Vidit Kalani MSCS 5610 Data Mining
Assignment 3 April 4, 2018

Titanic Dataset is a dimensional table in the Datasets package, with information about the future of travelers on social class, gender, age, survival etc. according to the future of Titanic. To make it suitable for the association rule mining, we rebuild the raw information as a titanic, where each row represents a person.

Function apriori () Mine frequent item sets, association rules or association hyperedges using the Apriori algorithm. The Apriori algorithm employs level-wise search for frequent item sets.

Default settings:

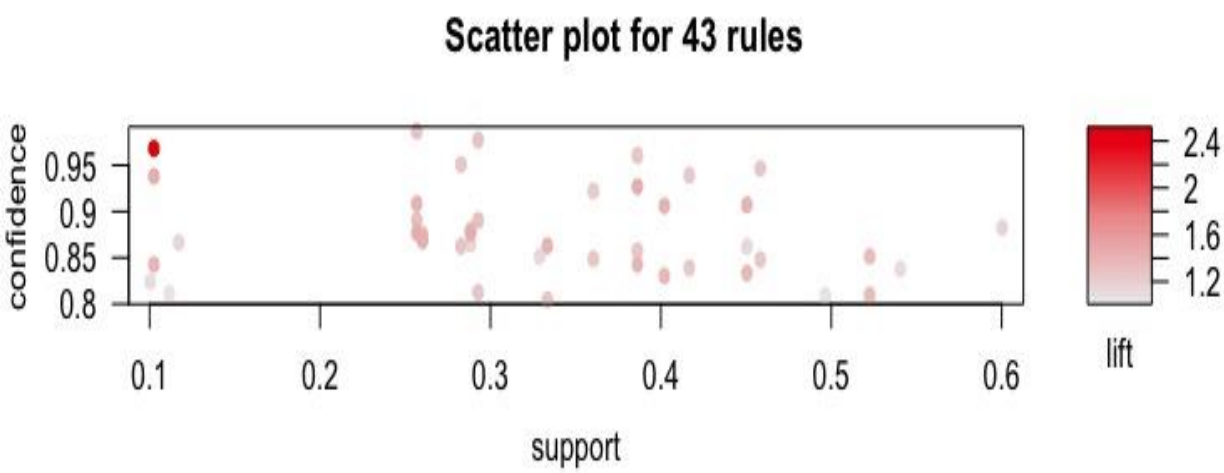
- minimum support: supp = 0.1
- minimum confidence: conf = 0.8
- maximum length of rules: maxlen = 10

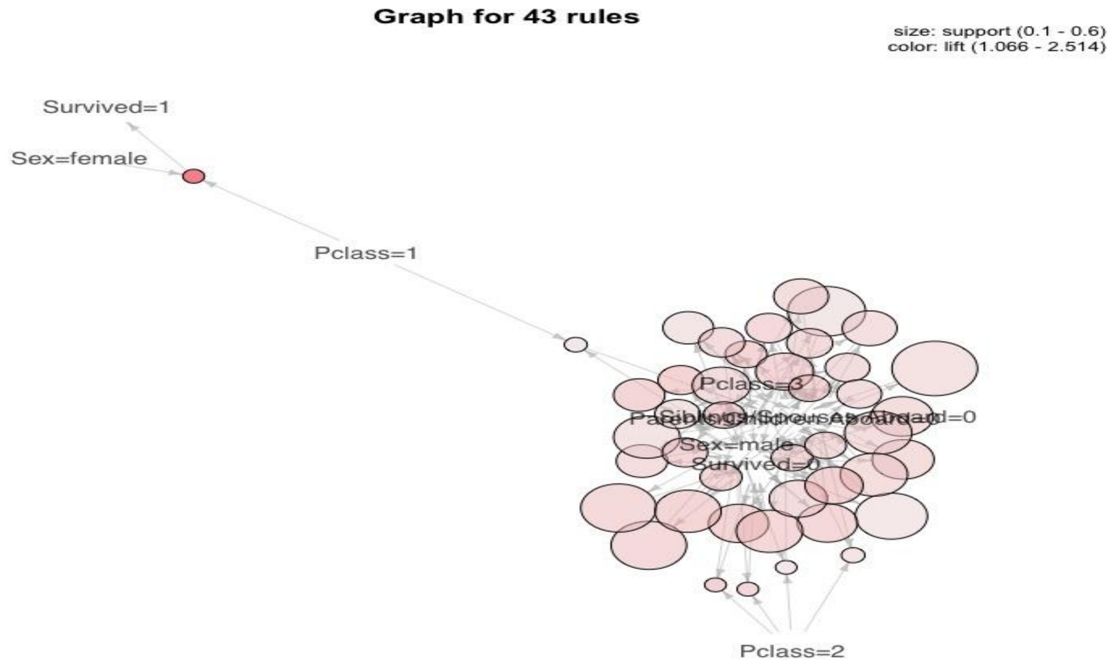
```
> rules.all <- apriori(titanic)
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target
        0.8    0.1    1 none FALSE            TRUE       5     0.1      1     10  rules
   ext
 FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 88

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[1254 item(s), 888 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 done [0.00s].
writing ... [43 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(rules.sorted)
     lhs                                   rhs            support confidence lift  count
[1]  {Class=2nd,Age=Child}              => {Survived=Yes} 0.011   1.000      3.096  24
[2]  {Class=2nd,Sex=Female,Age=Child}   => {Survived=Yes} 0.006   1.000      3.096  13
[3]  {Class=1st,Sex=Female}             => {Survived=Yes} 0.064   0.972      3.010 141
[4]  {Class=1st,Sex=Female,Age=Adult}   => {Survived=Yes} 0.064   0.972      3.010 140
[5]  {Class=2nd,Sex=Female}             => {Survived=Yes} 0.042   0.877      2.716  93
[6]  {Class=Crew,Sex=Female}            => {Survived=Yes} 0.009   0.870      2.692  20
[7]  {Class=Crew,Sex=Female,Age=Adult}  => {Survived=Yes} 0.009   0.870      2.692  20
[8]  {Class=2nd,Sex=Female,Age=Adult}   => {Survived=Yes} 0.036   0.860      2.663  80
[9]  {Class=2nd,Sex=Male,Age=Adult}     => {Survived=No}  0.070   0.917      1.354 154
[10] {Class=2nd,Sex=Male}               => {Survived=No}  0.070   0.860      1.271 154
[11] {Class=3rd,Sex=Male,Age=Adult}     => {Survived=No}  0.176   0.838      1.237 387
[12] {Class=3rd,Sex=Male}               => {Survived=No}  0.192   0.827      1.222 422
```

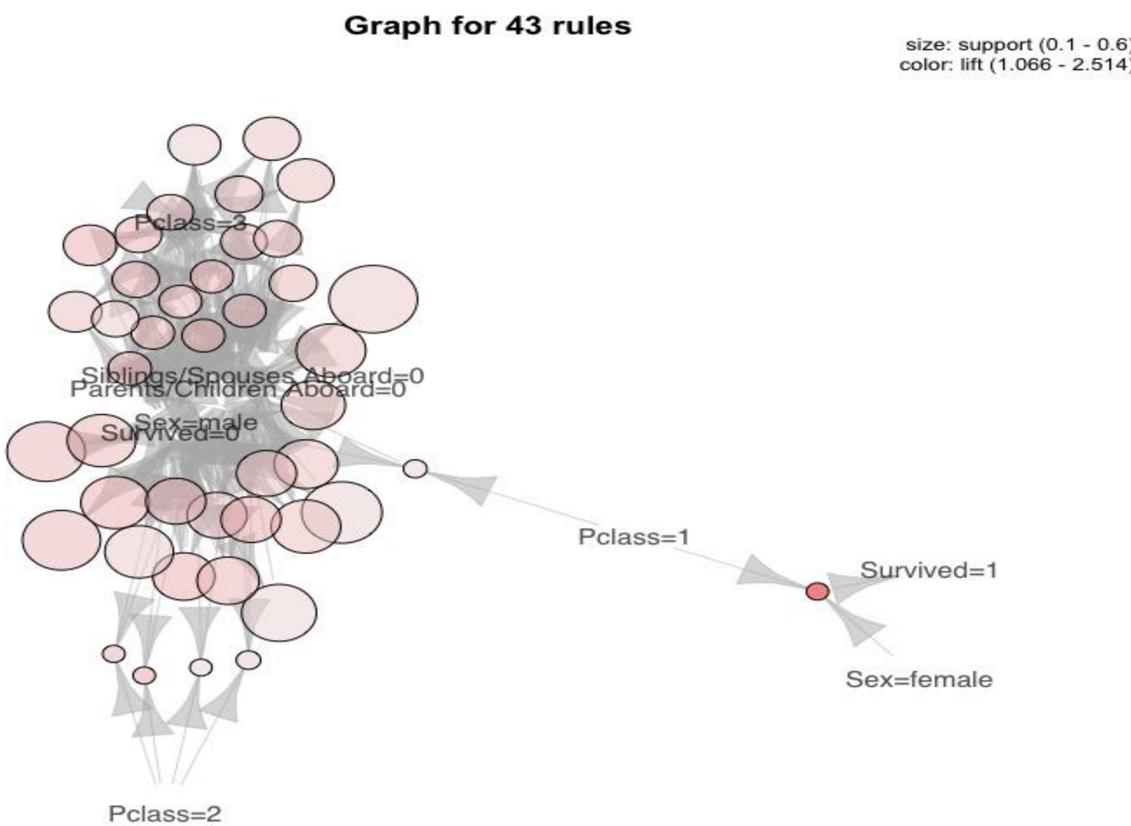
```
> inspect(rules.sorted[1:2])
    lhs                                 rhs            support confidence lift  count
[1] {Class=2nd,Age=Child}            => {Survived=Yes} 0.011   1          3.096 24
[2] {Class=2nd,Sex=Female,Age=Child} => {Survived=Yes} 0.006   1          3.096 13
```

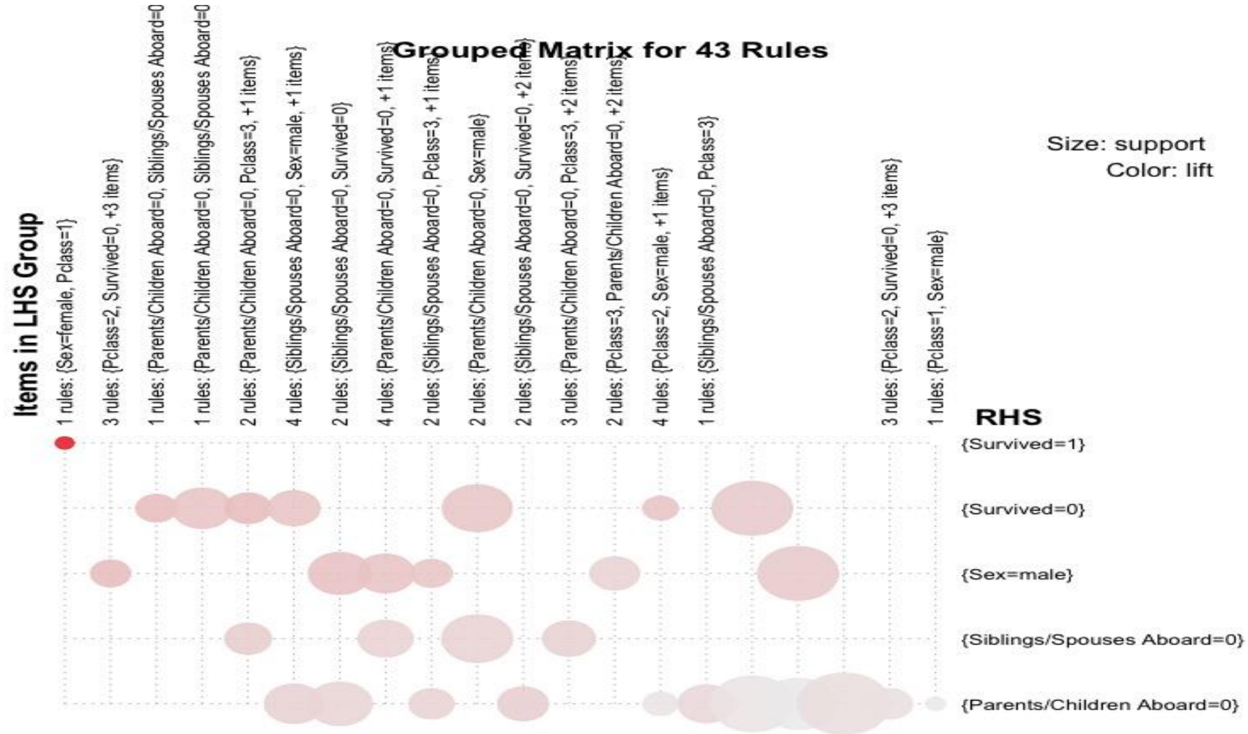
Rule: 2 does not give any special knowledge except Rule: 1, because rules: 1 tells us that all the 3nd class children have survived. When a rule (such as: 2) is a super rule of the second rule (: 1) and the former has the same or lower lift, the former rule (: 2) is considered unnecessary. I have the comparison 4, 7 and 8 rules respectively with other unnecessary rules 3, 6 and 5 in the above results.

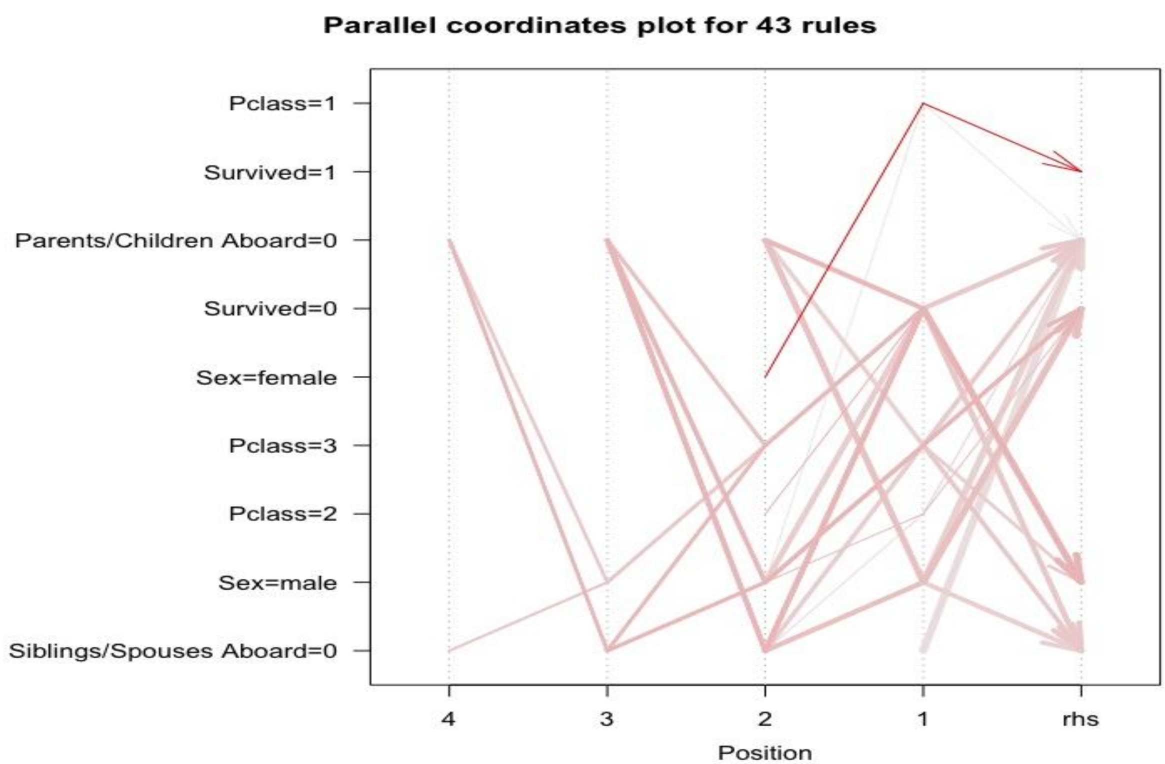
```
> rules_lift <- sort (rules, by="lift", decreasing=TRUE) # 'high-lift' rules.
> inspect(head(rules_lift))
    lhs                        rhs        support confidence lift  count
[1] {Fare=73.5}             => {Pclass=2} 0.006   1          4.826  5
[2] {Fare=21}               => {Pclass=2} 0.007   1          4.826  6
[3] {Fare=26.25}            => {Pclass=2} 0.007   1          4.826  6
[4] {Fare=10.5}             => {Pclass=2} 0.027   1          4.826 24
[5] {Fare=13}               => {Pclass=2} 0.047   1          4.826 42
[6] {Survived=0,Fare=73.5}  => {Pclass=2} 0.006   1          4.826  5
```




Graph for 43 rules
size: support (0.1 - 0.6)
color: lift (1.066 - 2.514)
Survived=1
Sex=female
Pclass=1
Pclass=3
Sex=male
Survived=0
Pclass=2


Graph for 43 rules
size: support (0.1 - 0.6)
color: lift (1.066 - 2.514)
Pclass=3
Siblings/Spouses Aboard=0
Parents/Children Aboard=0
Sex=male
Survived=0
Pclass=1
Survived=1
Sex=female
Pclass=2


Grouped Matrix for 43 Rules
Size: support
Color: lift
Items in LHS Group
1 rules: {Sex=female, Pclass=1}
3 rules: {Pclass=2, Survived=0, +3 items}
1 rules: {Parents/Children Aboard=0, Siblings/Spouses Aboard=0
1 rules: {Parents/Children Aboard=0, Siblings/Spouses Aboard=0
2 rules: {Parents/Children Aboard=0, Pclass=3, +1 items}
4 rules: {Siblings/Spouses Aboard=0, Sex=male, +1 items}
2 rules: {Siblings/Spouses Aboard=0, Survived=0}
4 rules: {Parents/Children Aboard=0, Survived=0, +1 items}
2 rules: {Siblings/Spouses Aboard=0, Pclass=3, +1 items}
2 rules: {Parents/Children Aboard=0, Sex=male}
2 rules: {Siblings/Spouses Aboard=0, Survived=0, +2 items}
3 rules: {Parents/Children Aboard=0, Pclass=3, +2 items}
2 rules: {Pclass=3, Parents/Children Aboard=0, +2 items}
4 rules: {Pclass=2, Sex=male, +1 items}
1 rules: {Siblings/Spouses Aboard=0, Pclass=3}
3 rules: {Pclass=2, Survived=0, +3 items}
1 rules: {Pclass=1, Sex=male}
RHS
{Survived=1}
{Survived=0}
{Sex=male}
{Siblings/Spouses Aboard=0}
{Parents/Children Aboard=0}


Parallel coordinates plot for 43 rules
Pclass=1
Survived=1
Parents/Children Aboard=0
Survived=0
Sex=female
Pclass=3
Pclass=2
Sex=male
Siblings/Spouses Aboard=0
4
3
2
1
rhs
Position