Paper 142-2007

People, Pattern, and Place: SAS/GRAPH Data Display of Immigration Activity Across the United States

Wendy B. Dickinson, Ringling School of Art and Design, Sarasota, FL
Anthony J. Onwuegbuzie, University of South Florida, Tampa, FL
Constance V. Hines, University of South Florida, Tampa, FL

ABSTRACT

Federal immigration policy is at the forefront of current public and legislative scrutiny across the United States. Questions of economics, ethnicity, and population movement comprise this public discussion.

The 2005 Year Book of Immigration Statistics and the 2005 American Community Survey, (United States Census Bureau, 2006) offer nationwide, empirical data by city, state, and region pertaining to immigration to the U.S and population mobility patterns. Using multiple SAS/GRAPH procedures, we developed a visual exploration of these quantitative data to help identify patterns and discern themes not readily apparent or easily accessible in the numeric format. Historically, quantitative data have facilitated graphical display techniques by offering a visual one-to-one correspondence of number to graphical element. Today, PROC GMAP, within SAS/GRAPH, provides an opportunity for one-to-one geographical correspondence of data to physical location.

Our presentation is constructed to provide an enhanced kinetic display, using a video monitor and laptop to present each type of graphical procedure within a PowerPoint presentation. By showing the SAS code first and then the resultant graph, we provide a visual and practical link between procedure, code, and output.

CONTEXT

Overview of Graphical Display

> As knowledge increases among mankind, and transactions multiply, it becomes more and more desirable to abbreviate and facilitate the modes of conveying information from one person to another, and from one individual to many.

While written by William Playfair in 1801, the idea of conveying information still is regarded as timely and valuable in the contemporary present. Images can be visual renditions or representations of ideas, dimensions, and events. Representing statistical ideas and information is a complex task—and as Tufte (1990) stated, “all communication between the readers of an image must take place on a two-dimensional surface” (p. 12).

Historically, quantitative data has facilitated graphical display techniques—offering a visual one-to-one correspondence of number to graphical element. A typical example of this one-to-one correspondence is the scatterplot, with each Cartesian coordinate pair represented by a plotted point, providing a visual summary of the measure of association between the variables of interest. Today, Proc GMAP, within SAS/GRAPH, provides an opportunity for one-to-one geographical correspondence of data to physical location.

Graphical Communication: Summary and Exposure

Graphs, like other communication forms, serve different purposes, with the most common goals being visual summary and exposure (Friendly, 1995). Summary refers to the visual inventory of data created to present the viewer with an informed understanding of the underlying information. Depending on the initial format or source, visual summary may embrace a variety of structural composites. Powsner and Tufte (1997) advocated the creation of effective visual summaries, while encouraging the researcher to retain a diversity of methods for data representation.

Information (and meaning) is conveyed through the composition and integration of multiple lines within an image. A familiar example of multiple lines within an image would be geographical maps: using multiple lines, with differentiated line thickness, colors, and weight to portray three-dimensional information on a two-dimensional surface. Tufte (2006) noted “maps show information with differentiated lines all the time, with greater richness” (p. 71). By incorporating differentiated line qualities within SAS data display techniques, we can illustrate research information with enhanced clarity, thereby providing a visual summary of the phenomena of interest.

Immigration Policy

Federal immigration policy is at the forefront of current public and legislative scrutiny across the United States. Questions of economics, ethnicity, and population movement comprise this public discussion. It has been noted that the volume of immigrants to the U.S. in the past two decades mirrors that which occurred in the first decade of the 1900’s, over one million immigrants annually (Spain, 1999). While the influx of immigrants at the beginning and end of the 20th century sparked debates on many similar issues such as how these newcomers would fit into the U.S. society, and what the impact on the society would be, there are nonetheless distinct differences in the immigration patterns between these two periods that are worth examining. In the early 1900’s most immigrants came from the European countries and Canada and arrived in the U.S. by sea. They settled in the northeastern United States and the Midwest. The 1980’s and 1990’s experienced much greater diversity in its immigrant pool. Central America (Mexico, in particular) and Asia have emerged as major countries of origin of these immigrants who now come to the U.S. by air, sea, and land and settle in a wider range of cities (Spain, 1999). Martin and Midgley (1994) note among other things, that this latter wave of immigrants differs significantly in ethnicity, education, and skills when compared to native-born Americans. In order, then, to inform the discussion about the impact that current immigration patterns may have on the U.S. in terms of demographic, socioeconomic, cultural, and political trends, an initial first step is to have a sense of the destination points of current immigrants across the U.S. landscape as well as the mobility patterns of the U.S. population, in general. Two data sources are particularly useful for these purposes.

The 2005 Yearbook of Immigration Statistics published by the Office of Immigration Statistics (OIS) provides detailed statistics on U. S. immigrants compiled in five different areas: legal permanent residents, refugees and asylees, naturalizations, nonimmigrant admissions, and enforcement actions.

A second source, the 2005 American Community Survey (ACS), recently released by the United States Census Bureau offers nationwide, empirical data by city, state, and region. The four major topics in the ACS are (1) demographic characteristics (including gender, age, households by type, race, and Hispanic origin); (2) social characteristics (including school enrollment,

educational attainment, marital status, fertility, veteran status, disability status, residence one year ago, place of birth, U.S. citizenship status, year of entry, language spoken at home); (3) economic characteristics (including employment status, occupation, industry, income, and poverty status); and (4) housing characteristics; including occupancy, units in structure, year structure built, number of rooms, number of bedrooms, and occupants per room (United States Bureau of the Census, 2006). Individual and aggregate data were collected from residents and communities to provide a comprehensive, aggregate numeric summary of these four topical areas.

Data drawn from these two sources were used to provide visual mappings of areas of concentration that serve as destination points for recent immigrants to the U.S. and patterns of mobility of the U.S. population within a one-year period.

PURPOSE

Using SAS/GRAPH procedures, we developed a visual exploration of these quantitative data to help identify patterns and discern themes not readily apparent or easily accessible in the numeric summary format. Our presentation is constructed to provide an enhanced kinetic display, utilizing a video monitor and laptop to present each type of graphical procedure within a power point presentation. By showing first the SAS code, and then the resultant graph, we thus provide a visual and practical link between procedure, code, and graphical output.

METHOD

Variables of interest selected from the two databases are shown in Table 1.
Using these databases, the *Yearbook of Immigration Statistics (OIS) and* the *American Community Survey* (ACS), data were input from each source as shown in Figure 1, Algorithmic Flowchart. Variable values were matched by state and year to provide a more complete picture of United States immigration activity. The unit of analysis for the data was the state.

Utilizing this programming schematic, both 2- and 3-dimensional graphs were developed. Within PROC GMAP, the two- dimensional graphing procedure produces the choropleth mapping; whereas the three-dimensional graphing procedures produce block, prism, and surface maps. We utilize the GMAP procedure to “produce maps, summarize data that vary by physical area, show trends and variations between geographic area, and highlight regional differences or extremes” (SAS Institute, 2004, p. 996).

Table 1.

Algorithmic Variables

Variable name	SAS variable name	Type of variable	Data base
Legal permanent resident status	LEGPERM	categorical	OIS
Country of last residence	LASTRESC	categorical	OIS
Year	YEAR	continuous	OIS, ACS
State	STATE	categorical	OIS, ACS
Educational level	EDUCLVL	categorical	ACS
Mobility	MOBILITY	continuous	ACS
Non-Immigrant admissions	NONIMMAD	continuous	OIS

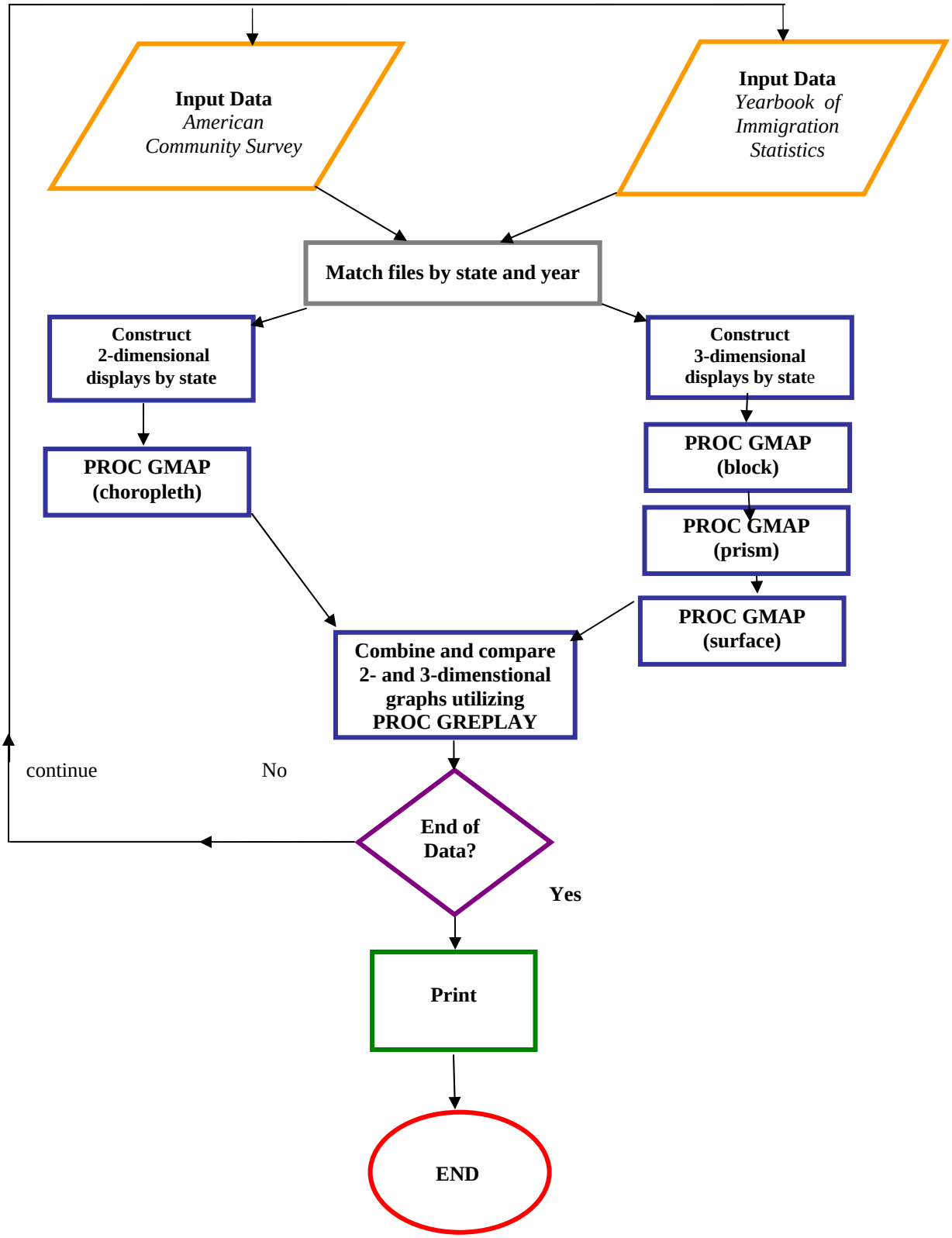

Figure 1. Algorithmic Flowchart

RESULTS AND DISCUSSION

Graphical Display: SAS Sample Code by Procedure

SAS code was developed to invoke the SAS/GRAPH GMAP procedure for both two-dimensional (choropleth) and three-dimensional (block, prism, and surface) maps. The two-dimensional graph (Figure 2) uses the same geographic reference structure as the three-dimensional prism graph (Figure 3), but lacks the added attribute of depth. Figure 2 and Figure 3 both use the same color scheme for consistency of design. However, in Figure 3, the addition of depth provides a visual indicator of amount based on the "lifting" of each state to reflect frequency of persons obtaining legal permanent resident status. Thus, it is easier to discern differences in frequency with the three-dimensional prism graph, as opposed to the two-dimensional choropleth mapping. The SAS code for each type of graph is presented, as well as the corresponding Figure.

SAS code for Proc GMAP Choropleth maps: Figure 2

```
/*GRAPH Figure 2*/
/* choropleth map : 2- dimensional PROC GMAP */
/* Define titles and footnotes for map */
title1 'Persons Obtaining Legal Permanent Resident Status';
title2 'Frequency By State, 2005';
footnote j=r 'Data source:Office of Immigration Statistics';

proc gmap map = maps.us data = legalres;
id state;
choro res2005/ coutline = gray;
run;
quit;
```

SAS code for Proc GMAP Prism maps: Figure 3

```
/*GRAPH Figure 3*/
/*prism map : 3- dimensional PROC GMAP*/
/* Define titles and footnotes for map */
title1 'Persons Obtaining Legal Permanent Resident Status';
title2 'Frequency By State, 2005';
footnote j=r 'Data source:Office of Immigration Statistics';

proc gmap map = maps.us data = legalres;
id state;
prism res2005/ coutline = gray
                xlight= 5
                xview= .75
                zview = 5;
run;
quit;
```

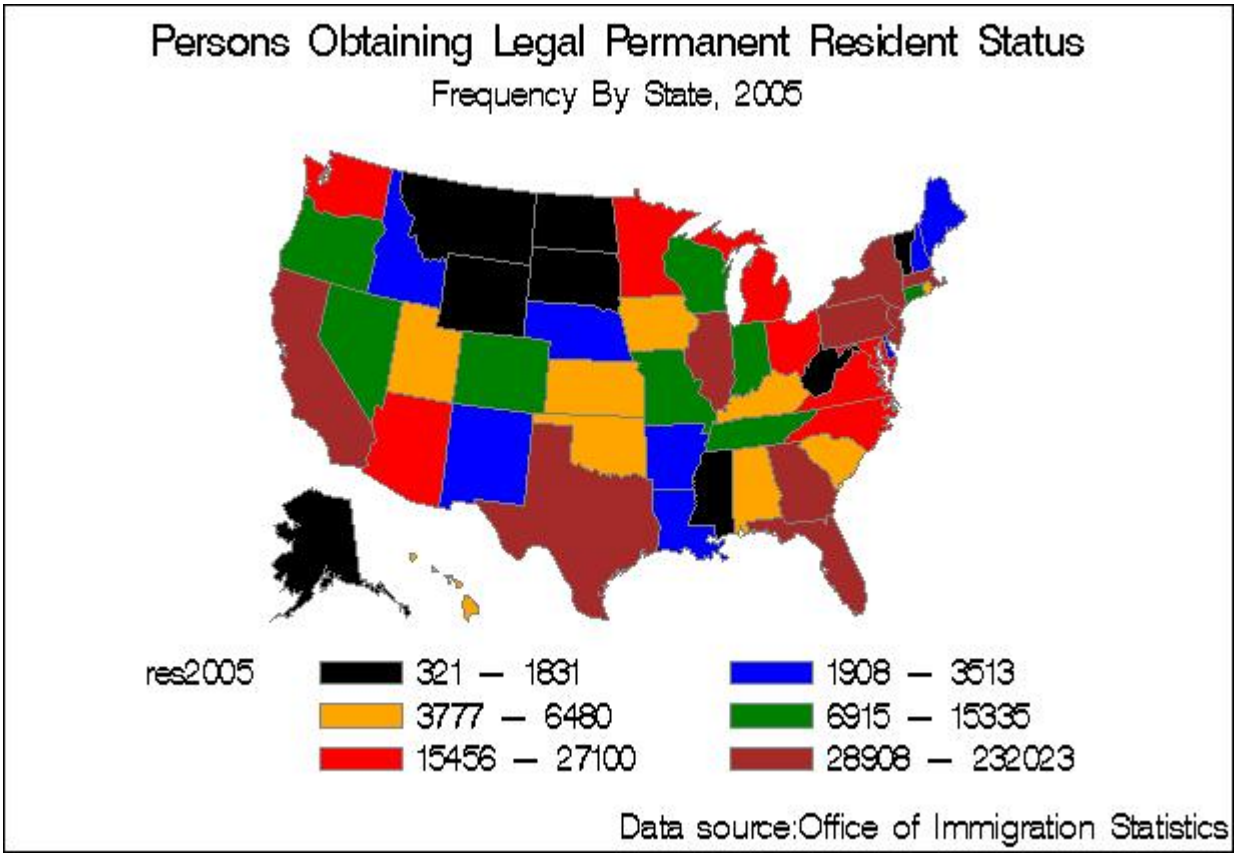

Figure 2. Persons Obtaining Legal Permanent Resident Status

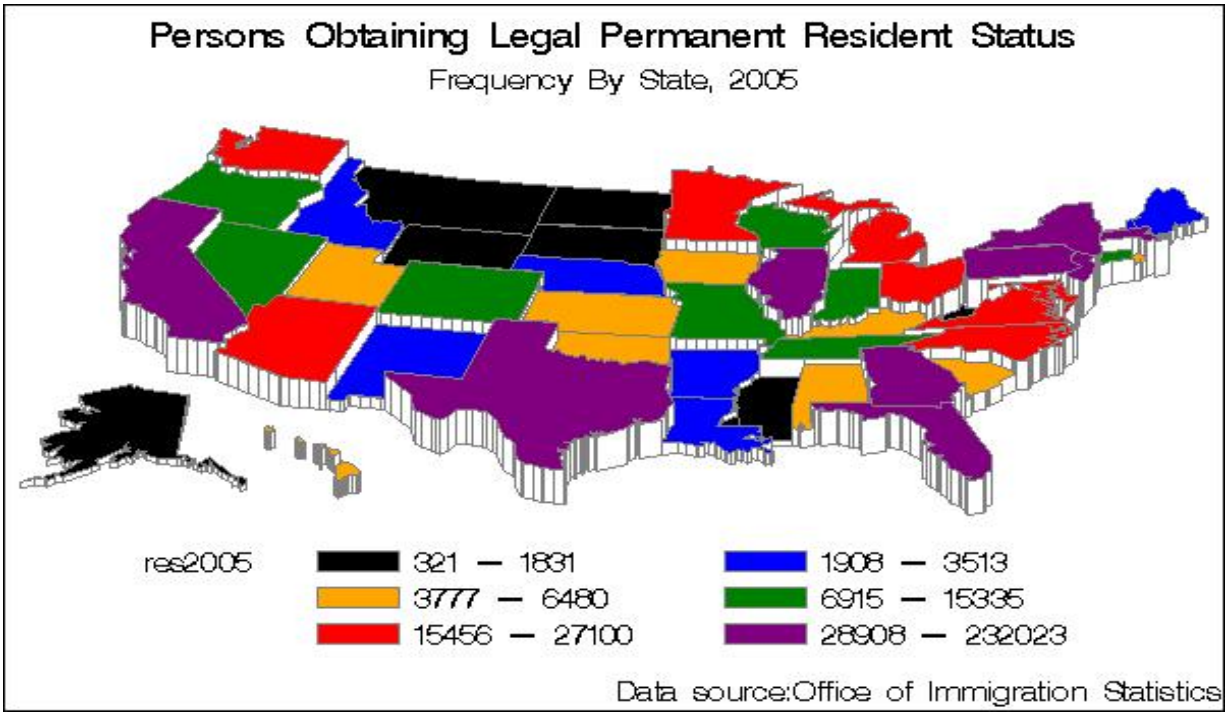

Figure 3. Persons Obtaining Legal Permanent Resident Status

Form I-94 "is a non-immigrant visa waiver arrival and departure form that must be completed and signed upon entry into the United States" (United States Immigration Support, 2007). Categories of admission include tourist and business traveler waivers, students, temporary workers, diplomats, and other classes. Figure 4, a block map, shows the resultant graph from the code presented below. The block map shows the large numbers of nonimmigrant admissions present in the Northeast states, and along the eastern coast of the United States. Higher blocks indicating larger numbers of admissions also are seen along the border states (Washington, Texas, California, and New Mexico), and Hawaii.

SAS code for Proc GMAP Block maps: Figure 4

```
/*GRAPH Figure 4*/
/* block map: 3-dimensional PROC GMAP */
/* Define titles and footnotes for map */

title1 'NonImmigrant Admissions:I-94 status';
title2 'Frequency By State, 2005';
footnote j=r 'Data source:Office of Immigration Statistics';
pattern9 value=mempty color=black repeat=50;
proc gmap map=maps.us data=legalres;
  id state;
  block nonresad / discrete
            area=1
            nolegend
            shape=block
            coutline=black
            cblkout=black
            woutline=3;
run;
quit;
```

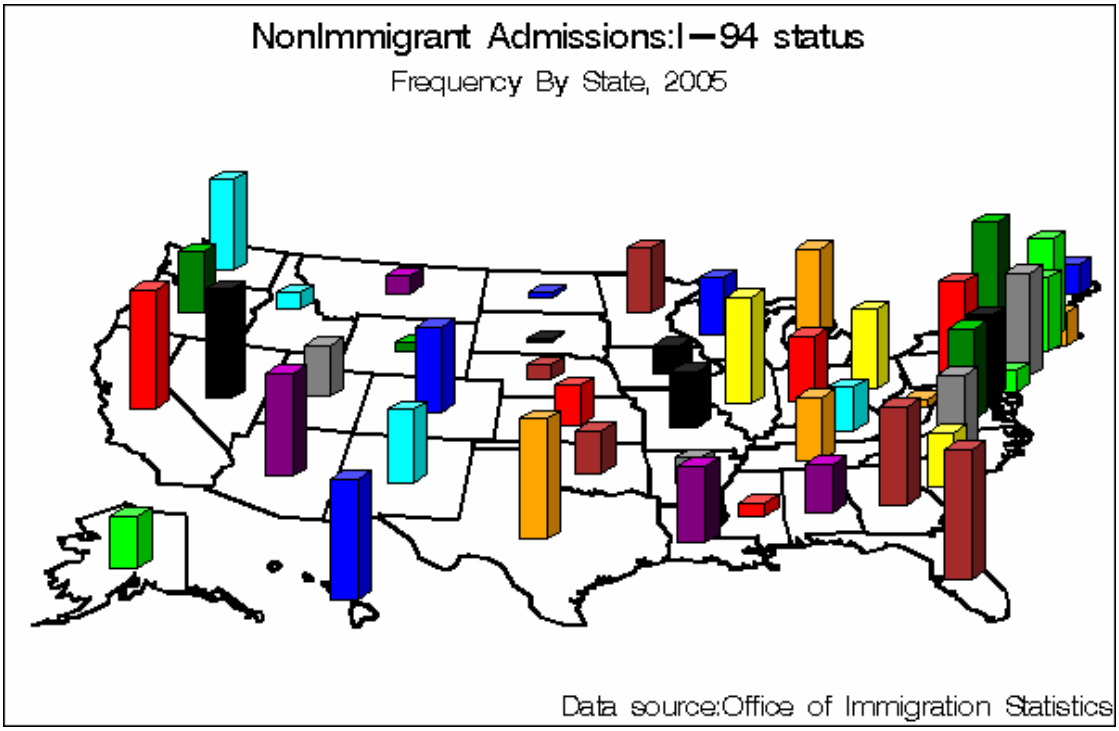

Figure 4. Nonimmigrant Admissions: I-94 Status

SAS code for Surface Maps – Figures 5 and 6

The code written to produce surface maps, as shown in Figures 5 and 6, is presented below. Surface maps display a “spike” based on the unit of analysis. In our example, unit of analysis is state. Each state has a “spike” drawn to represent the frequency of the variable of interest. The height of the spike reflects the reported value of the variable -- the higher the spike, the greater the value. This conceptual congruence provides a visual “yardstick” within and between graphs to enhance communication of the underlying relationship.

Figure 5 displays the numbers of persons obtaining legal permanent resident status by state. In Figure 5, we can see the largest numbers of people are shown by three of the border states (California, Texas, and Florida); and the Northeast states, particularly New York.
In Figure 6, mobility is more closely distributed. However, a westerly mobility trend is observed, with more people moving west across the country, and moving towards the borders from the interior states.

```
/*GRAPH Figure 5*/
/*surface map : 3-dimensional PROC GMAP */
 /* Define titles and footnotes for map */
title1 'Persons Obtaining Legal Permanent Resident Status';
title2 'Frequency By State, 2005';
footnote j=r 'Data source: Office of Immigration Statistics';

proc gmap map=maps.us data=legalres;
id state;
surface res2005/nlines=100;
run;
quit;

/*GRAPH Figure 6 */
/* surface map: 3-dimensional Proc GMAP*/
/* Define titles and footnotes for map */
/*American Community Survey*/
/* Mobility =Percent of People who lived in a different state one year ago/
title1 'Percent of People who lived in a different state one year ago';
title2 'Mobility by State, 2005';
footnote j=r 'Data source: American Community Survey';
proc gmap map=maps.us data=legalres;
id state;
surface mobility/nlines=100;
run;
quit;
```

SAS Code: Proc GREPLAY

This code specified four separate graphs produced by the authors to be combined together and printed on one page. By combining these 4 graphs together, small multiples are created which allow for ease of information comparison between graphs. Templates used for our graphs involved the use of a grid, separated into either 4 or 2 panels. SAS reads from left-to-right, top-to-bottom, so the numbering for the 4-panel grid creates a two-by- two matrix grid. Composite graphs constructed using Proc GREPLAY provided an effective visual summary by enhancing comparisons among data strata.

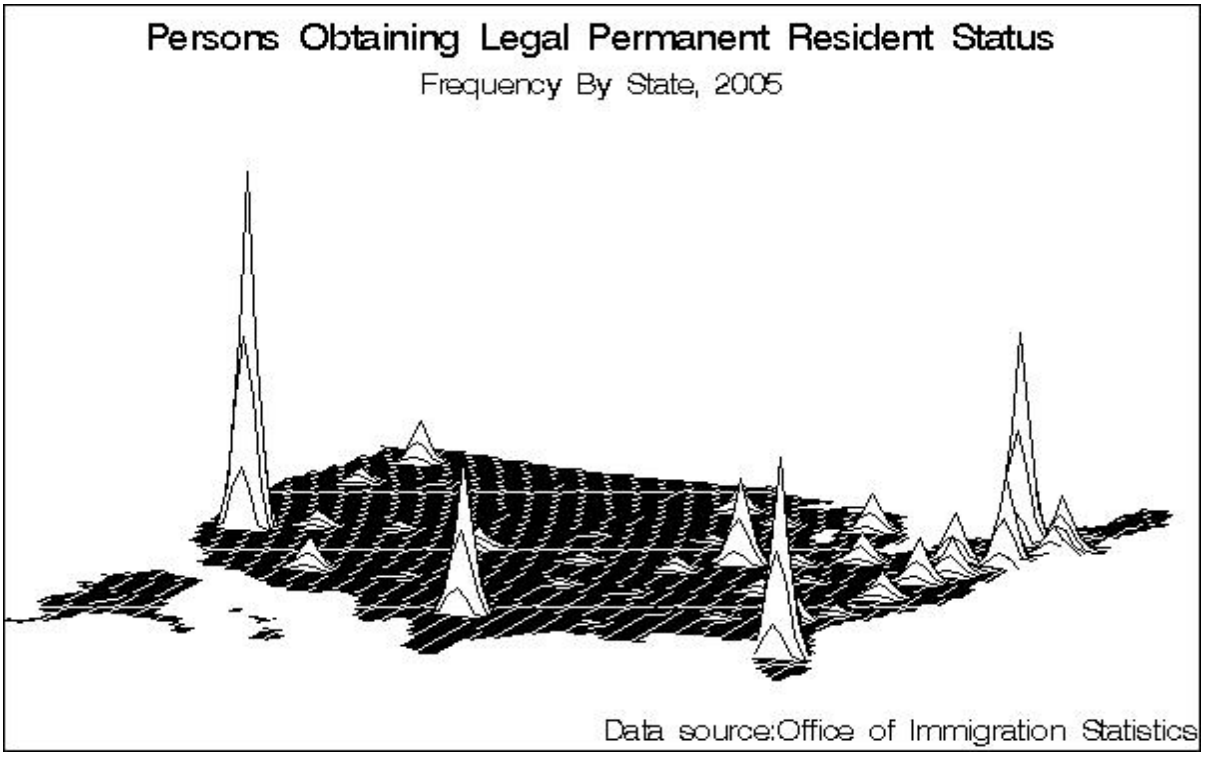

Figure 5. Persons Obtaining Legal Permanent Resident Status

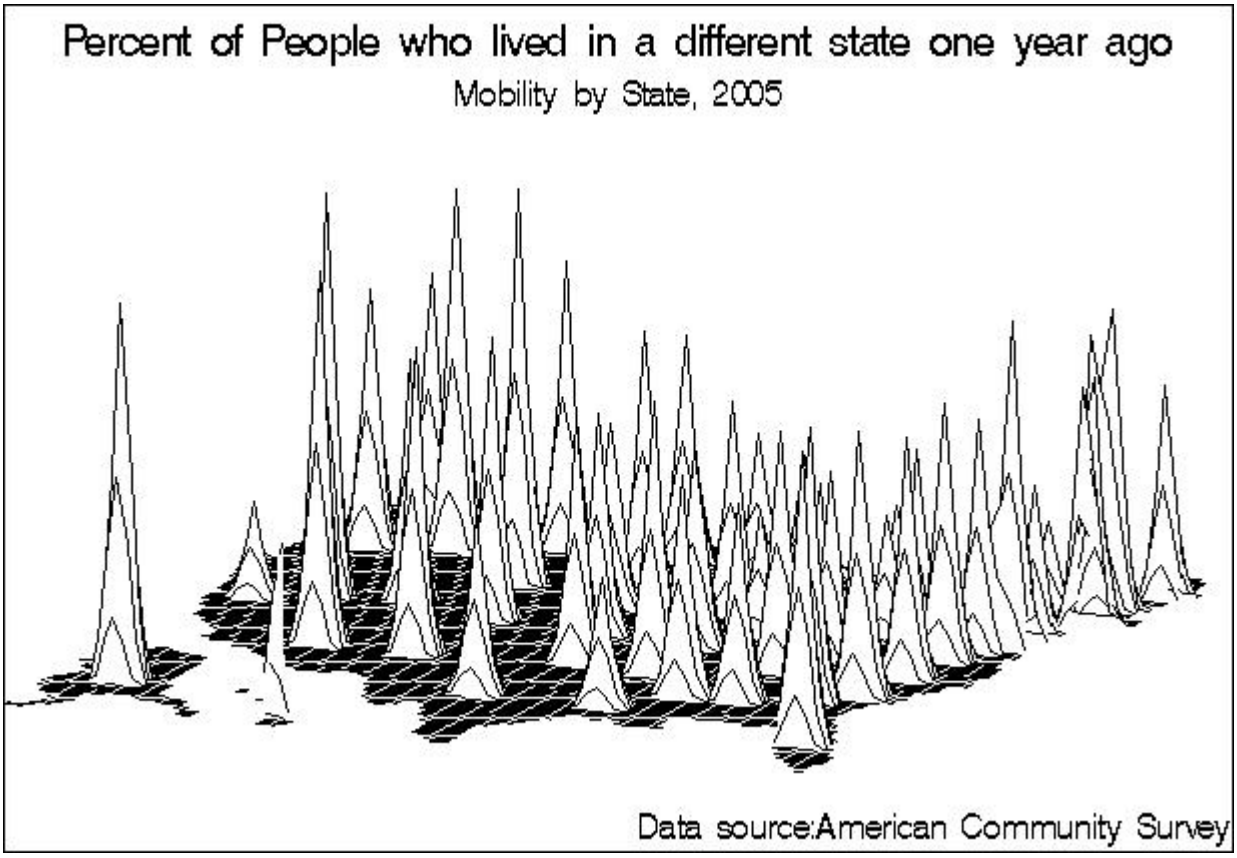

Figure 6. Mobility by State, 2005

Conclusions

With increased access to technology and use of advanced graphing software, the "capacity to produce more efficient and easily understood graphical information displays" is a powerful visual tool for data analysis (Dickinson, 2001, p.4). Indeed, "the overwhelming premise of visual display is that of communication" (Dickinson, 2000, 2002, page 1). Wainer (1992) reminds us of the power of graphical displays; stating the "unrelenting forcefulness inherent in the character of a good graphic is its greatest virtue" (p. 14).

By utilizing two- and three-dimensional displays, we were able to highlight immigration and mobility patterns occurring throughout the United States. For example, in the visual displays shown above, it was easy to see that the geographic areas favored for current immigrant destination are in the southeast, south, southwest, and northeast United States. The graphical displays can help us discover patterns and recognize important truths about our data not readily apparent in a table or text. Graphs can easily show us patterns that might not have been seen otherwise (Wainer, 1990, 1997). As Tukey (1972, 1989) declared, the greatest possibilities of visual display lie in the vibrancy and the accessibility of the intended message. Utilizing SAS/GRAPH provides powerful techniques for contemporary visual display, and enhanced transmission of the underlying meaning and message.

REFERENCES

Dickinson, W. B. (2000). *Escaping Flatland: Chernoff's Faces Revisited*. Doctoral Thesis. UMI Order Number: AAI9968808
Retrieved February 28, 2007, from
http://portal.acm.org/citation.cfm?id=931541&jmp=cit&coll=GUIDE&dl=GUIDE&CFID=15596083&CFTOKEN=41111326#CIT

Dickinson, W.B. (2001). *Envisioning Kinaalda: Navaho magic, mystery, and myth*.
Retrieved February 28, 2007, from
http://www2.sas.com/proceedings/sugi25/25/po/25p228.pdf

Dickinson, W. B. (2002). *Escaping Flatland*.
Retrieved February 28, 2007, from
http://www2.sas.com/proceedings/sugi26/p195-26.pdf

Friendly, M. (1995). Conceptual and visual models for categorical data. *The American Statistician, 49*, 153-160.

Martin, P., & Midgley, E. (1994, September). Immigration to the United States: Journey to an uncertain destination. *Population Bulletin*, 1–40.

Playfair, W. (1801, 2005). Commercial and Political Atlas and Statistical Breviary. Edited and introduced by Howard Wainer and Ian Spence. Cambridge University Press: New York, NY.

Powsner, S. M., & Tufte, E. (1997) Summarizing clinical psychiatric data. Psychiatric Services, 48, 1458-1459.

SAS Institute Inc. (2004). *SAS/GRAPH 9.1 Reference, Volumes 1, 2 and 3.* Cary, NC: SAS Institute Inc.

Spain, Daphne (1999). The debate in the United States over immigration. In Changing America: The United States population in transition. *U.S. Society & Values, 4*(2), 17-20

Tufte, E. (1990) *Envisioning information*. Cheshire, CT: Graphics Press.

Tufte, E. (2001). *The visual display of quantitative information*. CT: Graphics Press.

Tufte, E. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.

Tukey, J. W. (1972). Some graphic and semigraphic displays. In T.A. Bancroft (Ed.) *Statistical papers in honor of George W. Snedecor* (pp.293-316). Ames, IA: The Iowa State University Press

Tukey, J. W. (1989). Data-based graphics: Visual display in the years to come. *Proceedings of the American Statistical Association, 84*, 366-381.

United States. (2006). *Department of Homeland Security. Yearbook of Immigration Statistics: 2005*. Washington, D.C.: U.S. Dept. of Homeland Security, Office of Immigration Statistics.

United States Immigration Support. (2007). Retrieved February 28, 2007, from http://www.usimmigrationsupport.org/i94.html

Wainer, H. (1990). Graphical visions from William Playfair to John Tukey. *Statistical Science, 5*, 340-346.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14-23.

Wainer, H. (1997). *Visual Revelations*. Mahweh, NJ: Lawrence Erlbaum Associates, Inc.

ACKNOWLEDGMENTS
The authors greatly appreciate the assistance of Lisa Adkins in the preparation of this document.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

Dr. Wendy Dickinson
Coordinator of Mathematics
Ringling School of Art and Design
2700 North Tamiami Trail
Sarasota, FL 34234
Fax: 941-953-5071
E-mail: wdickins@ringling.edu