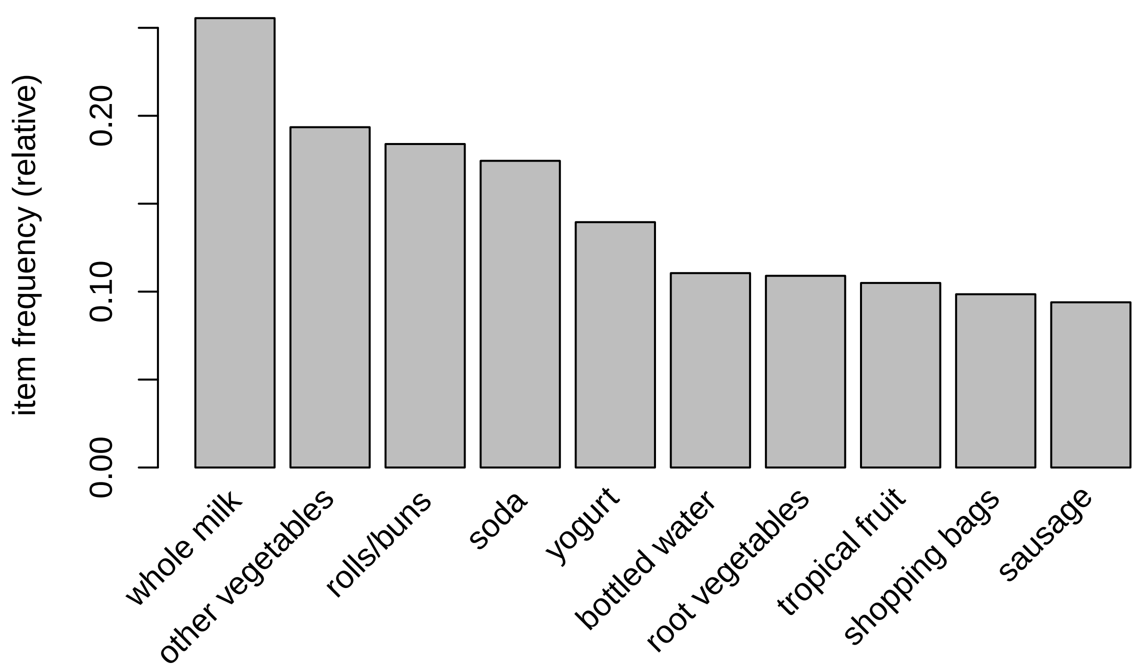
machine learning exercise NO.8 Association rule mining

Yolanda Cheung/Geer Zhang

8/14/2023

##Association rule mining

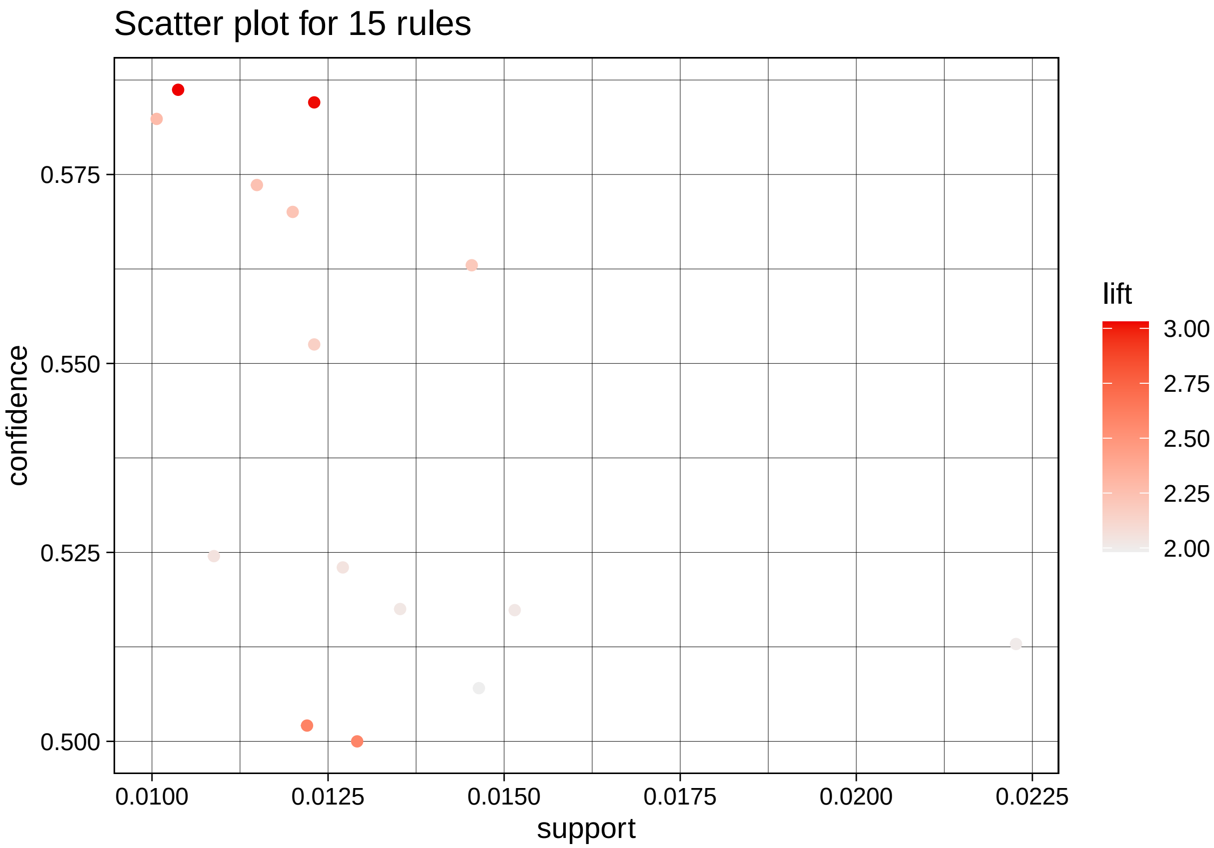
```
## Loading required package: Matrix

##
## Attaching package: 'arules'

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




```
## Apriori
```

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

```
##     lhs                                  rhs                support
## [1] {citrus fruit, root vegetables}   => {other vegetables} 0.01037112
## [2] {root vegetables, tropical fruit} => {other vegetables} 0.01230300
## [3] {rolls/buns, root vegetables}     => {other vegetables} 0.01220132
## [4] {root vegetables, yogurt}         => {other vegetables} 0.01291307
## [5] {curd, yogurt}                    => {whole milk}       0.01006609
## [6] {butter, other vegetables}        => {whole milk}       0.01148958
##     confidence coverage   lift     count
## [1] 0.5862069  0.01769192 3.029608 102
## [2] 0.5845411  0.02104728 3.020999 121
## [3] 0.5020921  0.02430097 2.594890 120
## [4] 0.5000000  0.02582613 2.584078 127
## [5] 0.5823529  0.01728521 2.279125  99
## [6] 0.5736041  0.02003050 2.244885 113
```



This analysis delves into a dataset containing shopping transactions from a grocery store. By examining this dataset, we aim to uncover patterns and relationships between different items that customers tend to purchase together.

1. Understanding the Most Popular Items:Initially, we visualized the 10 most frequently purchased items. This gave us an immediate understanding of the core items that drive customers to the store. These could be staple items that every household commonly needs or special items that are unique to this grocery store.

2. Discovering Shopping Patterns with Association Rules: We applied the Apriori algorithm, a prominent method for finding items that are often purchased together. For instance, if bread and butter frequently appear in the same transaction, the algorithm will recognize this pattern.To ensure the quality and relevance of our findings, we focused on patterns that appear in at least 1% of all transactions (support) and have at least 50% likelihood that if a customer buys item A, they will also buy item B (confidence).

3. Prioritizing the Most Relevant Patterns:After extracting potential shopping patterns, we ranked them by their 'lift' value. Lift indicates the strength of a rule over random co-occurrence. For instance, a lift value greater than 1 for the bread and butter rule would indicate that bread and butter are purchased together more frequently than if the two items were bought independently.

4. Visual Representation for Clearer Insights:Lastly, we visualized these associations, providing a clear, visual representation of the relationships between products. Such a scatter plot allows stakeholders to immediately identify the strongest associations without delving into the numbers.

In summary, by analyzing the shopping transactions, we've gained insights into the most popular items and the relationships between products. Such insights can drive marketing strategies, optimize store layout, and even influence inventory decisions.