ESRI International User Conference 2010

Paper # 1523

Session Title: Growing Local Businesses: Spatial Analysis Fosters Regional Development

Paper Title

Spatial Analysis of Knowledge-based Occupation Clusters

Prepared by

Christine Nolan, Senior Associate, Purdue Center for Regional Development

Indraneel Kumar, Regional Planner: GIS and Spatial Analysis, Purdue Center for Regional Development

- The occupation cluster and mapping part of this paper is based on research supported by the U.S. Economic Development Administration: *Crossing the Next Regional Frontier: Information and Analytics Linking Regional Competitiveness to Investment in a Knowledge-Based Economy, 2009* *http://www.statsamerica.org/innovation/report_next_regional_frontier_2009.html*. Part of the spatial analysis component was executed after completion of the grant project and presented at the Indiana State GIS Conference, 2010.

Spatial Analysis of Knowledge-based Occupation Clusters

Abstract
Knowledge economy, technology based industries, and knowledge workers are the current buzz words in regional economic development. This project uses GIS and spatial analytic techniques to study the distribution of occupation clusters in the US. The clusters are derived by applying statistical clustering techniques on the knowledge and skills components of the occupations. A total of 15 occupation clusters are identified and six technology-based clusters are analyzed in detail to study their growth, specialization, and spatial clustering. A variety of maps and visuals are prepared for the US and four case study areas including Economic Growth Regions 6 & 11 in Indiana, WAEM (Western Alabama and eastern Mississippi), and Riverlands (portions of Wisconsin, Illinois, and Iowa). This project is part of a larger research project "Crossing the Next Regional Frontier", funded by the Economic Development Administration. The grant partners include Purdue University, Indiana University, Rural Policy Research Institute (RUPRI), Economic Development Specialists Inc. (EMSI), and the Strategic Development Group Inc. (SDG).

Acknowledgements:
The authors would like to thank Matthew Baller, Purdue Center for Regional Development (PCRD); Rachel Justis, Indiana Business Research Center (IBRC); and Sean Moore and Ana D. Ojanen, Rural Policy Research Institute (RUPRI) for mapping assistance. We are also grateful to Economic Modeling Specialists Inc. (EMSI), a project partner, for providing data. Thanks are due to IBRC, our project partner, for developing and maintaining the project website.

Introduction:
Knowledge-based assets are increasingly becoming important in the new economy. Prospective businesses and industries seek for regions that have access to a skilled and educated labor force. Unlike the manufacturing industries of yesteryear, today's high-technology based industries require human resources with specialized skills and advanced education. Now a region's assets include its institutions, inventions, inventors, educational attainment of residents, access to funds for research and development, skilled labor pool, optic fiber, etc. Occupation cluster analysis is one way of assessing the knowledge and skills of the labor force and giving insights into regions that have specializations in certain types of skills.

1. What are knowledge occupation clusters?

As part of a recent study conducted for the U.S. Economic Development Administration, the Purdue Center for Regional Development developed a set of 15 knowledge-based occupation clusters, one of which – health care and medical science – is also shown in the form of three sub-clusters for added accuracy in analysis (Table 1). National county-level data for the clusters, full cluster definitions, and the complete report are available online at www.statsamerica.org/innovation/.[1]

Table 1: Occupation Clusters Defined in the Study

	Occupation Cluster Name
1	Agribusiness and Food Technology
2	Arts, Entertainment, Publishing and Broadcasting
3	Building, Landscape and Construction Design
4	Engineering and Related Sciences
5	Health Care and Medical Science (Aggregate)
i	*Health Care and Medical Science (Medical Practitioners and Scientists)*
ii	*Health Care and Medical Science (Medical Technicians)*
iii	*Health Care and Medical Science (Therapy, Counseling, Nursing and Rehabilitation)*
6	Information Technology
7	Legal and Financial Services, and Real Estate
8	Managerial, Sales, Marketing and HR
9	Mathematics, Statistics, Data and Accounting
10	Natural Sciences and Environmental Management
11	Personal Services
12	Postsecondary Education and Knowledge Creation
13	Primary/Secondary and Vocational Education, Remediation & Social Services
14	Public Safety and Domestic Security
15	Skilled Production Workers: Technicians, Operators, Trades, Installers & Repairers

Unlike industry cluster's focus on value-chain operations (a fairly wide range of different industries and establishments linked through functions that they perform for each other such as buying inputs, providing expertise, etc.), knowledge occupation clusters are formed into groups of similar job functions and knowledge components such as bio-science, or engineering and mathematics.

However, *like* the industry classifications used for business and industry cluster identification (the North American Industry Classification System or NAICS codes), occupations are also classified and coded into a system: the Standard Occupation Classification System (SOC codes). This system is somewhat similar to the NAICS system. The current version of the SOC codes (2000 Version) contains 821 detailed occupations, which will increase to 840 detailed occupations in 2010 (Federal Register, Vol. 74, No. 12, 1-21-09). The detailed occupations are further classified into 449 broad occupations, 96 minor groups and 23 major groups.

---

[1] Research was conducted by the Purdue Center for Regional Development, the Indiana Business Research Center at Indiana University's Kelley School of Business, Strategic Development Group, Inc., the Rural Policy Research Institute, and Economic Modeling Specialists, Inc.

Additionally, SOC occupations are classified into a system called O*Net (Occupational Information Network) which describes the "content" and attributes of each occupation. O*Net classifies and describes occupations according to six major domains (see www.onetcenter.org/overview.html).

In this study, the research team has taken the O*Net "Worker Requirements" Domain (which classifies SOC occupations according to the degrees of skill, knowledge and education needed to fulfill the occupation requirements)[2] and has structured the framework for building occupation clusters based on an additional concept developed by the National O*Net Center: *Job Zones*.

A job zone is a group of occupations that are similar in these ways:

- How most people are able to enter the work
- How much overall experience people need to do the work
- How much education people need to do the work
- How much on-the-job training people need to do the work

Within the knowledge domain for O*NET occupational descriptions, there are 33 knowledge variables. Each variable has two measurements, namely: knowledge *level* and knowledge *importance*, which are measured on a scale of 1 to 7 and 0 to 5, respectively. In total, there are 66 knowledge measurements for each occupation (33 each for knowledge level and knowledge importance). The team found that knowledge importance scores were highly correlated with knowledge level scores; consequently, the 33 variables for knowledge importance were dropped from the analysis.

Method Used to Identify Occupation Clusters

A major goal in this study was to quantifiably identify and categorize occupations into clusters based on O*NET-SOC knowledge classifications and measurements. To achieve this goal, Ward's agglomerative hierarchical clustering algorithm (henceforth, Ward's algorithm) was utilized. Though other hierarchical and K-cluster algorithms were tested, Ward's generated the best results and past research (Feser 2003) has shown this to be the case. Ward's algorithm uses error sum of squares to minimize within-cluster variance. A detailed technical description of the clustering process is available at (http://www.statsamerica.org/innovation/reports/sections2/G.pdf).

Using 33 variables describing occupations' knowledge levels, 19 clusters were formed on the basis of a cluster analysis using Ward's algorithm. The knowledge variables refer to the squared knowledge level scores. Squaring the knowledge level scores allowed key

---

[2] National Center for O*NET Development: www.onetcenter.org/content.html

knowledge characteristics for each occupation to be highlighted when conducting Ward's algorithm.

The 19-cluster solution served as a baseline and was subsequently fine-tuned by scrutinizing each cluster for consistency. As a result, 90 occupations were re-allocated and some clusters were merged. Moreover, data constraints necessitated the formation of a cluster that pulls together all postsecondary educators, independent of specialization. The final result has been the identification of 15 clusters containing all occupations within O*Net Job Zones 3 to 5.

In addition, the team has provided a breakdown of the health care and medical science occupation cluster into three sub-clusters in order to facilitate greater accuracy in assessing regional strengths in this highly important occupation cluster:

• Medical Practitioners and Scientists

• Medical Technicians

• Therapy, Counseling, Nursing and Rehabilitation

Why is occupation cluster analysis important for regional economic development?

Occupation cluster analysis is a relatively new approach in regional development. In contrast to industry clusters that focus on what businesses produce, occupation clusters focus on the knowledge, skills and abilities of the individuals who work for those businesses. Occupation cluster analysis offers insights into the talent base of the regional workforce that go beyond the relatively simple measure of educational attainment (such as highest degree earned).

This project focused on identifying and mapping clusters of occupations for the United States as a whole and for two selected test regions in the state of Indiana, using counties as the basic unit of measurement, with an emphasis on occupations most closely associated with knowledge and innovation. The research team built upon the pioneering work of Markusen and Barbour (2003, 2007) and others (for example, Feser and Koo, 2001; Feser 2003) to develop a methodology for determining the following:

• The structure and composition of regional occupation clusters

• The competitive advantage of regions in terms of occupation and knowledge specializations

• The distribution of concentrations of regional occupation clusters that seem likely to enable or encourage the production of innovative behaviors

• The geographic distribution of concentrations of particular occupations within regional industry clusters

Research to date suggests that occupation clusters may be at least as important as industry clusters in driving regional competitive advantage, particularly in a situation of globalization where the skills, knowledge and capacity to innovate may be the most important factors that

distinguish one region from another.
(http://www.statsamerica.org/innovation/reports/sections2/3.pdf, pp43-44).

Developing a nationwide mapping of occupation clusters, with county-level data available for every U.S. county and the capability to aggregate counties to a regional level, serves as a powerful complement to an understanding of regional industry clusters, which was a major focus of a previous EDA-funded project conducted by partners in this research team. GIS maps showing occupation cluster employment, cluster specialization (location quotients over 1.2) and changes in degree of specialization, by county, for the entire United States can be found at (www.statsamerica.org/innovation/maps.html). The same website also displays similar maps for industry clusters throughout the nation.

Of the 15 occupation clusters, the research team identified six technology-based knowledge occupation clusters that are supposedly drivers of the regional innovation and economy. Henceforth known as the tech-clusters, they include Information Technology (IT); Engineering (ENG); Health Care and Medical Science (Medical Practitioners and Scientists) (MED); Mathematics, Statistics, Data and Accounting (MATH); Natural Sciences and Environmental Management (SCI); and Postsecondary Education and Knowledge Creation (ED). Although only eight percent of the national employment, the tech-clusters warranted additional analysis to study situation of jobs growth in the US as well as Indiana. Percent of jobs in the tech-clusters combined revealed that only a few states were above the national share of (8.2%) in 2007. They included MN, the states in the northeast corridor, UT and CO, CA and WA. Whereas TX was at the national average, 34 states were below the national average. The details are available at (http://www.statsamerica.org/innovation/reports/sections2/D.pdf, pp 216). In addition to analysis at the state level, spatial analysis of individual clusters at the regional level could illustrate patterns of clustering and dispersal at the regional level. Moreover, regions, as a group of few counties, are the proper spatial scale for planning for workforce development.

What is spatial analysis?

Spatial analysis looks into integrating geographic space into research and intellectual enquiry. The advent of computing technology, computer graphics, and finally the Geographic Information System (GIS) have increased the use of mapping and spatial analysis across various disciplines. According to Longley and Batty, in addition to advances in computers and graphics, spatial analysis emerged largely from the advancement of quantitative methods in geography, geology, earth sciences, regional science and macro-economics.[3] Contributions of different scholars from different disciplines have expanded spatial statistics as well as spatial econometrics tools. Now a new way of thinking, known as “spatial thinking”, is

[3] Longley, Paul A. and Michael Batty, Advanced Spatial Analysis: the CASA book of GIS (Redlands: ESRI Press, 2003), 3.

emerging within the researchers and practitioners. Spatial analysis extends a researcher's abilities to look beyond global statistics and parameters and account for unique individualities at the regional and local levels. In a workshop presented at the Center for Spatially Integrated Social Science, Fotheringham described spatial analysis tools, such as Geographically Weighted Regression (GWR), as a "spatial lens" through which researchers can investigate spatial heterogeneities.[4] Spatial analytic techniques integrated with GIS facilitate strategic planning and policy making and can therefore be a useful tool for decision making.

How can Geographic Information System (GIS) and spatial analysis assist in cluster studies?

As the name itself suggests, "clusters" have spatial characteristics. They form a spatial pattern where events or activities are located closely over the geographic space. Clusters can result from the first or the second order spatial processes. They can be generated from causal variables or due to interactions between each other.[5] Previous research has shown that industries, businesses, and occupations "cluster" over space to draw benefits from agglomeration and localization economies. The geographic proximity facilitates interactions. Sharing of resources, skills, knowledge and technologies are possible. The agglomeration also benefits economies of scale for suppliers, distributors, and helps create a labor force with adequate skills. Spatial analysis can reveal "causes and effects" of these interactions.

As mentioned previously, the six technology-based occupation clusters were studied in detail for the project. The Cluster data consisted of employment figures and location quotients in the 15 occupation clusters for 2001 and 2007 for each of the 3,140 counties in the U.S., Alaska, and Hawaii. This included 3,109 counties in the continental USA, 27 counties in Alaska and four counties in Hawaii. Maui and Kalawao, two counties in Hawaii, had to be combined to facilitate data creation. Our previous study of business and industry clusters presented at the 2006 ESRI International User Conference included cluster data from 2004 and only 3,108 counties in the continental USA were covered. Economic Modeling Specialists, Inc. (EMSI), a partner in the current occupation study prepared the employment and location quotient data for all the 3,140 counties.

The research team used ESRI's ArcGIS to develop GIS databases that included attaching tables to the county polygons and points comprising the county centroids. For the initial

---

[4] Fotheringham, Stewart, 2003 Workshops Video Clips- Geographically Weighted Regression and Associated Statistics, http://www.csiss.org/streaming_video/2003/. Accessed on May 17, 2010.

[5] De Smith, Michael, Michael Goodchild and Paul Longley. "First-and second-order processes." In *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools,* www.spatialanalysisonline.com. Accessed on May 20th, 2010.

mapping, we used shapefiles to facilitate GIS analysts using the data to create maps for presentations and reports in different locations in Indiana and Missouri. Recently the data has been uploaded into an ArcSDE Oracle database and now they are polygon and point feature classes. Eventually, the maps will be served by using the ArcGIS Server platform within a dynamic web-based interface.

In addition to county maps for the entire USA, the research team prepared detailed maps of occupation clusters plus infrastructure and amenities for the pilot regions. These include Economic Growth Region 6 (9 counties) in eastern Indiana and Economic Growth Region 11 (9 counties) in the southern Indiana. Additional regions included a 37county region in western Alabama and eastern Mississippi (WAEM) and Riverlands, a 17 county region spanning parts of Wisconsin, Illinois, and Iowa.

Mapping and graphics included:

- Dot-density employment maps at the county level for whole USA
- Graduated symbol mapping
- Choropleth maps for four pilot regions
- Infrastructure and amenity maps for four pilot regions

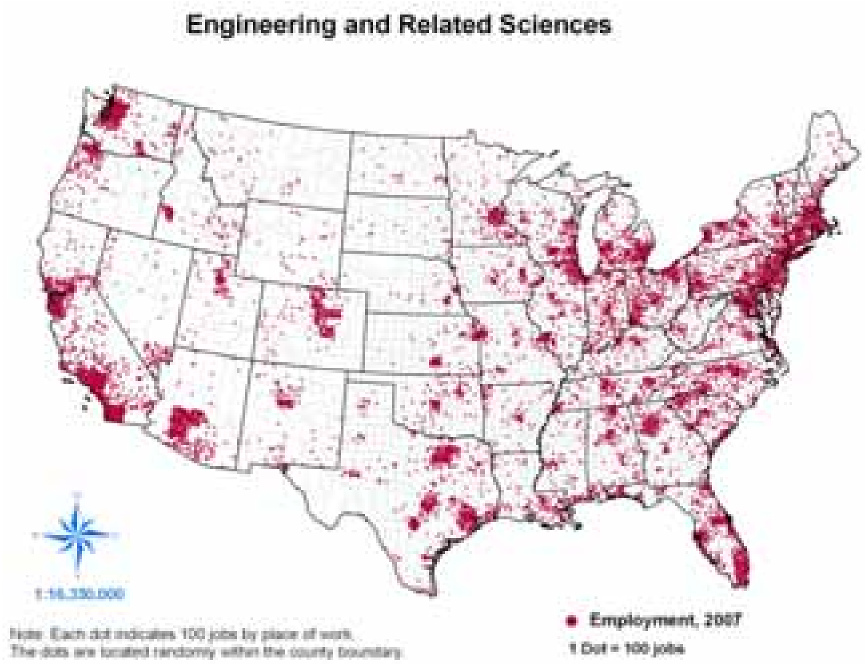
Exploratory Spatial Data Analysis (ESDA) included:

- Nearest-neighbor
- Standard deviational ellipse
- Spatial autocorrelation (Global and Local)

Dot-density Mapping:

Dot-density mapping is a useful way to show the quantity and geographic distribution where dots are generated randomly within the spatial unit, such as counties. Here, the dot-density maps show the employment counts. These maps can show discernible spatial distribution if made at a lower geographic level, such as counties, and observed at higher levels of geography, such as states and the census regions. The dot-density map of engineering occupations (Map 1) shows concentrations of engineers occurring primarily in urban and metropolitan areas with one exception - rural counties that are home to federal research laboratories. The manufacturing belt of the Midwest shows a higher concentration of engineers in 2007. Sometimes double-dot density maps can be used to show two different types of distributions. Map 2 shows distribution of engineering occupations along with agribusiness and food related occupations. The former shows concentration mainly in the urban areas, whereas the later has concentrations in urban as well as in rural areas.

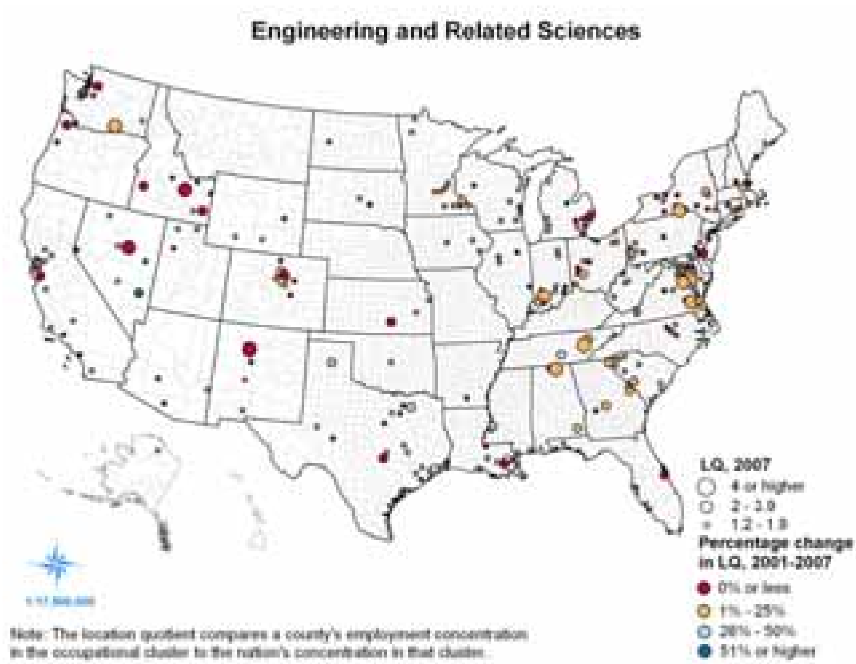
Map 1: Dot-density of Engineering & Related Sciences Occupation Cluster, 2007



Map 2: Double dot-density of Engineering vs. Agribusiness Occupation Clusters, 2007

Graduated-symbol & Other Types of Mapping:
Graduated-symbol is another way of mapping quantities. Here the symbols show range-of-quantities and larger symbols indicate larger quantities. Graduated symbol mapping was created for location-quotients of the occupation clusters where symbol sizes showed values for the location quotients (LQs) in 2007. The increase or decrease in the LQs from 2001-2007 was shown by different colors on the same map. This is achieved by using the "Multiple Attributes (quantity by category)" method where the range of percentage changes in LQ is shown by colors and the LQ size is shown by symbols.

Map 3: Graduated-symbol, Engineering and Related Sciences, 2007



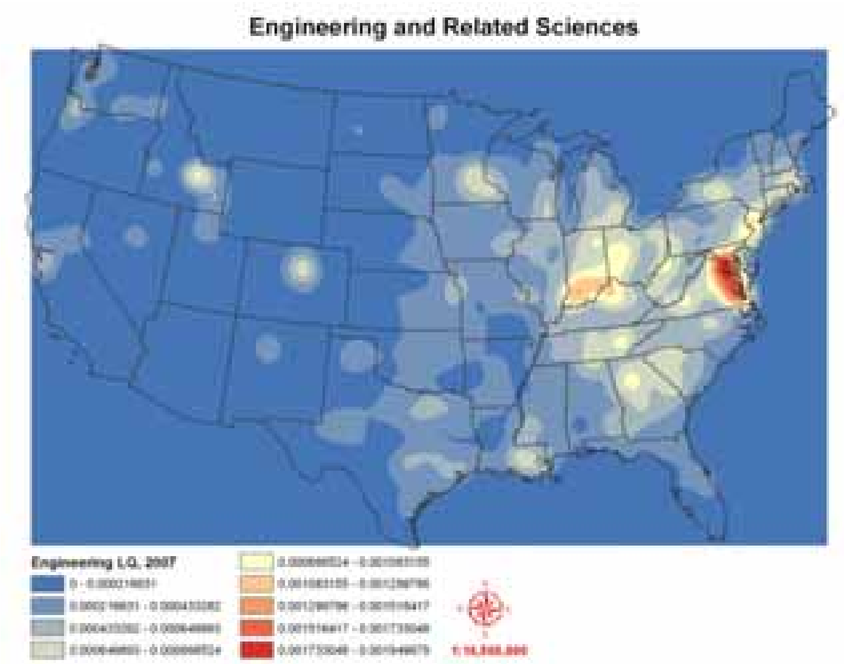
Kernel-density mapping is another way of spatial-smoothing by which values from a point dataset can be redistributed into a surface. Compared to other methods of spatial smoothing, such as the Floating Catchment Area (FCA), in kernel estimation nearby data points are weighted more than the points that are farther away.[6] In case of a lattice dataset, such as counties or census tracts, kernel can be used for spatial smoothing. If the data is a point-pattern, such as disease occurrences, kernel can be used for spatial interpolation.

Here, ESRI's ArcGIS Spatial Analyst Extension is used to create kernel-densities of location quotients (LQs) of the engineering occupation cluster, which are attached to the county centroids. Various combinations of search-radius, area-unit, and output cell sizes

[6] Wang, Fahui, Quantitative Methods and Applications in GIS (New York: CRC Press, 2006), 37.

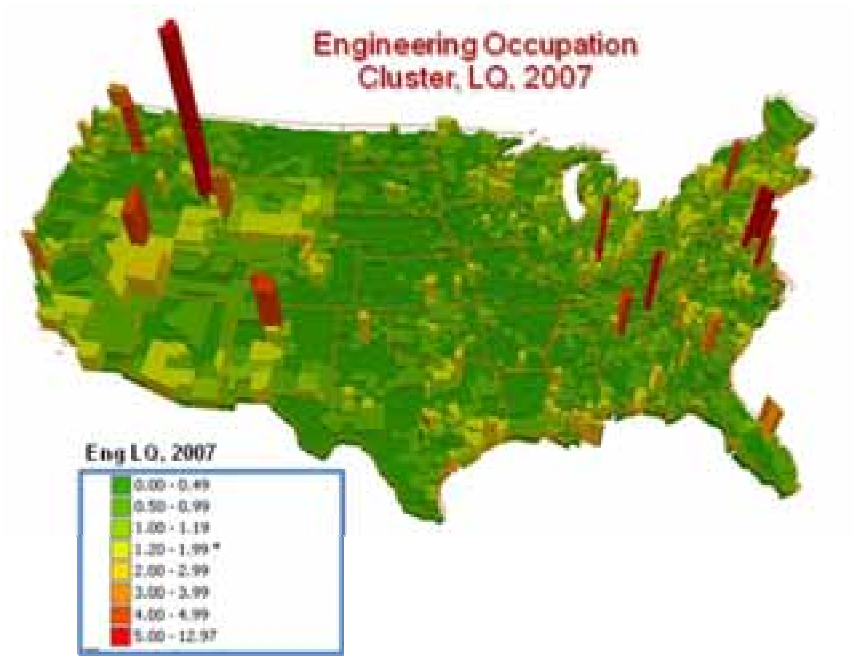
were used to experiment with the maps. Finally, a search radius of 125 km, the average of distances up to five nearest-neighbors, was found appropriate. “HawthsTools” was used to create matrices of the nearest-neighbor distances. Since LQ values of all the 3,109 counties are used, areas in Minnesota, Michigan, Indiana, Ohio, Idaho, Colorado, California, etc. emerge. A high concentration of engineers is seen in the Bos-Wash Corridor, which is Virginia, Washington DC, Maryland, New Jersey, Massachusetts, etc. Engineers do have a higher concentration in these states; however the higher density can in part be explained due to the “edge-effects”. Spatial analysis accounts for the neighbors but jurisdictions located on the edges might not have as many neighbors as jurisdictions lying in the interior parts of the study area. This may skew the “mean value” so very high or low values on the edges require further investigation.

Map 4: Kernel mapping, Engineering and Related Sciences, 2007

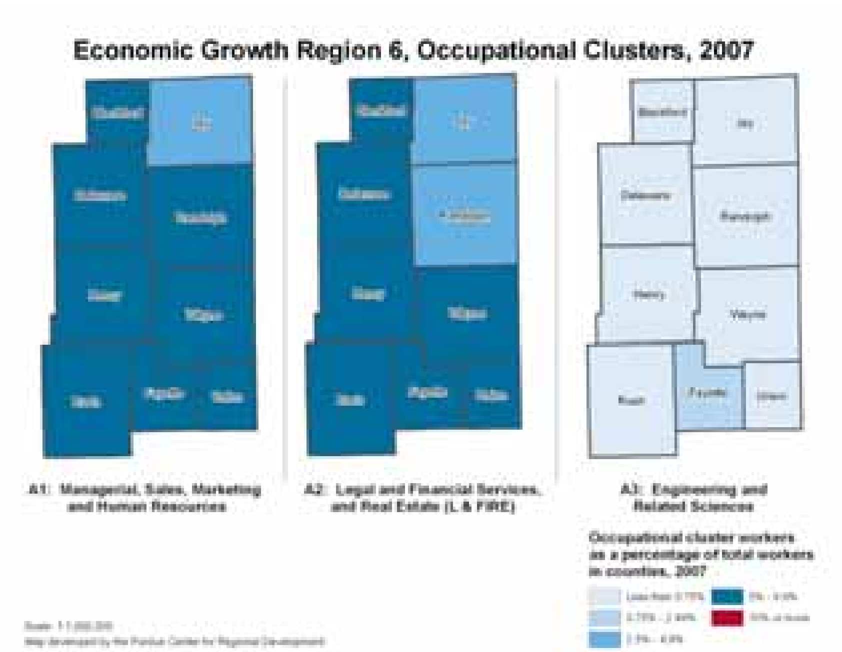


A 3-D map (Map 5) can reveal outliers or counties with very high concentration of engineering occupations. Here ESRI’s ArcScene is used to create 3-D features by using engineering occupation LQs of the counties. A lot of counties with high LQs are actually locations of research laboratories, such as naval research center in Martin County, Indiana; Oak Ridge National laboratories in Tennessee; Huntsville in Alabama; Butte County in Idaho; and Sandoval County in New Mexico. Choropleth maps (Map 6) were created for the pilot regions and were used in the workshops with county economic development and workforce professionals. The team also created infrastructure and amenities maps (Map 7) for the pilot regions by using the National Transportation Atlas Database (NTAD). Use of the NTAD GIS data facilitated uniform infrastructure mapping for the four pilot regions located in different states. Transportation infrastructure and amenities are useful assets to determine competitive advantages of the regions.

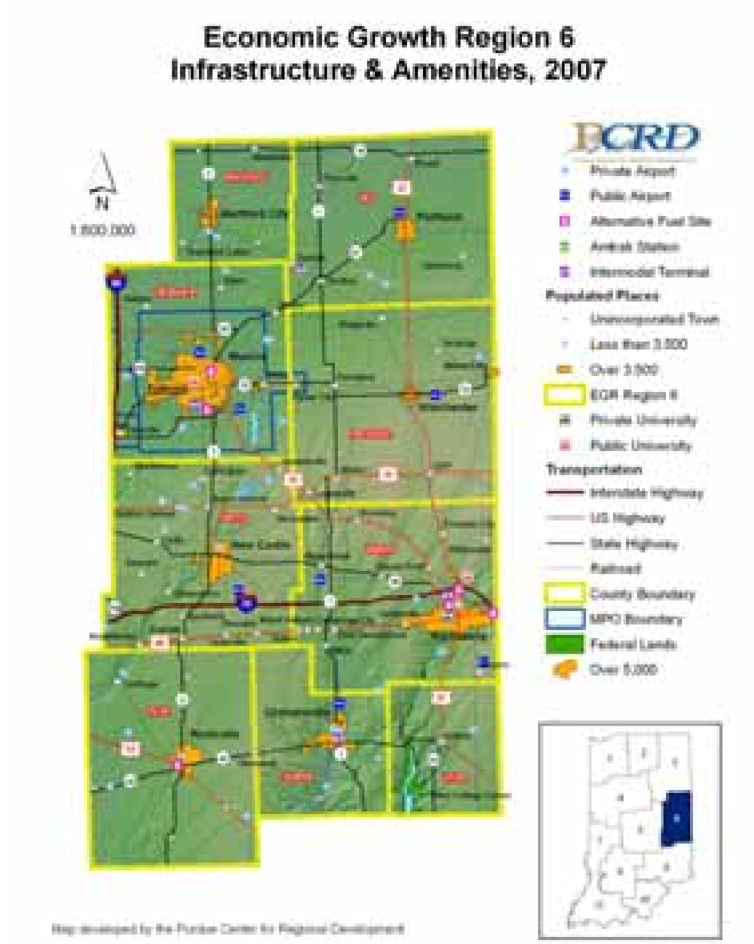
Map 5: 3-D map of Engineering and Related Sciences, 2007



Map 6: Indiana Economic Growth Region 6 (EGR 6), Pilot Region



Map 7: EGR 6, Infrastructure and Amenities



Exploratory Spatial Data Analysis:

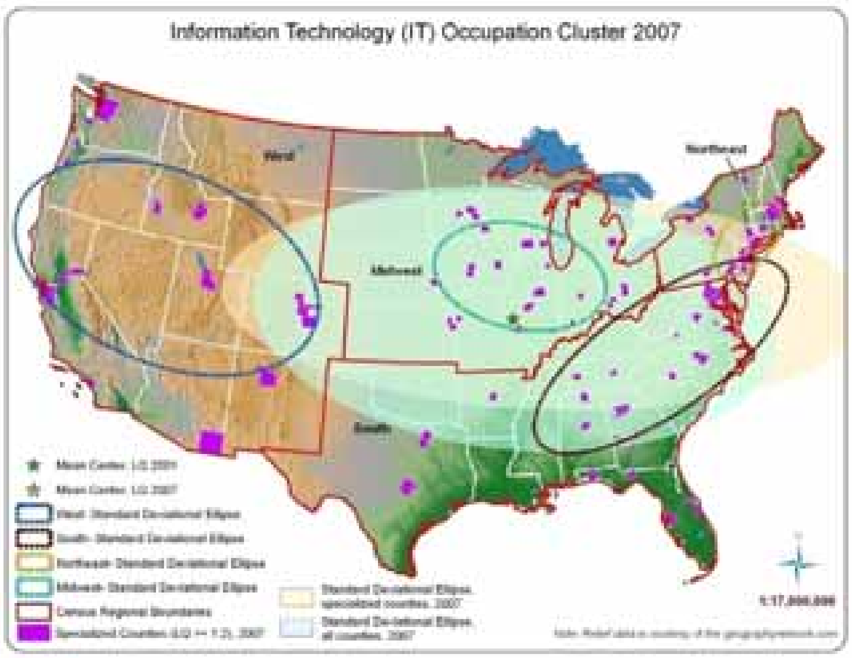
The average Nearest-Neighbor is a technique that can detect if a spatial dataset has a clustered or dispersed pattern. Here, counties that are specialized in Engineering occupations as well as Information Technology (IT) occupations are found to be clustered in 2001 as well as 2007. In case of IT occupation clusters, the Nearest-Neighbor Index has decreased showing more clustering from 2001-2007, whereas for the Engineering occupations, the index increased slightly. However, all the four scenarios have statistically significant Nearest-Neighbor indices showing a clustered pattern.

Table 1: Nearest-Neighbor Ratios

Occupation Cluster	LQ (Year)	Nearest-Neighbor Index (clustered < 1)
Information Technology (IT)	2007	0.577
Information Technology (IT)	2001	0.589
Engineering	2007	0.684
Engineering	2001	0.634

Standard-deviational ellipse is a way of demonstrating a directional distribution in the spatial dataset. Here, the directional distribution is not that prominent when applied to counties specialized in the IT occupation cluster at the U.S. level. However, the directional distribution becomes evident when applied at the census-region levels. For example, the distribution ellipse of counties specialized in IT occupations highlights the northeast corridor (the Bos-Wash Corridor).

Map 8: Standard-deviation Ellipse IT Occupation Cluster



Spatial autocorrelation is another way of assessing spatial relationships and verifying if the neighbors have similar or dissimilar values. Moran's I is a global measure of spatial autocorrelation and according to experts in spatial science, it is a powerful measure that can detect slight spatial relationships in the data. Mathematically, spatial autocorrelation is actually a correlation coefficient integrated with a spatial weight matrix. However, de Smith et al. caution that spatial correlations should not be implied as causation without examining the details.[7] As expected, at a global level, the Moran's I index has positive values for Engineering and IT occupations that are statistically significant for 2001 and 2007. For IT occupations LQ values, Moran's I was 0.44 in 2007. This means that counties with higher or lower concentration of IT occupations are co-locating over geographic space. This can imply a strong or weaker labor markets. For example, the

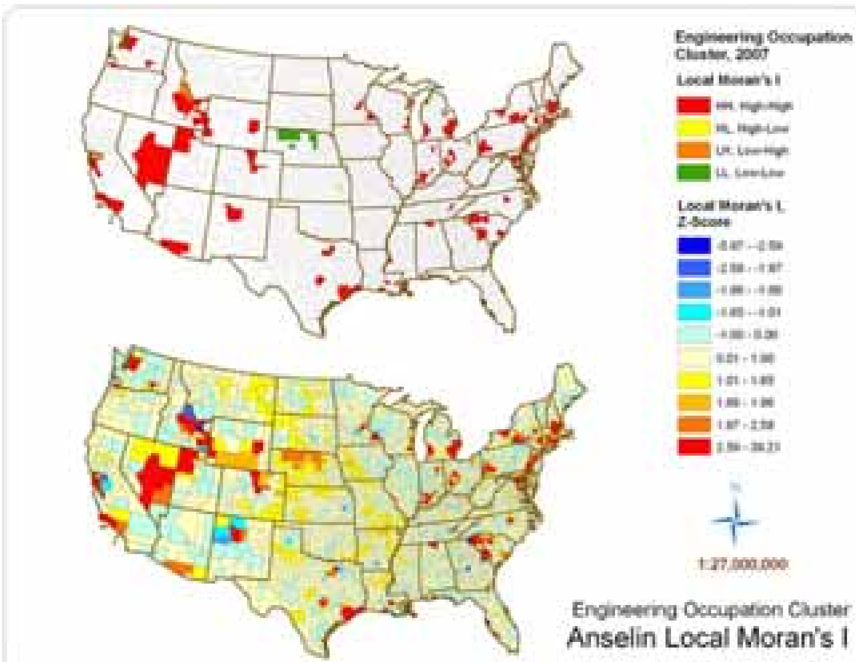
---

[7] De Smith, Michael, Michael Goodchild and Paul Longley. "Autocorrelation, Time Series and Spatial Analysis." In *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, www.spatialanalysisonline.com. Accessed on May 27th, 2010.

northeast corridor is a strong labor market for Engineering and IT occupation clusters, which can be attractive to prospective businesses and industries looking for those skills.

Compared to global spatial autocorrelation, local spatial autocorrelation indices can detect spatial variations in the correlation measures. Also known as the Local Indicators of Spatial Association (LISA) statistics, these indicators can be useful tools for policy makers. As applied in the Engineering occupation cluster LQ values for 2007, LISA could group counties that are high-high, low-low, high-low, and low-high. It can detect clustering as well as dispersal over geographic space. For regional planning purposes, it is useful to know “high-high” concentration or clustering of engineers as well as those locations that have “low-low” concentration or dispersed patterns.

Map 9: LISA, Engineering Occupation Cluster, 2007



Getis-Ord Gi* is an additional spatial statistics tool that can detect clustering of “high-high” and “low-low” values. In other words, it can detect the “hot-spots” and the “cold-spots”, which could be useful for regional planning purposes.

Conclusion: Occupation clusters can show concentration of knowledge, skills, and abilities for a county or a region, which could be a useful tool for economic development professionals. Spatial analysis can help determine spatial patterns within the data that can show clustering or dispersal over geographic space. A sequential step-by step spatial analysis and mapping can reveal valuable information for the regions. For economic development purposes, not only we should know the regions that are doing well but also the regions that are not doing so well. Spatial analytic tools help determine such patterns.

Glossary:

1) Location quotient (LQ) – The formula for LQ is

$$\text{LQ} = \frac{(R1/R2)}{(N1/N2)}$$

Where: R1 = Regional Employment in Industry X

R2 = Total Regional Employment

N1 = National Employment in Industry X

N2 = Total National Employment

If LQ < 1, region is less specialized in industry X, and needs to import goods to satisfy local demand; if LQ = 1, region produces just enough in industry X to satisfy local demand; and if LQ > 1, region is more specialized in industry X and exports the industry's output to other regions.

2) Nearest-Neighbor Index- It is the ratio of the actual mean distance between nearest neighbors in a given area to the expected distance of the random distribution with the same number of points.[8]

N-N Index $= \frac{(\overline{Do})}{(\overline{Dr})} = \frac{\overline{Do}}{1/2\sqrt{\frac{N}{A}}}$; Where Do is the mean distance, N is the number of points or counties and A is the area of continental U.S.

3) Moran's I- The formula is similar to the formula of correlation coefficient with spatial weight matrices or correlation coefficient between a variable and its spatial lag. $I = (N/So)\sum i \sum j W_{ij} \cdot Z_i \cdot Z_j / \sum_i Z_i^2$; Where Zi = $X_i - \mu$ and So = $\sum i \sum j * W_{ij}$
4) First and Second-order Spatial Process- Spatial data can have large variations in the mean values across the geographic space or some global trend known as the first-order spatial process. The second-order process is more localized where location of one event is influencing co-location of another event, such as distribution of contagious diseases or the technology-based occupations. It is important for spatial analyst to detect if a spatial clustering is due to first-order or second-order process or due to both the processes.
5) Local Moran's I- It was developed by Luc Anselin as Local Indicator of Spatial Association (LISA) statistic. Local Moran's I detect spatial clusters at the local level.

[8] Mahmood, Aslam and Moonis Raza, *Statistical Methods in Geographical Studies* (New Delhi: Rajesh Publications, 1986), 73.

The formula is $\frac{z_i}{s_x^2} * \sum_j [w_{ij}(x_j - \bar{x})]$; where Zi = $x_i - \mu$ and $s_x^2 = \sum_j (x_j - \bar{x})^2 / n$

References:

1) Anselin, Luc. *Exploring Spatial Data with GeoDa: A Workbook*. Spatial Analysis Laboratory, Department of Geography, University of Illinois, Urbana-Champaign, Urbana, IL.
2) ArcGIS 9.2 Desktop Help. http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=welcome
3) Cordes, Sam, Jerry Conover, Christine Nolan et al. "Crossing the Next Regional Frontier: Information and Analytics Linking Regional Competitiveness to Investment in a Knowledge-based Economy." Economic Development Administration, 2009, http://www.statsamerica.org/innovation/report_next_regional_frontier_2009.html.
4) Cordes, Sam, Jerry Conover, Christine Nolan, Brigitte Waldorf et al. "Unlocking Rural Competitiveness: The Role of Regional Clusters." Economic Development Administration, 2007, http://www.statsamerica.org/innovation/report_role_of_regional_clusters_2007.html.
5) Nolan Christine and Indraneel Kumar. "Industrial Cluster Analysis Using GIS: A Case Study of Indiana." Paper presented at the ESRI International User Conference, 2006.
6) De Smith, Michael, Michael Goodchild and Paul Longley. "First-and Second-Order Processes." In *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, http://www.spatialanalysisonline.com.
7) Feser, E., and J. Koo. 2001. *Labor-based industry clusters*. Unpublished manuscript, University of North Carolina at Chapel Hill, Department of City and Regional Planning.
8) Feser, E. 2003. What regions do rather than make: A proposed set of knowledge-based occupation clusters. *Urban Studies* 40 (10): 1937-1958.
9) Fotheringham, Stewart. *2003 Workshops Video Clips- Geographically Weighted Regression and Associated Statistics. CSISS - Summer Workshops 2003 Video Clips - Geographically Weighted Regression and Associated Statistics* (http://www.csiss.org/streaming_video/2003/).
10) Kumar, Indraneel. "Innovation Clusters: A Study of Patents and Citations." Paper presented at the ESRI International User Conference, 2008.
11) Mahmood, Aslam, and Moonis Raza. *Statistical Methods in Geographical Studies*. New Delhi: Rajesh Publications, 1986.
12) Markusen, A. and E. Barbour. 2003. California's occupational advantage. Working Paper No. 12. Accessed from International Relations and Security Network, www.isn.ethz.ch/isn/Digital_Library/Publications/.

13) Wang, Fahui. *Quantitative Methods and Applications in GIS*. New York: CRC Press, 2006.