ASSOCIATION RULE MINING

install.packages("arulesViz")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

install.packages("arules")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

library(arules)

## Loading required package: Matrix

##   
## Attaching package: 'arules'

## The following objects are masked from 'package:base':  
##   
## abbreviate, write

library(arulesViz)  
library(ggplot2)  
data("Groceries")  
summary(Groceries)

## transactions as itemMatrix in sparse format with  
## 9835 rows (elements/itemsets/transactions) and  
## 169 columns (items) and a density of 0.02609146   
##   
## most frequent items:  
## whole milk other vegetables rolls/buns soda   
## 2513 1903 1809 1715   
## yogurt (Other)   
## 1372 34055   
##   
## element (itemset/transaction) length distribution:  
## sizes  
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16   
## 2159 1643 1299 1005 855 645 545 438 350 246 182 117 78 77 55 46   
## 17 18 19 20 21 22 23 24 26 27 28 29 32   
## 29 14 14 9 11 4 6 1 1 1 1 3 1   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 2.000 3.000 4.409 6.000 32.000   
##   
## includes extended item information - examples:  
## labels level2 level1  
## 1 frankfurter sausage meat and sausage  
## 2 sausage sausage meat and sausage  
## 3 liver loaf sausage meat and sausage

#total number of transactions  
tot\_trans <- 9835\*169\*0.02609146  
inspect(Groceries[1:3])

## items   
## [1] {citrus fruit,   
## semi-finished bread,   
## margarine,   
## ready soups}   
## [2] {tropical fruit,   
## yogurt,   
## coffee}   
## [3] {whole milk}

itemFrequency(Groceries[,1])

## frankfurter   
## 0.05897306

#total number of transactions for the particular item  
part\_item <- 0.05897306 \*9835  
#item frequency for the items with support count of 10%  
itemFrequencyPlot(Groceries,support = 0.10)

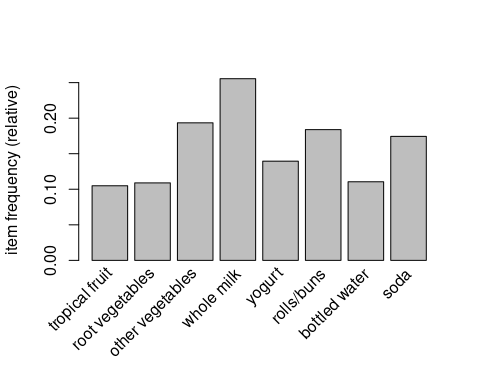


Fig.1

#most occuring top 10 items  
itemFrequencyPlot(Groceries,topN = 10)

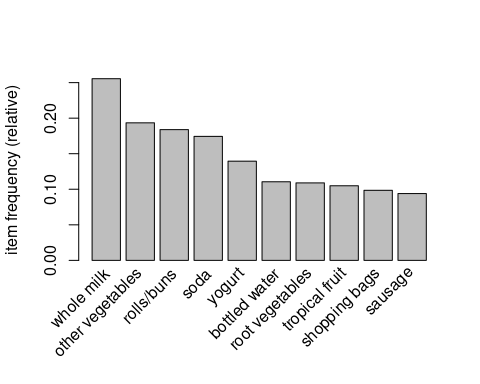


Fig.2

#association rule1  
apriori(Groceries,parameter = list(support = 0.002,confidence = 0.5,minlen = 2)) -> rule1

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.5 0.1 1 none FALSE TRUE 5 0.002 2  
## maxlen target ext  
## 10 rules TRUE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 19   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].  
## sorting and recoding items ... [147 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 done [0.00s].  
## writing ... [1098 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

inspect(head(rule1,5))

## lhs rhs support confidence coverage   
## [1] {cereals} => {whole milk} 0.003660397 0.6428571 0.005693950  
## [2] {jam} => {whole milk} 0.002948653 0.5471698 0.005388917  
## [3] {specialty cheese} => {other vegetables} 0.004270463 0.5000000 0.008540925  
## [4] {rice} => {other vegetables} 0.003965430 0.5200000 0.007625826  
## [5] {rice} => {whole milk} 0.004677173 0.6133333 0.007625826  
## lift count  
## [1] 2.515917 36   
## [2] 2.141431 29   
## [3] 2.584078 42   
## [4] 2.687441 39   
## [5] 2.400371 46

inspect(head(sort(rule1,by = "lift",5)))

## lhs rhs support confidence coverage lift count  
## [1] {butter,   
## hard cheese} => {whipped/sour cream} 0.002033554 0.5128205 0.003965430 7.154028 20  
## [2] {beef,   
## citrus fruit,   
## other vegetables} => {root vegetables} 0.002135231 0.6363636 0.003355363 5.838280 21  
## [3] {citrus fruit,   
## tropical fruit,   
## other vegetables,   
## whole milk} => {root vegetables} 0.003152008 0.6326531 0.004982206 5.804238 31  
## [4] {citrus fruit,   
## other vegetables,   
## frozen vegetables} => {root vegetables} 0.002033554 0.6250000 0.003253686 5.734025 20  
## [5] {beef,   
## tropical fruit,   
## other vegetables} => {root vegetables} 0.002745297 0.6136364 0.004473818 5.629770 27  
## [6] {root vegetables,   
## yogurt,   
## bottled water} => {tropical fruit} 0.002236909 0.5789474 0.003863752 5.517391 22

plot(rule1,method = "grouped")

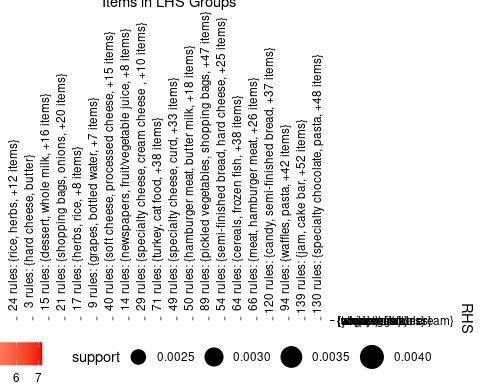


Fig.3

#association rule2  
apriori(Groceries,parameter = list(support = 0.007,confidence = 0.6)) -> rule2

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.6 0.1 1 none FALSE TRUE 5 0.007 1  
## maxlen target ext  
## 10 rules TRUE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 68   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].  
## sorting and recoding items ... [104 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 done [0.00s].  
## writing ... [4 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

inspect(head(rule2,4))

## lhs rhs support confidence coverage lift count  
## [1] {root vegetables,   
## butter} => {whole milk} 0.008235892 0.6377953 0.01291307 2.496107 81  
## [2] {butter,   
## yogurt} => {whole milk} 0.009354347 0.6388889 0.01464159 2.500387 92  
## [3] {tropical fruit,   
## other vegetables,   
## yogurt} => {whole milk} 0.007625826 0.6198347 0.01230300 2.425816 75  
## [4] {root vegetables,   
## other vegetables,   
## yogurt} => {whole milk} 0.007829181 0.6062992 0.01291307 2.372842 77

plot(rule2)

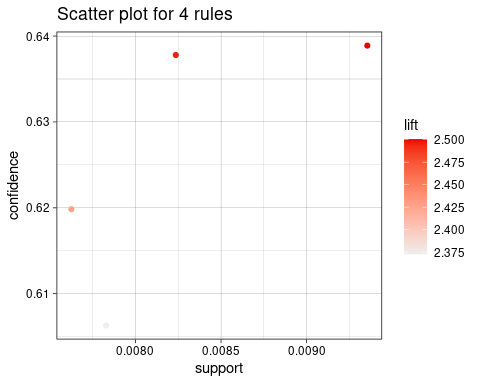


Fig.4

plot(rule2,method = "grouped")

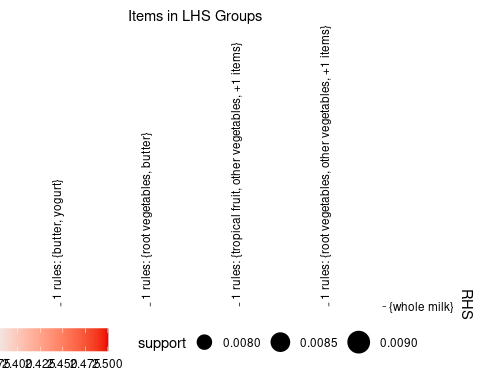


Fig.5