Office Locating / Target Marketing: Multiple Analyses for Determining Office Locations and Target Marketing

Justin Niebuhr [1,2]
[1] *Department of Resource Analysis, Saint Mary's University, Winona, MN 55987;*
[2] *H & R Block District Office, Roseville, MN 55113*

Keywords: Market Area Analysis, GIS, Geographic Information Systems, Geodemographics, Spatial Marketing, Target Market Area

Abstract

This paper examines the implementation of geographic information systems (GIS) in the targeting of a market and the sighting of office locations. Here, the locations of current clients were compared to the office they used; client density maps were made for each office to help determine location client densities; and multiple buffer analysis was done in support of the density map. Lastly travel distance was calculated. With these factors it was shown that office location couldn't be based on current clients. Target market areas were then defined to show areas where offices would have the most effect. This was accomplished through the use of a weighted model based on total number of returns, market percentage, age of filer, number of schedule C filed and competition. Increased numbers of clients would be expected from these target areas of the model. The placement of offices in these target areas would most likely increase the effectiveness of the office.

Introduction

In business, location is everything. The idea of location being everything creates a major use of GIS in the business world. GIS offers the ability to successfully determine a good location based on selection criteria. This project was initially undertaken as a way to determine new office locations or improve existing locations based solely on current client locations. After initial analysis, it was determined that many factors are involved in selecting a good location and that current client locations were insufficient. This project works with some of the factors used in selecting office locations. However, it was found that locating an office is very complex and is affected by many factors.

Background

According to the Internal Revenue Service (IRS), the state of Minnesota had 2,298,213 tax returns filed in the 1998 tax year. H & R Block serves 16.9 million clients throughout the US with about 1 in 7 IRS tax returns handled by H& R Block. There are 10,588 offices worldwide and about 80% of Americans live within 10 miles of an H& R Block office (H&R BLOCK, 2001).

This study concentrates on a district, which is made up of several offices that service clients throughout the Minneapolis St. Paul metro area of Minnesota. The district is comprised of eleven offices in the northeast corner of

the metro area with over 16,000 clients using the district (Figure 1).

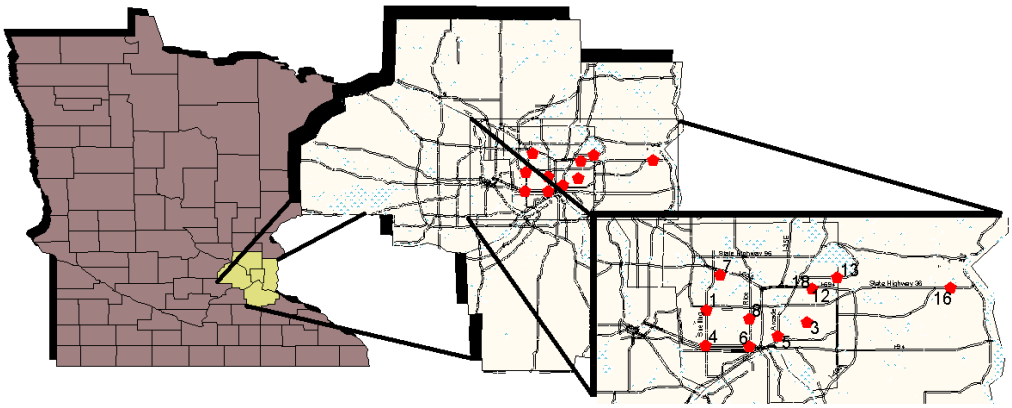

Figure 1. Study area in the state of Minnesota and the five counties that form it.

Data Collection / Assembly

The client data used were from the 1998 tax year and was obtained from the H & R Block district office. Due to the sensitivity of the data only the client addresses were used in this project. IRS demographic data by zip code was obtained from the IRS statistical information center for the 1998 tax year. Census files for 1999 were obtained for the United States Census bureau including the roads coverage. A polygon theme of approximate zip code boundaries was assembled based on the 1999 tiger census file road coverage using on-screen digitizing in ArcView.

Software / Extension

The project was completed in ArcView 3.2a software developed by Environmental Systems Research Institute Inc. (ESRI). The multiple analysis required extensions to ArcView including the Network Analyst, Spatial Analyst and Model Builder for Spatial Analyst, which are also developed by (ESRI). Other extensions, not developed by ESRI used in the completion of this project include Tiger Reader by MAPCLICK, X-Tools by Mike DeLaunne, Geoprocessing by Jeffrey Lane and Count Points in Poly by Yingming Zhou.

Methods

The initial steps in the project involved looking at the location of clients and then examining the relation of the clients to the office they patronize. This process was started with the geocoding of the data and then moved into a density analysis, multiple buffer analysis and travel distance analysis. The second step was to look at available tax related demographic data and how it could be modeled to find target client areas.

The conversion of address data to latitude and longitude coordinates required the use of geocoding. Geocoding is the use of a reference theme as a way to locate the location of an address. Address matching is a process that compares the reference theme and the address to determine whether they are the same (ESRI ArcView 3.2a Help file, 2000). Once a

match is made a point theme is made of locations of the matched addresses. If no match is found the point is not geocoded and thus has no coordinates.

Address geocoding was performed in ArcView using the 1999 tiger census file road coverage as the reference theme. The geocoding properties of the theme were set to US Street with Zone using the address and zip code fields. The client address data were initially cleaned to remove apartment numbers from the address field to provide more exact matches. The geocoding preferences of the ArcView geocoding process, including spelling sensitivity, minimum match score and minimum score to be considered a candidate were set to provide for the most matches with the least amount of errors. Spelling sensitivity has a default value of 80 if a higher number was used the less likely additional candidates would be retrieved. Minimum match score defines the score required to be considered a match from 0 - 100 it has a default value of 60. The minimum score to be considered a candidate with a default value of 30 ignores candidates that scored below the minimum score. The only change was a minimum match score which was changed from 60 to 50. The eleven H&R Block district office locations and other tax service providers (competitors) were also geocoded using the road coverage.

The goecoding of the clients provided for a point map of the clients. This excluded any clients not geocoded or falling outside the five county study area. From this point map, several analyses were performed including client density, multiple buffer and client distance traveled. These analyses were all performed using the geocoded client data, office locations, and the 1999 tiger census road file.

Initial Analysis

The first analysis was a density analysis to show concentrations of clients. The density analysis was accomplished using the ArcView Spatial Analyst extension using the Calculate Density function. This was done for the clients as a whole for the district and for each separate office. Histograms were also made to show the distribution of the different density levels.

The second step was to better show client locations compared to the office. A multiple buffer analysis was performed around each office. The multiple buffer analysis was developed for each office separately; this was accomplished by buffering each office at different distances. The distances were set at one and a half, three and five miles. The first two distances were chosen based on the fact that the H & R Block offices are all about three miles from the next closest office. The last distance was chosen to take in a larger area. The ring effect is completed by clipping out the center of the five-mile buffer with the three-mile buffer and the three-mile buffer with the mile and a half buffer. This provides a bulls-eye effect around each office. Next a point counting extension called count points with poly was used to count the number of the office's clients that were within each ring. Then the same extension was run to count the number of all of the offices clients in each ring around the offices. This was performed for each office. This provided for a comparison between possible and actual number of clients using an office based on distance. This is also known as cannibalism of the

clients by the offices in other words offices attracting clients away from neighboring offices. This could be caused by many factors for example transportation or advertisement. As another way to assess the ring analysis, the measurement of travel distance for each client of each office was found.

The Find Closest Facility function in the ArcView Network Analyst extension was then used to measure the distance the client traveled to an office. Normally, the Find Closest Facility function would define the office as the "facility" and the client as the "event". For this study, the function was manipulated so that the office was the "event" and the client was the "facility". This process required entering the maximum number of clients geocoded for that office as the number of facilities to find. By doing it this way, the manipulation measured every client's distance with out missing any clients. The distance traveled by the clients and the directions of travel were extracted using the text file option that is included with in the extension. The text file was then cleaned to remove the direction information and leave the total distance traveled for each customer. The text file was imported into Excel as a tab delineated file. Once in Excel an average distance traveled for each customer was calculated for each office and the district.

Extended Analysis

After looking at the current customer location data the next step was to look at potential customer data. This was executed with the use of IRS tax demographic data and the functions of ESRI's Model Builder extension for Spatial Analyst. Some initial analysis was performed in preparation for the use of the model builder extension. This analysis also provided some very useful information on current customer patterns.

Initial Model Analysis

The clients were then examined by area groups. This was accomplished by using zip code boundaries to group the clients. The count point with poly extension was used to find the number of clients per zip code for each office and the district as a whole. Number of clients per zip code was calculated as another way to show client distribution and to be used to calculate market percentage. Market percentage was calculated by using the total number of returns filed by zip code and the number of clients per zip code. The data used were IRS data for the 1998 tax year based on zip code and the number of clients per zip code.

The next step was to take the geocoded competitor location data and use the service area function of ESRI's Network Analyst extension. This was to provide a way to convey an area that the competitor is servicing. This was done opposite of the ring analysis with the use of a service area instead of a buffer. The service area distance was set for 4.34 miles based on the predetermined average travel distance of H & R Block clients. This provided a way of showing an approximated service area of the competitors based on the known travel habits of the tax customers.

The weighted model was developed at this point to bring factors together to show areas of interest in the targeting of clients. Many factors could be included in the determination including total number of returns filed, current market percentage, age of the

filer, number of schedule C forms filed by zip code, average gross income or number of competitors. This was performed at the zip code level using the IRS data for the 1998 tax year.

The Model Builder extension in ArcView was used in the development of the model. Model Builder is a raster-based environment working with Spatial Analyst. The extension works by first designing a layout of the model. The design of the model for this project used five factors: total number of returns, market percentage, age of client, number of schedule C filed, and the areas of competition predetermined at 4.34 miles. These factors were determined in consultation with the district manager. The determination of the weighted values was also done in conjunction with the district manager. With the use of the Spatial Analyst extension the data first need to be converted from vector to raster format. In the conversion to raster the data were classified into discrete categories or classes. For this project each data set was classified into ten classes. The use of ten classes was used at every step of the model.

The initial age data were broken down into four separate age groups. To narrow the age factor down into one theme the separate themes were given a weighted ranking and a new theme was developed. To rank the separate age group themes, a higher rank was given to the younger aged groups. Then within the age groups higher overall numbers were given a higher rank. This weighting of the age themes provided for the determining of prime age areas by including all age groups and their number in the model instead of just choosing one age group as a factor in the model.

The narrowing of the age to one theme brings the model to the five main factors. The five factors were then given a weighted ranking with total number of returns weighted the highest at 50%, market percentage was ranked at 22%, age of filers 14%, number of Schedule C 7% and areas of competition at 7%. Within each weighted factor the ten classifications of the themes were ranked separately. For example, areas with high numbers were assigned ten; areas with low numbers were assigned a negative ten. This, however, was opposite for the competitor areas where a class showing a high concentration of competitors was given a negative 10 and low concentrations were given a positive ten. This model and the ranking scheme were developed with the district manager to incorporate age of filer, number of schedule C, market percentage and competition besides the total number of returns filed in the determination of areas of interest. The output was also put into ten classes to provide for a more graduated view.

Results / Discussion

The geocoding of the client data provided for an average of 82% of the clients matched or geocoded. The predetermined acceptable number of clients or customers geocoded was set at 80% (McMullin, Year Unknown). The average was however 82% with a low of 63% and a high of 89% (Table 1). These numbers however can be explained due to several things, one being the use of P.O. boxes as addresses by clients, addresses that fall outside of the study area, and none matching addresses. Another factor in the geocoding of clients is office 16 located on the Minnesota side of the Minnesota /

Table 1. Geocoding statistics for each office based on the criteria of Arc View and totals.

Office Number	Good Match	%Good Match	Partial Match	%Partial Match	No Match	%Not Matched	Total
1	2310	67.56	570	16.67	539	15.76	3419
3	778	67.30	210	18.17	168	14.53	1156
4	1918	74.03	348	13.43	325	12.54	2591
5	1234	69.84	331	18.73	202	11.43	1767
6	964	70.16	230	16.74	180	13.10	1374
7	564	67.06	118	14.03	159	18.91	841
8	552	69.43	123	15.47	120	15.09	795
12	1135	66.61	213	12.50	356	20.89	1704
13	216	65.65	43	13.07	70	21.28	329
16	660	44.96	255	17.37	553	37.67	1468
18	823	65.47	153	12.17	281	22.35	1257
Totals	11154	66.79	2594	15.53	2953	17.68	16701

Wisconsin border and used by a fairly large number of Wisconsin residents. Wisconsin clients were not geocoded and this lowered the percentage geocoded for that office. Overall, of the total 16,701 clients for the district, 13,748 clients were geocoded.

The goecoding of the clients provided a point map of the clients (Figure 2). This excluded any clients not geocoded or falling outside the five county area, as previously stated. From this point map, client density, distances rings and client distance traveled analysis were performed with relative ease.

Initial Analysis Results

The analysis of client density with the Spatial Analyst provided for areas of client density for each of the offices and the district. The Spatial Analyst gives a grid as an output with the cells showing

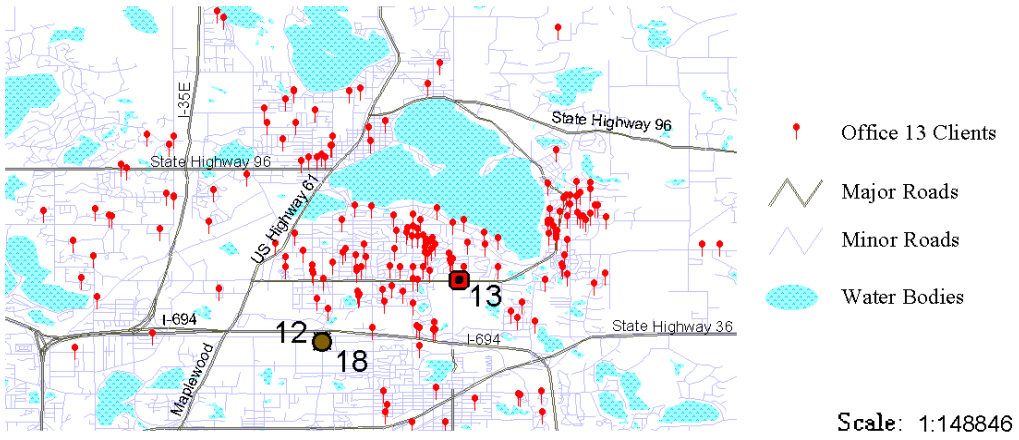

Figure 2. Geocoding point map for office 13 done in ArcView

the density of an area. The density analysis shows where the client concentrations were located. This helped to show general locations of the majority of the clients for the district and for each separate office. The analysis revealed that the patches of density did not always center on the location of the office that the clients used. However, a beneficial location is defined by more than simply where the location of the clients concentration. Several of the offices did not have well defined areas of density, especially offices seven, 18 and 12 which showed a variable distribution of their client density (Figure 3). The absence of a smooth curve in the histogram supports the variable distribution of the clients. In contrast offices such as five, display a very well defined concentration, which presented a very smooth curve on the histogram (Figure 4). One reason behind this may be the difference between the downtown area and the more suburban areas of the metro area, which is somewhat evident between the offices. Office four has a notable offset of client density from the location of the office (Figure 5). This office is also located in a downtown area.

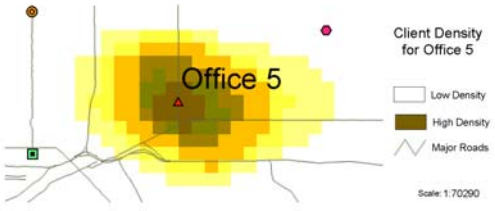

Figure 4. Client density for office five

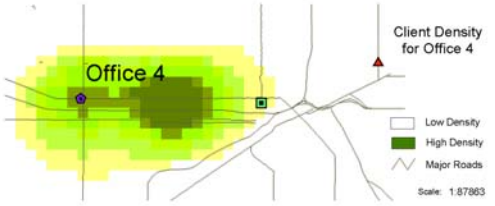

Figure 5. Client density for office four

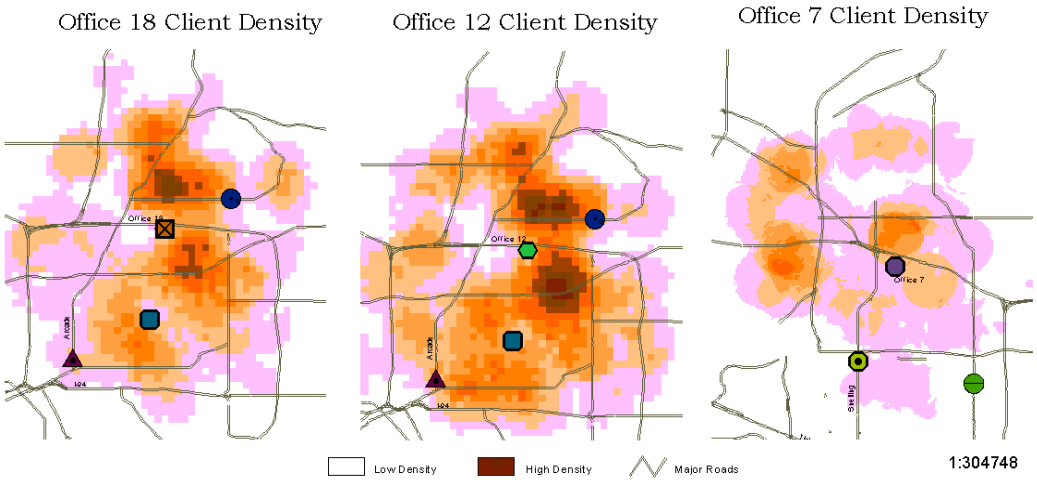

Figure 3. Client density for offices seven, 12 and 18 done in Spatial Analyst

The ring analysis provided for a lot of collaborating information to support the density analysis about each office even though both analyses are based on different principles. Service areas could have been made at each distance for this analysis instead of the buffer pattern. However, this was not done to keep the area of the rings the

same size for each office. The ring analysis also showed many differences between the offices, especially in the distance clients were from the offices. The distance the offices were drawing clients from varied from office to office. For example, office three had 333 clients within 1.5 miles, 344 clients from 1.5 to three miles and 151 clients from three to five miles. In contrast, office 18 had 83 within 1.5 miles, 218 clients from 1.5 to three miles and 353 clients from three to five miles. Office three had a total of 1,156 geocoded clients and office 18 had a total of 1,257 geocoded clients. This shows office three's clients were close to the office, while office 18's clients traveled a greater distance. The integration of the district client data into the ring analysis brought about very important information about the office. The ring analysis showed large gaps between the actual and possible number of clients based on distance from the office (Figure 6). On average at one and a half miles from any given office 280 clients are being cannibalized by another office. However, this was variable depending on the office, with a low of 28 at office 16 and a high of 619 at office six. The low cannibalization at office 16 is to be expected based on its location.

The travel distance analysis done using the Network Analyst produced results that complimented the previously determined information of the other analyses. The travel distance analysis determined that on average client traveled a distance of 4.34 miles to any given office (Figure 7). This was also dependent on the office, with office seven's clients showing an average travel distance of 6.28 miles and office five's clients only traveling 2.05 miles on average. The example previously given between office three and 18 was also supported in this analysis by an increase of almost two miles in travel distance between the two offices.

Final Analysis Results

At this point the project turns from the first focus of the client to the identifying of potential clients with the IRS demographic data. The initial analysis of number of clients per zip code provided a way of showing a gradient of client numbers and their locations in a better

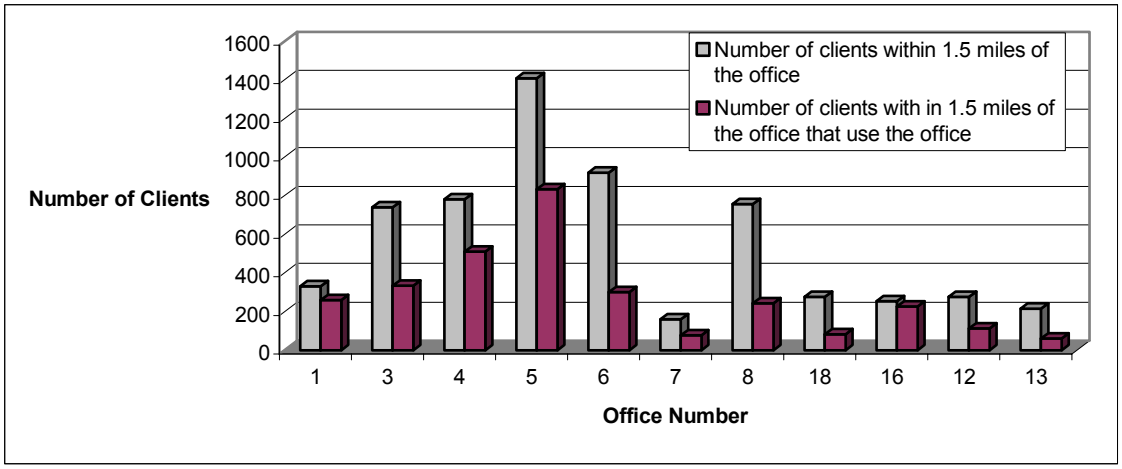

Figure 6. Comparison of number of clients using office within 1.5 miles to the total number of clients within 1.5 miles.

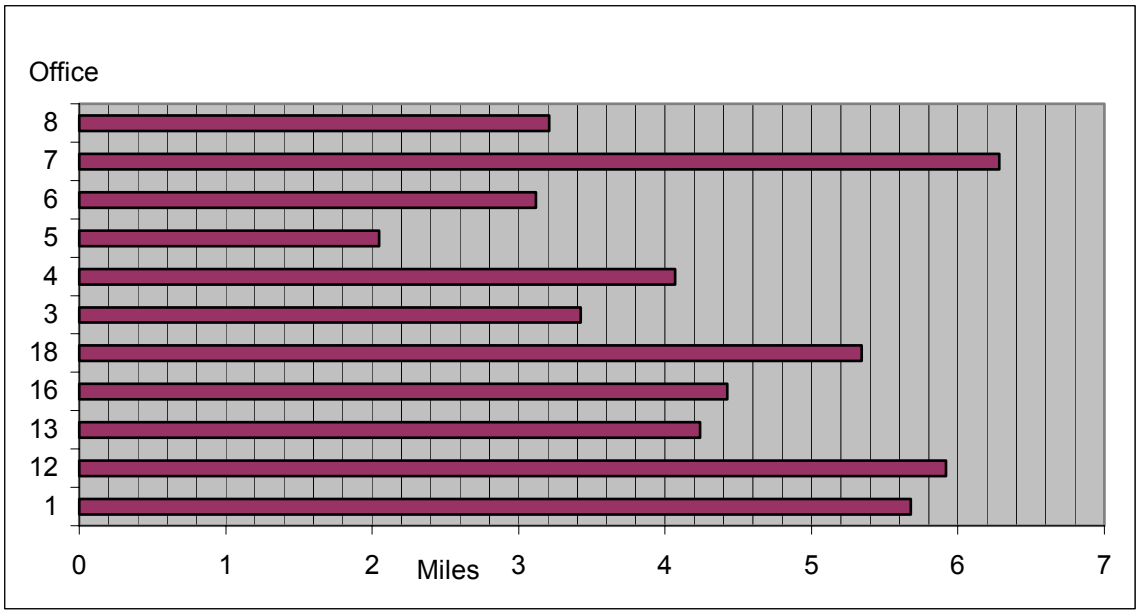

Figure 7. Average distance traveled by clients to an office.

way than the point map. With the number of clients per zip code calculated the next step was to calculate market percentage. The calculation of market percentage provided information that was not previously available by just looking at the clients (Figure 8)**.** Market percentage is also highly beneficial information because it shows the effect the business is having in certain zip codes. Market percentage helped in making some initial predictions about potential client target areas although this does not figure in the actual total number possible, which leads into the use of a model.

The mapping of competition showed areas of higher and lower impact of competition. The use of the service area instead of the buffer was based mainly on the known average travel distance of clients. The using of a service area makes it harder to visualize, but is more useful in the model than just a point location. Based on early findings of willingness to travel, this may not be a factor that should be considered in the model. It was included though, but was weighted low as to not be of much effect on the results of the model. The

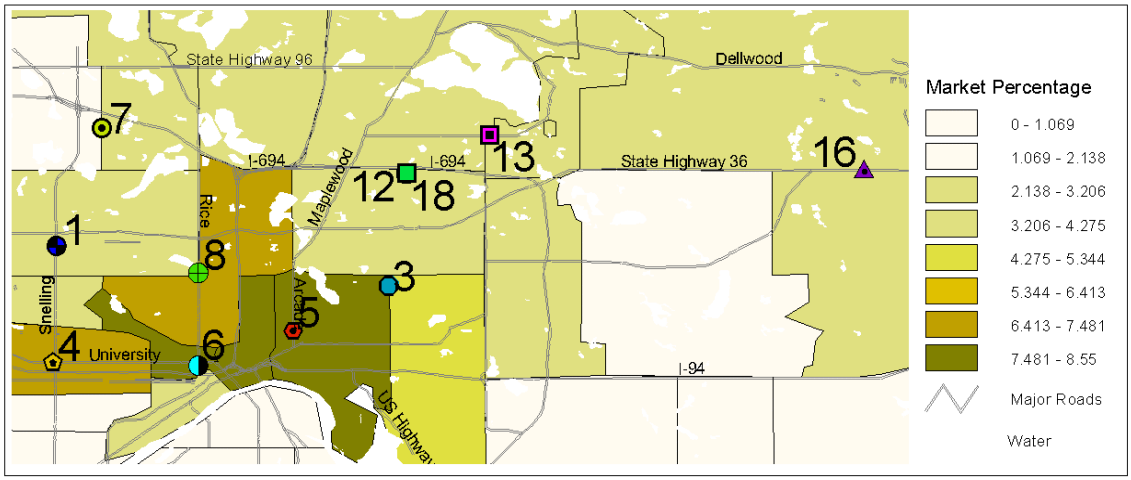

Figure 8. H & R Block district market percentage for the 1998 tax season.

competition data represented as service areas help to show density with areas of higher and lower competition. If competition was decided to be a larger factor, its weight in the model could easily be increased.

The model produced a compiled age group theme, which was made up of four original age groups. The model produced grid coverages of each factor to allow it to perform the weighted analysis (Figure 9). The model produced a final resulting weighted map showing the combination of the five factors at their set weighted percentage. This map represents areas where the specific clients would be at a higher percentage (Figure 10). The values of the map represent the ranking of the factors involved. If the movement of an office into an area improves access to clients of an area the areas of higher client potential would be areas to move offices to.

geocoded clients within three miles of any one office and 42 percent within one and half miles of any one office. These factors made the use of current client location as a factor in determining good or better office locations not valid. Another thing to consider is that the majority of the offices are about three miles apart.

The initial analysis determined that current clients could not be used as a factor in the placement of offices. The use of demographics of taxpayers was the second factor studied. In this analysis factors including age, number of returns filed and number of schedule C filed were used as another way in helping to determine good or better office locations. Also included in the analysis was the market percentage based on the current clients. The inclusion of market percentage provided a way to incorporate current clients. These two analyses together found out information

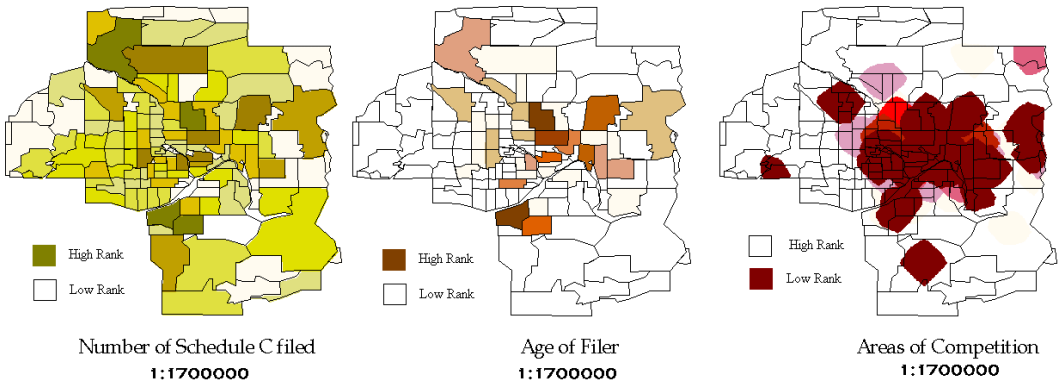

Figure 9. Three of the factors included in the model number of Schedule C filed, Age of filer and Areas of Competition.

The initial analysis found that clients were traveling more than four miles on average to any one office, with clients traveling more than four miles to an office and with over 75 percent of the

on the travel habits of the customer and then put that information to use in determining good and better office locations.

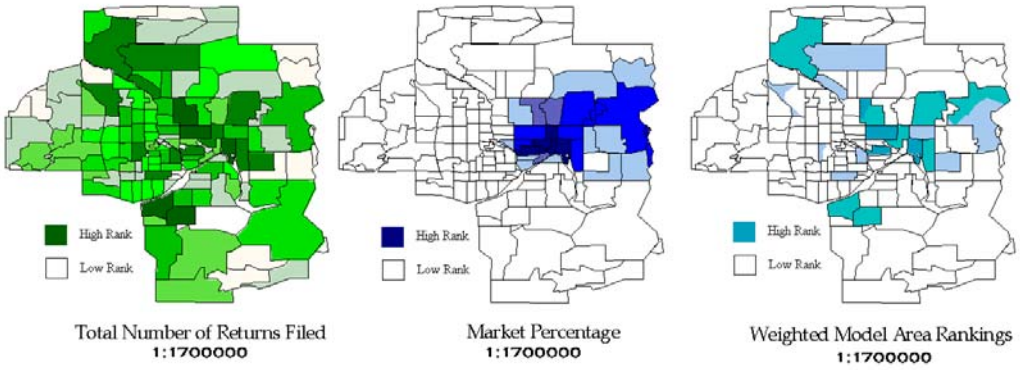

Figure 10. The two major factors of the model total number of returns filed and market percentage. The output area rankings of the model.

Further Discussion

This project did not come to a decisive conclusion about office locations, which may not be possible based on geographic related information. It did however provide a lot of very useful derived information about potential and current clients. This information could be used in many further studies about clients such as a temporal study of client movement and density pattern changes.

Subjects/Factors that could have been performed or included in this study depend on what factors are considered important. A customer survey could be done as a way to determine the reasoning behind the client's choice of an office. Many factors could be looked at in a study like this including demographics, zoning, and transportation. For example, did the clients take the bus to get to the office? Were the clients in the area for other reasons such as shopping, or do they just prefer a certain preparer?

The use of the model as a way to determine good or better office location has some merits; it includes current client as market percentage and it finds areas of high numbers of possible clients. However, it also takes into account competition, which has little effect when a business is the leader in the field. The inclusion of schedule C in the model narrows the type of clients. It may be better to run it without the schedule C factor, then figure in the schedule C factor in the training of personnel at office in areas of high numbers of schedule C forms.

Conclusion

This analysis was unable to provide any recommendations about office locations based on current clients because of the cannibalism between offices, the willingness of clients to travel and the many other factors affecting their selection of an office. Some recommendations could be made about office location based on the output of the model and the possible clients. The use of these types of analysis in a study such as this provides a lot of useful information. However, the many factors involved make it difficult to make a substantial conclusion.

Acknowledgments

I would like to thank H & R Block for the opportunity to demonstrate the possibilities of GIS in the mapping of their clients for a district. I would also like to thank the district manager for insight in the development of the model. I would like to pass an appreciative thank you to Saint Mary's University for the use of their facilities and equipment. I would also like at this time to thank my peers in the Resource Analysis program at Saint Mary's University for any insight and support given. A special thanks to the people who developed the software capabilities used to perform the analysis in this project.

References

ESRI ArcView 3.2a Help file. 2000. What is geocoding? Environmental Systems Research Institute Inc. (ESRI) Redlands CA

H & R Block. 2001. Company Information. http://www.hrblock.com/about/company_info/index.html

H & R Block. 2001. Corporate Facts. http://www.hrblock.com/press_relations/press-facts.html

Internal Revenue Service, Information Services, Martinsburg Computing Center, Development Center Branch. Unpublished Data. http://ftp.fedworld.gov/pub/irs-soi/98in24mn.xls.

Internal Revenue Service, Statistics of Income Division. http://ftp.fedworld.gov/pub/irs-soi/91zp24mn.xls.

Longley, Paul and Clarke, Graham. 1995. GIS for Business and Service Planning. John Wiley & Sons Inc New York.

McMullin, Shaun K. Year Unknown. Where are your Customers? Raster Based Modeling for Customer Prospecting. Graduate Thesis University of Washington Seattle, WA.

Mullen, Kevin and Kinsley, Michael. District Office of Research and Analysis and Electronic Tax Administration Connecticut-Rhode Island District Internal Revenue Service, Statistics of Income Division. e-file demographics. http://ftp.fedworld.gov/pub/irs-eta/mn.xls.

US Census Bureau. 1999. Topologically Integrated Geographic Encoding and Referencing system (TIGER) /Line Files for Minnesota. http://www.census.gov/geo/www/tiger/rd_2ktiger/MN/.

Yahoo Yellow Pages. 2001. Tax Service Providers. http://yp.yahoo.com/.