Application Project
CATT Traffic Data Visualization and Analysis

Group Member
Milad Gholami: milad621@gmail.com
Peixin Gao: gaopei@umd.edu

Introduction
In this project we used Spotfire to visualize and analyze data provided by Michael VanDaniker, the Visualization Programs Manager at Center for Advanced portdfation Technology (CATT) Laboratory, University of Maryland.

Data Source
We choose the data from CATT for two reasons. The first one is that we are now living in the Washington DC Metropolitan area and we have been suffering from traffic jams, which makes us curious about what the traffic pattern is in this area, what are reasons that cause such problem, and how the traffic events or incidents change along time, in both hours and in days. The second reason is that We could achieve CATT’s data via RITIS (Regional Integrated Transportation Information System). And the data provided by Michael VanDaniker from CATT contains information about accidents, construction events, disabled vehicles, and so on for multiple states, including Maryland. With all the traffic information, we could find a lot of useful results.

Software Platform
We select Spotfire as the software platform to conduct the information visualization because it is easy to learn and use. Besides, among the demos on Spotfire website, we found one project called “Arlington Virginia County Crime Analysis”. In the demo, they used Spotfire to plot the map of Arlington county with information about how often each type of crime has occurred and where, and illustrated the counts of different type of crimes. They also showed the temporal and timestamp analysis, as well as district breakdown. We think that with data from CATT, we could do similar analysis and visualization, and find useful insights.

Data Sets
From RITIS, we could get traffic incident data from 14 regions for a maximum of 3 months span.
In our project, we exported all traffic event data for Maryland for the 3 months span from July 10th to October 10th of 2013. This data contains 27,684 traffic events of 18 unique incident types: Alert, Animal Struck, Collision, Congestion, Disabled Vehicle, Emergency Roadwork, Fatalities Involved, Flood, Incident, Injuries Involved, Obstructions, Road Maintenance Operations, Special Event, Strong Winds, Traffic Congestion, Traffic Signal Not Working, Utility Work and Vehicle on Fire. Each incident in the data contains a variety of metadata which can be used for analysis.

Data Processing

Before applying the data for visualization, we first processed the data to extract useful information. The original data we got from RITIS has location of event as an address. However, we are focused on the main roads so we extracted the Highway, Interstates, or important roads from the address by a python script. Also the original data had duration of the event as strings like: “Less than a minute”, “1 day 2 hours 24 minutes”, “1 hour 1 minute”, etc. We replaced this with the amount of duration in minutes.

When analyzing the data, we notice that several event types not only have a small amount of occurrences but also have a quite short duration, such as Animal Struck, Congestion, Emergency road work, Fatalities involved, Flood, etc. So we considered them as not important and filtered them out in most of the charts in this report.

Findings and Insights

1. The Whole Picture

The first figure shows the distribution of these traffic events and their total duration. The length of duration is reflected by the darkness of the bar.
From Figure 1 we can notice that the most frequently happened events are Disabled Vehicle and Road Maintenance, however Road Maintenance has a much longer duration than the Disabled Vehicle.

In order to verify our idea of differences in counts and durations, we make two more figures.
From Figure 2 and Figure 3, we can find the same result, that the distribution of events’ total durations and counts are very different, and we think the total duration information is more important that the count of the event. Thus in the following analysis, we use total duration of an event as a metric of its influence on the traffic.

2. Variations Within a Day

Here we try to analyze the events during different hours of a day. In the next figure, we can see the distribution of different events per day.
We can see the two peaks for Road Maintenance Operations and Disabled Vehicles which both are around the traffic peak hours (rush hours). One fact we can see here is that (we guess), we see many Road Maintenance Operations starting in the beginning of the day and it keeps their staff busy so they are less probable to start Road Maintenance Operations during the day. Also, we will see how different roads have different patterns later.

Figure 5 shows the duration of the events per hour of the day.
As we mentioned before, seems like heavy Road Maintenance Operations will start only at the beginning of the day or beginning of the night.

3. Variation Within a Week

Next figure shows the distribution of different events over days of a week.
As we could expect, everything gets lower during the weekends. We seem to have more disabled vehicle events on Monday, it gets lower the next day and then it starts increasing again till Thursday. From this chart we get the insight that it is safer and faster when driving on weekends.

In the next figure, we can see the duration of the events per day.

An interesting fact we can see here is that, it takes longer for maintenance event happened on Monday compared to other days. Seems like there is an effort to finish road maintenance works before the weekend. Also, we start to see more traffic signal not working events closer to the weekends.

4. Patterns in Major Roads
There are a lot of roads in the data. So we filter out roads with less events. We only compare I-95, I-70, I-695, US 50, MD 295, I-895, I-97, MD 200, and I-270. Here is the event distribution over these roads:
And in the next figure we can see the duration of these events.

We can see that on some roads, events have much less duration. Also they have the same number of events compared to some other roads. As an example, I-70 much more number of events than I-895; however, they both have similar duration time which is aggregated. We guess
this means that maintenance and road support for I-70 are faster and more efficient than the ones for I-895.

**Evaluation about Spotfire**

Spotfire is a powerful tool to visualize the data and it inspires us to find more insights via the visualizations. The data we have is high-dimensional, with Spotfire, we can find the correlations and differences among the dimensions and visualize the relations. And the visualizations generated by Spotfire are aesthetic, which is another advantage of Spotfire. Besides, it is not very difficult to learn the tool, which allows us to work on the application project efficiently.

But the tool is not perfect, as far as we see, in the following aspects.

First of all, the map chart is really hard to use. We wanted to import a map into Spotfire and show visualization information on the map, just as in the demo called “Arlington Virginia County Crime Analysis” that Spotfire provides. However, we could only import the shapefile into Spotfire but could not visualize our data on the map, and there is no instruction on how to make it. Finally we gave up the idea after many times of failures.

Besides, we found a problem when working on the project, that within a Spotfire file, all charts, even all pages, share the same data set and the same filter setting, which can be very inconvenient in many cases. For example in a single page, it is hardly possible to compare the difference on traffic events distribution between two interstates. And if we want to make charts using different filter settings, we have to open more than one Spotfire files. This should be the most imperfect place on Spotfire that we have found via using it. We admit that we may be ignorant on the some functions in Spotfire that can do the work, but we argue that such function should be more obvious to find.

At last, we are hoping that there is a tutorial about advanced techniques on using Spotfire, which could guide us to make better use of the software.

**References**

Regional Integrated Transportation Information System (RITIS) ([https://www.ritis.org/](https://www.ritis.org/))