**Student id: 200110688**

**Name: Akshay vasant Nayak**

**R**. Bonus Question (R implementation) (**4% of grade**) (Please note that this is

completely optional; use your time wisely as the implementation may take time).

(You can use any 2-d data, real or simulated for implementation; test data will be

provided later to answer part b of this question)

(**a**) Implement G-Means (paper is provided under additional resources) (Algorithm 1,

listed on page 3). (submit code as separate file; make single zip file)

(**b**) Generate 2-d plots (scatter plots and draw ellipsoids) (data will be provided later),

include these plots as part of h/w solution)

**PART B**

**NOTE: 2d plots for test data has been plotted using significance level:1**

**(the six plots correspond to the clusters detected during the corresponding iterations of the algorithm)**

**The plots for this model has been attached in the zip file.**

**NOTE: THE CODE GENERATES PLOTS AND STORES IT IN A DIRECTORY SPECIFIED BY THE FOLLOWING COMMMAND**

**mypath <- file.path("C:","Users","AkshayN","Desktop","fds","r\_codes","hw5\_papers","cluster\_plots",paste("cluster\_generated\_during\_iteration\_0.jpg", sep = ""))**

**PLEASE UNCOMMENT THESE LINES AND SPECIFY THE DIRECTORY TO STORE PLOTS IF YOU WANT TO VIEW THEM .**

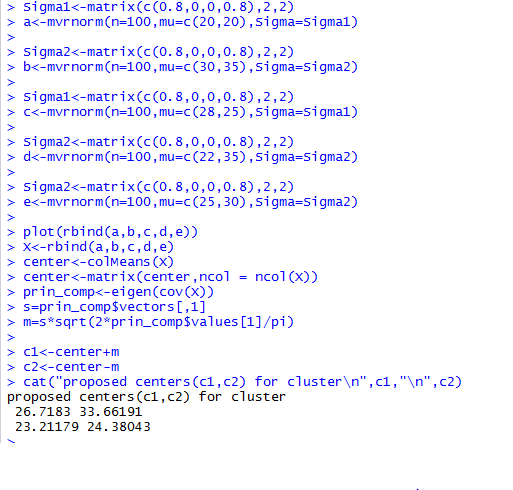
**THEY HAVE BEEN MARKED IN 2 SECTIONS AS SHOWN BELOW**

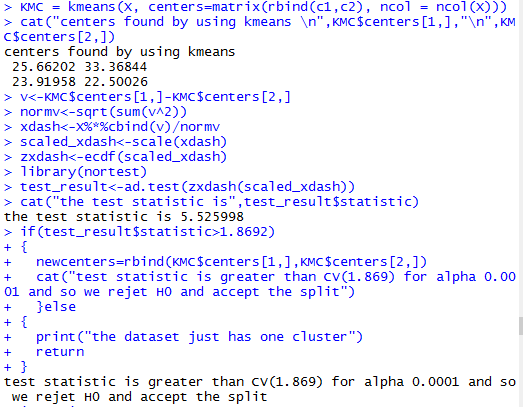
**\*\*\*\*\*\*\*\*\*\*\*\*\***

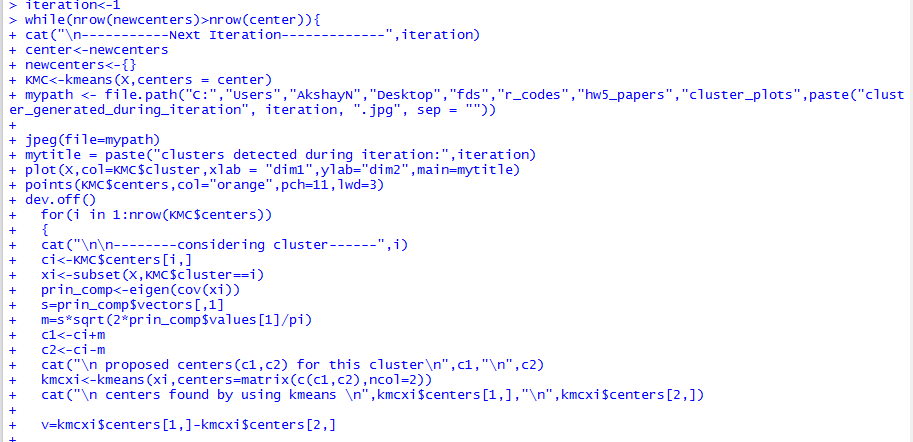
**LINES OF CODE**

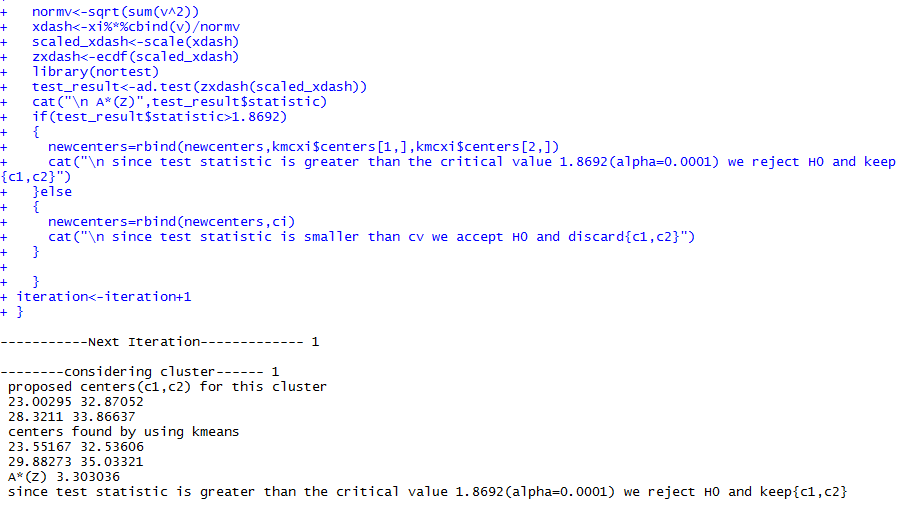
**\*\*\*\*\*\*\*\*\*\*\*\*\*\***

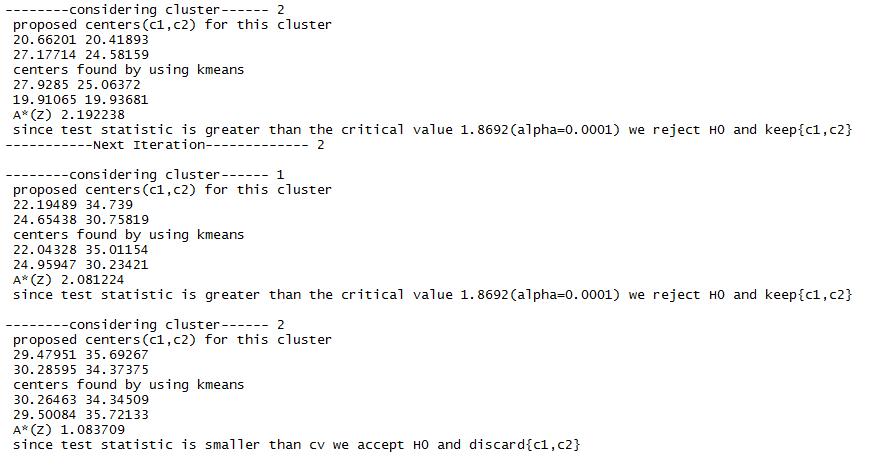
**Screenshots for sample 2d data**:

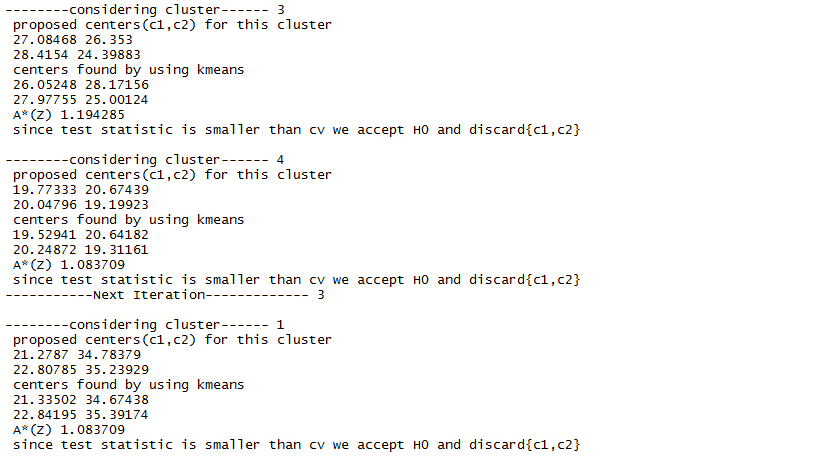


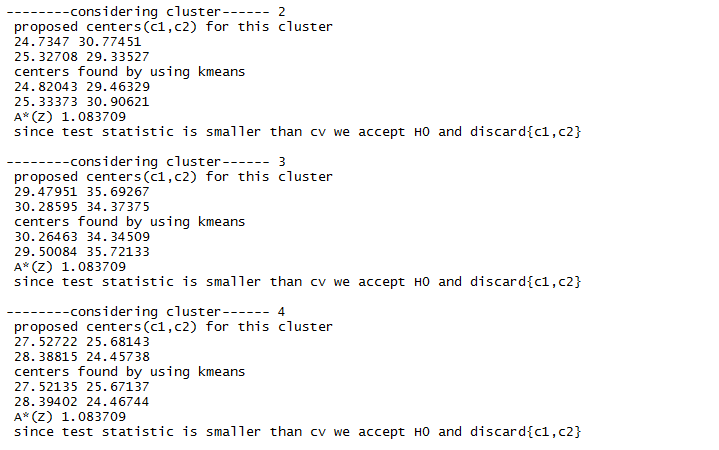


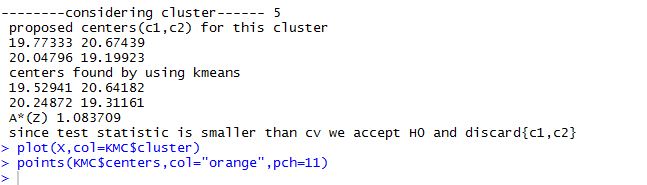


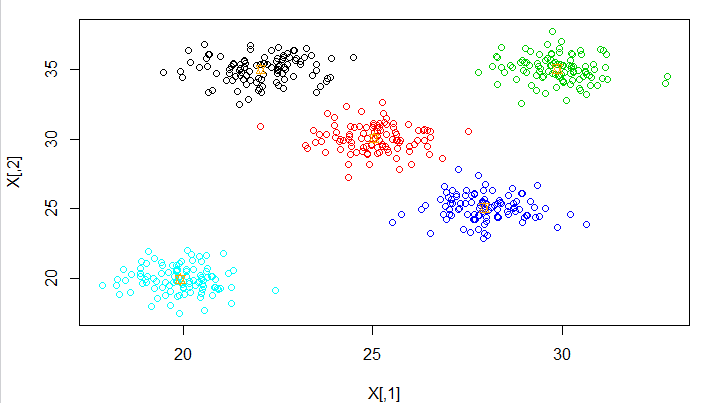




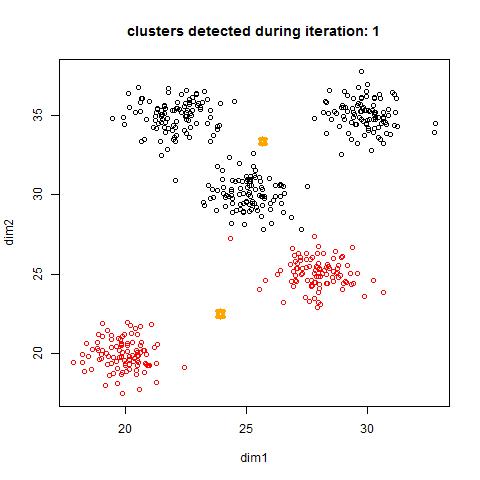


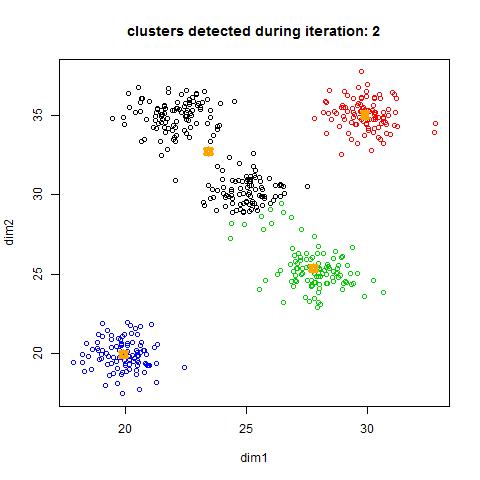


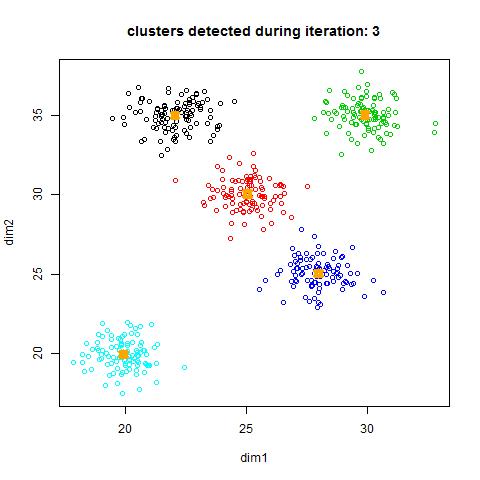




Plots generated during the iteration

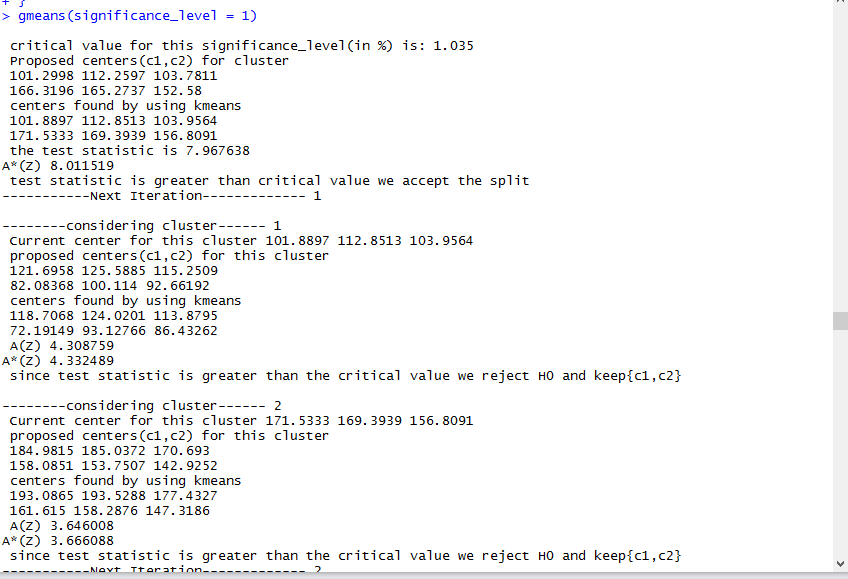


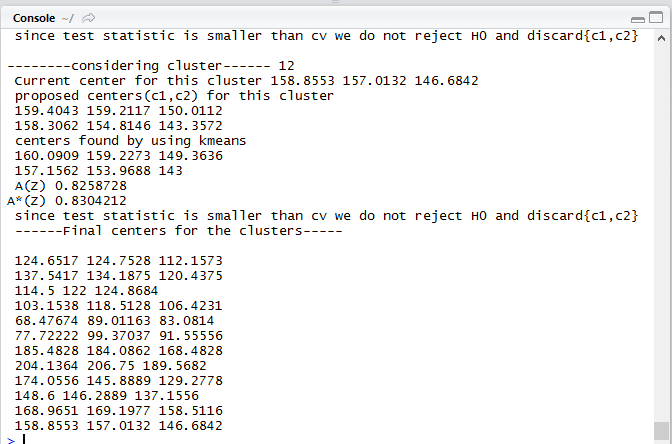


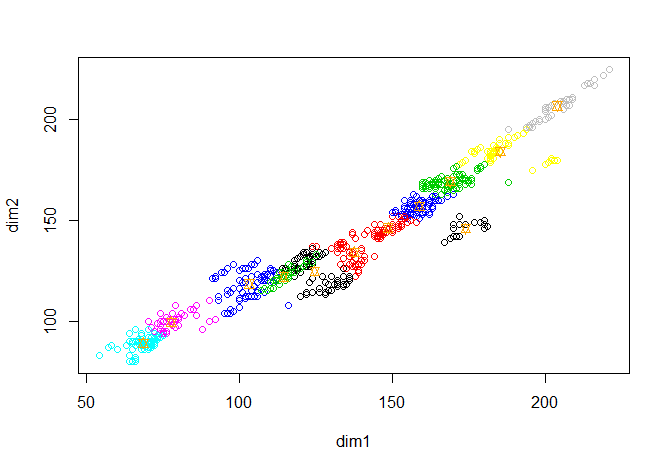


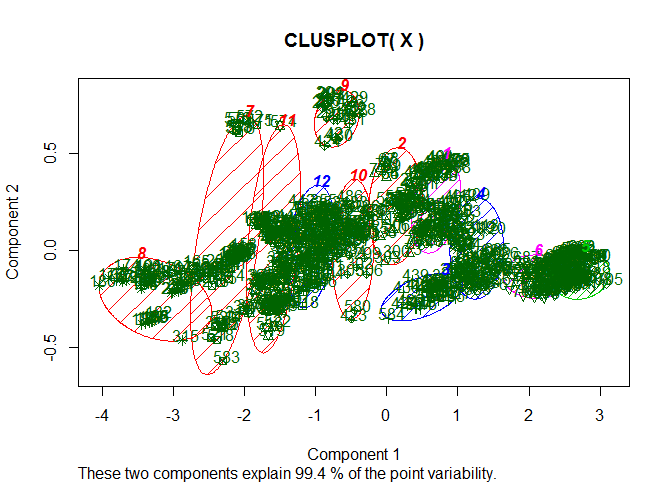
**Gmeans algorithm gives 12 clusters on the test data when we select 1% significance level. The critical value for which is 1.035.**

**Screenshots for the sample output for significance level:1**



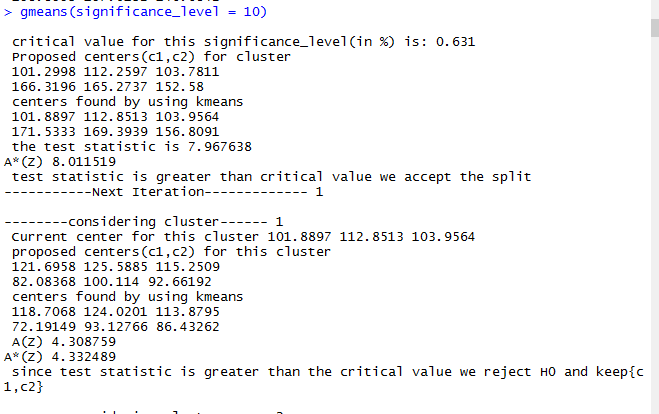


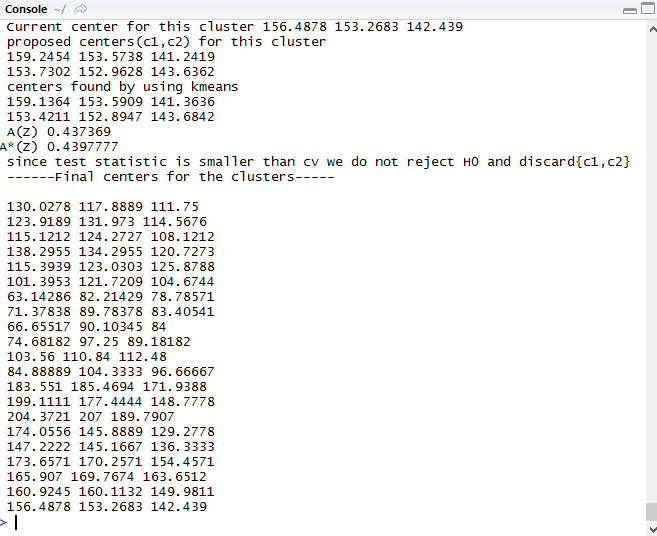


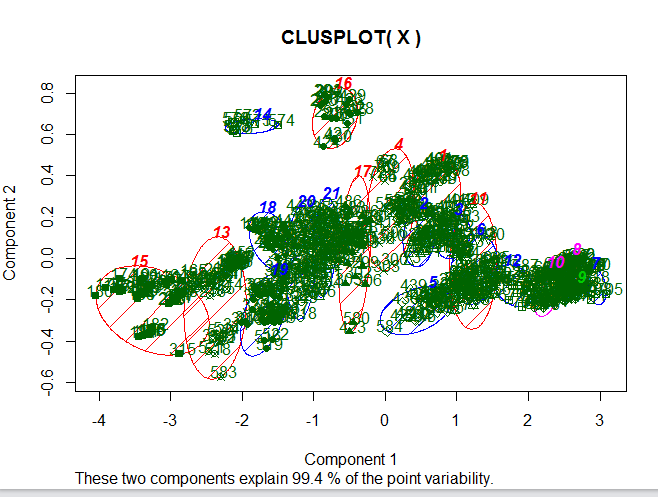
****

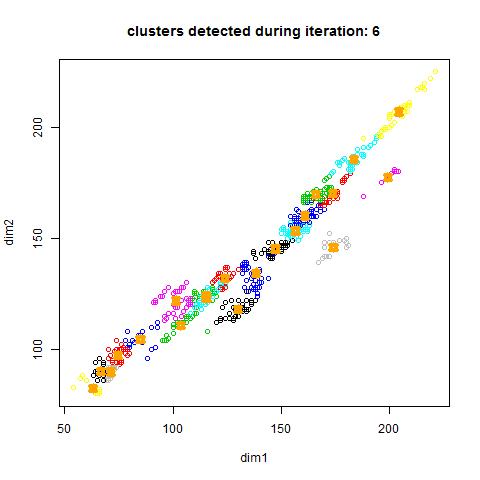
**Sample output :2**

**Significance level 10%**



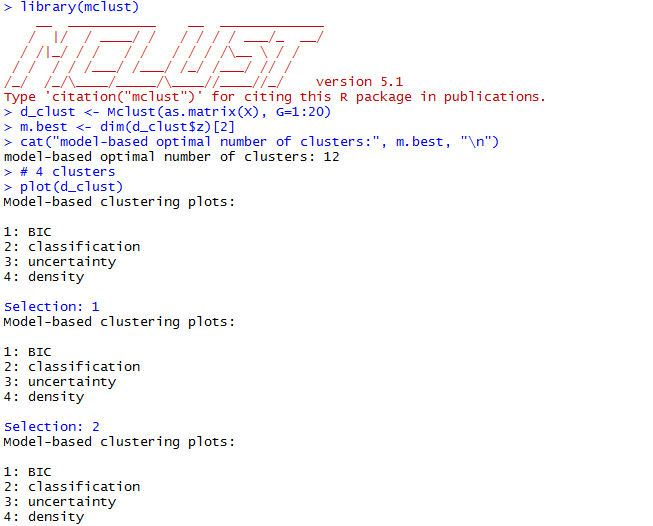




****

**The clusters generated during the iterations have been attached in the zip file and have been labelled accordingly.**

**Checking other clustering algorithms :**



**Optimal number of clusters suggested by model based clustering is 12.**