Aggregation in Spatial Data Management: Prerequisite Database Concepts for GIS Skills in Retail Marketing

Peter Y. Wu
wu@rmu.edu
Department of Computer and Information Systems
Robert Morris University
Pittsburgh, PA 15108, USA

Abstract

With myriads of application areas, the geographical information system (GIS) has been proven to be a very valuable tool for the intelligent use of information. Many education institutions are adding GIS courses to the Information Systems curriculum. GIS being fundamentally a tool to work with spatial information for problem solving, GIS skills require a multi-disciplinary background. Sufficiently preparing students to learn GIS skills becomes a challenge in course design. We investigate the prerequisite database concepts needed in the GIS skills, focusing on spatial data management and business applications in retail marketing. Starting with comparing spatial join to table join, we then analyze different spatial data processing functions. The concept of aggregation becomes essential to the understanding of spatial join and the dissolve operation in spatial processing as well. These GIS functions all have common applications in retail marketing, among many other areas. We contend that while a complete database course as prerequisite may not be necessary, covering these fundamental database concepts is essential in the first GIS course.

Keywords: GIS, Aggregation, Spatial Data Management, Spatial Join, Dissolve.

1. INTRODUCTION

With colossal volumes of data readily available and more or less ubiquitously accessible through the internet, the information systems education curriculum must also be changing its focus to the intelligent use of information for businesses today. (Gorgone et al., 2002; Caputo, Kovacs, & Turchek, 2004) Geographic information system (GIS), fueled by lower cost of hardware and the availability of data, has proliferated in the market place, seeing tremendous growth in commerce, government and industry, (Gewin, 2004; Donahue, 2008; DiBiase, Corbin & Fox, 2010) and proven to be a viable tool working with spatial data for business problem solving. (Boasson, Boasson, & Tastle, 2004; Miller, 2007; North, Fee, & Bytnar, 2009; Cimons, 2011) In the past decade, many schools are adding GIS courses to their information systems curriculum. (Glover, 2005; Reames, 2005; Wu & Kohun, 2005; Sinton & Lund, 2007; Wu, 2007; Kolvoord, 2010) One of the design issues about a GIS course is the appropriate prerequisites for the course. Depending on the context of the curriculum design, there may be a few different approaches. A complete database systems course will serve well to prepare students for the GIS course, but some opt to avoid the situation of cascading prerequisites in the curriculum, and favor the inclusion of fundamental database concepts as preparatory in the GIS course (Wu & Rathswohl, 2011). The approach is important to curriculum design. On the one hand we may minimize prerequisite courses, and on the other we may allow first year graduate students from other backgrounds to get into the course more readily.

In this paper, we investigate the prerequisite database concepts needed in the GIS skills for spatial data management, focusing on some retail marketing application issues. The next section starts with using GIS for information visualization. Quite often, we apply an outer join operation to bring the needed information into the digital map so that the GIS can visually present the information. However, when the data records are not indexed in the same way we formulated the geographical features on the map, the table join is not applicable. Sections 3 and 4 then discuss two different approaches in applying the spatial join: point-to-polygon and polygon-to-point. In both cases, while the join operation is based on spatial relationship, we also need to understand how aggregation of attribute data occurs. Section 5 goes on to discuss another common spatial processing function in business application, the dissolve operation. How to use the aggregate function in the processing may then become apparent. In the last section, we conclude with listing these prerequisite database concepts, and argue to include data record grouping and aggregation as preparatory for the GIS course.

2. VISUALIZE SPATIAL DATA WITH JOIN

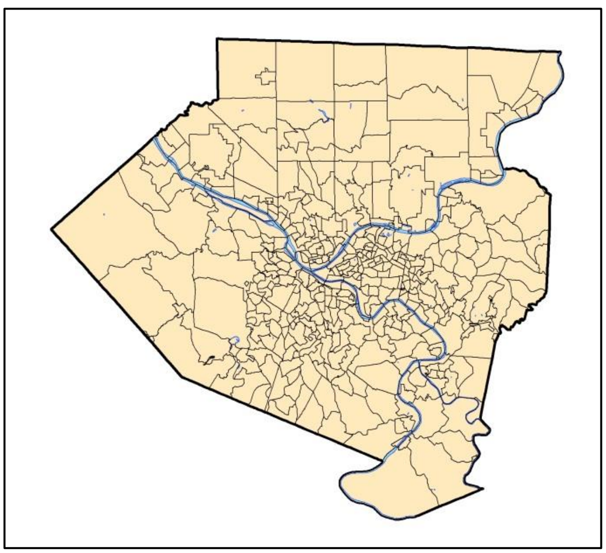

Figure 1: Census Tracts in Allegheny County

A major focus of retail marketing is the potential customers: What do they want to buy? Where are they located? GIS is an excellent tool to visualize these opportunities on a map (Wu & Rathswohl, 2011). The demographics information is readily published by the Census Bureau. The entire nation of the United States is subdivided into tracts and then into blocks for the collection of census data. The maps are also available on the website administered by ESRI (ESRI, 2013). Figure 1 shows the map of 416 census tracts in Allegheny County, Pennsylvania.

For the map, the GIS maintains a data structure with one record for each geometric feature. In this case, the feature class is that of polygons, one for each census tract. While there is an extremely rich set of information gathered for each of these census tracts - available at the census bureau web site (www.census.gov) - we need to select the appropriate data and relate them to the respective census tracts. To locate the potential customers in retail marketing, we want visualize the information about the average per capita income in each census tract on the map. We apply the simple table join, technically an outer join, to associate the chosen data to the census tract, indexed by the Tract ID.

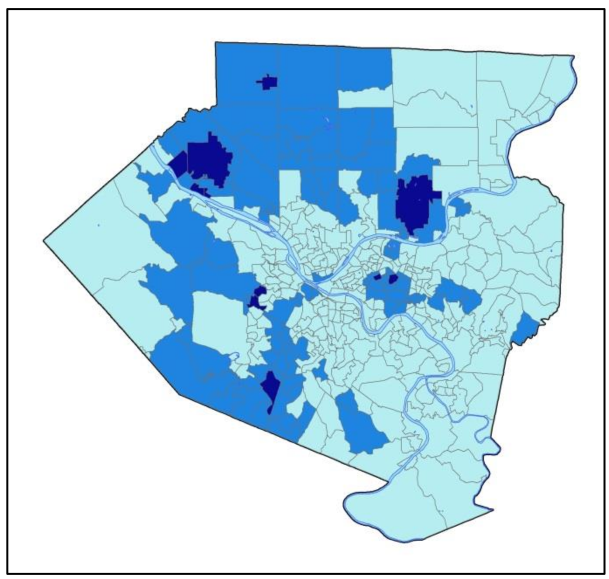

Figure 2: Per Capita Income by Census Tract

Making use of the database system at the core of the GIS, a simple SQL select statement can select those census tracts at any range of average per capita income to be highlighted for visualization. With progressively changing levels of grey or saturation called a color ramp, the GIS presents the information visualized in a choropleth map. Figure 2 is an example of the choropleth map illustrating the census tracts of Allegheny County, shaded in light, medium, and dark blue for per capita annual income levels respectively at 25,000 or less, 25,000 to 50,000, and above 50,000 (in US dollars). The GIS provides the tool for us to conveniently set up and adjust the class intervals in the color ramp of the choropleth map (Gorr & Kurland, 2007).

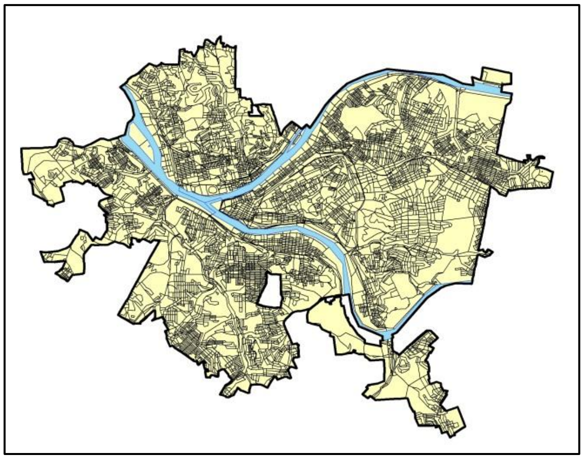

Figure 3: Census Blocks in of Pittsburgh

The Census Bureau also provides demographics data for the smaller census blocks. Figure 3 illustrates the city of Pittsburgh in Pennsylvania, consisting of 7492 census blocks. We may also obtain the population information for each age group at the census block level. With table join, we bring the data to the map feature, that is, to the attribute data record for each census block. The GIS will allow us to highlight the senior population (of age 65 and higher) in each census block in a choropleth map. Figure 4 presents the map, shading the census blocks in red where the senior population is more than 15.

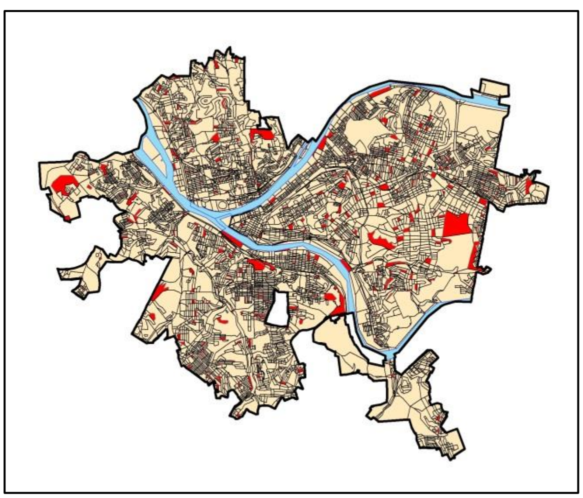

Figure 4: Senior Population by Census Block

The GIS skill requires the student to be reasonably familiar with the relational data model, the concept of simple SQL select function, and the outer join operation of two tables.

The Census data provides us with a great deal of demographics information about people by the census tracts and by the census blocks. However, when we are running a business, our administrative regions may coincide neither with census tracts nor census blocks. There is no appropriate attribute data (such as Tract ID or Block ID) for us to apply table join to bring the information into our map. The spatial data management capability of the GIS, often referred to as spatial join, can work with the geospatial relationship in the data. We will discuss different approaches in more details.

3. ATTRIBUTE DATA FROM THE NEIGHBORHOOD

Suppose we want to examine our business competition: to study how our business may compete in each municipality of the city. We first gather the addresses of the store locations, our own as well as those of competitors. We may conveniently apply address geocoding to produce a map layer of point features of these store locations. Address geocoding is a GIS function to convert coded information such as postal address into map location (Wu & Rathswohl, 2010). Figure 5 illustrates a map of the store locations in city, as a map of point features.

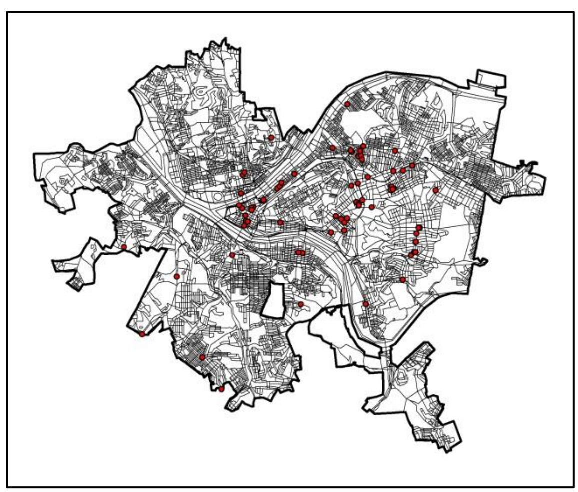

Figure 5: Store locations in the city

 ISSN: 2167-1435
v30 n2576

To analyze the competitor situation, we need to know more about the neighborhood in which every store is located. Suppose we want to know the senior population in the census block where a store is located. While there is no appropriate way to apply table join, we can see that visually when we overlay the two map layers together, locating the census block where each store is. The GIS does this with a spatial join.

In this case, the spatial join applies the census block attribute data to each store location: whenever a store is located inside a census block, the data from the census block is then also attributed to the point feature of the store. The operation joins polygon data to point features. Since each point is only located in one polygon, we can determine the senior population in the neighborhood of each store location. Figure 6 shows a map of store locations marked in red those stores in a census block with senior population more than 15.

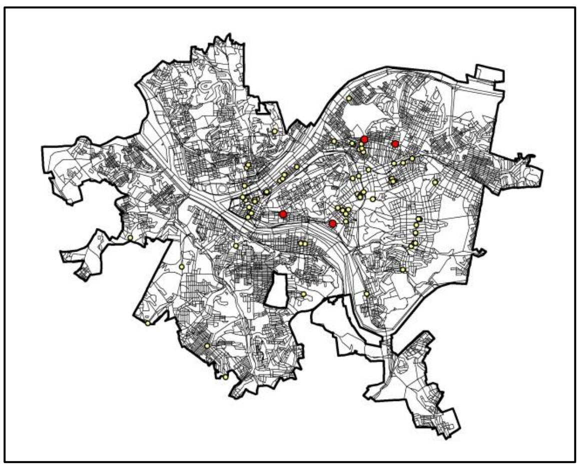

Figure 6: Stores at blocks with senior population>15

4. AGGREGATE DATA TO THE NEIGHBORHOOD

Since each municipality may have its own tax arrangement and incentives to encourage certain lines of business, we may have set up business administration by the municipality. For any municipality with at least one of our stores, we set up an administrative office there. On the other hand, we also are not so concerned about competition from stores far away from our store, such as those outside of the municipality. We therefore want to analyze each case of competition by the municipality. To do this, we need to bring to the information of each municipality, the data about each store, both our stores and those of our competitor.

Figure 7 shows the map of municipalities in the city of Pittsburgh. Again, there is no appropriate key attribute to apply table join to bring the information about each store to this map while the two sets of data are related spatially, and the GIS can do this with a different spatial join.

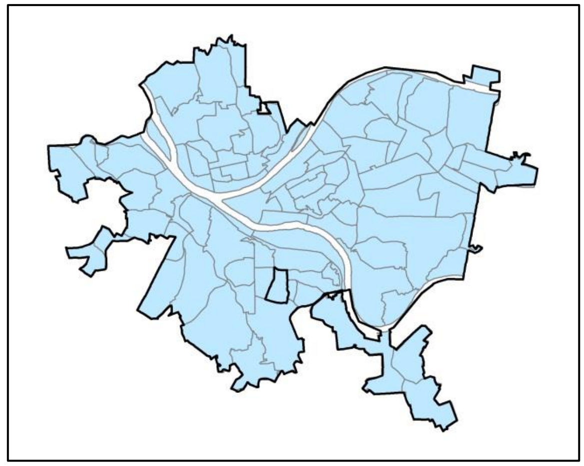

Figure 7: Municipalities of Pittsburgh

In this case, the spatial join applies the attribute data of each store to the municipality in which the store resides: information about a store becomes attributed to the polygon feature of the municipality containing the store. The operation joins point data to polygon features. However, each polygon may contain more than one point feature. A municipality may contain several stores.

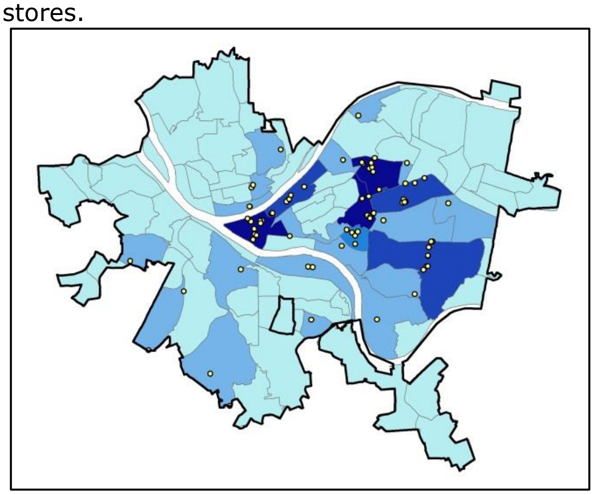

Figure 8: Number of stores by municipality

When multiple stores are grouped together in the same municipality, we need a way to properly aggregate the attribute data of the different stores to the record for the municipality. The same concept of aggregate function applies here: for a municipality with multiple stores, we may sum up the total service capacity the stores together may serve customers; or sum up the total quantity of stocking inventory. Figure 8 shows a choropleth map illustrating the number of stores in each municipality, with different shades of blue for the five class intervals: 0,1-2,3-5,6-8,9-10.

The GIS skill requires the student to understand the concept of aggregate functions and which ones to use when the spatial operation organizes the point data records into sub-groups. The basic understanding of statistics is necessary, but familiarity with the concepts of using "GROUP BY" in SQL will be most helpful.

5. DISSOLVE REGIONS INTO ONE

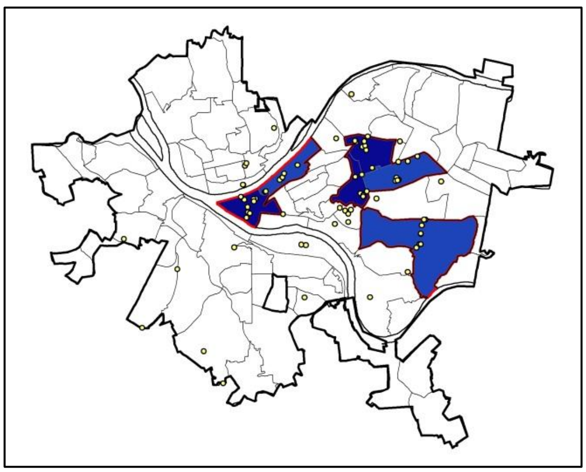

Figure 9: Dissolve - consolidation for administration

Dissolve is a GIS operation to merge multiple regions together on a polygon map layer, in the processing aggregating attribute data into one collection. Similar spatial processing is done when we may consolidate business operations. For various reasons such as dealing with tax incentives by different municipal governments, or unifying counter measures against same competitors in retail business, it may be desirable to merge business units in different municipalities into one. Suppose we want to identify those municipalities in which we are working against the same competitor in business. On the map, we want to dissolve these polygons into one region whenever they share the same specific attribute data value. Figure 9 illustrates the dissolving of municipalities into one when we have 6 or more competitor stores in the municipality.

The operation is technically not a spatial join, but it groups together records based on some common attribute data values. Polygons are dissolved into one when they are grouped together. But in the process, the same data attribute aggregation is applied. The GIS skill requires the same understanding of aggregate functions when database records are grouped together, merging into one.

6. CONCLUSION

While many schools are adding GIS courses to their IS curriculum, we focused on issues in the detailed course design (Wu & Kohun, 2005). Particularly when a first database course is not practicable as prerequisite, there is a need to identify the necessary prerequisite concepts in order to prepare the students well. We started with applying outer join to bring attribute data into the digital map of geographical features for appropriate visualization. But when there is applicable index in the data for the geographical features, we will need to be using the spatial data management capabilities of the GIS. While the geometry involved in the functions is intriguing, the fundamental database concepts are necessary to understand some of these functions. In two different spatial join operations, as well as the dissolve operation in spatial data processing, the concepts of aggregate function and grouping are essential. In the examples of retail marketing applications, the GIS skills involved all depend on the student's preparation in understanding these fundamental database concepts. These include handling of the data tables, the logic of simple SQL select, join operation, the use of aggregate function when organizing features into sub-groups. Figure 10 is a table listing these prerequisite concepts.

	Prerequisite Knowledge
1	Data Table and Keys
2	SELECT logic based on attribute data
3	JOIN
4	Aggregate functions with GROUP BY

Figure 10: Prerequisite Database Concepts

7. REFERENCES

Boasson, E, Boasson, V & Tastle, WJ (2004) "A New Tool in IS Management: Geographic

Information Systems." *Proceedings of ISECON 2004, v.21*, Newport, RI, §.3124.

Caputo, D, Kovacs, P & Turchek, J (2004) "Defining the Essential Skill and Functional Areas of Study in IT as Measured by a Survey of Field Professionals." *Proceedings of ISECON 2004, v.21,* Newport, RI, SS.2215.

Cimons, M. (2011), "Geospatial Technology as a Core Tool: Impacts Everything from Navigating to Law Enforcement," published in *U.S. News and World Report*, May 11, 2011.

DiBiase D., T. Corbin, T. Fox, J. Francica, K. Green, J.Jackson, G. Jeffress, B. Jones, B. Jones, J. Mennis, K. Schuckman, C. Smith, J. van Sickle. (2010), "The New Geospatial Technology Competency Model: Bringing Workforce Needs into Focus," *URISA Journal*, Vol.22, No.2, pp.55-72.

Donahue, G. (2008) "Healthy Perspectives," *American City and County*, Vol.123, No.2, p.24.

Gewin, V. (2004) "Mapping Opportunities." *Nature*, Vol.427, pp.376-377.

Glover, B. (2005), "Curriculum Update for GIS Programs in Texas Community Colleges," *Proceedings of 5th Annual ESRI Education User Conference*, July 2005, San Diego, CA.

Gorgone, J.T., G.B. Davis, J.S. Valacich, H. Topi, D.L. Feinstein, H.E. Longenecker Jr. (2002), "IS 2002 Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems," Association of Information Systems, 2002.

Gorr, WL and Kurland, KS. (2007), "Designing Maps." in *Learning and Using Geographic Information Systems*, Thomason Course Technology, pp. 137-176.

Kolvoord, Robert. (2010) "What Are They Thinking: Students' Spatial Thinking Skills," *Proceedings of 7th ESRI Education User Conference*, Aug 2010, San Diego, CA.

Miller, F.L. (2007) *GIS Tutorial for Marketing,* ESRI Press, Redlands, CA.

North, M.A., S.B. Fee, J.M. Bytnar. (2009) "Relational Data Modeling to Enhance GIS-Based Visual Information Systems," *Issues in Information Systems* 10(2), pp.226-230.

Reames, S. (2005) "Business Geographic Information Systems - A Course in Business Geomapping," *Proceedings of ISECON 2005 v.22* (Columbus OH): § 2334.

Sinton, D.S. and J.L. Lund. (2007) *Understanding Place: GIS and Mapping Across the Curriculum*, ESRI Press, Redlands, CA.

Traore, M. (2010) "GIS in Merchandising Education: Visualizing Customer Data for Decision Making," *Proceedings of 10th ESRI Education User Conference*, August 2010, San Diego, CA.

Wu, P.Y. & Kohun, F.G. (2005) "Designing Geographic Information System Courses in the IS Curriculum," *Proceedings of ISECON 2005 v.22* (Columbus OH): § 2564.

Wu, P.Y. (2007) "Introducing Geographic Information Systems into the IS Curriculum: GIS Tutorial and Preparation Workshop," *Proceedings of ISECON 2007 v.24* (Pittsburgh PA): §3732.

Wu, P.Y. & E.J. Rathswohl. (2010) "Address Matching: an Expert System and Decision Support Application for GIS," Proceedings of ISECON 2010 v.27, (Nashville, TN): §1399.

Wu, P.Y and E.J. Rathswohl. (2011) "Visualizing Opportunities: GIS Skills for Retail Marketing," *Information Systems Education Journal*, Vol.9, No.6 (Jul 2011), pp.44-50. ISSN:1545-679X

ISSN: 2167-1435
v30 n2576