

02_analisis_exploratiu

Soulaiman el Hamri

2025-04-22

Contents

Introducció	1
1. Estadístiques descriptives dels contaminants	2
2. Distribució per contaminant	3
3. Evolució temporal: mitjanes diàries i comparativa per dècades	4
3.1 Preparació de dades diàries i càlcul de mitjana mòbil	4
3.2 Visualització de l'evolució per cada contaminant	5
4. Comparativa per tipus d'estació: trànsit vs urbana	8
5. Correlacions entre contaminants i variables meteorològiques	10
6. Anàlisi complementària	13
6.1 Anàlisi estacional (per mesos)	13
6.2 Detecció de dies crítics (valors extrems)	14
6.3 Resums per estació geogràfica	16
7. Conclusions generals	18

Introducció

Aquesta anàlisi exploratòria té com a objectiu entendre els patrons temporals i espacials dels principals contaminants atmosfèrics (NO_2 , $\text{PM}10$, $\text{PM}2.5$, O_3) a la ciutat de Barcelona, a partir de les dades preprocessades de meteorologia i qualitat de l'aire. En aquest document es realitzen les següents tasques:

- **Estadístiques descriptives** per a cadascun dels contaminants: càlcul de mitjanes, mediana, quartils, desviació estàndard, etc.
- **Visualitzacions de la distribució** de les concentracions (histogrames).
- **Anàlisi temporal** mitjançant mitjanes diàries i suavitzat amb finestra mòbil (30 dies), amb comparatives per dècades.
- **Comparativa per tipus d'estació** (trànsit vs. urbana de fons).
- **Anàlisi de correlacions** entre contaminants i variables meteorològiques.
- **Estudi estacional** per mesos, detecció de dies crítics i resum per estació geogràfica.

Els resultats d'aquesta exploració seran la base per a la següent fase de visualitzacions interactives.

Llegim les dades ja preprocessades per meteocat (dades meteorològiques) i contaminants.

```
df_meteo <- read_csv("../data/processed/meteocat/meteocat_2021_2025_bcn_processed.csv")
df_contaminants <- read_csv("../data/processed/contaminants/contaminants_bcn_filtrat.csv")
```

Convertim les dades horàries en un format “long” per facilitar l'anàlisi.

```
hour_columns <- sprintf("h%02d", 1:24)
contaminants_long <- df_contaminants %>%
  pivot_longer(
    cols = all_of(hour_columns),
    names_to = "hora",
    values_to = "valor"
  ) %>%
  mutate(
    hora = parse_number(hora),
    data = ymd(data),
    datetime = data + hours(hora - 1),
    mes = month(data, label = TRUE, abbr = TRUE),
    any = year(data)
  ) %>%
  drop_na(valor)
```

1. Estadístiques descriptives dels contaminants

Calculem estadístiques bàsiques per cadascun dels contaminants.

```
contaminants_long %>%
  group_by(contaminant) %>%
  summarise(
    n = n(), mínim = min(valor), q1 = quantile(valor, 0.25),
    mediana = median(valor), mitjana = mean(valor),
    q3 = quantile(valor, 0.75), màxim = max(valor), "NA" = sum(is.na(valor))
  ) %>%
  knitr::kable(digits = 2)
```

contaminant	n	mínim	q1	mediana	mitjana	q3	màxim	NA
NO2	1538632	0	18	36	41.37	59	483	0
O3	1205738	0	16	41	42.44	64	271	0
PM10	680474	0	15	23	28.30	35	1608	0
PM2.5	55076	1	6	9	11.08	14	180	0

Observació:

- **NO2:** Amb 1.538.632 observacions, la mediana és de $36 \mu\text{g}/\text{m}^3$ i la mitjana de $41.37 \mu\text{g}/\text{m}^3$, indicant una distribució asimètrica amb episodis d'alta contaminació (valor màxim de 483). Aquesta variabilitat es relaciona amb l'emissió de NO2 principalment pel trànsit i condicions atmosfèriques desfavorables (p.e. inversió tèrmica).

- **O3:** Amb 1.205.738 registres, la proximitat entre mediana ($41 \text{ }\mu\text{g}/\text{m}^3$) i mitjana ($42,44 \text{ }\mu\text{g}/\text{m}^3$) suggereix una distribució equilibrada. L'ozó a nivell del sòl es forma en condicions d'alta irradiació solar, habituals en els mesos d'estiu, i pot afectar la salut en períodes d'intensa exposició.
- **PM10:** Amb 680.474 mesures, la mediana és de $23 \text{ }\mu\text{g}/\text{m}^3$ i la mitjana de $28,30 \text{ }\mu\text{g}/\text{m}^3$. Tot i que la majoria de les concentracions són moderades, s'observa un valor màxim extrem de $1.608 \text{ }\mu\text{g}/\text{m}^3$, indicant episodis puntuals amb alta concentració que poden tenir un fort impacte en la salut respiratòria.
- **PM2.5:** Amb 55.076 observacions, la mediana és de $9 \text{ }\mu\text{g}/\text{m}^3$ i la mitjana de $11,08 \text{ }\mu\text{g}/\text{m}^3$, amb episodis d'alta concentració (valor màxim de 180). Aquestes partícules fines són especialment perilloses per la seva capacitat de penetrar profundament en els pulmons i fins i tot en el sistema sanguini.

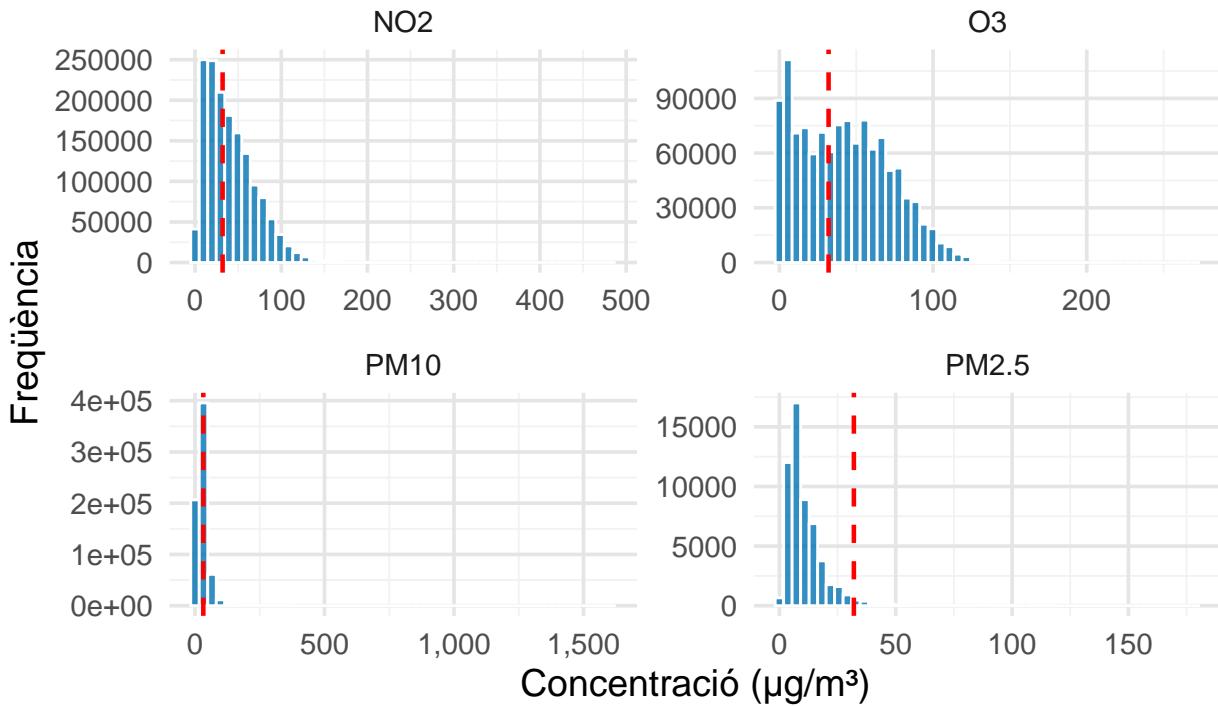
2. Distribució per contaminant

Visualitzem la distribució de les concentracions per cada contaminant mitjançant histogrames.

```
ggplot(contaminants_long, aes(x = valor)) +
  geom_histogram(bins = 50, fill = "#0072B2", color = "white", alpha = 0.8) +
  # Afegim una línia vertical per la mediana de cada grup (cada facet rep la seva pròpia mediana)
  geom_vline(aes(xintercept = median(valor, na.rm = TRUE)),
             color = "red", linetype = "dashed", linewidth = 0.8) +
  facet_wrap(~ contaminant, scales = "free") +
  labs(
    title = "Distribució horària dels contaminants",
    subtitle = "Línia vermella indica la mediana per cada contaminant",
    x = "Concentració ( $\mu\text{g}/\text{m}^3$ )",
    y = "Freqüència",
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5),
    panel.grid.major = element_line(color = "grey90"),
    panel.grid.minor = element_line(color = "grey95")
  ) +
  scale_x_continuous(labels = scales::comma)
```

Distribució horària dels contaminants

Línia vermella indica la mediana per cada contaminant



La gràfica mostra la **distribució de les concentracions** dels principals contaminants atmosfèrics: **NO2**, **O3**, **PM10** i **PM2.5**, expressades en $\mu\text{g}/\text{m}^3$.

- Els contaminants **NO2** i **PM10** mostren distribucions clarament asimètriques, amb episodis de concentracions altes que enerven la mitjana i aporten una cua dreta llarga.
- En **O3**, l'acumulació de valors és més concentrada al voltant de la mediana, tot i que existeix un rang ampli fins a 200–300 $\mu\text{g}/\text{m}^3$.
- **PM2.5** tendeix a concentrar la majoria de registres al voltant de valors relativament baixos, però no es descarten esdeveniments on les partícules fines pugen per sobre de 100 $\mu\text{g}/\text{m}^3$.

En resum, la gràfica mostra que cada contaminant té la seva pròpia dinàmica i rang de variabilitat. La mediana (línia vermella) ajuda a veure on es localitza la major part de les observacions, i la cua a la dreta en NO2, PM10 i PM2.5 ens indica que, tot i que la majoria de registres no són excessivament alts, hi ha episodis concrets on les concentracions pugen substancialment, fet important per avaluar riscos de salut i planificar mesures de mitigació.

3. Evolució temporal: mitjanes diàries i comparativa per dècades

3.1 Preparació de dades diàries i càlcul de mitjana mòbil

```
df_diari <- contaminants_long %>%
  mutate(
```

```

data = as.Date(datetime),
decada = paste0(floor(year(data) / 10) * 10, "s")
) %>%
group_by(data, contaminant, decada) %>%
summarise(valor_mitja = mean(valor, na.rm = TRUE), .groups = "drop") %>%
arrange(contaminant, data) %>%
group_by(contaminant) %>%
mutate(valor_suavitzat = zoo::rollmean(valor_mitja, k = 30, fill = NA)) %>%
ungroup()

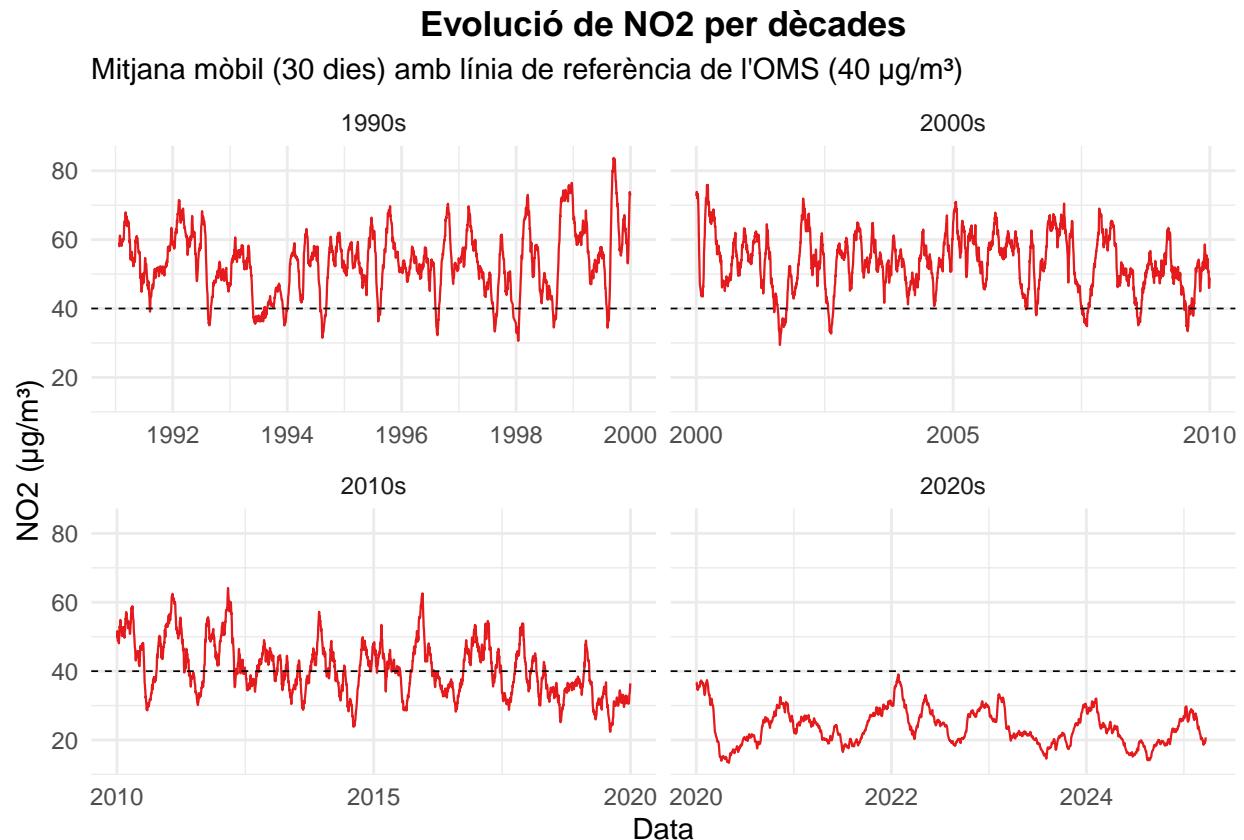
```

3.2 Visualització de l'evolució per cada contaminant

```

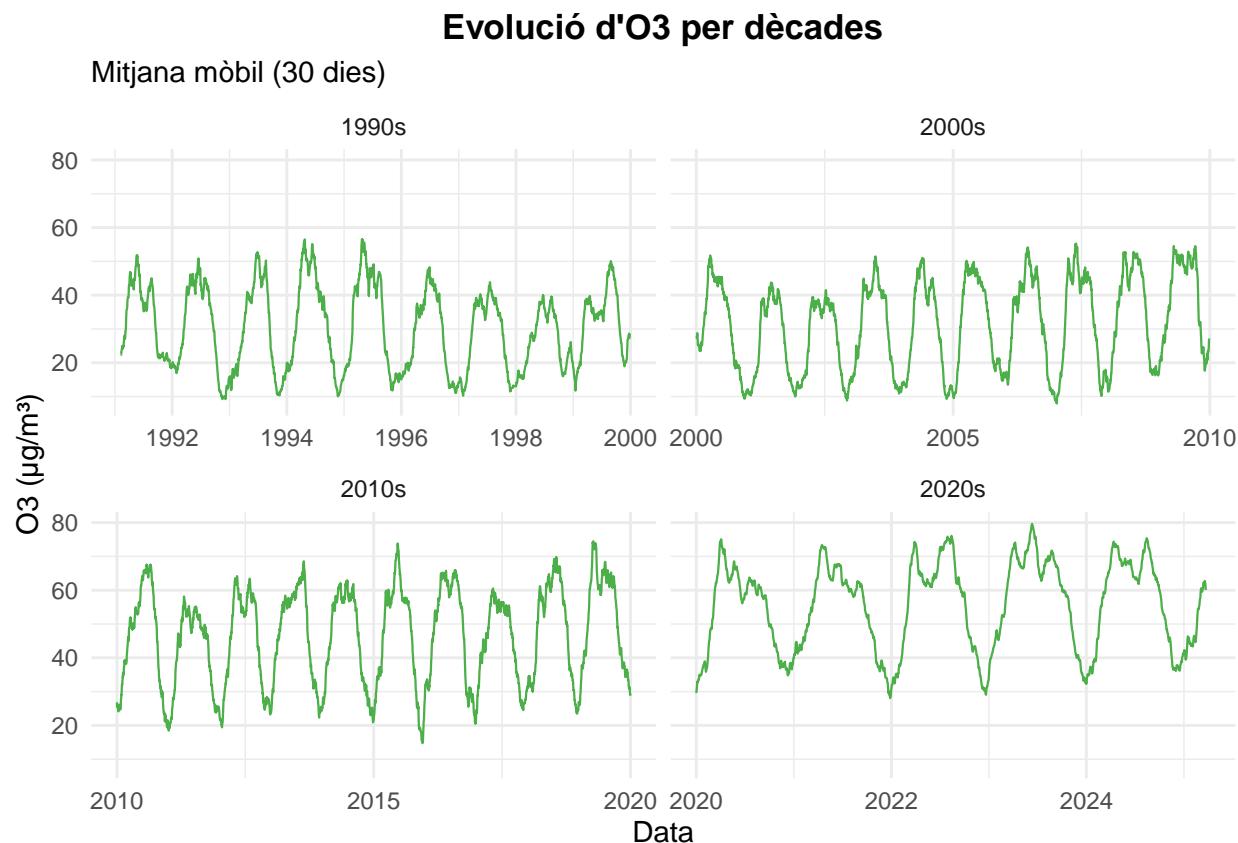
# Evolució per NO2 amb línia de referència de 40 µg/m³ (OMS)
ggplot(df_diari %>% filter(contaminant == "NO2"),
aes(x = data, y = valor_suavitzat)) +
geom_line(color = "#E41A1C", size = 0.4) +
geom_hline(yintercept = 40, linetype = "dashed", color = "black", size = 0.3) +
facet_wrap(~decada, scales = "free_x") +
labs(title = "Evolució de NO2 per dècades",
subtitle = "Mitjana mòbil (30 dies) amb línia de referència de l'OMS (40 µg/m³)",
x = "Data", y = "NO2 (µg/m³)") +
theme_minimal() +
theme(plot.title = element_text(face = "bold", hjust = 0.5))

```



La gràfica del NO₂ revela una evolució temporal on es poden observar, en primer lloc, fluctuacions diàries suavitzades amb una mitjana mòbil de 30 dies, la qual cosa facilita la detecció de tendències generals. En aquest sentit, s'identifiquen pics puntuals que superen significativament els valors mitjans, habitualment associats a condicions meteorològiques desfavorables com inversions tèrmiques o vents dèbils que impedeixen la dispersió dels contaminants. No obstant això, en períodes més recents es pot apreciar una tendència a la baixa, fet que podria reflectir l'efecte de les normatives ambientals més estrictes i la modernització del parc automobilístic amb vehicles menys contaminants. En conjunt, tot i que la concentració mitjana de NO₂ ha millorat, la persistència d'episodis amb altes concentracions subratlla la necessitat de continuar implementant mesures per la millora de la qualitat de l'aire i la protecció de la salut pública.

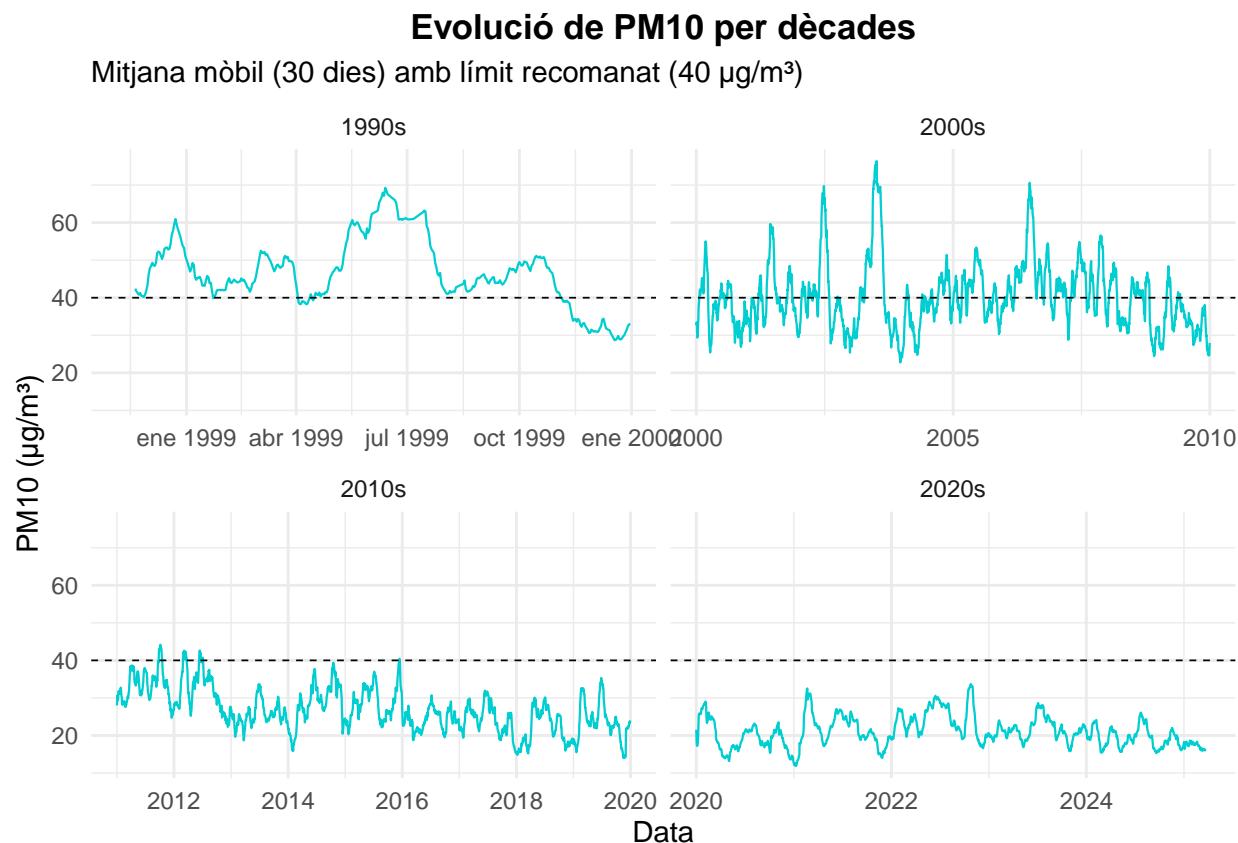
```
# Evolució per O3
ggplot(df_diari %>% filter(contaminant == "O3"),
       aes(x = data, y = valor_suavitzat)) +
  geom_line(color = "#4DAF4A", size = 0.4) +
  facet_wrap(~decada, scales = "free_x") +
  labs(title = "Evolució d'O3 per dècades",
       subtitle = "Mitjana mòbil (30 dies)",
       x = "Data", y = "O3 ( $\mu\text{g}/\text{m}^3$ )") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", hjust = 0.5))
```



La gràfica de l'O₃ mostra un patró clarament **estacional**, amb valors més alts durant els mesos calorosos i més baixos a l'hivern, un fet esperable atès que l'ozó troposfèric es genera principalment per reaccions fotoquímiques en presència d'irradiació solar. L'evolució al llarg de les diferents dècades indica que, tot i que el cicle anual es manté força constant—amb pics que poden acostar-se als 70–80 $\mu\text{g}/\text{m}^3$ —no s'observa una variació de fons molt marcada en la mitjana mòbil; més aviat, segueix el mateix patró repetitiu de

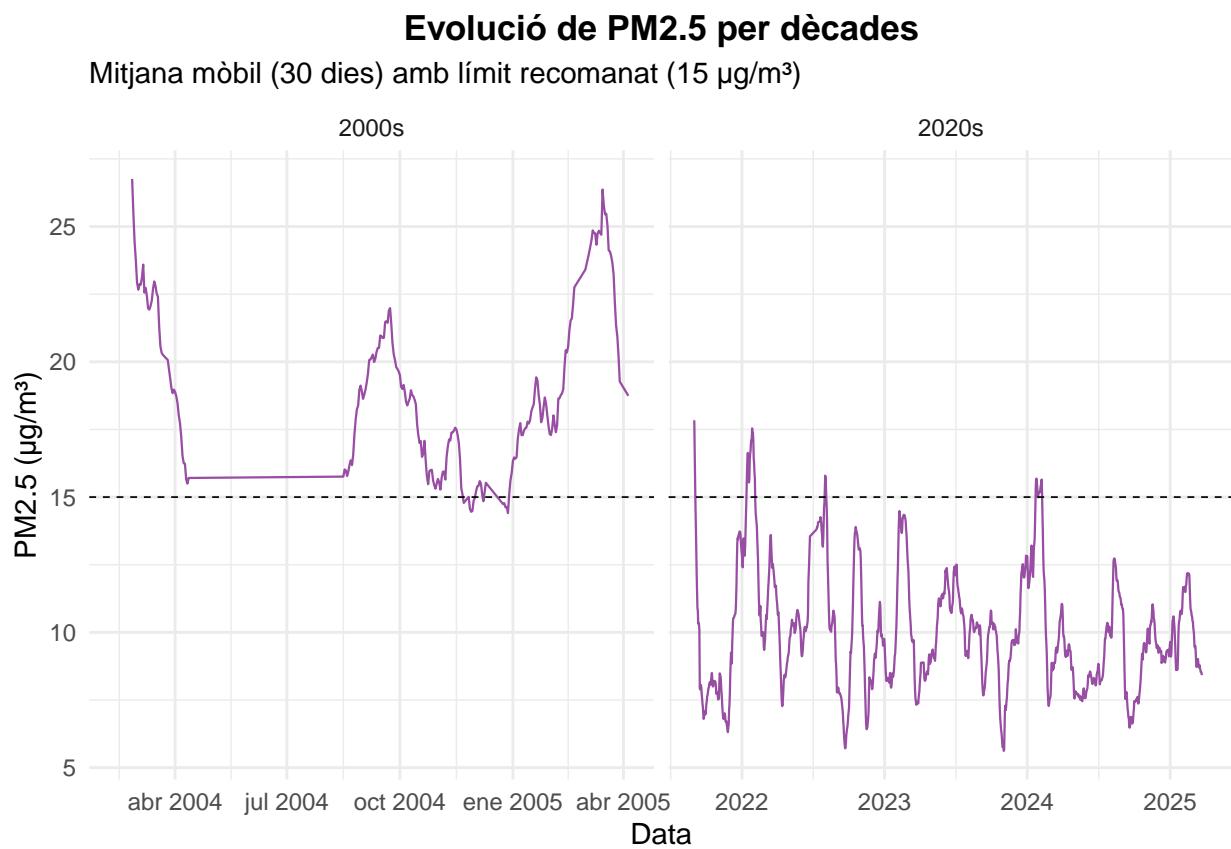
pujada i davallada en funció de les condicions meteorològiques. Cal destacar que, si bé la concentració d'ozó podria reduir-se lleugerament en determinats períodes a causa de canvis en les emissions de precursores (com òxids de nitrogen i compostos orgànics volàtils), aquesta reducció no és tan notòria en la sèrie temporal, la qual cosa suggereix que, en comparació amb altres contaminants, l'ozó manté un comportament cínic molt marcat any rere any.

```
# Evolució per PM10 amb límit recomanat de 40 µg/m³
ggplot(df_diari %>% filter(contaminant == "PM10"),
       aes(x = data, y = valor_suavitzat)) +
  geom_line(color = "#00CED1", size = 0.4) +
  geom_hline(yintercept = 40, linetype = "dashed", color = "black", size = 0.3) +
  facet_wrap(~decada, scales = "free_x") +
  labs(title = "Evolució de PM10 per dècades",
       subtitle = "Mitjana mòbil (30 dies) amb límit recomanat (40 µg/m³)",
       x = "Data", y = "PM10 (µg/m³)") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", hjust = 0.5))
```



La gràfica, que mostra l'evolució de PM10 en mitjana mòbil de 30 dies al llarg de diverses dècades, permet observar com, durant els anys noranta i principis del 2000, els nivells se situaven sovint al voltant dels 30–40 µg/m³ i ocasionalment superaven el límit recomanat de 40 µg/m³ (línia discontinua), fet que indica episodis puntuals de contaminació més elevada; a mesura que avança el temps, cap als períodes més recents, es percep una tendència a la baixa i una estabilització per sota o prop del valor de referència, possiblement degut a la implementació de normatives ambientals, millores tecnològiques en el sector del transport i canvis en la mobilitat urbana, tot i que encara es poden produir pics ocasionals associats a factors meteorològics i a fonts locals de partícules.

```
# Evolució per PM2.5 amb límit recomanat de 15 µg/m³
ggplot(df_diari %>% filter(contaminant == "PM2.5"),
       aes(x = data, y = valor_suavitzat)) +
  geom_line(color = "#984EA3", size = 0.4) +
  geom_hline(yintercept = 15, linetype = "dashed", color = "black", size = 0.3) +
  facet_wrap(~decada, scales = "free_x") +
  labs(title = "Evolució de PM2.5 per dècades",
       subtitle = "Mitjana mòbil (30 dies) amb límit recomanat (15 µg/m³)",
       x = "Data", y = "PM2.5 (µg/m³)") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", hjust = 0.5))
```



La gràfica reflecteix l'evolució de les partícules fines **PM2.5**, mesurades en mitjana mòbil de 30 dies, i compara els nivells observats amb el valor recomanat ($15 \mu\text{g}/\text{m}^3$) en diferents dècades. Al voltant de 2004–2005, s'observen pics que superen clarament el llindar de $15 \mu\text{g}/\text{m}^3$, assolint valors superiors a $20\text{--}25 \mu\text{g}/\text{m}^3$, fet que indica episodis de contaminació elevada. En canvi, en períodes més recents (cap al 2020 en endavant), la majoria de registres es mantenen per sota o prop de la línia de referència, suggerint una possible millora en la qualitat de l'aire. Tot i això, encara s'aprecien pics puntuals que s'aproximen als $15 \mu\text{g}/\text{m}^3$, la qual cosa subratlla la necessitat de continuar vigilant i aplicant mesures de control per assegurar que els nivells de PM2.5 romanguin dins de valors saludables.

4. Comparativa per tipus d'estació: trànsit vs urbana

Explorem com varien els valors mitjans segons el **tipus d'estació**.

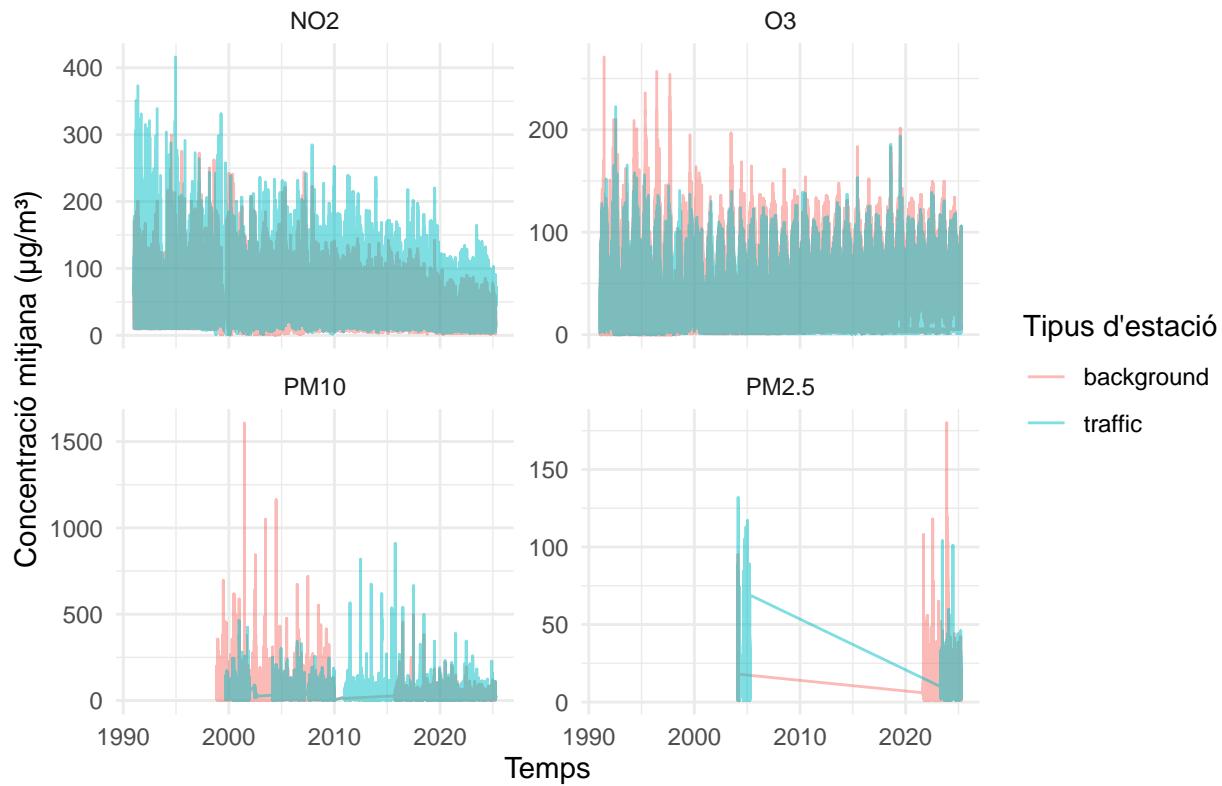
```

df_tipus <- contaminants_long %>%
  mutate(tipus = as.factor(tipus_estacio)) %>%
  group_by(datetime, contaminant, tipus) %>%
  summarise(valor_mitja = mean(valor, na.rm = TRUE), .groups = "drop")

ggplot(df_tipus, aes(x = datetime, y = valor_mitja, color = tipus)) +
  geom_line(alpha = 0.5) +
  facet_wrap(~contaminant, scales = "free_y") +
  labs(title = "Comparació entre tipus d'estació",
       x = "Temps", y = "Concentració mitjana (µg/m³)",
       color = "Tipus d'estació") +
  theme_minimal()

```

Comparació entre tipus d'estació



En aquesta visualització, s'hi comparen les concentracions mitjanes horàries de quatre contaminants (**NO₂**, **O₃**, **PM10** i **PM2.5**) en funció del **tipus d'estació**—concretament, estacions de fons urbà (**background**) i estacions situades en zones de trànsit (**traffic**)—al llarg del temps. A grans trets, s'hi observen els patrons següents:

- **NO₂:** Les estacions de trànsit (línia en color blau/verd) mostren clarament valors més elevats que les de fons urbà (línia en rosat), ja que el diòxid de nitrogen es genera majoritàriament per combustió de vehicles. Tot i que en èpoques recents sembla apreciar-se un lleuger descens, segueix existint una diferència notable entre ambdós tipus d'estació.
- **O₃:** De manera contrària, en molts intervals de temps es detecten concentracions més altes a les estacions de fons urbà que a les de trànsit. Això succeeix perquè, al costat de vies amb molt trànsit, els

òxids de nitrogen (NO, NO₂) poden “titrar” (destruir) part de l'ozó. Per tant, en zones més allunyades de les fonts de NO, els nivells d'O₃ tendeixen a ser superiors.

- **PM10:** Les estacions de trànsit també soLEN presentar valors més alts que les de fons, reflectint la contribució del trànsit rodat (desgast de pneumàtics, frenades, pols en suspensió) a l'augment de partícules de mida superior a 10 µm. En dècades recents es pot observar una lleugera millora, però encara es mantenen pics puntuals.
- **PM2.5:** Tot i que hi ha menys dades en algunes franges de temps, en general també es percepEN concentracions més elevades a la zona de trànsit que a la de fons. Les partícules fines (inferiors a 2,5 µm) tenEN un impacte significatiu en la salut, i l'activitat del trànsit n'és una font rellevant, juntament amb altres processos de combustió.

En definitiva, la comparativa evidencia que les estacions de **trànsit** registren concentracions clarament més elevades de contaminants primaris (NO₂, partícules) degut a l'emissió directa dels vehicles, mentre que les estacions de **fons urbà** soVENT mostren nivells relativament menors per a aquests contaminants però podEN presentar concentracions més altes d'**O₃**, a causa dels processos fotoquímics i l'efecte de titració del NO en zones de trànsit. Aquestes diferèNCIES remarquEN la necessitat de desplegar estratèGIES ambientals i de mobilitat que tinguin en compte tant la contaminació local (trànsit) com la de caràcter més difús (ozó, partícules fines), per tal de millorar globalMENT la qualitat de l'aire urbà.

5. Correlacions entre contaminants i variables meteorològiques

Integrem les dades diàries de meteorologia i contaminació per a calcular la matriu de correlacions.

```
# Agreguem les dades meteorològiques diàries
meteo_diari <- df_meteo %>%
  mutate(data = as.Date(DATA_LECTURA)) %>%
  group_by(data, ACRÒNIM) %>%
  summarise(valor = mean(VALOR, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = ACRÒNIM, values_from = valor)

# Ajustem les dades de contaminació, prèvies al càlcul de la mitjana
contam_diari <- df_diari %>%
  pivot_wider(names_from = contaminant, values_from = valor_mitja)

# Unim les dades per data
df_cor <- inner_join(meteo_diari, contam_diari, by = "data")

# Transformem les columnes a numèriques, excepte la columna 'data'
df_cor_numeric <- df_cor %>%
  select(-data) %>%
  mutate(across(everything(), as.numeric))

# Calculem la matriu de correlació utilitzant "pairwise.complete.obs"
cor_matrix <- cor(df_cor_numeric, use = "pairwise.complete.obs")

# Visualitzem la matriu de correlació amb ggcovrplot

# Reemplaça els NA/Nan/Inf amb 0 si cal reordenar les variables sense errors
cor_matrix[!is.finite(cor_matrix)] <- 0

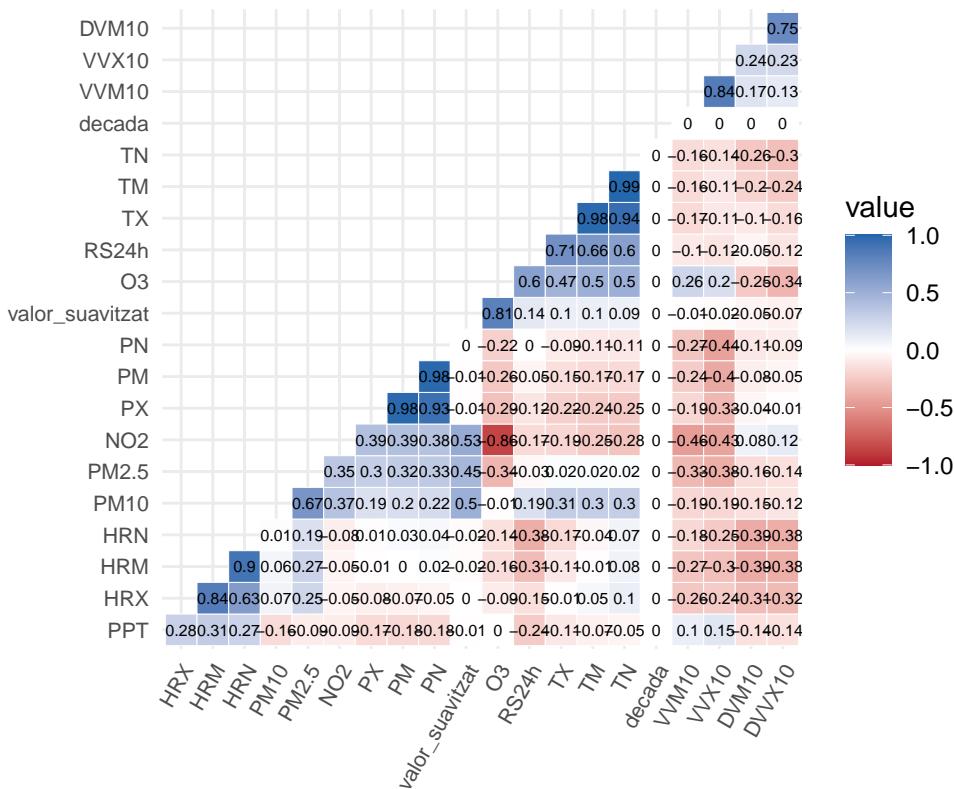
ggcovrplot(
  cor_matrix,
```

```

hc.order = TRUE,           # Ordena les variables amb agrupament jeràrquic
method = "square",         # Representa cada correlació en forma de quadrat
type = "lower",            # Només mostra la part inferior de la matriu
lab = TRUE,                # Mostra els valors numèrics dins de cada quadrat
digits = 2,                 # Arrodoneix a 2 decimals
lab_size = 2,               # Ajusta la mida del text de les correlacions
ggtheme = theme_minimal(),
outline.col = "white"    # Contorn dels quadrats en blanc
) +
scale_fill_gradient2(
  low = "#B2182B",        # Correlacions negatives en vermell
  mid = "white",           # Valor zero en blanc
  high = "#2166AC",        # Correlacions positives en blau
  midpoint = 0,
  limit = c(-1, 1)
) +
# Fixem la relació d'aspecte per no deformar la matriu
coord_fixed() +
theme(
  # Títol centrat i en negreta
  plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
  # Rotació i mida del text a l'eix X per evitar solapaments
  axis.text.x = element_text(angle = 60, vjust = 1, hjust = 1, size = 8),
  # Mida del text a l'eix Y
  axis.text.y = element_text(size = 8)
) +
labs(
  title = "Matriu de correlació: variables meteorològiques vs contaminants"
)

```

Matriu de correlació: variables meteorològiques vs contaminants



La matriu de correlacions mostra de manera sintètica les relacions entre variables meteorològiques (com la temperatura màxima **TX**, la temperatura mínima **TN**, el vent a 10 m **VVM10** i **VVX10**, la humitat **HRN** / **HRX**, la precipitació **PPT**, etc.) i diversos contaminants atmosfèrics (**NO2**, **O3**, **PM10**, **PM2.5**, etc.). A continuació es destaquen els aspectes més rellevants:

1. Relació entre contaminants:

- **PM10** i **PM2.5** tenen una correlació positiva elevada (valor al voltant de 0,7), fet que indica que sovint augmenten o disminueixen plegades.
- **NO2** i **PM10** també presenten una correlació positiva moderada (entorn de 0,4), cosa coherent amb el trànsit rodat com a font comuna d'aquests contaminants.
- **O3** mostra correlacions negatives amb NO2, PM10 i PM2.5 (entorn de -0,3), consistent amb el fenomen de “titració” de l’ozó per l’NO i amb el fet que, en condicions de molt trànsit, tendeix a haver-hi menys O3 a prop de les fonts emissives.

2. Influència de la temperatura i la radiació solar:

- Les variables relacionades amb la **temperatura** (p. ex. TX, TN) solen correlacionar-se positivament amb O3. Aquest fet reflecteix que, sota condicions de temperatura elevada i alta irradiació, es faciliten els processos fotoquímics que generen ozó.
- Sovint, contaminants com NO2 poden correlacionar-se negativament amb la temperatura, ja que l’ús de calefacció o la inversió tèrmica a l’hivern pot afavorir-ne l’acumulació en èpoques fredes.

3. Paper de la precipitació i la humitat:

- La **precipitació** (PPT) i les variables d'humitat (HRN, HRX) mostren correlacions negatives amb la majoria de contaminants. Això és conseqüència de l'efecte de rentatge (washout) que redueix la concentració de partícules i gasos contaminants en dies de pluja o humitat elevada.
- Quan plou o la humitat és alta, els contaminants tendeixen a deposar-se més ràpidament, fet que ajuda a netear l'aire.

4. Vent i dispersió:

- Variables com **VVM10** o **VVX10** (velocitat del vent) poden tenir correlacions lleugerament negatives amb NO₂, PM10 o PM2.5, reflectint que un vent més intens normalment dispersa els contaminants, reduint-ne la concentració local.
- Si el vent és molt feble, la dispersió és menor, afavorint l'acumulació de contaminants.

La matriu reflecteix dinàmiques típiques de la contaminació urbana i la meteorologia:

- **O₃** augmenta quan hi ha més temperatura i radiació, però decreix a prop de fonts de NO₂.
- **PM10 i PM2.5** s'acompanyen mútuament i també correlacionen amb NO₂ per la font comuna del trànsit.
- **La pluja i el vent** afavoreixen la dispersió i la reducció de concentracions de contaminants.

Aquests patrons són els habituals en àrees urbanes, i reforcen la idea que les condicions meteorològiques influeixen de manera significativa en la qualitat de l'aire, a més de les característiques d'emissió pròpies de cada contaminant.

6. Anàlisi complementària

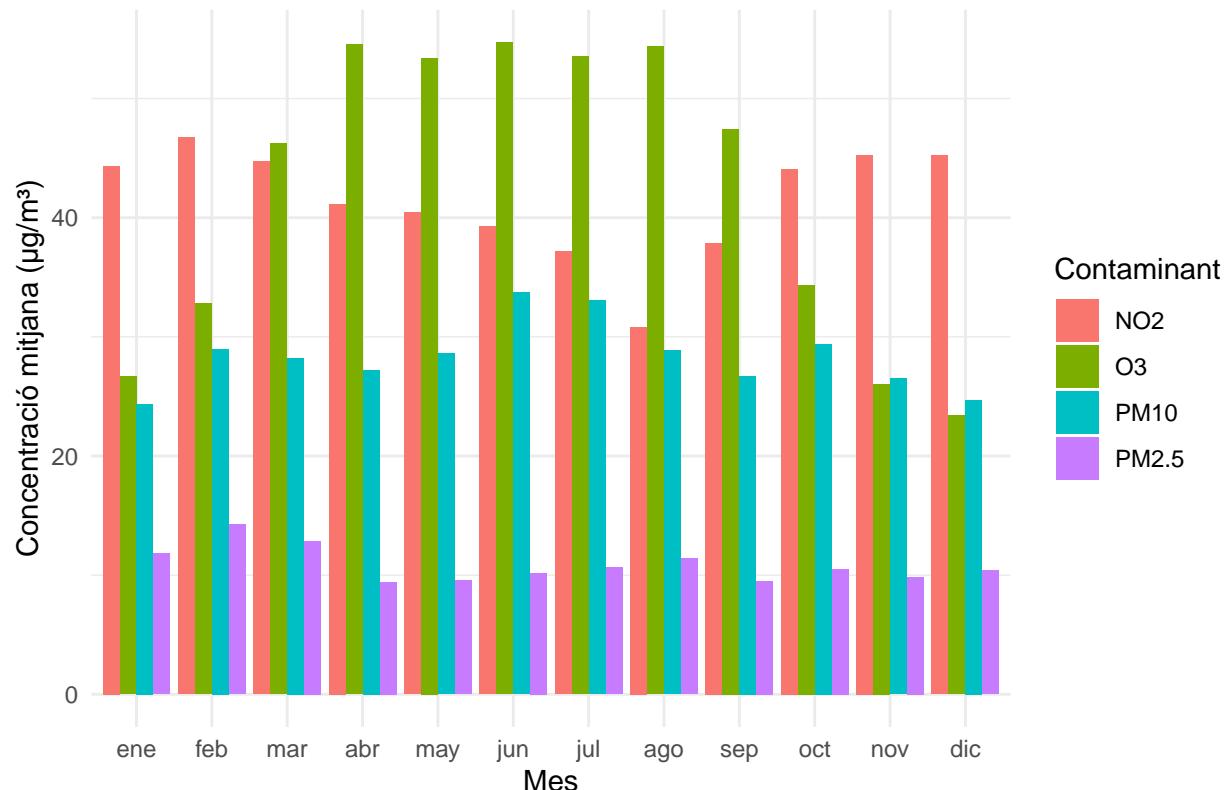
6.1 Anàlisi estacional (per mesos)

Calculem la mitjana mensual per cada contaminant per detectar patrons estacionals.

```
contaminants_mes <- contaminants_long %>%
  group_by(mes, contaminant) %>%
  summarise(valor_mitja = mean(valor, na.rm = TRUE), .groups = "drop")

ggplot(contaminants_mes, aes(x = mes, y = valor_mitja, fill = contaminant)) +
  geom_col(position = "dodge") +
  labs(title = "Mitjana mensual de contaminants",
       x = "Mes", y = "Concentració mitjana (µg/m³)",
       fill = "Contaminant") +
  theme_minimal()
```

Mitjana mensual de contaminants



La gràfica mostra la concentració mitjana mensual dels quatre contaminants (**NO₂**, **O₃**, **PM10** i **PM2.5**) al llarg de l'any, posant de manifest una clara estacionalitat: en general, els valors més alts es troben en els mesos de primavera-estiu (aproximadament de maig a agost/setembre) i els més baixos entre tardor i hivern. Aquest patró és especialment notori en el cas de l'**ozó (O₃)**, que augmenta amb la irradiació solar i les temperatures elevades, però també es pot observar que **NO₂**, **PM10** i **PM2.5** registren valors relativament elevats durant els mesos càlids. Això pot deure's, entre altres factors, a la menor dispersió atmosfèrica en períodes anticiclònics, a l'increment del trànsit o l'activitat turística a l'estiu, o a la reducció de la pluja, que dificulta el rentatge de l'aire. D'altra banda, la disminució de les concentracions a l'hivern s'explica, en part, per una major freqüència de pluja i vents, fet que afavoreix la dispersió i la deposició de contaminants. En conjunt, aquesta anàlisi mensual confirma la influència decisiva dels factors meteorològics i de l'activitat humana en la variació estacional de la contaminació a la ciutat.

6.2 Detecció de dies crítics (valors extrems)

Identifiquem aquells dies en què s'han registrat els valors màxims per cada contaminant.

```
# 1. Calcular els dies extrems per cada contaminant
dies_extrems <- contaminants_long %>%
  # Per cada dia i per cada contaminant, mantenim la fila amb el valor més alt
  group_by(data, contaminant) %>%
  slice_max(valor, n = 1, with_ties = FALSE) %>%
  ungroup() %>%
  # Per cada contaminant, triem les 5 observacions amb els valors més alts
  group_by(contaminant) %>%
  slice_max(valor, n = 5, with_ties = FALSE) %>%
  arrange(contaminant, desc(valor))
```

```

# 2. Agregar les dades meteorològiques diàries
meteo_diari <- df_meteo %>%
  mutate(data = as.Date(DATA_LECTURA)) %>%
  group_by(data, ACRÒNIM) %>%
  summarise(valor_meteo = mean(VALOR, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = ACRÒNIM, values_from = valor_meteo)

# 3. Unir les dades per data: afegim les variables meteorològiques als dies extrems
dies_extrems_meteo <- dies_extrems %>%
  left_join(meteo_diari, by = "data")

# 4. Filtrar només les columnes rellevants per a l'anàlisi
dies_extrems_meteo %>%
  select(data, contaminant, nom_estacio, valor, mes, any, TM, TN, TX, PPT, VVM10, VVX10) %>%
  knitr::kable()

```

data	contaminant	nom_estacio	valor	mes	any	TM	TN	TX	PPT	VVM10	VVX10
1995-11-08	NO2	Barcelona (St. Gervasi)	483	nov	1995	NA	NA	NA	NA	NA	NA
1992-06-26	NO2	Barcelona (Sagrera)	422	jun	1992	NA	NA	NA	NA	NA	NA
1994-12-14	NO2	Barcelona (St. Gervasi)	416	dic	1994	NA	NA	NA	NA	NA	NA
1993-03-19	NO2	Barcelona (St. Gervasi)	393	mar	1993	NA	NA	NA	NA	NA	NA
1995-10-09	NO2	Barcelona (St. Gervasi)	389	oct	1995	NA	NA	NA	NA	NA	NA
1991-06-15	O3	Barcelona (Poblenou)	271	jun	1991	NA	NA	NA	NA	NA	NA
1996-06-06	O3	Barcelona (Poblenou)	257	jun	1996	NA	NA	NA	NA	NA	NA
1997-09-04	O3	Barcelona (Poblenou)	254	sep	1997	NA	NA	NA	NA	NA	NA
1992-07-17	O3	Barcelona (St. Gervasi)	240	jul	1992	NA	NA	NA	NA	NA	NA
2020-07-10	O3	Barcelona (Palau Reial)	239	jul	2020	NA	NA	NA	NA	NA	NA
2001-06-24	PM10	Barcelona (Sants)	1608	jun	2001	NA	NA	NA	NA	NA	NA
2017-06-23	PM10	Barcelona (Poblenou)	1409	jun	2017	NA	NA	NA	NA	NA	NA
2017-06-24	PM10	Barcelona (Poblenou)	1263	jun	2017	NA	NA	NA	NA	NA	NA
2004-06-24	PM10	Barcelona (Sants)	1245	jun	2004	NA	NA	NA	NA	NA	NA
2007-06-24	PM10	Barcelona (Sants)	1185	jun	2007	NA	NA	NA	NA	NA	NA
2023-11-19	PM2.5	Barcelona (Palau Reial)	180	nov	2023	15.450	12.050	20.675	0.0000001	9000006.833333	
2004-02-21	PM2.5	Barcelona (Gràcia - Sant Gervasi)	132	feb	2004	NA	NA	NA	NA	NA	NA

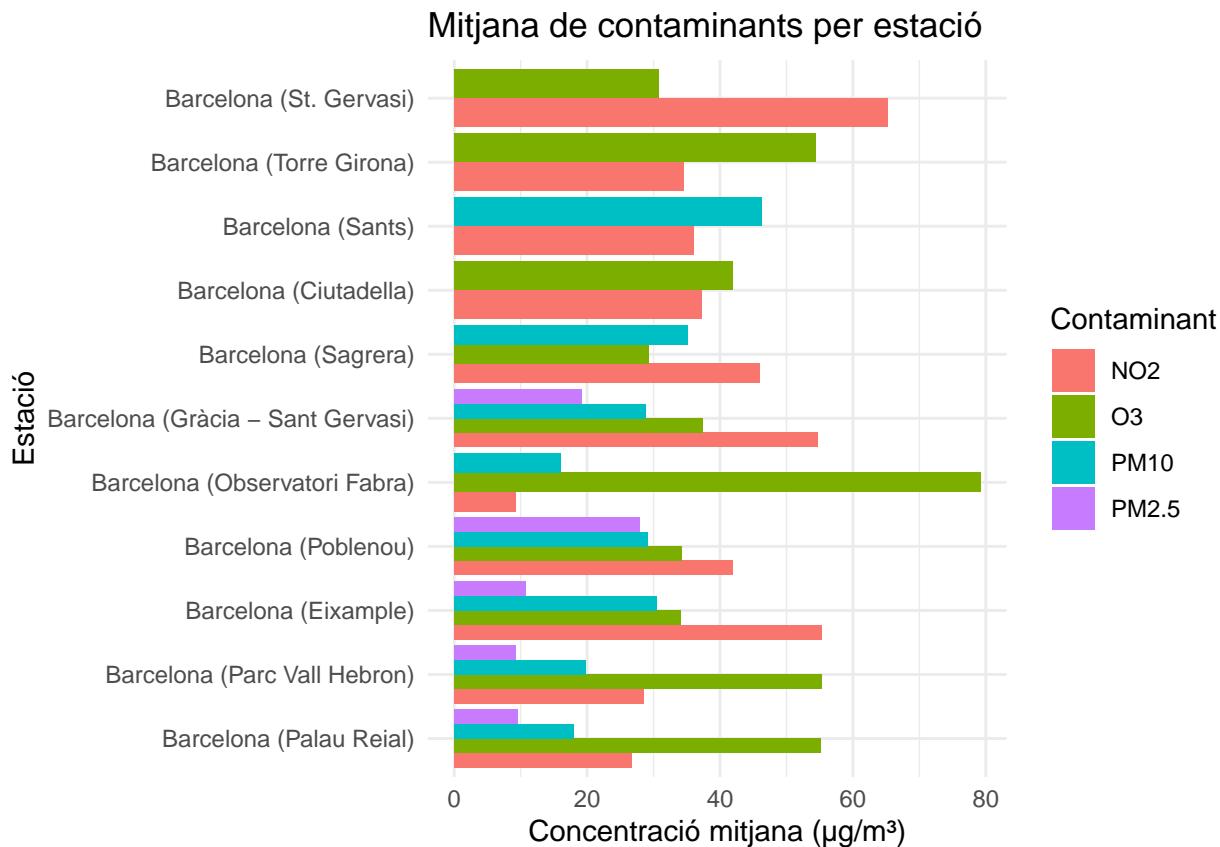
data	contaminant	nom_estacio	valor	mes	any	TM	TN	TX	PPT	VVM10	VVX10
2023-11-29	PM2.5	Barcelona (Palau Reial)	120	nov	2023	13.025	11.475	15.075	0.4333333	1.7000006	3.000000
2022-07-24	PM2.5	Barcelona (Palau Reial)	118	jul	2022	28.125	24.125	32.800	0.0000000	2.7666679	9.000000
2005-01-13	PM2.5	Barcelona (Gràcia - Sant Gervasi)	117	ene	2005	NA	NA	NA	NA	NA	NA

6.3 Resums per estació geogràfica

Resum per cadascuna de les estacions (per exemple, nom_estacio) dels valors mitjans i la variabilitat dels contaminants.

```
resum_estacio <- contaminants_long %>%
  group_by(nom_estacio, contaminant) %>%
  summarise(mitjana = mean(valor, na.rm = TRUE),
            desviacio = sd(valor, na.rm = TRUE), .groups = "drop")

ggplot(resum_estacio, aes(x = reorder(nom_estacio, mitjana), y = mitjana, fill = contaminant)) +
  geom_col(position = "dodge") +
  coord_flip() +
  labs(title = "Mitjana de contaminants per estació",
       x = "Estació", y = "Concentració mitjana (µg/m³)",
       fill = "Contaminant") +
  theme_minimal()
```



En aquesta gràfica s'hi representen les concentracions mitjanes dels quatre principals contaminants atmosfèrics (**NO₂**, **O₃**, **PM10** i **PM2.5**) per diverses estacions de mesurament a Barcelona. Alguns punts clau que s'hi poden observar:

1. Diferències entre estacions de trànsit i de fons urbà:

- Les estacions ubicades en zones de trànsit intens (com Eixample, Sants, etc.) presenten nivells de **NO₂** relativament més elevats. Aquesta dada reflecteix l'origen principalment vehicular d'aquest contaminant.
- Les estacions més allunyades del trànsit o situades en zones més altes (com Observatori Fabra) mostren menys acumulació de NO₂ i, en canvi, tendeixen a tenir valors més alts d'**O₃**, ja que hi ha menys titració de l'ozó per òxids de nitrogen.

2. O₃ més alt en estacions d'altitud o allunyades del trànsit:

- Observatori Fabra, per exemple, presenta una barra verda més gran, indicativa d'una concentració mitjana més alta d'ozó, consistent amb els patrons típics en llocs amb menys emissió directa de NO₂.

3. Distribució de PM10 i PM2.5:

- En gairebé totes les estacions, **PM10** supera **PM2.5** en valors mitjans, fet habitual atès que les partícules més grans (fins a 10 µm) soLEN provenir de múltiples fonts (trànsit, obres, pols en suspensió), mentre que les partícules fines (<2,5 µm) es concentren sobretot en processos de combustió i reaccions químiques secundàries.
- No obstant, les concentracions de **PM2.5** no són negligibles, i s'haurien de controlar especialment per l'impacte que tenen en la salut respiratòria.

4. Variabilitat entre barris:

- Algunes estacions situades en zones cèntriques o amb més densitat de població (com Sants, Eixample, Poblenou) tendeixen a registrar nivells més alts de NO₂ i PM10, a causa de la concentració de trànsit i activitat urbana.
- D'altres, com Palau Reial o Sant Gervasi, tot i que també registren NO₂, poden mostrar un perfil mixt (fons urbà / trànsit), reflectint la variabilitat urbana de Barcelona.

7. Conclusions generals

Aquest estudi exploratori ha permès obtenir una visió detallada i contextualitzada de l'evolució i el comportament dels principals contaminants atmosfèrics a la ciutat de Barcelona al llarg dels darrers anys. A partir de les dades horàries i diàries de qualitat de l'aire i variables meteorològiques, es poden extreure les conclusions següents:

1. Patrons de concentració i distribució:

Els contaminants **NO₂** i **PM10** mostren distribucions marcadament assimètriques, amb una gran quantitat de registres concentrats en valors moderats però amb presència d'episodis puntuals d'alta contaminació. En canvi, **O₃** i **PM2.5** presenten distribucions més equilibrades, tot i que també registren valors extrems en determinats períodes.

2. Dinàmica temporal i tendències:

S'ha observat una clara **estacionalitat** en la concentració de contaminants, especialment en el cas de l'ozó, que augmenta durant els mesos amb més irradiació solar. Per altra banda, les **mitjanes mòbils** per dècades permeten detectar **tendències decreixents** en NO₂ i PM10, possiblement atribuïbles a l'aplicació de polítiques ambientals, renovació del parc automobilístic o canvis en la mobilitat urbana.

3. Diferències entre tipus d'estacions:

Les **estacions de trànsit** registren sistemàticament nivells més elevats de contaminants primaris com NO₂, PM10 i PM2.5, mentre que les **estacions de fons urbà** mostren valors més alts d'ozó a causa del fenomen de titració i l'absència d'emissions directes de NO. Aquesta divergència posa de manifest la necessitat de tenir en compte la **tipologia d'estació** per a la interpretació adequada dels nivells de contaminació.

4. Relació amb les condicions meteorològiques:

Les **correlacions entre variables meteorològiques i contaminants** confirmen patrons coneguts: la **temperatura** i la **irradiació solar** es correlacionen positivament amb l'ozó, mentre que la **precipitació** i el **vent** tenen un efecte de dispersió o "rentatge" sobre la majoria de contaminants. Aquests resultats reforçen la importància de les condicions meteorològiques en la **dispersió o acumulació** de contaminants en l'àmbit urbà.

5. Detecció de dies crítics i estacionalitat:

La identificació dels **dies amb valors extrems** permet delimitar episodis puntuals de risc ambiental, molts dels quals coincideixen amb períodes d'estabilitat atmosfèrica, baixa ventilació o absència de precipitació. L'estudi **mensual** de les concentracions reforça la necessitat d'establir **estratègies de mitigació adaptades estacionalment**, especialment en els mesos de major risc.

6. Variabilitat geogràfica dins de la ciutat:

Les anàlisis per estació mostren una **variabilitat considerable entre barris**, amb concentracions més elevades en zones densament urbanitzades i de trànsit intens com l'Eixample o Sants. En canvi, estacions com l'Observatori Fabra o Sant Gervasi presenten perfils diferenciats, més propis de zones amb menys activitat emissora directa.