UCD-Griffin-MC2 Submission Summary

Kevin S. Griffin*

University of California at Davis

ABSTRACT

This paper describes the approach and the visual analytics (VA) tool developed to solve the VAST 2014 Challenge: Mini-Challenge 2. For this challenge, a methodical approach was taken by analyzing each GAStech employees daily activities. Open source software, commercial application programming interfaces (APIs), and custom software were combined to create a very robust VA tool used in this challenge. The tool provides visualization and animation of the tracking data, an overview of credit card and loyalty card transactions, and on-demand details of the aforementioned data. Finally, a novel feature was implemented in the VA tool to insert missing location data for credit card and loyalty card transactions in real-time.

1 INTRODUCTION

The analysis and tool development was driven by the characteristics of the data that was provided for this challenge. Visualization components were developed to leverage these unique attributes. These components were then integrated into a single tool that proved critical for identifying suspicious patterns of behavior that could've led to the missing GAStech employees. Section 2 describes the characteristics that informed the analytical approach and development, section 3 gives an overview of the VA tool, and section 4 summarizes the tool development and analysis.

2 APPROACH

2.1 Geospatial Data

The Keyhole Markup Language (KML) [2] format of the Kronos Island and Abila shapefiles were used. This format was chosen so that the geographic data could be visualized using Google's very robust Maps Application Programming Interface (API) [1] (See Figure 1). The Google Maps API provided seamless integration of the geographic data with its Layers interface. More specifically, the KML Layer was used to render Abila and Kronos on the map. As an added benefit of using this API, panning, zooming, and the ability to drill down into the data to see additional information, like street names, was provided without any additional development effort.

2.2 Tracking Data

The *car-assignments.csv* and the *gps.csv* files were placed into separate tables in a MySQL database [3]. These tables were then joined on *car id* to create a single table containing the vehicle tracking data and the associated employee information. The main characteristics of this dataset that motivated my visualization approach are the size of the data (613,086 rows), the spatio-temporal nature of the data, and the completeness of the data. The Google Maps API proved ideal for visualizing this data because of its robust support for drawing markers, paths, and info windows on the map according to its latitude/longitude values. An *Options* filter within the VA tool was developed to filter this large dataset based on employee name and a specified date range. The *Options* filter also allows the analyst to show or hide markers/paths, specify the number of markers/paths to display, and to animate the placement of the markers on the map to simulate the employee's vehicle movements. Figure 2 shows how this data is visualized within the VA tool.

*e-mail: kevgriffin@ucdavis.edu

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

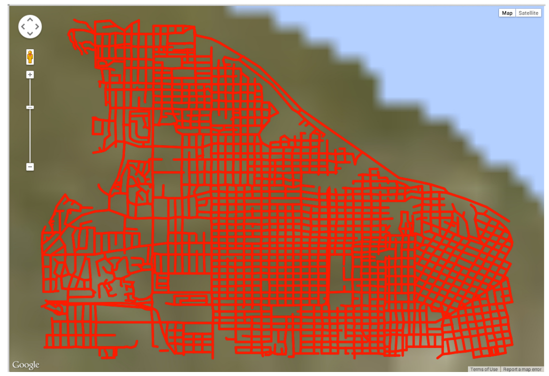

Figure 1: Kronos and Abila KML Layers

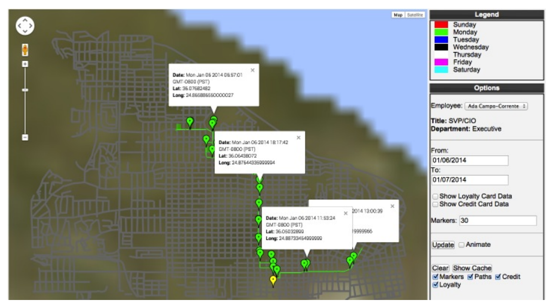

Figure 2: Vehicle Tracking Data

2.3 Transaction Data

2.3.1 Credit and Debit Card Data

The credit and debit card transaction data was small ($\sim$ 1500 rows) and incomplete (missing geospatial coordinates for the business locations). These characteristics, along with its temporal nature, helped inform my visualization approach for this dataset. Google Map's circle shapes were used to represent credit/debit card transactions. Circles, and shapes in general, are just objects on the map tied to latitude and longitude coordinates [1]. The size of the circle is proportional to the purchase amount and the color of the circle represents the day of the week (see Figure 3). Since the location data was missing for the purchase locations, an interactive learning approach was used to automatically populate and save the latitude/longitude coordinates. As the user drags the shape to the correct location on the map, the final mouse location is converted to latitude/longitude coordinates and added to the credit/debit card data. Because these coordinates are integrated into the original data, analyst are spared the wasted time from having to figure out the correct location for these shapes during each subsequent analysis. These positions are initially determined from the tourist map and reference markers (tracking data). Finally, clicking on the circles will display an information window that displays the business location name, timestamp of the purchase, the price of the purchase, and the automatically generated latitude/longitude coordinates of the business.

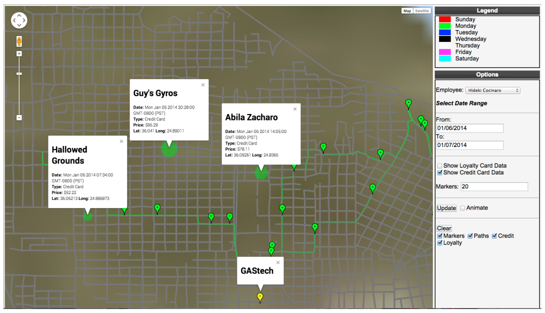

Figure 3: Credit Card Data

2.3.2 Loyalty Card Data

The main characteristics of the loyalty card dataset that motivated my visualization choices were its size ($\sim$ 1400 rows), it is temporal nature, and the dataset wasn't complete (missing geospatial coordinates for the business location and time of purchase). Google Map's rectangle shapes were used to represent loyalty card transactions. The size of the rectangle was mapped to the loyalty card transaction value and the color represents the day of the week (see Figure 4). The location data was missing from this dataset also, so interactive learning was used to automatically populate and save the latitude/longitude coordinates just like in the credit/debit card dataset. Clicking on the rectangles will display an information window that shows the business location name, date of the transaction, the price of the transaction, and the latitude/longitude coordinates of the business location.

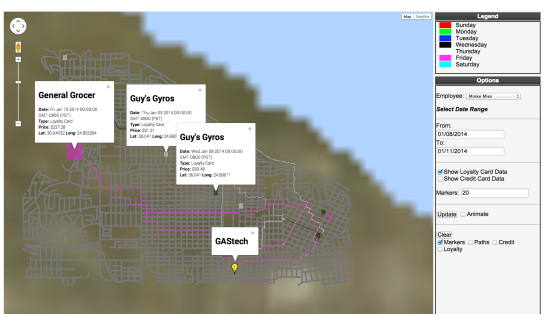

Figure 4: Loyalty Card Data

3 Visual Analytics Tool

Figure 5 shows the VA Tool created to solve the VAST Challenge 2014: Mini-Challenge 2. The main display area uses the Google Maps API to display all of the geospatial data (vehicle tracking, business locations, Abila and Kronos KML data). The right side of the tool displays the *Legend* and *Options* user interface (UI) components. For the *Legend* UI, colored boxes and labels were used to depict the day of the week. The color mappings found in the *Legend* UI were used to determine the color of all paths, markers, circles (credit/debit cards), and squares (loyalty cards) visualized in the main display area.

The *Options* UI contain the components used to filter and customize the presentation of the data provided for this challenge. The first UI component, employee filter, utilizes a drop-down box to select an employee to filter on and also show their title and department in the labels added below the employee filter. A date filter is provided to allow an analyst to filter the data based on a single day or a range of days. Below that, the "Show ..." check boxes can be used to show or hide the credit/debit card and/or the loyalty card shape markers. The *Markers* input text field has two different uses depending on the UI state. If *animate* checkbox (see description below) is selected, then the *Markers* input text field is used to represent the number of animation steps. If the *animate* check box is not selected, then the input text field is used to represent the number of vehicle tracking markers to visualize along the path. The *Animate* check box is used to request animation of the vehicle tracking markers. The *Clear* button removes all items selected in the associated check boxes from the main display area.

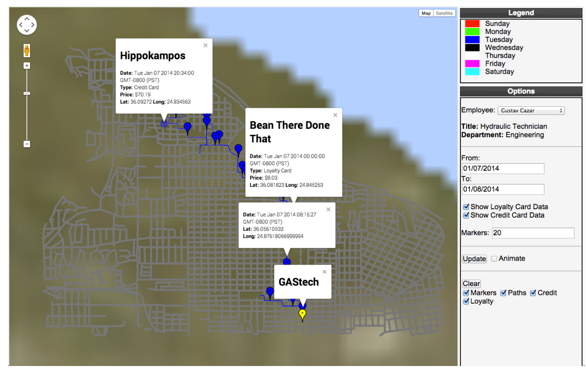

Figure 5: VA Tool

4 Conclusion

The VA Tool implemented for this challenge was inspired by several characteristics of the data (size, spatio-temporal nature, and completeness). The vehicle tracking data contained over 600,000 records so filtering based on employee name, date, and marker count was implemented to help show the data in a more meaningful and manageable way. The temporal nature of the data inspired the use of animation to show the order in which the employee traversed different locations. The geospatial data included in the dataset was visualized well using the Google Maps API. Finally, to handle missing/incomplete data, like GPS coordinates for credit card purchases, an interactive learning approach was implemented. This approach turned out to be a huge benefit since subsequent analysis using these locations maintain the correct coordinates and save the user from the additional work of having to disambiguate the location every time.

References

[1] Google Maps Javascript API. https://developers.google.com/maps/documentation/javascript.

[2] KML Reference. https://developers.google.com/kml/documentation/kmlreference.

[3] MySQL: The World's Most Popular Open Source Database. http://www.mysql.com.

[4] N. Q. Zhu. *Data Visualization with D3.js Cookbook*. Packt Publishing, 2013.