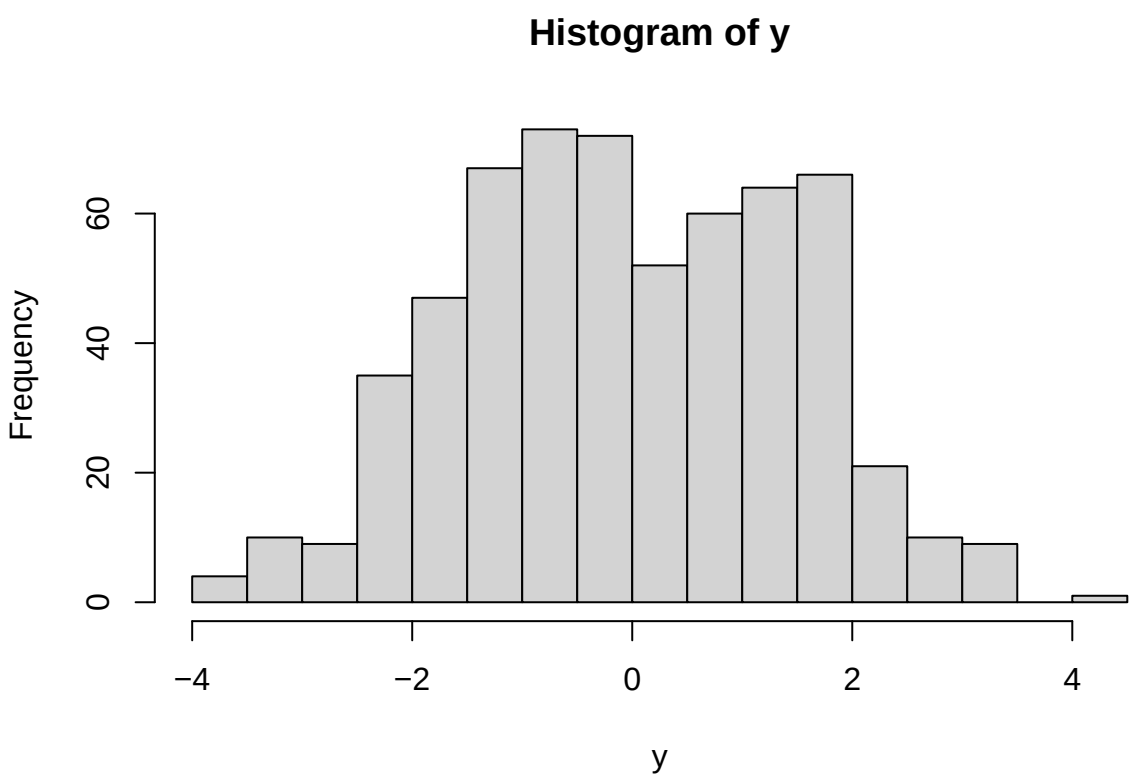
VI

Guanyu

2025-08-10

```
devtools::load_all()
library(WASABI.ext)
library(BNPmix)
library(mcclust)
library(salso)
library(superheat)
library(ggplot2)
```

```
set.seed(12345)
mu <- c(-1.1, 1.1)
prop <- c(0.5, 0.5)
n <- 600
components <- sample(1:2, size = n, replace = TRUE, prob = prop)
y <- rnorm(n, mean = mu[components], sd = 1)
hist(y, breaks = 20)
```



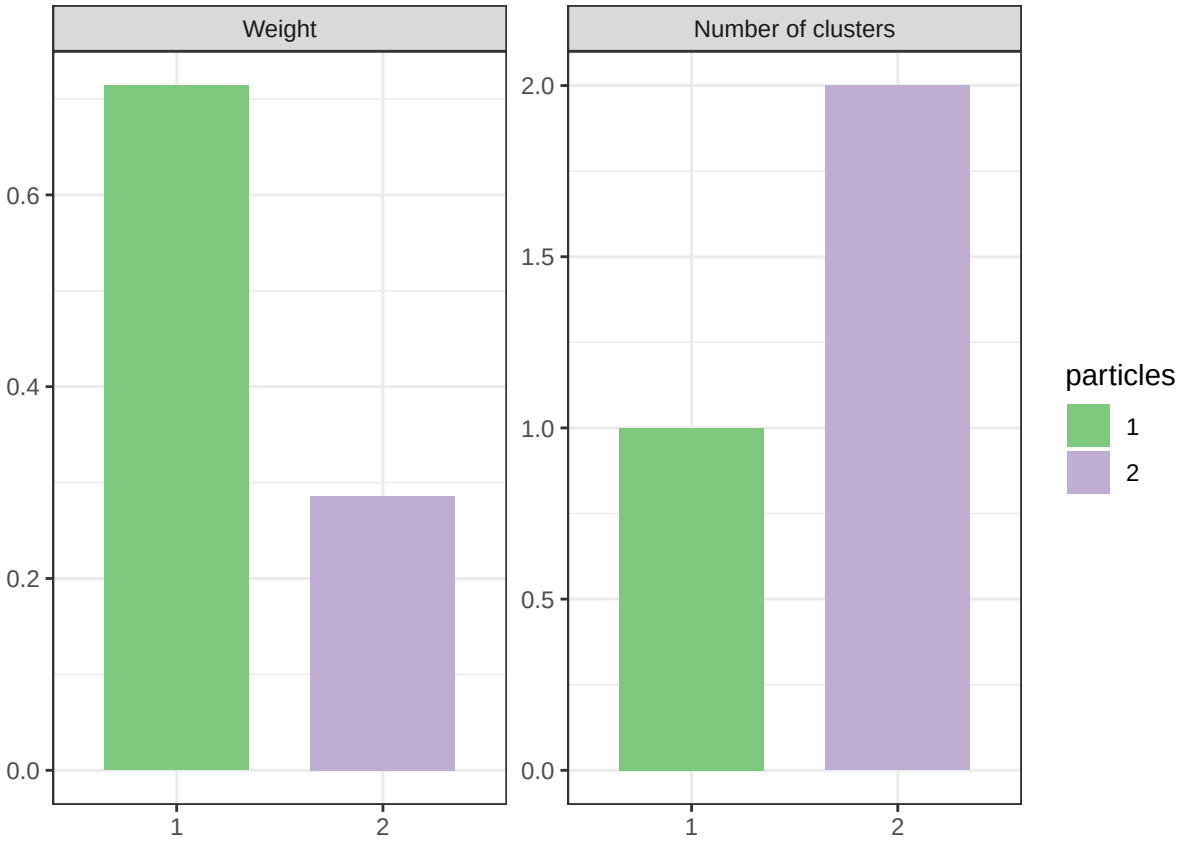

```
est_model <- BNPmix::PYdensity(y = y,
                               mcmc = list(niter = 15000,
```

```
                                        nburn = 5000,
                                        model = "LS",
                                        print_message = FALSE),
                                output = list(out_type = "FULL",
                                              out_param = TRUE))
cls.draw = est_model$clust
z_minVI <- salso::salso(cls.draw)
table(z_minVI)
```

```
## z_minVI
##   1
## 600
```

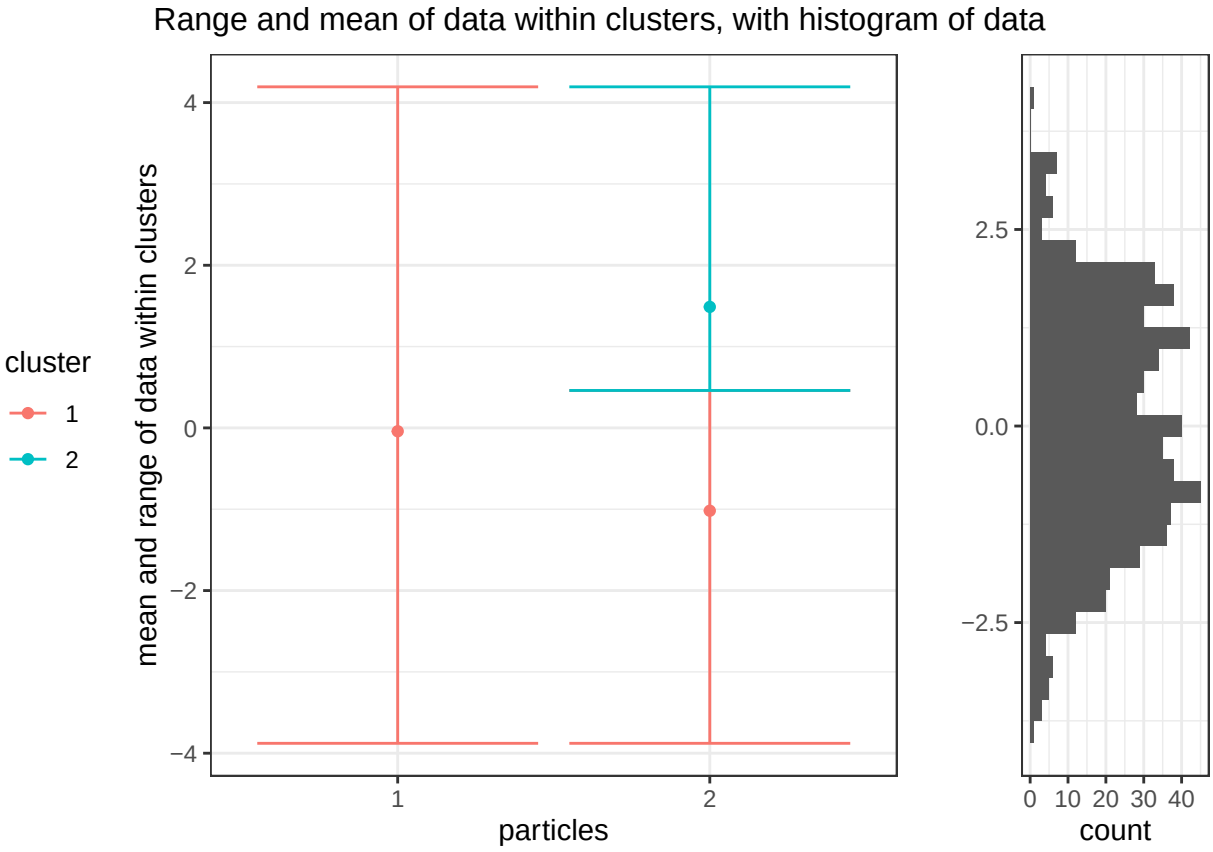
```
psm=mcclust::comp.psm(cls.draw+1)
```

```
out_WASABI <- WASABI(cls.draw, psm = psm, L = 2,
                     method.init = "topvi", method = "salso", loss = "VI")

out_WASABI_ms <- WASABI_multistart(cls.draw, psm = psm, L = 2,
                                   multi.start = 20, ncores = 4,
                                   mini.batch = 150,
                                   max.iter = 10, extra.iter = 4,
                                   method.init = "++", method = "salso",
                                   loss = "VI")
if(out_WASABI_ms$wass.dist < out_WASABI$wass.dist){
  out_WASABI <- out_WASABI_ms
}
ggsummary(out_WASABI)
```

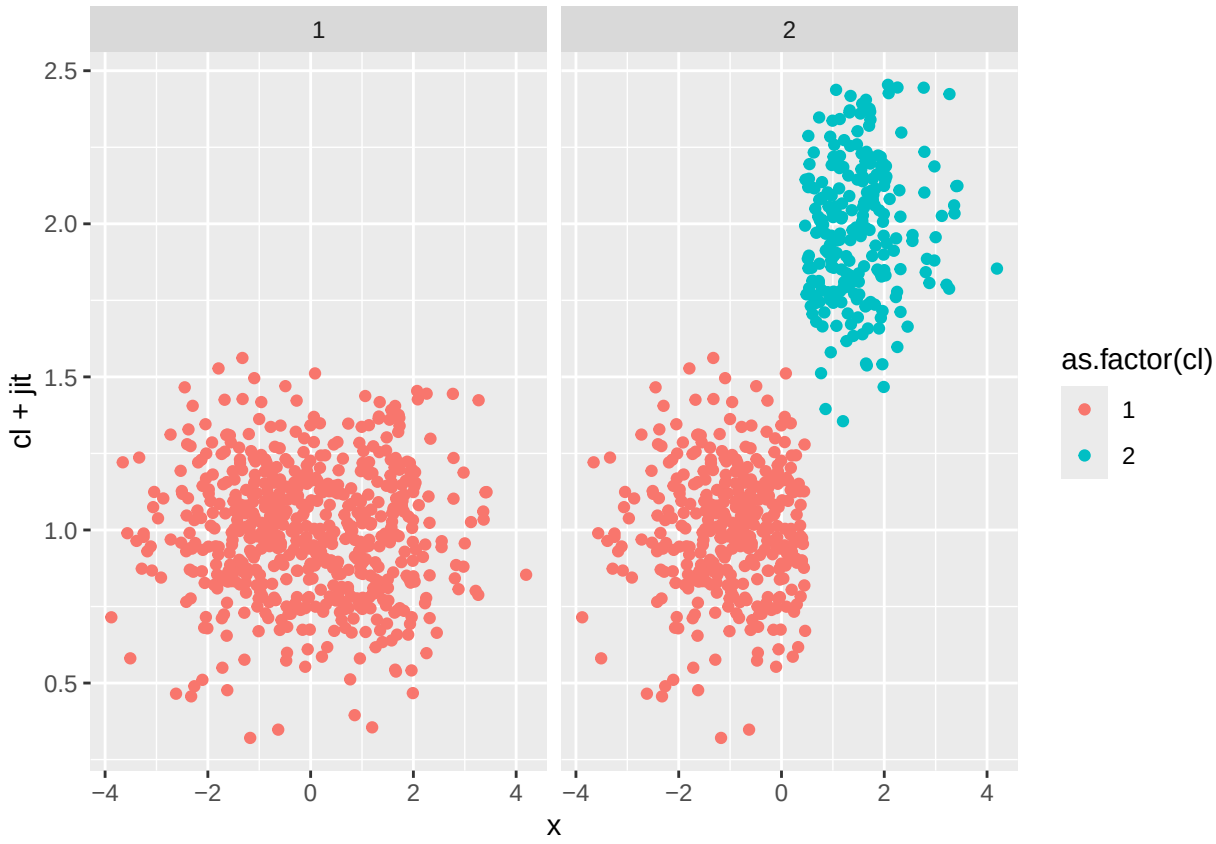


```
ggrange_hist(out_WASABI, y)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

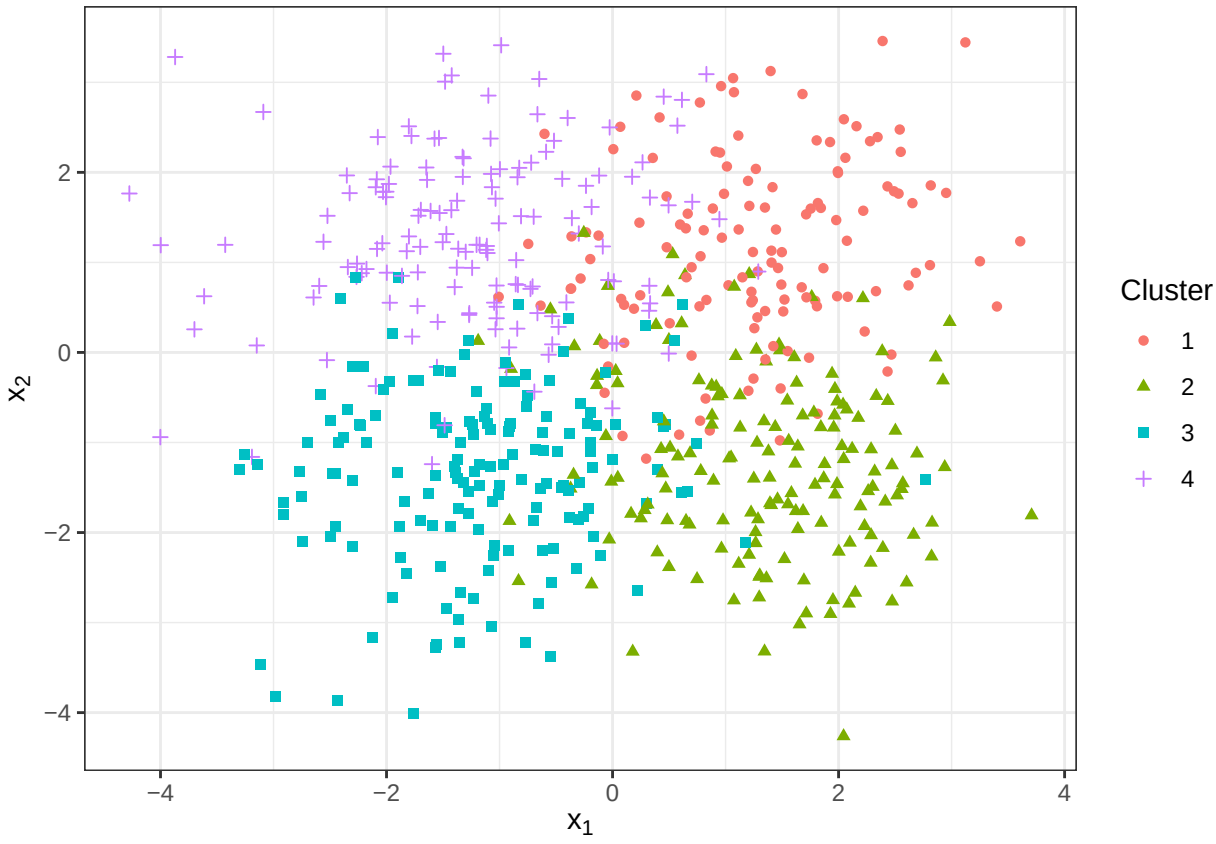



```
ggscatter_grid(out_WASABI, y)
```




```
m = 1.25
n = 600
p = 2
Kt = 4

set.seed(4321)

Y=matrix(rnorm(p*n),n,p)
usim=runif(n)
ind=ifelse(usim<1/4,1,ifelse(usim<1/2,2,ifelse(usim<3/4,3,4)))
Y[ind==1,] = Y[ind==1,] +m
Y[ind==2,1] = Y[ind==2,1] + m; Y[ind==2,2] = Y[ind==2,2] - m;
Y[ind==3,] = Y[ind==3,] -m
Y[ind==4,1] = Y[ind==4,1] - m; Y[ind==4,2] = Y[ind==4,2] + m;

cls.true = ind
```

```
library(ggplot2)
ggplot() +
  geom_point(aes(x = Y[,1],
                 y = Y[,2],
                 colour = as.factor(cls.true),
                 shape  = as.factor(cls.true))) +
  theme_bw() + guides(colour=guide_legend(title="Cluster"),
                      shape = guide_legend(title="Cluster")) +
  xlab(expression("x"[1])) + ylab(expression("x"[2]))
```




```
set.seed(4321)
### Parameters for DP mixture
alpha = 1
# using Fraley and Raftery recommendation
a_x=rep((p+2)/2,p)
khat = 4
b_x= rep(mean(apply(Y,2,var))/(khat^(2/p))/2,p)

### Parameters for MCMC function
S=10000
thin = 1
tot = S*thin
burnin= 5000

est_model <- BNPmix::PYdensity(y = Y,
                        mcmc = list(niter = burnin + tot,
                                    nburn = burnin,
                                    model = "DLS",
                                    hyper = FALSE
                                    ),
                        prior = list(
                          k0 = 0.1*rep(1,p),
                          a0 = a_x,
                          b0 = b_x,
                          strength = alpha,
                          discount = 0),
                        output = list(out_type = "FULL", out_param = TRUE))
```

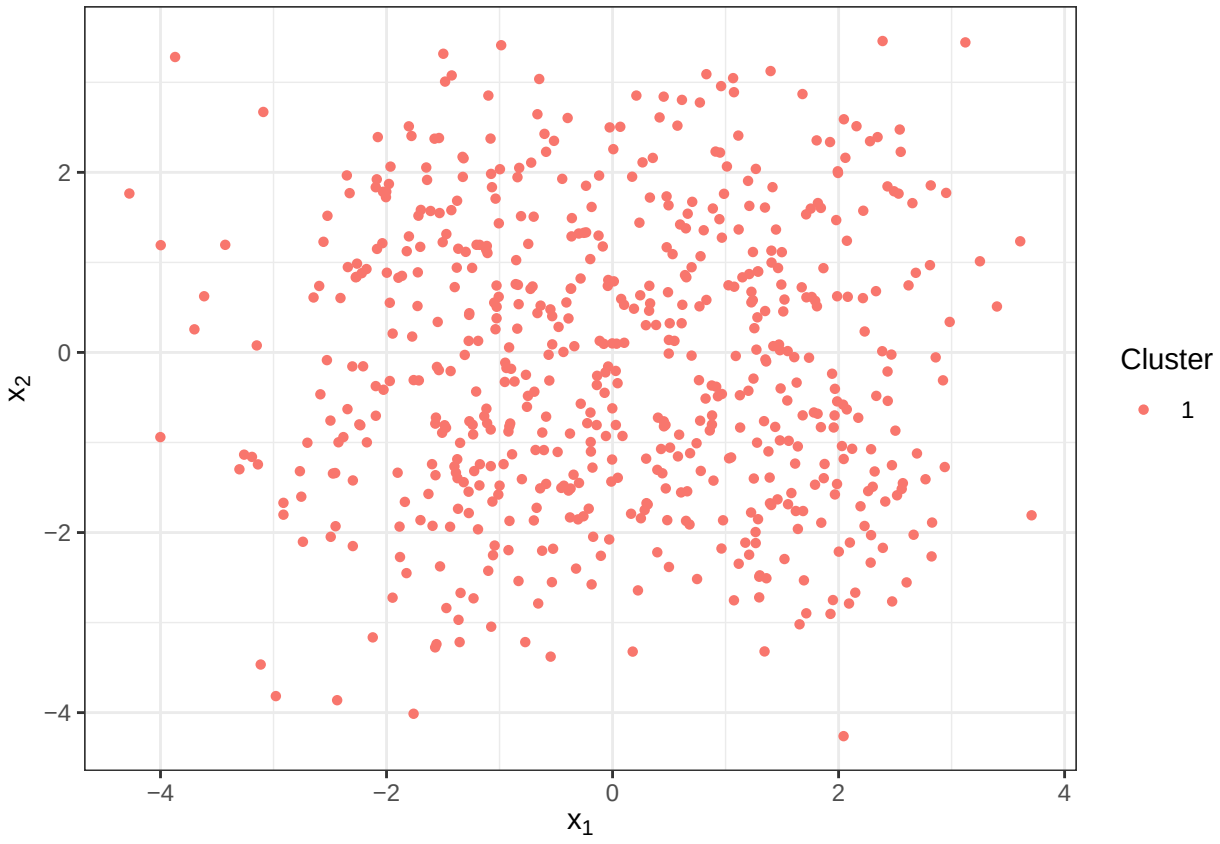
```
## Completed:   1500/15000 - in 0.527192 sec
## Completed:   3000/15000 - in 1.01564 sec
## Completed:   4500/15000 - in 1.49332 sec
## Completed:   6000/15000 - in 2.12149 sec
## Completed:   7500/15000 - in 2.73618 sec
## Completed:   9000/15000 - in 3.42126 sec
## Completed:   10500/15000 - in 4.12399 sec
## Completed:   12000/15000 - in 4.78706 sec
## Completed:   13500/15000 - in 5.42639 sec
## Completed:   15000/15000 - in 6.09012 sec
##
## Estimation done in 6.09016 seconds
```

```
#> Completed:   1100/11000 - in 0.478067 sec
#> Completed:   2200/11000 - in 0.939248 sec
#> Completed:   3300/11000 - in 1.39589 sec
#> Completed:   4400/11000 - in 1.84335 sec
#> Completed:   5500/11000 - in 2.32279 sec
#> Completed:   6600/11000 - in 2.80815 sec
#> Completed:   7700/11000 - in 3.22682 sec
#> Completed:   8800/11000 - in 3.7245 sec
#> Completed:   9900/11000 - in 4.19597 sec
#> Completed:   11000/11000 - in 4.83649 sec
#>
#> Estimation done in 4.83655 seconds
cls.draw = est_model$clust
psm=mcclust::comp.psm(cls.draw+1)
```

```
z_minVI <- salso::salso(cls.draw)
table(z_minVI)
```

```
## z_minVI
##   1
## 600
```

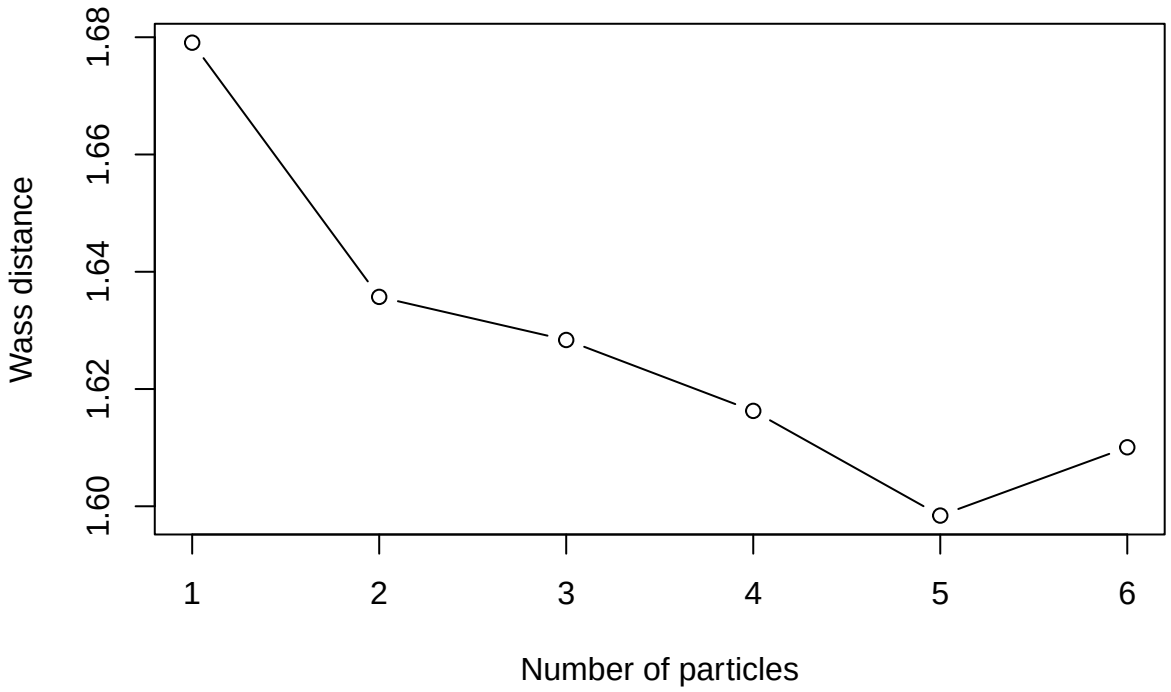
```
df = data.frame(x1 = Y[,1],
                x2 = Y[,2],
                Cluster = z_minVI)
df$Cluster = as.factor(df$Cluster)

ggplot(df)+
  geom_point(aes(x = x1, y = x2, color = Cluster, shape = Cluster)) +
  ylab(expression("x"[2]))+xlab(expression("x"[1]))+
  theme_bw()
```



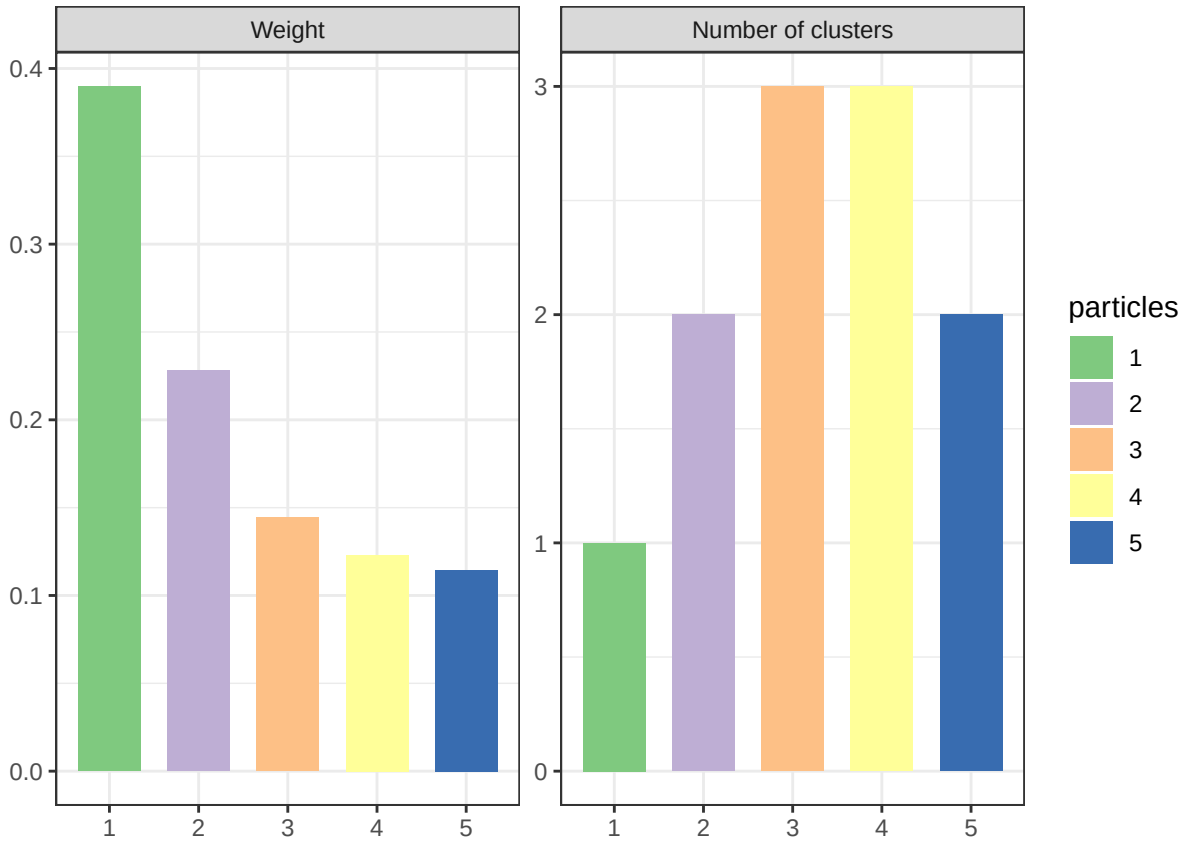

```
set.seed(123)
out_elbow <- elbow(cls.draw, L_max = 6, psm = psm,
                   multi.start = 1,
                   method.init = "topvi", method = "salso",
                   loss = "VI")
```

```
## Completed  1 / 6
## Completed  2 / 6
## Completed  3 / 6
## Completed  4 / 6
## Completed  5 / 6
## Completed  6 / 6
```

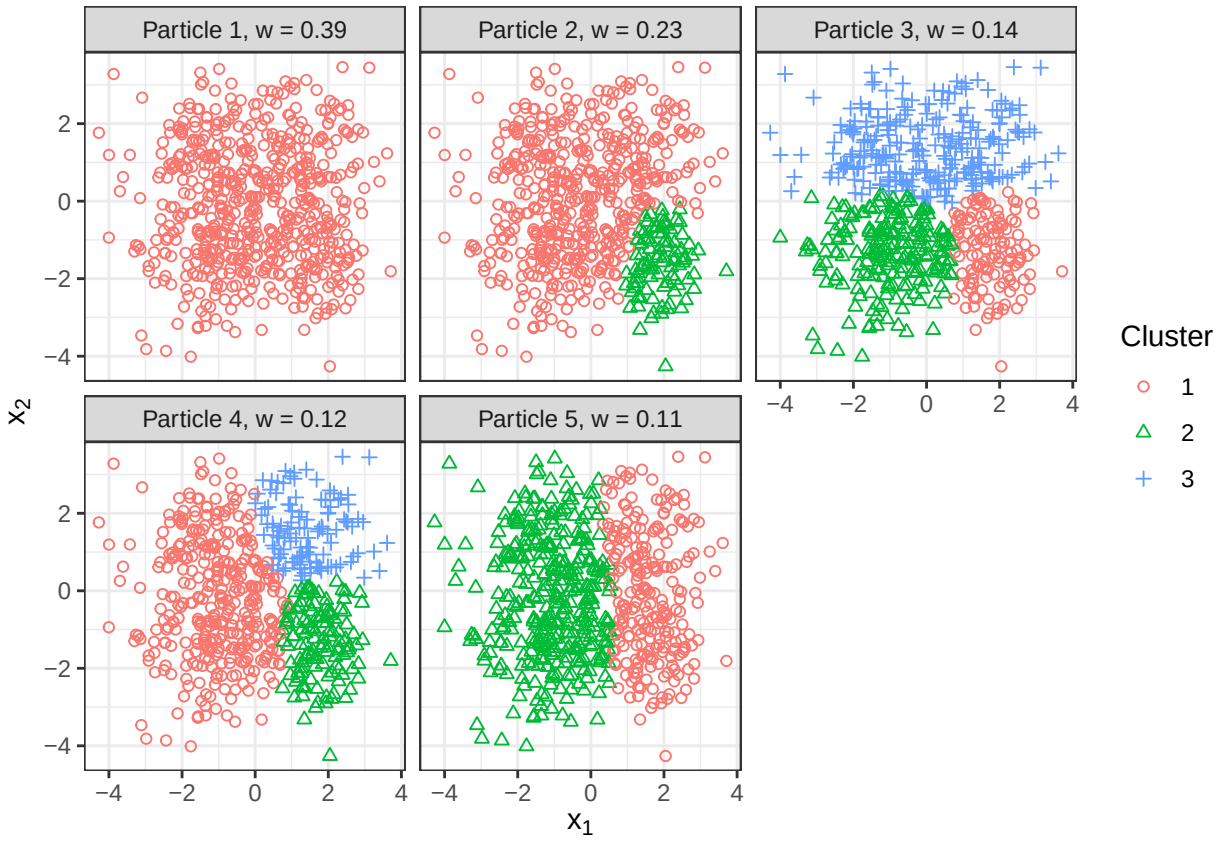
```
plot(out_elbow$wass_vec, type = "b", ylab = "Wass distance", xlab = "Number of particles")
```



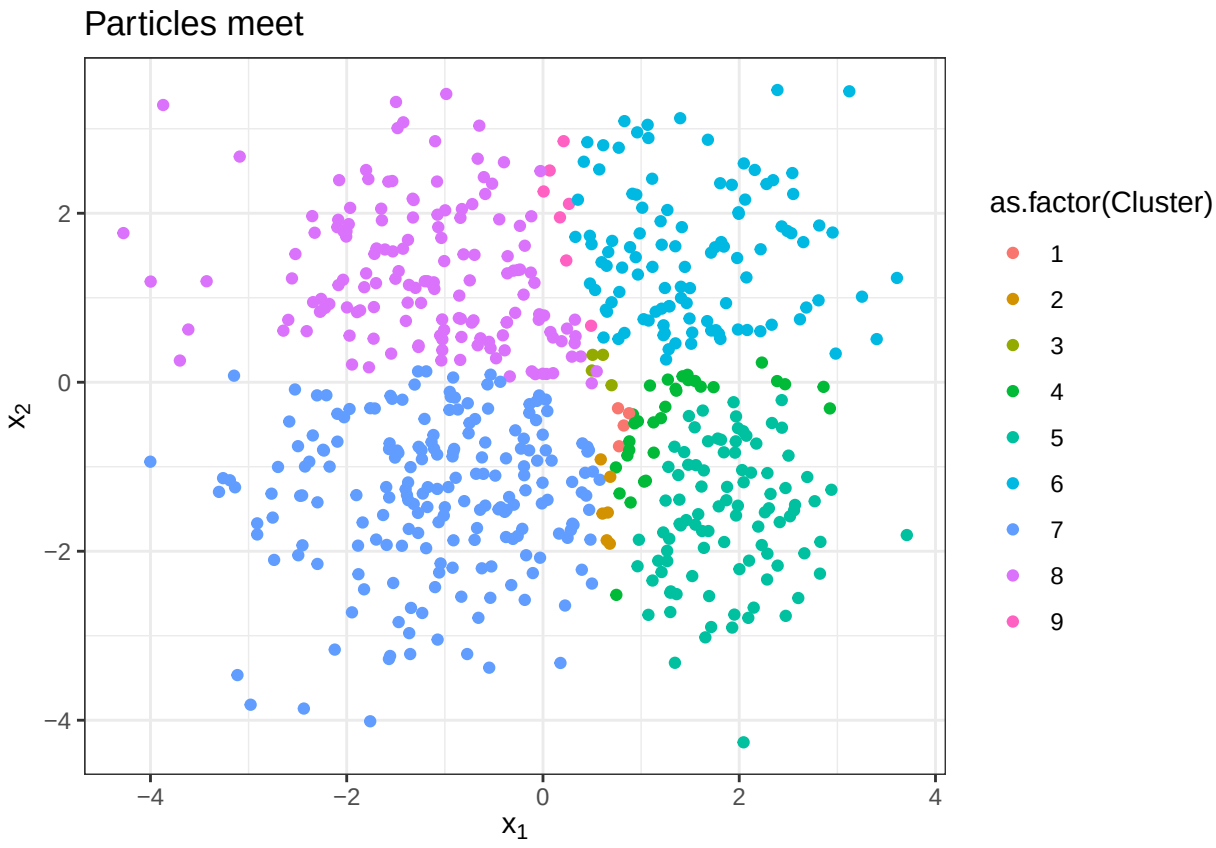

```
L = 5
output_WASABI <- out_elbow$output_list[[1]]
output_WASABI_mb = WASABI_multistart(cls.draw, psm,
                                     multi.start = 25, ncores = 4,
                                     method.init ="++", add_topvi = FALSE,
                                     method="salso", L=L,
                                     mini.batch = 200,
                                     max.iter= 10, suppress.comment=FALSE,
                                     swap_countone = TRUE,
                                     seed = 54321, loss = "VI")
if(output_WASABI_mb$wass.dist < output_WASABI$wass.dist){
  output_WASABI <- output_WASABI_mb
}
ggsummary(output_WASABI)
```




```
ggscatter_grid2d(output_WASABI, Y)
```




```
output_meet = cls.meet(output_WASABI$particles)
z_meet = output_meet$cls.m
df_tmp <- data.frame(x1 = Y[,1],
           x2 = Y[,2],
           Cluster = z_meet)
ggplot(df_tmp) +
  geom_point(aes(x = x1,y = x2,
                 color = as.factor(Cluster))) +
  theme_bw() +
  xlab(expression("x"[1])) + ylab(expression("x"[2]))+
  ggtitle("Particles meet")+theme(legend.box = "horizotal")
```


Particles meet
2
0
−2
−4
x2
−4
−2
0
2
4
x1
as.factor(Cluster)
1
2
3
4
5
6
7
8
9