

ProyectoTD2021

Albert Ramos Nadal, Mar García Álvarez, Radia Ben Ghazala, Sergio Piquer Esteve

5/4/2021

Contents

Introducción	2
Instalacion y carga de Librerías de R	2
Descarga de los ficheros.	2
Fusion de los datos	3
Planteamiento de las preguntas	3
Conversion del dataframe a tidy data	4
Estudio previo del data frame:	4
Respuestas a las preguntas:	6
1.¿Qué zona tiene mas gases tóxicos?	6
2.¿Qué gas está más presente en la ciudad?	7
3.¿Los niveles de gases tóxicos son constantes durante los años? ¿Hay alguna zona donde los niveles de contaminación han bajado en los últimos años ?	10
4.¿Los niveles de CO son constantes en toda la ciudad o son diferentes de una estación a otra?	16
5.Las zonas industriales con fabricas,al ser zonas donde hay mucha quema de combustibles y reacciones químicas(algo que podemos observar del nivel de PM2.5), ¿tendrán los niveles más altos de contaminación ?	17
6 a. ¿Qué agente contaminante ha sido el menos registrado en todas las zonas? (viéndolo como de qué contaminante se ha registrado la menor cantidad por cada zona)	18
6 b) Para poder hacernos una idea de la fiabilidad y la precisión de los resultados anteriores debemos comprobar que el número de muestras de las que se dispone para hallar los valores estadísticos ha sido lo suficientemente grande como para que dicho estadístico tenga un “valor real”. Para ello vamos a mirar qué agente contaminante ha sido el menos registrado en todas las zonas a nivel de veces que se ha registrado sobre el total de días (valores perdidos).	20

7. Se sabe que en el territorio nacional, en un plano general, durante los meses de confinamiento a raíz de la pandemia de la Covid-19 hubo una reducción de la cantidad de gases contaminantes emitidos a la atmósfera debido a las restricciones de movilidad a la que nos vimos sujetos. ¿Hay evidencias de ello si nos basamos en las mediciones que han hecho las estaciones valencianas cuyos datos estamos estudiando?	22
Conclusión:	24

Introducción

Uno de los problemas mas graves a los que nos enfrentamos hoy en dia es la contaminación del aire que causa diversos problemas a nuestra sociedad como es la destrucción de la capa de ozono, el calentamiento global o incluso efectos negativos sobre la salud de las personas. Por esta razón el ayuntamiento de Valencia ha decidido monitorizar las emisiones de gases contaminantes en diferentes zonas de Valencia mediante estaciones de control de calidad del aire. A partir de los datos recogidos por estas estaciones hemos hecho un análisis exhaustivo para llegar a diferentes conclusiones usando las herramientas proporcionadas por la plataforma R

Instalacion y carga de Librerias de R

```
packages <-c("tidyverse", "readr", "dplyr", "kableExtra", "VIM")

package.check <- lapply(packages, FUN = function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = TRUE)
    library(x, character.only = TRUE)
  }
})
```

Descarga de los ficheros.

```
#borramos todas las variables que hay en el environment.

rm(list = ls())

#cargamos los nombres de los ficheros
files <- list.files(path="data/",full.names = T)

#descargamos cada uno de los ficheros y vamos renombrando los nombres de las columnas

f1<-as.data.frame(read_delim(files[1], ";", escape_double = FALSE, trim_ws = TRUE, col_names = T))
names(f1) <- c("fecha", "PM2.5", "SO2", "NO", "NO2", "PM10", "NOx", "Ozono")

f2<-as.data.frame(read_delim(files[2], ";", escape_double = FALSE, trim_ws = TRUE, col_names = T))
names(f2) <- c("fecha", "PM2.5", "PM1", "SO2", "CO", "NO", "NO2", "PM10", "NOx", "Ozono")
```

```
f3<-as.data.frame(read_delim(files[3], ";", escape_double = FALSE, trim_ws = TRUE, col_names = T))
names(f3) <- c("fecha", "PM2.5", "Xileno", "SO2","CO", "NO", "NO2", "PM10", "NOx", "Ozono" ,
"Tolueno", "Benceno1", "Ruido")

f4<-as.data.frame(read_delim(files[4], ";", escape_double = FALSE, trim_ws = TRUE, col_names = T))
names(f4) <- c("fecha", "PM2.5", "SO2", "NO", "NO2", "PM10","Ni", "NOx", "Ozono", "As","Pb", "Cd")

f5<-as.data.frame(read_delim(files[5], ";", escape_double = FALSE, trim_ws = TRUE, col_names = T))
names(f5) <- c("fecha", "PM2.5", "SO2","CO", "NO", "NO2", "PM10", "NOx", "Ozono" ,"velocidad")

f6<-as.data.frame(read_delim(files[6], ";", escape_double = FALSE, trim_ws = TRUE, col_names = T))
names(f6) <- c("fecha", "SO2", "NO", "NO2", "PM10","Ni", "NOx", "Ozono", "As","Pb", "BaP","Cd")
```

Fusion de los datos

```
datos1 <- bind_rows(f1,f2,f3,f4,f5,f6)
```

Observamos que hay archivos en cuyo data frame hay valores faltantes en todas las observaciones de una misma columna. Es decir, no todas las estaciones de control de calidad del aire tienen en cuenta o detectan exactamente los mismos agentes contaminantes.

Planteamiento de las preguntas

Las preguntas que podemos hacer en el apartado 3, antes de convertirlo en tidy, con los valores proporcionados son:

- 1) ¿Qué zona tiene más gases tóxicos ?
- 2) ¿Qué gas está más presente en la ciudad?
- 3) ¿Los niveles de gases tóxicos son constantes durante los años ?
- 4) ¿En qué zona se ha notado menos contaminación ?
- 5) ¿Los niveles de CO son constantes en toda la ciudad o son diferentes en una estación o en otra ?
- 6) Las zonas industriales con fábricas, al ser zonas donde hay mucha quema de combustibles y reacciones químicas(algo que podemos observar del nivel de PM2.5) ¿tendrá los niveles más altos de contaminación ?
- 7) ¿Qué agente contaminante ha sido el menos registrado en todas las zonas ?
- 8) Se sabe que en el territorio nacional, en un plano general, durante los meses de confinamiento a raíz de la pandemia de la Covid-19 hubo una reducción de la cantidad de gases contaminantes emitidos a la atmósfera debido a las restricciones de movilidad a la que nos vimos sujetos. ¿Hay evidencias de ello si nos basamos en las mediciones que han hecho las estaciones valencianas cuyos datos estamos estudiando?

Conversion del dataframe a tidy data

convertimos el dataframe en tidy y eliminamos las filas que contienen valores perdidos del dataframe ya que una observación sin

```
datos <- gather(datos1, key = "contaminante", value="valor", -fecha, factor_key = F )

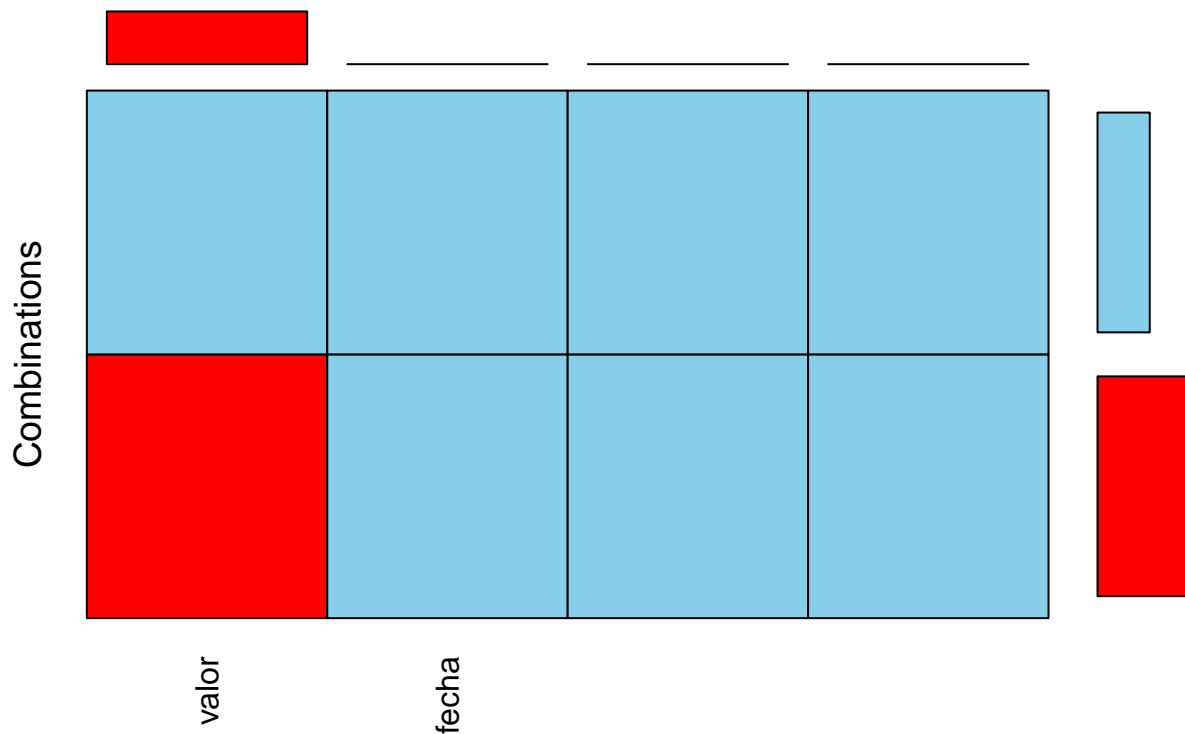
doc1 <- replicate(nrow(f1), "Estación1")
doc3 <- replicate(nrow(f2), "Estación3")
doc4 <- replicate(nrow(f3), "Estación4")
doc5 <- replicate(nrow(f4), "Estación5")
doc6 <- replicate(nrow(f5), "Estación6")
doc7 <- replicate(nrow(f6), "Estación7")

lugar <- c(doc1, doc3, doc4, doc5, doc6, doc7)

datos$Localizacion <- lugar
datos$Localizacion <- as.factor(datos$Localizacion)
datos$contaminante <- as.factor(datos$contaminante)
datos$fecha <- as.Date(datos$fecha, "%d/%m/%Y")
datos$valor <- as.numeric(datos$valor)
```

Estudio previo del data frame:

```
aggr(datos, prop = FALSE, combined = TRUE, numbers = TRUE, sortVars = TRUE, sortCombs = TRUE)
```



```
##
## Variables sorted by number of missings:
##   Variable  Count
##   valor    188827
##   fecha      0
##   contaminante  0
##   Localizacion  0
```

```
#quitamos los valores de NA
datos <- na.omit(datos)
```

Como podemos observar, el data frame tiene 188827 valores nulos en la columna de valor, ya que al juntarlos todos, en algunos hay columnas que en otros no existían y por eso se rellenan con NA's.

```
glimpse(datos)
```

```
## Rows: 103,602
## Columns: 4
## $ fecha      <date> 2014-01-01, 2014-01-02, 2014-01-03, 2014-01-04, 2014-...
## $ contaminante <fct> PM2.5, PM2.5, PM2.5, PM2.5, PM2.5, PM2.5, PM2.5, PM2.5...
## $ valor      <dbl> 1, 2, 2, 1, 2, 2, 8, 12, 27, 29, 20, 18, 8, 2, 3, 4, 4...
## $ Localizacion <fct> Estación1, Estación1, Estación1, Estación1, Estación1,...
```

Este data frame ya ordenado consta de 103602 filas/observaciones sin datos nulos.

```
summary(datos)
```

```
##      fecha      contaminante      valor      Localizacion
## Min.   :2014-01-01  Ozono   :14839  Min.    : 0.00  Estación1:17400
## 1st Qu.:2015-09-30  NO     :14697  1st Qu.: 6.00  Estación3:19834
## Median :2017-08-19  NOx    :14697  Median  :17.00  Estación4:20336
## Mean   :2017-07-28  NO2    :14695  Mean    :24.85  Estación5:16135
## 3rd Qu.:2019-05-21  SO2    :14526  3rd Qu.:37.00  Estación6:15765
## Max.   :2021-01-31  PM10   :12732  Max.    :331.00  Estación7:14132
##                                     (Other):17416
```

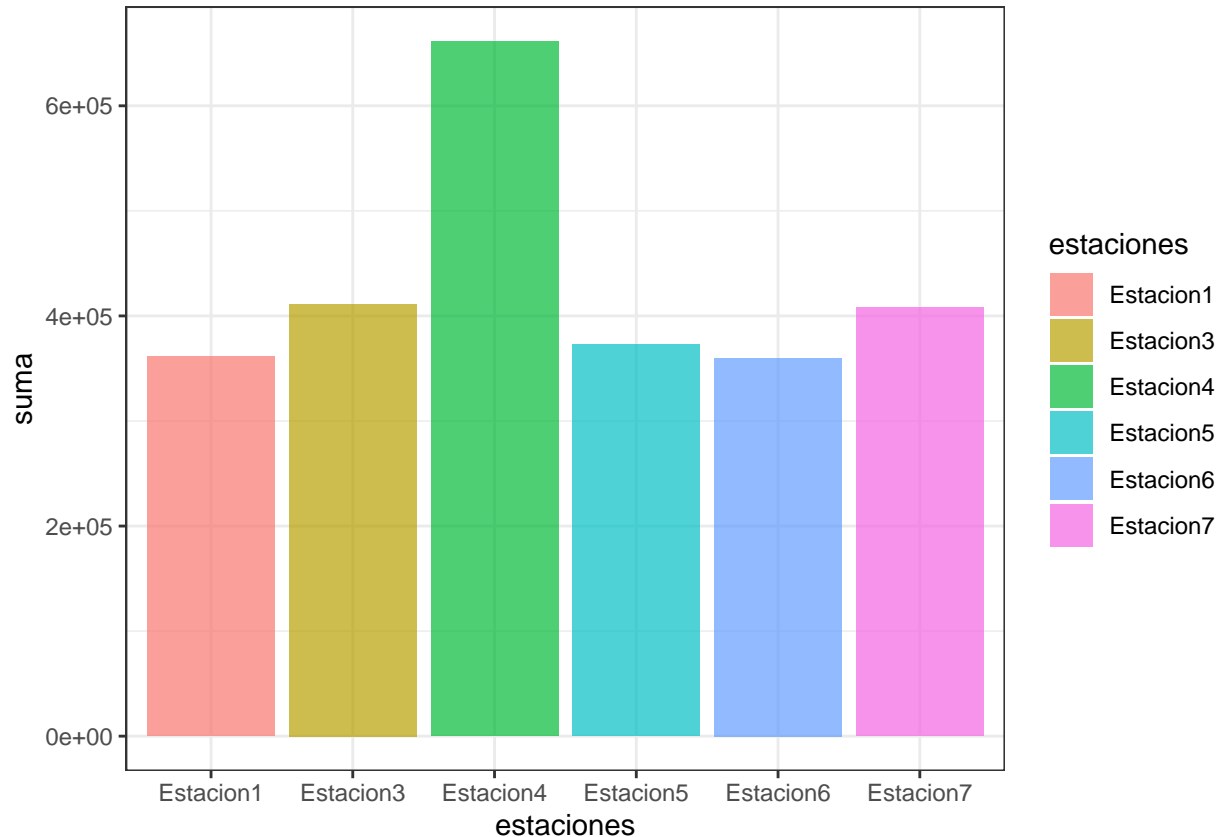
Los datos que vamos a estudiar van desde el 1 de enero de 2014 hasta el 31 de enero de 2021. Estudiaremos 6 estaciones, cada una corresponde a una zona diferente de Valencia: Estación 1 = Universidad politécnica Estación 3 = Molí del Sol Estación 4 = Pista de Silla Estación 5 = Vivers Estación 6 = Avda. França Estación 7 = Boulevard Sud

Respuestas a las preguntas:

1.¿Qué zona tiene mas gases tóxicos?

```
lista<-datos[grep("Estación1",datos$Localizacion),]
a<-sum(as.numeric(lista$valor), na.rm = TRUE)
lista3<-datos[grep("Estación3",datos$Localizacion),]
b<- sum(as.numeric(lista3$valor), na.rm = TRUE)
lista4<-datos[grep("Estación4",datos$Localizacion),]
c<-sum(as.numeric(lista4$valor), na.rm = TRUE)
lista5<-datos[grep("Estación5",datos$Localizacion),]
d<-sum(as.numeric(lista5$valor), na.rm = TRUE)
lista6<-datos[grep("Estación6",datos$Localizacion),]
e<-sum(as.numeric(lista6$valor), na.rm = TRUE)
lista7<-datos[grep("Estación7",datos$Localizacion),]
f<-sum(as.numeric(lista7$valor), na.rm = TRUE)
suma<-c(a,b,c,d,e,f)
estaciones=c("Estacion1","Estacion3","Estacion4","Estacion5","Estacion6","Estacion7")

#para hacer la gráfica:
df1<-data.frame(suma,estaciones)
ggplot(df1,aes(x=estaciones,y=suma),breaks=seq(0, 10, by=1))+geom_col(aes(fill=estaciones),
```



```
max(suma)
```

```
## [1] 661344
```

Como podemos observar, la zona que presenta más gases tóxicos es la estacion 4, la que hace referencia a la Pista de Silla. Esto puede ser debido a que se trata de una carretera, por lo que el tránsito de vehículos hace que se acumule mayor cantidad de gases tóxicos en el ambiente por la quema de combustible, que produce gases como NO, NOx, NO2 (todos los que sean óxidos de nitrógeno), CO y otros gases que no están presentes en nuestra tabla. También contamos con la contaminación del suelo por el petróleo (que produce gases tales como SO2(óxidos de azufre) y CO(monóxido de carbono)).

2.¿Qué gas está más presente en la ciudad?

```
unique(datos$contaminante)
```

```
## [1] PM2.5    SO2      NO       NO2      PM10     NOx      Ozono
## [8] PM1      CO      Xileno   Tolueno  Bencenol Ruido    Ni
## [15] As       Pb      Cd       velocidad
## 19 Levels: As BaP Bencenol Cd CO Ni NO NO2 NOx Ozono Pb PM1 PM10 ... Xileno
```

```
lista<-datos[grep("PM2.5",datos$contaminante),]
a<-sum(as.numeric(lista$valor), na.rm = TRUE)
```

```

lista3<-datos[grep("SO2",datos$contaminante),]
b<- sum(as.numeric(lista3$valor),na.rm=TRUE)
lista4<-datos[grep("NO",datos$contaminante),]
c<-sum(as.numeric(lista4$valor), na.rm = TRUE)
lista5<-datos[grep("NO2",datos$contaminante),]
d<-sum(as.numeric(lista5$valor), na.rm = TRUE)
lista6<-datos[grep("PM10",datos$contaminante),]
e<-sum(as.numeric(lista6$valor), na.rm = TRUE)
lista7<-datos[grep("NOx",datos$contaminante),]
f<-sum(as.numeric(lista7$valor), na.rm = TRUE)
lista8<-datos[grep("Ozono",datos$contaminante),]
g<-sum(as.numeric(lista8$valor), na.rm = TRUE)
lista9<-datos[grep("PM1",datos$contaminante),]
h<-sum(as.numeric(lista9$valor), na.rm = TRUE)
lista10<-datos[grep("CO",datos$contaminante),]
i<-sum(as.numeric(lista10$valor), na.rm = TRUE)
lista11<-datos[grep("Xielno",datos$contaminante),]
j<-sum(as.numeric(lista11$valor), na.rm = TRUE)
lista12<-datos[grep("Tolueno",datos$contaminante),]
k<-sum(as.numeric(lista12$valor), na.rm = TRUE)
lista13<-datos[grep("Bencenol",datos$contaminante),]
l<-sum(as.numeric(lista13$valor), na.rm = TRUE)
lista14<-datos[grep("Ruido",datos$contaminante),]
m<-sum(as.numeric(lista14$valor), na.rm = TRUE)
lista15<-datos[grep("Ni",datos$contaminante),]
n<-sum(as.numeric(lista15$valor), na.rm = TRUE)
lista16<-datos[grep("As",datos$contaminante),]
o<-sum(as.numeric(lista16$valor), na.rm = TRUE)
lista17<-datos[grep("Pb",datos$contaminante),]
p<-sum(as.numeric(lista17$valor), na.rm = TRUE)
lista18<-datos[grep("Cd",datos$contaminante),]
q<-sum(as.numeric(lista18$valor), na.rm = TRUE)
lista19<-datos[grep("velocidad",datos$contaminante),]
r<-sum(as.numeric(lista19$valor), na.rm = TRUE)
lista20<-datos[grep("BaP",datos$contaminante),]
s<-sum(as.numeric(lista20$valor), na.rm = TRUE)
contam<-c("PM2.5", "SO2", "NO", "NO2", "PM10", "NOx", "Ozono", "PM1", "CO", "Xileno", "Tolueno",
"Bencenol", "Ruido", "Ni", "As", "Pb", "Cd", "velocidad", "BaP")
sumas<-c(a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s)
sumas

```

```

## [1] 138881 52385 1208104 395584 252604 646094 742673 279869 0
## [10] 0 833 236 150310 731 127 0 6 348
## [19] 0

```

```

#para hacer la gráfica:
df<-data.frame(contam,sumas)
df

```

```

##      contam  sumas
## 1      PM2.5 138881
## 2         SO2  52385
## 3         NO 1208104

```

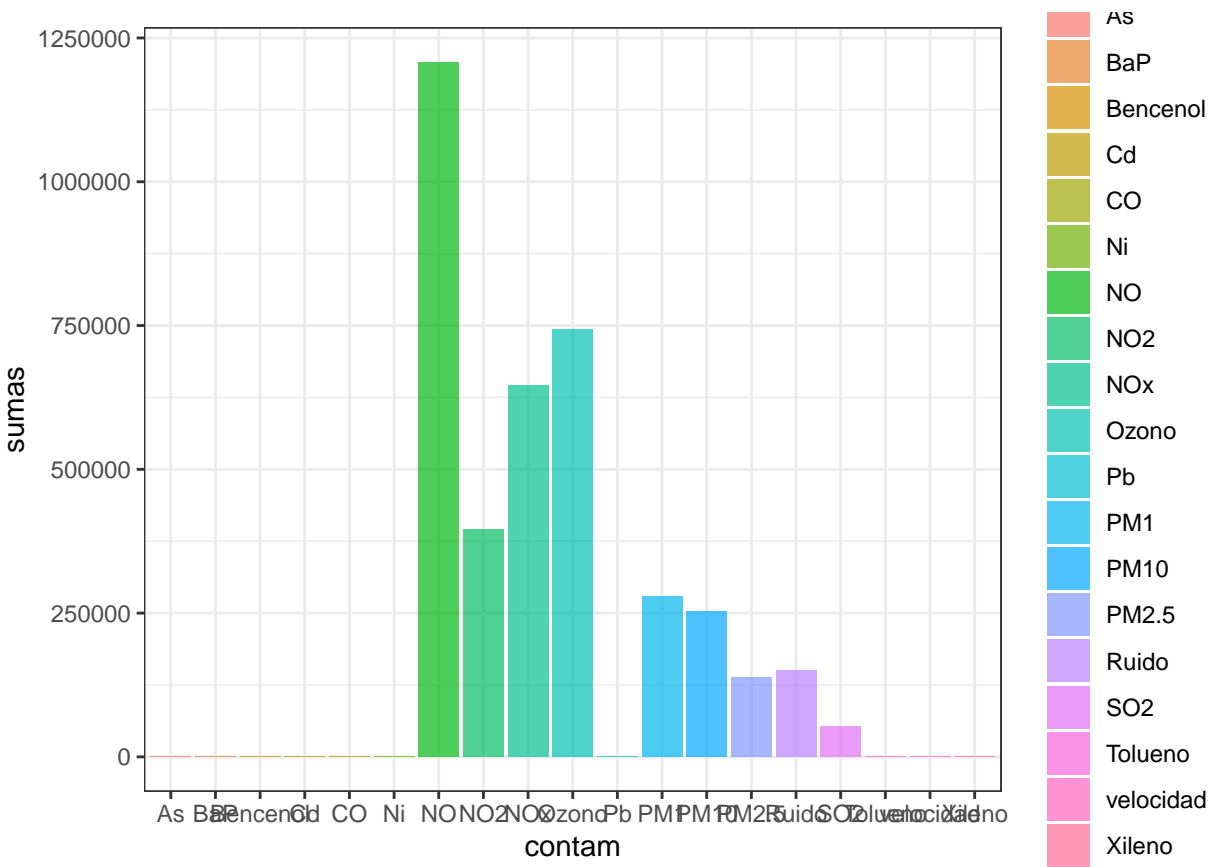


```
## 4      NO2  395584
## 5      PM10 252604
## 6      NOx  646094
## 7      Ozono 742673
## 8      PM1  279869
## 9      CO   0
## 10     Xileno 0
## 11     Tolueno 833
## 12     Bencenol 236
## 13     Ruido 150310
## 14     Ni    731
## 15     As    127
## 16     Pb    0
## 17     Cd    6
## 18     velocidad 348
## 19     BaP   0
```

```
max(sumas)
```

```
## [1] 1208104
```

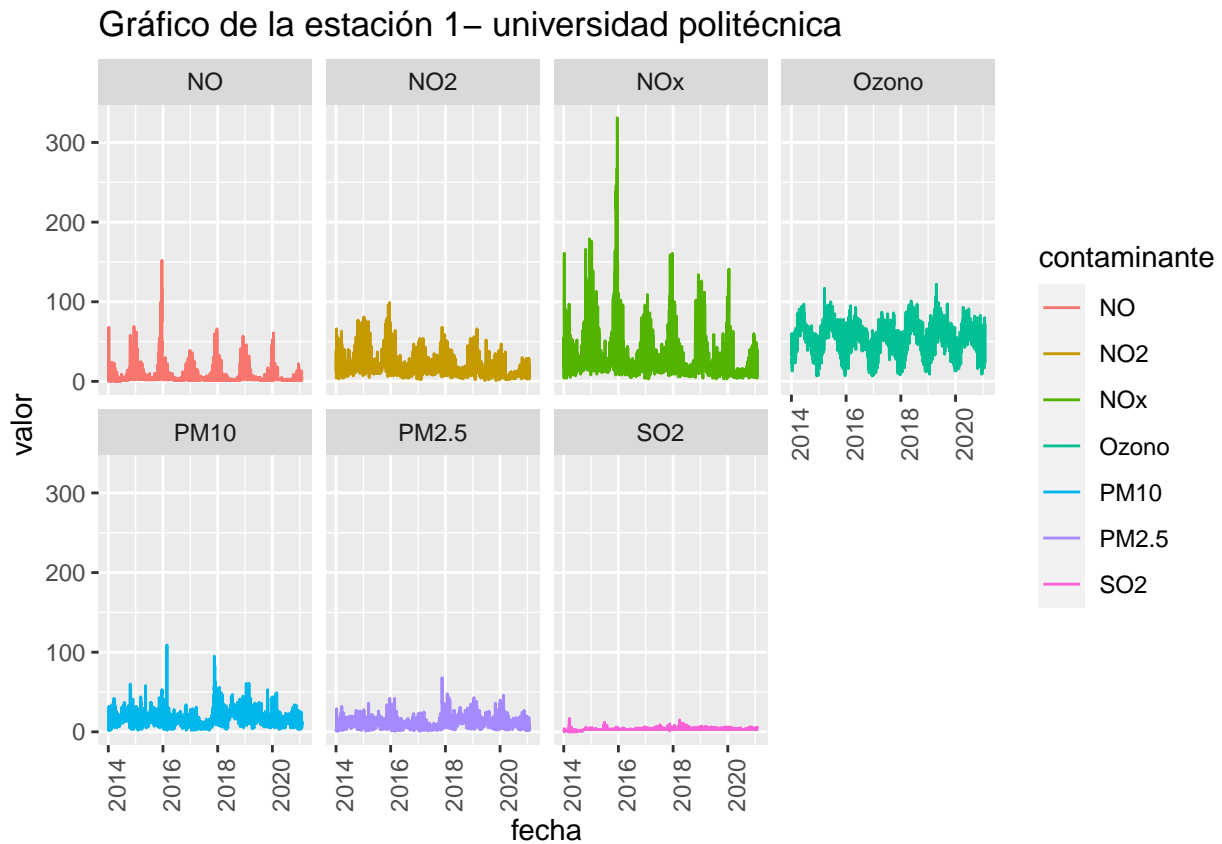
```
ggplot(df, aes(x=contam, y=sumas), breaks=seq(0, 10, by=1))+geom_col(aes(fill=contam), alpha=0.7)+theme_bw()
```



Según nuestros datos, el gas más presente en la ciudad es el NO (monóxido de nitrógeno), que se libera a la atmósfera desde el escape de vehículos motorizados, la combustión del carbón, petróleo, o gas natural, y otros muchos procesos como la detonación de dinamita.

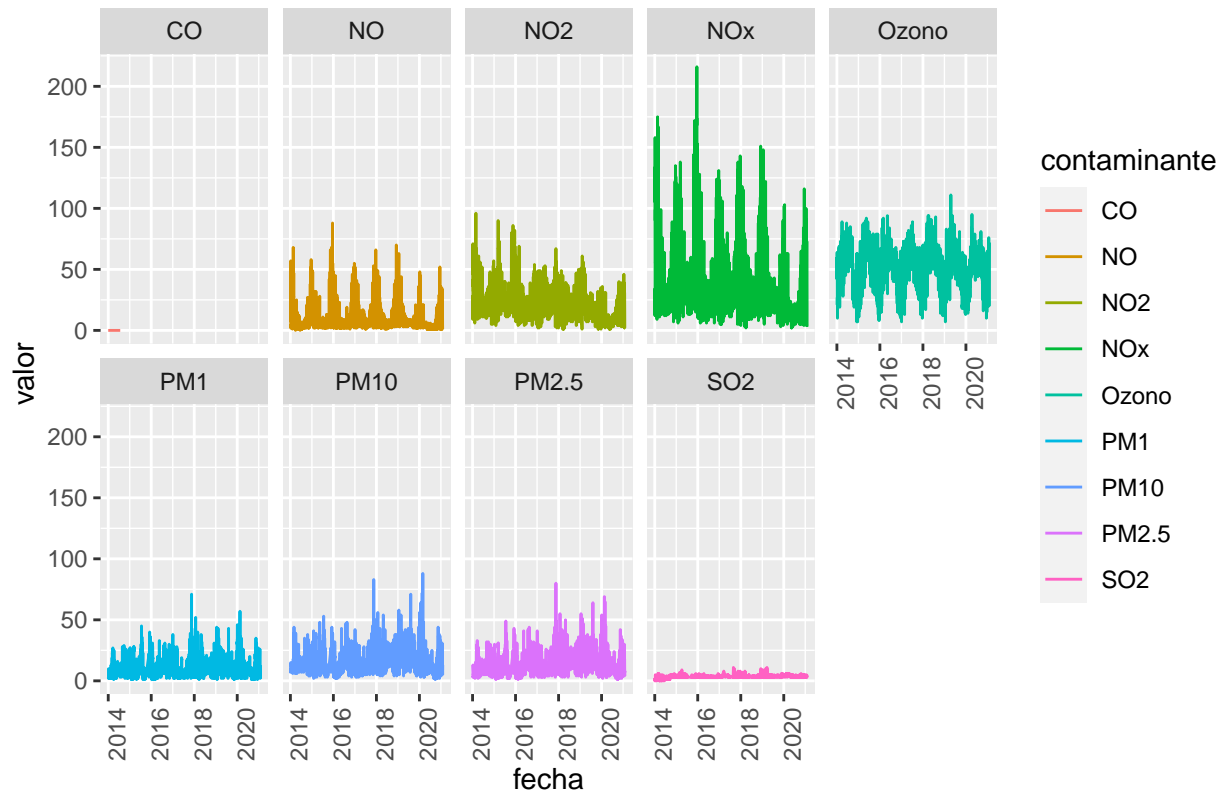
3. ¿Los niveles de gases tóxicos son constantes durante los años? ¿Hay alguna zona donde los niveles de contaminación han bajado en los últimos años?

```
#estación1
x<-as.data.frame(datos[datos$Localizacion=="Estación1",])
ggplot(x, aes(x=fecha, y=valor, colour=contaminante))+geom_line()+facet_wrap(~ contaminante, nrow = 2)+
  labs(title="Gráfico de la estación 1- universidad politécnica")+theme(axis.text.x=element_text(angle=90))
```



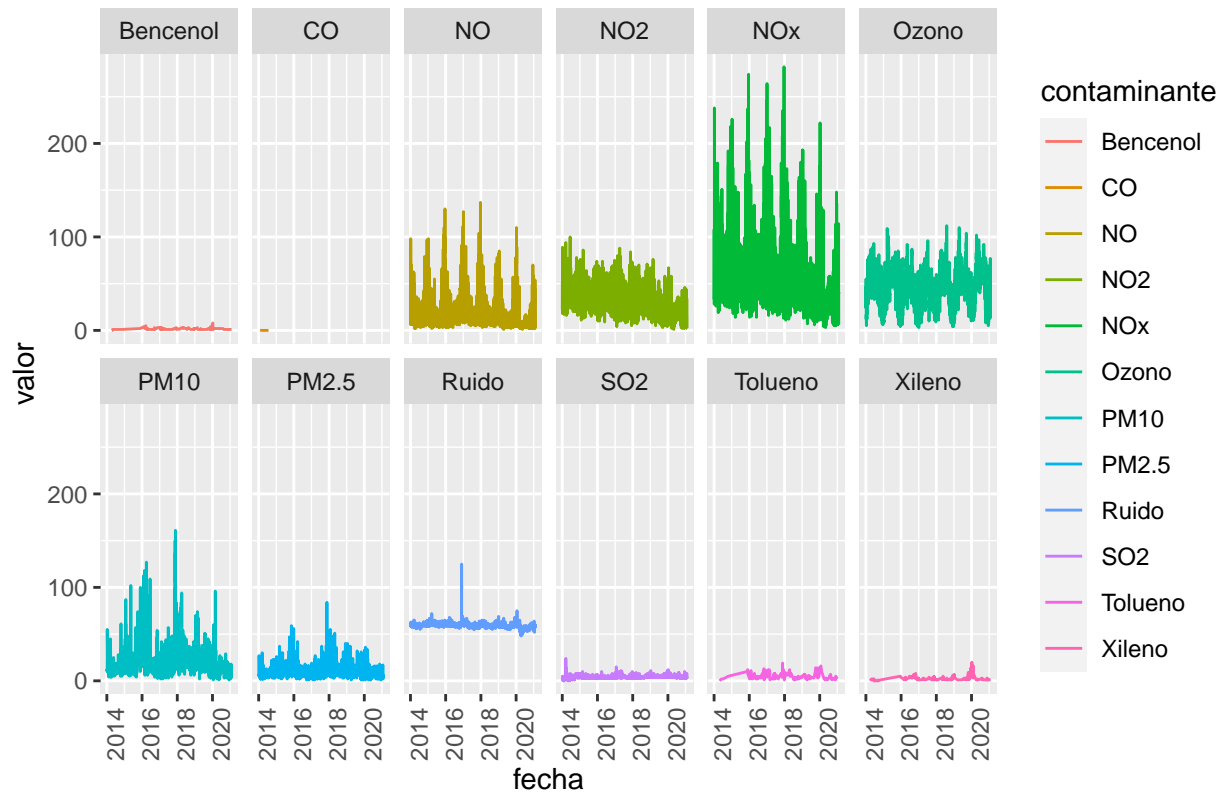
```
#estación3
x2<-as.data.frame(datos[datos$Localizacion=="Estación3",])
ggplot(x2, aes(x=fecha, y=valor, colour=contaminante))+geom_line()+facet_wrap(~ contaminante, nrow = 2)+
  labs(title="Gráfico de la estación 3 - molí del sol")+theme(axis.text.x=element_text(angle=90))
```

Gráfico de la estación 3 – molí del sol



```
#estación4
x3<-as.data.frame(datos[datos$Localizacion=="Estación4",])
ggplot(x3, aes(x=fecha, y=valor, colour=contaminante))+geom_line()+facet_wrap(~ contaminante, nrow = 2)
  labs(title="Gráfico de la estación 4 - Pista de Silla")+theme(axis.text.x=element_text(angle=90))
```

Gráfico de la estación 4 – Pista de Silla



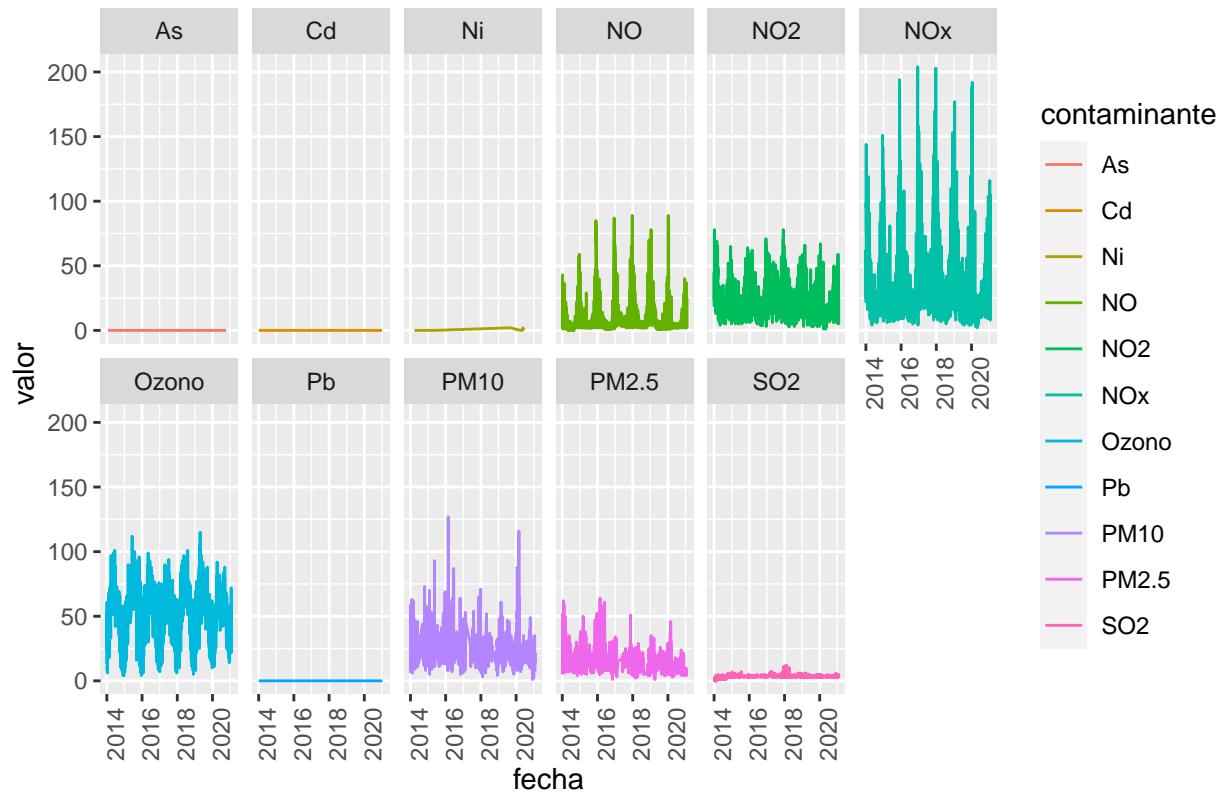
```
#estación5
```

```
x4<-as.data.frame(datos[datos$Localizacion=="Estación5",])
```

```
ggplot(x4, aes(x=fecha, y=valor, colour=contaminante))+geom_line()+facet_wrap(~ contaminante, nrow = 2)
```

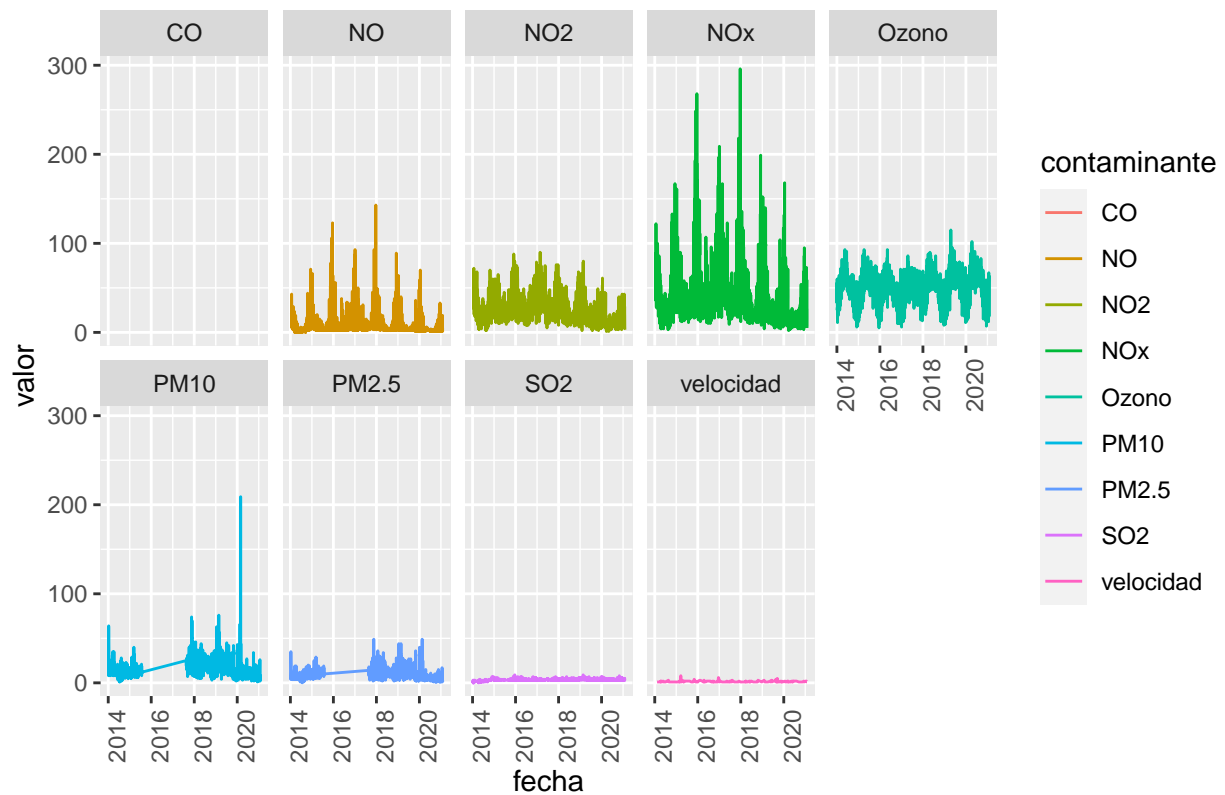
```
labs(title="Gráfico de la estación 5 - Vivers")+theme(axis.text.x=element_text(angle=90))
```

Gráfico de la estación 5 – Vivers



```
x5<-as.data.frame(datos[datos$Localizacion=="Estación6",])
ggplot(x5, aes(x=fecha, y=valor, colour=contaminante))+geom_line()+facet_wrap(~ contaminante, nrow = 2)
labs(title="Gráfico de la estación6 - Avda. França")+theme(axis.text.x=element_text(angle=90))
```

Gráfico de la estación6 – Avda. França



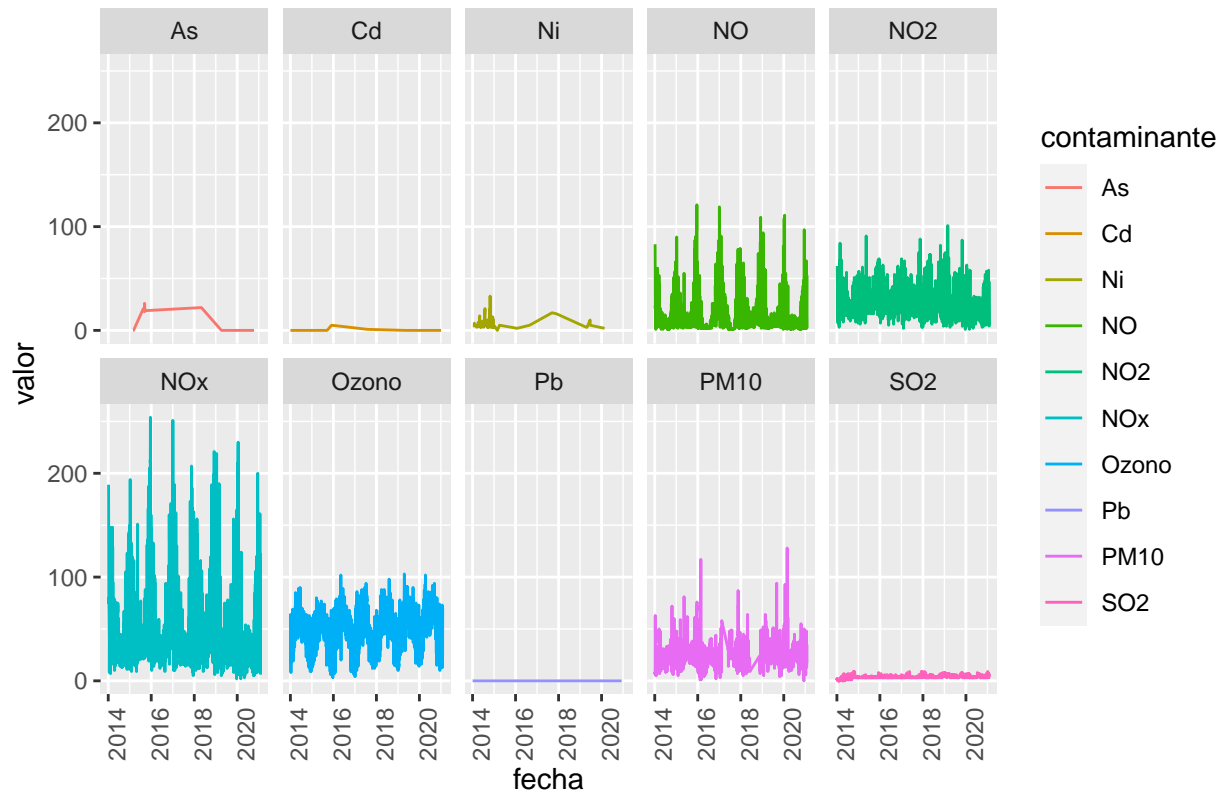
```
#estación7
```

```
x6<-as.data.frame(datos[datos$Localizacion=="Estación7",])
```

```
ggplot(x6, aes(x=fecha, y=valor, colour=contaminante))+geom_line()+facet_wrap(~ contaminante, nrow = 2)
```

```
labs(title="Gráfico de la estación 7 - Boulevard Sud")+theme(axis.text.x=element_text(angle=90))
```

Gráfico de la estación 7 – Boulevard Sud



Estudio estación1:

Como podemos observar, en esta estación el agente contaminante que más oscilaciones tiene es el NOx, con un pico muy alto a principios de 2016, al igual que el NO. El ozono oscila de manera muy similar a lo largo del año para cada año, aunque en 2019 incrementó un poco. Es fácil observar que en esta estación ningún gas es constante durante todo el año, todos presentan oscilaciones mayores o menores a lo largo del año.

También podemos llegar a la conclusión de que todos los agentes contaminantes, excepto el SO2 y el Ozono han bajado en el último año.

Estudio estación3:

En esta estación los niveles de NO son mayores, debido a que es una zona industrial, aunque las oscilaciones son muy similares a las de la primera estación. En general aumenta la presencia de todos los agentes contaminantes y aparece uno nuevo, el PM1, que al igual que los demás va oscilando según la época del año.

En esta estación también han disminuido un poco los valores de los agentes en los últimos meses.

Estudio estación4:

Aquí podemos observar que los niveles de NOx se disparan al comenzar cada año, así como de todos los otros agentes contaminantes, y aparece uno nuevo, el ruido. Esto ocurre porque esta estación se encuentra situada en una autopista y como hemos mencionado anteriormente, la quema de combustible contamina mucho y hace aumentar los valores de estos agentes. Además añadimos nuevos contaminantes como son el Tolueno y el Xileno, que también proceden de la combustión de gasolina y diésel.

Estudio estación5:

En esta estación podemos observar que los agentes que menos oscilaciones tienen son As, Cd, Ni y Pb, pero esto no quiere decir que no sean importantes, debido a su gran toxicidad. En esta zona aumentan

considerablemente los niveles de Ozono, esto puede ser porque para que se lleve a cabo reacción fotoquímica que lo genera, es necesaria una cierta distancia para generar O₃ a partir de sus precursores, cosa que en un parque como viveros se da.

Al contrario de lo que podríamos pensar, dado que es una zona verde, en esta estación no ha disminuido apenas los niveles de agentes contaminantes a lo largo de este último año, aunque durante el tiempo que estuvimos encerrados por COVID-19 si que se nota un buen descenso.

Estudio estación6:

En esta estación lo que más nos puede llamar la atención es que el PM₁₀ y el PM_{2.5} desde mediados de 2015 hasta casi finales de 2018 aumenta de manera muy uniforme, poco a poco, al contrario que los demás agentes, que presentan oscilaciones bastante fuertes.

En esta ocasión, volvemos a observar que todos los agentes formados por nitrógeno disminuyen su cantidad en los últimos meses.

Estudio estación7:

En esta estación vuelven a apreocer los agentes As, Cd, Ni y Pb. Pero esta vez presentan un mayor cambio a lo largo de los años. El As aumenta, obtiene valores más altos entre 2015 y 2018, aunque se queda más o menos constante en una meseta. El Cd aumenta un poco entre finales de 2015 y 2017. El Ni presenta oscilaciones hasta 2015, a partir de este año aumenta su valor de manera más uniforme, pero en 2018 vuelve a disminuir.

Según aparece en este gráfico, los valores de los agentes descienden muy poco.

4.¿Los niveles de CO son constantes en toda la ciudad o son diferentes de una estación a otra?

Centrémonos ahora en nuestra salud:

El monóxido de carbono (CO) es uno de los gases contaminantes más nocivos para el ser humano, pues reduce la capacidad de la sangre de transportar oxígeno a los tejidos del cuerpo.

```
#Estación 3. Molí de Sol
f2$CO <- sub(",", ".", f2$CO, fixed = TRUE)
f2$CO <- as.numeric(f2$CO)
mean(f2$CO, na.rm = T)
```

```
## [1] 0.1534792
```

```
#Estación 4. Pista de Silla
f3$CO <- sub(",", ".", f3$CO, fixed = TRUE)
f3$CO <- as.numeric(f3$CO)
mean(f3$CO, na.rm = T)
```

```
## [1] 0.1950575
```

```
#Estación 6. Avda. França
f5$CO <- sub(",", ".", f5$CO, fixed = TRUE)
f5$CO <- as.numeric(f5$CO)
mean(f5$CO, na.rm = T)
```

```
## [1] 0.1593086
```


Vemos que la estación que presenta unos niveles más altos de CO es la de la Pista se Silla, que es una autopista. Los niveles en Molí del Sol y en la avenida de Francia son muy similares. Estos resultados tienen sentido, pues la principal fuente antropogénica de CO es la quema de combustibles fósiles, como la que se produce en los motores de los vehículos que circulan por esa vía.

El dióxido de azufre (SO_2) contribuye a la acidificación a los ecosistemas mediante lo que se conoce como lluvia ácida, además de que sus efectos sobre la salud humana se dejan sentir de manera especial sobre el sistema respiratorio y el funcionamiento de los pulmones, pudiendo causar también irritación ocular. En todas las estaciones meteorológicas se ha medido la cantidad de este gas en el aire. Veamos los resultados obtenidos:

```
mean(f1$SO2, na.rm = TRUE)
```

```
## [1] 3.344828
```

```
mean(f2$SO2, na.rm = TRUE)
```

```
## [1] 3.403226
```

```
mean(f3$SO2, na.rm = TRUE)
```

```
## [1] 4.263704
```

```
mean(f4$SO2, na.rm = TRUE)
```

```
## [1] 3.407188
```

```
mean(f5$SO2, na.rm = TRUE)
```

```
## [1] 3.495363
```

```
mean(f6$SO2, na.rm = TRUE)
```

```
## [1] 3.724544
```

Al ser la quema de combustibles fósiles la principal fuente de este contaminante, vemos que es la estación de la autovía (Pista de Silla) la que ha registrado una mayor cantidad de este; 4.26 micro gramos por metro cúbico.

5.Las zonas industriales con fabricas,al ser zonas donde hay mucha quema de combustibles y reacciones químicas(algo que podemos observar del nivel de PM2.5), ¿tendrán los niveles más altos de contaminación ?

```
sum(f1$PM2.5,na.rm = T)
```

```
## [1] 28442
```

```
sum(f2$PM2.5,na.rm = T)
```

```
## [1] 36175
```

```
sum(f3$PM2.5,na.rm = T)
```

```
## [1] 28709
```

```
sum(f4$PM2.5,na.rm = T)
```

```
## [1] 28074
```

```
sum(f5$PM2.5,na.rm = T)
```

```
## [1] 17481
```

```
#sum(f5$PM2.5,na.rm = T) La estación de Boulevard Sur no detecta contaminantes de tipo PM2.5
```

Como podemos ver, la estación con mayor valor de PM2.5 es la de Molí del Sol, la cual está al lado del Polígono Industrial Verge de la Salud. Este resultado era esperable ya que en este polígono hay varias fábricas donde se realiza tanto quema de combustibles como reacciones químicas a gran escala.

6 a. ¿Qué agente contaminante ha sido el menos registrado en todas las zonas? (viéndolo como de qué contaminante se ha registrado la menor cantidad por cada zona)

```
#Antes de nada convertiremos a numérico todas las variables que deban serlo:
```

```
f2$CO <- sub(",", ".", f2$CO, fixed = TRUE)  
f2 <- f2 %>% mutate_at(vars(CO), as.numeric)
```

```
f3$CO <- sub(",", ".", f3$CO, fixed = TRUE)  
f3$Tolueno <- sub(",", ".", f3$Tolueno, fixed = TRUE)  
f3$Benceno <- sub(",", ".", f3$Benceno, fixed = TRUE)  
f3 <- f3 %>% mutate_at(vars(CO, Tolueno, Benceno), as.numeric)
```

```
f4$As <- sub(",", ".", f4$As, fixed = TRUE)  
f4$Pb <- sub(",", ".", f4$Pb, fixed = TRUE)  
f4$Cd <- sub(",", ".", f4$Cd, fixed = TRUE)  
f4$Ni <- sub(",", ".", f4$Ni, fixed = TRUE)  
f4 <- f4 %>% mutate_at(vars(As, Pb, Cd, Ni), as.numeric)
```

```
f5$CO <- sub(",", ".", f5$CO, fixed = TRUE)  
f5$CO <- as.numeric(f5$CO)
```

```
f6$As <- sub(",", ".", f6$As, fixed = TRUE)  
f6$Pb <- sub(",", ".", f6$Pb, fixed = TRUE)  
f6$Cd <- sub(",", ".", f6$Cd, fixed = TRUE)
```

```

f6$BaP <- sub(",", ".", f6$BaP, fixed = TRUE)
f6$Ni <- sub(",", ".", f6$Ni, fixed = TRUE)
f6 <- f6 %>% mutate_at(vars(As, Pb, Cd, BaP, Ni), as.numeric)

# Solo lo comprobaremos para aquellos agentes contaminantes que aparezcan en todas las
#estaciones o en la mayoría de ellas (en 4 o más).

#Los contaminantes que se detectan en todas las estaciones son NOx, SO2, NO, NO2, Ozono y PM10.
#PM2.5 está ausente en la estación 7, que es la de Boulevard Sud, pero se encuentra
#en todas las demás, luego también la tendremos en cuenta para el estudio.

#Universitat Politècnica
media1 <- f1 %>% select(c(SO2, NO, NO2, NOx, Ozono, PM10, PM2.5))%>%summarise_if(is.numeric,

#Molí del Sol
media3 <- f2 %>% select(c(SO2, NO, NO2, NOx, Ozono, PM10, PM2.5))%>%summarise_if(is.numeric,

#Pista de Silla
media4 <- f3 %>% select(c(SO2, NO, NO2, NOx, Ozono, PM10, PM2.5))%>%summarise_if(is.numeric,

#Vivers
media5 <- f4 %>% select(c(SO2, NO, NO2, NOx, Ozono, PM10, PM2.5))%>%summarise_if(is.numeric,

#Av. de França
media6 <- f5 %>% select(c(NOx, SO2, NO, NO2, Ozono, PM10, PM2.5))%>%summarise_if(is.numeric,

#Boulevard Sud
media7 <- f6 %>% select(c(SO2, NO, NO2, NOx, Ozono, PM10))%>%summarise_if(is.numeric,
mean, na.rm=TRUE)

res <- bind_rows(media1, media3, media4, media5, media6, media7)
row.names(res)<- c("UPV","Molí del Sol","Pista de Silla", "Vivers","Av. França","Boulevard Sud")
res

```

```

##           SO2           NO           NO2           NOx           Ozono           PM10           PM2.5
## UPV          3.344828    6.366225    21.46490    30.85038    54.24690    17.13478    11.40874
## Molí del Sol  3.403226    9.198532    23.15830    36.96453    50.59517    16.78240    14.61616
## Pista de Silla 4.263704    20.714568    35.73431    67.26214    44.84537    21.20189    11.47901
## Vivers        3.407188    8.187396    23.36728    35.50710    51.49860    24.13490    14.96482
## Av. França    3.495363    8.745424    26.80740    39.81489    50.26624    16.24206    10.09879
## Boulevard Sud 3.724544    14.425145    30.75393    52.66543    48.51797    25.21742         NA

```

```

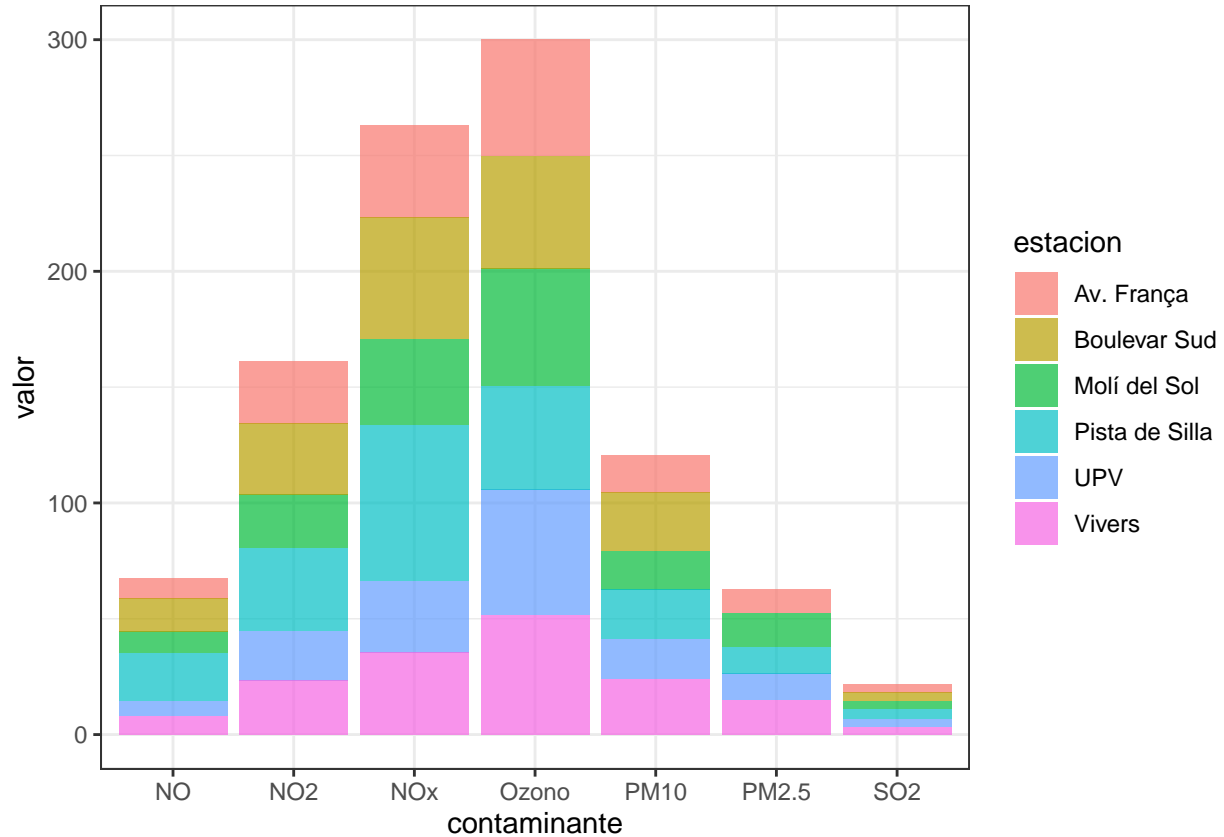
#lo representamos en un gráfico

res2 <- res
res2$estacion <- c("UPV","Molí del Sol","Pista de Silla", "Vivers","Av. França","Boulevard Sud")
res2$estacion <- as.factor(res2$estacion)

```

```
res2 <- res2 %>% gather(key = contaminante, value = valor, -estacion)
```

```
ggplot(res2,aes(x=contaminante,y=valor),breaks=seq(0, 10, by=1))+geom_col(aes(fill=estacion),
```



A partir de la tabla anterior podemos concluir que la zona cuyo nivel de calidad del aire es registrado por la Estación de la Universitat Politècnica es la más limpia, pues es la que presenta unos niveles de contaminación más bajos si nos centramos en los contaminantes más comunmente estudiados.

6 b) Para poder hacernos una idea de la fiabilidad y la precisión de los resultados anteriores debemos comprobar que el número de muestras de las que se dispone para hallar los valores estadísticos ha sido lo suficientemente grande como para que dicho estadístico tenga un “valor real”. Para ello vamos a mirar qué agente contaminante ha sido el menos registrado en todas las zonas a nivel de veces que se ha registrado sobre el total de días (valores perdidos).

Miraremos el contaminante que tiene más valores perdidos en cada estación, es decir, aquel #del que hay menos registros, para ver si suele ser el mismo en toda la ciudad:

```
na1 <- sapply(f1, function(x) sum(is.na(x)));na1
```

```
## fecha PM2.5 SO2 NO NO2 PM10 NOx Ozono
## 0 68 125 68 68 68 68 62
```

```
#Más NA en SO2
na3 <- sapply(f2, function(x) sum(is.na(x)));na3
```

```
## fecha PM2.5 PM1 SO2 CO NO NO2 PM10 NOx Ozono
## 0 98 101 93 288 120 122 96 120 46
```

```
#Más NA en CO
na4 <- sapply(f3, function(x) sum(is.na(x)));na4
```

```
## fecha PM2.5 Xileno SO2 CO NO NO2 PM10
## 0 61 728 154 215 29 29 130
## NOx Ozono Tolueno Bencenol Ruido Benceno
## 29 208 731 752 41 752
```

```
#Más NA en CO
na5 <- sapply(f4, function(x) sum(is.na(x)));na5
```

```
## fecha PM2.5 SO2 NO NO2 PM10 Ni NOx Ozono As Pb Cd
## 0 701 212 181 181 620 1725 181 74 1724 1726 1724
```

```
#Más NA en PM2.5
na6 <- sapply(f5, function(x) sum(is.na(x)));na6
```

```
## fecha PM2.5 SO2 CO NO NO2 PM10 NOx
## 0 822 181 152 149 149 822 149
## Ozono velocidad
## 74 0
```

```
#Más NA en PM10 y PM2.5
na7 <- sapply(f6, function(x) sum(is.na(x)));na7
```

```
## fecha SO2 NO NO2 PM10 Ni NOx Ozono As Pb BaP Cd
## 0 100 147 147 923 1801 147 88 1795 1801 2407 1801
```

```
#Más NA en PM10

# Se aprecia que en las dos estaciones donde se registran algunos elementos pesados (5 y 7),
# como As, Pb o Cd hay una gran cantidad de valores faltantes de estos, llegando a haber incluso
# más valores perdidos que observaciones numéricas.

# Por otro lado vemos también que la estación de la Universitat Politècnica es la que contiene,
# con diferencia, menos valores perdidos en conjunto.

# Las observaciones de partículas en suspensión de 2.5 y 10 micras (PM2.5 y PM10) también
# presentan una gran cantidad de valores perdidos en las estaciones de Viveros,
# Boulevard Sud y Av. de França.

# En un plano general no podemos decir que haya un parámetro que destaque por la gran cantidad
# de valores perdidos que hay en sus observaciones, siendo esto común para todas las estaciones.
```

Se aprecia que en las dos estaciones donde se registran algunos elementos pesados (5 y 7), como As, Pb o Cd hay una gran cantidad de valores faltantes de estos, llegando a haber incluso más valores perdidos que observaciones numéricas.

Por otro lado vemos también que la estación de la Universitat Politècnica es la que contiene, con diferencia, menos valores perdidos en conjunto.

Las observaciones de partículas en suspensión de 2.5 y 10 micras (PM2.5 y PM10) también presentan una gran cantidad de valores perdidos en las estaciones de Viveros, Boulevard Sud y Av. de França.

En un plano general no podemos decir que haya un parámetro que destaque por la gran cantidad de valores perdidos que hay en sus observaciones, siendo esto común para todas las estaciones.

7. Se sabe que en el territorio nacional, en un plano general, durante los meses de confinamiento a raíz de la pandemia de la Covid-19 hubo una reducción de la cantidad de gases contaminantes emitidos a la atmósfera debido a las restricciones de movilidad a la que nos vimos sujetos. ¿Hay evidencias de ello si nos basamos en las mediciones que han hecho las estaciones valencianas cuyos datos estamos estudiando?

El 1 de mayo se publica en el BOE una Orden que establece que apartir del 2 de mayo se puede salir a hacer deporte y a pasear. A su vez, hay unos pocos territorios que avanzan de fase (véanse las fases de la desescalada: https://es.wikipedia.org/wiki/Confinamiento_por_la_pandemia_de_COVID-19_en_Espa%C3%B1a#Desescalada).

El 21 de junio acaba el estado de alarma y todos los territorios pasan a la nueva normalidad, por lo que la mayoría de ellos ya estaban en las fases finales de la desescalada a mediados de ese mes. Es por eso que nos hemos basado en los datos recogidos desde la declaración del estado de alarma (14 de mayo) hasta el 31 de mayo. Es decir, contamos con los datos de un mes y medio de confinamiento:

```
# Seleccionamos las observaciones correspondientes a las fechas establecidas como de confinamiento:
```

```
covid1<- slice(f1, 2242:2319)
covid3 <- slice(f2, 2256:2333)
covid4 <- slice(f3, 2249:2325)
covid5 <- slice(f4, 2255:2332)
covid6 <- slice(f5, 2235:2312)
covid7 <- slice(f6, 2243:2320)
```

```
# Media de cada contaminante durante el confinamiento:
```

```
conf1 <- covid1 %>%summarise_if(is.numeric, mean, na.rm=TRUE);conf1
```

```
##      PM2.5  SO2      NO      NO2      PM10      NOx      Ozono
## 1 10.52564 3.24 2.012821 5.74359 12.46154 8.333333 62.15385
```

```
conf3 <- covid3 %>%summarise_if(is.numeric, mean, na.rm=TRUE);conf3
```

```
##      PM2.5      PM1      SO2      CO      NO      NO2      PM10      NOx
## 1 12.60256 9.525641 4.216667 0.1333333 3.192308 6.833333 13.80769 11.78205
##      Ozono
## 1 61.29487
```

```
conf4 <- covid4 %>% select(-`Ruido`)%>%summarise_if(is.numeric, mean, na.rm=TRUE);conf4
```

```
##      PM2.5      SO2      CO      NO      NO2      PM10      NOx      Ozono
## 1 6.917808 4.662338 0.1383562 4.493506 11.06494 9.30137 17.63636 66.07246
##      Tolueno      Benceno
## 1 2.327778 0.5578947
```

```
conf5 <- covid5 %>% summarise_if(is.numeric, mean, na.rm=TRUE);conf5
```

```
##      PM2.5      SO2      NO      NO2      PM10      Ni      NOx      Ozono
## 1 9.448718 3.038961 2.844156 7.961039 17.35065 0.8489474 11.85714 65.5974
##      As      Pb      Cd
## 1 0.1694737 0.008947368 0.06289474
```

```
conf6 <- covid6 %>% summarise_if(is.numeric, mean, na.rm=TRUE);conf6
```

```
##      PM2.5      SO2      CO NO      NO2      PM10      NOx      Ozono
## 1 6.435897 3.705128 0.1179487 2 7.474359 9.666667 10.65385 71.23077
```

```
conf7 <- covid7 %>% summarise_if(is.numeric, mean, na.rm=TRUE);conf7
```

```
##      SO2      NO      NO2      PM10      Ni NOx      Ozono      As      Pb BaP
## 1 3.192308 4.444444 14.81944 19.02941 1.4475 21.5 65.97436 0.20775 0.0095 0.04
##      Cd
## 1 0.09675
```

```
# Selección de fechas anteriores al confinamiento y media por contaminante:
no1 <- slice(f1, 1:2241) %>% summarise_if(is.numeric, mean, na.rm = TRUE);no1
```

```
##      PM2.5      SO2      NO      NO2      PM10      NOx      Ozono
## 1 11.72636 3.333961 6.914864 23.19144 17.74646 33.41141 53.84686
```

```
no3 <- slice(f2, 1:2255) %>% summarise_if(is.numeric, mean, na.rm = TRUE);no3
```

```
##      PM2.5      PM1      SO2      CO      NO      NO2      PM10      NOx
## 1 15.08457 11.35909 3.371219 0.1547038 9.795316 24.65588 17.4132 39.35363
##      Ozono
## 1 50.29022
```

```
no4 <- slice(f3, 1:2248) %>% summarise_if(is.numeric, mean, na.rm = TRUE);no4
```

```
##      PM2.5      SO2      CO      NO      NO2      PM10      NOx      Ozono
## 1 12.05467 4.270296 0.2025828 22.23503 38.13868 22.69097 71.99595 43.65944
##      Tolueno      Ruido      Benceno
## 1 5.541667 60.07342 1.734679
```

```
no5 <- slice(f4, 1:2254) %>% summarise_if(is.numeric, mean, na.rm = TRUE);no5
```

```
##      PM2.5      SO2      NO      NO2      PM10      Ni      NOx      Ozono
## 1 15.65258 3.433953 8.570053 24.24603 24.91667 2.369447 36.94319 50.92009
##      As      Pb      Cd
## 1 0.3206916 0.0312344 0.07542185
```

```
no6 <- slice(f5, 1:2234) %>% summarise_if(is.numeric, mean, na.rm = TRUE);no6
```

```
##      PM2.5      SO2      CO      NO      NO2      PM10      NOx      Ozono  
## 1 11.18997 3.46068 0.1619933 9.623501 28.93381 17.7952 43.27434 49.3579
```

```
no7 <- slice(f6, 1:2242) %>% summarise_if(is.numeric, mean, na.rm = TRUE);no7
```

```
##      SO2      NO      NO2      PM10      Ni      NOx      Ozono      As  
## 1 3.690233 15.23216 32.19125 25.63649 4.49634 55.32826 47.16944 0.6278241  
##      Pb      BaP      Cd  
## 1 0.01376947 0.1107246 0.1697508
```

Conclusión:

Después de hacer este trabajo y haber manipulado los datos hemos descubierto que la principal causa de las emisiones de gases contaminantes son por culpa del ser humano ya que como hemos comprobado los valores más altos de contaminación son los de las estaciones que están cerca de zonas industriales con fábricas, carreteras o ciudades.