ICE--Visual Analytics for Transportation Incident Datasets

Michael L. Pack, Krist Wongsuphasawat, Michael VanDaniker, and Darya Filippova
University of Maryland, College Park
Center for Advanced Transportation Technology Laboratory
College Park, MD 20742
Email: PackML@umd.edu

Abstract

Transportation systems are being monitored at an unprecedented scope resulting in tremendously detailed traffic and incident databases. While the transportation community emphasizes developing standards for storing this incident data, little effort has been made to design appropriate visual analytics tools to explore the data, extract meaningful knowledge, and represent results. Analyzing these large multivariate geospatial datasets is a non-trivial task.

A novel, web-based, visual analytics tool called ICE (Incident Cluster Explorer) is proposed as an application that affords sophisticated yet user-friendly analysis of transportation incident datasets. Interactive maps, histograms, two-dimensional plots and parallel coordinates plots are four visualizations that are integrated together to allow users to simultaneously interact with and see relationships between multiple visualizations. Accompanied by a rich set of filters, users can create custom conditions to filter data and focus on a smaller dataset. Due to the multivariate nature of the data, a rank-by-feature framework has been expanded to quantify the strength of relationships between the different fields.

Keywords: Information Visualization; Visual Analytics, Data Mining, Knowledge Discovery

1. Introduction

Traffic Management Centers (TMCs) throughout the country generate detailed logs of hundreds of traffic incidents each day. These logs include data about the time and location of an incident, number of vehicles involved, lane closures, weather and road conditions, incident severity, etc. While the transportation community emphasizes developing standards for storing incident data, little effort has been made to design appropriate visual analytics tools to analyze the data, extract meaningful knowledge, and represent results.

Exploring these data-rich logs to glean any significant meaning can be an overwhelming task. Incident data is inherently difficult to explore due to its multivariate nature. Inferring the causal relationships of and trends within incident data can be problematic without using statistical methods. Many current tools are geared towards either geospatial analysis only or non-geospatial analysis only, and do not fully encompass all aspects of incident data. Furthermore, these tools are often either too complex for an untrained user, or are rigid in their design and allow only limited functionality.

The challenge is so daunting that many state Departments of Transportation (DOTs) are either forced to hire dedicated IT staff to help facilitate data mining requests, or worse, they simply do not attempt to perform any meaningful exploration of the data. For county and city DOTs, the problem is compounded due to lower budgets and a lack of resources. Even when budgets allow for analysis, experts must often make detailed data requests, wait for staff to generate queries, return data, review the results, and then revise the request and start again. This slow process is often frustrating and does not afford one to truly explore data, ask questions, and often fails to yield meaningful knowledge. Clearly, transportation professionals at the national, state, and local levels need better tools to empower them to accomplish sophisticated analytical data mining in an efficient and effective manner. These tools need to afford a user to more readily discover trends and patterns that would not normally be obvious.

A novel, web-based, visual analytics tool called ICE is proposed as an application that affords sophisticated yet user-friendly analysis of transportation incident datasets. The tool provides the user with an intuitive suite of functionality that includes data filtering, geospatial visualizations, statistical ranking functions, and multi-dimensional data exploration capabilities.

While each function is powerful by itself, ICE integrates them together allowing users to simultaneously interact with and see relationships between multiple visualizations.

The ICE application's rank-by-feature framework introduces interesting distributions and relationships according to many criteria: correlation coefficients, uniformity of distribution, number of outliers, etc. This framework is particularly unique in that most of the ranking criteria in the current literature are suitable only for numerical variables; however, transportation incident datasets consist mainly of categorical variables. This paper also proposes several novel ranking criteria for categorical data.

2. Related work

Most states provide access to summaries of their transportation incident data in various forms on-line. *(1, 2, 3, 4)* However, these summaries are typically pre-generated reports that do not allow for any interactivity or individual analysis.

The Fatality Analysis Reporting System (FARS) website is sponsored by the National Highway Traffic Safety Administration and allows users to create highly customized data queries. *(5)* While the FARS website represents a substantial leap forward in on-line data retrieval, the FARS tool leaves the burden of visualization and analysis up to the individual user. That is, the user must download raw data, and then perform graphing and statistical functions independent of the website application.

In a 2007 presentation given to the Safety Data Systems Task Group of the American Association of State Highway and Transportation Officials (AASHTO) it was noted that new data analysis tools were needed to evaluate transportation incident data. These tools needed to be: GIS-based, simple,

978-1-4244-4116-7/09/$25.00 ©2009 IEEE

straightforward, and provide for analysis of data in existing state data formats, be availability to a variety of user levels, and be expandable. *(6)*

To best allow a user to glean the most information from a dataset, one must first present the user with an “overview” (table, map, or other picture) of the entire data set. Next the user must be able to “zoom” and “filter” the data. As this occurs, the user must be able to access “details” about the data that remains on demand. Only from this top-down approach can the user begin to recognize patterns, realize what questions should be further probed, and notice trends that would otherwise go unnoticed. *(7)* However, if any of these tools are to be successfully implemented, they must be designed with simplicity, speed, and ease-of-use in mind.

Many local and state DOTs have sought such functionality from Geographic Information Systems (GIS) like ESRI’s ArcGIS suite of products. *(8)* These highly specialized geospatial tools allow for extremely complex spatial data analysis, management, and cartography; however, these powerful products have complex interfaces that can require months and even years of extensive training to fully master. Other tools like CommonGIS *(9)* and GeoVista *(10)* attempt to make data visualization simpler, yet in doing so lose the granularity and power that is needed to perform true transportation incident data analysis. Similarly, commercial data visualization products such as Spotfire *(11)* and Tableau *(12)* include mapping subcomponents. However, both tools show data points as single icons which results in occlusion and overcrowding for large datasets such as transportation incident data. Dykes, et al. explores population density data using Google Earth in conjunction with other open-source products *(13)*. In their example, the Google Earth API *(14)* represents a single entry in the data as a pushpin on the map. As with Spotfire and Tableau, the use of pushpins does not work well with clustered data due to problems with occlusion.

Heat maps are a method for representing spatial data that reveal the high-occurrence areas without obscuring the general view. Other GIS Tools *(15, 16, 8)* have used heat maps successfully for a number of purposes, though the difficulties in analyzing transportation incident data with these types of geospatially oriented tools has already been addressed. The geospatial and heat map functionality found within these applications need to be merged with more user-friendly statistical analysis tools.

In addition to problems relating to ease of use and occlusion, transportation incident datasets are composed of multidimensional data. Dealing with multidimensionality has been a challenge to researchers in many disciplines due to difficulties comprehending data in more than three dimensions while searching for relationships, outliers, clusters, and gaps. This challenge is well recognized and has been dubbed "the curse of high dimensionality". Seo and Shneiderman *(17)* present a conceptual framework for relationship detection called the rank-by-feature framework and demonstrate its capabilities in the Hierarchical Clustering Explorer (HCE). HCE ranks all possible axis-parallel projections against the selected criterion and presents the result in a color-coded grid; however, the ranking criteria in HCE are only applicable to numerical variables. Many variables in transportation incident datasets are categorical (e.g., weather could have values: "dry", "rainy", "cloudy", etc.), which indicates the need for a categorical variable ranking methodology. Continued work on HCE by Seo and Gordish-Dressman suggests estimating the relative significance between two categorical variables based on chi-square test. *(18)* Later sections of this paper further explore relationships between categorical variables and discuss several new ranking criteria.

3. Description of the interface

ICE follows the general design guidelines “overview first, zoom and filter, then details-on-demand” as described by Shneiderman. *(7)* The screen space is distributed into three sections as seen in figure 1. On the left is the control panel, which contains filter and ranking panels. The *filter panel* contains a rich set of filters that allows users to narrow down the dataset. The *ranking panel* adopts the idea of rank-by-feature framework, allowing users to rank the variables or relationships between variables by various criteria. The main area on the right is horizontally divided into two resizable sections. By default, the top-right section is dedicated to *the map,* featuring two displaying modes. Users can select between *icon* mode, plotting every incident as a circle on the map, or *heat* mode, using the heat map algorithm to draw color regions on the map. The bottom-right section locates a traditional tabular view of data and three visualization components: *histogram*, *two-dimensional (2D) plot*, and *parallel coordinates plot*. The histogram shows distributions of the dataset for a selected variable. The 2D plot visualizes relationships between a selected pair of variables in *scatter plot* mode and *grid* mode. The *parallel coordinates plot* provides an overview of distributions and relationships of multiple variables. To provide flexibility, users can also drag and drop to move components between the top-right and bottom-right section.

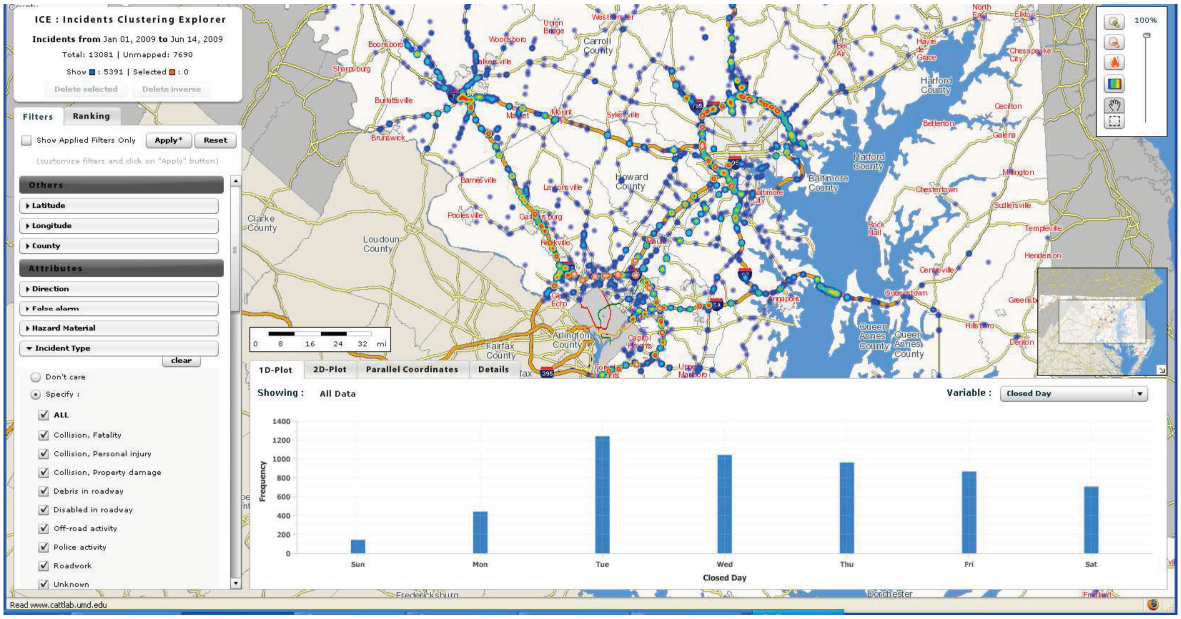
A significant feature of ICE is that all of these seemingly individual components are tied together. Filtering the dataset using the left side of the screen will immediately affect all other components. Selecting a particular incident or set of incidents in one component will also highlight identical selections in other components. Figure 1 shows that while viewing a histogram, one can select a cluster of points on the map which immediately highlights those data points on the histogram. Similarly, clicking on an area of the histogram highlights those corresponding data points on the map.

The following sections describe in greater detail some of the individual components of ICE. While each component is described separately, any interaction with one component ultimately affects the other component windows.

3.1 Mapping

Each accident is plotted on a built in map, taking advantage of the significance of the geospatial nature of the data. The primary method for exploring regions of interest is through zooming and panning. **Error! Reference source not found.** There are two mapping modes available: *icon* mode and *heat* mode. Users can switch between the modes using the controls along the top of the map. In *icon* mode, points are rendered as colored dots. Clicking on a dot brings up a details window displaying specific details for that particular accident.

While rendering each incident as an icon is reasonable in certain domains, it introduces several problems when dense regions need to be analyzed. Occlusion is a major issue when dealing with nearby icons, especially when the map is at the



ICE : Incidents Clustering Explorer

Incidents from Jan 01, 2009 to Jun 14, 2009

Total: 13081 | Unmapped: 7690

Show 🟦 : 5391 | Selected 🟧 : 0

Delete selected | Delete inverse

Filters | Ranking

□ Show Applied Filters Only | Apply* | Reset

(customize filters and click on "Apply" button)

Others
- Latitude
- Longitude
- County

Attributes
- Direction
- False alarm
- Hazard Material
- Incident Type

clear

○ Don't care

● Specify :

☑ ALL
☑ Collision, Fatality
☑ Collision, Personal injury
☑ Collision, Property damage
☑ Debris in roadway
☑ Disabled in roadway
☑ Off-road activity
☑ Police activity
☑ Roadwork
☑ Unknown

1D-Plot | 2D-Plot | Parallel Coordinates | Details

Showing : All Data

Variable : Closed Day

Read www.cattlab.umd.edu

Figure 1: The Incident Cluster Explorer user interface

farthest zoom levels. A region with only a few points can look like it has the same number of points as a dense region because it is difficult—if not impossible—to count overlapping points. The *heat* mode solves this problem by assigning each point a sphere of influence that dissipates when moving away from the point. Spheres of influence have an additive effect on each other, and each pixel is colored based on the total influence it receives. Regions with the most influence will be brightly colored in reds while less interesting areas will be assigned a lower wavelength color like blue.

The heat map algorithm used in ICE is based off of open source work by Corunet *(19)*, which uses heat maps to analyze web page clicks. The heat map gradient used by ICE transitions from low density to high through blue, cyan, green, yellow, orange, red, and white. These colors were selected because their order is associated with increasing temperature, and the relatively large number of distinct colors allows for the user to differentiate slight variations in density without much effort. The heat map uses highly saturated colors because they contrast well with the relatively pale background colors of the underlying map.

3.2 Ranking

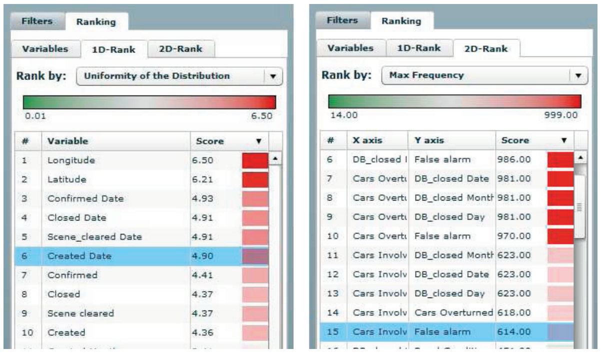
Because of the high-dimensionality of transportation incident data, ICE adopts the idea of a rank-by-feature framework from the HCE *(17, 18)* in ICE's ranking panel to help users find interesting one-dimensional (1D) and two-dimensional (2D) distributions. Users are allowed to select a ranking criterion and then sort the results (figure 2). Selecting a variable or a pair of variables from the ranking panel will show the corresponding histogram or 2D plot in the lower right window of the application as seen in figure 3.

While HCE illustrates many ranking criteria for 1D and 2D distributions, it is primarily focused on numerical variables. Transportation incident datasets consist of 3 types of variables:

- Numerical (N) variables: latitude (-77.124323), number of vehicles involved (0, 1, 2, … , n)
- Date-Time (D) variables: start date (12 Jan 2005), start time (10.45am)
- Categorical (C) variables: road condition (dry, rain, snow, etc.), direction (north, south, east, west, inner loop, etc.)

However, most of the variables are of type C. This presents a challenge as the existing HCE ranking criteria does not provide meaningful results when applied to variables of type C for both 1D and 2D rankings.

One approach to deal with this challenge is to convert categorical variables into ordinal variables. Ordinal variables are categorical variables that can be ordered by some criteria. By ordering the possible values of a categorical variable, a numerical value can be obtained: the "rank" of the value. Consider an example where, road condition (dry, rain, snow) can be ranked by the relative risk of driving in that condition. In this example, "dry" would be safest, whereas a rainy road would be more slippery, but not as much as when it snows. Thus, one may assign increasing numerical values 1, 2, 3 to represent each road condition. ICE developers decided to give



Filters | Ranking

Variables | 1D-Rank | 2D-Rank

Rank by: Uniformity of the Distribution

0.01 | 6.50

#	Variable	Score
1	Longitude	6.50
2	Latitude	6.21
3	Confirmed Date	4.93
4	Closed Date	4.91
5	Scene_cleared Date	4.91
6	Created Date	4.90
7	Confirmed	4.41
8	Closed	4.37
9	Scene cleared	4.37
10	Created	4.36

Filters | Ranking

Variables | 1D-Rank | 2D-Rank

Rank by: Max Frequency

14.00 | 999.00

#	X axis	Y axis	Score
6	DB_closed I	False alarm	986.00
7	Cars Overtu	DB_closed Date	981.00
8	Cars Overtu	DB_closed Month	981.00
9	Cars Overtu	DB_closed Day	981.00
10	Cars Overtu	False alarm	970.00
11	Cars Involv	DB_closed Month	623.00
12	Cars Involv	DB_closed Date	623.00
13	Cars Involv	DB_closed Day	623.00
14	Cars Involv	Cars Overturned	618.00
15	Cars Involv	False alarm	614.00

Figure 2: Screen captures of 1D and 2D rankings

the user the ability to order categorical data numerically which ultimately allows for standard HCE methods to be used.

In addition to the five existing 1D ranking criteria from HCE, ICE adds the five new criteria seen in Table 1 to rank 1D distributions for categorical variables. These criteria are based on frequency (F_k)—the number of incidents in each category in each variable. Let X be a categorical variable. X has n possible values X_1 to X_n. F_k is the number of incidents in which $X = X_k$. Using F_k, five additional ranking criteria seen in Table 1are used to rank 1D distributions for categorical variables. Once a 1D ranking criteria is chosen, the user can click on any of the resulting variables to automatically generate a histogram which then appears in the lower right section of the application.

Table 1: Five ranking criteria in ICE

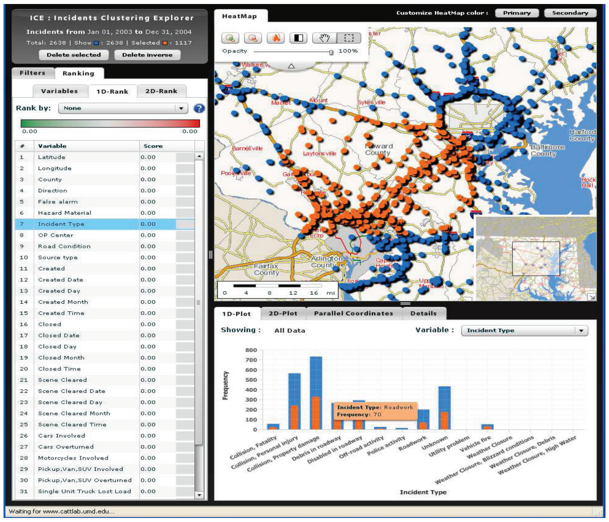
Ranking	Formula
Maximum Frequency (0 to n)	$Maximum(F)$
Number of Potential Outliers (0 to n)	$Count(Z(F)>1.5 \text{ or } Z(F)<-1.5)$
Percentage of Empty Area (0 to 100)	$Count(F>0) / Count(All\ F)*100$
Number of Existing Values (0 to n)	$Count(F>0)$
Standard Deviation of Distinct Frequencies (0 to n)	$Standard\ Deviation\ (F>0)$

The existing HCE work presents seven 2D ranking criteria, but only one ranking criteria for 2D C-C relationships—a contingency coefficient based on the chi-square test. However, the five 1D ranking criteria proposed earlier can also be adapted to rank 2D C-C relationships. These additions improve the flexibility of the rank-by-feature framework providing more options to users. Instead of using F_k, the frequency now becomes Fij, the number of incidents in each value pair in each relationship. Let X and Y be categorical variables. X has n possible values X_1 to X_n while Y has m possible values Y_1 to Y_m. F_{ij} is the number of incidents in which $X = X_i$ and $Y = Y_j$. Using F_{ij}, the five ranking criteria seen in Table 1 can now be used to rank 2D C-C relationships. Once a 2D ranking criteria is chosen, the user can then click on any of the resulting variable pairs to automatically generate a number of interactive plots which appear in the lower right section of the application.

3.3 Histograms

The histogram panel can be accessed from the 1D distribution window and shows variations and clusters in the number of incidents for each variable. The histogram panel is highly interactive in several ways. Figure 3 depicts a histogram for "incident type" in the data set. When the user clicks on any of the histogram bars, those incidents related to that category of incident are highlighted in the map interface. Conversely, the user may select a set of incidents on the map and the incidents selected are highlighted within the histogram as a highlighted subset.

One can also examine the incident frequencies in a monthly, weekly, daily or hourly fashion. First, clicking on certain types of histograms highlights incidents associated with those values in the map. Additionally, temporal data is represented as a hierarchy of values (minutes, hours, days, months, years), and temporal histograms in ICE support a logical way for interacting with this hierarchy. The initial view displays the correlation between incident numbers and the month of the year in which those incidents occurred.



ICE : Incidents Clustering Explorer

Incidents from Jan 01, 2003 to Dec 31, 2004

Total: 2638 | Show 🟦 : 2638 | Selected 🟧 : 1117

Delete selected | Delete inverse

Filters | Ranking

Variables | 1D-Rank | 2D-Rank

Rank by: None

0.00 0.00

#	Variable	Score
1	Latitude	0.00
2	Longitude	0.00
3	County	0.00
4	Direction	0.00
5	False alarm	0.00
6	Hazard Material	0.00
7	Incident Type	0.00
8	OP Center	0.00
9	Road Condition	0.00
10	Source type	0.00
11	Created	0.00
12	Created Date	0.00
13	Created Day	0.00
14	Created Month	0.00
15	Created Time	0.00
16	Closed	0.00
17	Closed Date	0.00
18	Closed Day	0.00
19	Closed Month	0.00
20	Closed Time	0.00
21	Scene Cleared	0.00
22	Scene Cleared Date	0.00
23	Scene Cleared Day	0.00
24	Scene Cleared Month	0.00
25	Scene Cleared Time	0.00
26	Cars Involved	0.00
27	Cars Overturned	0.00
28	Motorcycles Involved	0.00
29	Pickup,Van,SUV Involved	0.00
30	Pickup,Van,SUV Overturned	0.00
31	Single Unit Truck Lost Load	0.00

Showing : All Data

Variable : Incident Type

Incident Type: Roadwork
Frequency: 70

Waiting for www.cattlab.umd.edu...

Figure 3: Interactive histograms and maps are linked together

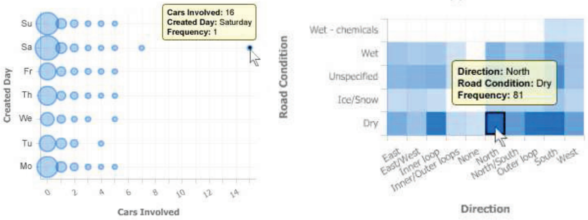
Control-Clicking on a particular month zooms in via a drill-down animation effect and displays the information by day of the month. Users can further zoom in to view the number of incidents spread over a 24-hour period. The zooming process is animated to prevent users from becoming disoriented by rapid changes in the histograms.

ICE's histograms use a "breadcrumb trail" which is a horizontal list of labels used for keeping track of one's location within a series of views. Clicking on any labeled category in the breadcrumb trail will zoom back directly to that particular zoom level.

All subsequent sections of this paper show other forms of graphs and visualizations that can be viewed from the same section of ICE as the histogram panel.

3.4 Two-dimensional (2D) plots

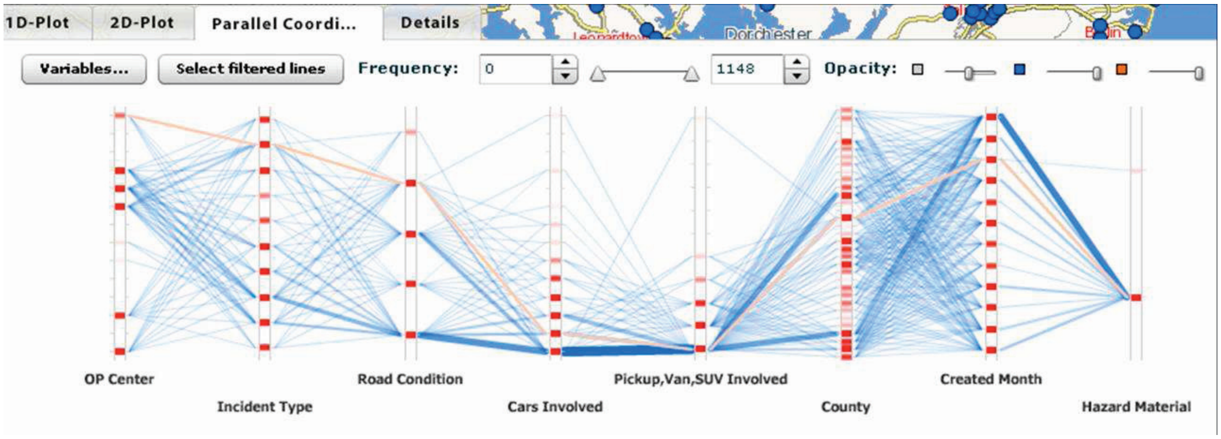
Traditional scatter plots can visualize the relationship between two variables by drawing elements on *(x,y)* coordinates. ICE adopts this idea and combines it with the idea of using a color grid—colored tiles where color shades represent values. Placing mouse over a colored tile or circle will show more information about the item. For every 2D plot, users can choose between two modes: *Scatter Plot* or *Grid* mode. Figure 4 shows the exact same 2D plot in both scatter plot mode and grid mode.



Cars Involved: 16
Created Day: Saturday
Frequency: 1

Direction: North
Road Condition: Dry
Frequency: 81

Figure 4: Two-dimensional plotting in ICE



Frequency: 0 1148

Figure 5: Parallel Coordinate Plot

In the *scatter plot* mode, incidents are represented by circles. The size of each circle represents frequency of occurrence. Big circles will have higher frequencies than smaller circles. Placing one's mouse over a circle brings up additional details about that data and highlights those specific incident records on the map. To provide an alternative view, the *Grid* mode is introduced. Both *X* and *Y* axes are divided into rectangular bins. Each section of the grid, or bin, is then colored according to the number of incidents that fall into the given area. Dark colors represent high values while light colors represent low values. This technique can prevent problems with occlusion that might arise in the scatter plot mode. Figure 4 shows examples of both types of plots.

3.5 Parallel coordinates plot

Parallel coordinates plot (Figure 5) is a way to visualize multivariate data on a plane. Each plot can have as many vertical axes as the dataset has variables. ICE allows the user to specify which variables to include within the plot. A single incident record is represented as a line that goes across all axes from left to right; on each axis, the line goes through the point corresponding to the record's value for that field.

Using parallel coordinates plot, it is easy to see clusters in the data. If many lines converge at one point on any given axis, then many incident records have that particular value in common for that particular variable. On the other hand, if there are only a few lines going through other points on the axis, the values corresponding to these lines are outliers in the original dataset. Examining both outliers and clusters may open new insights into the data.

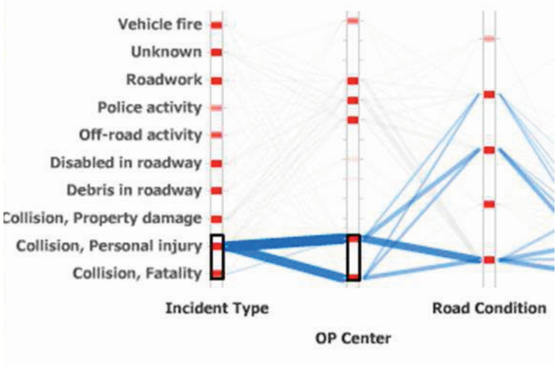


Figure 6: Filtering the Parallel Coordinate Plot

As with the icon mode in the ICE map, overcrowding may become an issue with large data sets. Therefore, ICE gives users control over both line thickness and transparency. Additionally, heat maps have been added along the axes to identify the clusters of lines. The heat map overlays can give a clear idea about the distribution of the values for any particular axis. The radius of the heat maps may also be adjusted by the user to prevent occlusion resulting from close data points. Users can also graphically filter incidents on the plot by using the mouse to select certain values on the plots. Multiple variables can be filtered simultaneously as shown in Figure 6.

Parallel coordinates plot in ICE can display both numerical and categorical fields as well as dates and timestamps. As with the other components in ICE, "brushing and linking" is used to highlight selected items on the map when they are selected in the parallel coordinates window and vice versa. For example, if the parallel coordinates plot shows a high concentration of converging lines in a particular axis, it would be reasonable to assume those accidents are somehow correlated with that property. Selecting those accidents on the plot will highlight the same accidents in all the other views including the map, histogram and 2D plot.

4. Implementation details

All incident data is stored in a PostgreSQL database. ICE's server tier is written in ColdFusion and the front-end is written in Flex. The system requires that Adobe Flash Player 9 or higher be installed on the client machine.

There are several hundred thousand incidents in the current database. The current unoptomized ICE application is capable of visualizing approximately 15,000 simultaneous incidents which is sufficient to view several year's worth of freeway accident data for the state of Maryland.

5. User evaluation

Though a formal usability analysis study has yet to be conducted on the ICE application, many transportation professionals have experimented with the tool and have provided valuable feedback. Each user was given a short, but detailed overview of the tool introducing them to the features. Each evaluator commented on performance, usability, and applicability.

All evaluators were impressed with the supporting analytical components that the tool provided and were particularly intrigued by the 2D and parallel coordinates plot since they provided a systematic way to explore relationships within data. Positive comments were also received on the many "brushing and linking" features that tied each of the three main panels of the application together making it easier to explore relationships and find patterns. Users highly appreciated the fact that the application was extremely user friendly, the interface was well thought-out, required almost no training, and was web-accessible. One reviewer, however, commented that while heat maps represent the overall distribution of incidents in the region quite well, the heat maps themselves could be misleading since they (1) are only representing straight counts of incidents rather than rates of incidents based on road size and volumes, and (2) the heat maps aggregate accidents on adjoining roads together due to their proximity to one another.

6. Future work

Given the feedback on heat maps, developers are working to improve the heat map algorithms so that the user can have control of the heat map granularity, specify color schemes, and breakpoints. While the application is currently running off of incident records from the Maryland CHART program, it has been developed to be easily extensible to other datasets with minimal programming effort. However, a simple data import tool allowing other datasets to be automatically integrated into the application is desired. ICE could be adapted to work on non-transportation related multivariate datasets.

7. Conclusions

ICE affords sophisticated yet user-friendly analysis of transportation incident datasets. Interactive map, histogram, two dimensional plot and parallel coordinates plot are featured visualizations that are integrated together to allow users to simultaneously interact with and see relationships between multiple visualizations. Users can create custom conditions to filter data and focus on a smaller dataset. Due to the multivariate nature of the data a rank-by-feature framework has been adopted and further expanded to quantify the strength of relationships between the different fields describing the data. The ICE application will allow transportation professionals to spend less time and energy worrying about the pure mechanics and economics of asking questions of the data, and afford them the opportunities to determine which questions to ask, seek out new answers, and derive knowledge from these vast data sources.

8. Acknowledgments

The authors wish to thank the Maryland State Highway Administration for their support. A special thanks to Dr. Ben Shneiderman at the University of Maryland for his guidance and feedback throughout the design and development process.

9. References

[1] GDOT. *Crash Analysis, Statistics & Information Notebook*. http://www.dot.state.ga.us/statistics/CrashData/Pages/casi.aspx. Accessed Jul. 27, 2008.

[2] Hammond, P. J. *Measures, Markers and Mileposts*. WSDOT's quarterly report to the Governor, the Legislature, and the Washington State Transportation Commission on transportation programs and department management, 2008.

[3] University of Maryland's National Study Center for Trauma and EMS. *Maryland Traffic Safety Facts 2006*. http://medschool.umaryland.edu/nscfortrauma/traffic.asp. Accessed Jul. 27, 2008.

[4] NHTSA. *Maryland – Toll of Motor Vehicle Crashes.* http://www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/STSI/24_MD/2006/24_MD_2006.htm. Accessed Jul. 27, 2008.

[5] FARS. *FARS Encyclopedia*. http://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed Jul. 27, 2008.

[6] Welch, T. AASHTO Safety Data Systems Task Group. Presented at AASHTO Standing Committee on Highway Traffic Safety 2007 Spring Meeting, MO, 2007.

[7] Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, In *Proceedings IEEE Symposium on Visual Languages*, 1996.

[8] ArcGIS. http://www.esri.com/software/arcgis/. Accessed Jul. 18, 2008.

[9] Andrienko, N., and G. Andrienko. Interactive Visual Tools to Explore Spatio-Temporal Variation. *AVI '04: Proceedings of the working conference on advanced visual interfaces*, ACM, NY, 2004.

[10] MacEachren, A., X. Dai, F. Hardisty, D. Guo, and G. Lengerich. Exploring high-d spaces with multiform matrices and small multiples. *IEEE Proceedings of the International Symposium on Information Visualization*, 2003.

[11] Spotfire. http://www.spotfire.com. Accessed Apr. 20, 2008

[12] Tableau. http://www.tableausoftware.com. Accessed Apr.20,2008.

[13] Dykes J., A. Slingsby, and K. Clarke. Interactive Visual Exploration of a Large Spatio-Temporal Dataset: Reflections on a Geovisualization Mashup. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 13, No. 6, 2007.

[14] Google. *Google Earth API*. http://code.google.com/apis/earth/. Accessed Jul. 18, 2008.

[15] Fisher, D. Hotmap: Looking at Geographic Attention. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 13, No. 6, 2007.

[16] Universal Mind. *Universal Mind: Demos*. http://www.universalmind.com/demo/launchpad.cfm. Accessed May 5, 2008.

[17] Seo J. and B. Shneiderman. A Rank-by-Feature Framework for Interactive Exploration of Multidimensional Data. *Information Visualization*, Vol. 4, No.2, 2005.

[18] Seo J. and H. Gordish-Dressman. Exploratory Data Analysis with Categorical Variables: An Improved Rank-by-Feature Framework and a Case Study. *International Journal of Human-Computer Interaction*, Vol. 23, No.3, 2007.

[19] Corunet. *How to make heat maps.* http://blog.corunet.com/english/how-to-make-heat-maps. Accessed Apr. 20, 2008.