Fields Exercises

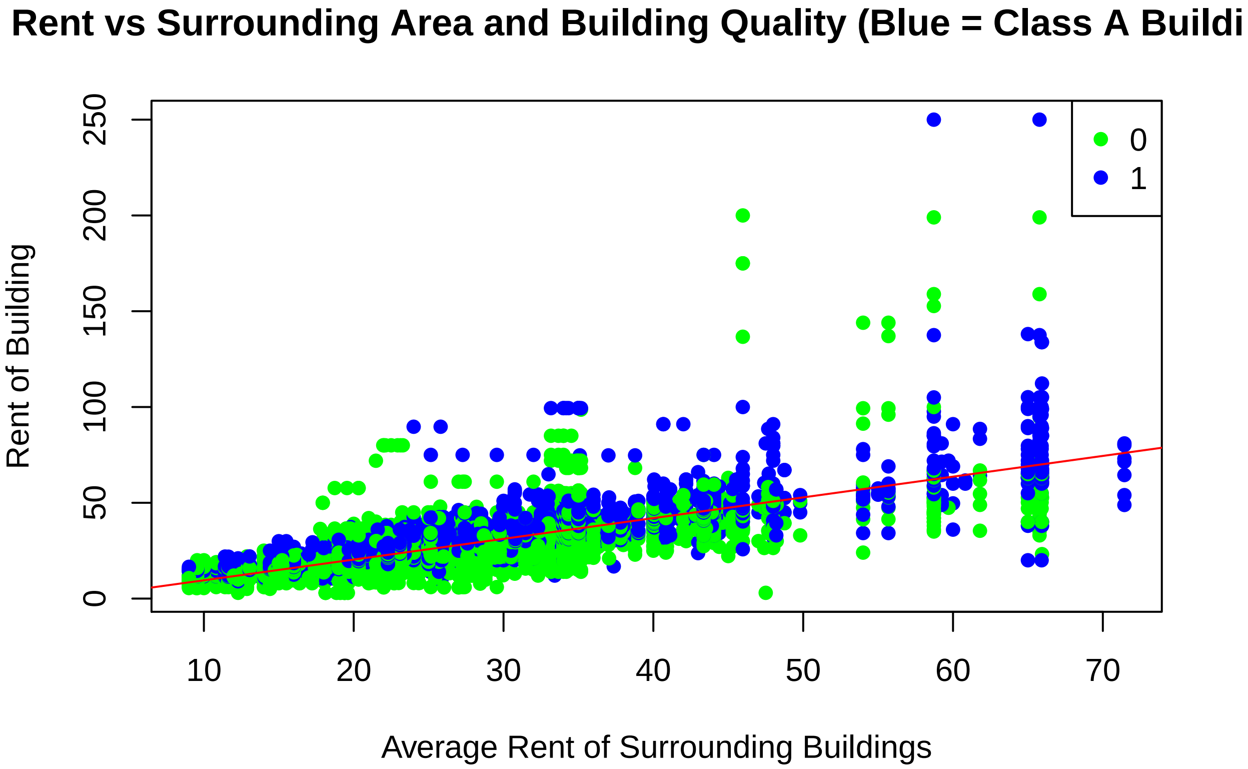
Scott Fields

8/11/2021

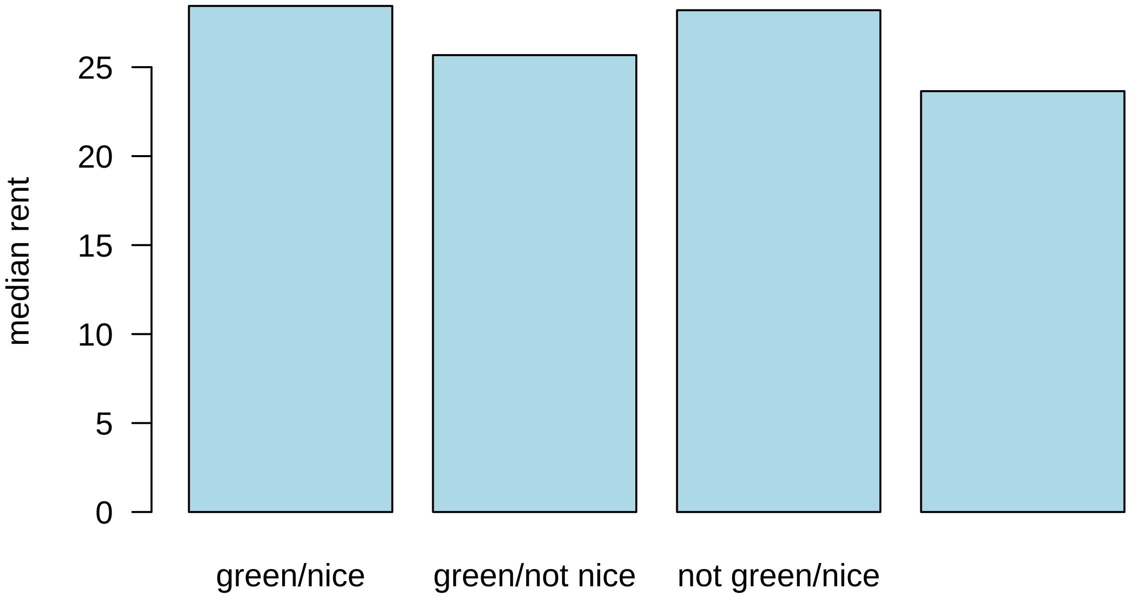
Scott Fields full code at: https://github.com/shfields/ML-Exercises

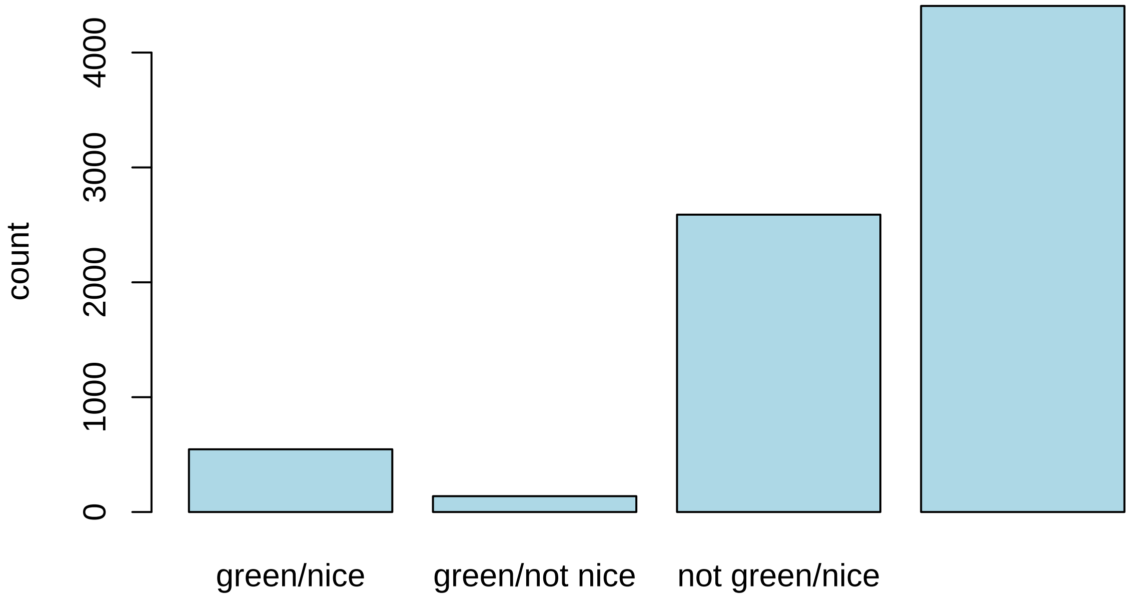
Visual Story Telling Part 1

I disagree with the previous conclusions as there seem to be some unaccounted for variables in his process. As we can see below the rent of a building strongly correlates with the price of buildings around it and then within similar neighborhoods, the most expensive buildings are consistently Class A.



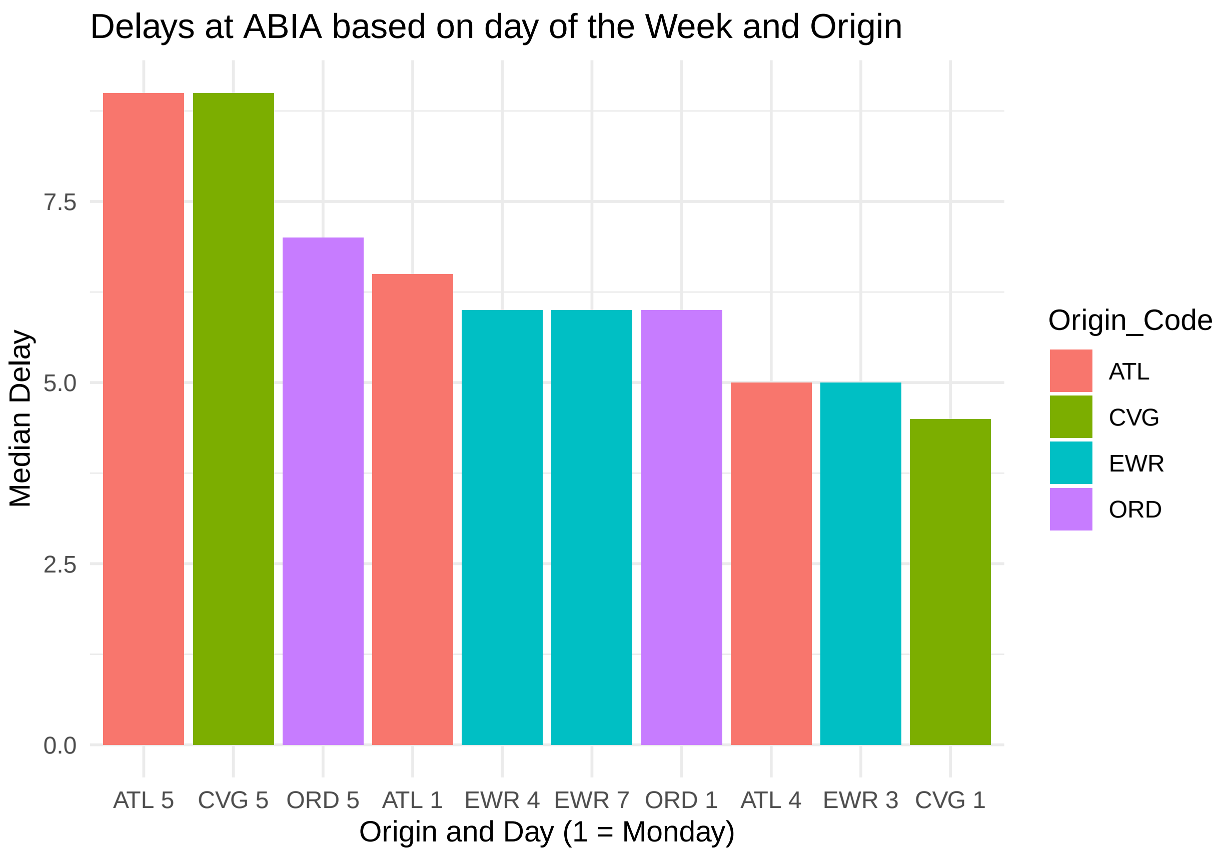
Additionally we can see that while green buildings do have a higher average rent than non green buildings that seems to be because most green buildings are Class A. It is a reasonable conclusion that if you build a nice building in a nice area you will get a high rent, making a green building may help, but its hard to say whether it would cover the extra costs.


median rent
0
5
10
15
20
25
green/nice
green/not nice
not green/nice



Visual Story Telling Part 2

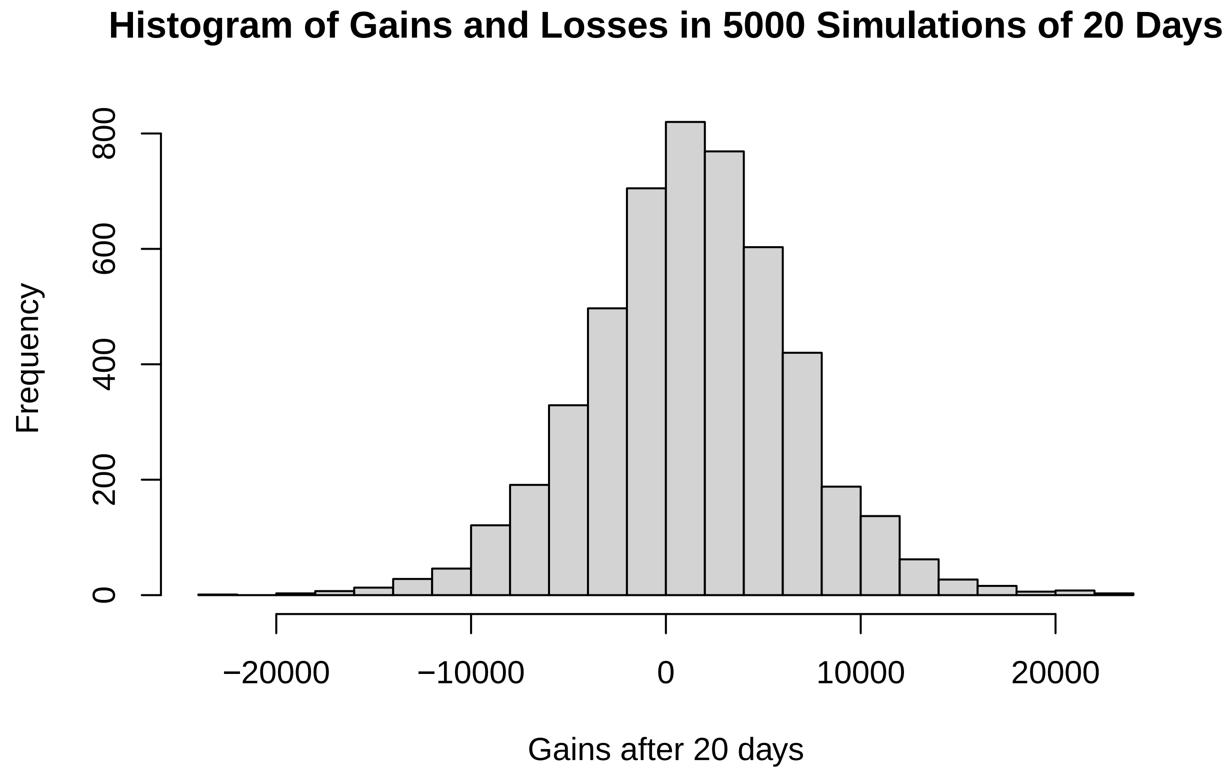
The 10 worst flights coming in to Austin by Origin and Day are from just 4 cities and 6 of the worst are on Mondays and Fridays.



Portfolio Modeling

The three portfolios I made were a large growth portfolio, a technology portfolio, and an oil and gas portfolio. The large growth portfolio was made up of SPY (S&P 500), Vanguard Total World, and QQQ, a NASDAQ composite. The Tech collection was made up of Vanguard IT ETF, a cloud computing ETF, a Semiconductors ETF, and a FinTech ETF. THe O&G portfolio was made up of a US oil ETF, a US natural gas ETF, and a BRENT (European Oil) ETF.

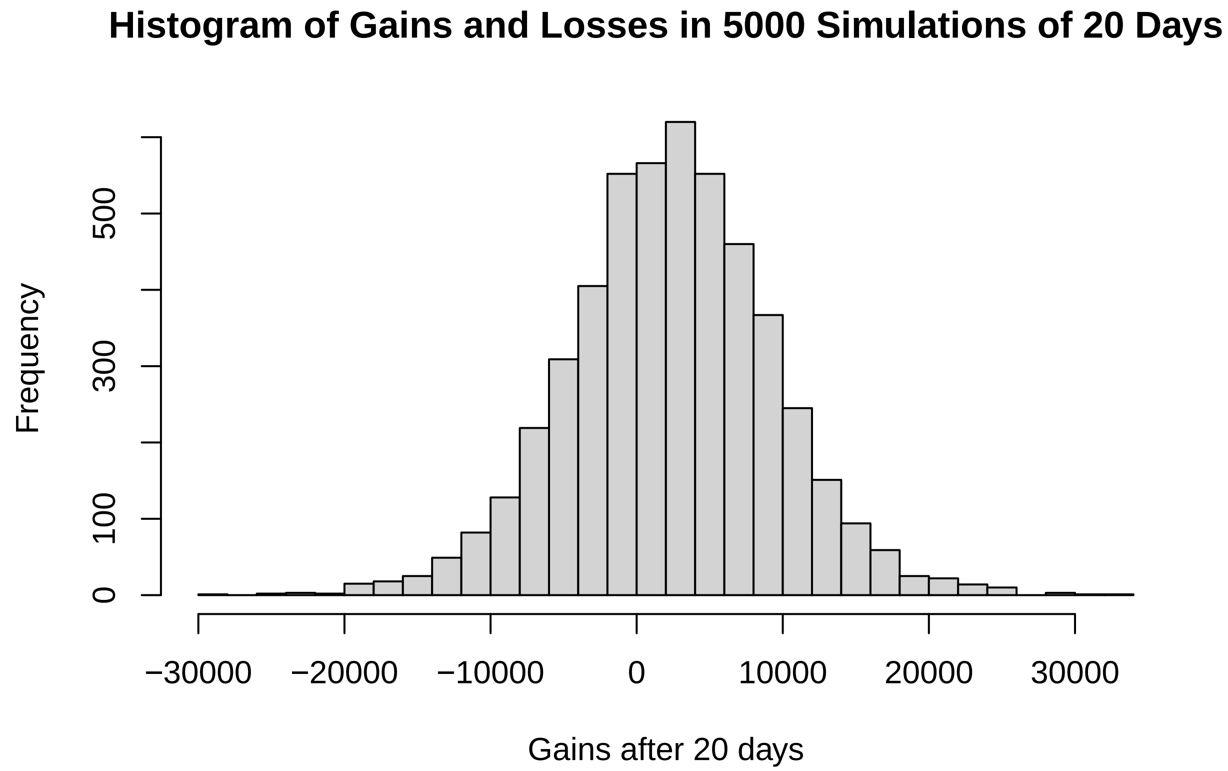
Large Growth
Large growth stocks performed well with an average gain of over $1,000, people normally invest in these because of how steady their growth is as they encapsulate the whole market.



```
## [1] "Average Gain: $ 1327.5"
```

```
## [1] "Value at Risk at 5%: -7641.99"
```

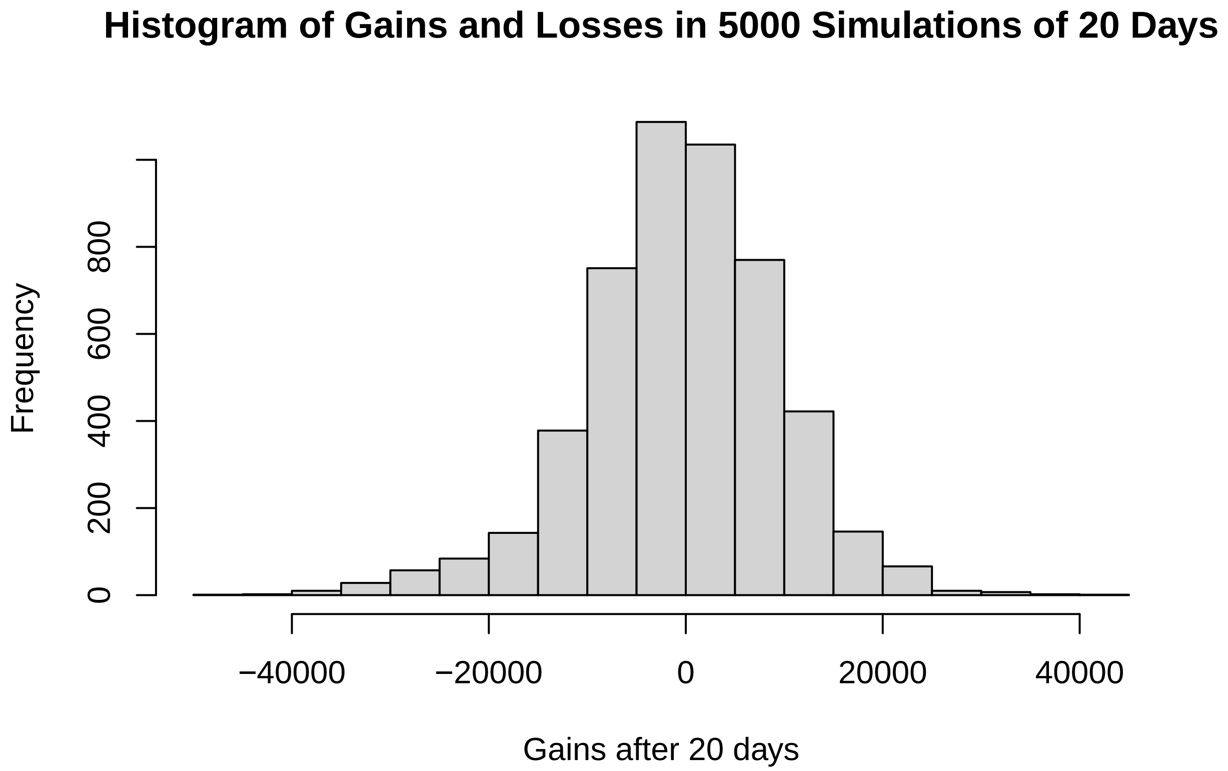
Technology
I made one portfolio technology ETFs because they have a tendency to make a lot of gains while also being quite volatile and that's exactly what we got from the simulations. While the average return was higher than large growth, the 5% worst case scenario was also worse, showing a lot of risk from this portfolio.



```
## [1] "Average Gain: $ 2345.78"
```

```
## [1] "Value at Risk at 5%: -9003.2"
```

Oil and Gas
I wanted one of the portfolios to be based on commodities because they have a tendency to be very stable relative to the stock market but this oil and gas portfolio performed very poorly with its average return and VaR being significantly worse than both Large Growth and Technology portfolios.



```
## [1] "Average Gain: $ -427.61"
```

```
## [1] "Value at Risk at 5%: -17122.52"
```

Market Segmentation

I applied a hierarchical clustering model to group customers into 8 main clusters with 5 of them being sizable enough that I would recommend focusing on them as markets.

Below are the clusters with the number of customers in each.

```
## Loading required package: Rcpp
```

```
## Loading required package: RcppZiggurat
```

```
##
## Attaching package: ’Rfast’
```

```
## The following object is masked from ’package:dplyr’:
##
##     nth
```

```
##    1    2    3    4    5    6    7    8
##  496 6038  284  859  130   16   49   10
```

Starting with the biggest cluster, #2, we're going to be able to learn the least about these customers because there are so many of them. Below we can see the top three classes of tweets this cluster sends and it seems like the biggest group of customers for NutrientH20 are college students that play video games. So marketing to young people and understanding online culture is a must for this brand.

```
## college_uni
##   0.0457463
```

```
## online_gaming
##    0.03981552
```

```
##    chatter
## 0.03183217
```

CLuster 1 on the other hand seems to be focused on people living an active healthy lifestyle. Marketing should emphasize the health benefits that I assume NutrientH20 has.

```
## health_nutrition
##         2.470149
```

```
## personal_fitness
##         2.417278
```

```
## outdoors
## 1.949973
```

Cluster 3 seems to be focused on art and entertainment, if further research found that there was a specific tv show or movie series was particularly resonant with this market, a celebrity endorsement from said tv show or movie might go a long way.

```
##     art
## 3.79679
```

```
##  tv_film
## 2.502765
```

```
##    crafts
## 0.9332342
```

Cluster 4 seems to mostly use twitter for talking about politics and news so there might not be much to learn here from a marketing standpoint.

```
## politics
## 1.834511
```

```
##     news
## 1.762441
```

```
##   travel
## 1.202322
```

Cluster 5 seems to be the classic stereotypical “Faith, Football, Family” guy, so marketing towards parents and sports fans may be a wise choice.

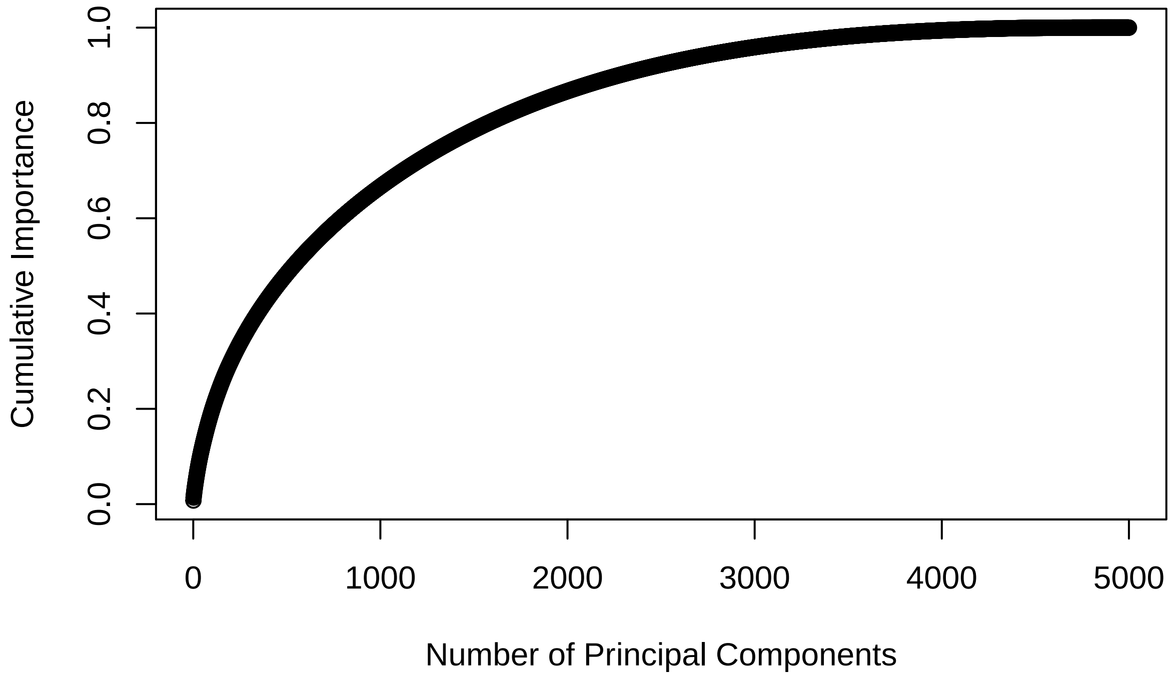
```
## religion
## 3.621921
```

```
## parenting
##   3.52405
```

```
## sports_fandom
##      3.398765
```

Austhor Attribution

For this problem I began by reading in all the training files and converted all words to lower case, removed numbers, punctuation, and white space, and removed stop words. Then I created a Document Term Matrix of each file and removed sparse terms at a 95% level.I added a column with the authors name and then combined all the matricies using bind_rows(). Then I repeated the entire process wit the test data but limited their matricies to the columns from the training set and combined the train and test data to perform Principal Components Analysis. Below is the chart of cumulative importanmce as the number of components increased.



I decided to use the first 1,000 Principal Components as my predictors to reduce dimensionality while still retaining about 67% of total importance and prediction power. I trained my [2500 row X 1000 column] training set on an xgboost model that used 5 fold cross validation on 50 rounds, and had an out of fold training accuracy of almost 89%

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
##         1  44  0  0  0  0  0  0  0  0  0  0  0  0  0  4  1  0  0  0  1  0  0  3
##         2   0 45  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0  0  0
##         3   0  0 43  0  0  0  0  0  0  1  0  0  0  0  2  0  0  0  0  0  0  4  0
##         4   0  0  0 42  0  0  0  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         5   0  0  1  0 44  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  1  0
##         6   0  0  0  0  0 46  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         7   1  0  2  0  0  1 45  0  0  0  0  0  0  0  1  0  3  0  0  0  0  0  0
##         8   0  0  0  1  0  0  0 39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         9   0  0  0  0  0  0  0  0 48  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         10  0  0  0  0  0  0  0  0  0 46  0  1  0  0  0  0  1  0  0  0  0  0  0
##         11  0  0  0  0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0  0  1  0
##         12  0  1  0  0  0  0  0  0  0  1  0 41  0  0  0  0  0  0  0  0  0  0  0
##         13  0  0  0  0  0  0  0  0  0  0  0  0 44  1  0  8  0  0  0  0  0  0  0
##         14  0  0  0  0  0  0  0  0  0  0  0  0  0 48  0  0  0  0  0  0  0  1  0
##         15  1  0  0  0  1  0  0  0  0  0  0  0  0  0 38  0  0  0  0  0  0  0  0
##         16  0  0  0  0  0  0  0  0  0  0  0  0  5  0  0 39  0  0  0  0  0  0  0
##         17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 46  0  0  1  0  0  0
##         18  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0 45  0  1  0  0  0
##         19  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 50  0  0  0  0
##         20  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  2  0 43  0  0  2
##         21  0  0  0  2  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0 49  0  0
##         22  0  1  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0 41  0
##         23  2  0  1  0  0  0  1  0  0  0  0  0  0  0  0  0  0  1  0  1  0  0 41
##         24  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  3  0  0  0
##         25  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         26  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         27  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0
##         28  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         29  0  0  1  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  1  0
##         30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         31  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         32  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  1
##         33  0  2  0  0  0  0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0
##         34  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  1  3
##         35  0  0  0  1  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         36  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         37  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         38  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         40  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
##         41  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0
##         42  0  0  0  0  0  0  1  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
##         43  2  0  0  0  0  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0
##         44  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
##         45  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         46  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         47  0  0  0  3  1  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         48  0  0  0  0  0  0  0  0  0  1  0  3  0  0  0  1  0  0  0  0  0  0  0
##         49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         50  0  0  0  0  2  0  0  1  0  0  0  0  1  0  0  1  0  0  0  0  0  0  0
##           Reference
## Prediction 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
##          1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  3  0  0  0
##          2  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0  0  0  0  0
##          3  0  0  0  0  0  1  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##          4  0  3  0  0  0  0  2  0  2  0  1  0  0  0  0  2  0  0  0  0  0  0  0
##          5  0  1  0  0  0  0  0  0  0  0  0  1  0  0  0  2  0  0  0  0  0  1  0
##          6  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##          7  0  0  0  0  0  0  0  0  0  0  0  1  0  1  0  0  0  0  0  1  0  0  0
##          8  2  0  0  0  0  0  0  0  1  0  0  0  0  0  0  1  3  0  0  0  0  0  0
##          9  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         10  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  1  0  0  0  0  0
##         11  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         12  0  0  1  0  0  0  1  0  0  3  0  0  0  0  0  0  0  0  4  0  1  0  0
##         13  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
##         14  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         15  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0
##         16  0  0  0  1  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0
##         17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         18  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         19  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         20  0  0  0  0  0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0
##         21  0  0  0  0  0  0  0  0  0  0  0  2  0  0  5  0  0  0  0  0  0  2  0
##         22  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         23  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         24 44  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  0  0  0  0  0
##         25  0 46  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0
##         26  0  0 49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0
##         27  0  0  0 49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         28  0  0  0  0 48  0  0  0  0  0  0  0  0  2  0  0  0  0  1  1  0  0  0
##         29  0  0  0  0  0 45  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         30  0  0  0  0  0  0 44  0  0  0  0  1  0  0  0  0  2  0  0  0  0  0  0
##         31  0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         32  0  0  0  0  0  1  0  0 44  0  0  0  0  0  2  0  0  0  0  0  0  1  2
##         33  0  0  0  0  0  0  0  0  0 39  0  0  1  0  0  0  0  0  0  0  0  0  0
##         34  0  0  0  0  0  0  0  0  0  0 45  0  0  0  0  0  0  0  0  0  0  0  0
##         35  0  0  0  0  0  0  0  0  0  0  0 45  0  0  0  0  0  0  0  0  0  0  0
##         36  0  0  0  0  0  0  0  0  0  0  0  0 47  0  0  0  0  0  1  0  4  0  0
##         37  0  0  0  0  1  0  2  0  0  0  0  0  0 47  0  0  0  0  0  0  1  0  0
##         38  0  0  0  0  0  0  0  0  0  0  1  0  0  0 43  0  0  0  0  0  0  0  0
##         39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 45  0  0  0  0  0  0  0
##         40  2  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0 43  0  0  0  0  0  0
##         41  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 43  0  0  0  0  0
##         42  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 39  0  4  0  0
##         43  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 44  0  0  0
##         44  0  0  0  0  0  0  0  0  0  1  0  0  2  0  0  0  0  2  1  0 38  0  0
##         45  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 46  0
##         46  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 48
##         47  1  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
##         48  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  1  3  0  1  0  0
##         49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  0  0  0  0
##         50  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##           Reference
## Prediction 47 48 49 50
##         1   0  0  0  0
##         2   0  0  0  0
##         3   0  0  0  0
##         4   0  0  0  0
##         5   0  0  0  1
##         6   6  0  0  0
##         7   0  0  0  0
##         8   0  0  0  0
##         9   0  0  0  0
##         10  0  4  0  0
##         11  0  0  0  0
##         12  0  1  0  0
##         13  0  0  0  1
##         14  0  0  0  0
##         15  0  0  0  0
##         16  0  0  0  2
##         17  0  0  0  0
##         18  0  0  0  0
##         19  1  0  0  0
##         20  0  0  0  0
##         21  0  0  0  0
##         22  0  0  0  0
##         23  0  0  0  0
##         24  0  0  0  0
##         25  0  0  0  0
##         26  0  1  0  0
##         27  0  0  0  0
##         28  0  0  0  0
##         29  0  0  1  0
##         30  0  0  0  0
##         31  0  0  0  0
##         32  0  0  0  0
##         33  0  1  0  0
##         34  0  0  0  0
##         35  1  0  0  0
##         36  0  0  0  0
##         37  0  2  0  0
##         38  0  0  0  0
##         39  0  0  2  0
##         40  0  0  1  0
##         41  0  0  0  0
##         42  0  2  0  0
##         43  0  0  0  0
##         44  0  0  0  0
##         45  0  0  1  0
##         46  0  0  0  0
##         47 42  0  0  0
##         48  0 39  0  0
##         49  0  0 45  0
##         50  0  0  0 46
##
```

```
## Overall Statistics
##
##                Accuracy : 0.8872
##                  95% CI : (0.8741, 0.8993)
##     No Information Rate : 0.02
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8849
##
##  Mcnemar’s Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
## Sensitivity            0.8800   0.9000   0.8600   0.8400   0.8800   0.9200
## Specificity            0.9951   0.9976   0.9959   0.9947   0.9963   0.9976
## Pos Pred Value         0.7857   0.8824   0.8113   0.7636   0.8302   0.8846
## Neg Pred Value         0.9975   0.9980   0.9971   0.9967   0.9975   0.9984
## Prevalence             0.0200   0.0200   0.0200   0.0200   0.0200   0.0200
## Detection Rate         0.0176   0.0180   0.0172   0.0168   0.0176   0.0184
## Detection Prevalence   0.0224   0.0204   0.0212   0.0220   0.0212   0.0208
## Balanced Accuracy      0.9376   0.9488   0.9280   0.9173   0.9382   0.9588
##                      Class: 7 Class: 8 Class: 9 Class: 10 Class: 11 Class: 12
## Sensitivity            0.9000   0.7800   0.9600    0.9200    0.9800    0.8200
## Specificity            0.9955   0.9967   1.0000    0.9967    0.9996    0.9947
## Pos Pred Value         0.8036   0.8298   1.0000    0.8519    0.9800    0.7593
## Neg Pred Value         0.9980   0.9955   0.9992    0.9984    0.9996    0.9963
## Prevalence             0.0200   0.0200   0.0200    0.0200    0.0200    0.0200
## Detection Rate         0.0180   0.0156   0.0192    0.0184    0.0196    0.0164
## Detection Prevalence   0.0224   0.0188   0.0192    0.0216    0.0200    0.0216
## Balanced Accuracy      0.9478   0.8884   0.9800    0.9584    0.9898    0.9073
##                      Class: 13 Class: 14 Class: 15 Class: 16 Class: 17
## Sensitivity             0.8800    0.9600    0.7600    0.7800    0.9200
## Specificity             0.9951    0.9996    0.9988    0.9963    0.9996
## Pos Pred Value          0.7857    0.9796    0.9268    0.8125    0.9787
## Neg Pred Value          0.9975    0.9992    0.9951    0.9955    0.9984
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0176    0.0192    0.0152    0.0156    0.0184
## Detection Prevalence    0.0224    0.0196    0.0164    0.0192    0.0188
## Balanced Accuracy       0.9376    0.9798    0.8794    0.8882    0.9598
##                      Class: 18 Class: 19 Class: 20 Class: 21 Class: 22
## Sensitivity             0.9000    1.0000    0.8600    0.9800    0.8200
## Specificity             0.9992    0.9996    0.9971    0.9951    0.9971
## Pos Pred Value          0.9574    0.9804    0.8600    0.8033    0.8542
## Neg Pred Value          0.9980    1.0000    0.9971    0.9996    0.9963
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0180    0.0200    0.0172    0.0196    0.0164
## Detection Prevalence    0.0188    0.0204    0.0200    0.0244    0.0192
## Balanced Accuracy       0.9496    0.9998    0.9286    0.9876    0.9086
##                      Class: 23 Class: 24 Class: 25 Class: 26 Class: 27
## Sensitivity             0.8200    0.8800    0.9200    0.9800    0.9800
## Specificity             0.9971    0.9971    0.9992    0.9992    0.9996
## Pos Pred Value          0.8542    0.8627    0.9583    0.9608    0.9800
## Neg Pred Value          0.9963    0.9976    0.9984    0.9996    0.9996
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0164    0.0176    0.0184    0.0196    0.0196
## Detection Prevalence    0.0192    0.0204    0.0192    0.0204    0.0200
## Balanced Accuracy       0.9086    0.9386    0.9596    0.9896    0.9898
##                      Class: 28 Class: 29 Class: 30 Class: 31 Class: 32
## Sensitivity             0.9600    0.9000    0.8800    0.9800    0.8800
## Specificity             0.9984    0.9984    0.9988    1.0000    0.9967
## Pos Pred Value          0.9231    0.9184    0.9362    1.0000    0.8462
## Neg Pred Value          0.9992    0.9980    0.9976    0.9996    0.9975
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0192    0.0180    0.0176    0.0196    0.0176
## Detection Prevalence    0.0208    0.0196    0.0188    0.0196    0.0208
## Balanced Accuracy       0.9792    0.9492    0.9394    0.9900    0.9384
##                      Class: 33 Class: 34 Class: 35 Class: 36 Class: 37
## Sensitivity             0.7800     0.900    0.9000    0.9400    0.9400
## Specificity             0.9976     0.998    0.9984    0.9976    0.9976
## Pos Pred Value          0.8667     0.900    0.9184    0.8868    0.8868
## Neg Pred Value          0.9955     0.998    0.9980    0.9988    0.9988
## Prevalence              0.0200     0.020    0.0200    0.0200    0.0200
## Detection Rate          0.0156     0.018    0.0180    0.0188    0.0188
## Detection Prevalence    0.0180     0.020    0.0196    0.0212    0.0212
## Balanced Accuracy       0.8888     0.949    0.9492    0.9688    0.9688
##                      Class: 38 Class: 39 Class: 40 Class: 41 Class: 42
## Sensitivity             0.8600    0.9000    0.8600    0.8600    0.7800
## Specificity             0.9992    0.9992    0.9976    0.9996    0.9967
## Pos Pred Value          0.9556    0.9574    0.8776    0.9773    0.8298
## Neg Pred Value          0.9971    0.9980    0.9971    0.9971    0.9955
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0172    0.0180    0.0172    0.0172    0.0156
## Detection Prevalence    0.0180    0.0188    0.0196    0.0176    0.0188
## Balanced Accuracy       0.9296    0.9496    0.9288    0.9298    0.8884
##                      Class: 43 Class: 44 Class: 45 Class: 46 Class: 47
## Sensitivity             0.8800    0.7600    0.9200    0.9600    0.8400
## Specificity             0.9980    0.9971    0.9980    1.0000    0.9963
## Pos Pred Value          0.8980    0.8444    0.9020    1.0000    0.8235
## Neg Pred Value          0.9976    0.9951    0.9984    0.9992    0.9967
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0176    0.0152    0.0184    0.0192    0.0168
## Detection Prevalence    0.0196    0.0180    0.0204    0.0192    0.0204
## Balanced Accuracy       0.9390    0.8786    0.9590    0.9800    0.9182
##                      Class: 48 Class: 49 Class: 50
## Sensitivity             0.7800    0.9000    0.9200
## Specificity             0.9947    0.9992    0.9980
## Pos Pred Value          0.7500    0.9574    0.9020
## Neg Pred Value          0.9955    0.9980    0.9984
## Prevalence              0.0200    0.0200    0.0200
## Detection Rate          0.0156    0.0180    0.0184
## Detection Prevalence    0.0208    0.0188    0.0204
## Balanced Accuracy       0.8873    0.9496    0.9590
```

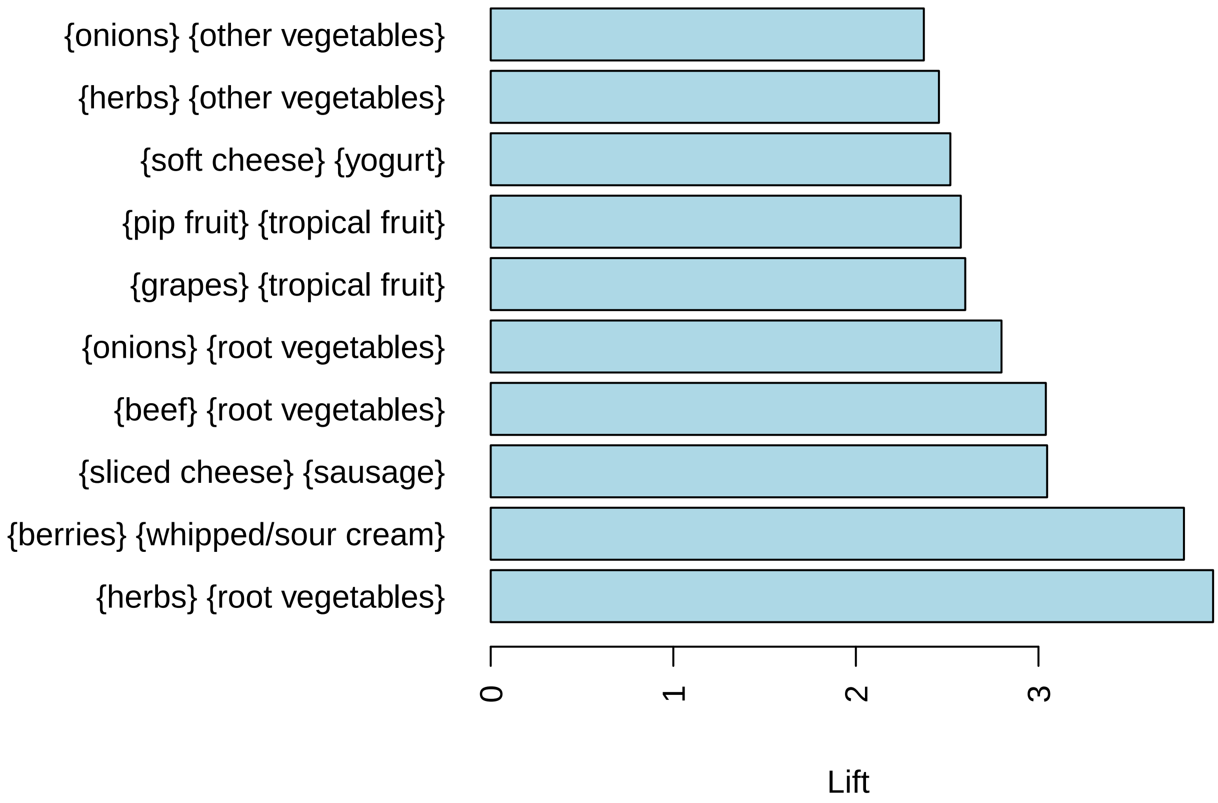
On the test set the model achieved an accuracy of 57% which is a significant improvement of 2% accuracy of blind guessing.

```
## Confusion Matrix and Statistics
```

```
##
##           Reference
## Prediction  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
##          1 19  0  0  1  0  1  1  1  5  0  0  0  0  0  7  0  1  5  0  2  0  0  6
##          2  0 15  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0
##          3  0  0 35  1  0  0  0  0  0  1  1  0  0  0  2  0  0  1  0  0  0  8  0
##          4  0  0  2 35  0  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  3  0  0
##          5  0  0  0  2 13  0  0  2  0  0  3  0  0  0  0  0  0  1  3  1  1  0  0
##          6  0  0  0  0  0 12  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##          7  0  0  0  0  0  1 22  0  0  0  0  0  0  0  1  0  4  1  0  0  0  0  1
##          8  0  0  0  0  0  0  0 25  0  0  0  0  0  0  0  0  0  0  0  1  1  0  0
##          9  0  1  0  1  0  0  0  0 26  1  0  0  0  0  0  0  0  0  0  0  0  0  0
##         10  0  3  0  0  0  0  0  0  0 27  0  0  0  0  0  2  0  0  0  2  0  0  0
##         11  0  0  0  0 35  0  0  0  0  0 16  0  0  0  0  0  0  0  0  0  0  0  0
##         12  0  1  0  0  0  0  0  0  2  1  3  7  2  0  0  2  0  0  0  0  0  0  0
##         13  0  0  0  0  0  0  0  0  0  0  0  0 27  0  0 14  0  0  2  0  0  0  0
##         14  0  1  0  0  0  0  0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0
##         15 19  0  4  0  0  0  0  0  9  0  0  0  0  0 27  0  0  0  0  2  0  2  0
##         16  0  0  0  0  0  0  0  0  0  0  0  0 10  0  0 24  0  0  0  0  0  0  0
##         17  1  0  0  0  0  0  7  0  0  0  0  1  0  0  0  1 42  1  0  1  3  0  0
##         18  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1 34  0  6  0  1  2
##         19  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 44  0  0  0  0
##         20  0  0  0  1  0  0  3  0  0  0  0  0  0  0  0  0  0  0  0 28  0  0 10
##         21  0  1  2  0  0  1  0  3  0  0  0  0  0  0  0  1  1  0  0  0 29  0  0
##         22  0  0  4  0  2  0  0  1  0  0  2  0  0  0  0  0  0  0  0  0  0 33  0
##         23  1  0  0  0  0  0  8  0  0  0  2  1  0  0  0  0  0  0  0  2  0  0 24
##         24  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         25  0  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0  0  0  0
##         26  0  0  0  0  0  0  0  0  0  0  2  0  3  0  0  0  0  0  0  0  0  0  0
##         27  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0
##         28  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
##         29  0  0  2  0  0  0  0  0  0  1  0  0  0  0  2  0  0  0  0  0  0  2  0
##         30  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1
##         31  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         32  1  0  0  1  0  1  0  1  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
##         33  0 10  0  0  0  0  0  0  0  2  1  7  0  0  0  0  0  1  0  0  0  0  0
##         34  0  0  0  0  0  0  0  7  0  0  2  0  0  0  0  0  0  0  0  0  0  0  1
##         35  0  0  0  1  0  4  0  1  0  0  0  0  0  0  0  0  0  0  0  0  9  0  0
##         36  0  2  0  0  0  0  0  0  0  0  3  4  0  0  0  1  0  0  1  0  0  0  0
##         37  0  0  0  0  0  1  2  0  0  0  0  0  0  1  0  0  1  0  0  1  0  0  2
##         38  0  0  0  4  0  1  0  3  1  0  3  0  0  0  0  0  0  0  0  0  4  0  0
##         39  0  0  0  0  0  0  1  2  0  0  0  0  0  0  1  0  0  0  0  0  0  1  0
##         40  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         41  2  0  0  0  0  0  0  0  0  2  1  3  0  0  0  1  0  0  0  0  0  1  1
##         42  0  5  0  0  0  0  0  0  4  2  2  2  0  0  0  0  0  0  0  0  0  0  0
##         43  5  0  0  0  0  0  0  0  0  0  1  0  0  0 10  0  0  4  0  0  0  0  0
##         44  0  4  0  0  0  0  0  0  0  1  2  1  0  0  0  0  0  0  0  0  0  1  0
##         45  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         46  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  4  0  1  1
##         47  0  0  1  1  0 25  5  0  3  0  0  1  1  0  0  1  0  0  0  0  0  0  0
##         48  1  7  0  0  0  0  1  1  0 12  0 14  0  0  0  0  0  0  0  0  0  0  0
##         49  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  1
##         50  0  0  0  0  0  0  0  0  0  0  1  0  7  0  0  2  0  0  0  0  0  0  0
##           Reference
## Prediction 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
##          1  0  0  1  0  4  0  0  0  0  0  1  0  0  0  0  1  0  0  0  5  1  0  1
##          2  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  3  0  0  0  0
##          3  0  0  0  0  0  4  0  0  0  1  0  0  0  0  0  0  0  3  0  0  0  0  0
##          4  0  2  0  0  0  0  0  0  0  0  2  8  0  0  0  1  0  0  0  0  0  0  0
##          5  0  2  0  0  0  1  3  0  1  0  0  0  2  0  0  1  0  0  0  0  0  0  1
##          6  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##          7  0  0  0  0  3  1  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  1
##          8  3  2  1  0  0  0  1  0  3  0  2  2  0  0  0  0  4  0  0  0  1  1  1
##          9  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  3  0  0  0
##         10  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  1  0  0  0  2
##         11  0  0  0  0  0  1  0  0  0  7  0  0  0  0  0  0  0  0  2  0  0  0  0
##         12  0  0  0  0  0  0  0  0  0  7  0  0  0  0  0  1  0  5  3  0  4  0  0
##         13  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0
##         14  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         15  0  0  0  0  2  0  0  0  0  0  1  0  0  0  0  0  0  0  0  3  0  0  0
##         16  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0
##         17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  3  0  0  0
##         18  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  1
##         19  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0
##         20  0  0  0  0  0  0  0  0  2  0  0  0  0  1  0  0  0  0  0  0  0  0  1
##         21  1  0  0  0  0  0  1  0  0  0  3 10  0  0  2  0  0  0  0  0  0  0  0
##         22  0  0  1  0  0  9  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
##         23  0  0  0  0  5  3  0  0  0  0  2  0  0  0  1  2  0  0  0  0  2  1  1
##         24 37  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  9  0  0  0  0  1  0
##         25  0 37  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         26  0  0 29  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         27  0  0  0 41  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         28  0  0  0  1 13  0  0  0  0  0  0  0  0 10  0  0  0  0  0  1  0  0  0
##         29  0  0  2  0  0 26  1  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
##         30  1  1  0  0  2  0 38  0  0  0  0  0  0  0  0  2  1  0  0  0  1  0  0
##         31  0  0  0  0  0  0  0 45  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         32  0  1  0  0  0  1  0  0 42  0  3  4  0  0  1  0  0  0  0  0  0  4  2
##         33  0  0  0  1  0  0  0  2  0 16  0  0  2  0  0  0  0  0 11  0  2  0  0
##         34  1  0  0  0  0  0  0  0  0  0 34  0  0  0  0  0  0  0  0  0  0  1  0
##         35  0  2  0  0  0  0  1  0  0  0  1 22  0  0  0  0  0  0  0  0  0  0  0
##         36  0  0  0  0  0  0  0  0  0  0  0  0 24  0  0  0  0  7  2  0  9  0  0
##         37  0  0  0  1 21  0  0  0  0  0  0  0  0 31  0  1  0  0  0  0  1  0  0
##         38  0  2  0  0  0  0  0  0  2  0  0  0  0  1 46  1  0  1  1  0  0  2  0
##         39  0  0  0  1  0  0  0  0  0  0  1  0  0  0  0 31  0  0  0  0  0  0  0
##         40  2  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0 29  0  0  0  0  0  0
##         41  0  0  1  0  0  3  0  0  0  4  0  0  1  0  0  0  0 17  1  2  2  0  0
##         42  0  0  3  0  0  0  0  0  0  7  0  0  0  0  0  0  0  0 10  1  0  0  0
##         43  0  0  0  0  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0 28  0  0  3
##         44  0  0  7  1  0  0  0  1  0  3  0  0 19  0  0  0  0 13  8  0 26  0  0
##         45  5  0  1  0  0  0  1  0  0  0  0  0  0  0  0  0  7  0  0  0  0 38  0
##         46  0  0  0  0  0  0  0  0  0  0  0  1  0  1  0  0  0  0  0  0  0  0 36
##         47  0  1  0  0  0  0  1  0  0  0  0  0  1  1  0  2  0  0  1  0  0  2  0
##         48  0  0  3  1  0  1  0  2  0  4  0  0  1  0  0  0  0  3  6  1  1  0  0
##         49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  4  0  0  0  3  0  0  0
##         50  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##           Reference
## Prediction 47 48 49 50
##          1  0  0  2  0
##          2    0  1  0  0
##          3    0  0  0  0
##          4    0  0  0  0
##          5    2  1  0  0
##          6    9  0  0  0
##          7    0  0  0  0
##          8    0  0  0  1
##          9    0  0  0  0
##          10   3  2  0  0
##          11   0  0  0  0
##          12   0  3  0  0
##          13   0  0  0 12
##          14   3  0  0  0
##          15   0  0  0  0
##          16   0  0  0  8
##          17   0  0  0  0
##          18   0  0  0  0
##          19   0  0  0  0
##          20   0  0  0  0
##          21   0  0  0  0
##          22   0  0  0  0
##          23   0  0  0  0
##          24   0  0  0  0
##          25   0  0  0  0
##          26   0  0  0  0
##          27   0  0  0  0
##          28   0  2  0  0
##          29   0  0  0  0
##          30   0  0  2  0
##          31   0  0  0  3
##          32   3  0  1  0
##          33   0  8  0  0
##          34   0  0  0  0
##          35   0  0  0  0
##          36   0  0  0  0
##          37   2  0  0  3
##          38   0  0  0  0
##          39   0  0  0  0
##          40   0  0  0  0
##          41   0  0  0  1
##          42   0  3  0  0
##          43   0  0  0  0
##          44   0  7  0  0
##          45   0  0  0  0
##          46   0  0  0  0
##          47  28  0  1  0
##          48   0 23  0  0
##          49   0  0 44  0
##          50   0  0  0 22
##
## Overall Statistics
##
##                Accuracy : 0.5704
##                  95% CI : (0.5507, 0.5899)
##     No Information Rate : 0.02
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5616
##
##  Mcnemar’s Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
## Sensitivity            0.3800   0.3000   0.7000   0.7000   0.2600   0.2400
## Specificity            0.9812   0.9947   0.9910   0.9914   0.9890   0.9963
## Pos Pred Value         0.2923   0.5357   0.6140   0.6250   0.3250   0.5714
## Neg Pred Value         0.9873   0.9858   0.9939   0.9939   0.9850   0.9847
## Precision              0.2923   0.5357   0.6140   0.6250   0.3250   0.5714
## Recall                 0.3800   0.3000   0.7000   0.7000   0.2600   0.2400
## F1                     0.3304   0.3846   0.6542   0.6604   0.2889   0.3380
## Prevalence             0.0200   0.0200   0.0200   0.0200   0.0200   0.0200
## Detection Rate         0.0076   0.0060   0.0140   0.0140   0.0052   0.0048
## Detection Prevalence   0.0260   0.0112   0.0228   0.0224   0.0160   0.0084
## Balanced Accuracy      0.6806   0.6473   0.8455   0.8457   0.6245   0.6182
##                      Class: 7 Class: 8 Class: 9 Class: 10 Class: 11 Class: 12
## Sensitivity            0.4400   0.5000   0.5200    0.5400    0.3200    0.1400
## Specificity            0.9943   0.9902   0.9971    0.9931    0.9816    0.9861
## Pos Pred Value         0.6111   0.5102   0.7879    0.6136    0.2623    0.1707
## Neg Pred Value         0.9886   0.9898   0.9903    0.9906    0.9861    0.9825
## Precision              0.6111   0.5102   0.7879    0.6136    0.2623    0.1707
## Recall                 0.4400   0.5000   0.5200    0.5400    0.3200    0.1400
## F1                     0.5116   0.5051   0.6265    0.5745    0.2883    0.1538
## Prevalence             0.0200   0.0200   0.0200    0.0200    0.0200    0.0200
## Detection Rate         0.0088   0.0100   0.0104    0.0108    0.0064    0.0028
## Detection Prevalence   0.0144   0.0196   0.0132    0.0176    0.0244    0.0164
## Balanced Accuracy      0.7171   0.7451   0.7586    0.7665    0.6508    0.5631
##                      Class: 13 Class: 14 Class: 15 Class: 16 Class: 17
## Sensitivity             0.5400    0.9800    0.5400    0.4800    0.8400
## Specificity             0.9878    0.9984    0.9829    0.9914    0.9927
## Pos Pred Value          0.4737    0.9245    0.3913    0.5333    0.7000
## Neg Pred Value          0.9906    0.9996    0.9905    0.9894    0.9967
## Precision               0.4737    0.9245    0.3913    0.5333    0.7000
## Recall                  0.5400    0.9800    0.5400    0.4800    0.8400
## F1                      0.5047    0.9515    0.4538    0.5053    0.7636
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0108    0.0196    0.0108    0.0096    0.0168
## Detection Prevalence    0.0228    0.0212    0.0276    0.0180    0.0240
## Balanced Accuracy       0.7639    0.9892    0.7614    0.7357    0.9163
##                      Class: 18 Class: 19 Class: 20 Class: 21 Class: 22
## Sensitivity             0.6800    0.8800    0.5600    0.5800    0.6600
## Specificity             0.9951    0.9996    0.9927    0.9894    0.9918
## Pos Pred Value          0.7391    0.9778    0.6087    0.5273    0.6226
## Neg Pred Value          0.9935    0.9976    0.9910    0.9914    0.9931
## Precision               0.7391    0.9778    0.6087    0.5273    0.6226
## Recall                  0.6800    0.8800    0.5600    0.5800    0.6600
## F1                      0.7083    0.9263    0.5833    0.5524    0.6408
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0136    0.0176    0.0112    0.0116    0.0132
## Detection Prevalence    0.0184    0.0180    0.0184    0.0220    0.0212
## Balanced Accuracy       0.8376    0.9398    0.7763    0.7847    0.8259
##                      Class: 23 Class: 24 Class: 25 Class: 26 Class: 27
## Sensitivity             0.4800    0.7400    0.7400    0.5800    0.8200
## Specificity             0.9873    0.9959    0.9988    0.9967    0.9996
## Pos Pred Value          0.4364    0.7872    0.9250    0.7838    0.9762
## Neg Pred Value          0.9894    0.9947    0.9947    0.9915    0.9963
## Precision               0.4364    0.7872    0.9250    0.7838    0.9762
## Recall                  0.4800    0.7400    0.7400    0.5800    0.8200
## F1                      0.4571    0.7629    0.8222    0.6667    0.8913
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0096    0.0148    0.0148    0.0116    0.0164
## Detection Prevalence    0.0220    0.0188    0.0160    0.0148    0.0168
## Balanced Accuracy       0.7337    0.8680    0.8694    0.7884    0.9098
##                      Class: 28 Class: 29 Class: 30 Class: 31 Class: 32
## Sensitivity             0.2600    0.5200    0.7600    0.9000    0.8400
## Specificity             0.9939    0.9955    0.9947    0.9988    0.9898
## Pos Pred Value          0.4643    0.7027    0.7451    0.9375    0.6269
## Neg Pred Value          0.9850    0.9903    0.9951    0.9980    0.9967
## Precision               0.4643    0.7027    0.7451    0.9375    0.6269
## Recall                  0.2600    0.5200    0.7600    0.9000    0.8400
## F1                      0.3333    0.5977    0.7525    0.9184    0.7179
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0052    0.0104    0.0152    0.0180    0.0168
## Detection Prevalence    0.0112    0.0148    0.0204    0.0192    0.0268
## Balanced Accuracy       0.6269    0.7578    0.8773    0.9494    0.9149
##                      Class: 33 Class: 34 Class: 35 Class: 36 Class: 37
## Sensitivity             0.3200    0.6800    0.4400    0.4800    0.6200
## Specificity             0.9808    0.9951    0.9922    0.9882    0.9849
## Pos Pred Value          0.2540    0.7391    0.5366    0.4528    0.4559
## Neg Pred Value          0.9860    0.9935    0.9886    0.9894    0.9922
## Precision               0.2540    0.7391    0.5366    0.4528    0.4559
## Recall                  0.3200    0.6800    0.4400    0.4800    0.6200
## F1                      0.2832    0.7083    0.4835    0.4660    0.5254
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0064    0.0136    0.0088    0.0096    0.0124
## Detection Prevalence    0.0252    0.0184    0.0164    0.0212    0.0272
## Balanced Accuracy       0.6504    0.8376    0.7161    0.7341    0.8024
##                      Class: 38 Class: 39 Class: 40 Class: 41 Class: 42
## Sensitivity             0.9200    0.6200    0.5800    0.3400    0.2000
## Specificity             0.9894    0.9971    0.9976    0.9894    0.9882
## Pos Pred Value          0.6389    0.8158    0.8286    0.3953    0.2564
## Neg Pred Value          0.9984    0.9923    0.9915    0.9866    0.9837
## Precision               0.6389    0.8158    0.8286    0.3953    0.2564
## Recall                  0.9200    0.6200    0.5800    0.3400    0.2000
## F1                      0.7541    0.7045    0.6824    0.3656    0.2247
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0184    0.0124    0.0116    0.0068    0.0040
## Detection Prevalence    0.0288    0.0152    0.0140    0.0172    0.0156
## Balanced Accuracy       0.9547    0.8086    0.7888    0.6647    0.5941
##                      Class: 43 Class: 44 Class: 45 Class: 46 Class: 47
## Sensitivity             0.5600    0.5200    0.7600    0.7200    0.5600
## Specificity             0.9898    0.9722    0.9939    0.9955    0.9804
## Pos Pred Value          0.5283    0.2766    0.7170    0.7660    0.3684
## Neg Pred Value          0.9910    0.9900    0.9951    0.9943    0.9909
## Precision               0.5283    0.2766    0.7170    0.7660    0.3684
## Recall                  0.5600    0.5200    0.7600    0.7200    0.5600
## F1                      0.5437    0.3611    0.7379    0.7423    0.4444
## Prevalence              0.0200    0.0200    0.0200    0.0200    0.0200
## Detection Rate          0.0112    0.0104    0.0152    0.0144    0.0112
## Detection Prevalence    0.0212    0.0376    0.0212    0.0188    0.0304
## Balanced Accuracy       0.7749    0.7461    0.8769    0.8578    0.7702
##                      Class: 48 Class: 49 Class: 50
## Sensitivity             0.4600    0.8800    0.4400
## Specificity             0.9759    0.9963    0.9959
## Pos Pred Value          0.2805    0.8302    0.6875
## Neg Pred Value          0.9888    0.9975    0.9887
## Precision               0.2805    0.8302    0.6875
## Recall                  0.4600    0.8800    0.4400
## F1                      0.3485    0.8544    0.5366
## Prevalence              0.0200    0.0200    0.0200
## Detection Rate          0.0092    0.0176    0.0088
## Detection Prevalence    0.0328    0.0212    0.0128
## Balanced Accuracy       0.7180    0.9382    0.7180
```

Association Rules

For this problem I decided to set a minimum support of .005 to encourage even rare pairs and a confidence of .2 to find pairs that really were exclusive to each other. I found that people like to buy similar products to each other that aren't necessarily compliments like onions and other vegetables or grapes and tropical fruits. I was surprised that the second highest lift belonged to berries and whipped/sour cream which does not seem like a good combination to me, but maybe its a regional thing.


{onions} {other vegetables}
{herbs} {other vegetables}
{soft cheese} {yogurt}
{pip fruit} {tropical fruit}
{grapes} {tropical fruit}
{onions} {root vegetables}
{beef} {root vegetables}
{sliced cheese} {sausage}
{berries} {whipped/sour cream}
{herbs} {root vegetables}
0
1
2
3
Lift