Milestone Three

Week Seven - Project Two

Ross L. Kim-Schreck

DSC680 Applied Data Science

Professor Iranitalab

2024.07.21

Causes of Air Pollution in the Republic of Korea

Business Problem

Air pollution has a major effect on every facet of society. It affects the economy in that it causes governments to spend more on energy. Fine dust can cause equipment to malfunction rendering it in need of repair or replacement. Fine dust is obviously very bad for one's health; it causes people illness and has long-term repercussions regarding health. This, in turn, affects businesses as an unhealthy workforce is an ineffective one. Air pollution, above all, has a massive effect on the environment; it damages vegetation such as trees and crops in a variety of ways. It causes fewer yields and reduces crop growth. If the government could implement laws and regulations to reduce the amount of air pollution, it would aid in the restoration of the economy, assist businesses in their related problems, improve the overall health of the population, and most significantly improve the environment and ecosystem.

Background / History

Since the Republic of Korea started industrialization in the early 20th century, economic output grew exponentially. Prior to this period, the Empire of Japan occupied the peninsula from 1910 until the end of world war two. During this time, Japan built up Korea's infrastructure including roads and railways. In 1945 when Korea was partitioned into North and South, Japan had been developing the North as an industrial zone due to its lack of fertile soil for agriculture; the South was developed as an agrarian economy. The government of the Republic of Korea (South Korea) which I will refer to as 'the ROK' decided to take an economy-first initiative. While technically a liberal democracy, the government was, in fact, a military dictatorship. During the post-war period, Korea's population also skyrocketed to nearly 50 million people. Korea began extensive construction projects to accommodate its population. Nearly one half of the population is concentrated in the Seoul metropolitan area. Despite having adequate public transportation systems, most Koreans prefer to use private forms of transportation. There is an ever-growing environmental problem of fine dust. Much like smog, this has the effect of damaging the air quality. Koreans generally blame neighboring China as being the primary cause of this environmental problem. I believe that the majority of fine dust does not in fact come from China but rather from car exhaust due to the unbalanced number of vehicles in the city.

Data Explanation

After getting my data from Kaggle and KOSIS (Korean Statistical Information Service), which is the Korean government's official website. I initially used two datasets: 'Air_Quality_Seoul.csv' which contains data of SO2 (sulfur dioxide), NO2 (nitrogen dioxide), O3 (ground-level Ozone), CO (carbon monoxide), PM10 (particles with a diameter of 10 micrometers or less, and PM2.5 (particles with a diameter of 2.5 or smaller) and 'cities_air_quality_water_pollution.18-10-2021.csv' which contains data of major cities worldwide. After performing EDA, I came to realize that these datasets are inadequate and too unorganized to perform any meaningful transformations. The first dataset, 'Air_Quality_Seoul.csv', has columns labeled for time-series data formats, but the data is not in sequential order. Reorganizing this dataset would require significant time and effort. The second dataset, 'cities_air_quality_water_pollution.18-10-2021.csv', lacks details of the types of air pollution. It only includes the amount of air pollution collectively. I added three datasets from KOSIS: 03 - 'measurement _info.csv', 04 - 'measurement _item _info.csv', and 05 - 'measurement _station _info.csv'. Dataset three includes measurement dates, locations, and average amounts of pollution. Dataset four includes rates of chemical composition in the atmosphere. Dataset five includes station names, locations, and coordinates.

In milestone three, I added 11 new datasets which include the following: dataset six (106_DT_106N_03_0200045_20240703151242.csv) - timeline of records from 2010 to 2023, dataset seven (106_ DT_ 106N_ 03_ 0200076 _20240703151123.csv) - timeline of records from 2010 to 2023 and regions in the ROK, dataset eight (Measurement _summary.csv) - measurement dates, locations, PM2.5, PM10, SO2, NO2, O3, and CO, dataset nine (National _Air _Pollutant _Emissions _20240703151513.csv) - timeline of records from 1999 to 2021 and locations, dataset 10 (seoul _air

_1988_2021.csv) - measurement dates, locations, PM2.5, PM10, SO2, NO2, O3, and CO, dataset 11 (SeoulHourlyAvgAirPollution.csv) - locations, PM2.5, and PM10, dataset 12 (116 _DT _MLTM _962 _20240712093545.csv) - statistics on type of streets in the Seoul metropolitan area, dataset 13 (Average _Daily_Traffic_by_Road_and_Vehicle_Types_20240712093400.csv) - statistics on the daily traffic by road and vehicle types, dataset 14 (Yearly _ADT_20240712093451.csv) - road types from 2003 to 2023, dataset 15 (input_data_spatial_panel.csv) - statistics on weather patterns and chemical composition of the air, and dataset 16 (car_fuel_emissions.csv) - statistics on car fuel emissions in the ROK *(appendix 2.1) (appendix 2.2) (appendix 2.3) (appendix 2.4)*.

Methods

My target variable is PM2.5 which is particles that are less than 2.5 micrometers in diameter. I will be comparing this variable to the following: SO2 (sulfur dioxide), NO2 (nitrogen dioxide), O3 (o-zone or trioxygen), CO (carbon monoxide), and PM10 (particles that are between 2.5 and 10 micrometers in diameter). I tested my datasets using linear and logistic regression models and gradient boosting classifiers. I initially tested for accuracy and consistency between the first two datasets by comparing RMSE and R2 scores. I also used linear regression for time-series data analysis. Unfortunately the datasets were too inadequate to build models that could accurately forecast weather and other environmental patterns based on historical data. In the next milestone, I plan to use these techniques on data containing economic trends. The first two datasets returned very high RMSE and low R2 scores. According to the box plot, dataset three contains too many outliers *(appendix 1.6)*. Also due to the structure of the data, I was unable to run a linear regression model. I ran the same regression model on the fourth and fifth datasets and returned much more realistic results. The fourth dataset returned an RMSE of 6.2 and R2 of 0.09 which is far from ideal, but can be adjusted with additional transformations *(appendix 1.8)*. The fifth dataset returned an RMSE of 0.07 and R2 of -0.4 which is also not ideal, but can possibly be adjusted with additional transformations.

For milestone three, I started off by transforming datasets six to 11. Dataset 11 is in Korean, but I just left it due to time constraints; I figured I could explain it in English in the presentation. I also changed the column names of each dataset and confirmed the data types for visualization. I was able to visualize datasets six, seven, and eight. I went through the same process with datasets 12, 13, and 14. For dataset 12, I changed the types to float for visualization. I was able to visualize dataset 12 after type conversion in a number of ways. I then went through the same process with dataset 15. After creating a boxplot of dataset 15, I realized there is a large column of unnamed data points which was skewing the overall boxplot. I removed the column ‘unnamed’ and created an additional boxplot; this made observation of outliers clearer. I replaced non values with the mode for every column and then ran a linear regression model on dataset 15 to return an RMSE of 2.4 and R2 of 1.0 with the target variable being ID2 as a test. I then ran the same model on humidity levels to return an RMSE of 14.3 and R2 of 0.5 which are much better than the previous results of linear regression on the first two datasets. I assume any abnormalities are due to the large number of unnamed data points in the removed column. For dataset 16, I used the library joblib for pickler objects as dataset 16 had columns that were unpicklable. I created functions 'async' and 'start methods' for multiprocessing and pickle serialization. I set the 'loky_pickler' to use pickle serialization from stdlib and to not pass the desired function 'func_async' as it is not picklable. I then assigned the relevant variables for the gradient boosting classifier that returned an accuracy of 0.944 which is quite accurate; however, with more time, I need to validate this process. It seems that the values are ambiguous, and this requires further investigation.

Analysis

The bar chart of the first dataset shows that carbon emissions are overall far greater than other gasses *(appendix 1.1)*. The second dataset returned results for the amount of air pollution by city. According to the area chart, the United States and Europe have the cleanest air *(appendix 1.2)* *(appendix 1.3)*. This was confirmed in the tree map and bubble chart of the same data *(appendix 1.4) (appendix 1.5).* For the fourth dataset, the degrees of air quality were visualized as a line chart. It shows that over time, the air quality increasingly worsened *(appendix 1.9)*. The fourth dataset also supported the first result in a bar graph displaying the degrees of air quality *(appendix 1.11).* The fifth dataset did not return any visualizations of statistical significance. It only contained coordinates of locations where measurements were taken. I will not use this dataset in further analysis.

For milestone three, I created a number of visualizations for datasets six to 16. Dataset eight shows very few outliers and seems to be a reliable source *(appendix 2.09)*. According to the bar chart *(appendix 2.10)*. PM2.5 is significantly higher than PM10 and other emissions. This analysis solidifies the already known fact that fine dust is overly abundant in the Seoul metropolitan area. *(appendix 2.14)* shows the abundance of small particles (PM2.5) in different regions of Seoul with up to 800,000 in some areas. This unit is ambiguous as fine dust is measured in μg/m³ (micrograms per cubic meter). I can only assume this is high when compared to other regions. Dataset eight also has few outliers for PM2.5 *(appendix 2.16)*. After creating a bubble plot of dataset ten, I realized that the N/A column is too large rendering this dataset useless. Cleaning it would take too much time *(appendix 2.18)*. *(appendix 2.19)* and *(appendix 2.20)* show the amount of fine dust per hour in all regions of Seoul. The peaks are in the downtown areas at nearly 7,500. Once again this measurement is ambiguous as fine dust is measured in μg/m³. I was unable to find a conversion method for these two datasets. Dataset 12 contains statistics on the amounts and types of road and highways in Seoul from 2005 to 2023. The number of roads has significantly increased since 2005 *(appendix 2.21)*. Dataset 13 contains additional statistics on the number of roads *(appendix 2.23)*. Dataset 15, which contains data on the atmospheric conditions and composition, shows relatively consistent data with few outliers *(appendix 2.24) (appendix 2.25)*. Dataset 16 contains statistics on the amount of vehicle emissions. A line graph demonstrates the steady increase from 1999 to 2014 *(appendix 2.26)*.

Conclusion

I can conclude that the air quality in the Republic of Korea is significantly lower than in other countries. I can also conclude that the rate of carbon emissions is far greater than other gasses worldwide. Due to the lack of specific contents of all 15 datasets, the reasons for these rates are still yet inconclusive. Further analysis with additional datasets is required. The amount of PM2.5 compared to PM10 and other emissions is significantly higher and, according to the data, is steadily increasing. According to Scientific American Inc., 52 percent of fine dust comes from domestic factories while 34 percent comes from western China. I question these figures as I think fine dust from vehicle exhaust is a major factor. Traffic has worsened and the number of roads have increased correlating with the continuing increase in fine dust in the Seoul metropolitan area.

Assumptions

It was assumed that the main culprit of the fine-dust problem in Korea is due to the large number of vehicles, particularly individuals' personal modes of transport. It is also assumed that China's pollution does not play a key role in the low air quality in Korea. The greatest emitter is Koreans' personal methods of transportation. I will need to consider other metrics to confirm these assumptions. As stated above, an estimated 34 percent of fine dust comes from China while 34 percent comes from domestic factories. I assume these figures to be wrong. Based on my research, a significant amount of fine dust likely comes from vehicle exhaust. This was demonstrated in my graph showing vehicle emissions *(appendix 2.26)*.

Limitations

In the previous datasets, gas emitters were not included. I will need to find additional data containing statistics on the number of vehicles in operation in both the ROK and China. I will also need to find similar data on other neighboring countries such as Japan.

Several datasets only include statistics on a one to two year timespan. In order to conduct the research required, I will likely need to obtain datasets that cover the entire modern history of the Republic of Korea. This would span from the end of the Korean war until the present day.

I was unable to find datasets that cover the ideal timespan. The datasets that spanned the furthest back were to 1993. I was also unable to conduct similar research on other regions such as China and Japan. The quality and organization of many datasets were below par. I had to undergo further transformations and for some datasets, cleaning the data would consume too much time; thus, I did not use certain datasets.

Challenges

There doesn't seem to be complete datasets with metrics specific to this topic. Finding additional datasets containing type of emitters and the extent to which they affect the environment could be difficult. I will need to find datasets with traffic, weather, degree of industrialization, etc. to conclude this research. After obtaining these additional datasets, I will need to transform if necessary, test, and make comparisons.

I was still unable to find datasets containing data specific to or relating to the main emitters. Despite being able to find correlations among traffic volume, number of roads constructed, and the amount of fine dust, I need other metrics in order to further verify this. One of the biggest challenges was finding the necessary data for this research. There are far more metrics that should be considered in order to conduct this type of research. My data sources were inadequate. I was also unable to find datasets from reliable Chinese sources such as the government of China.

Future Uses / Additional Applications

I have built several models that have specific inputs. I could apply this to any region as long as I have the proper inputs for those regions, particularly regions with abnormally high rates of air pollution. I also hope to apply certain models used in this study to water quality in various regions. I used gradient boosting classifiers and other regression models in this milestone. I also collected data on the amounts of sulfur dioxide and nitrogen dioxide of different regions as these metrics seem to be abundantly available. I would need to see what metrics are available in other datasets for other regions besides the ROK.

Recommendations / Implementation

I used other methods to test the accuracy of the data for the datasets that I had available. If I had access to more time-series datasets, I would apply predictive modeling techniques such as linear and logistic regression and categorical grouping of data by using ARIMA, K-Means clustering, and PCA. I would recommend finding metrics pertaining to the political and social aspects of the topic as I think politics plays a major role. Other metrics to consider would be how rapidly regions have developed and industrialized, GDP of OECD nations, other economic indicators, other climatic and environmental factors, terrain types of different regions, and demographic considerations such as population densities and urbanization.

Ethical Assessment

The origin of the data should be considered. Data-producing entities could have skewed data sources. One reason for this is that certain aspects could be emphasized or on the contrary, omitted for that particular entity's benefit. I could make a comparison with the issue of global warming and how particular companies and industries manipulate scientific data for their own political or economic motives.

The data could be affected by a number of factors considering the subject matter. There are many ways in which this could become an issue in this research. There are methods that could be used to tackle the problem that would have the greatest effect on certain sectors of industry but minimal effect on the population or environment. The question is, how ethical is it to prioritize in such ways?

I think that I should choose datasets that allow me to take an objective approach in this research. Based on my observations, the datasets used in this research were not skewed either accidentally or deliberately. As per the above mention of taking political and social aspects into consideration, I think that ethical issues could be addressed. I do think that politics and the economy are intertwined, and because of this, ethical protocol is sometimes not followed, especially by political entities. Governments like that of the Republic of Korea would likely skew data in order to avoid sacrificing its economic growth.

References

2021, 2024.07.07, National Air Pollutant Emissions

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0100007&conn_path=I2&language=en

2021, 2024.07.07, Particulate Matter (PM-10) Emissions

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_99_2300011&conn_path=I2&language=en

2021, 2024.07.07, Particulate Matter (PM-2.5) Emissions

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_99_2300020&conn_path=I2&language=en

2023.12, 2024.07.07, Degree of Air Pollution by Measurement Network (Fine dust(PM2.5))

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0200176&conn_path=I2&language=en

2023.12, 2024.07.07, FineDust(PM2.5) by Month and City

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0200145&conn_path=I2&language=en

2023.12, 2024.07.07, Degree of Air Pollution by Measurement Network (Fine dust(PM10))

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0200076&conn_path=I2&language=en

2023.12, 2024.07.07, FineDust(PM10) by Month and City

https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0200045&conn_path=I2&language=en

Attribution 4.0 International (CC BY 4.0)

https://www.kaggle.com/datasets/joebeachcapital/world-air-quality/data

SaveEcoBot and EcoCity

https://www.kaggle.com/code/vbmokin/air-quality-city-daily-data-forecasting

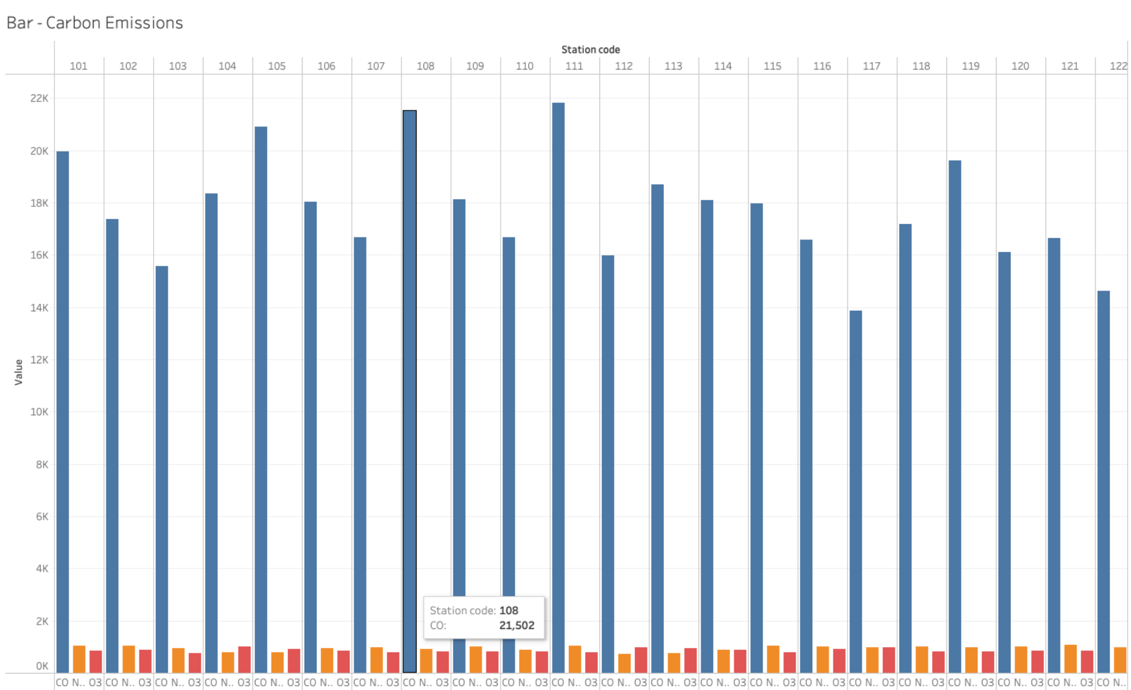
JUN, MYUNG-JIN (2022). Input_Data_Spatial_Panel. figshare. Dataset. https://doi.org/10.6084/m9.figshare.21699698.v1

2024 Scientific American, A Division of Springer Nature America Inc.

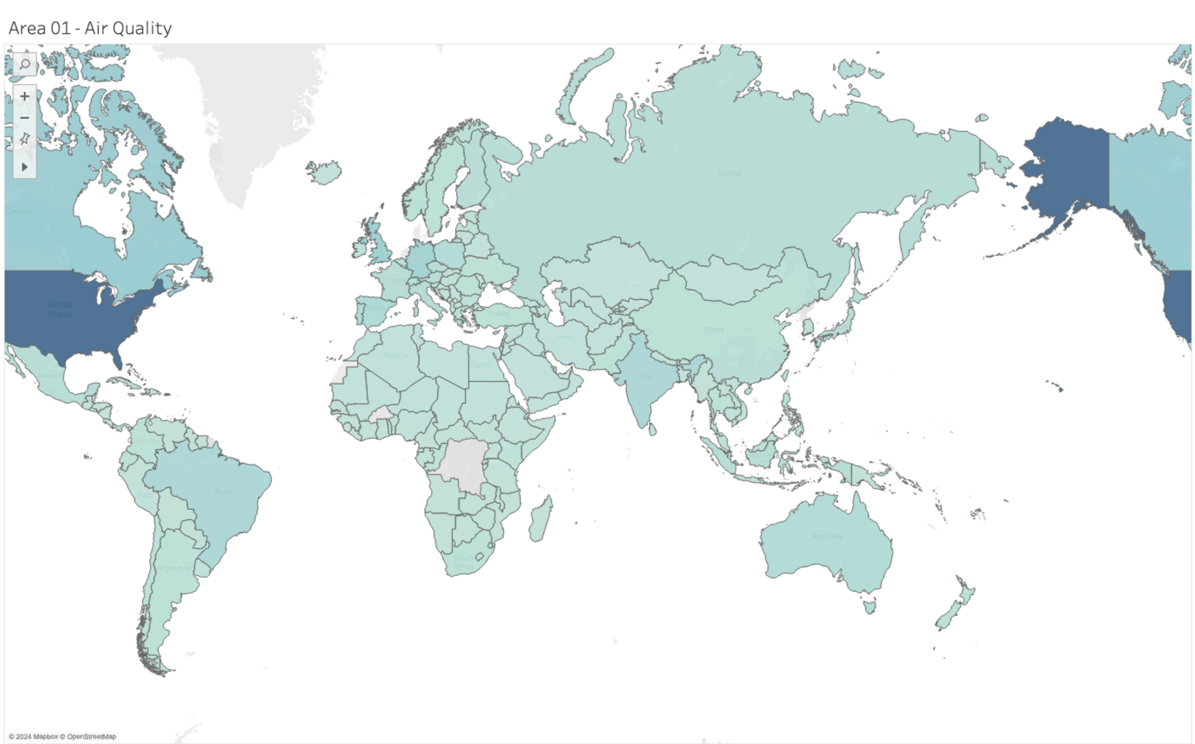
https://www.scientificamerican.com/article/what-air-pollution-in-south-korea-can-teach-the-world-about-misinformation/

Appendix

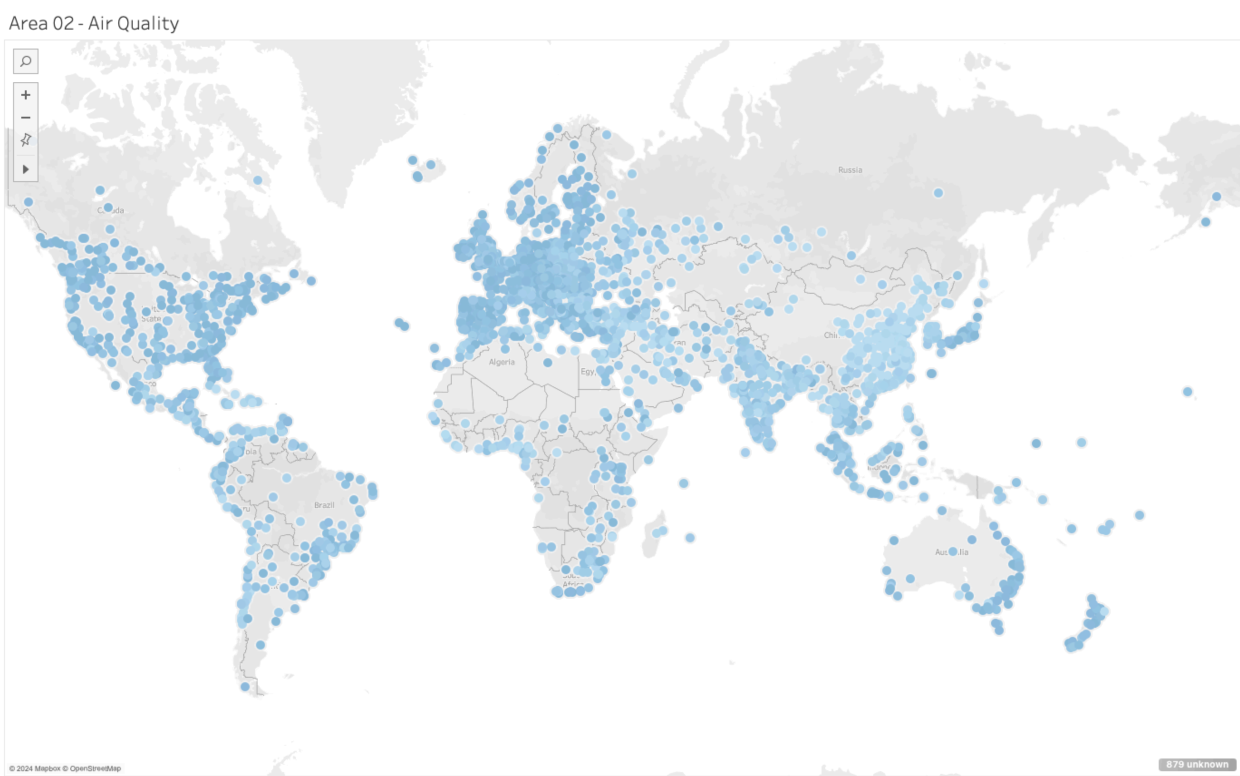
01.01



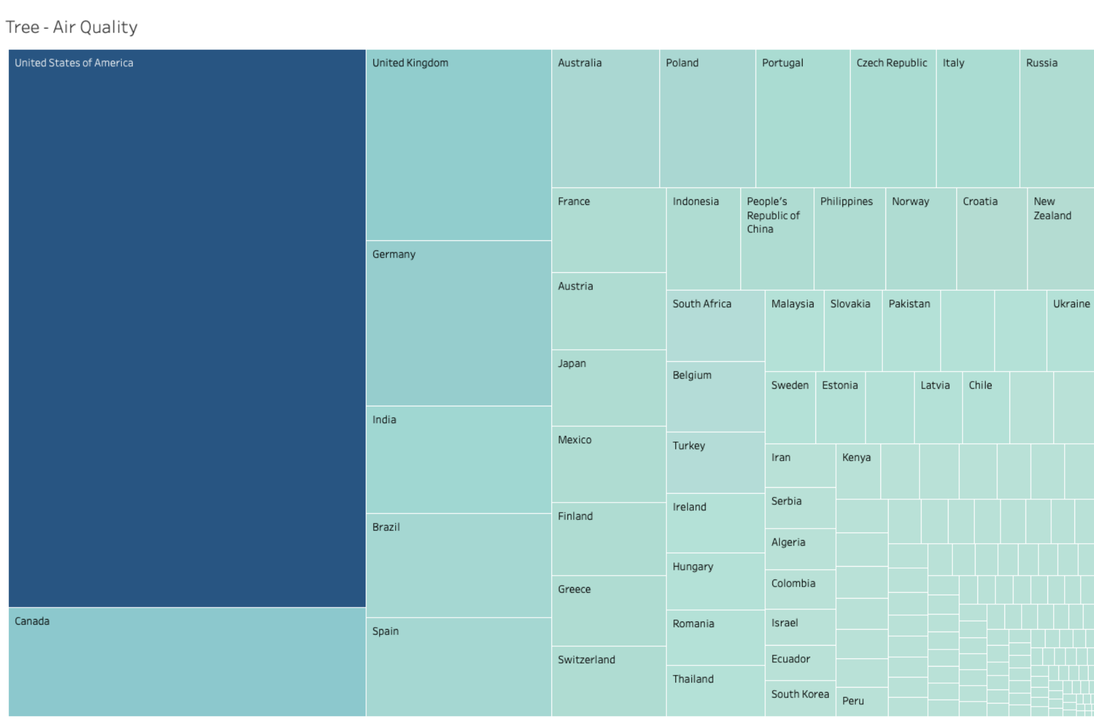
01.02



01.03

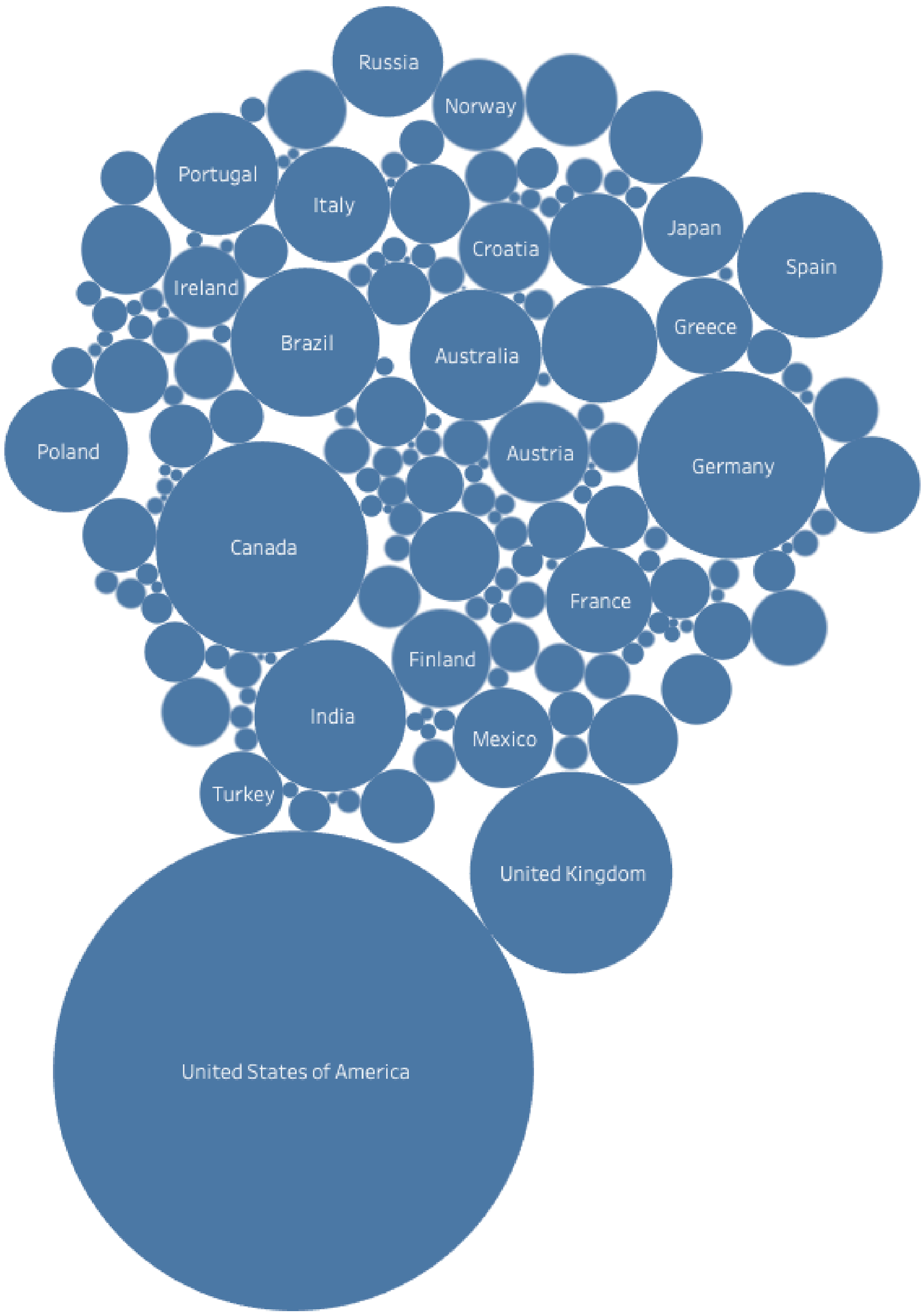

Area 02 - Air Quality
Russia
Algeria
Brazil
© 2024 Mapbox © OpenStreetMap
879 unknown

01.04

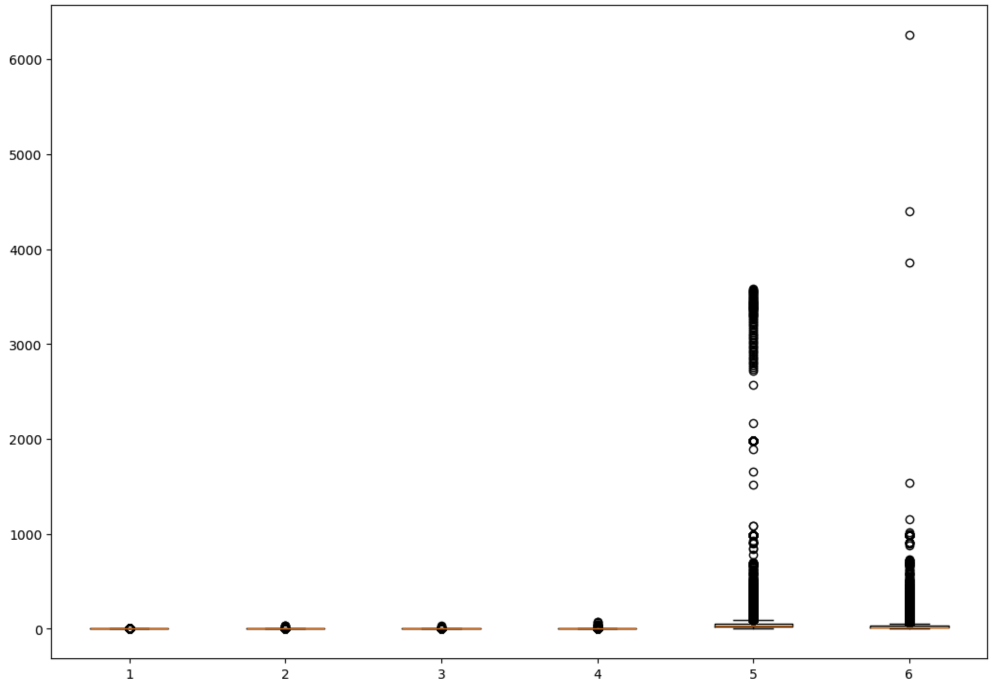

Tree - Air Quality
United States of America
Canada
United Kingdom
Germany
India
Brazil
Spain
Australia
Poland
Portugal
Czech Republic
Italy
Russia
France
Indonesia
People's Republic of China
Philippines
Norway
Croatia
New Zealand
Austria
South Africa
Malaysia
Slovakia
Pakistan
Ukraine
Japan
Belgium
Sweden
Estonia
Latvia
Chile
Mexico
Turkey
Iran
Kenya
Finland
Ireland
Serbia
Algeria
Hungary
Colombia
Greece
Romania
Israel
Ecuador
Switzerland
Thailand
South Korea
Peru

01.05

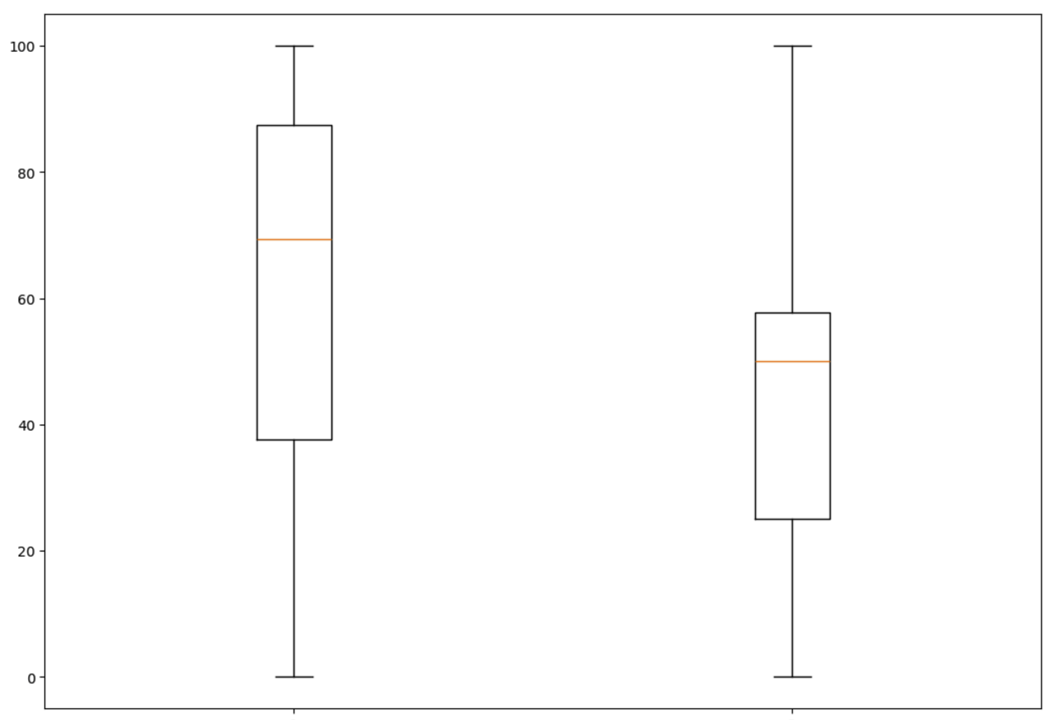
Bubble - Air Quality



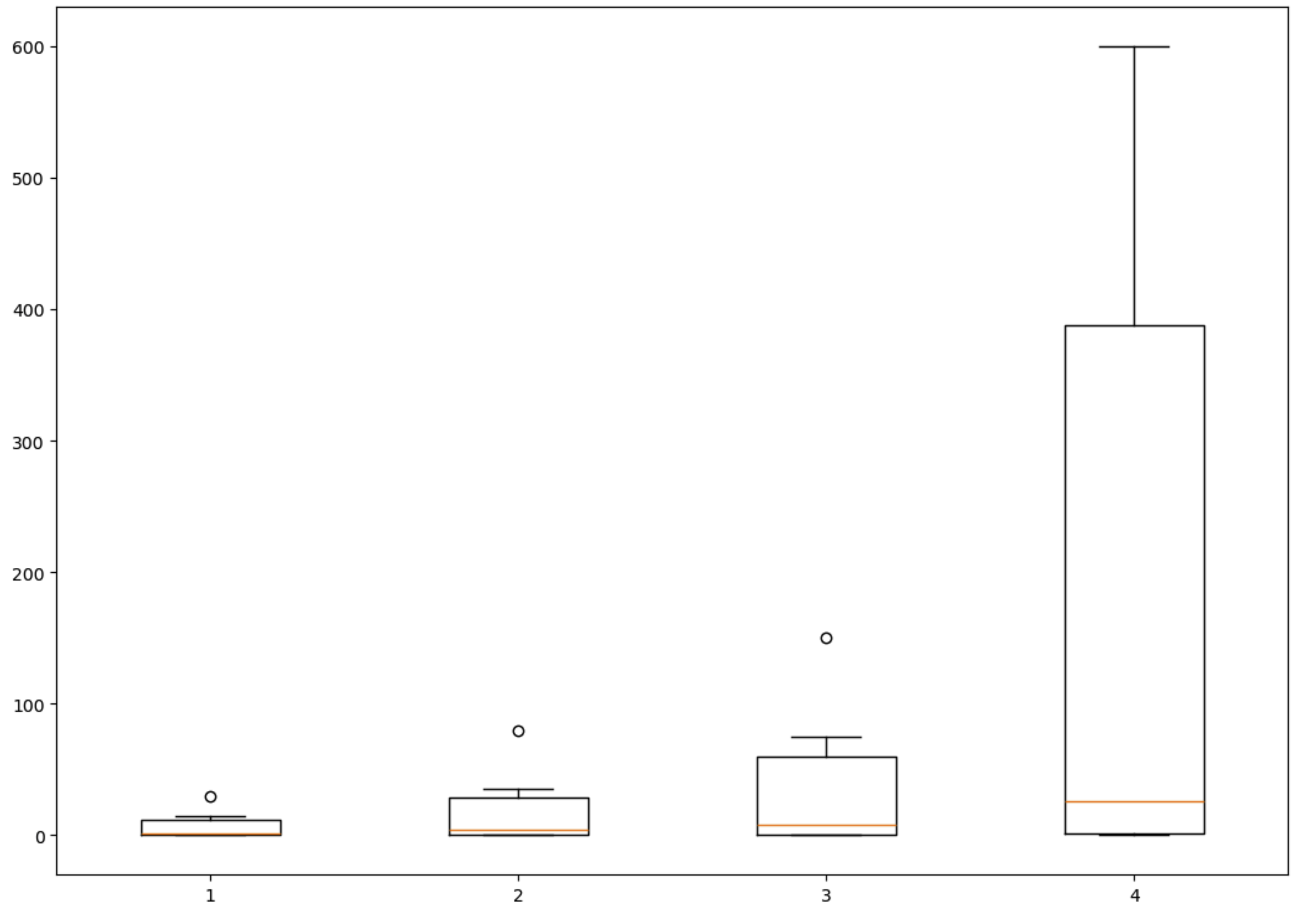
01.06



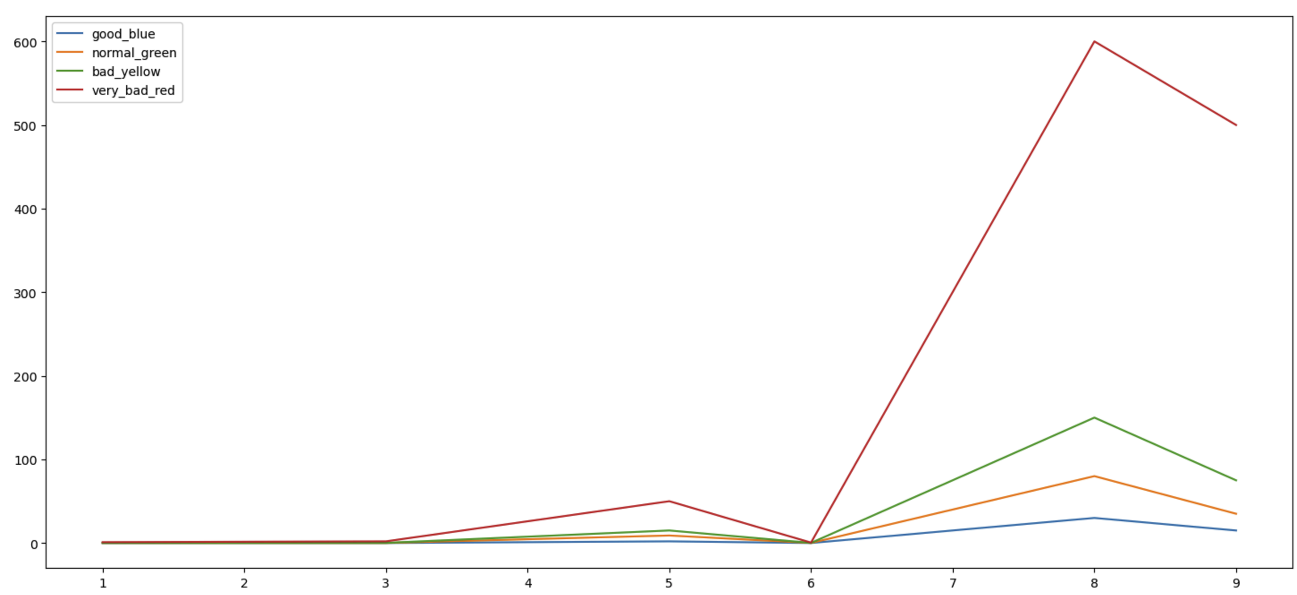
01.07



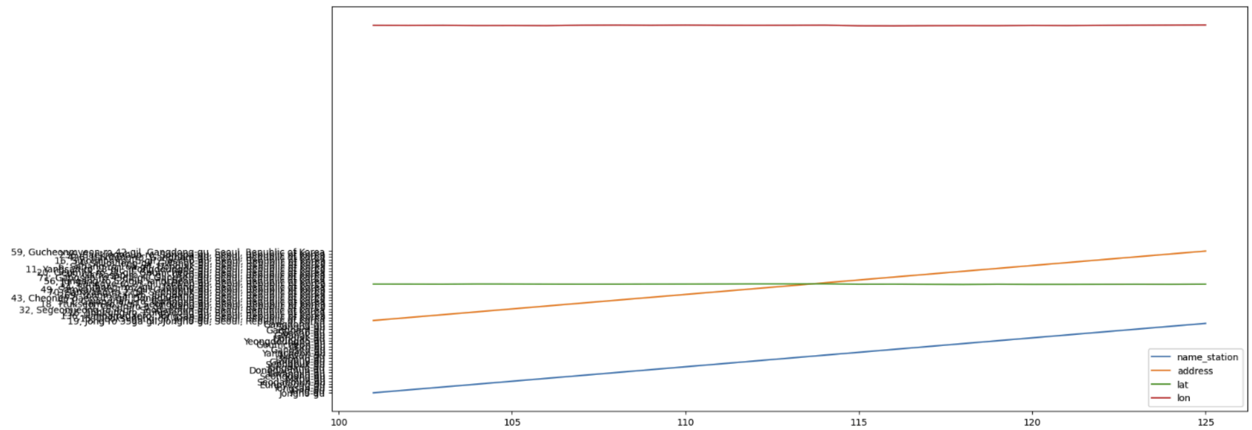
01.08


600
500
400
300
200
100
0
1
2
3
4

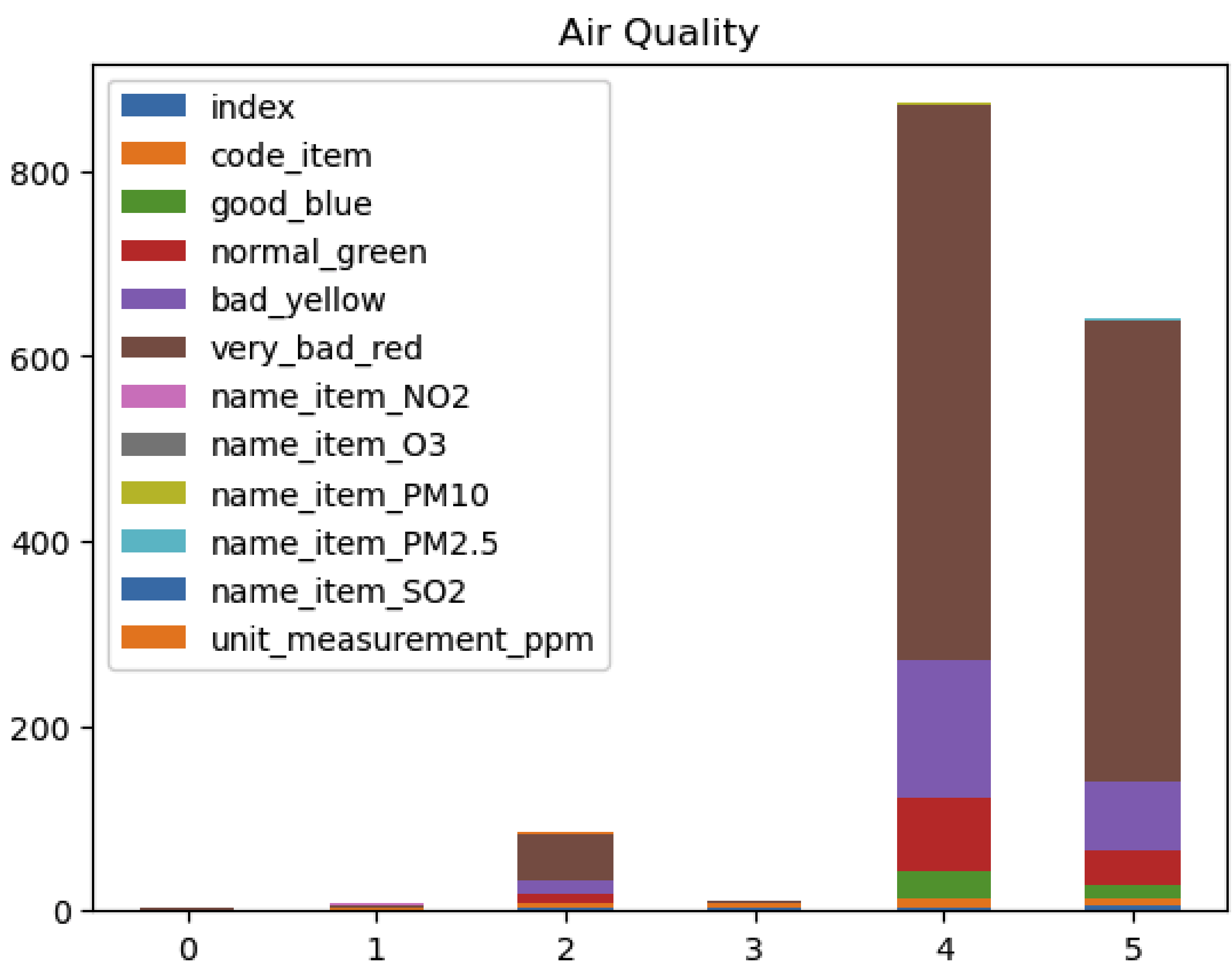
01.09


good_blue
normal_green
bad_yellow
very_bad_red
600
500
400
300
200
100
0
1
2
3
4
5
6
7
8
9

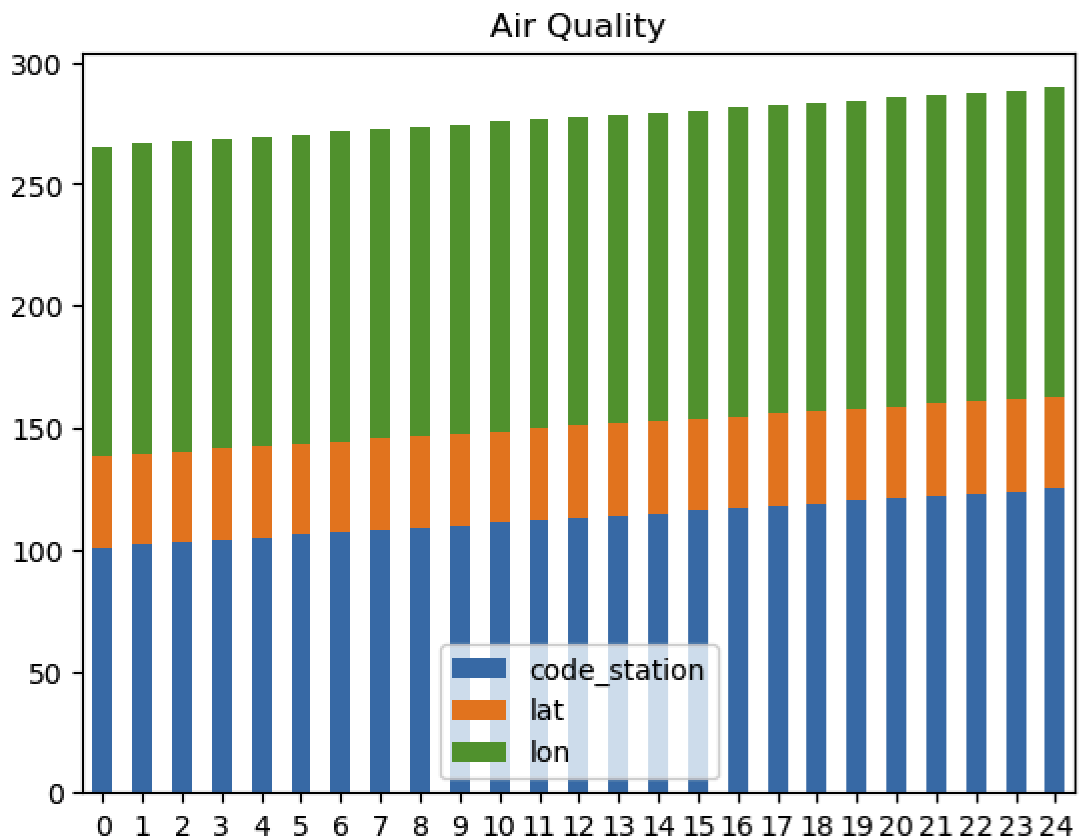
01.10


name_station
address
lat
lon
100
105
110
115
120
125

01.11


Air Quality
index
code_item
good_blue
normal_green
bad_yellow
very_bad_red
name_item_NO2
name_item_O3
name_item_PM10
name_item_PM2.5
name_item_SO2
unit_measurement_ppm
0
200
400
600
800
0
1
2
3
4
5

01.12


Air Quality
300
250
200
150
100
50
0
code_station
lat
lon
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

02.01

```python
# 06.02.17.01
# read csv
# assign variable
# dt06

dt06_dt01____________________00 = pd.read_csv('106_DT_106N_03_0200045_20240703151242.csv')
```

```python
# 06.02.17.02
# read csv
# assign variable
# dt07

dt07_dt02____________________00 = pd.read_csv('106_DT_106N_03_0200076_20240703151123.csv')
```

```python
# 06.02.17.03
# read csv
# assign variable
# dt08

dt08_summary_measurement_____00 = pd.read_csv('Measurement_summary.csv')
```

```python
# 06.02.17.04
# read csv
# assign variable
# dt09

dt09_nat_emissions___________00 = pd.read_csv('National_Air_Pollutant_Emissions_20240703151513.csv')
```

```python
# 06.02.17.05
# read csv
# assign variable
# dt10

dt10_seoul_air_______________00 = pd.read_csv('seoul_air_1988_2021.csv')
```

```python
# 06.02.17.06
# read csv
# assign variable
# dt11

dt11_seoul_ave_air___________00 = pd.read_csv('SeoulHourlyAvgAirPollution.csv')
```

02.02

```
# 06.02.23.01
# read csv
# assign variable
# dt12

dt12_streets_________________00 = pd.read_csv('116_DT_MLTM_962_20240712093545.csv')
```

```
# 06.02.23.02
# read csv
# assign variable
# dt13

dt13_ave_traffic_____________00 = pd.read_csv('Average_Daily_Traffic_by_Road_and_Vehicle_Types_20240712093400.csv')
```

```
# 06.02.23.03
# read csv
# assign variable
# dt14

dt14_yearly__________________00 = pd.read_csv('Yearly_ADT_20240712093451.csv')
```

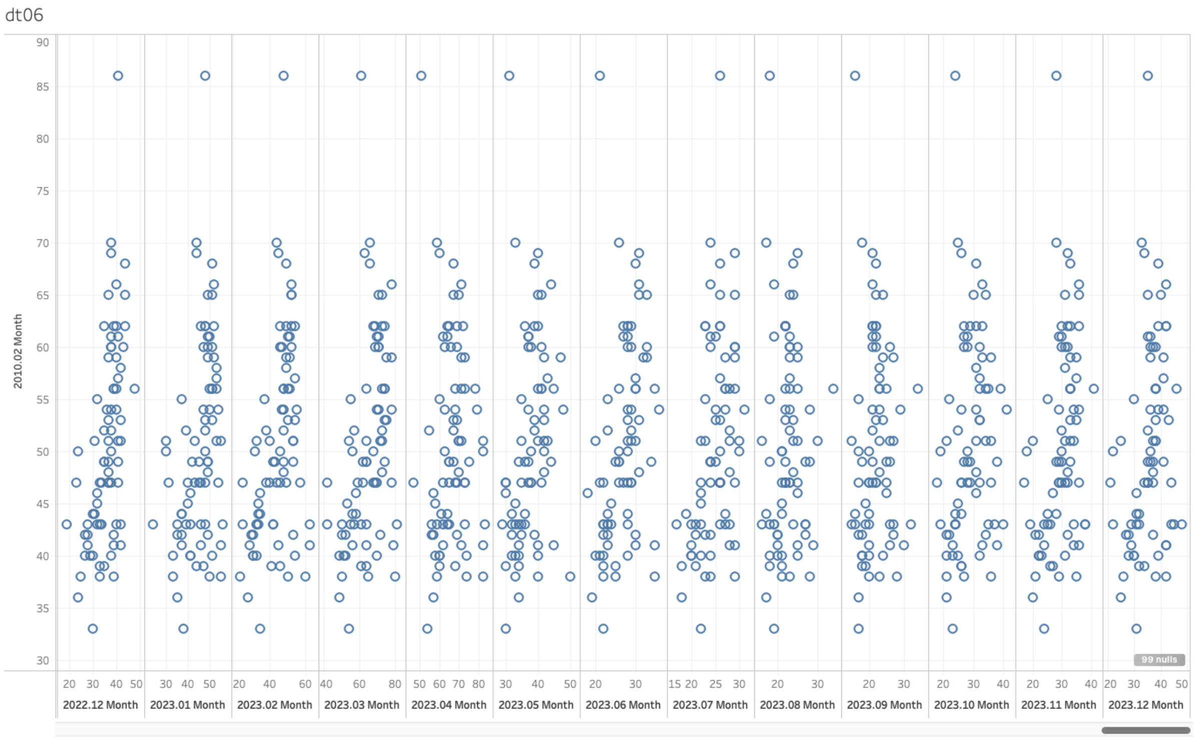
02.03

```
# 06.02.32.01
# read csv
# assign variable
# dt15

dt15_spatial_________________00 = pd.read_csv('input_data_spatial_panel.csv')
```

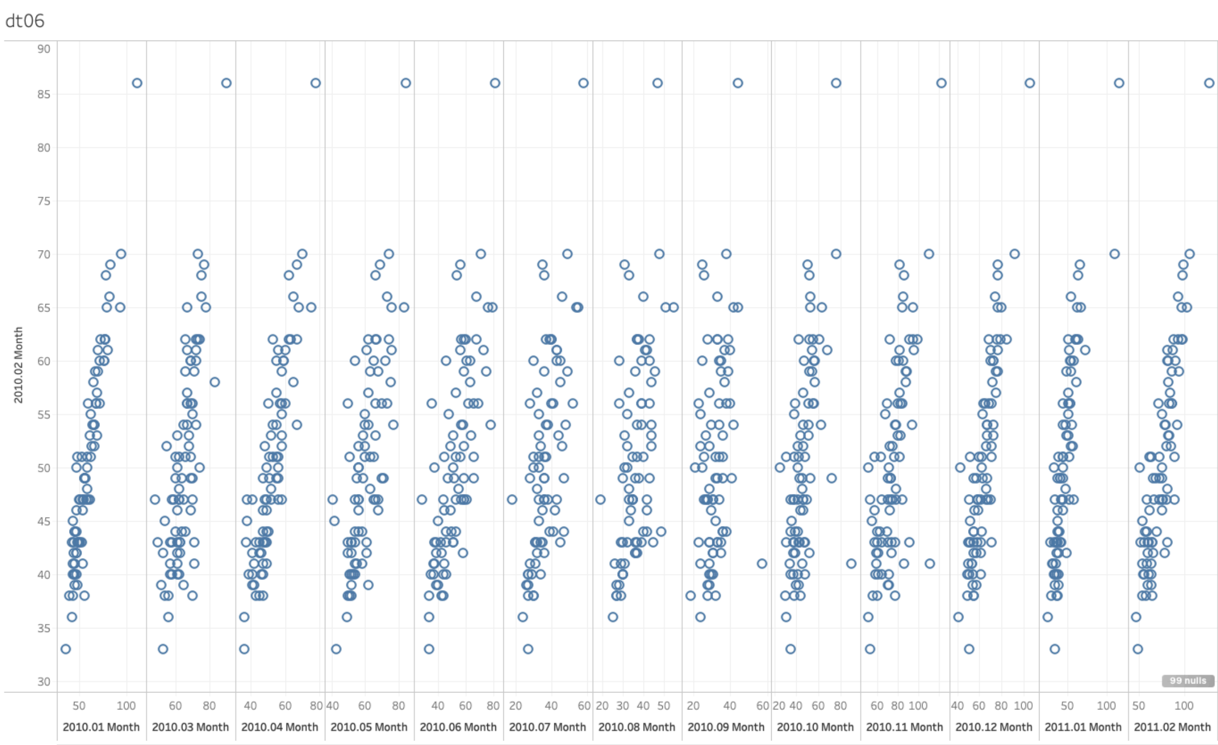
02.04

```
# 06.02.35.01
# read csv
# assign variable
# dt16

dt16_emissions_______________00 = pd.read_csv('car_fuel_emissions.csv')
```

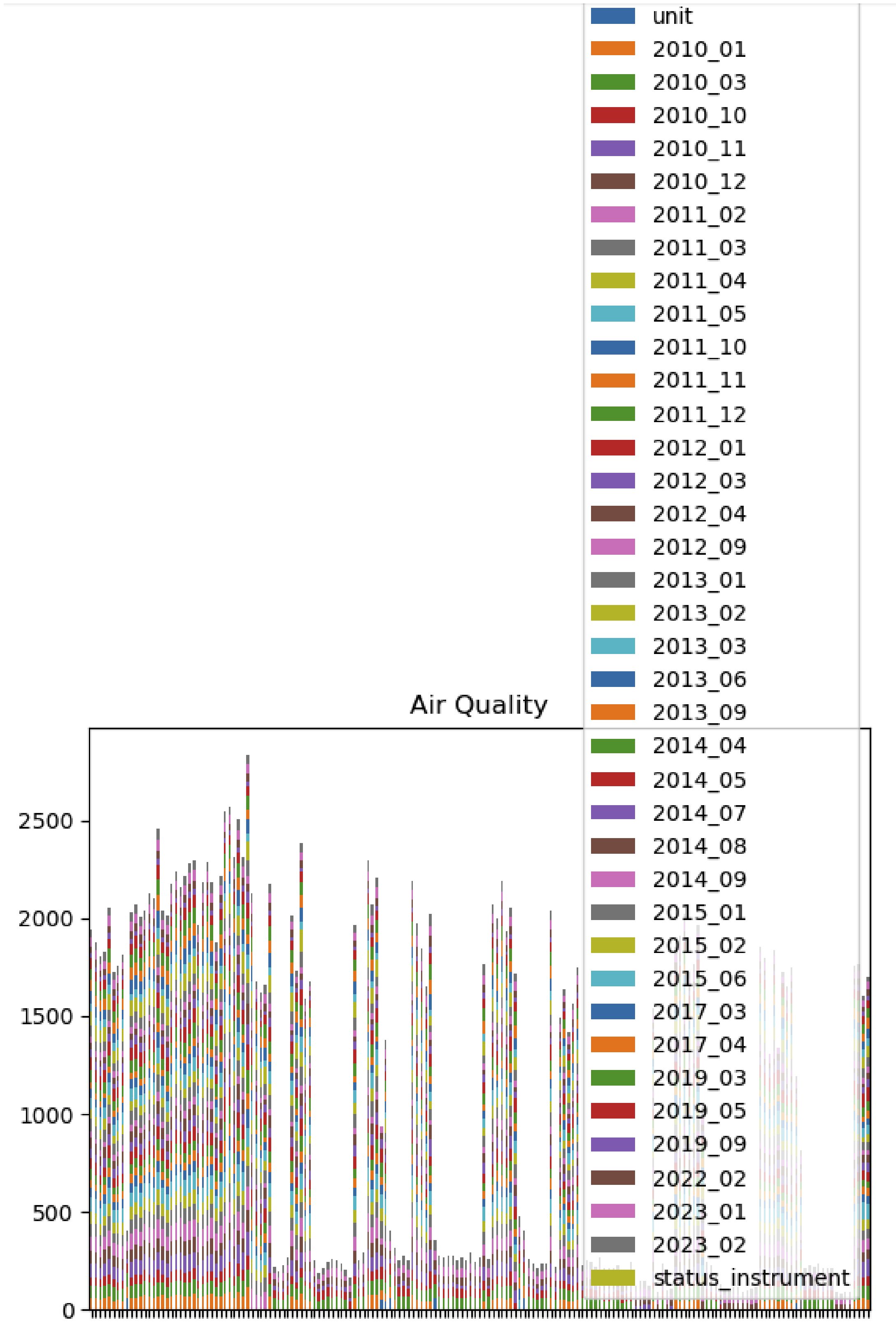
02.05


dt06
2010.02 Month
99 nulls
2022.12 Month
2023.01 Month
2023.02 Month
2023.03 Month
2023.04 Month
2023.05 Month
2023.06 Month
2023.07 Month
2023.08 Month
2023.09 Month
2023.10 Month
2023.11 Month
2023.12 Month

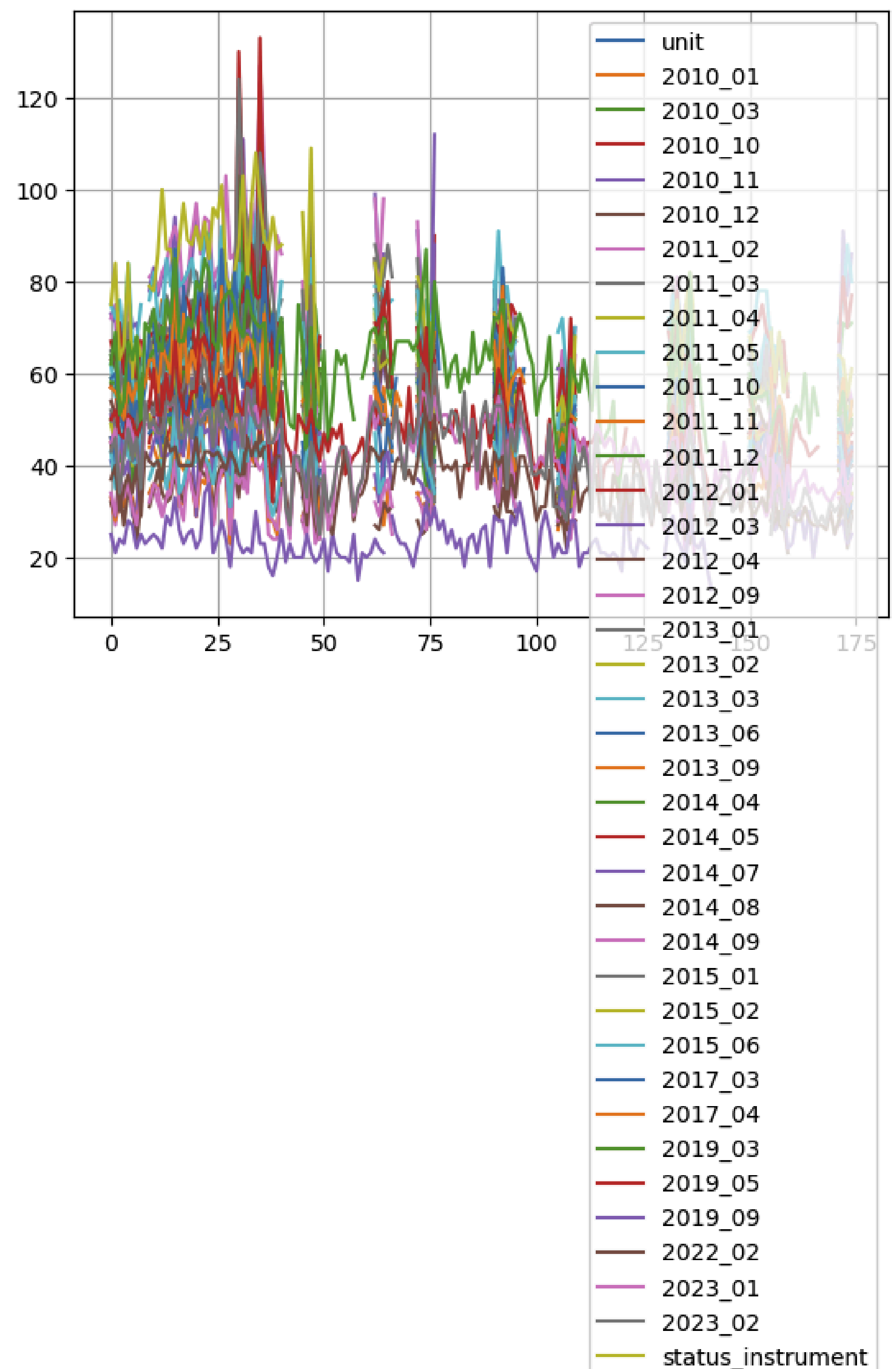
02.06


dt06
2010.02 Month
99 nulls
2010.01 Month
2010.03 Month
2010.04 Month
2010.05 Month
2010.06 Month
2010.07 Month
2010.08 Month
2010.09 Month
2010.10 Month
2010.11 Month
2010.12 Month
2011.01 Month
2011.02 Month

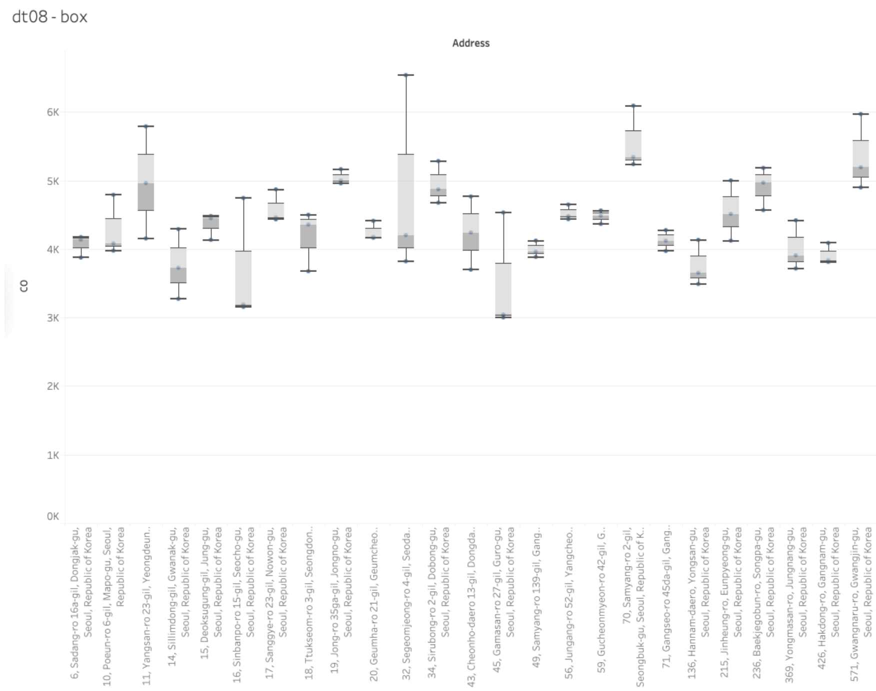
02.07


Air Quality
2500
2000
1500
1000
500
0
unit
2010_01
2010_03
2010_10
2010_11
2010_12
2011_02
2011_03
2011_04
2011_05
2011_10
2011_11
2011_12
2012_01
2012_03
2012_04
2012_09
2013_01
2013_02
2013_03
2013_06
2013_09
2014_04
2014_05
2014_07
2014_08
2014_09
2015_01
2015_02
2015_06
2017_03
2017_04
2019_03
2019_05
2019_09
2022_02
2023_01
2023_02
status_instrument

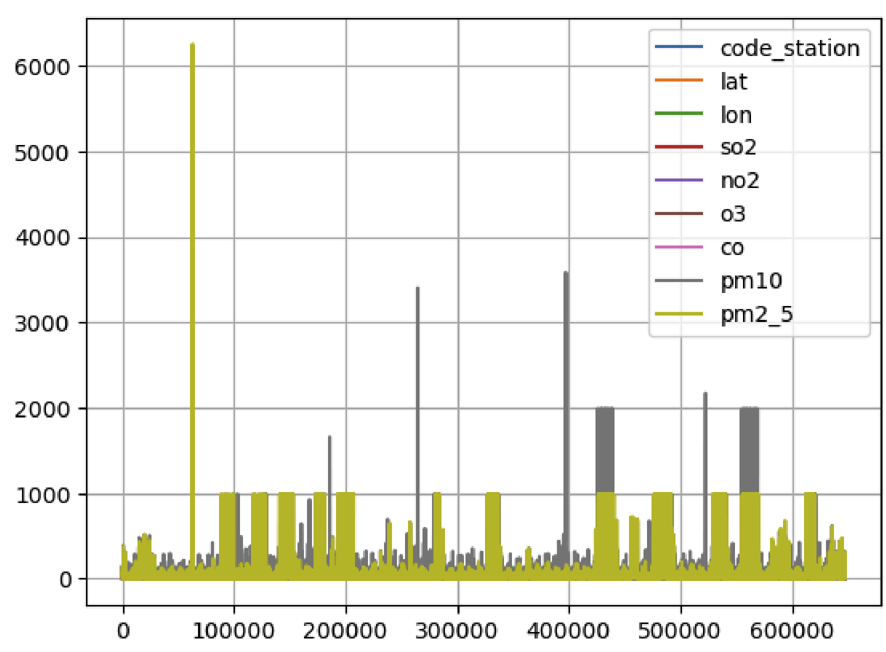
02.08


120
100
80
60
40
20
0
25
50
75
100
125
150
175
unit
2010_01
2010_03
2010_10
2010_11
2010_12
2011_02
2011_03
2011_04
2011_05
2011_10
2011_11
2011_12
2012_01
2012_03
2012_04
2012_09
2013_01
2013_02
2013_03
2013_06
2013_09
2014_04
2014_05
2014_07
2014_08
2014_09
2015_01
2015_02
2015_06
2017_03
2017_04
2019_03
2019_05
2019_09
2022_02
2023_01
2023_02
status_instrument

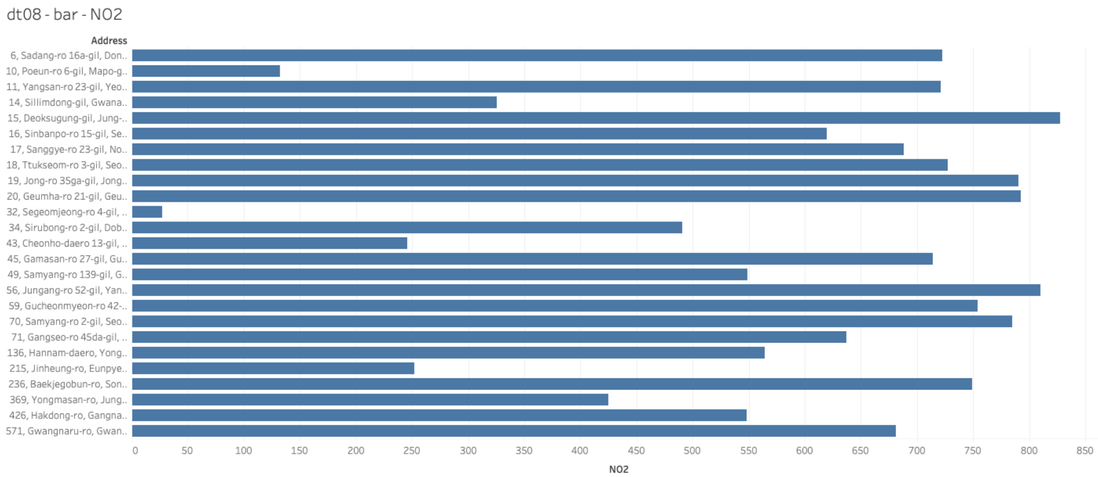
02.09


dt08 - box
Address
6K
5K
4K
3K
2K
1K
0K
co
6, Sadang-ro 16a-gil, Dongjak-gu, Seoul, Republic of Korea
10, Poeun-ro 6-gil, Mapo-gu, Seoul, Republic of Korea
11, Yangsan-ro 23-gil, Yeongdeun..
14, Sillimdong-gil, Gwanak-gu, Seoul, Republic of Korea
15, Deoksugung-gil, Jung-gu, Seoul, Republic of Korea
16, Sinbanpo-ro 15-gil, Seocho-gu, Seoul, Republic of Korea
17, Sanggye-ro 23-gil, Nowon-gu, Seoul, Republic of Korea
18, Ttukseom-ro 3-gil, Seongdon..
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republic of Korea
20, Geumha-ro 21-gil, Geumcheo..
32, Segeomjeong-ro 4-gil, Seoda..
34, Sirubong-ro 2-gil, Dobong-gu, Seoul, Republic of Korea
43, Cheonho-daero 13-gil, Dongda..
45, Gamasan-ro 27-gil, Guro-gu, Seoul, Republic of Korea
49, Samyang-ro 139-gil, Gang..
56, Jungang-ro 52-gil, Yangcheo..
59, Gucheonmyeon-ro 42-gil, G..
70, Samyang-ro 2-gil, Seongbuk-gu, Seoul, Republic of K..
71, Gangseo-ro 45da-gil, Gang..
136, Hannam-daero, Yongsan-gu, Seoul, Republic of Korea
215, Jinheung-ro, Eunpyeong-gu, Seoul, Republic of Korea
236, Baekjegobun-ro, Songpa-gu, Seoul, Republic of Korea
369, Yongmasan-ro, Jungnang-gu, Seoul, Republic of Korea
426, Hakdong-ro, Gangnam-gu, Seoul, Republic of Korea
571, Gwangnaru-ro, Gwangjin-gu, Seoul, Republic of Korea

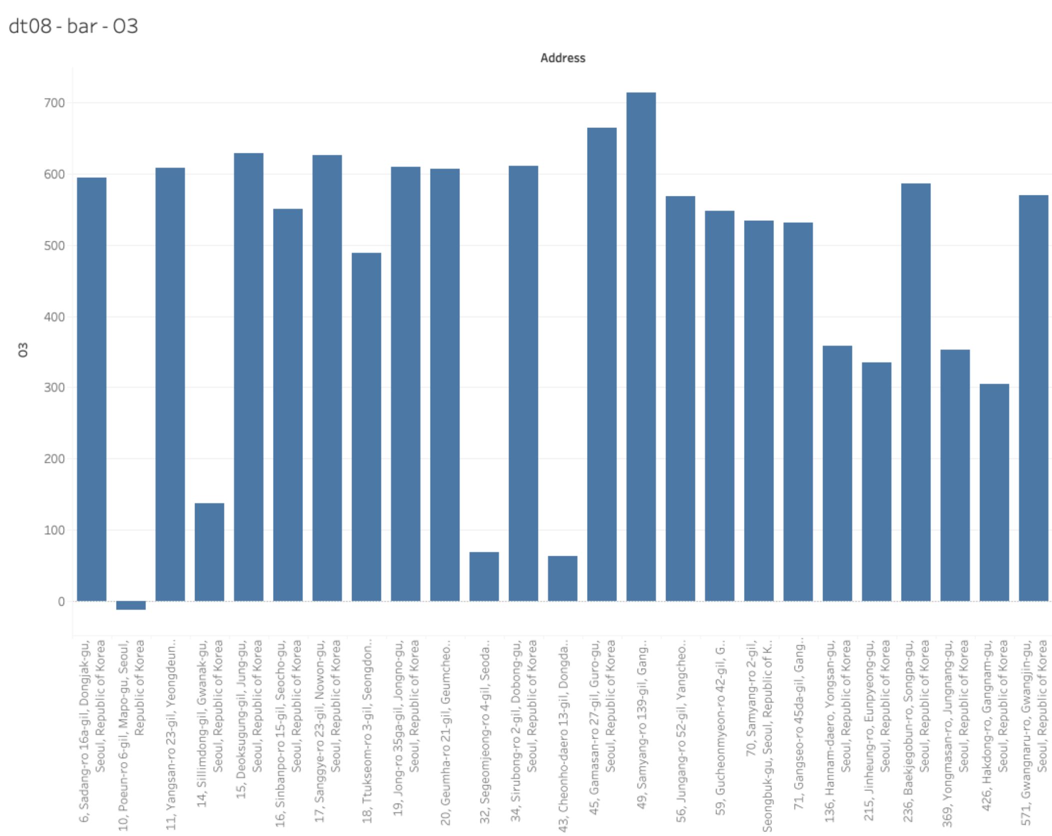
02.10


6000
5000
4000
3000
2000
1000
0
0
100000
200000
300000
400000
500000
600000
code_station
lat
lon
so2
no2
o3
co
pm10
pm2_5

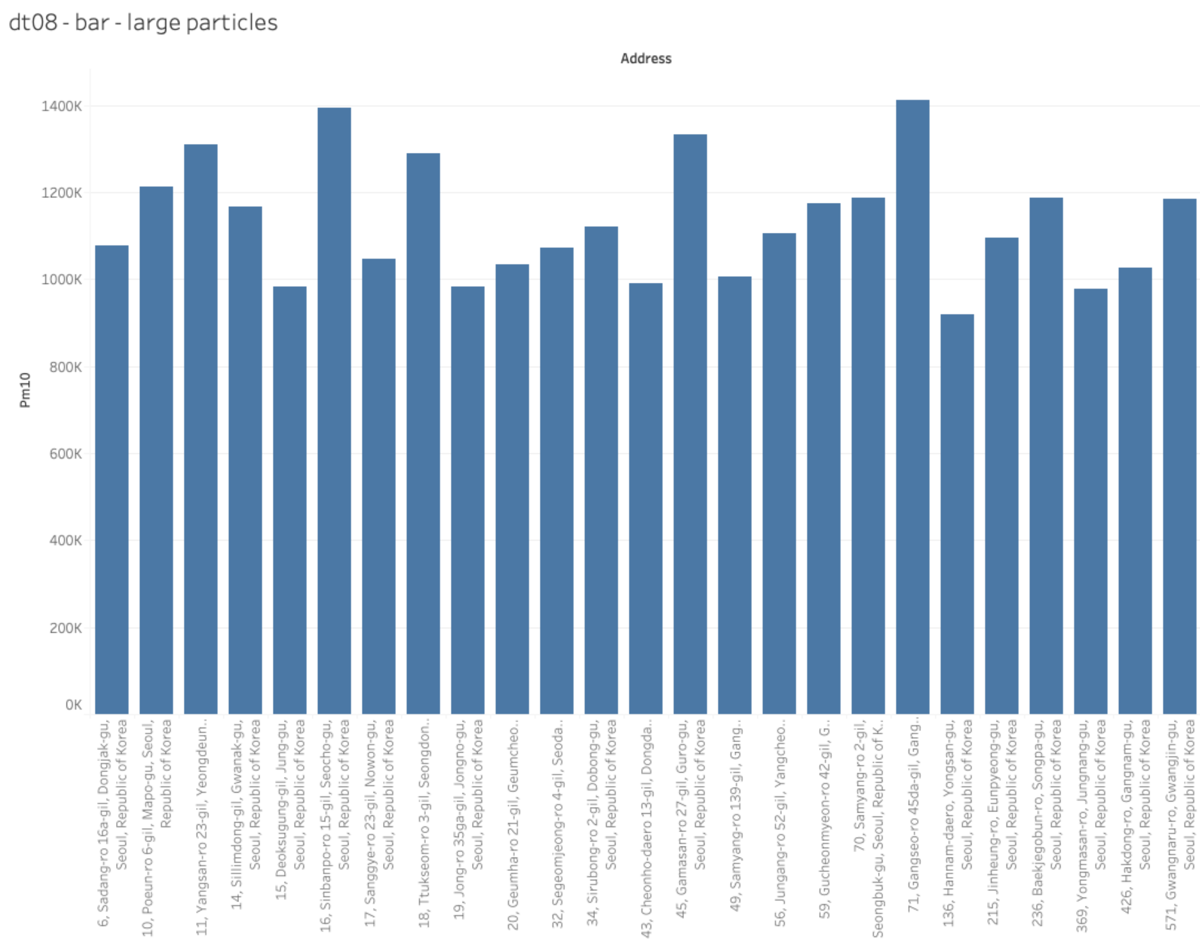
02.11


dt08 - bar - NO2
Address
6, Sadang-ro 16a-gil, Don..
10, Poeun-ro 6-gil, Mapo-g..
11, Yangsan-ro 23-gil, Yeo..
14, Sillimdong-gil, Gwana..
15, Deoksugung-gil, Jung-..
16, Sinbanpo-ro 15-gil, Se..
17, Sanggye-ro 23-gil, No..
18, Ttukseom-ro 3-gil, Seo..
19, Jong-ro 35ga-gil, Jong..
20, Geumha-ro 21-gil, Geu..
32, Segeomjeong-ro 4-gil, ..
34, Sirubong-ro 2-gil, Dob..
43, Cheonho-daero 13-gil, ..
45, Gamasan-ro 27-gil, Gu..
49, Samyang-ro 139-gil, G..
56, Jungang-ro 52-gil, Yan..
59, Gucheonmyeon-ro 42-..
70, Samyang-ro 2-gil, Seo..
71, Gangseo-ro 45da-gil, ..
136, Hannam-daero, Yong..
215, Jinheung-ro, Eunpye..
236, Baekjegobun-ro, Son..
369, Yongmasan-ro, Jung..
426, Hakdong-ro, Gangna..
571, Gwangnaru-ro, Gwan..
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850
NO2

02.12

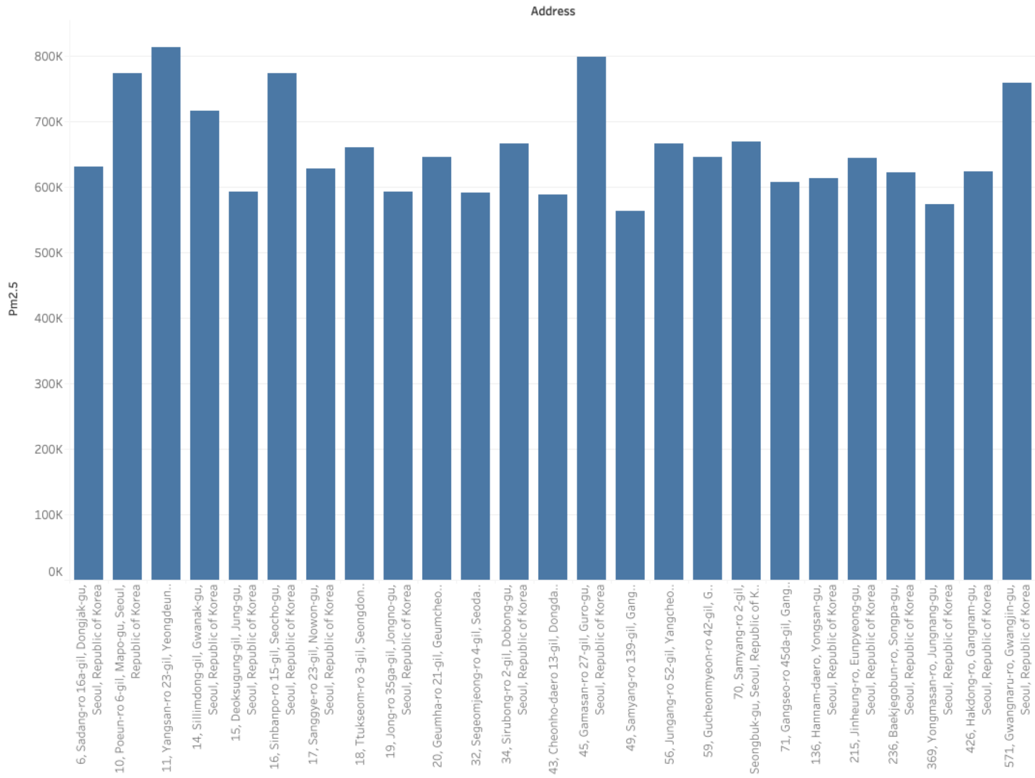

dt08 - bar - O3
Address
700
600
500
400
300
200
100
0
O3
6, Sadang-ro 16a-gil, Dongjak-gu, Seoul, Republic of Korea
10, Poeun-ro 6-gil, Mapo-gu, Seoul, Republic of Korea
11, Yangsan-ro 23-gil, Yeongdeun..
14, Sillimdong-gil, Gwanak-gu, Seoul, Republic of Korea
15, Deoksugung-gil, Jung-gu, Seoul, Republic of Korea
16, Sinbanpo-ro 15-gil, Seocho-gu, Seoul, Republic of Korea
17, Sanggye-ro 23-gil, Nowon-gu, Seoul, Republic of Korea
18, Ttukseom-ro 3-gil, Seongdon..
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republic of Korea
20, Geumha-ro 21-gil, Geumcheo..
32, Segeomjeong-ro 4-gil, Seoda..
34, Sirubong-ro 2-gil, Dobong-gu, Seoul, Republic of Korea
43, Cheonho-daero 13-gil, Dongda..
45, Gamasan-ro 27-gil, Guro-gu, Seoul, Republic of Korea
49, Samyang-ro 139-gil, Gang..
56, Jungang-ro 52-gil, Yangcheo..
59, Gucheonmyeon-ro 42-gil, G..
70, Samyang-ro 2-gil, Seongbuk-gu, Seoul, Republic of K..
71, Gangseo-ro 45da-gil, Gang..
136, Hannam-daero, Yongsan-gu, Seoul, Republic of Korea
215, Jinheung-ro, Eunpyeong-gu, Seoul, Republic of Korea
236, Baekjegobun-ro, Songpa-gu, Seoul, Republic of Korea
369, Yongmasan-ro, Jungnang-gu, Seoul, Republic of Korea
426, Hakdong-ro, Gangnam-gu, Seoul, Republic of Korea
571, Gwangnaru-ro, Gwangjin-gu, Seoul, Republic of Korea

02.13


dt08 - bar - large particles
Address
Pm10
1400K
1200K
1000K
800K
600K
400K
200K
0K

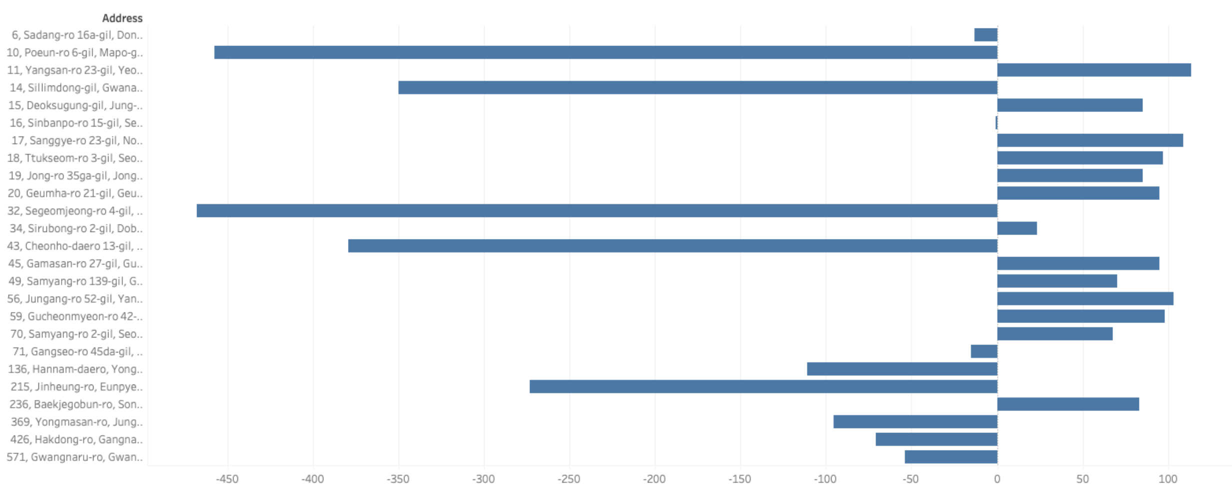
02.14

dt08 - bar - small particles

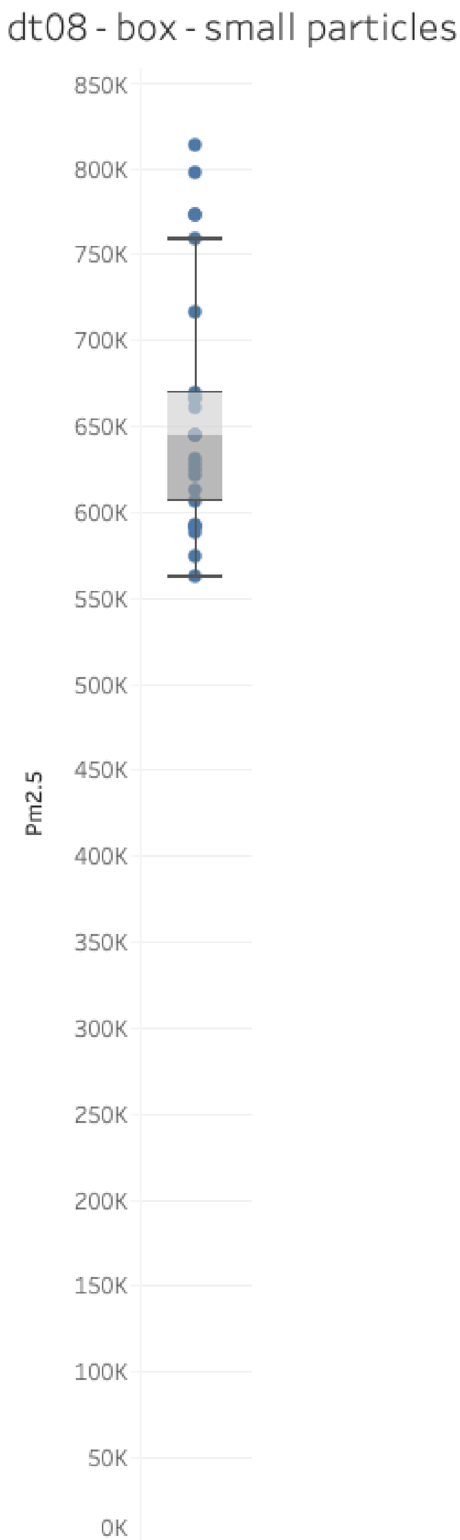


02.15

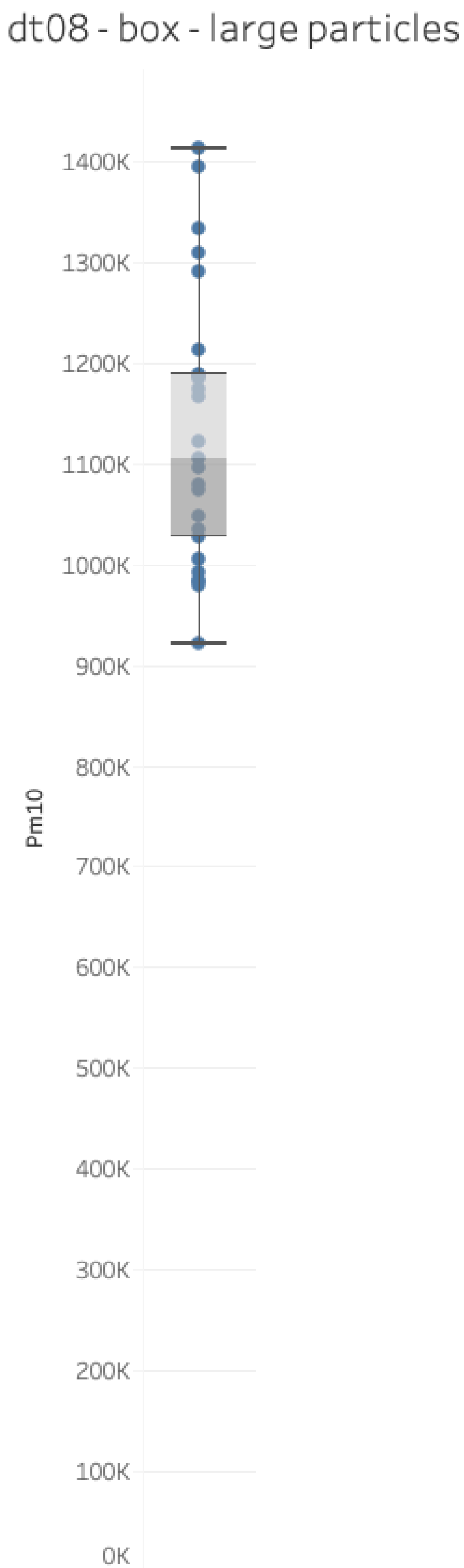
dt08 - bar - SO2



02.16

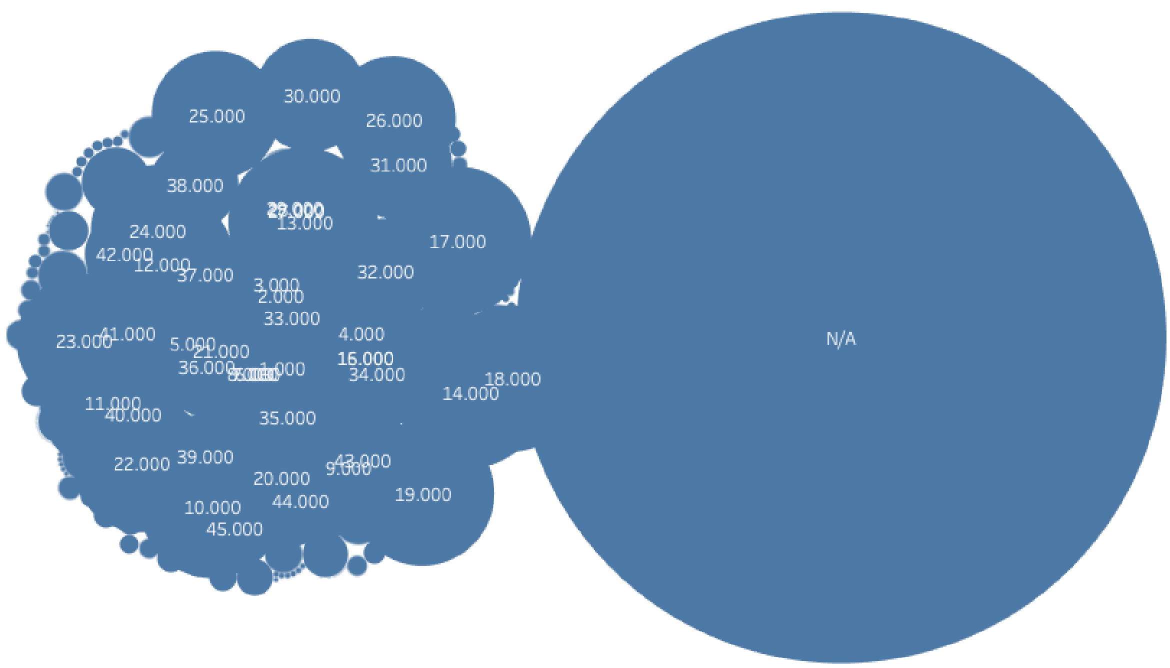

dt08 - box - small particles
850K
800K
750K
700K
650K
600K
550K
500K
450K
400K
350K
300K
250K
200K
150K
100K
50K
0K
Pm2.5

02.17

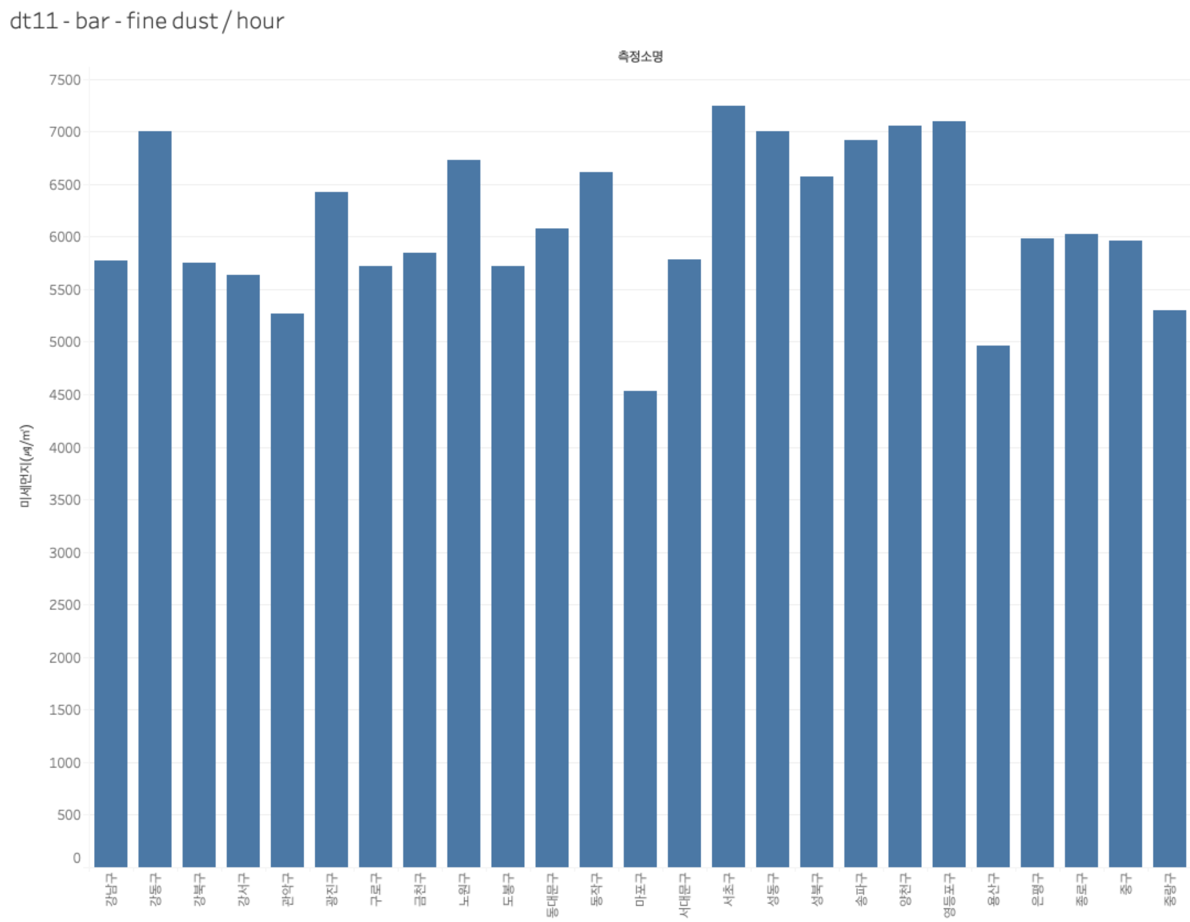

dt08 - box - large particles
1400K
1300K
1200K
1100K
1000K
900K
800K
700K
600K
500K
400K
300K
200K
100K
0K
Pm10

02.18

dt10 - bubble - fine particles



02.19

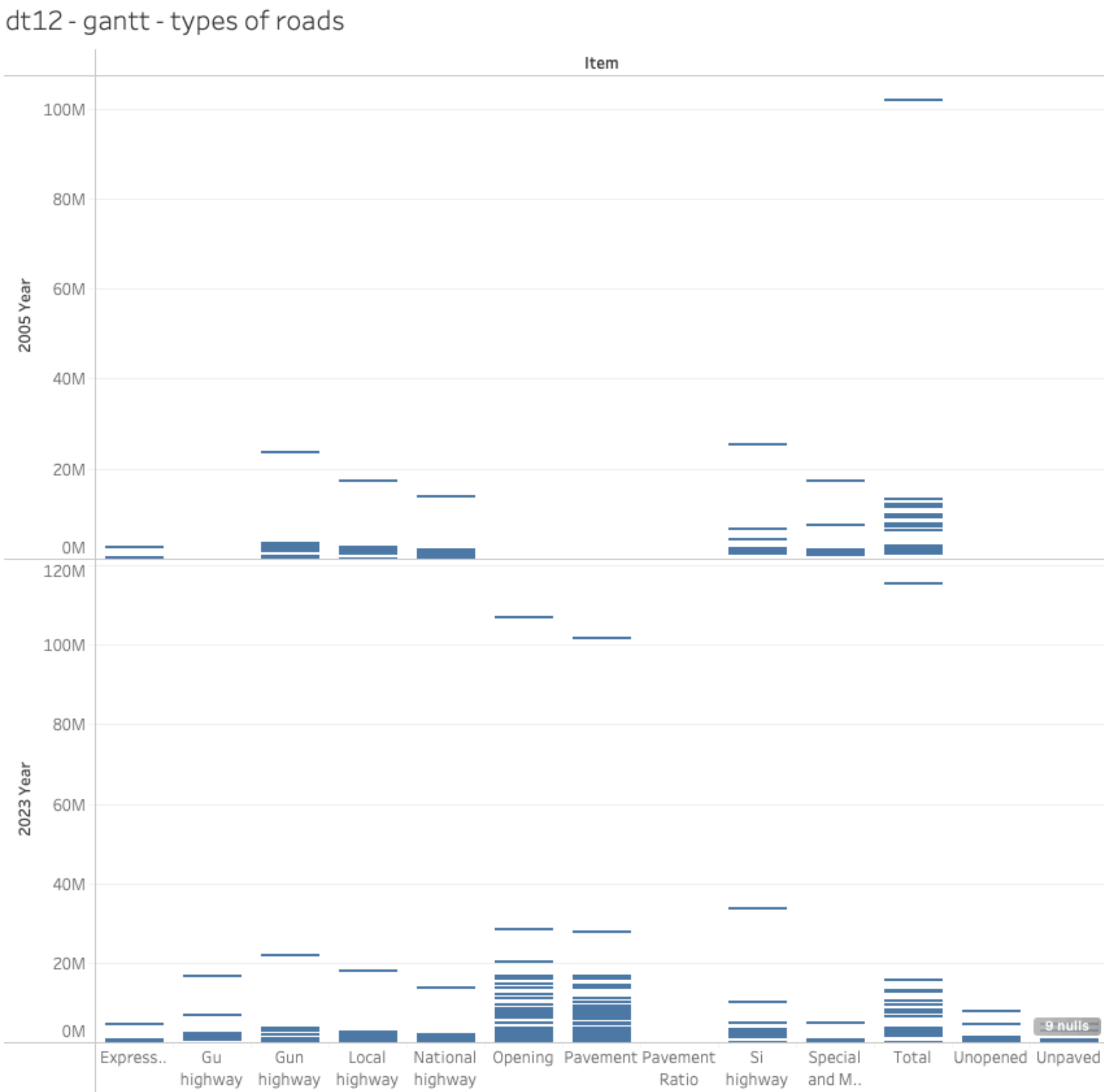

dt11 - bar - fine dust / hour
측정소명
미세먼지(㎍/㎥)
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
강남구
강동구
강북구
강서구
관악구
광진구
구로구
금천구
노원구
도봉구
동대문구
동작구
마포구
서대문구
서초구
성동구
성북구
송파구
양천구
영등포구
용산구
은평구
종로구
중구
중랑구

02.20

dt11 - chart - fine dust by region

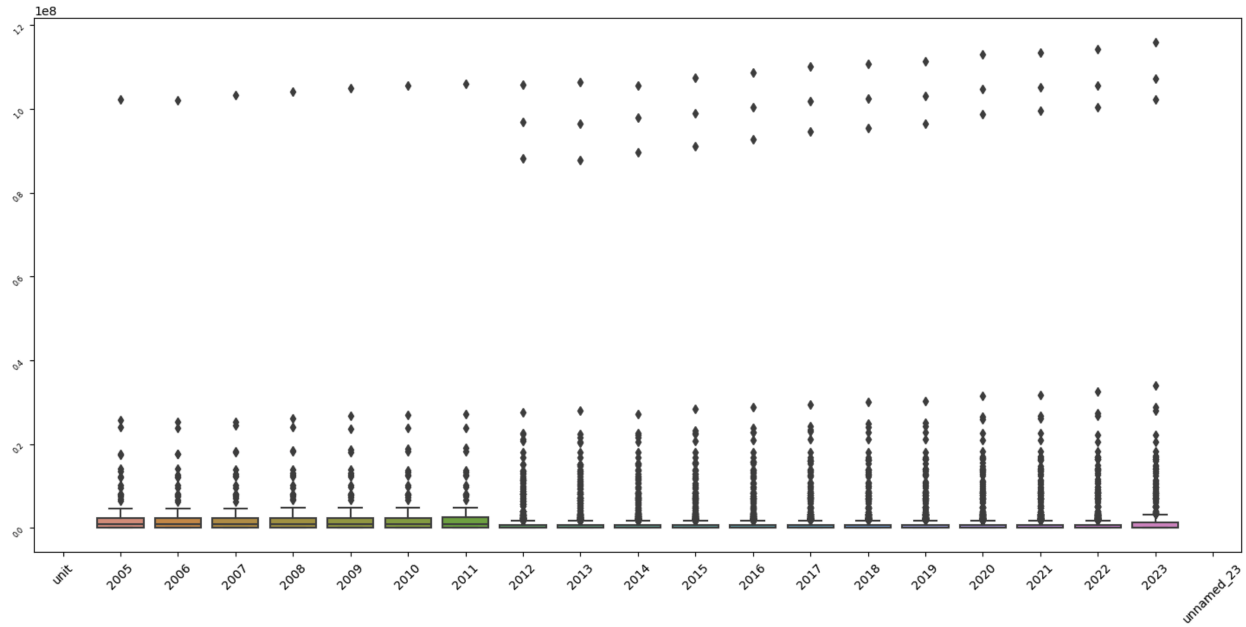
측정소명	
강남구	5,767
강동구	7,006
강북구	5,756
강서구	5,636
관악구	5,268
광진구	6,424
구로구	5,717
금천구	5,846
노원구	6,734
도봉구	5,720
동대문구	6,077
동작구	6,608
마포구	4,536
서대문구	5,779
서초구	7,249
성동구	7,006
성북구	6,573
송파구	6,917
양천구	7,058
영등포구	7,097
용산구	4,960
은평구	5,982
종로구	6,026
중구	5,957
중랑구	5,295

02.21

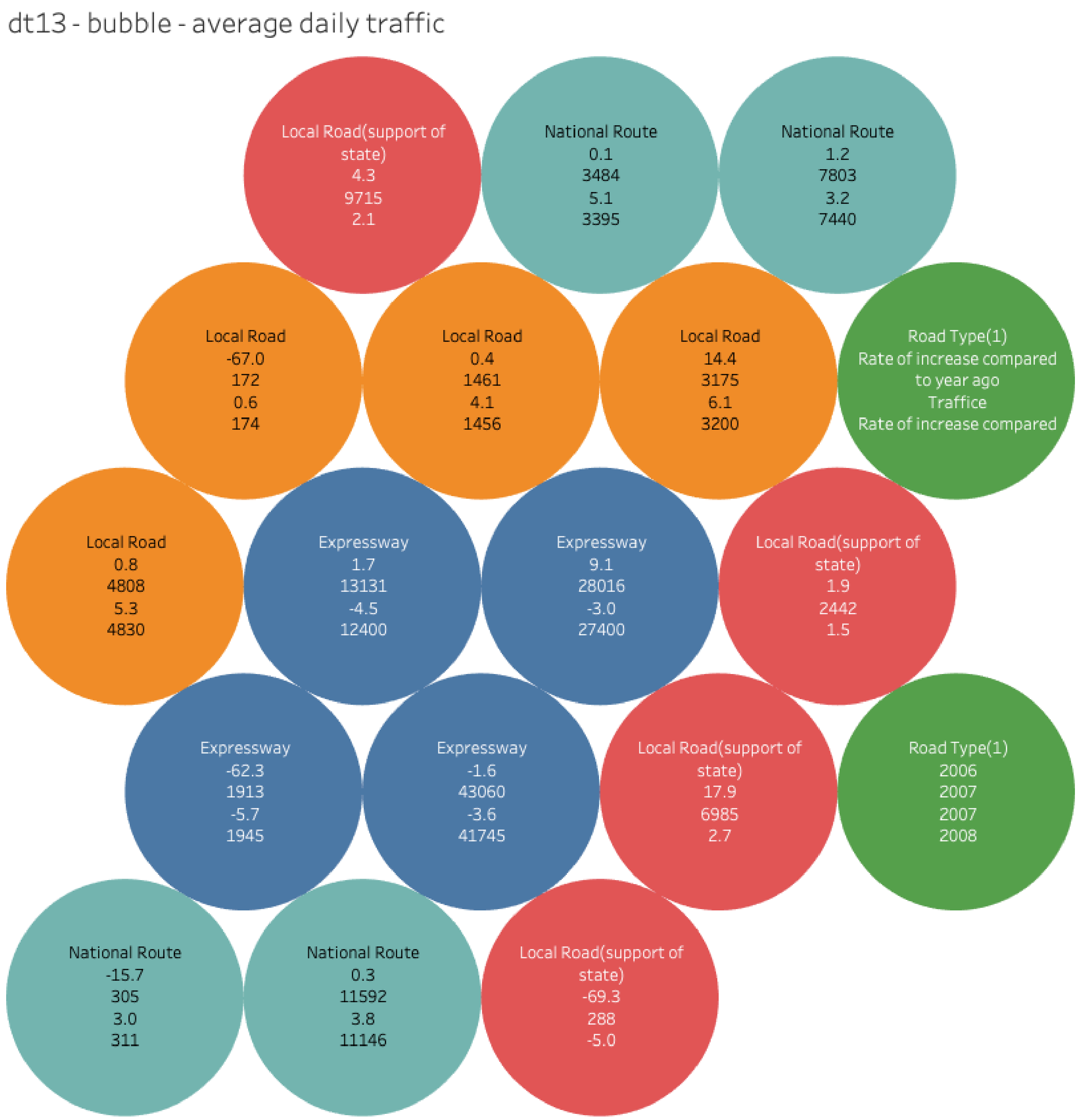


dt12 - gantt - types of roads
Item
2005 Year
2023 Year
100M
80M
60M
40M
20M
0M
120M
9 nulls
Express..
Gu highway
Gun highway
Local highway
National highway
Opening
Pavement
Pavement Ratio
Si highway
Special and M..
Total
Unopened
Unpaved

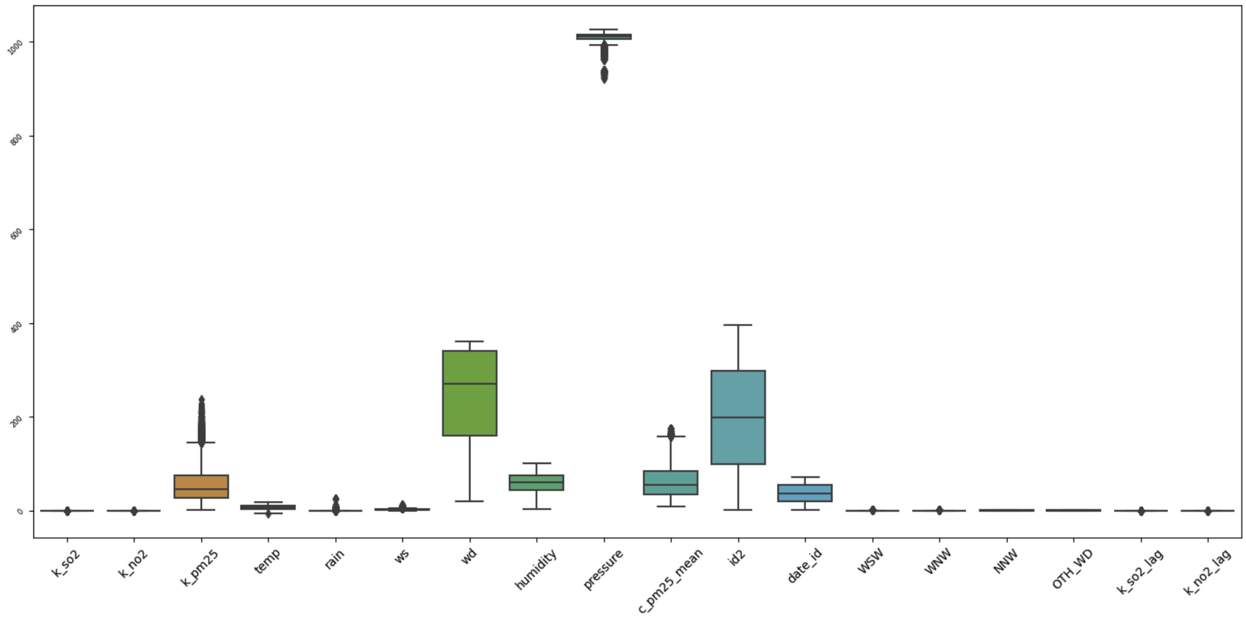
02.22


1e8
1.2
1.0
0.8
0.6
0.4
0.2
0.0
unit
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
unnamed_23

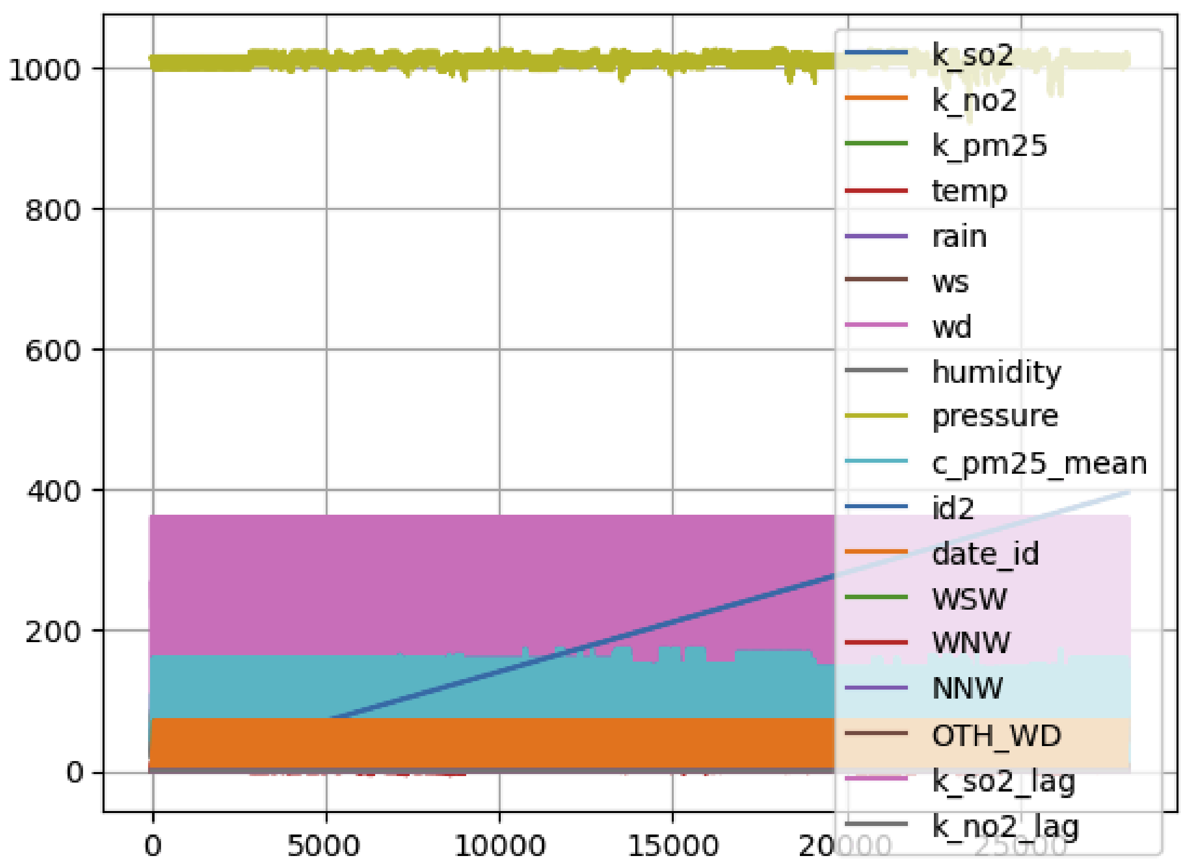
02.23


dt13 - bubble - average daily traffic
Local Road(support of state) 4.3 9715 2.1
National Route 0.1 3484 5.1 3395
National Route 1.2 7803 3.2 7440
Local Road -67.0 172 0.6 174
Local Road 0.4 1461 4.1 1456
Local Road 14.4 3175 6.1 3200
Road Type(1) Rate of increase compared to year ago Traffice Rate of increase compared
Local Road 0.8 4808 5.3 4830
Expressway 1.7 13131 -4.5 12400
Expressway 9.1 28016 -3.0 27400
Local Road(support of state) 1.9 2442 1.5
Expressway -62.3 1913 -5.7 1945
Expressway -1.6 43060 -3.6 41745
Local Road(support of state) 17.9 6985 2.7
Road Type(1) 2006 2007 2007 2008
National Route -15.7 305 3.0 311
National Route 0.3 11592 3.8 11146
Local Road(support of state) -69.3 288 -5.0

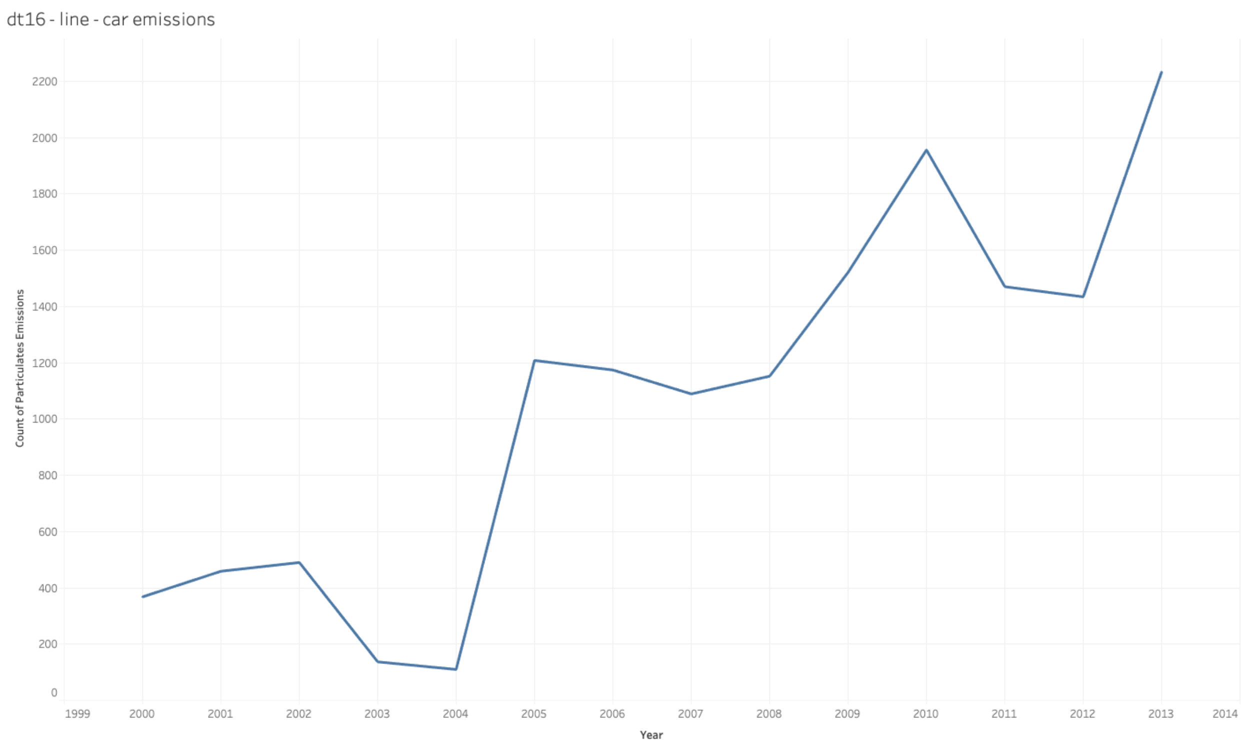
02.24


k_so2
k_no2
k_pm25
temp
rain
ws
wd
humidity
pressure
c_pm25_mean
id2
date_id
WSW
WNW
NNW
OTH_WD
k_so2_lag
k_no2_lag

02.25


1000
800
600
400
200
0
0
5000
10000
15000
20000
25000
k_so2
k_no2
k_pm25
temp
rain
ws
wd
humidity
pressure
c_pm25_mean
id2
date_id
WSW
WNW
NNW
OTH_WD
k_so2_lag
k_no2_lag

02.26


dt16 - line - car emissions
Count of Particulates Emissions
Year
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

Dataset 01:

‘Air_Quality_Seoul.csv’ as ‘dt01_Air_Quality_Seoul________00’

Dataset 02:

‘cities_air_quality_water_pollution.18-10-2021.csv’ as

‘dt02_Air_Quality_cities_______00’

Dataset 03:

'Measurement_info.csv' as ‘dt03_measurement_info_________00’

Dataset 04:

'Measurement_item_info.csv' as ‘dt04_measurement_info_item___00’

Dataset 05:

'Measurement_station_info.csv' as ‘dt05_measurement_info_stn____00’

Dataset 06:

'106_DT_106N_03_0200045_20240703151242.csv' as

‘Dt06_dt01______________________00’

Dataset 07:

'106_DT_106N_03_0200076_20240703151123.csv' as

‘Dt07_dt02______________________00’

Dataset 08:

'Measurement_summary.csv' as ‘dt08_summary_measurement_____00’

Dataset 09:

'National_Air_Pollutant_Emissions_20240703151513.csv' as

‘Dt09_nat_emissions____________00’

Dataset 10:

'Seoul_air_1988_2021.csv' as ‘dt10_seoul_air________________00’

Dataset 11:

'SeoulHourlyAvgAirPollution.csv' as ‘dt11_seoul_ave_air____________00’

Dataset 12:

'116_DT_MLTM_962_20240712093545.csv' as

‘Dt12_streets___________________00’

Dataset 13:

'Average_Daily_Traffic_by_Road_and_Vehicle_Types_20240712093400.csv'

as ‘dt13_ave_traffic______________00’

Dataset 14:

'Yearly_ADT_20240712093451.csv' as ‘dt14_yearly___________________00’

Dataset 15:

'Input_data_spatial_panel.csv' as ‘dt15_spatial__________________00’

Dataset 16:

‘Car_fuel_emissions.csv’ as ‘dt16_emissions________________00’

Ten questions an audience would ask you:

1. Are there other regions in the ROK that you should include in this research?
    a. The Seoul metropolitan area should be the primary region in this research as roughly half of the population of the ROK inhabits this area with over 23 million. Busan would be the second largest metropolitan area with over four million followed by Daegu with over two million residents.
2. What demographic aspects affect your decisions on the number of inputs?
    a. I first and foremost considered population density. Seoul has one of the highest population densities in East Asia with over 23 million inhabitants. This is also the region in Korea with the highest concentration of traffic.
3. In what ways does the degree of industrialization of a region affect your research?
    a. Although I do suspect that the degree of industrialization plays an enormous role in this research, I was unable to make comparisons with other regions. This is definitely worth exploring in further research. This goes hand in hand with the development of a country. This not only has economic implications but also social.

4. How much does the rate of industrialization of a region affect your research?
    a. Similarly to question three, I do believe that not only the degree of industrialization, but also the rate of industrialization plays a major role in this research. I think that it is likely that developing nations produce the most carbon emissions compared to undeveloped and developed countries. Unfortunately, I was unable to obtain data that could address the issue.
5. How much does urbanization affect your results?
    a. Urbanization played an enormous role in this research. It was obvious that rural areas were less polluted than urban areas. This also helps to prove that the majority of particulate matter is domestically produced. Not only is this backed by the data, one could simply travel to the rural areas of the ROK to see that the sky is significantly clearer.
6. What other metrics could you use in your research?
    a. It would be interesting to include political and social metrics in this research. I think that politics play a key role in this issue. The fact that Korea, in some ways, is still a developing country along with most of China, I think that addressing social aspects would be relevant. For example, despite having adequate and affordable public transportation, Koreans still prefer to drive a

personal automobile. The mindset of the general population of a developing country would, in most cases, choose to show off their social status rather than protect the environment.

7. How much does air quality and water quality correlate?
   a. I was unable to make a direct comparison between the two. I hope to tackle the issue of water quality in my next project. I do suspect, however, that water quality does not correlate to air pollution in developed nations.
8. What are the major indicators of the largest emitters based on your research?
   a. Based on my research, the most significant emitter of particulates were vehicles. The data clearly showed that there is a correlation between the amount of roads constructed in the target region, the amount of traffic in that region, and the amount of air pollution in that region.
9. How much does politics affect your research?
   a. Political and social aspects are both two metrics that would be worth exploring. Unfortunately, I was unable to find relevant data to address the issues of air quality. I suspect that both politics and social issues play profound roles in the amount of air pollution and the issue of fine dust. I think that for social reasons, fine dust is more of an issue in developing countries

10. After deep insights, what other regions would you consider researching in the future?

    a. It would be worth exploring any region that has higher rates of air pollution. I think that North Korea would be very relevant in this research. I would suspect that North Korea, due to its demographics, has a substantially lower rate of fine dust compared to its South Korean counterparts. I would also consider researching populated parts of the United States and making a comparison to rural areas. This could shed light on the social impact between developed and developing countries.