

고려대학교
빅데이터 연구회

KU-BIG

Credit Card Fraud Detection

유현우 박정진 정희정 송예은 심정은 양수형



목 차

I

Outlier / Novelty / Anomaly

II

About Dataset

III

Approach

1) Outlier / Novelty / Anomaly

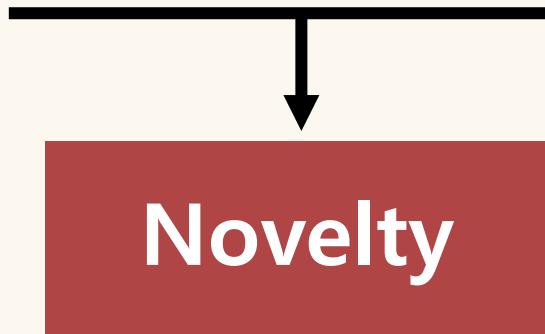
Outliers are also referred to as abnormalities, discordants, deviants, or **anomalies** in the data mining and statistics literature.

(Source: "Outlier Analysis" (Springer), Charu Aggarwal, 2017, <http://charuaggarwal.net/outlierbook.pdf>)

Outlier = Anomaly

2) Novelty detection

The training data **is not polluted by outliers** and we are interested in detecting whether a **new observation is an outlier**.



(Source: https://scikit-learn.org/stable/modules/outlier_detection.html)

Novelty Detection 활용 데이터

The screenshot shows a dataset page from Kaggle. At the top left is a red 'Dataset' icon. To its right, the text 'Dataset' is displayed. In the top right corner, there is a button with an upward arrow and the number '2842'. Below these, the title 'Credit Card Fraud Detection' is prominently shown in large white text. Underneath the title, a subtitle reads 'Anonymized credit card transactions labeled as fraudulent or genuine'. A blurred background image shows several credit cards. On the left side of the main content area, there is a blue icon of a line graph. Next to it, the text 'Machine Learning Group - ULB' is followed by a small dot and 'updated a year ago (Version 3)'. Below this section, there are several navigation links: 'Data' (which is underlined in blue), 'Kernels (2,031)', 'Discussion (40)', 'Activity', and 'Metadata'. To the right of these links are two buttons: 'Download (66 MB)' and a blue 'New Kernel' button. Further down, there is a section with a balance scale icon and the text 'Database: Open Database, Contents: Database Contents'. To the right of this, there is a tag icon followed by the text 'finance, crime, machine learning'.

Dataset

Credit Card Fraud Detection

Anonymized credit card transactions labeled as fraudulent or genuine

Machine Learning Group - ULB • updated a year ago (Version 3)

Data Kernels (2,031) Discussion (40) Activity Metadata Download (66 MB) New Kernel :

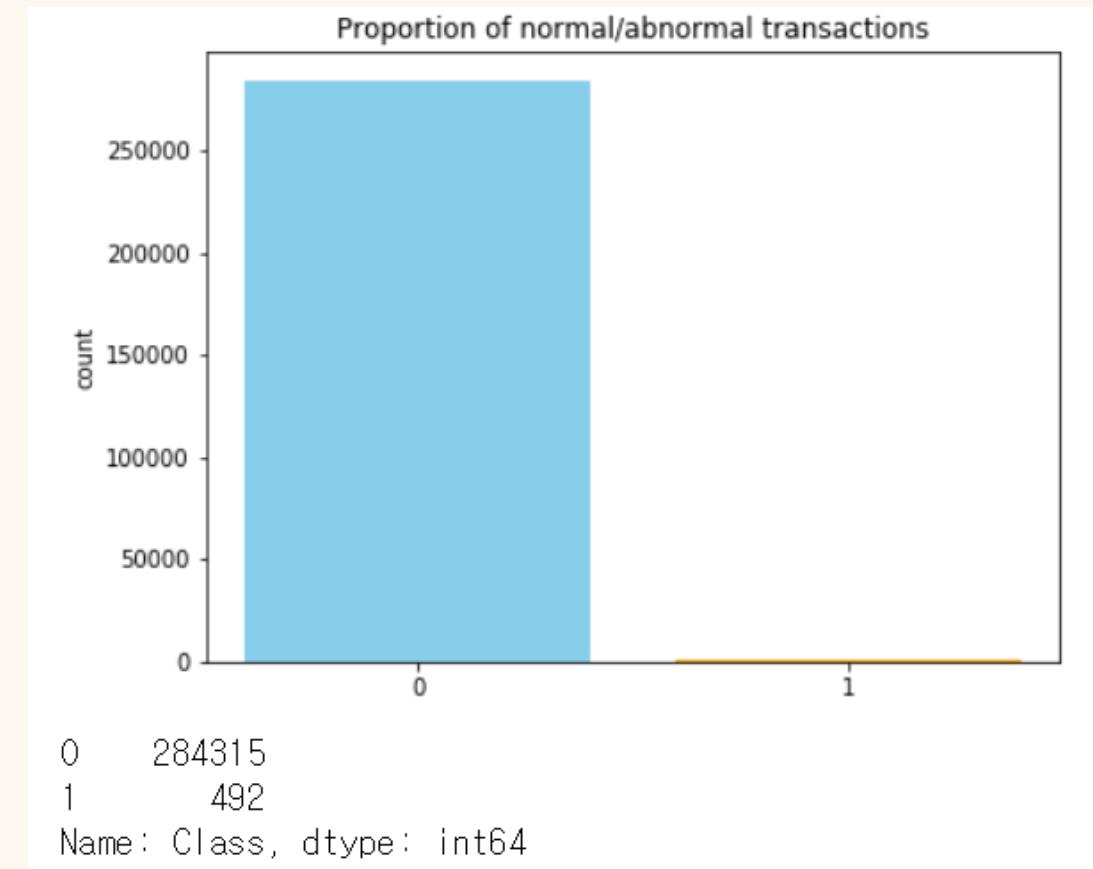
Database: Open Database, Contents: Database Contents

finance, crime, machine learning

1) Description of Data

Credit Card Dataset : (From Kaggle)

- Highly unbalanced data
- 492 frauds out of 284,807 transactions

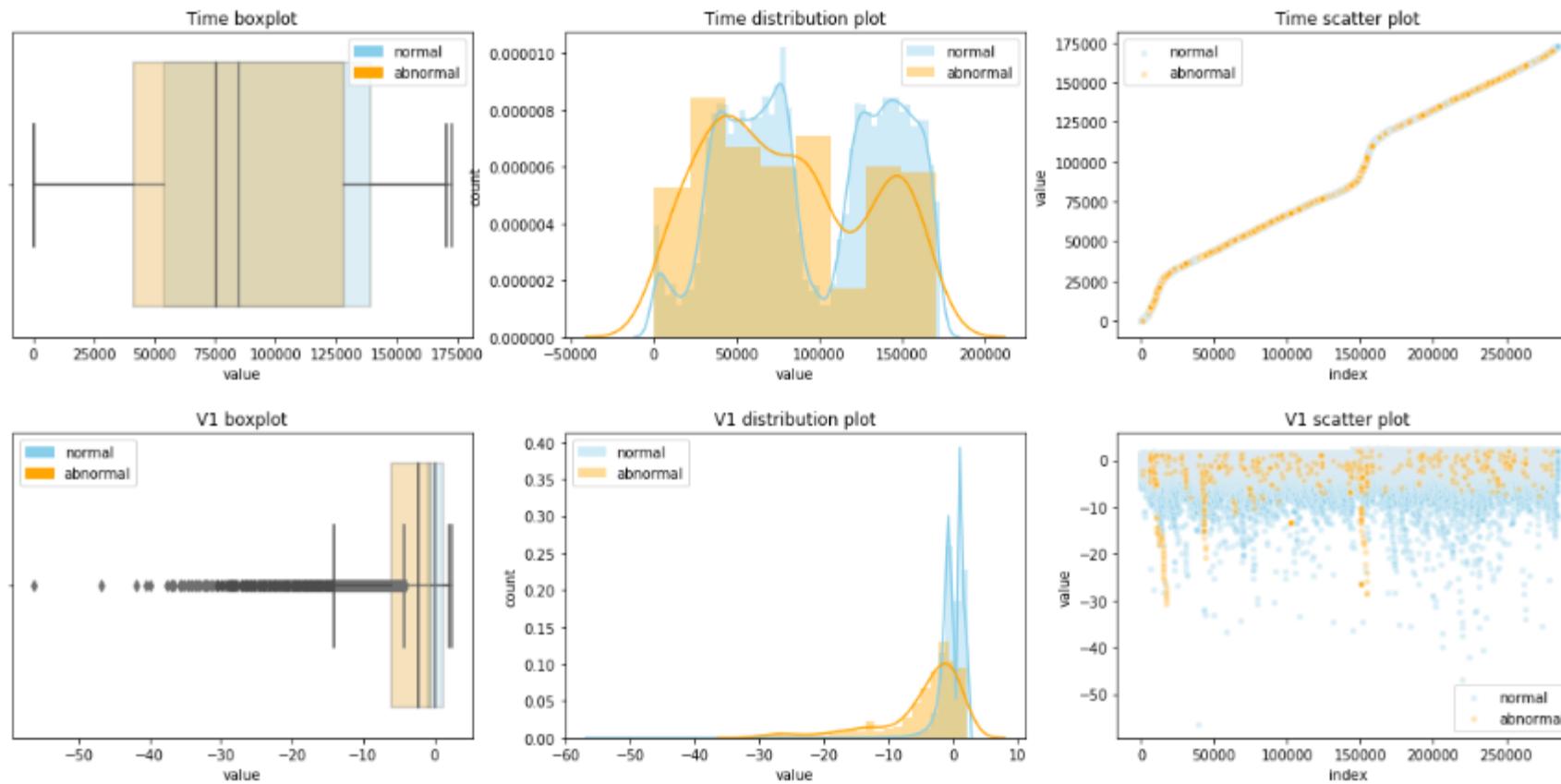


2) Description of Data

Input variable :

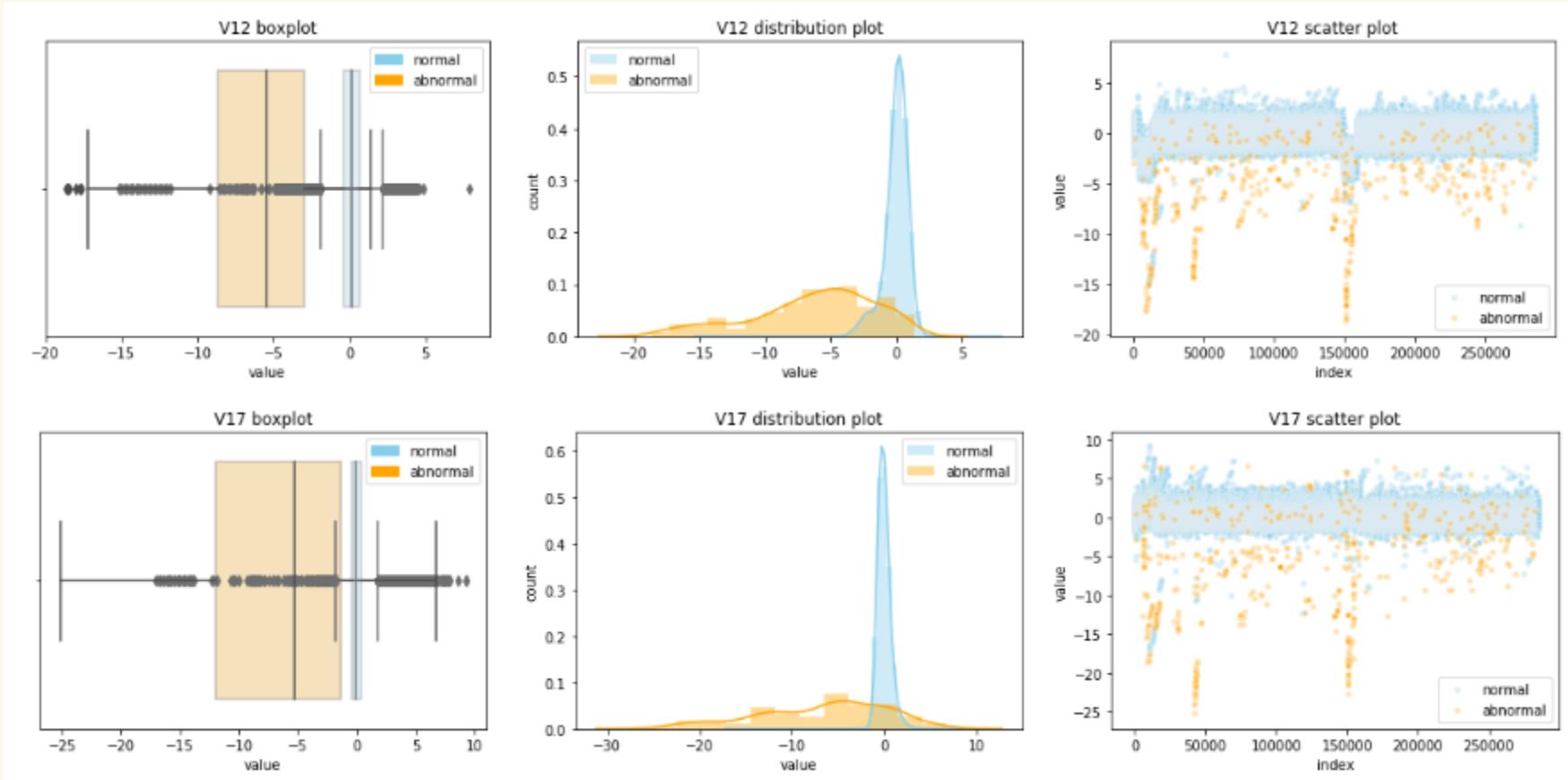
- 28 input variable V, result of PCA transformation.
- Time : seconds (about 48 hours)
- Amount : transaction amount.

3) Description of Data

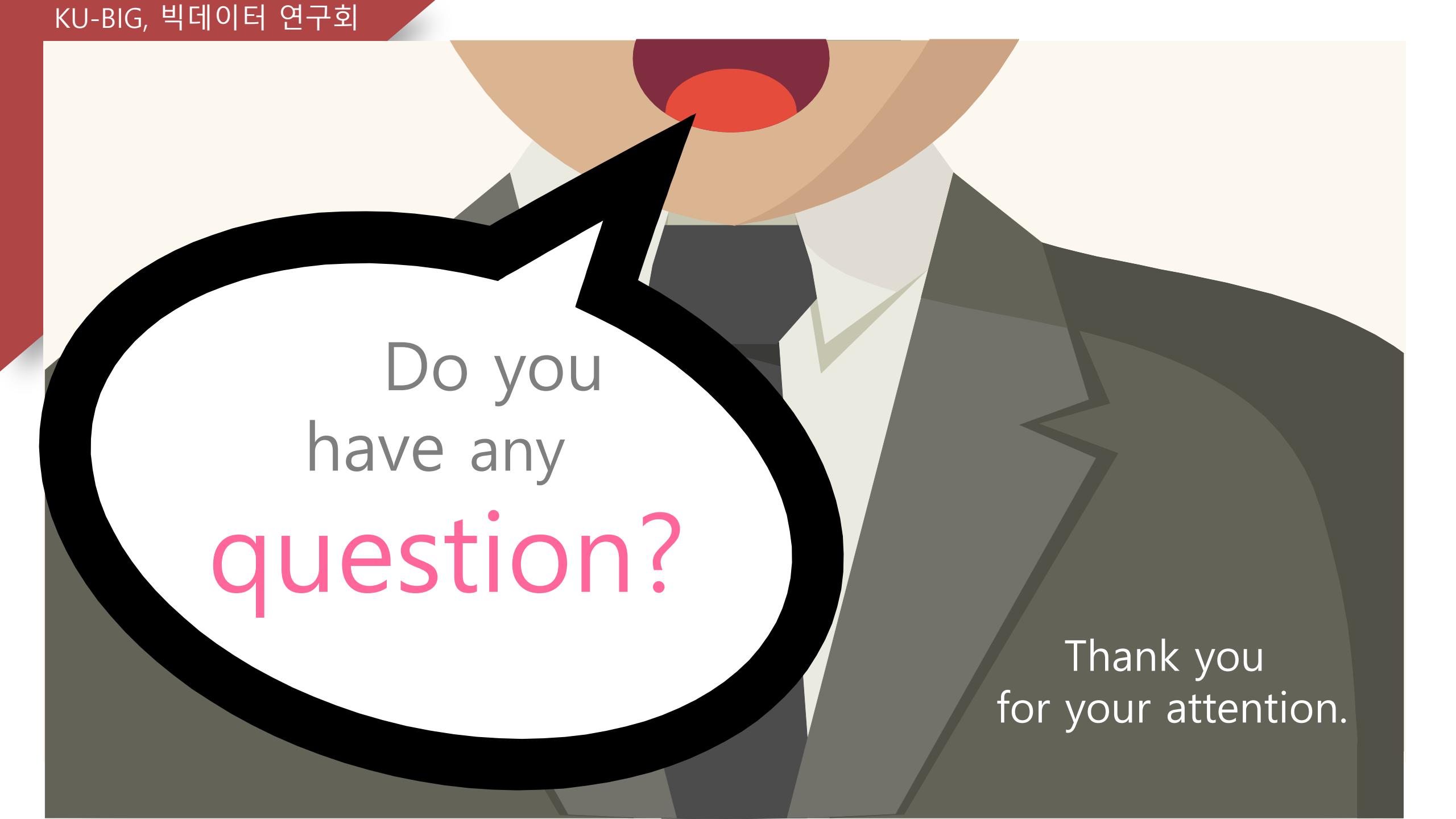


Plot of Time and V1

4) Description of Data



Plot of V12 and V17



Do you
have any
question?

Thank you
for your attention.