

```

R Console
Terms      Docs
analysis   1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
applications 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
code       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
computing  0 0 1 1 0 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0
data       1 1 0 0 2 0 0 0 0 0 1 2 1 1 1 0 1 0 0 0 0 0 0
examples   0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

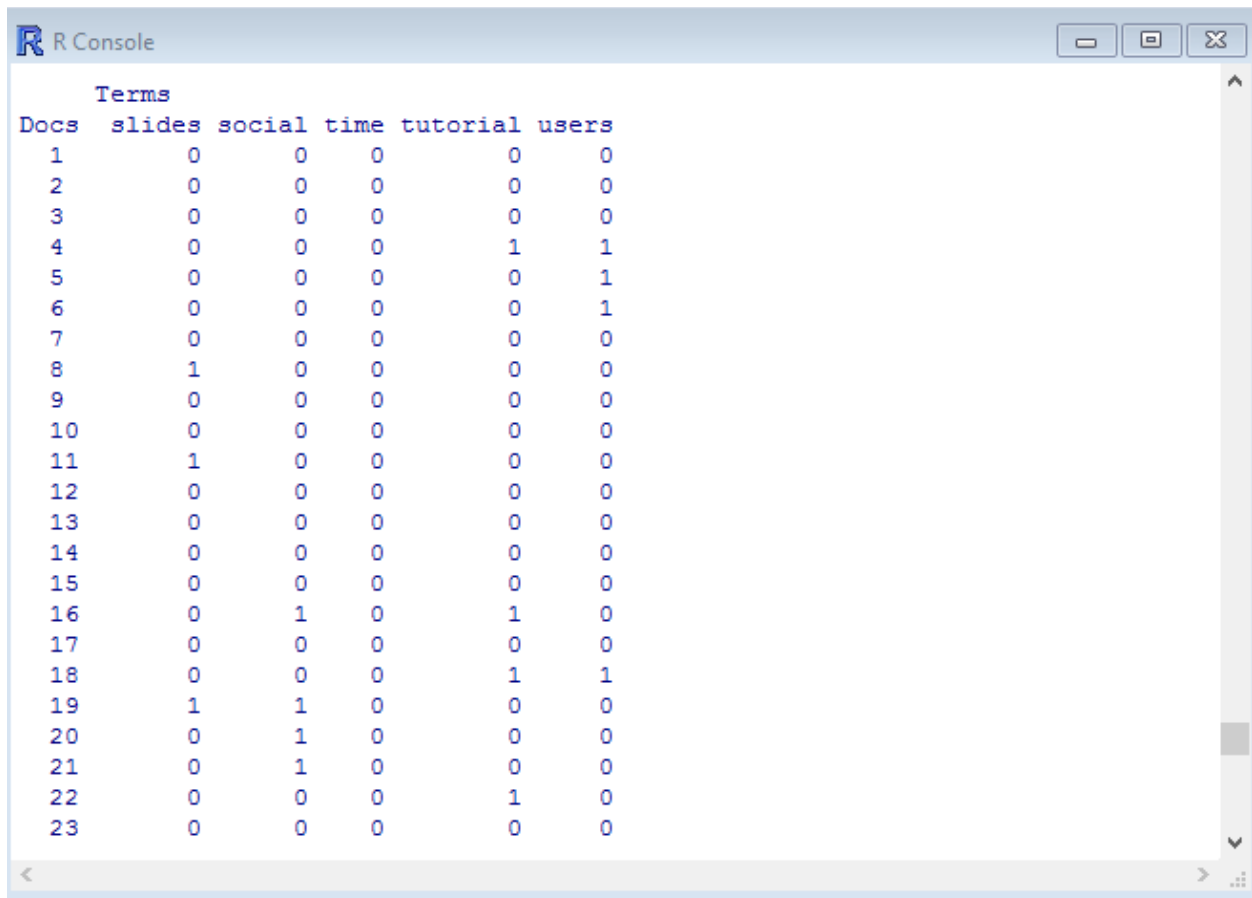
Terms      Docs
analysis   25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
applications 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0
code       1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
computing  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
data       0 0 0 1 0 0 0 1 0 0 1 1 0 0 0 0 0 1 0 0 1
examples   0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0

Terms      Docs
analysis   46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66
applications 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
code       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
computing  0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
data       1 1 0 0 1 0 0 0 1 1 0 0 1 1 0 1 0 1 0 1 1
examples   0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0

```

#create tData variable with transposed data set \*(Note that terms are now columns and docs are rows)

```
tData <- t(termDocMatrix)
```



The image shows an R Console window with a table of data. The table has 6 columns labeled 'Terms' (slides, social, time, tutorial, users) and 23 rows labeled 'Docs' (1 to 23). The data is a binary matrix where 1 indicates a term is present in a document and 0 indicates it is not.

Docs	slides	social	time	tutorial	users
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	1	1
5	0	0	0	0	1
6	0	0	0	0	1
7	0	0	0	0	0
8	1	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	1	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	0	0	0	0
16	0	1	0	1	0
17	0	0	0	0	0
18	0	0	0	1	1
19	1	1	0	0	0
20	0	1	0	0	0
21	0	1	0	0	0
22	0	0	0	1	0
23	0	0	0	0	0

#install arules package

```
install.packages("arules")
```

#load library

```
library("arules")
```

#install arulesViz package

```
install.packages("arulesViz")
```

```
#load arulesViz library
```

```
library(arulesViz)
```

```
ruleset <- apriori(tData,parameter=list(support=0.05,confidence=.5))
```

```
> ruleset <- apriori(tData,parameter=list(support=0.05,confidence=.5))
Apriori

Parameter specification:
 confidence minval smax arem aval originalSupport support minlen maxlen
          0.5   0.1    1 none FALSE             TRUE   0.05      1    10
target      ext
 rules FALSE

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

Absolute minimum support count: 7

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[21 item(s), 154 transaction(s)] done [0.00s].
sorting and recoding items ... [21 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [20 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
Warning message:
In asMethod(object) :
  matrix contains values other than 0 and 1! Setting all entries != 0 to 1.
> |
```

```
#look at rules of ruleset
```

```
summary(ruleset)
```

```

> summary(ruleset)
set of 20 rules

rule length distribution (lhs + rhs):sizes
 2  3
15  5

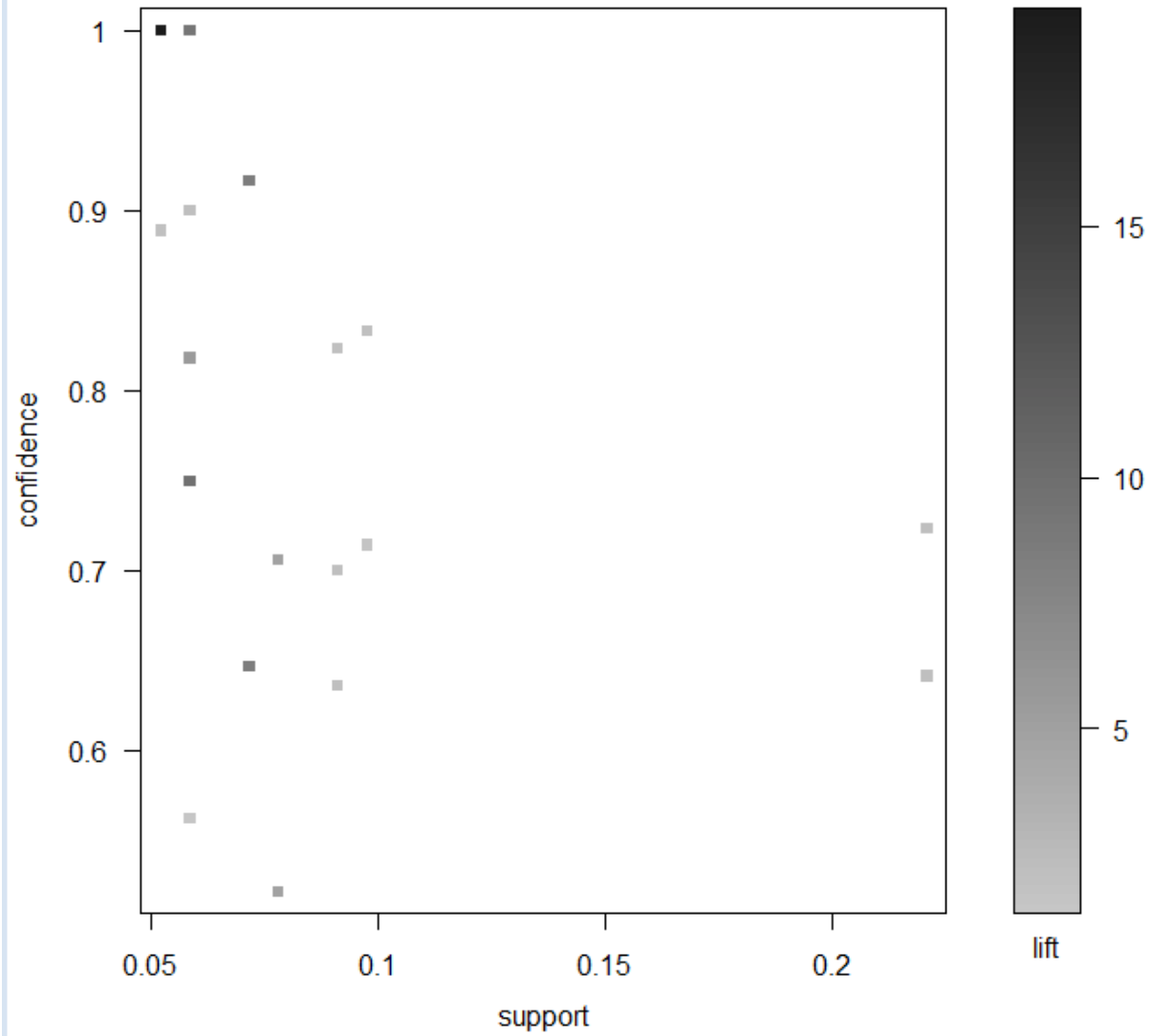
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      2.00   2.00   2.00   2.25   2.25   3.00

summary of quality measures:
      support      confidence      lift
Min.   :0.05195  Min.   :0.5217  Min.   : 1.238
1st Qu.:0.05844  1st Qu.:0.6868  1st Qu.: 1.974
Median :0.07143  Median :0.7500  Median : 3.414
Mean   :0.08571  Mean   :0.7767  Mean   : 5.623
3rd Qu.:0.09091  3rd Qu.:0.8917  3rd Qu.: 8.304
Max.   :0.22078  Max.   :1.0000  Max.   :19.250

mining info:
      data ntransactions support confidence
tData      154      0.05      0.5

```

Plot(ruleset)

**Scatter plot for 20 rules**

inspect(ruleset)

```
R Console
matrix contains values other than 0 and 1! Setting all entries != 0 to 1.
> plot(ruleset)
> inspect(ruleset)
```

	lhs	rhs	support	confidence	lift
1	{code}	=> {r}	0.05194805	0.8888889	1.955556
2	{computing}	=> {r}	0.05844156	0.9000000	1.980000
3	{series}	=> {time}	0.05194805	1.0000000	19.250000
4	{time}	=> {series}	0.05194805	1.0000000	19.250000
5	{social}	=> {network}	0.07142857	0.9166667	8.303922
6	{network}	=> {social}	0.07142857	0.6470588	8.303922
7	{social}	=> {analysis}	0.05844156	0.7500000	5.021739
8	{slides}	=> {r}	0.05844156	0.5625000	1.237500
9	{network}	=> {analysis}	0.07792208	0.7058824	4.726343
10	{analysis}	=> {network}	0.07792208	0.5217391	4.726343
11	{examples}	=> {r}	0.09090909	0.8235294	1.811765
12	{package}	=> {r}	0.09740260	0.7142857	1.571429
13	{users}	=> {r}	0.09740260	0.8333333	1.833333
14	{mining}	=> {data}	0.22077922	0.7234043	2.101967
15	{data}	=> {mining}	0.22077922	0.6415094	2.101967
16	{network,social}	=> {analysis}	0.05844156	0.8181818	5.478261
17	{analysis,social}	=> {network}	0.05844156	1.0000000	9.058824
18	{analysis,network}	=> {social}	0.05844156	0.7500000	9.625000
19	{mining,r}	=> {data}	0.09090909	0.7000000	2.033962
20	{data,r}	=> {mining}	0.09090909	0.6363636	2.085106

```
> |
```

#Rules 3 and 4 show the highest lift, meaning these are the most interesting

#rules. LHS and RHS occur less frequently and support and confidence are

#high, with the series/time association occurring 100% of the time. This is

#useful to find frequent itemsets.