An Analysis of the Change in Student Body Geographic Distribution Resulting From a Relocation of the School Campus

Benjamin Schlawin^{1,2}

¹Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, Minnesota 55987; ²Fox Valley Lutheran High School, Appleton, Wisconsin 54915

Keywords: Geographic Information Science (GIS), Geocode, Zip Code, Distance Analysis, Education, and Relocation.

Abstract

This study involved the analysis of ten years of graduates from Fox Valley Lutheran High School in Appleton, Wisconsin. The study focus was to determine how the spatial distribution of the graduates was affected by the relocation of the school campus. The spatial and statistical analyses performed on the data revealed no significant differences between the graduates from the two sites that could be attributed to the relocation.

Introduction

The objective of this study was to determine if the relocation of Fox Valley Lutheran High School to a new location had an effect on the geographic distribution of its student body. This study tested the hypothesis that there would be no change in the student body geographic distribution due to the school's relocation.

The key data set for this study included a listing of the address and graduation year for all graduates from 1996 to 2005. The data set was divided into two subsets, the five-year period prior to the relocation, and the five-year period after the relocation.

Background

Fox Valley Lutheran High School is a parochial high school, and as a result is not confined by political boundaries as public schools are by their district lines. Other factors limit the area from which this school can draw students. These include commute distance/time, faith, commitment, and tuition.

The school in many ways

operates like a public school district. It has a federation of congregations that supply students and monetary support. Many of these congregations operate grade schools that serve as a feeder system for the high school.

Methods

Technology

The Geographic Information Science (GIS) software packages used for this study included ArcGIS 9.0, ArcGIS 9.1 and ArcView 3.3 along with the appropriate extensions. The Microsoft Office suite of programs was used for data conversion, documentation, and communication. SPSS 11.0 was used for statistical analysis. The Internet was critical for research, data transfer and acquisition, and e-mail communication.

Projection

The source data sets came in a variety of spatial projections and geographic coordinate systems. Each data set was converted to the following projection before it was used. The consistent use of this coordinate system assured accurate overlay of map components and consistent analysis results with the GIS software.

Projected Coordinate System: NAD_1983_UTM_Zone_16N Projection: Transverse_Mercator False_Easting: 500000.00000000 False_Northing: 0.00000000 Central_Meridian: -87.00000000 Central_Meridian: -87.00000000 Scale_Factor: 0.99960000 Latitude_Of_Origin: 0.00000000 Linear Unit: Meter (1.000000) Geographic Coordinate System: GCS_North_American_1983 Datum: D_North_American_1983 Prime Meridian: 0

Data

Sources

The following sources provided the base data sets that were used during the study.

Fox Valley Lutheran High School Graduate addresses Graduate's year of graduation Student's parish/congregation Parish address

Census Bureau Tiger Line Data files Zip code coverage files

ESRI Data Files 2004 road data files Zip code point file

Limitations

Study Area

The data sets for this study were confined to a six-county area surrounding the two school sites. The six counties involved in this study were Brown, Calumet, Outagamie, Shawano, Waupaca, and Winnebago located in northeast Wisconsin (Figure 1).



Figure 1. Locator map for the study area showing the six-county study area and the new high school – red star.

The counties were chosen as a convenient geographic reference for limiting the study area. They contained the vast majority of the student body, and provided easy access to road, zip code, and census data. This allowed for easier analysis by automatically eliminating students outside a reasonable commute distance such as foreign students and students from Florida. California, and Texas. The number of students eliminated because of these constraints was not significant (17 students of 1777 students). Eliminating these students prevented the skewing of analysis involving distances, centroids, and geographic means.

Confining the study area to the six-county area simplified the questions of map projection, data analysis, and data correlation. This also eliminated questions such as "What was the local address for students who were not able to commute on a daily basis?" for which data was unavailable.

Geocode/Zip Code

The study area contains a significant amount of rural area, which presented problems for the geocoding process. It had inconsistent zip code polygons, alternate road names, and a unique addressing system.

Zip codes are designed as linear routes for mail delivery. Polygons are frequently drawn around the area serviced by the linear routes for the same zip code (U.S. Census Bureau, 2005). The conversion from line to area works well in metropolitan areas, but not as well in rural areas. Much of the study area is rural with the metropolitan areas rapidly expanding into it (American Fact Finder, 2005). The result is that delivery routes of one zip code intrude into polygons of a different zip code. This anomaly prevented several addresses from being matched to an address location by the software.

The rapid expansion of the urban area near the school has led to the construction of new roads and the expansion of existing roads. This required the newest road database available for complete geocoding of addresses. The accurate updating of these road databases is always an issue (U.S. Postal Service, 2005). To further complicate matters rural roads in this area often have alternate names. These names were not always available as part of the road database. The Mapquest website was used to explore alternate road names and locate cities (Mapquest, 2005).

Eastern Wisconsin uses a grid system for some of its addresses. An example of one of these grid addresses would be W123456 N12 Anystreet. The road database does not always account for this. Addresses in this area are often a mix of grid addresses and conventional addresses on the same road. Other addresses are a combination of the grid system and conventional addresses like N1234 Anystreet. These issues accounted for a large number of the unmatched addresses. Data Normalization

The normalization of the student database with the road database was a problem. One example of this was the mismatch of road names like County Hwy C as opposed to Cty Rd C.

Another issue was the inclusion of student addresses in the database that were unable to be geocoded. These included addresses outside the study area (17), PO boxes (25), and no address (13).

Geocoding Results

The geocoding software was not able to match 672 (38%) addresses of the 1777 total addresses. Manual matching matched 461 (69%) of the 672 unmatched addresses. This left 211 (12%) addresses unable to be geocoded (Table 1). The breakdown of geocoding results by year shows no single year had an inordinate number of addresses unmatched (Table 2).

Table 1. Geocoding of student addresses results.

Total Number of Students	1777
Software Matched Addresses	1105
Manually Matched Addresses	461
Unmatched Addresses	211

A systematic approach was used for the manual matching of addresses. First, the zip code was ignored, second the street number was ignored, and finally an alternate name/spelling for the street was explored. When these steps produced a list of choices, the choices were highlighted on the map. The city portion of the address along with the address style was used to determine the most reasonable choice. If no reasonable choice was ascertained the address was left unmatched. A sampling of the manually matched addresses was submitted to the school for verification. No errors in address placement were reported.

Year	Unmatched	Total	%
	Student	Graduates	Un-
	Addresses		Matched
1996	18	175	10
1997	21	159	13
1998	12	116	10
1999	20	207	9.6
2000	29	204	14
2001	23	178	12.9
2002	24	184	13
2003	27	180	15
2004	23	200	8
2005	14	174	12

Table 2. Geocoding results by graduation year.

Manipulations of Data

The original student body data set and congregation data set were obtained in Excel format. The Excel format data sets were converted to dbf IV (Database IV) format. This conversion allowed the GIS software to import the data sets for use in the geocoding process and to integrate the data sets into a shapefile.

The geocoding process performed with ArcGIS 9.0 plots points based on the address field from the student data set correlated with the address fields from the road file. The software estimates the location of the address along a road segment and places a point there. The shapefile that results shows the approximate geographic location of the student addresses. Much of the analyses performed for this project were based on this shapefile.

The results of various spatial queries were entered into an Excel spreadsheet and SPSS 11.0. Much of the statistical analyses and graphical output were generated using SPSS 11.0.

Quality Control

The shapefiles that resulted from the

geocoding process were used to produce several maps with different student subsets selected. One subset was manually matched students while another subset included students whose addresses were geocoded by the software. The resulting maps of student address locations were checked for location accuracy. No incorrectly matched addresses were found.

The conflicting and missing data in the student Excel spreadsheet and congregation Excel spreadsheet were edited to improve the matching results.

Analysis

Zip Code

The student data set was first explored by geocoding the students to their zip code. This was done to verify the extent of the study area and to test the appropriateness of limiting the study area to the six-county area. Zip code polygons were obtained from the census bureau. Figure 2 shows the zip code polygons for the six-county area with the six-county area overlaid.



Figure 2. Zip code polygon file for the sixcounty study area.

Geocoding only the zip code places a single point for each address at the post office for that zip code. Figure 3 illustrates the result of this stacking by showing only one point for each zip code that has a student address regardless of the number of student addresses found with that zip code. This analysis showed 1729 of the 1777 students in the study had an address with a zip code located within the six-county area. This affirmed that the study area could be limited to the six-county area without significant loss of student data. It also coincided with distance parameters that would allow for a daily commute.



Figure 3. The zip code polygon file for the sixcounty study area with red spots showing the student body geocoded by zip code.

Mean Center

The mean center is the geographic center for all the points in the data set being analyzed. Figure 4 shows the location of the old school as a blue star and the new school as a red star. The corresponding colored circles represent the geographic mean center for the graduates from each school site. The distance between the mean centers for the two schools is 0.18 miles while the actual distance between the two schools is 2.2 miles. The farthest graduate for either school fell within a forty mile buffer ring. The mean center plot suggests no appreciable shift in the two subsets of students.



Figure 4. Geographic mean for the graduates of each school. Red star is the new school while the blue star is the old school with the corresponding circles representing the mean center.

Central Feature

The central feature chooses the most central point from a set of points based on Euclidean distance. The point subsets used were the geocoded addresses for the graduates from each school location. The distance between the central points for the two sets was 0.15 miles. The central points were added to Figure 4 to create Figure 5. The blue triangle in Figure 5 represents the central most student from the old school site. The red triangle is the central most student graduating from the new site. More variation was shown between the central student plots for any two graduating classes from a site than between the overall combined data. The central point analysis was consistent with the mean center analysis, neither suggesting an appreciable difference between the two subsets.

Quadrant

The quadrant analysis divided the sixcounty area into quadrants using the new school location as the center point for the quadrants. The axes were oriented North-South and East-West (Figure 6). This orientation of the axes divides Lake Winnebago to the south of the schools in half. It prevents any quadrant from having an excessive area that cannot be populated. It also means three of the four quadrants have a major lake in them.



Figure 5. Central feature placed on Figure 4. The triangle's represents the central most student from graduates of that school with red for the new school and blue for the old school.



Figure 6. The quadrants boundaries are in red. The six-county study area is in blue with students plotted as black dots.

Over half of the student addresses were in the Southwest quadrant (Table 3). This quadrant has a large urban area within ten miles of the new school site. The quadrants to the west (NW and SW) remained consistent in the number of students they supplied while the quadrants to the east (NE and SE) showed a gain in graduates for the new school site. The urban areas in these quadrants are experiencing rapid growth (American Fact Finder, 2005).

A Chi-Square test was used to determine if the observed frequencies in Table 3 deviated significantly from the expected frequencies. A 1:1 ratio was used for the test since the assumption was that there is no difference between the students from the two sites.

graduation	site.		
Quadrant	Old	New	Total

Table 3. Student breakdown by quadrant and

Quadrant	Old	New	Total
	School	School	
SW	416	417	833
SE	151	166	317
NE	82	110	192
NW	113	117	230

The Chi-Square test showed the NE quadrant results were significantly different between the two schools (p<.05). Possible reasons for this deviation will be explored later in the paper. The remaining three quadrants and total subset comparisons showed equal frequencies (p>.05).

The move of the new school site to the north allowed easier access from the major urban center in the northeast quadrant. The completion of a limited access highway also improved access. The urban area to the east of the school site is a rapidly developing area, which has shown more than 18% population increase (American Fact Finder, 2005). The completion of the limited access highway from the south to near the new school site has improved access from the urban areas in the SE quadrant. Figure 7 shows the improved highway segments in yellow. The improved access and population growth of these areas suggest an explanation for the new school's gain in graduates from the NE and SE quadrants.



Figure 7 School access shown with major roads in coral with newly changed sections of them in yellow. Minor roads are represented in gray. The old school site is a blue star and the new school site is a red star.

Congregation

A federation of local congregations operates the high school. The student database had a congregation field associated with many of the students. No congregation was listed for 264 of the 1777 total students with an additional 112 students not belonging to a federation congregation. This resulted in 376 students (2%) who were not associated with a geocoded congregation.

Students were listed as members of 38 different congregations. The total number of students any congregation contributed during the ten-year study period ranged from 0 to 141 students.

Figure 8 depicts the distribution of the congregations listed in the student database. During the ten-year period of this study eight congregations sent more students to the old school than the new school while 27 congregations sent more students to the new school than the old school. The remaining congregations sent the same number to each school. Table 4 lists a breakdown for the number of congregations by quadrant. These numbers mirror the number of students found in the quadrant and provide a backdrop for the following analysis.



Figure 8. The location of the congregations is shown as yellow spots. Blue lines show ten-mile buffers and quadrants and gray lines the road system.

Table 4. The number of congregations in each quadrant.

Quadrant	Number of Congregations
SW	15
SE	10
NE	8
NW	9

The ten congregations that contributed the greatest number of students were analyzed using the quadrant analysis. Five congregations were from the southwest quadrant, two from the northwest quadrant, and three from the southeast quadrant. A buffer analysis also showed that nine of the ten congregations were within ten miles of the school sites. The only congregation outside the ten-mile ring was within twenty miles and located in an urban setting (Figure 9). The location of these congregations is consistent with the travel limitations and the potential of urban areas to have congregations of larger membership.

Mapping the seventeen

congregations that contributed less than ten graduates during the ten-year study period revealed these congregations are located in more rural areas and at a greater distance. Only two congregations fell within the ten-mile buffer (Figure 9).



Figure 9. Red triangles depict congregations with the lowest total student contribution. Black squares depict congregations with the highest total student contribution. Blue lines illustrate ten-mile rings and quadrants.

A quadrant analysis was applied to congregations that had more students graduate from the new school than from the old school. It was also applied to those that had more graduates from the old school than from the new school. No trends were observed in either case.

Students by Distance

The GIS software was used to geocode the addresses of the students and then select subsets of those students based on their year of graduation and distance from the site of their graduation. This data was used to create maps and collect data that was subjected to statistical analyses using the SPSS 11.0 software.

The ring buffers were set at onemile intervals for the first ten miles and at five-mile intervals there after (all distances referred to in this section are Euclidean unless otherwise noted). Each school site had a set of buffer rings created around it. These buffers and the students located within them provide the base data for the analyses performed in this section. Figure 10 illustrates the study area imposed on a set of ten-mile buffers centered at the new school site. No students from the study area were found in the outer (50 mile buffer) ring.



Figure 10. The study area shown with ten-mile buffers around the new school site with the red quadrant lines.

Figure 11 plots all the students within a given distance from the school site of their graduation. The difference in origin of the two plots can be explained by the more rural setting of the new school. The old school urban setting allowed for more students to be found in the first few miles from the site. A definite flattening of the curve can be observed around the fifteen-mile mark. This led to further analysis.

A set of one-mile buffers from one mile to forty miles was created around each school site. The number of students in each one-mile buffer were counted. The first subset was the old school graduates buffered from the old school and the second subset was the new school graduates buffered from the new school. This data is graphed in Figure 12. The relocation to a more rural area affected the number of students within the first two miles. After the first two miles both data sets show a steady decline toward the fifteen-mile ring. A bubble in the decline has a peak at about twenty miles (Figure 12). This corresponds with two urban areas, one in the NE quadrant and one in the NW quadrant. This data suggests that if the population density were consistent the number of students per mile would steadily decline with distance. This analysis shows the forty-mile distance to be the upper limit for commuting, but not for school attendance.



Figure 11. A graphic comparison of number of graduates within a distance from their graduation site for each school site.



Figure 12. A plot of the number of students per mile vs. miles. The old school students are the dark blue line and the new school students the purple line.

The buffer data sets were plotted using a curvilinear cubic regression

model. There was a strong correlation between the regression model and the students from the new school as well as from the old school (Figures 13 and 14). The coefficient of determination between the data and the model was 0.99. A simple regression analysis compares two variables to see if a relationship exists between them. The regression model will show if a dependency between the two variables exists (Zar, 1984).



Figure 13. The results of applying the curved linear regression model to student distance from the new school site.



Figure 14. The results of applying the curved linear regression model to student distance from the old school site.

A further analysis was performed

using the same series of five-mile buffers that were created around each site. The number of students found in the five-mile range of the buffer ring rather than the number of students found from the school site (center) to the buffer distance were used for this analysis. The same mathematical model was applied and similar results were obtained. A paired T-test was also performed. This test is commonly employed to test for correlation and variances between two populations (Zar, 1984). These results indicate a strong correlation between the population from the old school and the population from the new school.

Conclusion

The result of the various analyses showed strong support for the hypothesis that there is not a significant difference between the subset of students that graduated from the old school compared to those that graduated from the new school. More variation occurs between the individual year subsets than between the two site subsets. The contribution of students to the high school by congregation yielded no discernable pattern of change.

The quadrant analysis did reveal interesting facts. The southern half and more specifically the southwest quadrant contain the urban areas closest to the school sites. These urban areas are the only ones within seven miles of either school site. Analysis shows that half of all the graduates come from within 7.5 miles of either school. It also showed that over half of all the graduates came from the southwest quadrant. The presence of the urban area explains the skewing of students to the southwest. The presence of the lake also presents a limiting factor in the southern quadrants.

The northern quadrants show fewer students. These students generally commute farther due to the more rural nature of the area. The urban areas in these quadrants are smaller and outside the seven-mile buffer. When urban areas do occur in these rural areas an increase in the number of students commuting from that distance was noted. Changes detected in the student subsets might best be explained by population growth and road improvements as opposed to the relocation.

Acknowledgements

I would like to thank the members of the Resource Analysis Department for their professionalism and guidance as well as my family members for their support and assistance in completion of this project. I would also like to thank Fox Valley Lutheran High School for sharing the data, which made this project possible.

References

American Fact Finder. http://factfinder. census.gov/home/. Retrieved October 7,2005.

Mapquest. http://www.mapquest.com. Retrieved September 28, 2005.

U.S. Census Bureau. http://www.census. gov/geo/www/tiger/tigermap. Retrieved September 25, 2005.

U.S. Postal Service. http://hdusps.esecurecare.net. Retrieved September 12, 2005.

Zar, Jerrold. 1984. *Biostatistical* Analysis 2nd Edition. Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07458.