

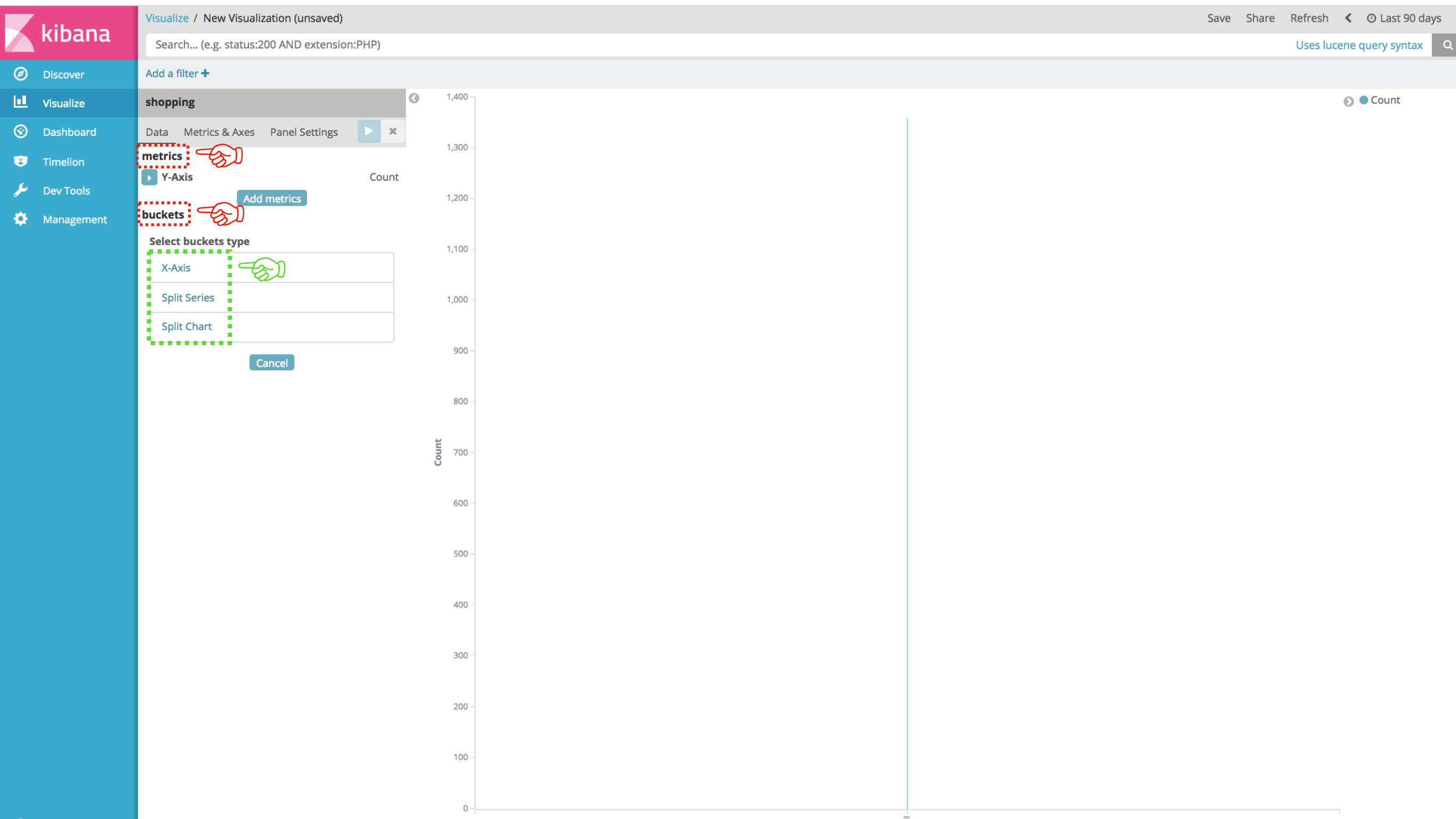
# Visualization 안내

**우선 Visualization 몇 가지를 보자**

## Data Table

[illegible]

# Area Map



metrics, buckets, x-axis, split series, split chart, split rows, split table ...

시각화하려는 문제는 명확한데,

어디에 들어가서 어떻게 조작해야되는지 모르겠다

# 1. 큰 틀은 비슷하다

**metrics** : sum, avg, min, max 등 수치 연산을 수행하는 부분

**buckets** : 위의 metrics를 적용할 그룹을 정의하는 부분

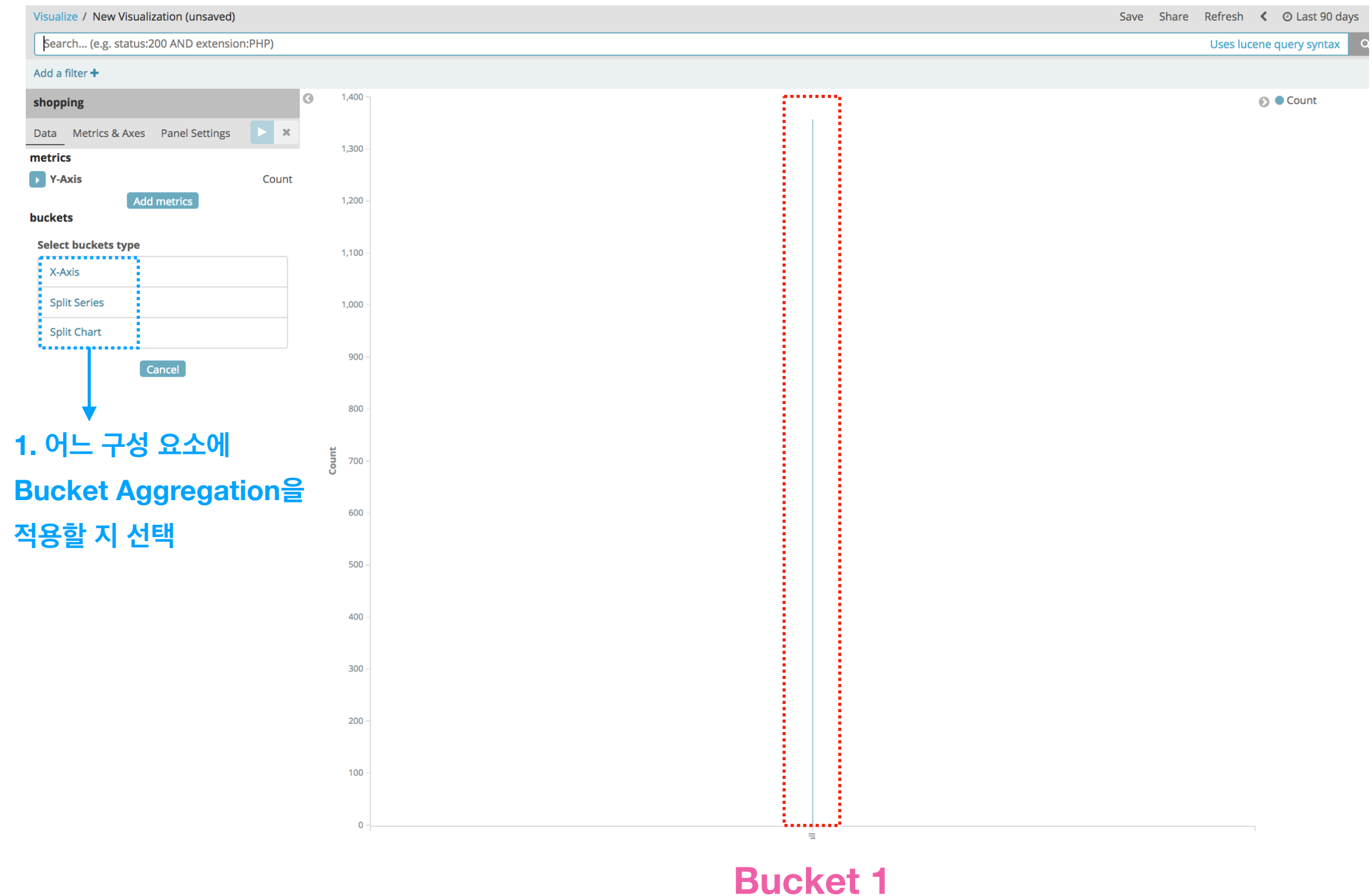
예: 전국 학생들의 지역별 평균 키를 구한다고 하자

키의 평균을 구하는 작업 : metrics

학생들을 지역별로 나누는 작업 : buckets

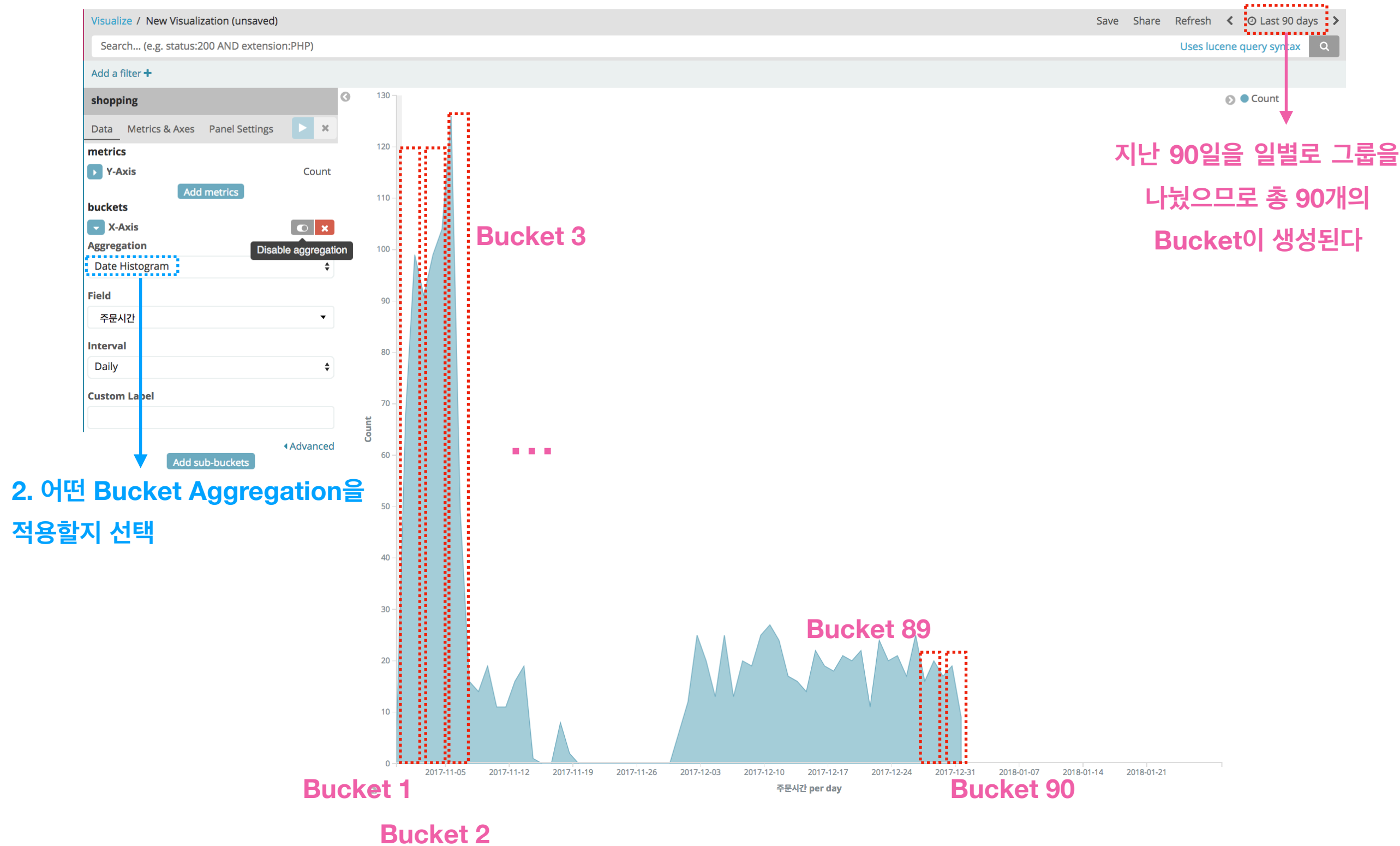
**2. 개별적 구성요소는 Visualization Type 마다 상이할 수 있다**

# 대표적인 buckets type 몇 개를 살펴보자 X-Axis Before

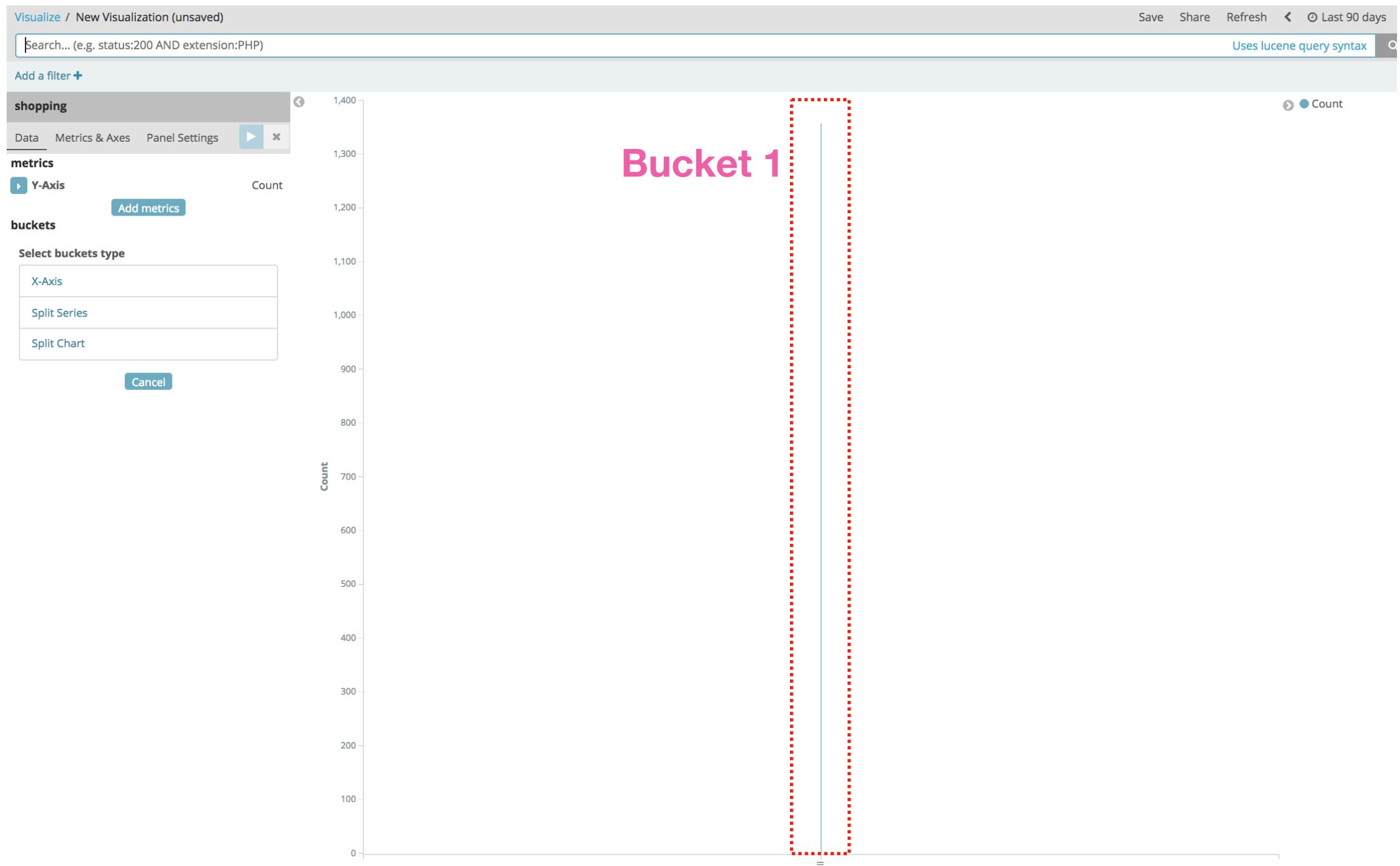




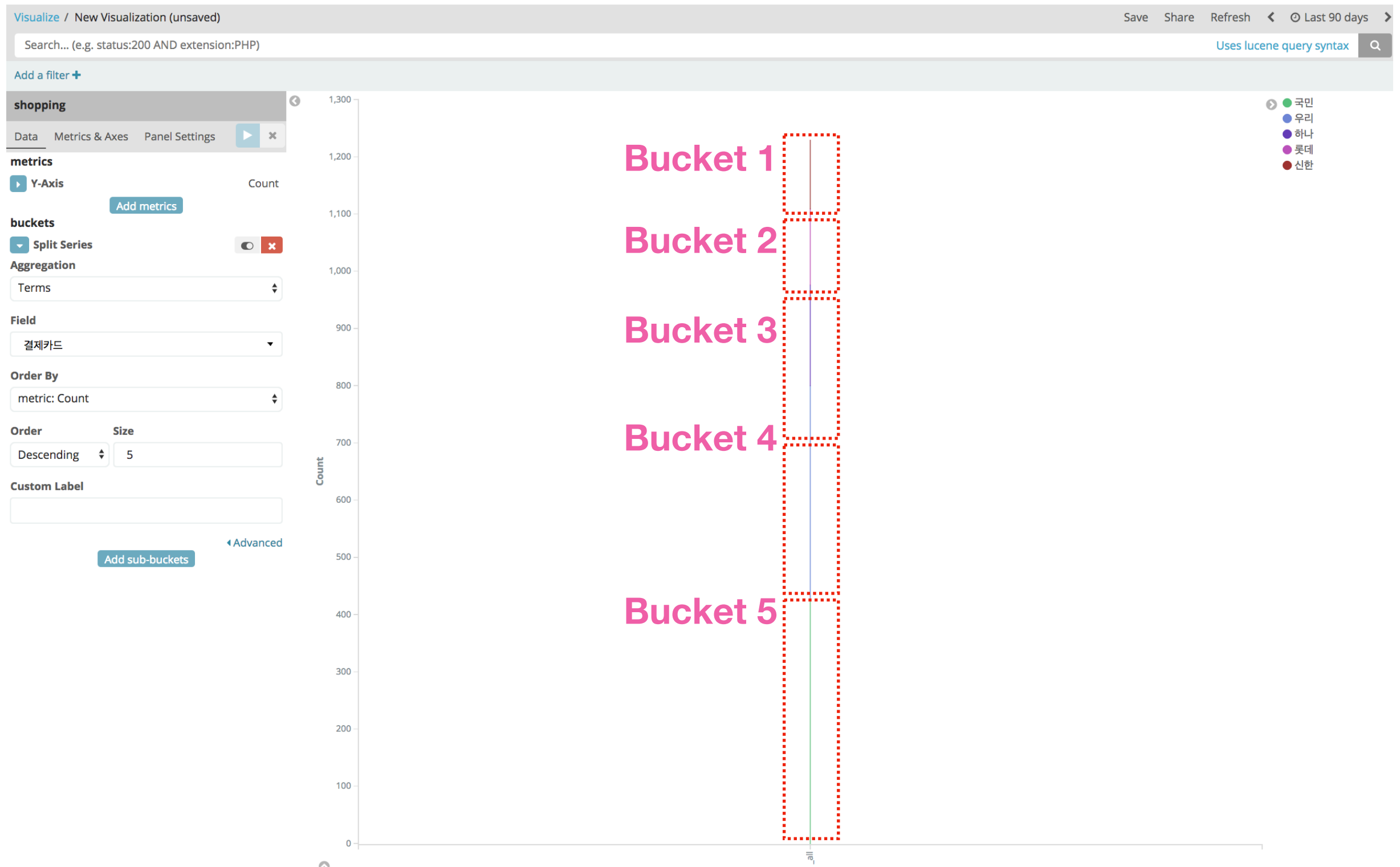
# 대표적인 buckets type 몇 개를 살펴보자 X-Axis After



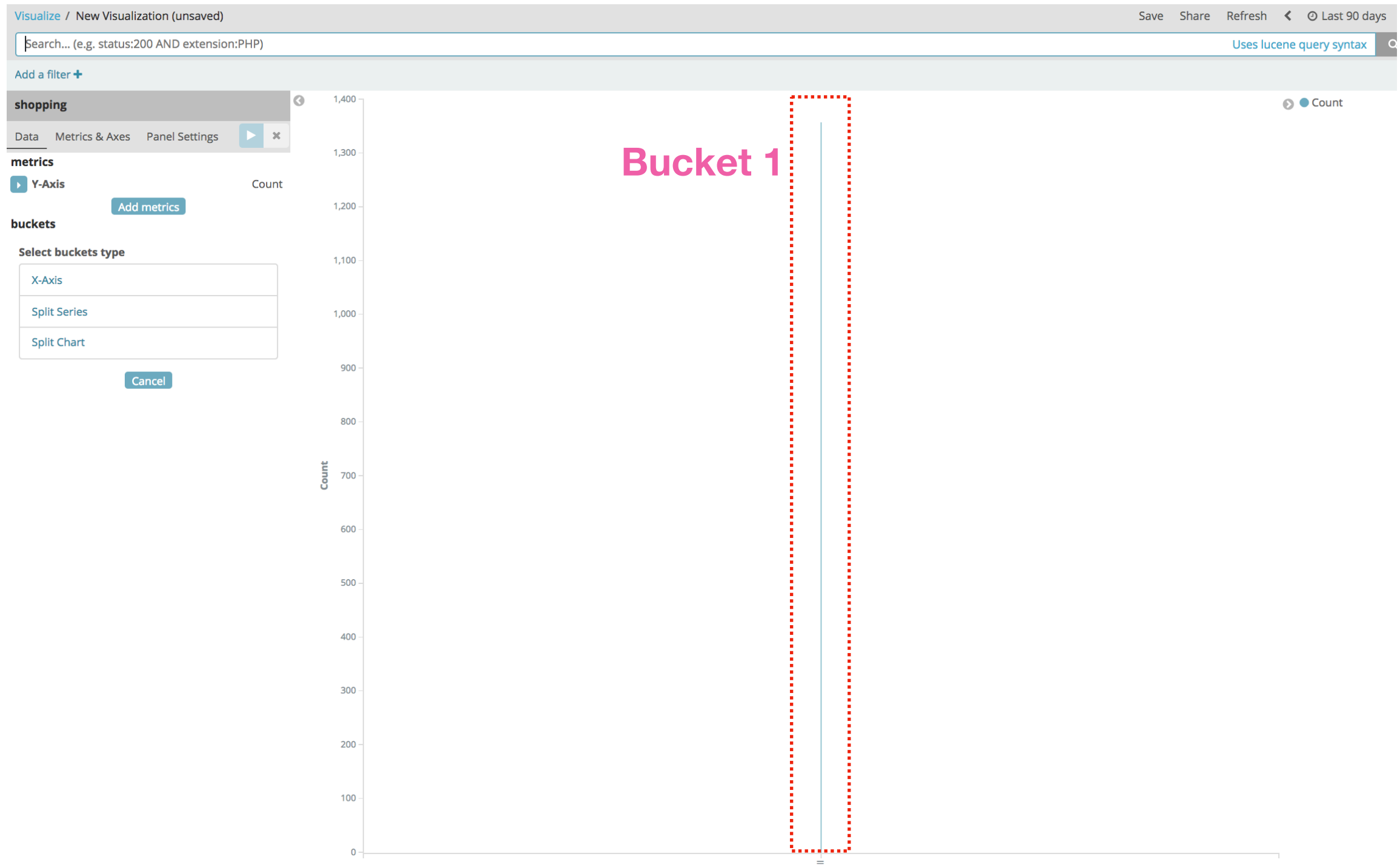
# 대표적인 buckets type 몇 개를 살펴보자 Split Series Before



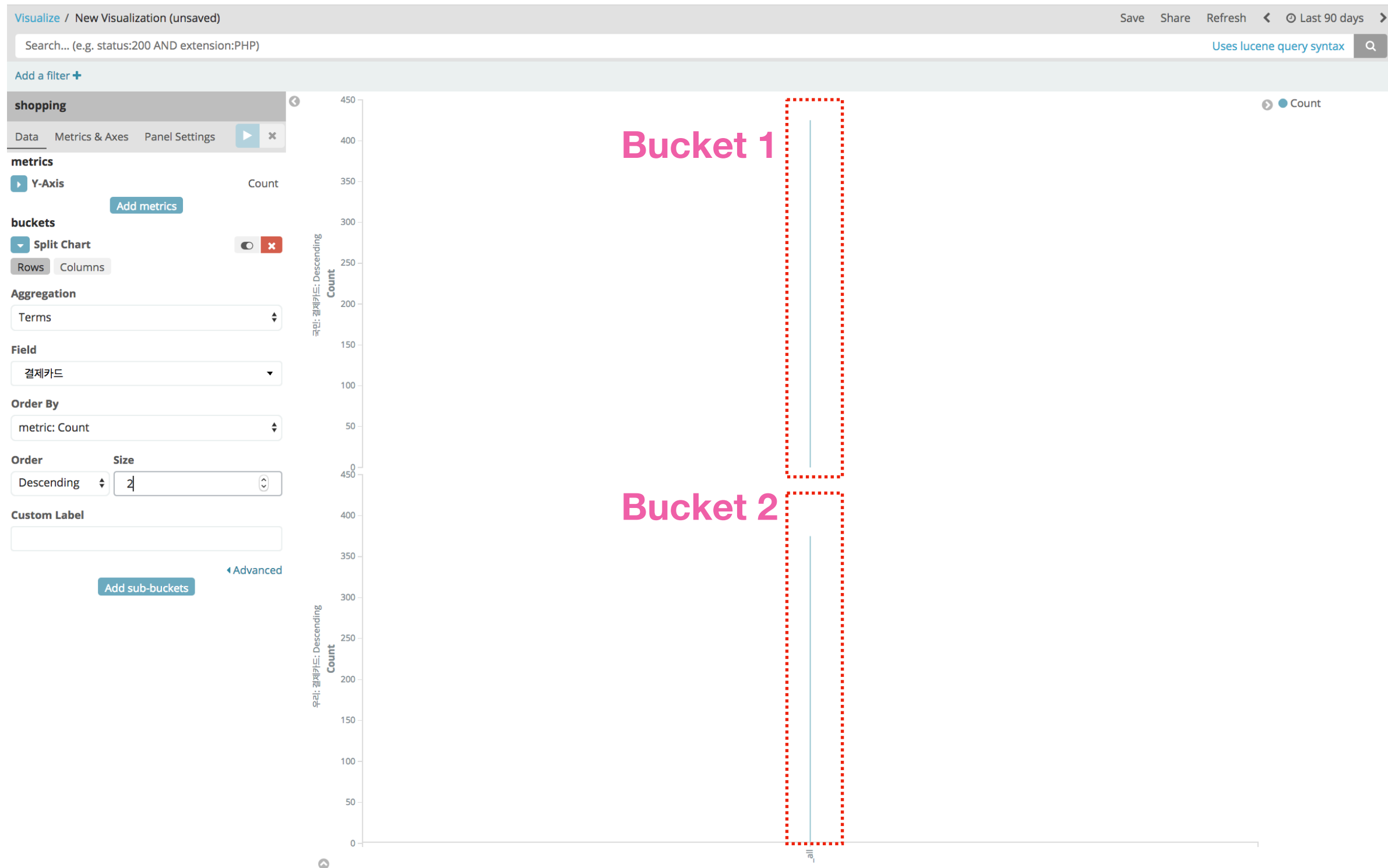
# 대표적인 buckets type 몇 개를 살펴보자 Split Series After



# 대표적인 buckets type 몇 개를 살펴보자 Split Chart Before



# 대표적인 buckets type 몇 개를 살펴보자 Split Chart After



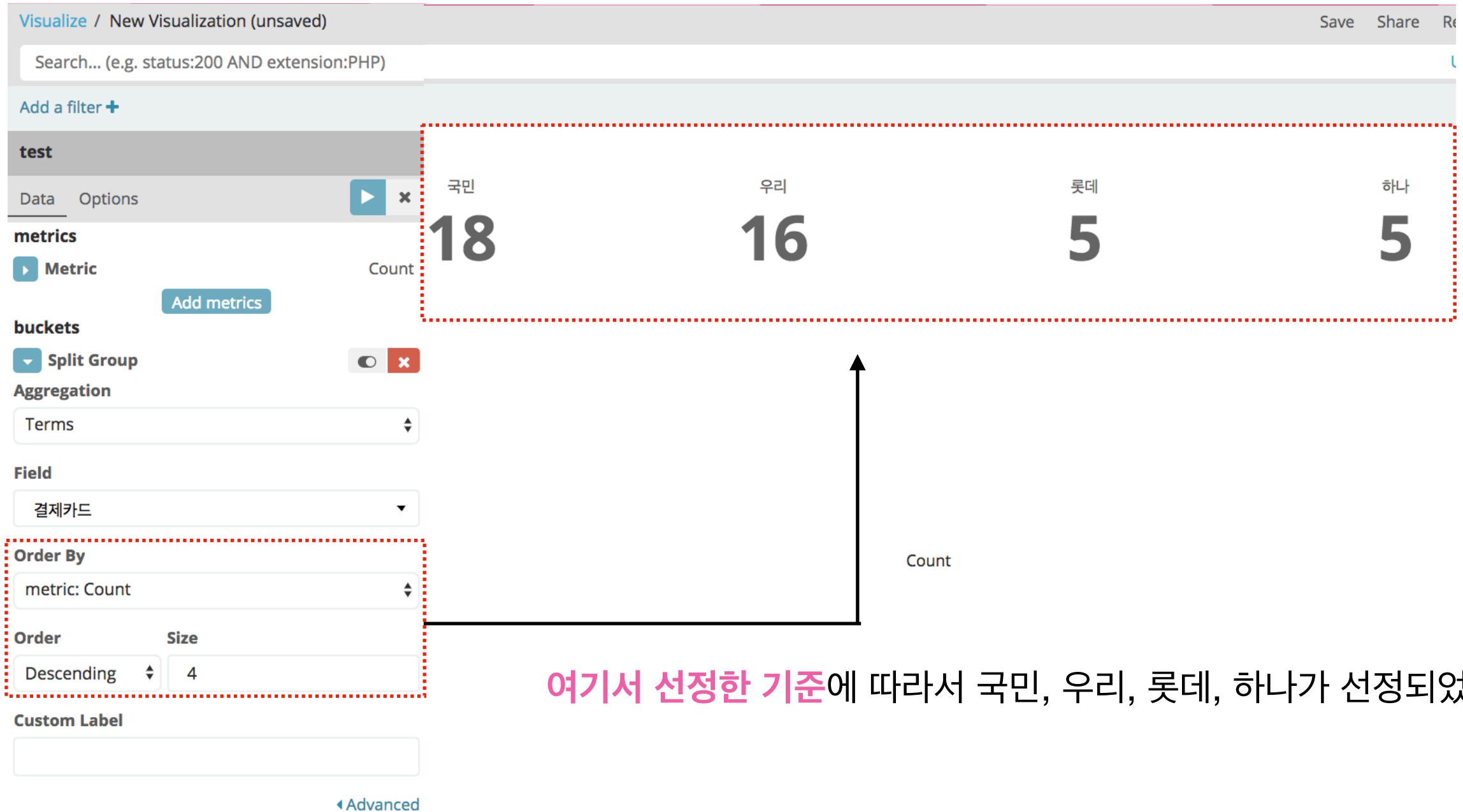
보통은 이 작업의 반복이지만 **Term Aggregation**으로  
Bucket을 나눌 경우 한 단계 더 고려해야한다

- 결제카드 별 사용자 수를 구한다고 하자.
- 모든 결제카드에 대해 구할 수 있지만 특정 4개 카드에 대해서만 본다고 하자.
- 이 때 특정한 카드 4개는 어떻게 선정할까?



이를 위해 Term Aggregation 내에서

**Bucket을 선정하기 위한 Aggregation**을 수행한다



여기서 **선택한 기준**에 따라서 국민, 우리, 롯데, 하나가 선정되었다.



test

Data Options

metrics

▶ Metric Count

Add metrics

buckets

▼ Split Group

Aggregation

Terms

Field

결제카드

Order By

metric: Count

Order

Descending

Size

4

Custom Label

1. 결제카드로 Bucket을 구분해서...

2. Bucket 별 Count를 구하고...

4. 상위 4개를 선정해라

3. Count가 큰 순으로 정렬해서...

국민 | 하나 | 신한 | 롯데 | 시티 | 우리 ...

국민 | 하나 | 신한 | 롯데 | 시티 | 우리 ...  
18 | 5 | 3 | 5 | 2 | 16 ...

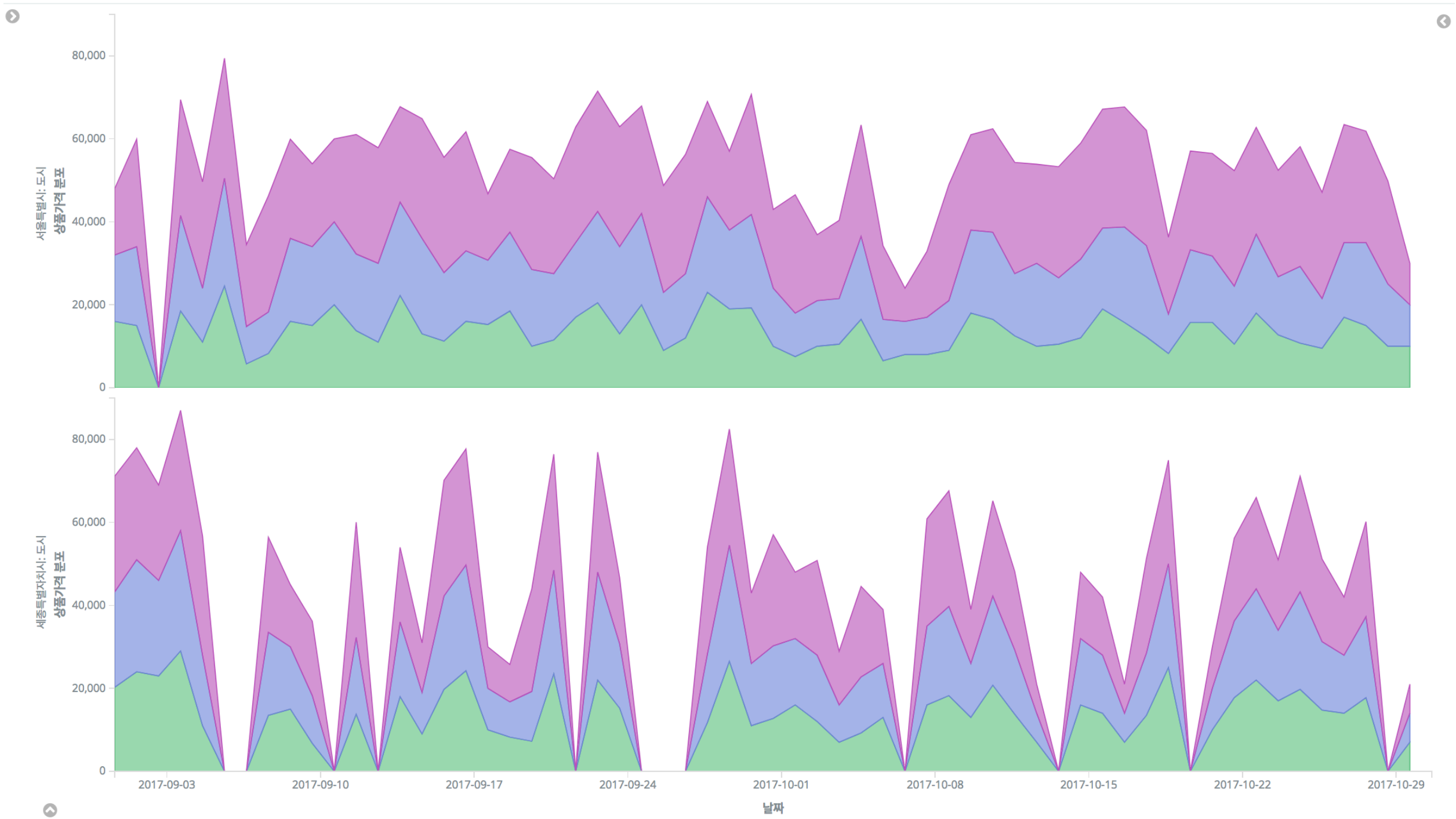
국민 | 우리 | 롯데 | 하나  
18 | 16 | 5 | 5

국민 | 우리 | 롯데 | 하나 | 신한 | 시티 ...  
18 | 16 | 5 | 5 | 3 | 2 ...

**그렇다면 Visualization 문제가 주어지면 어떤 flow로 생각해야 할까?**

1. 문제에서 **metrics** 영역과 **buckets** 영역으로 구분한다
2. **metrics**와 **buckets** 내에서 사용할 aggregation을 선택한다
3. term aggregation으로 **bucket**을 나눌 경우 sorting을 위한 aggregation을 정의한다

**예시를 통해 어떻게 적용하는지 보자**



## 데이터

- Index : shopping
- Time Range : 2017년 9월 1일 ~ 2017년

## 문제

- “상품가격”의 합이 가장 큰
- “고객주소\_시도” 2개의
- “상품가격”의 25분위, 50분위, 95분위를
- “주문시간”을 기준으로 daily로 표시하고
- week2\_{id}\_area로 저장하세요

## 사용한 Aggregation

- Metrics (Y-Axis) : Percentiles Aggregation
- Buckets
  - Split Chart (Rows) : Terms Aggregation
  - X-Axis : Date Histogram Aggregation

## 사용 필드

- 상품가격
- 고객주소\_시도
- 주문시간

## 문제에서 **metrics** 영역과 **buckets** 영역으로 구분한다

### 문제

- “상품가격”의 합이 가장 큰

- “고객주소\_시도” 2개의

- “상품가격”의 25분위, 50분위, 95분위를

- “주문시간”을 기준으로 daily로 표시해라

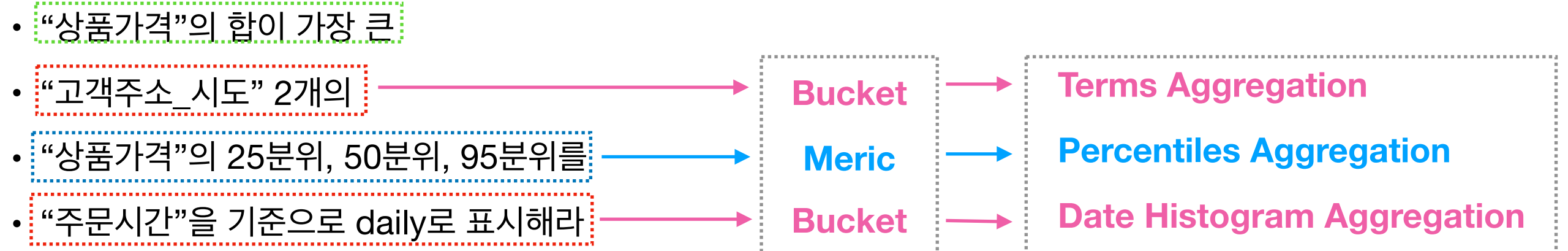
Bucket

Metric

Bucket

## metrics와 buckets 내에서 사용할 aggregation을 선택한다

### 문제





term aggregation으로 **bucket**을 나눌 경우 sorting을 위한 aggregation을 정의한다

