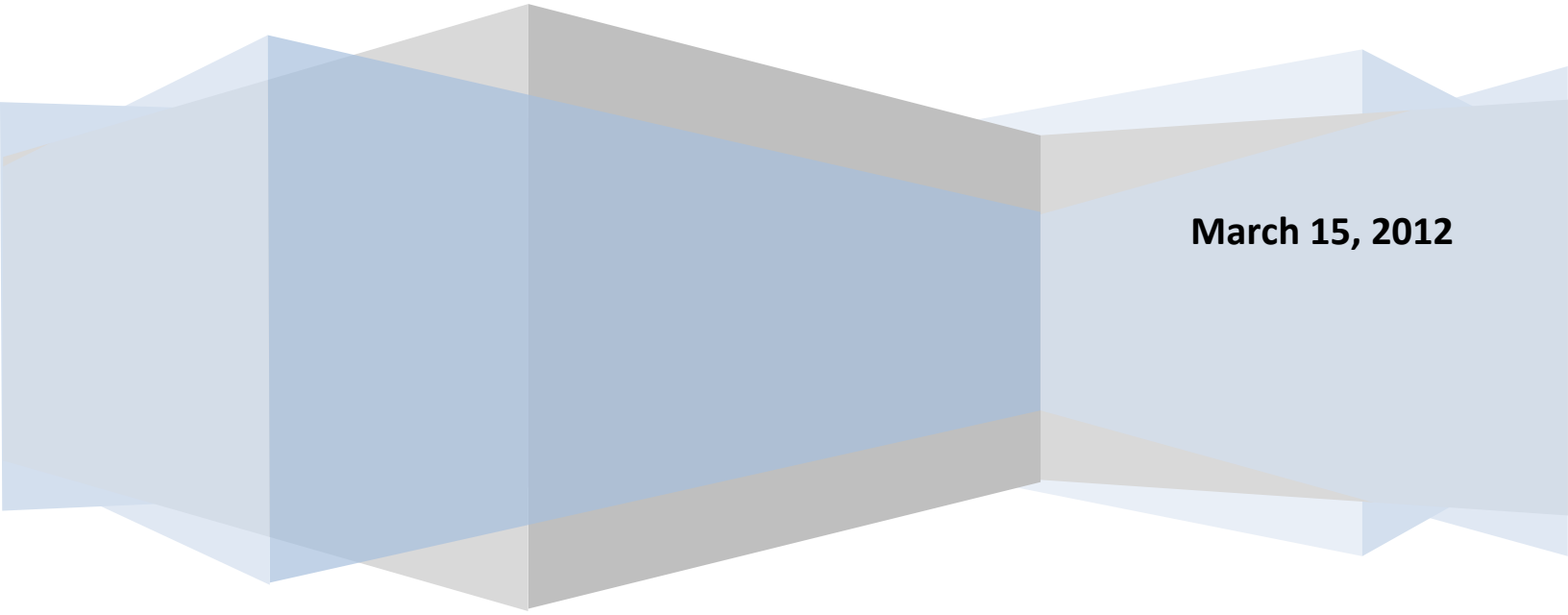
Censuses and Surveys of Governments: A Workshop on the Research and Methodology behind the Estimates

Visualizing Data from Government Census and Surveys: Plans for the Future

Kerstin Edwards



March 15, 2012

Visualizing Data from Government Census and Surveys:
Plans for the Future

Kerstin Edwards
Special Projects Staff, Governments Division
U.S. Census Bureau[1], Washington, DC 20233

1. Introduction

In 2007, the U.S. Census Bureau's State and Local Government Statistics Program underwent an external evaluation by the Committee on National Statistics, National Academy of Sciences. The Committee's 2007 report confirmed that the Governments Division (GOVS) data on state and local governments are the only comprehensive source on the fiscal welfare of state and local governments and provided recommendations for improving, building, and enhancing GOVS programs (National Research Council, 2007). The Committee recommended providing graphical analysis, mapping tools, and more explanatory material to assist users in understanding and using GOVS data.

In addition, the Census Bureau committed to promoting transparency, participation, and collaboration, by taking steps to increase public access to government data in accordance with the Open Government Directive from the White House in 2009. Developing and publishing data visualizations is one way to increase public access by expanding the audience of our data and aiding in data user comprehension.

In August 2011, the Census Bureau invited Nathan Yau, author of *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics* (Yau, 2011) and flowingdata.com, to visit the Census Bureau and share his insight into data visualization techniques. Yau collaborated with staff to improve data visualization efforts throughout the Census Bureau.

Also in 2011, GOVS established a small team of staff to identify and develop ways to implement new data visualization techniques using data from the censuses and surveys of state and local governments. The team aims to create visualizations that are informative, efficient, aesthetic, and novel (Iliinsky, 2010). Visualizations that are informative convey information to the reader. Visualizations that are efficient display only relevant information that is needed to convey the intended information. Visualizations that are aesthetic make appropriate use of colors, images, lines, and shapes to assist in conveying the intended information. Visualizations that are novel offer new ways of looking at data or new insights into the story that the data tell.

GOVS has developed the following types of visualizations using some of our most recent data: 1) information graphics, 2) static and interactive tree maps, 3) percentage change maps, 4)

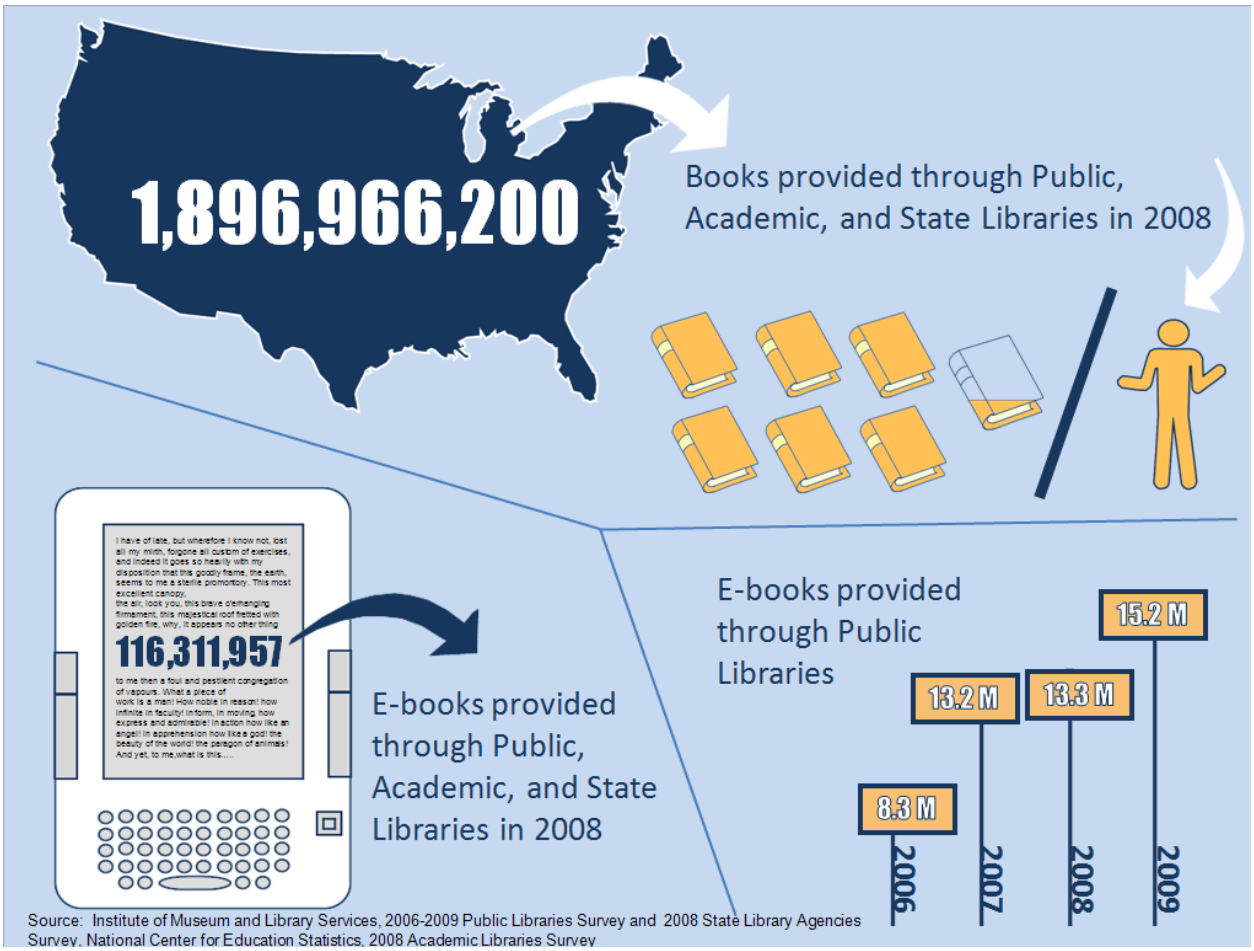
[1] *Disclaimer: This report is released to inform interested parties of research and to encourage discussion of work in progress. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.*

animated multi-year density maps, and 5) animated bubble plots. GOVS intends to use these and other visualizations to tell the story of the 2012 Census of Governments. In addition, research into new visualization techniques is ongoing.

2. Information Graphics

Information graphics use familiar objects to represent data. Information graphics can be used to highlight key findings and provide contextual information to aid understanding. For example, Figure 1 shows an information graphic, developed by GOVS, depicting libraries borrowing data.

Figure 1.



According to Figure 1, 1.9 billion books were provided through Public, Academic, and State Libraries in 2008, equating to 6.2 books per person in the United States. Displaying the number of books on a map of the United States, clearly denotes the geographic area covered by the data. Furthermore, visually displaying the total number of books per capita, provides additional contextual information to aid the reader in understanding the information presented.

Of the 1.9 billion books borrowed, 116.3 million were E-books. Displaying this figure on an illustration of an E-reader helps the user understand the term 'E-book,' a term that may be unfamiliar to many users because of its relatively recent invention. The figure also depicts how public libraries' lending of E-books has increased over a four-year period as E-reading devices have grown in popularity, a topic of particular interest to those in the library and publishing industries.

Information graphics can be easily understood by people with all levels of expertise and knowledge on a subject because of the simplicity of the visualization. While many of our users are able to appropriately interpret traditional data visualizations such as pie charts, bar graphs, or histograms, other users may not due to a lack of exposure or education on how to interpret graphs. In addition, information graphics can be used to grab users' attention to key findings and entice users to delve deeper into the subject to find additional information. For these reasons, GOVS will actively look for opportunities to use information graphics in outreach materials and in future publications.

Figure 1 was created using a combination of MS Office PowerPoint 2007 and SnagIT v10.

3. Tree Maps

A tree map is 'an area-based visualization where the size of each rectangle represents a metric' (Yau, 2011). Adding color to a tree map adds another useful dimension that helps communicate important information. Rectangles grouped together by color provide additional detail about the subcomponents that comprise a whole.

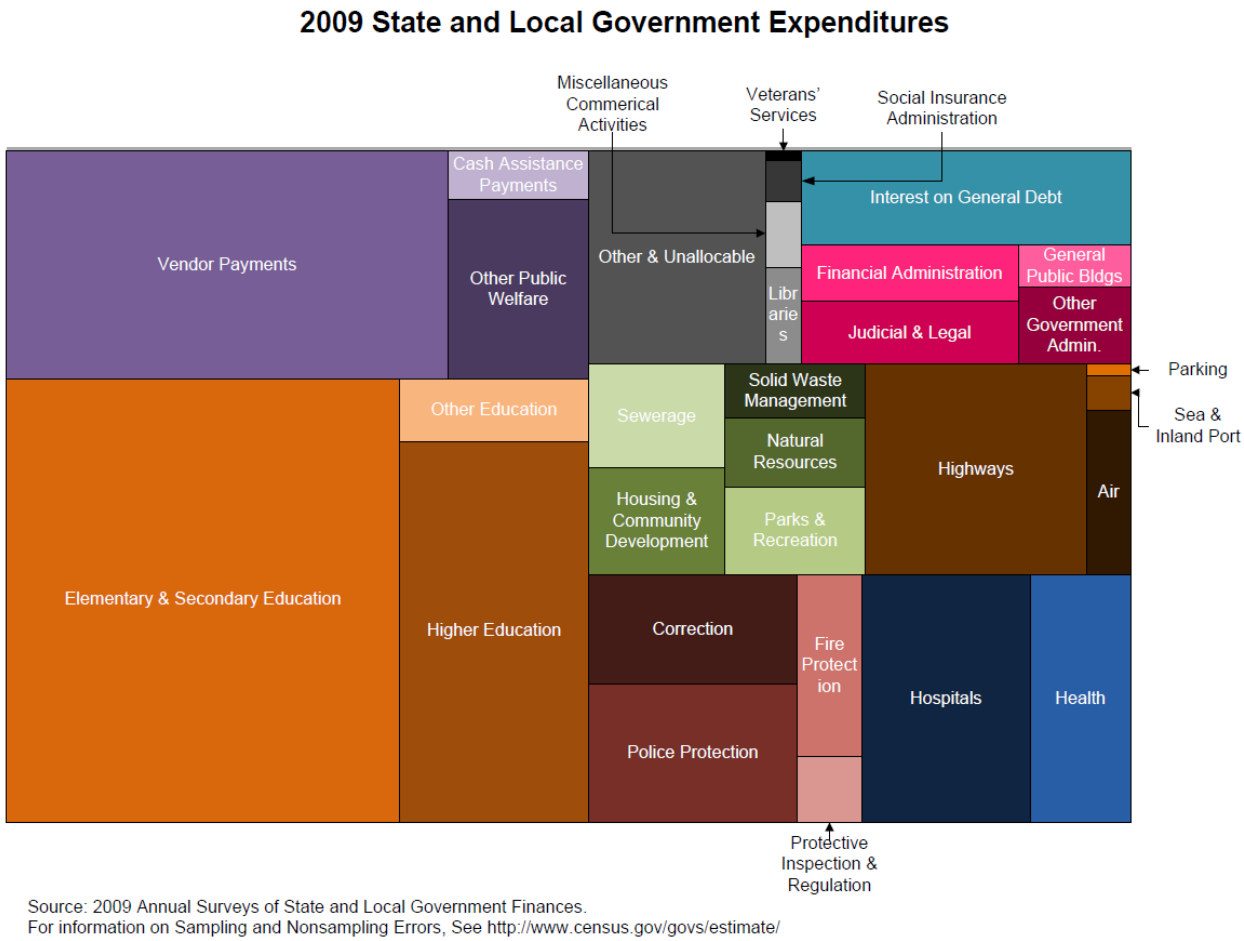
Tree maps provide more detail than a pie chart because subgroups can be clustered together using proximity and color. Tree maps are most useful when there are subsections that need to be displayed and when there are a large number of categories as well as categories that represent a small proportion of the data. This is because it is easier for the human eye to gauge and compare the area of a rectangle than a slice of a pie chart.

GOVS developed a series of tree maps to visually display the proportions that each government function and government activity receives in revenue or expends based on the 2009 Annual Survey of State and Local Government Finances. Figure 2 displays the distribution of state and local government expenditures by government function and activity in 2009.

Each rectangle represents a government activity. The size of the rectangle represents the total dollar spent by state and local governments on that activity. Each government activity is grouped with like activities and aggregated to make up government functions. The color of

each rectangle indicates the government function, where the larger rectangle includes smaller rectangles all in one color family. Grouped together, the aggregate of the smaller rectangles depict one government function. For example, Figure 2 displays 'highways,' 'parking,' 'sea and inland port,' and 'air' activities in varying shades of brown. The sum of these four activities constitutes the total dollars spent on transportation, a government function.

Figure 2.



Source: 2009 Annual Surveys of State and Local Government Finances.
For information on Sampling and Nonsampling Errors, See http://www.census.gov/govs/estimate/

According to the tree map in Figure 2, in 2009, state and local governments spent the most on elementary and secondary education activities. Furthermore, the function that state and local governments spent the most on was education, which includes elementary and secondary education, higher education, and other education activities.

Figure 2 was created using R, an open-source statistical software package for the initial creation of the chart, and MS Visio Professional 2007 for formatting and labeling. R outputs a tree map with rectangles in shades of red or green. Starting from the output from R, MS Visio Professional 2007 is used to change the colors of rectangles and font choices (type, color, and size) for the labels.

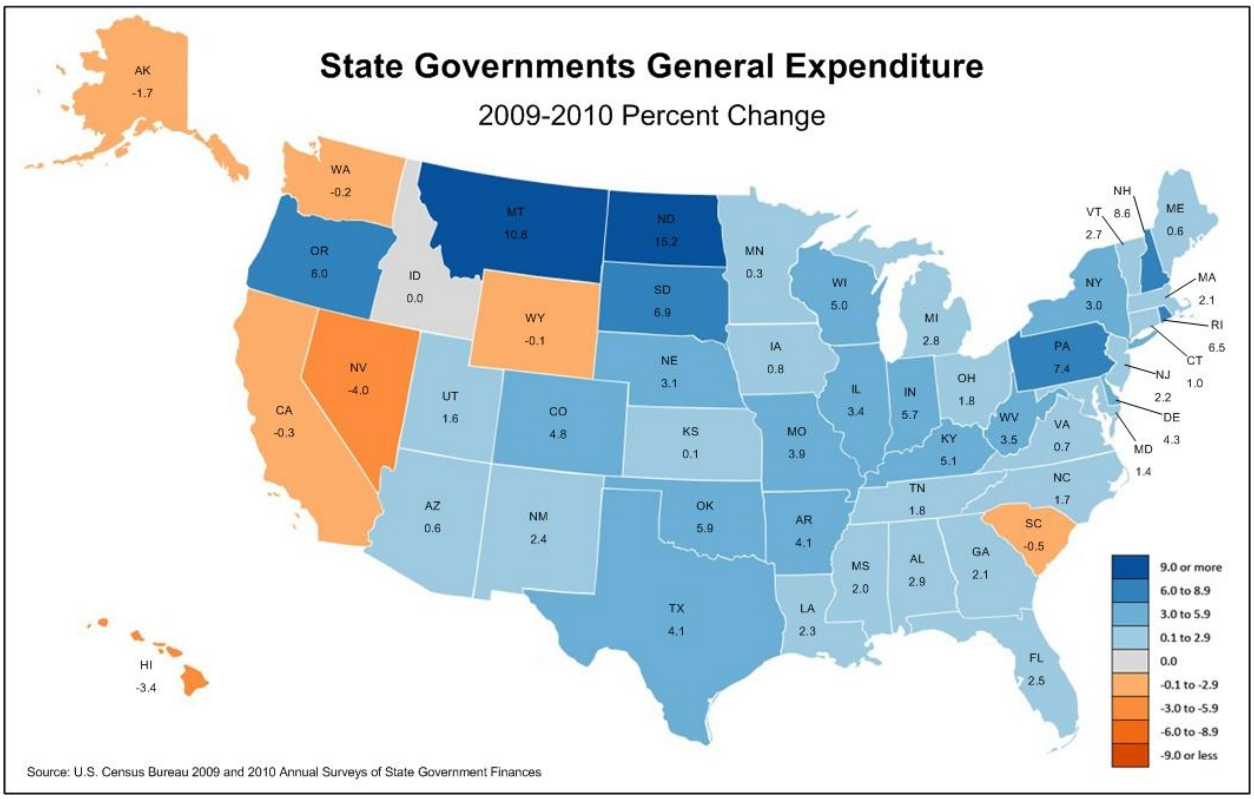
GOVS has expanded the usability of our tree maps by making them interactive. In an interactive tree map, a user is able to click on a category and find additional related information such as definitions, quality metrics, and possibly in the future additional visualizations specific to the category.

4. Percentage Change Maps

Displaying census and survey data on a United States map is an intuitive way to visualize the data because GOVS publishes the data by state. Displaying the data on a map allows the user to identify patterns or theorize possible explanations based on their knowledge of local events that may have had an effect on a specific state or region.

Figure 3 shows percent change between 2009 and 2010 in general expenditures by state based on the Annual Survey of State Government Finances.

Figure 3.



When the data are displayed geographically a user can easily identify the seven states that decreased in general expenditure as well as the varying degrees of increases that the remaining states saw allowing identification of any regional patterns.

Figure 3 was created using Python, an open source programming language, to create the initial maps. Adobe Photoshop was used to make cosmetic changes such as placement, formatting, and creating the legend. Finally, MS Visio Professional 2007 was used to add the text including the title and source statement.

GOVS plans to produce an analytical paper in the spring of 2012 which will discuss the percent changes between 2009 and 2010 for various revenue and expenditure categories of state governments. This will include several maps, each displaying data for different revenue and expenditure categories, but using the same legend. These maps will allow users to easily compare the percent changes between 2009 and 2010 for different revenue and expenditure categories by state. Furthermore, GOVS will explore creating an interactive application that would allow a user to select the revenue or expenditure categories or year(s) of interest to display.

5. Animated Multi-Year Density Maps

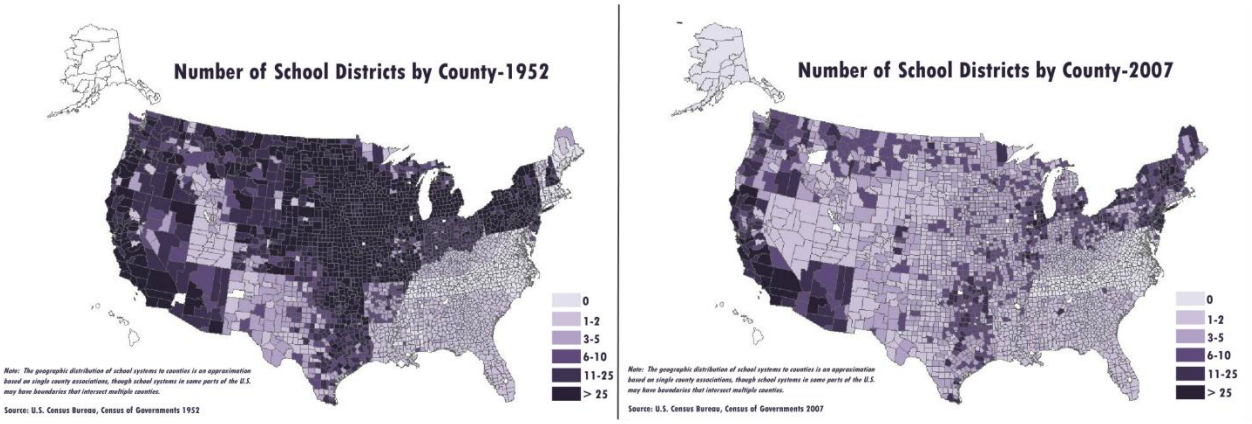
In addition to displaying year-to-year percentage changes on a map, GOVS has developed several density maps. A map displaying the density of local governments by county area was published using results from the 2007 Census of Governments. In addition, GOVS has begun developing an animated multi-year density map to display counts by county area for multiple years. By displaying each of these maps chronologically by year, the change in a single variable over time is highlighted.

Figure 4 displays a screenshot of two maps that display the number of school districts by county for 1952 and 2007. The geographic distribution of school systems to counties is an approximation based on single county associations, though school systems in some parts of the United States may have boundaries that intersect multiple counties. The screenshot was pulled from an animation that cycles through maps for every five years between 1952 to 2007. When the animation is displayed, a data user will notice the map change from dark shades of purple in 1952 to lighter shades of purple in 2007, signifying a decrease in the number of school districts per county.

The maps in Figure 4 were created using SAS and then formatted and animated using Adobe Flash Professional CS5.

This visualization will be released in the fall of 2012 in conjunction with the release of the 2011 Government Units Survey and will include the most recent count of school districts by county. A similar animation will be made using data on the number of special districts over time.

Figure 4.



6. Animated Bubble Plot

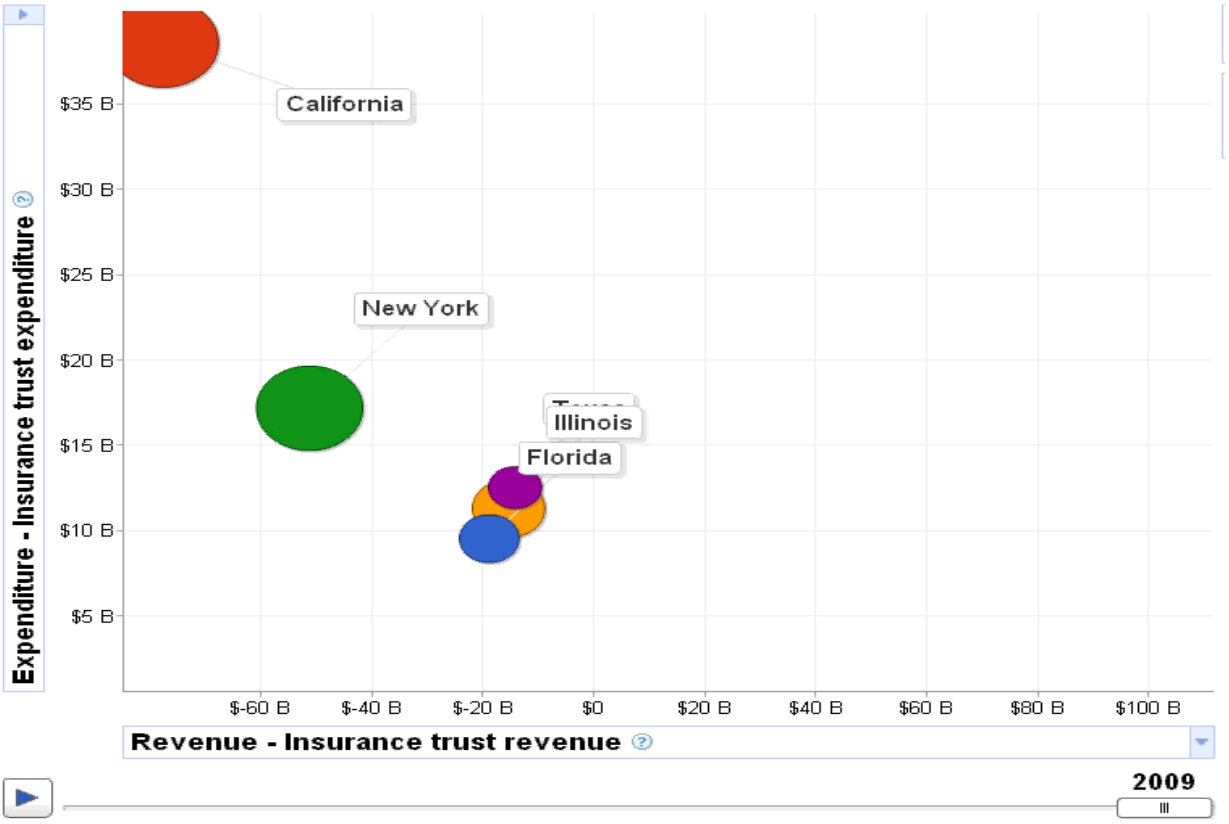
A bubble plot can be used to reveal relationships between multiple variables. One variable can be displayed on each axis, a third variable can be displayed as the size of the bubble, and a fourth variable can be displayed by the color of the bubble. In animated bubble plots, a fifth variable can be displayed by moving bubbles over time.

GOVS partnered with Google to include data from the Annual Survey of State government Finances in Google's Public Data Explorer ("State Government Finances," 2012). Using this tool, a user can create an animated bubble plot by selecting the variables of interest. Figure 5 is an example of an animated bubble plot created with this tool.

The plot shows insurance trust revenue plotted against insurance trust expenditure for five states. Each state is assigned a colored bubble and the size of the bubble represents total debt at the end of the fiscal year. The plot is animated to show how these variables change between 1998 and 2009.

GOVS will continue to identify data variables that may be correlated and where the relationship can be clearly depicted by a bubble plot (animated or otherwise). GOVS will also explore creating customized bubble plots using Adobe Flash Professional CS5 to customize the display and formatting options such as color, shape, and layout.

Figure 5.



Source: Annual Survey of State Government Finances, 1998-2009. Data are not subject to sampling error. Data are available at www.census.gov/govs/state/

7. Graphical Summary

For the 2007 Census of Governments, GOVS developed a graphical summary that was released in February 2012. This summary includes data on state and local government revenues, expenditures, taxes, education, retirement, and employment and contains traditional graphs such as bar charts, histograms, and pie charts to highlight key findings from the 2007 Census of Governments.

Using data from the upcoming 2012 Census of Governments, GOVS will develop a more modern graphical summary. Unlike the 2007 Graphical Summary, GOVS plans to structure the graphical summary topically rather than grouping data by survey. For example, the graphical summary may have a section on 'Transportation' and include revenue and expenditure data from the state and local government finance survey, employment and payroll information from the state and local government employment survey, and retirement data from the state and local retirement survey all related to the transportation government function. The graphical

summary will also include newly developed techniques for visualizing data rather than only the traditional graphs used in the 2007 version.

8. Ongoing Research

GOVS continues to develop ideas and research innovative methods for communicating information about the data we collect on state and local governments. Many of the visualizations that we are currently working on are based on variations of established visualization methods including step plots, stacked area charts, and donut charts, resulting in unique and innovative ways of helping our data users to better understand our data, find patterns, and highlight key findings.

Conclusion

GOVS has committed to develop and publish data visualizations in order to increase transparency and broaden the audience of our data. GOVS has developed or is developing several different types of data visualizations and will continue to research and create innovative ways to tell the story of state and local government finances while maintaining the integrity of the data.

References

National Research Council. (2007). *State and Local Government Statistics at a Crossroads.* Panel on Research and Development Priorities for the U.S. Census Bureau's State and Local Government Statistics Program. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Yau, N. (2011). *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics.* Indianapolis: Wiley Publishing, Inc.

Illinsky, N. (2010). On Beauty. In J. Steele & N. Illinsky (Eds.), *Beautiful Visualization: Looking at Data Through the Eyes of Experts* (pp. 1-13). Sebastopol: O'Reilly Media, Inc.

State Government Finances in the U.S. (February 29, 2012). *Google Public Data Explorer.* Retrieved from http://www.google.com/publicdata/explore?ds=i6b2dd9bq9ljq_