**Data Visualization Lab(L13+L14)**

**Name: Alpanshu Kataria**

**Reg. No.: 18BCE1267**

**Faculty: [Dr.Parvathi.R.](https://lms.vit.ac.in/course/view.php?id=815" \l "section-3)**

Lab 4

**Task:**

Take a dataset and apply association rule learning.

**Solution:**

**Dataset:**

To demonstrate the association rule learning in a real biomedical case-study, I’ve used a transactional healthcare data representing [a subset of the Head and Neck Cancer Medication data](https://umich.instructure.com/files/1678540/download?download_frd=1),.It consists of inpatient medications for head and neck cancer patients.

**Link:** [http://www.socr.umich.edu/people/dinov/courses/DSPA\_notes/11\_Apriory\_AssocRuleLearning.html#72\_Step\_2\_-\_exploring\_and\_preparing\_the\_data](http://www.socr.umich.edu/people/dinov/courses/DSPA_notes/11_Apriory_AssocRuleLearning.html" \l "72_Step_2_-_exploring_and_preparing_the_data)

**Dataset sample:**



**Key Points:**

We apply an iterative approach or level-wise search where k-frequent itemsets are used to find itemsets. To improve the efficiency of level-wise generation of frequent itemsets, an important property is used called Apriori property which helps by reducing the search space. Apriori Property – All non-empty subset of frequent itemset must be frequent. The key concept of Apriori algorithm is its anti-monotonicity of support measure. Parameters

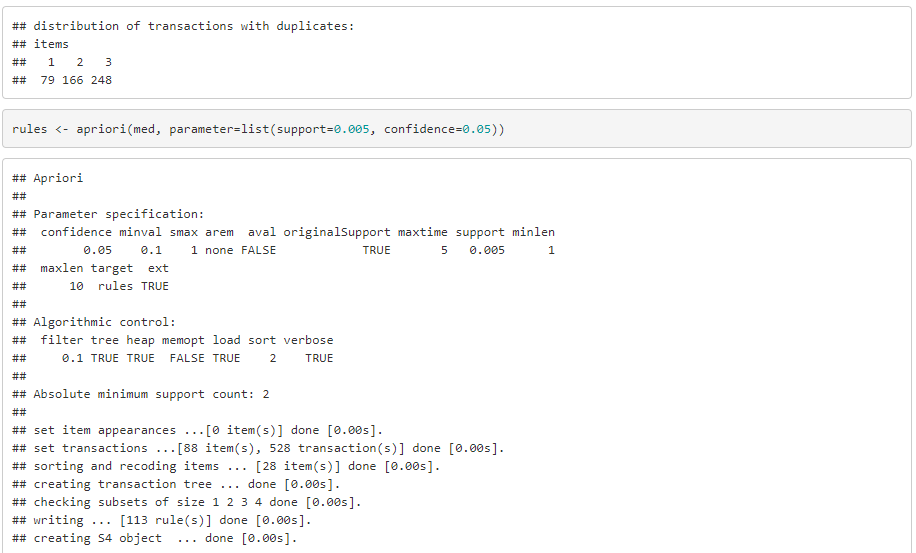
1. **Support:** This measure gives an idea of how frequent an itemset is in all the transactions.

2. **Confidence:** This measure defines the likeliness of occurrence of consequent on the cart given that the cart already has the antecedents.

3. **Lift:** Lift controls for the support (frequency) of consequent while calculating the conditional probability of occurrence of {Y} given {X}.

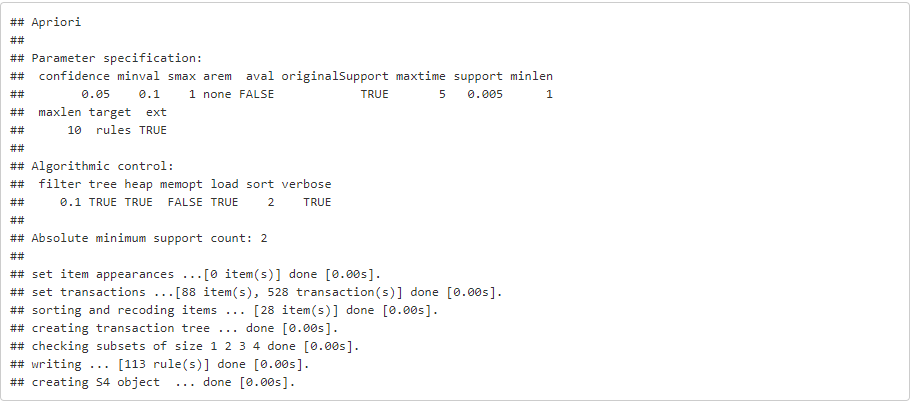
**Code:**

We’ve loaded our dataset now we’ll apply apriori algorithm with support as 0.005 and confidence as 0.05 the data





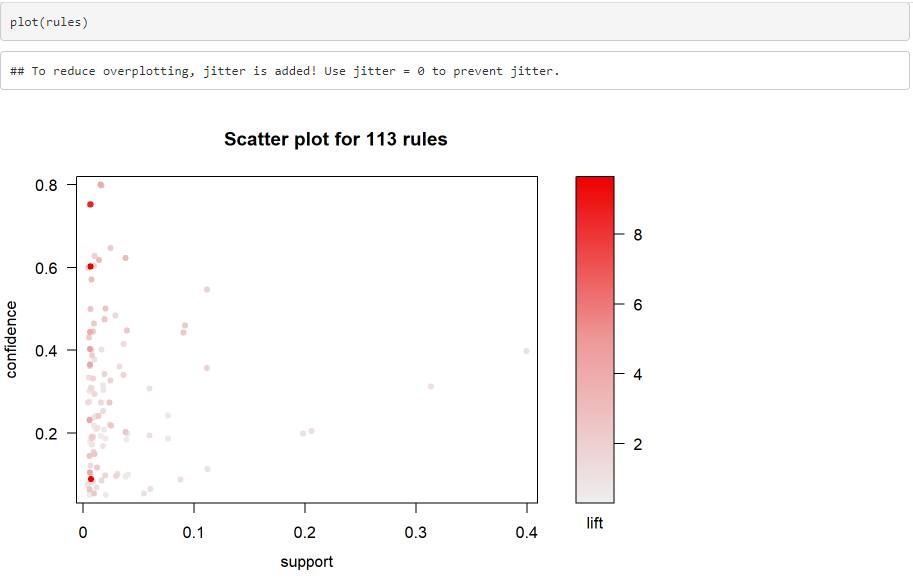
The values of support and confidence are user defined and based on these values rules are generated.





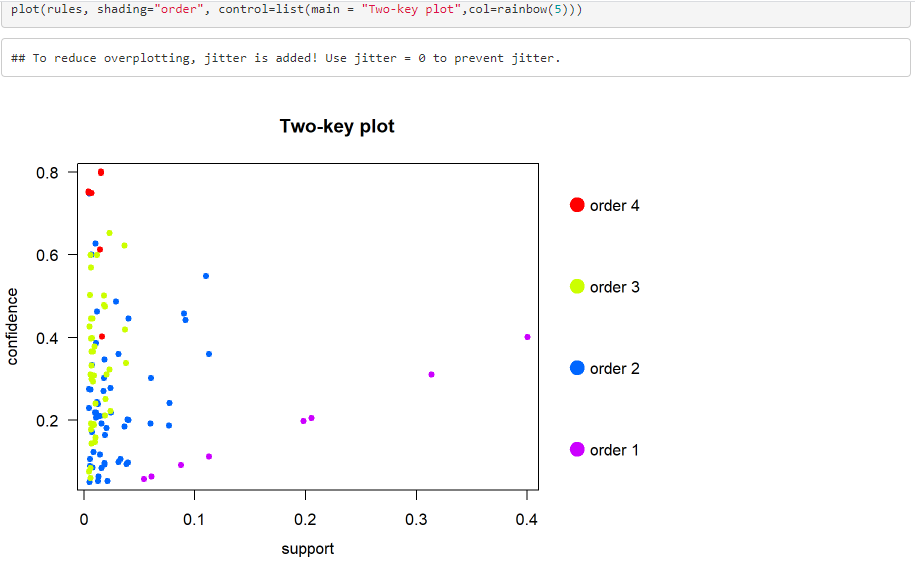
In our case 113 rules are generated.

Now we’ll plot them.



Above scatter plot shows the 113 rules generated by the apriori algorithm for our dataset with support=0.005, confidence=0.05

Now we’ll group the rules based on order.



This visualization method draws a two dimensional scatterplot with different measures of interestingness (support and confidence) on the

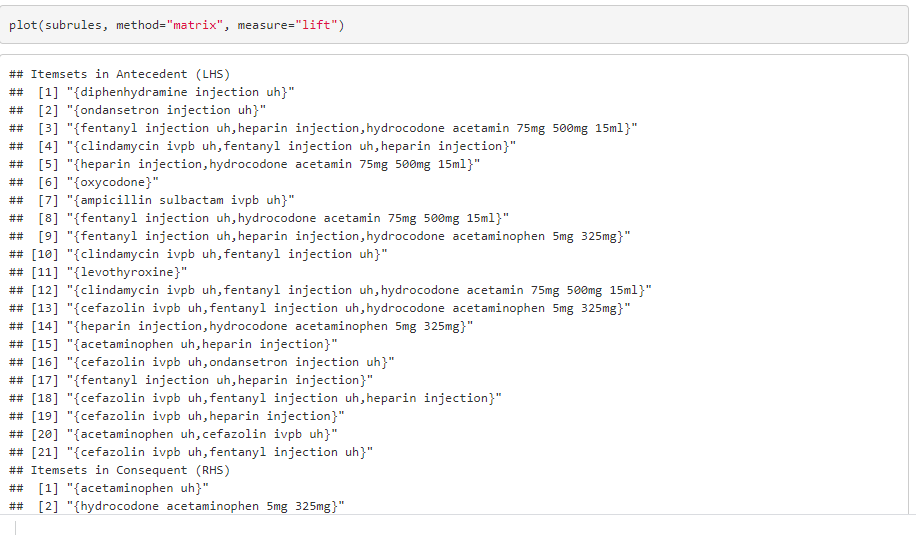
axes and a third measure (parameter “shading”) is represented by the points color. There is a special value for shading called “order”. With

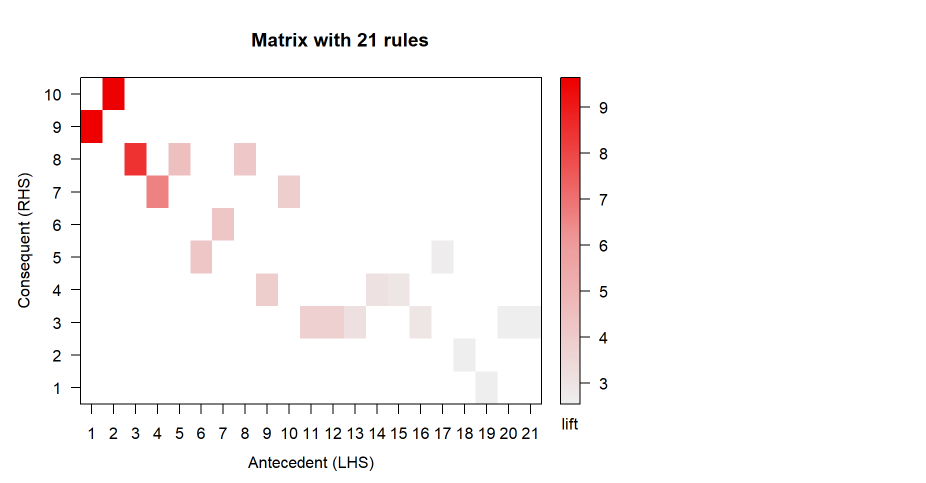
this value the color of the points represents the length (order) of the rule. This is used for two-key plots

2D matrix with shading.

The following techniques work better with fewer rules for better visualization

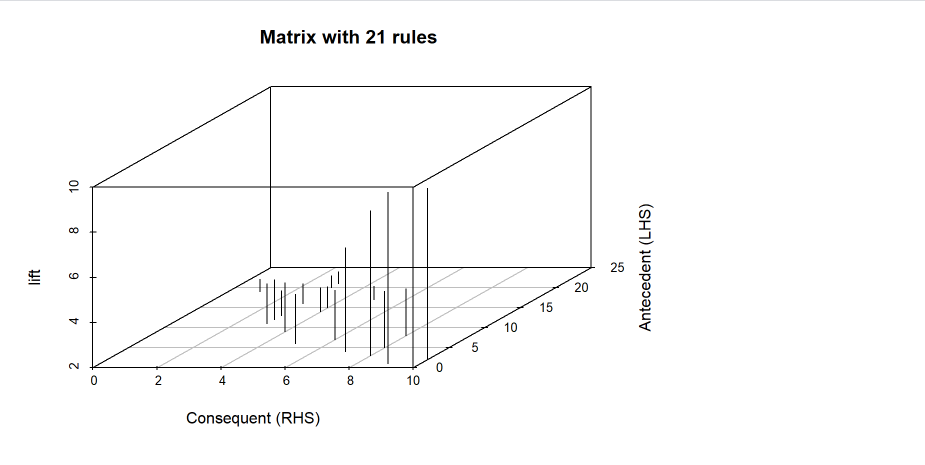
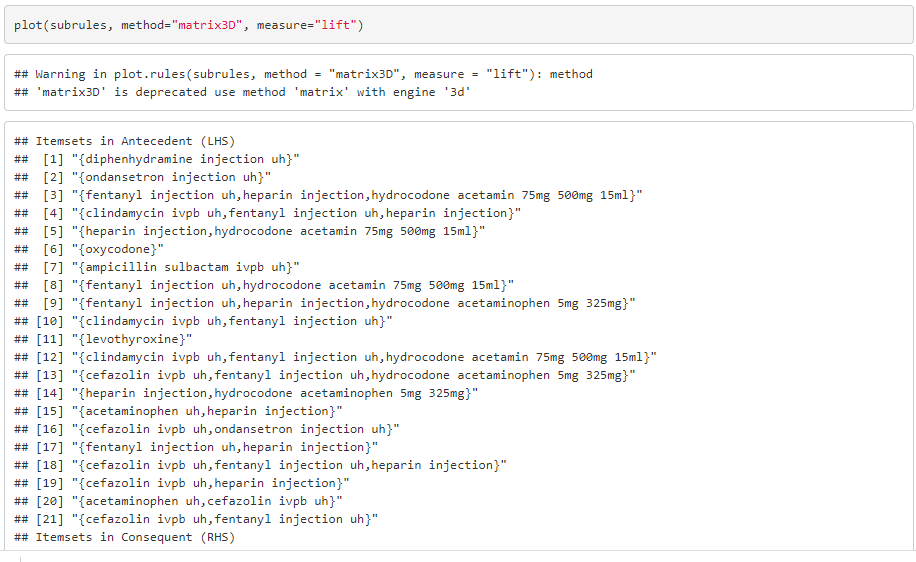




It arranges the association rules as a matrix with the itemsets in the antecedents on one axis and the itemsets in the consequent on the

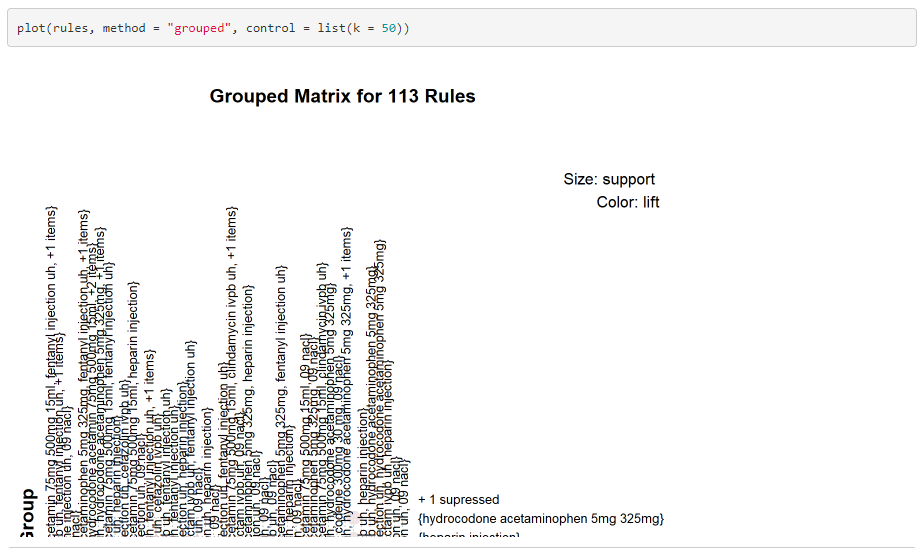
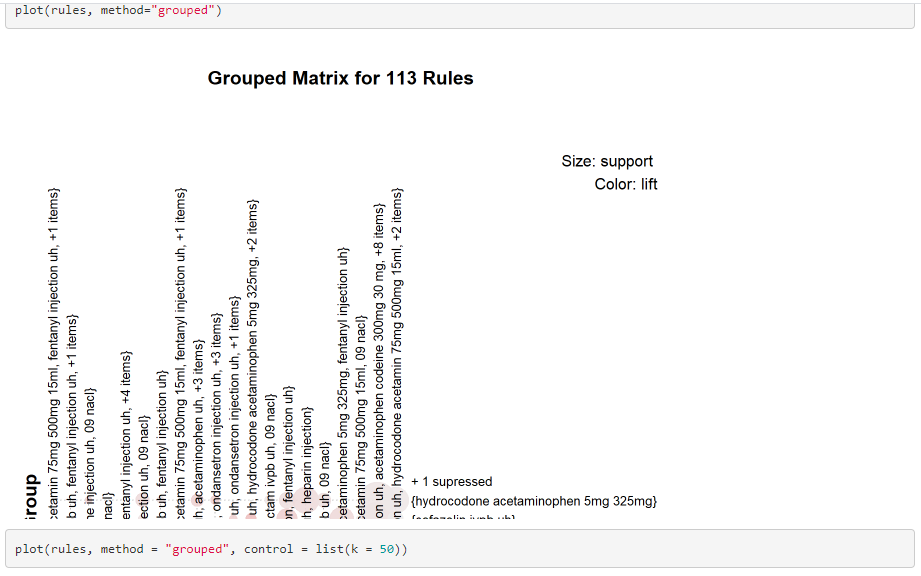
other

:3D Matrix visulization

Matrix3D Arranges the association rules as a matrix with the itemsets in the antecedents on one axis, the itemsets in the consequent on the

other and lift in the third dimension

:grouped matrix plot

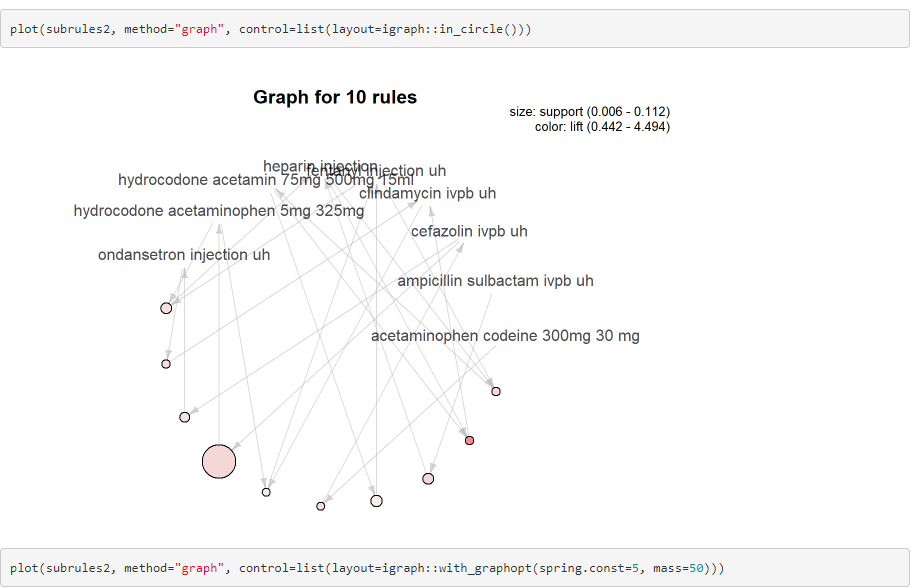
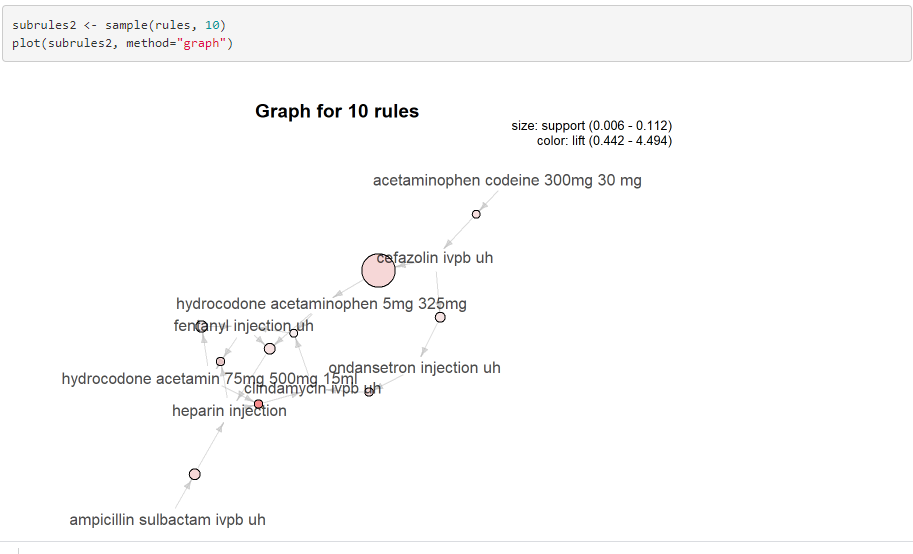


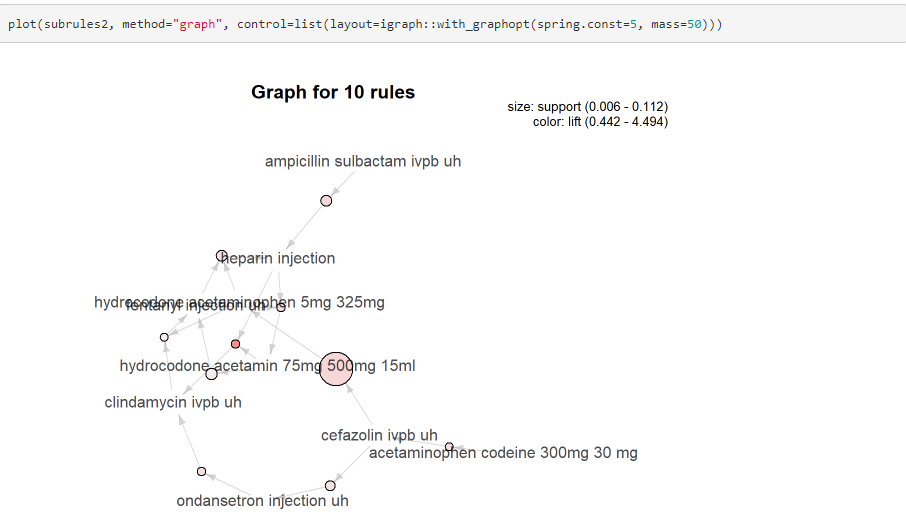
Grouped matrix-based visualization Antecedents (columns) in the matrix are grouped using clustering. Groups are represented by the most

interesting item (highest ratio of support in the group to support in all rules) in the group. Balloons in the matrix are used to represent with

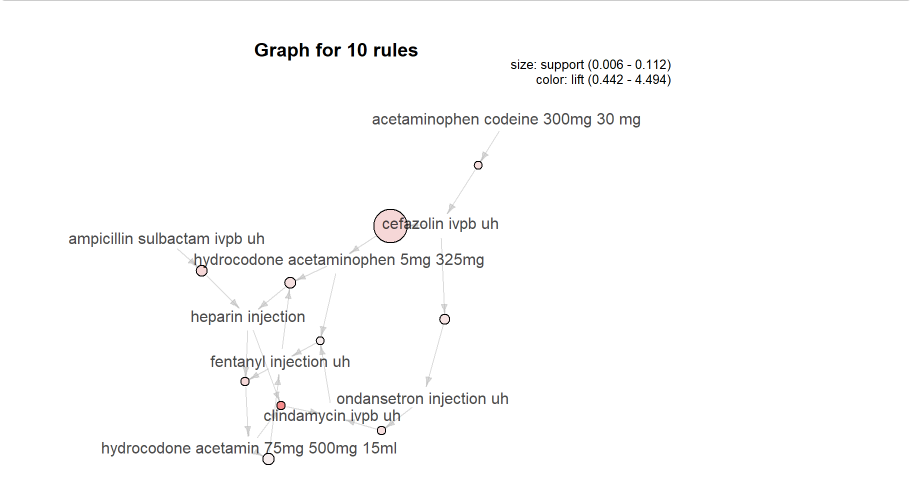
what consequent the antecedents are connected

:Visualization of rules using Graphs







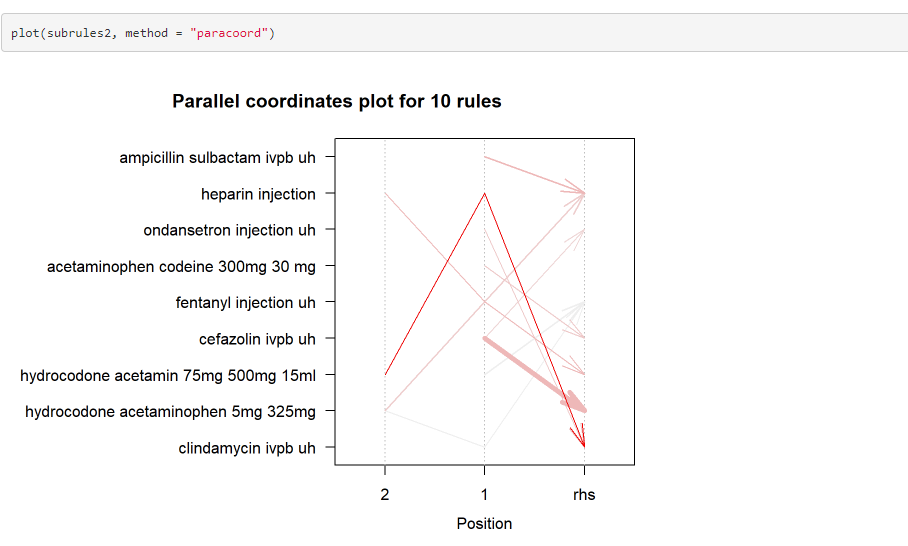


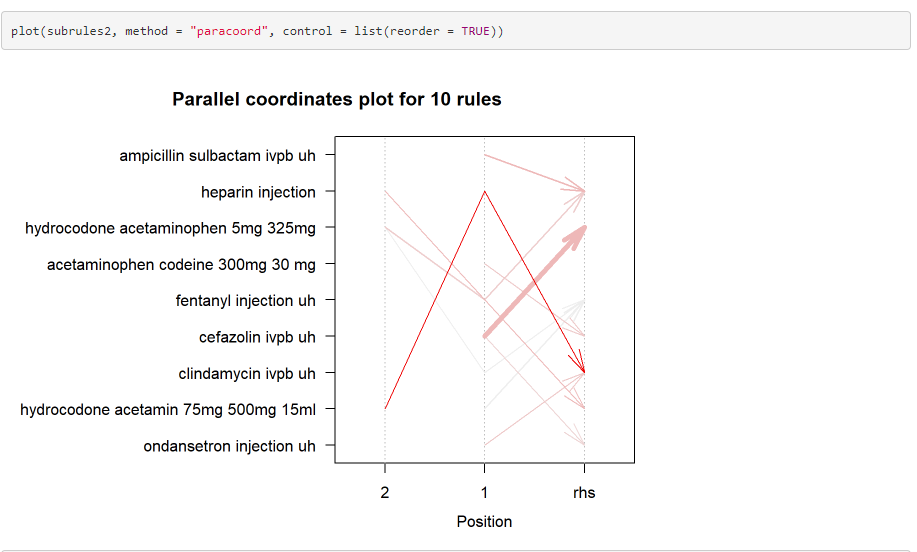
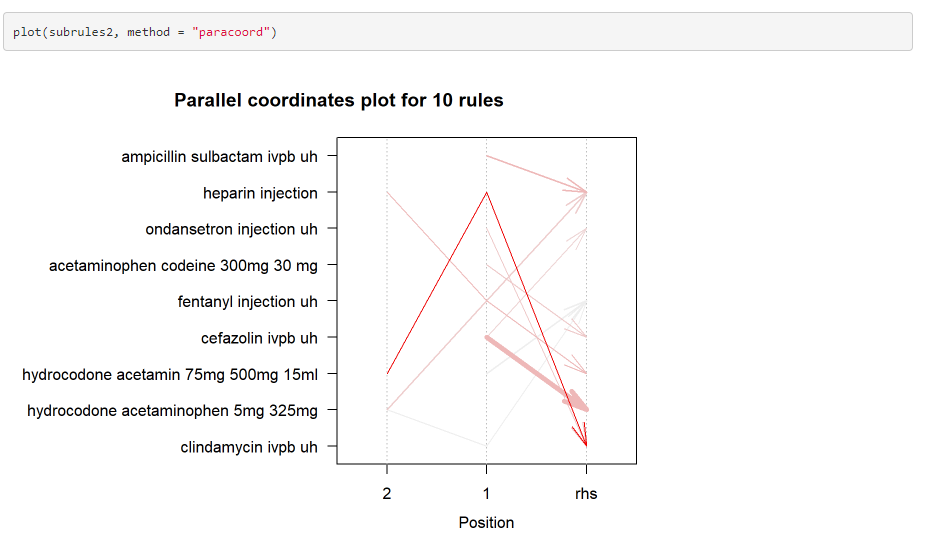
It represents the rules (or itemsets) as a graph with items as labeled vertices, and rules (or itemsets) represented as vertices connected to

items using arrows. For rules, the LHS items are connected with arrows pointing to the vertex representing the rule and the RHS has an

arrow pointing to the item

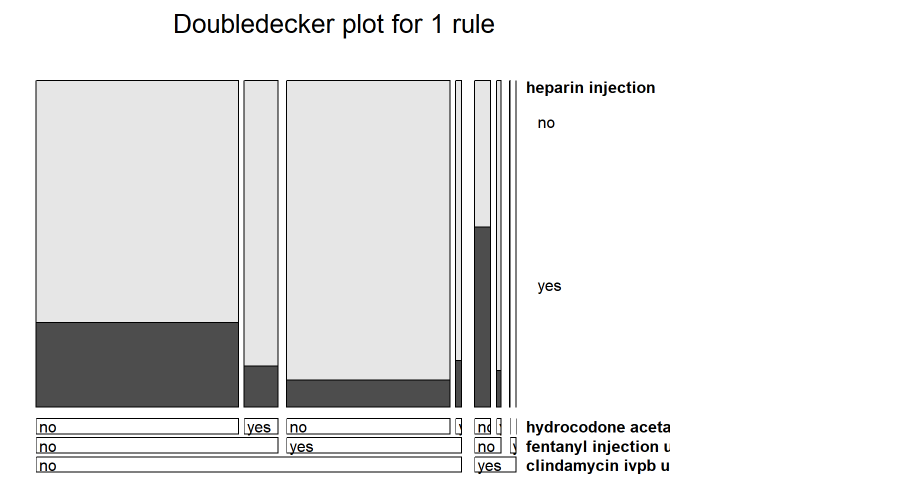
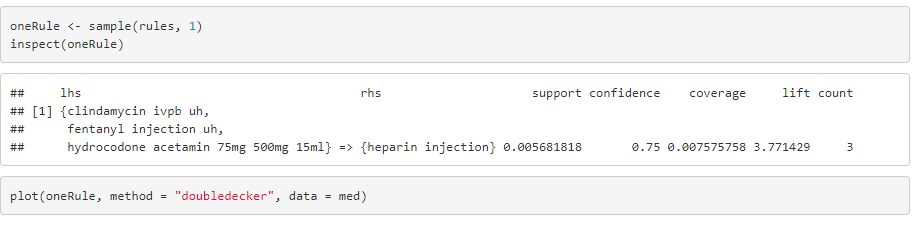
: Visualization of rules using parallel coordinates



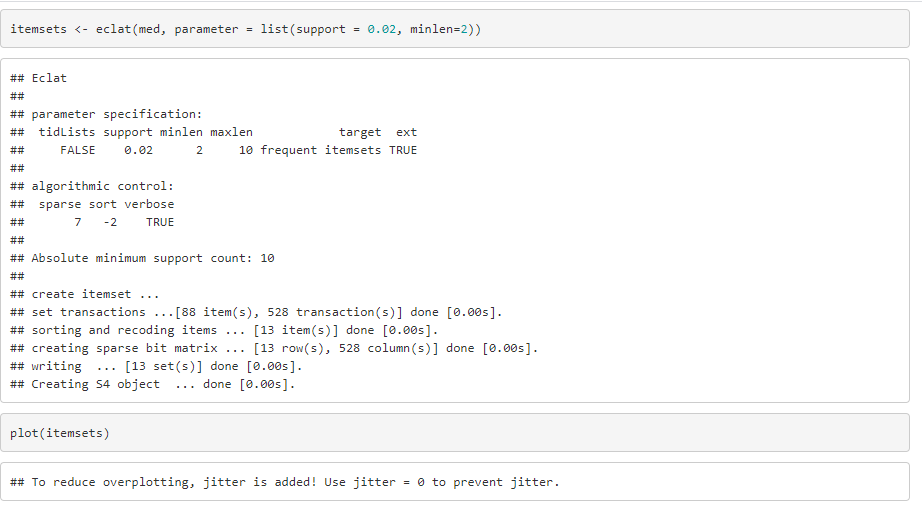


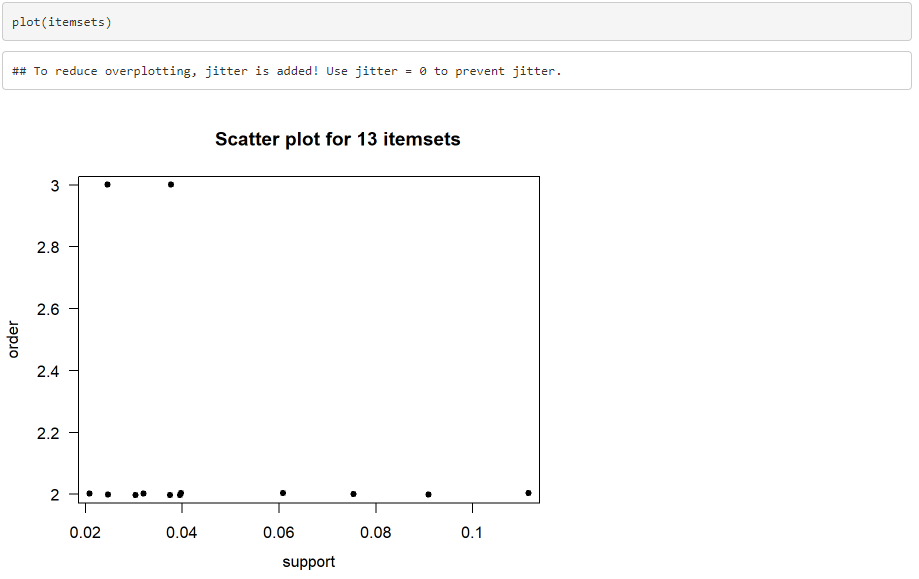
It represents the rules as a parallel coordinate plot.

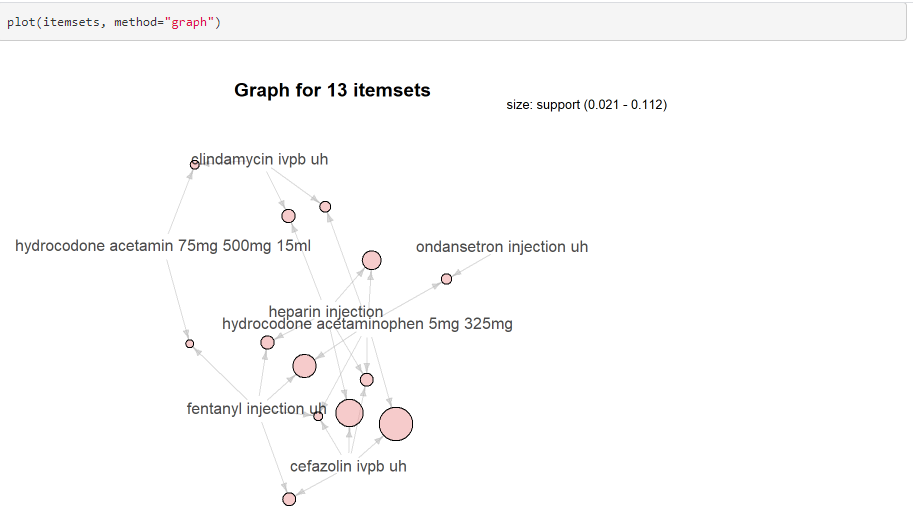
:Visualization of one rule using doubledecker plot

It represents a single rule as a doubledecker or mosaic plot. Parameter data has to be specified to compute the needed contingency table.

:Itemset Visualization as Graph



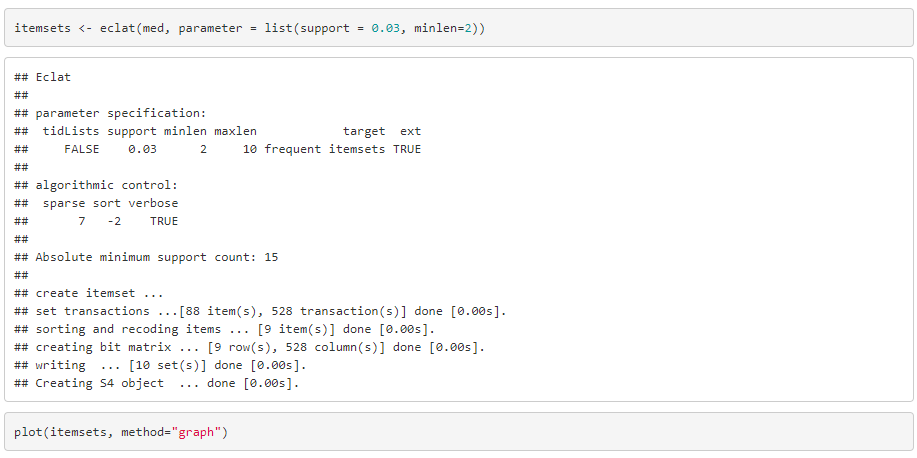


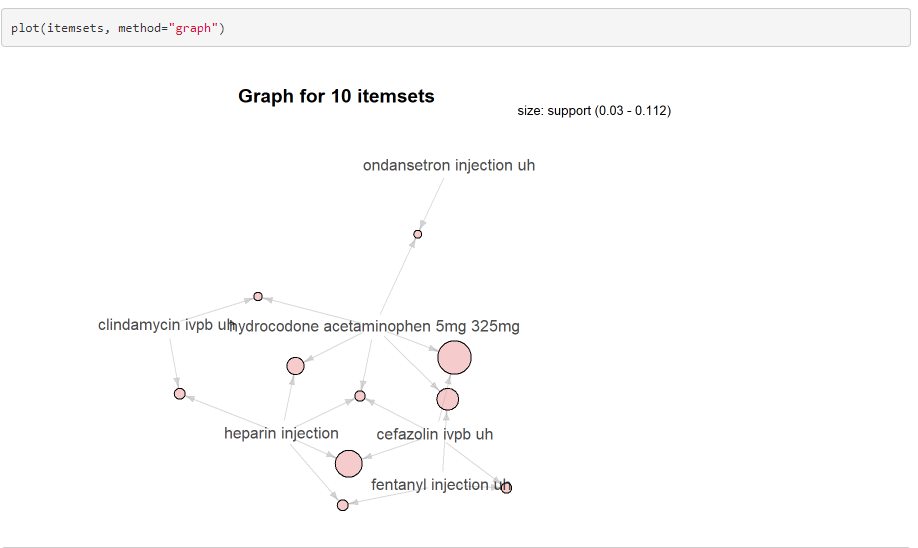


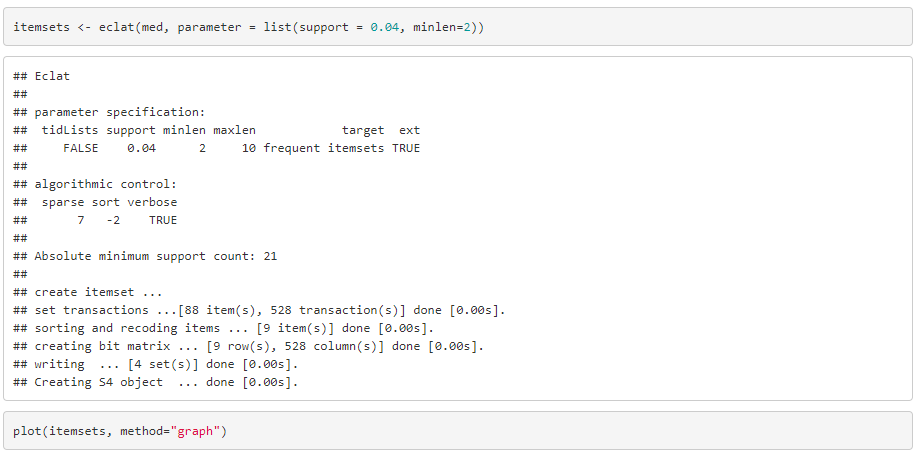
The eclat() takes in a transactions object and gives the most frequent items in the data based the support argument. The maxlen defines the

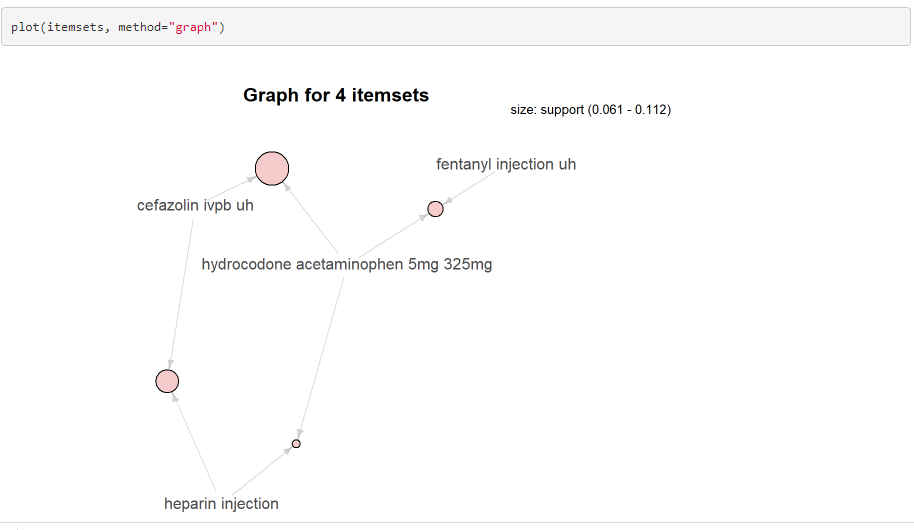
maximum number of items in each itemset of frequent items.This plot used to represent the most frequent items in the dataset as graph with

support value as 0.03 and length as 2





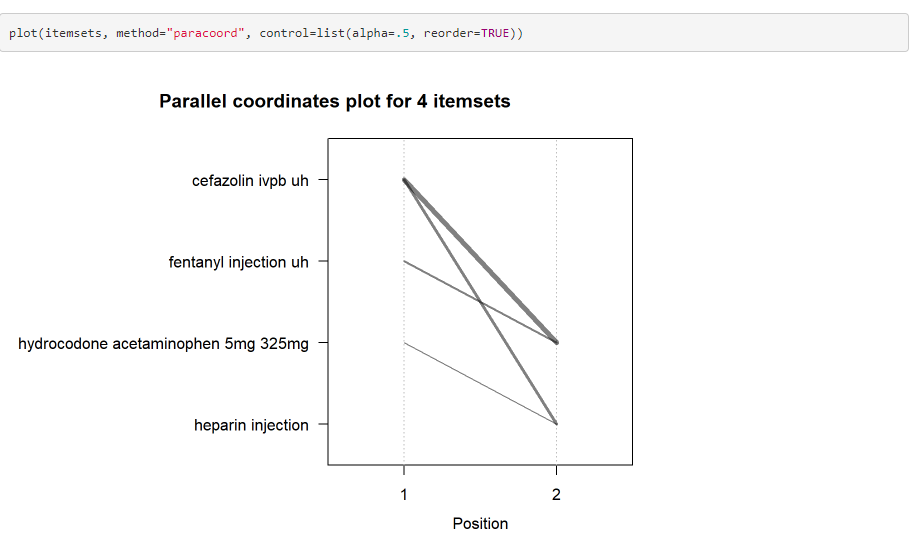


The eclat() takes in a transactions object and gives the most frequent items in the data based the support argument. The maxlen defines the

maximum number of items in each itemset of frequent items.This plot used to represent the most frequent items in the dataset as graph with

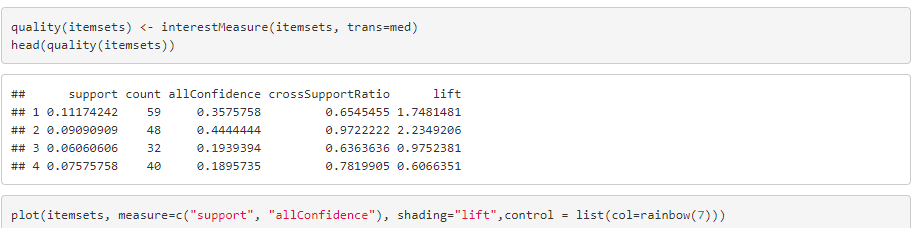
support value as 0.03,0.04 respectively and length as 2

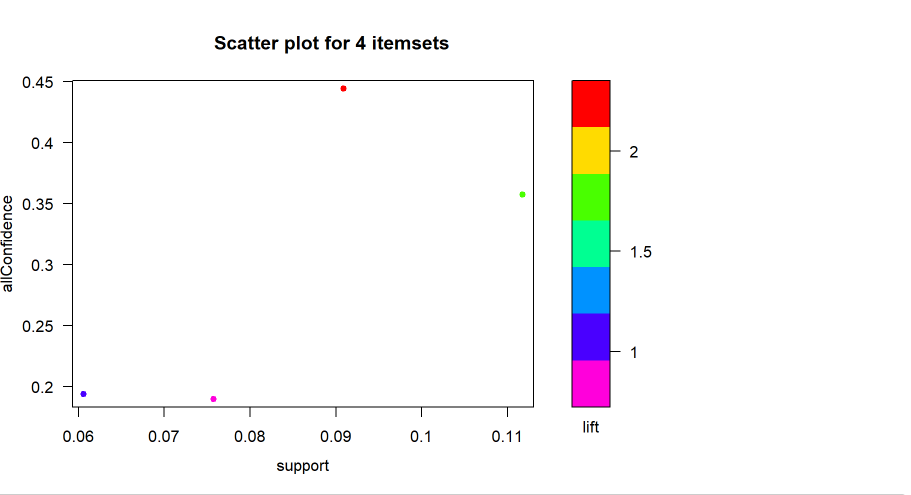
:Itemset visualization using parallel coordinates



It represents the itemset as a parallel coordinate plot.

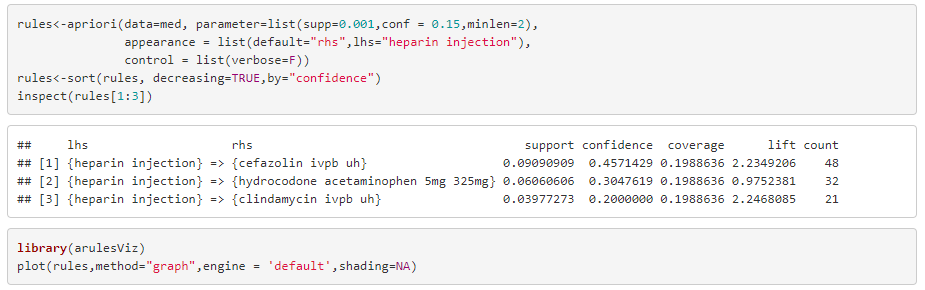
:Add more quality measures to the scatterplot

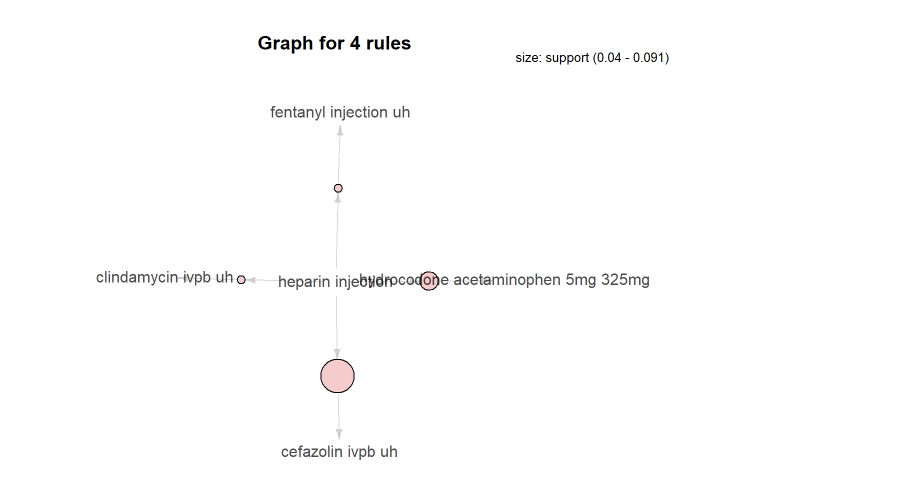




It represents the itemsets with various support and confidence values and with rainbow color using the lift parameter

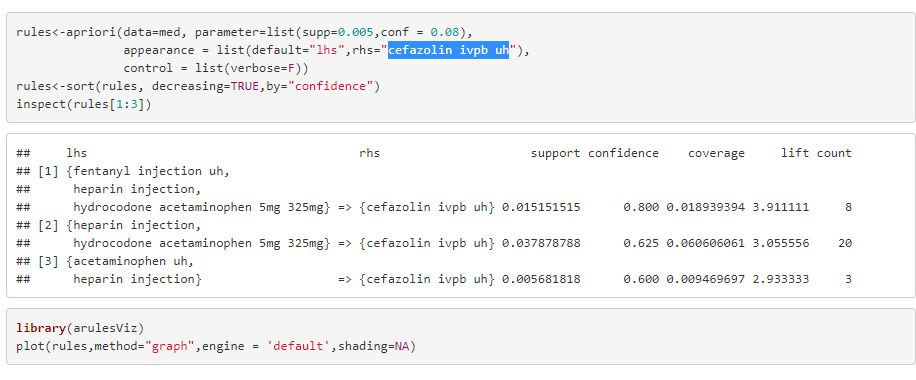
:Visulization of rules with Left hand side value as heparin injection

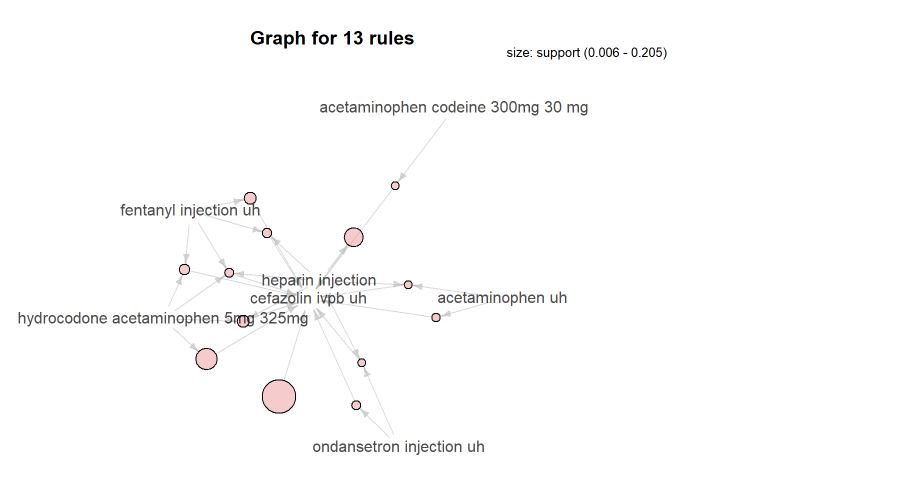




It is used find out the Customers who bought ‘Whole Milk’ also bought other item

:Visualization of rules with Right Hand Side value as cefazolin ivpb uh





It represents the what customers had purchased before buying ‘cefazolin ivpb uh’. This will help you understand the patterns that led to the purchase of ‘cefazolin ivpb uh’

:Create an item frequency plot for top 20 items

