Traffic Trender: Exploring Bottlenecks at a Systematic Level

Chris Musialek  
musialek@umd.edu

Chenglin Chen  
chenglin@cs.umd.edu

Hua He  
huah@cs.umd.edu

Sheng Zha  
szha@cs.umd.edu

Richard Johnson  
rjohns8g@cs.umd.edu

ABSTRACT
The quantity of well-structured traffic incident data collected by government organizations continues to expand at an alarming rate. While the number of transportation information visualization applications has grown over the past several years, there remain few infovis applications dedicated to helping road planners analyze trends of traffic data over long periods of time. In this paper, we present Traffic Trender, a systematic level traffic bottleneck viewer that combines a zoomable Treemap, a set of line charts, and a rich set of filters to uniquely assist planners in identifying insights. In particular, we attempt to enhance the ability of these analysts to find trends in traffic incidents per geographic region and over time, and to assist them easily identify outliers and anomalies. We then reflect on how our design makes progress towards these goals by providing results of a usability study which identify continued problems and areas for improvement.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous; H.5.2 [Information Interfaces and Presentation]: User Interfaces

Keywords
Information visualization, transportation, traffic trends, usability studies, exploratory data analysis

1. INTRODUCTION AND MOTIVATION
Road planners are regularly faced with the difficult task of making important determinations as to the prioritization and allocation of limited transportation funds for road redesign, paving, and improvements. Many of these types of projects incur substantial financial costs. Because state and local transportation agencies continue to be stretched for budget dollars, more and more are turning to data analysis and visualization to identify the worst areas of congestion. In particular, planners are interested in whether certain geographic regions are getting better or worse over time, and whether there are temporal/seasonal patterns in bottlenecks. They also struggle to find ways to visualize outliers, anomalies, and traffic spikes across long periods of time.

Luckily, the quantity of well-structured data capturing real-time traffic conditions is also growing exceptionally fast. The emergence of GPS in many new models of vehicles has provided unprecedented accuracy and granularity of real-time traffic conditions. While many good examples of traffic incident tools are beginning to emerge to help visualize real-time traffic conditions, few if any are devoted to visualizing these data over time. We hypothesize that planners could benefit significantly from taking a full system level approach to viewing these data.

In this paper, we present Traffic Trender, a systematic level traffic bottleneck viewer that combines a zoomable Treemap, a set of time-based line charts, and a rich set of filters to uniquely assist planners with these particular problems. By allowing the planner to take a holistic view of the traffic bottleneck system, we provide a way for them to be able to more easily answer important questions such as:

- Are certain geographic regions getting better or worse over time?
- Are there temporal and/or seasonal patterns in bottlenecks, and how can one easily find outliers, anomalies, and abnormal patterns?

Our application also provides an easy way to compare locations, minimizing cognitive load, and at the same time preserves flexibility for the user to choose the data metrics they would like to compare (described below).
2. DATA

Our system consumes post-processed probe data from moving vehicles of traffic speeds at specific locations across six states for a period of two years. This raw data was processed to calculate individual traffic bottlenecks and produce a couple of key data points associated with it. A bottleneck is simply defined as a 60% or more reduction in the normal speed of a particular location for a period of 5 minutes or more. Additional data points are also calculated from each bottleneck. They are maximum length of the bottleneck in miles from the origin location, the bottleneck duration in minutes, and the number of occurrences that a particular event occurred. From this, another derived metric is created, called impact factor, defined below.

\[
\text{Impact Factor} = (\text{Maximum Length} \times \text{bottleneck duration} \times \text{number of occurrences})
\]

With these metrics, we averaged their values per month per location. In total, we produced a total of 1651 locations across the six states with these statistics per month over two years.

3. DESIGN MOTIVATION

There are several reasons why we chose a Treemap to be the focal point of our application.

The Treemap, first developed by Ben Shneiderman [17], provides a simple yet powerful way for comparing nodes enclosed in a hierarchy. By design it recursively subdivides area into rectangles. This fits the traffic location bottleneck data very well because every location is contained within a larger geographic boundary (a location is part of a road, which is part of a county, which is part of a state). Utilizing the three associated data metrics (Section 2) and applying them to the Treemap, we felt it would provide a great way for planners to identify outliers and anomalies easily, as well as view these sets of statistics in a compact, holistic way, something that the Treemap also does very well.

We took inspiration from several other Treemap implementations, such as Smart Money’s “Map of the Market” [2], combining the breadth view we desired with an interactive capability to zoom and filter as needed for in-depth data inspection. This also seemed to be a natural fit for our data and use case, as it provided interesting ways that planners could zoom and explore for interesting trends. We combined this with a checkbox location filter to remove specific roads or counties that may be skewing system-level comparative results, a key challenge due to the loss of fidelity in working with aggregated data averages. [4]

4. RELATED WORK

The past few years have seen an increasing number of applications devoted to visualizing the vastly growing amount of traffic sensor data. The CATT Lab at the University of Maryland, College Park has developed a suite of tools to identify major bottlenecks, report on travel time reliability, and display other congestion conditions and traffic incidents [1]. Their congestion scan tool provides a useful heatmap allowing the exploration of the rise and fall of congested conditions on a stretch of road. Fervor provides four integrated visualizations for exploring traffic data, including interactive maps, histograms, two-dimensional plots, and parallel coordinates [12].

However, most of these tools focus on visualizing real-time data, and do not provide features allowing the visualization of trends over an extended period of time. At this time, we are currently only aware of work being done at the Federal Highway Administration called the Highway Performance Monitoring System (HPMS), although we were not able to view their work at the time of this paper.

5. TRAFFIC TRENDER DESIGN

The Traffic Trender is designed to explore trends in traffic bottlenecks over time based on different zoom levels. Our tool consists of four major components, namely Treemap, line charts, filter control panel and details-on-demand display, as show in the figure.

Since our visualization is web-based, we implement the backend and the frontend [7], each with different techniques. Our backend is implemented with the MySQL database and Java programming language, and our frontend interface is implemented with D3/JavaScript. The reason we make our visualization tool web based is that we would like to make it platform independent, and make it easier to use. We host our visualization tool on Amazon cloud services 1, and users only need to type in our http address 2 in their web browser to start using it.

We will describe the design features of our four major components in the follow subsections.

5.1 TreeMap

The Treemap visualization [3] is not only useful for analyzing the correlation among three important attributes (impact factor, duration and maximum length) of our traffic data but also very powerful for refining and selecting the needed traffic data after filtering. Our content-based Treemap visualization [15] is two-dimensional, and rectangles in our Treemap visualization represents individual locations in one road, and all rectangles are grouped together according to their geographic information, and they are firstly grouped based on which road they are located, then the county and the state. In Figure 2, we show our Treemap in the county level under the state of Maryland, all counties within Maryland are shown and rectangles representing the specific locations within Maryland are grouped together based on their county information.

Our Treemap visualization can be easily customized by users, and we provide three options for users to specify the size and the color attribute of each rectangle in the Treemap [14]. For example, if a user want to understand the correlation between the maximum length and the maximum duration of our traffic bottleneck data, they can select the size of the rectangles to be the length and the color to be the duration in the Treemap control panel.

By providing such functionality for users, they can easily explore their work at the time of this paper.

1URL: http://aws.amazon.com/ec2/
2URL: http://www.traffic-trender.com
We also provide the functionality of changing the specific data selection range and the zoom level for users in our Treemap. Due to the reason that our dataset is hierarchical as described in the previous section, we have four zoom levels in our Treemap accordingly. All four levels (state, county, road and individual location) are shown in Figure 1, 2, 3, 4. By default, our Treemap is on the state level (as shown in Figure 1), where all states are shown. By just a simple mouse clicking, our users can navigate into a certain state if they click on a specific state rectangle. The same applies when the Treemap is on the county level (Figure 2) and they can then zoom into road level (Figure 3) and so on. Whenever the user changes the zoom level, the Treemap automatically updates itself so that users are able to focus specifically on the requested locations.

The ability to change the zoom level of our dataset enables users to visually emphasize the traffic bottleneck locations that are within a specific range, which could be inside a state, or county and others. In addition, such easiness of changing zoom levels by simple mouse clicking can greatly improve the effectiveness of our Treemap visualization.

5.2 Line Chart

Our second major component is the line charts. We provide two line charts for users to focus on specific locations or aggregated information among individual locations, like Treemap both of our line charts can be customized as well based on the three attributes of our data, namely impact factor, maximum length and duration. The following sections describe the two line charts.

5.2.1 Top-10 Location Line Chart

Based on current zoom level of the Treemap, and users’ selections in the filter control panel, our top-10 location line chart can automatically display, as the name suggests, the top individual locations in the selected range over time by selecting the traffic bottleneck locations and ranking them in order. The x-axis shows the time over the last two years [11], and the y-axis shows one of the three user-specified attributes (impact factor, maximum length or duration) based on which the line chart do the ranking [5].

The top-10 location line chart is generated based on current zoom level, current filter selections and the user-specified attributes. We first of all find all location within the selected range, and then rank the locations based on one of the three user-specified attributes. Then we select the top 10 of these locations on the line chart.

We designed our line chart so that it simply follows the selection of Treemap and users do not need to make extra movements in order to use this visualization. For example, when users click on the Treemap and zoom from state level into county level, our line chart will automatically change its scope and recalculate the new top-10 locations. We use different colors to represent each individual lines, and if users are interested in details about each location, they can move...
Figure 2: County Level under State MD

Figure 3: Road Level

Figure 4: Location Level
their mouse to specific lines, details of each road name are shown in the upper side of this visualization.

5.2.2 Aggregated Line Chart
The aggregated line chart is provided to answer questions, such as what is the difference over time between states MD, VA and DE, or the difference between Howard and Prince George counties. The aggregated line chart is based on the zoom level, and show only the aggregated lower level aggregated information, for example if the zoom level is at the state level and it is under state MD, then only the counties under MD will be shown in the aggregated line chart where each line represents a county. For another example, if the current zoom level is at the county level, then only the aggregated road information is shown where each line represents a road (not individual locations). The difference between this visualization and our top-10 location line chart is that, the top-10 location chart always displays each individual locations while this visualization can provide higher level information.

Besides the above difference, all other visual features stay exactly the same as the top-10 location line chart, this line chart follows the zoom level selection of Treemap and users do not need to make extra movements to use it, and individual lines are represented in different colors, details can be retrieved once users move their mouse to specific lines.

5.3 Filter Control Panel
Our third major component is the filter control panel. As the name suggests, it can help the users to select what they want and display them in the Treemap and line chart. We organize the filter content in a hierarchical tree [9], the structure starts from the state, then county, and finally down to road. This list is inclusive, users can make selections and click the “apply” button, the Treemap and line chart visualizations will change accordingly.

The filter selections along with the zoom level selection are the two data manipulation mechanisms that we provide for users to explore our traffic bottleneck data set.

5.4 Details On Demand Display
Our final major component is a details on demand display. Whenever a user selects a location in the line chart, or move over the rectangles in the Treemap, the detail on demand display on the right of our interface will show the specific information based on users’ selections.

This display can be shown in Figure 5 as an example. Our detail on demand display has two components the top part will show the basic property information of the current selection (either a Treemap or a linechart location). If the selection is the location in the line chart, the bottom component will show all the values (based on one of the three metrics that user selected) over the past two years.

6. EVALUATION
In order to assess the effectiveness of our Traffic Data analysis tool, we designed and performed an experimental user study. The evaluation scenario best matching our purposes is the User Experience (UE) test case [10]. The goal of our experiment [19] was:

1. to observe the users’ reaction to the visualization and interactive features offered by Traffic Trender
2. to indentify the strengths and weaknesses of Traffic Trender
3. to get suggestions for future improvements.

6.1 Experiment Design
6.1.1 subject
Our test population consisted of two graduate students in Computer Science (cs) and six domain experts from the center for advanced transportation technology (CATT). Both groups used Traffic Trender to perform tasks on the same traffic data set provided by Michael Pack, Director of CATT Lab.

6.1.2 Experimental Setup and Procedure
The testing protocol was as the following:

1. First, the user received a 2-minute general introduction of our software and the dataset, followed by a 5-minute demo of the software. The user then asked any question he or she might have.
2. The user was asked to perform 6 tasks within 15 minutes. While interacting with the system, the user was encouraged to think out loud [16]. The 6 timed tasks were shown in Table 1.
3. After finishing the tasks, the user then filled out a post experiment questionnaire [6] with 10 closed-ended questions (shown in Table 2) and 4 open-ended questions (shown in Table 3). The first 10 questions were on a 9-point Likert scale with 9 being positive. These questions were about the learnability, interface design, navigation, responsiveness and the usefulness of the features. The 4 open-ended questions were used to identify useful features that Traffic Trender had or missed, and features that need to be reworked.

The entire experiment was recorded in both screen capture video and audio.

<table>
<thead>
<tr>
<th>Table 1: Six Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: Use the TreeMap visualization to represent size with Impact Factor, represent color with Duration, show all individual locations. Which state has the worst bottleneck situation (i.e. the highest Impact Factor) during the year 2010 and 2011?</td>
</tr>
<tr>
<td>T2: Use the TreeMap visualization to represent size with Duration, represent color with Impact Factor, zoom in the TreeMap and show locations only in MD. Which location in MD has the worst bottleneck situation (i.e. the highest impact factor) during the year 2010 and 2011?</td>
</tr>
<tr>
<td>T3: Use the TreeMap visualization to represent size with Maximum Length, represent color with Duration, use the filter panel to show locations only in State ‘SC’ and County ‘GREENVILLE’. Which location in GREENVILLE, SC has the bottleneck with the longest duration during the year 2010 and 2011?</td>
</tr>
<tr>
<td>T4: Use the Top 10 Locations LineChart visualization to show the top 10 locations based on Impact Factor for all locations in State ‘VA’. Among these top 10 locations, which location has the bottleneck with the highest impact factor in Aug. 2011?</td>
</tr>
<tr>
<td>T5: Use the Top 10 Locations LineChart visualization to show the top 10 locations based on Duration for all locations inside ‘Richmond , VA’ and with Road_name ‘I-195’. Among these top 10 locations, which location has the bottleneck with the longest duration between Apr. 2011 and Dec. 2011?</td>
</tr>
<tr>
<td>T6: Use the TreeMap to zoom in to ‘Frederick, MD’, set the y-axis of the Aggregated Boundaries LineChart to be Impact Factor. Between Jan. 2010 and Dec. 2011, are certain roads getting better or worse over time?</td>
</tr>
</tbody>
</table>

### 6.2 Results

Overall, all 8 participants appeared to perform well and finished all 6 tasks within 10 minutes.

**Task1** seemed straightforward to all participants. 7 out of 8 participants answered the question correctly. The 4th participant (p4) forgot to click the apply button after changing the setting of size for Treemap, so he got a wrong answer. p2 was confused about the meaning of the color of Treemap (i.e. which color represented high value? [18]). And p7 suggested the ability to choose other color to present the data because “red-green” doesn’t work well with people with vision disability.

For **Task2**, 7 out of 8 participants answered the question correctly. p2 forgot to click the apply button after changing the setting for Treemap from “duration for color” to “impact factor for color”, and this caused the wrong answer. p6 was confused about the hierarchal structure of the Treemap at first, but he figure it out after a while. And 4 participants said a “details on demand” feature for the Treemap would be really helpful.

For **Task3**, all 8 participants got the correct answer but 4 participants seemed to having trouble using the filter menu to unselect efficiently. And p6 said it is hard to tell the difference between two shapes of red color.

For **Task4**, all 8 participants answered the question correctly but they all noticed the problem of crowded labels in the line chart and 4 participants complained about the lack of zoom-in for the x-axis.

For **Task5**, all 8 participants got the correct answer. 6 participants used the filter menu to filter out most of the unwanted locations, while the other 2 used the Treemap to zoom in to the desired region.

For **Task6**, all 8 participants answered the question correctly. They seemed to have no problem switching between the two line charts. And p5 suggested a “search” function on the control panel.

### 6.3 Feedback based on Close-Ended Questionnaire

We show the results on the close-ended questions below [13]:

<table>
<thead>
<tr>
<th>Table 2: 10 Close-Ended Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: How do you rate your overall performance on the tasks given?</td>
</tr>
<tr>
<td>Q2: Did you find the interface simple?</td>
</tr>
<tr>
<td>Q3: Do you think the interface is easy to navigate?</td>
</tr>
<tr>
<td>Q4: Do you think the interface responds quickly?</td>
</tr>
<tr>
<td>Q5: How was your experience with the Screen resolution?</td>
</tr>
<tr>
<td>Q6: How was your experience with the Display colors of the visual display?</td>
</tr>
<tr>
<td>Q7: How was your experience with the Scrolling and Other Feedback of the visual display?</td>
</tr>
<tr>
<td>Q8: Did you find the TreeMap Visualization helpful?</td>
</tr>
<tr>
<td>Q9: Did you find the two LineChart Visualizations helpful?</td>
</tr>
<tr>
<td>Q10: Did you find the Filtering panel helpful?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3: 4 Open-Ended Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>For <strong>task2</strong>, 7 out of 8 participants answered the question correctly. p2 forgot to click the apply button after changing the setting for Treemap from “duration for color” to “impact factor for color”, and this caused the wrong answer. p6 was confused about the hierarchal structure of the Treemap at first, but he figure it out after a while. And 4 participants said a “details on demand” feature for the Treemap would be really helpful.</td>
</tr>
<tr>
<td>For <strong>task3</strong>, all 8 participants got the correct answer but 4 participants seemed to having trouble using the filter menu to unselect efficiently. And p6 said it is hard to tell the difference between two shapes of red color.</td>
</tr>
<tr>
<td>For <strong>task4</strong>, all 8 participants answered the question correctly but they all noticed the problem of crowded labels in the line chart and 4 participants complained about the lack of zoom-in for the x-axis.</td>
</tr>
<tr>
<td>For <strong>task5</strong>, all 8 participants got the correct answer. 6 participants used the filter menu to filter out most of the unwanted locations, while the other 2 used the Treemap to zoom in to the desired region.</td>
</tr>
<tr>
<td>For <strong>task6</strong>, all 8 participants answered the question correctly. They seemed to have no problem switching between the two line charts. And p5 suggested a “search” function on the control panel.</td>
</tr>
</tbody>
</table>
For **Q2** (UI), the average score is 6.88 out of 9, with a standard deviation of 1.46. Two Traffic domain experts didn’t find the interface simple.

For **Q3** (Navigation), the average score is 7.50 out of 9, with a standard deviation of 1.31.

For **Q4** (Responsiveness), the average score is 7.00 out of 9, with a standard deviation of 1.60.

For **Q5** (Screen resolution), the average score is 7.12 out of 9, with a standard deviation of 1.73. P1 wanted a larger display of the visualization.

For **Q6** (Color), the average score is 7.38 out of 9, with a standard deviation of 0.92. All participants except p5 seemed to like the colorful display, especially the Treemap visualization.

For **Q7** (Scrolling), the average score is 8.14 out of 9, with a standard deviation of 0.90. This shows the Scrolling and other Feedback of the visual display were good.

For **Q8** (Treemap helpfulness), the average score is 8.75 out of 9, with a standard deviation of 0.46. All participants liked the Treemap visualization very much.

For **Q9** (Line chart helpfulness), the average score is 7.13 out of 9, with a standard deviation of 1.25. P1 complained that the line chart was “Too busy” and 5 other participants suggested the “details on demand” feature to us.

For **Q10** (Filter menu helpfulness), the average score is 7.38 out of 9, with a standard deviation of 0.92. All other area in the Treemap will turn grey. Now if users click a line in the line chart, the corresponding block in the Treemap will be highlighted and its thickness increases a little so it stands out.

### 6.4 Feedback based on Open-Ended Questions

From the 4 open-ended questions, we got the following feedback:

For **Q1** (Tool understandable), all 8 participants answered “yes”. P4 said “could probably teach myself with a bit more time.” P7 said “yes, but in the long run it would benefit from some help icons to explain the data and how Treemaps work. This is a pretty standard feature for most visualization tools.”

For **Q2** (most useful feature), all 8 participants answered “Treemap” as the most useful. P1 liked the ability to change Treemap primary/secondary attribute, and P1 also liked the Time graphs (Line Chart). P7 liked the “Zooming in on Treemap to trouble spots” feature very much.

For **Q3** (missing features), “details on demand” was mentioned 4 times, “zooming on line graphs” was mentioned 3 times, “legends for the Treemap and Line Charts” was mentioned 3 times, “un-clicking all for the filter menu” was mentioned 3 times. There were other answers such as “explanation of impact factor”, “explanation of Treemap color”, “help icon”, “showing the direction of the road”, “date filtering”, “search function” and “one more level in the Treemap - Region”.

For **Q4** (features need to be reworked), all participants mentioned “the label of the lines in Line Chart”, 4 participants suggested the “apply” button. Other things suggested were “option to order Treemap alphabetically”, “cross ref locations to a map”, “give the order of the Top 10 locations in the line chart”, “change the name of the Aggregated Boundaries to something else”.

### 6.5 Key Evaluation Insights and Changes

All of the participants in the user evaluation study reacted favorably to the overall performance of Traffic Trender and provided valuable feedbacks. Some of the suggestions were within the scope of our tool development and we have made changes after getting the feedbacks from the first round of usability tests. For example:

1. All participants mentioned the labels of the lines in both line charts were hard to read so we got rid of the labels. Instead we added the “Details on demand” feature in the line chart to provide information.

2. One participant complained that the line charts were too busy and it is difficult to tell the lines apart. We added a “highlight” function for that. Now when the mouse hovers over a line, that line is highlighted and its thickness increases a little so it stands out.

3. Originally there were three “apply” buttons on the UI which confused some participants during the usability test. We changed that. Now the “apply” buttons for both Treemap and line chart are gone and the selectors for these two sections work automatically. Users only need to click “apply” if they change the filter menu settings.

4. Cross reference from line chart to Treemap was added. Now if users click a line in the line chart, the corresponding block in the Treemap will be highlighted and all other area in the Treemap will turn grey.

Some other features are not going to be implemented in this project, but are worth exploring in the future:

1. By integrating other datasets in our visualization, we can answer the following question: “Are there other factors that might be affecting the bottlenecks? (weather, incident, construction, etc.)” This could be really helpful.

2. Date filtering for the data, and the ability to zoom in the x-axis of the line chart.

3. A “search” function.

4. A “help” menu to provide information on both the data and the visualization tools.

5. The ability to use other combinations of color (not only red-white-green) to represent the data in the Treemap.
REFERENCES


