



# Contaminación acústica en el barrio de Ruzafa

Tratamiento de Datos. Grado en Ciencia de Datos - UV

Miguel Caurín

Javier Martínez

Carlos Gila



# Índice

- Importación de los datos
- Análisis exploratorio
- Acondicionamiento de los datos
  - Detección y corrección de outliers
- Gráficos
  - Análisis univariante/bivariante
  - Gráficos de nivel sonoro



# Pero antes de comenzar...

Numeramos las calles para que sea más fácil trabajar con ellas:

1. Calle Sueca Esq. Denia

2. Calle Cádiz, 16

3. Calle Cádiz, 3

4. Calle Cuba, 3

5. Calle Sueca, 2

6. Calle Sueca, 61

7. Calle Sueca, 32

8. Calle Carles Cervera, Chaflán  
Reina Doña María

9. Calle Salvador Abril Chaflán  
Maestro José Serrano

10. Calle Vivons Chaflán Cádiz

11. Calle Carles Cervera, 34

12. Calle Puerto Rico, 21

13. Calle Doctor Serrano, 21

14. Calle General Prim Chaflán  
Donoso Cortés



# Importación de los datos

- Descargamos los 14 ficheros con toda la información
- Guardamos los nombres de todas las calles
- Fusionamos todos los datos en un solo data frame

```
{r}
nombre_estaciones_corto <-
c(" Sueca Esq. Denia",
  " Cádiz, 16",
  " Cádiz, 3",
  " Cuba, 3",
  " Sueca, 2",
  " Sueca, 61",
  " Sueca, 32",
  " Carles Cervera, Chaflán Reina Doña María",
  " Salvador Abril Chaflán Maestro José Serrano",
  " Vivons Chaflán Cádiz",
  " Carles Cervera, 34",
  " Puerto Rico, 21",
  " Doctor Serrano, 21",
  " General Prim Chaflán Donoso Cortés")
```

```
{r}
nombres <- paste0('df',1:14)

for(x in nombres){
  assign(x,read_csv(str_c("data/",x,".csv"),
                        col_types = cols(dateObserved =
                                          col_datetime(format = "%Y-%m-%d"))))
}

vector_dfs <- list(df1,df2,df3,df4,df5,df6,df7,df8,df9,df10,df11,df12,df13,df14)
```



# Análisis exploratorio

- Utilizamos glimpse y head para echar un primer vistazo a nuestros datos
- Utilizamos kable para observar estadísticos básicos

Var2	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
recvTime	2020-09-18 12	2021-01-18 00	2021-06-09 00	2021-07-04 09	2021-12-11 00	2022-04-19 00
LAeq	49.60	58.50	60.40	60.76	62.10	91.30
LAeq_d	51.3	59.6	61.7	Inf	63.3	Inf
LAeq_den	52.60	61.40	63.40	63.99	65.80	96.40
LAeq_e	46.3	57.8	60.1	Inf	63.2	Inf
LAeq_n	40.10	51.40	54.00	54.72	57.50	90.00
dateObserved	2020-09-17 00	2021-01-17 00	2021-06-08 00	2021-07-03 09	2021-12-10 00	2022-04-18 00



# Acondicionamiento de los datos

- Eliminamos datos innecesarios
- Renombramos las variables
- Representamos la variables en columnas
- Representamos las observaciones en filas

id	Fecha_dato	id_sensor	Dia_medidas	Estacion	Franja_horaria	Nivel_sonoro
1	2020-11-18 13:40:52	T248652-daily	2020-11-17	1	1min	62.4
2	2020-11-19 00:00:44	T248652-daily	2020-11-18	1	1min	58.9
3	2020-11-20 00:00:21	T248652-daily	2020-11-19	1	1min	59.3
4	2020-11-21 00:00:22	T248652-daily	2020-11-20	1	1min	59.4
5	2020-11-22 00:02:25	T248652-daily	2020-11-21	1	1min	58.9
6	2020-11-23 00:00:27	T248652-daily	2020-11-22	1	1min	56.9

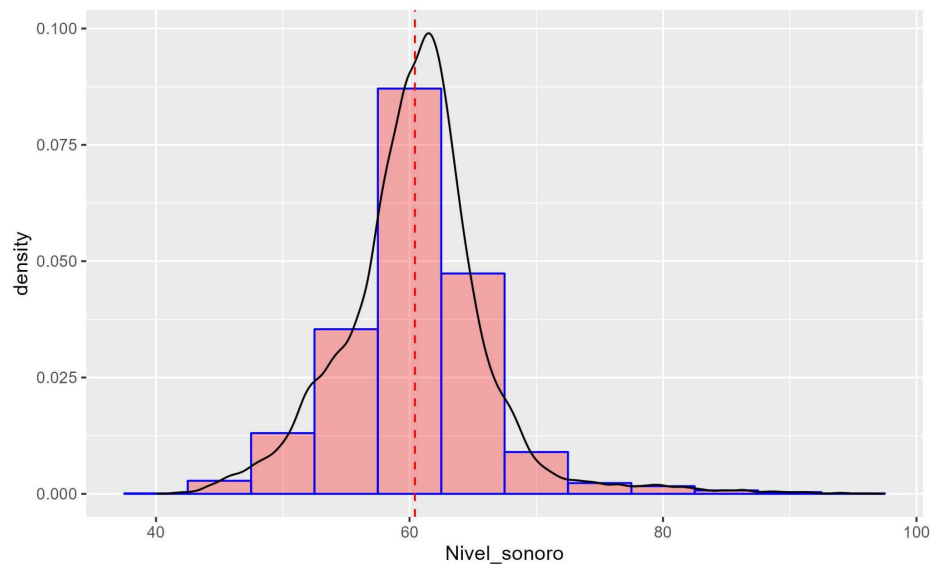


# Detección y corrección de outliers

- No puede haber registros con nivel sonoro infinito.
- Reemplazamos estos registros por la media del resto de registros.  
(Al tener tantos datos su alteración no supondrá cambios)

```
pos <- which(df$Nivel_sonoro==Inf)
df$Nivel_sonoro[pos]<-mean(df$Nivel_sonoro[-pos])
```

# Análisis univariante



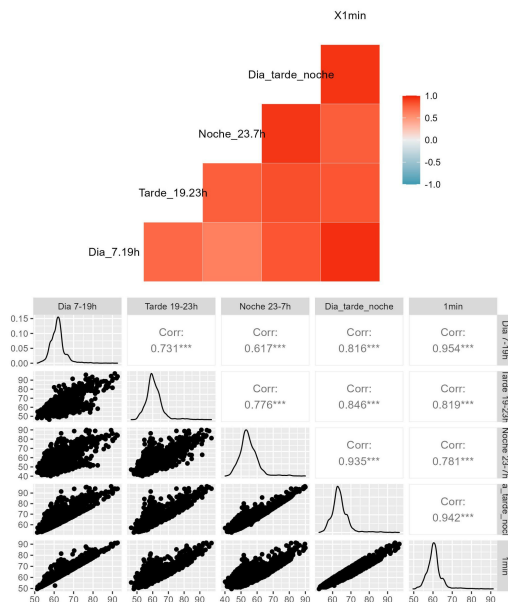
- Hacemos un histograma de densidad
- Hacemos uso de la función table

Var1	Freq
Dia 7-19h	6554
Tarde 19-23h	6554
Noche 23-7h	6554
Dia_tarde_noche	6554
1min	6554

Var1	Freq
1	2385
2	2415
3	2420
4	2275
5	1635
6	2435
7	2310
8	2360
9	2340
10	2420
11	2475
12	2365
13	2480
14	2455



# Análisis bivariante



- Utilizamos ggcorr y ggpairs para observar la correlación
- Utilizamos cov para obtener las matrices de covarianza

```
pearson <- cov(df$biv %>% select("Dia 7-19h", "Tarde 19-23h", "Noche 23-7h",
                                "Dia_tarde_noche", "1min"),
              method = "pearson")
kable(pearson) %>% kable_styling(bootstrap_options = c("striped", "bordered",
                                                       "hover", "condensed", "responsive"))
```

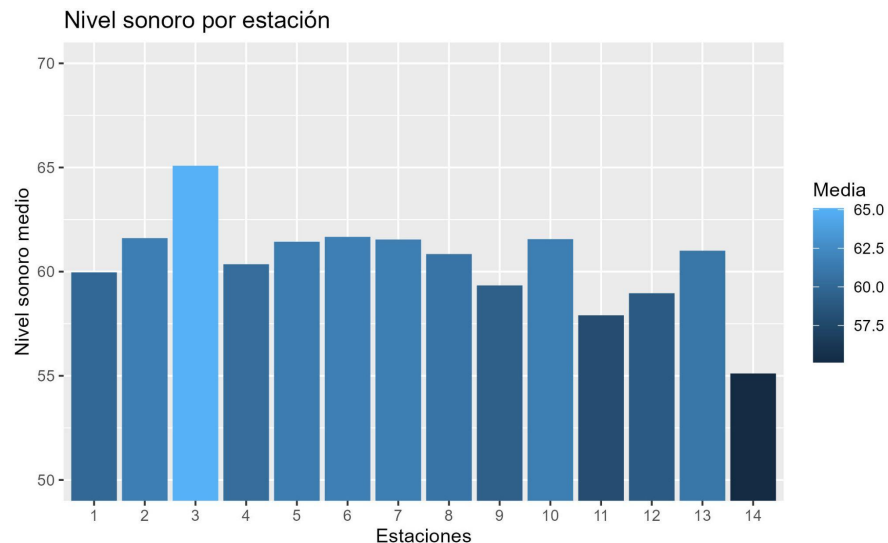
	Día 7-19h	Tarde 19-23h	Noche 23-7h	Día_tarde_noche	1min
Día 7-19h	15.70265	15.12466	14.05723	15.53460	15.94219
Tarde 19-23h	15.12466	27.25563	23.27546	21.20775	18.02823
Noche 23-7h	14.05723	23.27546	33.02982	25.79746	18.93591
Día_tarde_noche	15.53460	21.20775	25.79746	23.06050	19.07408
1min	15.94219	18.02823	18.93591	19.07408	17.78236

```
spearman <- cov(df$biv %>% select("Dia 7-19h", "Tarde 19-23h", "Noche 23-7h",
                                  "Dia_tarde_noche", "1min"),
               method = "spearman")
kable(spearman) %>% kable_styling(bootstrap_options = c("striped", "bordered",
                                                         "hover", "condensed", "responsive"))
```

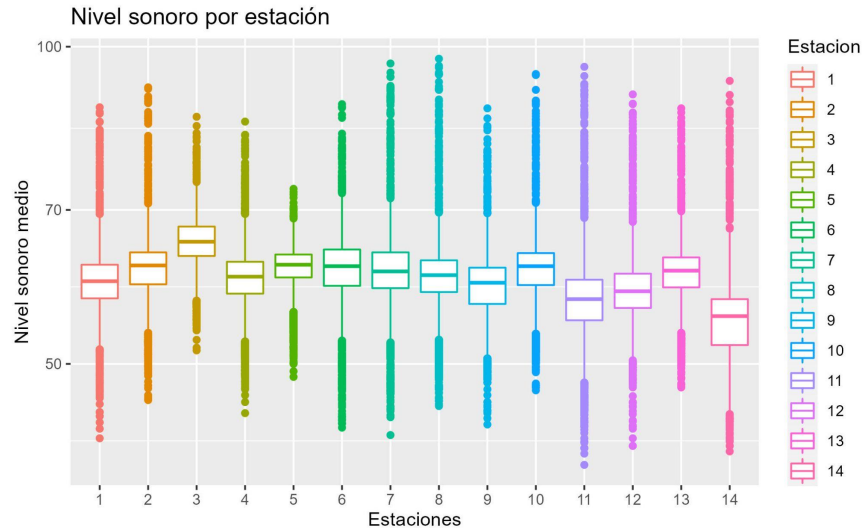
	Día 7-19h	Tarde 19-23h	Noche 23-7h	Día_tarde_noche	1min
Día 7-19h	3579689	2331063	1734941	2624739	3385012
Tarde 19-23h	2331063	3579900	2634406	2911871	2726074
Noche 23-7h	1734941	2634406	3579956	3259649	2419767
Día_tarde_noche	2624739	2911871	3259649	3579819	3193370
1min	3385012	2726074	2419767	3193370	3579685

# Nivel sonoro por estación

Representamos el nivel sonoro por estación con un gráfico de barras cambiando el color según su intensidad



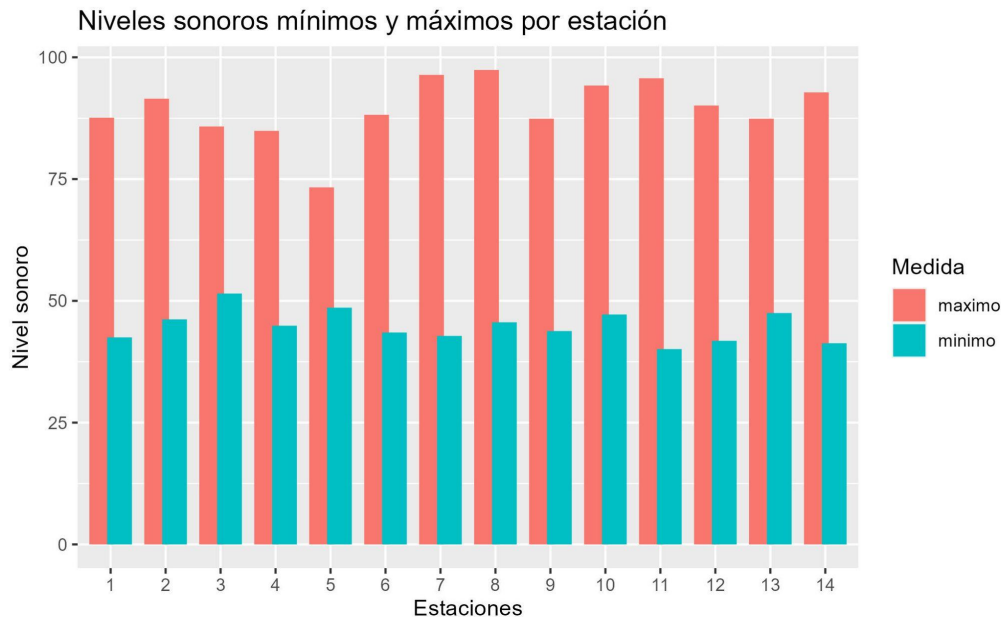
# Nivel sonoro por estación



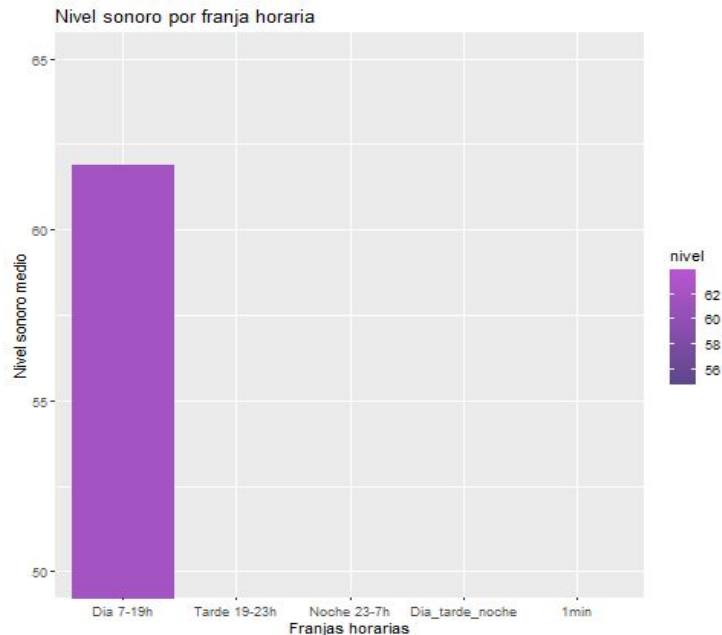
También lo representamos mediante boxplots, con los que podemos observar los datos de una manera distinta

# Niveles sonoros máximos y mínimos

Representamos en un mismo gráfico los niveles sonoros máximos y mínimos de cada estación



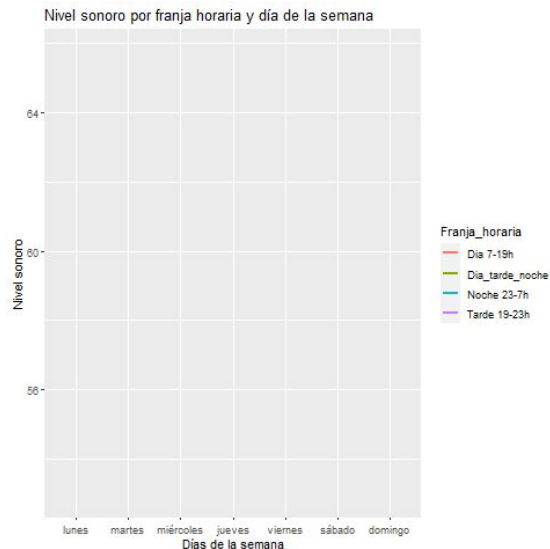
# Nivel sonoro por franja horaria



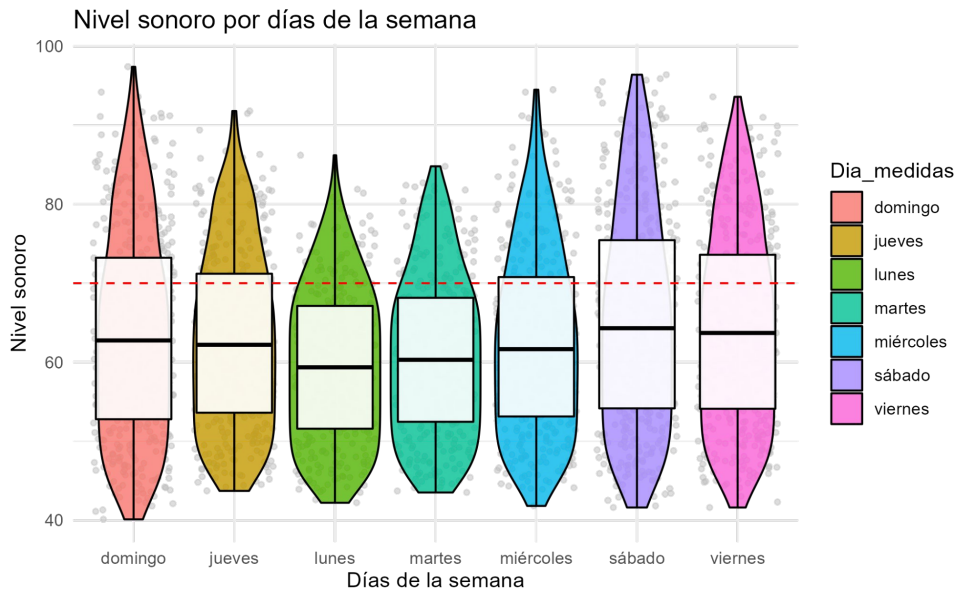
Representamos el nivel de sonido por franja horaria con un gráfico de barras cambiando el color según el nivel sonoro y lo animamos con gganimate

# Nivel sonoro por franja horaria y día de la semana

Hacemos un gráfico de líneas mostrando el nivel sonoro por franja horaria para cada día de la semana y lo animamos con gganimate



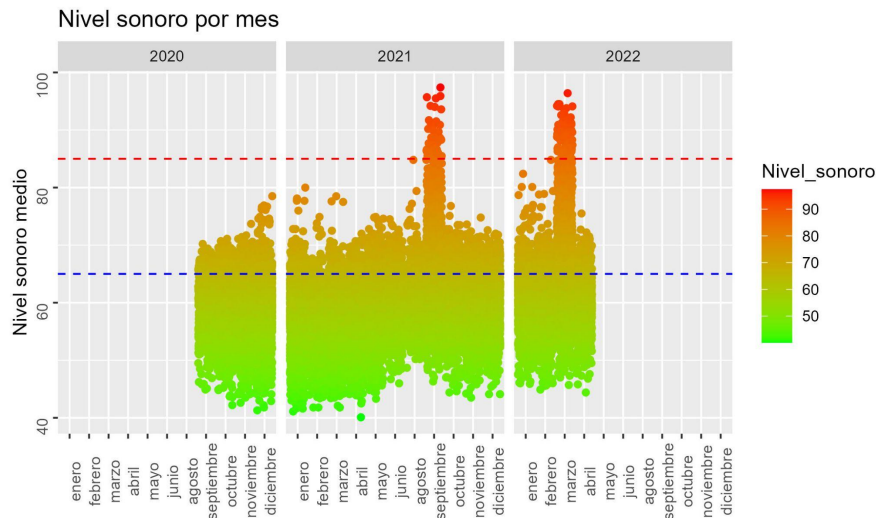
# Nivel sonoro por día de la semana



- Hacemos un gráfico de violín junto a un boxplot y un jitter
- Marcamos en rojo el nivel en el que comienza a afectar a nuestra concentración

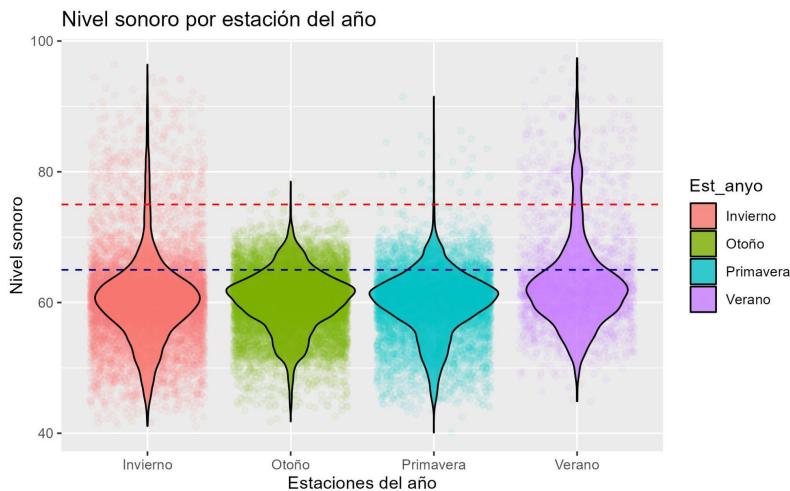
# Nivel sonoro por mes

- Usamos un jitter con facet\_grid para diferenciar por año
- Marcamos en rojo el nivel en el que empieza la pérdida auditiva crónica
- Marcamos en azul el umbral del nivel recomendado para una vida sana





# Nivel sonoro por estación del año

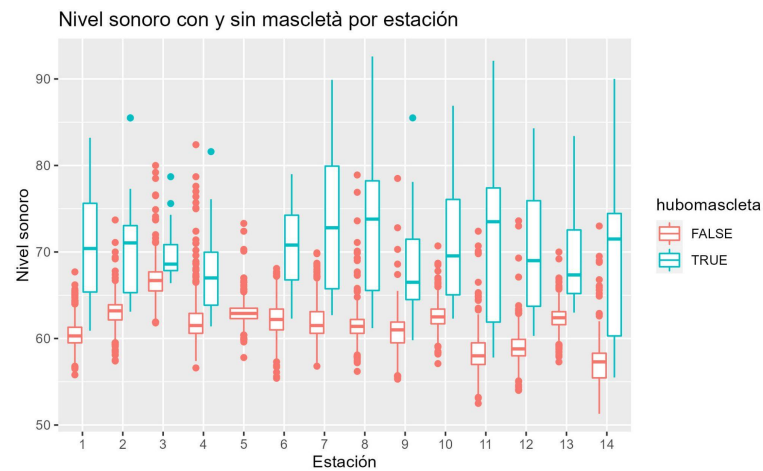


- Hacemos un gráfico de violín con un jitter de fondo
- Marcamos en azul el umbral a partir del cual el sonido se considera ruido
- Marcamos en rojo el nivel en el que el ruido comienza a ser dañino

# Nivel sonoro con mascletà

Hacemos boxplots que muestran el nivel de sonido en cada estación con y sin mascletà

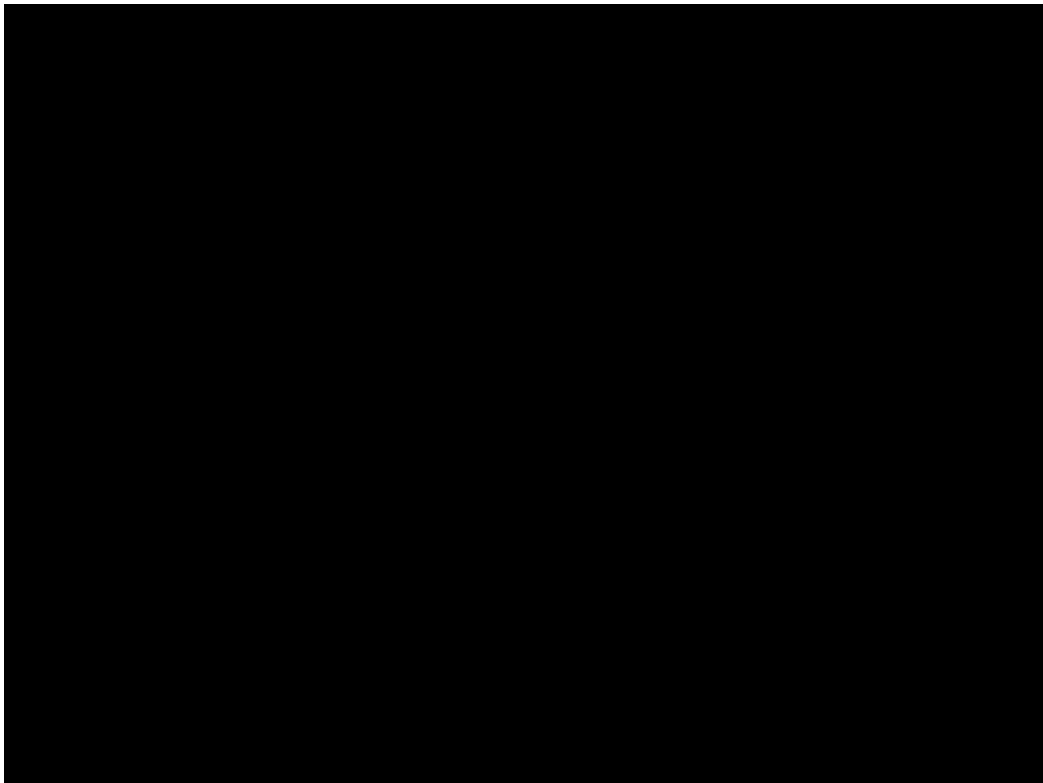
No hay datos de días de mascletà en la calle Sueca, 2





# Mapa interactivo

Mapa de Ruzafa que  
muestra el nivel  
sonoro en cada  
estación del barrio





# Conclusiones

- Muchos posibles enfoques
- Apertura de negocios
- Hostelería y ocio
- Bienestar y comodidad
- Contaminación acústica



# ¿Dudas?

