

Ejercicios_VisualizacionDatos_Ainhoa

Ainhoa Calleja Rodriguez

2024-12-07

Cargas paquetes

Como este Rmd es un archivo común para todos los ejercicios, lo primero que hago es cargar los paquetes necesarios para realizarlos. En mi caso suelo cargarlos con una función, ya que debido a mis necesidades suelo tener que cargar o instalar numerosos paquetes.

```
# Packages needed
packages <- c("ggplot2", "dplyr", "tidyr", "yaml", "readxl", "ggradar", "tibble", "scales", "grid")

# Function to install (if needed) and load packages
install_and_load <- function(pkg){
  if (!require(pkg, character.only = TRUE)) {
    install.packages(pkg, dependencies = TRUE)
    library(pkg, character.only = TRUE)
  }
}

# Apply the function to each package
lapply(packages, install_and_load)
```

Para la carga de los archivos de datos necesarios, mi directorio principal está incluido en la parte yaml del Rmd que contiene lo siguiente:

```
# Mostrar el YAML completo
yaml_content <- rmarkdown::yaml_front_matter(knitr::current_input())
yaml_raw <- yaml::as.yaml(yaml_content)
cat("Contenido del YAML:\n", yaml_raw)
```

```
## Contenido del YAML:
## title: Ejercicios_VisualizacionDatos_Ainhoa
## author: Ainhoa Calleja Rodriguez
## date: '2024-12-07'
## output:
## pdf_document:
## fig_height: 6
## fig_width: 8
## main_dir: /Users/ainhoacalleja/Documents/cursos/GEODATA_UVA/1_Gestion_Visualizacion_Datos/Tema3_Visualizacion
```

De esta forma sólo tengo que indicar mi directorio principal y asignar el archivo de datos al nombre que voy a utilizar para cargarlos. He decidido hacer los ejercicios 3, 4, 6 y 8 por tanto los archivos necesarios son los siguientes:

```
# Leer el directorio desde el YAML
main_dir <- rmarkdown::yaml_front_matter(knitr::current_input())$main_dir
file1 <- file.path(main_dir, "Quercus_raiz.csv") #Ejercicio 3
file2 <- file.path(main_dir, "Try2023713193415 traits of 34 hebeaceous species in Sherbrooke.xlsx") #Ejercicio 4
file3 <- file.path(main_dir, "Gasteropodos_ambiente.csv") #Ejercicio 6
file4 <- file.path(main_dir, "Polinizadores.csv") #Ejercicio 8
```

Ejercicios

Ejercicio 3: Producción roble

En un ensayo sobre producción de robles (*Quercus spp.*) en vivero se ha evaluado el crecimiento de las plántulas de diferentes especies en diferentes sustratos comerciales (Quercus_raiz.csv). Se te pide elaborar un gráfico que resuma el estado de las plántulas en función del sustrato señalando la distribución de datos por cuartiles y la existencia de datos anómalos en su caso. El gráfico debe tener calidad suficiente para ser incluido en una publicación divulgativa del sector. Adjuntar el script utilizado.

1. Carga de los datos:

```
quercus <- read.csv2(file1, header = TRUE)
str(quercus)
```

```
## 'data.frame': 279 obs. of 4 variables:
## $ Especie : chr "Quercus.petraea" "Quercus.robur" "Quercus.robur" "Quercus.robur" ...
## $ Longitud_tallo: num 16 34.5 20.5 49 23 31 49.2 40 5 31 ...
## $ Longitud_raiz : int 31 57 33 33 31 50 35 51 23 58 ...
## $ Sustrato : chr "TIPO A" "TIPO A" "TIPO A" "TIPO A" ...
```

```
summary(quercus)
```

```
## Especie      Longitud_tallo Longitud_raiz      Sustrato
## Length:279   Min. : 5.0    Min. :10.00   Length:279
## Class :character 1st Qu.:15.0   1st Qu.:25.00   Class :character
## Mode :character  Median :25.9   Median :36.00   Mode :character
##              Mean :26.3    Mean :34.99
##              3rd Qu.:37.0   3rd Qu.:43.00
##              Max. :50.0    Max. :60.00
```

2. Modificaciones.

Por preferencias personales sustituyo el punto que aparece en los nombres científicos de las especies en la columna especie por un guión bajo “_”.

```
quercus$Especie <- gsub("\\.", "_", quercus$Especie)
head(quercus)
```

```
##           Especie Longitud_tallo Longitud_raiz Sustrato
## 1 Quercus_petraea          16.0           31 TIPO A
## 2 Quercus_robur           34.5           57 TIPO A
## 3 Quercus_robur           20.5           33 TIPO A
## 4 Quercus_robur           49.0           33 TIPO A
## 5 Quercus_robur           23.0           31 TIPO A
## 6 Quercus_robur           31.0           50 TIPO A
```

3. Gráfico.

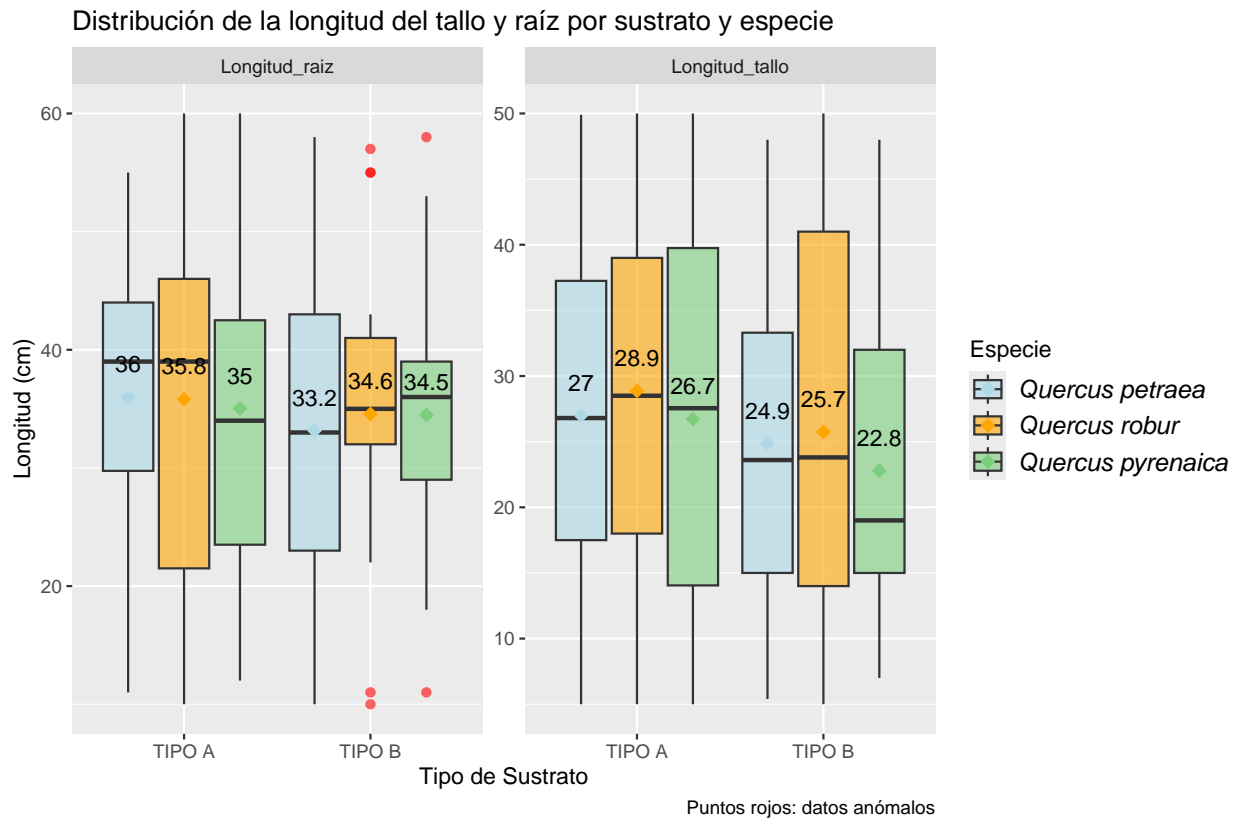
Modifico los datos para darles el formato necesario para incluir tanto la longitud de la raíz como la del tallo en el gráfico. El gráfico elegido ha sido un boxplot para incluir cuartiles así como medias por especie. Además en el gráfico he añadido un comando para destacar en punto de color rojo aquellos datos que puedan ser anómalos. Añado la media a cada boxplot para especie y longitud del tallo o raíz, con un punto del mismo color que la especie así como el valor de la media en texto.

```
# Crear una columna para el tipo de medida (Tallo vs Raíz)
quercus_longitud <- quercus %>%
  pivot_longer(cols = c(Longitud_tallo, Longitud_raiz),
    names_to = "Medida",
    values_to = "Valor")

quercus_plot <- ggplot(quercus_longitud, aes(x = Sustrato, y = Valor, fill = Especie)) +
  geom_boxplot(outlier.color = "red", outlier.shape = 16, outlier.size = 2, alpha = 0.6,
    width = 0.9) +
  stat_summary(fun = mean, geom = "point", aes(color = Especie), shape = 18, size = 3,
    position = position_dodge(width = 0.9)) +
  stat_summary(fun = mean, geom = "text", aes(label = round(after_stat(y), 1)),
    color = "black", vjust = -1.5, position = position_dodge(width = 0.9)) +
  labs(
    title = "Distribución de la longitud del tallo y raíz por sustrato y especie",
    x = "Tipo de Sustrato",
    y = "Longitud (cm)",
    caption = "Puntos rojos: datos anómalos"
  ) +
  scale_color_manual(
    values = c("Quercus_petraea" = "lightblue", "Quercus_robur" = "palegreen3",
      "Quercus_pyrenaica" = "orange"),
    labels = gsub("_", " ", c("Quercus_petraea", "Quercus_robur", "Quercus_pyrenaica"))
  ) +
  scale_fill_manual(
    values = c("Quercus_petraea" = "lightblue", "Quercus_robur" = "palegreen3",
      "Quercus_pyrenaica" = "orange"),
    labels = gsub("_", " ", c("Quercus_petraea", "Quercus_robur", "Quercus_pyrenaica"))
  ) +
  facet_wrap(~Medida, scales = "free_y") +
  theme(
    legend.position = "right",
```

```
legend.text = element_text(size = 12, face = "italic")
)
```

`quercus_plot` *#para ver el gráfico al compilar el pdf*



```
# Ruta y nombre del archivo a guardar
output_file <- file.path(main_dir, "3_grafico_quercus.png")

ggsave(
  filename = output_file,
  plot = quercus_plot,
  dpi = 300,
  width = 10,
  height = 8,
  units = "in"
)
```

El resultado es un gráfico que nos indica que la longitud del tallo es mayor con el sustrato tipo A y para el tallo no existen datos anómalos. En el caso de la longitud de la raíz es mayor para el sustrato tipo A aunque la diferencia no es tan destacable para *Quercus robur* y *Quercus pyrenaica*, aunque en estos dos últimos casos existen algunos datos anómalos que podrían haber influido en el resultado.

Ejercicio 4: Comparativa fisiológica y morfológica a nivel foliar

En un estudio de ecología vegetal se quiere realizar una comparativa fisiológica y morfológica a nivel foliar de tres especies: *Dactylis glomerata*, *Festuca rubra* y *Erigeron canadensis* (Yuanzhi Li, 2015)*. Se pide un gráfico radar de calidad publicable en una revista científica internacional en lengua inglesa mostrando la descripción a escala foliar de los tres taxones. Adjuntar el script utilizado.

1. Carga de los datos:

```
Try_dat <- read_excel(file2, sheet = "trait data")
```

```
head(Try_dat)
```

```
## # A tibble: 6 x 16
##   Species      VH      TB      LT      LA      LDMC      SLA      LCC      LNC      MPR      SSD      SDMC
##   <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Acalypha r~ 0.036   28  0.145  62.8  29.8  26.1  402.  25.3  2.83  208.  0.074
## 2 Agrostis g~ 0.182  610.  0.18  665.  23.1  36.7  404.  16.3  6.43  233.  0.086
## 3 Centaurea ~ 0.36  1146.  0.382  577.  16.7  25.9  424.  13.8  6.49  286.  0.155
## 4 Cerastium ~ 0.138  184.  0.334  130.  23.6  17.3  411.  19.2  6.40  499.  0.059
## 5 Chenopodiu~ 0.138  196.  0.326  251.  16.4  34.9  389.  21.3  3.88  170.  0.033
## 6 Cichorium ~ 0.103 1099.  0.242 1627.  17.3  37.6  392.  16.7  5.73  146.  0.049
## # i 4 more variables: SRL <dbl>, RB <dbl>, SM <dbl>, SS <dbl>
```

```
str(Try_dat)
```

```
## tibble [34 x 16] (S3: tbl_df/tbl/data.frame)
## $ Species: chr [1:34] "Acalypha rhomboidea" "Agrostis gigantea" "Centaurea cyanus" "Cerastium tomen
## $ VH      : num [1:34] 0.036 0.182 0.36 0.138 0.138 0.103 0.173 0.088 0.158 0.061 ...
## $ TB      : num [1:34] 28 610 1146 184 196 ...
## $ LT      : num [1:34] 0.145 0.18 0.382 0.334 0.326 0.242 0.16 0.185 0.246 0.454 ...
## $ LA      : num [1:34] 62.8 665.1 577.3 130.5 250.6 ...
## $ LDMC    : num [1:34] 29.8 23.1 16.7 23.6 16.4 ...
## $ SLA     : num [1:34] 26.1 36.7 25.9 17.3 34.9 ...
## $ LCC     : num [1:34] 402 404 424 411 389 ...
## $ LNC     : num [1:34] 25.3 16.3 13.8 19.2 21.3 ...
## $ MPR     : num [1:34] 2.83 6.43 6.49 6.39 3.88 ...
## $ SSD     : num [1:34] 208 233 286 499 170 ...
## $ SDMC    : num [1:34] 0.074 0.086 0.155 0.059 0.033 0.049 0.063 0.043 0.036 0.129 ...
## $ SRL     : num [1:34] 188 223 107 327 307 ...
## $ RB      : num [1:34] 3.8 173.9 344.2 68.9 54 ...
## $ SM      : num [1:34] 0.72 0.115 4.135 0.385 0.675 ...
## $ SS      : num [1:34] 0.62 0.791 2.539 0.477 0.549 ...
```

```
summary(Try_dat)
```

```
##   Species      VH      TB      LT
## Length:34      Min.    :0.02400      Min.    : 28.0      Min.    :0.1000
## Class :character 1st Qu.:0.07375      1st Qu.: 161.5      1st Qu.:0.1807
## Mode  :character Median :0.12500      Median : 237.5      Median :0.2400
```

```
##           Mean   :0.13332   Mean   : 466.6   Mean   :0.2417
##           3rd Qu.:0.16100   3rd Qu.: 588.4   3rd Qu.:0.2858
##           Max.    :0.36000   Max.    :3744.0   Max.    :0.4740
##           LA           LDMC           SLA           LCC
##   Min.    : 52.92   Min.    :12.11   Min.    :11.19   Min.    :383.2
##   1st Qu.: 154.06   1st Qu.:19.71   1st Qu.:22.90   1st Qu.:402.3
##   Median : 263.40   Median :23.53   Median :29.04   Median :414.5
##   Mean    : 456.11   Mean    :23.86   Mean    :28.98   Mean    :415.2
##   3rd Qu.: 561.00   3rd Qu.:28.03   3rd Qu.:33.10   3rd Qu.:423.0
##   Max.    :3066.08   Max.    :42.34   Max.    :58.17   Max.    :472.5
##           LNC           MPR           SSD           SDMC
##   Min.    :10.62   Min.    : 2.828   Min.    :126.5   Min.    :0.01600
##   1st Qu.:15.92   1st Qu.: 4.414   1st Qu.:171.8   1st Qu.:0.04475
##   Median :18.91   Median : 5.681   Median :224.1   Median :0.05950
##   Mean    :22.33   Mean    : 6.380   Mean    :225.5   Mean    :0.06865
##   3rd Qu.:28.18   3rd Qu.: 6.420   3rd Qu.:251.3   3rd Qu.:0.08550
##   Max.    :41.91   Max.    :22.096   Max.    :499.3   Max.    :0.15800
##           SRL           RB           SM           SS
##   Min.    : 35.35   Min.    : 3.80   Min.    : 0.0350   Min.    :0.3700
##   1st Qu.:108.53   1st Qu.: 55.29   1st Qu.: 0.2112   1st Qu.:0.5555
##   Median :186.74   Median : 99.13   Median : 0.5925   Median :0.7950
##   Mean    :187.27   Mean    :182.96   Mean    : 1.1869   Mean    :0.9798
##   3rd Qu.:223.64   3rd Qu.:220.93   3rd Qu.: 0.9962   3rd Qu.:1.1445
##   Max.    :455.12   Max.    :811.12   Max.    :13.5050   Max.    :2.6330
```

2. Modificaciones.

Como se trata de un estudio a nivel foliar, las características que me interesa analizar son todas las que aparecen bajo la categoría “Leaf traits” que son “LT -> leaf thickness; LA -> leaf area; LDMC -> leaf dry matter; SLA -> leaf area/leaf dry mass; LCC -> leaf carbon concentration; LNC -> leaf carbon concentration; LNC -> leaf nitrogen concentration; MPR -> maximum photosynthesis rate”. A continuación reduzco el datadrame a las 3 especies de interés (*Dactylis glomerata*, *Festuca rubra* y *Erigeron canadensis*).

```
species_of_interest <- c("Dactylis glomerata", "Festuca rubra", "Erigeron canadensis")
small_dat <- Try_dat %>%
  filter(Species %in% species_of_interest) %>%
  select(Species, LT, LA, LDMC, SLA, LCC, LNC, MPR)
```

3. Gráfico.

Las variables de interés están medidas en diferentes unidades y escalas por lo que es necesario normalizarlas para poder representarlas en el gráfico de radar. Con la función `rescale()` normalizamos para que el rango de la variable esté entre 0 y 1, y así poder comparar todas las variables en el gráfico de radar. Existen otras formas de normalización como Z-score para normalizar todas las variables de forma que tengan una distribución normal con media 0 y desviación estándar de 1. En este último caso podrían existir valores negativos por lo que habría que hacerlos positivos para poder representarlos en el gráfico de radar. En mi caso he optado por la opción `rescale()`.

```
radar_dat <- small_dat %>%
  mutate(across(-Species, rescale)) # Normalize the data except 'Species'

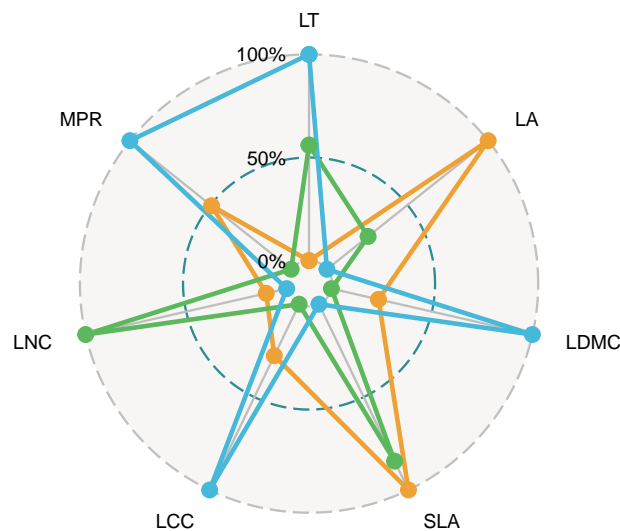
radar_plot <- ggradar(radar_dat,
```

```

axis.label.size = 3,
group.line.width = 1,
group.point.size = 3,
grid.label.size = 4,
group.colours = c("#EEA236", "#5CB85C", "#46B8DA"),
legend.position = "bottom") +
theme(legend.text = element_text(face = "italic")) +
labs(caption = "Plot acronyms represent, i.e.: \nLT --> leaf thickness\nLA --> leaf area\nLDMC --> leaf dry mass")
theme(plot.caption = element_text(size = 8, face = "italic", hjust = 0)) # Adjust caption position

```

radar_plot



—●— *Dactylis glomerata* —●— *Erigeron canadensis* —●— *Festuca rubra*

Plot acronyms represent, i.e.:
 LT --> leaf thickness
 LA --> leaf area
 LDMC --> leaf dry matter
 SLA --> leaf area/leaf dry mass
 LCC --> leaf carbon concentration
 LNC --> leaf nitrogen concentration
 MPR --> maximum photosynthesis rate

```

# Ruta y nombre del archivo a guardar
output_file2 <- file.path(main_dir, "4_grafico_radar.pdf")

ggsave(
  filename = output_file2,
  plot = radar_plot,
  width = 10,
  height = 8
)

```

El gráfico muestra las diferencias entre las tres especies en cuanto a varios parámetros. Por ejemplo, *Dactylis glomerata* muestra mayor LA y consecuentemente mayor SLA, mientras que *Erigeron canadensis* tiene mayor LNC y *Festuca rubra* destaca en cuanto a mayor MPR,LT, LDMC Y LCC.

Ejercicio 6: Diversidad gasterópodos

Se ha realizado un estudio de diversidad gasterópodos terrestres en un parque natural mediante transectos nocturnos. Se ha registrado, para cada evento de muestreo, la humedad relativa del aire y la temperatura media diarias. Los resultados para una determinada especie se muestran en Gasteropodos_ambiente.csv. ¿Existe algún patrón climático asociado a la mayor abundancia de la especie? Se pide responder a esta pregunta empleando material gráfico listo para ser publicado en una revista ambientalista mensual. Adjuntar el script utilizado.

1. Datos

```
Gdata <- read.csv2(file3, header = TRUE)
str(Gdata)
```

```
## 'data.frame': 199 obs. of 3 variables:
## $ adultos_Deroceras.transecto: int 5 6 4 3 7 5 7 3 2 1 ...
## $ humedad_media : num 51.2 51.2 51.2 51.2 51.2 ...
## $ temp_media : num 27.4 27.2 27 25.8 25.6 ...
```

```
summary(Gdata)
```

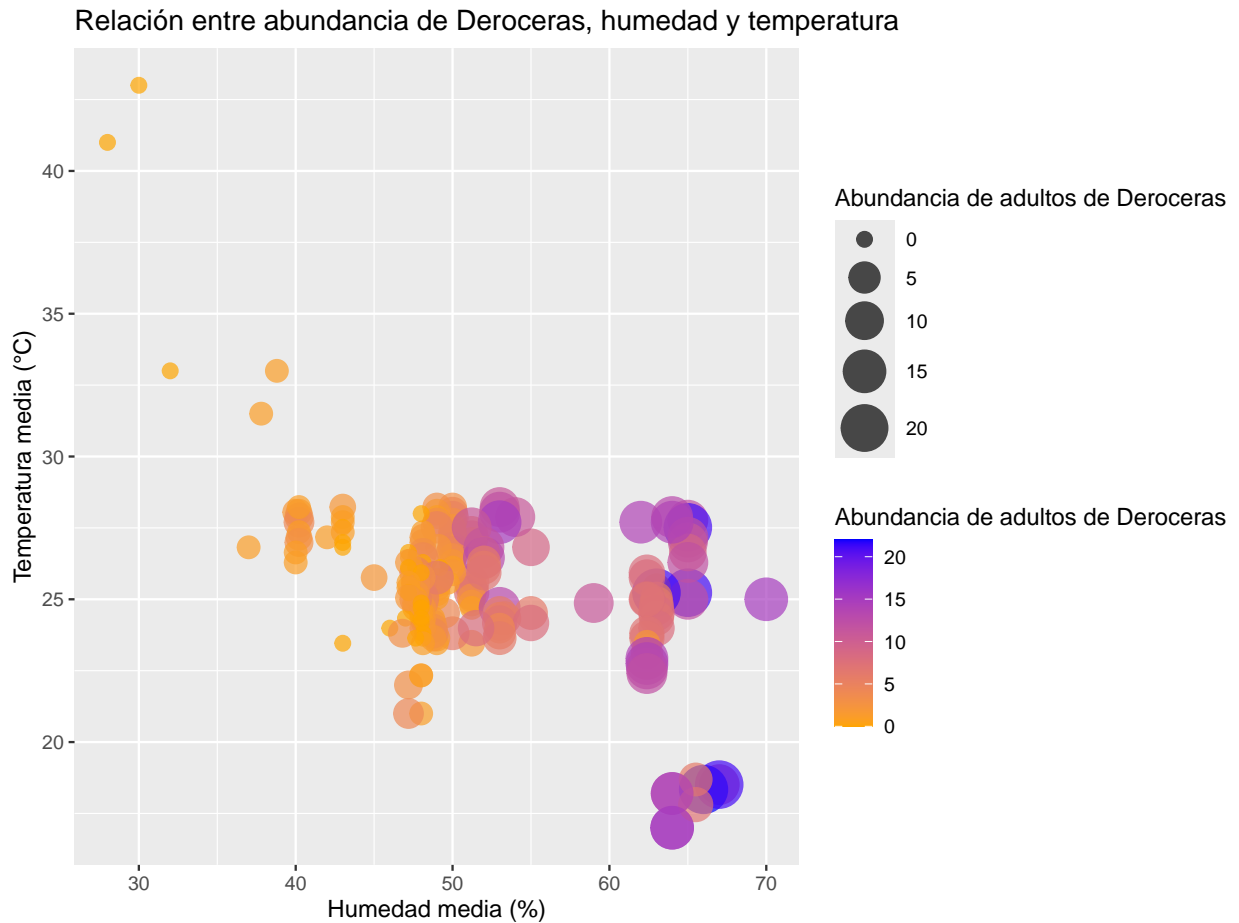
```
## adultos_Deroceras.transecto humedad_media temp_media
## Min. : 0.000 Min. :28.00 Min. :17.00
## 1st Qu.: 1.000 1st Qu.:47.83 1st Qu.:24.34
## Median : 3.000 Median :49.10 Median :25.40
## Mean : 4.804 Mean :51.58 Mean :25.51
## 3rd Qu.: 7.000 3rd Qu.:54.50 3rd Qu.:26.99
## Max. :22.000 Max. :70.00 Max. :43.00
```

2. Gráfico

En este caso el gráfico que he decidido hacer incluye tanto la variable humedad como la de temperatura media, para poder hacer una valoración conjunta sobre la abundancia de gasterópodos.

```
Gplot <- ggplot(Gdata, aes(x = humedad_media, y = temp_media, size = adultos_Deroceras.transecto, color =
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(3, 10)) +
  scale_color_gradient(low = "orange", high = "blue") +
  labs(title = "Relación entre abundancia de Deroceras, humedad y temperatura",
    x = "Humedad media (%)",
    y = "Temperatura media (°C)",
    size = "Abundancia de adultos de Deroceras",
    color = "Abundancia de adultos de Deroceras")
```

```
Gplot
```

```
# Ruta y nombre del archivo a guardar
output_file2 <- file.path(main_dir, "6_Gplot.pdf")

ggsave(
  filename = output_file2,
  plot = Gplot,
  width = 10,
  height = 7,
  dpi = 300
)
```

En respuesta a la pregunta del enunciado del ejercicio, el gráfico muestra mayor abundancia de gasterópodos cuando la humedad media es superior al 60% y la temperatura media es inferior a 30 grados. En todo caso es llamativa la escasez de gasterópodos para la mencionada humedad cuando la temperatura es de alrededor de 20 grados para volver a aumentar cuando la temperatura es inferior a 20 grados.

Ejercicio 8: Polinizadores

En un invernadero se quiere evaluar la efectividad de la liberación de polinizadores criados en cautividad (*Bombus* sp.) sobre la producción de fruto de una determinada especie de arbusto mediterráneo para el

que se dispone de tres variedades genéticamente diferenciadas (Polinizadores.csv). Se pide determinar el momento óptimo de liberación de los insectos polinizadores, para ello se requiere de un gráfico explicativo que pueda utilizarse en las campañas de divulgación con agricultores y dueños de viveros forestales locales. Adjuntar el script.

1. Datos

```
df_poli <- read.csv2(file4, header = TRUE)
str(df_poli)
```

```
## 'data.frame': 163 obs. of 4 variables:
## $ Variedad : chr "multiflora" "macrocarpa" "multiflora" "microcarpa" ...
## $ Edad_maduracion_meses: num 10 2.8 10 5.6 7 3 4 4.7 6 11 ...
## $ Floracion_juliana : int 33 29 59 19 38 35 31 13 33 53 ...
## $ Riego : chr "normal" "normal" "normal" "doble" ...
```

```
summary(df_poli)
```

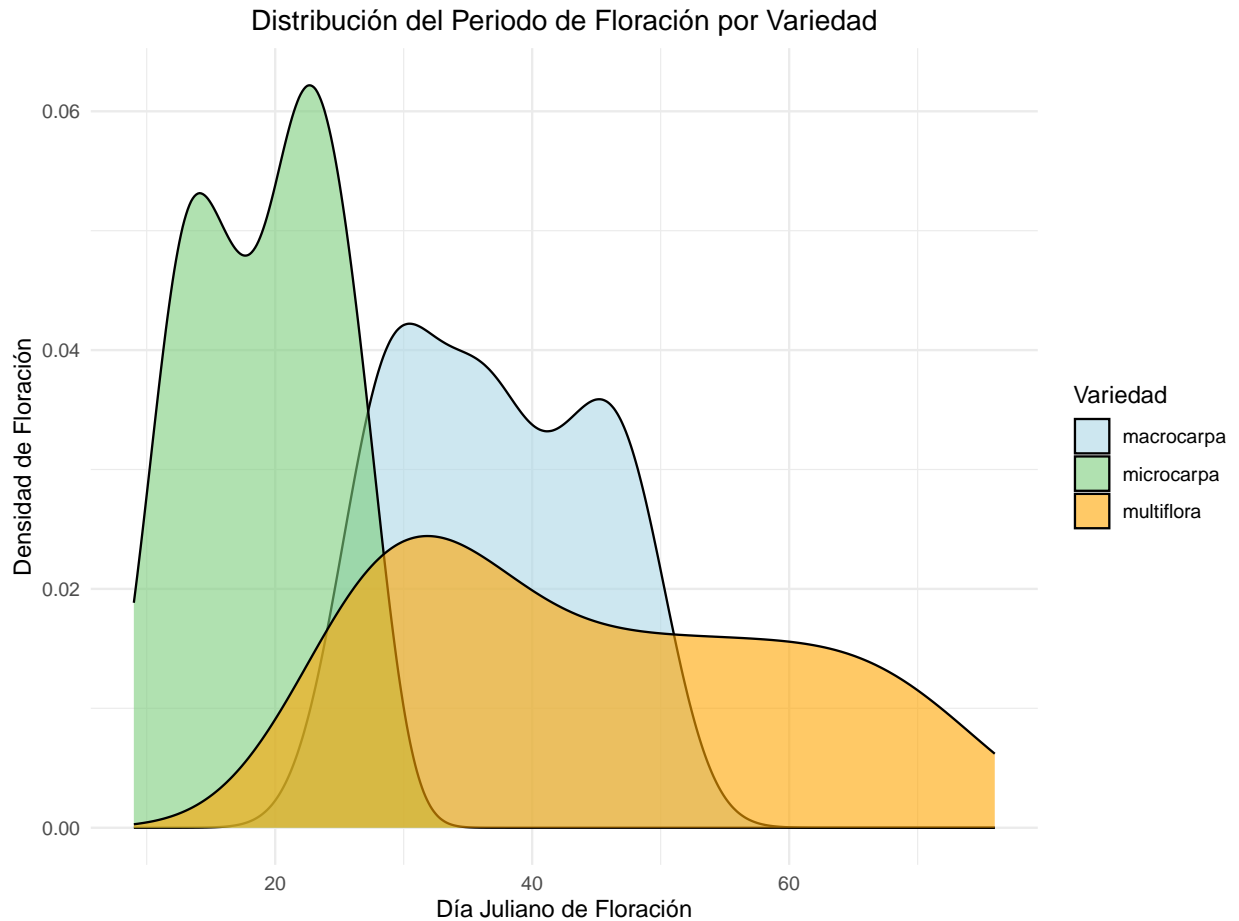
```
## Variedad Edad_maduracion_meses Floracion_juliana Riego
## Length:163 Min. : 1.000 Min. : 9.00 Length:163
## Class :character 1st Qu.: 4.500 1st Qu.:23.00 Class :character
## Mode :character Median : 6.000 Median :29.00 Mode :character
## Mean : 6.708 Mean :33.27
## 3rd Qu.: 8.550 3rd Qu.:43.00
## Max. :21.000 Max. :76.00
```

2. Gráfico

En este caso he preferido realizar un gráfico de densidades para cada variedad. Posteriormente lo he dividido en gráfico de densidades para cada variedad y tipo de riego.

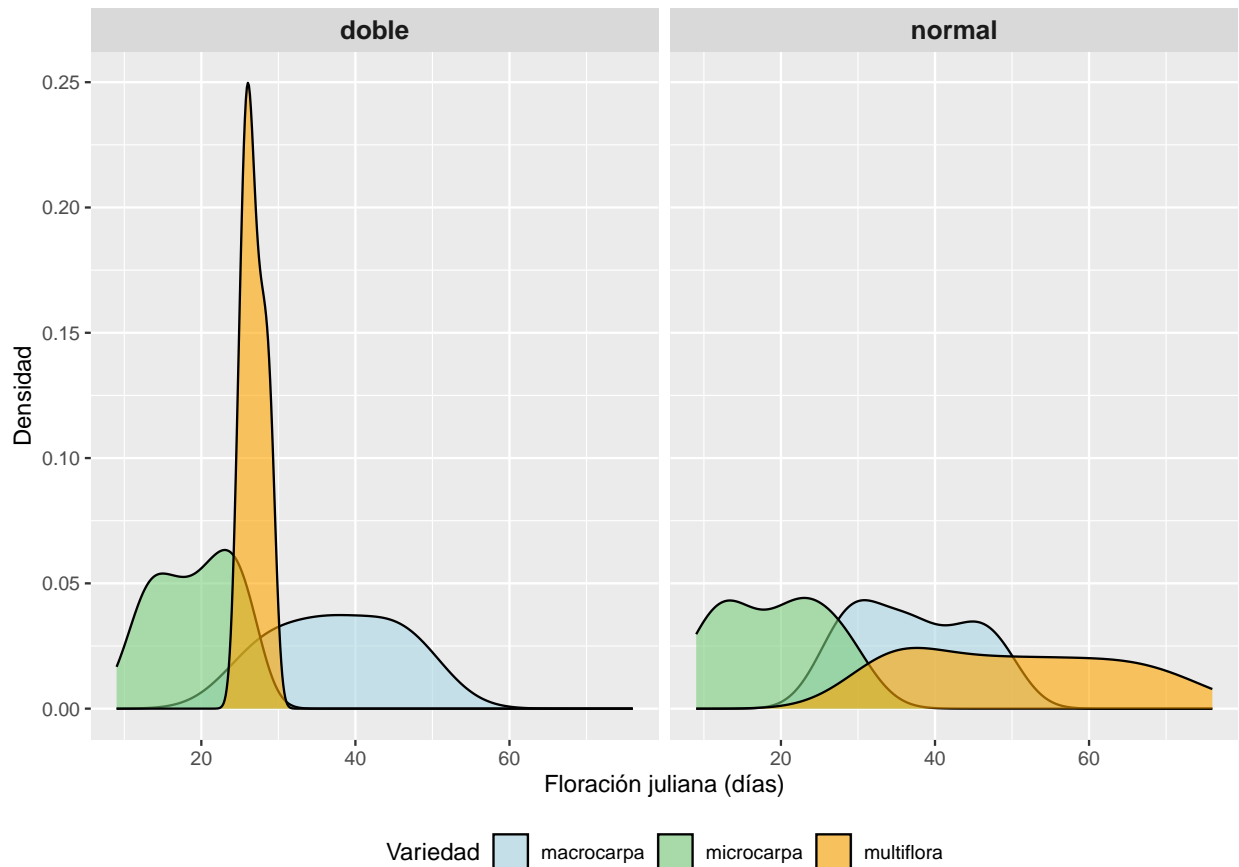
```
plot_poli1 <- ggplot(df_poli, aes(x = Floracion_juliana, fill = Variedad)) +
  geom_density(alpha = 0.6) +
  labs(title = "Distribución del Periodo de Floración por Variedad",
    x = "Día Juliano de Floración",
    y = "Densidad de Floración",
    fill = "Variedad") +
  theme_minimal() +
  scale_fill_manual(values = c("macrocarpa" = "lightblue", "microcarpa" = "palegreen3", "multiflora" = "orange"),
    theme(plot.title = element_text(hjust = 0.5)))

plot_poli1
```



```
plot_poli2 <- ggplot(df_poli, aes(x = Floracion_juliana, fill = Variedad)) +
  geom_density(alpha = 0.6) +
  facet_wrap(~ Riego) +
  labs(
    title = "Distribución de floración juliana según variedad y riego",
    x = "Floración juliana (días)",
    y = "Densidad",
    fill = "Variedad"
  ) +
  scale_fill_manual(values = c("macrocarpa" = "lightblue", "microcarpa" = "palegreen3", "multiflora" =
  theme(
    strip.text = element_text(size = 12, face = "bold"),
    legend.position = "bottom"
  )
plot_poli2
```

Distribución de floración juliana según variedad y riego



```
# Ruta y nombre del archivo a guardar
output_poli1 <- file.path(main_dir, "8_grafico_poli1.png")
output_poli2 <- file.path(main_dir, "8_grafico_poli2.png")

ggsave(
  filename = output_poli1,
  plot = plot_poli1,
  width = 10,
  height = 7,
  dpi = 300
)

ggsave(
  filename = output_poli2,
  plot = plot_poli2,
  width = 10,
  height = 7,
  dpi = 300
)
```

En ambos casos los gráficos muestran que la suelta de polinizadores debería hacerse entre los 20 y 30 días, si la suelta se hace para obtener una polinización simultanea para todas las especies. Si nos centramos en la suelta específica por especie y riego, entonces las sueltas de polinizadores deberían hacerse del siguiente modo: - Macrocarpa: el periodo de suelta puede hacerse entre los 30 y 40 días independientemente del riego

ya que la floración es similar en ambos casos. - Microcarpa: el periodo de suelta de polinizadores debe hacerse entre los 15 y 25 días, independientemente del riego ya que la floración es similar. - Multiflora: en este caso la floración es dependiente del tipo de riego. En el caso del riego doble, la suelta debe hacerse entre los días 25 y 27. En cambio para el riego normal el periodo de suelta puede hacerse entre los días 35 y 70.
