

## Índice

- Abstract
- Objetivos del estudio
- Materiales y metodos
- Resultados
- Discusión limitaciones y conclusiones del estudio.

## Abstract

En este trabajo, he realizado la exploración de datos usados en el paper "Metabotypes of response to bariatric surgery independent of the magnitude of weight loss", usando en R, principalmente, la librería SummarisedExperiment para crear un objeto y así explorar datos metabolómicos en distintos pacientes en cinco tiempos distintos.

## Objetivos del estudio

El objetivo del estudio es la exploración de datos metabolómicos usando un objeto SummarisedExperiment, en este caso, la comparación de diferentes valores metabolómicos en pacientes en cinco tiempos distintos.

## Materiales y metodos

He usado R, SummarisedExperiment. Para la obtención de los datos, he instalado Git en mi sistema para así usarlo dentro de Rstudio haciendo check en 'Create a git repository for this project', al crear mi proyecto de R. Anteriormente he instalado la aplicación Desktop de Github, para así clonar el github proporcionado por la PEC <https://github.com/nutrimetabolomics/metaboData/> e introducirlo en mi repositorio git, dentro de este repositorio se encuentra la carpeta Datasets, donde se encuentra la carpeta 2018-MetabotypingPaper. En esta carpeta se encuentran los datos que he utilizado para mi estudio, y una descripción de ellos. Añadir que uno de los datos, A4information\_S006, estaba en el formato equivocado (.hlm), sin embargo al acceder al archivo donde se describen los datos, obtenemos un repositorio independiente de estos datos, donde A4information\_S006 esta en el formato csv, por lo tanto lo he descargado de manera individual para reemplazar los datos con el formato equivocado.

Para crear un objeto SummarisedExperiment es necesario una matriz de datos, assay, he usado DataValues\_S013.csv. He eliminado su primera columna ya que era redundante y he dividido el dataset en dos, assay\_data, que son todos los datos numéricos de distintos metabolitos y biométricas, y sample\_info, las primeras 5 columnas de DataValues\_S013, con valores como el sexo de los pacientes, el grupo donde están, he usado assay\_data como mi matriz de datos, y sample\_info como mi Col\_data, añadir que ha sido necesario transponer assay\_data. Además he usado el dataset DataInfo\_S013.csv, eliminando su primera columna por la misma razón, para así introducirla como mis metadatos de mi objeto SummarisedExperiment.

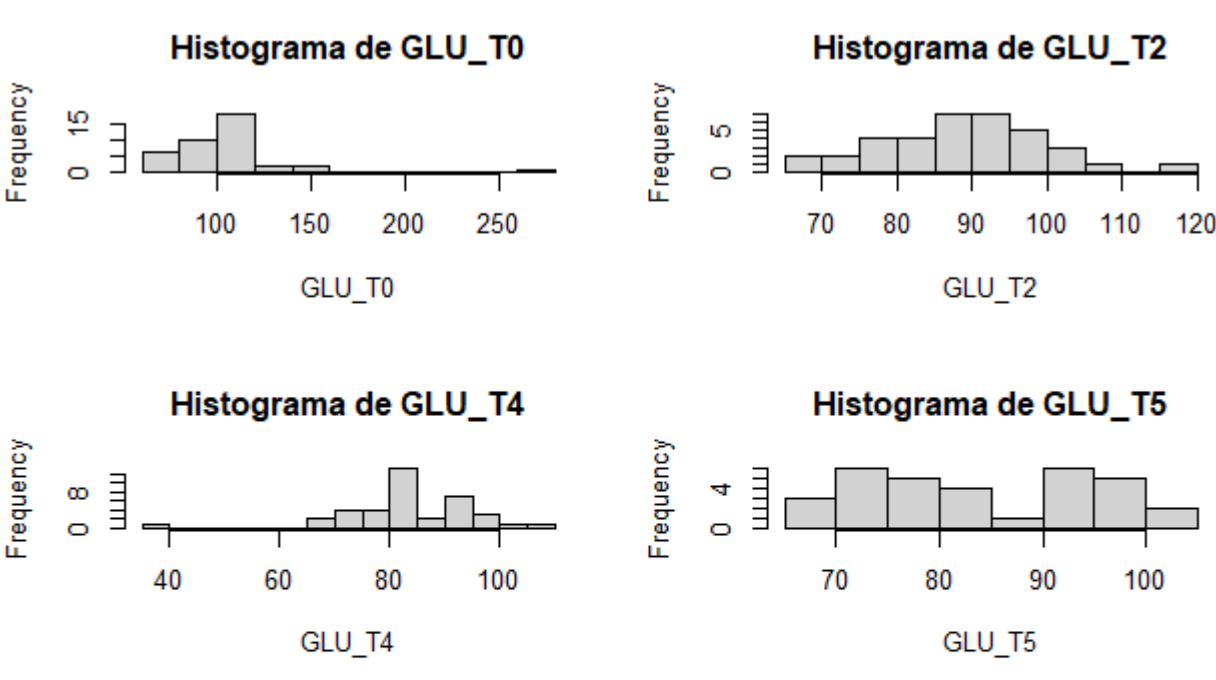
A la hora de explorar los datos, ya que se dividen en tiempos distintos, he tomado los niveles de glucosa, colesterol, peso, LDL, HDL, y aminoácidos Leucina y Prolina para hacer histogramas con cada uno de ellos, además, he realizado un PCA con los valores de mi assay eliminando los valores NA, ya que es necesario para realizar un PCA en R, con el cual he realizado un análisis de las desviaciones estándar de mis componentes y he comparado los primeros dos componentes de mi PCA en una gráfica de puntos usando ggplot2.

Finalmente, he creado un archivo Rda 'objetoSE.Rda' con mi objeto SummarisedExperiment, un archivo csv con mis datos para el assay de este, 't\_assay\_data.csv' y un archivo markdown con los metadatos de mi objeto SummarisedExperiment, además de crear un repositorio en <https://github.com> donde se encuentran estos datos

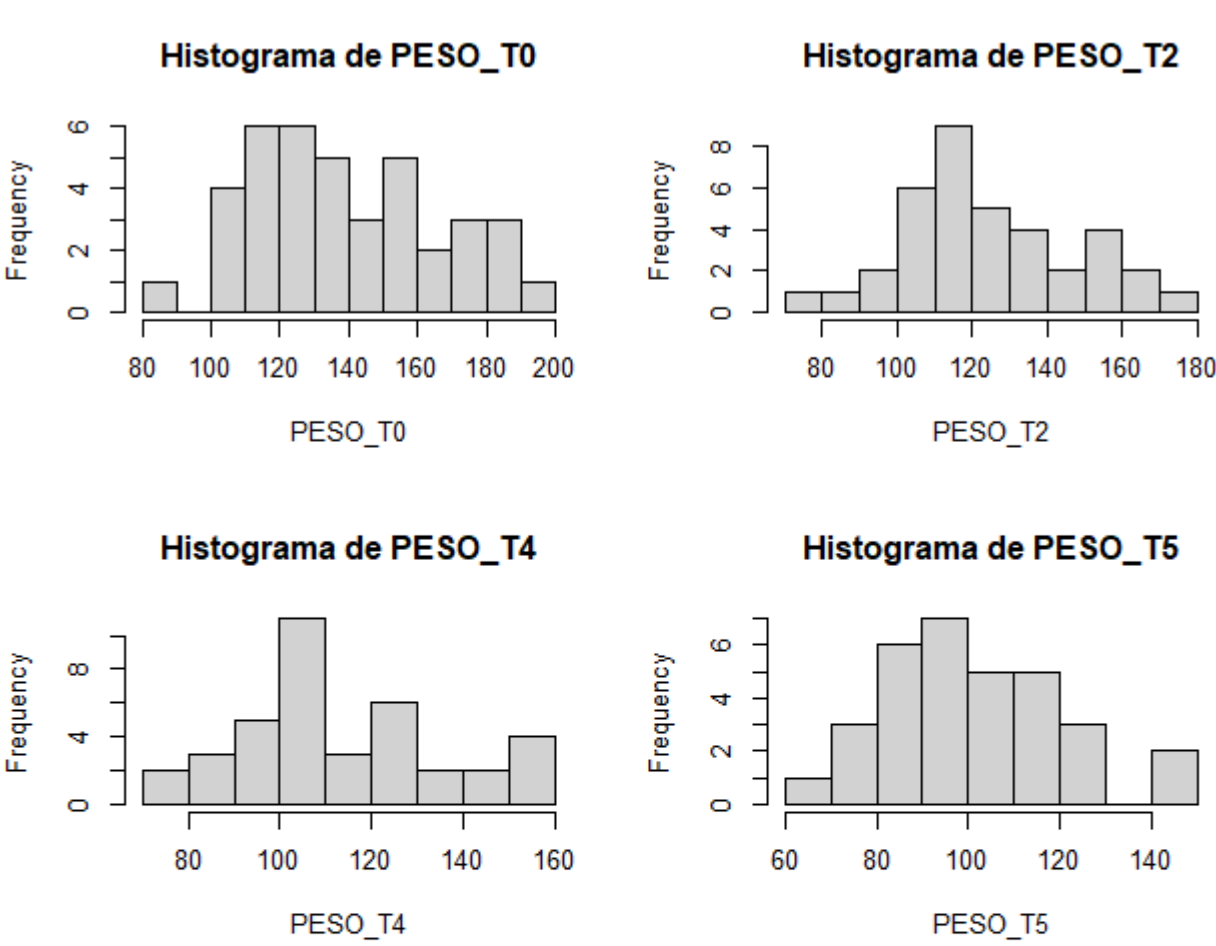
## Resultados

Tenemos los histogramas de los distintos valores metabolómicos mencionados anteriormente:

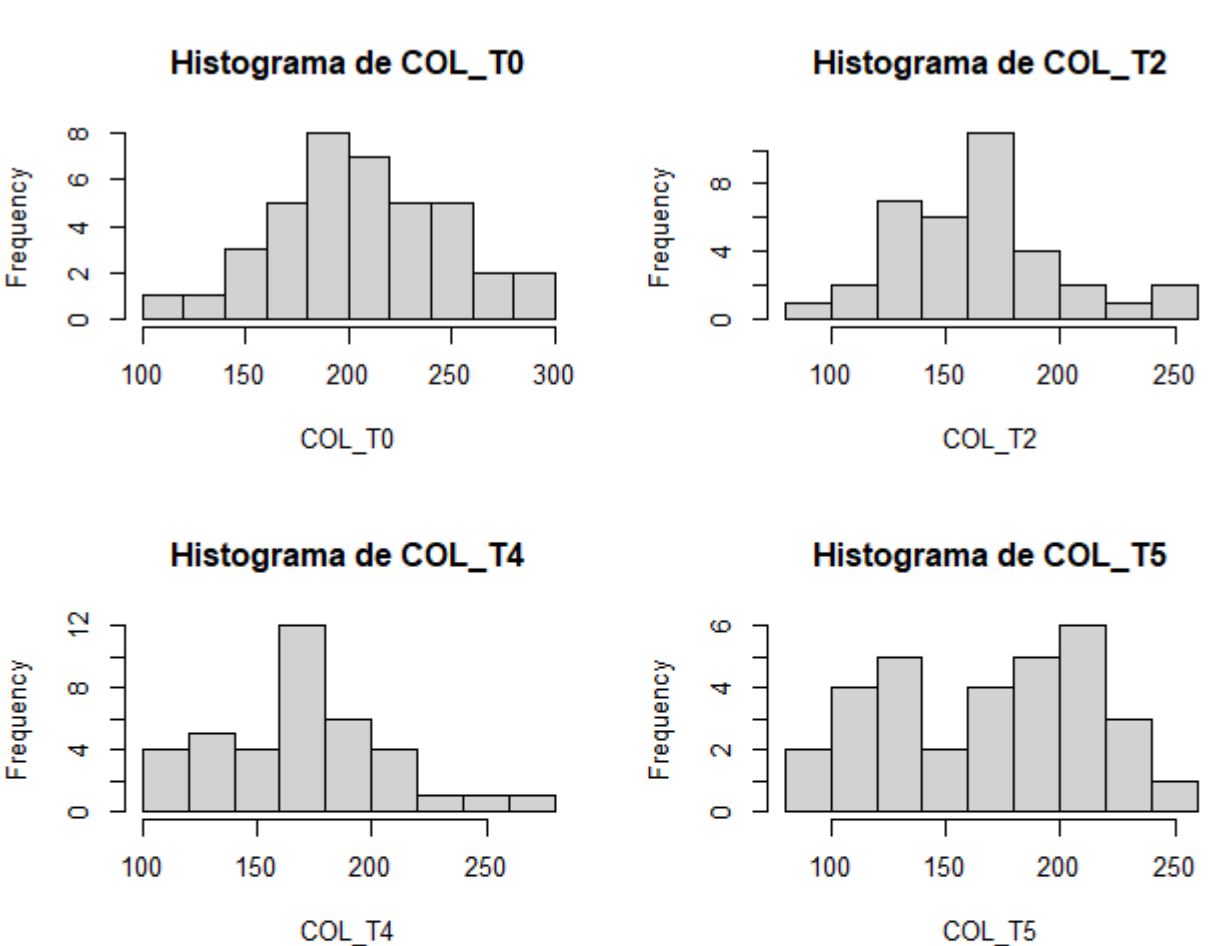
Glucosa



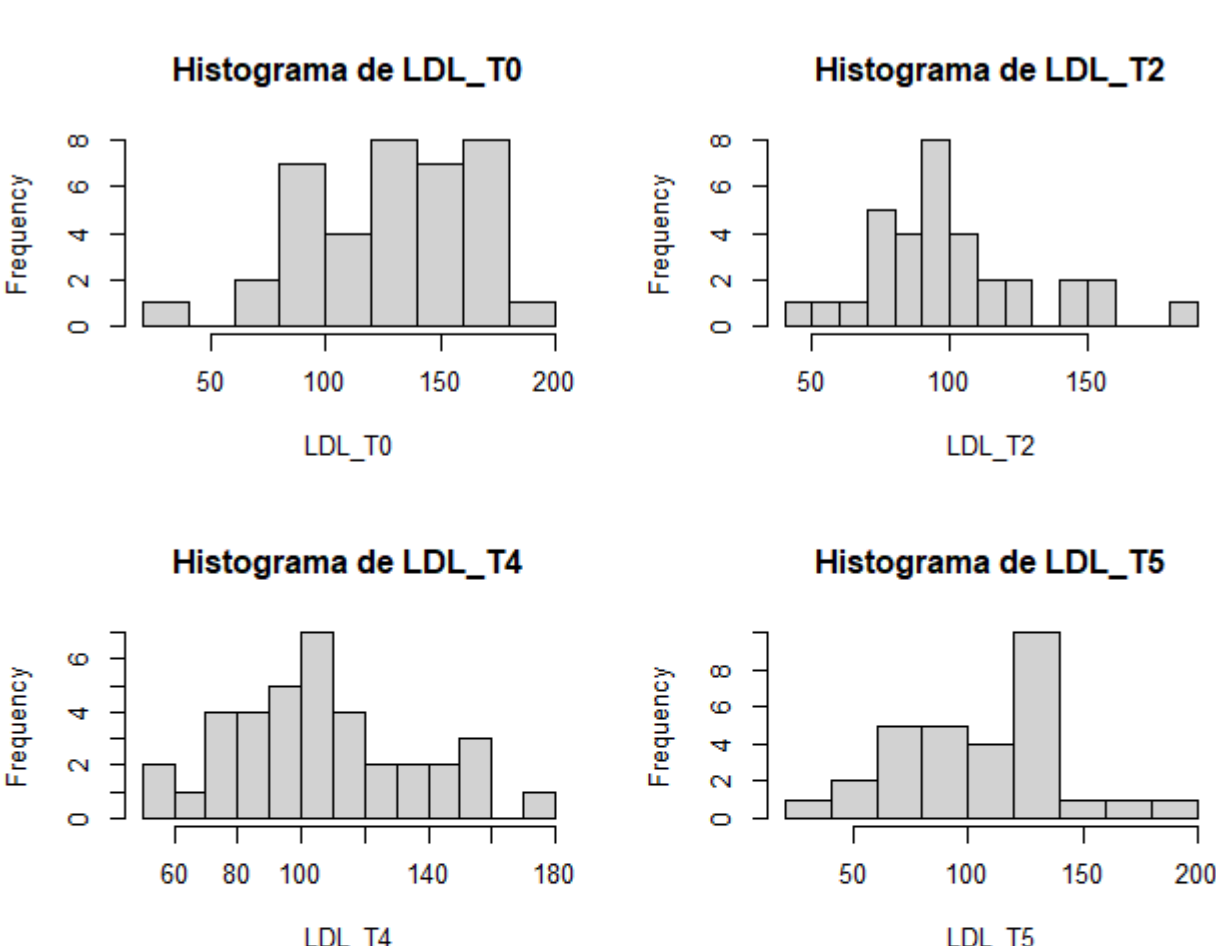
Peso



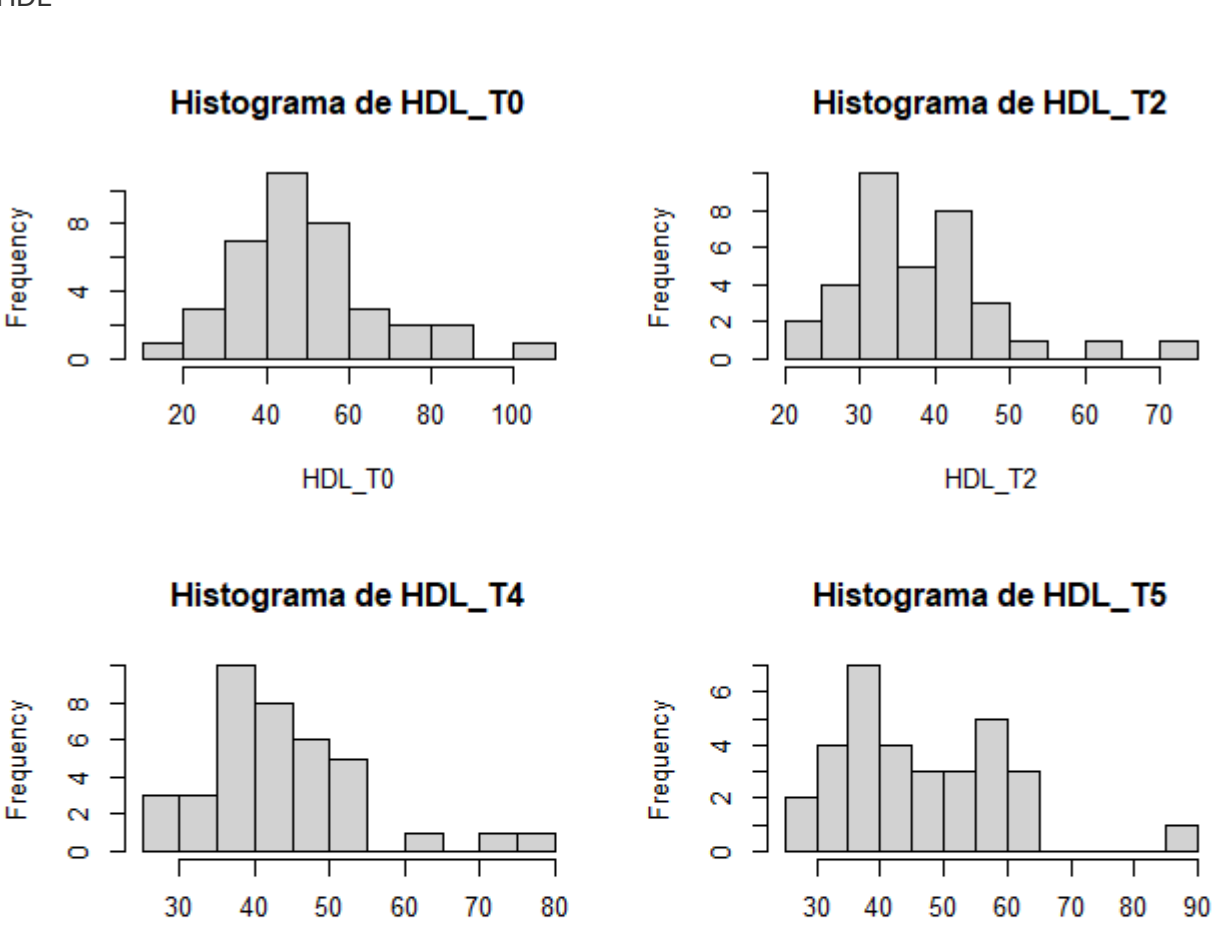
Colesterol



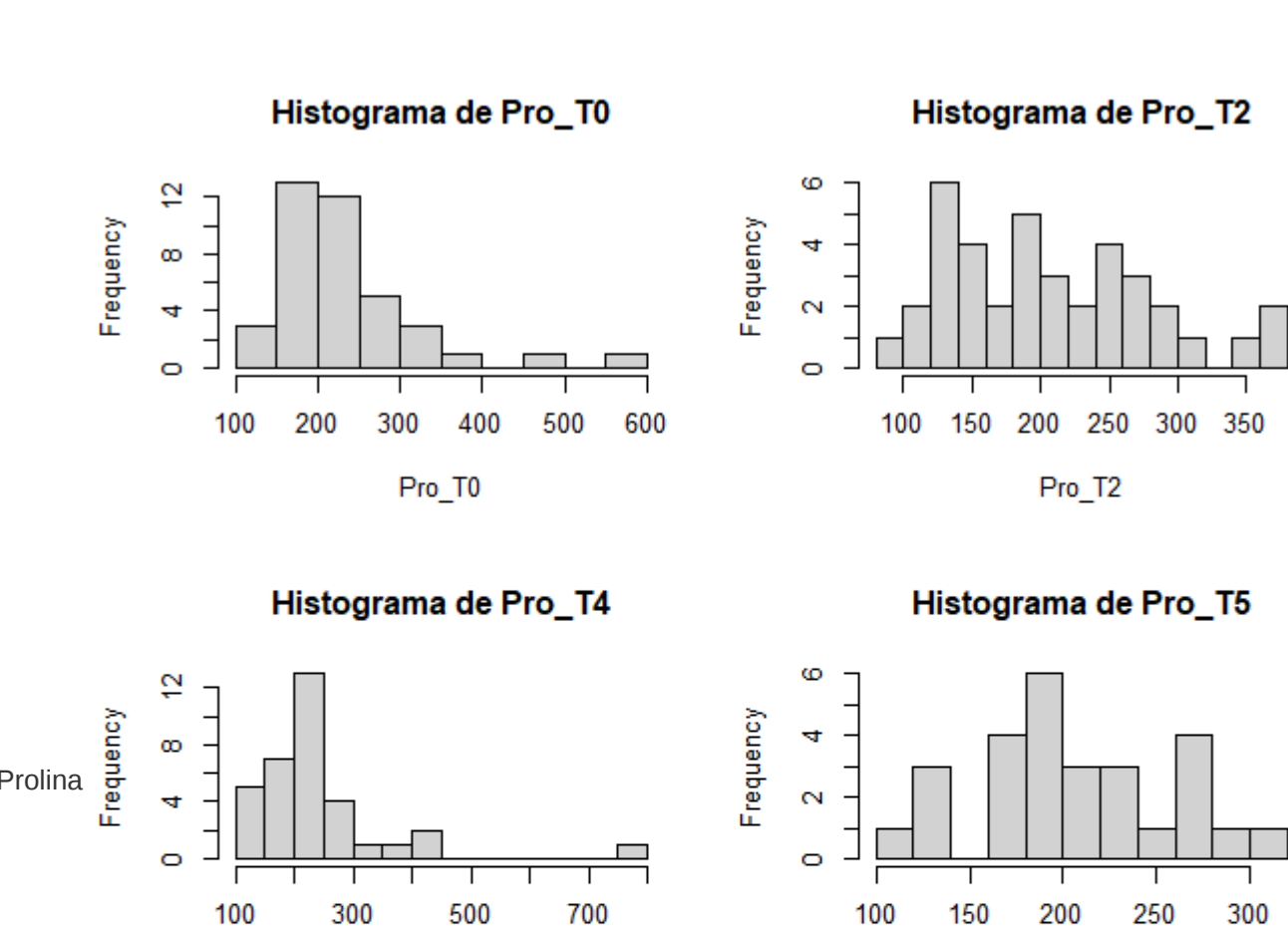
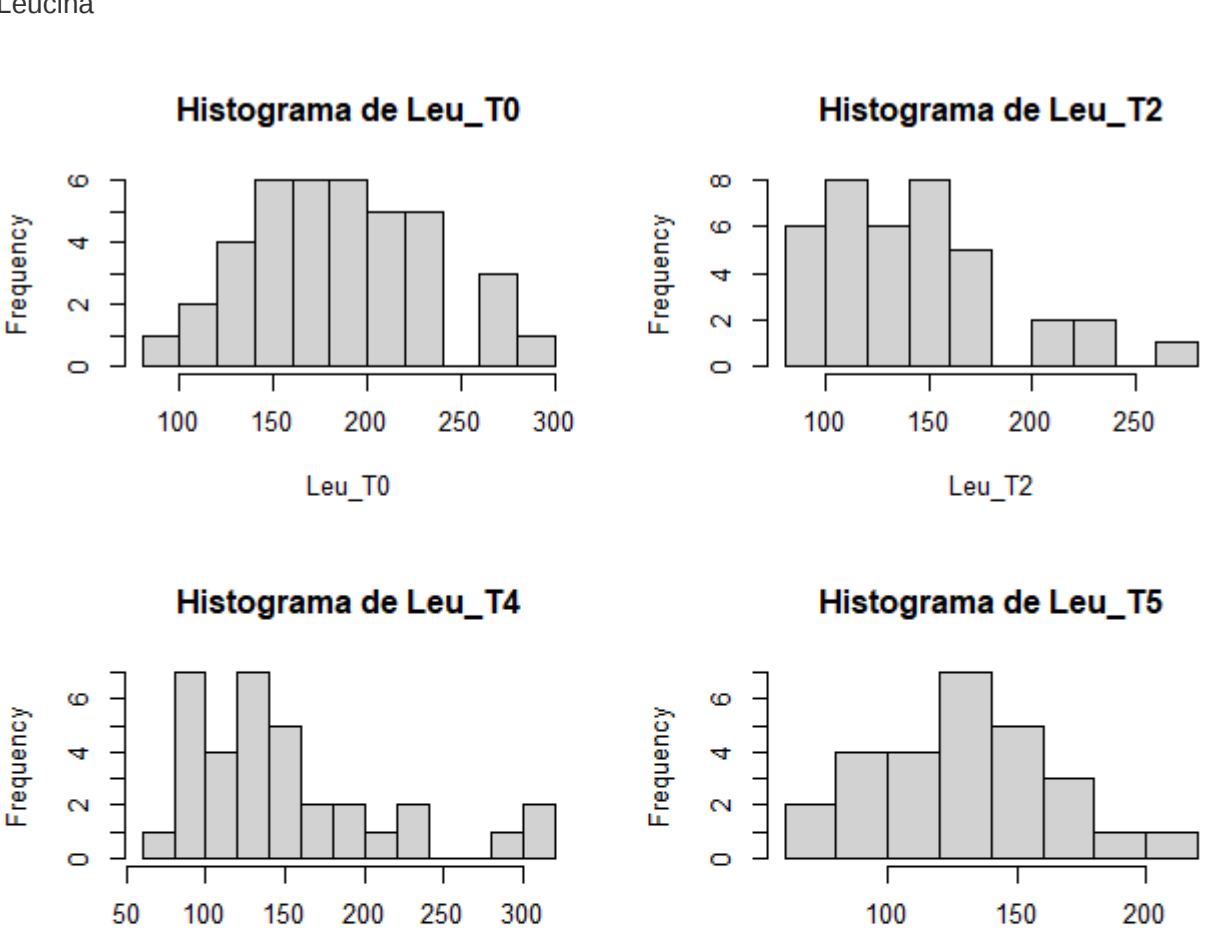
LDL



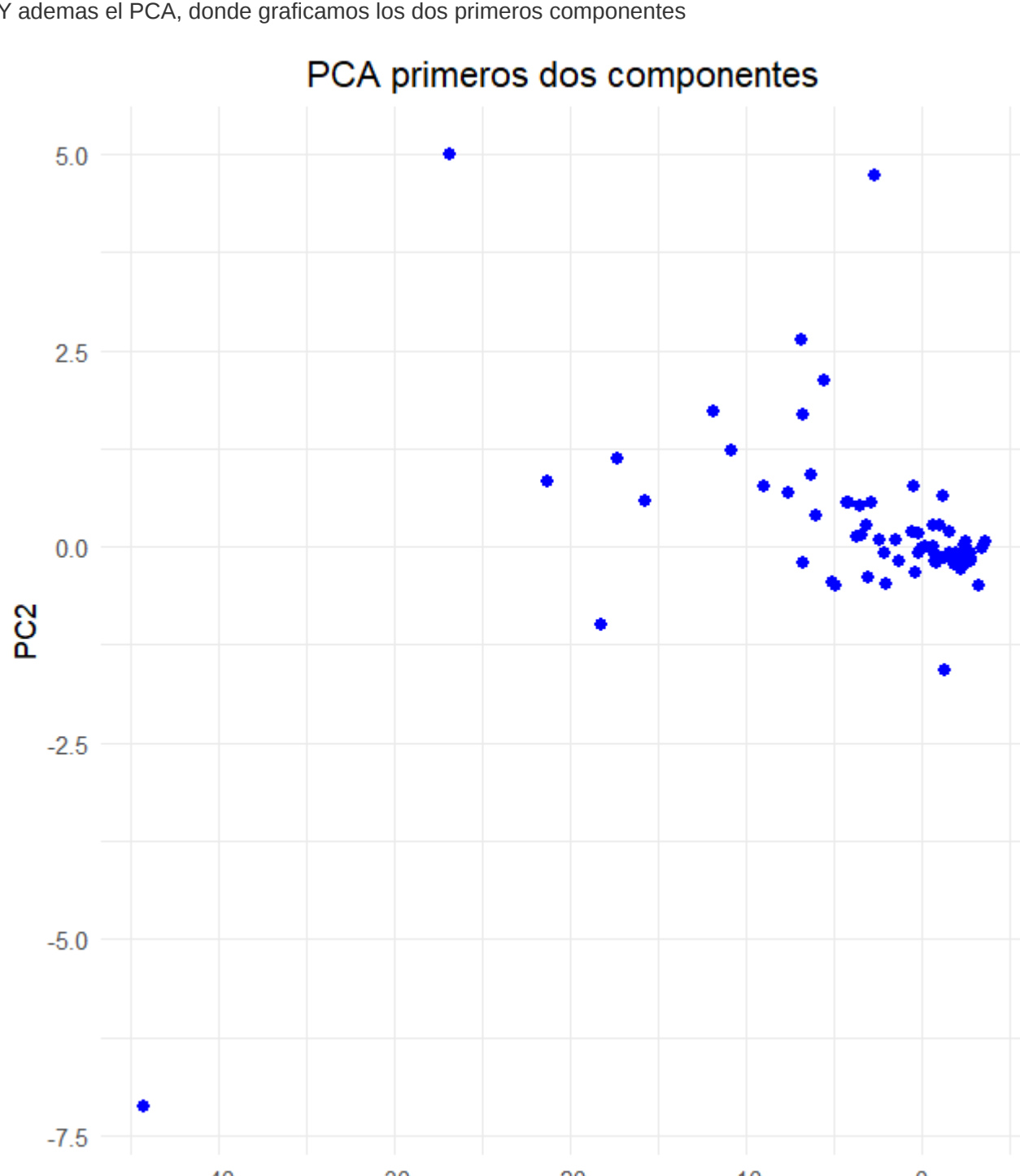
HDL



Leucina



Y además el PCA, donde graficamos los dos primeros componentes



Haciendo un análisis de las desviaciones estándar, vemos que el primer componente tiene un valor bastante más alto que el resto, siendo este el que explica más variabilidad de mis datos que el resto de componentes:

Desviaciones estándar de primeros componentes.

6.05651017	0.93691557	0.63958283	0.45302891	0.37518763	0.35427193	0.31126768
------------	------------	------------	------------	------------	------------	------------

## Discusión y limitaciones y conclusiones del estudio.

Mis limitaciones principales a la hora de realizar esta PEC han sido mi falta de conocimiento a la hora de usar el paquete SummarisedExperiment, ha consumido mucho tiempo tras mucha prueba y error, lo mismo a la hora de realizar el PCA. Sin embargo he manejado git y github de forma adecuada

Link al github: [https://github.com/AGD99/Gutierrez\\_Dominguez\\_Adrian\\_PEC1](https://github.com/AGD99/Gutierrez_Dominguez_Adrian_PEC1)