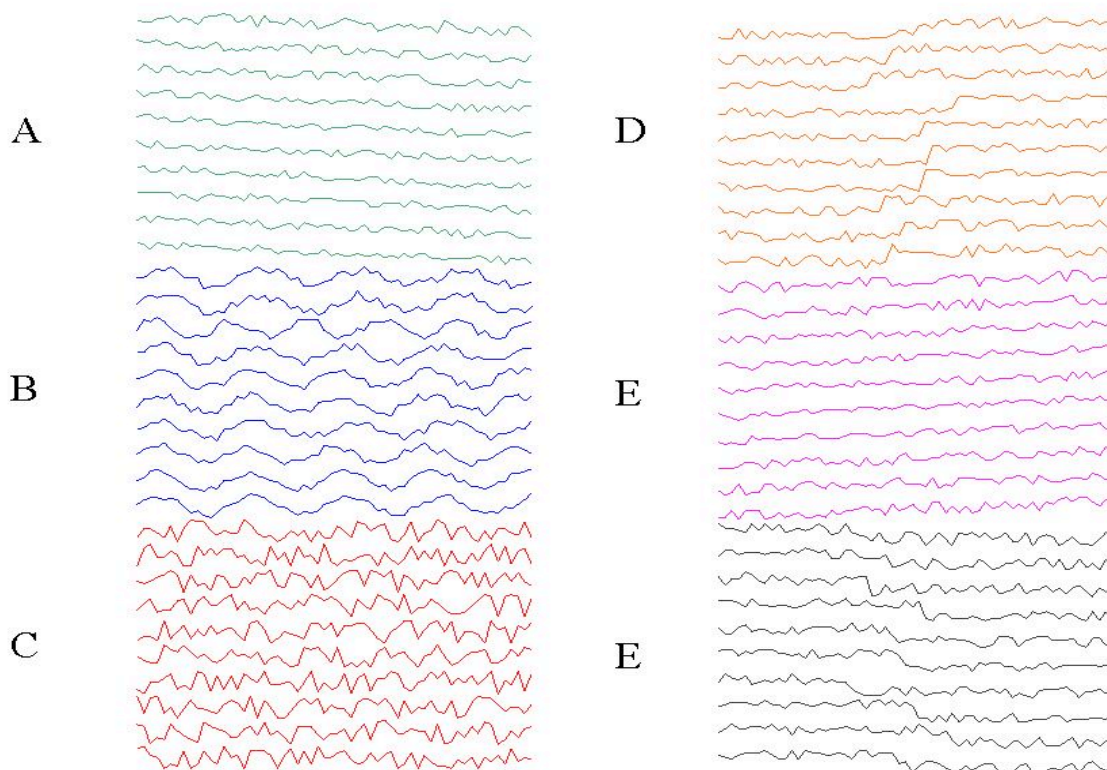


CS 445 Data Mining HW4
Due: Nov 30, 2010

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) Programming: Implement **k-means** clustering algorithm to find six clusters in the data.
 - 2) Use **k-means** module included in **RapidMiner** to generate six clusters.
 - a. Load the input data *synthetic_control_data.txt* using Import → Data → Read CSV (uncheck the box: use first row as attribute names)
 - b. Add k-means module using Modeling → Clustering and Segmentation → K-Means (set k to 6, and use the default values for all other parameters)
 - c. Generate output file using Export → Data → Write Excel (The input of this module should come from the output of k-means module. Choose the appropriate one!)
 - d. Specify the name of your excel file. After you click the run button, a new file with the specified name will be created.
 - 3) Programming: Visualize the results generated by your own program and the RapidMiner, respectively. Use six different colors to represent the six clusters you discovered in task 1 and task 2 (like the image shown above).
- **Deliverables:**
 - (5 points) Your own program files.
 - (10 points) Two result files in the format of <record number, cluster number> in each row. For example, <457, 2> means the 457th control chart from the input file belongs to the 2nd cluster.
 - (10 points) Two snap shots showing your visualization results.
 - Include all the files into a single .zip file, name it as ***YourLastName.zip*** and **submit your file using VISTA.**