Group 1 ShinyVA - A Shimy Application for Crime Detection

Li Nan
Singapore Management University
nan.li.2020@mitb.smu.edu.sg

Li Yueting
Singapore Management University
ytli.2020@mitb.smu.edu.sg

ABSTRACT
This is the abstract.

It consists of two paragraphs.

1. INTRODUCTION
This paper is based on analytics for the Mini-Challenge 1 & Mini-Challenge 2 of the VAST Challenge 2021.
In a fiction scenario,a gas-production company named Tethys-based GAStech has been operating a natural gas production site in the island country of Kronos and 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. And in January, 2014, the leaders of GAStech are celebrating their new-found fortune, but 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. For this fiction scenario, we focus on mini-challenge1 and mini-challenge2 of VAST Challenge 2021.

- Mini-challenge1: Mini-challenge1 provides (adding more. . .)
- Mini-challenge2 Mini-challenge2 provides 4 datasets to data analysts for exploration. These datasets contain 3 aspects. (Transaction/GPS/Car Assignment) Through data manipulation and visualization, it can be detected anomalies that appear in unmatched transaction records between loyalty card and debit/credit card and gps data also shows the unfamiliar movement for some employees. Through the background stated, this paper develops interactive visualization approaches to provide evidence and suspicious behaviors of GAStech employees. The introduction of this paper is followed by an explanation of our motivation and objectives in Section 2, then followed by Section 3 which details the data used and methodology selected. And Section 4 provides a visual overview of the final application and finally section 5 provides the conclusion and insights for this paper.

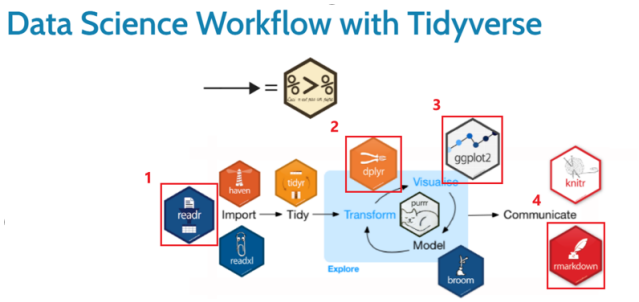
2. MOTIVATION AND OBJECTIVES
This project was motivated by a desire to i) identify the complex relationships among all of these people and organizations. ii) track data for the two weeks leading up to the disappearance, as well as credit card transactions and loyalty card usage data. To achieve these aims, the interactive tools are developed to addresses the following requirements: (adding graphs and utility)

- What is the suspicious pattern in company cars for employees' personal and professional use? - i) utilize geographical visualization techniques to create car/truck routes of GAStech employees based on gps dataset and also label parking points of car (parking more than 5 minutes) to dig out suspicious movement patterns.
- how to utilize gps data to match card owners and the debit/credit card owners - ii) Create interactive transaction data table to track unusual credit/debit card transactions records of GAStech employees and compare them with car parking points shown in visualization map.

3. METHODOLOGY
The paper utilize data in three aspects: i) Find relationships between different data tables and try to use different data manipulation methods (e.g. group and filter data) to deepen understandings of data story. ii) Based on the understanding of data, find explicit and effective data visualization methods to deal with varied data types (e.g. data table – currency data; spatial data – geospatial visualization). iii) Point out and try to give reasonable explanations and insights of data anomalies based on data visualization tools. To enable readers to have deeper understanding about data manipulation in this project, an explanation of the different methods used follows.

3.1 Data Manipulation Flow
The raw data manipulation is followed in the following 4 steps by using tidyverse package in R (shown in Figure 1.1): i) Step 1 is to read varied dataset (e.g. csv file, spatial data set and rds file) by using readr package. ii) Step 2 is focus on data preparation which was conducted in R using base R and the dplyr package to narrow the scope of the data by

summarizing data (e.g. extract transaction data by grouping location and day), convert data type (e.g. character type to factor type), and fix garbled character (e.g. Karterina's Cafe). iii) Step 3 is to use different visualization packages to create visualization graphs and tools. In this project, tmap is used for geospatial visualization and shiny is intended for interactive tools. iv) Step 4 is to write reproducible report by using r markdown and create poster by using posterdown package.



References

[1] Fenner, M. 2012. One-click science marketing. *Nature Materials*. 11, 4 (Mar. 2012), 261–263.

[2] Meier, R. 2012. *Professinal Android 4 Application Development*. John Wiley & Sons, Inc.