ABSTRACT

ROGERS, GEOFFREY J. Combining Glyph Based and Topographical Techniques to Visualize United States Congressional Earmarks. (Under the direction of Christopher G. Healey, Ph.D.)

Recent advances in computing technologies have created a vast quantity of unprocessed data, and visualizations assist in extracting knowledge from this ever increasing data repository. Current glyph based visualization techniques are effective at representing various multidimensional data sets; however, they abstract the data from their underlying context. Topographical map based techniques retain the underlying context of the data set, but are limited in how many attributes can be displayed at a given map location. This thesis presents a novel information visualization approach for displaying multidimensional data with underlying topographical context. The technique addresses three problem areas within visualization: 1) displaying multiple attributes for a region or location on a map, 2) allowing the viewer to make comparisons among neighboring data points, and 3) maintaining a relative encoding of data across regions that allows for direct comparison of data across multiple areas. The technique developed to meet these requirements involves a combination of glyph based and topographical visualization techniques. Starting with a topographical map of the United States as a foundation, glyphs are used to denote individual states and regions within states. The hue, saturation, luminance and height of the glyphs are varied to represent different attributes of the data set. Visualizations were generated which show U.S. congressional earmark requests in a single fiscal year as well as trends between fiscal years. This technique is not limited to U.S. political delineation, but can accommodate multidimensional data from any domain that has underlying geographical context.
Combining Glyph Based and Topographical Techniques to Visualize United States Congressional Earmarks

by
Geoffrey J. Rogers

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Science

Computer Science
Raleigh, North Carolina
2011

APPROVED BY:

__________________________  ___________________________
Robert St. Amant, Ph.D.     Benjamin Watson, Ph.D.

_________________________
Christopher G. Healey, Ph.D.
Chair of Advisory Committee
BIOGRAPHY

Geoffrey Rogers is originally from Cuyahoga Falls, Ohio. He graduated from Virginia Commonwealth University in 2008 with a Bachelor of Science in Computer Science and a minor in Mathematics. He began graduate studies at North Carolina State University in 2009 with research interests in computer graphics and artificial intelligence.
ACKNOWLEDGEMENTS

I would like to thank my committee chair, Christopher G. Healey, Ph.D., for his mentorship and for welcoming all questions and concerns. I also thank my committee members, Benjamin Watson, Ph.D. and Robert St. Amant, Ph.D. for their critical evaluation and feedback.

I would also like to thank the Center for Responsive Politics and Taxpayers for Common Sense for providing the earmark data necessary for this project. This service, as well as the many others they provide, are invaluable assets for public knowledge.

Finally, I am most grateful to my wife, Samantha Rogers, for her unwavering support, encouragement and motivation.
# TABLE OF CONTENTS

List of Tables ................................................................. v
List of Figures ................................................................. vi

Chapter 1 Introduction ....................................................... 1

Chapter 2 Congressional Earmark Spending ............................ 3

Chapter 3 Previous Visualization Techniques .......................... 7
  3.1 Charts and Tables .................................................... 7
  3.2 Glyph Based Techniques ........................................... 10
  3.3 Information Visualization Techniques ......................... 12
  3.4 Moving Forward ..................................................... 21

Chapter 4 Methods ........................................................... 22
  4.1 Data Collection and Processing ................................. 22
  4.2 Earmarks in a Single Fiscal Year ............................... 23
  4.3 Earmarks between Two Fiscal Years ......................... 26

Chapter 5 Results .............................................................. 27
  5.1 Earmarks in a Single Fiscal Year ............................... 27
  5.2 Earmarks between Two Fiscal Years ......................... 33

Chapter 6 Discussion and Future Work .................................. 36

References ........................................................................... 39
LIST OF TABLES

Table 4.1  Attributes calculated from the earmark data set. . . . . . . . . . . . . . . . 23
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Graph of the budgets for the 14 major departments funded through discretionary spending. [11]</td>
<td>4</td>
</tr>
<tr>
<td>3.1</td>
<td>Table from the Center for Responsive Politics reporting earmark dollar totals for Ohio’s congressional members during the 2010 fiscal year [7].</td>
<td>8</td>
</tr>
<tr>
<td>3.2</td>
<td>Bubble chart representing state total dollar earmark requests. Larger circles represent higher earmark dollar request. [22].</td>
<td>9</td>
</tr>
<tr>
<td>3.3</td>
<td>Illustration of the Ebbinghaus illusion. Both middle circles are the same size; however, the circle surrounded by smaller circles (right) appears larger.</td>
<td>9</td>
</tr>
<tr>
<td>3.4</td>
<td>Set of Chernoff faces. Each face represents a different data point, with facial features encoding a specific attribute from the data set. [8].</td>
<td>11</td>
</tr>
<tr>
<td>3.5</td>
<td>Glyph based visualization of weather conditions across the United States. Angle of glyph represents amount of rain, size represents pressure, density represents wind strength, and color represents temperature. [13, 18].</td>
<td>11</td>
</tr>
<tr>
<td>3.6</td>
<td>Treemap representing the 2010 congressional earmark requests by state. State size represents total dollar amount of earmark requests for the state. Saturation of the state denotes the per-capita dollar amount requested by the state [21].</td>
<td>13</td>
</tr>
<tr>
<td>3.7</td>
<td>An example of multiple views using electronic geographical maps. Each image portrays a different level of detail of the target location [15].</td>
<td>15</td>
</tr>
<tr>
<td>3.8</td>
<td>Demonstration of the fisheye technique using three different sized “lens” areas [38].</td>
<td>16</td>
</tr>
<tr>
<td>3.9</td>
<td>Map of the world using an area cartogram with country area resized to represent 2008 total gross domestic product [24].</td>
<td>17</td>
</tr>
<tr>
<td>3.10</td>
<td>Choropleth maps used by the Center for Responsive Politics to represent total and per-capita earmark requests in each state during the 2010 congressional fiscal year [6].</td>
<td>18</td>
</tr>
<tr>
<td>3.11</td>
<td>Map showing number of incarcerated individuals in the United States using proportional symbol mapping. A larger disc above the state indicates a larger number of incarcerated individuals [14].</td>
<td>19</td>
</tr>
<tr>
<td>3.12</td>
<td>Visualization of multidimensional data using data-driven spots. Each attribute is represented using a specific size and color spot, which are then merged together to create the final map [28].</td>
<td>20</td>
</tr>
<tr>
<td>3.13</td>
<td>Illustration of attribute blocks applied to a section of North America. Each of the four colors represents a different attribute from the data set. Blue represents potential evaporation, green represents soil moisture-holding capacity, red represents temperature and purple represents precipitation. Higher saturation indicates a higher observed value. [23].</td>
<td>20</td>
</tr>
<tr>
<td>4.1</td>
<td>The visual features used to encode attributes from the U.S. congressional earmark request data set.</td>
<td>25</td>
</tr>
</tbody>
</table>
Figure 5.1 Visualization of Louisiana’s 2010 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over the state shows Senate per-capita earmark requests.  

Figure 5.2 Visualization of East North Central States’ 2010 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over each state shows Senate per-capita earmark requests. State height reflects total earmark requests for the state, not normalized for state population.  

Figure 5.3 Visualization of United States 2009 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over each state shows Senate per-capita earmark requests. State height reflects total earmark requests for the state, not normalized for state population.  

Figure 5.4 Visualization of United States 2010 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over each state shows Senate per-capita earmark requests. State height reflects total earmark requests for the state, not normalized for state population.  

Figure 5.5 Visualization of the change in congressional earmark requests between the 2009 and 2010 fiscal years for districts in Ohio. Color denotes the political party that experienced an increase the proportion of earmark money requested by that district. Saturation denotes the difference in per-capita earmarks between the two fiscal years. A textured ‘x’ indicates that there was an increase in earmark requests between the two years; absence of this texture indicates a decrease.  

Figure 5.6 Visualization of the change in congressional earmark requests between the 2009 and 2010 fiscal years for the United States. Color denotes the political party that experienced an increase the proportion of earmark money requested by that district. Saturation denotes the difference in per-capita earmarks between the two fiscal years. A textured ‘x’ indicates that there was an increase in earmark requests between the two years; absence of this texture indicates a decrease.
Chapter 1

Introduction

Recent advances in computing technologies have created a vast quantity of unprocessed data, and visualizations assist in extracting knowledge from this ever increasing data repository. Visualizations allow the viewer to gain insight into a data set either by answering specific questions or by identifying facts and correlations which were previously unknown [29]. Visualizations are helpful because they reduce the cognitive effort required to gain insight into a data set, allowing for rapid information processing. Visual properties capable of being analyzed at first glance include color, saturation, size and orientation [17]. By representing multidimensional data in these forms, visualizations are able to greatly condense the quantity of data presented simultaneously.

Multidimensional visualizations work by mapping a data set to an image that is presented to the viewer. Formally, a data set $D$ contains $n$ data elements, $e_i$, such that $D = \{e_1, e_2, ..., e_n\}$. $D$ represents the set $A$ which contains $m$ attributes, $A_j$, present in the data set, such that $A = \{A_1, A_2, ..., A_m\}$, where $m > 1$. Each element $e_i \in D$ encodes values for all $m$ attributes in $A$; therefore, $e_i = \{a_{i,1}, a_{i,2}, ..., a_{i,m}\}$, where $a_{i,j} \in A_j$. The image used to visualize the data set $D$ applies a set $V$ of $m$ visual features, $V = \{V_1, V_2, ..., V_m\}$. The mapping $M(V, \phi)$ defines how to encode each attribute $A_j \in A$ using a visual property $V_j \in V$, where $\phi_j$ identifies the specific mapping from $A_j \rightarrow V_j$ [16]. For example, if a data set has elements with three attributes
(e.g., temperature, wind speed, humidity), each of these attributes would be mapped to a different visual feature in the image (e.g., color, height, texture). An effective visualization uses a mapping that allows the viewer to rapidly analyze the data and draw accurate conclusions.

Visualization is typically divided into two categories: scientific visualization and information visualization. Scientific visualization involves the representation of physical objects or data recorded from scientific experiments. This category commonly includes meteorological, chemical, biological and geographical data. Scientific visualizations typically take the form of volume and surface renderings, which accurately reflect the physical item being measured. On the other hand, information visualization is used to represent abstract data, such as text, images, database records and qualitative information [29].

This thesis presents a novel information visualization approach for displaying multidimensional data with underlying topographical context. While I utilize congressional earmark spending data to present the technique, the visualization developed is not domain restricted. The contributions of my visualization technique are to:

- display multiple attributes for a region or location on a map,
- allow the viewer to make comparisons among neighboring data points,
- maintain a relative encoding of data across regions that allows for direct comparison of data across multiple areas.

The technique developed to meet these requirements involves visualization of multidimensional data on a topographical map of the United States using glyphs. The hue, saturation, luminance and height of individual states and districts are varied to represent different attributes of the data set.
Chapter 2

Congressional Earmark Spending

The United States Congress, the legislative branch of the United States government, is divided into the Senate and House of Representatives. The United States Senate is the upper house of Congress and is comprised of 100 elected members with six-year terms. Each of the 50 states has two Senate members, giving all states equal representation. State Senators represent the entire state from which they are elected [33]. The United States House of Representatives is the lower house of Congress and is comprised of 435 elected members with two-year terms [34, 36]. Representation in the House of Representatives varies based on state population, with each state guaranteed at least one member. States are divided into districts, with the number of districts in a state equal to its number of members in the House. Members in the House of Representatives represent only the district from which they are elected [34].

One responsibility of the United States Congress is to create and pass an annual budget for the United States federal government. Money in the United States budget is divided into two categories, non-discretionary and discretionary. Non-discretionary spending, also known as mandatory spending, is spending that is required by law. Programs covered by non-discretionary spending include Medicare, Medicaid and Social Security. These programs are commonly referred to as “entitlements,” as any individual meeting the requirements of the program is eligible, or entitled, to receive its benefits. Discretionary spending is spending that
is not required by law, and is left to the discretion of Congress. Discretionary spending is currently determined through twelve appropriations bills. The amount spent in each of these bills varies annually, and is determined by the subcommittee in which the bill originates [2, 10]. Programs funded through discretionary spending include Health and Human Services, Education and Defense (Figure 2.1; [11]). Non-discretionary and discretionary spending for the 2009 fiscal year totaled 2.1 trillion and 1.2 trillion dollars, respectively [12].

![Figure 2.1: Graph of the budgets for the 14 major departments funded through discretionary spending. [11].](image)

Congressional earmarks are also assigned under discretionary spending. An earmark is a congressional appropriation of funds for use on a specific project. Earmarks are requested by individual members of Congress and are typically included in appropriations bills, although they may exist in any piece of legislation. Earmarks are assigned from existing agency budgets, not in addition to. These projects or programs normally benefit the congressional member’s home state or district. The exact dollar amount spent on congressional earmarks varies annually, with fiscal year 2010 earmarks totaling 15.9 billion dollars [30].
Recent legislation has required public disclosure of all approved earmarks from members of Congress [26], which has created a vast amount of new, unorganized data. This new legislation intended to provide greater transparency in the discretionary spending of taxpayer money on earmarks. Earmarks are not distributed evenly across states and districts, with some areas receiving larger sums of money than others. Public knowledge regarding where earmark money is spent could increase the accountability of congressional members to their constituents. In order for the newly available earmark data to be useful, techniques must be developed that allow the viewer to rapidly and accurately extract information and draw conclusions from the data set. Some relationships of interest in the data include: 1) which states receive the most earmark funds; 2) whether or not a state’s funding amounts correspond to its population; and 3) the distribution of earmark requests across political parties. Given the interest in earmark requests, I developed a visualization technique that:

- indicates the political party of the congressional member requesting the majority of a district’s earmark funds,
- denotes when earmark requests for a given district are split relatively equally across political parties,
- shows which states received the most congressional earmark dollars in absolute dollar amounts, independent of state population,
- shows which districts received the most congressional earmarks in per-capita dollar amounts,
- shows the relative distribution of earmark requests between a state’s Senators,
- maintains a relative encoding of data across states that allows for direct comparison of data.

In addition, my technique is capable of representing changes in congressional earmark requests between two fiscal years. This second visualization shows how earmark requests trended between the two years and can:
• identify the political party that experienced an increase in the proportion of earmark money requested in a district,

• depict the difference in per-capita earmark dollars requested between the two fiscal years,

• indicate if a district had an increase or decrease in per-capita earmark dollars between the two fiscal years.
Chapter 3

Previous Visualization Techniques

3.1 Charts and Tables

Multidimensional data are commonly presented through the use of tables. In fact, the data used to generate the images in this thesis were provided in tabular form. Figure 3.1 shows a table reporting earmark requests for each congressional district in the state of Ohio [7]. While tables can effectively summarize and present multidimensional data, they are not conducive to extracting relationships among variables in the data set. A table showing all pair-wise differences between attributes in even a small data set would quickly grow to an unmanageable size. Although data presented in this fashion contain sufficient information to draw relationships, the time and memory requirements placed on the viewer are too high to allow for rapid information gain.
Bubble charts may be used to visualize data by representing each data element as a circle, where circle size corresponds to the attribute value at the represented data point. This technique has been applied to earmark data by representing state total earmark values with a circle for each state (Figure 3.2; [22]). Bubble charts allow for comparison of a single attribute among data elements; multiple attributes are not represented. Also, some circles are too small to allow for identification of the data element they represent, resulting in the inability to identify all elements’ values. Bubble charts place data circles randomly; therefore, no spatial neighborhood information is displayed. Data sets visualized with bubble charts may suffer bias from the Ebbinghaus illusion [27]. Two circles of equal size are placed on a map, with one circle surrounded by larger circles and the second circle surrounded by smaller circles. Although the two center circles are identical, the one that is surrounded by larger circles appears smaller (Figure 3.3). When bubble charts are used to represent data, viewers may incorrectly determine the value denoted by the circle due to the size of its neighboring circles.

Figure 3.1: Table from the Center for Responsive Politics reporting earmark dollar totals for Ohio’s congressional members during the 2010 fiscal year [7].

<table>
<thead>
<tr>
<th>Representative</th>
<th>State</th>
<th>Earmark Value (in $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercy Kaptur (D-Ohio)</td>
<td>OH</td>
<td>$71,301,300</td>
</tr>
<tr>
<td>Tim Ryan (D-Ohio)</td>
<td>OH</td>
<td>$35,400,000</td>
</tr>
<tr>
<td>Michael R. Turner (R-Ohio)</td>
<td>OH</td>
<td>$28,113,200</td>
</tr>
<tr>
<td>Steven C. LaTourette (R-Ohio)</td>
<td>OH</td>
<td>$27,277,000</td>
</tr>
<tr>
<td>Betty Sue Sutter (D-Ohio)</td>
<td>OH</td>
<td>$24,672,700</td>
</tr>
<tr>
<td>Dennis J. Kucinich (D-Ohio)</td>
<td>OH</td>
<td>$20,533,000</td>
</tr>
<tr>
<td>Mary Jo Kilroy (D-Ohio)</td>
<td>OH</td>
<td>$18,215,500</td>
</tr>
<tr>
<td>Zachary T. Space (D-Ohio)</td>
<td>OH</td>
<td>$13,725,000</td>
</tr>
<tr>
<td>Steve Driehaus (D-Ohio)</td>
<td>OH</td>
<td>$10,937,300</td>
</tr>
<tr>
<td>Marcia L. Fudge (D-Ohio)</td>
<td>OH</td>
<td>$10,000,000</td>
</tr>
<tr>
<td>Charlie Wilson (D-Ohio)</td>
<td>OH</td>
<td>$9,998,000</td>
</tr>
<tr>
<td>John A. Boccieri (D-Ohio)</td>
<td>OH</td>
<td>$7,607,000</td>
</tr>
<tr>
<td>Robert E. Lattea (R-Ohio)</td>
<td>OH</td>
<td>$4,990,000</td>
</tr>
<tr>
<td>Jean Schmidt (R-Ohio)</td>
<td>OH</td>
<td>$3,274,000</td>
</tr>
<tr>
<td>James D. Jordan (R-Ohio)</td>
<td>OH</td>
<td>$2,400,000</td>
</tr>
</tbody>
</table>
Figure 3.2: Bubble chart representing state total dollar earmark requests. Larger circles represent higher earmark dollar request. [22].

Figure 3.3: Illustration of the Ebbinghaus illusion. Both middle circles are the same size; however, the circle surrounded by smaller circles (right) appears larger.
3.2 Glyph Based Techniques

Glyphs are symbols or icons used to visualize one or more attributes in a data set. The visual appearance of the glyph is changed according to the values and elements it represents. Glyph shape, size, color, height, location and orientation are commonly varied in glyph based visualizations [3, 37].

Star glyphs can be used to visualize multidimensional data. In a star glyph, each dimension of the data set is represented using a ray extending from the star’s center. Each star represents a different data point or group of data points with the length of each ray being proportional to the value of the dimension it represents [19]. Star glyphs are useful in providing clear groupings of different data points while showing which data points do not fit within any cluster; however, star glyphs do not readily allow for comparison of attributes among data points.

Chernoff faces are a more advanced form of star glyphs which use a glyph of a human face to represent each data point. Each facial feature encodes a different attribute observed at the data point [8]. A group of Chernoff faces is then presented, allowing the viewer to make comparisons between elements in the data set (Figure 3.4). Chernoff faces are advantageous because they allow for the rapid comparison among data points, as points with similar values for a particular attribute will have similar facial features. Chernoff faces are limited by the number of faces that can be presented simultaneously. Additionally, any underlying geographic context from the data set is not included.

When multiple characteristics are varied on a glyph, extensive amounts of data may be contained within a single visualization. Weather conditions for the contiguous United States have been visualized using a combination of multiple glyph techniques (Figure 3.5, [13, 18]). In this example, each glyph has been designed to represent a painted brush stroke. Individual glyphs represent the data for the area they cover; however, visual relationships between different glyphs allow the viewer to draw regional conclusions from the data set. This technique allows for some geographical context, as the glyphs appear similar to the geography they represent when viewed together.
Figure 3.4: Set of Chernoff faces. Each face represents a different data point, with facial features encoding a specific attribute from the data set. [8].

Figure 3.5: Glyph based visualization of weather conditions across the United States. Angle of glyph represents amount of rain, size represents pressure, density represents wind strength, and color represents temperature. [13, 18].
3.3 Information Visualization Techniques

Treemaps are an alternative approach for visualizing multidimensional data that are particularly useful for hierarchical (tree-structured) data sets. With this technique, a main rectangle box represents the root of the tree. From there, each child of the root is depicted as a rectangle inside the main box. Each interior rectangle is subdivided recursively until all children are represented [29]. Treemaps have been used to visualize earmark data (Figure 3.6; [21]). In this treemap, the main rectangle represents all earmark dollars requested by members of Congress. The main rectangle is divided to represent individual states, with rectangle size representing the state’s total earmark requests (i.e., larger rectangles indicate more total earmark dollars). The saturation of the state rectangles shows the per-capita dollars requested in the state (i.e., higher saturation reflects more per-capita earmark dollars). This example does not include district-level information, which could be added by further subdividing each state rectangle into its congressional districts. Although this technique shows both total and per-capita earmark requests, the underlying geographical context is lost, making it difficult to identify regional patterns.
Figure 3.6: Treemap representing the 2010 congressional earmark requests by state. State size represents total dollar amount of earmark requests for the state. Saturation of the state denotes the per-capita dollar amount requested by the state [21].
Multiple views using electronic geographical maps are commonly employed for the visualization of multidimensional data using maps. This technique works by creating multiple renderings of an area using different levels of detail. Each level of the visualization contains relevant information for the given level and scope of the map. Figure 3.7 shows a multiple views map using three different levels of detail. Figure 3.7(a) has the largest spatial extent, with the entire continental United States visible. Figure 3.7(b) begins to move closer to the target region, and more information about the major roads and nearby cities becomes visible. Figure 3.7(c) is the closest to the target region. In this view, very detailed information about the target area (e.g., street and building names) can be obtained. Higher levels of detail reduce the viewer’s ability to draw regional relationships. Therefore, this technique is most useful when information about the context of the target area is not necessary.

An alternative to multiple views is a fisheye transformation [38]. Only one map view is presented with this technique, shown at a predetermined level of detail. Detailed information about specific regions of the map is obtained through the use of a fisheye lens that is moved over the map by the viewer. When placed over an area of interest, the lens bubbles up the selected region and displays more detailed information, while maintaining the context of the region relative to the rest of the map. The level of detail can be varied, with higher levels resulting in more pronounced distortions of the map surface. The size of the distortion area can be changed to best fit the size of the map and the level of detail required. Figure 3.8 shows a map of a city in Japan with three different sized distortion areas. Figure 3.8(a) has a distortion area that is too small to glean any new information from the map. Figure 3.8(b) has a distortion area that is too large for the map and removes most of the contextual information. Figure 3.8(c) has the best sized distortion for the given map because it provides a good balance of detail and context. Although this technique maintains some contextual information about the target location, relative distances between points as well as information regarding direction between points are lost [4].
Figure 3.7: An example of multiple views using electronic geographical maps. Each image portrays a different level of detail of the target location [15]
Another technique for the visualization of data on maps is an area cartogram. With this technique, locations with higher values for the target attribute have greater area when displayed on the map. Figure 3.9 shows a map of the world which has been distorted to represent gross domestic product (GDP); countries with higher GDP have larger areas than countries with lower GDP. Although an area cartogram could be used in the visualization of congressional earmarks, with area showing the amount of earmark dollars received by each state, several problems exist with this approach. Area cartograms can create severe distortion, making it difficult to locate individual items on the map. Depending on the values associated with each state, some states may be impossible to locate on the warped map. This problem can be seen in
Figure 3.9 where some Eastern European countries are pinched from view. Lastly, a projection of this type would be limited to showing only one attribute.

![Figure 3.9: Map of the world using an area cartogram with country area resized to represent 2008 total gross domestic product [24].](image)

Two techniques that would preserve state and district shape and area are choropleth maps and proportional symbol mapping; however, these maps also traditionally depict a single attribute [27]. A choropleth map has shaded or colored areas showing the value of the displayed attribute (Figure 3.10). Proportional symbol mapping places a symbol on top of each data point, with the size of the symbol indicating the value observed at the point. For example, the circle on top of each state in Figure 3.11 indicates the total number of incarcerated individuals in the state, with large circles corresponding to a larger number of individuals. Proportional symbol maps also suffer bias from the Ebbinghaus illusion.

In order to display multiple attributes with these techniques, a series of maps can be presented, with each map representing a different attribute from the data set [27]. The Center for Responsive Politics utilizes this approach by presenting total and per-capita earmark requests.
on separate choropleth maps (Figure 3.10; [6]). While this technique provides a general picture of earmark requests aggregated at the state level, it does not allow for detailed district level analysis or provide insight to Senate earmark contributions independent from districts. Additionally, presenting a series of maps is less effective than providing a single map that displays all attributes [25].

![Choropleth maps](image)

(a) Color denotes total earmark amounts

(b) Color denotes per-capita earmark amounts

Figure 3.10: Choropleth maps used by the Center for Responsive Politics to represent total and per-capita earmark requests in each state during the 2010 congressional fiscal year [6].
Unlike choropleth and proportional symbol maps, data-driven spots are capable of visualizing multiple attributes in a single map. Data-driven spots work by presenting multiple layers on the same map, with each layer representing a different attribute from the data set (Figure 3.12). Each layer is assigned a spot with a specific size and color for the attribute it represents. The layers are then merged together, with some layers more visible than others at a given location on the map [1]. This technique would not work well for data sets with high regional variability. If layers with important information were omitted in key map locations, the generated map would have an incomplete representation of the underlying data (i.e., regional variability would be lost).

Alternatively, attribute blocks, which are commonly used to display multiple variables over a geographical region, would not under-represent variability as seen with data-driven spots. In
Figure 3.12: Visualization of multidimensional data using data-driven spots. Each attribute is represented using a specific size and color spot, which are then merged together to create the final map [28].

This technique, a block is divided into a number of sections equal to the number of variables being presented. Each section is colored to represent a different attribute from the data set. This attribute block is then tiled across the entire map, with each block showing all attributes for the region it covers (Figure 3.13; [23]). While attribute blocks effectively show trends over a region, they do not allow for detailed analysis of smaller map segments.

Figure 3.13: Illustration of attribute blocks applied to a section of North America. Each of the four colors represents a different attribute from the data set. Blue represents potential evaporation, green represents soil moisture-holding capacity, red represents temperature and purple represents precipitation. Higher saturation indicates a higher observed value. [23].
3.4 Moving Forward

All of these approaches have advantages and disadvantages. While the glyph based techniques are effective at representing various data sets, they abstract the data from their underlying context. These techniques, when applied to data with a topographical context, would require additional enhancements to represent information about the data’s physical location with respect to the underlying geography as well as other data points. However, glyph based techniques allow for the representation of multiple attributes at a single location. Topographical map based techniques retain the underlying context of the data set, but are limited in how many attributes can be displayed at a given map location. An effective solution for the visualization of multi-dimensional data with an underlying topographical context will involve combining characteristics from both types of visualizations.
Chapter 4

Methods

4.1 Data Collection and Processing

Obtaining earmark requests for each member of Congress was not straightforward. Although legislation requires all members of Congress to report their earmark requests, the law does not stipulate where the reports must be published [20]. Originally, earmark disclosures were reported only in print form and were held at the Library of Congress. Later legislation required the online reporting of earmarks; however, no single outlet was provided. Consequently, congressional earmark requests are scattered across approximately 550 different websites.

The effort required for data collection was greatly reduced by the Center for Responsive Politics (CFRP), which has compiled all earmark requests into a central database [7]. For each state, data reported on the CFRP website include the congressional members’ names and their earmark request totals. In order to incorporate district level visualizations, data from House members also needed to be associated with the district number within the representative’s home state. To achieve this, a set of scripts was developed that matched congressional members to their district number. Data for linking a member of Congress to their home district was acquired through the United States House of Representatives website [35].

A fiscal year spanning across multiple sessions of Congress further complicated data col-
lection. During election years, a district’s House representative may switch mid-fiscal year, sometimes involving a change in political party. To accommodate these situations, my technique sums fiscal year earmark requests from both district representatives to determine the total earmark requests of the district. In cases where district members differ by political party, the political party associated with each set of earmark requests is maintained such that the final color of the district represents the majority earmark contributor.

Given the starting data set, and my specific objectives, a total of ten attributes were calculated from the earmark data set. This included six state level attributes and four district level attributes (Table 4.1).

<table>
<thead>
<tr>
<th>State Attributes</th>
<th>District Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total earmark dollars requested</td>
<td>Per-capita earmark dollars requested</td>
</tr>
<tr>
<td>Percentage of Republican requests</td>
<td>Percentage of Republican requests</td>
</tr>
<tr>
<td>Percentage of Democratic requests</td>
<td>Percentage of Democratic requests</td>
</tr>
<tr>
<td>Percentage of Independent requests</td>
<td>Percentage of Independent requests</td>
</tr>
<tr>
<td>Per-capita requests for each Senator</td>
<td></td>
</tr>
<tr>
<td>Sum of per-capita requests for both Senators</td>
<td></td>
</tr>
</tbody>
</table>

### 4.2 Earmarks in a Single Fiscal Year

The visualization of congressional earmark requests uses a topographical map of the United States as its foundation. The map intuitively depicts where earmark requests originated (e.g., requests from North Carolina congressional members are represented on the state of North Carolina). This design removes the need for additional encoding to capture geographic information. However, insufficient information is provided at the state level to draw detailed conclusions from the data set.
To allow for more detailed analysis, each state is divided into its House districts. Each district is treated as a glyph and is assigned a specific color and saturation to represent earmark requests for the district. District color is based on the party of the majority earmark contributor for the given district. I used blue for Democrats, red for Republicans and green for Independents. These color assignments are commonly associated with the major political parties in U.S. public media [9], which helps reduce the amount of time necessary to understand the map encoding. A district may have a different color than the political party of the district’s House representative. This can happen in two situations: 1) when Senate earmark requests outweigh the House member requests, and 2) when the district switched parties during the most recent election cycle. In the latter case, the opposite color is normally due to the 2010 fiscal year earmarks for the district being primarily requested by the former House member. District saturation is proportional to the per-capita earmark dollars requested in the district; a highly saturated district had more per-capita dollars requested by the district’s congressional members. Highly saturated districts have darker, stronger colors while lightly saturated districts have paler, muted colors (Figure 4.1).

The amount of earmark dollars requested by a district $d_i$ is calculated by summing the earmarks requested by its House member $h_i$ and a percentage of the state’s Senators’ earmark requests $s$. This additional Senate amount is proportional to the number of districts in the state. If a state has $n$ districts, then each district would have $\frac{1}{n}$ of the Senate earmark requests added to the district total. Using a proportion of the Senators’ earmark requests is necessary because Senate earmarks are not tagged to a particular district. This proportion assumes that each Senator represents and benefits all of the state’s districts equally. Formally, the per-capita earmark total for a congressional district $d_i$ is calculated as:

$$d_i = \frac{h_i + \frac{s}{n}}{p_i}$$

where $p_i$ is the population of $d_i$ and is equal to state population divided by $n$, as each state district is supposed to contain an equal percentage of the state’s population [32].
The coloring scheme denotes the political party that contributed the majority of the district’s earmark requests. When a district’s House member and Senators are from different political parties and both contribute a high amount of earmark requests, a district may not have a distinct majority contributor. Earmark requests are assumed to be evenly split between two political parties when each party contributes more than 40% of the district’s earmark total. To show when earmark requests are evenly split, districts are depicted with a small ‘x’ texture (Figure 4.1).

To capture information regarding earmark requests independent of population size, state height varies relative to total earmark requests for the state. Total earmark requests at the state level are calculated as the sum of all Senate and House earmark requests for the fiscal year. States with higher earmark requests in absolute terms will appear taller than the surrounding states.

Figure 4.1: The visual features used to encode attributes from the U.S. congressional earmark request data set.

Dividing the state into its congressional districts does not capture the total amount of
earmark money requested by the state's Senators. To show this information, a small pie-chart glyph is shown with every state. This glyph is sectioned to show the percentage of the Senate earmarks that came from each Senator and has a saturation directly proportional to the Senators’ per-capita earmark requests (Figure 4.1).

4.3 Earmarks between Two Fiscal Years

This technique can also be used to visualize differences in district per-capita earmark requests between two fiscal years. Only the differences in per-capita district and Senate earmark requests are shown. By normalizing the earmark requests for population size, we are better able to compare districts of different sizes. The technique utilizes the same foundation as the single year visualization; however, I modified the method used to determine district color and saturation.

The district is colored by the political party that experienced an increase in the proportion of earmark money requested by that district. For instance, a red district indicates that the proportion of Republican earmark requests increased between the two fiscal years. Likewise, a blue district indicates that the proportion of Democratic earmark requests increased between the two fiscal years. The change in proportion may be due to an increase in requests by one party as well as a decrease in requests by the opposite party. The color does not denote the major earmark contributing party for the district.

District saturation indicates the strength of the change in per-capita earmark requests for the district. Regardless of political party, higher saturation indicates a larger change in per-capita earmark dollars. An ‘x’ texture is applied when this change is positive (i.e., the district experienced an increase in per-capita earmarks between the two fiscal years). Districts that experienced a decrease in per-capita earmarks are untextured.

A similar technique is used to color the Senate discs. The disc is colored to show the political party that experienced an increase in the proportion of Senate earmarks for the state. The disc is then shaded and textured using the same technique applied to districts. Requests from both Senators are combined when calculating the color and saturation of the Senate disc.
Chapter 5

Results

5.1 Earmarks in a Single Fiscal Year

Two sets of visualizations were generated, one for each of the 2009 and 2010 Congress fiscal years. These fiscal years begin on 1 October of the previous calendar year and end on 30 September of the current year. Figure 5.1 shows Louisiana’s 2010 fiscal year earmark requests and is a comprehensive example of the visualization technique. In the image, four of Louisiana’s congressional districts have the majority of earmark requests from Republican congressional members, while three congressional districts received the majority of their requests from Democratic congressional members (denoted with four red and three blue districts). Louisiana’s seven House members received different amounts of earmark requests; therefore, the districts are saturated unequally. The Senate disc shows that the state has a Senator from each political party, with the Democratic Senator requesting approximately twice the amount of earmark dollars as the Republican Senator (Senate glyph is $\frac{2}{3}$ blue and $\frac{1}{3}$ red). In one of the state’s districts, the distribution of earmark requests was split relatively evenly, indicated by the small ‘x’ texturing across the district. This image effectively represents the distribution of earmark requests within the state.

Although visualizations of individual states can be useful in understanding differences in
Figure 5.1: Visualization of Louisiana’s 2010 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over the state shows Senate per-capita earmark requests.

earmark requests among a state’s districts, multiple states must be viewed together to make regional comparisons. For example, within the East North Central Region of the United States [31], Illinois had the smallest absolute earmark request total, while Michigan had the largest total, since these states appear the shortest and the tallest, respectively (Figure 5.2). The Senate discs for Wisconsin and Illinois are not sectioned, which indicates that only one of the state Senators requested earmark funds. Although the earmark request total varies widely across these states, similar saturation reveals that the per-capita requests are relatively equal.
Several districts in this region also have about equal distribution of Republican and Democratic earmark requests, as is indicated by the ‘x’ texturing.

While regional patterns appear when visualizing congressional earmark requests for groups of states, more general conclusions and patterns can be observed when the country is viewed as a whole (Figures 5.3, 5.4). During fiscal 2010, a large earmark request total did not equate to a large amount of earmark requests in per-capita terms for some states. For example, California is taller than all of the other states; however, its districts have low saturation (i.e., the districts are pale), indicating a low per-capita amount. Although California’s Senate and House members bring in large amounts of earmark money, the state’s large population causes the dollars to spread thin across the population. The opposite situation can be observed in West Virginia, which has a low total earmark request total, and thus, appears shorter than other states, but with highly saturated districts. West Virginia has a small population relative to other states, resulting in a disproportionately large amount of per-capita earmark requests. Other states, such as North Dakota, Mississippi and Iowa, have large earmark request totals as well as large per-capita requests. These states sit taller than their surrounding states and have highly saturated districts.

Mississippi is an example where the districts are colored differently than the political party of their House representative (Figure 5.4). All of Mississippi’s districts are colored red; however, the state has three Democratic House members. The percentage of Senate earmark dollars received by these districts exceeds the earmark dollars received by the district’s House representative. Thus, the Republican party is the majority earmark contributor in these districts.
Figure 5.2: Visualization of East North Central States’ 2010 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over each state shows Senate per-capita earmark requests. State height reflects total earmark requests for the state, not normalized for state population.
Figure 5.3: Visualization of United States 2009 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over each state shows Senate per-capita earmark requests. State height reflects total earmark requests for the state, not normalized for state population.
Figure 5.4: Visualization of United States 2010 fiscal year congressional earmark requests delineated by district. The map is colored by requesting party and saturated by per-capita amounts. A textured ‘x’ pattern shows when district per-capita totals are evenly split between two political parties (i.e., each party contributes more than 40% of the district’s total). The disc over each state shows Senate per-capita earmark requests. State height reflects total earmark requests for the state, not normalized for state population.
5.2 Earmarks between Two Fiscal Years

A map displaying the change in per-capita earmark requests between 2009 and 2010 for the state of Ohio (Figure 5.5) provides a representative sample for this technique. From this image, the viewer can see that most of Ohio’s districts experienced a decrease in earmark requests, with only two districts receiving more money in 2010 than in 2009 (only two districts are textured). The amount of decrease in earmark requests varied across the districts, illustrated by the difference in district saturation. Moreover, the untextured Senate disc reveals a decrease in per-capita Senate earmark requests. The majority of the state is shaded blue, indicating that the proportion of earmark dollars requested by the Democratic party increased between the fiscal years.

Figure 5.5: Visualization of the change in congressional earmark requests between the 2009 and 2010 fiscal years for districts in Ohio. Color denotes the political party that experienced an increase the proportion of earmark money requested by that district. Saturation denotes the difference in per-capita earmarks between the two fiscal years. A textured ‘x’ indicates that there was an increase in earmark requests between the two years; absence of this texture indicates a decrease.
A visualization of the United States gives a better representation of the change between the two fiscal years (Figure 5.6). Very few districts on this map are textured, indicating that the vast majority of districts received fewer per-capita earmark dollars in 2010 compared to 2009. Additionally, a large portion of the map is blue, reflecting an increase in the percentage of earmark dollars requested by Democrats between 2009 and 2010. This can be attributed to a decrease in Republican requests and/or an increase in Democratic requests. It is important to note that the blue color does not signify that the Democratic party is the majority earmark contributor in these districts, but that the shift in political party for earmarks is trending toward the Democratic party.
Figure 5.6: Visualization of the change in congressional earmark requests between the 2009 and 2010 fiscal years for the United States. Color denotes the political party that experienced an increase the proportion of earmark money requested by that district. Saturation denotes the difference in per-capita earmarks between the two fiscal years. A textured ‘x’ indicates that there was an increase in earmark requests between the two years; absence of this texture indicates a decrease.
Chapter 6

Discussion and Future Work

I present a new technique for visualizing multidimensional scalar data over a topographical map and apply it to a dataset of U.S. congressional earmark requests. Earmark requests are presented in both total and per-capita dollar amounts, with the political party of the major contributor being highlighted. My technique is also capable of displaying differences in congressional earmark requests between fiscal years through the use of a trend map.

This visualization improves on existing techniques for visualizing multidimensional data on a map by addressing limitations of those methods. Unlike traditional choropleth maps, my technique represents multidimensional data on a single map surface. Although my technique does utilize a small disc above each state, all discs are the same size; therefore, it does not suffer from perceptual illusions that can exist in bubble charts and proportional symbol maps. My technique visualizes all regional data from all attributes on the map and thus does not under-represent variability, a problem that can occur with data-driven spots. No previous technique is capable of representing the number of attributes contained within the earmark data set, while maintaining the underlying geographical context and allowing for rapid information analysis.

The images resulting from this technique indicate that state size does not correspond to earmark requests in either absolute or per-capita earmark dollars. This implies that other factors determine the amount of earmark money a state or district receives. One hypothesis
is that districts represented by influential members of Congress will receive higher amounts of earmark requests. Tall and highly saturated states, indicating high absolute and per-capita totals, support this hypothesis. From the map, North Dakota, Mississippi and Hawaii fit this description. Coincidentally, during the 2010 fiscal year, the Senate Appropriations Committee Chairman was Daniel Inouye (D-Hawaii) and the ranking member was Thad Cochran (R-Mississippi) [5]. Senators from North Dakota, Iowa, California and Texas also served on the appropriations committee.

A limitation of my technique is that specific earmark requests are not shown within each district, and therefore, cannot be extracted from the visualization. For example, a one million dollar earmark request for a new school appears the same as a one million dollar earmark request for a baseball stadium renovation. Future work should expand this technique in order to answer the following questions:

- What is the distribution of earmark project types within a congressional district?
- Does a correlation exist between geographical area or political party and the type of earmark projects requested?
- Are there differences in the amount of earmark requests in certain categories (e.g., education, defense, social services) between political parties or geographic regions?
- Does the primary source of campaign funds during the most recent election influence the type of earmarks a member of Congress requests?

As implemented, the current system does not allow for the level of detail necessary to answer these questions. Future iterations could display only earmark totals from selected categories instead of total earmark dollars from congressional members. This visualization would address the first two questions; however, the detail necessary to draw correlations with campaign contributions would not be provided.

It is important to note that the per-capita earmark amounts reported in these visualizations may be biased. Although each district should contain an equal percentage of a state’s
population, district boundaries were last drawn based on 2000 U.S. Census data. Populations may have changed over the past ten years and may no longer be distributed equally across state districts. With the recent 2010 U.S. Census, some states are gaining or losing congressional members. Other states, while maintaining the same number of congressional members, are re-drawing their district boundaries to better reflect the distribution of the state’s population. This change will require that the technique be updated with the most recent congressional boundaries before it can be used to represent 2012 fiscal year earmark requests. Additionally, the change in district boundaries will reduce the ability to generate trend maps between the 2012 fiscal year and previous years.

While the earmark data required U.S. congressional district and state boundaries, this technique is not restricted to U.S. political delineation. It can also be applied to multidimensional data sets from any domain. This technique accommodates maps from various world regions with any pre-defined set of delineations (e.g., geographical or ecological boundaries).
REFERENCES


[33] United States Constitution, Art I, §1, cl 1-3.


