

Dimensionality reduction

Exercices

PCA

1. Load the `nutrimouse` data from the `mixOmics` R package and investigate its structure.

```
library(mixOmics)
```

A data object provided by an R package can be loaded with `data` . Its structure can be obtained with `str` , `length` , `dim` , etc.

```
data("nutrimouse")  
## display the structure of the nutrimouse object  
str(nutrimouse)
```

```

## List of 4
## $ gene      : 'data.frame':   40 obs. of  120 variables:
## ..$ X36b4    : num [1:40] -0.42 -0.44 -0.48 -0.45 -0.42 -0.43 -0.53 -0.49 -0.36
-0.5 ...
## ..$ ACAT1    : num [1:40] -0.65 -0.68 -0.74 -0.69 -0.71 -0.69 -0.62 -0.69 -0.66
-0.62 ...
## ..$ ACAT2    : num [1:40] -0.84 -0.91 -1.1 -0.65 -0.54 -0.8 -1 -0.91 -0.74 -0.79
...
## ..$ ACBP     : num [1:40] -0.34 -0.32 -0.46 -0.41 -0.38 -0.32 -0.44 -0.37 -0.39
-0.36 ...
## ..$ ACC1     : num [1:40] -1.29 -1.23 -1.3 -1.26 -1.21 -1.13 -1.22 -1.29 -1.15 -
1.21 ...
## ..$ ACC2     : num [1:40] -1.13 -1.06 -1.09 -1.09 -0.89 -0.79 -1 -1.06 -1.08 -0.
82 ...
## ..$ ACOTH    : num [1:40] -0.93 -0.99 -1.06 -0.93 -1 -0.93 -0.94 -1.05 -0.88 -0.
92 ...
## ..$ ADISP    : num [1:40] -0.98 -0.97 -1.08 -1.02 -0.95 -0.97 -0.94 -1.02 -0.98
-0.99 ...
## ..$ ADSS1    : num [1:40] -1.19 -1 -1.18 -1.07 -1.08 -1.07 -1.05 -1.16 -1.05 -1
...
## ..$ ALDH3    : num [1:40] -0.68 -0.62 -0.75 -0.71 -0.76 -0.75 -0.67 -0.75 -0.66
-0.69 ...
## ..$ AM2R     : num [1:40] -0.59 -0.58 -0.66 -0.65 -0.59 -0.55 -0.66 -0.66 -0.53
-0.62 ...
## ..$ AOX      : num [1:40] -0.16 -0.12 -0.16 -0.17 -0.31 -0.23 -0.09 -0.22 -0.06
-0.23 ...
## ..$ BACT     : num [1:40] -0.22 -0.32 -0.32 -0.32 -0.31 -0.29 -0.25 -0.21 -0.15
-0.2 ...
## ..$ BIEN     : num [1:40] -0.89 -0.88 -0.89 -0.77 -0.97 -0.84 -0.86 -0.9 -0.74 -
0.76 ...
## ..$ BSEP     : num [1:40] -0.69 -0.6 -0.7 -0.67 -0.68 -0.55 -0.67 -0.66 -0.6 -0.
58 ...
## ..$ Bcl.3    : num [1:40] -1.18 -1.07 -1.17 -1.12 -0.93 -1.08 -1.03 -1.01 -1.01
-1.1 ...
## ..$ C16SR    : num [1:40] 1.66 1.65 1.57 1.61 1.66 1.7 1.58 1.62 1.72 1.55 ...
## ..$ CACP     : num [1:40] -0.92 -0.87 -1.02 -0.89 -0.93 -0.97 -0.97 -0.96 -0.85
-0.95 ...
## ..$ CAR1     : num [1:40] -0.97 -0.92 -0.98 -0.97 -1.06 -1.03 -0.91 -1.11 -0.85
-0.99 ...
## ..$ CBS      : num [1:40] -0.26 -0.36 -0.4 -0.39 -0.35 -0.31 -0.32 -0.4 -0.26 -
0.39 ...
## ..$ CIDEA    : num [1:40] -1.21 -1.17 -1.29 -1.18 -1.15 -1.14 -1.16 -1.26 -1.12
-1.08 ...
## ..$ COX1     : num [1:40] -1.11 -1.06 -1.17 -1.03 -0.99 -1.03 -1.15 -1.18 -0.94
-1.07 ...
## ..$ COX2     : num [1:40] -1.18 -1.06 -1.14 -1.13 -1.1 -1.16 -1.06 -1.24 -1.23 -
1.09 ...
## ..$ CPT2     : num [1:40] -0.87 -0.87 -0.95 -0.88 -0.91 -0.92 -0.86 -0.93 -0.82
-0.88 ...
## ..$ CYP24    : num [1:40] -1.37 -1.14 -1.3 -1.27 -1.2 -1.11 -1.12 -1.3 -1.14 -1.
08 ...
## ..$ CYP26    : num [1:40] -1.21 -1.12 -1.22 -1.18 -1.16 -1.1 -1.07 -1.23 -1.1 -
1.1 ...
## ..$ CYP27a1  : num [1:40] -0.71 -0.62 -0.78 -0.71 -0.69 -0.6 -0.69 -0.81 -0.62 -
0.62 ...

```

```

## ..$ CYP27b1 : num [1:40] -1.31 -1.14 -1.29 -1.27 -1.2 -1.15 -1.17 -1.28 -1.13 -
1.15 ...
## ..$ CYP2b10 : num [1:40] -1.23 -1.2 -1.32 -1.23 -1.22 -1.1 -1.07 -1.26 -1.19 -
1.1 ...
## ..$ CYP2b13 : num [1:40] -1.19 -1.06 -1.25 -1.13 -1.1 -1.07 -1.2 -1.37 -1.15 -
1.11 ...
## ..$ CYP2c29 : num [1:40] -0.06 -0.2 -0.3 -0.07 -0.29 -0.28 -0.1 -0.1 0.18 -0.33
...
## ..$ CYP3A11 : num [1:40] -0.09 -0.34 -0.45 -0.11 -0.51 -0.55 -0.18 -0.25 0.06 -
0.4 ...
## ..$ CYP4A10 : num [1:40] -0.81 -0.88 -0.71 -0.65 -1.16 -0.99 -0.62 -0.82 -0.48
-0.79 ...
## ..$ CYP4A14 : num [1:40] -0.81 -0.84 -0.98 -0.41 -1.16 -1.09 -0.76 -0.87 -0.37
-0.95 ...
## ..$ CYP7a : num [1:40] -0.77 -0.71 -0.93 -0.8 -0.71 -0.74 -0.76 -0.88 -0.77 -
0.77 ...
## ..$ CYP8b1 : num [1:40] -0.77 -0.63 -0.53 -0.73 -0.51 -0.55 -0.57 -0.63 -0.6 -
0.66 ...
## ..$ FAS : num [1:40] -0.41 -0.37 -0.3 -0.59 -0.06 0.18 -0.16 0.04 -0.53 0.0
8 ...
## ..$ FAT : num [1:40] -1.03 -0.98 -1.03 -1.06 -0.99 -0.99 -0.89 -1.08 -1.04
-0.91 ...
## ..$ FDFT : num [1:40] -0.98 -0.92 -1.04 -1 -0.99 -1 -1.02 -0.97 -1.03 -0.95
...
## ..$ FXR : num [1:40] -0.93 -0.87 -1 -0.9 -0.89 -0.89 -0.86 -1.01 -0.81 -0.9
1 ...
## ..$ G6PDH : num [1:40] -1.22 -1.09 -1.28 -1.19 -1.16 -0.96 -1.15 -1.26 -1.13
-1.03 ...
## ..$ G6Pase : num [1:40] -0.46 -0.63 -1.06 -0.71 -0.58 -0.49 -0.51 -0.61 -0.38
-0.6 ...
## ..$ GK : num [1:40] -0.71 -0.67 -0.68 -0.75 -0.62 -0.59 -0.59 -0.66 -0.68
-0.47 ...
## ..$ GS : num [1:40] -1.24 -1.22 -1.36 -1.21 -1.22 -1.16 -1.15 -1.31 -1.16
-1.19 ...
## ..$ GSTa : num [1:40] 0 -0.05 -0.13 -0.09 -0.02 -0.11 -0.06 -0.04 0.03 -0.02
...
## ..$ GSTmu : num [1:40] 0.02 -0.05 -0.19 0.03 -0.23 -0.05 -0.22 -0.07 0.23 -0.
14 ...
## ..$ GSTpi2 : num [1:40] 0.45 0.3 0.18 0.36 0.3 0.17 0.12 0.48 0.53 0.01 ...
## ..$ HMGCoAred : num [1:40] -0.95 -0.86 -0.96 -1.02 -0.7 -0.76 -1 -0.88 -0.96 -0.7
...
## ..$ HPNCL : num [1:40] -0.65 -0.69 -0.75 -0.61 -0.66 -0.56 -0.61 -0.71 -0.53
-0.6 ...
## ..$ IL.2 : num [1:40] -0.94 -0.94 -1.16 -0.97 -0.93 -0.96 -0.96 -0.85 -0.84
-0.95 ...
## ..$ L.FABP : num [1:40] 0.24 0.27 0.17 0.16 0 0.23 0.18 0.18 0.2 0.2 ...
## ..$ LCE : num [1:40] 0.09 0.06 -0.05 0.01 -0.07 -0.1 -0.03 -0.08 0.12 -0.1
...
## ..$ LDLr : num [1:40] -0.82 -0.68 -0.82 -0.94 -0.73 -0.74 -0.8 -0.83 -0.81 -
0.72 ...
## ..$ LPK : num [1:40] -0.32 -0.39 -0.38 -0.38 -0.17 -0.14 -0.35 -0.13 -0.32
-0.24 ...
## ..$ LPL : num [1:40] -1.01 -0.97 -1.11 -0.99 -1.05 -0.99 -0.93 -1.07 -0.94
-0.95 ...
## ..$ LXRA : num [1:40] -0.82 -0.82 -0.91 -0.85 -0.83 -0.79 -0.77 -0.84 -0.75
-0.78 ...

```

```

## ..$ LXRb      : num [1:40] -1 -0.95 -1.16 -1.01 -1.01 -0.99 -0.98 -1.04 -0.98 -0.
99 ...
## ..$ Lpin      : num [1:40] -0.87 -0.97 -0.95 -1 -0.57 -0.51 -0.81 -0.83 -0.83 -0.
48 ...
## ..$ Lpin1     : num [1:40] -0.85 -0.99 -0.94 -1.02 -0.53 -0.51 -0.81 -0.87 -0.82
-0.49 ...
## ..$ Lpin2     : num [1:40] -0.85 -0.87 -0.9 -0.88 -0.72 -0.68 -0.8 -0.9 -0.68 -0.
67 ...
## ..$ Lpin3     : num [1:40] -1.23 -1.12 -1.25 -1.18 -1.12 -1.09 -1.04 -1.23 -1.13
-1.11 ...
## ..$ M.CPT1    : num [1:40] -1.15 -1.06 -1.26 -1.1 -1.11 -1.14 -1.08 -1.19 -1.06 -
1.09 ...
## ..$ MCAD      : num [1:40] -0.6 -0.62 -0.7 -0.59 -0.69 -0.66 -0.53 -0.66 -0.45 -
0.62 ...
## ..$ MDR1      : num [1:40] -1.15 -1.1 -1.26 -1.13 -1.11 -1.09 -1.09 -1.19 -1.06 -
1.1 ...
## ..$ MDR2      : num [1:40] -0.77 -0.65 -0.86 -0.77 -0.7 -0.69 -0.81 -0.81 -0.69 -
0.75 ...
## ..$ MRP6      : num [1:40] -0.99 -0.85 -0.9 -0.95 -0.91 -0.84 -0.88 -1.02 -0.83 -
0.86 ...
## ..$ MS        : num [1:40] -1.11 -1.06 -1.2 -1.09 -1.09 -1.09 -0.99 -1.16 -1.06 -
0.98 ...
## ..$ MTHFR     : num [1:40] -0.96 -0.99 -1.1 -0.95 -0.93 -0.96 -0.88 -1.03 -1.01 -
0.95 ...
## ..$ NGFiB     : num [1:40] -1.21 -1.08 -1.24 -1.12 -1.11 -1.04 -1.02 -1.21 -1.11
-1.04 ...
## ..$ NURR1     : num [1:40] -1.21 -1.1 -1.32 -1.11 -1.14 -1.18 -1.1 -1.26 -1.14 -
1.09 ...
## ..$ Ntcp      : num [1:40] -0.49 -0.45 -0.44 -0.54 -0.47 -0.46 -0.55 -0.5 -0.44 -
0.43 ...
## ..$ OCTN2     : num [1:40] -1.15 -1.15 -1.2 -1.17 -1.19 -1.11 -1.08 -1.21 -1.05 -
1.08 ...
## ..$ PAL       : num [1:40] -1.32 -1.25 -1.16 -1.25 -1.24 -1.02 -1.04 -1.27 -0.93
-0.92 ...
## ..$ PDK4      : num [1:40] -1.16 -1.16 -1.27 -1.16 -1.13 -1.08 -1.14 -1.24 -1.19
-1.04 ...
## ..$ PECCI     : num [1:40] -0.68 -0.69 -0.92 -0.71 -0.83 -0.81 -0.79 -0.85 -0.58
-0.82 ...
## ..$ PLTP      : num [1:40] -1.1 -0.99 -1.03 -1.08 -0.98 -0.89 -1.05 -1.07 -1.02 -
0.85 ...
## ..$ PMDCI     : num [1:40] -0.52 -0.52 -0.6 -0.52 -0.71 -0.69 -0.55 -0.57 -0.46 -
0.69 ...
## ..$ PON       : num [1:40] -0.52 -0.55 -0.65 -0.64 -0.57 -0.63 -0.56 -0.65 -0.6 -
0.64 ...
## ..$ PPARa     : num [1:40] -0.93 -0.86 -0.95 -0.97 -0.94 -0.95 -0.9 -1.12 -0.88 -
0.95 ...
## ..$ PPARd     : num [1:40] -1.51 -1.59 -1.71 -1.57 -1.53 -1.56 -1.49 -1.57 -1.58
-1.54 ...
## ..$ PPARg     : num [1:40] -1.06 -1.02 -1.14 -1.05 -1.09 -1.01 -1 -1.13 -0.97 -1.
07 ...
## ..$ PXR       : num [1:40] -0.99 -0.96 -1.1 -0.99 -1 -1.03 -0.93 -1.07 -0.98 -0.9
6 ...
## ..$ Pex11a    : num [1:40] -1 -1.02 -1.2 -1 -0.95 -1.07 -1.05 -1.02 -1 -1.01 ...
## ..$ RARa      : num [1:40] -1.2 -1.06 -1.16 -1.17 -1.15 -1.13 -1.09 -1.24 -1.03 -
1.09 ...
## ..$ RARb2     : num [1:40] -1.19 -1.11 -1.23 -1.16 -1.14 -1.07 -1.09 -1.18 -1.12

```

```

-1.1 ...
## ..$ RXRa      : num [1:40] -0.67 -0.59 -0.68 -0.72 -0.78 -0.62 -0.65 -0.76 -0.55
-0.67 ...
## ..$ RXRb2     : num [1:40] -0.95 -0.95 -1.07 -0.95 -0.98 -0.94 -0.92 -1.03 -0.94
-0.95 ...
## ..$ RXRg1     : num [1:40] -1.16 -1.1 -1.21 -1.1 -1.11 -1.03 -1.07 -1.19 -1.05 -
1.04 ...
## ..$ S14       : num [1:40] -0.93 -0.86 -0.84 -1.05 -0.65 -0.4 -0.73 -0.62 -0.99 -
0.25 ...
## ..$ SHP1      : num [1:40] -1.1 -0.97 -1.09 -1.03 -1.13 -0.98 -0.95 -1.21 -0.93 -
0.97 ...
## ..$ SIAT4c    : num [1:40] -1.07 -0.97 -1.04 -0.99 -0.94 -0.93 -0.89 -1.04 -0.93
-0.95 ...
## ..$ SPI1.1    : num [1:40] 1.19 1.15 1.09 1.07 1.22 1.05 1.15 1.18 1.21 1.04 ...
## ..$ SR.BI     : num [1:40] -0.84 -0.86 -0.95 -0.95 -1.06 -0.8 -0.83 -1 -0.83 -0.7
7 ...
## ..$ THB       : num [1:40] -0.79 -0.85 -0.92 -0.79 -0.84 -0.86 -0.8 -0.86 -0.83 -
0.85 ...
## ..$ THIOL     : num [1:40] -0.18 -0.15 -0.24 -0.15 -0.35 -0.29 -0.22 -0.23 -0.17
-0.18 ...
## ..$ TRa       : num [1:40] -1.48 -1.46 -1.58 -1.54 -1.46 -1.44 -1.32 -1.56 -1.46
-1.35 ...
## ..$ TRb       : num [1:40] -1.07 -1 -1.16 -1.11 -1.01 -1 -0.97 -1.08 -1.02 -0.98
...
## ..$ Tpalpha   : num [1:40] -0.69 -0.74 -0.81 -0.74 -0.82 -0.76 -0.72 -0.76 -0.65
-0.83 ...
## ..$ Tpbeta    : num [1:40] -1.11 -1.09 -1.14 -1.04 -1.2 -1.05 -1 -1.16 -0.91 -1.0
7 ...
## .. [list output truncated]
## $ lipid      : 'data.frame': 40 obs. of 21 variables:
## ..$ C14.0     : num [1:40] 0.34 0.38 0.36 0.22 0.37 1.7 0.35 0.34 0.22 1.38 ...
## ..$ C16.0     : num [1:40] 26.4 24 23.7 25.5 24.8 ...
## ..$ C18.0     : num [1:40] 10.22 9.93 8.96 8.14 9.63 ...
## ..$ C16.1n.9 : num [1:40] 0.35 0.55 0.55 0.49 0.46 0.66 0.36 0.29 0.44 0.9 ...
## ..$ C16.1n.7 : num [1:40] 3.1 2.54 2.65 2.82 2.85 7.26 3.6 3.27 2.36 7.01 ...
## ..$ C18.1n.9 : num [1:40] 17 20.1 22.9 21.9 21.4 ...
## ..$ C18.1n.7 : num [1:40] 2.41 3.92 3.96 2.52 2.96 8.99 2.15 1.99 1.81 8.85 ...
## ..$ C20.1n.9 : num [1:40] 0.26 0.23 0.26 0 0.3 0.36 0.25 0.31 0 0.21 ...
## ..$ C20.3n.9 : num [1:40] 0 0 0.19 0 0.27 2.89 0 0 0 2.03 ...
## ..$ C18.2n.6 : num [1:40] 8.93 14.98 16.06 13.89 14.55 ...
## ..$ C18.3n.6 : num [1:40] 0 0.3 0.27 0 0.27 2.66 0 0 0 0 ...
## ..$ C20.2n.6 : num [1:40] 0 0.3 0.33 0 0.23 0 0 0 0 0 ...
## ..$ C20.3n.6 : num [1:40] 0.78 1.64 1.51 1.1 1.58 0.81 0.68 0.72 1.07 0.59 ...
## ..$ C20.4n.6 : num [1:40] 3.07 15.34 13.27 3.92 11.85 ...
## ..$ C22.4n.6 : num [1:40] 0 0.58 0.54 0 0.32 0 0 0 0 0 ...
## ..$ C22.5n.6 : num [1:40] 0 2.1 1.77 0 0.44 0.56 0 0 0 0.39 ...
## ..$ C18.3n.3 : num [1:40] 5.97 0 0 0.49 0.42 0 8.4 6.01 0.55 0 ...
## ..$ C20.3n.3 : num [1:40] 0.37 0 0 0 0 0 0.42 0.39 0 0 ...
## ..$ C20.5n.3 : num [1:40] 8.62 0 0 2.99 0.3 0 7.37 7.96 3.13 0 ...
## ..$ C22.5n.3 : num [1:40] 1.75 0.48 0.22 1.04 0.35 2.13 2.05 2.33 1.65 0 ...
## ..$ C22.6n.3 : num [1:40] 10.39 2.61 2.51 14.99 6.69 ...
## $ diet       : Factor w/ 5 levels "coc","fish","lin",...: 3 5 5 2 4 1 3 3 2 1 ...
## $ genotype   : Factor w/ 2 levels "wt","ppar": 1 1 1 1 1 1 1 1 1 1 ...

```

```
## check dimensions
lapply(nutrimouse, dim) # apply function dim to each element in list nutrimouse
```

```
## $gene
## [1] 40 120
##
## $lipid
## [1] 40 21
##
## $diet
## NULL
##
## $genotype
## NULL
```

```
lapply(nutrimouse, length) # apply function length to each element in list nutrimouse
```

```
## $gene
## [1] 120
##
## $lipid
## [1] 21
##
## $diet
## [1] 40
##
## $genotype
## [1] 40
```

2. Take the gene expression dataset in *samples x variables* matrix format. Investigate their distribution.

```
## get gene expression data structure
str(nutrimouse$gene)
```

```
## 'data.frame':    40 obs. of  120 variables:
## $ X36b4      : num  -0.42 -0.44 -0.48 -0.45 -0.42 -0.43 -0.53 -0.49 -0.36 -0.5 ...
## $ ACAT1      : num  -0.65 -0.68 -0.74 -0.69 -0.71 -0.69 -0.62 -0.69 -0.66 -0.62 ...
## $ ACAT2      : num  -0.84 -0.91 -1.1 -0.65 -0.54 -0.8 -1 -0.91 -0.74 -0.79 ...
## $ ACPBP      : num  -0.34 -0.32 -0.46 -0.41 -0.38 -0.32 -0.44 -0.37 -0.39 -0.36 ...
## $ ACC1       : num  -1.29 -1.23 -1.3 -1.26 -1.21 -1.13 -1.22 -1.29 -1.15 -1.21 ...
## $ ACC2       : num  -1.13 -1.06 -1.09 -1.09 -0.89 -0.79 -1 -1.06 -1.08 -0.82 ...
## $ ACOTH      : num  -0.93 -0.99 -1.06 -0.93 -1 -0.93 -0.94 -1.05 -0.88 -0.92 ...
## $ ADISP      : num  -0.98 -0.97 -1.08 -1.02 -0.95 -0.97 -0.94 -1.02 -0.98 -0.99 ...
## $ ADSS1      : num  -1.19 -1 -1.18 -1.07 -1.08 -1.07 -1.05 -1.16 -1.05 -1 ...
## $ ALDH3      : num  -0.68 -0.62 -0.75 -0.71 -0.76 -0.75 -0.67 -0.75 -0.66 -0.69 ...
## $ AM2R       : num  -0.59 -0.58 -0.66 -0.65 -0.59 -0.55 -0.66 -0.66 -0.53 -0.62 ...
## $ AOX        : num  -0.16 -0.12 -0.16 -0.17 -0.31 -0.23 -0.09 -0.22 -0.06 -0.23 ...
## $ BACT       : num  -0.22 -0.32 -0.32 -0.32 -0.31 -0.29 -0.25 -0.21 -0.15 -0.2 ...
## $ BIEN       : num  -0.89 -0.88 -0.89 -0.77 -0.97 -0.84 -0.86 -0.9 -0.74 -0.76 ...
## $ BSEP       : num  -0.69 -0.6 -0.7 -0.67 -0.68 -0.55 -0.67 -0.66 -0.6 -0.58 ...
## $ Bcl.3      : num  -1.18 -1.07 -1.17 -1.12 -0.93 -1.08 -1.03 -1.01 -1.01 -1.1 ...
## $ C16SR      : num  1.66 1.65 1.57 1.61 1.66 1.7 1.58 1.62 1.72 1.55 ...
## $ CACP       : num  -0.92 -0.87 -1.02 -0.89 -0.93 -0.97 -0.97 -0.96 -0.85 -0.95 ...
## $ CAR1       : num  -0.97 -0.92 -0.98 -0.97 -1.06 -1.03 -0.91 -1.11 -0.85 -0.99 ...
## $ CBS        : num  -0.26 -0.36 -0.4 -0.39 -0.35 -0.31 -0.32 -0.4 -0.26 -0.39 ...
## $ CIDEA      : num  -1.21 -1.17 -1.29 -1.18 -1.15 -1.14 -1.16 -1.26 -1.12 -1.08 ...
## $ COX1       : num  -1.11 -1.06 -1.17 -1.03 -0.99 -1.03 -1.15 -1.18 -0.94 -1.07 ...
## $ COX2       : num  -1.18 -1.06 -1.14 -1.13 -1.1 -1.16 -1.06 -1.24 -1.23 -1.09 ...
## $ CPT2       : num  -0.87 -0.87 -0.95 -0.88 -0.91 -0.92 -0.86 -0.93 -0.82 -0.88 ...
## $ CYP24      : num  -1.37 -1.14 -1.3 -1.27 -1.2 -1.11 -1.12 -1.3 -1.14 -1.08 ...
## $ CYP26      : num  -1.21 -1.12 -1.22 -1.18 -1.16 -1.1 -1.07 -1.23 -1.1 -1.1 ...
## $ CYP27a1    : num  -0.71 -0.62 -0.78 -0.71 -0.69 -0.6 -0.69 -0.81 -0.62 -0.62 ...
## $ CYP27b1    : num  -1.31 -1.14 -1.29 -1.27 -1.2 -1.15 -1.17 -1.28 -1.13 -1.15 ...
## $ CYP2b10    : num  -1.23 -1.2 -1.32 -1.23 -1.22 -1.1 -1.07 -1.26 -1.19 -1.1 ...
## $ CYP2b13    : num  -1.19 -1.06 -1.25 -1.13 -1.1 -1.07 -1.2 -1.37 -1.15 -1.11 ...
## $ CYP2c29    : num  -0.06 -0.2 -0.3 -0.07 -0.29 -0.28 -0.1 -0.1 0.18 -0.33 ...
## $ CYP3A11    : num  -0.09 -0.34 -0.45 -0.11 -0.51 -0.55 -0.18 -0.25 0.06 -0.4 ...
## $ CYP4A10    : num  -0.81 -0.88 -0.71 -0.65 -1.16 -0.99 -0.62 -0.82 -0.48 -0.79 ...
## $ CYP4A14    : num  -0.81 -0.84 -0.98 -0.41 -1.16 -1.09 -0.76 -0.87 -0.37 -0.95 ...
## $ CYP7a      : num  -0.77 -0.71 -0.93 -0.8 -0.71 -0.74 -0.76 -0.88 -0.77 -0.77 ...
## $ CYP8b1     : num  -0.77 -0.63 -0.53 -0.73 -0.51 -0.55 -0.57 -0.63 -0.6 -0.66 ...
## $ FAS        : num  -0.41 -0.37 -0.3 -0.59 -0.06 0.18 -0.16 0.04 -0.53 0.08 ...
## $ FAT        : num  -1.03 -0.98 -1.03 -1.06 -0.99 -0.99 -0.89 -1.08 -1.04 -0.91 ...
## $ FDFT       : num  -0.98 -0.92 -1.04 -1 -0.99 -1 -1.02 -0.97 -1.03 -0.95 ...
## $ FXR        : num  -0.93 -0.87 -1 -0.9 -0.89 -0.89 -0.86 -1.01 -0.81 -0.91 ...
## $ G6PDH      : num  -1.22 -1.09 -1.28 -1.19 -1.16 -0.96 -1.15 -1.26 -1.13 -1.03 ...
## $ G6Pase     : num  -0.46 -0.63 -1.06 -0.71 -0.58 -0.49 -0.51 -0.61 -0.38 -0.6 ...
## $ GK         : num  -0.71 -0.67 -0.68 -0.75 -0.62 -0.59 -0.59 -0.66 -0.68 -0.47 ...
## $ GS         : num  -1.24 -1.22 -1.36 -1.21 -1.22 -1.16 -1.15 -1.31 -1.16 -1.19 ...
## $ GSTa       : num  0 -0.05 -0.13 -0.09 -0.02 -0.11 -0.06 -0.04 0.03 -0.02 ...
## $ GSTmu      : num  0.02 -0.05 -0.19 0.03 -0.23 -0.05 -0.22 -0.07 0.23 -0.14 ...
## $ GSTpi2     : num  0.45 0.3 0.18 0.36 0.3 0.17 0.12 0.48 0.53 0.01 ...
## $ HMGCoAred : num  -0.95 -0.86 -0.96 -1.02 -0.7 -0.76 -1 -0.88 -0.96 -0.7 ...
## $ HPNCL      : num  -0.65 -0.69 -0.75 -0.61 -0.66 -0.56 -0.61 -0.71 -0.53 -0.6 ...
## $ IL.2       : num  -0.94 -0.94 -1.16 -0.97 -0.93 -0.96 -0.96 -0.85 -0.84 -0.95 ...
## $ L.FABP     : num  0.24 0.27 0.17 0.16 0 0.23 0.18 0.18 0.2 0.2 ...
## $ LCE        : num  0.09 0.06 -0.05 0.01 -0.07 -0.1 -0.03 -0.08 0.12 -0.1 ...
## $ LDLr       : num  -0.82 -0.68 -0.82 -0.94 -0.73 -0.74 -0.8 -0.83 -0.81 -0.72 ...
## $ LPK        : num  -0.32 -0.39 -0.38 -0.38 -0.17 -0.14 -0.35 -0.13 -0.32 -0.24 ...
```

```

## $ LPL : num -1.01 -0.97 -1.11 -0.99 -1.05 -0.99 -0.93 -1.07 -0.94 -0.95 ...
## $ LXRa : num -0.82 -0.82 -0.91 -0.85 -0.83 -0.79 -0.77 -0.84 -0.75 -0.78 ...
## $ LXRB : num -1 -0.95 -1.16 -1.01 -1.01 -0.99 -0.98 -1.04 -0.98 -0.99 ...
## $ Lpin : num -0.87 -0.97 -0.95 -1 -0.57 -0.51 -0.81 -0.83 -0.83 -0.48 ...
## $ Lpin1 : num -0.85 -0.99 -0.94 -1.02 -0.53 -0.51 -0.81 -0.87 -0.82 -0.49 ...
## $ Lpin2 : num -0.85 -0.87 -0.9 -0.88 -0.72 -0.68 -0.8 -0.9 -0.68 -0.67 ...
## $ Lpin3 : num -1.23 -1.12 -1.25 -1.18 -1.12 -1.09 -1.04 -1.23 -1.13 -1.11 ...
## $ M.CPT1 : num -1.15 -1.06 -1.26 -1.1 -1.11 -1.14 -1.08 -1.19 -1.06 -1.09 ...
## $ MCAD : num -0.6 -0.62 -0.7 -0.59 -0.69 -0.66 -0.53 -0.66 -0.45 -0.62 ...
## $ MDR1 : num -1.15 -1.1 -1.26 -1.13 -1.11 -1.09 -1.09 -1.19 -1.06 -1.1 ...
## $ MDR2 : num -0.77 -0.65 -0.86 -0.77 -0.7 -0.69 -0.81 -0.81 -0.69 -0.75 ...
## $ MRP6 : num -0.99 -0.85 -0.9 -0.95 -0.91 -0.84 -0.88 -1.02 -0.83 -0.86 ...
## $ MS : num -1.11 -1.06 -1.2 -1.09 -1.09 -1.09 -0.99 -1.16 -1.06 -0.98 ...
## $ MTHFR : num -0.96 -0.99 -1.1 -0.95 -0.93 -0.96 -0.88 -1.03 -1.01 -0.95 ...
## $ NGFiB : num -1.21 -1.08 -1.24 -1.12 -1.11 -1.04 -1.02 -1.21 -1.11 -1.04 ...
## $ NURR1 : num -1.21 -1.1 -1.32 -1.11 -1.14 -1.18 -1.1 -1.26 -1.14 -1.09 ...
## $ Ntcp : num -0.49 -0.45 -0.44 -0.54 -0.47 -0.46 -0.55 -0.5 -0.44 -0.43 ...
## $ OCTN2 : num -1.15 -1.15 -1.2 -1.17 -1.19 -1.11 -1.08 -1.21 -1.05 -1.08 ...
## $ PAL : num -1.32 -1.25 -1.16 -1.25 -1.24 -1.02 -1.04 -1.27 -0.93 -0.92 ...
## $ PDK4 : num -1.16 -1.16 -1.27 -1.16 -1.13 -1.08 -1.14 -1.24 -1.19 -1.04 ...
## $ PECI : num -0.68 -0.69 -0.92 -0.71 -0.83 -0.81 -0.79 -0.85 -0.58 -0.82 ...
## $ PLTP : num -1.1 -0.99 -1.03 -1.08 -0.98 -0.89 -1.05 -1.07 -1.02 -0.85 ...
## $ PMDCI : num -0.52 -0.52 -0.6 -0.52 -0.71 -0.69 -0.55 -0.57 -0.46 -0.69 ...
## $ PON : num -0.52 -0.55 -0.65 -0.64 -0.57 -0.63 -0.56 -0.65 -0.6 -0.64 ...
## $ PPARa : num -0.93 -0.86 -0.95 -0.97 -0.94 -0.95 -0.9 -1.12 -0.88 -0.95 ...
## $ PPARd : num -1.51 -1.59 -1.71 -1.57 -1.53 -1.56 -1.49 -1.57 -1.58 -1.54 ...
## $ PPARg : num -1.06 -1.02 -1.14 -1.05 -1.09 -1.01 -1 -1.13 -0.97 -1.07 ...
## $ PXR : num -0.99 -0.96 -1.1 -0.99 -1 -1.03 -0.93 -1.07 -0.98 -0.96 ...
## $ Pex11a : num -1 -1.02 -1.2 -1 -0.95 -1.07 -1.05 -1.02 -1 -1.01 ...
## $ RARa : num -1.2 -1.06 -1.16 -1.17 -1.15 -1.13 -1.09 -1.24 -1.03 -1.09 ...
## $ RARb2 : num -1.19 -1.11 -1.23 -1.16 -1.14 -1.07 -1.09 -1.18 -1.12 -1.1 ...
## $ RXRa : num -0.67 -0.59 -0.68 -0.72 -0.78 -0.62 -0.65 -0.76 -0.55 -0.67 ...
## $ RXRb2 : num -0.95 -0.95 -1.07 -0.95 -0.98 -0.94 -0.92 -1.03 -0.94 -0.95 ...
## $ RXRg1 : num -1.16 -1.1 -1.21 -1.1 -1.11 -1.03 -1.07 -1.19 -1.05 -1.04 ...
## $ S14 : num -0.93 -0.86 -0.84 -1.05 -0.65 -0.4 -0.73 -0.62 -0.99 -0.25 ...
## $ SHP1 : num -1.1 -0.97 -1.09 -1.03 -1.13 -0.98 -0.95 -1.21 -0.93 -0.97 ...
## $ SIAT4c : num -1.07 -0.97 -1.04 -0.99 -0.94 -0.93 -0.89 -1.04 -0.93 -0.95 ...
## $ SPI1.1 : num 1.19 1.15 1.09 1.07 1.22 1.05 1.15 1.18 1.21 1.04 ...
## $ SR.BI : num -0.84 -0.86 -0.95 -0.95 -1.06 -0.8 -0.83 -1 -0.83 -0.77 ...
## $ THB : num -0.79 -0.85 -0.92 -0.79 -0.84 -0.86 -0.8 -0.86 -0.83 -0.85 ...
## $ THIOL : num -0.18 -0.15 -0.24 -0.15 -0.35 -0.29 -0.22 -0.23 -0.17 -0.18 ...
## $ TRa : num -1.48 -1.46 -1.58 -1.54 -1.46 -1.44 -1.32 -1.56 -1.46 -1.35 ...
## $ TRb : num -1.07 -1 -1.16 -1.11 -1.01 -1 -0.97 -1.08 -1.02 -0.98 ...
## $ Tpalpha : num -0.69 -0.74 -0.81 -0.74 -0.82 -0.76 -0.72 -0.76 -0.65 -0.83 ...
## $ Tpbeta : num -1.11 -1.09 -1.14 -1.04 -1.2 -1.05 -1 -1.16 -0.91 -1.07 ...
## [list output truncated]

```

```

## check if there are missing values
any(is.na(nutrimouse$gene))

```

```
## [1] FALSE
```



```
## investigate each variable
summary(nutrimouse$gene[, 1])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.5800 -0.5025 -0.4600 -0.4552 -0.4200 -0.3000
```

```
colors <- rainbow(20, alpha=1)
plot(density(scale(nutrimouse$gene[, 1], center=T, scale=F)),
      col=colors[1], xlim=c(-0.5,0.5), ylim=c(0,8))
sapply(2:20, function(i) {
  lines(density(scale(nutrimouse$gene[, i], center=T, scale=F)), col=colors[i])
})
```

```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
##
## [[6]]
## NULL
##
## [[7]]
## NULL
##
## [[8]]
## NULL
##
## [[9]]
## NULL
##
## [[10]]
## NULL
##
## [[11]]
## NULL
##
## [[12]]
## NULL
##
## [[13]]
## NULL
##
## [[14]]
## NULL
##
## [[15]]
## NULL
##
## [[16]]
## NULL
##
## [[17]]
## NULL
##
## [[18]]
## NULL
##
```

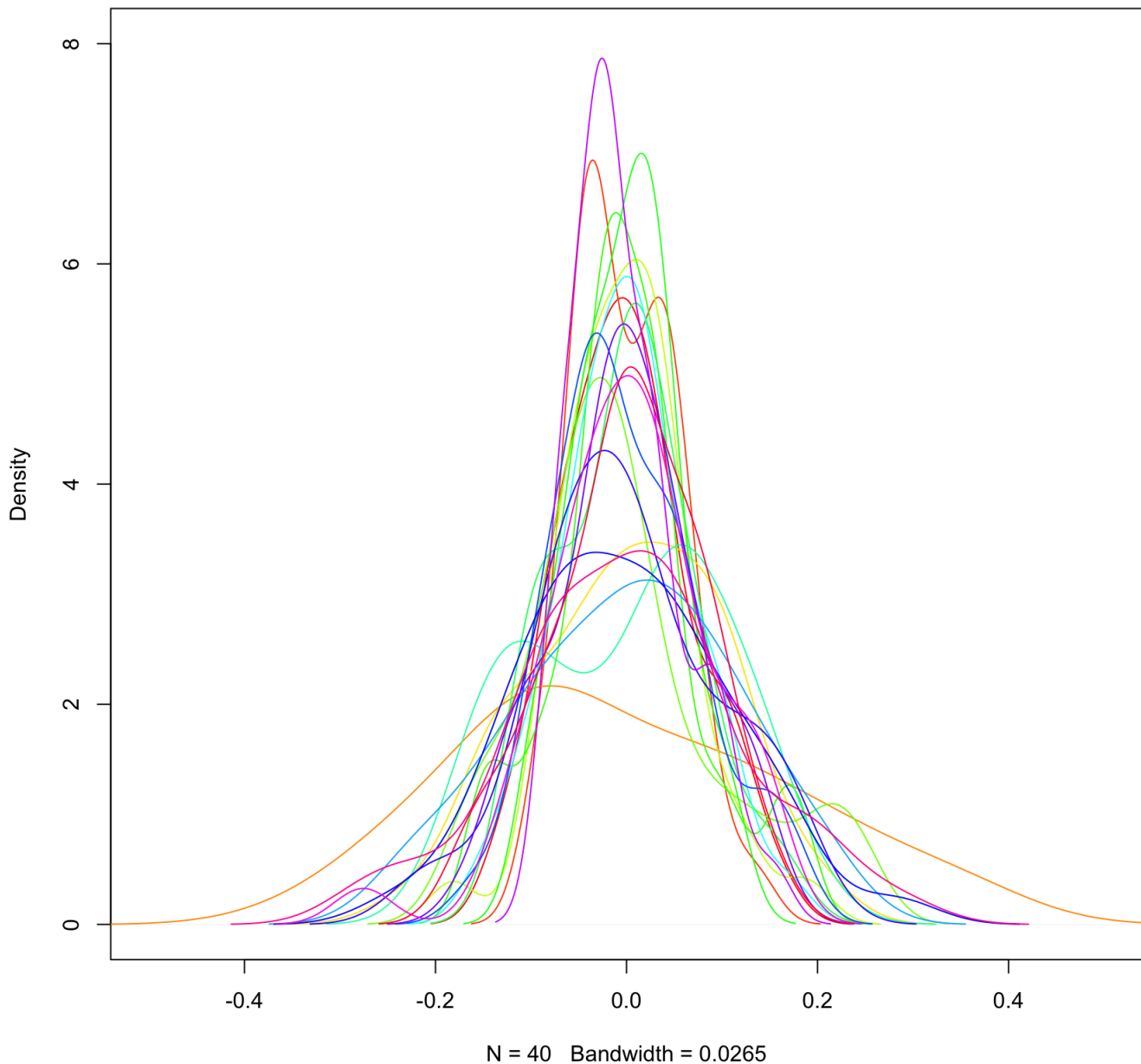
```
## [[19]]  
## NULL
```

```
apply(nutrimouse$gene, 2, summary)
```

##	X36b4	ACAT1	ACAT2	ACBP	ACC1	ACC2	ACOTH	ADISP
## Min.	-0.58000	-0.75000	-1.10000	-0.66000	-1.4400	-1.2000	-1.06000	-1.08000
## 1st Qu.	-0.50250	-0.69000	-0.88000	-0.50250	-1.3000	-1.0900	-0.95000	-1.02000
## Median	-0.46000	-0.66000	-0.79500	-0.42500	-1.2600	-1.0450	-0.92000	-0.97000
## Mean	-0.45525	-0.65525	-0.76675	-0.43375	-1.2585	-1.0280	-0.91075	-0.97825
## 3rd Qu.	-0.42000	-0.62000	-0.64500	-0.35500	-1.2200	-0.9875	-0.88000	-0.94000
## Max.	-0.30000	-0.52000	-0.39000	-0.24000	-1.0700	-0.7900	-0.73000	-0.87000
##	ADSS1	ALDH3	AM2R	AOX	BACT	BIEN	BSEP	Bcl.3
## Min.	-1.19000	-0.9900	-0.780	-0.4800	-0.44000	-1.16000	-0.9000	-1.22000
## 1st Qu.	-1.14000	-0.9100	-0.670	-0.3175	-0.32250	-0.99000	-0.7600	-1.10250
## Median	-1.07500	-0.7850	-0.630	-0.2300	-0.30000	-0.92000	-0.7000	-1.06500
## Mean	-1.07575	-0.8100	-0.628	-0.2505	-0.28275	-0.92125	-0.6910	-1.05875
## 3rd Qu.	-1.03500	-0.7475	-0.590	-0.1675	-0.23500	-0.85500	-0.6275	-1.01000
## Max.	-0.91000	-0.6200	-0.460	-0.0400	-0.11000	-0.64000	-0.5100	-0.91000
##	C16SR	CACP	CAR1	CBS	CIDEA	COX1	COX2	CPT2
## Min.	1.55000	-1.2600	-1.1900	-0.5600	-1.3300	-1.18000	-1.280	-1.2000
## 1st Qu.	1.59000	-1.0325	-0.9900	-0.4450	-1.2325	-1.09250	-1.180	-1.0100
## Median	1.61000	-0.9800	-0.9100	-0.4000	-1.1700	-1.05500	-1.130	-0.9450
## Mean	1.62675	-0.9845	-0.9135	-0.3995	-1.1840	-1.04975	-1.135	-0.9565
## 3rd Qu.	1.65250	-0.9375	-0.8475	-0.3375	-1.1400	-1.01000	-1.090	-0.8800
## Max.	1.78000	-0.8300	-0.6300	-0.2600	-1.0700	-0.88000	-1.040	-0.8200
##	CYP26	CYP27a1	CYP27b1	CYP2b10	CYP2b13	CYP2c29	CYP3A11	CYP4A10
## Min.	-1.3200	-0.88000	-1.350	-1.32000	-1.37000	-0.52000	-1.02000	-1.33000
## 1st Qu.	-1.2225	-0.78500	-1.245	-1.23000	-1.19250	-0.28250	-0.71250	-1.15250
## Median	-1.1500	-0.73000	-1.180	-1.20000	-1.14000	-0.14000	-0.53000	-1.05000
## Mean	-1.1560	-0.72725	-1.200	-1.18475	-1.14575	-0.14725	-0.50825	-0.97975
## 3rd Qu.	-1.1000	-0.67000	-1.150	-1.15000	-1.09750	-0.03000	-0.38500	-0.81750
## Max.	-0.9600	-0.59000	-0.990	-1.04000	-0.96000	0.18000	0.06000	-0.48000
##	CYP4A14	CYP7a	CYP8b1	FAS	FAT	FDFT	FXR	G6PDH
## Min.	-1.2900	-0.9300	-1.01000	-1.05000	-1.0900	-1.17000	-1.0600	-1.30000
## 1st Qu.	-1.1500	-0.8000	-0.76000	-0.67000	-1.0400	-1.02000	-0.9525	-1.20250
## Median	-1.0800	-0.7700	-0.67000	-0.49000	-0.9950	-0.99000	-0.9000	-1.15000
## Mean	-0.9930	-0.7695	-0.68225	-0.45175	-0.9910	-0.98075	-0.9105	-1.15125
## 3rd Qu.	-0.8925	-0.7400	-0.59000	-0.22500	-0.9475	-0.93750	-0.8775	-1.10750
## Max.	-0.1500	-0.6100	-0.50000	0.18000	-0.7500	-0.81000	-0.7600	-0.96000
##	G6Pase	GK	GS	GSTa	GSTmu	GSTpi2	HMGCoAred	HPNCL
## Min.	-1.06000	-0.9600	-1.3800	-0.4300	-0.440	0.00000	-1.0700	-0.97000
## 1st Qu.	-0.82000	-0.8000	-1.3025	-0.1525	-0.200	0.12000	-0.9700	-0.75000
## Median	-0.69000	-0.7000	-1.2250	-0.0900	-0.140	0.21000	-0.9300	-0.69000
## Mean	-0.69825	-0.7145	-1.2325	-0.1030	-0.119	0.22975	-0.9135	-0.69375
## 3rd Qu.	-0.53500	-0.6200	-1.1675	-0.0350	-0.050	0.33250	-0.8750	-0.60750
## Max.	-0.38000	-0.4600	-1.1200	0.0400	0.230	0.55000	-0.7000	-0.53000
##	IL.2	L.FABP	LCE	LDLr	LPK	LPL	LXRa	LXRb
## Min.	-1.1600	-0.4600	-0.26000	-0.9600	-0.570	-1.11000	-0.9100	-1.1600
## 1st Qu.	-1.0025	-0.0750	-0.10000	-0.8525	-0.395	-1.03000	-0.8400	-1.0225
## Median	-0.9450	0.0600	-0.06000	-0.8200	-0.350	-0.99000	-0.8150	-0.9900
## Mean	-0.9505	0.0340	-0.05275	-0.8195	-0.344	-0.99075	-0.8115	-0.9960
## 3rd Qu.	-0.8975	0.1825	0.00000	-0.7675	-0.295	-0.95000	-0.7775	-0.9675
## Max.	-0.8200	0.2800	0.12000	-0.6800	-0.130	-0.86000	-0.6500	-0.8400
##	Lpin	Lpin1	Lpin2	Lpin3	M.CPT1	MCAD	MDR1	MDR2
## Min.	-1.13000	-1.10000	-1.140	-1.2900	-1.29000	-0.7300	-1.30000	-0.92000
## 1st Qu.	-0.85500	-0.87000	-0.910	-1.1975	-1.16500	-0.6600	-1.16250	-0.83000
## Median	-0.72500	-0.76000	-0.855	-1.1450	-1.12000	-0.6200	-1.12000	-0.78000
## Mean	-0.75325	-0.76475	-0.849	-1.1475	-1.12575	-0.6050	-1.13425	-0.77875
## 3rd Qu.	-0.61500	-0.64000	-0.775	-1.0975	-1.09000	-0.5575	-1.09000	-0.71750

## Max.	-0.48000	-0.49000	-0.670	-0.9800	-0.96000	-0.4200	-0.99000	-0.65000
##	MRP6	MS	MTHFR	NGFiB	NURR1	Ntcp	OCTN2	PAL
## Min.	-1.09000	-1.20000	-1.1000	-1.29000	-1.32000	-0.6500	-1.28000	-1.3200
## 1st Qu.	-1.00250	-1.11000	-1.0025	-1.20000	-1.21000	-0.4925	-1.19000	-1.2550
## Median	-0.95500	-1.06500	-0.9700	-1.12000	-1.14000	-0.4400	-1.15000	-1.2000
## Mean	-0.94775	-1.06075	-0.9720	-1.12925	-1.16125	-0.4370	-1.13925	-1.1445
## 3rd Qu.	-0.87750	-1.00750	-0.9300	-1.07750	-1.10750	-0.3675	-1.08000	-1.0075
## Max.	-0.83000	-0.88000	-0.8800	-0.91000	-0.95000	-0.2500	-1.04000	-0.8900
##	PDK4	PECI	PLTP	PMDCI	PON	PPARa	PPARd	PPARG
## Min.	-1.28000	-1.11000	-1.15000	-1.07000	-0.7100	-1.1400	-1.7100	-1.190
## 1st Qu.	-1.17250	-0.92250	-1.09250	-0.94250	-0.6325	-1.0225	-1.5900	-1.090
## Median	-1.13000	-0.84000	-1.05000	-0.76500	-0.5800	-0.9500	-1.5600	-1.055
## Mean	-1.13525	-0.84725	-1.03625	-0.76725	-0.5825	-0.9660	-1.5595	-1.052
## 3rd Qu.	-1.08000	-0.79750	-0.99750	-0.60000	-0.5375	-0.9000	-1.5100	-1.010
## Max.	-1.01000	-0.58000	-0.85000	-0.44000	-0.4500	-0.8300	-1.4300	-0.900
##	PXR	Pexlla	RARa	RARb2	RXRa	RXRb2	RXRg1	S14
## Min.	-1.13000	-1.2000	-1.30000	-1.3000	-0.7800	-1.070	-1.2300	-1.05000
## 1st Qu.	-1.03000	-1.0500	-1.18250	-1.1900	-0.6725	-1.000	-1.1425	-0.98000
## Median	-0.99000	-1.0200	-1.13000	-1.1350	-0.6350	-0.960	-1.1000	-0.85500
## Mean	-0.99225	-1.0220	-1.13325	-1.1445	-0.6360	-0.964	-1.0955	-0.80675
## 3rd Qu.	-0.94750	-0.9875	-1.07500	-1.0900	-0.5875	-0.935	-1.0500	-0.65750
## Max.	-0.84000	-0.9000	-0.97000	-0.9900	-0.4900	-0.780	-0.9000	-0.25000
##	SHP1	SIAT4c	SPI1.1	SR.BI	THB	THIOL	TRa	TRb
## Min.	-1.21000	-1.16000	0.96000	-1.060	-0.9200	-0.900	-1.670	-1.22000
## 1st Qu.	-1.07500	-0.99000	1.03750	-0.920	-0.8500	-0.590	-1.510	-1.11000
## Median	-0.99000	-0.96000	1.07500	-0.830	-0.8200	-0.345	-1.460	-1.06000
## Mean	-1.00675	-0.96225	1.09075	-0.843	-0.8170	-0.411	-1.457	-1.05425
## 3rd Qu.	-0.94750	-0.92750	1.15000	-0.800	-0.7875	-0.230	-1.395	-0.99750
## Max.	-0.78000	-0.84000	1.23000	-0.610	-0.6900	-0.030	-1.220	-0.92000
##	Tpalpha	Tpbeta	UCP2	UCP3	VDR	VLDLr	Waf1	ap2
## Min.	-1.00000	-1.310	-1.0800	-1.27000	-1.30000	-1.19000	-1.3000	-1.3700
## 1st Qu.	-0.86000	-1.200	-1.0025	-1.15250	-1.18000	-1.09250	-1.1500	-1.2225
## Median	-0.83000	-1.140	-0.9800	-1.11000	-1.12000	-1.05500	-1.1300	-1.1900
## Mean	-0.81825	-1.130	-0.9660	-1.10775	-1.13175	-1.05325	-1.1235	-1.1880
## 3rd Qu.	-0.76000	-1.065	-0.9275	-1.05000	-1.08000	-1.01000	-1.0875	-1.1475
## Max.	-0.65000	-0.910	-0.7600	-0.92000	-0.94000	-0.91000	-0.9400	-1.0800
##	apoA.I	apoB	apoC3	apoE	c.fos	CHMGCoAS	cMOAT	eif2g
## Min.	0.5400	-0.2700	-0.49000	0.86000	-1.22000	-1.24000	-1.0200	-1.230
## 1st Qu.	0.6575	-0.2000	-0.39000	0.98000	-1.15000	-1.10250	-0.8950	-1.100
## Median	0.7200	-0.1700	-0.34000	1.04000	-1.11000	-1.03000	-0.8700	-1.055
## Mean	0.7295	-0.1675	-0.34075	1.02825	-1.10525	-1.01375	-0.8485	-1.058
## 3rd Qu.	0.8100	-0.1450	-0.30000	1.07000	-1.06000	-0.91000	-0.7875	-1.020
## Max.	0.9200	0.0100	-0.18000	1.18000	-0.98000	-0.78000	-0.6900	-0.840
##	hABC1	i.BABP	i.BAT	i.FABP	i.NOS	mABC1	mHMGCoAS	
## Min.	-1.25000	-0.8900	-1.89000	-1.300	-1.4300	-0.9800	-0.5800	
## 1st Qu.	-1.17250	-0.8325	-1.74250	-1.170	-1.2850	-0.9200	-0.3000	
## Median	-1.13500	-0.8000	-1.69000	-1.140	-1.2400	-0.8700	-0.2100	
## Mean	-1.13825	-0.7935	-1.69775	-1.122	-1.2460	-0.8765	-0.2210	
## 3rd Qu.	-1.09750	-0.7475	-1.66000	-1.075	-1.2075	-0.8375	-0.1275	
## Max.	-0.98000	-0.6700	-1.55000	-0.930	-1.0900	-0.8000	0.0600	

```
density.default(x = scale(nutrimouse$gene[, 1], center = T, scale = F))
```



PLS

1. Perform PLS (`mixOmics::pls`) and investigate the output, sample distribution and variable relationship with plots.

```
pls.res <- pls(X=nutrimouse$gene, Y=nutrimouse$lipid, ncomp=2, scale=TRUE, mode="canonical")  
max(abs(scale(nutrimouse$gene, center=T, scale=T) - pls.res$X))
```

```
## [1] 0
```

```
max(abs(scale(nutrimouse$lipid, center=T, scale=T) - pls.res$Y))
```

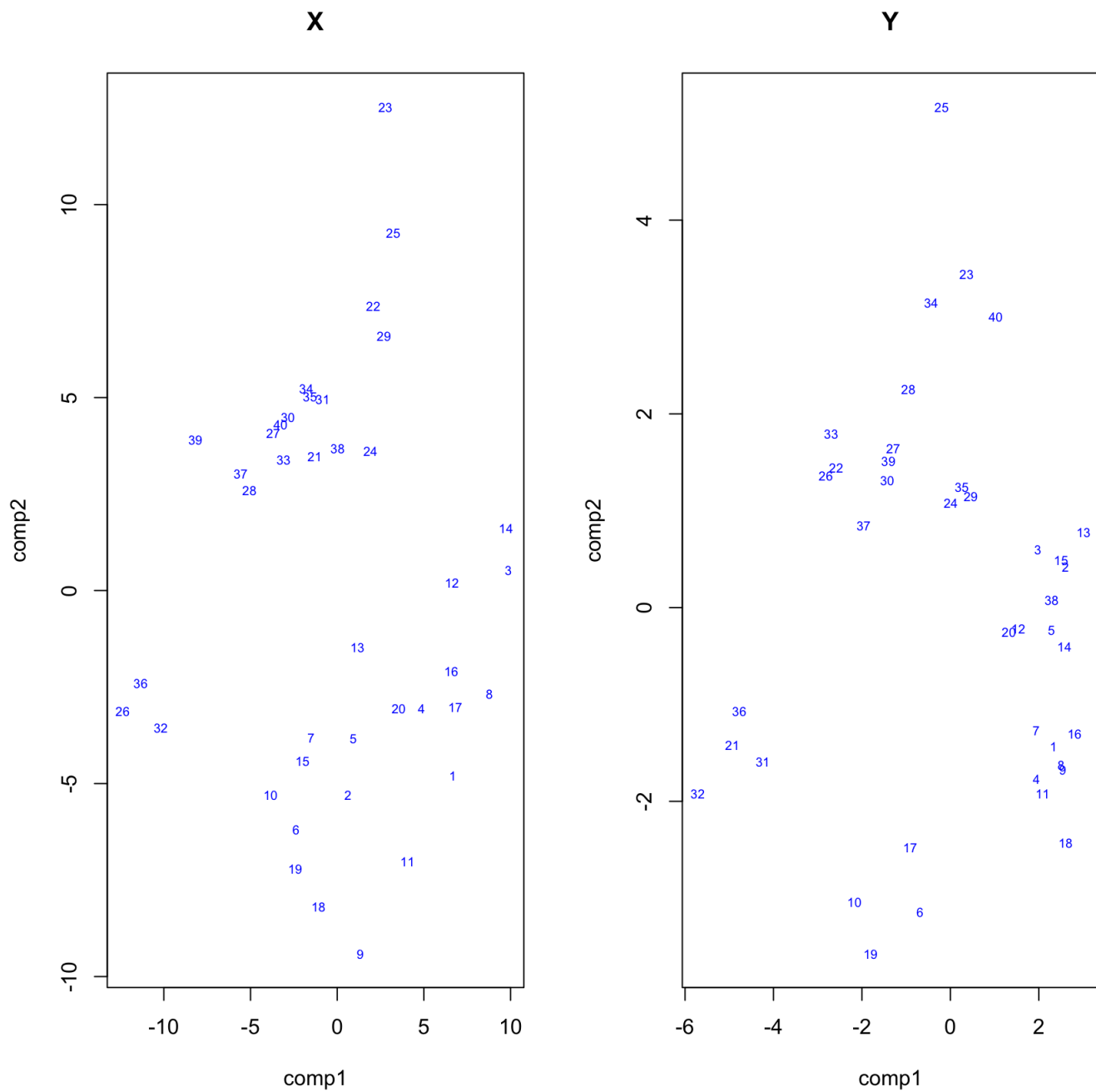
```
## [1] 0
```

The sample distribution plot can be performed with **variates**, sample coordinates in the new reference (rotated axes) for each of the two blocks.

```
str(pls.res$variates)
```

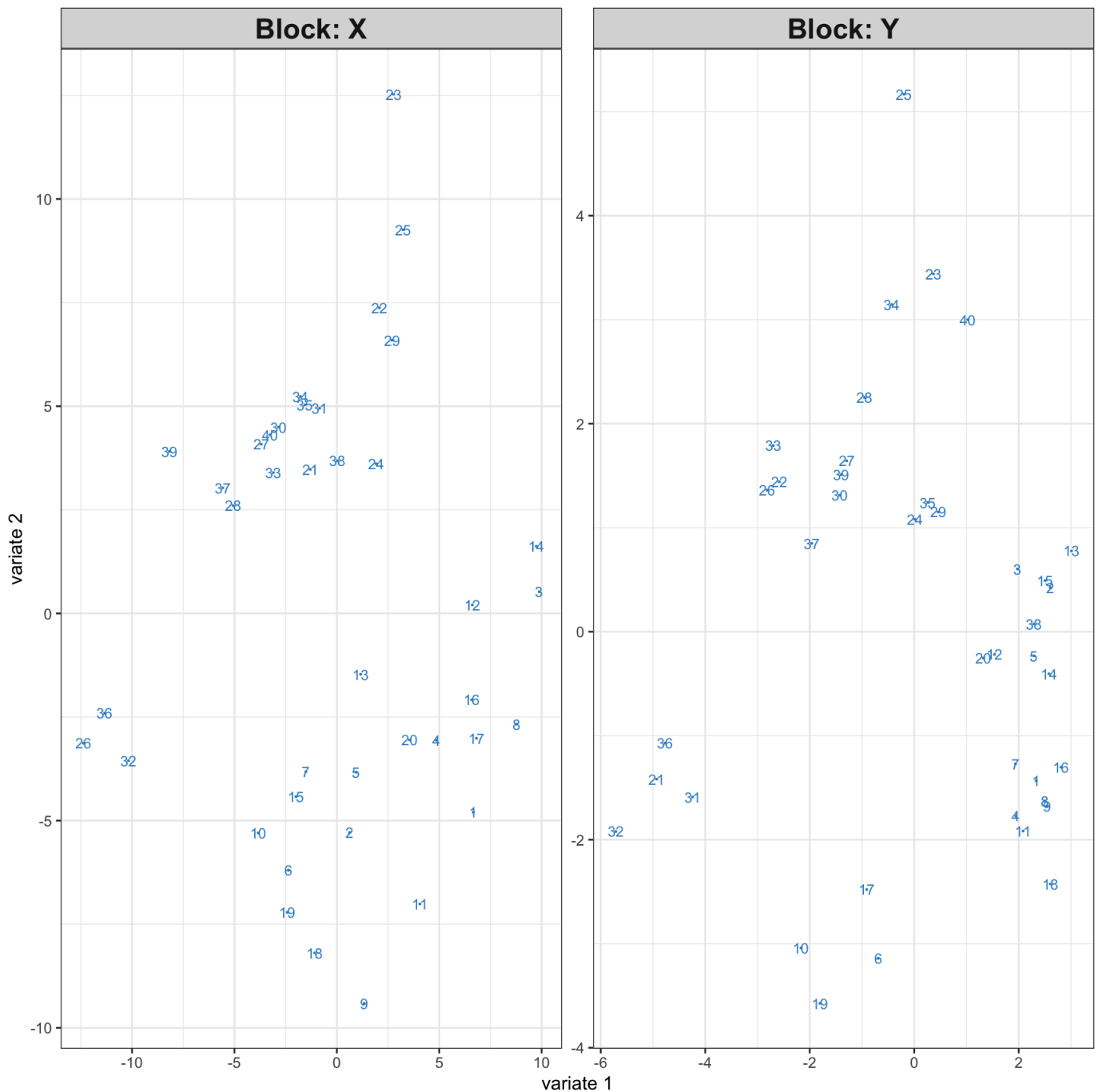
```
## List of 2
## $ X: num [1:40, 1:2] 6.659 0.614 9.876 4.864 0.934 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:40] "1" "2" "3" "4" ...
## .. ..$ : chr [1:2] "comp1" "comp2"
## $ Y: num [1:40, 1:2] 2.33 2.6 1.98 1.94 2.29 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:40] "1" "2" "3" "4" ...
## .. ..$ : chr [1:2] "comp1" "comp2"
```

```
PCx <- "comp1"
PCy <- "comp2"
par(mfrow=c(1,2))
plot(pls.res$variates$X[, PCx], pls.res$variates$X[, PCy], xlab=PCx, ylab=PCy, main="X", type='n')
text(pls.res$variates$X[, PCx], pls.res$variates$X[, PCy], rownames(pls.res$variates$X), col='blue', cex=0.6)
plot(pls.res$variates$Y[, PCx], pls.res$variates$Y[, PCy], xlab=PCx, ylab=PCy, main="Y", type='n')
text(pls.res$variates$Y[, PCx], pls.res$variates$Y[, PCy], rownames(pls.res$variates$Y), col='blue', cex=0.6)
```



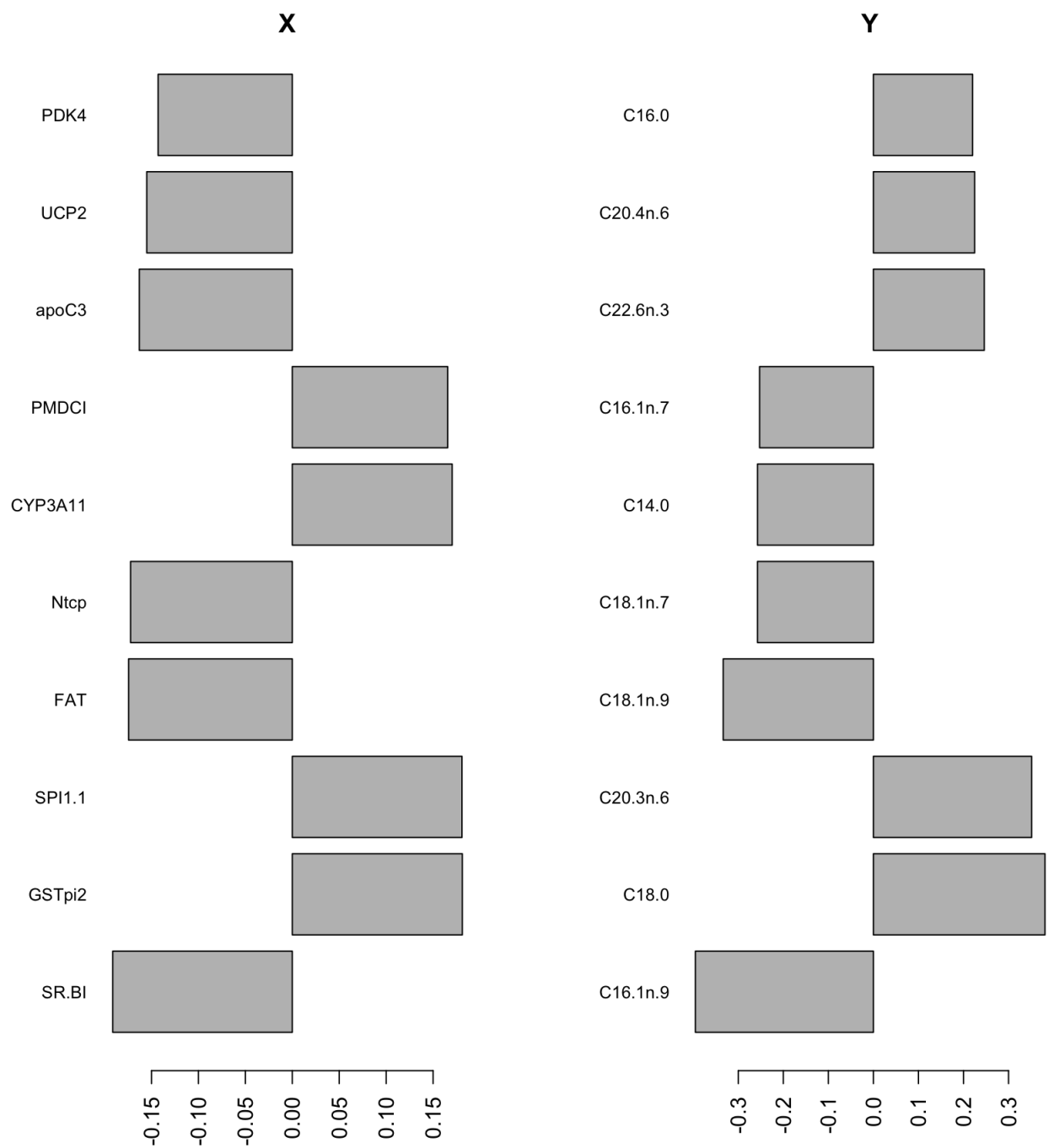
which is also produced with `plotIndiv`.

```
plotIndiv(pls.res)
```

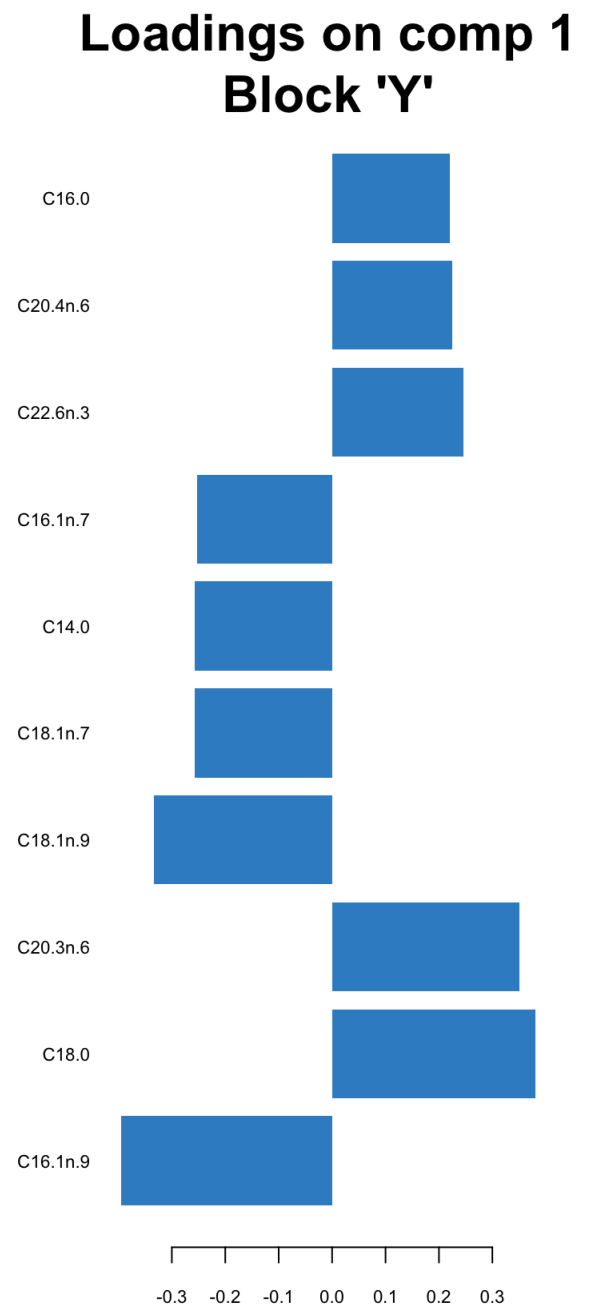
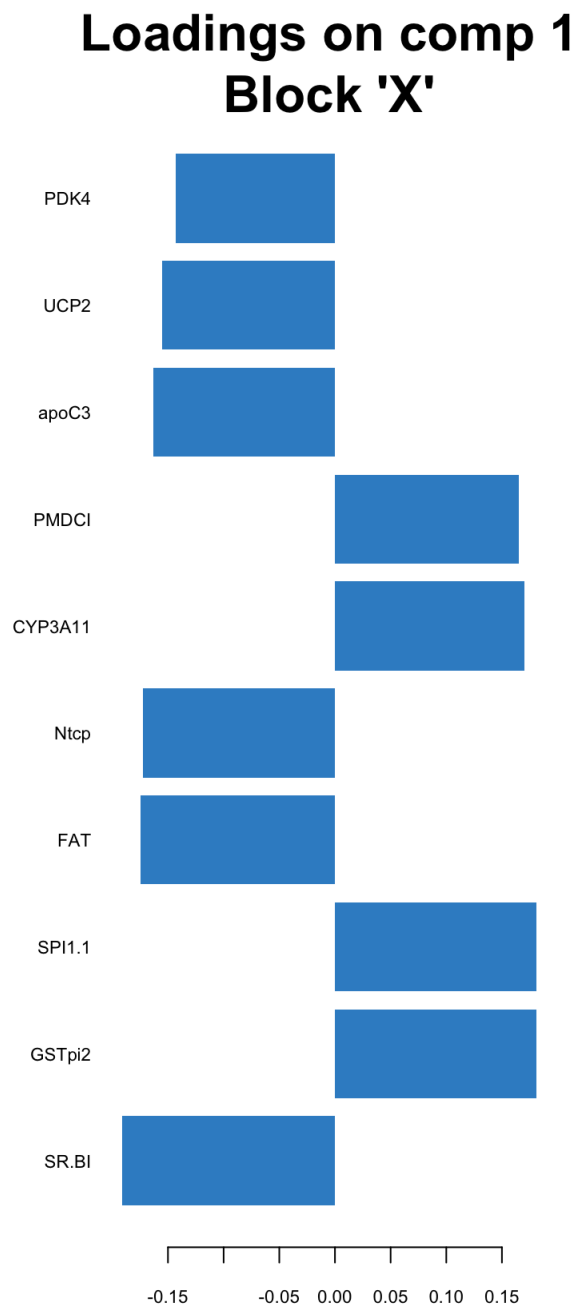
Loading plot: plot of variables' contribution in each data block to each variate, after deflating more *important* variates.

```
par(mfrow=c(1,2), las=2, mar=c(4,8,1,1))
loadings.ind.X <- order(abs(pls.res$loadings$X[, "comp1"]), decreasing = T)
barplot(head(pls.res$loadings$X[loadings.ind.X, "comp1"], 10), main="X", horiz = T, cex.names=0.8)
loadings.ind.Y <- order(abs(pls.res$loadings$Y[, "comp1"]), decreasing = T)
barplot(head(pls.res$loadings$Y[loadings.ind.Y, "comp1"], 10), main="Y", horiz = T, cex.names=0.8)
```



which is the same as with `plotLoadings` .

```
plotLoadings(pls.res, ndisplay = 10)
```

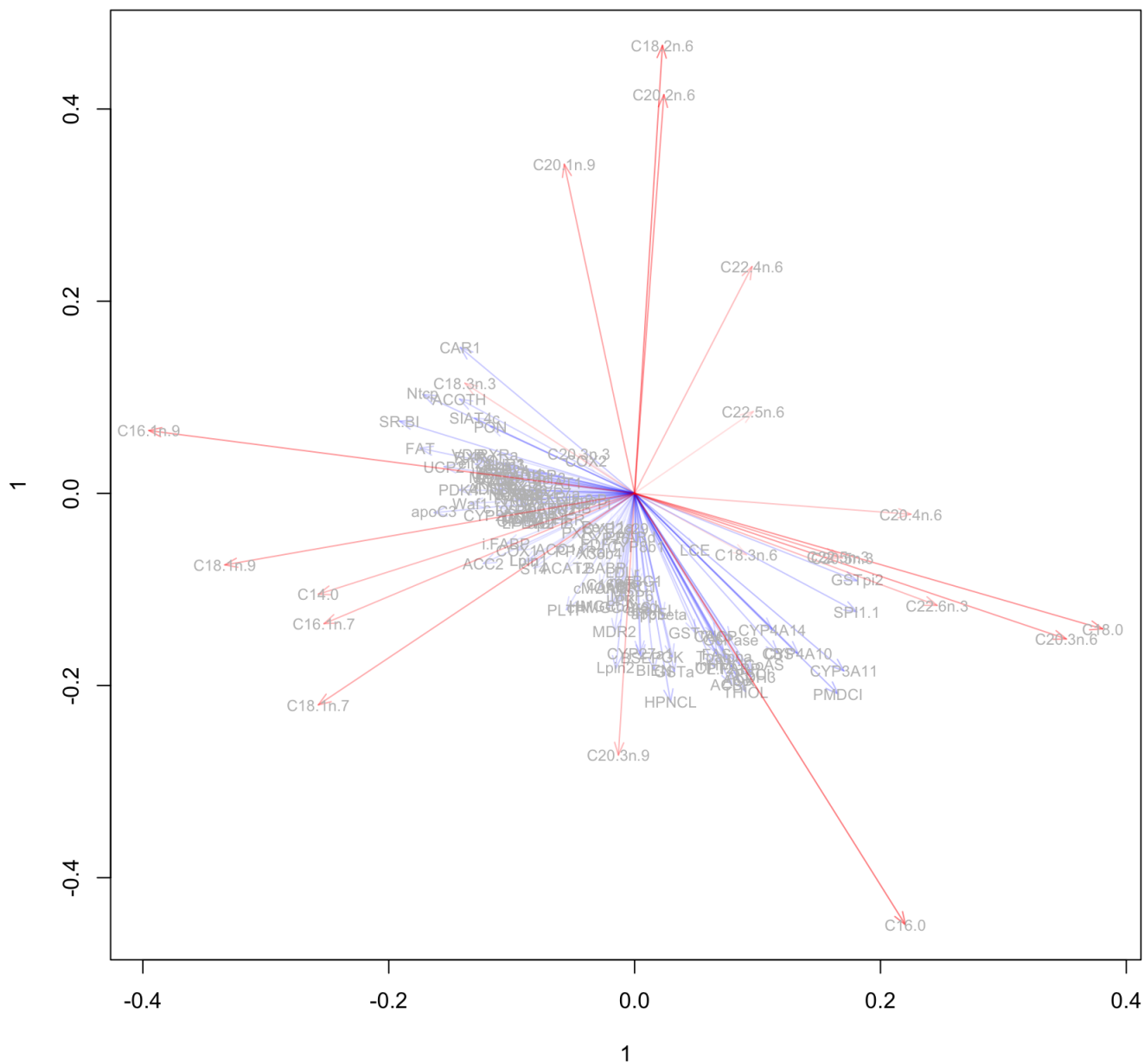


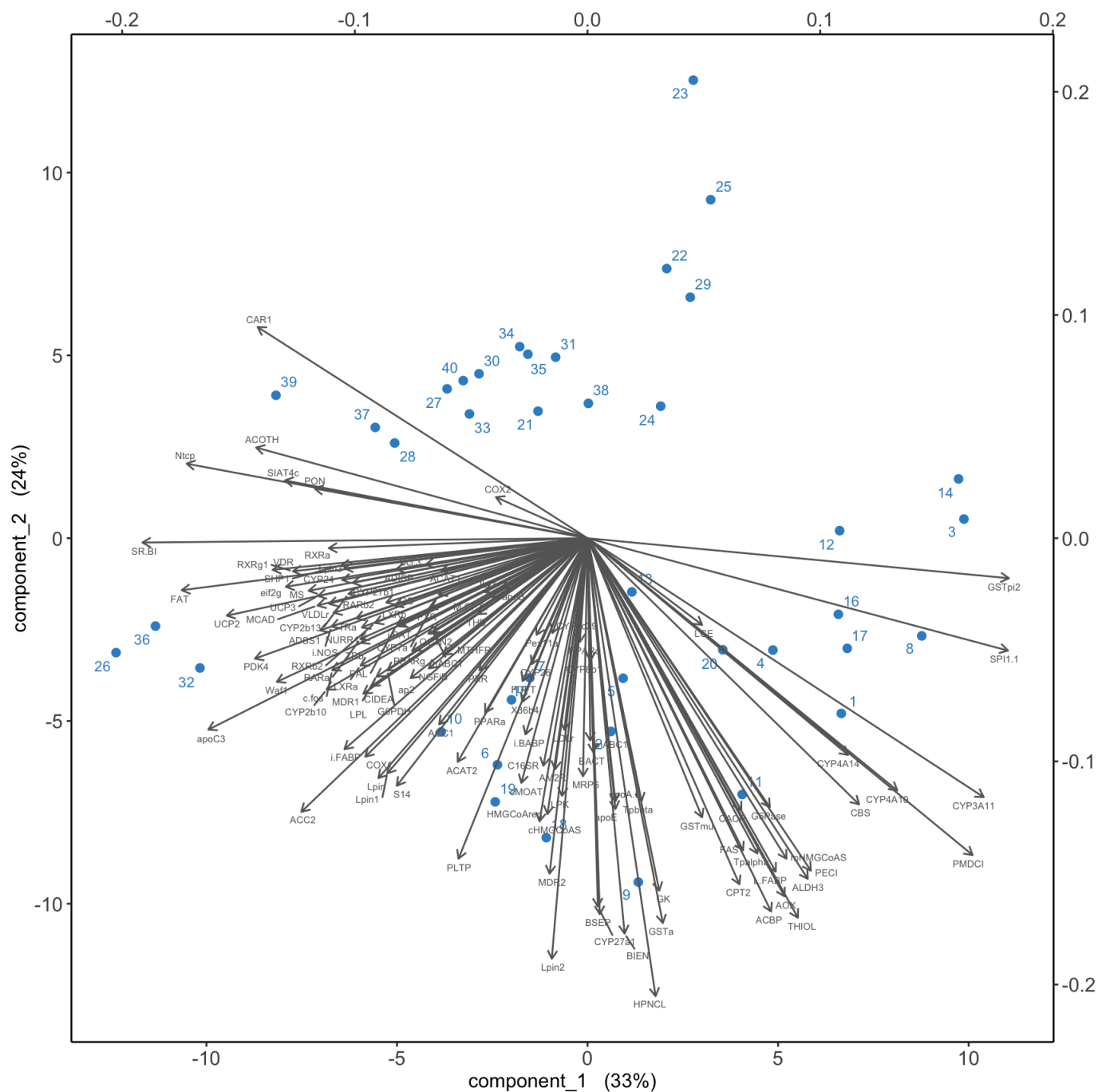
The plot of variable relationship could be obtained from **loadings.star**.

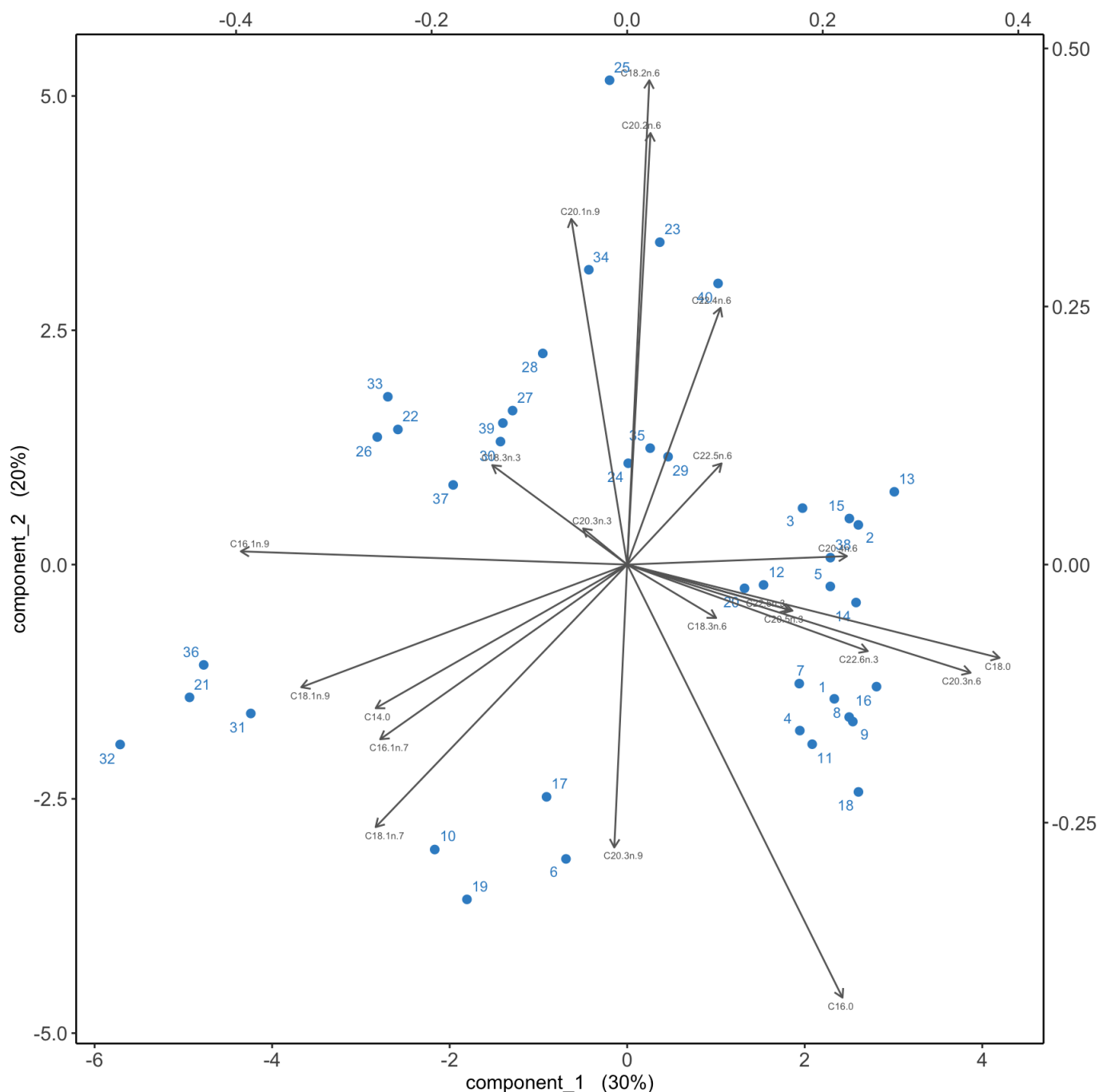
```

names(pls.res$loadings.star) <- c("X", "Y")
colnames(pls.res$loadings.star$X) <- colnames(pls.res$loadings.star$Y) <- c(PCx, PCy)
plot(1,1,type='n',
      xlim=range(c(pls.res$loadings.star$X[, PCx],pls.res$loadings.star$Y[, PCx])),
      ylim=range(c(pls.res$loadings.star$X[, PCy],pls.res$loadings.star$Y[, PCy])))
arrows(0, 0, pls.res$loadings.star$X[, PCx], pls.res$loadings.star$X[, PCy],
       length=0.1, angle=20, col=rgb(0,0,1,alpha=apply(pls.res$loadings.star$X[, c(PC
x, PCy)], 1, norm, "2")))
text(pls.res$loadings.star$X[, PCx],
     pls.res$loadings.star$X[, PCy],
     rownames(pls.res$loadings.star$X), col='grey', cex=0.7)
arrows(0, 0, pls.res$loadings.star$Y[, PCx], pls.res$loadings.star$Y[, PCy],
       length=0.1, angle=20, col=rgb(1,0,0,alpha=apply(pls.res$loadings.star$Y[, c(PC
x, PCy)], 1, norm, "2")))
text(pls.res$loadings.star$Y[, PCx],
     pls.res$loadings.star$Y[, PCy],
     rownames(pls.res$loadings.star$Y), col='grey', cex=0.7)
plotVar(pls.res)

```







2. Observe the difference between the two modes *regression* and *canonical* of PLS.

```
pls.reg.res <- pls(X=nutrimouse$gene, Y=nutrimouse$lipid, ncomp=2, scale=TRUE, mode
="regression")
```

CCA

1. Perform CCA (`mixOmics::rcc`) between 20 genes and all lipids. Investigate correlations, sample distribution and variable relationship with plots.

The gene expression data is reduced to 20 genes so that the number of variables is less than the number of samples, to perform an unregularized CCA.

```
nutrimouse$gene_selected <- as.matrix(nutrimouse$gene[, 1:20])
str(nutrimouse$gene_selected)
```



```
## num [1:40, 1:20] -0.42 -0.44 -0.48 -0.45 -0.42 -0.43 -0.53 -0.49 -0.36 -0.5 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:40] "1" "2" "3" "4" ...
## ..$ : chr [1:20] "X36b4" "ACAT1" "ACAT2" "ACBP" ...
```

```
cca.res <- rcc(X=nutrimouse$gene_selected, Y=as.matrix(nutrimouse$lipid), ncomp=2)
max(abs(nutrimouse$gene - cca.res$X))
```

```
## [1] 2.9
```

```
max(abs(nutrimouse$lipid - cca.res$Y))
```

```
## [1] 0
```

```
str(cca.res)
```

```

## List of 11
## $ call      : language rcc(X = nutrimouse$gene_selected, Y = as.matrix(nutrimouse$lipid), ncomp = 2)
## $ X        : num [1:40, 1:20] -0.42 -0.44 -0.48 -0.45 -0.42 -0.43 -0.53 -0.49 -0.36 -0.5 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:40] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:20] "X36b4" "ACAT1" "ACAT2" "ACBP" ...
## $ Y        : num [1:40, 1:21] 0.34 0.38 0.36 0.22 0.37 1.7 0.35 0.34 0.22 1.3 8 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:40] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:21] "C14.0" "C16.0" "C18.0" "C16.1n.9" ...
## $ ncomp    : num 2
## $ method   : chr "ridge"
## $ cor      : Named num [1:20] 1 1 0.999 0.996 0.981 ...
##   ..- attr(*, "names")= chr [1:20] "1" "2" "3" "4" ...
## $ loadings :List of 2
##   ..$ X: num [1:20, 1:2] 1.408 4.802 3.235 -7.373 -0.724 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:20] "X36b4" "ACAT1" "ACAT2" "ACBP" ...
##   .. .. ..$ : NULL
##   ..$ Y: num [1:21, 1:2] 1.111 -0.143 -0.462 -1.02 -0.09 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:21] "C14.0" "C16.0" "C18.0" "C16.1n.9" ...
##   .. .. ..$ : NULL
## $ variates :List of 2
##   ..$ X: num [1:40, 1:2] -1.203 -1.25 -0.831 0.338 -0.119 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:40] "1" "2" "3" "4" ...
##   .. .. ..$ : NULL
##   ..$ Y: num [1:40, 1:2] -1.203 -1.25 -0.831 0.338 -0.119 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:40] "1" "2" "3" "4" ...
##   .. .. ..$ : NULL
## $ names    :List of 4
##   ..$ sample : chr [1:40] "1" "2" "3" "4" ...
##   ..$ colnames:List of 2
##   .. ..$ X: chr [1:20] "X36b4" "ACAT1" "ACAT2" "ACBP" ...
##   .. ..$ Y: chr [1:21] "C14.0" "C16.0" "C18.0" "C16.1n.9" ...
##   ..$ blocks : chr [1:2] "X" "Y"
##   ..$ data    : chr [1:2] "nutrimouse$gene_selected" "as.matrix(nutrimouse$lipid)"
## $ lambda    : Named num [1:2] 0 0
##   ..- attr(*, "names")= chr [1:2] "lambda1" "lambda2"
## $ prop_expl_var:List of 2
##   ..$ X: Named num [1:2] 0.00132 0.0024
##   .. ..- attr(*, "names")= chr [1:2] "comp1" "comp2"
##   ..$ Y: Named num [1:2] 0.0184 0.0299
##   .. ..- attr(*, "names")= chr [1:2] "comp1" "comp2"
## - attr(*, "class")= chr "rcc"

```

```
cca.res$cor
```

```
##           1           2           3           4           5           6           7
## 1.00000000 1.00000000 0.99922446 0.99607902 0.98142435 0.95641141 0.89083472
##           8           9          10          11          12          13          14
## 0.88959894 0.78648273 0.76470925 0.75189350 0.66984945 0.63240310 0.53662009
##          15          16          17          18          19          20
## 0.49948385 0.34852831 0.33274136 0.27818295 0.22569639 0.03783839
```

The sample distribution plot can be performed with **variates**, sample coordinates in the new reference (rotated axes) for each of the two blocks.

```
str(cca.res$variates)
```

```
## List of 2
## $ X: num [1:40, 1:2] -1.203 -1.25 -0.831 0.338 -0.119 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:40] "1" "2" "3" "4" ...
## .. ..$ : NULL
## $ Y: num [1:40, 1:2] -1.203 -1.25 -0.831 0.338 -0.119 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:40] "1" "2" "3" "4" ...
## .. ..$ : NULL
```

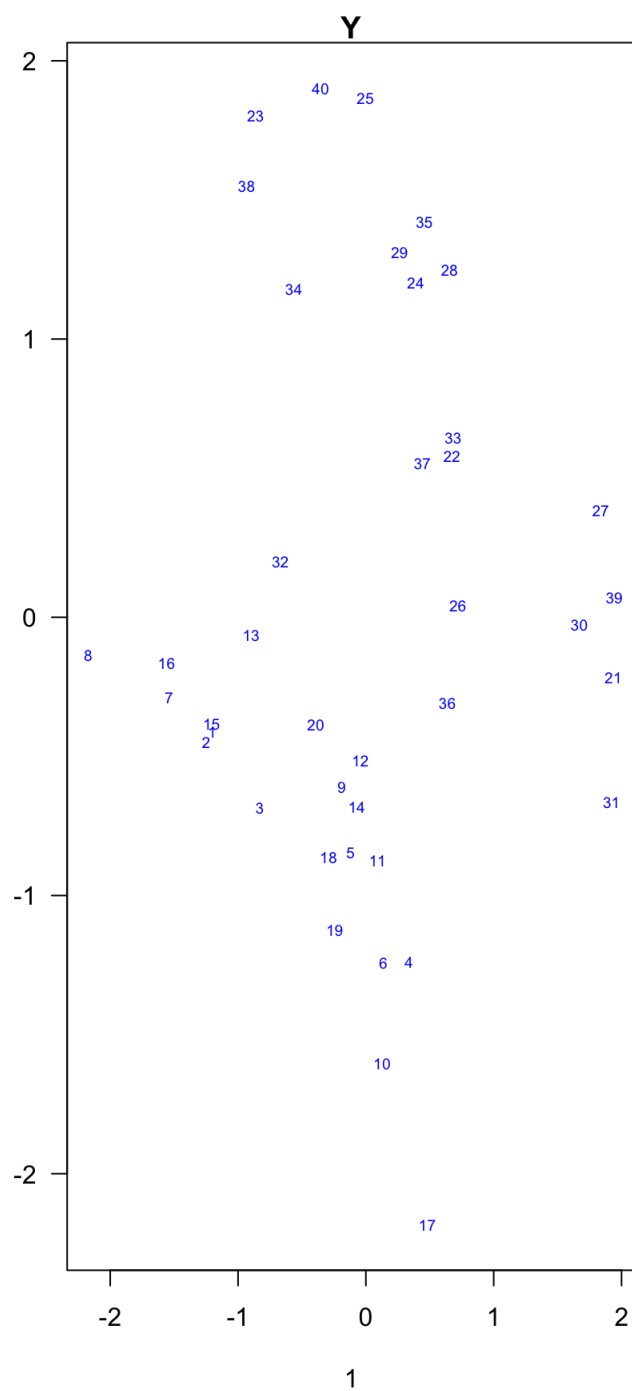
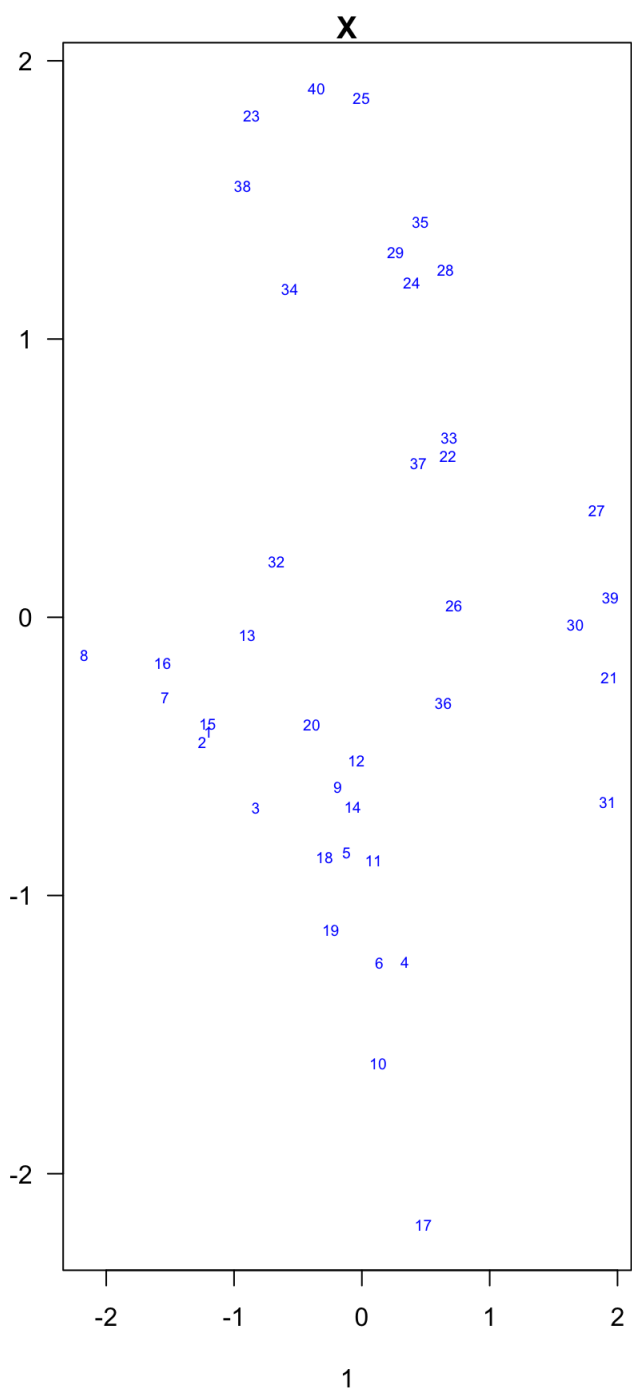
```
PCx <- 1
PCy <- 2
par(mfrow=c(1,2), las=1, mar=c(4,3,1,1))
plot(cca.res$variates$X[, PCx], cca.res$variates$X[, PCy], xlab=PCx, ylab=PCy, main="X", type='n')
text(cca.res$variates$X[, PCx], cca.res$variates$X[, PCy], rownames(cca.res$variates$X), col='blue', cex=0.6)
plot(cca.res$variates$Y[, PCx], cca.res$variates$Y[, PCy], xlab=PCx, ylab=PCy, main="Y", type='n')
text(cca.res$variates$Y[, PCx], cca.res$variates$Y[, PCy], rownames(cca.res$variates$Y), col='blue', cex=0.6)
cor(cca.res$variates$X[,1], cca.res$variates$Y[,1])
```

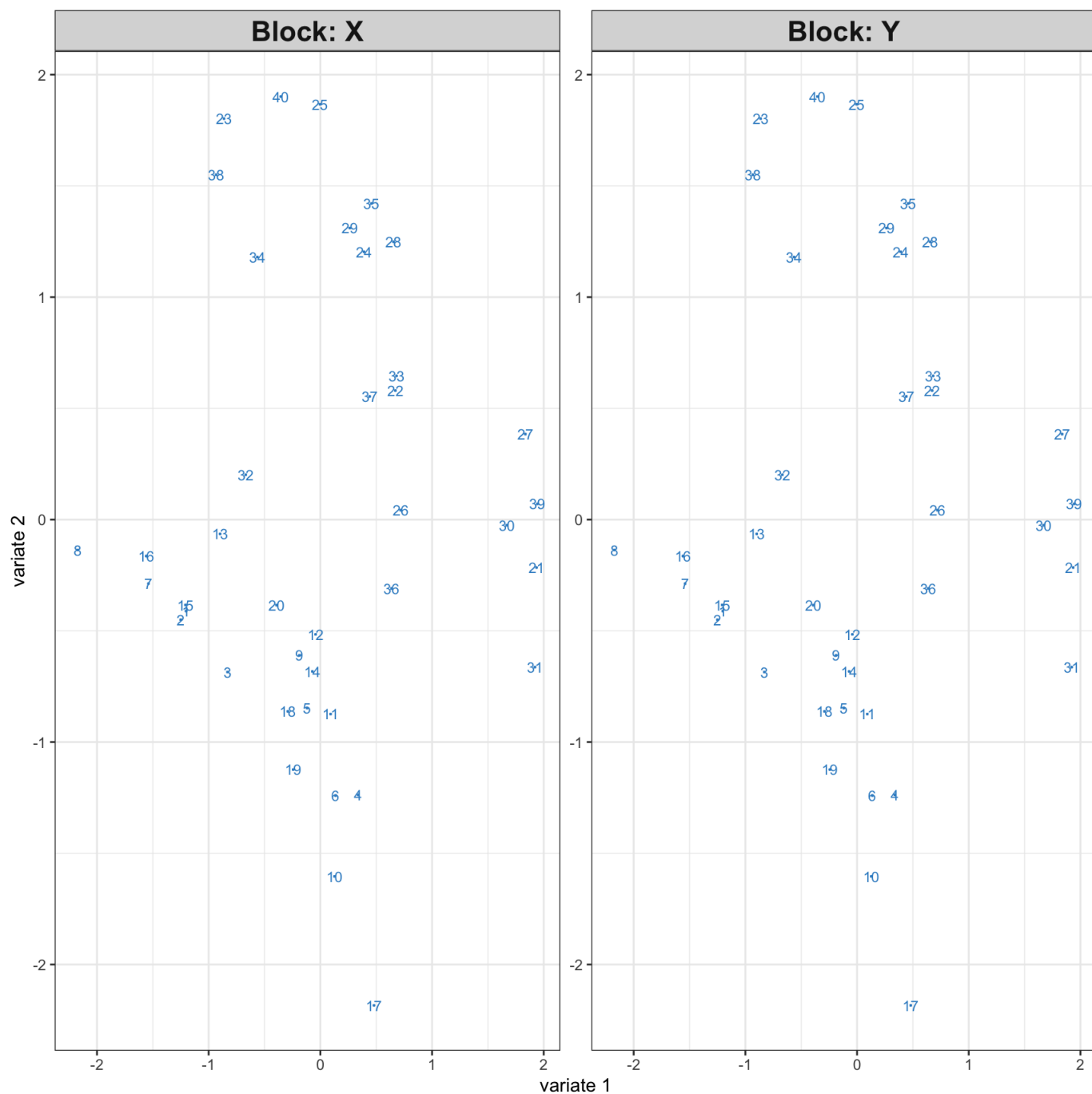
```
## [1] 1
```

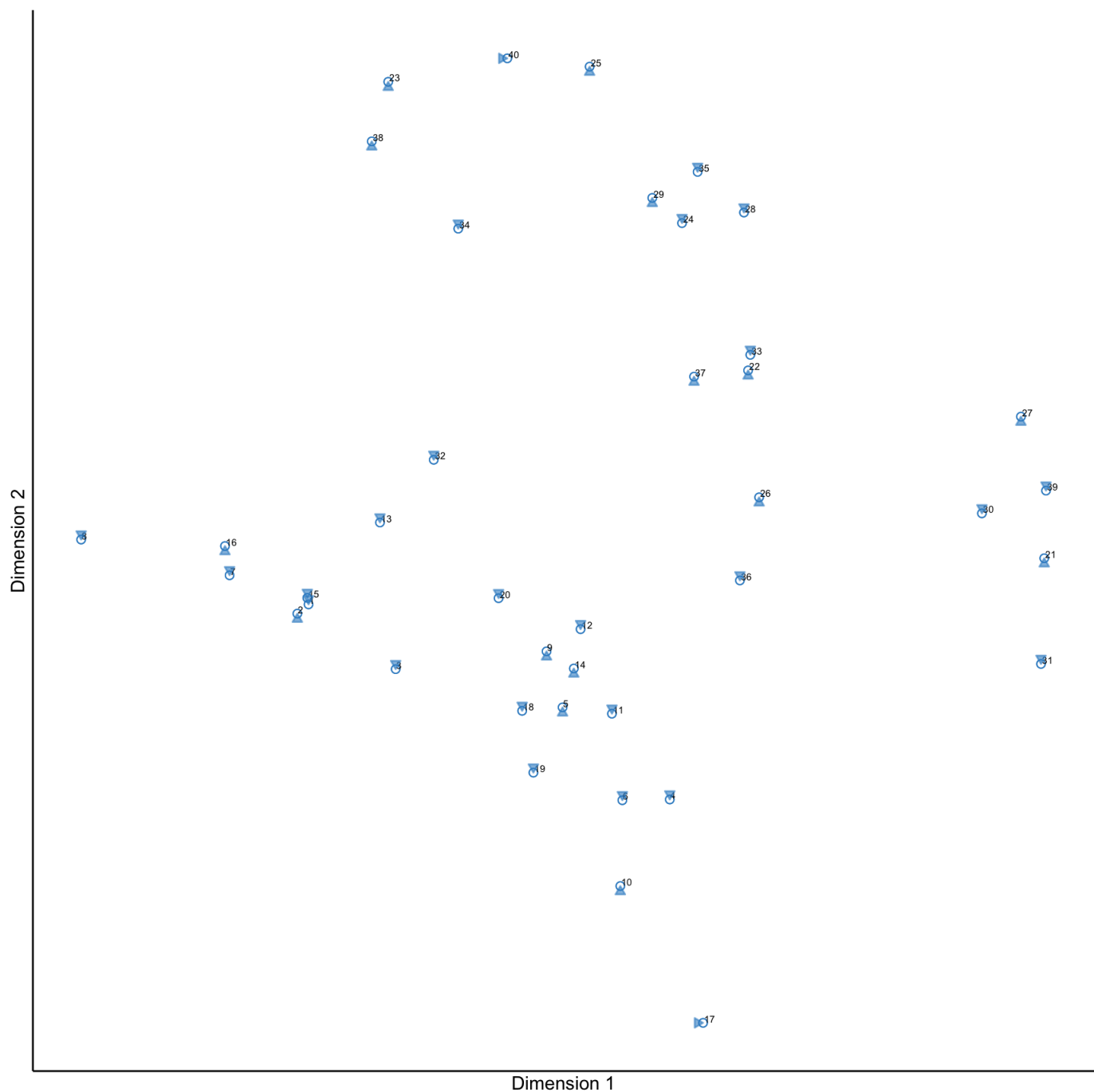
```
cor(cca.res$variates$X[,2], cca.res$variates$Y[,2])
```

```
## [1] 1
```

```
plotIndiv(cca.res)
plotArrow(cca.res)
```







Variable relationship is obtained from **loadings** or with `plotVar` .

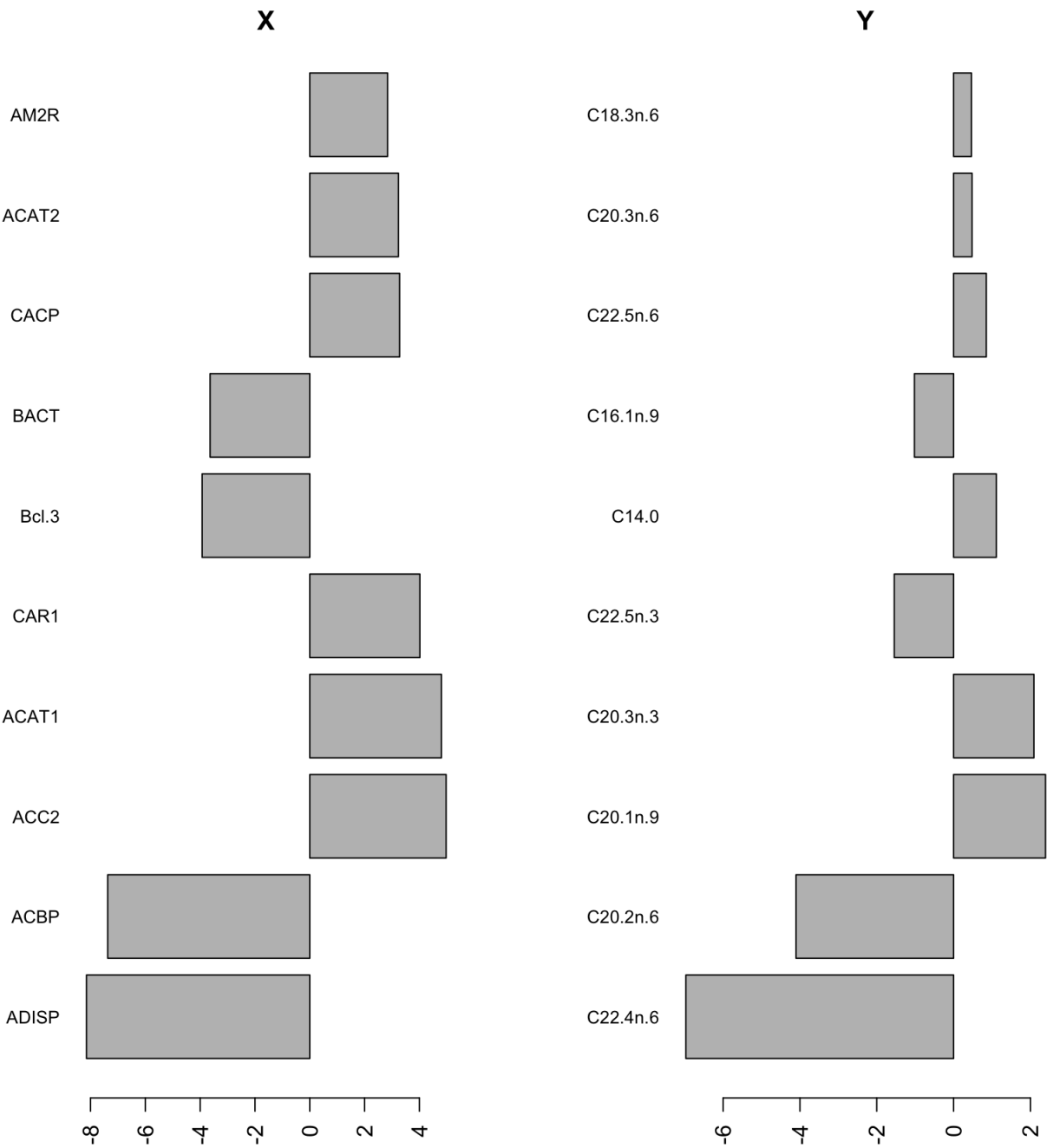
```
par(mfrow=c(1,2), las=2, mar=c(4,8,1,1))
loadings.ind.X <- order(abs(cca.res$loadings$X[, 1]), decreasing = T)
barplot(head(cca.res$loadings$X[loadings.ind.X, 1], 10), main="X", horiz = T, cex.names=0.8)
loadings.ind.Y <- order(abs(cca.res$loadings$Y[, 1]), decreasing = T)
barplot(head(cca.res$loadings$Y[loadings.ind.Y, 1], 10), main="Y", horiz = T, cex.names=0.8)
max(abs(cca.res$variates$X - scale(cca.res$X, center=T, scale=F) %*% cca.res$loadings$X))
```

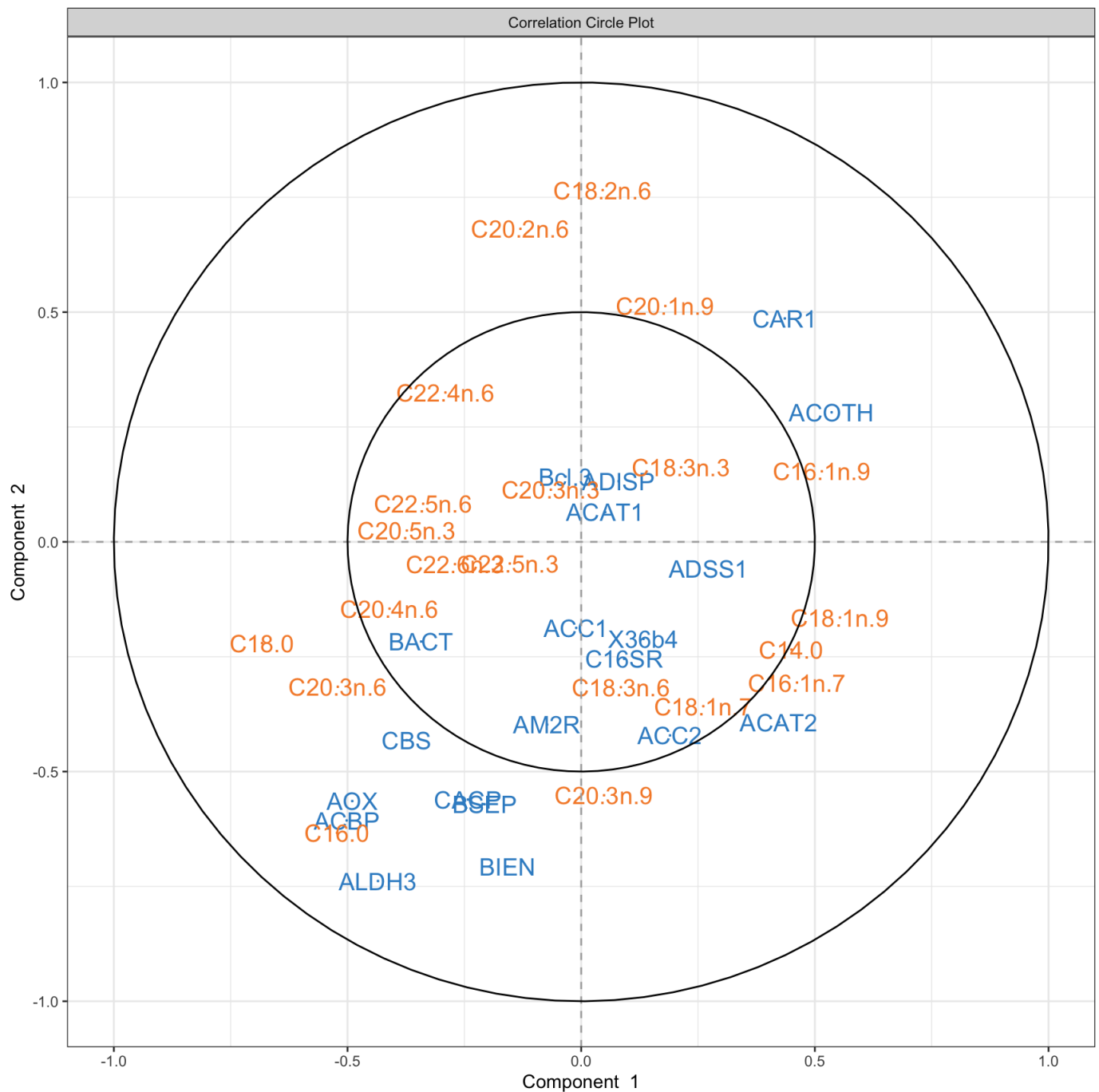
```
## [1] 0
```

```
max(abs(cca.res$variates$Y - scale(cca.res$Y, center=T, scale=F) %*% cca.res$loadings$Y))
```

```
## [1] 0
```

```
plotVar(cca.res)
```





2. Perform CCA with scaled datasets and observe the difference

```
cca.res.scale <- rcc(X=scale(nutrimouse$gene_selected, center=T, scale=T),
                    Y=scale(nutrimouse$lipid, center=T, scale=T), ncomp=2, method =
"shrinkage")
max(abs(cca.res.scale$cor - cca.res$cor))
```

```
## [1] 0.4797719
```

```
max(abs(cca.res.scale$variates$X - cca.res$variates$X))
```

```
## [1] 3.631284
```

```
max(abs(cca.res.scale$variates$Y - cca.res$variates$Y))
```



```
## [1] 3.841865
```

```
max(abs(cca.res.scale$loadings$X - cca.res$loadings$X))
```

```
## [1] 8.113192
```

```
max(abs(cca.res.scale$loadings$Y - cca.res$loadings$Y))
```

```
## [1] 16.07948
```

3. Perform regularized CCA with all genes and lipids.

```
rcca.res <- rcc(X=nutrimouse$gene, Y=nutrimouse$lipid, ncomp=2, method="shrinkage")
```