Reading the data

First, we use the library "arules" as it is necessary to do association rule mining. Next, we use read.transactions on groceries.txt, separating based on comma and setting a basket format.

```r
library(arules)  # has a big ecosystem of packages built around it
```

```
## Warning: package 'arules' was built under R version 3.2.2
```

```
## Loading required package: Matrix
##
## Attaching package: 'arules'
##
## The following objects are masked from 'package:base':
##
##     %in%, write
```

```r
groceries <- read.transactions("data/groceries.txt", format = 'basket', sep = ',')
```

Creating the rules

Next, we run the apriori algorithm on the groceries data. A support of .005 means that the algorithm will filter out any item sets that do not occur in at least .005 of the transactions. The confidence of .55 means that the left hand side will result in the right hand side at least 55% of the time. Finally, the maxlen of 5 means that the total number of items in both the right hand side and the left hand side can add up to at most 5.

```r
groceriesrules <- apriori(groceries, parameter=list(support=.005, confidence=.59, maxlen=5))
```

```
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen maxlen
##        0.59    0.1    1 none FALSE            TRUE   0.005      1      5
##  target   ext
##   rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## apriori - find association rules with the apriori algorithm
## version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [120 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [23 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
# Look at the output
inspect(groceriesrules)
```

```
##    lhs                      rhs                    support confidence     lift
## 1  {onions,
##      root vegetables}       => {other vegetables} 0.005693950  0.6021505 3.112008
## 2  {curd,
##      tropical fruit}        => {whole milk}       0.006507372  0.6336634 2.479936
## 3  {domestic eggs,
##      margarine}             => {whole milk}       0.005185562  0.6219512 2.434099
## 4  {butter,
##      domestic eggs}         => {whole milk}       0.005998983  0.6210526 2.430582
## 5  {butter,
##      whipped/sour cream}    => {whole milk}       0.006710727  0.6600000 2.583008
## 6  {bottled water,
##      butter}                => {whole milk}       0.005388917  0.6022727 2.357084
## 7  {butter,
##      tropical fruit}        => {whole milk}       0.006202339  0.6224490 2.436047
## 8  {butter,
##      root vegetables}       => {whole milk}       0.008235892  0.6377953 2.496107
## 9  {butter,
##      yogurt}                => {whole milk}       0.009354347  0.6388889 2.500387
## 10 {domestic eggs,
##      pip fruit}             => {whole milk}       0.005388917  0.6235294 2.440275
## 11 {domestic eggs,
##      tropical fruit}        => {whole milk}       0.006914082  0.6071429 2.376144
## 12 {domestic eggs,
##      root vegetables}       => {whole milk}       0.008540925  0.5957447 2.331536
## 13 {pip fruit,
##      whipped/sour cream}    => {other vegetables} 0.005592272  0.6043956 3.123610
## 14 {pip fruit,
##      whipped/sour cream}    => {whole milk}       0.005998983  0.6483516 2.537421
## 15 {fruit/vegetable juice,
##      other vegetables,
##      yogurt}                => {whole milk}       0.005083884  0.6172840 2.415833
## 16 {other vegetables,
##      root vegetables,
##      whipped/sour cream}    => {whole milk}       0.005185562  0.6071429 2.376144
## 17 {other vegetables,
##      pip fruit,
##      root vegetables}       => {whole milk}       0.005490595  0.6750000 2.641713
## 18 {pip fruit,
##      root vegetables,
##      whole milk}            => {other vegetables} 0.005490595  0.6136364 3.171368
## 19 {other vegetables,
##      pip fruit,
##      yogurt}                => {whole milk}       0.005083884  0.6250000 2.446031
## 20 {citrus fruit,
##      root vegetables,
##      whole milk}            => {other vegetables} 0.005795628  0.6333333 3.273165
## 21 {root vegetables,
##      tropical fruit,
##      yogurt}                => {whole milk}       0.005693950  0.7000000 2.739554
```

```
## 22 {other vegetables,
##    tropical fruit,
##    yogurt}                   => {whole milk}      0.007625826  0.6198347 2.425816
## 23 {other vegetables,
##    root vegetables,
##    yogurt}                   => {whole milk}      0.007829181  0.6062992 2.372842
```

Looking at the output, we can see that 23 rules were generated. Also, the right hand side always has only one item in it, and is either whole milk or other vegatables. This is a rather large set of rules, so we create subsets next to narrow down the number of rules.

Creating a subset based on lift

Next, we create a subset using a lift of at least 3.

```
## Choose a subset
inspect(subset(groceriesrules, subset=lift > 3))
```

```
##   lhs                    rhs                   support confidence      lift
## 1 {onions,
##    root vegetables}    => {other vegetables} 0.005693950  0.6021505 3.112008
## 2 {pip fruit,
##    whipped/sour cream} => {other vegetables} 0.005592272  0.6043956 3.123610
## 3 {pip fruit,
##    root vegetables,
##    whole milk}         => {other vegetables} 0.005490595  0.6136364 3.171368
## 4 {citrus fruit,
##    root vegetables,
##    whole milk}         => {other vegetables} 0.005795628  0.6333333 3.273165
```

This subset contains only 4 rules, and the right hand side is always other vegetables. With this subset, we can that people who buy a variety of types of fruit and vegetables are also likely to buy other types of vegetables.

Creating a subset based on confidence

Next, we create a subset based on confidence, using a confidence of .65.

```
inspect(subset(groceriesrules, subset=confidence > 0.65))
```

```
##   lhs                    rhs              support confidence      lift
## 1 {butter,
##    whipped/sour cream} => {whole milk} 0.006710727       0.660 2.583008
## 2 {other vegetables,
##    pip fruit,
##    root vegetables}    => {whole milk} 0.005490595       0.675 2.641713
## 3 {root vegetables,
##    tropical fruit,
##    yogurt}             => {whole milk} 0.005693950       0.700 2.739554
```

This subset contains just three rules, and the right hand side is always whole milk. Therefore, we can infer that whole milk is the most common item to be included if we know about certain other items in the grocery basket.