나의 사진은 내가 지난 과거에 한 일을 알고 있다

- 사진 데이터 분석해 보기:최규민

최규민 소개

웃음지수 [98.9171 97.4664 97.0466 96.7611 96.6105]

웃음지수 [0.575181 0.626017 0.751368 0.782068 0.810508]

데이터 덕질을 좋아하는 개발자 입니다.

시작하기에 앞서

Quantify Self(자아 정량화)

에 대하여 알아보자

Quantified Self

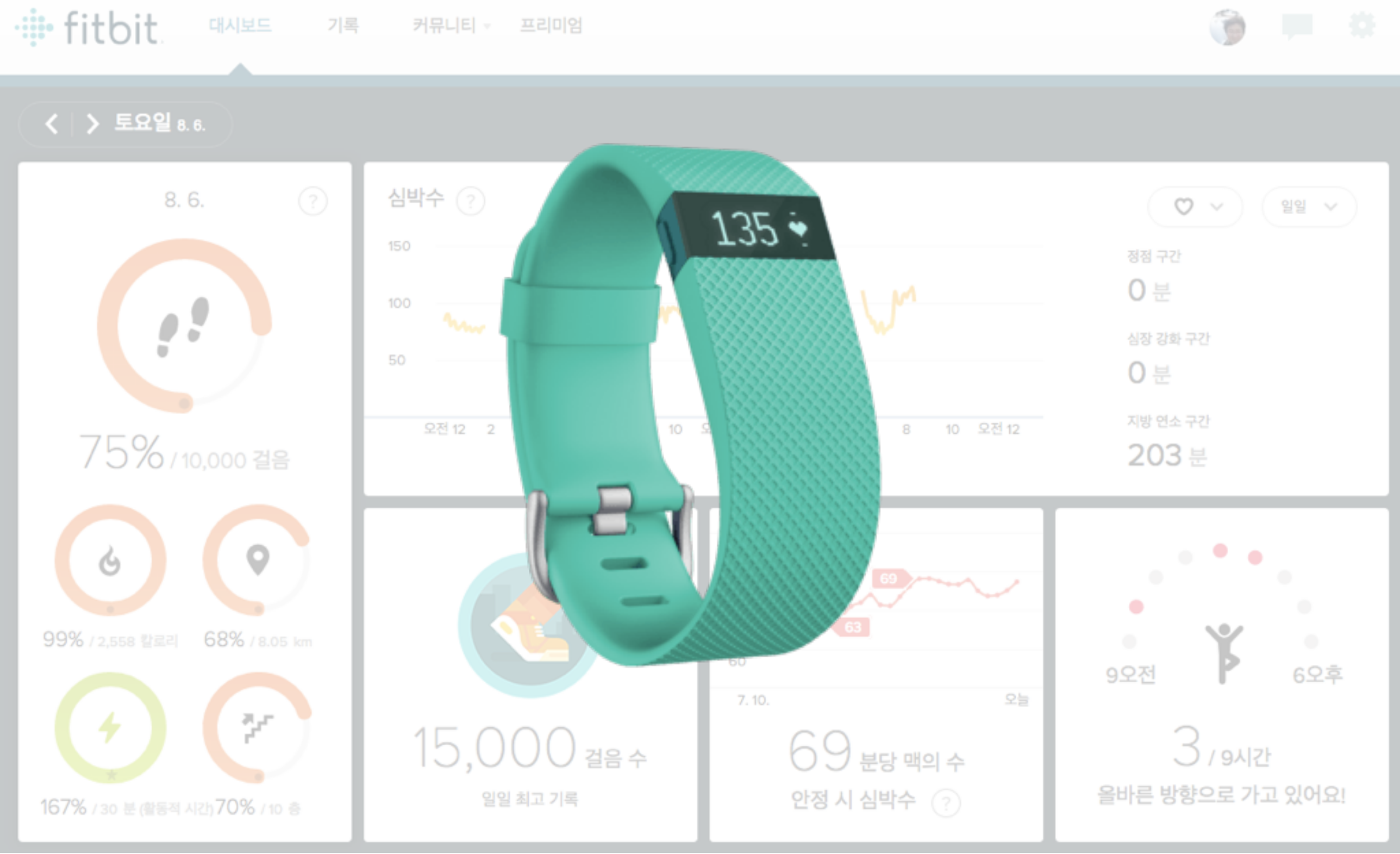
개인의 모든 일상을 측정하고 숫자로 표현하는 것..

오늘의 주제

사진 데이터 분석도 자아 정량화의 한가지 수단

그래서 잠시
저의 데이터 덕질 ㅈㅏ..랑을

피트니스 APP으로 LifeTracking



평균 11시 29분 잠이 들고

6시간 3분 잠을 자며
수면중 14번 깨어나며
수명 효율은 93.4%

하루에
9,880 걸음
7Km 이동
소소한 23분간 운동
활발한 20분간 운동
=
2,816Kcal
를 소모합니다

6시 25분에 일어나고

평균 11시 29분 잠이 들고

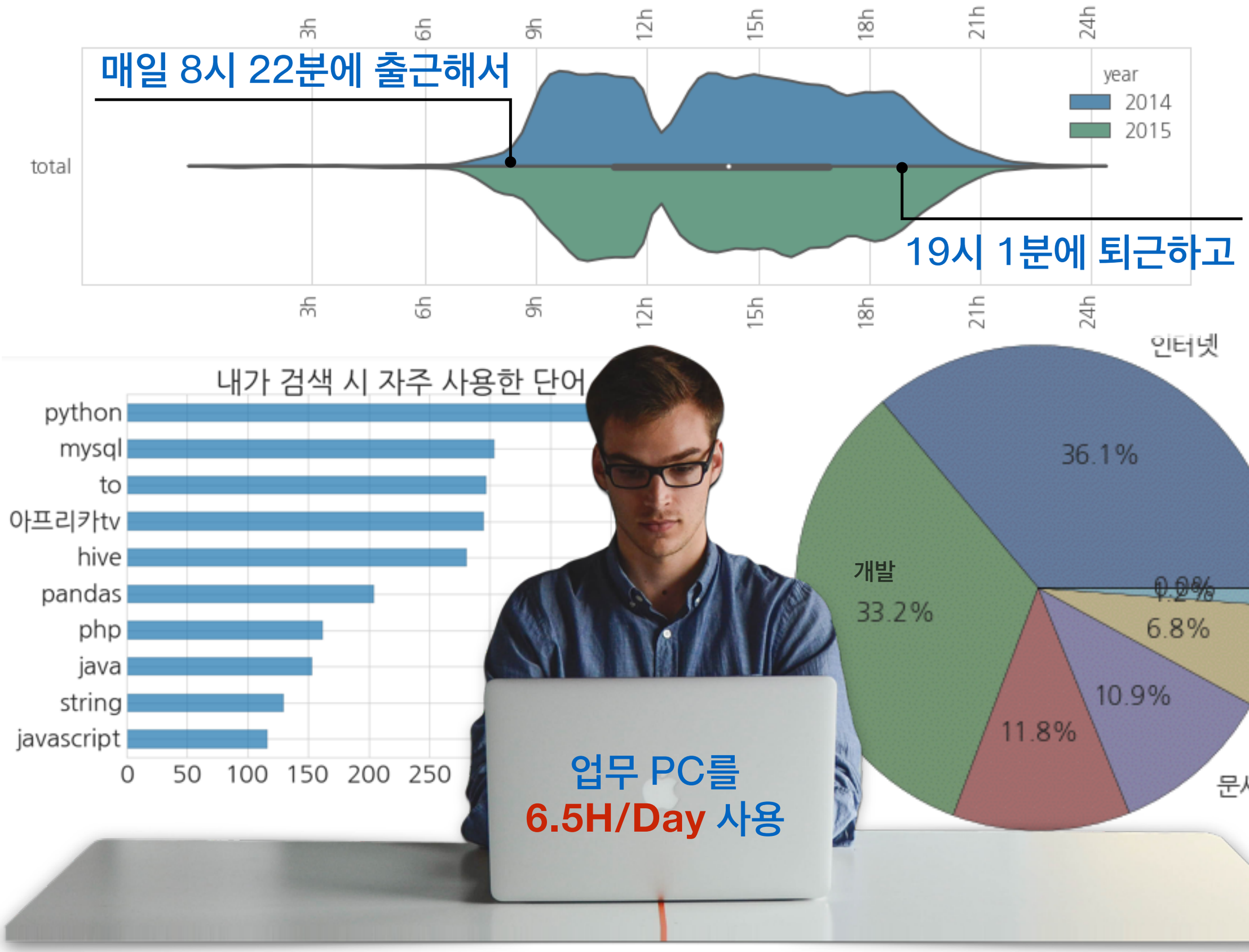
6시간 3분 잠을 자며
수면중 14번 깨어나며
수명 효율은 93.4%

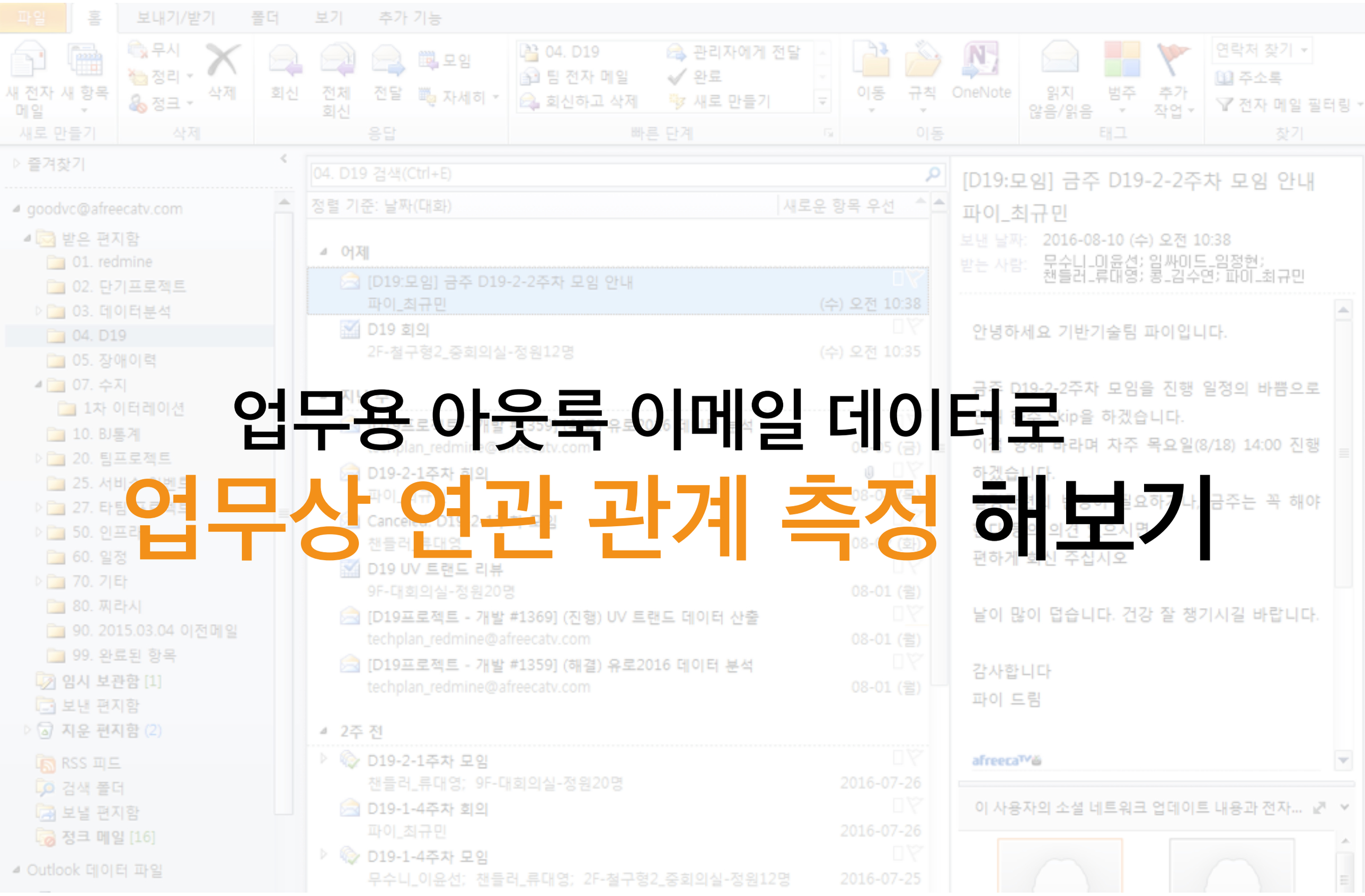
하루에
9,880 걸음
7Km 이동
소소한 23분간 운동
활발한 20분간 운동
||
2,816Kcal
를 소모합니다

6시 25분에 일어나고

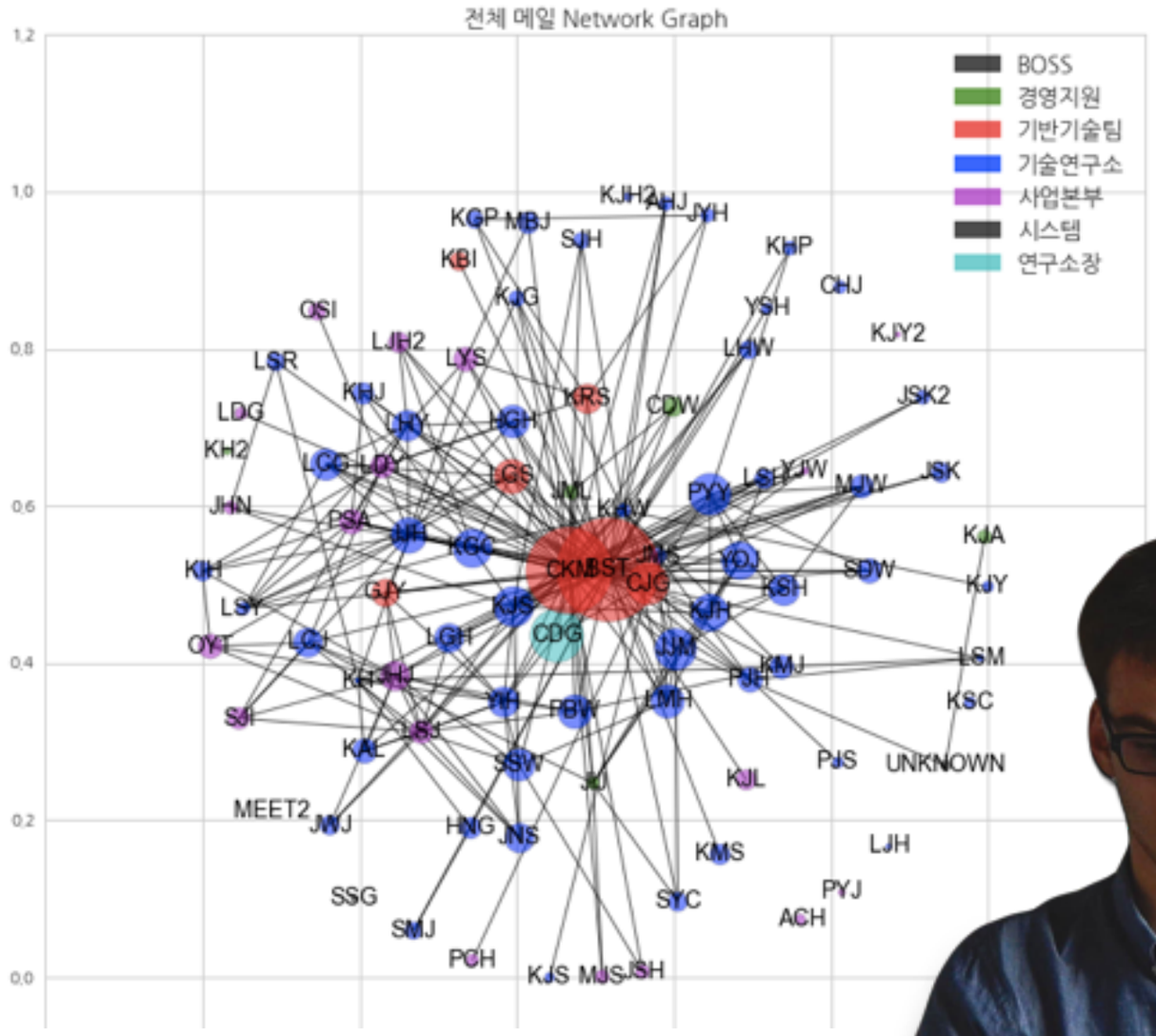
= 건강한 돼지는
될수 있음

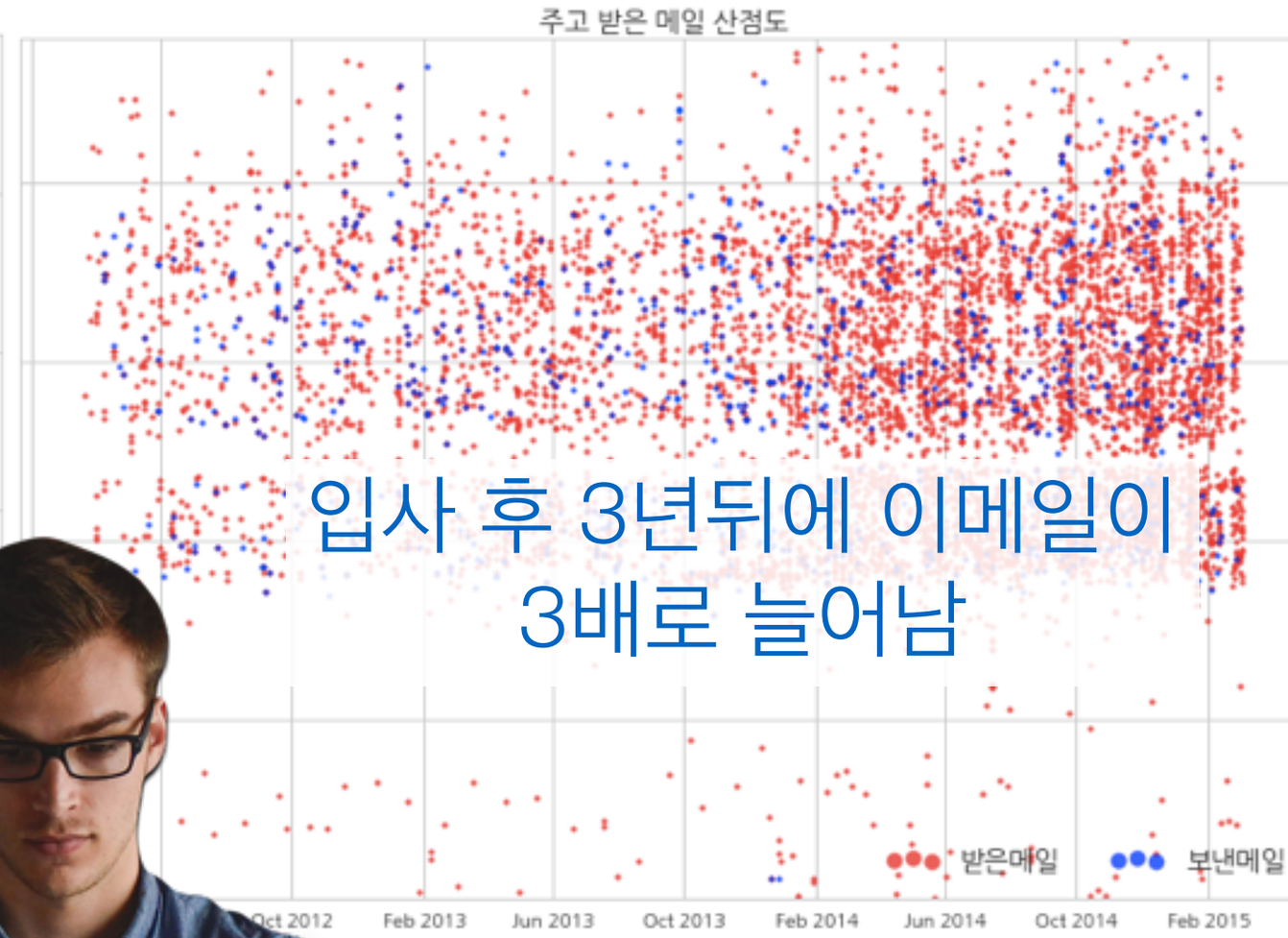
업무 PC Monitoring Agent로
일의 생산성 측정하고 있습니다.


매일 8시 22분에 출근해서
19시 1분에 퇴근하고
total
year
2014
2015
3h
6h
9h
12h
15h
18h
21h
24h
내가 검색 시 자주 사용한 단어
python
mysql
to
아프리카tv
hive
pandas
php
java
string
javascript
0
50
100
150
200
250
인터넷
36.1%
개발
33.2%
0.0%
1.2%
기타
PC운영
6.8%
Uncateg
10.9%
11.8%
문서
업무 PC를
6.5H/Day 사용



업무용 아웃룩 이메일 데이터로 업무상 연관 관계 측정 해보기



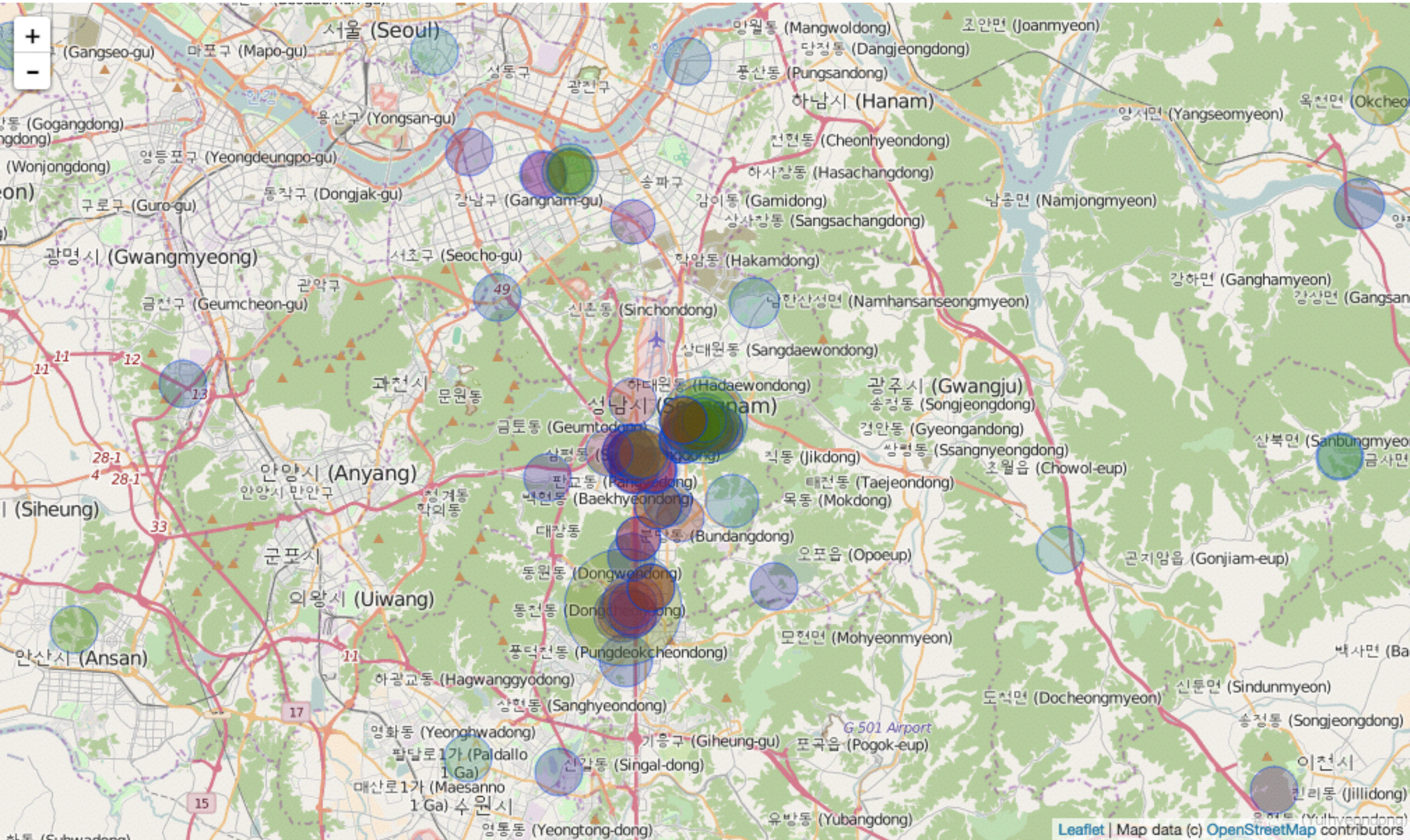


입사 후 3년뒤에 이메일이
3배로 늘어남

이메일을
9.7건 / Day
주고받고

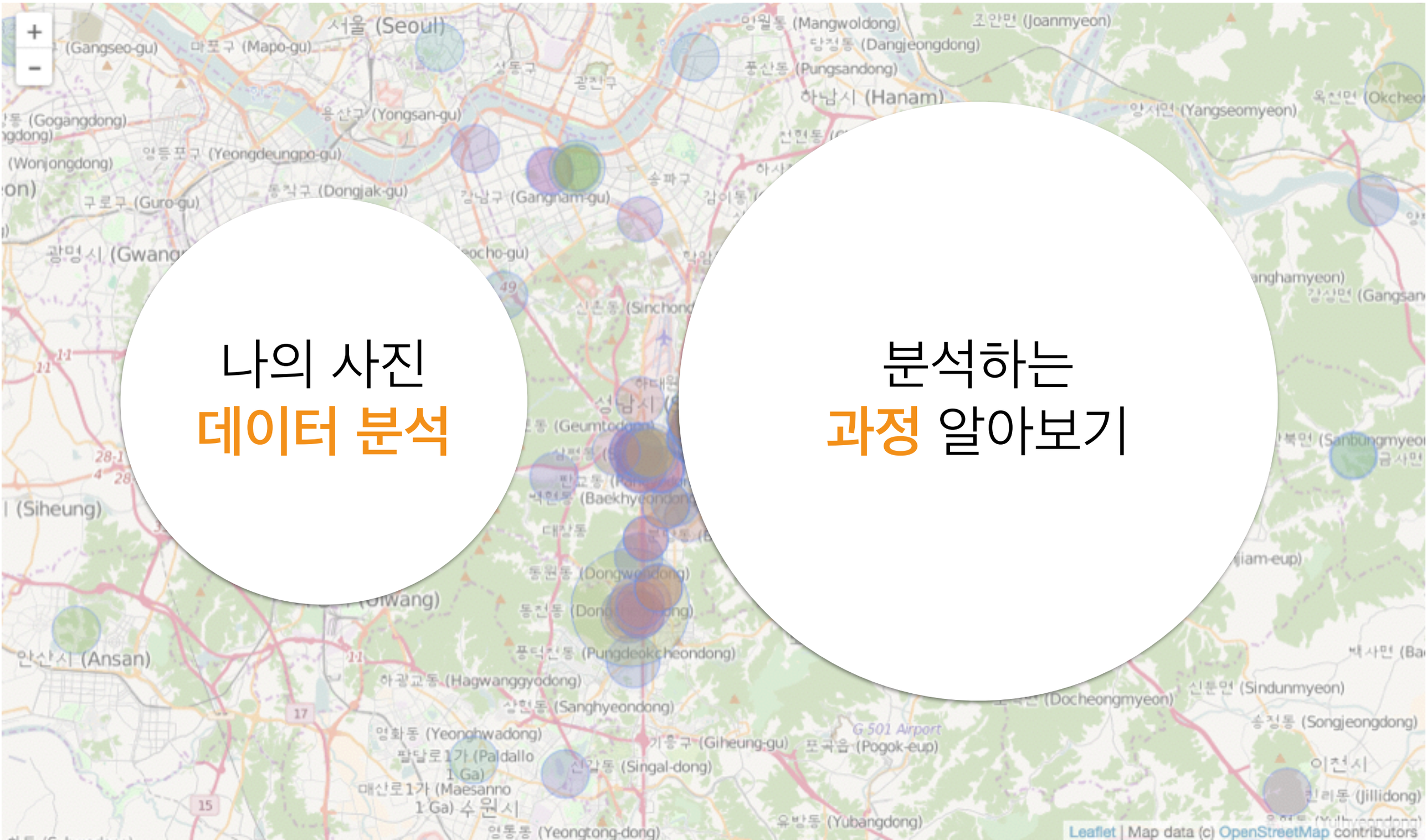
스마트폰으로 찍은 사진으로 나의 생활 반경측정 해보기

나의 생활 반경은?



Leaflet | Map data (c) OpenStreetMap contributors

오늘은 할 이야기?



나의 사진
데이터 분석

분석하는
과정 알아보기

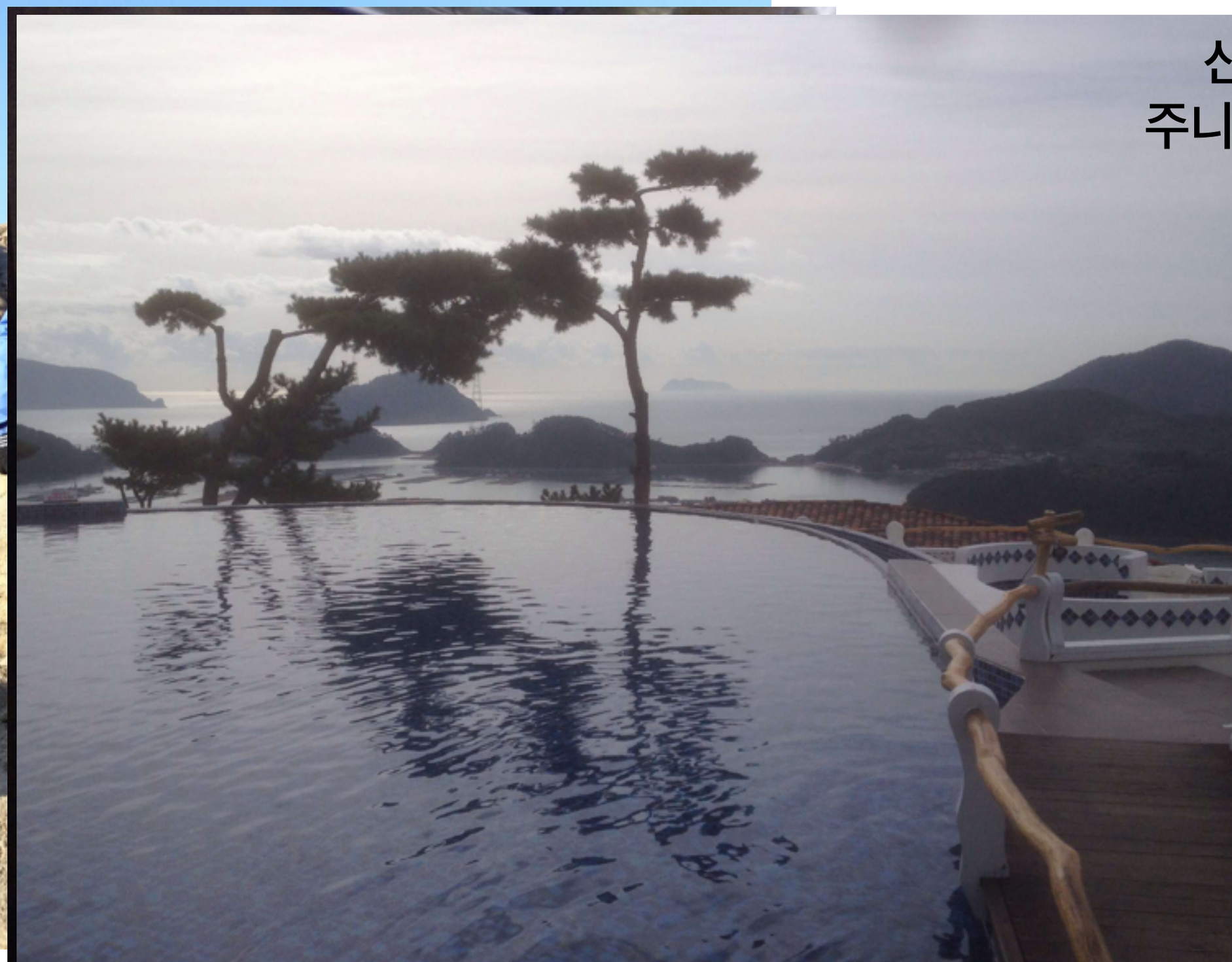
언제 사진을찍을까요?

저의 **개인적인 경우** 입니다.

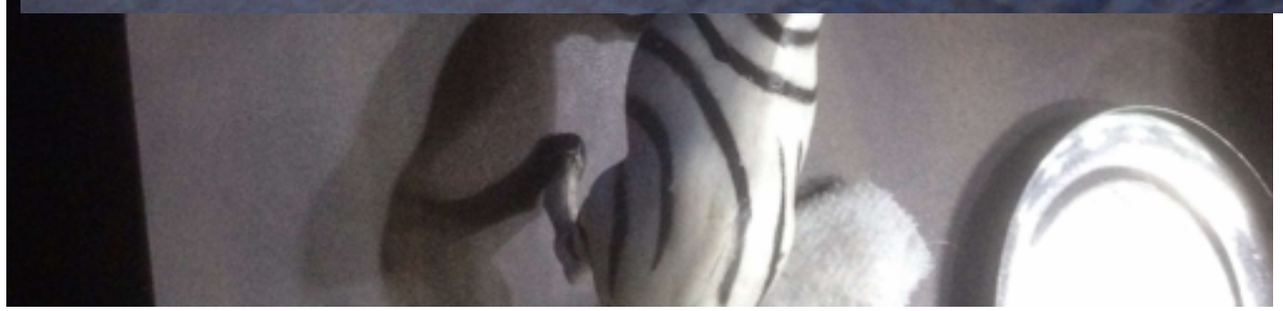
산 정상에 올랐을때

산 정상에 올랐을때

규민 주니어가 동물에 밥줄때



산에 정상에 올랐을때
주니어가 동물에 밥줄때
멋진곳에 왔을때



산에 정상에 올랐을때
주니어가 동물에 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때
머리를 깍았을때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때
머리를 깍았을때
증거 남길때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때
머리를 깍았을때
증거 남길때
캠핑 할때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때
머리를 깍았을때
증거 남길때
캠핑 할때
옷 샀을때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때
머리를 깍았을때
증거 남길때
캠핑 할때
옷 샀을때
발표 할때

저에게 사진은
중요한 순간을
기록합니다.

산에 정상에 올랐을때
밥줄때
을때
때
때
때
때
할때
발표 할때

산에 정상에 올랐을때
주니어가 동물 밥줄때
멋진곳에 왔을때
맛난 것을 먹을때
온라인으로 살 물건 봤을때
머리를 깍았을때
증거 남길때
캠핑 할때
발표 할때

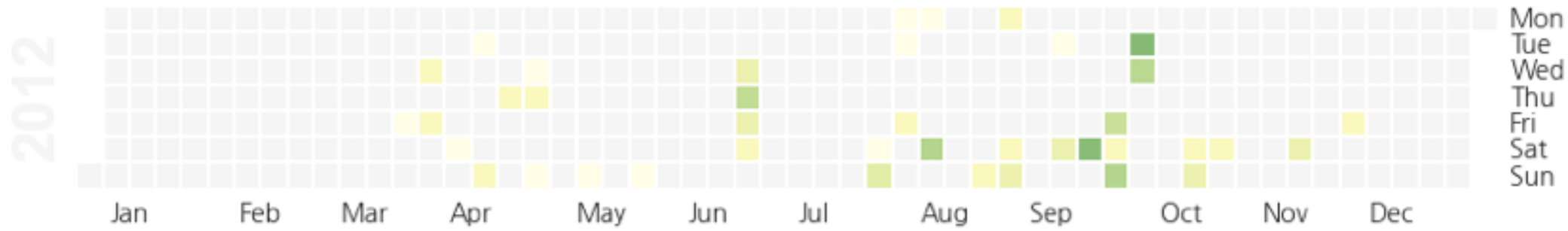
그래서 사진 데이터를 파

봤습니다.

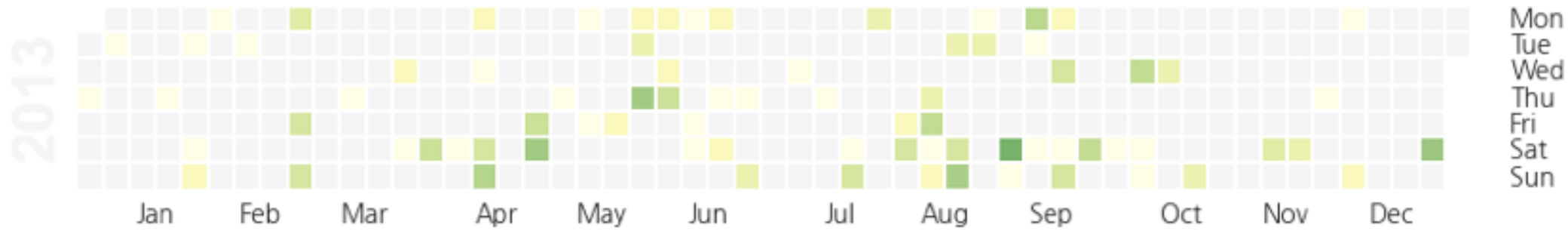
IPhone으로 ‘2012년 3월 ~ 2016년 7월’ 찍은
사진 1,595장을 분석

사진을 얼마나 찍었는가?

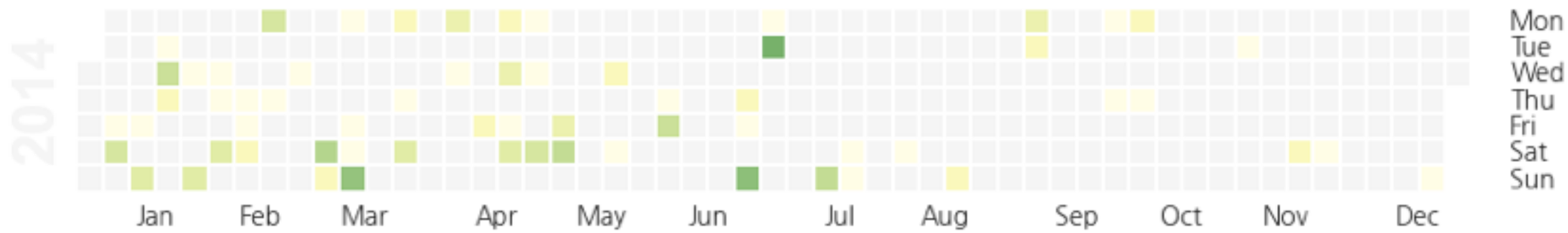
138장



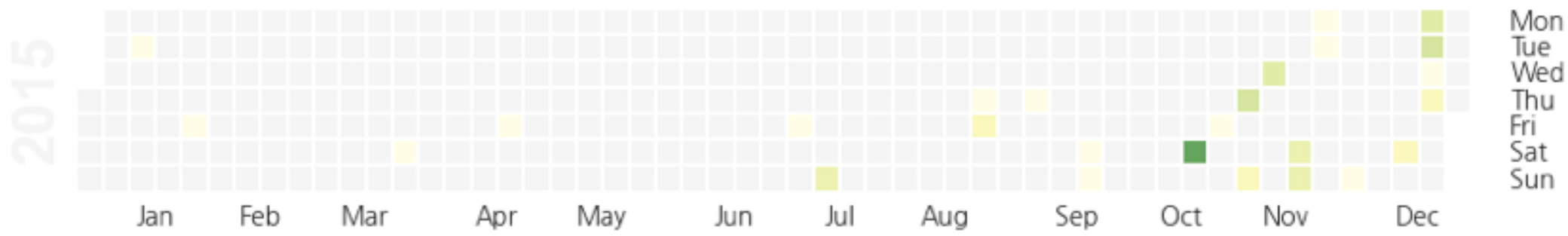
262장



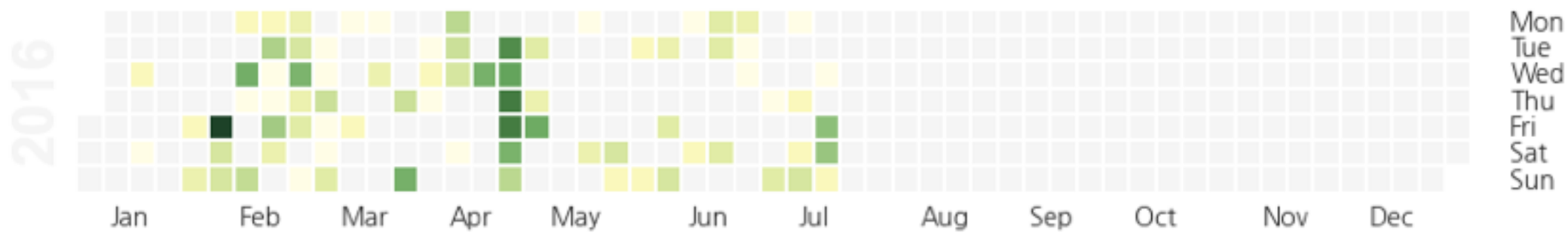
193장



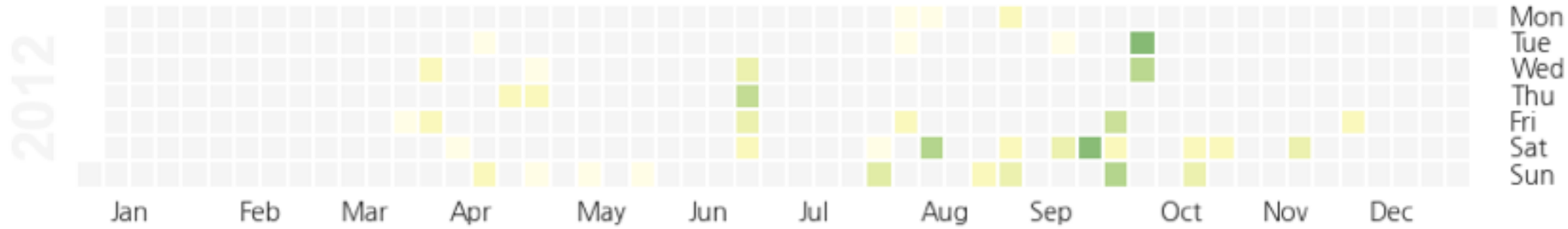
83장



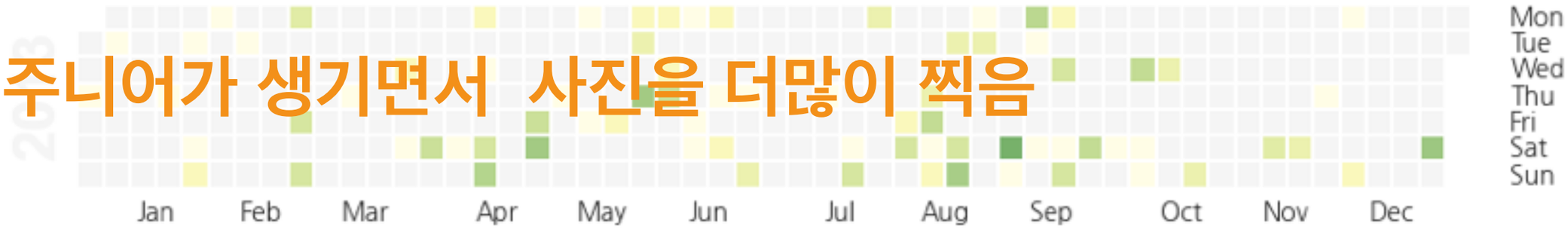
919장



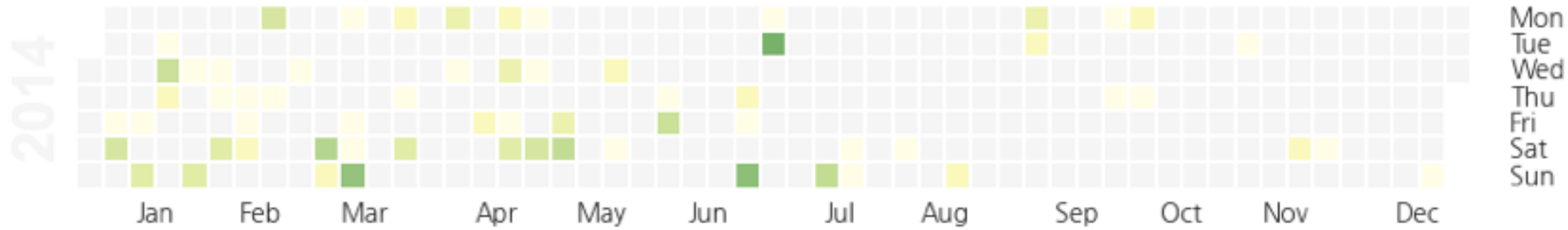
138장



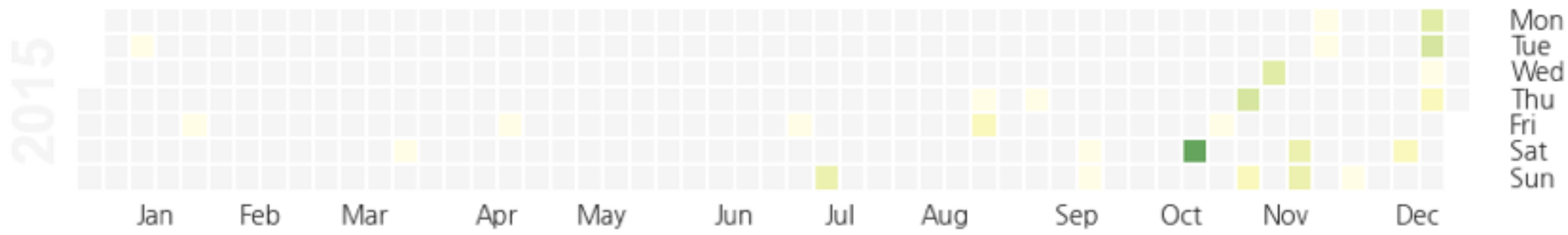
262장 : 주니어가 생기면서 사진을 더많이 찍음



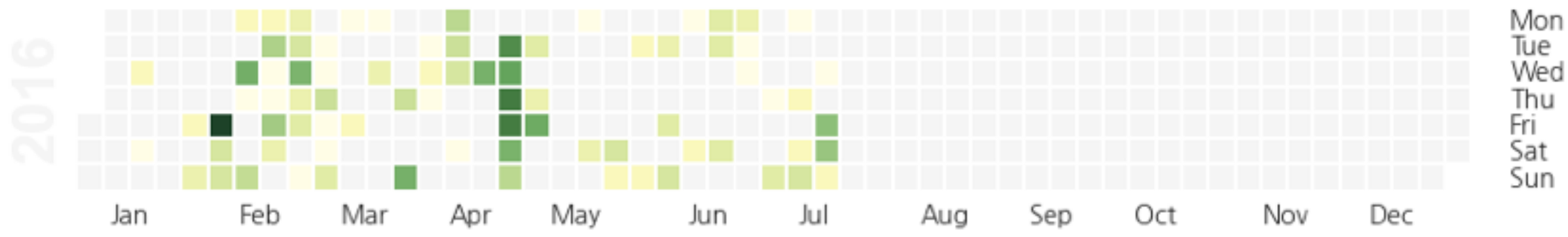
193장



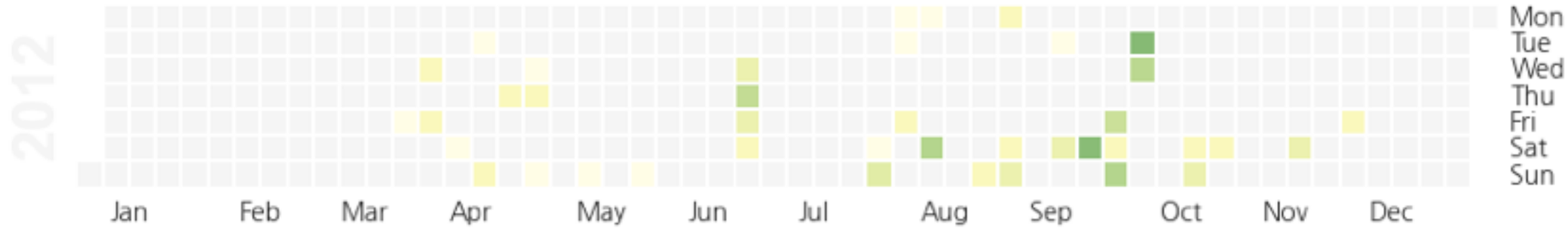
83장



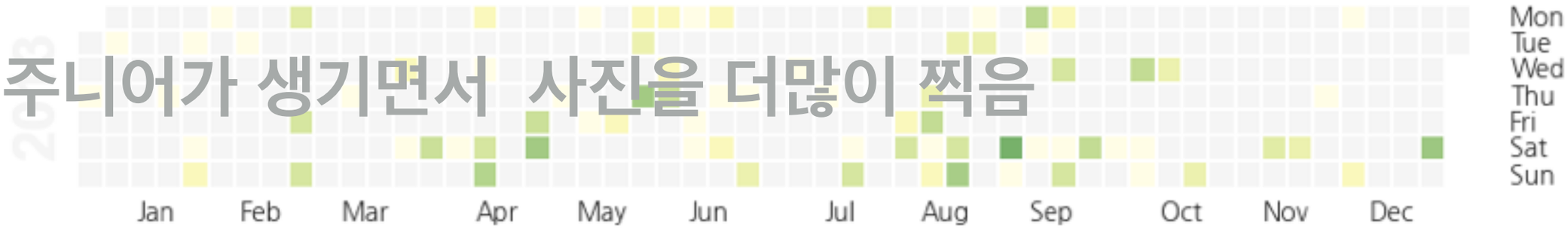
919장



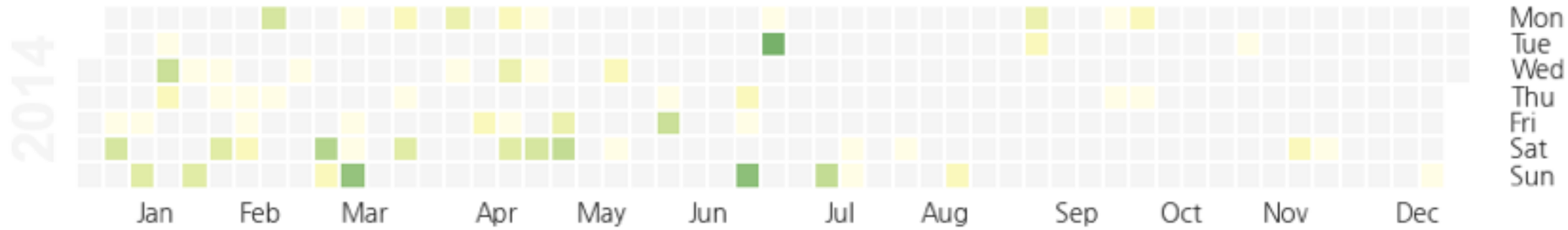
138장



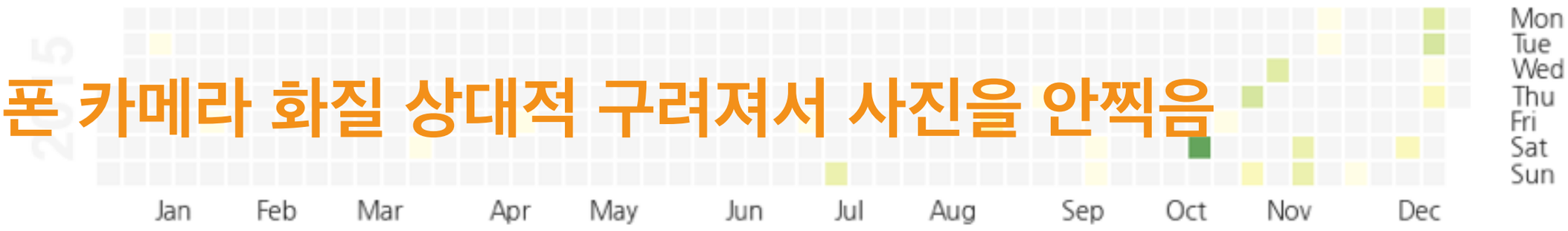
262장 : 주니어가 생기면서 사진을 더많이 찍음



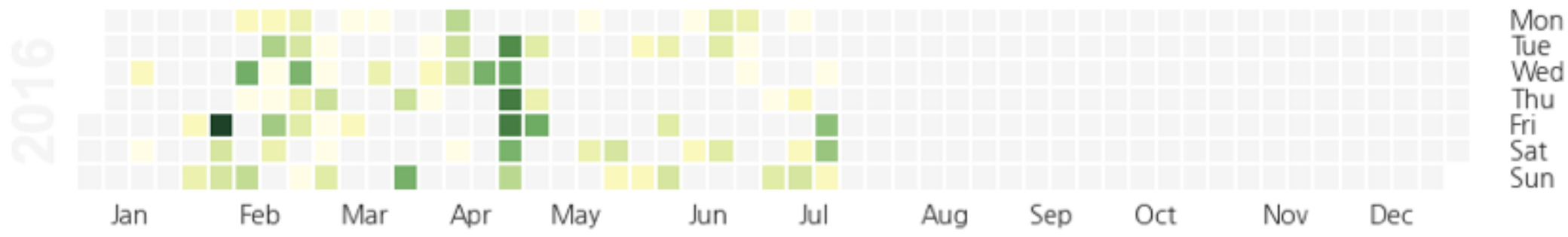
193장



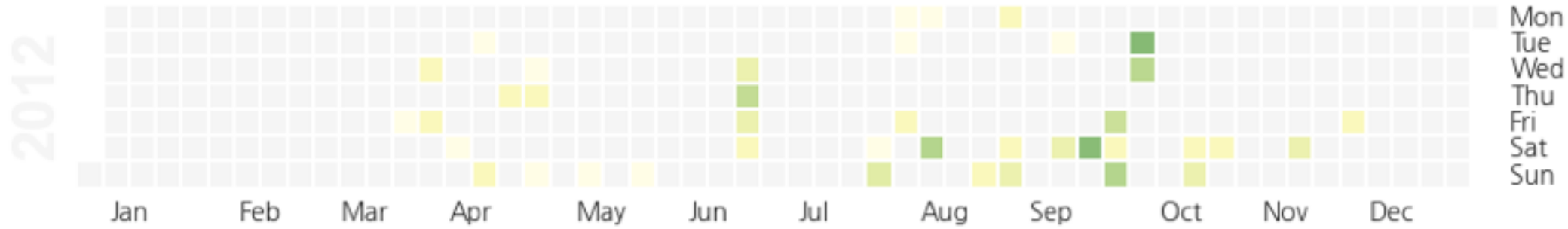
83장 : 폰 카메라 화질 상대적 구려져서 사진을 안찍음



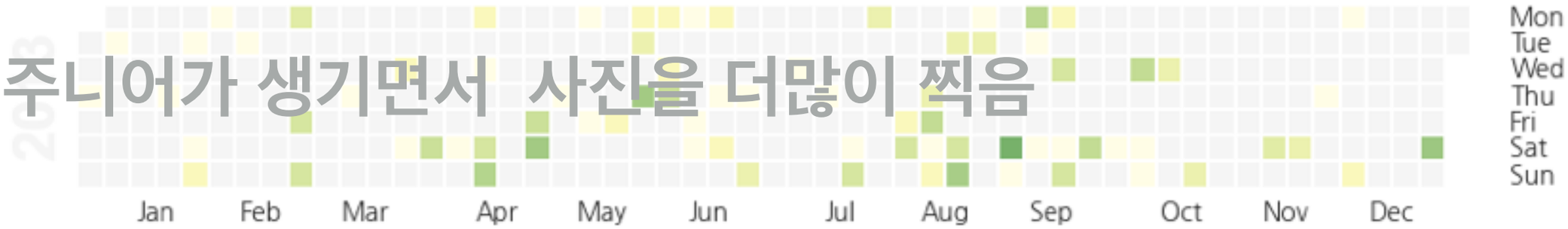
919장



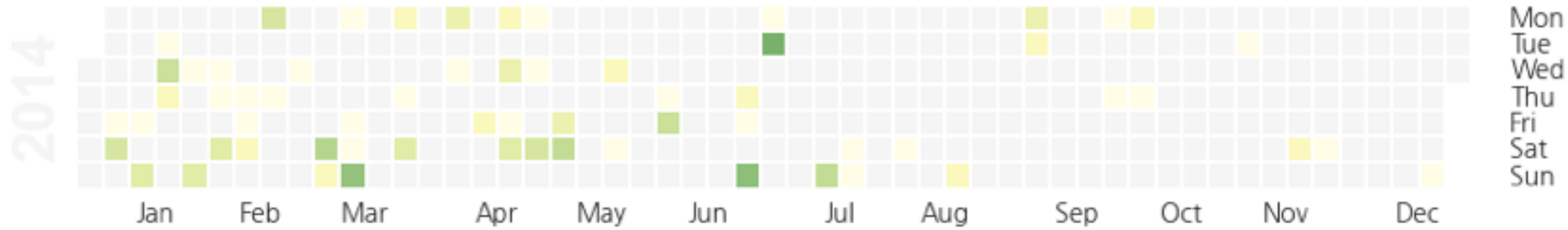
138장



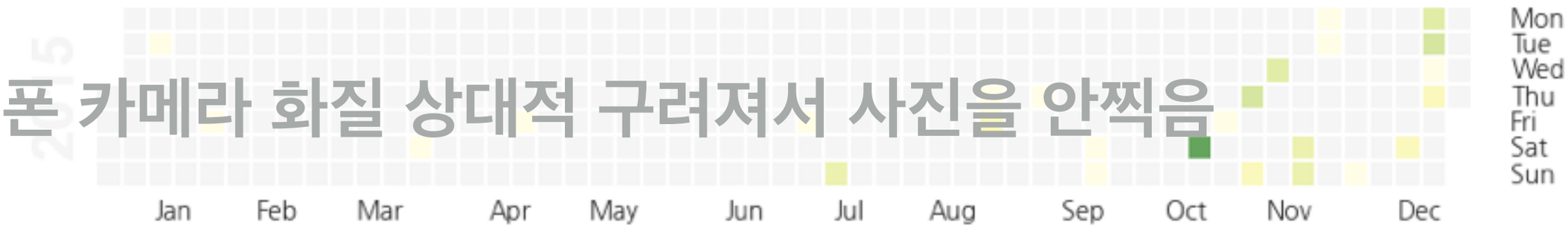
262장 : 주니어가 생기면서 사진을 더많이 찍음



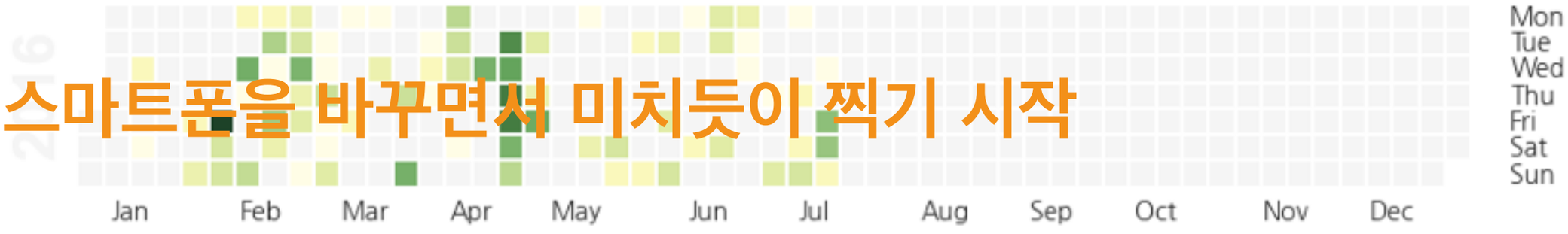
193장



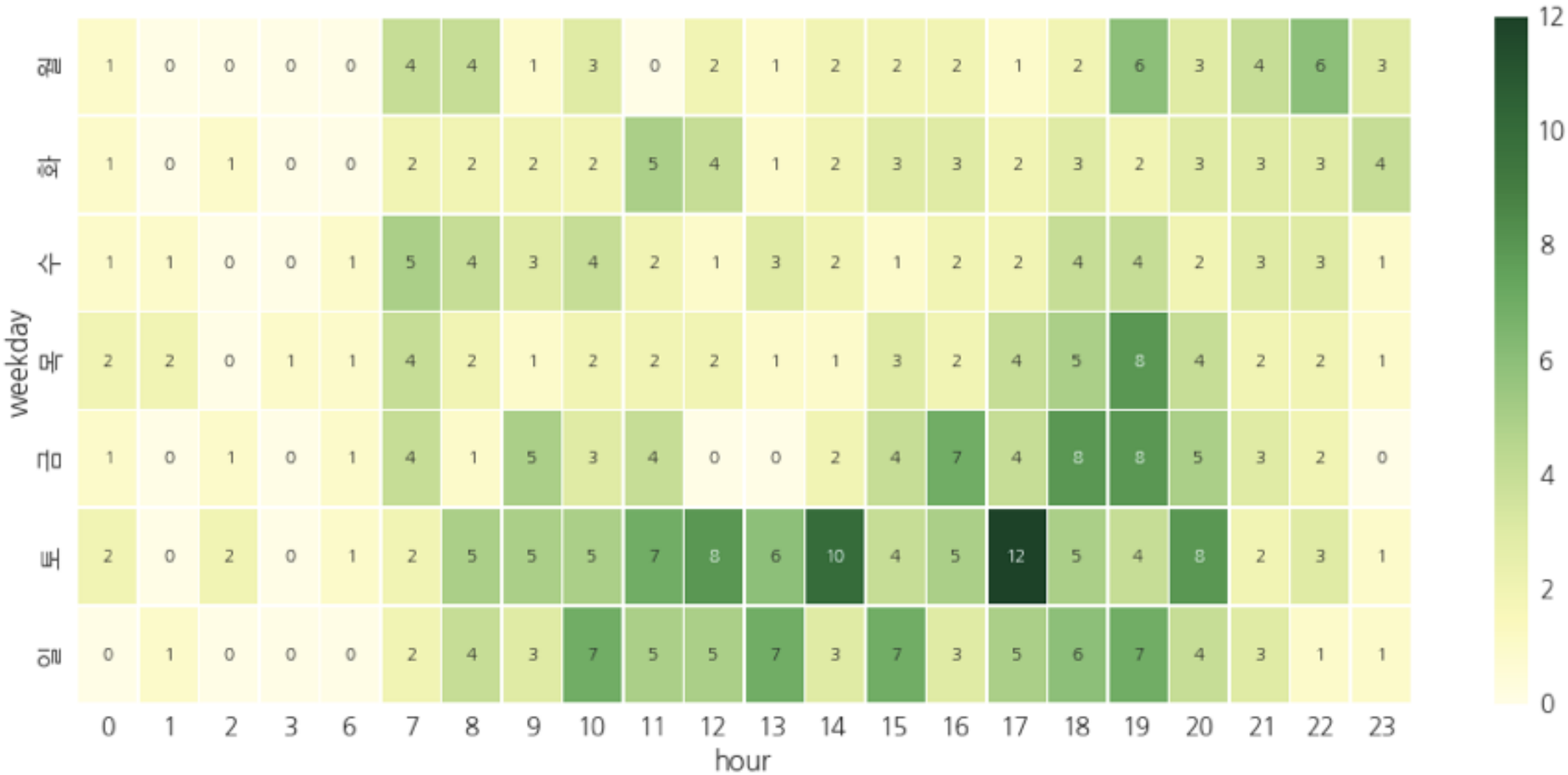
83장 : 폰 카메라 화질 상대적 구려져서 사진을 안찍음



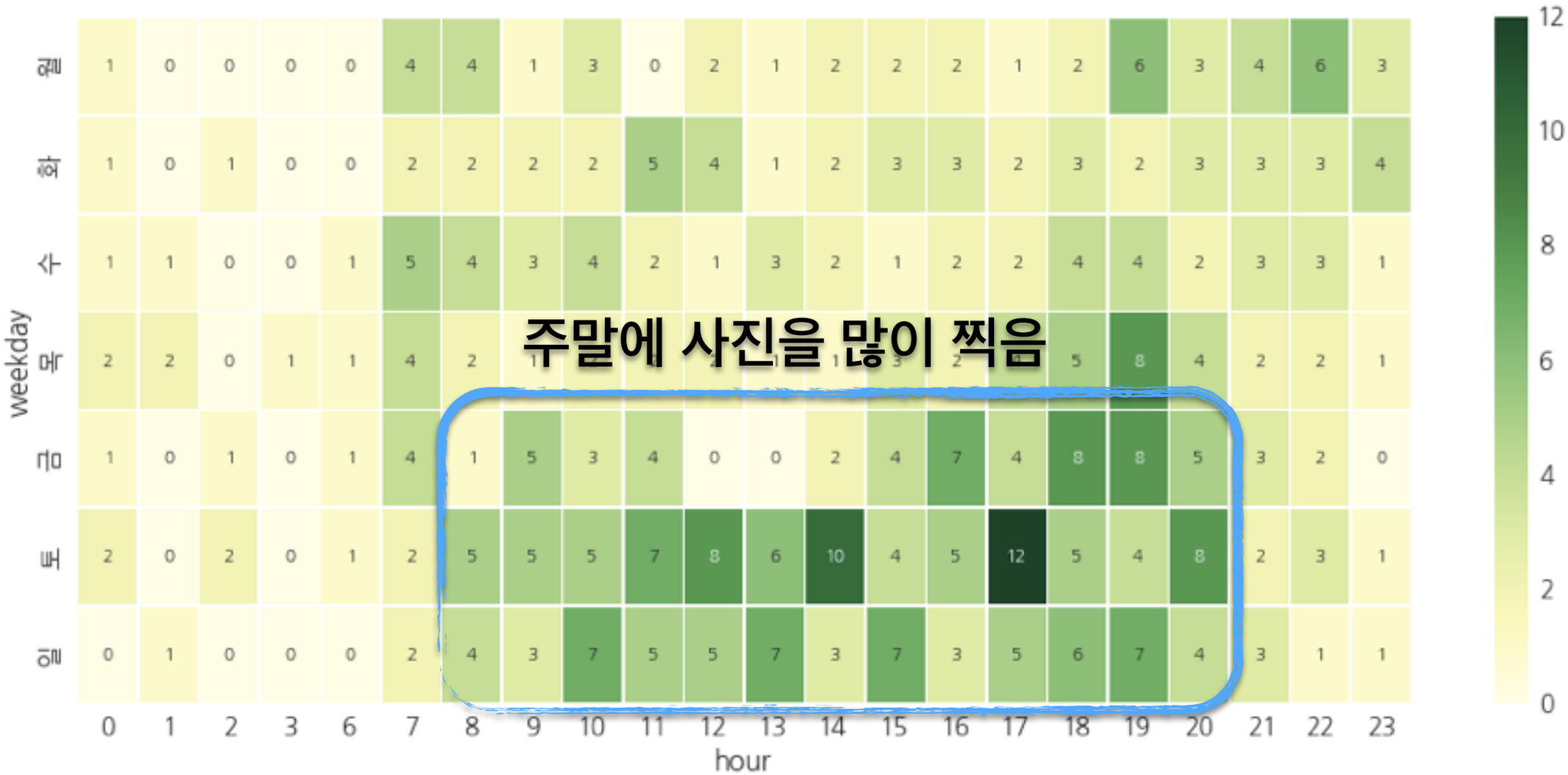
919장 : 스마트폰을 바꾸면서 미치듯이 찍기 시작



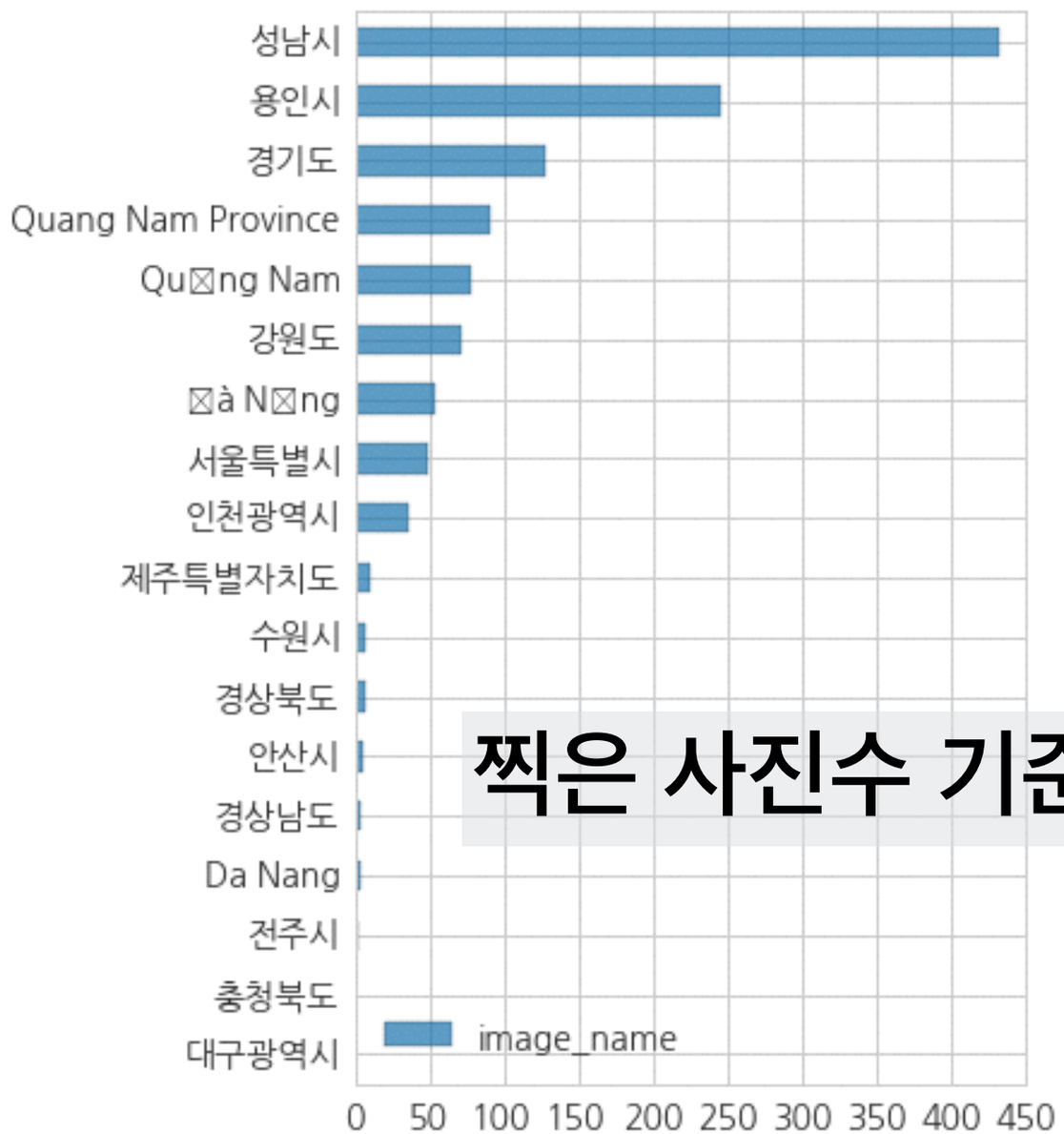
시간과 요일에 따라서는 어떻게 다른가?



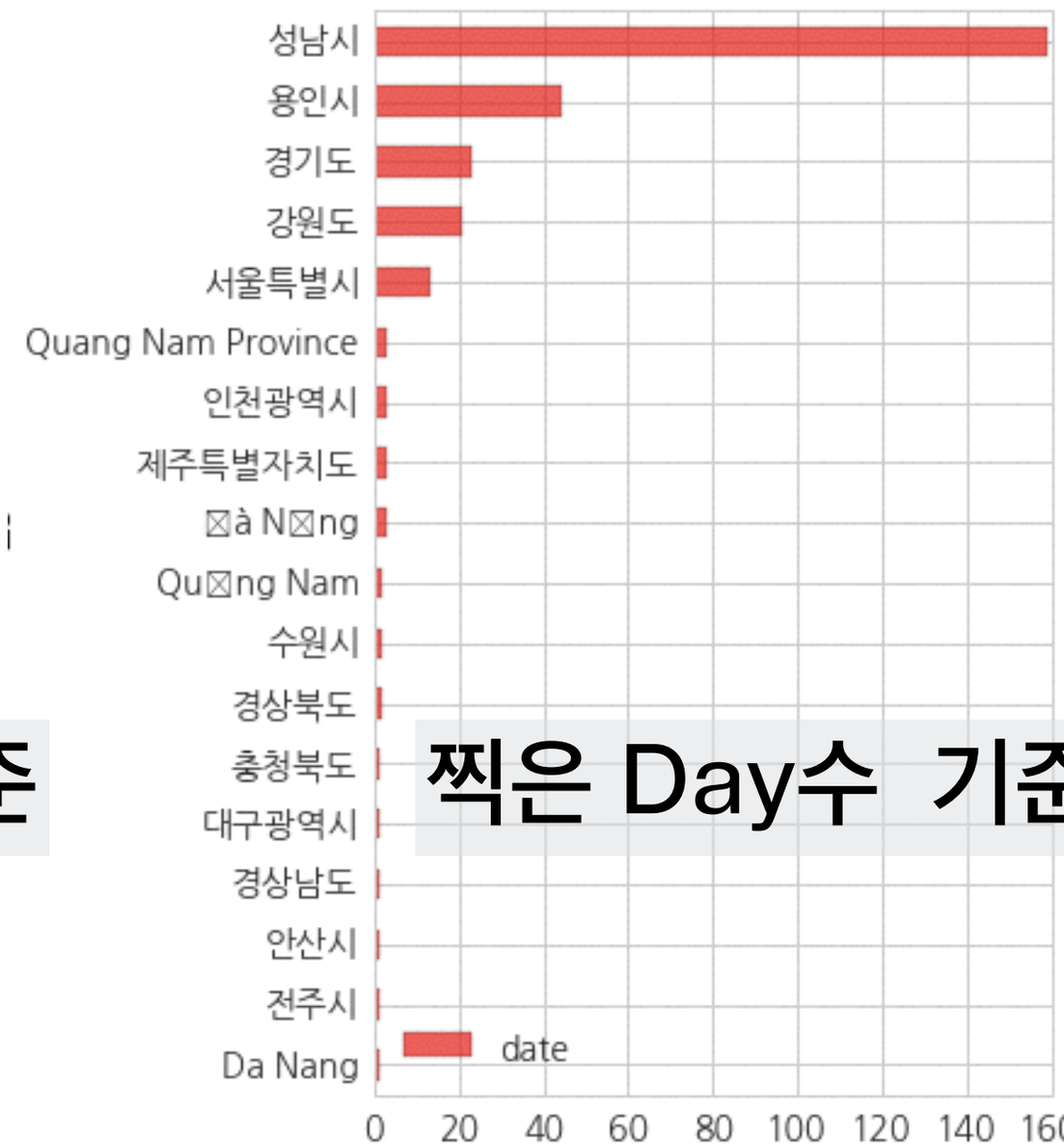
시간과 요일에 따라서는 어떻게 다른가?



어디에서 많이 찍었는가?



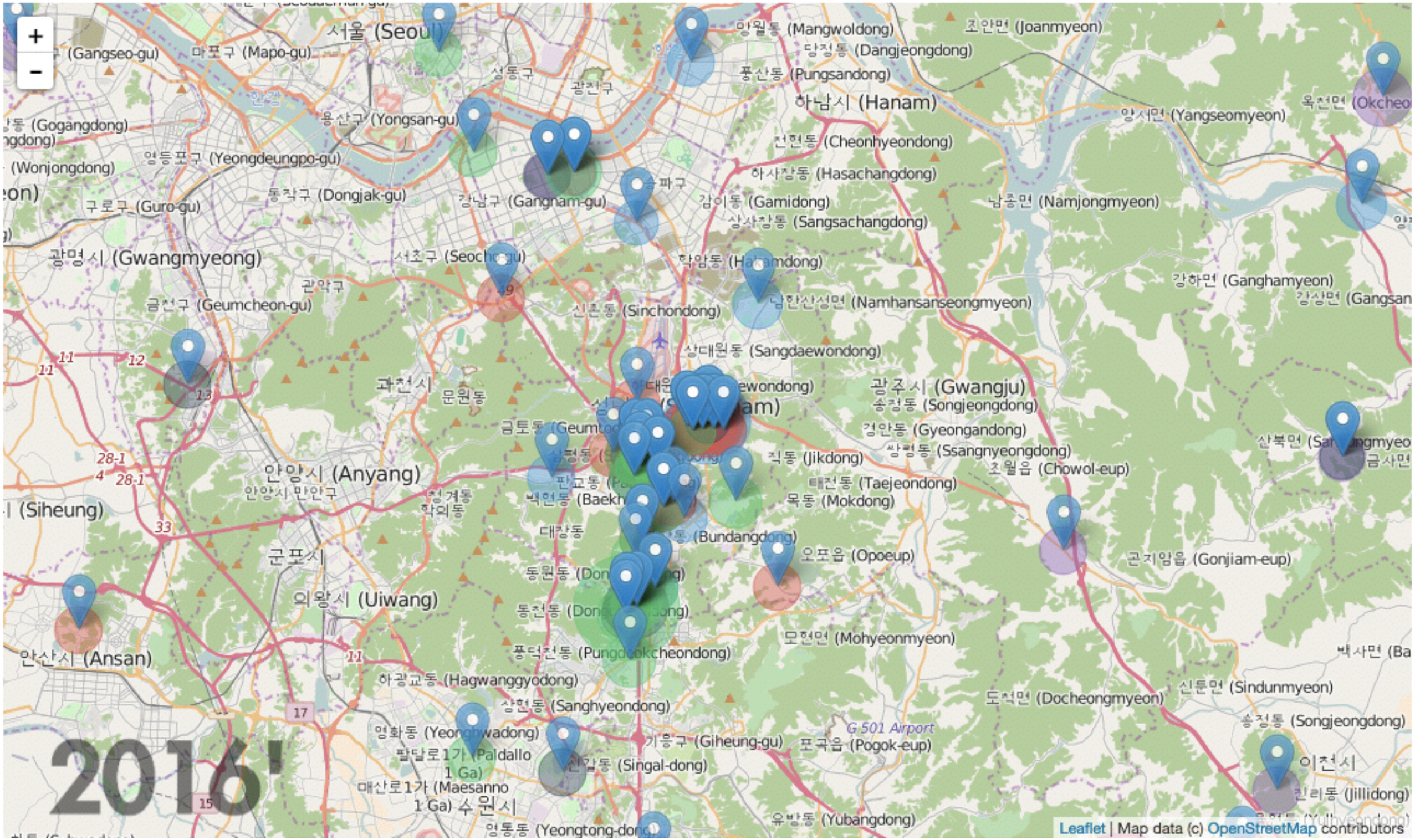
찍은 사진수 기준



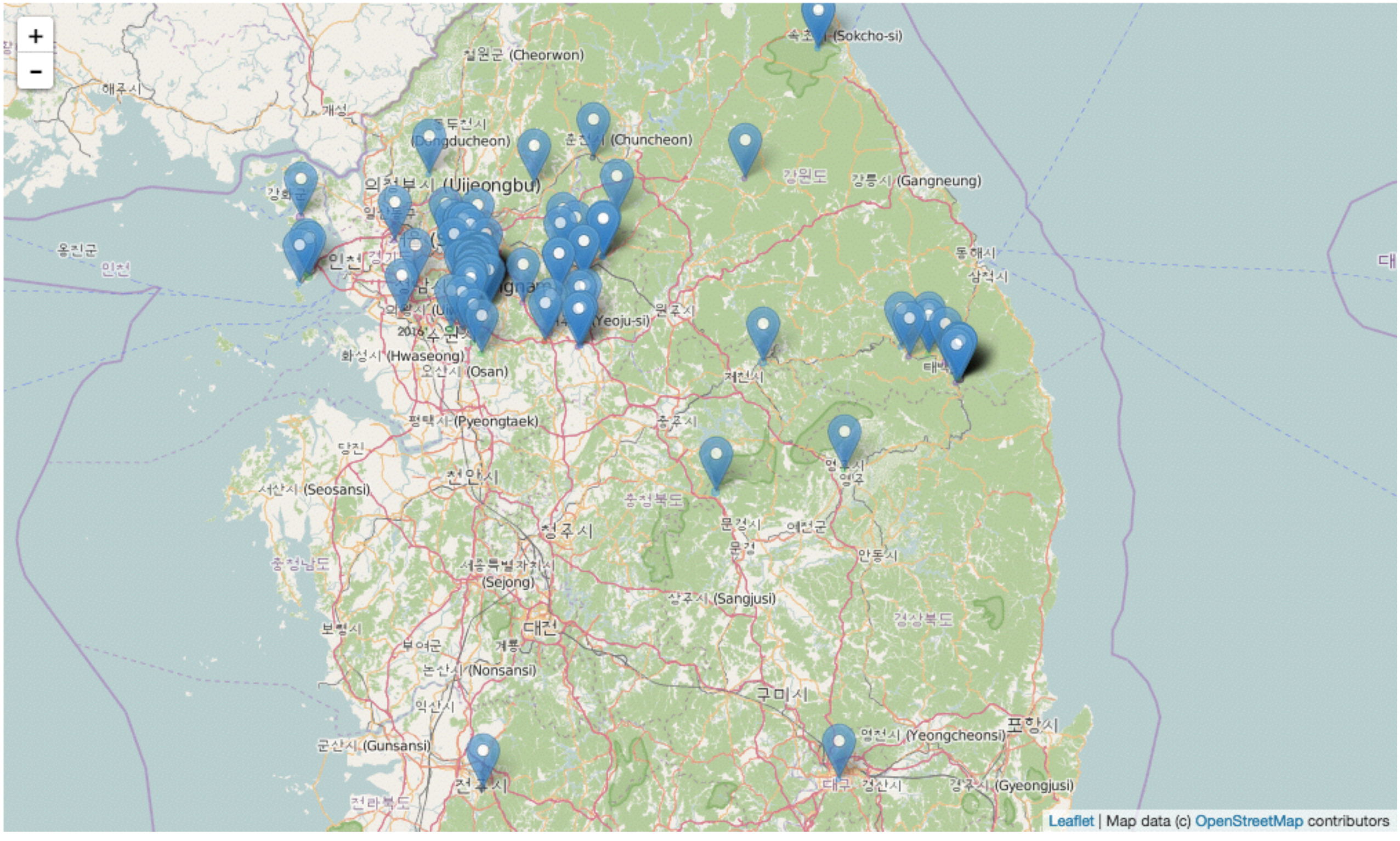
찍은 Day수 기준

나의 활동 무대는?

야탑, 판교, 수지 주요 생활권

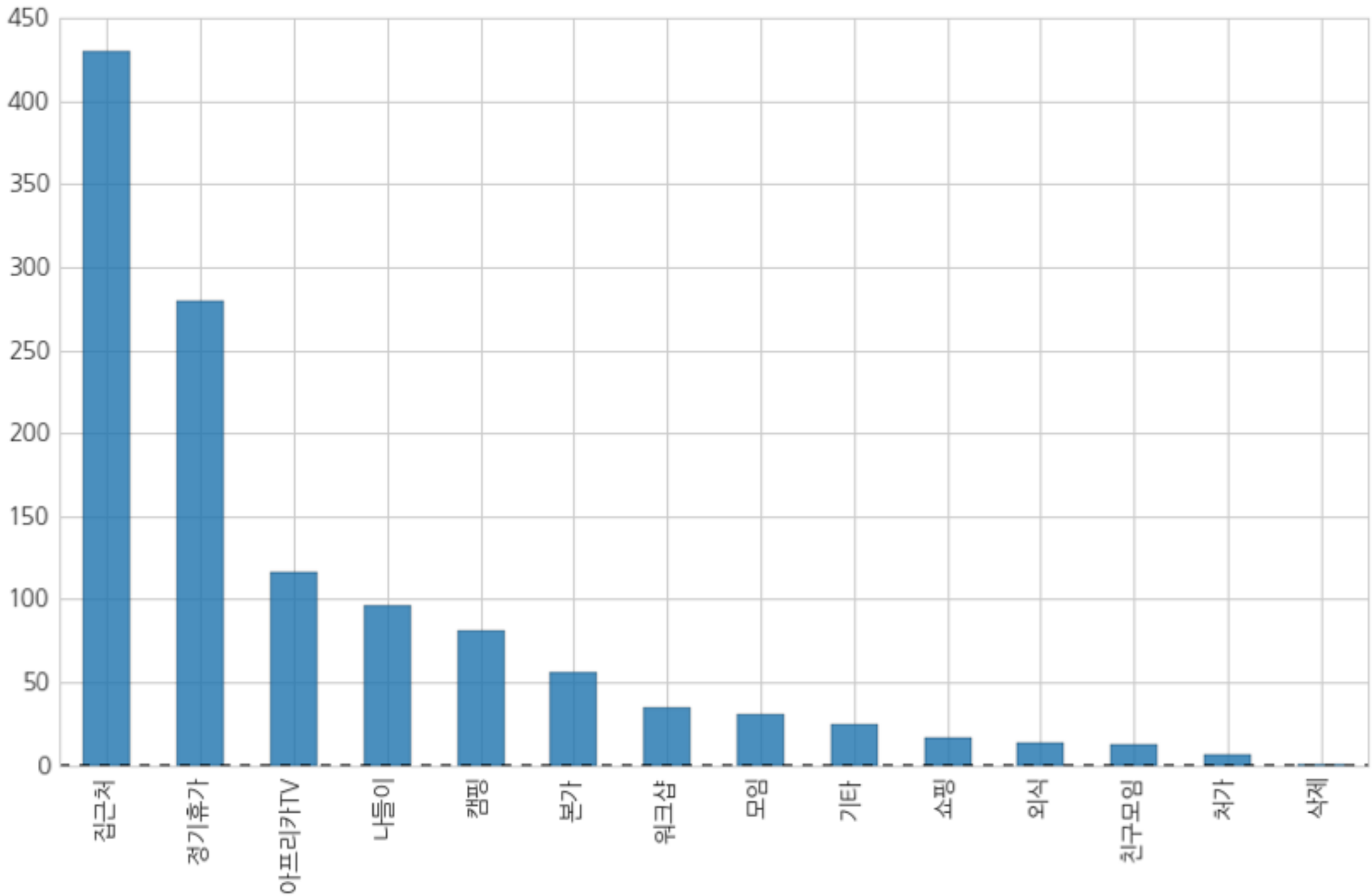


전국적으로는 태백,여주,양평

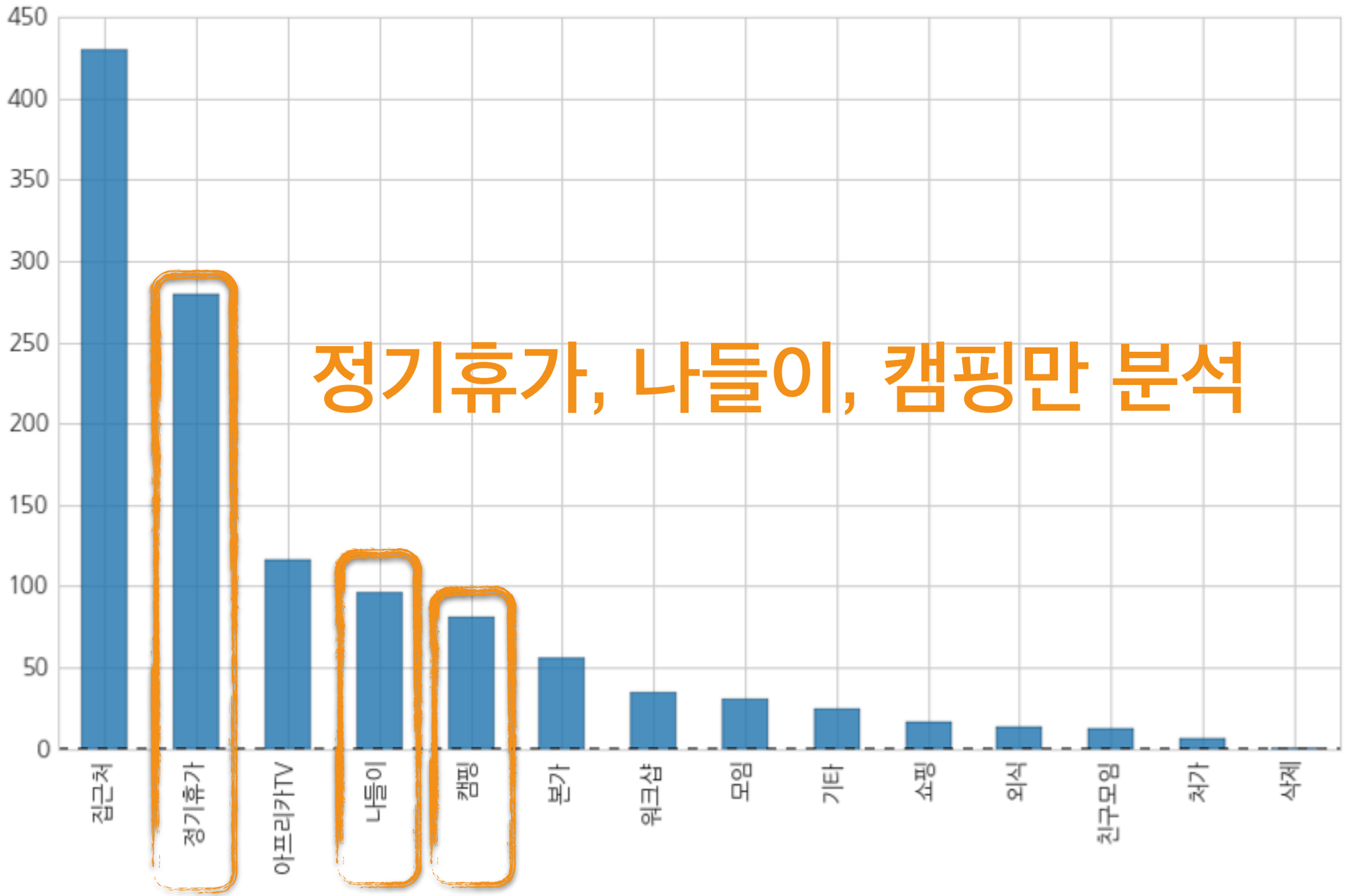


여행은 어디로 갔을까?

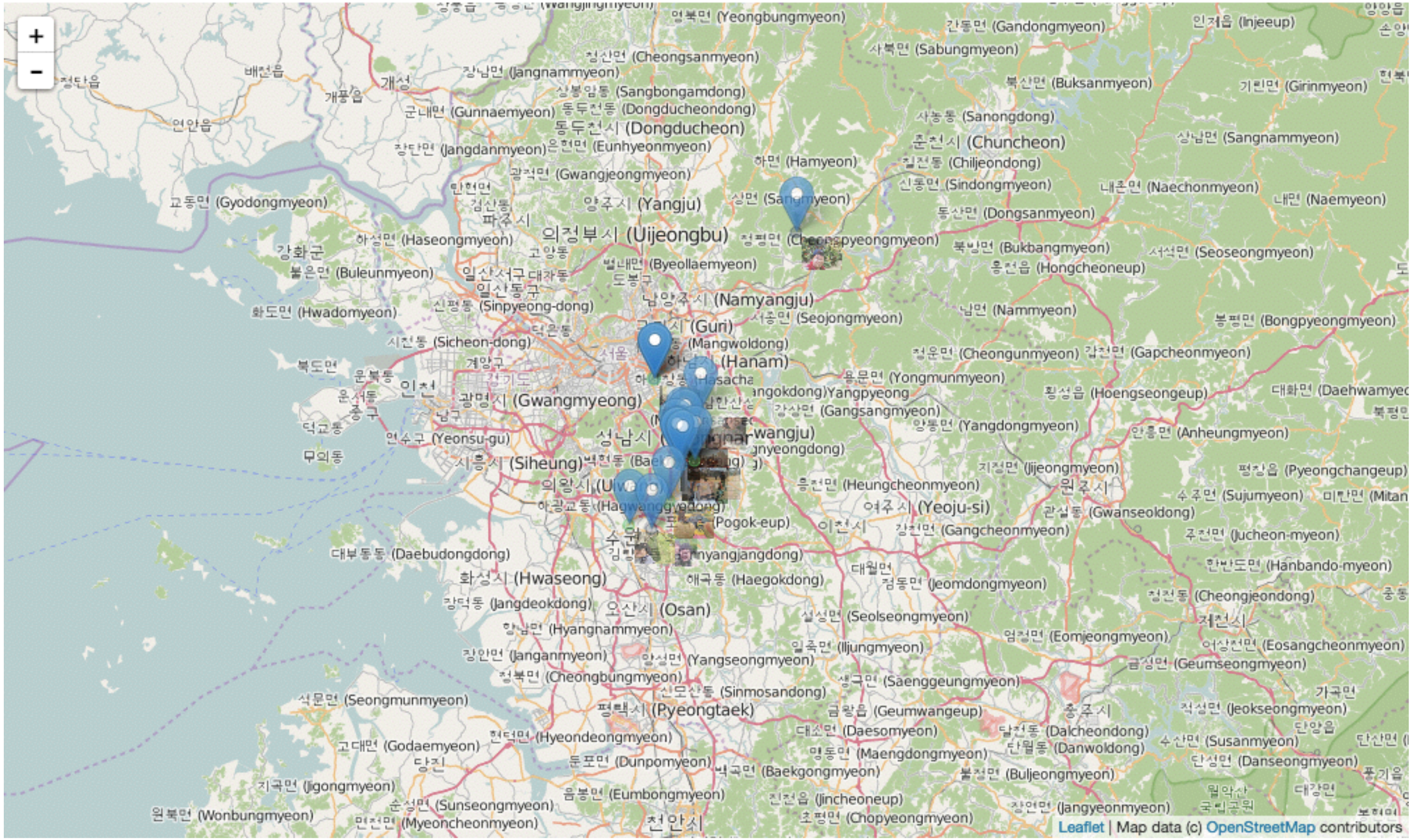
사진의 위치 정보로 분류하면



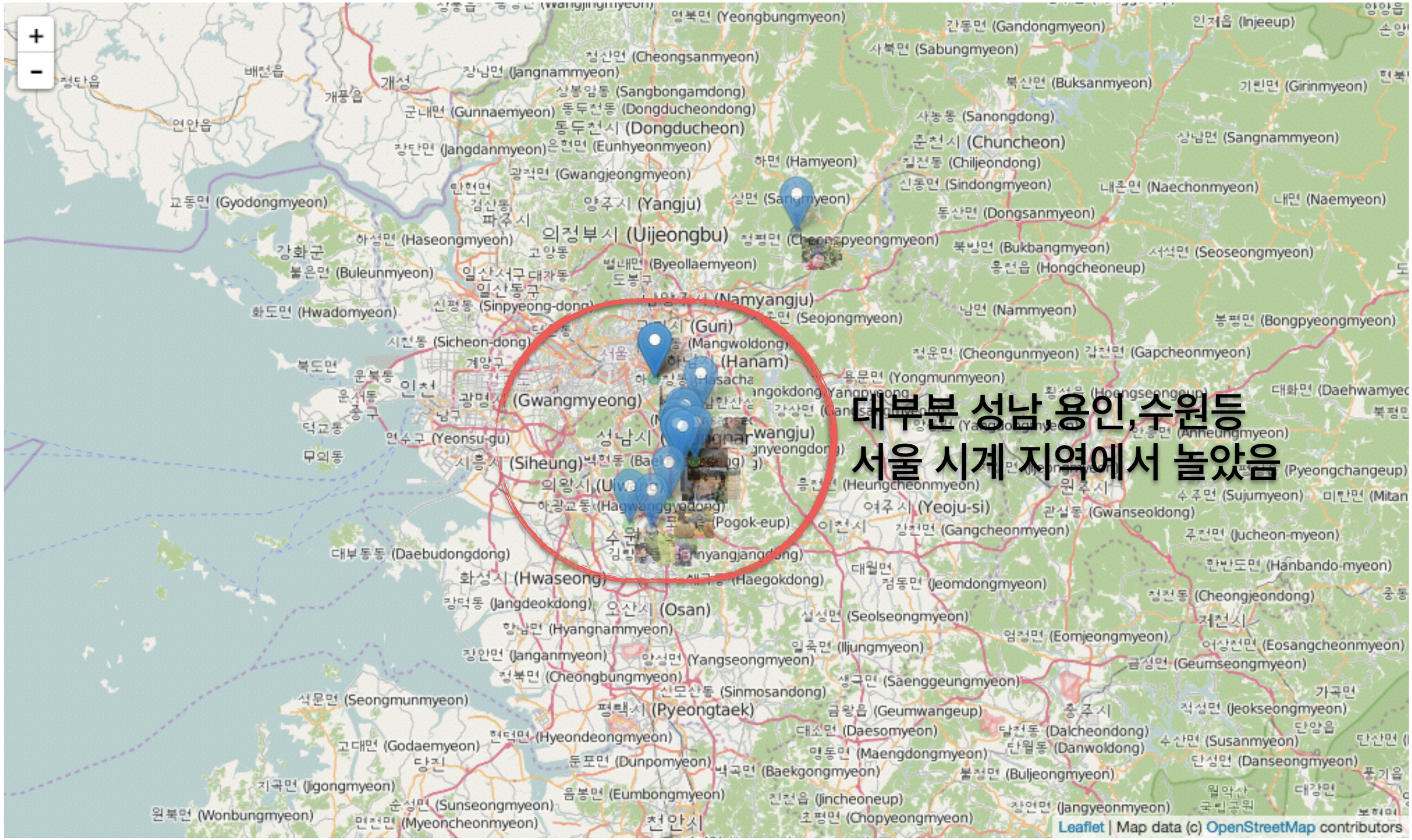
여행관련된 항목만 선택


450
400
350
300
250
200
150
100
50
0
집근처
정기휴가
아프리카TV
나들이
캠핑
본가
워크샵
모임
기타
쇼핑
외식
친구모임
처가
삭제
정기휴가, 나들이, 캠핑만 분석

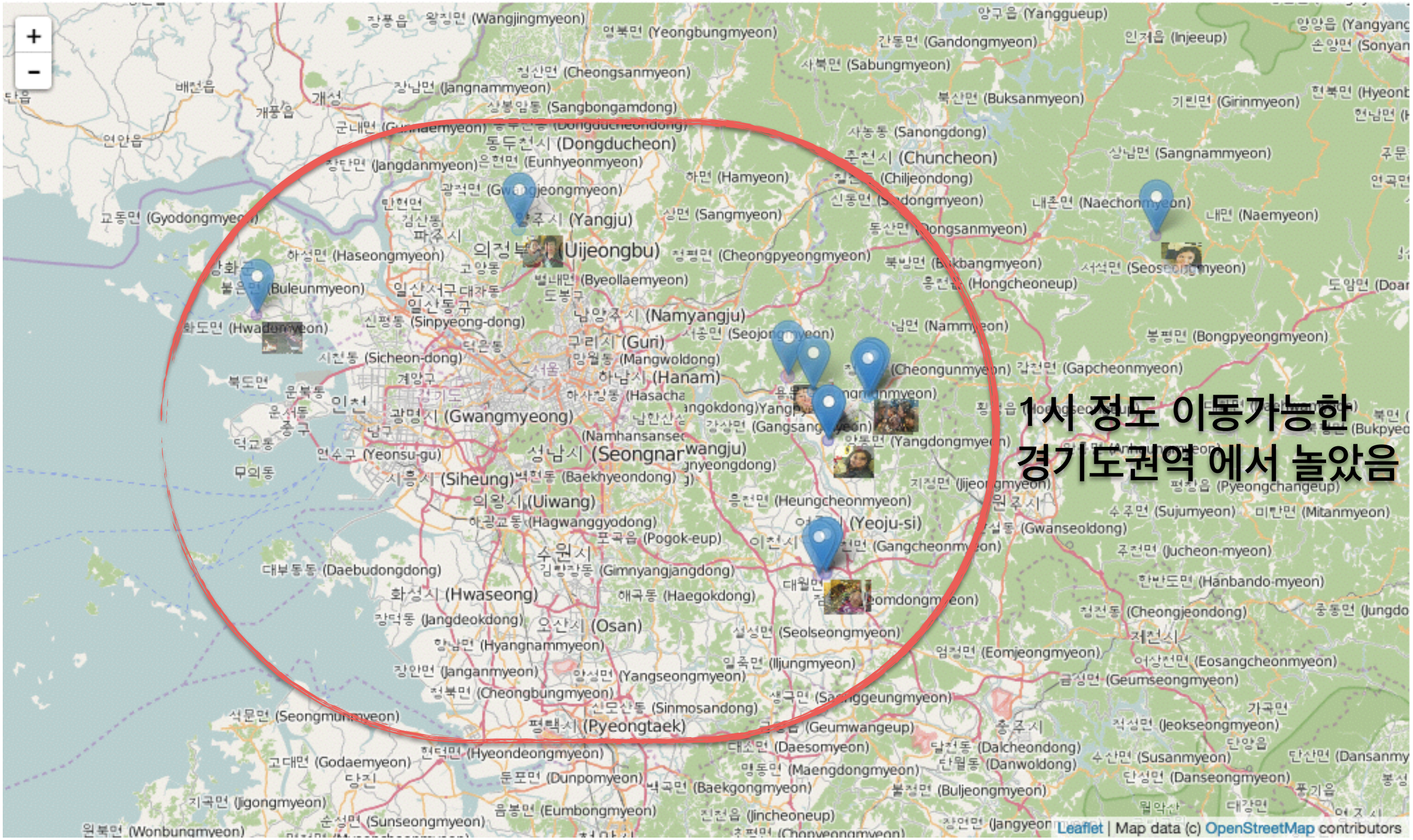
당일치기 나들이를 간곳은?



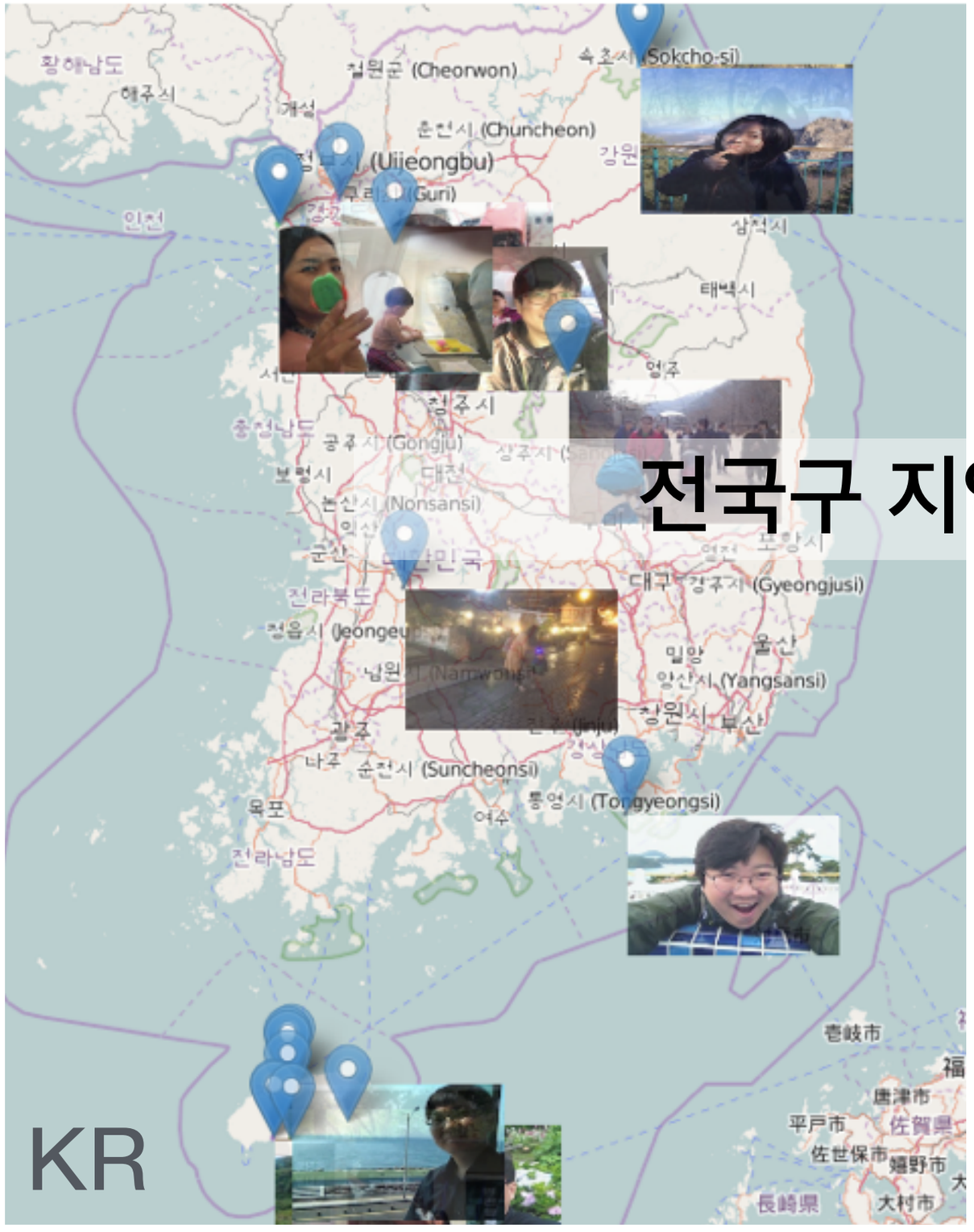
당일치기 나들이를 간곳은?



1~2박 캠핑을 간곳은?


1시 정도 이동가능한
경기도권역 에서 놀았음
Leaflet | Map data (c) OpenStreetMap contributors

장박 정기휴가를 간곳은?



전국구 지역에서 놀았음

여행에 주어진 기간

≈

이동 거리가 비례

사진의 얼굴을 분석해 보면?

미소 천사는 뉴구?

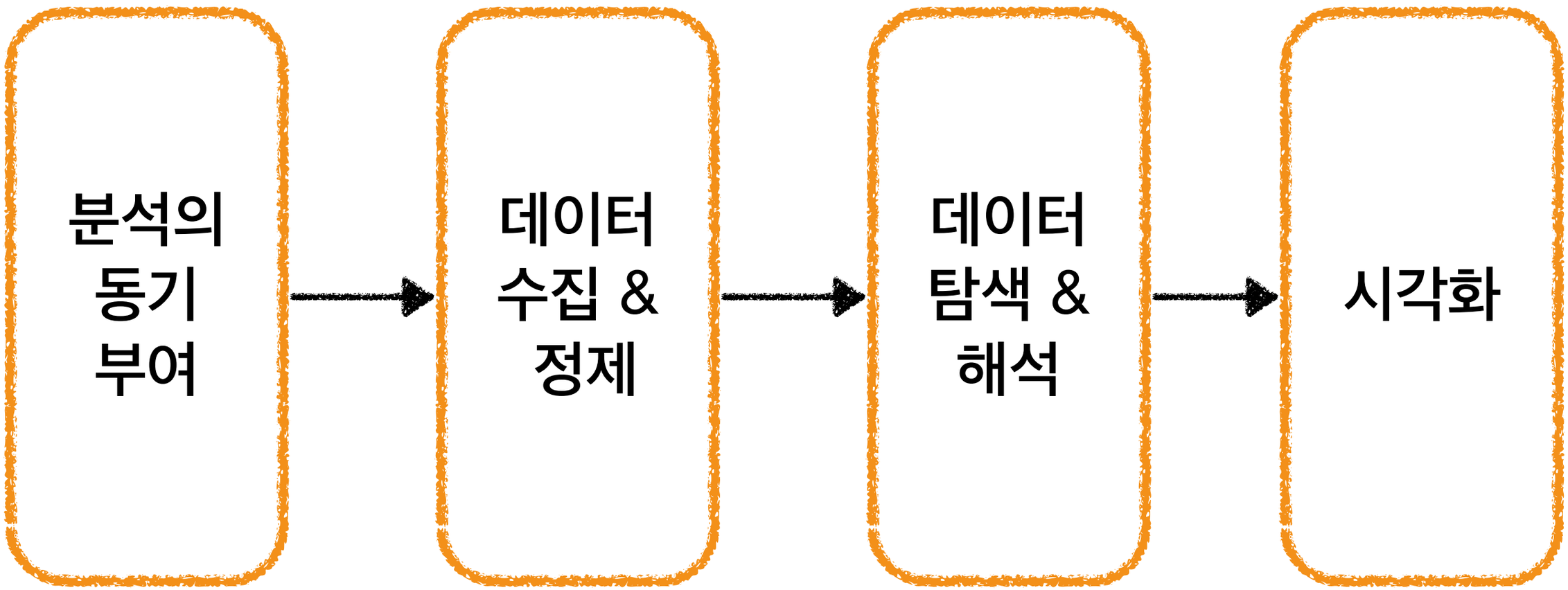
```
웃음지수 [ 99.228   97.7974  97.0466  96.8194  96.7165  96.4635  96.1923  96.0079
  95.7384  95.6591  95.2065  94.7131  94.4222  94.4134  94.301   93.9689
  93.3251  93.118   93.0164  92.7536]
```

노안 얼굴은 뉴구?

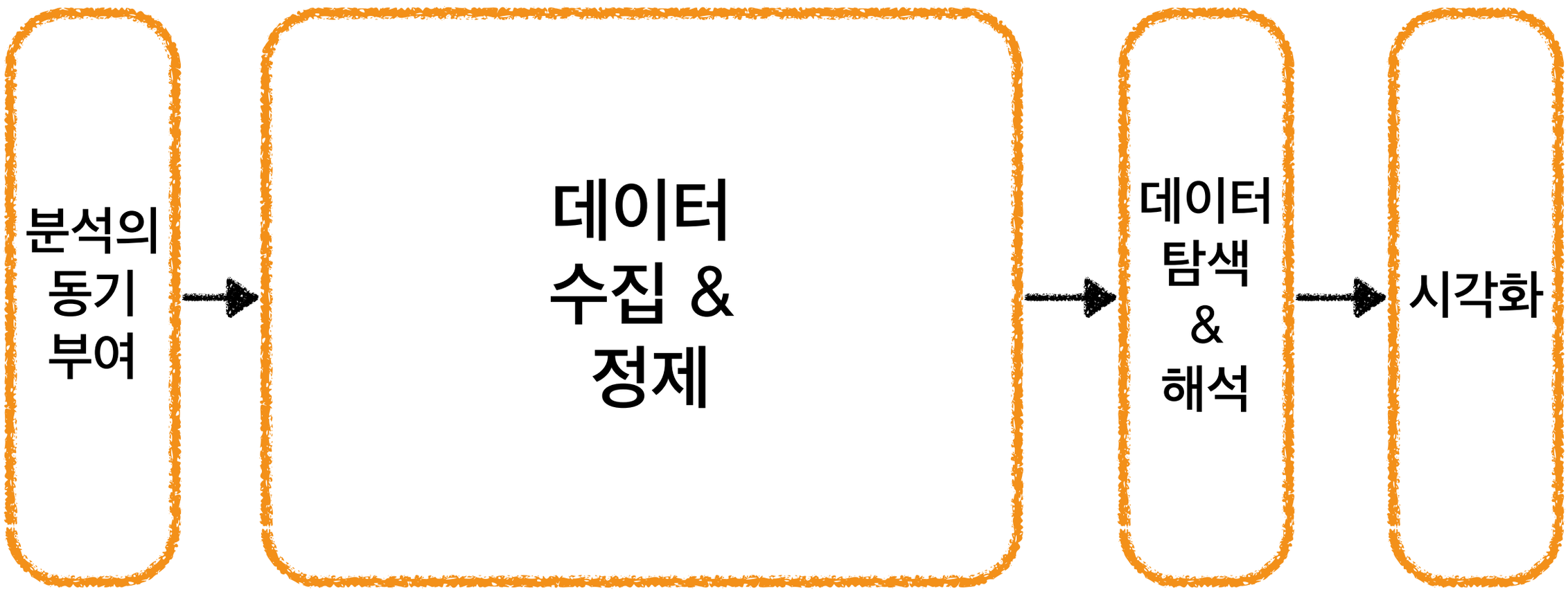
```
예측 나이 [ 59.  56.  56.  56.  56.  51.  51.  50.  46.  46.  45.  45.  45.  44.  44.
  44.  44.  43.  43.  43.]
```

그럼 이제
어떻게 데이터를 팠는지
알아 보겠습니다.

전체의 과정



소비되는 시간



분석의 동기 부여

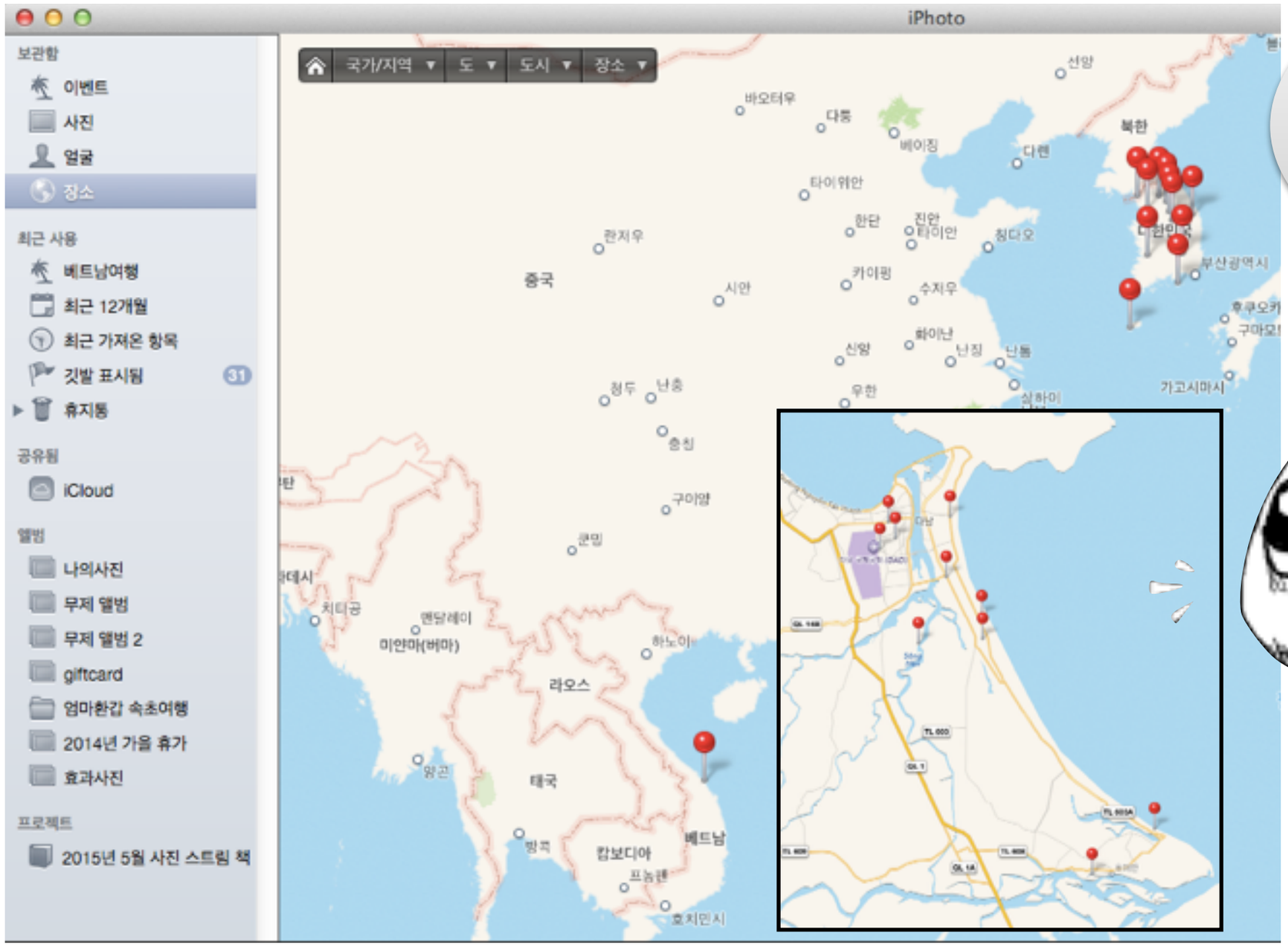
동기 부여 → 수집 & 정제 → 탐색&해석 → 시각화

저의 사진분석을 하기된 계기는?

Mac IPhoto로
장소별 사진 보기 기능 봤을때!!

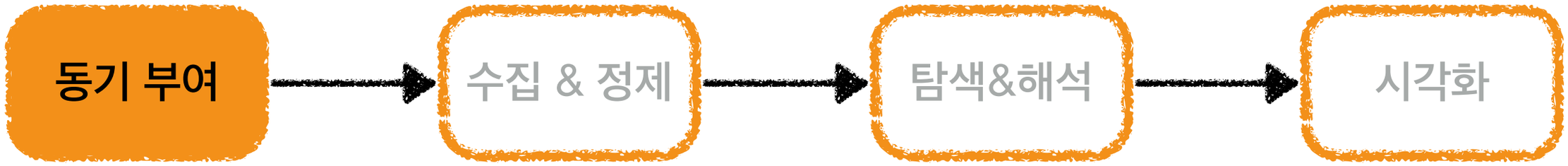
동기 부여 → 수집 & 정제 → 탐색&해석 → 시각화

IPhoto의 장소탭을 여는 순간!!



내가 지금까지
어디 갔는지
여기에 다나오네

쩐...짠다!



이런 데이터를 보고
분석을 안 할 수야 없지!!

동기 부여 → 수집 & 정제 → 탐색&해석 → 시각화

최규민

brunch.co.kr/@goodvc78

글쓰기

내 브런치

작가의 서랍

통계

브런치 홈

브런치 나우

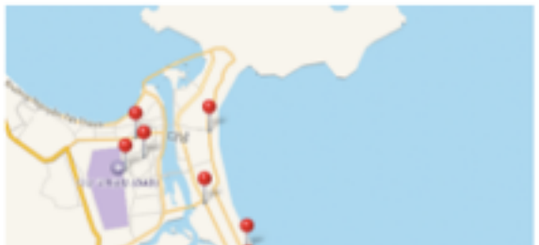
피드

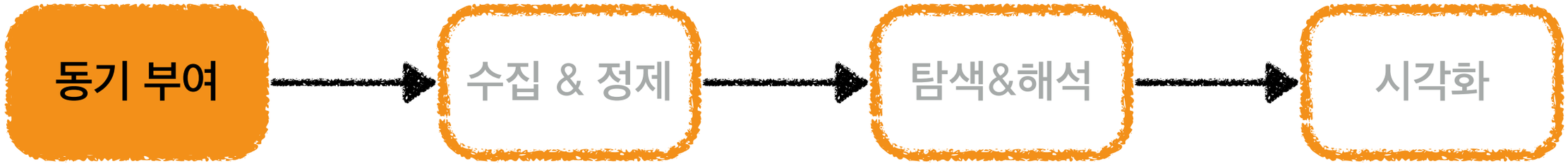
매거진

설정 로그아웃

사진이란 중요한 순간을 기록하기 위한 좋은 수단이다. 특별한 여행 중이라면 순간순간의 소중함, 자랑질, 찌질함을 남기기 위해 더욱더 많이 찍게 된다. 그래서 이번에 Me, Wife, Junior와 함께 한 4박 6일 베트남 다낭 여행에서 아이폰으로 찍은 259장 사진에 담긴 소중한 기록으로 나의 베트남 다낭 가족 여행을 재구성해 보고자 한다. (다른 전용 카메라로 찍은 사진도 많았지만 위치정보가 없어 IPhone의 사진만을 대상으로 하였다.)

사실 이렇게 분석해 보고자 했던 개시(??)는 맥북으로 여행 사진을 옮기면서 iPhoto의 "장소" 탭을 눌렀을 때 사진이 찍힌 장소가 베트남이 나타나면서 이거다!! 사진으로 나의 여행 경로를 자동으로 그려줄 수 있겠네! 하는 필이 팍! 들면서 이다. (그런데 여행 경로는 다 기억이 나는데! 왜 이것을 굳이 할까 생각을 해 봤지만 그냥 왠지 하고 싶었다. 그리고 잘했구나 하고 결론을 냈다.)





데이터에서 확인하고자 목표

- 내가 **주로 생활하는 지역**을 데이터로 확인하고
- 내가 **여행 다녀 온 곳의 타임라인**을 그려보자

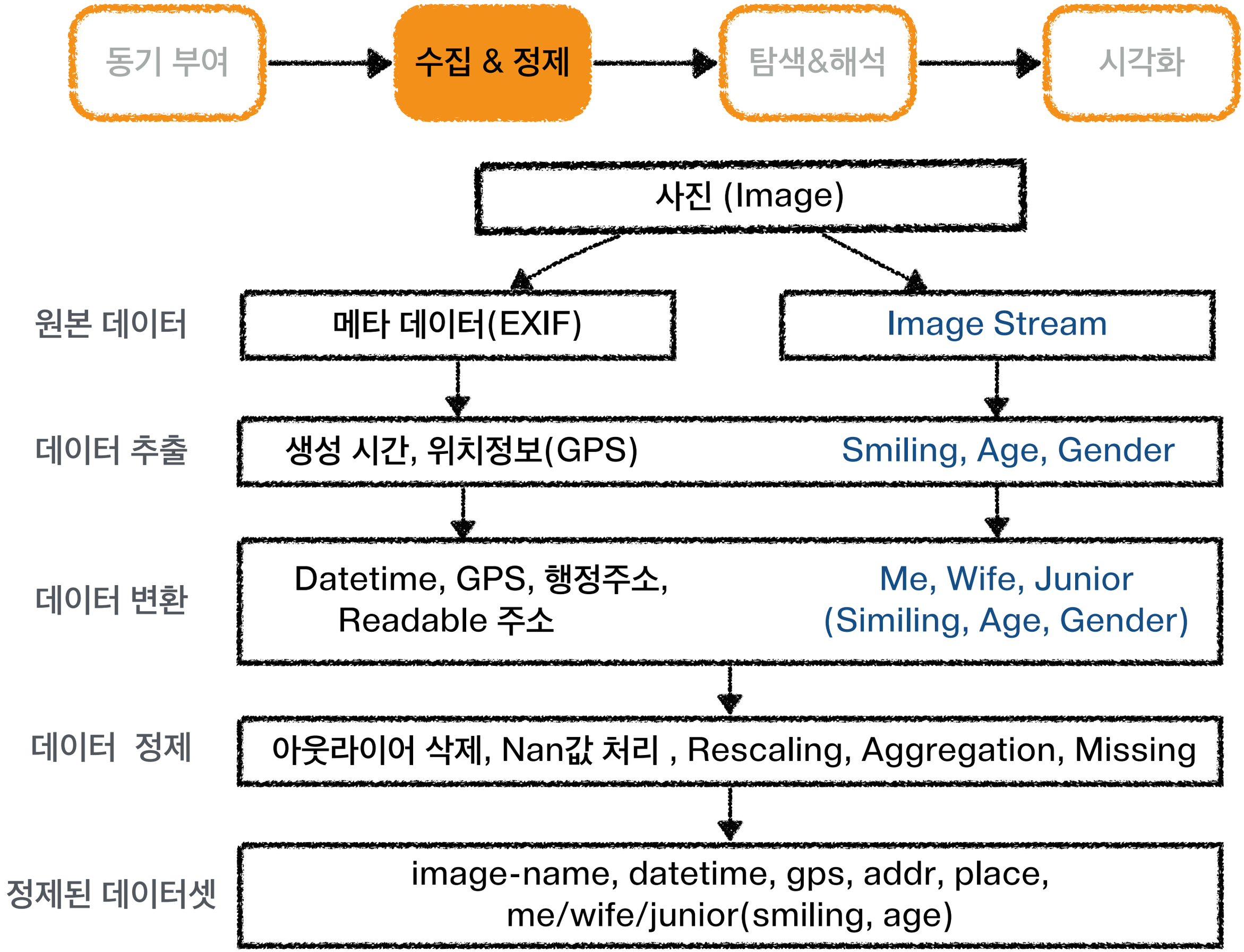
(이 타임라인을 여행루트로 추천해 주면 좋겠다)

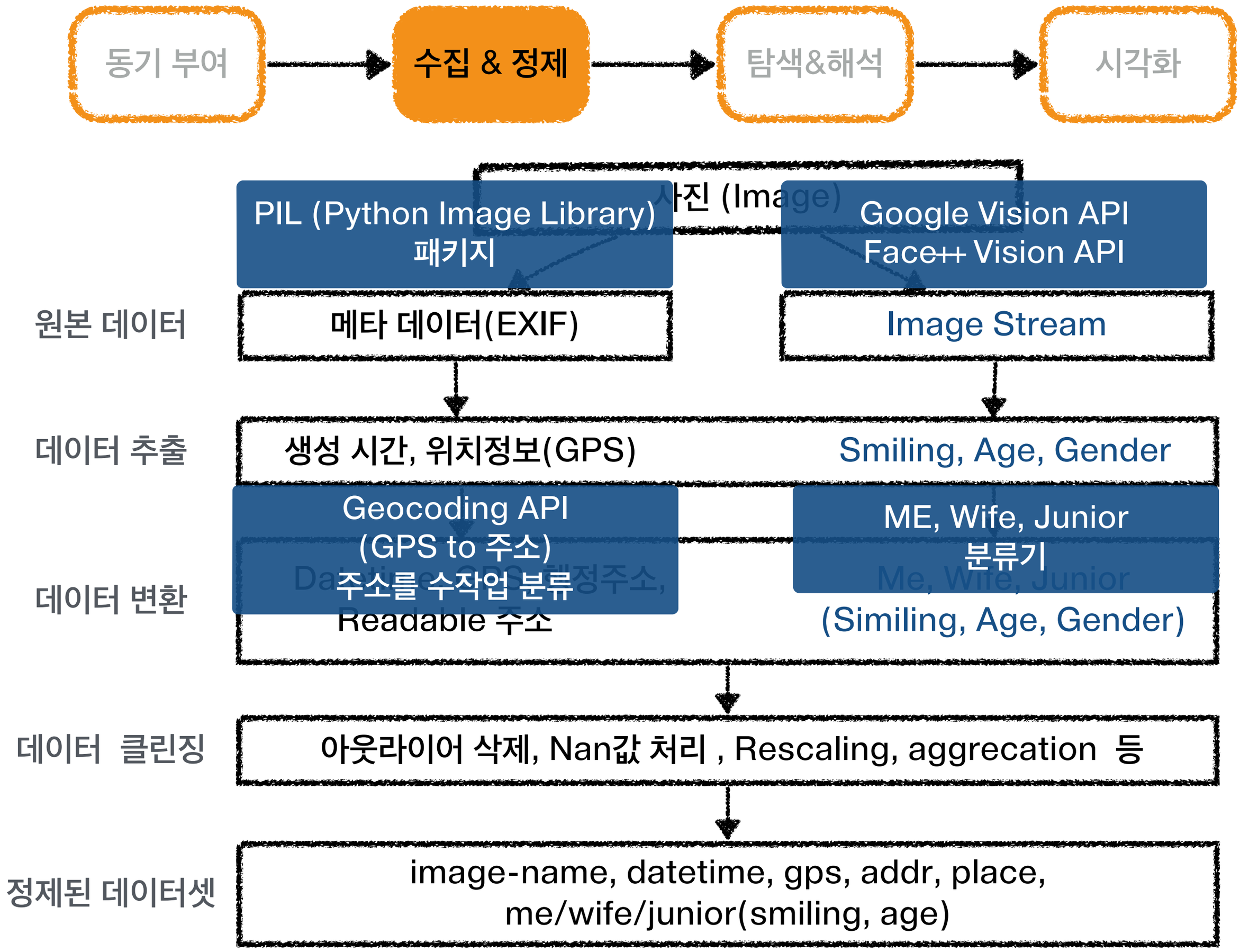
분석의
동기
부여
데이터
수집 &
정제
데이터
탐색 &
해석

동기 부여 → **수집 & 정제** → 탐색&해석 → 시각화

원자료(사진)에서 탐색에 용이한 테이블 데이터로 변환

	875	384	318	144
image_name	IMG_0958.JPG	IMG_0445.JPG	IMG_0379.JPG	IMG_0205.JPG
datetime	2012-03-23 18:53:24	2013-05-04 12:10:05	2014-01-10 18:41:54	2015-07-19 11:03:31
latitude	37.40967	37.51267	37.40134	37.27649
longitude	127.1343	127.6368	127.1015	127.065
model	iPhone 4S	iPhone 4S	iPhone 4S	iPhone 4S
resolution	480X640	2448X3264	2448X3264	2448X3264
year	2012	2013	2014	2015
year_day	83.78681	124.5069	10.77847	200.4604
date	2012-03-23	2013-05-04	2014-01-10	2015-07-19
addr	경기도_성남시_분당구_야탑3동	KR_경기도_양평군_용문면	경기도_성남시_분당구_삼평동	경기도_수원시_영통구_하동
place	야탑_이전집근처	양평_캠핑장	판교_회사	수원_호수공원
class	집근처	캠핑	아프리카TV	나들이
person	2	1	0	0
me_smiling	7.18004	34.7251	0	0
me_age	34	29	0	0
wife_smiling	43.4948	0	0	0
wife_age	24	0	0	0
junior_smiling	0	0	0	0
junior_age	0	0	0	0


동기 부여
수집 & 정제
탐색&해석
시각화
사진 (Image)
원본 데이터
메타 데이터(EXIF)
Image Stream
데이터 추출
생성 시간, 위치정보(GPS)
Smiling, Age, Gender
데이터 변환
Datetime, GPS, 행정주소,
Readable 주소
Me, Wife, Junior
(Similing, Age, Gender)
데이터 정제
아웃라이어 삭제, Nan값 처리 , Rescaling, Aggregation, Missing
정제된 데이터셋
image-name, datetime, gps, addr, place,
me/wife/junior(smiling, age)


동기 부여
수집 & 정제
탐색&해석
시각화
사진 (Image)
PIL (Python Image Library) 패키지
Google Vision API
Face++ Vision API
원본 데이터
메타 데이터(EXIF)
Image Stream
데이터 추출
생성 시간, 위치정보(GPS)
Smiling, Age, Gender
데이터 변환
Geocoding API (GPS to 주소)
주소를 수작업 분류
Datetime, GPS, 행정주소, Readable 주소
ME, Wife, Junior 분류기
Me, Wife, Junior (Similing, Age, Gender)
데이터 클린징
아웃라이어 삭제, Nan값 처리 , Rescaling, aggrecation 등
정제된 데이터셋
image-name, datetime, gps, addr, place, me/wife/junior(smiling, age)

동기 부여 → **수집 & 정제** → 탐색&해석 → 시각화

```
pic_ds[::200]
```

	image_name	datetime	latitude	longitude	model	resolution	year	year_day	date	addr	place	class	person
0	IMG_0001.JPG	2016-02-05 15:29:17	0.000000	0.000000	unknown	960X720	2016	36.645139	2016-02-05	—			0
200	IMG_0261.JPG	2014-05-03 21:19:26	37.332272	126.828042	iPhone 4S	2448X3264	2014	123.888194	2014-05-03	경기도_안산시_단원구_와동	안산_	친구모임	0
400	IMG_0461.JPG	2013-04-21 18:44:22	37.128167	129.046000	iPhone 4S	2448X3264	2013	111.780556	2013-04-21	KR_강원도_태백시_철암동	태백_본가	본가	0
600	IMG_0668.JPG	2012-09-22 17:43:00	37.231333	127.542833	iPhone 4S	2448X3264	2012	266.738194	2012-09-22	KR_경기도_여주군_가남면	여주_캠핑장	캠핑	0
800	IMG_0882.JPG	2016-02-05 15:29:17	0.000000	0.000000	unknown	1377X1032	2016	36.645139	2016-02-05	—			1
1000	IMG_1089.JPG	2016-02-10 18:58:56	37.341367	127.097794	iPhone 6s	1932X2576	2016	41.790278	2016-02-10	경기도_용인시_수지구_동천동	수지_집근처	집근처	3
1200	IMG_1424.JPG	2016-04-19 10:56:46	37.455672	126.446869	iPhone 6s	3024X4032	2016	110.455556	2016-04-19	KR_인천광역시_중구_운서동	인천_공항	정기휴가	0

동기 부여 → 수집 & 정제 → **탐색&해석** → 시각화

Pandas의 describe()로 기본 Stat 보기

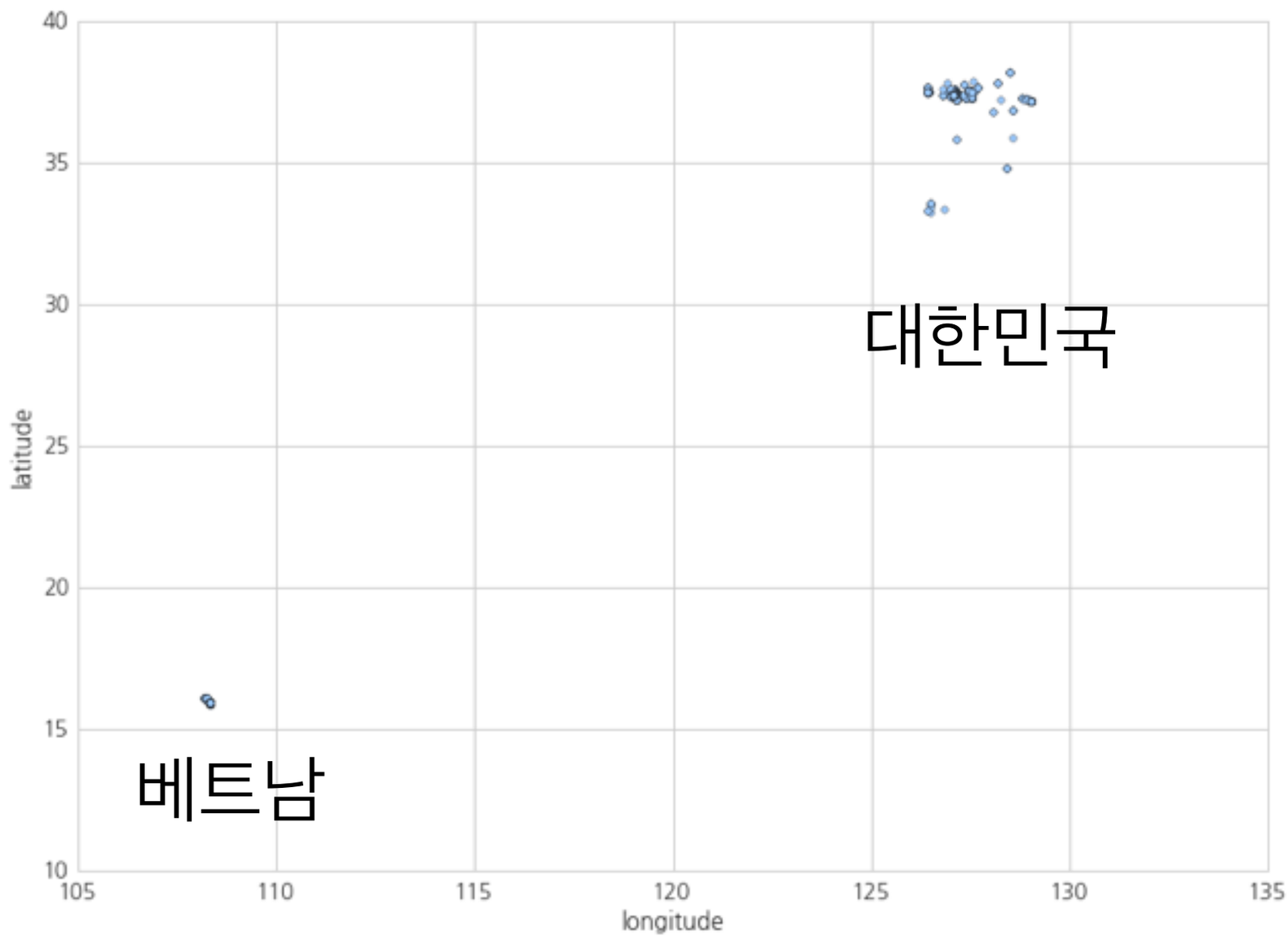
```
pic_ds.describe().T
```

	count	mean	std	min	25%	50%	75%	max
latitude	1576	25.425478	15.949763	0.000000	15.876232	37.338508	37.410667	38.176833
longitude	1576	94.366223	53.049088	0.000000	108.245649	127.097625	127.137333	129.050400
year	1576	2014.853426	1.458231	2012.000000	2013.000000	2016.000000	2016.000000	2016.000000
year_day	1576	133.038906	89.833956	3.050694	47.577778	112.777083	192.809722	362.789583
person	1576	0.720812	1.045977	0.000000	0.000000	0.000000	1.000000	11.000000
me_smiling	1576	5.792091	18.592427	0.000000	0.000000	0.000000	0.000000	98.917100
me_age	1576	4.246827	11.374909	0.000000	0.000000	0.000000	0.000000	62.000000
wife_smiling	1576	12.378644	27.947059	0.000000	0.000000	0.000000	0.000000	99.228000
wife_age	1576	5.555203	11.474560	0.000000	0.000000	0.000000	0.000000	56.000000
junior_smiling	1576	5.128215	15.712543	0.000000	0.000000	0.000000	0.000000	94.713100
junior_age	1576	0.689721	1.526751	0.000000	0.000000	0.000000	0.000000	9.000000

동기 부여 → 수집 & 정제 → **탐색&해석** → 시각화

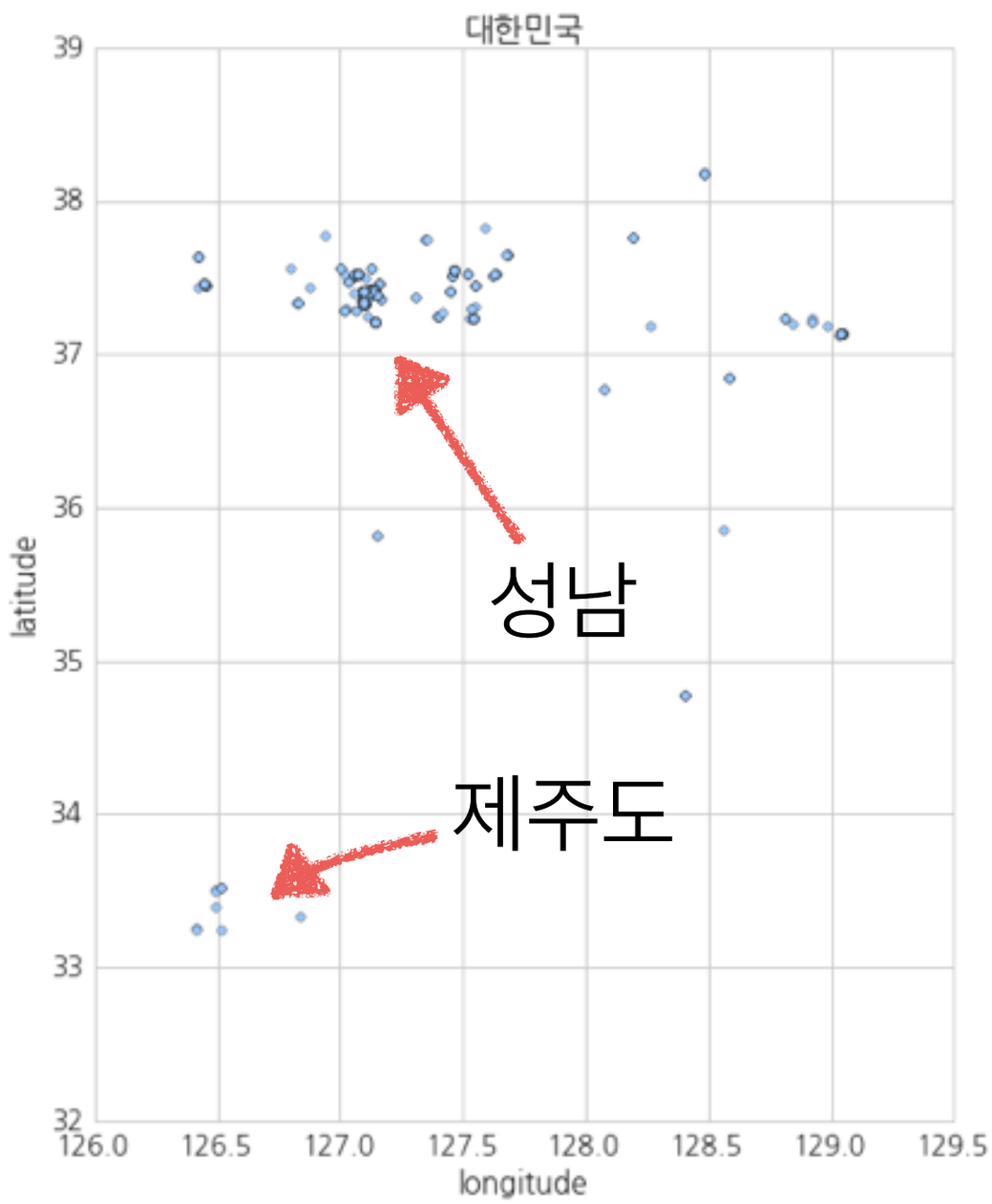
지리 좌표를 Scatter Plotting하기 (Matplotlib)

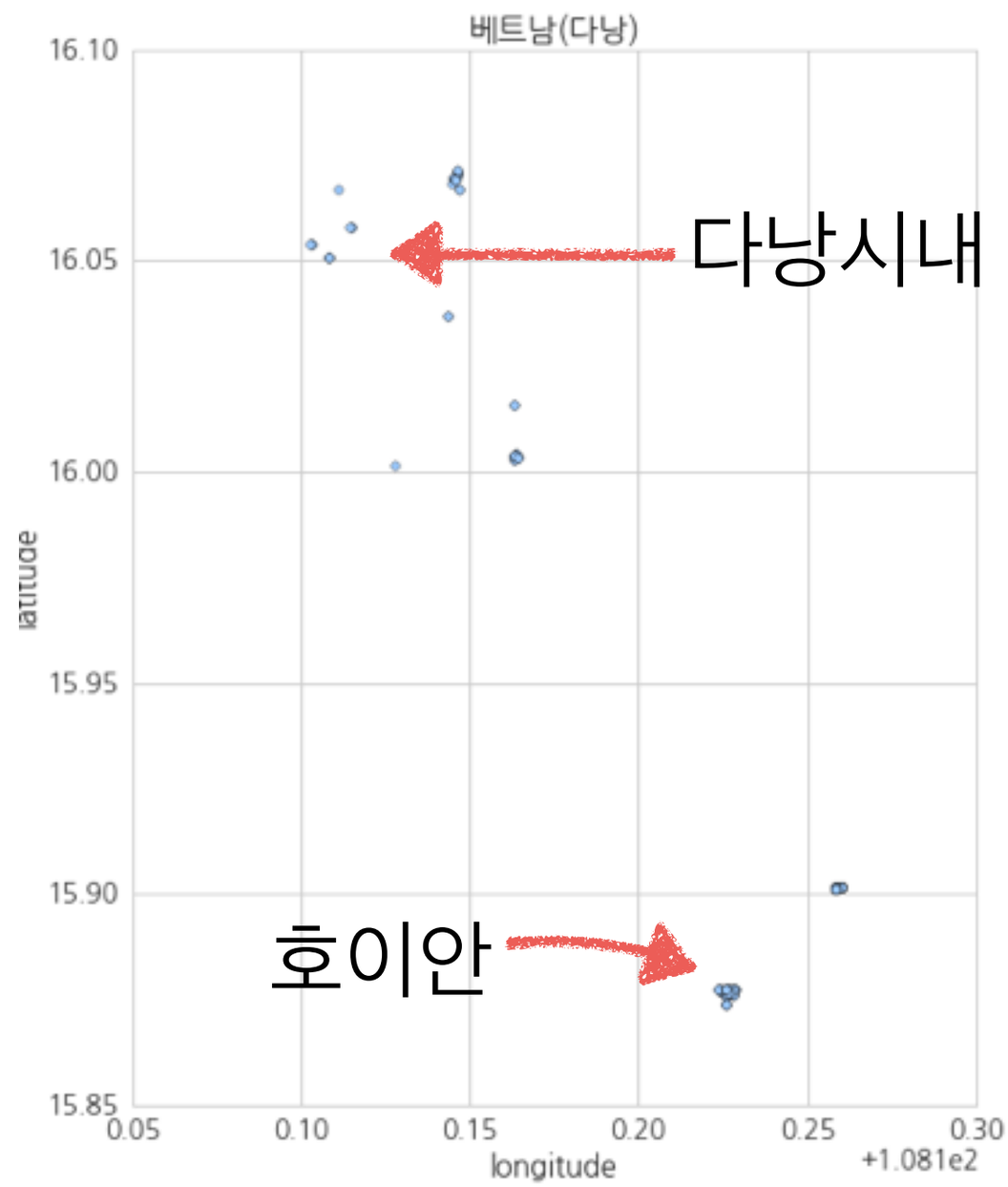
```
ax = pic_ds[pic_ds.latitude>0].plot(kind='scatter', y='latitude',x='longitude', figsize=(14,10))
```



동기 부여 → 수집 & 정제 → **탐색&해석** → 시각화

지리 좌표를 Scatter Plotting하기 (Matplotlib)

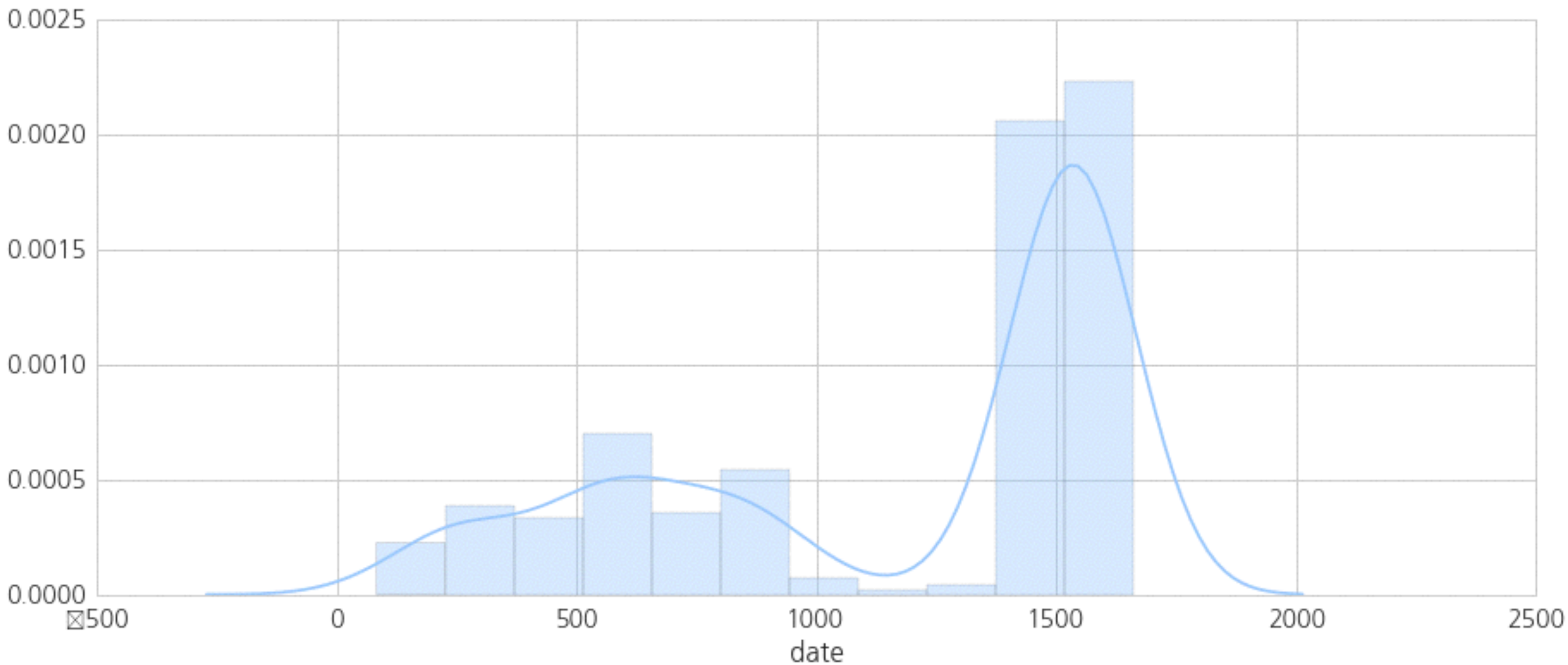


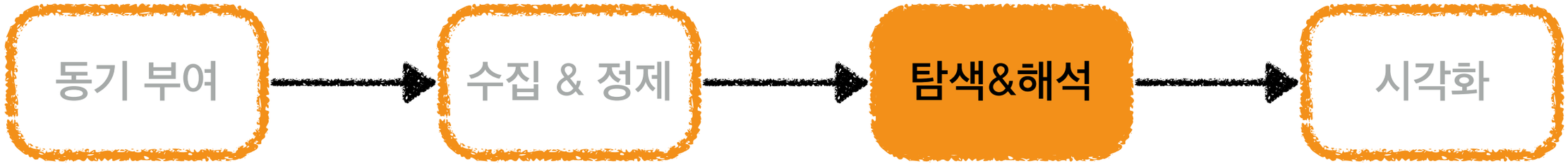


동기 부여 → 수집 & 정제 → **탐색&해석** → 시각화

찍은 사진의 날짜별 분포 그려보기 (Seaborn 패키지)

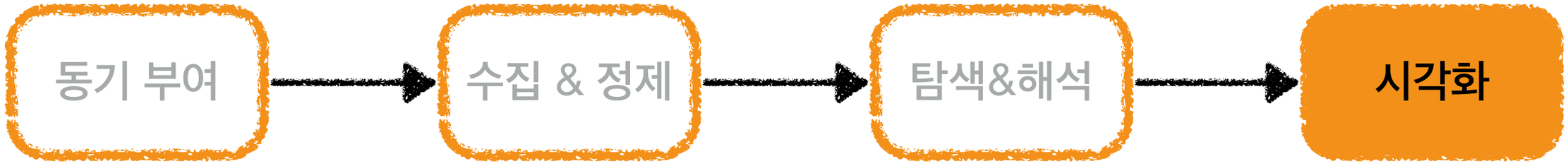
```
plt.figure(figsize=(15,6))
ax = sns.distplot(pic_ds.date.apply(lambda x: (x-datetime.date(2012,1,1)).days))
```





찍은 사진의 날짜별 분포 그려보기 (Seaborn 패키지)

```
plt.figure(figsize=(15,6))
ax = sns.distplot(pic_ds.date.apply(lambda x: (x-datetime.date(2012,1,1)).days))
```



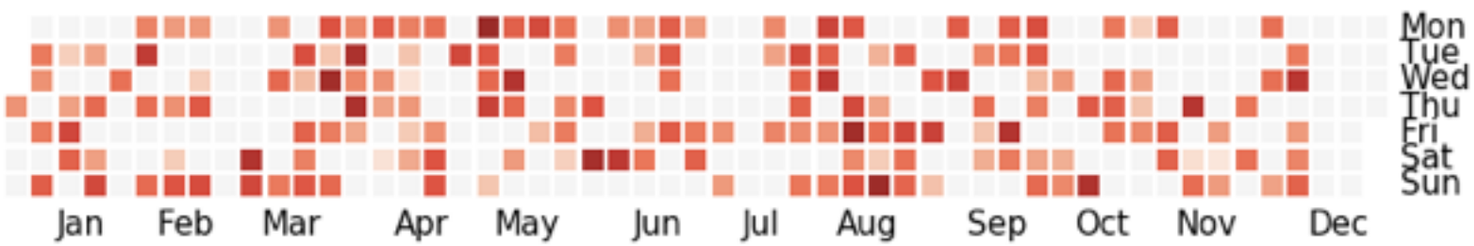
데이터의 시각화는 데이터의 속성에 맞는 표현 방식을 선택

동기 부여 → 수집 & 정제 → 탐색&해석 → **시각화**

사진을 찍은 1-D(시간) 데이터의 표현

Calendar heatmaps from Pandas time series data

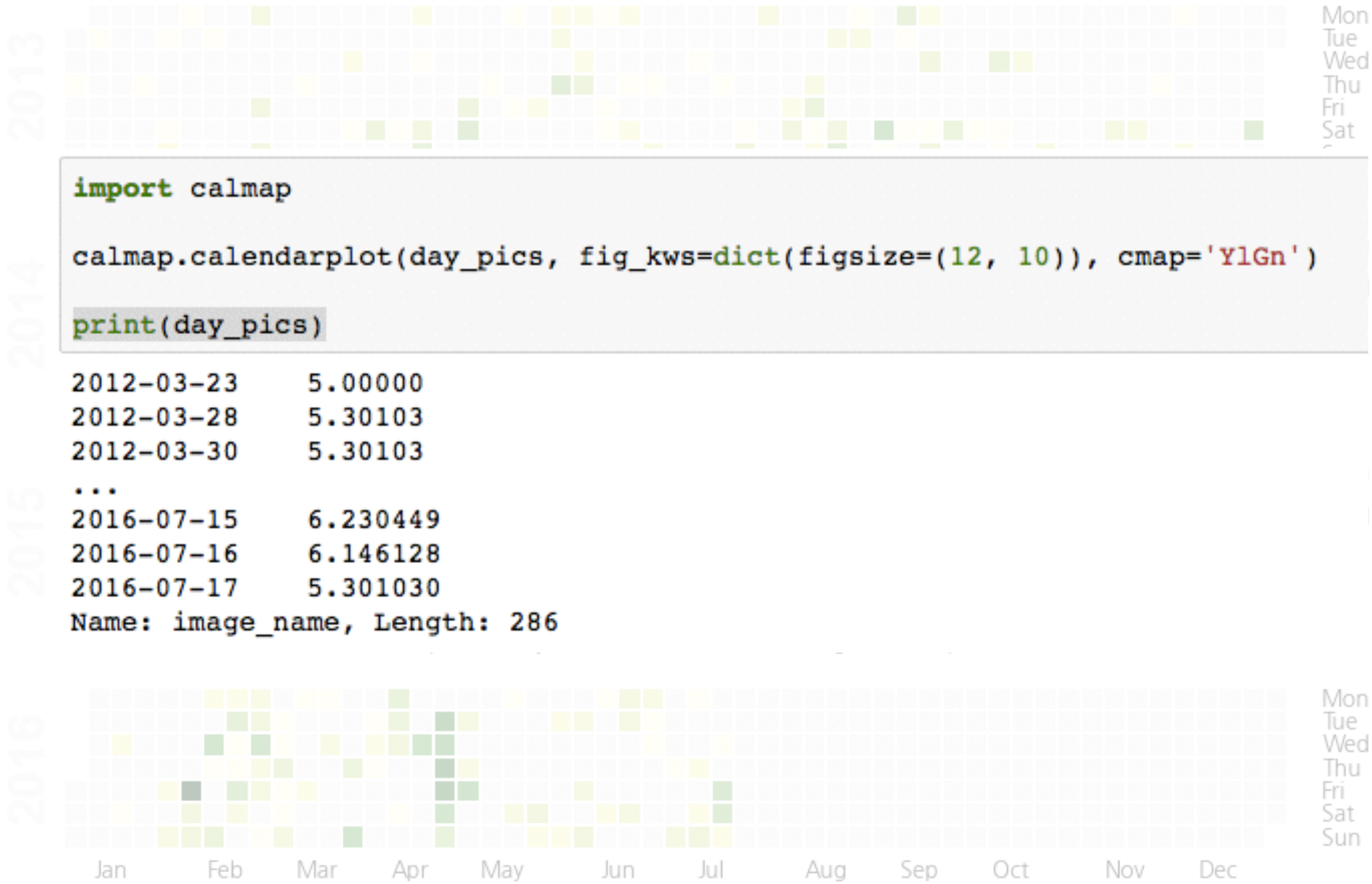
Plot Pandas time series data sampled by day in a heatmap per calendar year, similar to GitHub's con
using matplotlib.



http://pythonhosted.org/calmap/

	pic count
date	
2012-03-23	1
2012-03-28	2
2012-03-30	2
2012-04-07	1
2012-04-10	1
...	...
2016-07-10	5
2016-07-13	1
2016-07-15	17
2016-07-16	14
2016-07-17	2

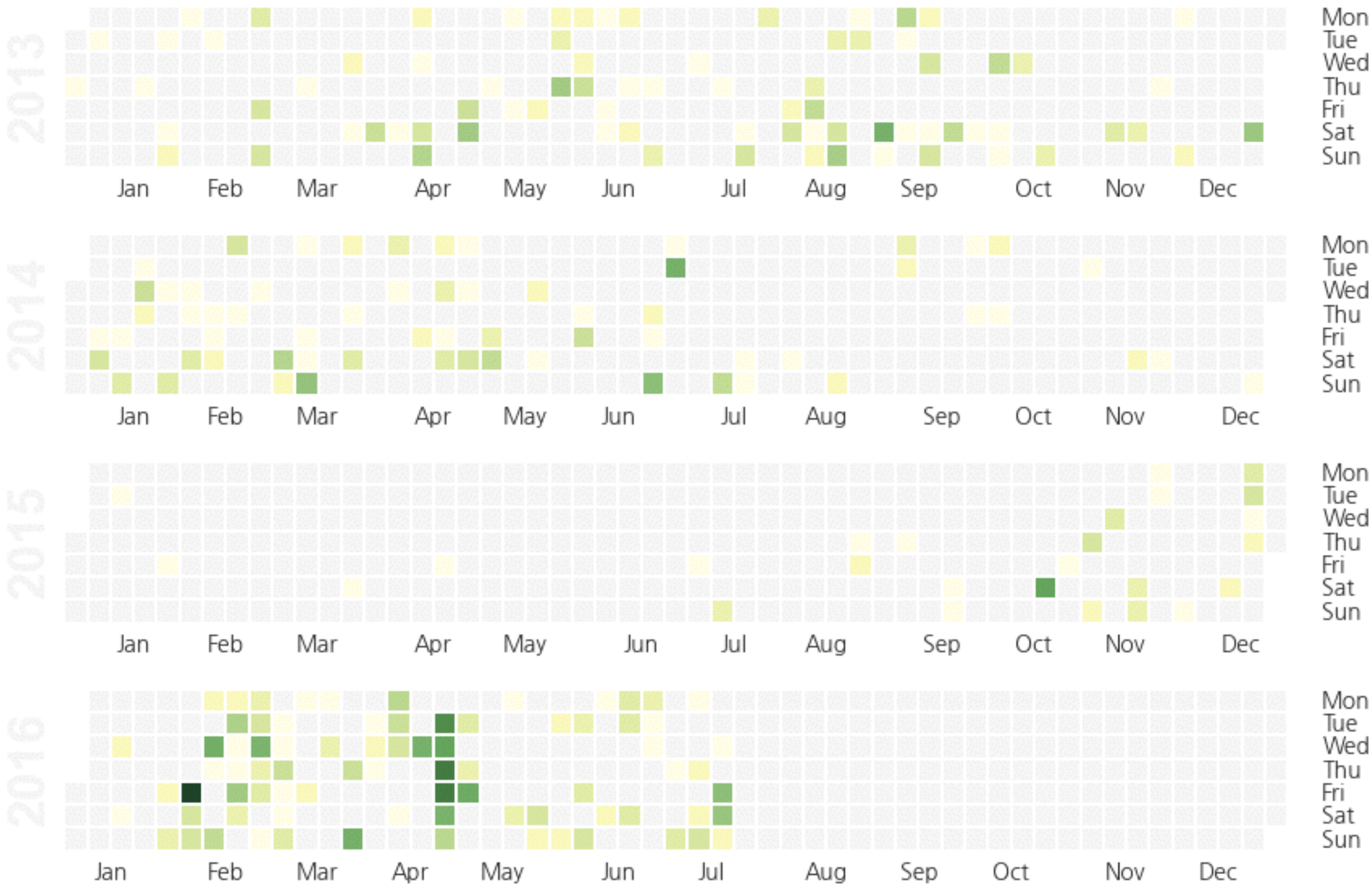
동기 부여 → 수집 & 정제 → 탐색&해석 → **시각화**




```python
import calmap

calmap.calendarplot(day_pics, fig_kws=dict(figsize=(12, 10)), cmap='YlGn')

print(day_pics)
```

```
2012-03-23    5.00000
2012-03-28    5.30103
2012-03-30    5.30103
...
2016-07-15    6.230449
2016-07-16    6.146128
2016-07-17    5.301030
Name: image_name, Length: 286
```

동기 부여 → 수집 & 정제 → 탐색&해석 → **시각화**



동기 부여 → 수집 & 정제 → 탐색&해석 → **시각화**

사진 2-D(지리좌표) 데이터 표현

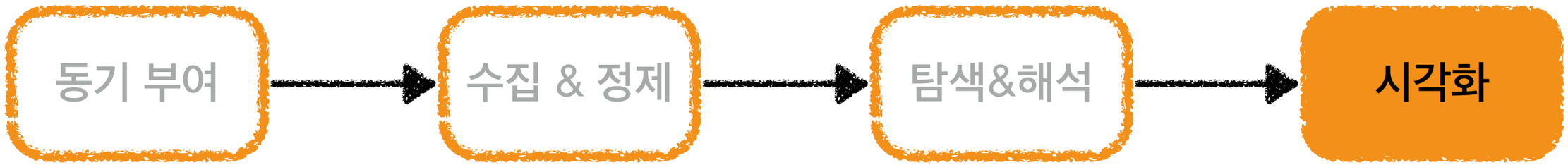
ipyleaflet

https://github.com/ellisonbg/ipyleaflet

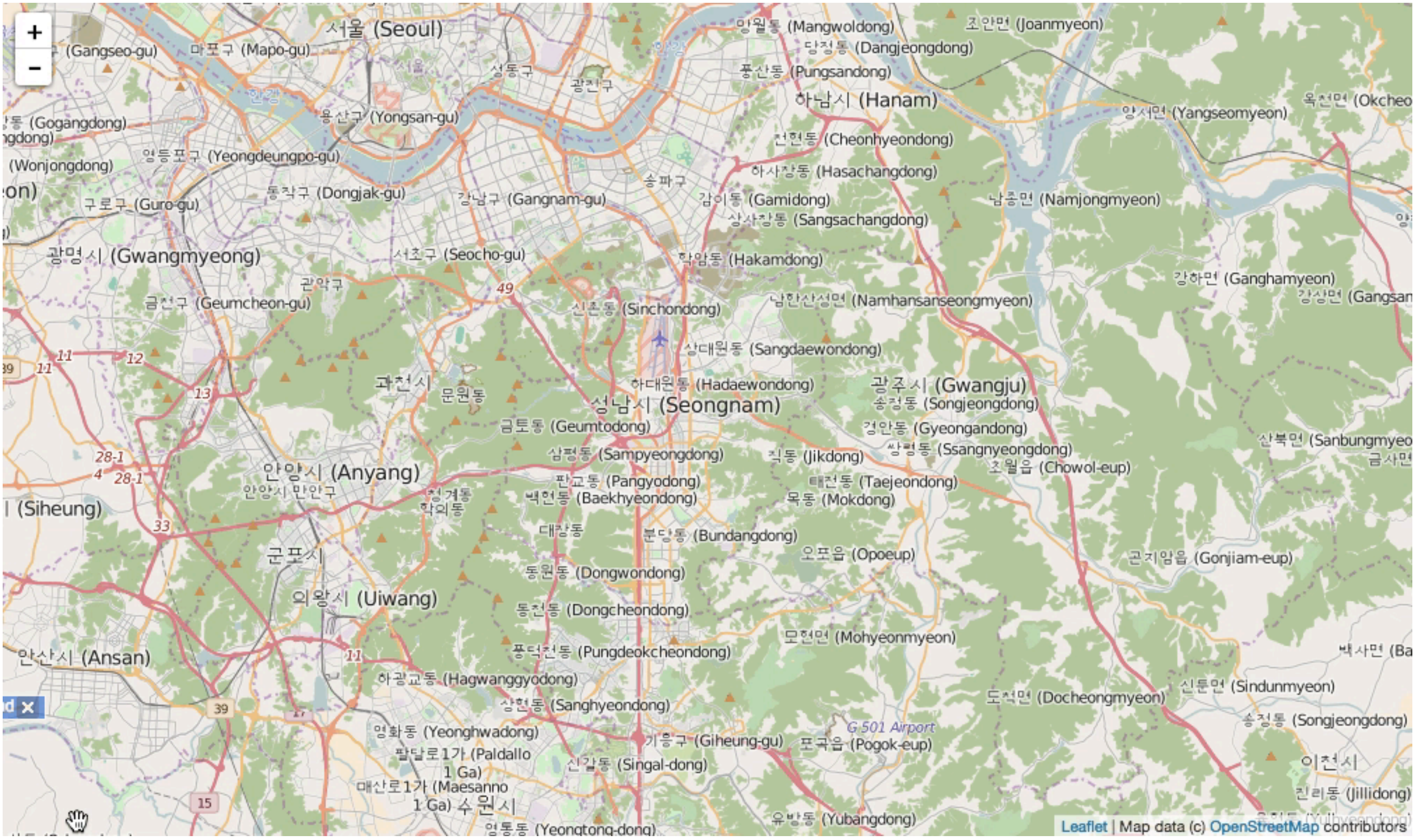
```
from ipyleaflet import Map

Map(center=[34.6252978589571, -77.34580993652344], zoom=10)
```

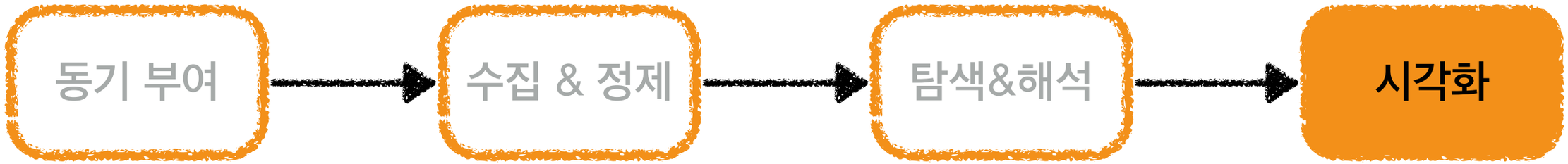
Java Script 기반으로 Jupyter Notebook에서

Interactive UX 제공

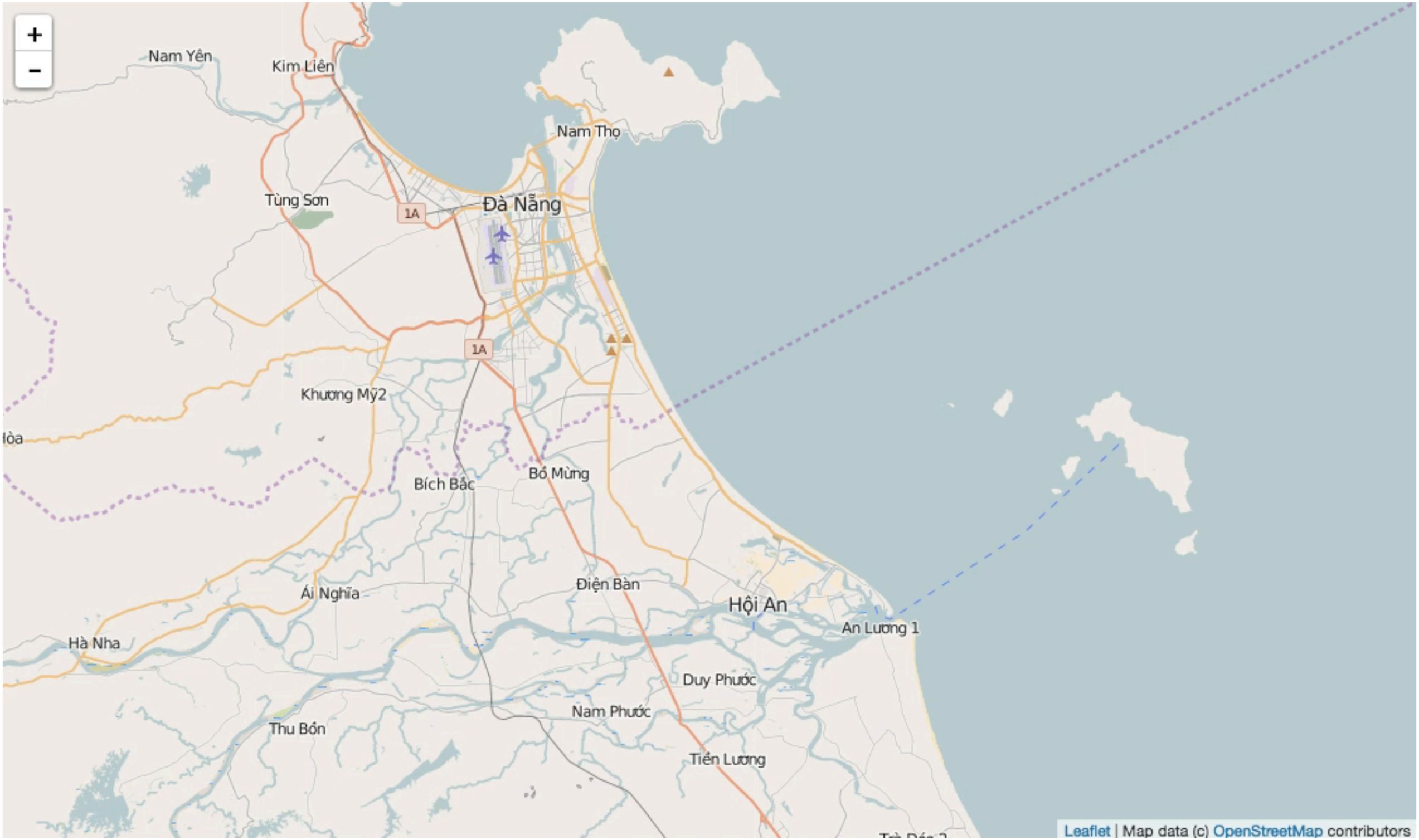


My Picture
2012’ to 2016’

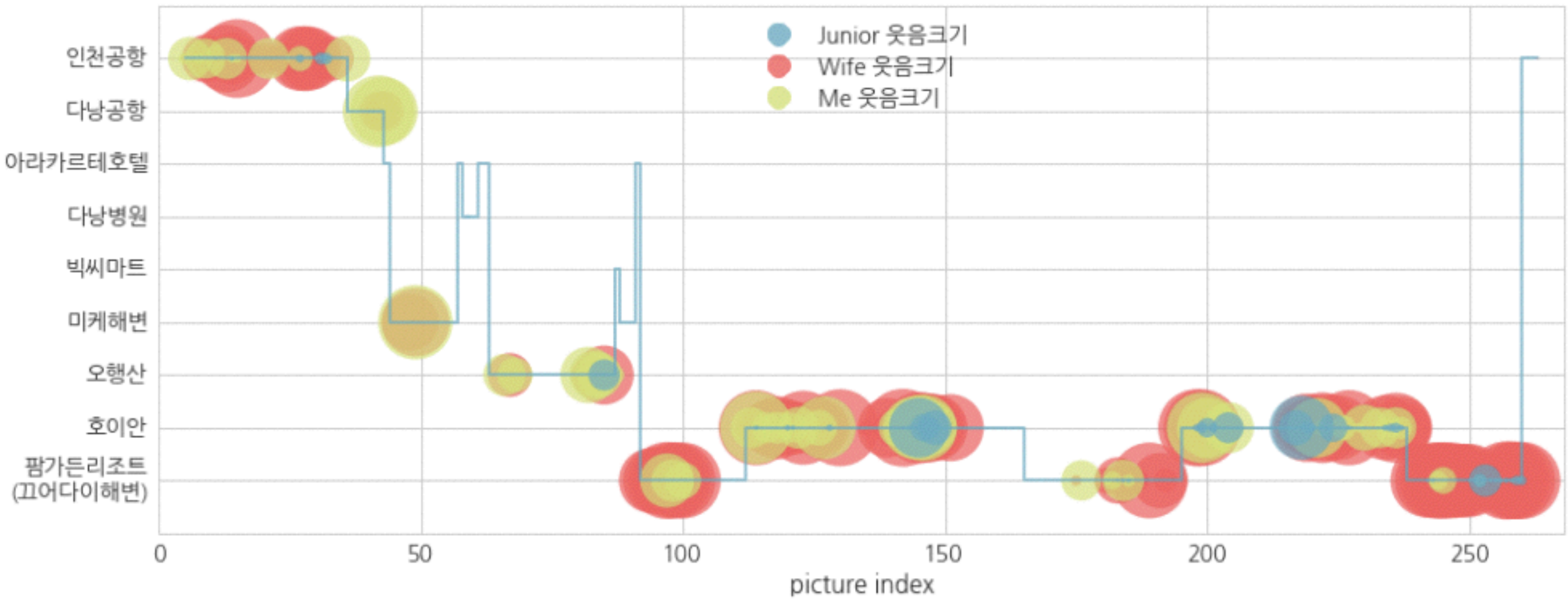

서울 (Seoul)
마포구 (Mapo-gu)
(Gangseo-gu)
성동구
광진구
용산구 (Yongsan-gu)
영등포구 (Yeongdeungpo-gu)
동작구 (Dongjak-gu)
구로구 (Guro-gu)
강남구 (Gangnam-gu)
송파구
광명시 (Gwangmyeong)
서초구 (Seocho-gu)
관악구
금천구 (Geumcheon-gu)
과천시
문원동
성남시 (Seongnam)
하대원동 (Hadaewondong)
상대원동 (Sangdaewondong)
신촌동 (Sinchondong)
금토동 (Geumtodong)
삼평동 (Sampyeongdong)
판교동 (Pangyodong)
백현동 (Baekhyeondong)
안양시 (Anyang)
안양시 만안구
청계동
학의동
군포시
의왕시 (Uiwang)
안산시 (Ansan)
하광교동 (Hagwanggyodong)
상현동 (Sanghyeondong)
영화동 (Yeonghwadong)
팔달로1가 (Paldallo 1 Ga)
매산로1가 (Maesanno 1 Ga)
수원시
영통동 (Yeongtong-dong)
기흥구 (Giheung-gu)
신갈동 (Singal-dong)
대창동
분당동 (Bundangdong)
동원동 (Dongwondong)
동천동 (Dongcheondong)
풍덕천동 (Pungdeokcheondong)
(Siheung)
(Gogangdong)
(Wonjongdong)
망월동 (Mangwoldong)
당정동 (Dangjeongdong)
풍산동 (Pungsandong)
하남시 (Hanam)
조안면 (Joanmyeon)
양서면 (Yangseomyeon)
옥천면 (Okcheo
천현동 (Cheonhyeondong)
하사창동 (Hasachangdong)
감이동 (Gamidong)
상사창동 (Sangsachangdong)
남종면 (Namjongmyeon)
학암동 (Hakamdong)
강하면 (Ganghamyeon)
남한산성면 (Namhansanseongmyeon)
광주시 (Gwangju)
송정동 (Songjeongdong)
경안동 (Gyeongandong)
쌍령동 (Ssangnyeongdong)
직동 (Jikdong)
조월읍 (Chowol-eup)
태전동 (Taejeondong)
목동 (Mokdong)
산북면 (Sanbungmyeo
금사면
오포읍 (Opoeup)
곤지암읍 (Gonjiam-eup)
모현면 (Mohyeonmyeon)
백사면 (Ba
신둔면 (Sindunmyeon)
도척면 (Docheongmyeon)
송정동 (Songjeongdong)
G 501 Airport
포곡읍 (Pogok-eup)
이천시
진리동 (Jillidong)
유방동 (Yubangdong)
Leaflet | Map data (c) OpenStreetMap contributors



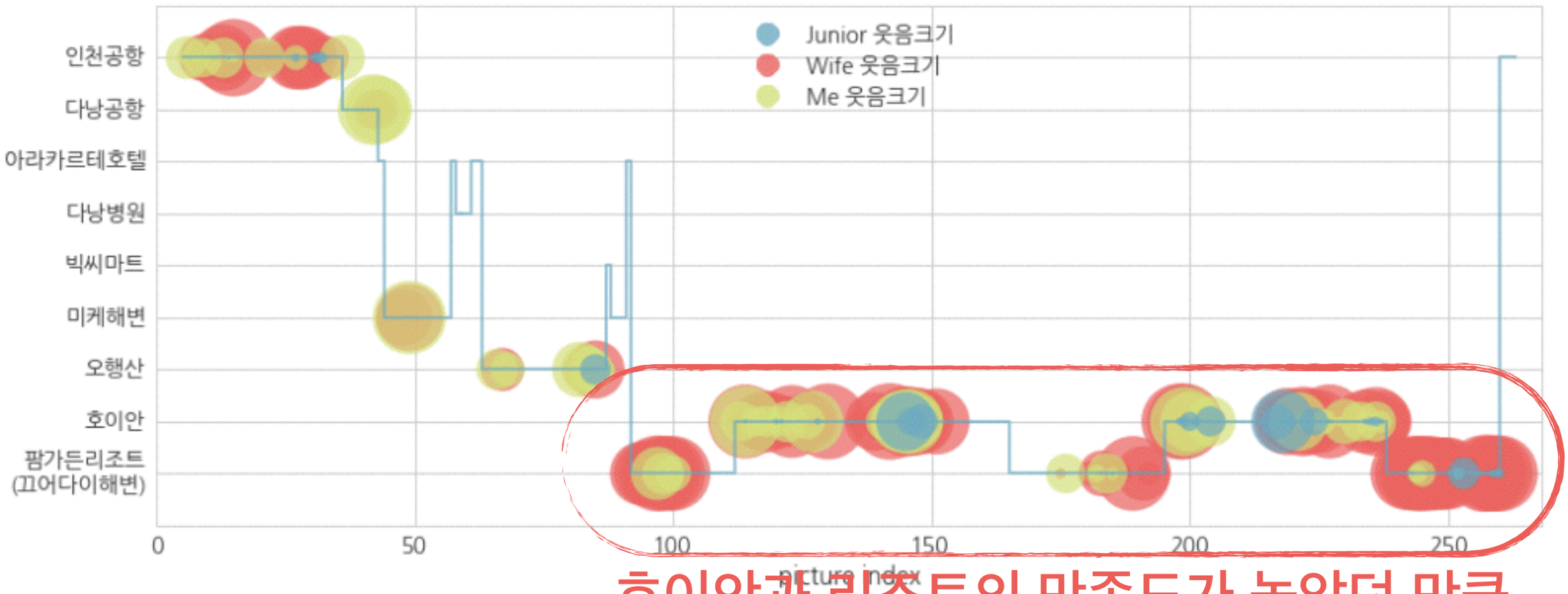
2016 My 베트남 여행


Nam Yên
Kim Liên
Nam Thọ
Tùng Sơn
Đà Nẵng
1A
1A
Khương Mỹ2
Bích Bắc
Bồ Mừng
Điện Bàn
Ái Nghĩa
Hội An
An Lương 1
Hà Nha
Duy Phước
Nam Phước
Thu Bồn
Tiền Lương
Leaflet | Map data (c) OpenStreetMap contributors

언제 가장 많이 웃었는가?


Junior 웃음크기
Wife 웃음크기
Me 웃음크기
인천공항
다낭공항
아라카르테호텔
다낭병원
빅씨마트
미케해변
오행산
호이안
팜가든리조트
(끄어다이해변)
0
50
100
150
200
250
picture index

여행 루트별 스마일지수



호이안과 리조트의 만족도가 높았던 만큼
잘웃고 사진도 많이 찍은듯 한다

마시면서

데이터 광부 모십니다!

문의 : goodvc@afreecatv.com

afreecaTV

감사합니다.

goodvc78@gmail.com