Interactive Visualization of New Jersey Gang Data
Manfred MinimairSeton Hall University

by

Manfred Minimair

Abstract

This article describes the design and functionality of an online visualization software of data from a survey on gang activities in New Jersey municipalities. The visualization enables the user to explore the distribution of numbers of gang sets across different municipalities in New Jersey, and study certain derived information. The purpose of the visualization is to make data from the gang survey easily and universally accessible through some engaging visual display, to facilitate seamless exploration of the data, and to thus foster discourse on the data among experts and the general public. In order to achieve these goals, bubble charts, enhanced by various interactive devices, are proposed as the fundamental visualization technique used in the software design. The resulting interactive visualization software can be accessed on one of Seton Hall University's WEB servers and runs in the user's WEB browser.

Street gangs have been and are being intensely studied (Knox, 2006; Thrasher & Knox, 2000). Their criminal activities present unique challenges to society (Urrea, 2011). Some of the investigations of gang activities, such as (Brown & Benedict, 2009; Decker & Pyrooz, 2010; Esbensen, Deschenes, & Winfree, 1999; Rojek, Petrocelli, & Oberweis, 2010; Yearwood & Hayes, 2001), deal with large data collections which are available to researchers and, in various instances, are also available to the general public on the WEB, such as (U.S. Department of Justice, 2011; University of Maryland, 2011a). An abundance of available data on gangs and their activities naturally is of great importance to gang researchers and society at large because it may be a source of new insights for dealing with street gangs. But this abundance can also be overwhelming and does not automatically lead to significant new findings about gangs. Advanced computing tools can greatly help in studying

the data. In particular, interactive data visualizations (Card, Mackinlay, & Shneiderman, 1999; Few, 2004; Tufte, 2001) allow users to explore and query the data in uniquely engaging visual ways. These visual modes of investigation can assist in deepening our understanding of gang-related data. Appropriately designed visualizations available on the WEB also reach a wide audience, and therefore can foster the discourse among researchers and non-experts on important issues in gang criminality.

The objective of the current paper is to describe the design and the functionality of my interactive online visualization of some data from the 2010 New Jersey Street Gang Survey conducted by the New Jersey State Police (New Jersey State Police, 2011). In 2010 the analysts Dean Baratta and Peter Lynch of the New Jersey State Police made me aware of the survey and suggested that I develop some online visualization for the data. While designing the visualization in 2010, Mr. Baratta and Mr. Lynch gave some feedback on the survey and on usability of the software I developed, but they did not participate in the design process.

The visualization supports exploring the distribution of numbers of gang sets across different municipalities in New Jersey and studying certain derived information, such as rankings and relative geographic location. The purpose of the visualization is to make data from the gang survey easily and universally accessible through some engaging visual display, to facilitate seamless exploration of the data, and to thus foster discourse on the data among experts and the general public.

For the design, I propose bubble charts as the fundamental visualization technique in order to achieve these goals. The bubbles correspond to municipalities and their sizes to the number of gang sets. Some interactive features allow to arrange the bubbles in different views and to highlight certain properties, like ranking.

The resulting interactive visualization software can be accessed on one of Seton Hall University's WEB servers (Minimair, 2011). It runs via the Adobe Flash Player in the user's WEB browser and remotely accesses data stored on the university's server.

This paper consists of the following sections in this order: Section Gang Survey and Data introduces the data used in the visualization. Section Visualization Design and Implementation describes my design and implementation of the visualization. Section Related Works discusses how the current work is related to others. Section Discussion reviews some natural strengths and limitations of the current software and points to potential avenues to expand the current work in the future, followed by the Conclusion section.

Gang Survey and Data

This section describes the data that has been used for creating the visualization.

In 2010 the New Jersey State Police conducted a survey of all municipal law enforcement agencies on street gang presence and activities. The survey used two

different forms, namely, an initial survey and a gang-specific survey for each gang that has been reported as present in the municipality (New Jersey State Police, 2011; University of Maryland, 2011a). I used both surveys in order to extract certain data for the visualization.

From the following questions in the initial survey, I deduced the number of gang sets present in a municipality in 2010:

> 14. During the past 12 months, which Bloods gang sets (if any) were present in your jurisdiction? Please check all that apply.
>
> 15. During the past 12 months, which Crips gang sets (if any) were present in your jurisdiction? Please check all that apply.
>
> 16. During the past 12 months, which OTHER gang sets (if any) were present in your jurisdiction? If there are no other gangs in your jurisdiction, please check 'Other' and type 'None' in the box below.

Each question is accompanied with a list of names of gang sets that participants of the survey can check off. Furthermore, each question also gives an empty field where participants can enter additional names.

Furthermore, the following question in the gang-specific survey allowed me to find the percentage of gang sets that have resident members in the municipality:

> 10. During the previous 12 months, were Transient and/or Resident members of this gang PRESENT in your jurisdiction? (You can check more than one box).

The answer choices were: Transient members, Resident members, Neither, Don't know. In the visualization, I used the sum of the percentages of gang sets of which some members were resident and whose members' residency was unknown. This sum gives a reliable upper bound on the percentage of gang sets of which some members were resident.

The visualization also presents the number of gang sets relative to the population size of the municipality. I obtained the population size data from the WEB site of the U.S. Census Bureau (U.S. Census Bureau, 2011).

Furthermore, the visualization incorporates some data on the geographic locations and shapes of municipalities in New Jersey. This data has already been used in Tauberer (2007b) for the 2007 New Jersey Street Gang Survey and is also available on the WEB site of the U.S. Census Bureau. For more on why this data is used for the current visualization see Section Preliminaries on Bubble Charts.

Visualization Design and Implementation

This section starts with some preliminary information on bubble charts which are used in my visualization for the New Jersey street gang survey. Then it describes how my visualization software is designed and how users can interact with it. The section concludes with an overview of the implementation of the software.

Preliminaries on Bubble Charts

Bubble charts are a well-known visualization technique which over recent years have become increasingly available in business analytics software and on the Web (Gapminder Foundation, 2011; IBM Research, 2011). These charts show some bubbles, that is, disks, arranged in a plane. For examples of bubble charts see Figure 1 through Figure 5. The bubbles usually represent some group of entities that are of interest to the user of the visualization. In my visualization for the New Jersey street gang survey, the bubbles represent the municipalities of New Jersey. Usually the sizes of the bubbles correspond to some numeric quantity. For the gang survey, the sizes correspond to the number of gang sets encountered in the municipality represented by the bubble. The bubbles can be colored to convey additional information through varying colors. For the gang survey, colors represent different counties or also ranking. The locations of the bubbles in the plane can be used to represent additional information associated with each bubble. For example, in one viewing mode of my visualization, the bubbles are arranged in a spiral sorted by their sizes. In another viewing mode, the bubbles are arranged roughly corresponding to the geographic location of the corresponding municipality in the state of New Jersey. This geographic arrangement is called Dorling cartogram after the developer of an algorithm for creating such cartograms (Dorling, 1996). Professor Dorling's algorithm requires geographic locations and shapes of the municipalities. Therefore the location and shape data is needed by the visualization which has been mentioned in Section Gang Survey and Data.

Design

The visualization is designed to allow the user to answer a number of questions: What are municipalities with many gang sets? What is the ranking of municipalities and the geographic distribution across the State of New Jersey? How many gang sets were found in specific municipalities? How do specific municipalities compare to each other and to New Jersey municipalities at large? The subsequent paragraphs elaborate on how the visualization addresses these objectives.

Notice that the following paragraphs refer to some screen shots of the visualization that have been included in figures. Since Journal of Gang Research does not print figures in color, all screen shots have been converted to grayscale. While colors are necessary to adequately interpret the data represented in the screen shots, the grayscale is sufficient to convey the intentions of the design in this paper.

The visualization consists of a bubble chart that can be manipulated in various ways. Each bubble represents one municipality in New Jersey. Initially (Figure 1) all the bubbles have the same color (gray) and their sizes represent the

numbers of gang sets found in the corresponding municipalities. Also initially, the bubbles are sorted and arranged in a spiral with the largest bubble in the center. The area with the bubble chart incorporates a search box that allows locating individual municipalities. A slider and an input field allow magnifying the chart to desired levels. Furthermore, the user can select the area visible under magnification by dragging this area into focus. Above the bubble chart, there are various buttons that allow changing the data shown by the bubbles ("data" row of buttons), the arrangement of the bubbles ("view" row of buttons), and the coloring of the bubbles ("color" row of buttons).

The initial arrangement of the bubbles in a spiral with the largest bubble in the center naturally focuses the user's attention on the municipality with the largest number of gang sets, and on similarly sized ones grouped around it. More fine-grained information on the ranking of the municipalities can be added by coloring the bubbles. One button in the "color" row of buttons allows highlighting the top-10 bubbles (Figure 2.) Another button colors the bubbles according to their percentile rank.

When the user hovers with the mouse pointer over a bubble, the name of the municipality and the number of gang sets is displayed. Individual bubbles can be selected by clicking on them. Multiple ones can be selected simultaneously by holding the Ctrl-key and clicking, as items are commonly selected in computer desktop environments. Another click on the bubble, while holding the Ctrl-key, clears the selection. All selections can be cleared simultaneously by clicking on the background area of the bubble chart. The selections stay active as the bubble chart is manipulated.

The initial view with the largest bubbles in the center draws the attention of the user to these largest bubbles. Since some of the larger bubbles appear of similar size, it is hard to precisely rank all bubbles by visual means. Therefore the visualization allows rearranging the bubbles in a spiral where the bubbles are sorted from smallest to largest (sort button in the "view" row of buttons.) This arrangement neatly lines up the bubbles according to their sizes while losing the focus on the largest ones (Figure 3.)

For understanding the geographic distribution of the bubbles, the visualization offers a cartogram view (cartogram button in the "view" row of buttons) and a color scheme according to New Jersey counties (counties button in the "color" row of buttons.) The cartogram button arranges the bubble in a Dorling cartogram such that their locations roughly correspond to their relative geographic locations (Figure 4.) The counties button assigns a color to each bubble according to the county the corresponding municipality is located in. Since there are 21 counties in New Jersey and 21 distinct colors would overwhelm the user (Ware, 2008), I divided the counties into four groups, namely, North-East, North-West, South-East, South-West. Each group contains five to six counties. I assigned a color to each

fig 1

fig 2

fig 3

fig 4

group, namely, brown, orange, green, and respectively purple. Subsequently, I assigned a different shade of the color to each county in the corresponding group. This color assignment lets the user quickly identify in which part of the state a municipality is located. The different shades of brown, orange, green, and purple may also help a user who is familiar with New Jersey geography relate the municipality to its county.

When the visualization is started, the sizes of the bubbles correspond to the absolute number of gang sets found in the corresponding municipalities. It is also important to understand this number in relation to the number of inhabitants of the municipalities because considering the number relative to the population size may affect the ranking. Therefore the "data" row of button enables the user to switch between the absolute number of gang sets and the number of gang sets per 10,000 inhabitants. For example, East Orange has an absolute number of 57 gang sets. There are 8.75 gang sets per 10,000 inhabitants. When switching from the absolute to the relative number of gang sets, one can observe that some municipalities move down and some move up in the ranking. For example, East Orange moves down from the upper 5-percentile to the upper 10-percentile. Some other municipalities move up because they have a small population and a relatively large number of gang sets. This raises the issue whether the visualization fairly represents these communities. Being an extreme case, the municipality Walpack has been designated as an outlier. It is a small municipality of 39 people located in the Delaware Water Gap recreational area. One gang set of a motorcycle gang has been found in this area. Therefore Walpack's number of gang sets relative to the population is about 256.4 (=10000/39.) This is an exceptionally large number when compared to the other towns where the second largest one is 43.1 for Highstown. Other municipalities may also qualify as outliers depending on one's understanding of the specific presence of street gangs in these municipalities. But the current visualization only considers Walpack as an outlier.

Besides studying the number of gang sets relative to the population size, the visualization also enables the user studying it with respect to residency of gang members. That is, the visualization offers a heat map for the proportions of gang sets with resident members in the municipalities. This heat map can be activated with the percentage resident button in the "color" row of buttons. The more intense the red of a bubble is the higher is the percentage of gang sets that have resident members in the corresponding municipality (Figure 5.)

Individual municipalities can be found easily in the bubble chart either by hovering over bubbles with the mouse pointer and zooming or by locating them in the search box. The bubbles can be selected by holding the Ctrl-key and clicking on them, as it has been already mentioned earlier. The bubbles are also selected through locating them via the search box. The selection stays active while manipulating the bubble chart. Transitions between different views are animated, thus, helping the user to track how the locations of the selected bubbles change. This animated

selection facility supports comparing individual municipalities and seeing them in the context of New Jersey municipalities at large.

Implementation

My visualization software is implemented according to a standard client-server model. That is, the graphics and interactive functions of the visualization are generated by a program at the client computer in the user's Web browser. The data is kept on a server computer at Seton Hall University from which the client program downloads any data that is needed. The client program is initially loaded into the user's Web browser when first accessing the Web address of the visualization (Minimair, 2011).

I implemented the client program in the programming language ActionScript. Therefore the program runs in Adobe Flash Player on the client computer. In this implementation I used the visualization toolkit Flare (Heer, Card, & Landay, 2005; UC Berkeley Visualization Lab, 2011) which has also been used for the popular WEB site ManyEyes (IBM Research, 2011).

The server computer keeps the data in a Postgre database (PostgreSQL Global Development Group, 2011). The server makes the data available through the interface provided by the communication toolkit amfPHP (Silex Labs, 2011).

Related Works

This section discusses two related works and elaborates on the differences to my visualization software.

Sunlight Foundation

The New Jersey State Police also conducted a survey on street gangs in 2007. At an event sponsored by Sunlight Foundation, 1818 N Street NW, Suite 300, Washington, DC, 20036, a non-profit organization that supports government transparency, a team of volunteers created a WEB site in order to present the 2007 survey's data (Tauberer, 2007a). The WEB site (Tauberer, 2007b) offers two main entry points for exploring the survey data, namely, Municipalities and Gangs.

The Municipalities view consists of a heat map of New Jersey municipalities reflecting the number of gangs found in each municipality. This heat map is a geographic map of New Jersey, where the areas of the municipalities are colored in a red shade. The darker the red appears, the more gangs were found in the municipality. The only interactive features of the map are zooming and scrolling. Other statistical information is presented in textual form, such as number of homicides, gang-related incidents at schools, and the names of gangs found in each municipality.

The Gangs view shows a map where the areas of the municipalities are colored according to the gang or gang set which has been identified as most problematic in the corresponding municipality. Nine different colors corresponding to nine different gangs or gang sets are shown on the map. As with the Municipalities view, the interactive features of the map are limited to zooming and scrolling. The user can also click on the gang names in the legend, but not on the corresponding colors on the map, in order to focus on the corresponding gang. Subsequently, the map shows the municipalities that have reported the gang as present shaded in pink.

University of Maryland

A team of graduate students in software engineering from University of Maryland created a WEB site for the data from the 2010 street gang of the New Jersey State Police (University of Maryland, 2011a). The WEB site (University of Maryland, 2011b) offers two main entry points for exploring the survey data, namely, Crime and Gangs.

The Crime view allows selecting different types of crimes. Then the site shows a geographic map of New Jersey where the municipalities affected by this type of crime are colored in red. The map does not allow any interactive manipulation.

The Gangs view allows selecting a gang set. Then the site shows a map of New Jersey where the municipalities where the gang set was found are colored in red. The map does not allow any interactive manipulation. Other information about the gang set is presented in textual form, such as number of towns where active, approximate number of members, and types of violent crimes.

Differences of the Related Works to My Work

The Sunlight Foundation and University of Maryland sites primarily present data in textual form. Graphical visualizations and interactivity with the graphics are kept at a minimal level. The only graphical elements on both sites are geographic maps where areas of municipalities are colored according to some data. The Sunlight Foundation site contains some heat maps with elementary interactivity for zooming and scrolling. The maps of the University of Maryland site do not allow any direct manipulation by the user. Both sites seem to be designed to efficiently retrieve information about individual municipalities, gangs, or crimes. But they seem not to be designed for comparing different municipalities.

In contrast, my site (Minimair, 2011) primarily focuses on graphical representation of data and the textual information is kept at a minimal level. The visualization uses bubble charts rather than geographic maps. The site is highly interactive supporting exploration of the data. The user can search for municipalities via a search box and by scanning the bubble chart with the mouse

pointer. Furthermore, the user can switch to different views of the same data set allowing emphasizing different aspects of the data. The software carries out some of these switches of the views through animations in order to help the user to maintain a mental connection between the different views. A design goal of the visualization is comparing the number of gang sets found in different municipalities which does not seem to be the focus of the Sunlight Foundation and University of Maryland sites.

Discussion

This section discusses some strengths, design challenges, and limitations of my visualization.

Strengths

There are some strengths shared by interactive online visualizations in general, including mine. One of these strengths is that the data can be explored in an engaging visual way. Users may arrive at insights about the data much quicker than by studying tables of numbers. Another of these strengths is being easily accessible by anybody on the Web and thus potentially reaching a large audience.

In addition, a particular strength of the visualization designed by me is to support some natural and important ways of exploring the data. These include comparing municipalities and seeing them in relation to all New Jersey municipalities (see Section Design for more details.) In comparison, the designs of other related works do not address these questions as thoroughly (see Related Works.)

Challenges and Limitations

Certain challenges in designing the visualization come from the desire to represent the municipalities and their data in a fair way. Subsequently two such challenges are examined more closely and some limitations are pointed out.

First, a large number of gang sets in a municipality may naturally lead users to guess that travel to or living at this municipality is particularly dangerous. Therefore, the visualization could affect some municipalities' reputations if it gave some unexpected insights. This may even lead to economic impacts because people may favor or avoid the municipality because of its reputation. Recognizing this fairness issue, I designed the visualization to allow exploring the data from different perspectives. I incorporated different ways to consider the number of gang sets, that is, as an absolute number, relatively to the population size, and also with respect to the proportion of gang sets with resident members. Furthermore, I added the capability of excluding outliers from the visualization. After consulting the New Jersey State Police investigators who made me aware of the data, I decided to designate the municipality Walpack as an outlier. This

notion of outlier is clearly subjective. Therefore, it would be useful to allow the user to decide how outliers are filtered by the software. Currently, this desirable feature is not available.

Second, there are a large number of New Jersey municipalities of which the survey reports zero gang sets. The visualization does not show well how large this number is in relation to the number of municipalities of which the survey reports one or more gang sets. This is due to representing the municipalities with zero gang sets by very small bubbles. It would be useful to have another view of the data that shows that the municipalities with gang sets are a minority. Fur such a view it may be necessary to incorporate other visualization techniques such as bar charts.

Conclusion

The design and implementation of my visualization of data from the 2010 New Jersey Street Gang Survey by the New Jersey State Police has been described. The objective of this visualization is to make some of the data from the survey easily and universally accessible on the WEB through some engaging visual interface.

The current work can be extended in various ways. One future direction is expanding the visualization to cover more survey data, and to thus make more data available visually. In particular, it would be interesting to allow users to view information on the types and numbers of crimes committed by street gangs. Another future direction may be to customize the visualization to serve specific needs of gang researchers or the police force trying to decide where and how to focus law enforcement resources. Such customization would require analyzing the needs and the work flow of such practitioners.

References

Brown, B., & Benedict, W. R. (2009). Growing pains and fear of gangs: A case study of fear of gangs at school among Hispanic high school students. *Applied Psychology in Criminal Justice, 5*(2), 139-164.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization : using vision to think.* San Francisco, Calif.: Morgan Kaufmann Publishers.

Decker, S. H., & Pyrooz, D. C. (2010). On the validity and reliability of gang homicide: A comparison of disparate sources. *Homicide Studies: An Interdisciplinary & International Journal, 14*(4), 359-376. doi: 10.1177/1088767910385400

Dorling, D. (1996). *Area Cartograms: Their Use and Creation* (Vol. 59).

Esbensen, F.-A., Deschenes, E. P., & Winfree, L. T., Jr. (1999). Differences

between gang girls and gang boys: Results from a multisite survey. *Youth & Society, 31*(1), 27-53. doi: 10.1177/0044118x99031001002

Few, S. (2004). *Show me the numbers : designing tables and graphs to enlighten*. Oakland, Calif.: Analytics Press.

Gapminder Foundation. (2011). Gapminder Retrieved December 15, 2011, from http://www.gapminder.org/

Heer, J., Card, S. K., & Landay, J. A. (2005). *prefuse: a toolkit for interactive information visualization*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA.

IBM Research. (2011). ManyEyes Retrieved December 15, 2011, from http://www-958.ibm.com/software/data/cognos/manyeyes/

Knox, G. W. (2006). *An introduction to gangs* (6th ed.): New Chicago School Press.

Minimair, M. (2011). Visualization: 2010 New Jersey Gang Survey Retrieved December 1, 2011, from http://sciris.shu.edu/njgangsurvey2010/countgangsets.html

New Jersey State Police. (2011). New Jersey State Police Publish New Street Gang Survey. *2011 New Jersey State Police News Release* Retrieved November 30, 2011, from http://www.nj.gov/njsp/news/pr012811.html

PostgreSQL Global Development Group. (2011). PostgreSQL Retrieved December 15, 2011, from http://www.postgresql.org/

Rojek, J., Petrocelli, M., & Oberweis, T. (2010). Recent patterns in gang prevalence: A two state comparison. *Journal of Gang Research, 18*(1), 1-17.

Silex Labs. (2011). amfPHP Retrieved Decmeber 15, 2011, from http://www.silexlabs.org/amfphp/

Tauberer, J. (2007a). About the New Jersey Gang Survey Viewer Retrieved December 1, 2011, from http://njgangsurvey.civicimpulse.com/about

Tauberer, J. (2007b). New Jersey Gang Survey Viewer Retrieved December 1, 2011, from http://njgangsurvey.civicimpulse.com/

Thrasher, F. M., & Knox, G. W. (2000). *The gang : a study of 1,313 gangs in Chicago* (2000 ed.). Peotone, Ill.: New Chicago School Press.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, Conn.: Graphics Press.

U.S. Census Bureau. (2011). U.S. Census Bureau Retrieved December 1, 2011, from http://www.census.gov

U.S. Department of Justice. (2011). National Youth Gang Survey Retrieved November 30, 2011, from http://www.nationalgangcenter.gov/About/Surveys-and-Analyses

UC Berkeley Visualization Lab. (2011). flare Retrieved December 15, 2011, from

http://flare.prefuse.org/
University of Maryland. (2011a). Gangs in New Jersey: 2010 New Jersey State Police Survey Information & Links Retrieved November 30, 2011, from http://njgangsurvey.com/links.aspx
University of Maryland. (2011b). New Jersey State Police Gang Survey Retrieved November 30, 2011, from http://njgangsurvey.com/
Urrea, C. (2011). *Gangs: Society's problem or a problem caused by society?* (72), ProQuest Information & Learning, US.
Ware, C. (2008). *Visual thinking for design*: Morgan Kaufmann.
Yearwood, D. L., & Hayes, R. (2001). A statewide assessment of gangs in the public schools: Origins, membership and criminal activities. *Journal of Gang Research, 8*(4), 1-12.

About the Author

Manfred Minimair, Ph.D., is a faculty at Department of Mathematics and Computer Science of Seton Hall University, South Orange, New Jersey. He has spearheaded the development of interdisciplinary undergraduate and graduate certificate programs in data visualization and analysis at Seton Hall University. He has taught various courses in data visualization, data analysis and computer graphics development at his department. His research includes applications of data visualization and analysis in diverse areas.