π-Flow: Flow & charts for the study of population movement

Benoit Gaudin, Mike Bennett, Brendan Sheehan, Aaron J Quigley, *Member, IEEE*

Abstract— π-Flow is a series of pieces of demonstration software for the visualisation of characteristics of human population dynamics which are widely studied in both economics and social science. Our visualisations are based on a 1% sample of the 2002 census in the USA. The modern census includes one aspect of historical data and it is this data namely, "*Which county did you live in 5 years ago*", that we use to study population flow through the use of the adapted MapFlow algorithm in π-Flow. In addition we present an attractive/repulsive visualisation that shows an aggreate view of net popultion gain and loss. Finally, we present an established symbolic technique based on a pie-chart per state based visualisation that aids in the study of the ethnic breakdown across the lower 48 United States.

Index Terms—Map Flow, Pie Chart, Visualisation, Graph Drawing.

1 Introduction

To understand the current state of a population governments or business typically collect statistical data about this population through sampling, survey or census. A census generally attempts to include everyone in a given country, so as to build up an accurate understanding of the breakdown of the population including age, race, employment and location. As a result census data is collected and processed to ensure it is anonymous. While such anonymous records can sometimes be linked to an individuals' identity, the census doesn't provide a means to track individuals across multiple census events. As such, much of the demographic data collected in a single census is primarily suitable for economic, governmental and marketing purposes. The study of the longitudinal changes in such data has given rise to the field of demography. This area of study illustrates the changing structure of population due to births, deaths and *migration*. While migration patterns can be studied in a textual or tabular form our interest is in the use of graphics to aid comprehension, as human beings are inherently skilled at understanding data in visual forms.

For the purposes of this study we are concerned with the temporal data and the multidimensional data within the census data set. We are also concerned with the imposition of a network model (a tree) for the display of the migratory pattern on a state-by-state basis. Naive approaches to the display of statistical or relational data under pinned by a well understood geometry (namely a map) often suffer from readability problems to the varying sizes of both populations and regions of interest. We show in π-Flow that we can trade off positional accuracy for improved readability in the display of state based ethnic diversity, countrywide state attractiveness and unattractiveness and the detailed population flows from state-to-state.

2 Data Manipulation

The π-Flow suite of visualisations is developed using standard data-visualisation pipeline architecture. We first pre-processed the 1% public use microdata sample (PUMS) from the 2002 Census using a Java parser. This pre-processing allowed us to load a series of tables (on a per state basis) into an optimised MySQL 5.0 database. The visualisation is produced using Java2D and the JUNG toolkit.

• *Mike Bennett, Brendan Sheehan, Benoit Gaudin and Aaron Quigley are with the Imaging, Visualisation & Graphics Laboratory, University College Dublin, E-Mail: {mike.bennet, brendan.sheehan, benoit.gaudin, aquigley}@ucd.ie*

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

2.1 FlowMap

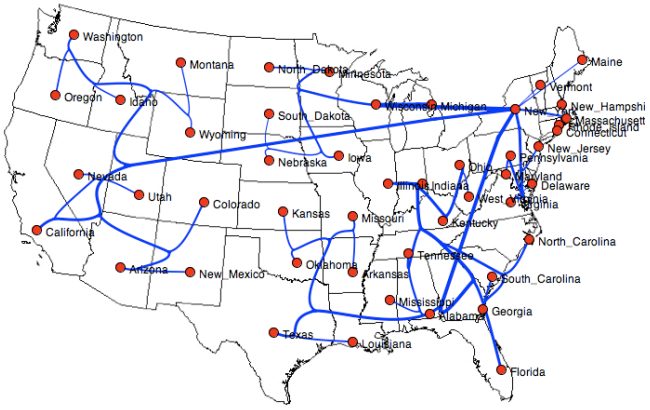

Fig. 1. FlowMap showing the population flow in one 5-year period from Michigan to all other states in the lower 48 states (without dynamic node movement on).

Fig. 2. FlowMap showing the population flow from California with intelligent node distortion.

Flow map layout is a visualization technique used in cartography to illustrate the volume of movement of objects among locations [3]. Flow maps can be used to depict migrations, invasions [3], trade and other data where there is flow from one source to many destinations as shown in Figures 1 and 2. The main advantage of a flow map is

that it reduces clutter by merging edges, which significantly improves the overall readability of the graph. Good flow map layouts generally have the following characteristics; intelligent distortion of destinations to ensure that flow lines have enough room to get to their destinations, merging edges that share destinations and intelligent edge routing. We implement an adapted version of Phan et al's flow map layout algorithm [2] to visualize migration flow between states based on the census data. The positions of the nodes do not always match the position of the state. Instead the state node remains in proximity of the state location but is positioning so as to improve readability and to emphasise the flow or thickness of the edge. Here the thickness of the lines is represented on a logarithm scale where the thickest lines represent the movement of 21,341 people and the thinnest just 19 people. Figure 2 clearly shows the major movement from people from California to it neighbouring states and notably to some of the east coast states.

2.2 Attractive/Repulsive States

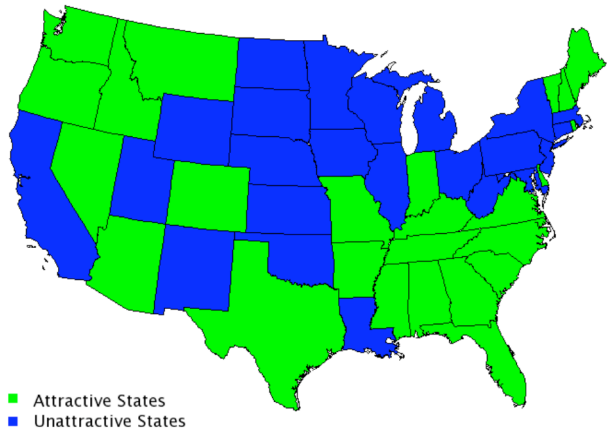

Fig. 3. Coarse Visualisation, which displays the inter-census measure to indicate which states have had a net-loss or net-gain due to migration.

This coarse grained visualisation is intended for use with either the flowmap or state based ethnic breakdown.

2.3 State based ethnic breakdown

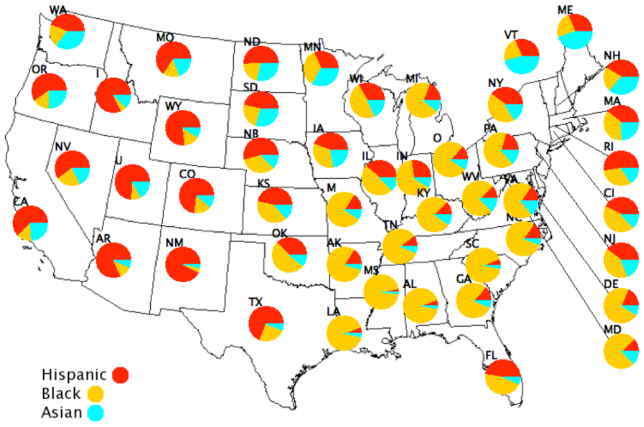

Fig. 4. Pie Chart Visualisation of the lower 48 states showing the proportions of Hispanic, Black and Asian ethnicities.

The visual display of absolute statistical values is not appropriate when comparative study of the data is important. Simkin and Hastie have demonstrated that proportion judgements based on pie charts (circles divided into wedges to represent percentages of the whole) are very accurate when users are asked to judge proportion of the whole [4]. Their analysis of this perceptual process included details of scanning, projection and superimposition. In answer to the question, where do particular ethnicities tend to congregate we have decided to implement a state based pie-chart visualisation where the user can decide which ethnicities are included or not in a particular exploration. Figure 4 demonstrates a countrywide map with three coded ethnicities displayed; these include Black, Asian and Hispanic.

3 Discussion

While Figure 3 and 4 are course-grained visualisations, clear and surprising patterns already emerge. The population growth in the northwest and the southeast is clear. The settlement of Asian populations on the West coast of the United States is clear. The continued expansion of Hispanic populations from the southern states through the western states emerges as a significant pattern. And the concentration of African American populations in the southern correlated with historical factors such as slavery. Figure 1 and 2 show the utility

4 Conclusion

π-Flow shows three comparative visualisations for the study and exploration of population makeup and movement. The aim of these visualisations is to act as a simple end for the study of a voluminous data set such as the census record without overwhelming the user. Future work aims to expand the methods to a multi-granularity approach where states can be opened to a lower level of granularity such as regions or counties.

Acknowledgements

The authors wish to thank the European Union Marie Curie research program and the Embark/Microsoft Research (IRCSET Ireland).

References

[1] Phan D., Xiao L., Yeh R. B., Hanrahan P., Winograd T.: Flow Map Layout, pp. 219-224, Proceedings of the IEEE Symposium on Information Visualization 2005.

[2] Madadhain, J.O., Fisher, D., Smyth, P., White, S., Boey, Y.B.: Analysis and visualization of network data using JUNG. http://jung.sourceforge.net (2005)

[3] Minard C. J, "Map representing the losses over time of French Army troops during the Russian Campaign, 1812-1813", figure 58 in The Visual Display of Quantitative Information, Edward R. Tufte, (Cheshire, California: Graphic Press, 1983)

[4] Simkin, D. and Hastie R., "An information processing Analysis of Graph Perception" (1987), Journal of the American Statistical Association, Vol. 52, pp. 454-465.