

```

---
title: "Anomalías de los volúmenes en embalses de España"
subtitle: "Actividad "
author:
  - name: Rubén Garrido Hidalgo
    email: rubengh2002@gmail.com
format: html
editor: source
---

**Objetivo**: Mostrar a nivel nacional y por cuenca el número de embalses
con sus respectivos anomalías del volumen de agua embalsada en percentiles.

"! Could not find tools necessary to compile a package" -\> https://cran.r-project.org/bin/windows/Rtools/rtools44/rtools.html

## Paquetes

```{r}
#| message: false
#| warning: false

# instalamos los paquetes si hace falta
if(!require("tidyverse")) install.packages("tidyverse")
if(!require("janitor")) install.packages("janitor")
if(!require("patchwork")) install.packages("patchwork")
if(!require("RODBC")) install.packages("RODBC")

# paquetes
library(tidyverse)
library(readxl)
library(janitor)
library(patchwork)
library(RODBC)
library(rmarkdown)

```

## Descarga e importación de datos

```{r}
# URL base de datos
url <- "https://www.miteco.gob.es/content/dam/miteco/es/agua/temas/
evaluacion-de-los-recursos-hidricos/boletin-hidrologico/Historico-de-
embalses/BD-Embalses.zip"

# descarga
tempf <- tempfile() # archivo temporal
download.file(url, tempf)
unzip(tempf) # descomprimos

# abrimos conexión con la mdb
conn <- odbcConnectAccess2007("BD-Embalses.mdb")

# leemos las tablas disponibles
subset(sqlTables(conn), TABLE_TYPE == "TABLE")

# importamos la tabla

```

```

emb <- sqlFetch(conn, "T_Datos Embalses 1988-2025")
```

## Preparación de los datos

```{r}
# limpiamos los nombres de columnas + convertimos en número columnas con
nombre agua
# + convertimos la fecha en clase Date
emb <- clean_names(emb) %>%
  mutate(across(starts_with("agua"),
    ~ parse_number(., locale = locale(decimal_mark = ","))),
    fecha = ymd(fecha))
# data structure
str(emb) # datos semanales
```

```{r}
# conversión del agua embalsada en percentiles
ecdf2 <- function(d) ecdf(d)(d)

# ejemplo
ecdf2(rnorm(100, 15, 20)) * 100

# nueva columna agua embalsada en percentiles
emb <- group_by(emb, embalse_nombre) %>%
  mutate(agua_perc = ecdf2(agua_actual))
head(emb)
```

## Nivel nacional

```{r}
# definiciones clases
percentile_breaks = c(0, 0.05, 0.1, 0.25, .4, .6, .75, 0.9, 0.95, 1)

# añadimos el año hidrológico Oct - Sep
emb <- mutate(emb, hy = ifelse(month(fecha) > 9, year(fecha)+1,
year(fecha)))

# filtramos el año actual
res_nacional <- filter(emb, hy == 2024) %>%
  mutate(agua_cat = cut(agua_perc,
    percentile_breaks,
    include.lowest = TRUE)) %>%
  group_by(fecha, .drop = FALSE) %>%
  count(agua_cat)

head(res_nacional, 10)

# clases de agua
lev_agua <- levels(res_nacional$agua_cat)

# complementamos el conteo de 0
# calculamos el número de embalses en %
res_nacional <- mutate(res_nacional,
  rel = n/sum(n),
  agua_cat = factor(agua_cat, lev_agua))

```

```
head(res_nacional, 10)
```
```

¿Cuántos embalses son más seco de lo normal?

```
```{r}
ymin_actual <- ungroup(res_nacional) %>%
  filter(fecha == max(fecha)) %>%
  arrange(agua_cat) %>%
  pull(rel) %>%
  .[1:4] %>%
  sum()

ymin_actual * 100
```
```

### A nivel nacional

```
```{r}
#| fig-height: 6
g0 <- ggplot(res_nacional,
  aes(fecha, rel, fill = agua_cat)) +
  geom_area()

g1 <- g0 + annotate("linrange",
  x = max(res_nacional$fecha) + days(4),
  linewidth = .6,
  ymax = 1,
  ymin = 1-ymin_actual) +
  annotate("text",
    x = max(res_nacional$fecha) + days(25),
    y = .95,
    label = scales::percent(ymin_actual, 1)) +
  annotate("text",
    x = ymd("2024-02-10"),
    y = .52,
    label = "normal",
    fontface = "bold")

g2 <- g1 + scale_fill_brewer(palette = "BrBG",
  labels = ~ scales::number(., scale = 100),
  guide = guide_colorsteps()) +
  scale_x_date(breaks = seq(ymd("2023-11-01"), ymd("2024-10-01"),
"2 month"),
    date_labels = "%b %y") +
  scale_y_continuous(breaks = c(0, .25, .5, .75, 1),
    labels = scales::label_percent()) +
  labs(x = NULL,
    y = "Proporción de embalses",
    fill = "Percentil",
    title = "Nacional",
    tag = str_wrap("más seco de lo normal", 10)) +
  coord_cartesian(expand = FALSE,
    clip = "off") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
    legend.margin = margin(),
```

```

        legend.ticks = element_line(color = "black",
linewidth = .8),
        legend.title.position = "top",
        legend.title = element_text(hjust = 0),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.justification = .1,
        legend.key.height = unit(0.3, "lines"),
        legend.key.width = unit(3, "lines"),
        plot.tag.position = c(1,.9),
        plot.tag = element_text(size = 8),
        axis.ticks.length = unit(.5, "mm"),
        plot.title = element_text(size = 14, vjust = -3,
                                margin = margin()),
        axis.ticks = element_line(),
        plot.margin = margin(10, 10, 20, 10),
        aspect.ratio = 1)

g2
```

## Nivel cuencas hidrográficas

```{r}
# agrupar por cuenca del año 2024
res_cuenca <- filter(emb, hy == 2024) %>%
  mutate(agua_cat = cut(agua_perc,
                        percentile_breaks,
                        include.lowest = TRUE)) %>%
  group_by(ambito_nombre, fecha, .drop = FALSE) %>%
  count(agua_cat)

# complementar todas las combinaciones de conteo
res_cuenca <- mutate(res_cuenca,
  rel = n/sum(n),
  agua_cat = factor(agua_cat, lev_agua),
  nembalses = sum(n))

# excluimos cuencas con menos de 10 embalses
res_cuenca_sel <- filter(res_cuenca, nembalses >= 10)

# fijamos el orden
ord <- unique(res_cuenca_sel$ambito_nombre)
ord <- ord[c(4, 8, 2:3, 10, 6:7, 5, 9, 1)]

res_cuenca_sel <- mutate(res_cuenca_sel,
  ambito_nombre = factor(ambito_nombre, ord))
```

```{r}
#| fig-height: 6
#| fig-width: 13

# parte basica
p0 <- ggplot(res_cuenca_sel,
  aes(fecha, rel,
      fill = agua_cat)) +
  geom_area()

```

```
# añadimos ajustes en scales
p1 <- p0 + scale_fill_brewer(palette = "BrBG", guide = guide_colorsteps()) +
  scale_x_date(breaks = seq(ymd("2023-11-01"), ymd("2024-10-01"),
    "3 month"),
    date_labels = "%b",
    expand = expansion(0)) +
  scale_y_continuous(breaks = c(0, .25, .5, .75, 1),
    labels = ~scales::number(., scale = 100),
    expand = expansion()) +
  facet_wrap(ambito_nombre~., ncol = 5,
    scales = "free_x",
    labeller = labeller(ambito_nombre =
label_wrap_gen(width = 20,
    multi_line =
TRUE)))

# ajustes de estilo
p2 <- p1 + labs(x = NULL, y = NULL, fill = NULL) +
  coord_cartesian(clip = "off") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
    panel.spacing.x = unit(1, "lines"),
    legend.position = "none",
    axis.text.y = element_blank(),
    axis.ticks.length = unit(.5, "mm"),
    strip.text = element_text(face = "bold"),
    strip.clip = "off",
    axis.ticks = element_line(),
    aspect.ratio = 1)

p2 # pequeños multiples
```


```
```{r}
#| fig-height: 6
#| fig-width: 12

# fecha ultima actualización
update_date <- format(max(res_nacional$fecha), "%d %B %Y")

# patrón de combinación
pattern <- 'AB'

# construcción final
final <- wrap_plots(A = g2, B = p2, design = pattern) +
  plot_annotation(title = "EMBALSES 2023-2024", # titulo global
    subtitle = str_glue("Actualizado el {update_date}"),
    caption = "Dominic Royé (@dr_xeo) | Datos: MITECO",
    theme = theme(plot.title = element_text(size = 25,
    hjust = .5,
    colour =
"#01665e",
    face = "bold",
    margin =
margin(10, 5,
```


```

```

5, 5)),

                                plot.subtitle = element_text(hjust = 0,
  vjust = 24,
  size = 8),
                                plot.caption = element_text(hjust = 0, size
= 8)))

final
```

```{r}
# exportamos
ggsave("embasles.png", final,
        height = 6, width = 12,
        units = "in",
        bg = "white")
```

## 1) Gráfico del año hidrologico 1994-1995

```{r}
#| message: false
#| warning: false

# instalamos los paquetes si hace falta
if(!require("tidyverse")) install.packages("tidyverse")
if(!require("janitor")) install.packages("janitor")
if(!require("patchwork")) install.packages("patchwork")
if(!require("RODBC")) install.packages("RODBC")

# paquetes
library(tidyverse)
library(readxl)
library(janitor)
library(patchwork)
library(RODBC)
```

## Descargamos los datos

```{r}
# URL base de datos
url <- "https://www.miteco.gob.es/content/dam/miteco/es/agua/temas/
evaluacion-de-los-recursos-hidricos/boletin-hidrologico/Historico-de-
embalses/BD-Embalses.zip"

# descarga
tempf <- tempfile() # archivo temporal
download.file(url, tempf)
unzip(tempf) # descomprimos

# abrimos conexión con la mdb
conn <- odbcConnectAccess2007("BD-Embalses.mdb")

# leemos las tablas disponibles

```

```

subset(sqlTables(conn), TABLE_TYPE == "TABLE")

# importamos la tabla
emb <- sqlFetch(conn, "T_Datos Embalses 1988-2025")
```

## Preparamos los datos:

```{r}
# Limpiamos los nombres de columnas, convertimos en numérico las columnas
con nombre agua y convertimos la fecha en clase Date

if (is.null(emb) || nrow(emb) == 0) {
  stop("La tabla no contiene datos o no se cargó correctamente.")
}

emb <- clean_names(emb) %>%
  mutate(across(starts_with("agua"),
    ~ parse_number(., locale = locale(decimal_mark = ","))),
    fecha = ymd(fecha))

# data structure
str(emb) # datos semanales
```

## Vamos a proceder a convertir el agua embalsada en percentiles y a crear
una nueva columna con esos datos

```{r}

# Convertimos el agua embalsada en percentiles
ecdf2 <- function(d) ecdf(d)(d)

# Ejemplo
ecdf2(rnorm(100, 15, 20)) * 100

# Creamos una nueva columna de agua embalsada en Percentiles
emb <- group_by(emb, embalse_nombre) %>%
  mutate(agua_perc = ecdf2(agua_actual))
head(emb)
```

## A continuación vamos a definir las distintas clases del eje y y el año
hidrológico con el que vamos a trabajar

```{r}
# Definimos clases
percentile_breaks = c(0, 0.05, 0.1, 0.25, .4, .6, .75, 0.9, 0.95, 1)

# Añadimos el año hidrológico Oct - Sep
emb <- mutate(emb, hy = ifelse(month(fecha) > 9, year(fecha)+1,
year(fecha)))

# Filtramos el año 1994
res_nacional <- filter(emb, hy == 1995) %>%
  mutate(agua_cat = cut(agua_perc,
    percentile_breaks,
    include.lowest = TRUE)) %>%
  group_by(fecha, .drop = FALSE) %>%

```

```

count(agua_cat)

head(res_nacional, 10)

# clases de agua
lev_agua <- levels(res_nacional$agua_cat)

...

## Calculamos el porcentaje de número de embalses:

```{r}
# calculamos el número de embalses en %
res_nacional <- mutate(res_nacional,
                        rel = n/sum(n),
                        agua_cat = factor(agua_cat, lev_agua))

head(res_nacional, 10)

ymin_actual <- ungroup(res_nacional) %>%
  filter(fecha == max(fecha)) %>%
  arrange(agua_cat) %>%
  pull(rel) %>%
  .[1:4] %>%
  sum()
ymin_actual * 100

...

## Representación Gráfica:

```{r}

g0 <- ggplot(res_nacional,
             aes(fecha, rel, fill = agua_cat)) +
  geom_area()

g1 <- g0 + annotate("linetype",
                  x = max(res_nacional$fecha) + days(4),
                  linewidth = .6,
                  ymax = 1,
                  ymin = 1-ymin_actual) +
  annotate("text",
          x = max(res_nacional$fecha) + days(25),
          y = .95,
          label = scales::percent(ymin_actual, 1)) +
  annotate("text",
          x = ymd("1994-10-04"),
          y = .6,
          label = "normal",
          fontface = "bold",
          size = 4) #Añado el comando size para ajustar el tamaño de la
palabra normal

```



```

g2 <- g1 + scale_fill_brewer(palette = "BrBG",
                             labels = ~ scales::number(., scale = 100),
                             guide = guide_colorsteps()) +
  scale_x_date(breaks = seq(ymd("1994-10-04"), ymd("1995-09-26"), "2
month"),
               date_labels = "%b %y") +
  scale_y_continuous(breaks = c(0, .25, .5, .75, 1),
                     labels = scales::label_percent()) +
  labs(x = NULL,
       y = "Proporción de embalses",
       fill = "Percentil",
       title = "Nacional",
       tag = str_wrap("más seco de lo normal", 5)) +
  coord_cartesian(expand = FALSE,
                  clip = "off") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        legend.margin = margin(),
        legend.ticks = element_line(color = "black", linewidth = .8),
        legend.title.position = "top",
        legend.title = element_text(hjust = 0),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.justification = .1,
        legend.key.height = unit(0.3, "lines"),
        legend.key.width = unit(3, "lines"),
        plot.tag.position = c(1, .9),
        plot.tag = element_text(size = 8),
        axis.ticks.length = unit(.5, "mm"),
        plot.title = element_text(size = 14, vjust = -3,
                                   margin = margin()),
        axis.ticks = element_line(),
        plot.margin = margin(10, 10, 20, 10),
        aspect.ratio = 1)

```

g2

```

Importamos esta representación a un archivo png:

```{r}

```

ggsave("año_hidrológico.png", g2,
       height = 6, width = 12,
       units = "in",
       bg = "white")

```

```

## 2) Múltiples pequeños a nivel nacional de toda la serie temporal

## Agrupamos por cuenca del año hidrológico 1994-1995

```{r}

```

res_cuenca <- filter(emb, hy == 1995) %>%
  mutate(agua_cat = cut(agua_perc,
                        percentile_breaks,
                        include.lowest = TRUE)) %>%
  group_by(ambito_nombre, fecha, .drop = FALSE) %>%
  count(agua_cat)

```

```

```{r}

## Completamos la combinaciones del conteo y excluimos las cuencas con
menos de diez embalses

```{r}
# Completamos el conteo
res_cuenca <- mutate(res_cuenca,
                      rel = n/sum(n),
                      agua_cat = factor(agua_cat, lev_agua),
                      nembalses = sum(n))

# Exclusión de cuencas con menos de diez embalses
res_cuenca_sel <- filter(res_cuenca, nembalses >= 10)

```{r}

## Fijamos el orden para la subdivisión en pequeños múltiples

```{r}
ord <- unique(res_cuenca_sel$ambito_nombre)
ord <- ord[c(4, 8, 2:3, 10, 6:7, 5, 9, 1)]

res_cuenca_sel <- mutate(res_cuenca_sel,
                        ambito_nombre = factor(ambito_nombre, ord))

#| fig-height: 6
#| fig-width: 13

```{r}

## Una vez que tenemos todos los datos necesarios vamos a representar los
datos:

```{r}
p0 <- ggplot(res_cuenca_sel,
             aes(fecha, rel,
                 fill = agua_cat)) +
  geom_area()

# añadimos ajustes en scales
p1 <- p0 + scale_fill_brewer(palette = "BrBG", guide = guide_colorsteps())
+
  scale_x_date(breaks = seq(ymd("1994-10-04"), ymd("1995-09-26"), "3
month"),
              date_labels = "%b",
              expand = expansion(0)) +
  scale_y_continuous(breaks = c(0, .25, .5, .75, 1),
                    labels = ~scales::number(., scale = 100),
                    expand = expansion()) +
  facet_wrap(ambito_nombre~., ncol = 5,
            scales = "free_x",
            labeller = labeller(ambito_nombre = label_wrap_gen(width = 20,
  multi_line
= TRUE))))

# ajustes de estilo

```

```

p2 <- p1 + labs(x = NULL, y = NULL, fill = NULL) +
  coord_cartesian(clip = "off") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.spacing.x = unit(1, "lines"),
        legend.position = "none",
        axis.text.y = element_blank(),
        axis.ticks.length = unit(.5, "mm"),
        strip.text = element_text(face = "bold"),
        strip.clip = "off",
        axis.ticks = element_line(),
        aspect.ratio = 1)

p2 # pequeños multiples del año hidrológico 1994-1995

```
Importamos los pequeños múltiples del año hidrológico:
```{r}
ggsave("pequeños_multiples.png", p2,
       height = 6, width = 12,
       units = "in",
       bg = "white")
```

## BONOS. Gráfico polar a nivel nacional

## Cargamos las librerías necesarias

```{r}

# Cargar librerías necesarias
library(tidyverse)
library(janitor)
library(lubridate)
library(RODBC)
```

## Filtramos los datos del año hidrológico 1994-1995 y calculamos la suma
total de agua embalsada:

```{r}
datos_nacionales <- emb %>%
  mutate(hy = ifelse(month(fecha) > 9, year(fecha) + 1, year(fecha))) %>% #
Año hidrológico
  filter(hy == 1995) %>%
  group_by(month = month(fecha, label = TRUE)) %>% # Agrupamos por mes
  summarise(total_embalsada = sum(agua_actual, na.rm = TRUE))
```

## Calculamos proporciones nacionales hidrológicas y ángulos para el
gráfico polar:

```{r}
datos_nacionales <- datos_nacionales %>%
  mutate(proporcion = total_embalsada / sum(total_embalsada), #Proporción
         nacional
         angle = 2 * pi * as.numeric(month) / 12) # Conversión de meses a
radianes
```

```

## Una vez que tenemos los datos necesarios para nuestra representación gráfica procedemos a realizar el gráfico polar del año hidrológico:

```
```{r}
g_polar <- ggplot(datos_nacionales, aes(x = month, y = proporcion, fill =
month)) +
  geom_bar(stat = "identity", width = 1, color = "white") + #Barras
apiladas
  coord_polar(start = 0) + # Gráfico polar, empieza 0 par que así podamos
empezar a leerlo en el sentido de las agujas edl reloj desde enero
  scale_fill_brewer(palette = "Set3") +
  labs(title = "Distribución Nacional - Año Hidrológico 1994-1995",
        x = NULL, y = NULL, fill = "Mes",
        caption = "Fuente: Datos del MITECO") +
  theme_minimal() +
  theme(axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        panel.grid = element_blank(),
        legend.position = "bottom")

# Visualizar BrBG
print(g_polar)

```
```

Importamos este gráfico hidrológico polar:

```
```{r}
ggsave("gráfico_polar.png", g2,
        height = 6, width = 12,
        units = "in",
        bg = "white")
```
```

## He cambiado la paleta de colores empleada en clase y en los apartados anteriores "BrBG", porque no se adaptaba bien a los datos y poseía pocos colores para todos los meses del año hidrológico. He encontrado la paleta de colores "Set3" la cual hace la visualización mucho más clara y sencilla pues emplea colores más vivos con tonos más diferenciados que aquellos empleados en la paleta de colores "BrBG".