STA380 Exercise 2 - Michael Lundy

Michael Lundy

August 16, 2016

Problem 1 - Flights at ABIA

```
library(arules)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
##
##     abbreviate, write
```

```
library(mosaic)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:arules':
##
##     intersect, recode, setdiff, setequal, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
## Loading required package: mosaicData
```

```
## Warning: failed to assign NativeSymbolInfo for lhs since lhs is already
## defined in the 'lazyeval' namespace
```

```
## Warning: failed to assign NativeSymbolInfo for rhs since rhs is already
## defined in the 'lazyeval' namespace
```

```
##
## The 'mosaic' package masks several functions from core packages in order
to add additional features.
## The original behavior of these functions should not be affected by this.

##
## Attaching package: 'mosaic'

## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following objects are masked from 'package:arules':
##
##     inspect, lhs, rhs, sample

## The following object is masked from 'package:Matrix':
##
##     mean

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cov, D, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```
library(ggplot2)
abia=read.csv("ABIA.csv")
names(abia)
```

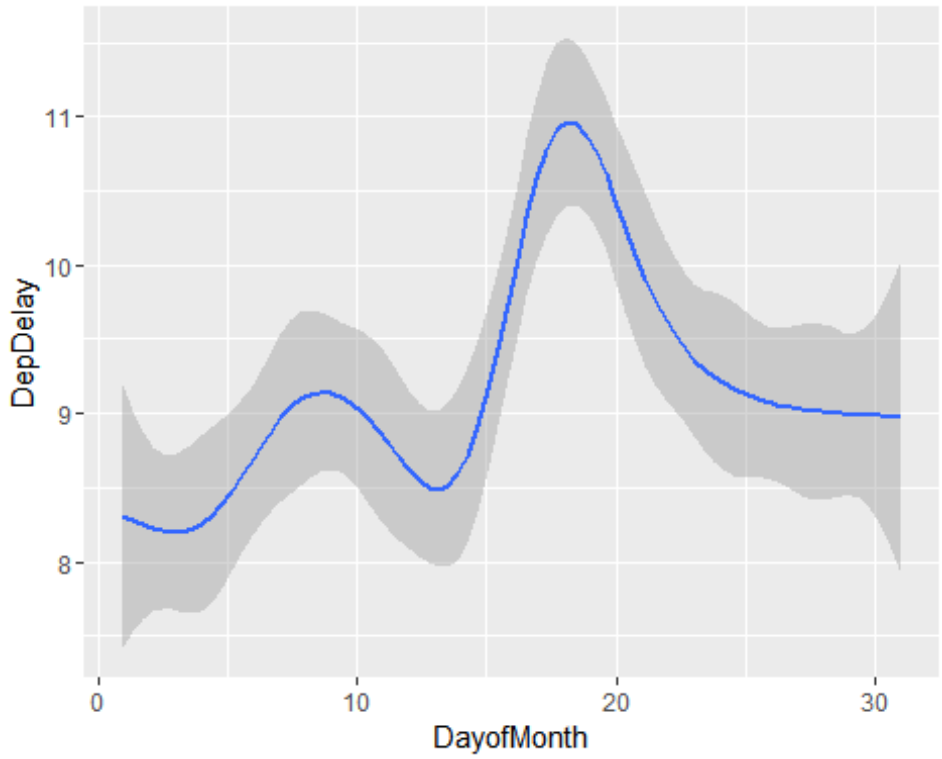
```
##  [1] "Year"              "Month"             "DayofMonth"
##  [4] "DayOfWeek"         "DepTime"           "CRSDepTime"
##  [7] "ArrTime"           "CRSArrTime"        "UniqueCarrier"
## [10] "FlightNum"         "TailNum"           "ActualElapsedTime"
## [13] "CRSElapsedTime"    "AirTime"           "ArrDelay"
## [16] "DepDelay"          "Origin"            "Dest"
## [19] "Distance"          "TaxiIn"            "TaxiOut"
## [22] "Cancelled"         "CancellationCode"  "Diverted"
## [25] "CarrierDelay"      "WeatherDelay"      "NASDelay"
## [28] "SecurityDelay"     "LateAircraftDelay"
```

```
attach(abia)
```

Average departure delay for each day of the month

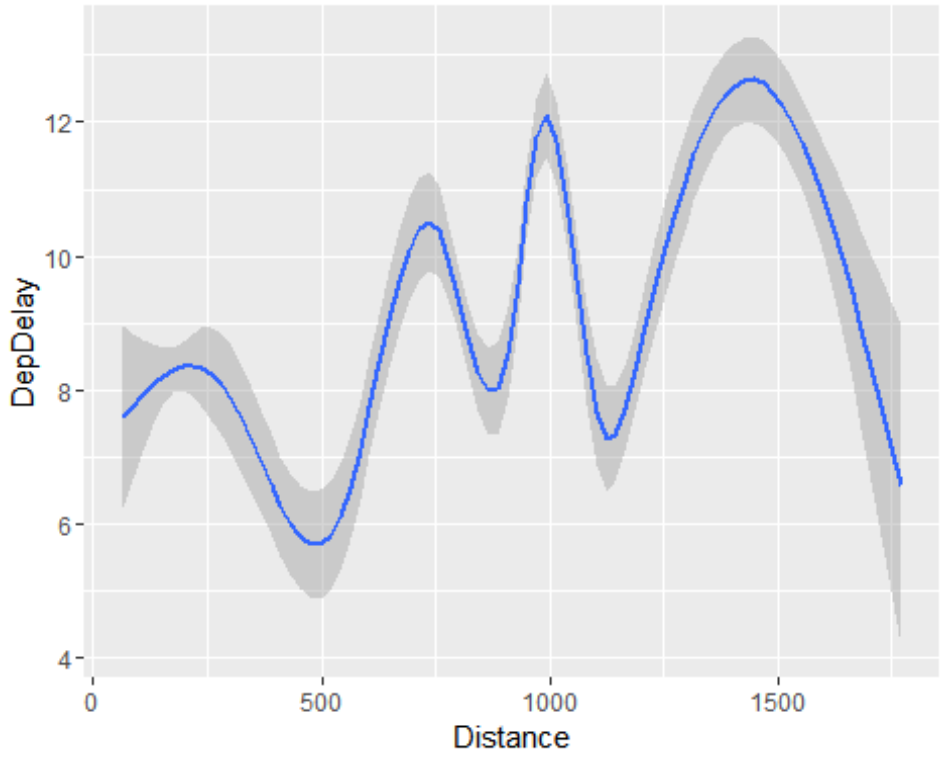
```
ggplot(abia, aes(DayofMonth, DepDelay)) + geom_smooth()
```

```
## Warning: Removed 1413 rows containing non-finite values (stat_smooth).
```




```
ggplot(abia, aes(Distance, DepDelay)) + geom_smooth()
```

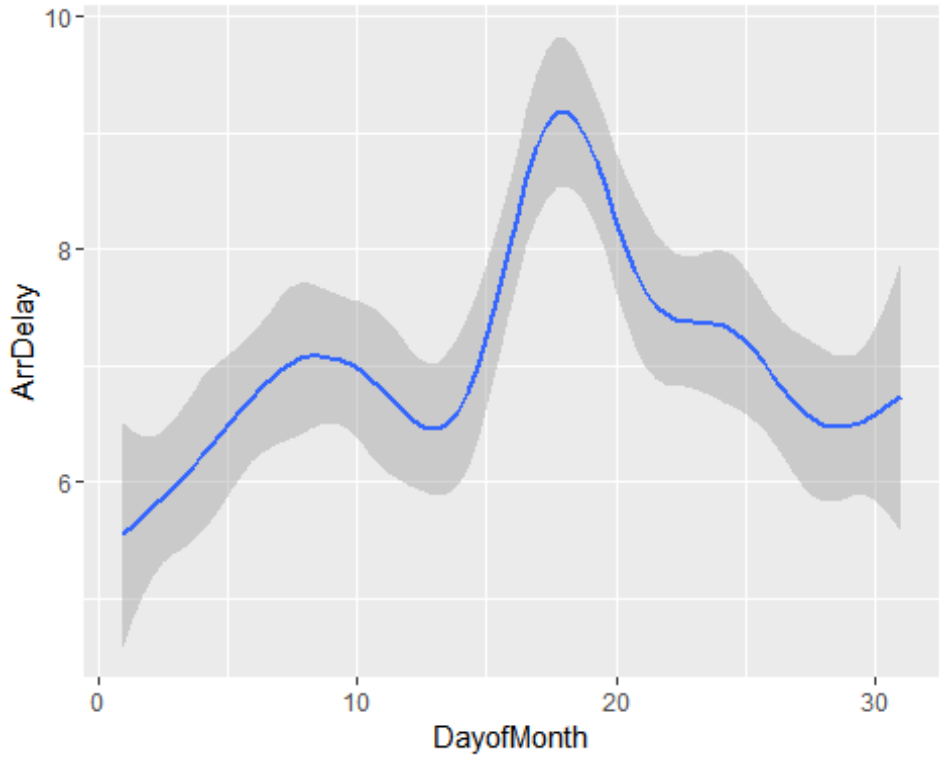
```
## Warning: Removed 1413 rows containing non-finite values (stat_smooth).
```



Average arrival delay for each day of the month

```
ggplot(abia, aes(DayofMonth, ArrDelay)) + geom_smooth()
```

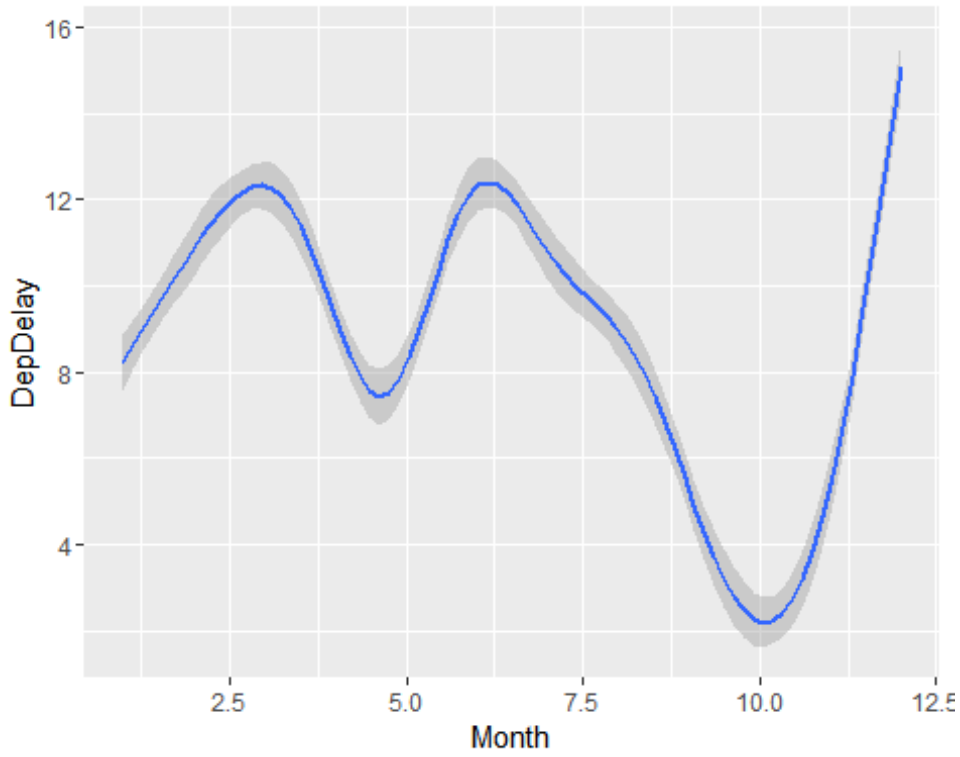
```
## Warning: Removed 1601 rows containing non-finite values (stat_smooth).
```



Average departure delay per month

```
ggplot(abia, aes(Month, DepDelay)) + geom_smooth()
```

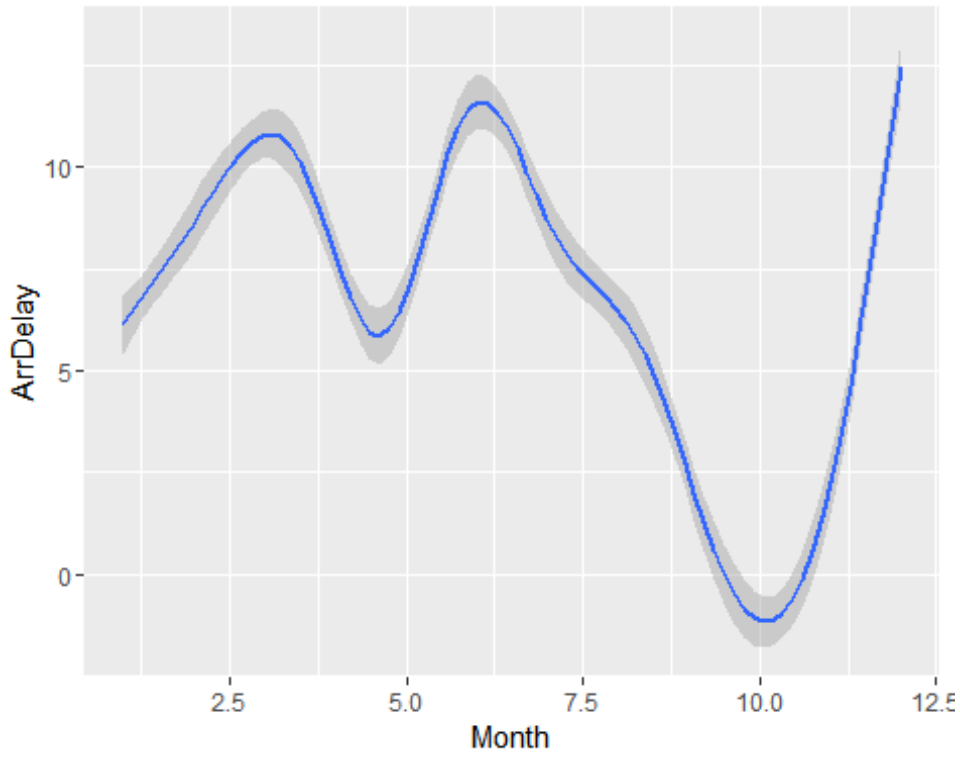
```
## Warning: Removed 1413 rows containing non-finite values (stat_smooth).
```



Average arrival delay per month

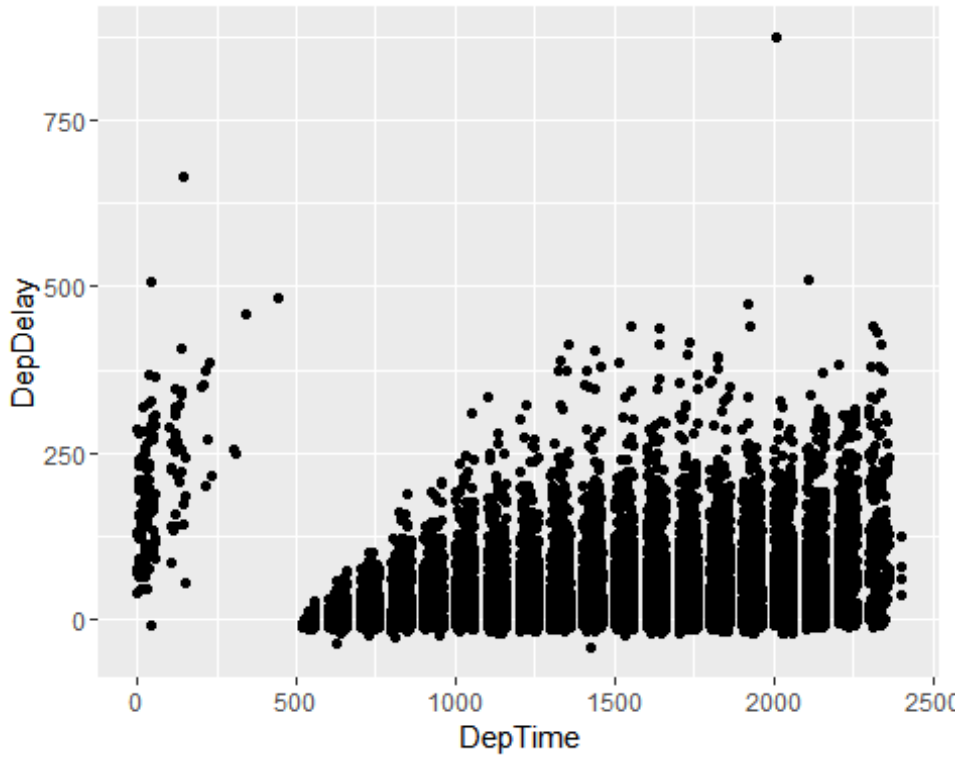
```
ggplot(abia, aes(Month, ArrDelay)) + geom_smooth()
```

```
## Warning: Removed 1601 rows containing non-finite values (stat_smooth).
```



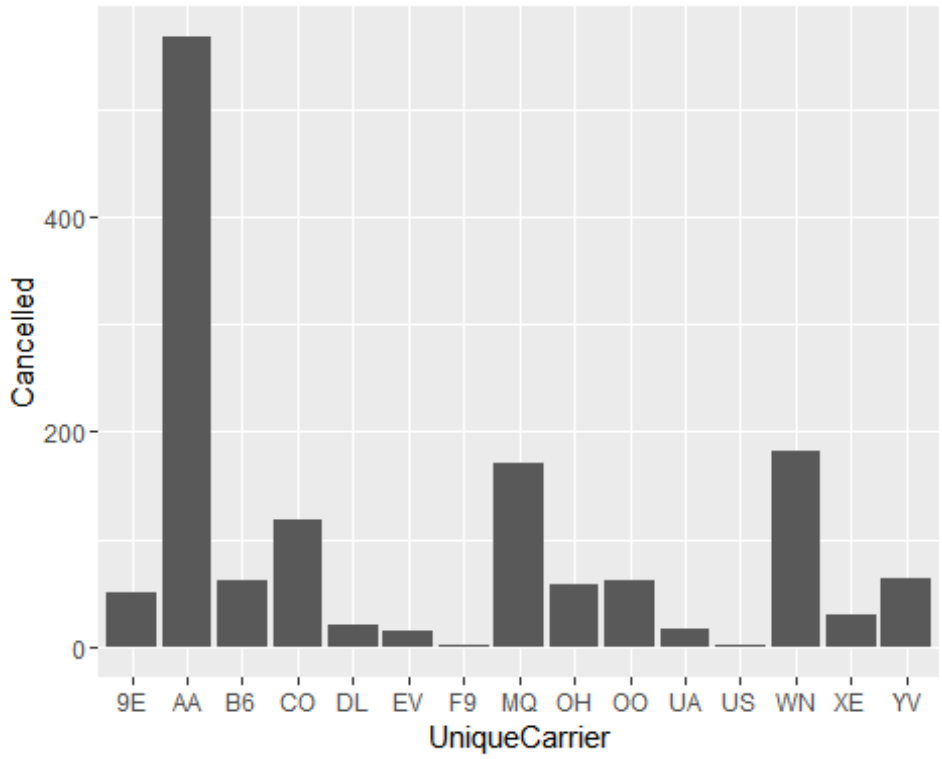
A plot of all of the departure delays versus the time they occured at in the day

```
ggplot(abia, aes(DepTime, DepDelay)) +
  geom_jitter()

## Warning: Removed 1413 rows containing missing values (geom_point).
```



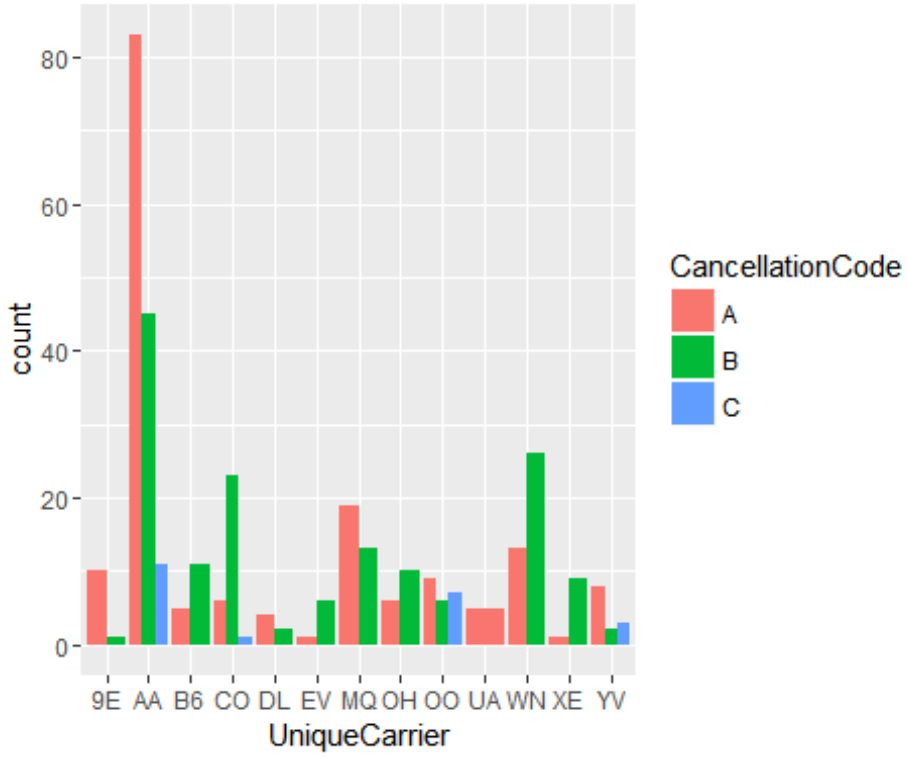
The count of flight cancellations per flight carrier

```
df1 <- abia[,c('UniqueCarrier', 'Cancelled')]
only_cancellations <- df1[df1$Cancelled == 1,]
ggplot(only_cancellations, aes(x = UniqueCarrier, y = Cancelled)) +
geom_bar(stat='identity')
```



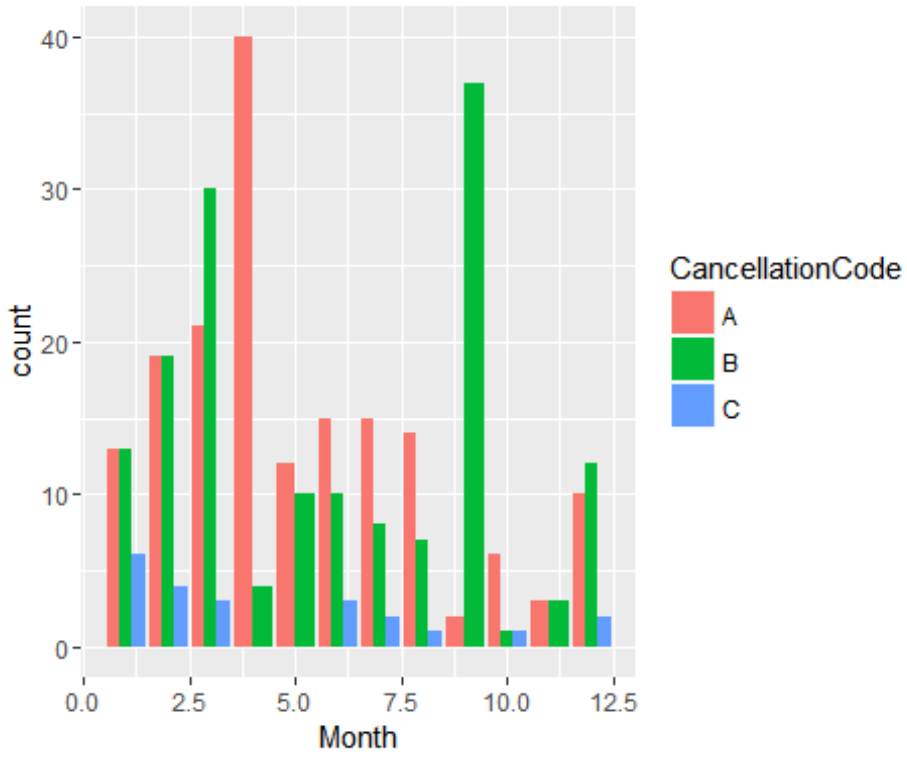
The count of each type of cancellation per flight carrier

```
df2 <- abia[,c('UniqueCarrier', 'CancellationCode')]
df2_A <- df2[df2$CancellationCode == c('A','B','C','D'),]
ggplot(df2_A, aes(x=UniqueCarrier, fill=CancellationCode)) +
geom_bar(position='dodge')
```

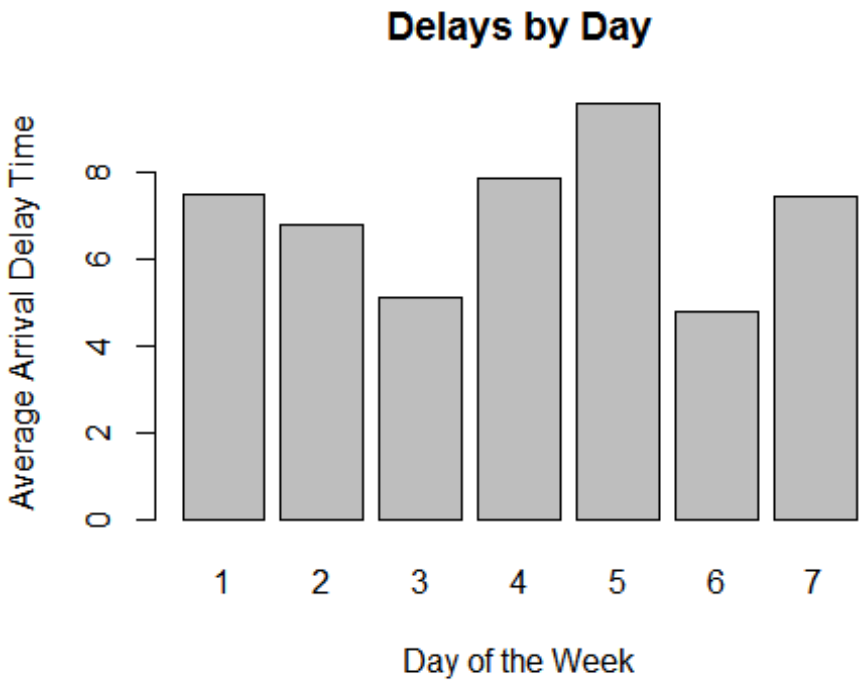


The count of each type of cancellation for each month

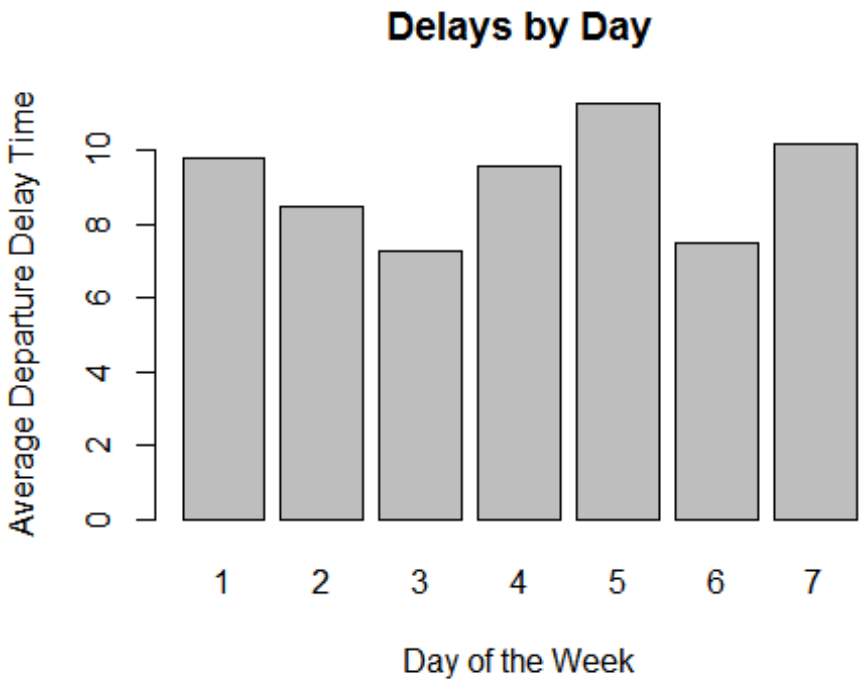
```
df3 <- abia[,c('Month', 'CancellationCode')]
df3_A <- df3[df3$CancellationCode == c('A','B','C','D'),]
ggplot(df3_A, aes(x=Month, fill=CancellationCode)) +
geom_bar(position='dodge')
```



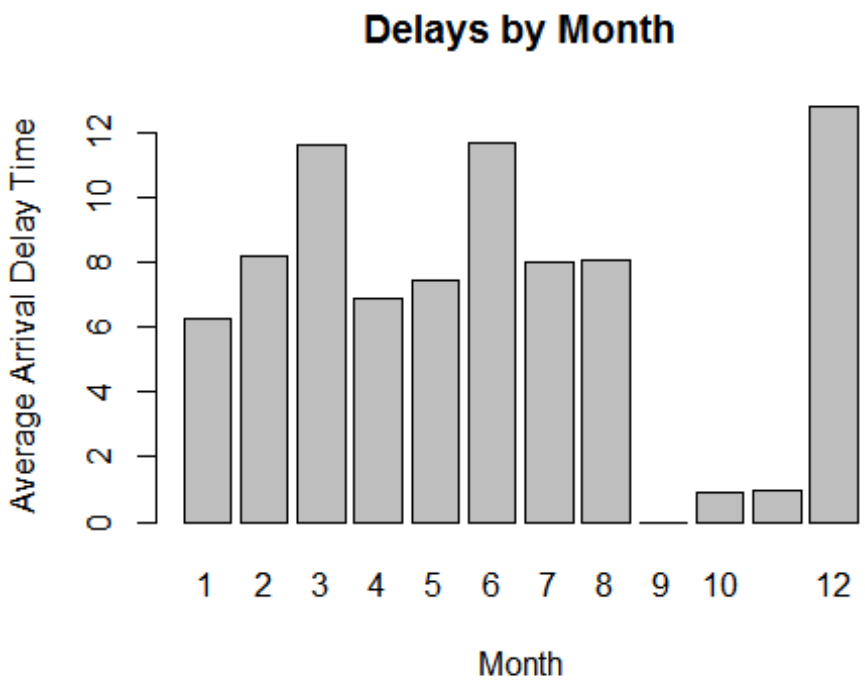

```
df4 <- abia[, c("ArrDelay","DayOfWeek")]
head(df4)

##   ArrDelay DayOfWeek
## 1      339         2
## 2       -9         2
## 3       -1         2
## 4      -23         2
## 5       -6         2
## 6        2         2

a <- aggregate(ArrDelay ~ DayOfWeek, data=df4, FUN=mean)
barplot(a$ArrDelay, names.arg=a$DayOfWeek, ylab = "Average Arrival Delay
Time", xlab = "Day of the Week", main = "Delays by Day")
```



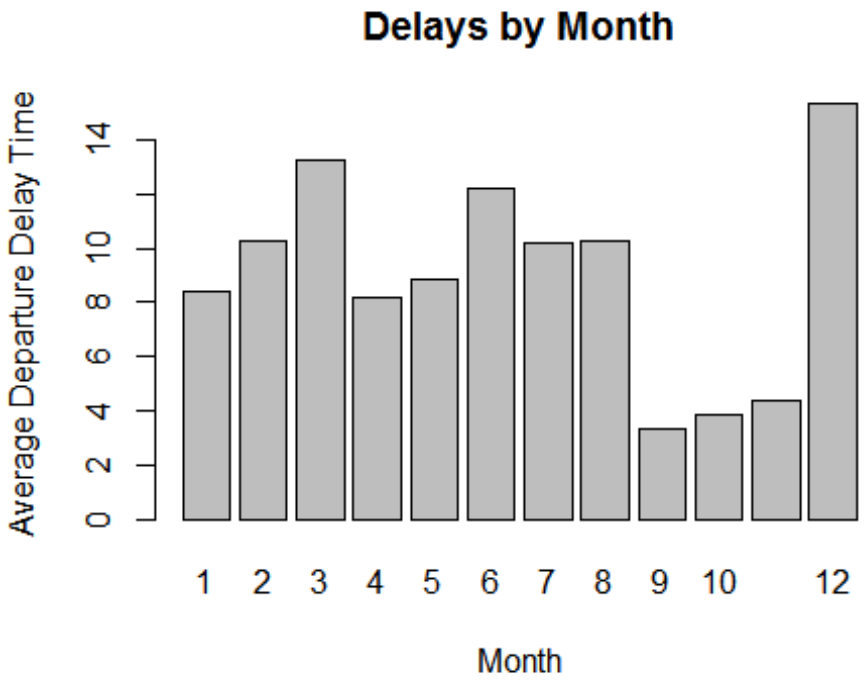

```
df5 <- abia[, c("DepDelay","DayOfWeek")]
head(df5)

##   DepDelay DayOfWeek
## 1      345         2
## 2       -5         2
## 3        0         2
## 4       -4         2
## 5        1         2
## 6       -9         2

b <- aggregate(DepDelay ~ DayOfWeek, data=df5, FUN=mean)
barplot(b$DepDelay, names.arg=b$DayOfWeek, ylab = "Average Departure Delay
Time", xlab = "Day of the Week", main = "Delays by Day")
```




```
df6 <- abia[, c("ArrDelay","Month")]
head(df6)

##   ArrDelay Month
## 1      339     1
## 2       -9     1
## 3       -1     1
## 4      -23     1
## 5       -6     1
## 6        2     1

c <- aggregate(ArrDelay ~ Month, data=df6, FUN=mean)
barplot(c$ArrDelay, names.arg=c$Month, ylab = "Average Arrival Delay Time",
xlab = "Month", main = "Delays by Month")
```




```
df7 <- abia[, c("DepDelay","Month")]
head(df7)

##   DepDelay Month
## 1      345     1
## 2       -5     1
## 3        0     1
## 4       -4     1
## 5        1     1
## 6       -9     1

d <- aggregate(DepDelay ~ Month, data=df7, FUN=mean)
barplot(d$DepDelay, names.arg=d$Month, ylab = "Average Departure Delay Time",
xlab = "Month", main = "Delays by Month")
```



Problem 2 - Author Attribution

```
library(tm)

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##     annotate

##
## Attaching package: 'tm'

## The following object is masked from 'package:mosaic':
##
##     inspect

## The following object is masked from 'package:arules':
##
##     inspect

library(e1071)
library(caret)

##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:mosaic':
##
##     dotPlot

library(glmnet)

## Loading required package: foreach

## Loaded glmnet 2.0-5

library(SnowballC)
library(class)
library(randomForest)

## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

## The following object is masked from 'package:dplyr':
##
##     combine

source('textutils.R')

readerPlain=function(fname){
  readPlain(elem=list(content=readLines(fname)), id=fname, language='en')}


author_dirs=Sys.glob('~/MSBA/Summer/Predictive Models 2/STA380-
master/data/ReutersC50/C50train/*')
file_list=NULL
labels=NULL

for(author in author_dirs){
  author_name=substring(author, first=20)
  files_to_add= Sys.glob(paste0(author,'/*.txt'))
  file_list=append(file_list, files_to_add)
  labels=append(labels, rep(author_name, length(files_to_add)))
}

all_docs = lapply(file_list, readerPlain)
names(all_docs) = file_list

my_corpus = Corpus(VectorSource(all_docs))
```

```
names(my_corpus) = file_list


my_corpus = tm_map(my_corpus, content_transformer(tolower)) # make everything
lowercase
my_corpus = tm_map(my_corpus, content_transformer(removeNumbers)) # remove
numbers
my_corpus = tm_map(my_corpus, content_transformer(removePunctuation)) #
remove punctuation
my_corpus = tm_map(my_corpus, content_transformer(stripWhitespace)) ## remove
excess white-space
my_corpus = tm_map(my_corpus, content_transformer(removeWords),
stopwords("SMART"))

#my_corpus <- tm_map(my_corpus, stemDocument)
DTM = DocumentTermMatrix(my_corpus)
DTM = removeSparseTerms(DTM, 0.975)

X = as.matrix(DTM)
X = X/rowSums(X)
X = idf.weight(X)
```

In order to create a model that takes into account all possible words, words that weren't in the training set but were in the test set were added to the training set, and vice versa. This way, all potential words were available for creating and testing the model.

Test Set

```
author_test=Sys.glob('~/MSBA/Summer/Predictive Models 2/STA380-
master/data/ReutersC50/C50test/*')
filelist = NULL
labels2 = NULL

for(author in author_test){
  authorname=substring(author, first=20)
  filestoadd= Sys.glob(paste0(author,'/*.txt'))
  filelist=append(filelist, filestoadd)
  labels2=append(labels2, rep(authorname, length(filestoadd)))
}

alldocs = lapply(filelist, readerPlain)
names(alldocs)=filelist

mycorpus = Corpus(VectorSource(alldocs))
names(mycorpus)=filelist

mycorpus = tm_map(mycorpus, content_transformer(tolower)) # make everything
lowercase
mycorpus = tm_map(mycorpus, content_transformer(removeNumbers)) # remove
numbers
```

```
mycorpus = tm_map(mycorpus, content_transformer(removePunctuation)) # remove
punctuation
mycorpus = tm_map(mycorpus, content_transformer(stripWhitespace)) ## remove
excess white-space
mycorpus = tm_map(mycorpus, content_transformer(removeWords),
stopwords("SMART"))

DTM2=DocumentTermMatrix(mycorpus)
DTM2=removeSparseTerms(DTM2,0.975)

x2=as.matrix(DTM2)

x2=x2/rowSums(x2)
x2=idf.weight(x2)

words=colnames(X)
words2=colnames(x2)

W=words[!(words %in% words2)]
W2=words2[!(words2 %in% words)]

words_matrix=matrix(0,nrow=nrow(x2), ncol=length(W))
colnames(words_matrix)=W

words_matrix2=matrix(0,nrow=nrow(X), ncol=length(W2))
colnames(words_matrix2)=W2

train_matrix=cbind(X,words_matrix2)
test_matrix=cbind(x2,words_matrix)
```

Naive Bayes Model

```
test_matrix=as.data.frame(test_matrix)
train_matrix=as.data.frame(train_matrix)

nb = naiveBayes(x=train_matrix,y=as.factor(labels),laplace=1)
predNB=predict(nb,test_matrix)

actual = rep(1:50,each=50)

TestTable = table(predNB,actual)
correct = 0
for (i in seq(1,50)){
  correct = correct + TestTable[i,i]
}

accuracy = correct/2500
accuracy
```

```
## [1] 0.4004
```

The Naive Bayes model prediction accuracy is somewhat low, despite being much better than randomly guessing. A different model may have better predictive accuracy.

Random Forest Model

```
rand = randomForest(y=as.factor(labels), x=train_matrix,ntrees=100)
pr = predict(rand, test_matrix, type = "response")

TestTable2 = table(pr, actual)

correct2 = 0
for (i in seq(1,50)){
  correct2 = correct2 + TestTable2[i,i]
}

accuracy2 = correct2/2500
accuracy2
```

```
## [1] 0.5468
```

The random forest model was a good bit better at about 54%.

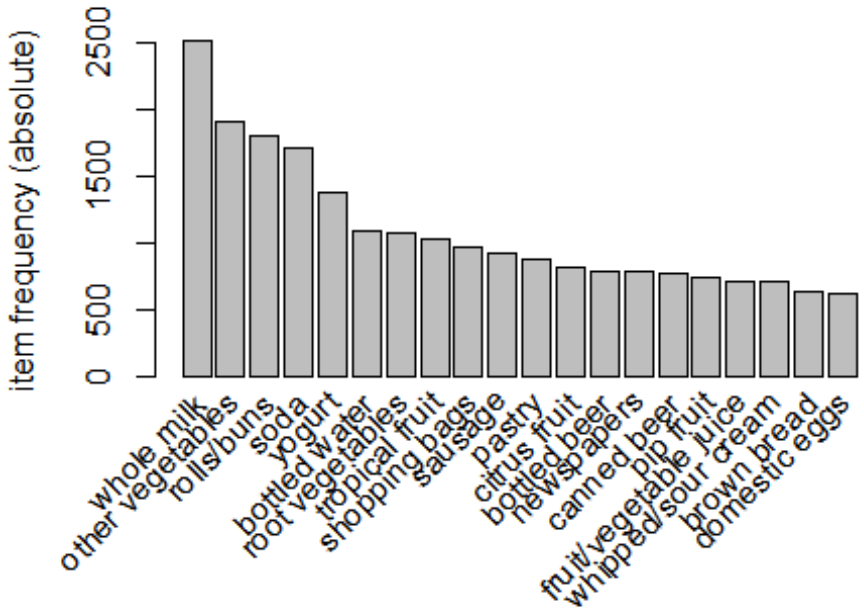
Problem 3 - Association Rule Mining

```
library(arules)
library(arulesViz)

## Loading required package: grid

library(datasets)
groceries <- read.transactions(file = 'groceries.txt', format='basket',
sep=',')
summary(groceries)

## 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
```

Plot the top 20 most frequent items found in baskets.

```
itemFrequencyPlot(groceries, topN=20, type='absolute')
```



Find all of the associations that have a support of at least 0.001 and a confidence of 0.8. A support of at least 0.001 means that we only want to create associations with items that are found in at least 0.1% of the transactions. Items that occur any less than this most likely can't form any valuable insights due to their low number of occurences. A confidence of 0.8 means that we only want associations that actually occur in the data of baskets 80% of the time. The lift displayed shows how much more likely a given item will occur (rhs) if a basket contains the given set of items (lhs).

```
rules <- apriori(groceries, parameter = list(support = 0.001, confidence =
0.8, maxlen=4))

## Apriori
##
```

```
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen maxlen
##         0.8    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 ... [258 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
options(digits=2)
arules::inspect(rules)
```

```
##     lhs                           rhs               support confidence
lift
## 1   {liquor,
##      red/blush wine}           => {bottled beer}      0.0019       0.90
11.2
## 2   {cereals,
##      curd}                     => {whole milk}        0.0010       0.91
3.6
## 3   {cereals,
##      yogurt}                   => {whole milk}        0.0017       0.81
3.2
## 4   {butter,
##      jam}                      => {whole milk}        0.0010       0.83
3.3
## 5   {bottled beer,
##      soups}                    => {whole milk}        0.0011       0.92
3.6
## 6   {house keeping products,
##      napkins}                  => {whole milk}        0.0013       0.81
3.2
## 7   {house keeping products,
##      whipped/sour cream}       => {whole milk}        0.0012       0.92
3.6
## 8   {pastry,
##      sweet spreads}            => {whole milk}        0.0010       0.91
3.6
## 9   {curd,
```

```
##      turkey}                    => {other vegetables}  0.0012       0.80
4.1
## 10  {rice,
##      sugar}                     => {whole milk}        0.0012       1.00
3.9
## 11  {butter,
##      rice}                      => {whole milk}        0.0015       0.83
3.3
## 12  {domestic eggs,
##      rice}                      => {whole milk}        0.0011       0.85
3.3
## 13  {bottled water,
##      rice}                      => {whole milk}        0.0012       0.92
3.6
## 14  {rice,
##      yogurt}                    => {other vegetables}  0.0019       0.83
4.3
## 15  {mustard,
##      oil}                       => {whole milk}        0.0012       0.86
3.4
## 16  {canned fish,
##      hygiene articles}          => {whole milk}        0.0011       1.00
3.9
## 17  {fruit/vegetable juice,
##      herbs}                     => {other vegetables}  0.0012       0.80
4.1
## 18  {herbs,
##      shopping bags}             => {other vegetables}  0.0019       0.83
4.3
## 19  {herbs,
##      tropical fruit}            => {whole milk}        0.0023       0.82
3.2
## 20  {herbs,
##      rolls/buns}                => {whole milk}        0.0024       0.80
3.1
## 21  {chocolate,
##      pickled vegetables}        => {whole milk}        0.0012       0.86
3.4
## 22  {grapes,
##      onions}                    => {other vegetables}  0.0011       0.92
4.7
## 23  {margarine,
##      meat}                      => {other vegetables}  0.0017       0.85
4.4
## 24  {hard cheese,
##      oil}                       => {other vegetables}  0.0011       0.92
4.7
## 25  {butter milk,
##      onions}                    => {other vegetables}  0.0013       0.81
4.2
```

```
## 26  {butter milk,
##      pork}                    => {other vegetables}  0.0018       0.86
4.4
## 27  {onions,
##      waffles}                 => {other vegetables}  0.0012       0.80
4.1
## 28  {curd,
##      hamburger meat}          => {whole milk}        0.0025       0.81
3.2
## 29  {bottled beer,
##      hamburger meat}          => {whole milk}        0.0017       0.81
3.2
## 30  {other vegetables,
##      specialty cheese,
##      yogurt}                  => {whole milk}        0.0013       0.81
3.2
## 31  {root vegetables,
##      tropical fruit,
##      turkey}                  => {other vegetables}  0.0012       0.80
4.1
## 32  {root vegetables,
##      turkey,
##      whole milk}              => {other vegetables}  0.0012       0.80
4.1
## 33  {butter,
##      rice,
##      root vegetables}         => {whole milk}        0.0010       1.00
3.9
## 34  {other vegetables,
##      rice,
##      tropical fruit}          => {whole milk}        0.0010       0.83
3.3
## 35  {rice,
##      root vegetables,
##      yogurt}                  => {other vegetables}  0.0014       0.88
4.5
## 36  {rice,
##      root vegetables,
##      yogurt}                  => {whole milk}        0.0014       0.88
3.4
## 37  {other vegetables,
##      rice,
##      root vegetables}         => {whole milk}        0.0018       0.82
3.2
## 38  {rice,
##      whole milk,
##      yogurt}                  => {other vegetables}  0.0015       0.83
4.3
## 39  {frozen fish,
##      pip fruit,
```

```
##      tropical fruit}           => {other vegetables}  0.0010       0.83
4.3
## 40  {frozen fish,
##      pip fruit,
##      whole milk}               => {other vegetables}  0.0011       0.85
4.4
## 41  {frozen fish,
##      root vegetables,
##      yogurt}                   => {whole milk}        0.0012       0.86
3.4
## 42  {frozen fish,
##      other vegetables,
##      yogurt}                   => {whole milk}        0.0012       0.86
3.4
## 43  {curd,
##      herbs,
##      root vegetables}          => {whole milk}        0.0012       0.86
3.4
## 44  {domestic eggs,
##      herbs,
##      other vegetables}         => {whole milk}        0.0010       0.83
3.3
## 45  {fruit/vegetable juice,
##      herbs,
##      other vegetables}         => {whole milk}        0.0010       0.83
3.3
## 46  {fruit/vegetable juice,
##      herbs,
##      whole milk}               => {other vegetables}  0.0010       0.91
4.7
## 47  {citrus fruit,
##      herbs,
##      tropical fruit}           => {other vegetables}  0.0010       0.83
4.3
## 48  {citrus fruit,
##      herbs,
##      tropical fruit}           => {whole milk}        0.0011       0.92
3.6
## 49  {citrus fruit,
##      herbs,
##      root vegetables}          => {other vegetables}  0.0013       0.81
4.2
## 50  {citrus fruit,
##      herbs,
##      root vegetables}          => {whole milk}        0.0013       0.81
3.2
## 51  {herbs,
##      root vegetables,
##      shopping bags}            => {other vegetables}  0.0011       0.85
4.4
```

```
## 52  {herbs,
##      root vegetables,
##      tropical fruit}           => {whole milk}        0.0015       0.88
3.5
## 53  {herbs,
##      tropical fruit,
##      yogurt}                   => {whole milk}        0.0010       0.83
3.3
## 54  {herbs,
##      other vegetables,
##      tropical fruit}           => {whole milk}        0.0013       0.81
3.2
## 55  {herbs,
##      rolls/buns,
##      root vegetables}          => {whole milk}        0.0015       0.83
3.3
## 56  {semi-finished bread,
##      tropical fruit,
##      yogurt}                   => {other vegetables}  0.0013       0.81
4.2
## 57  {detergent,
##      other vegetables,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 58  {baking powder,
##      tropical fruit,
##      yogurt}                   => {whole milk}        0.0011       0.85
3.3
## 59  {flour,
##      other vegetables,
##      sugar}                    => {whole milk}        0.0012       0.86
3.4
## 60  {flour,
##      root vegetables,
##      whipped/sour cream}       => {whole milk}        0.0017       1.00
3.9
## 61  {flour,
##      rolls/buns,
##      root vegetables}          => {other vegetables}  0.0010       0.83
4.3
## 62  {butter,
##      domestic eggs,
##      soft cheese}              => {whole milk}        0.0010       1.00
3.9
## 63  {butter,
##      soft cheese,
##      yogurt}                   => {whole milk}        0.0012       0.80
3.1
## 64  {domestic eggs,
##      root vegetables,
```

```
##      soft cheese}              => {other vegetables}  0.0010       0.83
4.3
## 65  {domestic eggs,
##      root vegetables,
##      soft cheese}              => {whole milk}        0.0010       0.83
3.3
## 66  {soft cheese,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0012       0.92
4.8
## 67  {root vegetables,
##      soft cheese,
##      whipped/sour cream}       => {whole milk}        0.0012       0.92
3.6
## 68  {citrus fruit,
##      root vegetables,
##      soft cheese}              => {other vegetables}  0.0010       1.00
5.2
## 69  {grapes,
##      pork,
##      whole milk}               => {other vegetables}  0.0010       0.83
4.3
## 70  {citrus fruit,
##      fruit/vegetable juice,
##      grapes}                   => {tropical fruit}    0.0011       0.85
8.1
## 71  {grapes,
##      root vegetables,
##      tropical fruit}           => {other vegetables}  0.0012       0.80
4.1
## 72  {grapes,
##      tropical fruit,
##      yogurt}                   => {other vegetables}  0.0014       0.82
4.3
## 73  {grapes,
##      tropical fruit,
##      whole milk}               => {other vegetables}  0.0020       0.80
4.1
## 74  {meat,
##      tropical fruit,
##      whole milk}               => {other vegetables}  0.0010       0.83
4.3
## 75  {meat,
##      root vegetables,
##      yogurt}                   => {other vegetables}  0.0012       0.86
4.4
## 76  {curd,
##      frozen meals,
##      yogurt}                   => {whole milk}        0.0011       0.85
3.3
```

```
## 77  {frankfurter,
##      frozen meals,
##      tropical fruit}           => {other vegetables}  0.0010       0.91
4.7
## 78  {frankfurter,
##      frozen meals,
##      tropical fruit}           => {whole milk}        0.0010       0.91
3.6
## 79  {frankfurter,
##      frozen meals,
##      other vegetables}         => {whole milk}        0.0012       0.80
3.1
## 80  {butter,
##      frozen meals,
##      tropical fruit}           => {whole milk}        0.0010       0.91
3.6
## 81  {frozen meals,
##      root vegetables,
##      tropical fruit}           => {whole milk}        0.0011       0.85
3.3
## 82  {frozen meals,
##      tropical fruit,
##      yogurt}                   => {whole milk}        0.0016       0.80
3.1
## 83  {butter,
##      hard cheese,
##      yogurt}                   => {whole milk}        0.0013       0.81
3.2
## 84  {hard cheese,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0010       0.91
4.7
## 85  {hard cheese,
##      root vegetables,
##      whipped/sour cream}       => {other vegetables}  0.0013       0.81
4.2
## 86  {hard cheese,
##      tropical fruit,
##      yogurt}                   => {whole milk}        0.0014       0.82
3.2
## 87  {butter milk,
##      dessert,
##      yogurt}                   => {whole milk}        0.0013       0.81
3.2
## 88  {butter milk,
##      pork,
##      whole milk}               => {other vegetables}  0.0010       0.91
4.7
## 89  {butter milk,
##      fruit/vegetable juice,
```

```
##      pip fruit}                => {other vegetables}  0.0010       0.91
4.7
## 90  {butter milk,
##      pip fruit,
##      root vegetables}          => {other vegetables}  0.0012       0.86
4.4
## 91  {butter milk,
##      other vegetables,
##      pastry}                   => {yogurt}            0.0012       0.80
5.7
## 92  {butter milk,
##      sausage,
##      yogurt}                   => {whole milk}        0.0011       0.85
3.3
## 93  {butter milk,
##      root vegetables,
##      yogurt}                   => {whole milk}        0.0015       0.88
3.5
## 94  {candy,
##      rolls/buns,
##      root vegetables}          => {other vegetables}  0.0010       0.83
4.3
## 95  {frozen vegetables,
##      ham,
##      whole milk}               => {other vegetables}  0.0010       0.83
4.3
## 96  {ham,
##      pip fruit,
##      tropical fruit}           => {other vegetables}  0.0016       0.89
4.6
## 97  {frankfurter,
##      root vegetables,
##      sliced cheese}            => {whole milk}        0.0010       0.91
3.6
## 98  {frankfurter,
##      sliced cheese,
##      yogurt}                   => {whole milk}        0.0010       0.83
3.3
## 99  {butter,
##      sliced cheese,
##      whipped/sour cream}       => {whole milk}        0.0012       0.92
3.6
## 100 {pip fruit,
##      sausage,
##      sliced cheese}            => {yogurt}            0.0012       0.86
6.1
## 101 {coffee,
##      oil,
##      yogurt}                   => {other vegetables}  0.0010       0.91
4.7
```

```
## 102 {citrus fruit,
##      fruit/vegetable juice,
##      oil}                      => {other vegetables}  0.0011       0.85
4.4
## 103 {fruit/vegetable juice,
##      oil,
##      tropical fruit}           => {other vegetables}  0.0012       0.86
4.4
## 104 {oil,
##      root vegetables,
##      shopping bags}            => {whole milk}        0.0010       0.83
3.3
## 105 {oil,
##      root vegetables,
##      tropical fruit}           => {other vegetables}  0.0017       0.85
4.4
## 106 {frozen vegetables,
##      onions,
##      root vegetables}          => {other vegetables}  0.0013       0.87
4.5
## 107 {curd,
##      onions,
##      yogurt}                   => {whole milk}        0.0011       0.85
3.3
## 108 {napkins,
##      onions,
##      root vegetables}          => {other vegetables}  0.0010       0.91
4.7
## 109 {napkins,
##      onions,
##      whole milk}               => {other vegetables}  0.0012       0.80
4.1
## 110 {butter,
##      domestic eggs,
##      onions}                   => {whole milk}        0.0011       0.85
3.3
## 111 {bottled water,
##      butter,
##      onions}                   => {whole milk}        0.0010       0.83
3.3
## 112 {butter,
##      onions,
##      tropical fruit}           => {whole milk}        0.0012       0.86
3.4
## 113 {butter,
##      onions,
##      root vegetables}          => {whole milk}        0.0017       0.85
3.3
## 114 {butter,
##      onions,
```

```
##      yogurt}                   => {whole milk}        0.0013       0.81
3.2
## 115 {citrus fruit,
##      onions,
##      root vegetables}          => {other vegetables}  0.0014       0.82
4.3
## 116 {onions,
##      root vegetables,
##      tropical fruit}           => {other vegetables}  0.0016       0.89
4.6
## 117 {berries,
##      butter,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 118 {berries,
##      butter,
##      sausage}                  => {whole milk}        0.0010       0.91
3.6
## 119 {curd,
##      hamburger meat,
##      other vegetables}         => {whole milk}        0.0016       0.80
3.1
## 120 {butter,
##      hamburger meat,
##      whipped/sour cream}       => {whole milk}        0.0012       0.80
3.1
## 121 {hamburger meat,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0010       0.91
4.7
## 122 {hamburger meat,
##      root vegetables,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 123 {butter,
##      hygiene articles,
##      napkins}                  => {whole milk}        0.0010       0.91
3.6
## 124 {hygiene articles,
##      napkins,
##      tropical fruit}           => {whole milk}        0.0012       0.80
3.1
## 125 {hygiene articles,
##      margarine,
##      rolls/buns}               => {whole milk}        0.0010       0.83
3.3
## 126 {butter,
##      hygiene articles,
##      pip fruit}                => {whole milk}        0.0010       1.00
3.9
```

```
## 127 {butter,
##      citrus fruit,
##      hygiene articles}         => {whole milk}        0.0010       0.83
3.3
## 128 {bottled water,
##      butter,
##      hygiene articles}         => {whole milk}        0.0012       0.86
3.4
## 129 {butter,
##      hygiene articles,
##      tropical fruit}           => {whole milk}        0.0012       0.92
3.6
## 130 {butter,
##      hygiene articles,
##      root vegetables}          => {whole milk}        0.0014       0.82
3.2
## 131 {domestic eggs,
##      hygiene articles,
##      tropical fruit}           => {whole milk}        0.0012       0.92
3.6
## 132 {hygiene articles,
##      tropical fruit,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 133 {hygiene articles,
##      root vegetables,
##      whipped/sour cream}       => {whole milk}        0.0010       1.00
3.9
## 134 {hygiene articles,
##      pip fruit,
##      sausage}                  => {whole milk}        0.0013       0.81
3.2
## 135 {hygiene articles,
##      pip fruit,
##      root vegetables}          => {whole milk}        0.0010       1.00
3.9
## 136 {citrus fruit,
##      hygiene articles,
##      root vegetables}          => {whole milk}        0.0012       0.86
3.4
## 137 {hygiene articles,
##      root vegetables,
##      yogurt}                   => {whole milk}        0.0012       0.86
3.4
## 138 {long life bakery product,
##      other vegetables,
##      salty snack}              => {whole milk}        0.0010       0.83
3.3
## 139 {salty snack,
##      tropical fruit,
```

```
##      whipped/sour cream}       => {other vegetables}  0.0012       0.80
4.1
## 140 {pip fruit,
##      salty snack,
##      yogurt}                   => {whole milk}        0.0011       0.85
3.3
## 141 {salty snack,
##      tropical fruit,
##      yogurt}                   => {other vegetables}  0.0013       0.81
4.2
## 142 {root vegetables,
##      salty snack,
##      yogurt}                   => {other vegetables}  0.0012       0.86
4.4
## 143 {cream cheese,
##      domestic eggs,
##      sugar}                    => {whole milk}        0.0011       1.00
3.9
## 144 {cream cheese,
##      other vegetables,
##      sugar}                    => {whole milk}        0.0015       0.94
3.7
## 145 {beef,
##      root vegetables,
##      sugar}                    => {other vegetables}  0.0011       0.85
4.4
## 146 {curd,
##      domestic eggs,
##      sugar}                    => {whole milk}        0.0010       1.00
3.9
## 147 {butter,
##      sugar,
##      whipped/sour cream}       => {other vegetables}  0.0010       0.83
4.3
## 148 {butter,
##      sugar,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 149 {citrus fruit,
##      domestic eggs,
##      sugar}                    => {whole milk}        0.0014       0.93
3.7
## 150 {domestic eggs,
##      sugar,
##      tropical fruit}           => {whole milk}        0.0011       0.92
3.6
## 151 {domestic eggs,
##      sugar,
##      yogurt}                   => {whole milk}        0.0014       0.93
3.7
```

```
## 152 {citrus fruit,
##      sugar,
##      whipped/sour cream}       => {whole milk}        0.0011       0.85
3.3
## 153 {root vegetables,
##      sugar,
##      whipped/sour cream}       => {whole milk}        0.0012       0.92
3.6
## 154 {bottled water,
##      other vegetables,
##      sugar}                    => {whole milk}        0.0013       0.81
3.2
## 155 {pork,
##      rolls/buns,
##      waffles}                  => {whole milk}        0.0010       0.91
3.6
## 156 {rolls/buns,
##      waffles,
##      whipped/sour cream}       => {whole milk}        0.0017       0.81
3.2
## 157 {rolls/buns,
##      root vegetables,
##      waffles}                  => {whole milk}        0.0016       0.84
3.3
## 158 {long life bakery product,
##      napkins,
##      whipped/sour cream}       => {whole milk}        0.0010       0.91
3.6
## 159 {long life bakery product,
##      napkins,
##      tropical fruit}           => {whole milk}        0.0012       0.92
3.6
## 160 {long life bakery product,
##      napkins,
##      other vegetables}         => {whole milk}        0.0012       0.86
3.4
## 161 {butter,
##      long life bakery product,
##      whipped/sour cream}       => {other vegetables}  0.0012       0.80
4.1
## 162 {butter,
##      long life bakery product,
##      sausage}                  => {whole milk}        0.0010       0.91
3.6
## 163 {long life bakery product,
##      sausage,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 164 {long life bakery product,
##      whipped/sour cream,
```

```
##      yogurt}                   => {whole milk}        0.0017       0.81
3.2
## 165 {long life bakery product,
##      root vegetables,
##      tropical fruit}           => {other vegetables}  0.0011       0.85
4.4
## 166 {dessert,
##      sausage,
##      whipped/sour cream}       => {other vegetables}  0.0010       0.83
4.3
## 167 {dessert,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0011       0.92
4.7
## 168 {cream cheese,
##      curd,
##      root vegetables}          => {other vegetables}  0.0017       0.85
4.4
## 169 {cream cheese,
##      domestic eggs,
##      napkins}                  => {whole milk}        0.0011       1.00
3.9
## 170 {cream cheese,
##      pork,
##      yogurt}                   => {whole milk}        0.0010       0.83
3.3
## 171 {cream cheese,
##      frankfurter,
##      yogurt}                   => {whole milk}        0.0010       0.83
3.3
## 172 {cream cheese,
##      margarine,
##      whipped/sour cream}       => {yogurt}            0.0010       0.83
6.0
## 173 {butter,
##      cream cheese,
##      whipped/sour cream}       => {whole milk}        0.0011       0.85
3.3
## 174 {butter,
##      cream cheese,
##      root vegetables}          => {yogurt}            0.0010       0.91
6.5
## 175 {butter,
##      cream cheese,
##      root vegetables}          => {whole milk}        0.0010       0.91
3.6
## 176 {cream cheese,
##      domestic eggs,
##      whipped/sour cream}       => {whole milk}        0.0012       0.86
3.4
```

```
## 177 {cream cheese,
##      domestic eggs,
##      pip fruit}                => {whole milk}        0.0011       0.85
3.3
## 178 {citrus fruit,
##      cream cheese,
##      domestic eggs}            => {whole milk}        0.0016       0.89
3.5
## 179 {cream cheese,
##      domestic eggs,
##      yogurt}                   => {whole milk}        0.0013       0.81
3.2
## 180 {cream cheese,
##      pip fruit,
##      whipped/sour cream}       => {whole milk}        0.0013       0.93
3.6
## 181 {citrus fruit,
##      cream cheese,
##      whipped/sour cream}       => {other vegetables}  0.0013       0.81
4.2
## 182 {cream cheese,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0014       0.88
4.5
## 183 {cream cheese,
##      pip fruit,
##      sausage}                  => {whole milk}        0.0010       0.91
3.6
## 184 {citrus fruit,
##      cream cheese,
##      root vegetables}          => {other vegetables}  0.0012       0.92
4.8
## 185 {butter,
##      chicken,
##      whipped/sour cream}       => {whole milk}        0.0012       0.80
3.1
## 186 {chicken,
##      domestic eggs,
##      sausage}                  => {whole milk}        0.0012       0.86
3.4
## 187 {chicken,
##      pastry,
##      root vegetables}          => {other vegetables}  0.0010       0.83
4.3
## 188 {butter,
##      tropical fruit,
##      white bread}              => {yogurt}            0.0011       0.85
6.1
## 189 {butter,
##      tropical fruit,
```

```
##      white bread}              => {other vegetables}  0.0011       0.85
4.4
## 190 {butter,
##      root vegetables,
##      white bread}              => {other vegetables}  0.0010       0.83
4.3
## 191 {butter,
##      root vegetables,
##      white bread}              => {whole milk}        0.0011       0.92
3.6
## 192 {tropical fruit,
##      whipped/sour cream,
##      white bread}              => {other vegetables}  0.0017       0.85
4.4
## 193 {root vegetables,
##      whipped/sour cream,
##      white bread}              => {other vegetables}  0.0014       0.88
4.5
## 194 {root vegetables,
##      whipped/sour cream,
##      white bread}              => {whole milk}        0.0013       0.81
3.2
## 195 {chocolate,
##      napkins,
##      root vegetables}          => {whole milk}        0.0012       0.80
3.1
## 196 {chocolate,
##      domestic eggs,
##      sausage}                  => {whole milk}        0.0010       0.83
3.3
## 197 {bottled beer,
##      coffee,
##      yogurt}                   => {whole milk}        0.0010       0.83
3.3
## 198 {butter,
##      coffee,
##      whipped/sour cream}       => {whole milk}        0.0012       0.92
3.6
## 199 {coffee,
##      domestic eggs,
##      root vegetables}          => {whole milk}        0.0010       0.91
3.6
## 200 {citrus fruit,
##      frozen vegetables,
##      napkins}                  => {whole milk}        0.0011       0.85
3.3
## 201 {frozen vegetables,
##      napkins,
##      other vegetables}         => {whole milk}        0.0013       0.81
3.2
```

```
## 202 {frozen vegetables,
##      margarine,
##      rolls/buns}               => {whole milk}        0.0013       0.87
3.4
## 203 {citrus fruit,
##      frozen vegetables,
##      fruit/vegetable juice}    => {whole milk}        0.0016       0.84
3.3
## 204 {beef,
##      butter,
##      curd}                     => {whole milk}        0.0010       0.83
3.3
## 205 {beef,
##      butter,
##      tropical fruit}           => {yogurt}            0.0010       0.83
6.0
## 206 {beef,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0014       0.88
4.5
## 207 {beef,
##      tropical fruit,
##      whipped/sour cream}       => {whole milk}        0.0013       0.81
3.2
## 208 {curd,
##      margarine,
##      rolls/buns}               => {whole milk}        0.0013       0.81
3.2
## 209 {butter,
##      curd,
##      domestic eggs}            => {other vegetables}  0.0010       0.83
4.3
## 210 {butter,
##      curd,
##      domestic eggs}            => {whole milk}        0.0011       0.92
3.6
## 211 {butter,
##      curd,
##      pip fruit}                => {whole milk}        0.0012       0.80
3.1
## 212 {butter,
##      citrus fruit,
##      curd}                     => {whole milk}        0.0011       0.92
3.6
## 213 {curd,
##      domestic eggs,
##      other vegetables}         => {whole milk}        0.0028       0.82
3.2
## 214 {curd,
##      pip fruit,
```

```
##      whipped/sour cream}       => {whole milk}        0.0018       0.82
3.2
## 215 {bottled beer,
##      napkins,
##      rolls/buns}               => {whole milk}        0.0012       0.80
3.1
## 216 {butter,
##      napkins,
##      whipped/sour cream}       => {whole milk}        0.0014       0.82
3.2
## 217 {bottled water,
##      butter,
##      napkins}                  => {whole milk}        0.0011       0.85
3.3
## 218 {butter,
##      napkins,
##      yogurt}                   => {whole milk}        0.0011       0.85
3.3
## 219 {domestic eggs,
##      napkins,
##      tropical fruit}           => {whole milk}        0.0013       0.81
3.2
## 220 {bottled beer,
##      pork,
##      rolls/buns}               => {whole milk}        0.0011       0.85
3.3
## 221 {butter,
##      pork,
##      whipped/sour cream}       => {whole milk}        0.0014       0.88
3.4
## 222 {butter,
##      pork,
##      yogurt}                   => {whole milk}        0.0014       0.82
3.2
## 223 {butter,
##      other vegetables,
##      pork}                     => {whole milk}        0.0022       0.85
3.3
## 224 {fruit/vegetable juice,
##      pork,
##      tropical fruit}           => {yogurt}            0.0010       0.83
6.0
## 225 {pip fruit,
##      pork,
##      soda}                     => {other vegetables}  0.0011       0.85
4.4
## 226 {bottled beer,
##      domestic eggs,
##      margarine}                => {whole milk}        0.0010       0.91
3.6
```

```
## 227 {brown bread,
##      domestic eggs,
##      root vegetables}          => {whole milk}        0.0015       0.83
3.3
## 228 {brown bread,
##      pip fruit,
##      whipped/sour cream}       => {other vegetables}  0.0011       1.00
5.2
## 229 {brown bread,
##      sausage,
##      whipped/sour cream}       => {other vegetables}  0.0010       0.83
4.3
## 230 {brown bread,
##      pip fruit,
##      root vegetables}          => {other vegetables}  0.0013       0.81
4.2
## 231 {brown bread,
##      pip fruit,
##      root vegetables}          => {whole milk}        0.0014       0.88
3.4
## 232 {butter,
##      margarine,
##      tropical fruit}           => {yogurt}            0.0011       0.85
6.1
## 233 {domestic eggs,
##      fruit/vegetable juice,
##      margarine}                => {whole milk}        0.0011       0.92
3.6
## 234 {domestic eggs,
##      margarine,
##      whipped/sour cream}       => {whole milk}        0.0012       0.80
3.1
## 235 {bottled water,
##      margarine,
##      whipped/sour cream}       => {whole milk}        0.0010       0.83
3.3
## 236 {margarine,
##      rolls/buns,
##      whipped/sour cream}       => {whole milk}        0.0016       0.89
3.5
## 237 {butter,
##      domestic eggs,
##      whipped/sour cream}       => {whole milk}        0.0016       0.84
3.3
## 238 {butter,
##      domestic eggs,
##      pip fruit}                => {whole milk}        0.0012       0.86
3.4
## 239 {butter,
##      pip fruit,
```

```
##      whipped/sour cream}       => {whole milk}        0.0018       0.90
3.5
## 240 {bottled water,
##      butter,
##      whipped/sour cream}       => {whole milk}        0.0012       0.86
3.4
## 241 {butter,
##      soda,
##      whipped/sour cream}       => {other vegetables}  0.0013       0.93
4.8
## 242 {butter,
##      pastry,
##      pip fruit}                => {other vegetables}  0.0013       0.93
4.8
## 243 {bottled water,
##      butter,
##      pip fruit}                => {whole milk}        0.0013       0.81
3.2
## 244 {butter,
##      pip fruit,
##      root vegetables}          => {whole milk}        0.0017       0.81
3.2
## 245 {butter,
##      citrus fruit,
##      tropical fruit}           => {whole milk}        0.0016       0.80
3.1
## 246 {citrus fruit,
##      newspapers,
##      root vegetables}          => {other vegetables}  0.0016       0.84
4.4
## 247 {domestic eggs,
##      pastry,
##      whipped/sour cream}       => {other vegetables}  0.0012       0.86
4.4
## 248 {domestic eggs,
##      tropical fruit,
##      whipped/sour cream}       => {whole milk}        0.0018       0.90
3.5
## 249 {domestic eggs,
##      pip fruit,
##      sausage}                  => {whole milk}        0.0014       0.82
3.2
## 250 {domestic eggs,
##      pastry,
##      tropical fruit}           => {whole milk}        0.0013       0.81
3.2
## 251 {domestic eggs,
##      pastry,
##      root vegetables}          => {other vegetables}  0.0012       0.86
4.4
```

```
## 252 {fruit/vegetable juice,
##      pastry,
##      whipped/sour cream}       => {yogurt}            0.0012       0.80
5.7
## 253 {fruit/vegetable juice,
##      tropical fruit,
##      whipped/sour cream}       => {other vegetables}  0.0019       0.90
4.7
## 254 {pip fruit,
##      sausage,
##      whipped/sour cream}       => {whole milk}        0.0012       0.80
3.1
## 255 {citrus fruit,
##      pastry,
##      whipped/sour cream}       => {whole milk}        0.0015       0.88
3.5
## 256 {bottled water,
##      sausage,
##      whipped/sour cream}       => {other vegetables}  0.0013       0.81
4.2
## 257 {citrus fruit,
##      pastry,
##      root vegetables}          => {other vegetables}  0.0015       0.88
4.6
## 258 {pastry,
##      root vegetables,
##      shopping bags}            => {other vegetables}  0.0011       0.85
4.4
```

Here, I remove all of the associations found from the model that are redundant. That is, these associations actually display associations that have already been established and are therefore repetitive and unnecessary.

```
subset.matrix <- is.subset(rules, rules)
subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
redundant <- colSums(subset.matrix, na.rm=T) >= 1
rules.pruned <- rules[!redundant]
rules<-rules.pruned
```

Here, I find the top five associations for one of the more popular items, rolls/buns, from the newly pruned list of associations. The top association for rolls/buns appears to be when a basket contains newspapers and spread cheese. This makes sense, as this appears to be a purchase of roll/buns in a breakfast/early day meal purchase. Another strong association if with beef, tropical fruit, whole milk, and yogurt. This appears to be a more rounded purchase, that makes sense mainly due to the beef.

```
rules<-apriori(data=groceries, parameter=list(supp=0.001,conf = 0.08,
minlen=3),
               appearance = list(default="lhs",rhs="rolls/buns"),
               control = list(verbose=F))
```

```
rules<-sort(rules, decreasing=TRUE,by="confidence")
arules::inspect(rules[1:5])
```

```
##   lhs                     rhs          support confidence lift
## 1 {newspapers,
##    spread cheese}      => {rolls/buns}  0.0012       0.75  4.1
## 2 {beef,
##    tropical fruit,
##    whole milk,
##    yogurt}             => {rolls/buns}  0.0013       0.68  3.7
## 3 {citrus fruit,
##    pastry,
##    whipped/sour cream,
##    whole milk}         => {rolls/buns}  0.0010       0.67  3.6
## 4 {candy,
##    chocolate,
##    soda}               => {rolls/buns}  0.0012       0.63  3.4
## 5 {chocolate,
##    napkins,
##    other vegetables}   => {rolls/buns}  0.0010       0.62  3.4
```

Next, I found the flipped association. That is, what purchases a basket with rolls/buns led to the most. However, the confidence and lifts from these associations are fairly low, because all of the associations are for other popular items, meaning that these other items in the basket aren't necessarily there because of the rolls/buns, but rather just because they are popular items.

```
rules<-apriori(data=groceries, parameter=list(supp=0.001,conf =
0.15,minlen=2),
               appearance = list(default="rhs",lhs="rolls/buns"),
               control = list(verbose=F))
rules<-sort(rules, decreasing=TRUE,by="confidence")
arules::inspect(rules[1:5])
```

```
##   lhs             rhs                support confidence lift
## 5 {rolls/buns} => {whole milk}       0.057   0.31       1.2
## 4 {rolls/buns} => {other vegetables} 0.043   0.23       1.2
## 2 {rolls/buns} => {soda}             0.038   0.21       1.2
## 3 {rolls/buns} => {yogurt}           0.034   0.19       1.3
## 1 {rolls/buns} => {sausage}          0.031   0.17       1.8
```