# Step 0: Install and load R libraries.

# 1. Required libraries: arules, arulesViz, and RColorBrewer. Install them if they are not already installed

library(arules)

library(arulesViz)

library(RColorBrewer)

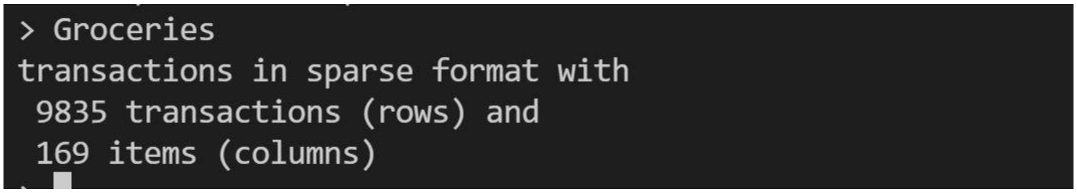
# 2. The arulesViz and RColorBrewer libraries will assist in visualizing the data.

# 3. For more information on those libraries, type ?LibraryName in you preferred R IDE.

# Step 1: Load the dataset and Explore the data.

data("Groceries")

# a. Type Groceries in the console



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

# b. How many transactions are there in the dataset?

9835 transactions

# c. How many items are there in the Groceries dataset?

169 items

# 2. Most popular items:

arules::itemFrequencyPlot(Groceries, topN = 20,col = brewer.pal(8, 'Dark2'), main =

'Relative Item Frequency Plot', type = "relative", ylab = "Item Frequency (Relative)")

# a. What is the most popular item?

whole milk

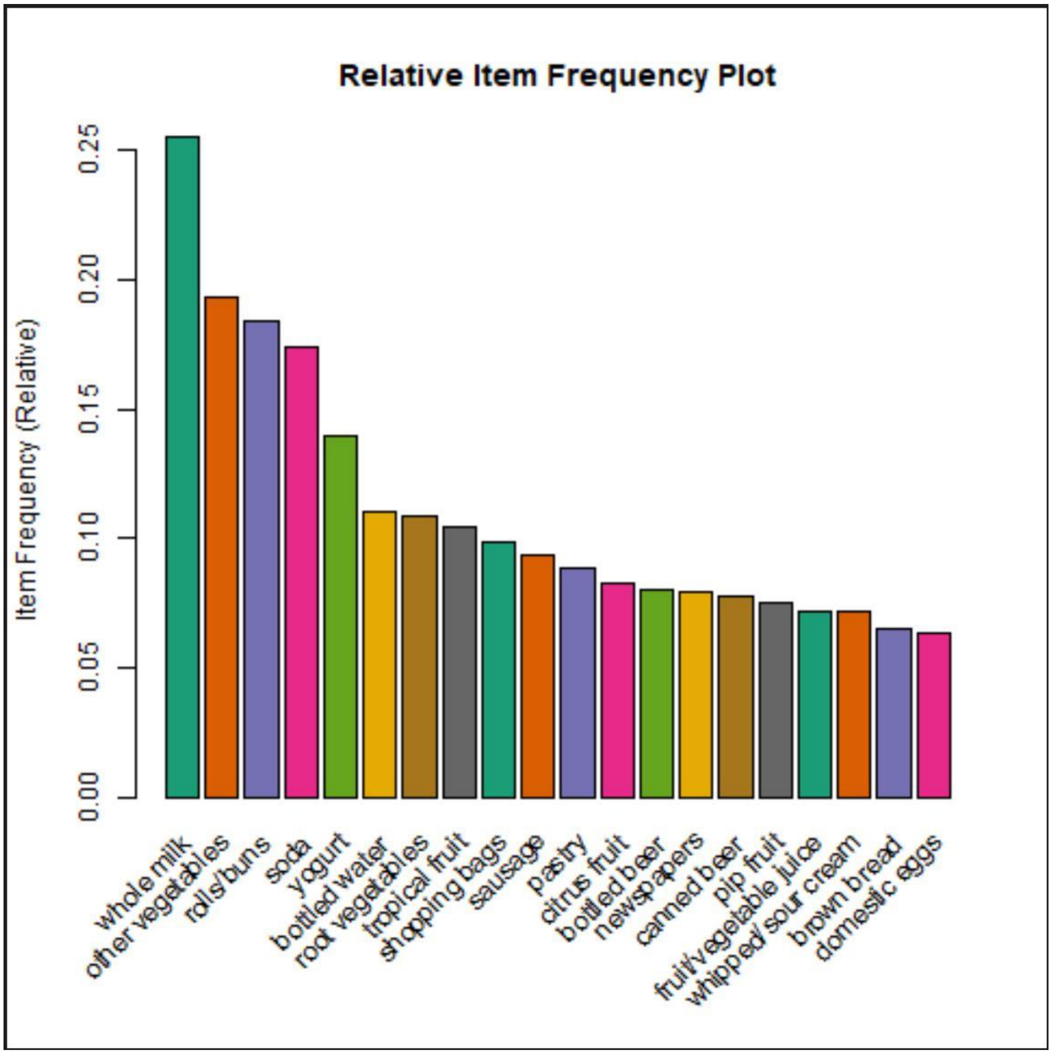
# b. What is the least popular item?

domestic eggs

# c. How many items occur more than 15% by count? What are they?

4 items

d. Save the plot.

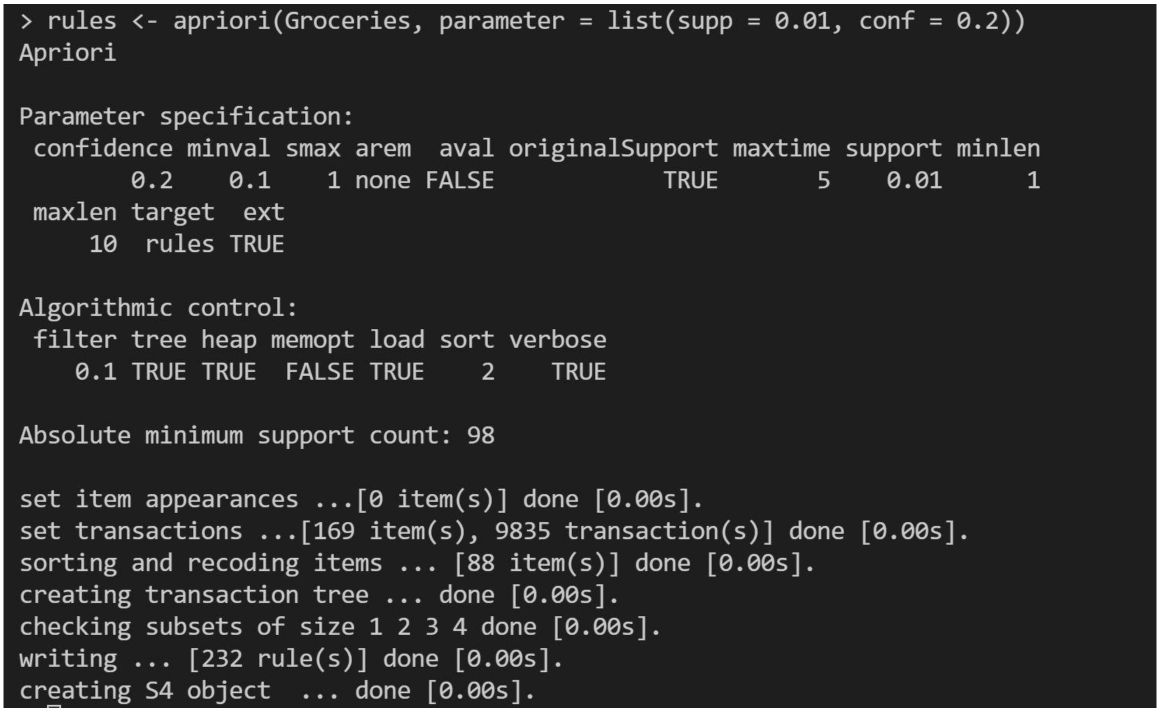


# Step 2: Use the apriori library to extract the rules.

# 1. Assign the association output:

rules <- apriori(Groceries, parameter = list(supp = 0.01, conf = 0.2))

# a. Capture the output. Save it.



```
> rules <- apriori(Groceries, parameter = list(supp = 0.01, conf = 0.2))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
        0.2    0.1    1 none FALSE            TRUE       5    0.01      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: 98

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 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 ... [232 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

# b. What is the minimum itemset length?

1

# c. What is the maximum itemset length?

10

# d. What is the confidence?

0.2

# e. What is the support?

0.01

# f. How many rules were generated?

232 rules

# g. Use summary(rules) to find out how many rules have one items, two items, three items?

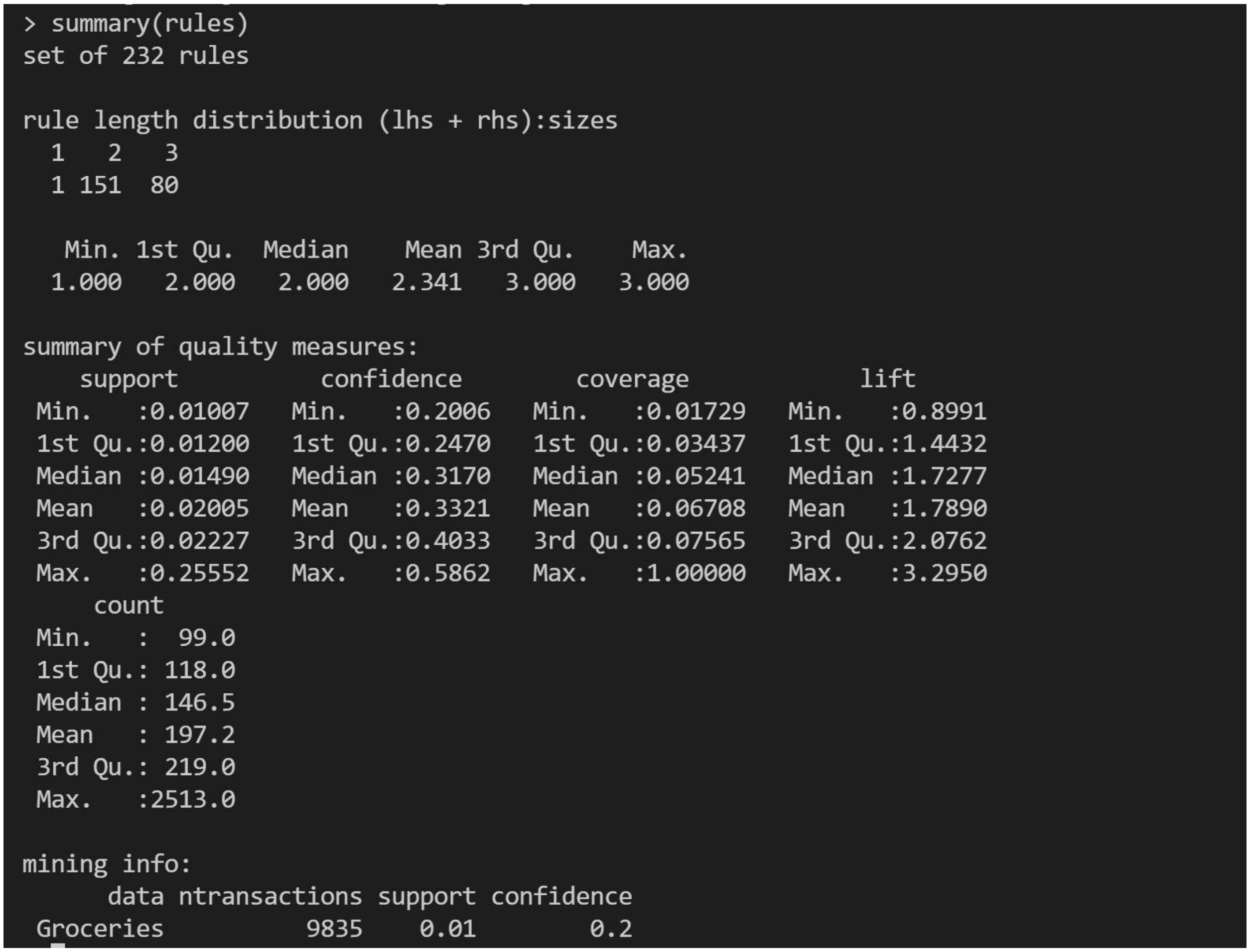
summary(rules)

1 = 1

2 = 151

3 = 80

# i. Save the summary(rules) output.

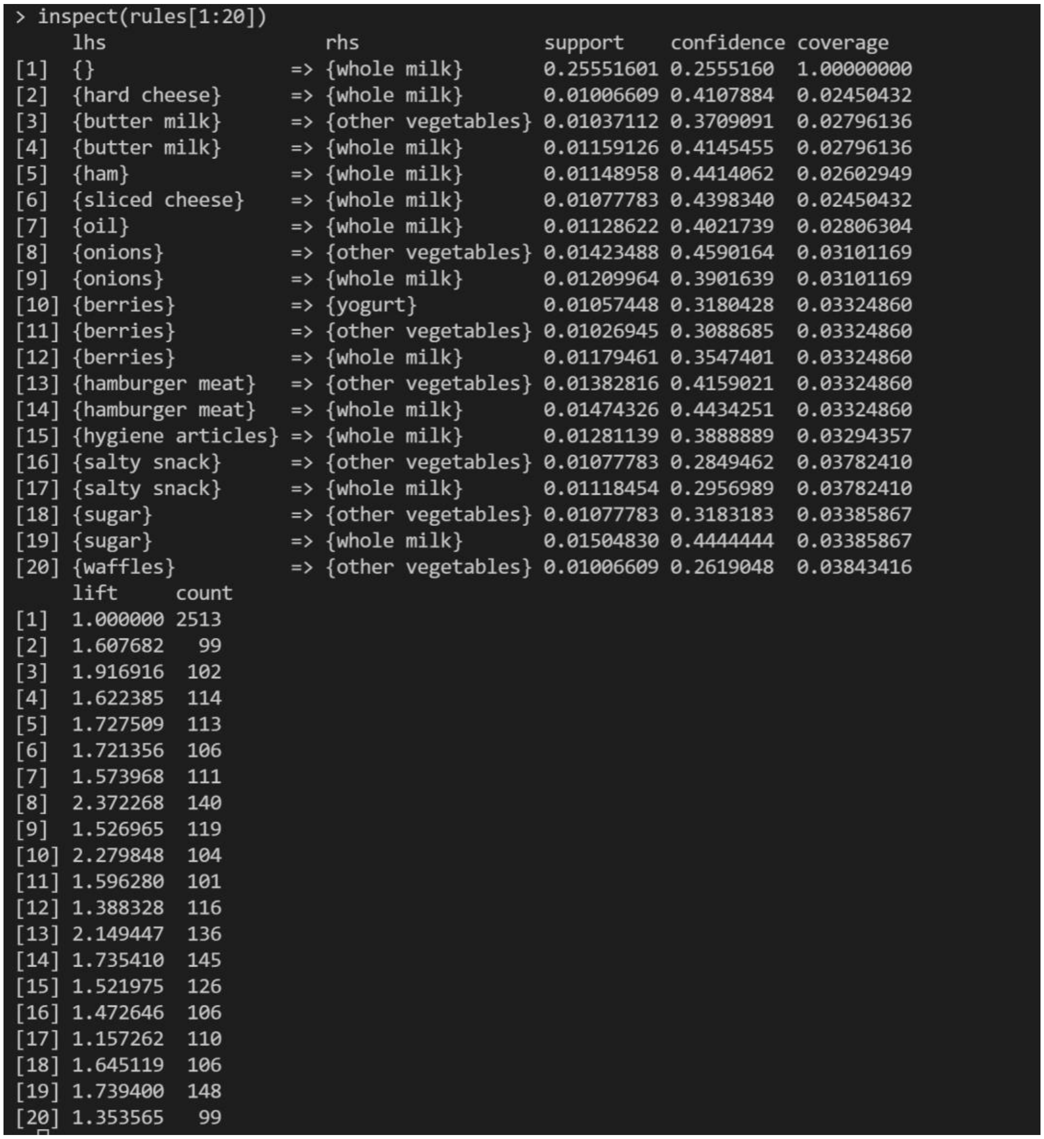


```
> summary(rules)
set of 232 rules

rule length distribution (lhs + rhs):sizes
  1   2   3 
  1 151  80 

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   2.000   2.000   2.341   3.000   3.000 

summary of quality measures:
    support          confidence        coverage            lift       
 Min.   :0.01007   Min.   :0.2006   Min.   :0.01729   Min.   :0.8991  
 1st Qu.:0.01200   1st Qu.:0.2470   1st Qu.:0.03437   1st Qu.:1.4432  
 Median :0.01490   Median :0.3170   Median :0.05241   Median :1.7277  
 Mean   :0.02005   Mean   :0.3321   Mean   :0.06708   Mean   :1.7890  
 3rd Qu.:0.02227   3rd Qu.:0.4033   3rd Qu.:0.07565   3rd Qu.:2.0762  
 Max.   :0.25552   Max.   :0.5862   Max.   :1.00000   Max.   :3.2950  
     count       
 Min.   :  99.0  
 1st Qu.: 118.0  
 Median : 146.5  
 Mean   : 197.2  
 3rd Qu.: 219.0  
 Max.   :2513.0  

mining info:
      data ntransactions support confidence
 Groceries          9835    0.01        0.2
```

# 3. Let us inspect the top 20 rules: inspect(rules([1:20])

inspect(rules[1:20])

inspect(sort(rules[1:20], by="lift"))



```
> inspect(rules[1:20])
     lhs                   rhs                support    confidence coverage
[1]  {}                 => {whole milk}       0.25551601 0.2555160  1.00000000
[2]  {hard cheese}      => {whole milk}       0.01006609 0.4107884  0.02450432
[3]  {butter milk}      => {other vegetables} 0.01037112 0.3709091  0.02796136
[4]  {butter milk}      => {whole milk}       0.01159126 0.4145455  0.02796136
[5]  {ham}              => {whole milk}       0.01148958 0.4414062  0.02602949
[6]  {sliced cheese}    => {whole milk}       0.01077783 0.4398340  0.02450432
[7]  {oil}              => {whole milk}       0.01128622 0.4021739  0.02806304
[8]  {onions}           => {other vegetables} 0.01423488 0.4590164  0.03101169
[9]  {onions}           => {whole milk}       0.01209964 0.3901639  0.03101169
[10] {berries}          => {yogurt}           0.01057448 0.3180428  0.03324860
[11] {berries}          => {other vegetables} 0.01026945 0.3088685  0.03324860
[12] {berries}          => {whole milk}       0.01179461 0.3547401  0.03324860
[13] {hamburger meat}   => {other vegetables} 0.01382816 0.4159021  0.03324860
[14] {hamburger meat}   => {whole milk}       0.01474326 0.4434251  0.03324860
[15] {hygiene articles} => {whole milk}       0.01281139 0.3888889  0.03294357
[16] {salty snack}      => {other vegetables} 0.01077783 0.2849462  0.03782410
[17] {salty snack}      => {whole milk}       0.01118454 0.2956989  0.03782410
[18] {sugar}            => {other vegetables} 0.01077783 0.3183183  0.03385867
[19] {sugar}            => {whole milk}       0.01504830 0.4444444  0.03385867
[20] {waffles}          => {other vegetables} 0.01006609 0.2619048  0.03843416
     lift     count
[1]  1.000000 2513
[2]  1.607682   99
[3]  1.916916  102
[4]  1.622385  114
[5]  1.727509  113
[6]  1.721356  106
[7]  1.573968  111
[8]  2.372268  140
[9]  1.526965  119
[10] 2.279848  104
[11] 1.596280  101
[12] 1.388328  116
[13] 2.149447  136
[14] 1.735410  145
[15] 1.521975  126
[16] 1.472646  106
[17] 1.157262  110
[18] 1.645119  106
[19] 1.739400  148
[20] 1.353565   99
```

# a. What top three association rules have the highest Lift? Name each three association rule numbers, lefthand side (lhs) and right-hand side (rhs).

rule	left hand side (lhs)	right hand side (rhs)
8	onions	other vegetables
10	berries	yogurt
13	hamburger meat	other vegetables

# b. Comment on what you believe rule #1 means.

Whole milk is the most popular item and it can be paired with anything

# c. Is your answer supported by what you found to be the most popular item in the question 1.2.a in Step 1?

yes

# d. What is the support, confidence, and lift for a lhs of berries leading to a rhs of yogurt according to the output?

lhs	rhs	support	confidence	lift
berries	yogurt	0.01057448	0.3180428	2.279848

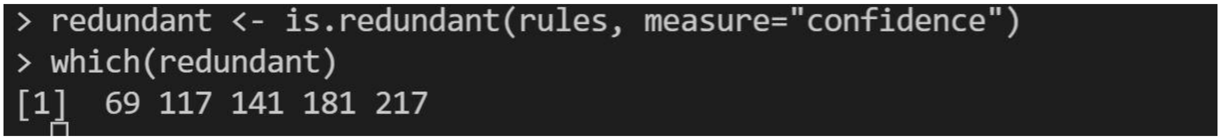
# Step 3: Finding redundancy and pruning the association rules:

# 1. To find redundant association rules. We will focus on confidence, but feel free to try support and left as well.

redundant <- is.redundant(rules, measure="confidence")

which(redundant)

# c. Save the output.



```
> redundant <- is.redundant(rules, measure="confidence")
> which(redundant)
[1]  69 117 141 181 217
```

# d. How many rules are redundant based on this criterion? Which ones? List them out.

Redundant are based on 5 rules. They are 69, 117, 141, 181, 217.

# 2. Let us do some pruning based on what was found above:

rules.pruned <- rules[!redundant]

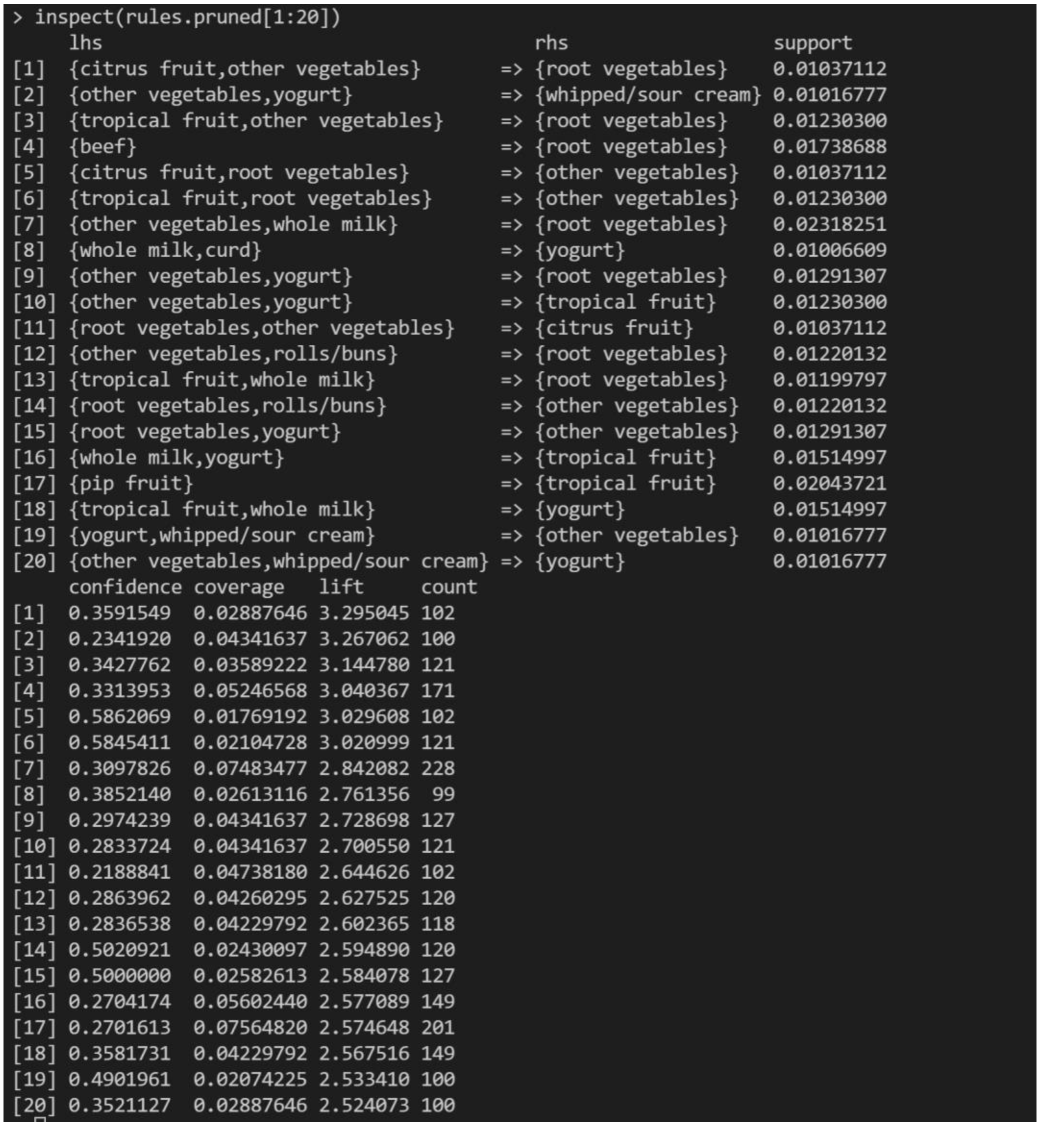
rules.pruned <- sort(rules.pruned, by="lift")

inspect(rules.pruned)

# d. It is a long list. Printout the top 20 with inspect(rules.pruned[1:20])

# i. Save the output.

inspect(rules.pruned[1:20])



```
> inspect(rules.pruned[1:20])
     lhs                                      rhs                  support
[1]  {citrus fruit,other vegetables}       => {root vegetables}    0.01037112
[2]  {other vegetables,yogurt}             => {whipped/sour cream} 0.01016777
[3]  {tropical fruit,other vegetables}     => {root vegetables}    0.01230300
[4]  {beef}                                => {root vegetables}    0.01738688
[5]  {citrus fruit,root vegetables}        => {other vegetables}   0.01037112
[6]  {tropical fruit,root vegetables}      => {other vegetables}   0.01230300
[7]  {other vegetables,whole milk}         => {root vegetables}    0.02318251
[8]  {whole milk,curd}                     => {yogurt}             0.01006609
[9]  {other vegetables,yogurt}             => {root vegetables}    0.01291307
[10] {other vegetables,yogurt}             => {tropical fruit}     0.01230300
[11] {root vegetables,other vegetables}    => {citrus fruit}       0.01037112
[12] {other vegetables,rolls/buns}         => {root vegetables}    0.01220132
[13] {tropical fruit,whole milk}           => {root vegetables}    0.01199797
[14] {root vegetables,rolls/buns}          => {other vegetables}   0.01220132
[15] {root vegetables,yogurt}              => {other vegetables}   0.01291307
[16] {whole milk,yogurt}                   => {tropical fruit}     0.01514997
[17] {pip fruit}                           => {tropical fruit}     0.02043721
[18] {tropical fruit,whole milk}           => {yogurt}             0.01514997
[19] {yogurt,whipped/sour cream}           => {other vegetables}   0.01016777
[20] {other vegetables,whipped/sour cream} => {yogurt}             0.01016777
     confidence coverage   lift     count
[1]  0.3591549  0.02887646 3.295045 102
[2]  0.2341920  0.04341637 3.267062 100
[3]  0.3427762  0.03589222 3.144780 121
[4]  0.3313953  0.05246568 3.040367 171
[5]  0.5862069  0.01769192 3.029608 102
[6]  0.5845411  0.02104728 3.020999 121
[7]  0.3097826  0.07483477 2.842082 228
[8]  0.3852140  0.02613116 2.761356  99
[9]  0.2974239  0.04341637 2.728698 127
[10] 0.2833724  0.04341637 2.700550 121
[11] 0.2188841  0.04738180 2.644626 102
[12] 0.2863962  0.04260295 2.627525 120
[13] 0.2836538  0.04229792 2.602365 118
[14] 0.5020921  0.02430097 2.594890 120
[15] 0.5000000  0.02582613 2.584078 127
[16] 0.2704174  0.05602440 2.577089 149
[17] 0.2701613  0.07564820 2.574648 201
[18] 0.3581731  0.04229792 2.567516 149
[19] 0.4901961  0.02074225 2.533410 100
[20] 0.3521127  0.02887646 2.524073 100
```

# e. What are the top three rules after pruning based on lift? List them.

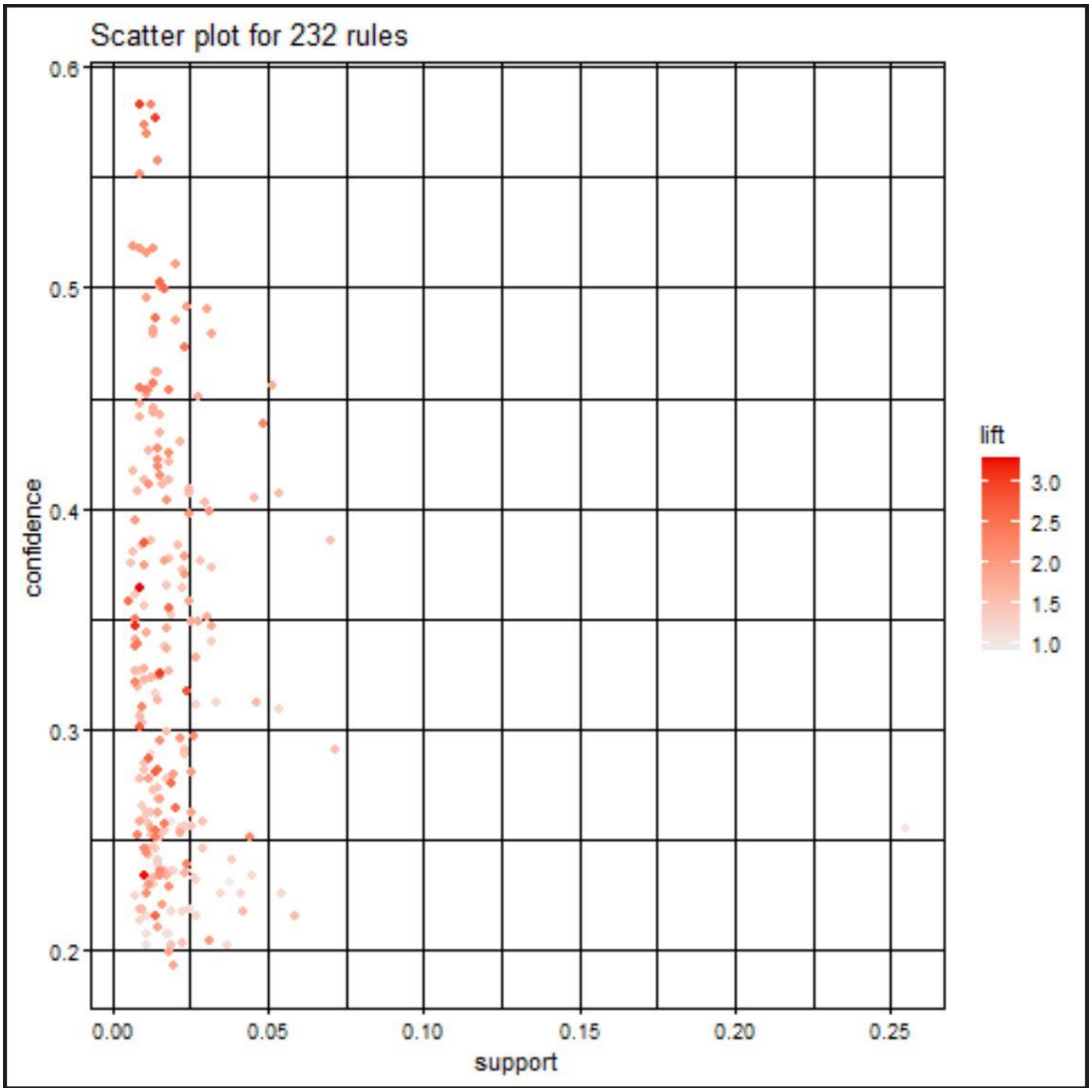
lhs	rhs
citrus fruit, other vegetables	root vegetables
other vegetables, yogurt	whipped/sour cream
tropical fruit, other vegetables	root vegetables

# Step 4: Visualizing the association rules:

# 1. Make sure you have loaded arulesViz.

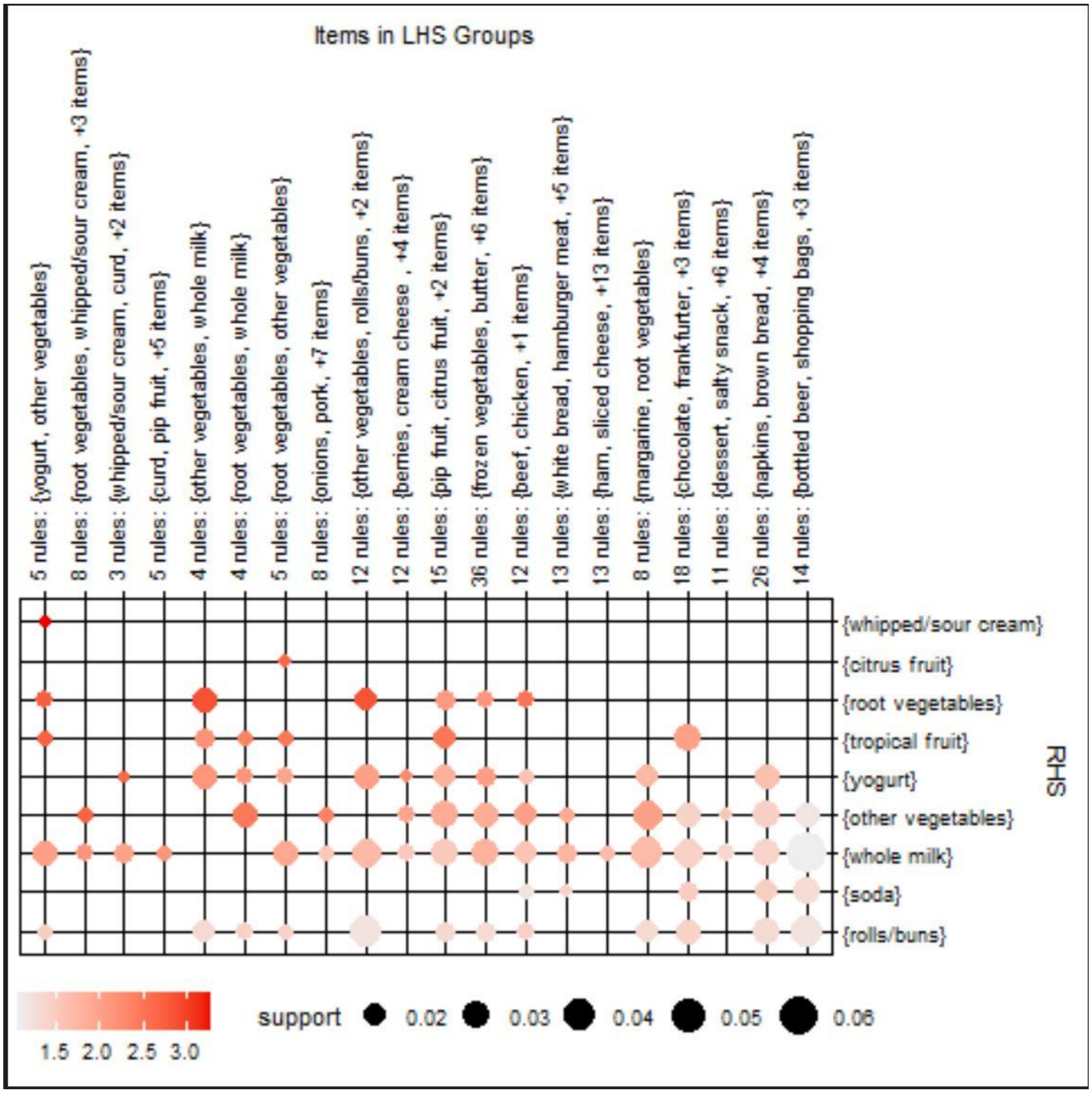
plot(rules)

# a. Save the plot.



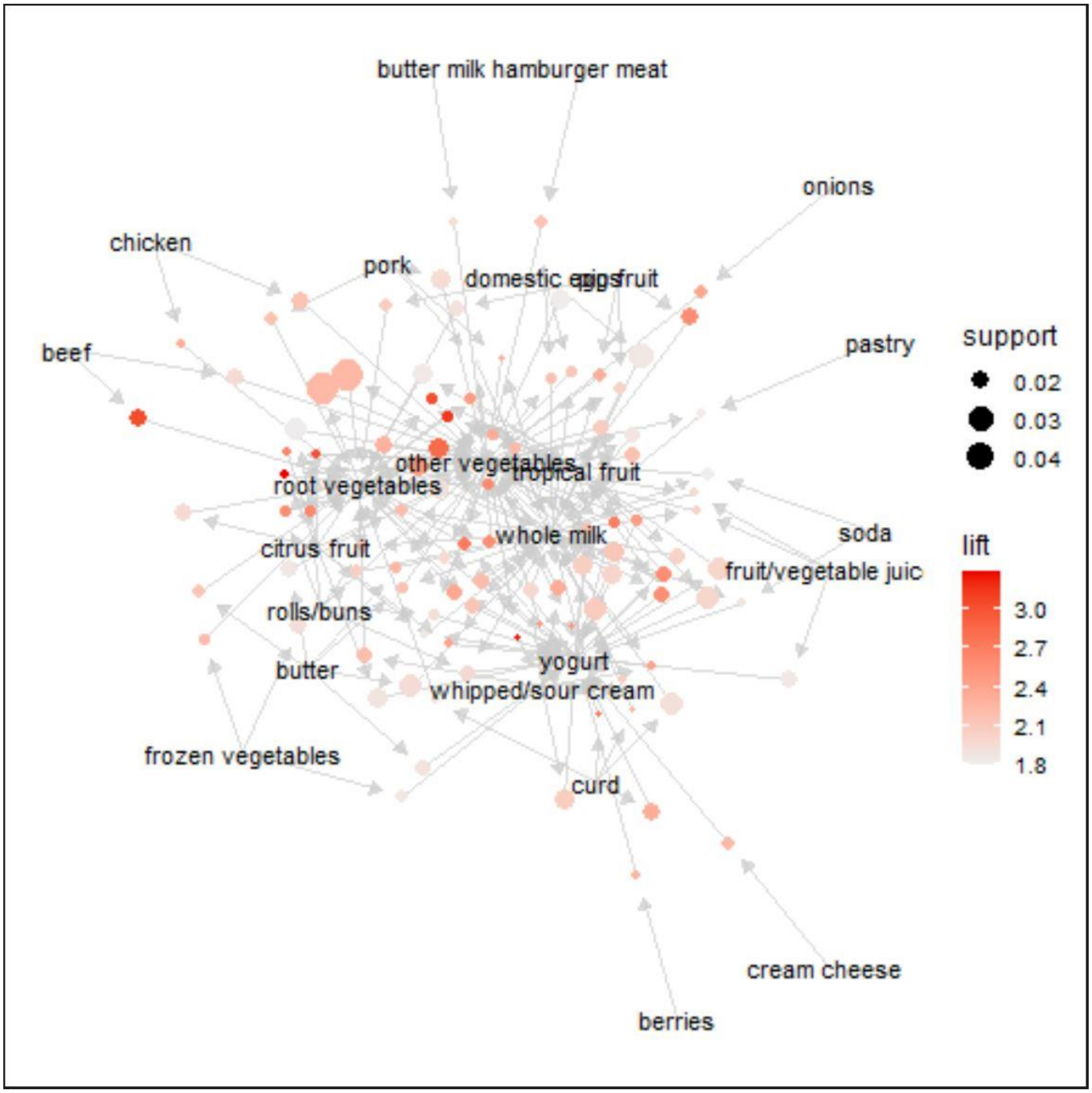
plot(rules, method="grouped")

# a. Save the plot.



plot(rules, method="graph")

# a. Save the plot.



# 5. What items have the most associations based on the plots? Does that confirm your understanding of the association rules in the Groceries dataset? Any surprises?

Based on the plots following items have most associations:

Whole milk,
root vegetables,
yogurt whipped/ sour cream,
other vegetables

Yes, the plot does help me confirm my understanding. There are no surprises.