

# Development of a Custom GIS Application for Viewing Ethnographic Data

Jim Detwiler

## Abstract

This paper discusses the use of GIS in an ethnographic study being conducted at Penn State University. Its focus is on the development of an ArcMap extension using ArcObjects and Visual Basic that provides the social scientists in the study (generally not GIS users) easy access to various forms of geographic information through custom tools and forms. The extension provides access to secondary data such as census variables, transportation networks, crime data and tax parcel data. It also enables the researchers to view data collected by the ethnographers in a geographic context through neighborhood walkthrough layers that are hyper-linked to the original field notes and layers that map the neighborhood resources used by families in the study. This extension enables the researchers to understand the neighborhoods in which their families live in a way that would not be possible without GIS.

## Background

*Welfare, Children, and Families: A Three-City Study* is a longitudinal study of children and their caregivers in low-income families that were living in low-income neighborhoods in three cities in 1999. The purpose of the study is to investigate the consequences of policy changes resulting from the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA), a piece of legislation that represents a significant change in U.S. welfare policy. The survey was designed to provide information on the health, cognitive, behavioral, and emotional development of children and on their primary caregivers' labor force behavior, welfare experiences, family lives, use of social service, health, and well-being.<sup>1</sup>

The three cities under study are Boston, Chicago, and San Antonio. Over 175 individuals from 6 research institutions have been involved. The study has three components:

- a longitudinal survey of about 2400 low-income families,
- a focused look at the development of children in a sub-sample of approximately 700 of the survey families,
- an ethnographic study of 215 low-income families and the neighborhoods in which they reside.

All of the information collected by the ethnographic teams at the three cities is sent to Penn State. The information is most commonly in the form of field notes that summarize an ethnographer's observations. Several thousand of these field notes have been transferred electronically to Penn State over the course of the study. The majority of the

<sup>1</sup> – from *Welfare, Children, & Families: A Three-City Study* Public Release, <http://www.jhu.edu/~welfare/release.html>, accessed May 25, 2004.

notes are comprised of interviews with family members, while a smaller proportion cover observations of the neighborhoods in which the families live.

The role of the Penn State team has been to prepare the field notes for analysis by coding them and importing them into a software package called NUD\*IST. NUD\*IST facilitates the analysis of qualitative data like textual field notes by allowing a researcher to query all paragraphs from all coded files having to do with a particular topic of interest, such as child nutrition or crime in the neighborhood.

### **The Role of GIS in the Study**

The researchers working on the Three City Study are faculty, research associates, and graduate students in various social science disciplines such as sociology, anthropology, and family studies. As such, very few of them have experience with GIS. However, the principal investigator at the Penn State site realized the potential value in adding GIS to the study proposal and contracted the services of the Geographic Information Analysis Core (GIA Core), a unit within the university's Population Research Institute (PRI).

The GIA Core was written into the Three City Study proposal for two primary reasons:

1. Maps are able to convey information about families and their neighborhoods much more effectively than a simple narrative;
2. GIS provides the best tools for exploring the neighborhood context in which the study families live.

Using information collected by the family ethnographers (locations of homes and of resources used by the families in their day-to-day lives), it is possible to construct "typical day" maps for each family in the study. Figure 1 shows one such map. These maps have been requested often by the members of the ethnographic team because of their effectiveness in illustrating the adversity faced by many low-income mothers.

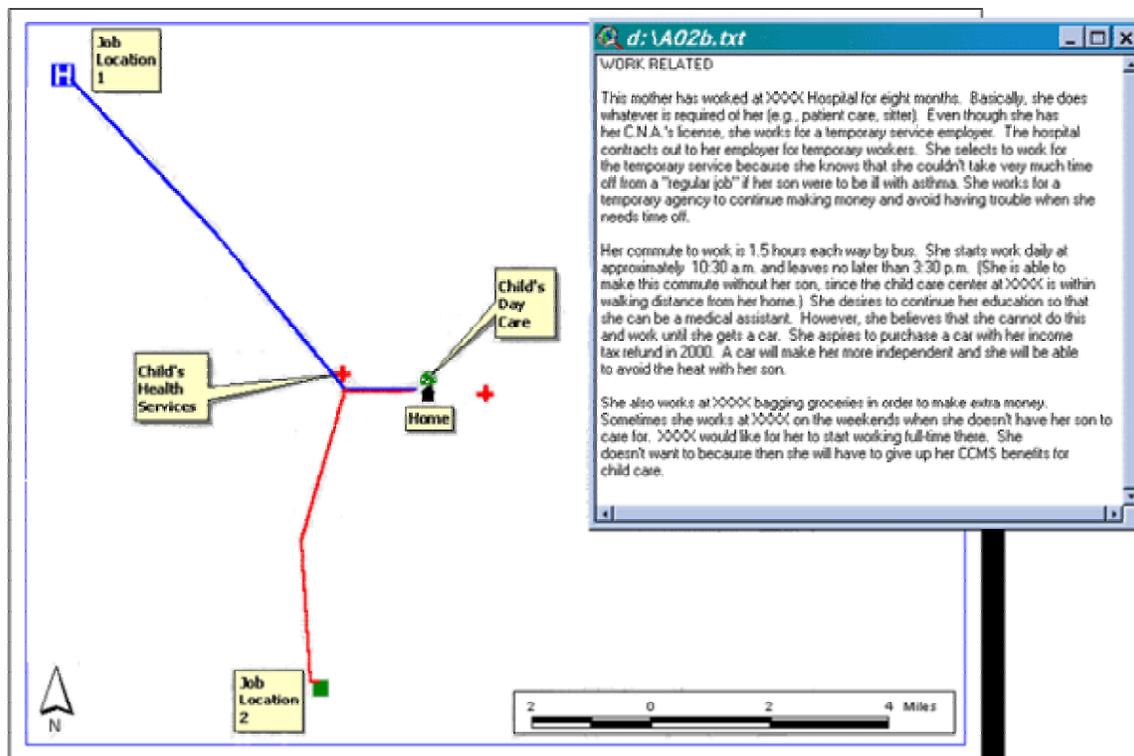


Figure 1. Map of a typical day for one of the mothers in the study. All locational identifiers (such as the road network) have been removed to ensure the confidentiality of the participant.

Maps like Figure 1 do an excellent job of serving the first function described above; that is, providing powerful illustrations. However, meeting the second goal of exploring neighborhood context called for a more proactive, creative approach.

### The Welfare Project Extension for ArcMap

To construct an accurate picture of the context in which the study families lived, the GIA Core collected a number of different types of data to augment the information collected by the family ethnographers. These data include:

- transportation networks,
- census data,
- land parcel information,
- school locations and enrollment information,
- locations of criminal activity.

Another important data source was the hundreds of files containing field notes from the neighborhood ethnography. These notes typically cover an ethnographer's walk or drive through a neighborhood or an interview with an important member of the community such as a social worker, property manager, etc.

One of the challenges of this project has been that the social scientists carrying out the study have expertise in welfare policy and its impact on low-income families, but are

largely unfamiliar with GIS. Likewise, the GIS experts working on the project have little experience working in the subject area and may not recognize how GIS can best be used to contribute to the project. To address this difficulty, the GIA Core developed a custom application that allows the project's non-GIS users to explore the collected neighborhood data without having to spend an inordinate amount of time learning how to use GIS software. The rest of this paper will demonstrate how this application works.

Originally developed in Avenue for ArcView 3.x, the application now exists as an extension for ArcMap. All tools are provided through a toolbar containing five buttons and tools (see Figure 2).

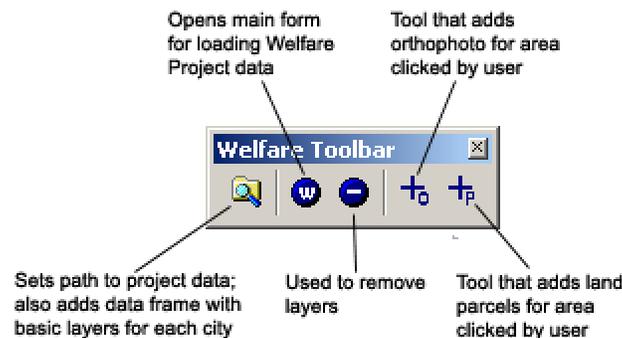


Figure 2. The Welfare Project Toolbar

Because of confidentiality concerns, neither the locations of the families nor the neighborhoods in which they live may be displayed in any publications. Therefore, in order to illustrate the workings of the extension, data for three other cities (Pittsburgh, Lancaster, and Philadelphia in Pennsylvania) have been extracted and a hypothetical set of family locations, neighborhoods, and neighborhood field notes have been created.

#### *Set Path button*

Upon clicking the Set Path button and specifying the path to the project data, a data frame with base layers is added for each of the three cities (see Figure 3). These base layers include the following:

- city limits;
- neighborhoods in which the study participants live;
- major roads - extracted from the city's overall road dataset according to the U.S. Census Bureau's feature class code (CFCC) which differentiates highways from arterial streets and local roads, etc.;
- an ortho-photo index layer – each rectangular polygon in this layer depicts the extent of an ortho-photo image covering part of the city while a field in its attribute table stores the file names of the corresponding images;
- a land parcel index layer – like the ortho-photo index layer, it depicts the extent of various tiles in an overall city parcel database.

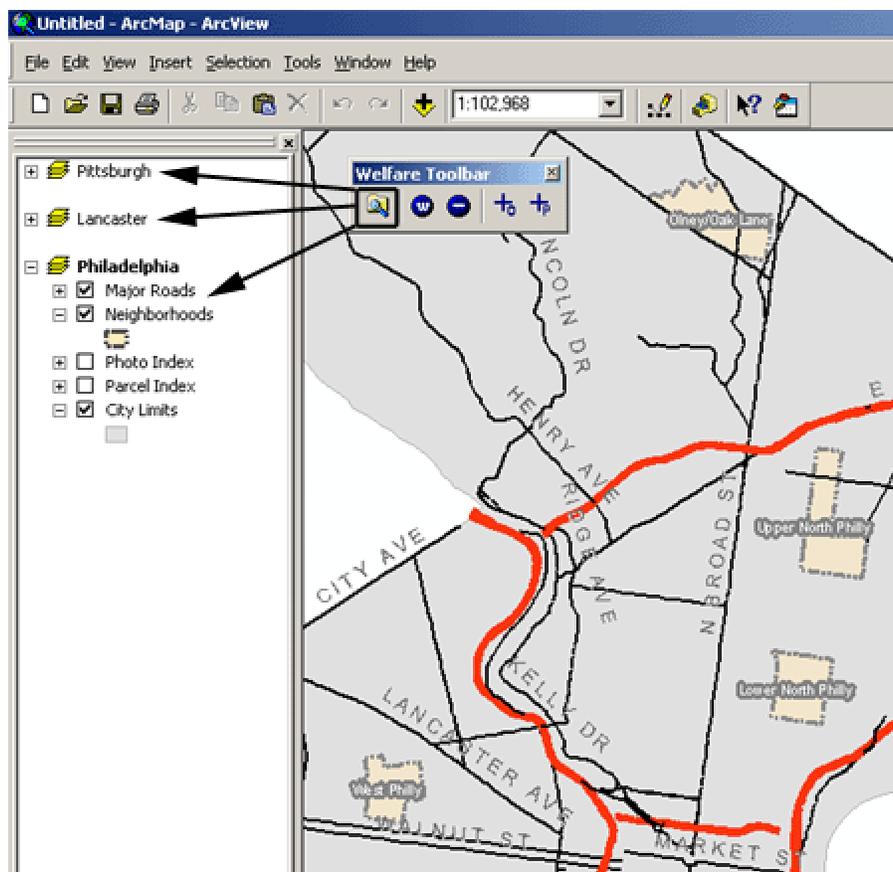


Figure 3. Data frames and base layers are automatically added and symbolized after the user specifies the path to project data.

All layers are drawn in a custom symbology and the neighborhoods and major roads are labeled with their names. In addition, unimportant legend entries and inactive data frames are collapsed in the Table of Contents to reduce clutter.

#### *Ortho-photo tool*

The Ortho-photo tool is enabled only if the Photo Index layer is present in the active data frame. The key element in the tool's design is a field in the index layer's attribute table that stores the name of the file associated with that index layer feature, or tile. When the user clicks on the map display using this tool, the code triggered by the tool's MouseUp event first determines which of the index layer features is intersected by the clicked point location. The code then checks the value in the name field to determine which image should be added to the data frame (see Figure 4). The image layer is moved to the bottom of the data frame's layer collection so that it does not obscure any of the vector layers. Its legend entry is unimportant and is subsequently hidden.

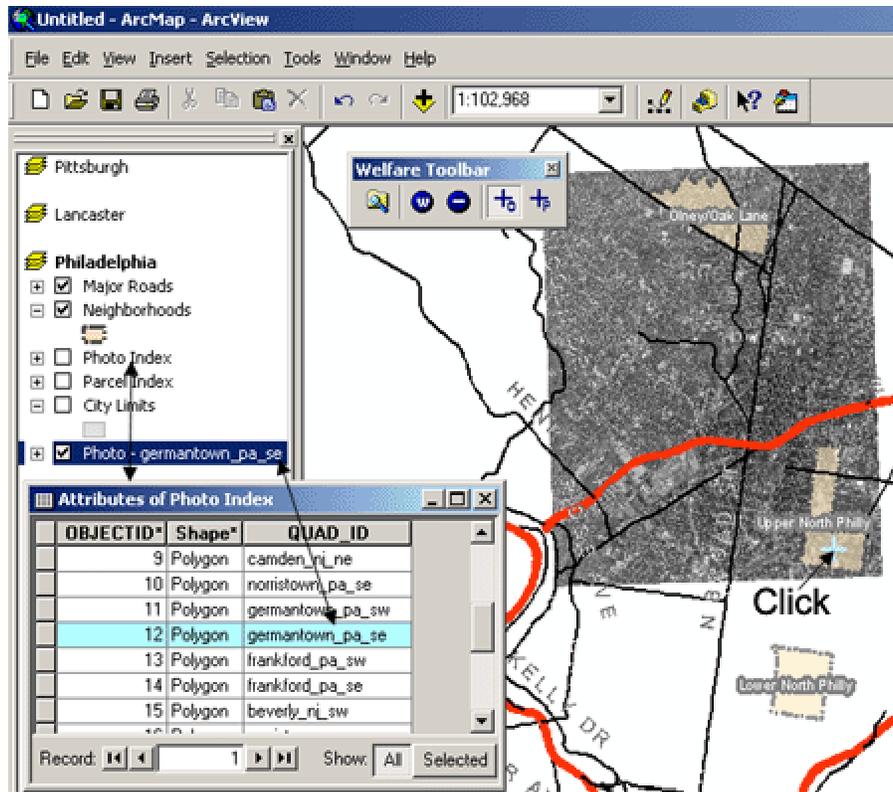


Figure 4. Using the ortho-photo tool to add a photo for an area of interest. The tool works by identifying which Photo Index tile is clicked and adding the image associated with that tile.

The ortho-photos were all downloaded freely from state data clearinghouses. The index layers were created manually.

#### Parcel tool

The Parcel tool operates on much the same design as the Ortho-photo tool. Each tile in the Parcel Index layer is associated with a feature class in a personal geodatabase. The value from the name field determines which feature class should be added to the data frame. The same basic symbol is applied to all features in layers added using the Parcel tool. As with the Ortho-photo tool, this tool is meant to add a layer that will serve as a backdrop for other layers.

For the cities of Boston and San Antonio, the parcel datasets were obtained for a nominal fee through the local jurisdictions responsible for maintaining land records: the City of Boston and Bexar County (San Antonio). Cook County, IL was in the middle of a parcel conversion project when the data for this project were collected and was not distributing parcel data for Chicago. However, we managed to obtain a land use dataset (in which the unit of analysis was closer to the scale of a city block than an individual parcel), from the Northeastern Illinois Planning Commission.

*Remove Layers button* 

This button opens a form that lists all of the non-base layers (those layers not added using the Set Path button) in the active data frame. The user then selects which layers to remove from the data frame from this list. This method of removing layers is a bit more user-friendly than the default ArcMap functionality.

*Welfare Data button* 

The most useful control on the project toolbar is the Welfare Data button. Clicking this button opens a form that provides access to the various datasets listed above.

Along the top of the form are three radio buttons, one for each of the study cities. Clicking on a city's radio button activates the data frame for that city, expands the layer listing in the Table of Contents for that data frame, and collapses the listings for the other two data frames.

Beneath the city names is a list box that displays a list of the study neighborhoods in the selected city. To the right and below the neighborhood list box are controls for adding data. All of these controls are disabled until the user selects one or more neighborhoods. The steps involved in adding a dataset are summarized in Figure 5. After selecting a neighborhood, the user selects a data type from the set of radio buttons, and finally, a layer from the list in the adjacent list box. The list of layers is dependent on which of the data types is selected.

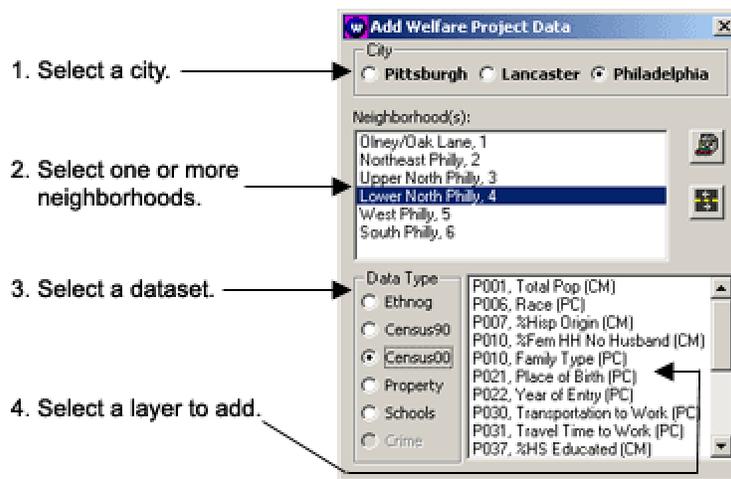


Figure 5. Using the Welfare Data form to add a project dataset.

*Ethnog data option*

The “Ethnog” option is used to view information from the neighborhood ethnography component of the study. As mentioned above, the notes written by the neighborhood ethnographers typically describe what they observed during a walk or drive through a neighborhood. Upon reading a few of these notes, it became clear that the walk-throughs

could be parsed into smaller pieces, with each piece being associated with a segment of the overall route. These segments could be mapped as features in a feature class and the name of the text file associated with each feature could be stored in the feature class's attribute table to enable the use of hyper-linking to view the associated notes.

Figure 6 shows an example in which one of these ethnographic data layers has been added to the map. When the user clicks on the "Ethnog" radio button, a search is performed for all observations that were made in the selected neighborhood and a list of these observations is added to the adjacent list box. Each observation in the list is named using an abbreviation of the city name, the ID number of the neighborhood, and the date the observation was made. This corresponds to the name of the folder that stores all of the individual text files that together make up the complete text of the observation. Selecting an item from the list adds a layer that shows all the features having associated text on the selected observation date.

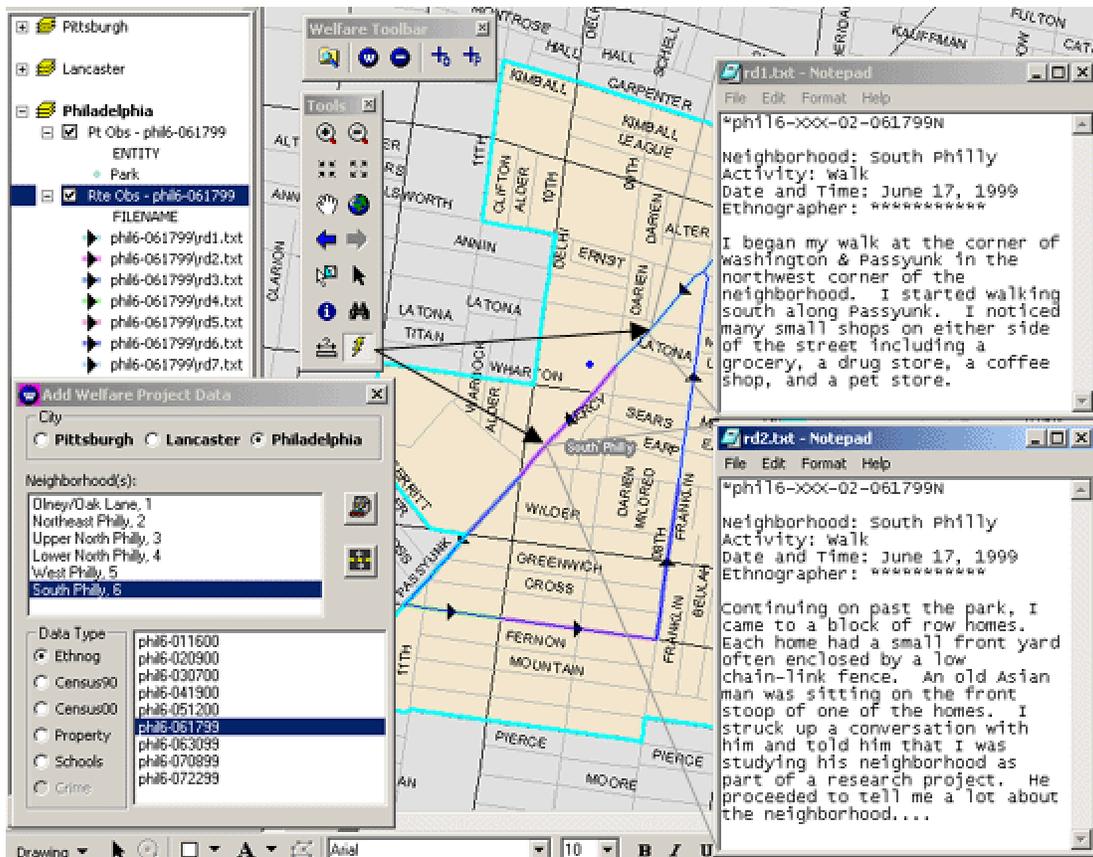


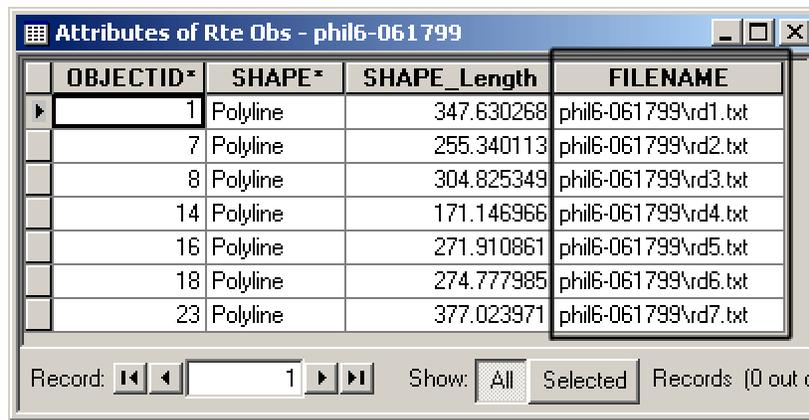
Figure 6. Viewing ethnographic field notes that have been split into smaller pieces and associated with segments of an overall walking route. Here observations made in the South Philly neighborhood on 6/17/99 are viewed by clicking on route segments with the Hyperlink tool.

For all of the field notes that have been geocoded, three geodatabase feature classes store the features that have associated text; one for point observations, one for line observations, and one for polygon observations. The most common feature type is a line, since most of the text refers to what the ethnographer observed as he/she was moving

along a street. Sometimes the ethnographer will stop and write a good deal of text about a particular location or the field note will actually be an interview with an important member of the community. In these cases, the text will be associated with a point or a polygon, depending on how well the location can be delineated.

Within the attribute table of each of the three feature classes is a field that stores the name and relative path to the text file associated with each feature (see Figure 7). The application code specifies this as the hyperlink field and sets the map document's hyperlink path based on the user's response to the Set Path button prompt.

The application adds only those features associated with the selected observation date (filtering out features from other dates) by setting up a definition query that identifies features for which the beginning of the filename value matches the item selected in the list box on the form.



OBJECTID*	SHAPE*	SHAPE_Length	FILENAME
1	Polyline	347.630268	phil6-061799\rd1.txt
7	Polyline	255.340113	phil6-061799\rd2.txt
8	Polyline	304.825349	phil6-061799\rd3.txt
14	Polyline	171.146966	phil6-061799\rd4.txt
16	Polyline	271.910861	phil6-061799\rd5.txt
18	Polyline	274.777985	phil6-061799\rd6.txt
23	Polyline	377.023971	phil6-061799\rd7.txt

Figure 7. The relative path to the text file associated with each segment of an observation route is stored in the feature class attribute table.

### *Census data options*

The next two radio buttons (Census90 & Census00) allow the user to add preset layers showing block group information from the 1990 and 2000 Censuses. These data were extracted from Geolytics' Census CD products. The layers include basic data like total population and population counts by race, but also poverty indicators like percentage high school educated, percentage of children on public assistance, per capita income, etc. Each item in the list of available layers includes a two-letter abbreviation indicating the type of symbology used; CM for a basic choropleth map, PC for a pie chart map, and BC for a bar chart map.

For all pie chart and bar chart maps, the charts are sized according to an appropriate total population field to help the user determine if a pattern in the charts is truly significant. Choropleth maps that depict percentages are constructed by normalizing the count field by the total population field (e.g., normalizing the count of female-headed households in which no husband is present by the total number of households), eliminating the need to

calculate and store these values in a separate field. Figure 8 shows an example of using the Welfare Data form to add two layers from the 2000 Census layers; one a pie chart layer depicting race, the other a choropleth layer depicting total population.

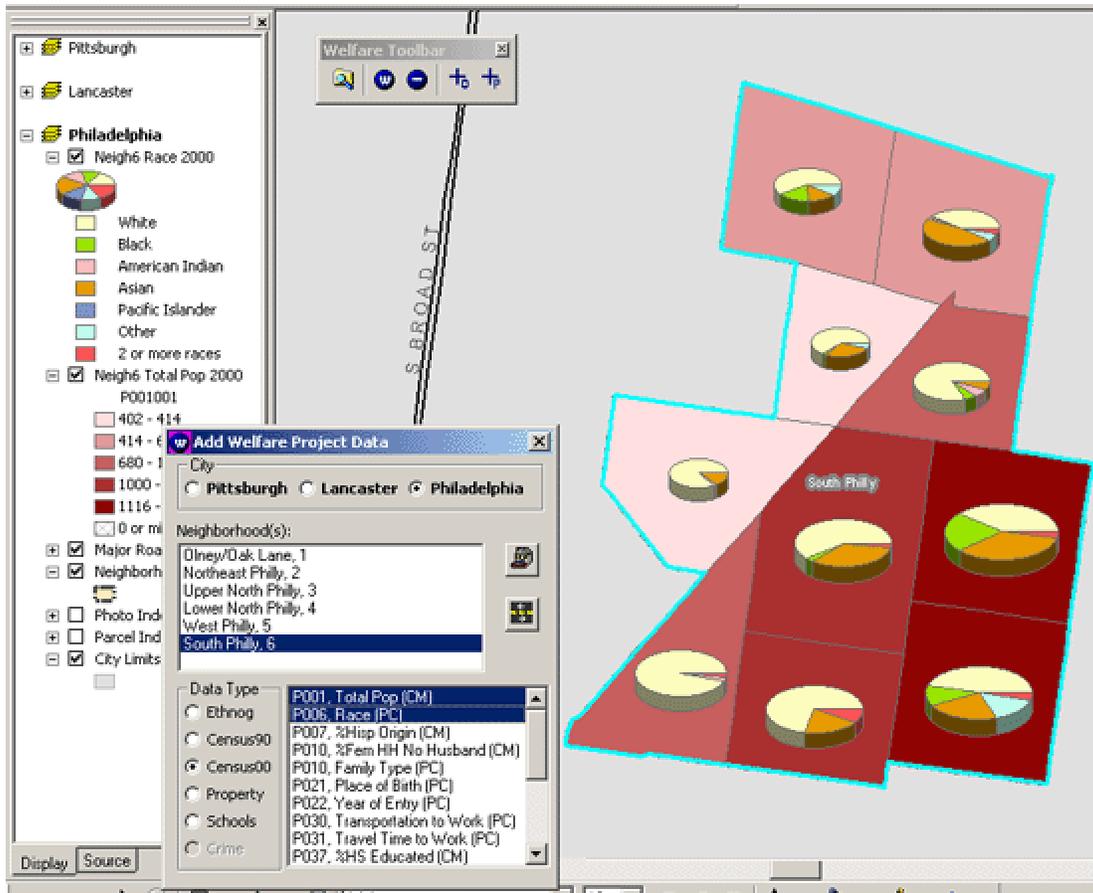


Figure 8. Using the Welfare Data form to add layers from the 2000 Census.

### *Property data option*

The next category of data available to users is Property. Clicking this radio button produces a list of layers that are built upon the same land parcel dataset discussed above in the section on the Parcel tool. While the Parcel tool simply draws all parcels using the same symbol, the preset layers in the Property category provide the user with a number of thematic views of the data.

For all three cities, the user is able to add a layer depicting land use. The land use classification system varied between cities, so all parcels were re-classified into a standard system using basic categories that were easy for users to understand. These categories included: high-density residential, low-density residential, commercial, industrial, tax-exempt, etc.

As mentioned above, the city block-scale land use dataset was the only property information we were able to acquire for Chicago. This means land use is the only layer available for Chicago under the Property category in the Welfare Project form. However, parcel data were available for Boston and San Antonio. The set of available layers is different for each city, since each city's dataset had its own set of parcel attributes.

Figure 9 shows partial listings of the available property layers for each of the three cities. Under the heading "All Properties" are thematic maps that show all properties (parcels) in the selected neighborhood. In Boston, the only such map available is "Value per Square Foot." However, the San Antonio data provides a number of attributes that may be of interest to the study including "Heat Type," "Air Conditioning," etc.



Figure 9. Partial listings of Property layers available for each of the three cities.

Both the Boston and San Antonio datasets also include a field that identifies the type of property using a system that's a bit more specific than the land use classification. In Boston, these property types include schools, places of worship, benevolent/social foundations, museums/zoos, etc. San Antonio has a similar property type code, though its codes are even more extensive than Boston's. Because these codes help to identify neighborhood resources that are available to low-income families, the various property types were manually extracted from the overall dataset by neighborhood, stored in their own feature classes, and made available through the project extension. This makes it possible for a user to quickly view all places of worship in Lancaster's neighborhood 1, for example, as shown in Figure 10. (As noted above, maps of neighborhoods in the actual study neighborhoods of Boston, Chicago, and San Antonio may not be shown because of confidentiality concerns.)

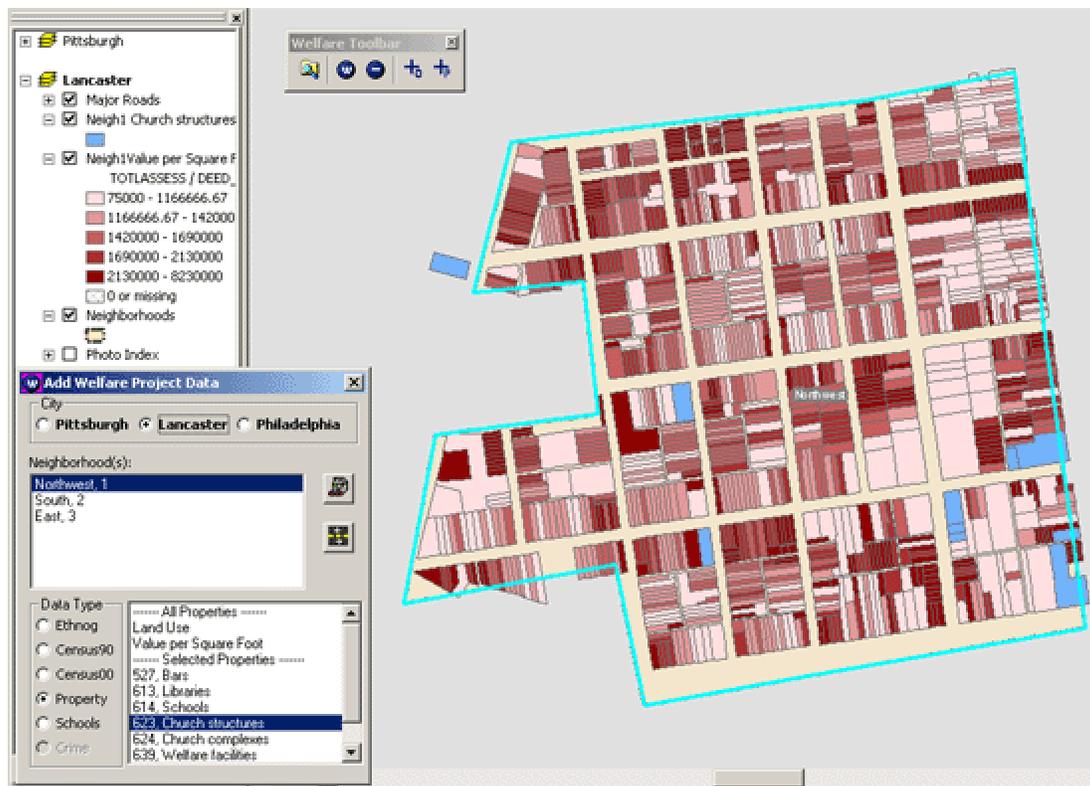


Figure 10. Viewing places of worship above a map of parcel value per square foot.

Looking at Figure 10, you may notice that one of the places of worship is actually outside the neighborhood. The reason for this is that it would be unrealistic to expect families to use only the resources within our rather arbitrary neighborhood boundaries. When extracting these special property types, parcels were selected both within the neighborhood and up to a mile beyond the neighborhood boundary.

### *Schools data option*

Data on schools was downloaded from the National Center for Education Statistics (NCES) website. Their Common Core of Data provides information on school enrollment, student/teacher ratio, racial composition, state expenditures per student, etc. for all public elementary and secondary schools in the U.S. Conveniently, it also includes the latitude/longitude position of each school, making it easy to geocode this information.

For each neighborhood in the study, schools inside and within a mile of the neighborhood were extracted and stored in a geodatabase feature class, much like the process used for the property types. Figure 11 shows the layers that are available from the NCES dataset along with an example of adding a “School Type” layer which breaks the schools up into elementary, middle, and high, using different colors for each category. The “Num Students” layer displays each school in the vicinity of the neighborhood using a point

whose size is proportional to the school's enrollment. The "Racial Makeup" layer displays a pie chart symbol for each school that reflects the school's racial composition.

### *Crime data option*

The last dataset made available through the set of radio buttons is a crime dataset downloaded from the San Antonio Police Department. Similar datasets for Boston and Chicago were not found. The SAPD data includes the type of crime and State Plane coordinates of all crimes committed in the city over a month in 2000. For burglaries, the date and time of day are also included. Viewing the spatial pattern of crimes within study neighborhoods can provide yet another way for researchers to understand the context in which the families are living. Figure 12 shows a map of violent crimes that occurred in one of the San Antonio neighborhoods during the downloaded time period. (All elements that would give away the location of the neighborhood have been removed.) Intuitive symbols from ESRI's Crime Analysis symbol set are used to depict the crimes.

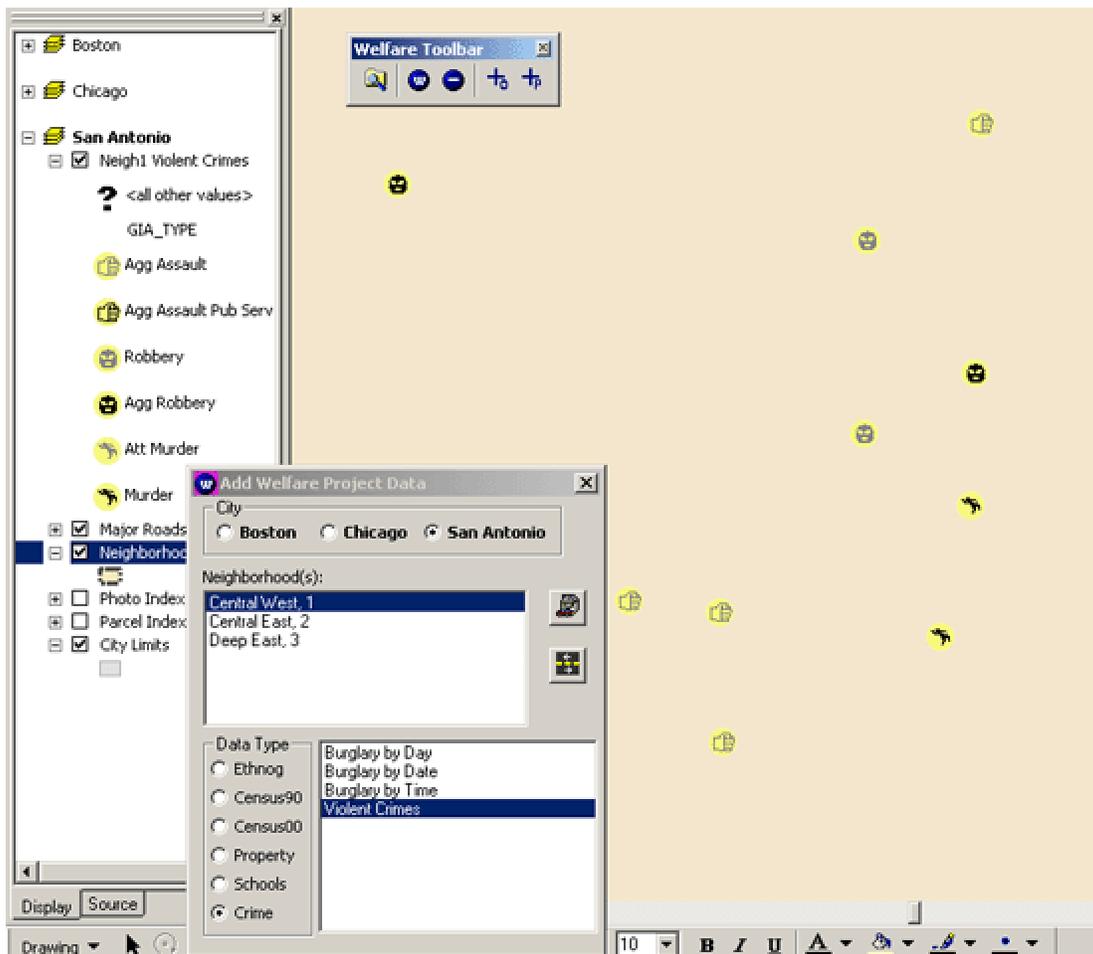


Figure 12. Viewing violent crime information for a neighborhood in San Antonio.

The other three available crime layers provide a way to explore the temporal dimension of crime within the neighborhoods, at least for burglaries. Clicking on the "Burglary by

Day” choice prompts the user to choose the day(s) of the week that are of interest. Clicking on “Burglary by Date” allows the user to specify a time period of interest by entering a “from” and “to” date. Similarly, the “Burglary by Time” choice allows the user to specify a “from” and “to” time of day.

### *Homes & Roads buttons*

Two more datasets are provided through the two command buttons to the right of the neighborhood list box. The first of these buttons adds a layer showing the locations of the study families in the selected neighborhood(s). The second button adds a layer showing all of the roads (not just major roads) in the selected neighborhood(s) with the roads being labeled with their name. Figure 13 shows the locations of imaginary study families in one of the Philadelphia neighborhoods drawn on top of the neighborhood’s roads layer.

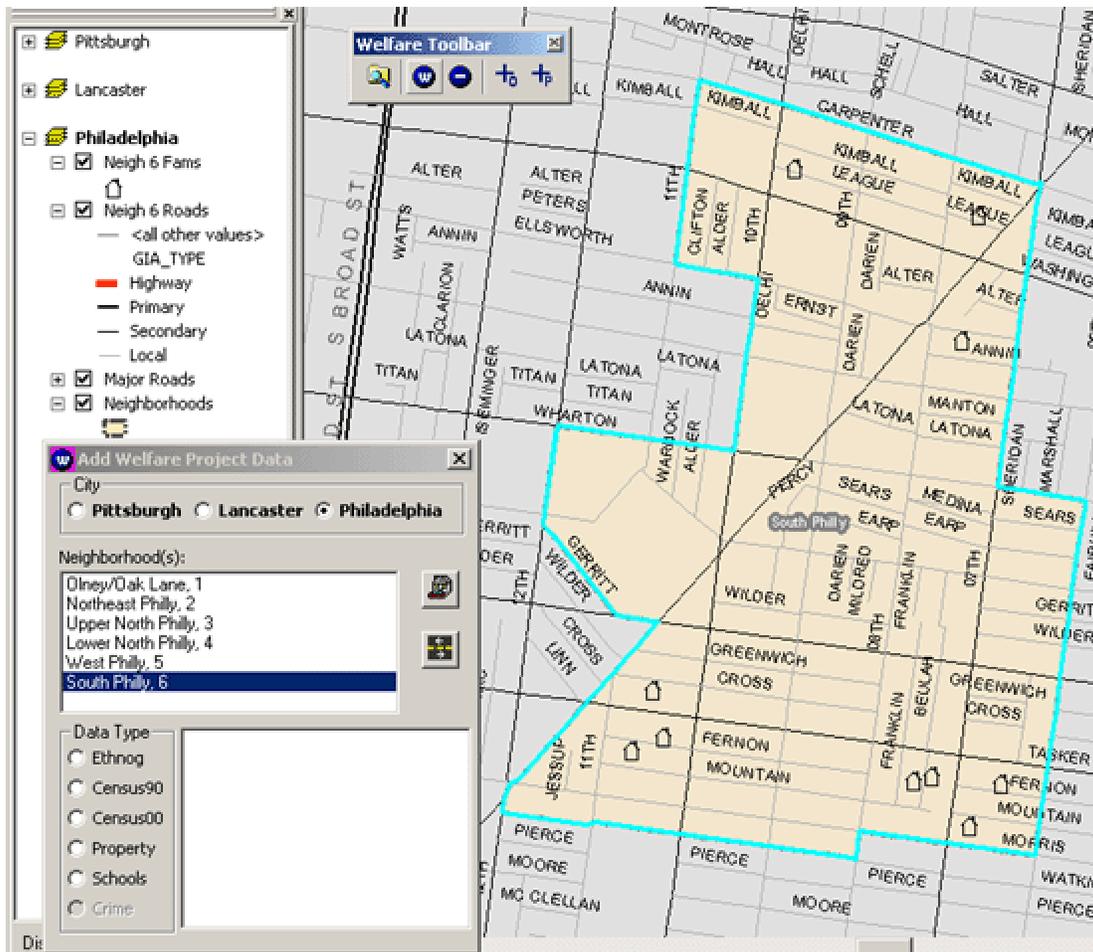


Figure 13. Family locations (imaginary in this example) and roads are added using the two command buttons to the right of the neighborhood list box.

## Summary

Development of the Welfare Project Extension for ArcMap has illuminated a few important themes that may be of use to other application developers:

- Hyperlinking is an underutilized tool in GIS. In this particular project, hyperlinking was used to provide access to field notes written by ethnographers during their visits to the study neighborhoods. Geocoding the field notes has greatly increased their value by making it possible to read them in a spatial context.
- GIS projects often team non-GIS users with knowledge in the project's subject matter with people GIS experts. It is sometimes difficult to bridge the knowledge gap between these two groups to ensure that GIS is used effectively in the project. The approach used in this project was to develop custom tools that enabled GIS novices to view important contextual information quickly and with just a small amount of training.
- Related to the issue of being teamed with non-GIS users is the notion of working proactively vs. reactively. The general tendency in such situations is to simply provide whatever products are requested (e.g., "I need a map of..."). Developing a custom application that provides a simplified, intuitive user interface is one example of a more proactive approach. Putting the GIS tools and data into the hands of the domain experts enables them to get a better sense of how geographic information and analysis can help their project.

## Acknowledgements

I would like to thank everyone on the project team who gave advice on the application, helped to gather data, and of course, provided funding support, especially Stephen Matthews, Linda Burton, Alan Benjamin, Brian McManus, and Jason Smith.

## Author Information

Jim Detwiler  
Geographic Information Analysis Core  
Population Research Institute  
Penn State University  
University Park, PA  
detwiler@essc.psu.edu