paskek_q3

Question 3

```
library(arules)
```

```
## Warning: package 'arules' was built under R version 3.2.2
```

```
## Loading required package: Matrix
##
## Attaching package: 'arules'
##
## The following objects are masked from 'package:base':
##
##     %in%, write
```

```
library(tm)
```

```
## Loading required package: NLP
##
## Attaching package: 'tm'
##
## The following object is masked from 'package:arules':
##
##     inspect
```

```
grocery = read.transactions("C:/Users/Kathleen/Downloads/STA380-master/STA380-master/data/groceries.txt
grocerytrans = as(grocery, "transactions")
summary(grocerytrans)
```

```
## transactions as itemMatrix in sparse format with
##  9835 rows (elements/itemsets/transactions) and
##  169 columns (items) and a density of 0.02609146
##
## most frequent items:
##       whole milk other vegetables       rolls/buns             soda
##             2513             1903             1809             1715
##           yogurt          (Other)
##             1372            34055
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
## 2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55
##   16   17   18   19   20   21   22   23   24   26   27   28   29   32
##   46   29   14   14    9   11    4    6    1    1    1    1    3    1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   1.000   2.000   3.000   4.409   6.000  32.000
##
## includes extended item information - examples:
##             labels
## 1 abrasive cleaner
## 2 artif. sweetener
## 3   baby cosmetics
```

Using the Apriori algorithm

This algorithm is used for frequent item set mining, i.e. basket analysis, and association between factors, i.e. products bought together. This algorithm allows us to determine which products will be purchased together or what else might be bought by people who purchase a grouping of other items. This algorithm produces 5688 rules

```
groceryrules <- apriori(grocerytrans,parameter=list(support=.001, confidence=.5, maxlen=6))
```

```
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen maxlen
##         0.5    0.1    1 none FALSE            TRUE   0.001      1      6
##  target   ext
##   rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## apriori - find association rules with the apriori algorithm
## version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
## 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 5 6 done [0.01s].
## writing ... [5668 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

From here we are able to run the inspect function to see which products are often bought together, i.e. the consumers basket. This allows further investigatino into the types of customers shopping and what tends to be purchased together. The following code would be used for this purpose. inspect("groceryrules") The output would show which products can be grouped together as common baskets. *However; the code will not knit.