

Market basket Analysis

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy.

Association Rules are widely used to analyze retail basket or transaction data, and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules.

Abstract

Association rule mining is a popular data mining method to discover interesting relationships between variables in large databases. An extensive toolbox is available in the R-extension package arules. However, mining association rules often results in a vast number of found rules, leaving the analyst with the task to go through a large set of rules to identify interesting ones. Sifting manually through extensive sets of rules is time-consuming and strenuous. Visualization and especially interactive visualization has a long history of making large amounts of data better accessible. The R-extension package arulesViz provides most popular visualization techniques for association rules. In this paper, we discuss recently added interactive visualizations to explore association rules and demonstrate how easily they can be used in arulesViz via a unified interface. With examples, we help to guide the user in selecting appropriate visualizations and interpreting the results.

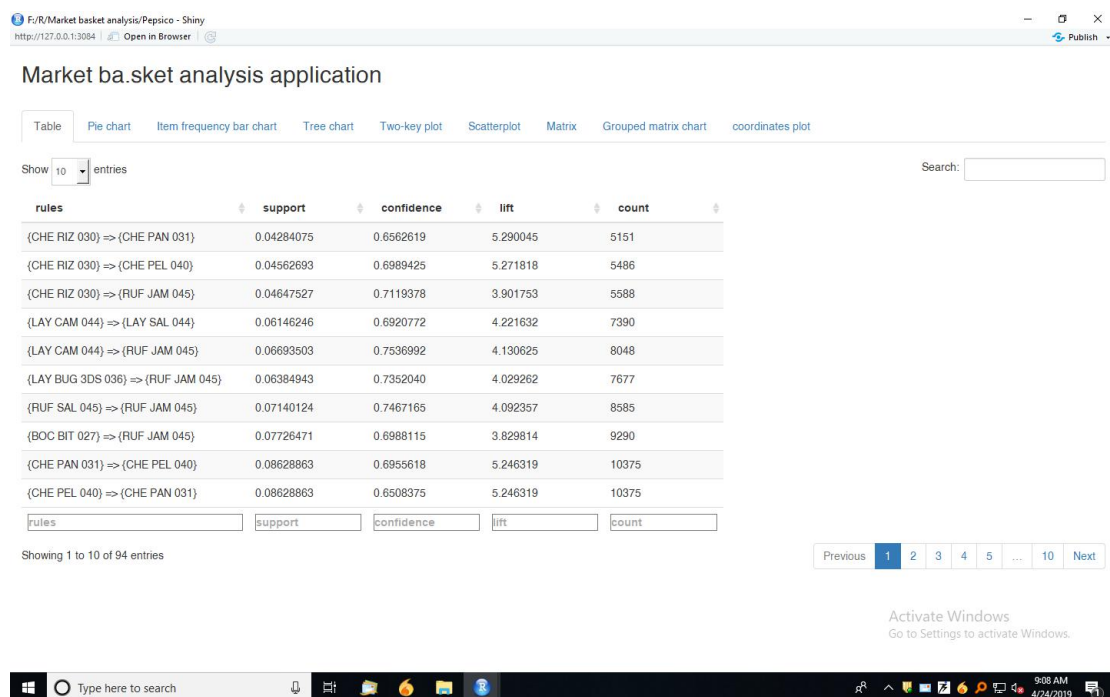


Fig Interactive data table for the mined rule set.

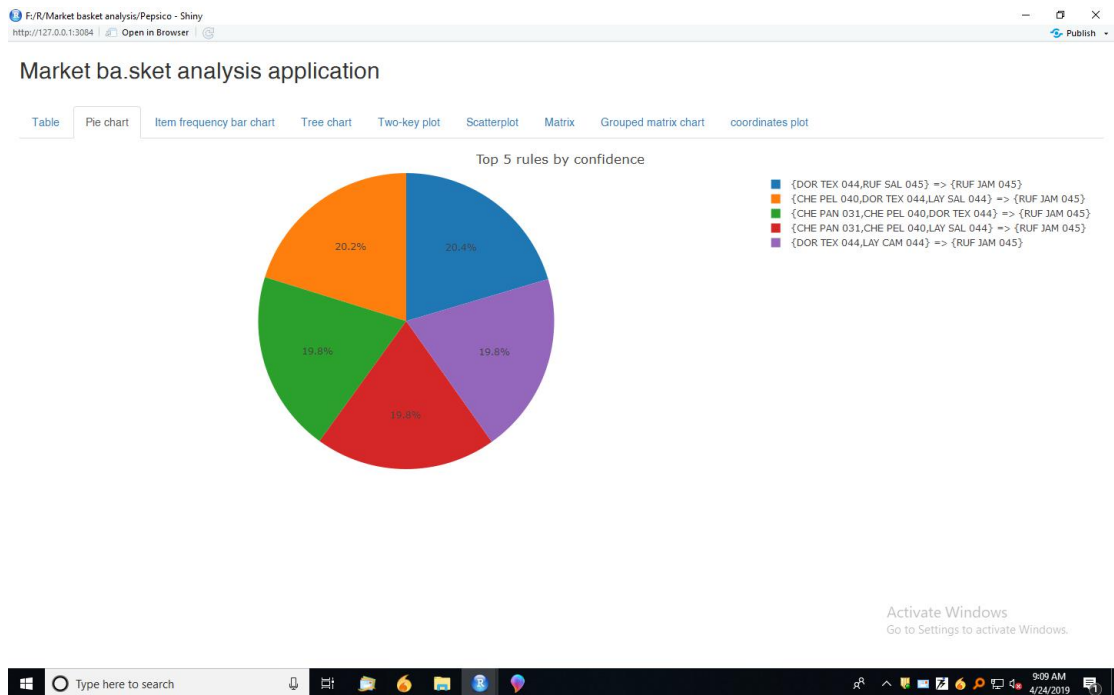


Fig 10.3.2 Top 5 rules by confidence

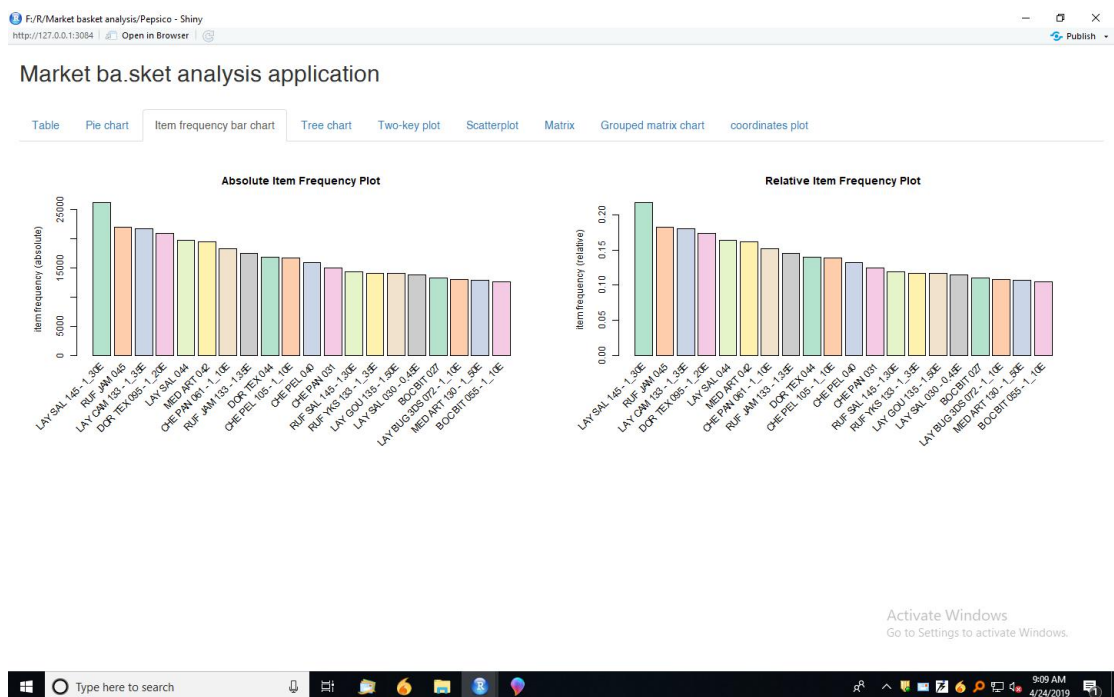


Fig Item frequency plot.

Graph-based visualization

Graph-based techniques concentrate on the relationship between individual items in the rule set. arulesViz offers an interactive visualization based on package visNetwork (Almende B.V. et al., 2017)

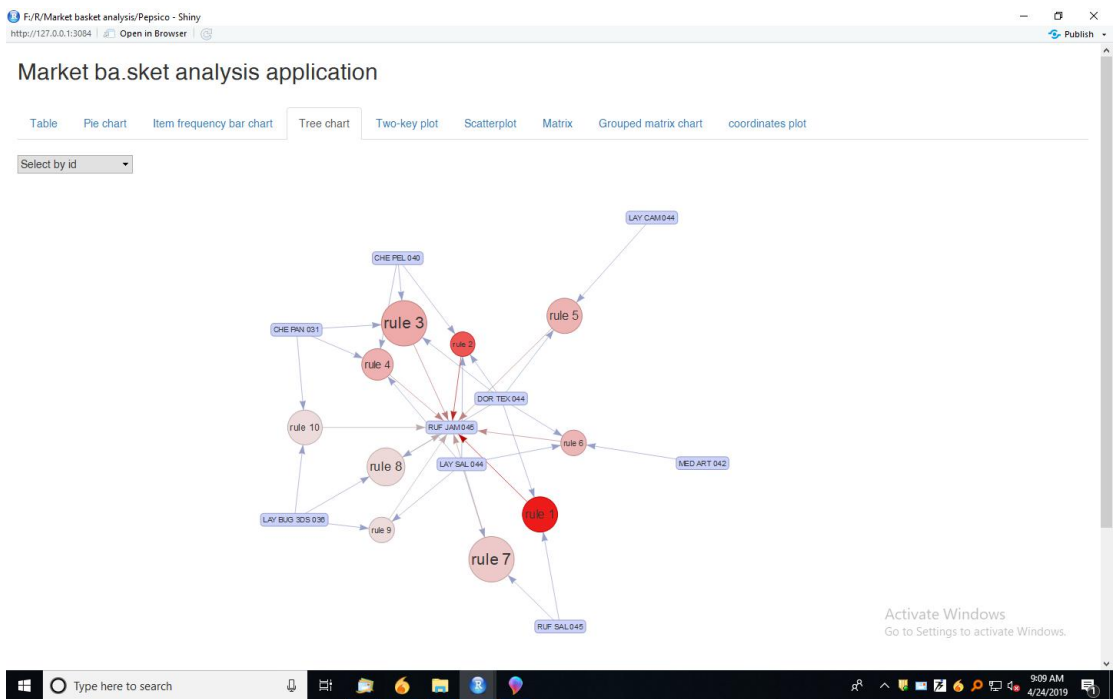


Fig Top 10 rules Tree Graph 01.

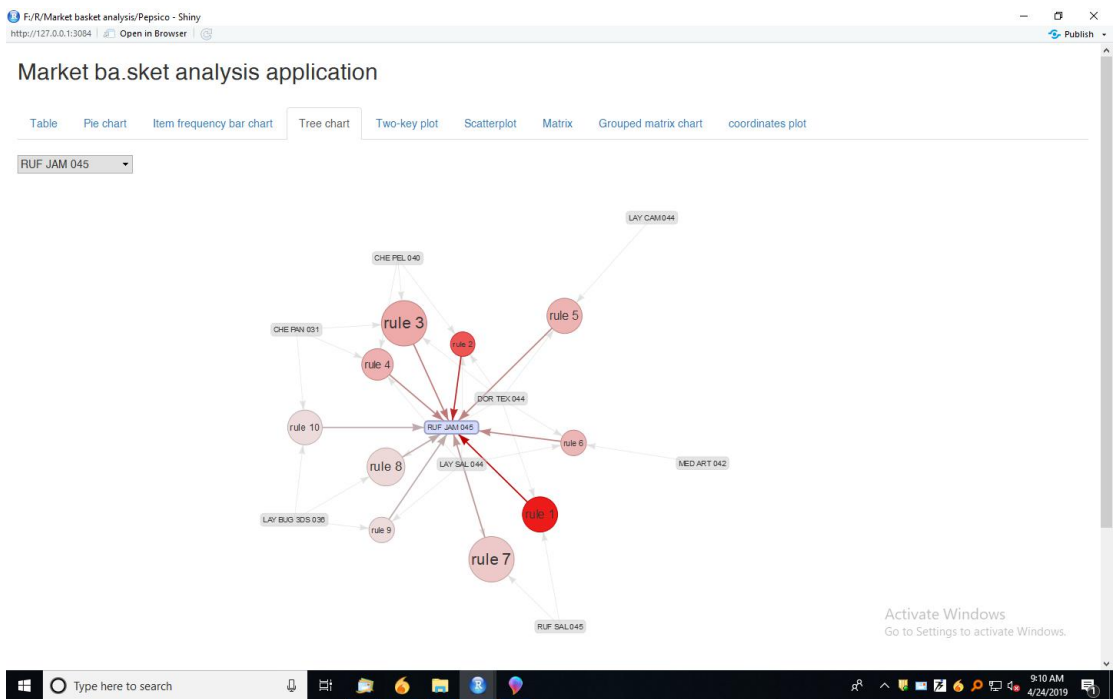


Fig Item association with rules.

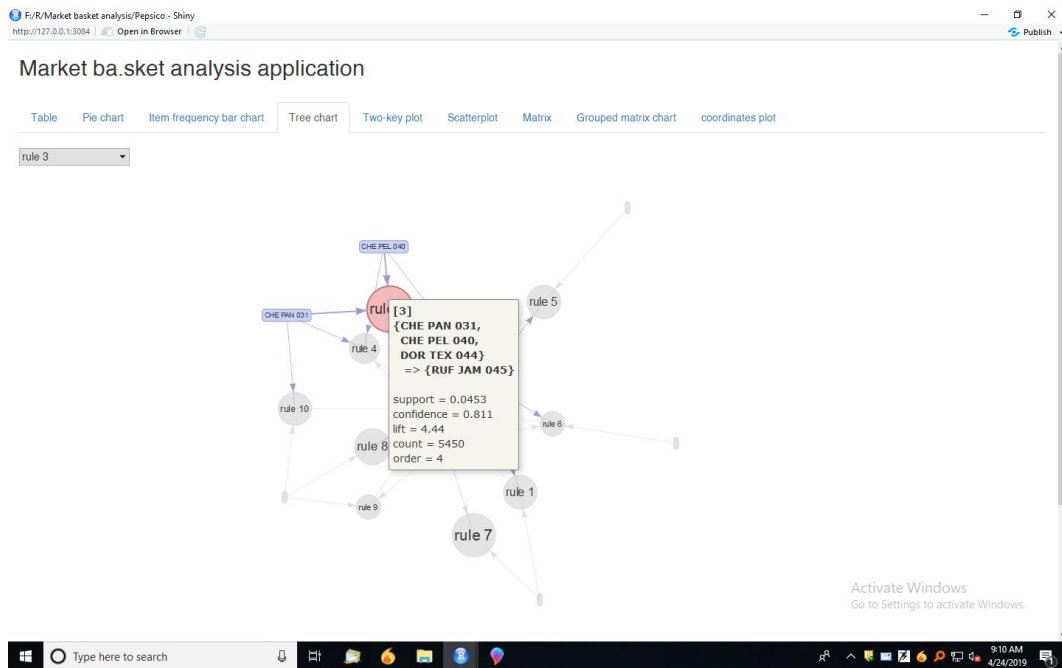


Fig Rule associated with item.

Scatter plot

A straight-forward visualization of association rules is to use a scatter plot with two interest measures on the axes. Such a presentation can be found already in an early paper by Bayardo, Jr. and Agrawal (1999) when they discuss sc-optimal rules.

The default method for plot() for association rules in arulesViz is a scatter plot using support and confidence on the axes. In addition a third measure (default: lift) is used as the color (gray level) of the points. A color key is provided to the right of the plot.

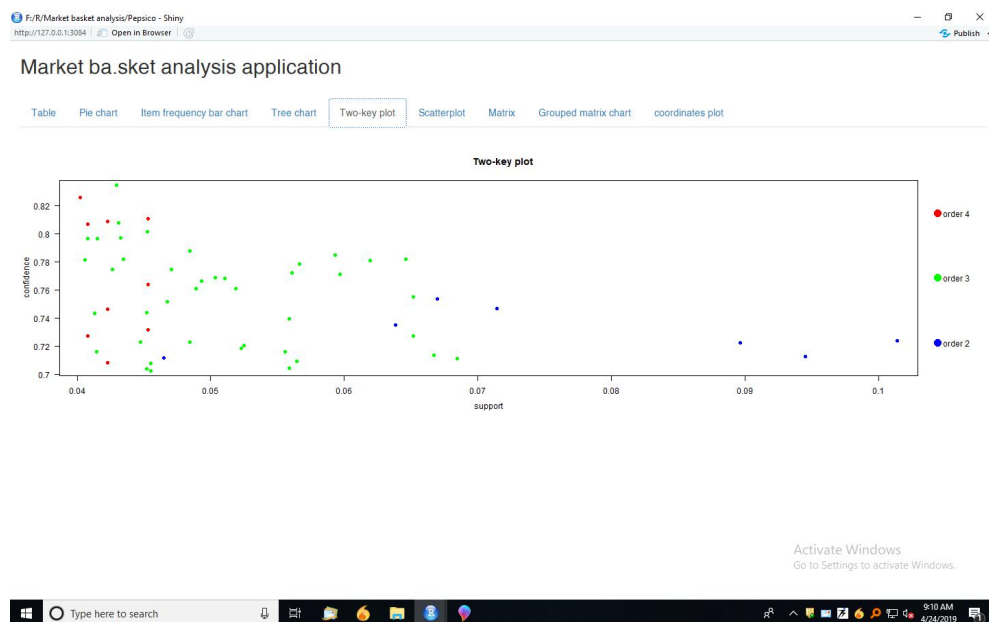


Fig Two key plot.

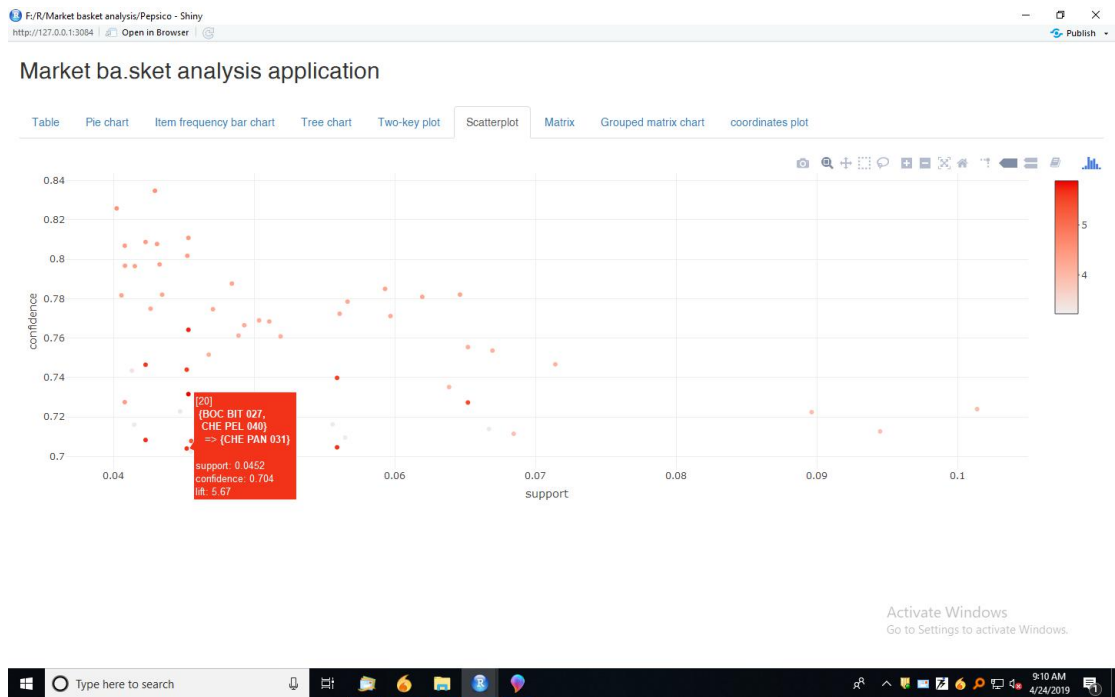


Fig Plotly scatter plot.

Matrix-based visualization

Matrix-based visualization creates a matrix with unique antecedent and consequent itemsets forming the columns and rows, respectively. The matrix contains the values for a interest measure selected by the analyst and is visualized using matrix shading. The order of rows and columns is arbitrary, however, to improve the ability to analyze the data, we suggest in arulesViz to reorder the matrix such that the row averages decrease from top to bottom and the column averages decrease from left to right. This pushes the rules with higher values of interestingness to the top-left position in the plot. Here we plot the rules using the method matrix and render it using an HTML widget.

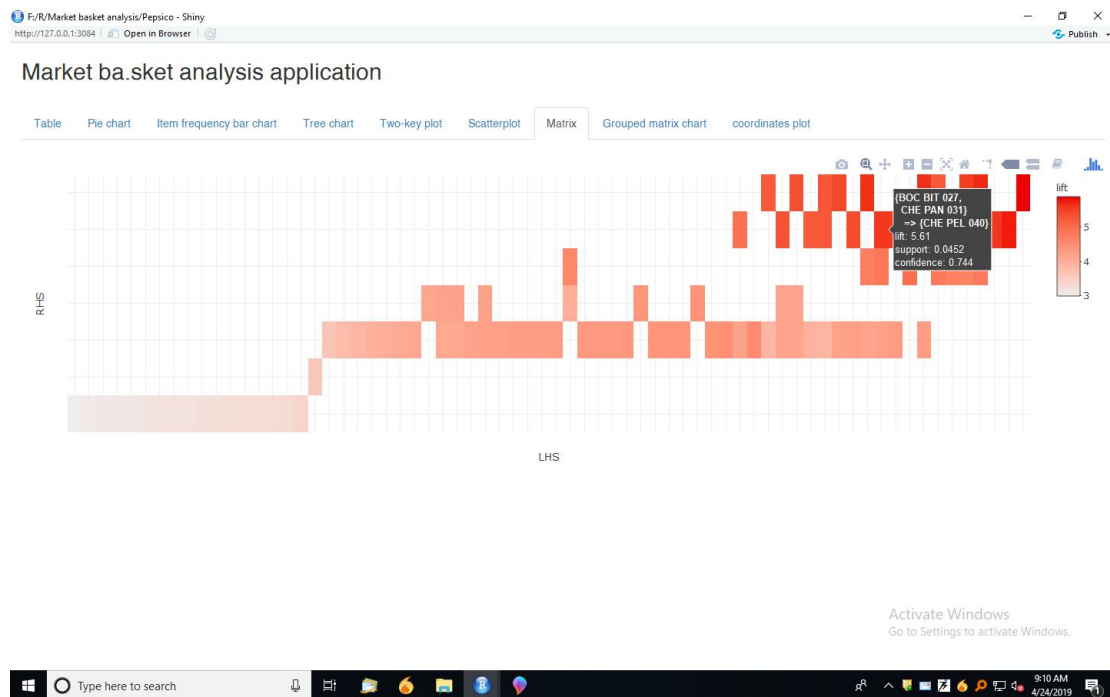


Fig Plottly matrix plot.

Matrix-based visualization is limited in the number of rules it can visualize effectively since large sets of rules typically also have large sets of unique antecedents. This would require the analyst to repeatedly zoom in and out. Grouped matrix-based visualization (Hahsler and Karpienko, 2016) enhances matrix-based visualization by grouping antecedents of rules via clustering and sorting rules by “interestingness” to handle a larger number of rules. Grouped rules are presented as an aggregate in a matrix that is visualized as a balloon plot.

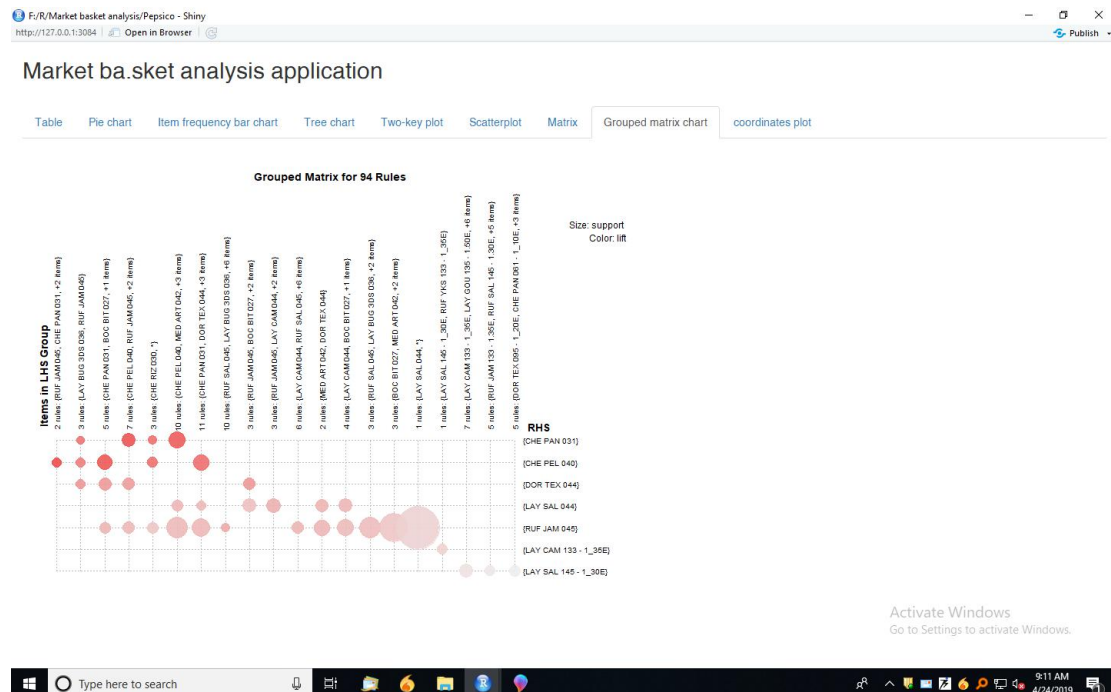


Fig Grouped matrix plot

Conclusion

While each of the three basic groups of visualizations reveal similar information, they focus on different properties of the analyzed rule set and some can visualize larger rule sets than others effectively. A single visualization by itself is typically not sufficient to understand all aspects of a rule set, but repeated use of different methods can lead the analyst to deeper insight into the data. Analysis typically starts with creating a scatter plot to inspect the rules with extreme values for support, confidence and lift. The advantage is that this visualization can deal with relatively large rule sets.

Alternatively, an interactive table can also be used for this task. A grouped matrix plot can then be used to inspect the rules and group of rules. Finally, graph-based visualization can be employed to get a deeper understanding of a smaller set of rules and items. This visualization is especially useful to present found results because it is easy to understand for non-analysts. *arulesViz* makes this process easier by providing a simple and unified plot method, where different visualizations can be explored by just changing the method argument. Many methods also can create JavaScriptbased HTML widgets (Vaidyanathan et al., 2017) which can be saved as HTML files, included in web-based applications using shiny (Chang et al., 2017), or used in interactive documents created with R markdown (Allaire et al., 2017)