

R Shiny Developing for Analysis GASTech Employee Disappearance

LUO Haoyu
Singapore Management University
haoyu.luo.2020@mitb.smu.edu.sg

HUANG Shan
Singapore Management University
shan.huang.2020@mitb.smu.edu.sg

ABSTRACT

Developing Technologies improves the efficiency of collecting data. Big data is around our life which record our life and behaviors, while how to make better use of it and deliver a good data story is still a challenge. Utilize appropriate tools could help office bureau directly and clearly get meaningful observations and insights. Shiny which is a R package provide the function that enable users easy to build interactive web apps straight from R. it achieves the Interaction, Analyze. Communication at the same time. In this paper, the MC challenge2 of identifying anomalies and suspicious behaviors in "VAST Challenge 2021" will be as case to be analyzed. We will develop a prototype interactive data dashboard app by shiny to complete the analyze. The methodology, web interface and future work are discussed.

It consists of two paragraphs.

1. 1.INTRODUCTION

Nowadays, data as an important record medium, has been employed in various fields and industries, which is helpful to explore valuable understanding. For example, supermarket could identify different customers segment by analyzing transactions records. The government could get the traffic situation which disclosed by GPS data. Apart from it, with huge data source, combined different data has been more and more popular in data analysis. Like police station, usually analyse vast huge data to track the suspicious behaviours and clue.

Refer to the analysed case background, transaction record in loyalty and credit card, movement of car used by employee in GPS data would be provided. In order to identifying anomalies and suspicious behaviors with strong evidence, we present a user-side dynamic and interactive data analytics dashboard. We hope to provide police station and relative bureau with a behavior and geo visual analytics tool for them to detect the desired and unexpected pattern from transaction and spatial-temporal movement data.

This paper reports on our research and development process of layout and implement a web-enabled user-based visual analytics tool for desired analysis requirement and visualization delivery. The designed application includes 3 sections. Section 1 provides a basic exploratory analysis of transaction data, the bar chart used to disclose the popularity of transaction location, the density line would show the transaction distribution of day and time, link_active would be employed to connect the two different type plots. Section 2 aim to mapping the gps data and visualize the employee's life pattern. The employee's life patten would be shown by Gantt chart. It would regard as a main analysis part of detecting anomaly behaviors. The section 3 would show the methodology of refer location with latitude and longitude. The cluster method would be used to group the closing data.

Above the sections would be discussed in this report, apart from it, the motivation, review on past works, dem and future direction would be introduced in this paper.

2. 2.MOTIVATION OF THE APPLICATION

Our research and visualization work were motivated by the practical analysis requirement which are short of interactive and communicate for users to make customized input selection and combine different data analysis. It aims to support police station and relative official bureau to detect suspicious movement and transaction activity. Below is the specific planned design for the application:

- To be interactive combined transaction location plot and transaction data & day plot.
- Visualize gps data on map to support movement analysis.
- Confirm the movement stop point in gps data with specified location.
- To provide an interactive Gannt plot that can display each person's transaction history and car movement pattern, at the same time the user's target observation could connect to the map's relative observation.
- User-friendly which enable users could select input in terms theirs own demand and requirement.
- To be able to be employed in more broaden fields.

3. 3.REVIEW AND CRITIC ON PAST WORKS

Most of the past work, there are not enough design for interactive and connective. Each of the plot usually static and separately which it's hard for users to make customized analysis and comprehensive analysis.

And for our own past works, the inconsistent of design thinking and real technology skills is the main problem of the overall work. Apart from it, away from the main topic and pay more attention on details is another point that influence the whole project output.

4. 4.DESIGN FRAMEWORK.

The design of the application follows closely our data analysis process we used. Exploratory analysis on transaction data, gps data analysis, further analysis of combined data. It consists of three major tabpanels namely: transaction analysis, cluster analysis, gps analysis to disclose different observation from different aspects of the data to the user.

Firstly, the panel on the left side is the selection input area which provide the user's customized selection. On the right side, there is the main panel which mainly provide the visualization work (Figure.1).

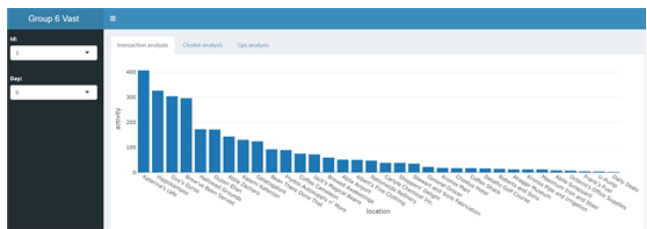


Figure.1: Overview of shiny app

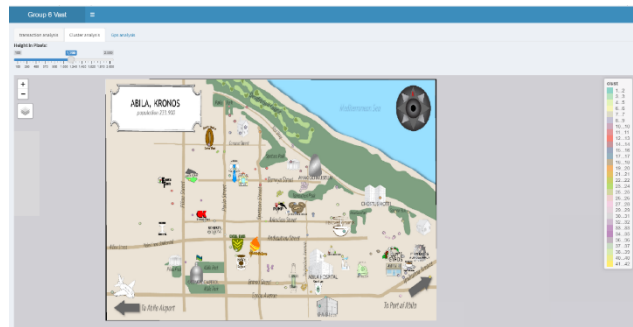
Below is the detail description:

- For tabpanel1 (Figure2), At the top most of the panel is the bar chart of transaction location frequency, by clicking different bar, below the bar chart will show two density plot which plot the transaction distribution by day and time separately in terms of the specific location selected by click.



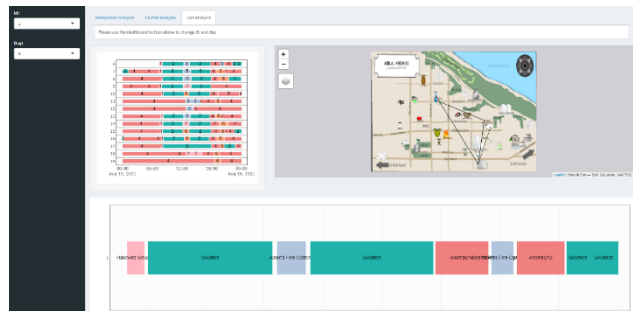
ure.2: Tabpanel1 of shiny app

- For tabpanel2 (Figure3), the top line is the selection input of zoom which could enlarge or smaller the map size. Below is a map which plot the clustered point on the map. It would be an necessary part to support users connect gps point with provided location name on the map.



ure.3: Tabpanel2 of shiny app

- For tabpanel3 (Figure4), firstly, on the top left, there is a Gantt which is an overview of each employee's movement pattern in each day. It could be a good way to find the unnormal activity compared with other days. On the top right, it's a map with movement path and stop point. On the bottom panel, there is a specified Gantt on specified day to meet user's detail analysis.



ure.4: Tabpanel3 of shiny app

5. 5.DEMONSTRATION - USE CASE

By using cluster analysis, timeline and map visualizations for the GPS data, we can get very clear visualization of each members' activity and visited place on each day. This can be also applied to various GPS data. Like tracking the route and finding anomalies for visitors in a zoo, and also the most popular route and places for a playground.

6. 6.NEW INSIGHT ENABLED IN OUR APPLICATION

The highlight insight we want to stress is the Gantt plot. We used cluster method to group the gps data point to confirm the specified location, plot the activity period by Gantt in each day for each person. At the same time, color the plot by activity type. overview of the plot, it would be more efficiency for audience to identify the anomaly in activities and time compared with other day records.

7. 7.FUTURE WORK

The current system we developed is based on a specified case, the aspect of data is limited. For the extend development, add more broaden data resources would be more powerful to support analysis. Apart from it, more interactive function added would be a plus to make more detail and comprehensive analysis.

Also, after improving our application in the future, we hope it could be used in more broaden fields apart from police station, like analyze specified area's citizen's