Pexel and Heatmap Visual Analysis of Multidimensional Gun/Homicide Data

Scott D. Rothenberger*, John E. Wenskovitch Jr†, G. Elisabeta Marai†
*Department of Statistics, †Department of Computer Science
University of Pittsburgh

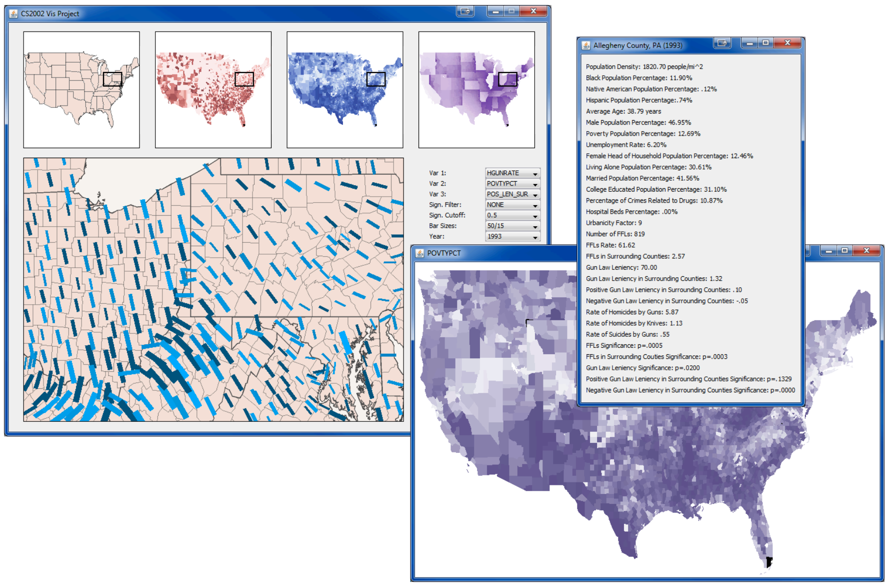

Figure 1: Interactive visual analysis of gun/homicide correlations, including a detailed view map with pexels and linked heatmaps, county-level details-on-demand, and an expanded heatmap popup.

Abstract

We present a visual analysis tool for mining correlations in county-level, multidimensional gun/homicide data. The tool uses 2D pexels, heatmaps, linked-views, dynamic queries and details-on-demand to analyze annual county-level data on firearm homicide rates and gun availability, as well as various socio-demographic measures. A statistical significance filter was implemented as a visual means to validate exploratory hypotheses. Results from expert evaluations indicate that our methods outperform typical graphical techniques used by statisticians, such as bar graphs, scatterplots and residual plots, to show spatial and temporal relationships. Our visualization has the potential to convey the impact of gun availability on firearm homicides to the public health arena and the general public.

1 Introduction

Homicide is one of the leading causes of death in the United States, and it is the leading cause of death for people between the ages of 15 and 34. The vast majority of homicides are committed with a firearm. Researchers in the public health domain, as well as in criminology and sociology, seek to analyze the relationships between any given U.S. county's gun homicides, the number of federal firearms licensees (FFLs), and gun laws of the surrounding region. An effective and interactive visualization of these relationships could also inform the general public of the severity of FFLs as a homicide risk factor that can be changed.

We obtained access to a confidential dataset compiled from various sources, including the National Center for Health Statistics, the ATF, the FBI, and the Census Bureau. The dataset contained annual county-level data for 3,142 U.S. counties from 1993-1999. Our focus was gun availability, homicide, and the multiple socio-demographic measures in the dataset. The relationships of interest in this dataset had originally been established on a national level [1]. However, the statistics researchers wished to perform county-level statistical analyses, as well as create a county-level informative visualization of the data and statistical findings. Prior visualizations available for this dataset were simple bar and line graphs [1] [2].

We developed a visualization system to display and mine multiple variables from the raw data as well as statistical findings on geographic maps. Our design employs heatmaps and 2D pexels in a linked-views design [3] [4]. Details-on-demand and dynamic queries facilitate the interactive exploration of the data. Our methods allow the users to make predictions based on visual inference and to validate such hypotheses by filtering data according to statistical significance (p-value based).

2 Methods

175 heatmaps were generated from the county-level data using SAS; the heatmaps were then stored in 700 different PNG files. 525 of the images are thumbnails, used as small multiples. The other 175 are larger files that can be explored in a separate pane.

{sdr49, jew51, marai}@pitt.edu

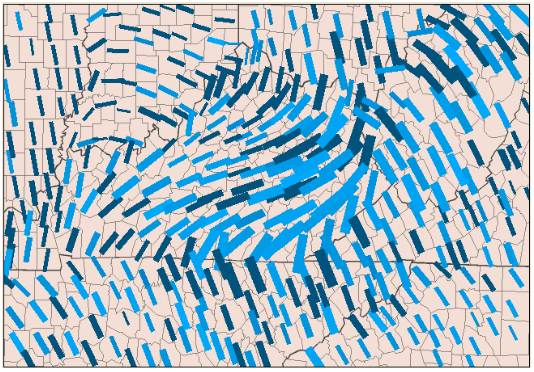

Figure 2. 2D Pexel mapping of Gun Homicide Rate (Value), Poverty Percent (Size), and Surrounding Leniency Laws (Orientation) – a distance-weighted measure of the difference in gun law leniency between a given county and surrounding counties. Kentucky 1993.

Details-on-demand are provided for more than 25 individual variables associated with that county for the selected year. The dropdown boxes allow for dynamic queries. The pexels vary in color (strictly by value), in size, and in orientation (Figure 2). Flexible attribute-mapping is an advantage of the system, as the effectiveness of a given display can be subjective and user-specific. With the addition of a time variable (the selected year), two spatial variables (the latitude and longitude of the county), and a significance variable (dropdown-menu selectable), seven variables of information can be displayed at a time. Furthermore, the maximum and minimum bar sizes can be adjusted to account for the variable sizes of counties in different areas of the country.

In order to implement a significance filter on the county-level, we developed our own statistical model. We fit a negative binomial generalized linear mixed model (GLMM) to the data, using the gun homicide rate as the response variable and five gun availability measures as covariates of interest, adjusting for 15 confounding variables. An AR(1) temporal correlation structure and a Gaussian spatial correlation structure for random county effects were employed. P-values were then obtained for the covariates of interest and implemented in our significance filter.

3 Evaluation and Results

Two professors and two Ph.D. students from the Department of Statistics provided expert feedback on the system. Exploratory tasks using the raw data were performed, and the significance filter was used as a means to validate hypotheses. Inferences were made about the association between homicide rate and FFLs based on visual inspection of the glyphs. After formulating a hypothesis, the p-value significance filter for the effect of FFLs was used at various cutoffs, allowing the user to see if the hypothesis was supported.

3 out of 4 users formed the correct hypothesis for question 1(a): County gun homicide rate tends to increase as number of FFLs increase, showing a positive association. They also gave virtually the same explanation: as glyph size increases (FFLs), the color value tends to increase as well. One user was unable to determine any association until seeing the significance filter implemented later on. However, all four users reached the correct hypothesis for 1(b): Homicide gun rate tends to increase as the urbanity of the county increases, showing a positive association. Their explanations were the same: The glyphs tend to have higher color value when they are more positively sloped in orientation (urban).

Overall, the expert users were impressed and surprised by the non-standard representations of statistical data. Feedback about the 2D pexel method revealed some interesting results. Three of the four users reported that orientation was by far the most helpful in detecting relationships. The effectiveness of value and size varied substantially depending on the task, the region shown, the user, and the variables chosen. One user commented that the glyph's color value was "not terribly discriminatory and just gave a sense of high and low values with no in-between." The univariate heatmaps were a great supplement to the 2D glyph view. As one user said, "It allows one to see the data from two different perspectives." A context-specific comment regarding orientation provided insight into the advantage of our detail view with respect to visualizing spatial relationships: "Using orientation to display surrounding gun leniency provides a clear picture of how firearms can flow across borders." A typical bar graph or scatterplot would not be able show such relationships.

The significance filter was a success for this project. Being statisticians, the expert users naturally understood its utility.

4 Discussion and Conclusion

Healey's work on preattentive perception, as well as related discussions on perceptual-cognitive issues, may help to explain why value and size were useful in some cases but uninformative in others. Orientation was most likely very helpful due to the inherent spatial relationships in this dataset; many of the detail-view displays have the appearance of vector fields or streamlines. Our 2D glyph technique was chosen for the sake of effectiveness and ease-of-use with a short learning curve, although the use of 3D pexels in future work may allow for clearer perception of significant relationships in this dataset.

Interestingly, for many cases in which the relationships in the raw data were not initially perceptible, turning the significance filter on and back off helped. Once the users saw what kinds of patterns in the glyph attributes were significant, they were able to detect such patterns more easily. It is worth mentioning that one expert user did not fully understand the dataset and relationships we intended to show. Yet, he was able to navigate the system, make hypotheses, and see the intended relationships – even though he did not realize what they meant.

In conclusion, our case study explores associations between gun availability and firearm homicide rates through combined visualization and statistical means. With further tuning, such an approach has the potential to convey the significance of these relationships to the public in an effective manner. The field of statistics would benefit greatly from learning about new visualization techniques, and this project (part of an interdisciplinary visualization course) is a step in that direction.

References

[1] D. Wiebe, R. Krafty, C. Koper, M. Nance, M. Elliot, C. Branas (2009). Homicide and Geographic Access to Gun Dealers in the United States, *BMC Public Health*, 9:199.

[2] M. Miller et al (2007). State-level homicide victimization rates in the US in relation to survey measures of household firearm ownership, 2001-2003, *Social Science & Medicine* 2007;64: 656-664

[3] Healey, C. G. and Enns, J. T. 1999. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Trans. Visual. Comput. Graph. 5,* 2, 145-167.

[4] Healey, C. G., Tateosian, L., Enns, J. T., and Remple, M. 2004. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Trans. Graph. 23*, 1, 64–96.