Crime Detection by Movement and Network Visualization

Miao Lu
Singapore Management University
lmiao.2020@mitb.smu.edu.sg

Zhang Xinru
Singapore Management University
xrzhang.2020@mitb.smu.edu.sg

Zhong Ping
Singapore Management University
pingzhong.2020@mitb.smu.edu.sg

ABSTRACT

In the roughly twenty years that Tethys-based GAStech has been operating a natural gas production site in the island country of Kronos, it has produced remarkable profits and developed strong relationships with the government of Kronos. However, GAStech has not been as successful in demonstrating environmental stewardship.

In January 2014, the leaders of GAStech are celebrating their new-found fortune as a result of the initial public offering of their very successful company. In the midst of this celebration, several employees of GAStech go missing. An organization known as the Protectors of Kronos (POK) is suspected in the disappearance, but things may not be what they seem.

In the meanwhile, we can obtain the GPS movement data of each car, credit card and loyalty card transaction data and car owner information of GAStech employees. Then we need to act as an expert in visual analytics to visualize and analysis these datasets to figure out the suspicious locations involved in this kidnapping. In the meanwhile, as students of ISSS608 Visual Analytics and Applications, we need to get equipped with related competences of identifying proper R packages and graph category, developing the prototype and performing adequate discuss and analysis, aiming to increase the efficiency of both static and interactive visualization.

1. INTRODUCTION

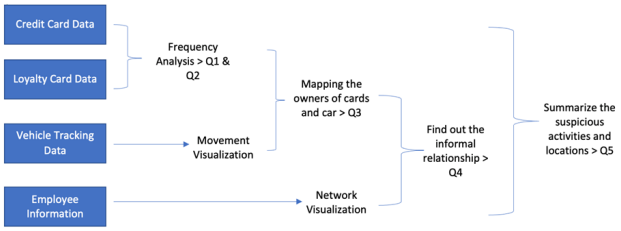
In order to find out clues and detect suspects, our project is composed of three main part.

The first one is frequency analysis, which analyzes the transaction volume of every store per day and per hour. It helps to recognize the most popular store and what time they are popular.

The second part is ownership analysis, which contains movement visualization, transaction record scatter plot and cards ownership parrel plot separately. It brings geographic data containing dates and times to life by adding it to a map as a tracking layer [2]. It also helps track employees anytime to map employee with their credit card and find suspicious activity. Then we use the parallel plot to view the ownership mapping result. Besides, it helps to compare the track of employees to find out their informal relationships.

The third part is network analysis, which helps visualize distance and density of relationships of employees interacting frequently with each other in a social network [1]. It helps to understand all informal relationships among GAStech employees and helps to find possible suspicious staff and locations.

At last, we will combine these three parts to summarize the suspicious locations.



2. MOTIVATION

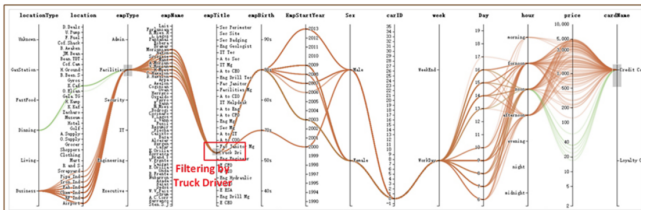
To analyze and visualize our data, we use package *shiny* to build an interactive web application via R studio.[5] When preparing data, we use *tidyverse*, *dplyr* and *tidyr* to manipulate and tidy data. We use package *clock* to work with date-time data.

In frequency analysis, we use *plotly* to create interactive graphs of customer volume.
For movement visualization, we use *tmap* to build map by spatial data distributions; we use package *sf* to encode spatial vector data; we use *igraph* to visualize data.[4]

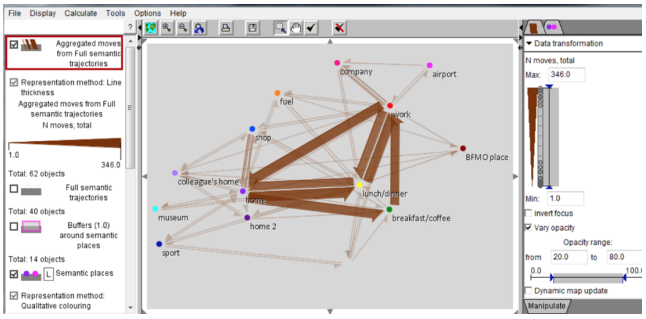
In parallel plot page, we use *cdparcoord* to build parrel chart to create parallel chart to visualize the mapping relationships among credit card, loyalty card, car ID and employee names. For network analysis, we use *tidygraph* and *ggraph* to provides a tidy API for network manipulation[3].

3. LITERATURE REVIEW

Our project drew inspiration from previous works by Mini challenge 2014, the datasets of Mini challenge 2014 is similar to the project. Team from Central South University plot the parallel graph in their project, it is just what we want to plot to show the correspondence between the loyalty card, credit card and the owner's name. However, in their parallel graph, there are too many columns included which would be confusing and meaningless.

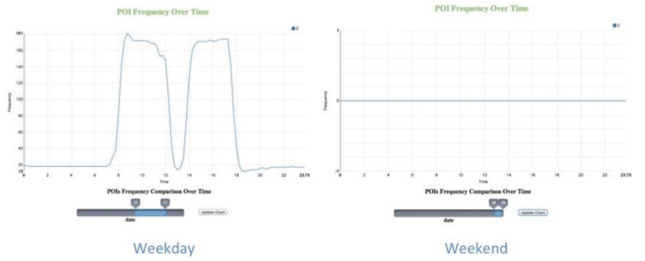


Team from Fraunhofer IAIS and City University London also implement a very significant graph called link graph. They treat place classification as the point and treat the number of movements as the link line between the points.



The drawback is that the graph would be too messy to be plotted in the link graph if the data is large enough or if the interactions between each place are too many. Plus, the graph can look better if they use shiny, since there are so many parameters setting in this tool which lack of aesthetic layout.

Team from Virginia Tech implement time series graph to plot the visit frequency of POIs (Place of Interest). This is a good way to differentiate visit times between weekend and weekday. However, one challenge is that people need to clarify and decide the POIs on their own, if the judgement of POIs is incorrect, then they are unlikely to get useful insights from the graph.



4. DESIGN FRAMEWORK

4.1 Data preparation

All the data wrangling and cleaning is performed in R. we joined the credit card data and loyalty together by same price, location, timestamp to map the ownership of credit card and loyalty card. The joined data can be used to extract the explicitly relationships between two types of cards as well as implicitly relationship.

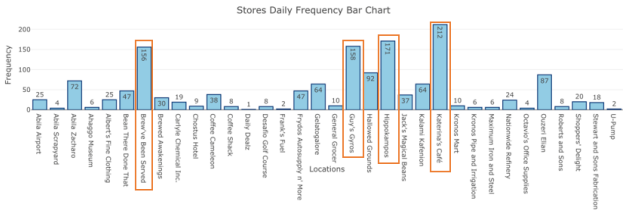
Card ownership data is a new data extracted from both joined transaction data and car assignment data. This new data contains 7columns, the first four columns is the basic information of employees, the last 3 columns are their carid, credit card number and loyalty card number. This data can be extracted with help of the map tracking and some other estimation.

Edges data is another new data we created based on the observation from mapping track and consume history. Each row represents an informal relationship we recognize. The first and second columns are employee's car ID which represents employees. The third column is the frequency of the activity. The fourth column is the type of the relationship.

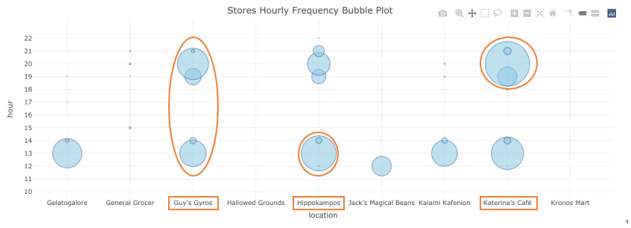
4.2 Exploratory spatial data analysis

4.2.1 Location Frequency Analysis

In this chapter, we use frequency bar chart and bubble plot to do the location frequency analysis. Among those 2 weeks, we can observe that Katerina's Café, Hippokampos, Guy's Gyros, Brew've Been Served are the more popularity locations almost in each day.

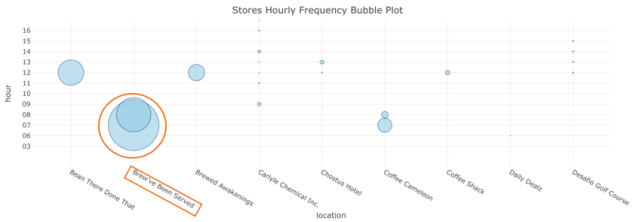


For the popular time, consumption is concentrated at 7am-8am, 12pm-1pm and 7pm -8pm, respectively corresponding to breakfast, lunch and dinner time. As for these four most popular stores, Katerina's café, Hippokampos and Brew've Been Served have no morning transaction, which means they are open from 12pm every day. Among them three, Katerina Café is most popular in 8 pm dinner time, and Hippokampos is most popular in 1pm lunch time, while Guy's Gyros is as popular at 1pm lunch time as it is at 8pm dinner time.

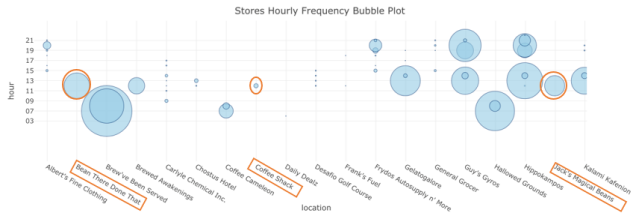


Different from above three stores, Brew've Been Served only has morning transaction, which means it may be a breakfast bar. As a result, it is most popular in 7-8am. In addition, our Shiny application also allows users to view the daily fre-

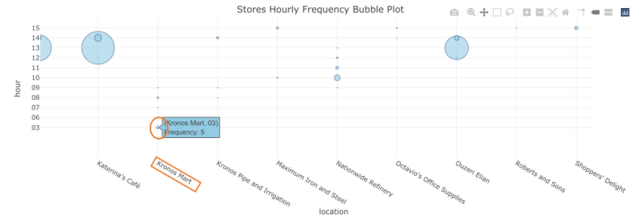
quency and hourly distribution of each store in each day, which generally follows same pattern as their overall popularity.



Furthermore, we also observed some anomalies from the frequency analysis. At first, the transactions of Bean There Done That, Coffee Shack and Jack's Magical Beans are all recorded at 12pm.



At second, there are some transaction of Kronos Mart happened at 3am, which is suspicious and abnormal because seldom mart will open till 3am.



However, adding in the vehicle data, we found not all people went to Bean There Done That, Coffee Shack and Jack's Magical Beans at 12pm, which means that we also observed some employees drove to these three stores can consumed at other time. And there is no GPS patten shows somebody went to Kronos Mart at 3am. So, it is possible that these anomalies were caused by the time of order record of these stores' ERP-POS system. In other words, Bean There Done That, Coffee Shack and Jack's Magical Beans may record their orders one time each day at 12pm, and the order record time of Kronos Mart may get delayed for several hours. As a result, these transactions cannot represent that somebody consumed at that time. As for recommendations, we suggest updating these stores' ERP-POS system to record orders more accurately or collecting the order record rules in each store to understand these anomalies better.

4.2.2 *Ownership Analysis*

Movement Visualization

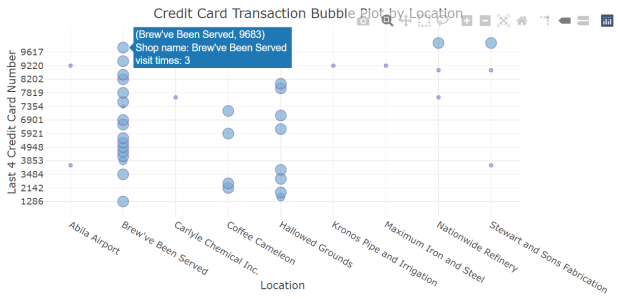
Spatial map plotting and path tracking is performed in this section, it is a frequently used graph throughout all questions. There are a lot of implicit and explicit information under the map. The interaction of the map can be split up into 3 dimensions including the carid, day of the date and hour of the date.

The map is in an overwhelming position in the project, plus, putting the map in the shiny application can highlight the advantages of it in a large degree. With the help of controller, it helps solve the location of each store in question 2, for example we realized that there are some stores which are not displayed in the original Abila map picture but there is some purchase information from credit card, in this case we can use map to track each car's transaction activities and compare it with the credit card purchase data, therefore accounting for both normal and abnormal transaction activities.

Not only question 2, but the map also includes a lot of implicit information, some of them might be the hypothesis and assumption based on the abnormal path comparing to regular routine on the daily basis. For example, CFO, COO and CEO regularly gather together in golf course at the weekend, which may indicate their relationship maybe team building or interest negotiation.

Scatter Plot

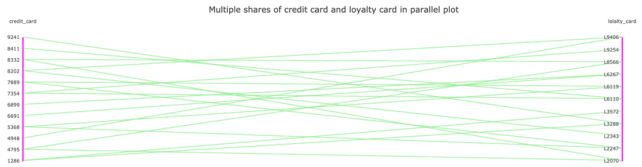
The scatter plot is the graph we create by using PLOTLY package, and it aims to show the transaction record in each location in selected time period. We put the scatter plot and the map together and use one filter of time to map the transaction activities of each car then we can get the ownership mapping of carid and credit card number.



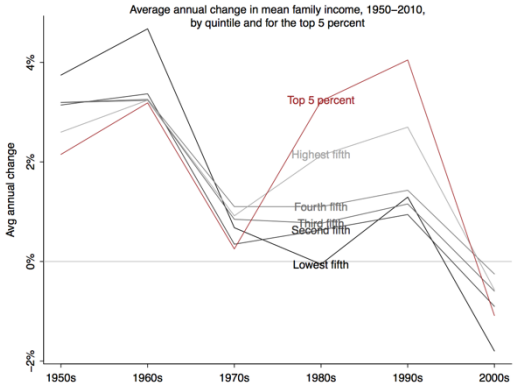
Parallel graph

After getting the ownership mapping of Car ID, Credit Card Number and Loyalty Card Number, we add two parallel graphs to visualize the ownership. The first one is about

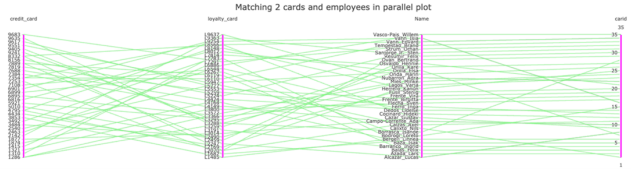
all ownership; another is for viewing the cross-used card activities, which will reveal some potential relationship. In addition, we can still select certain Car ID to view its particular information.



Extensional usage of parallel graph is to show the frequency by using the width of the line, in our cases, we just want to look at the relationship, so we did not include this function. Apart from that, parallel graph is also able to be used as a time series graph to plot the changes over time.



In our project the, based on the previous implementation and judgment, we get a comprehensive data, hence the parallel plot become more diversity compared to only use credit card data and loyalty card data, like the second parallel plot shown as below.



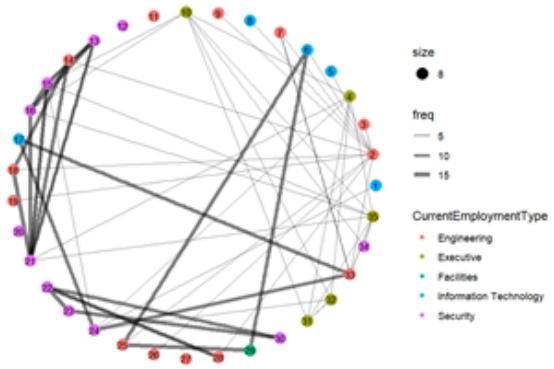
With the new graph, more information is revealed to us, not only all data matches perfectly, but this graph also helps us to figure out the suspicious behavior by combining map tracking. For example, cc 9951 credit card purchased at late night, since we have the matching graph, we are able to know who did this quickly, and along the way, we can look at the tracking path, we found out that the car did not pass by at that time, then we would say that the Kronos Mart or the person might be suspicious in a sense.

datetime	location	price	last4ccnum
2014-01-10 09:30:00	Kronos Mart	203.91	7688
2014-01-12 03:39:00	Kronos Mart	277.26	8156
2014-01-13 03:00:00	Kronos Mart	147.30	5407
2014-01-13 08:01:00	Kronos Mart	159.06	6816
2014-01-14 08:20:00	Kronos Mart	58.85	6899
2014-01-16 07:30:00	Kronos Mart	298.83	7108
2014-01-17 08:08:00	Kronos Mart	286.24	1415
2014-01-19 03:13:00	Kronos Mart	87.66	3484
2014-01-19 03:45:00	Kronos Mart	194.51	9551
2014-01-19 03:48:00	Kronos Mart	150.36	8332

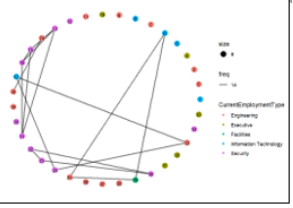
4.2.3 Network Analysis

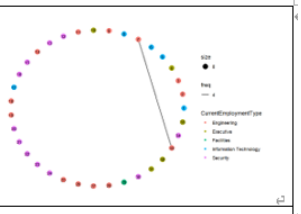
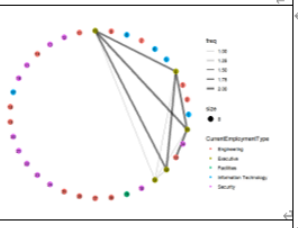
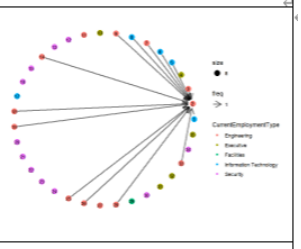
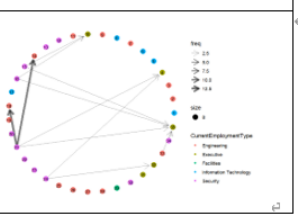
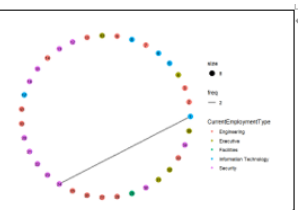
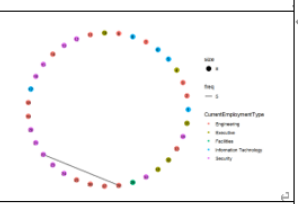
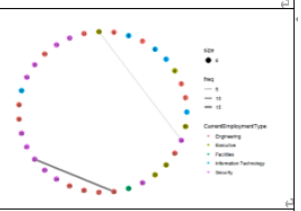
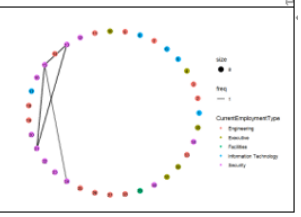
In this chapter, we will reveal some informal relationships of GAStech employees. Some relationships are detected from mapping tracking, some relationships are detected from comparing mapping tracking and transaction history.

Each node represents an employee, and each edge represents informal relationship we recognized. In this network graph, we use each employee's car id represents them. In total, there are 29 employees have 9 kinds of informal relationships out of all 35 employs. The color of node represents the department of employee. Edges' width implies the frequency of informal activities, wilder the edge is, closer the relationship is.



Overall, there are 9 kinds of informal relationship shown as table 1.

Relationship	Employee	Directed	
Live together	14,18 17,24,33 6,25,29 22,23,30 13,15,16,21	False	

Visit hotel privately	7,33	False	
Play golf together	4,10,32,33,35 4,10,31	False	
Gather at Lars's house	3->2 5->2 6->2 7->2 8->2 9->2 14->2 18->2 19->2 25->2 26->2 28->2 33->2	True	
Suspicious/midnight visit	15->10 15->35 16->10 16->35 21->14 21->18 21->35 24->32 24->4 34->35 33->35	True	
Share one credit card	1,24	False	
Eat at restaurant together	22,28	False	
Share one loyalty card	10,24 22,28	False	
Meet at wasteland	13,15 13,21 15,21 15,24 21,13	False	

4.2.4 Suspicious location analysis

Combining with all the visualization we made before, in this chapter, we will conclude top 11 suspicious locations that may be involved in this kidnapping, which are marked as red cross in the map below.

Location 1: House of owner of Car 2

As the largest concentration of activity, Azada Lars (Car2)'s house got the most visitors in the past two weeks, which involving the whole engineering team and IT team. In other words, there was a team gathering that includes 15 employees in the evening of 1.10, which raised the highest possibility of occurrence of suspicious activities.

Location 2-6: From security team's spay

From our observation, we found some members of security team including the owners of car 15,16,21,24,33,34 that was called "attacker team" in the following went to the houses of all executive members excluding CEO to monitor them, whose houses should be highly possible to become the location of suspicious activities. In addition, we found the owner of car 14 and 18 were also monitored.

- Location 2-5: Houses of executive department

The reason why we think it's a monitoring is that these surveillants usually arrived early than the house owner and left later than them, which does not fit the rules of guesting and visiting. In addition, they usually took turns to monitor executives, such as 16 came after 15 left, which is very suspicious.

In addition, the root of this kidnapping is the environmental issues, so I guess the security team was aimed to find some clues from the executives, who was thought to be informed of related issues. And this also explained why Environmental Safety Adviser got monitored by most people.

- Location 6: House of the owners of Car 14 and 18

Other than the executives, we found the house of the owners of Car14 and Car18 was also got monitored by Osvaldo Hennie (Car21) Perimeter Controller from Security team. Although the reason is not so obvious, the highly monitor

sequences that are almost high to monitor everyday make it another highly suspicious location.

Location 7: Chostus hotel

The hotel is another place that is possible for the occurrence of conspiracy because of its confidentiality. At first, Orilla Elsa (Car7) and Tempestad Brand (Car33) went to hotel together twice a week. Such a high frequency makes this place very suspicious. Secondly, Gastech's CEO also stayed in Chostus hotel, which mean that there should be the most important and confidential documents and information. And it became more suspicious that 7 and 33 went to here at Friday while CEO went out for work then, which left the space for them to slip into CEO (Car31)'s room. What will make it more possible is that 33 monitored Environmental Safety Adviser before, which means that he is the member of "attacker team", and 7 may be just his pretense of coming to hotel.

Location 8-10: Site controller's working places

In Security department, some site controllers and perimeter need to go to the site out of company to investigate the security level sometimes, which was shown at map as short time go-out without transaction records before lunch break. As a result, those sites provide a secret and reasonable place for the security team members to meet and communicate, which is mainly involved Ferro Inga (Car13), Bodrogi Loreto (Car15), Osvaldo Hennie (Car21) and Mies Minke (Car24). Among them, 15, 21,24 are all the members of "attacker team", which increased the suspicion of these site controller's working places.

Location 11: Store with high transaction frequency before the kidnapping

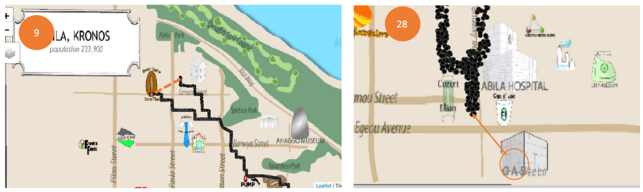
In the night at 1.19, last night before disappearance, there was a surge of transaction in Guy's Gyro that was up to 6 orders, which is far high than previous days. So, we list it as the last suspicious place, which maybe involve the suspect's final plan before the kidnapping.

5. DISCUSSION

To put it in a nutshell, we did a basic frequency of transaction data at first to conclude that Katerina's Café Hippokampos, Guy's Gyros, Brew've Been Served are the more popularity locations and they are most popular at breakfast, lunch and dinner time. In addition, we also noticed some anomalies in some stores' ERP-POS system. At second, we make a movement visualization for the vehicle GPS data to get the routes of each car. Then combining the routes of each car and each stores' transaction record, we can observe the ownership of car, credit card and loyalty card information, and we can also notice some anomalies of sharing cards. At third, combing the employee information, daily routes and cards information, we can guess their informal relationship including living together, visiting hotel privately, playing golf together, department gathering, suspicious/midnight visit, dining together, sharing cards and meeting at wasteland. At last, combing all analysis and visualization together, we summarized 12 suspicious location including house of owner of Car 2, 14, 18, and executive team, Chostus hotel, site controller's working places and some stores with abnormal transaction frequency.

6. LIMITATION AND FUTURE WORK

However, there are also some limitation of our work can be improved. At first, some employee's GPS may get broken. Car 9's GPS cannot collect all data, we can see the start point is Bean There Done That and end point is his actual residence from the patten on 1/13, which means that the vehicle data between home and Bean There Done That was gone. In addition, Car 28's GPS points were upper than the actual position, which can be proved by the fact that his parking location is not in Gastech but the upper position. As a result, we cannot get the accurate route of Car 9 and Car 28, and it is hard for us to tell it is broken accidentally or intentionally that they want to cover their real routes, which means that we cannot evaluate the suspicion of the owners of Car 9 and Car 28.



At second, the vehicle data is not continuous and lack of interactivity among different cars. In other words, there is no GPS data if the car stopped. Ideally, if we get the GPS data in very single minutes, we can select a specific time to check whether some employees are stay together. For example, if the GPS data is continuous, we can see car7 and car8 are in Katerina Café when we select a time 8:45pm in the filter, in which case we can observe the interactivities of different cars.

In the future, if these limitations can be overcome by interviewing the suspicions and implementing more data wrangling techniques, we can draw a more accurate guess and improve the fitness of our visualization application.

References

[1] A.J., H. 2021. Graph Analytics:Types,Tools and Top 10 Use cases. *Nature Materials*. 11, (Aug. 2021).

[2] Documentation, A. 2012. What is Tracking Analyst? *Nature Materials*. 11, (Mar. 2012).

[3] H. 2016. Supreme Annotationsr. *Nature Materials*. 11, (Mar. 2016).

[4] 2020. Geovisualization - an overview | ScienceDirect Topics. *Nature Materials*. 11, (Mar. 2020).

[5] 2020. Shiny - Unicode characters in Shiny apps. *Nature Materials*. 11, (Mar. 2020).