

# Social Media Search

# CONTENTS

---



- 01 UCI Big Data Group
- 02 Project
- 03 Silicon Valley Tour

# 01. UCI Big Data Group

- \* Big Data Group
  - Big Data Group in University of California, Irvine
- \* AsterixDB
  - Apache AsterixDB is a scalable, open source Big Data Management System
  - <https://asterixdb.apache.org/index.html>



# 02. PROJECT

## 1) Introduction

### \* Start From Jianfeng's Demo

**Asterix<sup>DB</sup>**  
Explore

### Query Builder

**Keyword**  
sister

**Location**  
Enter a location  
Location Selection

**Limit Date Range**  
2015-11-11 Start Date  
2016-01-07 End Date

**Result Group Resolution**  
By State By County By City

Submit Clear

☐ Submit asynchronously?

```
use dataverse twitter; create temporary dataset  
tmp_tweets(type_tweet) primary key id; insert into  
dataset tmp_tweets ( let $region := create-  
rectangle(create-  
point(-126.58535156250002,17.848305449542117),
```

Google

Mexico: Map data ©2016 Google, INEGI | Terms of Use | Report a map error

1

# Tweets

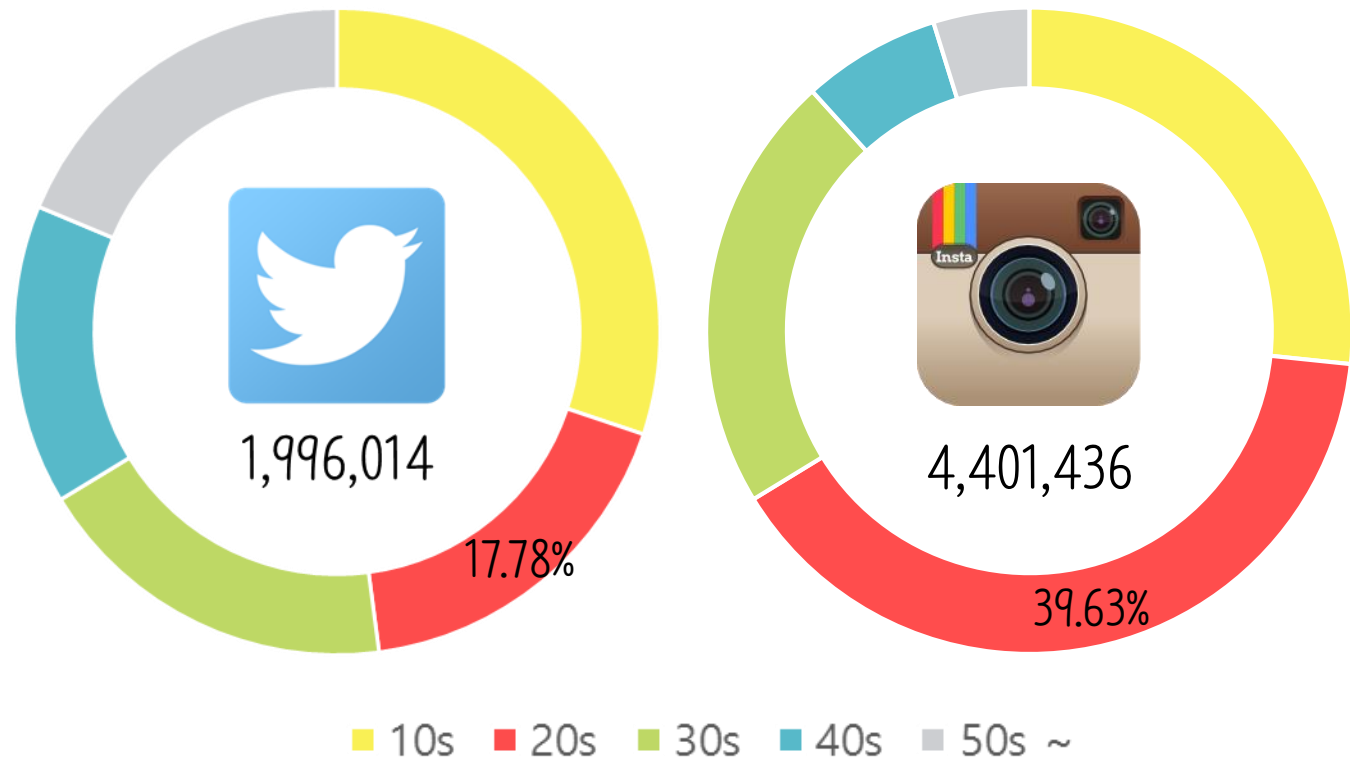
234

ly sends some data to Mozilla so that we can improve your experience.

# 02. PROJECT

## 1) Introduction

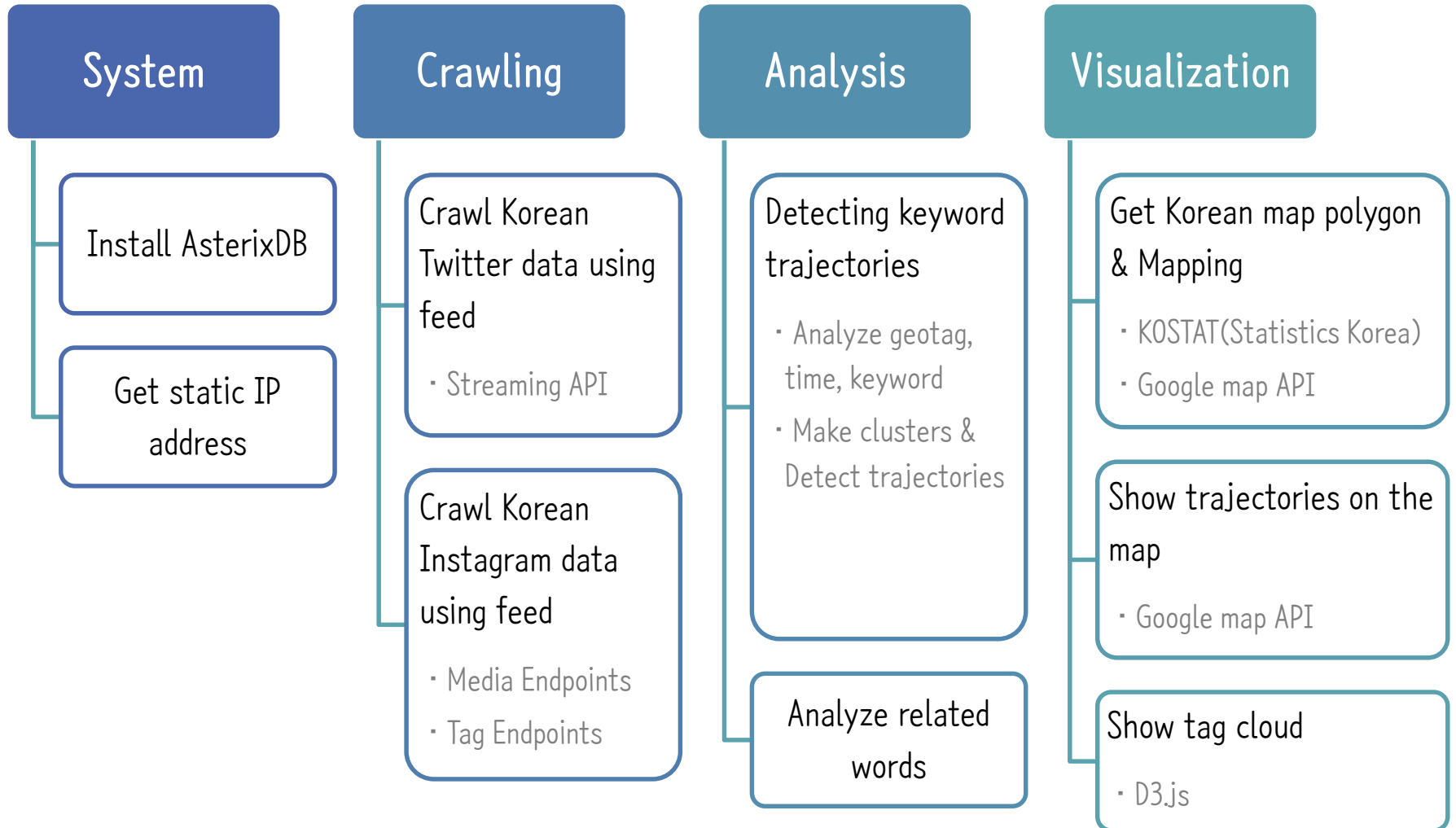
- \* Focusing on Korea
  - Demographic composition % of Korean social networks



Source: Appranker, Korea, March, 2015

# 02. PROJECT

## 2) Social Media Search



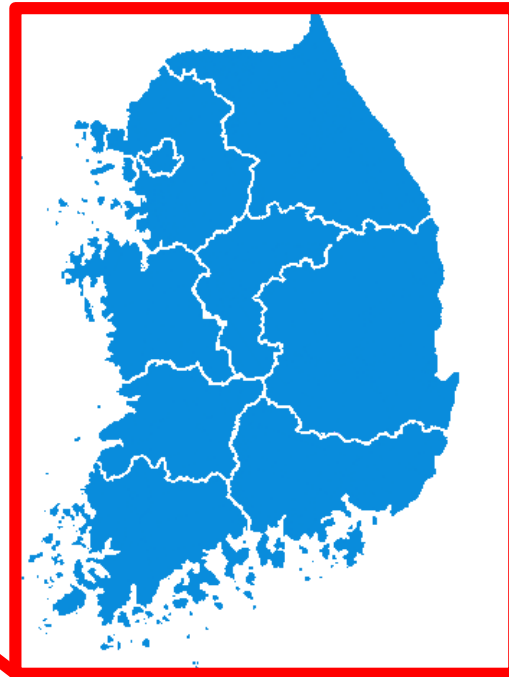
# 02. PROJECT

## 2) Social Media Search

\* Twitter Streaming API



Longitude: 125.749512  
Latitude: 33.815666



Longitude: 125.749512  
Latitude: 33.815666

Longitude: 125.991211  
Latitude: 33.063924



Longitude: 127.177734  
Latitude: 33.651208

# 02. PROJECT

## 2) Social Media Search

### \* Instagram Media & Tag Endpoints



GET /tags/search

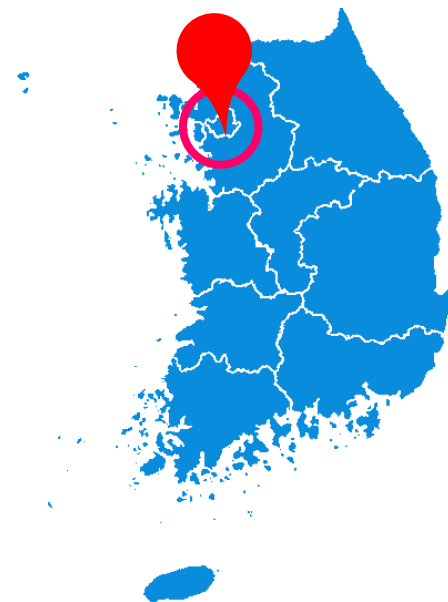
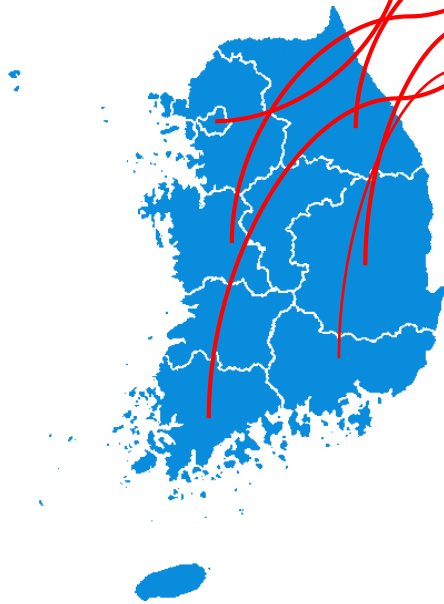
[https://api.instagram.com/v1/tags/search?q=snowy&access\\_token=ACCESS-TOKEN](https://api.instagram.com/v1/tags/search?q=snowy&access_token=ACCESS-TOKEN)

RESPONSE ▾

GET /locations/search

[https://api.instagram.com/v1/locations/search?lat=48.858844&lng=2.294351&access\\_token=ACCESS-TOKEN](https://api.instagram.com/v1/locations/search?lat=48.858844&lng=2.294351&access_token=ACCESS-TOKEN)

RESPONSE ▾



# 02. PROJECT

## 2) Social Media Search

### \* ADM & Data

```
create type type_instagram_user as open{
  id: string,
  full_name: string,
  profile_picture : string,
  username : string
}

create type type_instagram_place as open{
  geo_location : point,
  full_name : string,
  id : string
}

create type type_instagram as open{
  id : int64,
  create_at : datetime,
  link: string,
  text_msg : string,
  hashtags : {{ string }} ?,
  user : type_instagram_user,
  place : type_instagram_place
}

create dataset ds_instagram(type_instagram) primary
create index text_idx on ds_instagram(text_msg) type
create index time_idx on ds_instagram(create_at) type
```

```
{
  "id": 1183127639881124098,
  "create_at": datetime("2016-02-12T15:50:27.000Z"),
  "link": "https://www.instagram.com/p/BBrUFHrD4kC/",
  "text_msg": "알바 끝나고 집 가는길에 예뻐서\n찍은 우리학교 지하철  
입구\nd\n#가천대 #지하철에서 #일상스냅 #데일리인스타 #풍경사진 #daily  
#instadaily #dailyphoto #맛팔로우 #알바개꿀",
  "hashtags": {
    { "알바개꿀",
      "지하철에서",
      "풍경사진",
      "데일리인스타",
      "맛팔로우",
      "가천대" }
  },
  "user": {
    "id": "1778975993",
    "full_name": "김영수",
    "profile_picture":
https://igcdn-photos-e-a.akamaihd.net/hphotos-ak-xtf1/t51.2885-19/s150x150/12446005\_1566249323700588\_1286404259\_a.jpg,
    "username": "0_songg"
  },
  "place": {
    "geo_location": point("127.126944444,37.447777778"),
    "full_name": "Seongnam-si, Gyeonggi-do",
    "id": "237870886"
  }
}
```

# 02. PROJECT

---

## 2) Social Media Search

### \* Twitter Data

- Number of tweets : About 500,000
- Size of adm : 700MB
- Period of collecting : 2016/01/25 ~

### \* Instagram Data

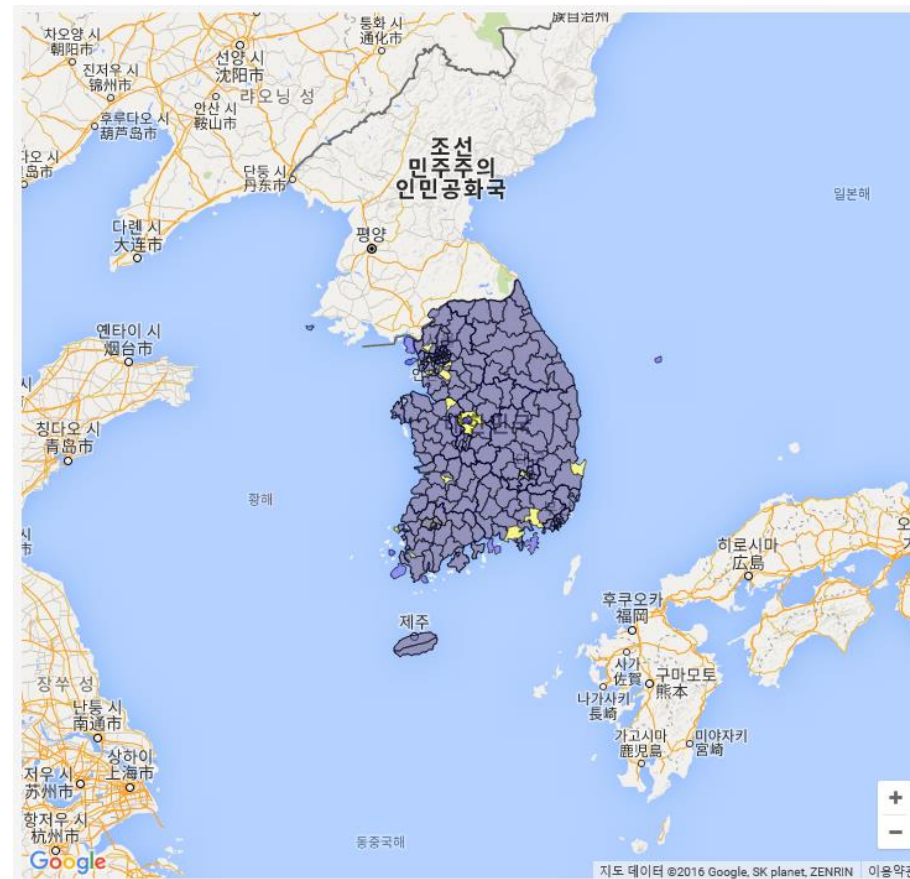
- Number of instagram : About 200,000
- Size of adm : 169MB
- Period of collecting : 2016/02/07 ~ 2016/02/14

# 02. PROJECT

## 2) Social Media Search

### \* Korean Map

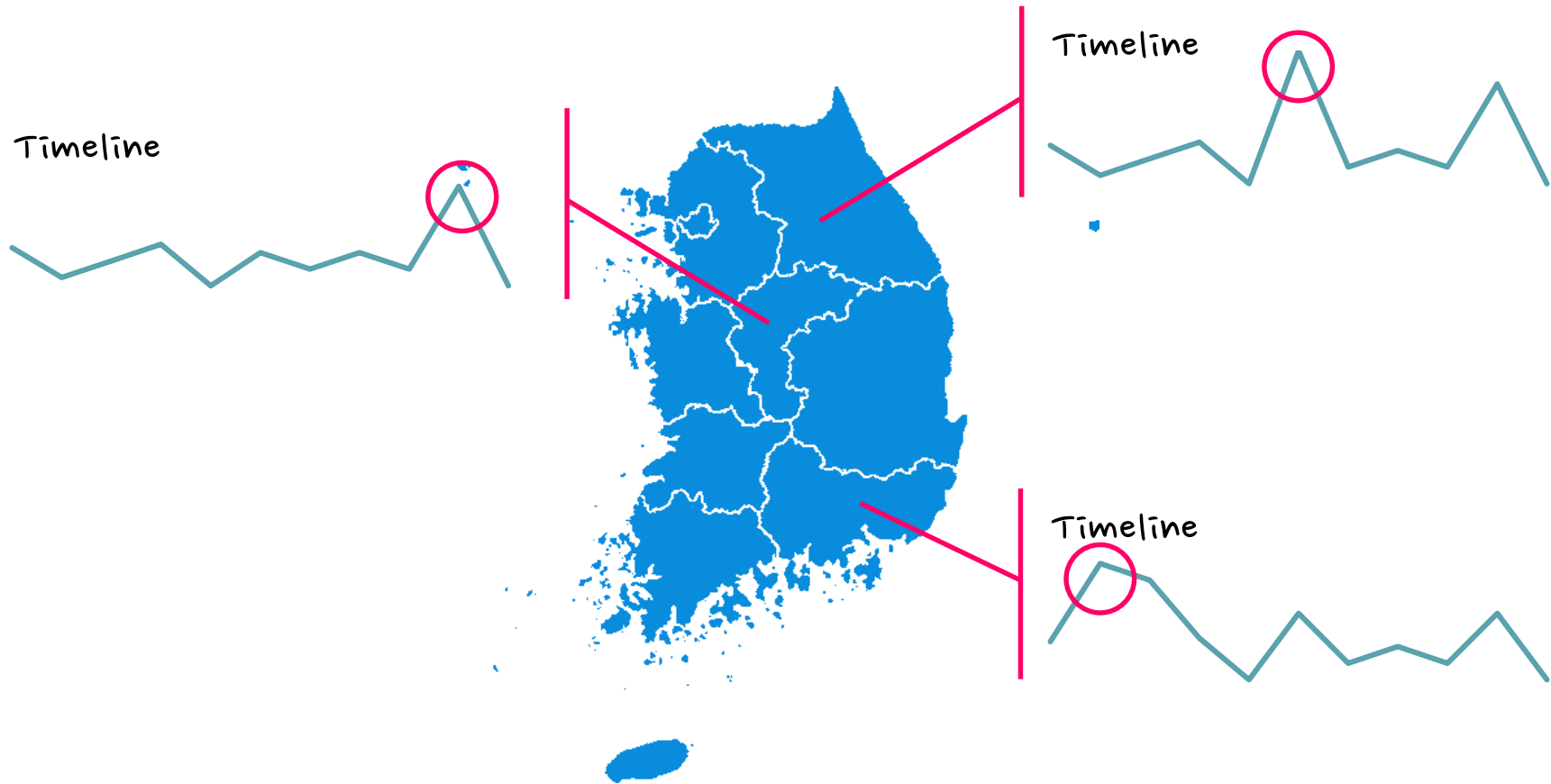
- Use data from KOSTAT
- Province / County



# 02. PROJECT

## 2) Social Media Search

- \* Trajectory
  - Analyze trajectory about keyword

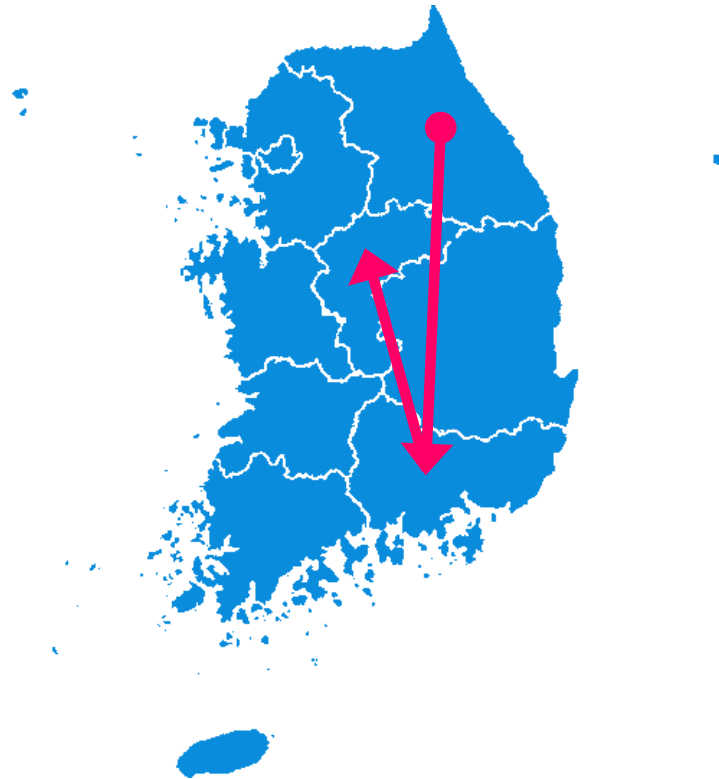


# 02. PROJECT

---

## 2) Social Media Search

\* Trajectory



# 02. PROJECT

## 2) Social Media Search

### \* Trajectory

- Aql for analyzing trajectory

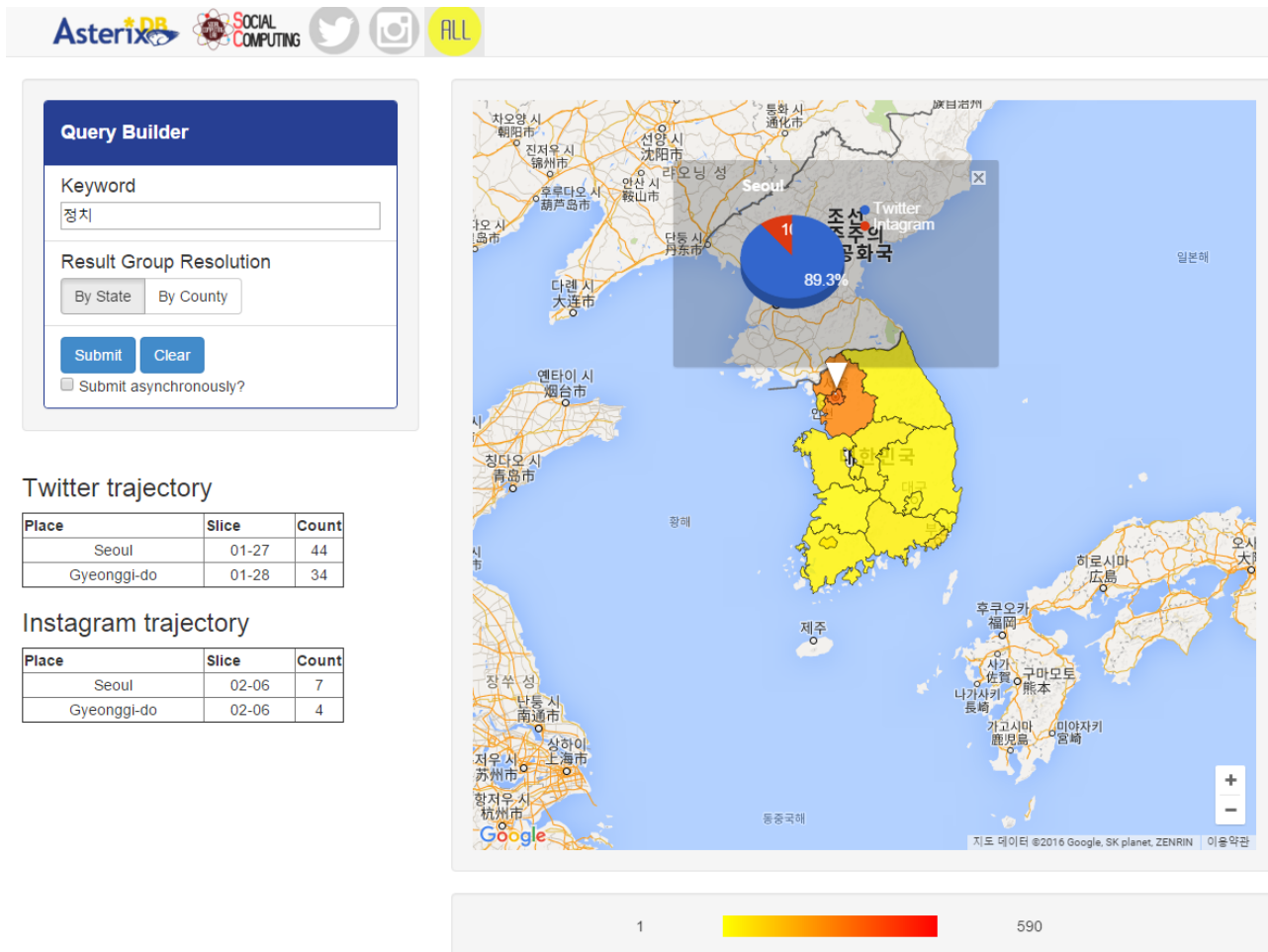
```
aql.push('for $t in dataset tmp_instagram');

if (polygon) {
  aql.push('where ' + selectAreaByPolygon(polygon));
  aql.push('group by $g := substring-before($t.place.full_name, ",") with $t ');
  aql.push('return {');
  aql.push('    "place":$g, ');
  aql.push('    "data": for $tmp in dataset tmp_instagram');
  aql.push('        where substring-before($tmp.place.full_name, ",") = $g ');
  aql.push('        group by $c := print-datetime($tmp.create_at, "MM-DD") with $tmp ');
  aql.push('        let $count := count($tmp) ');
  aql.push('        order by $count desc');
}
else {
  aql.push('group by $g := substring-after($t.place.full_name, ", ") with $t ');
  aql.push('return {');
  aql.push('    "place":$g, ');
  aql.push('    "data": for $tmp in dataset tmp_instagram');
  aql.push('        where substring-after($tmp.place.full_name, ", ") = $g ');
  aql.push('        group by $c := print-datetime($tmp.create_at, "MM-DD") with $tmp ');
  aql.push('        let $count := count($tmp) ');
  aql.push('        order by $count desc');
}
aql.push('    return {"slice":$c, "count":$count, "content":$tmp}');
aql.push('};\n');
```

# 02. PROJECT

## 2) Social Media Search

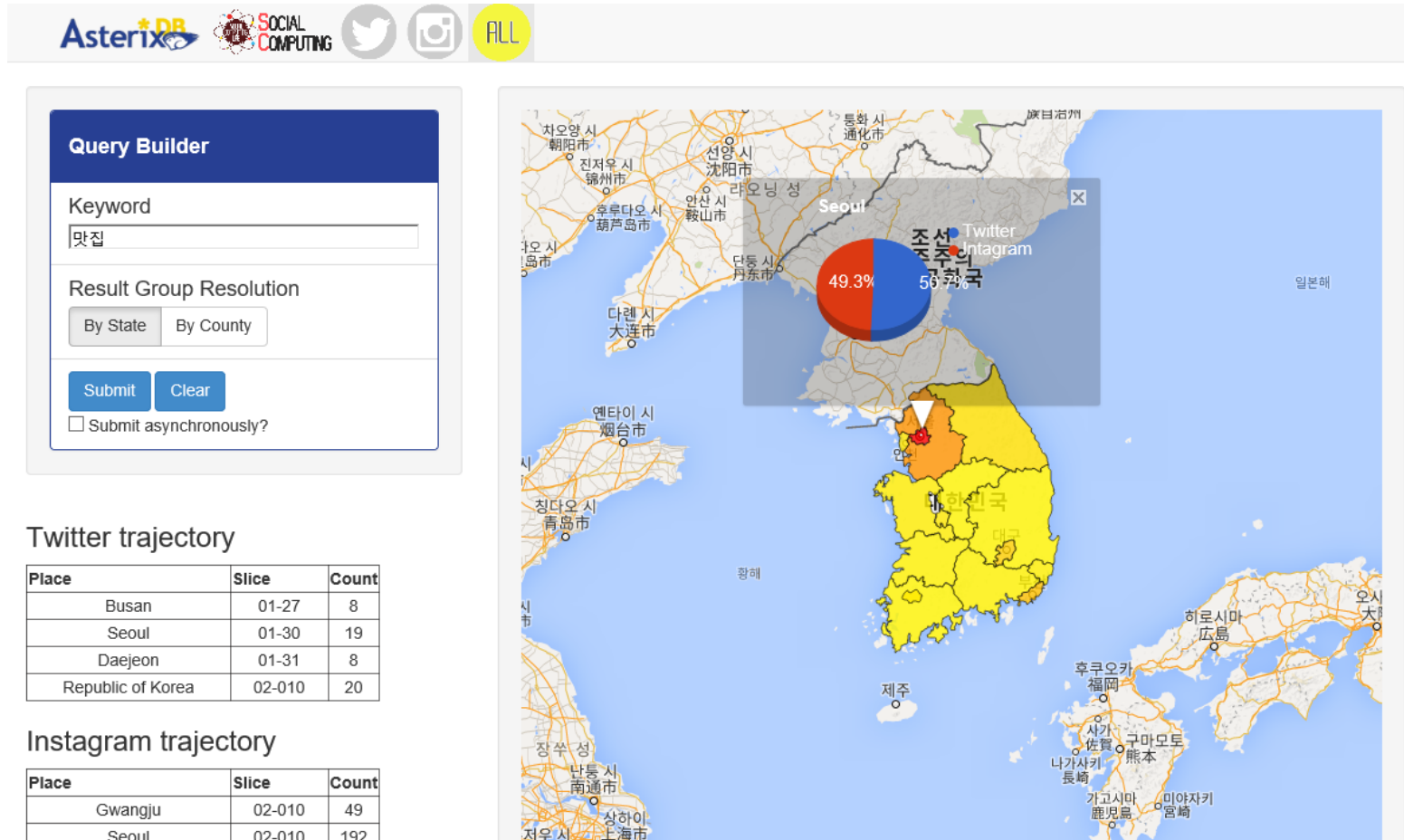
### \* Twitter vs. Instagram



# 02. PROJECT

## 2) Social Media Search

### \* Twitter vs. Instagram



# 02. PROJECT

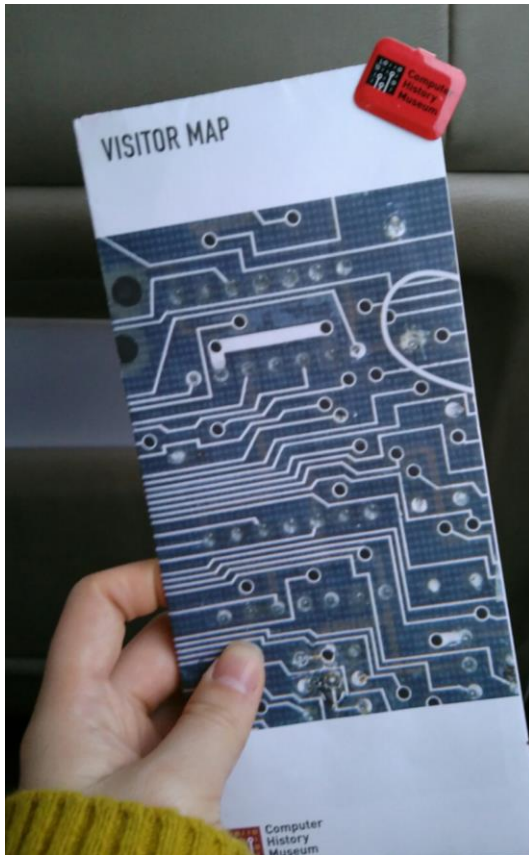
---

## 3) Future Work

- \* **Feed**
  - Collecting data using feed
- \* **Trajectory**
  - Improve the function to analyze trajectory
- \* **Social Context Aware**
  - Expand the function to apply social context awareness
  - Using location / weather / time
- \* **Opinion Mining**
  - Analyze positive / negative

# 03. Silicon Valley Tour

Computer History Museum



Intel



Apple



# 03. Silicon Valley Tour

\* Google



# THANK YOU