Displaying a comprehensive data overview is one of the most challenging tasks in visual analytics. Methods to achieve this are, for the most part, dependent on the type of the data and the motivation for analysis. The Recovery Act report dataset includes two intrinsic hierarchies, both of which should be included in any overview display. Treemaps have proved very useful for displaying hierarchical data. We present a tool which synchronizes between two treemaps and implements a novel approach to highlighting corresponding nodes across the treemaps. Interesting explorations can be saved and compared in further analysis.

Keywords
Information visualization, treemap, polyarchy, multiple hierarchy, Recovery Act

1. INTRODUCTION
In February 2009, President Obama signed an economic stimulus package into law, dedicating $787 billion to create jobs and give a boost to the economy, with the provision that the distribution of the money would be completely transparent. 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 for 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. In addition, while many provide some interaction, most do not allow users the full control over the data visualized.

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, who 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 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. It accomplishes this through the use of two synchronized 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 synchronized as well as highlighting. In addition to the synchronized treemap design, we incorporated a few other features to improve exploration techniques. Besides the core recipient data available through recovery.gov, we included census data for each county. We found that many of insights found using this data also made use of geographical and demographic statistics, suggesting that their inclusion would be useful. We also developed a method to highlight inconsistent data, of which there was a significant amount in the recipient reports. Lastly, we added the abil-
ity to save snapshots of the current state of the treemap for later comparison.

The rest of the paper is organized as follows. In Section 2, we discuss related work. Section 3 provides our process of task analysis including a detailed illustration of Spotfire’s ability to support similar exploration to TreeCover. Section 4 explains implementation the tool in detail, while Section 5 offers some sample results. Section 6 contains our conclusion and future work.

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 (Figure 1). The maps can be zoomed in to state and zip code level and can be colored by a few different attributes. 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 are limited.

Many other websites offer similar tools to those of recovery.gov. The Federal Procurement Data System1, 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 2. Most of the site is devoted to text articles, but tabular and geographical visualizations of spending progress area offered as well. Other sites following the geographical/tabular trend include the Wall Street Journal [6] and msnbc (Figure 2)[7].

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. Polychology Visualization allows representation of intersecting hierarchies [11]. 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 [10]. However, they are presented as a graph theoretic rather than visual concept and, as such, are not particularly helpful in building our tool.

We found treemaps to be the most powerful approach toward visualization of multi-attribute, multi-hierarchical records like the stimulus data. Burch and Diehl introduce the “Trees in a Treemap” technique to represent trees with an associated taxonomy [8]. 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 are not well described besides for 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 [9]. 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.

One final treemap-based hierarchical representation is that of Wood et al. [12, 13, 14]. A specialized ordering is used to facilitate special and temporal locality so that the layout of nodes is more intuitive. The 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. TreeCover therefore uses two treemaps to represent the data. In this way, both hierarchies are transparently represented, and each can display its own attribute values. Cross-hierarchical exploration is supported through synchronization of the treemaps.

3. TASK ANALYSIS

To design our tool, we first needed to determine the chief goals of stimulus data visualization. As recipient reports of 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 analyze headlines of relevant news articles in

---

1 https://www.fpds.gov/
2 http://www.propublica.org/ion/stimulus
order to understand the process of journalists analyzing the data. Further, we analyzed the data with Spotfire \(^3\), one of the most versatile visual analytic tools. These exploratory tasks helped us to understand the data better and to develop the concept of our tool.

### 3.1 Analysis On News Headlines

Journalists are the most prominent group investigating Recovery Act data. From their work we can infer what people want to know and how the data can be analyzed. We collected headlines from the first two weeks of November 2009, just after the recipient reports had been publicized. We searched by ‘Recovery Act’ in the Google News search engine\(^4\) and focused on articles referring the recipient reports. While doing so, we found some interesting patterns of analytic tasks done on the Recovery Act report.

Most prominently, many finding focused on a State/County-wise 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” \([4]\) It is noteworthy that state/county-wise comparisons require aggregation of projects in each state or county.

The second pattern we perceived was 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 referred high-school graduation rates, infant mortality rates, unemployment rates, and juvenile justice incarceration to pick 5 worst cities for youth and compared numbers of jobs created by Recovery Act in those places\([1]\). Both of those patterns indicate that the Agency and Spatial hierarchies are equally important when analyzing Recovery Act data.

The third pattern we noticed 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\([3]\); however, other cases more complicated than simple mistakes can be identified. For instance, ‘Number of jobs created’, which is an important gauge of success, can be interpreted in different ways. Several news headlines pointed out projects that created too many or Too few jobs\([2]\)\([5]\). Usually, invalid values are either resolved or ignored in information visualization, although they have significant importance especially for a federal governmental website.

Summing up the findings above, we came up with the idea of bi-hierarchical data exploration in Figure 3. While Recovery Act funding is distributed to recipients along the Agency tree, the information of recipients is also aggregated by county and state in combination with census data.

### 3.2 Spotfire Tryout

Before designing our own tool, we tried to analyze the recipient reports data of contracts \(^5\) with an existing application. Spotfire was chosen because of its wide set of features dealing with multivariate data. There were four purposes of doing this analysis. First of all, analyzing the actual data gives deeper understandings about the data. Second, we wanted to know the capability and limitation of current visualization techniques on this specific data. Lastly, we could find opportunities of improving existing visualization techniques with new ideas.

Additional to the recipient report, We also incorporated State and County QuickFacts dataset\(^6\) from US Census Bureau into the recipient report data. The census data consists of wide variety of demographic profiles of each county including Population, Infant deaths, Housing unit, Household income and Unemployment rate. (For the full list, see Appendix A)

*The most effective job creators are questionable.* Knowing the fact that there are projects with unlikely high number of jobs created, we drew a treemap visualization (Figure 4) showing which departments or states are related to those projects. Colors of elements represent money per job which means how much money has been spent for creating a job in each project. Filtering by an attribute is crucial because small important outliers are easily be overlooked in treemap. We found that treemap visualization is suitable for the recipient report data.

\(^3\)http://spotfire.tibco.com/

\(^4\)http://news.google.com/

\(^5\)http://www.recovery.gov/FAQ/Pages/DownloadCenter.aspx

\(^6\)http://quickfacts.census.gov/qfd/download_data.html
A shortcoming of the highlighting technique is that it often misleads by coloring the entire area of each corresponding element no matter how much portion is actually related.

4. TreeCcovery
We designed TreeCcovery both to streamline the exploration process available through Spotfire and to offer additional features for data analysis. The chief component of the tool is, of course, the dual treemap display (Figure 6). 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 synchronize the displays. Essentially, filtering occurs simultaneously (Figure 7); zooming in on one treemap, which amounts to filtering on 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 synchronized automatically. In TreeCcovery, users can zoom in and out on both treemaps in any arbitrary order, and the recipient leaves included in the layout will remain coordinated.

TreeCcovery relies on a modified version of the recovery.gov dataset. 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 that are used for filtering, size-by, and color-by settings. All saved images are shown as thumbnails in the a shoebox area. Users can select and view any number of the saved images side by side in a separate window.

Besides the main treemap functionality, TreeCcovery provides additional features for data exploration. Firstly, details of the recipients that comprise each node are available when selected. A simple table displays the data for each recipient, including values for all attribute fields in our data set. This fairly basic view is enhanced by the highlighting of invalid data rows. TreeCcovery 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 a shoebox area. Users can select and view any number of the saved images side by side in a separate window.

TreeCcovery’s main innovation lies in its highlighting capability. This feature completes the synchronization 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 treemap that share any child recipients with the selected node. The square’s size is proportional to 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 an 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, as long as the two treemaps are sized by the same attribute. Users can change the 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.

In addition to the filtering capabilities provided through zooming, TreeCcovery also allows users to customize 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. Each treemap can be sized and colored on a chosen attribute as well. In this case, the attributes for each hierarchy are independent, although users must determine whether their chosen attributes pro-duce meaningful views. These capabilities mirror those offered by Spotfire.

Figure 7: Two treemaps share the identical data whose filter is controlled by zooming activity on either treemap.
Figure 4: Both treemaps show same data with different hierarchies. The size of each element represents amount of award money. In the agency treemap (left), recipients in light color create jobs with unlikely small money. In the state treemap (right), counties in light color are where those effective job creators are located.

Figure 5: In the scatter plot (right below) counties with higher (>20%) percentage of population over 65 (percentage) are selected. Corresponding counties and projects are also highlighted in both treemaps.
also used census information so we felt it would be very useful to integrate that information directly into the tool. Currently, TreeCovery only displays data for contracts, but it is readily extensible to include the full dataset.

5. INSIGHTS
To measure the utility of TreeCovery, It was run on high resolution and bigger display and recovery contract data were explored to find insights. The analysis was done only for few hours but generated few fruitful insights. More insights and issues in the tool are expected to come when it will be used by many users for longer time. Here is a list of 4 insights found in the small session of exploration task.

Washington State is one of the major recipients of money distributed by the Department of Energy (Figure 8). TreeCovery was run with the contract data and both treemaps were sized by the Award amount. Department of Energy received the highest amount of contract money which was easily visible in the Agency treemap. Department of Energy node was selected that quickly highlighted the related counties in the Spatial treemap. The highlighted area of each county shows the fraction of money given by the Department of Energy. Washington State received the highest contract amount, especially because of its two highly funded counties. So, highlighting feature enable users to perform the quick overview and motivates them to do further exploration.

DC State, the federal hub, received highest amount of contract money from General Service Administration Department (Figure 9). Both Treemaps were size by the Award Amount and General Service Administration Department (GSA) node was selected in Agencies treemap. Different counties got highlighted in the Spatial Treemap and as expected, DC had received the highest amount of contract money. GSA node was double clicked to perform zoom-in operation(Figure 10). One of the biggest projects of GSA was getting executed in DC and all of its recipients (Prime and Sub-Prime) are in DC.

Georgia State is creating more jobs while getting very lesser contract award amount (Figure 11). To see the utility of Shoebox in analysis, Spatial treemap was sized by the Award Amount and snap shot was taken. One more snapshot was taken after sizing Spatial treemap by Job Created. Both images were selected in the ShoeBox and right clicked to open the comparison window. Left TreeMap shows money distribution and right TreeMap shows Jobs Creation. Usually, States getting higher money are creating more jobs although it’s not proportional as various projects were not completed or half completed so the actual reported number of jobs will be different in the future. Georgia (GA) looked perfect outlier as more jobs were getting created with very lesser contract amount.
Figure 8: Washington State (Highlighted) received highest contract amount from the Department of Energy

Figure 9: DC State (Highlighted) received the highest contract amount from General Service Administration Department
6. CONCLUSION AND FUTURE WORK

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 synchronized treemaps to accomplish exactly that task. We use coordinated zooming and filtering and finely tuned highlighting to streamline exploration across the two hierarchies. The two incorporates a number of other features to aid in customization and flexibility of the display. Insights that would difficult or impossible to see with previously available tools become readily apparent when TreeCovery is used to visualize the data.

While TreeCovery provides some innovative features and encompasses many exploration aids, it can, of course, be greatly improved. The shoebox feature could potentially allow more extensive comparison among saved treemaps if the saved views were more interactive. The future version of TreeCovery will allow entire treemaps to be saved and loaded for viewing, rather than just a screenshot. Future versions will also include support for data manipulation, including for 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.

7. ACKNOWLEDGMENTS
This project was We would like to thank people of Synteractive.com who provided Recovery Act data, and Dr. Shneiderman for constant regards and helps.

8. REFERENCES
APPENDIX
A. CENSUS DATA
downloaded from http://quickfacts.census.gov/qfd/download_data.html
Population
Population over 65
Female Percentage
White people percentage
Black people percentage
Hispanic percentage
Infant deaths
High school graduate percentage
Bachelor degree percentage
Housing unit
Housing unit percent change
Median household income
People in poverty
labor force
unemployment rate
number of firms
women-owned firms percentage