

POMA (1.5.0): Exploratory Data Analysis Report

Username

diciembre, 2021

Contents

1	Know your data	2
1.1	Summary Table	2
1.2	Samples by Group	3
2	Normalization Plots	4
3	Group Distribution Plots	5
4	Outlier Detection	6
5	High Correlated Features ($r > 0.97$)	7
6	Heatmap and Clustering	8
7	Principal Component Analysis	9

1 Know your data

- Your data have **224** samples, **113** features and **3** groups, that are ‘CRC’, ‘Healthy’, ‘Polyp’. Furthermore, **4** covariates have been found in your data. These covariates are ‘age_at_consent’, ‘gender’, ‘smoking_condition’, ‘alcohol_consumption’.
- A **0%** of values in your data are NAs (missing values).
- A **0%** of values in your data are zeros.
- Removed from the exploratory data analysis **0** features that only have zeros.
- Removed from the exploratory data analysis **0** features that have zero variance.

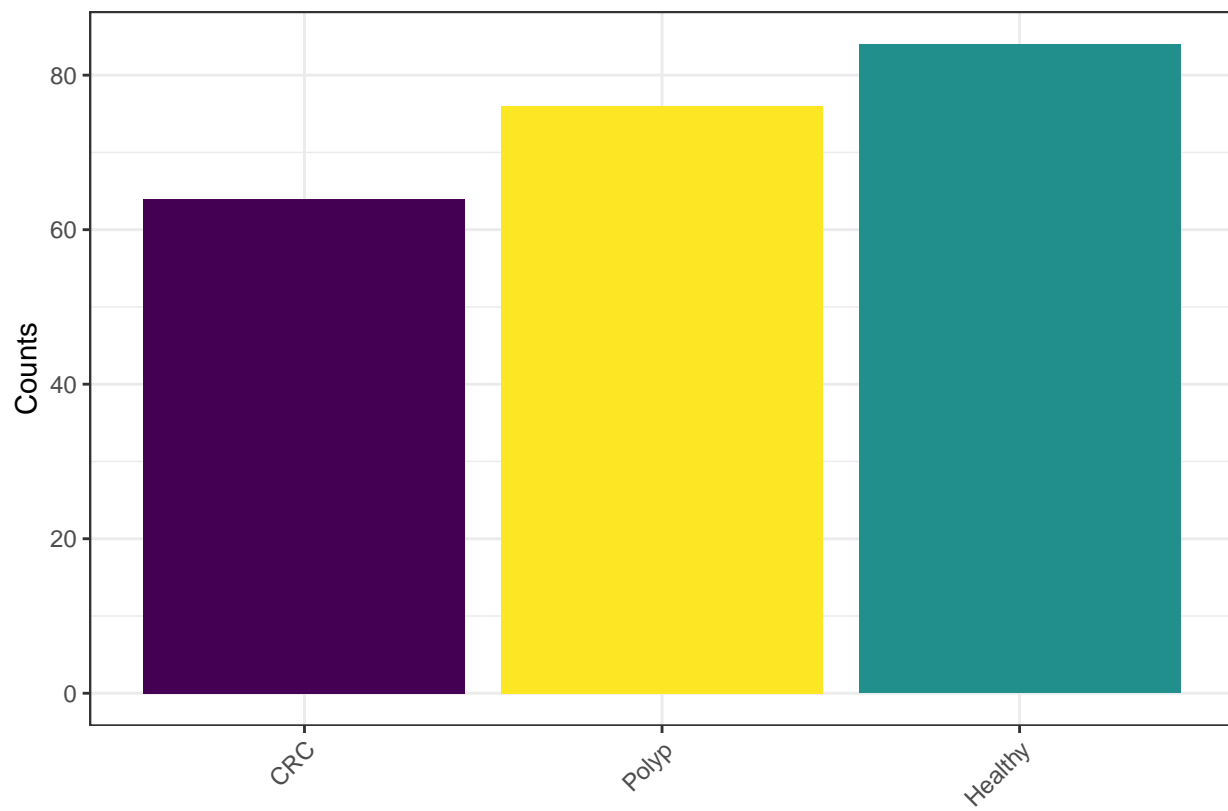
1.1 Summary Table

Samples	Features	Covariates
224	113	4

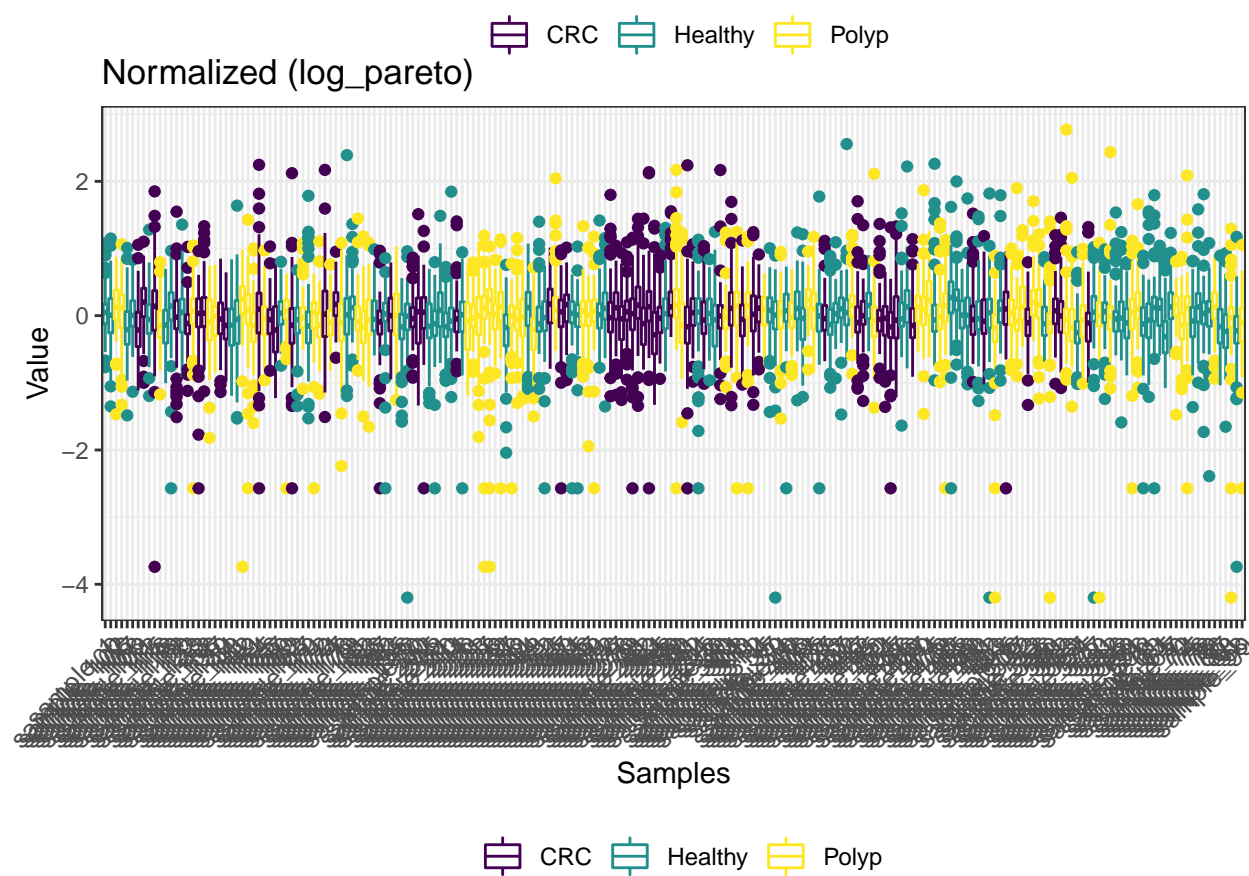
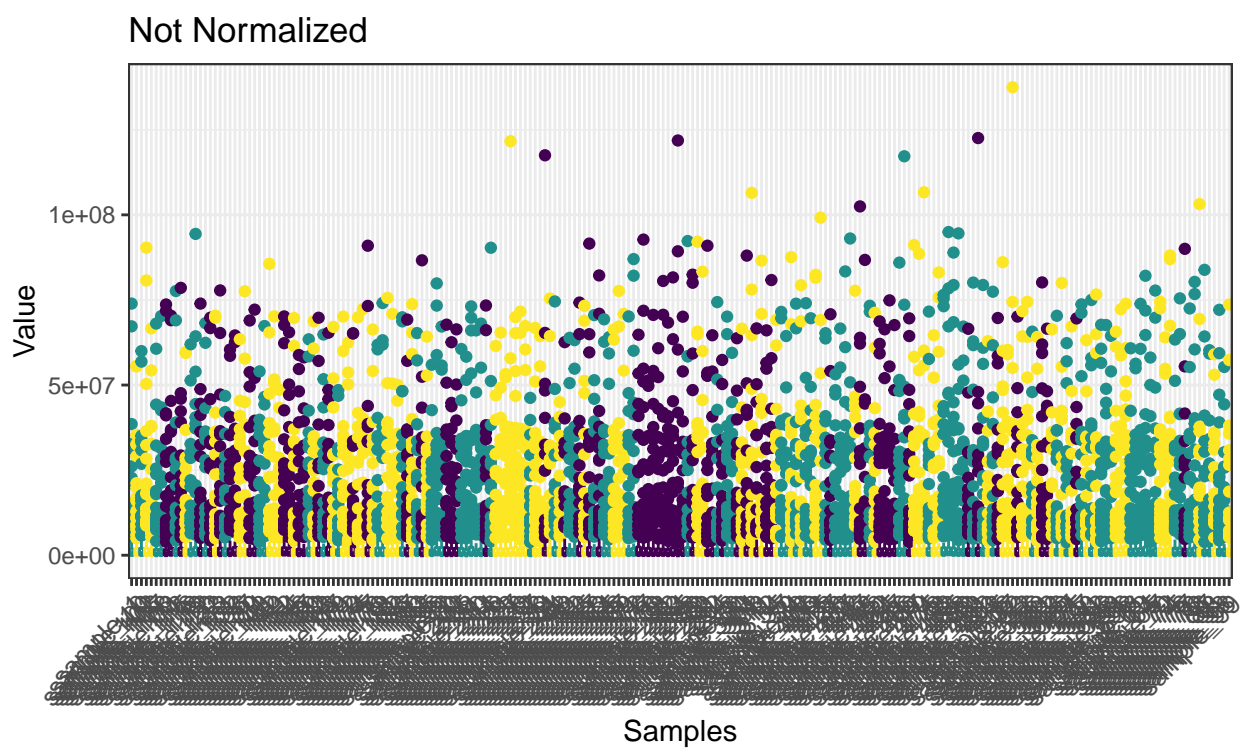
Counts_Zero	Percent_Zero
0	0 %

Counts_NA	Percent_NA
0	0 %

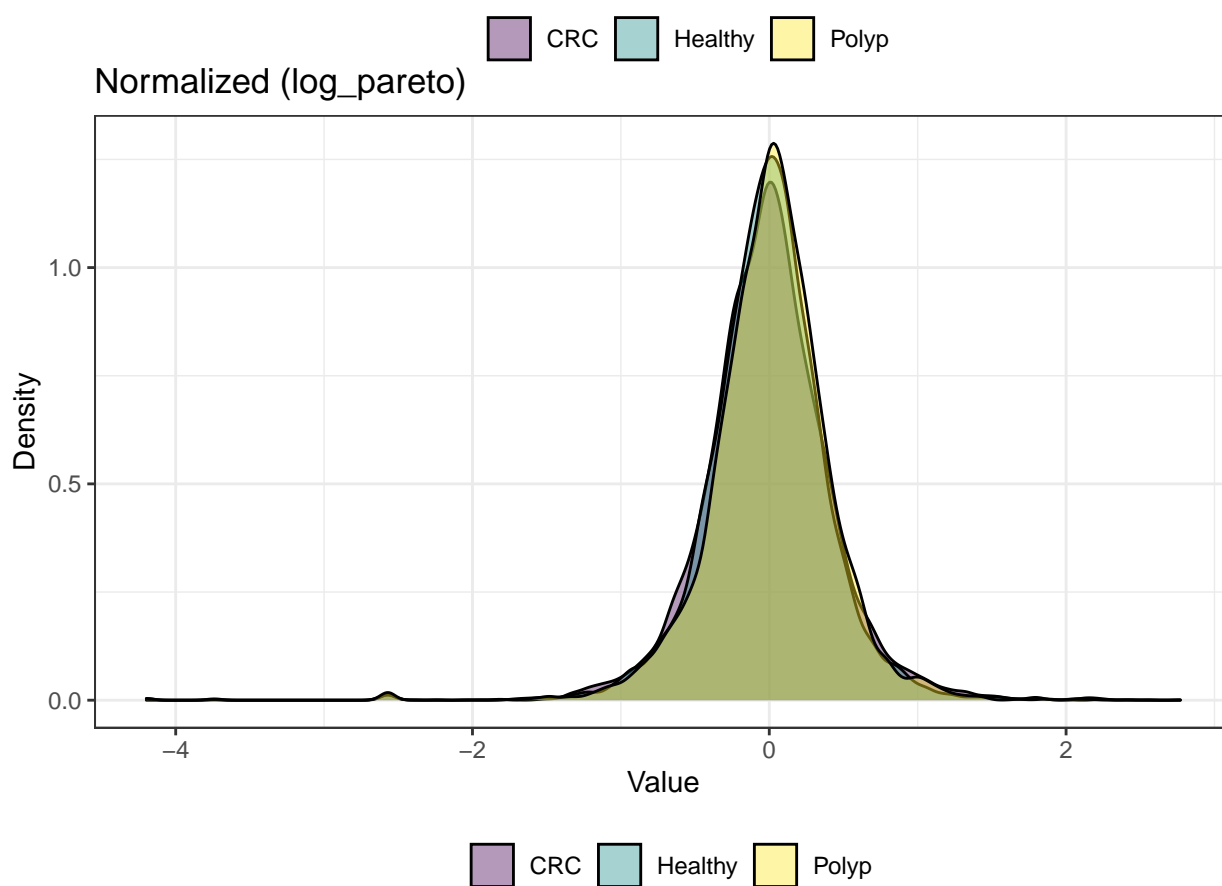
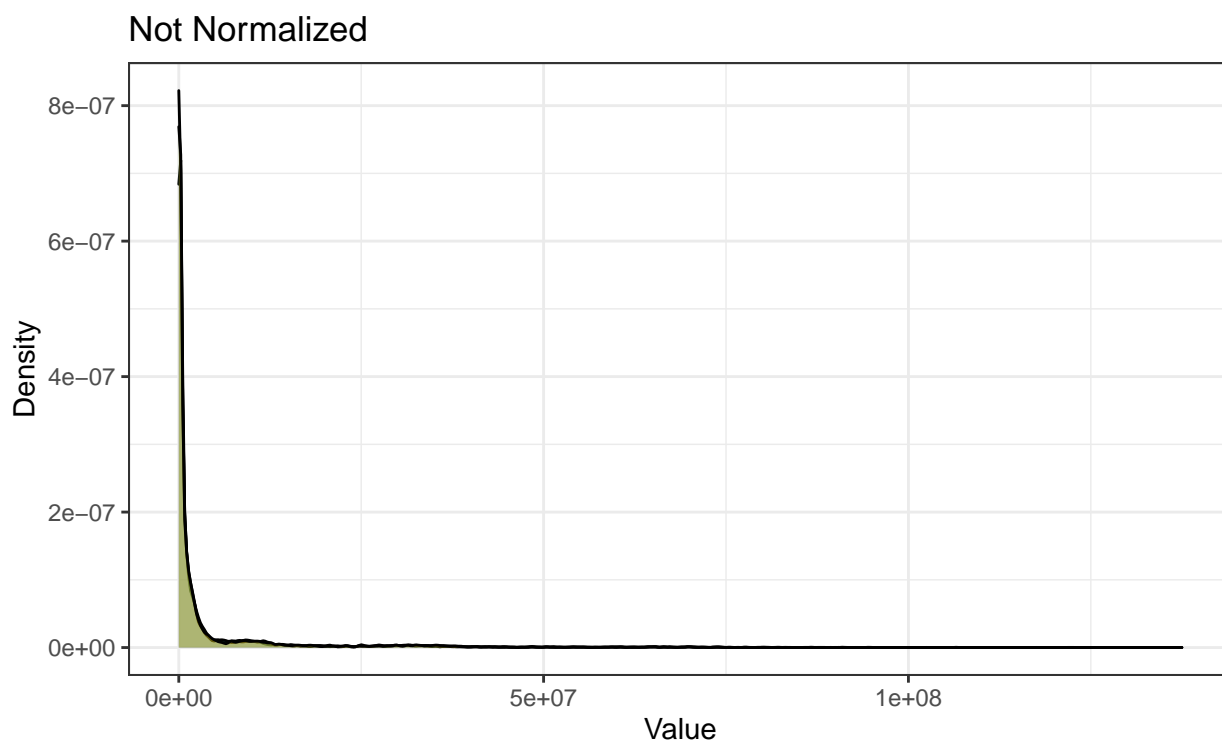
1.2 Samples by Group



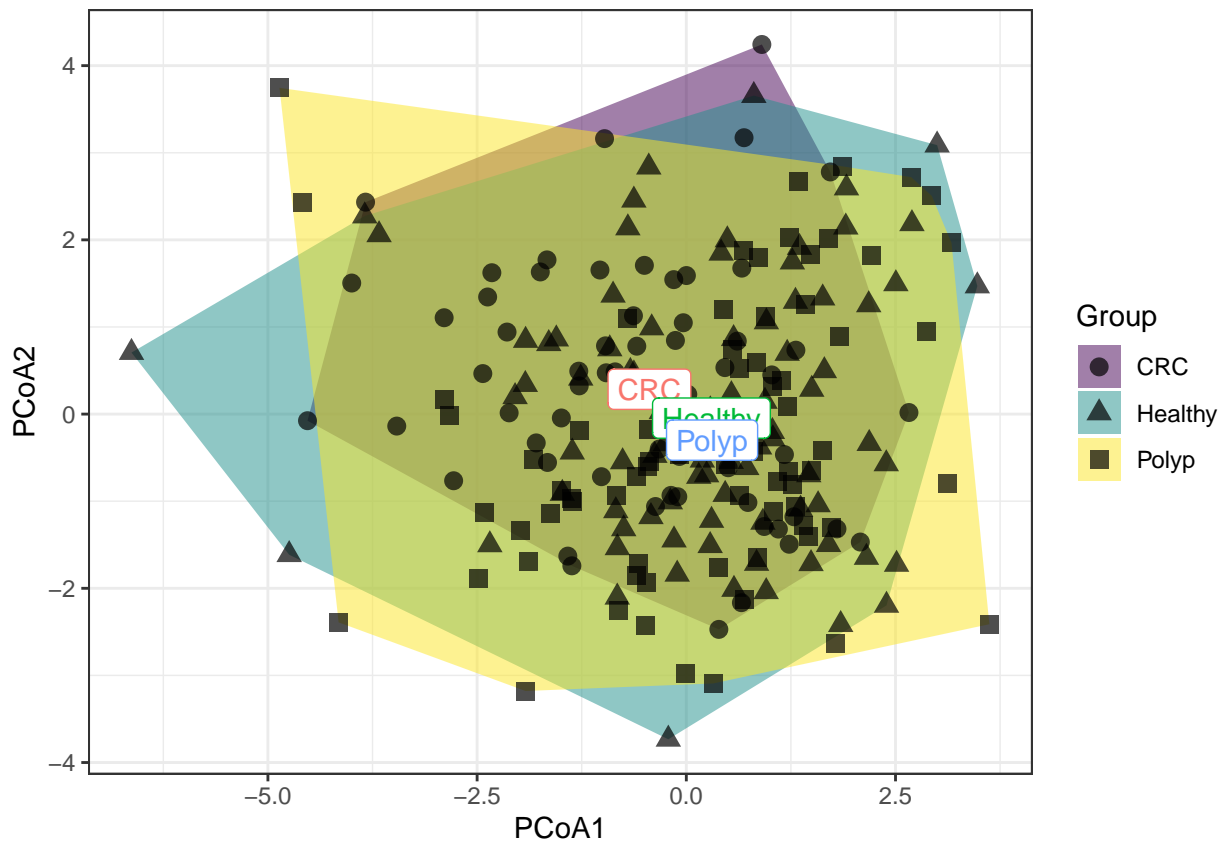
2 Normalization Plots



3 Group Distribution Plots



4 Outlier Detection



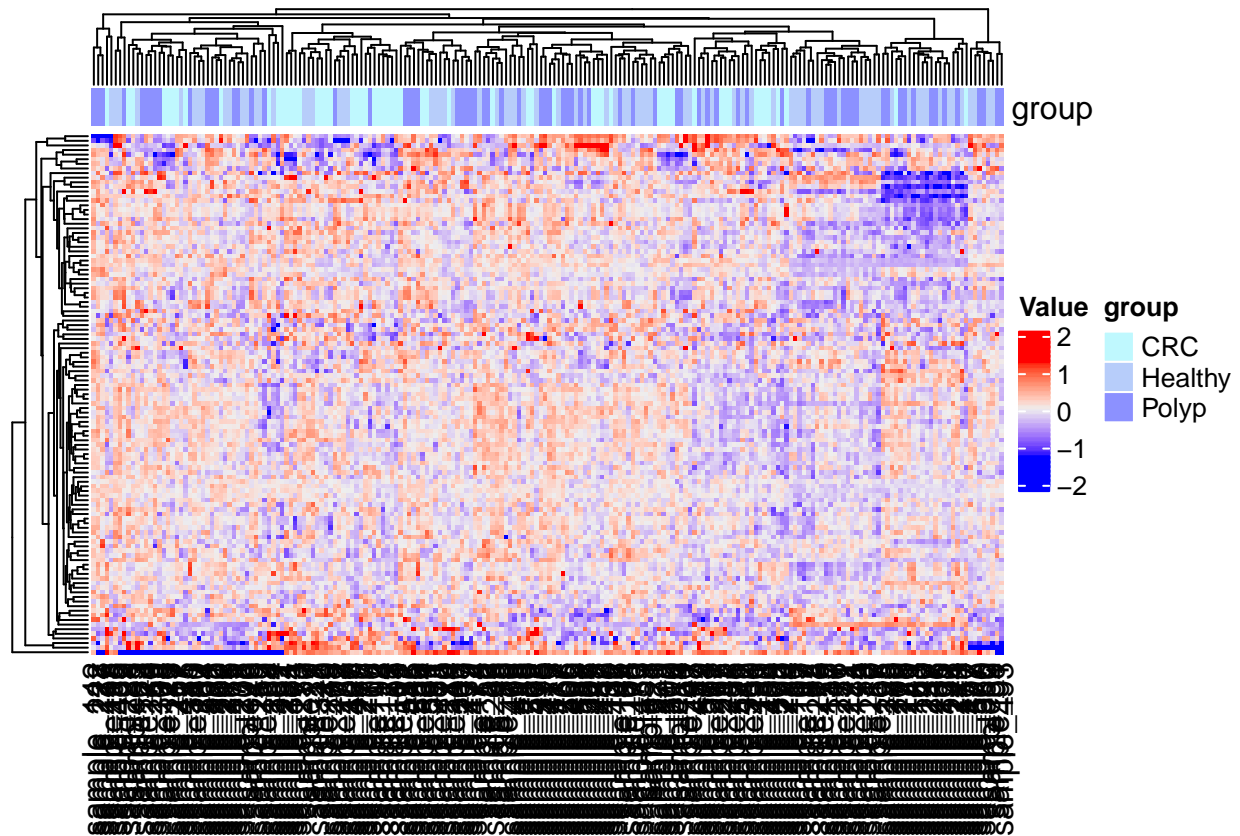
16 possible outliers detected in your data. These outliers are 'sample_133', 'sample_152', 'sample_168', 'sample_24', 'sample_250', 'sample_77', 'sample_131', 'sample_199', 'sample_258', 'sample_483', 'sample_566', 'sample_93', 'sample_10', 'sample_175', 'sample_190', 'sample_598'.

sample	group	distance_to_centroid	limit_distance
sample_133	CRC	7.685837	6.742215
sample_152	CRC	8.091521	6.742215
sample_168	CRC	6.859004	6.742215
sample_24	CRC	7.042991	6.742215
sample_250	CRC	7.036088	6.742215
sample_77	CRC	7.508360	6.742215
sample_131	Healthy	7.668372	6.801927
sample_199	Healthy	9.124536	6.801927
sample_258	Healthy	7.661949	6.801927
sample_483	Healthy	7.497505	6.801927
sample_566	Healthy	8.151007	6.801927
sample_93	Healthy	6.884839	6.801927
sample_10	Polyp	10.491807	6.845334
sample_175	Polyp	7.228950	6.845334
sample_190	Polyp	8.191032	6.845334
sample_598	Polyp	8.190130	6.845334

5 High Correlated Features ($r > 0.97$)

There are **2** high correlated feature pairs in your data. These features are **'aconitate - citraconic_acid'**, **'lactate - oxalic_acid'**.

6 Heatmap and Clustering



7 Principal Component Analysis

