AssociationRuleAnalysis

Zhichao Yuan(zxy180004)

April 16, 2019

Required Library

We need following libraries

```
require(readr)
require(arules)
require(dplyr)
```

Preprocessing data

First, We read data from file from original data file which is downloaded from kaggle. For the data file "order_products___train.csv", if we read the whole data, we cannot process it in the PC momery later, so we just use the first 20000 rows in the data.

Then, we process and extract useful informations into transaction_product_data and transaction_department_data.

Finally, we convert them to classes of transactions, product_as_transaction and department_as_transaction.

```
# preprocessing data

# read data from file
order_products_data<-read.csv("https://github.com/lypf2018/InstacartMarketBasketAnalysisKaggle/raw/maste
products_data<-read.csv("https://github.com/lypf2018/InstacartMarketBasketAnalysisKaggle/raw/master/data
department_data<-read.csv("https://github.com/lypf2018/InstacartMarketBasketAnalysisKaggle/raw/master/da

# getting order and product id from order_products_data
order_product_id_data<-order_products_data[,1:2]
# get product id, name and department id from products_data
product_department_data<-products_data[,c(1:2,4)]

# merging to transection information data and sorting
product_department_with_name_data<-merge(product_department_data, department_data, by="department_id")
transaction_data<-merge(order_product_id_data, product_department_with_name_data, by="product_id")
transaction_data<-arrange(transaction_data, order_id)

# extracting product analysis data needed
product_analysis_data<-transaction_data[,c(2,4)]
department_analysis_data<-transaction_data[,c(2,5)]

# Splitting the data
transaction_product_data <- split(product_analysis_data$product_name, product_analysis_data$order_id)
transaction_department_data <- split(department_analysis_data$department, department_analysis_data$orde

# Converting data to a class of transactions
product_as_transaction = as(transaction_product_data,"transactions")
department_as_transaction = as(transaction_department_data,"transactions")
```

Finding frequent itemsets and Association rules

1. Frequent itemsets for products in orders dataset.

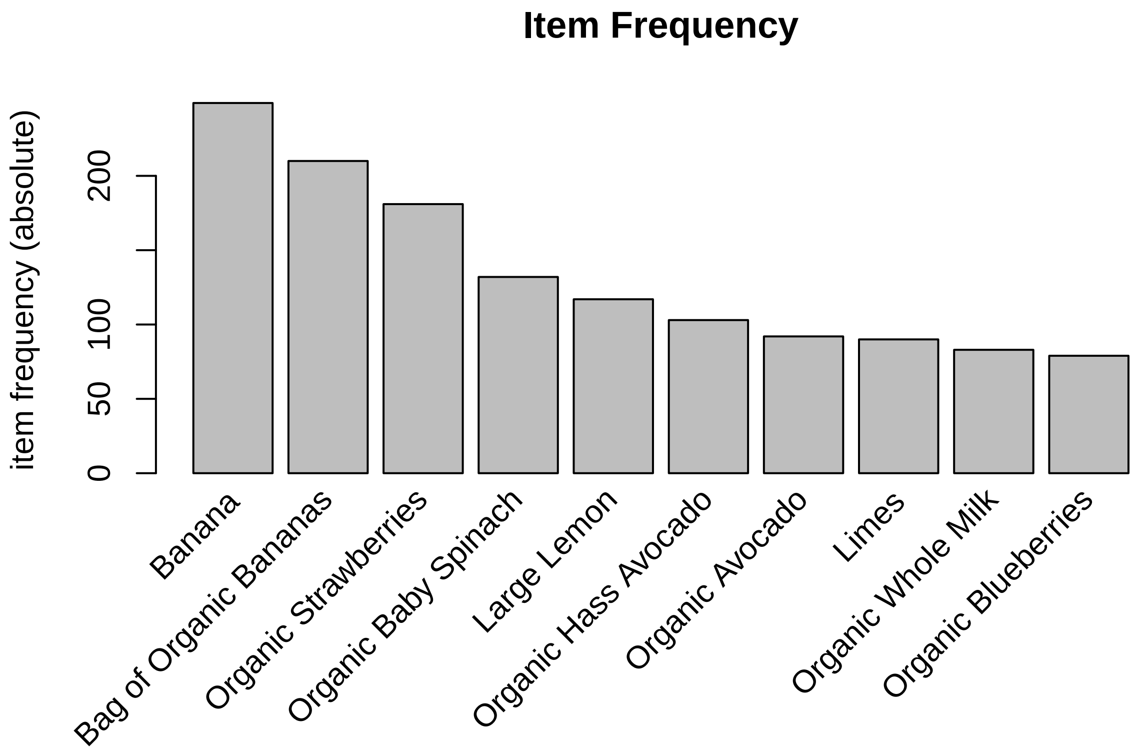
```
# 1. Frequent itemsets for products in orders dataset.
frequentItems <- eclat(product_as_transaction, parameter = list(supp = 0.07, maxlen = 15))
```

```
## Eclat
##
## parameter specification:
##  tidLists support minlen maxlen            target   ext
##     FALSE    0.07      1     15 frequent itemsets FALSE
##
## algorithmic control:
##  sparse sort verbose
##       7   -2    TRUE
##
## Absolute minimum support count: 132
##
## create itemset ...
## set transactions ...[7338 item(s), 1886 transaction(s)] done [0.02s].
## sorting and recoding items ... [3 item(s)] done [0.00s].
## creating sparse bit matrix ... [3 row(s), 1886 column(s)] done [0.00s].
## writing  ... [3 set(s)] done [0.00s].
## Creating S4 object  ... done [0.00s].
```

```
inspect(frequentItems)
```

```
##     items                    support    count
## [1] {Banana}                 0.13202545 249
## [2] {Bag of Organic Bananas} 0.11134677 210
## [3] {Organic Strawberries}   0.09597031 181
```

```
# plot frequent itemsets for products
itemFrequencyPlot(product_as_transaction, topN=10, type="absolute", main="Item Frequency")
```



2. Association rules for products in orders dataset.

```
# 2. Association rules for products in orders dataset.
rules <- apriori(product_as_transaction, parameter = list(supp = 0.001, conf = 0.5, maxlen=15))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      15  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 1
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[7338 item(s), 1886 transaction(s)] done [0.01s].
## sorting and recoding items ... [2910 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 done [0.04s].
## writing ... [24663 rule(s)] done [0.01s].
## creating S4 object  ... done [0.01s].
```

```
# 'high-confidence' rules.
rules_conf <- sort(rules, by="confidence", decreasing=TRUE)
# show the support, lift and confidence for all rules
inspect(head(rules_conf))
```

```
##     lhs                                     rhs                          support confidence      li
## [1] {April Fresh Liquid Fabric Softener} => {Hass Avocados}          0.001060445          1  44.9047
## [2] {Cheez-It Baked Snack Crackers}      => {Strawberry Twizzlers}   0.001060445          1 628.6666
## [3] {Hearts Of Palm}                     => {Golden Pineapple}       0.001060445          1 314.3333
## [4] {Hearts Of Palm}                     => {Organic Strawberries}   0.001060445          1  10.4198
## [5] {Oranges}                            => {Bag of Organic Bananas} 0.001060445          1   8.9809
## [6] {Pink Lady Apples}                   => {Organic Strawberries}   0.001060445          1  10.4198
```

3. Frequent itemsets for departments in orders dataset.

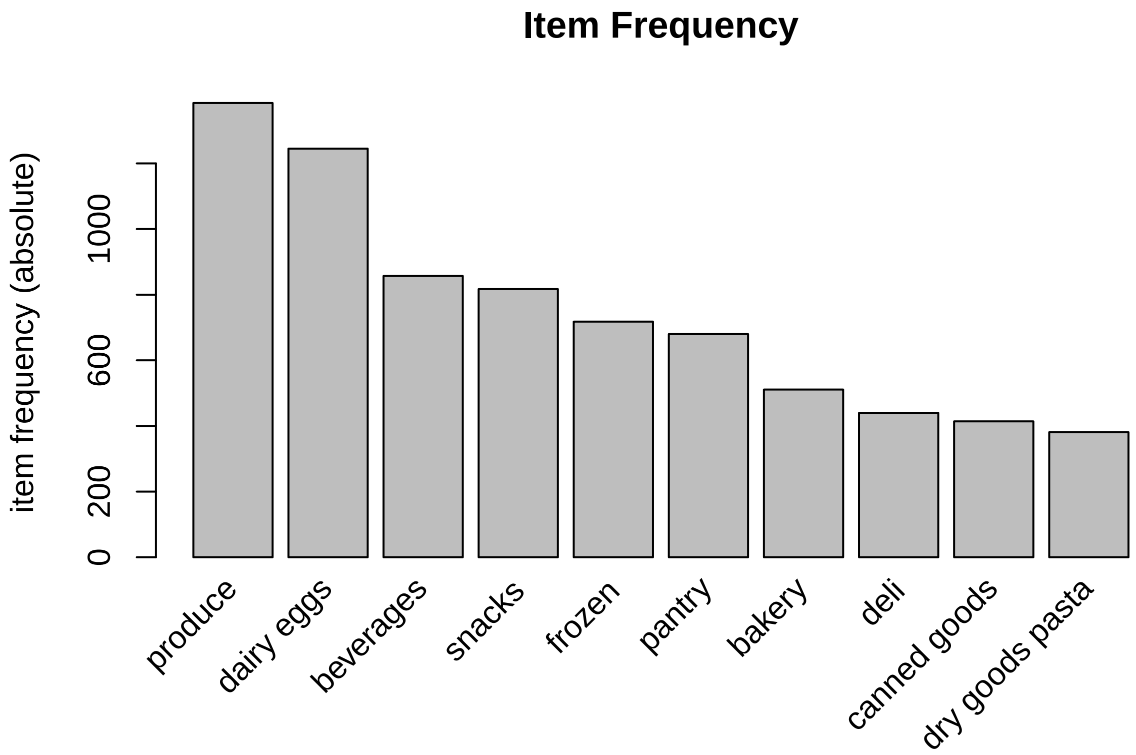
```
# 3. Frequent itemsets for departments in orders dataset.
frequentItems <- eclat(department_as_transaction, parameter = list(supp = 0.07, maxlen = 15))
```

```
## Eclat
##
## parameter specification:
##  tidLists support minlen maxlen            target   ext
##     FALSE    0.07      1     15 frequent itemsets FALSE
##
## algorithmic control:
##  sparse sort verbose
##       7   -2    TRUE
##
## Absolute minimum support count: 132
##
## create itemset ...
## set transactions ...[21 item(s), 1886 transaction(s)] done [0.00s].
## sorting and recoding items ... [14 item(s)] done [0.00s].
## creating bit matrix ... [14 row(s), 1886 column(s)] done [0.00s].
## writing  ... [220 set(s)] done [0.00s].
## Creating S4 object  ... done [0.00s].
```

```
inspect(frequentItems)
```

```
##       items                                          support    count
## [1]   {personal care,produce}                        0.07370095  139
## [2]   {dairy eggs,personal care}                     0.07423118  140
## [3]   {beverages,dairy eggs,household}               0.07476140  141
## [4]   {dairy eggs,household,produce}                 0.08536585  161
## [5]   {household,produce}                            0.10763521  203
## [6]   {dairy eggs,household}                         0.11558855  218
## [7]   {beverages,household}                          0.10233298  193
## [8]   {household,snacks}                             0.08536585  161
## [9]   {frozen,household}                             0.07529162  142
## [10]  {breakfast,produce,snacks}                     0.07317073  138
## [11]  {breakfast,dairy eggs,snacks}                  0.07900318  149
## [12]  {beverages,breakfast,dairy eggs}               0.07264051  137
## [13]  {breakfast,dairy eggs,produce}                 0.10657476  201
## [14]  {breakfast,produce}                            0.12354189  233
## [15]  {breakfast,dairy eggs}                         0.12884411  243
## [16]  {beverages,breakfast}                          0.08536585  161
## [17]  {breakfast,snacks}                             0.09013786  170
## [18]  {breakfast,frozen}                             0.07741251  146
## [19]  {bakery,meat seafood,produce}                  0.07476140  141
## [20]  {bakery,dairy eggs,meat seafood}               0.07051962  133
## [21]  {meat seafood,pantry,produce}                  0.08059385  152
## [22]  {dairy eggs,meat seafood,pantry}               0.07529162  142
## [23]  {dairy eggs,frozen,meat seafood,produce}       0.07688229  145
## [24]  {frozen,meat seafood,produce}                  0.08801697  166
## [25]  {dairy eggs,frozen,meat seafood}               0.08112407  153
## [26]  {meat seafood,produce,snacks}                  0.07847296  148
## [27]  {dairy eggs,meat seafood,snacks}               0.07104984  134
## [28]  {beverages,meat seafood,produce}               0.08271474  156
## [29]  {beverages,dairy eggs,meat seafood}            0.07476140  141
## [30]  {dairy eggs,meat seafood,produce}              0.13361612  252
## [31]  {meat seafood,produce}                         0.16436903  310
## [32]  {dairy eggs,meat seafood}                      0.14475080  273
## [33]  {beverages,meat seafood}                       0.09013786  170
## [34]  {meat seafood,snacks}                          0.08589608  162
## [35]  {frozen,meat seafood}                          0.09437964  178
## [36]  {meat seafood,pantry}                          0.08748674  165
## [37]  {bakery,meat seafood}                          0.08006363  151
## [38]  {canned goods,dry goods pasta,produce}         0.08059385  152
## [39]  {canned goods,dairy eggs,dry goods pasta}      0.07741251  146
## [40]  {bakery,dry goods pasta,produce}               0.07529162  142
## [41]  {bakery,dairy eggs,dry goods pasta}            0.07158006  135
## [42]  {dairy eggs,dry goods pasta,pantry,produce}    0.08801697  166
## [43]  {dry goods pasta,pantry,produce}               0.09915164  187
## [44]  {dairy eggs,dry goods pasta,pantry}            0.09331919  176
## [45]  {dairy eggs,dry goods pasta,frozen,produce}    0.08960764  169
## [46]  {dry goods pasta,frozen,produce}               0.10074231  190
## [47]  {dairy eggs,dry goods pasta,frozen}            0.09862142  186
## [48]  {dairy eggs,dry goods pasta,produce,snacks}    0.07953340  150
## [49]  {dry goods pasta,produce,snacks}               0.09066808  171
## [50]  {dairy eggs,dry goods pasta,snacks}            0.08854719  167
## [51]  {beverages,dairy eggs,dry goods pasta,produce} 0.07529162  142
## [52]  {beverages,dry goods pasta,produce}            0.08907741  168
## [53]  {beverages,dairy eggs,dry goods pasta}         0.08483563  160
## [54]  {dairy eggs,dry goods pasta,produce}           0.14581124  275
## [55]  {dry goods pasta,produce}                      0.17497349  330
## [56]  {dairy eggs,dry goods pasta}                   0.16383881  309
## [57]  {beverages,dry goods pasta}                    0.10286320  194
## [58]  {dry goods pasta,snacks}                       0.10286320  194
## [59]  {dry goods pasta,frozen}                       0.11346766  214
## [60]  {dry goods pasta,pantry}                       0.10763521  203
## [61]  {bakery,dry goods pasta}                       0.08218452  155
## [62]  {deli,dry goods pasta}                         0.07423118  140
## [63]  {canned goods,dry goods pasta}                 0.08854719  167
## [64]  {bakery,canned goods,dairy eggs,produce}       0.07635207  144
## [65]  {bakery,canned goods,produce}                  0.08271474  156
## [66]  {bakery,canned goods,dairy eggs}               0.07900318  149
## [67]  {canned goods,dairy eggs,pantry,produce}       0.09756098  184
## [68]  {canned goods,pantry,produce}                  0.11346766  214
## [69]  {canned goods,dairy eggs,pantry}               0.10657476  201
## [70]  {canned goods,pantry,snacks}                   0.07264051  137
## [71]  {canned goods,dairy eggs,frozen,produce}       0.08907741  168
## [72]  {canned goods,frozen,produce}                  0.09756098  184
## [73]  {canned goods,dairy eggs,frozen}               0.09915164  187
## [74]  {canned goods,dairy eggs,produce,snacks}       0.08801697  166
## [75]  {canned goods,produce,snacks}                  0.10074231  190
## [76]  {canned goods,dairy eggs,snacks}               0.09597031  181
## [77]  {beverages,canned goods,dairy eggs,produce}    0.08377519  158
## [78]  {beverages,canned goods,produce}               0.09703075  183
## [79]  {beverages,canned goods,dairy eggs}            0.09437964  178
## [80]  {canned goods,dairy eggs,produce}              0.15800636  298
## [81]  {canned goods,produce}                         0.19353128  365
## [82]  {canned goods,dairy eggs}                      0.17656416  333
## [83]  {beverages,canned goods}                       0.11081654  209
## [84]  {canned goods,snacks}                          0.11293743  213
## [85]  {canned goods,frozen}                          0.11081654  209
## [86]  {canned goods,pantry}                          0.12460233  235
## [87]  {bakery,canned goods}                          0.08589608  162
## [88]  {canned goods,deli}                            0.07211029  136
## [89]  {bakery,dairy eggs,deli,produce}               0.07900318  149
## [90]  {bakery,deli,produce}                          0.08589608  162
## [91]  {bakery,dairy eggs,deli}                       0.08642630  163
## [92]  {dairy eggs,deli,pantry,produce}               0.08642630  163
## [93]  {deli,pantry,produce}                          0.09437964  178
## [94]  {dairy eggs,deli,pantry}                       0.09437964  178
## [95]  {deli,frozen,produce,snacks}                   0.07264051  137
## [96]  {dairy eggs,deli,frozen,snacks}                0.07158006  135
## [97]  {dairy eggs,deli,frozen,produce}               0.09597031  181
## [98]  {deli,frozen,produce}                          0.10657476  201
## [99]  {dairy eggs,deli,frozen}                       0.10392365  196
## [100] {beverages,deli,frozen}                        0.07264051  137
## [101] {deli,frozen,snacks}                           0.08218452  155
## [102] {beverages,deli,produce,snacks}                0.07370095  139
## [103] {beverages,dairy eggs,deli,snacks}             0.07423118  140
## [104] {dairy eggs,deli,produce,snacks}               0.10180276  192
## [105] {deli,produce,snacks}                          0.11399788  215
## [106] {dairy eggs,deli,snacks}                       0.11505832  217
## [107] {beverages,deli,snacks}                        0.08218452  155
## [108] {beverages,dairy eggs,deli,produce}            0.09597031  181
## [109] {beverages,deli,produce}                       0.10922587  206
## [110] {beverages,dairy eggs,deli}                    0.10763521  203
## [111] {dairy eggs,deli,produce}                      0.16755037  316
## [112] {deli,produce}                                 0.19671262  371
## [113] {dairy eggs,deli}                              0.19194062  362
## [114] {beverages,deli}                               0.12248144  231
## [115] {deli,snacks}                                  0.13520679  255
## [116] {deli,frozen}                                  0.11876988  224
## [117] {deli,pantry}                                  0.10710498  202
## [118] {bakery,deli}                                  0.09756098  184
## [119] {bakery,pantry,produce,snacks}                 0.07317073  138
## [120] {bakery,dairy eggs,pantry,snacks}              0.07476140  141
## [121] {bakery,dairy eggs,pantry,produce}             0.10127253  191
## [122] {bakery,pantry,produce}                        0.11134677  210
## [123] {bakery,dairy eggs,pantry}                      0.10975610  207
## [124] {bakery,beverages,pantry}                       0.07317073  138
## [125] {bakery,pantry,snacks}                          0.08059385  152
## [126] {bakery,dairy eggs,frozen,produce,snacks}       0.07423118  140
## [127] {bakery,frozen,produce,snacks}                  0.08006363  151
## [128] {bakery,dairy eggs,frozen,snacks}               0.07900318  149
## [129] {bakery,beverages,dairy eggs,frozen,produce}    0.07264051  137
## [130] {bakery,beverages,frozen,produce}               0.07794274  147
## [131] {bakery,beverages,dairy eggs,frozen}            0.07688229  145
## [132] {bakery,dairy eggs,frozen,produce}              0.11346766  214
## [133] {bakery,frozen,produce}                         0.12354189  233
## [134] {bakery,dairy eggs,frozen}                      0.12089077  228
## [135] {bakery,beverages,frozen}                       0.08483563  160
## [136] {bakery,frozen,snacks}                          0.09013786  170
## [137] {bakery,beverages,dairy eggs,produce,snacks}    0.07529162  142
## [138] {bakery,beverages,produce,snacks}               0.08377519  158
## [139] {bakery,beverages,dairy eggs,snacks}            0.08271474  156
## [140] {bakery,dairy eggs,produce,snacks}              0.11983033  226
## [141] {bakery,produce,snacks}                         0.13626723  257
## [142] {bakery,dairy eggs,snacks}                      0.13096501  247
## [143] {bakery,beverages,snacks}                       0.09490986  179
## [144] {bakery,beverages,dairy eggs,produce}           0.11399788  215
## [145] {bakery,beverages,produce}                      0.12990456  245
## [146] {bakery,beverages,dairy eggs}                   0.12566278  237
## [147] {bakery,dairy eggs,produce}                     0.19883351  375
## [148] {bakery,produce}                                0.23276776  439
## [149] {bakery,dairy eggs}                             0.22110286  417
## [150] {bakery,beverages}                              0.14793213  279
## [151] {bakery,snacks}                                 0.15694592  296
## [152] {bakery,frozen}                                 0.13732768  259
## [153] {bakery,pantry}                                 0.12301166  232
## [154] {dairy eggs,frozen,pantry,produce,snacks}       0.08059385  152
## [155] {frozen,pantry,produce,snacks}                  0.08695652  164
## [156] {dairy eggs,frozen,pantry,snacks}               0.08854719  167
## [157] {beverages,dairy eggs,frozen,pantry,produce}    0.07741251  146
## [158] {beverages,frozen,pantry,produce}               0.08218452  155
## [159] {beverages,dairy eggs,frozen,pantry}            0.08430541  159
## [160] {dairy eggs,frozen,pantry,produce}              0.12990456  245
## [161] {frozen,pantry,produce}                         0.14369035  271
## [162] {dairy eggs,frozen,pantry}                      0.14422057  272
## [163] {beverages,frozen,pantry}                       0.09119830  172
## [164] {frozen,pantry,snacks}                          0.09915164  187
## [165] {beverages,dairy eggs,pantry,produce,snacks}    0.08642630  163
## [166] {beverages,pantry,produce,snacks}               0.09490986  179
## [167] {beverages,dairy eggs,pantry,snacks}            0.09809120  185
## [168] {dairy eggs,pantry,produce,snacks}              0.13467656  254
## [169] {pantry,produce,snacks}                         0.15535525  293
## [170] {dairy eggs,pantry,snacks}                      0.15376458  290
## [171] {beverages,pantry,snacks}                       0.11081654  209
## [172] {beverages,dairy eggs,pantry,produce}           0.13043478  246
## [173] {beverages,pantry,produce}                      0.14793213  279
## [174] {beverages,dairy eggs,pantry}                   0.15111347  285
## [175] {dairy eggs,pantry,produce}                     0.24125133  455
## [176] {pantry,produce}                                0.29321315  553
## [177] {dairy eggs,pantry}                            0.28048780  529
## [178] {beverages,pantry}                             0.17921527  338
## [179] {pantry,snacks}                                0.18557794  350
## [180] {frozen,pantry}                                0.16648993  314
## [181] {beverages,dairy eggs,frozen,produce,snacks}   0.09384942  177
## [182] {beverages,frozen,produce,snacks}              0.10604454  200
## [183] {beverages,dairy eggs,frozen,snacks}           0.10763521  203
## [184] {dairy eggs,frozen,produce,snacks}             0.14581124  275
## [185] {frozen,produce,snacks}                        0.16648993  314
## [186] {dairy eggs,frozen,snacks}                     0.16755037  316
## [187] {beverages,frozen,snacks}                      0.12725345  240
## [188] {beverages,dairy eggs,frozen,produce}          0.14528102  274
## [189] {beverages,frozen,produce}                     0.16648993  314
## [190] {beverages,dairy eggs,frozen}                  0.16595970  313
## [191] {dairy eggs,frozen,produce}                    0.26139979  493
## [192] {frozen,produce}                               0.31071050  586
## [193] {dairy eggs,frozen}                            0.30169671  569
## [194] {beverages,frozen}                             0.19830329  374
## [195] {frozen,snacks}                                0.20307529  383
## [196] {beverages,dairy eggs,produce,snacks}          0.15641569  295
## [197] {beverages,produce,snacks}                     0.18716861  353
## [198] {beverages,dairy eggs,snacks}                  0.18716861  353
## [199] {dairy eggs,produce,snacks}                    0.26193001  494
## [200] {produce,snacks}                               0.32820785  619
## [201] {dairy eggs,snacks}                            0.31389183  592
## [202] {beverages,snacks}                             0.24231177  457
## [203] {beverages,dairy eggs,produce}                 0.26086957  492
## [204] {beverages,produce}                            0.33457052  631
## [205] {beverages,dairy eggs}                         0.31601273  596
## [206] {dairy eggs,produce}                           0.53499470 1009
## [207] {produce}                                      0.73382821 1384
## [208] {dairy eggs}                                   0.66012725 1245
## [209] {beverages}                                    0.45440085  857
## [210] {snacks}                                       0.43319194  817
## [211] {frozen}                                       0.38069989  718
## [212] {pantry}                                       0.36055143  680
## [213] {bakery}                                       0.27094380  511
## [214] {deli}                                         0.23329799  440
## [215] {canned goods}                                 0.21951220  414
## [216] {dry goods pasta}                              0.20201485  381
## [217] {meat seafood}                                 0.18080594  341
## [218] {breakfast}                                    0.15376458  290
## [219] {household}                                    0.17285260  326
## [220] {personal care}                                0.10975610  207
```

```
# plot frequent itemsets for products
itemFrequencyPlot(department_as_transaction, topN=10, type="absolute", main="Item Frequency")
```



4. Association rules for departments in orders dataset.

```
# 4. Association rules for departments in orders dataset.
rules <- apriori(department_as_transaction, parameter = list(supp = 0.001, conf = 0.5, maxlen=15))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target   ext
##      15  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 1
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[21 item(s), 1886 transaction(s)] done [0.00s].
## sorting and recoding items ... [21 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 10 11 12 done [0.01s].
## writing ... [189723 rule(s)] done [0.03s].
## creating S4 object  ... done [0.08s].
```

```
# 'high-confidence' rules.
rules_conf <- sort(rules, by="confidence", decreasing=TRUE)
# show the support, lift and confidence for all rules
inspect(head(rules_conf))
```

```
##     lhs                       rhs               support     confidence
## [1] {babies,bulk}          => {pantry}          0.001060445 1
## [2] {babies,bulk}          => {snacks}          0.001060445 1
## [3] {babies,bulk}          => {beverages}       0.001060445 1
## [4] {babies,bulk}          => {dairy eggs}      0.001060445 1
## [5] {babies,bulk}          => {produce}         0.001060445 1
## [6] {bulk,international} => {dry goods pasta} 0.001060445 1
##     lift     count
## [1] 2.773529 2
## [2] 2.308446 2
## [3] 2.200700 2
## [4] 1.514859 2
## [5] 1.362717 2
## [6] 4.950131 2
```