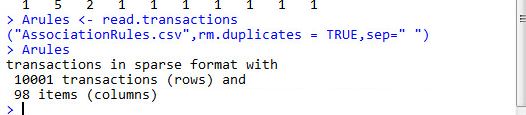
CSC570Q Data Science Essentials

LAB 4: Analytic Methods, part 2

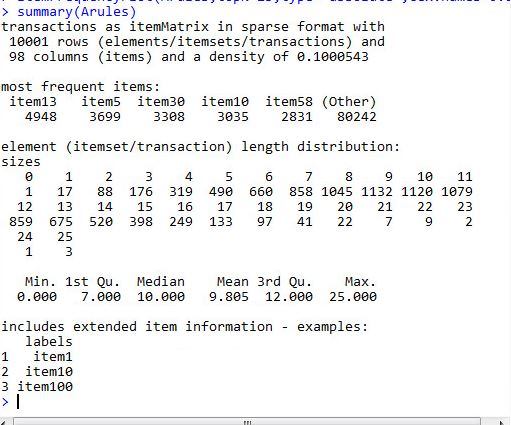
Association Rules

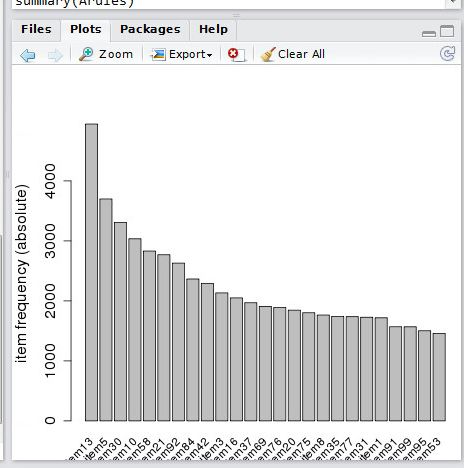
* Import the AssociationRules.csv transaction data file.



* Create a frequent item plot, and a frequent item table.

1. Determine the most frequent item bought in the store.

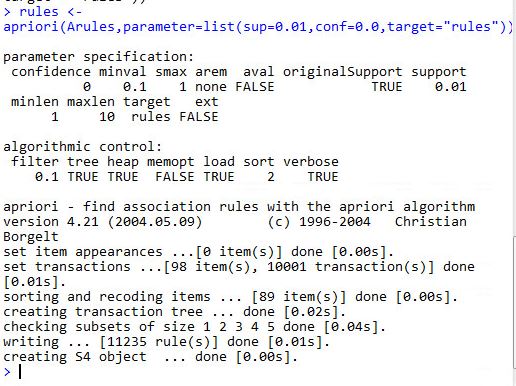




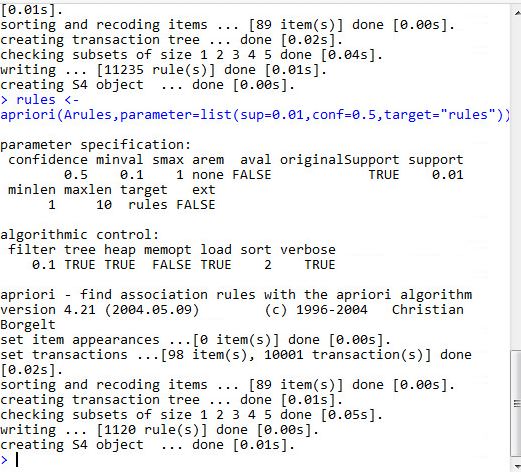
1. How many items were bought in the largest transaction?

* Mine the Association rules with a minimum Support of 1% and a minimum Confidence of 0%.

1. How many rules appear in the data?



1. How many rules are observed when the minimum confidence is 50%.

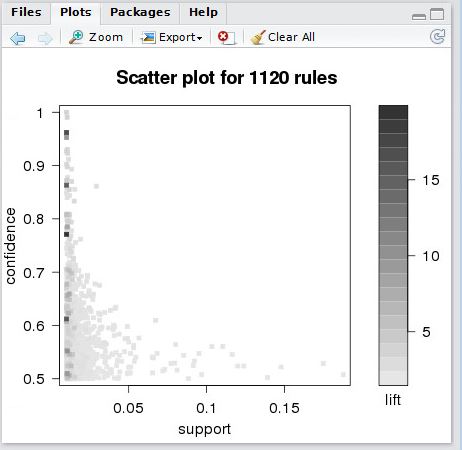


1. Explain how the specified confidence impacts the number of rules.

* Create a scatter plot comparing the parameters support and confidence on the axis, and lift with shading.

1. Identify the positioning of the interesting rules.

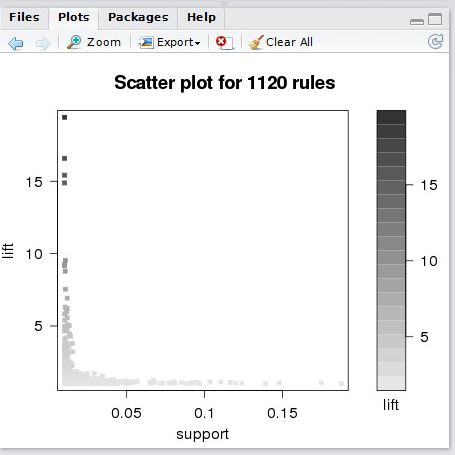
Plot(rules)

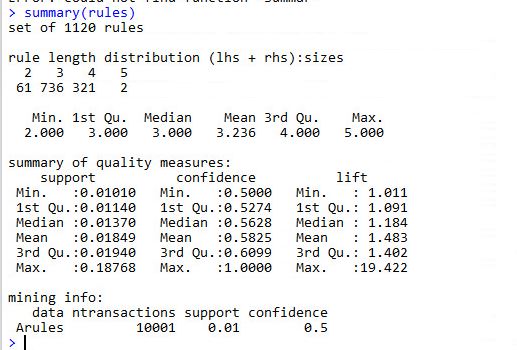


* Compare support and lift.

1. Create a scatter plot measuring support vs. lift; record your observations.

plot(rules,method="scatterplot",measure=c("support","lift"))





1. Where are the rules located that would be considered interesting and useful?

The rules plotted between the region Support =0.05 and lift=5.

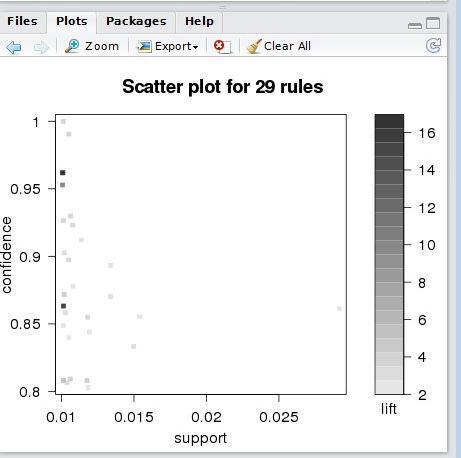
1. One downside to the Apriori algorithm, is that extraneous rules can be generated that are not particularly useful. Identify where these rules are located on the graph. Explain the relationship between the expected observation of these itemsets and the actual observation of the itemsets.
2. Using the interaction tool for a scatter plot, identify 3 rules that appear in at least 10% of the transactions by coincidence.

* Identify the most interesting rules by extracting the rules in which the Confidence is >0.8. Observe the output of the data table for the most interesting rules.

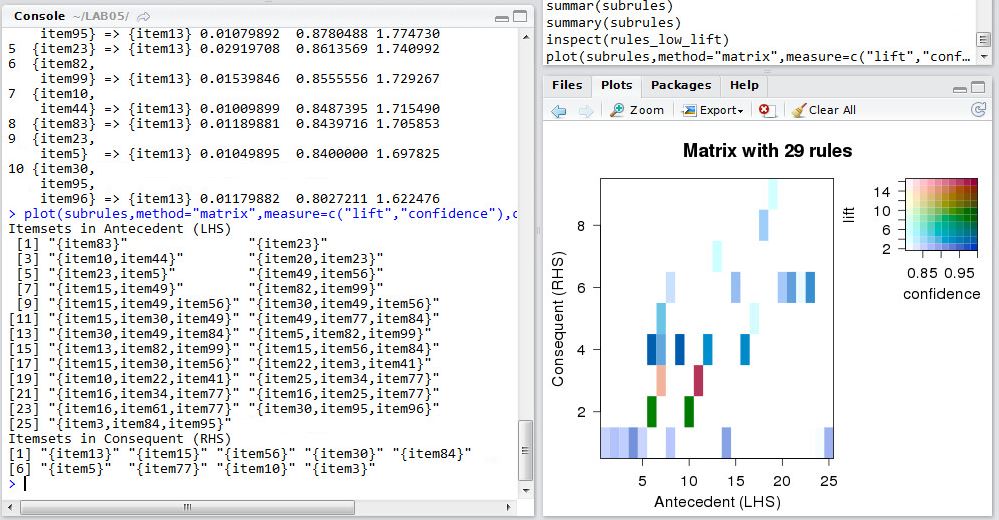
>subrules <- rules[quality(rules)$confidence>0.8]

> plot(subrules,control=list(jitter=2))

1. Sort the rules stating the highest lift first. Provide the 10 rules with the lowest lift. Do they appear to be coincidental? Why or why not?



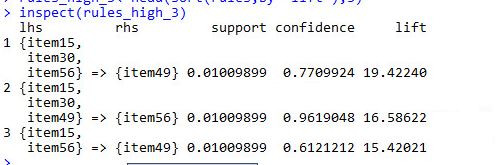
* Create a Matrix-based visualization of two measures with colored squares. The two measures should compare confidence and lift (have recorded = FALSE). Note that 4 interesting rules stand out on the graph.



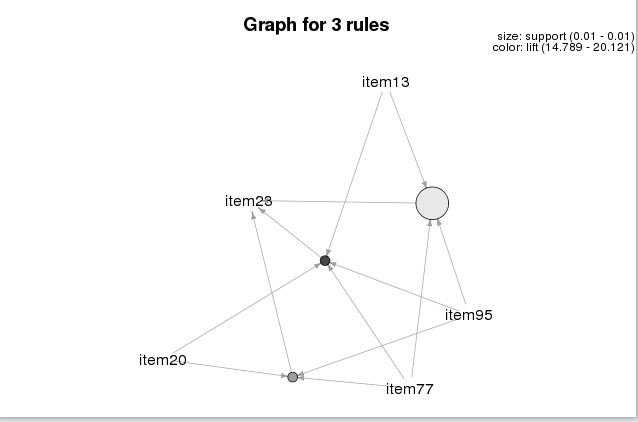
1. Identify these rules and explain their appearance.
2. What can you infer about rules represented by a dark blue color?

* Extract the three rules with the highest lift.

1. Record the Rules. Explain why these rules vary from the rules in Step 3.

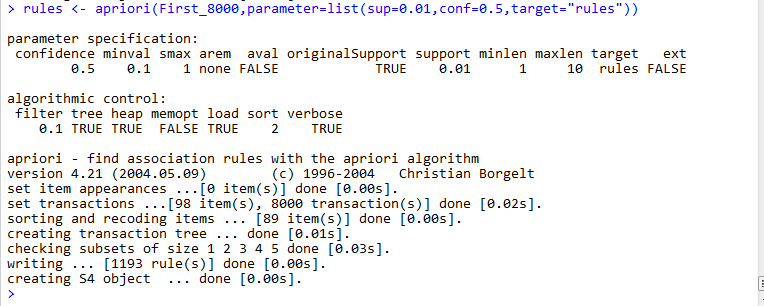
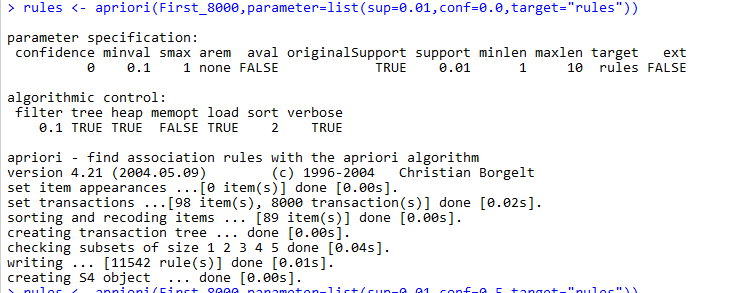


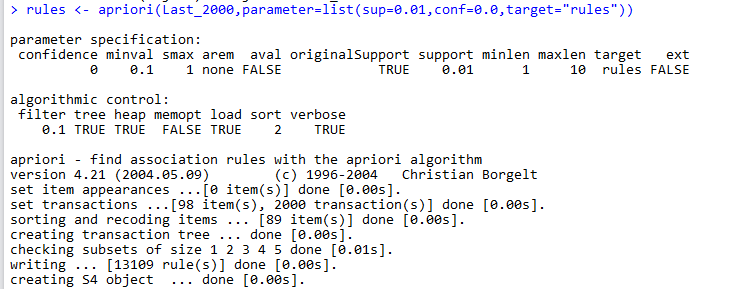
1. Create a Graph-based visualization with items and rules as vertices.

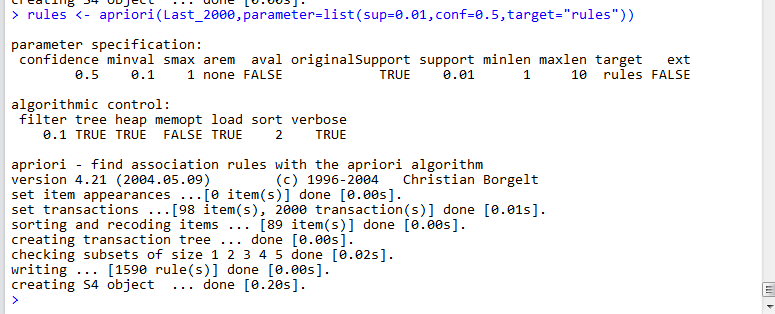


1. Based on your observations, explain how you would expect association rules to relate to order (i.e. the number of items contained in the rule).

* Create a training set from the first 8,000 transactions. Create a testing set from the last 2,000 transactions. Run the algorithm on each dataset. Compare the results.







1. Justify that the relationships we see are not just an artifact of the data we started with.

Can we conclude that the association rules we found are actually true in the population we are studying?

Yes we can conclude that rule we found are actually true in population.