Lab3

Table of contents

# 1. Grupo:

* Eduardo Amorim Vilela de Salis - 11226805
* Débora Mauricio Kono - 9896754
* Douglas Silva Cardoso - 11766990

# 2. Regras de Associação

## 2.1 Pacotes Utilizados

Loading required package: Matrix

Attaching package: 'arules'

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

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.0 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ ggplot2 3.4.1 ✔ tibble 3.2.0  
✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
✔ purrr 1.0.1   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ tidyr::expand() masks Matrix::expand()  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
✖ tidyr::pack() masks Matrix::pack()  
✖ dplyr::recode() masks arules::recode()  
✖ tidyr::unpack() masks Matrix::unpack()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors  
------------------------------------------------------------------------------  
  
You have loaded plyr after dplyr - this is likely to cause problems.  
If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
library(plyr); library(dplyr)  
  
------------------------------------------------------------------------------  
  
  
Attaching package: 'plyr'  
  
  
The following objects are masked from 'package:dplyr':  
  
 arrange, count, desc, failwith, id, mutate, rename, summarise,  
 summarize  
  
  
The following object is masked from 'package:purrr':  
  
 compact

## 2.2 Dados

retail <- read\_excel('Online Retail.xlsx')  
retail <- retail[complete.cases(retail), ]  
retail <- retail |>  
 mutate(  
 Description = as.factor(Description),  
 Country = as.factor(Country),  
 Date = as.Date(InvoiceDate),  
 TransTime = format(InvoiceDate,"%H:%M:%S"),  
 InvoiceNo = as.numeric(as.character(InvoiceNo)))

Warning in eval(cols[[col]], .data, parent.frame()): NAs introduced by coercion

glimpse(retail)

Rows: 406,829  
Columns: 10  
$ InvoiceNo <dbl> 536365, 536365, 536365, 536365, 536365, 536365, 536365, 53…  
$ StockCode <chr> "85123A", "71053", "84406B", "84029G", "84029E", "22752", …  
$ Description <fct> "WHITE HANGING HEART T-LIGHT HOLDER", "WHITE METAL LANTERN…  
$ Quantity <dbl> 6, 6, 8, 6, 6, 2, 6, 6, 6, 32, 6, 6, 8, 6, 6, 3, 2, 3, 3, …  
$ InvoiceDate <dttm> 2010-12-01 08:26:00, 2010-12-01 08:26:00, 2010-12-01 08:2…  
$ UnitPrice <dbl> 2.55, 3.39, 2.75, 3.39, 3.39, 7.65, 4.25, 1.85, 1.85, 1.69…  
$ CustomerID <dbl> 17850, 17850, 17850, 17850, 17850, 17850, 17850, 17850, 17…  
$ Country <fct> United Kingdom, United Kingdom, United Kingdom, United Kin…  
$ Date <date> 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-0…  
$ TransTime <chr> "08:26:00", "08:26:00", "08:26:00", "08:26:00", "08:26:00"…

#retail |> glimpse()  
transactionData <- ddply(retail,c("InvoiceNo","Date"),function(df1)paste(df1$Description, collapse = ","))  
transactionData$InvoiceNo <- NULL  
transactionData$Date <- NULL  
colnames(transactionData) <- c("items")  
  
  
transactionData |> write.csv("market\_basket\_transactions.csv", quote = FALSE,row.names = FALSE)

### 2.2.1 Transações

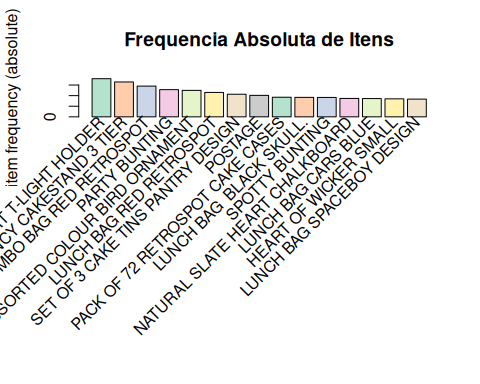
**COMENTÁRIOS**: Nesta etapa o número de transações e item difere do que foi apresentado no documento. Obtivemos um número menor de transações e um número bem menor de items. Isso faz com que os próximos resultados sejam um pouco diferentes. Deixamos os códigos para mostrar que o pré-processamento foi igual

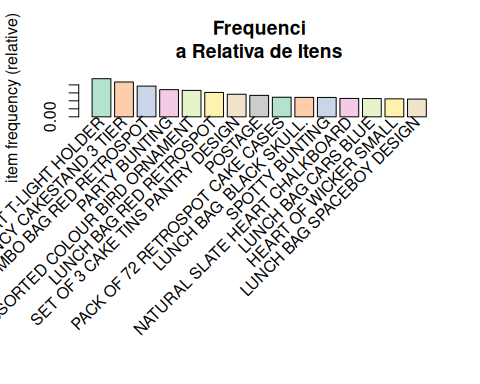
transactions in sparse format with  
 18839 transactions (rows) and  
 7887 items (columns)

**COMENTÁRIOS**: A densidade foi um pocuo maior da apresentada no documento do laboratório. Mas os demais resultados são praticamente iguais

transactions as itemMatrix in sparse format with  
 18839 rows (elements/itemsets/transactions) and  
 7887 columns (items) and a density of 0.002260834   
  
most frequent items:  
WHITE HANGING HEART T-LIGHT HOLDER REGENCY CAKESTAND 3 TIER   
 1798 1644   
 JUMBO BAG RED RETROSPOT PARTY BUNTING   
 1450 1282   
 ASSORTED COLOUR BIRD ORNAMENT (Other)   
 1249 328499   
  
element (itemset/transaction) length distribution:  
sizes  
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16   
1578 867 762 773 768 721 660 652 648 586 621 532 510 532 555 525   
 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32   
 470 442 483 425 396 319 310 276 241 255 230 218 223 215 173 163   
 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48   
 143 146 139 112 118 89 117 96 97 89 93 67 66 68 65 61   
 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64   
 64 53 67 43 42 50 43 37 31 40 30 27 28 18 26 25   
 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80   
 20 27 25 25 15 20 20 13 16 16 12 16 12 7 9 14   
 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96   
 15 12 8 9 11 11 14 8 6 5 6 12 6 4 4 3   
 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112   
 6 5 2 4 2 5 4 3 2 2 6 3 4 3 2 1   
 113 114 116 117 118 120 121 122 123 125 126 127 131 132 133 134   
 3 1 4 3 3 1 2 2 1 3 2 2 1 1 2 1   
 140 141 142 143 145 146 147 150 154 157 168 171 177 178 180 202   
 1 2 2 1 1 2 1 1 3 2 2 2 1 1 1 1   
 204 228 236 249 250 285 320 400 419   
 1 1 1 1 1 1 1 1 1   
  
 Min. 1st Qu. Median Mean 3rd Qu. Max.   
 1.00 5.00 13.00 17.83 23.00 419.00   
  
includes extended item information - examples:  
 labels  
1 1 HANGER  
2 10 COLOUR SPACEBOY PEN  
3 12 COLOURED PARTY BALLOONS

Loading required package: RColorBrewer





**COMENTÁRIOS**: Itens mais frequentes devem ser observados com maior cuidado pelo varejista. No entando , é importante ressaltar que o varejista pode querer observar essas informações em uma janela mais recente das vendas. Pressupõe-se que o padrão de compras não mudou ao longo do 1 ano que se tem registro na base

[1] "Data mínima: "

[1] "2010-12-01"

[1] "Data máxima: "

[1] "2011-12-09"

# A tibble: 305 × 2  
 Date n  
 <date> <int>  
 1 2010-12-01 1968  
 2 2010-12-02 2044  
 3 2010-12-03 1117  
 4 2010-12-05 2724  
 5 2010-12-06 1974  
 6 2010-12-07 1133  
 7 2010-12-08 2021  
 8 2010-12-09 1822  
 9 2010-12-10 1449  
10 2010-12-12 1448  
# ℹ 295 more rows

### 2.2.2 Gerando Regras

**COMENTÁRIOS**: Aumentamos o suporte para 0.002, isso alterou bastante o número de regras criadas. Enquanto para um suporte de 0.001 temos 116493 , enquanto para um suporte de 0.002 temos bem menos regras (8280)

summary(apriori(tr, parameter = list(supp=0.001, conf=0.8,maxlen=10)))

Apriori  
  
Parameter specification:  
 confidence minval smax arem aval originalSupport maxtime support minlen  
 0.8 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: 18   
  
set item appearances ...[0 item(s)] done [0.00s].  
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.06s].  
sorting and recoding items ... [2455 item(s)] done [0.01s].  
creating transaction tree ... done [0.00s].  
checking subsets of size 1 2 3 4 5 6 7 8 9 10

Warning in apriori(tr, parameter = list(supp = 0.001, conf = 0.8, maxlen =  
10)): Mining stopped (maxlen reached). Only patterns up to a length of 10  
returned!

done [0.18s].  
writing ... [116493 rule(s)] done [0.03s].  
creating S4 object ... done [0.02s].

set of 116493 rules  
  
rule length distribution (lhs + rhs):sizes  
 2 3 4 5 6 7 8 9 10   
 111 3378 10947 29980 39875 23872 6860 1249 221   
  
 Min. 1st Qu. Median Mean 3rd Qu. Max.   
 2.000 5.000 6.000 5.826 7.000 10.000   
  
summary of quality measures:  
 support confidence coverage lift   
 Min. :0.001009 Min. :0.8000 Min. :0.001009 Min. : 8.382   
 1st Qu.:0.001062 1st Qu.:0.8333 1st Qu.:0.001168 1st Qu.: 18.897   
 Median :0.001168 Median :0.8750 Median :0.001327 Median : 23.917   
 Mean :0.001323 Mean :0.8870 Mean :0.001499 Mean : 48.813   
 3rd Qu.:0.001380 3rd Qu.:0.9310 3rd Qu.:0.001592 3rd Qu.: 39.552   
 Max. :0.022453 Max. :1.0000 Max. :0.027921 Max. :607.710   
 count   
 Min. : 19.00   
 1st Qu.: 20.00   
 Median : 22.00   
 Mean : 24.92   
 3rd Qu.: 26.00   
 Max. :423.00   
  
mining info:  
 data ntransactions support confidence  
 tr 18839 0.001 0.8  
 call  
 apriori(data = tr, parameter = list(supp = 0.001, conf = 0.8, maxlen = 10))

association.rules <- apriori(tr, parameter = list(supp=0.002, conf=0.8,maxlen=10))

Apriori  
  
Parameter specification:  
 confidence minval smax arem aval originalSupport maxtime support minlen  
 0.8 0.1 1 none FALSE TRUE 5 0.002 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: 37   
  
set item appearances ...[0 item(s)] done [0.00s].  
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.07s].  
sorting and recoding items ... [1920 item(s)] done [0.01s].  
creating transaction tree ... done [0.00s].  
checking subsets of size 1 2 3 4 5 6 7 8 9 done [0.06s].  
writing ... [8280 rule(s)] done [0.01s].  
creating S4 object ... done [0.00s].

summary(association.rules)

set of 8280 rules  
  
rule length distribution (lhs + rhs):sizes  
 2 3 4 5 6 7 8 9   
 95 688 1810 2261 2000 1161 256 9   
  
 Min. 1st Qu. Median Mean 3rd Qu. Max.   
 2.0 4.0 5.0 5.2 6.0 9.0   
  
summary of quality measures:  
 support confidence coverage lift   
 Min. :0.002017 Min. :0.8000 Min. :0.002017 Min. : 8.456   
 1st Qu.:0.002123 1st Qu.:0.8333 1st Qu.:0.002442 1st Qu.: 18.907   
 Median :0.002389 Median :0.8711 Median :0.002707 Median : 22.586   
 Mean :0.002706 Mean :0.8774 Mean :0.003094 Mean : 43.382   
 3rd Qu.:0.002813 3rd Qu.:0.9130 3rd Qu.:0.003291 3rd Qu.: 60.085   
 Max. :0.022453 Max. :1.0000 Max. :0.027921 Max. :495.763   
 count   
 Min. : 38.00   
 1st Qu.: 40.00   
 Median : 45.00   
 Mean : 50.98   
 3rd Qu.: 53.00   
 Max. :423.00   
  
mining info:  
 data ntransactions support confidence  
 tr 18839 0.002 0.8  
 call  
 apriori(data = tr, parameter = list(supp = 0.002, conf = 0.8, maxlen = 10))

lhs rhs support confidence coverage lift count  
[1] {ART LIGHTS} => {FUNK MONKEY} 0.002017092 1.0000000 0.002017092 495.76316 38  
[2] {FUNK MONKEY} => {ART LIGHTS} 0.002017092 1.0000000 0.002017092 495.76316 38  
[3] {CHOCOLATE SPOTS} => {SWISS ROLL TOWEL} 0.002282499 1.0000000 0.002282499 392.47917 43  
[4] {SWISS ROLL TOWEL} => {CHOCOLATE SPOTS} 0.002282499 0.8958333 0.002547906 392.47917 43  
[5] {BLACK TEA} => {SUGAR JARS} 0.002441743 1.0000000 0.002441743 204.77174 46  
[6] {BLACK TEA} => {COFFEE} 0.002441743 1.0000000 0.002441743 59.42902 46  
[7] {FRENCH BLUE METAL DOOR SIGN 0} => {FRENCH BLUE METAL DOOR SIGN 2} 0.002707150 0.8225806 0.003291045 154.96597 51  
[8] {FRENCH BLUE METAL DOOR SIGN 0} => {FRENCH BLUE METAL DOOR SIGN 1} 0.002813313 0.8548387 0.003291045 135.33031 53  
[9] {GREEN 3 PIECE POLKADOT CUTLERY SET} => {RED 3 PIECE RETROSPOT CUTLERY SET} 0.002494825 0.8103448 0.003078720 177.51263 47  
[10] {NURSERY A} => {B} 0.002919476 1.0000000 0.002919476 342.52727 55

### 2.2.3 Limitar o número e o tamanho das regras

**COMENTÁRIOS**: Aqui fomos ainda mais agressivos , queremos apenas aqueles que possuem 95% de confiança . Para compensar aumentamos o max len para 5

shorter.association.rules <- apriori(tr, parameter = list(supp=0.001,  
conf=0.95,maxlen=5))

Apriori  
  
Parameter specification:  
 confidence minval smax arem aval originalSupport maxtime support minlen  
 0.95 0.1 1 none FALSE TRUE 5 0.001 1  
 maxlen target ext  
 5 rules TRUE  
  
Algorithmic control:  
 filter tree heap memopt load sort verbose  
 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
  
Absolute minimum support count: 18   
  
set item appearances ...[0 item(s)] done [0.00s].  
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.06s].  
sorting and recoding items ... [2455 item(s)] done [0.01s].  
creating transaction tree ... done [0.00s].  
checking subsets of size 1 2 3 4 5

Warning in apriori(tr, parameter = list(supp = 0.001, conf = 0.95, maxlen =  
5)): Mining stopped (maxlen reached). Only patterns up to a length of 5  
returned!

done [0.16s].  
writing ... [6163 rule(s)] done [0.02s].  
creating S4 object ... done [0.01s].

inspect(shorter.association.rules[1:10])

lhs rhs support confidence  
[1] {WOBBLY CHICKEN} => {METAL} 0.001486278 1   
[2] {WOBBLY CHICKEN} => {DECORATION} 0.001486278 1   
[3] {DECOUPAGE} => {GREETING CARD} 0.001220872 1   
[4] {BILLBOARD FONTS DESIGN} => {WRAP} 0.001539360 1   
[5] {WOBBLY RABBIT} => {METAL} 0.001804767 1   
[6] {WOBBLY RABBIT} => {DECORATION} 0.001804767 1   
[7] {ART LIGHTS} => {FUNK MONKEY} 0.002017092 1   
[8] {FUNK MONKEY} => {ART LIGHTS} 0.002017092 1   
[9] {CHOCOLATE SPOTS} => {SWISS ROLL TOWEL} 0.002282499 1   
[10] {BLACK TEA} => {SUGAR JARS} 0.002441743 1   
 coverage lift count  
[1] 0.001486278 376.7800 28   
[2] 0.001486278 376.7800 28   
[3] 0.001220872 330.5088 23   
[4] 0.001539360 607.7097 29   
[5] 0.001804767 376.7800 34   
[6] 0.001804767 376.7800 34   
[7] 0.002017092 495.7632 38   
[8] 0.002017092 495.7632 38   
[9] 0.002282499 392.4792 43   
[10] 0.002441743 204.7717 46

### 2.2.4 Removendo Regras Redundantes

[1] 7159

[1] 1121

[1] 6163

[1] 2482

### 2.2.5 **10 Regras (após subset e remover redundâncias) de tamanho igual a 3**

lhs rhs support confidence coverage lift count  
[1] {CHILDS GARDEN RAKE BLUE,   
 CHILDS GARDEN SPADE PINK} => {CHILDS GARDEN SPADE BLUE} 0.001327034 1 0.001327034 265.33803 25  
[2] {SWISS CHALET TREE DECORATION,   
 WOODEN STAR CHRISTMAS SCANDINAVIAN} => {WOODEN HEART CHRISTMAS SCANDINAVIAN} 0.001061627 1 0.001061627 49.83862 20  
[3] {CHILDS GARDEN BRUSH BLUE,   
 CHILDS GARDEN SPADE PINK} => {CHILDS GARDEN SPADE BLUE} 0.001167790 1 0.001167790 265.33803 22  
[4] {REGENCY CAKE FORK,   
 REGENCY TEA PLATE GREEN} => {REGENCY TEA PLATE ROSES} 0.001114709 1 0.001114709 62.58804 21  
[5] {DOLLY GIRL CHILDRENS CUP,   
 POSTAGE} => {DOLLY GIRL CHILDRENS BOWL} 0.001008546 1 0.001008546 172.83486 19  
[6] {CHRISTMAS GINGHAM TREE,   
 CHRISTMAS RETROSPOT STAR WOOD} => {CHRISTMAS GINGHAM STAR} 0.001008546 1 0.001008546 174.43519 19  
[7] {GLITTER HEART DECORATION,   
 WOODEN TREE CHRISTMAS SCANDINAVIAN} => {WOODEN STAR CHRISTMAS SCANDINAVIAN} 0.001008546 1 0.001008546 54.76453 19  
[8] {GLITTER HEART DECORATION,   
 WOODEN TREE CHRISTMAS SCANDINAVIAN} => {WOODEN HEART CHRISTMAS SCANDINAVIAN} 0.001008546 1 0.001008546 49.83862 19  
[9] {LUNCH BAG DOILEY PATTERN,   
 PINK VINTAGE SPOT BEAKER} => {BLUE VINTAGE SPOT BEAKER} 0.001008546 1 0.001008546 131.74126 19  
[10] {BLUE GIANT GARDEN THERMOMETER,   
 HERB MARKER ROSEMARY} => {HERB MARKER BASIL} 0.001008546 1 0.001008546 112.80838 19

### 2.2.6 **10 Regras (após subset e remover redundâncias) de tamanho igual a 3**

lhs rhs support confidence coverage lift count  
[1] {PARTY PIZZA DISH BLUE POLKADOT,   
 PARTY PIZZA DISH GREEN POLKADOT} => {PARTY PIZZA DISH PINK POLKADOT} 0.001114709 0.9545455 0.001167790 438.60200 21  
[2] {DRAWER KNOB CERAMIC RED,   
 RED STRIPE CERAMIC DRAWER KNOB} => {BLUE STRIPE CERAMIC DRAWER KNOB} 0.001167790 0.9565217 0.001220872 55.96246 22  
[3] {FRENCH BLUE METAL DOOR SIGN 0,   
 FRENCH BLUE METAL DOOR SIGN 9} => {FRENCH BLUE METAL DOOR SIGN 7} 0.001857848 0.9722222 0.001910929 257.96753 35  
[4] {FRENCH BLUE METAL DOOR SIGN 0,   
 FRENCH BLUE METAL DOOR SIGN 9} => {FRENCH BLUE METAL DOOR SIGN 2} 0.001857848 0.9722222 0.001910929 183.15694 35  
[5] {FRENCH BLUE METAL DOOR SIGN 6,   
 FRENCH BLUE METAL DOOR SIGN 9} => {FRENCH BLUE METAL DOOR SIGN 7} 0.001857848 0.9722222 0.001910929 257.96753 35  
[6] {FRENCH BLUE METAL DOOR SIGN 5,   
 FRENCH BLUE METAL DOOR SIGN 9} => {FRENCH BLUE METAL DOOR SIGN 7} 0.002070174 0.9750000 0.002123255 258.70458 39  
[7] {FRENCH BLUE METAL DOOR SIGN 3,   
 FRENCH BLUE METAL DOOR SIGN 9} => {FRENCH BLUE METAL DOOR SIGN 1} 0.002017092 0.9743590 0.002070174 154.25167 38  
[8] {FRENCH BLUE METAL DOOR SIGN 0,   
 FRENCH BLUE METAL DOOR SIGN 7} => {FRENCH BLUE METAL DOOR SIGN 2} 0.002070174 0.9750000 0.002123255 183.68025 39  
[9] {FRENCH BLUE METAL DOOR SIGN 0,   
 FRENCH BLUE METAL DOOR SIGN 7} => {FRENCH BLUE METAL DOOR SIGN 1} 0.002070174 0.9750000 0.002123255 154.35315 39  
[10] {FRENCH BLUE METAL DOOR SIGN 0,   
 FRENCH BLUE METAL DOOR SIGN 5} => {FRENCH BLUE METAL DOOR SIGN 2} 0.002123255 0.9523810 0.002229418 179.41905 40

### 2.2.7 Regras (após subset e remover redundâncias) que possuem o item de maior frequência

lhs rhs support confidence coverage lift count  
[1] {GLASS STAR FROSTED T-LIGHT HOLDER,   
 KNITTED UNION FLAG HOT WATER BOTTLE} => {WHITE HANGING HEART T-LIGHT HOLDER} 0.001008546 0.95 0.001061627 9.953865 19  
[2] {KNITTED UNION FLAG HOT WATER BOTTLE,   
 SET 7 BABUSHKA NESTING BOXES} => {WHITE HANGING HEART T-LIGHT HOLDER} 0.001008546 1.00 0.001008546 10.477753 19  
[3] {CANDLEHOLDER PINK HANGING HEART,   
 HEART IVORY TRELLIS LARGE} => {WHITE HANGING HEART T-LIGHT HOLDER} 0.001008546 0.95 0.001061627 9.953865 19

### 2.2.8 Encontrando Regras Relacionadas a Itens Fornecidos

* Metal

Apriori  
  
Parameter specification:  
 confidence minval smax arem aval originalSupport maxtime support minlen  
 0.8 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: 18   
  
set item appearances ...[1 item(s)] done [0.00s].  
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.06s].  
sorting and recoding items ... [2455 item(s)] done [0.01s].  
creating transaction tree ... done [0.00s].  
checking subsets of size 1 2 3 4 5 6 7 8 9 10

Warning in apriori(tr, parameter = list(supp = 0.001, conf = 0.8), appearance =  
list(default = "lhs", : Mining stopped (maxlen reached). Only patterns up to a  
length of 10 returned!

done [0.31s].  
writing ... [5 rule(s)] done [0.04s].  
creating S4 object ... done [0.01s].

lhs rhs support confidence coverage   
[1] {WOBBLY CHICKEN} => {METAL} 0.001486278 1 0.001486278  
[2] {WOBBLY RABBIT} => {METAL} 0.001804767 1 0.001804767  
[3] {DECORATION} => {METAL} 0.002654069 1 0.002654069  
[4] {DECORATION, WOBBLY CHICKEN} => {METAL} 0.001486278 1 0.001486278  
[5] {DECORATION, WOBBLY RABBIT} => {METAL} 0.001804767 1 0.001804767  
 lift count  
[1] 376.78 28   
[2] 376.78 34   
[3] 376.78 50   
[4] 376.78 28   
[5] 376.78 34

* Sugar

Apriori  
  
Parameter specification:  
 confidence minval smax arem aval originalSupport maxtime support minlen  
 0.95 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: 18   
  
set item appearances ...[1 item(s)] done [0.00s].  
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.06s].  
sorting and recoding items ... [2455 item(s)] done [0.01s].  
creating transaction tree ... done [0.00s].  
checking subsets of size 1 2 3 4 5 6 7 8 9 10

Warning in apriori(tr, parameter = list(supp = 0.001, conf = 0.95), appearance  
= list(default = "lhs", : Mining stopped (maxlen reached). Only patterns up to  
a length of 10 returned!

done [0.22s].  
writing ... [97 rule(s)] done [0.04s].  
creating S4 object ... done [0.01s].

lhs rhs support confidence coverage lift count  
[1] {SET 3 RETROSPOT TEA} => {SUGAR} 0.012261797 1 0.012261797 81.55411 231  
[2] {RED RETROSPOT MUG,   
 SET 3 RETROSPOT TEA} => {SUGAR} 0.001061627 1 0.001061627 81.55411 20  
[3] {RED RETROSPOT SUGAR JAM BOWL,   
 SET 3 RETROSPOT TEA} => {SUGAR} 0.001751685 1 0.001751685 81.55411 33  
[4] {RED RETROSPOT BUTTER DISH,   
 SET 3 RETROSPOT TEA} => {SUGAR} 0.002017092 1 0.002017092 81.55411 38  
[5] {BREAD BIN DINER STYLE RED,   
 SET 3 RETROSPOT TEA} => {SUGAR} 0.001061627 1 0.001061627 81.55411 20  
[6] {RED SPOTTY BISCUIT TIN,   
 SET 3 RETROSPOT TEA} => {SUGAR} 0.003609533 1 0.003609533 81.55411 68

* Coffe

Apriori  
  
Parameter specification:  
 confidence minval smax arem aval originalSupport maxtime support minlen  
 0.95 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: 18   
  
set item appearances ...[1 item(s)] done [0.00s].  
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.06s].  
sorting and recoding items ... [2455 item(s)] done [0.01s].  
creating transaction tree ... done [0.00s].  
checking subsets of size 1 2 3 4 5 6 7 8 9 10

Warning in apriori(tr, parameter = list(supp = 0.001, conf = 0.95), appearance  
= list(default = "lhs", : Mining stopped (maxlen reached). Only patterns up to  
a length of 10 returned!

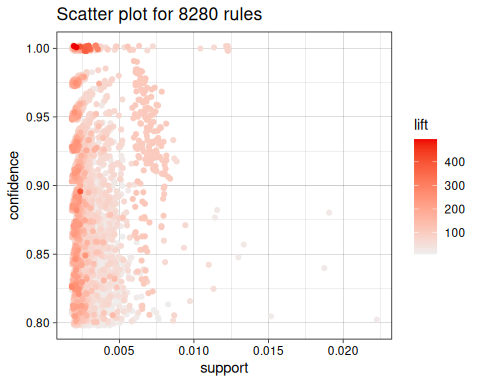
done [0.21s].  
writing ... [135 rule(s)] done [0.04s].  
creating S4 object ... done [0.01s].

lhs rhs support confidence coverage   
[1] {BLACK TEA} => {COFFEE} 0.002441743 1 0.002441743  
[2] {WHITE TEA} => {COFFEE} 0.003397208 1 0.003397208  
[3] {SUGAR JARS} => {COFFEE} 0.004883486 1 0.004883486  
[4] {SET 3 RETROSPOT TEA} => {COFFEE} 0.012261797 1 0.012261797  
[5] {SUGAR} => {COFFEE} 0.012261797 1 0.012261797  
[6] {BLACK TEA, SUGAR JARS} => {COFFEE} 0.002441743 1 0.002441743  
 lift count  
[1] 59.42902 46   
[2] 59.42902 64   
[3] 59.42902 92   
[4] 59.42902 231   
[5] 59.42902 231   
[6] 59.42902 46

## 2.3 Visualizando Regras de Associação

### 2.3.1 Gráfico de Dispersão

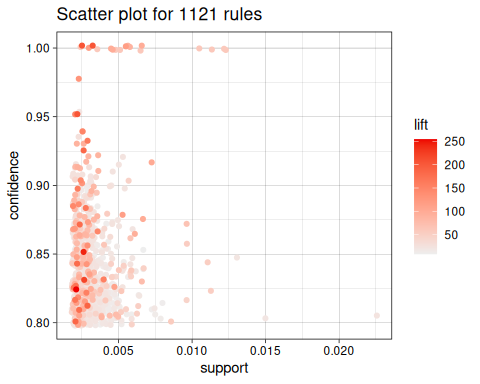
To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



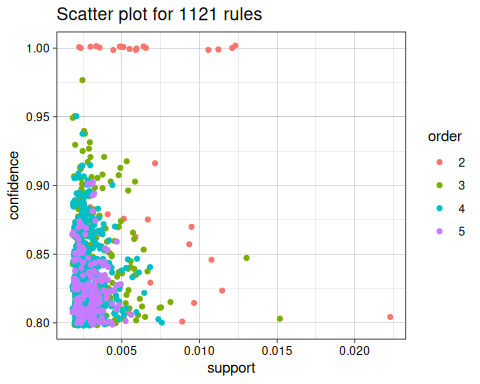
To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.

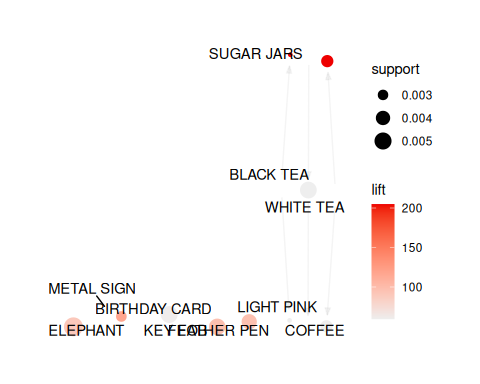


To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.

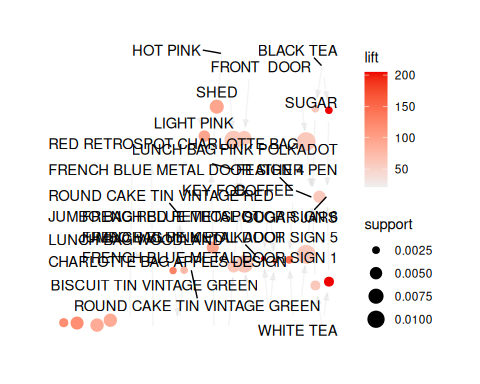


## 2.4 Visualizações Baseadas em Grafos

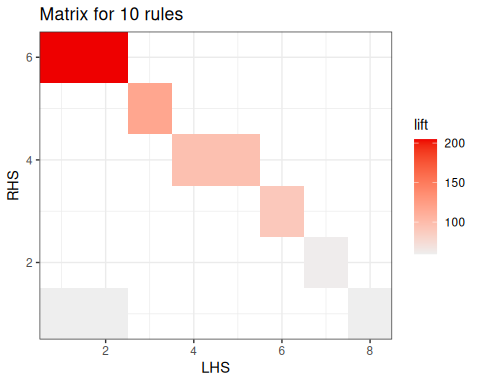
Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider  
increasing max.overlaps



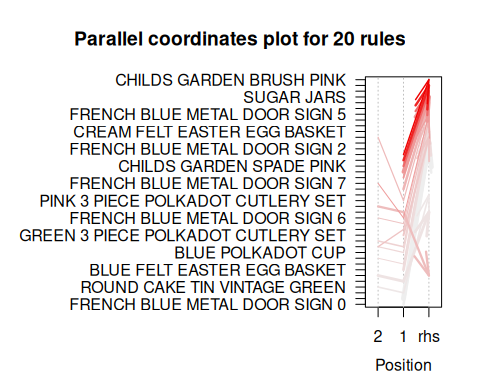
Warning: ggrepel: 15 unlabeled data points (too many overlaps). Consider  
increasing max.overlaps



Itemsets in Antecedent (LHS)  
[1] "{BLACK TEA}" "{WHITE TEA}" "{CUPCAKE SINGLE HOOK}"  
[4] "{COAL BLACK}" "{LIGHT PINK}" "{ELEPHANT}"   
[7] "{GARAGE DESIGN}" "{SUGAR JARS}"   
Itemsets in Consequent (RHS)  
[1] "{COFFEE}" "{KEY FOB}" "{BIRTHDAY CARD}" "{FEATHER PEN}"   
[5] "{METAL SIGN}" "{SUGAR JARS}"



## 2.5 Representação de Regra Individual



# 3. Conclusões

Neste laboratório, trabalhamos utilizando algoritmos de regras de associação, em específico o “APRIORI”. Esse tipo de modelo é amplamente utilizado na mineração de dados para descobrir padrões de associação em conjuntos de transações/cestas de produtos (inclusive para juntar tópicos de palavras). Esses algoritmos procuram identificar relações entre itens frequentemente coocorrentes, permitindo a extração de informações valiosas para tomada de decisões de negócios. Sua aplicação permite a extração de informações relevantes para apoiar a tomada de decisões de negócios em várias áreas, proporcionando oportunidades de otimização e melhoria de resultados.

Com base em regras de associação identificadas pelo APRIORI, uma empresa de varejo pode criar promoções personalizadas para incentivar a compra conjunta de produtos frequentemente associados, aumentando as vendas e a satisfação do cliente. Além disso, o APRIORI pode auxiliar na identificação de grupos de clientes com características semelhantes, permitindo a criação de campanhas de marketing direcionadas e eficientes. Para além do marketing, o APRIORI poderia auxiliar um processo de categorização de produtos com base nas palavras que aparecem em seus títulos, automatizando um processo que geralmente é inteiramente manual e subjetivo. Pensando em finanças, essa potente ferramenta pode ser utilizada para analisar o portfólio de gestores de investimento, identificando ativos que geralmente são utilizados em conjunto para implantar estratégias de investimento.

O conjunto de dados “Online-Retail” do UCI Machine Learning Repository oferece uma oportunidade interessante de explorar a análise de vendas e o comportamento dos clientes em um ambiente de varejo online. Ao aplicar técnicas de mineração de dados e análises adequadas, é possível obter insights valiosos para aprimorar as estratégias de negócios e impulsionar o sucesso da empresa. Foi possível encontrar algumas relações interessantes:

* Produtos de uma mesma linha “FRENCH BLUE METAL DOOR” são geralmente comprados em conjunto
* Itens de decoração são compratdos com itens de metal
* Compras de ART LIGHT E FUNK MONKEY estão bem relacionadas
* Apesar de obvio, o algorítmo foi capaz de encontrar relações em
  + produtos natalinos (CHRISTMAS GINGHAM TREE,WOODEN STAR CHRISTMAS SCANDINAVIAN,CHRISTMAS GINGHAM TREE)
  + Produtos relacioandos a chá e café (xírcares, talheres, pires, jarras para açúcar)
    - O Café inclusive é comprado como consequência de CHÁ