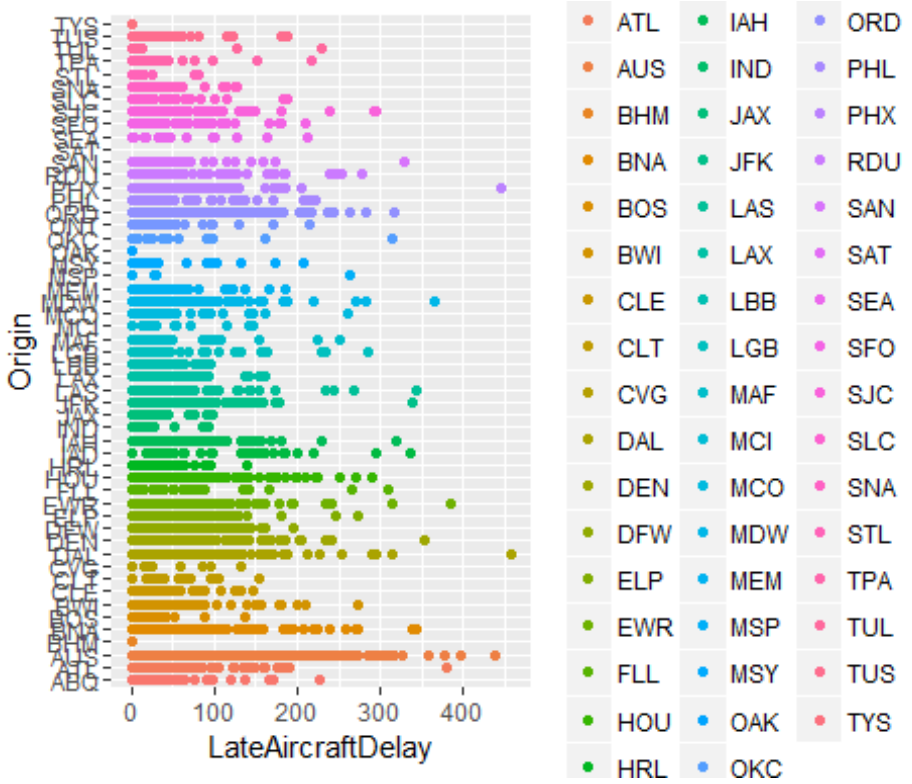
Homework 2 - Coker

Flights at ABIA

```
## Warning: Removed 79513 rows containing missing values (geom_point).
```



This figure plots the Late Aircraft Delay against the Origin, color coded based on the origin. Certain airports have very low instances of delay such as TYS, SAT, OAK, MSP, and BHM. Whereas other airpots have much higher instances of delay including ORD, MDW, IAH, JFK, IAH, HOU, DFW, DEN, and DAL.

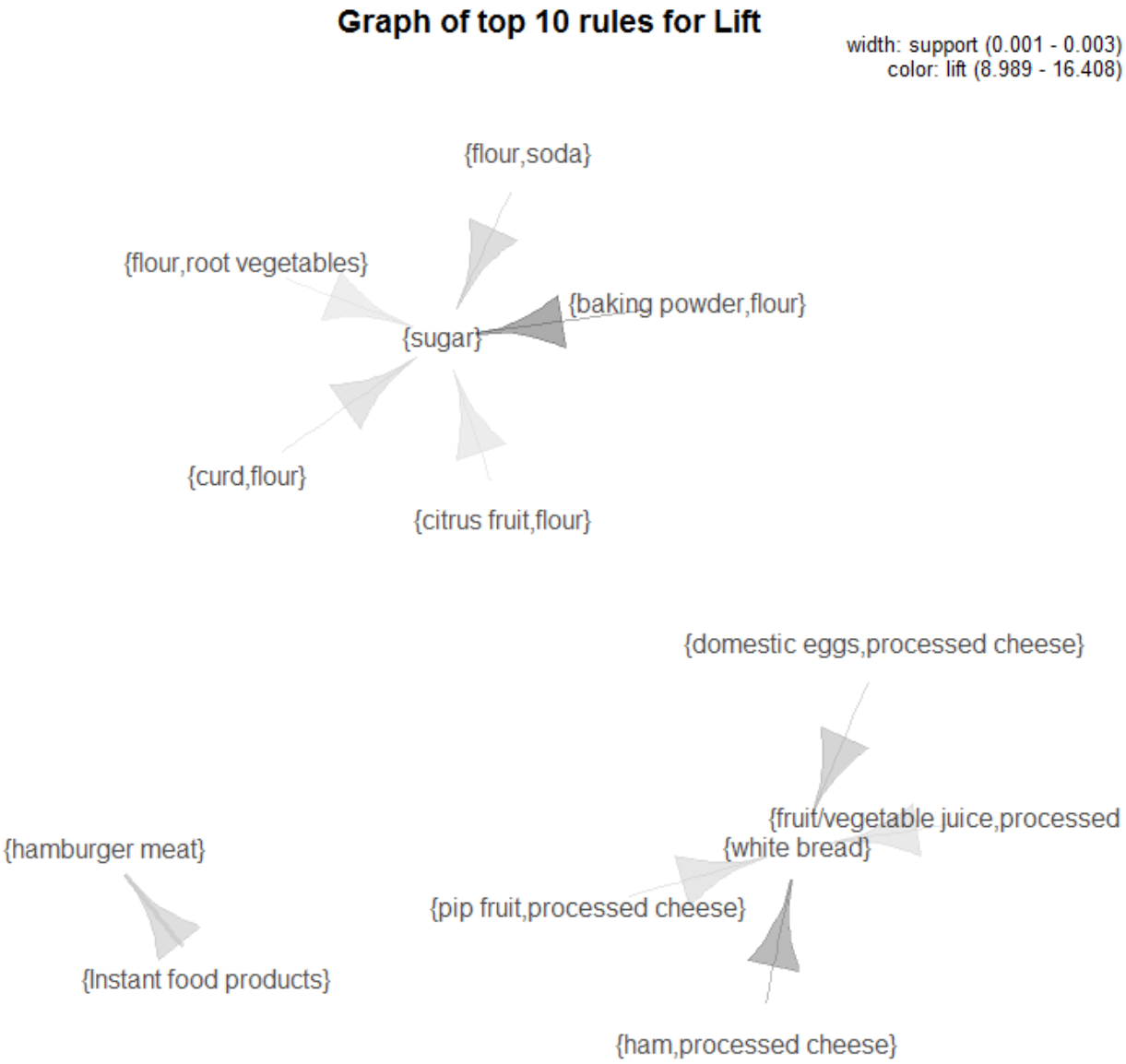
Author attribution

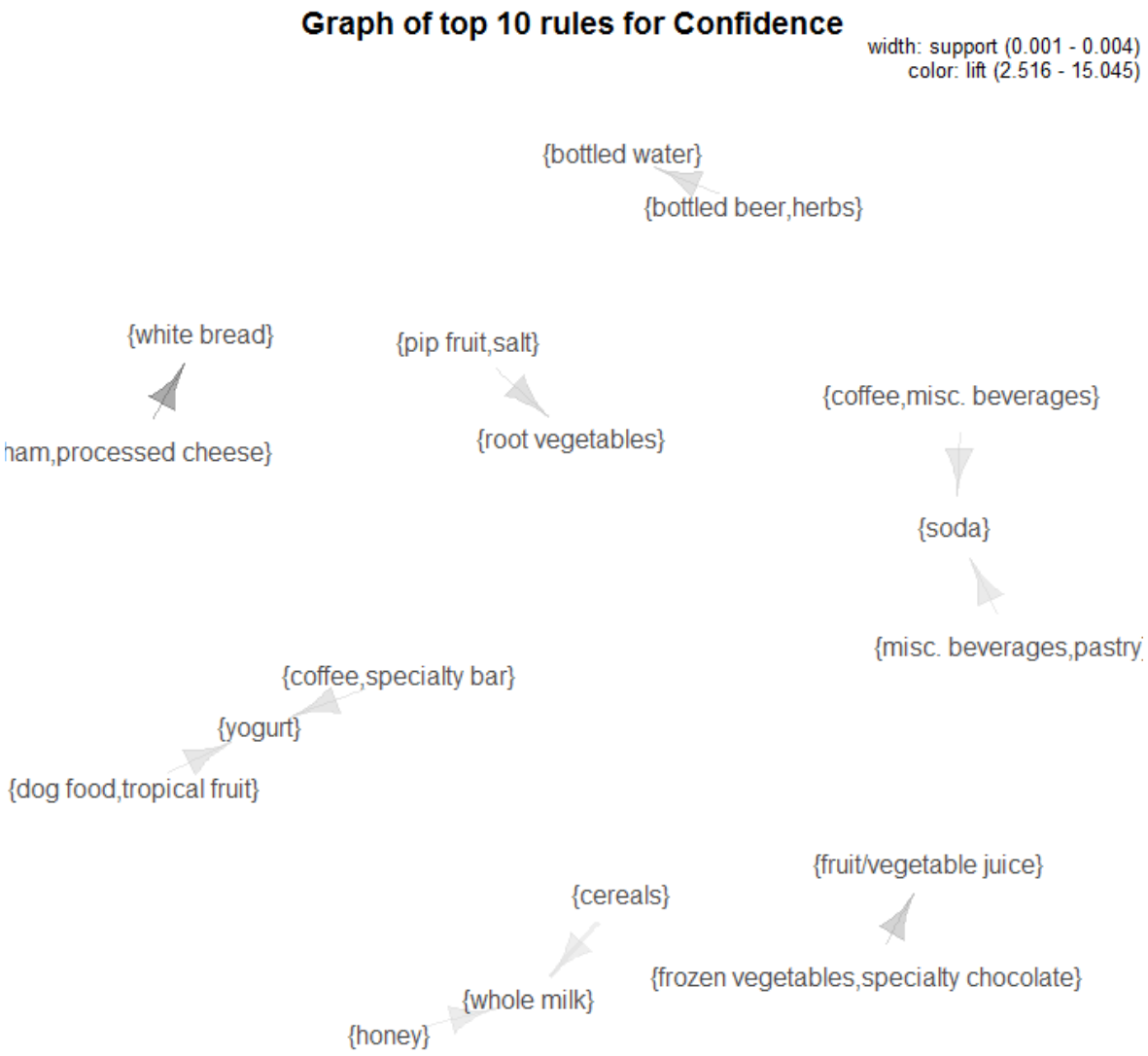
First, I chose to run a Naive Bayes model for my predictions using the caret and kLaR libraries, which had approximately 95% accuracy.

Next, I chose was KNN, which had approximately 60% accuracy.

Practice with association rule mining

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen maxlen
##         0.3    0.1    1 none FALSE            TRUE   0.001      1      4
##  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: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.01s].
## writing ... [12181 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```


Graph of top 10 rules for Lift
width: support (0.001 - 0.003)
color: lift (8.989 - 16.408)
{flour,soda}
{flour,root vegetables}
{baking powder,flour}
{sugar}
{curd,flour}
{citrus fruit,flour}
{domestic eggs,processed cheese}
{fruit/vegetable juice,processed
{white bread}
{hamburger meat}
{pip fruit,processed cheese}
{Instant food products}
{ham,processed cheese}



Support for the association wule mining is set to 0.001, which was chosen to allow for a high number of transactions. When support is set at a much higher fraction, the number of transactions that it specifies becomes very low.

Confidence is set at 0.3, which was chosen again to allow for a high number of patterns to be exposed for inspection.

The two graphs above show the top 10 rules for lift and confidence.

The graphs show that sugar associated with flour, white bread is associated with processed foods, and hamburgre meat is associated with instant food products. The graphs also show that wholk milk is asscoiated cereals and soda is associated with beverages.

These discovered item sets make a lot of sense seeing as how these items would naturally be bought together.