Let's see what's in the basket

MARKET BASKET ANALYSIS IN R

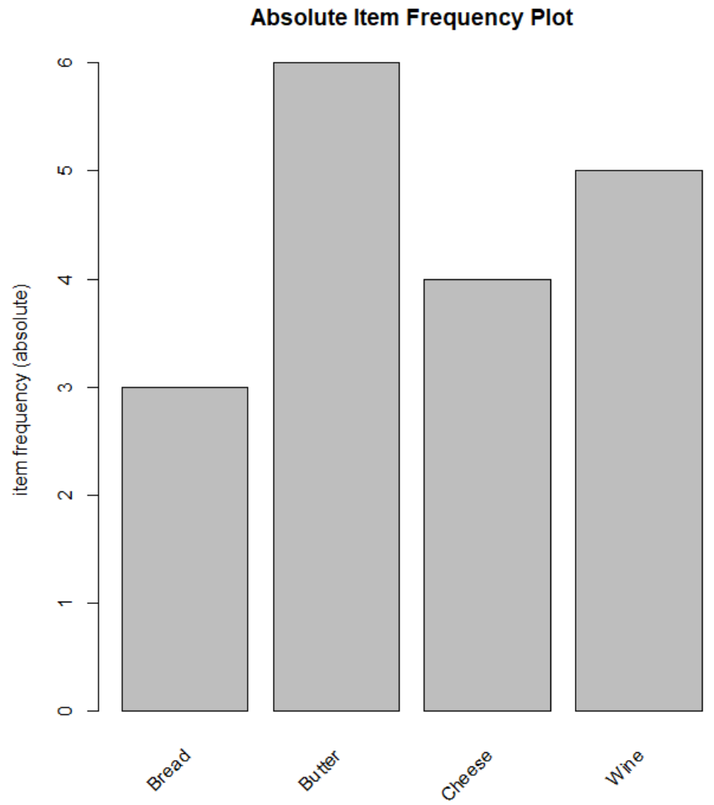
Christopher Bruffaerts
Statistician

datacamp

Visualizing items

TID	Transaction
1	{Bread, Butter, Cheese, Wine}
2	{Bread, Butter, Wine}
3	{Bread, Butter}
4	{Butter, Cheese, Wine}
5	{Butter, Cheese}
6	{Cheese, Wine}
7	{Butter, Wine}

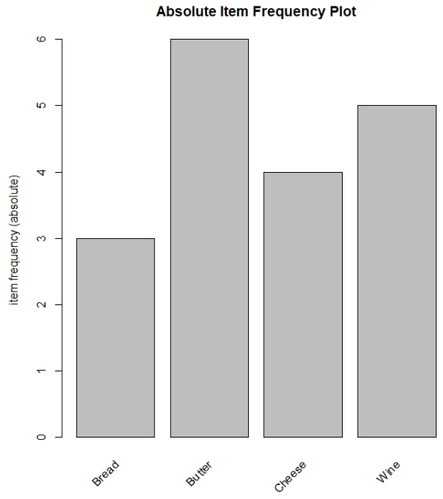
Item Frequency Plot



Visualizing items in R

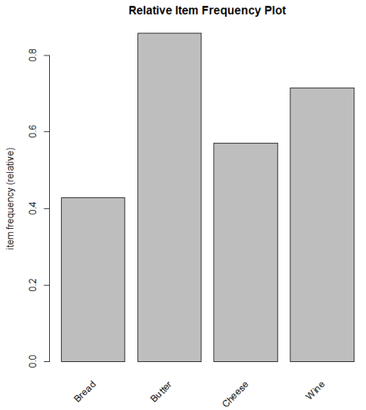
Absolute type

```
itemFrequencyPlot(data_trx,
                  main='Absolute Item Frequency Plot',
                  type="absolute"
)
```



Relative type

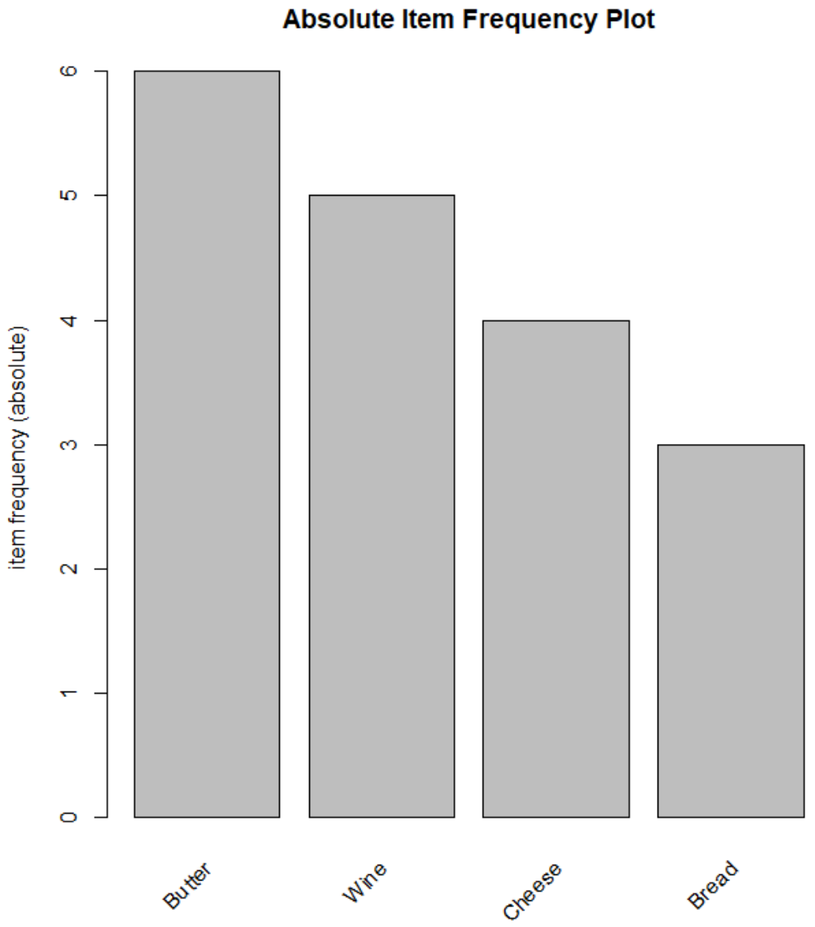
```
itemFrequencyPlot(data_trx,
                  main='Relative Item Frequency Plot',
                  type="relative"
)
```



Top items

Reordering and filtering with topN

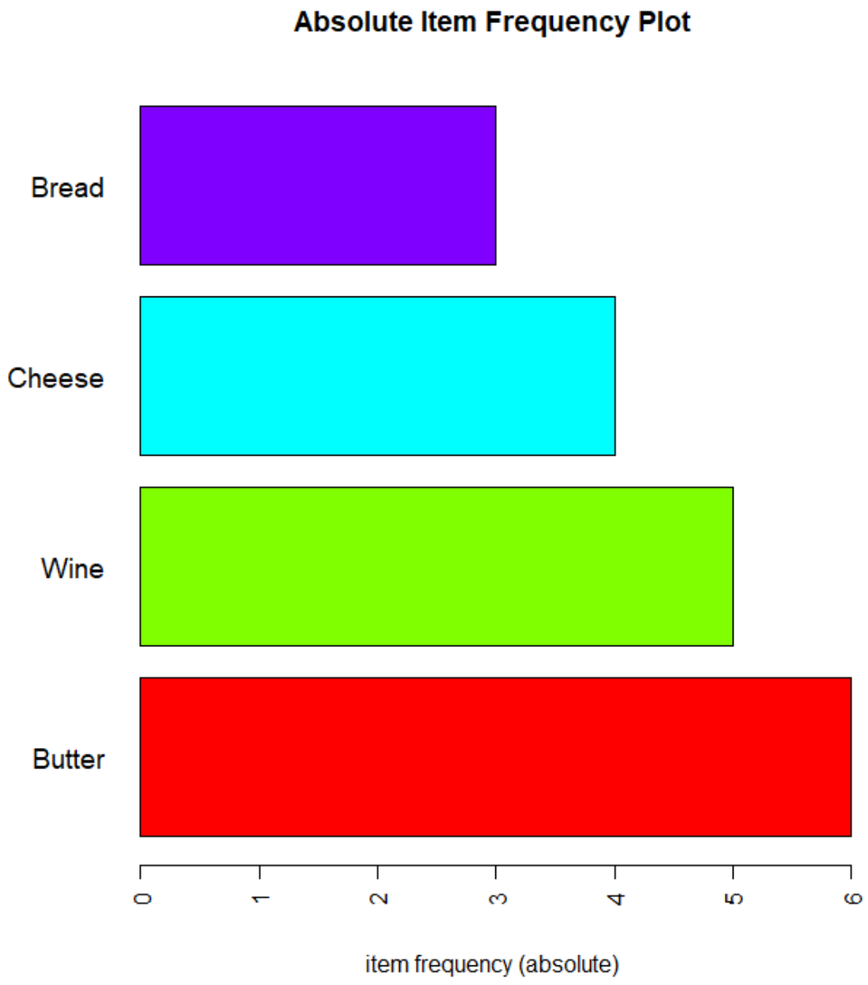
```
itemFrequencyPlot(
  data_trx,
  topN = 4,
  main = 'Absolute Item Frequency Plot',
  type = "absolute"
)
```



Further customization

Flip & customize plot

```
itemFrequencyPlot(
  data_trx,
  topN = 4,
  main = 'Absolute Item Frequency Plot',
  type = "absolute",
  col = rainbow(4),
  ylab = "",
  cex.names = 1.2,
  horiz = TRUE
)
```



Let's plot items!

MARKET BASKET ANALYSIS IN R

Visualizing metrics

MARKET BASKET ANALYSIS IN R

Christopher Bruffaerts
Statistician

datacamp

Interactive table with metrics

Rules from Grocery store

```
library(arules)
rules = apriori(data_trx,
                parameter = list(
                  supp = 3/7, conf=0.6, minlen=2
                )
inspect(rules)
```

HTML table

Show 10 entries Search:

	LHS	RHS	support	confidence	lift	count
	All	All	All	All	All	All
[1]	{Bread}	{Butter}	0.429	1.000	1.167	3.000
[2]	{Cheese}	{Wine}	0.429	0.750	1.050	3.000
[3]	{Wine}	{Cheese}	0.429	0.600	1.050	3.000
[4]	{Cheese}	{Butter}	0.429	0.750	0.875	3.000
[5]	{Wine}	{Butter}	0.571	0.800	0.933	4.000
[6]	{Butter}	{Wine}	0.571	0.667	0.933	4.000

Showing 1 to 6 of 6 entries Previous 1 Next

Interactive table

```
library(arulesViz)
inspectDT(rules)
```

Scatter plots introduction

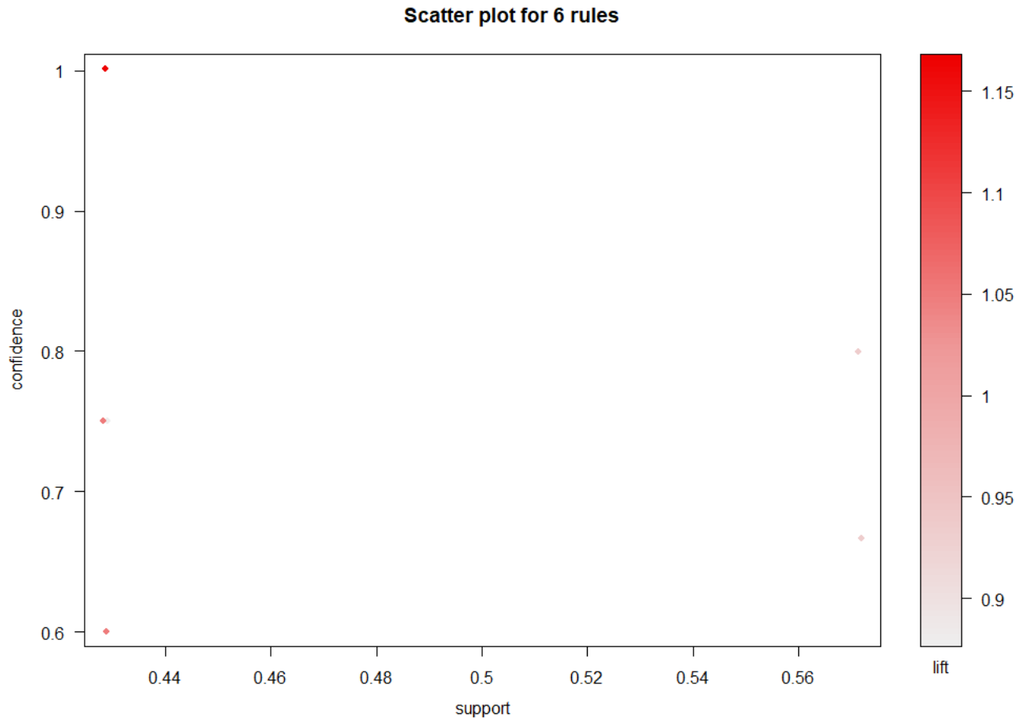
Inspection of the rules

```
inspect(rules)
```

```
    lhs         rhs      support   confidence lift      count
[1] {Bread}  => {Butter} 0.4285714 1.0000000  1.1666667 3
[2] {Cheese} => {Wine}   0.4285714 0.7500000  1.0500000 3
[3] {Wine}   => {Cheese} 0.4285714 0.6000000  1.0500000 3
[4] {Cheese} => {Butter} 0.4285714 0.7500000  0.8750000 3
[5] {Wine}   => {Butter} 0.5714286 0.8000000  0.9333333 4
[6] {Butter} => {Wine}   0.5714286 0.6666667  0.9333333 4
```

Scatterplot from arulesViz

```
plot(rules)
```



Flexible arules plots

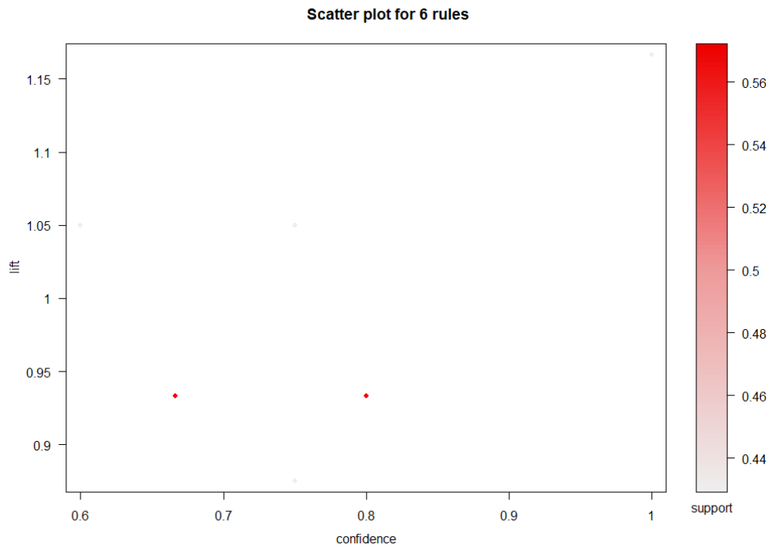
Options of the plot

```
plot(rulesObject, measure, shading, method)
```

- `rulesObject` : the rules object to be plotted
- `measure` : Measures for rule interestingness (Support, Confidence, lift,...)
- `shading` : Measure used to color points.
- `method` : Visualization method to be used (`"scatterplot"` , `"matrix"` , `"two-key plot"` , `"matrix3D"`)

Example

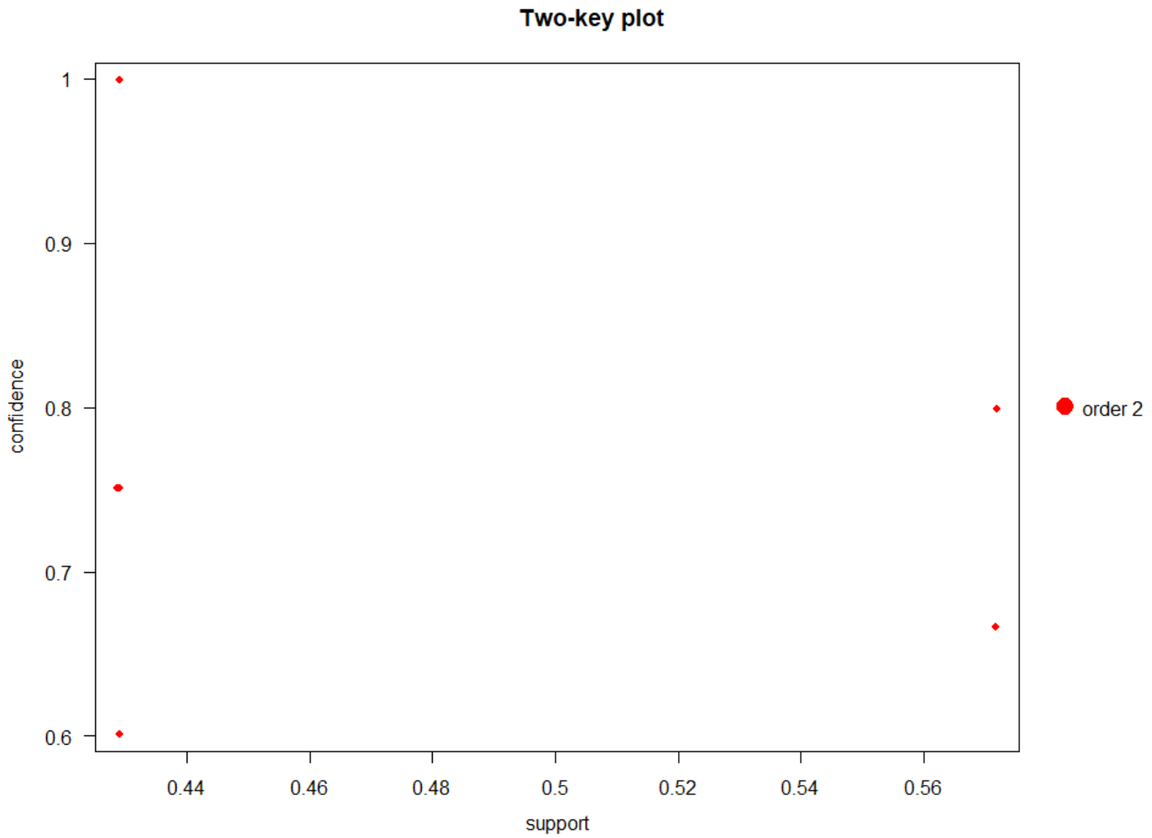
```
plot(rules, measure = c("confidence", "lift"),
     shading = "support",
     method = "scatterplot")
```



Other arules plots

Two-key plot

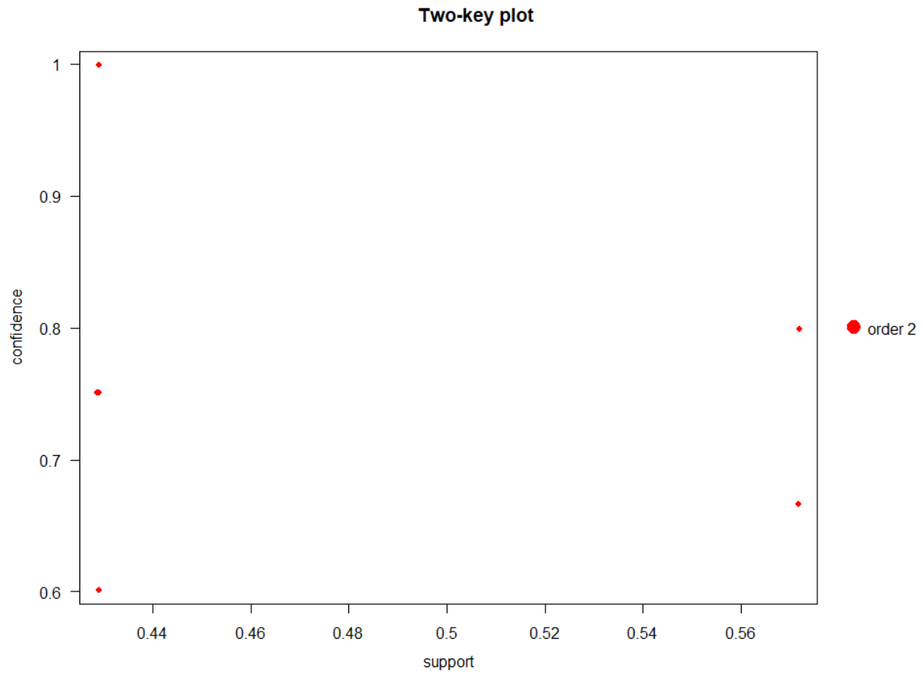
```
plot(rules, method = "two-key plot")
```



Jittering your plots

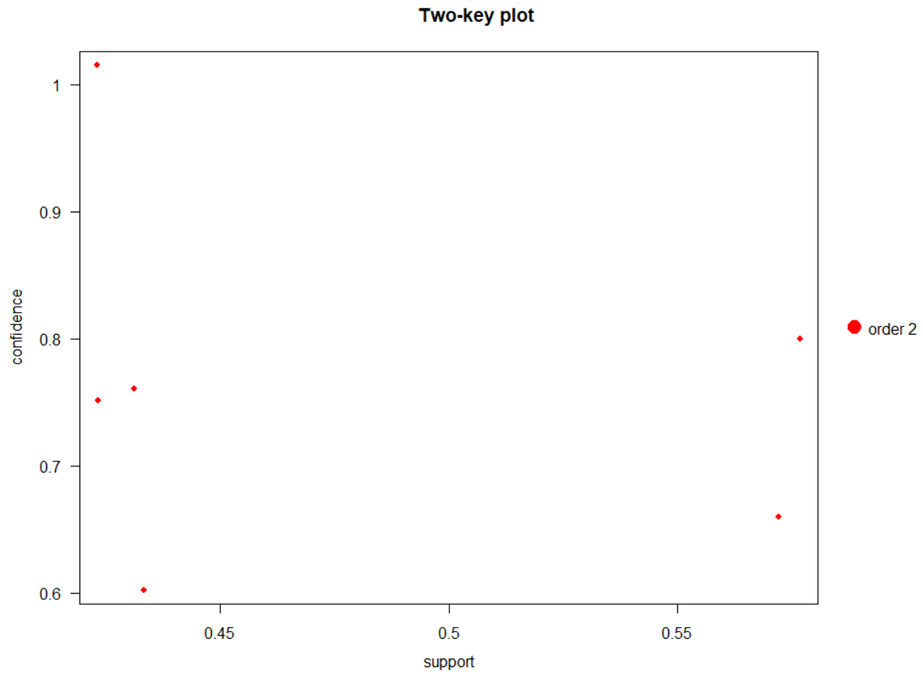
No jittering

```
plot(rules, method = "two-key plot")
```



With jittering

```
plot(rules, method = "two-key plot",jitter = 2)
```

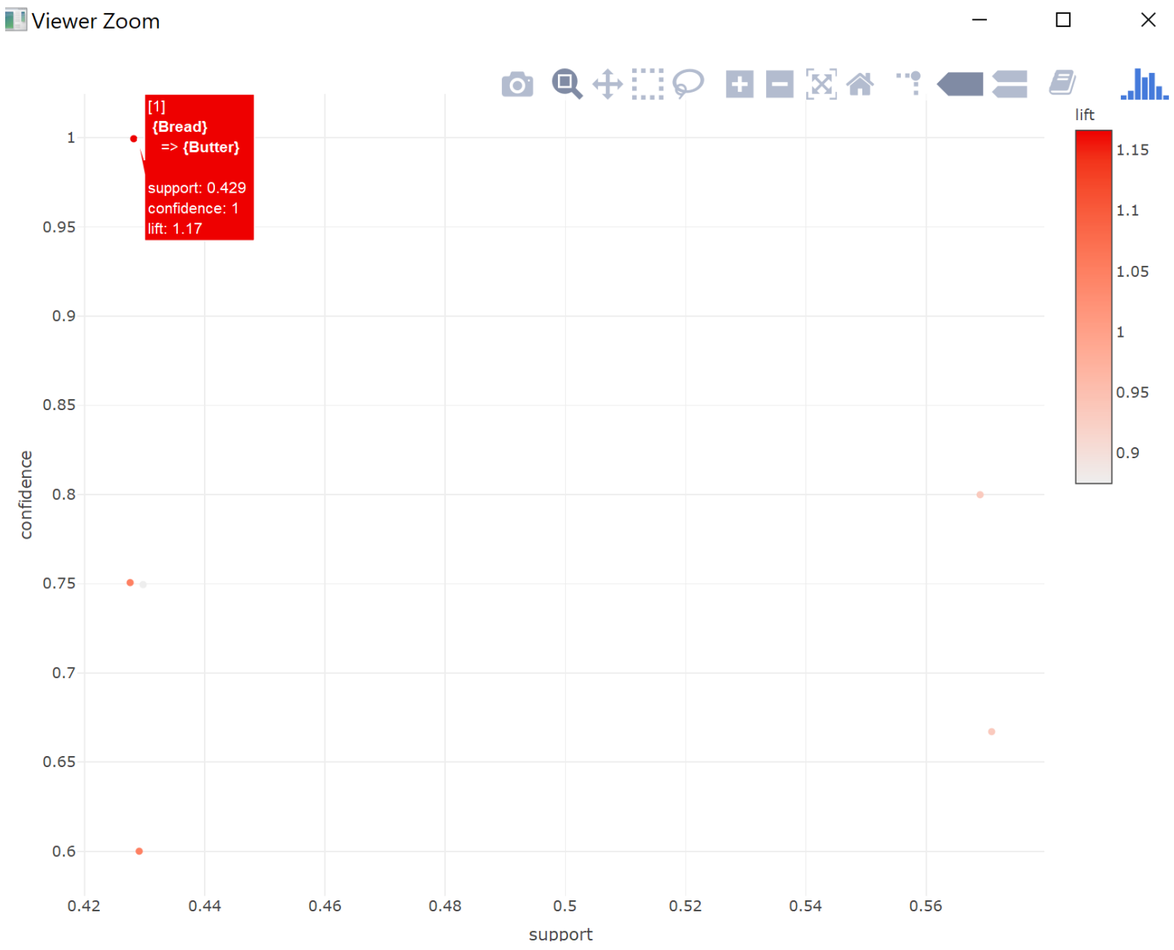


Interactive arules plots

Interactive rules

```
plot(rules, engine = "plotly")
```

From static to interactive



Let's visualize metrics!

MARKET BASKET ANALYSIS IN R

From rules to graph based visualizations

MARKET BASKET ANALYSIS IN R

Christopher Bruffaerts
Statistician

datacamp

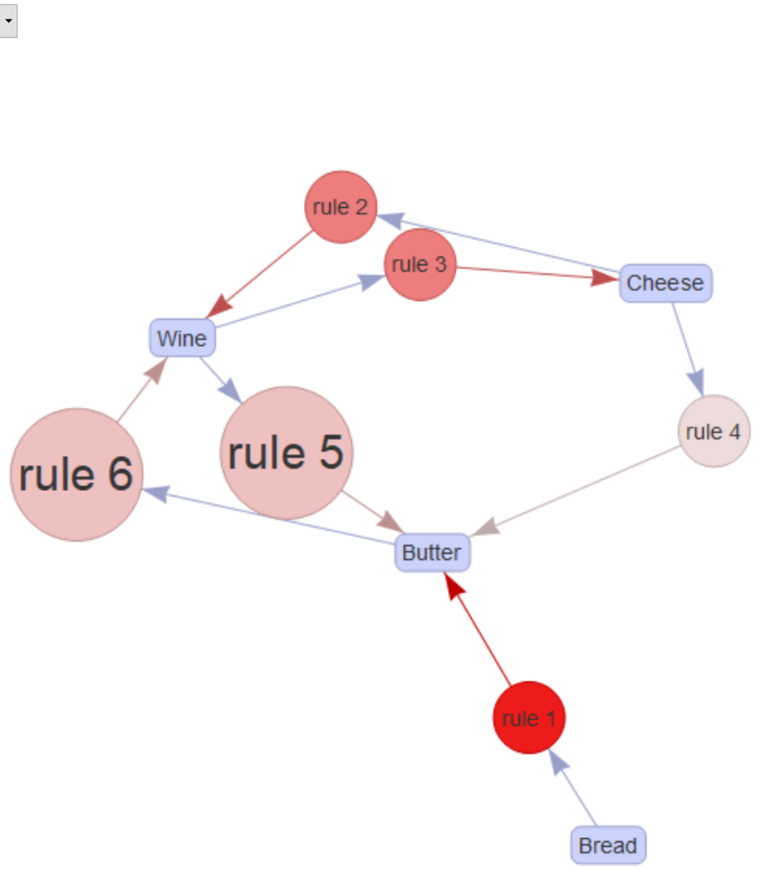
Visualizing rules

Interactive rules

```
rules_html = plot(rules, method = "graph",
                  engine = "htmlwidget")
rules_html
```

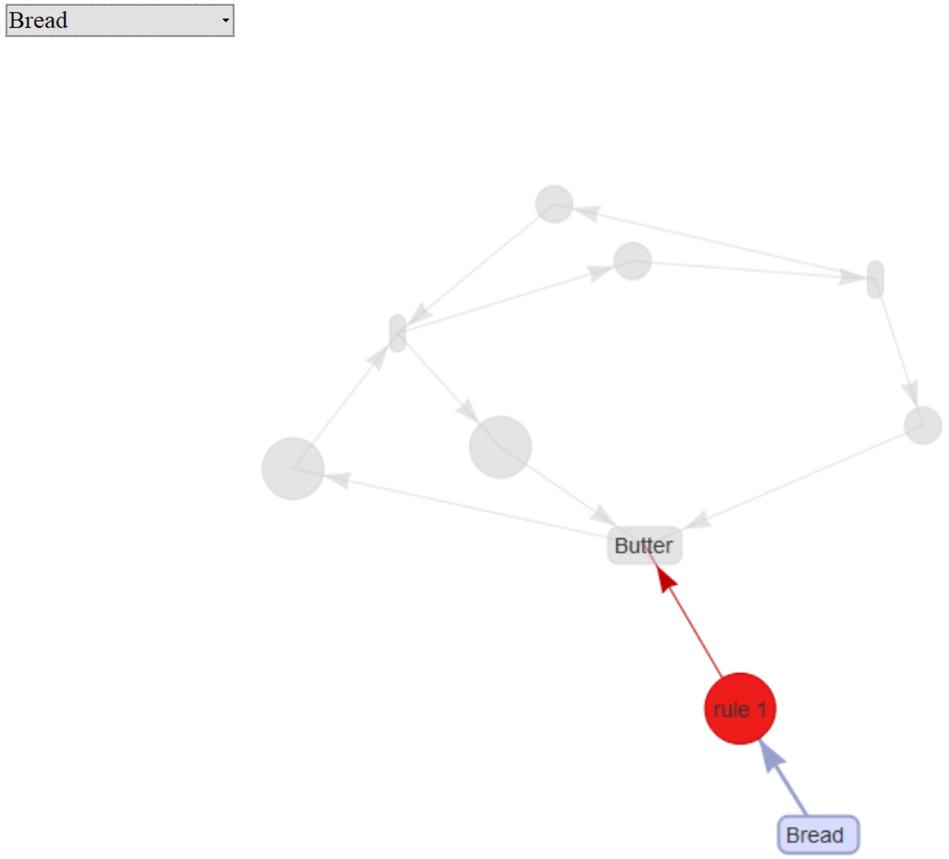
Save the HTML widget

```
library(htmlwidgets)
saveWidget(rules_html, file = "rules_grocery.html")
```

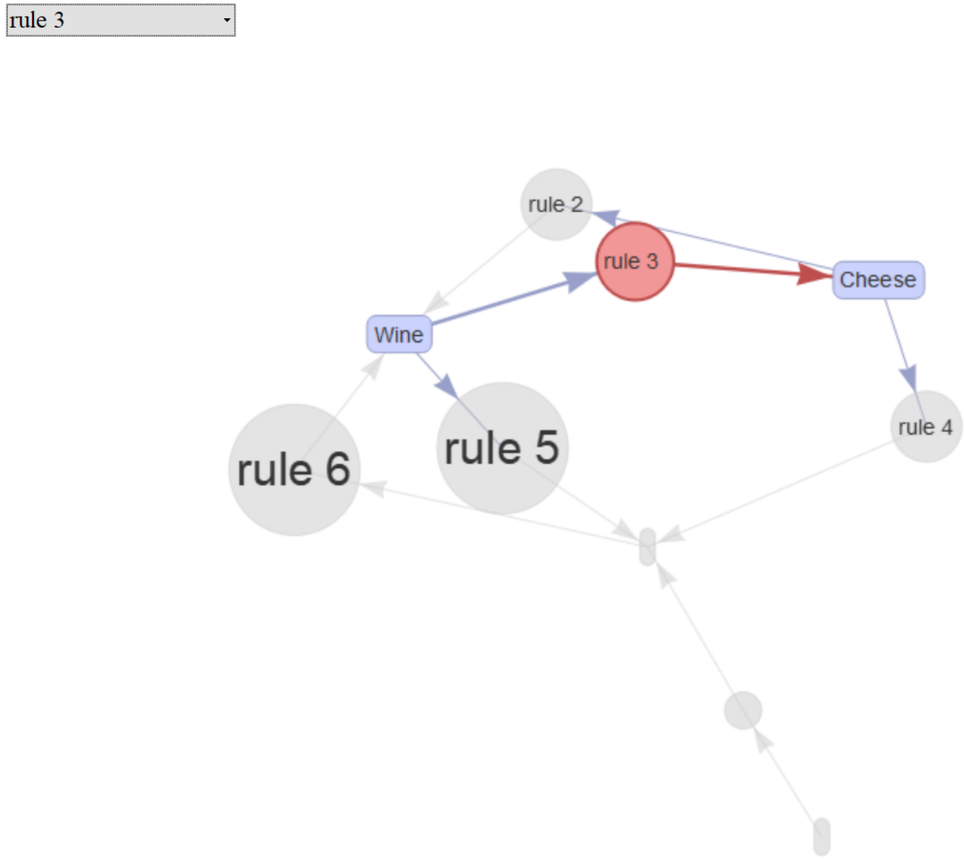


Selecting items and rules from the graph

Select the item Bread



Select Rule 3



Graphs and subgraphs (1)

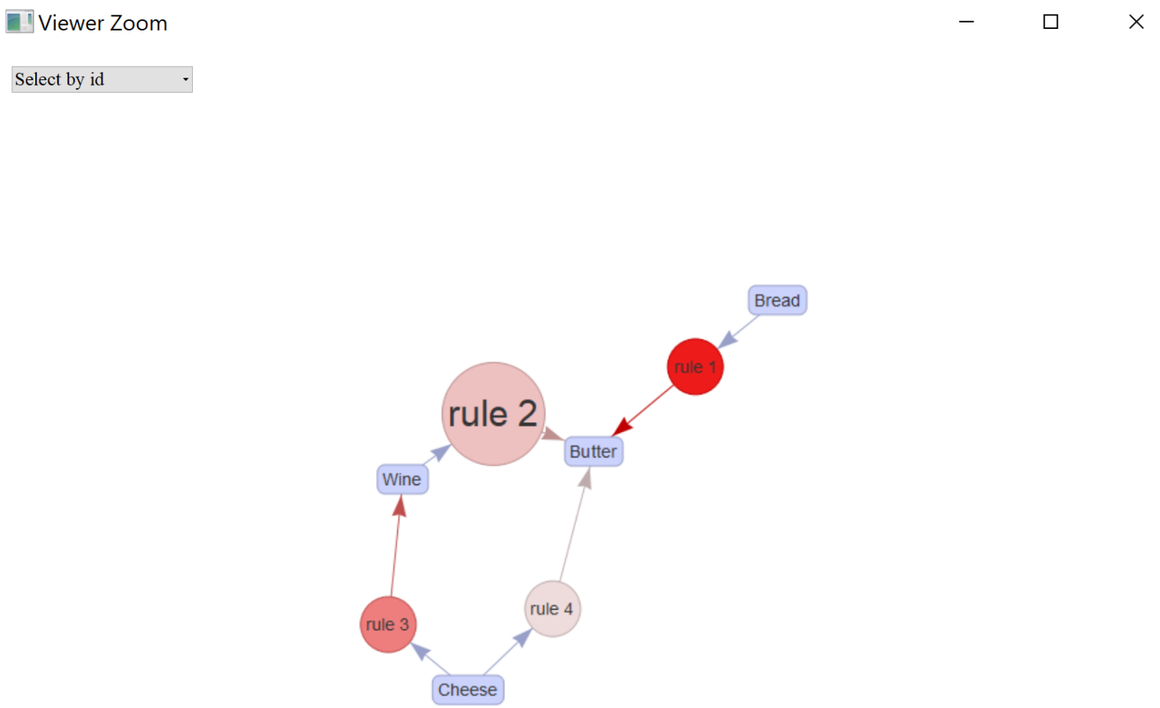
Sort rules by confidence

```
top4subRules =
    head(sort(rules, by = "confidence"), 4)
inspect(top4subRules)
```

```
    lhs         rhs      support   confidence lift      count
[1] {Bread}  => {Butter} 0.4285714 1.00       1.1666667 3
[2] {Wine}   => {Butter} 0.5714286 0.80       0.9333333 4
[3] {Cheese} => {Wine}   0.4285714 0.75       1.0500000 3
[4] {Cheese} => {Butter} 0.4285714 0.75       0.8750000 3
```

Plot high-confidence rules

```
plot(top4subRules, method = "graph",  engine = "htmlwidget")
```



Graphs and subgraphs (2)

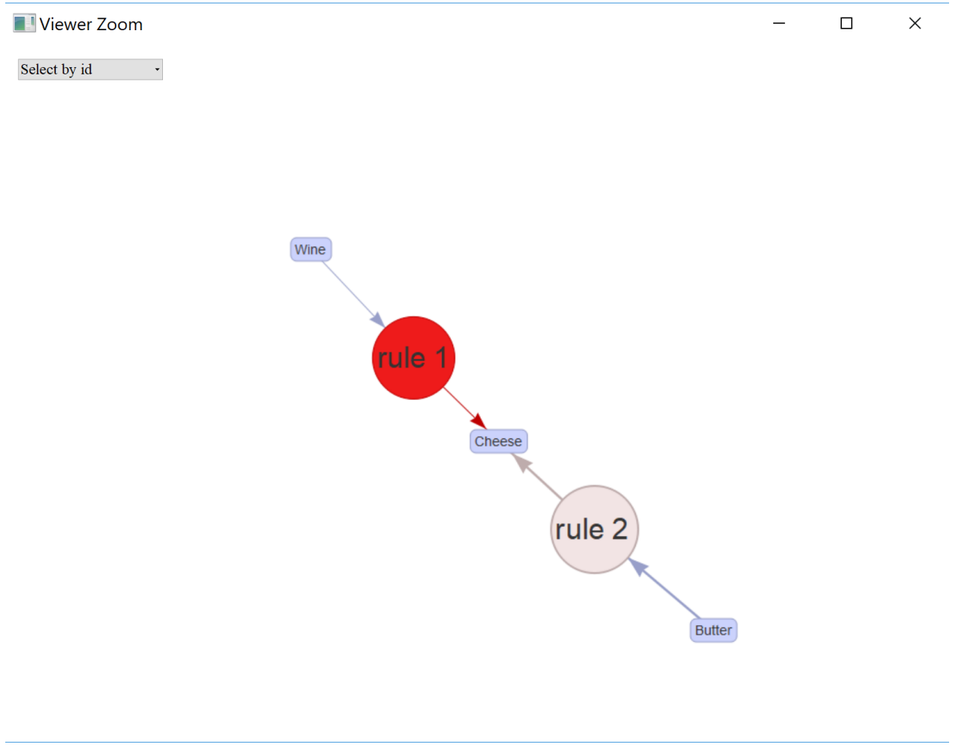
Inspect cheesy rules

```
C_rules =
    apriori(data = data_trx,
            parameter = list(supp = 3/7,
                             conf = 0.2,
                             minlen = 2),
            appearance = list(rhs = "Cheese"))
inspect(C_rules)
```

```
    lhs          rhs       support   confidence lift  coun
[1] {Wine}   => {Cheese} 0.4285714 0.6        1.050 3
[2] {Butter} => {Cheese} 0.4285714 0.5        0.875 3
```

Plot cheesy rules

```
plot(C_rules, method = "graph",  engine = "htmlwidget")
```



Save as Graph

Saving your graph

```
saveAsGraph(rules, file = "rules.graphml")
```

Let's rule!

MARKET BASKET ANALYSIS IN R

Alternative rule plots

MARKET BASKET ANALYSIS IN R

Christopher Bruffaerts
Statistician

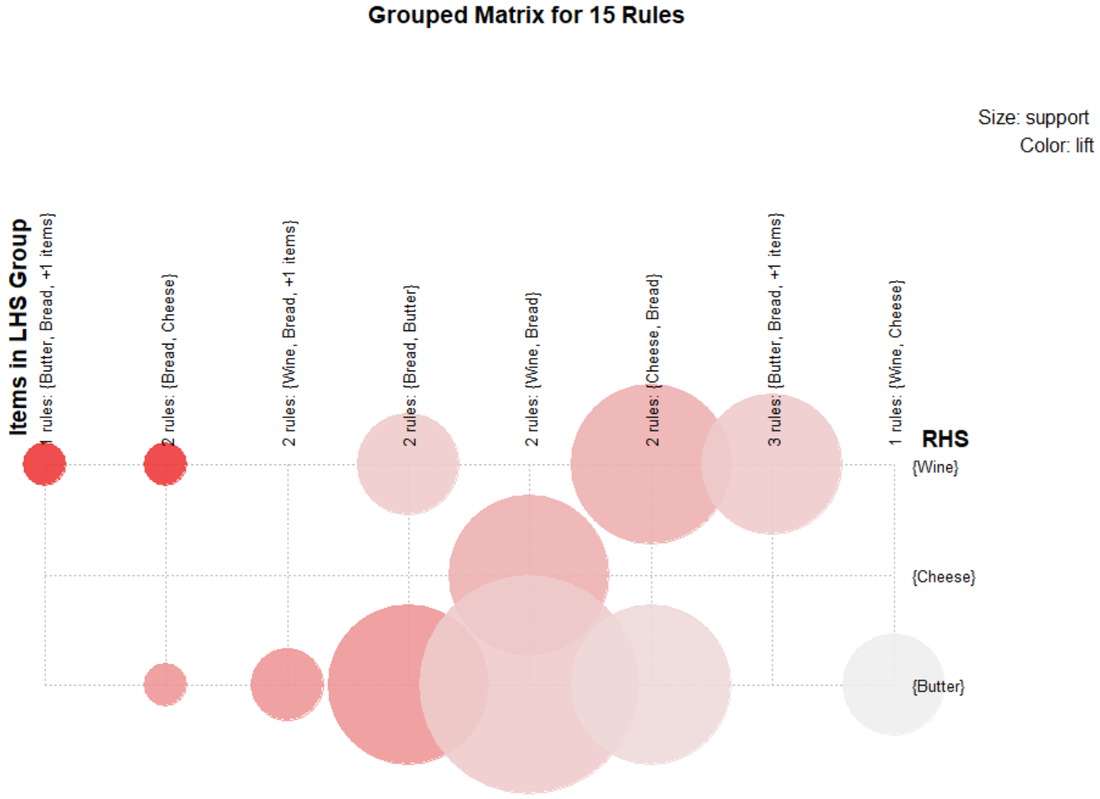
Group-based matrix visualizations

Rule extraction with apriori

```
rules = apriori(data_trx,
                parameter = list(
                  supp = 1/7,
                  conf = 0.6,
                  minlen = 2)
)
```

Method grouped

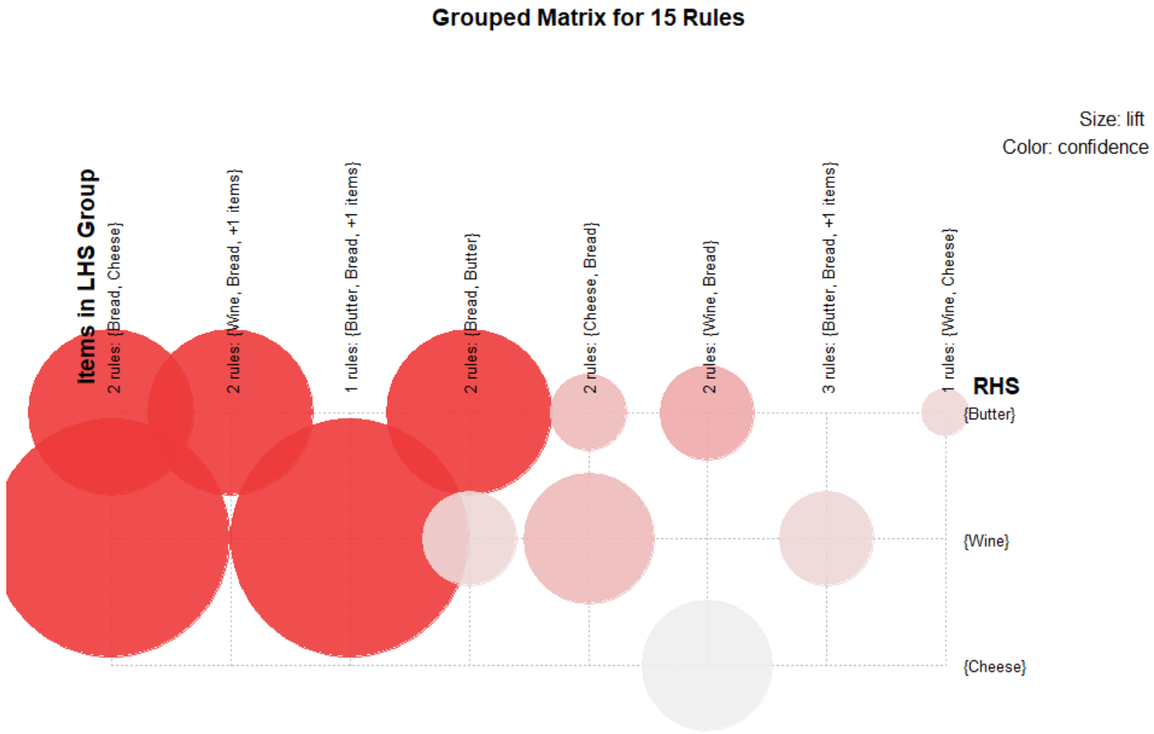
```
plot(rules, method = "grouped")
```



Group-based matrix visualizations

Grouped matrix with different metrics

```
plot(rules,
     method = "grouped",
     measure = "lift",
     shading = "confidence")
```



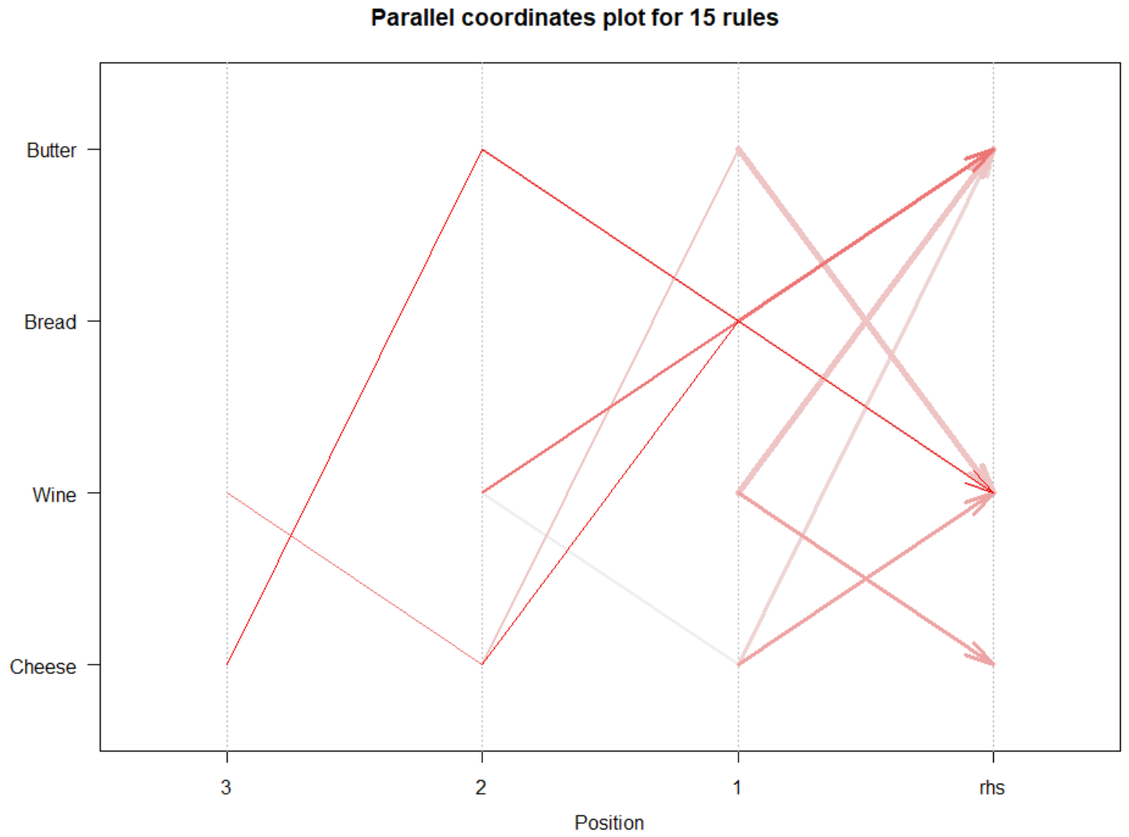
Parallel coordinate plots

Generating rules and calling the plot

```
plot(rules, method = "paracoord")
```

```
     lhs                rhs          support  confidence lift  count
[1]  {Bread}            => {Wine}    0.28     0.66       0.93  2
[2]  {Bread}            => {Butter}  0.42     1          1.16  3
[3]  {Cheese}           => {Wine}    0.42     0.75       1.05  3
[4]  {Wine}             => {Cheese}  0.42     0.6        1.05  3
[5]  {Cheese}           => {Butter}  0.42     0.75       0.87  3
[6]  {Wine}             => {Butter}  0.57     0.8        0.93  4
[7]  {Butter}           => {Wine}    0.57     0.66       0.93  4
[8]  {Bread,Cheese}     => {Wine}    0.14     1          1.4   1
[9]  {Bread,Cheese}     => {Butter}  0.14     1          1.16  1
[10] {Bread,Wine}       => {Butter}  0.28     1          1.16  2
...
```

Parallel coordinates plot



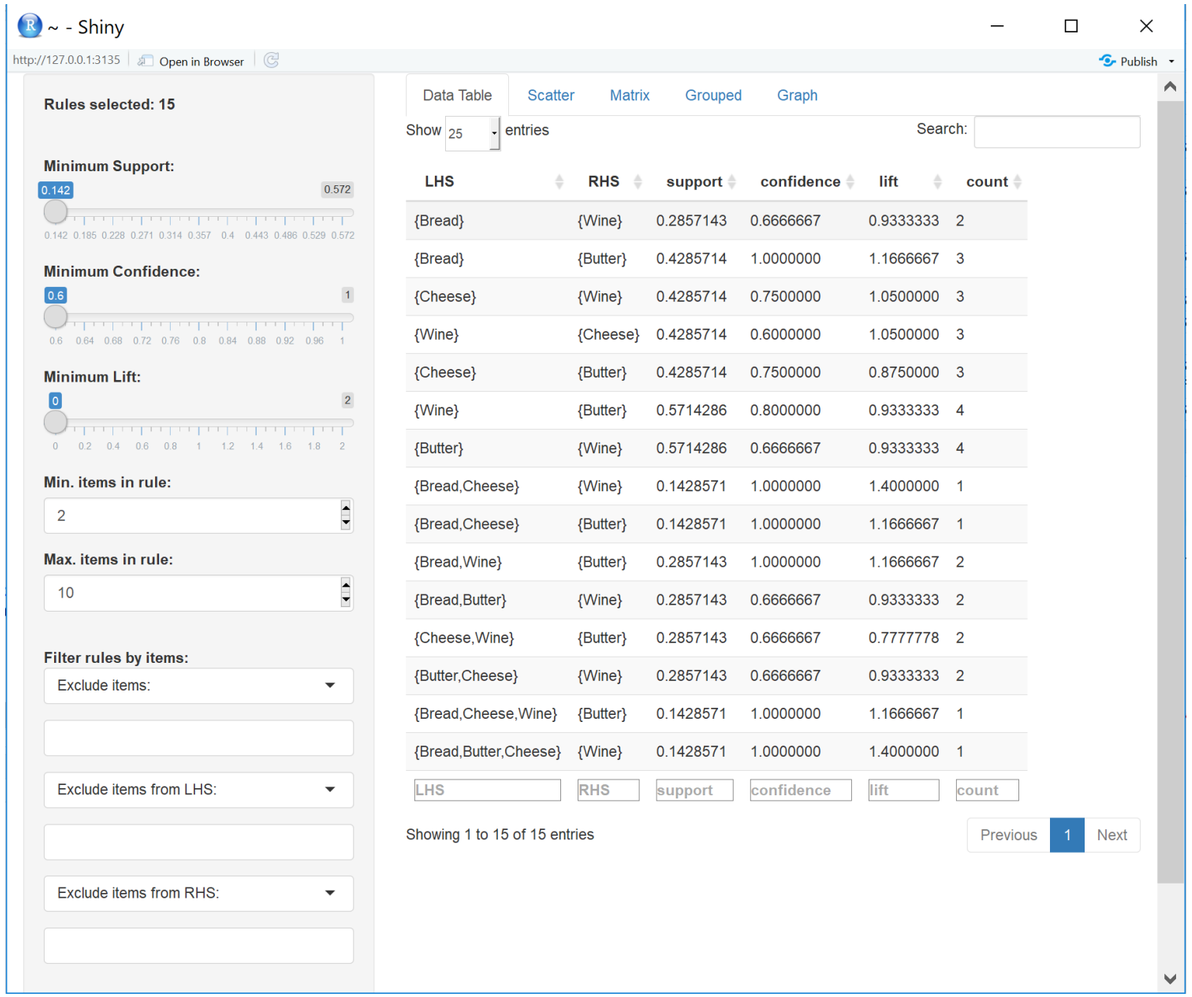
ruleExplorer: the Swiss Army knife

Shiny app

```
ruleExplorer(rules)
```

Available plots:

- Data table
- Scatter
- Matrix
- Grouped
- Graph



More on Shiny

[1] [Link to track:](https://www.datacamp.com/tracks/shiny-fundamentals-with-r)

Let's ruleExplore!

MARKET BASKET ANALYSIS IN R