Lab 09

Riley Payung

4/6/2020

```
#Change the support to 0.1

frequentItems <- eclat (Groceries, parameter = list(supp = 0.07, maxlen = 15)) # calculates supp
ort for frequent items
```

```
## 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: 688
##
## create itemset ...
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [18 item(s)] done [0.00s].
## creating sparse bit matrix ... [18 row(s), 9835 column(s)] done [0.00s].
## writing  ... [19 set(s)] done [0.00s].
## Creating S4 object  ... done [0.00s].
```

```
inspect(frequentItems)
```

```
##      items                          support    count
## [1]  {other vegetables,whole milk} 0.07483477  736
## [2]  {whole milk}                  0.25551601 2513
## [3]  {other vegetables}            0.19349263 1903
## [4]  {rolls/buns}                  0.18393493 1809
## [5]  {yogurt}                      0.13950178 1372
## [6]  {soda}                        0.17437722 1715
## [7]  {root vegetables}             0.10899847 1072
## [8]  {tropical fruit}              0.10493137 1032
## [9]  {bottled water}               0.11052364 1087
## [10] {sausage}                     0.09395018  924
## [11] {shopping bags}               0.09852567  969
## [12] {citrus fruit}                0.08276563  814
## [13] {pastry}                      0.08896797  875
## [14] {pip fruit}                   0.07564820  744
## [15] {whipped/sour cream}          0.07168277  705
## [16] {fruit/vegetable juice}       0.07229283  711
## [17] {newspapers}                  0.07981698  785
## [18] {bottled beer}                0.08052872  792
## [19] {canned beer}                 0.07768175  764
```

```
frequentItems <- eclat (Groceries, parameter = list(supp = 0.1, maxlen = 15)) # calculates suppo
rt for frequent items
```

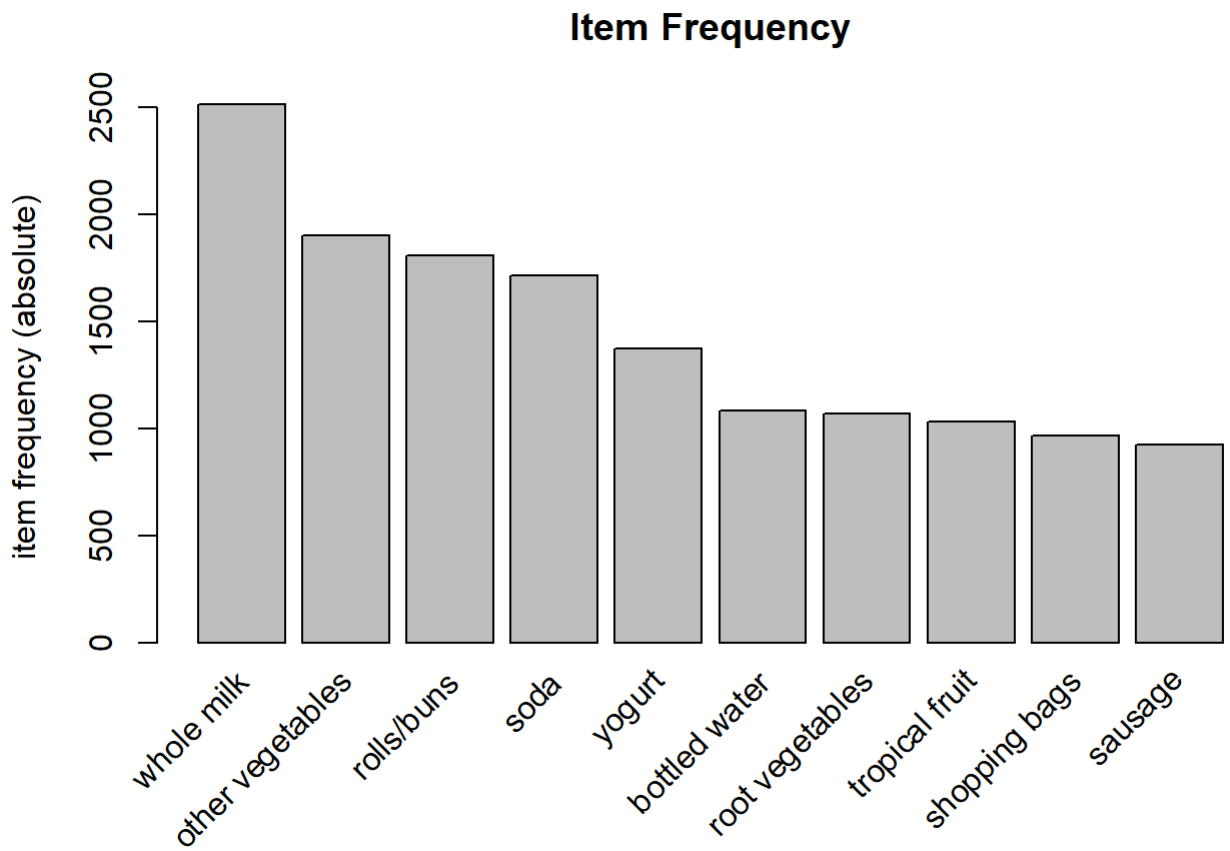
```
## Eclat
##
## parameter specification:
##  tidLists support minlen maxlen            target   ext
##     FALSE     0.1      1     15 frequent itemsets FALSE
##
## algorithmic control:
##  sparse sort verbose
##       7   -2    TRUE
##
## Absolute minimum support count: 983
##
## create itemset ...
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [8 item(s)] done [0.00s].
## creating bit matrix ... [8 row(s), 9835 column(s)] done [0.00s].
## writing  ... [8 set(s)] done [0.00s].
## Creating S4 object  ... done [0.00s].
```

```
inspect(frequentItems)
```

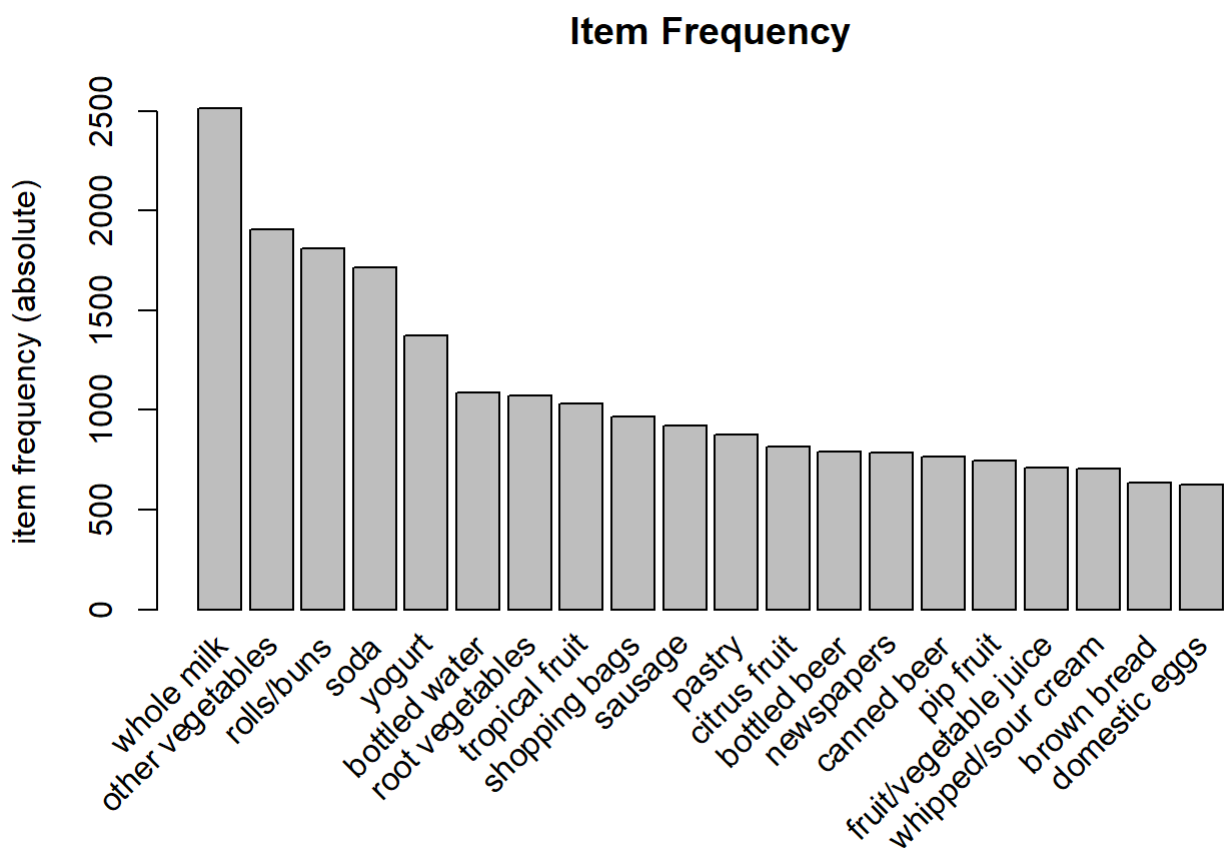
```
##     items              support   count
## [1] {whole milk}       0.2555160 2513
## [2] {other vegetables} 0.1934926 1903
## [3] {rolls/buns}       0.1839349 1809
## [4] {yogurt}           0.1395018 1372
## [5] {soda}             0.1743772 1715
## [6] {root vegetables}  0.1089985 1072
## [7] {tropical fruit}   0.1049314 1032
## [8] {bottled water}    0.1105236 1087
```

The data completely changes. Originally it contained a compound transaction {other vegetables,whole milk} which is now gone since the support was not greater than 0.1. In fact the support was 0.0748, hence why we were able to see the itemset before. {sausage} itemset is also gone since its support was also not greater than or equal to 0.1. In fact, its support was 0.0939, just below the threshold. The data also completely drops everything after the 9th itemset. The data is ordered based on support, where compound itemsets are displayed first.

```
#Change to TopN=20
itemFrequencyPlot(Groceries, topN=10, type="absolute", main="Item Frequency")
```




```
#Change to TopN=20
itemFrequencyPlot(Groceries, topN=20, type="absolute", main="Item Frequency")
```



Changing the topN to 20 just increases the amount of displayed data. Nothing new here.

```
rules <- apriori (Groceries, parameter = list(supp = 0.001, conf = 0.5)) # Min Support as 0.001, confidence as 0.8.
```

```
## 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
##      10  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.01s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.02s].
## writing ... [5668 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

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

```
##     lhs                     rhs                    support confidence     lift count
## [1] {rice,
##      sugar}              => {whole milk}       0.001220132          1 3.913649    12
## [2] {canned fish,
##      hygiene articles}   => {whole milk}       0.001118454          1 3.913649    11
## [3] {root vegetables,
##      butter,
##      rice}               => {whole milk}       0.001016777          1 3.913649    10
## [4] {root vegetables,
##      whipped/sour cream,
##      flour}              => {whole milk}       0.001728521          1 3.913649    17
## [5] {butter,
##      soft cheese,
##      domestic eggs}      => {whole milk}       0.001016777          1 3.913649    10
## [6] {citrus fruit,
##      root vegetables,
##      soft cheese}        => {other vegetables} 0.001016777          1 5.168156    10
```

```
rules <- apriori (Groceries, parameter = list(supp = 0.1, conf = 0.5)) # Min Support as 0.001, c
onfidence as 0.8.
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.1      1
##  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: 983
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.01s].
## sorting and recoding items ... [8 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

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

rules_conf just ends up empty, however, if we change support to 0.01, the list completely changes. The itemsets are completely different.

```
rules <- apriori (Groceries, parameter = list(supp = 0.01, conf = 0.5)) # Min Support as 0.001,
 confidence as 0.8.
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5    0.01      1
##  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: 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 [0.01s].
## writing ... [15 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

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

```
##     lhs                                rhs                support
## [1] {citrus fruit,root vegetables}   => {other vegetables} 0.01037112
## [2] {tropical fruit,root vegetables} => {other vegetables} 0.01230300
## [3] {curd,yogurt}                    => {whole milk}       0.01006609
## [4] {other vegetables,butter}        => {whole milk}       0.01148958
## [5] {tropical fruit,root vegetables} => {whole milk}       0.01199797
## [6] {root vegetables,yogurt}         => {whole milk}       0.01453991
##     confidence lift     count
## [1] 0.5862069  3.029608 102
## [2] 0.5845411  3.020999 121
## [3] 0.5823529  2.279125  99
## [4] 0.5736041  2.244885 113
## [5] 0.5700483  2.230969 118
## [6] 0.5629921  2.203354 143
```

```
#Give me the high lift rules

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

```
##     lhs                                rhs                support
## [1] {citrus fruit,root vegetables}   => {other vegetables} 0.01037112
## [2] {tropical fruit,root vegetables} => {other vegetables} 0.01230300
## [3] {root vegetables,rolls/buns}     => {other vegetables} 0.01220132
## [4] {root vegetables,yogurt}         => {other vegetables} 0.01291307
## [5] {curd,yogurt}                    => {whole milk}       0.01006609
## [6] {other vegetables,butter}        => {whole milk}       0.01148958
##     confidence lift     count
## [1] 0.5862069  3.029608 102
## [2] 0.5845411  3.020999 121
## [3] 0.5020921  2.594890 120
## [4] 0.5000000  2.584078 127
## [5] 0.5823529  2.279125  99
## [6] 0.5736041  2.244885 113
```

```
#Change support to .1
#Tells us who bought whole milk in addition to groceries

rules <- apriori (data=Groceries, parameter=list (supp=0.001,conf = 0.08), appearance = list (de
fault="lhs",rhs="whole milk"), control = list (verbose=F)) # get rules that lead to buying 'whol
e milk'
rules_conf <- sort (rules, by="confidence", decreasing=TRUE) # 'high-confidence' rules.
inspect(head(rules_conf))
```

```
##     lhs                    rhs              support confidence     lift count
## [1] {rice,
##      sugar}              => {whole milk} 0.001220132          1 3.913649    12
## [2] {canned fish,
##      hygiene articles}   => {whole milk} 0.001118454          1 3.913649    11
## [3] {root vegetables,
##      butter,
##      rice}               => {whole milk} 0.001016777          1 3.913649    10
## [4] {root vegetables,
##      whipped/sour cream,
##      flour}              => {whole milk} 0.001728521          1 3.913649    17
## [5] {butter,
##      soft cheese,
##      domestic eggs}      => {whole milk} 0.001016777          1 3.913649    10
## [6] {pip fruit,
##      butter,
##      hygiene articles}   => {whole milk} 0.001016777          1 3.913649    10
```

```
#Change support to .1
#Tells us who bought groceries in addition to whole milk


rules <- apriori (data=Groceries, parameter=list (supp=0.01,conf = 0.15,minlen=2), appearance =
 list(default="rhs",lhs="whole milk"), control = list (verbose=F)) # those who bought 'milk' als
o bought..
rules_conf <- sort (rules, by="confidence", decreasing=TRUE) # 'high-confidence' rules.
inspect(head(rules_conf))
```

```
##     lhs             rhs                support    confidence lift      count
## [1] {whole milk} => {other vegetables} 0.07483477 0.2928770  1.5136341 736
## [2] {whole milk} => {rolls/buns}       0.05663447 0.2216474  1.2050318 557
## [3] {whole milk} => {yogurt}           0.05602440 0.2192598  1.5717351 551
## [4] {whole milk} => {root vegetables}  0.04890696 0.1914047  1.7560310 481
## [5] {whole milk} => {tropical fruit}   0.04229792 0.1655392  1.5775950 416
## [6] {whole milk} => {soda}             0.04006101 0.1567847  0.8991124 394
```

Answer Section

1. What are the top 20 items being purchased

The top 20 items that are being purchased are whole milk, other vegetables, rolls/buns, soda, yogurt, bottled water, root vegetables, tropical fruit, shopping bags, sausage, pastry, citrus fruit, bottled beer, newspapers, canned beer, pip fruit, fruit/vegetable juice, whipped/sour cream, brown bread, and domestic eggs.

2. Give me the the top 10 rules given a 10% min support and .5 confidence

- [1] {whole milk} => {other vegetables} 0.07483477 0.2928770 1.5136341 736
- [2] {whole milk} => {rolls/buns} 0.05663447 0.2216474 1.2050318 557
- [3] {whole milk} => {yogurt} 0.05602440 0.2192598 1.5717351 551
- [4] {whole milk} => {root vegetables} 0.04890696 0.1914047 1.7560310 481
- [5] {whole milk} => {tropical fruit} 0.04229792 0.1655392 1.5775950 416
- [6] {whole milk} => {soda} 0.04006101 0.1567847 0.8991124 394

3. Tell us what people bought whole milk in addition to groceries

People who were buying vegetables, tropical fruits, soda, yogurt, rolls/buns, just generally when they go grocery shopping, they are buying whole milk in addition to the other items on their lists.

4. Tells us what people groceries in addition to whole milk.

Rolls/Buns, Yougurt, Soda, tropical fruits, bottled water, root vegetables. Most buy other vegetables along with their milk.

5. What story does this tell you about purchasing whole milk and groceries as a whole.

People who purchase other groceries at the grocery store also tend to buy whole milk. Generally, people tend to buy whole milk in most of the transactions this store handles.