Bakery Market Basket Analysis

Stone Shi

11/18/2019

Contents

This dataset is transactions from a bakery.

1. Collect the data

```
library(arules)
library(tidyverse)

bread <- read.transactions("BreadBasket_DMS.csv", cols = c(3,4), sep = ",", format = "single", quote =
```

2. Exploratory Data Analysis

2.1 Basic Summary Output

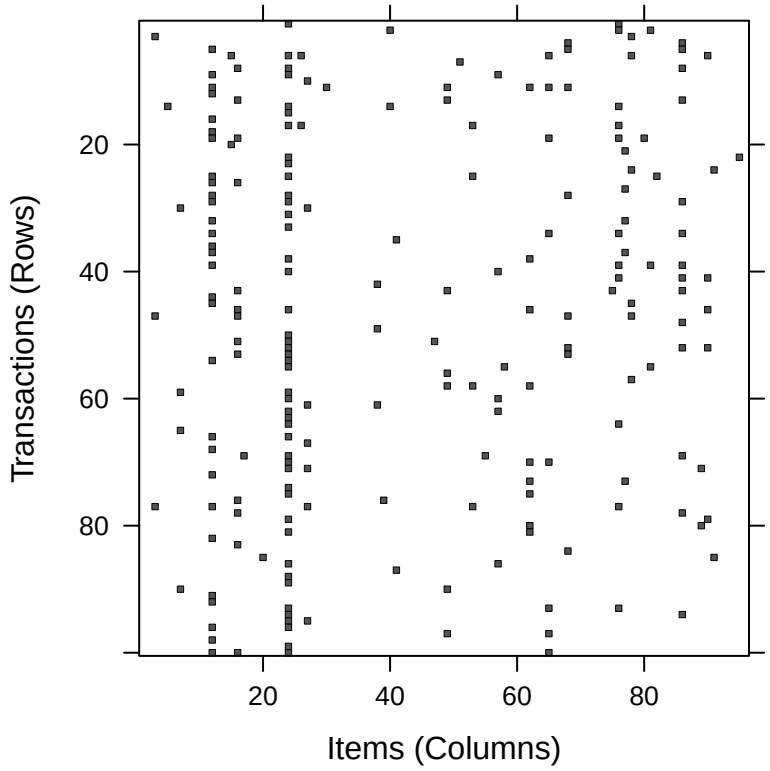
Let's take a look at the summary stats for the dataset. Through the summary output, we can see that there are in total **6614** transactions and **104** items in my dataset. The top 5 frequent items are: *Coffee*, *Bread*, *Tea*, *Cake* and *Pastry*. The length distribution tells us there are 2556 transactions contain one item, 2154 transactions contain two items, 1078 transactions contain 3 items, etc. The maximum count of item in a single transaction is 10.

```
summary(bread)
```

```
## transactions as itemMatrix in sparse format with
##  9532 rows (elements/itemsets/transactions) and
##  96 columns (items) and a density of 0.02146388
##
## most frequent items:
##  Coffee   Bread     Tea    Cake  Pastry (Other)
##    4528    3097    1350     983     815    8868
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10
## 3775 3101 1537  738  246   96   27    6    2    4
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   2.061   3.000  10.000
##
## includes extended item information - examples:
##                     labels
## 1               Adjustment
## 2 Afternoon with the baker
## 3                Alfajores
##
## includes extended transaction information - examples:
##   transactionID
## 1             1
## 2            10
## 3           100
```

2.2 Transactions Exploration

```
image(sample(bread,100))
```



Let's look at our first five transactions.

```
inspect(bread[1:100])
```

```
##       items                                      transactionID
## [1]   {Bread}                                    1
## [2]   {Medialuna,Scandinavian}                   10
## [3]   {Bread}                                    100
## [4]   {Chimichurri Oil,Scandinavian}             1000
## [5]   {Bread,Truffles}                           1001
## [6]   {Brownie,Focaccia}                         1002
## [7]   {Bread,Coffee}                             1003
## [8]   {Art Tray,Coffee,Cookies,Tea}              1004
## [9]   {Coffee}                                   1005
## [10]  {Bread}                                    1006
## [11]  {Alfajores,Coffee,Coke}                    1007
## [12]  {Bread}                                    1008
## [13]  {Bread}                                    1009
## [14]  {Coffee,Pastry}                            101
## [15]  {Bread,Coffee,Medialuna,Pastry}            1010
## [16]  {Coffee}                                   1011
## [17]  {Bread,Keeping It Local}                   1012
## [18]  {Bread}                                    1013
## [19]  {Coffee,Scandinavian}                      1014
## [20]  {Bread,Farm House,Medialuna,Pastry}        1015
## [21]  {Medialuna}                                1016
## [22]  {Coffee}                                   1017
## [23]  {Bread,Farm House}                         1018
## [24]  {Bread}                                    1019
```

```
## [25]  {Farm House}                                        102
## [26]  {Bread,Keeping It Local}                            1020
## [27]  {Bread,Medialuna}                                   1021
## [28]  {Tea}                                               1022
## [29]  {Coffee,Keeping It Local,Pastry}                    1023
## [30]  {Coffee}                                            1024
## [31]  {Tea}                                               1025
## [32]  {Hot chocolate}                                     1026
## [33]  {Sandwich,Tea}                                      1027
## [34]  {Bread}                                             1028
## [35]  {Brownie,Medialuna,Muffin}                          1029
## [36]  {Bread,NONE}                                        103
## [37]  {Coffee,Muffin}                                     1030
## [38]  {Bread}                                             1031
## [39]  {Coffee,Medialuna}                                  1032
## [40]  {Coffee}                                            1033
## [41]  {Coffee,Medialuna}                                  1034
## [42]  {Coffee}                                            1035
## [43]  {Bread,Brownie,Coffee,Hot chocolate}                1036
## [44]  {Bread}                                             1037
## [45]  {Bread,NONE}                                        1038
## [46]  {Brownie,Coffee}                                    1039
## [47]  {Coffee,Medialuna,Muffin}                           104
## [48]  {Coffee,Ella's Kitchen Pouches,Medialuna,Pastry}    1040
## [49]  {Jam}                                               1041
## [50]  {Bread}                                             1042
## [51]  {Ella's Kitchen Pouches}                            1043
## [52]  {Coffee,Muffin}                                     1044
## [53]  {Coffee,Tea,Truffles}                               1046
## [54]  {Coffee}                                            1047
## [55]  {Brownie,Pastry}                                    1048
## [56]  {Bread,Brownie}                                     1049
## [57]  {Cake,Coffee,NONE,Tea}                              105
## [58]  {Bread,Coffee}                                      1050
## [59]  {Bread,Pastry}                                      1051
## [60]  {Alfajores,Bread}                                   1052
## [61]  {Brownie,Muffin,Scandinavian}                       1053
## [62]  {Coffee}                                            1054
## [63]  {Fudge}                                             1055
## [64]  {Brownie,Tea}                                       1056
## [65]  {Bread}                                             1057
## [66]  {Bread,Muffin}                                      1058
## [67]  {Scandinavian}                                      1059
## [68]  {Coffee,Pastry}                                     106
## [69]  {Bread}                                             1060
## [70]  {Coffee}                                            1061
## [71]  {Bread}                                             1062
## [72]  {Coffee,Soup}                                       1063
## [73]  {Soup}                                              1064
## [74]  {Alfajores,Bread,Coffee}                            1065
## [75]  {Farm House}                                        1066
## [76]  {Chimichurri Oil,Farm House}                        1068
## [77]  {Farm House}                                        1069
## [78]  {Coffee,Tea}                                        107
```

```
## [79]  {Sandwich}                                  1070
## [80]  {Bread,Soup}                                1071
## [81]  {Art Tray,Coffee,Frittata,Sandwich}         1072
## [82]  {Fairy Doors,Scandinavian}                  1073
## [83]  {Cookies,Mineral water}                     1074
## [84]  {Brownie,Coffee}                            1075
## [85]  {Bread,Hearty & Seasonal}                   1076
## [86]  {Bread,Hearty & Seasonal}                   1077
## [87]  {Focaccia,Scandinavian}                     1078
## [88]  {Farm House,Muffin}                         108
## [89]  {Fudge}                                     1080
## [90]  {Cake,Coffee,Tea}                           1081
## [91]  {Coffee}                                    1082
## [92]  {Truffles}                                  1083
## [93]  {Brownie,Coffee}                            1084
## [94]  {Scandinavian}                              1085
## [95]  {Coffee,Muffin}                             1086
## [96]  {Coke,Fudge,Hot chocolate,Soup}             1087
## [97]  {Alfajores,Juice,Sandwich}                  1088
## [98]  {Alfajores,Coffee}                          1089
## [99]  {Bread,Muffin,Pastry}                       109
## [100] {Brownie,Coffee,Hot chocolate,Juice}        1090
```

Let's get to know the frequency of our first five items from our transactions.

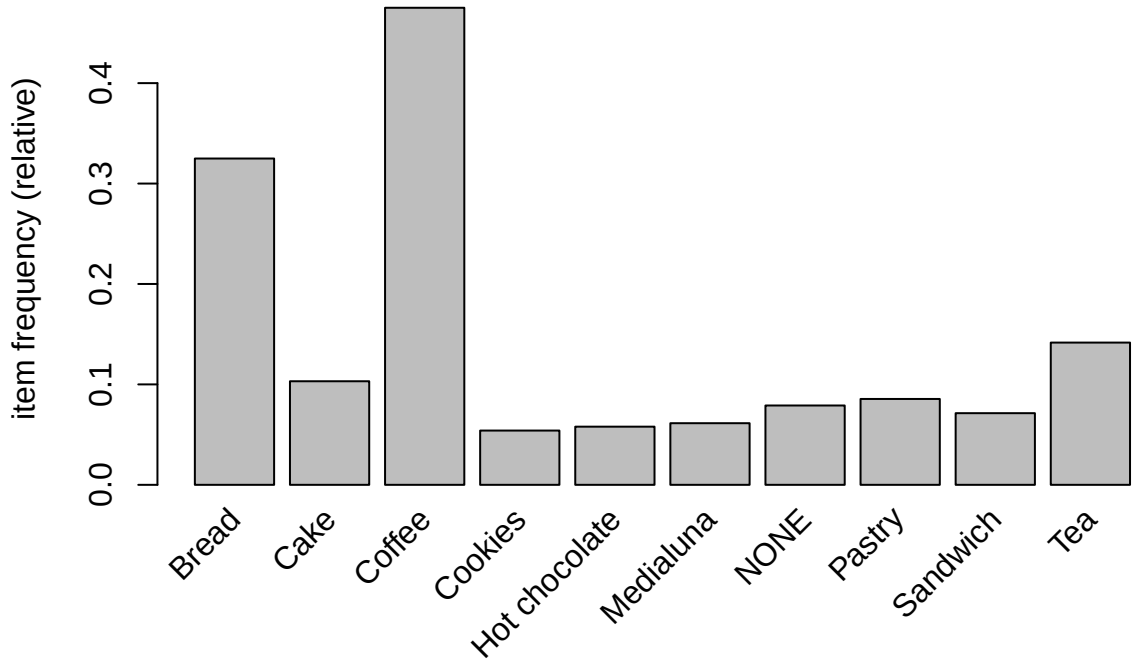
```
itemFrequency(bread[,1:5])
```

```
##               Adjustment Afternoon with the baker               Alfajores
##             0.0001049098             0.0045111204             0.0360889635
##          Argentina Night                 Art Tray
##             0.0007343684             0.0039865715
```

Then we can also utilize the *itemFrequencyPlot()* function to visualize the frequency of our items.

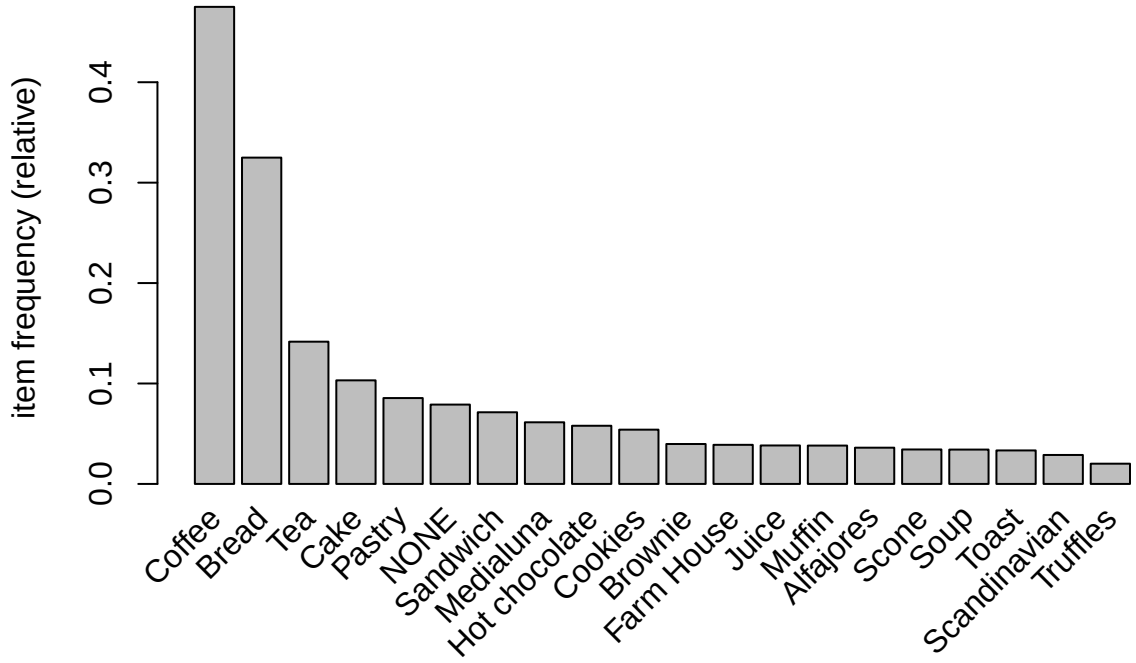
Let's produce a histogram of items with at least 5 percent support.

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



We can also produce a histogram for the top 20 frequent item sets.

```
itemFrequencyPlot(bread, topN = 20)
```



#3. Market Basket Analysis The *arules* pakcage in R has a limited number of functions for exploring the data. So in order to continue our market basket analysis, we need to create a dataframe of the items and their purchase frequency. Let's take a look at the whole itemFrequency.

```
itemFrequency(bread)
```

```
##                  Adjustment      Afternoon with the baker
```

```
##               0.0001049098               0.0045111204
##                  Alfajores             Argentina Night
##               0.0360889635               0.0007343684
##                   Art Tray                      Bacon
##               0.0039865715               0.0001049098
##                   Baguette                  Bakewell
##               0.0159462862               0.0050356693
##               Bare Popcorn                     Basket
##               0.0005245489               0.0006294587
##              Bowl Nic Pitt                      Bread
##               0.0002098196               0.3249055812
##              Bread Pudding          Brioche and salami
##               0.0004196391               0.0003147293
##                    Brownie                       Cake
##               0.0397608057               0.1031263114
##              Caramel bites       Cherry me Dried fruit
##               0.0003147293               0.0003147293
##               Chicken sand                Chicken Stew
##               0.0001049098               0.0129039026
##            Chimichurri Oil                 Chocolates
##               0.0002098196               0.0009441880
##            Christmas common                     Coffee
##               0.0011540076               0.4750314729
##            Coffee granules                       Coke
##               0.0007343684               0.0193033991
##                    Cookies                     Crepes
##               0.0540285355               0.0006294587
##                     Crisps  Drinking chocolate spoons
##               0.0014687369               0.0008392782
##                   Duck egg              Dulce de Leche
##               0.0012589173               0.0013638271
##                       Eggs       Ella's Kitchen Pouches
##               0.0029374738               0.0017834662
##                  Empanadas        Extra Salami or Feta
##               0.0007343684               0.0039865715
##                Fairy Doors                 Farm House
##               0.0002098196               0.0389215275
##                   Focaccia                   Frittata
##               0.0056651280               0.0084976920
##                      Fudge                Gift voucher
##               0.0148971884               0.0001049098
##            Gingerbread syrup                    Granola
##               0.0009441880               0.0029374738
##               Hack the stack          Half slice Monster
##               0.0002098196               0.0006294587
##            Hearty & Seasonal                      Honey
##               0.0104909778               0.0006294587
##               Hot chocolate                       Item
##               0.0579101972               0.0001049098
##                        Jam              Jammie Dodgers
##               0.0148971884               0.0131137222
##                      Juice            Keeping It Local
##               0.0382920688               0.0066093160
##               Kids biscuit           Lemon and coconut
```

```
##                   0.0012589173                 0.0006294587
##                      Medialuna                Mighty Protein
##                   0.0613722199                 0.0011540076
##                  Mineral water                     Mortimer
##                   0.0140579102                 0.0005245489
##                         Muesli                       Muffin
##                   0.0008392782                 0.0381871590
##               My-5 Fruit Shoot                    Nomad bag
##                   0.0018883760                 0.0008392782
##                           NONE                Olum & polenta
##                   0.0789970625                 0.0001049098
##                       Panatone                       Pastry
##                   0.0005245489                 0.0855014687
##             Pick and Mix Bowls                      Pintxos
##                   0.0012589173                 0.0006294587
##                        Polenta                     Postcard
##                   0.0001049098                 0.0010490978
## Raspberry shortbread sandwich                     Raw bars
##                   0.0003147293                 0.0001049098
##                          Salad                     Sandwich
##                   0.0103860680                 0.0713386488
##                   Scandinavian                        Scone
##                   0.0288501888                 0.0343054973
##                       Siblings                    Smoothies
##                   0.0002098196                 0.0080780529
##                          Soup                Spanish Brunch
##                   0.0342005875                 0.0180444817
##                         Spread                   Tacos/Fajita
##                   0.0002098196                 0.0011540076
##                        Tartine                          Tea
##                   0.0048258498                 0.1416281997
##                       The BART                    The Nomad
##                   0.0001049098                 0.0060847671
##                         Tiffin                       Toast
##                   0.0153168275                 0.0333613093
##                       Truffles                      Tshirt
##                   0.0201426773                 0.0022031053
##               Valentine's card                  Vegan Feast
##                   0.0013638271                 0.0016785564
##                  Vegan mincepie              Victorian Sponge
##                   0.0054553084                 0.0007343684
```

Now, we can convert the data to a data frame.

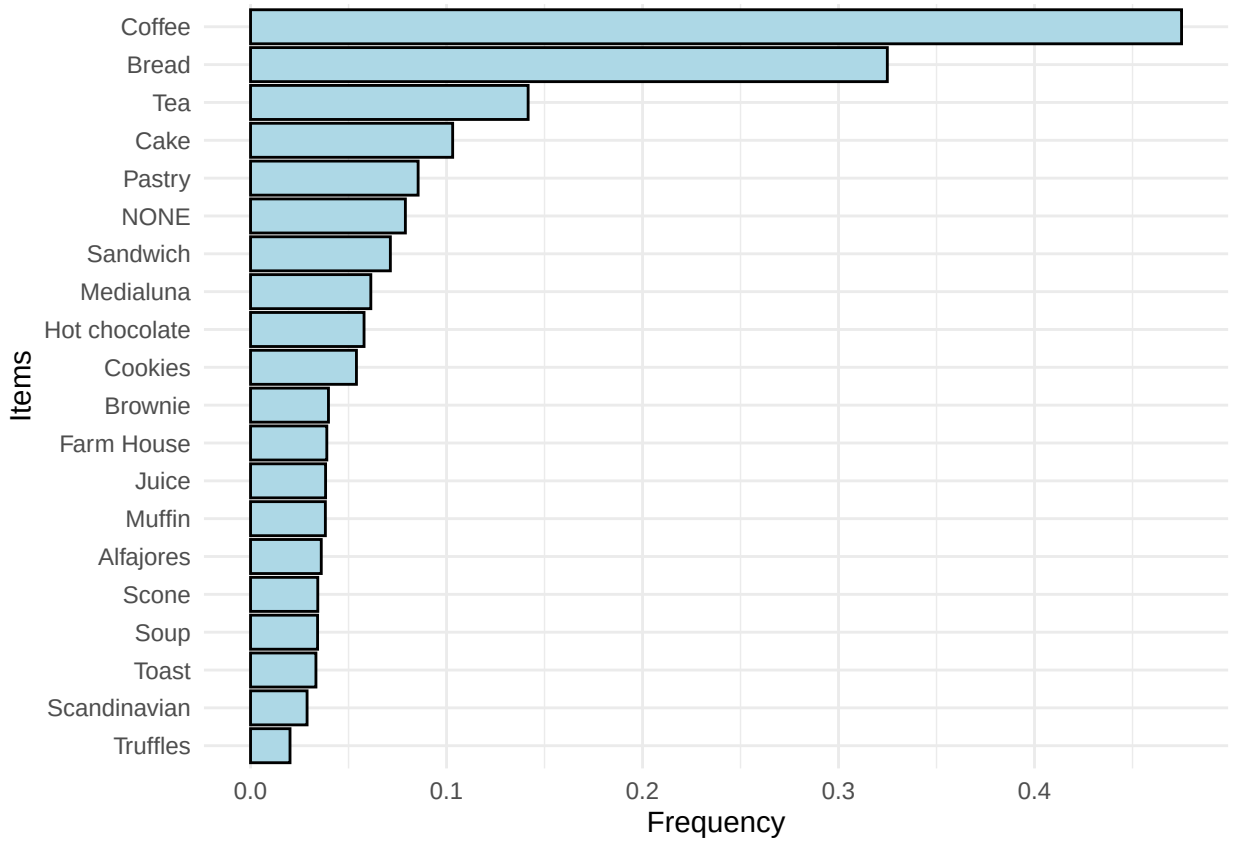
```
bakery.frequency <-
  data.frame(
    Items = names(itemFrequency(bread)),
    Frequency = itemFrequency(bread),
    row.names = NULL
  )

head(bakery.frequency)
```

```
##                       Items    Frequency
```

```
## 1               Adjustment 0.0001049098
## 2 Afternoon with the baker 0.0045111204
## 3                Alfajores 0.0360889635
## 4          Argentina Night 0.0007343684
## 5                 Art Tray 0.0039865715
## 6                    Bacon 0.0001049098
```

10 most frequently bought items at the store.

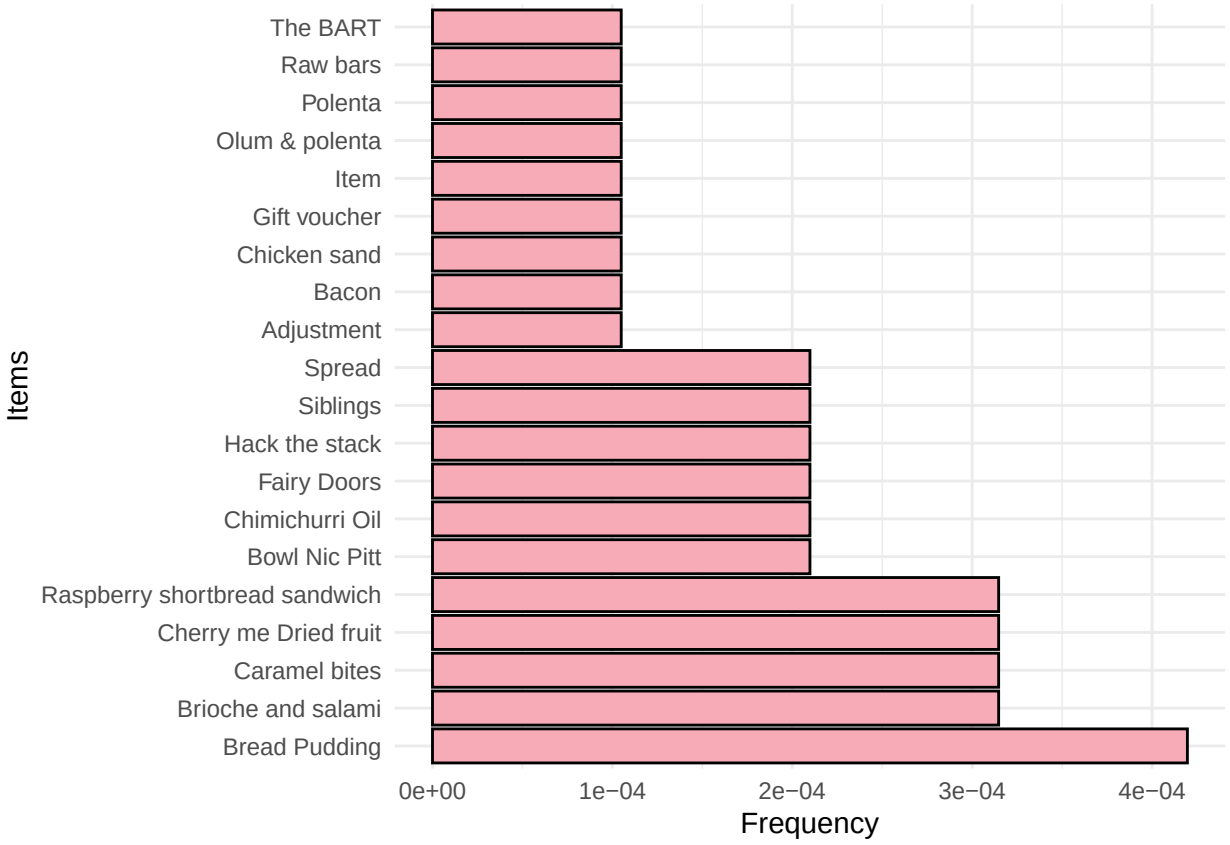
```
bakery.frequency %>%
  arrange(desc(Frequency)) %>%
  slice(1:20) %>%
  ggplot() +
  geom_col(aes(x = reorder(Items, Frequency), y = Frequency), fill = "lightblue", color = "black")+
  labs(x = "Items") +
  coord_flip() +
  theme_minimal() +
  theme(legend.position = "none")
```



10 least frequently bought items at the store.

```
bakery.frequency %>%
  arrange(Frequency) %>%
  slice(1:20) %>%
  ggplot() +
```

```
  geom_col(aes(x = reorder(Items, -Frequency), y = Frequency), fill = "#f6abb6", color = "black")+
  labs(x = "Items") +
  coord_flip() +
  theme_minimal() +
  theme(legend.position = "none")
```



3. Train the model

I will use **apriori()** function to generate some rules.

```
bakery_rules <- apriori(bread, parameter = list(support = 0.01, confidence = 0.5, minlen = 2))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5    0.01      2
##  maxlen target   ext
##      10  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
```

```
## Absolute minimum support count: 95
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[96 item(s), 9532 transaction(s)] done [0.00s].
## sorting and recoding items ... [31 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [12 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

4. "Evaluate" the Model's Performance.

4.1 Understand the rules

Let's first get the summary of the bakery association rules.

```
summary(bakery_rules)
```

```
## set of 12 rules
##
## rule length distribution (lhs + rhs):sizes
##  2
## 12
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       2       2       2       2       2       2
##
## summary of quality measures:
##     support          confidence          lift           count
##  Min.   :0.01081   Min.   :0.5072   Min.   :1.068   Min.   :103.0
##  1st Qu.:0.02022   1st Qu.:0.5260   1st Qu.:1.107   1st Qu.:192.8
##  Median :0.02869   Median :0.5334   Median :1.123   Median :273.5
##  Mean   :0.03051   Mean   :0.5533   Mean   :1.165   Mean   :290.8
##  3rd Qu.:0.03900   3rd Qu.:0.5564   3rd Qu.:1.171   3rd Qu.:371.8
##  Max.   :0.05434   Max.   :0.7044   Max.   :1.483   Max.   :518.0
##
## mining info:
##   data ntransactions support confidence
##  bread          9532    0.01        0.5
```

Then we can take a look at the rules we have:

```
inspect(bakery_rules)
```

```
##      lhs                 rhs      support    confidence lift     count
## [1]  {Spanish Brunch} => {Coffee} 0.01080571 0.5988372  1.260626 103
## [2]  {Toast}          => {Coffee} 0.02349979 0.7044025  1.482854 224
## [3]  {Scone}          => {Coffee} 0.01793957 0.5229358  1.100844 171
## [4]  {Alfajores}      => {Coffee} 0.01951322 0.5406977  1.138235 186
## [5]  {Juice}          => {Coffee} 0.02045741 0.5342466  1.124655 195
## [6]  {Cookies}        => {Coffee} 0.02801091 0.5184466  1.091394 267
```

```
## [7]  {Medialuna}      => {Coffee} 0.03493496 0.5692308  1.198301 333
## [8]  {Hot chocolate}  => {Coffee} 0.02937474 0.5072464  1.067816 280
## [9]  {Sandwich}       => {Coffee} 0.03797734 0.5323529  1.120669 362
## [10] {Pastry}         => {Coffee} 0.04720940 0.5521472  1.162338 450
## [11] {NONE}           => {Coffee} 0.04206882 0.5325365  1.121055 401
## [12] {Cake}           => {Coffee} 0.05434326 0.5269583  1.109312 518
```

From the rules we have above, we can convey some key information:

- 59% of the customers who bought a spanish brunch also bought a coffee.
- 70% of the customers who bought a toast also bought a coffee.

Looks like most of the customers who come to the bakery will buy a cup of coffee!

4.2. Visualize the rules

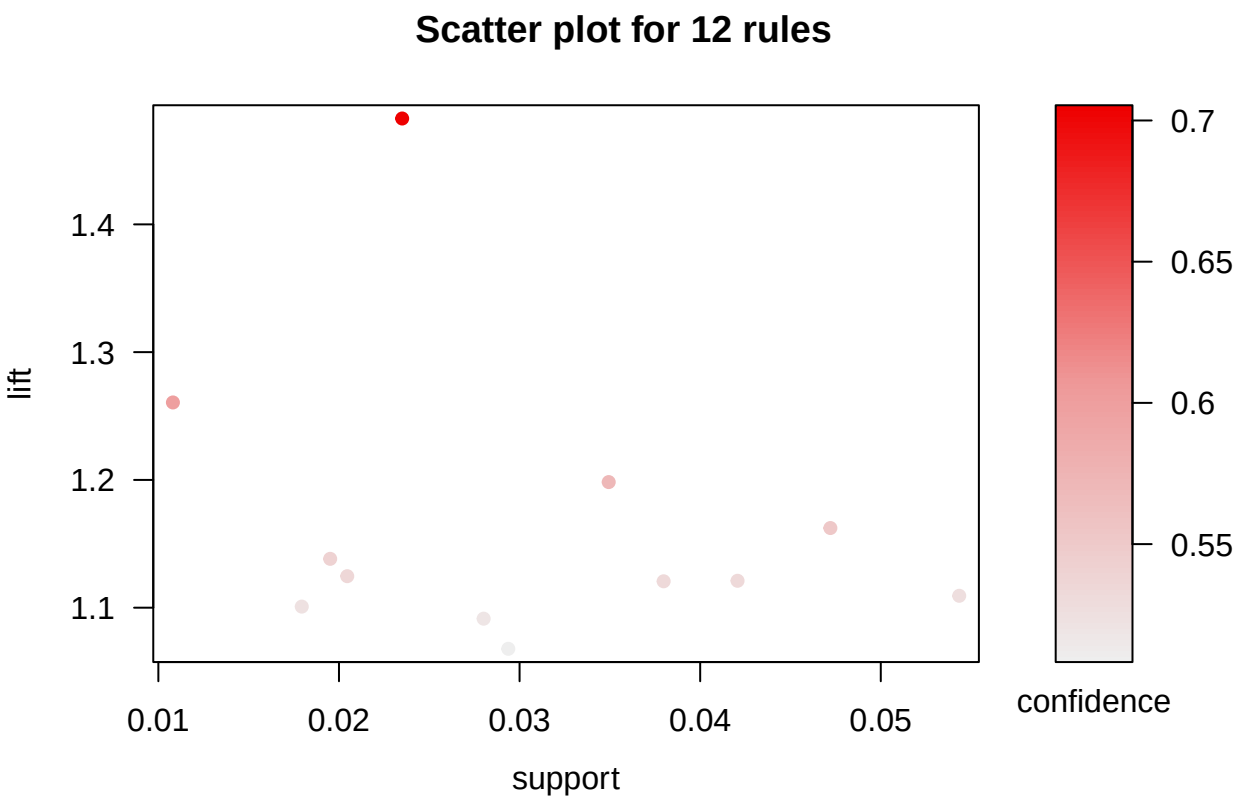
We can also **visualize** our rules.

```
library(arulesViz)
```

```
## Loading required package: grid
```

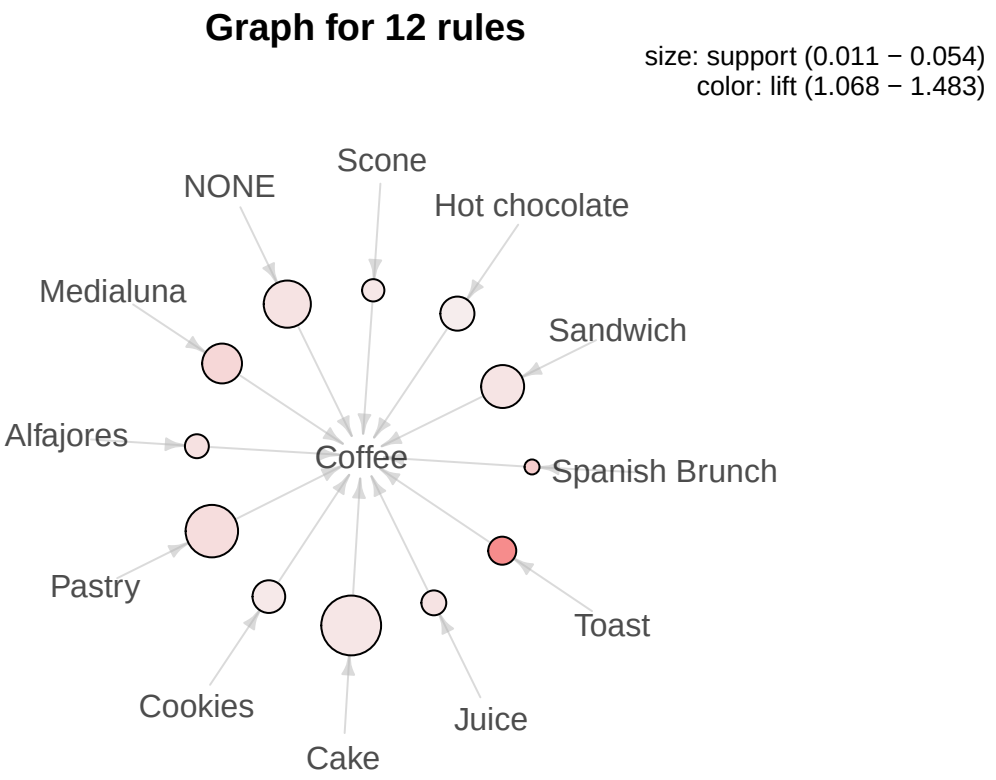
```
## Registered S3 method overwritten by 'seriation':
##   method         from
##   reorder.hclust gclus
```

```
plot(bakery_rules, measure = c("support", "lift"), shading = "confidence")
```



We can use another visualization to represent our rules.

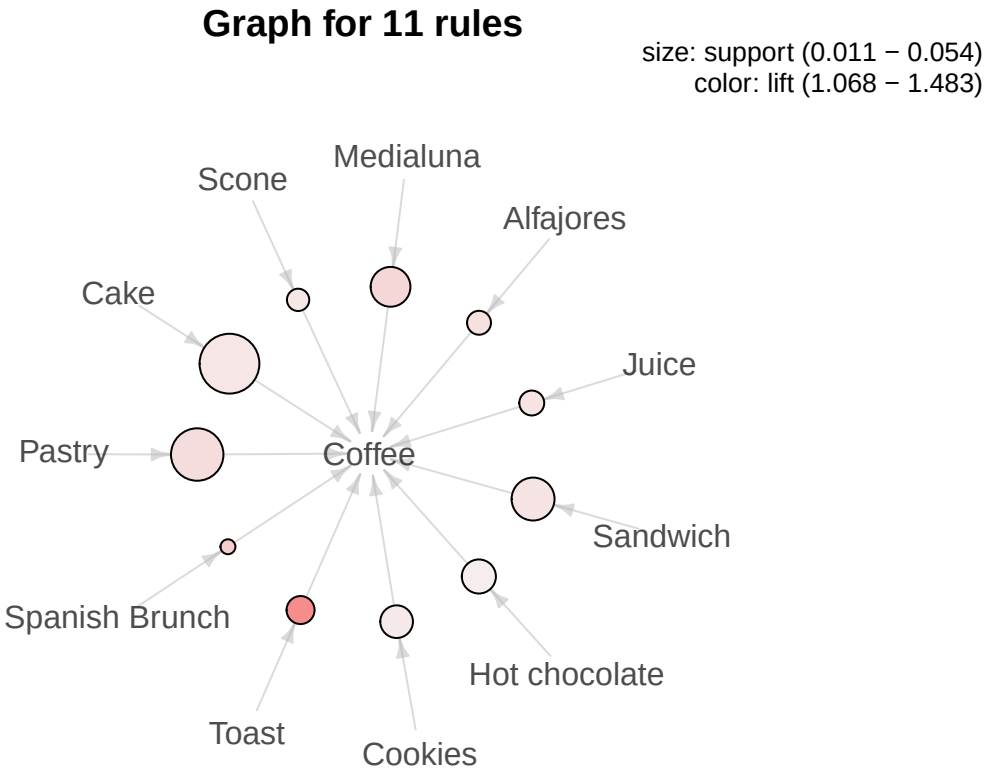
```
plot(bakery_rules, method = "graph")
```



5. "Improve" the Model's Performance

I saw the item "NONE" a lot of times through our analysis, so I want to eliminate it.

```
bakery_rules %>%
  subset(!lhs %in% "NONE") %>%
  plot(., method = "graph")
```



Now our rules look good!