**Cluster**

張家倫

Q36034138

**環境與程式語言**

RStudio, R 3.1.3, Windows 7

**R dependent library**

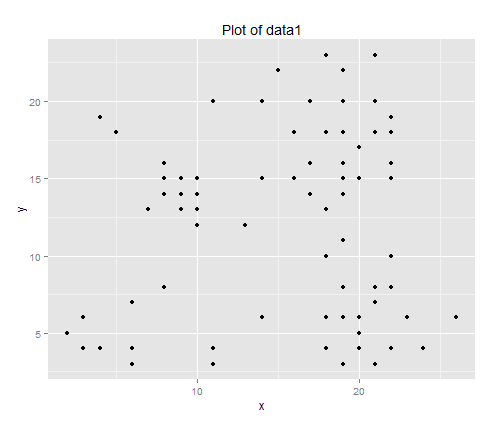
RWeka // install with following command > install.packages("RWeka")

ggplot2 // plot library

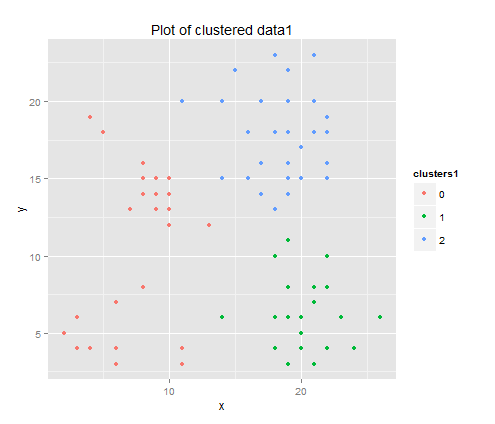
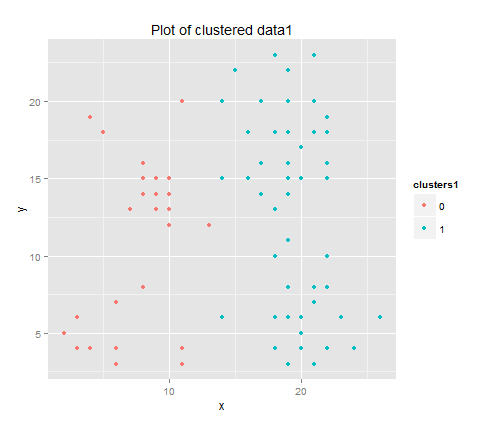
使用kmeans作為分群的演算法，要決定分成幾群，可以先將test data畫圖，再決定分成n群，將結果存成txt檔，步驟:

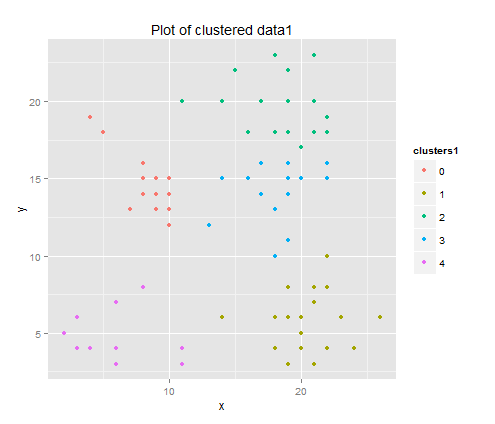
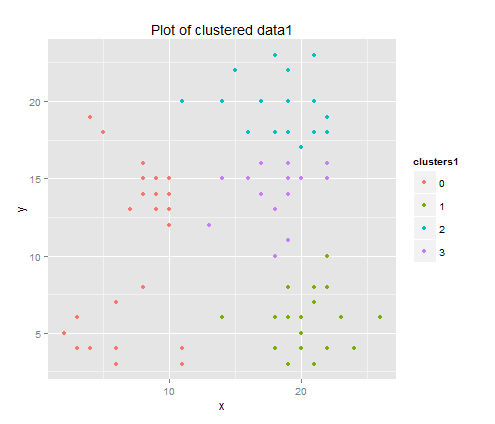
1. Plot training data set
2. 決定分成n群 (n視training data set分布情形而定)
3. 開始使用演算法分群
4. 輸出output.txt檔

* **Data set 1 - 0\_random.txt**

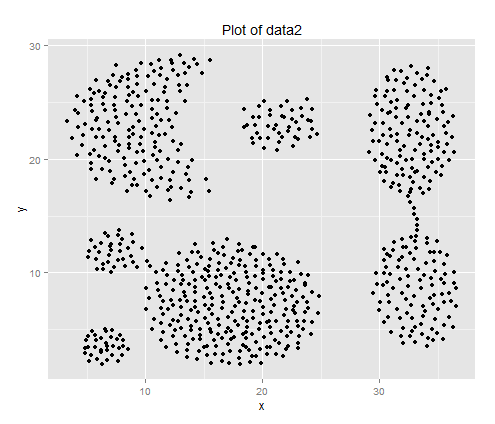


1. 決定分別分成2, 3, 4, 5群

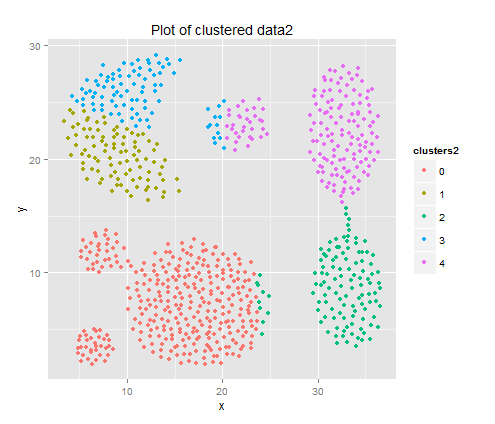
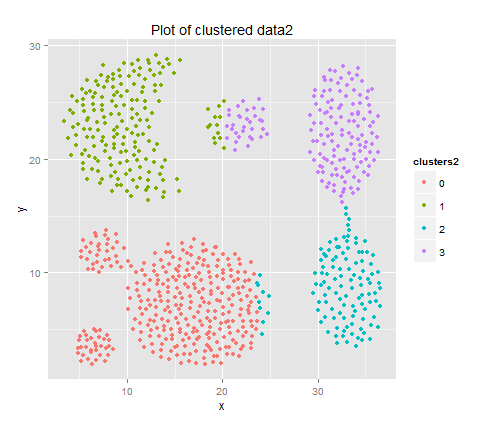


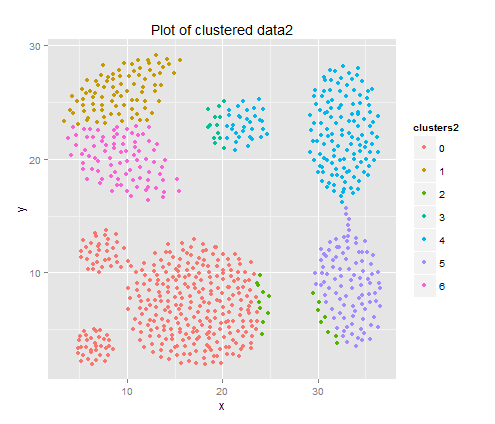
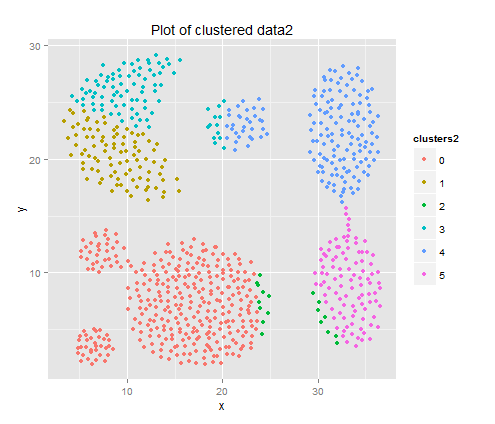


* **Data set 2 - 1\_random.txt**



1. 決定分別分成4,5,6,7群





使用kmeans來對data set 2作分群結果不太理想，可以考慮使用density-based cluster algorithm。

* **Output**

最後data set1 and data set2輸出的檔案為output1.txt及output2.txt。

* **Execute**

**>** source(“hw4.R”)