Fine dust and disease

Hong, yubin

Park, Eun Chong

Park Hyunsu

Han Ji Soo

Contents

Goal of project

01. Goal of project

Purpose

Find relevance between patient numbers and environmental variables
(temperature, fine dust, etc ...)

In summary,

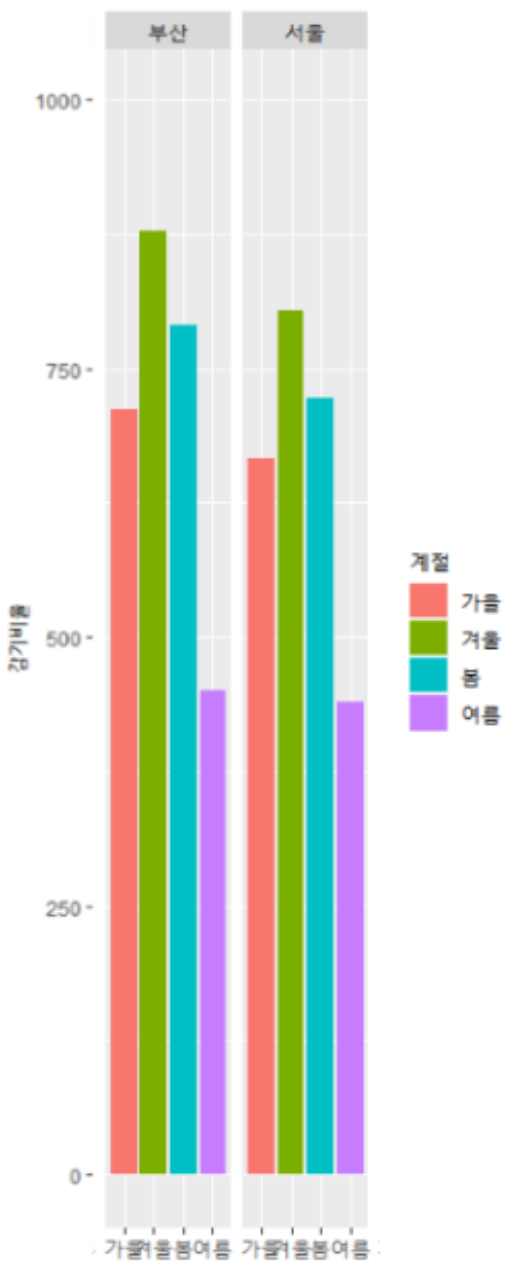
we raised awareness of the problem by showing

the relationship between disease and the environmental variables

to the citizens

Target region: Seoul, Busan

two most populous regions in Korea.



01. Goal of project

Data detail

1. Actual clinical data such as colds, asthma, skin diseases, eye diseases
   - Public data portal

2. Seoul, Busan Fine Dust Data
   - Public Health and Environmental Information System

3. Seoul, Busan Daily Weather Data

4. Additional data
- Population and population density by region – National indicator system
- longitude, latitude in Seoul, Busan city

연도	시도	감기	눈병	천식	피부병
20140101	서울	5992	247	715	246
20140101	부산	3158	91	554	138
20140101	대구	2190	96	370	85
20140101	인천	3426	83	437	138

측정 시간	측정항목					
	초미세먼지 PM-2.5(㎍/㎥)	미세먼지 PM-10(㎍/㎥)	아황산가스 SO2(ppm)	오존 O3(ppm)	이산화질소 NO2(ppm)	일산화탄소 CO(ppm)
환경기준	35/24/h이하	100/24h이하	0.15/1h이하	0.1/1h이하	0.1/1h이하	25/1h이하
09월 01일	16	20	0.01	0.023	0.02	0.2
02일	16	22	0.016	0.019	0.022	0.2

01. Goal of project

Data detail

1. Actual clinical data such as colds, asthma, skin diseases, eye diseases
   - Public data portal

2. Seoul, Busan Fine Dust Data
   - Public Health and Environmental Information System

날짜	시군구	천식환자수	감기환자수	눈병환자수	피부병환자수	초미세먼지	미세먼지	아황산가스	오존	이산화질소	일산화탄소
date	city	Asthma Patients	Cold patients	Eye disease patients	Skin disease patients	Ultrafine dust	fine dust	SO3	O3	NO2	CO
20140101	Gangnam-gu	15	205	11	8	29	117	0.01	0.016	0.031	0.5
20140102	Gangnam-gu	324	4916	393	551	23	54	0.008	0.01	0.041	0.5
20140103	Gangnam-gu	285	4471	395	532	30	59	0.008	0.003	0.055	0.7
20140104	Gangnam-gu	247	3659	311	404	25	49	0.007	0.006	0.041	0.5
20140105	Gangnam-gu	26	310	13	14	25	42	0.006	0.007	0.039	0.5
20140106	Gangnam-gu	378	5442	461	654	44	72	0.009	0.004	0.059	0.8
20140107	Gangnam-gu	226	3825	364	499	56	87	0.011	0.003	0.063	0.8
20140108	Gangnam-gu	239	3640	353	372	46	75	0.009	0.012	0.039	0.6
20140109	Gangnam-gu	215	3416	272	379	8	20	0.005	0.025	0.019	0.2
20140110	Gangnam-gu	259	3935	418	475	19	34	0.007	0.007	0.042	0.4
20140111	Gangnam-gu	215	3611	266	413	33	52	0.008	0.005	0.052	0.6

01. Goal of project

Data detail

3. Seoul, Busan Daily Weather Data (27 variables)

평균기온..C.	일강수량.mm.	최대.순간.풍속.m.s.	최대.풍속.m.s.	평균.풍속.m.s.	풍정합.100m.	평균.이슬점온도..C.	평균.상대습도...	평균.증기압.hPa.
Average temperature	Daily precipitation.mm.	Max. Instantaneous Wind Speed ??m.s.	Max. wind speed m.s.	Average wind speed m.s.	Wind.100m.	Average dew point temperature	Relative humidity	Mean.steam pressure.hPa.
4	0	14.8	7.8	3.7	3218	5.5	51.3	4.2
1.5	0	6	4	2	1685	12.9	35.4	2.3
2.3	0	7.8	5.4	2.5	2148	6.2	54.4	4
0.1	0	7.6	4.7	2.4	2087	10.3	47.1	2.8

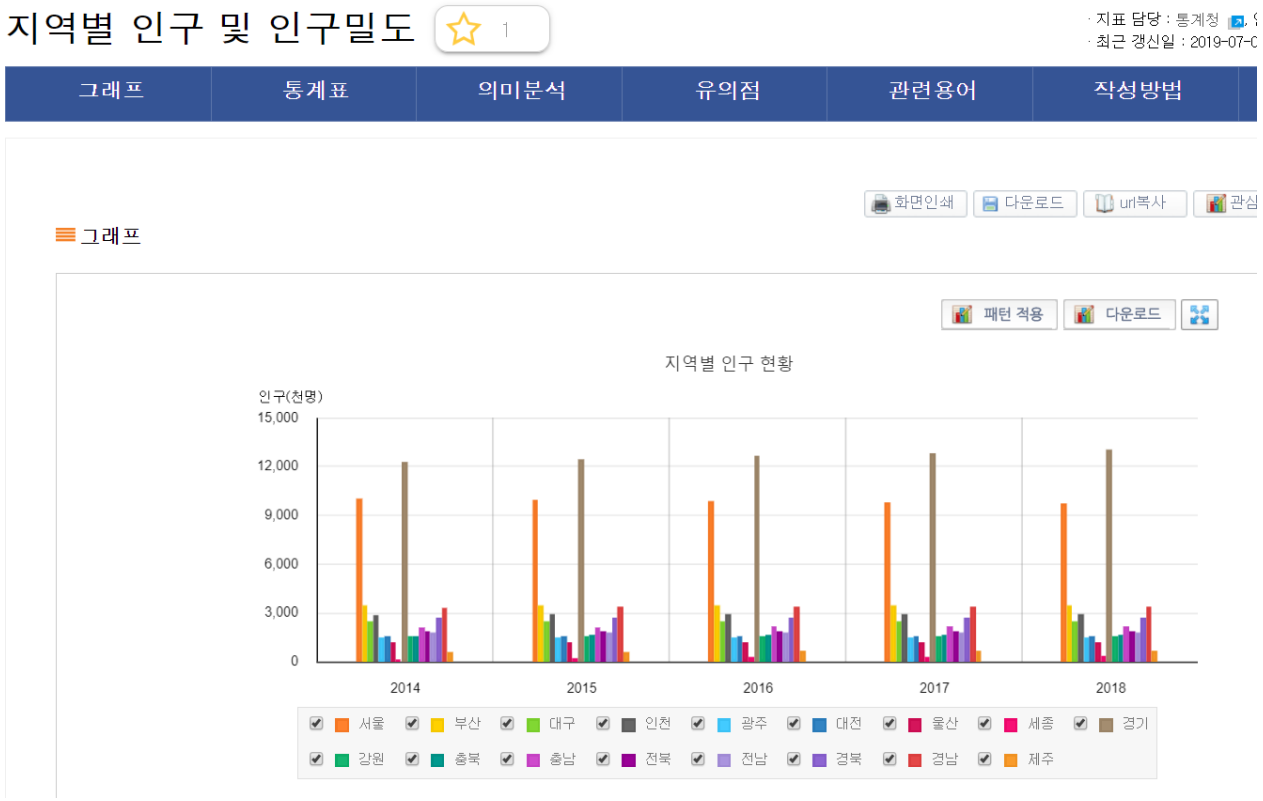
평균.현지기압.hPa.	최고.해면기압.hPa.	최저.해면기압.hPa.	평균.해면기압.hPa.	가조시간.hr.	합계.일조시간.hr.	X1시간.최다일사.시각.hhmi.	X1시간.최다일사량.MJ.m2.	합계.일사량.MJ.m2.
Average.Hydration.hPa.	Max.atmospheric pressure.hPa.	Min. Sea pressure hPa.	Average. Sea pressure hPa.	Autumn time.hr.	Total sunshine hours.hr.	Most Illustrative Time.hhmi.	Most Illustrative Time.hhmi.m2.	Total.insolation.MJ.m2
1003.6	1018.3	1012.4	1014.2	9.6	6.4	1200	1.13	6.17
1010.4	1023.6	1018.4	1021.2	9.6	9	1200	1.6	9.18
1006.9	1020.9	1014.8	1017.6	9.7	2.7	1200	1.22	4.69
1008.8	1021.2	1017.9	1019.7	9.7	8.7	1200	1.46	8.14

평균.전운량.1.10.	평균.중하층운량.1.10.	평균.지면온도..C.	최저.초상온도..C.	평균.5cm.지중온도..C.	평균.10cm.지중온도..C.	X0.5m.지중온도..C.	X1.0m.지중온도..C.	X1.5m.지중온도..C.
Average clouding volume 1.10.	Average.Middle Lower Clouds.1.10.	Average ground temperature.	Minimum portrait temperature	Average 5 cm Ground Temperature	Average 10 cm Ground Temperature	0.5m Ground temperature	1m Ground temperature	1.5m Ground temperature
2.1	1.9	0.8	4.3	0.3	0.2	2.6	6.3	9.2
0.3	0	0	5.4	0.1	0.2	2.6	6.2	9.1
4.6	2	0.1	5.1	0.2	0.1	2.6	6.1	8.9
0.9	0.5	0.2	6.4	0	0	2.6	6	8.8

01. Goal of project

Data detail

4. Additional data
- Population and population density by region – National indicator system
- longitude, latitude in Seoul, Busan city



시군구	경도	위도
Gangnam-gu	127.0664091	37.4959854
Gangdong-gu	127.1464824	37.5492077
Gangbuk-gu	127.0147158	37.6469954
Gangseo-gu	126.8226561	37.5657617
Gwanak-gu	126.9438071	37.4653993
Gwangjin-gu	127.0857528	37.5481445
Guro-gu	126.858121	37.4954856
Geumcheon-gu	126.9001546	37.4600969
Nowon-gu	127.0771201	37.655264
Dobong-gu	127.0317674	37.6658609
Dongdaemun-gu	127.0507003	37.5838012
dongjag-gu	126.9443073	37.4965037
Mapo-gu	126.9087803	37.5622906
Seodaemun-gu	126.9356665	37.5820369
Seocho-gu	127.0378103	37.4769528
Seongdong-gu	127.0409622	37.5506753
Seongbuk-gu	127.0232185	37.606991

Data preparation

02. Data preparation

Data Visualization tools

1. Data preprocessing
(excel) basic technique
(R studio) basic technique

2. ggplot2 package
(ggplot2) correlogram

(ggplot2) Scatter plot

3. R shiny package & Leaflet
(R shiny) Interactive visualization

(R shiny) Polygon mapping

(R shiny) Image, Text output

(Leaflet) + (R shiny)
Interactive Mapping

02. Data preparation

Data processing procedures

4 patient variables + 5 fine dust variables + 27 environmental variables + location information

= 40 variables in total

Data processing

1. Use **only average** value among minimum, maximum and average values
   (there is no big difference for maximum, minimum)
   (If there is no average value, use the maximum value)

2. New **air index(AQHI - Air Quality Health Index)** added to integrate fine dust information

$$AQHI = (\frac{1000}{10.4}) \times [(e^{0.000537 \times O_3} - 1) + (e^{0.000871 \times NO_2} - 1) + (e^{0.000487 \times PM_{2.5}} - 1)]$$

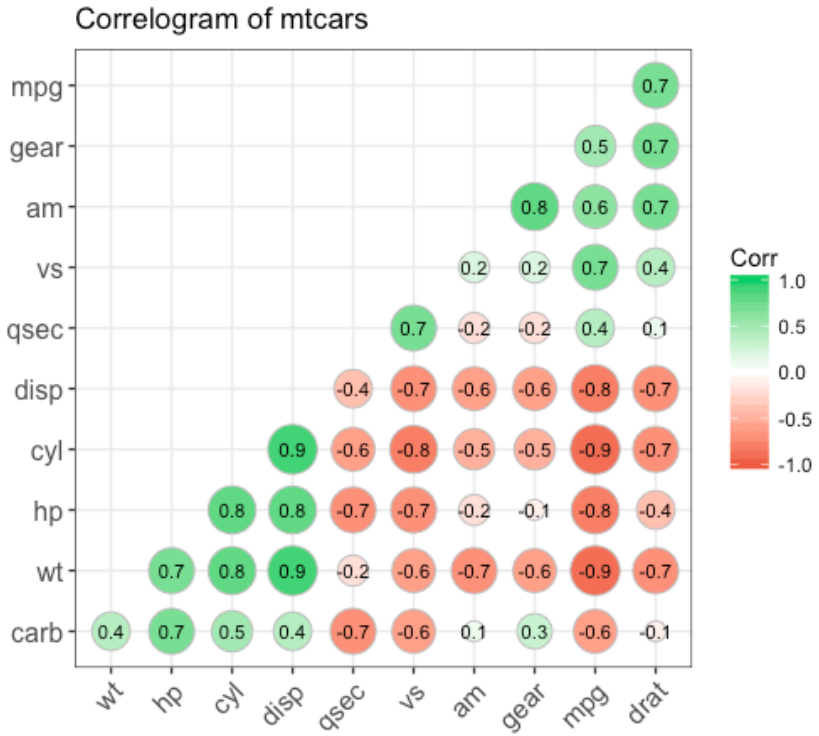
3. Use **Correlogram** to choose and analysis meaningful variables

02. Data preparation

Correlogram in R

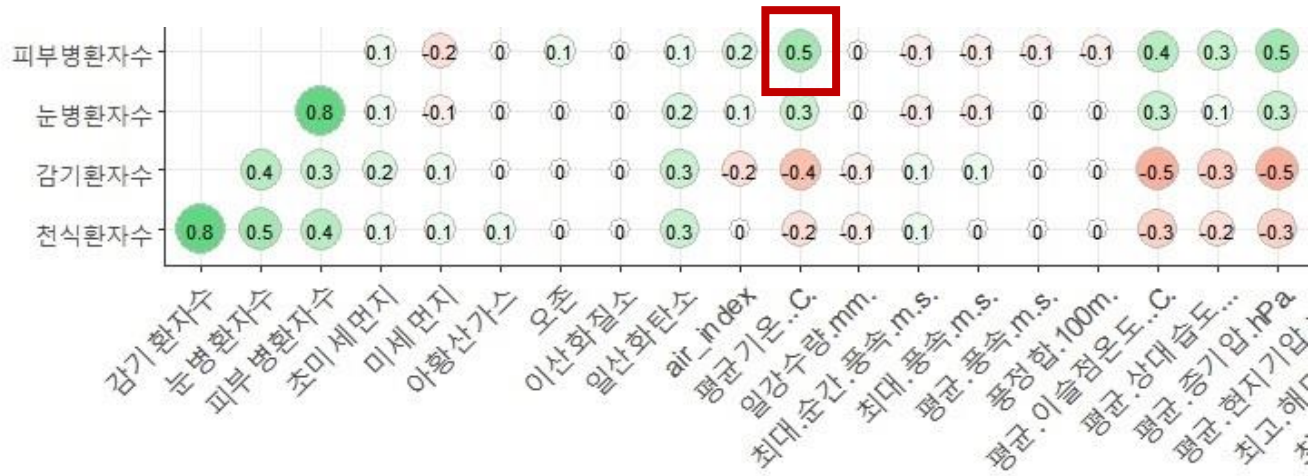
A correlogram or correlation matrix allows to analyze the relationship between each pair of numeric variables in a dataset.
Green = positive proportional relationship / Red = negative inverse relationship

→ **So we find relationship each variable, removed variable has not relationship.**



Correlogram in Seoul Data

- For example, skin disease and average temperature are 0.5. This means that the higher the temperature, the greater the number of skin disease patients.



02. Data preparation

Correlogram in R

After excluding weekends and Mondays, we use the data +1 and -1 day to see the delay

1. Correlogram after removing weekends and Mondays

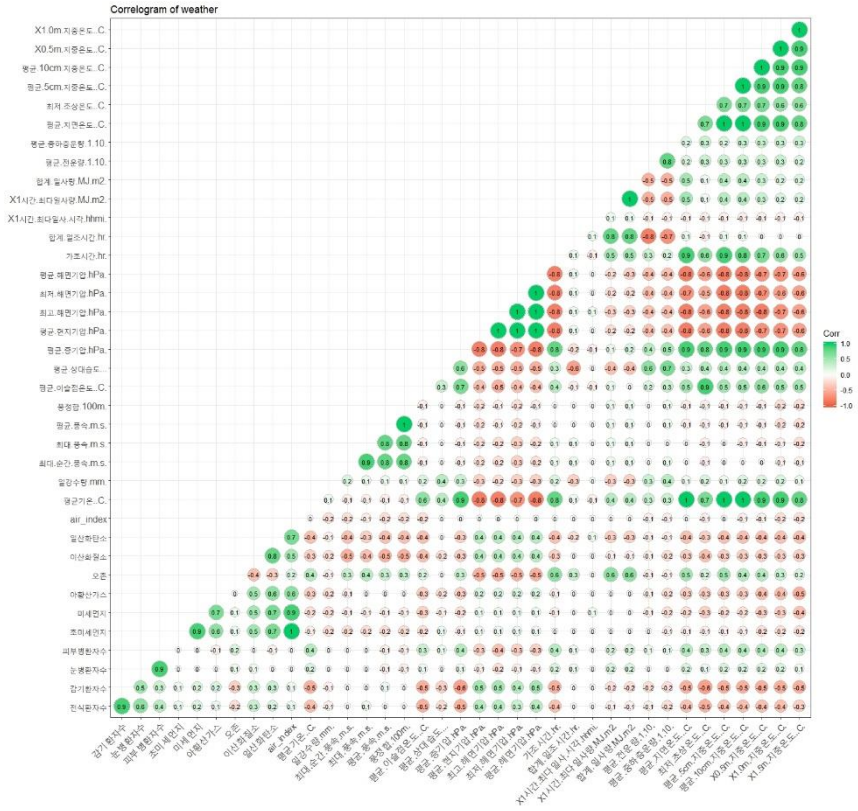
2. Correlogram after total date +1 (+delay)

3. Correlogram after total date -1 (-delay)

In case 3, the correlation becomes clear and the explanation for each variable is reasonably complete.

02. Data preparation

Correlogram in R



1.The higher the **temperature / relative humidity / sunshine time / average vapor pressure,**
-> The number of cold and asthma patients decreases
and the number of skin and eye diseases increases.

2. The higher the **local air pressure / sea level air pressure**
-> The number of colds and asthma patients increases
and the number of skin and eye patients decreases.

3. Among <u>fine dust</u>
-> **Nitrogen dioxide** has a positive proportional relationship to cold and asthma (low impact)
-> **Carbon monoxide** has a positive proportional relationship with cold, asthma (high impact)

02. Data preparation

Correlogram in R

4 kinds of patients number in 2014 ~ 2016

1. Cold
2. Asthma
3. Skin disease
4. Eye disease

10 Representative Environmental Variables

1. Average temperature
2. Dew point temperature
3. Relative humidity
4. Autumn time
5. Average vapor pressure
6. carbon monoxide
7. Nitrogen Dioxide
8. Sulfur dioxide
9. Local air pressure
10. air_index

02. Data preparation

Correlogram in R

We look closely at the relationship between **patient numbers** and **environmental variables.**
The following relationships and reasons were discussed. In addition, related news and information were collected and organized.

Examples)

1. cold-eye disease

Adenovirus, a cold virus, is the virus that causes eye disease.

2. Autumn time and cold / asthma
Sun Shower can Increase Immunity
Vitamin D deficiency due to lack of sunlight increases the risk of respiratory infections up to five times. Because without vitamin D, your immune system isn't working.

3. Pressure and cold / asthma
Barometric Pressure and Diaphragm Muscle Exercise At high pressures, the alveoli can supply more oxygen to the blood. When the air pressure is low, the oxygen concentration is low in the blood-> white blood cells do not function well, so the virus multiplies faster-> catches a cold.

4. Skin disease and humidity
If the humidity is high, the skin disease caused by the fungus will be well taken.

Visualization Result

03. Results

Data Visualization Results

We visualized all areas of Seoul and Busan.
In addition, these results allow the user to select and confirm by themselves.
We used ggplot2 & R shiny & Leaflet to visualize the following:
In addition, Tab-panel was used to group related data together.

1. **Polygon mapping** for all areas of Seoul and Busan

2. **Correlogram** check for all areas of Seoul and Busan

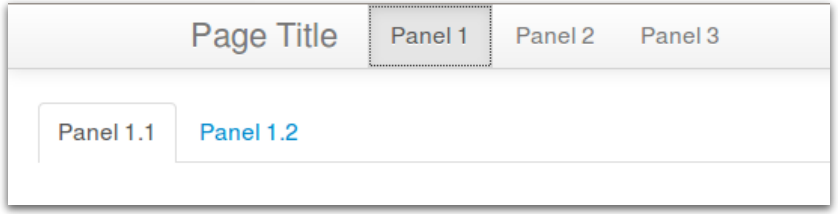
3. **Scatter plot** check for all areas of Seoul and Busan

4. **Interactive visualization** of all areas of Seoul and Busan
(scatter plot + mapping)

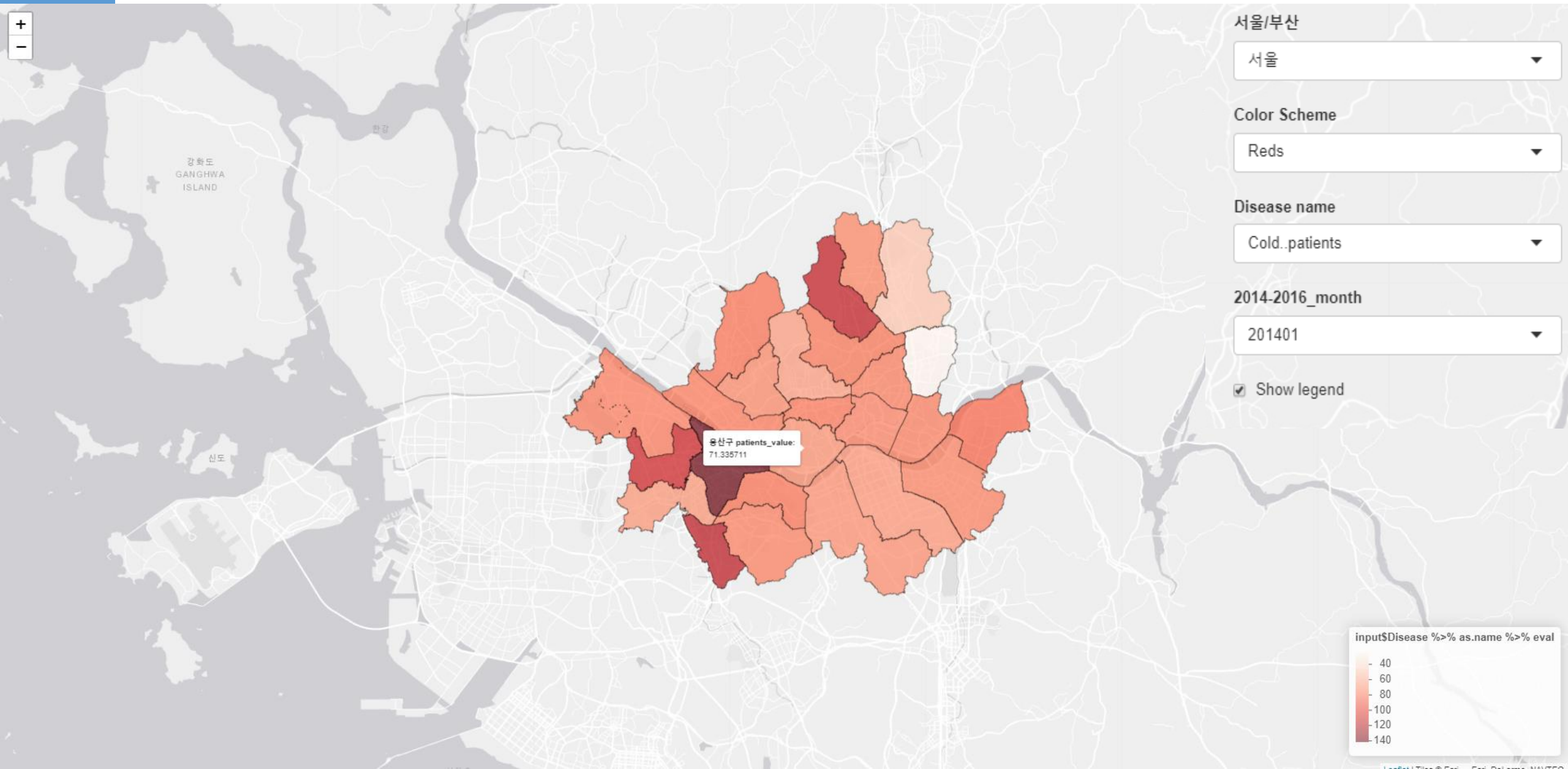
5. Visualize relevant **news and information** using **images output**

6. (Additional) **Scatter plot** for Seoul and Busan **each regions**.

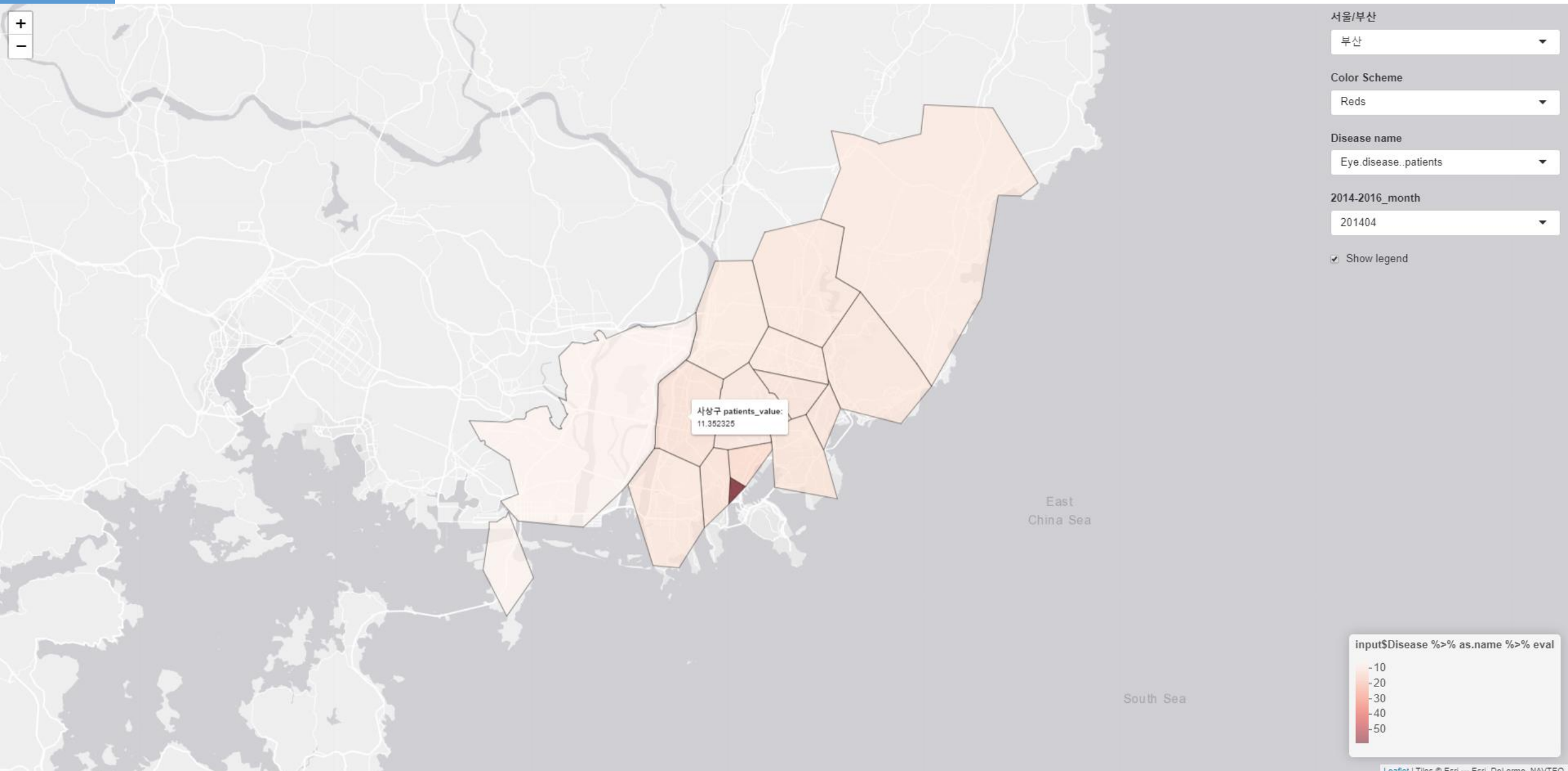
Tab-panel



03. Results

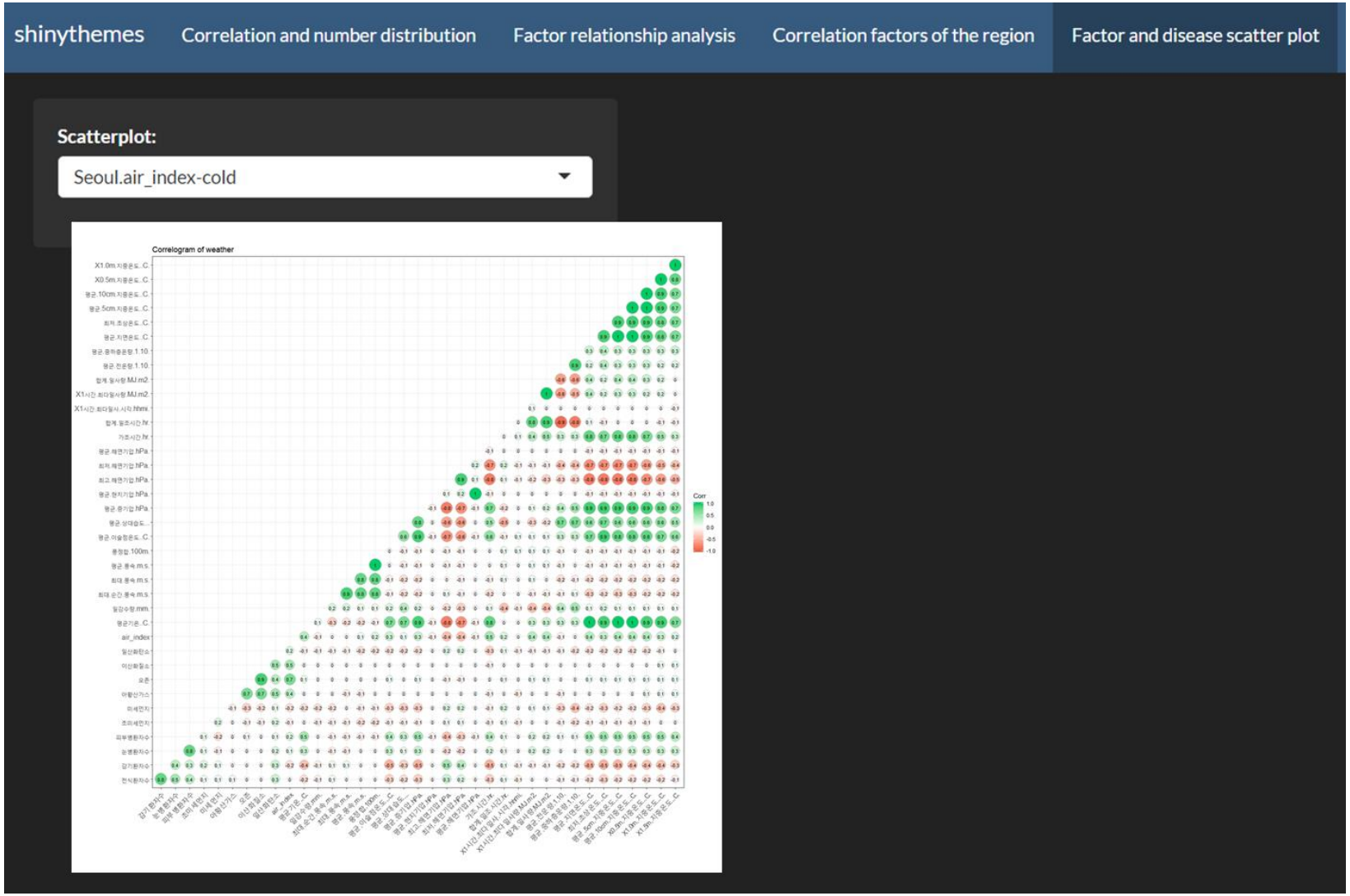

서울/부산
서울
Color Scheme
Reds
Disease name
Cold..patients
2014-2016_month
201401
Show legend
용산구 patients_value:
71.335711
input$Disease %>% as.name %>% eval
40
60
80
100
120
140
강화도
GANGHWA
ISLAND
한강
신도

03. Results


서울/부산
부산
Color Scheme
Reds
Disease name
Eye.disease..patients
2014-2016_month
201404
Show legend
사상구 patients_value:
11.352325
East
China Sea
South Sea
input$Disease %>% as.name %>% eval
10
20
30
40
50

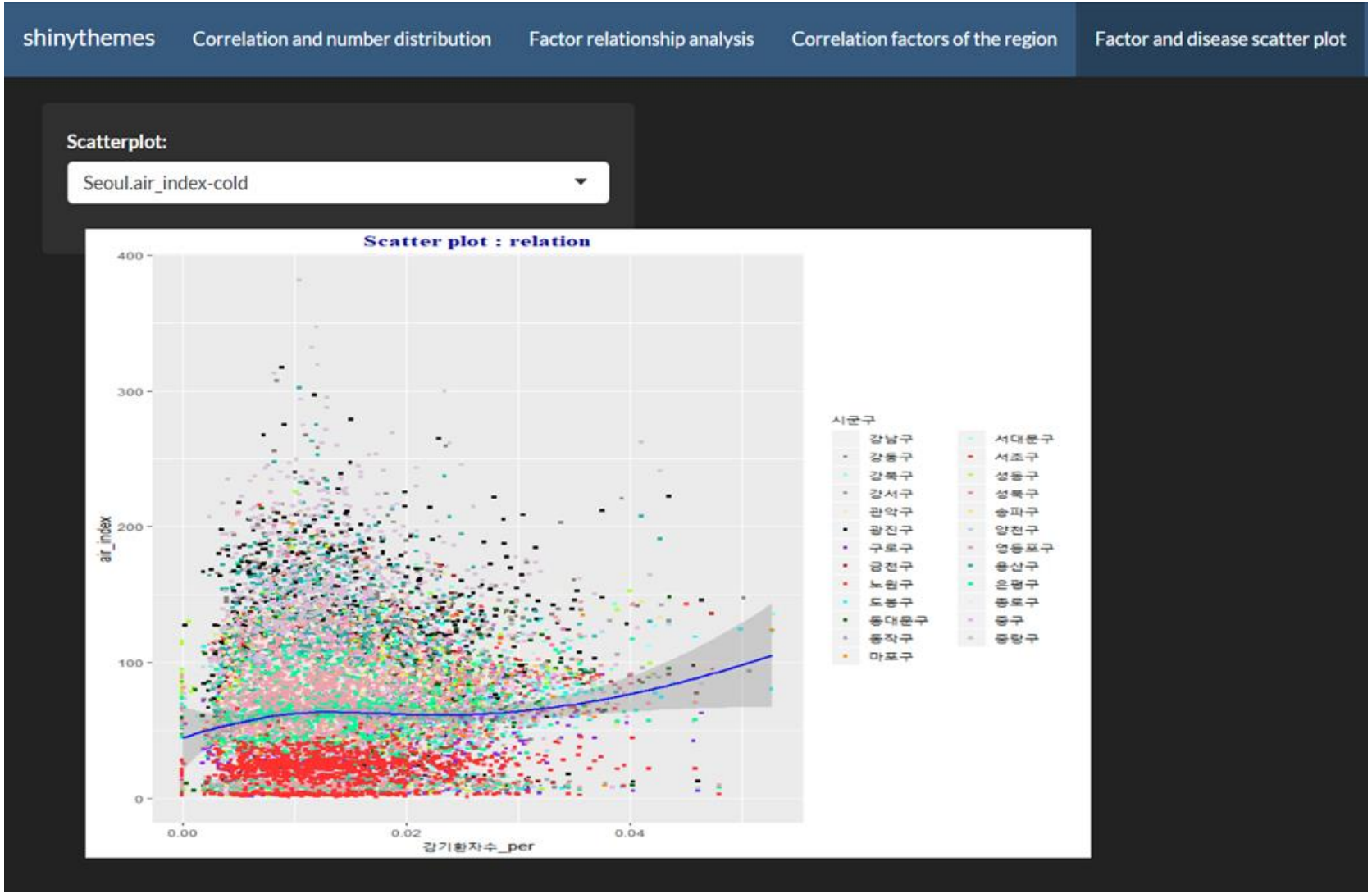
03. Results

(2) Correlogram



03. Results

(3) Scatter plot



03. Results

(4) Interactive visualization

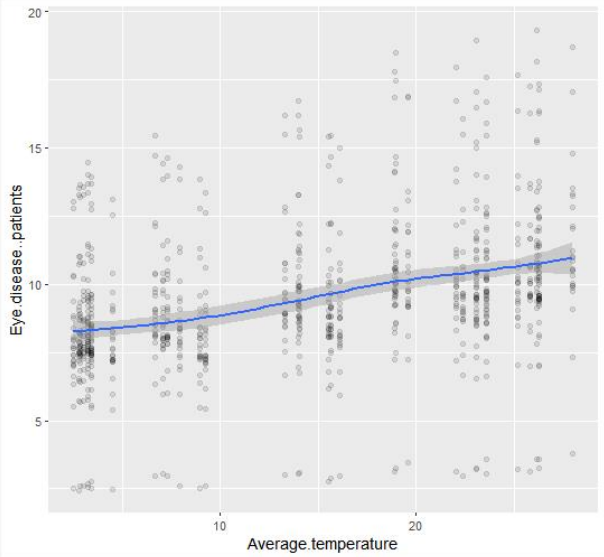
shinythemes | Correlation and number distribution | Factor relationship analysis | Correlation factors of the region | Factor and disease scatter plot

Diseases:

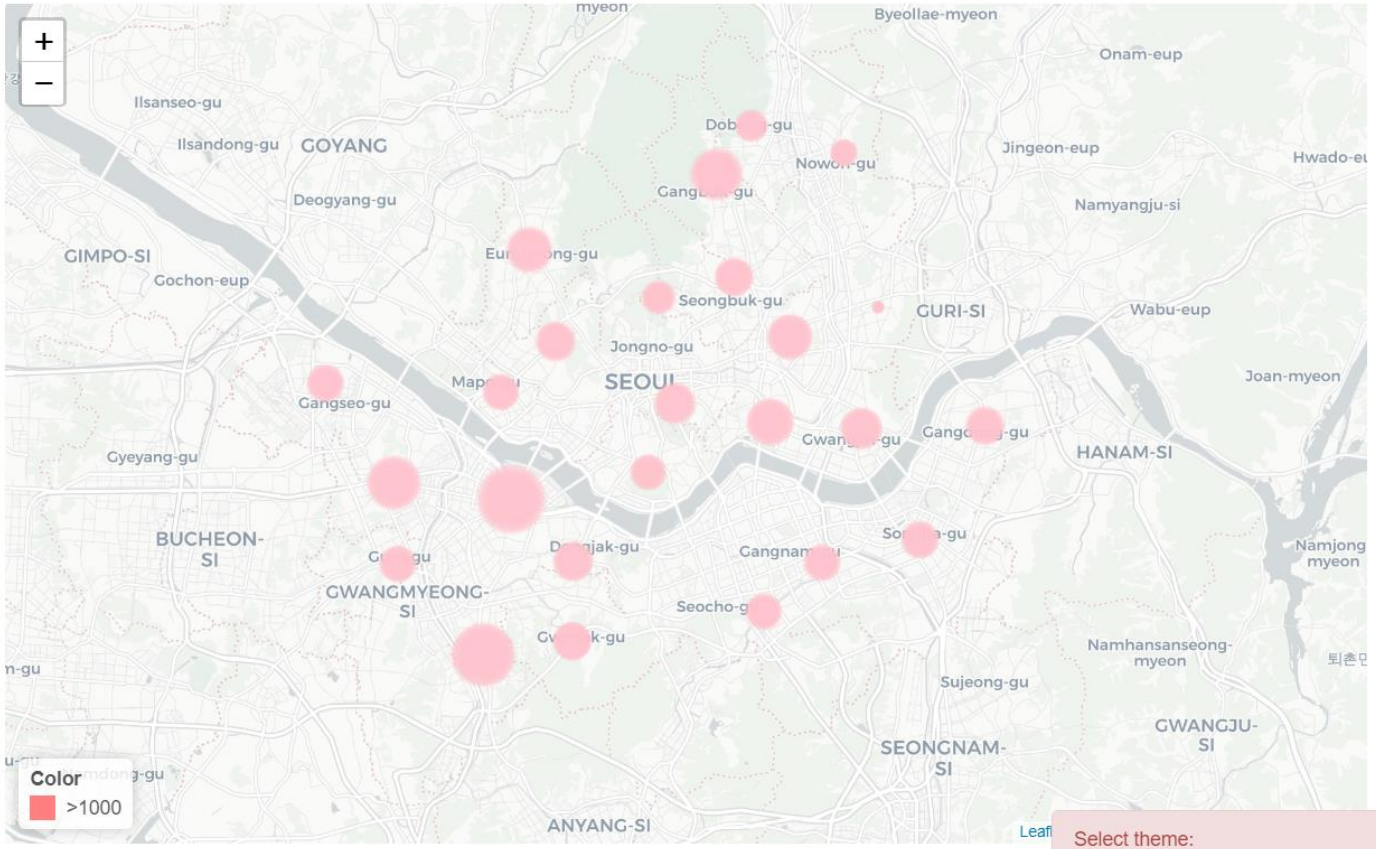
Eye disease Patient

Factor:

Average Temperature



Correlation = 0.352. Note: If the relationship between the two variables is not linear, the correlation coefficient will not be meaningful.



03. Results

(5) Image output

shinythemes | Correlation and number distribution | Factor relationship analysis | Correlation factors of the region | Factor and disease scatter plot

picture:

- Temperature-Cold
- Temperature-Asthma
- Temperature-Skin
- Temperature-Eye
- Suntime-Cold
- Sumtime-Asthma
- Suntime-Skin
- Suntime-Eye
- Airpressure-Asthma
- Airpressure-Skin
- Dust-Cold
- Dust-Asthma
- Ozone-Cold
- Ozone-Asthma
- Ozone-Skin
- Ozone-Eye
- CO-Cold
- CO-Asthma
- Humidity-Cold
- Humidity-Asthma
- Humidity-Skin
- Humidity-Eye

Select theme: superhero

독감과 **감기 바이러스**는 일종의 단백질 입자인데

실내 습도가 낮아질수록 이 바이러스들이 공기 중에 둥둥 떠다니며 호흡기에 침투해 사람들은 감염시킵니다.

→ 환기와 적정 습도를 유지하는 것이 중요합니다.

[출처: LG blog]

03. Results

(6) Scatter plot in each regions

select region & city:

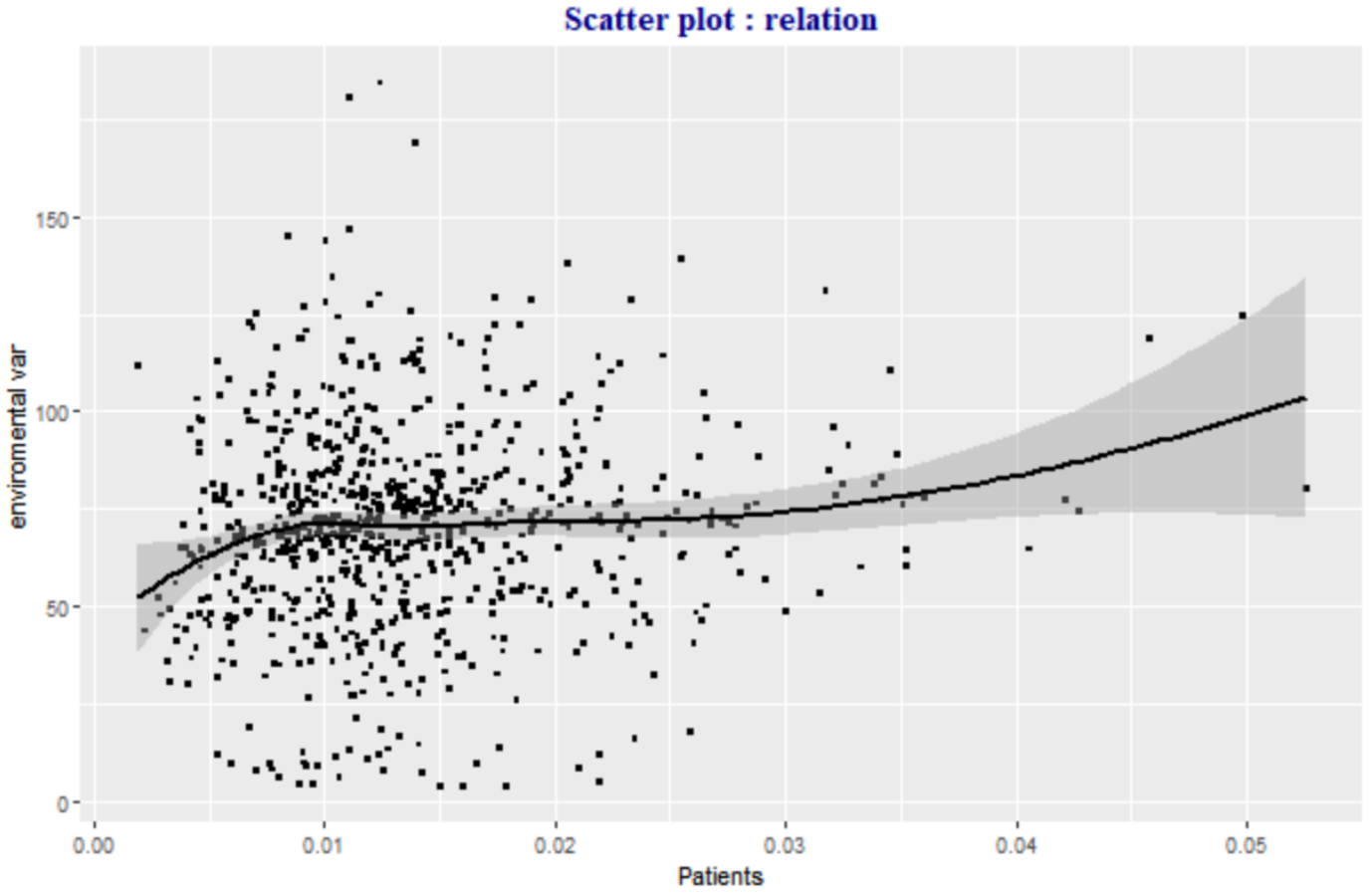
Seoul_Gangnam-gu

Patient:

Cold.patients_per

environmental var:

air_index



03. Results

Final version in R shiny

Project final version released

1. **Polygon mapping** for all areas of Seoul and Busan
2. **Correlogram** check for all areas of Seoul and Busan
3. **Scatter plot** check for all areas of Seoul and Busan
4. **Interactive visualization** of all areas of Seoul and Busan (scatter plot + mapping)
5. Visualize relevant **news and information** using **images output**

Tab-panel

6. (Additional) **Scatter plot** for Seoul and Busan **each regions**.

Actual Demo presentation

Discussions

04. Discussion

Stakeholder's opinions

1. Boas Medical Doctor

Purpose: We thought that the most relevant place for disease-related information at Handong University was Boas Medical, which actually cares about students. Therefore, We decided to show the results of the project to doctor and seek advice on improvements and future applications.

Date: Monday, December 2, 2019

Subject: Handong Boas Medical Doctor

Method: Schedule a Meeting by Mail, Present Project Results, and Ask for Comments

04. Discussion

Stakeholder's opinions

1. Boas Medical Doctor

conclusion:

1) One of the strengths is that it uses big data to analyze one of the factors that affect disease. I think this would be very good to alert patients.

2) Another strength is that doctors can use it to communicate medical information to patients. When explaining to patients through visualization, they will be more understanding.

3) An application is to provide people with medical knowledge by using touchscreen panels in the patient reception area.

4) This is beyond the traditional method of delivering medical information through photographs, and it is easy to understand that users can interactively touch and visualize the change of information, graphs and correlations.

04. Discussion

Stakeholder's opinions

2. Professor of Life Sciences (Professor Ahn Tae-jin)

Purpose: Because it is a project that deals with the relationship between disease and fine dust, we thought that the opinion of experts in life sciences is necessary. In particular, We decided to consult with Professor Biodata-Lab, who understands both biology and ICT convergence.

Date: Wednesday, December 4, 2019

Subject: Prof. Tae Jin Ahn of Biodata-Lab, Life Sciences

Method: Schedule a Meeting by Mail, Present Project Results, and Ask for Comments

04. Discussion

Stakeholder's opinions

2. Professor of Life Sciences (Professor Ahn Tae-jin)

conclusion:

As the explanation of the conclusions is weak, a general explanation should be included based on statistical analysis. For example, if you find that carbon monoxide is associated with colds and asthma, you should use a regression analysis to establish a null hypothesis and show whether it is statistically significant whether carbon monoxide is actually associated with colds or asthma.

04. Discussion

Stakeholder's opinions

3. Handong student’s comments

Purpose: We also asked for feedback from the same student to see if the project results were clear and if there were any improvements that could be made from the presentation.

Date: Tuesday, December 3, 2019

Subject: Handong University Student 2 Group

Method: After meeting personally, ask for feedback on project results

04. Discussion

Stakeholder's opinions

3. Handong student’s comments

conclusion:

Group 1: Lee 16, Gang 18

> Correlogram is a good visualization, but it's not very accessible. Therefore, it would be better to set up stakeholders as experts and journalists who deal with diseases and environmental issues.

Group 2: Han 16

> The theme setting is excellent but the visualization is less readable. Additionally, because the overall color is gray and the parts represented by dots are not very noticeable, it is better to visualize it in a different color.

04. Discussion

Data visualization conclusion & thinking

We did visualization to show the correlation between fine dust and disease.

And we could actually see some connection.

I wish our team's visualization would be more complementary to make it understand easier for people who have never seen it before

Thank you