Factor Analysis of Retailers

```
##Install packages
options(repos = c(CRAN = "https://cran.r-project.org"))

install.packages('nFactors')
```

```
The downloaded binary packages are in
    /var/folders/_f/jxf6gqq91bg2n4kxz46lwv200000gn/T//RtmpgRumxh/downloaded_packages
```

```
#install.packages('dplyr')
install.packages('GPArotation')
```

```
The downloaded binary packages are in
    /var/folders/_f/jxf6gqq91bg2n4kxz46lwv200000gn/T//RtmpgRumxh/downloaded_packages
```

```
install.packages('gplots')
```

```
The downloaded binary packages are in
    /var/folders/_f/jxf6gqq91bg2n4kxz46lwv200000gn/T//RtmpgRumxh/downloaded_packages
```

```
install.packages('RColorBrewer')
```

```
The downloaded binary packages are in
    /var/folders/_f/jxf6gqq91bg2n4kxz46lwv200000gn/T//RtmpgRumxh/downloaded_packages
```

```r
## load packages and set seed
library(nFactors)
```

```
Loading required package: lattice
```

```
Attaching package: 'nFactors'
```

```
The following object is masked from 'package:lattice':

    parallel
```

```r
library(dplyr)
```

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

```
The following objects are masked from 'package:stats':

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
library(GPArotation)
library(gplots)
```

```
Attaching package: 'gplots'
```

```
The following object is masked from 'package:stats':

    lowess
```

```r
library(RColorBrewer)
set.seed(1)
```

```r
## Read in Factor Analysis Data
retailer_survey <- read.table(file = "../data/retailersurvey.csv", header = TRUE, sep = ",
```

```r
head(retailer_survey)
```

```
  Quality Last Fit Latest Trends Stylish Value Bargain Worth Satisfied Purchase
1       2    1   1      7      7       5     3       3     4         6        7
2       7    7   7      3      4       3     1       1     1         2        1
3       3    4   2      3      3       5     6       7     6         2        2
4       4    5   3      7      7       7     1       2     2         7        7
5       4    4   4      3      3       3     1       1     1         6        6
6       3    4   3      7      5       6     1       1     1         1        3
  Recommend Retailer
1         7       CR
2         3       CR
3         1       CR
4         6       CR
5         6       CR
6         2       CR
```

```r
dim(retailer_survey)
```

```
[1] 600  13
```

```r
table(retailer_survey$'Retailer')
```

```
  A   B   C  CR   D   E
100 100 100 100 100 100
```

```r
retailer.factor.mean <- aggregate(.~Retailer, data=retailer_survey, mean)
retailer.factor.mean
```

```
  Retailer Quality Last  Fit Latest Trends Stylish Value Bargain Worth
1        A    5.07 5.02 4.83   3.31   3.47    3.50  2.56    2.75  2.57
2        B    3.06 3.18 3.15   5.79   5.57    5.61  1.97    2.29  2.17
3        C    4.61 4.57 4.48   2.65   2.81    2.78  5.52    5.36  5.29
4       CR    5.03 4.88 4.85   4.63   4.63    4.53  3.00    3.11  3.12
5        D    4.17 4.10 4.22   2.85   3.23    3.11  4.00    4.10  3.98
6        E    2.79 2.82 2.96   4.03   4.14    3.89  2.79    5.47  2.91
  Satisfied Purchase Recommend
1      4.74     4.62      4.60
2      3.46     3.55      3.58
3      3.84     3.84      3.80
4      4.60     4.52      4.50
5      3.52     3.52      3.54
6      5.41     5.22      5.28
```

```r
## Determine the Number of Factors
# chekcing the eigenvalues of the factors
retailer_factors <- select(retailer_survey, -Retailer) # this removes the retailer column
eigen(cor(retailer_factors))$values
```

```
 [1] 3.27396254 2.90493678 2.66726790 2.04691003 0.29372650 0.17848047
 [7] 0.17199170 0.15689070 0.08864872 0.07600663 0.07177044 0.06940759
```

There are 4 factor values over 1, so I want to do a 4-factor solution.

```r
## Run 4 Factor Analysis
factanal(retailer_factors, factors = 4)
```

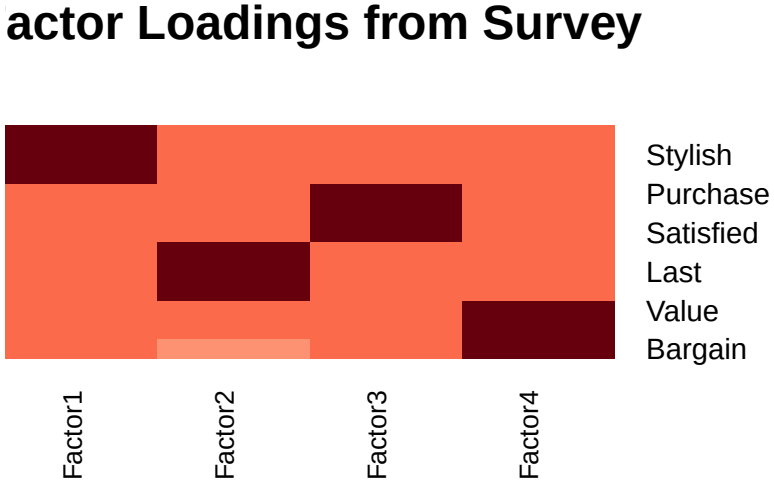
```
Call:
factanal(x = retailer_factors, factors = 4)

Uniquenesses:
  Quality      Last       Fit    Latest    Trends   Stylish     Value   Bargain
    0.014     0.183     0.178     0.021     0.157     0.163     0.014     0.354
    Worth Satisfied  Purchase Recommend
    0.170     0.024     0.178     0.166

Loadings:
          Factor1 Factor2 Factor3 Factor4
```

```
Quality            0.992
Last               0.904
Fit                0.906
Latest     0.983                -0.109
Trends     0.913
Stylish    0.908                -0.105
Value     -0.110                 0.981
Bargain                          0.793
Worth                            0.903
Satisfied                 0.987
Purchase                  0.905
Recommend                 0.911

               Factor1 Factor2 Factor3 Factor4
SS loadings      2.656   2.641   2.638   2.442
Proportion Var   0.221   0.220   0.220   0.203
Cumulative Var   0.221   0.441   0.661   0.865

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 22.86 on 24 degrees of freedom.
The p-value is 0.528
```

The 12 factors seem to load well on the 4 factors and are sufficient for the data. The positive loading are used with the Factors. Since there are partial loading, I need to see if items load better on single factors using rotation.

```
## 4 Factor Analysis and Oblique Rotation
retailer.fa <- factanal(retailer_factors, factors = 4, rotation = 'oblimin', scores = 'Bar
#oblimin = oblique
#bartlette is saving the previous scores
```

```
## Heatmap of Factor Loadings
# analyze using a heatmap of factor scores
heatmap.2(retailer.fa$loadings, col = brewer.pal(9, 'Reds'), trace = 'none', key = FALSE,
```



The heat map shows that the Oblique rotation strengthened the factor loadings and reduced cross loading. I need to name the factors by identifying statements that load highly on each factor. (name is subjective)

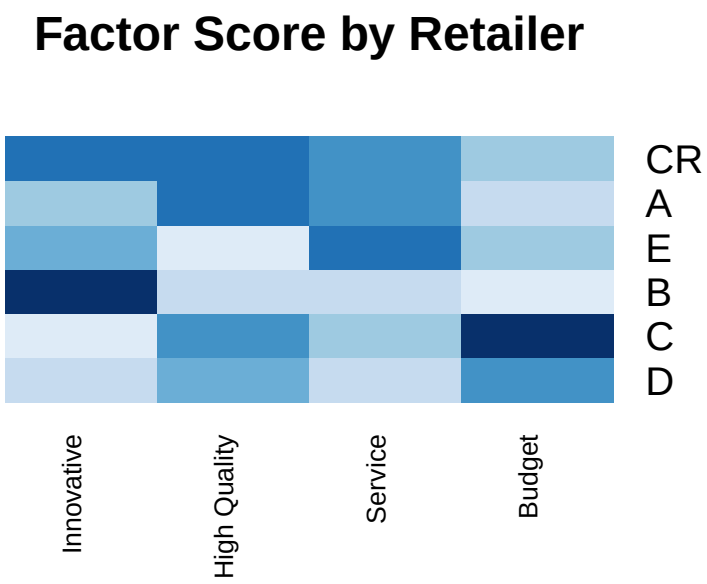
Factor 1: Innovative; Factor 2: High Quality; Factor 3: Service; Factor 4: Budget

```
## Aggregate Factor Scores by Retailer
retailer.scores <- data.frame(retailer.fa$scores)
retailer.scores$retailer <- retailer_survey$Retailer
retailer.fa.mean <- aggregate(. ~ retailer, data = retailer.scores, mean)
rownames(retailer.fa.mean) <- retailer.fa.mean[,1]
retailer.fa.mean <- select(retailer.fa.mean, -retailer)
names(retailer.fa.mean) <- c('Innovative', 'High Quality', 'Service', 'Budget')
retailer.fa.mean
```

```
   Innovative High Quality    Service     Budget
A  -0.25900898   0.45572219  0.2338283 -0.3774511
B   0.89220177  -0.50775233 -0.3781924 -0.6370296
C  -0.57843428   0.23954525 -0.2206203  1.0397720
CR  0.35620993   0.43960427  0.1641804 -0.1596213
D  -0.46779941   0.02693213 -0.3674417  0.3154521
E   0.05683097  -0.65405152  0.5682457 -0.1811221
```

These are the average factor scores for each retailer (Scores are standardized with mean = 0 and std dev = 1).If > 0 = higher than average, < 0 lower than average

```
## Heatmap of Retailer Scores
heatmap.2(as.matrix(retailer.fa.mean), col=brewer.pal(9, "Blues"), trace='none', key=FALSE
```



(Darker Blue means higher factor scores) CR rates high in consumer perceptions with Innovation and High Quality. B is most Innovative. A and C seem to have high quality products. E ranks highest in Service followed by CR and A. C is ranked highest on Budget.