Project 5 Part 4

Kaushal Khatiwada

2024-05-13

Use "Groceries" data available in the "datasets" package

A) Load "arules" and "arulesViz" libraries

```
library(arules)
```

```
## Warning: package 'arules' was built under R version 4.3.3
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.3.3
```

```
##
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
##
##     abbreviate, write
```

```
library(arulesViz)
```

```
## Warning: package 'arulesViz' was built under R version 4.3.3
```

B) Load "Groceries" data, check its structure and interpret it carefully

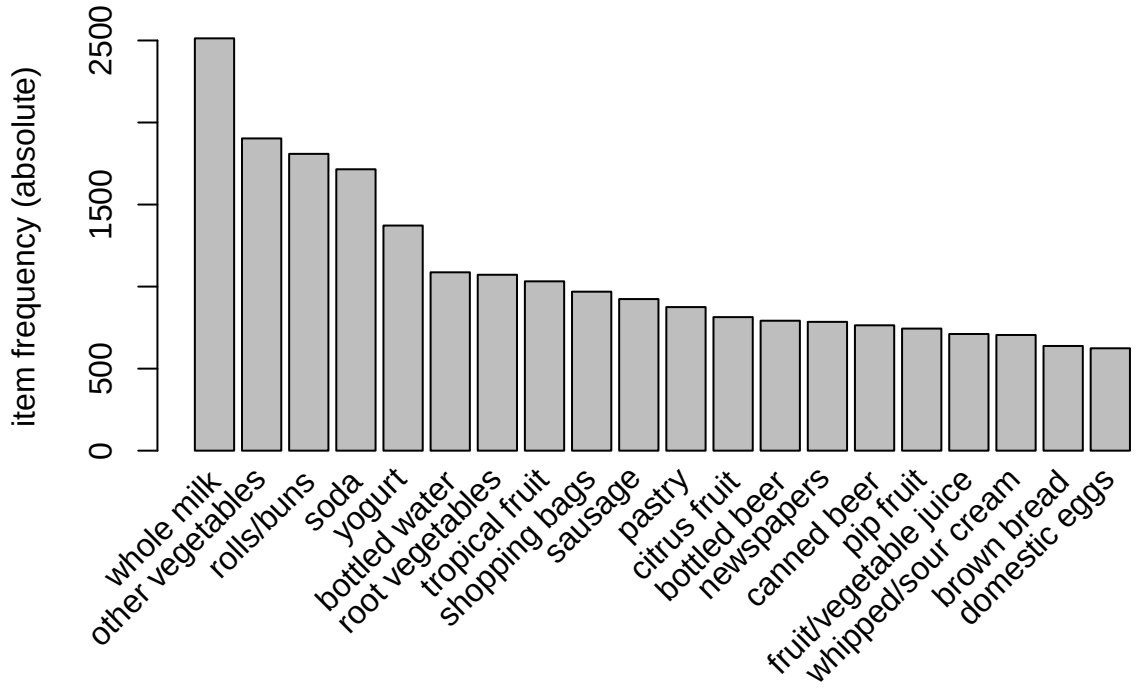
```
data(Groceries)
str(Groceries)
```

```
## Formal class 'transactions' [package "arules"] with 3 slots
##   ..@ data       :Formal class 'ngCMatrix' [package "Matrix"] with 5 slots
##   .. .. ..@ i       : int [1:43367] 13 60 69 78 14 29 98 24 15 29 ...
##   .. .. ..@ p       : int [1:9836] 0 4 7 8 12 16 21 22 27 28 ...
##   .. .. ..@ Dim     : int [1:2] 169 9835
##   .. .. ..@ Dimnames:List of 2
##   .. .. .. ..$ : NULL
##   .. .. .. ..$ : NULL
```

```
##   .. .. ..@ factors : list()
##   ..@ itemInfo   :'data.frame':  169 obs. of  3 variables:
##   .. ..$ labels: chr [1:169] "frankfurter" "sausage" "liver loaf" "ham" ...
##   .. ..$ level2: Factor w/ 55 levels "baby food","bags",..: 44 44 44 44 44 44 44 42 42 41 ...
##   .. ..$ level1: Factor w/ 10 levels "canned food",..: 6 6 6 6 6 6 6 6 6 6 ...
##   ..@ itemsetInfo:'data.frame':  0 obs. of  0 variables
```

C) Get Frequent Item frequencies using itemFrequencyPlot function and interpret it carefully

```
itemFrequencyPlot(Groceries, topN=20, type="absolute")
```



Whole milk is Frequenty used item

D) Set a apriori rule with support = 0.001 and confidence = 0.8 and interpret the output carefully

```
apriori_rule <- apriori(Groceries,
                        parameter = list(supp = 0.001,
                                         conf = 0.8,
                                         maxlen=10,
                                         target="rules"))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.01s].
## writing ... [410 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

E) Show the top five rules using inspect and round the results to two digits

```
five_rules<- inspect(apriori_rule[1:5])
```

```
##     lhs                           rhs            support     confidence
## [1] {liquor, red/blush wine} => {bottled beer} 0.001931876 0.9047619
## [2] {curd, cereals}          => {whole milk}   0.001016777 0.9090909
## [3] {yogurt, cereals}        => {whole milk}   0.001728521 0.8095238
## [4] {butter, jam}            => {whole milk}   0.001016777 0.8333333
## [5] {soups, bottled beer}    => {whole milk}   0.001118454 0.9166667
##     coverage    lift      count
## [1] 0.002135231 11.235269 19
## [2] 0.001118454  3.557863 10
## [3] 0.002135231  3.168192 17
## [4] 0.001220132  3.261374 10
## [5] 0.001220132  3.587512 11
```

F) Sort the rule by confidence in decreasing order

```
sorted_rules <- sort(apriori_rule, by = "confidence", decreasing = TRUE)
inspect(sorted_rules[1:5])
```

```
##     lhs                     rhs              support confidence    coverage     lift count
## [1] {rice,
##      sugar}              => {whole milk} 0.001220132          1 0.001220132 3.913649    12
## [2] {canned fish,
##      hygiene articles}   => {whole milk} 0.001118454          1 0.001118454 3.913649    11
```

```
## [3] {root vegetables,
##      butter,
##      rice}                => {whole milk} 0.001016777          1 0.001016777 3.913649    10
## [4] {root vegetables,
##      whipped/sour cream,
##      flour}               => {whole milk} 0.001728521          1 0.001728521 3.913649    17
## [5] {butter,
##      soft cheese,
##      domestic eggs}       => {whole milk} 0.001016777          1 0.001016777 3.913649    10
```

G) Use "whole milk" as target item and show the items in "lhs" with decreasing order of confidence and show the top five rules

```
milk_rule <- apriori(Groceries,
                     parameter = list(supp = 0.001,
                                      conf = 0.8),
                                      appearance = list(default="lhs",rhs="whole milk"))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.01s].
## writing ... [252 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
milk_sorted_rules <- sort(milk_rule, by = "confidence", decreasing = TRUE)
inspect(milk_sorted_rules[1:5])
```

```
##     lhs                     rhs              support confidence    coverage     lift count
## [1] {rice,
##      sugar}               => {whole milk} 0.001220132          1 0.001220132 3.913649    12
## [2] {canned fish,
##      hygiene articles}    => {whole milk} 0.001118454          1 0.001118454 3.913649    11
## [3] {root vegetables,
##      butter,
##      rice}                => {whole milk} 0.001016777          1 0.001016777 3.913649    10
```

```
## [4] {root vegetables,
##      whipped/sour cream,
##      flour}               => {whole milk} 0.001728521          1 0.001728521 3.913649    17
## [5] {butter,
##      soft cheese,
##      domestic eggs}       => {whole milk} 0.001016777          1 0.001016777 3.913649    10
```

H) Use "whole milk" as target item and show the items in "rhs" with decreasing order of confidence and show the top five rules

```
milk_rule1 <- apriori(Groceries,
                      parameter = list(supp = 0.001,
                                       conf = 0.8),
                      appearance = list(default="rhs"))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 done [0.00s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

I) Write summary and conclusion based on your findings above