

Usando R para examinar la calidad del aire en Barcelona

Jose M. Sallan – Universitat Politècnica de Catalunya

Introducción

Introducción

- **Open Data BCN** ofrece más de 550 conjuntos de datos públicos de la ciudad de Barcelona.
- **Uso R y tidyverse** para:
 - Obtener y limpiar datos abiertos sobre calidad del aire (2018–2025).
 - Analizar su evolución temporal y espacial.
- **Objetivo:**
 - Demostrar el potencial de los **datos abiertos** y de **R** para el análisis ambiental urbano.

Contexto y relevancia

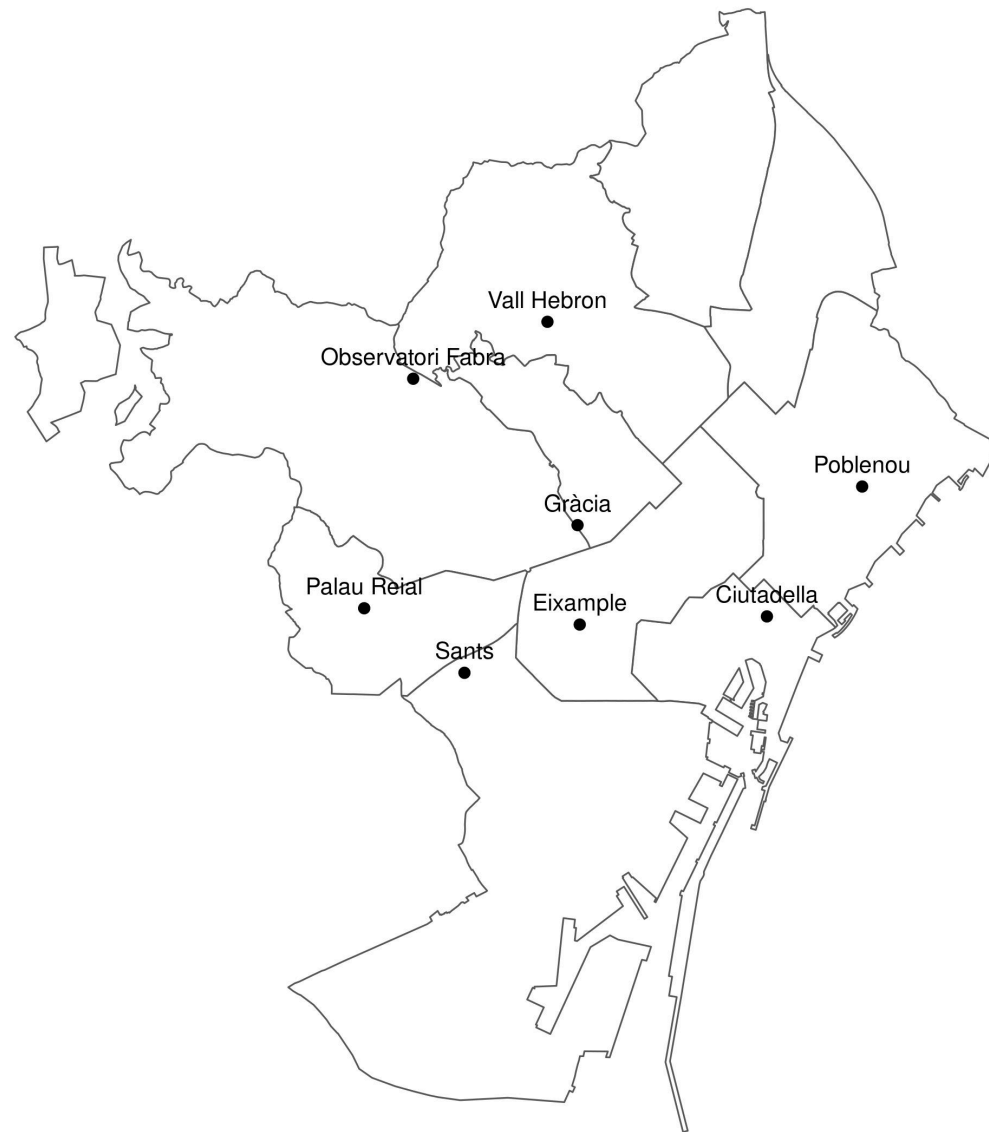
- Los portales de datos abiertos fomentan la transparencia y la investigación basada en evidencia.
- R es especialmente útil por su:
 - Ecosistema coherente (*tidyverse*).
 - Herramientas para análisis exploratorio y visualización.
 - Capacidad de reproducir y compartir resultados con **RMarkdown** y **blogdown**.

Adquisición y limpieza de datos

Datos de calidad del aire en Barcelona

Descripción	Nombre conjunto de datos
Contaminantes	<code>contaminants-estacions-mesura-qualitat-aire</code>
Estaciones	<code>qualitat-aire-estacions-bcn</code>
Datos horarios	<code>qualitat-aire-detall-bcn</code>

- Fuente: [Open Data BCN](#)
- Periodo: **2018–2025**
- Cobertura: 8 estaciones distribuidas en la ciudad.



Distribución geográfica de estaciones

- Las estaciones cubren la mayoría de distritos, excepto el norte (Sant Andreu, Nou Barris).
- Zonas de mayor densidad de población: **Eixample, Gràcia, Ciutadella.**
- Zonas menos pobladas: **Palau Reial, Observatori Fabra.**

Open Data BCN

- └─ contaminants-estacions-mesura-qualitat-aire
- └─ qualitat-aire-estacions-bcn
- └─ qualitat-aire-detall-bcn

```
[ckanr] Descubrimiento de recursos (.csv)
```

[readr + purrr] Descarga y lectura mensual (2018-2025)

[dplyr + purrr] Normalización de esquemas

- └ Mapear columnas 2018-2019 ↔ 2019-2025
- └ Arreglar separadores/encodings
- └ Armonizar IDs de estación

[tidyr::pivot_longer()] → Formato tidy (station, pollutant, datetime, value)

[dplyr::bind_rows()] Consolidación (≈3,15M filas) → guardar .RDS



EDA: mapa (ggplot)	Avisos/Episodios (dplyr + ggplot)	COVID-19 2020 vs 2021-2024 (ggplot)	Tablas/Gráficos (dplyr + ggplot)
-----------------------	--------------------------------------	--	-------------------------------------

Avisos/Episodios (dplyr + ggplot)	COVID-19 2020 vs 2021-2024 (ggplot)	Tablas/Gráficos (dplyr + ggplot)
--------------------------------------	--	-------------------------------------

COVID-19 2020 vs 2021-2024 (ggplot)	Tablas/Gráficos (dplyr + ggplot)
--	-------------------------------------

Tablas/Gráficos (dplyr + ggplot)

Publicación reproducible (rmarkdown/blogdown)

Análisis exploratorio de datos

Ejes de análisis

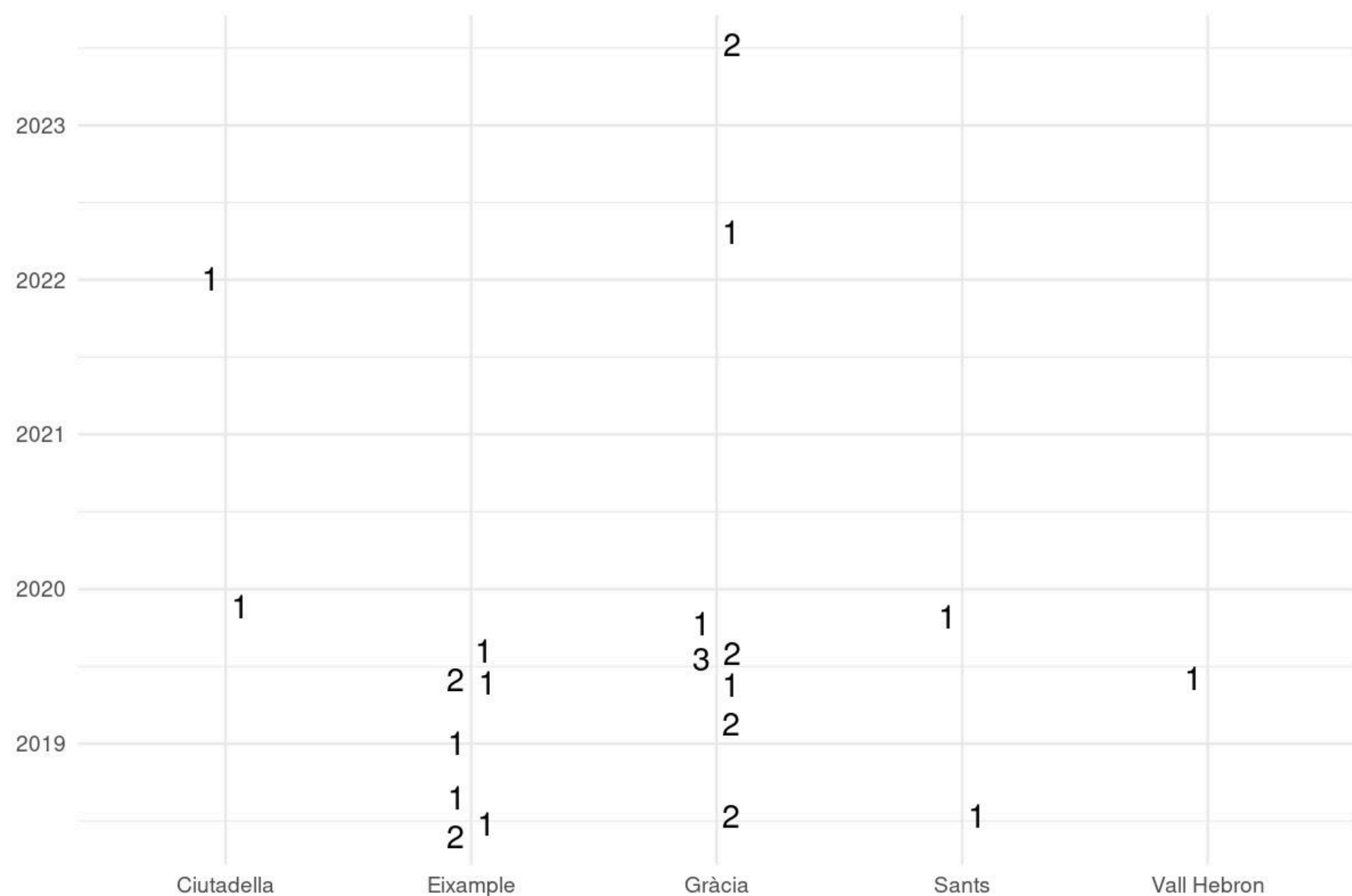
1. Avisos preventivos y episodios ambientales.
2. Impacto del confinamiento por COVID-19.

Avisos preventivos por contaminación

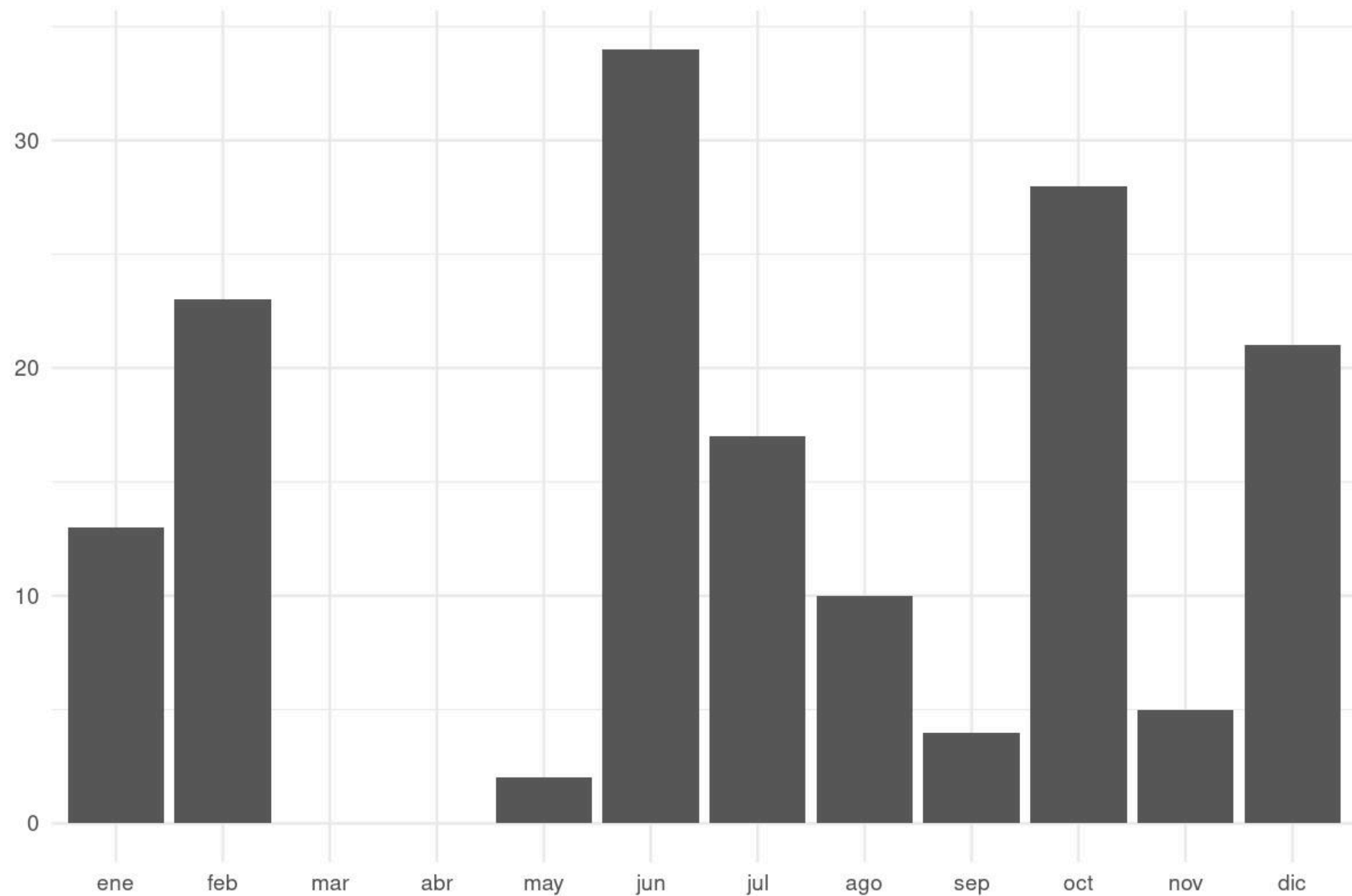
Criterios de aviso (Generalitat de Catalunya):

- $\text{NO}_2 > 160 \mu\text{g}/\text{m}^3$ (media horaria).
- $\text{PM}_{10} > 50 \mu\text{g}/\text{m}^3$ (media diaria).

Avisos por N02



Avisos por PM10



Avisos por PM10

- **127 días** con avisos desde 2018.
- Picos en **junio, octubre, diciembre y febrero**.
- Estaciones más afectadas: **Eixample, Poble Nou, Vall d'Hebron**.

Episodios ambientales

Criterios de episodio (Generalitat de Catalunya):

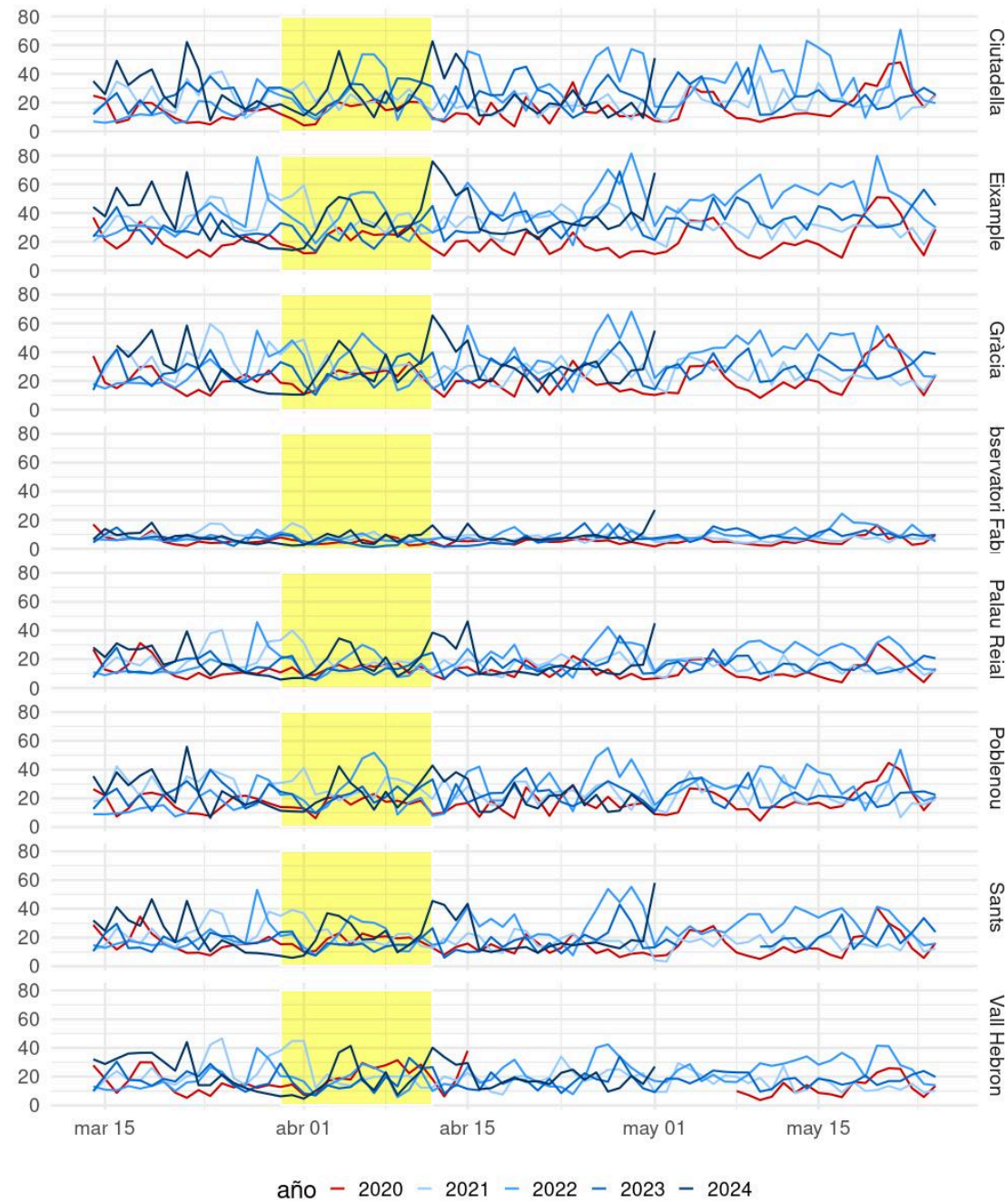
- $\text{NO}_2 > 200 \mu\text{g}/\text{m}^3$ (media horaria).
- $\text{PM}_{10} > 80 \mu\text{g}/\text{m}^3$ (media diaria).
- $\text{PM}_{10} > 50 \mu\text{g}/\text{m}^3$ (media diaria, tres días).

Episodios ambientales

- Un episodio de NO₂:
 - 28 de junio 2019
- Ocho episodios por PM₁₀, los más graves:
 - 17–18 diciembre 2019
 - 6 febrero 2021

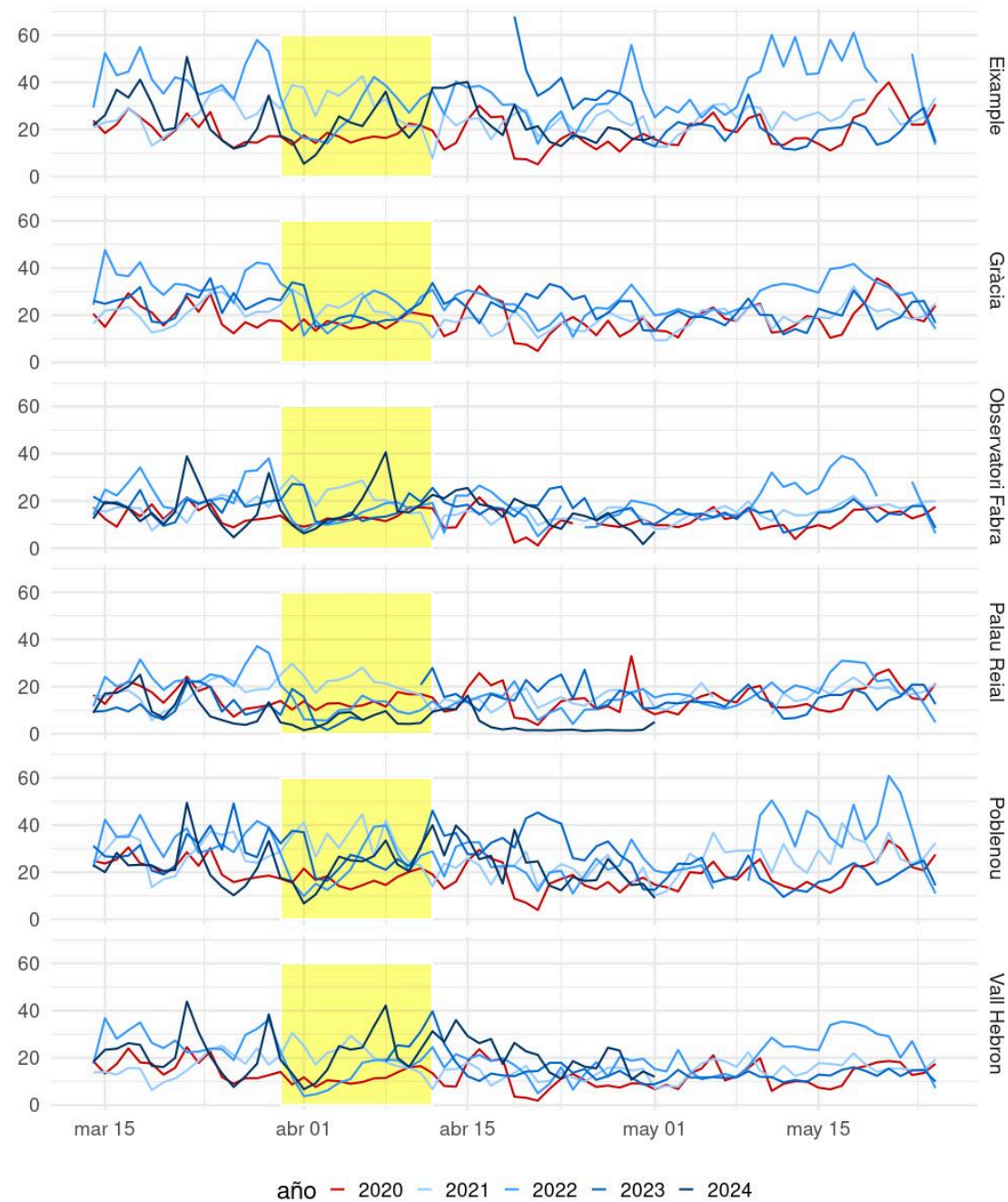
Impacto del confinamiento por COVID-19

- Confinamiento total: **14 marzo – 29 mayo 2020.**
- Reducción del tráfico hasta un **84%.**
- Oportunidad de *experimento natural*.
- Media diaria de NO₂ y PM₁₀ para mismo rango de fechas 2020-2024.



Evolución del NO2

- **NO2**: valores 2020 entre los más bajos.
- Estaciones con niveles más altos: **Eixample, Gràcia, Ciutadella**.
- Valores muy bajos en **Observatori Fabra**.



Evolución del PM10

- **PM10:** reducciones moderadas.
- No basta reducir movilidad: requiere políticas complementarias.

Conclusiones

Reproducibilidad y comunicación

- Proyecto completamente reproducible.
- Resultados parciales publicados en:
 - [Reading and Tidying Barcelona Air Quality Data](#)
 - [Warnings and Episodes of Pollution in Barcelona](#)
 - [Impact of COVID Lockdowns on Barcelona Air Quality](#)

Conclusiones

- R permite gestionar todo el flujo: *adquisición* → *limpieza* → *análisis* → *comunicación*.
- **Mayor desafío:** reorganización de datos heterogéneos a lo largo del tiempo.
- **Herramientas clave:**
 - `ckanr`, `tidyverse`, `rmarkdown`, `blogdown`.

Conclusiones

- Los datos abiertos requieren ingeniería de datos.
- Problema más relevante: inconsistencia de las series de datos en el tiempo.
- La reducción de tráfico no basta para mejorar la calidad del aire.