R Notebook

Code

Hide

```
install.packages("arules")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please downlo
ad and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/User/AppData/Local/R/win-library/4.2’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/arules_1.7-5.zip'
Content type 'application/zip' length 2126054 bytes (2.0 MB)
downloaded 2.0 MB
```

```
package ‘arules’ successfully unpacked and MD5 sums checked
Warning in install.packages :
  cannot remove prior installation of package ‘arules’
Warning in install.packages :
  problem copying C:\Users\User\AppData\Local\R\win-library\4.2\00LOCK\arules\libs\x64\arule
s.dll to C:\Users\User\AppData\Local\R\win-library\4.2\arules\libs\x64\arules.dll: Permission
denied
Warning in install.packages :
  restored ‘arules’

The downloaded binary packages are in
    C:\Users\User\AppData\Local\Temp\RtmpiYsaJ8\downloaded_packages
```

Hide

```
install.packages("arulesViz")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please downlo
ad and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/User/AppData/Local/R/win-library/4.2’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/arulesViz_1.5-1.zip'
Content type 'application/zip' length 1781449 bytes (1.7 MB)
downloaded 1.7 MB
```

```
package ‘arulesViz’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\User\AppData\Local\Temp\RtmpiYsaJ8\downloaded_packages
```

Hide

```
install.packages("RColorBrewer")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please downlo
ad and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/User/AppData/Local/R/win-library/4.2’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/RColorBrewer_1.1-3.zip'
Content type 'application/zip' length 55837 bytes (54 KB)
downloaded 54 KB
```

```
package ‘RColorBrewer’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\User\AppData\Local\Temp\RtmpiYsaJ8\downloaded_packages
```

Hide

```
# Loading Libraries
library(arules)
library(arulesViz)
library(RColorBrewer)
```

Hide

```
# import dataset
data(Groceries)
Groceries
```

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

Hide

```
summary(Groceries)
```

```
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             soda           yogurt
(Other)
            2513             1903             1809             1715             1372
34055

element (itemset/transaction) length distribution:
sizes
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   1
9   20   21
2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55   46   29   14   1
4    9   11
  22   23   24   26   27   28   29   32
   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>	level2 <fctr>	level1 <fctr>
1	frankfurter	sausage	meat and sausage
2	sausage	sausage	meat and sausage
3	liver loaf	sausage	meat and sausage

3 rows

Hide

```
class(Groceries)
```

```
[1] "transactions"
attr(,"package")
[1] "arules"
```

Hide

```
# using apriori() function
rules = apriori(Groceries, parameter = list(supp = 0.02, conf = 0.2))
```

```
Apriori

Parameter specification:
```

confidence <dbl>	minval <dbl>	s… <dbl>	ar… <chr>	aval <lgl>	originalSupport <lgl>	maxti… <dbl>	support <dbl>	minlen <int>
0.2	0.1	1	none	FALSE	TRUE	5	0.02	1

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: 196

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.01s].
sorting and recoding items ... [59 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [73 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

Hide

```
summary (rules)
```

```
set of 73 rules

rule length distribution (lhs + rhs):sizes
 1  2  3
 1 66  6

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   2.000   2.068   2.000   3.000

summary of quality measures:
    support         confidence        coverage            lift            count
 Min.   :0.02003   Min.   :0.2006   Min.   :0.04342   Min.   :0.8991   Min.   : 197.0
 1st Qu.:0.02257   1st Qu.:0.2369   1st Qu.:0.07168   1st Qu.:1.3112   1st Qu.: 222.0
 Median :0.02664   Median :0.3079   Median :0.09395   Median :1.5570   Median : 262.0
 Mean   :0.03424   Mean   :0.3187   Mean   :0.11739   Mean   :1.6061   Mean   : 336.8
 3rd Qu.:0.03589   3rd Qu.:0.3868   3rd Qu.:0.11052   3rd Qu.:1.8502   3rd Qu.: 353.0
 Max.   :0.25552   Max.   :0.5129   Max.   :1.00000   Max.   :2.8421   Max.   :2513.0

mining info:
```

data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
Groceries	9835	0.02	0.2

1 row | 1-5 of 5 columns

Hide

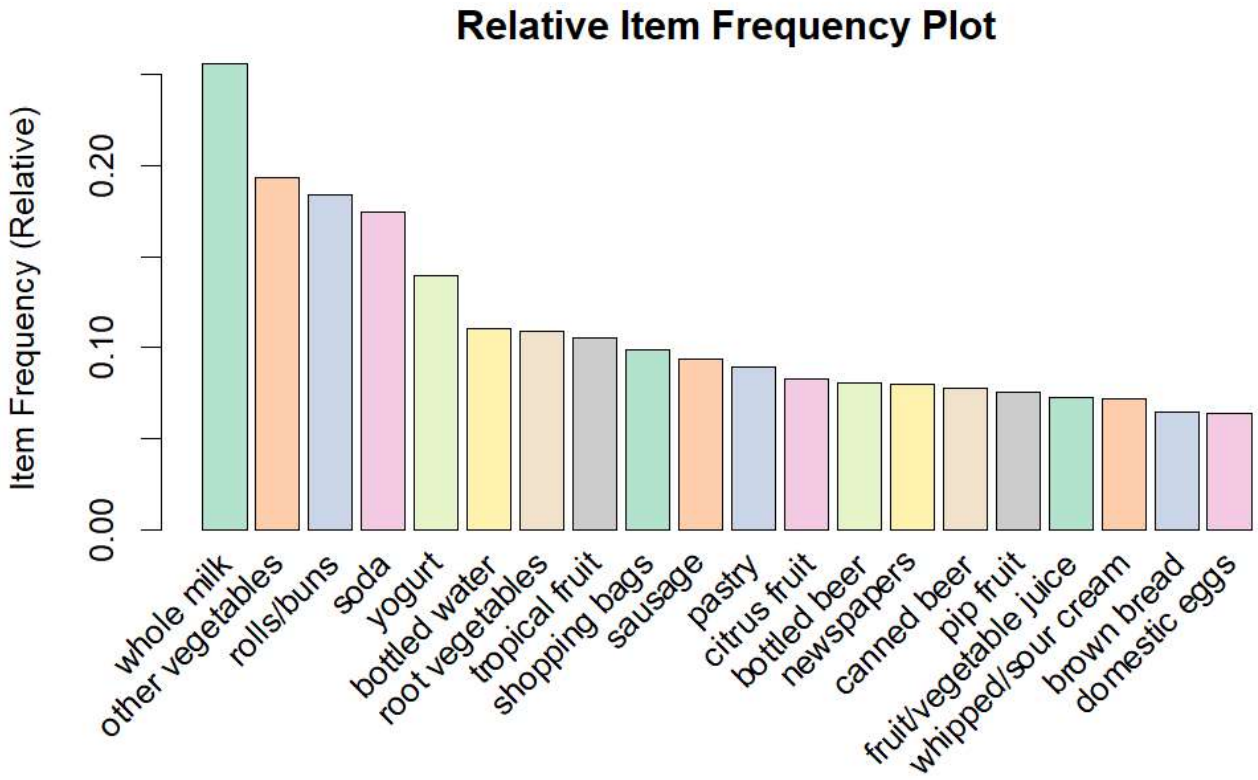
```
# using inspect() function
inspect(rules[1:10])
```

	lhs <chr>		rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>
[1]	{}	=>	{whole milk}	0.25551601	0.2555160	1.00000000
[2]	{frozen vegetables}	=>	{whole milk}	0.02043721	0.4249471	0.04809354
[3]	{beef}	=>	{whole milk}	0.02125064	0.4050388	0.05246568
[4]	{curd}	=>	{whole milk}	0.02613116	0.4904580	0.05327911
[5]	{pork}	=>	{other vegetables}	0.02165735	0.3756614	0.05765125
[6]	{pork}	=>	{whole milk}	0.02216573	0.3844797	0.05765125
[7]	{frankfurter}	=>	{whole milk}	0.02053889	0.3482759	0.05897306
[8]	{bottled beer}	=>	{whole milk}	0.02043721	0.2537879	0.08052872
[9]	{brown bread}	=>	{whole milk}	0.02521607	0.3887147	0.06487036
[10]	{margarine}	=>	{whole milk}	0.02419929	0.4131944	0.05856634

1-10 of 10 rows

Hide

```
# using itemFrequencyPlot() function
arules::itemFrequencyPlot(Groceries, topN = 20,
                          col = brewer.pal(8, 'Pastel2'),
                          main = 'Relative Item Frequency Plot',
                          type = "relative",
                          ylab = "Item Frequency (Relative)")
```



Hide

```
itemsets = apriori(Groceries, parameter = list(minlen=2, maxlen=2,support=0.02, target="frequ
ent itemsets"))
```

```
Apriori

Parameter specification:
```

confidence <dbl>	minval <dbl>	s… <dbl>	ar… <chr>	aval <lgl>	originalSupport <lgl>	maxti… <dbl>	support <dbl>	minlen <int>
NA	0.1	1	none	FALSE	TRUE	5	0.02	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: 196

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.01s].
sorting and recoding items ... [59 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2
```

```
Warning: Mining stopped (maxlen reached). Only patterns up to a length of 2 returned!
```

```
 done [0.00s].
sorting transactions ... done [0.00s].
writing ... [61 set(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

Hide

```
summary(itemsets)
```

```
set of 61 itemsets

most frequent items:
      whole milk other vegetables           yogurt       rolls/buns             soda
(Other)
              25               17                9                9                9
53

element (itemset/transaction) length distribution:sizes
 2
61

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      2       2       2       2       2       2

summary of quality measures:
    support            count
 Min.   :0.02003   Min.   :197.0
 1st Qu.:0.02227   1st Qu.:219.0
 Median :0.02613   Median :257.0
 Mean   :0.02951   Mean   :290.3
 3rd Qu.:0.03223   3rd Qu.:317.0
 Max.   :0.07483   Max.   :736.0

includes transaction ID lists: FALSE

mining info:
```

data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
Groceries	9835	0.02	1

1 row | 1-5 of 5 columns

Hide

```
# using inspect() function
inspect(itemsets[1:10])
```

	items <chr>	support <dbl>	count <int>
[1]	{whole milk, frozen vegetables}	0.02043721	201
[2]	{beef, whole milk}	0.02125064	209
[3]	{whole milk, curd}	0.02613116	257
[4]	{pork, other vegetables}	0.02165735	213
[5]	{pork, whole milk}	0.02216573	218
[6]	{frankfurter, whole milk}	0.02053889	202
[7]	{whole milk, bottled beer}	0.02043721	201
[8]	{whole milk, brown bread}	0.02521607	248
[9]	{whole milk, margarine}	0.02419929	238
[10]	{other vegetables, butter}	0.02003050	197

1-10 of 10 rows

Hide

```
itemsets_3 = apriori(Groceries, parameter = list(minlen=3, maxlen=3,support=0.02, target="fre
quent itemsets"))
```

```
Apriori

Parameter specification:
```

confidence <dbl>	minval <dbl>	s… <dbl>	ar… <chr>	aval <lgl>	originalSupport <lgl>	maxti… <dbl>	support <dbl>	minlen <int>
NA	0.1	1	none	FALSE	TRUE	5	0.02	3

1 row | 1-10 of 12 columns

```

Algorithmic control:
```

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

1 row

```
Absolute minimum support count: 196

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.01s].
sorting and recoding items ... [59 item(s)] done [0.00s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3
```

```
Warning: Mining stopped (maxlen reached). Only patterns up to a length of 3 returned!
```

```
 done [0.00s].
sorting transactions ... done [0.00s].
writing ... [2 set(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

Hide

```
summary(itemsets_3)
```

```
set of 2 itemsets

most frequent items:
other vegetables       whole milk  root vegetables           yogurt      frankfurter
(Other)
               2                2                1                1                0
0

element (itemset/transaction) length distribution:sizes
3
2

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

summary of quality measures:
    support            count
 Min.   :0.02227   Min.   :219.0
 1st Qu.:0.02250   1st Qu.:221.2
 Median :0.02272   Median :223.5
 Mean   :0.02272   Mean   :223.5
 3rd Qu.:0.02295   3rd Qu.:225.8
 Max.   :0.02318   Max.   :228.0

includes transaction ID lists: FALSE

mining info:
```

	data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
	Groceries	9835	0.02	1

1 row | 1-5 of 5 columns

Hide

```
# using inspect() function
inspect(itemsets_3)
```

	items <chr>	support <dbl>	count <int>
[1]	{root vegetables, other vegetables, whole milk}	0.02318251	228
[2]	{other vegetables, whole milk, yogurt}	0.02226741	219

2 rows