

Análisis de datos ambientales

Rosario Huertas Henao

1/7/2021

EXTRACCIÓN Y CARGA DE DATOS

Carga de datos

```
datos2018 <- read_delim("datos201812.CSV",  
  ";", escape_double = FALSE, trim_ws = TRUE)
```

```
##  
## -- Column specification -----  
## cols(  
##   .default = col_character(),  
##   PROVINCIA = col_double(),  
##   ESTACION = col_double(),  
##   MAGNITUD = col_double(),  
##   ANO = col_double()  
## )  
## i Use 'spec()' for the full column specifications.
```

```
datos2019 <- read_delim("datos201912.CSV",  
  ";", escape_double = FALSE, trim_ws = TRUE)
```

```
##  
## -- Column specification -----  
## cols(  
##   .default = col_character(),  
##   PROVINCIA = col_double(),  
##   ESTACION = col_double(),  
##   MAGNITUD = col_double(),  
##   ANO = col_double()  
## )  
## i Use 'spec()' for the full column specifications.
```

```
datos2020 <- read_delim("datos202012.CSV",  
  ";", escape_double = FALSE, trim_ws = TRUE)
```

```
##  
## -- Column specification -----
```

```
## cols(
##   .default = col_character(),
##   PROVINCIA = col_double(),
##   ESTACION = col_double(),
##   MAGNITUD = col_double(),
##   ANO = col_double()
## )
## i Use 'spec()' for the full column specifications.
```

```
datos2021 <- read_delim("datos202103.CSV",
  ";", escape_double = FALSE, trim_ws = TRUE)
```

```
##
## -- Column specification -----
## cols(
##   .default = col_character(),
##   PROVINCIA = col_double(),
##   ESTACION = col_double(),
##   MAGNITUD = col_double(),
##   ANO = col_double()
## )
## i Use 'spec()' for the full column specifications.
```

Descripcion de los datos

Primero vemos que variables aparecen. Para hacernos una idea de la estructura hacemos un head.

```
names(datos2020)
```

```
## [1] "PROVINCIA"      "MUNICIPIO"      "ESTACION"       "MAGNITUD"
## [5] "PUNTO_MUESTREO" "ANO"            "MES"            "D01"
## [9] "V01"            "D02"            "V02"            "D03"
## [13] "V03"            "D04"            "V04"            "D05"
## [17] "V05"            "D06"            "V06"            "D07"
## [21] "V07"            "D08"            "V08"            "D09"
## [25] "V09"            "D10"            "V10"            "D11"
## [29] "V11"            "D12"            "V12"            "D13"
## [33] "V13"            "D14"            "V14"            "D15"
## [37] "V15"            "D16"            "V16"            "D17"
## [41] "V17"            "D18"            "V18"            "D19"
## [45] "V19"            "D20"            "V20"            "D21"
## [49] "V21"            "D22"            "V22"            "D23"
## [53] "V23"            "D24"            "V24"            "D25"
## [57] "V25"            "D26"            "V26"            "D27"
## [61] "V27"            "D28"            "V28"            "D29"
## [65] "V29"            "D30"            "V30"            "D31"
## [69] "V31"
```

```
head(datos2020)
```

```
## # A tibble: 6 x 69
```

```
## PROVINCIA MUNICIPIO ESTACION MAGNITUD PUNTO_MUESTREO ANO MES D01
## <dbl> <chr> <dbl> <dbl> <chr> <dbl> <chr> <chr>
## 1 28 079 4 1 28079004_1_38 2020 01 00009
## 2 28 079 4 1 28079004_1_38 2020 02 00008
## 3 28 079 4 1 28079004_1_38 2020 03 00000
## 4 28 079 4 1 28079004_1_38 2020 04 00000
## 5 28 079 4 1 28079004_1_38 2020 05 00005
## 6 28 079 4 1 28079004_1_38 2020 06 00004
## # ... with 61 more variables: V01 <chr>, D02 <chr>, V02 <chr>, D03 <chr>,
## # V03 <chr>, D04 <chr>, V04 <chr>, D05 <chr>, V05 <chr>, D06 <chr>,
## # V06 <chr>, D07 <chr>, V07 <chr>, D08 <chr>, V08 <chr>, D09 <chr>,
## # V09 <chr>, D10 <chr>, V10 <chr>, D11 <chr>, V11 <chr>, D12 <chr>,
## # V12 <chr>, D13 <chr>, V13 <chr>, D14 <chr>, V14 <chr>, D15 <chr>,
## # V15 <chr>, D16 <chr>, V16 <chr>, D17 <chr>, V17 <chr>, D18 <chr>,
## # V18 <chr>, D19 <chr>, V19 <chr>, D20 <chr>, V20 <chr>, D21 <chr>,
## # V21 <chr>, D22 <chr>, V22 <chr>, D23 <chr>, V23 <chr>, D24 <chr>,
## # V24 <chr>, D25 <chr>, V25 <chr>, D26 <chr>, V26 <chr>, D27 <chr>,
## # V27 <chr>, D28 <chr>, V28 <chr>, D29 <chr>, V29 <chr>, D30 <chr>,
## # V30 <chr>, D31 <chr>, V31 <chr>
```

TRANSFORMACIÓN Y LIMPIEZA DE DATOS

Transformacion del formato de los datos de 2020

Calculamos una columna día y una columna valor. Nos quedamos solo con los válidos.

```
datos_2020 <- datos2020 %>% select(-names(datos2020)[str_starts(names(datos2020),"V")])%>%
  gather(key = "dia", value = "valor", names(datos2020)[str_starts(names(datos2020),"D")]) %>%
  mutate(dia= substring(dia,2))

validos_2020 <- datos2020 %>%select(-names(datos2020)[str_starts(names(datos2020),"D")]) %>%
  gather(key = "dia", value = "validez",
    names(datos2020)[str_starts(names(datos2020),"V")]) %>% mutate(dia= substring(dia,2))

new_2020 <- inner_join(datos_2020, validos_2020, by = c("ANO", "ESTACION", "MAGNITUD", "MES", "MUNICIPIO"))

print("La tabla siguiente muestra el numero de registros validos y no validos del año 2020")
```

```
## [1] "La tabla siguiente muestra el numero de registros validos y no validos del año 2020"
```

```
table(new_2020$validez)
```

```
##
##      N      V
## 2019 54897
```

```
new_2020 <- new_2020 %>% filter(validez=="V")
```

Vemos como quedan los datos transformados

```
head(new_2020)
```

```
## # A tibble: 6 x 10
##   PROVINCIA MUNICIPIO ESTACION MAGNITUD PUNTO_MUESTREO ANO MES dia
##   <dbl> <chr>         <dbl>    <dbl> <chr>         <dbl> <chr> <chr>
## 1      28 079          4        1 28079004_1_38 2020 01 01
## 2      28 079          4        1 28079004_1_38 2020 02 01
## 3      28 079          4        1 28079004_1_38 2020 05 01
## 4      28 079          4        1 28079004_1_38 2020 06 01
## 5      28 079          4        1 28079004_1_38 2020 07 01
## 6      28 079          4        1 28079004_1_38 2020 08 01
## # ... with 2 more variables: valor <chr>, validez <chr>
```

Transformacion del formato de los datos de 2021

```
datos_2021 <- datos2021 %>% select(-names(datos2021)[str_starts(names(datos2021),"V")])%>%
  gather(key = "dia", value = "valor", names(datos2021)[str_starts(names(datos2021),"D")]) %>%
  mutate(dia= substring(dia,2))

validos_2021 <- datos2021 %>%select(-names(datos2021)[str_starts(names(datos2021),"D")]) %>%
  gather(key = "dia", value = "validez",
    names(datos2021)[str_starts(names(datos2021),"V")]) %>% mutate(dia= substring(dia,2))

new_2021 <- inner_join(datos_2021, validos_2021, by = c("ANO", "ESTACION", "MAGNITUD", "MES", "MUNICIPIO"))

print("La tabla siguiente muestra el numero de registros validos y no validos del año 2021")
```

```
## [1] "La tabla siguiente muestra el numero de registros validos y no validos del año 2021"
```

```
table(new_2021$validez)
```

```
##
##      N      V
## 1267 11846
```

```
new_2021 <- new_2021%>% filter(validez=="V")
```

```
head(new_2021)
```

```
## # A tibble: 6 x 10
##   PROVINCIA MUNICIPIO ESTACION MAGNITUD PUNTO_MUESTREO ANO MES dia
##   <dbl> <chr>         <dbl>    <dbl> <chr>         <dbl> <chr> <chr>
## 1      28 079          4        1 28079004_1_38 2021 01 01
## 2      28 079          4        1 28079004_1_38 2021 03 01
## 3      28 079          4        6 28079004_6_48 2021 01 01
## 4      28 079          4        6 28079004_6_48 2021 03 01
## 5      28 079          4        7 28079004_7_8 2021 01 01
## 6      28 079          4        7 28079004_7_8 2021 03 01
## # ... with 2 more variables: valor <chr>, validez <chr>
```

Transformacion del formato de los datos de 2018

```
datos_2018 <- datos2018 %>% select(-names(datos2018)[str_starts(names(datos2018), "V")]) %>%
  gather(key = "dia", value = "valor", names(datos2018)[str_starts(names(datos2018), "D")]) %>%
  mutate(dia= substring(dia,2))

validos_2018 <- datos2018 %>%select(-names(datos2018)[str_starts(names(datos2018), "D")]) %>%
  gather(key = "dia", value = "validez",
        names(datos2018)[str_starts(names(datos2018), "V")]) %>% mutate(dia= substring(dia,2))

new_2018 <- inner_join(datos_2018, validos_2018, by = c("ANO", "ESTACION", "MAGNITUD", "MES", "MUNICIPIO"))

print("La tabla siguiente muestra el numero de registros validos y no validos del año 2018")
```

```
## [1] "La tabla siguiente muestra el numero de registros validos y no validos del año 2018"
```

```
table(new_2018$validez)
```

```
##
##      N      V
## 1598 54388
```

```
new_2018 <- new_2018 %>% filter(validez=="V")
```

```
head(new_2018)
```

```
## # A tibble: 6 x 10
##   PROVINCIA MUNICIPIO ESTACION MAGNITUD PUNTO_MUESTREO ANO MES dia
##   <dbl> <chr>      <dbl>    <dbl> <chr>      <dbl> <chr> <chr>
## 1      28 079        4        1 28079004_1_38 2018 01 01
## 2      28 079        4        1 28079004_1_38 2018 02 01
## 3      28 079        4        1 28079004_1_38 2018 03 01
## 4      28 079        4        1 28079004_1_38 2018 04 01
## 5      28 079        4        1 28079004_1_38 2018 05 01
## 6      28 079        4        1 28079004_1_38 2018 06 01
## # ... with 2 more variables: valor <chr>, validez <chr>
```

Transformacion del formato de los datos de 2019

```
datos_2019 <- datos2019 %>% select(-names(datos2019)[str_starts(names(datos2019), "V")]) %>%
  gather(key = "dia", value = "valor", names(datos2019)[str_starts(names(datos2019), "D")]) %>%
  mutate(dia= substring(dia,2))

validos_2019 <- datos2019 %>%select(-names(datos2019)[str_starts(names(datos2019), "D")]) %>%
  gather(key = "dia", value = "validez",
        names(datos2019)[str_starts(names(datos2019), "V")]) %>% mutate(dia= substring(dia,2))

new_2019 <- inner_join(datos_2019, validos_2019, by = c("ANO", "ESTACION", "MAGNITUD", "MES", "MUNICIPIO"))

print("La tabla siguiente muestra el numero de registros validos y no validos del año 2019")
```

```
## [1] "La tabla siguiente muestra el numero de registros validos y no validos del año 2019"
```

```
table(new_2019$validez)
```

```
##
##      N      V
## 2042 54874
```

```
new_2019 <- new_2019%>% filter(validez=="V")
```

```
head(new_2019)
```

```
## # A tibble: 6 x 10
##   PROVINCIA MUNICIPIO ESTACION MAGNITUD PUNTO_MUESTREO ANO MES dia
##   <dbl> <chr>      <dbl>    <dbl> <chr>      <dbl> <chr> <chr>
## 1      28 079          4        1 28079004_1_38 2019 01 01
## 2      28 079          4        1 28079004_1_38 2019 02 01
## 3      28 079          4        1 28079004_1_38 2019 03 01
## 4      28 079          4        1 28079004_1_38 2019 04 01
## 5      28 079          4        1 28079004_1_38 2019 05 01
## 6      28 079          4        1 28079004_1_38 2019 06 01
## # ... with 2 more variables: valor <chr>, validez <chr>
```

Unificamos todos los registros históricos en un único dataframe:

```
data <- rbind(new_2018, new_2019)
data <- rbind(data, new_2020)
data <- rbind(data, new_2021)
View(data)
```

```
# Descargamos en un único fichero csv los datos
write.csv(data, "data.csv")
```

```
str(data)
```

```
## tibble [176,005 x 10] (S3: tbl_df/tbl/data.frame)
##  $ PROVINCIA      : num [1:176005] 28 28 28 28 28 28 28 28 28 28 ...
##  $ MUNICIPIO      : chr [1:176005] "079" "079" "079" "079" ...
##  $ ESTACION       : num [1:176005] 4 4 4 4 4 4 4 4 4 4 ...
##  $ MAGNITUD       : num [1:176005] 1 1 1 1 1 1 1 1 1 1 ...
##  $ PUNTO_MUESTREO: chr [1:176005] "28079004_1_38" "28079004_1_38" "28079004_1_38" "28079004_1_38" ..
##  $ ANO            : num [1:176005] 2018 2018 2018 2018 2018 ...
##  $ MES            : chr [1:176005] "01" "02" "03" "04" ...
##  $ dia            : chr [1:176005] "01" "01" "01" "01" ...
##  $ valor          : chr [1:176005] "00001" "00005" "00001" "00002" ...
##  $ validez        : chr [1:176005] "V" "V" "V" "V" ...
```

```
print("Provincias estudiadas")
```

```
## [1] "Provincias estudiadas"
```

```
unique(data$PROVINCIA)
```

```
## [1] 28
```

```
print("Municipios estudiadas")
```

```
## [1] "Municipios estudiadas"
```

```
unique(data$MUNICIPIO)
```

```
## [1] "079"
```

```
print("ESTACIONES estudiadas")
```

```
## [1] "ESTACIONES estudiadas"
```

```
length(unique(data$ESTACION))
```

```
## [1] 24
```

```
print("MAGNITUDES estudiadas")
```

```
## [1] "MAGNITUDES estudiadas"
```

```
length(unique(data$MAGNITUD))
```

```
## [1] 14
```

```
print("PUNTO_MUESTREO (combinaciones) estudiadas")
```

```
## [1] "PUNTO_MUESTREO (combinaciones) estudiadas"
```

```
length(unique(data$PUNTO_MUESTREO))
```

```
## [1] 153
```

Construimos la columna de fechas

Eliminamos columnas día, mes, año porque las cambiamos por fecha y validez, provincia y municipio porque solamente tienen un único valor

Fechas de inicio y fin de los datos

```
min(data$fecha)
```

```
## [1] "2018-01-01"
```

```
max(data$fecha)
```

```
## [1] "2021-03-31"
```

Separamos el punto de muestreo en estacion y forma de medida y quitamos punto de muestreo porque ya está separada su informacion

No nos quedamos con la tecnica, porque se ha usado siempre la de Fluorescencia ultravioleta.(cod. 38)
Traducimos los codigos a su significad, magnitudes y formas de medir.

```
data %<>% separate(PUNTO_MUESTREO, c("cod_estacion","magnitud","tecnica"), "_") %>%  
  select( -magnitud)
```

```
MAGNITUD <- c(1,6,7,8,9,10,12,14,20,30,35,37,38,39,42,43,44)
```

```
nombre_magnitud <- c("Dióxido de azufre",  
  "Monóxido de carbono",  
  "Monoxido de nitrógeno",  
  "Dióxido de nitrógeno",  
  "Partículas <2.5um",  
  "Partículas <10um",  
  "Oxidos de nitrógeno",  
  "Ozono",  
  "Tolueno",  
  "Benceno",  
  "Etilbenceno",  
  "Metaxileno",  
  "Paraxileno",  
  "Ortosileno",  
  "Hidrocarburos Totales(Hexano)",  
  "Metano",  
  "Hidrocarburos no metanicos (Hexano)")
```

```
diccionario <- as.data.frame(MAGNITUD, nombre_magnitud, rownames=FALSE)  
diccionario$nombre_magnitud <- nombre_magnitud
```

```
data <- left_join(data, diccionario)
```

```
## Joining, by = "MAGNITUD"
```

```
## # A tibble: 6 x 7
```

```
##   estacion magnitud cod_estacion tecnica valor fecha      nombre_magnitud  
##   <dbl>    <dbl> <chr>         <chr>  <dbl> <date>    <chr>  
## 1      4      1 28079004     38      1 2018-01-01 Dióxido de azufre  
## 2      4      1 28079004     38      5 2018-02-01 Dióxido de azufre
```



```
## 3      4      1 28079004      38      1 2018-03-01 Dióxido de azufre
## 4      4      1 28079004      38      2 2018-04-01 Dióxido de azufre
## 5      4      1 28079004      38      2 2018-05-01 Dióxido de azufre
## 6      4      1 28079004      38      3 2018-06-01 Dióxido de azufre
```

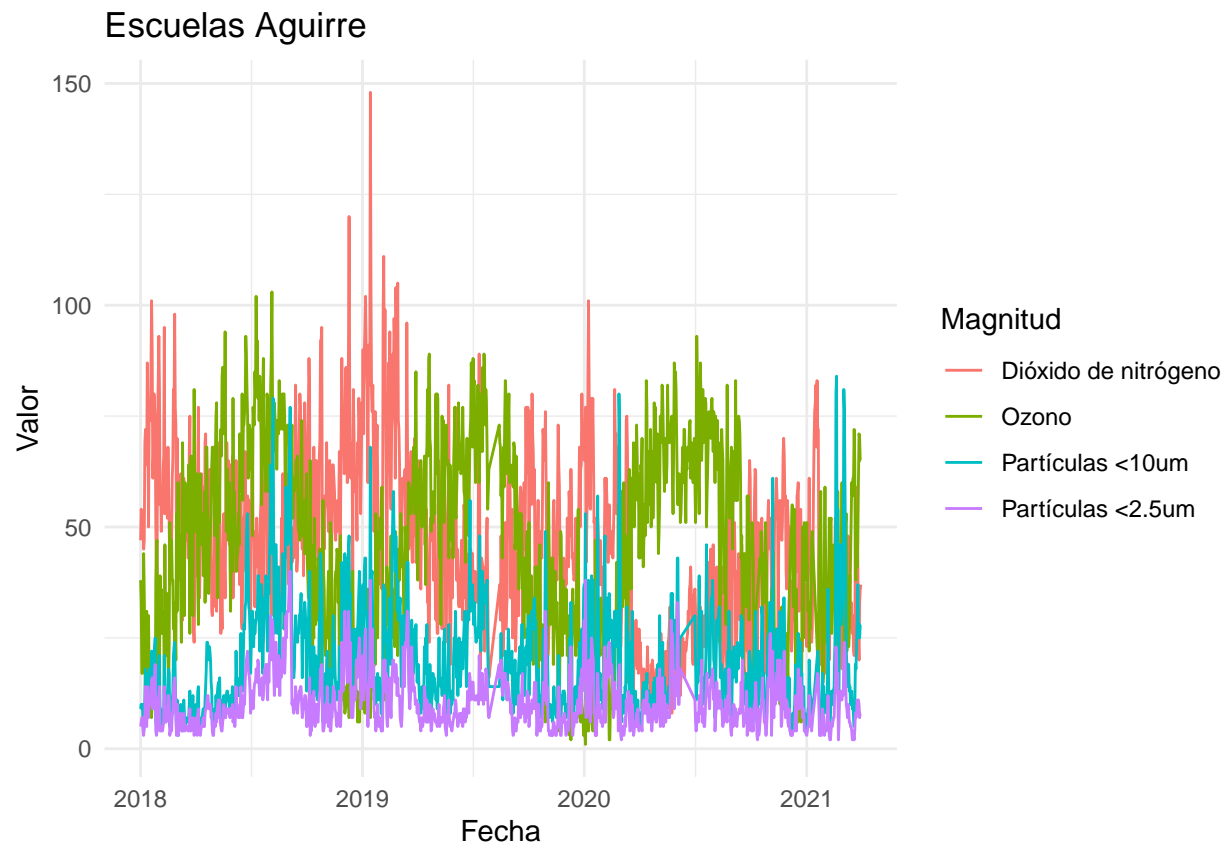
Guardamos en un dataframe las estaciones con los que trabajaremos así como los parámetros. Cuantos datos tengo por cada estación y parámetro

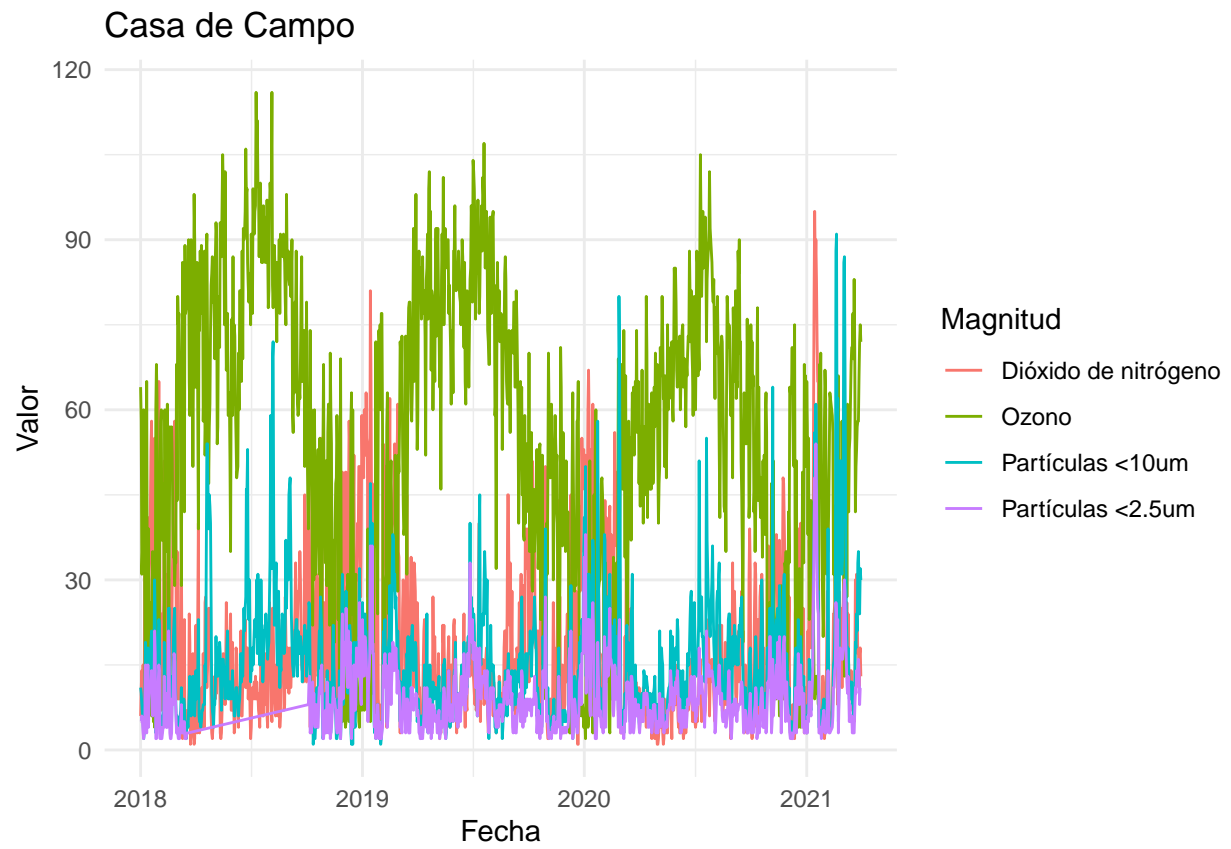
```
## # A tibble: 11 x 4
## # Groups:   estacion, magnitud, nombre_magnitud [11]
##   estacion magnitud nombre_magnitud      n
##   <dbl>    <dbl> <chr>                <int>
## 1      8      8      Dióxido de nitrógeno  1168
## 2      8      9      Partículas <2.5um    1129
## 3      8     10      Partículas <10um    1129
## 4      8     14      Ozono                1144
## 5     24      8      Dióxido de nitrógeno  1178
## 6     24      9      Partículas <2.5um    979
## 7     24     10      Partículas <10um    1172
## 8     24     14      Ozono                1177
## 9     60      8      Dióxido de nitrógeno  1184
## 10    60     10      Partículas <10um    1184
## 11    60     14      Ozono                1181
```

Series temporales Representando todas las magnitudes a la vez los puntos de estación 8,24,60

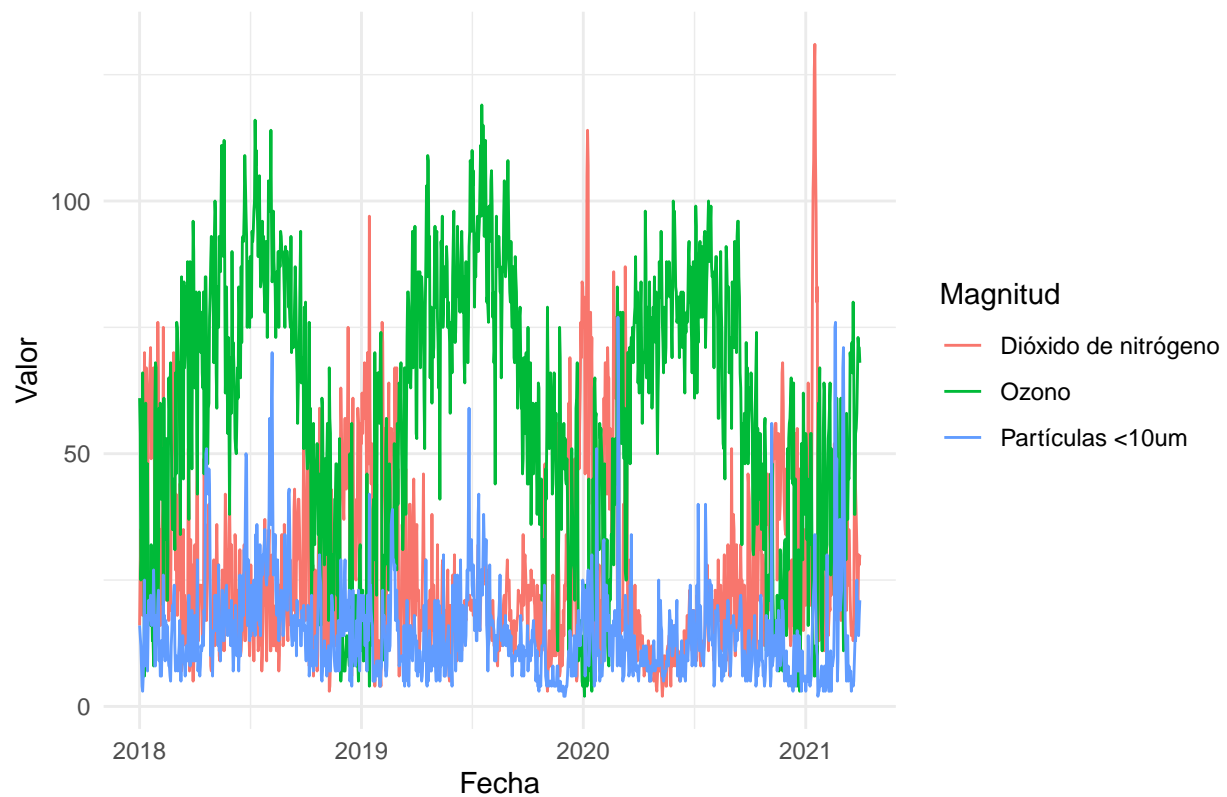
```
puntos<- unique(df$nombre_estacion)
for (punto in puntos){
muestra <- df%>% filter(nombre_estacion==punto)

plot <- ggplot(muestra, aes(x=fecha, y=valor)) +
  geom_line(aes(col=as.factor(nombre_magnitud)))+ theme_minimal()+ ggtitle(punto)+
  xlab("Fecha")+ylab("Valor")+labs(color="Magnitud")
print(plot)
}
```





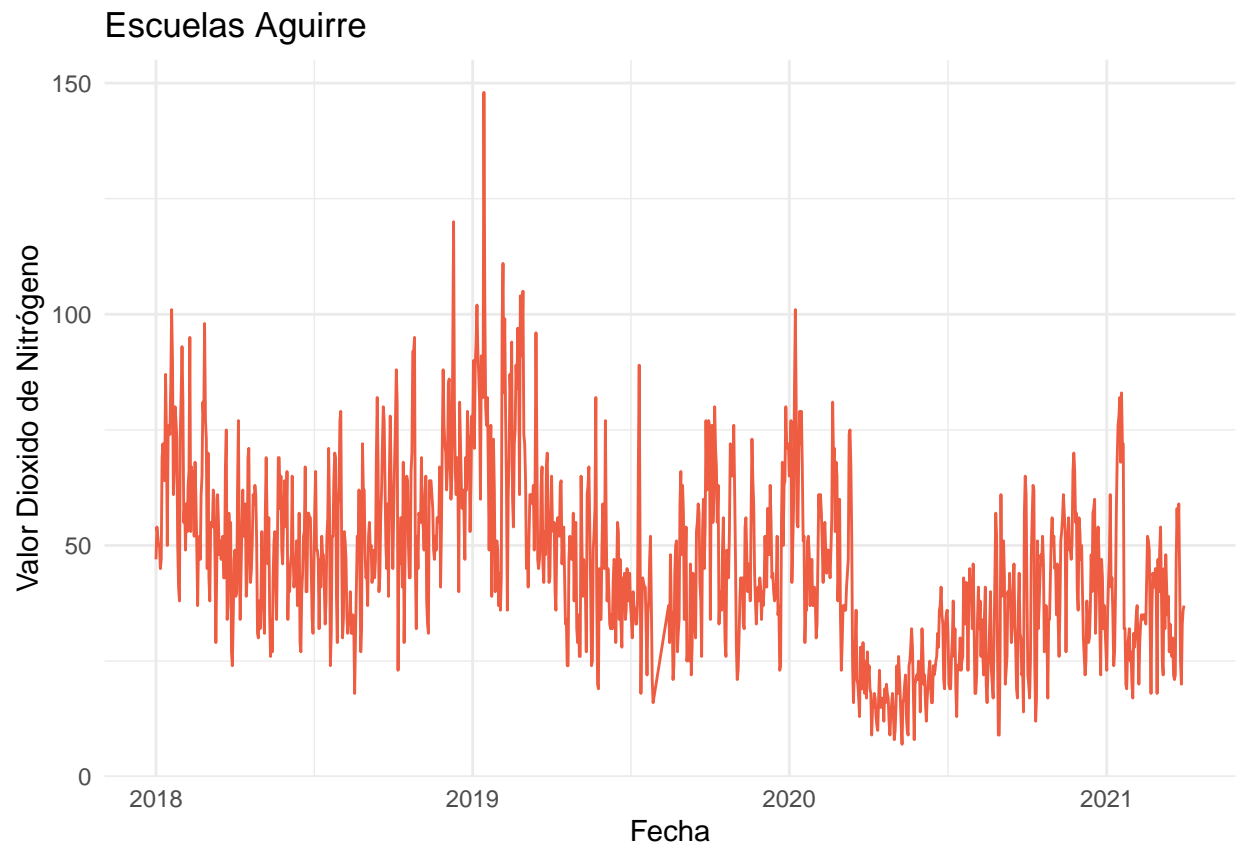
Plaza Tres Olivos

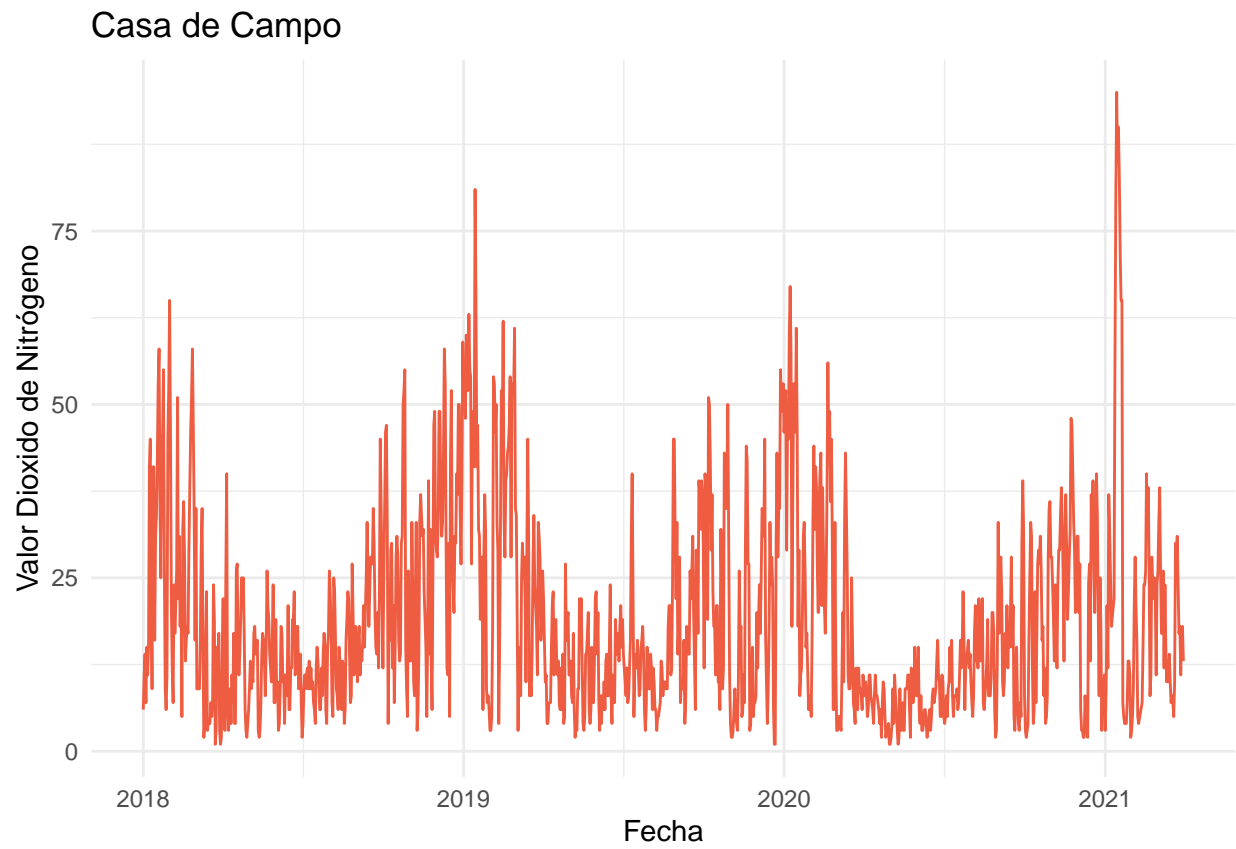


Series temporales Representando la Magnitud 8 - Datos brutos diarios

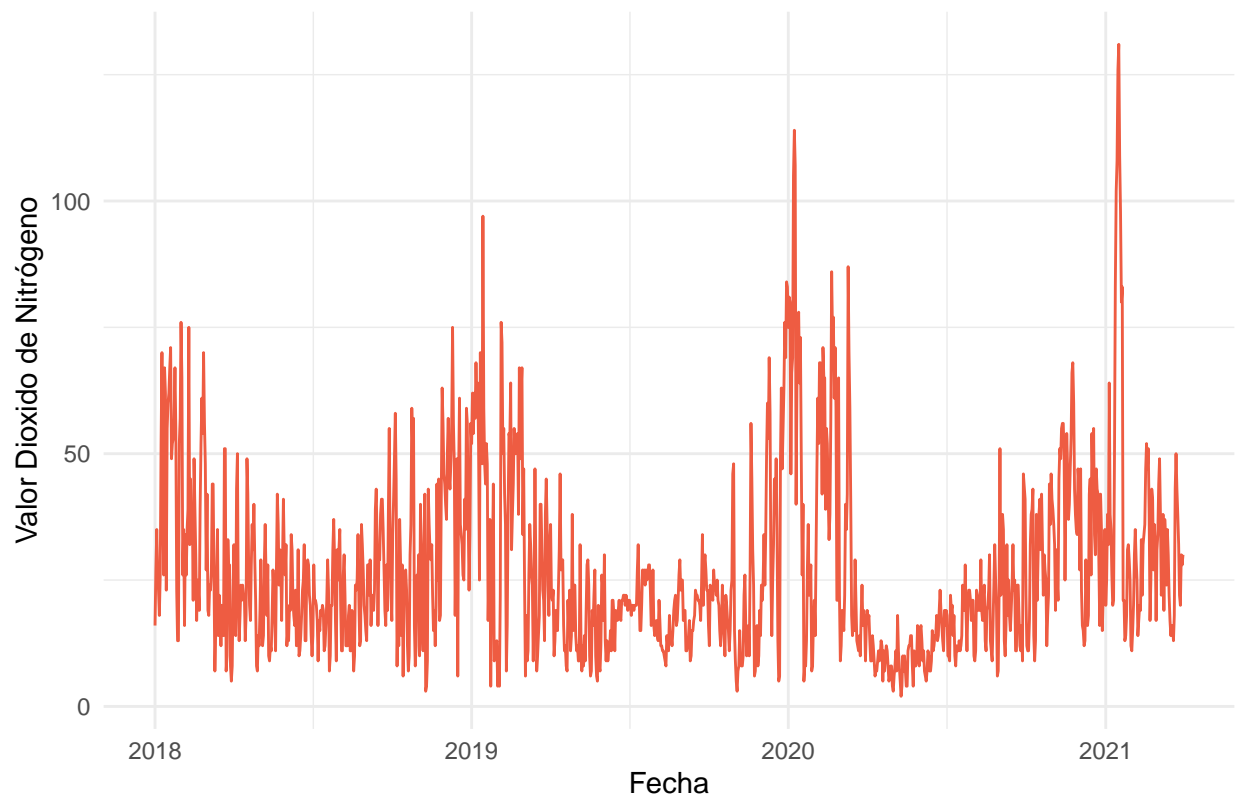
```
for (punto in puntos){
muestra_mag8_brutos <- df %>% filter(nombre_estacion==punto & magnitud==8)

plot_8_brutos <- ggplot(muestra_mag8_brutos, aes(x=fecha, y=valor)) +
  geom_line(color="tomato2") + theme_minimal() + ggtitle(punto) +
  scale_colour_manual(values = c("red")) + labs(color="") +
  xlab("Fecha") + ylab("Valor Dioxido de Nitrógeno")
print(plot_8_brutos)
}
```





Plaza Tres Olivos



Series temporales Representando la Magnitud 8 - Media Anual

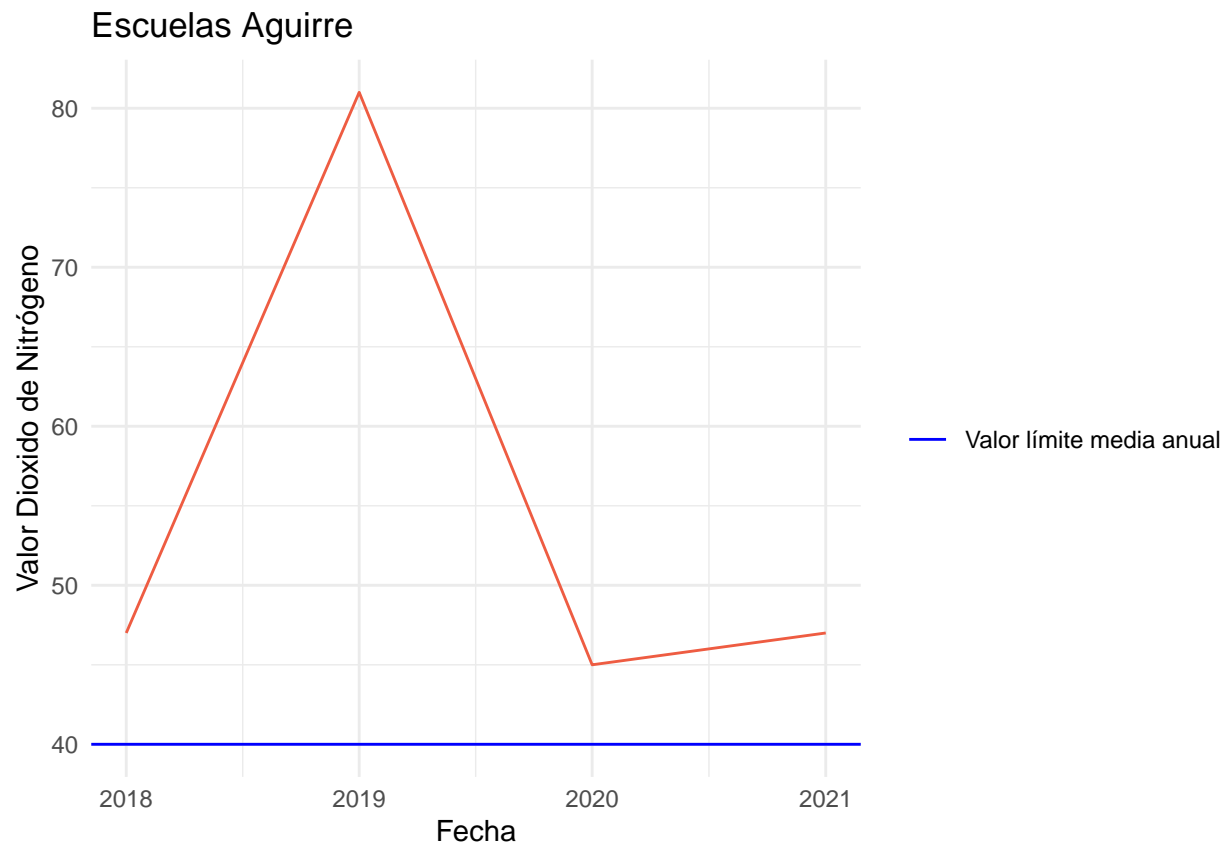
```
for (punto in puntos){
muestra_mag8 <- df %>% filter(nombre_estacion==punto & magnitud == 8)

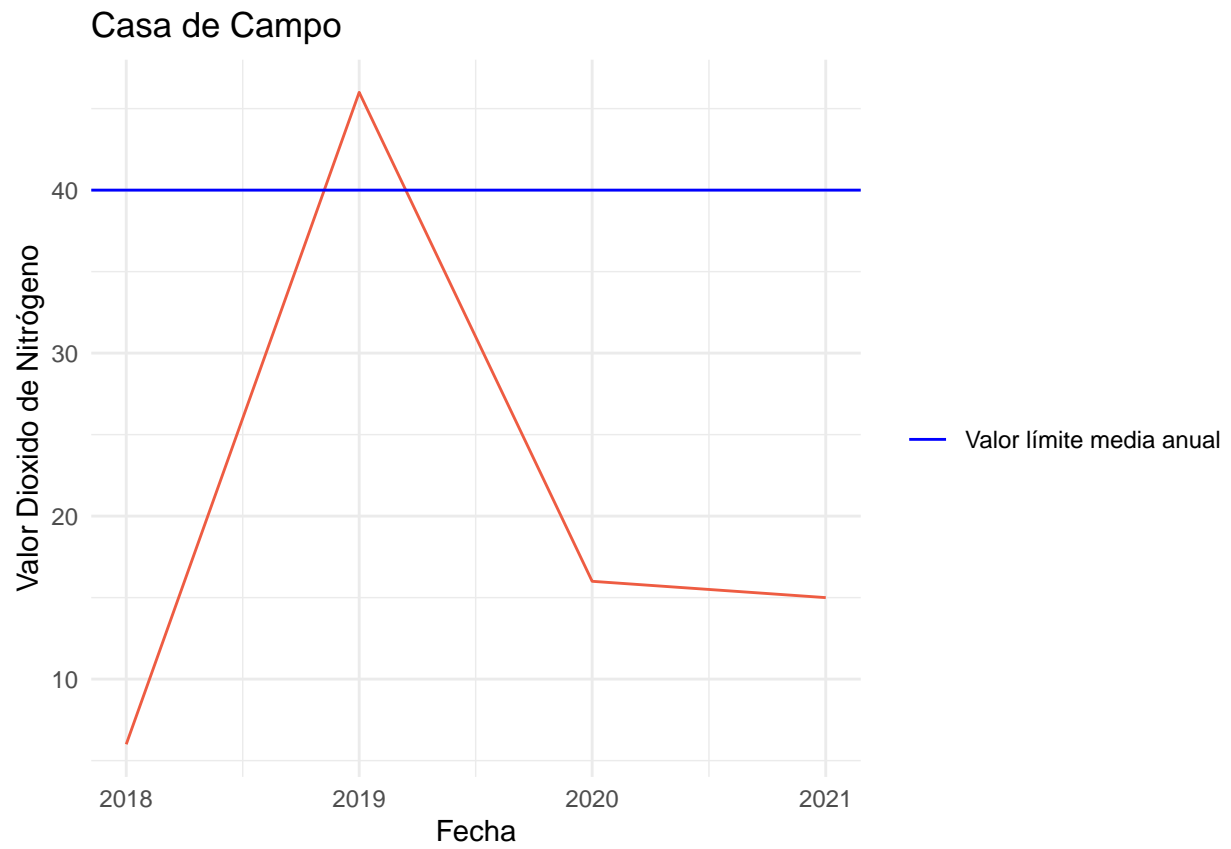
serie_muestra_mag8 <- ts(start= c(2018,1), end=c(2021,1),muestra_mag8$valor, frequency = 1)

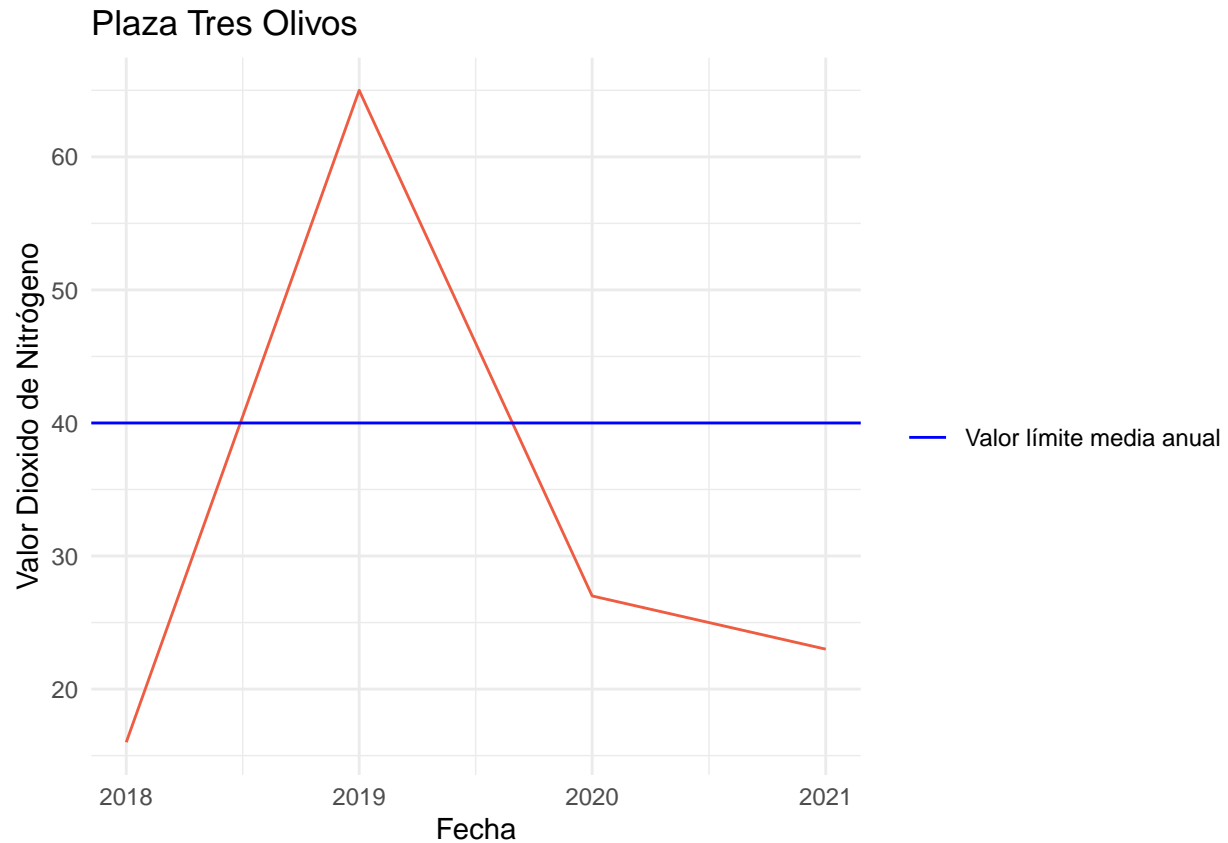
# Convierte de serie temporal a dataframe
mm8 <- ts_df(serie_muestra_mag8)

plot_8 <- ggplot(mm8, aes(x=time, y=value)) +
  geom_line(color="tomato2")+theme_minimal()+ ggtitle(punto)+
  geom_hline(aes(yintercept=40, color="Valor límite media anual"))+
  scale_colour_manual(values = c("blue"))+labs(color="")+
  xlab("Fecha")+ylab("Valor Dioxido de Nitrogeno")

print(plot_8)
}
```



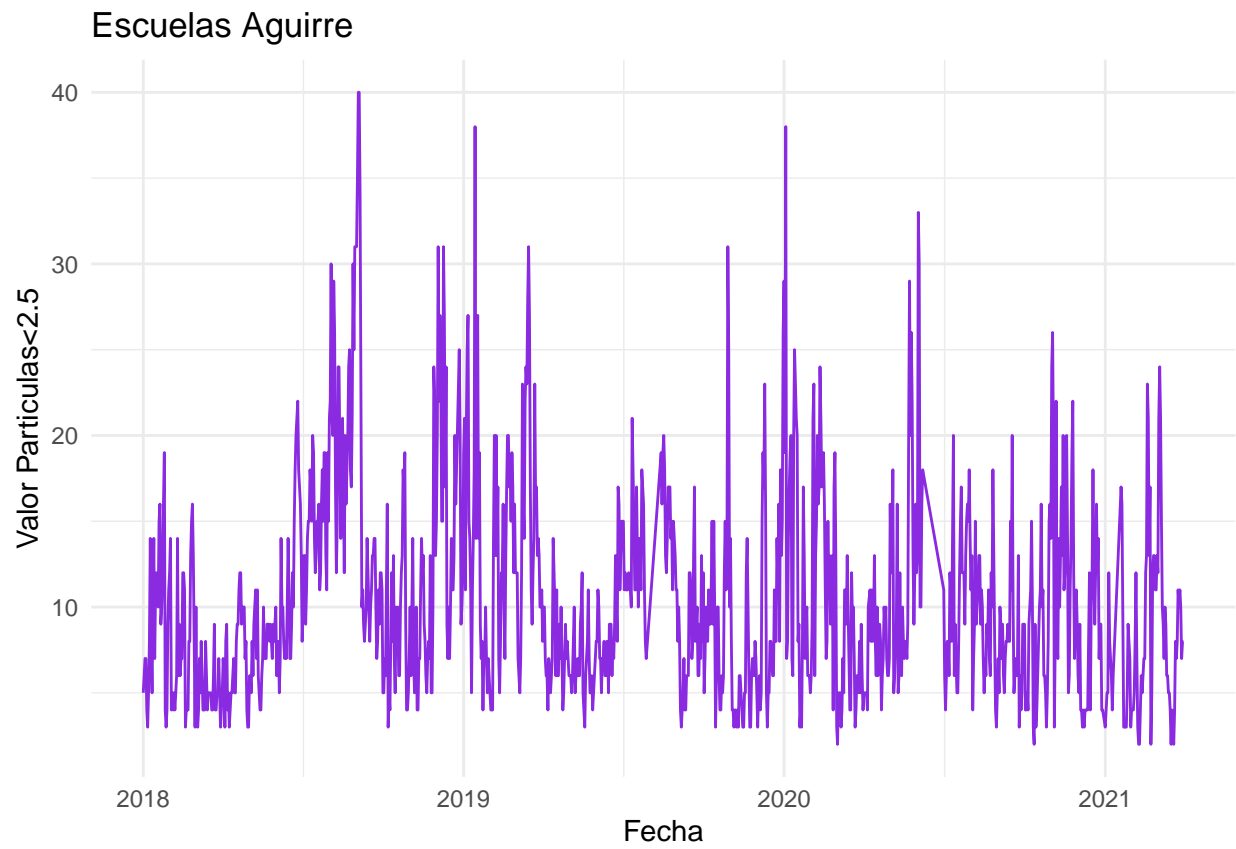


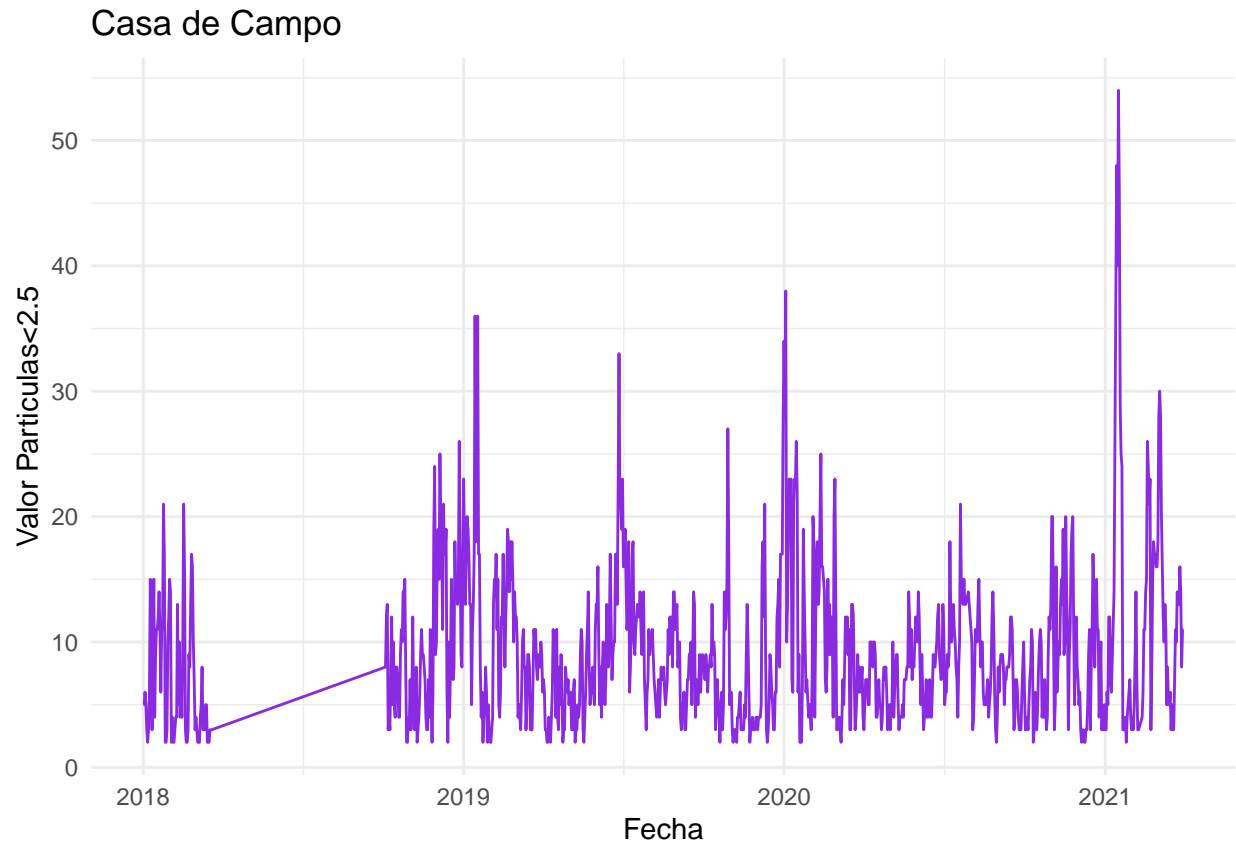


Series temporales Representando la Magnitud 9 - Datos brutos diarios

```
puntos3 <- c("Escuelas Aguirre", "Casa de Campo")
for (punto in puntos3){
  muestra_mag9_brutos <- df %>% filter(nombre_estacion==punto & magnitud==9)

  plot_9_brutos <- ggplot(muestra_mag9_brutos, aes(x=fecha, y=valor)) +
    geom_line(color="blueviolet") + theme_minimal() + ggtitle(punto) +
    scale_colour_manual(values = c("red")) + labs(color="") +
    xlab("Fecha") + ylab("Valor Particulas<2.5")
  print(plot_9_brutos)
}
```





Series temporales Representando la Magnitud 9 - Media Anual

```
puntos4 <- c("Escuelas Aguirre", "Casa de Campo")

for (punto in puntos4){

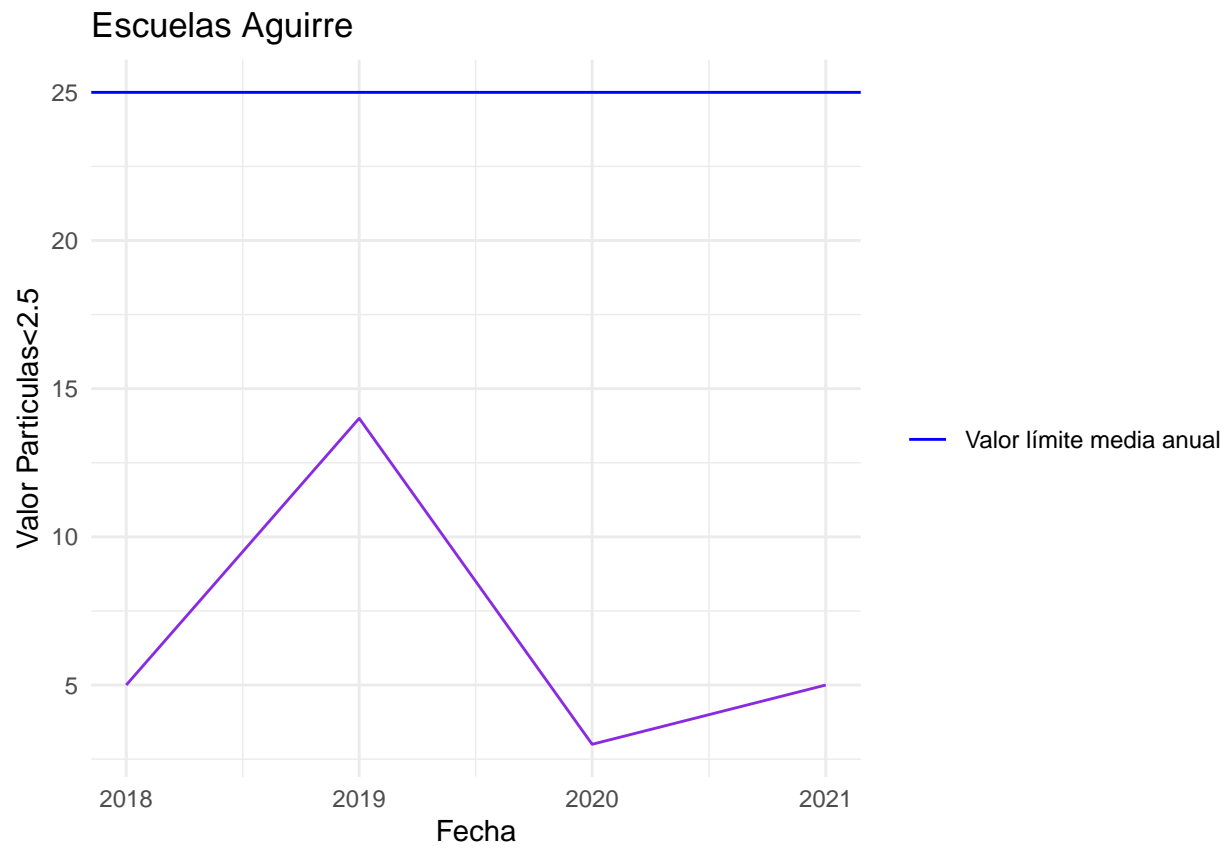
muestra_mag9 <- df%>% filter(nombre_estacion==punto & magnitud == 9)

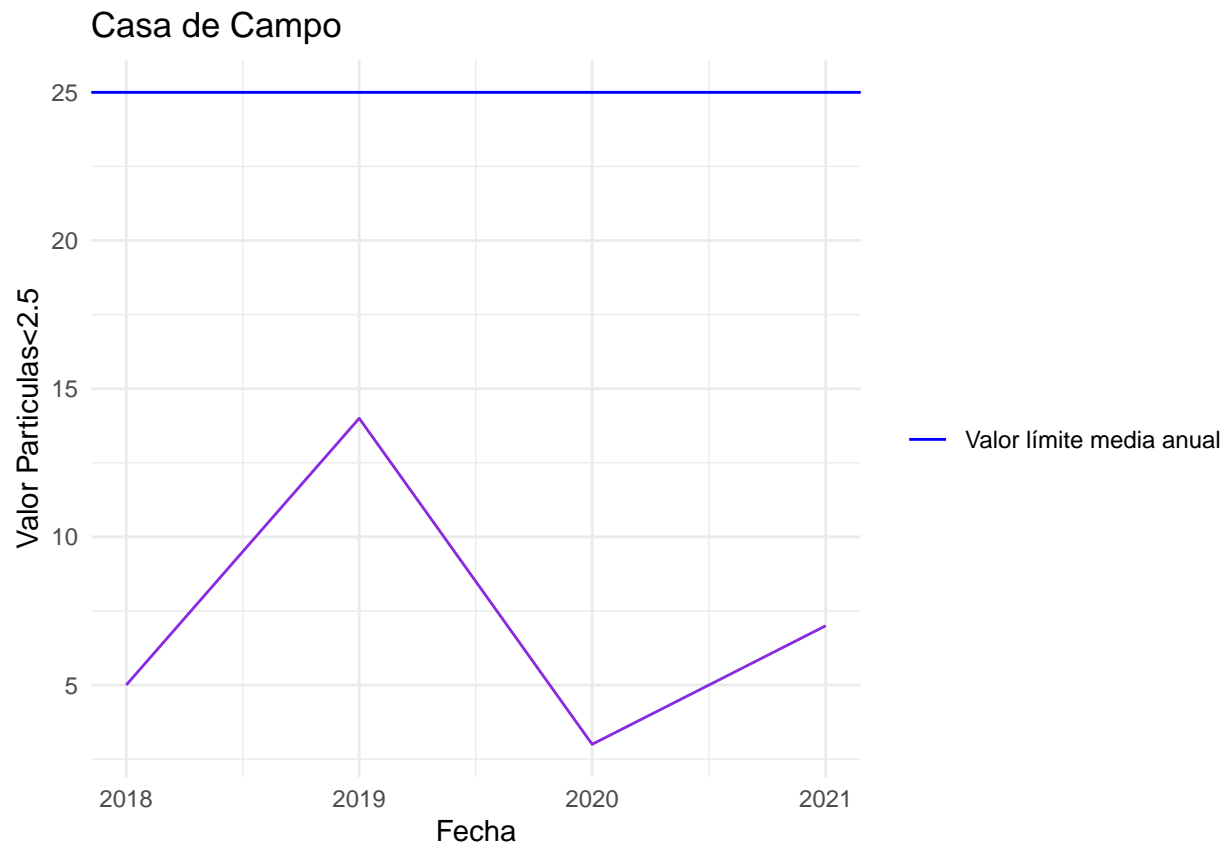
ss <- ts(start= c(2018,1), end=c(2021,1), muestra_mag9$valor, frequency = 1)

# Convierte de serie temporal a dataframe
mm9 <- ts_df(ss)

plot_9 <- ggplot(mm9, aes(x=time, y=value)) +
  geom_line(color="blueviolet")+theme_minimal()+ ggtitle(punto)+
  geom_hline(aes(yintercept=25, color="Valor límite media anual"))+
  scale_colour_manual(values = c("blue"))+labs(color="")+
  xlab("Fecha")+ylab("Valor Partículas<2.5")

print(plot_9)
}
```

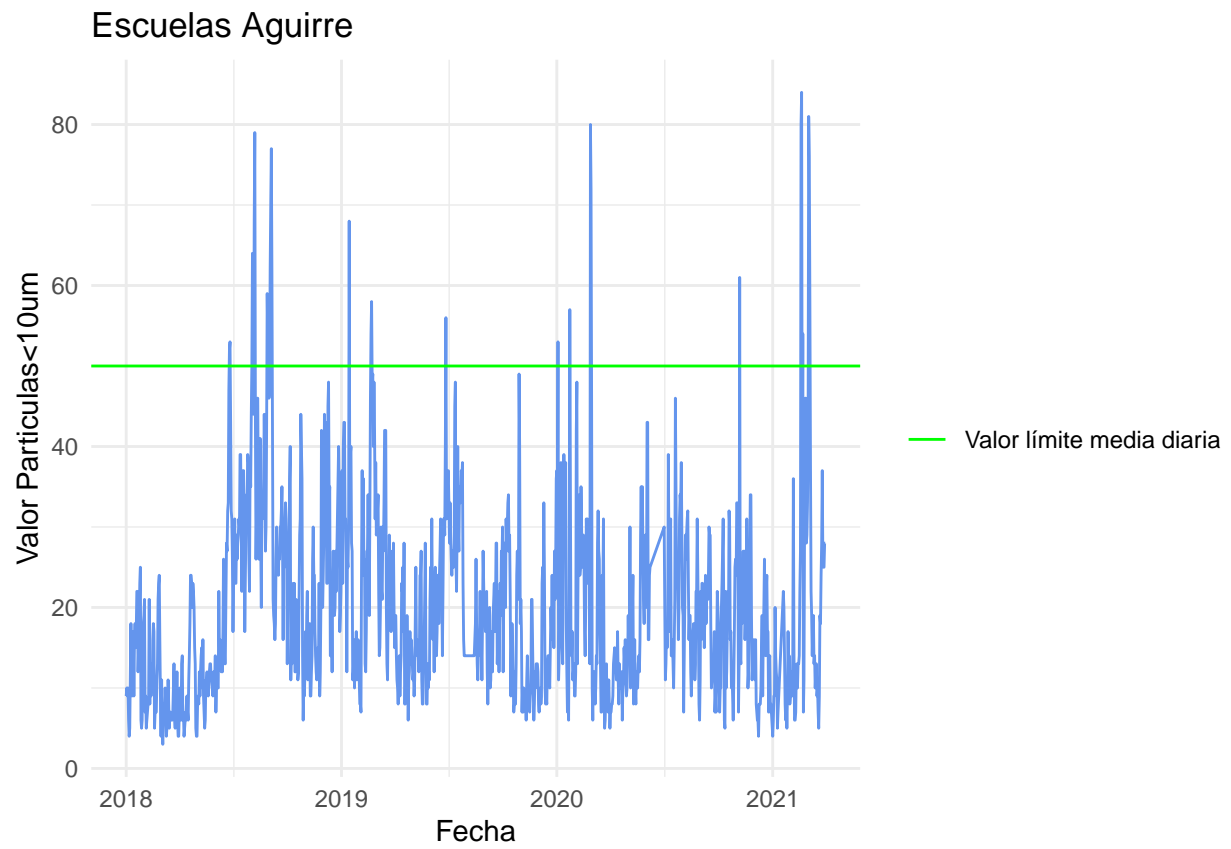


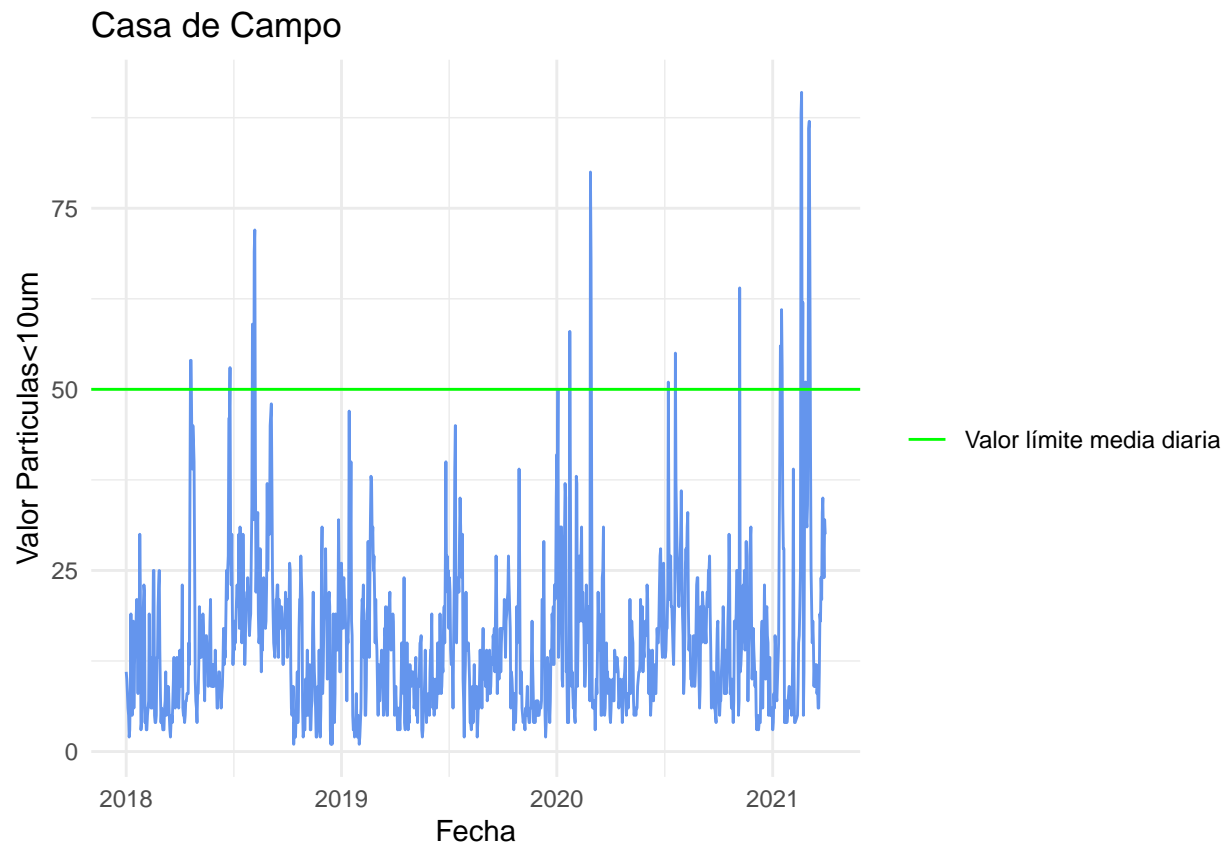


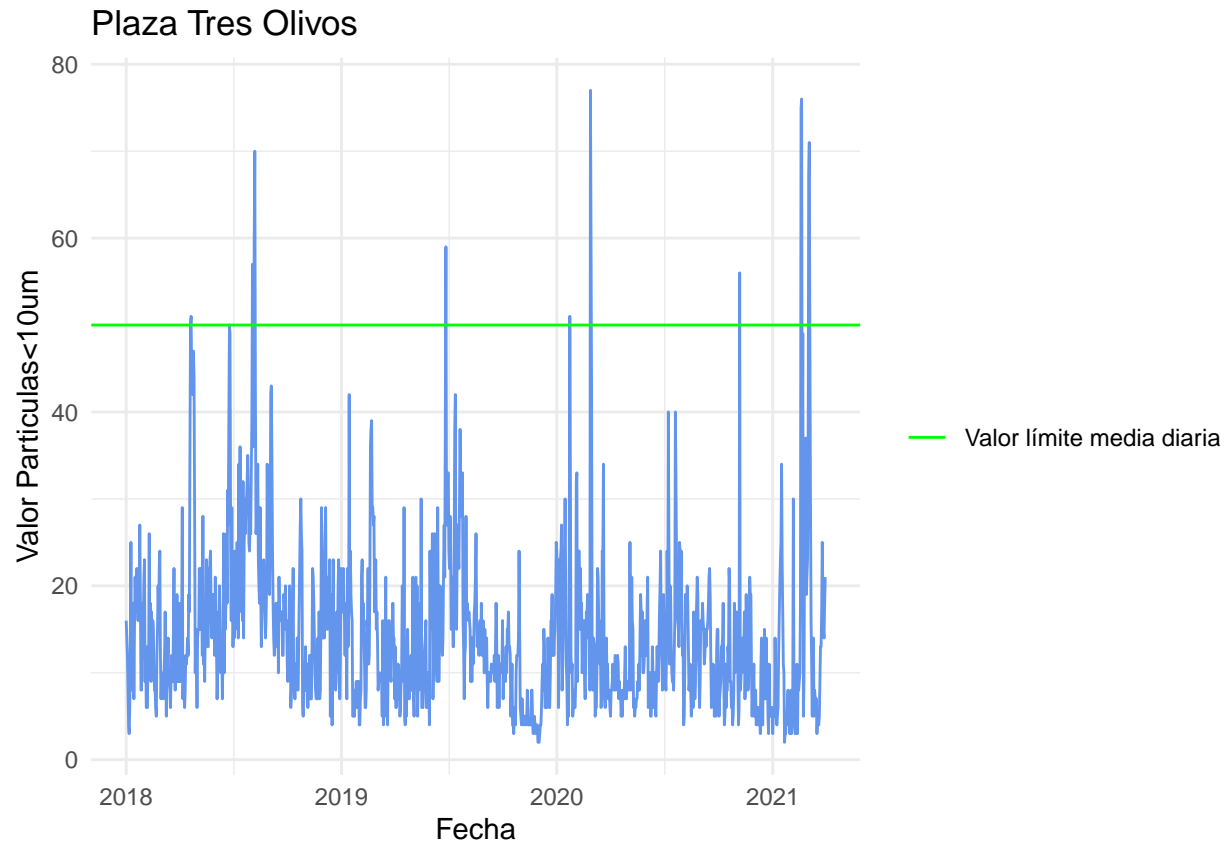
Series temporales Representando la Magnitud 10 - Datos brutos diarios

```
for (punto in puntos){
muestra_mag10 <- df%>% filter(nombre_estacion==punto & magnitud==10)

plot_10 <- ggplot(muestra_mag10, aes(x=fecha, y=valor)) +
  geom_line(color="cornflowerblue") + theme_minimal() + ggtitle(punto) +
  geom_hline(aes(yintercept=50, color="Valor límite media diaria")) +
  scale_colour_manual(values = c("green")) + labs(color="") +
  xlab("Fecha") + ylab("Valor Partículas<10um")
print(plot_10)
}
```







Series temporales Representando la Magnitud 10 - Media Anual

```
#library("tsbox")

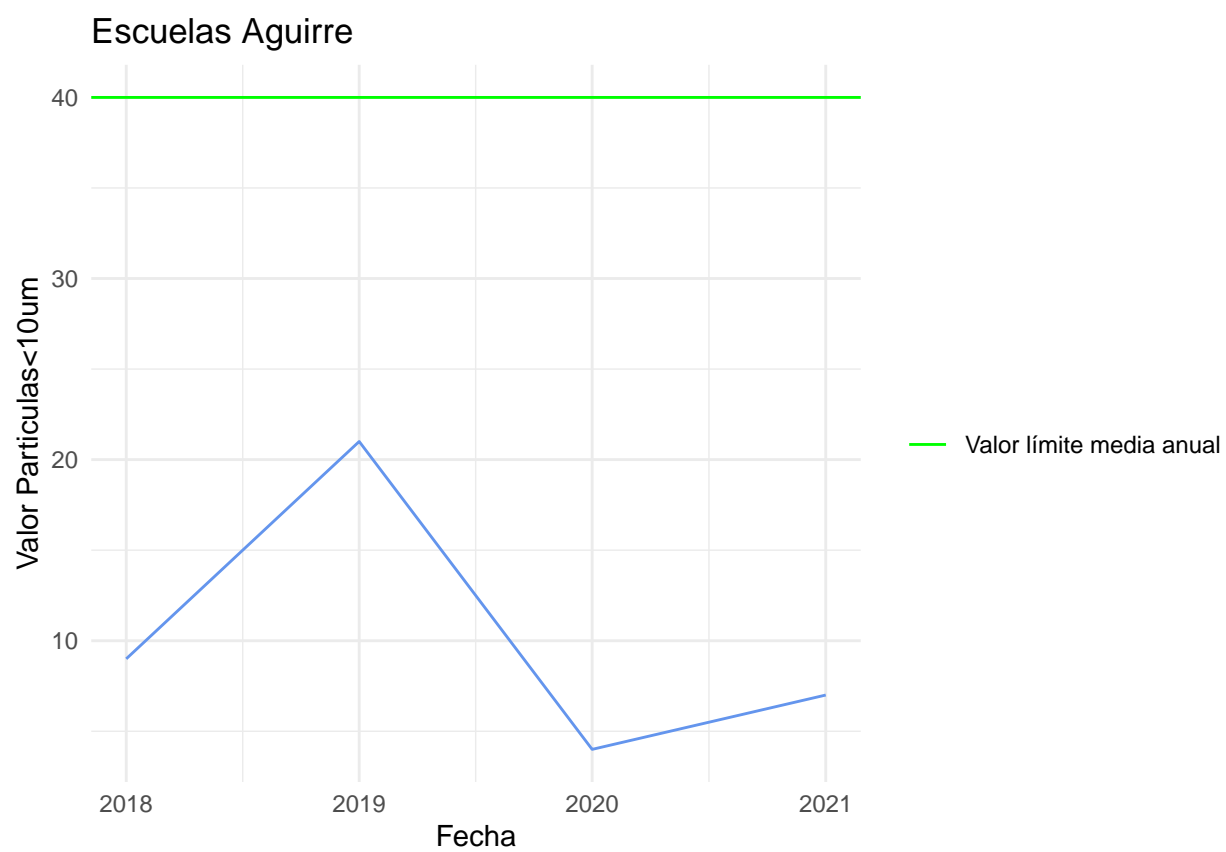
for (punto in puntos){
muestra_mag10 <- df%>% filter(nombre_estacion==punto & magnitud==10)

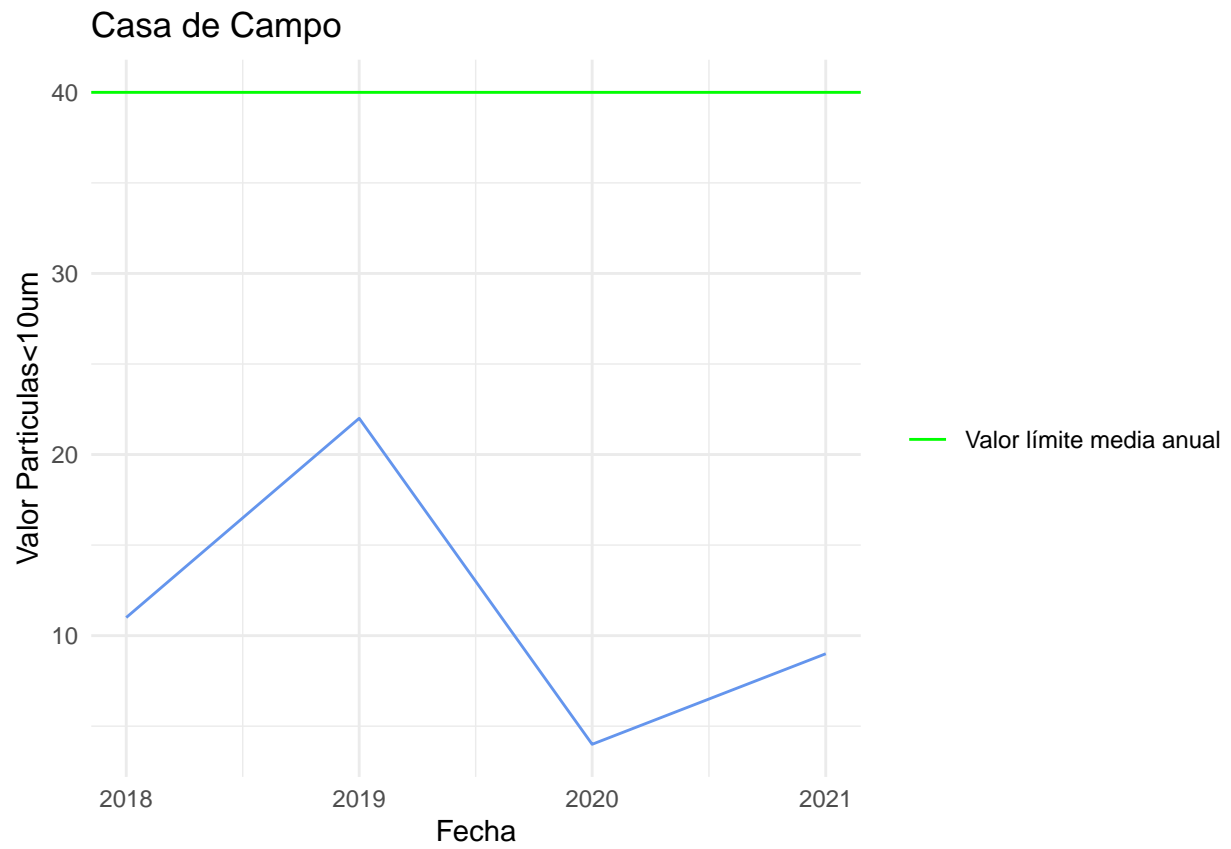
serie_muestra_mag10 <- ts(start= c(2018,1), end=c(2021,1),muestra_mag10$valor, frequency = 1)

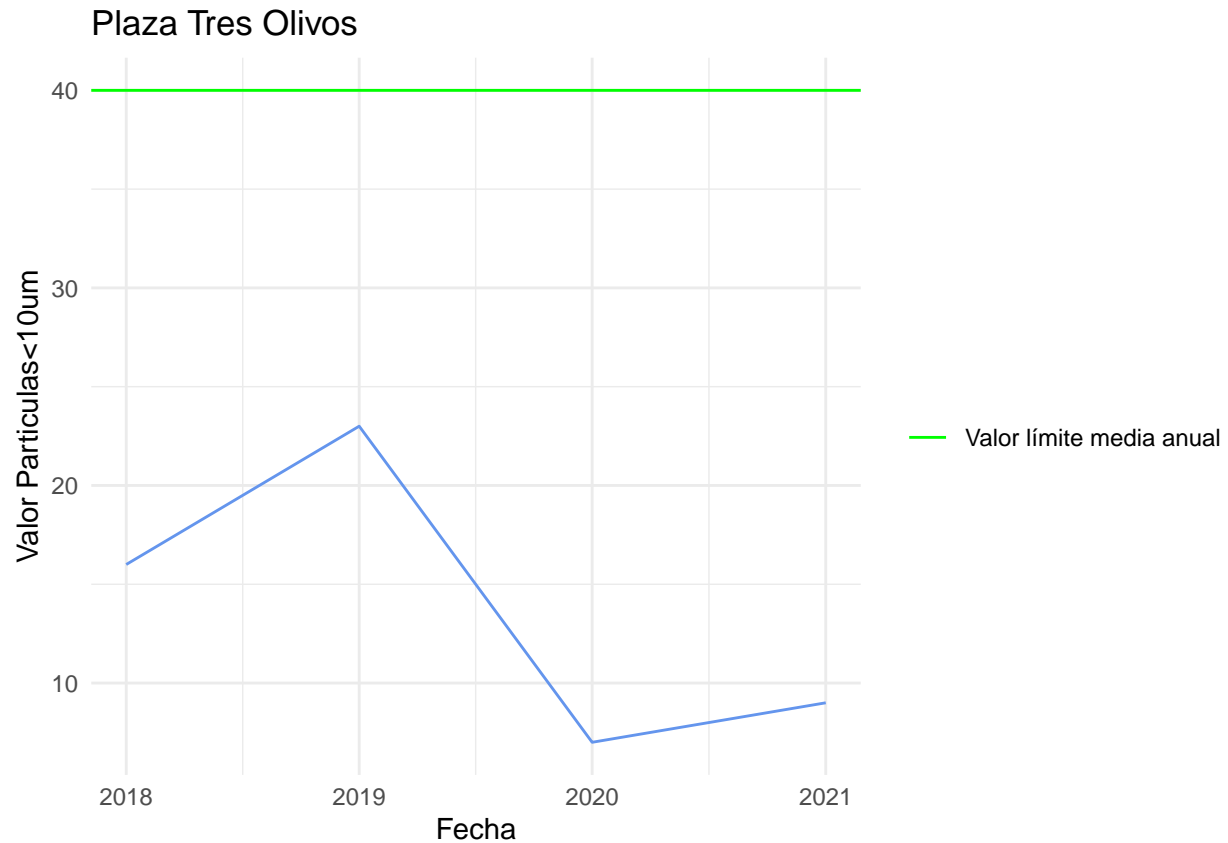
# Convierte de serie temporal a dataframe
mm10 <- ts_df(serie_muestra_mag10)

plot_10 <- ggplot(mm10, aes(x=time, y=value)) +
  geom_line(color="cornflowerblue")+ theme_minimal()+ ggtitle(punto)+
  geom_hline(aes(yintercept=40, color="Valor límite media anual"))+
  scale_colour_manual(values = c("green"))+labs(color="")+
  xlab("Fecha")+ylab("Valor Particulas<10um")
print(plot_10)
```

}



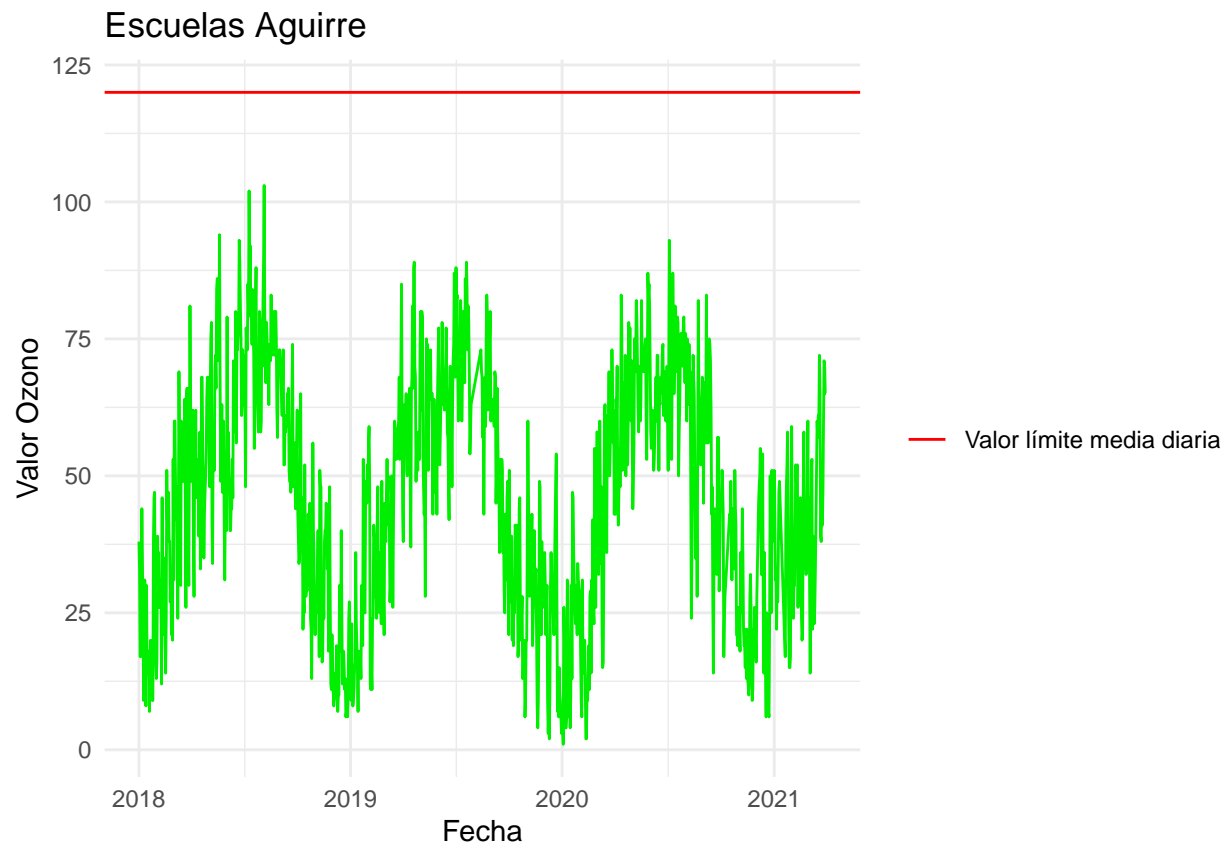


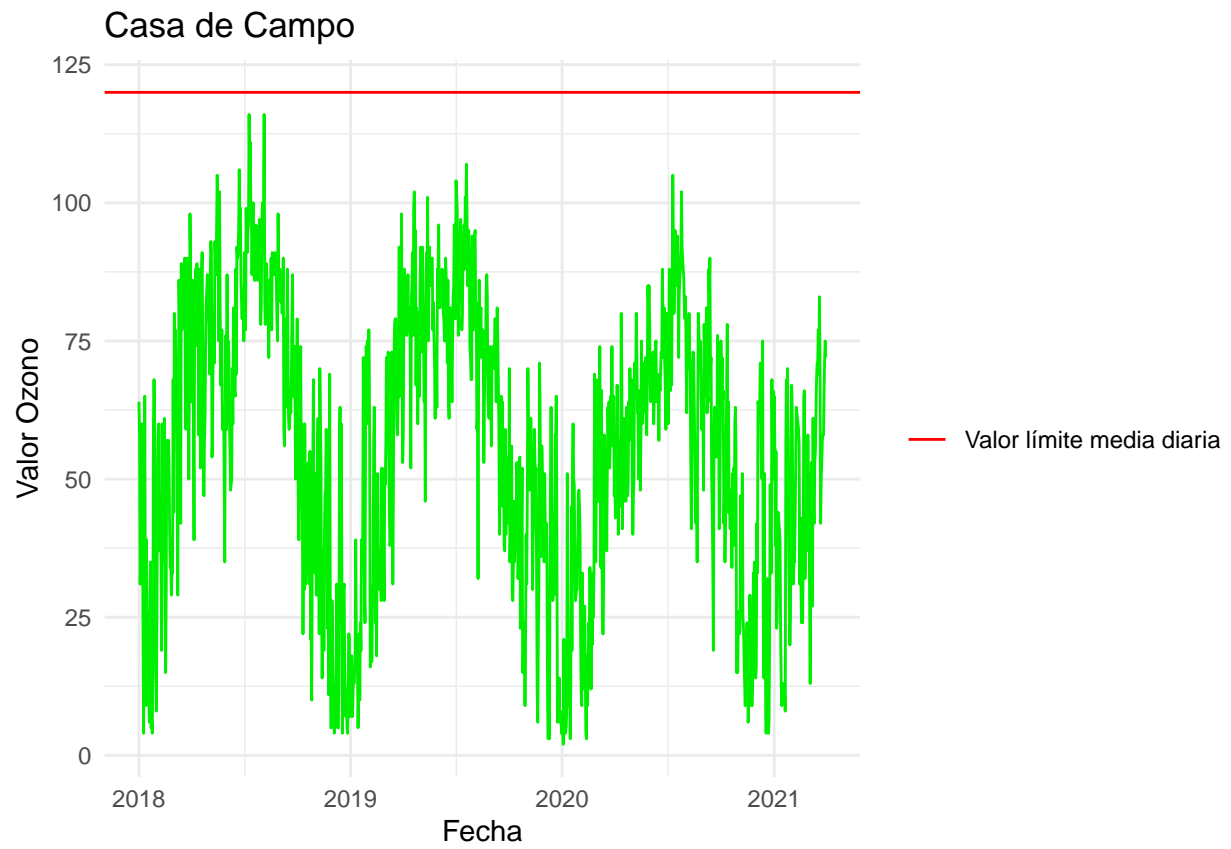


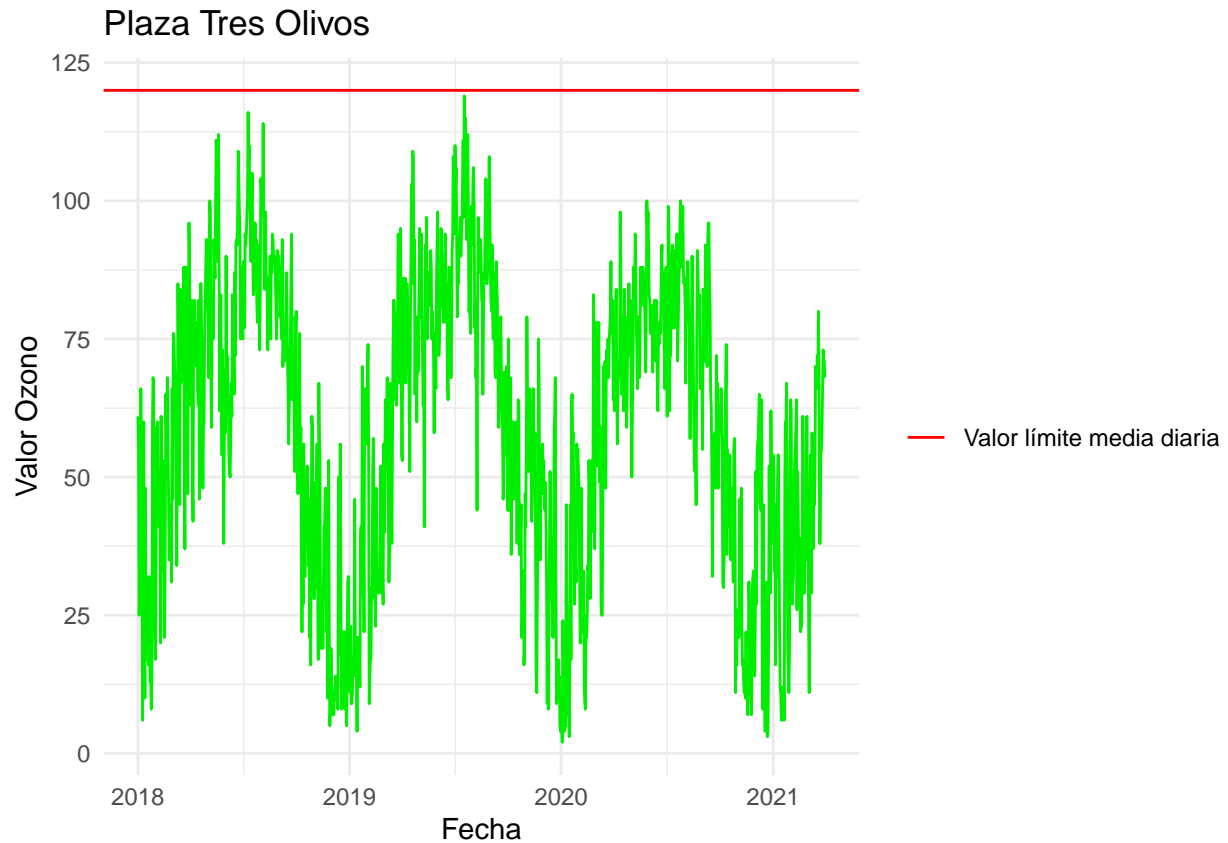
Series temporales Representando la Magnitud 14 - Datos brutos diarios

```
for (punto in puntos){
muestra_mag14 <- df%>% filter(nombre_estacion==punto & magnitud==14)

plot_14 <- ggplot(muestra_mag14, aes(x=fecha, y=valor)) +
  geom_line(color="green2")+ theme_minimal()+ ggtitle(punto)+
  geom_hline(aes(yintercept=120, color="Valor límite media diaria"))+
  scale_colour_manual(values = c("red"))+labs(color="")+
  xlab("Fecha")+ylab("Valor Ozono")
print(plot_14)
}
```



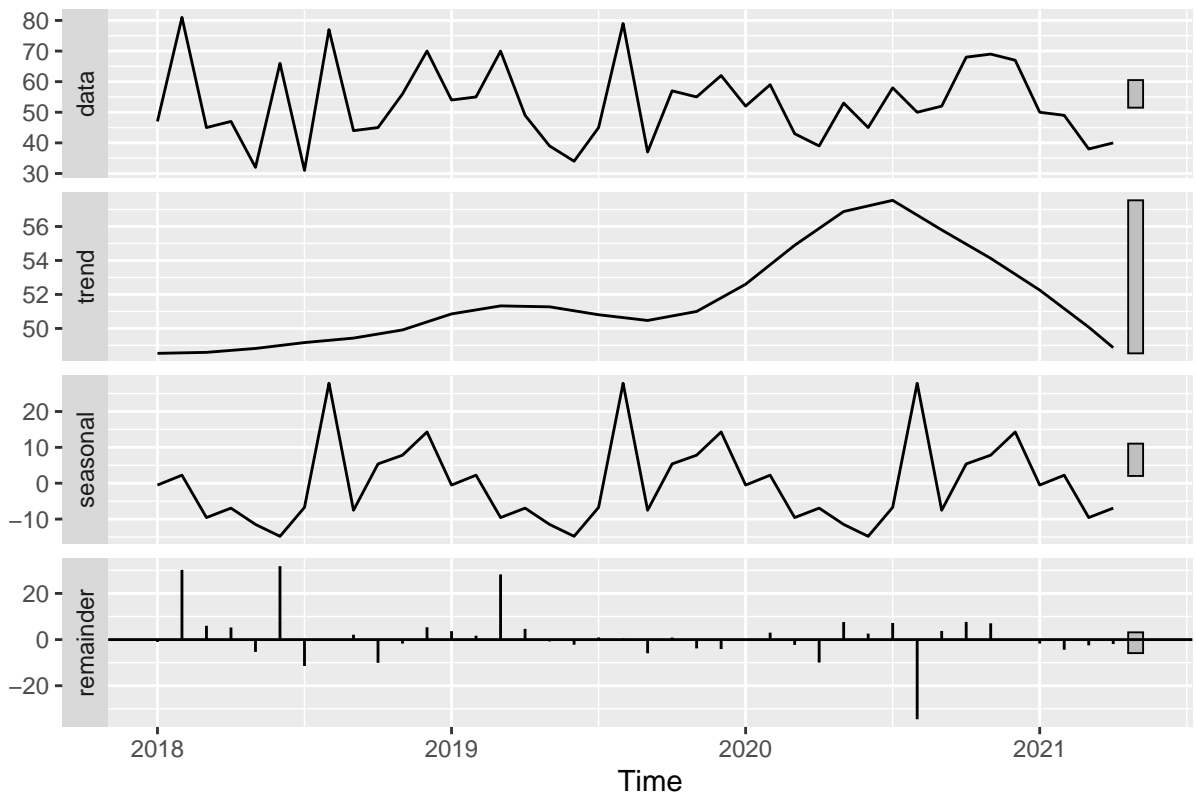




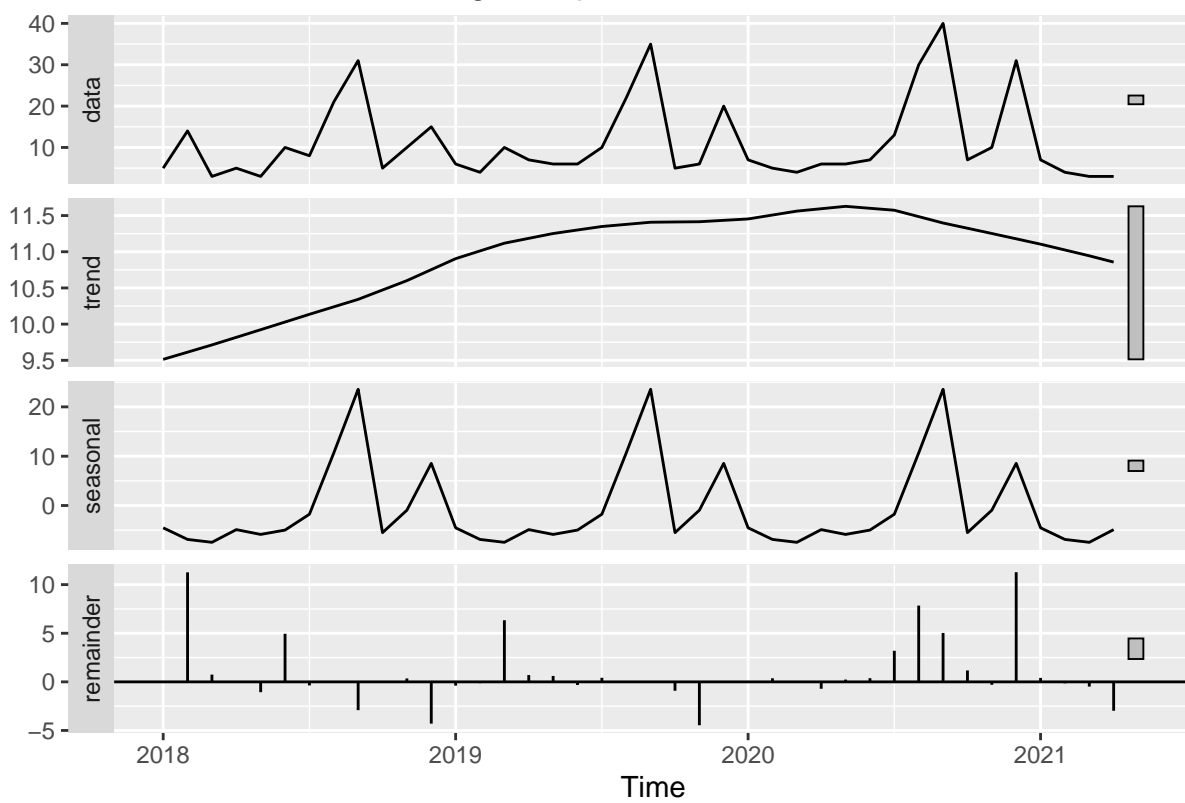
Descomposición de serie temporal mensual para la estación 8, Escuelas Aguirre.

Esta estación se encuentra ubicada entre Calle Alcalá y C/O'Donnell. Se trata de una estación de tipo “Urbana tráfico”. Hemos representado sus respectivos parámetros que son Dióxido de nitrógeno (8), Partículas <2.5um (9), Partículas <10um (10), Ozono (14).

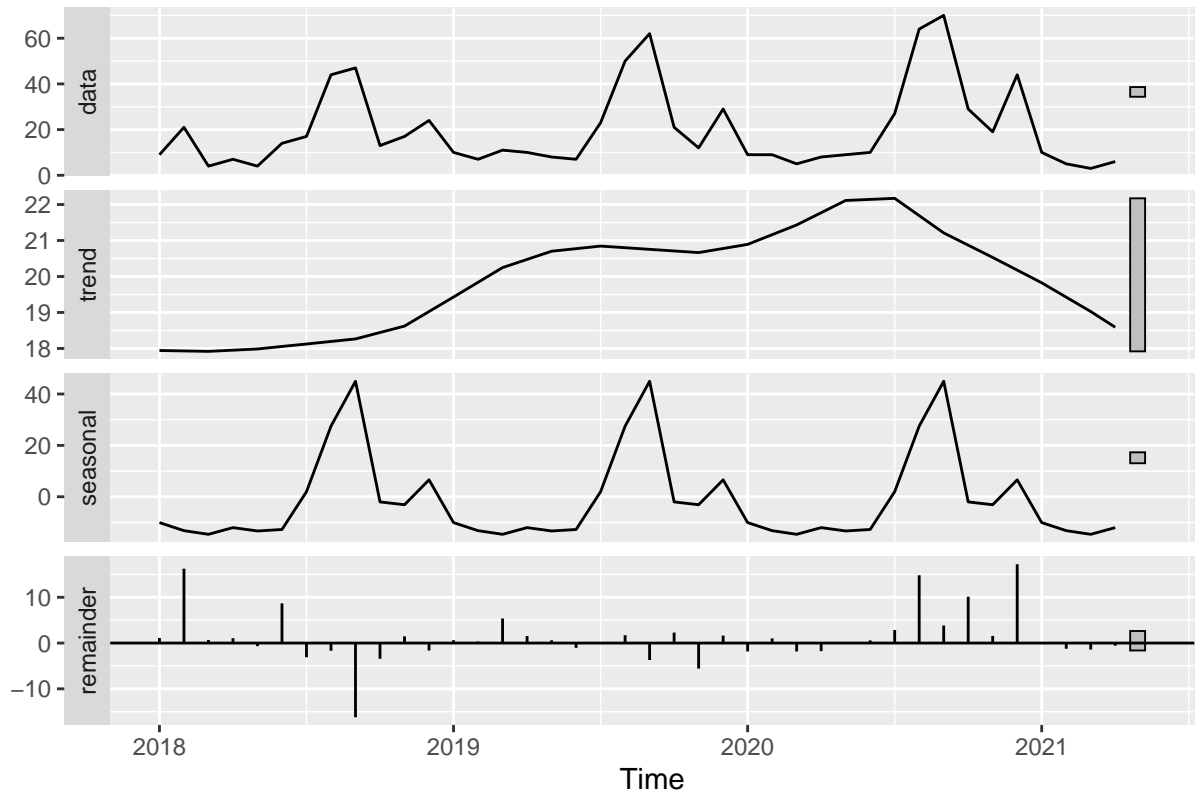
Estación 8, Escuelas Aguirre, parámetro Dióxido de Nitrógeno

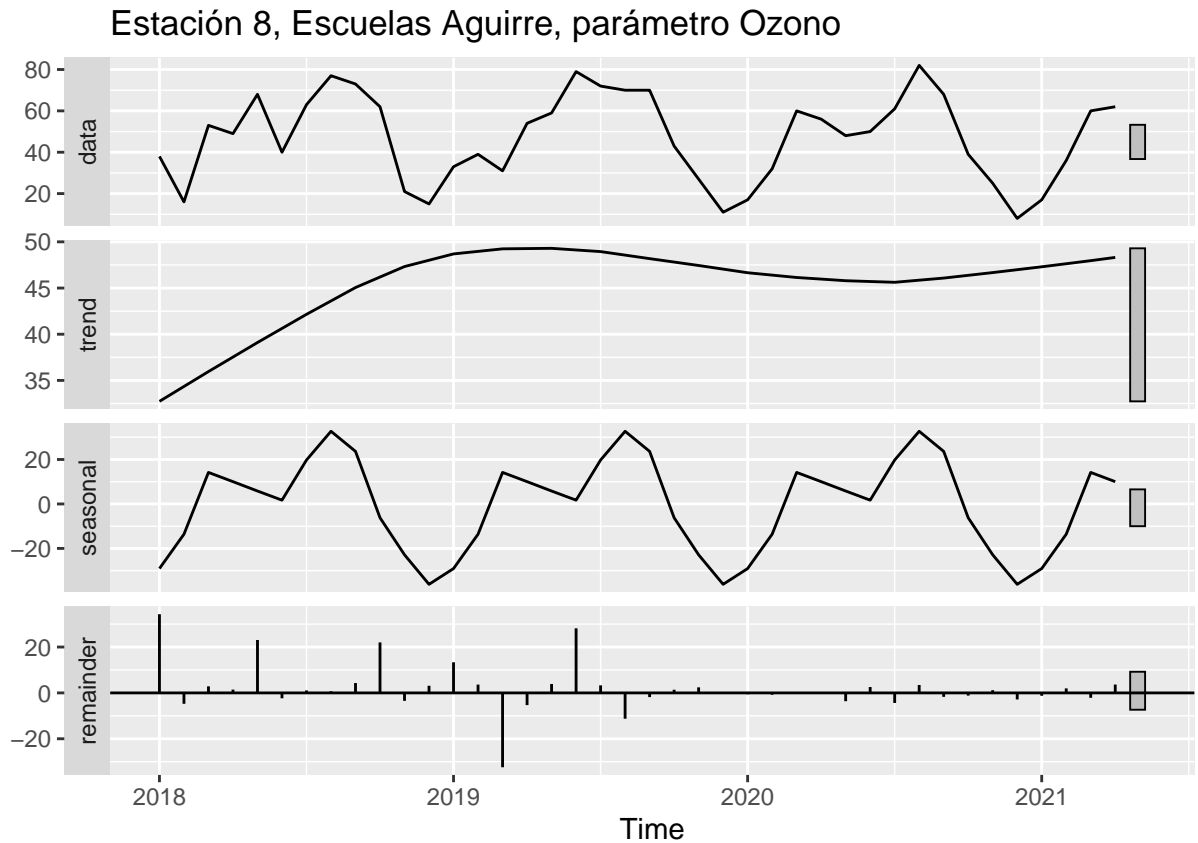


Estación 8, Escuelas Aguirre, parámetro Partículas <2.5um



Estación 8, Escuelas Aguirre, parámetro Partículas <10um

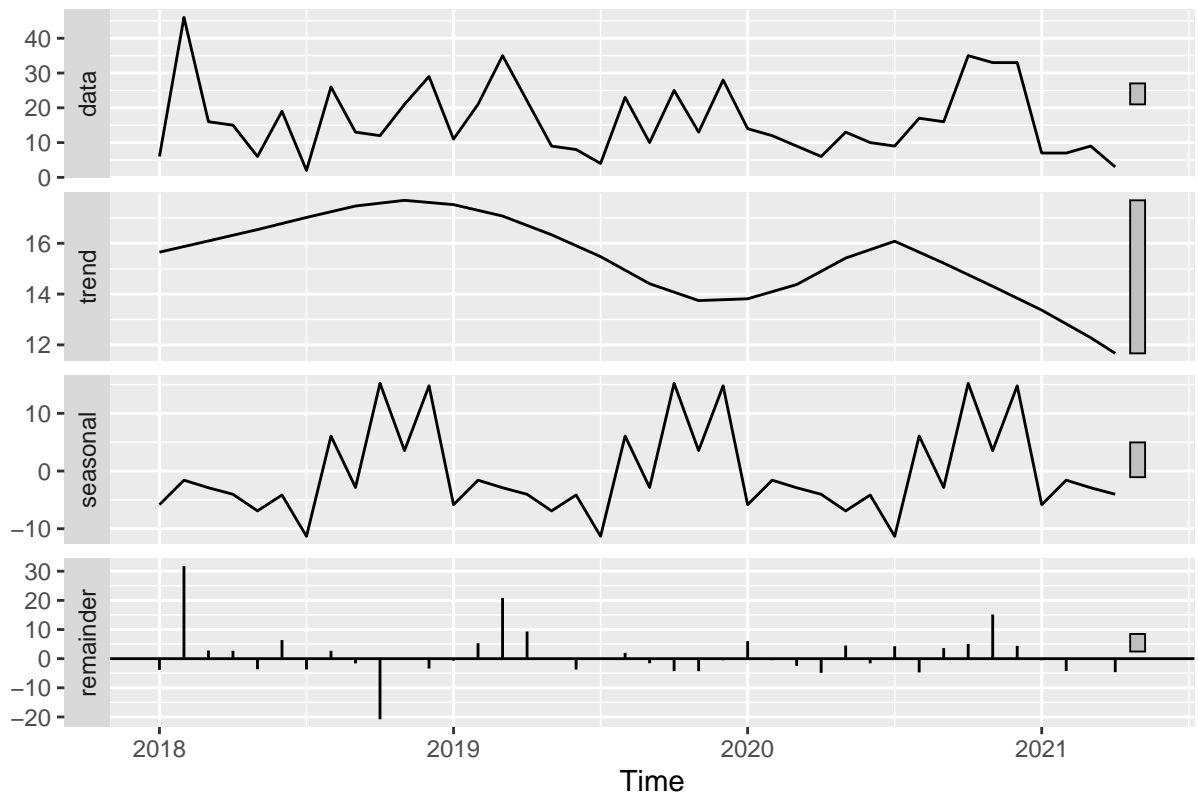




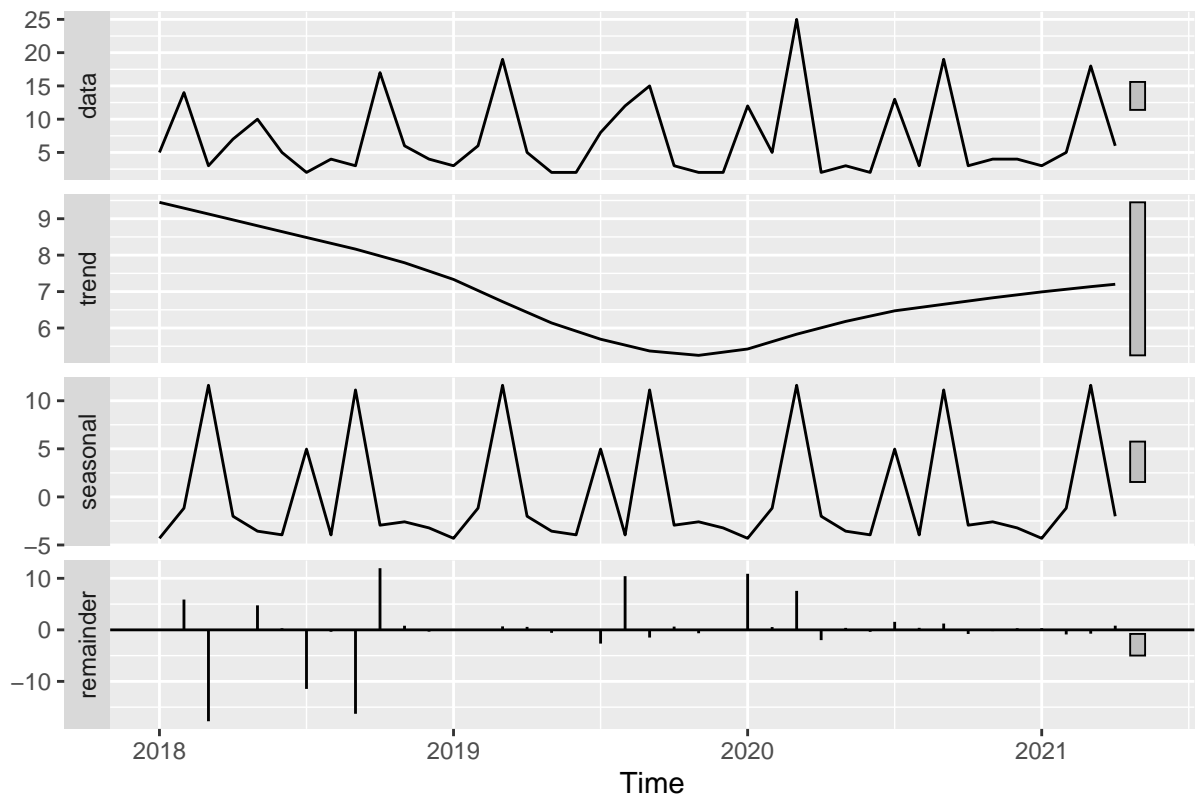
Descomposición de serie temporal mensual para la estación 24, Casa Campo.

Se trata de una estación de tipo “Suburbana”. Hemos representado sus respectivos parámetros que son Dióxido de nitrógeno (8), Partículas <2.5um (9), Partículas <10um (10), Ozono (14).

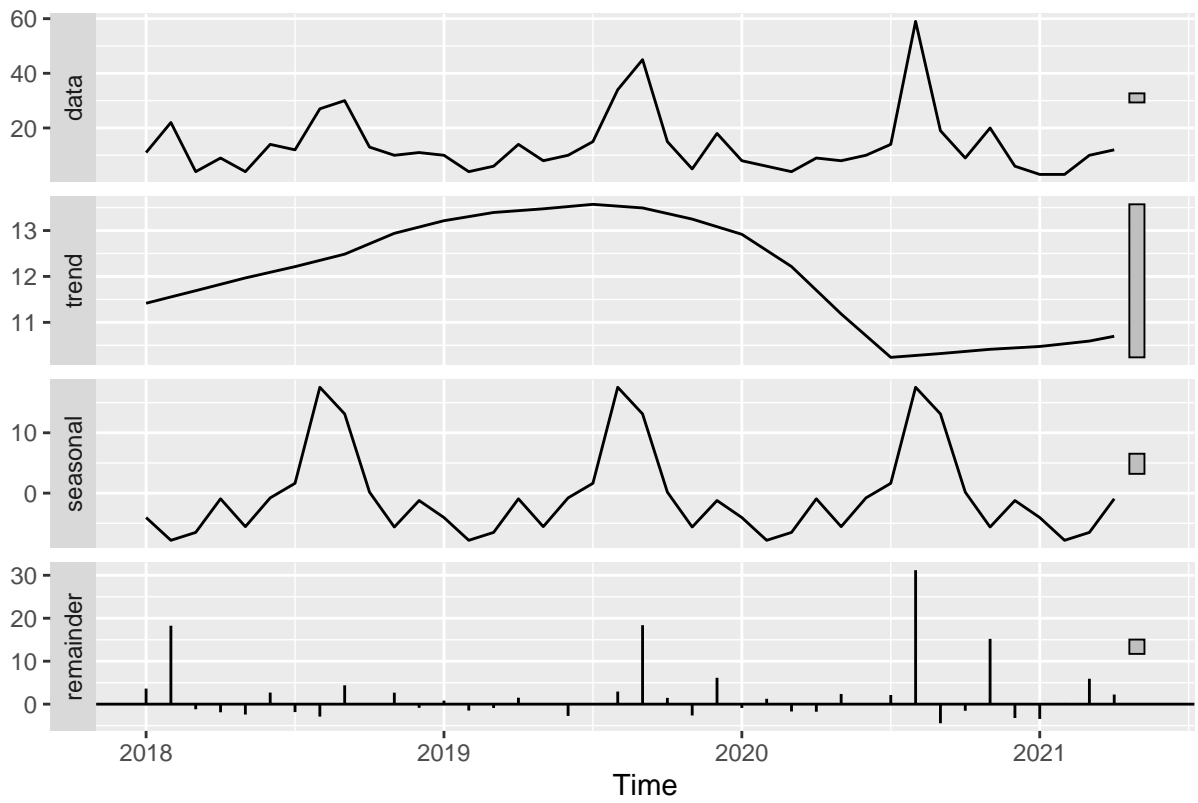
Estación 24, Casa de Campo, parámetro Dióxido de Nitrógeno

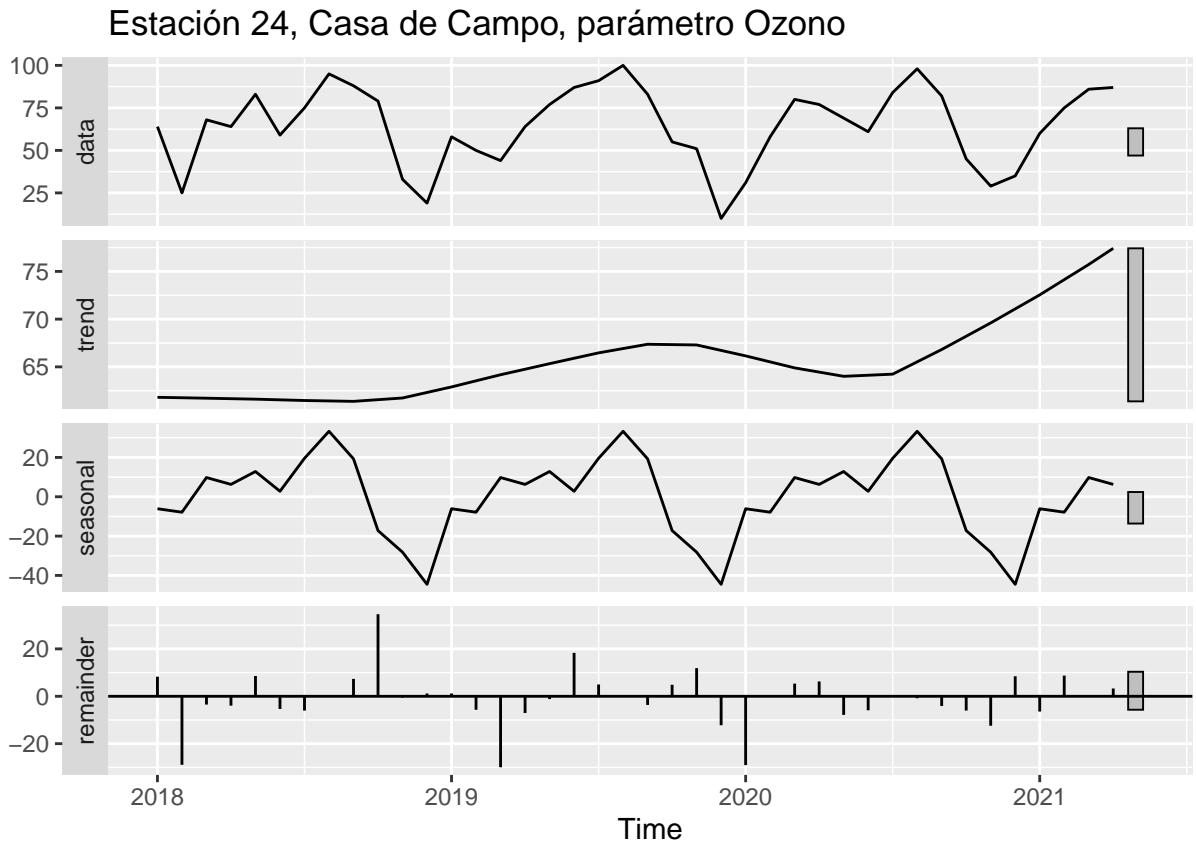


Estación 24, Casa de Campo, parámetro Partículas <2.5um



Estación 24, Casa de Campo, parámetro Partículas <10um

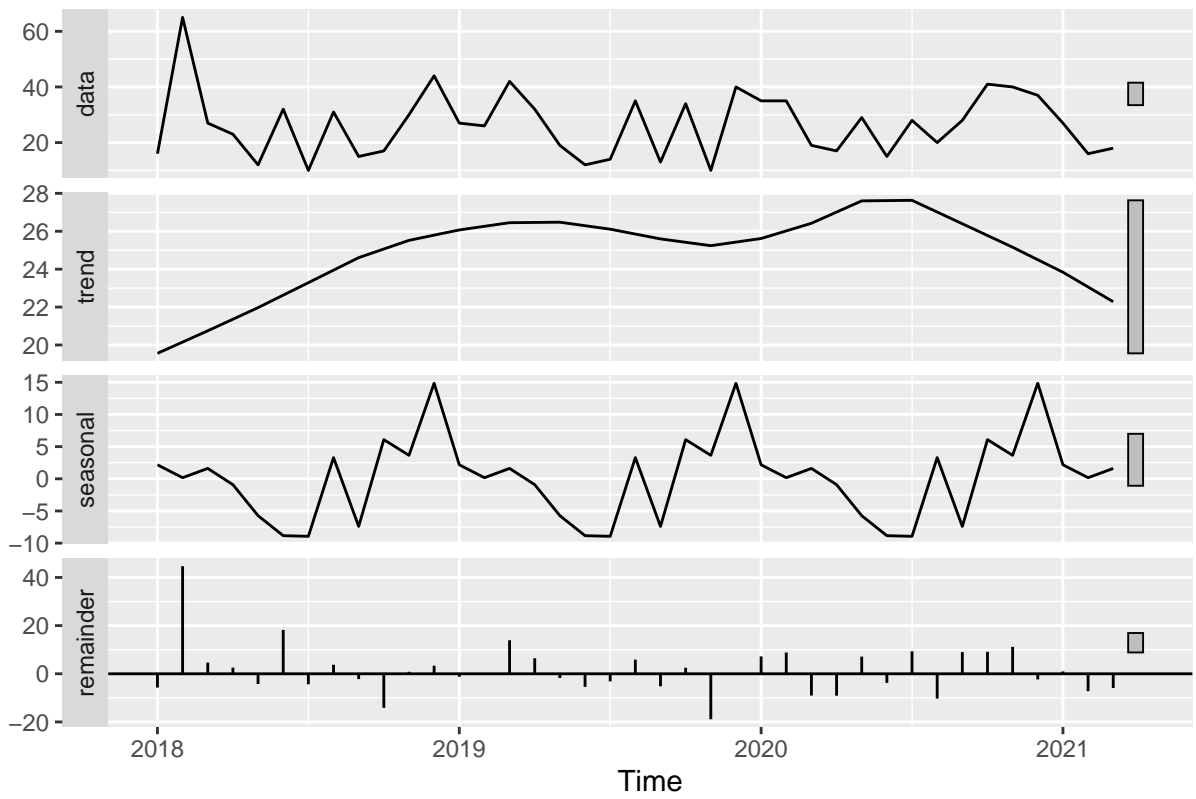




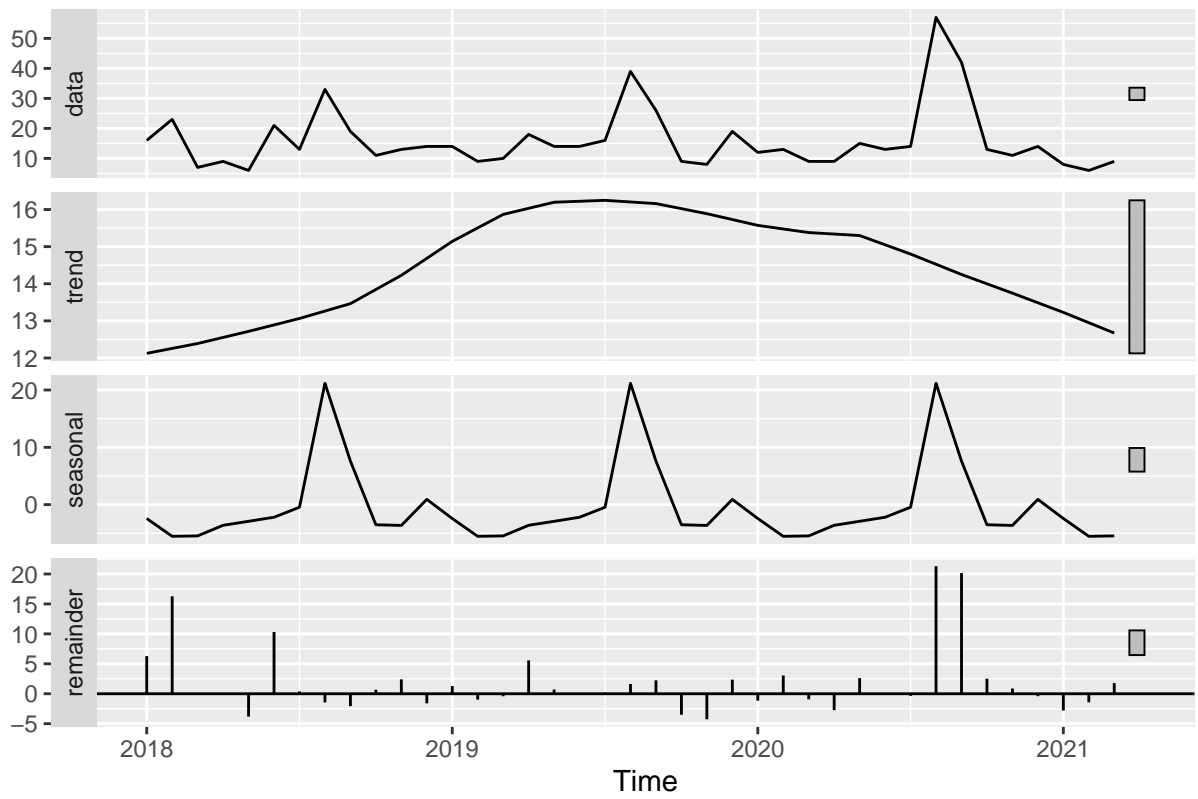
Descomposición de serie temporal mensual para la estación 60, Plaza Tres Olivos.

Se trata de una estación de tipo “Urbana Fondo”. Hemos representado sus respectivos parámetros que son Dióxido de nitrógeno (8), Partículas <10um (10), Ozono (14).

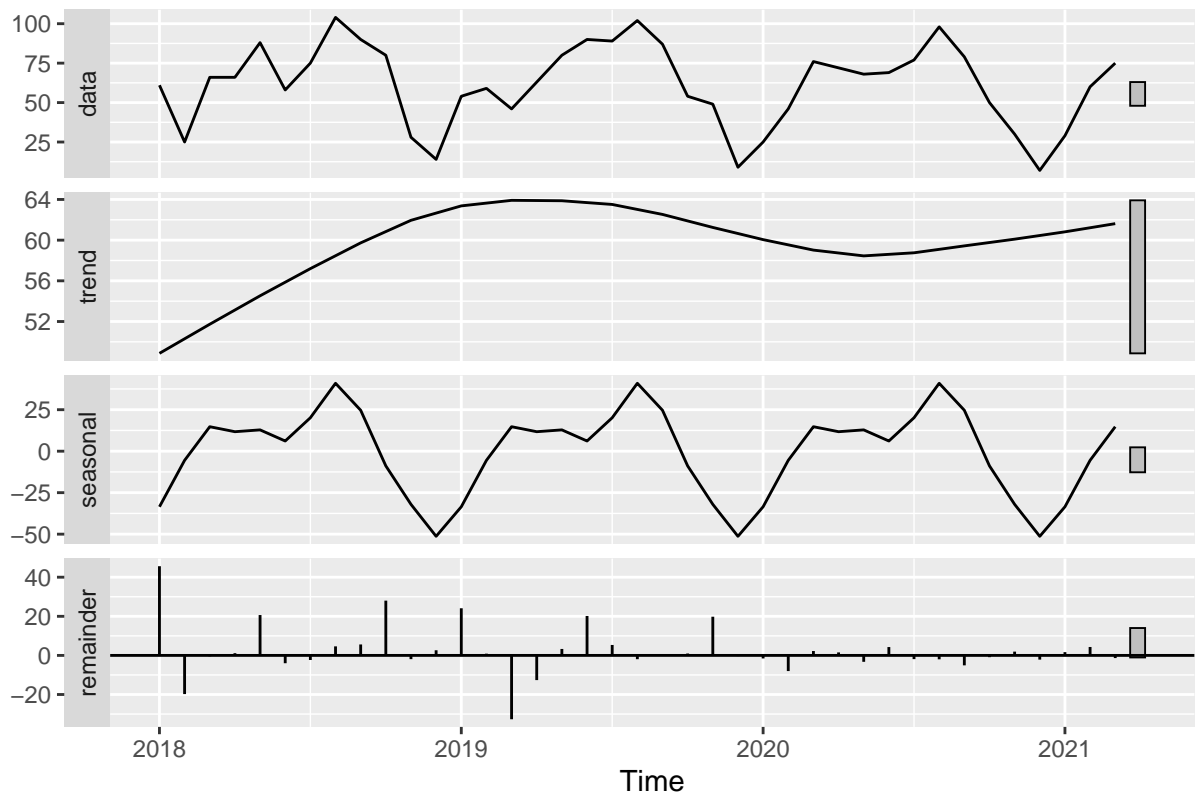
Estación 60, Plaza Tres Olivos, parámetro Dióxido de Nitrógeno



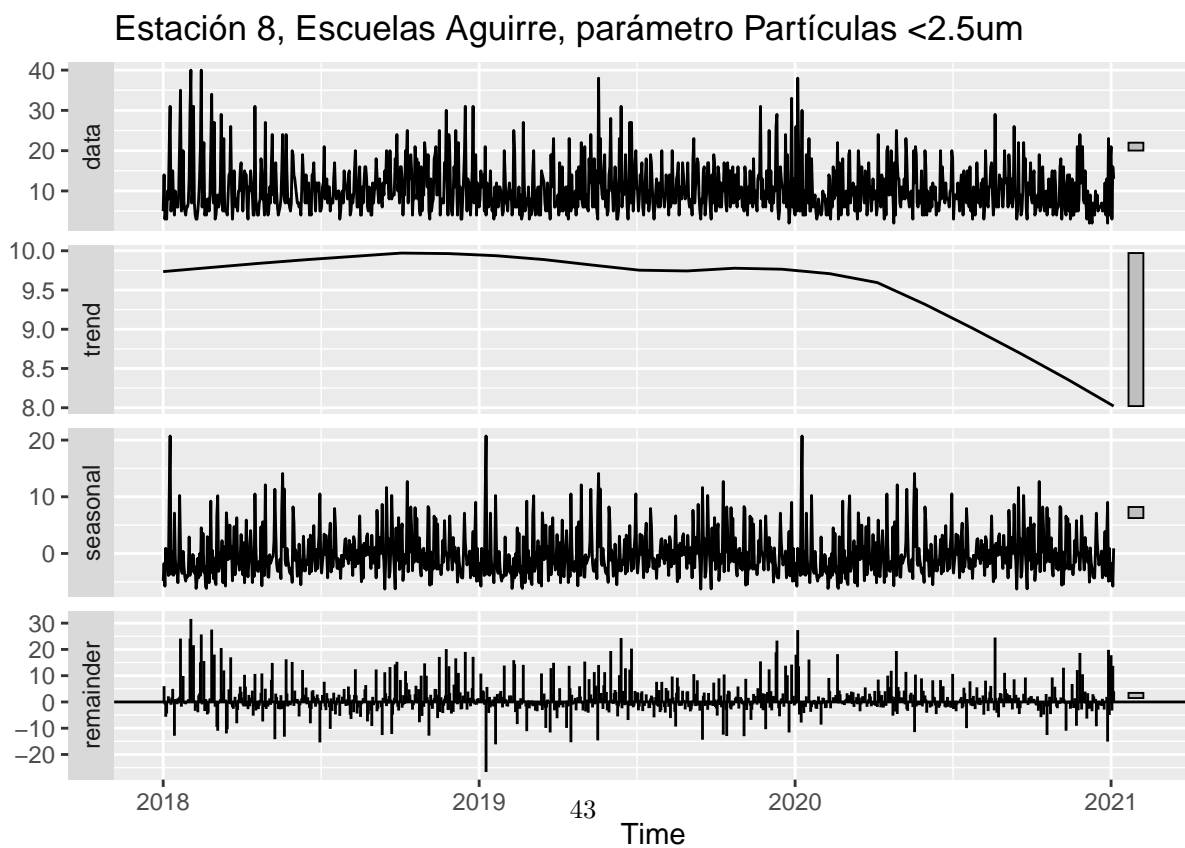
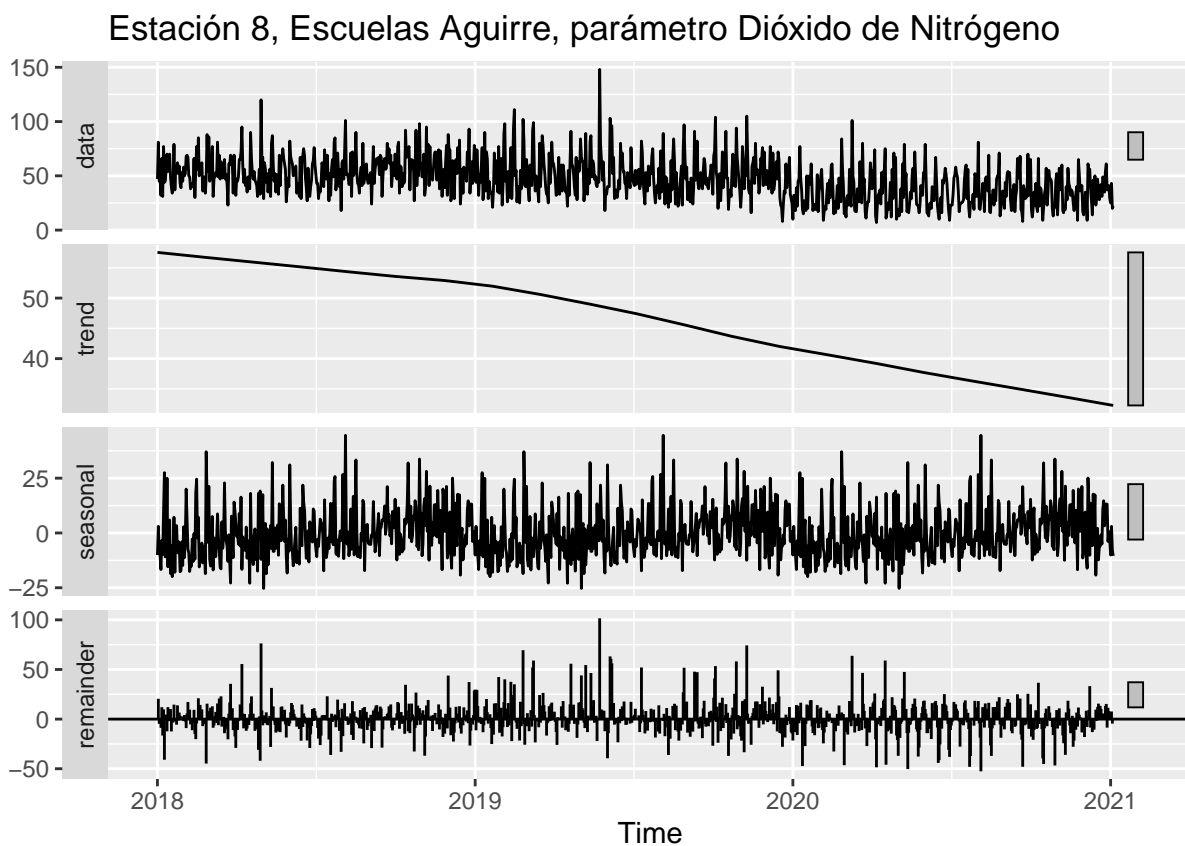
Estación 60, Plaza Tres Olivos, parámetro Partículas <10um



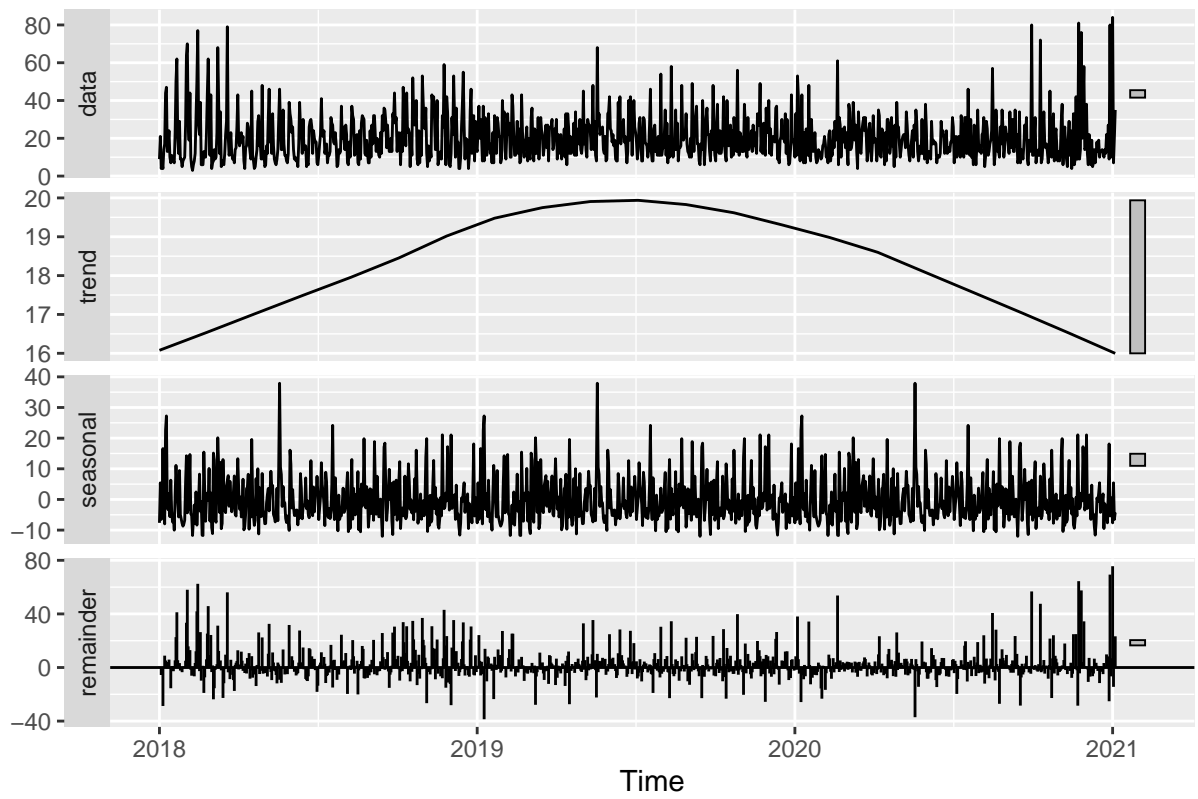
Estación 60, Plaza Tres Olivos, parámetro Ozono



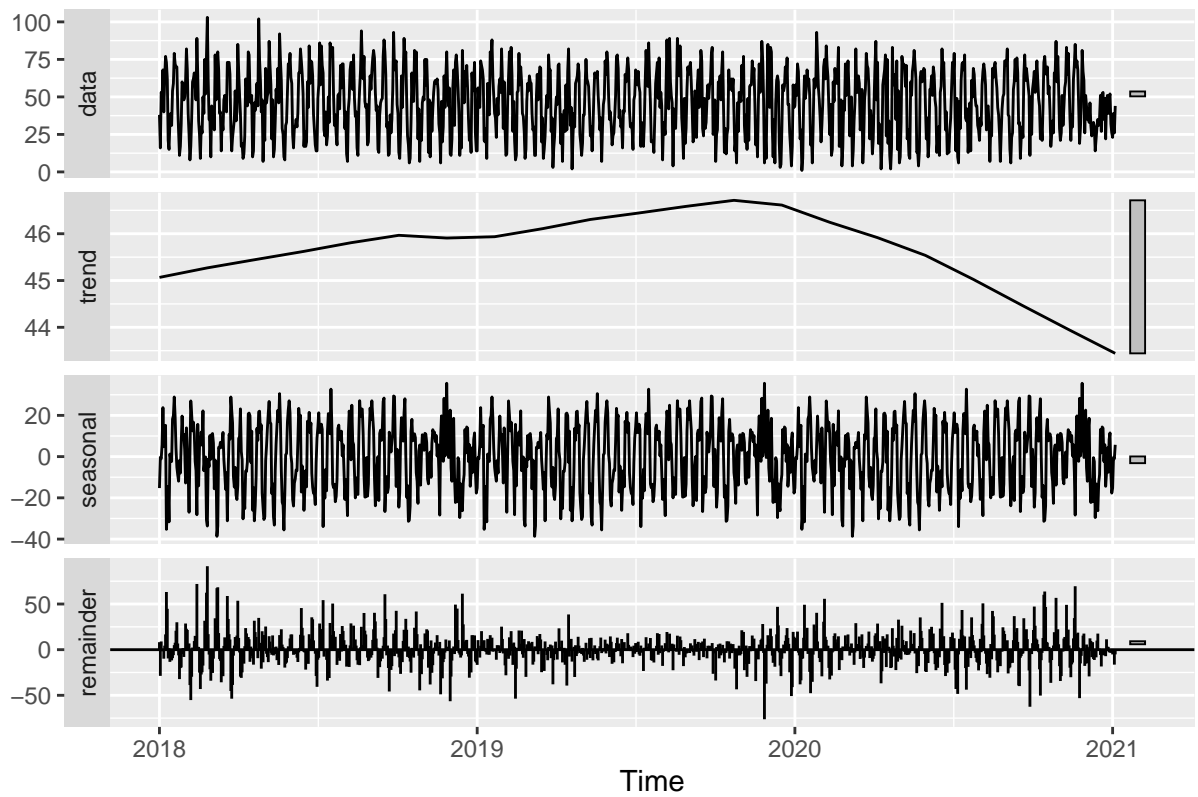
Descomposición de serie temporal diaria para la estación 8, Escuelas Aguirre.



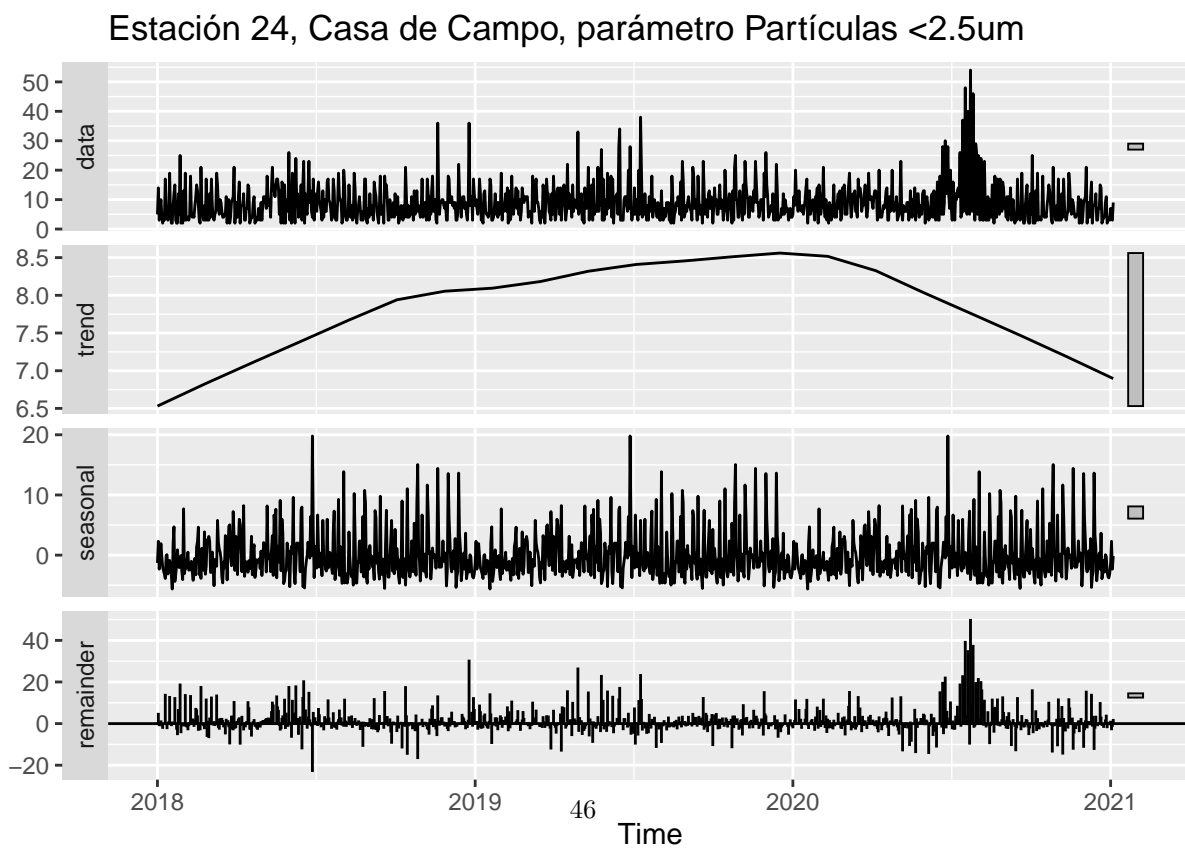
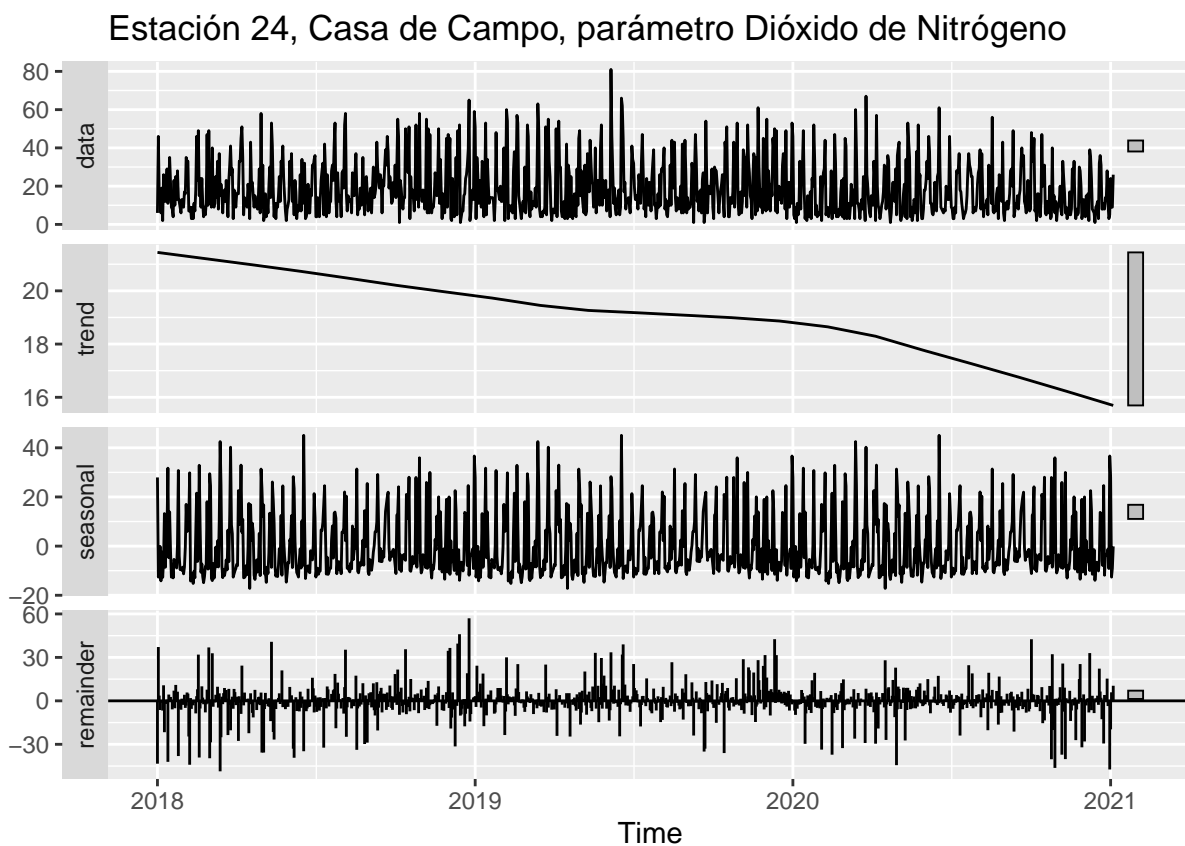
Estación 8, Escuelas Aguirre, parámetro Partículas <10um



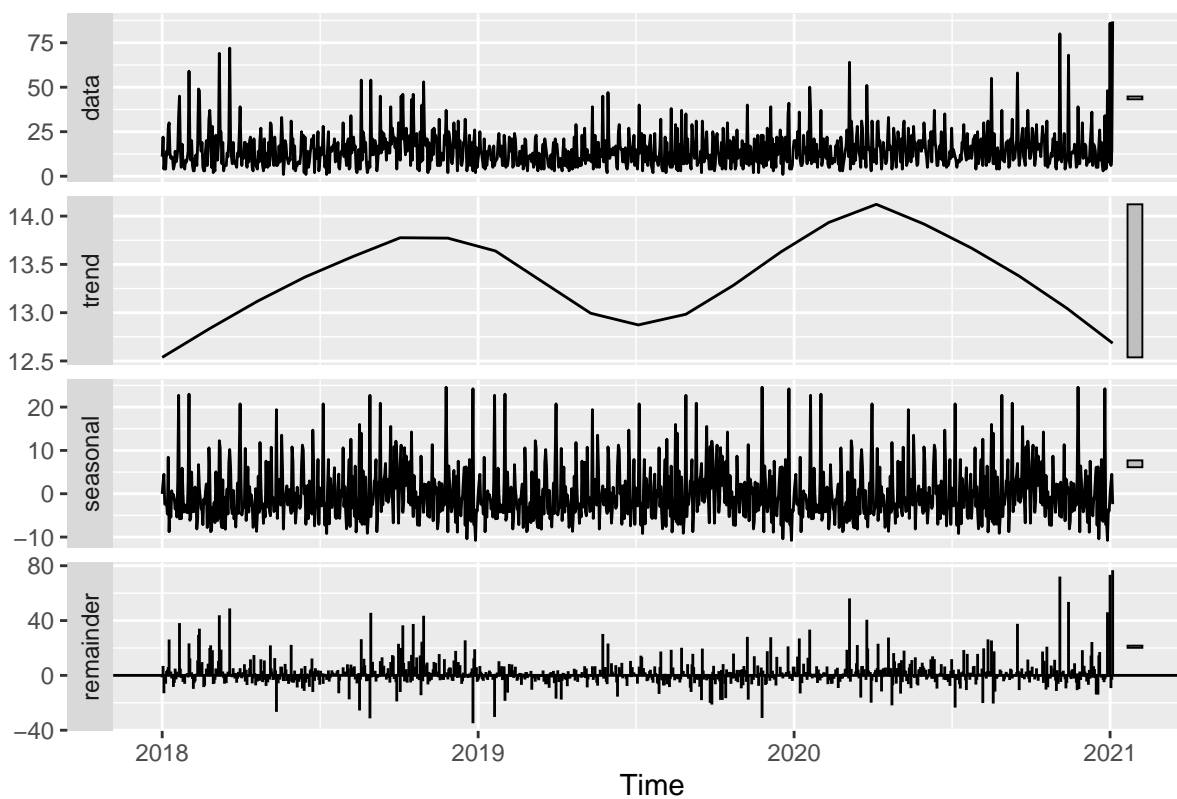
Estación 8, Escuelas Aguirre, parámetro Ozono



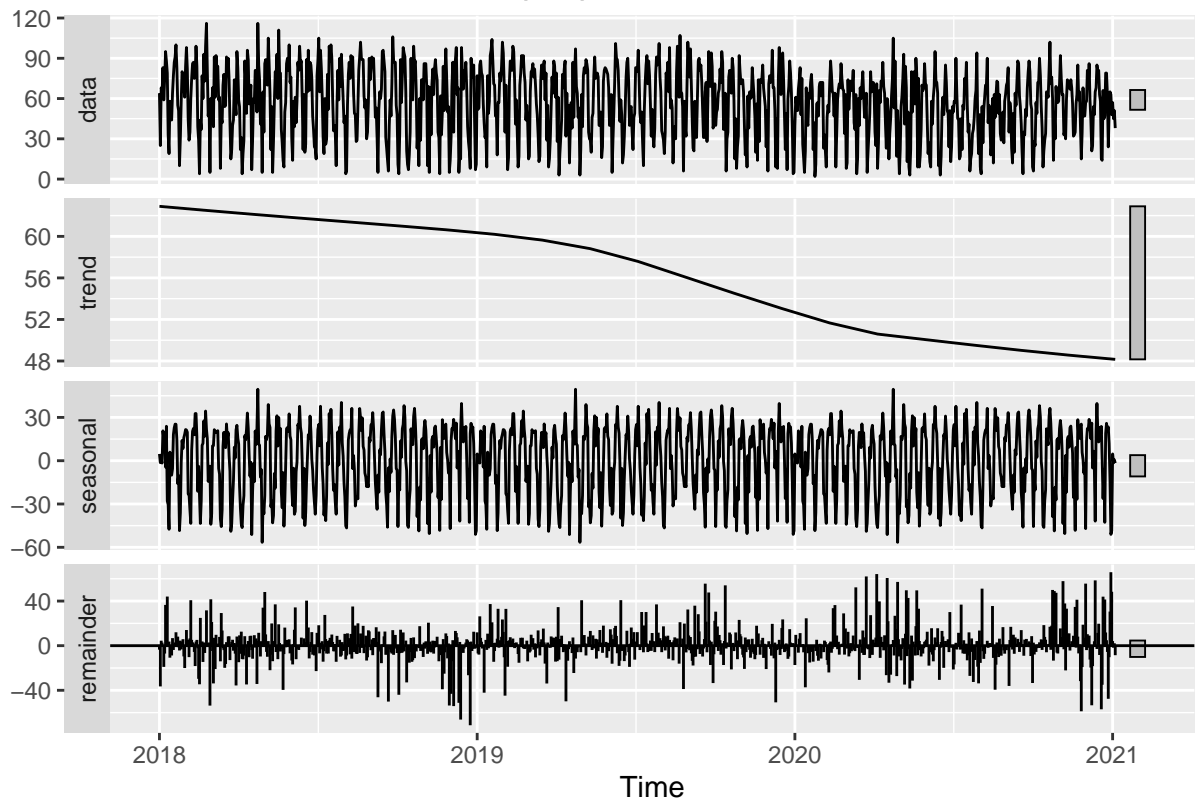
Descomposición de serie temporal diaria para la estación 24, Casa Campo.



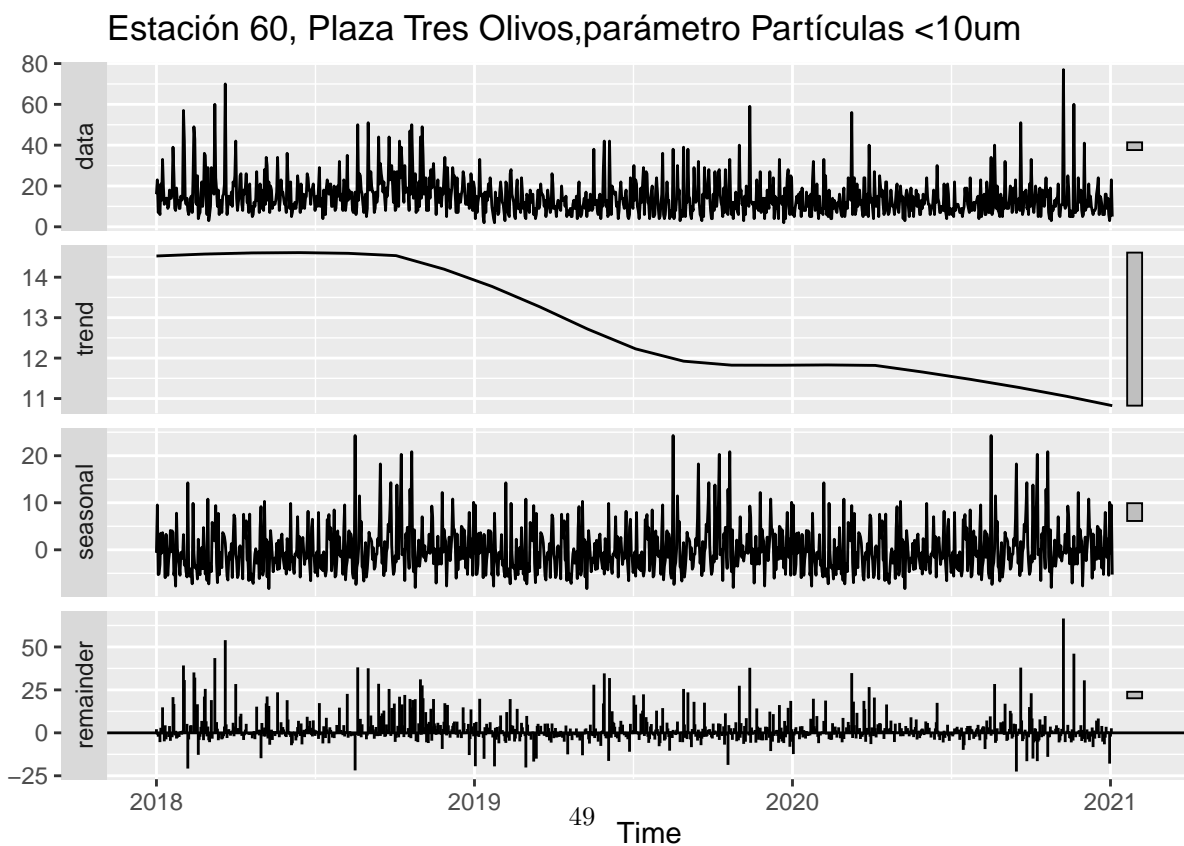
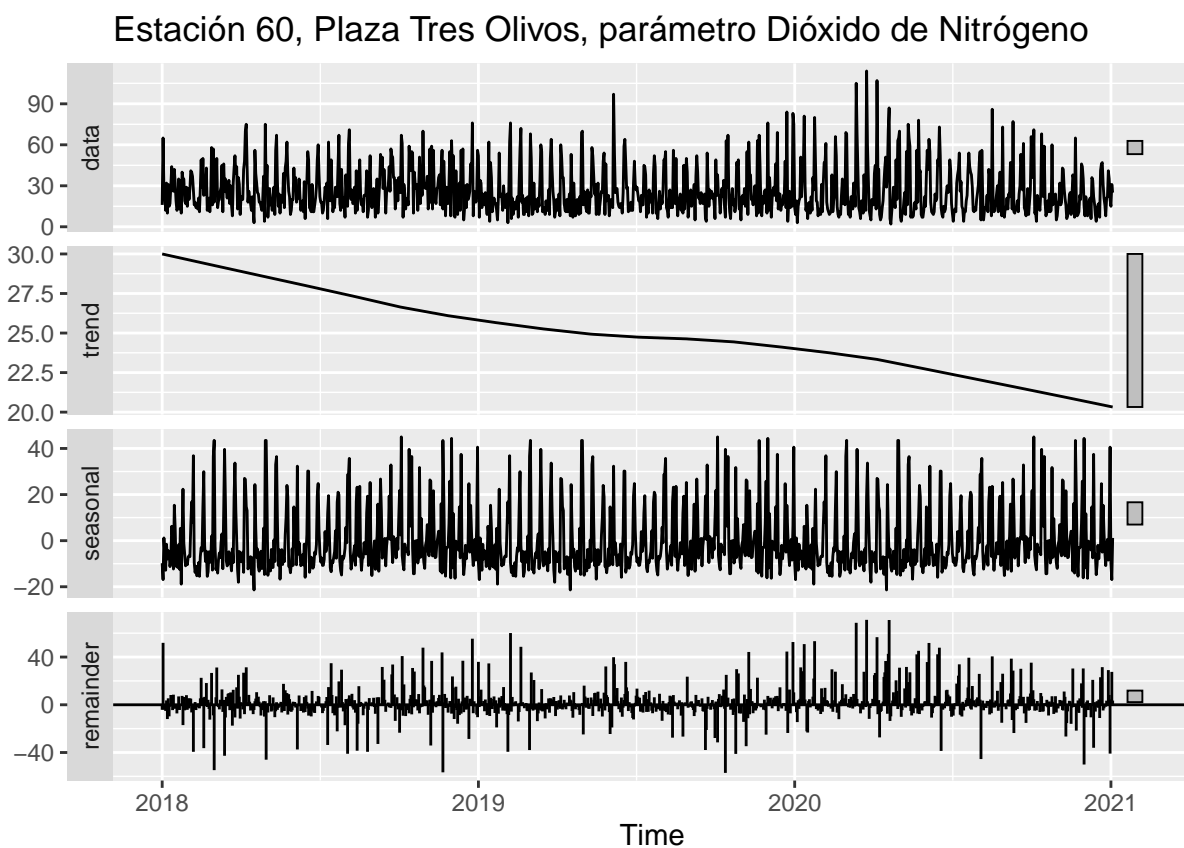
Estación 24, Casa de Campo, parámetro Partículas <10um



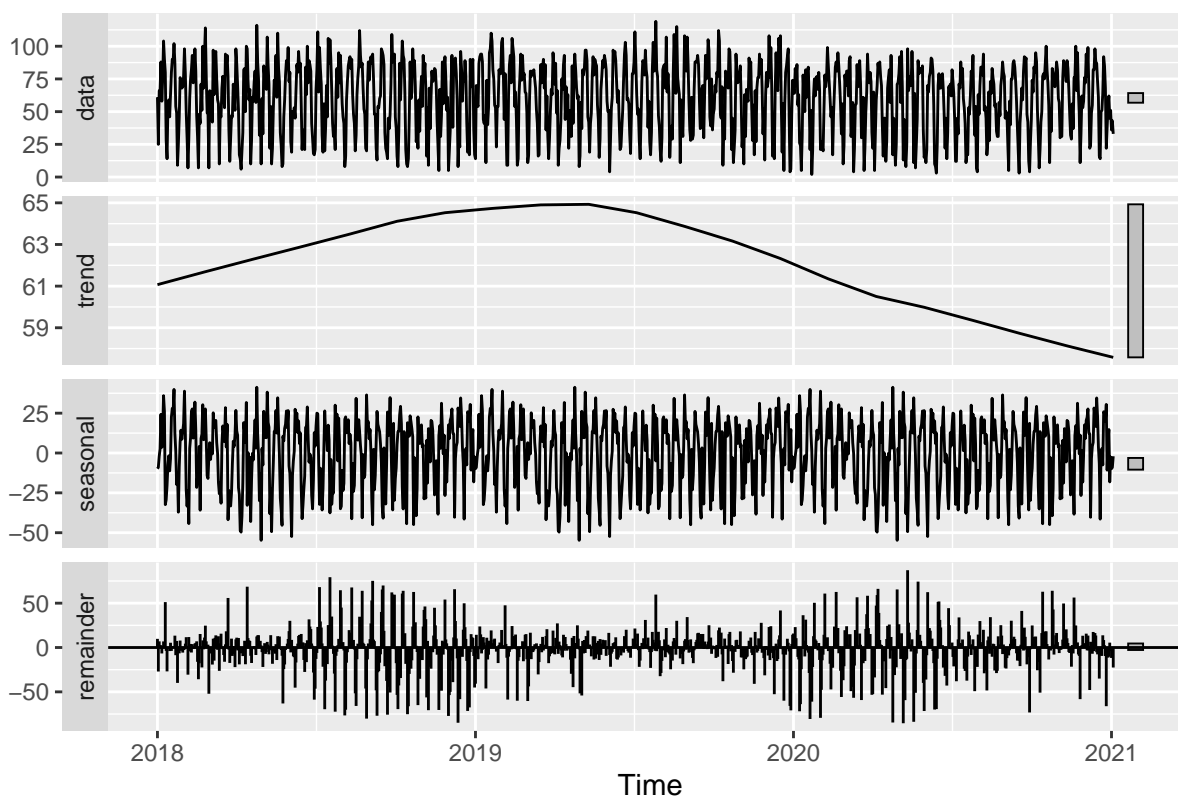
Estación 24, Casa de Campo, parámetro Ozono



Descomposición de serie temporal diaria para la estación 60, Plaza Tres Olivos.



Estación 60, Plaza Tres Olivos, parámetro Ozono



Modelos ARIMA

Series temporales de la estación 60 que se predicirán: magnitud 8= Dioxido de Nitrógeno, magnitud 10=Partículas<10, magnitud 14=ozono

Elaboramos el modelo magnitud 8, Dióxido de nitrógeno, en la estación 60

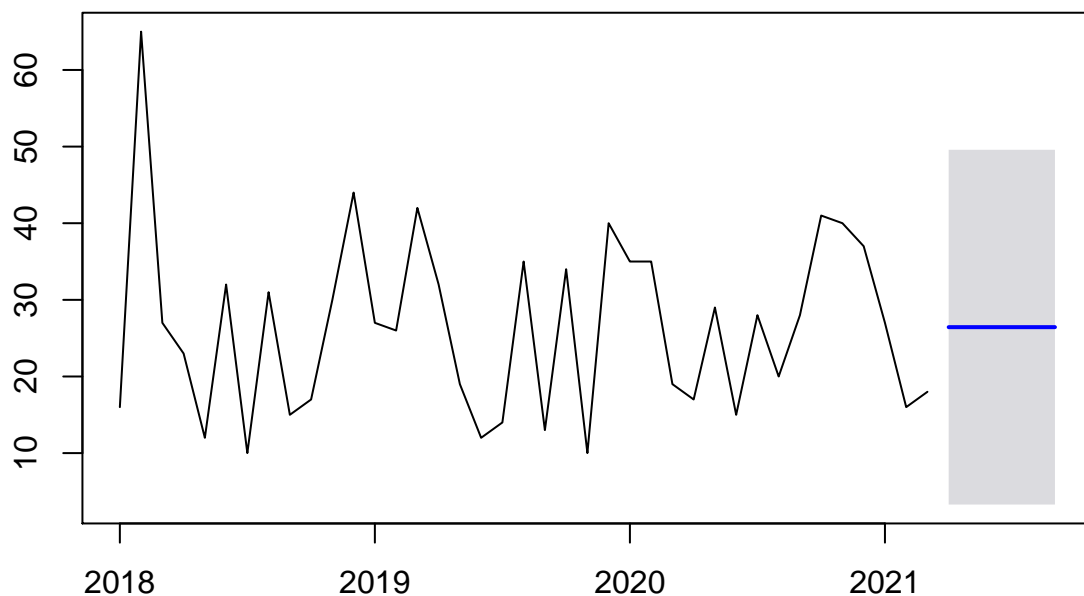
```
##
## ARIMA(2,0,2)(1,0,1)[12] with non-zero mean : Inf
## ARIMA(0,0,0) with non-zero mean : 306.5716
## ARIMA(1,0,0)(1,0,0)[12] with non-zero mean : 310.8688
## ARIMA(0,0,1)(0,0,1)[12] with non-zero mean : 310.917
## ARIMA(0,0,0) with zero mean : 375.1432
## ARIMA(0,0,0)(1,0,0)[12] with non-zero mean : 308.5791
## ARIMA(0,0,0)(0,0,1)[12] with non-zero mean : 308.602
## ARIMA(0,0,0)(1,0,1)[12] with non-zero mean : 311.0673
## ARIMA(1,0,0) with non-zero mean : 308.8182
## ARIMA(0,0,1) with non-zero mean : 308.8284
## ARIMA(1,0,1) with non-zero mean : 309.434
##
## Best model: ARIMA(0,0,0) with non-zero mean

## Series: tss_60_8
## ARIMA(0,0,0) with non-zero mean
```

```
##
## Coefficients:
##      mean
##      26.4359
## s.e.    1.8666
##
## sigma^2 estimated as 139.5:  log likelihood=-151.12
## AIC=306.24  AICc=306.57  BIC=309.57
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 1.585092e-14 11.65705 9.428008 -22.50909 45.66365 0.7667356
##              ACF1
## Training set -0.05178635
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 1.585092e-14 11.65705 9.428008 -22.50909 45.66365 0.7667356
##              ACF1
## Training set -0.05178635
```

Realizamos predicciones magnitud 8, Dióxido de nitrógeno, en la estación 60

Estación 60 – Predicción Dióxido de nitrógeno (8)



```
##
## Forecast method: ARIMA(0,0,0) with non-zero mean
##
## Model Information:
```

```
## Series: tss_60_8
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##          mean
##       26.4359
## s.e.    1.8666
##
## sigma^2 estimated as 139.5:  log likelihood=-151.12
## AIC=306.24   AICc=306.57   BIC=309.57
##
## Error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 1.585092e-14 11.65705 9.428008 -22.50909 45.66365 0.7667356
##              ACF1
## Training set -0.05178635
##
## Forecasts:
##      Point Forecast    Lo 95    Hi 95
## Apr 2021      26.4359 3.289819 49.58198
## May 2021      26.4359 3.289819 49.58198
## Jun 2021      26.4359 3.289819 49.58198
## Jul 2021      26.4359 3.289819 49.58198
## Aug 2021      26.4359 3.289819 49.58198
## Sep 2021      26.4359 3.289819 49.58198
```

Ruido blanco: es una secuencia de variables aleatorias de media cero, varianza constante y covarianzas nula. Esto significa que ningún modelo podrá ajustar los valores de la raíz cuadrada del parámetro de Dióxido de Nitrógeno.

Realizamos nuevo filtrado de fecha para separar la serie de Dióxido de Nitrógeno: guardamos sólo los datos desde Marzo de 2018 hasta Marzo 2020

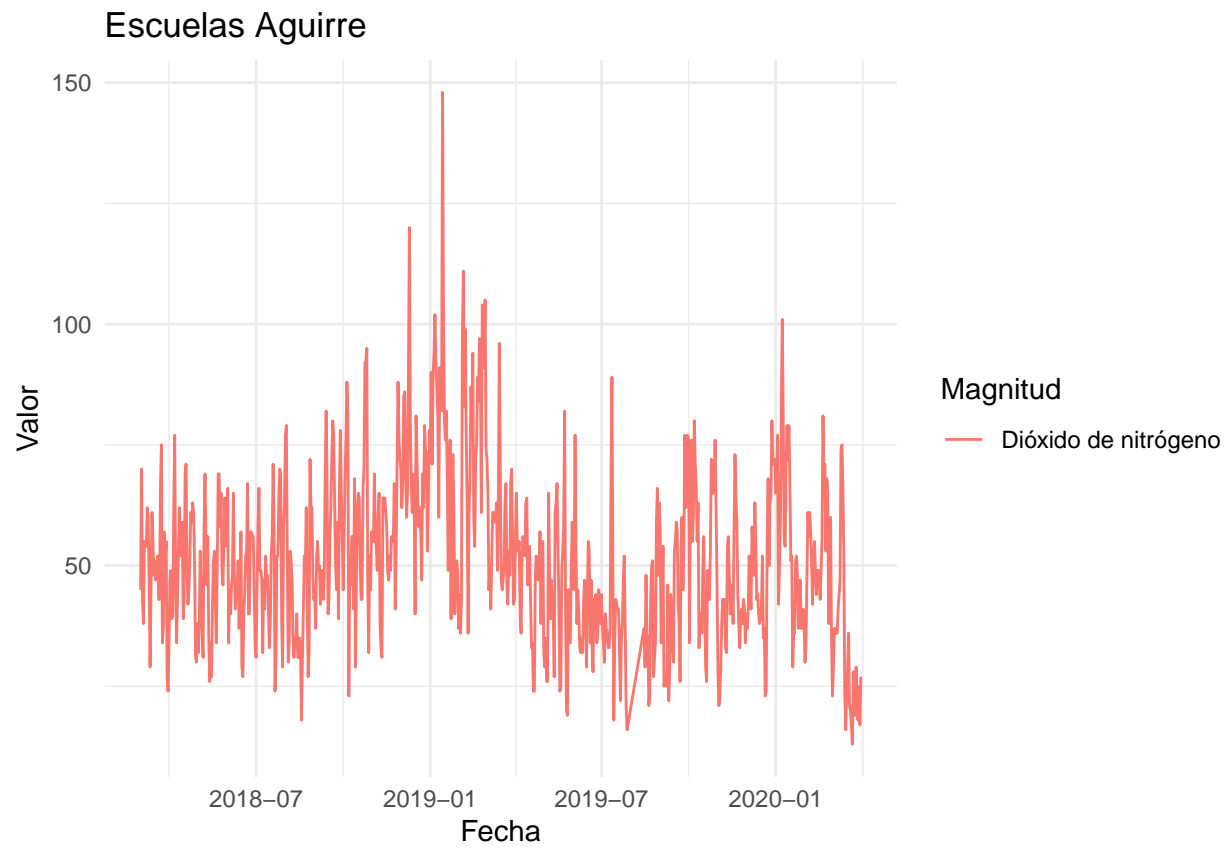
Construcción de las series desde Marzo de 2018 hasta Marzo de 2020, utilizamos dos años completos.

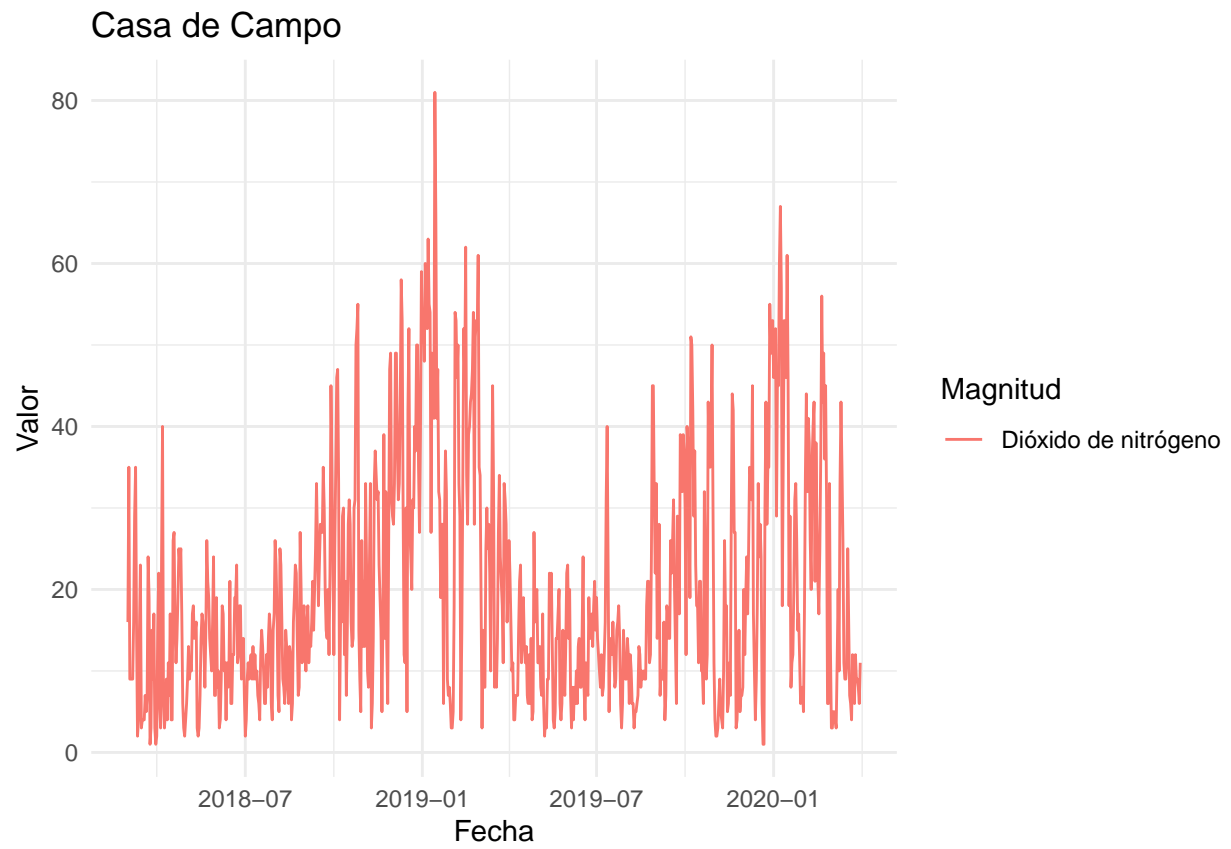
```
puntos5<- unique(df2$nombre_estacion)

for (punto in puntos5){

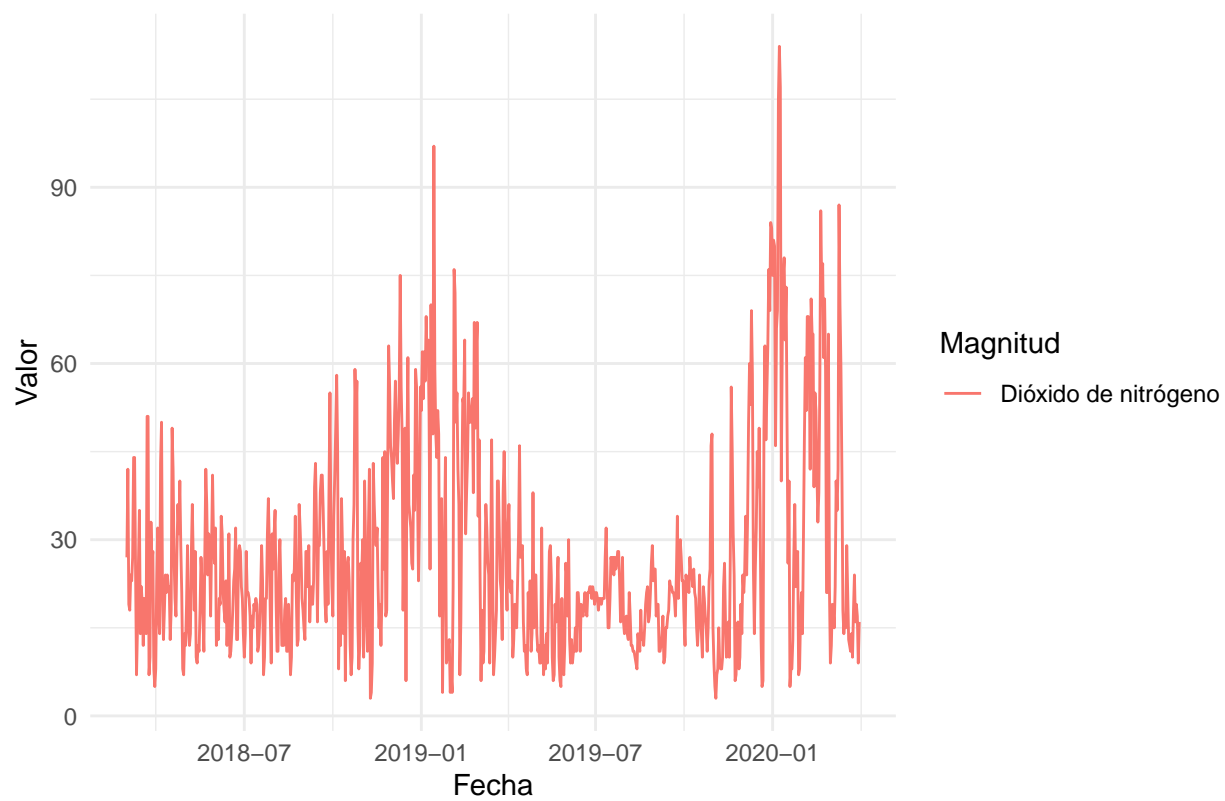
muestra_2 <- df2%>% filter(nombre_estacion==punto & magnitud == 8)

plot <- ggplot(muestra_2, aes(x=fecha, y=valor)) +
  geom_line(aes(col=as.factor(nombre_magnitud)))+ theme_minimal()+ ggtitle(punto)+
  xlab("Fecha")+ylab("Valor")+labs(color="Magnitud")
print(plot)
}
```

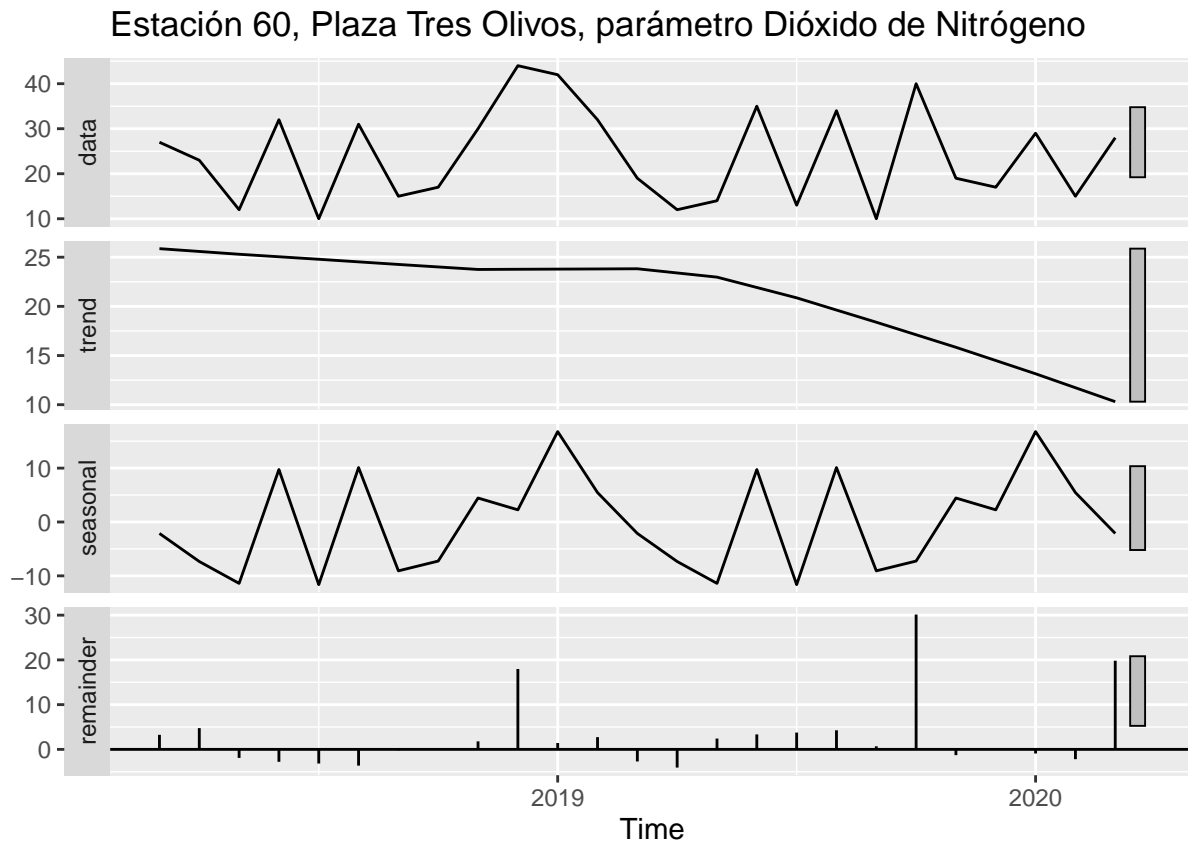




Plaza Tres Olivos



Descomposición de los datos de marzo 2018 - marzo 2020



Elaboramos el modelo magnitud 8, Dióxido de nitrógeno, en la estación 60

```
##
## ARIMA(2,0,2)(0,1,0)[12] with drift : Inf
## ARIMA(0,0,0)(0,1,0)[12] with drift : 23.81784
## ARIMA(1,0,0)(0,1,0)[12] with drift : 27.23986
## ARIMA(0,0,1)(0,1,0)[12] with drift : 27.23509
## ARIMA(0,0,0)(0,1,0)[12] : 22.29782
## ARIMA(1,0,1)(0,1,0)[12] with drift : Inf
##
## Best model: ARIMA(0,0,0)(0,1,0)[12]

## Series: tss2_60_8
## ARIMA(0,0,0)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## sigma^2 estimated as 0.2713: log likelihood=-9.97
## AIC=21.93 AICc=22.3 BIC=22.5
##
## Training set error measures:
## ME RMSE MAE MPE MAPE MASE
## Training set -1.917811 9.29326 5.442189 -16.41808 28.27401 0.5240626
## ACF1
```

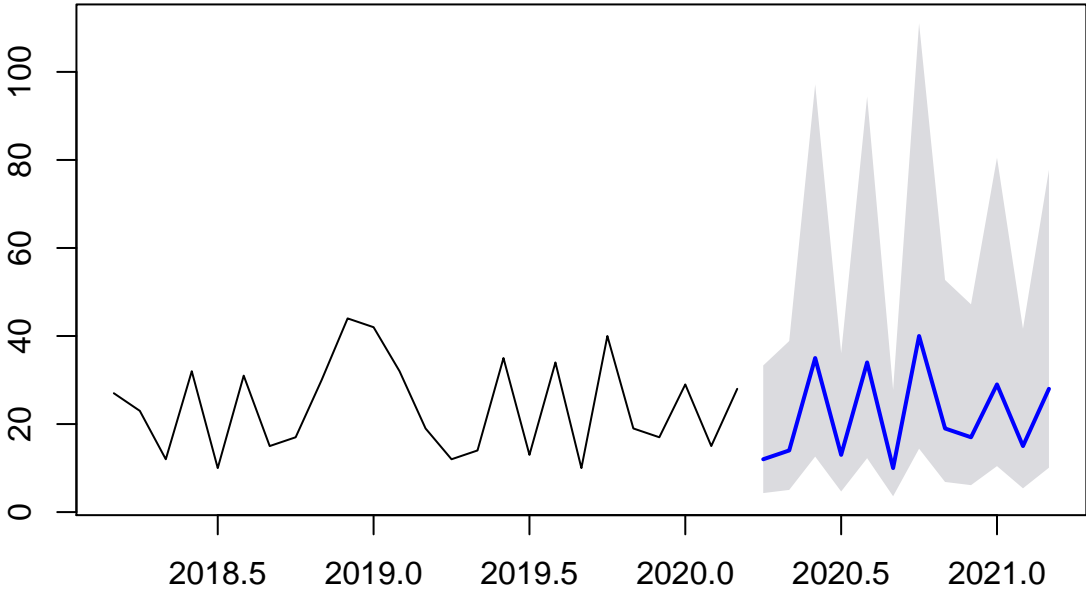


```
## Training set 0.1495496
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -1.917811 9.29326 5.442189 -16.41808 28.27401 0.5240626
##           ACF1
## Training set 0.1495496
```

Realizamos predicciones utilizando el modelo elaborado anteriormente del período marzo 2018 hasta marzo 2020

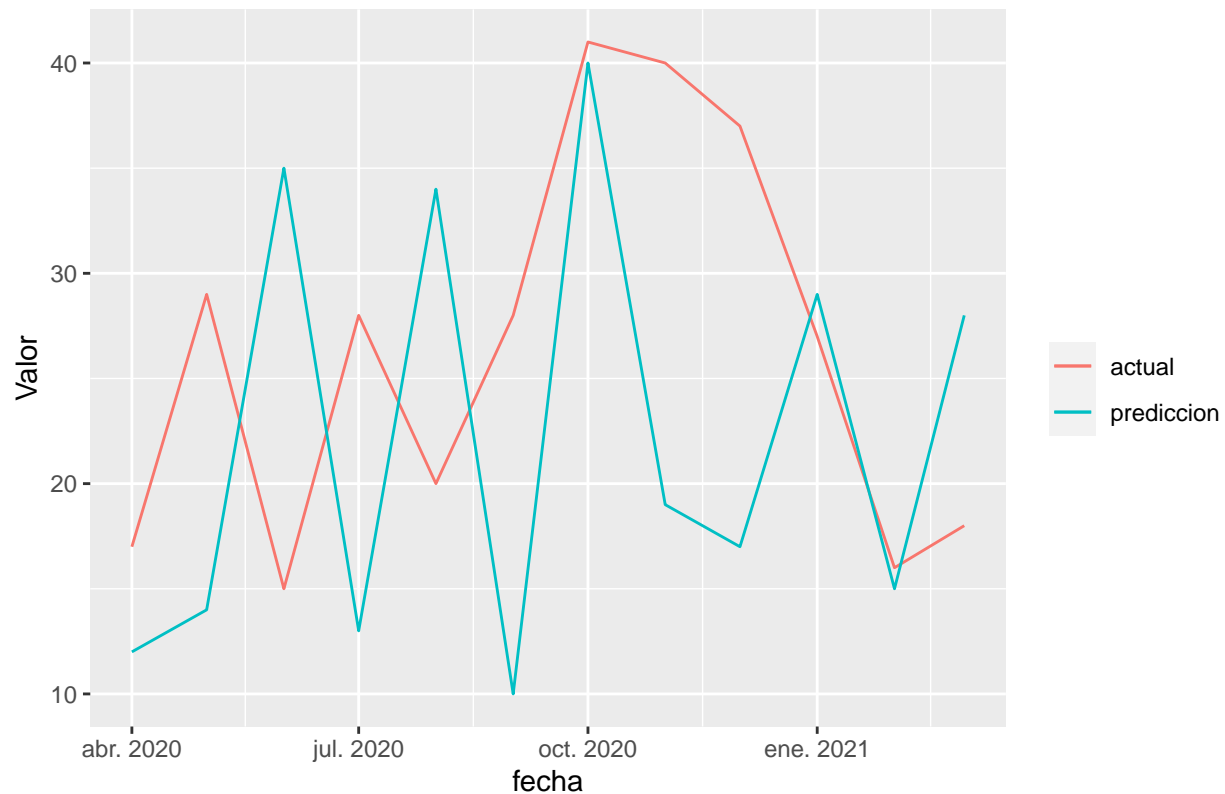
```
##
## Forecast method: ARIMA(0,0,0)(0,1,0)[12]
##
## Model Information:
## Series: tss2_60_8
## ARIMA(0,0,0)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## sigma^2 estimated as 0.2713: log likelihood=-9.97
## AIC=21.93 AICc=22.3 BIC=22.5
##
## Error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -1.917811 9.29326 5.442189 -16.41808 28.27401 0.5240626
##           ACF1
## Training set 0.1495496
##
## Forecasts:
##           Point Forecast      Lo 95      Hi 95
## Apr 2020           12  4.323159 33.30897
## May 2020           14  5.043686 38.86047
## Jun 2020           35 12.609215 97.15117
## Jul 2020           13  4.683423 36.08472
## Aug 2020           34 12.248951 94.37543
## Sep 2020           10  3.602633 27.75748
## Oct 2020           40 14.410531 111.02991
## Nov 2020           19  6.845002 52.73921
## Dec 2020           17  6.124476 47.18771
## Jan 2021           29 10.447635 80.49669
## Feb 2021           15  5.403949 41.63622
## Mar 2021           28 10.087372 77.72094
```

Estación 60 – Predicción Dióxido de nitrógeno (8)



Plot Comparativo de las valores reales vs valores predichos del Dióxido de nitrógeno(8)

Gráfico comparativo de los valores reales vs predichos del Dióxido de nitrógeno



Elaboración modelo magnitud 10, Partículas <10um, estación 60

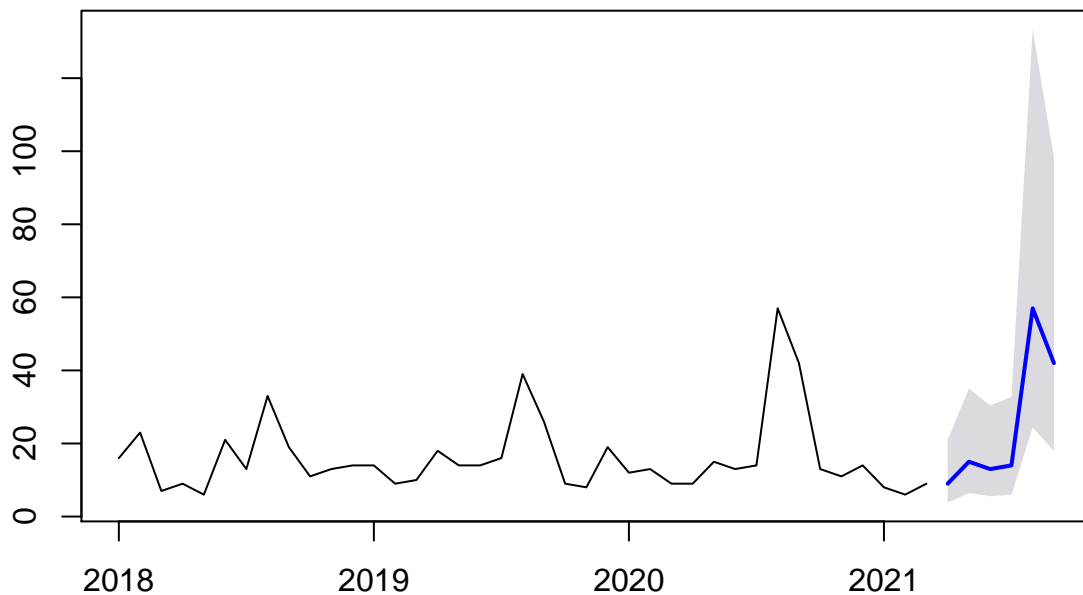
```
##
## ARIMA(2,0,2)(1,1,1)[12] with drift : Inf
## ARIMA(0,0,0)(0,1,0)[12] with drift : 35.97272
## ARIMA(1,0,0)(1,1,0)[12] with drift : 37.62663
## ARIMA(0,0,1)(0,1,1)[12] with drift : 38.70225
## ARIMA(0,0,0)(0,1,0)[12] : 33.63356
## ARIMA(0,0,0)(1,1,0)[12] with drift : 35.52406
## ARIMA(0,0,0)(0,1,1)[12] with drift : 36.72795
## ARIMA(0,0,0)(1,1,1)[12] with drift : Inf
## ARIMA(1,0,0)(0,1,0)[12] with drift : 37.81948
## ARIMA(0,0,1)(0,1,0)[12] with drift : 37.68143
## ARIMA(1,0,1)(0,1,0)[12] with drift : 40.446
##
## Best model: ARIMA(0,0,0)(0,1,0)[12]

## Series: tss_60_10
## ARIMA(0,0,0)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## sigma^2 estimated as 0.1878: log likelihood=-15.74
## AIC=33.47 AICc=33.63 BIC=34.77
```

```
##
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.6801317 5.892265 3.808337 -6.766323 26.91061 0.6947641
##           ACF1
## Training set 0.2698178
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.6801317 5.892265 3.808337 -6.766323 26.91061 0.6947641
##           ACF1
## Training set 0.2698178
```

Predicción la magnitud 10, Partículas <10um, estación 60

Estación 60 – Predicción Partículas <10um (10)



```
##
## Forecast method: ARIMA(0,0,0)(0,1,0)[12]
##
## Model Information:
## Series: tss_60_10
## ARIMA(0,0,0)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## sigma^2 estimated as 0.1878: log likelihood=-15.74
## AIC=33.47 AICc=33.63 BIC=34.77
##
## Error measures:
```

```

##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.6801317 5.892265 3.808337 -6.766323 26.91061 0.6947641
##              ACF1
## Training set 0.2698178
##
## Forecasts:
##      Point Forecast      Lo 95      Hi 95
## Apr 2021              9  3.848832 21.04535
## May 2021             15  6.414719 35.07558
## Jun 2021             13  5.559424 30.39884
## Jul 2021             14  5.987071 32.73721
## Aug 2021            57 24.375934 133.28720
## Sep 2021            42 17.961214  98.21162

```

Elaboración modelo magnitud 14, Ozono, estación 60

```

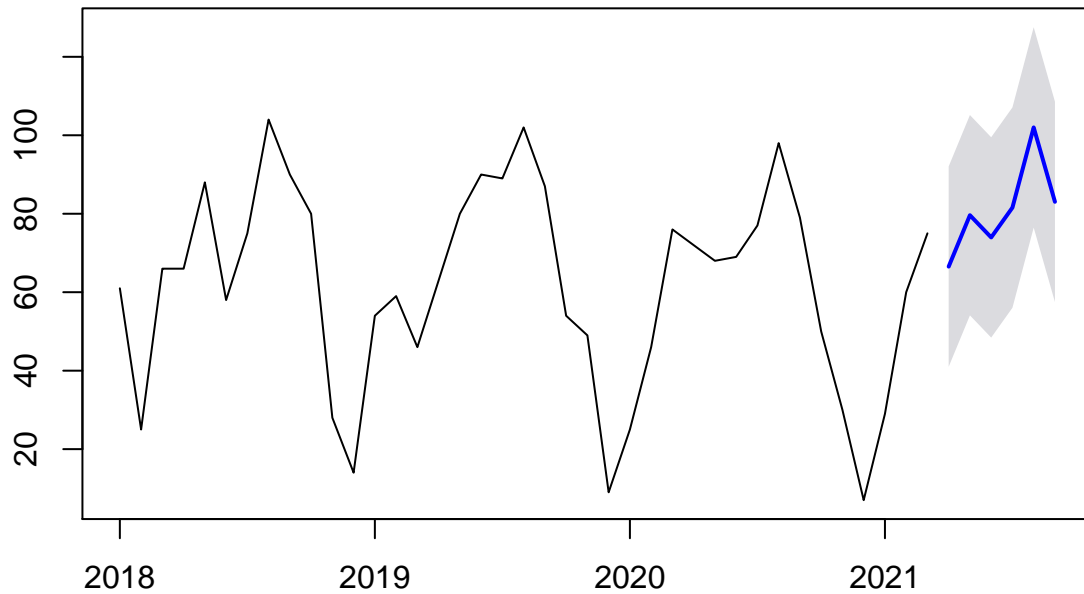
##
## ARIMA(2,0,2)(1,1,1)[12] with drift      : Inf
## ARIMA(0,0,0)(0,1,0)[12] with drift      : 23.54204
## ARIMA(1,0,0)(1,1,0)[12] with drift      : 26.29109
## ARIMA(0,0,1)(0,1,1)[12] with drift      : 26.67248
## ARIMA(0,0,0)(0,1,0)[12]                : 21.50968
## ARIMA(0,0,0)(1,1,0)[12] with drift      : 23.6254
## ARIMA(0,0,0)(0,1,1)[12] with drift      : 24.06747
## ARIMA(0,0,0)(1,1,1)[12] with drift      : Inf
## ARIMA(1,0,0)(0,1,0)[12] with drift      : 26.03625
## ARIMA(0,0,1)(0,1,0)[12] with drift      : 25.98164
## ARIMA(1,0,1)(0,1,0)[12] with drift      : Inf
##
## Best model: ARIMA(0,0,0)(0,1,0)[12]

## Series: tss_60_14
## ARIMA(0,0,0)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## sigma^2 estimated as 0.1199: log likelihood=-9.67
## AIC=21.35 AICc=21.51 BIC=22.65
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.9692425 13.81941 9.235886 -6.615966 19.21426 0.6985124
##              ACF1
## Training set -0.02903433
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.9692425 13.81941 9.235886 -6.615966 19.21426 0.6985124
##              ACF1
## Training set -0.02903433

```

Predicción la magnitud 14, Ozono, estación 60

Estación 60 – Predicción Ozono (14)



```
##
## Forecast method: ETS(A,N,A)
##
## Model Information:
## ETS(A,N,A)
##
## Call:
## ets(y = object, lambda = lambda, biasadj = biasadj, allow.multiplicative.trend = allow.multiplicati
##
## Smoothing parameters:
##   alpha = 0.0093
##   gamma = 1e-04
##
## Initial states:
##   l = 62.9391
##   s = -52.6565 -25.6945 -1.0254 20.3383 39.3008 18.8635
##       11.2607 16.9259 3.824 2.7415 -12.9175 -20.9608
##
## sigma: 13.0248
##
##      AIC      AICc      BIC
## 355.7507 376.6203 380.7041
##
## Error measures:
```

```

##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.6500867 10.42817 8.492121 -6.788008 19.18454 0.6422612
##               ACF1
## Training set -0.06655452
##
## Forecasts:
##      Point Forecast      Lo 95      Hi 95
## Apr 2021      66.52640 40.99831 92.05449
## May 2021      79.62783 54.09863 105.15703
## Jun 2021      73.96244 48.43213 99.49275
## Jul 2021      81.56540 56.03397 107.09683
## Aug 2021     102.00279 76.47025 127.53533
## Sep 2021      83.04125 57.50760 108.57491

```