

smandumu_A3

smandumu

4/18/2020

```
library(arules) #Arules, open source package is a powerful tool-set for mining associative rules
```

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

```
##
```

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

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

```
##
```

```
##      abbreviate, write
```

```
library(arulesViz)
```

```
## Loading required package: grid
```

```
## Registered S3 method overwritten by 'seriation':
```

```
##      method      from
```

```
## reorder.hclust gclus
```

```
groceries <- read.transactions("groceries_v2.csv",header=FALSE,format = "basket",sep = ",")
```

```
## Warning in readLines(file, encoding = encoding): incomplete final line found on
```

```
## 'groceries_v2.csv'
```

```
summary(groceries)
```

```
## transactions as itemMatrix in sparse format with
```

```
## 9834 rows (elements/itemsets/transactions) and
```

```
## 169 columns (items) and a density of 0.0260911
```

```
##
```

```
## most frequent items:
```

```
##      whole milk other vegetables      rolls/buns      soda
```

```
##           2513           1902           1809           1715
```

```
##           yogurt      (Other)
```

```
##           1372           34051
```

```
##
```

```
## element (itemset/transaction) length distribution:
```

```
## sizes
```

```
##      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16
```

```
## 2159 1643 1299 1005  854  645  545  438  350  246  182  117  78  77  55  46
```

```
##      17     18     19     20     21     22     23     24     26     27     28     29     32
```

```
##      29     14     14      9     11      4      6      1      1      1      1      3      1
```

```
##
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##      1.000  2.000   3.000   4.409   6.000  32.000
```

```
##
```

```
## includes extended item information - examples:
```

```
##           labels
```

```
## 1 abrasive cleaner
```

```
## 2 artif. sweetener
```

```
## 3   baby cosmetics
```

```
itemLabels(groceries) # The Items list
```

```
## [1] "abrasive cleaner"      "artif. sweetener"
## [3] "baby cosmetics"       "baby food"
## [5] "bags"                 "baking powder"
## [7] "bathroom cleaner"     "beef"
## [9] "berries"              "beverages"
## [11] "bottled beer"         "bottled water"
## [13] "brandy"               "brown bread"
## [15] "butter"               "butter milk"
## [17] "cake bar"             "candles"
## [19] "candy"                "canned beer"
## [21] "canned fish"          "canned fruit"
## [23] "canned vegetables"    "cat food"
## [25] "cereals"              "chewing gum"
## [27] "chicken"              "chocolate"
## [29] "chocolate marshmallow" "citrus fruit"
## [31] "cleaner"              "cling film/bags"
## [33] "cocoa drinks"         "coffee"
## [35] "condensed milk"       "cooking chocolate"
## [37] "cookware"             "cream"
## [39] "cream cheese"         "curd"
## [41] "curd cheese"          "decalcifier"
## [43] "dental care"          "dessert"
## [45] "detergent"            "dish cleaner"
## [47] "dishes"               "dog food"
## [49] "domestic eggs"        "female sanitary products"
## [51] "finished products"    "fish"
## [53] "flour"                "flower (seeds)"
## [55] "flower soil/fertilizer" "frankfurter"
## [57] "frozen chicken"       "frozen dessert"
## [59] "frozen fish"          "frozen fruits"
## [61] "frozen meals"         "frozen potato products"
## [63] "frozen vegetables"    "fruit/vegetable juice"
## [65] "grapes"               "hair spray"
## [67] "ham"                  "hamburger meat"
## [69] "hard cheese"          "herbs"
## [71] "honey"                "house keeping products"
## [73] "hygiene articles"     "ice cream"
## [75] "instant coffee"       "Instant food products"
## [77] "jam"                  "ketchup"
## [79] "kitchen towels"       "kitchen utensil"
## [81] "light bulbs"          "liqueur"
## [83] "liquor"               "liquor (appetizer)"
## [85] "liver loaf"           "long life bakery product"
## [87] "make up remover"      "male cosmetics"
## [89] "margarine"            "mayonnaise"
## [91] "meat"                 "meat spreads"
## [93] "misc. beverages"      "mustard"
```

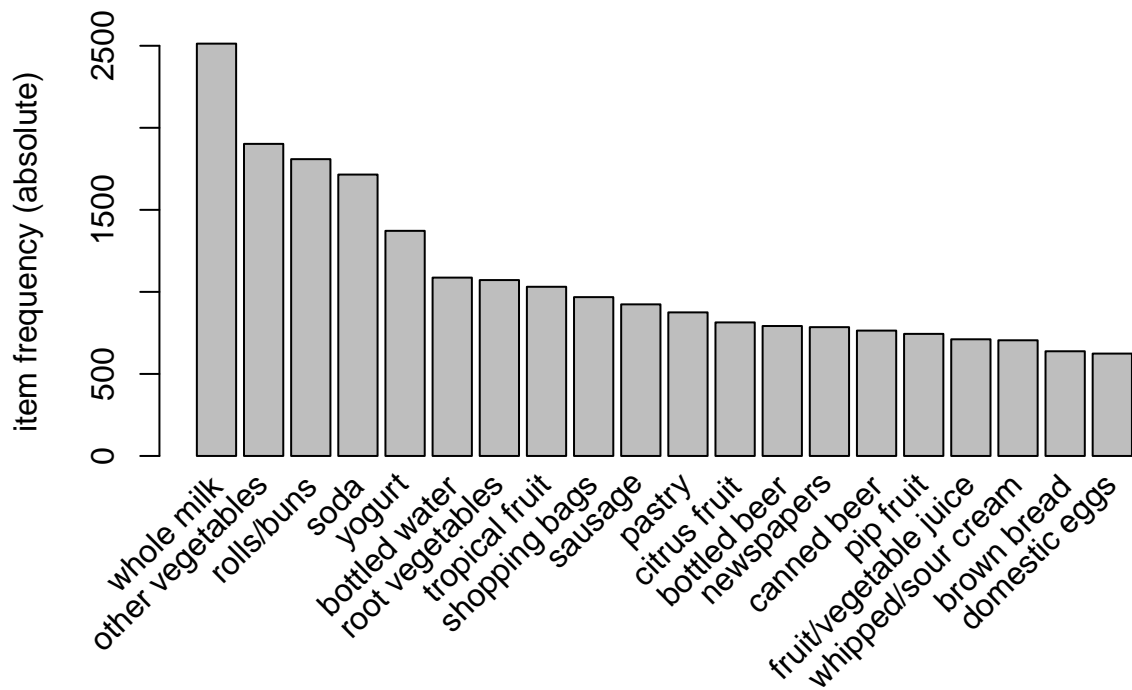
```
## [95] "napkins"           "newspapers"
## [97] "nut snack"         "nuts/prunes"
## [99] "oil"               "onions"
## [101] "organic products"  "organic sausage"
## [103] "other vegetables"  "packaged fruit/vegetables"
## [105] "pasta"             "pastry"
## [107] "pet care"          "photo/film"
## [109] "pickled vegetables" "pip fruit"
## [111] "popcorn"           "pork"
## [113] "pot plants"        "potato products"
## [115] "preservation products" "processed cheese"
## [117] "prosecco"          "pudding powder"
## [119] "ready soups"       "red/blush wine"
## [121] "rice"              "roll products"
## [123] "rolls/buns"        "root vegetables"
## [125] "rubbing alcohol"   "rum"
## [127] "salad dressing"    "salt"
## [129] "salty snack"       "sauces"
## [131] "sausage"           "seasonal products"
## [133] "semi-finished bread" "shopping bags"
## [135] "skin care"         "sliced cheese"
## [137] "snack products"    "soap"
## [139] "soda"              "soft cheese"
## [141] "softener"          "sound storage medium"
## [143] "soups"             "sparkling wine"
## [145] "specialty bar"      "specialty cheese"
## [147] "specialty chocolate" "specialty fat"
## [149] "specialty vegetables" "spices"
## [151] "spread cheese"     "sugar"
## [153] "sweet spreads"     "syrup"
## [155] "tea"               "tidbits"
## [157] "toilet cleaner"    "tropical fruit"
## [159] "turkey"            "UHT-milk"
## [161] "vinegar"           "waffles"
## [163] "whipped/sour cream" "whisky"
## [165] "white bread"        "white wine"
## [167] "whole milk"         "yogurt"
## [169] "zwieback"
```

```
summary(groceries)
```

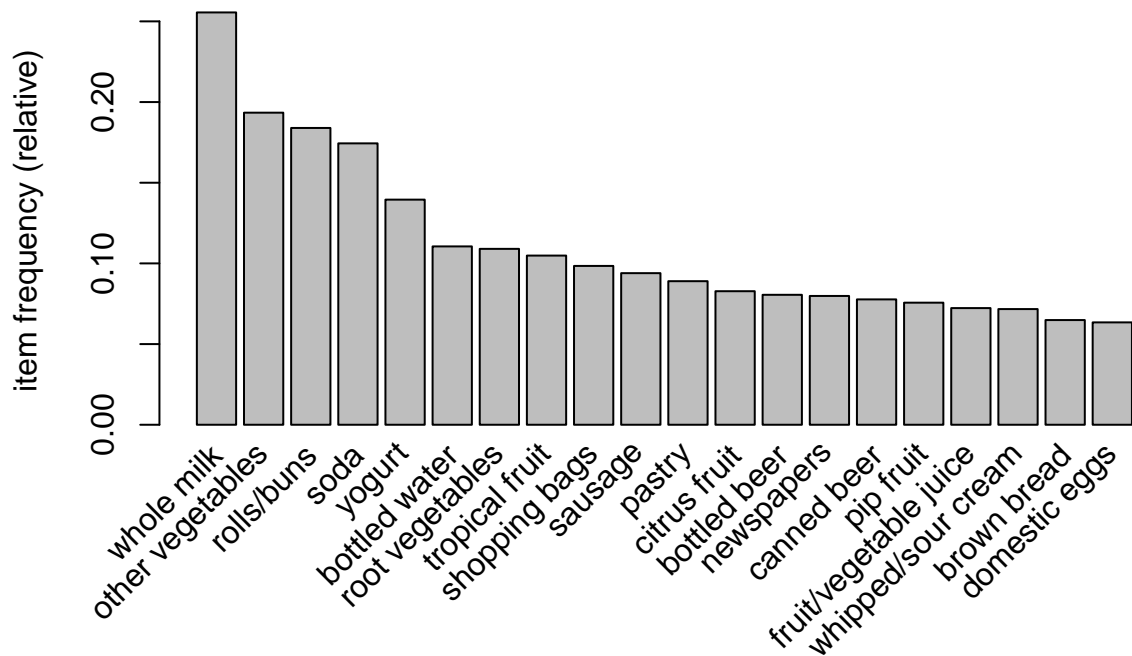
```
## transactions as itemMatrix in sparse format with
## 9834 rows (elements/itemsets/transactions) and
## 169 columns (items) and a density of 0.0260911
##
## most frequent items:
##      whole milk other vegetables      rolls/buns      soda
##           2513           1902           1809           1715
##           yogurt           (Other)
##           1372           34051
##
## element (itemset/transaction) length distribution:
## sizes
##      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16
## 2159 1643 1299 1005  854  645  545  438  350  246  182  117  78   77   55   46
```

```
##      17      18      19      20      21      22      23      24      26      27      28      29      32
##      29      14      14       9      11       4       6       1       1       1       1       3       1
##
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      1.000   2.000   3.000   4.409   6.000   32.000
##
## includes extended item information - examples:
##           labels
## 1 abrasive cleaner
## 2 artif. sweetener
## 3  baby cosmetics
```

```
itemFrequencyPlot(groceries,topN=20,type="absolute") #to look into the item frequency
```



```
itemFrequencyPlot(groceries,topN=20,type="relative") #to look into the item frequency
```



```
# frequent itemsets, association rules using the Apriori algorithm
(gm <- apriori(groceries, parameter=list(support=0.01, confidence=0.5)))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.5      0.1    1 none FALSE          TRUE      5     0.01      1
## maxlen target  ext
##     10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 98
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9834 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [15 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
##
## set of 15 rules
```

```
summary(gm)
```

```
## set of 15 rules
##
## rule length distribution (lhs + rhs):sizes
## 3
## 15
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3      3      3      3      3      3
##
## summary of quality measures:
##      support      confidence      lift      count
## Min. :0.01007 Min. :0.5000 Min. :1.984 Min. : 99.0
## 1st Qu.:0.01174 1st Qu.:0.5151 1st Qu.:2.036 1st Qu.:115.5
## Median :0.01230 Median :0.5245 Median :2.203 Median :121.0
## Mean :0.01316 Mean :0.5411 Mean :2.300 Mean :129.4
## 3rd Qu.:0.01403 3rd Qu.:0.5718 3rd Qu.:2.432 3rd Qu.:138.0
## Max. :0.02227 Max. :0.5862 Max. :3.031 Max. :219.0
##
## mining info:
##      data ntransactions support confidence
## groceries      9834      0.01      0.5
```

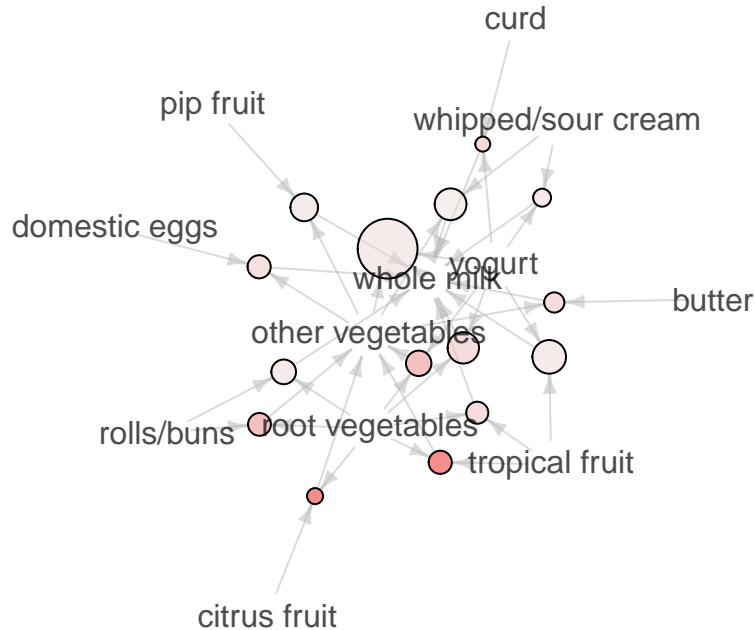
```
inspect(head(sort(gm, by = "lift"),5))
```

```
##      lhs      rhs      support
## [1] {citrus fruit,root vegetables} => {other vegetables} 0.01037218
## [2] {root vegetables,tropical fruit} => {other vegetables} 0.01230425
## [3] {rolls/buns,root vegetables} => {other vegetables} 0.01220256
## [4] {root vegetables,yogurt} => {other vegetables} 0.01291438
## [5] {curd,yogurt} => {whole milk} 0.01006711
##      confidence lift      count
## [1] 0.5862069 3.030893 102
## [2] 0.5845411 3.022280 121
## [3] 0.5020921 2.595990 120
## [4] 0.5000000 2.585174 127
## [5] 0.5823529 2.278893 99
```

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

Graph for 15 rules

size: support (0.01 – 0.022)
color: lift (1.984 – 3.031)



#Rules that lead to buying "whole milk"

```
(wmilk_rules<- apriori(data=groceries, parameter=list (support=0.01,confidence = 0.5), appearance = list
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.5      0.1    1 none FALSE          TRUE      5    0.01      1
## maxlen target  ext
##      10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 98
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9834 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [11 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
## set of 11 rules
```

```

# Rules wrt Whole Milk
inspect(head(sort(wmilk_rules,by="confidence"),3))

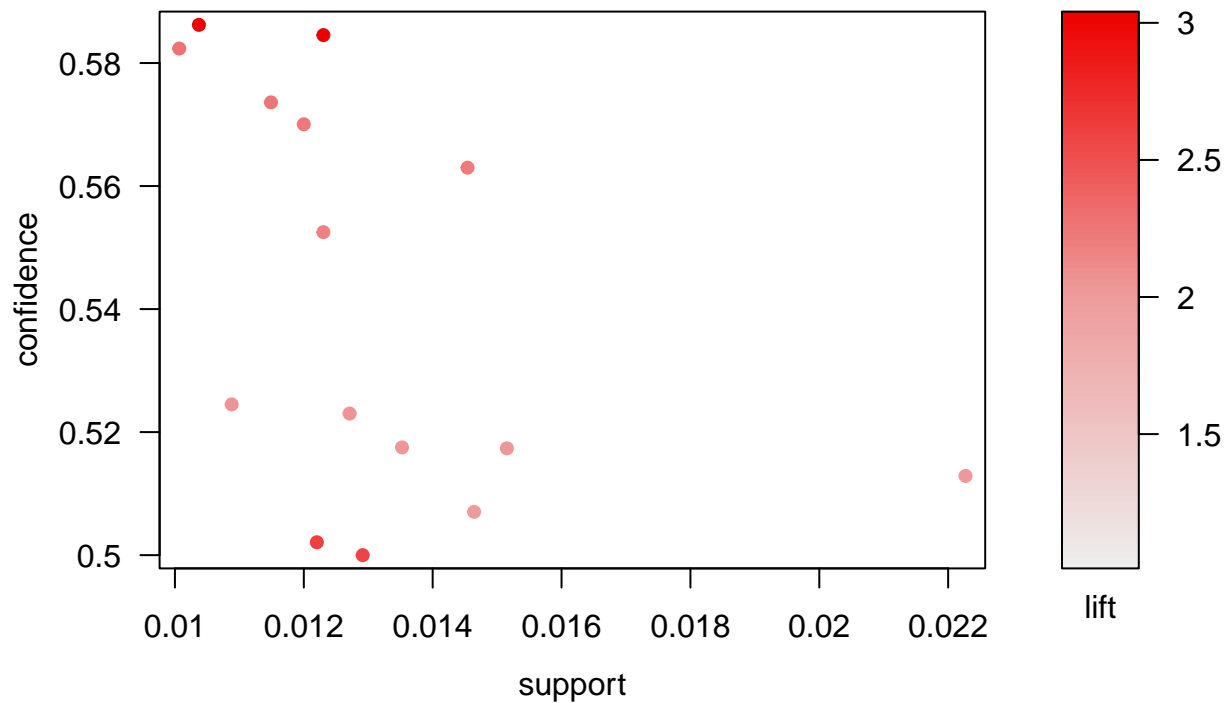
##      lhs                                rhs      support    confidence
## [1] {curd,yogurt}                      => {whole milk} 0.01006711 0.5823529
## [2] {butter,other vegetables}          => {whole milk} 0.01149075 0.5736041
## [3] {root vegetables,tropical fruit} => {whole milk} 0.01199919 0.5700483
##      lift      count
## [1] 2.278893   99
## [2] 2.244657  113
## [3] 2.230742  118

#Rules that lead to buying "whole milk" on LHS side
(wmilk_rules<- apriori(data=groceries, parameter=list (support=0.01,confidence = 0.5), appearance = list(
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0.5   0.1   1 none FALSE                TRUE      5    0.01    1
## maxlen target  ext
##      10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 98
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9834 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
## set of 0 rules

# We see 0 Rules associated
plot(gm)

```


Scatter plot for 15 rules



```
plot(gm,method = "matrix")
```

```
## Itemsets in Antecedent (LHS)
## [1] "{citrus fruit,root vegetables}"
## [2] "{root vegetables,tropical fruit}"
## [3] "{root vegetables,yogurt}"
## [4] "{rolls/buns,root vegetables}"
## [5] "{curd,yogurt}"
## [6] "{butter,other vegetables}"
## [7] "{domestic eggs,other vegetables}"
## [8] "{whipped/sour cream,yogurt}"
## [9] "{other vegetables,pip fruit}"
## [10] "{tropical fruit,yogurt}"
## [11] "{other vegetables,yogurt}"
## [12] "{other vegetables,whipped/sour cream}"
## Itemsets in Consequent (RHS)
## [1] "{whole milk}"      "{other vegetables}"
```

Matrix with 15 rules

