# Practice with Association Rule Mining

*Brooks Beckelman, Zack Bilderback, Davis Townsend*

Since the data was in a text file rather than a csv file, we had to handle the process of turning each row into a transaction differently from the example in class. There is a built in function called read.transactions that reads in text files line by line and converts each line to a single transaction which is exactly what we needed for the is quesiton.

The support parameter was set at .006 to create 464 rules rather than having the parameter set to something higher such as .05 which only made 5 rules. The confidence parameter was also set slightly low at .25 to allow for more rules. We can then subset rules based off each parameter if we want to be more strict on which rules we view.

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen maxlen
##        0.25    0.1    1 none FALSE            TRUE   0.006      1     10
##  target   ext
##   rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 59
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [109 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [464 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

First ten rules. Interestingly we see that someone with an empty basket is most likely to buy only whole milk however, we have a confidence level of .25

```
##     lhs                   rhs                 support      confidence
## 1  {}                  => {whole milk}        0.255516014 0.2555160
## 2  {pot plants}        => {whole milk}        0.006914082 0.4000000
## 3  {pasta}             => {whole milk}        0.006100661 0.4054054
## 4  {herbs}             => {root vegetables}   0.007015760 0.4312500
## 5  {herbs}             => {other vegetables}  0.007727504 0.4750000
## 6  {herbs}             => {whole milk}        0.007727504 0.4750000
## 7  {processed cheese}  => {whole milk}        0.007015760 0.4233129
## 8  {semi-finished bread} => {whole milk}      0.007117438 0.4022989
## 9  {beverages}         => {whole milk}        0.006812405 0.2617188
## 10 {detergent}         => {other vegetables}  0.006405694 0.3333333
##     lift
## 1  1.000000
## 2  1.565460
```

```
## 3   1.586614
## 4   3.956477
## 5   2.454874
## 6   1.858983
## 7   1.656698
## 8   1.574457
## 9   1.024275
## 10  1.722719
```

We can subset off rules as well if we want a higher support or confidence level.

```r
inspect(subset(basketrules, subset=support > .01 & confidence > 0.5)[1:10])
```

```
##     lhs                    rhs                     support confidence      lift
## 1   {curd,
##      yogurt}            => {whole milk}         0.01006609  0.5823529 2.279125
## 2   {butter,
##      other vegetables}  => {whole milk}         0.01148958  0.5736041 2.244885
## 3   {domestic eggs,
##      other vegetables}  => {whole milk}         0.01230300  0.5525114 2.162336
## 4   {whipped/sour cream,
##      yogurt}            => {whole milk}         0.01087951  0.5245098 2.052747
## 5   {other vegetables,
##      whipped/sour cream} => {whole milk}        0.01464159  0.5070423 1.984385
## 6   {other vegetables,
##      pip fruit}         => {whole milk}         0.01352313  0.5175097 2.025351
## 7   {citrus fruit,
##      root vegetables}   => {other vegetables}   0.01037112  0.5862069 3.029608
## 8   {root vegetables,
##      tropical fruit}    => {other vegetables}   0.01230300  0.5845411 3.020999
## 9   {root vegetables,
##      tropical fruit}    => {whole milk}         0.01199797  0.5700483 2.230969
## 10 {tropical fruit,
##      yogurt}            => {whole milk}         0.01514997  0.5173611 2.024770
```

Also, we can see which rules have the highest values in "lift", "confidence", or "support".

```r
##Sorted by lift
inspect(sort(basketrules, by = "lift")[1:5])
```

```
##   lhs                 rhs                     support confidence      lift
## 1 {herbs}          => {root vegetables}    0.007015760  0.4312500 3.956477
## 2 {berries}        => {whipped/sour cream} 0.009049314  0.2721713 3.796886
## 3 {other vegetables,
##    tropical fruit,
##    whole milk}     => {root vegetables}    0.007015760  0.4107143 3.768074
## 4 {beef,
##    other vegetables} => {root vegetables}  0.007930859  0.4020619 3.688692
## 5 {other vegetables,
##    tropical fruit}  => {pip fruit}         0.009456024  0.2634561 3.482649
```

```r
##Sorted by confidence
inspect(sort(basketrules, by = "confidence")[1:5])
```

```
##     lhs                           rhs            support     confidence
## 222 {butter,whipped/sour cream} => {whole milk} 0.006710727 0.6600000
## 230 {butter,yogurt}             => {whole milk} 0.009354347 0.6388889
## 226 {butter,root vegetables}    => {whole milk} 0.008235892 0.6377953
## 184 {curd,tropical fruit}       => {whole milk} 0.006507372 0.6336634
## 223 {butter,tropical fruit}     => {whole milk} 0.006202339 0.6224490
##     lift
## 222 2.583008
## 230 2.500387
## 226 2.496107
## 184 2.479936
## 223 2.436047
```

```r
##Sorted by support
inspect(sort(basketrules, by = "support")[1:5])
```

```
##     lhs                  rhs                  support    confidence
## 1   {}                => {whole milk}       0.25551601 0.2555160
## 150 {other vegetables} => {whole milk}       0.07483477 0.3867578
## 151 {whole milk}       => {other vegetables} 0.07483477 0.2928770
## 149 {rolls/buns}       => {whole milk}       0.05663447 0.3079049
## 148 {yogurt}           => {whole milk}       0.05602440 0.4016035
##     lift
## 1   1.000000
## 150 1.513634
## 151 1.513634
## 149 1.205032
## 148 1.571735
```