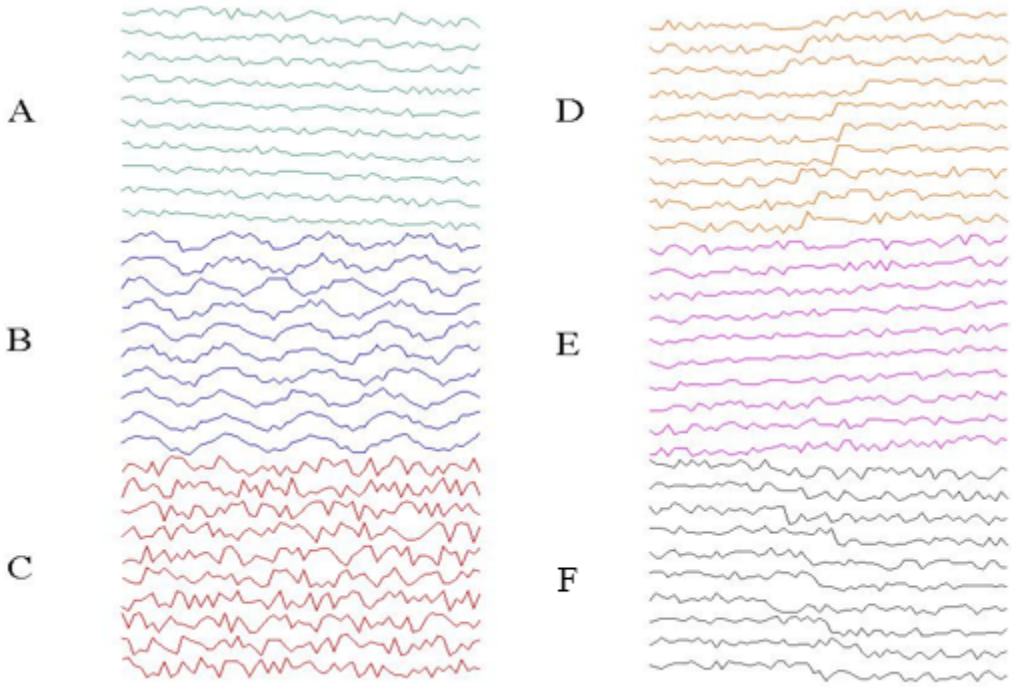
CSCD 429 Data Mining HW3 (40 points)

Clustering the control charts

- **Data Description:** The dataset *synthetic_control_data.txt* contains 600 examples of control chart time series data. The data is stored in an ASCII file, 600 rows, 60 columns, with a single chart per line. There are six different classes of control charts:
  - Normal
  - Cyclic
  - Increasing trend
  - Decreasing trend
  - Upward shift
  - Downward shift

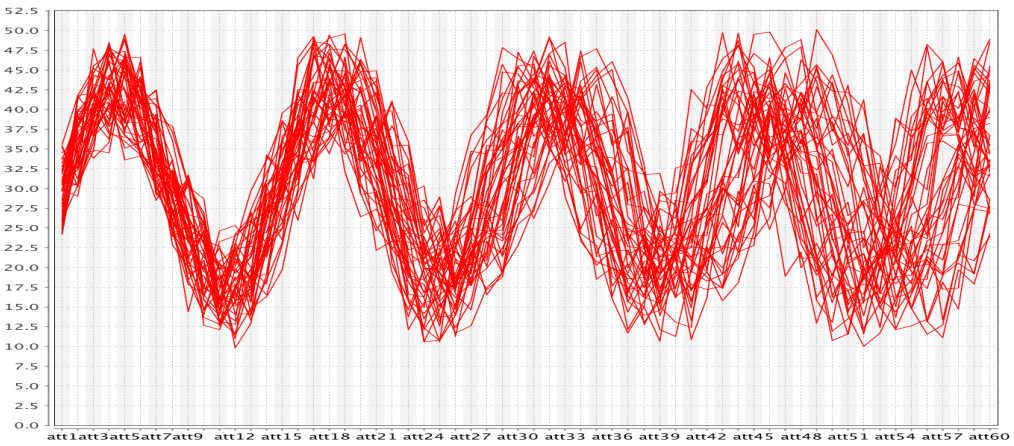
  The following image shows ten examples from each class: (A) Downward Trend. (B) Cyclic. (C) Normal. (D) Upward Shift. (E) Upward Trend. (F) Downward Shift.



Task Description:

1) **Clustering**: implement **k-means** clustering algorithm from scratch using 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.

2) **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.



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

Deliverables:

- (25 points) Workable program files and result files for Task 1.
- (5 points) Six images generated for Task 2 using RapidMiner.
- (10 points) Workable R code and result images for Task 3.
- Include all the files into a single .zip file and **submit your file via Canvas.**