

Lacaze-Corentin-PEC1

Corentin Lacaze

2025-03-31

Contents

1	1. Tabla de Contenidos : https://github.com/verdaturren/Lacaze-Corentin-PEC1.0	1
2	2. Resumen	2
3	3. Objetivos	2
4	4. Metodología	2
4.1	4.1 Selección del dataset	2
4.2	4.2 Preparación de los datos y creación de SummarizedExperiment	2
4.3	4.3 Análisis exploratorio	5
5	5. Resultados	5
5.1	5.1 PCA Modo POS	5
5.2	5.2 PCA Modo NEG	6
5.3	5.3 Heatmap Modo POS	7
5.4	5.4 Ejemplo de variación (Sujeto 1, modo POS)	9
5.5	5.5 En resumen, la ingestión de jugos parece afectar la cantidad de ciertos metabolitos.	9
6	6. Discusión	9
7	7. Conclusiones	9
8	8. Referencias	10
	enlace al repositorio	

1	1. Tabla de Contenidos : https://github.com/verdaturren/Lacaze-Corentin-PEC1.0
----------	---

2 2. Resumen

El presente informe describe el proceso de selección y análisis exploratorio de datos metabolómicos obtenidos a partir de espectrometría de masas en modo positivo y negativo. Los datos corresponden a un estudio en el que se investigan los cambios metabólicos tras el consumo de zumos de manzana y arándano en sujetos humanos. Se han creado objetos de la clase `SummarizedExperiment`, y se han aplicado técnicas de PCA y heatmap para obtener una visión general de los patrones metabolómicos. Asimismo, se analizan posibles variaciones individuales (comparando la línea basal con la ingesta de diferentes zumos). Finalmente, se discute la diferencia entre `SummarizedExperiment` y `ExpressionSet` y se presenta la interpretación biológica preliminar de los resultados.

3 3. Objetivos

1. **Justificar la elección del dataset metabolómico** (modos positivo y negativo) y describir brevemente el contexto biológico del estudio (consumo de zumos).
 2. **Crear un objeto** de clase `SummarizedExperiment` que contenga los datos y sus metadatos.
 3. **Realizar un análisis exploratorio** para obtener una visión global del dataset: PCA, heatmap, comparación de perfil entre condiciones.
 4. **Discutir** las diferencias entre `SummarizedExperiment` y `ExpressionSet`.
 5. **Interpretar** los resultados desde el punto de vista biológico.
-

4 4. Metodología

4.1 4.1 Selección del dataset

Se han utilizado dos archivos en formato Excel:

- `Gu_pos-metabolites_PA-WB.xlsx` (modo positivo del espectrometría de masas)
- `Gu_neg-metabolites_PA-WB.xlsx` (modo negativo del espectrometría de masas)

Estos datos representan picos de intensidad (peak area) de metabolitos en orina o plasma, antes y después del consumo de diferentes zumos. Cada columna corresponde a un sujeto y el sufijo (`b1`, `a1`, `c1`, etc.) indica momento (basal vs post-ingesta) y tipo de zumo (manzana vs arándano).

4.2 4.2 Preparación de los datos y creación de `SummarizedExperiment`

```
library(SummarizedExperiment)
library(readxl)
library(ggplot2)
library(pheatmap)
library(FactoMineR)
library(factoextra)
library(reshape2)
```

```

create_se_object <- function(df, mode) {
  # Quitar columnas innecesarias
  df <- df[, !(colnames(df) %in% c("METABOLITE_ID", "Units"))]

  # rowData = MetaboliteName
  rowData <- DataFrame(MetaboliteName = df$`Metabolite Name`)

  # Extraer matriz numérica (sin la 1era columna que es MetaboliteName)
  assay_data <- as.matrix(df[, -1])
  assay_data <- apply(assay_data, 2, as.numeric)

  SummarizedExperiment(
    assays = list(counts = assay_data),
    rowData = rowData,
    colData = DataFrame(Sample = colnames(assay_data)),
    metadata = list(mode = mode)
  )
}

run_pca <- function(se, mode) {
  mat <- t(assay(se))
  mat <- log2(mat + 1)
  pca <- PCA(mat, graph = FALSE)
  fviz_pca_ind(pca, title = paste("PCA - Modo", mode))
}

plot_heatmap <- function(se, mode) {
  mat <- assay(se)
  mat <- log2(mat + 1)

  mat <- t(scale(t(mat)))
  pheatmap(mat, main = paste("Heatmap - Modo", mode), show_rownames = FALSE)
}

# problema solve
analyze_subject_variation <- function(df, subject_id) {
  # Ex. b1, a1, c1
  cols <- paste0(c("b", "a", "c"), subject_id)

  df_subj <- df[, cols]

  rownames(df_subj) <- make.unique(as.character(df$`Metabolite Name`))

  df_subj <- as.data.frame(lapply(df_subj, as.numeric), row.names = rownames(df_subj))

  delta_apple <- log2(df_subj[[paste0("a", subject_id)]] + 1) -
    log2(df_subj[[paste0("b", subject_id)]] + 1)
  delta_cran <- log2(df_subj[[paste0("c", subject_id)]] + 1) -
    log2(df_subj[[paste0("b", subject_id)]] + 1)

  variations <- data.frame(
    Metabolito = rownames(df_subj),
    Delta_Manzana = delta_apple,

```

```

    Delta_Arandano = delta_cran
  )

df_melt <- melt(variations, id.vars = "Metabolito")
ggplot(df_melt, aes(x = variable, y = value, fill = variable)) +
  geom_boxplot() +
  labs(
    title = paste("Variación - Sujeto", subject_id),
    x = "Condición",
    y = "Log2 Change vs basal"
  ) +
  theme_minimal()
}

analyze_subject_variation <- function(df, subject_id) {
  # Ex. b1, a1, c1
  cols <- paste0(c("b", "a", "c"), subject_id)
  if (!all(cols %in% colnames(df))) {
    stop("Columnas faltantes para el sujeto ", subject_id)
  }

  df_subj <- df[, cols]

  # columnas en formato numérico
  rownames(df_subj) <- make.unique(as.character(df$`Metabolite Name`))

  df_subj <- as.data.frame(lapply(df_subj, as.numeric), row.names = rownames(df_subj))

  delta_apple <- log2(df_subj[[paste0("a", subject_id)]] + 1) -
    log2(df_subj[[paste0("b", subject_id)]] + 1)
  delta_cran <- log2(df_subj[[paste0("c", subject_id)]] + 1) -
    log2(df_subj[[paste0("b", subject_id)]] + 1)

  variations <- data.frame(
    Metabolito = rownames(df_subj),
    Delta_Manzana = delta_apple,
    Delta_Arandano = delta_cran
  )

  library(reshape2)
  df_melt <- melt(variations, id.vars = "Metabolito")

  library(ggplot2)
  ggplot(df_melt, aes(x = variable, y = value, fill = variable)) +
    geom_boxplot() +
    labs(
      title = paste("Variación - Sujeto", subject_id),
      x = "Condición",
      y = "Log2 Fold Change vs basal"
    ) +
    theme_minimal()
}

```

```

# Carga de los datasets y creación de SummarizedExperiment
file_pos <- "Gu_pos-metabolites_PA-WB.xlsx"
file_neg <- "Gu_neg-metabolites_PA-WB.xlsx"

df_pos <- read_excel(file_pos)
df_neg <- read_excel(file_neg)

# Creación de SummarizedExperiment
se_pos <- create_se_object(df_pos, "POS")
se_neg <- create_se_object(df_neg, "NEG")

```

4.3 4.3 Análisis exploratorio

4.3.1 4.3.1 PCA

Se aplicó una transformación $\log_2(x + 1)$ a las intensidades para estabilizar la varianza. Posteriormente se realizó un Análisis de Componentes Principales (PCA) con el paquete **FactoMineR**.

4.3.2 4.3.2 Heatmap

Se construyó un heatmap usando **pheatmap**, con escalado Z-score por metabolito. Las filas con demasiados valores perdidos se excluyeron para evitar errores de clustering.

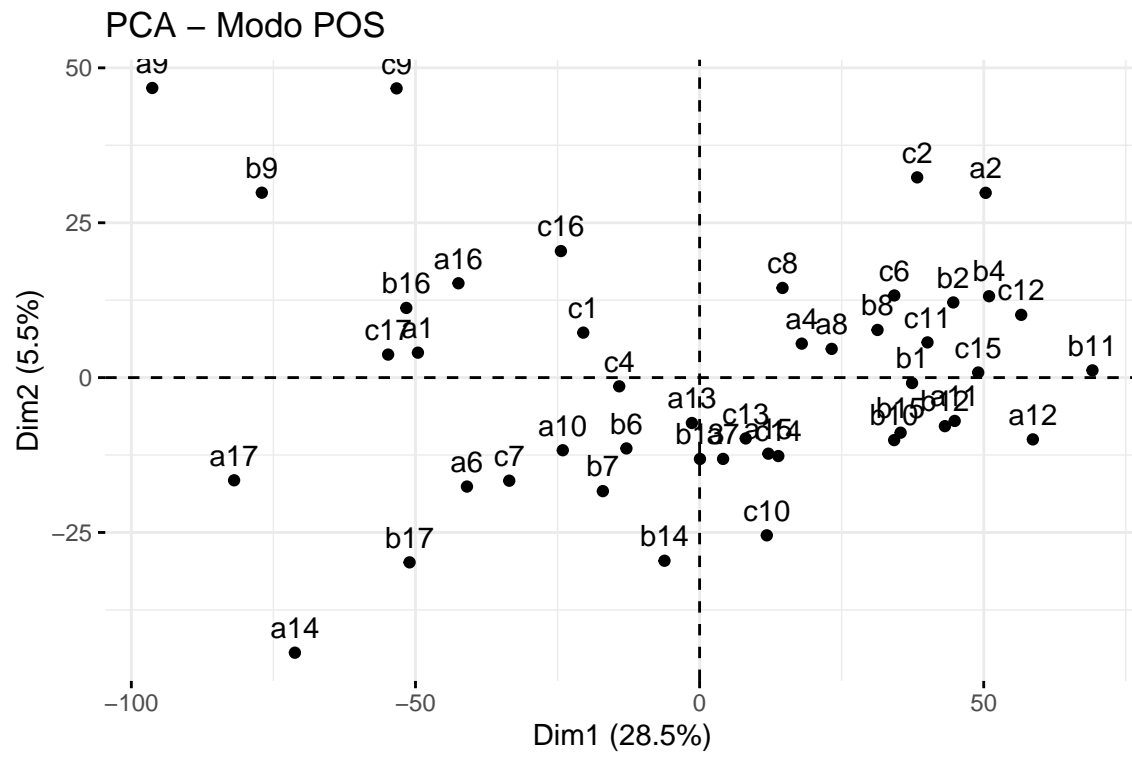
4.3.3 4.3.3 Comparación por sujeto (Ej. sujeto 1)

Para ilustrar la variación individual, se compararon los valores log2-fold-change entre la línea basal (**b1**) y post-ingesta de zumo de manzana (**a1**) o arándano (**c1**).

5 5. Resultados

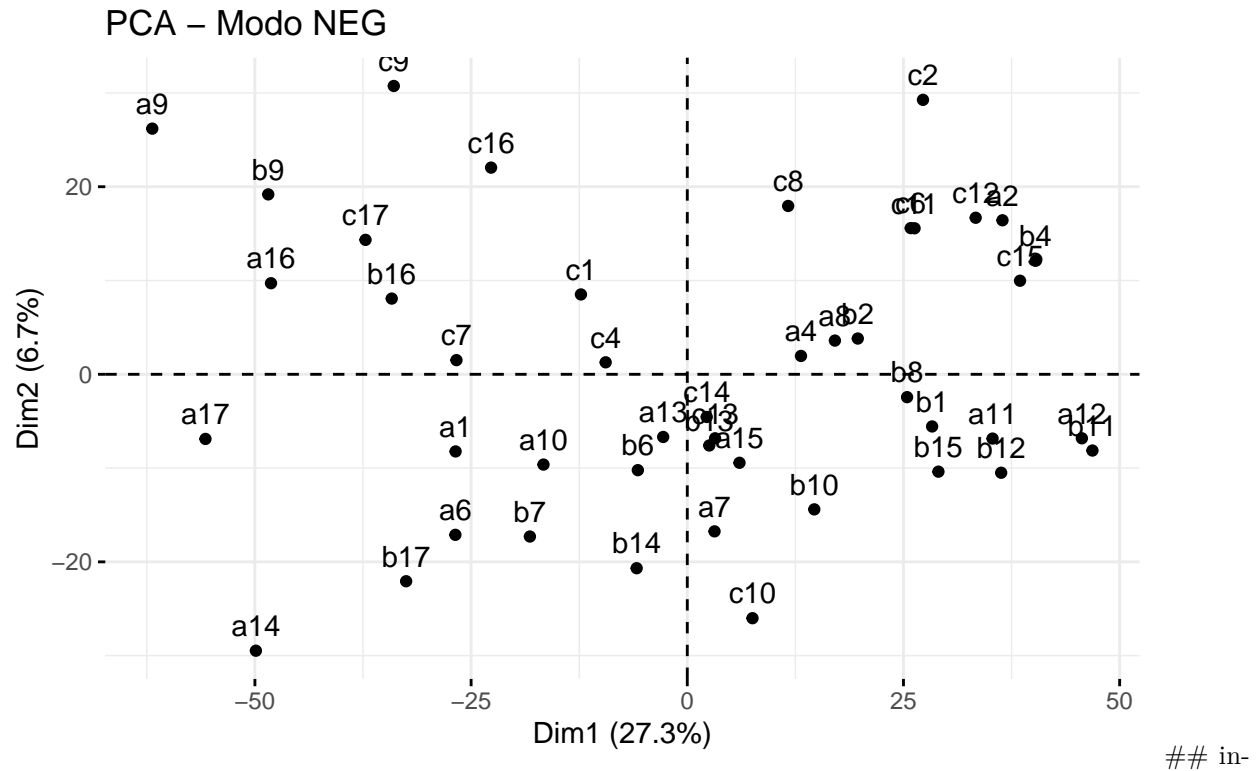
5.1 5.1 PCA Modo POS

```
run_pca(se_pos, "POS")
```



5.2 5.2 PCA Modo NEG

```
run_pca(se_neg, "NEG")
```



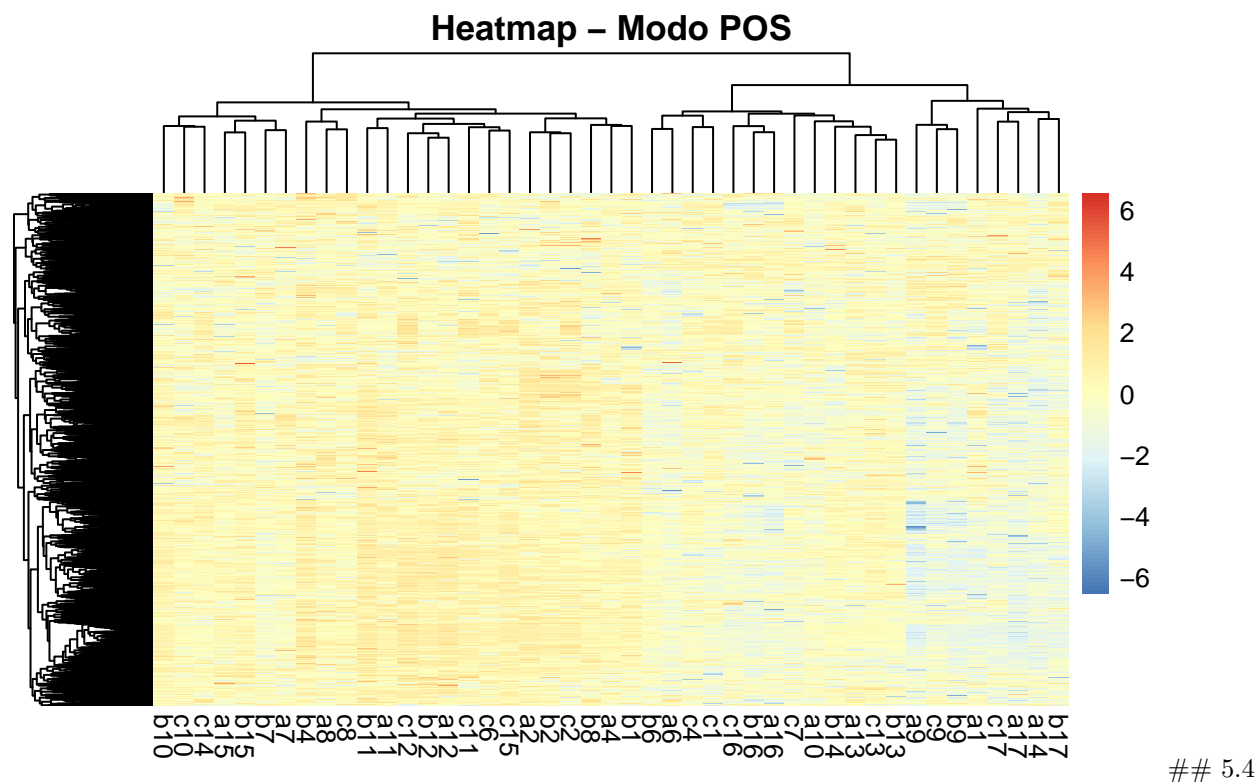
interpretación PCA Podemos extraer alguna información útil de este gráfico:

- No hay agrupamientos claros entre las diferentes condiciones del experimento, es decir, no hay grandes
- Algunos puntos están muy alejados, lo que sugiere una gran variación entre los individuos.

En resumen, los efectos parecen muy débiles y una gran variabilidad dificulta la interpretación de los datos.

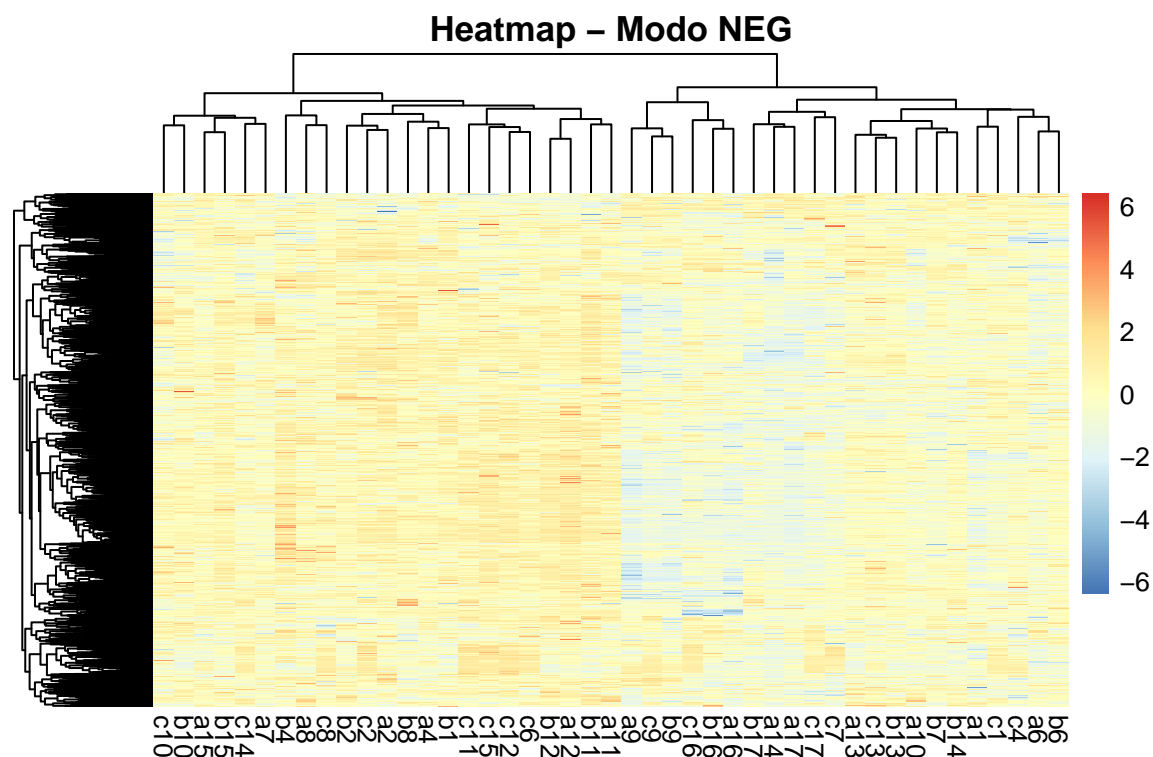
5.3 Heatmap Modo POS

```
plot_heatmap(se_pos, "POS")
```



Heatmap Modo NEG

```
plot_heatmap(se_neg, "NEG")
```

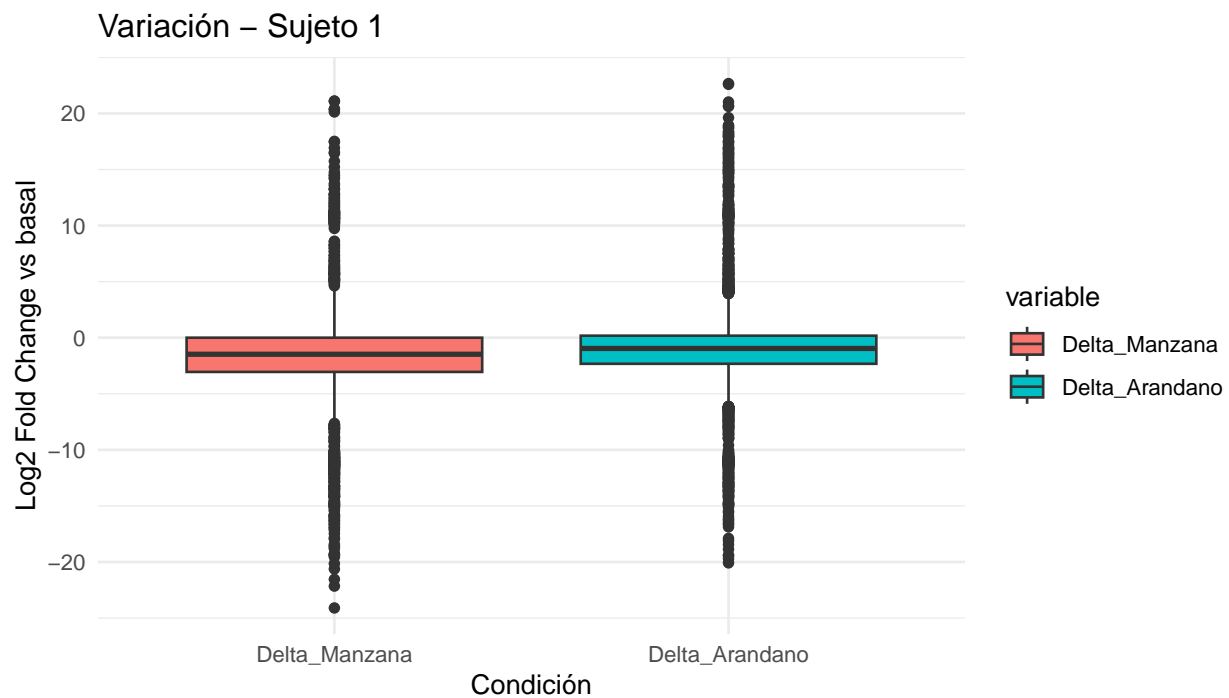


##interpretación heatmap Se aprecia una variación en algunos metabolitos tras la ingesta de zumo de

manzana o arándano, lo que sugiere un posible impacto de estos compuestos bioactivos. Se requeriría un análisis estadístico adicional para confirmar estos cambios.

5.4 Ejemplo de variación (Sujeto 1, modo POS)

```
analyze_subject_variation(df_pos, 1)
```



Interpretación:

El gráfico muestra varias cosas:

Las dos cajas están relativamente centradas en 0, lo que indica que, en general, hay poca variación en los metabolitos. Sin embargo, un gran número de puntos se encuentran fuera de la caja de bigotes, lo que indica una alta variabilidad.

5.5 En resumen, la ingestión de jugos parece afectar la cantidad de ciertos metabolitos.

6 6. Discusión

No podemos concluir claramente sobre el impacto de la ingesta de jugos en los metabolitos encontrados en las muestras; una gran variabilidad requiere una muestra más grande de participantes para mitigar esto.

7 7. Conclusiones

1. **Datos metabolómicos:** El uso de `SummarizedExperiment` facilita la gestión y exploración de datos.

2. **Análisis exploratorio:** La PCA y el heatmap evidencian cambios potenciales tras la ingesta de zumos.
 3. **Visión biológica:** Se observan posibles efectos de los polifenoles presentes en los zumos de manzana y arándano, sin poder ser formal en la conclusión.
-

8 8. Referencias

- metabolomicsWorkbench
- Bioconductor: SummarizedExperiment
- FactoMineR y factoextra para PCA.