Information at your finger tips: Exploring the US Census Data

Mike Sips*
Stanford University, USA

Jörn Schneidewind†
University of Konstanz, Germany

Daniel Keim‡
University of Konstanz, Germany

Stephen C. North§
AT&T Research Labs

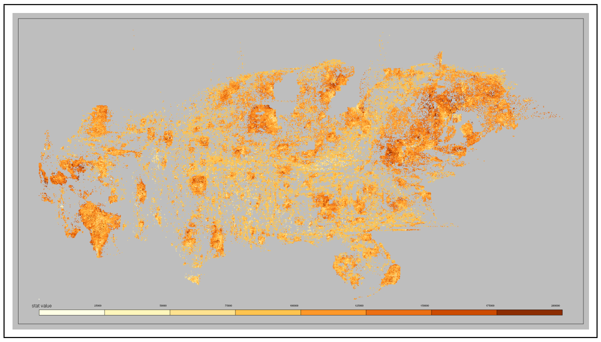


Figure 1: **United States, Year 2000 Median Household Income** – on the U.S. National Level plot there are high income clusters on the East Side of Central Park, and in suburbs of Chicago but not its downtown neighborhood. In the San Francisco area we can identify Silicon Valley; the income in this small area is significantly greater than average (Data=Block Level; Global Shape=Cartogram based on Household Distribution).

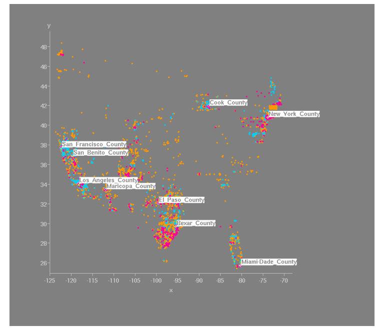
1 Introduction

The research reported here focuses on synthesizing geo-visual analytic methods that support analysis of large, multivariate, geospatial data sets. This contribution uses WALDO [4] as a component sharing and extending existing visual analysis methods. Specifically, we introduce non-overlapping pixilated displays with a novel form of a dynamic labeling of important or interesting geospatial pattern. We introduce a novel, component-based data exploration and analysis tools that integrate computational, visual, and cartographic approaches. We demonstrate the application of these tools to analysis of the census demographics characteristics in the US. The integrated approach is able to (1) perform multivariate analysis and present the result in pixilated displays in which color encodes statistical parameters and (2) provides a framework that scales with the data (from nation level to block level) and the analysis requirement through novel user integration. The methods are applied to a data set containing census demographics data on the block level. The data is extracted from the US Census Database provided by the US Census Bureau [6].

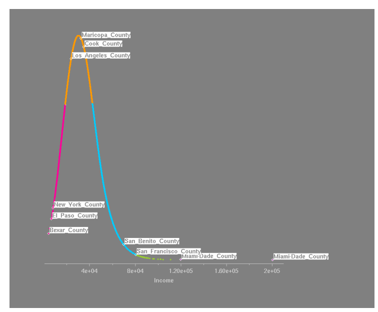
2 Synthesis

To address the unique challenges in analyzing and visualizing census demographics data, we combine non-overlapping pixilated displays that show the statistical parameters at block level and linked statistical displays based on XGobi [1]. We enhance all displays through a dynamic labeling of geo-pattern.

*e-mail: msips@graphics.stanford.edu
†e-mail: schneide@dbvis.inf.uni-konstanz.de
‡e-mail: keim@dbvis.inf.uni-konstanz.de
§e-mail: north@research.att.com



(a) Geographic Distribution – California, Texas, Florida and West Coast.



(b) Median Household Income Distribution – Majority has $40.000

Figure 2: **Linking and Brushing (Hispanic)** – two blocks with very rich households (> $120.000) are located in Miami Dade County, some few blocks with wealthy households ($80.000) are located in Bay Area (San Francisco and Benito County), households with low household incomes are located in Texas (El Paso and Bextar County) and West Coast (New York County).

2.1 Non-Overlapping Pixilated Displays

The basic idea of the PixelMap is to rescale certain map parts to fit better the dense 3d point clouds to unique positions on the output map, since spatial data are highly non-uniformly distributed in the real world. The goal is (a) to find areas with density in the two geographical dimensions (a_i^x, a_i^y) and (b) to allocate enough amount of pixels on the screen to place all data points of dense regions at unique positions close to each other. A detailed description of PixelMaps was published in [3]. Figure 3 shows the global and local distribution pattern of the major ethnic groups in the USA.

2.2 Linking and Brushing

There are many possibilities to visualize multi-dimensional data, each with their own strengths and weaknesses. The idea of linking and brushing is to combine different visualization methods to over-

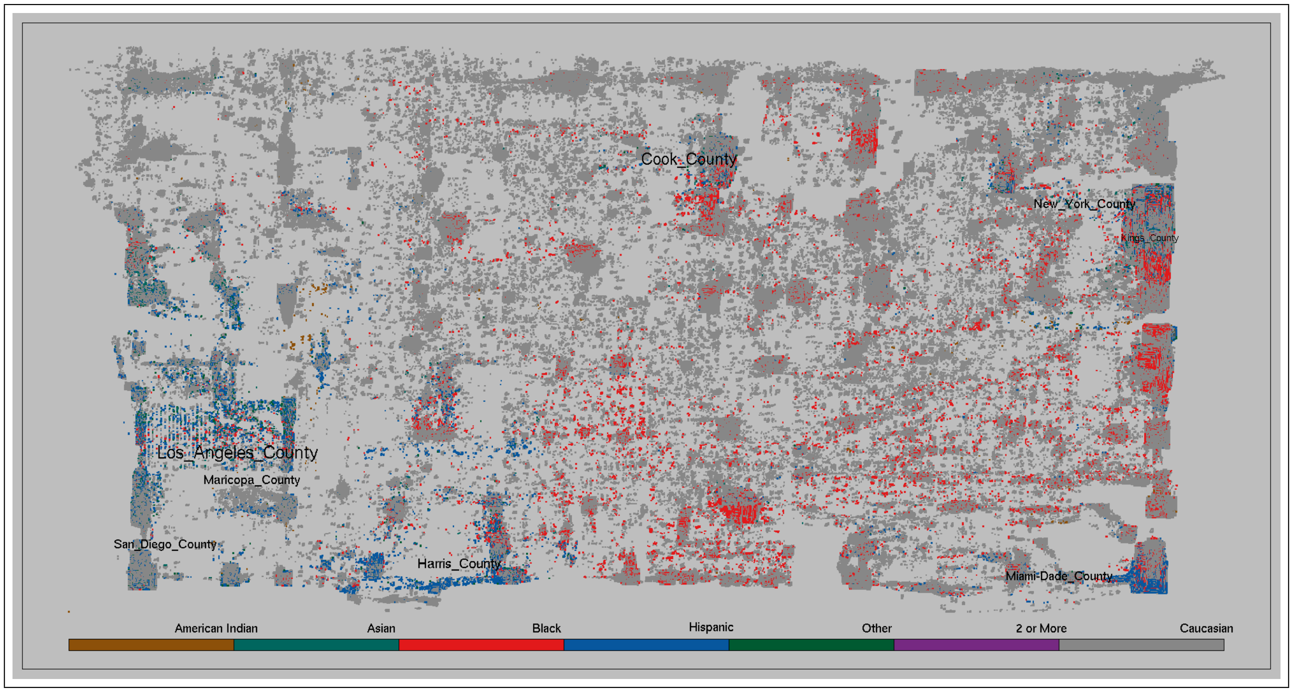


Figure 3: **Global and Local Distribution Pattern of the Major Ethnic Groups – Caucasian:** live in almost every neighborhood in the USA, **American Indian:** major pattern can be observed in Oregon, **Black:** major pattern at the West Coast and the South, **Hispanic:** Miami (Exile Cubans), West Coast, California and Texas (dynamic map labeling shows the eight most populated areas in the USA).

come the shortcomings of individual techniques. We linked and brushed a traditional map visualization with scatter plots of different projections of the census data. More precisely, we combined subsets of points in all projections by coloring and dynamic labeling of each data point. Figure 2 shows the relationship between median household income and geographic distribution.

3 Visual Analysis

3.1 Explicit Control of the Global Shape

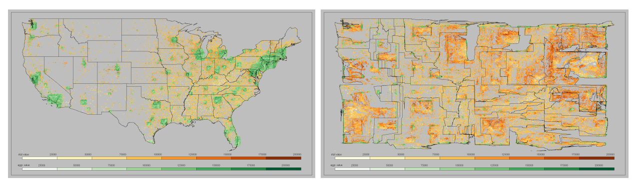
Recent information visualization research has addressed other types of transformation functions that make spatially-transformed maps with recognizable shapes. These types of spatial-transformation are called global shape functions. In particular, cartogram-based map distortion has been studied. On the other hand, cartogram-based distortion does not handle point sets readily.

The overall approach is to compute a global shape using CartoDraw [2], then place the data points into the global shape using PixelMap. Figure 1 United States, Year 2000 Median Household Income.

3.2 Scalability: From PDA to Wall-Size Displays

The resolution of the display space plays an important role in the degree of overlap and therefore in the PixelMap optimization. Nowadays, high-resolution pixilated displays are increasingly available in both wall-sized and desktop units.

Therefore, an appropriate trade-off between shape distortion and the degree of overlap for the PixelMap display has to be separately specified for each output resolution. Our Scalable PixelMaps approach addresses this request. A detailed description of our scalability issues was published in [5]. Figure 4 shows the scalability of our framework to different screen resolutions.



(a) PDA Resolution (b) Wall-Size Display

Figure 4: **From PDA to Wall-sized Displays** – appropriate trade-off between shape distortion and the degree of overlap for different output devices. The figure shows the US Census Year 2000 Median Household Income (block level) for different output devices

References

[1] AT&T Labs Research. XGOBI and XGVIS website, 2006. http://www.research.att.com/areas/stat/xgobi/, Sept. 2006.

[2] D. A. Keim, S. C. North, and C. Panse. Medial-Axis-based Cartograms. *IEEE Computer Graphics and Applications*, 25(3):60–68, May/June 2005.

[3] M. Sips, D. A. Keim, S. C. North, and C. Panse. Pixel based visual mining of geo-spatial data. *Computers & Graphics (CAG)*, 28(3):327–344, June 2004.

[4] M. Sips, C. Panse, D. A. Keim, and S. C. North. Visual data mining in large geo-spatial point sets. *IEEE Computer Graphics and Applications (CG&A), September-October 2004*, 24(5):36–44, September 2004.

[5] M. Sips, J. Schneidewind, D. A. Keim, and H. Schumann. Scalable pixel-based visual interfaces: Challenges and solutions. In *Information Visualization (IV 2006), July 5-7, London, United Kingdom*. IEEE Press, July 2006.

[6] United States Department of Commerce. Census Bureau website, 2006. http://www.census.gov/, March 2006.