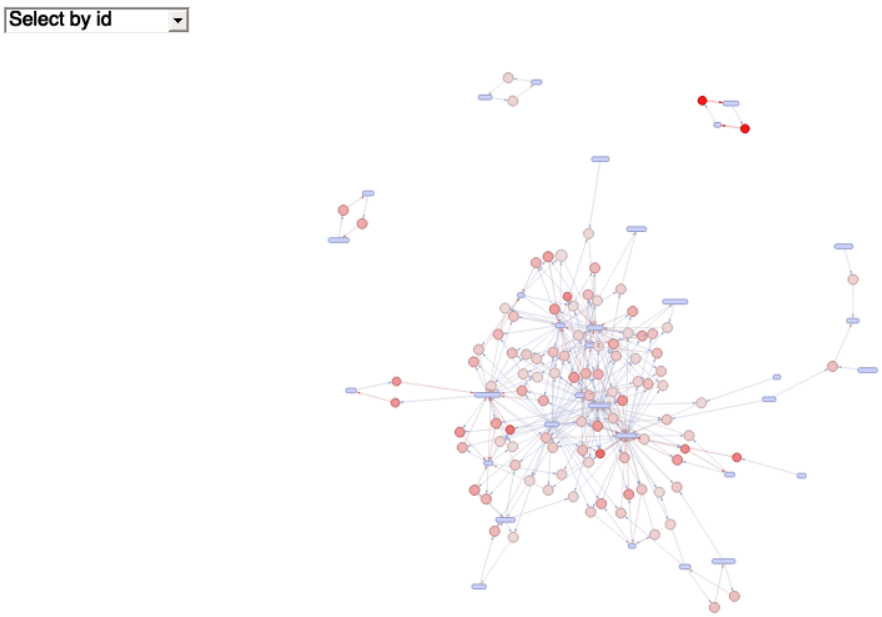
Association Rules

Scott Stempak, Alex Imhoff

To begin our analysis, we first wanted to look at some different values for support, confidence, and lift. We first generated association rules, specifying a minimum support of 0.5%, and minimum confidence of 10%. We then created a subset the rules to those with a lift of greater than 1.2. Finally, we visualized the rules above the lift threshold of 1.2, and then separately visualized the wider set of rules generated when we first initialized our rule set with minimum support of 0.5% and minimum confidence of 10%.

Association Rules with a Lift of at least 1.2 Network Visualization



The network created from our association rules with a lift of at least 1.2 shows many interesting trends. Chief among these is a complicated hub and spoke structure, where popular items are in the middle, connected to other grocery items, and a slew of rules. The item 'other vegetables' is among those elite items occupying everyone's grocery list, and it is involved in some simple rules, such as rule 10 (see below), which states that buying other vegetables, tropical fruit, and whole milk makes you liable to purchase root vegetables. These are all quintessential grocery store purchases, and most people who would shop for nutrient-dense fruits and vegetables can be reasonably expected to shop for milk and potatoes (root vegetables). One can imagine a Reese's Puffs connoisseur as less likely to spend meaningful time in the produce section of their local HEB.

Rule 10

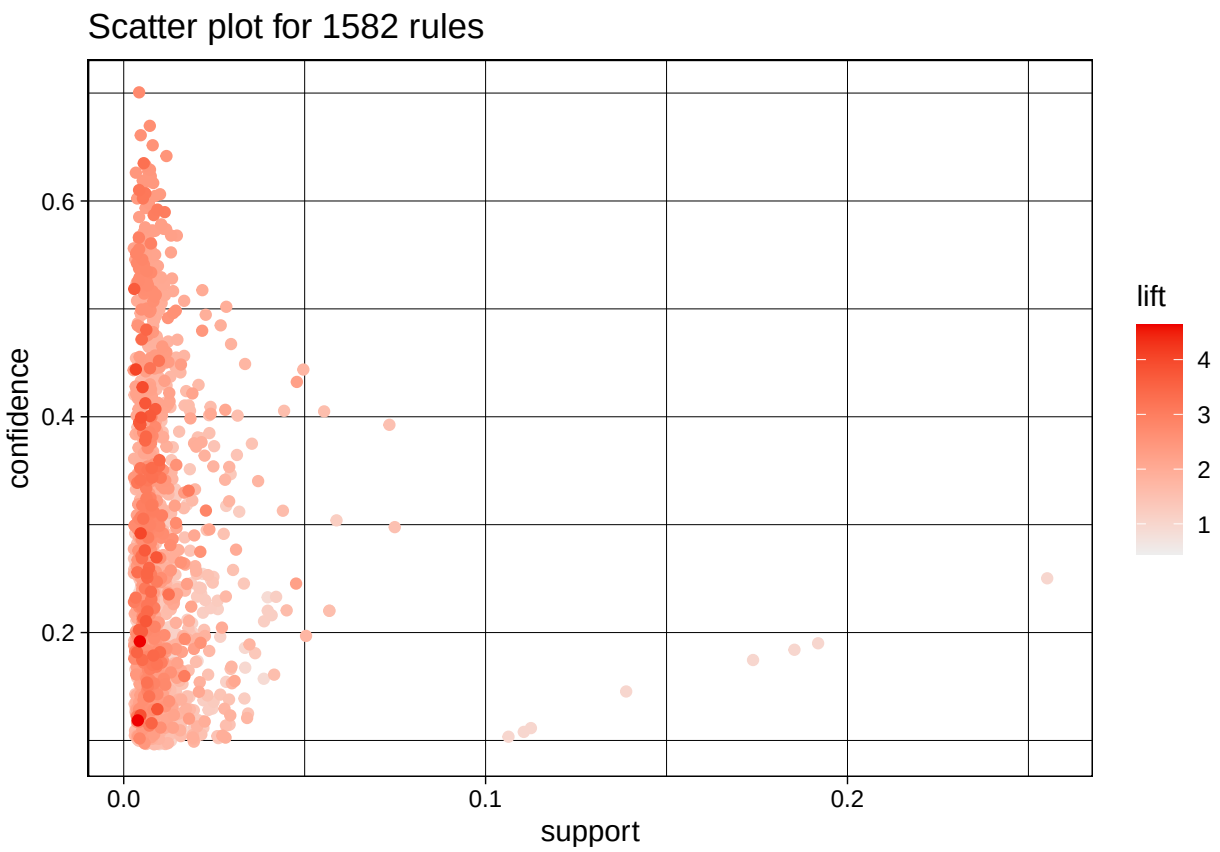
```
##     lhs                     rhs                   support confidence   coverage     lift count
## [1] {other vegetables,
##      tropical fruit,
##      whole milk}        => {root vegetables} 0.00701576  0.4107143 0.01708185 3.768074    69
```

The other vegetables item also forms some rules that are not nearly as likely to be taken for granted. For example, in rule 66 (see below), buying other vegetables and yogurt increases your likelihood of buying cream cheese by just over 3x. We read this as these vegetables going well with cream cheese as a snack dip, perhaps as a healthy option in a Super Bowl watch party; a Pippen to the Buffalo wing's Jordan. The addition of yogurt in that rule likely contributes to the lift, as it subsets the population of vegetable and cream cheese patrons to those who have the stomach (and appetite) for plenty of dairy.This dairy-related trend shows up consistently, as whole milk, whipped and sour cream, and yogurt are some of the other most central items in the graph, and two members of this lactose oligopoly show up in the majority of rules on the left side of the network.

Rule 66

```
##     lhs                             rhs             support     confidence
## [1] {other vegetables, yogurt} => {cream cheese } 0.005287239 0.1217799
##     coverage   lift     count
## [1] 0.04341637 3.071038 52
```

Confidence vs. Support plot for Association Rules



Taking a look at the Confidence vs. Support graph, it jumps out that there are very few rules with even 0.05 support. Confidence, however, maintains a solid density between 0.0-0.6. Lift, however, is the most interesting, as there seems to be more lift for items with lower support, and the darkest-colored points (with the highest lift) are more present below the 0.4 confidence mark. This makes it seem as though there are some rules for uncommon combinations of items that provide a very significant boost in buying another item. We imagine that human behaviors follow this trend depicted by the graph, as we often make grocery runs not just for our normal fill of food, but with specific plans in mind, such as buying cake and a lighter not just making you more likely to buy candles, but also to throw a birthday party.

It would be fascinating (and unfortunately difficult) to see how these grocery list association rules correspond to people's short-term plans, and longer term habits and lifestyles.