R을 이용한 **빅데이터 사례 1**

EXAMPLE1.R

날씨를 고려한 제주도 관광지 추천 알고리즘 개발

Development of Recommendation Algorithm for Attractions Considering Weather

김유림 /YURIM KIM (제1) [1], 임정현 /LIM JUNGHYUN (참여) [2], 이예지 /LEEYEJI (참여) [3], 윤상후 (교신) [4]

[1] 대구대학교
[2] 대구대학교
[3] 대구대학교
[4] 대구대학교

기상청

| 기상청 | 공모작제출 | 대회소식 |

2018 날씨 빅데이터 콘테스트

날씨 빅데이터, 우리의 내일을 제시하다.

대회기간
2018. 5. 21 ~ 2018. 8. 30

공모내용
기상기후 빅데이터를 활용한 데이터 분석(자유주제)

참가대상
대한민국 국민 누구나 (개인 또는 팀)
팀 구성 시 인원 제한은 없으며, 단체·법인명의 팀 참여 가능

Contents

DAEGU
UNIVERSITY

Background

- The weather in Korea peninsula has strong **locality property**, so that the weather might not be same even two different tourism sites in same city.

- Korea Meteorological Administrator are serving
  1) Tourism Climate index by city and district area.
  2) Weather forecasting at major tourism sites.

- However, it does not provide an enough information for travelers who wish to travel.

- We want to make an algorithm for tourist, which can help to decide the itinerary for a trip. The algorithm was consisted of **weather information and Social media.**
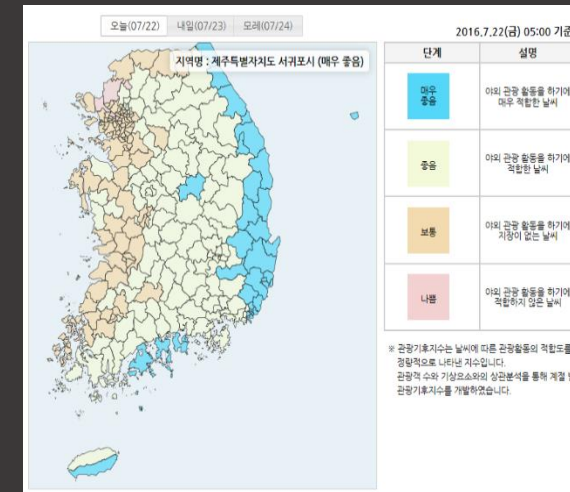
To do that we need to..

- Predict weather information at tourism sites

- Calculate tourism climate index

- Modify tourism climate index

    considering natural hazard circumstance

    considering property of tourism sites (indoor/ outdoor)

- Develop recommendation algorithm

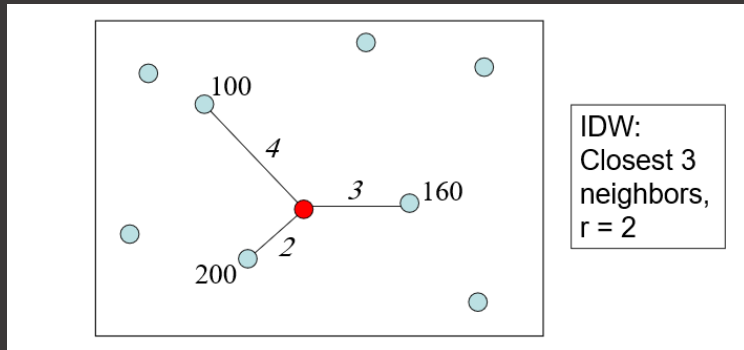   develop overall tourism score based on TCI & social media score
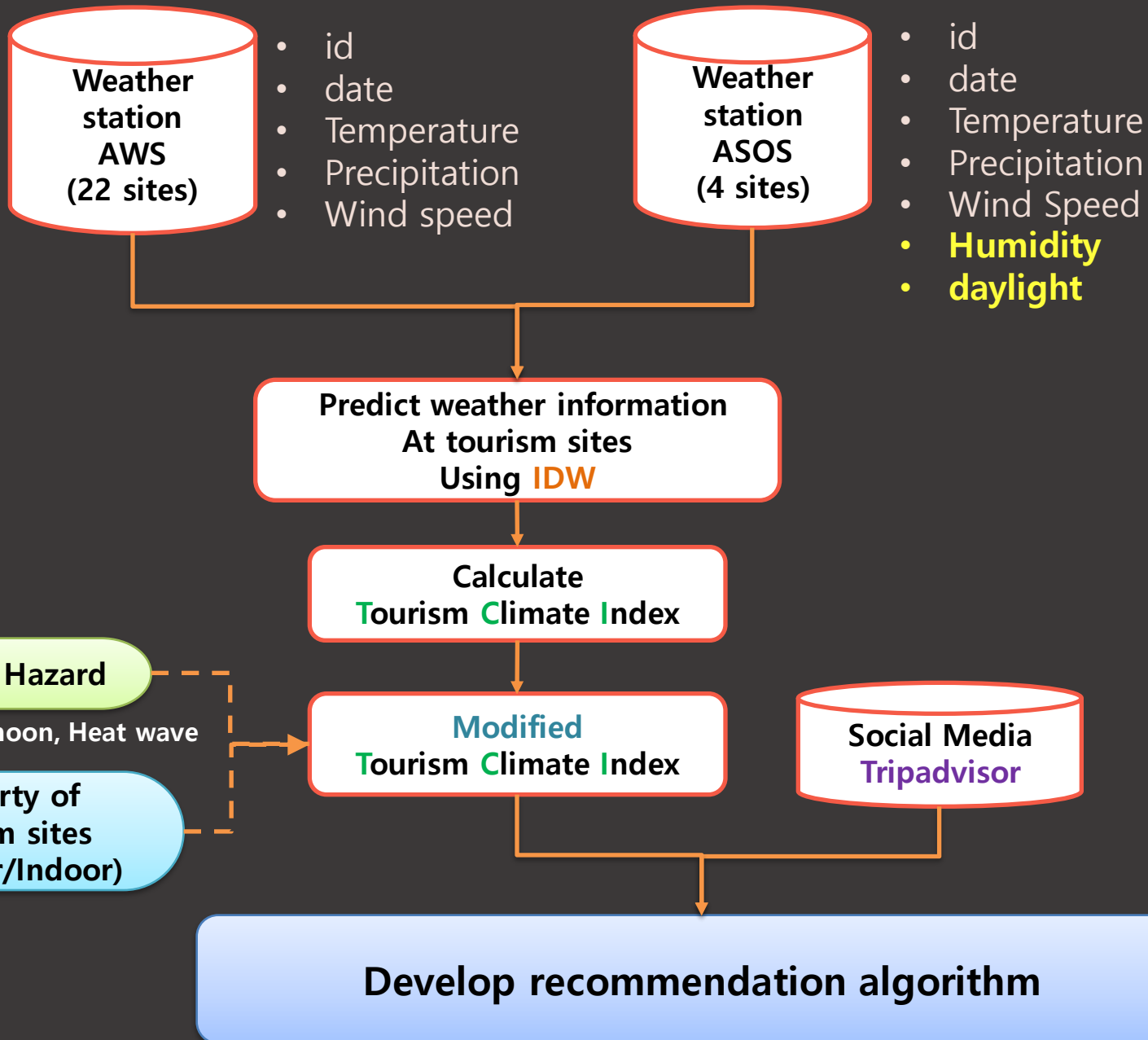
Methodology

Inverse Distance Weighted ; IDW

- IDW is an interpolation method which used in the field of geostatistics.

- The value of unobserved sites is depended on the inverse of distance from the known coordinates.
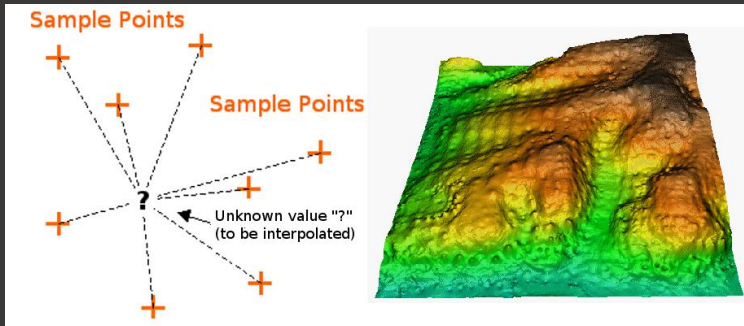
$$Z_p = \frac{\sum_{i=1}^{n} Z_i W_i}{\sum_{i=1}^{n} W_i} \quad W_i = \frac{1}{d_i^{\,p}}$$

Z_p : interpolation value of prediction point

Z_i : $location(x_i, y_i)$

W_i : $weight$

n : $number\ of\ reference\ values$

Methodology

Tourism Climate Index ; TCI (in case of Korea; Kim & Kim, 2014)

K-TCI formula (**summer**):
$$K - TCI = 2(3.07\boldsymbol{Cd} + 1.90\boldsymbol{Ca} + 3.27\boldsymbol{P} + 0.90\boldsymbol{W} + 0.86\boldsymbol{S})$$

Cd
(high noon,
The Thermal comfort)

- Maximum temperature
- Minimum relative humidity

Ca
(Daily average,
The Themal comfort)

- Average temperature
- Average relative humidity

P

- Total rainfall

W

- Average wind speed

S

- Duration of daylight

Range	Category
91~100	Ideal
81~90	Excellent
71~80	Very good
61~70	Good
51~60	Acceptable
41~50	Marginal
40~	Unfavorable

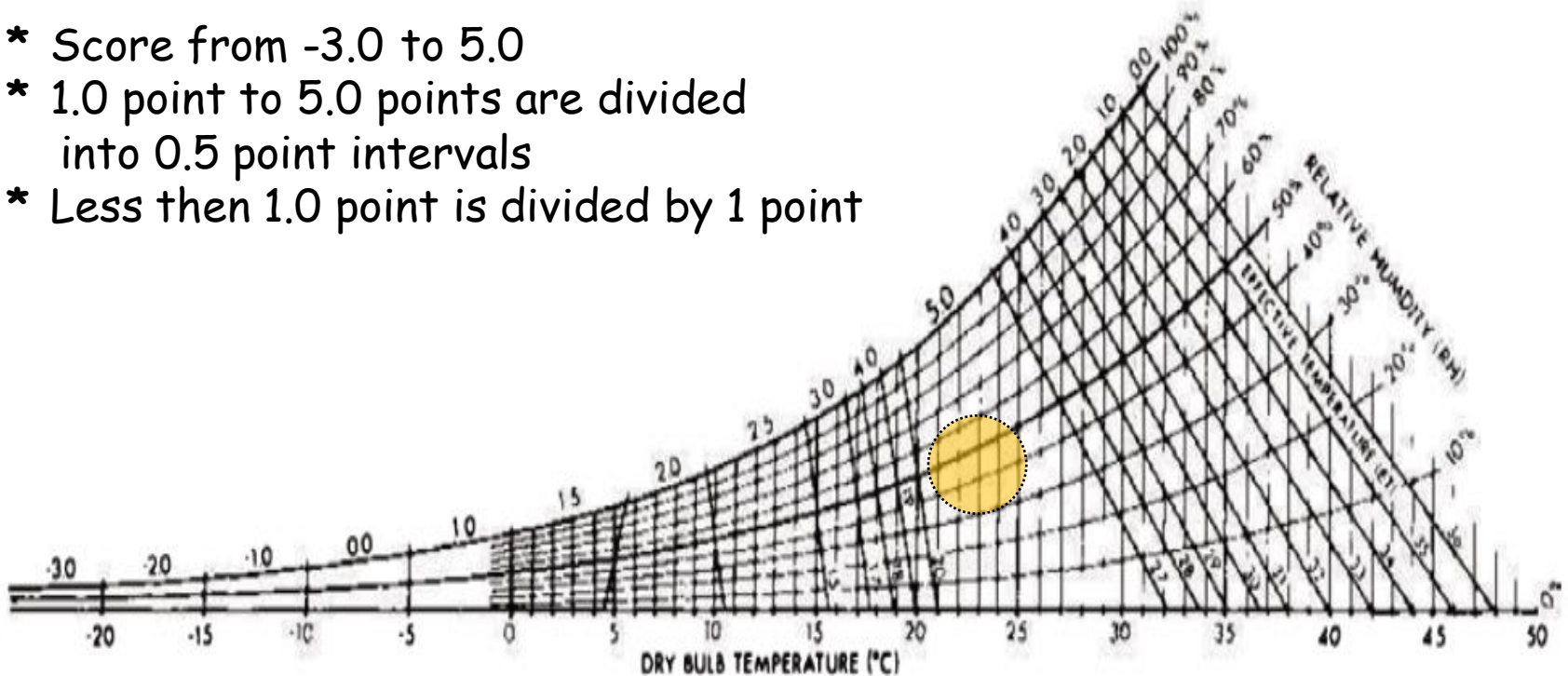Cd : Climate days, Ca : Climate average, P : Precipitation, W : Wind speed, S : Sunlight hours

DAEGU UNIVERSITY

Calculating score about Cd and Ca
- upon temperature and relative humidity.

* Score from -3.0 to 5.0
* 1.0 point to 5.0 points are divided into 0.5 point intervals
* Less then 1.0 point is divided by 1 point

Mieczkowski, Z. (1985). The tourism climatic index: a method of evaluating world climates for tourism. *The Canadian Geographer/Le Géographe canadien, 29*(3), 220-233.

Calculating score about P, W and S

Precipitation

Range	Score
- 5mm	5.0
5 – 10mm	4.5
10 – 15mm	4.0
15 – 20mm	3.5
20 – 25mm	3.0
25 – 30mm	2.5
30 – 35mm	2.0
35 – 40mm	1.5
40 – 45mm	1.0
45 – 50mm	0.5
50mm -	0.0

Wind Speed

Range	Score
- 0.8m/s	5.0
0.8 - 1.6m/s	4.5
1.6 - 2.51m/s	4.0
2.52 - 3.40m/s	3.5
3.41 - 5.50m/s	3.0
5.51 - 6.75m/s	2.5
6.75 - 8.00m/s	2.0
8.01 - 10.7m/s	1.0
10.8 -	0.0

Sunlight hours

Range	Score
10 hours -	5.0
9 hours – 10 hours	4.5
8 hours – 9 hours	4.0
7 hours – 8 hours	3.5
6 hours – 7 hours	3.0
5 hours – 6 hours	2.5
4 hours – 5 hours	2.0
3 hours – 4 hours	1.5
2 hours – 3 hours	1.0
1 hours – 2 hours	0.5
- 1hours	0.0

Methodology

Social media data : Tripadvisor

- The world's largest travel review websites.
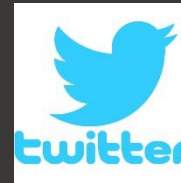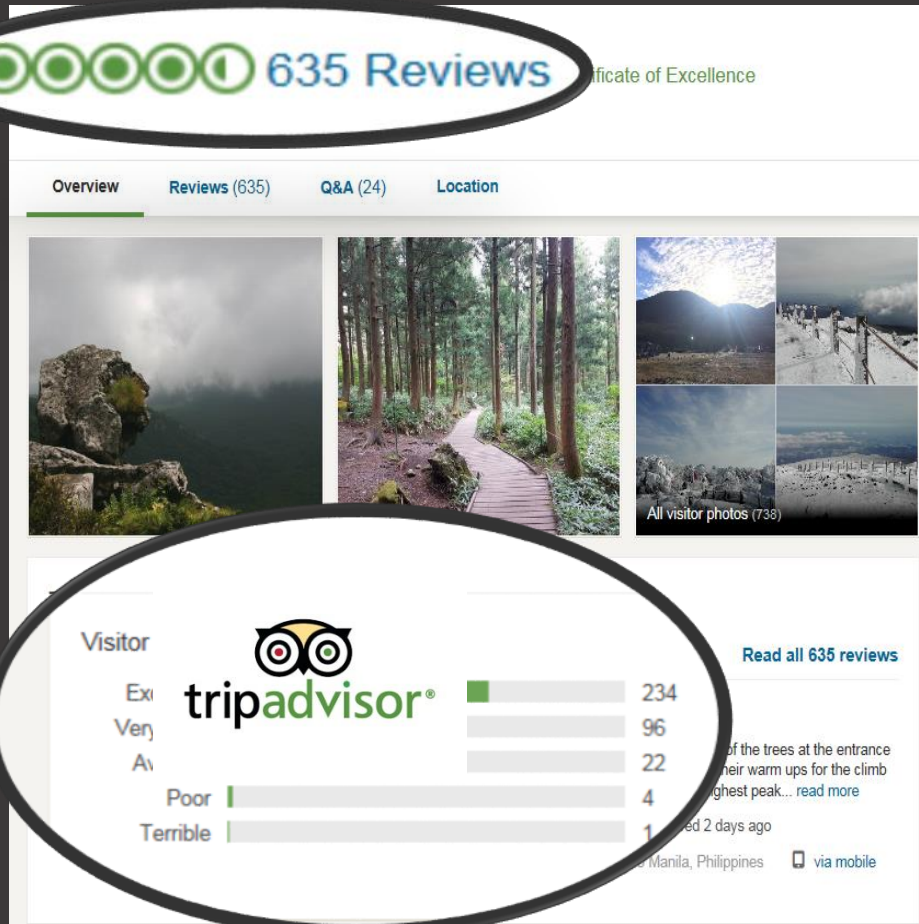
Twitter, Instagram, facebook were considered at the beginning of research.

However, they were limited in the collection of data because of the privacy.

Selection of tourism sites

Tourism sites : 305

97 sites

Reference : tripadvisor.com

Attraction	Category	Indoor/ outdoor	Lat	Lon
Seongsan Ilchulbong Tuffcone	Mountain	outdoor	33.46	126.93
Hallasan National Park	Mountain	outdoor	33.38	126.52
Udo	Island	outdoor	33.50	126.95
O'sulloc TeaMuseum	Ecological park	outdoor	33.31	126.29
Teddy bear Museum	Museum	indoor	33.25	126.41
Cheonjiyeon Falls	Mountain	outdoor	33.25	126.55
Jeongbang Falls	Mountain	outdoor	33.25	126.57
Manjanggul Cave	Mountain	outdoor	33.53	126.77
Alive Art Museum	Museum	indoor	33.26	126.41
Trick Art Museum	Museum	indoor	33.40	126.78
The Museum of Sex Health	Museum	outdoor	33.26	126.34
Jeju Flok Village Museum	Museum	indoor	33.32	126.84
Beach hyeopjjae	Beach	outdoor	33.39	126.23

. . .

The map of locations

Weather stations ASOS :   4 sites
                  AWS : 22 sites
Popular Tourism sites : 97 sites
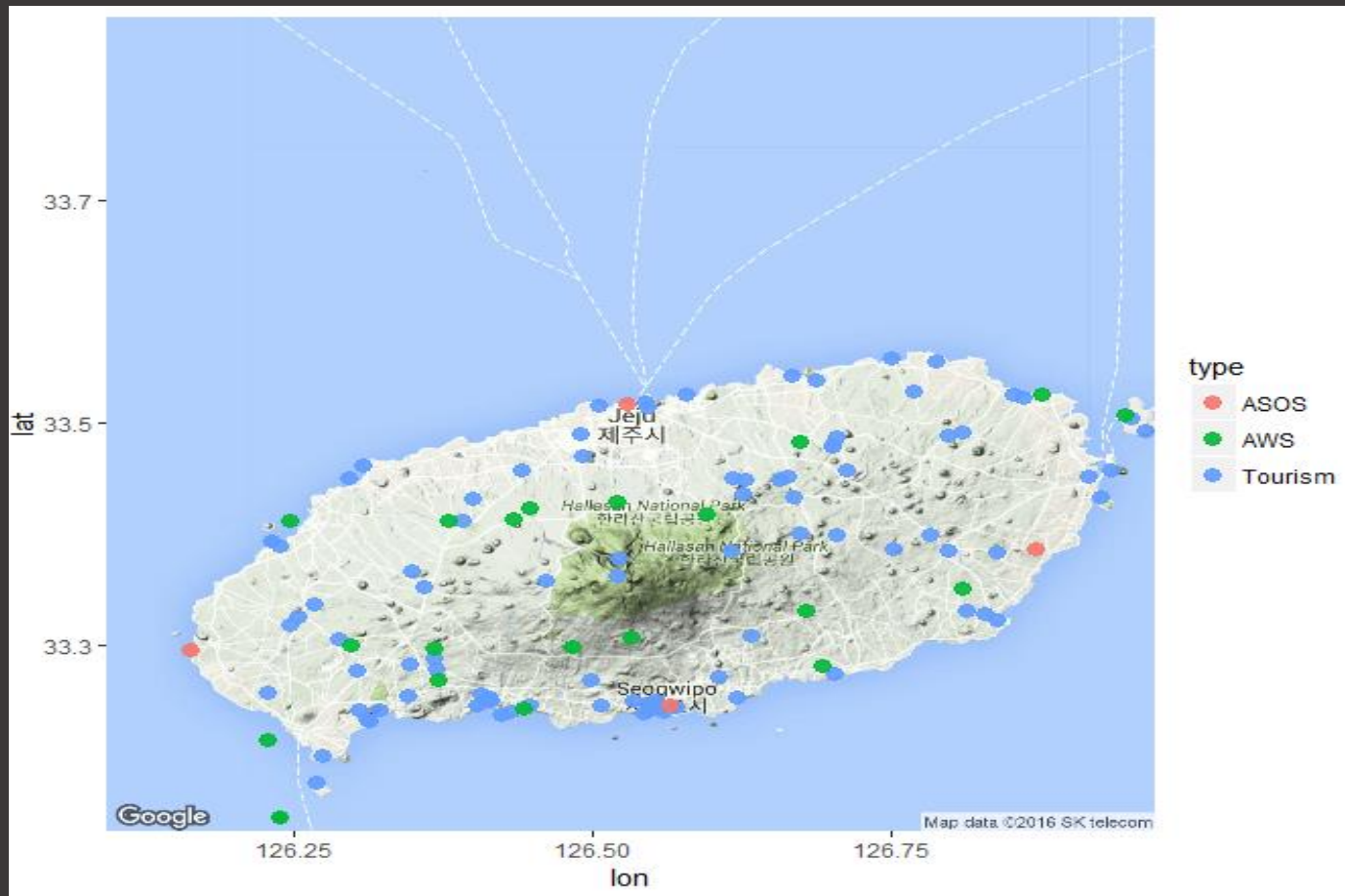
Data

Period : 2014-2015
Season : Summer (June, July, August)
Variable : Temperature, Precipitation, Wind speed,
        Relative humidity, Daylight

Example : June 1, 2014

| id	Average temperature	Maximum temperature	Daily precipitation	Average wind speed	Minimum relative humidity	Average relative humidity	Daylight hours
ASOS_184	22.4	25	0.1	1.8	37	65	3.5
ASOS_185	21.4	25.3	0	6	41	80	3.3
ASOS_188	20.3	23.9	0	2.5	64	89	4.4
ASOS_189	22.3	24.8	0	1.3	42	75	4.8
AWS_328	22.4	25.7	0	1.3			
AWS_329	21.3	27.8	0	2.2			
AWS_330	20.7	25.1	0	2.4			
AWS_685	21.3	26	0	2.5			
AWS_724	19.4	21	0	2.2			
⋮	⋮	⋮	⋮	⋮			
AWS_751	21.5	26	0	3.9			
AWS_752	21.4	25.3	0	3.9			

Predicting weather information at tourism sites.
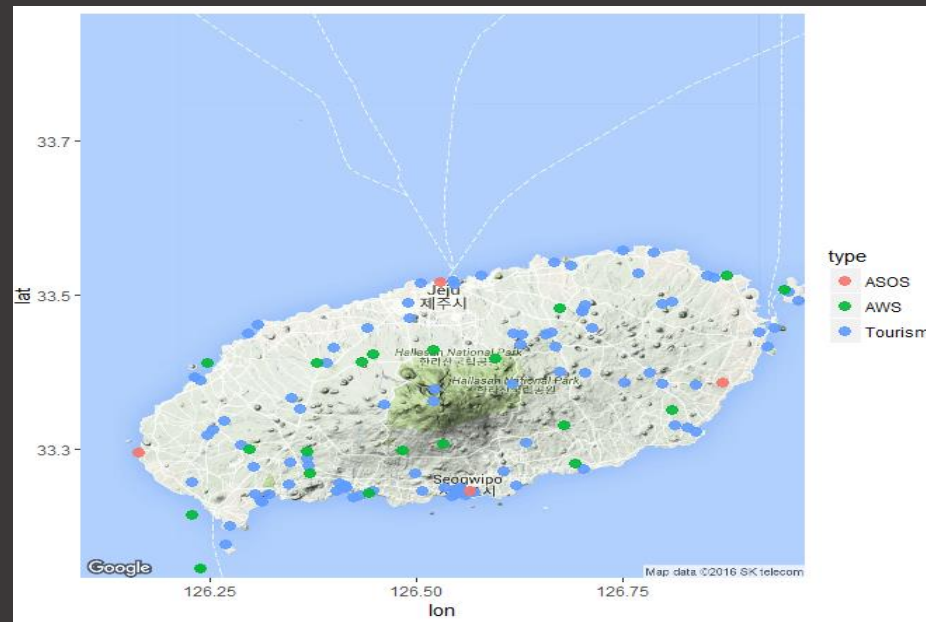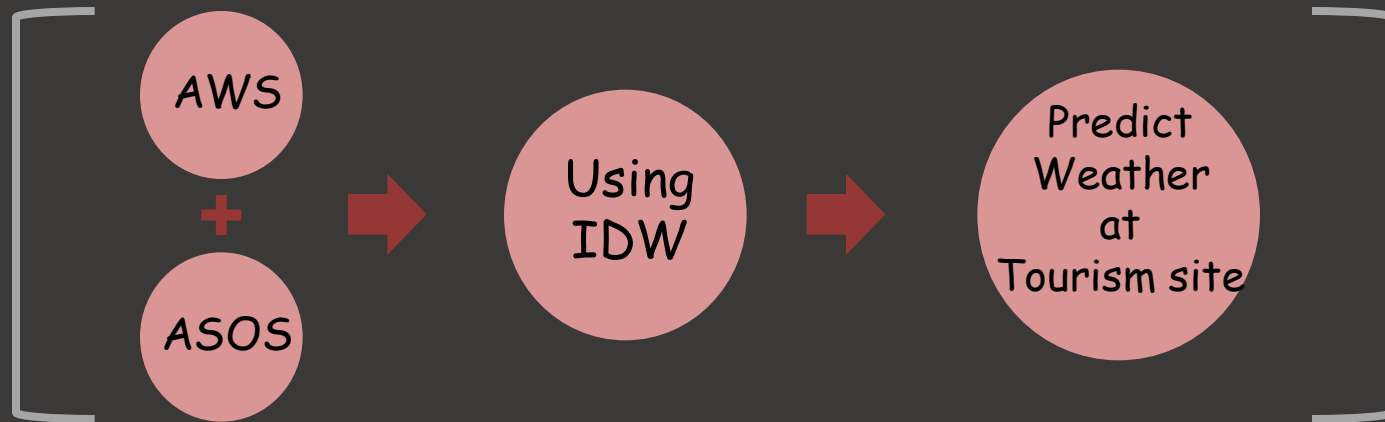
[On August 1, 2014] – Weather at tourism sites

attractions	Average temperature	Maximum temperature	Daily precipitation	Average wind speed	Minimum relative humidity	Average relative humidity	Daylight hours
Seongsan Ilchulbong Tuff Cone	24.48	26.11	26.63	4.00	81.60	91.79	0.09
Hallasan National Park	23.16	24.47	37.58	4.13	79.33	89.47	0.17
Udo	25.06	27.06	26.25	3.67	81.20	91.40	0.11
Hamdeok Beach	23.86	25.32	31.49	4.08	77.22	88.63	0.16
Jeongbang Falls	24.78	26.09	24.18	3.60	86.65	92.48	0.01
Jusangjeolli Cliff	23.88	25.19	33.90	3.51	82.59	91.28	0.25
Hyeopjae Beach	24.47	26.00	24.54	4.44	83.20	92.19	0.55
Cheonjiyeon Falls	23.84	25.14	32.20	3.88	85.62	92.14	0.05
Hallim Park	24.43	25.95	24.73	4.43	83.20	92.17	0.54
Manjanggul Cave	24.05	25.55	29.79	4.08	79.68	90.27	0.14
O'sulloc TeaMuseum	23.79	25.14	24.42	3.83	83.27	92.14	0.51
Eco Land Theme park	23.73	25.11	33.48	4.08	78.42	89.26	0.15
Teddy Bear Museum	23.83	25.10	32.94	3.56	82.43	91.24	0.27
Dongmun Market	23.82	25.15	31.13	4.08	83.91	91.60	0.09

. . .

▶Use IDW to predict daily weather at 97 locations.
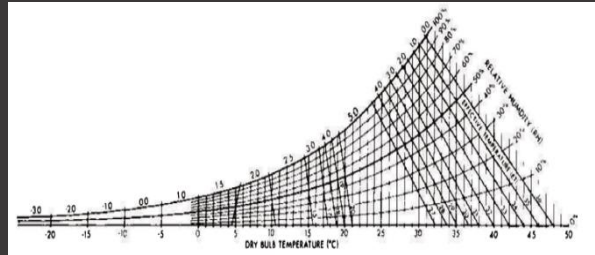
Range	Score
- 5mm	5.0
5 - 10mm	4.5
10 - 15mm	4.0
15 - 20mm	3.5
20 - 25mm	3.0
25 - 30mm	2.5
30 - 35mm	2.0
35 - 40mm	1.5
40 - 45mm	1.0
45 - 50mm	0.5
50mm -	0.0

Range	Score
- 0.8m/s	5.0
0.8 - 1.6m/s	4.5
1.6 - 2.51m/s	4.0
2.52 - 3.40m/s	3.5
3.41 - 5.50m/s	3.0
5.51 - 6.75m/s	2.5
6.75 - 8.00m/s	2.0
8.01 - 10.7m/s	1.0
10.8 -	0.0

Range	Score
10 hours -	5.0
9 hours – 10 hours	4.5
8 hours – 9 hours	4.0
7 hours – 8 hours	3.5
6 hours – 7 hours	3.0
5 hours – 6 hours	2.5
4 hours – 5 hours	2.0
3 hours – 4 hours	1.5
2 hours – 3 hours	1.0
1 hours – 2 hours	0.5
- 1hours	0.0

K-TCI formula (summer):

$$K - TCI = 2(3.07Cd + 1.90Ca + 3.27P + 0.90W + 0.86S)$$

Range	Category
91~100	Ideal
81~90	Excellent
71~80	Very good
61~70	Good
51~60	Acceptable
41~50	Marginal
40~	Unfavorable

DAEGU UNIVERSITY

K-TCI formula (summer):

$$K - TCI = 2(3.07Cd + 1.90Ca + 3.27P + 0.90W + 0.86S)$$

attractions	Average temperature	Maximum temperature	Daily precipitation	Average wind speed	Minimum relative humidity	Average relative humidity	Daylight hours	KTCI	Category
Seongsan Ilchulbong Tuff Cone	24.48	26.11	26.63	4.00	81.60	91.79	0.09	64.034	4_good
Hallasan National Park	23.16	24.47	37.58	4.13	79.33	89.47	0.17	57.362	5_acceptable
Udo	25.06	27.06	26.26	3.67	81.19	91.40	0.11	63.988	4_good
Hamdeok Beach	23.86	25.32	31.49	4.08	77.22	88.63	0.16	62.135	4_good
Jeongbang Falls	24.78	26.09	24.18	3.60	86.65	92.48	0.01	64.852	4_good
Jusangjeolli Cliff	23.88	25.12	33.90	3.51	82.59	91.28	0.25	59.691	5_acceptable
Hyeopjae Beach	24.47	26.00	24.53	4.44	83.20	92.19	0.55	65.577	4_good
Cheonjiyeon Falls	23.84	25.14	32.18	3.88	85.62	92.13	0.05	60.068	4_good
Hallim Park	24.43	25.95	24.73	4.43	83.18	92.17	0.54	65.463	4_good
Manjanggul Cave	24.05	25.54	29.79	4.08	79.68	90.27	0.14	62.627	4_good
O'sulloc TeaMuseum	23.79	25.14	24.42	3.84	83.27	92.14	0.51	65.880	4_good
Eco Land Theme park	23.73	25.11	33.45	4.08	78.42	89.26	0.15	60.676	4_good
Teddy Bear Museum	23.82	25.09	32.94	3.56	82.43	91.24	0.27	78.061	3_very good
Dongmun Market	23.82	25.15	31.13	4.07	83.91	91.59	0.09	61.093	4_good

. . .

Result (K-TCI map)
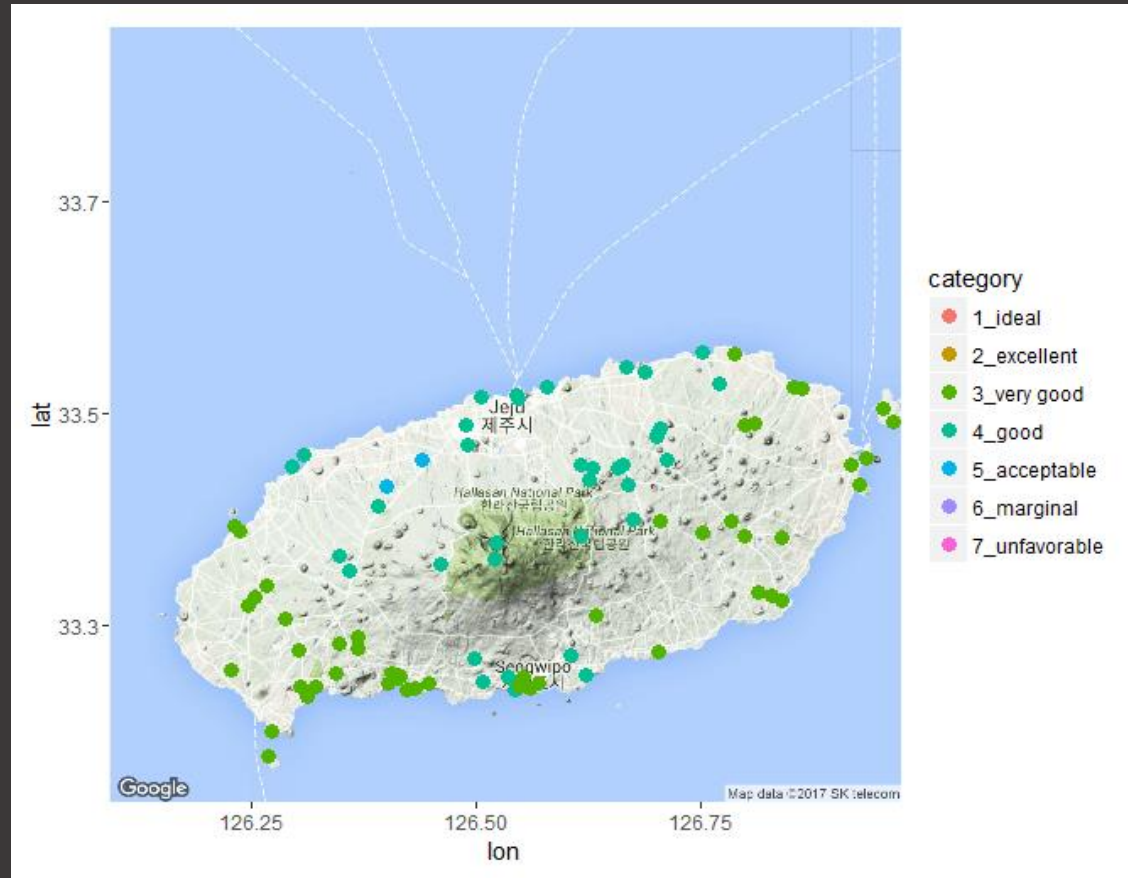
▶ **More practical** than currently available services by KMA

▶ K-TCI can be provide by **time zone** when hourly data is used.

Modifying K-TCI

‹ **To prohibit a traveling** when bad weather condition such as **Typhoon**.

1. Point where temperature is above **30 degrees** (unfavorable)
2. Above average wind speed of **14m/s** (unfavorable)
3. Above rainfall **110mm/day** (unfavorable)

	Watching	Warning
Heat wave	**highest temperature** is more than **33℃** for more than **two days**	**highest temperature** is more than **35℃** for more than **two days**
Gale	**wind speed** is more than **14m/s** or **instantaneous wind speed** is more than **20m/s**	**wind speed** is more than **21m/s** or **instantaneous wind speed** is more than **26m/s**
Downpour	6 hour precipitation is over **70mm** or 12 hour precipitation is over **110mm**	**6 hour** precipitation is over **110mm** or the **12hour** rainfall is over **180mm**

▲ **Natural hazard waring reference** in KMA

DAEGU UNIVERSITY

Affected by typhoon,

▶ originally 'acceptable' and 'marginal' is provided.

Example :

2014. 8. 3

typhoon : Nakri
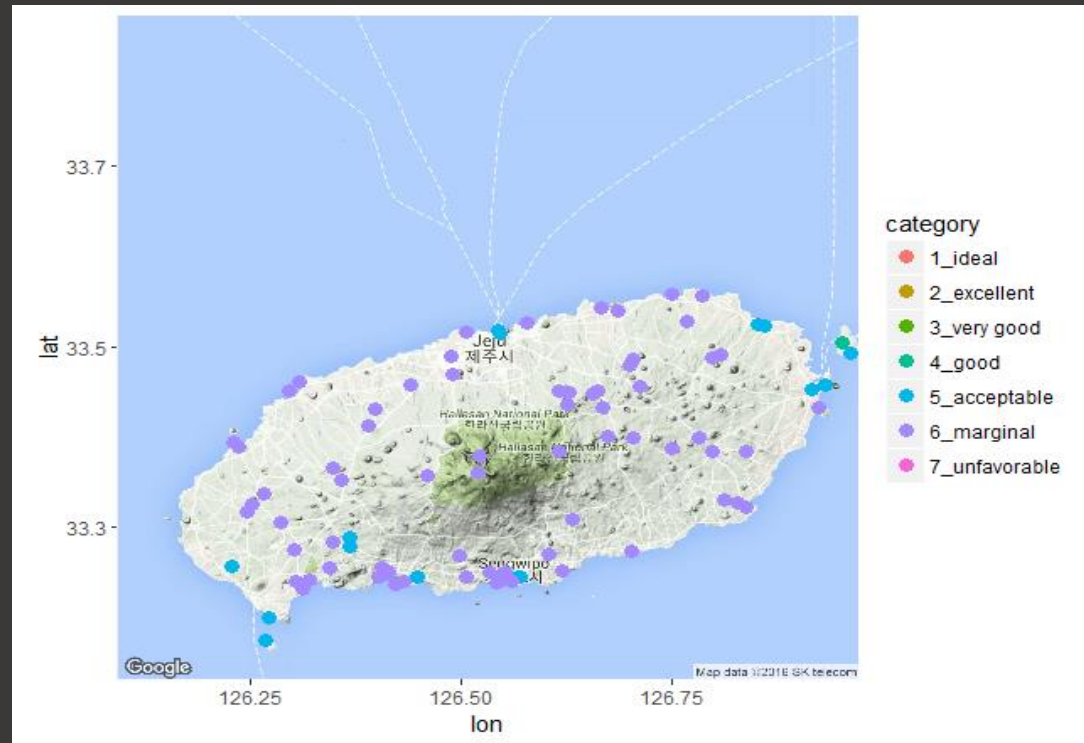
make 'unfavorable'

Before
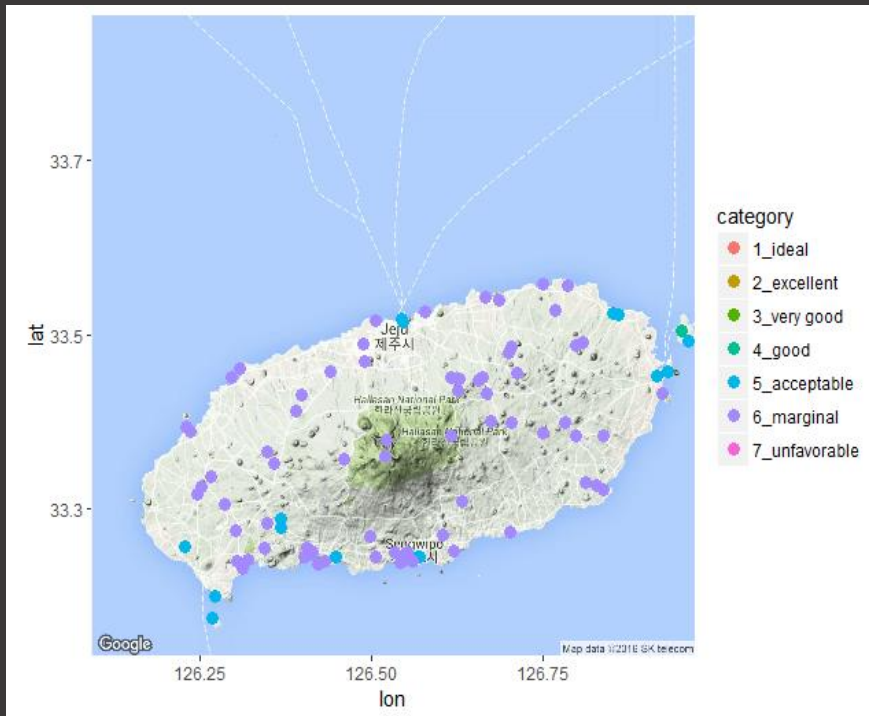
After
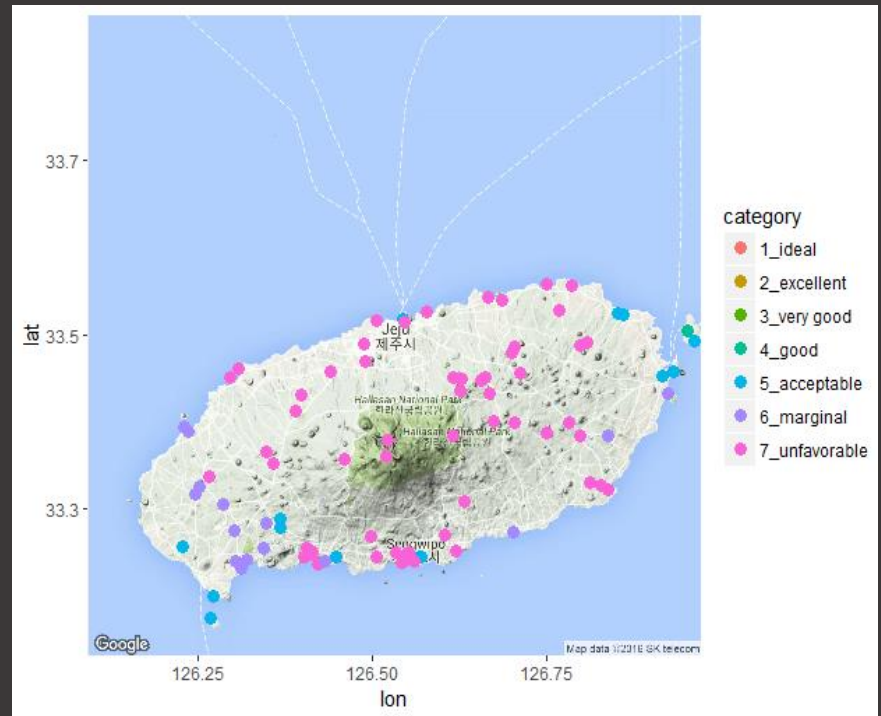
Most attractions appear on the map as 'unfavorable'

Modifying K-TCI

- indoor tourism sites and weather

Indoor attractions is less affected by weather

1. Except **precipitation** is over **110mm** and **wind speed** is over **8m/s**
2. Indoor attractions get **4.5 points** in **P** and **W**
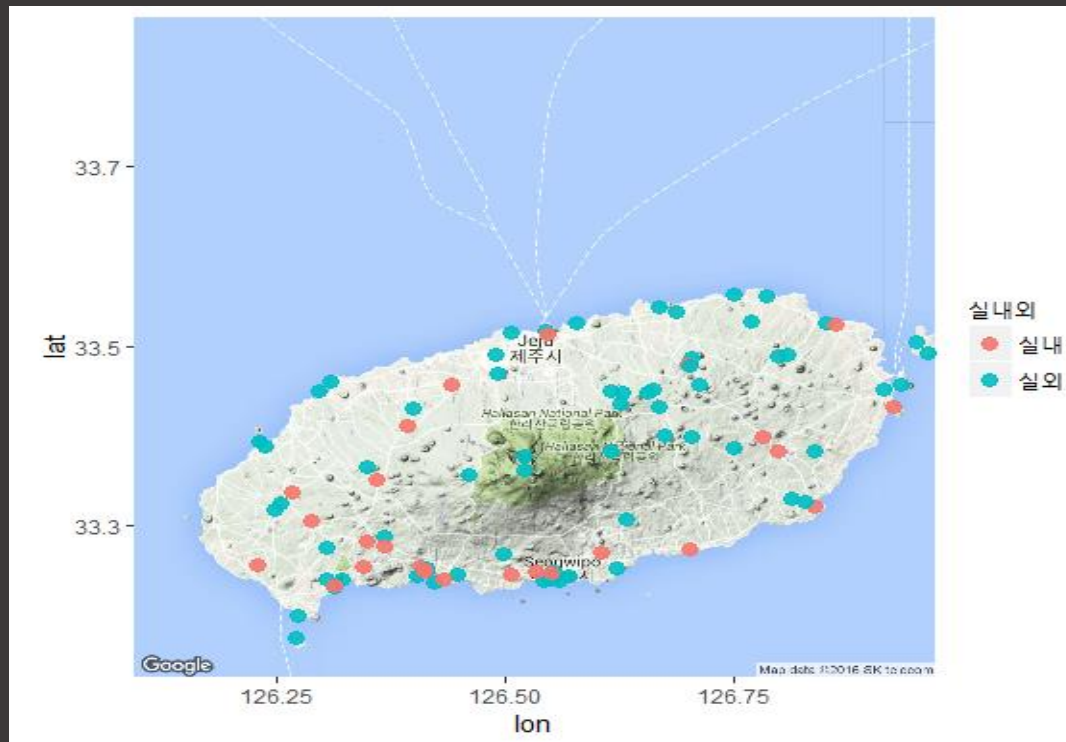
Before
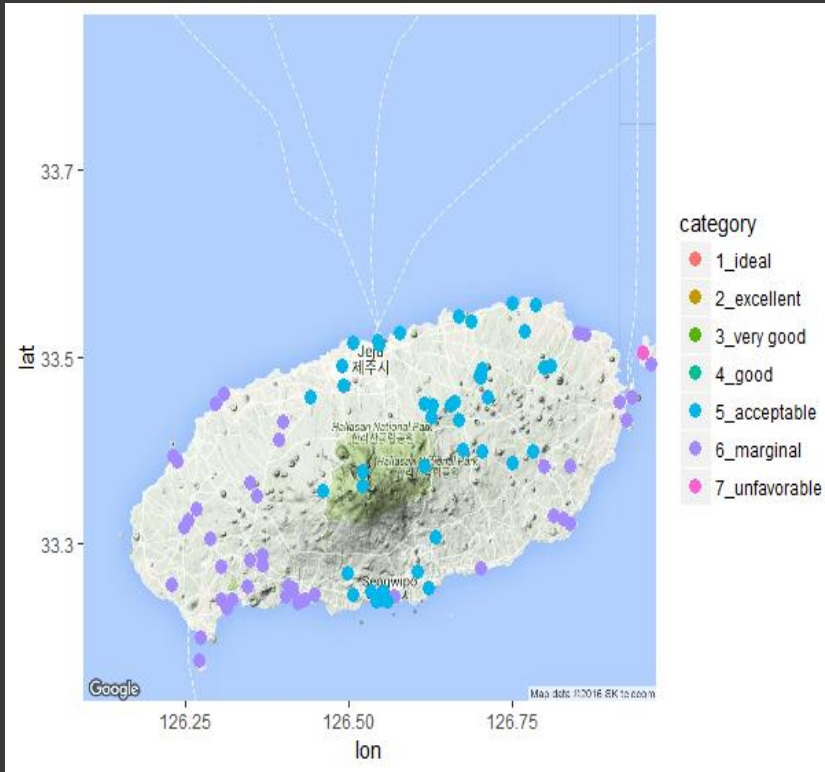
After
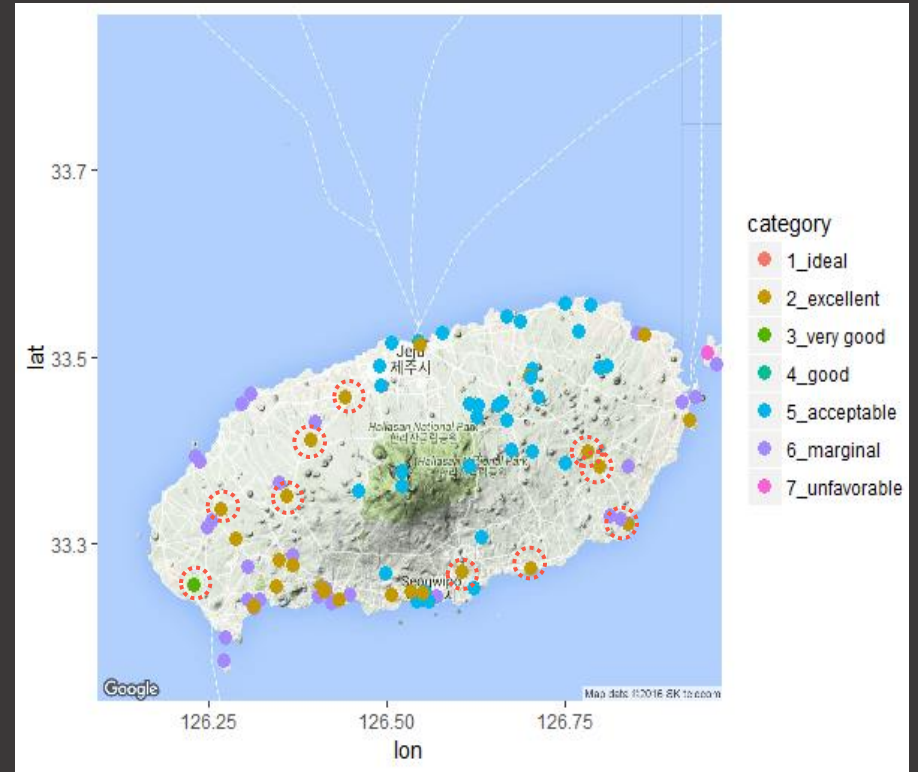
▶ **Indoor attractions improved between 1 and 3 grade**

Develop ranking formula

K-TCI & Tripadvisor ranking formula

score = (grade^α) * (The number of reviews^β) * (K-TCI score)

attractions	5	4	3	2	1	review
Seongsan Ilchulbong Tuff Cone	136	83	12	1	0	981
Hallasan National Park	56	13	7	0	0	518
Udo	112	36	6	0	0	440
Hamdeok Beach	55	23	3	0	0	167
Jeongbang Falls	28	39	10	0	0	349
Jusangjeolli Cliff	55	43	9	2	0	271
Hyeopjae Beach	55	33	3	0	0	190
Cheonjiyeon Falls	37	34	12	2	0	370
Hallim Park	20	17	6	1	0	262
Manjanggul Cave	19	17	6	4	0	608
O'sulloc TeaMuseum	59	59	46	8	3	438
Eco Land Theme park	35	36	4	0	0	142
Teddy Bear Museum	15	27	22	5	1	436
Dongmun Market	18	35	12	2	1	198
⋮						

Result (Sunny day)

August 27, 2014

Rank	KTCI standard		Tripadvisor standard		K-TCI & Tripadvisor	
1	Camellia Hill	O	Seongsan Ilchulbong Tuff Cone	O	O'sulloc TeaMuseum	O
2	Seobok park	O	Manjanggul Cave	O	Seongsan Ilchulbong Tuff Cone	O
3	Jeongbang Falls	O	Hallasan National Park	O	Jeongbang Falls	O
4	O'sulloc TeaMuseum	O	Udo	O	Udo	O
5	Hyeopjae Beach	O	O'sulloc TeaMuseum	O	Hallim Park	O
6	Beach of the geumneung	O	Teddy Bear Museum	I	Teddy Bear Museum	I
7	Songak Mountin	O	Cheonjiyeon Falls	O	Cheonjiyeon Falls	O
8	Jeju Glass Museum	O	Jeongbang Falls	O	Hyeopjae Beach	O
9	Hallim Park	O	Yongduam Rock	O	Jusangjeolli Cliff	O
10	Shell Museum of the world	I	Seongsan Ilchulbong Tuff Cone	O	Manjanggul Cave	O

UNIVERSITY

Result (Rainy and windy day)

August 24, 2015

Rank	KTCI standard		Tripadvisor standard		K-TCI & Tripadvisor	
1	Chocolate Museum	I	Seongsan Ilchulbong Tuff Cone	O	Teddy Bear Museum	I
2	Banglimwon	O	Manjanggul Cave	O	Alive Museum	I
3	Jumbo Village	I	Hallasan National Park	O	Trick Art Museum	I
4	The Museum of Sex Health	I	Udo	O	Aqua peullanet	I
5	World Automobile Museum	I	O'sulloc TeaMuseum	O	Ripley's Believe It of Not!	I
6	Chocolate Land	I	Teddy Bear Museum	I	Jeju Folk Village Museum	I
7	Teddy Bear Museum	I	Cheonjiyeon Falls	O	The Museum of Sex Health	I
8	Play K-POP	I	Jeongbang Falls	O	Haenyeo Museum	I
9	Alive Museum	I	Yongduam Rock	O	World Automobile Museum	I
10	Ripley's Believe It of Not!	I	Jusangjeolli Cliff	O	Tejium, Jeju	I

GU UNIVERSITY

Conclusion

- IDW, K-TCI, Tripadvisor is considered to develop weather recommendation algorithm

- The production is reflected in K-TCI to evaluate the tourism climate at the sightseeing spot

- To distinguish weather disaster from indoor and outdoor to create a tourism climate reflecting weather conditions

- Developed a recommendation algorithm by combining with K-TCI as well as Tripadvisor which includes travel information of people
- There are ways of predicting tourist attractions as well IDW as well as crazing, recent points, and so on. I did not compare which method is good, but it seems that approaching with IDW is not bad

- Weights were randomly given when combining Tripadvisor grade and review.
  In this part the future, we will give a clear weight to match the actual number of tourists

- Like the Korea Meteorological Administration, it is good to show it is good to show it in real time, but I could not create it in real time

- I predicted the weather at the tourist spot, but I could not evaluate the accuracy.

- It is possible to estimate the accuracy of the prediction by assuming that some arbitrary number is not observed and comparing the predicted value with the observed value.
  But that part has been omitted from this presentation, not the main concern.

- By showing the database of the tourist points on a monthly basis and linking the recommendation algorithm, you can get the beta corresponding to the power.

R을 이용한 **빅데이터 사례 2**

EXAMPLE2.R

Safety road service based on observed visibility

Contents

Background and Purpose

✓ The most important influence on Visibility in our traffic is fog.

✓ The visibility depends on the concentration of fog.

✓ This results in an increase in the risk of a traffic accident, influencing the driver's speed and type of driving.

출발 안갯속 106대 추돌…2명 사망·부상 60여명

출발 "짙은 안개로 가시거리 10여m 불과"

The west coast highway
one hundred six-car collision, 2015
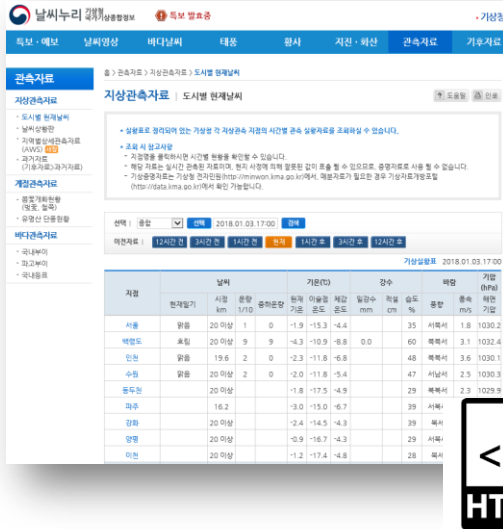
- ✓ According to the trend of traffic accidents by weather conditions, the 10-year average was 13.4 % on foggy days, higher than the clearer days (2.8 %).

- ✓ If you look at the service provided by the Korea Meteorological Administration, it is displayed locally.

- ✓ There is a limit to the combined information of road networks.

- ✓ Despite the dangers of the roadway fog, there is a lack of information service in Korea.

Crawling

: It is a technology to create new data by collecting only necessary data based on the unformatted data, url, and sauce provided by the website.

Real-time weather information (KMA)   **Dataset in R (for weather service contents)**
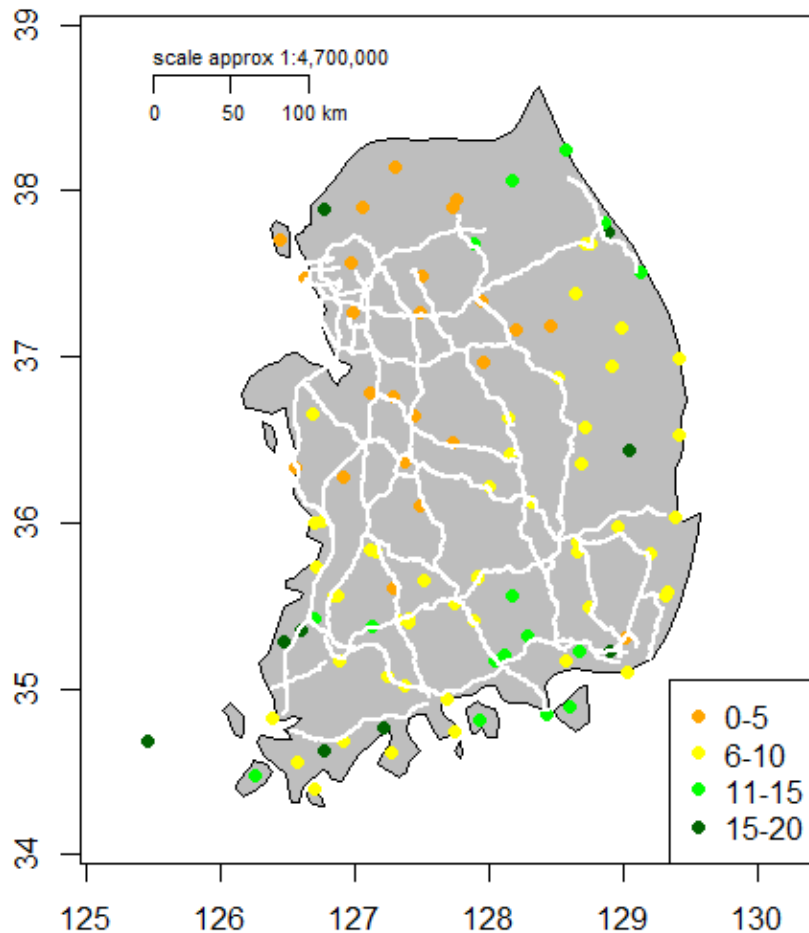

Crawling

Real-time weather data is presented in R using crawling.

Standard node link

: Electronic traffic maps designed to collect and provide traffic information. Compatibility and Effective Deployment and Management System
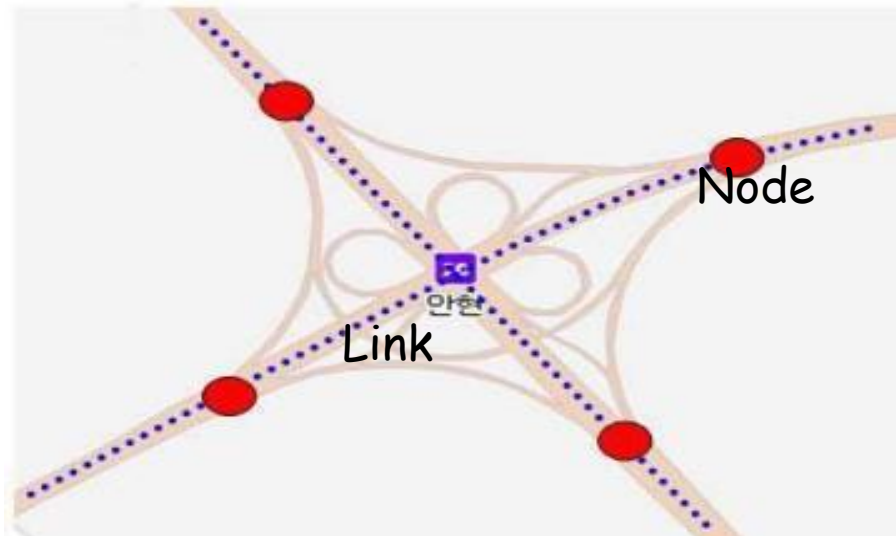
Node : Intersection of a chart or a graph
Link : A line connecting the dots

Domestic status
Standard node point count : 265,953
Standard node count : 6,953

IDW (Inverse distance weight method)

✓Inverse distance-based weight interpolation computes a **weighted average**,

$$\hat{Z}(S_0) = \frac{\sum_{i=1}^{n} w(s_i) Z(s_i)}{\sum_{i=1}^{n} w(s_i)}$$

✓ where $\hat{Z}(S_0)$ is a value of a prediction location (s_0) and $Z(s_i)$ is a value of observational location (s_i).

✓ The weights $w(s_i)$ for observations are computed according to their distance to the interpolation location, where are consisted by

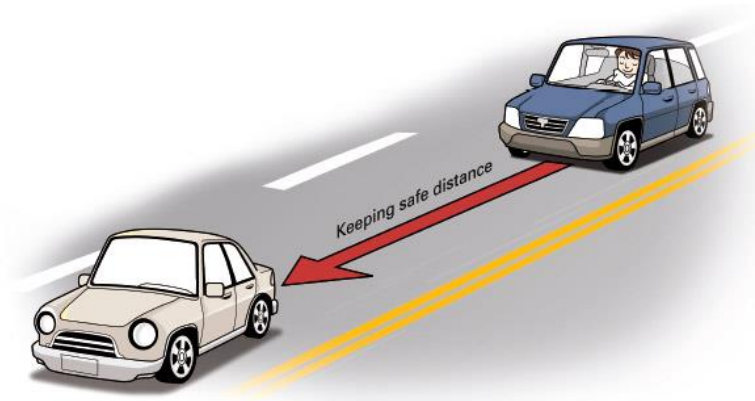✓ **Euclidean distance** and **an inverse distance weighting power** (p=2)

$$w(s_i) = ||s_i - s_o||^{-p}$$

The predicted values never **outside the range of observed values.**

Visibility

：Maximum distance to be able to see the target visually

$$x = \frac{\textit{Stopping distance}}{\textit{visibility}} \qquad , 0 \leq x \leq 50$$

■ Keeping safe distance adequately based on vehicle speed can help prevent collision, reduce driver fatigue and provide more time to respond to unexpected situations.

0	0.5	50
Safety	*Caution*	*Danger*

Stopping sight distance

: The determination of stopping sight distance requires the definition and consideration of seven design variables

- Perception-reaction time
- Driver eye height
- Object height
- Vehicle operating speed
- Pavement coefficient of friction
- Deceleration rates
- Roadway grade

$$d = d_1 + d_2 = \frac{V}{3.6}t + \frac{V^2}{254f} = 0.694V + \frac{V^2}{254\,f},$$
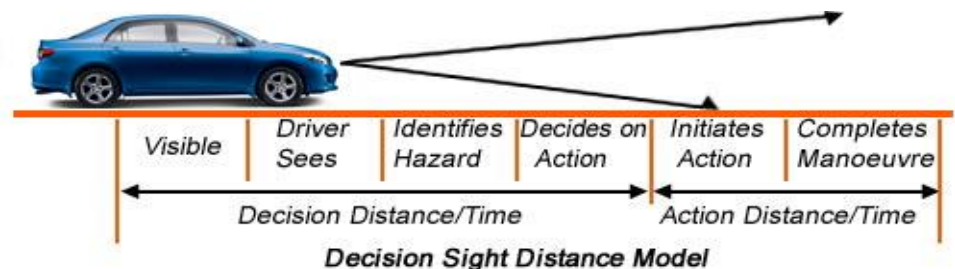
D: Stopping Sight Distance (m)
d_1: Reaction time Distance (m)
d_2: Braking Distance (m)
V: Speed (km/hr)
t: Reaction Time (2.5 sec)
f: Frictional coefficient (wet)

Visible	Driver Sees	Identifies Hazard	Decides on Action	Initiates Action	Completes Manoeuvre

Decision Distance/Time   Action Distance/Time

Decision Sight Distance Model

Example – Differences by friction factor

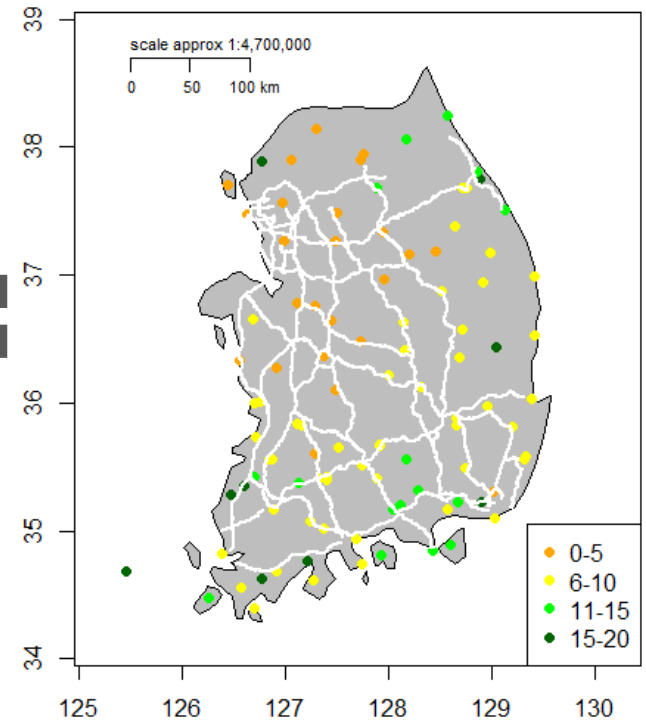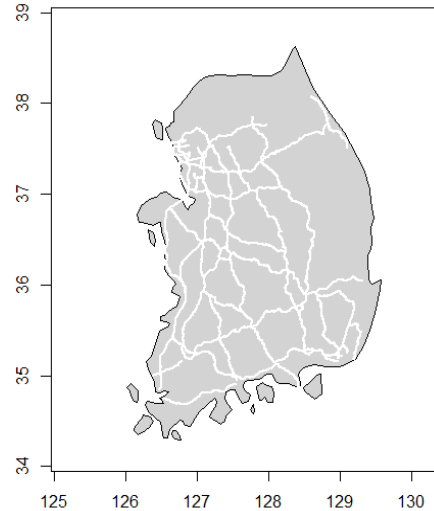**1) The stopping distance at dry road($V = 100$)**

$$d = 0.694V + \frac{V^2}{254\ f}, \qquad\qquad , f = 0.63$$

$$\cong 132$$

2) The stopping distance at wet road($V = 100$)

$$d = 0.694V + \frac{V^2}{254\ f}, \qquad\qquad , f = 0.28$$

$$\cong 210$$

When the road is set in asphalt, the force of friction varies with the weather. That is, stopping sight distance are affected by friction at the same speed.

# Data (criteria 2017.12.29)

**Real-time weather information
=> Crawling**

Standard node link

scale approx 1:4,700,000

| 0-5 |
| 6-10 |
| 11-15 |
| 15-20 |

- ✓ Observe it from a point within 2km.

- ✓ After assuming the missing value, the value is predicted.

- ✓ If observation value minus predict value is close to zero, it is predictable.

- ✓ However, this study failed to do so due to the lack of information on the visibility in the past.

Overlapping the highway with weather information

Real-time weather information => IDW

Standard node link
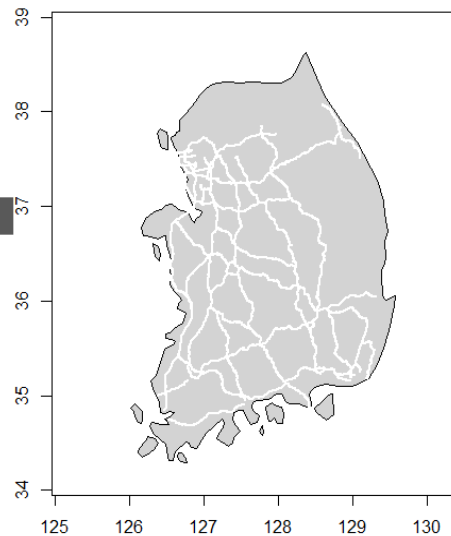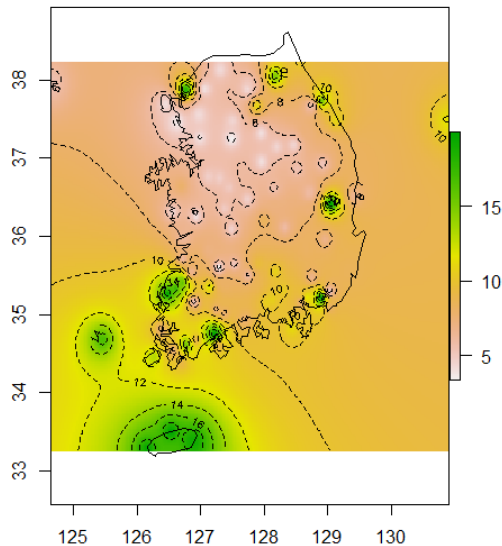
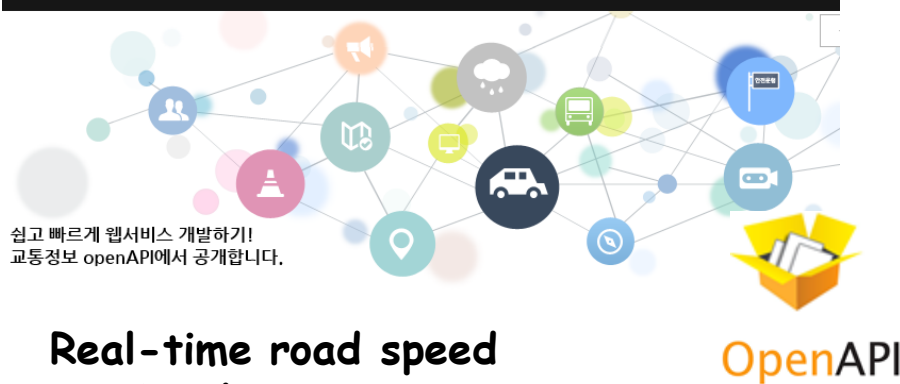**Node link to contour using IDW**
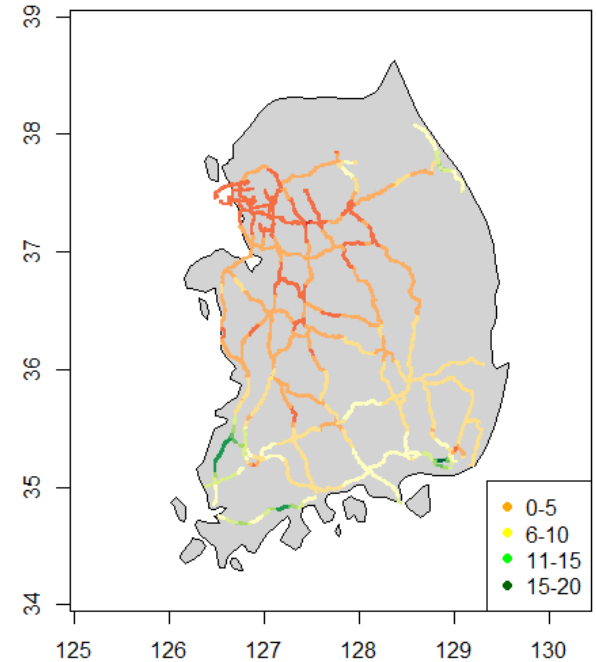
Real-time danger road position

Real-time road speed(ITS)
=> Crawling (open API)

Real-time danger road position

Real-time road speed
=> Crawling

```
<data>
    <roadsectionid>1520003501</roadsectionid>
    <avgspeed>91</avgspeed>
    <startnodeid>1520001301</startnodeid>
    <roadnametext>중앙고속도로</roadnametext>
    <traveltime>4</traveltime>
    <endnodeid>1520001300</endnodeid>
    <generatedate>20180104180027</generatedate>
</data>
<data>
    <roadsectionid>1520003502</roadsectionid>
    <avgspeed>91</avgspeed>
    <startnodeid>1540000100</startnodeid>
    <roadnametext>중앙고속도로</roadnametext>
    <traveltime>43</traveltime>
    <endnodeid>1520001301</endnodeid>
    <generatedate>20180104180027</generatedate>
```
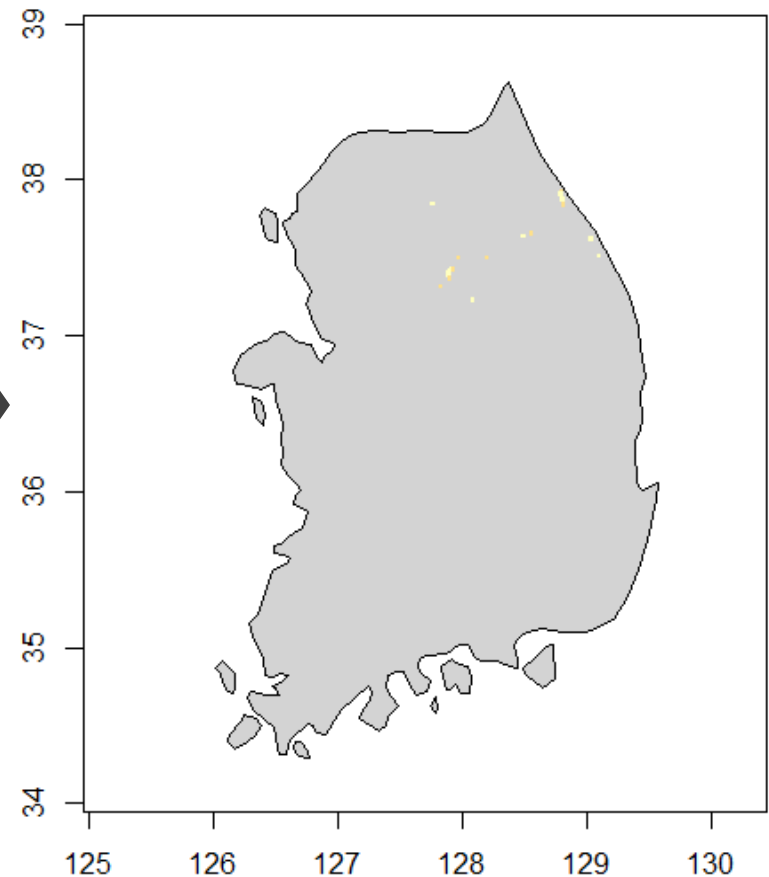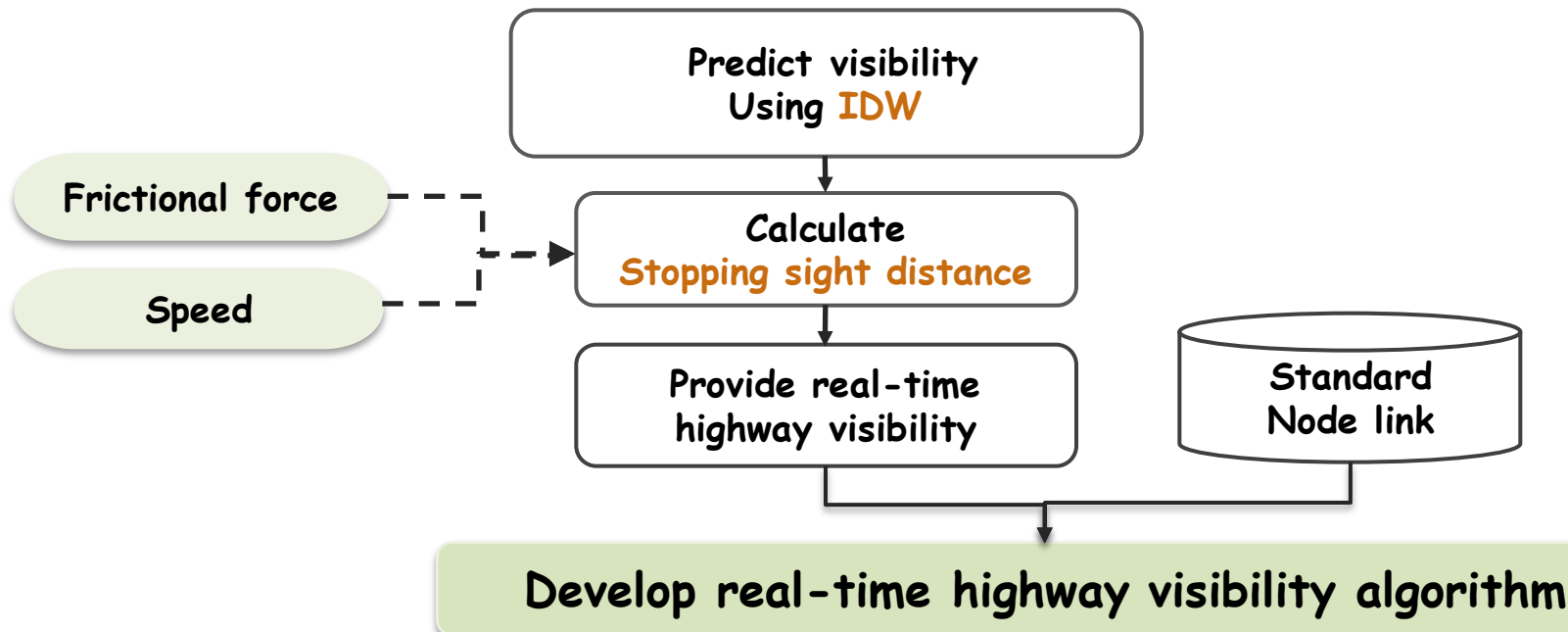
Stopping sight distance formula

$$d = d_1 + d_2 = \frac{V}{3.6}t + \frac{V^2}{254f} = 0.694V + \frac{V^2}{254\,f},$$

Conclusion

- IDW, Crawling, Node link were used to develop algorithm.
- IDW was used to predict visibility.
- Given weather information, stopping sight distance calculated.
- Real time highways can be checked for visibility distance, which improves stability.

R을 이용한 **빅데이터 사례 3**

EXAMPLE3.R

실시간 한반도 지진위험지도 생성 방법
A Method for Creating Real-Time Earthquake Hazard Map in Korean Peninsula

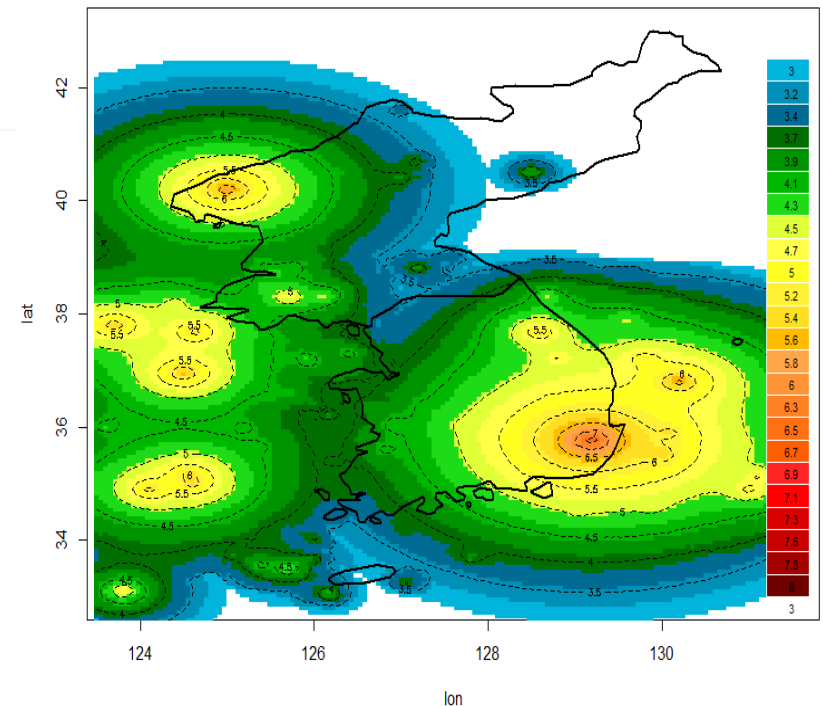오혜주 (제1) [1], 임지은 (참여) [2], 이예림 (참여) [3], 윤상후 (교신) [4]

[1] 대구대학교
[2] 대구대학교 통계학과
[3] 대구대학교
[4] 대구대학교

Earthquake risk of the Korean peninsula

https://prezi.com/t31c7uez7_3_/hazard-map-in-korean-peninsula/

감사합니다!