**NEW JERSEY INSTITUTE OF TECHNOLOGY**

|  |
| --- |
| Data Mining |
| Assignment-2 |
|  |
| **Garudadri,Sharavan Koushik** |
|  |

**THEORY AND UNDERSTANDING BEHIND EACH QUESTION IS WRITTEN AS COMMENTS IN THE \*.R POGRAM FILES**

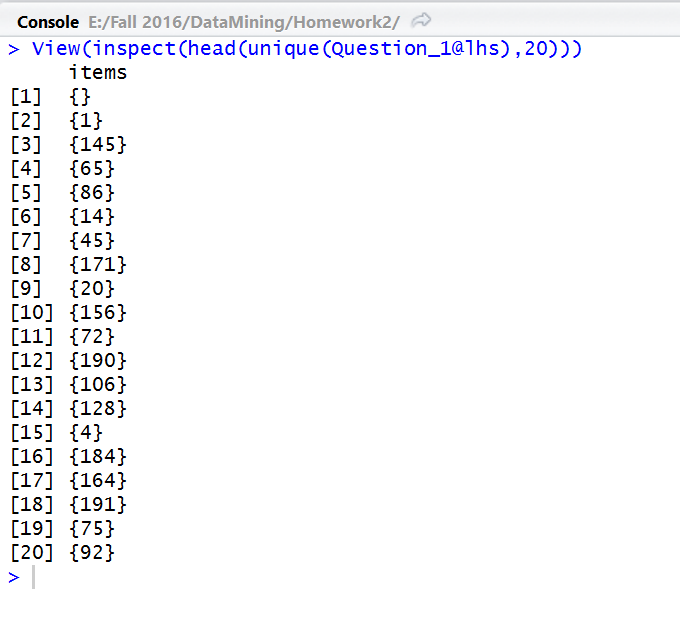
**Question 1:**

**Required Packages: arules and base packages of R**

**How to execute program:**

1. Open Rstudio
2. File->Open File , open Question\_1.R
3. Change the path of the resource files respective to the path in the machine executing the program
4. Execute each line individually or Ctrl-A and Run.
5. In the console paste the last line of File given to get top 20 patterns View(inspect(head(unique(Question\_1@lhs),20)))

**Output :**



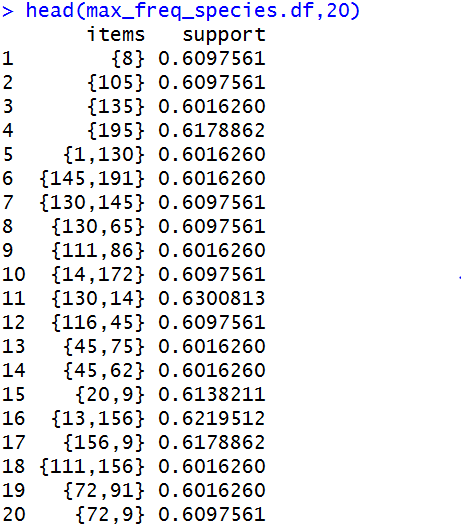
**Question 2:**

**Required Packages: arules and base packages of R**

**How to execute program:**

1. Open Rstudio
2. File->Open File , open Question\_2.R
3. Change the path of the resource files respective to the path in the machine executing the program
4. Execute each line individually or Ctrl-A and Run.

**Output displaying top 20 maximally frequent patterns:**



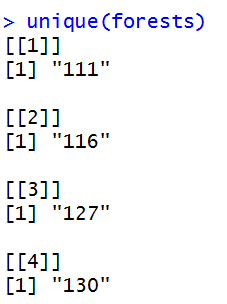
**Question3**

**Required Packages: arules,stringr and base packages of R**

**How to execute program:**

1. Open Rstudio
2. File->Open File , open Question\_3.R
3. Change the path of the resource files respective to the path in the machine executing the program
4. Execute each line individually or Ctrl-A and Run.

**Output**:



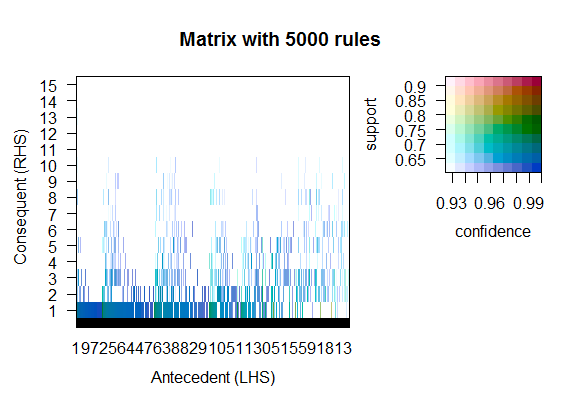
**Question4**

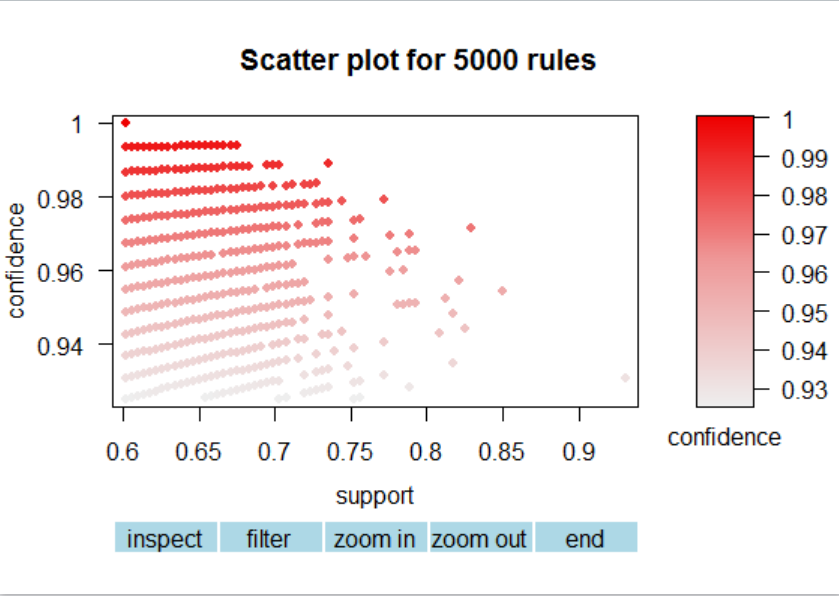
**Required Packages: arules and base packages of R**

**How to execute program:**

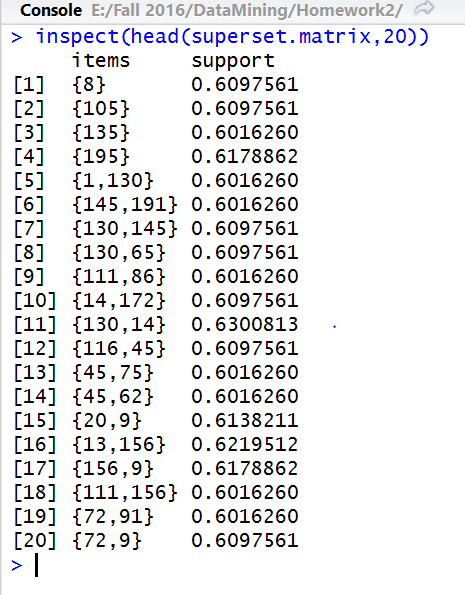
1. Open Rstudio
2. File->Open File , open Question\_4.R
3. Change the path of the resource files respective to the path in the machine executing the program
4. Execute each line individually or Ctrl-A and Run.
5. Click Finish on the plot division in the Rstudio to end the interactive graph , which is used to display how itemsets are scattered over support by confident ratio.
6. Itemsets are displayed after the scatter plot is executed and a matrix plot is displayed comparing the LHS => RHS

Ouput Plots:





Top 20 itemsets are displayed as output by pasting the following snippiet in console



**Question5**

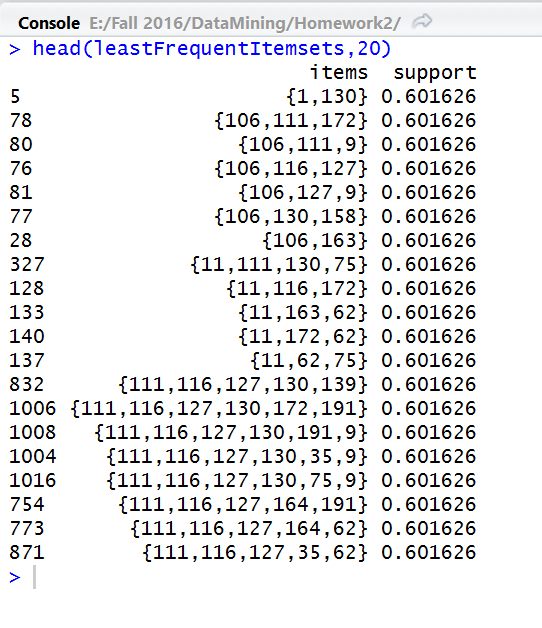
**Required Packages: arules and base packages of R**

**How to execute program:**

1. Open Rstudio
2. File->Open File , open Question\_5.R
3. Change the path of the resource files respective to the path in the machine executing the program
4. Execute each line individually or Ctrl-A and Run.

**Output**

The top 20 itemsets which are least frequent are shown as a data.frame by pasting the following snippet in the console



The follwing plot displays few itemssets that are least common in the maximally frequent patterns

