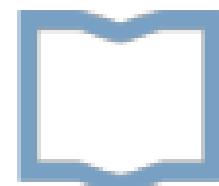


Grupo 11

Air Pollution in Seoul

Perez, Micaela
Fernandez , Silvia
De Mora , Nahuel
Oses , Joaquin



Licenciatura en
Ciencia de Datos
ECYT_UNSAM

Indice

01 Introducción

02 Objetivo

03 Limpieza

06 Análisis
exploratorio

10 Modelado

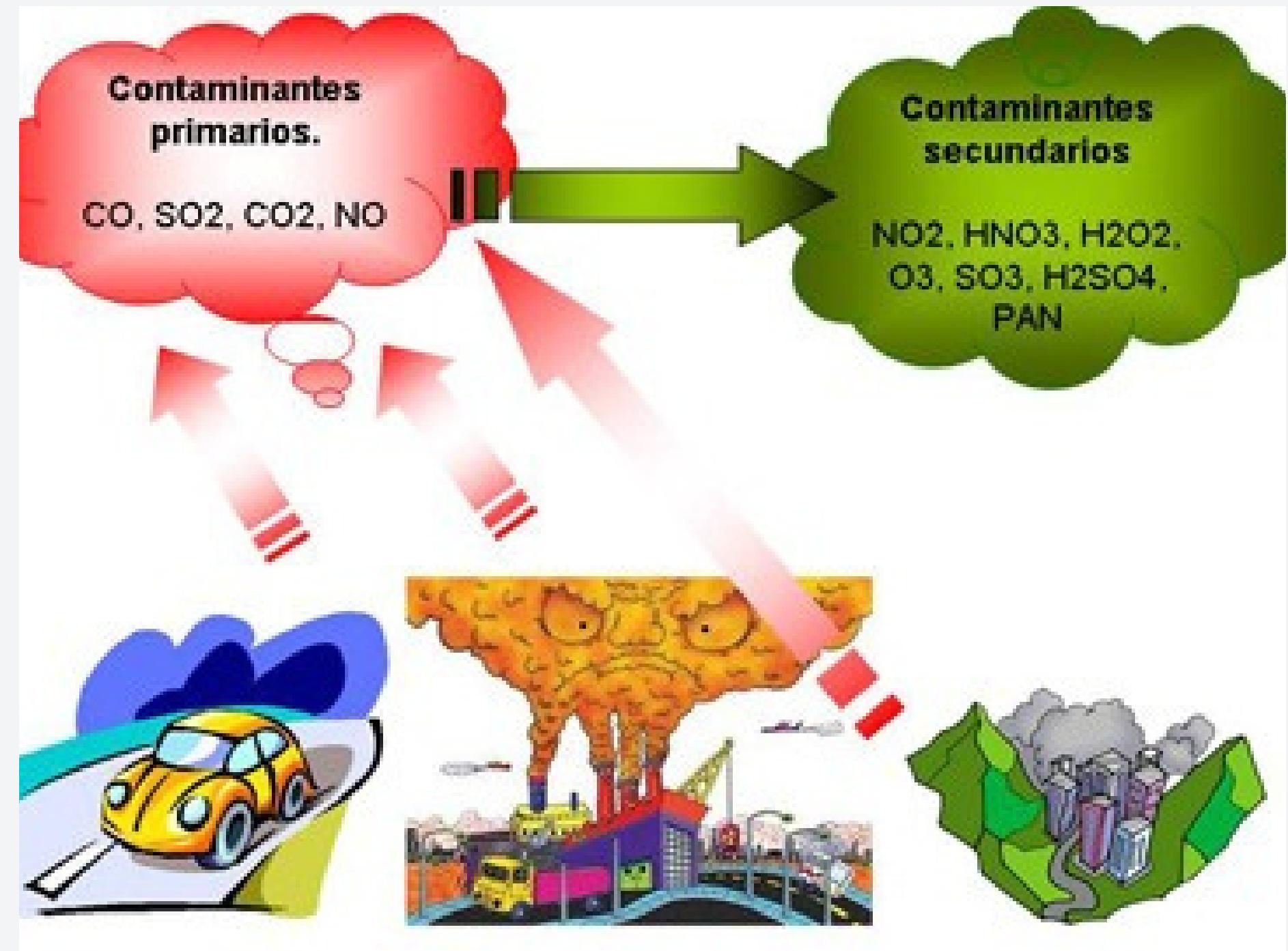
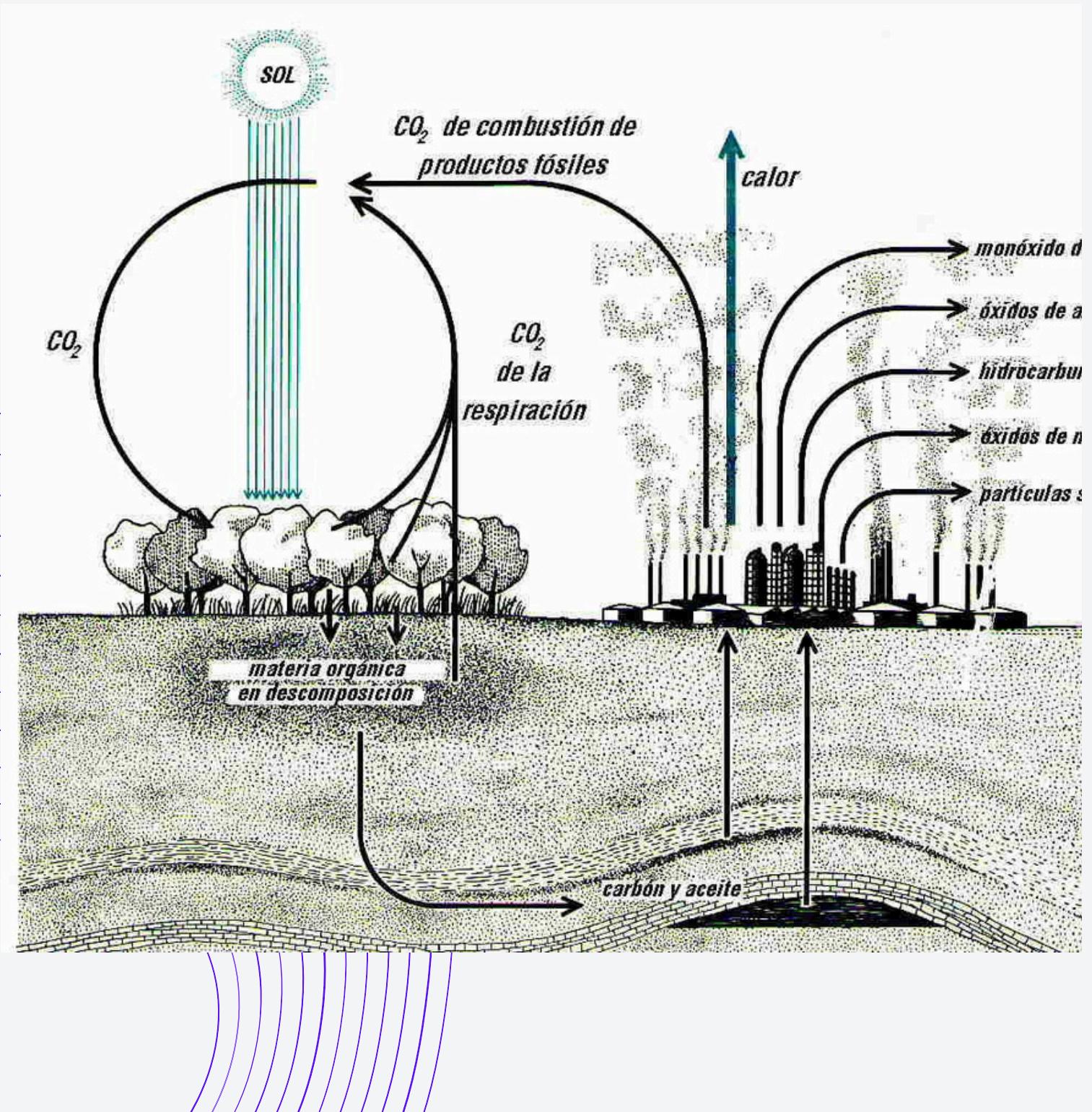
12 Conclusiones

15 Anexo



Introducción

Emisión de gases contaminantes





OBJETIVO

Lo que intentaremos responder

¿Existe una relación entre el incremento de las emisiones de NO₂, CO, SO₂ y O₃ en la región de Seúl a lo largo de las temporadas entre los años 2017 y 2019?



Limpieza

<https://www.kaggle.com>

Measurement_info.csv (124.38 MB):

Measurement date	Station code	Item code	Average value	Instrument status
1 2017-01-01	101	1	0.004	0
2 2017-01-01	101	3	0.059	0

Filas: 3,885,066

Columnas: 5

Measurement_item_info.csv (289 B)

Filas: 6
Columnas: 7

Item code	Item name	Unit of measurement	Good(Blue)	Normal(Green)	Bad(Yellow)	Very bad(Red)
1 1	SO2	ppm	0.02	0.05	0.15	1.0
2 3	NO2	ppm	0.03	0.06	0.20	2.0
3 5	CO	ppm	2.00	9.00	15.00	50.0

Measurement_station_info.csv (2.69 KB)

Station code	Station name(district)	Address	Latitude	Longitude
1 101	Jongno-gu	19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republic of Korea	37.57202	127.0050
2 102	Jung-gu	15, Deoksugung-gil, Jung-gu, Seoul, Republic of Korea	37.56426	126.9747
3 103	Yongsan-gu	136, Hannam-daero, Yongsan-gu, Seoul, Republic of Korea	37.54003	127.0049

Filas: 25

Columnas: 5

LIMPIEZA DE DATOS

- Cambio de nombre a las columnas

Fecha_hora	estacion_id	sust_quimica_id	promedio	estado_instrumento
2017-01-01	101	1	0.004	0

- Se separó la columna de fecha de medición en año, mes, día, hora, minuto, segundo,

estacion_id	sust_quimica_id	promedio	estado_instrumento	year	month	day	hrs

- Se cambiaron tipo de dato erróneo, se filtró solo gases y se las cambio a columnas

estacion_id	estado_instrumento	year	month	day	hrs	estaciones	SO2	NO2	CO	03
1	101	0	2017	1	1	0	0.004	0.059	1.2	0.0

- Con estacion_id como key, se hizo inner join

- No contenían datos NA .
- Se eliminaron los números negativos .



Análisis exploratorio

Filas: 2,558,092

Columns: 15

Numericas

Categóricas

Análisis para el modelado

Filas: 643,210

Columns: 16

```
$ year      <fct> 2017, 2017, 2017, 2017, 2017, 2017, 2017, 2017, 2017, 2017, 2017, ...
$ month     <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
$ day       <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
$ hour      <fct> 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, ...
$ station_code <fct> 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, ...
$ item_code  <fct> 1, 3, 5, 6, 1, 3, 5, 6, 1, 3, 5, 6, 1, 3, 5, 6, 1, 3, 5, 6, ...
$ item_name   <chr> "SO2", "NO2", "CO", "O3", "SO2", "NO2", "CO", "O3", "SO2", "O3", ...
$ average     <dbl> 0.004, 0.059, 1.200, 0.002, 0.004, 0.058, 1.200, 0.002, 0.00...
$ region      <chr> "Noreste", "Noreste", "Noreste", "Noreste", "Noreste", "Nore...
$ season      <chr> "Invierno", "Invierno", "Invierno", "Invierno", "Invierno", ...
$ station_name <chr> "Jongno-gu", "Jongno-gu", "Jongno-gu", "Jongno-gu", "Jongno-...
$ Latitude    <dbl> 37.57202, 37.57202, 37.57202, 37.57202, 37.57202, 37.57202, ...
$ time_day    <chr> "Madrugada", "Madrugada", "Madrugada", "Madrugada", "Madru...
$ Longitude   <dbl> 127.005, 127.005, 127.005, 127.005, 127.005, 127.005, 127.00...
$ address     <chr> "19, Jong-ro 35ga-gil, Jongno-gu, seoul, Republic of Korea", ...
```

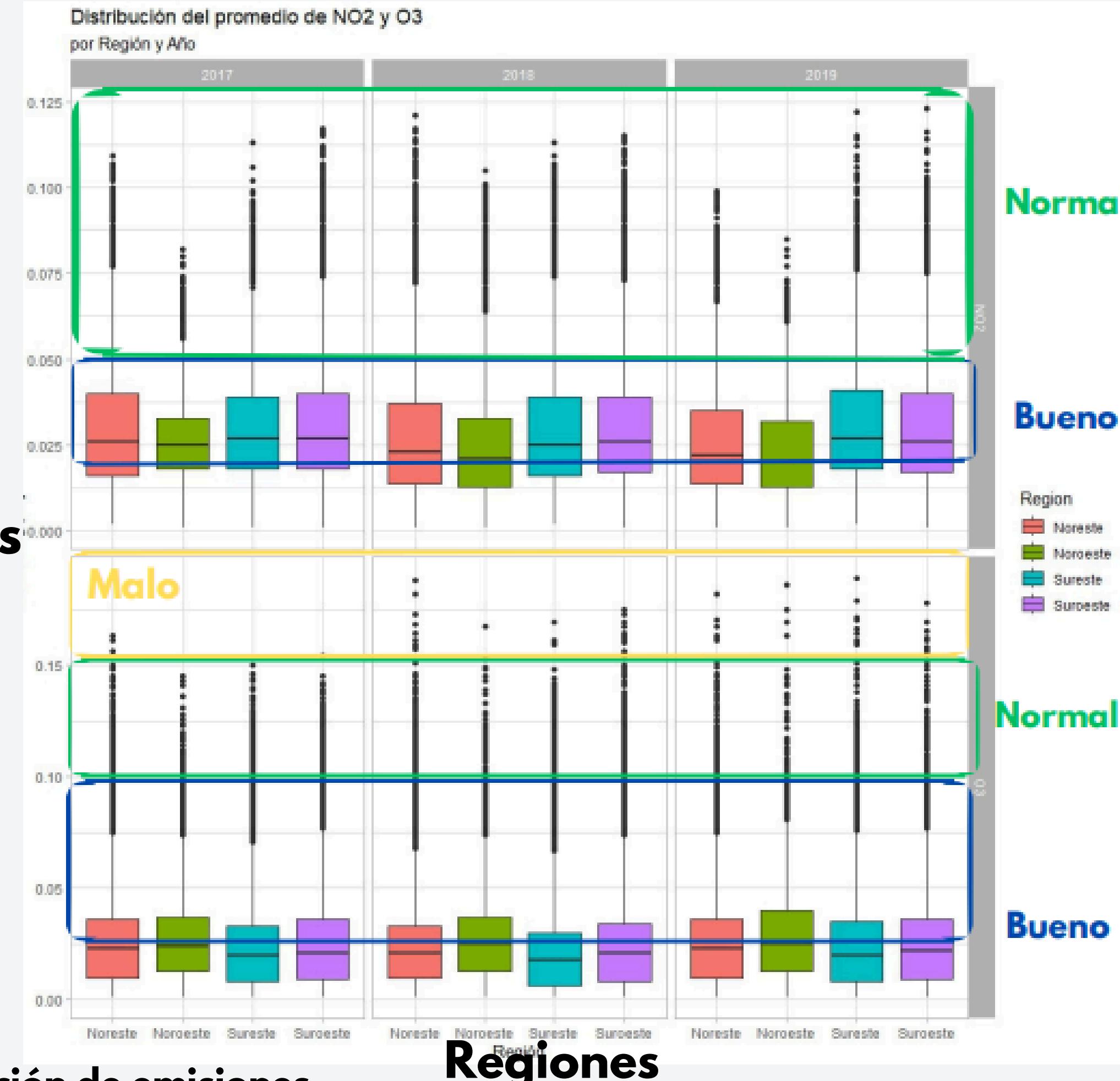
VISUALIZACIÓN DE DATOS



**Emisiones
[ppm]**

Análisis Exploratorio

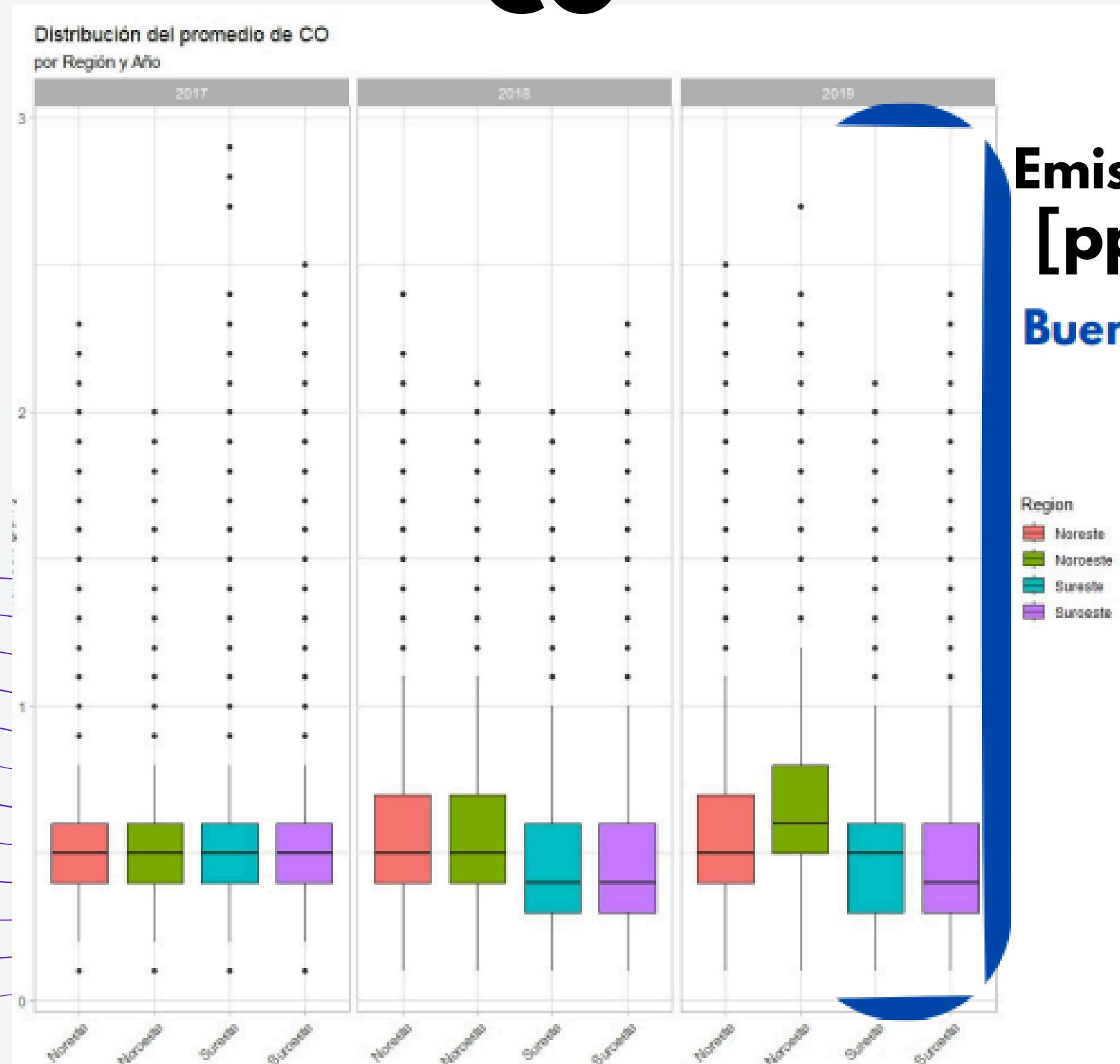
06



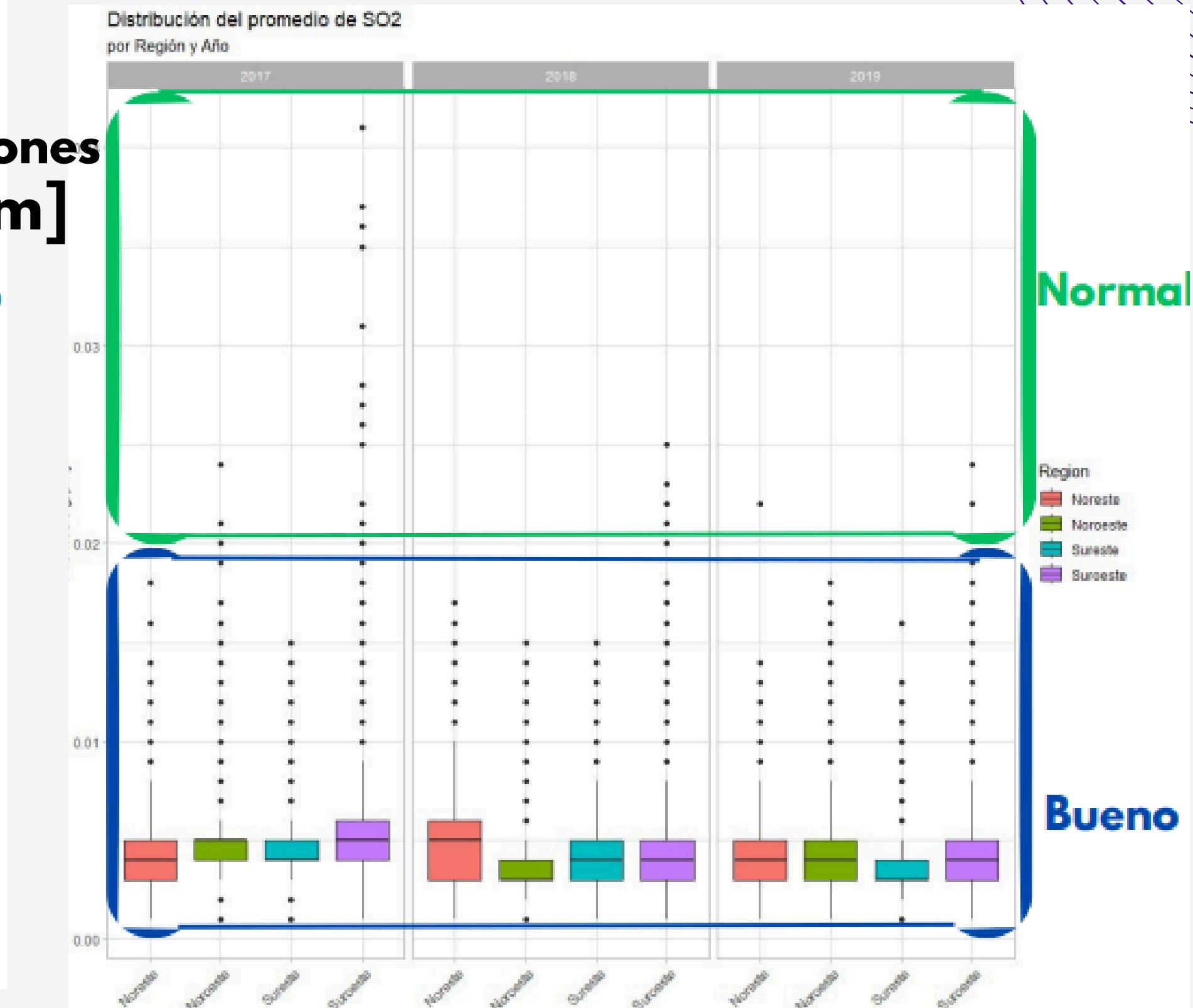
Análisis Exploratorio

07

CO



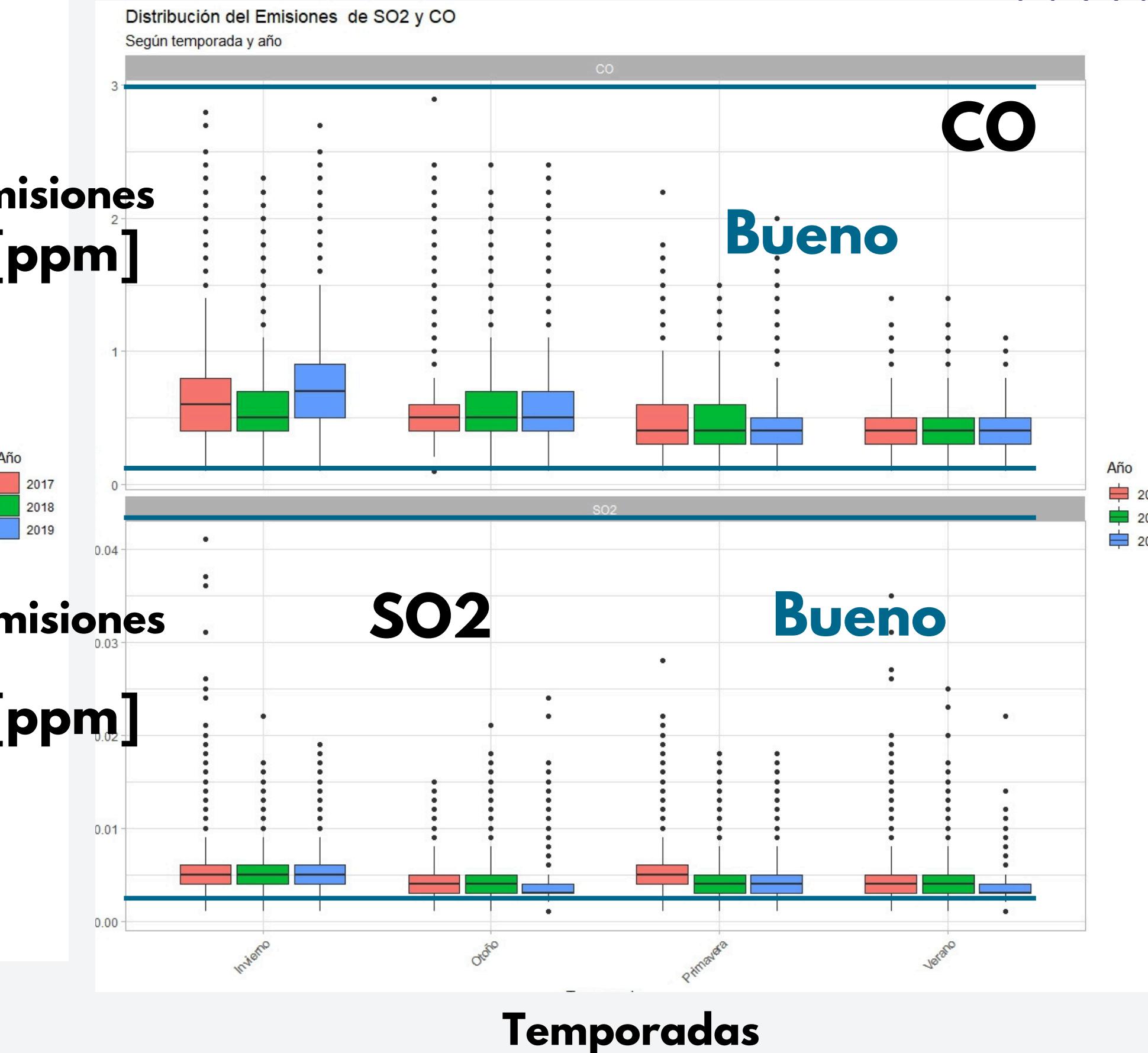
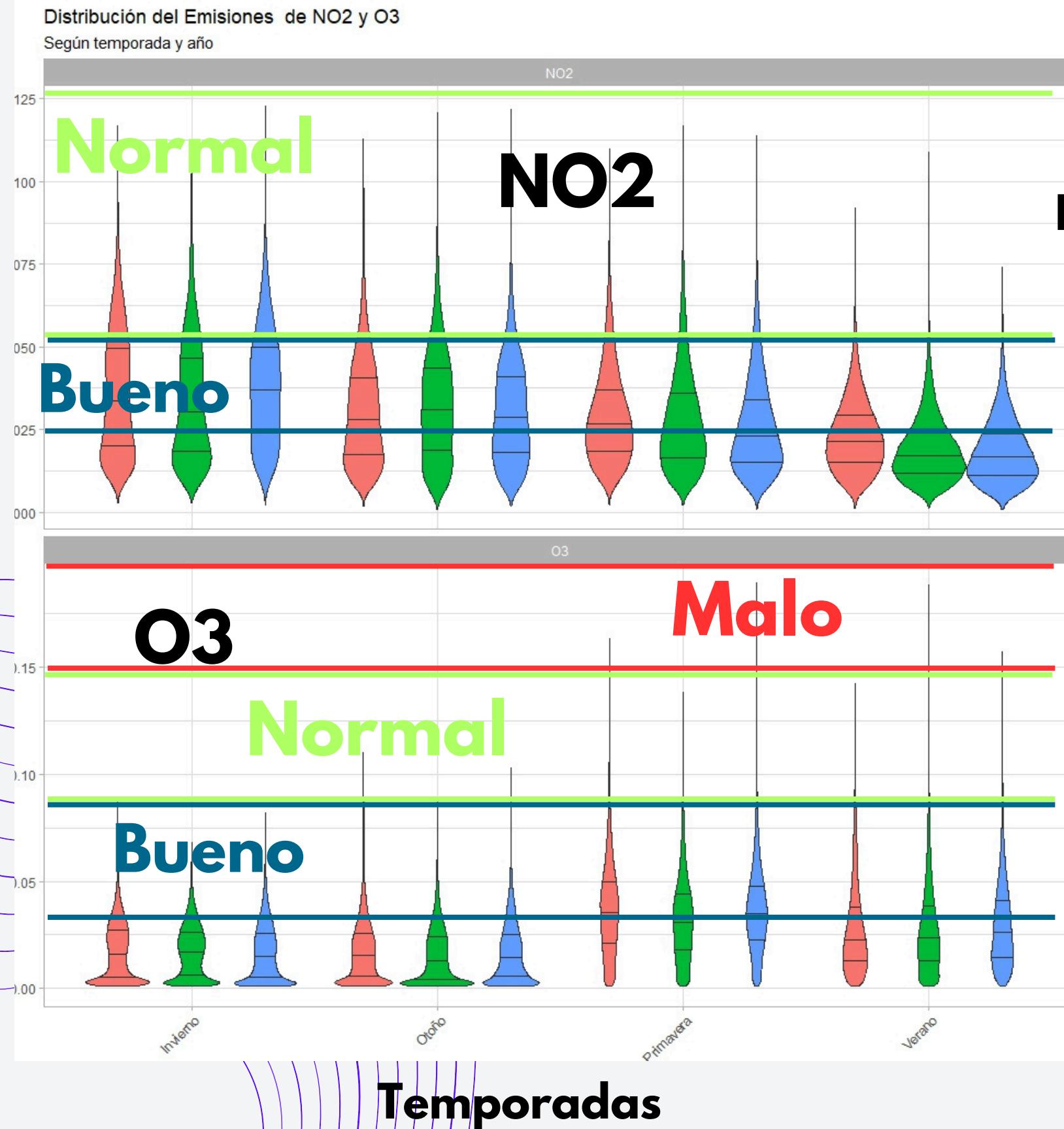
SO2



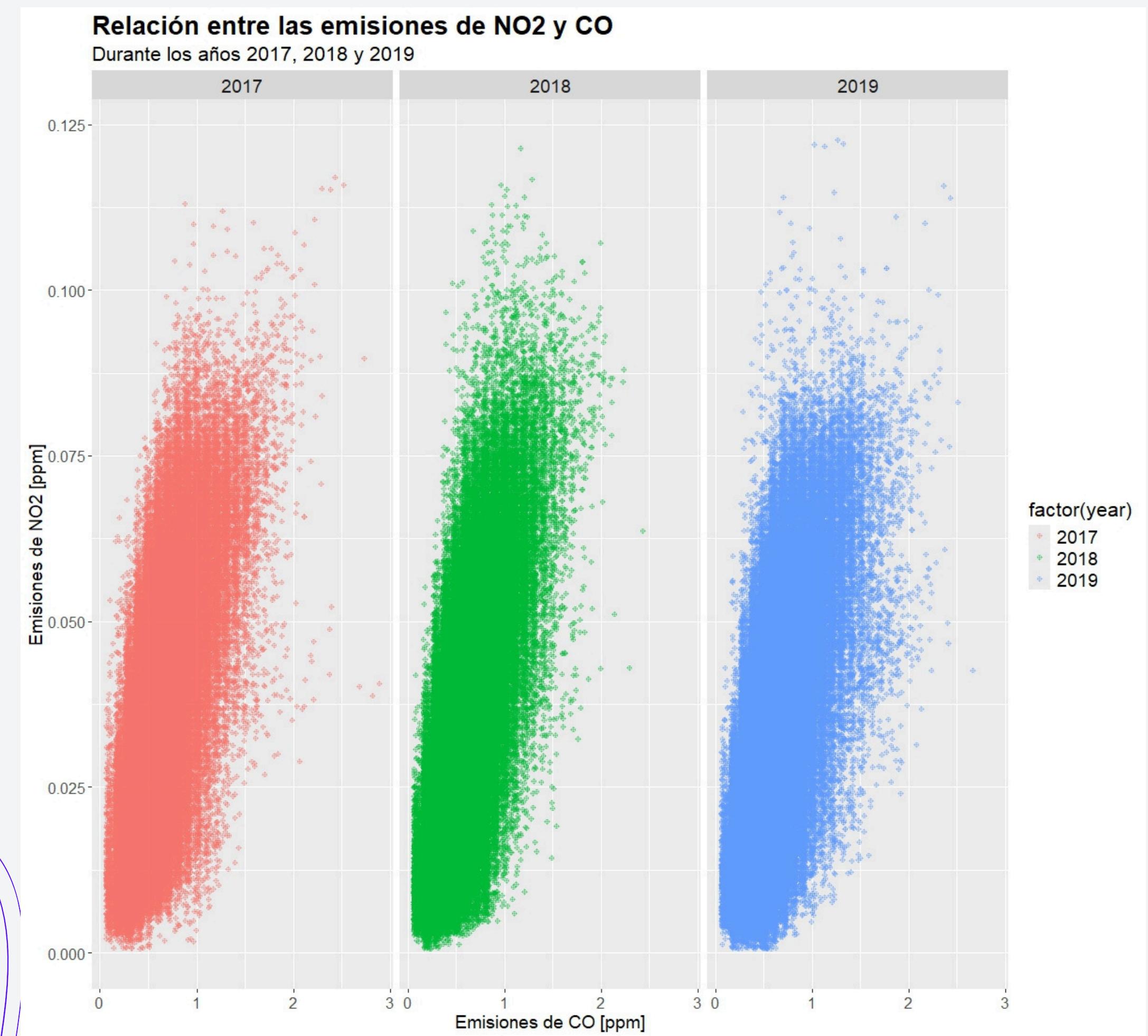
Análisis Exploratorio

Normal

08



NO2



CO

Modelado

¿Cómo se relacionan los incrementos en las emisiones de NO₂ con los incrementos en las emisiones de CO en la región de Seúl durante las temporadas entre los años 2017 y 2019?

Call: Modelo 4

lm(formula = NO2 ~ CO + season + region, data = gas_no2_co)

Residual standard error: 0.01108

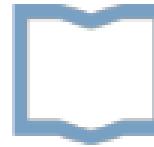
Adjusted R-squared: 0.5181

Call: Modelo 1

lm(formula = NO2 ~ CO , data = gas_no2_co)

Residual standard error: 0.01159

Adjusted R-squared: 0.4735



Comparación de modelados de las emisiones de NO₂ y CO por región

Durante las temporadas de los años 2017, 2018 y 2019

Tipo de Línea — Modelo 1 — Modelo 4 Temporada — Invierno — Otoño — Primavera — Verano

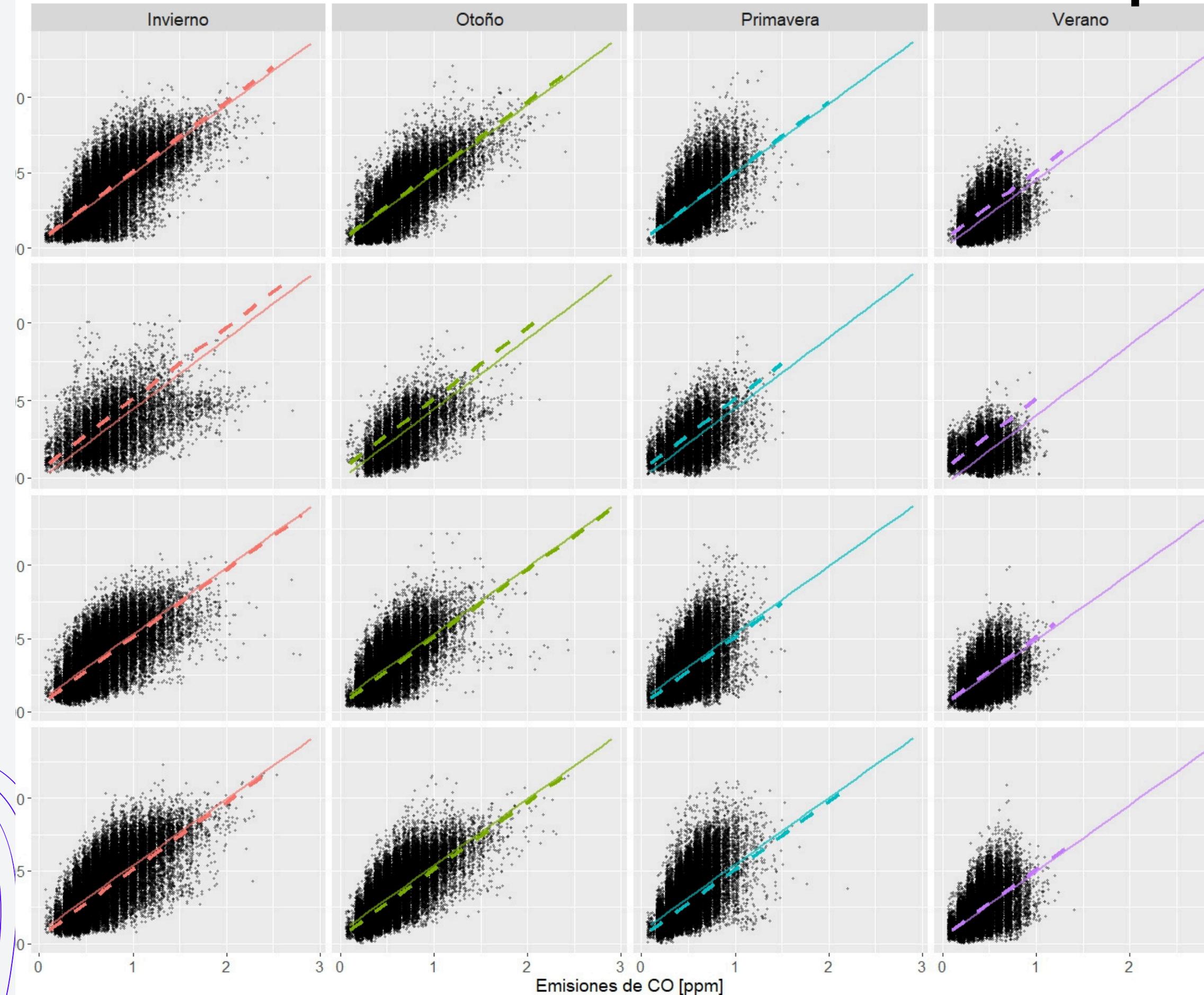
NO₂

Regiones

CO

Temporadas

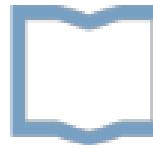
11





Conclusión

- **Los gases fueron bastante estables y sus rangos de emisiones fueron buenos**
- **Ozono con el paso de los años dejó de tener rangos malos**
- **Según la temporada del año a lo sumo llegan al rango NORMAL**
- **Al añadir las variables de temporada y region el modelo tuvo mejor R^2 y menos desvio standar**



¿Que más se podría hacer?

- Este análisis podría utilizarse para replicarlo en otros años, existen los mismos datasets pero de 2020 en adelante.
- Faltó analizar las partículas PM2.5 y PM10
- Se podría realizar los mismos pasos con datos de otros países/regiones
- Se podría analizar las emisiones y la relación con el aumento de la cantidad de autos por persona en alguna región target.
- Podríamos determinar cuáles son las zonas más industrializadas y analizar el crecimiento histórico urbano en las diferentes regiones
- Tipo de transporte que más utilizan

**GRACIAS POR
LA ATENCIÓN**



**SE ACEPTAN APLAUSOS
PERO NO PREGUNTAS**

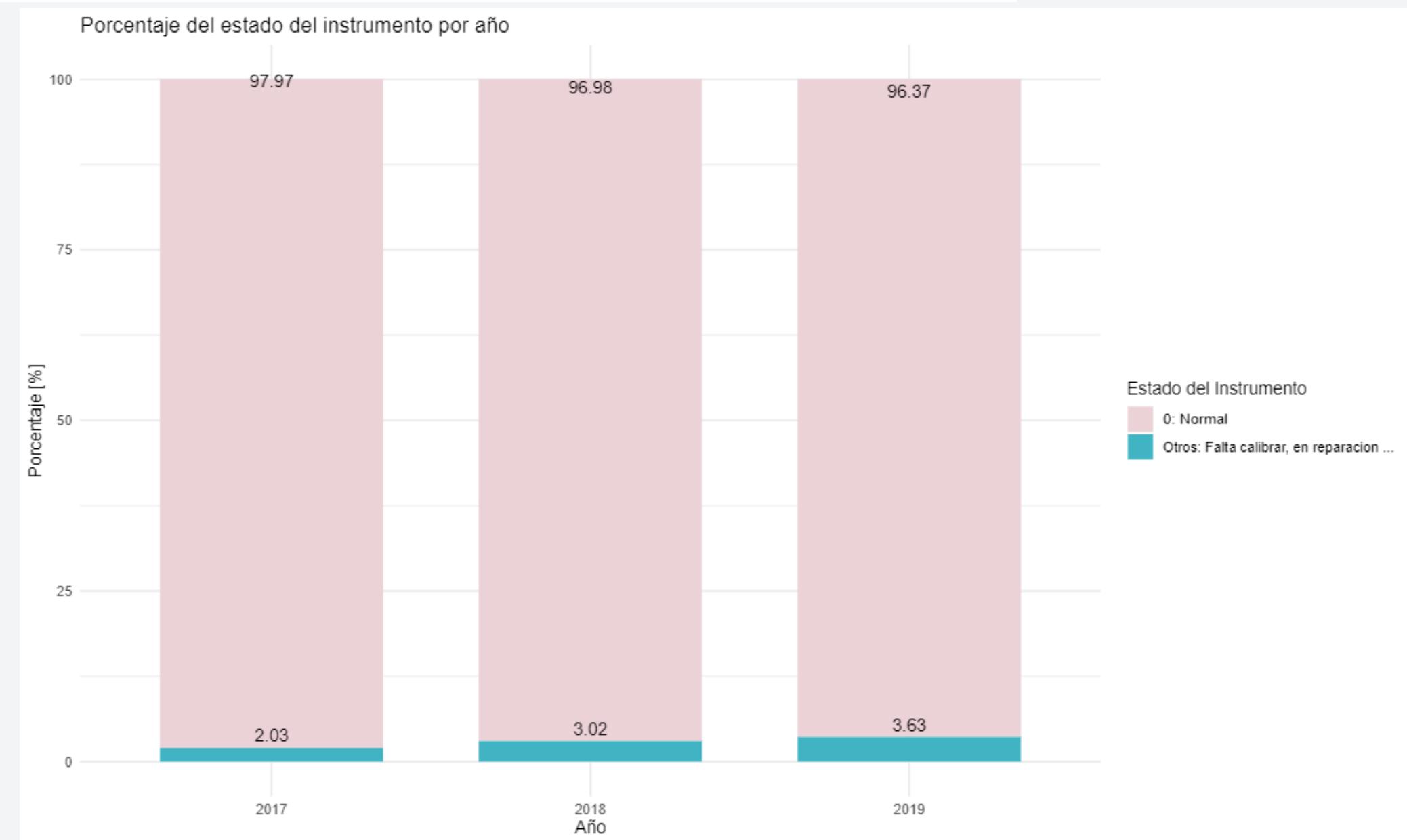
¡Muchas gracias!

Anexo

15

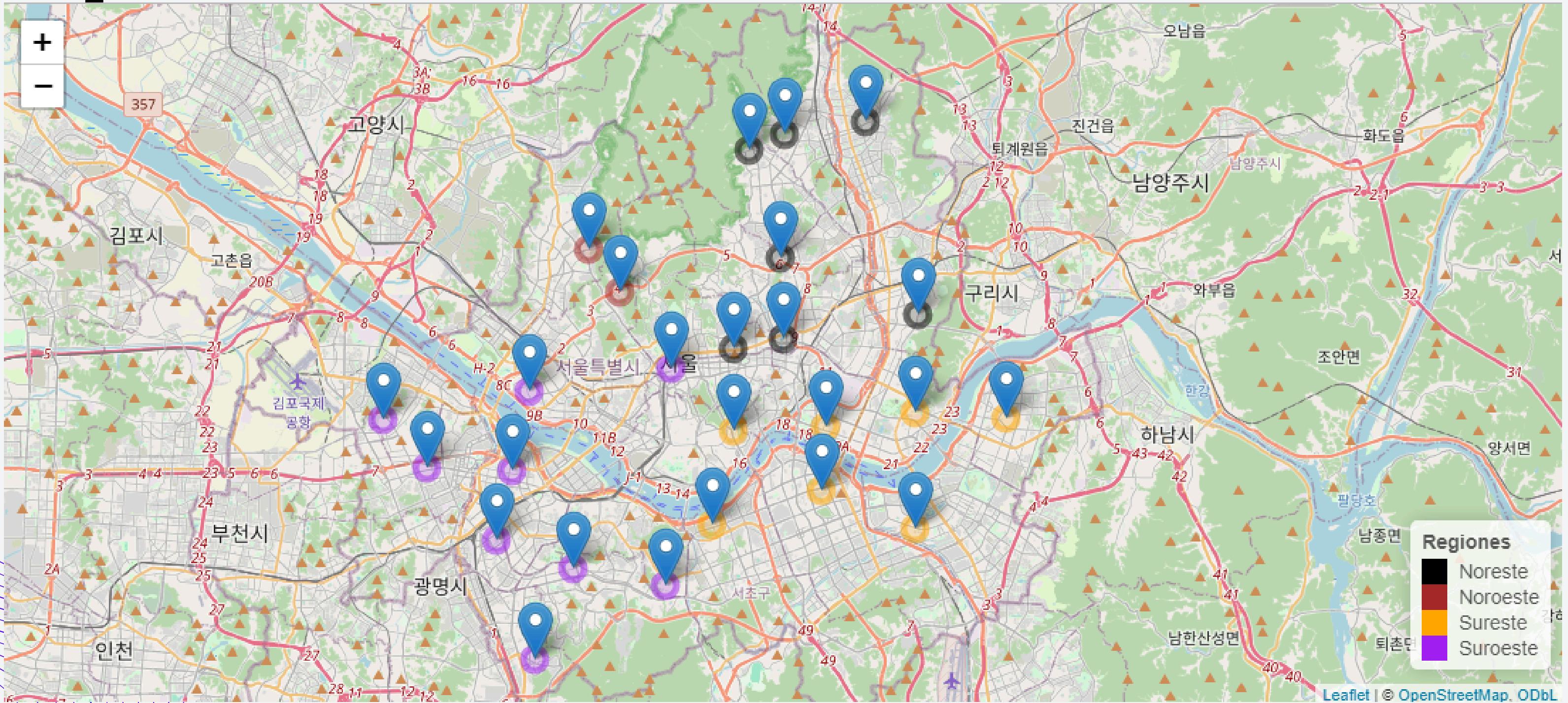
item_name (character)	unit (character)	Good(Blue) (double)	Normal(Green) (double)	Bad(Yellow) (double)	Very bad(Red) (double)
SO2	ppm	0.02	0.05	0.15	1.0
NO2	ppm	0.03	0.06	0.20	2.0
CO	ppm	2.00	9.00	15.00	50.0
O3	ppm	0.03	0.09	0.15	0.5
PM10	Microgram/m3	30.00	80.00	150.00	600.0
PM2.5	Microgram/m3	15.00	35.00	75.00	500.0

Análisis del Instrumento



Referencia de gases

Mapa con los distritos de Seúl



https://github.com/miqaapz98/T.P.F_ICD_Pollution_Seoul

Anexo

Anova

```

Analysis of Variance Table

Model 1: NO2 ~ CO
Model 2: NO2 ~ CO + region
Model 3: NO2 ~ CO * region
Model 4: NO2 ~ CO + season + region
Model 5: NO2 ~ CO * region * season
Model 6: NO2 ~ CO * season + region

  Res.Df   RSS Df Sum of Sq    F    Pr(>F)
1 635173 85.252
2 635170 80.190  3   5.0626 14294.08 < 2.2e-16 ***
3 635167 78.389  3   1.8011  5085.35 < 2.2e-16 ***
4 635167 78.035  0   0.3534
5 635143 74.984  24   3.0507  1076.70 < 2.2e-16 ***
6 635164 77.164 -21  -2.1800   879.31 < 2.2e-16 ***

---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1

```

Summary NO2~Co*region

```

Residual standard error: 0.01111 on 635167 degrees of freedom
(8035 observations deleted due to missingness)
Multiple R-squared:  0.5159,   Adjusted R-squared:  0.5159
F-statistic: 9.67e+04 on 7 and 635167 DF,  p-value: < 2.2e-16

```

VS

NO2 ~ Co + region + season

```

Residual standard error: 0.01108 on 635167 degrees of freedom
(8035 observations deleted due to missingness)
Multiple R-squared:  0.5181,   Adjusted R-squared:  0.5181
F-statistic: 9.755e+04 on 7 and 635167 DF,  p-value: < 2.2e-16

```