################################# Association Rule ############################

#Objective: Prepare rules for the all the data sets

#1) Try different values of support and confidence. Observe the change in

#number of rules for different support,confidence values

#2) Change the minimum length in apriori algorithm

#3) Visulize the obtained rules using different plots

#Data : groceries.csv

##################################################################################

install.packages("arules")

library("arules") # Used for building association rules i.e. apriori algorithm

**##Step1 : Data Exploration**

books<-read.csv('D:\\Shilpa\\Datascience\\Assignments\\Association Rule\\book.csv')

View(books)

# making rules using apriori algorithm

# Keep changing support and confidence values to obtain different rules

?apriori

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#

**#Step 2: Training a model on the data**

#install.packages("arulesViz")

library("arulesViz") # for visualizing rules

**#set better support and confidence levels to learn more rules**

**# Building rules using apriori algorithm support=0.02,confidence=0.5**

rules <- apriori(as.matrix(books),parameter=list(support=0.02, confidence = 0.5,minlen=5))

rules

#Apriori

#Parameter specification:

# confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext

#0.5 0.1 1 none FALSE TRUE 5 0.02 5 10 rules TRUE

#Algorithmic control:

# filter tree heap memopt load sort verbose

#0.1 TRUE TRUE FALSE TRUE 2 TRUE

#Absolute minimum support count: 40

#set item appearances ...[0 item(s)] done [0.00s].

#set transactions ...[11 item(s), 2000 transaction(s)] done [0.00s].

#sorting and recoding items ... [11 item(s)] done [0.00s].

#creating transaction tree ... done [0.00s].

#checking subsets of size 1 2 3 4 5 6 done [0.00s].

#writing ... [186 rule(s)] done [0.00s].

#creating S4 object ... done [0.00s].

**#Step 3: Evaluating model performance**

# summary of books association rules

summary(rules)

#set of 186 rules

#rule length distribution (lhs + rhs):sizes

#5 6

#160 26

#Min. 1st Qu. Median Mean 3rd Qu. Max.

#5.00 5.00 5.00 5.14 5.00 6.00

#summary of quality measures:

# support confidence coverage lift count

#Min. :0.02000 Min. :0.5000 Min. :0.02000 Min. : 1.751 Min. : 40.00

#1st Qu.:0.02450 1st Qu.:0.5762 1st Qu.:0.03062 1st Qu.: 2.087 1st Qu.: 49.00

#Median :0.02750 Median :0.6563 Median :0.04475 Median : 2.242 Median : 55.00

#Mean :0.03088 Mean :0.7224 Mean :0.04518 Mean : 2.571 Mean : 61.76

#3rd Qu.:0.03625 3rd Qu.:0.8727 3rd Qu.:0.05250 3rd Qu.: 2.357 3rd Qu.: 72.50

#Max. :0.05350 Max. :1.0000 Max. :0.08900 Max. :14.122 Max. :107.00

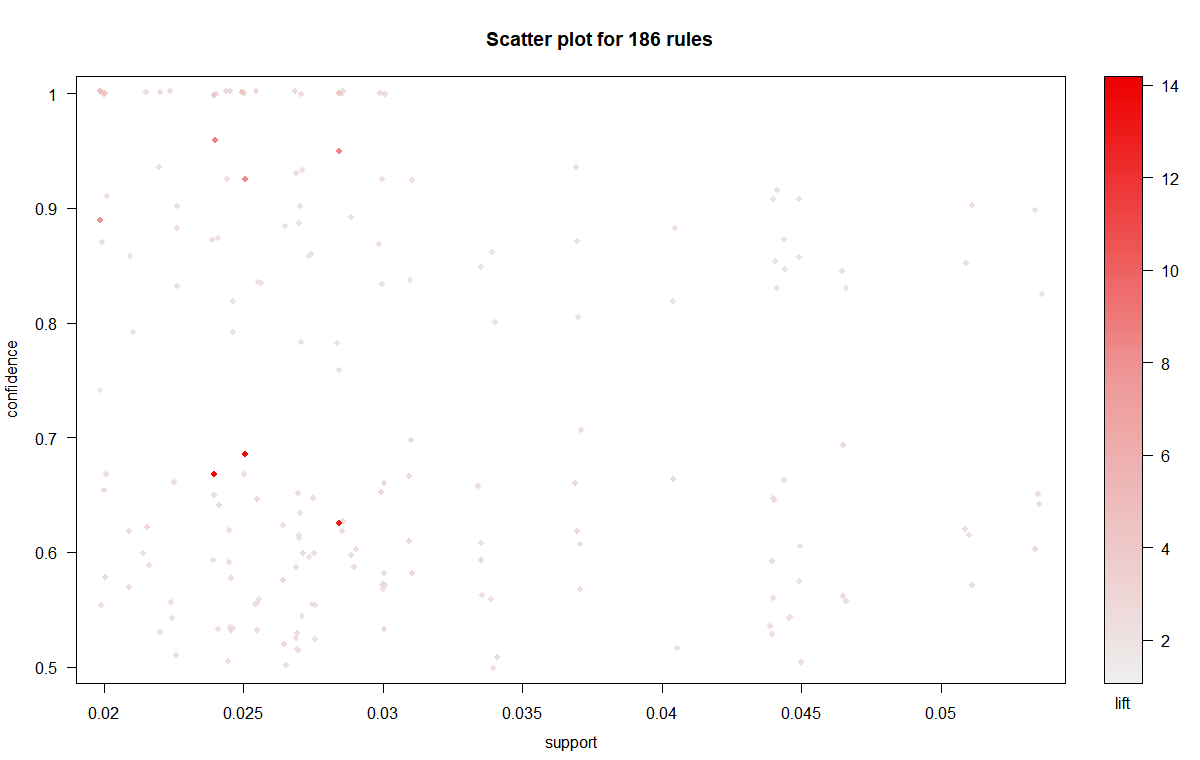
#mining info:

# data ntransactions support confidence

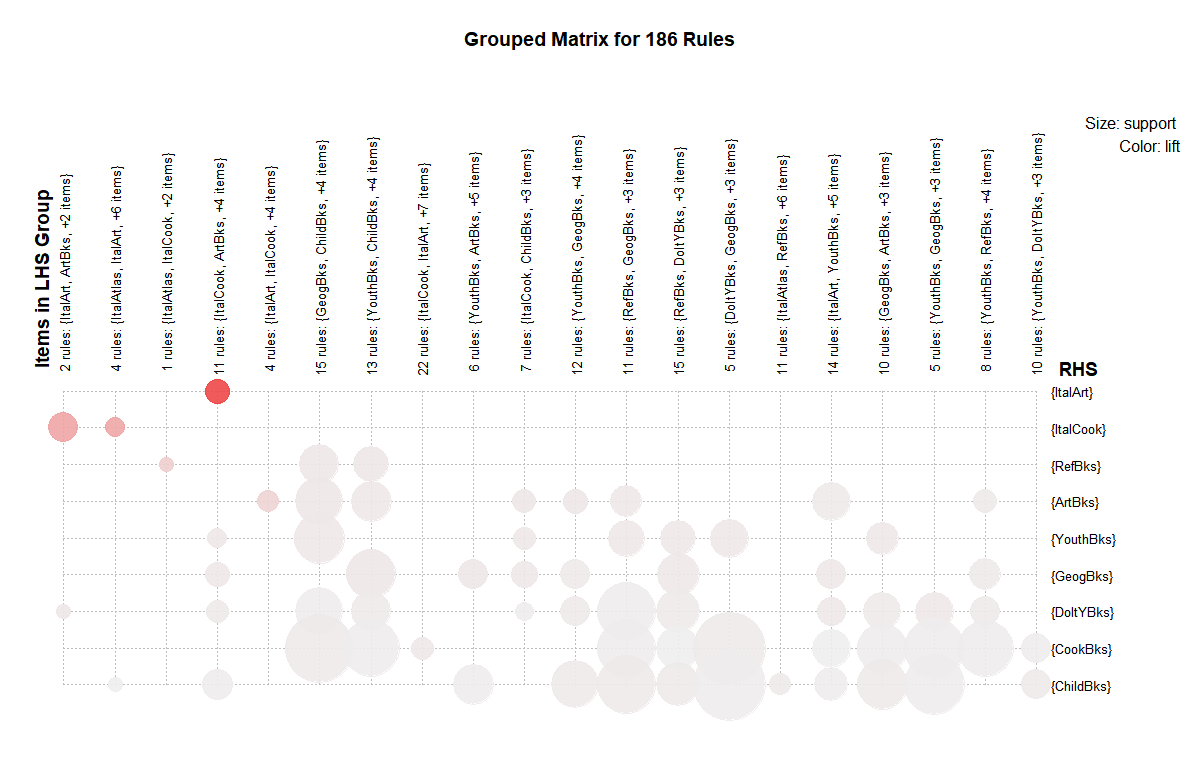
#as.matrix(books) 2000 0.02 0.5

**#Step 4: Visualization**

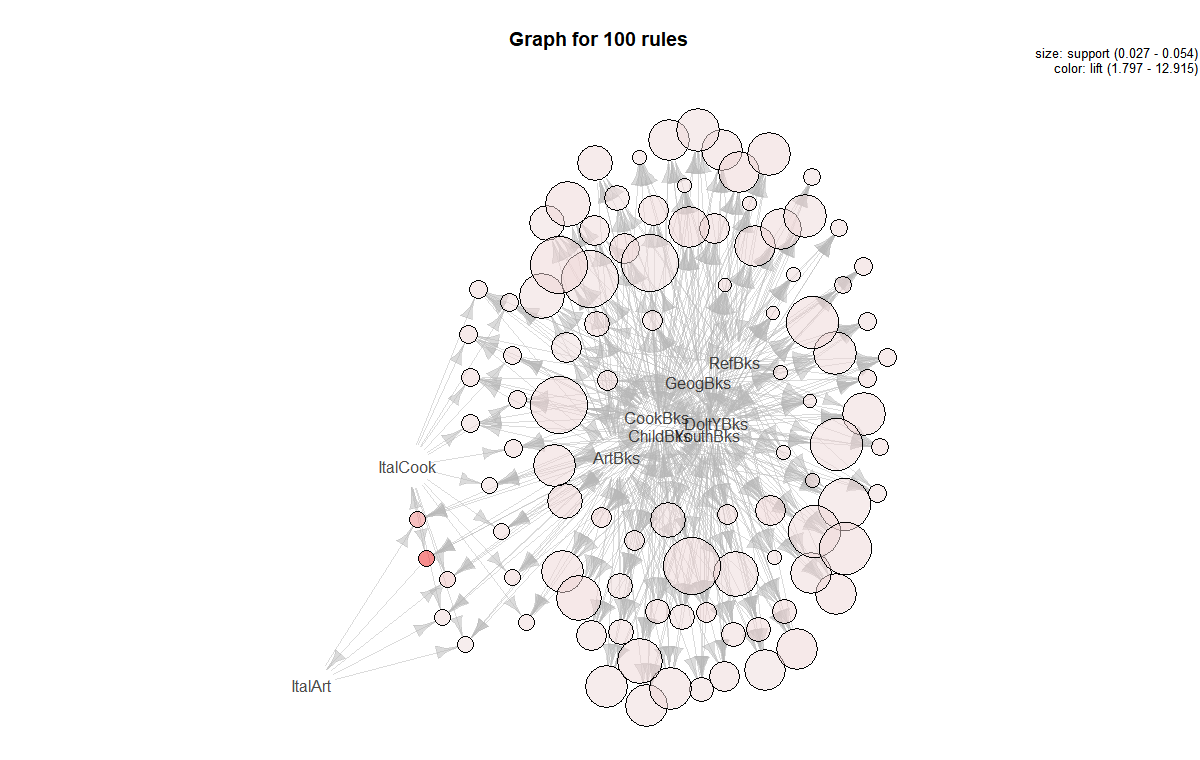
plot(rules,method = "scatterplot")



plot(rules,method = "grouped")



plot(rules,method = "graph")



#look at the first six rules

inspect(head(sort(rules, by = "lift")))

#lhs rhs support confidence coverage lift count

#[1] {CookBks,DoItYBks,ArtBks,ItalCook} => {ItalArt} 0.0250 0.6849315 0.0365 14.122299 50

#[2] {CookBks,ArtBks,GeogBks,ItalCook} => {ItalArt} 0.0240 0.6666667 0.0360 13.745704 48

#[3] {ChildBks,CookBks,ArtBks,ItalCook} => {ItalArt} 0.0285 0.6263736 0.0455 12.914920 57

#[4] {CookBks,ArtBks,GeogBks,ItalArt} => {ItalCook} 0.0240 0.9600000 0.0250 8.458150 48

#[5] {ChildBks,CookBks,ArtBks,ItalArt} => {ItalCook} 0.0285 0.9500000 0.0300 8.370044 57

#[6] {CookBks,DoItYBks,ArtBks,ItalArt} => {ItalCook} 0.0250 0.9259259 0.0270 8.157938 50

##################################################################################

#Association with support =0.01,confidence = 0.06,minlen=4

book\_rules<-apriori(as.matrix(books),parameter = list(support = 0.01,confidence = 0.06,minlen=4))

#Apriori

#Parameter specification:

# confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext

#0.06 0.1 1 none FALSE TRUE 5 0.01 4 10 rules TRUE

#Algorithmic control:

# filter tree heap memopt load sort verbose

#0.1 TRUE TRUE FALSE TRUE 2 TRUE

#Absolute minimum support count: 20

#set item appearances ...[0 item(s)] done [0.00s].

#set transactions ...[11 item(s), 2000 transaction(s)] done [0.00s].

#sorting and recoding items ... [11 item(s)] done [0.00s].

#creating transaction tree ... done [0.00s].

#checking subsets of size 1 2 3 4 5 6 7 done [0.00s].

#writing ... [1781 rule(s)] done [0.00s].

#creating S4 object ... done [0.04s].

**#Step : Evaluating model performance**

# summary of books association rules

summary(book\_rules)

#set of 1781 rules

#rule length distribution (lhs + rhs):sizes

#4 5 6 7

#800 695 258 28

#Min. 1st Qu. Median Mean 3rd Qu. Max.

#4.000 4.000 5.000 4.727 5.000 7.000

#summary of quality measures:

# support confidence coverage lift count

#Min. :0.01000 Min. :0.08904 Min. :0.0100 Min. : 1.019 Min. : 20.00

#st Qu.:0.01200 1st Qu.:0.51938 1st Qu.:0.0165 1st Qu.: 2.080 1st Qu.: 24.00

#Median :0.01450 Median :0.64912 Median :0.0250 Median : 2.320 Median : 29.00

#Mean :0.02032 Mean :0.66366 Mean :0.0366 Mean : 3.726 Mean : 40.65

#3rd Qu.:0.02400 3rd Qu.:0.87500 3rd Qu.:0.0465 3rd Qu.: 4.149 3rd Qu.: 48.00

#Max. :0.08900 Max. :1.00000 Max. :0.1495 Max. :23.023 Max. :178.00

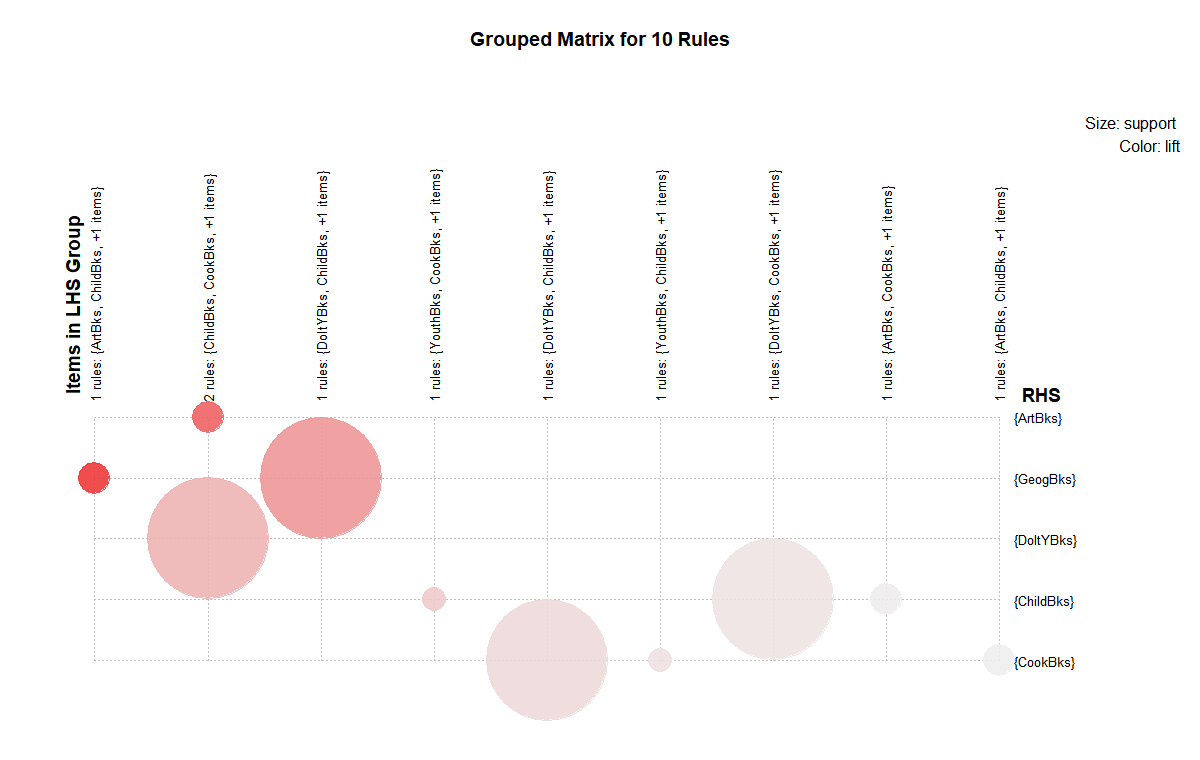
#mining info:

# data ntransactions support confidence

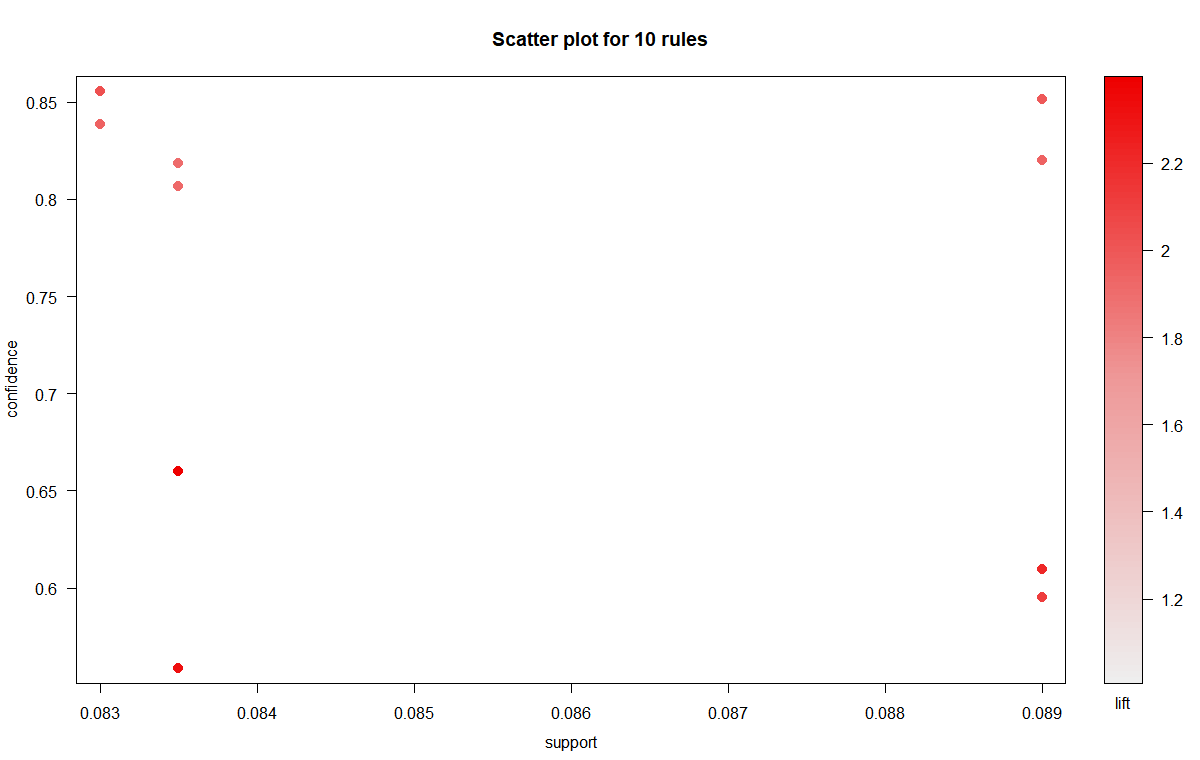
#s.matrix(books) 2000 0.01 0.06

**#Step: Visualization**

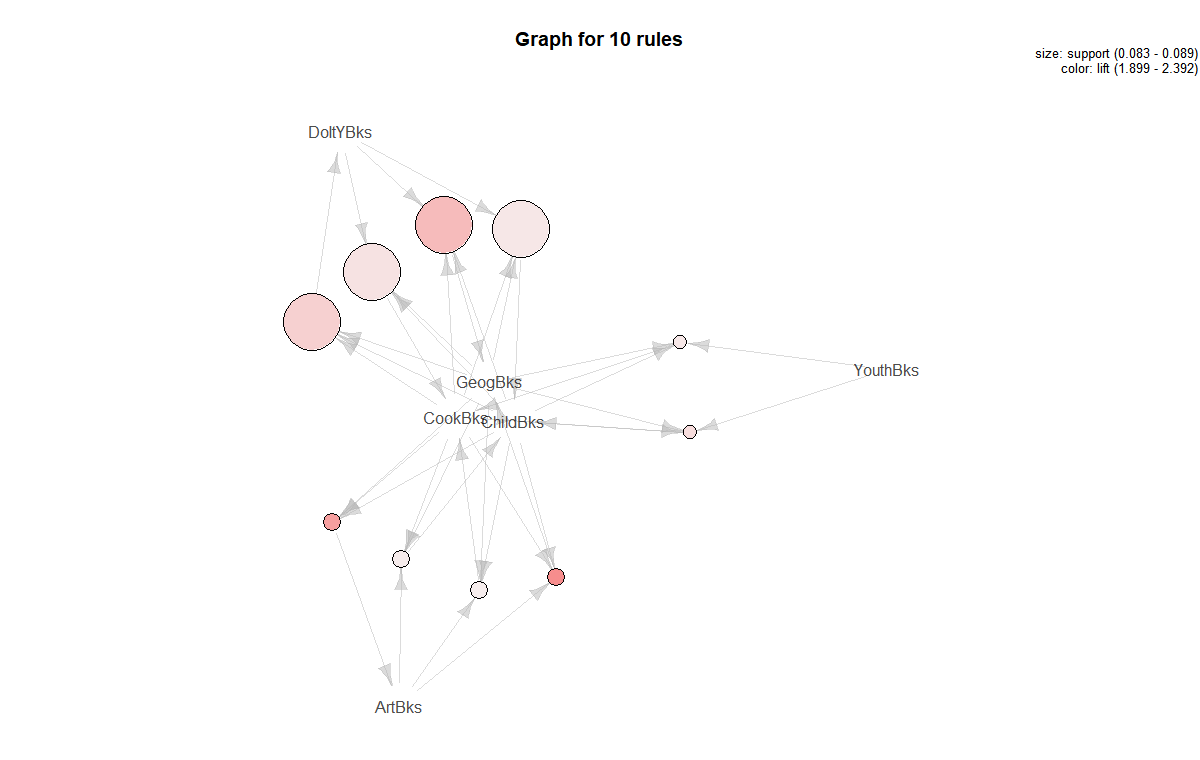
plot(head(sort(book\_rules), n = 10), method = "grouped", control = list(cex = 0.2))



plot(head(sort(book\_rules), n = 10), method = "scatterplot", control = list(cex = 0.8))



plot(head(sort(book\_rules), n = 10), method = "graph")



**#Step: Improving model performance**

#sorting books rules by lift

rules1 <- sort(book\_rules,by="lift")

inspect(rules1[1:4])

#lhs rhs support confidence coverage lift count

#[1] {RefBks,GeogBks,ItalArt} => {ItalAtlas} 0.0115 0.8518519 0.0135 23.02302 23

#[2] {RefBks,ArtBks,GeogBks,ItalArt} => {ItalAtlas} 0.0115 0.8518519 0.0135 23.02302 23

#[3] {ChildBks,RefBks,GeogBks,ItalArt} => {ItalAtlas} 0.0100 0.8333333 0.0120 22.52252 20

#[4] {ChildBks,RefBks,ArtBks,GeogBks,ItalArt} => {ItalAtlas} 0.0100 0.8333333 0.0120 22.52252 20

**#The first rule, with a lift of about 23.02302,**

**#implies that people who RefBks,GeogBks,ItalArt are nearly eight times more likely to ItalAtlas .**

**#writing the rules to a CSV file**

write(rules1, file="book\_rules.csv",sep=",")

getwd()