Visual Drill Down of Spatiotemporal Accident Data

Anil Kumar Motupalli, Pujith Sagar Edara, and Vishwanath Potluri, Member, IEEE

**Abstract**—These instructions give you guidelines for preparing papers for IEEE Computer Society Transactions. Use this document as a template if you are using Microsoft Word 6.0 or later. Otherwise, use this document as an instruction set. Please note that use of IEEE Computer Society templates is meant to assist authors incorrectly formatting manuscripts for final submission and does not guarantee how the final paper will be formatted by IEEE Computer Society staff. This template may be used for initial submissions; however, please consult the author submission guidelines for formatting instructions as most journals prefer single column format for peer review. An abstract should be 100 to 200 words for regular papers, no more than 50 words for short papers and comments, and should clearly state the nature and significance of the paper. Abstracts *must not* include mathematical expressions or bibliographic references. Please note that abstracts are formatted as left justified in our editing template (as shown here).

————————————————

1. Anil Kumar Motupalli, Email: [amotupal@asu.edu](mailto:amotupal@asu.edu), ASURITE: amotupal
2. Pujith Sagar Edara, Email: [pedara@asu.edu](mailto:pedara@asu.edu), ASURITE: pedara.
3. T.C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309. On leave from the National Research Institute for Metals, Tsukuba, Japan E-mail: [author@nrim.go.jp](mailto:author@nrim.go.jp).

—————————— ◆ ——————————

# 1 Introduction

T

TRAFFIC conditions are influenced by a lot of factors particularly in countries like The United States of America which is highly diverse in the environment, culture and technology used.

We would like to start our discussion with a small story that happened on my last vacation to colorado. In our return trip, we started with snowfall in Colorado, continued our drive-in rain and ended it in hot and Sunny Arizona. Environmental conditions contribute a lot to the driving conditions and can make driving challenging. For example, snowfall and rain can contribute to skidding and a bright sun can blind our vision. Driving conditions change from place to place with the season. Arizona winter is friendlier compared to driving compared to Seattle/Colorado winter where as a Seattle summer is more driving friendly compared to Arizona Summer. Every year National Highway Traffic Safety Administration(NHTSA) makes the Fatality Analysis Reporting system(FARS) data publicly available. This diversity in weather conditions will lead to interesting patterns in the Geographical analysis of accident data which led us to decide on analyzing the accident data related to The United States of America.

There are a lot of technological advancements in the last 10 years that assist drivers by enhancing the driving experience. Advanced mapping technologies with voice assisted navigation reduces the effort required to navigate and advancements in technology and Artificial Intelligence led to intelligent Driver Safety technologies like Automated collision prevention and obstacle detection. All these technologies should help to reduce the probability of accidents. However, the technological advancements have its own disadvantages along with the advantages. The increase in the mobile usage will lead to increase in distractions like texting and driving, mobile usage while driving. Along with the trends in the data, we also want to visualize any seasonal trends in the accidents. These technical advancements in the past 10 years will lead to interesting temporal patterns. Therefore, we are considering the 10 years of data i.e., 2006-2015 for the analysis.

FARS is a rich data set with a granular data about the accident, the vehicles involved in the accident and the persons involved in the accident for every accident that is reported. It also has data related to the factors that would have contributed to the accident. The granularity of the data will allow us to ask interesting questions on the data like How is the manner of collision correlated to weather condition in the incident? Therefore, we want to visualize FARS data using multivariate, geographical and temporal visualizations.

# 2 Related Work

## 2.1 Accident data

Previous work on accident data and their shortcomings.

## 2.2 Multi-Variate Visualization

There are multiple ways of visualizing multi-variate data. The parallel coordinate is one such way. It is useful for comparing multiple dimensions of data. It compares values of different dimensions for each data record and provides a detailed visualization. Such visualization can overwhelm a user and is not the primary purpose of our project’s multivariate visualization. Parallel-sets is a variation of parallel coordinate visualization, which emphasizes the correlation between the dimensions as a whole rather than individual data records, this will help the user to understand the correlation between the dimensions without difficulty and distraction.

## 2.3 Geographic Visualization

Previous work on geographic visualizations.

## Temporal visualization

Previous work on time series visualizations.

## Spatiotemporal Visualization

Previous work on time series visualization.

# 3 Design principles

## Data Discussion

Discuss our overwhelmingly large dataset and the challenges we faced while preparing this dataset for Visualization.

## Visual Variables

Location:

Parallel sets visualization is located at the top of the page to give an overview of the correlation of the factors as the data is aggregate for all the years in the dataset. It is oriented horizontally. Each horizontal axis represents a factor selected. The divergent color theme is utilized since the categories of the selected factor are nominal in nature. The color value of the ribbon increases when the mouse hovers over it to bring users attention to the information displayed in the tooltip. This increases the readability of the information.

Discuss why we chose some colors and other different visual variables.

## Questions that can be answered

Different questions our visualization system can answer. How we decided on using stacked area to alleviate problems with previous spatiotemporal visualization systems.

parallel sets are used for comparing categorical data of factors selected by the user. For example, user can compare factors like manner of collision (having ‘Front’, ‘Rear’, ‘Angle’, ‘Random’ categories) and weather conditions (having ‘Moderate’,’Rain’,’Snow’,’Fog’,’Winds’ categories) and can draw the conclusions that most of the accidents occur during rainy weather second only to ‘moderate’ weather conditions. From this observation, we can hypothesize that drivers are to be careful during rainy season and authorities should take necessary precautions to reduce fatalities involving these factors. Similarly, the user can compare different factors draw conclusions from them.

# System

## Technology stack

Write about the different technologies we used for developing this project like GeoJsons, crossfilterJS, DC.js, D3.js, python, Http-Server, d3.parsets.js and GitHub for team management

## Parallel sets

Parallel sets visualization is used for the purpose of visualizing any correlation between the factors nominal data with categorical sets. The discrete lines on the axis for a factor represents the categories and the width of the ribbons are scaled according to the frequency of the occurrence of categorical data in the dataset.

## Geographic Visualization

We are using Choropleth map to show accidents across different states in the USA. We are using two choropleth maps to show two different levels of Geographic Data. The first level of our Geographic view shows a Choropleth map of states and selecting a single state will show a county level Choropleth map for that state. We are using sequential colors for representing the accidents in the Choropleth map because the number of accidents is in ratio scale. We divided the states into seven bins with an equal width based on the number of accidents per 1000 population. We normalized the state and county values based on the number of accidents.

While developing this visualization, we faced few challenges and maintaining consistency between size and scale of state and county visualizations is the biggest we faced. Though we could find a set of county-level GeoJsons for all the states in the USA, they were not in a standard format. Therefore, we took the county level GeoJson for the entire country and processed it to produce a GeoJson for each state. However, the scale and position of each state were another problems we had to address with the county level Choropleth because the state maps are too small and they are rendering in their respective positions in the USA map instead of the center of the container which resulted in occlusion and absence of few state level maps. We have had to carefully handcraft the position and scale for each state to get a consumable county level maps for the states.

## Time Series Visualization

Our system adopts stacked area graph and heat map visualization techniques to visualize accidents based on various factors, in the time dimension. Our visualization helps the end user to perform time series analysis of accidents or deaths happening in a

## Interactive elements in the system

Most of our charts in our visualization system provide interactive elements to visually query and analyze the accidents data. Typical interactions on our chart include dynamic querying, brushing the charts, hovering on a map element. If we are unable to provide interactivity in any of our charts we enabled that by adding another visualization with another chart. Though we encode the data visually user can hover the mouse on the visual elements to see actual values. Every chart is connected to the other chart; therefore every chart will change according to the interactions on the other charts.

For parallel sets, the user can add new factors to the existing list of factors to generate a correlation between the factors which can help the user to develop insights on the data. The addition of factors can be done through a drop-down button with a list of factors from which user can choose. Each level of the visualization has an option to rearrange the categories based on the category frequency or alphabetic order of the category name. The user can hover over a particular ribbon that connects two categories of different factors and get the total aggregate and percentage of the frequency of occurrence of such combination in the dataset.

Bar charts are provided dynamically for all the selected factors of parallel sets to the right side of visualization which acts as drill-down for each selected category and also as a filter for the user to select a particular category from a factor to better understand its correlation to other factors.

In our choropleth map user can view the absolute number of accidents in a state by hovering mouse over that state. Selecting a state acts like a filter and it filters the entire data for by that state and the rest of the charts will be adjusted accordingly. Selecting a single state will show the county level map of the state and selecting multiple states will add those states to the stacked area chart for comparison.

## Limitations and possible extensions

Write about slow loading times of our project and possible extensions with enough time.

# 5 Results

Different observations we found using our system.

As a final result, we designed a parallel set visualization that can provide a correlation between various interesting factors that a user can select. Initially, at the start of the visualization, it provides an overall correlation between selected factors. The correlation between factors changes as the user interacts with other visualizations to gain insights. For example, when the user selects a particular year for analysis, this filter of data is applied to parallel sets as well. The user can hover over the correlation lines to know the magnitude of frequency of occurrence and its overall percentage in the dataset. This follows the Shneiderman’s visualization mantra of overview first and details on demand.

# 6 Conclusion

**References**