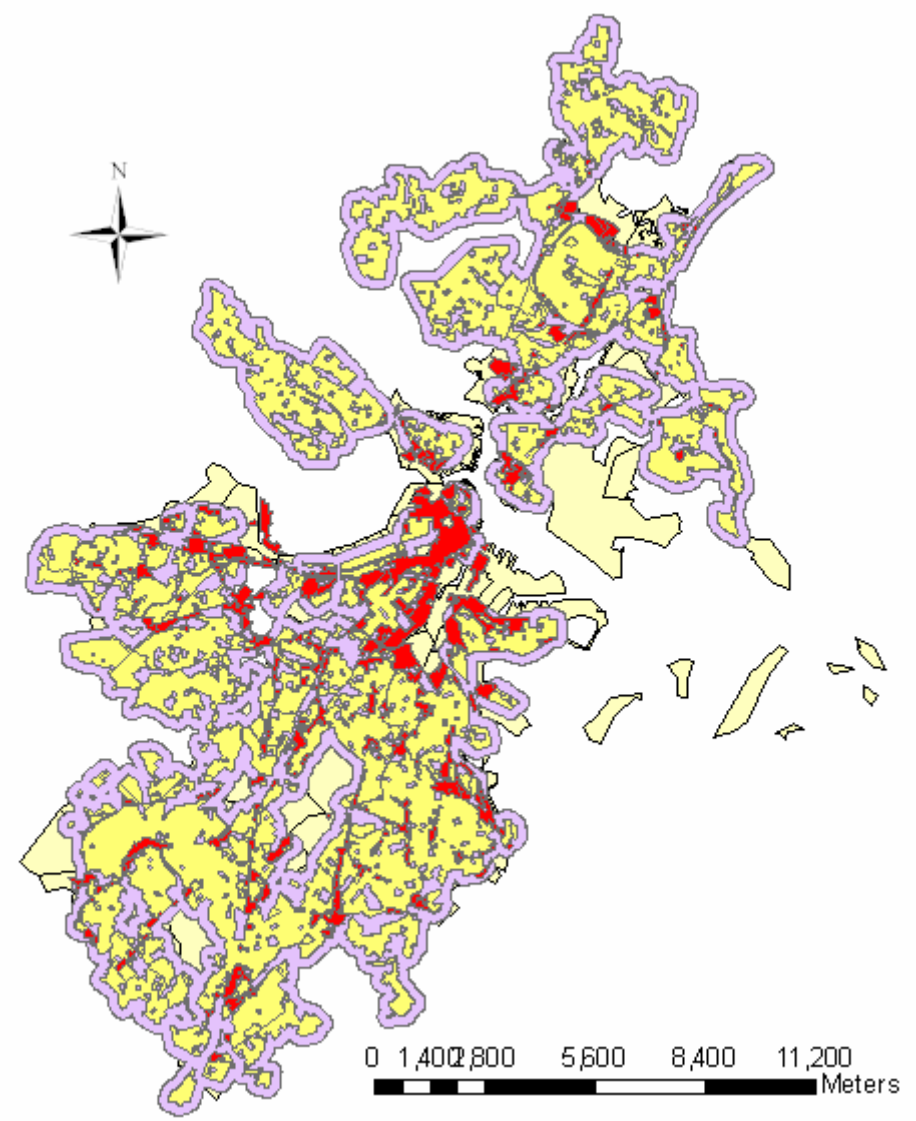


Figure 3: Land use in Suffolk County. Residential areas (all types and densities) are shown in yellow with a 250 meter buffer of these residential areas shown in light purple. Commercial areas are shown in red. Acceptable land use was defined as land that was in a commercial zone but also in the 250 meter residential buffer, so it represented an intersection of the red and purple layers.

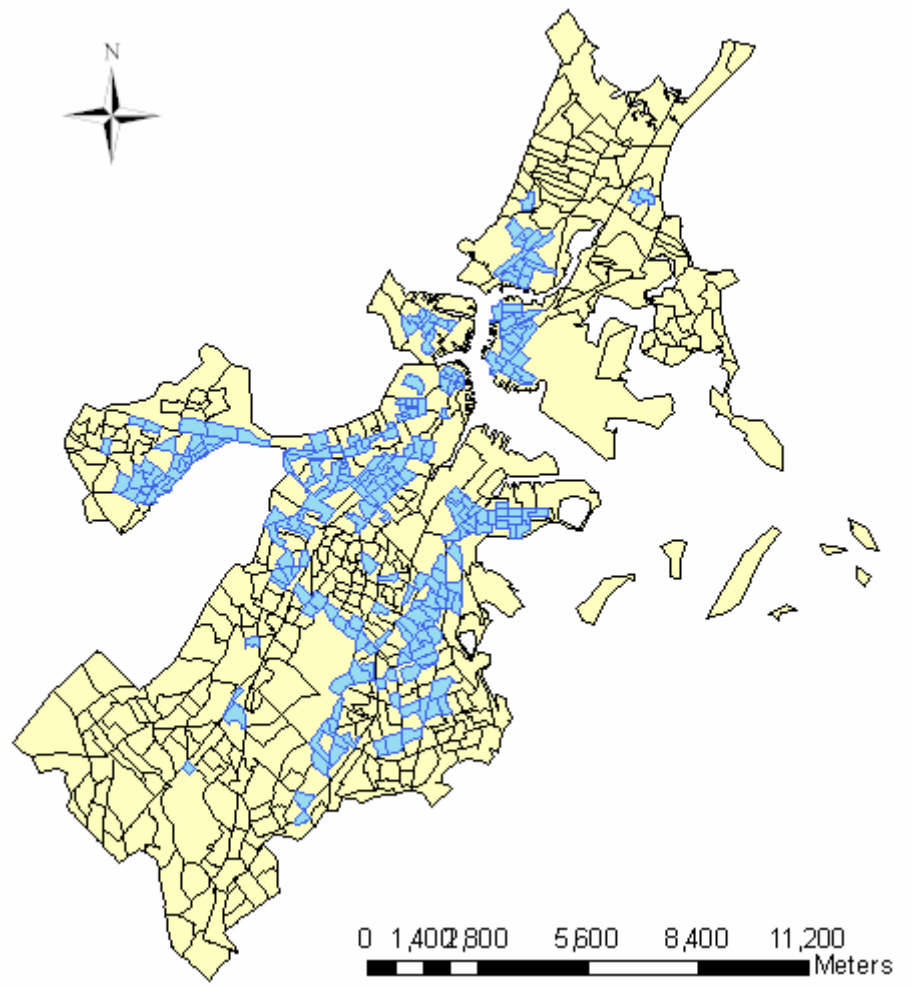


Figure 4: Suffolk County block groups with above average population density (shaded in blue). The average population density for the county was 8800 people per square kilometer, so the above map highlights block groups with density field greater than 8800.

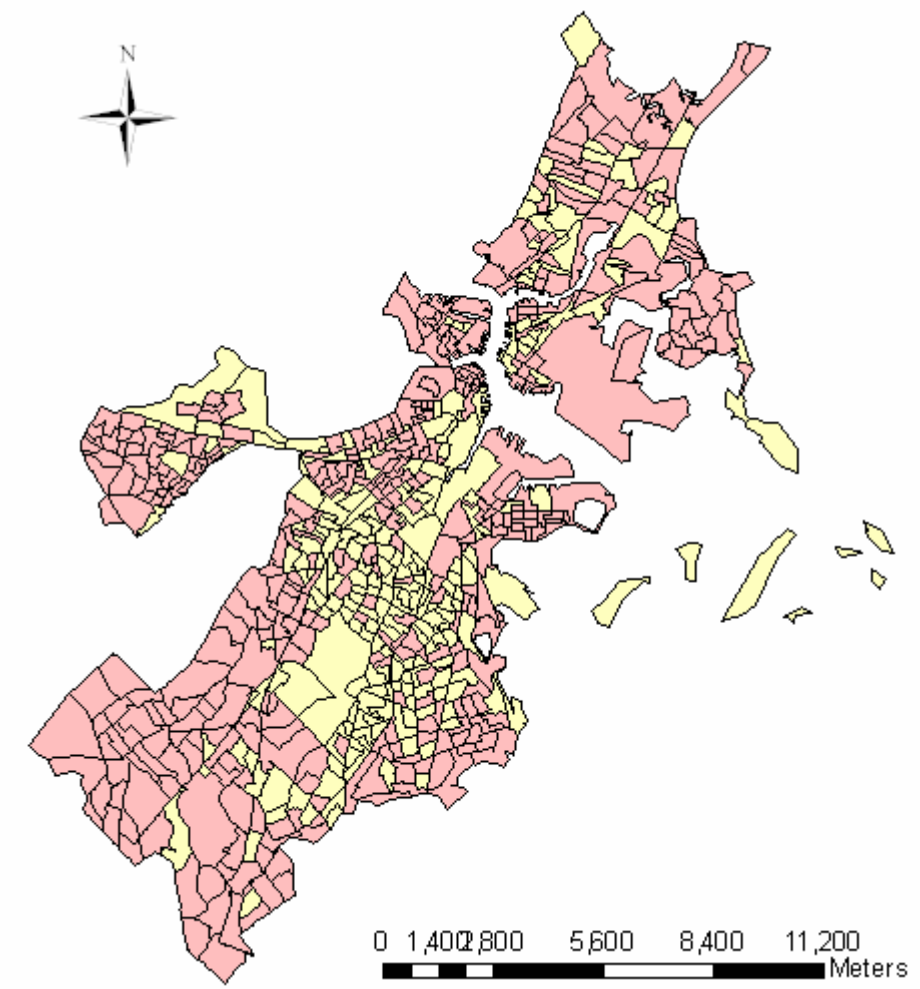


Figure 5: Suffolk County block groups with above-average employment (shaded in light red). Mean employment was 92.6%, so block groups with an employment rate greater than this value were included in the site suitability analysis.

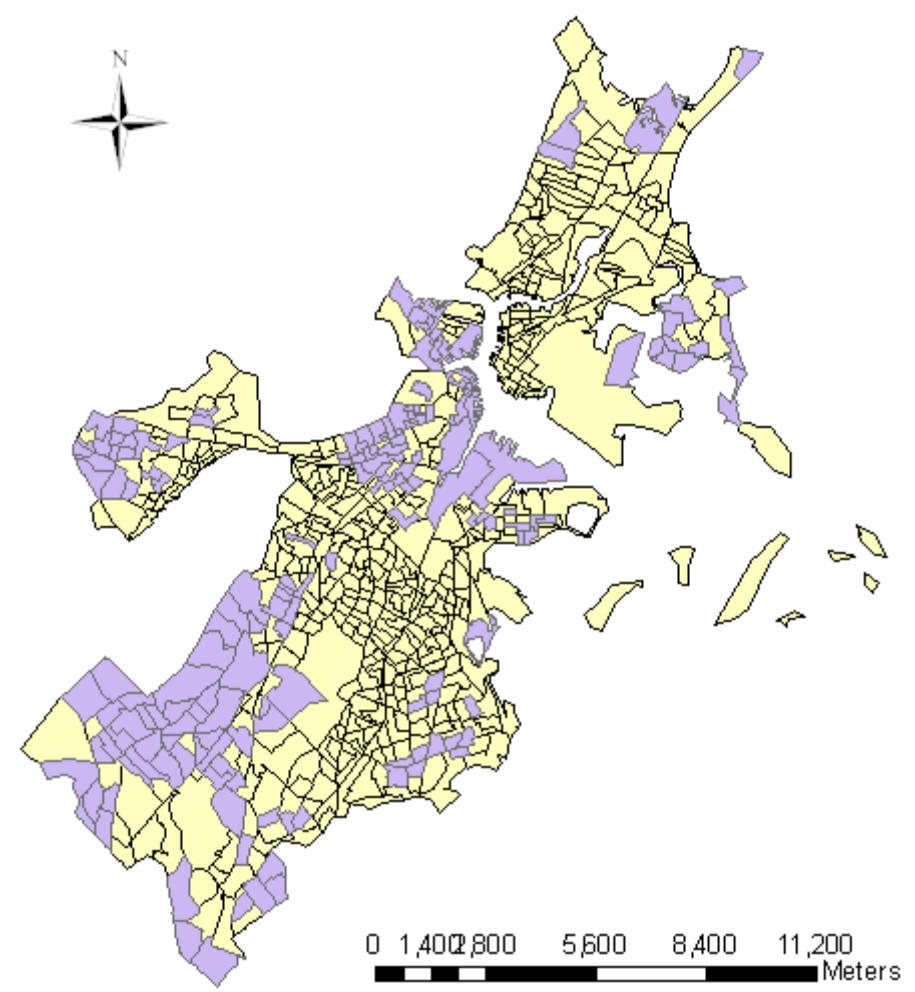


Figure 6: Suffolk County block groups with block groups with high median household income (shaded in purple). High is defined as greater than one standard deviation above the mean. The mean for the county was $41424 and the standard deviation was $8976.5, so the shaded block groups have a median household income of greater than ($41424 + 8976.5) = $50400.5

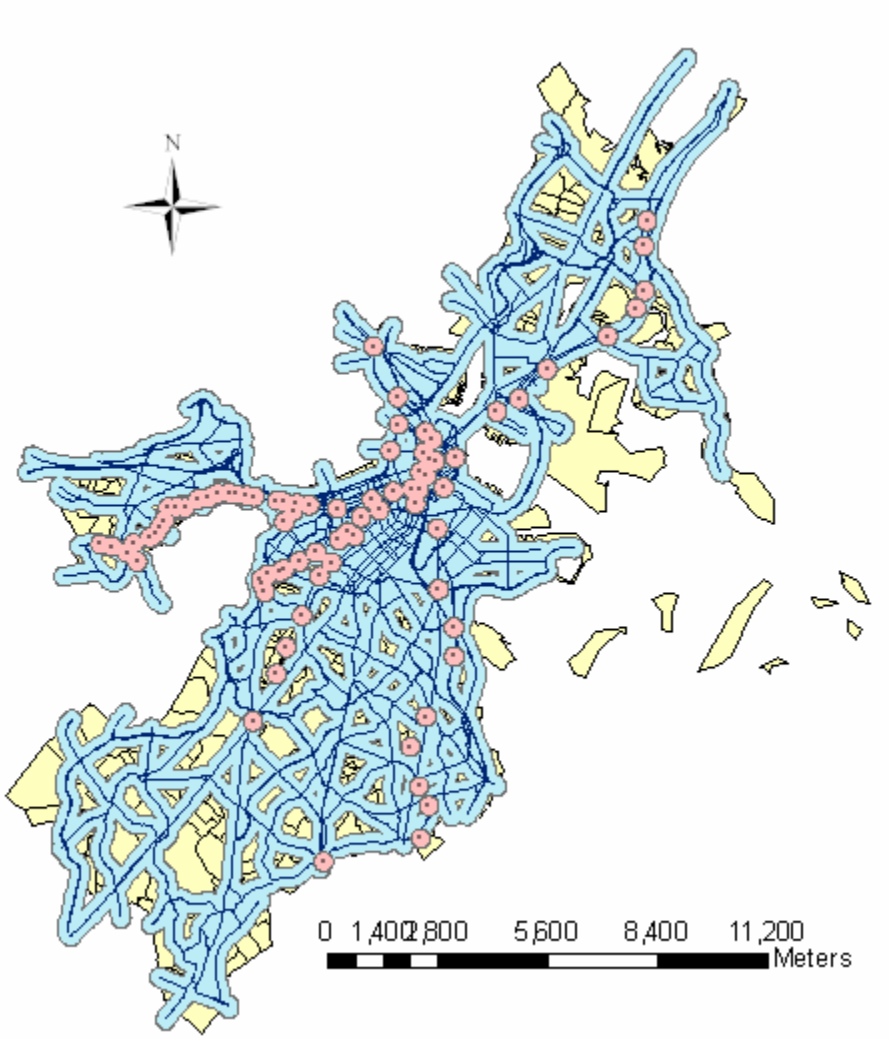
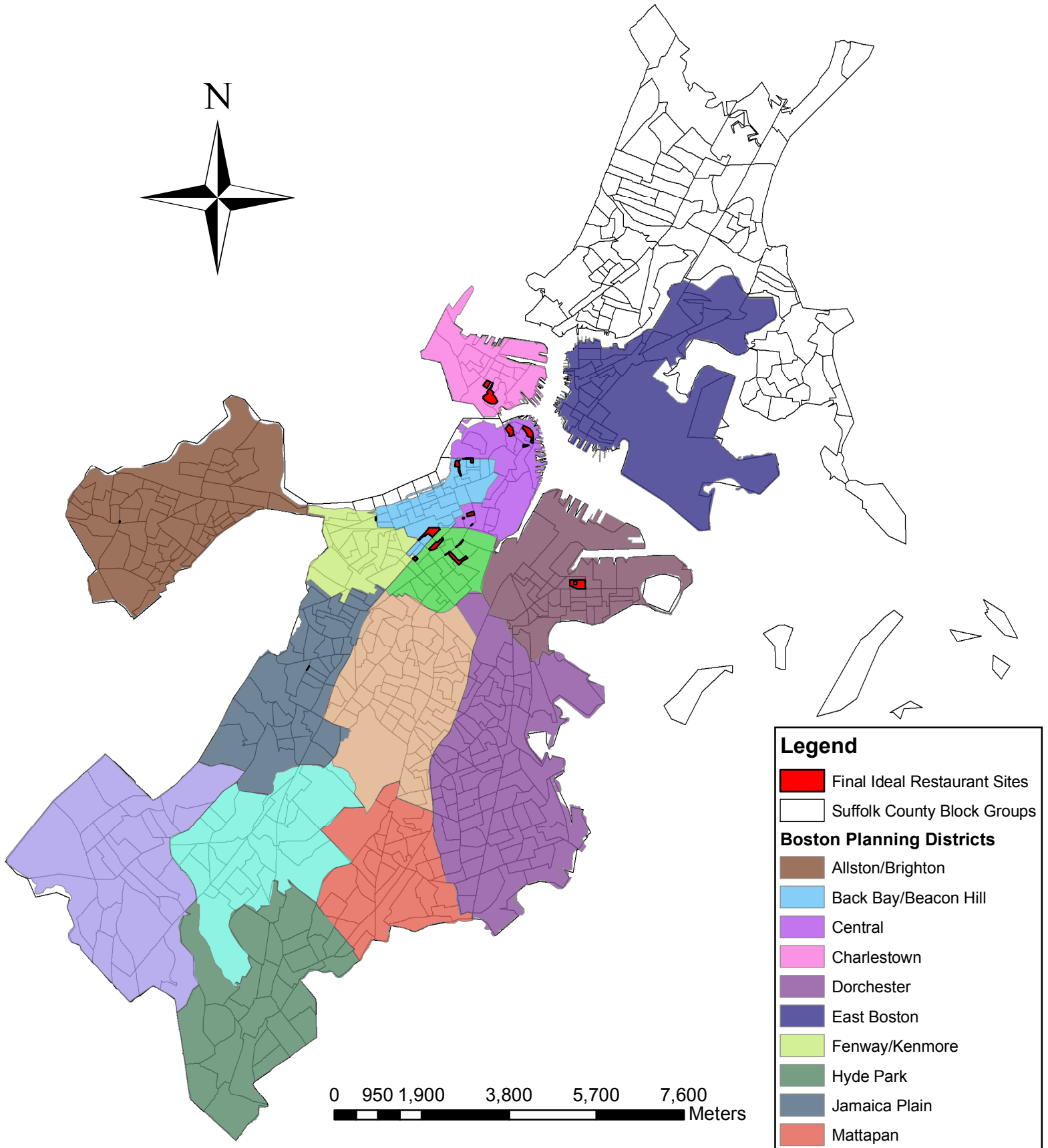


Figure 7: Transportation accessibility in Suffolk County. Major roads are shown in dark blue with a 250 meter buffer around these major roads shown in light blue. MBTA stations are shown as black points with the 250 meter buffer around MBTA stations shown in pink.

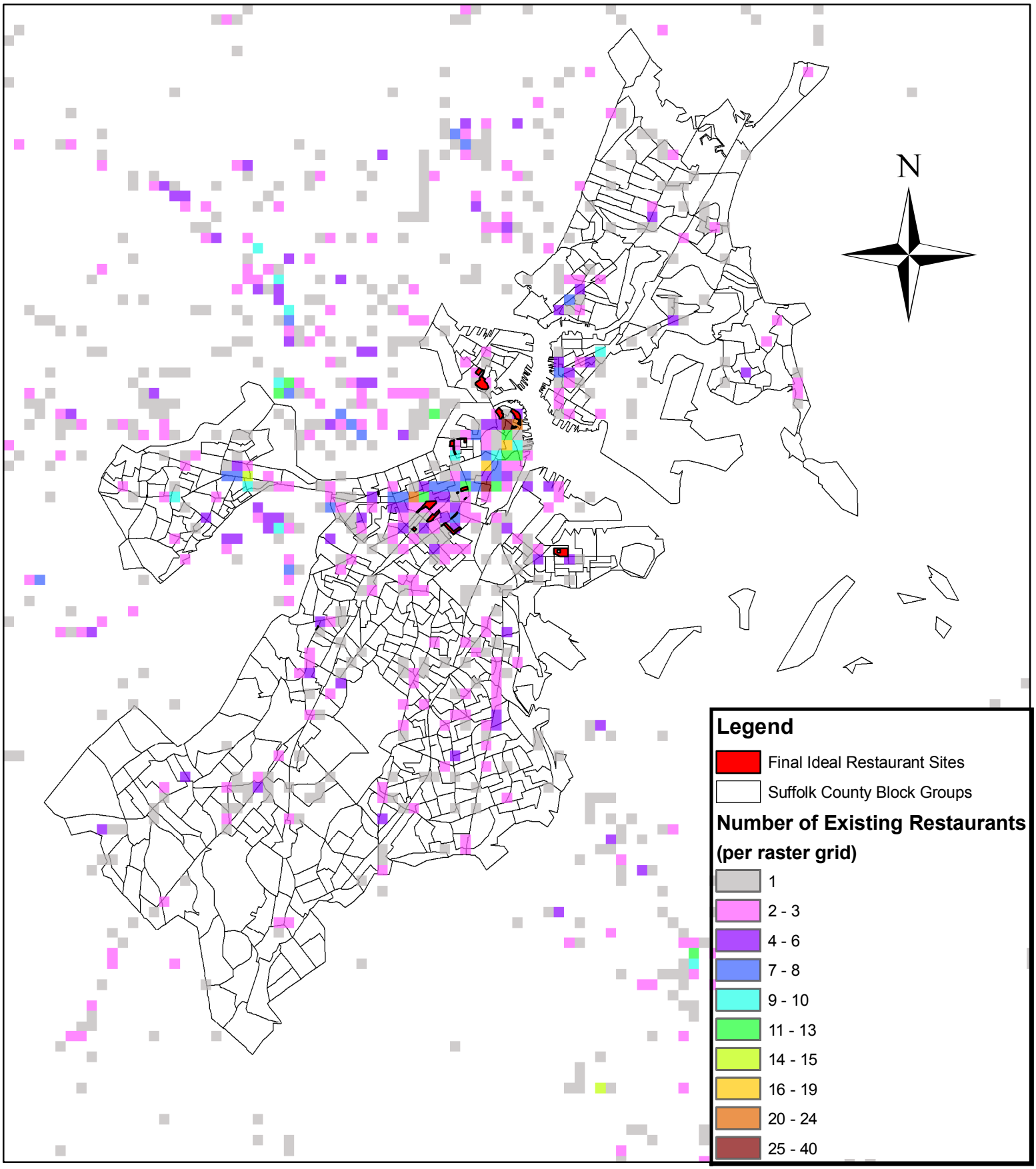
Map 1: Final Ideal Sites for a Suffolk County Restaurant over Boston Planning Districts



Sources:
US Bureau of the Census, 2000
MassGIS, 1998-2007
MIT GeoData Repository, 2000
MA State VMT Model Data, 2008

Praveen Subramani
(Athena: praveens)
December 10, 2008

Map 2: Final Ideal Sites for a Suffolk County Restaurant over Existing Restaurants Raster



Sources:
US Bureau of the Census, 2000
MassGIS, 1998-2007
MIT GeoData Repository, 2000
MA State VMT Model Data, 2008

Praveen Subramani
(Athena: praveens)
December 10, 2008