

CSCD 429-040 HW 4

Blake Chalpin

30 November 2021

Task 1

Question

Clustering: implement k-means clustering algorithm in JAVA to find six clusters from control chart data. Once the clusters are formed, extract the examples that belong to the same cluster into a .txt file. All together, your program should output six .txt files.

Solution

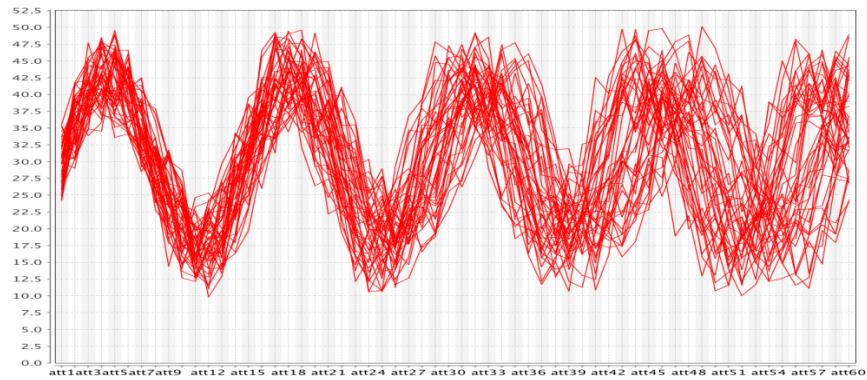
All workable Java files are under the folder `Task1/` within this zipped file. Output txt files are located in the sub-directory `Task1/cscd429-hw4/output/`

Note that the project was built and ran using Eclipse IDE. To build the run the project using Eclipse, simply open the folder `cscd429-hw4/` as a project.

Task 2

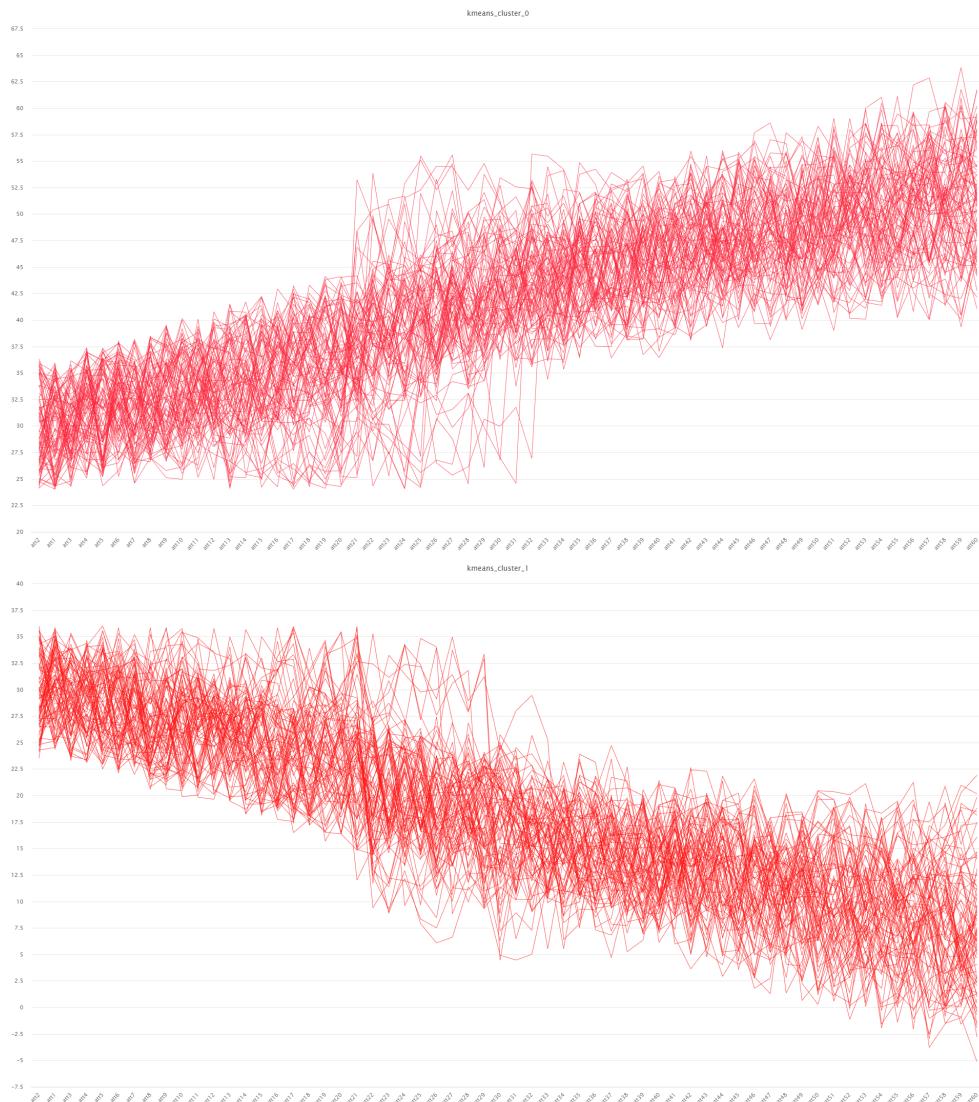
Question

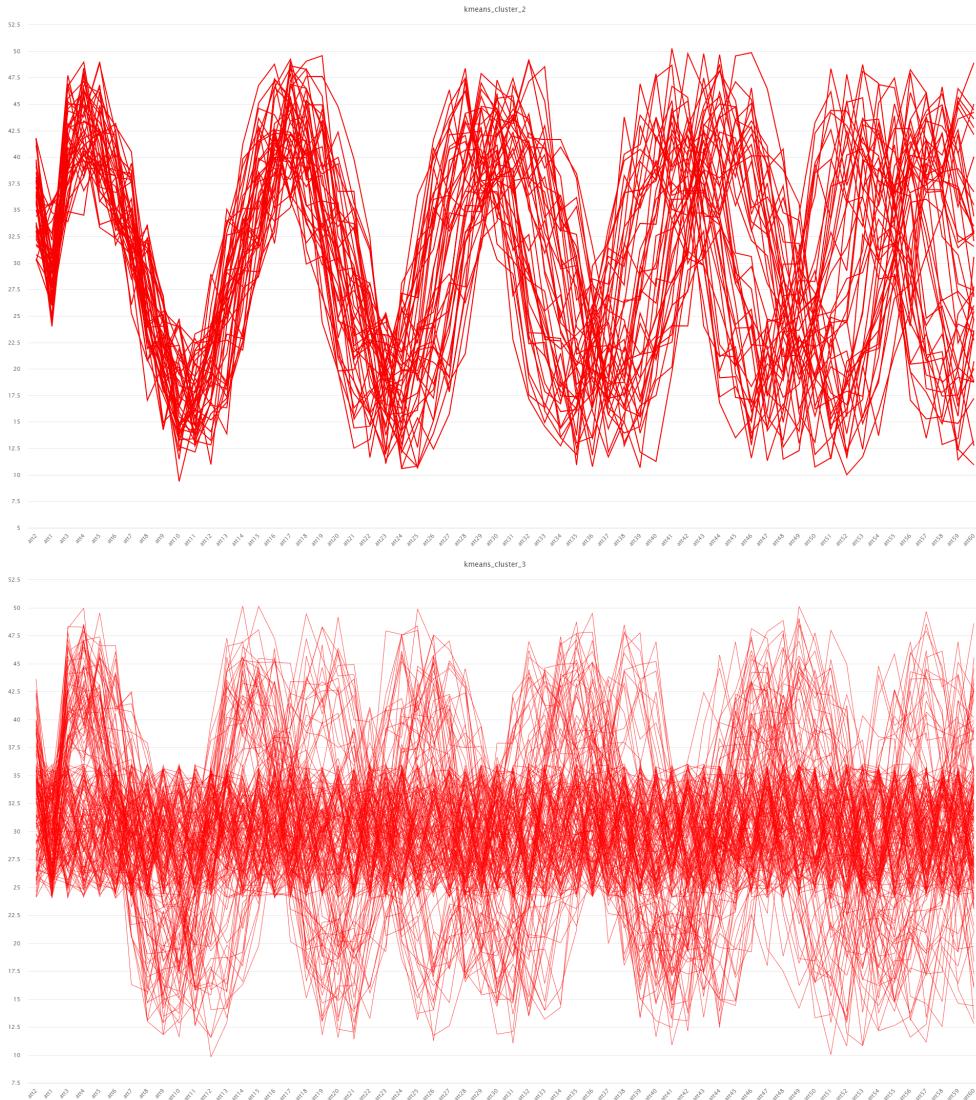
Visualization using RapidMiner: Use appropriate “chart view” to visualize the six clusters found from the previous step. As an example, the following graph is the visualization of one cluster using RapidMiner

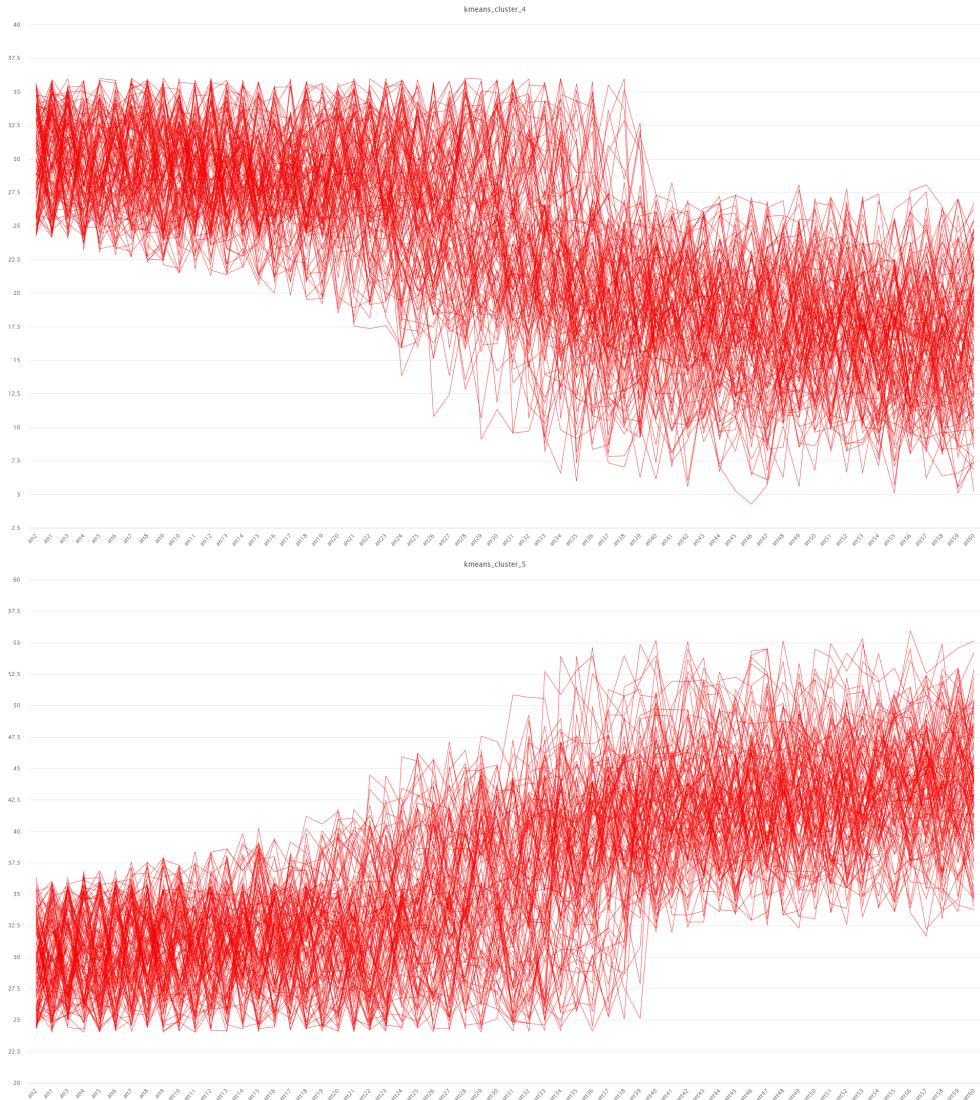


Solution

All RapidMiner cluster visualization output files are under the folder **Task2/** within this zipped file.







Task 3

Question

Clustering and Visualization using R: use R to generate six clusters from control chart data, and use R to visualize the six clusters.

Solution

Workable R script files and output captures are under the folder Task3/ within this zipped file.

R script:

```
# Libraries
install.packages("ggpubr")
library(ggpubr)

install.packages("factoextra")
library(factoextra)

# Load data
data <- read.table("synthetic_control_data.txt", header = FALSE)
str(data)

# Use kmeans to cluster the data
clustered_data <- kmeans(data, centers = 6)

# Plot the clusters
fviz_cluster(clustered_data, data = data,
             palette = c("#2E9FDF", "#00AFBB", "#E7B800",
                         "#c21e1e", "#7b6aa7", "#d2893d"),
             geom = "point",
             ellipse.type = "convex",
             ggtheme = theme_bw())
```

R cluster visualization output:

