INF0613 – Aprendizado de Máquina Não Supervisionado

Trabalho 1 - Regras de Associação

Leonardo Cesar Silva dos Santos Fernando Augusto Cardoso Candalaft

Neste primeiro trabalho vamos minerar Regras de Associação em uma base de dados que contém as vendas de uma padaria. A base de dados está disponível na página da disciplina no Moodle (arquivo `bakery.csv`).

Atividade 0 – Configurando o ambiente

Antes de começar a implementação do seu trabalho configure o *workspace* e importe todos os pacotes:

```
# Adicione os demais pacotes usados
# Bibliotecas usadas neste trabalho:
library(arules)
library(arulesViz)

# Configurando ambiente de trabalho:
setwd("~/workspace/mdc/04_aprendizado_nao_supervisionado_I/test01")
```

Atividade 1 – Análise Exploratória da Base de Dados (*3,0 pts*)

Dado um caminho para uma base de dados, leia as transações e faça uma análise Exploratória sobre elas. Use as funções `summary`, `inspect` e `itemFrequencyPlot`. Na função `inspect` limite sua análise às 10 primeiras transações e na função `itemFrequencyPlot` gere um gráfico com a frequência relativa dos 30 itens mais frequentes.

```
# Ler transações
bakery_path <- "./bakery.csv"
bakery_data <- read.transactions(bakery_path, format="basket", sep=",")

# Visualizando transações
inspect(bakery_data[1:10], linebreak=T)
```

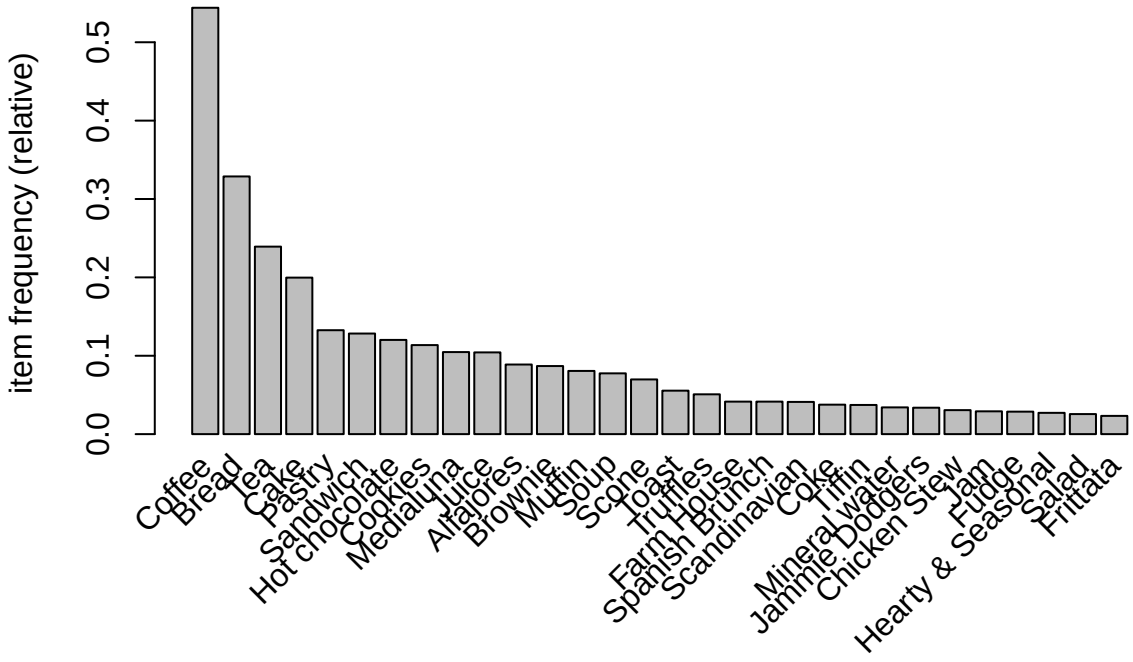
```
##      items
## [1]  {Coffee,
##       Vegan mincepie}
## [2]  {Farm House,
##       Muffin,
##       Tea}
## [3]  {Bread,
##       Ellas Kitchen Pouches,
##       Jam,
##       Juice,
##       Muffin}
## [4]  {Bread,
##       Juice,
```

```
##       Salad,
##       Sandwich}
## [5]  {Cake,
##       Coffee,
##       Sandwich,
##       Smoothies,
##       Soup}
## [6]  {Bread,
##       Medialuna}
## [7]  {Chocolates,
##       Coffee,
##       Tea}
## [8]  {Alfajores,
##       Brownie,
##       Medialuna}
## [9]  {Alfajores,
##       Coffee,
##       Fudge}
## [10] {Bread,
##       Pastry}
```

```
# Sumário da base
summary(bakery_data)
```

```
## transactions as itemMatrix in sparse format with
##  2579 rows (elements/itemsets/transactions) and
##  91 columns (items) and a density of 0.0352
##
## most frequent items:
##  Coffee   Bread     Tea    Cake  Pastry (Other)
##    1403     848     617     515     342    4532
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10
##   20  664 1041  591  189   52   15    4    2    1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0     2.0     3.0     3.2     4.0    10.0
##
## includes extended item information - examples:
##                     labels
## 1 Afternoon with the baker
## 2               Alfajores
## 3          Argentina Night
```

```
# Analisando a frequência dos itens
itemFrequencyPlot(bakery_data, topN=30, type="relative")
```



Análise

a) Descreva a base de dados discutindo os resultados das funções acima.

Resposta: Os dados utilizados ("bakery") apresentam valores transacionais no formato {bread, tea, etc.}. A base utilizada possui 2579 transações e um total de 91 itens. Constatamos por meio da sumarização das informações que os itens mais frequentes são Coffee, Bread, Tea, Cake e Pastry, com frequências 1403, 848, 617, 515 e 342, respectivamente. Coffee representa uma frequência relativa próxima de 60% e Bread representa uma frequência relativa próxima de 30%. Estes dois itens são os mais frequentes, algo que faz sentido, dado que são complementares na alimentação do dia a dia no contexto de uma padaria.

b) Ao gerarmos o gráfico de frequências, temos uma representação visual de uma informação já presente no resultado da função `summary`. Contudo, esse gráfico nos dá uma visão mais ampla da base. Assim podemos ver a frequência de outros itens em relação aos 10 mais frequentes. Quais informações podemos obter a partir desse gráfico (e da análise anterior) para nos ajudar na extração de regras de associação com o algoritmo `apriori`? Isto é, como a frequência dos itens pode afetar os parâmetros de configuração do algoritmo `apriori`?

Resposta: O algoritmo Apriori utiliza o que chamamos de propriedade anti-monotônica do suporte, isto é, usa o fato de que se um conjunto de itens é frequente, então todos os seus subcojuntos também são. Portanto, itens frequentes ajudam o método Apriori a percorrer um espaço menor de possíveis regras para um determinado limiar de suporte. O fato de termos poucos itens com alta frequência indica que devemos escolher com cuidado nossos valores de suporte e confiança no algoritmo Apriori, dado que um suporte relativamente alto pode não trazer informações de relação para itens mais raros, de modo que podemos perder informações relevante. Temos algo semelhante para um suporte alto, que pode acarretar em um valor elevado de regras de associação. No nosso problema iremos explorar valores de suporte baixos, dado que vamos buscar captar regras não triviais, uma vez que em uma padaria já é esperado que os clientes comprem café, pão, bolo etc.

Atividade 2 – Minerando Regras (*3,5 pts*)

Use o algoritmo `apriori` para minerar regras na base de dados fornecida. Experimente com pelo menos *3 conjuntos* de valores diferentes de suporte e confiança para encontrar regras de associação. Imprima as cinco regras com o maior confiança de cada conjunto escolhido. Lembre-se de usar seu conhecimento sobre a base,

obtido na questão anterior, para a escolha dos valores de suporte e confiança.

```
# Conjunto 1: suporte = 0.05 e confiança = 0.5
regras01 <- apriori(bakery_data,
                    parameter=list(supp=0.05, conf=0.5))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5    0.05      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: 128
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[91 item(s), 2579 transaction(s)] done [0.00s].
## sorting and recoding items ... [17 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [8 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
regras01
```

```
## set of 8 rules
```

```
inspect(regras01, linebreak=T)
```

```
##     lhs                rhs      support confidence coverage lift count
## [1] {}              => {Coffee}  0.5440      0.544    1.000 1.00  1403
## [2] {Medialuna}     => {Coffee}  0.0609      0.581    0.105 1.07   157
## [3] {Juice}         => {Coffee}  0.0539      0.517    0.104 0.95   139
## [4] {Cookies}       => {Coffee}  0.0655      0.577    0.114 1.06   169
## [5] {Hot chocolate} => {Coffee}  0.0659      0.548    0.120 1.01   170
## [6] {Pastry}        => {Coffee}  0.0748      0.564    0.133 1.04   193
## [7] {Sandwich}      => {Coffee}  0.0768      0.598    0.128 1.10   198
## [8] {Cake}          => {Coffee}  0.1117      0.559    0.200 1.03   288
```

```
# Conjunto 2: suporte = 0.005 e confiança = 0.7
regras02 <- apriori(bakery_data,
                    parameter=list(supp=0.005, conf=0.7))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.7    0.1    1 none FALSE            TRUE       5   0.005      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: 12
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[91 item(s), 2579 transaction(s)] done [0.00s].
## sorting and recoding items ... [44 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [10 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
regras02
```

```
## set of 10 rules
```

```
inspect(regras02, linebreak=T)
```

```
##      lhs                          rhs      support confidence coverage  lift count
## [1]  {Extra Salami or Feta}  => {Salad}  0.00620      0.800  0.00775 31.26    16
## [2]  {Extra Salami or Feta}  => {Coffee} 0.00698      0.900  0.00775  1.65    18
## [3]  {Keeping It Local}      => {Coffee} 0.00969      0.781  0.01241  1.44    25
## [4]  {Toast}                 => {Coffee} 0.03994      0.720  0.05545  1.32   103
## [5]  {Extra Salami or Feta,
##       Salad}                 => {Coffee} 0.00543      0.875  0.00620  1.61    14
## [6]  {Coffee,
##       Extra Salami or Feta}  => {Salad}  0.00543      0.778  0.00698 30.39    14
## [7]  {Salad,
##       Sandwich}              => {Coffee} 0.00543      0.824  0.00659  1.51    14
## [8]  {Juice,
##       Spanish Brunch}        => {Coffee} 0.00504      0.765  0.00659  1.41    13
## [9]  {Cookies,
##       Scone}                 => {Coffee} 0.00504      0.765  0.00659  1.41    13
## [10] {Juice,
##       Pastry}                => {Coffee} 0.00504      0.765  0.00659  1.41    13
```

```
# Conjunto 3: suporte = 0.001 e confiança = 0.7
regras03 <- apriori(bakery_data,
                    parameter=list(supp=0.001, conf=0.7))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.7    0.1    1 none FALSE            TRUE       5   0.001      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: 2
##
## set item appearances ...[0 item(s)] done [0.00s].
```

```
## set transactions ...[91 item(s), 2579 transaction(s)] done [0.00s].
## sorting and recoding items ... [75 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 done [0.00s].
## writing ... [185 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
regras03
```

```
## set of 185 rules
```

```
inspect(regras03, linebreak=T)
```

```
##       lhs                            rhs              support confidence coverage   lift count
## [1]   {Bread Pudding}             => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [2]   {Muesli}                    => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [3]   {Duck egg}                  => {Spanish Brunch} 0.00116      1.000  0.00116  24.10     3
## [4]   {Postcard}                  => {Tshirt}         0.00116      0.750  0.00155 241.78     3
## [5]   {Basket}                    => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [6]   {Mighty Protein}            => {Coffee}         0.00194      0.714  0.00271   1.31     5
## [7]   {Nomad bag}                 => {Bread}          0.00194      0.833  0.00233   2.53     5
## [8]   {Drinking chocolate spoons} => {Coffee}         0.00194      0.714  0.00271   1.31     5
## [9]   {Extra Salami or Feta}      => {Salad}          0.00620      0.800  0.00775  31.26    16
## [10]  {Extra Salami or Feta}      => {Coffee}         0.00698      0.900  0.00775   1.65    18
## [11]  {Keeping It Local}          => {Coffee}         0.00969      0.781  0.01241   1.44    25
## [12]  {Toast}                     => {Coffee}         0.03994      0.720  0.05545   1.32   103
## [13]  {Bread,
##        Drinking chocolate spoons} => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [14]  {Chocolates,
##        Juice}                     => {Hot chocolate}  0.00116      0.750  0.00155   6.24     3
## [15]  {Chocolates,
##        Hot chocolate}             => {Juice}          0.00116      0.750  0.00155   7.19     3
## [16]  {Chocolates,
##        Juice}                     => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [17]  {Mineral water,
##        Pick and Mix Bowls}        => {Juice}          0.00116      1.000  0.00116   9.59     3
## [18]  {Juice,
##        Pick and Mix Bowls}        => {Mineral water}  0.00116      0.750  0.00155  21.98     3
## [19]  {Crisps,
##        Juice}                     => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [20]  {Cake,
##        Eggs}                      => {Bread}          0.00116      1.000  0.00116   3.04     3
## [21]  {Granola,
##        Pastry}                    => {Bread}          0.00116      0.750  0.00155   2.28     3
## [22]  {Granola,
##        Pastry}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [23]  {Cookies,
##        My-5 Fruit Shoot}          => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [24]  {Extra Salami or Feta,
##        Juice}                     => {Salad}          0.00116      1.000  0.00116  39.08     3
## [25]  {Extra Salami or Feta,
##        Sandwich}                  => {Salad}          0.00116      1.000  0.00116  39.08     3
## [26]  {Cake,
##        Extra Salami or Feta}      => {Salad}          0.00116      1.000  0.00116  39.08     3
## [27]  {Bread,
```

```
##        Extra Salami or Feta}      => {Salad}          0.00233      1.000  0.00233  39.08     6
## [28]  {Extra Salami or Feta,
##        Salad}                     => {Coffee}         0.00543      0.875  0.00620   1.61    14
## [29]  {Coffee,
##        Extra Salami or Feta}      => {Salad}          0.00543      0.778  0.00698  30.39    14
## [30]  {Extra Salami or Feta,
##        Spanish Brunch}            => {Coffee}         0.00194      1.000  0.00194   1.84     5
## [31]  {Extra Salami or Feta,
##        Toast}                     => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [32]  {Extra Salami or Feta,
##        Juice}                     => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [33]  {Extra Salami or Feta,
##        Sandwich}                  => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [34]  {Cake,
##        Extra Salami or Feta}      => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [35]  {Cookies,
##        Keeping It Local}          => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [36]  {Keeping It Local,
##        Pastry}                    => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [37]  {Art Tray,
##        Spanish Brunch}            => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [38]  {Bakewell,
##        Juice}                     => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [39]  {Bakewell,
##        Cookies}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [40]  {Hot chocolate,
##        Vegan mincepie}            => {Coffee}         0.00271      0.778  0.00349   1.43     7
## [41]  {Cake,
##        Vegan mincepie}            => {Coffee}         0.00349      0.818  0.00427   1.50     9
## [42]  {Juice,
##        The Nomad}                 => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [43]  {Brownie,
##        Frittata}                  => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [44]  {Frittata,
##        Juice}                     => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [45]  {Cookies,
##        Frittata}                  => {Bread}          0.00116      0.750  0.00155   2.28     3
## [46]  {Brownie,
##        Smoothies}                 => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [47]  {Juice,
##        Smoothies}                 => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [48]  {Pastry,
##        Smoothies}                 => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [49]  {Hearty & Seasonal,
##        Hot chocolate}             => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [50]  {Hearty & Seasonal,
##        Sandwich}                  => {Coffee}         0.00427      0.846  0.00504   1.56    11
## [51]  {Cake,
##        Hearty & Seasonal}         => {Coffee}         0.00271      1.000  0.00271   1.84     7
## [52]  {Jam,
##        Medialuna}                 => {Bread}          0.00155      0.800  0.00194   2.43     4
## [53]  {Jam,
##        Juice}                     => {Bread}          0.00116      1.000  0.00116   3.04     3
## [54]  {Chicken Stew,
```

```
##        Salad}                    => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [55]  {Mineral water,
##        Salad}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [56]  {Salad,
##        Spanish Brunch}           => {Coffee}         0.00271      0.700  0.00388   1.29     7
## [57]  {Salad,
##        Toast}                    => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [58]  {Salad,
##        Sandwich}                 => {Coffee}         0.00543      0.824  0.00659   1.51    14
## [59]  {Cake,
##        Salad}                    => {Coffee}         0.00388      0.769  0.00504   1.41    10
## [60]  {Chicken Stew,
##        Tiffin}                   => {Bread}          0.00116      0.750  0.00155   2.28     3
## [61]  {Chicken Stew,
##        Pastry}                   => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [62]  {Jammie Dodgers,
##        Truffles}                 => {Cake}           0.00116      0.750  0.00155   3.76     3
## [63]  {Jammie Dodgers,
##        Truffles}                 => {Bread}          0.00116      0.750  0.00155   2.28     3
## [64]  {Jammie Dodgers,
##        Sandwich}                 => {Coffee}         0.00194      0.714  0.00271   1.31     5
## [65]  {Scandinavian,
##        Toast}                    => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [66]  {Hot chocolate,
##        Scandinavian}             => {Bread}          0.00116      0.750  0.00155   2.28     3
## [67]  {Hot chocolate,
##        Scandinavian}             => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [68]  {Scandinavian,
##        Tea}                      => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [69]  {Mineral water,
##        Tiffin}                   => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [70]  {Mineral water,
##        Spanish Brunch}           => {Bread}          0.00116      0.750  0.00155   2.28     3
## [71]  {Mineral water,
##        Toast}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [72]  {Medialuna,
##        Mineral water}            => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [73]  {Hot chocolate,
##        Mineral water}            => {Coffee}         0.00233      0.750  0.00310   1.38     6
## [74]  {Farm House,
##        Spanish Brunch}           => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [75]  {Farm House,
##        Toast}                    => {Coffee}         0.00233      1.000  0.00233   1.84     6
## [76]  {Alfajores,
##        Farm House}               => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [77]  {Farm House,
##        Juice}                    => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [78]  {Farm House,
##        Hot chocolate}            => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [79]  {Medialuna,
##        Tiffin}                   => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [80]  {Pastry,
##        Tiffin}                   => {Tea}            0.00194      0.714  0.00271   2.99     5
## [81]  {Coke,
```

```
##        Spanish Brunch}          => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [82]  {Coke,
##        Muffin}                  => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [83]  {Coke,
##        Cookies}                 => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [84]  {Coke,
##        Pastry}                  => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [85]  {Spanish Brunch,
##        Toast}                   => {Coffee}         0.00233      0.857  0.00271   1.58     6
## [86]  {Alfajores,
##        Spanish Brunch}          => {Tea}            0.00233      0.857  0.00271   3.58     6
## [87]  {Juice,
##        Spanish Brunch}          => {Coffee}         0.00504      0.765  0.00659   1.41    13
## [88]  {Pastry,
##        Spanish Brunch}          => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [89]  {Sandwich,
##        Spanish Brunch}          => {Coffee}         0.00465      0.750  0.00620   1.38    12
## [90]  {Scone,
##        Toast}                   => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [91]  {Soup,
##        Toast}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [92]  {Brownie,
##        Toast}                   => {Tea}            0.00116      0.750  0.00155   3.13     3
## [93]  {Alfajores,
##        Toast}                   => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [94]  {Pastry,
##        Toast}                   => {Coffee}         0.00427      0.846  0.00504   1.56    11
## [95]  {Cake,
##        Toast}                   => {Coffee}         0.00465      0.706  0.00659   1.30    12
## [96]  {Cookies,
##        Scone}                   => {Coffee}         0.00504      0.765  0.00659   1.41    13
## [97]  {Medialuna,
##        Soup}                    => {Coffee}         0.00194      0.714  0.00271   1.31     5
## [98]  {Juice,
##        Pastry}                  => {Coffee}         0.00504      0.765  0.00659   1.41    13
## [99]  {Pastry,
##        Sandwich}                => {Coffee}         0.00310      0.800  0.00388   1.47     8
## [100] {Extra Salami or Feta,
##        Salad,
##        Spanish Brunch}          => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [101] {Extra Salami or Feta,
##        Juice,
##        Salad}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [102] {Coffee,
##        Extra Salami or Feta,
##        Juice}                   => {Salad}          0.00116      1.000  0.00116  39.08     3
## [103] {Extra Salami or Feta,
##        Salad,
##        Sandwich}                => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [104] {Coffee,
##        Extra Salami or Feta,
##        Sandwich}                => {Salad}          0.00116      1.000  0.00116  39.08     3
## [105] {Cake,
##        Extra Salami or Feta,
```

```
##        Salad}                    => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [106] {Cake,
##        Coffee,
##        Extra Salami or Feta}     => {Salad}          0.00116      1.000  0.00116  39.08     3
## [107] {Bread,
##        Coffee,
##        Extra Salami or Feta}     => {Salad}          0.00155      1.000  0.00155  39.08     4
## [108] {Bread,
##        Sandwich,
##        Smoothies}                => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [109] {Bread,
##        Jam,
##        Medialuna}                => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [110] {Coffee,
##        Jam,
##        Medialuna}                => {Bread}          0.00116      1.000  0.00116   3.04     3
## [111] {Coffee,
##        Jam,
##        Tea}                      => {Bread}          0.00116      0.750  0.00155   2.28     3
## [112] {Cake,
##        Salad,
##        Sandwich}                 => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [113] {Cake,
##        Salad,
##        Tea}                      => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [114] {Chicken Stew,
##        Coke,
##        Truffles}                 => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [115] {Chicken Stew,
##        Coke,
##        Juice}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [116] {Chicken Stew,
##        Coke,
##        Sandwich}                 => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [117] {Chicken Stew,
##        Juice,
##        Sandwich}                 => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [118] {Cake,
##        Jammie Dodgers,
##        Truffles}                 => {Bread}          0.00116      1.000  0.00116   3.04     3
## [119] {Bread,
##        Jammie Dodgers,
##        Truffles}                 => {Cake}           0.00116      1.000  0.00116   5.01     3
## [120] {Jammie Dodgers,
##        Juice,
##        Muffin}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [121] {Cake,
##        Jammie Dodgers,
##        Juice}                    => {Bread}          0.00155      0.800  0.00194   2.43     4
## [122] {Jammie Dodgers,
##        Juice,
##        Tea}                      => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [123] {Cake,
##        Hot chocolate,
```

```
##        Jammie Dodgers}          => {Bread}          0.00116      0.750  0.00155   2.28     3
## [124] {Bread,
##        Jammie Dodgers,
##        Sandwich}                => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [125] {Mineral water,
##        Scone,
##        Truffles}                => {Bread}          0.00116      0.750  0.00155   2.28     3
## [126] {Bread,
##        Mineral water,
##        Truffles}                => {Scone}          0.00116      1.000  0.00116  14.33     3
## [127] {Bread,
##        Mineral water,
##        Scone}                   => {Truffles}       0.00116      0.750  0.00155  14.77     3
## [128] {Bread,
##        Scone,
##        Truffles}                => {Mineral water}  0.00116      0.750  0.00155  21.98     3
## [129] {Alfajores,
##        Cookies,
##        Mineral water}           => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [130] {Alfajores,
##        Mineral water,
##        Tea}                     => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [131] {Farm House,
##        Medialuna,
##        Tea}                     => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [132] {Cookies,
##        Tea,
##        Tiffin}                  => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [133] {Coffee,
##        Pastry,
##        Tiffin}                  => {Tea}            0.00116      1.000  0.00116   4.18     3
## [134] {Bread,
##        Sandwich,
##        Tiffin}                  => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [135] {Bread,
##        Cake,
##        Tiffin}                  => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [136] {Coke,
##        Juice,
##        Truffles}                => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [137] {Coke,
##        Sandwich,
##        Truffles}                => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [138] {Coke,
##        Sandwich,
##        Soup}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [139] {Bread,
##        Coke,
##        Soup}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [140] {Sandwich,
##        Spanish Brunch,
##        Truffles}                => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [141] {Bread,
##        Juice,
```

```
##        Spanish Brunch}          => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [142] {Sandwich,
##        Spanish Brunch,
##        Tea}                     => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [143] {Bread,
##        Sandwich,
##        Spanish Brunch}          => {Coffee}         0.00194      1.000  0.00194   1.84     5
## [144] {Juice,
##        Sandwich,
##        Truffles}                => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [145] {Bread,
##        Cookies,
##        Truffles}                => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [146] {Coffee,
##        Cookies,
##        Truffles}                => {Bread}          0.00116      1.000  0.00116   3.04     3
## [147] {Pastry,
##        Sandwich,
##        Truffles}                => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [148] {Bread,
##        Sandwich,
##        Truffles}                => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [149] {Cake,
##        Hot chocolate,
##        Toast}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [150] {Bread,
##        Pastry,
##        Toast}                   => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [151] {Bread,
##        Cake,
##        Toast}                   => {Tea}            0.00116      0.750  0.00155   3.13     3
## [152] {Cookies,
##        Juice,
##        Scone}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [153] {Cookies,
##        Hot chocolate,
##        Scone}                   => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [154] {Cake,
##        Cookies,
##        Scone}                   => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [155] {Bread,
##        Cookies,
##        Scone}                   => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [156] {Cake,
##        Hot chocolate,
##        Scone}                   => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [157] {Cake,
##        Sandwich,
##        Soup}                    => {Coffee}         0.00349      0.818  0.00427   1.50     9
## [158] {Alfajores,
##        Hot chocolate,
##        Muffin}                  => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [159] {Bread,
##        Medialuna,
```

```
##        Muffin}                    => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [160] {Muffin,
##        Pastry,
##        Tea}                       => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [161] {Bread,
##        Brownie,
##        Cookies}                   => {Coffee}         0.00194      0.714  0.00271   1.31     5
## [162] {Alfajores,
##        Cookies,
##        Tea}                       => {Juice}          0.00155      0.800  0.00194   7.67     4
## [163] {Alfajores,
##        Cookies,
##        Hot chocolate}             => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [164] {Alfajores,
##        Cookies,
##        Sandwich}                  => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [165] {Alfajores,
##        Cookies,
##        Tea}                       => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [166] {Alfajores,
##        Bread,
##        Pastry}                    => {Coffee}         0.00194      0.714  0.00271   1.31     5
## [167] {Hot chocolate,
##        Medialuna,
##        Pastry}                    => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [168] {Medialuna,
##        Pastry,
##        Tea}                       => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [169] {Bread,
##        Medialuna,
##        Sandwich}                  => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [170] {Cookies,
##        Juice,
##        Pastry}                    => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [171] {Cookies,
##        Hot chocolate,
##        Sandwich}                  => {Coffee}         0.00194      0.833  0.00233   1.53     5
## [172] {Cake,
##        Cookies,
##        Sandwich}                  => {Tea}            0.00116      1.000  0.00116   4.18     3
## [173] {Cake,
##        Cookies,
##        Sandwich}                  => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [174] {Cookies,
##        Sandwich,
##        Tea}                       => {Coffee}         0.00233      0.750  0.00310   1.38     6
## [175] {Cake,
##        Hot chocolate,
##        Pastry}                    => {Coffee}         0.00233      0.750  0.00310   1.38     6
## [176] {Bread,
##        Hot chocolate,
##        Pastry}                    => {Coffee}         0.00388      0.714  0.00543   1.31    10
## [177] {Cake,
##        Hot chocolate,
```

```
##       Sandwich}                => {Coffee}         0.00155      0.800  0.00194   1.47     4
## [178] {Cake,
##        Sandwich,
##        Tea}                    => {Coffee}         0.00349      0.818  0.00427   1.50     9
## [179] {Cake,
##        Sandwich,
##        Soup,
##        Tea}                    => {Coffee}         0.00155      1.000  0.00155   1.84     4
## [180] {Alfajores,
##        Cookies,
##        Juice,
##        Tea}                    => {Coffee}         0.00116      0.750  0.00155   1.38     3
## [181] {Alfajores,
##        Coffee,
##        Juice,
##        Tea}                    => {Cookies}        0.00116      1.000  0.00116   8.80     3
## [182] {Alfajores,
##        Coffee,
##        Cookies,
##        Tea}                    => {Juice}          0.00116      0.750  0.00155   7.19     3
## [183] {Coffee,
##        Cookies,
##        Juice,
##        Tea}                    => {Alfajores}      0.00116      0.750  0.00155   8.45     3
## [184] {Cake,
##        Cookies,
##        Sandwich,
##        Tea}                    => {Coffee}         0.00116      1.000  0.00116   1.84     3
## [185] {Cake,
##        Coffee,
##        Cookies,
##        Sandwich}               => {Tea}            0.00116      1.000  0.00116   4.18     3
```

```
regras02_sorted <- sort(regras02, by=c("lift", "confidence"), decreasing=T)
inspect(regras02_sorted[1:10], linebreak=T)
```

```
##      lhs                        rhs      support confidence coverage  lift count
## [1]  {Extra Salami or Feta}  => {Salad}  0.00620      0.800  0.00775 31.26    16
## [2]  {Coffee,
##       Extra Salami or Feta}  => {Salad}  0.00543      0.778  0.00698 30.39    14
## [3]  {Extra Salami or Feta}  => {Coffee} 0.00698      0.900  0.00775  1.65    18
## [4]  {Extra Salami or Feta,
##       Salad}                 => {Coffee} 0.00543      0.875  0.00620  1.61    14
## [5]  {Salad,
##       Sandwich}              => {Coffee} 0.00543      0.824  0.00659  1.51    14
## [6]  {Keeping It Local}      => {Coffee} 0.00969      0.781  0.01241  1.44    25
## [7]  {Juice,
##       Spanish Brunch}        => {Coffee} 0.00504      0.765  0.00659  1.41    13
## [8]  {Cookies,
##       Scone}                 => {Coffee} 0.00504      0.765  0.00659  1.41    13
## [9]  {Juice,
##       Pastry}                => {Coffee} 0.00504      0.765  0.00659  1.41    13
## [10] {Toast}                 => {Coffee} 0.03994      0.720  0.05545  1.32   103
```

```r
regras03_sorted <- sort(regras03, by=c("lift", "confidence"), decreasing=T)
inspect(regras03_sorted[1:10], linebreak=T)
```

```
##      lhs                          rhs      support confidence coverage  lift count
## [1]  {Postcard}                => {Tshirt} 0.00116       0.75  0.00155 241.8     3
## [2]  {Extra Salami or Feta,
##       Juice}                   => {Salad}  0.00116       1.00  0.00116  39.1     3
## [3]  {Extra Salami or Feta,
##       Sandwich}                => {Salad}  0.00116       1.00  0.00116  39.1     3
## [4]  {Cake,
##       Extra Salami or Feta}    => {Salad}  0.00116       1.00  0.00116  39.1     3
## [5]  {Bread,
##       Extra Salami or Feta}    => {Salad}  0.00233       1.00  0.00233  39.1     6
## [6]  {Coffee,
##       Extra Salami or Feta,
##       Juice}                   => {Salad}  0.00116       1.00  0.00116  39.1     3
## [7]  {Coffee,
##       Extra Salami or Feta,
##       Sandwich}                => {Salad}  0.00116       1.00  0.00116  39.1     3
## [8]  {Cake,
##       Coffee,
##       Extra Salami or Feta}    => {Salad}  0.00116       1.00  0.00116  39.1     3
## [9]  {Bread,
##       Coffee,
##       Extra Salami or Feta}    => {Salad}  0.00155       1.00  0.00155  39.1     4
## [10] {Extra Salami or Feta}    => {Salad}  0.00620       0.80  0.00775  31.3    16
```

Análises

a) Quais as regras mais interessantes geradas a partir dessa base? Justifique.

Resposta: As regras mais interessantes foram dadas pelo algoritmo Apriori considerando os valores de suporte = 0.005 e uma confiança = 0.7. Escolhemos esses valores de suporte e confiança dado que nosso conjunto de dados possuia poucos itens com valores frequentes e muitos com valores pouco frequentes e o intuito era capturar uma quantidade intermediária de regras de associação, isto é, um valor de suporte baixo para trazer mais regras e um valor de confiança alto para filtrar os casos de ocorrência. O intuito de tal ação foi capturar relacionamentos não explicítos nas nossas transações, dado que é senso comum que em uma padaria a expectativa é que Coffee e Bread sejam os itens mais consumidos, por exemplo. Classificamos as regras mais interessante como sendo as com um valor de lift elevado e uma confiança também elevada, listadas abaixo.

```
     lhs                          rhs      support confidence coverage  lift count
[1]  {Extra Salami or Feta}    => {Salad}  0.00620      0.800  0.00775 31.26    16
[2]  {Coffee,
      Extra Salami or Feta}    => {Salad}  0.00543      0.778  0.00698 30.39    14
[3]  {Extra Salami or Feta}    => {Coffee} 0.00698      0.900  0.00775  1.65    18
[4]  {Extra Salami or Feta,
      Salad}                   => {Coffee} 0.00543      0.875  0.00620  1.61    14
```

Atividade 3 – Medidas de Interesse (*3,5 pts*)

Vimos na aula que, mesmo após as podas do algoritmo `apriori`, ainda temos algumas regras com características indesejáveis como redundâncias e dependência estatística negativa. Também vimos algumas medidas que nos ajudam a analisar melhor essas regras como o lift, a convicção e a razão de chances. Nesta questão, escolha

um dos conjuntos de regras geradas na atividade anterior e o analise usando essas medidas. Compute as três medidas para o conjunto escolhido com a função `interestMeasure` e experimente ordenar as regras com cada uma das novas medidas.

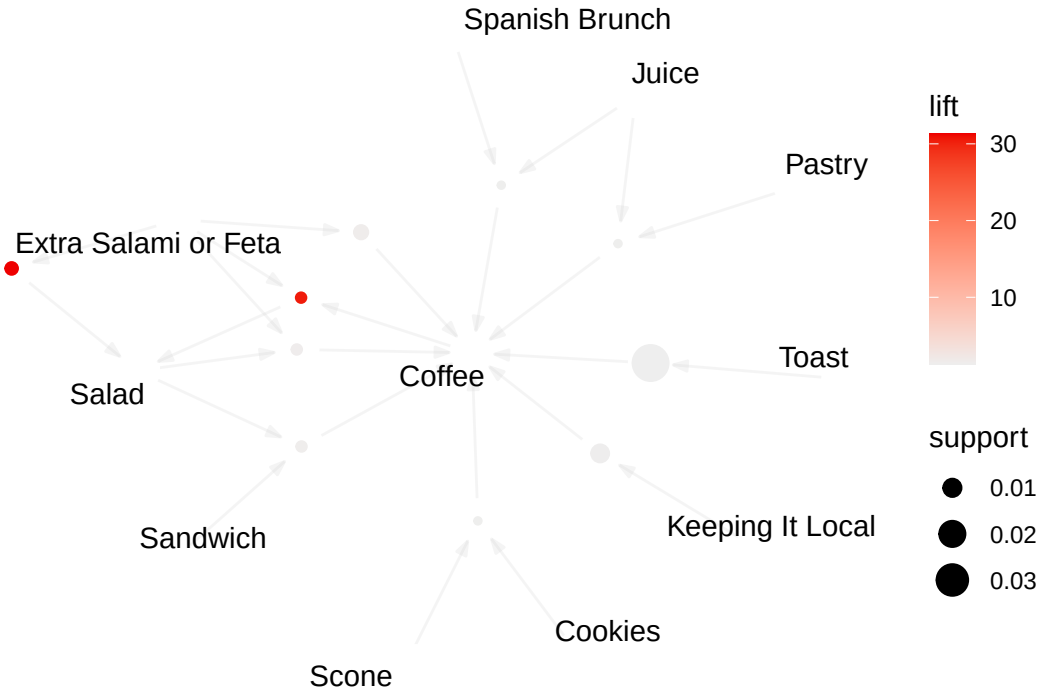
Dica: para adicionar as medidas em um conjunto de regras qualquer, você pode utilizar o comando `cbind` e a função `quality`:

```
quality(regras) <- cbind(quality(regras), interestMeasure(regras, measure=c("conviction", "oddsRatio"),
                                          transactions = transacoes))
```

```
# Compute as medidas de interesse
regras <- regras02
quality(regras) <- cbind(quality(regras),
                         interestMeasure(regras,
                                         measure=c("conviction", "oddsRatio"),
                                         transactions=bakery_data))
inspect(regras, linebreak=T)
```

```
##      lhs                          rhs      support confidence coverage  lift count conviction oddsRatio
## [1]  {Extra Salami or Feta}  => {Salad}  0.00620      0.800  0.00775 31.26    16       4.87    200.71
## [2]  {Extra Salami or Feta}  => {Coffee} 0.00698      0.900  0.00775  1.65    18       4.56      7.63
## [3]  {Keeping It Local}      => {Coffee} 0.00969      0.781  0.01241  1.44    25       2.08      3.03
## [4]  {Toast}                 => {Coffee} 0.03994      0.720  0.05545  1.32   103       1.63      2.25
## [5]  {Extra Salami or Feta,
##       Salad}                 => {Coffee} 0.00543      0.875  0.00620  1.61    14       3.65      5.91
## [6]  {Coffee,
##       Extra Salami or Feta}  => {Salad}  0.00543      0.778  0.00698 30.39    14       4.38    168.88
## [7]  {Salad,
##       Sandwich}              => {Coffee} 0.00543      0.824  0.00659  1.51    14       2.58      3.94
## [8]  {Juice,
##       Spanish Brunch}        => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [9]  {Cookies,
##       Scone}                 => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [10] {Juice,
##       Pastry}                => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
```

```
plot(regras, method= "graph")
```




```
# Apresente as regras ordenadas por lift
rlift_sorted <- sort(regras, by=c("lift"), decreasing=T)
inspect(rlift_sorted, linebreak=T)
```

```
##      lhs                          rhs      support confidence coverage  lift count conviction oddsRatio
## [1]  {Extra Salami or Feta}    => {Salad}  0.00620      0.800  0.00775 31.26    16       4.87    200.7
## [2]  {Coffee,
##       Extra Salami or Feta}    => {Salad}  0.00543      0.778  0.00698 30.39    14       4.38    168.8
## [3]  {Extra Salami or Feta}    => {Coffee} 0.00698      0.900  0.00775  1.65    18       4.56      7.6
## [4]  {Extra Salami or Feta,
##       Salad}                   => {Coffee} 0.00543      0.875  0.00620  1.61    14       3.65      5.9
## [5]  {Salad,
##       Sandwich}                => {Coffee} 0.00543      0.824  0.00659  1.51    14       2.58      3.9
## [6]  {Keeping It Local}        => {Coffee} 0.00969      0.781  0.01241  1.44    25       2.08      3.0
## [7]  {Juice,
##       Spanish Brunch}          => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.7
## [8]  {Cookies,
##       Scone}                   => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.7
## [9]  {Juice,
##       Pastry}                  => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.7
## [10] {Toast}                   => {Coffee} 0.03994      0.720  0.05545  1.32   103       1.63      2.2
```

```
# Apresente as regras ordenadas por convicção
rconvic_sorted <- sort(regras, by=c("conviction"), decreasing=T)
inspect(rconvic_sorted, linebreak=T)
```

```
##      lhs                          rhs      support confidence coverage  lift count conviction oddsRatio
## [1]  {Extra Salami or Feta}    => {Salad}  0.00620      0.800  0.00775 31.26    16       4.87    200.7
## [2]  {Extra Salami or Feta}    => {Coffee} 0.00698      0.900  0.00775  1.65    18       4.56      7.6
## [3]  {Coffee,
##       Extra Salami or Feta}    => {Salad}  0.00543      0.778  0.00698 30.39    14       4.38    168.8
## [4]  {Extra Salami or Feta,
##       Salad}                   => {Coffee} 0.00543      0.875  0.00620  1.61    14       3.65      5.9
```

```
## [5]  {Salad,
##       Sandwich}             => {Coffee} 0.00543      0.824  0.00659  1.51    14       2.58      3.94
## [6]  {Keeping It Local}     => {Coffee} 0.00969      0.781  0.01241  1.44    25       2.08      3.03
## [7]  {Juice,
##       Spanish Brunch}       => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [8]  {Cookies,
##       Scone}                => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [9]  {Juice,
##       Pastry}               => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [10] {Toast}                => {Coffee} 0.03994      0.720  0.05545  1.32   103       1.63      2.25
```

```
# Apresente as regras ordenadas por razão de chances
roddsr_sorted <- sort(regras, by=c("oddsRatio"), decreasing=T)
inspect(roddsr_sorted, linebreak=T)
```

```
##      lhs                         rhs      support confidence coverage  lift count conviction oddsRatio
## [1]  {Extra Salami or Feta}  => {Salad}  0.00620      0.800  0.00775 31.26    16       4.87    200.71
## [2]  {Coffee,
##       Extra Salami or Feta}  => {Salad}  0.00543      0.778  0.00698 30.39    14       4.38    168.88
## [3]  {Extra Salami or Feta}  => {Coffee} 0.00698      0.900  0.00775  1.65    18       4.56      7.63
## [4]  {Extra Salami or Feta,
##       Salad}                => {Coffee} 0.00543      0.875  0.00620  1.61    14       3.65      5.91
## [5]  {Salad,
##       Sandwich}             => {Coffee} 0.00543      0.824  0.00659  1.51    14       2.58      3.94
## [6]  {Keeping It Local}     => {Coffee} 0.00969      0.781  0.01241  1.44    25       2.08      3.03
## [7]  {Juice,
##       Spanish Brunch}       => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [8]  {Cookies,
##       Scone}                => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [9]  {Juice,
##       Pastry}               => {Coffee} 0.00504      0.765  0.00659  1.41    13       1.94      2.74
## [10] {Toast}                => {Coffee} 0.03994      0.720  0.05545  1.32   103       1.63      2.25
```

Análise

a) Quais as regras mais interessantes do conjunto? Justifique.

Resposta: As regras mais interessantes encontradas foram aquelas que possuiam valores altos em lift, conviction e oddsRatio, dadas abaixo:

```
    lhs                         rhs      support confidence coverage  lift count conviction oddsRatio
[1] {Extra Salami or Feta}  => {Salad}  0.00620      0.800  0.00775 31.26    16       4.87    200.72
[2] {Coffee, Extra Salami or Feta}  => {Salad}  0.00543      0.778  0.00698 30.39    14       4.38
```

Escolhemos tais regras por conta de explicitarem relacionamentos não triviais que podemos ter entre os produtos de uma padaria. Também levamos em conta o fato de que quanto maior o valor de lift mais certo é que os itens do conjunto esquerdo influenciem nos itens do conjunto direito. O mesmo conceito se aplica para a métrica OddsRatio e no caso da métrica conviction temos que valores entre 1 e 5 representam boas regras, que é o nosso caso.