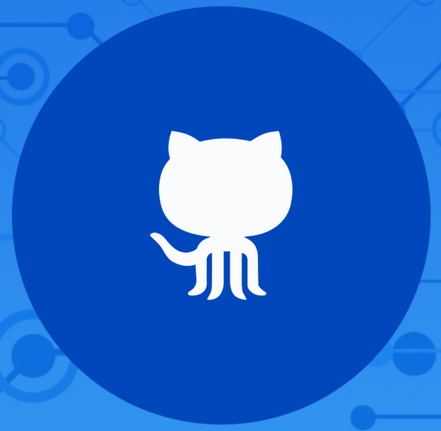


Side Project-Data Visualization

Event Trigger Root Cause Analysis

Roger Hao-Chuan Hsieh

Table of Content

Executive Summary

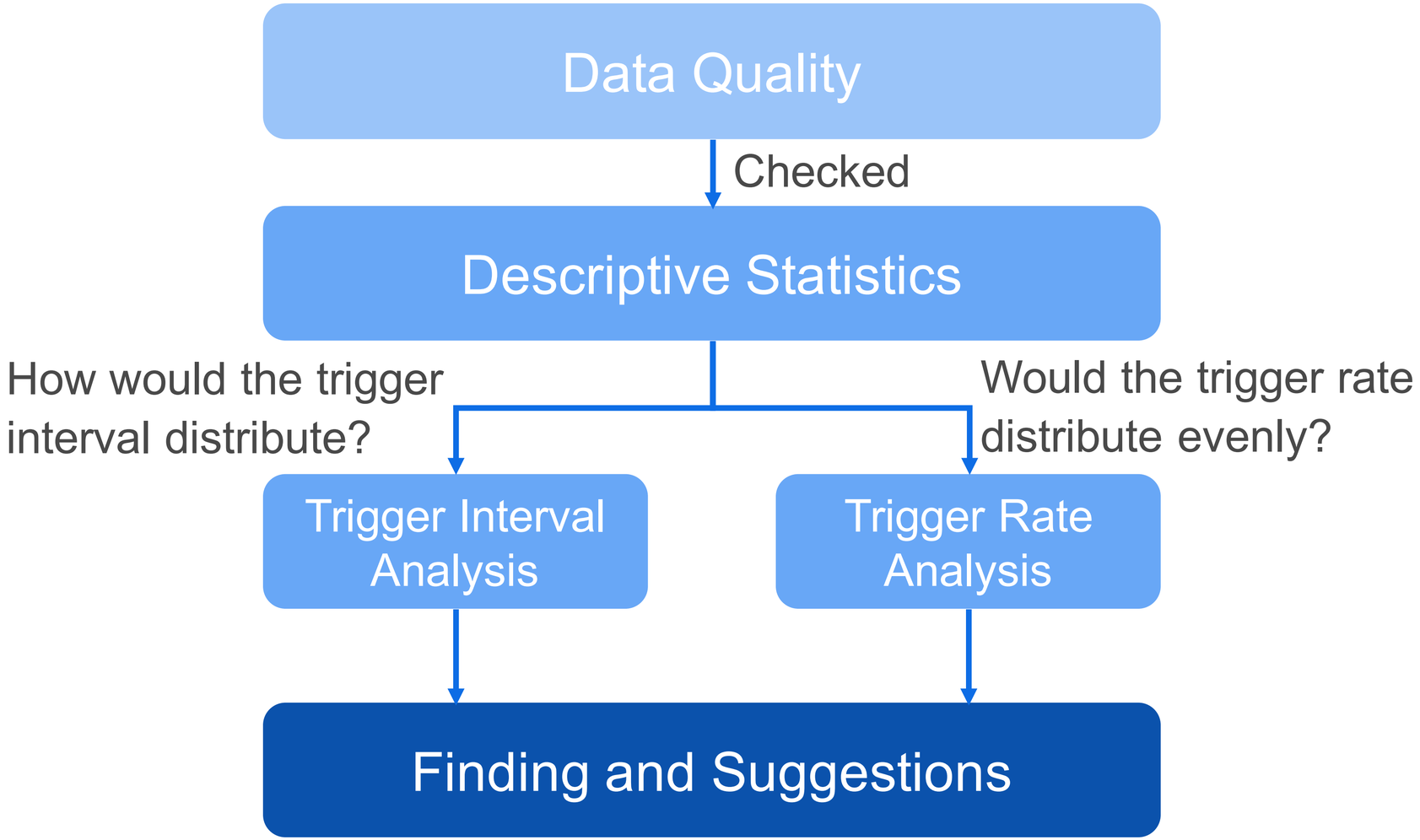
Executive Summary

- Event A log dataset
  - Cluster
  - Machine ID
  - Tracking ID (Primary Key)
  - Event Time
- Event B log dataset
  - Tracking ID (Primary Key)
  - Event Time

Analytics Object:
By using tracking ID as the primary key, we could join up two datasets and analyze the output pattern of B to see what the trigger point would be.

Analytics

Analytics Framework



Data Quality

The tracking ID columns in both event A dataset and event B dataset are checked and they are unique.

The observation count for event A dataset and event B dataset are recorded and the statistics are as follows:

- Event A log dataset
  - 5,256,536 observations
- Event B log dataset
  - 6,008 observations

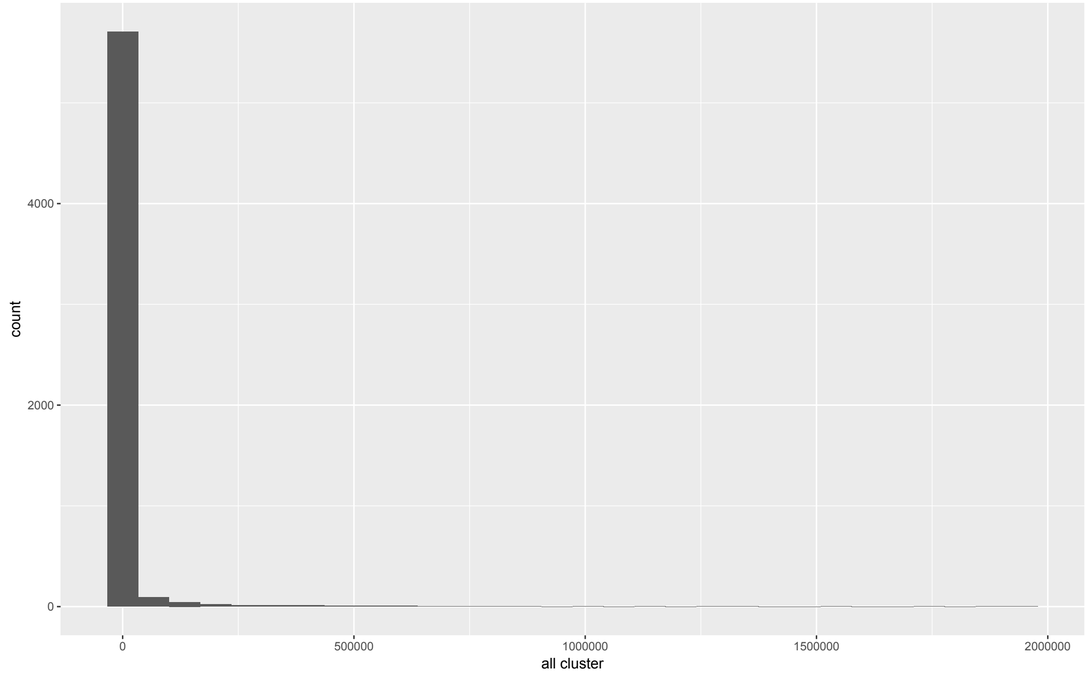
Descriptive Statistics

Cluster*	Event B Count	Event A Count	Event B Trigger Rate
6e65c	1,171	2,694,438	0.04346%
9a996	995	499,314	0.19927%
bafb4	835	112,645	0.74127%
e5818	1,411	494,709	0.28522%
b3a01	234	69,738	0.33554%
…	…	…	…

Considering both event B count and event B trigger rate, these 5 clusters are chosen as key clusters to be further analyzed.
Note that there are 66 records of event B not triggered by event A.

*Cluster names are abbreviated by their first 5 characters.

Trigger Interval Analysis



After our analysis, 90% of event B would be triggered within 395 seconds, and there is no significant difference among clusters.

Trigger Rate Analysis

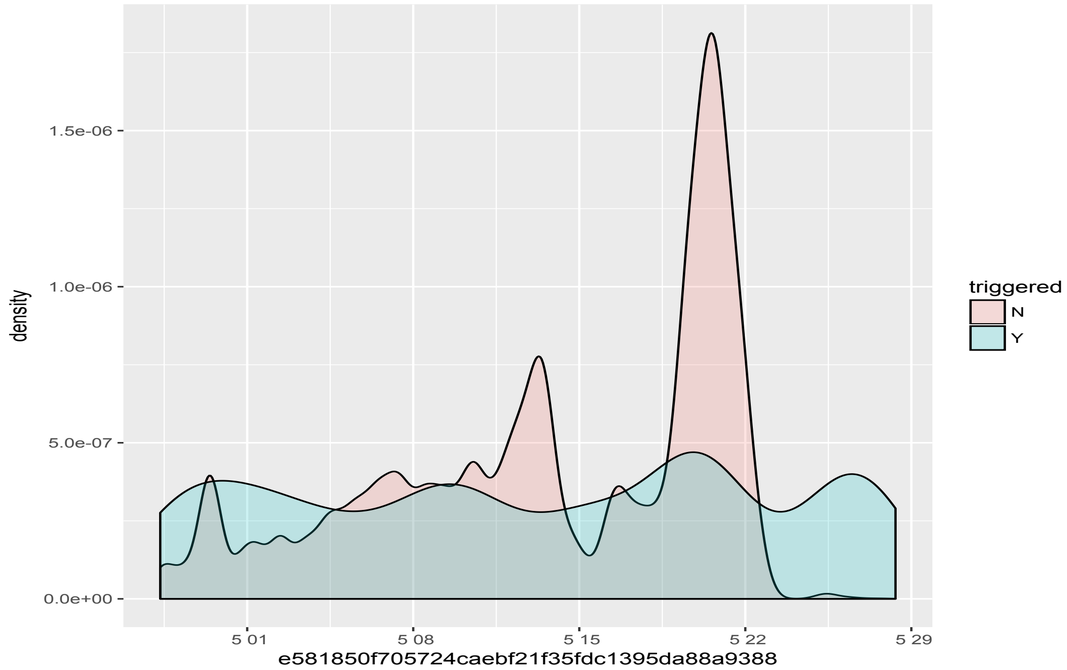
In this phase, we would visualize the density functions of event A frequency and event B frequency.

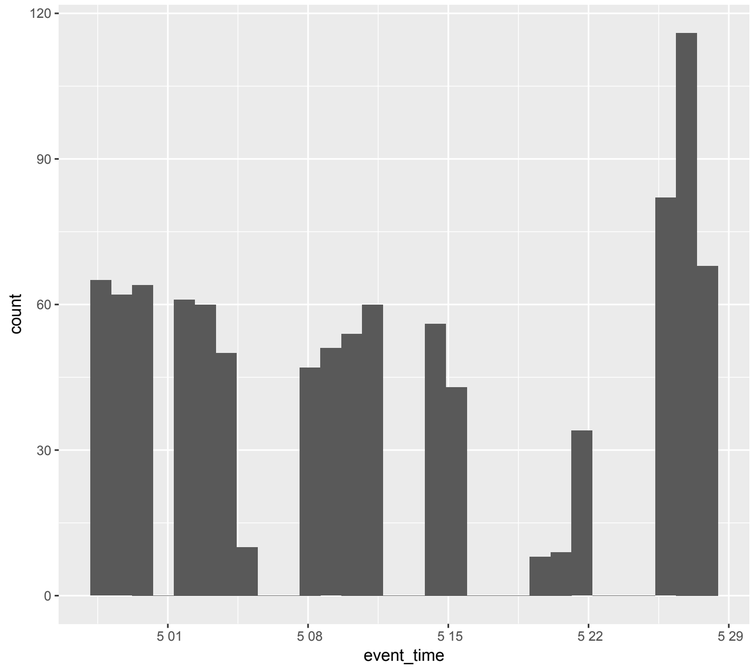
Consider the analysis result of trigger interval that 90% of event B would be triggered in a short period. The following clusters show special patterns and we would further investigate:

- e5818
- b3a01
- 6e65c
- bafb4
- 9a996

Trigger Rate Analysis

Cluster:e5818

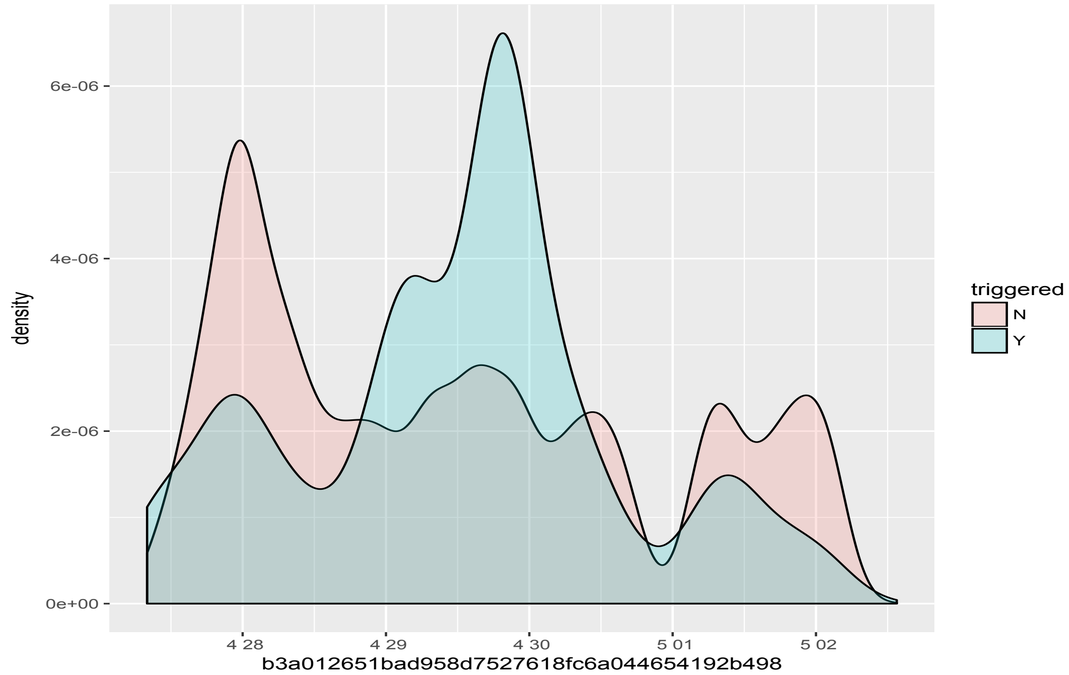


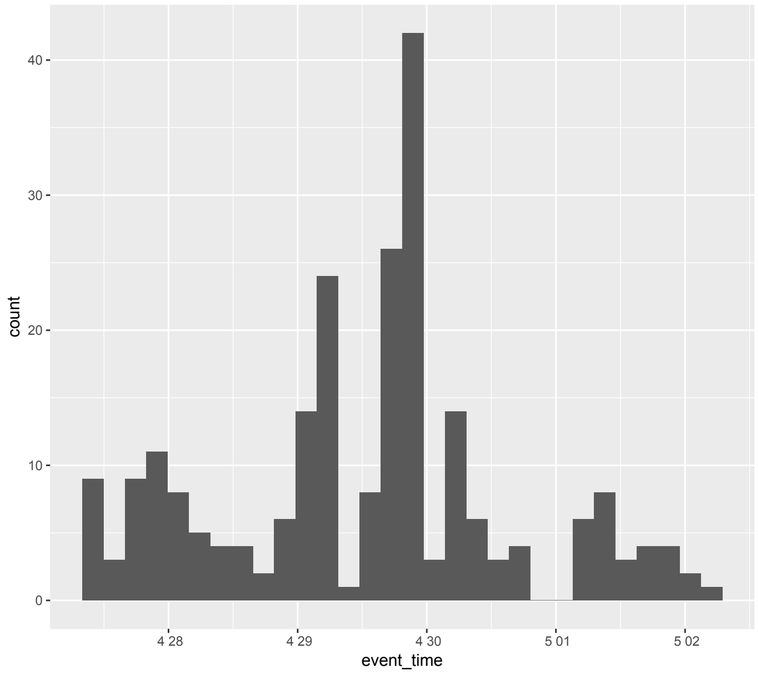


As shown in the density plot, the triggered frequency of event B is irrelevant to event A. We find that there is a key machine, machine A*, triggering most of the event B and the pattern matches.

Trigger Rate Analysis

Cluster:b3a01

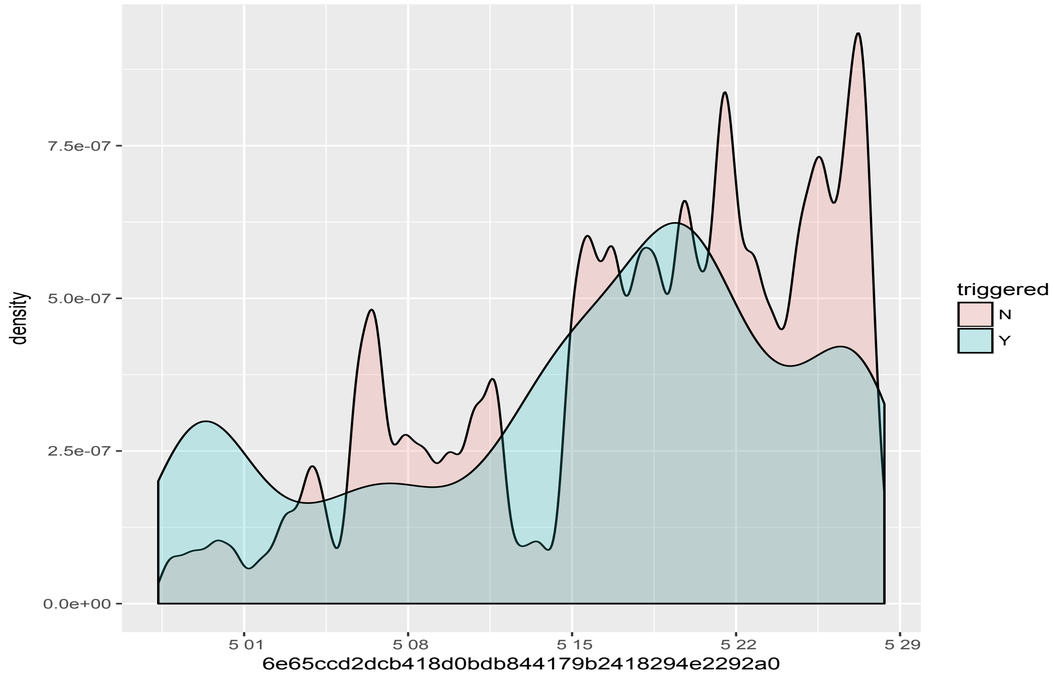


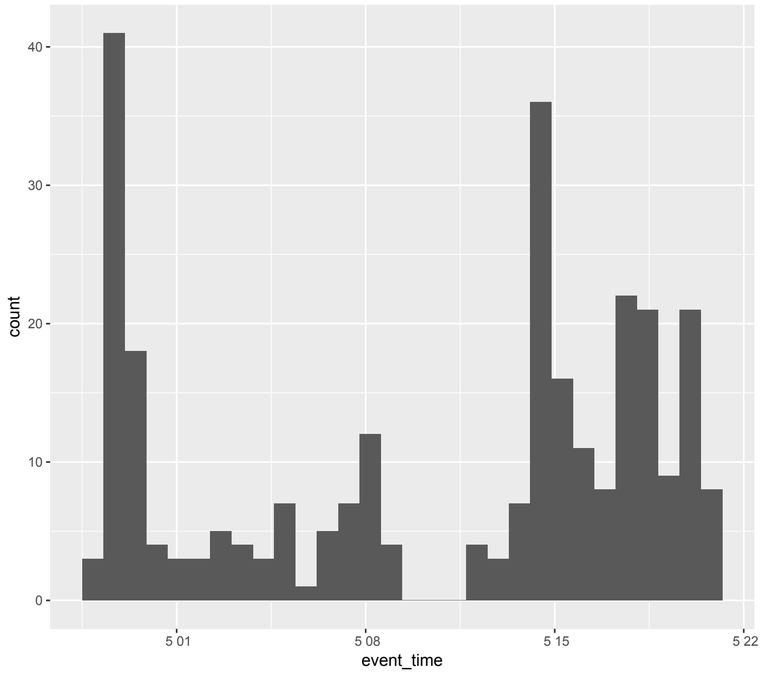


As shown in the density plot, there is a peak in late April for event B. We find that there is a key machine, machine B*, that its pattern totally fits the density figure of event B.

Trigger Rate Analysis

Cluster:6e65c

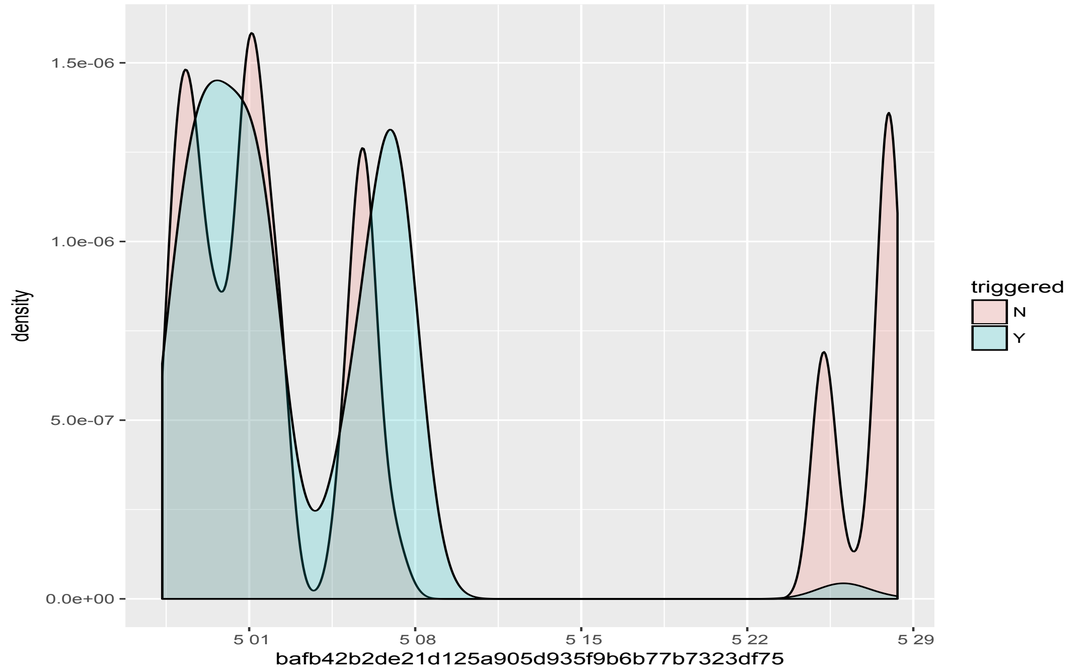


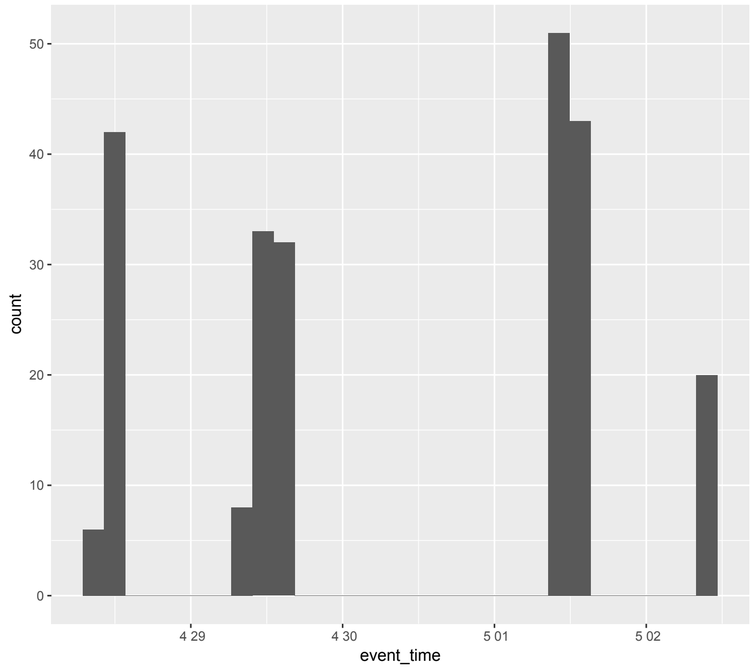


As shown in the density plot, there is a small vertex in late April for event B. We find that it is mainly contributed by machine C* and the pattern matches moderately.

Trigger Rate Analysis

Cluster:bafb4

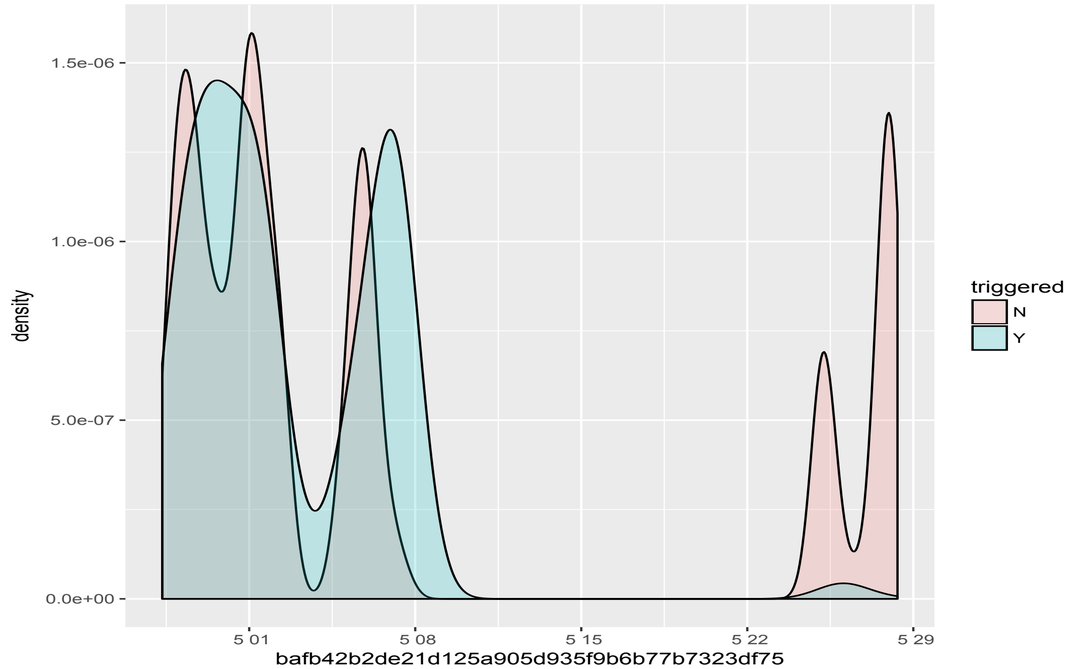


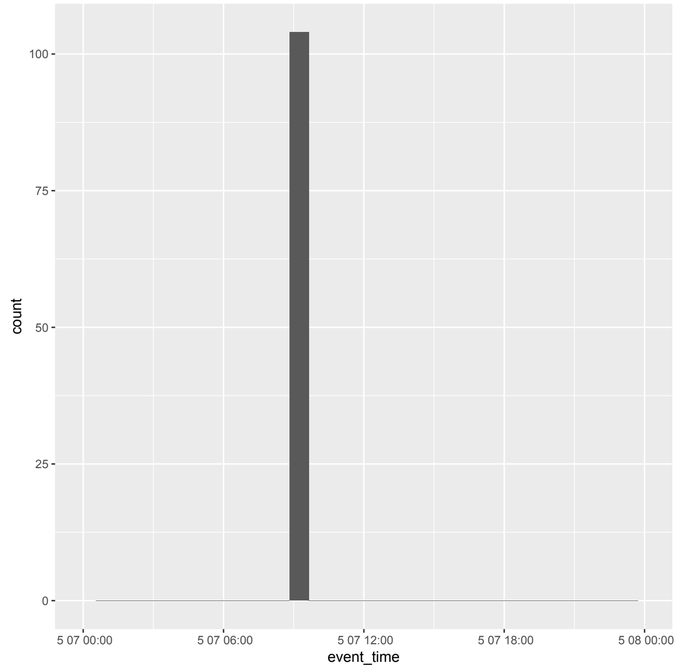


As shown in the density plot, there is one peak in late April for event B and it seems irrelevant to the frequency of event A. After our investigation, the peak is mainly contributed by machine D*.

Trigger Rate Analysis

Cluster:bafb4

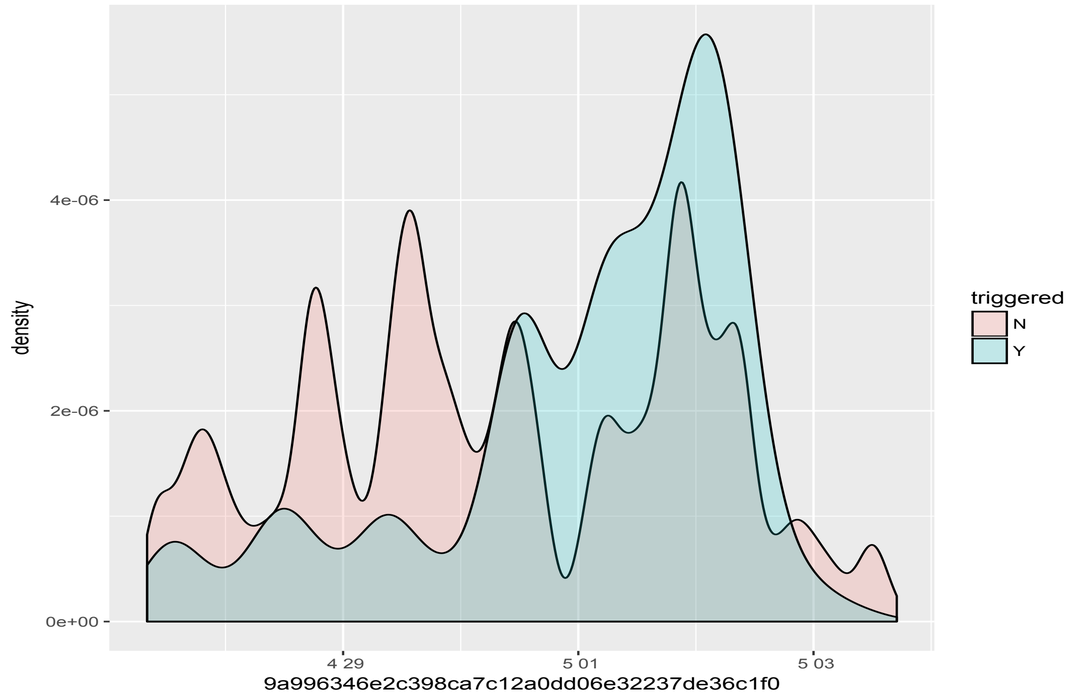


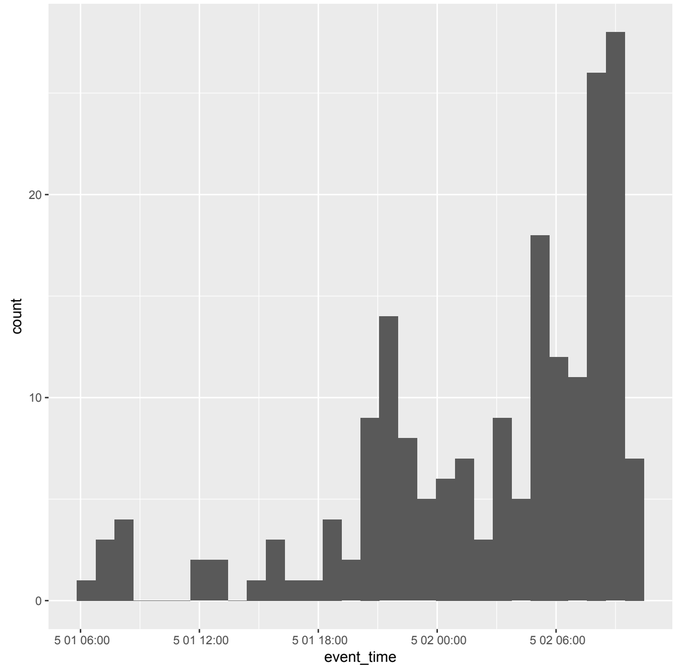


For another peak in early May, it is mainly triggered by the key machine, machine E*. There are 104 records of event B triggered within 36 minutes.

Trigger Rate Analysis

Cluster:9a996





As shown in the density plot, the triggered frequency of event B is not highly correlated with event A. We find that there is a key machine F*, of which pattern fit the distribution of event B.

Findings and Suggestions

Findings and Suggestions

According to all the analysis results, the following is our key findings:

- 90% of the event B would be triggered in a relatively short period. (395 seconds)
- There is a small portion of event B not triggered by event A. (66 records)
- There are some key machines triggering the major portion of event B.

Our suggestion would be further monitoring the behavior patterns of the key machines to improve our understandings of why event B would be triggered.

Appendix

Appendix

Key Machine

Machine	Actual Machine ID
Machine A	1645cbc3cfb3dc782fcd0816850c789196133fe1
Machine B	9a5d2ac40cfe19e56f625bb803fd7086882e660b
Machine C	cb1c56ec09f95d8df5496de8776dc26b07eb7ad9
Machine D	00d4ceae26b5905db33bca5a653c493b6df6ad01
Machine E	9e9ee26d5678cc7dda90c480e727f0e1032dc9df
Machine F	4559f6de0a7973aa141a02b53a251fed722a1473

Thank you for your attention!

Contact me via my LinkedIn: https://www.linkedin.com/in/hao-chuan-hsieh-2a5703113/