Coursework7

Fortunat Mutunda

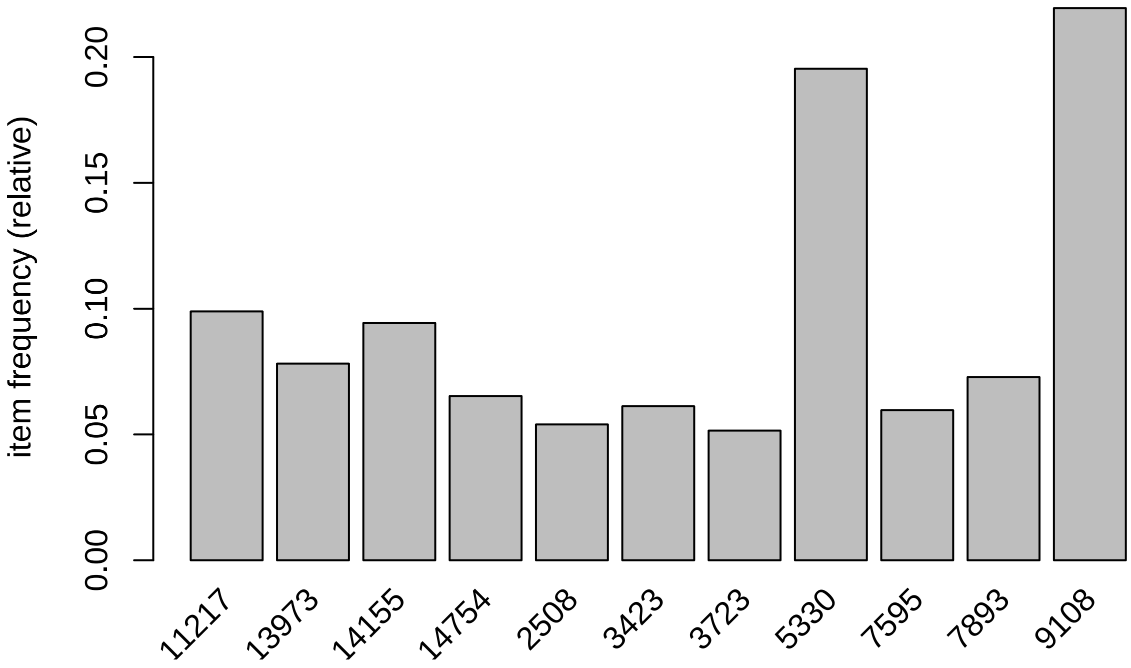
March 23, 2016

1. Report which tools you decided to use, how you used them, what were the first results. Also report the running times for the tools chosen.

comments: *Mine rules from the supermarket txt* 1. Run whatever tool you run and plot the running time 2. Get support and confidence 3. Which tool did you use , how did you use it how long did it run. 4. Discribe in general… show examples, max, min,… etc. 5. How did you get the result what params etc…

For this first tast I have selected the *arules* library because that's the one I am used to, and it's the one I've been using for the couple past homeworks. It is not only easy to use but it is also very intuitive and pretty straight forward. With its nice inuilt function `read.transaction()`

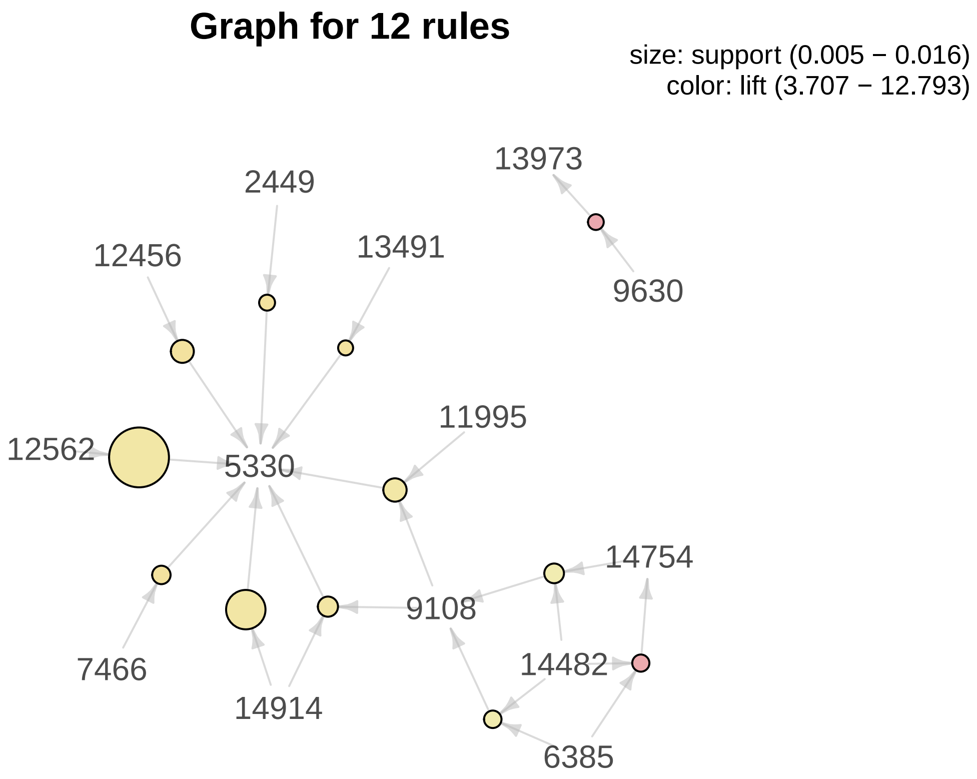
```
#df = read.csv("supermarket.txt", header = F, sep = " ")

#Starts from here
ptm <- proc.time()
trans = read.transactions("supermarket.txt", format = "basket" , sep = " ")
itemFrequencyPlot(trans, support = 0.05)
```



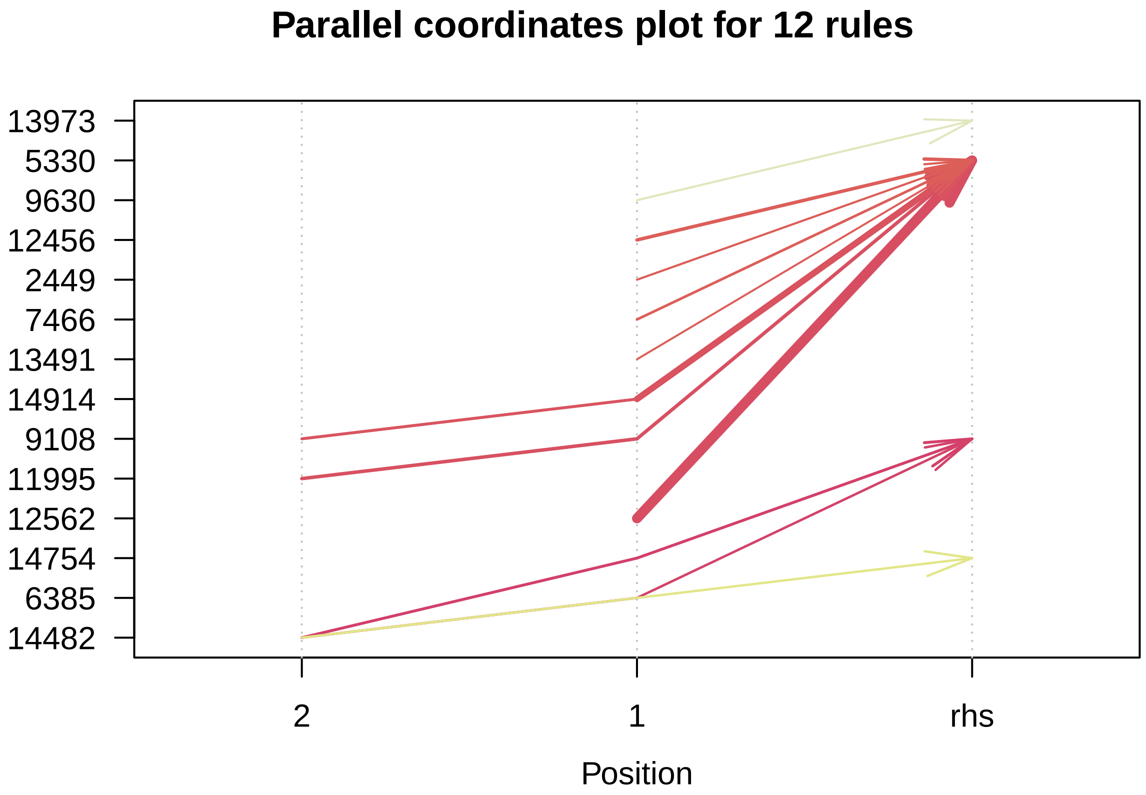
```
#rules
ruless <- apriori(trans, parameter= list(supp=0.005))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen maxlen
##         0.8    0.1    1 none FALSE            TRUE   0.005      1     10
##  target   ext
##   rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 126
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[15699 item(s), 25382 transaction(s)] done [0.04s].
## sorting and recoding items ... [186 item(s)] done [0.00s].
## creating transaction tree ... done [0.01s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [12 rule(s)] done [0.00s].
## creating S4 object  ... done [0.01s].
```

```
plot(ruless, method = "graph")
```



```
plot(ruless, method="paracoord", control=list(reorder=TRUE))
```



```r
proc.time() - ptm
```

```
##    user  system elapsed
##   4.743   0.050   4.807
```

2. Report overall 10 different high-support, high-confidence, high-lift rules; provide the respective contingency tables and scores. ***contengancy table*** if I can find a tool that get me a contengancy table. show the top 10 rules.

3. Discuss whether some other scores studied last week or in the lecture slides would help identify "more interesting" and different rules? ***Use something from last homewok*** no confidence or lift or anything try other maesures

4. Given the ability to discover frequent itemsets and association rules, propose a strategy to use these tools to study different customer segments, shops, shopping times, or specific products. ***What can you do with the data?*** Just write something, what method toools you have learnt so far to study diff rules… Suggest a plan…etc. There is nothing specific, just be creative… In which part can you use frequent set item or item set mining

5. Select some relatively high-support high-confidence rule (A->B) and based on that example describe the conditional probabilities P(A|B) and P(B|A), as well as the Bayes rule.

*** Just learn what a conditional probalities ***

6. (Bonus 2p) Run Krimp on same data, provide commands and describe your findings and compare to Frequent Itemset Mining + Association rules. (link to Krimp documentation)

Krimp