Government Information Quarterly 29 (2012) 212–222

Contents lists available at SciVerse ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

TreeCovery: Coordinated dual treemap visualization for exploring the Recovery Act

Miguel Rios-Berrios [1], Puneet Sharma [1], Tak Yeon Lee [1,*], Rachel Schwartz [1], Ben Shneiderman

Department of Computer Science, University of Maryland, USA

ARTICLE INFO

Available online 3 December 2011

Keywords:
Recovery Act
User interface
Visual analytic

ABSTRACT

The American Recovery and Reinvestment Act dedicated $787 billion to stimulate the U.S. economy and mandated the release of the data describing the exact distribution of that money. The dataset is a large and complex one; one of its distinguishing features is its bi-hierarchical structure, arising from the distribution of money through agencies to specific projects and the natural aggregation of awards based on location. To offer a comprehensive overview of the data, a visualization must incorporate both these hierarchies. We present TreeCovery, a tool that accomplishes this through the use of two coordinated treemaps. The tool includes a number of innovative features, including coordinated zooming and filtering and a proportional highlighting technique across the two trees. TreeCovery was designed to facilitate data exploration, and initial user studies suggest that it will be helpful in insight generation. RATB (Recovery Accountability and Transparency Board) has tested TreeCovery and is considering including the concept in their visual analytics.

1. Introduction

In February 2009, President Obama signed an economic stimulus package into law, dedicating $787 billion to create jobs and boost the economy, with the provision that the distribution of the money would be completely transparent. The growing recognition of the importance of design excellence in e-government applications (Fedorowicz & Dias, 2010) has raised attention to models for measuring user satisfaction (Verdegem & Verleye, 2009) and usability guidelines for e-government websites (Donker-Kuijer, de Jong, & Lentz, 2010). These concerns have taken on increased importance as the Obama administration expands its data availability efforts under the Open Government Directive. In fulfillment of this requirement, the agencies in charge of distributing the money and all recipients issued periodic reports detailing how the money they controlled was spent. These publicly available reports comprise a large amount of data, containing information about the effectiveness of the stimulus package, the general trends of distribution, and potentially interesting outliers.

Some effort has already been expended toward producing visualizations of this data that could assist in revealing such details. The government commissioned a website, Recovery.gov, dedicated to this purpose, and several independent journalism outlets have produced their own applications, all offering a particular take on the data. Most of the existing visualizations consist primarily of either tabular or geographical displays. The goals of this effort were to geographically display the distribution, allocation, and expenditure of stimulus recovery funds.

While the data lends itself well to geographical layout, given that states and counties are convenient schemas for chunking the data, exclusive use of maps cannot adequately portray alternate views of the monetary distribution. Specifically, money was distributed through 28 agencies, which assigned it to projects at their discretion; funding was placed in the charge of the prime recipient, who in turn funded sub-recipients and/or vendors as necessary for the project. Agencies naturally funded projects nationwide, and recipients for each project were not necessarily all located in the same area. This view of the data—an agency > project > recipient hierarchy—cannot be adequately conveyed by a geographical substrate.

Our tool, TreeCovery, offers a way to explore data both geographically and according to the monetary outlays. TreeCovery accomplishes this goal through the use of two coordinated treemaps, one drawn with a geographic hierarchy and the other one with levels corresponding to the agency > project > recipient money flow. While the views presented by the two treemaps differ, the underlying data remains identical at all times. Filtering is coordinated across the views and a proportional highlighting technique is used for coordination.

In addition to the coordinated treemap design, we incorporated a few other features to improve exploration techniques. From news articles about the Recovery Act we found many of them using demographical statistics such as population or unemployment rates. We thus included census data for each county, and made it possible to filter by demographic attributes. We also added the ability to save snapshots of the current state of the treemap for3 later comparison. Finally, we included support for emphasizing invalid data values.

* Corresponding author at: HCIL/UMIACS, University of Maryland, 2117 Hornbake Bldg, South Wing, College Park, MD 20742, USA. Fax: +1 301 314 9145.
E-mail addresses: mrios@cs.umd.edu (M. Rios-Berrios), puneet@cs.umd.edu (P. Sharma), tylee@umd.edu (T.Y. Lee), rachel13@cs.umd.edu (R. Schwartz), ben@cs.umd.edu (B. Shneiderman).
[1] The first four authors contributed equally to the project.

0740-624X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.giq.2011.07.004

The goals of this effort were to enable users to

- Explore the allocation of stimulus recovery funds by agency and states/counties;
- Identify extreme cases such as concentration of spending by an agency in one state or county;
- Discover unusual patterns of spending that show inequities by region, state, and county;
- Understand which agencies were most active in their state or county;
- Facilitate error detection and omissions in the data.

Of course, a powerful interface would also enable other tasks, which were beyond the initial planning of the designers. The larger goals are to empower policy makers, journalists, and citizen groups to have increased capacity to explore key data sets that are tied to national priorities.

TreeCovery has been tested by RATB and received positive feedback about the concept. Hopefully features of TreeCovery will be included into next version of RATB's visual analytic platform.

Treemaps are among the growing set of information visualization tools that could increase the analytic capabilities of government agency staffers, political analysts, journalists, and other interested citizens. The capacity to identify interesting patterns, clusters, gap, outliers, and others features is increasingly important in detecting fraud, ensuring fair allocation of resources, and refining policies to ensure effective use of public funds.

Section 2 discusses related work, while Section 3 provides an explanation of our analysis process, including a detailed illustration of Spotfire's[2] ability to support exploration of the Recovery Act data. Section 4 explains TreeCovery in detail, while Section 5 offers some sample insights found by the tool. Sections 6 and 7 suggest future work and offer conclusions.

2. Related work

Because the stimulus information is both newsworthy and publically available, many visualizations of the data are already available. First and foremost, recovery.gov offers geographical maps displaying award locality (Fig. 1). The maps can be zoomed in to state and zip code levels and show dots each representing a project colored by its award type—contract, grant, and loan. The site also offers some pie and bar chart summaries, as well as tabular data. While the basic information is thus available, interaction with the visualization and customization is limited.

Many other websites offer similar tools to those of recovery.gov. The Federal Procurement Data System,[3] which has the raw data available for download, offers a few selected slightly interactive visualizations as well; again, interactivity is limited and the data available through the visualizations is limited as well. The website ProPublica hosts a feature with Recovery Act information, Eye on the Stimulus.[4] Most of the site is devoted to text articles, but tabular and geographical visualizations of spending progress are offered as well. Other sites following the geographical/tabular trend include the *Wall Street Journal*[5] and msnbc. [6]

The existing visualizations of stimulus data, while informative, do not support exploratory analysis of the data. Because the data contains dual hierarchies, the geographical one and the monetary distribution, it can be most effectively portrayed using visualizations tailored toward this structure. Several methods of achieving this have been discussed in the literature. Polyarchy Visualization allows representation of intersecting hierarchies (Robertson, Cameron, Czerwinski, & Robbins, 2002). While useful, this approach does not provide well for visualization of multidimensional data, as is necessary for recovery.gov data. Multitrees describe the general structure of the data: nodes that are shared by multiple ancestor trees (Kules, Shneiderman, & Plaisant, 2003). However, they are presented as a graph theoretic rather than visual concept and, as such, are not particularly helpful in building our tool.

Treemaps are a space-filling method of visualizing large data sets, which have proved very useful for displaying hierarchical structure (Shneiderman, 1992). There exist many variations of treemaps specialized for different purposes and characteristics of data sets. Burch and Diehl introduce the "Trees in a Treemap" technique to represent trees with an associated taxonomy (Burch & Diehl, 2006). The taxonomy is represented as a treemap, and the related tree is drawn on top. In this representation, the two (or more) hierarchies are not represented in a symmetric fashion; attribute information about the hierarchy can be easily integrated, but the nodes of the trees show only their structural position. Therefore, this approach does not suit our data. Jern et al. suggest using a treemap visualization in combination with a cartographic one (Jern, Rogstadius, & Astrom, 2009). They demonstrate the value of multiple displays of the same data, but their treemap hierarchy is regional. The stimulus visualization requires a method that represents data in multiple hierarchies.

Another treemap-based hierarchical representation is that of Wood et al. (Slingsby, Dykes, Wood, & Crooks, 2009; Wood & Dykes, 2008; Wood, Slingsby, & Dykes, 2008). A specialized ordering is used to facilitate spatial and temporal locality so that the layout of nodes is more intuitive. Wood's approach combines two hierarchies (temporal and spatial) into a single treemap. We felt it was important for users to be able to explore each hierarchy in isolation, as well as in conjunction with the other. TreeCovery therefore uses two treemaps to represent the data using two different hierarchies, which is an instance of multiple coordinated visualizations.

Multiple coordinated visualizations help users explore complex data. There are four common types of coordination; Brushing and linking, Overview and detail view, Drill down, and Synchronized scrolling (North & Shneiderman, 2000). As supporting exploration of hierarchical data sets was the goal, TreeCovery focuses on the coordinated drill down that allows users to navigate down successive layers of a hierarchical database (Fredrikson, North, Plaisant, & Shneiderman, 1999). Among many cases of multiple coordinated visualizations, PairTrees (Kules et al., 2003) is an interesting case as it also utilizes treemaps to support exploratory data analysis on multiple hierarchical attributes. In PairTrees, two treemaps initially show overviews of the data set using two hierarchies based on aggregation. When an element is selected in one treemap, another treemap is automatically filtered by the selected element revealing meaningful relationships between the two hierarchies. Additional flexibility is gained in this special type of dynamic query by giving no fixed role of control and representation parts.

Using two coordinated treemaps raises the design question of how to use brushing that highlights selected elements on multiple coordinated visualizations. Brushing was first used in the PRIM9 system by Tukey et al. (Fisherkeller, Friedman, & Tukey, 1988) and now is applied in almost every interactive visualization environment (Roberts & Wright, 2006). While in most designs selected elements are present on every view, the two treemaps in TreeCovery have elements partially related to each other. For example, in existing applications such as Spotfire, given an agency selected in one view another view would highlight all the states awarded any amount of money from the agency. This, however, conflicts with the space-filling rule of treemap (size of an element is the sum of its children) as the size of the highlighted area does not necessarily match with the amount of money given. To address the problem, proportional representations have been used in bar charts and other visualizations (Shrinivasan & van Wijk, 2009),

[2] http://spotfire.tibco.com.
[3] https://www.fpds.gov.
[4] http://www.propublica.org/ion/stimulus.
[5] http://online.wsj.com/public/resources/documents/info-STIMULUS0903.html.
[6] http://www.msnbc.msn.com/id/33498869.

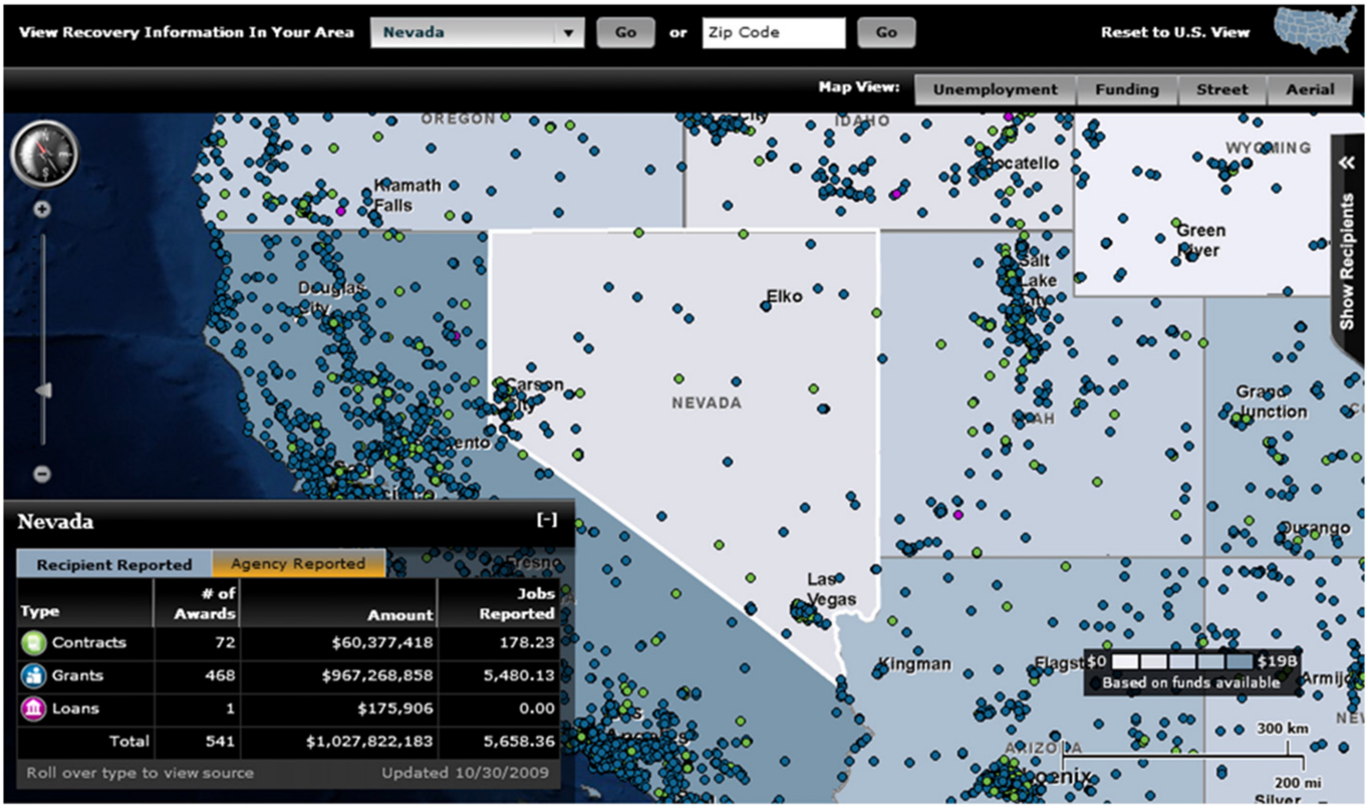

Fig. 1. Recovery.gov offers geographical maps showing each award as a dot.

but showing proportional representations in treemaps offers users a rapid means to discover extreme values in context. TreeCovery is the first attempt that represents actual proportion of related sub-elements in treemaps.

3. Analysis and methodology

To design our tool, we first needed to determine the chief goals of stimulus data visualization. As recipient reports of the Recovery Act had just been released, it was not easy to find end-users who had already done extensive work on the data. Thus, instead of using direct interviews or a survey, we decided to do reverse-engineering on relevant news articles in order to understand the process of journalists analyzing it. Further, we analyzed the7 data with Spotfire, one of the most versatile visual analytic applications. These exploratory tasks helped us understand the data better and develop the concept of our tool.

3.1. News insights

Journalists are the most prominent group interested in the Recovery Act data. ProPublica, for example, is a group of investigative journalists who analyze the raw Recovery Act data and produce original news articles that influence other news producers. Governmental organizations are quite responsive to those investigative journalists. From news articles we could infer what people want to know and how the data should be analyzed. More specifically, our research questions were, “What are the primary methods for analysis?” and “What kind of visualization can support the analysis?” We searched in the Google News search engine by the keyword “Recovery Act” for news reported during the first two weeks of November 2009—right after the recipient reports had been publicized on recovery.gov.

Most prominently, many findings focused on a State/County comparison. Although geographical region is not the main hierarchy of the Recovery Act plan, the most frequently asked question was something like, “How much money is given to our state/county?” For example, an article compares the amounts of awards given to two states: “Idaho Gets Four times More Stimulus Money in contracts Than Louisiana” (Glantz, 2009). It is noteworthy that state/county comparisons require aggregation of projects in each state or county.

The second insight was the usage of census data. In order to find states/counties in similar context or to validate fairness of funding from a specific agency, census data is quite useful. For example, an article (Donovan, 2009) referred to high-school graduation rates, infant mortality rates, unemployment rates, and juvenile justice incarceration to pick the 5 worst cities for youth and compared numbers of jobs created by the Recovery Act[7] in those places.

The third insight we found was validity checking to reveal unlikely numbers and non-existing categorical values. As each recipient report was submitted through an online form by the recipient, it is natural for the reports to have some errors due to simple mistakes. Non-existing congressional district codes are a typical case of simple mistakes criticized by CNN (Mooney, Bolduan, & Hanna, 2009); however, other more complicated cases can exist too. For instance, number of jobs created, which is an important gauge of success, can be interpreted in many different ways. Usually, invalid values are either resolved or filtered out in information visualization, although they have significant importance especially for a federal government website.

Summing up the findings above, we came up with the idea of bi-hierarchical data exploration in Fig. 2. While the Recovery Act funding is distributed to recipients along the agency tree, the information of

[7] http://news.google.com/news/search?aq=f&pz=1&cf=all&ned=us&hl=en&q=Recovery+Act.

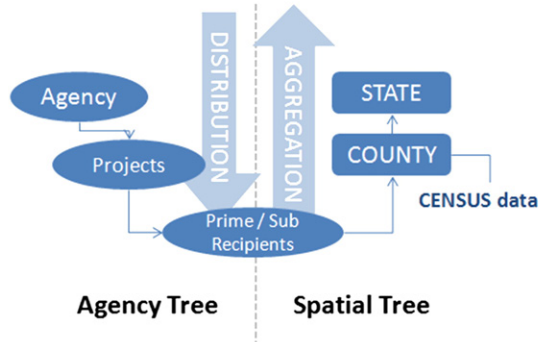

Fig. 2. The Recovery Act has two equally important hierarchies—agency and spatial trees.

recipients is also aggregated by county and state in combination with census data.

3.2. Spotfire insight

Prior to building our own tool, we analyzed the recipient data[8] with an existing visual analytic toolkit. Spotfire was chosen because of its wide set of features for dealing with multivariate data. This analysis had three purposes. Firstly, analyzing the actual data provided us a deeper understanding. Secondly, we understood the capabilities and limitations of current visualization techniques. Finally, the exploration was helpful in suggesting opportunities for improvements on existing visualization techniques.

In addition to the recipient report, we also incorporated the State and County QuickFacts dataset[9] from the U.S. Census Bureau. The census data consists of a wide variety of demographic profiles for each county, including Population, Infant deaths, Housing unit, Household income, and Unemployment rate.

The paragraphs below provide examples of the headlines we found by Spotfire analysis.

"The most effective job creators are suspect": Knowing that there are projects with an improbably high number of jobs created, we drew a treemap visualization (Fig. 3) showing which departments or states were related to those projects. Colors of elements represented money per job, which means how much money was spent for creating each job in a project. We had to filter projects by money per job in order that small important outliers be visible in the treemap. We found that treemap visualization was suitable for displaying job creation anomalies.

"Florida, the highest percentage of senior residence in the U.S., gets the most money from military sources": we tried to incorporate census data with recipient reports in this example. In Fig. 4, counties with high percentages of senior citizens are selected in the scatter plot. Then all projects in the selected counties were also highlighted by brushing and highlighting technique (Kules et al., 2003), and it was easy to see that Department of Defense and Air Force are major sources of the project funding. Brushing and highlighting is a useful way to interconnect multiple visualizations, however this method often misleads as it highlights the entire portion of the element. Thus even when one state contains only a tiny single project related to the brushed selection, it would look as if the state has a lot of related projects.

4. TreeCovery

We designed TreeCovery to be useful for investigative journalists and citizen watchdogs that have some domain knowledge and experience in data analysis. It streamlines the exploration process available through existing visualization techniques and adds more features for data analysis. This section elaborates on the development platform, data, and UI components of TreeCovery.

4.1. Software architecture

TreeCovery is implemented as a desktop application written in the Java programming language using the Prefuse visualization toolkit (Heer, Card, & Landay, 2005). We decided to use the Treemap implementation in the Prefuse toolkit due to its strength in the visualization area as well as its data management. Although the implementation was not robust enough, it served as a perfect base in which we could build our application.

TreeCovery's implementation has three main components: the container of the user interface with the controls and settings, the treemaps, and a data processing tool. When the data file (in CSV format) is loaded into the application, the data processing tool converts the tables into trees, by using Prefuse's tree implementation. The trees are fed into the treemap layouts, which are initialized with several internal settings as colors, fonts, and others, and then they are displayed in the container of the main user interface.

The filtering process in the treemaps is done using Prefuse predicates, which filter the underlying data table fast and efficiently by binding directly the visual sliders to the data that is stored in the memory. Prefuse provides in-memory querying of data, what was essential in our case because we are using real-time dynamic filters using several criteria. This created a limitation because the amounts of data we are handling are huge, and using this approach all the data had to fit in memory. We developed a version that stored the data in a SQL-based database and performed informal benchmarks. As TreeCovery relies on intensive real-time filtering and is well-known that database access is slower than in-memory access, we decided to continue using the former and mitigated memory bottlenecks by increasing the Java memory heap size in our application.

Other components such as the proportional highlighting visualization were developed from scratch. In this case the overlaid component is drawn according to the dimensions retrieved from a process that retrieves and re-sizes all the components of the original treemap according to the computed proportion. Another main component of the application that was built from scratch is the synchronization between treemaps, which was achieved by setting listeners that respond to events in any of the treemaps. They are monitoring each change, and when appropriately, they synchronize and show back the filtered or zoomed visualizations.

4.2. Data pre-processing

TreeCovery relies on the recipient report data from recovery.gov.[10] Information is available for each recipient of any stimulus dollars, including those attributes that are necessary to build the hierarchies (department, project, state, and county) as well as attributes (number of jobs created, award amount) that are used for filtering, size-by, and color-by, and other attributes that can be viewed in the details-on-demand panel. Our dataset includes a few interesting values for each record generated during the preprocess. For instance, we included money per job for each recipient, as that proportion may be more germane to the questions of stimulus effectiveness than either job creation or award amount alone. Besides the standard recipient data, TreeCovery also integrates census data fields such as population, education level, and unemployment rate. We found that many of the headlines discussing recovery data also used census information so we felt it would be very useful to integrate that information directly into the tool.

[8] http://www.recovery.gov/FAQ/Pages/DownloadCenter.aspx.
[9] http://quickfacts.census.gov/qfd/download.
[10] http://www.recovery.gov/FAQ/Pages/DownloadCenter.aspx.

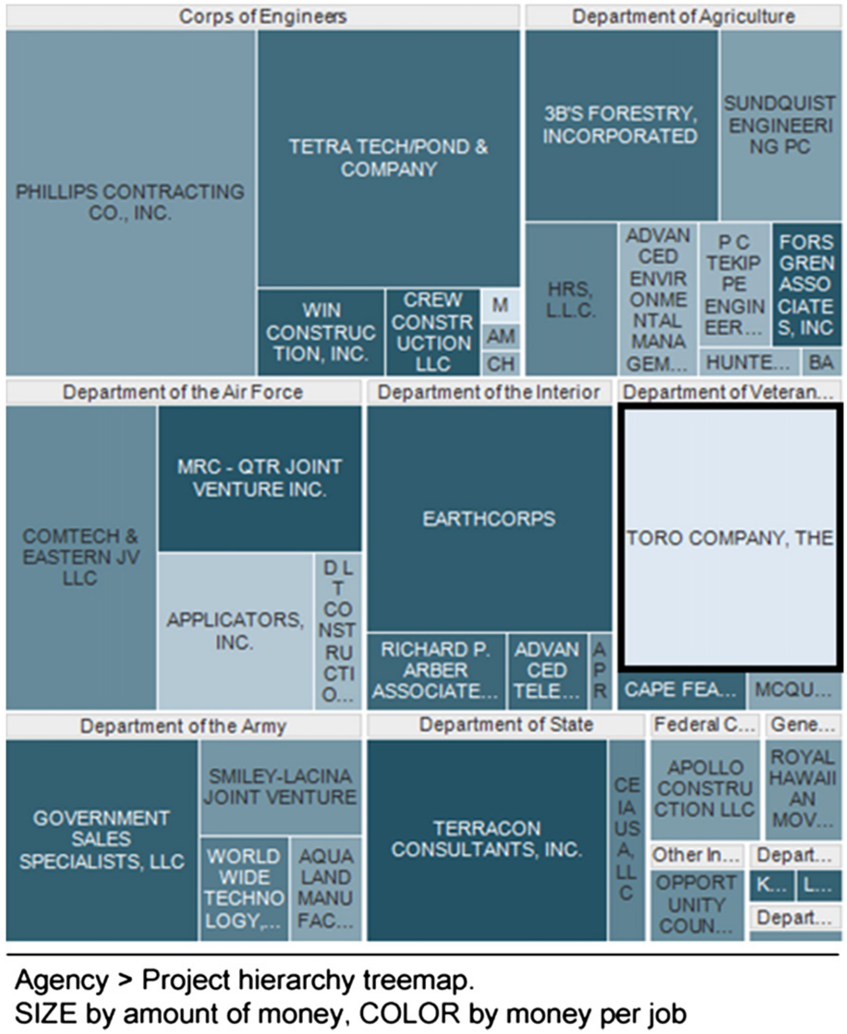

Fig. 3. Treemaps can show the distribution of recovery funds effectively with other attributes as well. Toro Company is an exceptional recipient which created a number of jobs with very small money.

In addition to this standard preprocess, three specific types of invalid values are searched and marked: invalid zip codes, invalid congressional districts, and projects that had no primary recipient, only sub-recipients. Invalid zip codes were simply those that do not exist in the standard zip codes table (253 recipients found). Invalid congressional districts were determined on a state-by-state basis, so that if a district did not exist in that recipient's state, it was flagged as invalid (38 project found). Finally, there were 561 projects with no prime recipient reported at all. All these errors were flagged and integrated into the TreeCovery display to be presented to RATB.

4.3. Visualization

The chief component of the tool is, of course, the dual treemap display (Fig. 5). The left treemap displays the agency (department > project) hierarchy, while the right one is spatial (state > county). Although the two share identical underlying data (the recipients) as their leaves, this level is never visible on the treemap. The shared leaf level data does, however, make it possible to coordinate the displays.

Essentially, filtering occurs simultaneously (Fig. 6); zooming in on one treemap, which amounts to filtering on 14 the recipient leaves of the zoomed node, causes the other treemap to be filtered on those same recipients. Although Spotfire allows users to both zoom and filter, the two actions are independent, so that side by side treemaps will not stay coordinated automatically. In TreeCovery, users can zoom in and out on both treemaps in any arbitrary order, and the recipient leaves included in the layout will remain coordinated.

In addition to the filtering capabilities provided through zooming, TreeCovery also allows customization of the treemap displays using a set of controls. Double-sided sliders allow filtering on various attributes of the recipients and the census information for their areas. One innovative feature we added was the ability to filter out invalid values found in the preprocessing stage. Each treemap can be sized and colored on a chosen attribute as well.

TreeCovery's main innovation lies in its proportional highlighting capability. This feature completes the coordination of the two treemaps. Although Spotfire highlights child nodes in a treemap when the parent is selected in another one, our highlighting technique is much more finely tuned.

When a node is clicked in one treemap, a highlighting square is placed inside all nodes in the other treemaps that share any child recipients with the selected node. The square's size is proportional to the ratio of the shared children's areas to the total area of the node. For example, if the selected node has a child with an area of 10, and its parent in the other tree has an area of 100, then the highlighted square will take up 1/10 of the parent in the other tree's area. In this way, the exact distribution of the size by attribute for the selected attribute across the other hierarchy becomes easily apparent. Using the default proportionality constant of 1, the total highlighted areas will, in fact, equal the area of the selected node. Users can change the

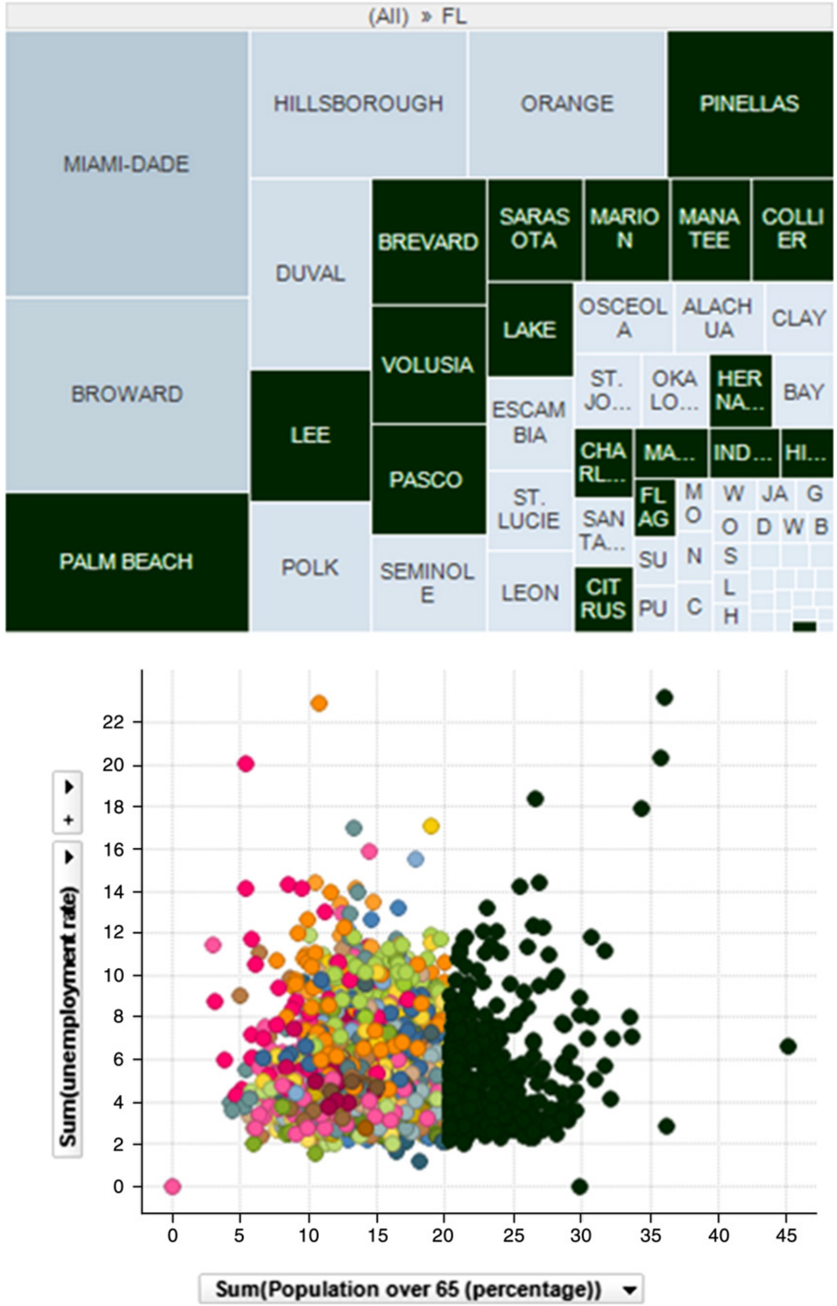

Fig. 4. Brushing and highlighting technique is useful for connecting multiple visualizations.

proportionality constant for sizing the squares, in case they want to emphasize the highlighted areas. They can also control the opacity of the squares, so that the underlying labels will be visible. This highly flexible highlighting capability displays extensive information about attribute distribution across hierarchies in a powerful and intuitive manner.

Besides the main treemap functionality, TreeCovery provides additional features for data exploration. Firstly, details of the recipients that comprise each node are available when selected. A Details-on-Demand table displays the data for each recipient, including values for all attribute fields in our data set. Each row of the table can be double-clicked to bring up the recovery.gov page on that project. TreeCovery also allows users to save the current view of a treemap for later viewing. The image of a single treemap is saved, along with the current filter, zoom, size-by, and color-by settings. All saved images are shown as thumbnails in the shoebox area. Users can select and view any number of the saved images side by side in a separate window.

5. Evaluation

5.1. Insights

To demonstrate the utility of TreeCovery, we give three examples of finding insights.

5.1.1. The major source of awards assigned to California (CA) is a few big projects from the Department of Education (Fig. 7)

TreeCovery was run with the contract, grant, and loan data. Both treemaps were sized by the amount of money, and colored by the number of jobs created. California received the highest amount of awards, which was easily visible in the spatial treemap (right). By selecting California, it quickly highlights the related awards in the agency treemap (left). The distribution of highlighted awards illustrates whether California received a few bigger awards or many

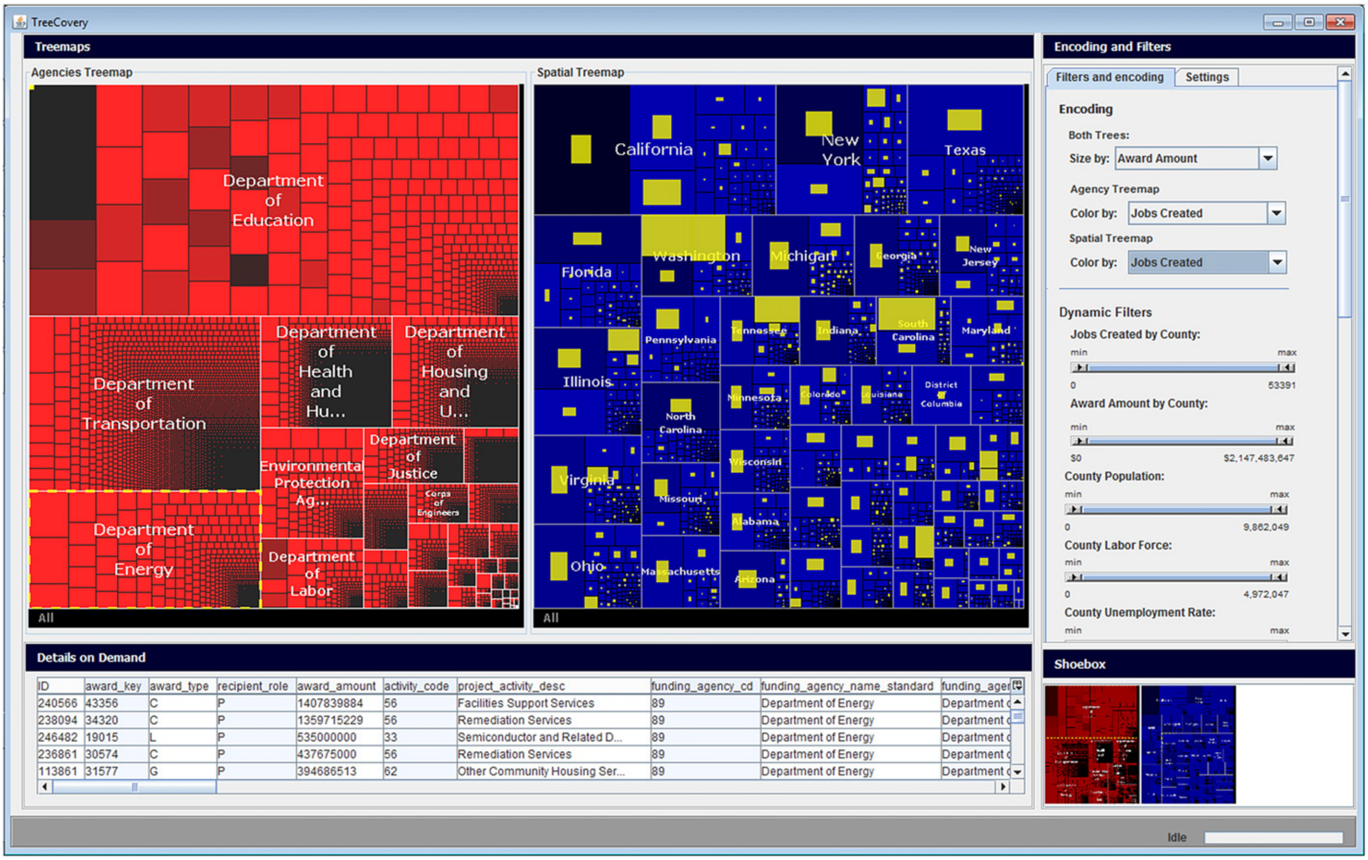

Fig. 5. TreeCovery User Interface with the Recovery Act data. The pair of treemaps is at the upper left, the control panels for Filters and Settings are at the upper right, the tabular data is at the lower left, and the shoebox with user selected screenshots is at the lower right.

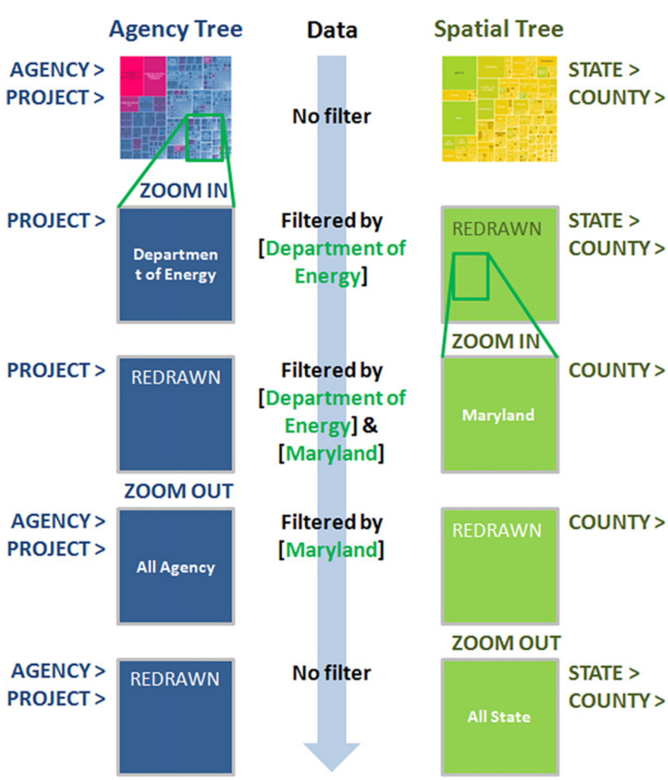

Fig. 6. Two treemaps share the identical data whose filter is controlled by zooming activity on either treemap.

smaller awards. From the Department of Education, California received the two biggest awards in a dark color that means very few jobs have been created by those projects. It is worthwhile to look into more details of them. The coordinated highlighting feature enables users to quickly examine the overview and motivates further exploration.

District of Columbia (DC), the federal hub, received the highest amount of contract money from the General Service Administration (Fig. 8). To illustrate TreeCovery's zooming feature, both treemaps were sized by the award amount and the General Services Administration (GSA) node was selected in the agency treemap. Various counties were highlighted in the spatial treemap, and it was clear that DC had received the highest amount of contract money. We zoomed in on the GSA node (Fig. 9) and found that one of the biggest projects of GSA was allocated to DC and all of its recipients (Prime and Sub-Prime) were in DC.

Georgia is creating more jobs while getting less contract award money (Fig. 10). To see the utility of the shoebox in analysis, the spatial treemap was sized by award amount and a snapshot was taken. Another snapshot was taken after sizing the spatial treemap by jobs created. Both images were selected in the shoebox and opened in the comparison window. In the figure, the left treemap shows the money distribution and the right treemap shows jobs creation. Usually, states getting more money create more jobs but Georgia (GA) stood out as an outlier, with more jobs created and a comparatively small contract award amount.

5.2. Usability evaluation

To find unexpected usability problems and insights for revision, we conducted a usability evaluation with four graduate students in

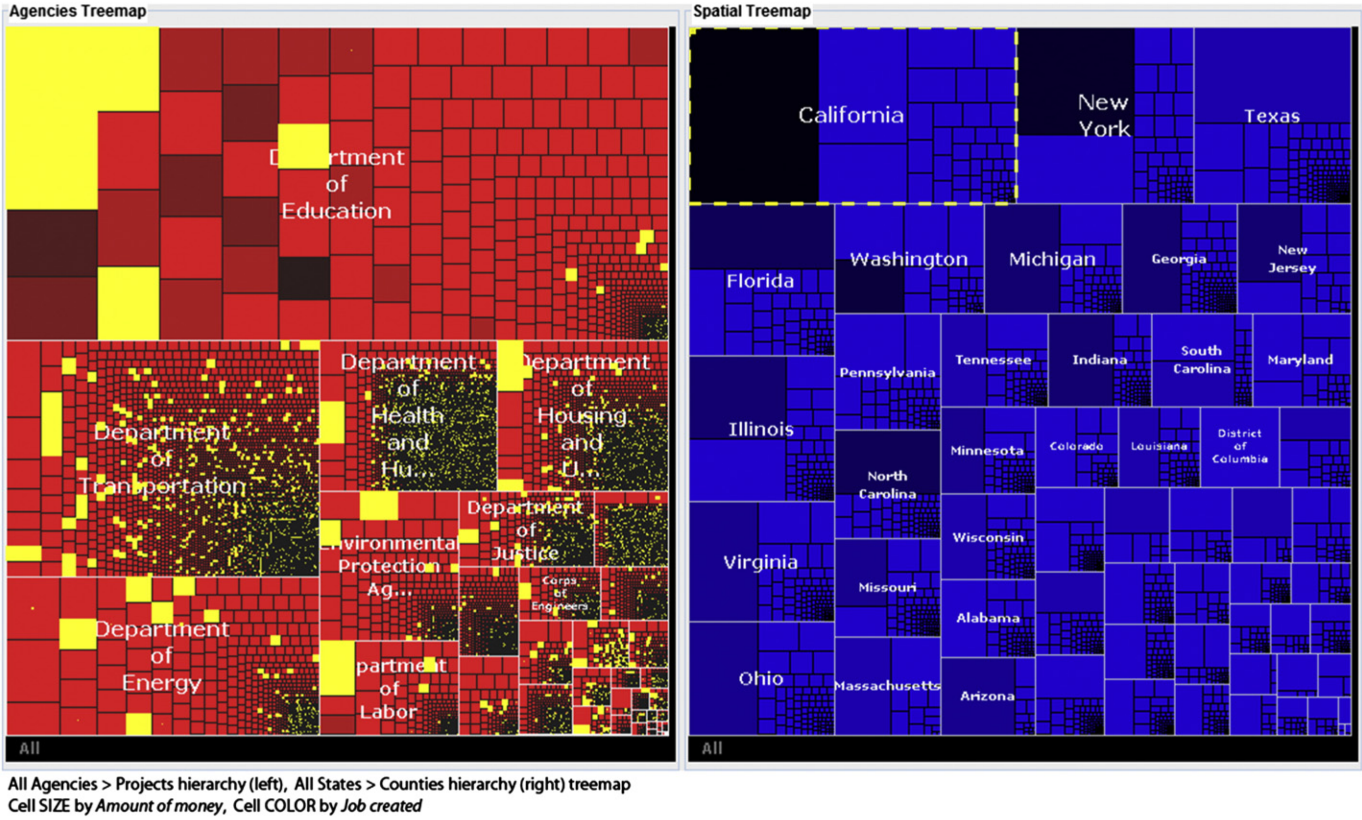

Fig. 7. The major source of awards assigned to California (highlighted) is a few big projects from the Department of Education.

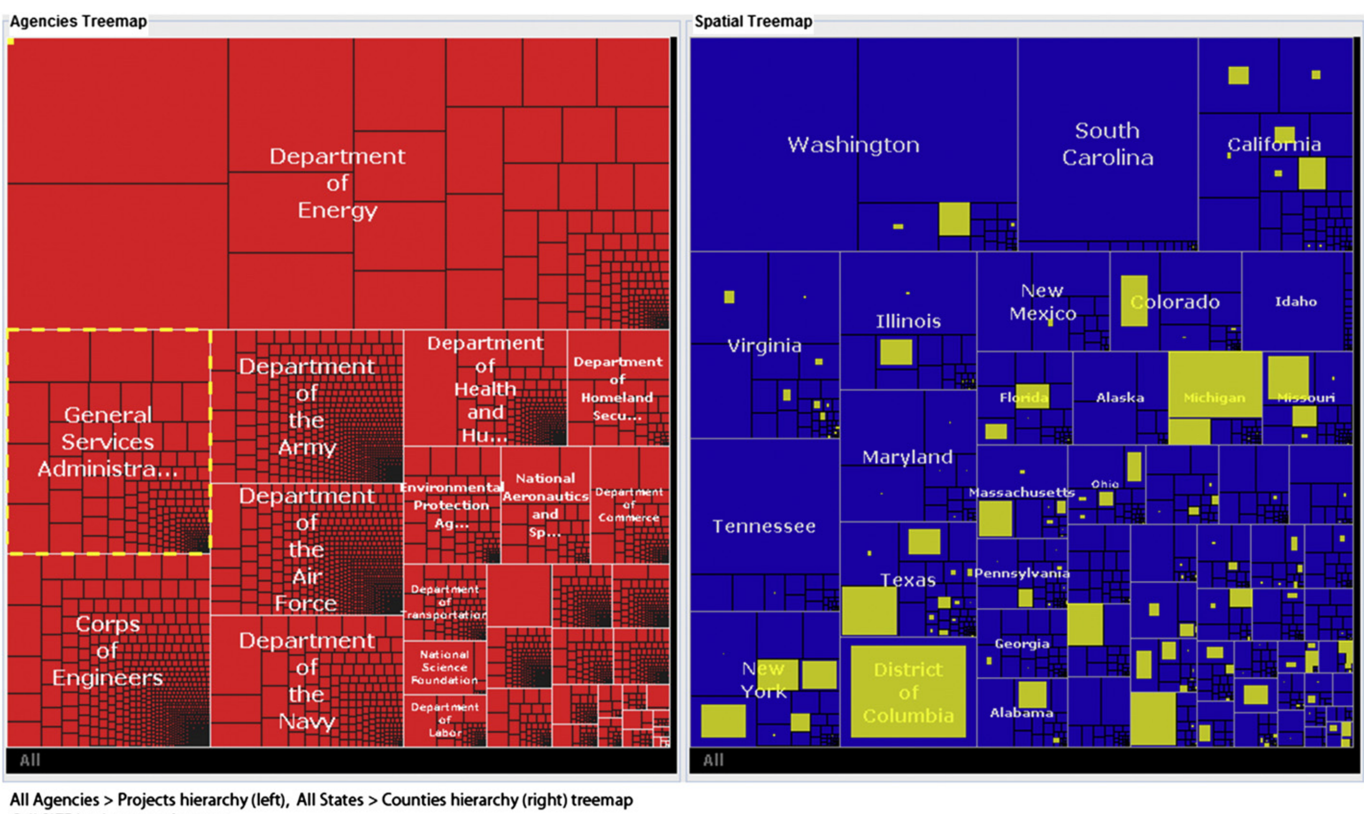

Fig. 8. Most money from the General Service Administration (selected in the left treemap) is given to the District of Columbia (right treemap).

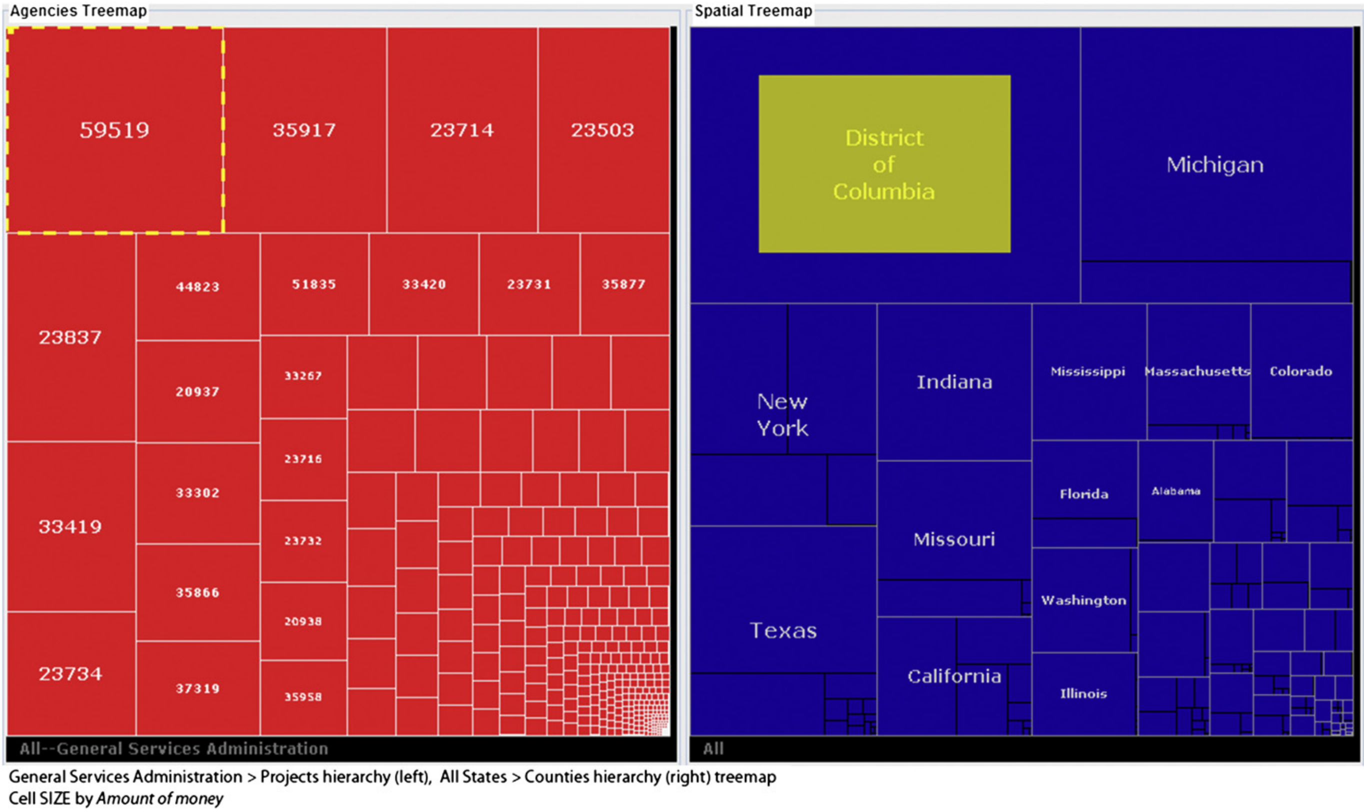

Fig. 9. The District of Columbia received the biggest projects from the GSA.

Computer Science department at University of Maryland. Each participant spent 10 min to learn the bi-hierarchical structure of Recovery Act data and the usage of TreeCovery. They then explored the data freely for 20 min, looking for interesting patterns and outliers. They were also asked to use think-aloud protocol. At the end of the evaluation, they gave us general remarks of the tool. Below we present a few usability issues and remarks.

5.2.1. The coordinated dual treemap interface proved to be usable

All the participants understood the dual treemap representation in a few seconds with brief explanation. They found the proportional highlighting very helpful to overview the hierarchical distribution on both axes. It also frequently raised further questions such as, "Why does Maryland receive a lot from the Department of Health and Human Services? Is it because of the National Institutes of Health?"

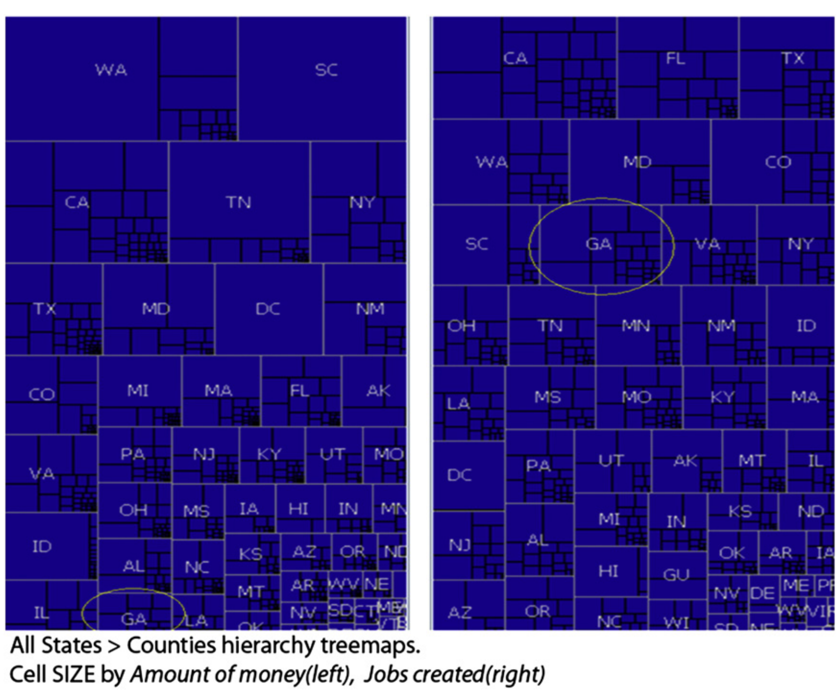

Fig. 10. Two treemaps with different settings are shown side-by-side in the Comparison window. Georgia (GA) stands out as it created more jobs with less money.

By the way, it was interesting to see that most participants tried to control the left treemap only for a short time. This is probably due to the familiarity the users have with layouts that have controls in the left and representations in the right side of the screen.

5.2.2. Some users get lost when the entire treemap is redrawn

During the evaluation participants often tried to apply dynamic filters on highlighted selections; however, as the current version of TreeCovery does not keep the selection when the entire treemap is redrawn, some participants got lost. Although we told them that the Shoebox can be used for comparing treemaps under multiple settings, they seem to need visual reference that relates the previous and the current status of the treemap.

5.2.3. Users need visual reference of the proportion of the current views to the entire data

Participants often wanted to know how big the size of the current treemap is within the entire data. One quick remedy can be a thumbnail window of the entire treemap with a smaller region representing the portion of current view.

5.3. Expert feedback

TreeCovery was presented to RATB and the recovery.gov development team at Synteractive.[11] Their feedback was very positive especially about the proportional highlight and coordinated exploration features. Both groups received the source code of TreeCovery, and are now evaluating the possibility of two practical applications. First, coordinated multiple treemaps can be incorporated into recovery.gov website by Synteractive as an exploration tool for citizens. Second, RATB can use TreeCovery application for their internal visual analytic tool and detect fraud, abuse, and waste.

6. Future work

While TreeCovery provides some innovative features and encompasses many exploration aids, it can, of course, be greatly improved. As observed during usability evaluation, we can make exploration easier by adding visual references of the size comparison between previous and current status. The shoebox feature could potentially allow more extensive comparison among saved treemaps if the saved views were more interactive. The next version of TreeCovery will allow the entire treemap to be saved and loaded for viewing, rather than just a screenshot. Future versions will also include support for data manipulation, including user-defined columns, in the manner of Spotfire. This will allow greater flexibility in the way users build treemaps. We would also like to incorporate an advanced color scheme, where the color gradient follows the distribution of the data, rather than staying linear. This will allow close but not identical values to be easily differentiable in color. These extra features will enhance the TreeCovery exploration experience, and hopefully lead to more insight generation.

Even though our application of coordinated dual treemaps was the visualization of the Recovery Act expenditures, the concept can be generalized to any multi-hierarchical dataset. One example is the representation of the money spent by U.S. government agencies in different industries. The hierarchies in this case could be agencies and projects in one treemap, while the category of industry (e.g. manufacturing, technology, etc.) including the companies in each category could be represented in another treemap.

The breadth of government agencies and topics of interest make coordinated dual treemaps a powerful tool for agency staffers, political analysts, journalists, and other interested citizens. The availability of numerous U.S. government datasets under the Open Government Directive has made powerful analysis and visualization tools even more valuable. While new users will have to learn dual treemap strategies, the payoffs in insights about patterns, clusters, gaps, and outliers warrant this additional effort.

7. Implications and recommendations

Our experience in implementing, showing, and evaluating TreeCovery demonstrates the capabilities of information visualization tools to enable policy makers, journalists, and citizen groups to conduct more effective explorations of policy-related data. TreeCovery is especially effective when there are dual hierarchies (e.g. geography and agency structures) and quantitative values (e.g. expenditures or jobs created). Users can find specific amounts for agencies and states/counties, compare spending, and see extreme and anomalous values, which might indicate data errors. Rapid exploration and visual displays also have the potential to enable users to discover fraud, abuse, or waste.

TreeCovery is just one tool, whose contribution is narrowly focused on a specific kind of data and set of queries. Other visual analytic tools will be needed for other kinds of data and other queries. However, the application of powerful visual analytic tools is changing the expectations for agency staffers and the informed public. Publication of data sets as databases or spreadsheets on sites such as data.gov becomes even more valuable when visual analytic tools are available to support exploration and discovery. However, the complexity of these tools and the tasks they support requires increased training which is designed to improve visual literacy and ensure proficient use of these novel tools. An important benefit is that visual analytic tools are especially effective in finding errors and missing values in large data sets. Detecting these flaws is especially important since critics are quick to use even a small number of erroneous data values in their attempts to discredit government programs.

8. Conclusion

The American Recovery and Reinvestment Act provided for a substantial sum of money, $787 billion, to be distributed with the goal of economic stimulus. Tracking that distribution involves a large, multi-attribute set that can be organized as a dual hierarchy of money flow and geographical allocation. Many visualizations of the stimulus data have already been developed, but none of them adequately portray this dual hierarchy or offer flexible exploration capabilities. Our tool, TreeCovery, uses coordinated treemaps to accomplish exactly that task. We use coordinated zooming and filtering and finely tuned highlighting to streamline exploration across the two hierarchies. The tool incorporates a number of other features to aid in customization and flexibility of the display. Insights that would be difficult or impossible to see with previously available tools become readily apparent when TreeCovery is used to visualize the data.

References

Burch, M., & Diehl, S. (2006). Trees in a treemap: Visualizing multiple hierarchies. *Proceedings of SPIE, 6060* (pp. 224–235).

Donker-Kuijer, M. W., de Jong, M., & Lentz, L. (2010). Usable guidelines for usable website? An analysis of five e-government heuristics. *Government Information Quarterly*, *27*, 254–263.

Donovan, R. (2009). The 5 worst cities for urban youth — ABC news. Retrived from http://abcnews.go.com/Politics/Politics/worst-cities-urbanyouth/story?id=9083935.

Fedorowicz, J., & Dias, M. (2010). A decade of design in digital government research. *Government Information Quarterly*, *27*, 1–8.

Fisherkeller, M. A., Friedman, J. H., & Tukey, J. W. (1988). Prim9: An interactive multidimensional data display and analysis system. *Proceedings of ACM Pacific' 1975* (pp. 140–145).

Fredrikson, A., North, C., Plaisant, C., & Shneiderman, B. (1999). Temporal, geographical and categorical aggregations viewed through coordinated displays: A case study with highway incident data. *Proceedings of the 1999 workshop on new paradigms*

[11] http://www.synteractive.com/.

in information visualization and manipulation in conjunction with the Eighth ACM international conference on information and knowledge management (pp. 26–34).
Glantz, A. (2009). Idaho gets four times more stimulus money in contracts than Louisiana — NAM. Retrived from http://news.newamericamedia.org.
Heer, J., Card, S. K., & Landay, J. A. (2005). Prefuse: A toolkit for interactive information visualization. *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, CHI '05* (pp. 421–430).
Jern, M., Rogstadius, J., & Astrom, T. (2009). Treemaps and choropleth maps applied to regional hierarchical statistical data. *2009 13th International Conference Information Visualisation Barcelona* (pp. 403–410).
Kules, B., Shneiderman, B., & Plaisant, C. (2003). Data exploration with paired hierarchical visualizations: Initial designs of PairTrees. *Proceedings of the 2003 annual national conference on digital government research*, *6*.
Mooney, A., Bolduan, K., & Hanna, J. (2009). Feds: False congressional districts removed from stimulus site — CNN.com. Retrived from http://www.cnn.com/2009/POLITICS/11/18/stimulus.district.errors/index.html.
North, C., & Shneiderman, B. (2000). Snap-together visualization: Can users construct and operate coordinated visualizations? *International Journal of Human Computer Studies*, *53*, 715–740.
Roberts, J. C., & Wright, M. (2006). Towards ubiquitous brushing for information visualization. *IEEE Proceedings of the Information Visualization (IV06)* (pp. 151–156).
Robertson, G., Cameron, K., Czerwinski, M., & Robbins, D. (2002). Polyarchy visualization: Visualizing multiple intersecting hierarchies. *Proceedings of the SIGCHI conference on human factors in computing systems: Changing our world, changing ourselves* (pp. 423–430).
Shneiderman, B. (1992). Tree visualization with treemaps: A 2-D space-filling approach. *ACM Transactions on Graphics*, *11*(1), 92–99.
Shrinivasan, Y., & van Wijk, J. (2009). Supporting exploration awareness in information visualization. *IEEE Computer Graphics and Applications*, *29*, 23–33.
Slingsby, A., Dykes, J., Wood, J., & Crooks, A. (2009). The role of layout and order in treemaps for showing spatial and temporal variation in house prices. *Abstract for the GeoViz Hamburg Workshop* Hamburg, Germany.
Verdegem, P., & Verleye, G. (2009). User-centered e-government in practice: A comprehensive model for measuring user satisfaction. *Government Information Quarterly*, *26*, 487–497.
Wood, J., & Dykes, J. (2008). From slice and dice to hierarchical cartograms: Spatial referencing of treemaps. *Proceedings of GIS Research UK* (pp. 1–8).
Wood, J., Slingsby, A., & Dykes, J. (2008). Using treemaps for variable selection in spatio-temporal visualization. *Information Visualization*, *7*(3).

Miguel Rios is a visualization scientist in Twitter Inc. There he builds tools to visualize and analyze Twitter's unique data sets and discovers insights that the company shares with the world. Before Twitter, Miguel worked as a research assistant in University of Maryland's Human Computer Interaction Lab where he also was a graduate student. Miguel's interest are large scale data visualization and information design.

Puneet Sharma holds a Masters Degree in Computer Science from University of Maryland College Park. His research interests include information visualization, Software Engineering and Computer Systems.

Tak Yeon Lee is a Ph.D. candidate in the area of Computer Science. He also holds a Master of Science in Industrial Design Engineering at Delft University of Technology, the Netherlands.

Rachel Schwartz graduated from the University of Maryland with an MSc in Computer Science in 2010. She is currently at Google.

Ben Shneiderman (http://www.cs.umd.edu/~ben) is a Professor in the Department of Computer Science and Founding Director (1983–2000) of the Human-Computer Interaction Laboratory (http://www.cs.umd.edu/hcil/) at the University of Maryland. He was elected as a Fellow of the Association for Computing (ACM) in 1997, a Fellow of the American Association for the Advancement of Science (AAAS) in 2001, and a Member of the National Academy of Engineering in 2010. He received the ACM SIGCHI Lifetime Achievement Award in 2001. He is the co-author with Catherine Plaisant of Designing the User Interface: Strategies for Effective Human-Computer Interaction (5th ed., 2010) http://www.awl.com/DTUI/. With Stu Card and Jock Mackinlay, he co-authored Readings in Information Visualization: Using Vision to Think (1999). His latest book, with Derek Hansen and Marc Smith, is Analyzing Social Media Networks with NodeXL (www.codeplex.com/nodexl, 2010).