Dynamic Graphics: An Interactive Analysis Of What Attaches People To Their Communities

Jessica M. Orth
Department of Statistics and Actuarial Science
University of Iowa

THE UNIVERSITY OF IOWA

I. Approach

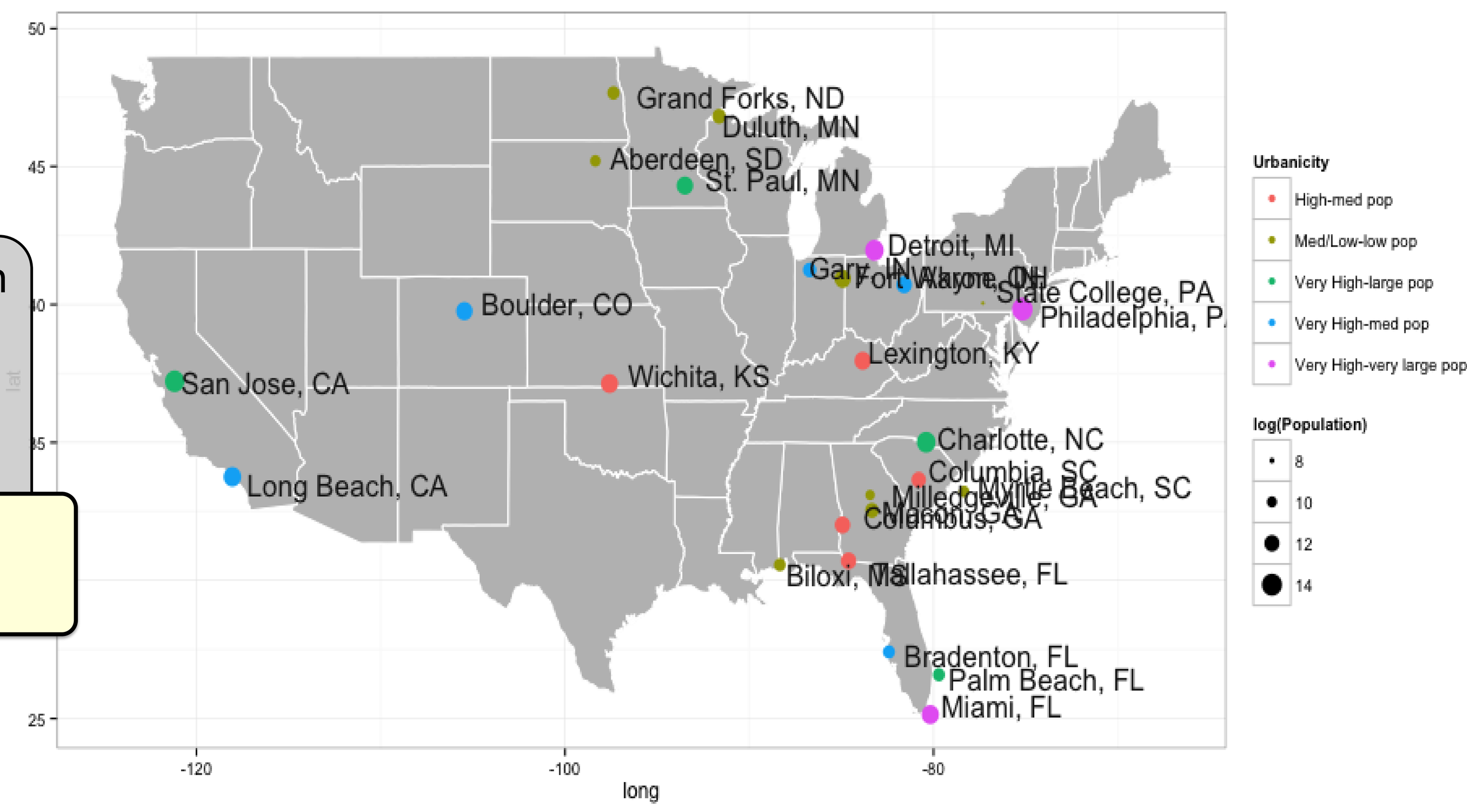
The Data
- 'Soul of the Community'
- Knight Foundation and Gallop
- Three years: 2008-2010
- 43,000 people surveyed in 26 communities across America

Index Variables
- Attachment, Loyalty, Passion, Basic Services, Leadership, Education, Safety, Aesthetics, Economy, Social Offerings, Community Offerings, Civic Involvement, Openness, Social Capital, Community Domains

Summary Statistics
- Means, standard deviations, proportion of high index variables, z-scores

Data Reduction and Data Mining
- Principal Component Analysis
- Multidimensional Scaling

Graphical Displays
- Interactive Motion Charts
- Average Hierarchical Cluster Analysis



Displaying multivariate data can be achieved in many ways through a variety of tools. Here we aim to emphasize the use of motion charts for displaying the trend analysis of time-dependent Principal Component Analysis and Multidimensional Scaling. It is well known that these methods are used as data reduction and data mining techniques in the analysis of multivariate data, but what happens when we introduce a time variable to these results? As will be seen, motion charts provide the tool to seamlessly merge these results throughout time and allow for dynamic and interactive interpretations of what attaches people to their communities.

We analyze the index variables from the 'Soul of the Community' survey conducted by the Knight Foundation and Gallop by looking at four different summary statistics: means, standard deviations, the proportion of high index variables, and z-scores. Means, standard deviations, and proportions are calculated for cities based on the index variables. The z-scores serve as an index themselves, providing information on each city's score for the original index variables: negative z-scores imply a lower score for the index variable and positive z-scores indicate a higher score for that city, relative to the overall score of the original index variable.

II. Key Drivers and Relationships Between them (PCA)

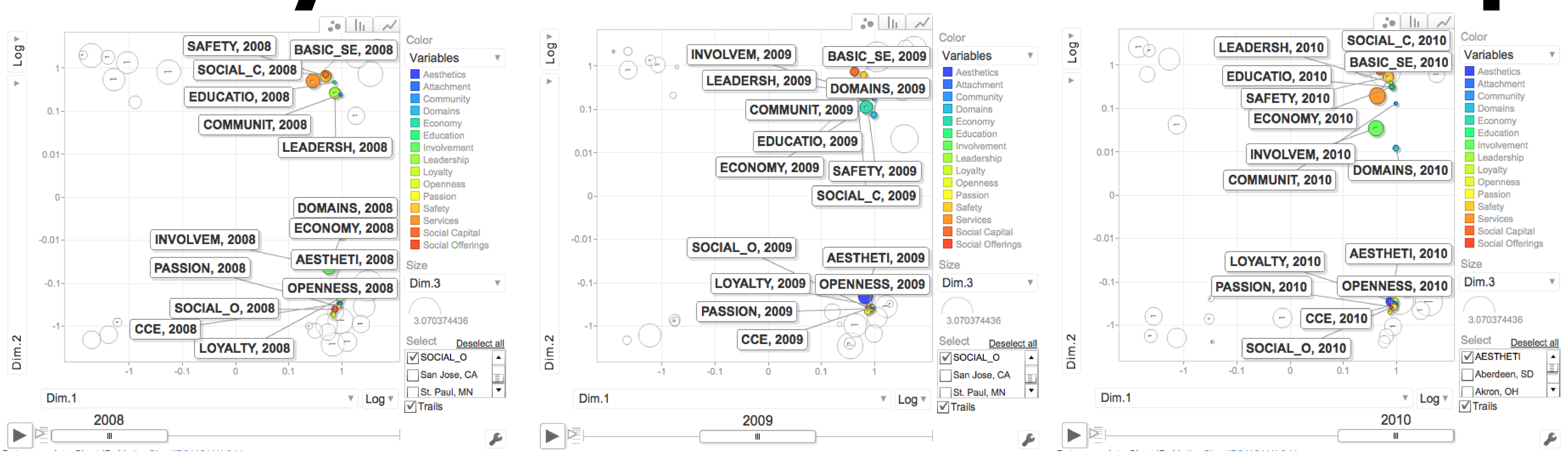


Figure 1: Means

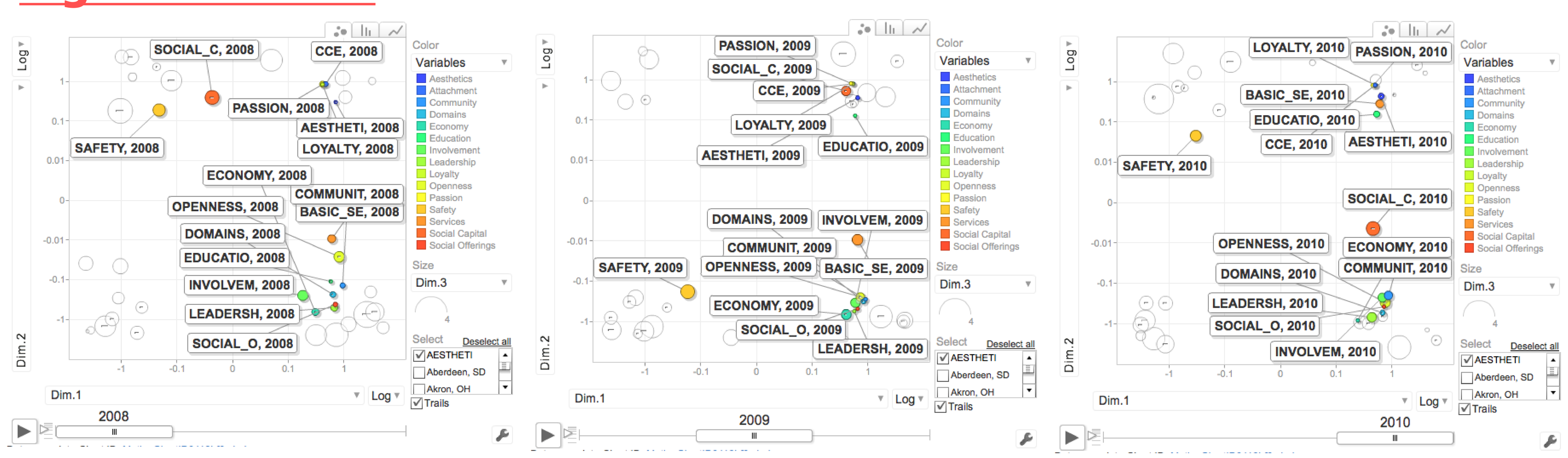


Figure 2: Standard Deviations

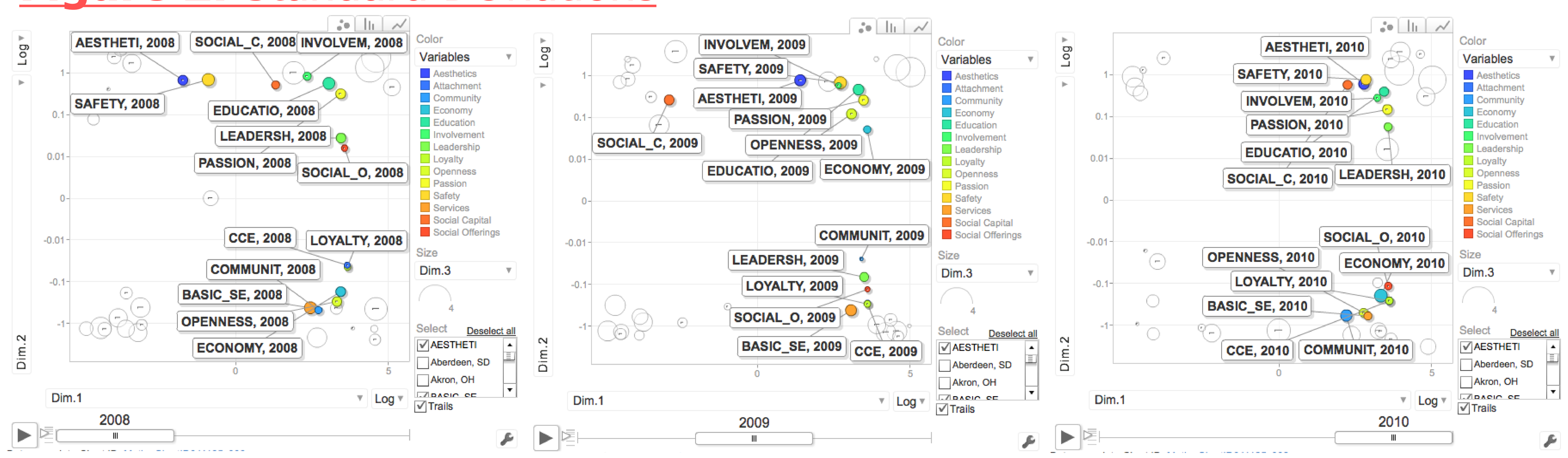


Figure 3: Proportions

It is often said that 'Beauty is in the eye of the beholder', so why not put the analysis in the hands of the user? One of the many beauties of motion charts is the capability to do just this. Why limit the results to a single graphical display? Motion charts allow for customizable analysis to suit the interests of multiple users.

While social offerings, openness, and aesthetics are found to be the leading drivers of community attachment by the Knight Foundation, we look at the relationship between these and the other index variables using Principal Component Analysis.

	Means	Standard Deviations	Proportions
Dimension 1	Overall drivers for attachment	Personal Assurance vs. Overall drivers for attachment	Personal Assurance vs. Overall drivers for attachment
Percentage of Variation Explained	2008: 54 2009: 58 2010: 62	2008: 32 2009: 38 2010: 34	2008: 35 2009: 42 2010: 39
Dimension 2	Economic Growth vs. Emotional Bond	Personal Assurance and Pride vs. Economic Growth	Emotional Bond vs. Economic Growth
Percentage of Variation Explained	2008: 15 2009: 14 2010: 13	2008: 23 2009: 25 2010: 27	2008: 20 2009: 15 2010: 20
Dynamic Drivers	Involvement, Economy, Domains	Safety, Social Capital, Education, Basic Services	Safety, Aesthetics, Social Capital, Leadership, Social Offering, Openness, Economy

III. Differences Between Communities

III.A. Multidimensional Scaling

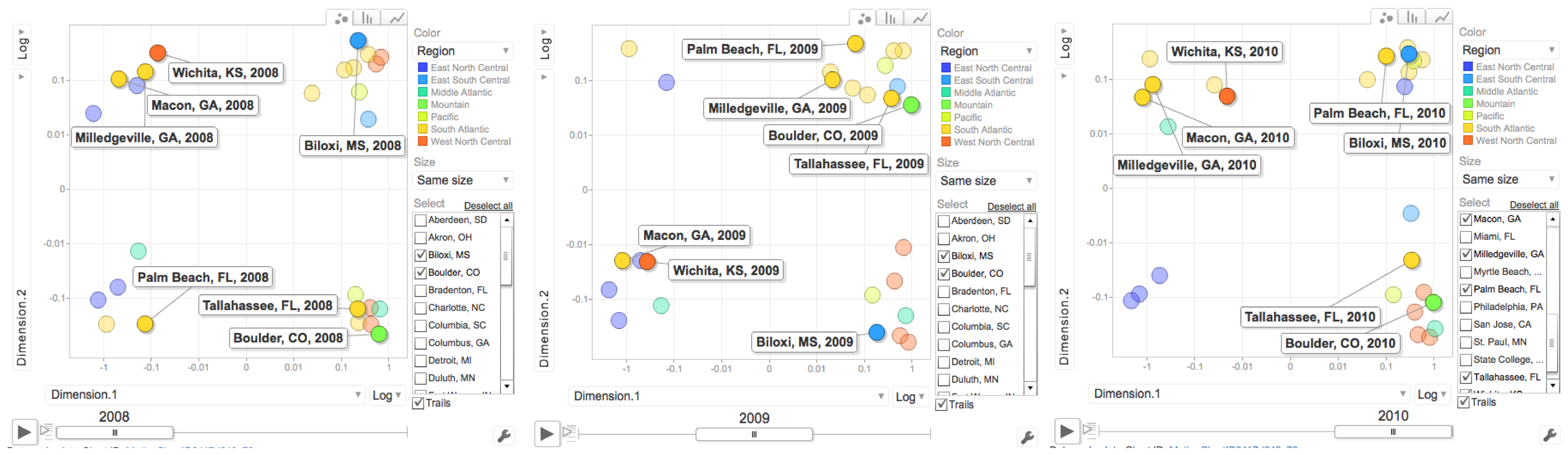


Figure 4: Means

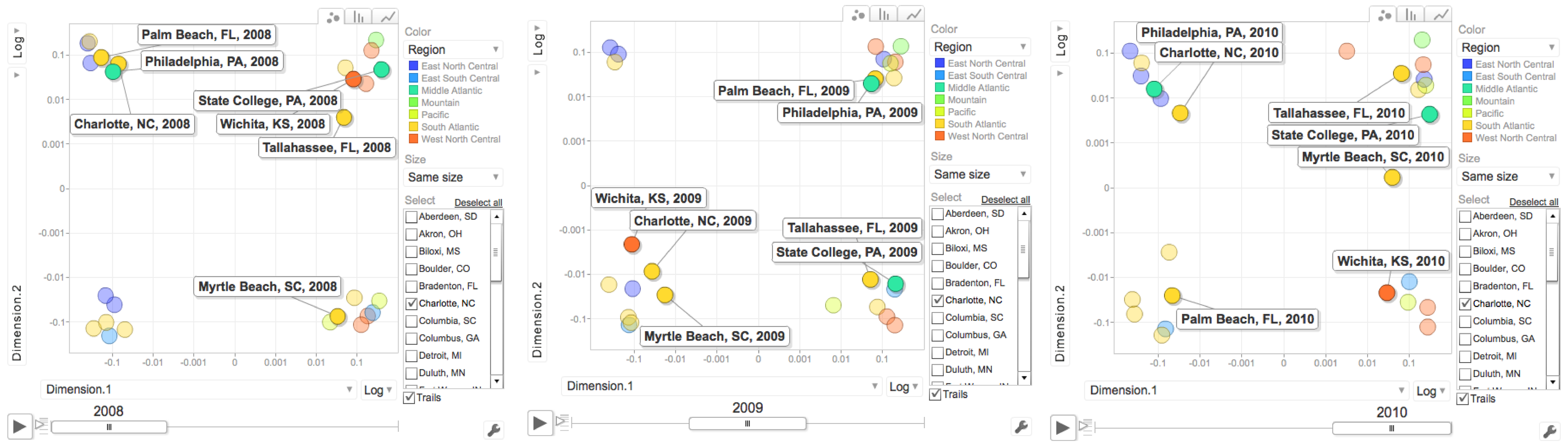


Figure 5: Standard Deviations

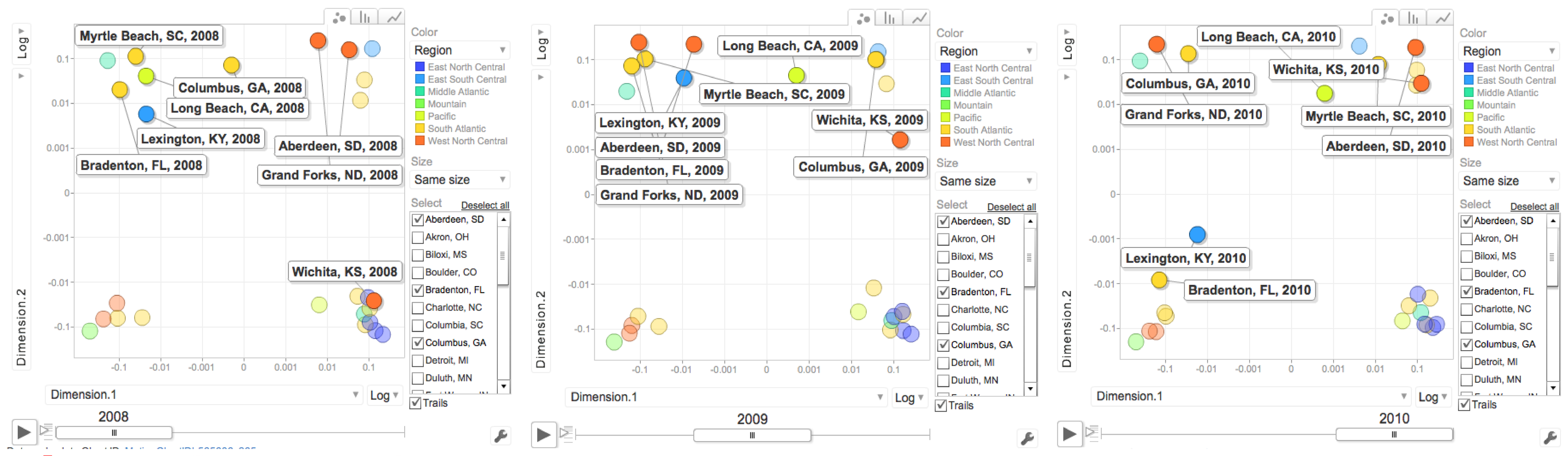


Figure 6: Proportions

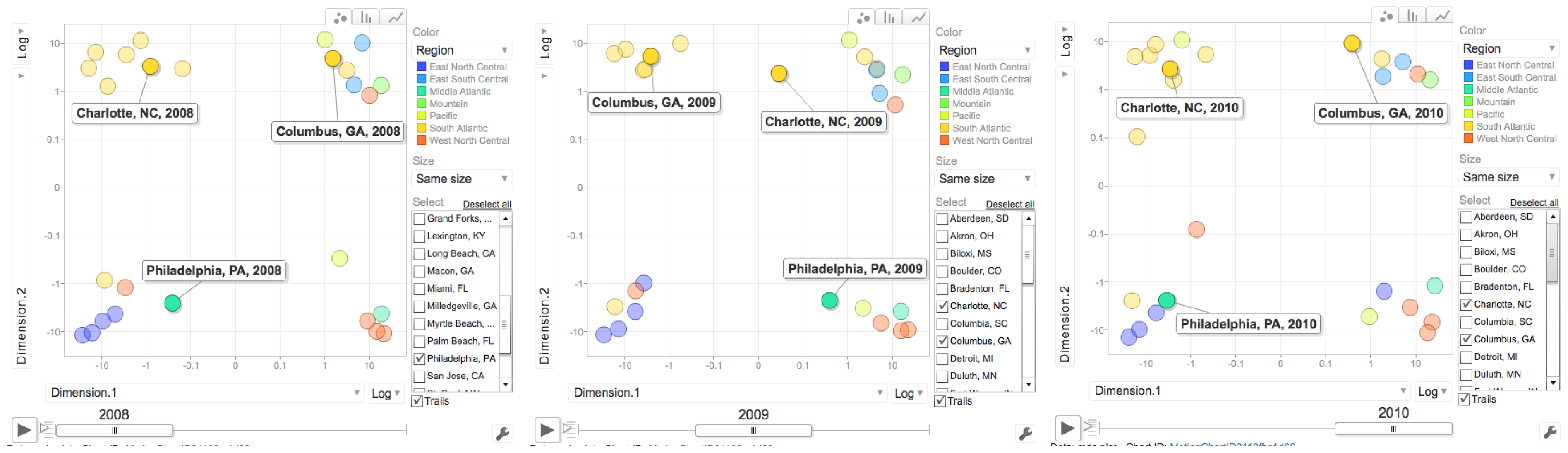


Figure 7: Z-Scores

The goal of Multidimensional Scaling is to provide a visual representation of the pattern of similarities and differences among the cities. We use the index variables to determine the relationships between the cities. Cities estimated to be very similar to each other in these characteristics are placed close to each other on the map, and those estimated to be very different from each other are placed far away from each other on the map.

These motion charts provide many different ways one can interpret the clusters and dimensions of the Multidimensional Scaling. In each figure, we can see distinct clusters of cities. We can group them by region or urbanicity to search for patterns in the clusters. Dynamic cites, which are those cities that move from cluster to cluster throughout the years, are marked on the charts.

Higher mean scores and proportion scores imply that the city scored higher across all index variables. A higher score in standard deviations implies that the responses for that city had more variation across the index variables, and higher z-scores indicate a higher city score relative to the original index variables.

III.B. Hierarchical Cluster Analysis

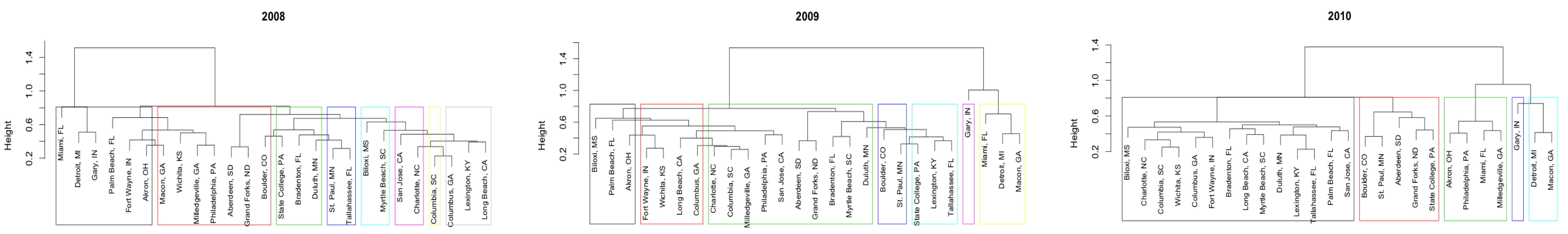


Figure 8

Another way we can observe the differences between the communities is to look at the results of average hierarchical cluster analysis. Figure 8 shows the dendrograms for each year, and the clusters of cities obtained by this method. Cutting each tree at 0.8, we can observe different numbers of clusters for each year, as well as different groupings of the cities throughout time.

IV. Conclusions and Future Research

We have demonstrated the use of motion charts in displaying the results of time-dependent multivariate analysis. Dynamic and interactive interpretations can be achieved and customized based on the interest of the user. Future research in this area will be to repeat the analyses based on subsets of the data by the suggested clusters to further understand the relationships between the index variables and cities, and to better characterize what attaches people to their communities.

References

- Markus Gesmann and Diego de Castillo. *Using the Google Visualisation API with R.* The R Journal, 3(2):40-44, December 2011.
- H. Wickham. *ggplot2: elegant graphics for data analysis.* Springer New York, 2009.
- Francois Husson, Julie Josse, Sebastien Le and Jeremy Mazet (2013). *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R.* R package version 1.24. http://CRAN.R-project.org/package=FactoMineR
- "What Attaches People to Their Communities? | Knight Soul of the Community." 03 June 2013 <http://www.soulofthecommunity.org/>.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.