

CSCI 585: Database Systems
Assignment: Weka and R
Name: Lohith Paripati (USC ID: 8299-5642-86)

FourClusters.R

```
#group1
x <- runif(100,0,1)
y <- runif(100,0,1)
f1 = data.frame(x, y)

#group2
x <- runif(100,3,4)
y <- runif(100,0,1)
f2 = data.frame(x, y)

# group3
x <- runif(100,0,1)
y <- runif(100,3,4)
f3 <- data.frame(x, y)

#group4
x <- runif(100,3,4)
y <- runif(100,3,4)
f4 <- data.frame(x, y)

#Bringing Group1 and Group2 together
m1 = merge(f1, f2, by = c("x","y"), all = TRUE)
#plot(m1)

#Bringing Group3 and Group4 together
m2 = merge(f3, f4, by = c("x","y"), all = TRUE)
#plot(m2)

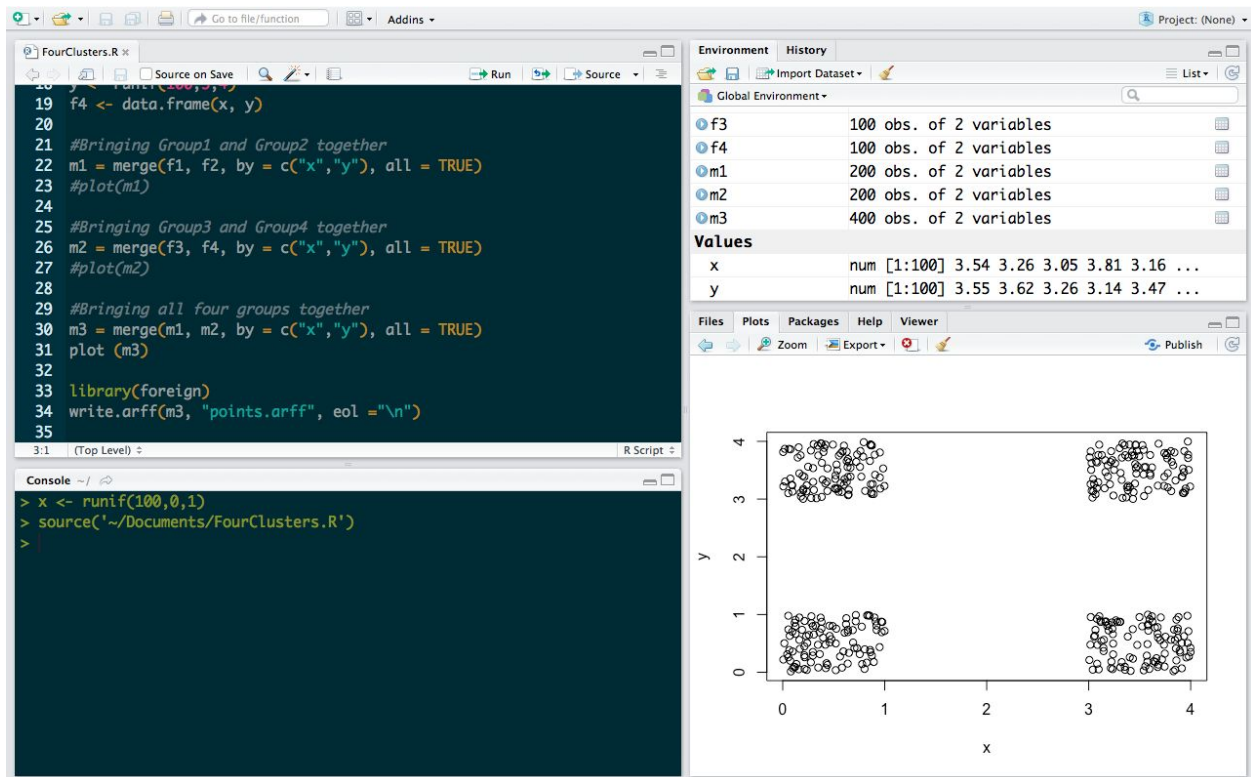
#Bringing all four groups together
m3 = merge(m1, m2, by = c("x","y"), all = TRUE)
```

plot (m3)

library(foreign)

write.arff(m3, "points.arff", eol = "\n")

Screengrab from Rstudio:



points.arff

@relation m3

@attribute x numeric

@attribute y numeric

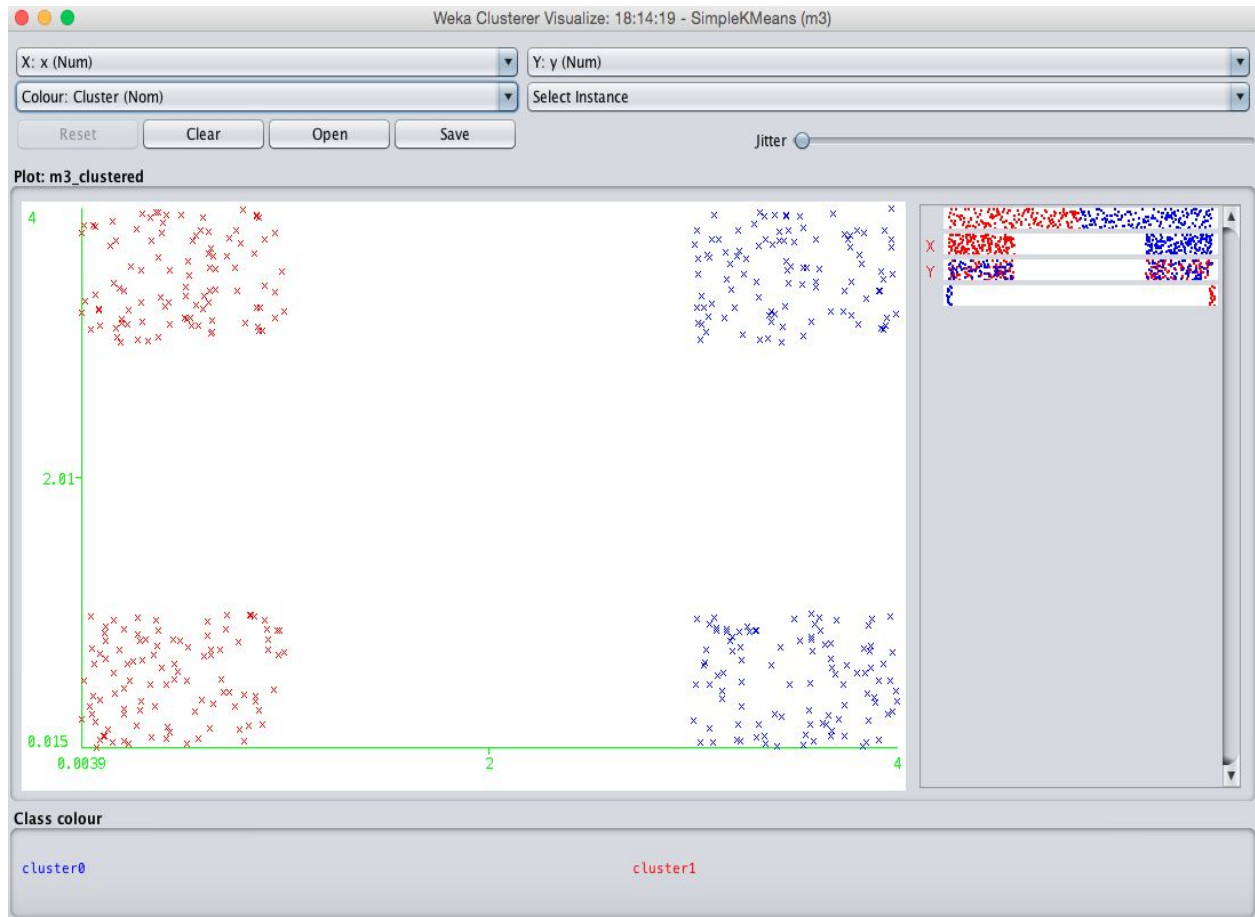
@data

0.00385620072484016,3.81536016310565
0.00554985413327813,0.219853724120185
0.00588478334248066,3.21942130522802
0.0159052389208227,0.502805310534313
0.0205389170441777,3.87013931502588

....

....

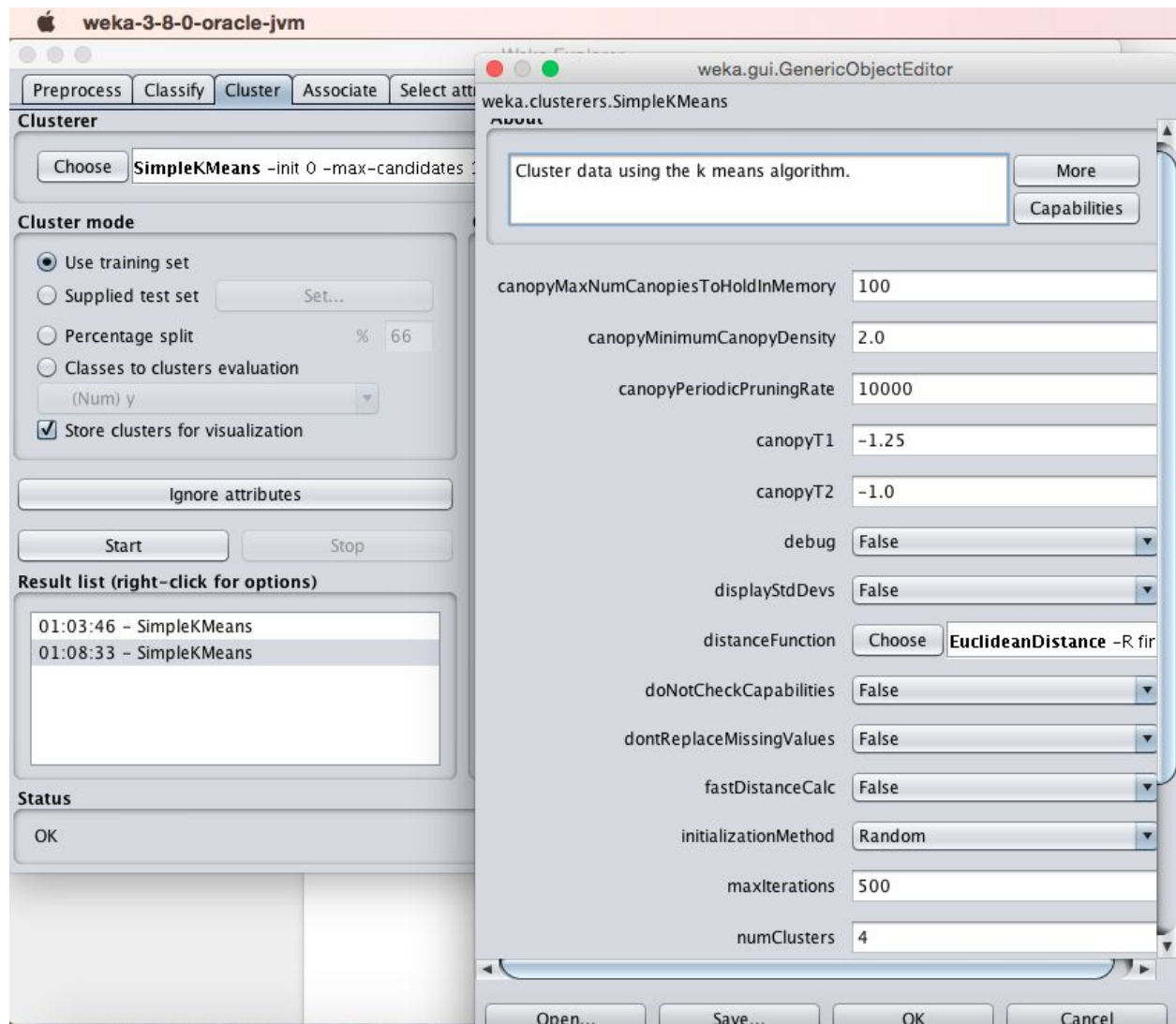
Screengrab from Weka:



Update in Assignment Question (29th June:) Show that 4 clusters are visualized in Weka stating Cluster 0, Cluster 1, Cluster 2 and Cluster 3.

<Check Next-Page>

To do that edit the properties of Cluster filter by right-clicking it and set the value as 4 for number of clusters attribute.



<Check Next-Page>

Now save the properties and visualize the cluster again. It will be as below -

