311306435

2023-07-18

```
r = getOption("repos")
r["CRAN"] = "http://cran.us.r-project.org"
options(repos = r)

install.packages("DirichletReg")
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'DirichletReg' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'DirichletReg'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\DirichletReg\libs\x64\DirichletReg.dll
## to
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\DirichletReg\libs\x64\DirichletReg.dll:
## Permission denied
```

```
## Warning: restored 'DirichletReg'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

```
install.packages("scatterplot3d") # Install
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'scatterplot3d' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

```
install.packages("MASS")
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'MASS' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'MASS'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\MASS\libs\x64\MASS.dll
## to C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\MASS\libs\x64\MASS.dll:
## Permission denied
```

```
## Warning: restored 'MASS'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

```
install.packages("MCMCprecision")
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'MCMCprecision' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'MCMCprecision'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\MCMCprecision\libs\x64\MCMCprecision.dll
## to
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\MCMCprecision\libs\x64\MCMCprecision.dll:
## Permission denied
```

```
## Warning: restored 'MCMCprecision'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

```
install.packages("ggpubr")
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'ggpubr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

```
library(gridExtra)
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggpubr':
##
##     get_legend
```

```
require(MCMCprecision)
```

```
## Loading required package: MCMCprecision
```

```
library(Formula)
library(DirichletReg)
```

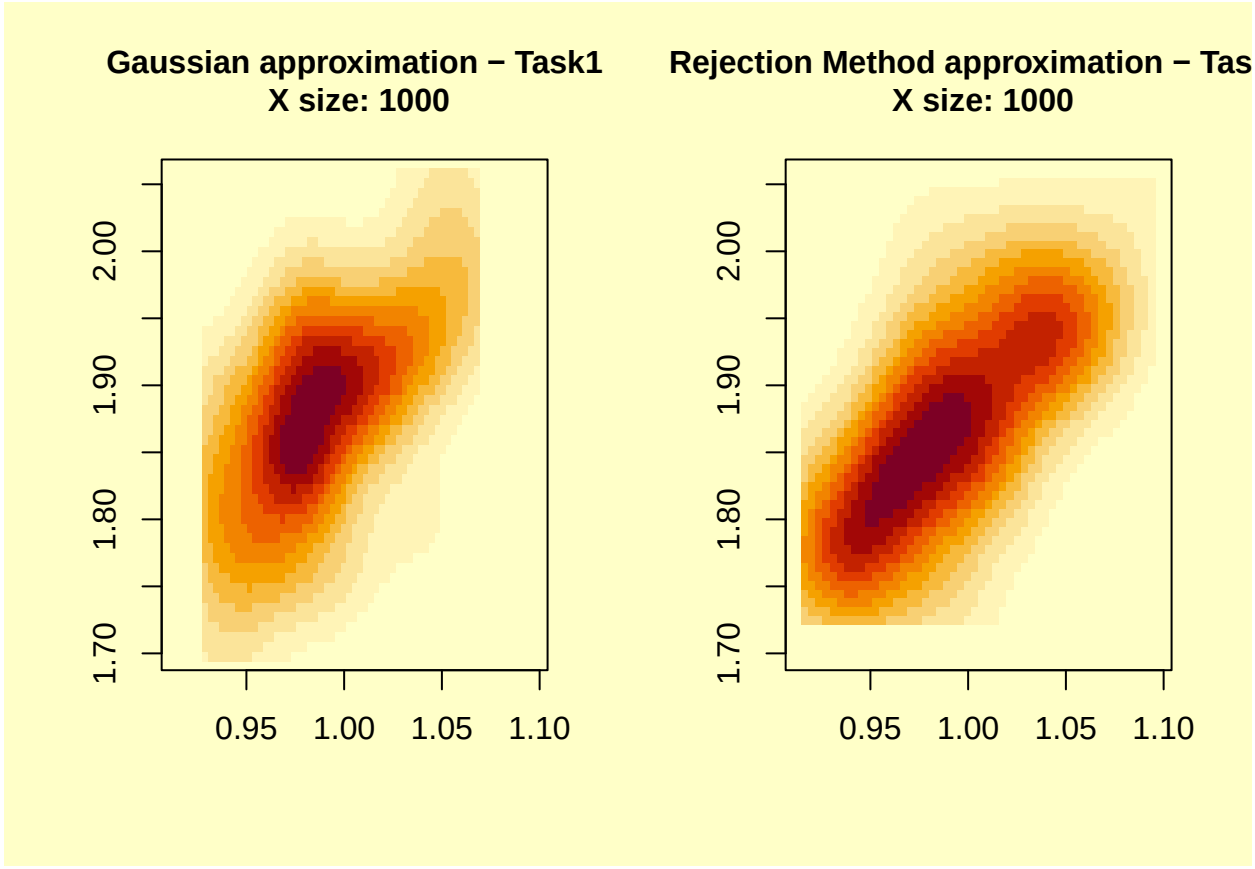
```
##
## Attaching package: 'DirichletReg'
```

```
## The following object is masked from 'package:MCMCprecision':
##
##     rdirichlet
```

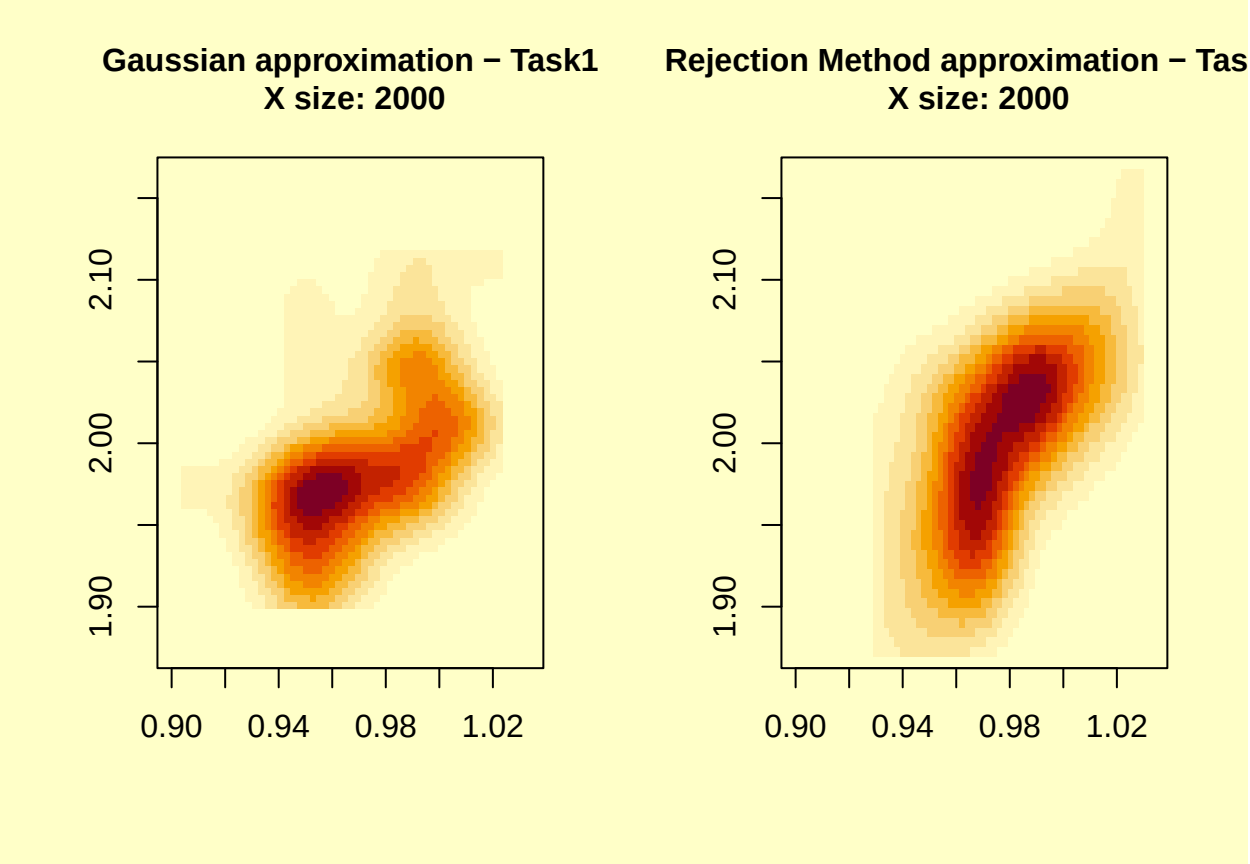
```r
library("scatterplot3d") # load
library(MASS)
```

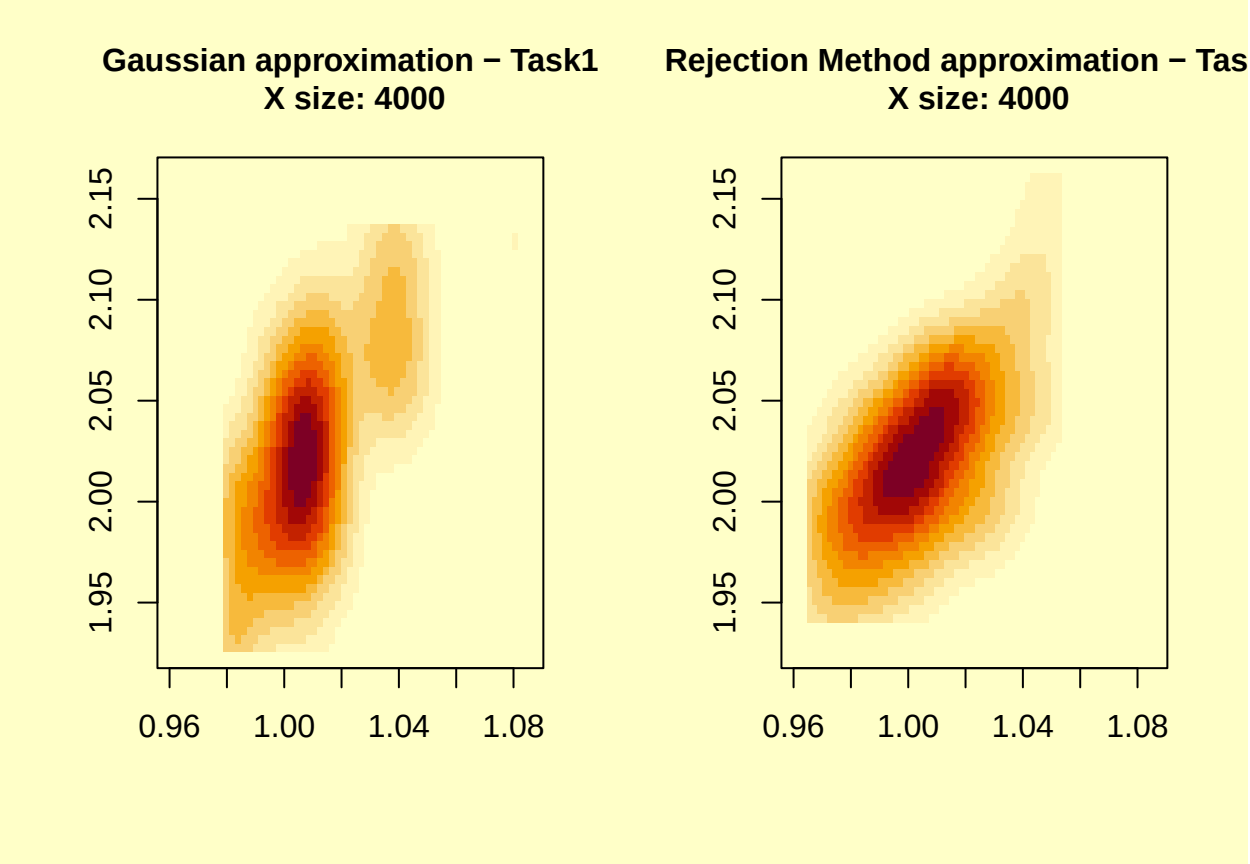
Question 1

Question 2



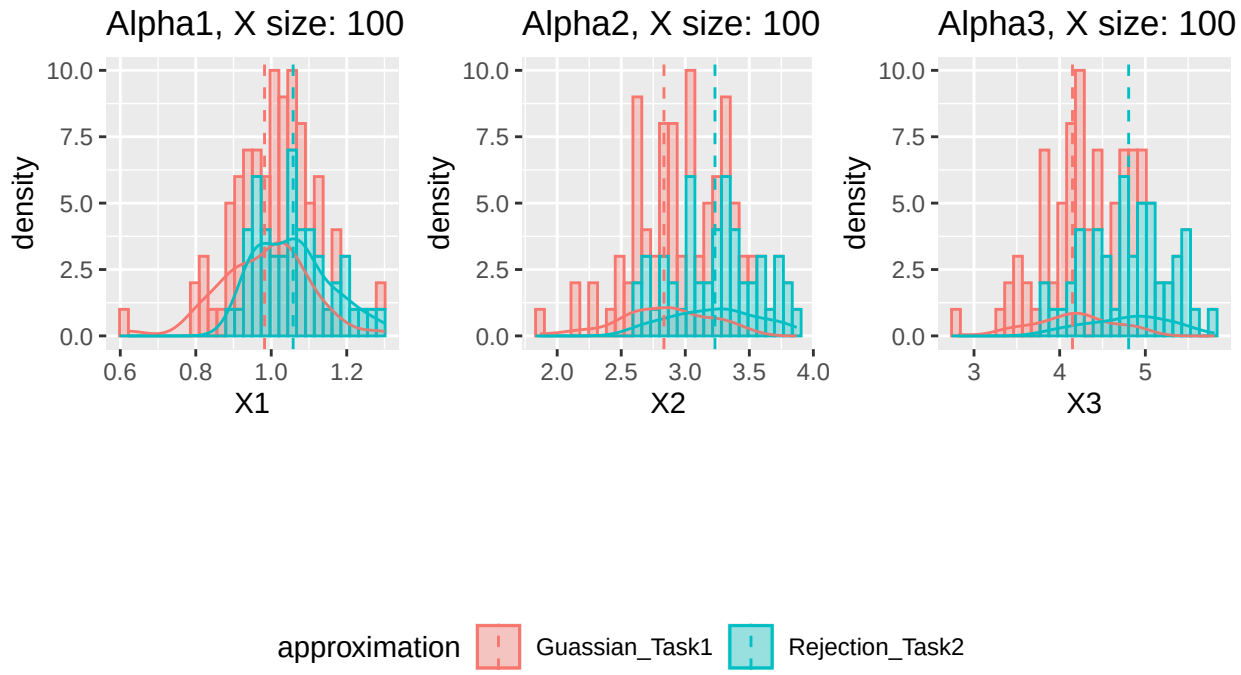
Question 3


Gaussian approximation − Task1
X size: 2000
Rejection Method approximation − Tas
X size: 2000
2.10
2.00
1.90
0.90
0.94
0.98
1.02

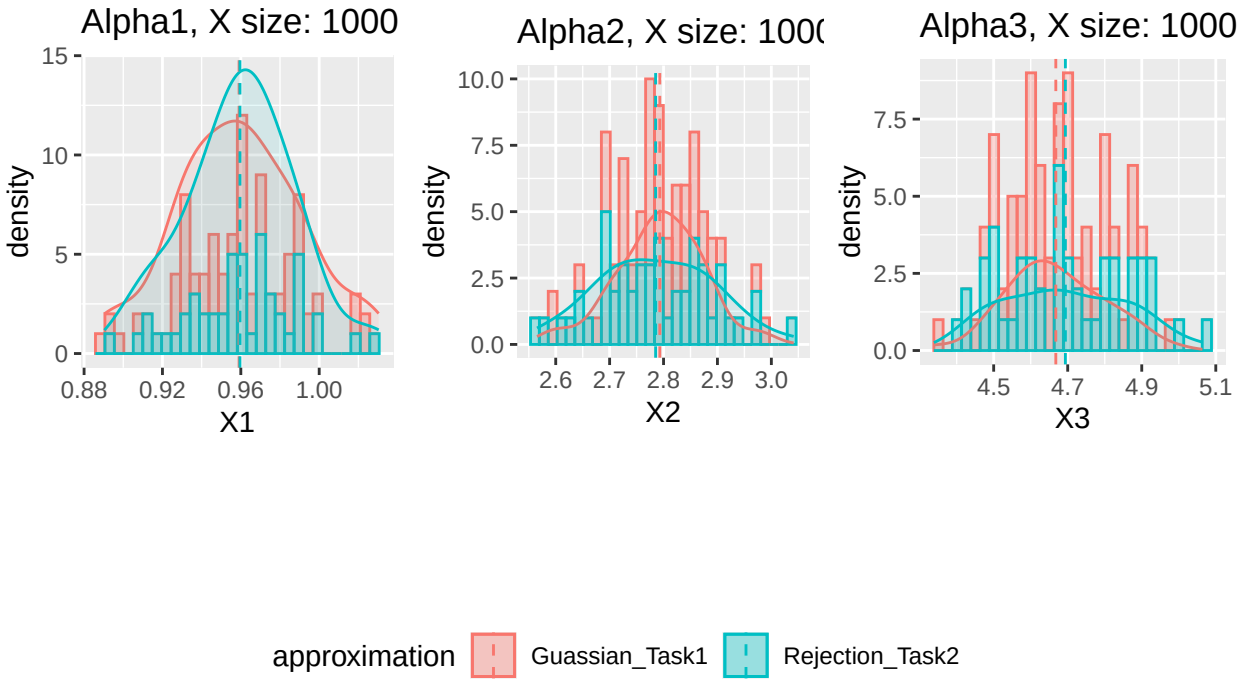


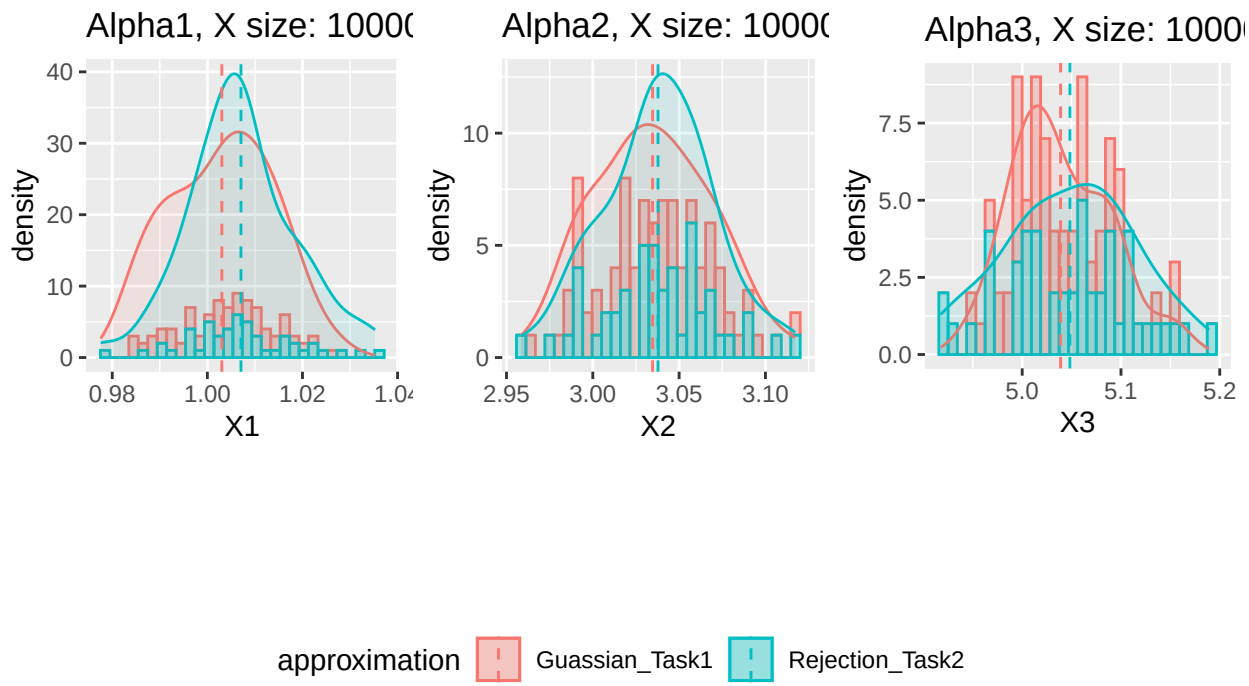

```
## Warning: package 'ggplot2' is in use and will not be installed

## Warning: package 'gridExtra' is in use and will not be installed

## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'plyr' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'plyr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\plyr\libs\x64\plyr.dll
## to C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\plyr\libs\x64\plyr.dll:
## Permission denied

## Warning: restored 'plyr'

##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages

##
## Attaching package: 'plyr'

## The following object is masked from 'package:ggpubr':
##
##     mutate

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

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



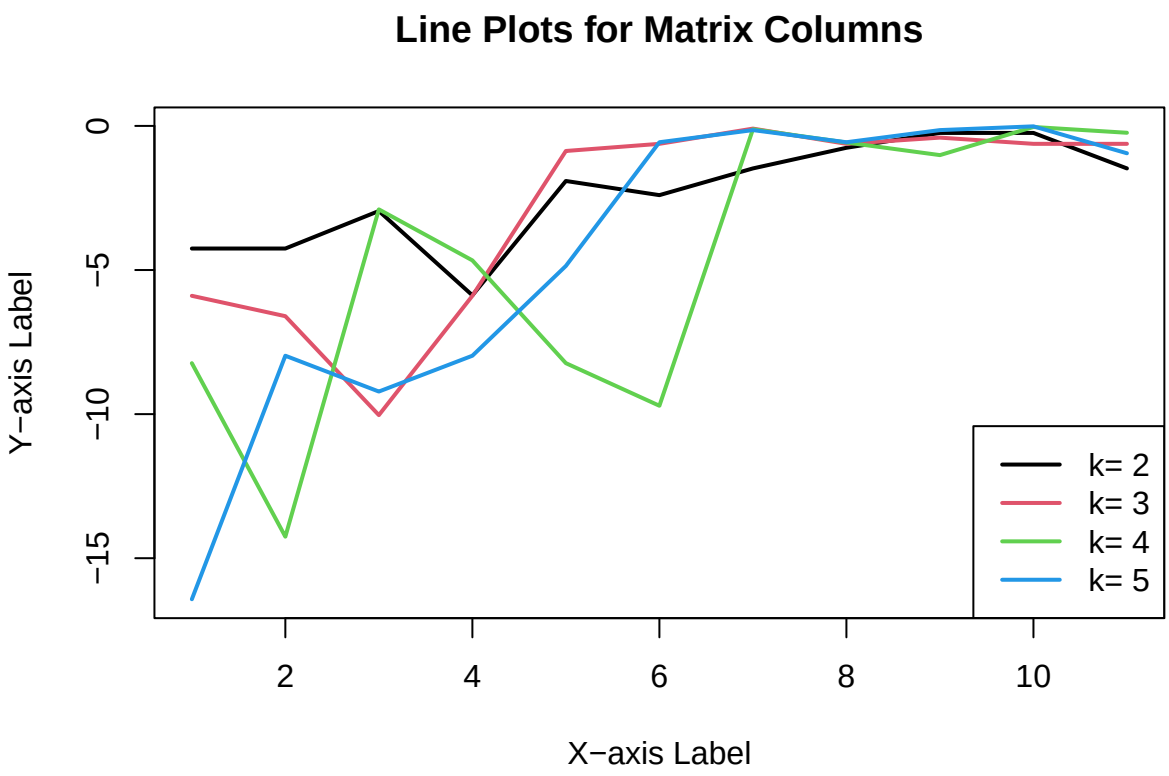

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Alpha1, X size: 1000
density
X1
Alpha2, X size: 100(
density
X2
Alpha3, X size: 1000
density
X3
approximation
Guassian_Task1
Rejection_Task2




```
## [1] 2
## [1] "x_size 64"
## [1] "x_size 128"
## [1] "x_size 256"
## [1] "x_size 512"
## [1] "x_size 1024"
## [1] "x_size 2048"
## [1] "x_size 4096"
## [1] "x_size 8192"
## [1] "x_size 16384"
## [1] "x_size 32768"
## [1] "x_size 65536"
## [1] 3
## [1] "x_size 64"
## [1] "x_size 128"
## [1] "x_size 256"
## [1] "x_size 512"
## [1] "x_size 1024"
## [1] "x_size 2048"
## [1] "x_size 4096"
## [1] "x_size 8192"
## [1] "x_size 16384"
## [1] "x_size 32768"
## [1] "x_size 65536"
## [1] 4
## [1] "x_size 64"
```

```
## [1] "x_size 128"
## [1] "x_size 256"
## [1] "x_size 512"
## [1] "x_size 1024"
## [1] "x_size 2048"
## [1] "x_size 4096"
## [1] "x_size 8192"
## [1] "x_size 16384"
## [1] "x_size 32768"
## [1] "x_size 65536"
## [1] 5
## [1] "x_size 64"
## [1] "x_size 128"
## [1] "x_size 256"
## [1] "x_size 512"
## [1] "x_size 1024"
## [1] "x_size 2048"
## [1] "x_size 4096"
## [1] "x_size 8192"
## [1] "x_size 16384"
## [1] "x_size 32768"
## [1] "x_size 65536"
##       results_per_k results_per_k results_per_k results_per_k
##  [1,]  5.569063e-05  1.275006e-06  5.862608e-09  3.770852e-17
##  [2,]  5.569063e-05  2.500012e-07  5.590327e-15  1.062496e-08
##  [3,]  1.115802e-03  9.236091e-11  1.270814e-03  6.078720e-10
##  [4,]  1.325967e-06  1.275006e-06  2.156358e-05  1.062496e-08
##  [5,]  1.229861e-02  1.350035e-01  5.862608e-09  1.386789e-05
##  [6,]  3.967294e-03  2.390730e-01  1.949205e-10  2.719136e-01
##  [7,]  3.354166e-02  8.079632e-01  7.659315e-01  7.166468e-01
##  [8,]  1.745330e-01  2.390730e-01  2.656871e-01  2.719136e-01
##  [9,]  5.713360e-01  3.929450e-01  9.707484e-02  7.166468e-01
## [10,]  5.713360e-01  2.390730e-01  9.188052e-01  9.667464e-01
## [11,]  3.354166e-02  2.390730e-01  5.786001e-01  1.123852e-01
```



Question 4

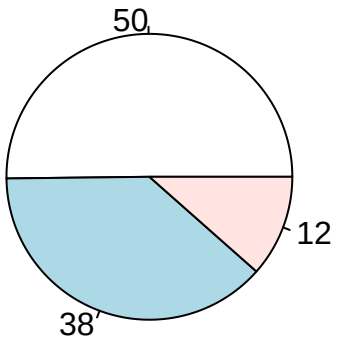
```
Q3_non_zeros <- rowSums(Q3 >= 0.001)
Q4_non_zeros <- rowSums(Q4 >= 0.001)
q3_labels <- round(table(Q3_non_zeros)/length(Q3_non_zeros)*100)
q4_labels <- round(table(Q4_non_zeros)/length(Q4_non_zeros)*100)
print(q3_labels)
```

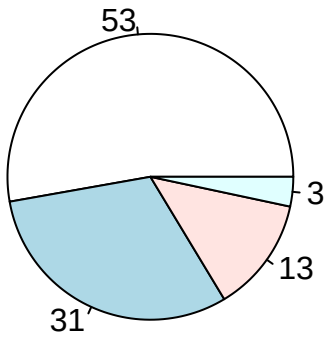
```
## Q3_non_zeros
##  1  2  3
## 50 38 12
```

```
print(q4_labels)
```

```
## Q4_non_zeros
##  1  2  3  4
## 53 31 13  3
```

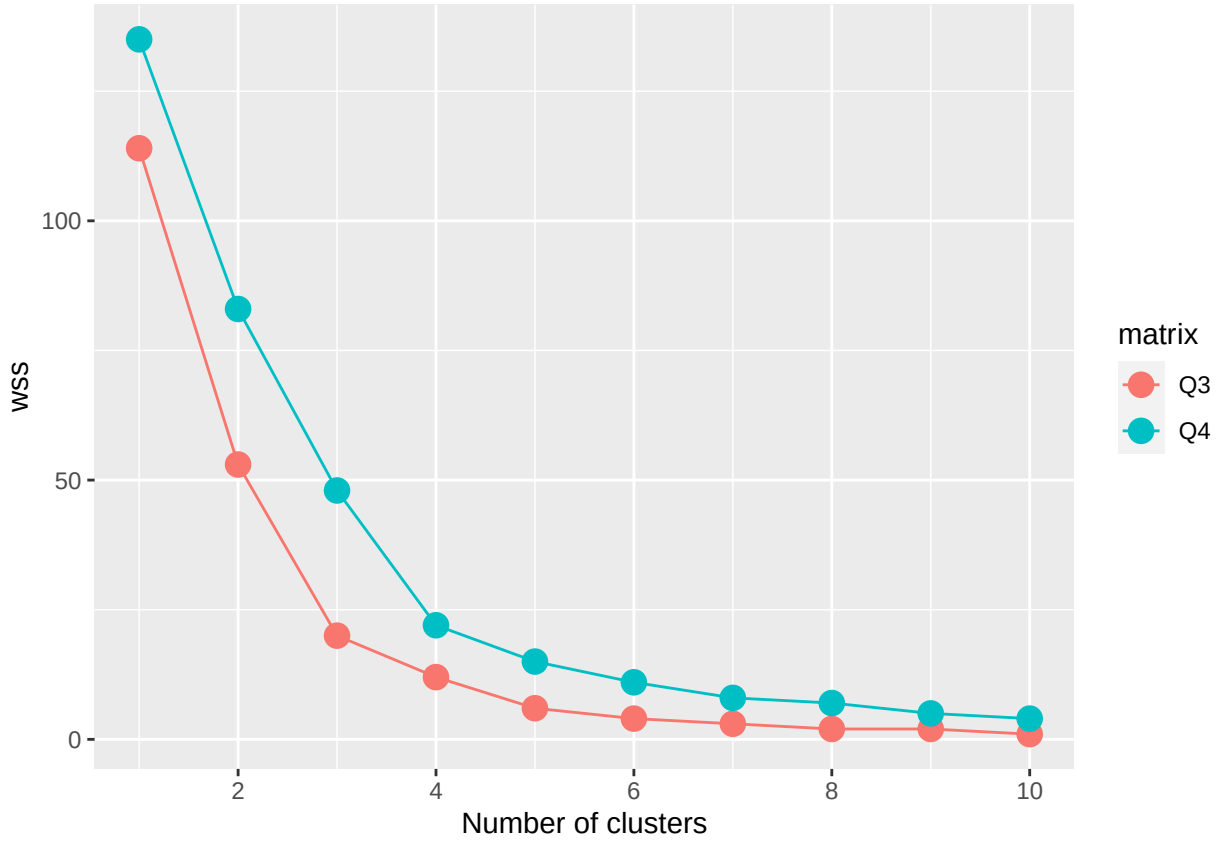
```
par(mfrow=c(1,2))
pie(table(Q3_non_zeros), main="Q3 non zero elements count", labels = q3_labels)
pie(table(Q4_non_zeros), main="Q4 non zero elements count", labels = q4_labels)
```





We can see that in both cases the samples are affected by one component of the distribution, although there are 3 or 4 components. Its look like adding the fourth component did not improved the explainability of the data that much.

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
```



The goal of that plot is to check what is the best k for k-means algorithm. By the elbow method, its look like it is best to select k=3 for Q3 and k=4 for Q4. By choosing k=8 to k=10 the loss is almost 0.

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'ggtern' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

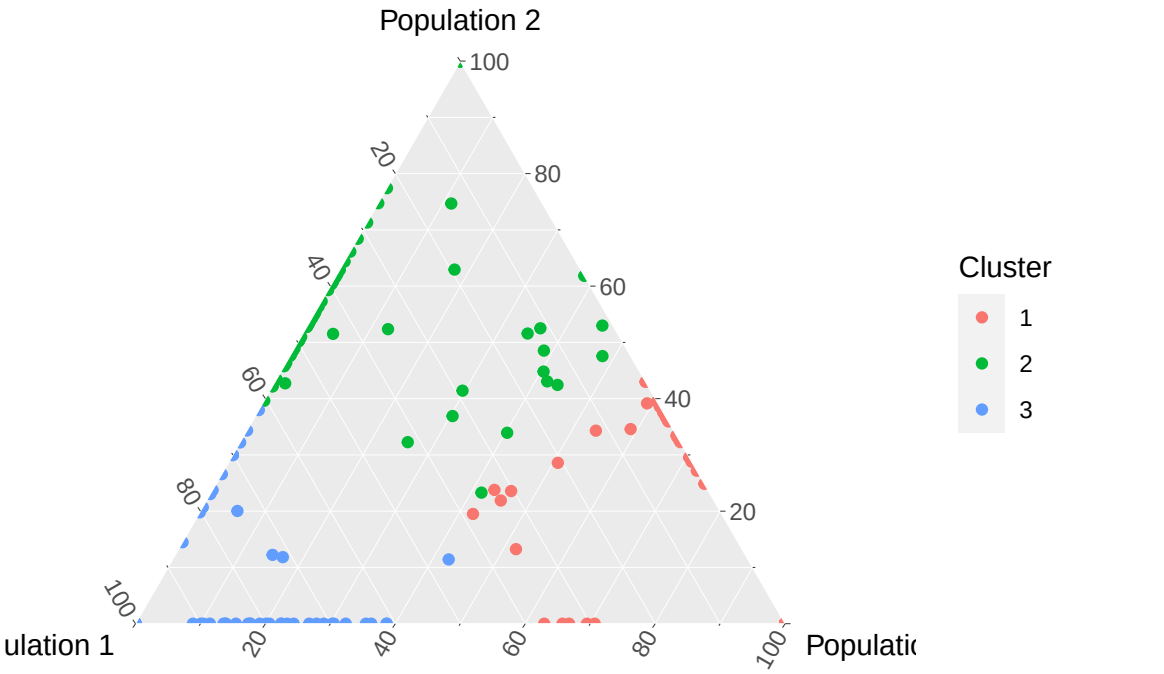
```
## Registered S3 methods overwritten by 'ggtern':
##   method           from
##   grid.draw.ggplot ggplot2
##   plot.ggplot      ggplot2
##   print.ggplot     ggplot2
```

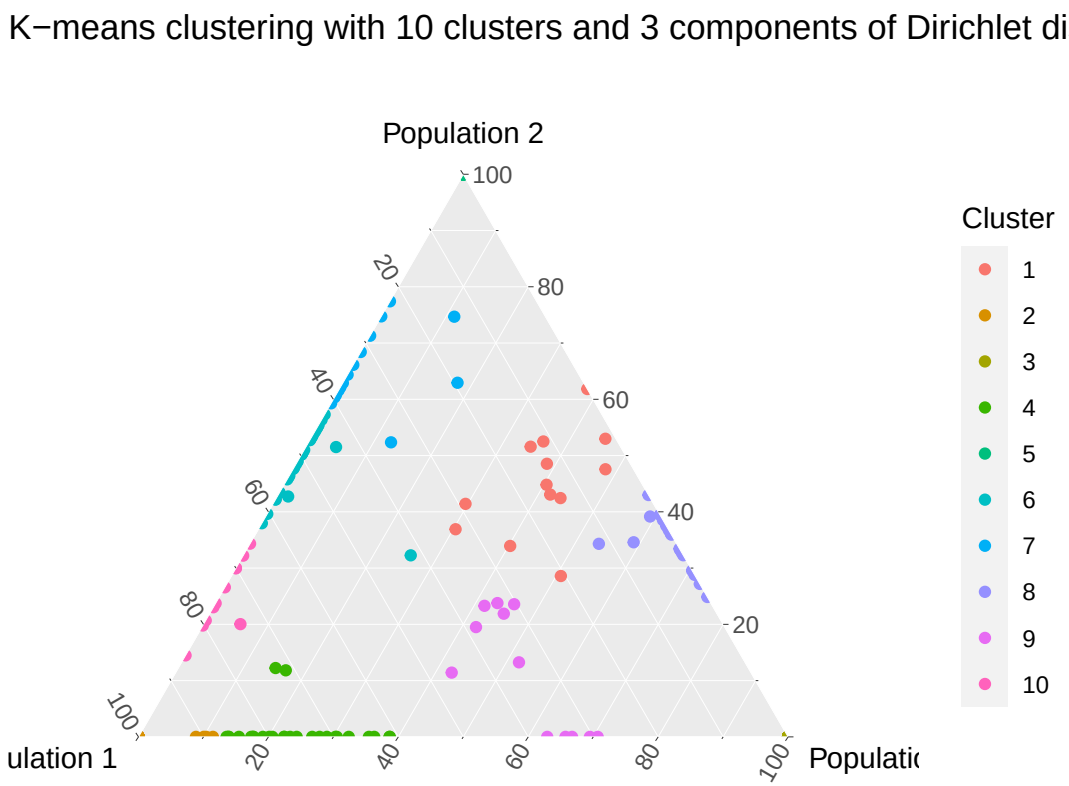
```
## --
## Remember to cite, run citation(package = 'ggtern') for further info.
## --
```

```
##
## Attaching package: 'ggtern'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     aes, annotate, ggplot, ggplot_build, ggplot_gtable, ggplotGrob,
##     ggsave, layer_data, theme_bw, theme_classic, theme_dark,
##     theme_gray, theme_light, theme_linedraw, theme_minimal, theme_void
```

```
## The following objects are masked from 'package:gridExtra':
##
##     arrangeGrob, grid.arrange
```

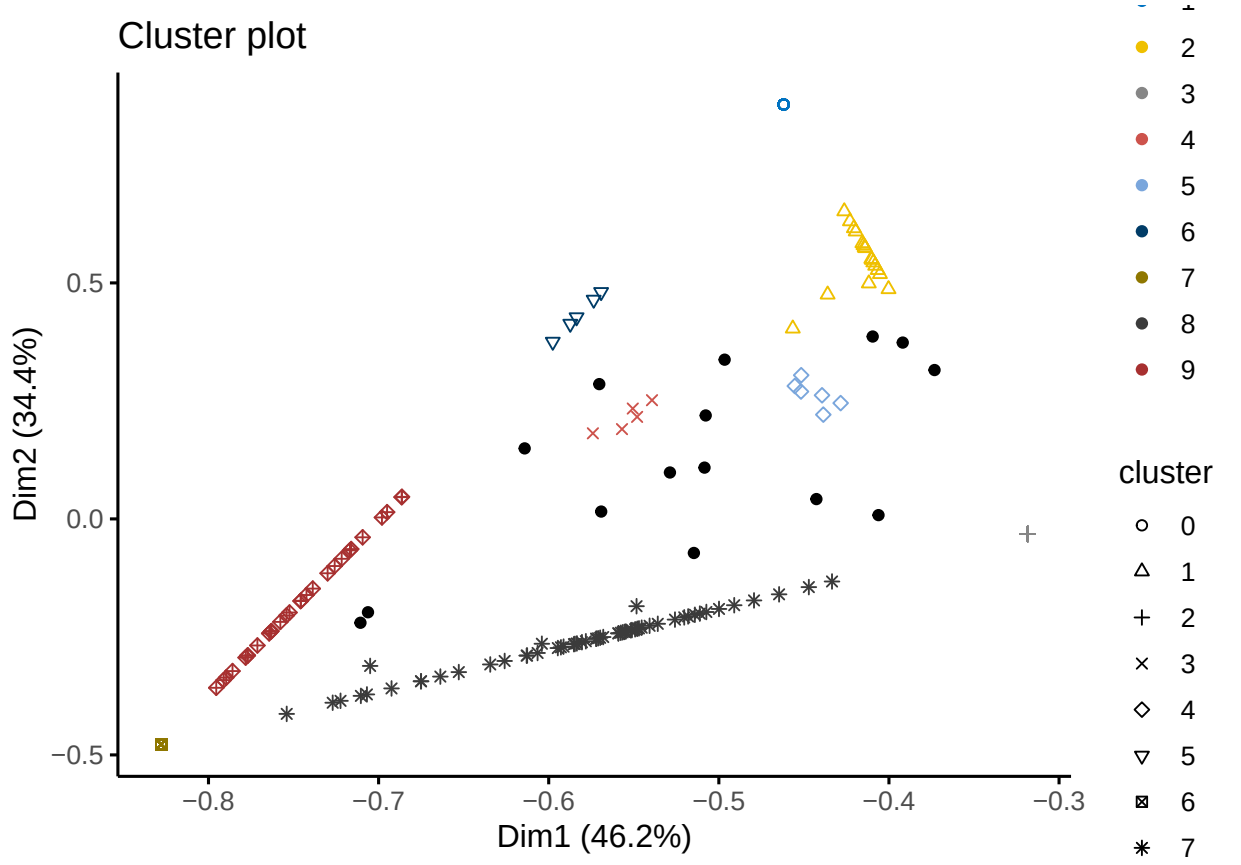




By looking at the trinary graph, its look like there are few different groups with different behaviors and distributions. Although the elbow method suggets to take for k-means k=3, I think it is much more suitable to take k=10. In that way, groups with no contribution in one of the components will attributed to separate cluster than those of affected from all the components like the points in the middle of the triangle. Yet, using k-means is not optimal as the centers are spatial and they separate gorups that suppose to be together like the ones in the left upper edge.

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'fpc' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages

## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'dbscan' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'dbscan'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\dbscan\libs\x64\dbscan.dll
## to
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\dbscan\libs\x64\dbscan.dll:
## Permission denied

## Warning: restored 'dbscan'

##
```

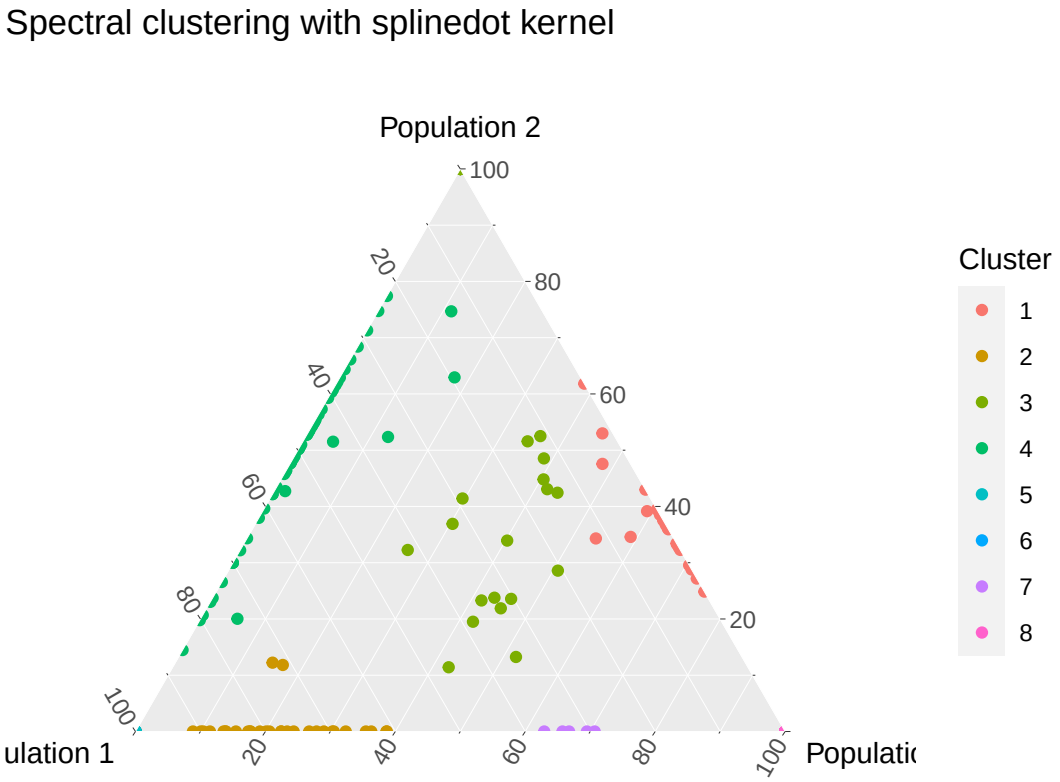
```
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
##
## Attaching package: 'dbscan'
## The following object is masked from 'package:fpc':
##
##     dbscan
## The following object is masked from 'package:stats':
##
##     as.dendrogram
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```



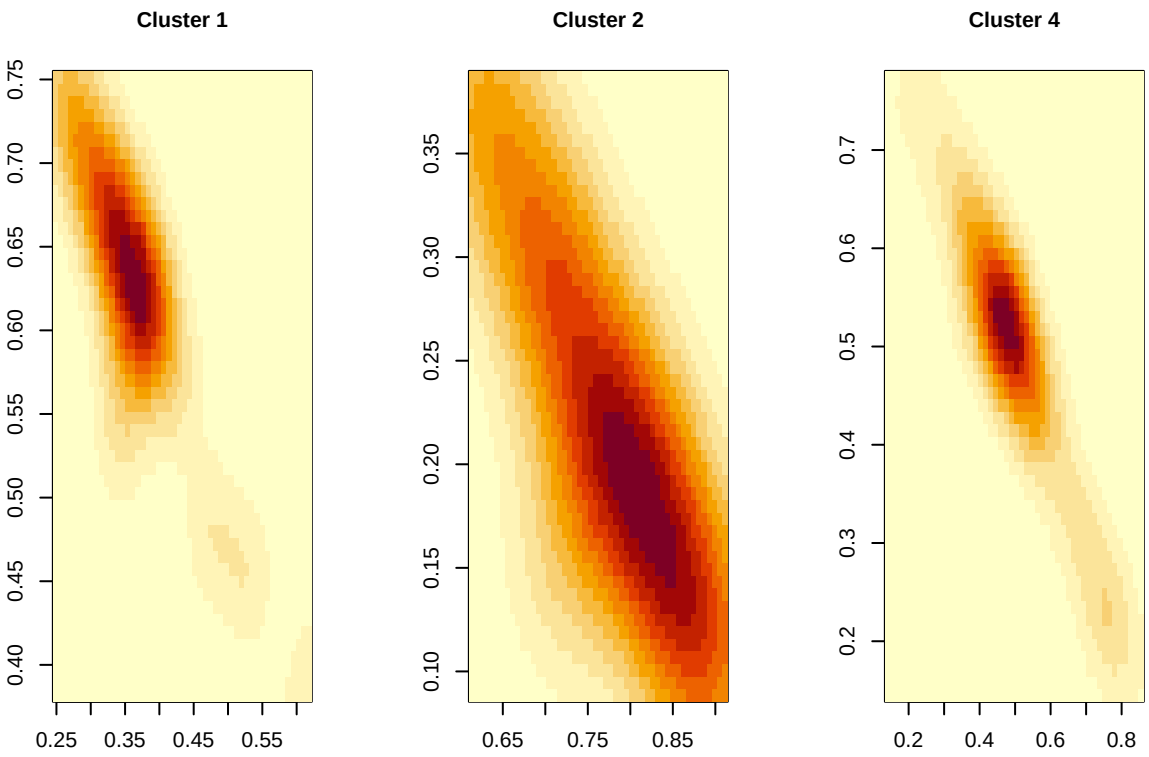
In this section we try to use DBSCAN in order to find better clusters. The assumption of DBSCAN is that clusters are densed in the space more than other areas in the space. It groups densely grouped data points into single cluster. It performs better on separating the edges to other groups than the points that affected from 2-3 components.

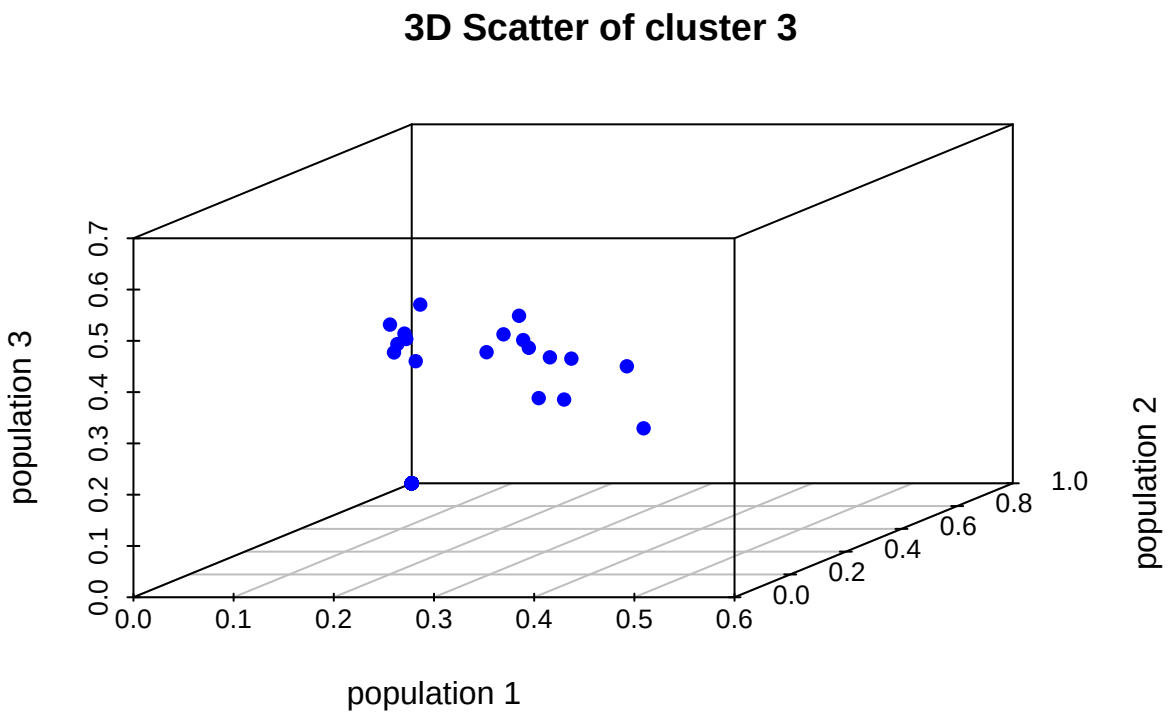
```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
## package 'kernlab' successfully unpacked and MD5 sums checked
## Warning: cannot remove prior installation of package 'kernlab'
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\kernlab\libs\x64\kernlab.dll
## to
```

```
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\kernlab\libs\x64\kernlab.dll:
## Permission denied

## Warning: restored 'kernlab'

##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages

##
## Attaching package: 'kernlab'

## The following object is masked from 'package:ggplot2':
##
##     alpha
```



In this section we apply spectral clustering with splinedot kernel. The goal was to find a kernel that separates the best between the zero components group to the the groups that are contributed from all the components. There are real 5 groups (and 3 more groups that catched single dots). This is the best clusters we got up to this point.


Cluster 1
Cluster 2
Cluster 4
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.25
0.35
0.45
0.55
0.35
0.30
0.25
0.20
0.15
0.10
0.65
0.75
0.85
0.7
0.6
0.5
0.4
0.3
0.2
0.2
0.4
0.6
0.8



By examine the cluster, we can find 3 kinds of clusters: 1. One component cluster - affected by one component only (Alpha3 = 0.99998 for example) 2. Two components cluster. We can see that it is approximately Guassian by looking at the heatmap 3. Three components cluster. We can see that it is approximately Guassian by looking at the 3DScatterplot

To sum up we had Cluster 5, 8 - one component Clusters 1, 2, 4, 7(small) - two components, approximately Guassian distributed cluster 3 - three components, approximately Guassian distributed

So we can say that Q3 composed from 5 major groups, each from specific as mentioned above.

```
install.packages("GGally")
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

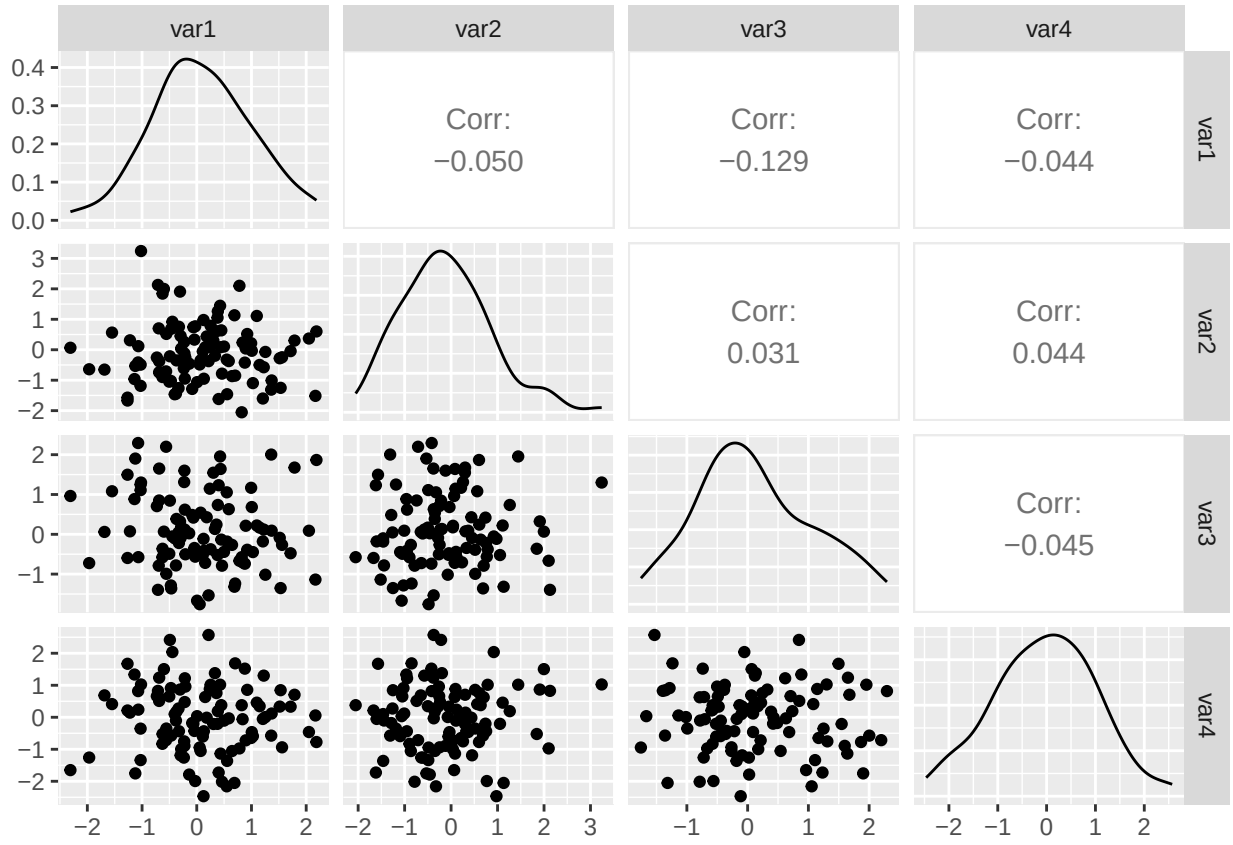
```
## package 'GGally' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
# Generate some sample data with four variables
set.seed(123)
n <- 100
data <- data.frame(
```

```
  var1 = rnorm(n),
  var2 = rnorm(n),
  var3 = rnorm(n),
  var4 = rnorm(n)
)

# Create the scatterplot matrix
ggpairs(data)
```




```
## Warning: package 'scatterplot3d' is in use and will not be installed
```

```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'klaR' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
```

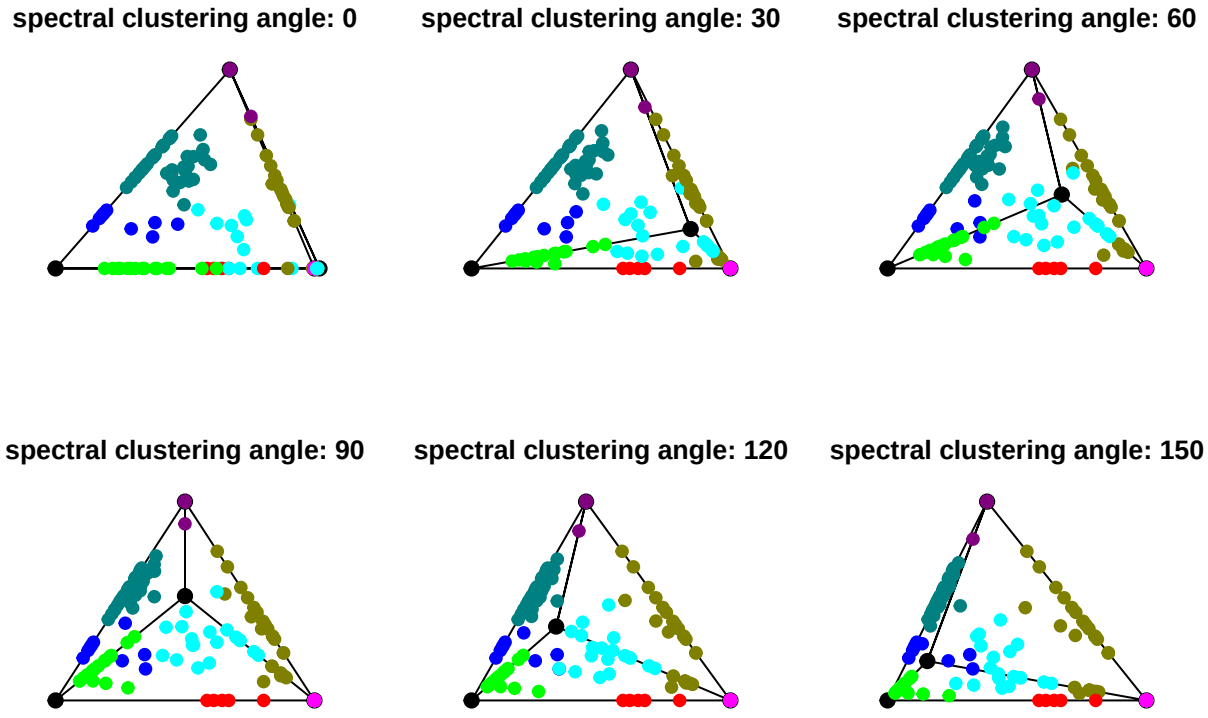
```
## Installing package into 'C:/Users/Yuval-PC/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

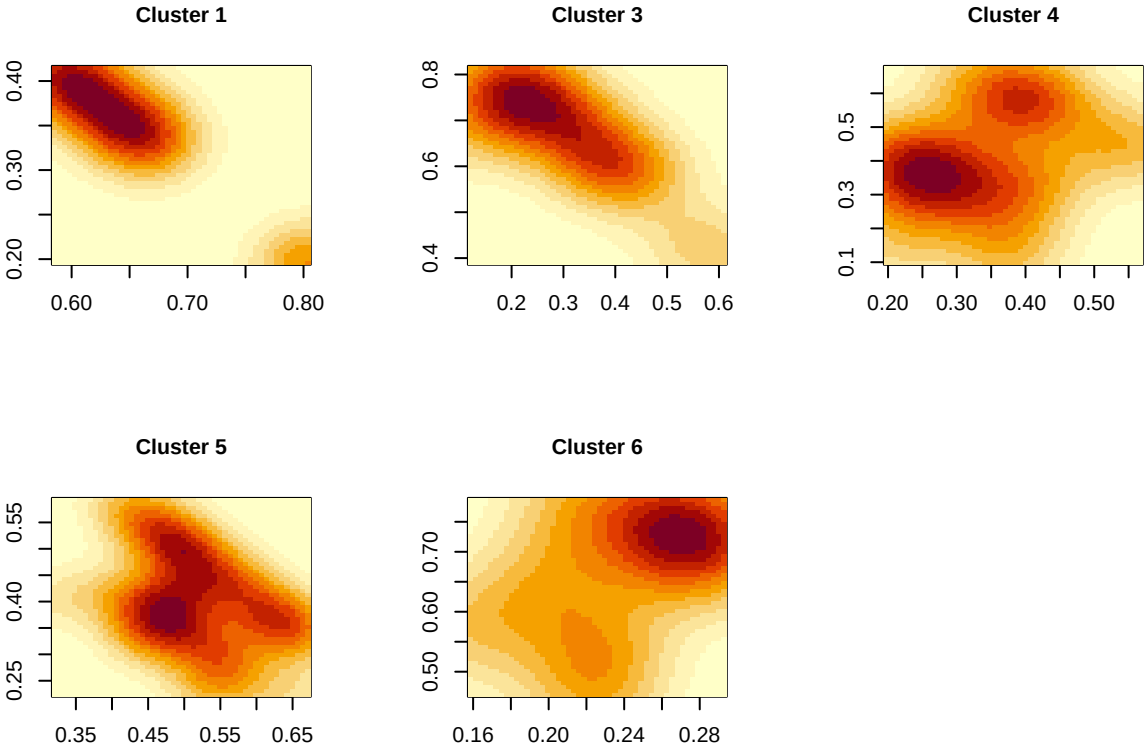
```
## package 'haven' successfully unpacked and MD5 sums checked
```

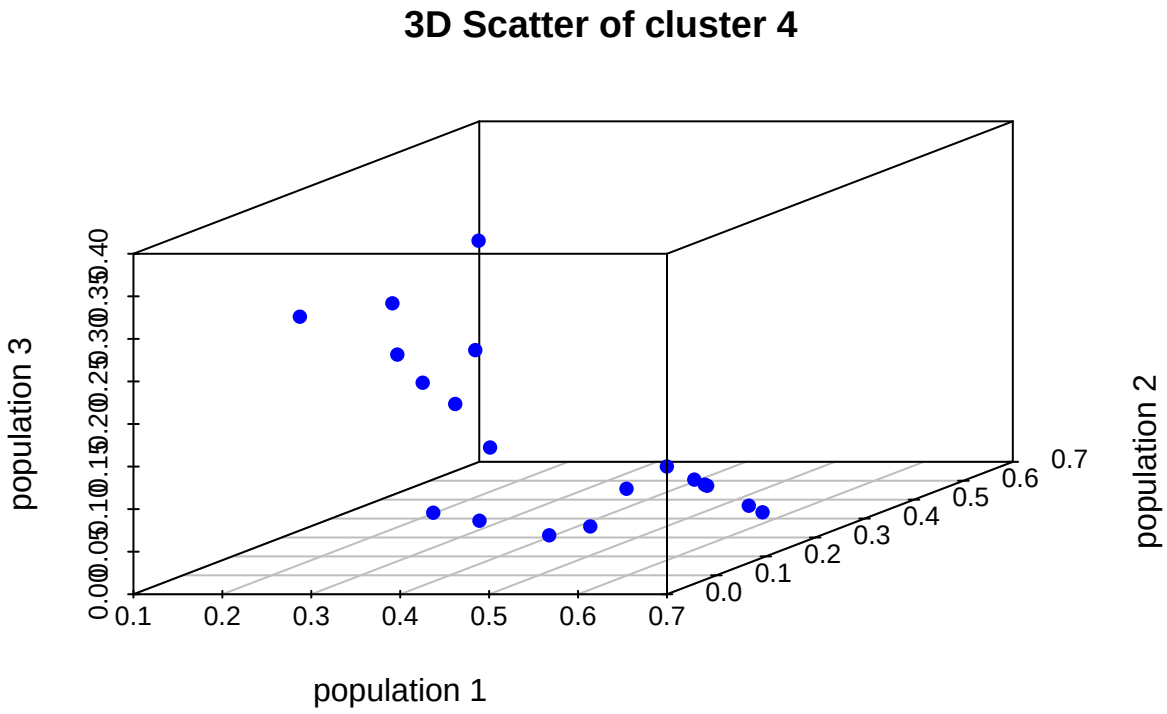
```
## Warning: cannot remove prior installation of package 'haven'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\00LOCK\haven\libs\x64\haven.dll
## to C:\Users\Yuval-PC\AppData\Local\R\win-library\4.3\haven\libs\x64\haven.dll:
```

```
## Permission denied
## Warning: restored 'haven'
##
## The downloaded binary packages are in
##  C:\Users\Yuval-PC\AppData\Local\Temp\RtmpCm5ser\downloaded_packages
## [1] 0
## [1] 30
## [1] 60
## [1] 90
## [1] 120
## [1] 150
```




Cluster 1
0.40
0.30
0.20
0.60
0.70
0.80
Cluster 3
0.8
0.6
0.4
0.2
0.3
0.4
0.5
0.6
Cluster 4
0.5
0.3
0.1
0.20
0.30
0.40
0.50
Cluster 5
0.55
0.40
0.25
0.35
0.45
0.55
0.65
Cluster 6
0.70
0.60
0.50
0.16
0.20
0.24
0.28



Like in Q3, we separate to 3 kind of groups one component - 8(V3) 7(V1) 9(V2) 10(V4) two components - 1, 2, 3, 5, 6 three components - 4

I contrast to the case of Q3 with 3 dimensions, here the clustering algorithms found it hard to separate the points into groups of 2 dimension and most of them spread over 3 dimensions. It is safe to say that the samples are taken from a mixture of distributions, but here it is harder to say what are the hidden distributions.