SFO Survey

Tao Tang (University of Notre Dame)

Part A

Address the Primary Questions of Interest for the SFO Executives

The SFO team have three (3) specific questions they want you to investigate.

- 1.Customers were asked to rate their opinion of the "SFO Airport as a whole" on a scale from 1 ("unacceptable") to 5 ("outstanding"). The executives want to know if there are patterns across the satisfied or dissatisfied customers based on demographic characteristics, such as sex, age group, and income level.
- 2.The executives also want to know if customer satisfaction can be broken down into different attributes of the airport. Knowing this will help the team target specific strengths or areas of improvement. The central feature the customer satisfaction survey is a 14-question portion of the survey asking customers to rate satisfaction with different aspects of the airport (see Question 6 in the data directory). The executives want you to perform a quantitative analysis to determine if there are broad themes that emerge from this part of the survey.
- 3.Free-response comments, either positive or negative, were collected in addition to the 14-item quantitative survey. The executives are not quite sure how to examine it without going through individual surveys one by one, but they want you to see if there are any concepts or insights that arise from these responses. Do the free responses relate to the findings in a) or b) at all?

Part A - 1

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(poLCA)
```

```
## Loading required package: scatterplot3d
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(ggcorrplot)
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```r
library(tidytext)
library(wordcloud2)
```

```r
sfo <- read.table('SFO_survey_withText.txt',stringsAsFactors = T, header=T)
```

```r
#Replace 6 with NA in Q6
cols <- c("Q6A", "Q6B", "Q6C", "Q6D", "Q6E", "Q6F","Q6G","Q6H","Q6I","Q6J","Q6K", "Q6L", "Q6M", "Q6N")
sfo[,cols][sfo[,cols]==6] <- NA

#Convert columns to factors
#sfo[,cols] <- lapply(sfo[,cols], function(x) as.factor(x))
```

```r
lca <- cbind(Q17, Q18, Q19,Q6N)~1

#2 classes LCA
set.seed(1989)
class2 <- poLCA(lca,
                sfo,
                nclass=2,
                maxiter=10000,
                tol = 1e-8,
                nrep = 1,
                verbose = F)

class2
```

```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $Q17
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)  Pr(6)  Pr(7)
## class 1:  0.0021 0.0174 0.1734 0.2660 0.2648 0.1921 0.0843
## class 2:  0.0068 0.1955 0.3717 0.0895 0.0959 0.1210 0.1195
##
```

```
## $Q18
##           Pr(1)  Pr(2)
## class 1:  0.5752 0.4248
## class 2:  0.5080 0.4920
##
## $Q19
##           Pr(1)  Pr(2)  Pr(3)  Pr(4) Pr(5)
## class 1:  0.000 0.2965 0.2942 0.4092 0.000
## class 2:  0.548 0.3787 0.0723 0.0000 0.001
##
## $Q6N
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
## class 1:  0.0017 0.0163 0.2586 0.5787 0.1447
## class 2:  0.0016 0.0111 0.1634 0.5628 0.2611
##
## Estimated class population shares
##  0.569 0.431
##
## Predicted class memberships (by modal posterior prob.)
##  0.5826 0.4174
##
## ============================================================
## Fit for 2 latent classes:
## ============================================================
## number of observations: 2422
## number of estimated parameters: 31
## residual degrees of freedom: 318
## maximum log-likelihood: -11559.29
##
## AIC(2): 23180.58
## BIC(2): 23360.14
## G^2(2): 365.2374 (Likelihood ratio/deviance statistic)
## X^2(2): 482.9958 (Chi-square goodness of fit)
##
```
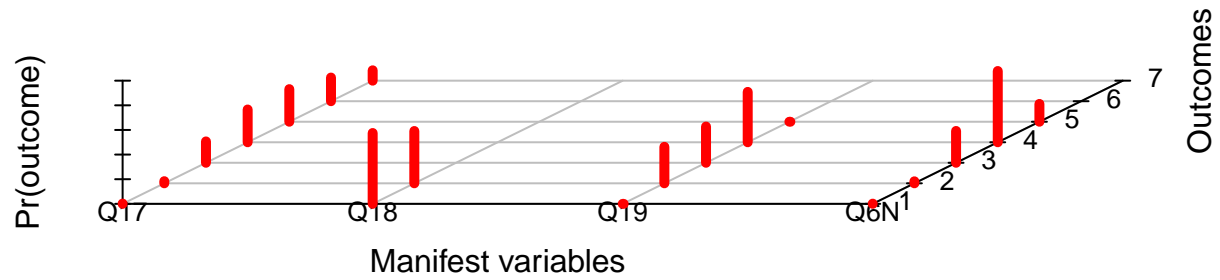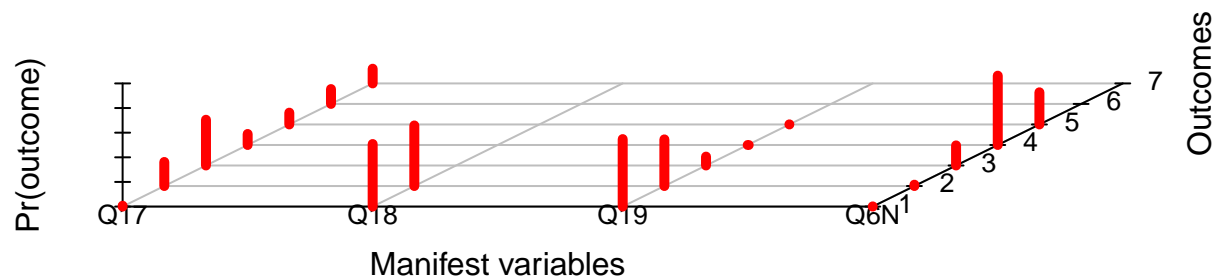
```r
plot(class2)
```

Class 1: population share = 0.569

Class 2: population share = 0.431


```
#3 classes LCA
set.seed(2000)
class3 <- poLCA(lca,
                sfo,
                nclass=3,
                maxiter=10000,
                tol = 1e-8,
                nrep = 1,
                verbose = F)

class3
```

```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $Q17
##          Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)  Pr(6)  Pr(7)
## class 1:  0.0000 0.0000 0.5996 0.1691 0.0653 0.0929 0.0731
## class 2:  0.0168 0.4840 0.0003 0.0000 0.1442 0.1718 0.1830
## class 3:  0.0024 0.0218 0.1406 0.2650 0.2834 0.1990 0.0877
##
## $Q18
##          Pr(1)  Pr(2)
## class 1:  0.5878 0.4122
## class 2:  0.3982 0.6018
## class 3:  0.5702 0.4298
```

```
## 
## $Q19
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
## class 1:  0.3981 0.5204 0.0815 0.0000 0.0000
## class 2:  0.6593 0.2660 0.0723 0.0000 0.0024
## class 3:  0.0000 0.2411 0.3102 0.4487 0.0000
## 
## $Q6N
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
## class 1:  0.0019 0.0155 0.2208 0.5566 0.2051
## class 2:  0.0010 0.0026 0.1014 0.5711 0.3240
## class 3:  0.0017 0.0169 0.2540 0.5812 0.1462
## 
## Estimated class population shares
##  0.3099 0.1711 0.519
## 
## Predicted class memberships (by modal posterior prob.)
##  0.2395 0.1713 0.5892
## 
## =============================================================
## Fit for 3 latent classes:
## =============================================================
## number of observations: 2422
## number of estimated parameters: 47
## residual degrees of freedom: 302
## maximum log-likelihood: -11515.79
## 
## AIC(3): 23125.57
## BIC(3): 23397.81
## G^2(3): 278.2359 (Likelihood ratio/deviance statistic)
## X^2(3): 363.1632 (Chi-square goodness of fit)
## 
```

```
plot(class3)
```

Class 1: population share = 0.31

Class 2: population share = 0.171

Class 3: population share = 0.519
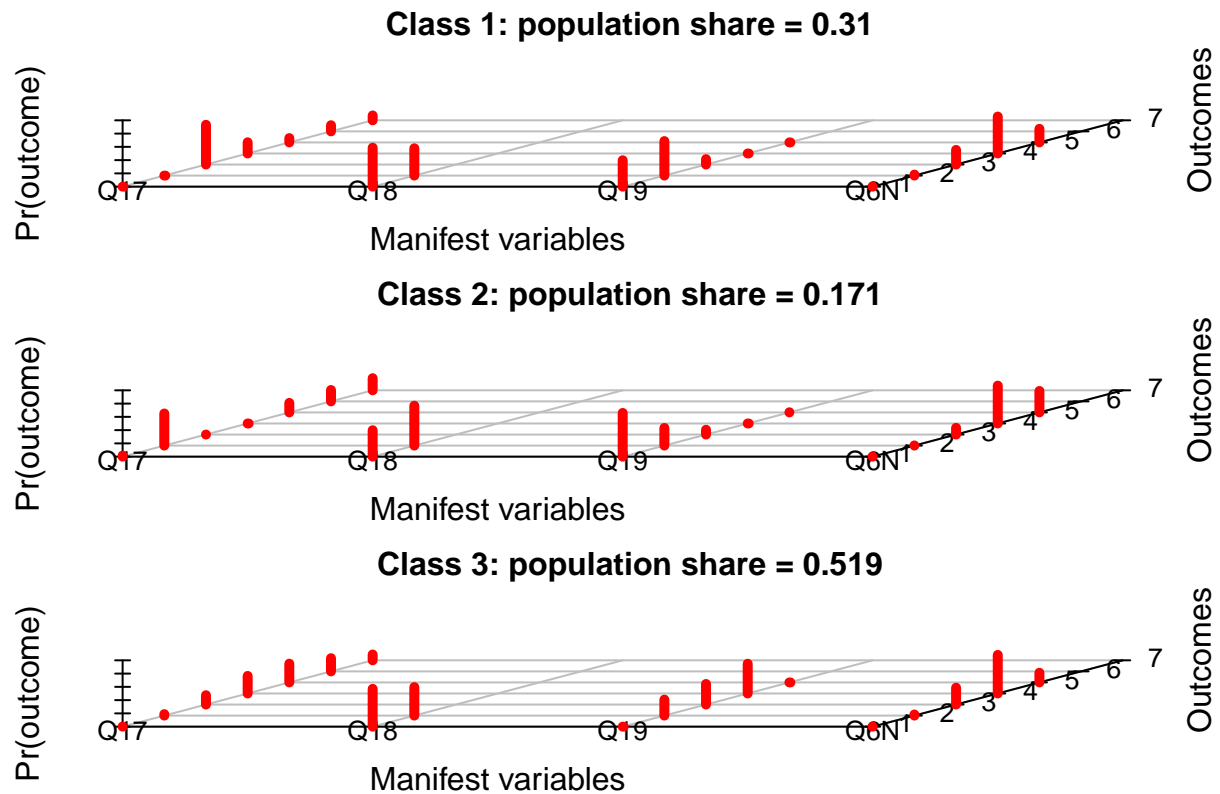

```r
#4 classes LCA
set.seed(3000)
class4 <- poLCA(lca,
                sfo,
                nclass=4,
                maxiter=10000,
                tol = 1e-8,
                nrep = 1,
                verbose = F)

class4
```
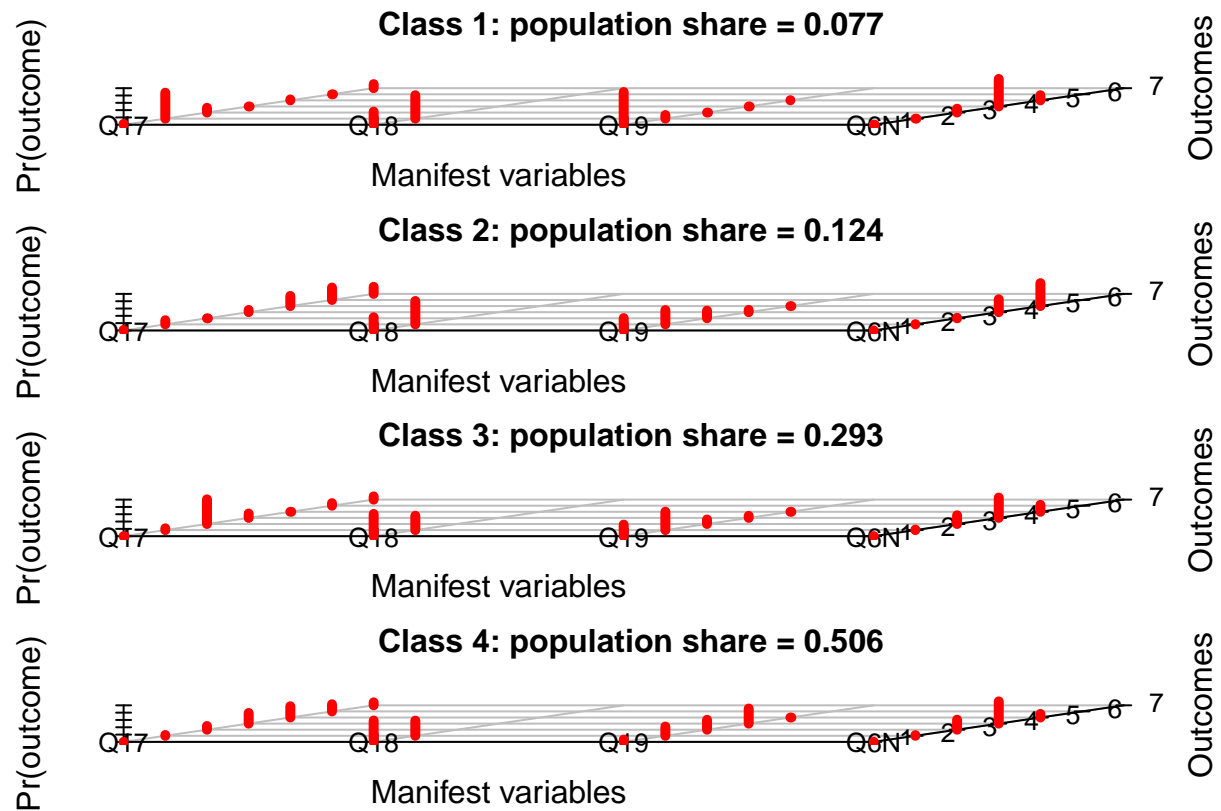
```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $Q17
##          Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)  Pr(6)  Pr(7)
## class 1: 0.0242 0.7099 0.1297 0.0000 0.0255 0.0000 0.1107
## class 2: 0.0144 0.1121 0.0000 0.0618 0.2774 0.3399 0.1944
## class 3: 0.0000 0.0478 0.6742 0.1193 0.0000 0.0704 0.0882
## class 4: 0.0009 0.0224 0.1019 0.2914 0.3076 0.1948 0.0809
##
## $Q18
##          Pr(1)  Pr(2)
## class 1: 0.3589 0.6411
## class 2: 0.3611 0.6389
```

```
## class 3:   0.6081 0.3919
## class 4:   0.5847 0.4153
##
## $Q19
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
## class 1:  0.8951 0.0996 0.0000 0.0000 0.0053
## class 2:  0.3390 0.3898 0.2090 0.0622 0.0000
## class 3:  0.3162 0.5039 0.1227 0.0572 0.0000
## class 4:  0.0637 0.2537 0.2703 0.4122 0.0000
##
## $Q6N
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
## class 1:  0.0000 0.0000 0.1060 0.7528 0.1412
## class 2:  0.0004 0.0061 0.0130 0.3525 0.6280
## class 3:  0.0012 0.0143 0.2379 0.5555 0.1911
## class 4:  0.0025 0.0180 0.2733 0.6076 0.0986
##
## Estimated class population shares
##   0.0773 0.1245 0.2927 0.5055
##
## Predicted class memberships (by modal posterior prob.)
##   0.0751 0.1049 0.2915 0.5285
##
## ============================================================
## Fit for 4 latent classes:
## ============================================================
## number of observations: 2422
## number of estimated parameters: 63
## residual degrees of freedom: 286
## maximum log-likelihood: -11483.32
##
## AIC(4): 23092.64
## BIC(4): 23457.56
## G^2(4): 213.3067 (Likelihood ratio/deviance statistic)
## X^2(4): 268.8398 (Chi-square goodness of fit)
##
```
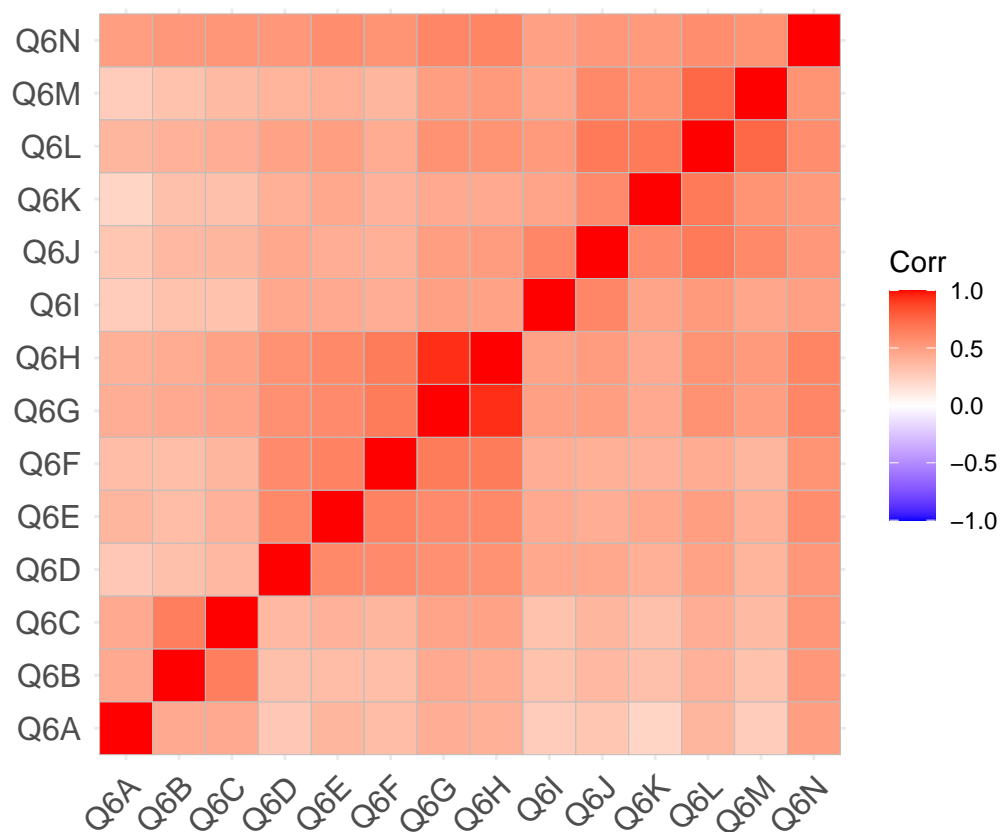
```r
plot(class4)
```

Class 1: population share = 0.077

Pr(outcome) — Manifest variables — Outcomes

Class 2: population share = 0.124

Pr(outcome) — Manifest variables — Outcomes

Class 3: population share = 0.293

Pr(outcome) — Manifest variables — Outcomes

Class 4: population share = 0.506

Pr(outcome) — Manifest variables — Outcomes
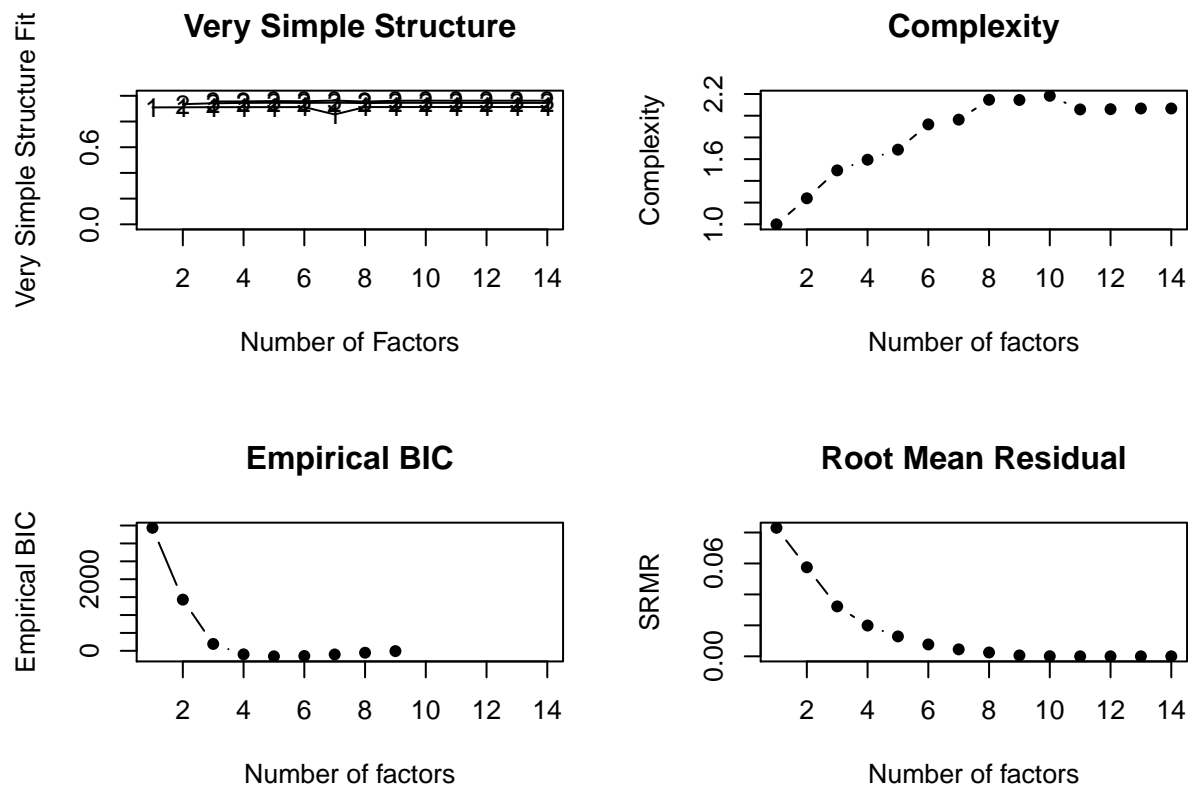
Part A-2

```
#select the columns starts with Q6
sfo_fa <- sfo %>% dplyr::select(cols)
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(cols)` instead of `cols` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```
#get the correlation plot
sfo_fa %>%
  cor(., use="pairwise.complete.obs") %>%
  ggcorrplot()
```

```
#nfactors
sfo_fa %>% nfactors(.,rotate = 'none')
```

	Very Simple Structure	Complexity

Very Simple Structure / **Complexity**

Empirical BIC / **Root Mean Residual**

```
## 
## Number of factors
## Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,
##     n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)
## VSS complexity 1 achieves a maximimum of Although the vss.max shows  10   factors, it is probably more
## VSS complexity 2 achieves a maximimum of 0.95  with  7  factors
## The Velicer MAP achieves a minimum of 0.04  with  3  factors
## Empirical BIC achieves a minimum of  -153.27  with  5  factors
## Sample Size adjusted BIC achieves a minimum of  -1.55  with  9  factors
## 
## Statistics by number of factors
##    vss1 vss2   map dof  chisq      prob sqresid  fit RMSEA    BIC  SABIC
## 1  0.91 0.00 0.040  77 8.5e+03  0.0e+00     5.2 0.91 0.184 7855.7 8100.3
## 2  0.91 0.93 0.041  64 5.3e+03  0.0e+00     3.7 0.93 0.160 4812.5 5015.9
## 3  0.91 0.94 0.039  52 3.4e+03  0.0e+00     2.6 0.96 0.141 2980.6 3145.8
## 4  0.91 0.94 0.052  41 8.5e+02 5.4e-152     2.2 0.96 0.078  519.8  650.1
## 5  0.91 0.95 0.063  31 4.5e+02  2.3e-76     1.9 0.97 0.065  202.1  300.6
## 6  0.91 0.94 0.070  22 2.8e+02  1.0e-46     1.5 0.97 0.060  102.8  172.7
## 7  0.85 0.95 0.101  14 1.1e+02  4.0e-16     1.4 0.98 0.045   -7.5   36.9
## 8  0.91 0.94 0.128   7 5.8e+01  4.5e-10     1.1 0.98 0.047    1.1   23.3
## 9  0.91 0.95 0.171   1 3.4e+00  6.7e-02     1.2 0.98 0.027   -4.7   -1.5
## 10 0.91 0.95 0.242  -4 3.4e-02       NA     1.1 0.98    NA     NA     NA
## 11 0.91 0.94 0.396  -8 8.0e-08       NA     1.2 0.98    NA     NA     NA
## 12 0.91 0.94 0.594 -11 9.4e-09       NA     1.2 0.98    NA     NA     NA
## 13 0.91 0.94 1.000 -13 3.1e-10       NA     1.2 0.98    NA     NA     NA
## 14 0.91 0.94    NA -14 3.1e-10       NA     1.2 0.98    NA     NA     NA
```

```
##    complex eChisq    SRMR  eCRMS    eBIC
## 1      1.0 4.1e+03 8.3e-02 0.0903 3436.2
## 2      1.2 1.9e+03 5.8e-02 0.0686 1429.5
## 3      1.5 6.1e+02 3.2e-02 0.0427  192.3
## 4      1.6 2.3e+02 2.0e-02 0.0296  -98.5
## 5      1.7 9.7e+01 1.3e-02 0.0220 -153.3
## 6      1.9 3.5e+01 7.7e-03 0.0156 -143.2
## 7      2.0 1.2e+01 4.5e-03 0.0114 -101.3
## 8      2.1 3.6e+00 2.5e-03 0.0089  -53.0
## 9      2.1 2.0e-01 5.8e-04 0.0056   -7.9
## 10     2.2 3.2e-03 7.3e-05     NA     NA
## 11     2.1 9.7e-09 1.3e-07     NA     NA
## 12     2.1 1.2e-09 4.6e-08     NA     NA
## 13     2.1 1.7e-11 5.4e-09     NA     NA
## 14     2.1 1.7e-11 5.4e-09     NA     NA
```

```
#parallel analysis
sfo_fa %>% fa.parallel(.,fa='fa', n.iter = 50)
```


Parallel Analysis Scree Plots
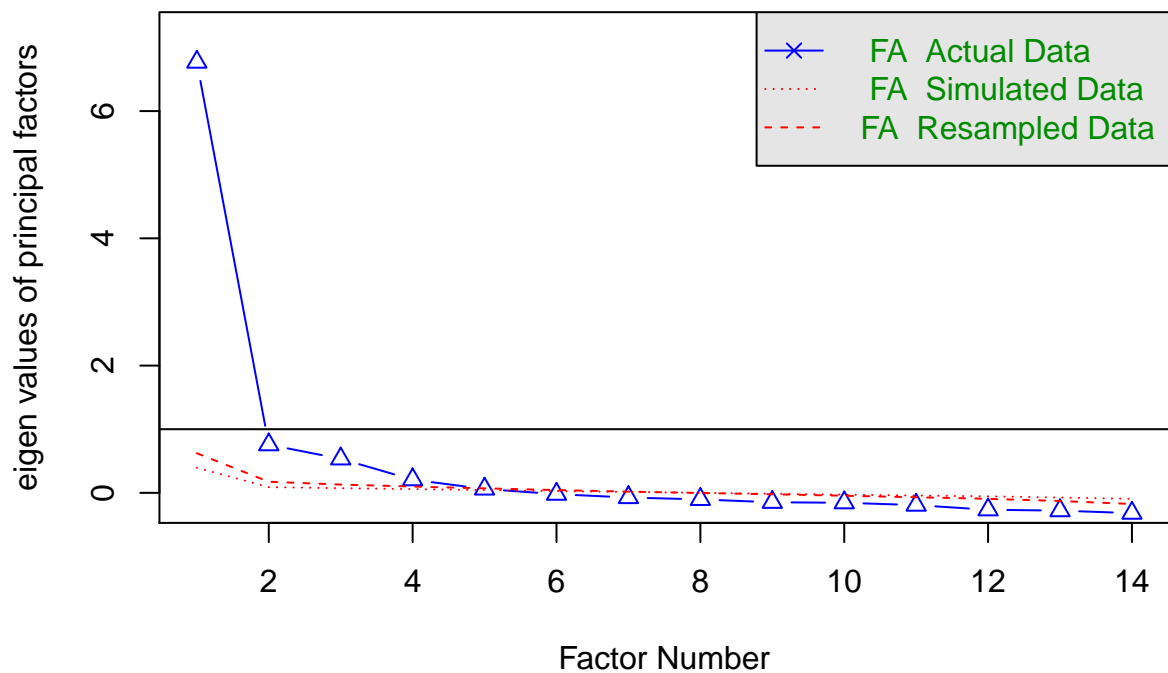
```
## Parallel analysis suggests that the number of factors =  5  and the number of components =  NA
```

```
fact5 <- sfo_fa %>% fa(., nfactors = 5)
```

```
## Loading required namespace: GPArotation
```

11

```
fact5$loadings
```

```
## 
## Loadings:
##       MR1   MR3   MR2   MR4   MR5
## Q6A         0.461       0.137 -0.115
## Q6B         0.850
## Q6C         0.741
## Q6D                     0.602  0.161
## Q6E               0.115 0.738
## Q6F   0.244             0.606
## Q6G   0.921
## Q6H   0.960
## Q6I                     0.173  0.602
## Q6J                            0.773
## Q6K               0.443 0.180  0.280
## Q6L         0.931
## Q6M   0.137       0.652        0.117
## Q6N         0.369 0.168 0.315  0.100
## 
##                   MR1   MR3   MR2   MR4   MR5
## SS loadings     1.874 1.627 1.555 1.466 1.106
## Proportion Var  0.134 0.116 0.111 0.105 0.079
## Cumulative Var  0.134 0.250 0.361 0.466 0.545
```

```
fa.diagram(fact5$loadings)
```

Factor Analysis
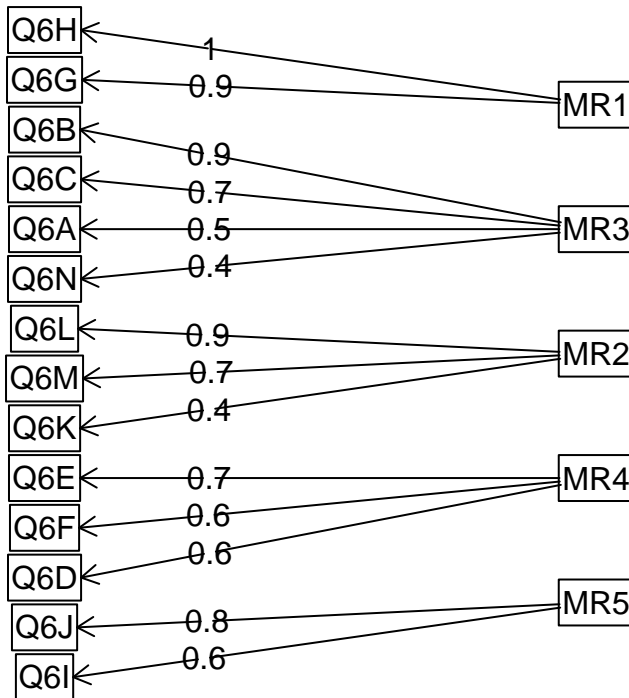


```
fact4 <- sfo_fa %>% fa(., nfactors = 4)
fact4$loadings
```

```
##
## Loadings:
##      MR2    MR1    MR3    MR4
## Q6A                 0.474  0.118
## Q6B                 0.850
## Q6C                 0.754
## Q6D  0.110                 0.636
## Q6E                        0.704
## Q6F         0.239          0.638
## Q6G         0.907
## Q6H         0.953
## Q6I  0.485                 0.222
## Q6J  0.745
## Q6K  0.714 -0.114          0.163
## Q6L  0.860
## Q6M  0.778  0.137         -0.115
## Q6N  0.249         0.373  0.306
##
##                 MR2   MR1   MR3   MR4
## SS loadings    2.729 1.834 1.664 1.512
## Proportion Var 0.195 0.131 0.119 0.108
## Cumulative Var 0.195 0.326 0.445 0.553
```

```
fa.diagram(fact4$loadings)
```

Factor Analysis
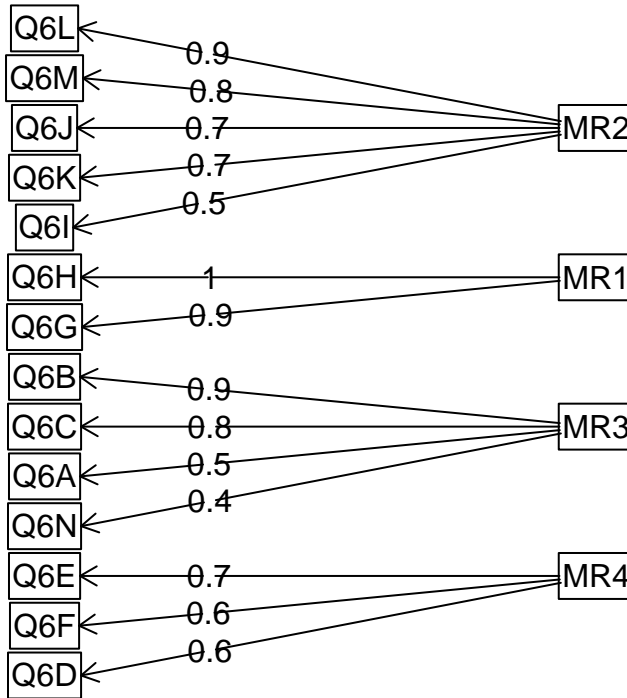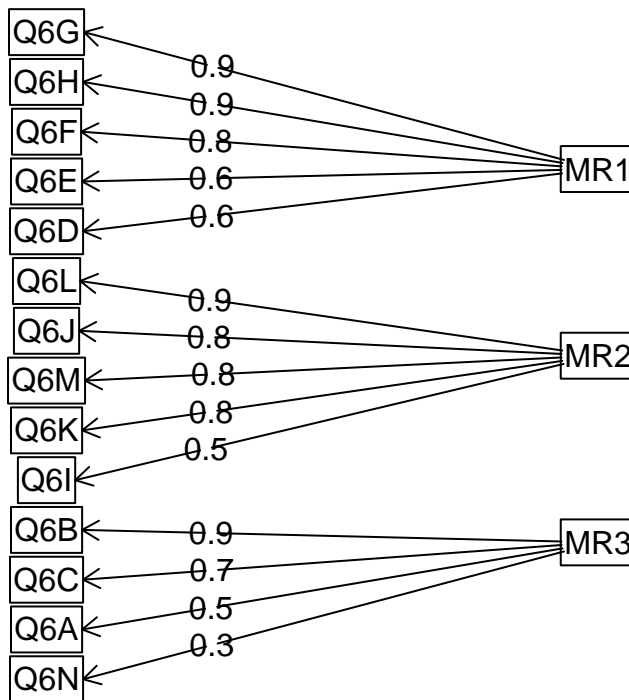


```
fact3 <- sfo_fa %>% fa(., nfactors = 3)
fact3$loadings
```

```
##
## Loadings:
##      MR1    MR2    MR3
## Q6A  0.204         0.460
## Q6B                0.850
## Q6C                0.749
## Q6D  0.612  0.142
## Q6E  0.642  0.104
## Q6F  0.831
## Q6G  0.895
## Q6H  0.885
## Q6I  0.214  0.511
## Q6J         0.779
## Q6K         0.758
## Q6L         0.888
## Q6M         0.779
## Q6N  0.316  0.270  0.341
##
##                 MR1    MR2    MR3
## SS loadings    3.256  2.943  1.618
```

```
## Proportion Var 0.233 0.210 0.116
## Cumulative Var 0.233 0.443 0.558
```

```r
fa.diagram(fact3$loadings)
```

Factor Analysis

3 factors gave the most Cumulative Var. We go with 3 factors.

Part A-3

```r
#replace "/" with ";"
sfo$Q7_text_All <- gsub('/',',',sfo$Q7_text_All)

#Visualize the most frequent words
sfo %>% dplyr::select(Q7_text_All) %>%
        unnest_tokens(word, Q7_text_All) %>%
        anti_join(stop_words) %>%
        count(word, sort=TRUE) %>%
        filter(n>50) %>%
        wordcloud2(shape='cardioid')
```

```
## Joining, by = "word"
```

Part B

Develop and Investigate Your Own Research Question

The SFO executives feel that additional insights can be gained from the customer satisfaction survey dataset. Based on your prior EDA deliverable and the topics we have discussed in class, develop an additional research question and execute a plan to evaluate it with these data using a method we covered this semester. Provide an appropriate explanation of your method of choice and how it applies to your question. If formal hypotheses are tested, clearly explain the results of these tests. If the method is more descriptive or data-driven, define how the results are evaluated, and provide sufficient output and data visuals to communicate the outcome. You don't need to fish for a "significant" finding here; even null or unexpected results can be useful if the hypothesis is reasonable.

```
library(stm)
```

```
## stm v1.3.6 successfully loaded. See ?stm for help.
##  Papers, resources, and other materials at structuraltopicmodel.com
```

```
library(tm)
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

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

```r
sfoDat <- textProcessor(documents=sfo$Q7_text_All, metadata = sfo)
```

```
## Building corpus...
## Converting to Lower Case...
## Removing punctuation...
## Removing stopwords...
## Removing numbers...
## Stemming...
## Creating Output...
```

```r
sfoPrep <- prepDocuments(documents=sfoDat$documents,
                         vocab = sfoDat$vocab,
                         meta = sfoDat$meta)
```
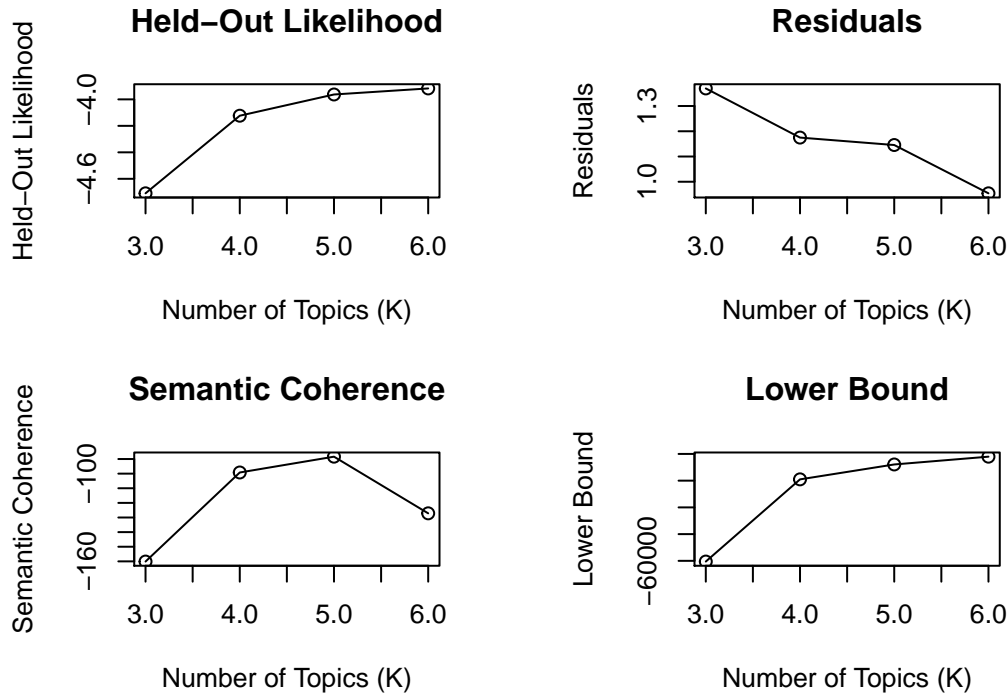
```
## Removing 6 of 227 terms (6 of 12498 tokens) due to frequency
## Your corpus now has 1557 documents, 221 terms and 12492 tokens.
```

```r
#find the best number of topics
kTest <- searchK(documents = sfoPrep$documents,
                 vocab = sfoDat$vocab,
                 K=c(3,4,5,6), verbose=FALSE, cores = 5)
```

```
## Using multiple-cores.  Progress will not be shown.
```

```r
plot(kTest)
```

Diagnostic Values by Number of Topics

We can settle with 5 topics based on the output above.

```
top5 <- stm(documents = sfoPrep$documents,
            vocab = sfoPrep$vocab,
            K=5, verbose = FALSE)
labelTopics(top5)
```

```
## Topic 1 Top Words:
##      Highest Prob: get, termin, long, airport, seat, free, crowdedmor
##      FREX: long, seat, accessdoesn't, airportoverloadeddidn't, cover, enoughdifficult, entir
##      Lift: cleannot, effici, morenot, open, shopsrestaur, long, seat
##      Score: long, get, accessdoesn't, airportoverloadeddidn't, cover, enoughdifficult, entir
## Topic 2 Top Words:
##      Highest Prob: comment, posit, general, sfo, area, facil, avail
##      FREX: posit, general, sfo, area, busi, everywher, foodwatersoda
##      Lift: allow, amenities-clock, appreci, atm', check-, curbsid, cut
##      Score: area, posit, sfo, general, comment, busi, everywher
## Topic 3 Top Words:
##      Highest Prob: securitycustom, inform, linesprocedur, longinefficientineffect, chang, enoughlack
##      FREX: securitycustom, inform, linesprocedur, longinefficientineffect, chang, enoughlack, inform
##      Lift: airtrainbart, electronicautom, humansfew, improv, sign, tofromon, chang
##      Score: chang, enoughlack, informationdisplay, rapid, screens-, smallnot, securitycustom
## Topic 4 Top Words:
##      Highest Prob: need, better, uniqu, restaur, food, shop, expens
##      FREX: better, uniqu, restaur, shop, healthier, low, qualityne
##      Lift: uniqu, air, areas-, artwork, artworkexhibitionschang, carts-, control
```

```
##        Score: better, uniqu, need, restaur, shop, food, qualityne
## Topic 5 Top Words:
##        Highest Prob: airport, find, signag, airlin, confusingsmallhard, gate, insid
##        FREX: find, signag, confusingsmallhard, gate, insid, airportdifficult, confusinghard
##        Lift: signag, airportdifficult, amen, awaynot, camera, clean, confusinghard
##        Score: find, signag, confusingsmallhard, gate, insid, shopsservic, airport
```

Topic 1: "Take too long get to airport and crowded" - commute to airport Topic 2: "Positive comment about facilities" -Facilities Topic 3: "Long and inefficient security custom line and information display" - Security Custom Topic 4: "Need better unique shops" -Shops & Restaur Topic 5: "Confusing signage inside airport" -Signage

```
plot(top5)
```

Top Topics

Expected Topic Proportions
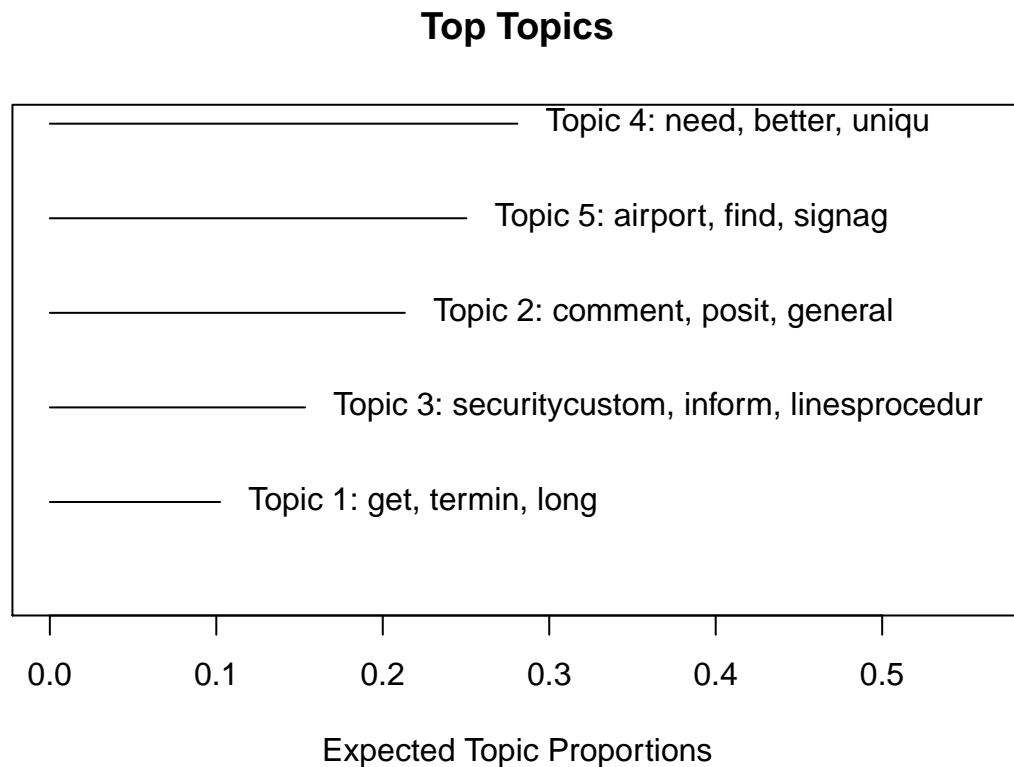
```
plot(top5, type='labels')
```

```
## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on ' get, termin, long, airport, seat, free,
## crowdedmor, accessdoesn't,' in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on ' get, termin, long, airport, seat, free,
## crowdedmor, accessdoesn't,' in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on ' get, termin, long, airport, seat, free,
## crowdedmor, accessdoesn't,' in 'mbcsToSbcs': dot substituted for <99>


## Warning in text.default(n/2, (i + 0.5), string, family = family, cex
## = text.cex): conversion failure on 'airportoverloadeddidn't, cover,
## enoughdifficult, entir, know, wifi-, avail,' in 'mbcsToSbcs': dot substituted
## for <e2>


## Warning in text.default(n/2, (i + 0.5), string, family = family, cex
## = text.cex): conversion failure on 'airportoverloadeddidn't, cover,
## enoughdifficult, entir, know, wifi-, avail,' in 'mbcsToSbcs': dot substituted
## for <80>


## Warning in text.default(n/2, (i + 0.5), string, family = family, cex
## = text.cex): conversion failure on 'airportoverloadeddidn't, cover,
## enoughdifficult, entir, know, wifi-, avail,' in 'mbcsToSbcs': dot substituted
## for <99>


## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'foodwatersoda, garden, hot, locker, loung,
## machin, needed-children', pet, play,' in 'mbcsToSbcs': dot substituted for <e2>


## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'foodwatersoda, garden, hot, locker, loung,
## machin, needed-children', pet, play,' in 'mbcsToSbcs': dot substituted for <80>


## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'foodwatersoda, garden, hot, locker, loung,
## machin, needed-children', pet, play,' in 'mbcsToSbcs': dot substituted for <99>
```

Topic 1:
get, termin, long, airport, seat, free, crowdedmor, accessdoesn...t,
airportoverloadeddidn...t, cover, enoughdifficult, entir, know, wifi–, avail,
far, airportno, around, awayconfusingtoo, difficulttak

Topic 2:
comment, posit, general, sfo, area, facil, avail, busi, everywher,
ersoda, garden, hot, locker, loung, machin, needed–children..., pet, play
restroom, sale

Topic 3:
ecuritycustom, inform, linesprocedur, longinefficientineffect, chang,
ghlack, informationdisplay, rapid, screens–, smallnot, awaydifficult, car,
enter, rental, rude, foconfusingemploye, far, get, airtrainbart, improv

Topic 4:
eed, better, uniqu, restaur, food, shop, expens, secur, healthier, low,
lityne, select, one, servicesamen, artwork, artworkexhibitionschang,
frequent, hous, starbuckspeetscoffe, type

Topic 5:
ort, find, signag, airlin, confusingsmallhard, gate, insid, termin, enough,
get, airportdifficult, confusinghard, correct, outsid, expens, personnel,
awaynot, enoughhard, park

```r
plot(top5, type='perspectives', topics = c(4,1))
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * :
## conversion failure on 'accessdoesn't' in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * :
## conversion failure on 'accessdoesn't' in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * :
## conversion failure on 'accessdoesn't' in 'mbcsToSbcs': dot substituted for <99>
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * : font
## metrics unknown for Unicode character U+2019
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * :
## conversion failure on 'airportoverloadeddidn't' in 'mbcsToSbcs': dot substituted
## for <e2>
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * :
## conversion failure on 'airportoverloadeddidn't' in 'mbcsToSbcs': dot substituted
## for <80>
```

```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * :
## conversion failure on 'airportoverloadeddidn't' in 'mbcsToSbcs': dot substituted
## for <99>
```
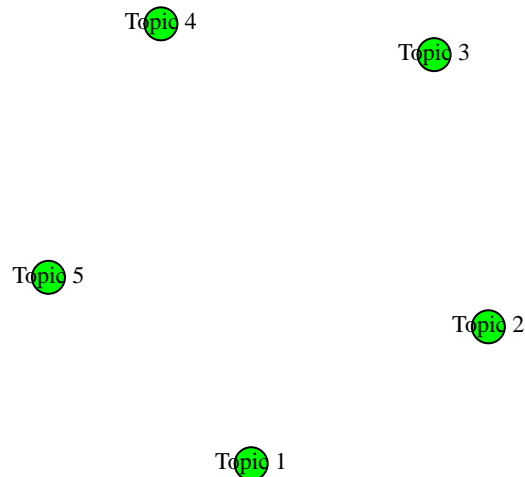
```
## Warning in text.default(diff, rand, model$vocab[words], cex = text.cex * : font
## metrics unknown for Unicode character U+2019
```

need

restaur   expens   termin
            secur
            qualityne
            healthier
     uniqu   cover
                  low entir
                    airportoverloadedtidh..t   long
           food   select
better   shop   get
                    enoughdifficult free
                          seat
                      airport
                    crowdedmor

                              accessdoesn...t

Topic 4   Topic 1

```
plot(topicCorr(top5))
```

Topic 4

Topic 3

Topic 5

Topic 2

Topic 1

No correlated relationship among these 5 topics

```
findThoughts(top5, texts = sfoPrep$meta$Q7_text_All, n=2)
```

```
##
##  Topic 1:
##       Airport Layout-Terminals too far away,confusing,too difficult,takes too long to get around air
##       Airport Layout-Terminals too far away,confusing,too difficult,takes too long to get around airp
##  Topic 2:
##        Other amenities-clocks, hand sanitizer, strollers, ATM's, curbside check-in, payphone   Other
##       General positive comments about SFO  Other Facilities Needed-children's play areas, garden area
##  Topic 3:
##        Information screens-Too small,not enough,lack information,displays change too rapidly  Rental
##       Information screens-Too small,not enough,lack information,displays change too rapidly  Security
##  Topic 4:
##       Need Starbucks,Peets,coffee house type restaurants (both before and after security)   Need bet
##       Food is low quality,need better, more, or healthier selections  Need better, more, or more uni
##  Topic 5:
##        Parking too far away,not enough,hard to find ,expensive  Signage inside airport confusing,smal
##       Parking too far away,not enough,hard to find ,expensive  Signage outside airport confusing,hard
```

```
#Topic 4
cloud(top5,topic = 4)
```

```
## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'tv'scomput' in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'tv'scomput' in 'mbcsToSbcs': dot substituted for <80>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'tv'scomput' in 'mbcsToSbcs': dot substituted for <99>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'tv'scomput' in 'mbcsToSbcs': dot substituted
## for <e2>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'tv'scomput' in 'mbcsToSbcs': dot substituted
## for <80>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'tv'scomput' in 'mbcsToSbcs': dot substituted
## for <99>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for Unicode character U+2019
```
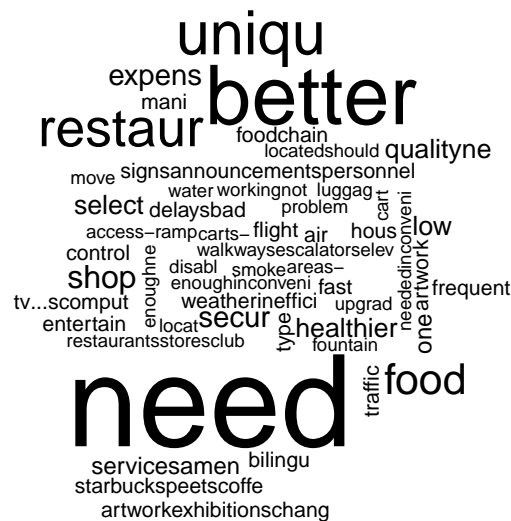


```
#Topic 5
cloud(top5,topic = 5)
```

airport

airportdifficult

confusingsmallhard

correct hook find outlet

expens gate train neg

explainedimprov inefficientrudenot

far duty-fre amen freecustom experi camera

local duti etc light oldoutd night enough

clean get facil electr

outsid rulescustom moreoth rule negat

upgradingairportawaynot

restrooms- measur enoughhard well small

termin look personnel

shopsservic insid miss park

signag airlin

confusinghard