Project Report G2

Group members:

1) Abdelrahman Abdallah Hassan

ID: 20221372522 Task: writing the report

2) Mohamed Gomaa Mostafa

ID: 20221324034 Task: association rules

3) Ahmed Yasser Battour

ID: 20221447255 Task: K-means

4) Youssef George Abdou

ID:20221372709 Task: data visualization

5) Marwan Ahmed Elsheikh

ID: 20221324034 Task: data visualization

Problem Description :

The problem was that a big amount of data about the goods that people bought from grocery stores around Egypt needed to be cleaned and processed to get useful information

Questions we needed to ask:-

a-what will the program do

b-what the input to the program will be

c-what the output from the program will be

A-What will the program do?

1.Clean the data:

by deleting any redundant data or any unnecessary column and to check for any data that isn't in it's place

2.Data visualization:

In part (i) we compare between the total money spent by people using cash and people using credit in their payment

From this we conclude which payment type is more likely to be used

In part(ii) we compare money spent by each age group

From this we conclude which age is more likely to buy from the grocery stores

In part(iii) we compare money spent by each city and arrange them from the greatest to the lowest

From this we conclude which city spends the most money in grocery stores

In part(iv):

We display and distribute the total spending

In part (c) we display the plots in one page all together

3.K-means by using total spending and the age:

We group the to different groups and every group has it's unique attribute

4.Association rule:

We create rules to know what is most probable item to be bought with another item

B-what the input to the program will be?

The input is number of: Clusters, min Support and min Confidence.

C- what the output from the program will be?

1.graphs.

2.grouping the data and elements in each group that has same attributes.

3. discover interesting relationships using Apriori algorithm

Data set description

This dataset is about a store that has many branches all over Egypt.

This data set includes 8 columns which are (items, age ,city, count, total spending, payment type, rnd, customer)

A)

```
library(dplyr)
library(arules)
library(ggplot2)
library(gtools)
library(states)
p<-read.csv("D:/grc.csv",stringsAsFactors = FALSE
str(p)
p$count<- NULL
p$rnd<-NULL
print(p)
```

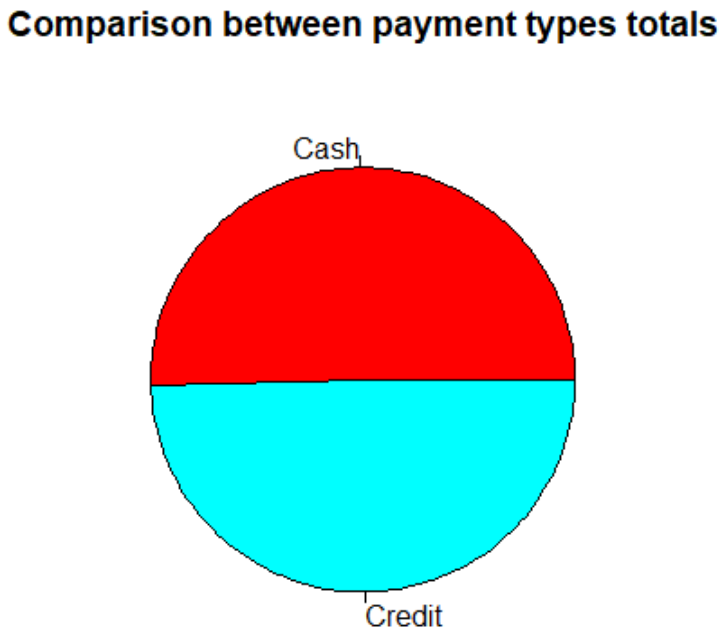
We read the data from the csv file and remove the columns count and rnd as they are not used in our program and we use str function to check on the data types of each column

B)

i)

```
namesx<-group_by(p,paymentType)
l<-summarise(x,TOtal=sum(total))
print(l)
pie(l$TOtal,main = "Comparison between payment types totals",labels = l$paymentType,col =rainbow(2))
```

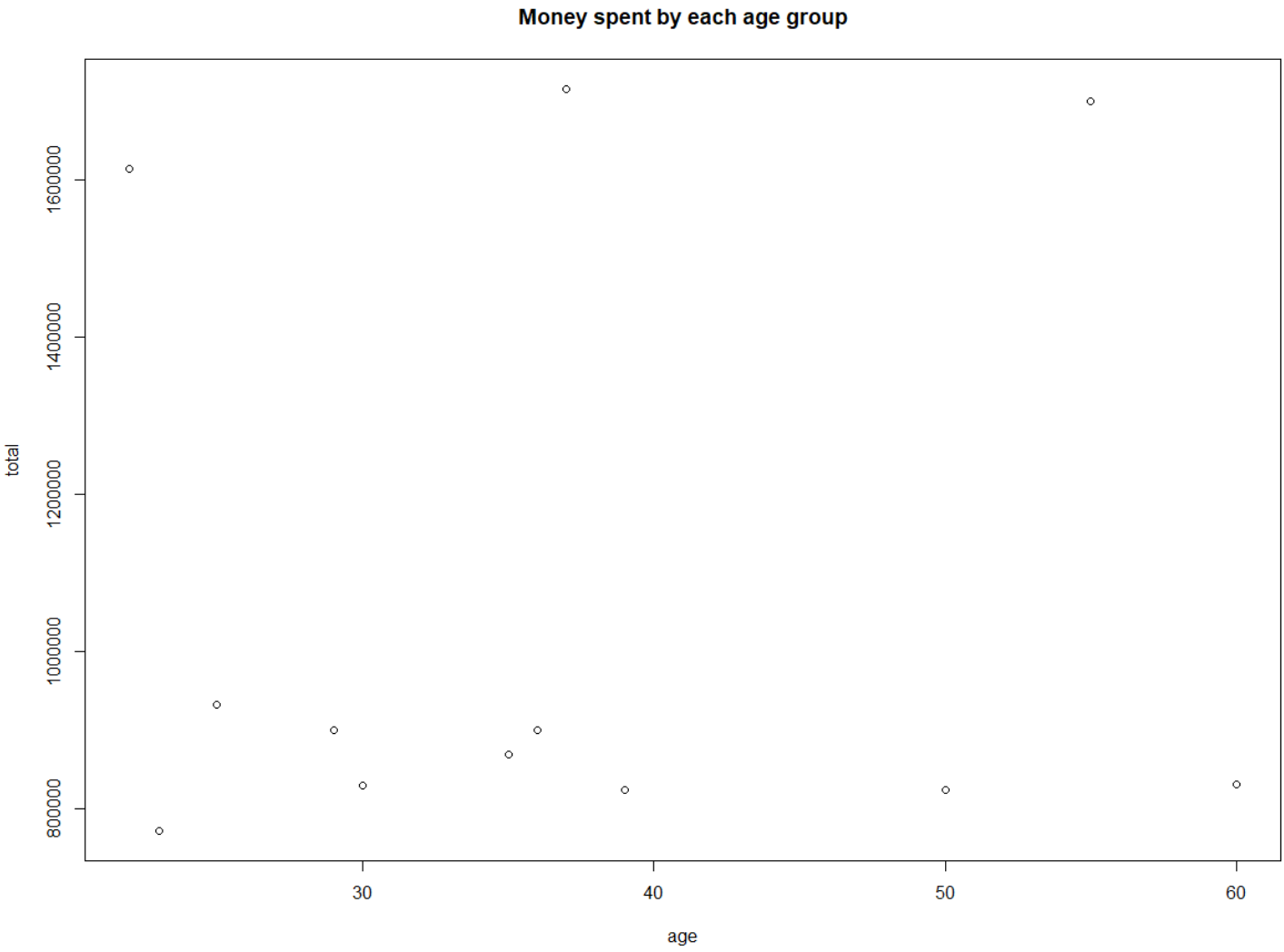
We use group_by function to give the sum of spendings by each payment type and display it in pie chart and we colour it to have two distinct parts in the pie



ii)

```
15  e<-group_by(p,age)
16  s<-summarise(e,total=sum(total))
17  print(s)
18  plot(s,main="Money spent by each age group")
```

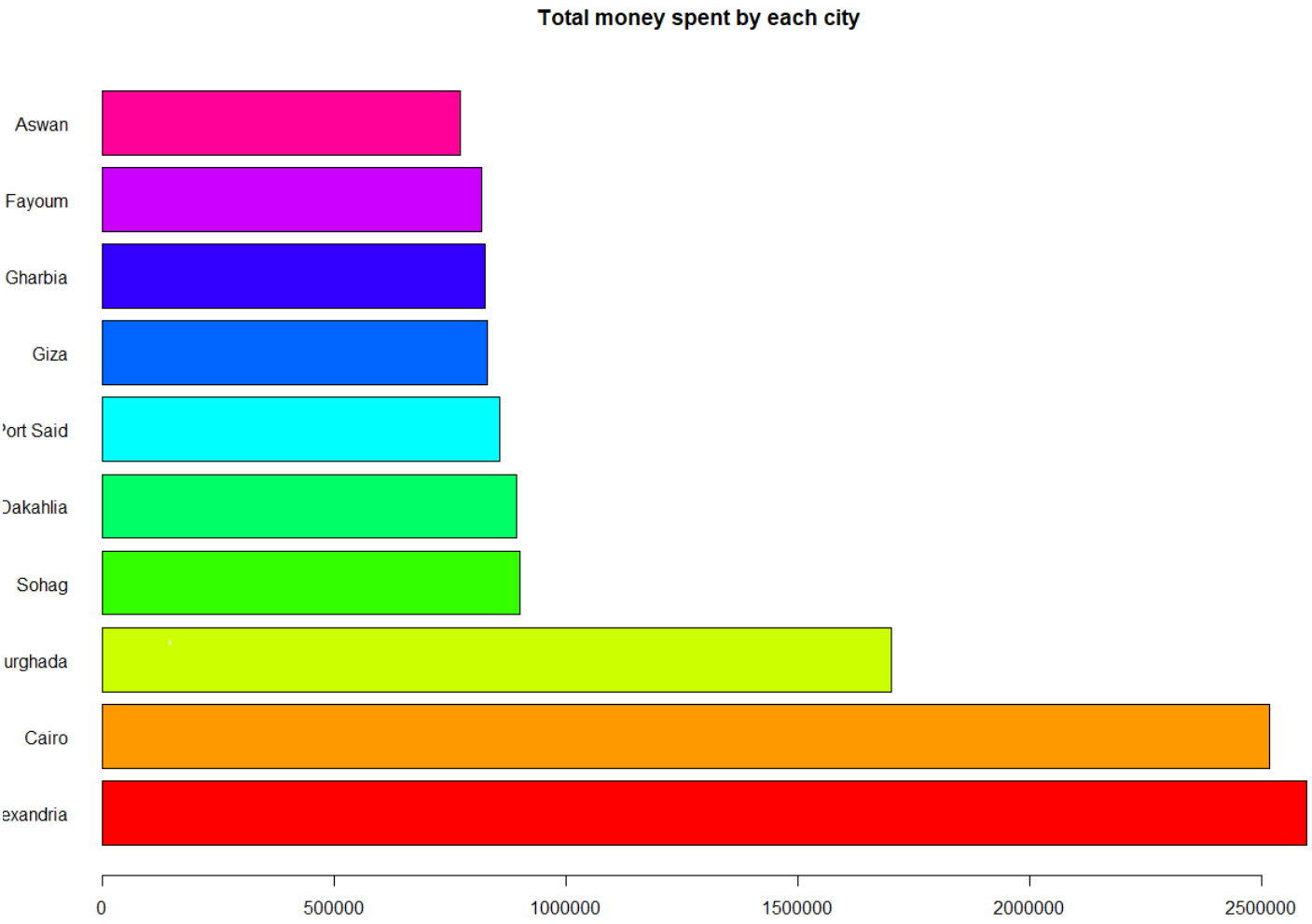
We use group_by function to sum the total spending of each age group and display it in a scatter plot



iii)

```
19  f<-group_by(p,city)
20  z<-summarise(f,TOTAL=sum(total))
21  print(z)
22  df <- z[order(z$TOTAL,decreasing = TRUE),]
23  barplot(df$TOTAL,names.arg = df$city,horiz = TRUE,las=1,main="Total money spent by each city", col = rainbow(10))
```

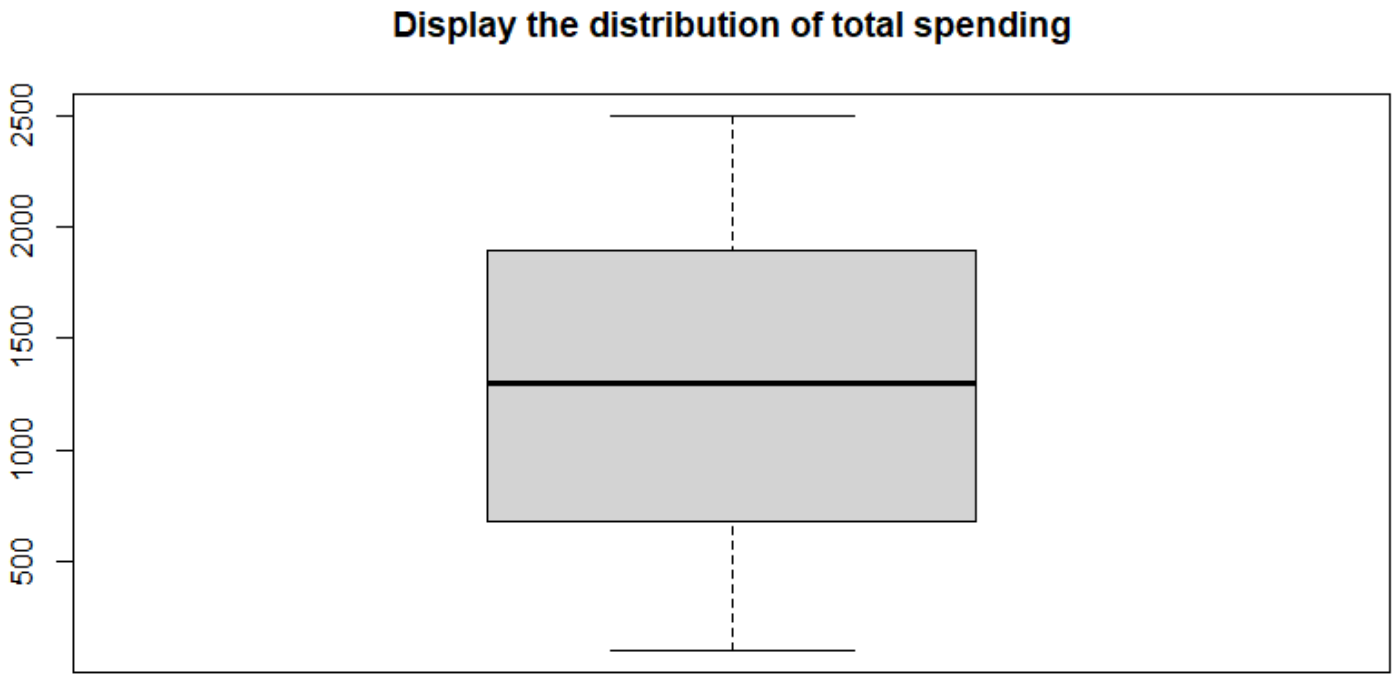
we use group_by function to count the total spending of each city and we order in using descending order while displaying it in a bar plot



iv)

```
24  boxplot(
25    x=p$total,
26    main="Display the distribution of total spending",
27    xlap="total"
28  )
```

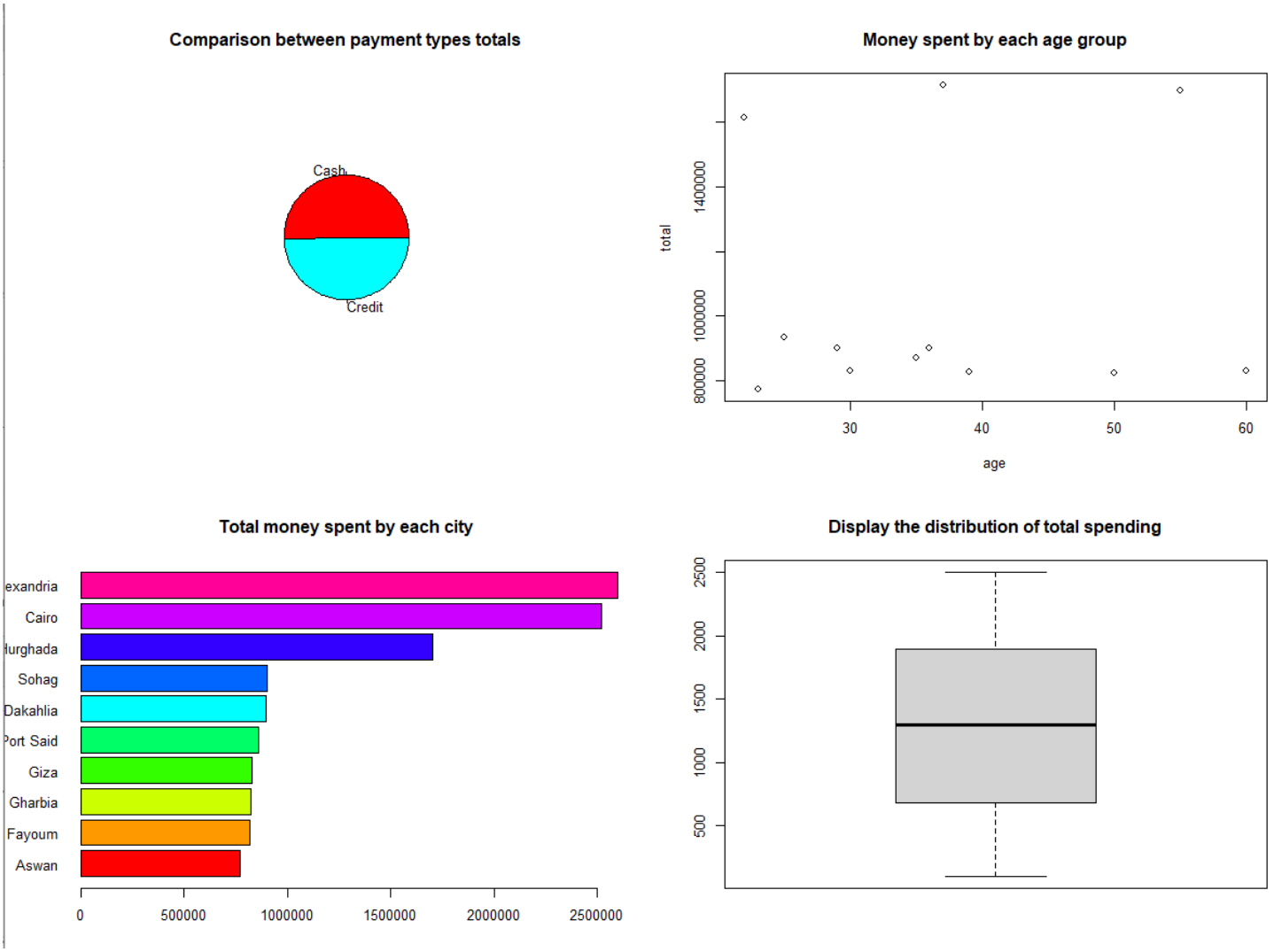
We use the box plot to display the total spending and it helps us get the mean of spendings of each person



c)

```
29  par(mfrow=c(2,2))
```

We use this function to display the plots in one paper



d)

```
xx<-cbind(p$total,p$age)
yy<-cbind(xx,p$customer)
n<-as.numeric(readline("enter the number of centrs between 2 and 4 only"))
if(n>=2&n<=4){
  zz<-kmeans(xx,centers = n)
  ff<-zz$cluster
  print(ff)
  print(zz)
  v<-cbind(yy,ff)
  v<-data.frame(v)
  print(v)
  View(v)
}
```

We use the kmeans built in function to separate the people into groups dependent on their age and total spending and the number of groups is decided by the user by entering the number of centroids and we create a table with name of the customer and their age and their total

e)

```
number_of_transaction<-length(p$items)
number_of_transaction
item<-strsplit(as.vector(p$items), ',')
products<-unique(unlist(item))
products
minsupport<-as.numeric(readline("enter minimum support between 0.001 and 1"))
if(minsupport>1&minsupport<0.001)
  print("invaild input")
minconfidence<-as.numeric(readline("enter minimum confidence between 0.001 and 1"))
if(minconfidence>1&minconfidence<0.001)
  print("invaild input")

sss<-as(item,"transactions")
rules<-apriori(sss,parameter=list(supp=minsupport,conf=minconfidence, minlen=2))
qq<-inspect(rules)
```

We use Apriori algorithm to find interesting relationship between item in the grocery store that the customers could buy together

Firstly we try minimum support and minimum confidence with the value 0.01 , and then we find very interesting relationship between the items bought in the store, more than 500 rule is made and we use them to determine which items is bought with

each other more often and we can take advantage of this by putting these items near each other so people wont go far to look for what they want

```
> qq<-inspect(rules)
      lhs                              rhs                              support    confidence coverage   lift      count
[1]   {hard cheese}                 => {whole milk}                     0.01006609 0.41078838 0.02450432 1.6076815  99
[2]   {whole milk}                  => {hard cheese}                    0.01006609 0.03939515 0.25551601 1.6076815  99
[3]   {butter milk}                 => {other vegetables}               0.01037112 0.37090909 0.02796136 1.9169159 102
[4]   {other vegetables}            => {butter milk}                    0.01037112 0.05359958 0.19349263 1.9169159 102
[5]   {butter milk}                 => {whole milk}                     0.01159126 0.41454545 0.02796136 1.6223854 114
[6]   {whole milk}                  => {butter milk}                    0.01159126 0.04536411 0.25551601 1.6223854 114
[7]   {ham}                         => {whole milk}                     0.01148958 0.44140625 0.02602949 1.7275091 113
[8]   {whole milk}                  => {ham}                            0.01148958 0.04496618 0.25551601 1.7275091 113
[9]   {sliced cheese}               => {whole milk}                     0.01077783 0.43983402 0.02450432 1.7213560 106
[10]  {whole milk}                  => {sliced cheese}                  0.01077783 0.04218066 0.25551601 1.7213560 106
[11]  {oil}                         => {whole milk}                     0.01128622 0.40217391 0.02806304 1.5739675 111
[12]  {whole milk}                  => {oil}                            0.01128622 0.04417031 0.25551601 1.5739675 111
[13]  {onions}                      => {other vegetables}               0.01423488 0.45901639 0.03101169 2.3722681 140
[14]  {other vegetables}            => {onions}                         0.01423488 0.07356805 0.19349263 2.3722681 140
[15]  {onions}                      => {whole milk}                     0.01209964 0.39016393 0.03101169 1.5269647 119
[16]  {whole milk}                  => {onions}                         0.01209964 0.04735376 0.25551601 1.5269647 119
[17]  {berries}                     => {yogurt}                         0.01057448 0.31804281 0.03324860 2.2798477 104
[18]  {yogurt}                      => {berries}                        0.01057448 0.07580175 0.13950178 2.2798477 104
[19]  {berries}                     => {other vegetables}               0.01026945 0.30886850 0.03324860 1.5962805 101
[20]  {other vegetables}            => {berries}                        0.01026945 0.05307409 0.19349263 1.5962805 101
[21]  {berries}                     => {whole milk}                     0.01179461 0.35474006 0.03324860 1.3883281 116
[22]  {whole milk}                  => {berries}                        0.01179461 0.04615997 0.25551601 1.3883281 116
[23]  {hamburger meat}              => {other vegetables}               0.01382816 0.41590214 0.03324860 2.1494470 136
[24]  {other vegetables}            => {hamburger meat}                 0.01382816 0.07146611 0.19349263 2.1494470 136
[25]  {hamburger meat}              => {whole milk}                     0.01474326 0.44342508 0.03324860 1.7354101 145
[26]  {whole milk}                  => {hamburger meat}                 0.01474326 0.05769996 0.25551601 1.7354101 145
[27]  {hygiene articles}            => {whole milk}                     0.01281139 0.38888889 0.03294357 1.5219746 126
[28]  {whole milk}                  => {hygiene articles}               0.01281139 0.05013928 0.25551601 1.5219746 126
[29]  {salty snack}                 => {other vegetables}               0.01077783 0.28494624 0.03782410 1.4726465 106
[30]  {other vegetables}            => {salty snack}                    0.01077783 0.05570152 0.19349263 1.4726465 106
[31]  {salty snack}                 => {whole milk}                     0.01118454 0.29569892 0.03782410 1.1572618 110
[32]  {whole milk}                  => {salty snack}                    0.01118454 0.04377238 0.25551601 1.1572618 110
[33]  {sugar}                       => {other vegetables}               0.01077783 0.31831832 0.03385867 1.6451186 106
[34]  {other vegetables}            => {sugar}                          0.01077783 0.05570152 0.19349263 1.6451186 106
[35]  {sugar}                       => {whole milk}                     0.01504830 0.44444444 0.03385867 1.7393996 148
[36]  {whole milk}                  => {sugar}                          0.01504830 0.05889375 0.25551601 1.7393996 148
[37]  {waffles}                     => {other vegetables}               0.01006609 0.26190476 0.03843416 1.3535645  99
[38]  {other vegetables}            => {waffles}                        0.01006609 0.05202312 0.19349263 1.3535645  99
[39]  {waffles}                     => {whole milk}                     0.01270971 0.33068783 0.03843416 1.2941961 125
[40]  {whole milk}                  => {waffles}                        0.01270971 0.04974135 0.25551601 1.2941961 125
[41]  {long life bakery product}    => {other vegetables}               0.01067616 0.28532609 0.03741739 1.4746096 105
[42]  {other vegetables}            => {long life bakery product}       0.01067616 0.05517604 0.19349263 1.4746096 105
[43]  {long life bakery product}    => {whole milk}                     0.01352313 0.36141304 0.03741739 1.4144438 133
[44]  {whole milk}                  => {long life bakery product}       0.01352313 0.05292479 0.25551601 1.4144438 133
[45]  {dessert}                     => {other vegetables}               0.01159126 0.31232877 0.03711235 1.6141636 114
[46]  {other vegetables}            => {dessert}                        0.01159126 0.05990541 0.19349263 1.6141636 114
[47]  {dessert}                     => {whole milk}                     0.01372649 0.36986301 0.03711235 1.4475140 135
[48]  {whole milk}                  => {dessert}                        0.01372649 0.05372065 0.25551601 1.4475140 135
[49]  {canned beer}                 => {shopping bags}                  0.01138790 0.14659686 0.07768175 1.4879052 112
[50]  {shopping bags}               => {canned beer}                    0.01138790 0.11558308 0.09852567 1.4879052 112
[51]  {canned beer}                 => {soda}                           0.01382816 0.17801047 0.07768175 1.0208356 136
[52]  {soda}                        => {canned beer}                    0.01382816 0.07930029 0.17437722 1.0208356 136
[53]  {canned beer}                 => {rolls/buns}                     0.01128622 0.14528796 0.07768175 0.7898878 111
[54]  {rolls/buns}                  => {canned beer}                    0.01128622 0.06135987 0.18393493 0.7898878 111
[55]  {cream cheese }               => {yogurt}                         0.01240468 0.31282051 0.03965430 2.2424123 122
```

```
[56]  {yogurt}                    => {cream cheese }           0.01240468 0.08892128 0.13950178 2.2424123 122
[57]  {cream cheese }             => {other vegetables}        0.01372649 0.34615385 0.03965430 1.7889769 135
[58]  {other vegetables}          => {cream cheese }           0.01372649 0.07094062 0.19349263 1.7889769 135
[59]  {cream cheese }             => {whole milk}              0.01647178 0.41538462 0.03965430 1.6256696 162
[60]  {whole milk}                => {cream cheese }           0.01647178 0.06446478 0.25551601 1.6256696 162
[61]  {chicken}                   => {root vegetables}         0.01087951 0.25355450 0.04290798 2.3262206 107
[62]  {root vegetables}           => {chicken}                 0.01087951 0.09981343 0.10899847 2.3262206 107
[63]  {chicken}                   => {other vegetables}        0.01789527 0.41706161 0.04290798 2.1554393 176
[64]  {other vegetables}          => {chicken}                 0.01789527 0.09248555 0.19349263 2.1554393 176
[65]  {chicken}                   => {whole milk}              0.01759024 0.40995261 0.04290798 1.6044106 173
[66]  {whole milk}                => {chicken}                 0.01759024 0.06884202 0.25551601 1.6044106 173
[67]  {white bread}               => {soda}                    0.01026945 0.24396135 0.04209456 1.3990437 101
[68]  {soda}                      => {white bread}             0.01026945 0.05889213 0.17437722 1.3990437 101
[69]  {white bread}               => {other vegetables}        0.01372649 0.32608696 0.04209456 1.6852681 135
[70]  {other vegetables}          => {white bread}             0.01372649 0.07094062 0.19349263 1.6852681 135
[71]  {white bread}               => {whole milk}              0.01708185 0.40579710 0.04209456 1.5881474 168
[72]  {whole milk}                => {white bread}             0.01708185 0.06685237 0.25551601 1.5881474 168
[73]  {chocolate}                 => {soda}                    0.01352313 0.27254098 0.04961871 1.5629391 133
[74]  {soda}                      => {chocolate}               0.01352313 0.07755102 0.17437722 1.5629391 133
[75]  {chocolate}                 => {rolls/buns}              0.01179461 0.23770492 0.04961871 1.2923316 116
[76]  {rolls/buns}                => {chocolate}               0.01179461 0.06412383 0.18393493 1.2923316 116
[77]  {chocolate}                 => {other vegetables}        0.01270971 0.25614754 0.04961871 1.3238103 125
[78]  {other vegetables}          => {chocolate}               0.01270971 0.06568576 0.19349263 1.3238103 125
[79]  {chocolate}                 => {whole milk}              0.01667514 0.33606557 0.04961871 1.3152427 164
[80]  {whole milk}                => {chocolate}               0.01667514 0.06526064 0.25551601 1.3152427 164
[81]  {coffee}                    => {rolls/buns}              0.01098119 0.18914186 0.05805796 1.0283085 108
[82]  {rolls/buns}                => {coffee}                  0.01098119 0.05970149 0.18393493 1.0283085 108
[83]  {coffee}                    => {other vegetables}        0.01342145 0.23117338 0.05805796 1.1947400 132
[84]  {other vegetables}          => {coffee}                  0.01342145 0.06936416 0.19349263 1.1947400 132
[85]  {coffee}                    => {whole milk}              0.01870869 0.32224168 0.05805796 1.2611408 184
[86]  {whole milk}                => {coffee}                  0.01870869 0.07321926 0.25551601 1.2611408 184
[87]  {frozen vegetables}         => {root vegetables}         0.01159126 0.24101480 0.04809354 2.2111759 114
[88]  {root vegetables}           => {frozen vegetables}       0.01159126 0.10634328 0.10899847 2.2111759 114
[89]  {frozen vegetables}         => {yogurt}                  0.01240468 0.25792812 0.04809354 1.8489235 122
[90]  {yogurt}                    => {frozen vegetables}       0.01240468 0.08892128 0.13950178 1.8489235 122
[91]  {frozen vegetables}         => {rolls/buns}              0.01016777 0.21141649 0.04809354 1.1494092 100
[92]  {rolls/buns}                => {frozen vegetables}       0.01016777 0.05527916 0.18393493 1.1494092 100
[93]  {frozen vegetables}         => {other vegetables}        0.01779359 0.36997886 0.04809354 1.9121083 175
[94]  {other vegetables}          => {frozen vegetables}       0.01779359 0.09196006 0.19349263 1.9121083 175
[95]  {frozen vegetables}         => {whole milk}              0.02043721 0.42494715 0.04809354 1.6630940 201
[96]  {whole milk}                => {frozen vegetables}       0.02043721 0.07998408 0.25551601 1.6630940 201
[97]  {beef}                      => {root vegetables}         0.01738688 0.33139535 0.05246568 3.0403668 171
[98]  {root vegetables}           => {beef}                    0.01738688 0.15951493 0.10899847 3.0403668 171
[99]  {beef}                      => {yogurt}                  0.01169293 0.22286822 0.05246568 1.5976012 115
[100] {yogurt}                    => {beef}                    0.01169293 0.08381924 0.13950178 1.5976012 115
[101] {beef}                      => {rolls/buns}              0.01362481 0.25968992 0.05246568 1.4118576 134
[102] {rolls/buns}                => {beef}                    0.01362481 0.07407407 0.18393493 1.4118576 134
[103] {beef}                      => {other vegetables}        0.01972547 0.37596899 0.05246568 1.9430662 194
[104] {other vegetables}          => {beef}                    0.01972547 0.10194430 0.19349263 1.9430662 194
[105] {beef}                      => {whole milk}              0.02125064 0.40503876 0.05246568 1.5851795 209
[106] {whole milk}                => {beef}                    0.02125064 0.08316753 0.25551601 1.5851795 209
[107] {curd}                      => {whipped/sour cream}      0.01047280 0.19656489 0.05327911 2.7421499 103
[108] {whipped/sour cream}        => {curd}                    0.01047280 0.14609929 0.07168277 2.7421499 103
[109] {curd}                      => {tropical fruit}          0.01026945 0.19274809 0.05327911 1.8368968 101
[110] {tropical fruit}            => {curd}                    0.01026945 0.09786822 0.10493137 1.8368968 101
[111] {curd}                      => {root vegetables}         0.01087951 0.20419847 0.05327911 1.8734067 107
[112] {root vegetables}           => {curd}                    0.01087951 0.09981343 0.10899847 1.8734067 107
[113] {curd}                      => {yogurt}                  0.01728521 0.32442748 0.05327911 2.3256154 170
[114] {yogurt}                    => {curd}                    0.01728521 0.12390671 0.13950178 2.3256154 170
[115] {curd}                      => {rolls/buns}              0.01006609 0.18893130 0.05327911 1.0271638  99
[116] {rolls/buns}                => {curd}                    0.01006609 0.05472637 0.18393493 1.0271638  99
[117] {curd}                      => {other vegetables}        0.01718353 0.32251908 0.05327911 1.6668288 169
[118] {other vegetables}          => {curd}                    0.01718353 0.08880715 0.19349263 1.6668288 169
[119] {curd}                      => {whole milk}              0.02613116 0.49045802 0.05327911 1.9194805 257
[120] {whole milk}                => {curd}                    0.02613116 0.10226821 0.25551601 1.9194805 257
[121] {napkins}                   => {tropical fruit}          0.01006609 0.19223301 0.05236401 1.8319880  99
[122] {tropical fruit}            => {napkins}                 0.01006609 0.09593023 0.10493137 1.8319880  99
[123] {napkins}                   => {soda}                    0.01199797 0.22912621 0.05236401 1.3139687 118
[124] {soda}                      => {napkins}                 0.01199797 0.06880466 0.17437722 1.3139687 118
[125] {napkins}                   => {yogurt}                  0.01230300 0.23495146 0.05236401 1.6842183 121
 [ reached 'max' / getOption("max.print") -- omitted 397 rows ]
```