

---

# 빅데이터를 이용한 통계그래픽스 TEAM PROJECT

---

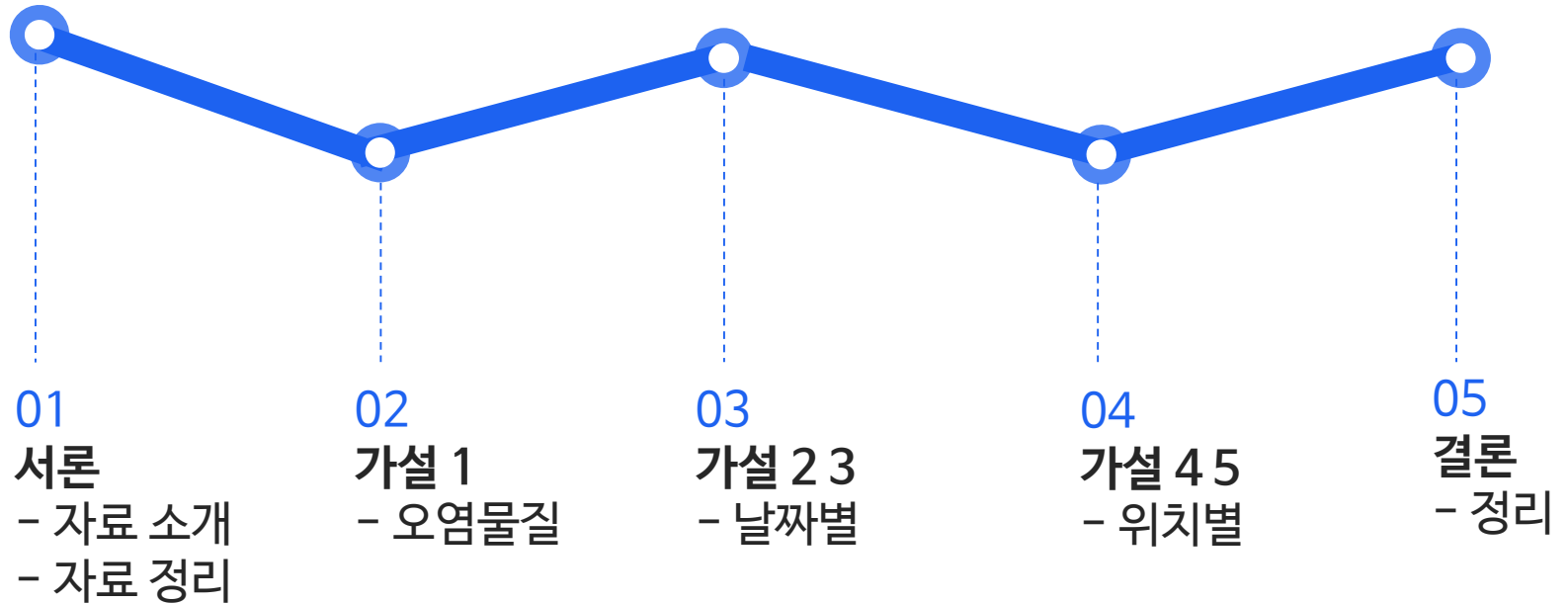
[Theme : Pollution]

4조

김송희 양보연 이하은  
정유진 주선미 최희원

# 목차

## CONTENTS

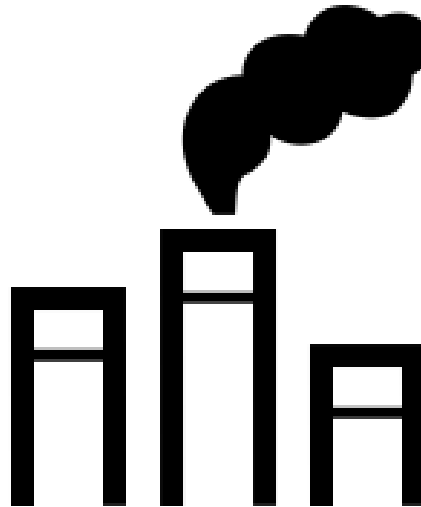


# 01 서론

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

## [ US Pollution ]



2000년부터 2016년까지 **미국 내 주·군·시별 4개 오염 물질**(NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, CO)의 관측 자료

자료 크기: 약 390MB / 관측치 수: **1746661** / 변수 개수: **29** (범주형 + 연속형)

# 01 서론 - 변수 정리

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

Variable		Type	Description
State_Code		int	미국 환경보호국(EPA)이 지정한 주 코드
County_Code		int	미국 환경보호국(EPA)이 지정한 군 코드
Site_Num		int	미국 환경보호국(EPA)이 지정한 관측 장소 코드
Address		char	관측 장소의 실제 주소
State		char	관측 장소의 주 명칭
County		char	관측 장소의 군 명칭
City		char	관측 장소의 시 명칭
Date_Local		date	관측 일자
NO2(이산화질소) O3(오존) SO2(이산화황) CO(일산화탄소)	_Units	char	이산화질소 측정 단위
	_Mean	num	하루 동안 측정된 이산화질소의 평균 농도
	_1st_Max_Value	num	하루 동안 측정된 이산화질소 농도 중 최대치
	_1st_Max_Hour	int	이산화질소 농도 최대치가 측정된 시간
	_AQI	int	하루 동안 측정된 이산화질소의 공기 품질 지표
Year		int	관측한 연도
Month		int	관측한 달
Day		int	관측한 날짜

Date\_Local  
2010-01-01  
Year      Month      Day

---

# 01 서론 - 관측치 정리

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

---



[ 관측치의 중복 ]

[ 이상치 ]

# 01 서론 - ① [ 관측치의 중복 ]

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

O3_Mean	O3_1st_Max_Value	O3_1st_Max_Hour	O3_AQI
0.0225	0.04	10	34
0.0225	0.04	10	34
0.0225	0.04	10	34
0.0225	0.04	10	34
0.013375	0.032	10	27
0.013375	0.032	10	27
0.013375	0.032	10	27
0.013375	0.032	10	27

▲ 데이터 일부 - O3

SO2_Mean	SO2_1st_Max_Value	SO2_1st_Max_Hour	SO2_AQI
3	9	21	13
3	9	21	13
2.975	6.6	23	NA
2.975	6.6	23	NA
1.958333	3	22	4
1.958333	3	22	4
1.9375	2.6	23	NA
1.9375	2.6	23	NA

▲ 데이터 일부 - SO2

# 01 서론 - ① [ 관측치의 중복 ]

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

```
Pollution_NO2 <- Pollution %>% select(State_Code, County_Code, Site_Num, Address, State,  
County, City, Date_Local, Year, Month, Day, NO2_Units, NO2_Mean,  
NO2_1st_Max_Value, NO2_1st_Max_Hour, NO2_AQI)
```

NO2

Pollution\_NO2

O3

Pollution\_O3

SO2

Pollution\_SO2

CO

Pollution\_CO

# 01 서론 - ① [ 관측치의 중복 ]

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

```
Pollution_NO2 <- Pollution_NO2 %>% distinct()
```

Year	Month	Day	NO2_Mean	NO2_1st_Max_Value
2000	1	1	19.041667	49
2000	1	1	19.041667	49
2000	1	1	19.041667	49
2000	1	1	19.041667	49
2000	1	2	22.958333	36

Year	Month	Day	NO2_Mean	NO2_1st_Max_Value
2000	1	1	19.041667	49
2000	1	2	22.958333	36
2000	1	3	38.125	51
2000	1	4	40.26087	74
2000	1	3	38.125	51
2000	1	4	40.26087	74
2000	1	4	40.26087	74
2000	1	4	40.26087	74
2000	1	4	40.26087	74



# 01 서론 - ②[이상치]

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

> summary(Pollution)

State_Code	County_Code	Site_Num	Address	State	County
Min. : 1.00	Min. : 1.00	Min. : 1	Length:1746661	Length:1746661	Length:1746661
1st Qu.: 6.00	1st Qu.: 17.00	1st Qu.: 9	Class :character	Class :character	Class :character
Median :17.00	Median : 59.00	Median : 60	Mode :character	Mode :character	Mode :character
Mean :22.31	Mean : 71.69	Mean :1118			
3rd Qu.:40.00	3rd Qu.: 97.00	3rd Qu.:1039			
Max. :80.00	Max. :650.00	Max. :9997			

Date_Local	Year	Month	Day	NO2_Units	NO2_Mean
Min. :2000-01-01	Min. :2000	Min. : 1.00	Min. : 1.00	Length:1746661	Min. : -2.00
1st Qu.:2004-11-23	1st Qu.:2004	1st Qu.: 4.00	1st Qu.: 8.00	Class :character	1st Qu.: 5.75
Median :2009-02-03	Median :2009	Median : 7.00	Median :16.00	Mode :character	Median : 10.74
Mean :2008-10-13	Mean :2008	Mean : 6.52	Mean :15.75		Mean : 12.82
3rd Qu.:2012-11-06	3rd Qu.:2012	3rd Qu.: 9.00	3rd Qu.:23.00		3rd Qu.: 17.71
Max. :2016-05-31	Max. :2016	Max. :12.00	Max. :31.00		Max. :139.54

NO2_AQI	O3_Units	O3_Mean	O3_1st_Max_Value	O3_1st_Max_Hour	O3_AQI
Min. : 0.0	Length:1746661	Min. :0.00000	Min. :0.0000	Min. : 0.00	Min. : 0.00
1st Qu.: 12.0	Class :character	1st Qu.:0.01787	1st Qu.:0.0290	1st Qu.: 9.00	1st Qu.: 25.00
Median : 23.0	Mode :character	Median :0.02587	Median :0.0380	Median :10.00	Median : 33.00
Mean : 23.9		Mean :0.02612	Mean :0.0392	Mean :10.17	Mean : 36.05
3rd Qu.: 33.0		3rd Qu.:0.03392	3rd Qu.:0.0480	3rd Qu.:11.00	3rd Qu.: 42.00
Max. :132.0		Max. :0.09508	Max. :0.1410	Max. :23.00	Max. :218.00

# 01 서론 - ②[ 이상치 ]

자료 소개 및 정리  
- US Pollution

데이터 출처

<http://slideplayer.com/slide/5675334/>

## Parts per million/billion (ppm & ppb)

$$\bullet \text{ ppm} = \frac{\text{mass solute}}{\text{volume solution}} \times 10^6$$

$$\text{or } \frac{\text{mg}}{\text{L}} = \text{ppm}$$

$$\bullet \text{ ppb} = \frac{\text{mass solute}}{\text{volume solution}} \times 10^9$$

$$\text{or } \frac{\mu\text{g}}{\text{L}} = \text{ppb}$$

Mass and volume units must match.

(g & mL)      or      (Kg & L)

AND

For very low  
concentrations:

$$\text{parts per trillion } \frac{\text{ng}}{\text{L}} = \text{ppt}$$

# 01 서론 - ②[ 이상치 ]

자료 소개 및 정리  
- US Pollution

데이터 출처  
<https://www.kaggle.com/sogun3/uspollution>  
논문 '관악과 시청의 요일별 오존농도'

```
Pollution_NO2 <- Pollution_NO2 %>% filter(NO2_Mean > 0,  
                                             NO2_1st_Max_Value > 0)
```

```
Pollution_SO2 <- Pollution_SO2 %>% filter(SO2_Mean > 0,  
                                             SO2_1st_Max_Value > 0)
```

```
Pollution_O3 <- Pollution_O3 %>% filter(O3_Mean > 0,  
                                           O3_1st_Max_Value > 0)
```

```
Pollution_CO <- Pollution_CO %>% filter(CO_Mean > 0,  
                                           CO_1st_Max_Value > 0)
```

## 02 본론-가설1

본론 소주제 소개

### 1 오염물질 자체 분석

1) 가설1



### 2 날짜별 오염물질 분석

1) 가설2

2) 가설3

### 3 위치별 오염물질 분석

1) 가설4

5) 가설5

## 02 본론-가설1



### 1. 오염물질 자체 분석

1) 가설1 "출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다."

Variables Description

Variable		Type	Description
State_Code		int	미국 환경보호국(EPA)이 지정한 주 코드
County_Code		int	미국 환경보호국(EPA)이 지정한 군 코드
Site_Num		int	미국 환경보호국(EPA)이 지정한 관측 장소 코드
Address		char	관측 장소의 실제 주소
State		char	관측 장소의 주 명칭
County		char	관측 장소의 군 명칭
City		char	관측 장소의 시 명칭
Date_Local		Date	관측 일자
NO2(이산화질소) O3(오존) SO2(이산화황) CO(일산화탄소)	+Units	char	이산화질소 측정 단위
	_Mean	num	하루 동안 측정된 이산화질소의 평균 농도
	_1st_Max_Value	num	하루 동안 측정된 이산화질소 농도 중 최대치
	1st_Max_Hour	int	이산화질소 농도 최대치가 측정된 시간
	_AQI	int	하루 동안 측정된 이산화질소의 공기 품질 지표
Year		int	관측한 연도
Month		int	관측한 달
Day		int	관측한 날짜

\*파란색: 기존 변수 / 빨간색: 추가 변수

### 〈가설 첫 번째〉

출퇴근시간의 오염물질의  
Max\_Value가 높게 측정될 것이다.

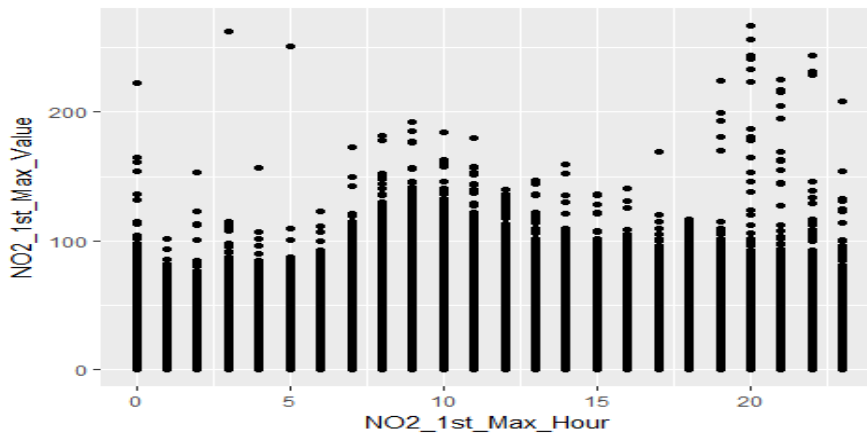
# 02 본론-가설1



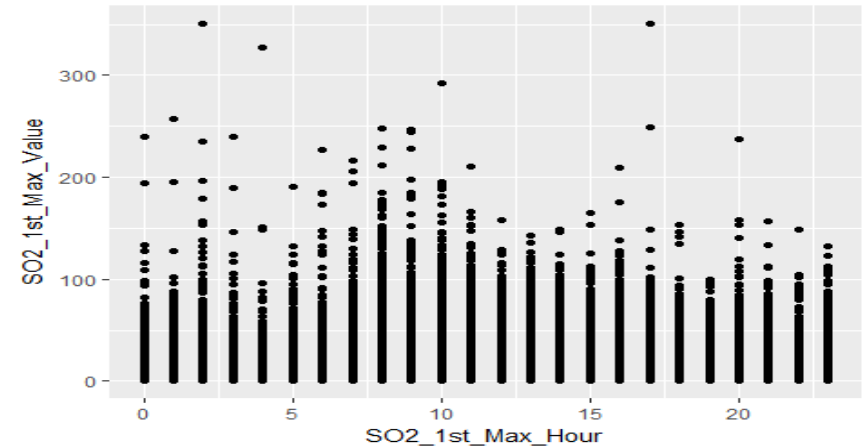
## 1. 오염물질 자체 분석

1) 가설1 "출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다."

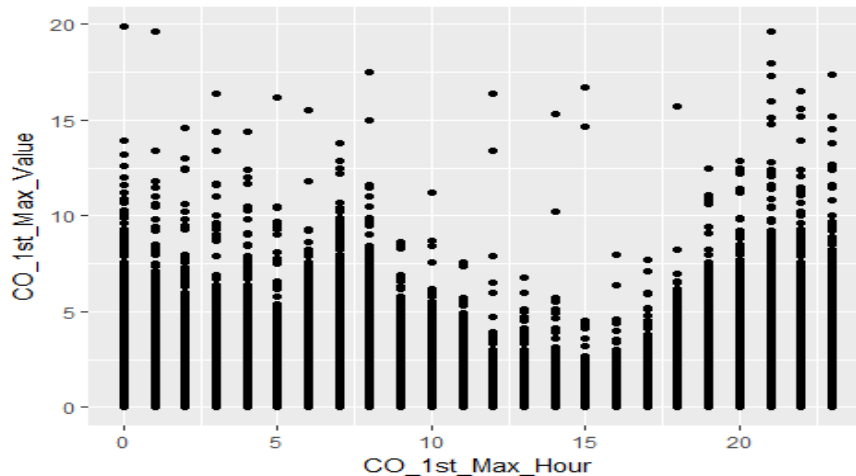
〈NO2〉 Hour 별 value 산점도



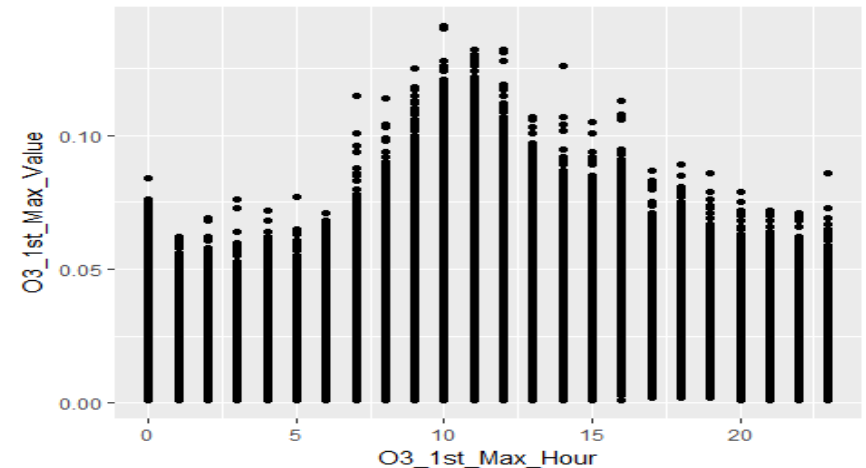
〈SO2〉



〈CO〉



〈O3〉



## 02 본론-가설1



### 1. 오염물질 자체 분석

1) 가설1 "출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다."

### Hour 별 value의 갯수

NO2_1 <sup>st</sup> _Max_Hour	Count
최소	2869
최대	39639

SO2_1 <sup>st</sup> _Max_Hour	Count
최소	8218
최대	110085

CO_1 <sup>st</sup> _Max_Hour	Count
최소	6313
최대	233666

O3_1 <sup>st</sup> _Max_Hour	Count
최소	1216
최대	113611

## 02 본론-가설1

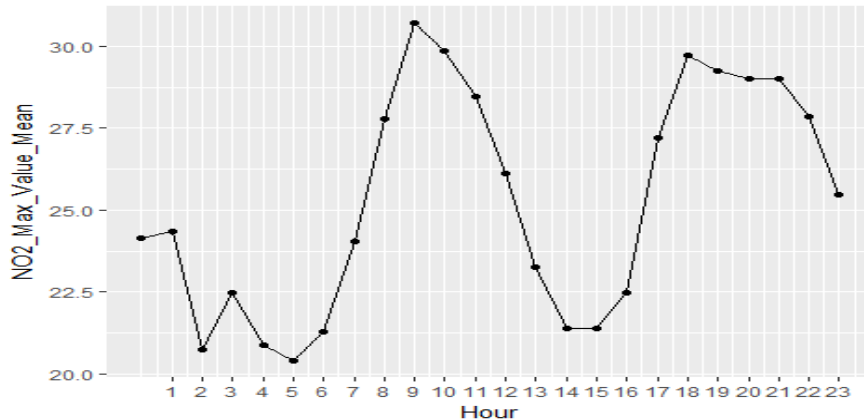
### 1. 오염물질 자체 분석

1) 가설1 “출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다.”

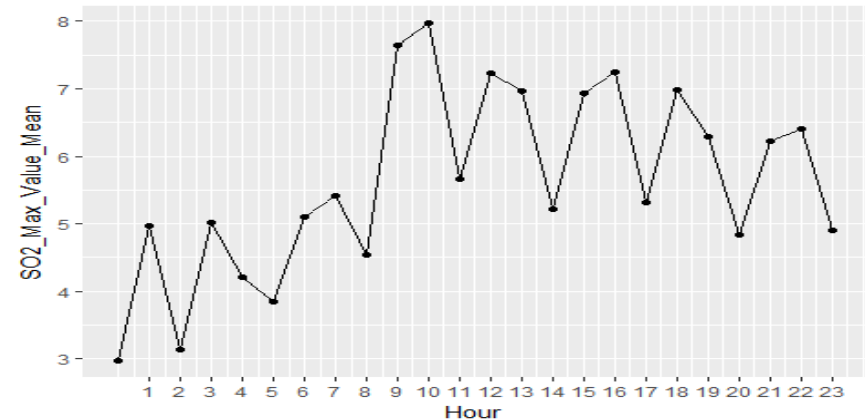


### Hour 별 value의 mean

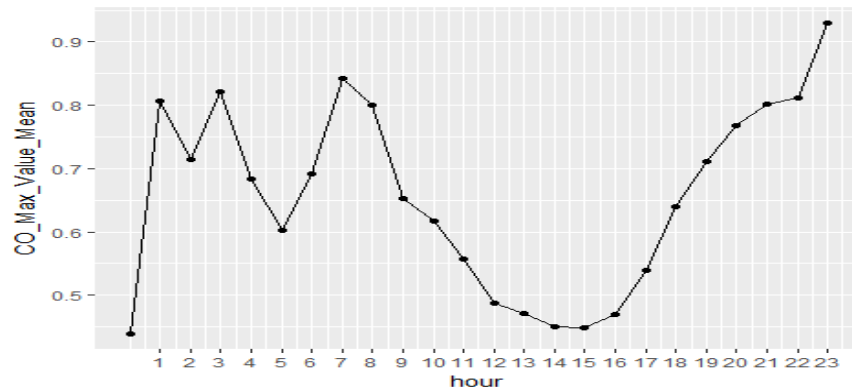
〈NO2〉 따름



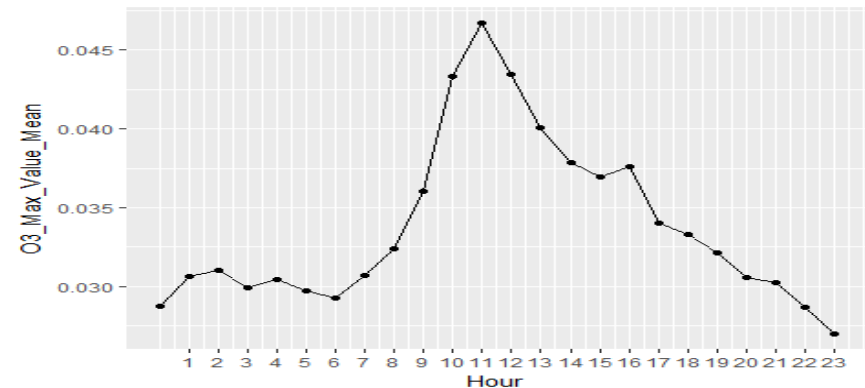
〈SO2〉 안따름



〈CO〉 따름



〈O3〉 정반대





## 02 본론-가설1

### 1. 오염물질 자체 분석

1) 가설1 “출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다.”

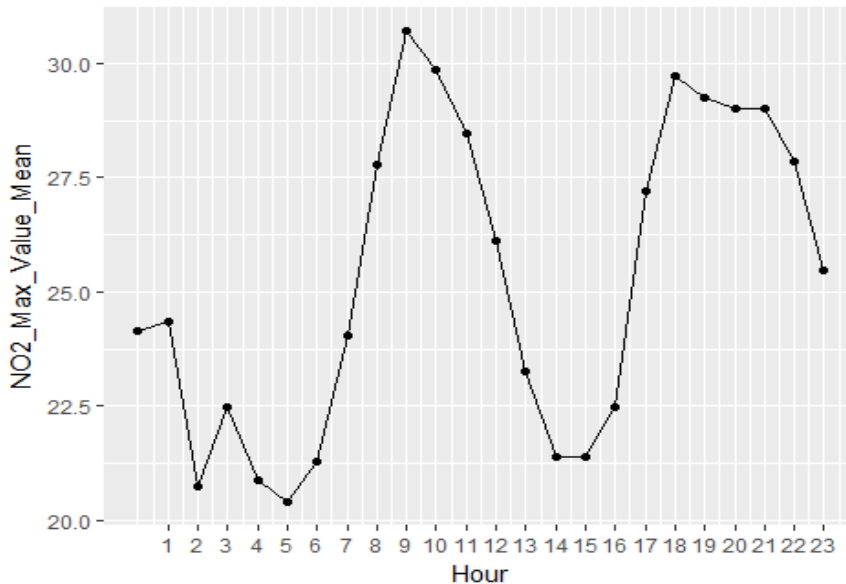
데이터 출처

[논문] 관악과 시청별 요일별 오존농도

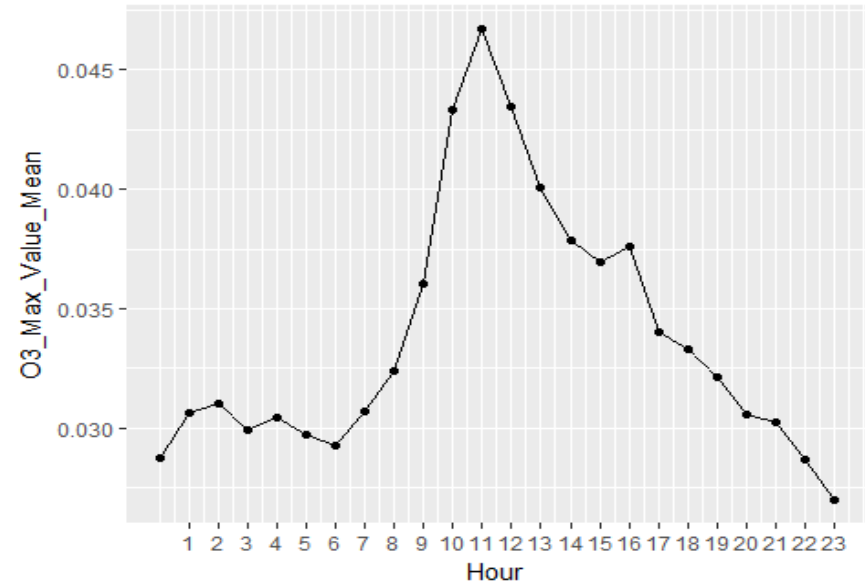
-김정화, 김용표-

### 결론 및 이유

#### 〈NO2〉



#### 〈O3〉



- NO2는 **교통량**에 영향을 많이 받기 때문에 출근시간에 높은 농도 형성
- NO2는 O3의 합성작용을 방해 -> 서로 반대의 결과 도출
- NO2는 출근시간에 농도가 **높고**, O3는 출근시간에 농도가 **낮다**.

## 02 본론-가설1

### 1. 오염물질 자체 분석

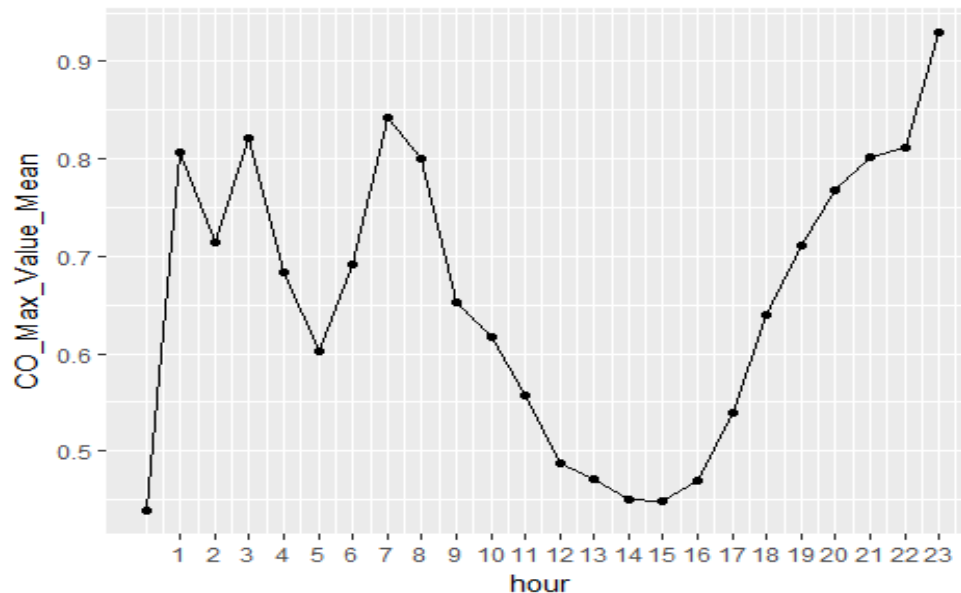
1) 가설1 "출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다."

데이터 출처

[논문] 관악과 시청별 요일별 오존농도

-김정화, 김용표-

### 결론 및 이유 〈CO〉



- CO 또한 출퇴근시간과 밤에 높은 농도를 달성하는 것을 볼 수 있음.
- CO는 **주거환경**에서 가장 높은 발생률을 보임  
→ 아침과 저녁에 높은 농도

## 02 본론-가설1

### 1. 오염물질 자체 분석

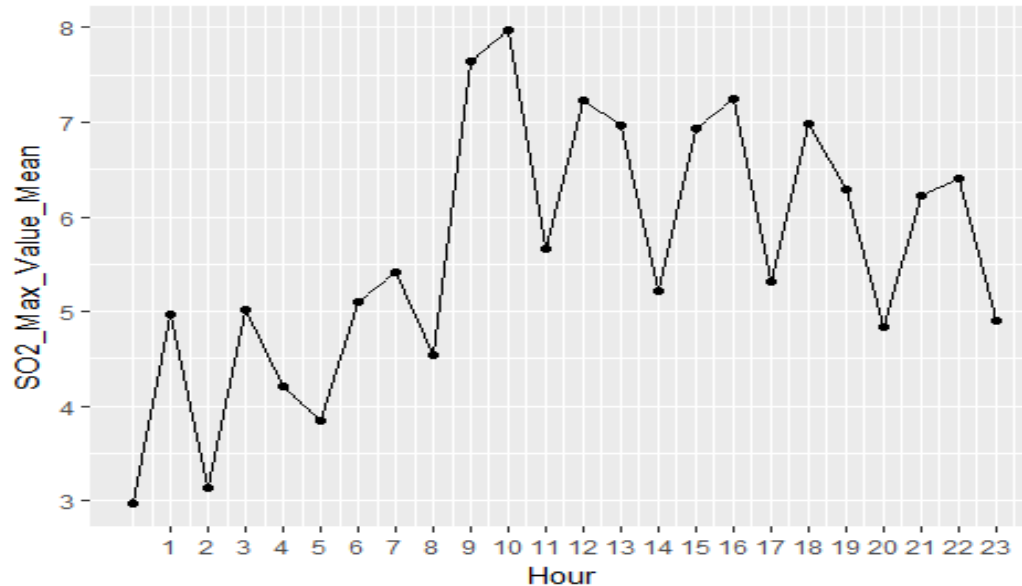
1) 가설1 "출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다."

데이터 출처

[논문] 관악과 시청별 요일별 오존농도

-김정화, 김용표-

### 결론 및 이유 〈SO2〉



- SO2는 출근시간 ~ 퇴근시간에 계속해서 높은 농도를 달성.
- SO2는 석탄, 경유, 중유 등에서 가장 높은 발생률을 보임  
→ 출퇴근시간, 노동시간에 높은 농도

## 03 본론-가설2,3

본론 소주제 소개

### 1 오염물질 자체 분석

1) 가설1



### 2 날짜별 오염물질 분석

1) 가설2

2) 가설3

### 3 위치별 오염물질 분석

1) 가설4

5) 가설5

## 03 본론-가설2,3

### 2. 날짜별 분석

\* 분석 내용 소개

- ▶ 2000년부터 2016년까지의 관측 기간 중, 2002년 6월 10일이 640건으로 가장 많았고, 2001년 2월 12일이 136건으로 가장 적었다.
- ▶ 최대값과 최소값의 차이가 약 500건  
→ 날짜별 자료분석에 있어서 mean 값을 취하여 살펴보는 것이 적절할 것이라 판단

Date_Local	Count
2002-06-10	640
2016-05-01	28

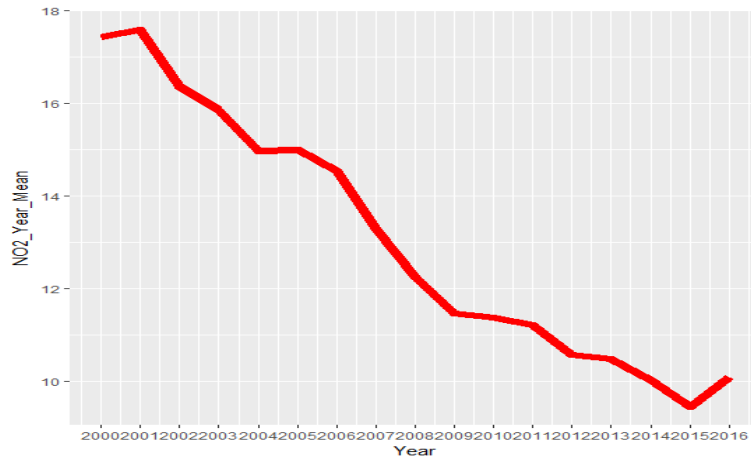
- ▶ **가설2.** 2010년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.
- ▶ **가설3.** 우리나라의 경우, 늦가을부터 겨울까지가 가장 대기오염이 심각하다고 여겨진다. 마찬가지로 미국도 늦가을부터 겨울까지의 오염물질의 농도가 높을 것이다.

# 03 본론-가설2,3 <4개의 오염물질을 Year별로 나눠서 분석>

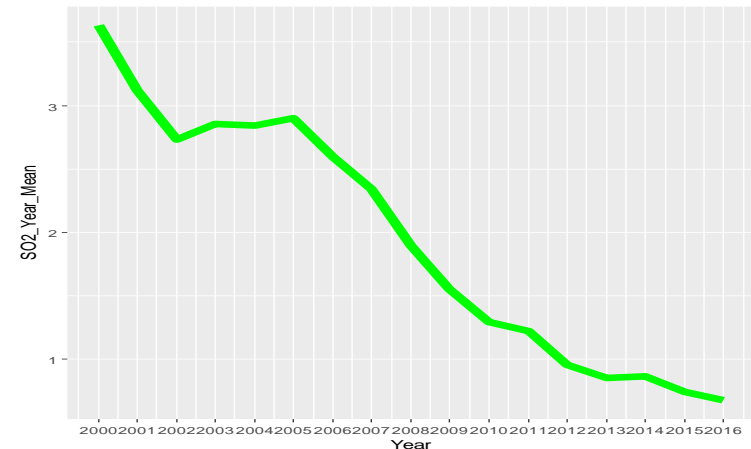
## 2. 날짜별 분석

1) 가설2 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.”

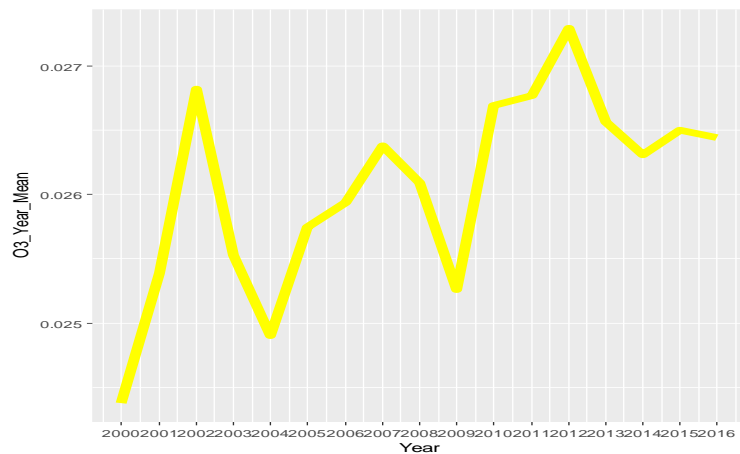
〈NO2〉



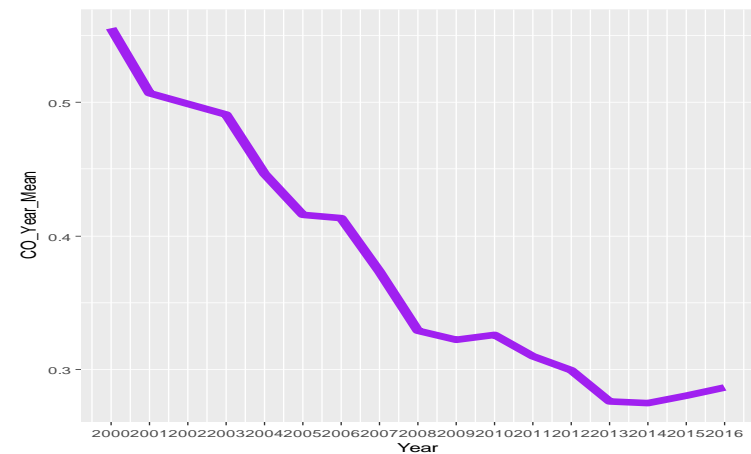
〈SO2〉



〈O3〉



〈CO〉

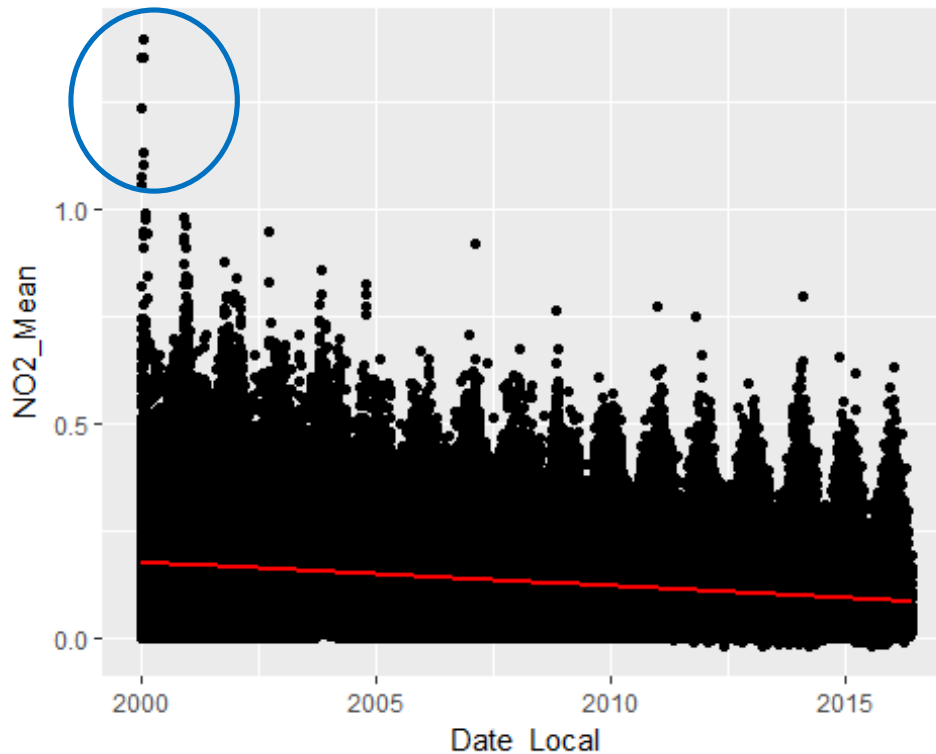


## 03 본론-가설2,3

### <NO2\_연도별>

#### 2. 날짜별 분석

1) 가설2 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.”



```
Pollution_NO2 %>% filter(NO2_Mean>0.01) %>% arrange(desc(NO2_Mean)) %>% select(Date_Local,  
NO2_Mean,everything()) %>% distinct()↵
```

```
## # A tibble: 398,622 x 16↵  
##   Date_Local NO2_Mean State_Code County_Code Site_Num↵  
##   <date>     <dbl>     <int>     <int>     <int>↵  
## 1 2000-01-19 1.3954167         4         13      3003↵  
## 2 2000-01-13 1.3533333         4         13      3003↵  
## 3 2000-01-10 1.3518750         4         13      3003↵  
## 4 2000-01-12 1.2333333         4         13      3003↵  
## 5 2000-01-15 1.1308333         4         13      3003↵  
## 6 2000-01-18 1.1013636         4         13      3003↵  
## 7 2000-01-11 1.0754546         4         13      3003↵  
## 8 2000-01-05 1.0550000         4         13      3003↵  
## 9 2000-02-04 0.9875000         4         13      3003↵  
## 10 2000-11-27 0.9813044         6         37     1103↵  
## # ... with 398,612 more rows, and 11 more variables: Address <chr>,↵  
## # State <chr>, County <chr>, City <chr>, Year <int>, Month <int>,↵  
## # Day <int>, NO2_Units <chr>, NO2_1st_Max_Value <dbl>,↵  
## # NO2_1st_Max_Hour <int>, NO2_AQI <int>↵
```

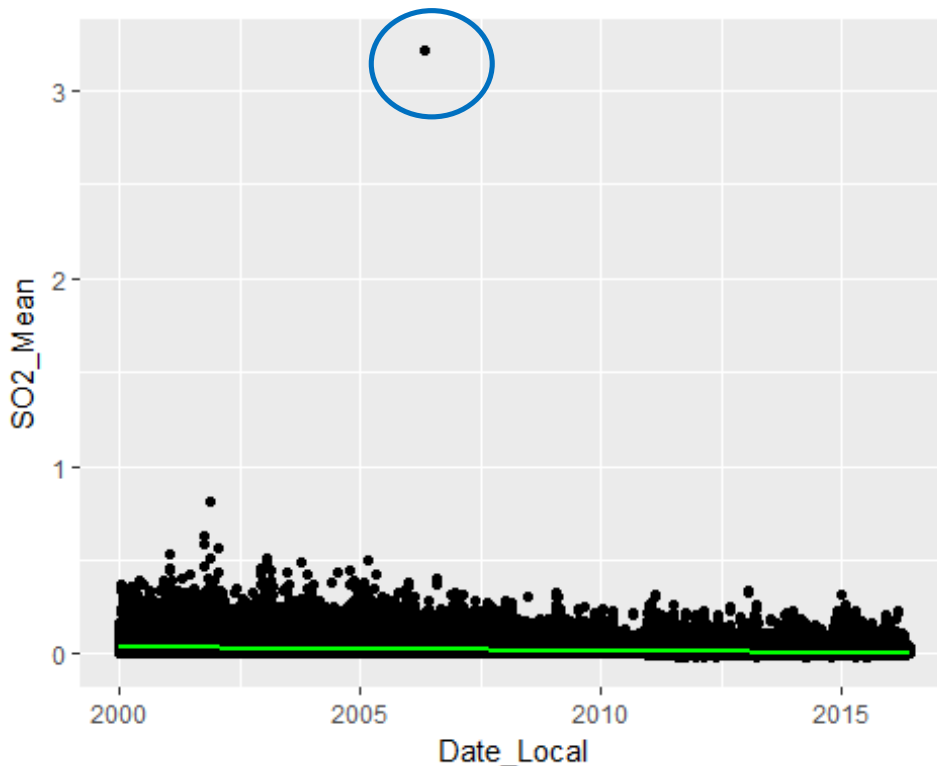
- ▶ NO2는 시간에 따라 오염물질의 평균 농도가 감소하고 있음을 확인할 수 있다.
- ▶ 또한 2000년에 NO2의 평균 농도가 매우 크게 관측된 날짜는 2000년 1월 19일이었음을 확인할 수 있다.

## 03 본론-가설2,3

### <SO2\_연도별>

#### 2. 날짜별 분석

1) 가설2 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.”



```
Pollution_SO2 %>% filter(SO2_Mean>0.03) %>% arrange(desc(SO2_Mean)) %>% distinct()
```

```
## # A tibble: 164,065 x 16↓  
##   Date_Local SO2_Mean State_Code County_Code Site_Num↓  
##   <date>     <dbl>     <int>     <int>     <int>↓  
## 1 2006-05-04 3.2162500      40         21      9002↓  
## 2 2006-05-04 3.2160000      40         21      9002↓  
## 3 2001-11-16 0.8125000      17        163       10↓  
## 4 2001-11-16 0.8121250      17        163       10↓  
## 5 2001-10-04 0.6315714      17        163       10↓  
## 6 2001-10-04 0.5856522      17        163       10↓  
## 7 2002-01-25 0.5608333      17        163       10↓  
## 8 2002-01-25 0.5603750      17        163       10↓  
## 9 2001-01-14 0.5350000      36          5       83↓  
## 10 2001-01-14 0.5346250      36          5       83↓  
## # ... with 164,055 more rows, and 11 more variables: Address  
## #   State <chr>, County <chr>, City <chr>, Year <int>, Month  
## #   Day <int>, SO2_Units <chr>, SO2_1st_Max_Value <dbl>,↓  
## #   SO2_1st_Max_Hour <int>, SO2_AQI <int>↓
```

- ▶ SO2는 평균 농도가 대략 비슷하게 분포하긴 하지만, 시간에 따라 오염물질의 평균 농도가 감소하고 있음을 확인할 수 있다.
- ▶ 또한 2006년경에 SO2의 평균 농도가 매우 크게 관측된 날짜는 2006년 5월 4일이었음을 확인할 수 있다.

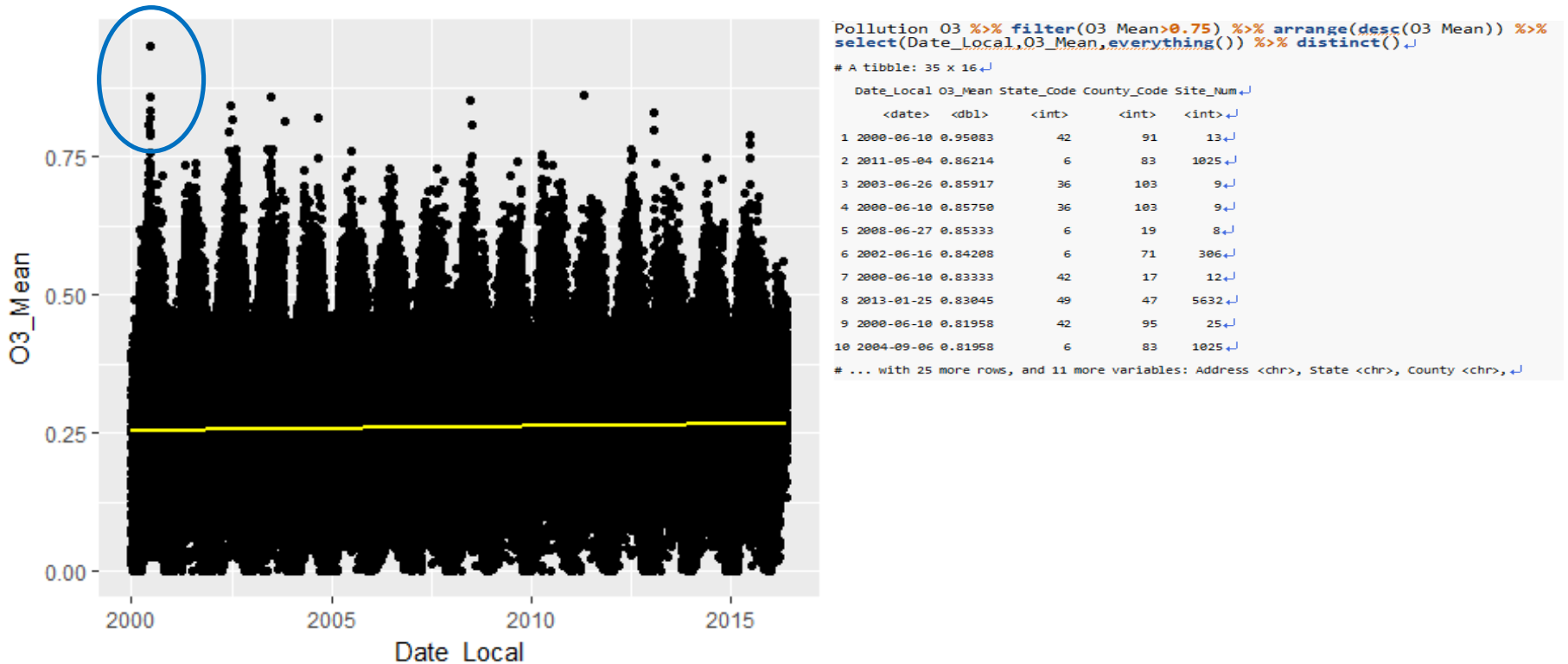


# 03 본론-가설2,3

## <O3\_연도별>

### 2. 날짜별 분석

1) 가설2 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.”



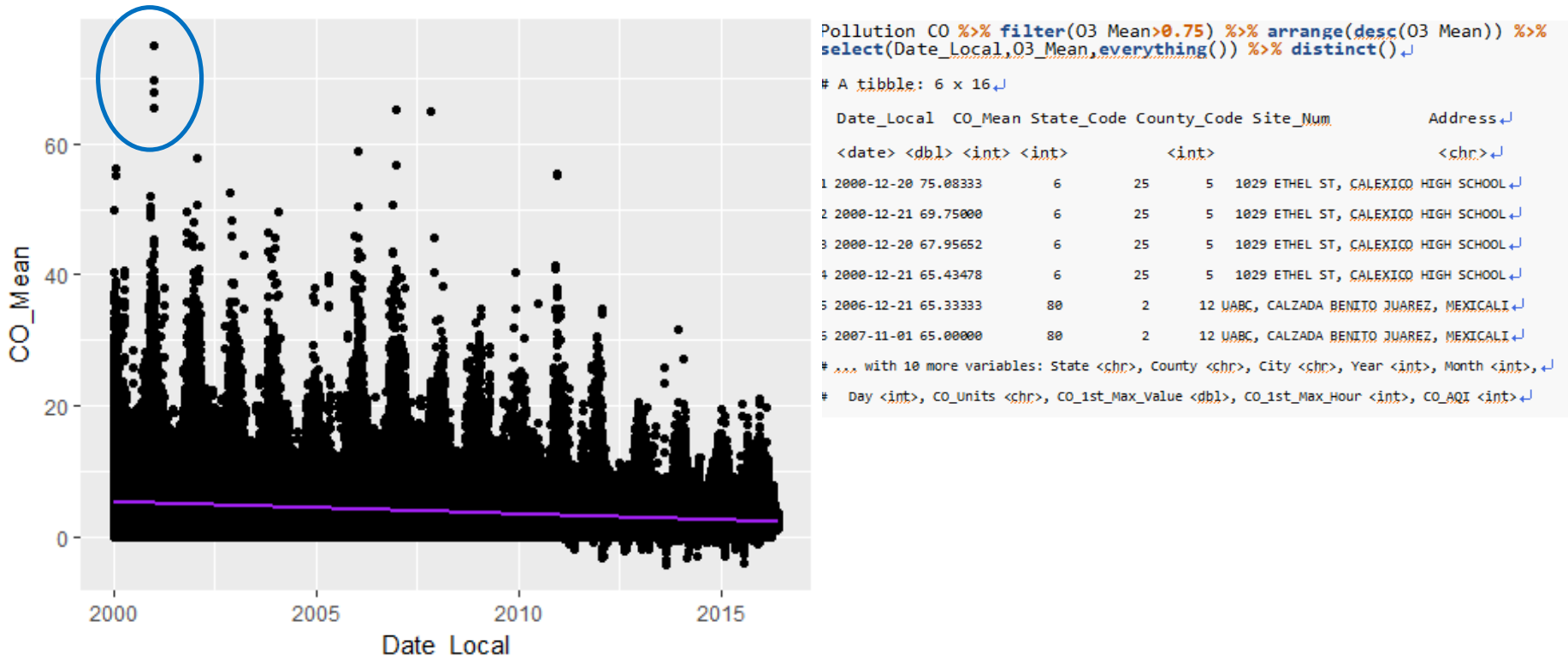
- ▶ O3는 평균 농도가 시간에 따라 증가하고 있음을 확인할 수 있다.
- ▶ 또한 2000년경에 O3의 평균 농도가 매우 크게 관측된 날짜는 2000년 6월 10일이었음을 확인할 수 있다.

## 03 본론-가설2,3

### <CO\_연도별>

#### 2. 날짜별 분석

1) 가설2 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.”



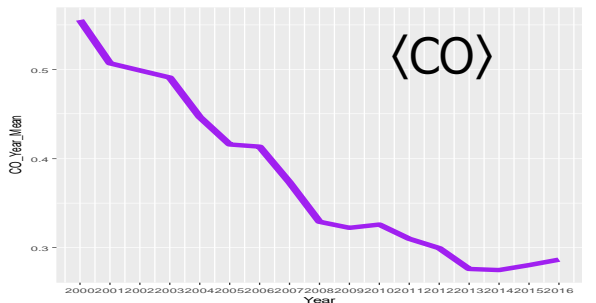
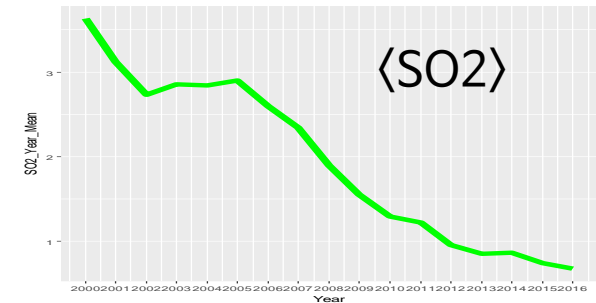
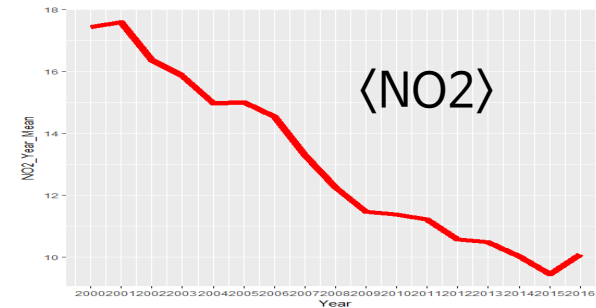
- ▶ CO는 시간에 따라 오염물질의 평균 농도가 감소하고 있음을 확인할 수 있다.
- ▶ 또한 CO의 평균 농도가 매우 크게 관측된 날짜는 2000년 12월 경이었음을 확인할 수 있다.

# 03 본론-가설2,3

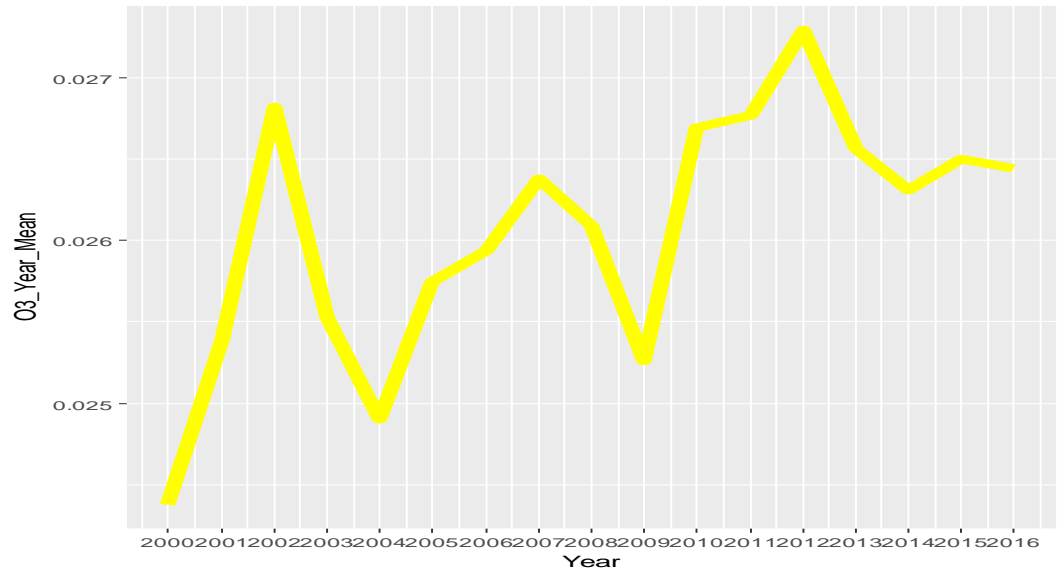
## <가설2에 대한 결론>

### 2. 날짜별 분석

1) 가설2 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다.”



### <O3>



▶ NO<sub>2</sub>, SO<sub>2</sub>, CO는 2010년에서 2016년으로 갈수록 점차 평균 농도가 **줄어 들고** 있는 반면, **O<sub>3</sub>**의 평균 농도는 점차 **증가**하고 있다.

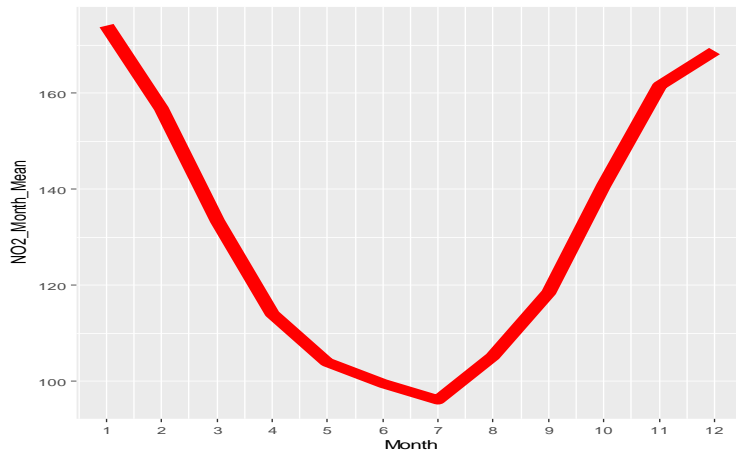
▶ 따라서 “2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것”이라는 가설은 O<sub>3</sub>에 대해서만 가능성이 인정되고 나머지 세 오염물질에 대해서는 오히려 감소할 것이라는 예측이 가능하므로, **가설은 기각된다.**

## 03 본론-가설2,3 <4개의 오염물질을 Month별로 나눠서 분석>

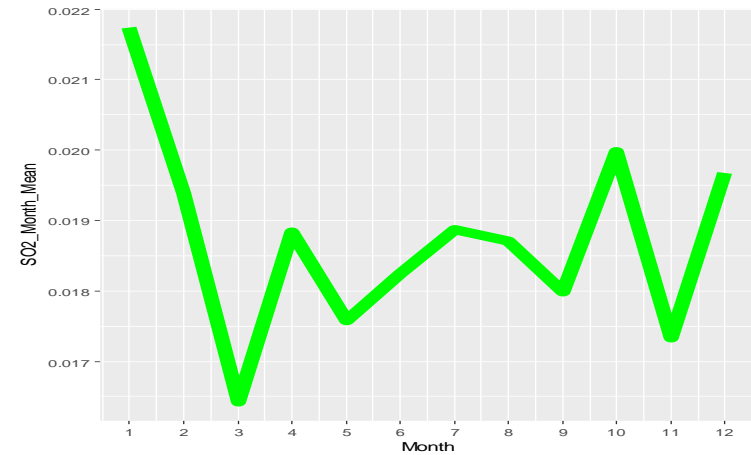
### 2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”

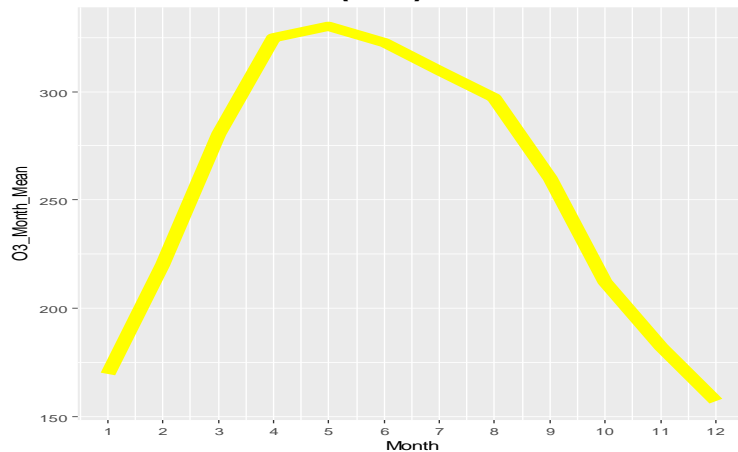
〈NO2〉



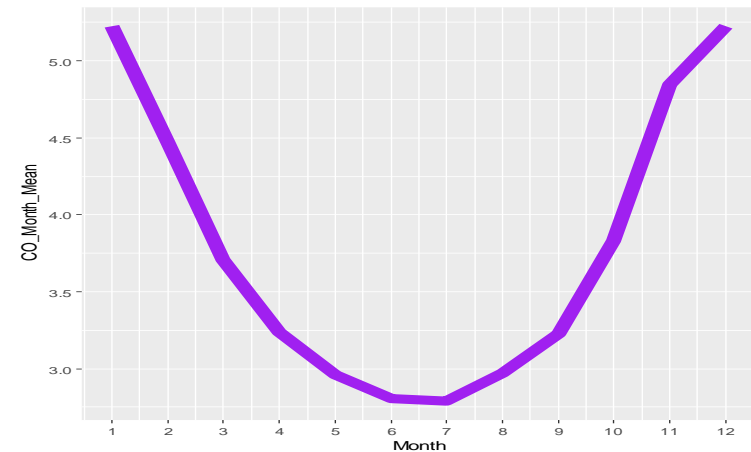
〈SO2〉



〈O3〉



〈CO〉



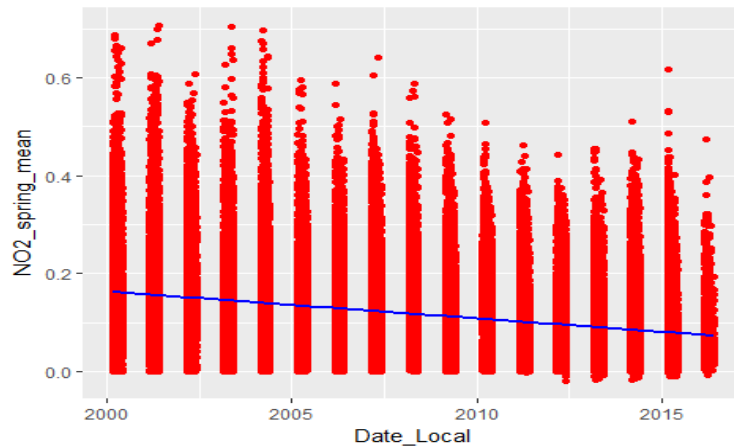
# 03 본론-가설2,3

## <NO2\_계절별>

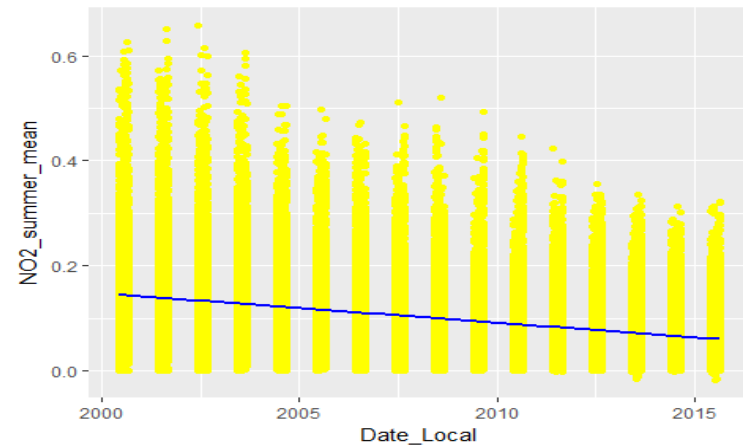
2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”

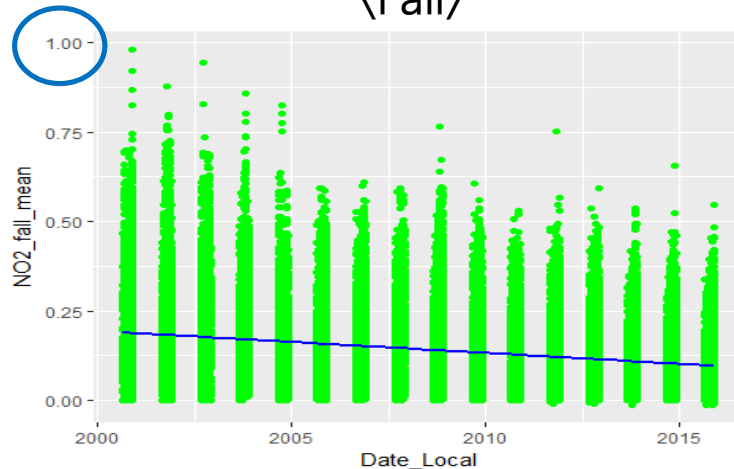
〈Spring〉



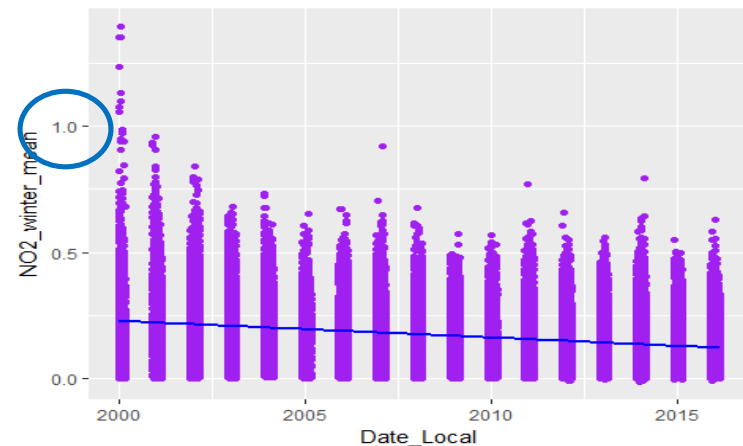
〈Summer〉



〈Fall〉



〈Winter〉



▶ NO2는 가을과 겨울철에 가장 오염물질의 농도가 높은 것으로 관찰된다.

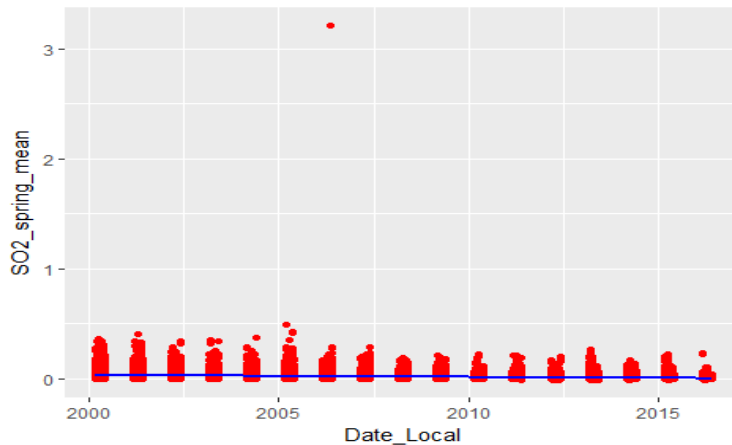
# 03 본론-가설2,3

## <SO2\_계절별>

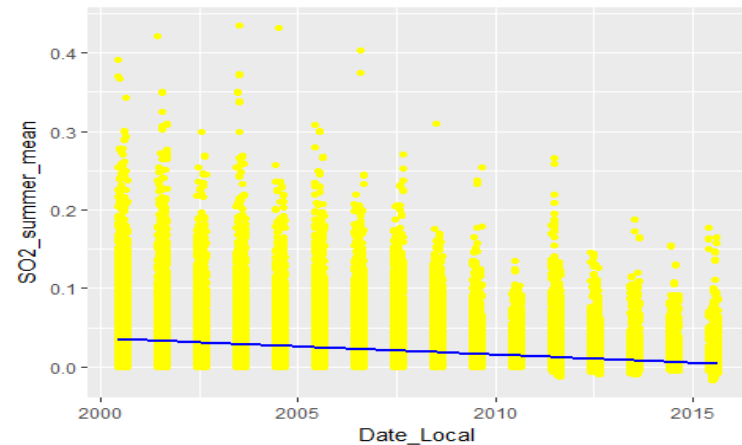
2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”

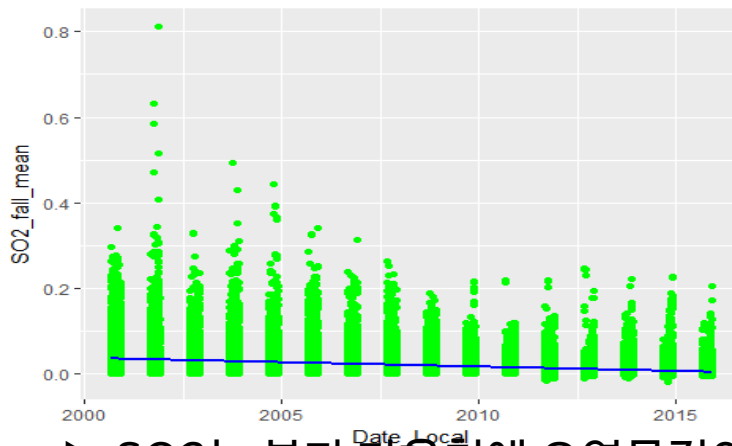
<Spring>



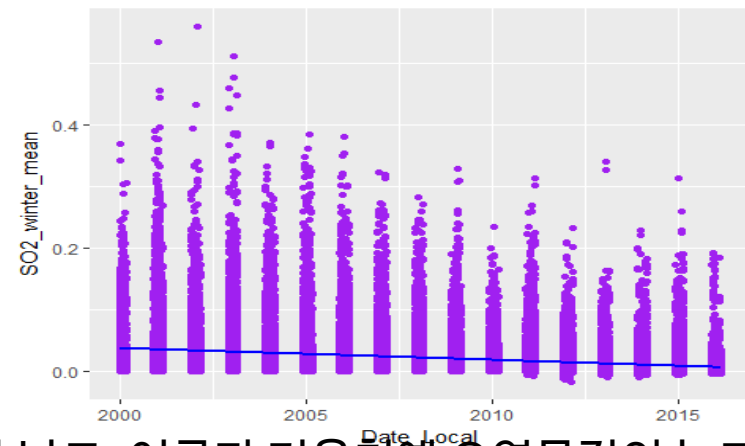
<Summer>



<Fall>



<Winter>



▶ SO2는 봄과 가을철에 오염물질의 농도가 낮고, 여름과 겨울철에 오염물질의 농도가 높은 것으로 관찰된다.

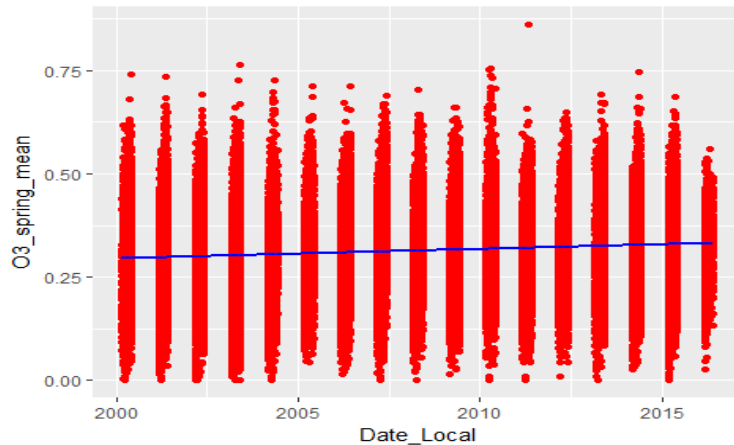
# 03 본론-가설2,3

## <O3\_계절별>

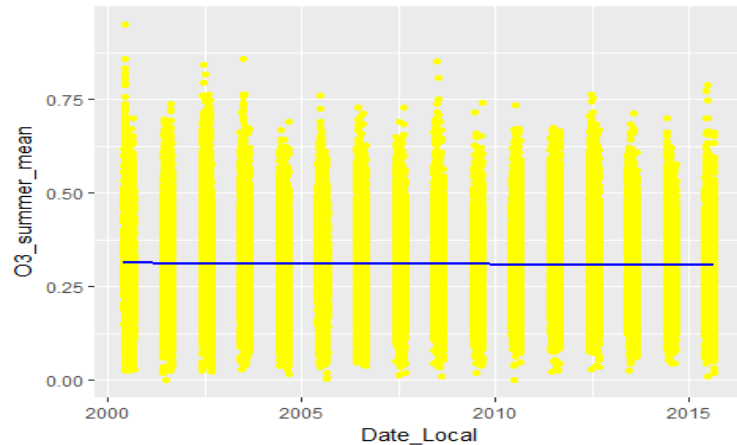
### 2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”

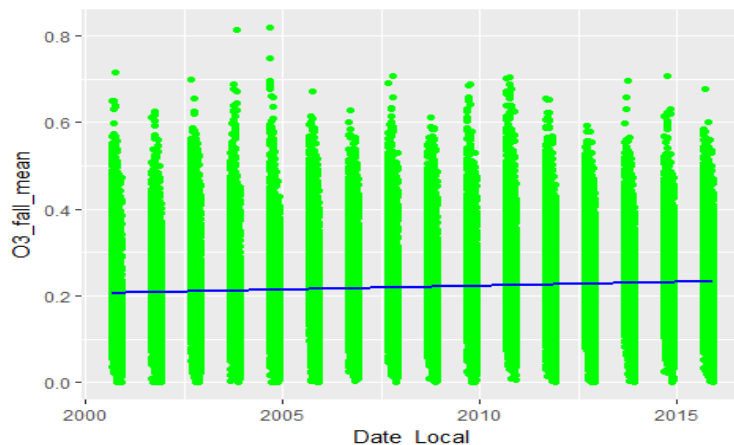
〈Spring〉



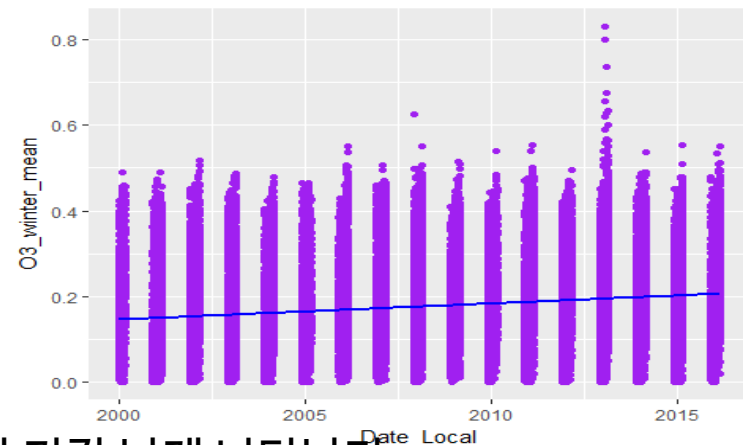
〈Summer〉



〈Fall〉



〈Winter〉



▶ O3는 겨울철에 오염물질의 평균농도가 가장 낮게 나타난다.

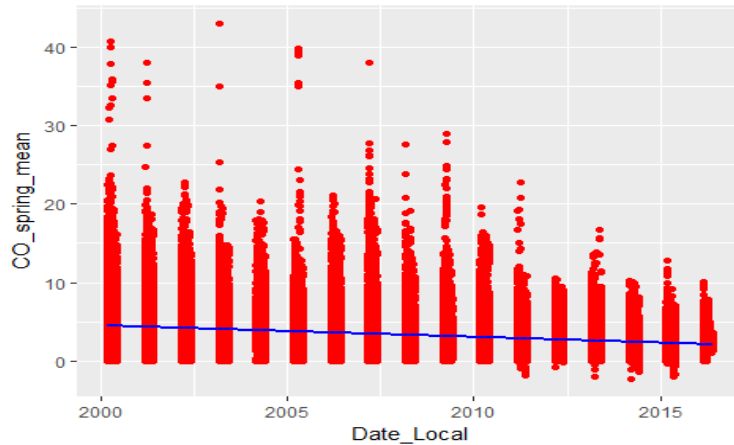
## 03 본론-가설2,3

### <CO\_계절별>

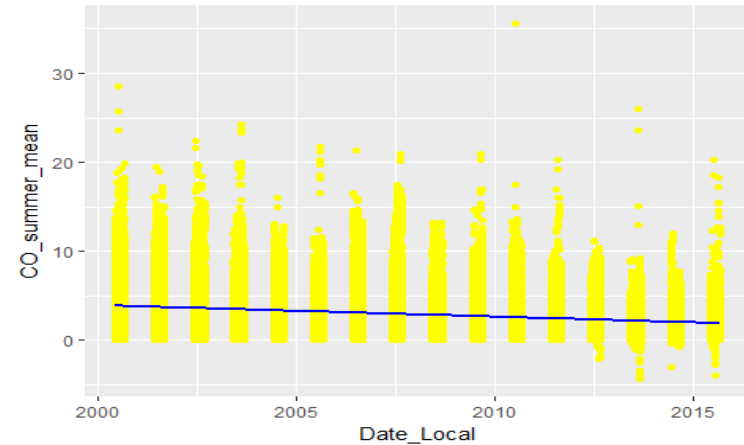
2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”

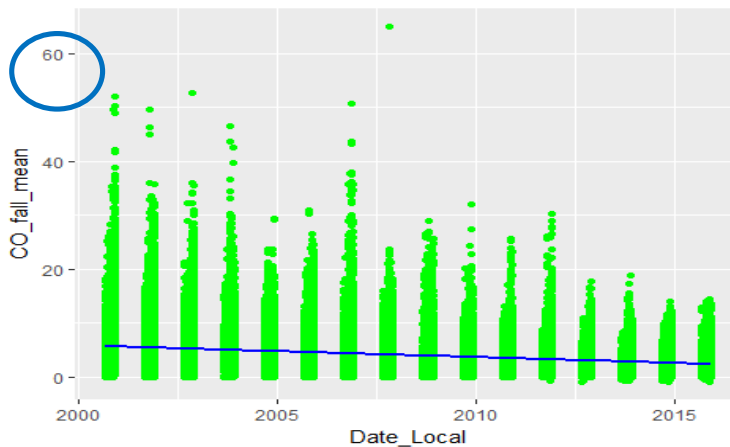
〈Spring〉



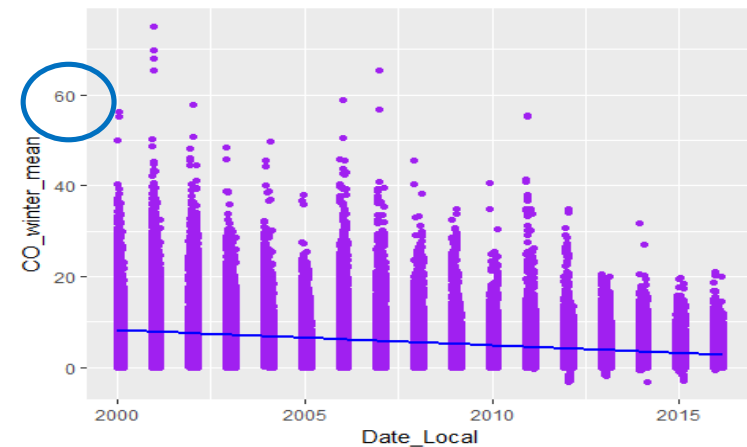
〈Summer〉



〈Fall〉



〈Winter〉



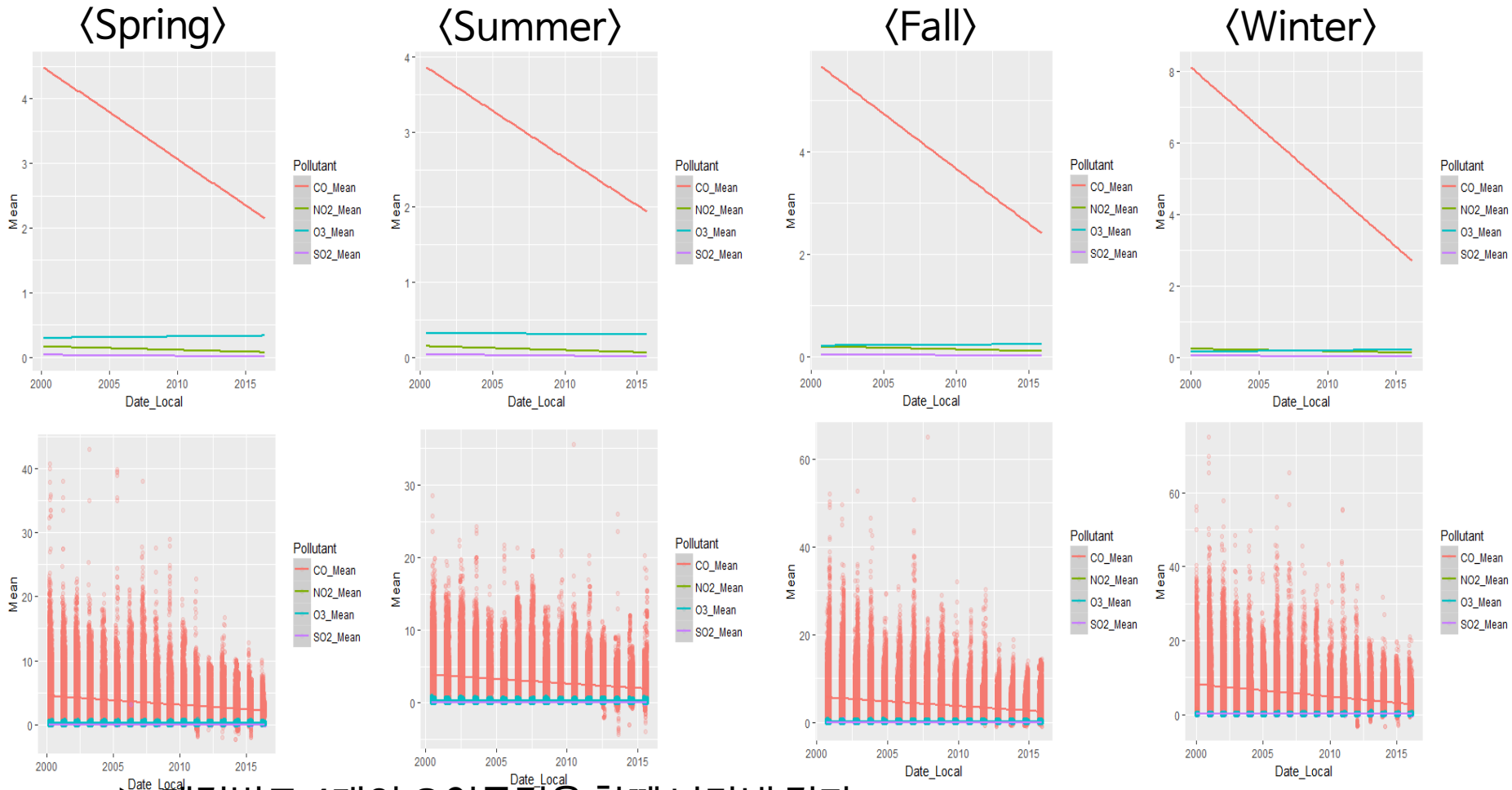
▶ CO는 가을과 겨울철에 오염물질의 평균농도가 높게 나타난다.



# 03 본론-가설2,3 <계절별-4개의 오염물질 다 함께 나타내기>

## 2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”



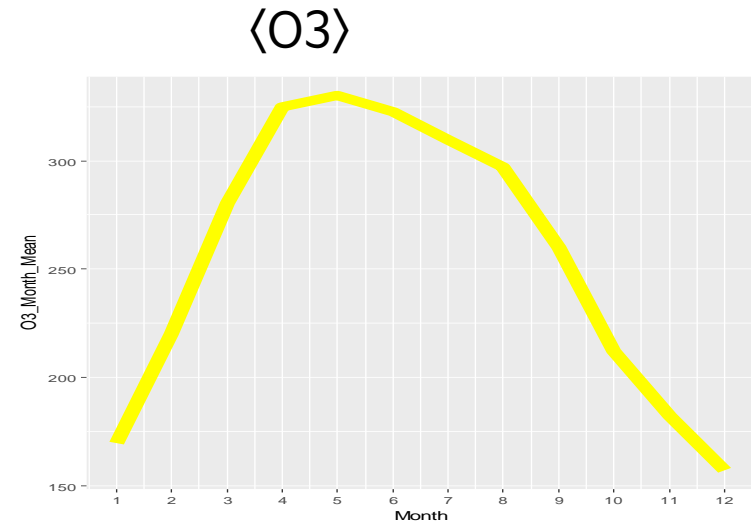
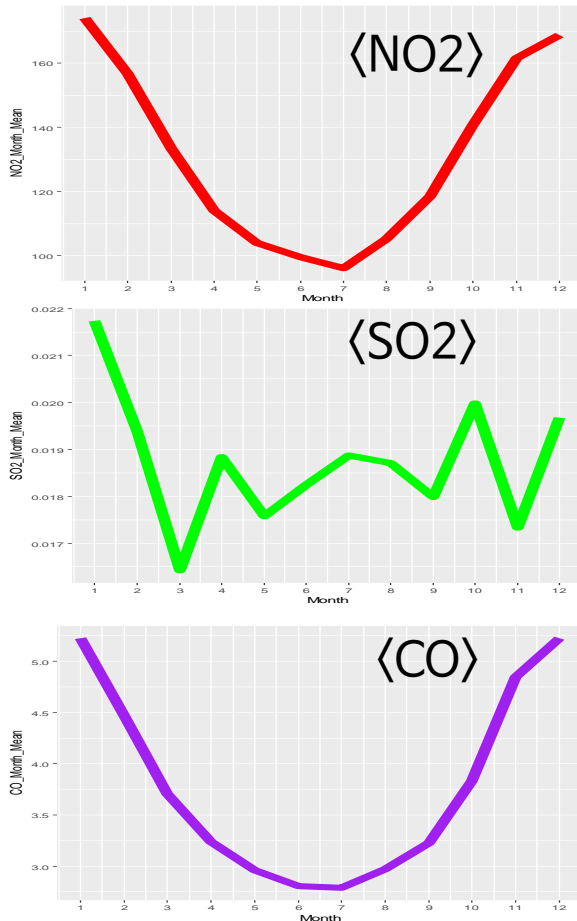
▶ 계절별로 4개의 오염물질을 함께 나타낸 결과, 공통적으로 시간이 흐름에 따라 CO의 평균 농도가 줄어들고 있었음을 확인할 수 있다.

# 03 본론-가설2,3

## <가설3에 대한 결론>

### 2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”



- ▶ NO2, SO2, CO는 여름에서 겨울로 갈수록 오염물질의 평균 농도가 **증가**하고 있는 반면, **O3**의 평균 농도는 점차 **감소**하고 있다.
- ▶ “늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것”이라는 가설은 NO2, CO, (SO2)에 대해서는 적당하지만 **O3에 대해서는 기각된다.**

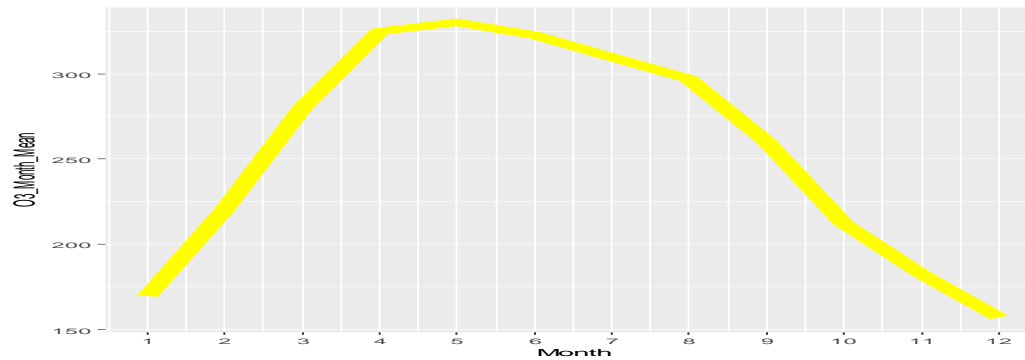
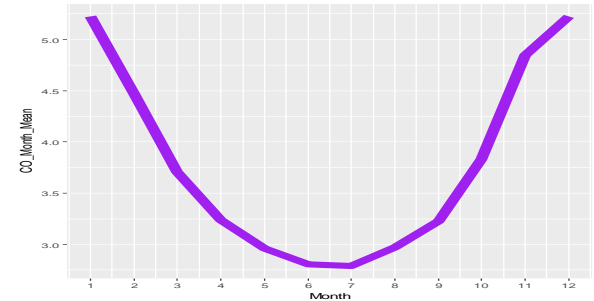
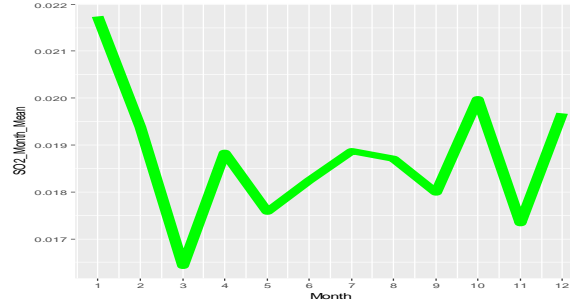
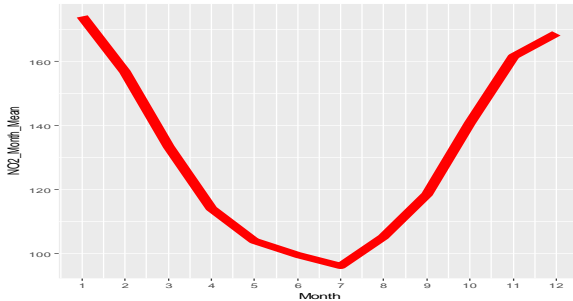
## 03 본론-가설2,3

### <가설3에 대한 결론+추가 원인분석>

#### 2. 날짜별 분석

2) 가설3 “미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다.”

- ▶ NO<sub>2</sub>, CO 그리고 O<sub>3</sub>가 서로 반대인 경향을 보인 이유는 **생성원인**일 확률이 높다.  
**NO<sub>2</sub>**가 강한 자외선을 받아 분해되고  
불안전하게 재결합 되는 **O<sub>3</sub>**는 자연적으로 여름철에 많았고  
**CO**의 가장 큰 발생장소인 거주지에서 겨울철 난방을 통해 가장 많은 양이 발생했다는  
근거자료를 통해 확인해볼 수 있다.



## 04 본론-가설4,5

본론 소주제 소개

### 1 오염물질 자체 분석

1) 가설1



### 2 날짜별 오염물질 분석

1) 가설2

2) 가설3

### 3 위치별 오염물질 분석

1) 가설4

5) 가설5

## 04 본론-가설4,5

### 3. 지역별 분석

1) 가설 4 “캘리포니아 주 안에서 공업지역(남동부)의 대기오염이 많을 것이다.”



〈가설 4〉  
캘리포니아 주 안에서  
공업지역(남동부)의 대기  
오염이 많을 것이다.

## 04 본론-가설4,5

### 3. 지역별 분석

1) 가설 4 “캘리포니아 주 안에서 공업지역(남동부)의 대기오염이 많을 것이다.”



### Mean 기준

NO<sub>2</sub> - *Burbank*(30.43051), *Bakersfield*(28.95377), *Fontana*(21.32776) 등 남부 공업도시  
SO<sub>2</sub> - 남부 지역인 *San diego*(3.306962), *Los Angeles*(1.567195) 등이 10개의 도시에 포함  
CO - *Hawthorne*(0.8577227), *Bakersfield*(0.8259681), *San diego*(0.8242527),  
*Burbank*(0.7470623) 등의 많은 도시들이 상위 10개 도시 중 포함  
O<sub>3</sub> - 남동부에 위치하는 *Victorville*(0.03338854), 중남부에 위치한 *Fresno*(0.03037633)

### AQI 기준

NO<sub>3</sub> - *San diego*(35.67786), *Bakersfield*(51.83486), *Fontana*(34.69798) 등  
남동부 지역에 위치  
SO<sub>2</sub> - *Hawthorne*(9.215022), *San diego*(8.854640), *Bakersfield*(8.633028) 등의  
많은 남부지역이 10개 도시 중 상위권에 위치  
CO - *Hawthorne*(17.165673), *Bakersfield*(16.074074) 등의  
남부 지역과 정확한 남동부 지역인 *Calexico*(14.712884) 등의 많은 도시들이 포함  
O<sub>3</sub> - 남부의 도시인 *Fontana*(55.98019), 중남부에 위치한 *Fresno*(52.80754),  
남동부에 위치한 *Victorville*(51.25039)

## 04 본론-가설4,5



### 3. 지역별 분석

1) 가설 4 “캘리포니아 주 안에서 공업지역(남동부)의 대기오염이 많을 것이다.”

NO<sub>2</sub> - California 지역 안의 NO<sub>2</sub>와 공기 오염 정도(AQI) 사이의 관계를 조사한 결과 조금의 미세한 차이는 있었지만 거의 비례하는 수준이다.

SO<sub>2</sub> - California 지역 안의 SO<sub>2</sub>와 공기 오염 정도(AQI) 사이의 관계를 조사한 결과 조금의 미세한 차이는 있었지만 거의 비례하는 수준이다.

CO - California 지역 안의 CO와 공기 오염 정도(AQI) 사이의 관계를 조사한 결과 조금의 미세한 차이는 있었지만 거의 비례하는 수준이다.

O<sub>3</sub> - California 지역 안의 O<sub>3</sub>와 공기 오염 정도(AQI) 사이의 관계를 조사한 결과 앞서 조사했던 물질들 보다는 O<sub>3</sub>와 AQI 사이의 관계가 크게 와 닿진 않았지만 아예 관련성 없는 결과는 아니다.

**결론 : 실제 높게 나온 곳 10개를 찾아서 실제 공업지역인지(캘리포니아 주 남동부, 실리콘 밸리 근처)를 살펴본 결과 10개의 모든 주가 그곳에 속하진 않았지만 과반수의 지역이 속해 있었다.**

## 04 본론-가설4,5

### 3. 지역별 분석

2) 가설 5 “공업지역이 많은 주가 대기오염이 높을 것이다.”



〈가설 5〉  
공업지역이 많은 주가  
대기오염이 높을 것이다.



## 04 본론-가설4,5



### 3. 지역별 분석

2) 가설 5 “공업지역이 많은 주가 대기오염이 높을 것이다.”

### Mean 기준

NO<sub>2</sub> - *Colorado(19.71049), Arizona(19.09904), New York(18.99439), Massachusetts(18.62205)*  
SO<sub>2</sub> - 북부의 *Alaska(6.092910), New York(4.819131), Kentucky(3.800549)*  
CO - *Arizona(0.4885040), Missouri(0.4684906), California(0.4578803), Colorado(0.4451940)*  
O<sub>3</sub> - *Utah(0.03206618), Nevada(0.03199683)*

### AQI 기준

NO<sub>3</sub> - *Arizona(36.16699), Colorado(36.07811), Michigan(31.89824)*  
SO<sub>2</sub> - *Michigan(19.20731), Missouri(18.59471), Ohio(16.06159)*  
CO - CO mean값과 거의 순서가 일치, *Arizona(9.174156), Colorado(7.760302), California(7.606910), Missouri(7.438742)*  
O<sub>3</sub> - *Utah(42.23614), Missouri(41.97449), Indiana(41.59765)*

## 04 본론-가설4,5



### 3. 지역별 분석

2) 가설 5 - 공업지역이 많은 주가 대기오염이 높을 것이다.

NO<sub>2</sub> - NO<sub>2</sub> 조사결과에서는 공업지역의 주의 수와 대기오염도가 큰 상관관계를 가졌다.

SO<sub>2</sub> - 반면 SO<sub>2</sub> 조사결과에서는 공업지역의 주의 수와 대기오염도와는 관계를 지을 만큼 관련도가 크지 않았다.

CO - CO 조사 결과 공업지역의 수와 대기오염도는 거의 비례한다고 볼 수 있다.

O<sub>3</sub> - O<sub>3</sub> 조사결과에서는 공업지역의 주의 수와 대기오염도와는 관계를 지을 만큼 관련도가 크지 않았다.

**결론 : 실제 농도가 높게 나온 주 10개 찾아서 대조해본 결과 대다수가 미국 남부 지방의 공업지역이었다. (검색 결과, 미국 공업지역은 대부분 오대호 연안/선벨트지역-노스캐롤라이나 주~남부 캘리포니아, 북쪽 지역이 많음)**

## 04 본론-가설4,5

### 3. 지역별 분석

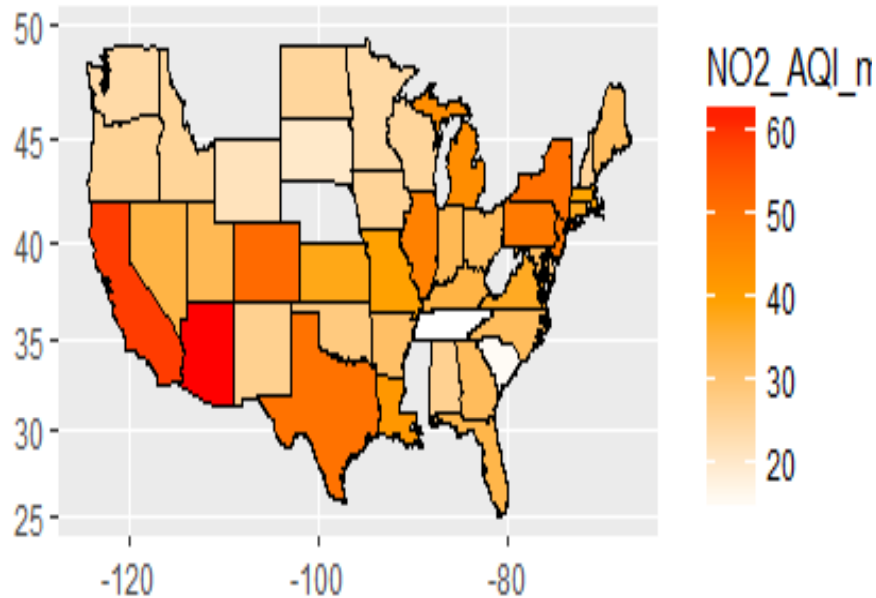
2) 가설 5 “공업지역이 많은 주가 대기오염이 높을 것이다.”



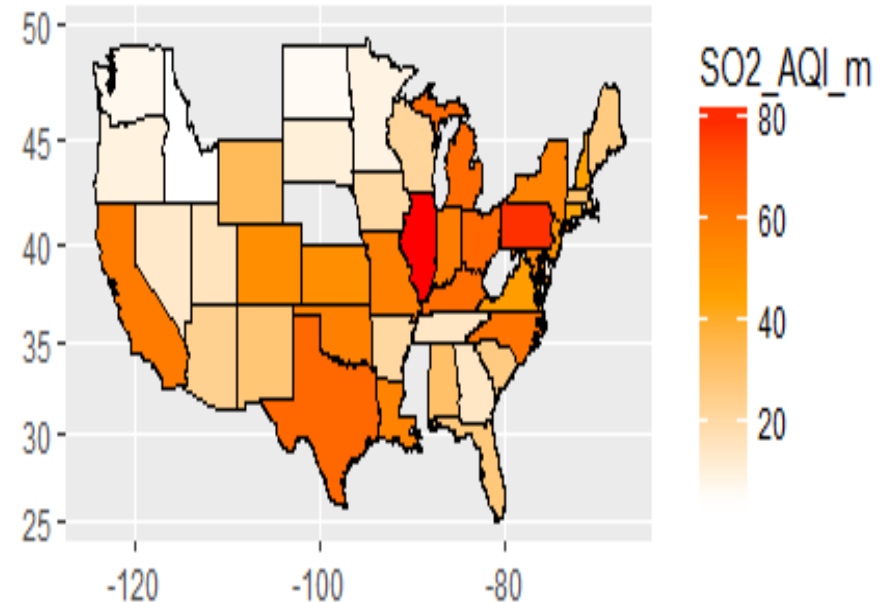
## 04 본론-가설4,5

### 3. 지역별 분석

2) 가설 5 “공업지역이 많은 주가 대기오염이 높을 것이다.”



1. Country Of Mexico (37.96057)
2. Arizona (36.16699),
3. Colorado (36.07811),



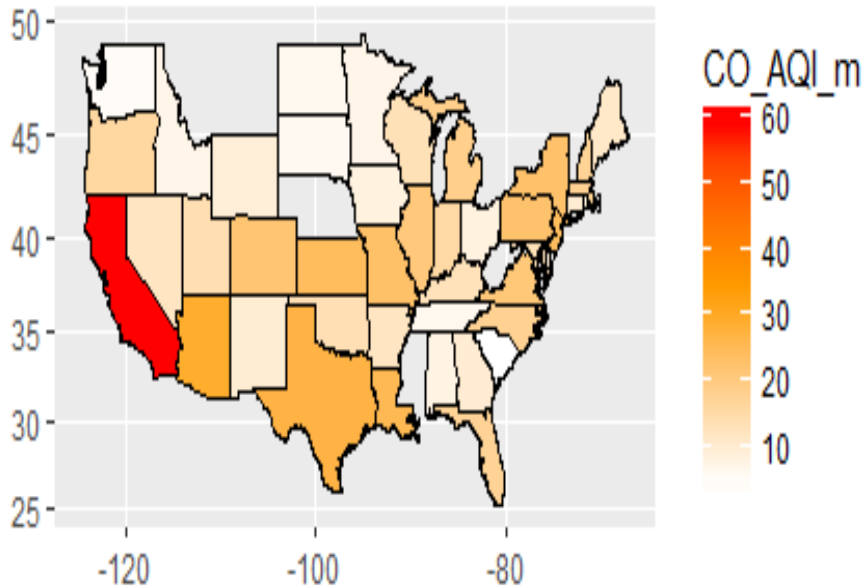
1. Michigan (19.20731),
3. Missouri (18.59471),
5. Ohio (16.06159)

## 04 본론-가설4,5

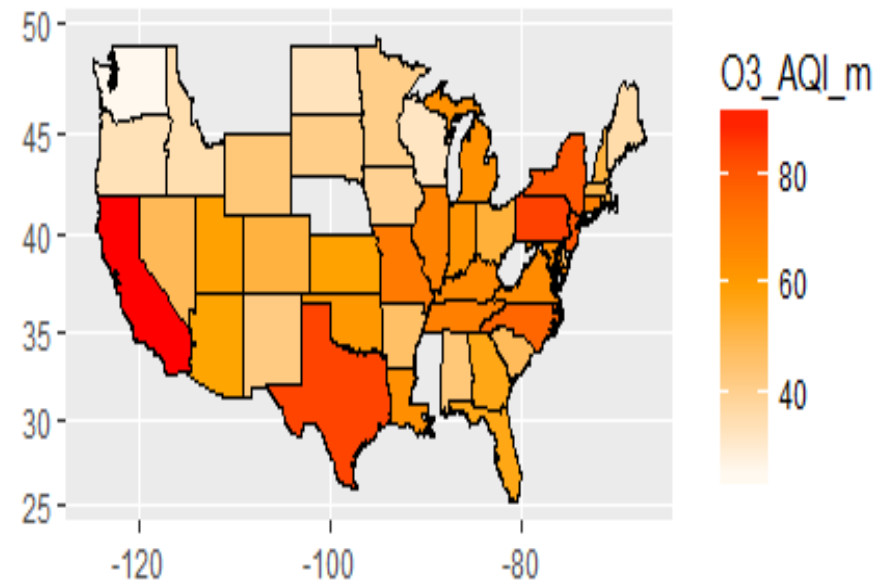


### 3. 지역별 분석

2) 가설 5 “공업지역이 많은 주가 대기오염이 높을 것이다.”



3. Arizona(9.174156),
4. Colorado(7.760302),
5. California(7.606910),
6. Missouri(7.438742)



4. Utah(42.23614),
5. Missouri(41.97449),
6. Indiana(41.59765)

## 05 결론



**가설 1.** "출퇴근시간의 오염물질의 Max\_Value가 높게 측정될 것이다."

▶ NO2와 CO만 출퇴근 시간의 오염물질이 높게 측정된다.

**가설 2.** "2000년에서 2016년으로 갈수록 오염물질의 농도가 높아질 것이다."

▶ NO2, SO2, CO는 점차 평균 농도가 줄어 들고 있는 반면, O3의 평균 농도는 점차 증가하고 있다.

**가설3.** "미국의 늦가을부터 겨울까지의 오염물질의 농도가 가장 높을 것이다."

▶ NO2, SO2, CO는 여름에서 겨울로 갈수록 오염물질의 평균 농도가 증가하고 있는 반면, O3의 평균 농도는 점차 감소하고 있다.

**가설 4.** "캘리포니아 주 안에서 공업지역(남동부)의 대기오염이 많을 것이다."

▶ 대기오염이 가장 높은 10개의 주 중 과반수의 지역이 공업지역(남동부)에 속해 있었다.

**가설 5.** "공업지역이 많은 주가 대기오염이 높을 것이다."

▶ 실제 오염 물질의 농도가 높게 나온 주 10개를 찾아 대조해본 결과 대다수가 미국 남부 지방의 공업지역이었다.

---

---

감사합니다♥

---

이상 4조의 발표를 마칩니다.  
한학기동안 정말 수고하셨습니다!