Week 14 IP- Association Rules

Billiah

2022-04-02

```
# Loading package
library(arules)
```

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

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

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

```
library(arulesViz)
```

```
df = read.transactions("http://bit.ly/SupermarketDatasetII")
```

```
## Warning in asMethod(object): removing duplicated items in transactions
```

```
head(df)
```

```
## transactions in sparse format with
##  6 transactions (rows) and
##  5729 items (columns)
```

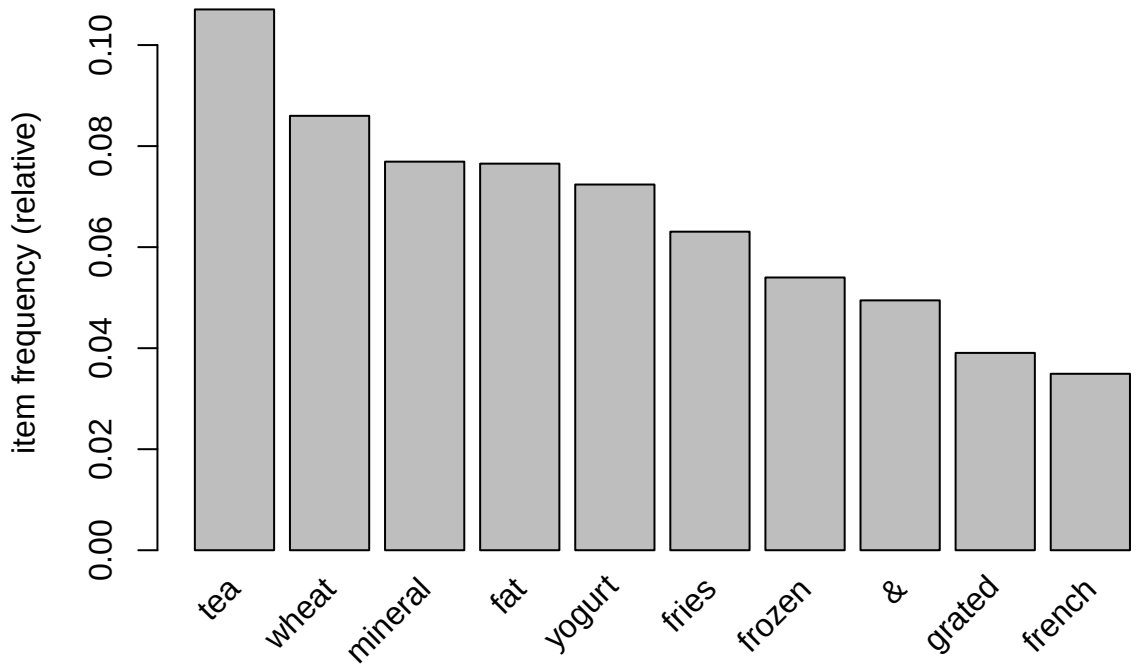
```
str(df)
```

```
## Formal class 'transactions' [package "arules"] with 3 slots
##   ..@ data       :Formal class 'ngCMatrix' [package "Matrix"] with 5 slots
##   .. .. ..@ i       : int [1:23299] 1087 1614 1705 1732 1993 2101 2105 2358 2444 3463 ...
##   .. .. ..@ p       : int [1:7502] 0 15 16 17 18 24 27 31 33 36 ...
##   .. .. ..@ Dim     : int [1:2] 5729 7501
##   .. .. ..@ Dimnames:List of 2
##   .. .. .. ..$ : NULL
##   .. .. .. ..$ : NULL
##   .. .. ..@ factors : list()
##   ..@ itemInfo   :'data.frame':  5729 obs. of  1 variable:
##   .. ..$ labels: chr [1:5729] "&" "accessories" "accessories,antioxydant" "accessories,champagne,fres
##   ..@ itemsetInfo:'data.frame':  0 obs. of  0 variables
```

```r
set.seed = 123 # Setting seed
a_rules = apriori(data = df,
                          parameter = list(support = 0.005,
                                           confidence = 0.5))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5   0.005      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: 37
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[5729 item(s), 7501 transaction(s)] done [0.03s].
## sorting and recoding items ... [73 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [30 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```r
itemFrequencyPlot(df, topN = 10)
```

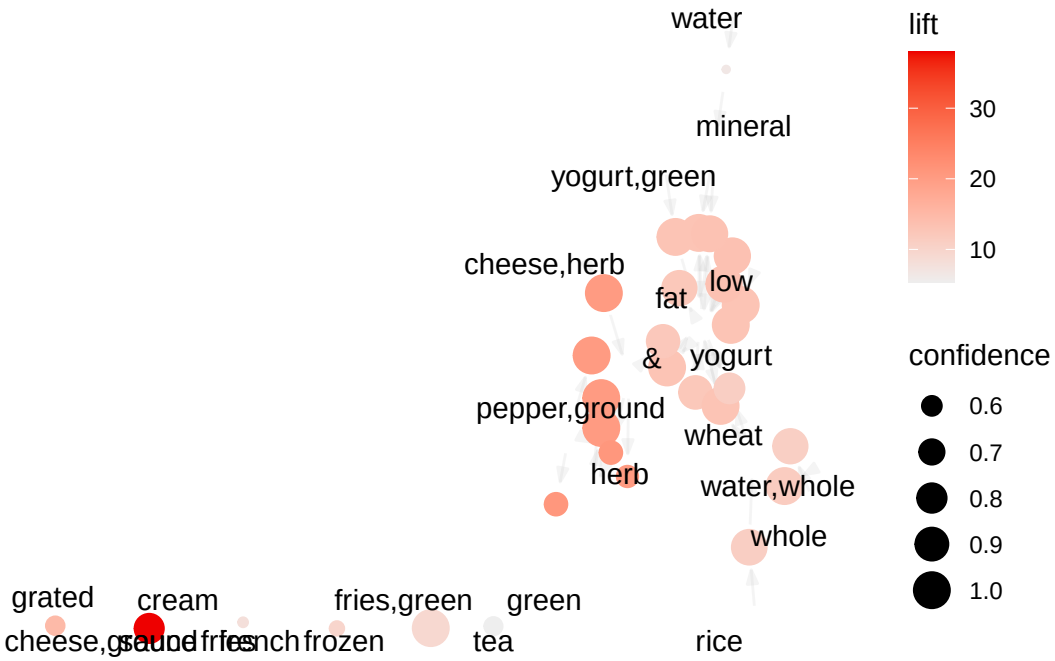



```
# Visualisation
inspect(sort(a_rules, by = 'lift')[1:10])
```

```
##      lhs                       rhs      support     confidence coverage
## [1]  {sauce}                => {cream}  0.010265298 0.7938144  0.012931609
## [2]  {pepper,ground}        => {herb}   0.005865885 0.6470588  0.009065458
## [3]  {&, pepper,ground}     => {herb}   0.005865885 0.6470588  0.009065458
## [4]  {cheese,herb}          => {&}      0.006532462 1.0000000  0.006532462
## [5]  {pepper,ground}        => {&}      0.009065458 1.0000000  0.009065458
## [6]  {herb}                 => {&}      0.030929209 1.0000000  0.030929209
## [7]  {&}                    => {herb}   0.030929209 0.6253369  0.049460072
## [8]  {herb, pepper,ground} => {&}      0.005865885 1.0000000  0.005865885
## [9]  {cheese,ground}        => {grated} 0.005199307 0.5820896  0.008932142
## [10] {low}                  => {yogurt} 0.005999200 0.9782609  0.006132516
##      lift     count
## [1]  37.92613  77
## [2]  20.92064  44
## [3]  20.92064  44
## [4]  20.21833  49
## [5]  20.21833  68
## [6]  20.21833 232
## [7]  20.21833 232
## [8]  20.21833  44
## [9]  14.90189  39
## [10] 13.51369  45
```

```
plot(a_rules, method = "graph",
     measure = "confidence", shading = "lift")
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



Tea is the best selling item, followed by wheat, mineral, fat and yoghurt, in that order.

1% of customers bought sauce and cream.

Confidence is that 79% of customers bought sauce and cream.

There is increased expectation that if someone buys sauce, they are likely to buy cream.If they buy pepper& ground, they are also likely to buy herb.