

Simultaneously Trivializing and Complicating GIS

By Joseph K. Berry (in GeoWorld, April 2012)¹

Several things seem to be coalescing in my mind (or maybe colliding is a better word). GIS has moved up the technology adoption curve from *Innovators* in the 1970s to *Early Adopters* in the 80s, to *Early Majority* in the 90s, to *Late Majority* in the 00s and is poised to capture the *Laggards* this decade. Somewhere along this progression, however, the field seems to have bifurcated along technical and analytical lines.

The lion's share of this growth has been GIS's ever expanding capabilities as a "technical tool" for corralling vast amounts of spatial data and providing near instantaneous access to remote sensing images, GPS navigation, interactive maps, asset management records, geo-queries and awesome displays. In just forty years GIS has morphed from boxes of cards passed through a window to a megabuck mainframe that generated page-printer maps, to today's sizzle of a 3D fly-through rendering of terrain anywhere in the world with back-dropped imagery and semi-transparent map layers draped on top—all pushed from the cloud to a GPS enabled tablet or smart phone. What a ride!

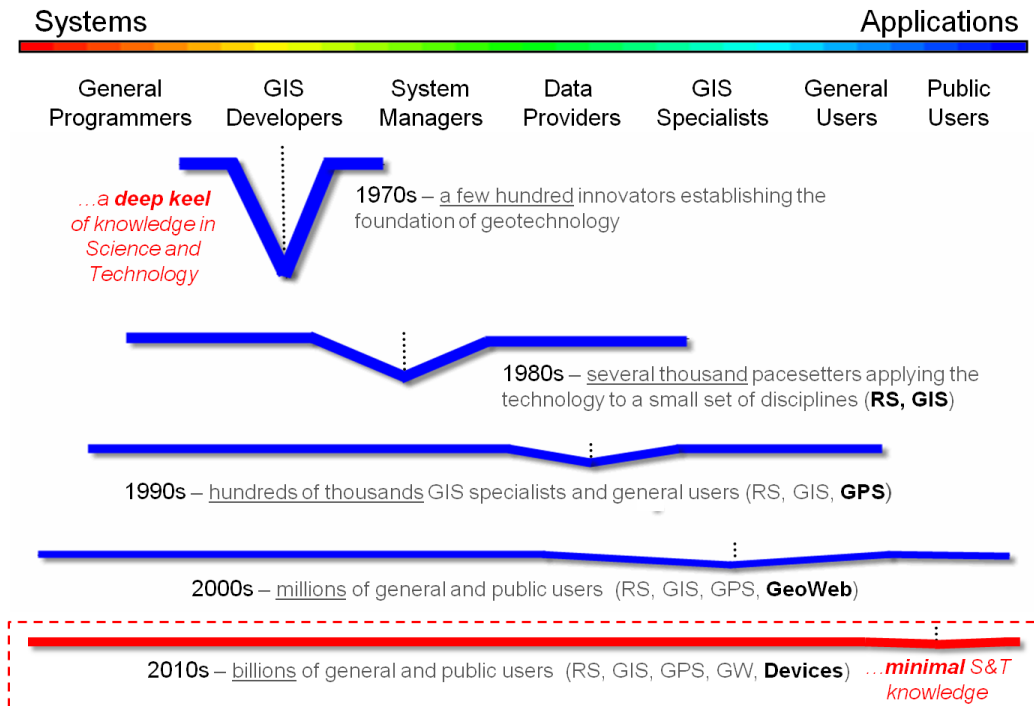


Figure 1. Changes in breadth and depth of the community.

However, GIS as an "analytical tool" hasn't experienced the same meteoric rise—in fact it might be argued that the analytic side of GIS has somewhat stalled over the last decade. I suspect that in large part this is due to the interests, backgrounds, education and excitement of the ever enlarging GIS tent. Several years ago (see figure 1 and author's note 1) I described the changes in breadth and depth of the community as flattening from the 1970s through the 2000s. By sheer numbers, the balance point has been shifting to the right toward general and public users with commercial systems responding to market demand for more technological advancements.

The 2010s will likely see billions of general and public users with the average depth of science and technology knowledge supporting GIS nearly “flatlining.” Success stories in quantitative map analysis and modeling applications have been all but lost in the glitz n’ flash of the technological whirlwind. The vast potential of GIS to change how society perceives maps, mapped data and their use in spatial reasoning and problem solving seems relatively derailed.

In a recent editorial in *Science* entitled *Trivializing Science Education*, Editor-in-Chief Bruce Alberts laments that “Tragically, we have managed to simultaneously trivialize and complicate science education” (author’s note 2). A similar assessment might be made for GIS education. For most students and faculty on campus, GIS technology is simply a set of highly useful apps on their smart phone that can direct them to the cheapest gas for tomorrow’s ski trip and locate the nearest pizza pub when they arrive. Or it is a Google fly-by of the beaches around Cancun. Or a means to screen grab a map for a paper on community-based conservation of howler monkeys in Belize.

To a smaller contingent on campus, it is career path that requires mastery of the mechanics, procedures and buttons of extremely complex commercial software systems for acquiring, storage, processing, and display spatial information. Both perspectives are valid. However neither fully grasps the radical nature of the digital map and how it can drastically change how we perceive and infuse spatial information and reasoning into science, policy formation and decision-making—in essence, how we can “think with maps.”

A large part of missing the mark on GIS’s full potential is our lack of “reaching” out to the larger science, technology, engineering and math (STEM) communities on campus by insisting 1) that non-GIS students interested in understanding map analysis and modeling must be tracked into general GIS courses that are designed for GIS specialists, and 2) that the material presented primarily focuses on commercial GIS software mechanics that GIS-specialists need to know to function in the workplace.

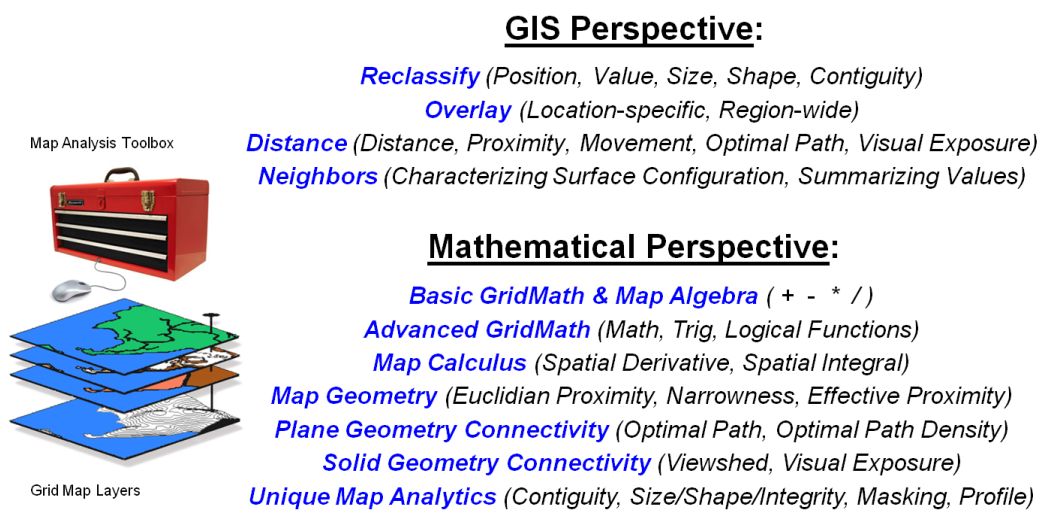


Figure 2. Alternative frameworks for quantitative map analysis.

Much of the earlier efforts in structuring a framework for quantitative map analysis has focused on how the analytical operations work within the context of *Focal*, *Local* and *Zonal* classification

by Tomlin, or even my own the *Reclassify, Overlay, Distance* and *Neighbors* classification scheme (see top portion of figure 2 and author’s note 3). The problem with these structuring approaches is that most STEM folks just want to understand and use the analytical operations properly—not appreciate the theoretical geographic-related elegance, or code the algorithm.

The bottom portion of figure 2 outlines restructuring of the basic spatial analysis operations to align with traditional mathematical concepts and operations (author’s note 4). This provides a means for the STEM community to jump right into map analysis without learning a whole new lexicon or an alternative GIS-centric mindset. For example, the GIS concept/operation of *Slope*= spatial “derivative”, *Zonal functions*= spatial “integral”, *Eucdistance*= extension of “planimetric distance” and the Pythagorean Theorem to proximity, *Costdistance*= extension of distance to effective proximity considering absolute and relative barriers that is not possible in non-spatial mathematics, and *Viewshed*= “solid geometry connectivity”.

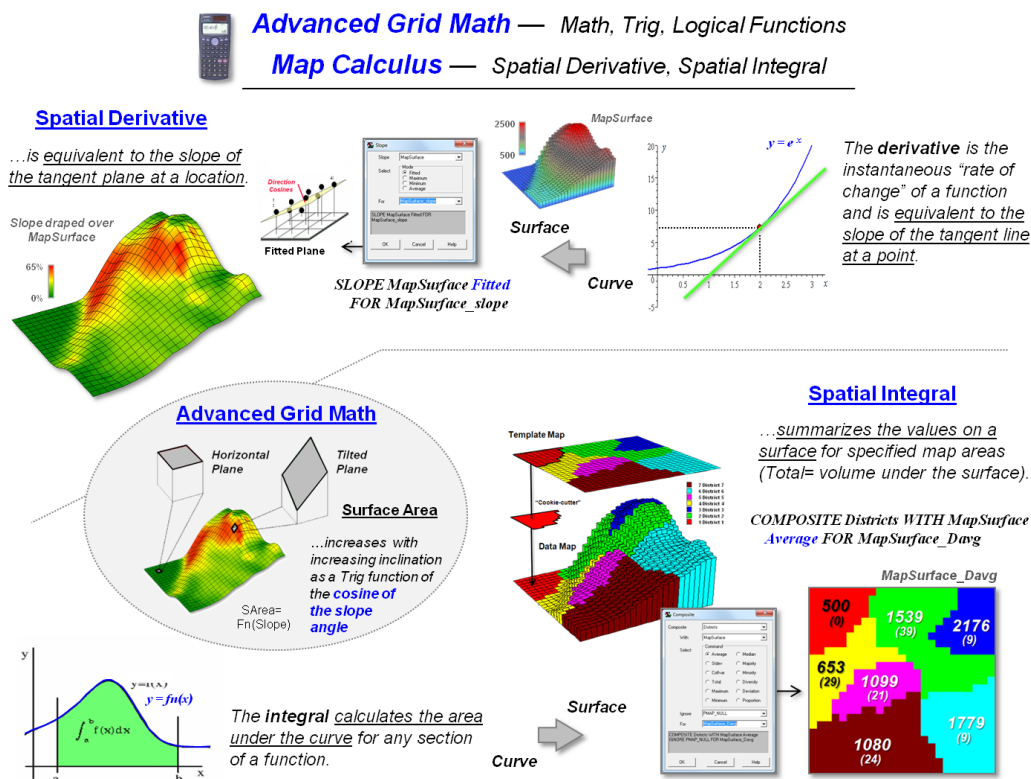


Figure 3. Conceptual extension of derivative, trigonometric functions and integral to mapped data and map analysis operations.

Figure 3 outlines the conceptual development of three of these operations. The top set of graphics identifies the *Calculus Derivative* as a measure of how a mathematical function changes as its input changes by assessing the slope along a curve in 2-dimensional abstract space—calculated as the “slope of the tangent line” at any location along the curve. In an equivalent manner the *Spatial Derivative* creates a slope map depicting the rate of change of a continuous map variable in 3-dimensional geographic space—calculated as the slope of the “best fitted plane” at any location along the map surface.

Advanced Grid Math includes most of the buttons on a scientific calculator to include trigonometric functions. For example, calculating the “cosine of the slope values” along a terrain surface and then multiplying times the planimetric surface area of a grid cell will solve for the increased real-world surface area of the “inclined plane” at each grid location.

The *Calculus Integral* is identified as the “area of a region under a curve” expressing a mathematical function. The *Spatial Integral* counterpart “summarizes map surface values within specified geographic regions.” The data summaries are not limited to just a total but can be extended to most statistical metrics. For example, the average map surface value can be calculated for each district in a project area. Similarly, the coefficient of variation ((Stdev / Average) * 100) can be calculated to assess data dispersion about the average for each of the regions.

By recasting GIS concepts and operations of map analysis within the general scientific language of math/stat we can more easily educate tomorrow’s movers and shakers in other fields in “spatial reasoning”—to think of maps as “mapped data” and express the wealth of quantitative analysis thinking they already understand on spatial variables.

Innovation and creativity in spatial problem solving is being held hostage to a trivial mindset of maps as pictures and a non-spatial mathematics that presuppose mapped data can be collapsed to a single central tendency value that ignores the spatial variability inherent in the data. Simultaneously, the “build it (GIS) and they will come (and take our existing courses)” educational paradigm is not working as it requires potential users to become a GIS’perts in complicated software systems.

GIS must take an active leadership role in “leading” the STEM community to the similarities/differences and advantages/disadvantages in the quantitative analysis of mapped data—there is little hope that the STEM folks will make the move on their own. Next month we’ll consider recasting spatial statistics concepts and operations into a traditional statistics framework.

Author’s Notes: 1) see “A Multifaceted GIS Community” in Topic 27, *GIS Evolution and Future Trends* in the online book *Beyond Mapping III*, posted at www.innovativegis.com. 2) Bruce Alberts in *Science*, 20 January 2012:Vol. 335 no. 6066 p. 263. 3) see “An Analytical Framework for GIS Modeling” posted at www.innovativegis.com/basis/Papers/Other/GISmodelingFramework/. 4) see “SpatialSTEM: Extending Traditional Mathematics and Statistics to Grid-based Map Analysis and Modeling” posted at www.innovativegis.com/basis/Papers/Other/SpatialSTEM/.

Part 2—

Infusing Spatial Character into Statistics

By Joseph K. Berry (in *GeoWorld*, May 2012)¹

The previous section discussed the assertion that we might be simultaneously trivializing and complicating GIS. At the root of the argument was the contention that “innovation and creativity in spatial problem solving is being held hostage to a trivial mindset of maps as pictures and a nonspatial mathematics that presuppose mapped data can be collapsed into a single central-tendency value that ignores the spatial variability inherent in data.”

The discussion described a mathematical framework that organizes the spatial analysis toolbox into commonly understood mathematical concepts and procedures. For example, the GIS concept/operation of *Slope*= spatial “derivative,” *Zonal functions*= spatial “integral,” *Eucdistance*= extension of “planimetric distance” and the Pythagorean Theorem to proximity, *Costdistance*= extension of distance to effective proximity considering absolute and relative barriers that is not possible in non-spatial mathematics, and *Viewshed*= “solid geometry connectivity.”

This section does a similar translation to describe a statistical framework for organizing the spatial statistics toolbox into commonly understood statistical concepts and procedures. But first we need to clarify the differences between spatial analysis and spatial statistics. *Spatial analysis* can be thought of as an extension of traditional mathematics involving the “contextual” relationships within and among mapped data layers. It focuses on geographic associations and connections, such as relative positioning, configurations and patterns among map locations.

Spatial statistics, on the other hand, can be thought of as an extension of traditional statistics involving the “numerical” relationships within and among mapped data layers. It focuses on mapping the variation inherent in a data set rather than characterizing its central tendency (e.g., average, standard deviation) and then summarizing the coincidence and correlation of the spatial distributions.

GIS Perspective:

Surface Modeling (*Density Analysis, Spatial Interpolation, Map Generalization*)
Spatial Data Mining (*Descriptive, Predictive, Prescriptive*)



Statistical Perspective:

Basic Descriptive Statistics (*Min, Max, Median, Mean, StDev, etc.*)
Basic Classification (*Reclassify, Binary/Ranking/Rating Suitability*)
Map Comparison (*Joint Coincidence, Statistical Tests*)
Unique Map Descriptive Statistics (*Roving Window Summaries*)
Surface Modeling (*Density Analysis, Spatial Interpolation*)
Advanced Classification (*Map Similarity, Maximum Likelihood, Clustering*)
Predictive Statistics (*Map Correlation/Regression, Data Mining Engines*)

Figure 1. Alternative frameworks for quantitative map analysis.

The top portion of figure 1 identifies the two dominant GIS perspectives of spatial statistics—*Surface Modeling* that derives a continuous spatial distribution of a map variable from point sampled data and *Spatial Data Mining* that investigates numerical relationships of map variables.

The bottom portion of the figure outlines restructuring of the basic spatial statistic operations to align with traditional non-spatial statistical concepts and operations (see author’s note). The first three groupings are associated with general descriptive statistics, the middle two involve unique spatial statistics operations and the final two identify classification and predictive statistics.

Figure 3 depicts the non-spatial and spatial approaches for characterizing the distribution of mapped data and the direct link between the two representations. The left side of the figure illustrates non-spatial statistics analysis of an example set of data as fitting a standard normal

curve in “data space” to assess the central tendency of the data as its average and standard deviation. In processing, the geographic coordinates are ignored and the typical value and its dispersion are assumed to be uniformly (or randomly) distributed in “geographic space.”

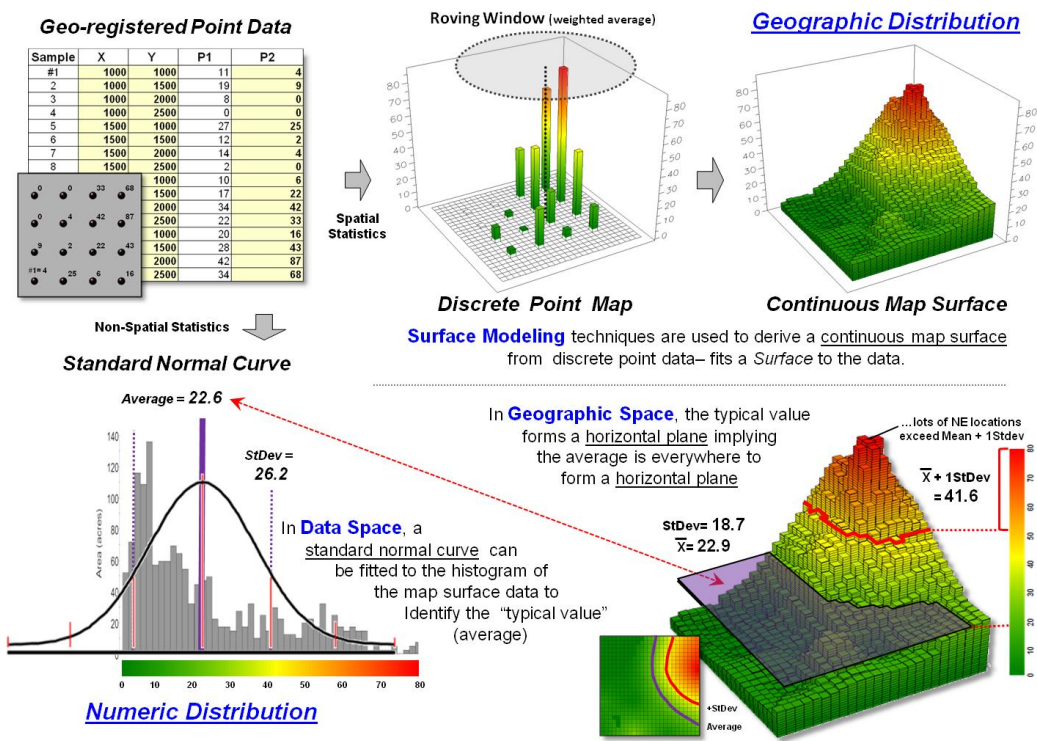


Figure 2. Comparison and linkage between spatial and non-spatial statistics

The top portion of figure 3 illustrates the derivation of a continuous map surface from geo-registered point data involving spatial autocorrelation. The discrete point map locates each sample point on the XY coordinate plane and extends these points to their relative values (higher values in the NE; lowest in the NW). A roving window is moved throughout the area that weight-averages the point data as an inverse function of distance—closer samples are more influential than distant samples. The effect is to fit a surface that represents the geographic distribution of the data in a manner that is analogous to fitting a SNV curve to characterize the data’s numeric distribution. Underlying this process is the nature of the sampled data which must be numerically quantitative (measurable as continuous numbers) and geographically isopleth (numbers form continuous gradients in space).

The lower-right portion of figure 3 shows the direct linkage between the numerical distribution and the geographic distribution views of the data. In geographic space, the “typical value” (average) forms a horizontal plane implying that the average is everywhere. In reality, the average is hardly anywhere and the geographic distribution denotes where values tend to be higher or lower than the average.

In data space, a histogram represents the relative occurrence of each map value. By clicking anywhere on the map, the corresponding histogram interval is highlighted; conversely, clicking anywhere on the histogram highlights all of the corresponding map values within the interval. By selecting all locations with values greater than + 1SD, areas of unusually high values are located—a technique requiring the direct linkage of both numerical and geographic distributions.

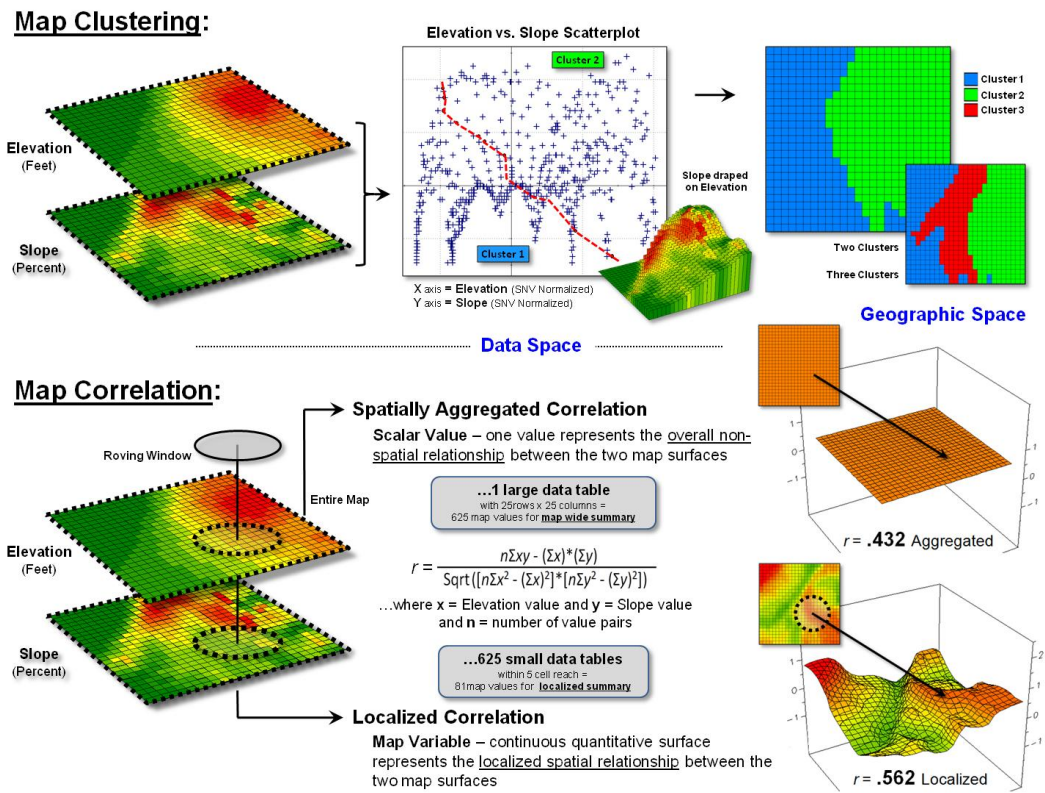


Figure 3. Conceptual extension of clustering and correlation to mapped data and analysis.

Figure 3 outlines two of the advance spatial statistics operations involving spatial correlation among two or more map layers. The top portion of the figure uses *map clustering* to identify the location of inherent groupings of elevation and slope data by assigning pairs of values into groups (called clusters) so that the value pairs in the same cluster are more similar to each other than to those in other clusters.

The bottom portion of the figure assesses map correlation by calculating the degree of dependency among the same maps of elevation and slope. Spatially “aggregated” correlation involves solving the standard correlation equation for the entire set of paired values to represent the overall relationship as a single metric. Like the statistical average, this value is assumed to be uniformly (or randomly) distributed in “geographic space” forming a horizontal plane.

“Localized” correlation, on the other hand, maps the degree of dependency between the two map variables by successively solving the standard correlation equation within a roving window to generate a continuous map surface. The result is a map representing the geographic distribution of the spatial dependency throughout a project area indicating where the two map variables are highly correlated (both positively, red tones; and negatively, green tones) and where they have minimal correlation (yellow tones).

With the exception of unique Map Descriptive Statistics and Surface Modeling classes of operations, the grid-based map analysis/modeling system simply acts as a mechanism to spatially organize the data. The alignment of the geo-registered grid cells is used to partition and arrange the map values into a format amenable for executing commonly used statistical

equations. The critical difference is that the answer often is in map form indicating where the statistical relationship is more or less than typical.

While the technological applications of GIS have soared over the last decade, the analytical applications seem to have flat-lined. The seduction of near instantaneous geo-queries and awesome graphics seem to be masking the underlying character of mapped data— that maps are numbers first, pictures later. However, grid-based map analysis and modeling involving Spatial Analysis and Spatial Statistics is, for the larger part, simply extensions of traditional mathematics and statistics. The recognition by the GIS community that quantitative analysis of maps is a reality and the recognition by the STEM community that spatial relationships exist and are quantifiable should be the glue that binds the two perspectives. That reminds me of a very wise observation about technology evolution—

“Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road.”
~Stewart Brand, editor of the *Whole Earth Catalog*

Author's Note: for a more detailed discussion, see “SpatialSTEM: Extending Traditional Mathematics and Statistics to Grid-based Map Analysis and Modeling” posted at www.innovativegis.com/basis/Papers/Other/SpatialSTEM/.

¹Both topics were first published in GeoWorld (April and May 2012) and subsequently compiled into the online book [Beyond Mapping III](#), Topic 24, *Overview of Spatial Analysis and Statistics* www.innovativegis.com/basis/MapAnalysis/Topic24/Topic24.htm

Joseph K. Berry
Keck Scholar in Geosciences, University of Denver
Adjunct Faculty in Natural Resources, Colorado State University
Principal, Berry and Associates // Spatial Information Systems
1701 Lindenwood Drive, Fort Collins, Colorado, 80525
Phone: 970-215-0825 Email: jberry@innovativegis.com
Website: <http://www.innovativegis.com/basis/>