GEOVISUALIZATION OF PUBLIC TRANSPORTATION USING THE SPACE-TIME CLOUD

RUSH J., KWAN M.P.
The Ohio State University, COLUMBUS, UNITED STATES

INTRODUCTION

Public transportation is expected to be efficient and effective in connecting an urban population. For lower-income residents who cannot afford an automobile or have irregular access to one, public transport facilitates getting to work, doctor's appointments, child care, school, or any number of critical activities. Public transport is heavily used by minority populations in the United States, with African-Americans making up 33% of trips (while making up 12% of US population), and is relied upon by lower income families, with those having household incomes of $25,000 a year or less making up about 35% of trips (American Public Transportation Association, 2007).

Despite numerous legislative initiatives to increase the mobility of low income populations, less than a third of the entry level jobs concentrated in suburbs are within a quarter mile of a public transportation stop and poor mobility remains a barrier to employment (Sanchez, 2007; US General Accounting Office [GAO], 1998). The prevailing theory is that there is a spatial mismatch between underemployed central urban residents, who tend to be of minority ethnicity and have low household income, to the entry level jobs which have moved out to the suburbs, which are designed for access from automobiles (Kain, 1968). Public transportation that facilitated a "reverse commute" from the urban core to the suburbs would be expected to increase the accessibility of residents to jobs and a higher overall quality of life.

Spatial coverage is not the whole story, however. Entry-level jobs are more likely to have shifts with schedules that do not align with the standard daytime work hours (Sanchez 2007, GAO 1998). This means that the timing and frequency of stops in a particular place needs to be considered to fully understand the access granted to nearby residents by the infrastructure. This perspective aligns with the philosophy of time geography, originating in the Lund School in the 1960s and championed by Torsten Hägerstrand (1970). The premise of time geography is that time is as important as space in determining the decisions that people are able to make, and thus strongly influences their quality of life.

In beginning to analyze the spatial and temporal characteristics of public transportation in a region, it would be prudent to first conduct some exploratory spatial data analysis (ESDA) to better form hypotheses about the data and more efficiently detect patterns (Anselin, 1993). Previous work has shown that 3D geovisualization is a capable method of ESDA for space-time data (e.g. Kwan & Lee, 2004), and this study attempts to employ a geovisualization environment in the time-geographic framework to detect patterns in the spatio-temporal provisioning of public transportation.

CHOOSING A GEOVISUALIZATION METHOD

Three visualization methods in particular have been frequently used in time-geographic analysis: space-time paths, activity density surfaces, and animated maps. The most fundamental of these is drawing space-time paths of moving objects (usually, people) in a time dimension that is drawn perpendicular to the two spatial dimensions (Hägerstrand, 1970). The technique allows seeing the entire range of positions that an object occupies over a span of time. The 3D space created by the combination of the time dimension with two spatial dimensions is most frequently called the space-time cube or space-time aquarium.

The space-time aquarium is a useful paradigm for viewing inherently multidimensional data without needing to sacrifice the temporal or spatial variation in the data to display it. Displaying space-time paths is perhaps the most faithful representation of mobility data, but when a large number of mobile objects are plotted, the display can become cluttered. Clutter can be reduced by showing only stationary activities, or stations, and removing the sloping parts of space-time paths that connect different activity locations and represent travel time from the display (Kwan, 2000). Removing travel information would not be appropriate in all cases, but in this scenario of analyzing the access to public transportation, it may be. If we conceptualize accessibility to public transport as spatio-temporal proximity to designated stops in the transport system, the paths between the stops are less important.

Another method that is useful for visualizing space-time data is density surfaces. Density surfaces are useful to displaying the concentration of activities in space, or alternatively in terms of distance from a location across time (Dykes & Mountain, 2003; Kwan, 2000) Because density surfaces are raster representations, they can scale to any data volume simply by recording larger values in the raster cells. The scalability is helpful in this case, as the data volume for public transport information can be quite high, but

the generalization of either the time dimension or one spatial dimension would be necessary to create a 2D raster surface.

Animated maps have a natural link to temporal data since they enable displays of data changing over time. Andrienko, Andrienko, and Gatalsky (2005) describe methods for displaying mobile data points with animated maps. Points could be drawn at their new positions at every time step in the animation, and their trajectories could be emphasized by following each point with a path that traces its previous positions. To reduce visual clutter resulting from overplotting of tracing paths, they could be made to disappear after a specified number of previous connected locations. Animated maps have are preferable to the plotting of space-time data on a static 2D map, since the data is spread across multiple frames of animation and creates less visual clutter. However, since only a small temporal range of the data is seen at a time, analysts must work to create associations across time mentally, rather than viewing them directly (Andrienko et al., 2005; Ware, 2008).

An ideal exploratory geovisualization method would combine the scalability of density surfaces with the full dimensionality of the space-time aquarium. Other researchers have come to a similar conclusion. Forer (1998) uses a 3D raster format that discretizes time as well as space. The resulting volumes are called taxels, a term derived from the volumetric equivalent voxel. The key purpose of the data structure in Forer's 1998 study was conducting efficient geometric queries on space-time paths and prisms, which required each individual person or object to be visualized one at a time. Follow up work (Forer & Huisman, 2000) showed an aggregate activity space for all individuals in the sample.

This research seeks to employ a method somewhere in between these two approaches. Forer and Huisman's (2000) aggregated activity space showed the potential of a 3D raster data structure to scale to large data volumes, but did not display variation across time. By following a similar procedure to Forer's 1998 study, space-time data can be discretized into a 3D raster, but rather than showing each object or person individually, they could be aggregated to show a total of space-time activity density. We call this approach the discretized space-time cloud.

CALCULATION OF THE DISCRETIZED SPACE-TIME CLOUD

The discretized space-time cloud was originally developed to support the analysis of human activity data (Rush, 2009). A computer program was built to convert the common structure of the human activity diary into a 3D raster representation. The procedure iterates on every unique person, and follows their activities throughout the time span. Based on a user-specified spatial and temporal resolution, the program marks the presence of the individual in the 3D raster cells that their space-time path would intersect with. The procedure is run for each individual in the dataset, summing the counts of individuals together into an aggregate representation of occupied space-time positions.

Transportation data is similar to activity diaries in that unique routes will have a sequential order of stops throughout a day. Where large datasets of activity diaries are expensive and time-consuming to collect, transportation data is now much easier to acquire due to the standardization of the General Transit Feed Specification (GTFS). The GTFS was created by Google to define a standard way of sharing public transit information so that it could be incorporated into its mapping products (Google, 2010). Public transport agencies are not required to make this data publicly downloadable, but many elect to. This study benefits from the openness of the Central Ohio Transit Authority (for Columbus, Ohio) and the Greater Cleveland Regional Transit Authority, which both provided their datasets for download (see http://www.gtfs-data-exchange.com/ for a collection of all known publicly-available transportation data).

The GTFS includes routes, stop locations, and stop times along routes, which are useful for plotting the spatio-temporal provisioning of public transport in a geovisualization environment. After joining these tables together, a table that is equivalent to a space-time activity diary is created, and can be loaded into the discretized space-time cloud software. The program provides an option of whether or not to connect activity locations with straight lines (as with space-time paths when precise GPS tracking data is not available), and this option is turned off for the processing of transportation data, so that only the stops are visualized.

SOFTWARE IMPLEMENTATION

To implement the calculation and display of the discretized space-time cloud, and linked system was created that leveraged the strengths of GIS software and advanced visualization. ESRI ArcGIS 9 is used in the first step of the analysis to handle the spatio-temporal input data. A Visual Basic for Applications (VBA) program is written in the ArcMap program to process the input data, create attribute or temporal queries on it, and output it as a 3D raster file that can be read by the visualization environment. For visualization, a space-time aquarium environment was designed in 3DVIA Virtools 4. The 3D

environment in a Virtools composition begins completely empty, but can be customized by manually adding 3D objects (e.g. a space-time aquarium), through a visual programming language that uses "building blocks," through a native scripting language, and through graphics shaders that determine how objects render on the screen (McCarthy & Callele, 2006).

This implementation is advantageous because it preserves all of the functionality of a desktop GIS system without having to re-invent it, while the external visualization environment provides a great deal more customization than desktop GIS products. An additional advantage of the selection of the Virtools software for visualization is that it runs in a web browser plugin, which helps remove barriers for academic openness. The program can be hosted on a web server, allowing readers of articles to use the same visualization environment that the authors used to reach their own conclusions.

DISPLAY OF THE DISCRETIZED SPACE-TIME CLOUD

After the VBA program in ArcMap creates a 3D raster representation of spatio-temporal public transport provisioning, it is displayed in Virtools as a point cloud in the space-time aquarium to preserve the dimensionality. In this case, we are looking at a typical weekday of public transportation availability, so the vertical temporal axis has a range of one day. Andrienko and Andrienko (2006) recommend that analysts "play" with different levels of aggregation, and Fotheringham (1989) recommends performing an analysis at multiple scales to determine the scale-dependence of the results. To that end, the user is able to customize the spatial and temporal resolution at which the data is rasterized at, which allows for experimentation to detect patterns at different granularities. The calculated area is set by drawing a rectangle around the interest area in ArcMap, and then specifying how many intervals in the X and Y dimension it should be divided into for rasterization. The user also sets the time interval in the software. The default is 100 intervals in the X and Y dimensions, and 15 minute intervals for a dataset of one day's data. At the default resolution, the 3D raster is made up of 960,000 taxels.

Each taxel that contains a stop along a public transit route is visualized in a 3D environment as a point in the aquarium. Rather than establish a single way to view the aggregated 3D raster data, the software supports several options that allow the user to experiment with the data and view it in different ways. The primary means of viewing the data is as a point cloud. By default, each point is partially transparent, and areas with a high density of activity will appear more opaque (Mountain, 2005; Theus, 2005). Each point is also symbolized by increasing size and decreasing transparency as the count of the taxel's transit stops increases, meaning the regions with clusters of taxels of high transport provisioning will appear the most opaque features in the display. The amount of transparency can be scaled by the user on the fly, which is particularly helpful for shifting the emphasis to and from dense (typically downtown) transportation regions and less frequented (typically peripheral) regions.

Even while adjusting the transparency, points closer to the user's viewpoint may obscure patterns that are further away. To better identify clusters and strip away noisy, low-frequency data, a thresholding technique was implemented. This allows the user to change the minimum value a taxel in the 3D raster must have in order to be displayed. Rather than determining the threshold at the rasterization stage, the threshold is applied when displaying the data, so that it can be interactively changed instead of needing to recalculate the data.

The default calculation of the discretized space-time cloud attributes sums of transit stops in each taxel. Depending on the data, patterns may not be easily detectable. Because this is an exploratory visualization technique, a 3D moving window filter was implemented to exaggerate the features and draw visual attention to them. It is calculated as a weighted average in a 3x3x3 moving window, achieving an effect similar to a blur filter in 2D image processing programs (figure 1). While computationally simpler, it achieves an effect analogous to the kernel density surfaces used to smooth activity density surfaces (e.g. Kwan, 2000; Kwan & Lee, 2004). The user is able to toggle between the filtered data with exaggerated features and the original data with meaningful data values.
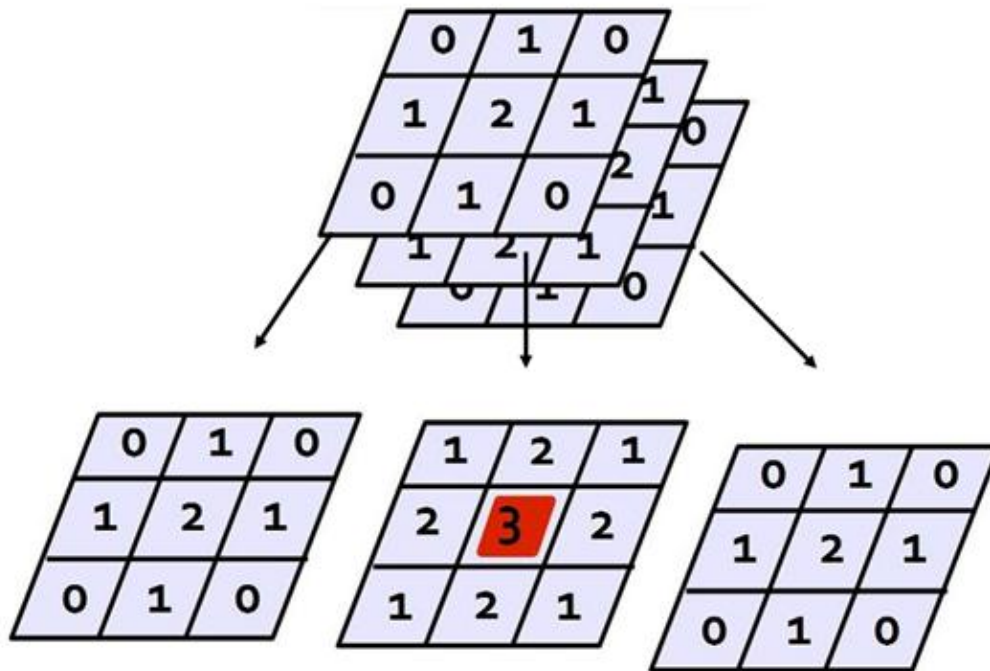
Figure 1 – Schematic of the 3x3x3 moving window filter. The value of each cell is multiplied by the value in the grids in the figure, then divided by the sum. The red highlighted cell is the center of the moving window, and is where the output value is recorded.

The filtered data in the space-time cloud is also functionally similar to the 3D kernel density surfaces created by Demšar and Virrantaus (2010) and displayed as isosurfaces. The key difference is the visual appearance (point cloud instead of surfaces) and calculation complexity. For the discretized space-time cloud, the 3D moving window filter algorithm runs right after the input GTFS or time activity data is converted to a 3D raster. The entire process of data rasterization and filtering takes less than two minutes on a computer with a 2 GHz processor for the Columbus dataset with 170,000 points. The greater demands of the computational geometry algorithms in the isosurface technique required 1.5 hours on an equivalent computer for a dataset of 240 points rasterized at a similar resolution.

Finally, the cube is completed by the inclusion of a base map plane that provides context for the spatio-temporal data in the aquarium (Kwan, 2000). This plane is mobile, so that finding the precise spatial locations that points in the cloud overlay is facilitated by bringing the plane close to the points of interest (Kraak, 2003).

Even with a movable base map plane, the time values along the vertical axis are less clear than the spatial axes that have a base map plotted along them. To provide a means of determining exactly what time particular data points lie at, a "selection ring" was added to the geovisualization environment. This ring can be moved up and down the time axis, and all points in the cloud that lie along its plane are selected and changed to a dark gray, completely opaque color. This makes the selected time slice highly visible regardless of the transparency settings effecting non-selected points. Along the edge of the ring (the walls of the space-time aquarium) are highly-visible text displays indicating the currently-selected time. Without the selection ring, the user would need to repeatedly experiment, realign the view angle, and move the base map plane (which would obscure earlier data) in order to approximate the time interval of a feature of interest.

RESULTS AND DISCUSSION

The discretized-space time cloud was employed as an exploratory geovisualization technique for the transportation data from the cities of Cleveland and Columbus, Ohio, USA. Like any visualization technique, it is ideally employed as a suite of tools, each of which have their strengths and weaknesses and jointly enable a greater understanding of the data. What follows is a brief discussion of the results of visual analysis of the two datasets, differences that were detected, and how further research might proceed.
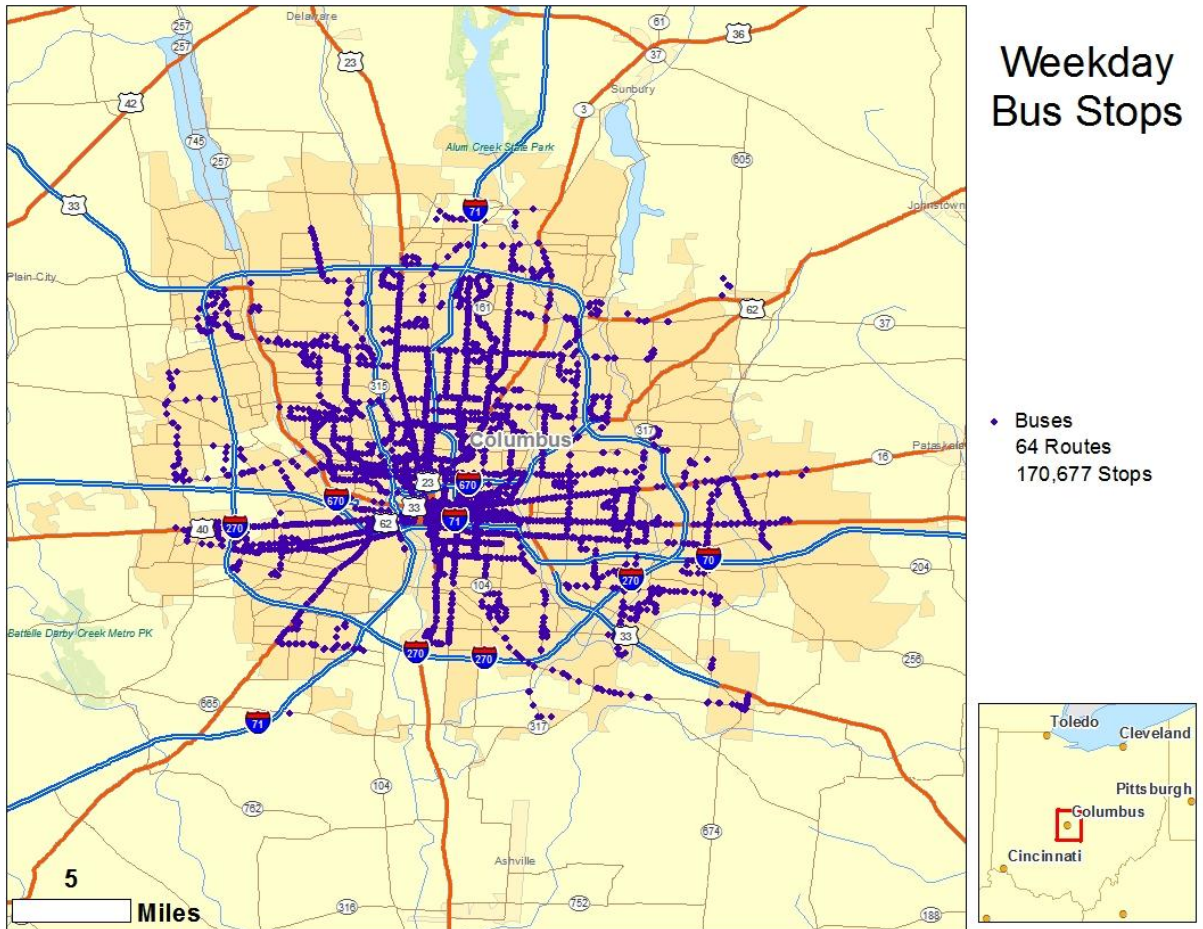
Figure 2 – Spatial distribution of bus service in Columbus, OH. Data as of January, 2011.
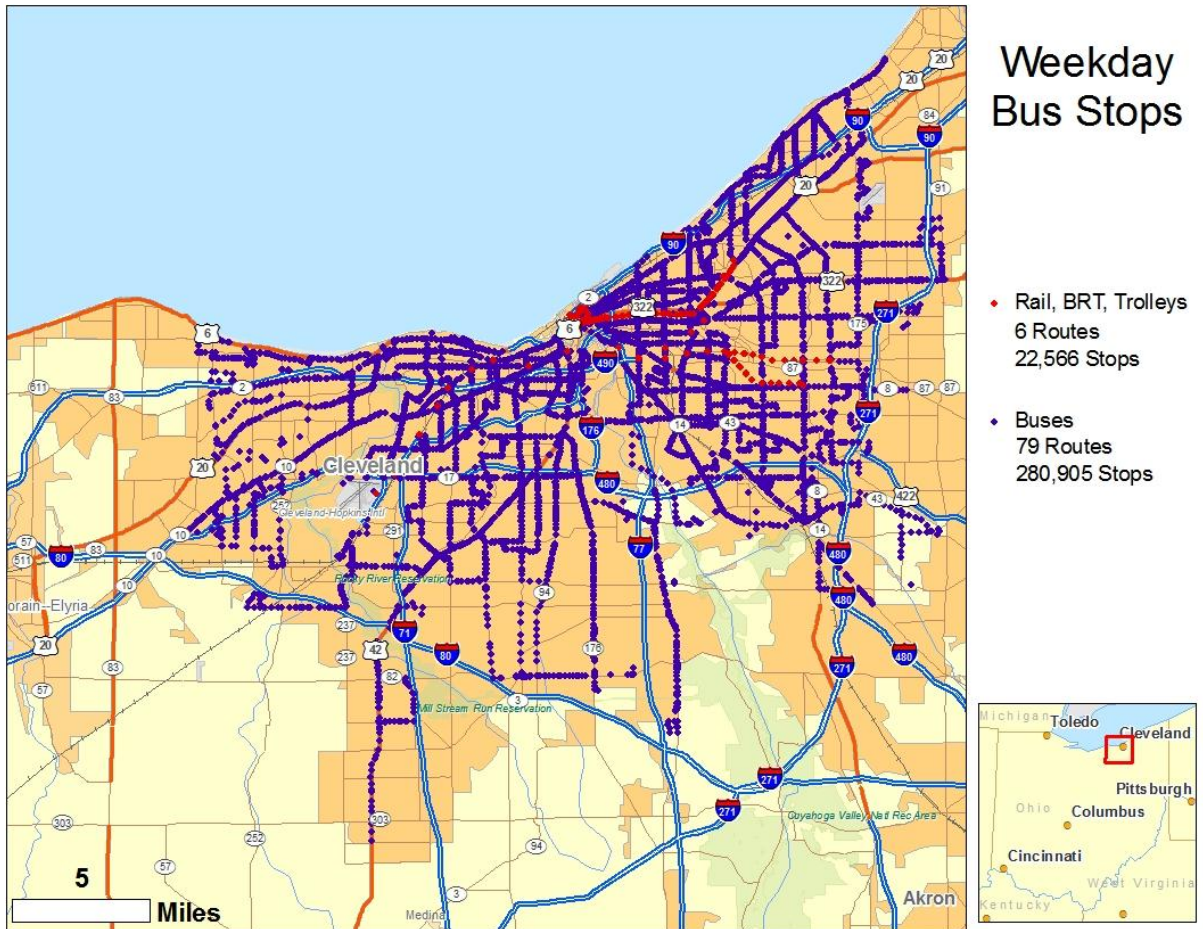
Figure 3 – Spatial distribution of bus service in Cleveland, OH. Data as of January, 2010.

The Columbus data is made up of 170,677 stop points from 64 routes (Figure 2), while the Cleveland data has 280,905 points from 79 routes (Figure 3). Columbus has only bus service for public transit, and all non-bus service was excluded from the geovisualization of the Cleveland data. In figures 4 and 5, we can see the spatio-temporal distribution of transportation in both cities.

In comparing figure 4 and 5, it is apparent that Columbus has more service in its periphery than Cleveland, with more arm-like features extending outward from the core of the city. In both figures, the 5 p.m. commuting time has been selected, but the service earlier in the day is also visible underneath the darkened points.
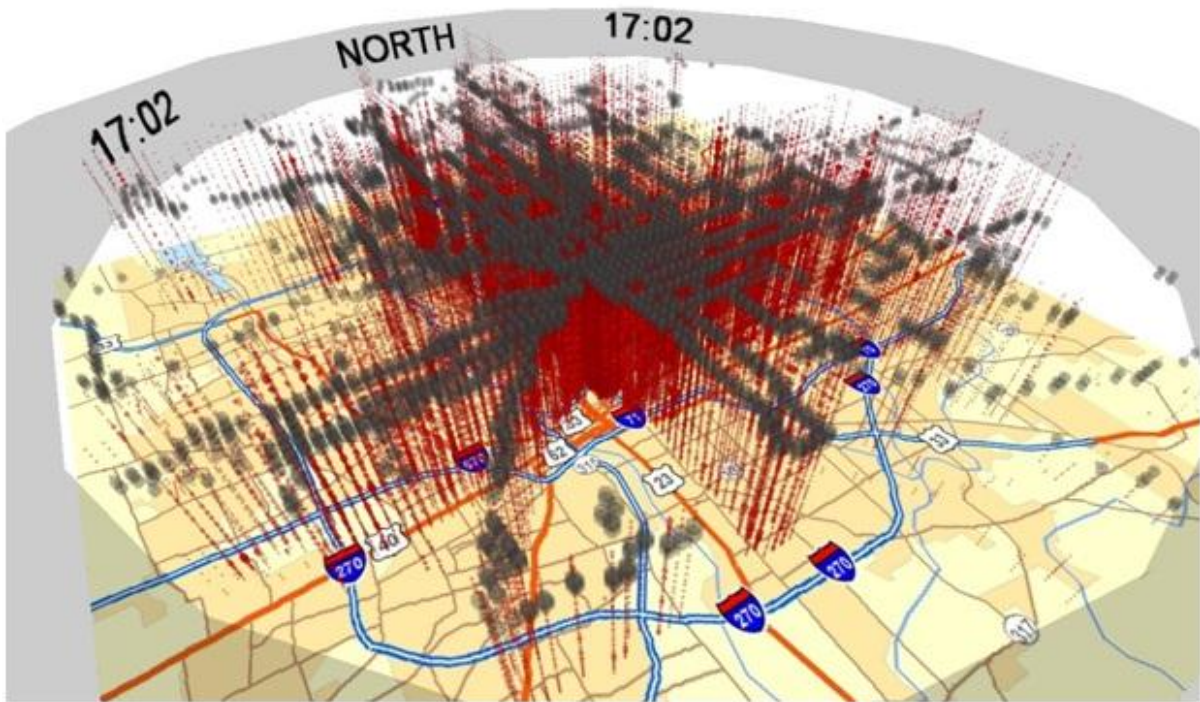
Figure 4 – Discretized space-time cloud of bus service in Columbus. Service around 5 p.m. is selected. This figure shows the entire dataset.
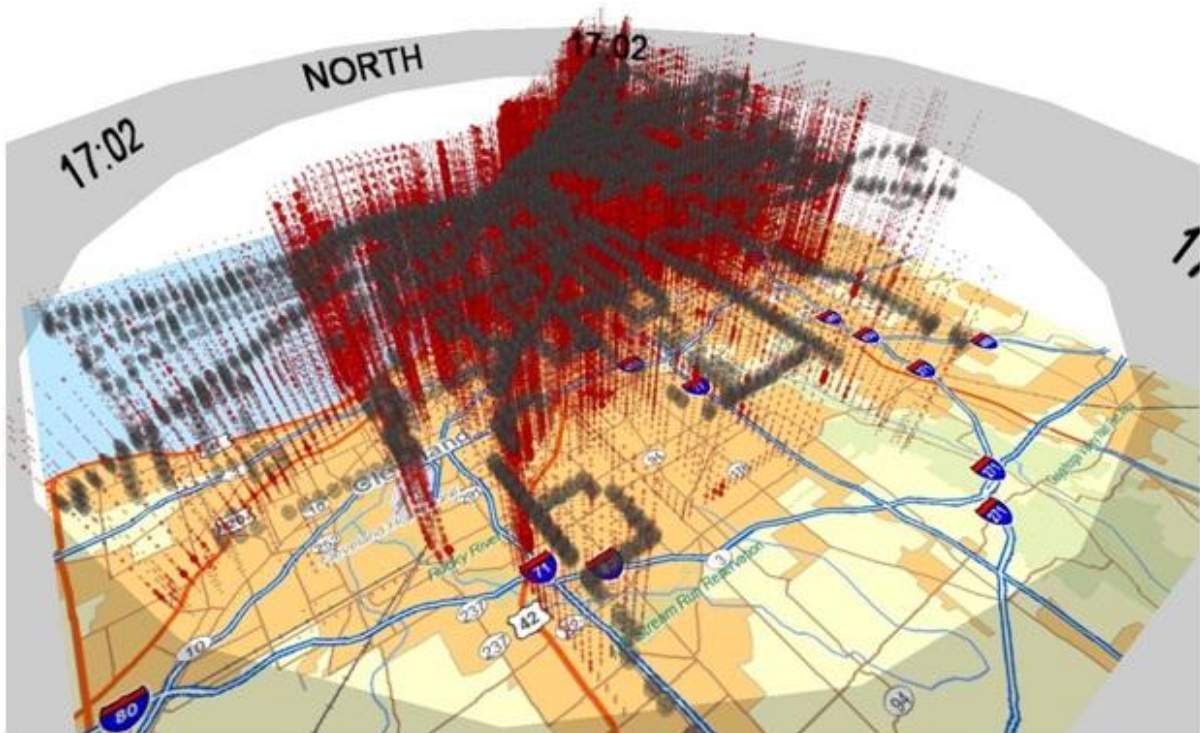

Figure 5 – Discretized space-time cloud of bus service in Cleveland. Service around 5 p.m. is selected. This figure shows the entire dataset.

Selection rings are employed in these figures to better locate the time of service. Layers of low-frequency service can be continually peeled away by increasing the threshold, which, as expected, shows a dense urban core of high-frequency service throughout the day in both cities (figures 6 and 7). After thresholding to a value of 7, the low-frequency service in Columbus and Cleveland has mostly been peeled away, showing a concentration of frequent service in the downtown cores. In Columbus, there is also a high-frequency north-south connection that local residents will recognize as the number 2 bus along High Street and an additional east-west link along Broad and Main Streets. In Cleveland, more locations are being

served in more directions after the low intensity service is thresholded away, but it appears that the arm-like features do not extend as far out into the periphery as the most frequent routes in Columbus. In both cities, service can be seen to begin increasing in frequency around 5 a.m. (that time interval is highlighted in figure 7 for Cleveland).
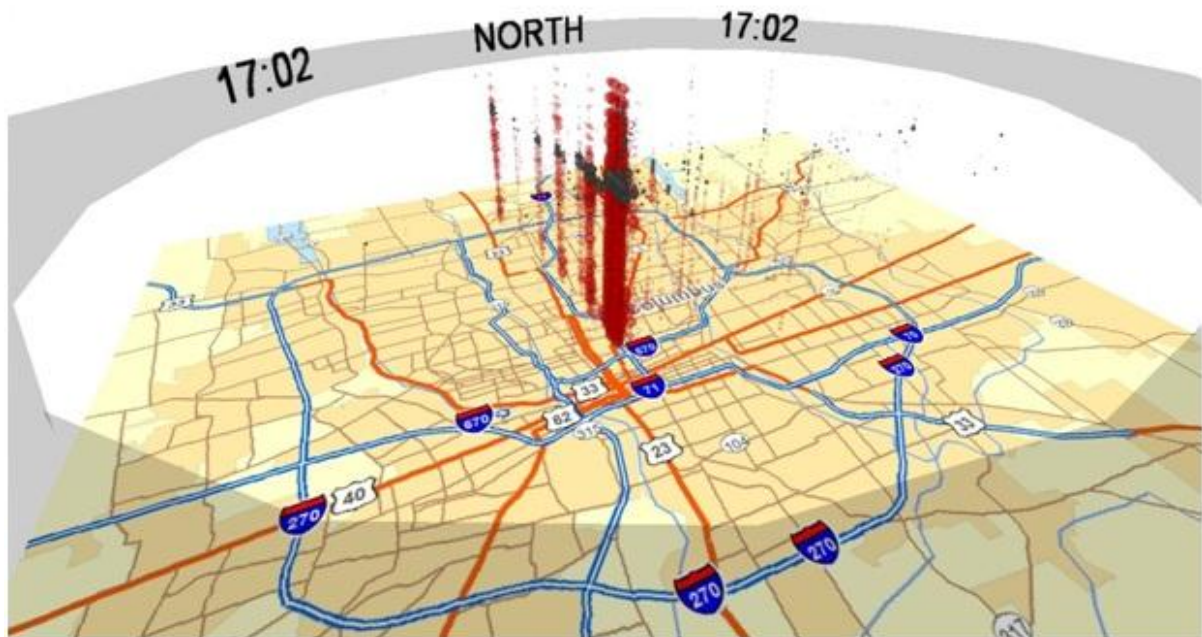

Figure 6 – Discretized space-time cloud in Columbus, with minimum values thresholded to seven.
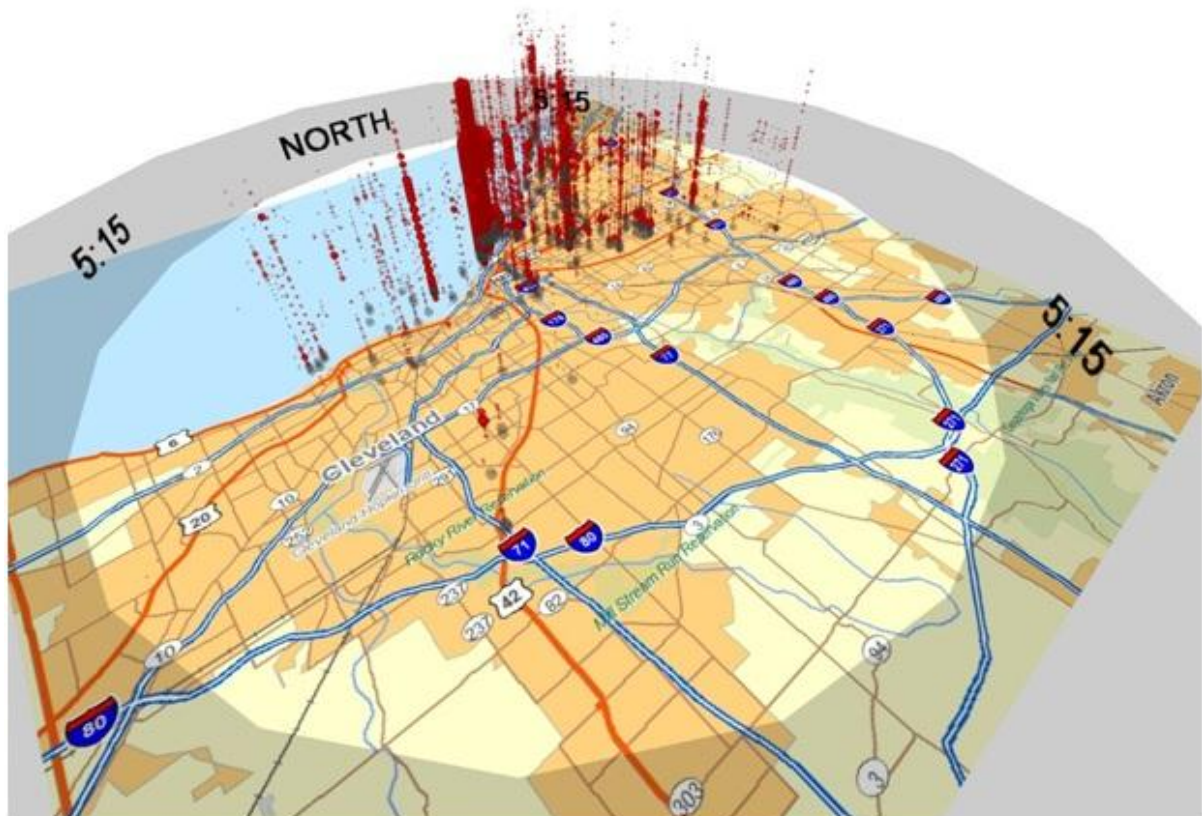

Figure 7 – Discretized space-time cloud in Cleveland, with minimum values thresholded to seven.

Finally, an example of filtered data is also provided for Columbus in figure 8. The selection ring is once again placed at 5 p.m., showing the effect of the filter on increasing the size of visible features. Looking beneath the selection, the unselected points have had their transparency increased from the default amount, showing the densest parts of the cloud to be the most opaque, and peripheral routes appearing more pale.

This effect is more pronounced in figure 9, where the viewpoint is positioned to look southeast on Cleveland from Lake Erie, and there are no selected points to distract from the cloud itself.
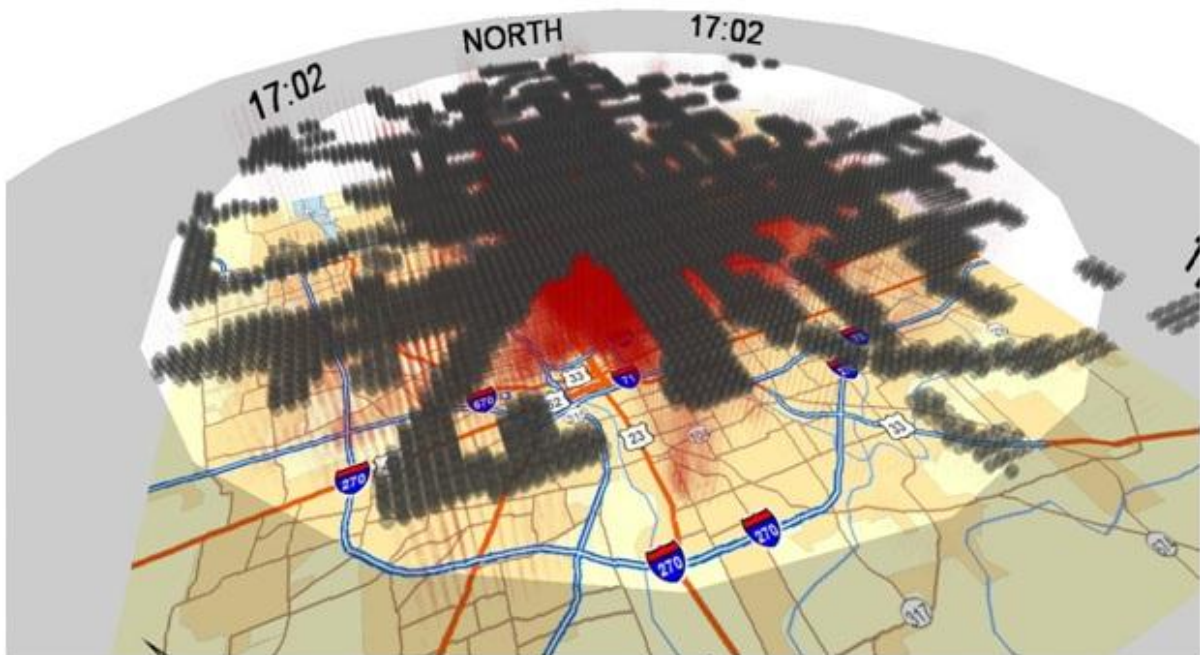

Figure 8 – Discretized space-time cloud in Columbus. Filtering has been enabled, and transparency of non-selected points has been increased.
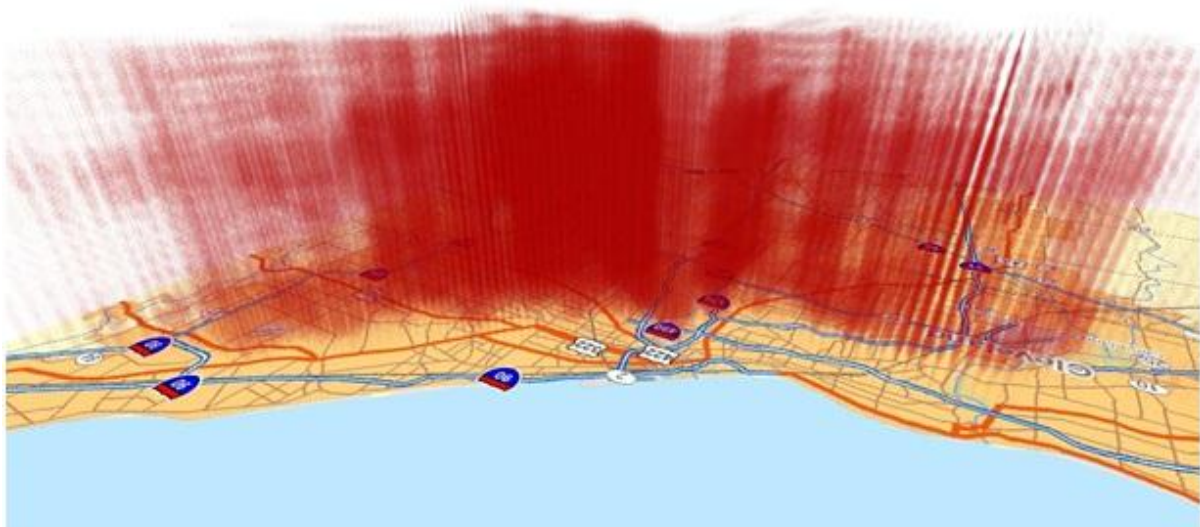

Figure 9 – Discretized space-time cloud in Cleveland, looking southeast from Lake Erie. Filtering has been enabled, and transparency of points has been increased.

This sort of visual analysis is helpful for becoming familiar with the transport data before doing more involved analysis. Of course, the next logical steps in a full transportation study would be to focus on areas that would have a particular need for public transport service, and see if they are being underserved. Quantitative measurements like accessibility calculations could help determine the extent to which population segments or particular places are underserved. A geovisualization method such as the one demonstrated here is helpful for understanding general patterns in the data, and putting more specific abstractions like statistical calculations into context. This paper suggests there is potential in this sort of geovisualization technique even outside human space-time activity data. In the future, the authors plan to increase the analytical capability of the software implementation, particularly with more bi-directional linking between the visualization program and the GIS, the creation of a normalized data option to better

support the comparison of multiple datasets, and the implementation of other spatio-temporal filters to investigate their effects on visual interpretation.

WORKS CITED

American Public Transportation Association. (2007). A Profile of Public Transportation Passenger Demographics and Travel Characteristics Reported in On-Board Survey.

Andrienko, N., Andrienko, G. & Gatalsky, P. (2005). Impact of Data and Task Characteristics on Design of Spatio-Temporal Data Visualization Tools. In Dykes, J., MacEachren, A., & Kraak, M.-J. (Eds.), Exploring Geovisualization (pp. 201–222). Elsevier.

Andrienko, N. & Andrienko, G. (2006). Exploratory Analysis of Spatial and Temporal Data. Springer.

Anselin, L. (1993). Exploratory Spatial Data Analysis and Geographic Information Systems. DOSES/EUROSTAT Workshop on New Tools for Spatial Analysis, ISEGI.

Forer, P. (1998). Geometric Approaches to the Nexus of Time, Space, and Microprocess: Implementing a Practical Model for Mundane Socio-Spatial Systems. In Egenhofer, M. & Golledge, R. (Eds.), Spatial and Temporal Reasoning in Geographic Information Systems (pp. 171–190). Oxford University Press.

Forer, P. & Huisman, O. (2000). Space, Time and Sequencing: Substitution at the Physical / Virtual Interface. In Janelle, D. & Hodge, D. (Eds.), Information, Place, and Cyberspace (pp. 73–90). Springer.

Fotheringham, A. S. (1989). Scale-independent spatial analysis. In Goodchild, M. and Gopal, S. (Eds.), Accuracy of Spatial Databases (pp. 221–228). Taylor & Francis.

Google. (2010). General Transit Feed Specification. Retrieved from http://code.google.com/transit/spec/transit_feed_specification.html

Hägerstrand, T. (1970). What about people in regional science?. Papers in Regional Science, 24(1), 7–24. Originally given as a talk at the Ninth European Congress of the Regional Science Association, 1969.

Kain, J. (1968). Housing segregation, Negro employment and metropolitan decentralization. Quarterly Journal of Economics, 82, 175-197.

Kraak, M. (2003). The space-time cube revisited from a geovisualization perspective. In Proceedings of the 21st International Cartographic Conference: Cartographic Renaissance, Durban, South Africa (pp. 1988–1996). International Cartographic Association.

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. Transportation Research C, 8, 185–203.

Kwan, M.-P. & Lee, J. (2004). Geovisualization of Human Activity Patterns Using 3D GIS: A Time-Geographic Approach. In Goodchild, M. & Janelle, D. (Eds.), Spatially Integrated Social Science (pp. 48-66). Oxford University Press.

McCarthy, C. & Callele, D. (2006). Virtools User Guide. Dassault Systèmes.

Rush, J. (2009). A Visualization Strategy for Analyzing High Volumes of Space-time Activity Data. (Master's Thesis). The Ohio State University, Columbus, Ohio, USA.

Theus, M. (2005). Statistical Data Exploration and Geographical Information Visualization. In Dykes, J., MacEachren, A., & Kraak, M.-J. (Eds.), Exploring Geovisualization (pp. 127–142). Elsevier.

US General Accounting Office. (1998). Welfare Reform: Transportation's Role in Moving From Welfare to Work (GAO/RCED-98-161), Washington,DC

Ware, C. (2008). Why Do We Keep Turning Time Into Space? In Hornsby, K. and Yuan, M. (Eds.), Understanding Dynamics of Geographic Domains. CRC Press.