Density Estimation 2

GMM. DBSCAN

Caballero Vergés Biel, Menzenbach Svenja and Reyes Illescas Kleber Enrique

2023-10-09

```
load("BikeDay.Rdata")
X <- as.matrix(day[day$yr==1,c(10,14)])
#pairs(X)
```

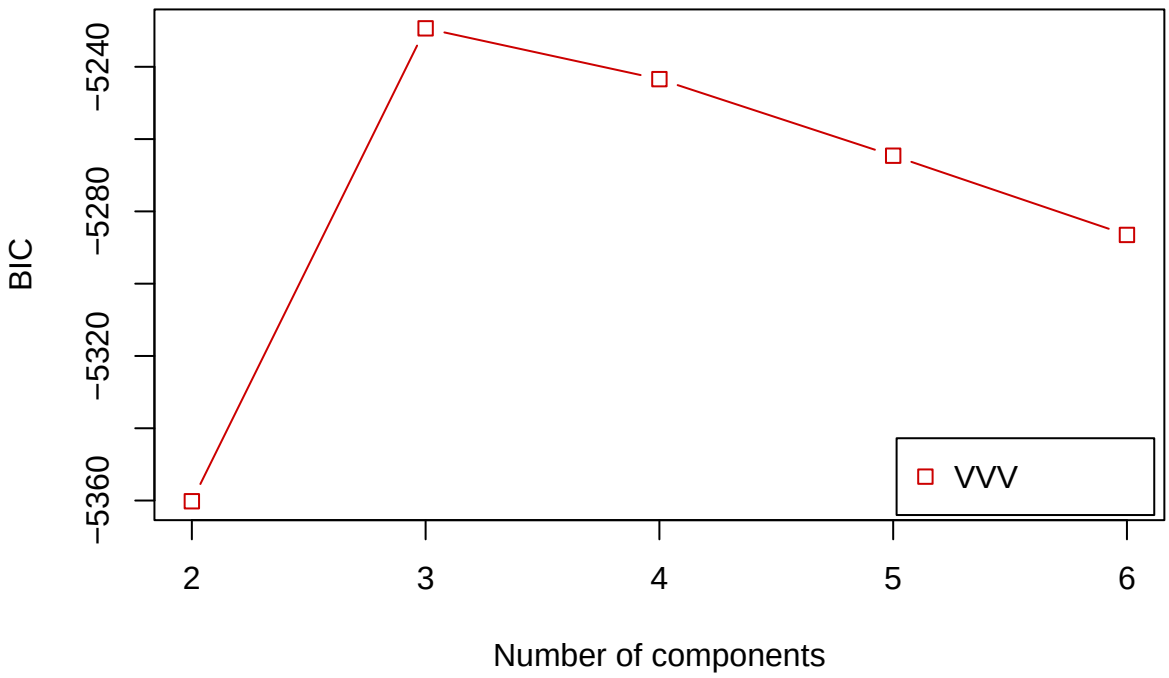
Questions

1.

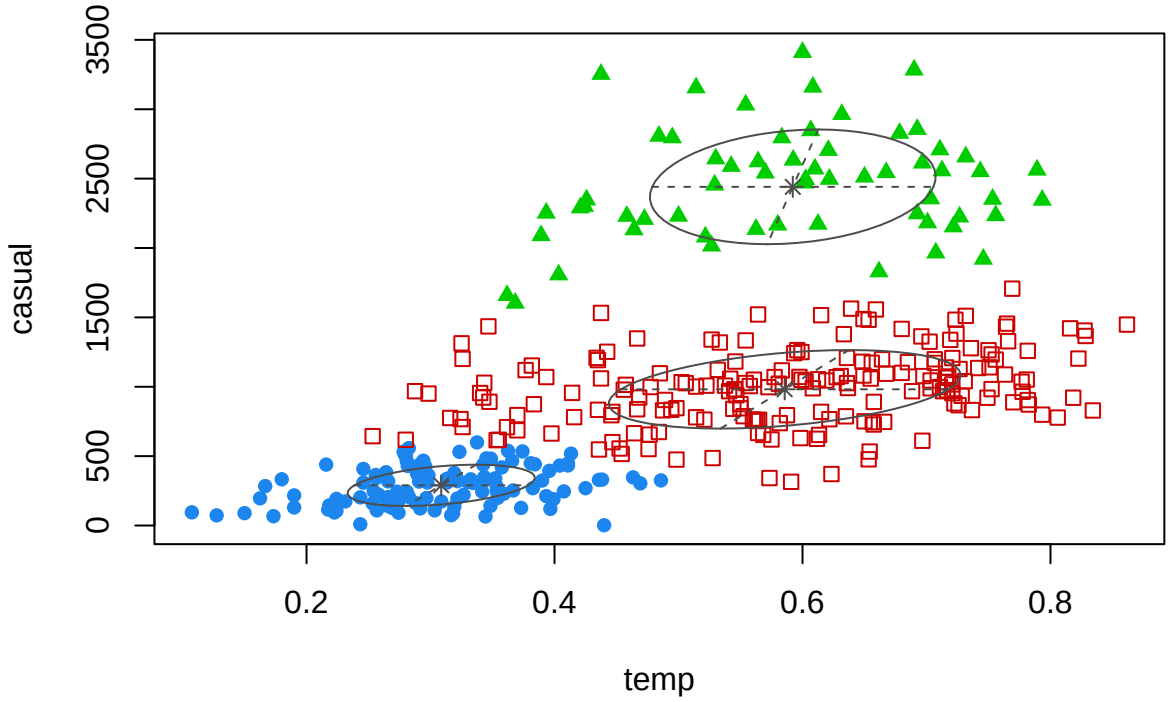
```
GMM_BIC <- Mclust(X,G=2:6, modelNames="VVV")
summary(GMM_BIC, parameters=F)
```

```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust VVV (ellipsoidal, varying volume, shape, and orientation) model with 3
## components:
##
##  log-likelihood   n df       BIC       ICL
##       -2564.509 366 17 -5229.362 -5261.588
##
## Clustering table:
##   1   2   3
## 111 195  60
```

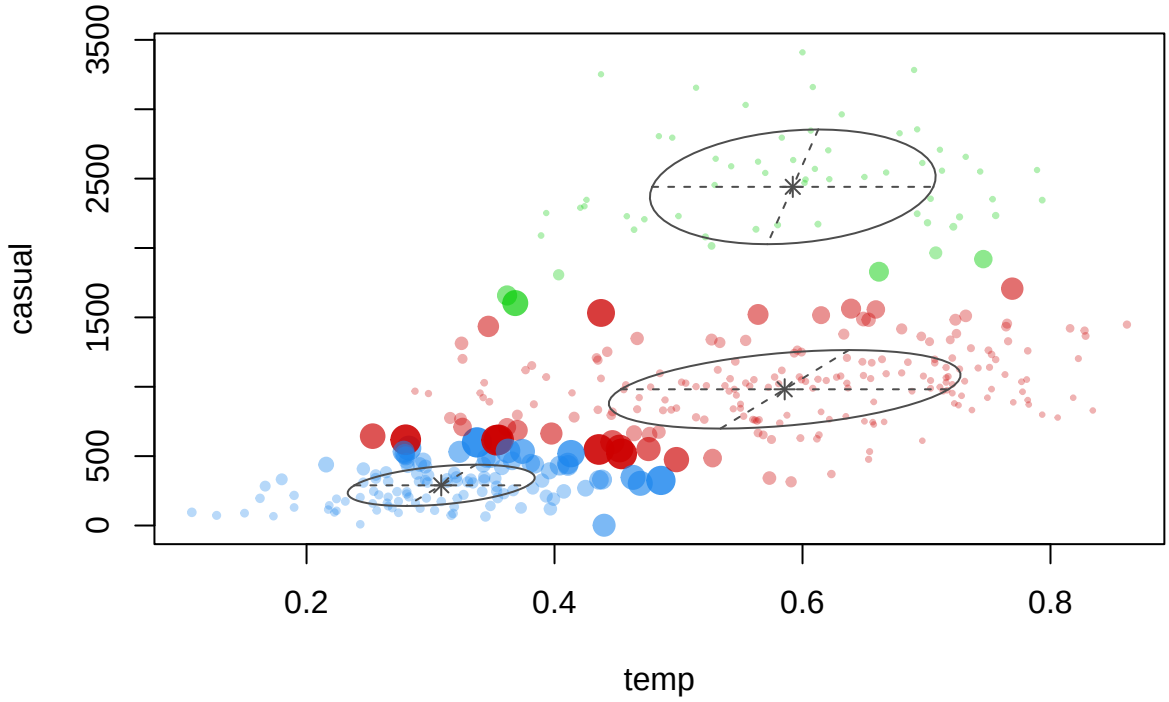
```
plot(GMM_BIC, what="BIC")
```



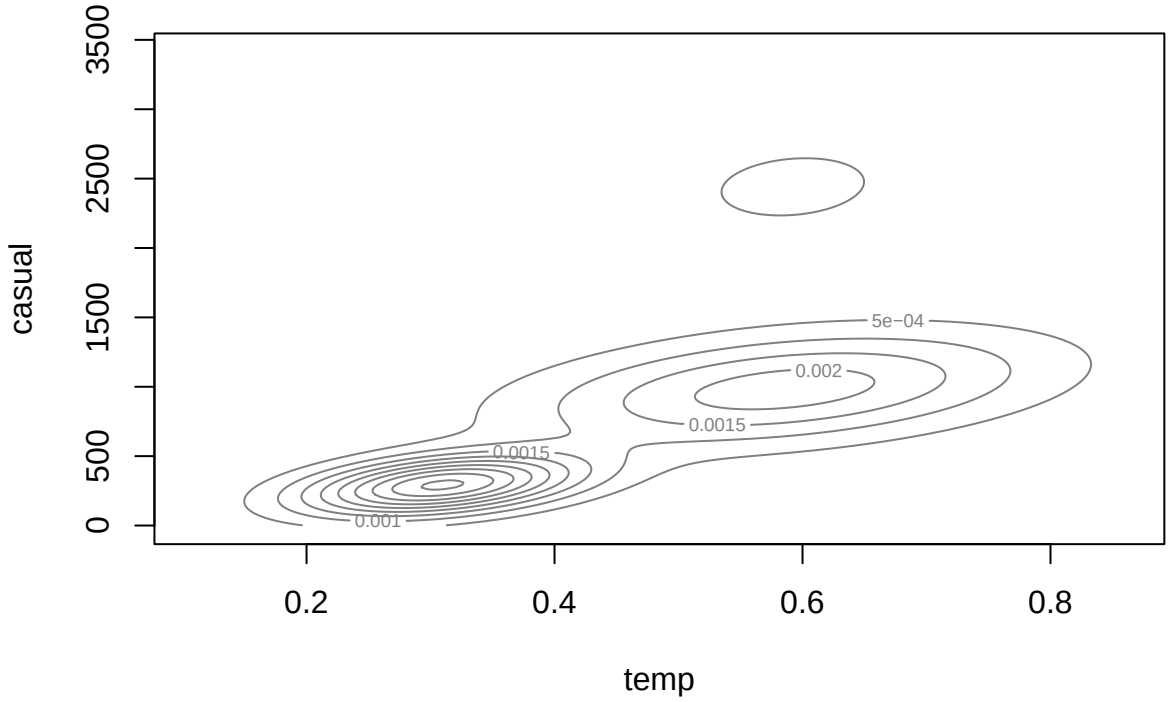

```
plot(GMM_BIC, what="classification")
```




```
plot(GMM_BIC, what="uncertainty")
```

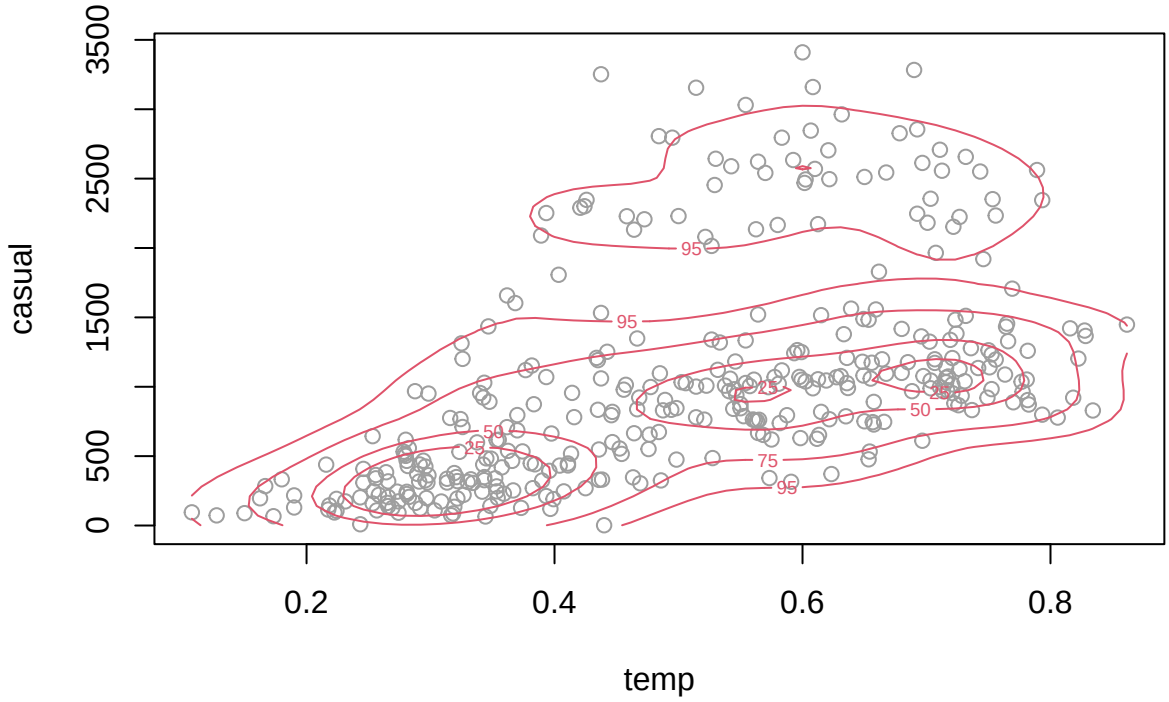



```
plot(GMM_BIC, what="density")
```



2.

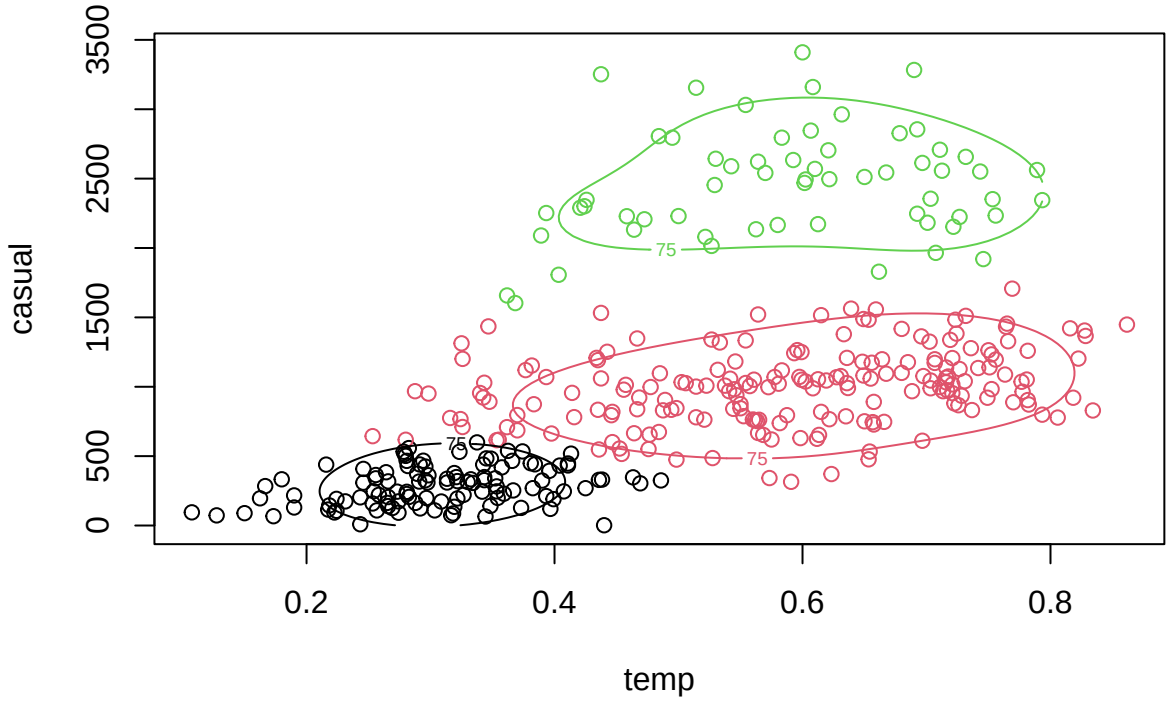
```
#?sm.density

a = 0.25

plot(X, col=8)
sm.density(X,h=a*c(sd(day$temp), sd(day$casual)),
           display="slice",col=2, props=c(25,50,75,95), add=TRUE)
```




```
# default props -> c(75,50,25)
```

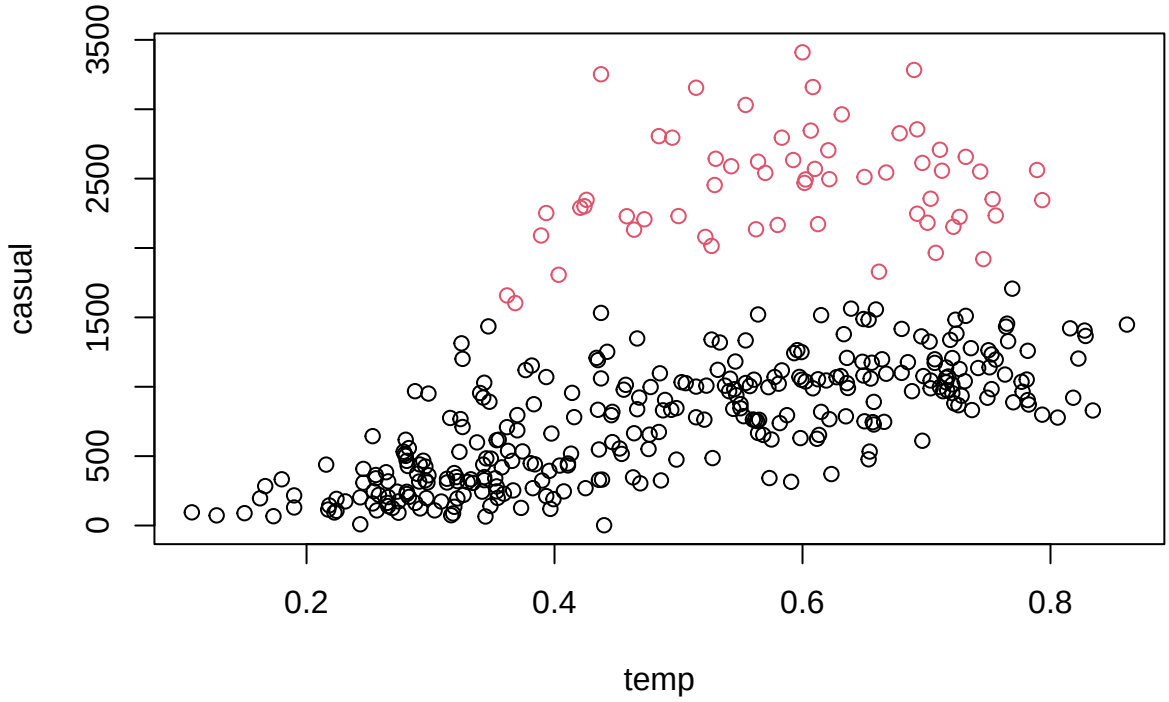
3.

```
k <- GMM_BIC$G
clust.ind <- GMM_BIC$classification

#plot(X, col=8)

plot(X,col=clust.ind)

a = 0.4

for (j in 1:k){
  cl.j <- (clust.ind==j)
  sm.density(X[cl.j,],h=a*c(sd(day$temp), sd(day$casual)),
             display="slice",props=c(75),
             col=j, cex=4, add=TRUE)
}
```



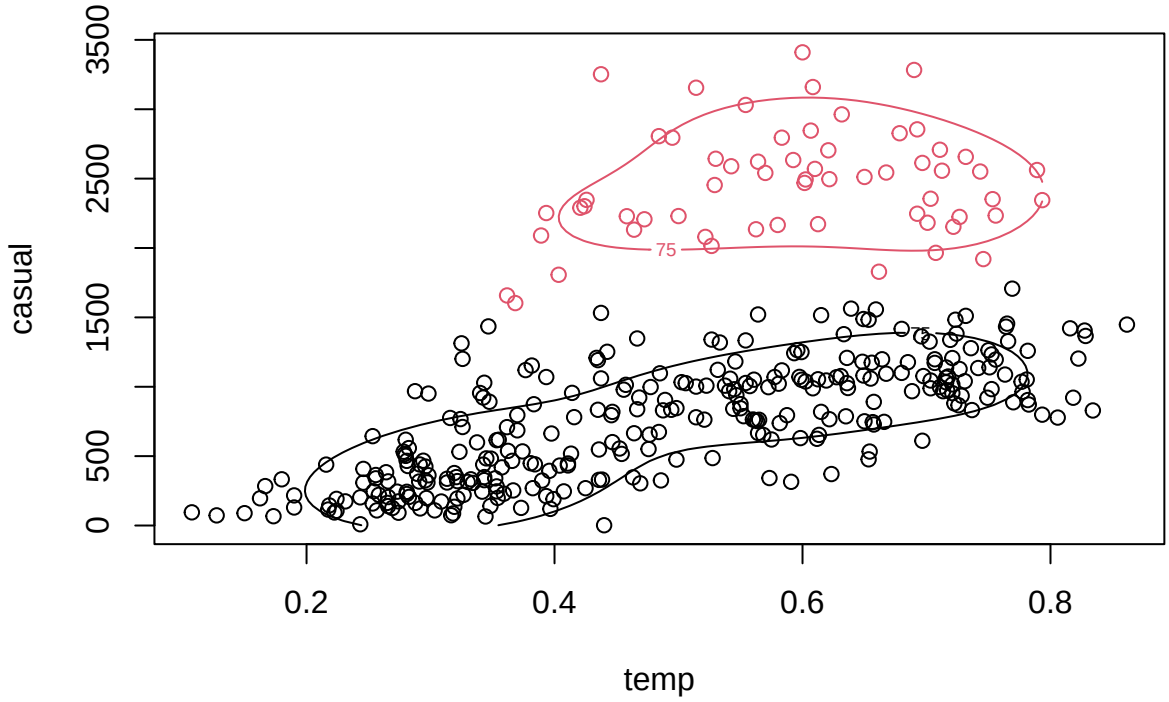
4.

```
#?fpc
cm <- mclustBIC(X,G=2:6,modelNames="VVV")
sum <- summary(cm, X)
cmnbhat <- mergenormals(X, sum, method="bhat")

plot(X, col=cmnbhat$clustering)
```



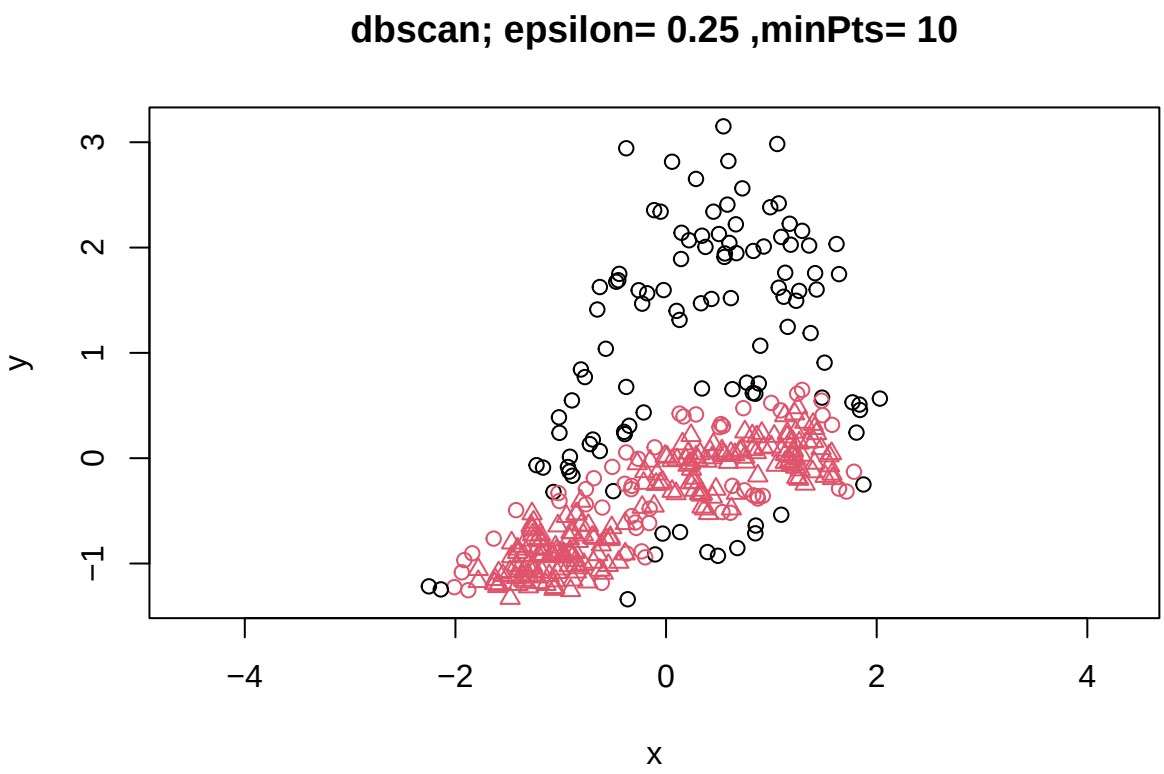
5.

```
k_new <- length(cmnbhat$clusternumbers)
clust_merged.ind <- cmnbhat$clustering

#plot(X, col=8)

plot(X,col=clust_merged.ind)

a = 0.4

for (j in 1:k_new){
  cl.j <- (clust_merged.ind==j)
  sm.density(X[cl.j,],h=a*c(sd(day$temp), sd(day$casual)),
             display="slice",props=c(75),
             col=j, cex=4, add=TRUE)
}
```

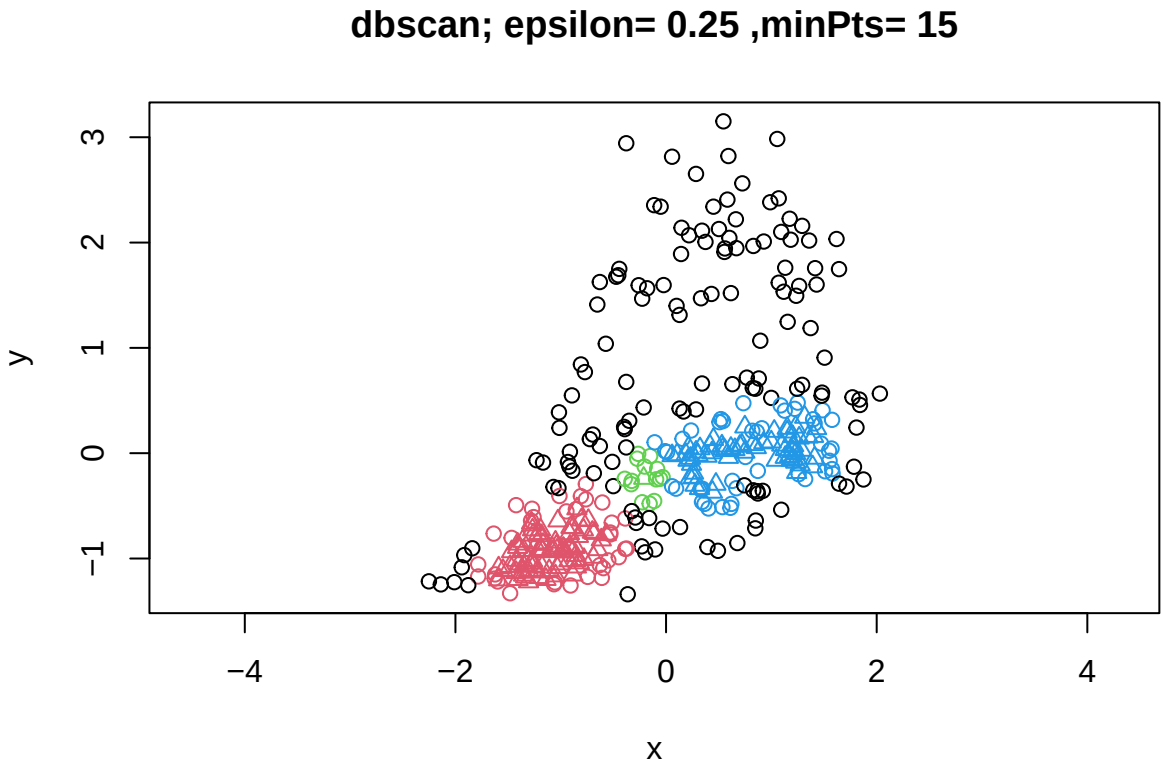


6.

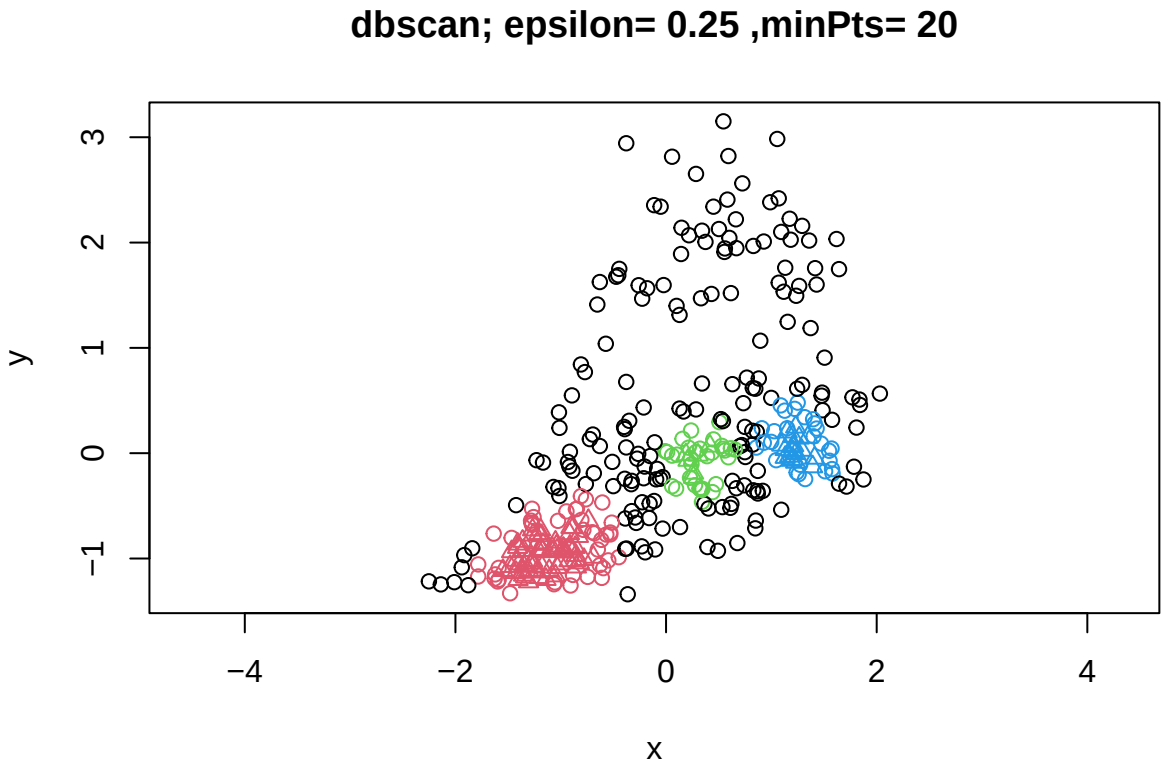
```
#?fpc::dbscan  / ?dbscan::dbscan
Xs <- scale(X)

eps_values <- c(0.25, 0.5)
minPts_values <- c(10, 15, 20)


for (eps in eps_values) {
  for (minPts in minPts_values) {

    dbscan_result <- fpc::dbscan(Xs, eps = eps, MinPts = minPts, showplot = 0)
    plot(dbscan_result, Xs, main=paste("dbscan; epsilon=",eps,",minPts=",minPts),
          xlab="x",ylab="y",asp=1)
    print(dbscan_result)
  }
}
```



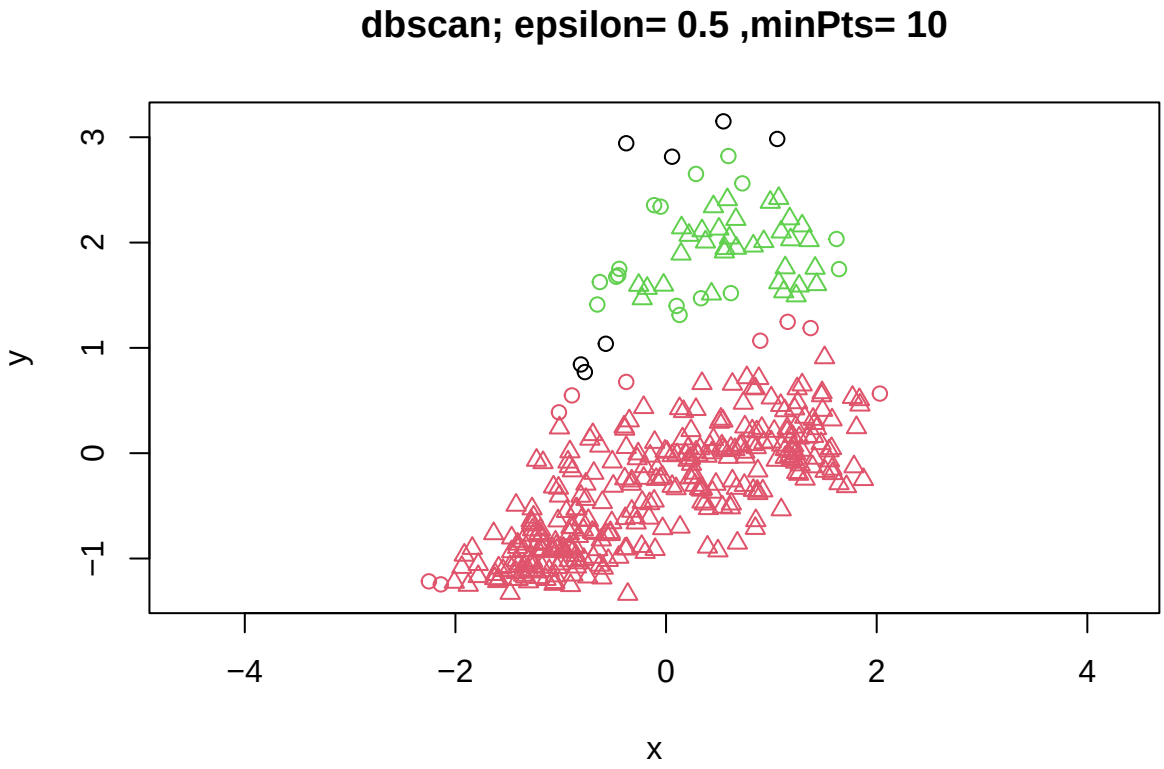

```
## dbscan Pts=366 MinPts=10 eps=0.25
##          0   1
## border 105  54
## seed     0 207
## total  105 261
```



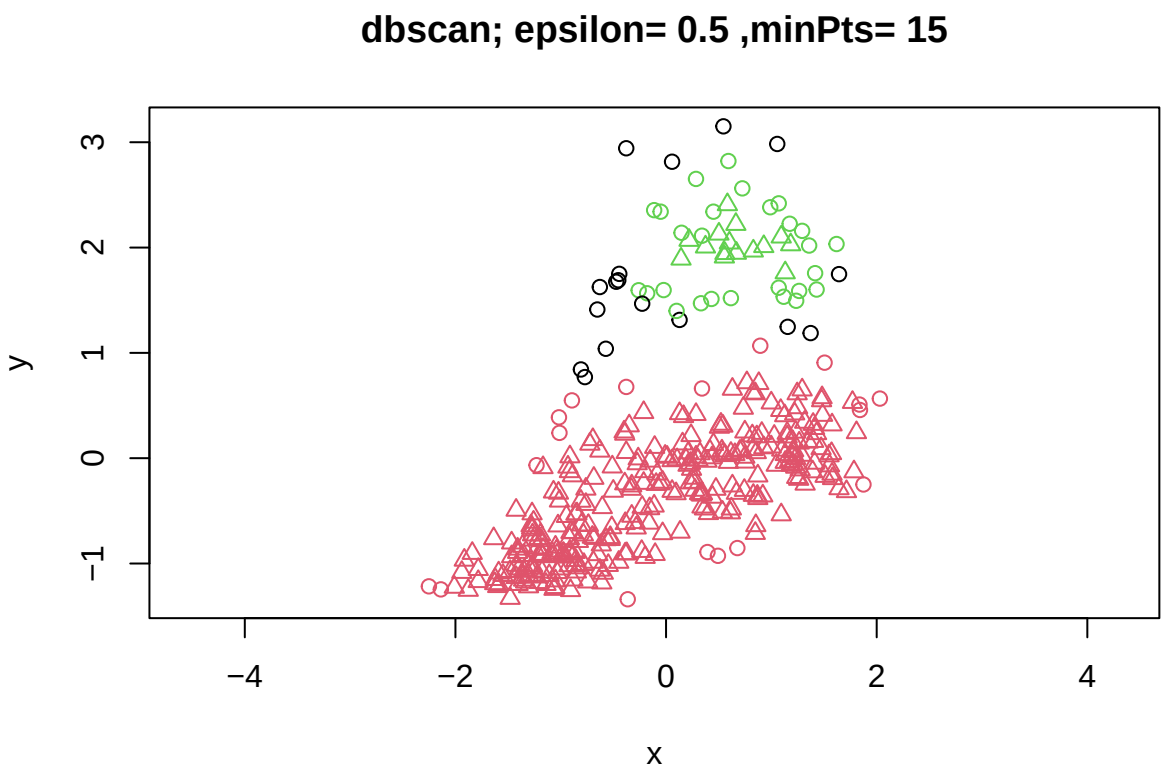

```
## dbscan Pts=366 MinPts=15 eps=0.25
##          0   1  2   3
## border 135  34 14  44
## seed     0  76  1  62
## total  135 110 15 106
```



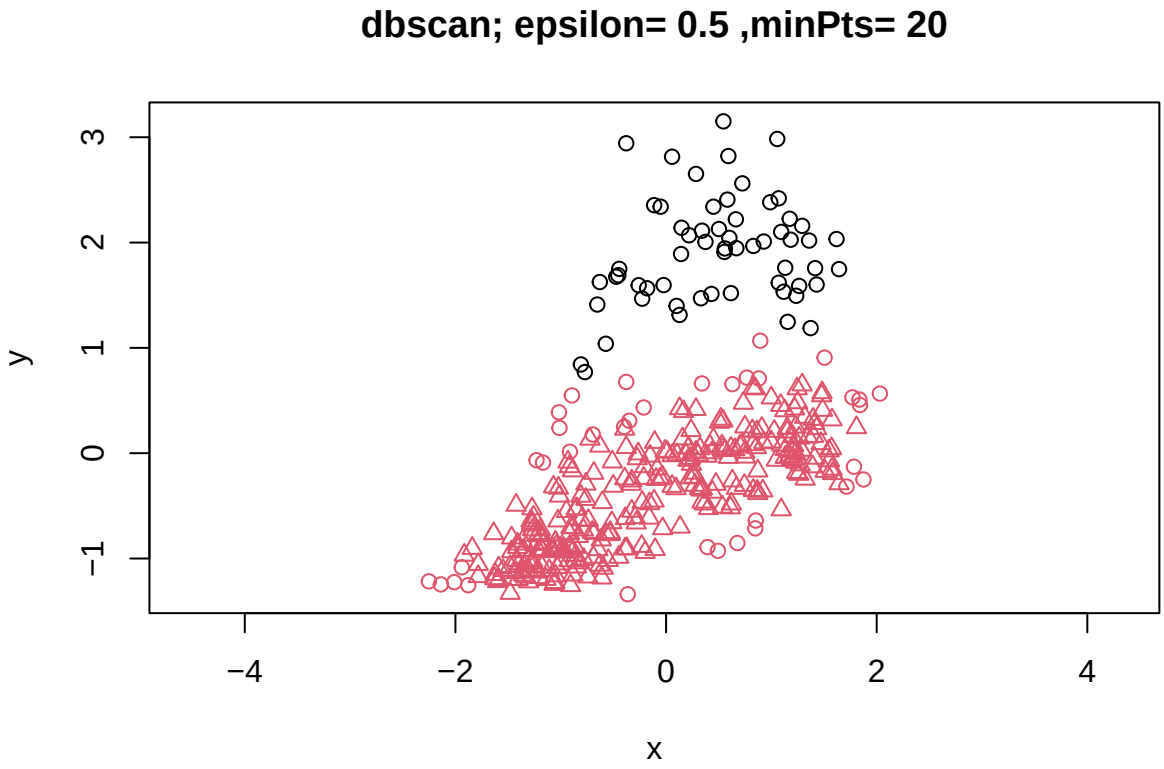

```
## dbscan Pts=366 MinPts=20 eps=0.25
##          0   1  2  3
## border 178  42 32 27
## seed     0  62  6 19
## total  178 104 38 46
```




```
## dbscan Pts=366 MinPts=10 eps=0.5
##        0   1  2
## border 7   9 16
## seed   0 300 34
## total  7 309 50
```

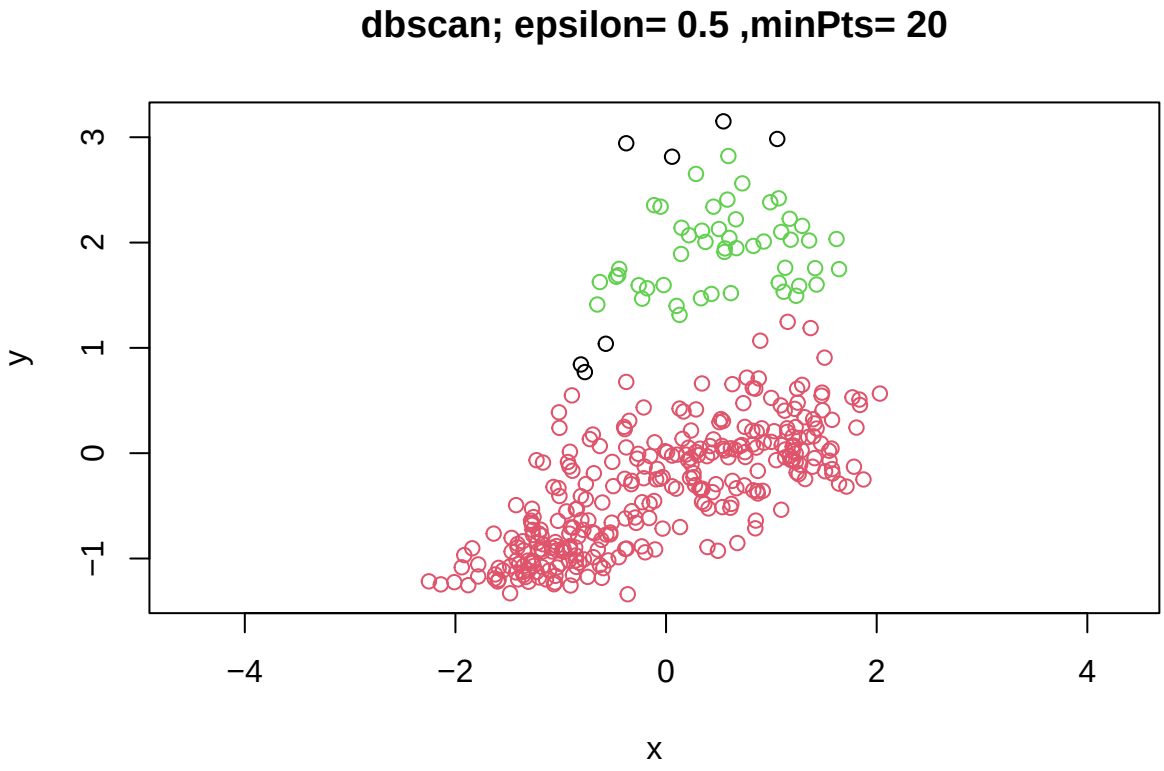



```
## dbscan Pts=366 MinPts=15 eps=0.5
##          0   1  2
## border 17  18 27
## seed    0 289 15
## total  17 307 42
```




```
## dbscan Pts=366 MinPts=20 eps=0.5
##         0   1
## border 59  35
## seed    0 272
## total  59 307
```

```
#Cross table
favorite_eps <- 0.5
favorite_minPts <- 10
dbscan_result <- dbscan::dbscan(Xs, eps = favorite_eps, minPts = favorite_minPts)

plot(dbscan_result, Xs, main=paste("dbscan; epsilon=",eps,",minPts=",minPts),
      xlab="x",ylab="y",asp=1)
```




```
cluster_assignments <- dbscan_result$cluster

comparison_data <- data.frame(
  DBSCAN = cluster_assignments,
  mergenormals = cmnbhat$clustering
)

cross_table <- table(comparison_data)

print(cross_table)
```

```
##       mergenormals
## DBSCAN   1   2
##      0   0   7
##      1 306   3
##      2   0  50
```

7.

```
#Explanation (Comment all group)
```