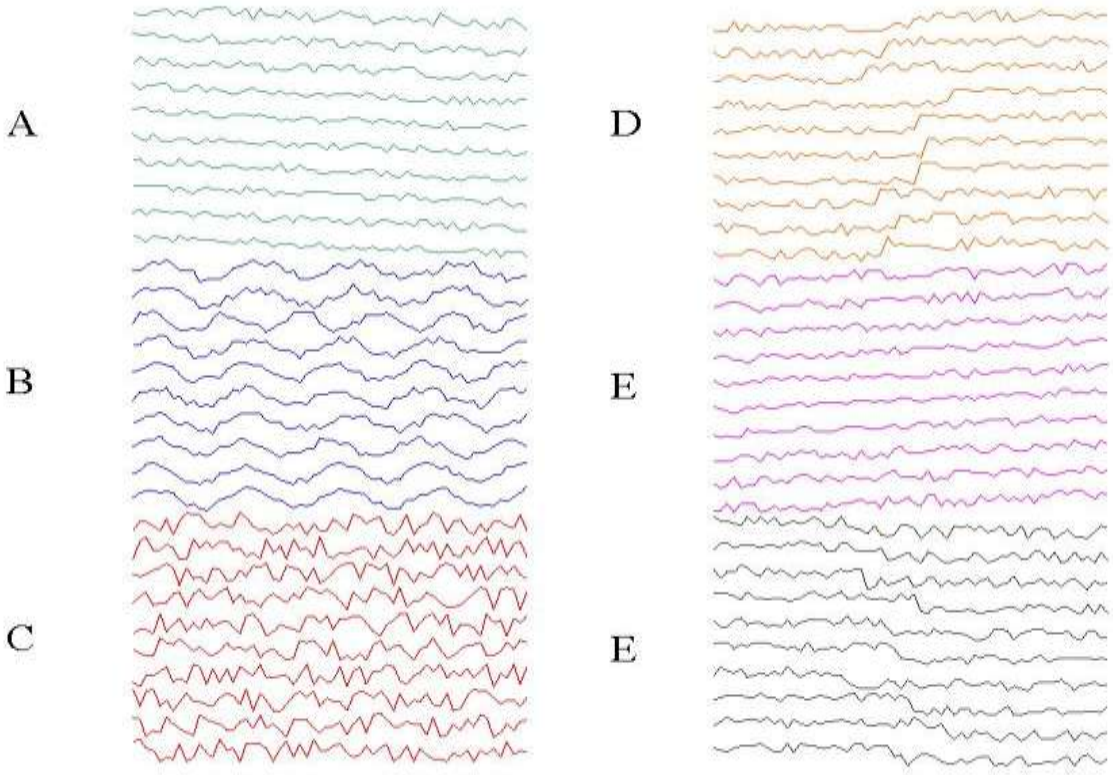
CSCD 429 Data Mining HW4 (30 points + 10 Extra)
Due: 11:59pm on March 13, 2015

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) Implement **k-means** clustering algorithm in **JAVA** to find **six** clusters in the data.
        a. Generate a result file in the format of <**record/row number, cluster number**> in each row. For example, <457, 2> means the 457th control chart from the input file belongs to the 2nd cluster.
        b. Extract all the examples that belong to the same cluster and save the examples into a CSV file. All together, your program should generate six CSV files.
    2) Use RapidMiner to visualize the results generated by your program. Use six different graphs to represent the six clusters you have discovered in task 1 (like the images shown above).
        a. Create a new process in RapidMiner, read one CSV file each time, and use the **appropriate** "Charts View" to visualize the cluster.
        b. Copy the image and paste it into your report.
    3) (**Extra**) In addition to use RapidMiner to visualize the clusters, develop your own JAVA program to show the visualization images of six clusters.
- **Deliverables**:
    - (15 points) Workable program files.
    - (5 points) The result file in the format of <**record/row number, cluster number**> in each row. For example, <457, 2> means the 457th control chart from the input file belongs to the 2nd cluster.
    - (5 points) Six CSV files.
    - (5 points) A report that includes six cluster images generated by RapidMiner.
    - (**10 Extra**) The images generated by your own JAVA program.
    - Include all the files into a single .zip file and **submit your file via Canvas.**