CSE578 Data Visualization Project Report: VAST 2021 Mini-Challenge 2

ALIMOHAMMAD BEIGI, CHAEHWAN YOON, JISOO LEE, NAYOUNG KIM, RENE MARCELINE, and SOHAM NAG

Our project is about investigating the disappearance of GAStech employees during a corporate celebration in the fictional island country of Kronos. An organization called Protectors of Kronos organization is a suspect in this incident. The project involves analyzing geospatial data of company cars and trucks, loyalty and credit card data of employees, and geospatial coordinates of Kronos. The aim is to identify anomalous activities and behaviors of employees and their corresponding locations. In this project, we develop a nested visualization design model that can be used to summarize the findings.

1 INTRODUCTION

In the fictional island of Kronos, GAStech employees have access to company cars for both personal and business use, while those without company cars can only use company trucks for business purposes. The vehicles are geospatially tracked without the knowledge of the employees, and this tracking data has been provided to law enforcement to aid in the investigation of the disappearance of GAStech employees during a corporate celebration.

The tracking data is only available for the two weeks prior to the disappearance, and unfortunately, there is no data for the day of the incident. Additionally, Kronos-based companies offer Kronos Kares benefit cards to GAStech employees, providing discounts and rewards for collecting information about their credit card purchases and preferences through loyalty cards. This data has been provided to investigators, but it does not contain any personal information.

As a visual analytics expert working with law enforcement, our task is to identify which employees made specific purchases and to detect any suspicious behavior patterns. The challenge is to deal with uncertain and incomplete data to make recommendations for further investigation.

We first pre-process the entire dataset by filtering unwanted information. Then, we use a geometric algorithm to match the tracking location data (time series) to the 2D image map. As our objective is to detect the anomalies, we implement 5 different visualizations while analyzing and handling computational complexity. As a result, we could uncover several significant abnormalities in the dataset, such as noticeable relationships between employees, and suspicious activities by individuals. The overall framework of our visualization is illustrated in Figure 1.

Authors' address: Alimohammad Beigi, abeigi@gmail.com; Chaehwan Yoon, cyoon8@asu.edu; Jisoo Lee, jlee815@asu.edu; Nayoung Kim, nkim48@asu.edu; Rene Marceline, rimarcel@asu.edu; Soham Nag, snag8@asu.edu.

Manuscript submitted to ACM

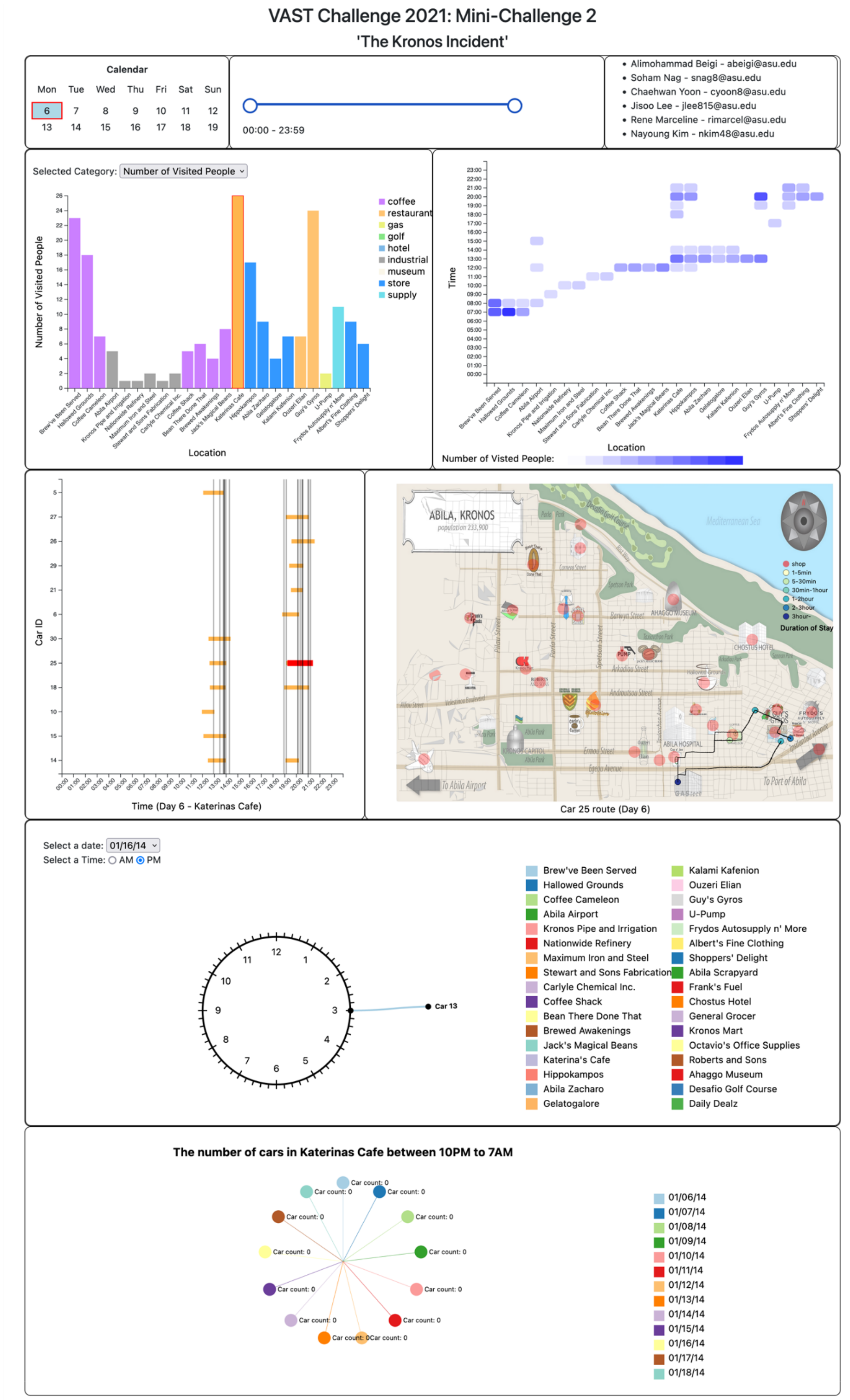

Fig. 1. The final visualization framework

2 VISUALIZATION DESIGN

Various visualizations such as maps, heatmaps, and bar charts can be used to identify popular locations based on saturation levels and credit card and loyalty card transactions. A bar chart with the estimated number of visitors can be used to compare different locations and determine the most frequented ones. The x-axis and lines help distinguish between places, and the time slider and select box allow for detailed data analysis.

Another visualization used to identify the most popular locations is a heatmap that displays the saturation of visited people at each location. A dot mark and color saturation for the number of people visited, an x-axis for category (location), and a y-axis for ordinal (time) are used in this visualization. Interactions such as tooltips showing the number of visited people and a Gantt chart for arrival and departure times are added to enhance user experience. Furthermore, the use of a time slider and calendar box aid in identifying the most popular place or location where the highest transaction amount occurred by clicking on each box (day).

The vehicle data can be analyzed using a Gantt chart, with the x-axis representing time and the y-axis representing car IDs. The line with lollipop ends representing arrival and departure can help distinguish between different cars and their stay durations at a particular location. By clicking on a line, a histogram of the car's stay at that location for two weeks can be displayed, aiding in identifying unusual patterns of vehicle movement and potential suspicious activity.

The methods for identifying a credit card owner include analyzing travel history, credit card purchase history, and vehicle data. Scatterplots, line charts, and connected scatterplots can be used to identify if a person has purchased from a particular store or not based on their travel history. Clustering stopped places and matching individuals based on their credit card and car data can provide valuable insights, and a connected scatterplot can be used to visualize this information and identify potential links between credit card owners and their purchases.

The goal of Task 4 is to identify informal relationships among GASTech personnel using a Clock chart and Network diagram. The visualization combines the two, with a color hue representing places and a tree network representing groups of people who met together, with arcs representing the time frame. Interactions include hovering over each leaf to display a tooltip with the name and location of the person. The chart provides an easy way to identify possible informal relationships among personnel.

Finally, We used a polar chart to detect suspicious activity by presenting the number of car stops during odd hours over a span of 14 days, filtered by location. The chart displays the number of car stops for all 14 days in relation to a chosen location, and the user can view the car ID by hovering over the circle on the chart. By combining the results with data from previous tasks, we were able to backtrack and establish evidence for any suspicious activity.

3 DESCRIPTION OF VAST MINI CHALLENGE

We aim to assist law enforcement in analyzing movement and trajectory data to gather valuable information that could aid in locating the missing employees. Based on this goal, we have identified the following analytical tasks:

- T1: Analyze credit and loyalty card data to determine frequent places and peak times, identify anomalies and suggest corrective actions.
- T2: Integrate vehicle information into credit and loyalty card data analysis, assess the impact on previously identified anomalies and identify inconsistencies or discrepancies among the data sources.
- T3: Determine the possibility of deducing owners of credit and loyalty cards with supporting evidence and acknowledge limitations and uncertainties in the approach and available data.
- T4: Detect possible informal connections between staffs with given data and present supporting evidence.

- T5: Identify 1-10 locations of suspicious activity and provide reasoning for suspicion.

4 DATASET DESCRIPTION

We utilized data from Mini-Challenge 2 of the 2021 VAST Challenge to solve various problems. The dataset included GPS data for each car, car assignments to employees, and card transaction data. Although a map image of Kronos and geospatial information about roads were provided, geospatial information about shops was missing. The data was discovered to be incomplete, it was challenging to process the data to be used.

To address this issue, we needed to know when someone visited a particular place and used which card. Since GPS data did not provide any information about the visited places, we had to process the data to extract information about the cars that stopped and parked. By analyzing the time interval of the timestamp in GPS data, we considered a car parked and stopped at a certain point if the time interval exceeded a specific threshold. We set six different thresholds, ranging from one minute to more than three hours, to determine the duration of the vehicle's stay. The "car parked point" data contained each car's ID, latitude range, longitude range, arrival time, departure time, and duration. To determine which shop each car visited, we manually found the geospatial data by overlapping map images and plotted lines of road data. By checking the overlapped area between the car's stopped area and shop location, we could map each car's visited shops. Finally, we mapped the card data to add information about the cards each employee might have used.

5 METHODOLOGY

5.1 Task 1: Identifying popular locations

Task 1 goal is to identify the most popular locations and find any abnormal pattern. To solve the task, we first need to define what is the measurement of popularity. Examples are the transaction amounts, counts of transactions, or the number of visited people. We visualized them using the drop-down on the bar chart to give insight to the users.

As a result of the visualization, we got a few inspirations. Summarizing the characteristics of each location during weekdays and weekends, it can be observed that people tend to drink coffee in the morning and have meals and coffee during lunchtime on weekdays. On weekends, people have the inclination to have lunch and dinner at restaurants, visit stores, and go to the museum or play golf. Through the bar chart, we pick out the 8 most popular locations: Katerinas Cafe, Brew've Been Served, Guy's Gyros, Hallowed Grounds, Ouzeri Elian, Frydos Autosupply n' More, Bean There Done That, and Albert's Fine Clothing. We also detected anomalies. For instance, Albert's Fine Clothing (store) had a purchase of over $1,000 between 19:00 and 20:00 on Thursday the 17th, which has not occurred at any other date or time. Frydos Autosupply n' More (supply) had a purchase of more than $10,000 at 19:00 on the 13th, which has not occurred at any other date or time.

5.2 Task 2: Incorporating vehicle data

In Task 2, we try to find out how our assessment of the anomalies in Task 1 changes based on combining the vehicle data with the analysis of the credit and loyalty card data. Once the user clicks on a location's bar in the Bar Chart from Task 1, we are using a Gantt Chart to visualize the various vehicles that visited that location on the selected day on the calendar and between the time interval shown on the time slider. We know about the most popular places at various hours of the day. Naturally, we expect to see a lot of cars visiting those locations at popular hours. If we come across a situation such that the bar chart suggests that a lot of people visited a particular location, but the Gantt chart shows only a few vehicles visiting the said location, it can be considered an anomaly. The inference that can be drawn from this observation is that someone made a very large transaction at that location which confused the bar chart since the bar chart depends on data that mimics 'usual expenditures' that the people demonstrated during the 14-day period. Additionally, the vertical bars denote transactions that occurred during the specified time period at the selected location. One of the initial observations that we made was that all the transactions that occurred in Brew've Been Served are offset from the time of vehicle visits by about 4-5 hours. This anomaly can be attributed to the payment system used in the shop.

5.3 Task 3: Owner of credit and loyalty card

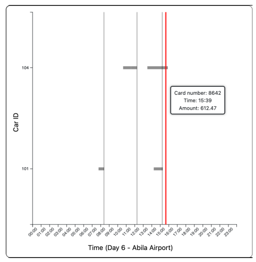

Fig. 2. Example of Gantt chart for Task 3

Task 2 and Task 3 are connected through the Car ID. When a user clicks on a car in the Gantt Chart, the trajectory of the vehicle is plotted for the entire day, including the places it visited. On the Gantt Chart, horizontal bars indicate when the cars stayed at the shop, while vertical lines show credit card transactions. By examining the intersection between the lines, we can infer the owner of each card. For instance, if there is only one horizontal line and one vertical line as shown in Figure 2, there is a high probability that the card belongs to the car owner. Users can refer to the car's route to obtain detailed information. However, the drawback of this approach is that if several individuals visit the same place together at the same time, it can be difficult to identify the owners of the cards.

5.4 Task 4: Informal relationships

The goal of the 4th task is to find potential informal relationships between the employees. To solve this task a clock chart + network diagram visualization has been chosen. The dataset is formed by combining credit card data, car assignments data, and GPS data. Through Python pre-processing, the final dataset is formed which has columns (Date, Time, AM_PM, Employee Name, Car_ID, and Location).

To be specific, the data is filtered by choosing a specific date from the date drop-down and selecting the time of the day (AM/PM) using the radio button. Then as a car from the 2nd visualization's chart is chosen, the data is filtered to show all the employees that met with the employee this car belongs to during the time of the day. The filtered data is converted into tree format, with the root node as the time and the child node as the car id. Then this tree structure is passed to another function to create the dendrogram. The root node is compared with the time on the clock and aligned accordingly round the clock. Hovering over the node shows information about the Name of the Employee and the Location visited. The different colored branches of the trees represent the different locations that have been visited. The positioning of the trees on the clock gives information about the time during which the employee visited the location.

Through this innovative visualization, we were able to specify several informal relationships. Vasco Pais (ID35), Ingrid Barranco (ID4), Ada Campo-Corrente (ID10) meet with each other at Desafio Golf Course on 12th and Orhan

Strum (ID32) joins them on the 19th. Isia Vann (ID 16) keeps visiting frydos Autosupply at different times of the day accompanied by Loreto Bodrogi (ID 15) on 8th,9th,10th and 12th.

5.5 Task 5: Suspicious Activities

We identified several abnormal activities by the employees. Here are the examples:

- On Jan 11, 2014, car ID 16 (Employee name : Vann Isia) visited Frydo's Auto Supply at 3:35 AM in the night and stayed there till 6 PM that day. Vann Isia also visits Frydo's Auto Supply almost everyday after office hours.
- Car ID 29 (Employee name : Ovan Bertrand) visited Brew've Been Served after 10 PM at night even though he had been the Katerina's Cafe just 2 hours earlier.
- Axel Calzas has spent $1239.41 at Albert's Fine Clothing on Jan 17th, such a high amount spent suddenly, leads to suspicion.
- On Jan 14th, Van Isia and Hennie Osvaldo visit Frydo's Auto Supply at odd hours.
- The meeting in the golf course between Vasco Pais, Ingrid Barranco, Ada Campo-Corrente and Orhan Strum, could have been to discuss the kidnapping under the pretext of playing golf, so Desafio Golf Course is a suspicious location as well.

6 DISCUSSION

We propose a visual analytics system that combines spatial and temporal data to improve situational awareness and enable analysis of diverse data sources. Our method was successfully used in the VAST mini-challenge 2 in 2021 to identify consumption and behavior patterns, as well as detect suspicious activities and groups. Our visualization system presents data in an integrated view, allowing analysts to examine individual movement paths and real-time spatial information, leading to better understanding of behavior patterns. Additionally, our system provides references for intelligence analysis by exploring correlations between multiple data sources from multiple perspectives. Given the complexity of analyzing diverse data sources, we suggest the following design takeaways and implications for future research. In the future, besides working on the problems above, we plan to investigate how our system solves the problems based on a more extensive database without intensely increasing the need for computational power. Furthermore, given that temporal-spatial data is widely applied in many fields, we plan to extend the current work using temporal-spatial map to other usage scenarios, such as automated driving and smart city.

7 TEAM MEMBERS

- Alimohammad Beigi (abeigi@gmail.com)
- Chaehwan Yoon (cyoon8@asu.edu)
- Jisoo Lee (jlee815@asu.edu)
- Nayoung Kim (nkim48@asu.edu)
- Rene Marceline (rimarcel@asu.edu)
- Soham Nag (snag8@asu.edu)