R Notebook

Code

Association Rules

For this portion of the assignment, we will be using data from Groceries, a dataset that can be found with the arules package. Each row in the file represents one buyer's purchases. This link provides some helpful templated examples for generating association rules: http://r-statistics.co/Association-Mining-With-R.html (http://r-statistics.co/Association-Mining-With-R.html)

1. Describe "Groceries" by answering following questions: What is the class of "Groceries"? How many rows and columns does Groceries contain?

Hide

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

```
Error in install.packages : Updating loaded packages
```

Hide

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

```
Error in install.packages : Updating loaded packages
```

Hide

```
library(arules)
library(arulesViz)
data(Groceries) # read the Groceries data
class(Groceries) # determine the class type
```

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

Hide

```
summary(Groceries) # inspect the rows and columns
```

```
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   1
8   19   20
2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55   46   29   1
4   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	level2	level1
	<chr>	<fctr>	<fctr>
1	frankfurter	sausage	meat and sausage
2	sausage	sausage	meat and sausage
3	liver loaf	sausage	meat and sausage

3 rows

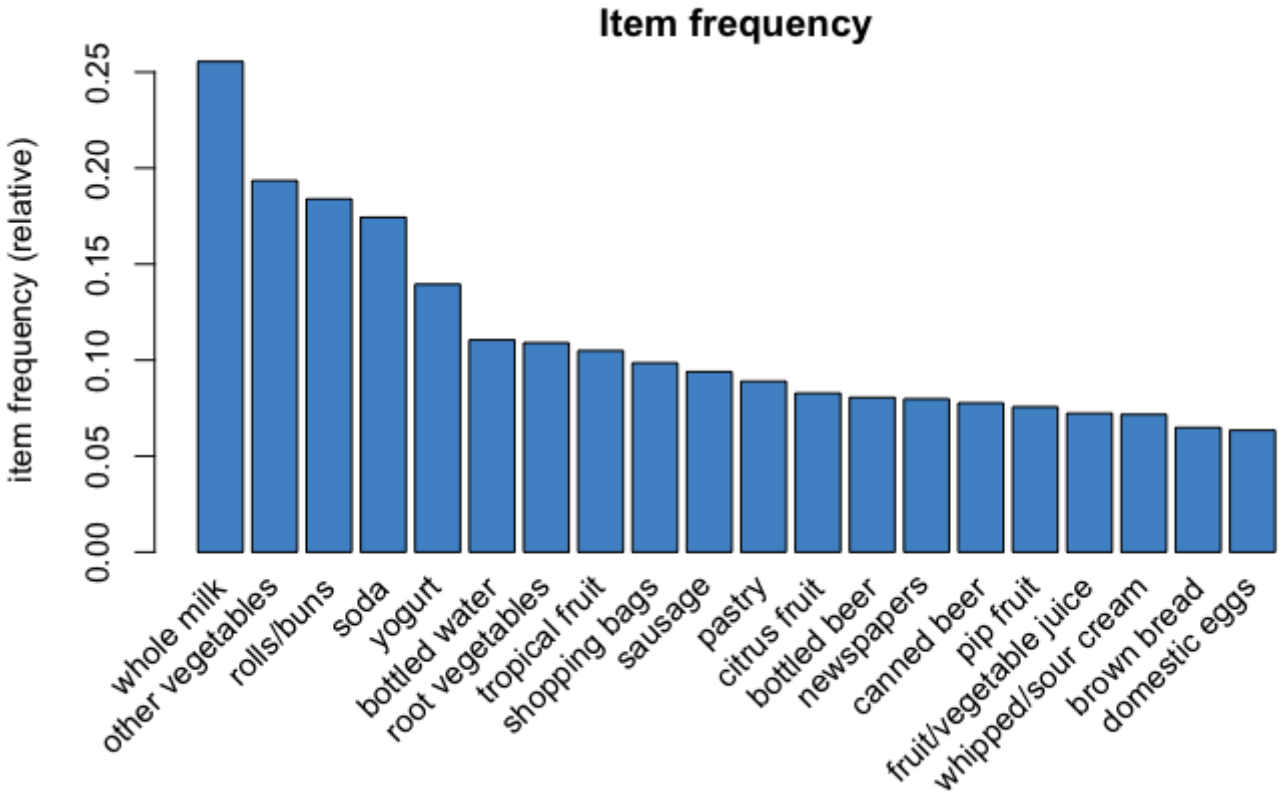
Hide

```
# The class of 'Groceries' is transactions and there are 9835 rows and 169 columns.
```

2. Generate an item frequency barplot for the grocery items with support rate greater than 0.05. Include a screenshot of your results, along with the code you used to do this.

Hide

```
itemFrequencyPlot(Groceries,
                  type="relative",
                  topN=20,
                  support=0.05,
                  horiz=F,
                  col='steelblue3',
                  xlab='',
                  main='Item frequency')
```



3. Now, create a subset of rules that contain your grocery item (you can find your item in the spreadsheet in Blackboard, in Class Discussions From Your Instructor). Select 4 different rules, (2 lhs and 2 rhs), and explain them in the way you would explain them to your roommate (I'm assuming your roommate is a smart person who is unfamiliar with data mining). Remember, every rule has three components: support, confidence, and lift. For each group of rules (grocery item on left-hand side, and grocery item on right-hand side), include a screenshot of your rules, along with the code you used to generate the rules. In a sentence or two, explain what meaning these rules might have for a supermarket retailer, such as Star Market. What could it do with this information?

Hide

```
# determine items for lhs
rules <- apriori (data=Groceries, parameter=list (supp=0.001,conf = 0.08),
                  appearance = list (default="lhs",rhs="whipped/sour cream"),
                  control = list (verbose=F)) # get rules that lead to buying 'whipped/s
our cream'
rules_conf_lhs <- sort (rules, by="confidence", decreasing=TRUE) # 'high-confidence' rul
es.
inspect(head(rules_conf_lhs))
```

```
    lhs                          rhs                     support confidence     lift coun
t
[1] {other vegetables,
     butter,
     sugar}                  => {whipped/sour cream} 0.001016777  0.7142857 9.964539    1
0
[2] {whole milk,
     butter,
     hard cheese}            => {whipped/sour cream} 0.001423488  0.6666667 9.300236    1
4
[3] {tropical fruit,
     other vegetables,
     butter,
     fruit/vegetable juice} => {whipped/sour cream} 0.001016777  0.6666667 9.300236    1
0
[4] {whole milk,
     curd,
     yogurt,
     cream cheese }          => {whipped/sour cream} 0.001118454  0.6470588 9.026700    1
1
[5] {butter,
     yogurt,
     hard cheese}            => {whipped/sour cream} 0.001016777  0.6250000 8.718972    1
0
[6] {curd,
     yogurt,
     sugar}                  => {whipped/sour cream} 0.001016777  0.6250000 8.718972    1
0
```

Hide

```
# determine items for rhs
rules <- apriori (data=Groceries, parameter=list (supp=0.001,conf = 0.08, minlen=2),
                  appearance = list(default="rhs",lhs="whipped/sour cream"),
                  control = list (verbose=F)) # those who bought 'whipped/sour cream' al
so bought..
rules_conf_rhs <- sort (rules, by="confidence", decreasing=TRUE)
inspect(head(rules_conf_rhs))
```

	lhs		rhs	support	confidence	lift	co…
	<fctr>		<fctr>	<dbl>	<dbl>	<dbl>	<int>
[1]	{whipped/sour cream}	=>	{whole milk}	0.03223183	0.4496454	1.759754	317
[2]	{whipped/sour cream}	=>	{other vegetables}	0.02887646	0.4028369	2.081924	284
[3]	{whipped/sour cream}	=>	{yogurt}	0.02074225	0.2893617	2.074251	204
[4]	{whipped/sour cream}	=>	{root vegetables}	0.01708185	0.2382979	2.186250	168
[5]	{whipped/sour cream}	=>	{rolls/buns}	0.01464159	0.2042553	1.110476	144
[6]	{whipped/sour cream}	=>	{tropical fruit}	0.01382816	0.1929078	1.838419	136

6 rows

Hide

```
# create rules
rules.sub1 <- subset(rules_conf_lhs, subset = lhs %in% c("whole milk", "whipped/sour cre
am"))
inspect(head(rules.sub1))
```

```
    lhs                   rhs                      support confidence     lift count
[1] {whole milk,
     butter,
     hard cheese}      => {whipped/sour cream} 0.001423488  0.6666667 9.300236    14
[2] {whole milk,
     curd,
     yogurt,
     cream cheese }    => {whipped/sour cream} 0.001118454  0.6470588 9.026700    11
[3] {other vegetables,
     whole milk,
     butter,
     soda}             => {whipped/sour cream} 0.001016777  0.6250000 8.718972    10
[4] {whole milk,
     butter,
     sliced cheese}    => {whipped/sour cream} 0.001220132  0.6000000 8.370213    12
[5] {root vegetables,
     whole milk,
     flour}            => {whipped/sour cream} 0.001728521  0.5862069 8.177794    17
[6] {citrus fruit,
     other vegetables,
     whole milk,
     cream cheese }    => {whipped/sour cream} 0.001118454  0.5789474 8.076521    11
```

Hide

```
rules.sub2 <- subset(rules_conf_lhs, subset = lhs %in% c("yogurt", "whipped/sour cream"
))
inspect(head(rules.sub2))
```

	lhs		rhs	sup
	<fctr>	<fctr>	<fctr>	<
[1]	{whole milk,curd,yogurt,cream cheese }	=>	{whipped/sour cream}	0.00111
[2]	{butter,yogurt,hard cheese}	=>	{whipped/sour cream}	0.00101
[3]	{curd,yogurt,sugar}	=>	{whipped/sour cream}	0.00101
[4]	{other vegetables,curd,yogurt,cream cheese }	=>	{whipped/sour cream}	0.00101
[5]	{hamburger meat,butter,yogurt}	=>	{whipped/sour cream}	0.00101
[6]	{root vegetables,yogurt,sliced cheese}	=>	{whipped/sour cream}	0.00122

6 rows | 1-6 of 7 columns

Hide

```
rules.sub3 <- subset(rules_conf_rhs, subset = rhs %in% c("other vegetables", "whipped/so
ur cream"))
inspect(head(rules.sub3))
```

	lhs		rhs	support	confidence	lift	co…
	<fctr>		<fctr>	<dbl>	<dbl>	<dbl>	<int>
[1]	{whipped/sour cream}	=>	{other vegetables}	0.02887646	0.4028369	2.081924	284

1 row

Hide

```
rules.sub4 <- subset(rules_conf_rhs, subset = rhs %in% c("tropical fruit", "whipped/sour
cream"))
inspect(head(rules.sub4))
```

	lhs		rhs	support	confidence	lift	co…
	<fctr>		<fctr>	<dbl>	<dbl>	<dbl>	<int>
[1]	{whipped/sour cream}	=>	{tropical fruit}	0.01382816	0.1929078	1.838419	136

1 row

Hide

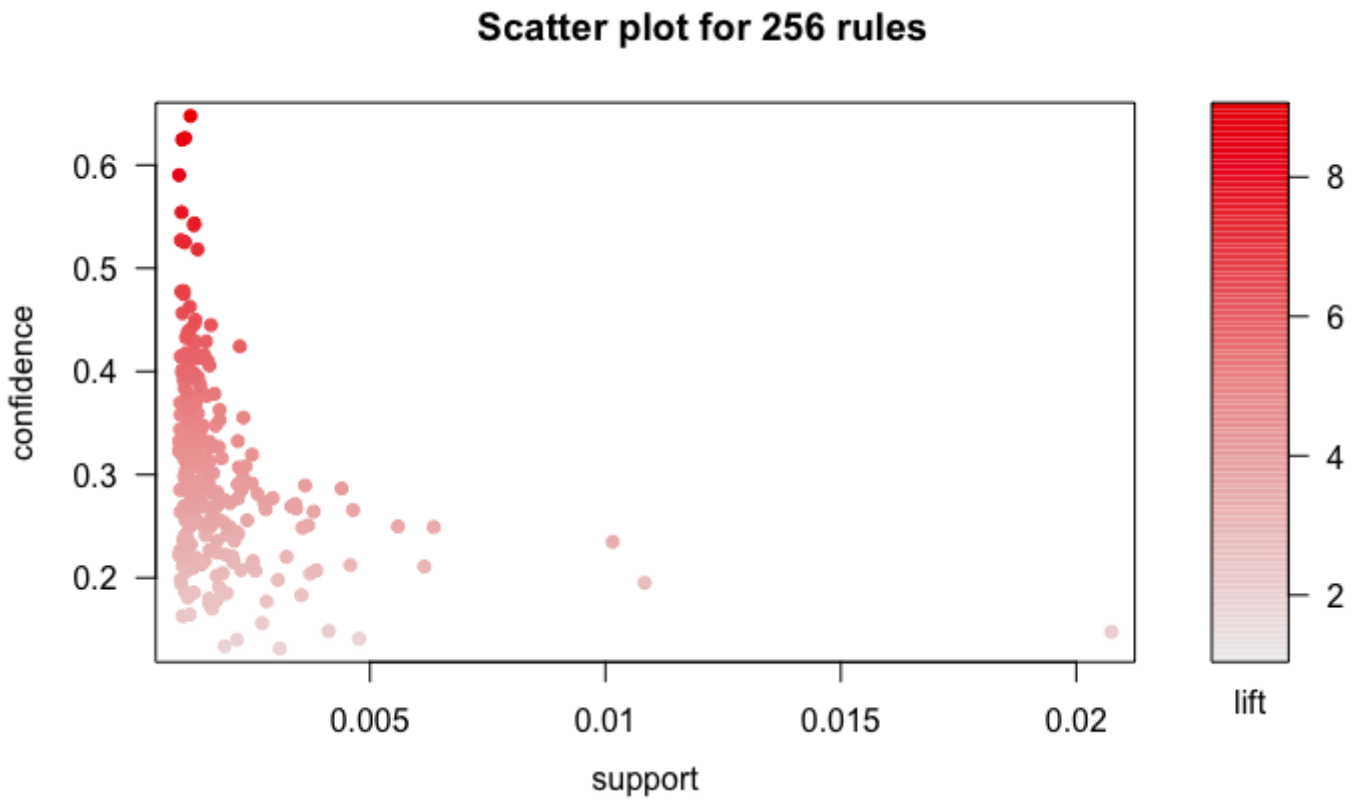
```
# When examining the LHS and RHS you are given an indication on what items are associate
d with the before and after purchase of whipped/sour cream. When we observe rules.sub1 f
or the left hand side or LHS, the buyer purchased milk will purchase whipped/sour cream
 9.3 times when observing the lift. In rules.sub4 observing the right hand side or RHS,
 purchasing whipped/sour cream will lead to 1.838 times purchasing tropical fruit. The s
ignificance behind these rules are numerous for supermarkets as it will allow them to or
ganize their food stock per store layout and manage their inventory more effectively. It
also allows buyers to shop faster.
```

4. Using the plot() function in the arulesViz package, generate a scatter plot of any three rules involving your grocery item. Include a screenshot of your plot, along with the code you used to generate the plot. Describe your results in a sentence or two.

Hide

```
plot(rules.sub2, interactive = F)
```

```
The parameter interactive is deprecated. Use engine='interactive' instead.
```



Hide

```
# When we examine rules.sub2 for for yogurt and whipped/sour cream, both ietms have a ve
ry low support but a high confidence. However the lift is high for both itmes meaning th
at the association of whipped/sour cream along with yogurt is highly likely.
```

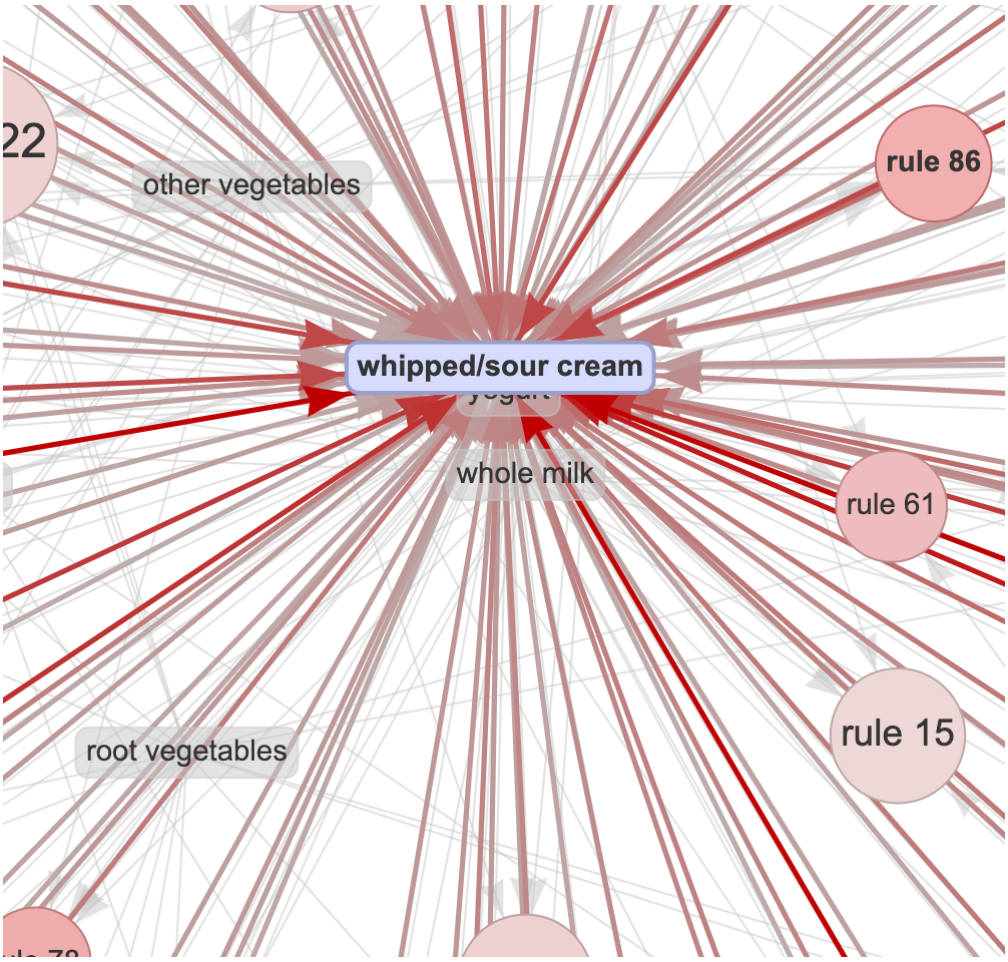
5. Again using the plot() function in the arulesViz package, generate a plot for any three of your rules. This time, add two more arguments to the function: method="graph", engine="htmlwidget". What do you see now? Include a screenshot of your plot, along with the code you used to generate the plot. Describe your results in a sentence or two.

Hide

```
plot(rules.sub2, method="graph", engine="htmlwidget")
```

```
Too many rules supplied. Only plotting the best 100 rules using lift (change control par
ameter max if needed)
```

whipped/sour cream



Hide

```
# When you filter the following graph by whipped/cream, you will find the closses asocia
ted items and rules associated with the item. We can see that yogurt and whipped/sour cr
eam are highly associated in rules.sub2.
```