Data Science Capstone: Metabolies in Health and Disease

Melba Torres

Metabolites in Health and Disease

Introduction

Metabolomics, the study of small molecules, offers a high-resolution window into an organism's phenotype. By measuring the presence and abundance of metabolites, we can gain quantitative insights into its physiological state. The data used for this project was produced by Casaro et al. their data was published along the study titled "Blood metabolomics and impacted cellular mechanisms during transition into lactation in dairy cows that develop metritis"[1]. This data can be found at the Metabolomics Workbench website under study ID ST002556.

Experimental Approach Summary

To investigate the metabolic changes associated with Metritis in Holstein dairy cows, Casaro et al. collected blood samples at three time points: prepartum (14 $\pm$ 6 days before calving), calving, and diagnosis (7 $\pm$ 2 days after calving). This dataset includes a total of 103 cows, 52 of which exhibited clinical signs of Metritis and thus were identified as "Met" as opposed to "Con" (Control). Metritis is a poly-microbial infection characterized by the presence of a reddish-brown, fetid discharge; Unlike other infections, recent research has shown that Metritis can be primarily attributed to shifts in the uterine microbiome rather the presence of specific pathogens[2]. The researchers analyzed these blood samples using untargeted gas chromatography time-of-flight mass spectrometry to obtain metabolic profiles.

Data Analysis Approach Summary

For this project, I decided to focus on "Calving", as this is the closest time point to diagnosis and the instance preceding metritis manifestations. This is also the time point that does not have a range, thus, measurements might be more comparable than for the pre-partum period which has a range of 8-20 days, which might lead to cows samples at day 8 to differ from

cows samples at day 20, even though they were grouped together based on time point and diagnosis.

Data and Approach Limitations

Metabolic profiles are made up of multiple variables that are, by the nature of metabolism, at least somewhat correlated, if not strongly. Untargeted gas chromatography time-of-flight mass spectrometry is a high-granularity approach, but it does have limitations, including a bias towards identifying more volatile compounds and incorrect identification of metabolites during downstream analysis of peaks.

It is also unlikely that all cows calved at the same time. According to the paper, samples were collected within a 24-hour range of calving. This range and the discrepancies in calving times could be large enough to introduce bias by adding time- and space-related noise to the data.

Additionally, as mentioned above, metritis is polybacterial and related to the microbiome. It is possible that multiple scenarios could lead to metritis and that cows are responding to specific microbiome products, leading to more variability in response.

Finally, just like people, each cow is different . Their behavior and personality could also be a source of noise. For example, a skittish cow might have a higher level of stress-related metabolites that are not related to infection compared to a more trusting cow. Collecting multiple samples at previous time points that allow for a "baseline" metabolic reading would be more appropriate.

Workflow Overview:

- Data Wrangling: Restructure data for statistical analysis and model fitting.
- Data transformation and scaling for for statistical analysis and model fitting.
- Group Differentiation Assessment or "is this even worth the time" test:
  Employ PERMANOVA to determine statistical differences between "Met" (metritis) and "Con" (control) groups.
- Metabolite Importance Identification:
  Utilize diverse approaches to pinpoint key metabolites:
  - Partial Least Squares Discriminant Analysis (PLS-DA)
  - Generalized Linear Model with Least Absolute Shrinkage and Selection Operator (GLM-LASSO)
  - Uni-variate analysis with t-tests
- Log-Fold Change Calculation:
  Quantify magnitude of change for identified important metabolites

Setup

```
pacman::p_load(tidyverse, performance, ggpubr, ggplot2, janitor, visdat, skimr, caret, res

fav_colors <- c("#e60200", "#e96000","#e94196","#ed5c9b","cornflowerblue", "#00cdff", "for

theme_set(theme_linedraw())
```

Data Wrangling and Exploration

The raw data file contains metabolite abundance measurements for 103 Holstein cows at the three previously mentioned time points. Each cow has a unique ID for identification. A total of 265 metabolites are included. Additionally, the dataset includes information on cow parity (primiparous or multiparous).

The data is organized so that each column represent a cow, with rows being either "factors" or metabolites.

Below are the first 6 rows and 10 columns of the raw data (with some modification for interpretability)

```
Raw_mod[1:6,1:10]
```

```
# A tibble: 6 x 10
  Samples             `10001_Calving_Con` `10038_Calving_Con` `9268_Calving_Con`
  <chr>               <chr>               <chr>               <chr>
1 Factors             Group:Con | Time:C~ Group:Con | Time:C~ Group:Con | Time:~
2 1_5-anhydroglucitol 10374               33155               30844
3 1-hexadecanol       520                 502                 879
4 1-monoolein         43073               5109                41738
5 1-monopalmitin      1054                1141                981
6 1-monostearin       1533                2390                1920
# i 6 more variables: `9280_Calving_Con` <chr>, `9509_Calving_Con` <chr>,
#   `9536_Calving_Con` <chr>, `9778_Calving_Con` <chr>,
#   `9794_Calving_Con` <chr>, `9802_Calving_Con` <chr>
```

During data wrangling, I restructured the data into an analysis-friendly version, where each row corresponds to a unique sample. Below is the annotated code.

```
# Since all the metabolite names begin with a number, add a prefix to metabolite names ("M
df_names <- paste("M",sep="_",
```

```
                df$X1[3:length(df$X1)])|>
  str_replace_all("-", "_") |>
  str_replace_all(" ","_")

# Transpose dataframe to have individuals as rows
tdf <- t(df) |>
  data.frame()

# Change column names
colnames(tdf) <- c("CowNumber", "Factor", df_names)

# Remove first row (names)
tdf <- tibble(tdf[-1,])

# Separate the items inside the column "Factors"
tdf <- separate_wider_delim(tdf, Factor, delim = " | ", names = c("Diagnosis", "Time", "Pa

# Clean up the data for the first 4 columns, rename "Group" to "Diagnosis"
tdf <- tdf %>%
  mutate(Diagnosis = str_remove_all(Diagnosis, "Group:"),
         Time = str_remove_all(Time, "Time:"),
         Parity= str_remove_all(Parity, "Parity:"),
         # Leave only the cow number
         CowNumber = str_remove_all(CowNumber, "_.*"))

tdf <- tdf |>
  mutate(across(!CowNumber & !Diagnosis & !Parity & !Time, .fns = as.numeric),
         Diagnosis = as.factor(Diagnosis),
         Time = factor(Time, ordered = T, levels =  c("Prepartum", "Calving", "Diagnosis")
         Parity = as.factor(Parity),
         CowNumber = as.factor(CowNumber))

# Remove entries where none of the metabolites have a value (equivalent to having no sampl
Cow_All <- tdf[rowSums(is.na(tdf[,5:269])) != ncol(tdf[,5:269]), ]
#dim(Cow_All) #309, 269

# Filter to only retain "calving" entries
Cow_calving <- Cow_All |>
  filter(Time == "Calving")

# First 6 rows and 10 columns
```

```
Cow_calving[1:6, 1:10]
```

```
# A tibble: 6 x 10
  CowNumber Diagnosis Time    Parity M_1_5_anhydroglucitol M_1_hexadecanol
  <fct>     <fct>     <ord>   <fct>                  <dbl>           <dbl>
1 10001     Con       Calving Mult                   10374             520
2 10038     Con       Calving Mult                   33155             502
3 9268      Con       Calving Mult                   30844             879
4 9280      Con       Calving Mult                   26916             431
5 9509      Con       Calving Mult                   23092             735
6 9536      Con       Calving Mult                   55551             489
# i 4 more variables: M_1_monoolein <dbl>, M_1_monopalmitin <dbl>,
#   M_1_monostearin <dbl>, M_2_5_dihydroxypyrazine <dbl>
```

Now we can look at the structure of the data using skim()

```
skim(Cow_calving[1:20,])
```

Table 1: Data summary

Name	Cow_calving[1:20,]
Number of rows	20
Number of columns	269
_____	
Column type frequency:	
factor	4
numeric	265
_____	
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
CowNumber	0	1	FALSE	20	100: 1, 100: 1, 926: 1, 928: 1
Diagnosis	0	1	FALSE	1	Con: 20, Met: 0
Time	0	1	TRUE	1	Cal: 20, Pre: 0, Dia: 0
Parity	0	1	FALSE	1	Mul: 20, Pri: 0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_1_5_anhydroglucitol	0	1	27432.65	9903.97	10374	21347.50	27012.50	31421.75	55551	
M_1_hexadecanol	0	1	592.95	165.05	315	498.25	552.0	682.50	903	
M_1_monoolein	0	1	27619.30	19894.79	2207	4882.25	27375.00	42071.75	70176	
M_1_monopalmitin	0	1	1149.80	198.98	886	1006.75	1097.5	1288.75	1618	
M_1_monostearin	0	1	1979.10	339.38	1463	1706.25	2007.5	2256.75	2530	
M_2_5_dihydroxypyrazine	0	1	539.75	168.94	261	421.25	513.0	633.75	995	
M_2_6_diaminopimelic_acid	0	1	441.60	159.93	258	337.00	394.5	488.50	902	
M_2_8_dihydroxyquinoline	0	1	641.35	268.26	297	446.25	604.5	745.75	1213	
M_2_aminobutyric_acid	0	1	50308.80	20030.70	20846	36899.25	48985.50	60151.25	105783	
M_2___deoxyguanosine	0	1	210.20	66.89	124	170.00	197.0	220.25	395	
M_2_deoxypentitol	0	1	889.55	304.91	537	612.50	892.0	1075.50	1671	
M_2_deoxytetronic_acid	0	1	2082.60	575.11	1159	1667.75	2103.0	2495.25	3195	
M_2_hydroxy_2_methylbutanoic_acid	0	1	4136.30	1225.04	1575	3310.25	4090.5	4994.25	6276	
M_2_hydroxybutanoic_acid	0	1	60559.45	25942.72	27756	42764.00	54782.50	77392.25	118516	
M_2_hydroxyglutaric_acid	0	1	2525.75	549.02	1727	2245.50	2430.5	2891.75	3611	
M_2_hydroxyhexanoic_acid	0	1	4785.20	4831.68	574	1449.00	2956.5	5989.00	18029	
M_2_hydroxyvaleric_acid	0	1	11280.15	3263.60	1166	10669.25	11490.50	13850.25	15916	
M_2_ketoadipic_acid	0	1	16462.60	9401.57	1790	6817.75	17976.50	21728.50	33466	
M_2_ketobutyric_acid	0	1	6656.80	954.74	5446	6065.00	6554.5	7058.00	9647	
M_2_ketoglucose_dimethylacetal	0	1	841.85	276.70	170	707.00	865.0	995.00	1306	
M_2_ketoisocaproic_acid	0	1	28757.15	6067.04	17219	24738.25	27605.00	32157.75	40095	
M_2_ketoisovaleric_acid	0	1	7258.70	4551.53	825	2614.75	7861.0	10666.25	15813	
M_2_methylglyceric_acid	0	1	1142.25	536.54	602	750.75	925.5	1511.25	2357	
M_2_monoolein	0	1	6636.30	12761.39	179	1056.75	2689.5	4447.75	47090	
M_2_monopalmitin	0	1	1641.20	411.88	983	1371.00	1545.5	1830.25	2699	
M_2_picolinic_acid	0	1	660.25	276.55	217	439.75	653.5	828.00	1183	
M_3___4_hydroxyphenyl_propionic_acid	0	1	709.90	333.20	352	556.50	633.0	808.50	1887	
M_3_aminoisobutyric_acid	0	1	7611.45	8420.59	729	1726.75	2606.0	11151.75	25154	
M_3_hydroxy_3_methylglutaric_acid	0	1	459.75	118.36	234	375.75	447.5	561.25	636	
M_3_hydroxybutyric_acid	0	1	896030.90	280169.85	526247	727127.25	916840.50	1005857.50	1786301	
M_3_hydroxypalmitic_acid	0	1	334.70	105.70	187	263.75	329.5	384.75	543	
M_3_hydroxypropionic_acid	0	1	8316.40	2561.23	5636	6478.25	7753.0	8991.75	15001	
M_3_phenyllactic_acid	0	1	1113.60	377.80	475	817.50	1081.5	1419.75	1685	
M_3_phosphoglycerate	0	1	251.50	59.01	171	211.25	237.0	293.50	421	
M_3_ureidopropionate	0	1	904.90	385.63	275	616.25	844.5	1218.50	1683	
M_4_aminobutyric_acid	0	1	283.70	180.30	58	199.00	233.5	301.25	969	
M_4_hydroxybutyric_acid	0	1	8412.75	3480.54	4317	5682.50	8494.0	9712.50	19325	
M_4_hydroxycinnamic_acid	0	1	12832.80	2977.04	8824	10494.25	11855.00	15239.50	18513	
M_4_hydroxyphenylacetic_acid	0	1	1016.15	458.21	501	697.25	957.0	1242.25	2499	
M_5_6_dihydrouracil	0	1	322.75	77.45	186	272.50	320.5	370.75	529	
M_5_aminovaleric_acid	0	1	4043.80	973.52	2154	3615.25	4192.0	4526.50	5975	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_5_hydroxynorvaline	0	1	1301.60	395.32	421	1107.75	1264.0	1495.25	2247	
M_5_methoxytryptamine	0	1	281.35	74.04	151	234.00	289.0	326.75	407	
M_9_myristoleate	0	1	11655.50	5405.45	3025	8070.50	10732.5	15229.50	25141	
M_aconitic_acid	0	1	1158.75	343.45	614	932.00	1110.0	1360.75	1783	
M_adenine	0	1	2603.75	6354.86	96	281.75	775.5	1609.50	28538	
M_adenosine	0	1	431.05	798.03	143	175.75	213.0	290.75	3770	
M_adenosine_5_monophosphate	0	1	312.50	92.63	129	233.00	323.5	377.25	474	
M_adipic_acid	0	1	5997.30	1199.49	4373	5306.25	5811.5	6381.75	9102	
M_alanine	0	1	604540.10	149979.33	339372	520934.00	554414.0	679061.00	967594	
M_alanine_alanine	0	1	5339.00	1676.31	2842	4077.75	5050.0	6039.00	9207	
M_allantoic_acid	0	1	52959.35	18164.82	23171	40936.25	52017.5	62704.75	82377	
M_alpha_aminoadipic_acid	0	1	1533.55	438.07	945	1232.50	1475.0	1666.50	2445	
M_alpha_ketoglutarate	0	1	2881.10	1556.51	151	1956.50	2920.5	3579.50	7366	
M_aminomalonate	0	1	8385.00	3813.84	2734	5695.25	7683.5	10595.75	16871	
M_arachidic_acid	0	1	6890.85	2312.88	5222	5631.75	6305.5	6818.50	13801	
M_arachidonic_acid	0	1	4747.80	1711.29	1622	3979.75	4327.5	5176.25	8985	
M_asparagine	0	1	12564.50	3629.97	8399	10058.75	10887.5	15327.75	20680	
M_aspartic_acid	0	1	8709.50	3634.39	4011	6910.75	8339.0	9373.00	19840	
M_behenic_acid	0	1	2750.55	788.29	1756	2193.25	2631.5	3117.75	5139	
M_benzoic_acid	0	1	82735.80	15015.26	56735	73446.75	82162.00	91203.50	117079	
M_beta_alanine	0	1	4032.00	519.33	2899	3694.75	4132.0	4354.75	5210	
M_beta_gentiobiose	0	1	687.70	537.90	142	267.00	591.0	759.00	2187	
M_beta_glutamic_acid	0	1	345.40	284.26	130	179.25	261.5	285.00	1049	
M_beta_glycerolphosphate	0	1	929.35	617.19	136	625.50	766.0	1041.50	3100	
M_beta_sitosterol	0	1	665.70	167.86	379	539.00	672.0	783.75	949	
M_butyrolactam	0	1	13338.10	2988.13	8931	10388.25	13848.0	14893.00	20193	
M_catechol	0	1	689.85	348.97	282	431.25	575.5	929.75	1539	
M_cellobiose	0	1	394978.45	89631.27	97078	280456.75	370295.5	514933.25	789475	
M_cerotinic_acid	0	1	600.75	182.28	319	503.75	569.0	634.75	1001	
M_cholesterol	0	1	306742.40	69701.66	175287	278219.00	296946.0	336483.00	455071	
M_cholesterone	0	1	155.95	47.89	78	119.75	151.5	193.50	247	
M_cholic_acid	0	1	7607.50	7388.52	988	2465.75	6498.0	9475.25	31993	
M_ciliatine	0	1	323.85	88.26	106	278.00	333.0	374.75	502	
M_cis_gondoic_acid	0	1	254.60	92.34	143	197.75	214.5	281.75	449	
M_citramalic_acid	0	1	1201.35	461.92	688	864.00	1158.5	1291.50	2332	
M_citric_acid	0	1	245689.55	105889.88	58003	163605.50	243477.0	283691.25	500928	
M_citrulline	0	1	17283.00	3404.47	12077	14736.25	16711.5	19937.50	22982	
M_conduritol_beta_epoxide	0	1	808.50	277.77	489	698.25	745.0	882.00	1798	
M_creatinine	0	1	330622.65	45213.72	247985	310858.50	328703.0	344042.50	439152	
M_cysteine	0	1	7621.25	2807.06	3478	5967.50	7418.5	9460.50	13221	
M_cysteine_glycine	0	1	1181.95	333.85	560	920.00	1160.0	1425.25	1679	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_cystine	0	1	4056.30	1404.11	1867	3119.00	3915.5	5015.25	7216	
M_cytidine	0	1	3492.90	1180.15	680	2998.50	3482.0	4037.25	5820	
M_cytosin	0	1	851.00	586.82	530	601.50	655.5	905.50	3190	
M_dehydroabietic_acid	0	1	1238.50	727.45	501	738.00	999.5	1365.75	3340	
M_deoxycholic_acid	0	1	763.40	618.01	234	442.75	579.5	856.00	2950	
M_D_erythro_sphingosine	0	1	310.05	98.91	182	232.50	288.5	358.00	545	
M_dihydroxyacetone	0	1	52811.30	10178.66	23074	49772.00	53448.0	58372.25	65212	
M_docosahexaenoic_acid	0	1	1440.20	344.66	870	1189.75	1525.5	1636.25	2224	
M_elaidic_acid	0	1	173267.50	98247.02	2844	6519.75	9512.0	170806.50	313752	
M_epsilon_caprolactam	0	1	1146.80	626.57	446	722.75	962.0	1302.00	2570	
M_erythritol	0	1	12794.55	5752.23	4826	8828.75	10159.5	16661.75	23779	
M_erythrose	0	1	670.60	224.34	339	545.25	616.0	809.75	1183	
M_ethanolamine	0	1	6505.60	1649.32	4154	5493.00	5883.5	7592.75	10725	
M_ethanol_phosphate	0	1	416.00	98.82	219	333.75	435.5	487.50	574	
M_fructose	0	1	236703.90	117827.30	3724	148795.25	270923.5	311124.25	429468	
M_fructose_1_phosphate	0	1	277.90	211.18	118	169.25	204.5	302.25	1065	
M_fructose_6_phosphate	0	1	201.20	64.81	119	148.75	186.5	250.50	347	
M_fucose	0	1	4363.15	711.52	3252	3930.25	4259.0	4710.75	6586	
M_fumaric_acid	0	1	2605.90	640.14	1190	2195.25	2647.0	3091.25	3624	
M_galactinol	0	1	961.90	323.36	666	727.50	882.0	1086.50	1895	
M_galactose_6_phosphate	0	1	179.45	42.57	96	151.50	183.0	202.25	264	
M_gluconic_acid	0	1	1506.10	402.99	946	1204.50	1382.0	1811.25	2243	
M_glucose	0	1	1362992.30	351077.48	780670	1154519.75	1340090.5	1544475.25	2169944	
M_glucose_1_phosphate	0	1	2048.70	1634.87	677	1280.25	1480.5	1869.75	7480	
M_glucose_6_phosphate	0	1	271.35	300.41	88	172.50	209.0	241.25	1528	
M_glutamic_acid	0	1	46272.10	13002.30	24900	36795.75	46038.05	52370.00	71533	
M_glutamine	0	1	438467.85	89270.69	266123	365535.25	452888.5	503771.00	610310	
M_glutaric_acid	0	1	779.25	199.42	503	603.50	773.5	883.25	1282	
M_glyceric_acid	0	1	13502.85	2721.20	8587	11652.75	13251.5	15662.50	18331	
M_glycerol	0	1	192682.65	58021.92	108928	140126.00	177302.0	234934.25	344597	
M_glycerol_3_galactoside	0	1	2075.15	1287.90	1176	1395.75	1679.0	2028.50	6121	
M_glycerol_alpha_phosphate	0	1	4136.40	1674.24	1684	3043.50	3676.5	4439.75	8138	
M_glycine	0	1	398196.25	133204.40	205560	312105.25	366424.0	414929.75	717146	
M_glycocyamine	0	1	1767.75	730.23	631	1062.75	1941.5	2169.00	2897	
M_glycolic_acid	0	1	12576.05	2816.89	8469	10364.25	12045.0	14519.75	18229	
M_glycyl_proline	0	1	413.45	98.43	196	363.50	412.5	487.50	547	
M_glycyl_tyrosine	0	1	403.95	120.88	259	305.25	356.0	510.50	637	
M_guanine	0	1	185.45	50.85	114	153.50	171.0	211.50	292	
M_guanosine	0	1	294.80	93.62	155	221.75	286.0	350.00	506	
M_heptadecanoic_acid	0	1	33071.50	10588.59	18209	26833.75	30563.0	37209.50	65178	
M_hippuric_acid	0	1	100607.65	39111.74	46337	64661.00	98600.5	132527.75	178350	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_histidine	0	1	74059.15	18370.13	35796	60606.00	68072.5	84041.75	116913	
M_homoserine	0	1	598.50	148.24	350	503.50	560.5	686.25	902	
M_hydrocinnamic_acid	0	1	32470.05	13924.25	3371	17910.00	37616.0	42601.00	52971	
M_hydroxycarbamate	0	1	7766.50	2225.16	2073	7285.75	8587.0	9384.25	10037	
M_hypoxanthine	0	1	523.40	191.98	188	398.00	517.5	617.75	859	
M_ile_ile	0	1	6676.00	3613.48	417	5861.00	6815.0	9051.25	13428	
M_indole_3_acetate	0	1	1210.90	510.41	520	951.25	1164.0	1266.50	2545	
M_indole_3_lactate	0	1	2371.35	946.96	869	1770.25	2226.0	2782.75	4216	
M_indole_3_propionic_acid	0	1	4848.85	3465.56	793	2675.00	3571.5	6461.50	15344	
M_indoxyl_sulfate	0	1	717.00	272.73	360	508.50	672.0	812.25	1492	
M_inosine	0	1	639.50	412.83	235	392.50	487.5	700.75	1816	
M_inositol_4_monophosphate	0	1	419.60	122.39	216	333.25	438.0	458.50	703	
M_isocitric_acid	0	1	4912.80	1692.47	2652	3922.50	4686.5	5093.75	9308	
M_isoheptadecanoic_acid	0	1	9985.15	3500.28	5418	7952.00	9359.5	11718.00	20579	
M_isoleucine	0	1	474981.20	128473.27	257598	366648.25	439965.5	575134.25	911675	
M_isolinoleic_acid	0	1	525.55	152.39	298	394.50	523.0	592.50	820	
M_isopentadecanoic_acid	0	1	20206.50	8378.37	6105	15410.75	22022.0	24165.25	35980	
M_isoribose	0	1	684.30	195.36	390	524.75	669.5	806.00	1234	
M_isothreonic_acid	0	1	5079.70	887.37	3634	4347.00	4765.5	5844.00	6497	
M_itaconic_acid	0	1	6045.40	3403.33	1393	2910.25	6113.0	8411.00	14174	
M_kynurenine	0	1	2863.85	1295.91	631	2210.50	2573.0	3414.00	6651	
M_lactamide	0	1	835.45	268.48	476	663.75	776.0	967.75	1403	
M_lactic_acid	0	1	734340.85	294894.44	196744	505191.25	649856.5	833168.25	1466775	
M_lactitol	0	1	2163.00	1141.53	513	1324.00	1933.0	2885.00	4475	
M_lactobionic_acid	0	1	476.25	479.67	127	210.00	269.0	527.25	1951	
M_lactose	0	1	541743.55	250532.26	96987	384121.00	512270.0	740093.50	1119622	
M_lanosterol	0	1	171.40	46.20	104	142.75	167.0	192.00	268	
M_lauric_acid	0	1	31186.70	8931.71	25314	28097.25	30721.5	33421.25	39950	
M_leucine	0	1	763991.90	197440.21	419288	615308.75	753934.0	882475.00	1114434	
M_lignoceric_acid	0	1	911.65	342.83	531	644.75	865.5	999.25	1856	
M_linoleic_acid	0	1	6991.75	2286.48	3290	6108.25	6571.0	7750.25	12618	
M_lithocholic_acid	0	1	246.95	46.65	162	217.00	245.5	278.00	320	
M_lysine	0	1	12310.05	3327.79	7195	9800.50	11414.5	15265.50	17711	
M_lyxitol	0	1	6267.95	1183.29	4697	5306.25	6016.5	7126.00	8738	
M_lyxose	0	1	1787.25	789.91	365	1307.00	1732.0	2172.50	4205	
M_maleimide	0	1	3820.30	698.55	1910	3486.75	3733.5	4378.50	4730	
M_malic_acid	0	1	3535.10	996.07	1697	2716.00	3447.5	4189.50	5263	
M_malonic_acid	0	1	345.45	103.07	76	297.00	346.0	416.50	524	
M_maltose	0	1	10235.20	13979.33	3042	3436.75	6032.0	9017.75	57810	
M_maltotriose	0	1	179.60	44.83	127	146.75	174.0	203.00	279	
M_mannose	0	1	135353.35	59986.58	46768	106580.75	124230.0	148184.25	260494	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_mannose_6_phosphate	0	1	205.70	66.68	140	153.00	180.5	243.25	359	
M_melibiose	0	1	229.05	97.48	122	147.25	204.5	286.50	504	
M_methanolphosphate	0	1	4486.20	1269.54	2033	3517.75	4524.5	5464.50	6784	
M_methionine	0	1	52115.10	11730.49	37322	44313.50	48931.0	58350.25	83664	
M_methionine_sulfoxide	0	1	16711.40	4895.09	7794	13289.75	17108.0	20501.00	24699	
M_methylmaleic_acid	0	1	264.80	68.09	102	234.00	267.0	302.00	390	
M_Mevalonic_acid	0	1	741.70	570.22	178	286.25	448.0	1145.00	2014	
M_myo_inositol	0	1	45795.20	11047.83	31467	34615.00	45581.0	52221.25	67417	
M_myristic_acid	0	1	14503.05	7061.38	6345	9437.25	12561.0	17701.25	32965	
M_N_acetylaspartic_acid	0	1	627.40	507.22	117	248.50	496.0	770.50	1719	
M_N_acetyl_D_galactosamine	0	1	574.55	213.32	352	471.00	546.0	605.25	1346	
M_N_acetylglycine	0	1	3216.70	2273.91	542	1193.00	3059.0	4161.00	7914	
M_N_acetylmannosamine	0	1	668.05	210.24	121	552.75	672.0	839.50	1054	
M_N_acetylornithine	0	1	5287.20	746.66	3570	4807.50	5380.0	5746.25	7082	
M_N_carbamoylaspartate	0	1	367.20	318.97	155	239.00	298.0	355.00	1640	
M_n_epsilon_trimethyllysine	0	1	714.00	229.46	437	549.75	656.5	819.50	1366	
M_nicotinamide	0	1	2241.75	1464.26	837	1337.00	1910.0	2272.50	6475	
M_nicotinic_acid	0	1	1304.30	3094.70	393	436.50	560.5	617.25	14396	
M_nonadecanoic_acid	0	1	2401.85	449.79	1608	2208.50	2362.0	2490.75	3782	
M_norvaline	0	1	5896.70	1010.67	4575	5244.00	5674.0	6317.50	9009	
M_O_acetylserine	0	1	317.75	177.05	130	200.25	276.0	322.00	720	
M_octadecanol	0	1	1419.25	152.63	1265	1344.00	1380.5	1457.50	1942	
M_octadecylglycerol	0	1	45463.40	80888.97	7194	419.50	637.0	63801.50	316563	
M_oleamide	0	1	1679.90	2515.98	154	833.00	1063.5	1406.00	12068	
M_oleic_acid	0	1	3140.75	2001.83	530	1457.25	3290.5	3985.50	8077	
M_O_phosphoserine	0	1	222.65	49.02	122	196.50	208.0	261.25	326	
M_ornithine	0	1	90142.00	26142.64	44413	70313.25	80107.5	113509.25	127888	
M_orotic_acid	0	1	326.95	93.56	164	252.25	343.0	378.00	552	
M_oxalic_acid	0	1	3158.20	1682.67	411	1989.25	3209.0	4235.25	6403	
M_oxoproline	0	1	418700.55	101652.32	248214	359450.25	421776.5	495474.75	704245	
M_palmitic_acid	0	1	161015.35	58245.32	98683	132241.25	154792.5	186204.75	310822	
M_palmitoleic_acid	0	1	4390.70	3106.84	1342	1994.00	3363.0	6160.50	13325	
M_pantothenic_acid	0	1	341.65	118.48	173	279.00	330.5	378.25	605	
M_parabanic_acid	0	1	1567.05	586.31	670	1227.00	1395.5	1819.75	3241	
M_pentadecanoic_acid	0	1	47683.70	15915.14	15056	46720.50	50806.5	57659.75	71562	
M_pentitol	0	1	311.15	346.78	65	104.25	163.5	332.75	1208	
M_phenaceturic_acid	0	1	36016.50	34329.56	6873	14935.00	24587.0	40312.00	135052	
M_phenylacetic_acid	0	1	3659.05	1567.23	1004	2680.50	3310.5	5010.25	6135	
M_phenylalanine	0	1	93225.45	24624.66	60389	76456.25	87740.0	106741.50	148584	
M_phenylethylamine	0	1	815.75	627.28	233	415.00	607.5	904.75	2578	
M_phosphate	0	1	65229.25	13293.04	44465	54462.25	63528.0	72629.00	99382	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_phosphoenolpyruvate	0	1	454.60	170.54	239	345.75	384.5	562.00	929	
M_phosphoethanolamine	0	1	668.95	314.13	212	502.75	597.5	804.00	1554	
M_p_hydroxylphenyllactic_acid	0	1	8485.05	16066.11	446	1327.75	1863.0	2830.50	64076	
M_phytanic_acid	0	1	1317.25	355.27	818	1100.75	1259.5	1502.50	2389	
M_pimelic_acid	0	1	2089.00	1064.96	536	1548.00	1960.5	2409.75	5668	
M_pipecolinic_acid	0	1	2184.70	1002.41	1037	1341.50	1842.0	3126.50	3763	
M_piperidone	0	1	412.75	307.18	165	267.00	322.5	394.25	1592	
M_proline	0	1	165262.74	77770.78	80098	130418.00	172598.5	201737.50	429854	
M_pseudo_uridine	0	1	18169.40	4072.13	12782	15132.75	17556.0	20772.00	29081	
M_p_tolyl_glucuronide	0	1	735.60	472.80	282	462.50	611.5	806.75	2282	
M_putrescine	0	1	7714.60	19612.88	1512	2206.25	2877.5	4075.25	90744	
M_pyrophosphate	0	1	4264.95	1616.34	2886	3529.00	4020.5	4305.00	10541	
M_pyrrole_2_carboxylic_acid	0	1	572.25	202.43	332	458.25	550.5	631.50	1235	
M_pyruvic_acid	0	1	26092.80	8809.06	10022	19860.75	25653.5	29787.50	49252	
M_quinolinic_acid	0	1	239.70	88.85	122	168.25	229.0	273.00	472	
M_ribitol	0	1	4501.10	3251.09	1428	2440.00	3036.5	4799.50	10450	
M_ribonic_acid	0	1	1377.40	676.54	269	696.25	1532.5	1921.25	2300	
M_ribose	0	1	7809.25	3815.25	3014	5705.00	6795.0	9206.50	19433	
M_ribose_5_phosphate	0	1	220.95	79.69	125	157.50	216.0	259.00	453	
M_ribulose_5_phosphate	0	1	197.15	54.08	117	148.00	203.0	233.50	316	
M_saccharic_acid	0	1	472.60	196.87	186	360.00	467.0	586.00	864	
M_sarcosine	0	1	69121.70	14313.53	37585	60933.25	73852.5	77418.25	90028	
M_serine	0	1	337051.70	120636.37	174505	256035.25	308042.5	367765.00	630914	
M_serotonin	0	1	1498.10	2427.55	306	463.25	654.0	989.50	9049	
M_shikimic_acid	0	1	3607.80	6104.23	1162	1507.25	1806.0	2500.25	28840	
M_sophorose	0	1	808.20	211.07	369	669.75	796.0	914.75	1318	
M_sorbitol	0	1	7922.65	3703.75	1372	5818.75	8113.5	9664.00	15958	
M_spermidine	0	1	3244.90	1691.61	853	1920.25	3089.0	4746.75	6782	
M_squalene	0	1	1523.90	542.67	664	1119.25	1584.0	1719.50	2602	
M_stearic_acid	0	1	1378027.28	386584.06	963382	1204735.50	1386205.0	1462781.25	2758592	
M_succinate_semialdehyde	0	1	1513.70	489.06	744	1174.50	1454.5	1896.50	2283	
M_succinic_acid	0	1	3653.70	893.37	2108	2951.25	3464.5	3998.25	5829	
M_sucrose	0	1	657.55	915.78	21	206.75	279.5	491.75	3137	
M_tagatose	0	1	4487.05	3067.47	858	2221.25	4153.5	6009.00	12262	
M_taurine	0	1	191.65	61.74	78	158.50	183.5	220.25	365	
M_threitol	0	1	2680.75	592.13	1655	2322.00	2568.5	3137.25	3913	
M_threonic_acid	0	1	4786.50	2068.83	2130	3349.50	4100.0	5919.75	9245	
M_threonine	0	1	56511.00	12061.32	31983	49938.75	54911.5	65574.75	74425	
M_thymidine	0	1	2316.65	510.01	1435	1960.00	2278.5	2552.75	3433	
M_thymine	0	1	848.15	226.29	420	671.25	858.5	1018.25	1219	
M_tocopherol_alpha_	0	1	10665.55	4106.58	5409	8018.75	9390.5	11945.75	20820	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_trans_4_hydroxyproline	0	1	43900.50	11897.02	26818	36398.75	40524.04	48985.75	69254	
M_trehalose	0	1	5829.60	9588.40	339	943.25	3078.5	4642.00	37123	
M_tryptophan	0	1	105999.55	57444.95	35824	80964.75	99418.01	136964.25	369210	
M_tryptophol	0	1	26408.00	5153.12	17352	23136.50	26703.03	30340.75	38684	
M_tyrosine	0	1	160433.85	54810.68	86706	131023.25	162789.01	186258.75	251048	
M_tyrosol	0	1	328.35	67.25	210	278.00	342.5	368.50	438	
M_UDP_GlcNAc	0	1	221.95	48.97	145	189.25	210.0	248.25	325	
M_undecanoic_acid	0	1	21817.35	8309.30	3465	19383.75	23370.02	28318.00	32794	
M_uracil	0	1	1175.30	448.79	633	876.25	1156.0	1321.50	2569	
M_urea	0	1	20726452	49273.66	130812	2824707	7531137	28371924	241082	
M_uric_acid	0	1	9957.60	3481.02	5497	7244.25	9231.0	11119.75	17802	
M_uridine	0	1	1123.25	298.32	540	931.75	1054.5	1409.25	1614	
M_urocanic_acid	0	1	356.10	165.29	201	250.00	313.0	372.00	912	
M_valine	0	1	771723.50	245864.04	84477	562884.75	694954.0	972794.00	1218133	
M_xanthine	0	1	200.90	46.74	129	175.50	196.0	217.00	320	
M_xanthosine	0	1	267.85	83.28	132	215.25	245.5	317.50	433	
M_xylitol	0	1	9406.45	1870.95	5785	8240.50	9247.0	10196.00	12913	
M_xylose	0	1	12964.60	2458.14	8496	11866.00	12310.5	14551.25	17065	
M_xylulose	0	1	1683.35	359.97	937	1456.00	1660.0	1892.25	2336	

There is significant skewness in the data, although this is expected for this kind of data, skewness can cause issues for analysis. To address this, I log transformed the data and opted for a log2, rather than the more common log10, to try to preserve more subtle biological signals that might be masked by a log10 transformation. Additionally, I used the autoscale() function to normalize the data by subtracting each observation by the mean and dividing it by the standard deviation. This is appropriate for this analysis because I am more interested in the *changes* in metabolites rather than amount or presence.

```
# Log Transform
Calving_log <- cbind(Cow_calving[1:4],
                log2(Cow_calving[5:ncol(Cow_All)]))

# Scale
Calving_scaled <- cbind(Calving_log[1:4],
                  scale(Calving_log[5:ncol(Cow_All)]))

# Check the structure again
skim(Calving_scaled[1:20,])
```

Table 4: Data summary

Name	Calving_scaled[1:20,]
Number of rows	20
Number of columns	269
_____	
Column type frequency:	
factor	4
numeric	265
_____	
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
CowNumber	0	1	FALSE	20	100: 1, 100: 1, 926: 1, 928: 1
Diagnosis	0	1	FALSE	1	Con: 20, Met: 0
Time	0	1	TRUE	1	Cal: 20, Pre: 0, Dia: 0
Parity	0	1	FALSE	1	Mul: 20, Pri: 0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_1_5_anhydroglucitol	0	1	-0.51	1.17	-3.37	-1.10	-0.36	0.11	1.91	
M_1_hexadecanol	0	1	-0.04	0.68	-1.44	-0.36	-0.12	0.38	1.04	
M_1_monoolein	0	1	0.06	1.02	-1.81	-1.10	0.45	0.83	1.29	
M_1_monopalmitin	0	1	0.25	0.96	-1.19	-0.45	0.05	0.99	2.31	
M_1_monostearin	0	1	-0.07	0.62	-1.08	-0.54	0.03	0.45	0.85	
M_2_5_dihydroxypyrazine	0	1	-0.39	1.04	-2.67	-1.07	-0.41	0.30	1.81	
M_2_6_diaminopimelic_acid	0	1	-0.15	0.84	-1.44	-0.73	-0.31	0.26	1.89	
M_2_8_dihydroxyquinoline	0	1	0.46	0.85	-0.93	-0.11	0.50	0.92	1.91	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_2_aminobutyric_acid	0	1	-0.16	0.72	-1.63	-0.59	-0.07	0.30	1.34	
M_2___deoxyguanosine	0	1	-0.40	0.74	-1.69	-0.85	-0.46	-0.17	1.38	
M_2_deoxypentitol	0	1	0.09	1.00	-1.27	-0.87	0.26	0.82	2.15	
M_2_deoxytetronic_acid	0	1	0.19	0.74	-1.21	-0.28	0.32	0.76	1.39	
M_2_hydroxy_2_methylbutanoic_acid	0	1	0.63	0.80	-1.56	0.21	0.72	1.19	1.74	
M_2_hydroxybutanoic_acid	0	1	-0.67	0.63	-1.70	-1.06	-0.69	-0.18	0.45	
M_2_hydroxyglutaric_acid	0	1	-0.37	0.60	-1.36	-0.63	-0.41	0.07	0.68	
M_2_hydroxyhexanoic_acid	0	1	0.25	0.83	-1.26	-0.44	0.19	0.82	1.79	
M_2_hydroxyvaleric_acid	0	1	-0.07	1.59	-6.39	0.02	0.24	0.78	1.18	
M_2_ketoadipic_acid	0	1	0.11	1.07	-2.66	-0.81	0.53	0.80	1.39	
M_2_ketobutyric_acid	0	1	0.14	0.42	-0.46	-0.13	0.12	0.35	1.32	
M_2_ketoglucose_dimethylacetal	0	1	0.45	0.79	-2.25	-0.27	0.63	0.87	1.36	
M_2_ketoisocaproic_acid	0	1	-0.03	0.97	-2.25	-0.62	-0.12	0.57	1.56	
M_2_ketoisovaleric_acid	0	1	-0.41	1.23	-3.23	-1.52	0.09	0.55	1.14	
M_2_methylglyceric_acid	0	1	0.05	0.93	-1.13	-0.66	-0.21	0.85	1.81	
M_2_monoolein	0	1	0.11	0.86	-1.34	-0.35	0.21	0.50	1.86	
M_2_monopalmitin	0	1	0.04	0.48	-0.91	-0.26	-0.02	0.31	1.08	
M_2_picolinic_acid	0	1	0.25	1.06	-2.11	-0.48	0.44	0.99	1.81	
M_3___4_hydroxyphenyl_propionic_acid	0	1	0.44	1.26	-1.58	-0.10	0.32	1.12	3.86	
M_3_aminoisobutyric_acid	0	1	0.44	1.18	-1.31	-0.43	-0.02	1.46	2.29	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_3_hydroxy_3_methylglutaric_acid	0	1	0.24	0.80	-1.63	-0.25	0.26	0.92	1.28	
M_3_hydroxybutyric_acid	0	1	-0.07	0.76	-1.30	-0.54	-0.22	0.36	1.95	
M_3_hydroxypalmitic_acid	0	1	-0.39	1.05	-2.14	-1.02	-0.29	0.22	1.35	
M_3_hydroxypropionic_acid	0	1	0.23	0.69	-0.67	-0.32	0.14	0.52	1.83	
M_3_phenyllactic_acid	0	1	0.11	1.01	-2.05	-0.57	0.18	0.94	1.41	
M_3_phosphoglycerate	0	1	0.37	0.84	-0.99	-0.20	0.24	1.04	2.39	
M_3_ureidopropionate	0	1	0.05	1.15	-2.57	-0.63	0.12	1.00	1.78	
M_4_aminobutyric_acid	0	1	0.19	0.67	-1.71	-0.10	0.10	0.43	1.95	
M_4_hydroxybutyric_acid	0	1	0.57	0.70	-0.53	-0.03	0.71	0.96	2.23	
M_4_hydroxycinnamic_acid	0	1	-0.17	0.71	-1.25	-0.71	-0.33	0.44	1.05	
M_4_hydroxyphenylacetic_acid	0	1	0.28	0.94	-1.22	-0.43	0.32	0.94	2.61	
M_5_6_dihydrouracil	0	1	-0.08	1.15	-2.57	-0.76	0.01	0.71	2.40	
M_5_aminovaleric_acid	0	1	0.06	0.99	-2.21	-0.25	0.31	0.60	1.66	
M_5_hydroxynorvaline	0	1	-0.07	1.28	-4.06	-0.49	0.00	0.63	2.13	
M_5_methoxytryptamine	0	1	-0.17	1.12	-2.41	-0.74	0.07	0.54	1.38	
M_9_myristoleate	0	1	-0.36	0.88	-2.47	-0.79	-0.33	0.29	1.15	
M_aconitic_acid	0	1	0.16	0.93	-1.65	-0.38	0.16	0.78	1.61	
M_adenine	0	1	-0.02	1.27	-1.94	-0.96	-0.04	0.63	3.24	
M_adenosine	0	1	0.13	1.37	-1.01	-0.63	-0.28	0.29	5.00	
M_adenosine_5_monophosphate	0	1	0.16	1.22	-2.94	-0.75	0.46	1.03	1.87	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_adipic_acid	0	1	0.36	0.97	-1.19	-0.19	0.28	0.77	2.60	
M_alanine	0	1	0.09	0.88	-1.90	-0.35	-0.12	0.61	1.90	
M_alanine_alanine	0	1	0.48	1.15	-1.69	-0.35	0.44	1.11	2.68	
M_allantoic_acid	0	1	0.64	0.88	-1.11	0.20	0.75	1.18	1.81	
M_alpha_aminoadipic_acid	0	1	-0.14	0.87	-1.60	-0.74	-0.15	0.25	1.50	
M_alpha_ketoglutarate	0	1	-0.25	1.43	-5.11	-0.57	0.14	0.50	1.79	
M_aminomalonate	0	1	0.27	1.16	-2.06	-0.37	0.32	1.06	2.13	
M_arachidic_acid	0	1	0.02	1.46	-1.32	-0.90	-0.26	0.17	4.12	
M_arachidonic_acid	0	1	-0.20	1.00	-2.91	-0.51	-0.28	0.20	1.67	
M_asparagine	0	1	0.45	0.88	-0.76	-0.16	0.10	1.22	2.21	
M_aspartic_acid	0	1	0.09	0.93	-1.58	-0.29	0.16	0.44	2.22	
M_behenic_acid	0	1	0.25	1.18	-1.62	-0.62	0.20	0.96	3.21	
M_benzoic_acid	0	1	0.55	0.85	-1.13	0.07	0.59	1.08	2.24	
M_beta_alanine	0	1	-0.22	0.50	-1.44	-0.52	-0.10	0.10	0.78	
M_beta_gentiobiose	0	1	0.03	1.17	-2.00	-1.02	0.21	0.59	2.23	
M_beta_glutamic_acid	0	1	-0.06	0.96	-1.23	-0.73	-0.15	-0.02	1.99	
M_beta_glycerolphosphate	0	1	-0.29	1.06	-3.23	-0.66	-0.32	0.20	2.04	
M_beta_sitosterol	0	1	0.13	1.03	-1.93	-0.56	0.29	0.89	1.63	
M_butyrolactam	0	1	-0.02	0.47	-0.80	-0.49	0.11	0.26	0.90	
M_catechol	0	1	0.26	0.86	-1.15	-0.39	0.12	1.01	1.92	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_cellobiose	0	1	-0.40	1.25	-4.73	-0.70	-0.22	0.35	1.09	
M_cerotinic_acid	0	1	-0.08	1.03	-2.13	-0.55	-0.12	0.25	1.84	
M_cholesterol	0	1	-0.11	0.88	-2.06	-0.37	-0.13	0.33	1.43	
M_cholesterone	0	1	0.11	1.13	-2.19	-0.66	0.17	1.04	1.91	
M_cholic_acid	0	1	0.00	0.91	-1.58	-0.72	0.23	0.60	1.77	
M_ciliatine	0	1	-0.08	0.90	-3.00	-0.38	0.12	0.44	1.24	
M_cis_gondoic_acid	0	1	-0.11	1.07	-1.81	-0.76	-0.49	0.39	1.91	
M_citramalic_acid	0	1	0.00	0.86	-1.20	-0.65	0.06	0.33	1.77	
M_citric_acid	0	1	0.66	0.74	-1.44	0.20	0.79	1.02	1.88	
M_citrulline	0	1	0.39	0.78	-0.94	-0.16	0.33	1.03	1.58	
M_conduritol_beta_epoxide	0	1	-0.10	0.94	-1.60	-0.44	-0.23	0.32	2.64	
M_creatinine	0	1	0.12	0.60	-1.13	-0.11	0.14	0.34	1.44	
M_cysteine	0	1	0.19	0.88	-1.37	-0.21	0.28	0.82	1.55	
M_cysteine_glycine	0	1	-0.27	0.93	-2.46	-0.92	-0.20	0.43	0.94	
M_cystine	0	1	0.12	0.86	-1.56	-0.35	0.18	0.76	1.62	
M_cytidine	0	1	-0.34	1.00	-3.76	-0.52	-0.18	0.14	0.94	
M_cytosin	0	1	0.50	1.18	-0.51	-0.16	0.09	1.00	4.55	
M_dehydroabietic_acid	0	1	0.23	1.04	-1.37	-0.57	0.06	0.70	2.56	
M_deoxycholic_acid	0	1	-0.26	0.98	-1.90	-0.84	-0.40	0.25	2.30	
M_D_erythro_sphingosine	0	1	-0.52	0.96	-2.06	-1.29	-0.62	0.07	1.39	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_dihydroxyacetone	0	1	0.05	0.82	-2.74	-0.08	0.17	0.47	0.85	
M_docosahexaenoic_acid	0	1	-0.37	0.91	-2.10	-0.96	-0.06	0.20	1.32	
M_elaidic_acid	0	1	0.18	1.19	-1.17	-0.65	-0.43	1.16	2.24	
M_epsilon_caprolactam	0	1	0.17	0.98	-1.46	-0.51	0.06	0.66	2.01	
M_erythritol	0	1	-0.47	0.86	-2.18	-1.01	-0.73	0.23	0.92	
M_erythrose	0	1	0.44	1.12	-1.77	-0.11	0.31	1.27	2.59	
M_ethanolamine	0	1	-0.18	1.09	-2.09	-0.82	-0.51	0.64	2.21	
M_ethanol_phosphate	0	1	-0.15	0.87	-2.24	-0.80	0.11	0.49	1.05	
M_fructose	0	1	-0.12	0.97	-2.45	-0.52	0.32	0.51	0.95	
M_fructose_1_phosphate	0	1	0.08	1.05	-1.26	-0.56	-0.20	0.57	3.02	
M_fructose_6_phosphate	0	1	-0.05	1.11	-1.74	-0.95	-0.15	0.89	2.05	
M_fucose	0	1	-0.78	0.70	-2.08	-1.21	-0.84	-0.38	1.17	
M_fumaric_acid	0	1	-0.02	0.79	-2.21	-0.43	0.12	0.57	1.03	
M_galactinol	0	1	-0.18	0.82	-1.08	-0.83	-0.30	0.29	1.84	
M_galactose_6_phosphate	0	1	0.06	0.94	-2.18	-0.47	0.24	0.62	1.62	
M_gluconic_acid	0	1	-0.37	0.72	-1.56	-0.89	-0.52	0.22	0.81	
M_glucose	0	1	-0.56	0.84	-2.22	-1.12	-0.54	0.01	1.33	
M_glucose_1_phosphate	0	1	-0.30	1.29	-2.27	-0.89	-0.57	-0.07	2.94	
M_glucose_6_phosphate	0	1	0.25	1.34	-2.03	-0.35	0.13	0.49	5.10	
M_glutamic_acid	0	1	-0.05	0.94	-1.93	-0.67	0.05	0.47	1.48	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_glutamine	0	1	0.15	0.79	-1.64	-0.45	0.35	0.75	1.47	
M_glutaric_acid	0	1	0.01	0.87	-1.42	-0.78	0.08	0.55	1.85	
M_glyceric_acid	0	1	-0.12	0.71	-1.59	-0.55	-0.11	0.46	1.00	
M_glycerol	0	1	-0.68	0.89	-1.98	-1.34	-0.74	-0.02	0.96	
M_glycerol_3_galactoside	0	1	0.21	1.39	-1.23	-0.69	-0.10	0.50	4.00	
M_glycerol_alpha_phosphate	0	1	0.27	0.92	-1.77	-0.31	0.15	0.61	2.10	
M_glycine	0	1	0.39	0.99	-1.15	-0.24	0.27	0.67	2.44	
M_glycocyamine	0	1	0.12	1.01	-1.82	-0.74	0.52	0.76	1.36	
M_glycolic_acid	0	1	-0.18	0.80	-1.51	-0.78	-0.25	0.42	1.24	
M_glycyl_proline	0	1	-0.17	0.69	-1.99	-0.42	-0.09	0.33	0.63	
M_glycyl_tyrosine	0	1	-0.51	0.72	-1.49	-1.09	-0.72	0.17	0.71	
M_guanine	0	1	0.36	1.10	-1.52	-0.29	0.16	1.05	2.39	
M_guanosine	0	1	-0.01	0.85	-1.58	-0.63	0.04	0.57	1.55	
M_heptadecanoic_acid	0	1	-0.63	0.82	-2.14	-1.08	-0.73	-0.19	1.35	
M_hippuric_acid	0	1	0.56	0.80	-1.22	-0.13	0.67	1.24	1.81	
M_histidine	0	1	0.07	0.81	-0.85	-0.55	-0.14	0.62	1.80	
M_homoserine	0	1	0.24	0.64	-1.08	-0.13	0.14	0.67	1.38	
M_hydrocinnamic_acid	0	1	0.59	0.69	-1.33	-0.06	0.91	1.06	1.31	
M_hydroxycarbamate	0	1	-0.18	1.13	-3.66	-0.18	0.27	0.52	0.70	
M_hypoxanthine	0	1	0.53	1.05	-1.88	-0.02	0.69	1.14	1.98	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_ile_ile	0	1	0.14	0.90	-1.94	0.30	0.44	0.68	1.02	
M_indole_3_acetate	0	1	0.22	0.88	-1.54	-0.16	0.30	0.49	2.09	
M_indole_3_lactate	0	1	-0.07	1.07	-2.41	-0.61	-0.03	0.54	1.59	
M_indole_3_propionic_acid	0	1	0.71	0.94	-1.40	0.22	0.60	1.39	2.55	
M_indoxyl_sulfate	0	1	-0.50	0.74	-1.79	-1.08	-0.50	-0.11	1.15	
M_inosine	0	1	-0.01	0.84	-1.31	-0.52	-0.19	0.37	1.84	
M_inositol_4_monophosphate	0	1	0.27	0.67	-1.09	-0.14	0.46	0.56	1.49	
M_isocitric_acid	0	1	0.47	0.77	-0.88	0.06	0.48	0.68	2.12	
M_isoheptadecanoic_acid	0	1	-0.64	0.86	-2.07	-1.10	-0.67	-0.09	1.35	
M_isoleucine	0	1	0.46	0.93	-1.34	-0.32	0.31	1.23	1.97	
M_isolinoleic_acid	0	1	-0.17	0.79	-1.63	-0.86	-0.08	0.27	1.16	
M_isopentadecanoic_acid	0	1	-0.36	0.49	-1.38	-0.53	-0.18	-0.09	0.28	
M_isoribose	0	1	-0.22	0.83	-1.78	-0.90	-0.18	0.37	1.64	
M_isothreonic_acid	0	1	-0.34	0.70	-1.62	-0.90	-0.54	0.28	0.70	
M_itaconic_acid	0	1	-0.12	1.13	-2.33	-1.07	-0.21	0.76	1.66	
M_kynurenine	0	1	0.37	0.97	-2.41	-0.07	0.36	0.92	2.24	
M_lactamide	0	1	-0.39	0.81	-1.72	-0.86	-0.46	0.11	1.07	
M_lactic_acid	0	1	-0.20	0.76	-1.24	-0.84	-0.31	0.20	1.38	
M_lactitol	0	1	-0.39	0.94	-2.53	-0.97	-0.34	0.31	1.04	
M_lactobionic_acid	0	1	-0.37	1.05	-1.73	-1.05	-0.70	0.21	2.00	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_lactose	0	1	-0.43	1.25	-5.05	-0.71	-0.24	0.36	1.04	
M_lanosterol	0	1	0.06	0.91	-1.54	-0.45	0.08	0.56	1.71	
M_lauric_acid	0	1	0.15	0.93	-1.38	-0.59	0.09	0.73	2.08	
M_leucine	0	1	0.35	1.11	-1.96	-0.40	0.43	1.08	2.03	
M_lignoceric_acid	0	1	0.05	0.71	-0.96	-0.55	0.06	0.36	1.64	
M_linoleic_acid	0	1	-0.31	1.00	-2.40	-0.56	-0.34	0.15	1.59	
M_lithocholic_acid	0	1	0.03	0.76	-1.55	-0.41	0.07	0.55	1.10	
M_lysine	0	1	0.45	0.89	-1.14	-0.16	0.32	1.25	1.72	
M_lyxitol	0	1	-0.24	0.77	-1.38	-0.87	-0.35	0.36	1.21	
M_lyxose	0	1	-0.22	0.86	-2.79	-0.59	-0.10	0.29	1.43	
M_maleimide	0	1	-0.12	0.84	-2.81	-0.41	-0.14	0.50	0.80	
M_malic_acid	0	1	-0.05	0.90	-2.15	-0.73	0.00	0.58	1.28	
M_malonic_acid	0	1	-0.04	0.90	-3.22	-0.24	0.09	0.50	1.00	
M_maltose	0	1	0.31	1.12	-1.76	-0.36	0.29	0.76	2.92	
M_maltotriose	0	1	-0.17	0.89	-1.35	-0.82	-0.18	0.39	1.57	
M_mannose	0	1	-0.42	0.65	-1.45	-0.93	-0.55	-0.11	1.30	
M_mannose_6_phosphate	0	1	0.46	1.17	-0.86	-0.52	0.12	1.28	2.80	
M_melibiose	0	1	0.04	1.13	-1.54	-1.01	-0.09	0.90	2.51	
M_methanolphosphate	0	1	-0.37	0.98	-2.68	-0.99	-0.21	0.38	1.05	
M_methionine	0	1	0.30	0.80	-0.89	-0.23	0.14	0.81	2.18	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_methionine_sulfoxide	0	1	0.13	0.73	-1.46	-0.28	0.29	0.69	1.11	
M_methylmaleic_acid	0	1	-0.09	1.26	-3.93	-0.45	0.10	0.62	1.69	
M_Mevalonic_acid	0	1	-0.04	0.90	-1.40	-0.84	-0.31	0.80	1.47	
M_myo_inositol	0	1	-0.65	0.60	-1.51	-1.28	-0.60	-0.26	0.37	
M_myristic_acid	0	1	-0.55	1.12	-2.29	-1.34	-0.65	0.17	1.67	
M_N_acetylaspartic_acid	0	1	-0.07	0.99	-1.63	-0.76	0.03	0.54	1.46	
M_N_acetyl_D_galactosamine	0	1	-0.53	0.64	-1.46	-0.85	-0.53	-0.32	1.38	
M_N_acetylglycine	0	1	-0.21	0.86	-1.77	-0.95	0.04	0.36	1.03	
M_N_acetylmannosamine	0	1	-0.34	0.95	-3.75	-0.58	-0.17	0.29	0.77	
M_N_acetylornithine	0	1	-0.39	0.60	-1.97	-0.74	-0.28	0.00	0.86	
M_N_carbamoylaspartate	0	1	0.25	1.46	-1.76	-0.51	0.13	0.64	5.07	
M_n_epsilon_trimethyllysine	0	1	0.27	0.85	-1.02	-0.36	0.15	0.79	2.25	
M_nicotinamide	0	1	0.54	0.98	-0.96	-0.11	0.53	0.84	2.74	
M_nicotinic_acid	0	1	0.16	1.27	-0.68	-0.51	-0.12	0.03	5.00	
M_nonadecanoic_acid	0	1	-0.37	0.62	-1.73	-0.61	-0.38	-0.19	1.28	
M_norvaline	0	1	-0.17	0.66	-1.18	-0.61	-0.28	0.18	1.67	
M_O_acetylserine	0	1	-0.40	0.84	-1.75	-1.00	-0.44	-0.17	1.24	
M_octadecanol	0	1	0.47	0.75	-0.37	0.09	0.30	0.71	2.91	
M_octadecylglycerol	0	1	-0.37	0.92	-1.33	-1.07	-0.93	0.63	1.19	
M_oleamide	0	1	0.20	1.42	-2.92	-0.22	0.18	0.63	4.09	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_oleic_acid	0	1	-0.56	0.89	-2.38	-1.19	-0.23	-0.01	0.82	
M_O_phosphoserine	0	1	0.14	0.88	-2.10	-0.25	-0.03	0.85	1.71	
M_ornithine	0	1	0.42	0.82	-1.41	-0.15	0.21	1.16	1.48	
M_orotic_acid	0	1	-0.13	0.92	-2.18	-0.83	0.14	0.44	1.64	
M_oxalic_acid	0	1	-0.50	0.86	-2.63	-0.81	-0.26	0.07	0.55	
M_oxoproline	0	1	-0.45	0.96	-2.29	-0.92	-0.31	0.30	1.04	
M_palmitic_acid	0	1	-0.53	0.91	-2.01	-1.05	-0.53	0.07	1.75	
M_palmitoleic_acid	0	1	-0.54	1.11	-2.09	-1.45	-0.61	0.36	1.60	
M_pantothenic_acid	0	1	-0.11	0.78	-1.52	-0.44	-0.06	0.24	1.31	
M_parabanic_acid	0	1	-0.08	0.97	-2.15	-0.56	-0.22	0.48	2.00	
M_pentadecanoic_acid	0	1	-0.25	0.96	-2.50	-0.12	0.05	0.32	0.77	
M_pentitol	0	1	-0.19	0.88	-1.28	-0.82	-0.39	0.30	1.54	
M_phenaceturic_acid	0	1	-0.23	0.96	-1.78	-0.87	-0.29	0.29	1.72	
M_phenylacetic_acid	0	1	0.38	1.02	-2.02	-0.04	0.40	1.23	1.64	
M_phenylalanine	0	1	0.34	1.08	-1.34	-0.36	0.22	1.04	2.43	
M_phenylethylamine	0	1	-0.03	1.19	-1.90	-0.85	-0.17	0.56	2.45	
M_phosphate	0	1	-0.58	0.64	-1.73	-1.09	-0.60	-0.17	0.82	
M_phosphoenolpyruvate	0	1	-0.03	0.84	-1.43	-0.54	-0.29	0.63	1.84	
M_phosphoethanolamine	0	1	0.01	1.01	-2.32	-0.40	-0.02	0.64	2.10	
M_p_hydroxylphenyllactic_acid	0	1	-0.19	1.02	-1.58	-0.75	-0.50	-0.18	2.18	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_phytanic_acid	0	1	0.20	0.86	-1.32	-0.31	0.15	0.75	2.33	
M_pimelic_acid	0	1	0.18	1.32	-3.19	-0.33	0.31	0.86	3.16	
M_pipecolinic_acid	0	1	0.23	1.18	-1.40	-0.75	0.06	1.40	1.87	
M_piperidone	0	1	0.05	1.22	-1.89	-0.68	-0.21	0.30	3.79	
M_proline	0	1	0.16	1.01	-1.89	-0.41	0.44	0.91	1.31	
M_pseudo_uridine	0	1	-0.01	0.92	-1.42	-0.69	-0.06	0.66	2.11	
M_p_tolyl_glucuronide	0	1	-0.70	0.54	-1.51	-1.01	-0.74	-0.48	0.57	
M_putrescine	0	1	0.15	1.20	-1.01	-0.50	-0.14	0.32	4.54	
M_pyrophosphate	0	1	-0.09	1.12	-1.47	-0.67	-0.15	0.13	3.71	
M_pyrrole_2_carboxylic_acid	0	1	-0.03	0.82	-1.32	-0.49	-0.01	0.35	2.09	
M_pyruvic_acid	0	1	0.19	0.61	-1.37	-0.18	0.26	0.52	1.39	
M_quinolinic_acid	0	1	-0.09	1.18	-2.12	-1.06	-0.05	0.54	2.35	
M_ribitol	0	1	-0.50	0.74	-1.61	-0.98	-0.72	-0.24	0.73	
M_ribonic_acid	0	1	0.01	0.79	-1.70	-0.60	0.34	0.61	0.82	
M_ribose	0	1	-0.46	1.09	-2.68	-1.04	-0.58	0.20	2.12	
M_ribose_5_phosphate	0	1	-0.31	0.87	-1.65	-1.05	-0.23	0.24	1.70	
M_ribulose_5_phosphate	0	1	0.03	0.94	-1.60	-0.81	0.24	0.71	1.72	
M_saccharic_acid	0	1	-0.03	1.05	-1.92	-0.48	0.15	0.66	1.54	
M_sarcosine	0	1	0.08	0.65	-1.55	-0.21	0.33	0.46	0.88	
M_serine	0	1	0.65	0.95	-0.55	-0.02	0.53	1.06	2.68	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_serotonin	0	1	-0.44	0.96	-1.54	-1.10	-0.72	-0.28	2.12	
M_shikimic_acid	0	1	0.32	1.32	-0.92	-0.45	-0.12	0.46	4.85	
M_sophorose	0	1	-0.25	0.74	-2.25	-0.66	-0.19	0.18	1.15	
M_sorbitol	0	1	-0.38	0.91	-2.86	-0.65	-0.14	0.14	0.91	
M_spermidine	0	1	-0.10	0.96	-1.94	-0.70	0.07	0.74	1.30	
M_squalene	0	1	0.02	1.13	-2.26	-0.71	0.33	0.57	1.80	
M_stearic_acid	0	1	-0.46	0.75	-1.81	-0.91	-0.37	-0.16	1.34	
M_succinate_semialdehyde	0	1	-0.23	0.83	-1.76	-0.69	-0.19	0.43	0.86	
M_succinic_acid	0	1	-0.19	0.87	-2.08	-0.86	-0.28	0.23	1.60	
M_sucrose	0	1	0.02	1.29	-2.98	-0.49	-0.16	0.41	2.48	
M_tagatose	0	1	0.12	0.91	-1.56	-0.44	0.32	0.76	1.61	
M_taurine	0	1	-0.03	0.72	-1.93	-0.35	-0.02	0.39	1.52	
M_threitol	0	1	0.25	0.99	-1.78	-0.28	0.16	1.05	2.02	
M_threonic_acid	0	1	0.05	0.97	-1.64	-0.59	-0.12	0.74	1.77	
M_threonine	0	1	0.60	0.73	-1.12	0.28	0.58	1.14	1.54	
M_thymidine	0	1	-0.10	0.77	-1.71	-0.61	-0.08	0.32	1.36	
M_thymine	0	1	0.02	0.95	-2.14	-0.62	0.18	0.73	1.32	
M_tocopherol_alpha_	0	1	0.00	0.91	-1.62	-0.58	-0.17	0.46	1.92	
M_trans_4_hydroxyproline	0	1	-0.29	0.94	-1.97	-0.85	-0.46	0.23	1.49	
M_trehalose	0	1	0.08	1.10	-1.66	-0.78	0.23	0.59	2.39	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
M_tryptophan	0	1	0.60	0.71	-1.23	0.23	0.59	1.18	1.56	
M_tryptophol	0	1	-0.66	1.04	-2.76	-1.26	-0.50	0.16	1.43	
M_tyrosine	0	1	0.14	0.89	-2.10	-0.53	0.29	0.80	1.26	
M_tyrosol	0	1	-0.29	0.63	-1.49	-0.70	-0.10	0.10	0.59	
M_UDP_GlcNAc	0	1	-0.03	0.77	-1.47	-0.52	-0.15	0.45	1.41	
M_undecanoic_acid	0	1	-0.01	1.11	-3.04	-0.01	0.34	0.68	0.94	
M_uracil	0	1	-0.01	1.03	-1.66	-0.71	-0.12	0.51	2.47	
M_urea	0	1	0.09	0.60	-1.04	-0.50	0.25	0.59	0.91	
M_uric_acid	0	1	-0.19	0.91	-1.65	-0.90	-0.25	0.25	1.52	
M_uridine	0	1	-0.13	0.93	-2.43	-0.63	-0.22	0.74	1.18	
M_urocanic_acid	0	1	0.12	1.06	-1.29	-0.67	-0.03	0.45	3.00	
M_valine	0	1	0.46	1.20	-1.38	-0.56	0.24	1.52	2.38	
M_xanthine	0	1	-0.24	0.69	-1.56	-0.59	-0.24	0.08	1.31	
M_xanthosine	0	1	-0.10	1.11	-2.47	-0.72	-0.26	0.67	1.79	
M_xylitol	0	1	-0.66	0.83	-2.58	-1.12	-0.65	-0.25	0.72	
M_xylose	0	1	-0.27	1.03	-2.42	-0.65	-0.45	0.43	1.28	
M_xylulose	0	1	-0.16	0.83	-2.26	-0.61	-0.12	0.36	1.15	

Before proceeding with the analysis, I dummy-coded the data for model fitting down the line.
- Diagnosis: 0 is Con and 1 is Met
- Parity: 0 is primiparous and 1 is multiparous

```r
Coded_calv <- Calving_scaled |>
  mutate(Diagnosis = as.numeric(Diagnosis)-1,
         Parity = as.numeric(Parity)-1)

# Print
Coded_calv[1:5,c(-1,-3)]
```

```
  Diagnosis Parity M_1_5_anhydroglucitol M_1_hexadecanol M_1_monoolein
1         0      0           -3.36996821      -0.2604090     0.8543952
2         0      0            0.28319146      -0.3431324    -1.0537918
3         0      0            0.05602253       0.9722826     0.8262146
4         0      0           -0.37227785      -0.7012112     0.3767564
5         0      0           -0.85407070       0.5521593    -1.5184625
  M_1_monopalmitin M_1_monostearin M_2_5_dihydroxypyrazine
1       -0.1798523      -0.9163973              -0.3257091
2        0.2800500       0.6479029              -0.6388007
3       -0.5960485      -0.1234495               0.3094925
4        1.1074626       0.3640961              -0.5288944
5        2.3054187      -1.0810379              -1.1347031
  M_2_6_diaminopimelic_acid M_2_8_dihydroxyquinoline M_2_aminobutyric_acid
1                -0.6197293                0.7313519            0.36654132
2                 1.1663619                0.4912174           -0.19420297
3                 0.2054108                0.5145700           -0.06174665
4                 1.8851100               -0.4492438            0.02089201
5                 0.4487685               -0.1746165            0.38978882
  M_2__deoxyguanosine M_2_deoxypentitol M_2_deoxytetronic_acid
1         -0.05180223         1.2768798              0.7180045
2         -0.40843151        -1.1123530             -1.2125496
3         -0.51590294         1.1107534             -0.6658086
4         -0.18148218         0.7853768              1.1257666
5         -0.85167378         0.2996328              0.3562294
  M_2_hydroxy_2_methylbutanoic_acid M_2_hydroxybutanoic_acid
1                         1.2564919               -1.4927491
2                         0.2759607               -0.8526483
3                         0.2355602               -0.3909984
4                         0.9815808               -1.0401694
5                        -1.5616382               -1.1006202
  M_2_hydroxyglutaric_acid M_2_hydroxyhexanoic_acid M_2_hydroxyvaleric_acid
1              -1.35777761                0.1701490              0.24145572
2              -0.40257175                0.6656920              0.17626116
3               0.07191239                0.3637872              0.08946548
4              -0.51468805               -0.4893106              0.20343403
```

```
5              -1.33387598                0.2124105              1.05423339
  M_2_ketoadipic_acid M_2_ketobutyric_acid M_2_ketoglucose_dimethylacetal
1          -1.0707097            0.1134815                      0.5014978
2           1.0869510           -0.1622772                      0.8490873
3           0.2513707            0.1634803                      0.3335519
4           1.3940858            0.2043874                      1.0819206
5           0.7619260            0.5802299                      1.1085935
  M_2_ketoisocaproic_acid M_2_ketoisovaleric_acid M_2_methylglyceric_acid
1             -0.26952225              -0.7498303              -0.2725749
2              1.41225444              -1.8312738              -0.7695737
3             -0.59958167              -0.6799025               0.8236961
4              0.07590225               0.8485043               0.7652616
5              1.56353489               0.7592835              -0.5350248
  M_2_monoolein M_2_monopalmitin M_2_picolinic_acid
1    0.15591692       -0.9109606         -0.3182158
2   -0.23204881        0.2592284          0.8976684
3    0.46721071        0.6889682          0.3905385
4   -1.06714652        1.0815005          1.4724712
5   -0.05702647        0.2845689          0.4860574
  M_3__4_hydroxyphenyl_propionic_acid M_3_aminoisobutyric_acid
1                           0.1672383               -0.1805865
2                           1.0666836                2.2658904
3                          -1.1622670               -0.6965806
4                           0.5325964               -0.4563774
5                           1.2640113               -0.4226035
  M_3_hydroxy_3_methylglutaric_acid M_3_hydroxybutyric_acid
1                         0.2054248              -1.2969974
2                         0.9857045              -0.2199784
3                         0.3031970               0.9824504
4                        -0.3719216              -0.4709746
5                         0.8982154              -0.5267962
  M_3_hydroxypalmitic_acid M_3_hydroxypropionic_acid M_3_phenyllactic_acid
1                0.1544318                 0.4509666             0.4493954
2               -0.6260963                 0.6704625             0.4837419
3               -0.5932068                -0.1397296             0.9318537
4                0.5924948                -0.3636902            -0.3910849
5               -0.2615333                -0.3786386            -2.0511950
  M_3_phosphoglycerate M_3_ureidopropionate M_4_aminobutyric_acid
1            1.0814070           -1.0084651           -0.09068190
2           -0.3439260           -0.6188089            0.57333495
3            2.3882809            0.9942514           -0.07781160
4           -0.8785183            1.4280953            0.89179834
5            0.9400435            0.1121541            0.02080549
```

```
  M_4_hydroxybutyric_acid M_4_hydroxycinnamic_acid M_4_hydroxyphenylacetic_acid
1              0.79199701                0.5912913                    -0.3737394
2              0.98122478               -1.0080083                     0.9337719
3             -0.46193767                0.7177165                    -0.3538297
4              0.00425402               -0.4737693                     0.5003788
5              0.93239101                0.2240023                     0.4537722
  M_5_6_dihydrouracil M_5_aminovaleric_acid M_5_hydroxynorvaline
1          -1.7541414             1.1086056            1.0209022
2           0.4332185            -1.3154472            0.6768659
3           0.7360111            -0.8515566           -0.1405978
4          -1.5435980             0.2945071           -4.0551041
5          -0.3801140            -0.1865576            0.1543300
  M_5_methoxytryptamine M_9_myristoleate M_aconitic_acid  M_adenine M_adenosine
1           -0.02103677      -1.82022133      -0.6500835  3.2427408  -0.7454412
2            1.38369603      -0.65357326      -0.9998539 -1.4120879   0.9470695
3           -2.41247788       0.45901385       1.0908796  0.3937163  -0.4388916
4            0.43894645       0.07051659       1.4072864  0.9850277   4.9961626
5            0.52230938      -0.84569771      -0.2848177  0.7086073   0.3082840
  M_adenosine_5_monophosphate M_adipic_acid  M_alanine M_alanine_alanine
1                   0.3974557    -1.1453313  1.8954282         1.1081926
2                   0.5898797     1.1062601 -0.7983402        -0.3979049
3                   1.6555744    -0.2272378 -0.3028871         0.8451089
4                  -0.7044170    -0.1551111  0.8503349         1.7983308
5                   0.5231037     0.4712706  1.5001446         0.5160193
  M_allantoic_acid M_alpha_aminoadipic_acid M_alpha_ketoglutarate
1        1.1404161                0.5059869            -5.1145132
2        0.5541899                0.5132127            -0.5297134
3        0.3710238               -0.8763180             0.5418376
4        1.7415050               -0.6676090             1.7851684
5       -0.9109578                1.3490562            -0.2707575
  M_aminomalonate M_arachidic_acid M_arachidonic_acid M_asparagine
1       0.3002641       -0.4141836       -0.650071304   1.20176170
2      -0.5791596       -0.8610091        0.002029076  -0.08414661
3       2.1258041       -0.9979701       -1.480912997   0.15702441
4       1.5575014        4.1205410        0.245263568   2.20693799
5       1.1362343        0.2536934        1.671949312   0.42081547
  M_aspartic_acid M_behenic_acid M_benzoic_acid M_beta_alanine
1      0.60992862     -1.0743847      1.0363532     0.06154180
2      0.01678389      0.4864695      1.5044117    -0.60707992
3     -0.38241465     -0.1206716      0.3226487    -0.02101342
4      2.22189506      0.5200353     -0.7821511    -0.45611602
5     -0.03188607      0.4703973      1.1913706    -0.32258688
  M_beta_gentiobiose M_beta_glutamic_acid M_beta_glycerolphosphate
```

```
1          0.9526749        -0.0004767582               -0.6088215
2         -1.2717380        -0.1056427600                0.6428373
3          0.5264204        -0.6824925150               -0.7196193
4          0.6703720        -0.1171421724                0.3530737
5          0.5221647         0.7640245210                0.1591149
  M_beta_sitosterol M_butyrolactam M_catechol M_cellobiose M_cerotinic_acid
1         0.8514307     -0.7992116 -0.1208157   -1.7285353        1.0870011
2         0.4576209      0.1246011  0.7365085   -0.8204150        1.8400398
3         0.5179451      0.4717036 -0.5719398   -0.3347631       -0.5036384
4        -1.9250124      0.2780190  1.1743543    0.6102633        1.7592515
5         0.4075564     -0.5733175 -1.1546647   -4.7294432       -0.2035032
  M_cholesterol M_cholesterone M_cholic_acid M_ciliatine M_cis_gondoic_acid
1   -0.43593393     -0.4538582     0.1583387  -0.9609760         -0.5313761
2    0.88836222      0.8853031    -0.3722532   0.9160971         -0.7539757
3   -0.31567151      0.2554295    -1.1637014  -0.6183954         -0.8542368
4   -0.02458731     -0.2628545     0.5871882   0.1360510         -0.1271844
5    1.15878631     -0.2895213    -1.5757356  -0.6399421         -0.0203477
  M_citramalic_acid M_citric_acid M_citrulline M_conduritol_beta_epoxide
1         0.2895285     0.7785461   1.18392356               -0.08834906
2         0.1697262     0.6390191   0.97204000                0.47516825
3        -0.9729860     0.8245930   0.07433367               -0.32707464
4        -0.5950480     1.8758848   1.58424490               -0.30463443
5         0.3294442     0.1562405   0.94820022               -0.34513882
  M_creatinine M_cysteine M_cysteine_glycine  M_cystine   M_cytidine  M_cytosin
1  -0.11835902  0.2895105         -0.9076828 -0.2315157 -1.390915070  0.4617010
2   0.36580463  0.8982146          0.7054910  1.0118858 -1.180688530 -0.2786562
3  -0.02979125  0.7485915          0.3399636  0.4435646 -0.213745550 -0.3884363
4   0.12161956 -1.2377559         -1.5875246 -0.8467246  0.005293929  1.3761372
5  -0.54498676  0.7873195         -0.6074260  0.2473530 -3.759069830 -0.4553746
  M_dehydroabietic_acid M_deoxycholic_acid M_D_erythro_sphingosine
1             1.7095253         -0.8437559              -1.0208735
2             1.9230398         -0.6200486              -2.0555273
3            -0.5404390         -1.3022971              -0.9472485
4            -0.6542019          0.4027776               0.9644473
5             0.1177422         -1.8958799              -0.9594001
  M_dihydroxyacetone M_docosahexaenoic_acid M_elaidic_acid
1        -0.21050615             -0.1113681      1.7076594
2         0.85286528             -0.2982946     -0.6131717
3        -0.80161164             -1.2948073     -0.7058017
4         0.22285782             -0.9062901     -0.6275519
5        -0.03842292              0.6609089     -1.1726178
  M_epsilon_caprolactam M_erythritol M_erythrose M_ethanolamine
1            2.00560845   -1.5599991 -0.52215178      0.7704233
```

```
2            0.01489439    0.0172441 -1.77222762      0.4401330
3           -1.46306183   -1.1303100  1.62410805     -0.7940903
4           -0.44599136    0.1774013  0.08410434     -0.4546881
5           -0.08013799    0.9193162  1.25938483     -0.6393923
  M_ethanol_phosphate M_fructose M_fructose_1_phosphate M_fructose_6_phosphate
1           0.1562347  0.1372635             1.17690673              1.1730598
2          -0.5075428  0.6146071            -0.87679174              1.3262604
3          -2.2445659 -0.6141080            -0.01559042              0.9992412
4          -0.9694290 -2.1360861             0.01890883              1.4969887
5           0.2327208  0.6626470            -0.50792941             -0.1767886
    M_fucose M_fumaric_acid M_galactinol M_galactose_6_phosphate
1 -1.1864747     -0.6534094  -0.98598849             -1.19082180
2 -1.5769712     -0.4178121  -0.65382942              1.61988528
3 -1.2843159      0.1968002  -0.93377508              0.74161902
4 -0.5940578      0.5780056  -0.28984506              0.11685456
5 -1.3225997     -0.4178121  -0.09754614              0.09555443
  M_gluconic_acid   M_glucose M_glucose_1_phosphate M_glucose_6_phosphate
1      -0.5974510 -0.85051839            -0.3379429            -0.8191661
2      -0.1483946 -0.04334105            -1.4691636             0.4944076
3      -0.2870172 -0.69774401            -0.7599575             0.1402794
4      -0.6866786 -0.24726833            -0.1981111             5.0955394
5      -0.4893629  0.36577704            -1.5093884             0.6926741
  M_glutamic_acid M_glutamine M_glutaric_acid M_glyceric_acid  M_glycerol
1      -0.3299023   0.9369245      -0.6940776     -0.19984628 -1.34630748
2       0.3865654  -0.2976545      -1.0251042     -0.64101103 -1.98269019
3      -0.7376197   0.5924404      -0.8262591      0.43862557 -0.03352473
4       0.5833614   1.4657600       0.4047708     -1.36457482 -1.32009997
5       1.1220966   0.7425941       0.9977422     -0.07360734 -0.83150403
  M_glycerol_3_galactoside M_glycerol_alpha_phosphate  M_glycine M_glycocyamine
1               -0.3653372                  1.1075037  0.3940840      0.3488093
2                0.6025126                  0.5912371 -1.1501530      0.6538324
3                3.9960176                 -0.2954653  2.1301582     -1.6282369
4                0.4933921                 -0.3760382  1.3707600      0.6819190
5                0.3363177                  2.0425741 -0.2108274     -0.4936491
  M_glycolic_acid M_glycyl_proline M_glycyl_tyrosine  M_guanine M_guanosine
1      -1.1588057       -0.3379071        -0.8726101  2.3916555   0.4883787
2       0.4721098        0.5593783        -0.9621564 -1.5236068  -0.4573282
3      -0.4259126       -0.7267728         0.1439978  0.9795977   0.2177898
4       1.2404402       -1.9936995         0.5281542 -0.2716535   1.0291737
5       0.3912507       -1.0293420         0.2940723  1.2507539   0.1030904
  M_heptadecanoic_acid M_hippuric_acid M_histidine M_homoserine
1           -2.1448887       0.8360159  -0.6651060  -0.07402840
2           -0.8468546       1.3179491  -0.2402193  -0.08920794
```

```
3            0.1944308      -0.4326396   1.8010905  -0.42457905
4           -0.6227645       0.3534554   1.4653186   0.96428136
5           -1.2006756      -0.1560235  -0.1156154   0.21681880
  M_hydrocinnamic_acid M_hydroxycarbamate M_hypoxanthine  M_ile_ile
1           1.04926419         -0.2838692      0.3823489  0.3595864
2           0.75639606          0.5330951      0.4909267 -1.3855880
3           0.06110016          0.2818101      1.1311201  0.8525904
4           1.13491436          0.5599487      1.9799641  0.5803869
5           0.53187453          0.3675082     -0.4426264 -1.8746904
  M_indole_3_acetate M_indole_3_lactate M_indole_3_propionic_acid
1        -0.03797132        -0.38341995                -0.1470767
2         0.64535322        -0.05518962                 0.4067831
3        -0.13585010         0.19707415                 0.3223401
4        -0.10275692        -2.41390440                 1.3903286
5         0.28690691        -1.66555015                 1.7007365
  M_indoxyl_sulfate  M_inosine M_inositol_4_monophosphate M_isocitric_acid
1       -1.05470692  0.3906329                 0.09522915       0.37035648
2        0.40609547 -0.9901112                 0.45794618       0.04864206
3       -1.79085946 -0.8477539                 1.04874458       1.14661161
4        0.02603255  0.3664824                -0.36159891       2.12268072
5       -1.39966751 -0.4666302                -1.01691071      -0.59525256
  M_isoheptadecanoic_acid M_isoleucine M_isolinoleic_acid
1              -2.0703208    1.1917318        -1.63129378
2              -0.6947915    0.2524083         0.06589207
3              -0.1428202    0.3692833         0.91361109
4              -0.9544328    0.6490713         0.16434322
5              -1.5248849    1.3588254        -0.37315446
  M_isopentadecanoic_acid M_isoribose M_isothreonic_acid M_itaconic_acid
1             -0.22752391  0.04892259         -0.9348818       0.4242789
2             -1.34929526  0.34945972         -0.5831956      -1.3313917
3             -0.12854188 -1.78413777         -1.2281060      -0.9477272
4             -0.07014310  1.63652412          0.3557530       1.0886947
5             -0.09734696  0.52599853         -0.6773310       0.3128195
  M_kynurenine M_lactamide M_lactic_acid M_lactitol M_lactobionic_acid
1   1.20050231   1.0682102     0.9175287 -1.5782419         0.03726024
2   0.07474276  -0.6891729    -0.3518117 -0.9051810        -0.74951456
3   0.34581041  -0.2403804    -0.3356656 -0.0614942        -0.87579969
4   2.23933719  -0.2958888    -0.9177608 -0.1733976        -1.29711353
5   1.19194468  -1.4973748    -1.2249143 -2.5344901         0.19817914
    M_lactose M_lanosterol M_lauric_acid M_leucine M_lignoceric_acid
1 -1.47305701   0.14594987    -1.0276551 0.8584178        -0.6672178
2 -0.66923250  -0.08337714     0.1498045 1.3589605         0.3590631
3 -0.20443876  -0.42308087    -0.6428169 0.3389721         0.8732268
```

```
4 -0.07752533   0.02276765     0.9046800 0.8303364         0.1007156
5 -5.05353031   1.01422212    -0.9504054 0.7630129        -0.3000501
  M_linoleic_acid M_lithocholic_acid  M_lysine  M_lyxitol   M_lyxose
1     -1.18101175          0.5133756 1.4658435 -1.0371747 -0.5802549
2     -0.01044262          0.3691992 1.2463851 -0.1283179 -0.1388368
3      0.11763327          1.0295823 0.5841591 -1.2652567  0.1999029
4     -0.51129216         -0.4627987 1.2582659  0.3819772 -1.1090864
5     -0.50890244         -0.3199453 1.3320920 -0.5860474 -0.9147926
  M_maleimide M_malic_acid M_malonic_acid   M_maltose M_maltotriose  M_mannose
1  -0.3887851   -1.0378012      0.6422811  0.08340288    0.54764701 -0.1268470
2  -0.2184829   -1.0751571      0.2549856 -1.05532361    0.38675092 -0.1672335
3   0.7967662    0.4282354     -3.2214453  2.51098238    1.56622397 -1.4483143
4  -0.8220037    0.4699345      0.2490933 -0.04715340    0.02226991 -0.8641386
5   0.8035191   -0.3164405     -0.5630673 -1.75537461    0.38675092  1.3015179
  M_mannose_6_phosphate M_melibiose M_methanolphosphate M_methionine
1             1.4840033  -0.5278468         -0.10210825   0.25859344
2             0.3853900  -1.2323037          0.05094833  -0.28487908
3            -0.2242014   0.3069307          0.23977896  -0.67551794
4             1.2167700  -0.4951723         -2.03873426   2.18222973
5            -0.8080228  -0.9331818          1.04860904  -0.06204024
  M_methionine_sulfoxide M_methylmaleic_acid M_Mevalonic_acid M_myo_inositol
1              0.6701615           1.6872099        0.2303306     -1.3214911
2              0.5867309           0.6984357       -0.9673796     -0.6210369
3             -0.5941180           0.3290063       -0.8160536     -1.2662932
4             -0.2376473           1.0504311        0.9602767      0.2723384
5              0.7923282          -0.2772885       -0.9132918     -1.3100045
  M_myristic_acid M_N_acetylaspartic_acid M_N_acetyl_D_galactosamine
1     -2.07182801               0.4965999                 -1.1210518
2     -1.53223877              -1.6260211                 -0.8342698
3     -0.05556912              -1.4873041                 -0.3775938
4     -0.55014985               1.4317183                 -0.7042319
5     -1.90354156              -0.8537698                 -0.7042319
  M_N_acetylglycine M_N_acetylmannosamine M_N_acetylornithine
1         0.9026367           -3.75024793          0.03068887
2        -0.5385030            0.06729968          0.15195271
3         0.3647937           -1.41808993         -0.13796361
4         0.3537938           -0.19478156         -0.64377138
5        -1.3098503           -0.59474856          0.10724479
  M_N_carbamoylaspartate M_n_epsilon_trimethyllysine M_nicotinamide
1              0.2989394                 -0.25457401      0.4551335
2             -0.5099638                 -0.67640693      0.6572972
3              5.0684712                 -0.01817377      0.4402860
4              0.7636956                  0.56252401      2.4836293
```

```
5              0.1290691                  1.05474114     -0.5075920
  M_nicotinic_acid M_nonadecanoic_acid M_norvaline M_O_acetylserine
1       0.10169175         -1.73054768  -0.5699841        1.0859079
2      -0.05501242         -0.39698160   0.9751778       -1.2644626
3      -0.11557955         -0.29933040   0.1708742       -0.3808273
4       4.99994090          0.06231628  -0.6740895       -0.5280629
5      -0.67976917         -0.35225884  -0.6904624        1.2408128
  M_octadecanol M_octadecylglycerol M_oleamide M_oleic_acid M_O_phosphoserine
1     0.5581060          -1.0486224  1.2655994 -2.375991441        -0.2216664
2     2.9050732          -0.9646616  1.5579205 -0.012072269        -0.1251122
3     0.4232771          -1.1582145  1.2915899  0.001431693         0.6656544
4     1.3880257           0.8672559 -2.3650770  0.040470882         1.7088412
5     0.3191904          -1.1297623 -0.6088598 -0.147767595         0.2560104
  M_ornithine M_orotic_acid M_oxalic_acid M_oxoproline M_palmitic_acid
1   1.4684654    -0.5655775   -0.61487050   -0.7345724      -1.8780940
2   1.3902184     0.6641217   -0.89715208   -0.5201574      -1.5668916
3   0.2121235    -0.2489794   -2.63435673   -1.7299839      -0.1112268
4   0.2008333    -1.1856139    0.05259026    0.8721857      -0.2987702
5   1.4846658    -1.0892009   -0.40275834    0.2926859      -1.1178133
  M_palmitoleic_acid M_pantothenic_acid M_parabanic_acid M_pentadecanoic_acid
1         -2.0898009         -0.3045885       -1.0762130         -1.789736729
2         -1.4623886         -0.3663989        0.5388089          0.334751003
3          0.1754039         -1.3648879       -0.5098530         -2.497158657
4         -0.6116802          0.1244314       -0.3112115          0.002931308
5         -1.9225581         -0.4059084       -0.2590210          0.259780206
  M_pentitol M_phenaceturic_acid M_phenylacetic_acid M_phenylalanine
1 -1.0755518          -1.1589781          0.40724851      0.48799269
2 -0.7860025           0.5518900          0.99781292      0.03928693
3 -0.2824824          -0.8516625         -0.31399659     -0.30763139
4  0.5081371           0.1995264          1.56990674      1.80001642
5  0.4201424           1.7210329          0.05484833      1.99587255
  M_phenylethylamine M_phosphate M_phosphoenolpyruvate M_phosphoethanolamine
1        -0.81817933  -0.6890315            -0.7373358           -0.03217397
2        -0.78006733   0.1749907             1.1310843            0.99191798
3        -0.03493265  -0.3511928            -0.5343424           -1.04184855
4         2.19925791  -1.4903936             1.8416938            0.35144673
5         0.41506750   0.8213543            -1.1757197            0.07368777
  M_p_hydroxylphenyllactic_acid M_phytanic_acid M_pimelic_acid
1                    -0.4310192     -0.07589535     -0.5026094
2                    -0.8964299      0.91340271     -3.1858505
3                    -1.5770251     -0.12540972      0.6186106
4                     1.3992560      0.59868747      1.2355156
5                     1.3588575     -0.50030359      0.2941338
```

```
  M_pipecolinic_acid M_piperidone  M_proline M_pseudo_uridine
1          1.3840536  -0.46552070  0.8081101       -0.8514650
2          1.3486908   0.08873153  0.8489594       -0.3742034
3         -0.8749711   1.88449725 -0.4299166       -0.8815604
4          0.2303140  -0.04604914  1.0429068        2.1077440
5          1.8714418  -1.88795159  1.3063993       -1.4176598
  M_p_tolyl_glucuronide M_putrescine M_pyrophosphate
1           -0.68529896   -0.1436487     -0.77992934
2           -0.08794388   -0.1356321     -0.63341562
3           -0.93077624   -0.3239625     -0.30968599
4           -0.79099284    4.5432719     -0.04788296
5            0.57145828   -0.5599552      0.25225892
  M_pyrrole_2_carboxylic_acid M_pyruvic_acid M_quinolinic_acid  M_ribitol
1                 -0.07015528      0.8134598         1.2045790 -0.7327927
2                  0.51141378      0.2273461         0.2086089  0.7326198
3                 -0.96698027      0.5046898        -0.2802000 -1.2873026
4                  2.08959928      0.7628273         1.3778291 -0.4936434
5                  0.23475219     -0.5049130         0.4902276 -0.6482959
  M_ribonic_acid   M_ribose M_ribose_5_phosphate M_ribulose_5_phosphate
1     0.81851010 -2.6773292           0.3336790             0.6631049
2     0.65794556 -0.9221945          -0.1222683            -0.1374284
3    -0.02523349 -0.5894622          -0.6325945            -0.3078468
4    -0.88056353 -1.1577747           1.7021606             0.7064062
5     0.61685267 -1.8125370          -1.4734270             0.6485455
  M_saccharic_acid M_sarcosine   M_serine M_serotonin M_shikimic_acid
1       -1.2650924   0.8805487 1.43630361  -1.1846207      -0.6642760
2       -1.9231935   0.7583840 0.09052448  -1.1595978       0.1846527
3        0.2014774   0.7776270 0.64140722   0.1641035       2.4131582
4       -0.1180919  -0.1350253 2.67961392   2.1161673       4.8514473
5        0.3612234   0.3149014 0.41807884  -1.1375620      -0.5510616
  M_sophorose M_sorbitol M_spermidine  M_squalene M_stearic_acid
1  -0.4292183 -0.3422135  -1.94134031  0.28230677    -1.80903984
2  -0.1577785  0.1376174  -1.21395572  0.18160196    -0.92213444
3   0.5425687 -0.8912497   0.74928029 -0.75269814    -0.26437464
4   0.6935034  0.5959543   0.74171006 -1.81829409    -0.07131519
5  -0.2315865 -0.5539072   0.05938843  0.07123866    -0.24133386
  M_succinate_semialdehyde M_succinic_acid M_sucrose M_tagatose   M_taurine
1               -0.6669824      -0.8624062 1.1239347  0.5149240  0.01230179
2               -0.4364999      -0.3417288 0.0914325  0.9061461 -0.30950399
3               -0.7593790      -0.5111762 2.3566887 -0.3115016 -0.16215856
4               -1.1926192       1.5983025 2.0552317 -1.1919628  1.51529548
5                0.6009728       0.1357547 0.9410475  0.9468004  1.01015704
   M_threitol M_threonic_acid M_threonine M_thymidine  M_thymine
```

```
1 -1.51275276      -0.1245121   1.3907714 -1.05543381 -0.5726816
2 -0.80153008       0.7359618   0.4441615 -1.71039181 -2.1361383
3  1.37681406      -0.8034797   1.0826496  0.03883692  0.3005806
4  1.66143937      -0.1120287   1.0676817  0.72562564  0.3404388
5  0.03602558      -0.5083999   0.3009971 -0.18497894 -0.7441436
  M_tocopherol_alpha_ M_trans_4_hydroxyproline M_trehalose M_tryptophan
1          0.31937701               -0.4343014   0.3266386    1.2103016
2          0.80289401               -0.6876805  -0.2329084    1.4244079
3         -0.03226271                0.3239268   2.1603470    0.3117203
4          1.10427243                1.3274347   0.3819941    0.7873166
5          1.91804822                1.3206414  -0.9222008    0.2019342
  M_tryptophol  M_tyrosine   M_tyrosol M_UDP_GlcNAc M_undecanoic_acid
1   -1.1508603  1.18057878  0.20567150   -0.4320725        -3.0366801
2    0.2181713  0.09906358 -0.12504608    1.1158951         0.7379430
3   -0.9526842 -0.53350022 -1.12587908   -0.6802750         0.3562312
4   -0.3043985  1.19598870  0.30787127   -1.4738348         0.9436968
5    0.1853477  0.98339419 -0.03463827   -0.5635774         0.1534784
    M_uracil     M_urea M_uric_acid  M_uridine M_urocanic_acid  M_valine
1  1.2086782 -0.4723718   0.1204685  0.2678856      3.00019869 1.9323010
2 -1.0951222  0.5922045  -1.0187700 -0.3730790      0.27558450 2.3813611
3  0.1233899  0.1880471  -1.0339795 -0.5067674     -0.21084637 0.2699949
4  0.1411073  0.5875685  -0.1996508  0.1526251     -0.87044082 0.2132927
5 -1.6562605  0.6709902  -1.6469204 -2.4284593     -0.09737279 0.9640943
   M_xanthine M_xanthosine  M_xylitol   M_xylose   M_xylulose
1 -0.23953932    1.3377155 -1.8660465 -1.9096656  0.001439424
2  1.31080641    0.2236528 -0.9545638 -1.7306073 -2.256587116
3 -0.94123044   -0.9120670  0.5859726 -0.7667792 -0.068809348
4  0.08236261   -0.6412629 -0.9809553  1.2823196 -0.443822951
5 -0.54419986   -0.6088061 -0.3494738 -2.4165022 -1.235137638
```

Data Exploration

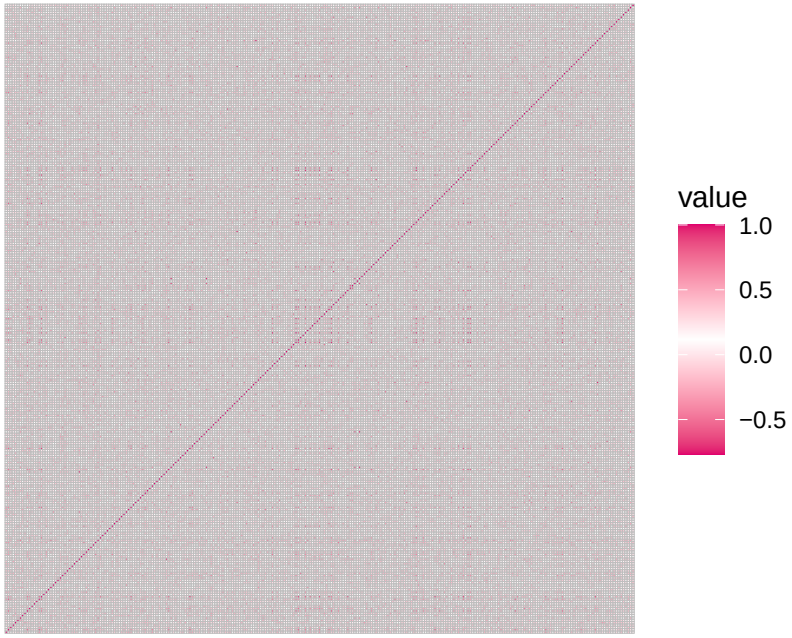
2d plots where two or three variables are being visualized can be informative and straightforward ways to visualize and explore data, unfortunately, visualizing relationships within datasets becomes increasingly challenging with a rising number of variables (and thus dimensions), which is the case for this dataset. For data exploration, I will be using multi-dimensional scaling and other methods.

Correlation

Given that the data I am working with is metabolites, and that metabolites are interconnected, obtaining some degree of correlation between them during metabolomic analysis highly likely, if

not unavoidable. To investigate the degree of correlation I (and by "I", I mean R) calculated a correlation matrix for all metabolites using Spearman correlation. I opted for Spearman, rather than the default choice, Pearson, because Spearman is able to capture monotonic relationship between variables, meaning it can detect whether two features increase or decrease together, regardless of the exact shape of their relationship while Pearson assumes linearity[3]. Finally, I produced a heat map to visualize the correlation matrix.

```
# Compute the correlation matrix
corr_matrix <- cor(Coded_calv[,c(-1,-3)], method = "spearman")

# Plot without names
ggcorrplot(corr_matrix, tl.cex = 0) +
  scale_fill_gradientn(values=c(1, .5, 0),
                       colours = c("#E0006D", "white", "#E0006D"))
```

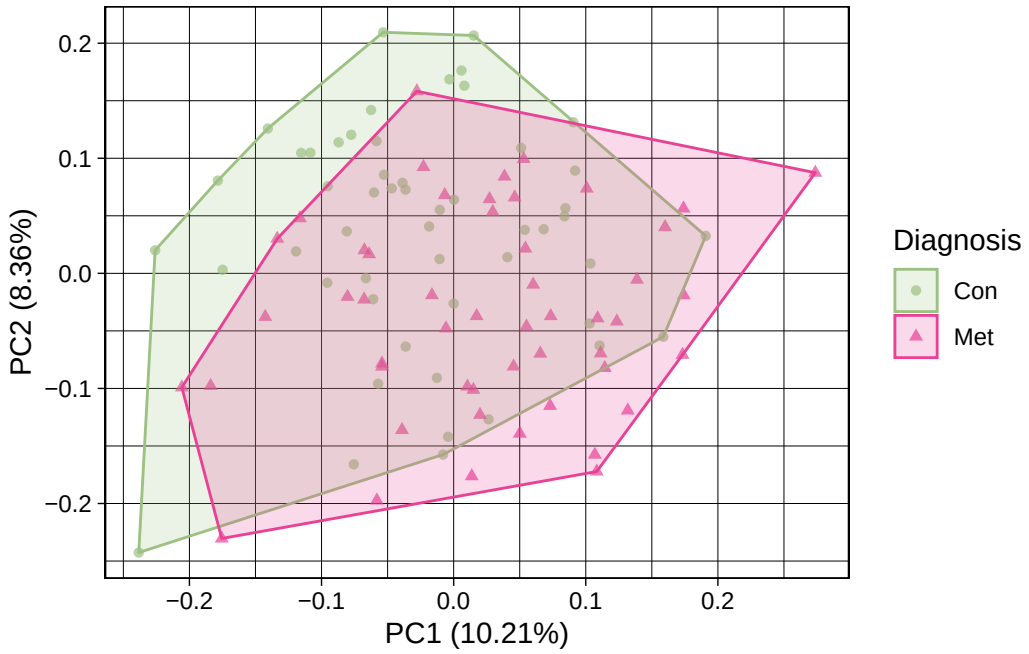


We can see above that there is correlation throughout the data, and some metabolites show strong correlation.

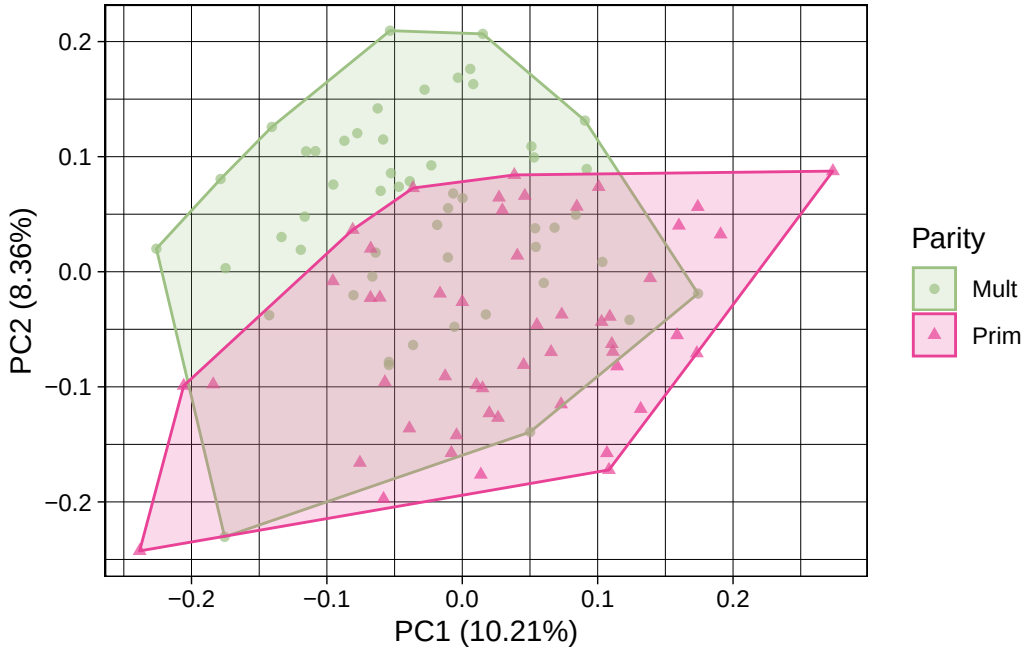
PCA

Now, let's plot this data using PCA. For the first plot, we'll see if cows seem to be grouped by "Diagnosis" or "Parity"

```r
# PCA
pca_out <- prcomp(Coded_calv[,c(-1,-3)])

pca_Diagnosis <- autoplot(pca_out,
         data = Calving_scaled,
         colour = 'Diagnosis',
         shape = 'Diagnosis',
         frame.colour = 'Diagnosis',
         alpha = .7,
         frame = T) +
  scale_color_manual(values = c(fav_colors[9], fav_colors[3])) +
  scale_fill_manual(values = c(fav_colors[9], fav_colors[3]))


pca_parity <- autoplot(pca_out,
         data = Calving_scaled,
         colour = 'Parity',
         shape = 'Parity',
         frame.colour = 'Parity',
         alpha = .7,
         frame = T) +
  scale_color_manual(values = c(fav_colors[9], fav_colors[3])) +
  scale_fill_manual(values = c(fav_colors[9], fav_colors[3]))

# Plot
pca_Diagnosis
```




```
pca_parity
```



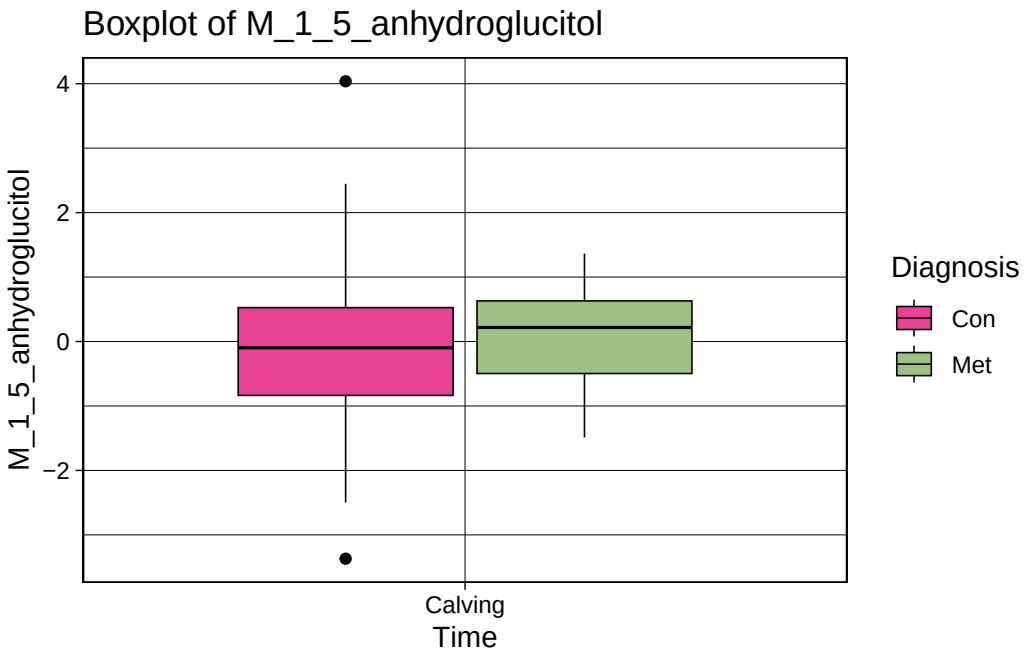
The first principal component accounts for 10.21% of variation, while the second one accounts

for 8.39%; Generally speaking, a PCA that achieves good separation is able to explain about 60-80% of variation in the first three components; The low variance explained in our plot is indicative that we do not have very well delineated groups and that there is overlap between cows with different characteristics. We can see that both diagnosis and parity exhibit significant overlap on the PCA plots. Notably, parity seems to show a slightly better separation between groups. Also notable is that there appear to be some outliers points, which may signify that some cows have a metabolic profile that is deviant even when compared to other cows in their group.

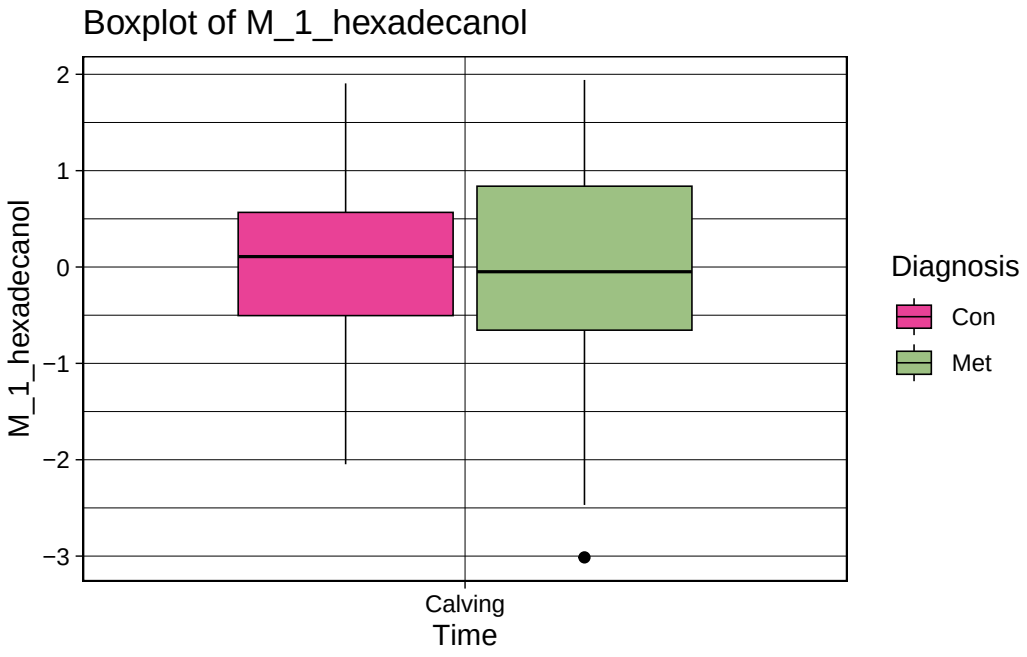
Shenanigans

Then, just for fun, I created a function that would iterate through the metabolites and create a plot for each metabolite with the fill corresponding to the either "Met" or "Con". Below you can find the first 4.

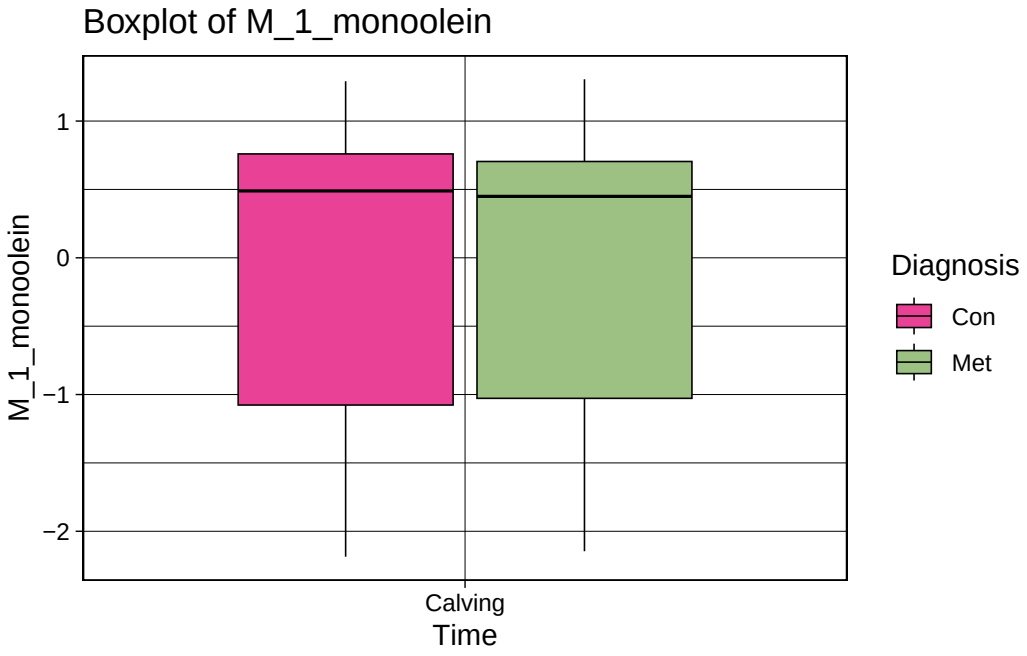
```
# Create a boxplot for each metabolite
metabolite_names <- names(Calving_scaled[5:ncol(Calving_scaled)])

# Create a function!
create_boxplot <- function(metabolite) {
  ggplot(Calving_scaled) +
    geom_boxplot(aes(x = Time, y = get(metabolite), fill = Diagnosis), color = "black", lw
    labs(title = paste0("Boxplot of ", metabolite),
         x = "Time",
         y = metabolite,
         color = "Diagnosis")+
    scale_fill_manual(values = c(fav_colors[3], fav_colors[9]))
}

# Apply the function to each metabolite and store the plots
boxplots <- lapply(metabolite_names, create_boxplot)

par(mfrow = c(2, 2))
boxplots[[1]]
```



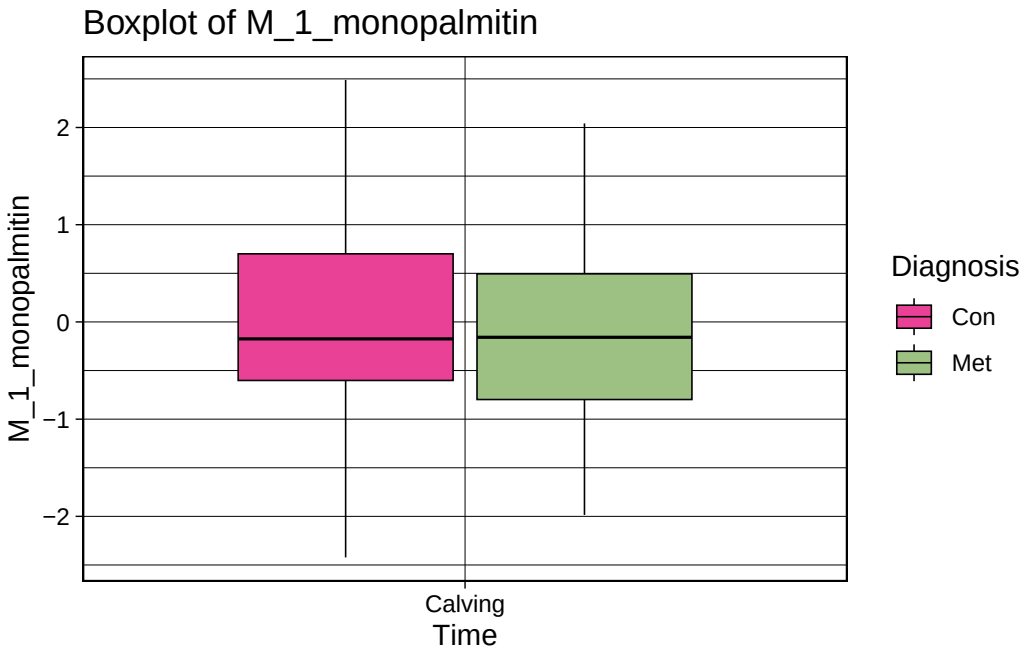

```
boxplots[[2]]
```




```
boxplots[[3]]
```




```
boxplots[[4]]
```




```
par(mfrow = c(1, 1))
```

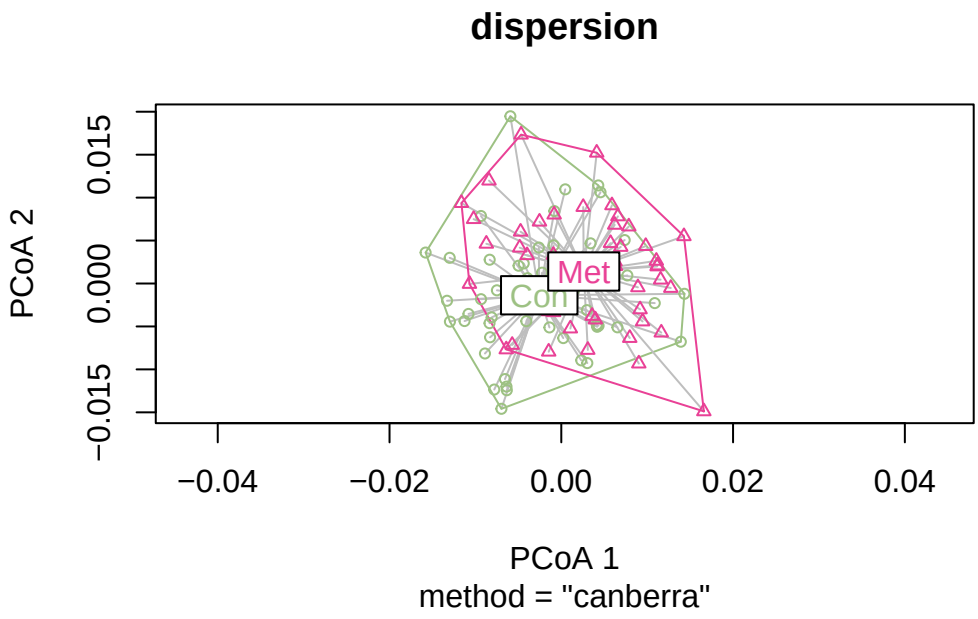
While the PCA plots visually suggest substantial overlap between Metritis and Control cows, I wanted to quantitatively test if these two groups could be considered different. To this end, I employed a PERMANOVA, the explanation and tutorial available on Youtube by creator MADHURAJ P K were referenced for this step.

PERMANOVA is suited for this data because it makes no assumptions about the underlying distribution of the data and it is robust towards multicollinearity. For PERMANOVA to be applicable though, there is one important assumption that I needed to test for: homogeneity of dispersion, or in other words, that the groups being compared have similar dispersion. To test this, I used the betasdisper() function before conducting the ANOVA.

For the betaspider() function, I used "Canberra" distance, which is a weighted form of the Manhattan distance. This is based on the paper by Dixon et al. in which the authors state that Canberra is one of the most repeatable measures for metabolomic data.

PERMANOVA - Homogeneity of Multivariate Dispersion

```
# Select the appropiate data
data_matrix <- Calving_log[5:ncol(Calving_log)]

# Distance
distance = "canberra"

# Check for Homogeneity of Multivariate Dispersion (Betaspider function)

# Distance Matrix
perm_dist <- vegdist(data_matrix, method = distance)

# Assumptions
dispersion <- betadisper(perm_dist, group = Calving_log$Diagnosis, type = "centroid")
plot(dispersion, col = c(fav_colors[9], fav_colors[3]))
```




```
anova(dispersion)
```

```
Analysis of Variance Table

Response: Distances
           Df     Sum Sq    Mean Sq F value Pr(>F)
Groups      1 0.00000218 2.1841e-06  0.2613 0.6103
Residuals 101 0.00084423 8.3588e-06
```

The p-value is 0.6103, much larger than 0.05, thus we reject fail to reject the null since there is not enough evidence to suggest that the dispersion for the groups is different. Now we can move on to the PERMANOVA using the adonis2() function

PERMANOVA

```
set.seed(575)
Perma_result <- adonis2(perm_dist ~ as.factor(Calving_log$Diagnosis),
                        data = data_matrix,
                        permutations = 10000)

Perma_result
```

```
Permutation test for adonis under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 10000

adonis2(formula = perm_dist ~ as.factor(Calving_log$Diagnosis), data = data_matrix, permutati
                                  Df SumOfSqs      R2      F    Pr(>F)
as.factor(Calving_log$Diagnosis)   1 0.001489 0.02708 2.8116 9.999e-05 ***
Residual                         101 0.053481 0.97292
Total                            102 0.054970 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The P-value for this test is 9.99e-05, this the data suggests that the Met and Con groups differ significantly.

Identifying Metabolites of importance

Next, I aimed to identify which metabolites are important to differentiate between the two groups, to do this, I used three approaches:

1. Variable Importance in Projection (VIP) Scoring:

First, I utilized Partial Least Squares Discriminant Analysis (PLS-DA) to identify metabolites with high VIP scores (>1). These scores highlight variables contributing most to group separation by assessing their covariance with the response (Metritis vs Control). Additionally, this model was be used for outcome prediction (although it did not perform very well, as discussed later).

2. Elastic Net for Feature Selection:

Next, I employed a penalized Generalized Linear Model (GLM). Similar to PLS-DA, LASSO will prioritize the most influential metabolites for model building, the retained variables can be extracted and used for subsequent steps.

3. Uni-variate T-test Confirmation:

Finally, aiming to replicate the approach in the reference paper, I performed individual t-tests for each metabolite comparing Metritis and Control groups. This provides complementary uni-variate confirmation of significant differences identified through multivariate methods.

PLS-DA

PLS-DA is a supervised machine learning method, meaning it incorporates response variables (class labels) during model fitting . This contrasts with unsupervised methods like PCA, which don't use class labels. Like PCA, PLS-DA reduces data dimensionality by identifying latent variables (components) that capture the most relevant information . However, PLS-DA focuses retaining the mot covariance between response and predictor variables in it's principal components. PLS-DA does not require independence between predictor variables or assume a distribution for the data.

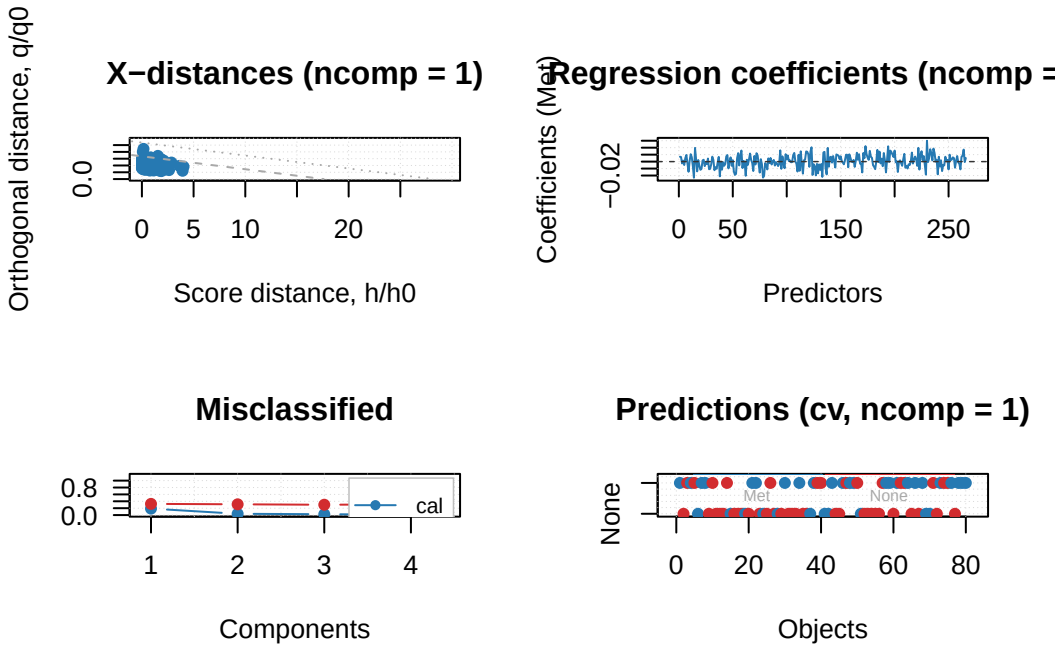
To fit the PLS-DA model, I followed this tutorial put together by Sergey *Kucheryavskiy* which used the plsda() function from the mdatools package which he authored. The code is shown below.

```
# We need to use the "coded" data here and turn it into a matrix
cod_tmp <- as.matrix(Coded_calv[,c(-1,-3)])
resp <- Coded_calv$Diagnosis

# Obrain a random index for splitting the data into training and validation
set.seed(575)
ind <- sample(x = 1:(nrow(cod_tmp)-2),
              size = 80)

# Save data
train_matrix <- cod_tmp[ind,-1]
train_response <- resp[ind] == 1 # Save the response variable at "true" if it is Met, or F

# Validation data
val_matrix <- cod_tmp[-ind,-1]
val_response <- resp[-ind] == 1 # Save the response variable at "true" if it is Met, or Fa
```

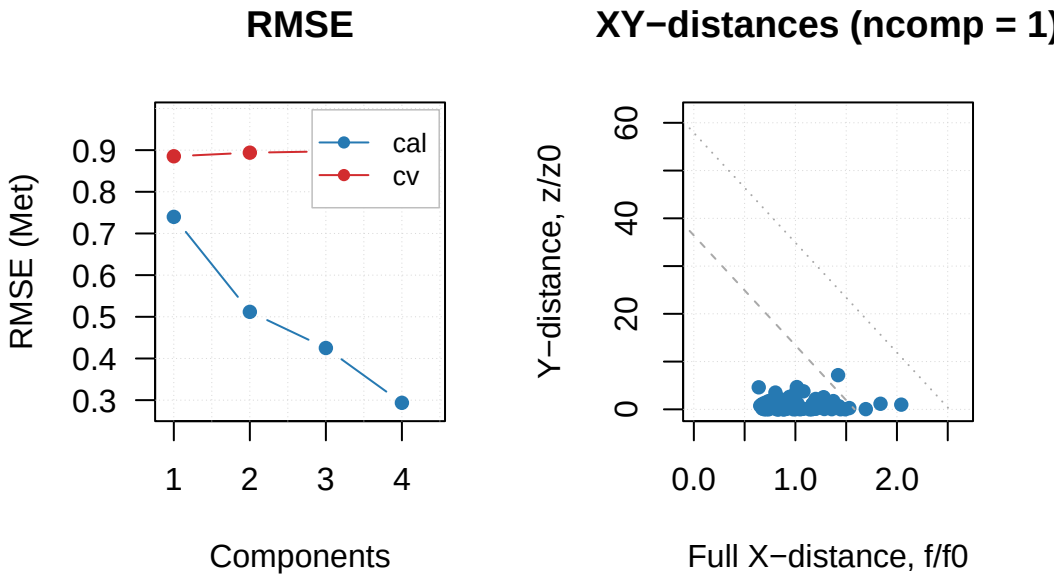
```
# Calibrate model
set.seed(575)
m.all <- mdatools::plsda(train_matrix,
                 train_response,
                 4, cv = 1,
                 classname = "Met",
                 center = F)

summary(m.all)
```

```
PLS-DA model (class plsda) summary
-----------------------------------
Info:
Number of selected components: 1
Cross-validation: full (leave one out)

Class #1 (Met)
    X cumexpvar Y cumexpvar TP FP TN FN Spec. Sens. Accuracy
Cal        8.56       45.26 30 10 35  5 0.778 0.857    0.812
Cv           NA          NA 24 15 30 11 0.667 0.686    0.675
```

```
plot(m.all)
```




```
par(mfrow = c(1, 2))
plotRMSE(m.all)
plotXYResiduals(m.all)
```



Below are the results of this first fit, for the calibration, the cumulative explained variance on the x-axis is 8.56 while it is 45.26 on Y. This model selected a total of 1 component. Now, it appears that our model is over-fitting as the calibrated model achieves an accuracy of ~81% but the cross validation resulted in an overall accuracy of 67.5%. While it is expected the data to do a bit better on the training set, but the difference seems too large here.

X	X cumxpvar	Y cumexpvar	TP	FP	TN	FN	Spec.	Sens.	Accuracy
Cal	8.56	45.26	30	10	35	5	0.778	0.857	0.812
Cv	NA	NA	24	15	30	11	0.667	0.686	0.675

I decided to check if outliers were affecting the results. In the tutorial, Kucheryavskiy references the paper titled "Detection of Outliers in Projection-Based Modeling" by Rodionova and Pomerantsev . Their approach consists of identifying outliers and removing them from the calibration set, and then re-fitting the model, this is repeated until there are no outliers; then, the removed points are predicted using the re-fitted model, and if the residuals are not outliers, then add them back into the calibration set and create a final model. Below is this process.

```
# Let's remove some outliers & repeat
outliers <- which(categorize(m.all) == "extreme")

# keep data for outliers on a separate matrix
Xo <- train_matrix[outliers, , drop = FALSE]
```

```
yo <- train_response[outliers]

# remove data for outliers from training data
X <- train_matrix[-outliers,]
y <- train_response[-outliers]

# make a new model for outlier free data #here
set.seed(575)
m.all <-  mdatools::plsda(X, y, 4, cv = 1, classname = "Met", center = F)
```

Let's repeat the process until we have no outliers

```
###
# 1
###

# Let's remove some outliers & repeat
outliers <- which(categorize(m.all) == "extreme")

# keep data for outliers on a separate matrix
Xo <- rbind(Xo, train_matrix[outliers, , drop = FALSE])
yo <- append(yo, train_response[outliers])

# remove data for outliers from training data
X <- X[-outliers,]
y <- y[-outliers]

# make a new model for outlier free data
set.seed(575)
m.all <-  mdatools::plsda(X, y, 4, cv = 1, classname = "Met", center = F)

# Check for outliers again
which(categorize(m.all) == "extreme")
```

```
[1]  2  3 26 43 70
```

```
###
# 2
###

# Let's remove some outliers & repeat
```

```r
outliers <- which(categorize(m.all) == "extreme")

# keep data for outliers on a separate matrix
Xo <- rbind(Xo, train_matrix[outliers, , drop = FALSE])
yo <- append(yo, train_response[outliers])

# remove data for outliers from training data
X <- X[-outliers,]
y <- y[-outliers]

# make a new model for outlier free data
set.seed(575)
m.all <-  mdatools::plsda(X, y, 4, cv = 1, classname = "Met", center = F)

# Check for outliers again
which(categorize(m.all) == "extreme")
```

```
[1]  3 16 45
```

```r
###
# 3
###

# Let's remove some outliers & repeat
outliers <- which(categorize(m.all) == "extreme")

# keep data for outliers on a separate matrix
Xo <- rbind(Xo, train_matrix[outliers, , drop = FALSE])
yo <- append(yo, train_response[outliers])

# remove data for outliers from training data
X <- X[-outliers,]
y <- y[-outliers]

# make a new model for outlier free data
set.seed(575)
m.all <-  mdatools::plsda(X, y, 4, cv = 1, classname = "Met", center = F)

# Check for outliers again
which(categorize(m.all) == "extreme")
```

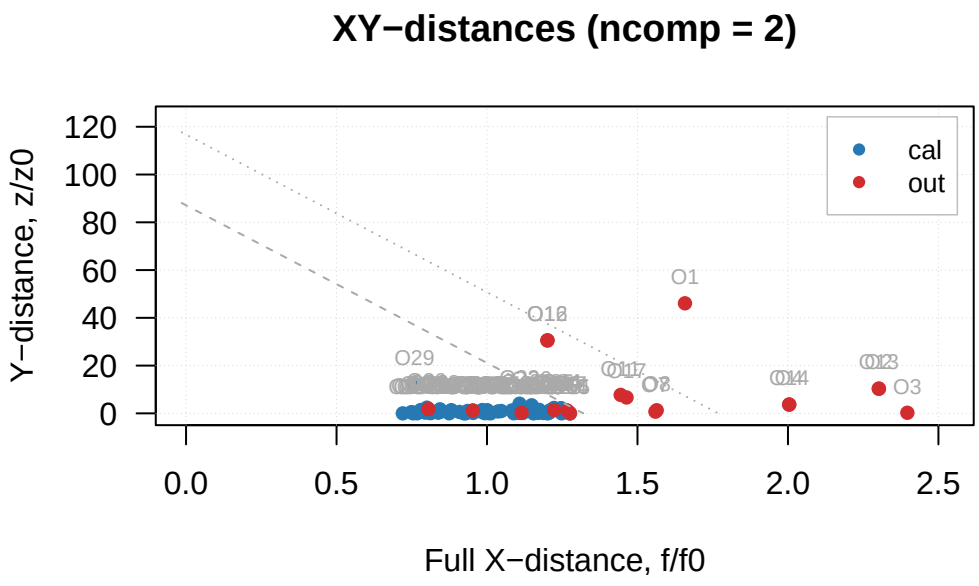
```
integer(0)
```

```
# We have removed all of the outliers

summary(m.all)
```

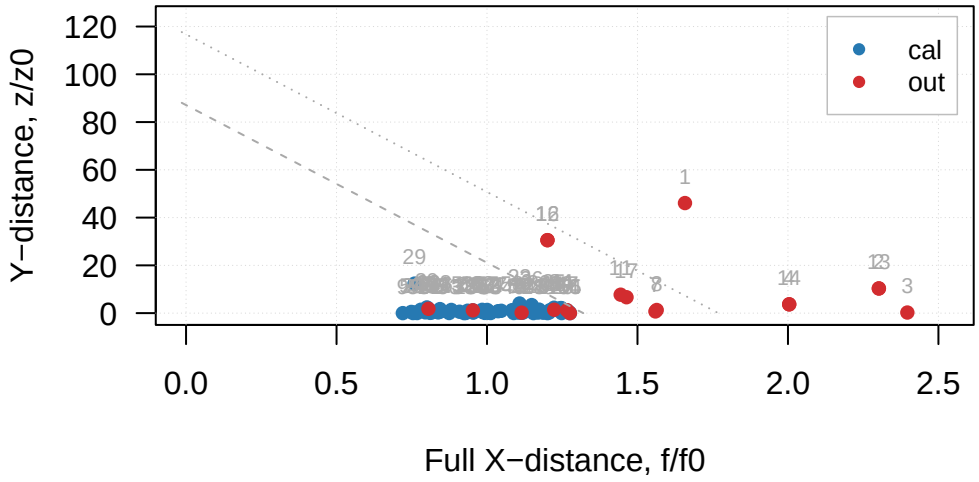
```
PLS-DA model (class plsda) summary
-----------------------------------
Info:
Number of selected components: 2
Cross-validation: full (leave one out)

Class #1 (Met)
    X cumexpvar Y cumexpvar TP FP TN FN Spec. Sens. Accuracy
Cal       14.64       82.91 26  0 35  1   1.0 0.963    0.984
Cv           NA          NA 20  7 28  7   0.8 0.741    0.774
```

This model without outliers seems to be a better fit as it performs better on the "new" data. Next, let's predict the outliers with the model

```
### Now predict them
set.seed(575)
res <- predict(m.all, Xo, yo)


# Now plot the residuals for the predicted outliers
plotXYResiduals(m.all, res = list("cal" = m.all$res$cal,
                                  "out" = res),
                show.labels = TRUE)
```




```
out_ind <- c(-1,-3,-14)

plotXYResiduals(m.all,
                res = list("cal" = m.all$res$cal, "out" = res),
                show.labels = TRUE, labels = "indices")
```



Oh there's as issue here! Per the paper, the next step would be to add back the outliers whose predicted values do not end up being outliers themselves, unfortunately, after working on this

for a while, I have not been able to actually figure out which ones those are, I have them labeled but it doesn't seem like the label actually corresponds to the index of the sample - I will look more into this since I do want to use this method for my own research, but for now there's just not enough time to figure it out. For the following prediction, I will use the model that does not include any outliers - which is acknowledge is not the recommended approach.

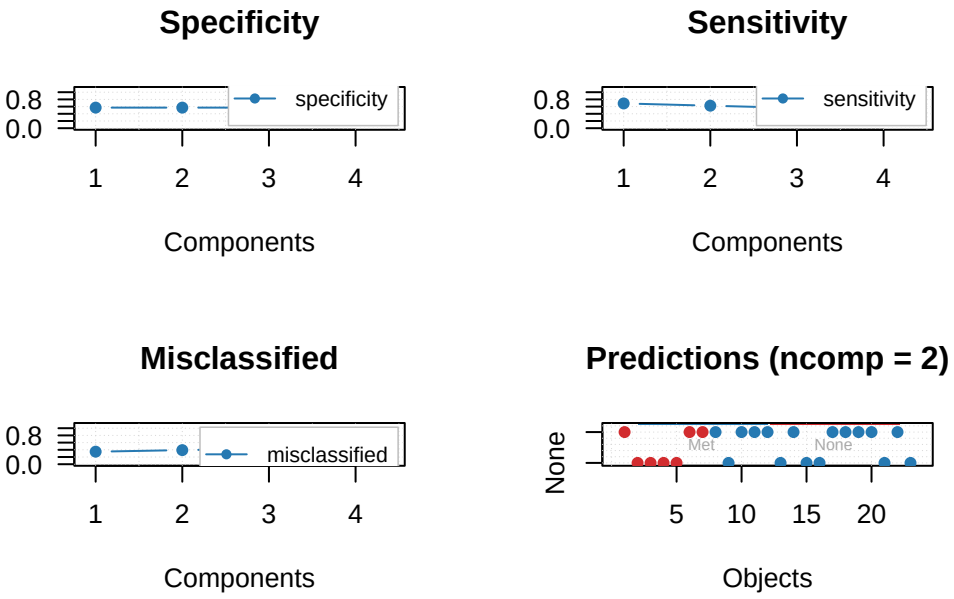
PLS-DA Prediction

Now we test the final model to predict the validation set and evaluate it's performance

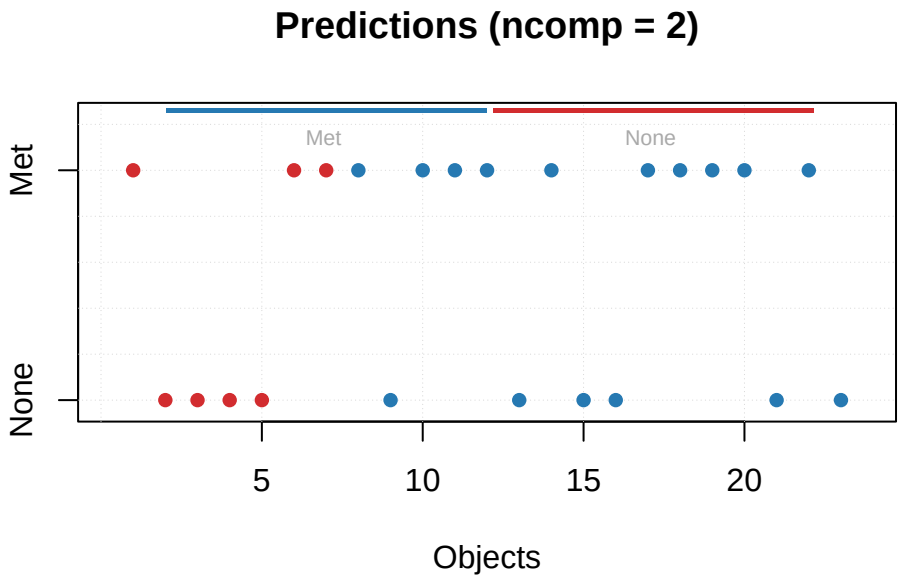
```
set.seed(575)
m.pred <- predict(m.all, val_matrix, val_response)
summary(m.pred)
```

```
PLS-DA results (class plsdares) summary:
Number of selected components: 2

Class #1 (Met):
       X expvar X cumexpvar Y expvar Y cumexpvar TP FP TN FN Spec. Sens.
Comp 1    6.734       6.734   19.160      19.160 11  3  4  5 0.571 0.688
Comp 2    3.009       9.743   -8.513      10.647 10  3  4  6 0.571 0.625
Comp 3    4.835      14.578    9.815      20.462  9  3  4  7 0.571 0.562
Comp 4    2.190      16.768    0.847      21.308 11  4  3  5 0.429 0.688
       Accuracy
Comp 1    0.652
Comp 2    0.609
Comp 3    0.565
Comp 4    0.609
```

```
par(mfrow = c(2, 2))
plotSpecificity(m.pred)
plotSensitivity(m.pred)
plotMisclassified(m.pred)
plotPredictions(m.pred)
```




```
par(mfrow = c(1, 1))

plotPredictions(m.pred)
```



This PLS-DA achieved ~61% accuracy when put to test on the validation set, this is only slightly better than what we would expect from simply guessing "Met" with a probability equal to its proportion in the data (~50%). In other words, the model struggled to accurately

predict for cows it hadn't encountered during training. This suggests limited generalizability and does raise concerns about the model's effectiveness for classifying unseen Holstein cow samples.

Despite the low predictive performance, some insights might still be gained from the analysis in the form of variable importance in projection (VIP) scores. As briefly discussed above, these scores highlight which metabolic features the model considers most influential in discriminating between "Met" and "Non-Met" states. However, given the model's low accuracy, it's important to interpret these VIP scores with caution as they may primarily reflect the specific characteristics of the training data (these Holstein cows) rather than providing reliable insights applicable to the broader Holstein population.

PLS-DA VIP Scores

Here, I extract the VIP scores for all input variables in the PSL-DA model (parity & all metabolites), then, I save the name and score of those exceeding a threshold of 1 (a common criterion for identifying important variables). This data frame will re-apear later in the project (see log fold change)

```
vipscore <- as.data.frame(vipscores(m.all)) |>
  rownames_to_column("Variable") |>
  arrange(desc(Met))

vipscore
```

```
                          Variable        Met
1                      M_serotonin 2.55885716
2                M_O_acetylserine 2.32016020
3          M_4_hydroxybutyric_acid 2.29874362
4                      M_tagatose 2.14597734
5                M_stearic_acid 2.05894839
6           M_methionine_sulfoxide 2.03554501
7                  M_benzoic_acid 2.00063522
8            M_heptadecanoic_acid 1.99546774
9          M_isoheptadecanoic_acid 1.95980883
10                     M_fructose 1.90496462
11       M_indole_3_propionic_acid 1.90329827
12                 M_ciliatine 1.89731518
13                   M_ribitol 1.89629605
14                 M_myo_inositol 1.85899015
15            M_pentadecanoic_acid 1.83292103
16            M_hydrocinnamic_acid 1.82027791
```

17	M_3_phosphoglycerate	1.80055896
18	M_pantothenic_acid	1.76092055
19	M_nonadecanoic_acid	1.73879653
20	M_ribose_5_phosphate	1.73651991
21	M_palmitic_acid	1.67481493
22	M_citric_acid	1.67330365
23	M_glycyl_tyrosine	1.66918915
24	M_arachidonic_acid	1.66137609
25	M_erythritol	1.63265591
26	M_p_hydroxylphenyllactic_acid	1.63080489
27	M_maltose	1.61319596
28	M_xylitol	1.60851013
29	M_glyceric_acid	1.59345225
30	M_cysteine	1.57439140
31	M_oxalic_acid	1.56707877
32	M_hippuric_acid	1.56544965
33	M_indoxyl_sulfate	1.55174853
34	M_phenylacetic_acid	1.54781513
35	M_2_hydroxy_2_methylbutanoic_acid	1.53968852
36	M_kynurenine	1.49112648
37	M_glucose_1_phosphate	1.49097179
38	M_glycerol	1.48040637
39	M_serine	1.47384870
40	M_alpha_ketoglutarate	1.45259488
41	M_threonic_acid	1.45246894
42	M_cholesterol	1.44574393
43	M_2_8_dihydroxyquinoline	1.42967239
44	M_2_hydroxybutanoic_acid	1.42412371
45	M_uric_acid	1.42191707
46	M_9_myristoleate	1.40259404
47	M_hypoxanthine	1.39711198
48	M_p_tolyl_glucuronide	1.38664059
49	M_spermidine	1.37044535
50	M_arachidic_acid	1.35921372
51	M_beta_glycerolphosphate	1.35856744
52	M_palmitoleic_acid	1.35628133
53	M_isocitric_acid	1.34491750
54	M_oleic_acid	1.31820319
55	M_3_aminoisobutyric_acid	1.30103766
56	M_myristic_acid	1.29356459
57	M_glycerol_alpha_phosphate	1.29234735
58	M_4_aminobutyric_acid	1.29146914
59	M_catechol	1.28913476

60	M_N_acetylglycine	1.27814443
61	M_lyxitol	1.27281429
62	M_ornithine	1.24957239
63	M_2_ketoglucose_dimethylacetal	1.24215504
64	M_allantoic_acid	1.22014038
65	M_isopentadecanoic_acid	1.21780689
66	M_tryptophol	1.19255068
67	M_lithocholic_acid	1.18469864
68	M_citrulline	1.17610873
69	M_linoleic_acid	1.15290324
70	M_fucose	1.14941734
71	M_saccharic_acid	1.13940306
72	M_3_hydroxypropionic_acid	1.13399274
73	M_glucose	1.13045342
74	M_guanine	1.11372047
75	M_alanine	1.11144163
76	M_nicotinamide	1.08048991
77	M_xylulose	1.06435033
78	M_xanthosine	1.06301744
79	M_glycerol_3_galactoside	1.06077900
80	M_indole_3_lactate	1.05402775
81	M_tryptophan	1.04750356
82	M_2_ketobutyric_acid	1.04132294
83	M_trans_4_hydroxyproline	1.03496288
84	M_5_methoxytryptamine	1.02329415
85	M_2_hydroxyhexanoic_acid	1.00680158
86	M_alanine_alanine	0.98511678
87	M_methylmaleic_acid	0.97282992
88	M_1_5_anhydroglucitol	0.96783693
89	M_D_erythro_sphingosine	0.96776201
90	M_3_phenyllactic_acid	0.96560748
91	M_citramalic_acid	0.96518217
92	M_2_methylglyceric_acid	0.96309312
93	M_glycine	0.95904630
94	M_threonine	0.95681456
95	M_ribose	0.95469682
96	M_glycolic_acid	0.94465828
97	M_methanolphosphate	0.93963464
98	M_tyrosine	0.92987929
99	M_erythrose	0.92183949
100	M_asparagine	0.91873990
101	M_tocopherol_alpha_	0.91357552
102	M_2__deoxyguanosine	0.89644837

103	M_phytanic_acid	0.89433523
104	M_inositol_4_monophosphate	0.89059486
105	M_1_monoolein	0.88407145
106	M_sorbitol	0.88369221
107	M_docosahexaenoic_acid	0.87485039
108	M_Mevalonic_acid	0.87437114
109	M_indole_3_acetate	0.86417461
110	M_guanosine	0.86117033
111	M_ethanol_phosphate	0.85998481
112	M_mannose	0.85821928
113	M_2_5_dihydroxypyrazine	0.85463695
114	M_valine	0.84598049
115	M_lactitol	0.84285140
116	M_aconitic_acid	0.83257799
117	M_uracil	0.83228086
118	M_sarcosine	0.82795141
119	M_succinate_semialdehyde	0.82226167
120	M_2_deoxytetronic_acid	0.82216828
121	M_inosine	0.81673661
122	M_cellobiose	0.81046926
123	M_N_acetylaspartic_acid	0.80925590
124	M_4_hydroxyphenylacetic_acid	0.80723689
125	M_ethanolamine	0.80515219
126	M_lysine	0.80502185
127	M_ribonic_acid	0.80494535
128	M_cysteine_glycine	0.80238306
129	M_homoserine	0.79813510
130	M_glutaric_acid	0.79613419
131	M_lyxose	0.79171162
132	M_5_6_dihydrouracil	0.78659408
133	Parity	0.78451084
134	M_2_aminobutyric_acid	0.78049520
135	M_pseudo_uridine	0.77968309
136	M_glycyl_proline	0.77666902
137	M_melibiose	0.75375388
138	M_3__4_hydroxyphenyl_propionic_acid	0.73616684
139	M_cerotinic_acid	0.73207828
140	M_lactamide	0.72870453
141	M_beta_glutamic_acid	0.70841708
142	M_lactose	0.70208782
143	M_pipecolinic_acid	0.70202392
144	M_glycocyamine	0.69790692
145	M_undecanoic_acid	0.69630924

146	M_histidine	0.69130447
147	M_aspartic_acid	0.68486301
148	M_taurine	0.68283202
149	M_isoleucine	0.67336609
150	M_octadecylglycerol	0.67324863
151	M_glutamine	0.65722499
152	M_sophorose	0.64806735
153	M_2_hydroxyglutaric_acid	0.64736856
154	M_beta_alanine	0.64639926
155	M_2_monopalmitin	0.64521708
156	M_phenylethylamine	0.63043260
157	M_quinolinic_acid	0.63040481
158	M_1_monopalmitin	0.62910815
159	M_3_hydroxybutyric_acid	0.62604606
160	M_phosphate	0.62367531
161	M_3_hydroxy_3_methylglutaric_acid	0.62082004
162	M_oleamide	0.61852753
163	M_norvaline	0.60405077
164	M_lactic_acid	0.60004444
165	M_isothreonic_acid	0.59900107
166	M_piperidone	0.59505160
167	M_tyrosol	0.58908691
168	M_3_hydroxypalmitic_acid	0.57943673
169	M_cholesterone	0.57885960
170	M_parabanic_acid	0.57863589
171	M_pyrrole_2_carboxylic_acid	0.57575579
172	M_2_ketoadipic_acid	0.57395770
173	M_maleimide	0.56368535
174	M_cytidine	0.56271365
175	M_fructose_6_phosphate	0.56119151
176	M_2_6_diaminopimelic_acid	0.56076645
177	M_isoribose	0.55781834
178	M_lauric_acid	0.55777746
179	M_trehalose	0.55209260
180	M_N_acetyl_D_galactosamine	0.54460084
181	M_adenosine_5_monophosphate	0.53462194
182	M_lactobionic_acid	0.53364775
183	M_epsilon_caprolactam	0.53097191
184	M_hydroxycarbamate	0.53039679
185	M_pentitol	0.52931758
186	M_thymine	0.51554898
187	M_N_acetylmannosamine	0.51450549
188	M_orotic_acid	0.51096904

189	M_gluconic_acid	0.50981994
190	M_lignoceric_acid	0.50581788
191	M_glutamic_acid	0.50561519
192	M_xylose	0.49508962
193	M_fructose_1_phosphate	0.48863024
194	M_putrescine	0.48763519
195	M_thymidine	0.48081871
196	M_methionine	0.47766462
197	M_pyruvic_acid	0.47760003
198	M_N_acetylornithine	0.47372115
199	M_cis_gondoic_acid	0.47110858
200	M_5_aminovaleric_acid	0.45799614
201	M_n_epsilon_trimethyllysine	0.45449260
202	M_O_phosphoserine	0.45031219
203	M_phenaceturic_acid	0.43958630
204	M_alpha_aminoadipic_acid	0.43821201
205	M_UDP_GlcNAc	0.43742692
206	M_cystine	0.42139859
207	M_adenine	0.41931495
208	M_ile_ile	0.41879960
209	M_uridine	0.41673371
210	M_5_hydroxynorvaline	0.39519366
211	M_N_carbamoylaspartate	0.39350249
212	M_1_hexadecanol	0.39170463
213	M_beta_gentiobiose	0.39081931
214	M_nicotinic_acid	0.38926832
215	M_squalene	0.38350078
216	M_succinic_acid	0.38151301
217	M_beta_sitosterol	0.38080415
218	M_octadecanol	0.37708141
219	M_malonic_acid	0.37626933
220	M_leucine	0.36821103
221	M_adenosine	0.36463062
222	M_galactose_6_phosphate	0.36254723
223	M_2_ketoisocaproic_acid	0.36051815
224	M_deoxycholic_acid	0.35784061
225	M_butyrolactam	0.35627420
226	M_isolinoleic_acid	0.35610180
227	M_creatinine	0.34377437
228	M_phosphoethanolamine	0.34239610
229	M_pyrophosphate	0.34232321
230	M_proline	0.34129102
231	M_urocanic_acid	0.34031026

```
232               M_glucose_6_phosphate 0.33358478
233                     M_pimelic_acid 0.32762417
234                         M_threitol 0.32405004
235                    M_1_monostearin 0.31326138
236                       M_lanosterol 0.31217608
237                      M_cholic_acid 0.31095003
238                      M_maltotriose 0.30580129
239                    M_phenylalanine 0.30407930
240            M_4_hydroxycinnamic_acid 0.30307217
241                  M_2_picolinic_acid 0.29232349
242                       M_galactinol 0.28331994
243             M_2_ketoisovaleric_acid 0.27964082
244                     M_behenic_acid 0.27434798
245                  M_dihydroxyacetone 0.27289301
246                     M_shikimic_acid 0.27255762
247               M_phosphoenolpyruvate 0.27014378
248                     M_elaidic_acid 0.26411826
249                     M_fumaric_acid 0.26147175
250                  M_2_deoxypentitol 0.25257652
251                      M_2_monoolein 0.23729016
252                          M_sucrose 0.23309548
253           M_conduritol_beta_epoxide 0.23210964
254                       M_oxoproline 0.22161059
255                       M_malic_acid 0.21095777
256                          M_cytosin 0.21074281
257              M_ribulose_5_phosphate 0.20116376
258             M_2_hydroxyvaleric_acid 0.19730015
259                M_3_ureidopropionate 0.19517811
260                         M_xanthine 0.19356596
261                    M_itaconic_acid 0.18690400
262                              M_urea 0.17088453
263                    M_aminomalonate 0.16762180
264               M_dehydroabietic_acid 0.10965846
265                      M_adipic_acid 0.06415806
266               M_mannose_6_phosphate 0.04669741
```

```
nrow(vipscore[vipscore$Met > 1,])
```

```
[1] 85
```

```r
# There are 85 metabolites with a VIP score greater than 1, these will be saved as a dataf

VIPdf <- vipscore[vipscore$Met > 1,]
```

GLM with Elastic Net

To fit a GLM that is penalized with elastic net I used the nestedcv[1] package which utilizes *glmnet* functions but also includes the (fantastic) nestedcv.train() function which makes it possible to fit a model using *nested* cross validation, for this analysis:

- I allowed the function to pick the best lambda.
- I used the filterFUN option to allow for filter the features based on their t-test values, this was supposed to facilitated model fitting and aligns with the subsequent uni variate t-test analysis.
- Used 10 outer and inner folds; this works as follows:
  - Outer Loop: The whole dataset is divided into 10 segments, each taking a turn as a holdout set to assess model performance on "new" data, while the remaining 9 segments are used in the inner loop to train the model
  - Inner Loop: Within each outer fold's training stage, the 9 segments are turned into 10 (10-fold inner loop) and used for tuning the model's hyperparameters, for this analysis, this is specifically the elastic net penalty lambda, which determines how harshly the penalization is. Alpha is set to 1 and does not get tuned.
  - The overall best performing is stored under "final_fit" inside the object
- Finally, to avoid overfitting and removing features that might nor be contributing strongly to the model's performance, I set the elastic net penalty alpha equal to 1, which corresponds to LASSO penalty, as opposed to 0 which corresponds to ridge regression; The former removes features, while the latter shrinks coefficients without removing features altogether.

  The final model retained 69 predictors which are shown below. However, after multiple iterations, I wasn't able to obtain a balanced accuracy higher than 63% - whlile lower than I was hoping, that the accuracy is consistent suggests that this is indeed the best this approach can do with the data and might just be a reflection on the lack of significant separation between groups.

```r
# We need to use the "coded" data here and turn it into a matrix
cod_tmp <- Coded_calv[,5:ncol(Coded_calv)]
resp <- Calving_scaled$Diagnosis
```

```r
# A holdout/validation set is not required for nested cross validation
tg <- expand.grid(lambda = exp(seq(log(2e-3), log(1e0), length.out = 6)),
                  alpha = 1)

set.seed(575)
lasso_fit <- nestcv.train(y = resp,
                          x = cod_tmp,
                          filterFUN = "ttest_filter",
                          method = "glmnet",
                          n_outer_folds = 10,
                          n_inner_folds = 10,
                          tuneGrid = tg)
```

```r
summary(lasso_fit)
```

```
Nested cross-validation with caret
Method:  glmnet
Filter:  ttest_filter
Outer loop:  10-fold cv
Inner loop:  10-fold cv
103 observations, 265 predictors
Con Met
 52  51

         alpha   lambda  n.filter
Fold 1       1  0.08326        59
Fold 2       1  0.02402        63
Fold 3       1  0.02402        63
Fold 4       1  0.08326        67
Fold 5       1  0.02402        53
Fold 6       1  0.02402        63
Fold 7       1  0.02402        54
Fold 8       1  0.08326        62
Fold 9       1  0.02402        57
Fold 10      1  0.02402        58

Final parameters:
  alpha   lambda
      1  0.08326

Result:
```

```
         Reference
Predicted Con Met
      Con  34  20
      Met  18  31
```

```
             AUC            Accuracy   Balanced accuracy
          0.6953              0.6311              0.6308
```

```
lasso_vars <- lasso_fit$final_vars
lasso_vars
```

```
 [1] "M_ribitol"                          "M_serine"
 [3] "M_indole_3_propionic_acid"          "M_serotonin"
 [5] "M_myo_inositol"                     "M_2_hydroxy_2_methylbutanoic_acid"
 [7] "M_2_hydroxybutanoic_acid"           "M_heptadecanoic_acid"
 [9] "M_erythritol"                       "M_isoheptadecanoic_acid"
[11] "M_oxalic_acid"                      "M_palmitic_acid"
[13] "M_citric_acid"                      "M_stearic_acid"
[15] "M_benzoic_acid"                     "M_citrulline"
[17] "M_hydrocinnamic_acid"               "M_sorbitol"
[19] "M_alanine_alanine"                  "M_xylitol"
[21] "M_allantoic_acid"                   "M_glycerol"
[23] "M_4_hydroxybutyric_acid"            "M_palmitoleic_acid"
[25] "M_hypoxanthine"                     "M_arachidonic_acid"
[27] "M_isocitric_acid"                   "M_2_hydroxyhexanoic_acid"
[29] "M_ornithine"                        "M_myristic_acid"
[31] "M_hippuric_acid"                    "M_O_acetylserine"
[33] "M_mannose"                          "M_glycerol_alpha_phosphate"
[35] "M_indoxyl_sulfate"                  "M_9_myristoleate"
[37] "M_phenylacetic_acid"                "M_erythrose"
[39] "M_threonine"                        "M_nonadecanoic_acid"
[41] "M_pentadecanoic_acid"               "M_ciliatine"
[43] "M_fructose_1_phosphate"             "M_spermidine"
[45] "M_alanine"                          "M_p_tolyl_glucuronide"
[47] "M_lactamide"                        "M_lyxitol"
[49] "M_linoleic_acid"                    "M_nicotinamide"
[51] "M_2_ketoadipic_acid"                "M_valine"
[53] "M_2_ketobutyric_acid"               "M_2_5_dihydroxypyrazine"
[55] "M_succinate_semialdehyde"           "M_indole_3_acetate"
[57] "M_lactic_acid"                      "M_tagatose"
[59] "M_adenosine_5_monophosphate"        "M_tryptophol"
```

```
[61] "M_3_phosphoglycerate"              "M_glyceric_acid"
[63] "M_fucose"                          "M_glycolic_acid"
[65] "M_ribose_5_phosphate"              "M_alpha_ketoglutarate"
[67] "M_3_aminoisobutyric_acid"          "M_asparagine"
[69] "M_beta_glycerolphosphate"
```

T-test

As the final step in my analysis pipeline, I performed an univariate t-test; that is, performed a t-test comparing one metabolite at a time in relation to the Diagnosis group. All tests followed the formula *Diagnosis* ~ *Metabolite*. The null hypothesis for each of this t-test is that there's no significant difference in the abundance of a given metabolite between the two groups. I selected a cutoff of 0.05; while I considered a stricted cutoff to minimize the likelihood of false positives, the goal of this analysis is to discovering potentially important metabolites and metabolic pathways and a stricter p-value might lead to the exclusion of real and important but subtle differences between groups.

As discussed in lecture (which was very helpful!) the obtained p-values are a measure of the probability of observing the data (or more extreme results) if the null hypothesis (no difference) were true.

- A p-value below the threshold indicates that the chance of observing this differences due to chance is fairly small, and thus leads to the rejection of the null; the interpretation of this conclusion is that the differences are likely real.
- On the other hand, a value above the 0.05 threshold indicates a higher chance of obtaining these values by chance and leads to the failure to reject the null; This means we don't have enough evidence to conclude a difference exists.

After the t-test, the metabolites with p-values smaller that 0.05 were retained in a variable called "sig_metabolites"

```
# For t-test, remove CowNumber and Time
t_data <- Calving_log |>
  dplyr::select(c(-CowNumber, -Time)) |>
  dplyr::mutate(Diagnosis = as.numeric(Diagnosis) - 1,
                Parity = as.numeric(Parity) - 1)


# Do the t-test for each metabolite
metabolite_names <- names(Calving_log[4:ncol(Calving_log)])
t.tests <- lapply(metabolite_names,
                  function(x) t.test(reformulate("Diagnosis", x),
```

```
                                    data = t_data))

# Create a vector containing all the p-values
p_list <- c()
for (i in 1:length(t.tests)){
  p_list <- append(p_list,(t.tests[[i]]$p.value))
}

# Get index for p-values smaller than 0.01 and obtain the entries
sig_metabolites <- metabolite_names[which(p_list < 0.05)]
```

Putting all the analyses together

Now, I retained the metabolites that were retained by the PLS-DA fit, the LASSO fit *and* had a p-value smaller than 0.05. This approach increased my confidence that these features are important for group separation

```
# Save the metabolites names for metabolites that had a high VIP score, were retained by L

validated_metabolites <- intersect(sig_metabolites, VIPdf$Variable)
validated_metabolites <- intersect(validated_metabolites, lasso_vars)

# Create a dataframe with colMeans for cows in the Met group, and a separate one with colM
met_df <- Calving_log |>
  filter(Diagnosis == "Met") |>
  dplyr::select(c(-Parity, -Time, - Diagnosis, -CowNumber)) |>
  colMeans()

con_df <- Calving_log |>
  filter(Diagnosis == "Con") |>
  dplyr::select(c(-Parity, -Time, - Diagnosis, -CowNumber)) |>
  colMeans()

tmp_df <- tibble(names = names(met_df), met = met_df, con = con_df)


# Now for plotting the log change modify the columns so that they can be read by ggplot

change_data <- tmp_df |>
  filter(names %in% validated_metabolites) |>
  mutate(met = as.numeric(met),
```

```
          con = as.numeric(con),
          change = con - met)
```

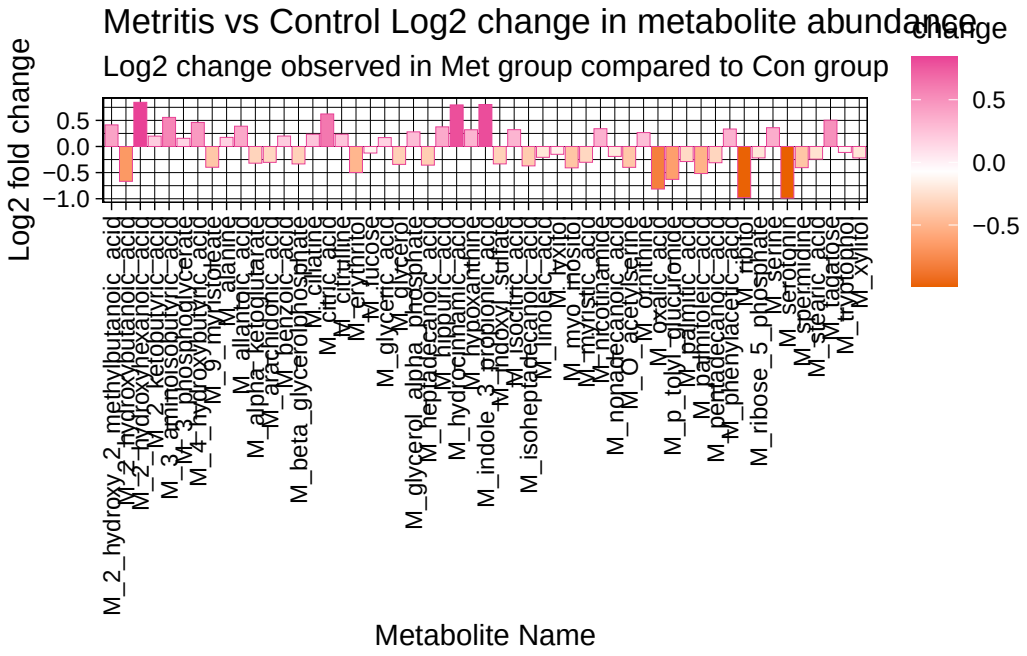
Log Fold Change and Conclusion

This final plot summarizes the metabolites significantly altered between the Met and Control groups, identified by all three employed statistical tests (PLS-DA, GLM-LASSO, and t-tests). It presents the log2 fold change, which reflects the magnitude of change in metabolite levels between the groups on a logarithmic scale. Positive values indicate higher abundance in the Met group, while negative values mean lower. For example, a log2 fold change of 0.5 signifies that the metabolite in Met is 1.5 times as abundant compared to Control, and vice versa for -0.5.

Metabolites with an absolute log2 fold increase value of 0.5 or higher are shown below (n=12).

[Thanks to user i.sudbery, who teaches bioinformatics at the University of Sheffieldat for providing a great explanation for this process on the biostar forum!]

```
ggplot() +
  geom_bar(aes(x = names,
               y = change,
               fill = change),
           stat = "identity",
           color = fav_colors[3],
           lwd = .1,
           data = change_data) +
  labs(title = "Metritis vs Control Log2 change in metabolite abundance",
       subtitle = "Log2 change observed in Met group compared to Con group") +
  xlab(label = "Metabolite Name") +
  ylab(label = "Log2 fold change") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_fill_gradientn(values=c(1, .5, 0), colours = c(fav_colors[3], "white", fav_colors[
```




```
change_data |>
  filter(abs(change) > 0.5) |>
  mutate(metabolite_name = str_replace(names, "M_", "")) |>
  select(metabolite_name, change) |>
  arrange(desc(change))
```

```
# A tibble: 12 x 2
   metabolite_name          change
   <chr>                     <dbl>
 1 2_hydroxyhexanoic_acid    0.846
 2 indole_3_propionic_acid   0.804
 3 hydrocinnamic_acid        0.796
 4 citric_acid               0.623
 5 3_aminoisobutyric_acid    0.557
 6 tagatose                  0.503
 7 palmitoleic_acid         -0.517
 8 p_tolyl_glucuronide      -0.629
 9 2_hydroxybutanoic_acid   -0.667
10 oxalic_acid              -0.814
11 ribitol                  -0.979
12 serotonin                -0.988
```

The appropriate next step for this analysis would be to put this information in a biologically relevant context by performing a functional enrichment analysis, this would translate the metabolite data into potential metabolic pathways that might be relevant for metritis susceptibility at time of calving for the cows included in this study.

As discussed before, the consistent lower performance of this dataset on "outside" data is of concern here though, and while might lead to biological insights, it is important to proceed with caution to avoid incorrectly generalizing to the entire population. We can conclude that changes between Met and Con were present, but I would find it troubling to assume that these changes represent a "Metritis" profile. More data and further analysis would be required to make these findings more generalizable.

References

1. Casaro, S. et al. Blood metabolomics and impacted cellular mechanisms during transition into lactation in dairy cows that develop metritis. J. Dairy Sci. 106, 8098–8109 (2023).
2. Metritis in Production Animals - Reproductive System. MSD Veterinary Manual https://www.msdvetmanual.com/reproductive-system/uterine-diseases-in-production-animals/metritis-in-production-animals.
3. https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf
4. PERMANOVA in R | Explanation and Implementation. (2023).
5. Dixon, P., Wu, L., Widrlechner, M. & Wurtele, E. Weighted distance measures for metabolomic data. (2009).
6. Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K. & Narasimhan, G. So you think you can PLS-DA? BMC Bioinformatics 21, 2 (2020).
7. Kucheryavskiy, S. PLS Discriminant Analysis | Getting started with mdatools for R.
8. Kucheryavskiy, S. mdatools – R package for chemometrics. Chemom. Intell. Lab. Syst. 198, 103937 (2020).
9. Rodionova, O. Ye. & Pomerantsev, A. L. Detection of Outliers in Projection-Based Modeling. Anal. Chem. 92, 2656–2664 (2020).
10. Lewis, M., Spiliopoulou, A. & Goldmann, K. nestedcv: Nested Cross-Validation with 'glmnet' and 'caret'. (2022).