Code running instruction:

It is important to note that although there are some datasets which we generated in some of our codes and then used them in others, one can generate all of those. In order to make sure that all the codes running fine without dependency to each others, we put all of the required dataset into a dataset folder for the judges.

1) File Finalcodefile.R

for this file, all you need to change some path and file name. the instruction for the changes are included in the beginning of the file. also, during the execution, it might need to install some r packages. the code might take about 30min to run depending on the computer.

2) File Cluster.R

This file contains all the codes for running clustering analysis and social network analysis. It contains a "path1" variable which should be set to data directory. Seven other datasets should be loaded to this code as well as the three original dataset. Also, one can run code "Finalcodefile.R" to get seven transition matrices which are used for social network analysis. One of these datasets "Kcode.csv" is our try to manually group error codes and the rest are transition matrices.

This file should run part by part so that one can see all the outputs. In the comments, there "PART 1" to "PART 13". If one wants run the code quickly he/she should run it part by part and see the results. Most of results are in the form of plots, graphs, and csv tables. The default directory is set for saving all the outputs. Most of the codes take few seconds to run on a normal laptop.

3) File SPatterns.R

All the codes for sequential pattern mining technique is in this file. This file is straight forward to run. Just change the "path1" to data folder and run the whole code in one run. The outputs in this code are mostly R tables. It is good to mention that because of sophisticated function we used in this code, it takes almost 30 minutes to run the whole code for different time span on a regular laptop. In order to give an estimation about the time we used a print function to show the increment of the for loop inside our function.

4) File ASRM.R

This file contains the whole codes for association rule mining. Output of this file is mostly graphs and R tables. In term of running time, you can run the whole code within 5 minutes in a regular laptop.

5) Association test (Chi-sq) and ANOVA-LSD-Boxplot

Please change directory when importing dataset. Visit_data and Pattern_data contains some new variables we produced. So use these variables to reproduce the results. Usually these codes take couple of hours to run.