From Statistical to Diagrammatic
Geo-spatial & time-based data visualization through parametric modeling

Ming Tang[1], Chris Auffrey[2], Mingming Lu[3]

[1]University of Cincinnati
SAID, College of DAAP
Cincinnati, OH. 45221
tangmg@ucmail.uc.edu

[2] University of Cincinnati
SOP, College of DAAP
Cincinnati, OH. 45221
auffrec@ucmail.uc.edu

[3] University of Cincinnati
SEEBME, College of Engineering
Cincinnati, OH. 45221
lumg@ucmail.uc.edu

Keywords: parametric, diagrammatic, Geo-spatial, visulization.

Abstract

The paper describes experiential learning outcomes in the application of geospatial data to generate diagrammatic representations. It also focuses on models and animations used in engineering, architecture and urban planning to visualize specific urban issues such as urban health, education, crime, and air pollution, as they are currently being confronted in major cities such as Cincinnati. The paper discusses how to reconstruct geospatial and time-based data into various representations. It describes the process of visualizing and representing databases with emerging parametric modeling and animation tools. Starting with statistical data, parametric design tools are used to decode and recode the social, cultural, economic and environmental complexity within the parametric equation. Computational tools for architects and urban designers such as Geographic Information Systems (GIS), Excel, Maya, Rhino and Grasshopper plugin are used to build platforms that allow parametric control of the rendered outcome.

1. INTRODUCTION

With the abundance of data visualization technology developed over the last few years, infographic techniques have evolved from simple charts and maps to sophisticated 3D models, animations and interactive media. As for designers, it is crucial to identify the appropriate visualization for the data set by investigating graphical features as well as their roles in the data representation. This paper examines the result of an ongoing collaborative interdisciplinary research project on geospatial data visualization involving faculty and students from School of Architecture and Interior Design (SAID), School of Planning (SOP), and School of Energy, Environmental, Biological and Medical Engineering (SEEBME) at University of Cincinnati.

Several student projects used parametric modeling methods to determine spatial patterns, clusters, networks, hierarchies, as well as the changing typology over time. Yet, the essential goal of this research is not constructing a mathematical model for the original data set, but rather applying a diagrammatic model and translating the complex data set into an abstract form. The geospatial data used, for factors such as air pollution, health, population, housing and employment, are presented through three dimensional and time-based media.

1.1. What is the statistical data visualization?

For architects and urban designers, a design process usually starts with data mining and site analysis. It is essential to collect geospatial information and visualize it in a meaningful way to stimulate design solutions. Although the data format might be different, the essential nature of various geospatial data at this stage is usually expressed as statistical distribution. The statistics such as mean, median and outliers need to be computed at the associated geographic unit (e.g. parcel, block, neighborhood, city, etc). Typically two possible distributions are compared by graphing quantities against each other and then studying the generated patterns. This allows the viewers to see whether the plot values are similar and if the two distributions are related.

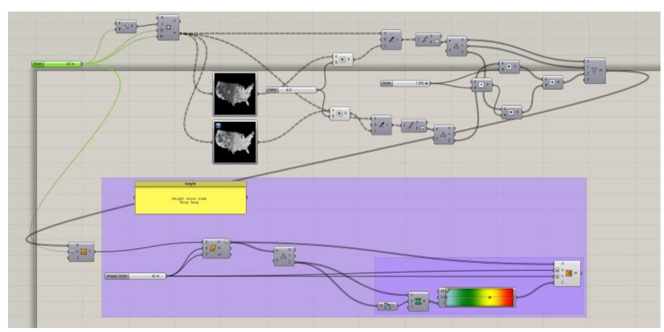
There are a variety of conventional methods for data presentation and infographics techniques such as tables, histograms, pie charts and bar graphs. However, architects and urban designers are always interested in finding alternative methods to visualize data from a designers' mindset. There are a growing number of creative design-related methods for visualizing data using the emerging digital tools in the built environment design field. Many of these new data representations are generated by directly feeding the statistical data into the digital design tools.

1.2. What is diagrammatic urban model?

Parametric modeling is increasingly implemented in genetic computing to create 3D models with a high degree of complexity, such as the parametric urbanism by Patrick Schumacher, *City Engine* by Pascal Mueller and urban ontology by José Beirão. Beirão described the process of using shape grammer and pattern to study the “topological relations between object classes, object types with specific topologies and attributes to condition their behavior and relationships” (Beirão 2008). Inspired by their theories and research, as well as data driven urban modeling techniques, a graduate course was developed to utilize a set of visualization principles to generate abstract urban models driven by a variety of datasets. The goal is to represent the statistical information, such as population census data with new methods, and explore the relationships among urban elements such as crime, pollution and various social parameters. “The geospatial database can provide a rich resource to optimize urban forms with respect to ecological performance criteria. The demographic, traffic and economic data from GIS provides the trace of activity and event parameters of the urban life process” (Tang 2011). In an experimental project based in the Cincinnati area, the population density data is defined as the only rule to control the height of each parcel (Figure 1). Although the generated model does not reflect any real typology in this area due to its singular parameter, it does allow the viewers to observe complex 3D urban patterns that communicate the information clearly. This is similar to how contour lines and digital elevation models (DEM) are used to represent 3D topography. These types of translations deliver meanings by manipulating data across various formats and structures, which allow points, lines, surfaces and masses to be interpreted as the diagrammatic objects. In this modeling approach, an abstract urban model is parametrically constructed by various inputs, which are not normally associated with any urban design vocabulary.

Figure 1: In order to integrate population density into parametric modeling tools so as to construct an urban model parametrically, the population density is visualized by the height of each block. Created by student Jesse Larkins.

There are two approaches that can be used to visualize urban data within a parametric model. First, 2D maps can be used to generate a 3D diagrammatic model. These thematic maps contain GIS data for zoning, population, transportation, and other spatial information. Parametric tools such as Maya or Grasshopper can read the 2D information embedded in the GIS maps based on the RGBA value of each pixel and execute corresponding modeling operation. The resultant model inherits all the geographical, social, demographic and design information from the 2D GIS maps. Integrating multiple 2D GIS maps allows modeling of changes in the data over time. This change can be effectively displayed as a continuously played video clip. (Figure 2)



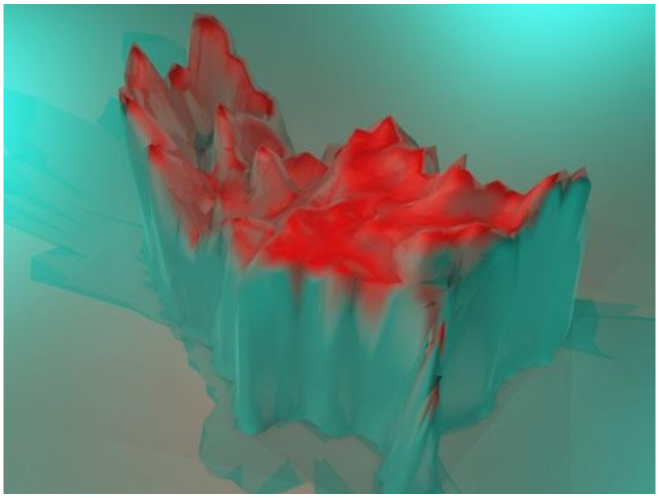


Figure 2: Population change of United States. Animated 3D form. Created by student Kuang Li.

The second approach involves using script to stream an Excel database directly to execute parametric modeling operations automatically and thus replacing more labor-intensive approaches. Designers only need to import the database, observe and evaluate the 3D result, and optimize the result by manipulating a limited number of variables. Because this process only uses text format data, the processing time from an Excel table to animation is easily manipulated back and forth, though it is less intuitive compared with previous mapping approaches (Figure 3).

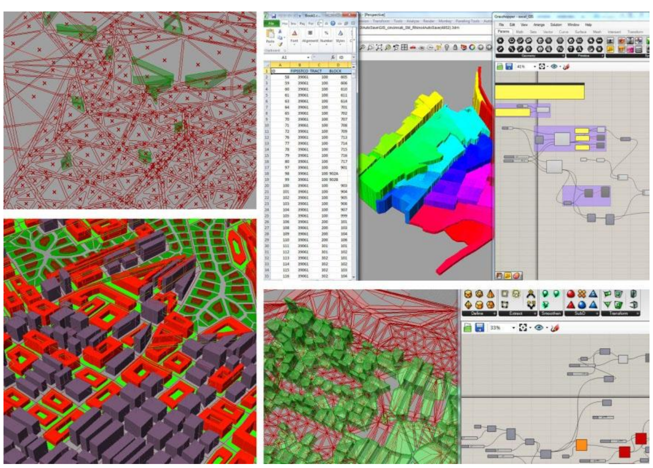


Figure 3: Excel files were streamed into parametric modeling process.

2. THE PARAMETRIC CONNECTION BETWEEN STATISTICAL DATA AND DIAGRAMMATIC URBAN MODEL

One of the objectives of connecting statistical data to a diagrammatic model is to create an engaging experience that allows designers to control the final form in a dynamically changing interface. This permits designers to visualize a large numbers of representation options in a relatively short period of time. As the parametric model is generated very quickly, it is advantageous for the designers to be aware of the following essential categories of visualization.

2.1. From time based data to animated form

The computational method is extended by exploring, collecting, analyzing, and visualizing urban information and interactively representing the information through digital technologies. As mentioned previously, a diagrammatic model allows complex conditions to be represented in a simplified form so that it might be more easily understood and managed. For instance, the designers would be able to read certain correlation patterns over a period of time.

Time-series data is frequently used in data visualization. Index charts and stacked graphs are the conventional ways to illustrate the relative changes over time. However, it is difficult to accurately interpret trends over a complex spatial pattern. An alternative way is to use morphing forms so the overall trends can be more easily interpreted. The 3D animation can become very powerful when it is used to illustrate geospatial data within a defined coordination system. The animated objects can be constructed based on the geographical geometries such as parcels, blocks, counties or states. Animation becomes a natural way to represent the changing data. Time and space can be depicted through the use of lines, shapes, colors and other visual elements. For instance, the animated magnetic field encodes data through the attractors within a geographical region (Figure 4). Using a variety of data, whether from a map sequence or multi-column Excel database, the resulting animation allows for a larger number of dimensions to be represented.

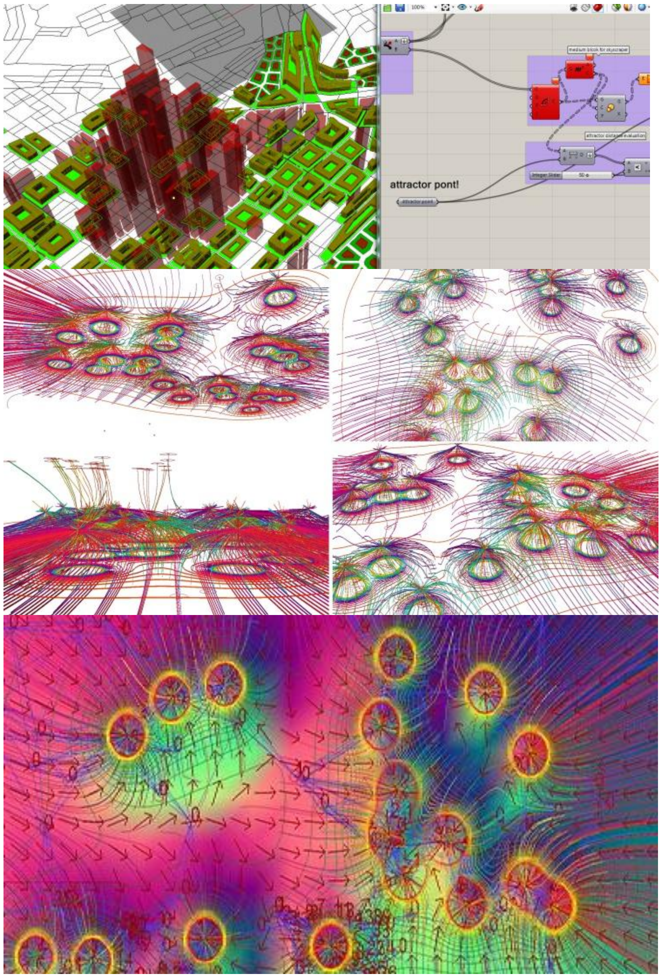


Figure 4: 3D field generated based on census data over time.

2.2. Hierarchy and networks

Simple lines with added values can be interpreted into many possible spatial features such as transportation networks or social connections, depending on their positions in the hierarchy. Geospatial data can be organized into polygon-based or line-based hierarchies. States, counties, tracts, blocks and parcels are common polygon hierarchies for data organization. Streams and transportation networks are common line-based hierarchies. One can also use space-filling variants such as point density to generate new hierarchies. Instead of generating hierarchies based on geographic organization, a large numbers of spatial nodes can be drawn as a point cloud and grouped based on their adjacency or distance. For instance, a Delaunay mesh grid can be constructed based on a group of points. This new hierarchy can be used to connect points into a web-like network to represent spatial linkage. Each point on the web has a (x,y,z) value, as well as its neighboring nodes' index values. By using color and saturation instead of text, values associated with the links can be perceived visually (Figure 5).

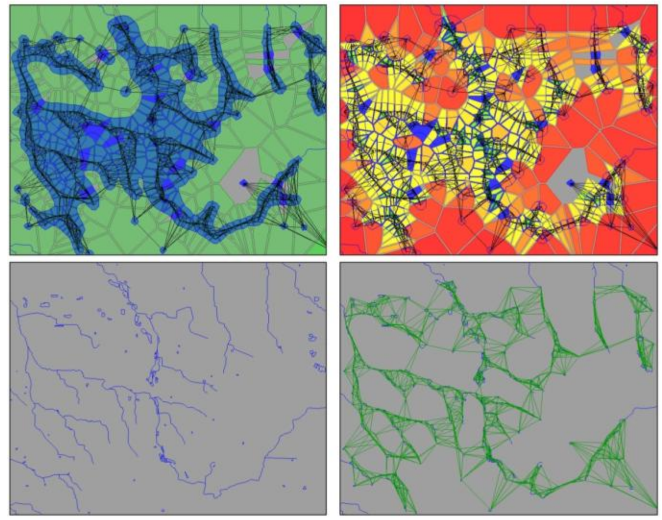


Figure 5: Spatial hierarchies of polygons, lines and points.

3. PROJECT

Based on a course taught at the University of Cincinnati in 2012, the authors investigated methods for integrating three disciplinary perspectives (Architecture, Planning and Environmental Engineering) within the context of data visualization and analysis. Students used theories, models and methods of modeling and visualization to examine various data representation. The course encouraged students to question the role of parametric design for integrating visualization as a diagrammatic instrument for inspiring hybrid conceptualizations of logical urban forms.

3.1. Project: Gender & education

Students started by collecting data about the educational attainment level by gender within one county based on the US Census data. Students were able to use GIS to symbolize the data as a dot density map with each dot representing five persons. Students then used GIS data of educational attainment level in dot density and extracted the center point of each dot in a Rhino program. Students defined point fields based on the clusters of points and extracted centroids from these clusters. Finally, a magnetic field was created in Grasshopper. At each level of education attainment, the magnetic field defined by the points changes from negative to positive. The fields move as the points move from the original educational level to the next education level.

Students created two videos: one using data for female educational attainment and the other using data for male educational attainment. In each video, students animated progressively from no school completed to doctorate degree using the following increments: nursery school to fourth grade, fifth to sixth grade, seventh to eighth grade, ninth grade, tenth grade, eleventh grade, twelfth grade with no diploma, high school graduate (and equivalency), some colleges with less than one year completed, some college with more than one year but no degree, associate's degree, bachelor's degree, master's degree, and professional degree (Figure 6).

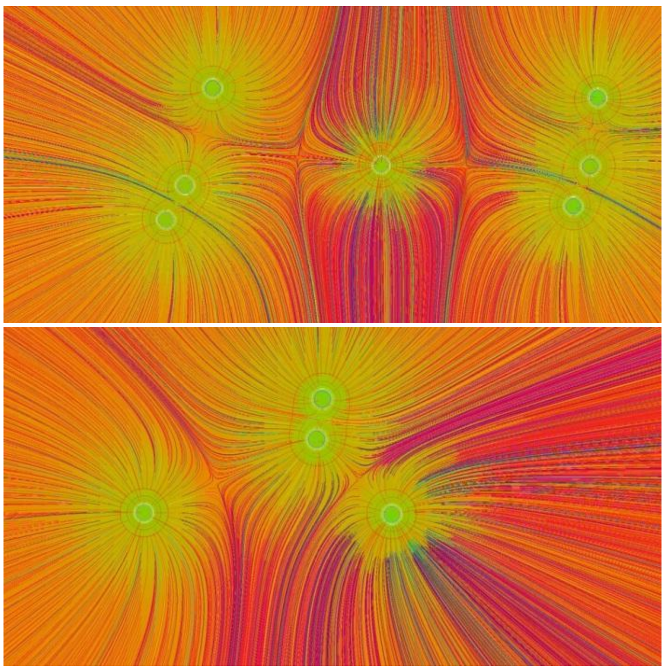


Figure 6: Changing field based on education level. Female, vs male. Created by student Sarah Kusuma.

3.2. Project: Urban crime

Students first located point-based crime data for the City of Cincinnati. Data for Hamilton County was downloaded as Excel files for 2007, 2008, and 2009. Excel was then used to compile and format the data for geocoding. Geocoding is the process of assigning each instance of reported crime to a point in space (on a map). This was done through a custom address locator, based on the street grid layer, which matches the street, house number, city and zipcode in the original crime table with a physical, real address existing in Hamilton County. Subsequently, only points occurring within the City of Cincinnati were selected and plotted. This data was used to calculate the incidence (density) of crime by type. Students then performed a basic kernel density analysis, examining and charting statistically significant clusters of the point-based data. To ensure an accurate and unbiased (to the extent possible) analysis of the clusters, the displayed properties of the resulting raster were of utmost importance. The 1/4 standard deviations settings were used, which resulted in 14 classes of clustering, with the first class (0-.00000019) being omitted for graphical clarity. The resulting raster images were exported and brought into Rhino and Grasshopper to construct a 3D morphing surface to represent changes in crime over time and space. This 3D surface was compelling and effectively displays spatial changes in crime over the years. The final result was rendered as several videos (Figure 7).

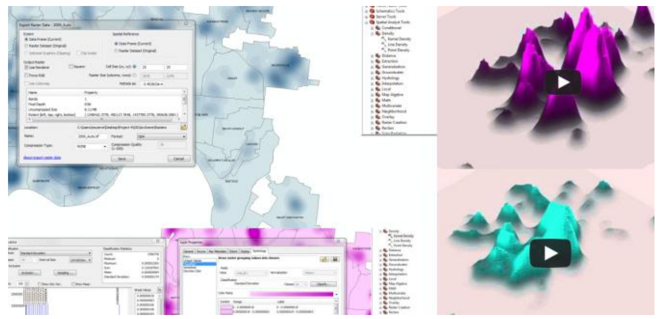


Figure 7: Crime data in Cincinnati, 2002 to 2010. Created by student Craig Moyer.

3.3. Project: Air pollution

In this project, students modeled air pollution in southwest Ohio using data collected from five air quality monitoring stations. The original PM2.5 air pollution data over the past 5 years was formatted as an Excel file. Students geocoded each monitoring station. Then the Excel data was streamed to control five spatial nodes in Rhino. By using one slider, the students manipulated the reading of the Excel file to drive the Z value of these nodes based on daily PM2.5 levels. By combining these spatial nodes, the students created a Grasshopper script to form a 3D surface. The surface became the representation of air pollution across the entire area. The next task was to communicate the PM2.5 daily values by adding a color code to the surface. Students reconstructed another recognizable form for the gradient color script, and created a sequence of stacking contours with darker on the top and lighter at the base. Animated texts were added to highlight values over a one year timeline. In the end, the time parameter was animated to load daily data and compiled as an animation (Figure 8).

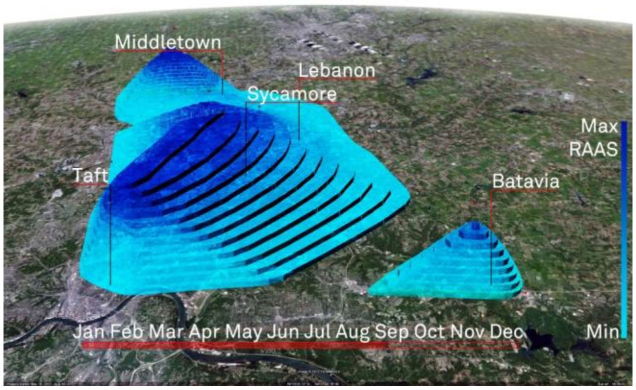


Figure 8: Daily PM5 Particle level in Cincinnati. Animation. Created by student Alexis Payen.

4. CONCLUSION

In all the experiments, social and environmental data from the U.S. Census Bureau were used to develop diagrammatic visualizations based on its geospatial information. The information was streamed and visualized as 3D animations with various shapes and colors to represent the quantifiable values in the database. Customized Grasshopper scripts, as well as advanced modeling tools such as Maya and Rhino were used to construct 3D diagrammatic models from the parameters. Also, information was extracted with image sampling processes and then used to drive the animation in the advanced rendering programs. As a result of the continuous transformation and representation, the visualization presented by the database, 2D maps and 3D models empowered each other.

Parametric modeling in the urban visulization field has allowed designers and planners to observe patterns and analyze spatial relationships. The gap between the abstract data and the graphical content is narrowing due to the increasing availability of digital tools for contructing interactive models based on input parameters. “It created an interesting notion to the parametric urbanism practice and further exploited the idea that design practice begins with the information” (Tang 2011). Incorporating variables from urban demographic, social and environmental data into parametric equations allows for a broader range of information to be integrated into 3D and animation formats. This new parametric procedure has worked as an interactive learning tool for assisting students to better visualize and understand complex urban environments and also serves as an efficient way to transfer the collected data into urban patterns and forms.

Acknowledgements

The participated students from the School of Architecture and Interior Design (SAID) at University of Cincinannati were: Alexis Payen, Anjali Patel, Craig Moyer, Dylan Fischer, Jesse Larkins, Kuang Li, Richard Huizenga, Melissa Long, Michael Rogovin, Nabi Hamdi Mortan, Ross Battoclette, Sarah Kusuma.

Reference

Beirão, José; Duarte, José; Stouffs, Rudi. Structuring a Generative Model for Urban Design: Linking GIS to Shape Grammars. Education and research in Computer Aided Architectural Design in Europe Confrence (Ecaade) 2008

Tang, M. City Generator: GIS Driven Genetic Evolution in Urban Simulation, Proceeding of the 96th ACSA Conference

Tang, M., Anderson, J. Information Urbanism---Parametric urbanism in junction with GIS data processing & fabrication. 2011 Annual Architectural Research Centers Consortium (ARCC) Spring Research Conference. Detroit, MI.