systeme informatique décisionnelle et data mining

Réalisé par : EL RHARROUBI Mohamed amine

Encadré par : Pr. Abdelhadi FENNAN

groceries

Code

Hide

```
require(arules)
require(Matrix)
```

Hide

```
data(package="arules")
Groc <- read.transactions("groceries.csv", sep = ",")
Groc
```

```
transactions in sparse format with
 9835 transactions (rows) and
 169 items (columns)
```

```
summary(Groc)
```

```
transactions as itemMatrix in sparse format with
 9835 rows (elements/itemsets/transactions) and
 169 columns (items) and a density of 0.02609146

most frequent items:
      whole milk other vegetables       rolls/buns
            2513             1903             1809
            soda           yogurt          (Other)
            1715             1372            34055

element (itemset/transaction) length distribution:
sizes
   1    2    3    4    5    6    7    8    9   10
2159 1643 1299 1005  855  645  545  438  350  246
  11   12   13   14   15   16   17   18   19   20
 182  117   78   77   55   46   29   14   14    9
  21   22   23   24   26   27   28   29   32
  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 <chr>
1	abrasive cleaner
2	artif. sweetener
3	baby cosmetics

3 rows

```
inspect(Groc[1:3])
```

	items <chr>
[1]	{citrus fruit,margarine,ready soups,semi-finished bread}
[2]	{coffee,tropical fruit,yogurt}
[3]	{whole milk}

3 rows

```
inspect(Groc[9000:9005])
```

```
    items
[1] {soda,
     tropical fruit,
     yogurt}
[2] {brown bread,
     margarine,
     whole milk}
[3] {baking powder,
     brown bread,
     butter,
     cat food,
     chocolate,
     citrus fruit,
     cream,
     curd,
     female sanitary products,
     frozen fish,
     frozen potato products,
     frozen vegetables,
     fruit/vegetable juice,
     hard cheese,
     house keeping products,
     hygiene articles,
     margarine,
     napkins,
     nut snack,
     pet care,
     pip fruit,
     pork,
     root vegetables,
     sausage,
     sliced cheese,
     soda,
     tropical fruit,
     whole milk,
     yogurt}
[4] {soda}
[5] {tropical fruit,
     whole milk}
[6] {bottled beer,
     candles,
     canned beer,
     chocolate,
     domestic eggs,
     long life bakery product,
     newspapers,
     pastry}
```

```
itemFrequency(Groc[,1])
```

```
abrasive cleaner
     0.003558719
```

Hide

```
itemFrequency(Groc[,100])
```

```
    onions
0.03101169
```

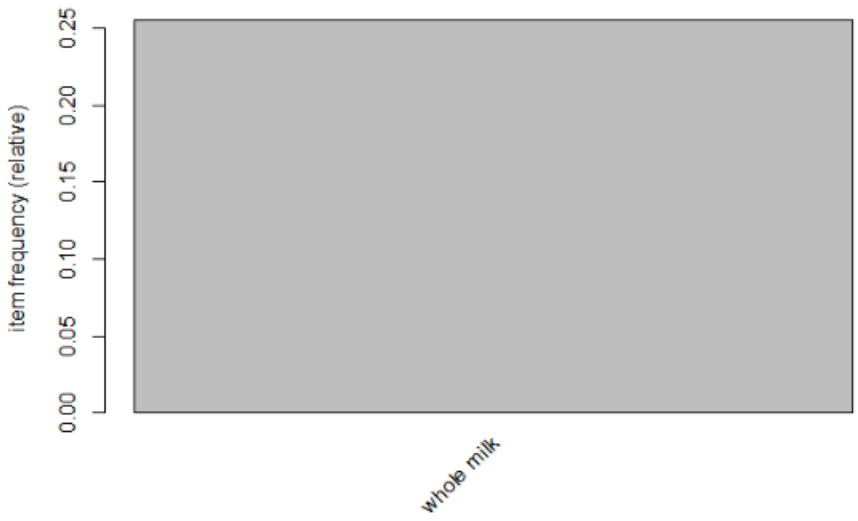
Hide

```
itemFrequency(Groc[,1:6])
```

```
abrasive cleaner artif. sweetener   baby cosmetics        baby food             bags    baking powder
    0.0035587189     0.0032536858     0.0006100661     0.0001016777     0.0004067107     0.0176919166
```

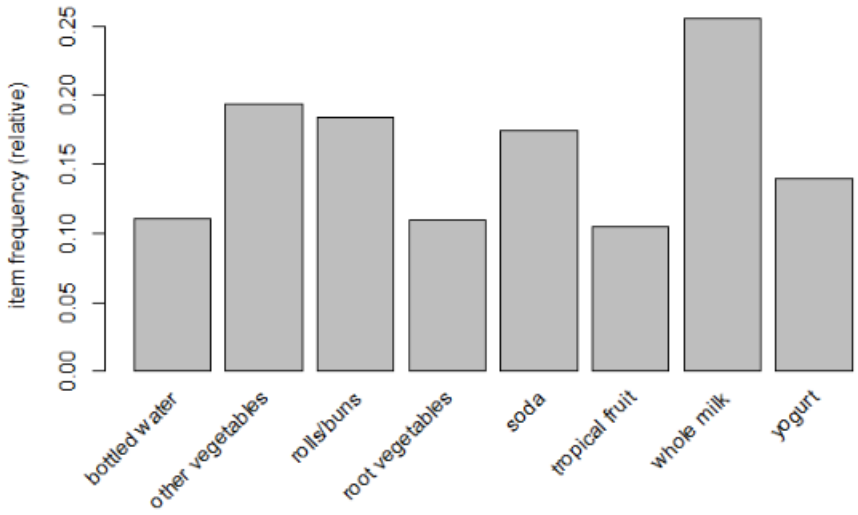
Hide

```
itemFrequencyPlot(Groc,support=0.2)
```



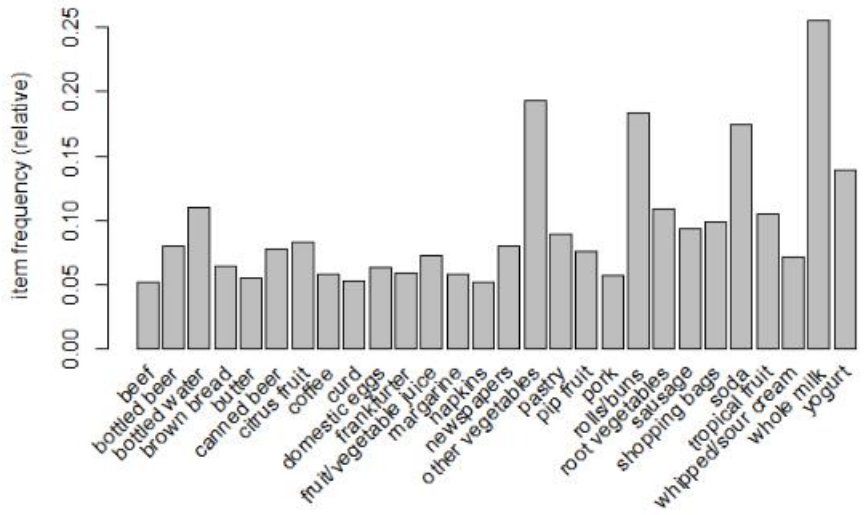
Hide

```
itemFrequencyPlot(Groc,support=0.1)
```



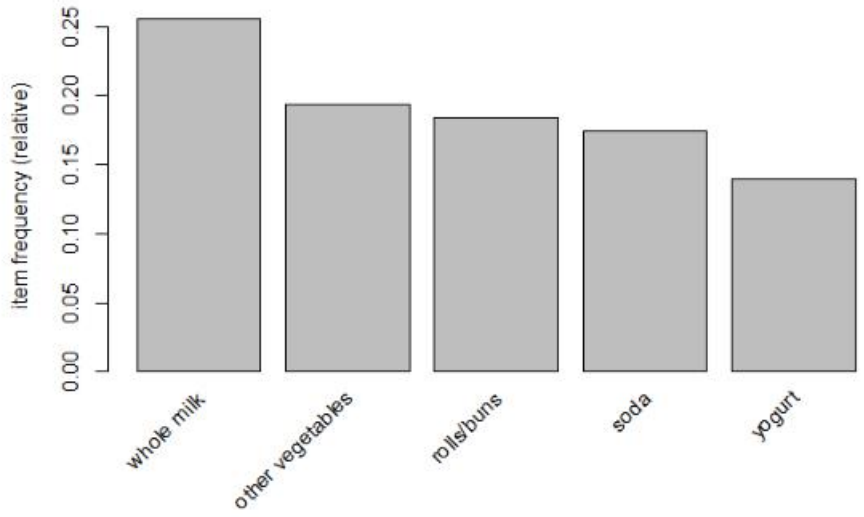
Hide

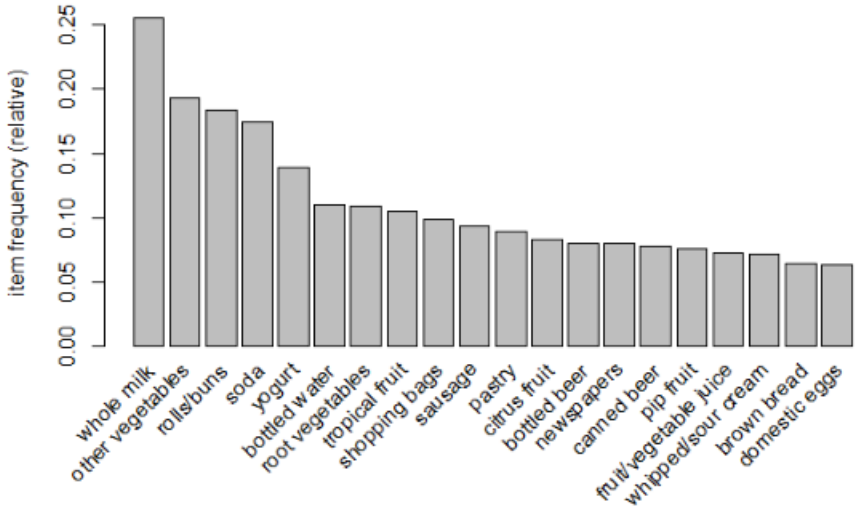
```
itemFrequencyPlot(Groc,support=0.05)
```



Hide

```
itemFrequencyPlot(Groc,topN=5)
```





Hide

```
m1 <- apriori(Groc,parameter = list(support=0.007, confidence=0.25, minlen=2))
```

```
Apriori

Parameter specification:
```

confidence <dbl>	minval <dbl>	smax <dbl>	arem <chr>	aval <lgl>	originalSupport <lgl>	maxtime <dbl>	support <dbl>	minlen <int>
0.25	0.1	1	none	FALSE	TRUE	5	0.007	2

1 row | 1-10 of 12 columns

```
Algorithmic control:
```

filter <dbl>	tree <lgl>	heap <lgl>	memopt <lgl>	load <lgl>	sort <int>	verbose <lgl>
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

1 row

```
Absolute minimum support count: 68

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.01s].
sorting and recoding items ... [104 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.01s].
writing ... [363 rule(s)] done [0.00s].
creating S4 object  ... done [0.02s].
```

Hide

```
summary(m1)
```

```
set of 363 rules

rule length distribution (lhs + rhs):sizes
  2   3   4
137 214  12

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   2.000   3.000   2.656   3.000   4.000

summary of quality measures:
    support           confidence        coverage           lift           count
 Min.   :0.007016   Min.   :0.2500   Min.   :0.01200   Min.   :0.9932   Min.   : 69.0
 1st Qu.:0.008134   1st Qu.:0.2962   1st Qu.:0.02166   1st Qu.:1.6060   1st Qu.: 80.0
 Median :0.009659   Median :0.3551   Median :0.02888   Median :1.9086   Median : 95.0
 Mean   :0.012945   Mean   :0.3743   Mean   :0.03675   Mean   :2.0072   Mean   :127.3
 3rd Qu.:0.013777   3rd Qu.:0.4420   3rd Qu.:0.04230   3rd Qu.:2.3289   3rd Qu.:135.5
 Max.   :0.074835   Max.   :0.6389   Max.   :0.25552   Max.   :3.9565   Max.   :736.0

mining info:
```

	data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
	Groc	9835	0.007	0.25

1 row

Hide

```
inspect(m1[1:2])
```

	lhs <chr>		rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{herbs}	=>	{root vegetables}	0.007015760	0.43125	0.01626843	3.956477	69
[2]	{herbs}	=>	{other vegetables}	0.007727504	0.47500	0.01626843	2.454874	76

2 rows

```
inspect(sort(m1,by="lift")[1:5])
```

	lhs <chr>		rhs <chr>	support <dbl>	confidence <dbl>
[1]	{herbs}	=>	{root vegetables}	0.007015760	0.4312500
[2]	{berries}	=>	{whipped/sour cream}	0.009049314	0.2721713
[3]	{other vegetables,tropical fruit,whole milk}	=>	{root vegetables}	0.007015760	0.4107143
[4]	{beef,other vegetables}	=>	{root vegetables}	0.007930859	0.4020619
[5]	{other vegetables,tropical fruit}	=>	{pip fruit}	0.009456024	0.2634561

5 rows | 1-6 of 8 columns

Hide

```
inspect(sort(m1,by="lift")[6:10])
```

	lhs <chr>		rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>
[1]	{beef,whole milk}	=>	{root vegetables}	0.008032537	0.3779904	0.02125064
[2]	{other vegetables,pip fruit}	=>	{tropical fruit}	0.009456024	0.3618677	0.02613116
[3]	{citrus fruit,other vegetables}	=>	{root vegetables}	0.010371124	0.3591549	0.02887646
[4]	{other vegetables,whole milk,yogurt}	=>	{tropical fruit}	0.007625826	0.3424658	0.02226741
[5]	{other vegetables,whole milk,yogurt}	=>	{root vegetables}	0.007829181	0.3515982	0.02226741

5 rows | 1-7 of 8 columns