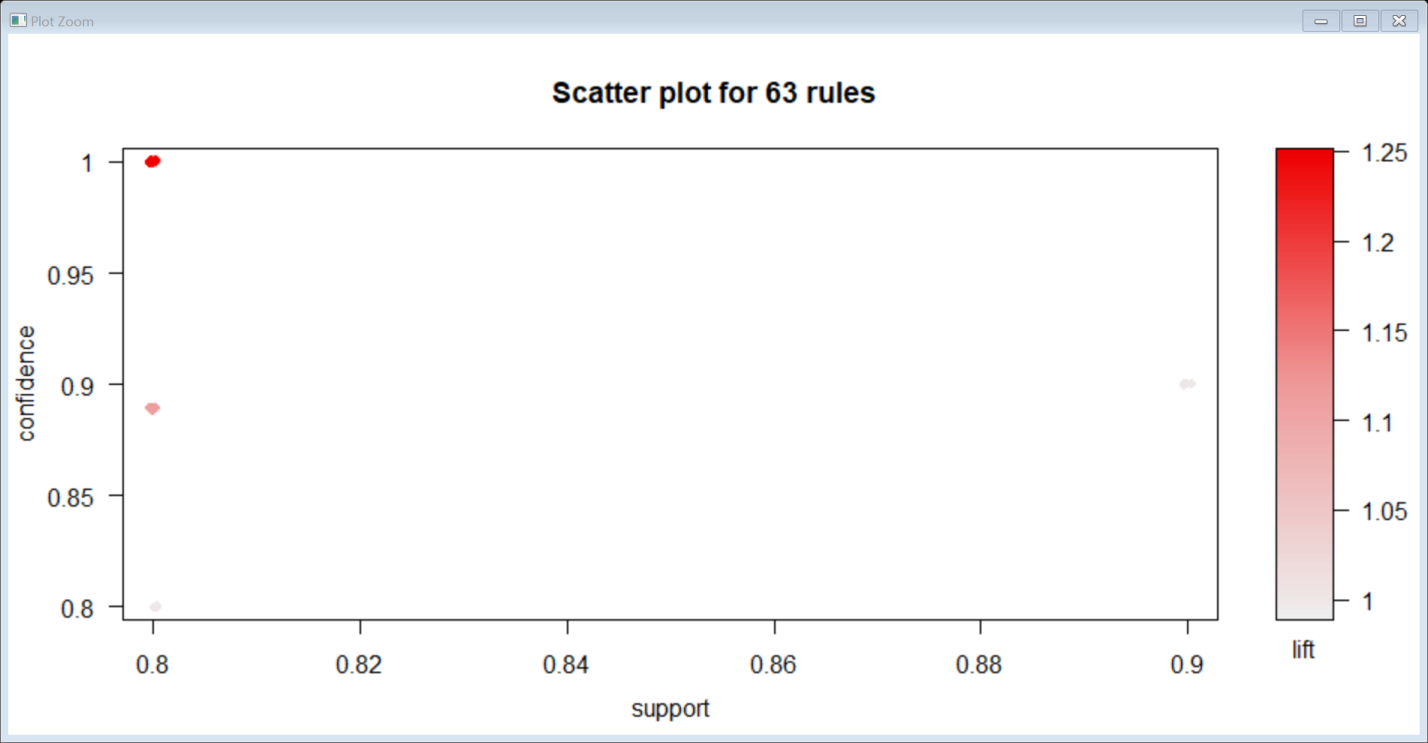
# Association Rules

**Business Objective:Movies watched**

Step 1: Install Packages arules and arulesViz

Step 2 : Since the dataset of variables contain categorical values convert them using factor.

Step 3 : Create rule using apriori Algorithum with sup =0.8 and conf =0.7 and do a scatter plot.



Rules = set of 63 rules

Step 4: Inspect the starting 6 values using head by using lift >1

lhs rhs support confidence lift count

[1] {LOTR1=0} => {LOTR2=0} 0.8 1 1.25 8

[2] {LOTR2=0} => {LOTR1=0} 0.8 1 1.25 8

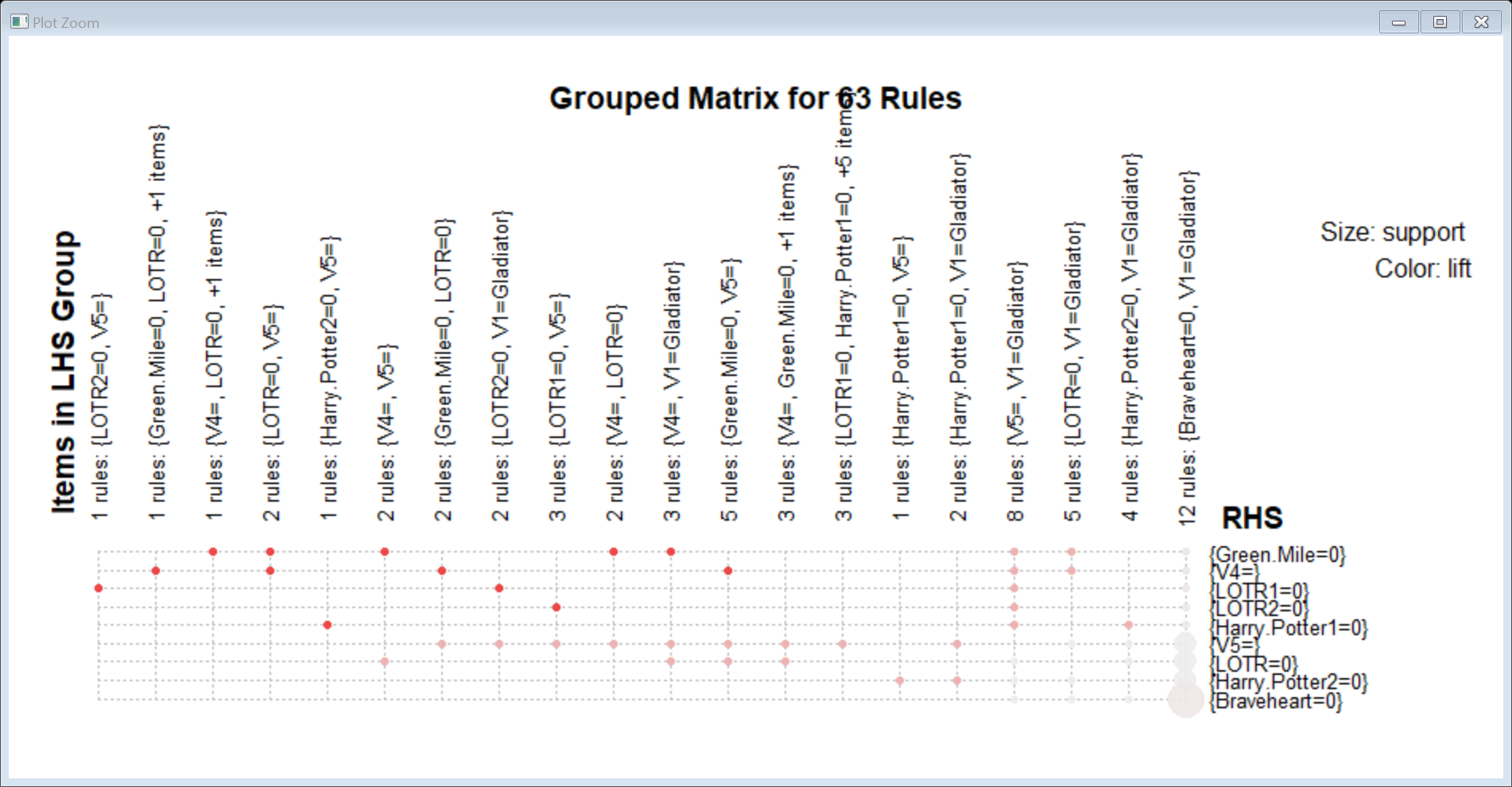
[3] {V4=} => {Green.Mile=0} 0.8 1 1.25 8

[4] {Green.Mile=0} => {V4=} 0.8 1 1.25 8

[5] {V5=,LOTR1=0} => {LOTR2=0} 0.8 1 1.25 8

[6] {V5=,LOTR2=0} => {LOTR1=0} 0.8 1 1.25 8

Step 5: Create another plot to determine the highest association



From the plot we can make the inference that 0 being watched and 1 being not watched

: V4,LOTR was watched along with Green Miles highest

: V4 and V5 was watched along the Green Miles

: Green mile was watched along with V4,Gladiator

Step 6: create a rules with rhs containing Patriot only

rhs = "Patriot=1"

lhs rhs support confidence lift count

[1] {V2=Patriot} => {Patriot=1} 0.6 1.0000000 1.666667 6

[2] {V1=Gladiator} => {Patriot=1} 0.6 1.0000000 1.666667 6

[3] {Gladiator=1,Green.Mile=0} => {Patriot=1} 0.6 1.0000000 1.666667 6

[4] {V4=,Gladiator=1} => {Patriot=1} 0.6 1.0000000 1.666667 6

[5] {Gladiator=1,LOTR=0} => {Patriot=1} 0.6 1.0000000 1.666667 6

[6] {Harry.Potter1=0,LOTR2=0,Green.Mile=0} => {Patriot=1} 0.6 1.0000000 1.666667 6

set of 2032 rules

table(Movies$Patriot)

0 1

4 6

Step 7 : Non Redundant values are below.

support confidence lift count

1 0.6 1 1.666667 6

2 0.6 1 1.666667 6

28 0.6 1 1.666667 6

32 0.6 1 1.666667 6

34 0.6 1 1.666667 6

173 0.6 1 1.666667 6

Conclusion:

: V4,LOTR was watched along with Green Miles highest

: V4 and V5 was watched along the Green Miles

: Green mile was watched along with V4,Gladiator

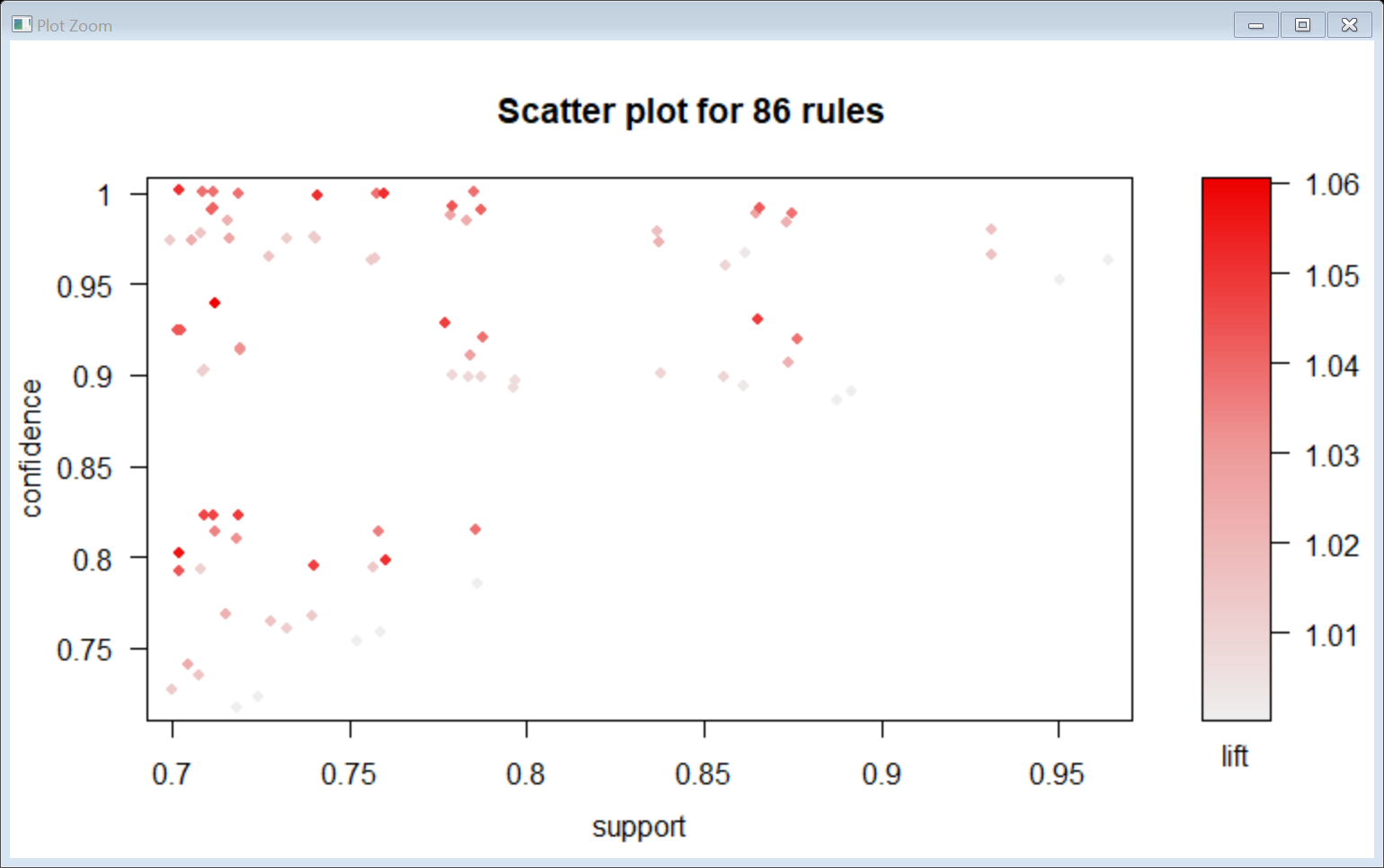
**Business Objective: Books**

Step 1: Install Packages arules and arulesViz

Step 2 : Since the dataset of variables contain categorical values convert them using factor.

Step 3 : Create rule using apriori Algorithum with sup =0.7 and conf =0.7 and do a scatter plot

set of 86 rules

****

Step 4: Inspect the starting 6 values using head by using lift >1

lhs rhs support confidence lift count

[1] {RefBks=0,

ItalArt=0} => {ItalCook=0} 0.7115 0.9398943 1.060230 1423

[2] {RefBks=0,

ItalAtlas=0,

ItalArt=0} => {ItalCook=0} 0.7115 0.9398943 1.060230 1423

[3] {ItalCook=0,

ItalArt=0} => {ArtBks=0} 0.7020 0.8018275 1.056426 1404

[4] {ArtBks=0} => {ItalArt=0} 0.7590 1.0000000 1.050972 1518

[5] {ItalArt=0} => {ArtBks=0} 0.7590 0.7976879 1.050972 1518

[6] {ArtBks=0,

ItalCook=0} => {ItalArt=0} 0.7020 1.0000000 1.050972 1404

Step 5 : Inspect the starting 6 values using head by using rule.

support confidence lift count

1 0.7180 0.7180 1 1436

2 0.7240 0.7240 1 1448

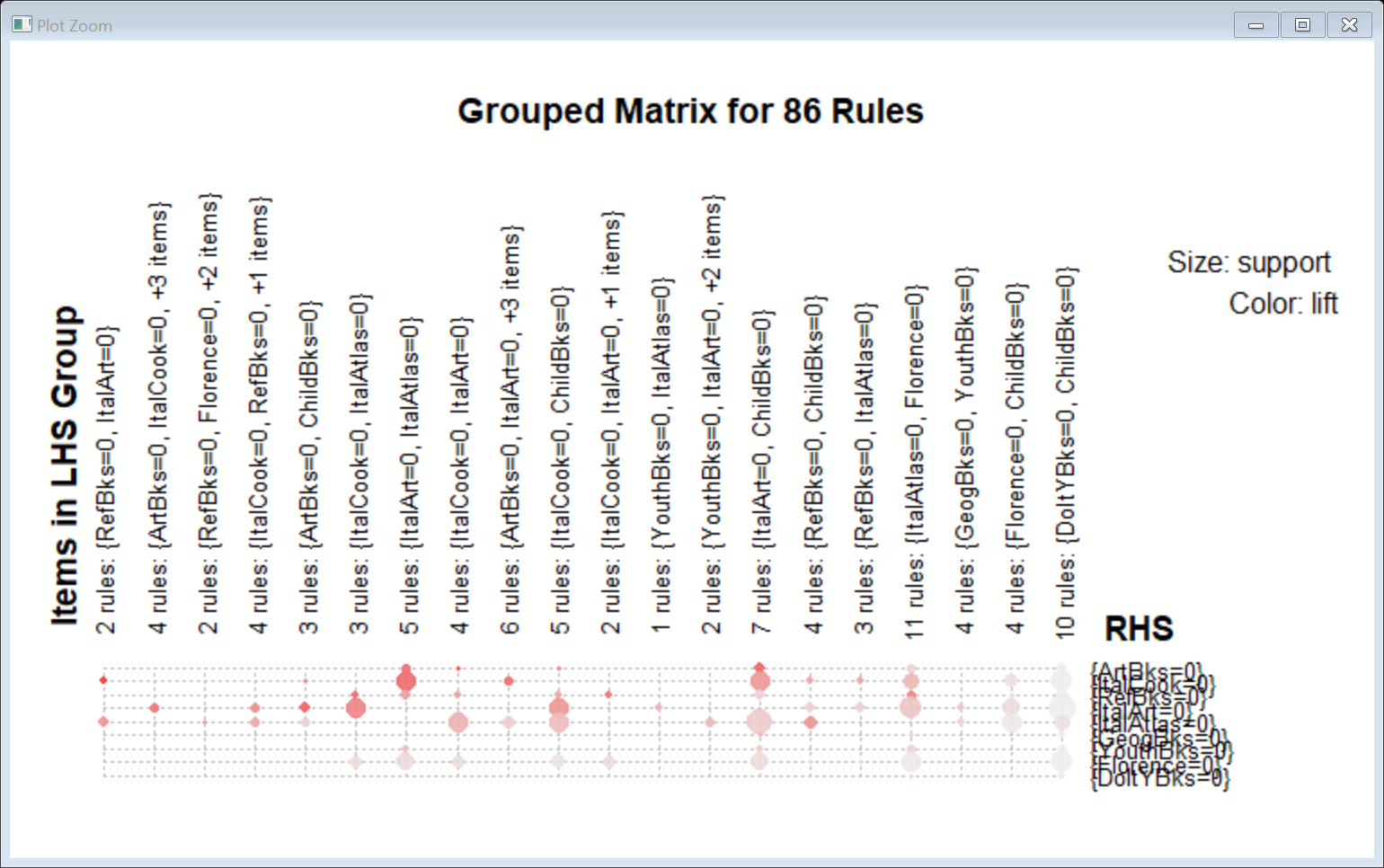
3 0.7525 0.7525 1 1505

4 0.7590 0.7590 1 1518

5 0.7855 0.7855 1 1571

6 0.8865 0.8865 1 1773

Step 6: Create another plot to determine the highest association



* RefBks,ItalArt goes along with ItalCook they can be placed together.
* ItalArt,ChildBks goes along with ArtBks they can be placed together.
* ItalArt,ItalAtlas goes along with ArtBks they can be placed together.

Step 7:Table with ChildBks observation

0 1

1154 846

**Business Objective: Groceries**

Step 1: Install Packages arules and arulesViz summary of groceries

citrus.fruit.semi.finished.bread.margarine.ready.soups

canned beer : 260

soda : 156

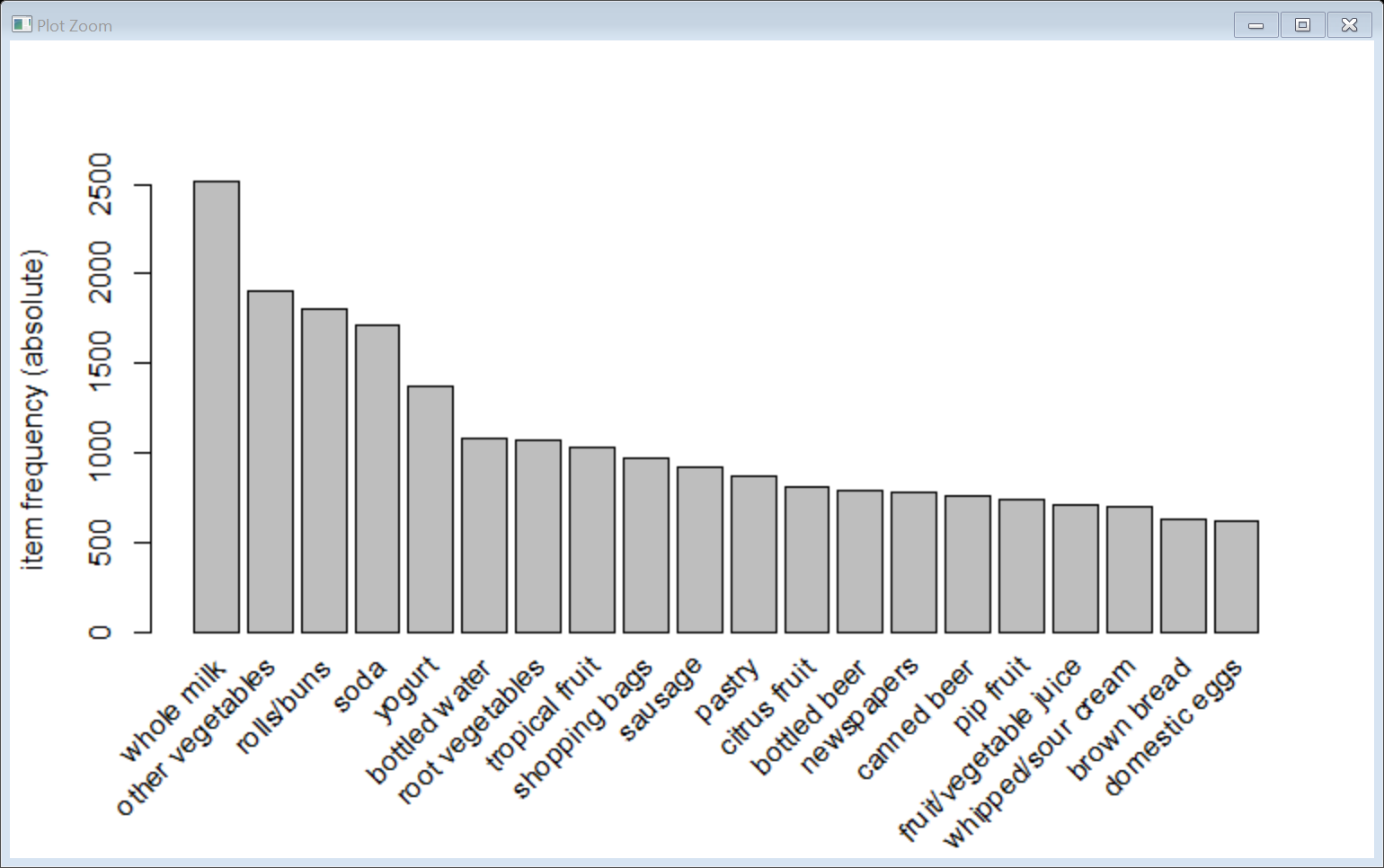
whole milk : 121

bottled beer : 120

rolls/buns : 109

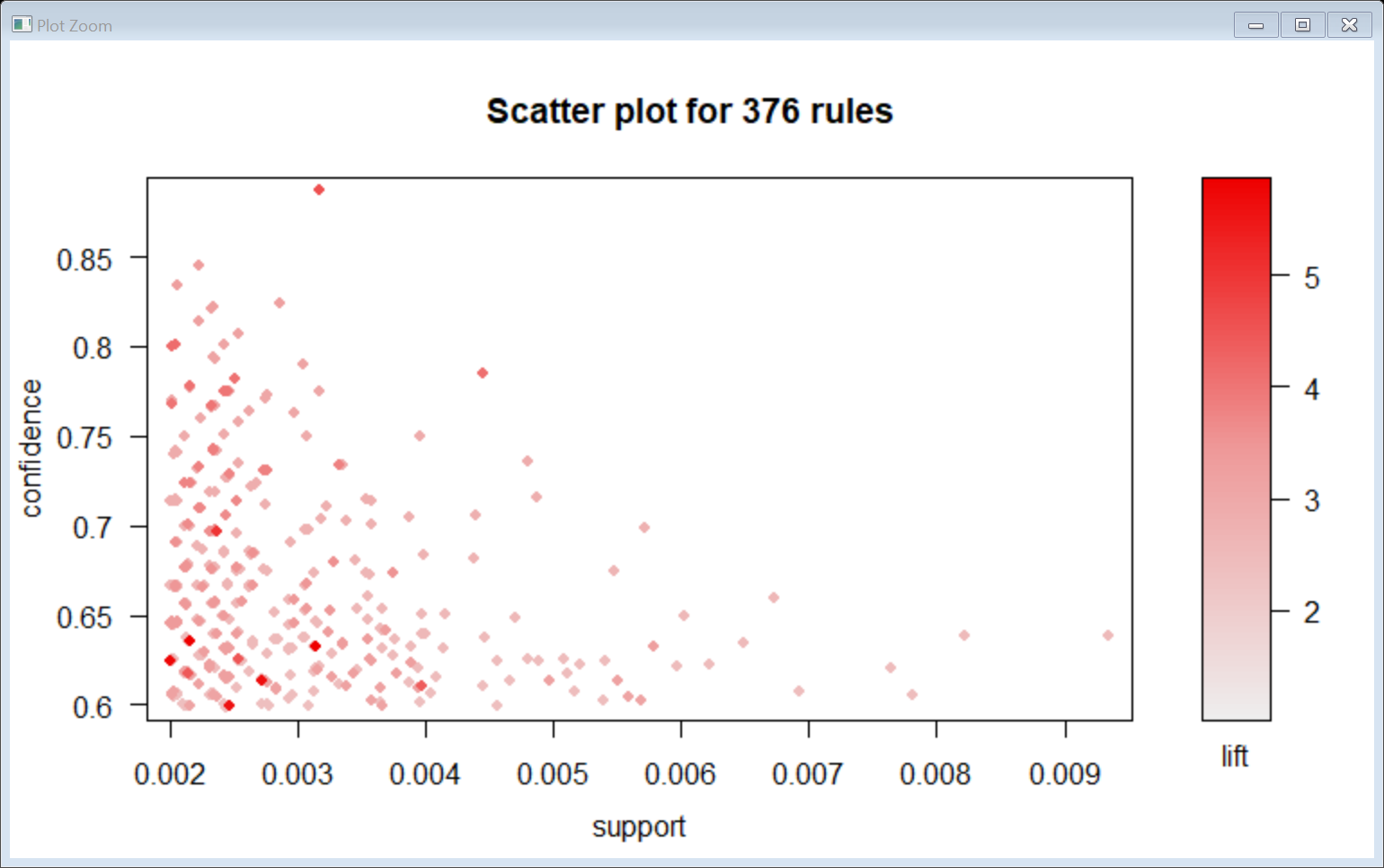
bottled water: 67

(Other) :9001

Step 2 : Create the frequency plot for the top 20 purchased item.

Step 3 : Create rule using apriori Algorithum with sup =0.002 and conf =0.6 and do a scatter plot

set of 376 rules



Step 4 : Inspect the starting 6 values using head by using rule.

support confidence lift count

1 0.003660397 0.6428571 2.515917 36

2 0.004677173 0.6133333 2.400371 46

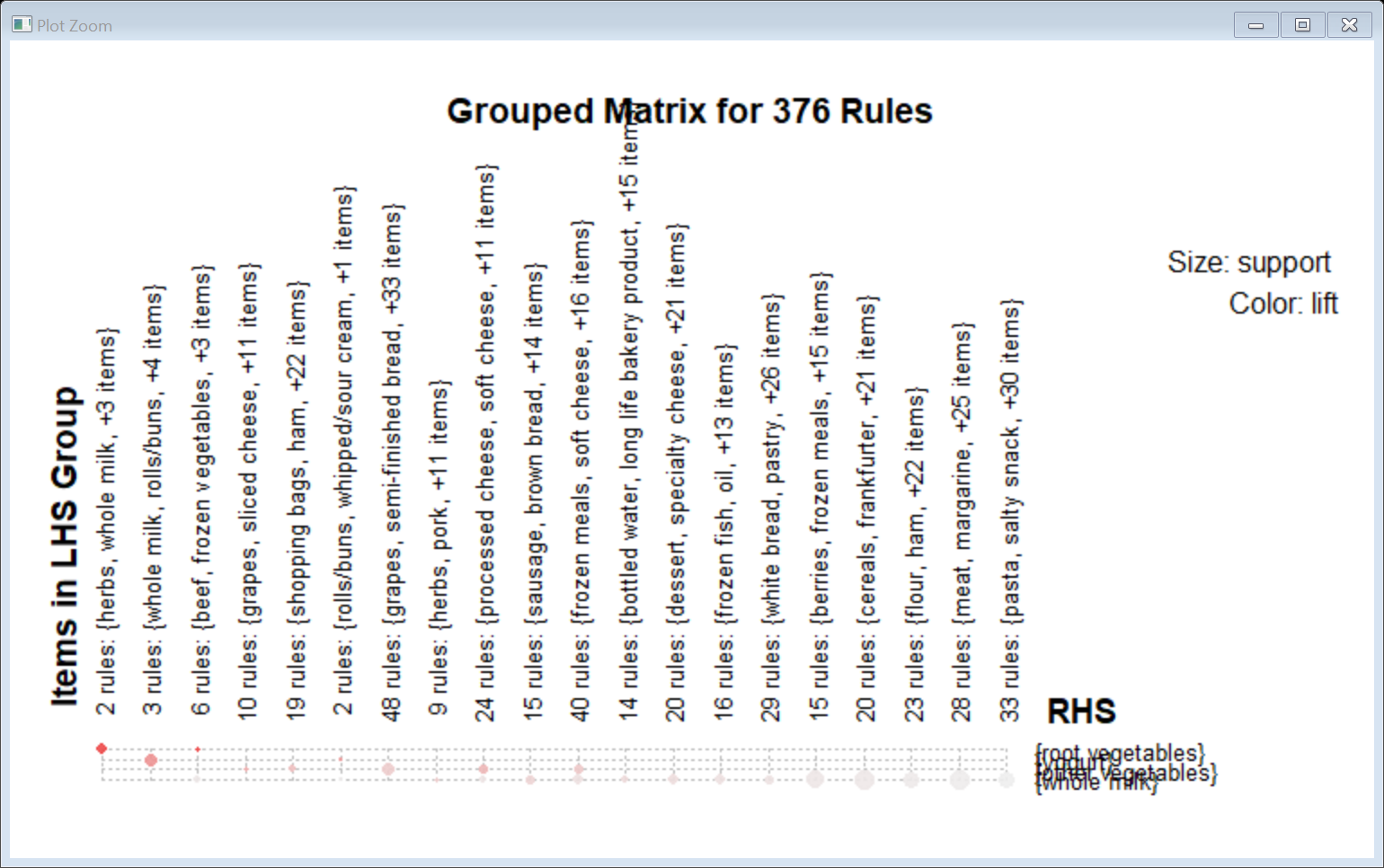
3 0.002033554 0.7142857 2.795464 20

4 0.002236909 0.7096774 3.667723 22

5 0.002440264 0.7741935 3.029922 24

6 0.002643620 0.6666667 2.609099 26

Step 5: Create another plot to determine the highest association



* Whole milk,Herbs along with root vegetables need to be kept together
* WholeMilk,rolls/buns along with yogurt needs to be kept together
* Beef , frozen vegetables alongside root vegetables needs to be kept together.

Step 6: Top 6 of the Highest association is below

[1] {root vegetables,

whole milk} => {other vegetables} 0.02318251 0.4740125 2.449770 228

[2] {root vegetables} => {other vegetables} 0.04738180 0.4347015 2.246605 466

[3] {whipped/sour cream} => {other vegetables} 0.02887646 0.4028369 2.081924 284

[4] {other vegetables,

yogurt} => {whole milk} 0.02226741 0.5128806 2.007235 219

[5] {butter} => {whole milk} 0.02755465 0.4972477 1.946053 271

[6] {curd} => {whole milk} 0.02613116 0.4904580 1.919481 257

Step 7: Non Redundant values are below.

support confidence lift count

14 0.02318251 0.4740125 2.449770 228

10 0.04738180 0.4347015 2.246605 466

7 0.02887646 0.4028369 2.081924 284

15 0.02226741 0.5128806 2.007235 219

5 0.02755465 0.4972477 1.946053 271

3 0.02613116 0.4904580 1.919481 257

Conclusion:

* Whole milk,Herbs,other vegetables along with root vegetables need to be kept together
* WholeMilk,rolls/buns,butter,sour cream along with yogurt needs to be kept together
* Beef , frozen vegetables alongside root vegetables needs to be kept together