

Milk (v5) split

Contents

Objectives	1
Analysis	1
Split	1
LMMS	1
C-section	1
Vaginal	2
Filter	3
C-section	3
Vaginal	4
PCA Clustering	5
C-section	5
Vaginal	10
sparse PCA Clustering	15
C-section	15
Vaginal	19
Results	23

```
library(tidyverse)
library(mixOmics)
walk(dir("~/Documents/timeOmics_dev/R/", pattern = ".R$", full.names = TRUE), source)
load("./milk_data.RData")
```

Objectives

Split data in devlivery mode. Re-run, modelisation, pca-clustering, get silhouette coef.

Analysis

Split

```
delivery_mode <- design %>% dplyr::select(BABY, Delivery) %>% unique
index.C <- rownames(OTU) %>% str_split("_") %>% map_chr(~.x[1]) %in%
  (delivery_mode %>% filter(Delivery == "C-section") %>% pull(BABY))
OTU_norm.C <- OTU[index.C,] %>% norm_OTU(AR = T)
OTU_norm.V <- OTU[!index.C,] %>% norm_OTU(AR = T)
```

LMMS

C-section

```

time_lmms.C <- rownames(OTU_norm.C) %>% str_split("_") %>% map_chr(~.x[2]) %>% as.numeric
sample_id = rownames(OTU_norm.C)

# cubic p-spline
spline.MILK.C.cubicpspline = lmms::lmmSpline(data = OTU_norm.C, time = time_lmms.C,
                                             sampleID = sample_id,
                                             basis = 'cubic p-spline', keepModels = T,
                                             numCores = 2)

spline.MILK.C.ppspline = lmms::lmmSpline(data = OTU_norm.C, time = time_lmms.C,
                                           sampleID = sample_id,
                                           basis = 'p-spline', keepModels = T,
                                           numCores = 2 )

spline.MILK.C.cubic = lmms::lmmSpline(data = OTU_norm.C, time = time_lmms.C,
                                       sampleID = sample_id,
                                       basis = 'cubic', keepModels = T,
                                       numCores = 2)

# summary
spline.MILK.C.cubicpspline@modelsUsed %>% table %>% as.data.frame() %>%
  set_names("ModelUsed", "Cubic P-spline") %>%
  left_join(spline.MILK.C.ppspline@modelsUsed %>% table %>% as.data.frame() %>%
    set_names("ModelUsed", "P-spline")) %>%
  left_join(spline.MILK.C.cubic@modelsUsed %>% table %>% as.data.frame() %>%
    set_names("ModelUsed", "Cubic")) %>%
  knitr::kable()

```

ModelUsed	Cubic P-spline	P-spline	Cubic
0	95	78	82
1	12	29	25

Vaginal

```

time_lmms.V <- rownames(OTU_norm.V) %>% str_split("_") %>% map_chr(~.x[2]) %>% as.numeric
sample_id = rownames(OTU_norm.V)

# cubic p-spline
spline.MILK.V.cubicpspline = lmms::lmmSpline(data = OTU_norm.V, time = time_lmms.V,
                                             sampleID = sample_id,
                                             basis = 'cubic p-spline', keepModels = T,
                                             numCores = 2)

spline.MILK.V.ppspline = lmms::lmmSpline(data = OTU_norm.V, time = time_lmms.V,
                                           sampleID = sample_id,
                                           basis = 'p-spline', keepModels = T,
                                           numCores = 2 )

spline.MILK.V.cubic = lmms::lmmSpline(data = OTU_norm.V, time = time_lmms.V,
                                       sampleID = sample_id,
                                       basis = 'cubic', keepModels = T,
                                       numCores = 2)

```

```
# summary
spline.MILK.V.cubicpspline@modelsUsed %>% table %>% as.data.frame() %>%
  set_names("ModelUsed", "Cubic P-spline") %>%
  left_join(spline.MILK.V.pspline@modelsUsed %>% table %>% as.data.frame() %>%
    set_names("ModelUsed", "P-spline")) %>%
  left_join(spline.MILK.V.cubic@modelsUsed %>% table %>% as.data.frame() %>%
    set_names("ModelUsed", "Cubic")) %>%
knitr::kable()
```

ModelUsed	Cubic P-spline	P-spline	Cubic
0	107	95	98
1	10	22	19

More straight lines in every delivery mode. Less in pspline.

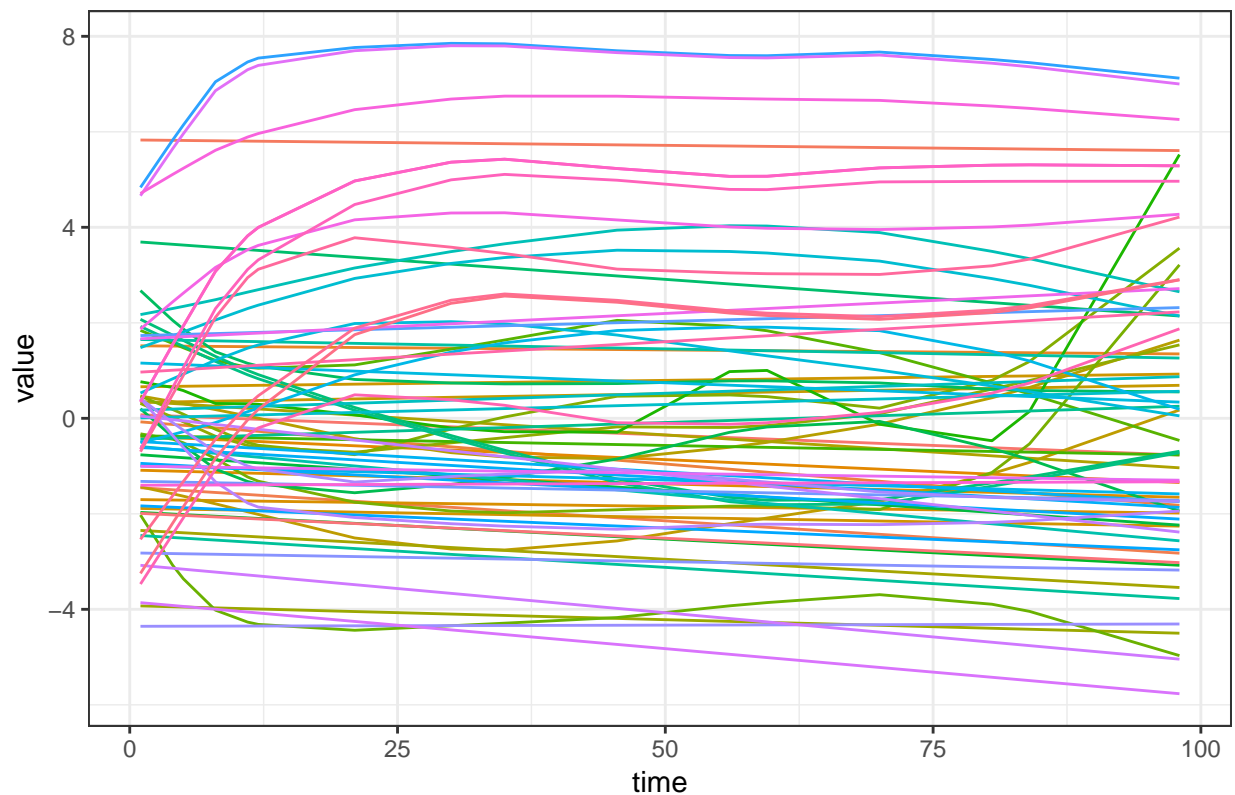
Filter

C-section

```
filter.spline.C.res <- wrapper.filter.splines(OTU_norm.C, spline.MILK.C.pspline)
index.filter.C <- (rownames(spline.MILK.C.pspline@predSpline) %in% filter.spline.C.res$to_keep) %>%
  which

spline.data.C <- spline.MILK.C.pspline@predSpline[index.filter.C,] %>% t %>% as.data.frame()
spline.data.C %>% rownames_to_column("time") %>%
  gather(Features, value, - time) %>% mutate(time =as.numeric(time)) %>%
  ggplot(aes(x=time, y = value, col = Features)) + geom_line() + theme_bw() +
  theme(legend.position = "none") + ggtitle("Modelled OTU evolution")
```

Modelled OTU evolution

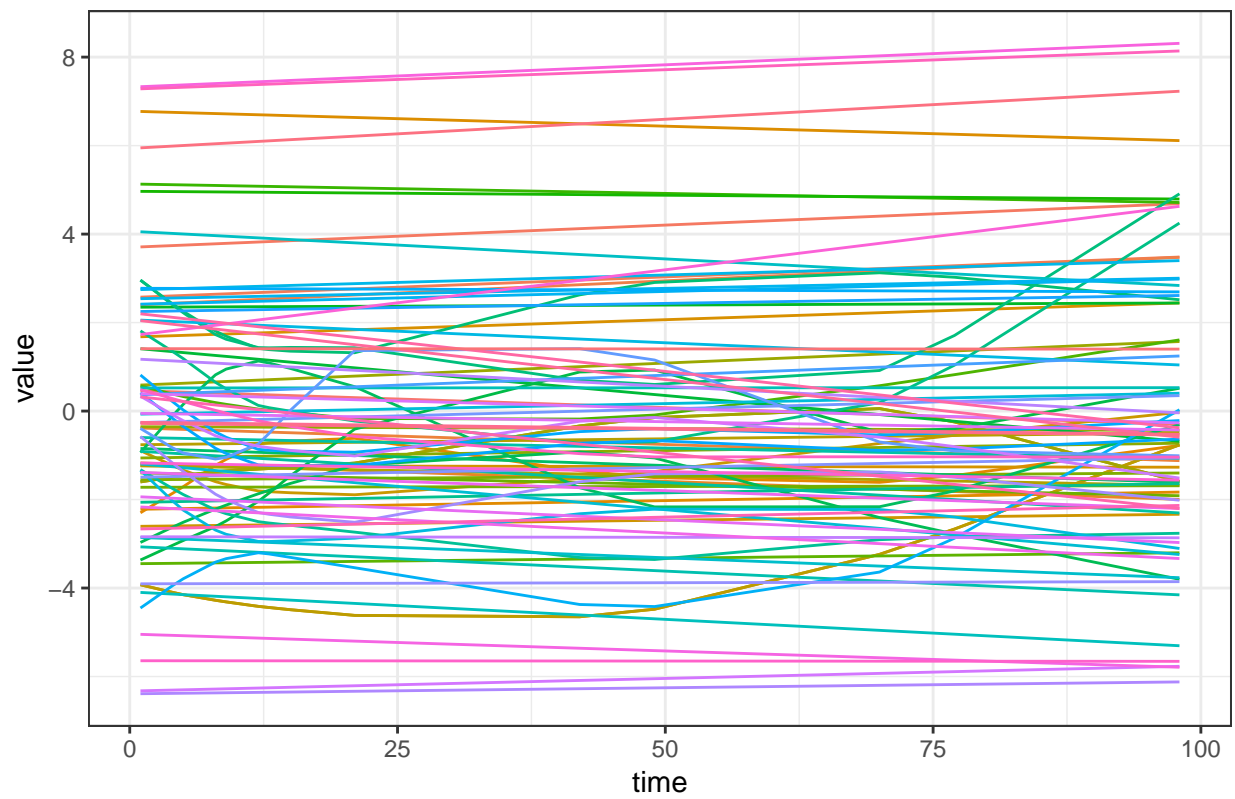


Vaginal

```
filter.spline.V.res <- wrapper.filter.splines(OTU_norm.V, spline.MILK.V.pspline)
index.filter.V <- (rownames(spline.MILK.V.pspline@predSpline) %in% filter.spline.V.res$to_keep) %>%
  which()

spline.data.V <- spline.MILK.V.pspline@predSpline[index.filter.V,] %>% t %>% as.data.frame()
spline.data.V %>% rownames_to_column("time") %>%
  gather(Features, value, - time) %>% mutate(time =as.numeric(time)) %>%
  ggplot(aes(x=time, y = value, col = Features)) + geom_line() + theme_bw() +
  theme(legend.position = "none") + ggtitle("Modelled OTU evolution")
```

Modelled OTU evolution

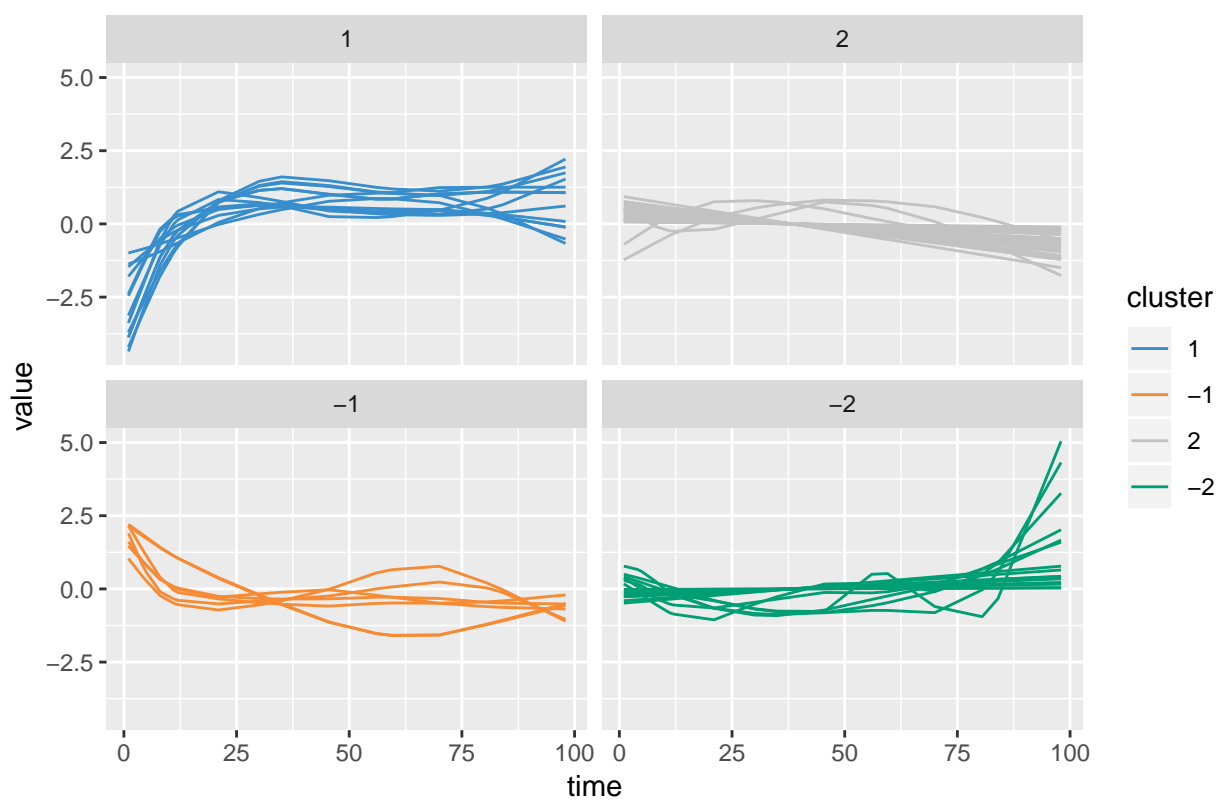


PCA Clustering

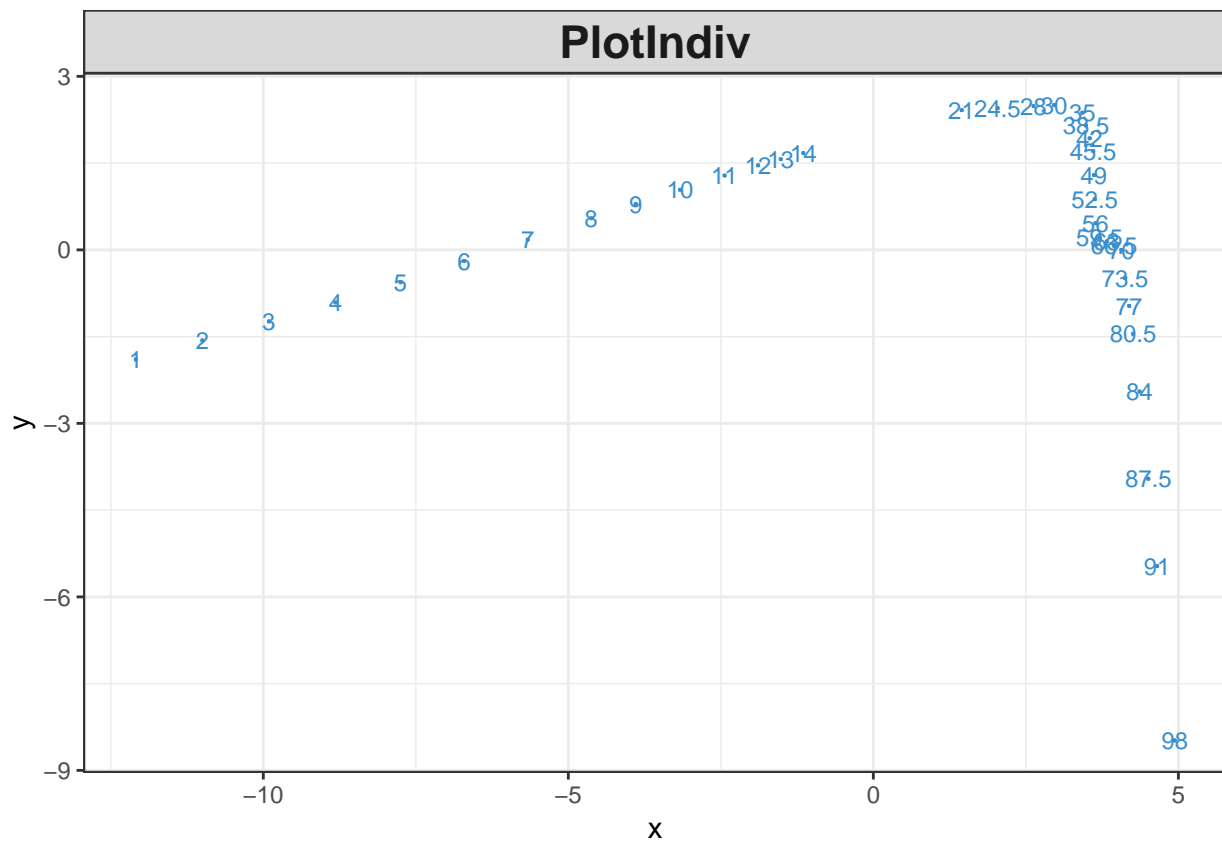
C-section

```
pca.res.C <- pca(spline.data.C, ncomp = 2, scale = F, center = T)
pca.plot(pca.res.C, title = "PCA, with lines, scale = F")
```

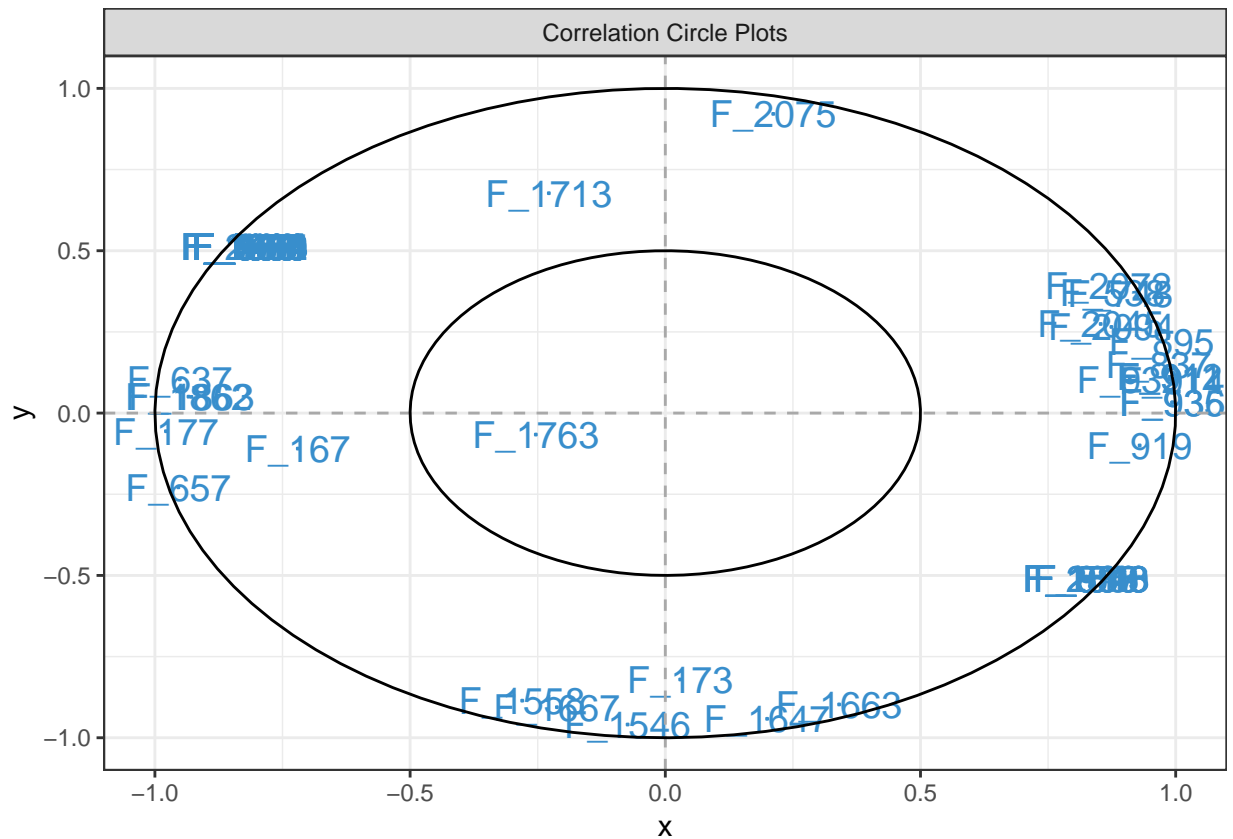
PCA, with lines, scale = F



```
plotIndiv(pca.res.C)
```



```
plotVar(pca.res.C)
```

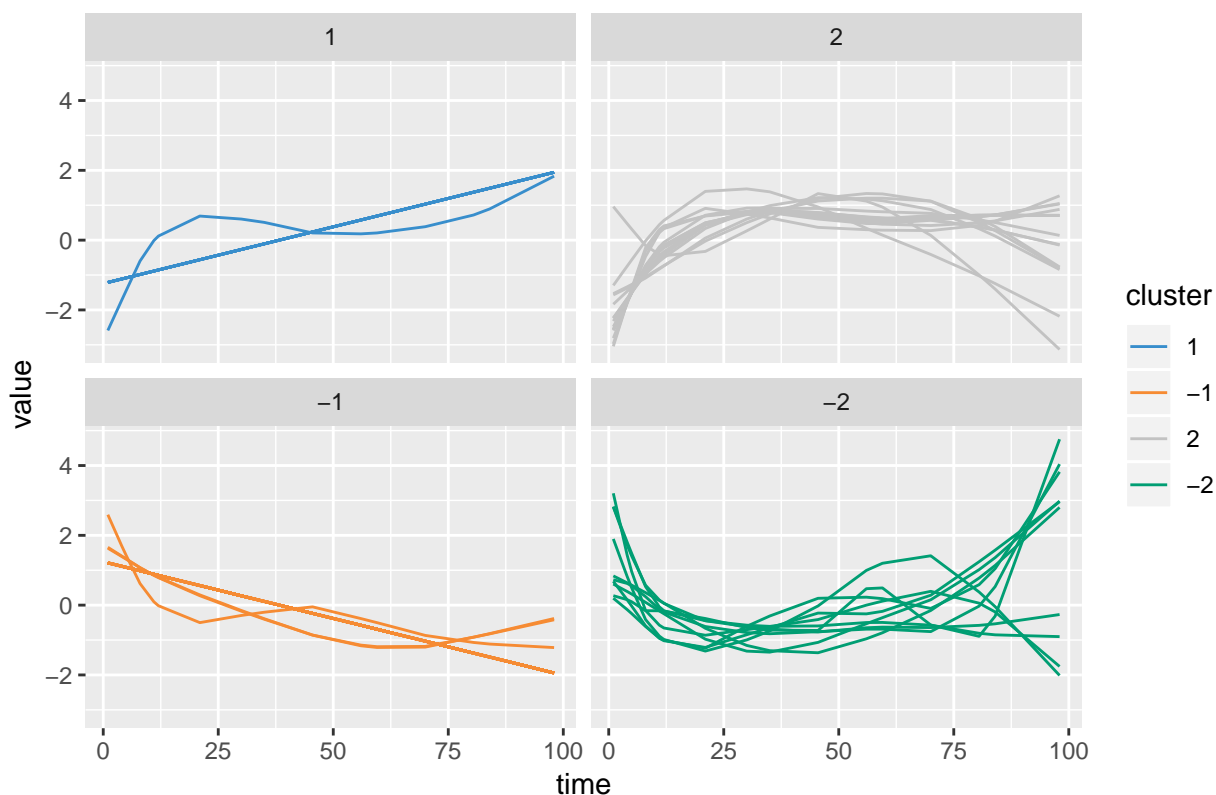


```
pca.res.C <- pca(spline.data.C, ncomp = 2, scale = T, center = T)
pca.get_cluster(pca.res.C) %>% pull(cluster) %>% table
```

```
## .
## -2 -1 1 2
## 10 35 11 15
```

```
pca.plot(pca.res.C, title = "PCA, with lines, scale = T")
```


PCA, with lines, scale = T



```
pca.res.C <- pca(spline.data.C, ncomp = 2, scale = F, center = T)
```

```
# silhouette coefficient for this clustering
```

```
wrapper.silhouette.pca(spline.data.C, ncomp = 2, scale = T, center=T)
```

```
## [1] 0.8426561
```

```
remove lines
```

```
spline.0.C <- spline.MILK.C.pspline@predSpline[spline.MILK.C.pspline@modelsUsed != 0, ] %>%  
  t %>% as.data.frame()  
# no filter needed
```

```
pca.res.0.C <- pca(spline.0.C, ncomp = 2, scale = T, center = T)
```

```
pca.get_cluster(pca.res.0.C) %>% pull(cluster) %>% table
```

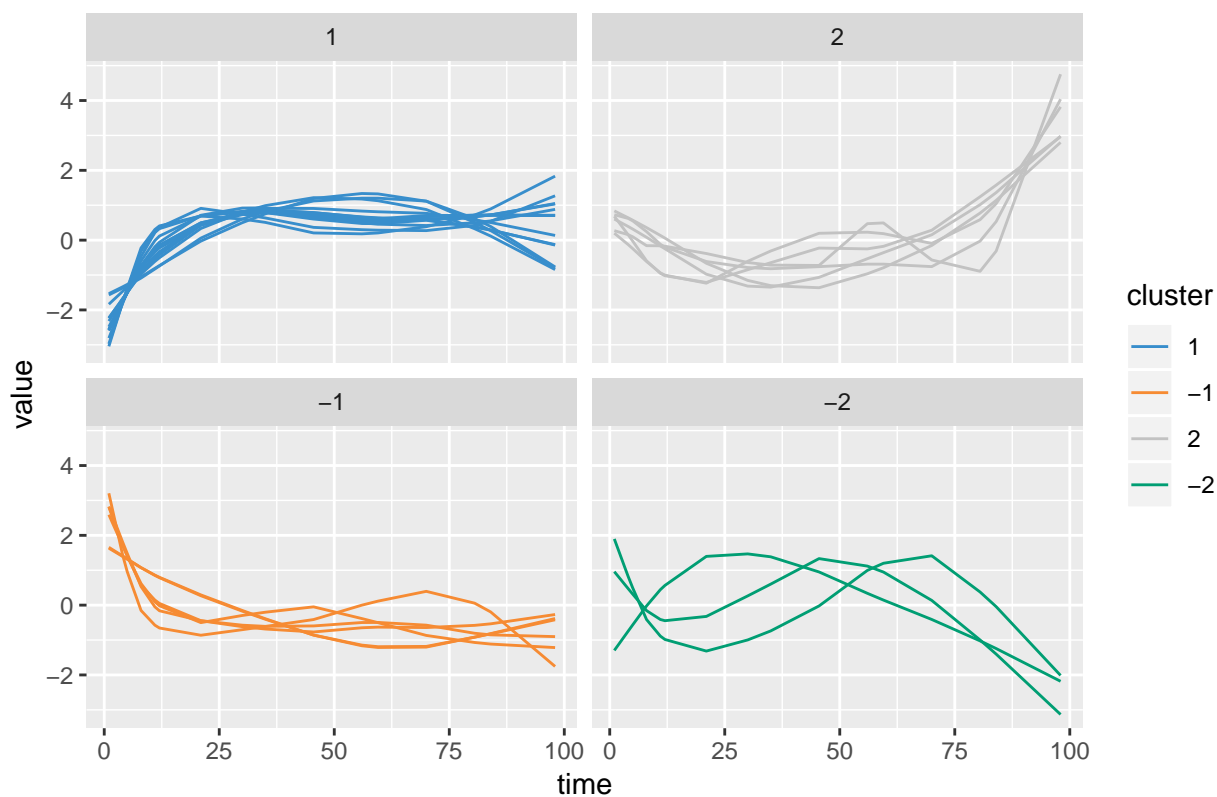
```
## .
```

```
## -2 -1 1 2
```

```
## 3 6 14 6
```

```
pca.plot(pca.res.0.C)
```

PCA



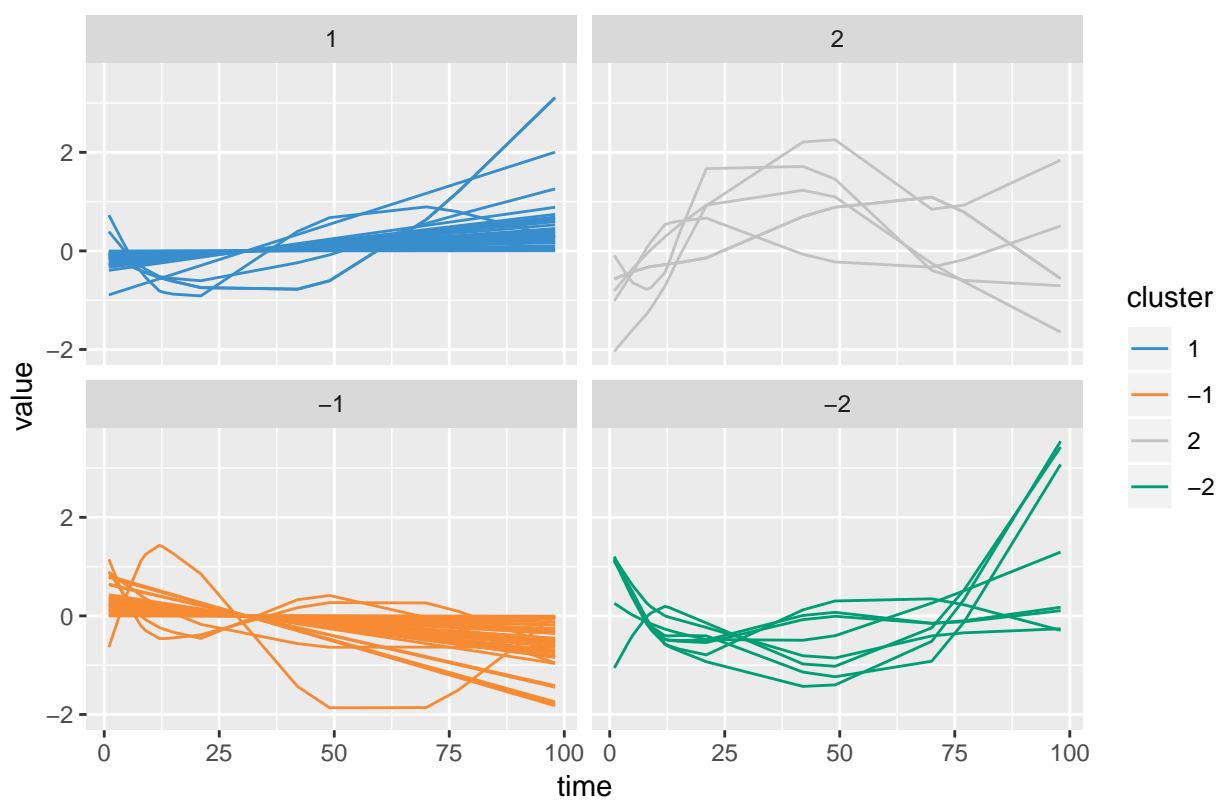
```
# silhouette coefficient for this clustering
wrapper.silhouette.pca(spline.0.C, ncomp = 2, scale = T, center=T)
```

```
## [1] 0.698184
```

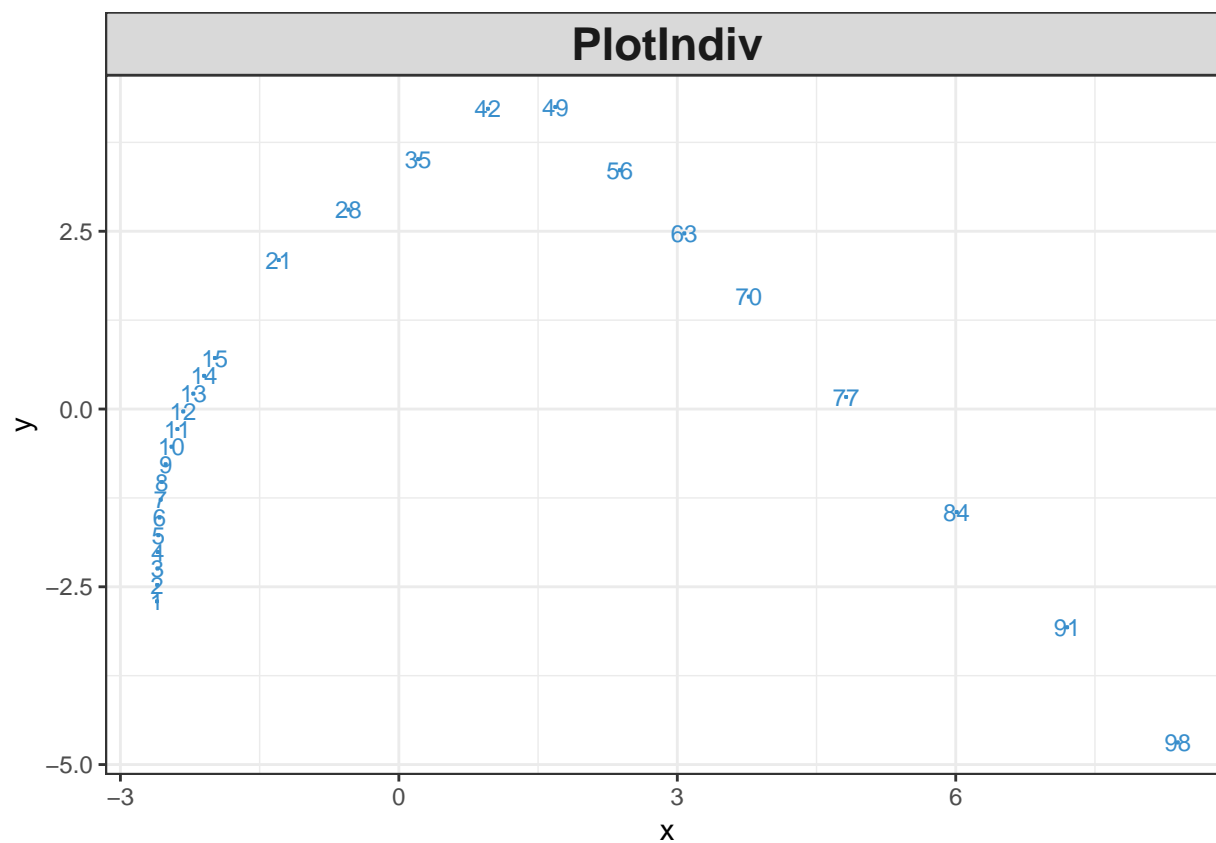
Vaginal

```
pca.res.V <- pca(spline.data.V, ncomp = 2, scale = F, center = T)
pca.plot(pca.res.V, title = "PCA, with lines, scale = F")
```

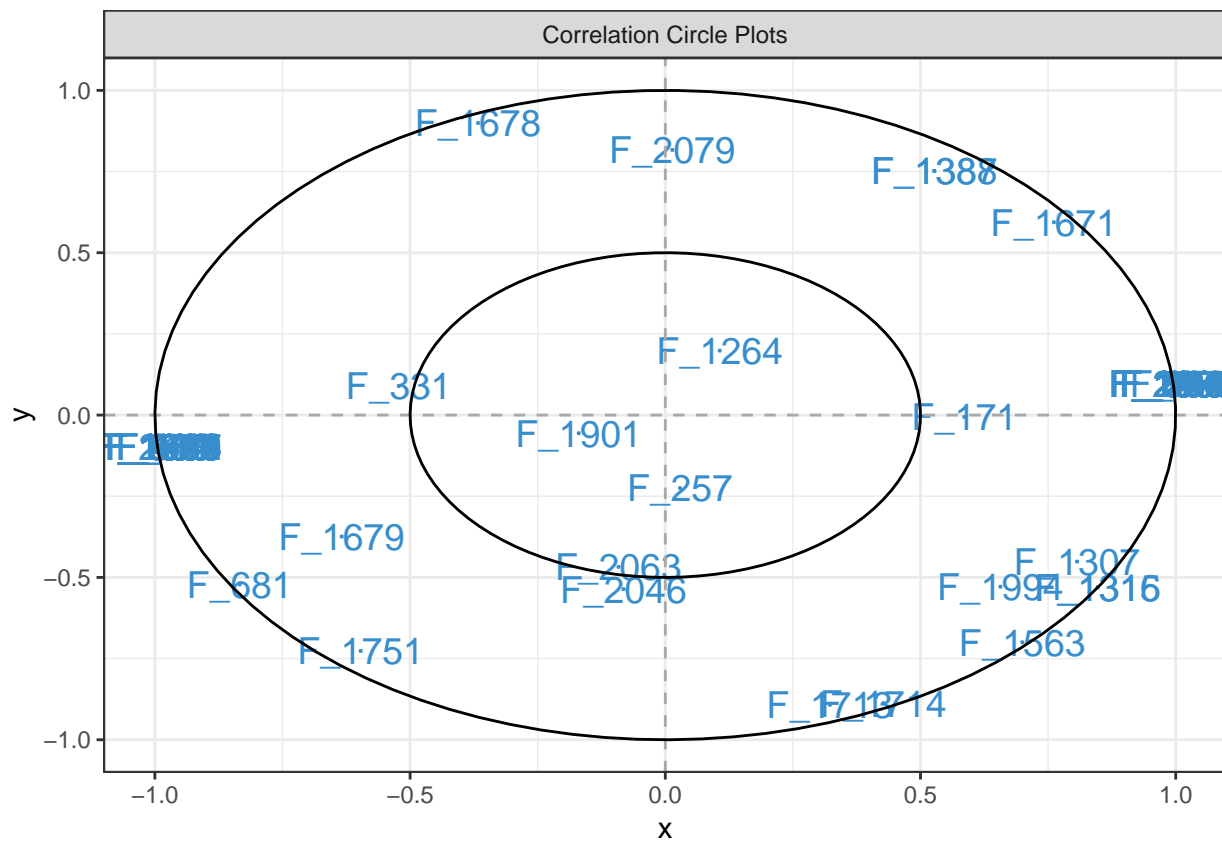
PCA, with lines, scale = F



```
plotIndiv(pca.res.V)
```



```
plotVar(pca.res.V)
```

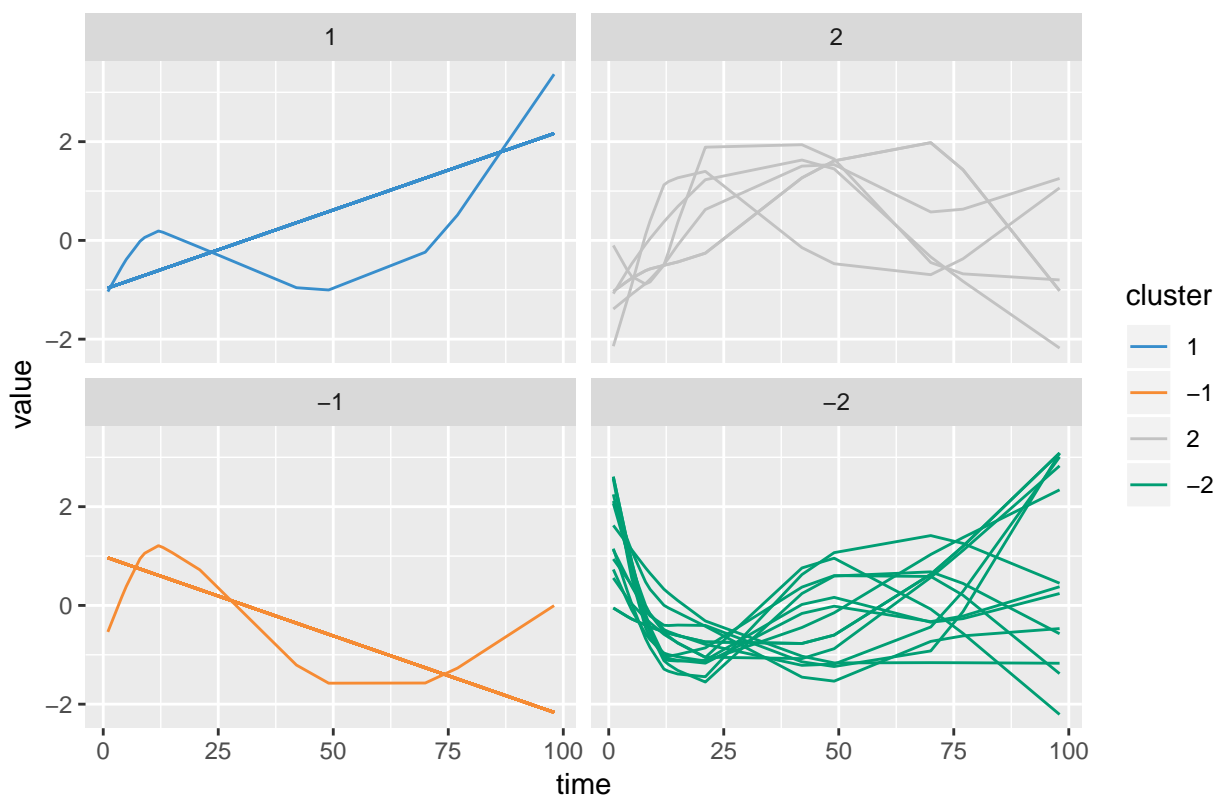


```
pca.res.V <- pca(spline.data.V, ncomp = 2, scale = T, center = T)
pca.get_cluster(pca.res.V) %>% pull(cluster) %>% table
```

```
## .
## -2 -1 1 2
## 14 38 32 6
```

```
pca.plot(pca.res.V, title = "PCA, with lines, scale = T")
```

PCA, with lines, scale = T



```
pca.res.V <- pca(spline.data.V, ncomp = 2, scale = F, center = T)
```

```
# silhouette coefficient for this clustering
```

```
wrapper.silhouette.pca(spline.data.V, ncomp = 2, scale = T, center=T)
```

```
## [1] 0.8730084
```

```
remove lines
```

```
spline.0.V <- spline.MILK.V.pspline@predSpline[spline.MILK.V.pspline@modelsUsed != 0, ] %>%  
  t %>% as.data.frame()  
# no filter needed
```

```
pca.res.0.V <- pca(spline.0.V, ncomp = 2, scale = T, center = T)
```

```
pca.get_cluster(pca.res.0.V) %>% pull(cluster) %>% table
```

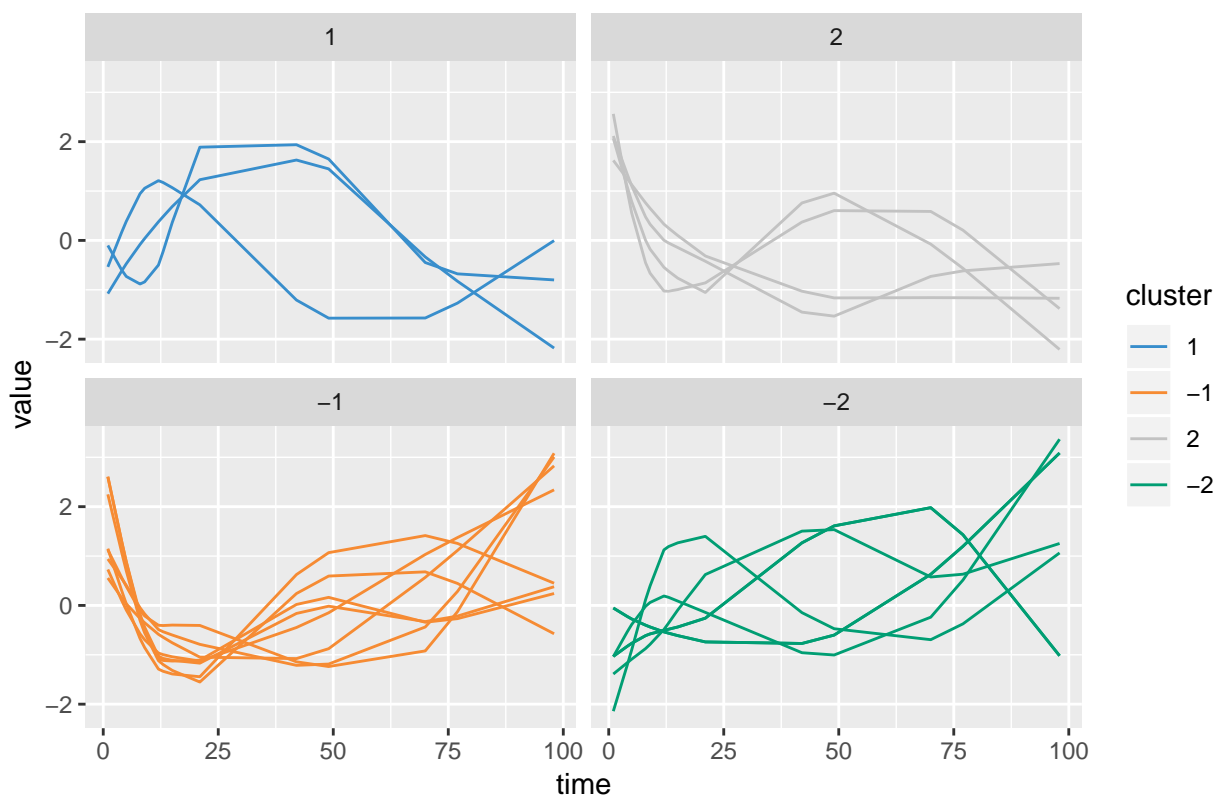
```
## .
```

```
## -2 -1 1 2
```

```
## 7 8 3 4
```

```
pca.plot(pca.res.0.V)
```

PCA



```
# silhouette coefficient for this clustering
wrapper.silhouette.pca(spline.0.V, ncomp = 2, scale = T, center=T)
```

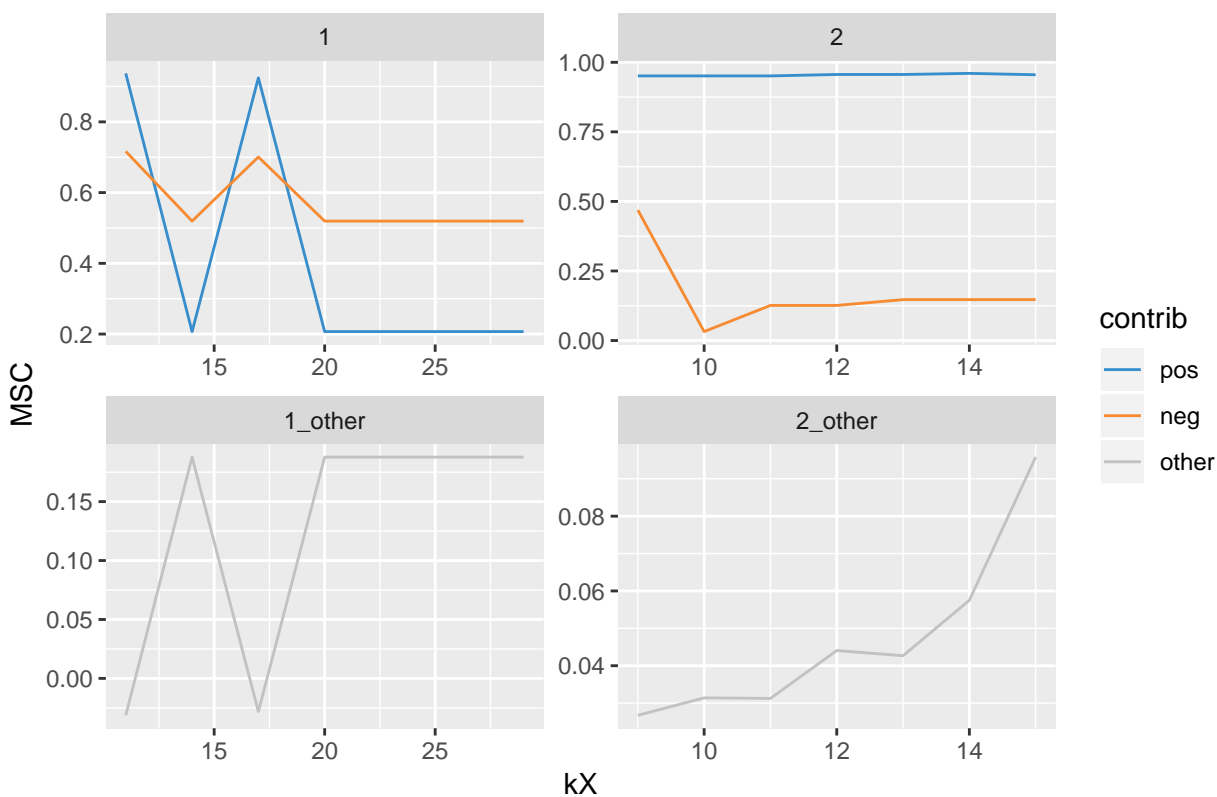
```
## [1] 0.4420804
```

sparse PCA Clustering

C-section

```
keepX = list(seq(11,29, 3), seq(9,15,1))
res.tune.spca.C <- tune.spca(X = spline.data.C, ncomp = 2, keepX = keepX)
tune.spca.choice.keepX(res.tune.spca.C, draw = T)
```

Tuning sPCA



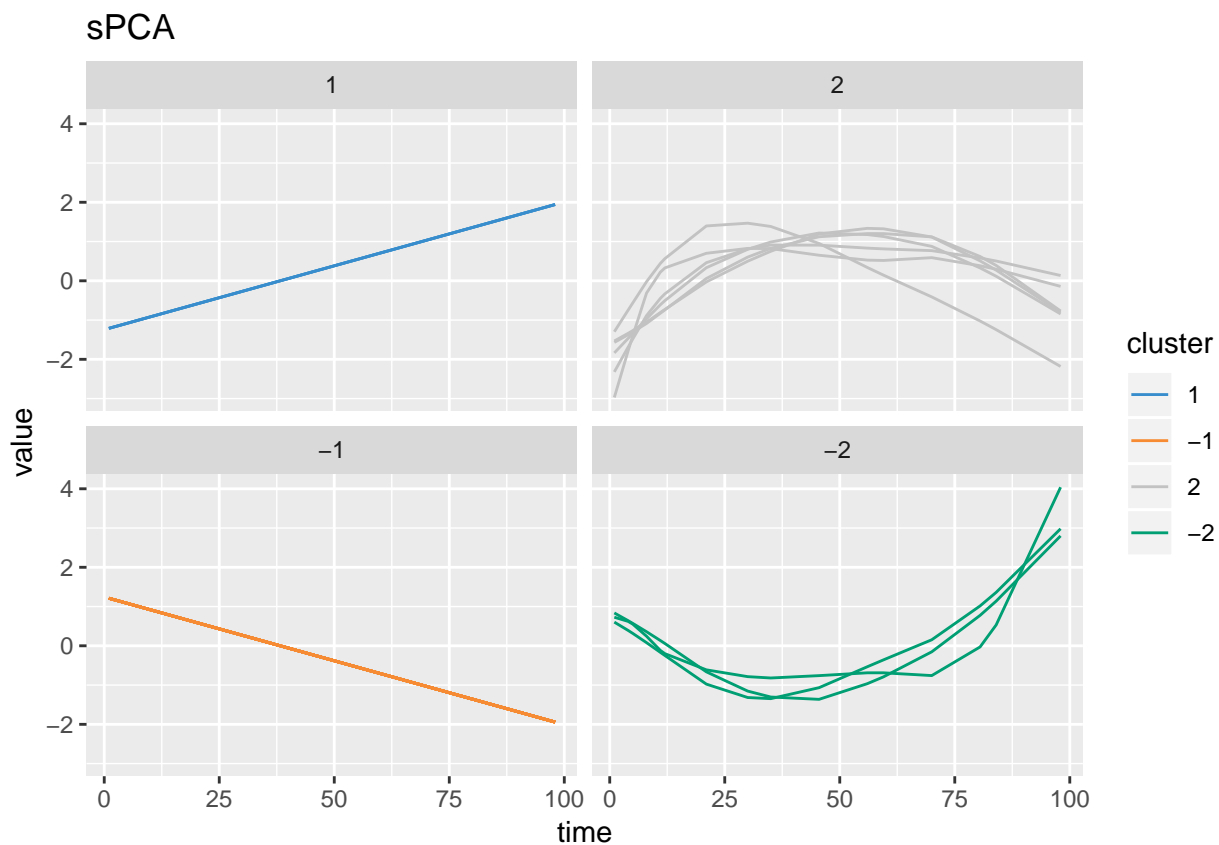
```
## [1] 11 NA
```

```
spca.res_f.C <- spca(spline.data.C, ncomp = 2, keepX = c(17,9))
```

```
wrapper.silhouette.spca(spline.data.C, keepX = c(17,9), ncomp = 2, scale = T, center=T)
```

```
## [1] 0.9527507
```

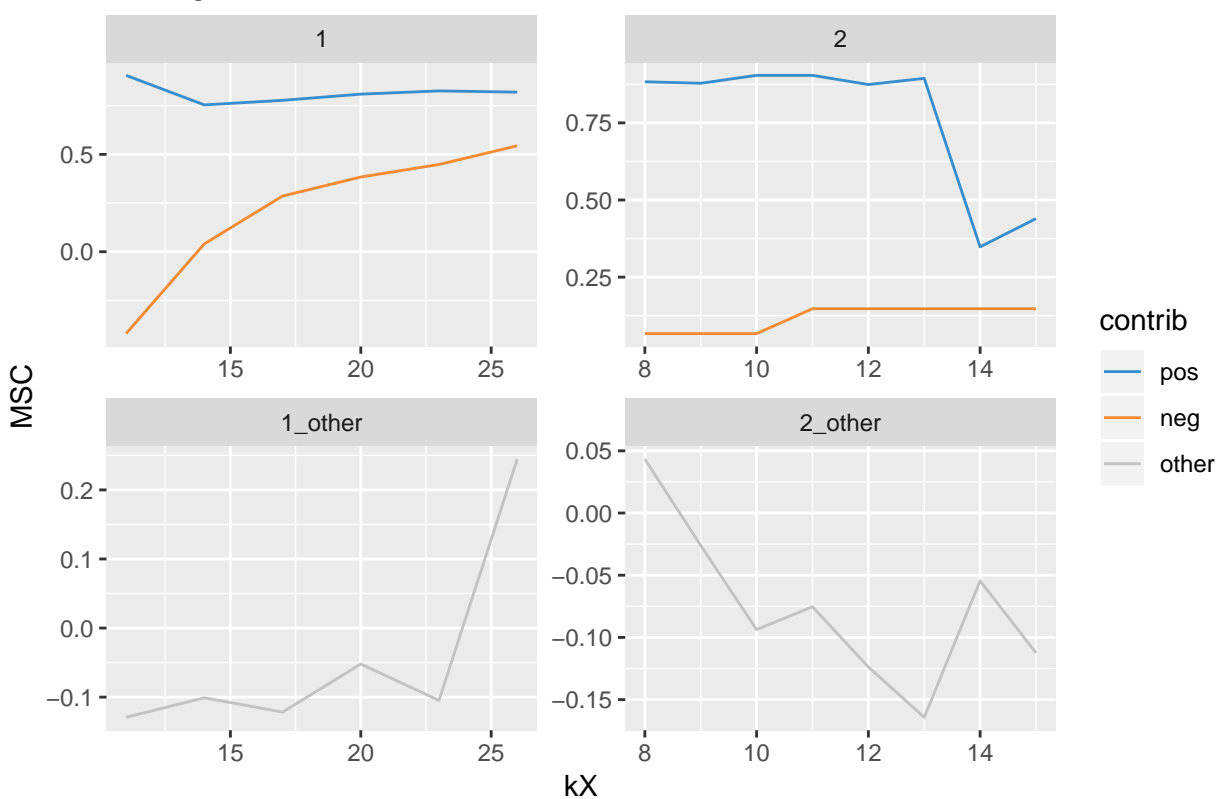
```
spca.plot(spca.res_f.C)
```

- remove lines :

```
keepX = list(seq(11,26, 3), seq(8,15,1))
res.tune.spca.0.C <- tune.spca(X = spline.0.C, ncomp = 2, keepX = keepX)
tune.spca.choice.keepX(res.tune.spca.0.C, draw = T)
```

Tuning sPCA



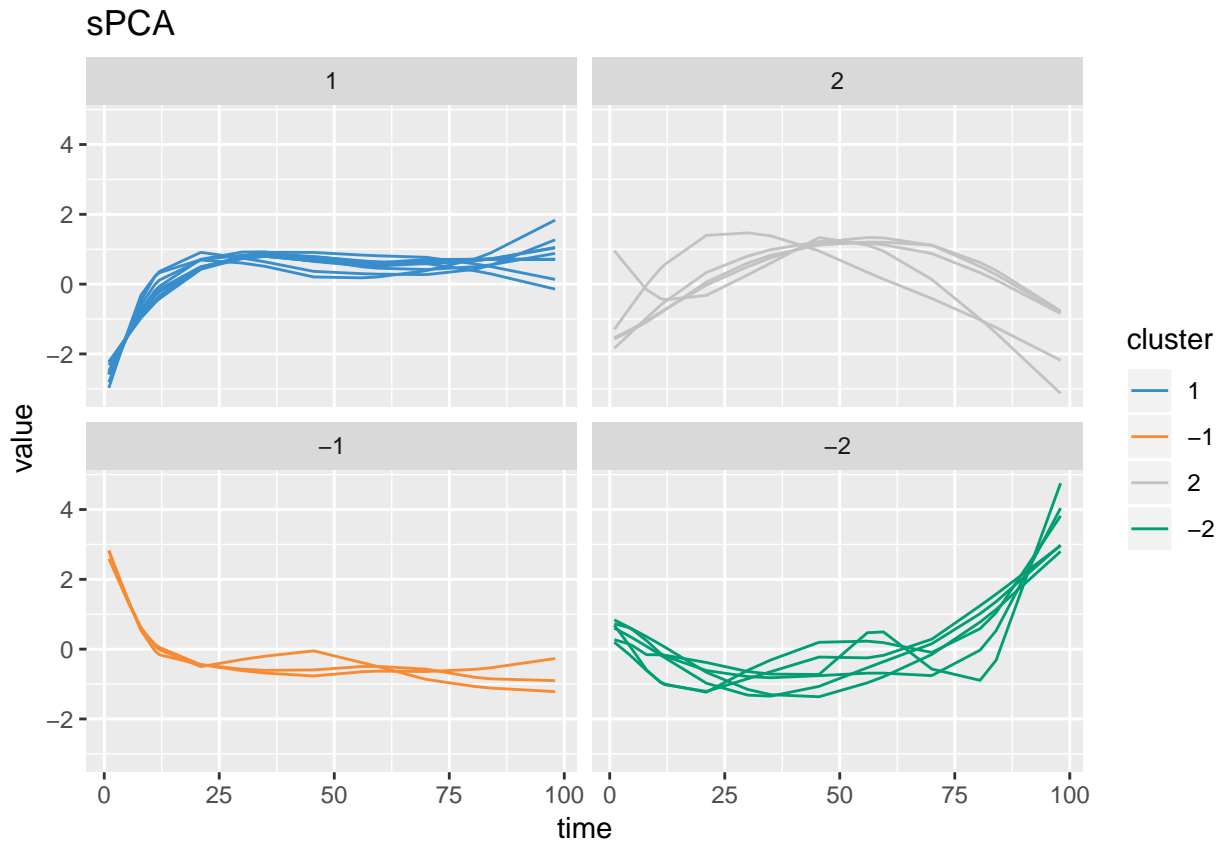
```
## [1] NA NA
```

```
spca.res_f <- spca(spline.0.C, ncomp = 2, keepX = c(13,13))
```

```
wrapper.silhouette.spca(spline.0.C, keepX = c(13,13), ncomp = 2, scale = T, center=T)
```

```
## [1] 0.7278328
```

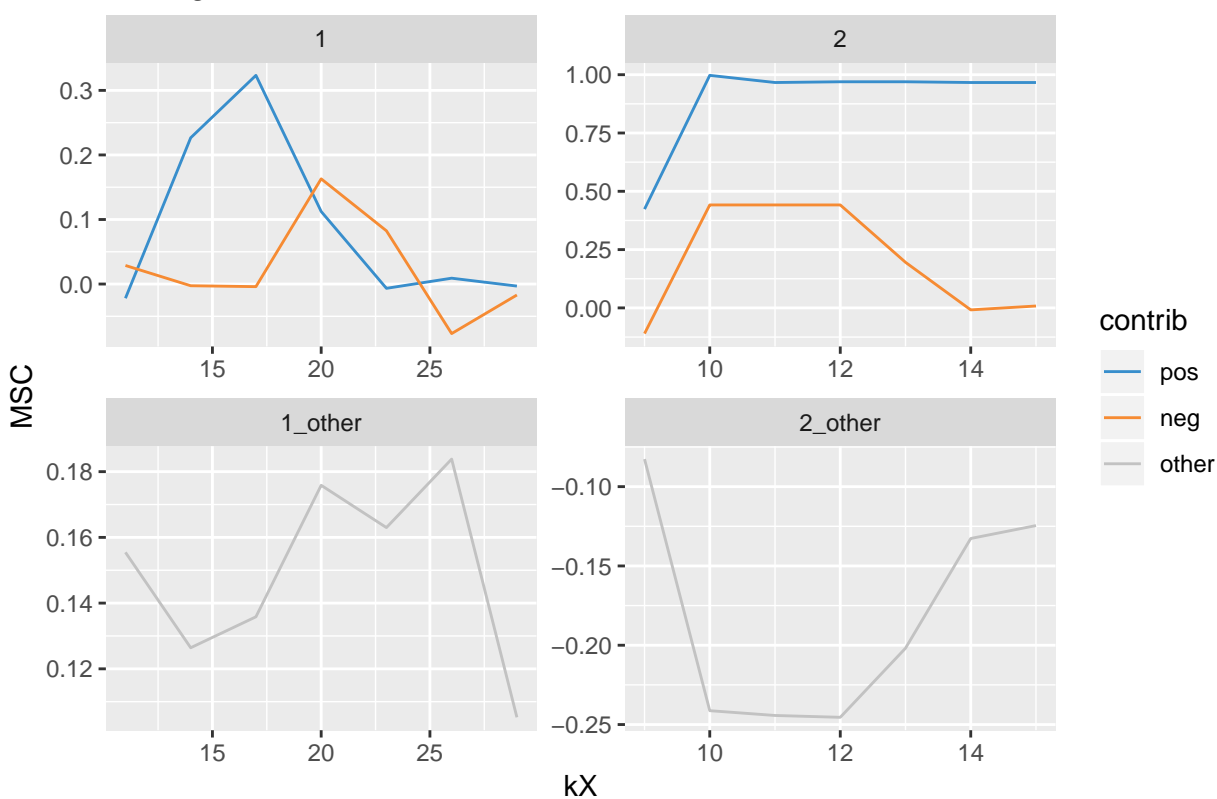
```
spca.plot(spca.res_f)
```



Vaginal

```
keepX = list(seq(11,29, 3), seq(9,15,1))
res.tune.spca.V <- tune.spca(X = spline.data.V, ncomp = 2, keepX = keepX)
tune.spca.choice.keepX(res.tune.spca.V, draw = T)
```

Tuning sPCA



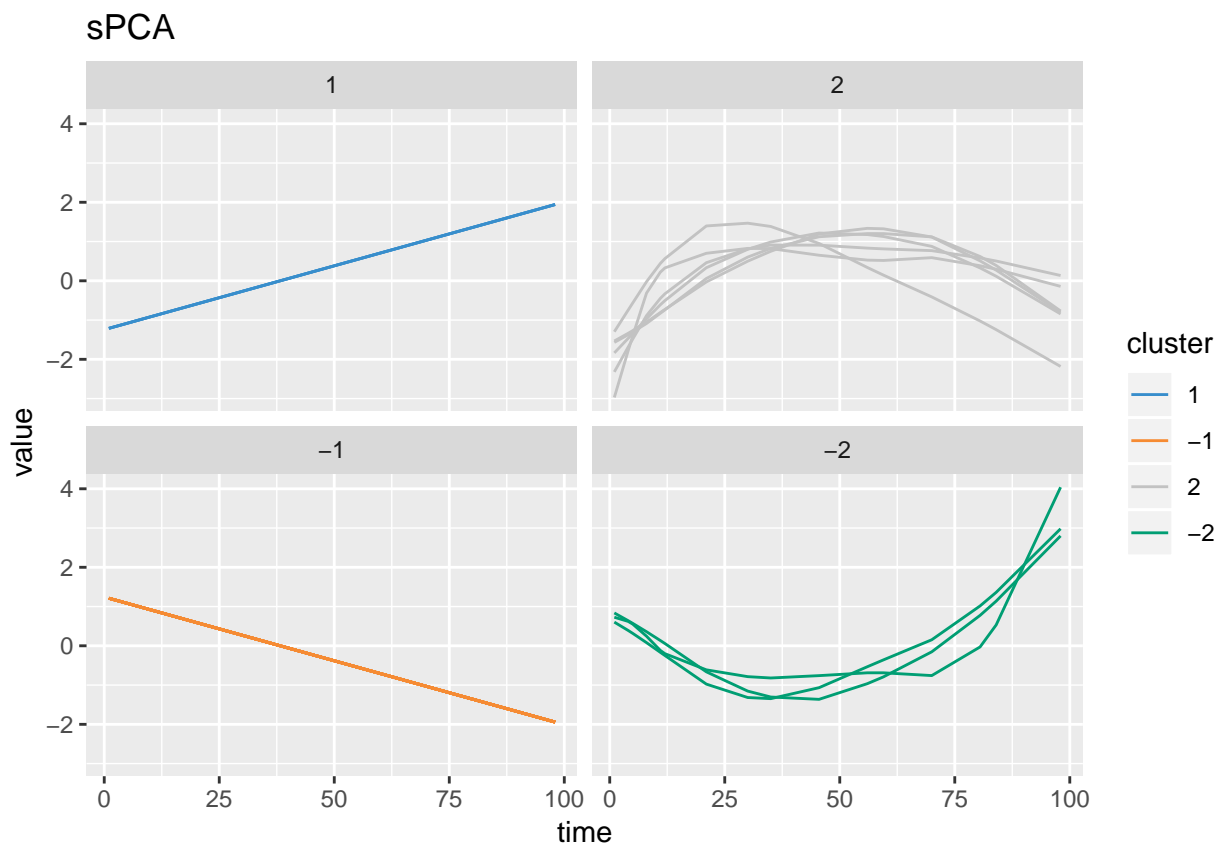
```
## [1] 17 10
```

```
spca.res_f.V <- spca(spline.data.V, ncomp = 2, keepX = c(17,10))
```

```
wrapper.silhouette.spca(spline.data.V, keepX = c(17,10), ncomp = 2, scale = T, center=T)
```

```
## [1] 0.8656932
```

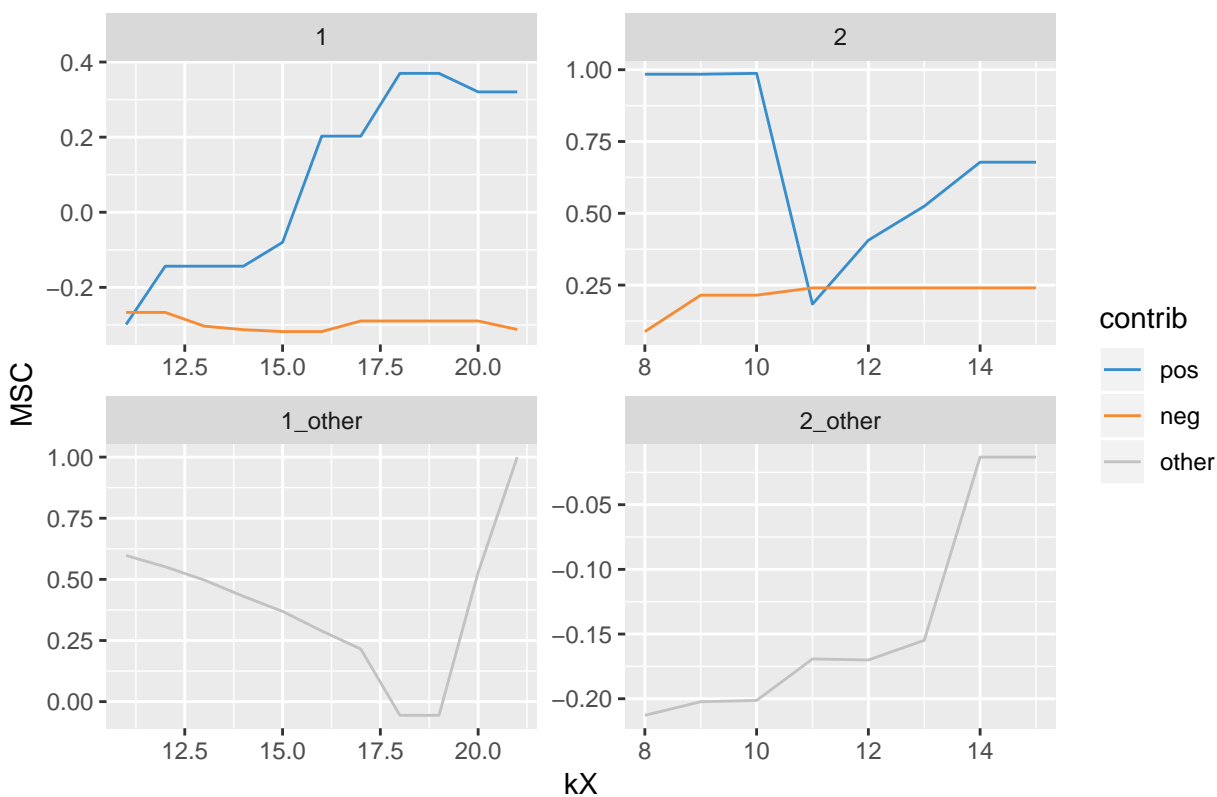
```
spca.plot(spca.res_f.C)
```



- remove lines :

```
keepX = list(seq(11,21,1), seq(8,15,1))
res.tune.spca.0.V <- tune.spca(X = spline.0.V, ncomp = 2, keepX = keepX)
tune.spca.choice.keepX(res.tune.spca.0.V, draw = T)
```

Tuning sPCA



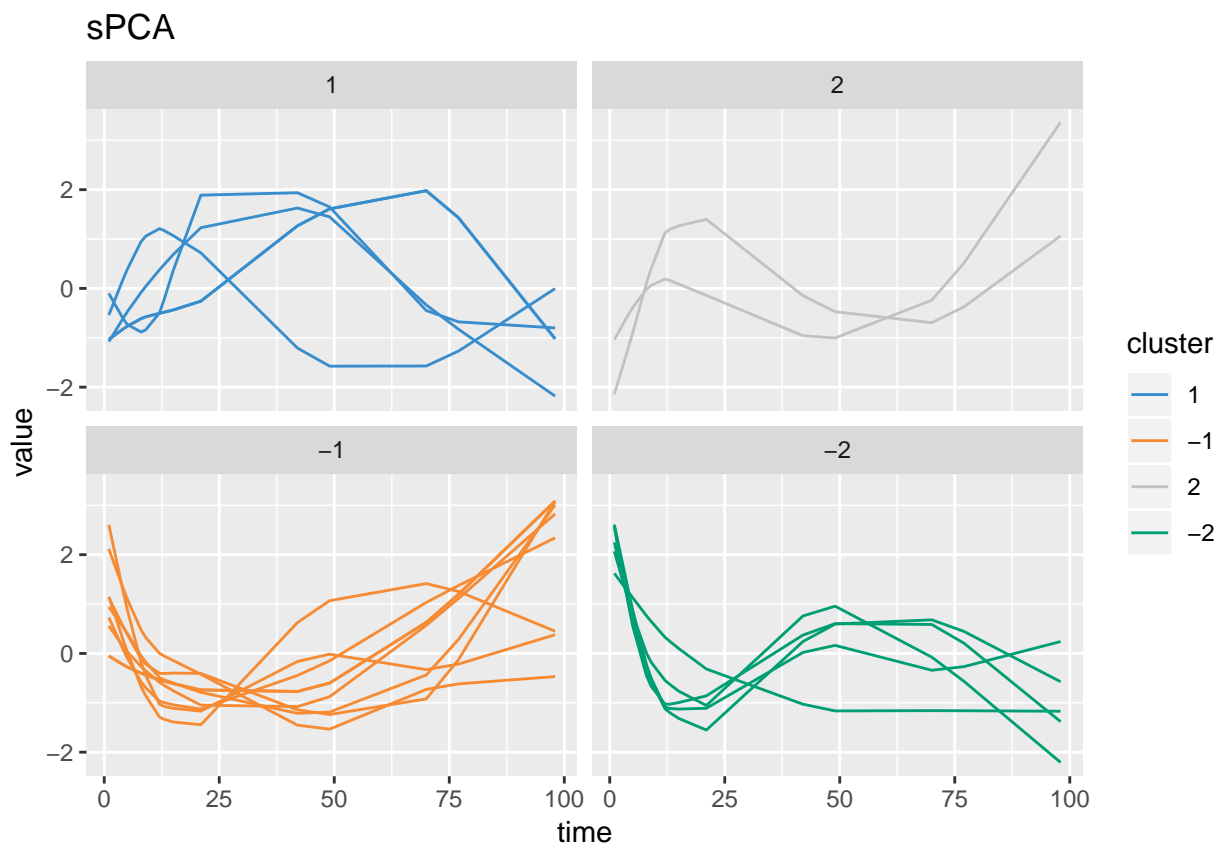
```
## [1] NA NA
```

```
spca.res_f <- spca(spline.0.V, ncomp = 2, keepX = c(18,10))
```

```
wrapper.silhouette.spca(spline.0.V, keepX = c(18,10), ncomp = 2, scale = T, center=T)
```

```
## [1] 0.5840128
```

```
spca.plot(spca.res_f)
```



Results

Silhouette coefficient :

	PCA	PCA w/o lines	sPCA	sPCA w/o lines
C-section	0.84	0.69	0.95	0.72
Vaginal	0.87	0.44	0.86	0.58
All	0.82	0.66	0.99	0.89

nb of molecules :

	0 (after filter)	1
C-section	42	29
Vaginale	68	22
All	54	47