2021 VAST Mini-Challenge 3

LI Zhen, Tang Yue, Qian Ziwei
[1] *Computing and Information System, Singapore Management University*

Introduction

In the country island of Kronos, the increasing noxious effects on health and farming have been related to the uncontrolled activity of GAStech, a natural gas operator, supported by corrupt government officials. On January 20th, 2014, a corporate meeting is held to celebrate the new-found fortune because of the initial public offering of the company. However, a series of rare events occur that lead to the disappearance of several employees. The Protectors of Kronos (POK), a social movement organization that has been fighting against water contamination and government corruption, is suspected in the disappearance.
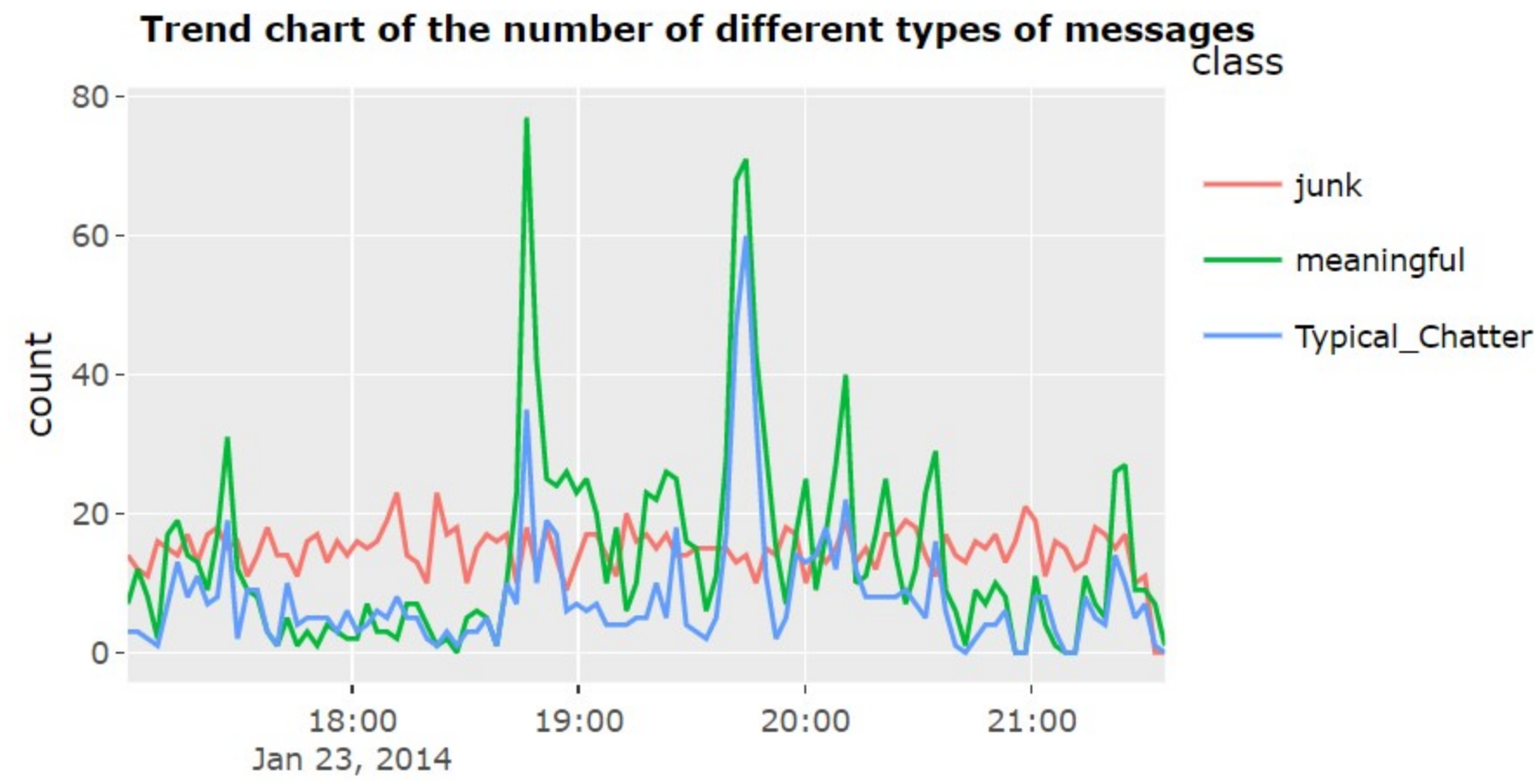
Objectives

- Identify meaningful, junk and chatter messages.
- Evaluate and show the risk level of events.
- Decide target places of responders.

Message Differece

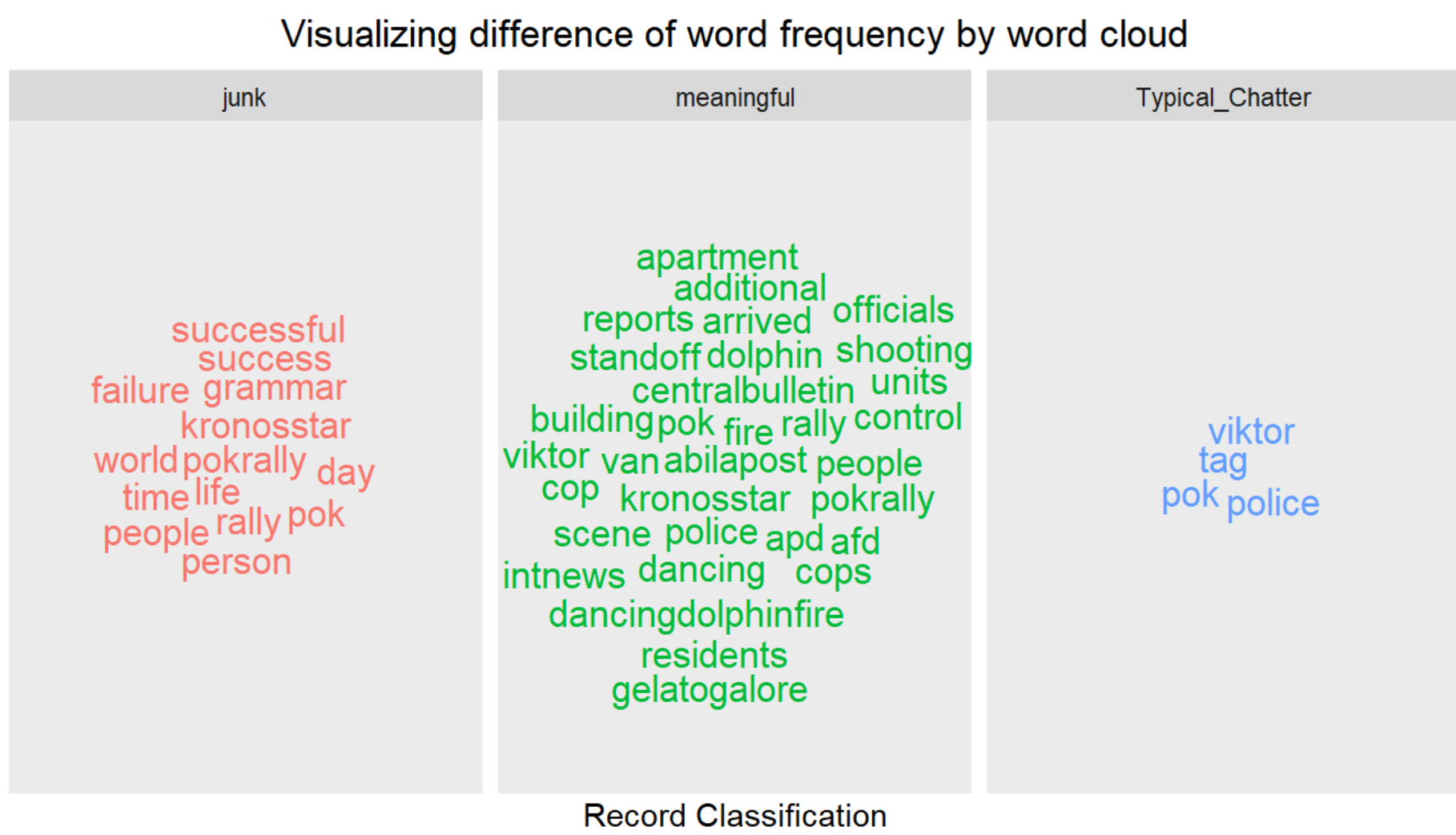| Class	Description
Junk	Advertisements or financial purpose reports.
Meaningful	Informative records which spread real news about events happen in Aliba.
Typical Chatter	Represents no inappropriate messages post online.

For meaningful news, there are two peaks around 18:40 and 19:40, and the corresponding events are the dancing dolphin fire incident and the black truck incident.For typical chatter, the peak at 19:40 coincides with the peak of meaningful messages.For junk, there was no obvious peak in the number of messages during the night period.
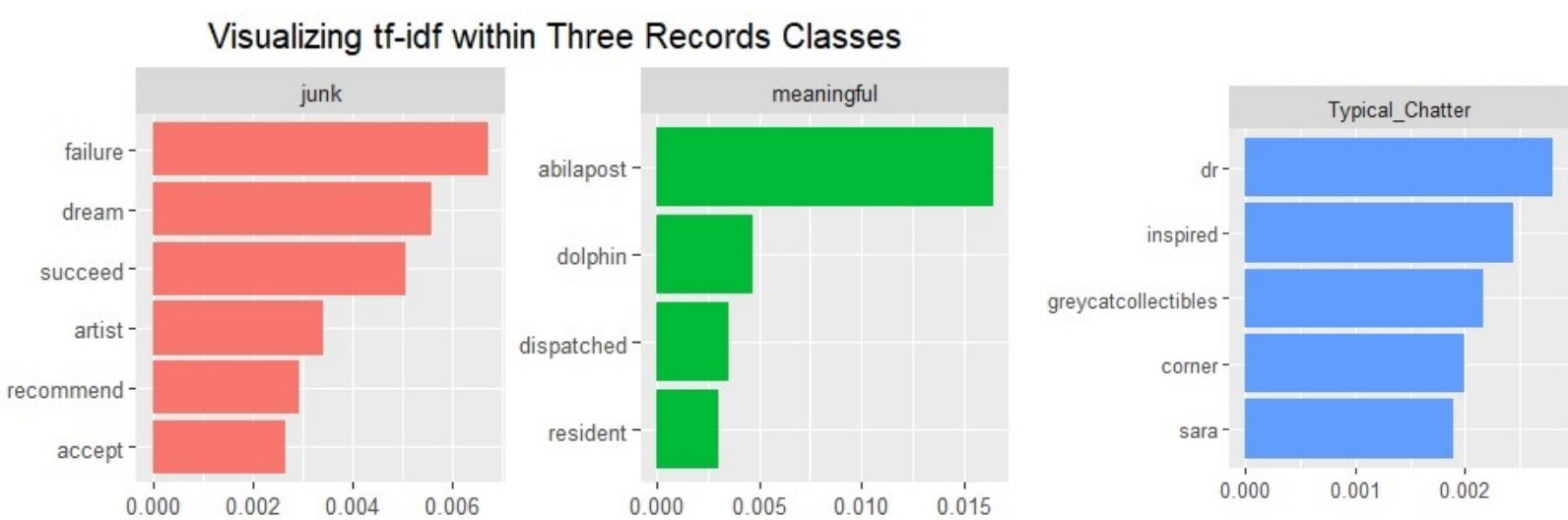

1.The difference of key words
1.1 Single word frequency

Junk records is more macroscopic. Meaningful records contains important social-event topic words, such as "fire,""rally," and "van." Comparably, the common words of typical-chatter are fewer and people usually end with a metatag "tag" if they post typical-chatter records.
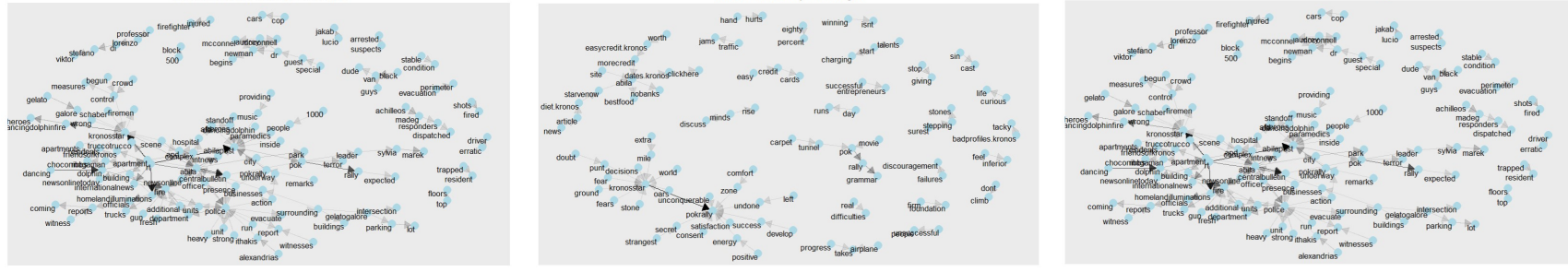

We also performed TF-IDF to find the less frequent but important words for each classification.

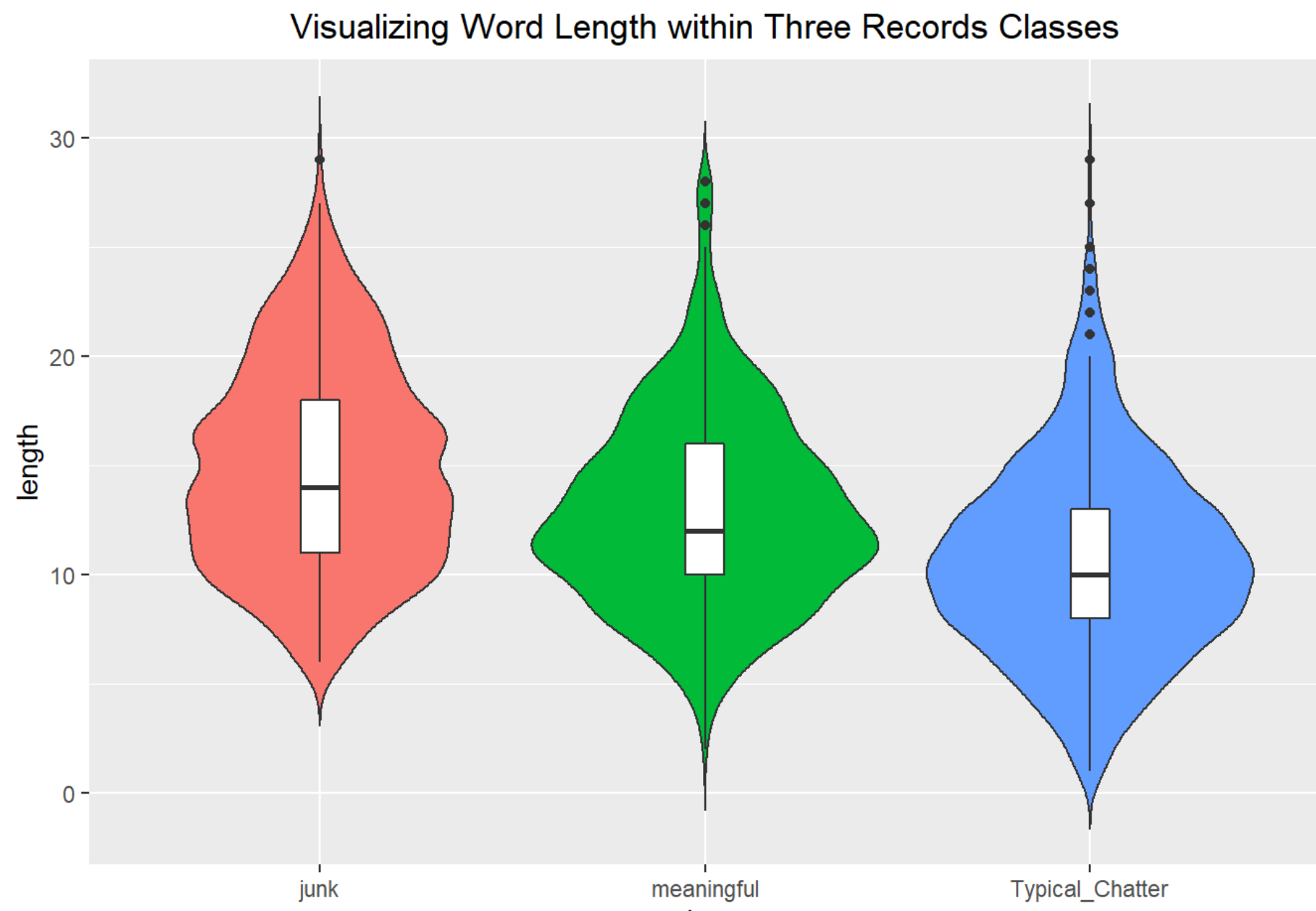### 1.2 The difference of co-occur words

From the charts, we can find that there are clear arrows in the meaningful messages and typical chatter messages while junk messages have least co-occur words.
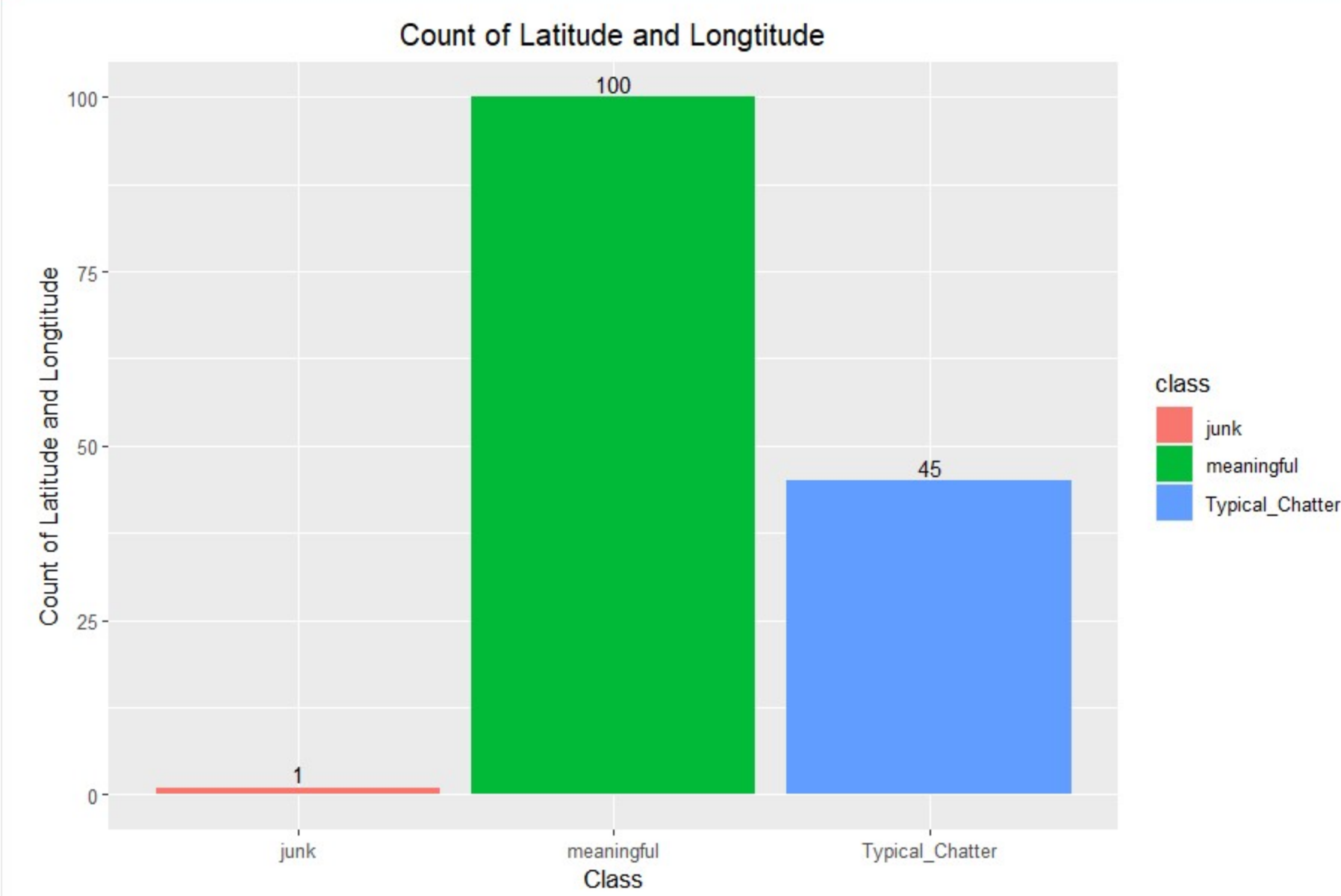

2. The difference of sentence length

The median of sentence length for Junk is largest. Moreover, the shape of the distribution indicates the weights of Typical_chatter and Meaningful record are highly concentrated around the median, but this is not distinct for Junk records.

3. The difference of geo information

In the messages blog dataset, we found that meaningful messages tend to have more geo information while junk messages have least geo information.
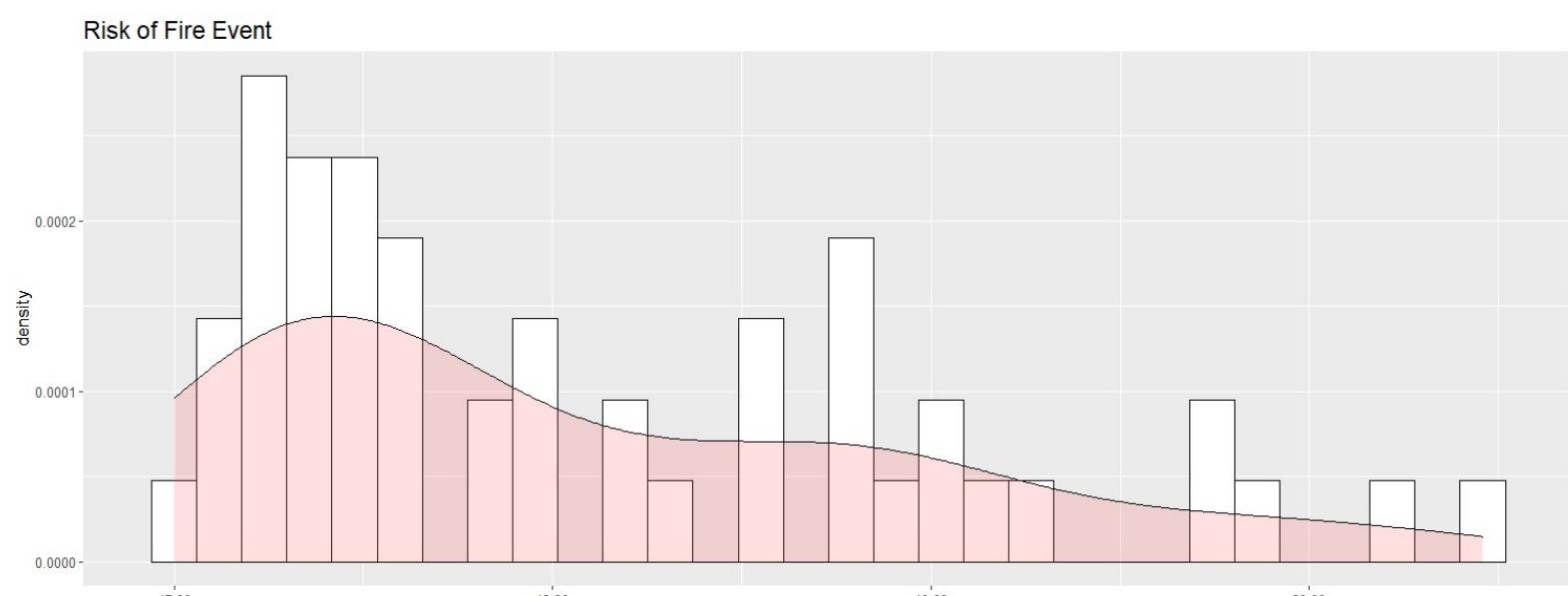

Risk of events

According to the Microblog data and emergency transcripts provided in dataset, there are 3 most risky events happened from 2014-01-23 17:00:00 to 2014-01-23 21:23:00. Based on the event developments, multiple phases are divided in a series of these incident processes.
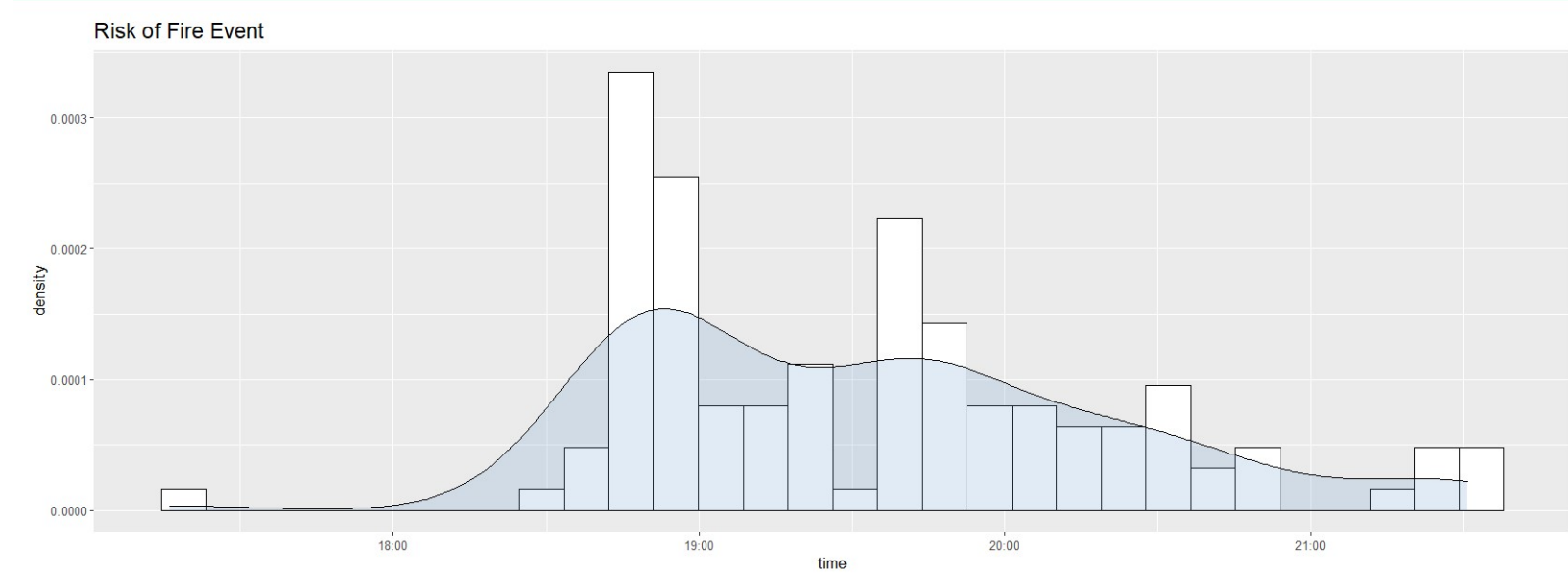
Event 1: POK Rally in Rally Park

With the events development, rally event experienced 3 different phases.

In the first event, we divided the phases based on speakers of speech. We found there are 5 important person, Sylvia Marek, Audrey McConnel Newman, Lucio Jakab, Lorenzo Di Stefano, Viktor-E.

Event 2: Fire Accident in Dancing Dolphin

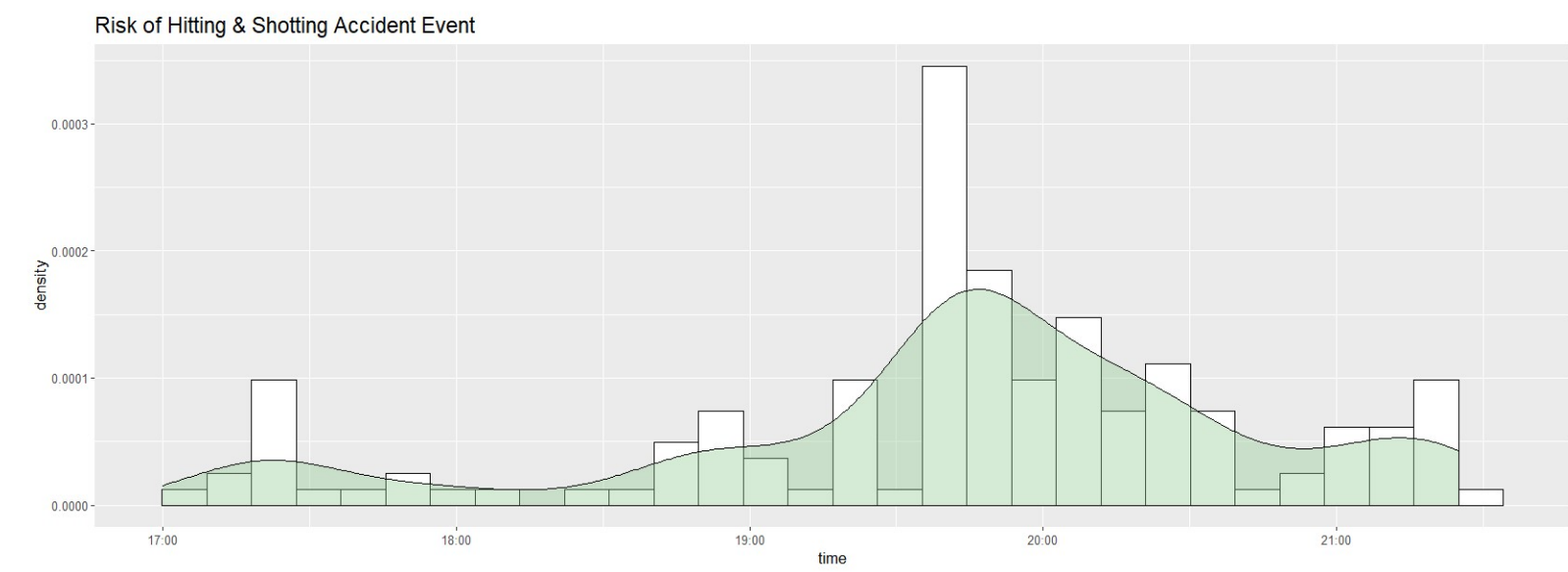The fire event started at 18:25 and involved a fire at the Dancing Dolphin apartment complex,experiencing 4 different phases.

- Phase 1: at 18:25pm, Fire Department got the fire call at Dancing Dolphin apartment.
- Phase 2: Fireman and police came to put out fire and evacuate the residents.
- Phase 3: around 20:25 officials stated fire was under control, but AbilaPOst reported that the top floor of dancing dolphin collapsed at 20:52.
- Phase 4: at 21:30, AFD reported an explosion at the Dancing Dolphin.
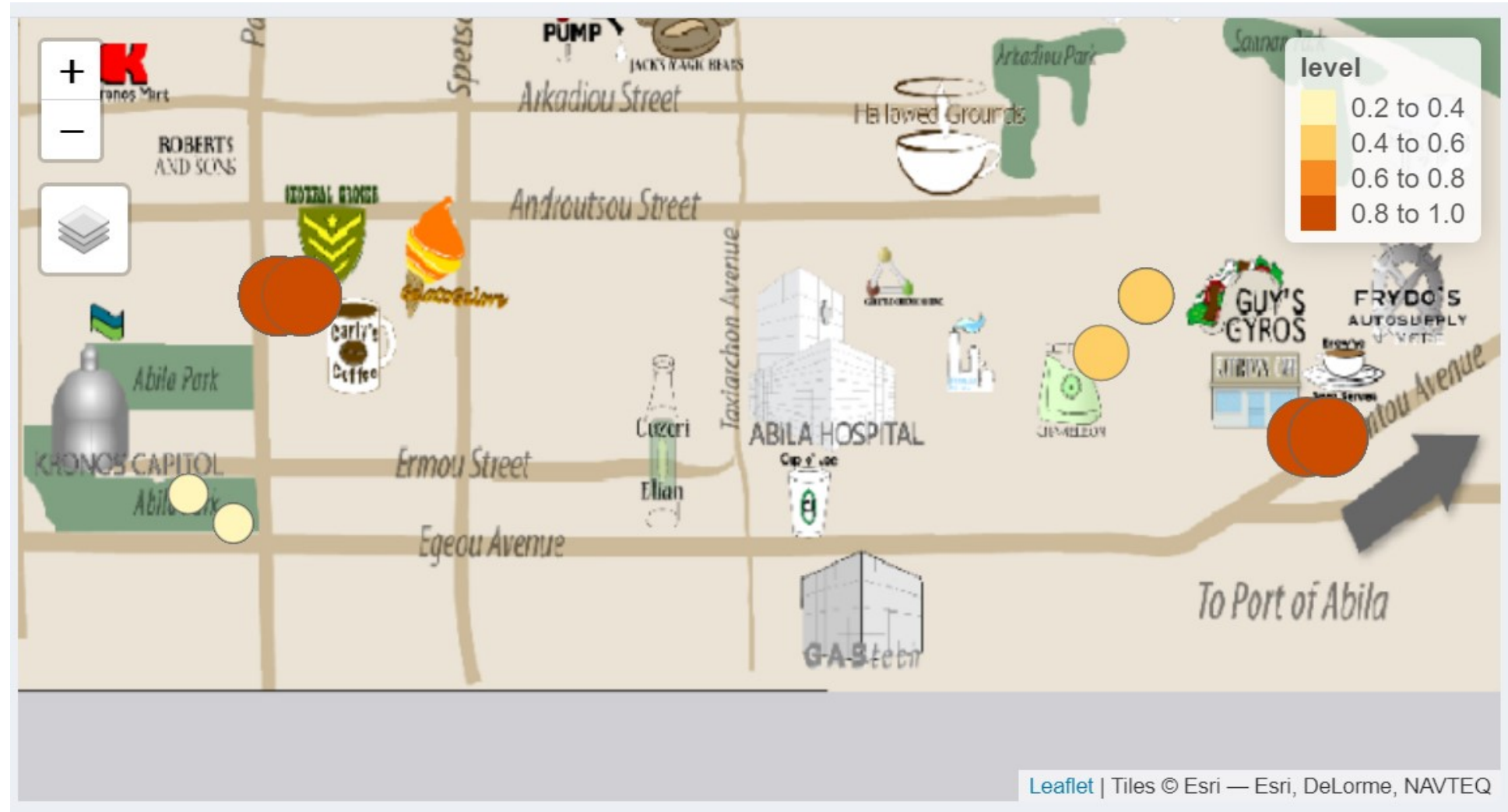
Event 3: Hitting & Shooting Event

Based on the event processes,hitting and shooting event could be divided into 4 phases.

- Phase 1: at 19:19, black van hit his car and later hit a bicyclist, then van escaped.
- Phase 2: police was chasing the crimes through the avenue.
- Phase 3: there was a shooting happened between polices and crimes, a cop was gunned down.
- Phase 4: at20:11, negotiations between shooter and police began. The standoff continues until about 21:00.
- Phase 5: at 21:13, two hostages were saved and two terrorists were arrested.

Responders

To send the responders more efficiently, we decided to send them to dancing dolphin department first. Since the there was fire happened there, the risk of this area is one of the highest. Moreover, compare to the "Black Van" event, "Dancing Dolphin" event is the first severe event happened on Jan 23, 2014. Besides the factor of timing, the geo distance between dancing dolphin and black van event is low. Therefore, it will be easier for responders to stop black van in the first stage of event and prevent the more severe results.

Conclusion

In this project, we utilize text analytics and visualization methods to present how to identify the Microblog messages from junk, meaningful and typical chatter, which would helpful for polices to tell useful information and discard unless information.

The 3 most risky events are simulated and evaluated from 17:00 to 21:30 pm. Since the fire event and hitting & shutting event have the most significant impacts for social security, it is recommended that polices and fireman are ready at all time to the region of Dancing Dolphin Apartment Complex and Souliou street between 17:00pm and 21:30pm. Fire Department prepare enough fire trucks and force for Dancing Dolphin Apartment and avoid fire spreading. For hitting & running event, it is quite important to restraint the escape of terrorists in hitting accident, which could avoid the shooting event.