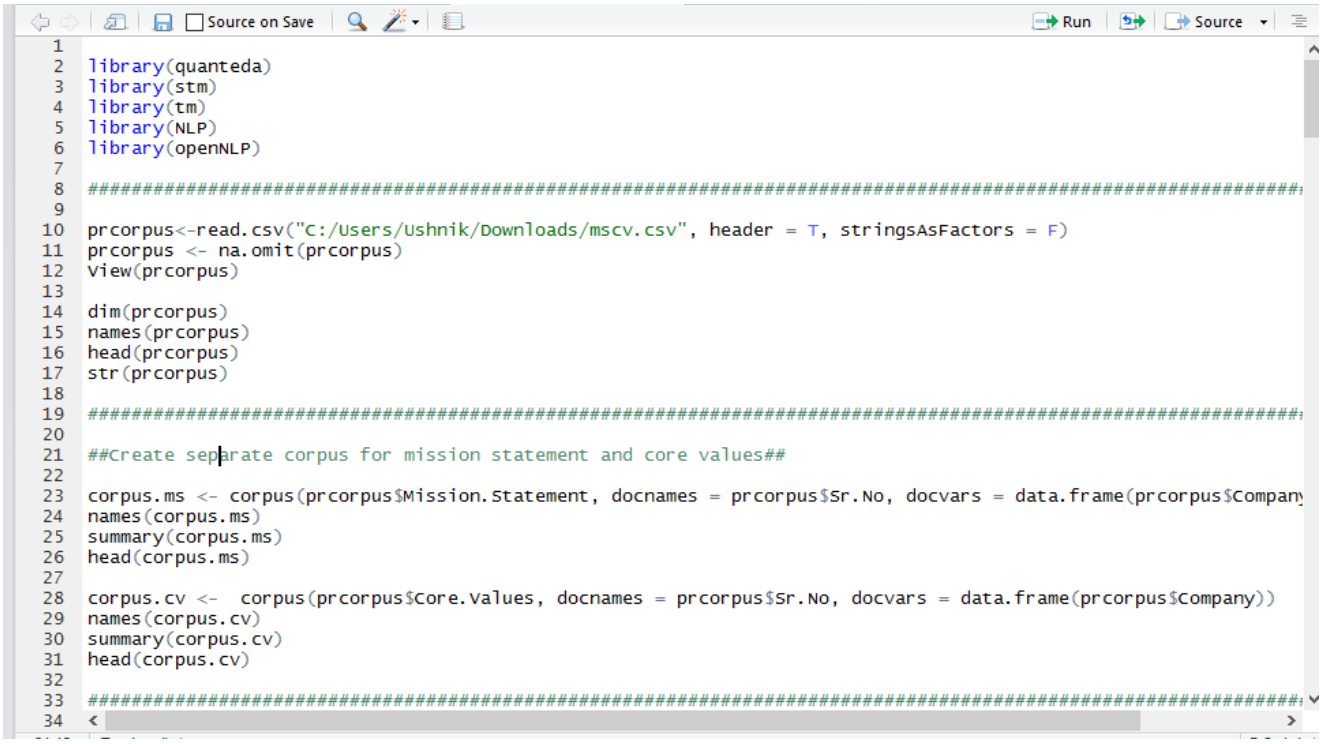
Name: Ushnik Dasgupta
NYU ID: N17591222

Business Analytics

Assignment 7

PART 1: Analyzing how to structure a firm for data-analysis readiness

Creating two corpuses - one for mission statements and one for core values:



```
library(quanteda)
library(stm)
library(tm)
library(NLP)
library(openNLP)

##############################################################################################################

prcorpus<-read.csv("C:/Users/Ushnik/Downloads/mscv.csv", header = T, stringsAsFactors = F)
prcorpus <- na.omit(prcorpus)
View(prcorpus)

dim(prcorpus)
names(prcorpus)
head(prcorpus)
str(prcorpus)

##############################################################################################################

##Create separate corpus for mission statement and core values##

corpus.ms <- corpus(prcorpus$Mission.Statement, docnames = prcorpus$Sr.No, docvars = data.frame(prcorpus$Company
names(corpus.ms)
summary(corpus.ms)
head(corpus.ms)

corpus.cv <-  corpus(prcorpus$Core.Values, docnames = prcorpus$Sr.No, docvars = data.frame(prcorpus$Company))
names(corpus.cv)
summary(corpus.cv)
head(corpus.cv)

##############################################################################################################
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Filter

	Sr.No	Company	Mission.Statement	Core.Values
1	1	Valen Analytics	Valen's mission is to help our clients achieve their g...	Live the Golden Rule: We treat our customers, emplo...
11	2	Bottome Line	Bottom Line's mission is to help low income, first gen...	Relationships : We are dedicated to building strong, ...
29	3	Trifacta	At Trifacta, we know that the key to winning with Big ...	1. Discovering is something of an umbrella term for t...
37	4	Cloudera	Cloudera's mission is to bring the power of Hadoop, ...	Be Open : At our core, we are open. We believe in tra...
60	5	Birst	Consume enterprise data with easy to use analytics f...	Broad and complete business visibility, fast. Real tim...
68	6	Gainsight	Gainsight is changing the way businesses understan...	HYPER-CONNECTED TEAM : We have offices in Silicon...
78	7	Google	Google's mission is to organize the world's informatio...	Focus on the user and all else will follow : Since the ...
95	8	Tableau Software	A mission that matters. We help people see and unde...	User Interface Design : We are lucky to have a really ...
105	9	Actifio	Actifio Enterprise fundamentally changes the nature ...	Achieving Business Resiliency means being ready w...
107	10	Hortonworks	Our mission is simple: To revolutionize and commodi...	We believe strongly in our core set of shared values....
131	11	Informatica	Informatica Big Data Management tackles those probl...	Accountability and Integrity : We are personally acco...
140	12	SAP	Our mission is to help every customer become a bes...	Gender Intelligence : Ensuring awareness of the ben...
152	13	SAS Institue	SAS delivers proven solutions that drive innovation ...	Approachable. Customer-Driven. Swift & Agile. Innov...
162	14	Microsoft	Empower every person and organization on the plan...	Growth mindset: At Microsoft, we're insatiably curiou...
172	15	Alpine Data	We exist to break down the barriers between people,...	Work with the latest innovations with ease. We embra...
178	16	MongoDB	MongoDB unleashes the power of software and data ...	Think Big, Go Far : We are big dreamers with a passio...
192	17	Teradata	Analytics and data unleash the potential of great co...	Passionate : We do great things because we believe i...
208	18	Oracle	Mission Statements and Vision Statements are writte...	Integrity: Oracle employees will demonstrate honest...

Showing 1 to 19 of 30 entries

```
Console ~/
> dim(prcorpus)
[1] 30  4
> names(prcorpus)
[1] "Sr.No"             "Company"           "Mission.Statement" "Core.Values"
>
```

Summary of Mission Statement Corpus:

```
33:1  (Top Level)
Console ~/
> summary(corpus.ms)
Corpus consisting of 30 documents.

 Text Types Tokens Sentences prcorpus.Company
    1    21     22         1  Valen Analytics
    2    23     26         1     Bottome Line
    3    16     17         1         Trifacta
    4    23     29         1         Cloudera
    5    24     26         2            Birst
    6    38     57         4        Gainsight
    7    40     46         3           Google
    8    86    131        12 Tableau Software
    9    37     46         2          Actifio
   10    68     92         4      Hortonworks
   11    59     81         2      Informatica
   12    73    125         5              SAP
   13    11     11         1     SAS Institue
   14    50     75         4        Microsoft
   15    25     25         1      Alpine Data
   16    11     11         1          MongoDB
   17    68     97         4         Teradata
   18    78    117         6           Oracle
   19    27     31         1           Splunk
   20    28     31         1           Amazon
   21    52     78         5              IBM
   22    22     27         1    Microstrategy
   23    32     42         2         Facebook
   24    22     24         1             Tamr
   25    28     31         2       Accenture
   26    23     24         1        Cognizant
   27     8      8         1              PwC
   28    17     18         1              E&Y
   29    26     31         1             KPMG
   30    36     47         2          Deloitte
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Summary of Core Values Corpus:

```
> summary(corpus.cv)
Corpus consisting of 30 documents.

 Text Types Tokens Sentences prcorpus.Company
    1    87    148        10  Valen Analytics
    2   176    357        15     Bottome Line
    3   286    578        25         Trifacta
    4   163    378        30         Cloudera
    5    45     59         7            Birst
    6    92    132         5        Gainsight
    7   287    557        19           Google
    8   151    255        19 Tableau Software
    9    58     86         3           Actifio
   10   183    362        35      Hortonworks
   11   132    285        16      Informatica
   12    60     82         4              SAP
   13    10     14         5     SAS Institue
   14   145    277        16       Microsoft
   15    80    120         7      Alpine Data
   16   119    197        13          MongoDB
   17    74    132        10         Teradata
   18   111    215        11           Oracle
   19   216    460        21           Splunk
   20   305    663        49           Amazon
   21    40     56         2              IBM
   22   106    166        10    Microstrategy
   23   146    237        16         Facebook
   24    67     96         4             Tamr
   25   103    176         4       Accenture
   26   152    264        17        Cognizant
   27   122    223        11              PwC
   28    23     33         3              E&Y
   29   105    195        12             KPMG
   30   100    173         8         Deloitte
```

Cleaning and Tokenizing Text Data:

```
32
33  ##################################################################################################################
34
35  ##Clean Mission Statement corpus##
36
37  corpus.ms <- toLower(corpus.ms, keepAcronyms = F)
38  corpus.ms
39  clean.corpus.ms <- tokenize(corpus.ms,
40                              removeNumbers = T,
41                              removePunct = T,
42                              removeSeparators = T,
43                              removeHyphens = T,
44                              removeSymbols = T,
45                              verbose = T,
46                              ngrams = 1)
47  head(clean.corpus.ms)
48
49  ##Clean Core Values Corpus##
50
51  corpus.cv <- toLower(corpus.cv, keepAcronyms = F)
52  corpus.cv
53  clean.corpus.cv <- tokenize(corpus.cv,
54                              removeNumbers = T,
55                              removePunct = T,
56                              removeSeparators = T,
57                              removeHyphens = T,
58                              removeSymbols = T,
59                              verbose = T,
60                              ngrams = 1)
61  head(clean.corpus.cv)
62
63  ##################################################################################################################
64
65
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Creating Document Feature Matrix and returning top 50 frequent words for Mission Statements and Core Values:

```
63  ##################################################################################################################
64
65  ##create document feature matrix from each of the clean corpuses, stem and return top 50 features##
66
67  ##dfm from Mission Statement corpus##
68
69  help(dfm)
70  dfm.ms <- dfm(clean.corpus.ms,
71                ignoredFeatures = stopwords("English"),
72                stem = T)
73  head(dfm.ms,30)
74  top.features.ms <- topfeatures(dfm.ms, n=50)
75  top.features.ms
76
77  ##dfm from Core Value Corpus##
78
79  dfm.cv <- dfm(clean.corpus.cv,
80                ignoredFeatures = stopwords("English"),
81                stem = T)
82  head(dfm.ms,30)
83  top.features.cv <- topfeatures(dfm.cv, n=50)
84  top.features.cv
85
86  ##################################################################################################################
87
```

```
> top.features.ms <- topfeatures(dfm.ms, n=50)
> top.features.ms
     data   mission   compani      busi      help     world     peopl    custom      valu      work      make
       25        16        14        14        12        11        10         9         9         9         8
enterpris technolog    inform    analyt    improv       use       way     organ     drive     build       can
        8         8         7         7         7         6         6         6         6         6         6
    innov     power   softwar   process    integr statement      lead    client    achiev   problem       big
        6         5         5         5         5         5         5         4         4         4         4
   hadoop   employe     deliv     everi     solut    servic    vision    provid      solv     decis   generat
        4         4         4         4         4         4         4         4         3         3         3
   storag    govern     cloud     chang      oper   success
        3         3         3         3         3         3
```

```
> top.features.cv <- topfeatures(dfm.cv, n=50)
> top.features.cv
   custom      work     peopl      data      make   success      take      valu      will    leader   employe
       53        42        32        29        28        27        25        24        23        23        22
   commit     innov      best   compani       new      team     everi communiti       can     thing     learn
       21        21        20        20        19        19        18        17        16        16        15
      one        us   product     think    believ      busi     oracl    client   passion     alway     focus
       15        15        15        15        15        15        15        15        14        14        14
     help      time       way     build   problem      need    better      open    divers     deliv    provid
       14        14        14        13        13        13        13        13        13        13        12
  respons    action    result       get     world      idea
       12        12        12        12        12        12
>
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Removing certain irrelevant words from each DFM:

```
86  ##################################################################################################################
87
88  ##Create custom dictionary and Remove/Filter selected irrelevant words from each corpus##
89
90  ## Mission Statement##
91
92  word.list.ms <- c("make", "world", "way", "can", "us", "use")
93  dfm.ms.1 <- dfm(clean.corpus.ms,
94                  ignoredFeatures = c(word.list.ms, stopwords("English")),
95                  stem = T)
96
97  topfeatures.ms.1<-topfeatures(dfm.ms.1, n=50)
98  topfeatures.ms.1
99
100 ##Core Values##
101
102 word.list.cv <- c("everi", "will", "take", "one", "us", "can")
103 dfm.cv.1 <- dfm(clean.corpus.cv,
104                 ignoredFeatures = c(word.list.cv, stopwords("English")),
105                 stem = T)
106
107 topfeatures.cv.1<-topfeatures(dfm.cv.1, n=50)
108 topfeatures.cv.1
109
110 ##################################################################################################################
```

```
>
> topfeatures.ms.1<-topfeatures(dfm.ms.1, n=50)
> topfeatures.ms.1
     data   mission   compani      busi      help     peopl    custom      valu      work enterpris technolog
       25        16        14        14        12        10         9         9         9         8         8
   inform    analyt    improv     organ     drive     build     innov     power     world   softwar   process
        7         7         7         6         6         6         6         5         5         5         5
   integr statement      lead    client    achiev   problem       big    hadoop   employe     deliv     everi
        5         5         5         4         4         4         4         4         4         4         4
    solut    servic    vision    provid      solv     decis   generat    storag    govern       use     cloud
        4         4         4         4         3         3         3         3         3         3         3
    chang      oper   success    cultur      time transform
        3         3         3         3         3         3
```

```
>
> topfeatures.cv.1<-topfeatures(dfm.cv.1, n=50)
> topfeatures.cv.1
   custom      work     peopl      data      make   success      valu    leader   employe    commit     innov
       53        42        32        29        28        27        24        23        22        21        21
     best   compani       new      team     everi communiti     thing     learn   product     think    believ
       20        20        19        19        18        17        16        15        15        15        15
     busi     oracl    client   passion     alway     focus      help      time       way     build   problem
       15        15        15        14        14        14        14        14        14        13        13
     need    better      open    divers     deliv    provid   respons    action    result       get     world
       13        13        13        13        13        12        12        12        12        12        12
     idea     share  individu    achiev   qualiti    person
       12        12        11        11        11        11
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Exploring Text Context:

```
110 ##########################################################################################################
111
112 ##Exploration of context##
113
114 ##Mission Statement##
115
116 kwic(clean.corpus.ms, "data", 2)
117 kwic(clean.corpus.ms, "company", 2)
118
119 ##Core Values##
120
121 kwic(clean.corpus.cv, "community", 2)
122 kwic(clean.corpus.cv, "work", 3)
123 kwic(clean.corpus.cv, "success", 2)
124
125 ##########################################################################################################
```

```
> kwic(clean.corpus.ms, "data", 2)
                  contextPre keyword              contextPost
 [1, 16]      by leveraging [    data ] to make
 [3, 12]           with big [    data ] lies with
  [5, 3] consume enterprise [    data ] with easy
 [5, 12]        birst share [    data ] across all
 [8, 12]   understand their [    data ] eight words
 [8, 35]         amount of [    data ] as in
 [8, 48]       within these [    data ] are huge
 [8, 66]           power of [    data ] at their
 [8, 86]         people use [    data ] to solve
 [8, 93]     make analyzing [    data ] fast and
  [9, 8]          nature of [    data ] protection and
 [9, 19]    actifio virtual [    data ] pipeline it
[10, 15]             of big [    data ] via open
[10, 76]        the world's [    data ] will be
 [11, 3]    informatica big [    data ] management tackles
[11, 16]         pillars of [    data ] management integration
[11, 41]         all things [    data ] future informatica
[11, 58]         secure big [    data ] to repeatably
[15, 10]     between people [    data ] and advanced
 [16, 8]       software and [    data ] for innovators
 [17, 3]      analytics and [    data ] unleash the
[17, 12]      analytics and [    data ] have become
[17, 30]         view their [    data ] as such
[17, 84]            to what [    data ] and analytics
[19, 12]       the machine [    data ] generated by
```

```
> kwic(clean.corpus.ms, "company", 2)
                  contextPre keyword       contextPost
 [7, 17]       useful our [ company ] has packed
[11, 26]         said the [ company ] which describes
[17, 43]     whether your [ company ] is working
[18, 29]             by a [ company ] or business
[20, 10] customer centric [ company ] to build
[21, 34]          for our [ company ] and the
[23, 17]             as a [ company ] we have
```

Name: Ushnik Dasgupta
NYU ID: N17591222

```
> kwic(clean.corpus.cv, "community", 2)
                contextPre   keyword           contextPost
  [2, 14] students schools [ community ] partners supporters
  [4, 23]         with the [ community ] that way
 [4, 172]          and the [ community ] we believe
 [4, 278]           in the [ community ] we have
 [10, 36]    customers and [ community ] we are
 [10, 85]           in our [ community ] and in
[10, 155]       engage our [ community ] partners and
[10, 199]     partners and [ community ] we strive
 [25, 38]           to the [ community ] as responsible
```

```
> kwic(clean.corpus.cv, "work", 3)
                          contextPre keyword               contextPost
  [1, 85]                 we enjoy our [    work ] most of all
 [2, 140]                aspect of our [    work ] we hold ourselves
 [4, 292]           how our customers [    work ] at our core
  [6, 39]                  idea of a [    work ] day is eight
 [7, 109]          believe they should [    work ] so well you
 [7, 377]               we continue to [    work ] on making it
 [8, 114]          question freedom to [    work ] without relying on
 [10, 45]              impact that our [    work ] can have on
 [10, 60]          to delivering great [    work ] that makes us
[10, 104]          of delivering great [    work ] clear and direct
[10, 289]                 to life than [    work ] we believe work
[10, 292]              work we believe [    work ] should be fun
 [11, 82]                 areas of our [    work ] we set the
 [12, 16]                women and men [    work ] more effectively together
[14, 167]           single mission we [    work ] together building on
  [15, 1]                             [    work ] with the latest
 [16, 75]          amount of important [    work ] and we are
[16, 154]          our convictions but [    work ] hard to ensure
 [18, 38]     teamwork oracle employees [    work ] together as a
[18, 113]                day to day [    work ] processes and seek
[19, 279]              the same time [    work ] hard play harder
  [20, 9]           the customer and [    work ] backwards they work
 [20, 12]          work backwards they [    work ] vigorously to earn
[20, 143]      diverse perspectives and [    work ] to disconfirm their
[20, 212]           coaching others we [    work ] on behalf of
 [22, 36]             people need to [    work ] well with each
 [22, 89]             world we value [    work ] life balance it's
[22, 110]          healthy balance of [    work ] and play this
 [25, 95]         diverse teams you'll [    work ] with diverse teams
[25, 146]               best ideas to [    work ] every day and
[26, 125] our environment collaboration [    work ] together to achieve
[26, 157]          share knowledge and [    work ] together the more
 [29, 27]             firms clients we [    work ] together we bring
[29, 142]      and perspectives through [    work ] in our communities
```

```
> kwic(clean.corpus.cv, "success", 2)
               contextPre keyword      contextPost
  [1, 43] our customer's [ success ] to provide
  [2, 93]    and achieve [ success ] in college
 [2, 150]      long term [ success ] of our
 [2, 259]      their own [ success ] excellence we
 [4, 107] our collective [ success ] we enable
 [4, 176]   believe that [ success ] for the
 [4, 181]     company is [ success ] for its
[11, 143]  our customers [ success ] our customers
[11, 146]  our customers [ success ] is our
[11, 149]         is our [ success ] we are
[11, 222]    doing their [ success ] is inseparable
 [17, 18]         to the [ success ] of our
 [17, 65]      value and [ success ] for our
[19, 350]    echelons of [ success ] we are
 [26, 14]     powers our [ success ] and exceeding
[26, 136] of cognizant's [ success ] is the
[27, 193]         to our [ success ]
>
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Checking for top bigrams:

```
##Checking for bigram significance##

##Mission Statement##

clean.corpus.ms1 <- tokenize(corpus.ms,
                             removeNumbers = T,
                             removePunct = T,
                             removeSeparators = T,
                             removeHyphens = T,
                             removeSymbols = T,
                             verbose = T, ngrams = 2)
dfm.bigram.ms <- dfm(clean.corpus.ms1,
                     ignoredFeatures = c(word.list.ms, stopwords("English")),
                     stem = T)

topfeatures.bigram.ms<-topfeatures(dfm.bigram.ms, n=50)
topfeatures.bigram.ms

##Core Values##

clean.corpus.cv1 <- tokenize(corpus.cv,
                             removeNumbers = T,
                             removePunct = T,
                             removeSeparators = T,
                             removeHyphens = T,
                             removeSymbols = T,
                             verbose = T, ngrams = 2)
dfm.bigram.cv <- dfm(clean.corpus.cv1,
                     ignoredFeatures = c(word.list.cv, stopwords("English")),
                     stem = T)

topfeatures.bigram.cv<-topfeatures(dfm.bigram.cv, n=50)
topfeatures.bigram.cv

##############################################################################################################
```

```
> topfeatures.bigram.ms<-topfeatures(dfm.bigram.ms, n=50)
> topfeatures.bigram.ms
             big_data         apache_hadoop          mission_stat          world's_lead       leading_compani
                    4                     3                     3                     3                     3
       solve_problem            data_manag        cloud_deliveri   professional_servic           vision_stat
                    2                     2                     2                     2                     2
     short_paragraph             core_valu        leading_provid information_technolog    technology_consult
                    2                     2                     2                     2                     2
    business_process      process_outsourc    outsourcing_servic          services_ded       companies_build
                    2                     2                     2                     2                     2
        build_strong         stronger_busi          valen's_miss        clients_achiev       leveraging_data
                    2                     2                     1                     1                     1
      informed_decis           bottom_lin           line's_miss             help_low            low_incom
                    1                     1                     1                     1                     1
        income_first          first_gener       generation_stud          students_get         college_gradu
                    1                     1                     1                     1                     1
              go_far               data_li       cloudera's_miss       hadoop_mapreduc    distributed_storag
                    1                     1                     1                     1                     1
  enterprise_internet     government_sector     consume_enterpris       enterprise_data           birst_shar
                    1                     1                     1                     1                     1
          share_data           data_across              teams_us           using_cloud             cloud_bas
                    1                     1                     1                     1                     1
```

```
> topfeatures.bigram.cv<-topfeatures(dfm.bigram.cv, n=50)
> topfeatures.bigram.cv
   oracle_employe          long_term        work_togeth          every_day       share_inform  customers_success
               10                  5                  5                  4                  4                  4
        high_qual          open_sourc          right_th customer_satisfact        member_firm  customers_employe
                3                  3                  3                  3                  3                  2
  employees_vendor     community_partn         highest_qu     purchased_item         might_want    attribute_field
                2                  2                  2                  2                  2                  2
   constraints_e.g         best_think         game_chang      search_result  continuous_improv        new_product
                2                  2                  2                  2                  2                  2
       really_tal          ask_quest         make_decis   delivering_great         great_work          team_memb
                2                  2                  2                  2                  2                  2
        good_idea       good_listen personally_account   highest_standard    products_servic     entire_compani
                2                  2                  2                  2                  2                  2
     highest_ratio       other's_idea    customer_obsess   diverse_perspect     global_compani       achieve_amaz
                2                  2                  2                  2                  2                  2
       amazing_th      alpine_chorus          think_big      skills_experi          work_hard consistently_treat
                2                  2                  2                  2                  2                  2
    solving_custom         best_peopl
                2                  2
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Wordcloud of most frequent words

```
160
161 ##############################################################################################################
162
163 ##WordCloud for Mission Statements##
164
165 library(wordcloud)
166 wordcloud(names(topfeatures.ms.1),
167           topfeatures.ms.1,max.words = 200,
168           scale = c(3,.1),
169           colors = brewer.pal(8,"Set1"))
170
171 ##WordCloud for Core Values##
172
173 wordcloud(names(topfeatures.cv.1),
174           topfeatures.cv.1,max.words = 200,
175           scale = c(3,.1),
176           colors = brewer.pal(8,"Set1"))
177
178 ##############################################################################################################
179
```

Wordcloud for Mission Statements:

Name: Ushnik Dasgupta
NYU ID: N17591222

Wordcloud for Core Values:

Checking for frequent word correlation:

```
##Specifying a correlation limit of 0.6##

##Mission Statement##

ms.tm <- convert(dfm.ms.1, to="tm")
findAssocs(ms.tm, c("data","mission","busi"), corlimit = 0.7)

##Core Values##

cv.tm <- convert(dfm.cv.1, to="tm")
findAssocs(ms.tm, c("work","peopl","employe"), corlimit = 0.7)


#########################################################################
```

Correlation for Mission Statements:

```
> findAssocs(ms.tm, c("data","mission","busi"), corlimit = 0.7)
$data
   time  number softwar
   0.78    0.76    0.74

$mission
open
0.79

$busi
oper
0.74
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Correlation for Core Values:

```
> cv.tm <- convert(dfm.cv.1, to="tm")
> findAssocs(ms.tm, c("work","peopl","employe"), corlimit = 0.7)
$work
    empow    person    planet    inspir      push    status       quo microsoft      also   freedom    achiev
     0.78      0.78      0.78      0.78      0.78      0.78      0.78      0.78      0.78      0.78      0.71
    everi       day    differ
     0.71      0.71      0.71

$peopl
     time       see     eight      word    amount       idc      huge  opportun     human   realiti      need
     0.78      0.76      0.76      0.76      0.76      0.76      0.76      0.76      0.76      0.76      0.76
fingertip   tableau     exact   product    analyz      fast    beauti     anyon
     0.76      0.76      0.76      0.76      0.76      0.76      0.76      0.76

$employe
    ensur      mean     empow    person    planet    inspir      push    status       quo microsoft      also
     0.86      0.86      0.81      0.81      0.81      0.81      0.81      0.81      0.81      0.81      0.81
  freedom
     0.81
```

Topic Modelling for Mission Statements:

```
library(stm)
help("textProcessor")
temp<-textProcessor(documents=prcorpus$Mission.Statement, metadata = prcorpus)
names(temp)
meta<-temp$meta
vocab<-temp$vocab
docs<-temp$documents
out <- prepDocuments(docs, vocab, meta)
docs<-out$documents
vocab<-out$vocab
meta <-out$meta

prevfit <-stm(docs , vocab ,
              K=10,
              verbose=TRUE,
              data=meta,
              max.em.its=10)

topics <-labelTopics(prevfit , topics=c(1:10))
topics #shows topics with highest probability words

#explore the topics in context.  Provides an example of the text
help("findThoughts")
findThoughts(prevfit, texts = prcorpus$Mission.Statement,  topics = c(1:2),  n = 3)

help("plot.STM")
plot.STM(prevfit, type="summary")
plot.STM(prevfit, type="labels", topics=c(1,5,10))
plot.STM(prevfit, type="perspectives", topics = c(6,10))

# to aid on assigment of labels & intepretation of topics
help(topicCorr)
library(igraph)
mod.out.corr <- topicCorr(prevfit)  #Estimates a graph of topic correlations
plot.topicCorr(mod.out.corr)
```

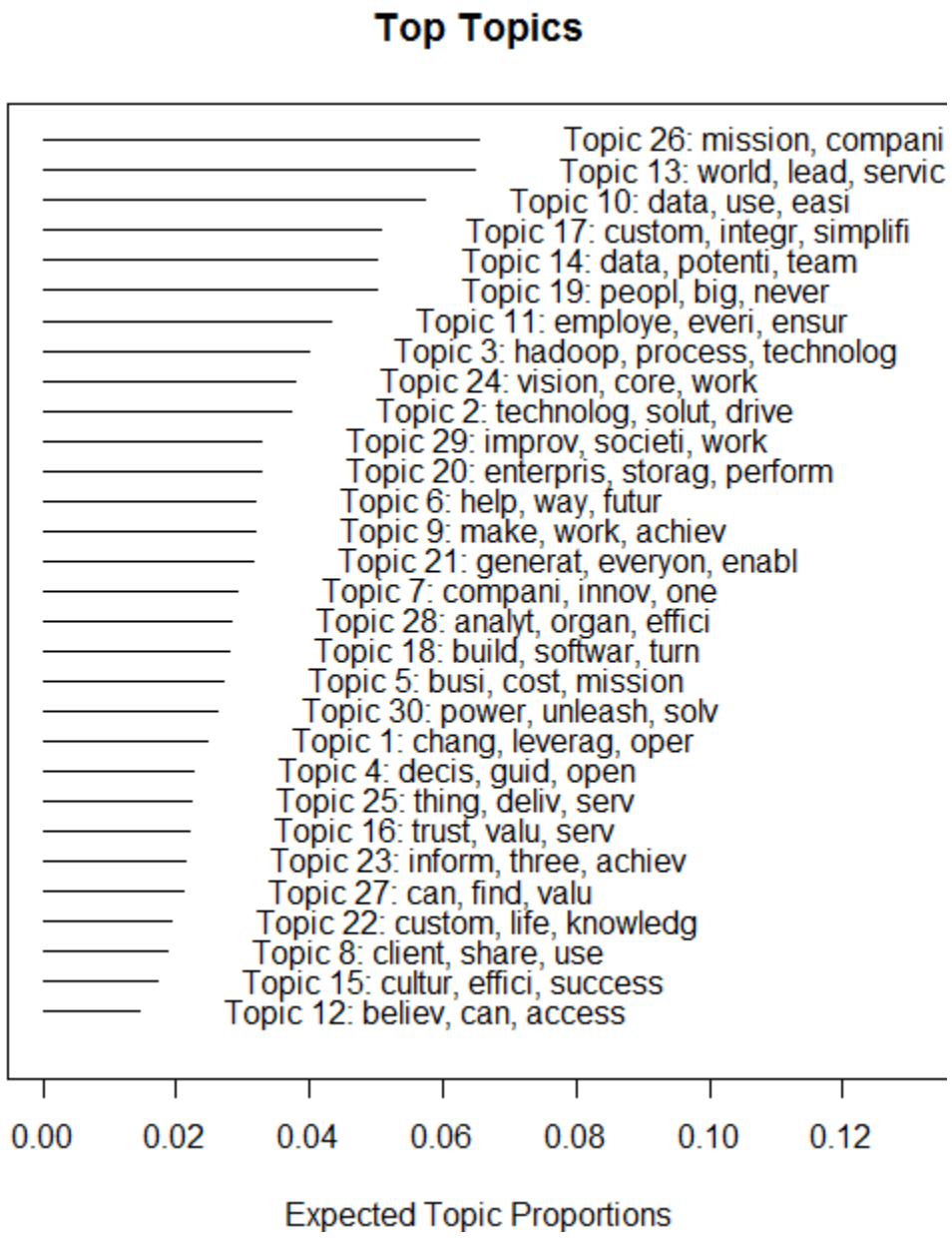
Name: Ushnik Dasgupta
NYU ID: N17591222

Displaying some of the top words with highest probability:

```
Topic 1 Top Words:
         Highest Prob: chang, leverag, oper, understand, everyth, key, success
         FREX: chang, leverag, understand, oper, everyth, key, power
         Lift: chang, leverag, understand, everyth, key, oper, success
         Score: chang, leverag, key, oper, understand, everyth, success
Topic 2 Top Words:
         Highest Prob: technolog, solut, drive, everi, today, dedic, deliv
         FREX: solut, technolog, today, everi, drive, dedic, deliv
         Lift: solut, today, technolog, drive, everi, dedic, deliv
         Score: solut, technolog, today, everi, drive, dedic, deliv
Topic 3 Top Words:
         Highest Prob: hadoop, process, technolog, provid, outsourc, open, will
         FREX: hadoop, process, provid, technolog, outsourc, big, open
         Lift: hadoop, process, provid, outsourc, will, reach, consum
         Score: hadoop, technolog, process, provid, reach, outsourc, will
Topic 4 Top Words:
         Highest Prob: decis, guid, open, day, valu, achiev, help
         FREX: decis, guid, open, day, valu, achiev, help
         Lift: guid, decis, day, open, achiev, valu, futur
         Score: decis, guid, day, achiev, open, valu, help
Topic 5 Top Words:
         Highest Prob: busi, cost, mission, provid, better, oper, challeng
         FREX: busi, cost, provid, oper, better, challeng, perform
         Lift: busi, cost, better, challeng, oper, provid, effici
         Score: busi, cost, challeng, better, oper, mission, provid
Topic 6 Top Words:
         Highest Prob: help, way, futur, innov, bring, becom, deliv
         FREX: help, way, futur, bring, becom, problem, success
         Lift: way, help, futur, bring, innov, becom, understand
         Score: help, way, futur, innov, bring, becom, one
Topic 7 Top Words:
         Highest Prob: compani, innov, one, bring, build, work, futur
         FREX: compani, innov, bring, futur, becom, serv, success
         Lift: innov, compani, one, bring, build, futur, work
         Score: compani, innov, one, bring, work, build, dedic
Topic 8 Top Words:
         Highest Prob: client, share, use, success, futur, integr, three
         FREX: client, share, use, global, success, futur, three
         Lift: share, client, use, success, three, consum, futur
```

Name: Ushnik Dasgupta
NYU ID: N17591222

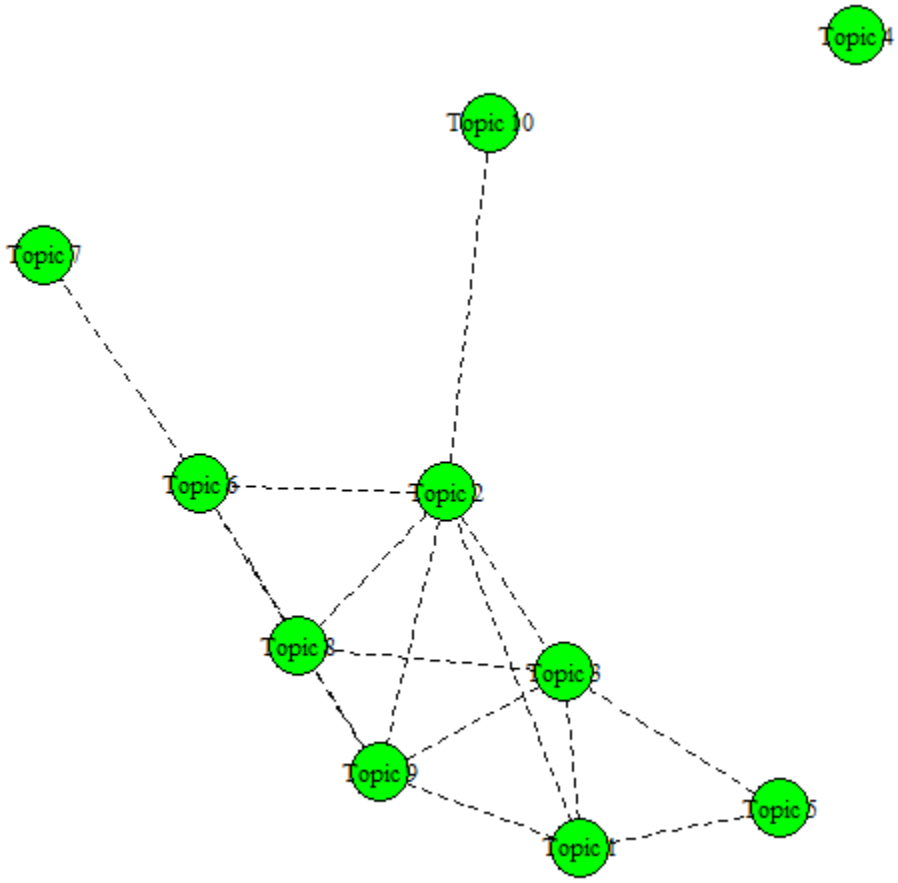
Probablity of top topics:



Having set a threshold of 70% for data analysis-readiness, a firm should include the highest probability words as displayed above from Topic 26 through Topic 6 (which approximately add up to 70%).

Name: Ushnik Dasgupta
NYU ID: N17591222

Topic Correlation Graph of first 10 companies:



Correlation of mission statements of the first 10 companies is as above. Mission statement of Company 4 is not related to any of the remaining 9.

Name: Ushnik Dasgupta
NYU ID: N17591222

Topic Modelling for Core Values:

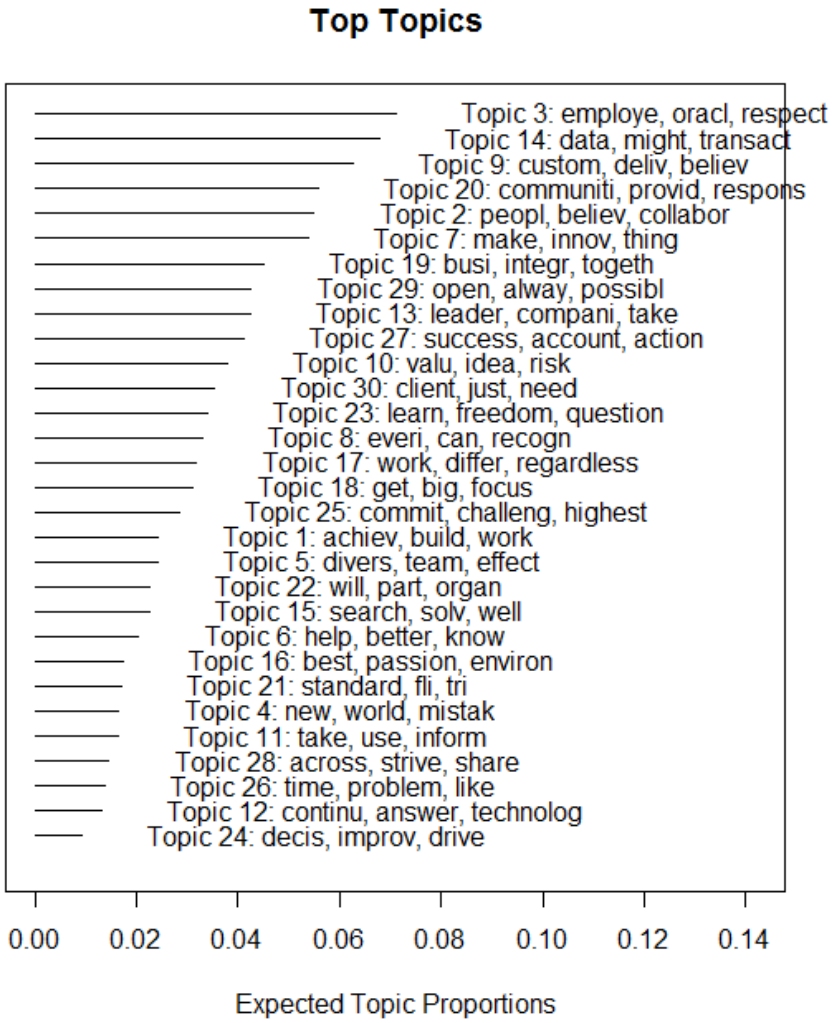
```
library(stm)
help("textProcessor")
temp<-textProcessor(documents=prcorpus$Core.Values, metadata = prcorpus)
names(temp)
meta<-temp$meta
vocab<-temp$vocab
docs<-temp$documents
out <- prepDocuments(docs, vocab, meta)
docs<-out$documents
vocab<-out$vocab
meta <-out$meta

prevfit <-stm(docs , vocab ,
              K=30,
              verbose=TRUE,
              data=meta,
              max.em.its=10)

topics <-labelTopics(prevfit , topics=c(1:30))
topics #shows topics with highest probability words

#explore the topics in context.  Provides an example of the text
help("findThoughts")
findThoughts(prevfit, texts = prcorpus$Core.Values,  topics = c(1:2),  n = 3)

help("plot.STM")
plot.STM(prevfit, type="summary")
plot.STM(prevfit, type="labels", topics=c(27,18,5))
plot.STM(prevfit, type="perspectives", topics = c(27,5))

# to aid on assigment of labels & intepretation of topics
help(topicCorr)
library(igraph)
mod.out.corr <- topicCorr(prevfit)  #Estimates a graph of topic correlations
plot.topicCorr(mod.out.corr)
```

Displaying high frequency words in core value statements of some of the companies:

```
Topic 1 Top Words:
         Highest Prob: chang, leverag, oper, understand, everyth, key, success
         FREX: chang, leverag, understand, oper, everyth, key, power
         Lift: chang, leverag, understand, everyth, key, oper, success
         Score: chang, leverag, key, oper, understand, everyth, success
Topic 2 Top Words:
         Highest Prob: technolog, solut, drive, everi, today, dedic, deliv
         FREX: solut, technolog, today, everi, drive, dedic, deliv
         Lift: solut, today, technolog, drive, everi, dedic, deliv
         Score: solut, technolog, today, everi, drive, dedic, deliv
Topic 3 Top Words:
         Highest Prob: hadoop, process, technolog, provid, outsourc, open, will
         FREX: hadoop, process, provid, technolog, outsourc, big, open
         Lift: hadoop, process, provid, outsourc, will, reach, consum
         Score: hadoop, technolog, process, provid, reach, outsourc, will
Topic 4 Top Words:
         Highest Prob: decis, guid, open, day, valu, achiev, help
         FREX: decis, guid, open, day, valu, achiev, help
         Lift: guid, decis, day, open, achiev, valu, futur
         Score: decis, guid, day, achiev, open, valu, help
Topic 5 Top Words:
         Highest Prob: busi, cost, mission, provid, better, oper, challeng
         FREX: busi, cost, provid, oper, better, challeng, perform
         Lift: busi, cost, better, challeng, oper, provid, effici
         Score: busi, cost, challeng, better, oper, mission, provid
Topic 6 Top Words:
         Highest Prob: help, way, futur, innov, bring, becom, deliv
         FREX: help, way, futur, bring, becom, problem, success
         Lift: way, help, futur, bring, innov, becom, understand
         Score: help, way, futur, innov, bring, becom, one
Topic 7 Top Words:
         Highest Prob: compani, innov, one, bring, build, work, futur
         FREX: compani, innov, bring, futur, becom, serv, success
         Lift: innov, compani, one, bring, build, futur, work
         Score: compani, innov, one, bring, work, build, dedic
Topic 8 Top Words:
         Highest Prob: client, share, use, success, futur, integr, three
         FREX: client, share, use, global, success, futur, three
         Lift: share, client, use, success, three, consum, futur
```

Name: Ushnik Dasgupta
NYU ID: N17591222

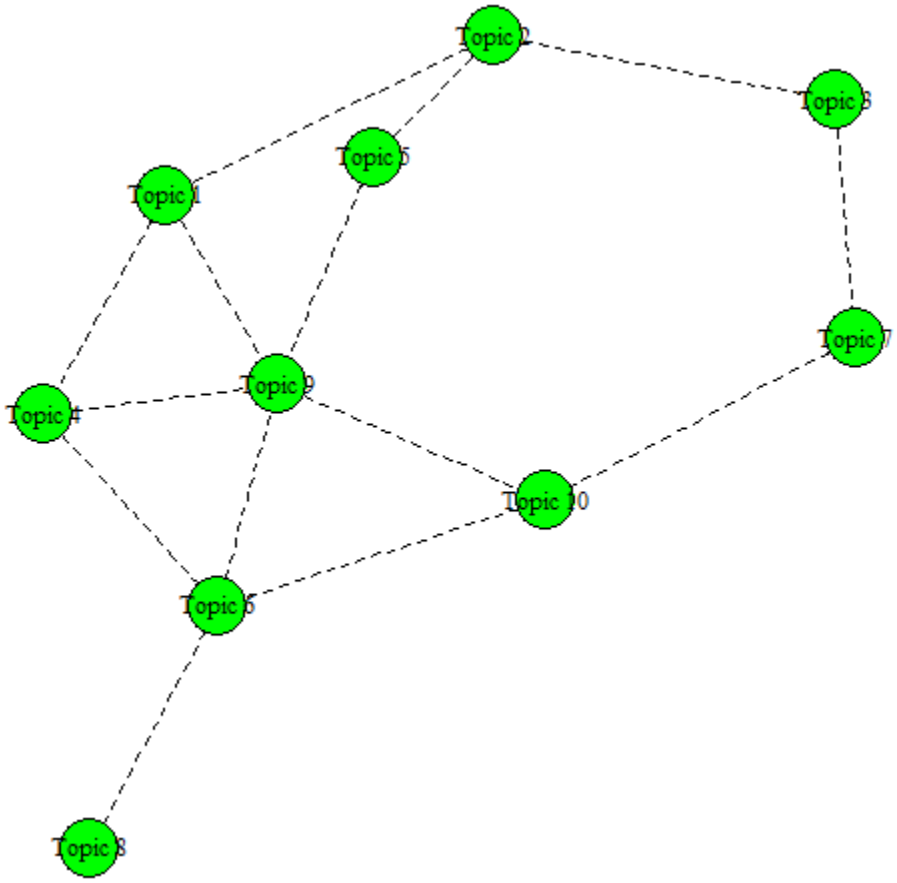
Probablity of top topics:



Having set a threshold of 70% for data analysis-readiness, a firm should include the highest probability words as displayed above from Topic 3 through Topic 27 (which approximately add up to 70%).

Name: Ushnik Dasgupta
NYU ID: N17591222

Topic Correlation Graph of Core Values of the first 10 companies:



I would recommend the CEO to analyze the topic correlations as well as the absence of any correlation (like the topics in mission statement of Company 4 when compared to the topics in mission statements of the first 10 companies) and check each of their business domains and models and figure out whether including words from their mission statement and core values would be a strategic fit.

Name: Ushnik Dasgupta
NYU ID: N17591222

PART 2: Analyzing Donald Trump's Speech Transcripts

The speeches analyzed here are:

- The "Immigration" Speech (August 31, 2016 in Phoenix)
- "Our Country Truly Needs a Great Leader" (June 16, 2016 in New York)
- "National Security" (September 7, 2016 in Philadelphia)

Creating a corpus and cleaning all the three speeches:

```
library(quanteda)
library(stm)
library(tm)
library(NLP)
library(openNLP)

prcorpus<-read.csv("C:/Users/Ushnik/Desktop/TrumpSpeechTranscript.csv", header = T, stringsAsFactors = F)
prcorpus <- na.omit(prcorpus)
View(prcorpus)

dim(prcorpus)
names(prcorpus)
head(prcorpus)
str(prcorpus)


speechcorpus <- corpus(prcorpus$Transcript, docnames = prcorpus$Sr..No., docvars = data.frame(prcorpus$Speech))
names(speechcorpus)
summary(speechcorpus)
head(speechcorpus)


speechcorpus<- toLower(speechcorpus, keepAcronyms = F)
speechcorpus
clean.corpus <- tokenize(speechcorpus,
                         removeNumbers = T,
                         removePunct = T,
                         removeSeparators = T,
                         removeHyphens = T,
                         removeSymbols = T,
                         verbose = T)
head(clean.corpus)
```

Filter

	Sr..No.	Speech	Transcript
1	Doc1	Immigration speech	Wow. Thank you. That's a lot of people, Phoenix, that'...
2	Doc2	Our Country Needs a Truly Great Leader	Wow. Whoa. That is some group of people. Thousand...
3	Doc3	National security in Philadelphia	Today, I am here to talk about three crucial words tha...

Name: Ushnik Dasgupta
NYU ID: N17591222

Creating a Document Freequency Matrix to return the 50 most frequently used words:

```
Console ~/
[2281] "and"          "inclusive"    "society"      "we"           "will"         "discard"
[2287] "the"          "failed"       "policies"     "and"          "division"     "of"
[2293] "the"          "past"         "and"          "embrace"      "true"         "american"
[2299] "change"       "to"           "rebuild"      "our"          "economy"      "rebuild"
[2305] "our"          "inner"        "cities"       "and"          "rebuild"      "our"
[2311] "country"      "we"           "will"         "bring"        "back"         "our"
[2317] "jobs"         "we"           "will"         "make"         "america"      "strong"
[2323] "again"        "we"           "will"         "make"         "america"      "safe"
[2329] "again"        "and"          "we"           "will"         "make"         "america"
[2335] "great"        "again"

> dfm.speech <- dfm(clean.corpus,
+               ignoredFeatures = stopwords("English"),
+               stem = T)
Creating a dfm from a tokenizedTexts object ...
   ... indexing documents: 3 documents
   ... indexing features: 2,192 feature types
   ... removed 146 features, from 174 supplied (glob) feature types
   ... stemming features (English), trimmed 473 feature variants
   ... created a 3 x 1573 sparse dfm
   ... complete.
Elapsed time: 0.08 seconds.
> top.features.speech <- topfeatures(dfm.speech, n=50)
> top.features.speech
    will       go    peopl  countri     know   immigr      one    great    trump      now     want     need
     151      111       98       79       66       57       54       53       52       49       49       46
     say    state      get  applaus     said     make       us      job  clinton     take     year     just
      46       43       40       39       38       38       38       36       36       36       34       33
    like      can    right     back    illeg   number    build american     come      new      law   happen
      32       32       32       31       31       31       30       29       29       28       28       28
    even     work   border     mani    think  hillari     unit  billion militari     good    china    world
      28       27       26       26       26       26       25       25       25       24       24       24
   offic   believ
      23       23
>
```

Creating a dictionary of irrelevant words and removing them from the corpus:

```
Console ~/
      32       32       32       31       31       31       30       29       29       28       28       28
    even     work   border     mani    think  hillari     unit  billion militari     good    china    world
      28       27       26       26       26       26       25       25       25       24       24       24
   offic   believ
      23       23
> word.list.speech <- c("will", "know", "say", "us", "take")
> dfm.speech.clean <- dfm(clean.corpus,
+                     ignoredFeatures = c(word.list.speech, stopwords("English")),
+                     stem = T)
Creating a dfm from a tokenizedTexts object ...
   ... indexing documents: 3 documents
   ... indexing features: 2,192 feature types
   ... removed 151 features, from 179 supplied (glob) feature types
   ... stemming features (English), trimmed 470 feature variants
   ... created a 3 x 1571 sparse dfm
   ... complete.
Elapsed time: 0.07 seconds.
>
> topfeatures.speech.clean<-topfeatures(dfm.speech.clean, n=50)
> topfeatures.speech.clean
      go    peopl  countri   immigr      one    great    trump      now     want     need    state      get
     111       98       79       57       54       53       52       49       49       46       43       40
 applaus     said     make      job  clinton     year     just     like      can    right     back    illeg
      39       38       38       36       36       34       33       32       32       32       31       31
  number    build american     come      new      law   happen     even     work   border     mani    think
      31       30       29       29       28       28       28       28       27       26       26       26
 hillari     unit  billion militari     good    china    world    offic   believ     also    enforc    obama
      26       25       25       25       24       24       24       23       23       22       22       22
million    thing
      22       22
>
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Checking for relevance with the word “obama”:

```
 [Doc3, 949]          and hillary [ clinton ] have also
 [Doc3, 998]          and hillary [ clinton ] defense spending
[Doc3, 1701]          under obama [ clinton ] our ballistic
[Doc3, 1799]            the obama [ clinton ] administration tried
[Doc3, 1923] capabilities hillary [ clinton ] has taught
[Doc3, 1964]  information hillary [ clinton ] put her
> kwic(clean.corpus, "obama", 5)
                                        contextPre keyword                                 contextPost
 [Doc1, 803]                  of the white house barack [   obama ] and hillary clinton weak weak
 [Doc1, 851]                   among the victims of the [   obama ] clinton open borders policy was
 [Doc1, 982]                      long who did not meet [   obama ] administration standards for removal and
[Doc1, 1570]       american citizens thank you president [   obama ] and hillary clinton have engaged
[Doc1, 1607]               arizona you know it president [   obama ] and hillary clinton support sanctuary
[Doc1, 2532]                  that we will terminate the [   obama ] administration's deadly and it is
[Doc1, 2576]                 that's over since alone the [   obama ] administration has allowed criminal aliens
[Doc1, 3874]           owe almost trillion doubled since [   obama ] took office our national debt
[Doc1, 4753]                     and never ever leave the [   obama ] clinton policy that's what it
[Doc2, 1324]                obamacare really kicks in in [   obama ] is going to be out
[Doc2, 1739]         cheerleader you know when president [   obama ] was elected i said well
[Doc2, 4021]              border with yemen you remember [   obama ] a year ago yemen was
 [Doc3, 502]                has no respect for president [   obama ] or hillary clinton sometimes it
 [Doc3, 763]                  what have the decisions of [   obama ] clinton produced libya is in
 [Doc3, 946]                but that's not all president [   obama ] and hillary clinton have also
 [Doc3, 995] unquestioned military strength under barack [   obama ] and hillary clinton defense spending
[Doc3, 1062]                  navy had ships when barack [   obama ] took office it had ships
[Doc3, 1136]      execute its current missions president [   obama ] plans to reduce the army
[Doc3, 1700]            art missile defense system under [   obama ] clinton our ballistic missile defense
[Doc3, 1798]                    and the middle east the [   obama ] clinton administration tried repeatedly to
>
```

Creating a new clean corpus of bigrams:

```
Source on Save                                                    Run   Source
54  clean.corpus.1 <- tokenize(speechcorpus,
55                             removeNumbers = T,
56                             removePunct = T,
57                             removeSeparators = T,
58                             removeHyphens = T,
59                             removeSymbols = T,
60                             verbose = T, ngrams = 2)
61  head(clean.corpus.1)
62
63  dfm.speech.1 <- dfm(clean.corpus.1,
64                      ignoredFeatures = stopwords("English"),
65                      stem = T)
66  top.features.speech.1 <- topfeatures(dfm.speech.1, n=50)
67  top.features.speech.1
68
69
58:45  (Top Level)                                                  R Script

Console ~/
> top.features.speech.1 <- topfeatures(dfm.speech.1, n=50)
> top.features.speech.1
   hillary_clinton          united_st      illegal_immigr       audience_memb        middle_east      criminal_alien
                26                 25                  19                  17                 11                   9
      trump_trump          open_bord     president_obama         law_enforc         want_trump         bring_back
                9                  8                   8                   8                  7                   7
        will_build          will_work            web_sit          net_worth     missile_defens immigration_system
                6                  6                   6                   6                  6                   5
  special_interest        sanctuary_c       obama_clinton           will_nev          number_on         right_peopl
                5                  5                   5                   5                  5                   5
     radical_islam     islamic_terror           trump_now       make_america           member_y        saudi_arabia
                5                  5                   5                   5                  5                   5
       nice_person          will_find      foreign_polici           will_ask   american_citizen          many_mani
                5                  5                   5                   5                  4                   4
  trump_administr           will_get           year_old          years_old          air_forc      billion_dollar
                4                  4                   4                   4                  4                   4
    american_work             one_th             day_on          even_know       common_sens     immigration_law
                4                  4                   4                   4                  4                   4
        right_now     tracking_system
                4                  4
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Checking for the relevance of bigram “hillary clinton”:

```
Console ~/
     radical_islam     islamic_terror          trump_now       make_america           member_y       saudi_arabia
                 5                  5                  5                  5                  5                  5
       nice_person          will_find     foreign_polici          will_ask   american_citizen         many_mani
                 5                  5                  5                  5                  4                  4
   trump_administr           will_get          year_old         years_old          air_forc     billion_dollar
                 4                  4                  4                  4                  4                  4
     american_work            one_th            day_on         even_know        common_sens    immigration_law
                 4                  4                  4                  4                  4                  4
         right_now    tracking_system
                 4                  4
> kwic(clean.corpus, "hillary clinton", 2)
                          contextPre         keyword           contextPost
  [Doc1, 758:759]         process called [ hillary clinton ] this includes
  [Doc1, 805:806]              obama and [ hillary clinton ] weak weak
[Doc1, 1475:1476]           close second [ hillary clinton ] for instance
[Doc1, 1572:1573]              obama and [ hillary clinton ] have engaged
[Doc1, 1609:1610]              obama and [ hillary clinton ] support sanctuary
[Doc1, 1643:1644]      executive amnesty [ hillary clinton ] has pledged
[Doc1, 1901:1902]              way while [ hillary clinton ] meets only
[Doc1, 2907:2908]              just like [ hillary clinton ] has evaded
[Doc1, 3223:3224]              so unfair [ hillary clinton ] has pledged
[Doc1, 3665:3666]          brutalized by [ hillary clinton ] brutalized homeland
[Doc1, 4239:4240]    countries secretary [ hillary clinton ] ignored this
[Doc1, 4533:4534]          never happens [ hillary clinton ] all talk
[Doc1, 5426:5427]           unfairly and [ hillary clinton ] is going
  [Doc3, 504:505]               obama or [ hillary clinton ] sometimes it
  [Doc3, 520:521]              east that [ hillary clinton ] didn't want
  [Doc3, 540:541]                 to war [ hillary clinton ] is just
  [Doc3, 698:699]              of before [ hillary clinton ] was sworn
  [Doc3, 948:949]              obama and [ hillary clinton ] have also
  [Doc3, 997:998]              obama and [ hillary clinton ] defense spending
[Doc3, 1922:1923]     cyber capabilities [ hillary clinton ] has taught
[Doc3, 1963:1964] classified information [ hillary clinton ] put her
>
```

Running a sentiment analysis:

Filter

	negative	positve
Doc1	5	26
Doc2	11	52
Doc3	0	3

Showing 1 to 3 of 3 entries

```
Console ~/
  [Doc3, 698:699]              of before [ hillary clinton ] was sworn
  [Doc3, 948:949]              obama and [ hillary clinton ] have also
  [Doc3, 997:998]              obama and [ hillary clinton ] defense spending
[Doc3, 1922:1923]     cyber capabilities [ hillary clinton ] has taught
[Doc3, 1963:1964] classified information [ hillary clinton ] put her
> speechdict<- dictionary(list(negative=c("detriment*", "bad*", "awful*", "terrib*", "horribl*,"),
+                          positve=c("good", "great", "super*", "excellent", "yay","vision",
+                                    "achieve","success")))
> speechdict
Dictionary object with 2 key entries.
 - negative: detriment*, bad*, awful*, terrib*, horribl*,
 - positve: good, great, super*, excellent, yay, vision, achieve, success
> speechsentiment<-dfm(clean.corpus,dictionary = speechdict)
Creating a dfm from a tokenizedTexts object ...
   ... indexing documents: 3 documents
   ... indexing features: 2,192 feature types
   ... applying a dictionary consisting of 2 keys
   ... created a 3 x 2 sparse dfm
   ... complete.
Elapsed time: 0.03 seconds.
> View(speechsentiment)
>
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Topic Modelling:

```
library(stm)
help("textProcessor")
temp<-textProcessor(documents=prcorpus$Transcript, metadata = prcorpus)
names(temp)
meta<-temp$meta
vocab<-temp$vocab
docs<-temp$documents
out <- prepDocuments(docs, vocab, meta)
docs<-out$documents
vocab<-out$vocab
meta <-out$meta

prevfit <-stm(docs , vocab ,
              K=3,
              verbose=TRUE,
              data=meta,
              max.em.its=3)

topics <-labelTopics(prevfit , topics=c(1:3))
topics #shows topics with highest probability words

#explore the topics in context.  Provides an example of the text
help("findThoughts")
findThoughts(prevfit, texts = prcorpus$Transcript,  topics = c(1:3),  n = 2)

help("plot.STM")
plot.STM(prevfit, type="summary")
plot.STM(prevfit, type="labels", topics=c(1,2,3))
plot.STM(prevfit, type="perspectives", topics = c(1,3))

# to aid on assigment of labels & intepretation of topics
help(topicCorr)
library(igraph)
mod.out.corr <- topicCorr(prevfit)  #Estimates a graph of topic correlations
plot.topicCorr(mod.out.corr)
```

Top words with highest frequency:

```
> topics <-labelTopics(prevfit , topics=c(1:3))
> topics
Topic 1 Top Words:
         Highest Prob: peopl, know, immigr, one, trump, say, state
         FREX: peopl, know, trump, one, say, immigr, said
         Lift: peopl, know, trump, say, said, one, get
         Score: peopl, know, immigr, trump, one, say, get
Topic 2 Top Words:
         Highest Prob: will, countri, great, now, need, applaus, just
         FREX: will, countri, great, need, now, applaus, can
         Lift: will, countri, great, need, now, applaus, can
         Score: will, countri, great, now, need, applaus, just
Topic 3 Top Words:
         Highest Prob: want, like, illeg, year, build, come, american
         FREX: want, illeg, american, hillari, year, clinton, like
         Lift: hillari, american, want, clinton, illeg, also, year
         Score: want, illeg, like, year, come, build, american
```

Name: Ushnik Dasgupta
NYU ID: N17591222

Top topics in each speech:

Topic 1:
peopl, countri, know, immigr, one, want, now, applaus, job, right, illeg, year, happen, work, mani, clinton, unit, believ, offic, million

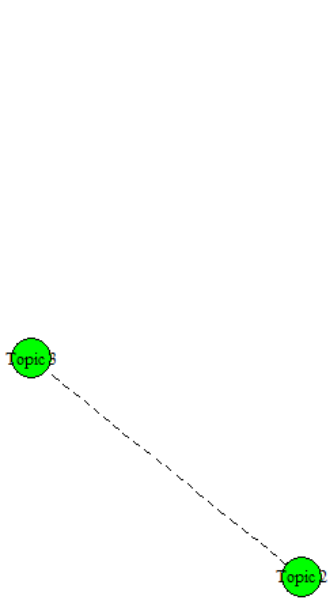
---

Topic 2:
great, need, state, take, just, can, number, law, even, american, militari, border, billion, also, world, realli, administr, never, bring, money

---

Topic 3:
will, trump, say, get, said, make, like, build, back, come, new, hillari, think, good, china, thing, enforc, got, time, love

Correlation of topics:



As from the topic correlation image alongside, Trump attributes his ability as a great leader (Speech 2: June 16) capable of restoring National security (Speech 3: September 7).

Name: Ushnik Dasgupta
NYU ID: N17591222

Report:

Donald Trump Presidential Campaign,

725 5th Ave, New York, NY 10022

Nov. 6, 2016

To: The Campaign Manager

From: UD

Subject: Mr. Trump's Linguistic Effectiveness

Mr. Trump has been able to connect to the American citizens with his speeches. Catchy slogans ("Make America Great Again"), short sentences and easy comprehension as opposed to the politically elite and difficult language used by other candidates, have been Mr. Trump's forte. He has been majorly positive in all his speeches most of which include important agendas like insufficient jobs, illegal immigration, Islamic terror, national security and the disorganized military and defense. Mr. Trump has been particularly able to address the people of the country as being great, and his potential and eagerness to guide the people of the country to unprecedented levels of economic and social stability. Furthermore, his act of reminding the American people that the nation needed a "Great Leader" through his speech on "National Security", almost two and a half months after the actual speech, was strategically brilliant.

Mr. Trump's approach towards professing his dissent against President Obama's initiatives and administration has been structured precisely to convey how differently he intends to tackle national and global issues. Mr. Trump has also spoken actively about Secretary Clinton and his profound feeling of her being a weak candidate for such a prestigious rank. He has charged her for illegally deleting classified information even after receiving a subpoena, and has identified the legal proceedings for it to be uncovered.

Lastly, Mr. Trump makes it very clear how America demands change, how he would work towards architecting a better America, and how Hillary Clinton would be just another four years of the not-so-good "Obama Rule".

Attachments.

DonaldTrumpSpeechAnalysis.R