# **BusinessAnalysis**

Arnav Ushnik Harshita November 23, 2016

#### PART 3

### Text Mining for Bill Gates and Melinda Gates Foundation data

In this document, we have done text mining on the nationwide campaign, #MyGivingStory and #Giving Tuesday conducted by the Bill and Melinda Gates Foundation. In the next part, we fit an LDA topic model to this topic. To fit the model, we use the R package Ida and we visualize the output using LDAvis. Using the visualization of LDAVis, we pick out the top topics and provide actionable insights for the next 2016 campaign.

```
library(quanteda)
## quanteda version 0.9.8.5
##
## Attaching package: 'quanteda'
  The following object is masked from 'package:base':
##
##
       sample
library(stm)
## stm v1.1.3 (2016-01-14) successfully loaded. See ?stm for help.
library(tm)
## Loading required package: NLP
## Attaching package: 'NLP'
## The following object is masked from 'package:quanteda':
##
##
       ngrams
##
## Attaching package: 'tm'
```

```
## The following objects are masked from 'package:quanteda':
##
## as.DocumentTermMatrix, stopwords
```

```
library(NLP)
library(openNLP)

story <- read.csv("/Users/arnavsomani/Desktop/NYU COURSE/sem 3/ba/r programming csv fil
es/story.csv")</pre>
```

str(story)

```
## 'data.frame':
                   823 obs. of 13 variables:
## $ Inclusion
                         : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ votes total
                        : int 33 1029 20 7 15 29 2 54 66 6 ...
## $ votes_gallery_.Calc.: int 4 72 2 2 1 3 0 27 22 1 ...
## $ votes local network : int 29 957 18 5 14 26 2 27 44 5 ...
## $ createdate
                        : Factor w/ 803 levels "11/1/15 14:32",..: 301 747 726 378 446
135 21 567 622 458 ...
## $ ipaddress
                        : Factor w/ 777 levels "100.15.14.250",..: 484 27 151 273 258
40 16 119 134 382 ...
## $ EIN
                        : Factor w/ 596 levels "01-0360745", "01-0450352",..: 88 225 59
0 108 108 24 1 232 554 274 ...
                         : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ Photo
## $ URL in Story .Calc. : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ matched zip
                   : Factor w/ 556 levels "","0","01331",..: 545 278 463 79 79 80
244 327 556 317 ...
## $ Story Text
                        : Factor w/ 810 levels "
                                                                                 Cure
PSP.org\\n \\nMy husband's diagnosis of PSP came as a terrible shock.\\nFor a number
"| truncated ,..: 512 120 30 781 734 638 248 84 92 94 ...
## $ give reason
                        : int 3 1 2 2 2 1 2 2 2 1 ...
## $ give method
                        : int 2 2 2 6 4 2 2 4 2 2 ...
```

```
storytext<-as.character(story$Story_Text)
class(storytext)</pre>
```

```
## [1] "character"
```

```
library(quanteda)
library(stm)
library(tm)
library(NLP)
library(openNLP)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

## ##

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

library(ggdendro)

##
## Attaching package: 'ggdendro'
```

```
library(cluster)
library(fpc)
```

## The following object is masked from 'package:quanteda':

## **Preprocessing**

segment

Creating a corpus with metadata

```
newscorpus<- corpus(storytext)
```

Exploring the corpus

```
names(newscorpus)
```

```
## [1] "documents" "metadata" "settings" "tokens"
```

summary(newscorpus)

```
## Corpus consisting of 823 documents, showing 100 documents.
##
##
       Text Types Tokens Sentences
##
                306
                        797
                                    21
      text1
##
      text2
                380
                      1072
                                    35
##
      text3
                216
                        413
                                    16
##
      text4
                214
                        409
                                    17
##
                        280
                                     8
      text5
                164
                                    37
##
      text6
                422
                      1043
##
      text7
                324
                        696
                                    22
##
                                    11
                156
                        272
      text8
##
                234
                        4\,7\,1
                                     6
      text9
                                     9
##
     text10
                123
                        208
##
     text11
                269
                        502
                                    23
##
     text12
                322
                        691
                                    19
                                     8
##
     text13
                109
                        181
##
                253
                        475
                                    14
     text14
##
                                    10
     text15
                140
                        276
##
                146
                        245
                                     8
     text16
##
                320
                        785
                                    36
     text17
##
                280
                                    15
     text18
                        615
##
                        228
                                    15
     text19
                147
##
                        271
                                    12
     text20
                158
                                    25
##
     text21
                268
                        621
##
     text22
                97
                        162
                                     5
                                    10
##
     text23
                155
                        326
##
     text24
                304
                        704
                                    24
##
     text25
                375
                        974
                                    31
##
                                    11
     text26
                258
                        542
##
     text27
                        280
                                     9
                161
                                    17
##
     text28
                234
                        447
##
     text29
                281
                        647
                                    15
##
     text30
                131
                        210
                                     8
                                     5
##
     text31
                76
                        122
##
     text32
                53
                        75
                                     5
##
     text33
                339
                        793
                                    31
                                     8
##
     text34
                106
                       180
##
     text35
                227
                        433
                                    13
##
                                    23
     text36
                265
                        525
                                     8
##
     text37
                173
                        312
##
     text38
                145
                        273
                                     6
##
     text39
                145
                        273
                                     6
                                    14
##
     text40
                148
                        243
##
     text41
                230
                        424
                                    20
##
     text42
                210
                        414
                                     6
                                     5
##
                       505
     text43
                235
                                     5
##
     text44
                        313
                164
##
     text45
                152
                        268
                                    16
##
     text46
                232
                       425
                                    18
##
     text47
                133
                        222
                                     7
##
                 39
                         61
                                     2
     text48
                                     2
##
     text49
                 43
                        86
##
     text50
                240
                        574
                                    21
```

```
##
               479
                      1278
                                    35
     text51
##
     text52
               139
                       226
                                     4
                                    46
##
               347
                       967
     text53
##
     text54
               360
                       828
                                    31
##
     text55
               413
                       896
                                    38
##
                                    10
     text56
               148
                       259
##
     text57
               171
                       315
                                    18
##
     text58
               209
                       396
                                    17
##
               376
                       846
                                    33
     text59
##
     text60
               127
                       228
                                    12
##
               214
                       394
                                    19
     text61
##
               203
                       403
                                    24
     text62
##
                                    36
     text63
               262
                       567
##
                                     9
               180
                       356
     text64
                                     9
##
     text65
               158
                       260
##
     text66
               148
                       250
                                    13
##
               244
                       525
                                    12
     text67
##
                       397
                                    13
     text68
               191
##
               140
                       207
                                    11
     text69
##
               156
                       257
                                    12
     text70
##
                                     9
                       429
     text71
               215
##
     text72
               193
                       426
                                     9
                                    27
##
                       556
     text73
               265
##
                                    12
     text74
               213
                       409
                                     9
##
     text75
               168
                       309
##
     text76
               304
                       722
                                    14
##
     text77
               140
                       258
                                    10
##
                                    10
     text78
                79
                       131
                                     6
##
     text79
               154
                       262
                                     5
##
     text80
                 54
                         76
                                     3
##
     text81
                         95
                 66
##
     text82
               178
                       351
                                    13
##
     text83
               240
                       499
                                    13
##
     text84
                                     6
               102
                       169
                                    31
##
     text85
               289
                       593
##
     text86
               117
                       194
                                     9
##
     text87
                91
                       150
                                     7
                                     9
##
                       303
     text88
               173
                                    34
##
     text89
               433
                      1108
                                     9
##
     text90
               162
                       302
                                     5
##
     text91
                70
                       102
##
     text92
               165
                       314
                                    11
##
                       171
                                     7
     text93
               112
##
     text94
               289
                       614
                                    26
##
     text95
               229
                       446
                                    12
                                    13
##
     text96
               173
                       310
##
                                    35
     text97
               399
                       961
##
                                    13
     text98
               184
                       355
##
                                    18
     text99
               336
                       620
##
    text100
               284
                       586
                                    28
##
             /Users/arnavsomani/Desktop/* on x86_64 by arnavsomani
## Source:
## Created: Wed Nov 23 16:59:41 2016
## Notes:
```

head(newscorpus,1)

##

#### text1

## "My Giving Story for 25:40 \\n(www.2540.org)\\n\\nFor many years I went through life not truly seeing the world and its problems. I knew there were people living in povert y; however it was easy to put those thoughts out of my head with the distractions of my daily routine. A couple of years ago, God removed the blinders from my eyes, and allow ed me to see how truly blessed my family and many other families here in America are com pared to other parts of the world. In the process, I was astonished to discover that ab out 1.5 billion people in our world live on less than \$1.25 per day. How does someone e ven live on \$1.25 per day, when many of us here in America are willing to pay more than \$1.25 for a cup of coffee? \\n\\nAfter God began revealing this to my husband and me, we felt called to do something to help. I just could not sit back anymore and preten

d I didn't know that others were living in such dire circumstances, and I truly wanted to do something to give back. My husband and I were led to an organization called 25:4 0 which helps orphans and underprivileged children in South Africa. We were drawn to th is organization, first and foremost because of the organization's name 25:40, drawn fr om Matthew 25:40 "Truly I tell you, whatever you did for one of the least of these bro thers and sisters of mine, you did for me.â€\u009d After hearing stories about many of the children they serve and how they were partnering with the community to help fix the problems with sustainable solutions, I felt compelled to join efforts with them and do something to help raise money for the least of these through this wonderful organizatio \\n\\nBeing a dancer for 18 years, I have much passion and love for dancing. Go d opened up a door to dance ministry a couple of years ago and through some prayer I dec ided to do a dance workshop over one summer to help benefit 25:40. It was such a blessi ng for me to be involved with dance again and at the same time such an amazing opportuni ty for me to use my time, talents, and passions to help these children in South Africa w ho so desperately needed our help. \\n\\nDuring the workshop I was also able to talk about some of the children that 25:40 serves. How some of these children lived in smal 1 huts with no floors or had to walk every day long distances to collect water for their The children would ask me if I was making these stories up since it was so har d for them to even imagine life as these children in South Africa know it. It was very heartwarming to see some of the children come to the workshop, emptying their own piggy banks to give back after hearing the stories of the children 25:40 serves. In addition to the money we were able to raise for 25:40 during the workshop, I know the workshop m ake a great impact on those children's hearts and also mine as well. ery blessed to have had the opportunity to give back and help children in South Africa t hrough 25:40. I learned through this giving experience that God is calling us all to do something, and in the process to be the hands and feet of Jesus. We can have such an a mazing impact on this world if we are mindful of what is going on within it and make the commitment that we will do what we can to help. I believe that when we give of our tim e, talents, and treasures we can also be a part of something much bigger than ourselves! And as we join forces with God to do something to help others, I think we will be amaz ed to see that not only will other's lives be helped and impacted because of it, but o ur own life will be changed as well as our hearts are transformed to be more like His. I know mine has and it is a beautiful thing! \\n"

Cleaning corpus: removes punctuation, digits, converts to lower case

```
## ...preserving Twitter characters (#, @)

## ...total elapsed: 0.0020000000000007 seconds.
```

## Starting tokenization...

## ...tokenizing texts

## ...total elapsed: 0.1329999999999 seconds.

## ...replacing Twitter characters (#, 0)

## ...total elapsed: 0.03599999999999 seconds.

## ...replacing names

## ...total elapsed: 0.00099999999999446 seconds.

## Finished tokenizing and cleaning 823 texts.

#### Exploring the clean corpus

head(cleancorpus,1)

```
## $text1
     [1] "my"
                              "giving"
##
                                                 "story"
                              "n"
##
     [4] "for"
                                                 "www"
                              "n"
                                                 "nfor"
     [7] "org"
##
                                                 "i"
##
    [10] "many"
                              "years"
                                                 "life"
##
    [13] "went"
                              "through"
    [16] "not"
                              "truly"
                                                 "seeing"
##
                                                 "and"
##
    [19] "the"
                              "world"
                                                 "i"
    [22] "its"
                              "problems"
##
##
    [25] "knew"
                              "there"
                                                 "were"
                                                 "in"
    [28] "people"
                              "living"
##
                                                 "it"
##
    [31] "poverty"
                             "however"
                              "easy"
                                                 "to"
##
   [34] "was"
##
    [37] "put"
                              "those"
                                                 "thoughts"
    [40] "out"
                              "of"
                                                 "my"
##
##
    [43] "head"
                              "with"
                                                 "the"
                              "of"
##
    [46] "distractions"
                                                 "my"
                                                 "a"
##
   [49] "daily"
                              "routine"
                              "of"
##
   [52] "couple"
                                                 "years"
   [55] "ago"
                              "god"
                                                 "removed"
##
##
   [58] "the"
                              "blinders"
                                                 "from"
                              "eyes"
                                                 "and"
    [61] "my"
##
                              "me"
                                                 "to"
##
    [64] "allowed"
    [67] "see"
                              "how"
                                                 "truly"
##
##
    [70] "blessed"
                              "my"
                                                 "family"
    [73] "and"
                              "many"
                                                 "other"
##
                              "here"
    [76] "families"
                                                 "in"
##
                              "are"
##
   [79] "america"
                                                 "compared"
   [82] "to"
                              "other"
                                                 "parts"
##
   [85] "of"
                             "the"
                                                 "world"
##
                              "the"
##
   [88]
         "in"
                                                 "process"
   [91] "i"
                              "was"
                                                 "astonished"
##
##
   [94] "to"
                              "discover"
                                                 "that"
                             "billion"
   [97] "about"
                                                 "people"
##
## [100] "in"
                              "our"
                                                 "world"
## [103] "live"
                              "on"
                                                 "less"
                             "per"
                                                 "day"
## [106] "than"
## [109] "how"
                              "does"
                                                 "someone"
## [112] "even"
                             "live"
                                                 "on"
                             "day"
                                                 "when"
## [115] "per"
## [118] "many"
                             "of"
                                                 "us"
                              "in"
## [121] "here"
                                                 "america"
                                                 "to"
## [124] "are"
                              "willing"
## [127] "pay"
                              "more"
                                                 "than"
                              "a"
## [130] "for"
                                                 "cup"
                                                 "n"
## [133] "of"
                              "coffee"
                              "god"
                                                 "began"
## [136] "nafter"
                             "this"
## [139] "revealing"
                                                 "to"
                                                 "and"
## [142] "my"
                              "husband"
## [145] "me"
                              "we"
                                                 "felt"
## [148] "called"
                              "to"
                                                 "do"
## [151] "something"
                              "to"
                                                 "help"
## [154] "i"
                              "just"
                                                 "could"
```

```
## [157] "not"
                             "sit"
                                                 "back"
## [160] "anymore"
                             "and"
                                                 "pretend"
                                                 "t"
## [163] "i"
                             "didnâ"
## [166] "know"
                             "that"
                                                 "others"
                             "living"
                                                 "in"
## [169] "were"
                             "dire"
## [172] "such"
                                                 "circumstances"
                             "i"
## [175] "and"
                                                 "truly"
## [178] "wanted"
                             "to"
                                                 "do"
                             "to"
                                                 "give"
## [181] "something"
## [184] "back"
                             "my"
                                                 "husband"
## [187] "and"
                             "i"
                                                 "were"
## [190] "led"
                             "to"
                                                 "an"
## [193] "organization"
                             "called"
                                                 "which"
## [196] "helps"
                                                 "and"
                             "orphans"
## [199] "underprivileged" "children"
                                                 "in"
## [202] "south"
                             "africa"
                                                 "we"
## [205] "were"
                             "drawn"
                                                 "to"
                                                 "first"
## [208] "this"
                             "organization"
## [211] "and"
                             "foremost"
                                                 "because"
## [214] "of"
                             "the"
                                                 "organizationâ"
## [217] "s"
                             "name"
                                                 "drawn"
## [220] "from"
                             "matthew"
                                                 "â"
                             "i"
## [223] "@truly"
                                                 "tell"
## [226] "you"
                             "whatever"
                                                 "you"
                             "for"
## [229] "did"
                                                 "one"
## [232] "of"
                             "the"
                                                 "least"
## [235] "of"
                                                 "brothers"
                             "these"
                                                 "of"
## [238] "and"
                             "sisters"
                                                 "did"
## [241] "mine"
                             "you"
## [244] "for"
                             "me.â"
                                                 "after"
## [247] "hearing"
                             "stories"
                                                 "about"
                             "of"
                                                 "the"
## [250] "many"
## [253] "children"
                             "they"
                                                 "serve"
## [256] "and"
                             "how"
                                                 "they"
## [259] "were"
                             "partnering"
                                                 "with"
## [262] "the"
                                                 "to"
                             "community"
## [265] "help"
                             "fix"
                                                 "the"
## [268] "problems"
                             "with"
                                                 "sustainable"
                             "i"
                                                 "felt"
## [271] "solutions"
## [274] "compelled"
                             "to"
                                                 "join"
                             "with"
## [277] "efforts"
                                                 "them"
## [280] "and"
                             "do"
                                                 "something"
## [283] "to"
                             "help"
                                                 "raise"
## [286] "money"
                             "for"
                                                 "the"
                                                 "these"
## [289] "least"
                             "of"
                             "this"
## [292] "through"
                                                 "wonderful"
                             "n"
## [295] "organization"
                                                 "nbeing"
                                                 "for"
## [298] "a"
                             "dancer"
                             "i"
## [301] "years"
                                                 "have"
## [304] "much"
                             "passion"
                                                 "and"
## [307] "love"
                             "for"
                                                 "dancing"
## [310] "god"
                             "opened"
                                                 "up"
## [313] "a"
                             "door"
                                                 "to"
## [316] "dance"
                             "ministry"
```

			,
## [319]	"couple"	"of"	"years"
## [322]	"ago"	"and"	"through"
## [325]	"some"	"prayer"	"i"
## [328]	"decided"	"to"	"do"
## [331]	"a"	"dance"	"workshop"
## [334]	"over"	"one"	"summer"
## [337]	"to"	"help"	"benefit"
## [340]	"it"	"was"	"such"
## [343]	"a"	"blessing"	"for"
## [346]	"me"	"to"	"be"
## [349]	"involved"	"with"	"dance"
## [352]	"again"	"and"	"at"
## [355]	"the"	"same"	"time"
## [358]	"such"	"an"	"amazing"
## [361]	"opportunity"	"for"	"me"
## [364]	"to"	"use"	"my"
## [367]	"time"	"talents"	"and"
## [370]	"passions"	"to"	"help"
## [373]	"these"	"children"	"in"
## [376]	"south"	"africa"	"who"
## [379]	"so"	"desperately"	"needed"
## [382]	"our"	"help"	"n"
## [385]	"nduring"	"the"	"workshop"
## [388]	"i"	"was"	"also"
## [391]	"able"	"to"	"talk"
## [394]	"about"	"some"	"of"
## [397]	"the"	"children"	"that"
## [400]	"serves"	"how"	"some"
## [403]	"of"	"these"	"children"
## [406]	"lived"	"in"	"small"
## [409]	"huts"	"with"	"no"
## [412]	"floors"	"or"	"had"
## [415]	"to"	"walk"	"every"
## [418]	"day"	"long"	"distances"
## [421]	"to"	"collect"	"water"
## [424]	"for"	"their"	"family"
## [427]	"the"	"children"	"would"
## [430]	"ask"	"me"	"if"
## [433]	"i"	"was"	"making"
## [436]	"these"	"stories"	"up"
## [439]	"since"	"it"	"was"
## [442]	"so"	"hard"	"for"
## [445]		"to"	"even"
## [448]	"imagine"	"life"	"as"
## [451]	"these"	"children"	"in"
## [454]	"south"	"africa"	"know"
## [457]	"it"	"it"	"was"
## [460]	"very"	"heartwarming"	"to"
## [463]	"see"	"some"	"of"
## [466]		"children"	"come"
## [469]		"the"	"workshop"
## [472]	"emptying"	"their"	"own"
## [475]	"piggy"	"banks"	"to"
## [478]	"give"	"back"	"after"

```
"the"
## [481] "hearing"
                                                 "stories"
## [484] "of"
                             "the"
                                                 "children"
                             "in"
                                                 "addition"
## [487] "serves"
## [490] "to"
                             "the"
                                                 "money"
                             "were"
                                                 "able"
## [493] "we"
## [496] "to"
                                                 "for"
                             "raise"
## [499] "during"
                             "the"
                                                 "workshop"
## [502] "i"
                             "know"
                                                 "the"
## [505] "workshop"
                             "make"
                                                 "a"
## [508] "great"
                             "impact"
                                                 "on"
                                                 "s"
## [511] "those"
                             "childrenâ"
## [514] "hearts"
                             "and"
                                                 "also"
## [517] "mine"
                             "as"
                                                 "well"
                             "ni"
                                                 "feel"
## [520] "n"
                             "blessed"
## [523] "very"
                                                 "to"
                                                 "the"
## [526] "have"
                             "had"
## [529] "opportunity"
                             "to"
                                                 "give"
## [532] "back"
                             "and"
                                                 "help"
## [535] "children"
                             "in"
                                                 "south"
                                                 "i"
## [538] "africa"
                             "through"
                                                 "this"
## [541] "learned"
                             "through"
## [544] "giving"
                             "experience"
                                                 "that"
                             "is"
## [547] "god"
                                                 "calling"
                             "all"
                                                 "to"
## [550] "us"
                                                 "and"
## [553] "do"
                             "something"
## [556] "in"
                             "the"
                                                 "process"
## [559] "to"
                             "be"
                                                 "the"
                                                 "feet"
                             "and"
## [562] "hands"
                             "jesus"
## [565] "of"
                                                 "we"
## [568] "can"
                             "have"
                                                 "such"
## [571] "an"
                                                 "impact"
                             "amazing"
## [574] "on"
                             "this"
                                                 "world"
## [577] "if"
                             "we"
                                                 "are"
## [580] "mindful"
                             "of"
                                                 "what"
                                                 "on"
## [583] "is"
                             "going"
## [586] "within"
                             "it"
                                                 "and"
## [589] "make"
                             "the"
                                                 "commitment"
## [592] "that"
                             "we"
                                                 "will"
                                                 "we"
## [595] "do"
                             "what"
                             "to"
## [598] "can"
                                                 "help"
                             "believe"
## [601] "i"
                                                 "that"
## [604] "when"
                             "we"
                                                 "give"
                             "our"
## [607] "of"
                                                 "time"
## [610] "talents"
                             "and"
                                                 "treasures"
## [613] "we"
                             "can"
                                                 "also"
                             "a"
## [616] "be"
                                                 "part"
## [619] "of"
                                                 "much"
                             "something"
## [622] "bigger"
                             "than"
                                                 "ourselves"
                             "as"
                                                 "we"
## [625] "and"
## [628] "join"
                             "forces"
                                                 "with"
## [631] "god"
                             "to"
                                                 "do"
## [634] "something"
                             "to"
                                                 "help"
                             "i"
## [637] "others"
                                                 "think"
                             "will"
                                                 "be"
## [640] "we"
```

```
## [643] "amazed"
                             "to"
                                                "see"
## [646] "that"
                             "not"
                                                "only"
                                                "s"
## [649] "will"
                             "otherâ"
                             "be"
## [652] "lives"
                                                "helped"
                             "impacted"
## [655] "and"
                                                "because"
                             "it"
                                                "but"
## [658] "of"
                             "own"
                                                "life"
## [661] "our"
                             "be"
## [664] "will"
                                                "changed"
## [667] "as"
                             "well"
                                                "as"
## [670] "our"
                             "hearts"
                                                "are"
## [673] "transformed"
                             "to"
                                                "be"
## [676] "more"
                             "like"
                                                "his"
## [679] "i"
                             "know"
                                                "mine"
## [682] "has"
                             "and"
                                                "it"
                             "a"
## [685] "is"
                                                "beautiful"
                             "n"
## [688] "thing"
```

#### creating document feature matrix from clean corpus

... lowercasing

##

```
##
## ... indexing documents: 823 documents
```

```
##
## ... indexing features:
```

```
## 17,103 feature types
```

```
##
```

```
## •••
```

```
## removed 161 features, from 174 supplied (glob) feature types
```

```
## ... created a 823 x 16942 sparse dfm
## ... complete.
## Elapsed time: 0.682 seconds.
```

```
head(dfm.simple,3)
```

```
## Document-feature matrix of: 823 documents, 16,942 features.
## (showing first 3 documents and first 6 features)
##
          features
## docs
           giving story n www org nfor
##
     text1
                2
                      1
                         7
                                  1
                0
                                  0
##
     text2
                      1 11
                              0
                                       0
##
     text3
                      1 13
                                       0
                0
```

#### Displaying most frequent terms in dfm

```
topfeatures<-topfeatures(dfm.simple, n=200)
topfeatures</pre>
```

##	n	give	â	time	life
##	3148	878	872	869	866
##	one	help	people	children	can
##	841	832	781	758	735
##	s	years	many	community	will
##	679	669	653	636	628
##	support	ni	family	day	like
##	591	562	555	552	535
##	work	just	school	us	back
##	529	528	526	523	515
##	world	year	first	giving	every
##	487	466	448	446	446
##	know	make	organization	love	t
##	444	443	436	426	424
##	way	get	also	much	new
##	414	395	385	374	370
##	now	others	home	families	need
##	360	358	354	353	348
##	even	see	lives	never	able
##	345	342	341	340	334
##	kids	hope	program	foundation	made
##	316	308	302	298	296
##	care	education	money	students	cancer
##	296	291	286	286	272
##	go	child	something	part	volunteer
##	269	263	258	257	256
##	experience	heart	girls	little	young
##	254	252	251	250	250
##	two	always	food	feel	well
##	246	245	241	236	234
##	started	friends	story	around	local
##	233	232	228	228	228
##	take	wanted	good	find	found
##	224	220	219	218	217
##	old	women	going	better	place
##	215	213	212	212	210
##	want	great	knew	came	nthe
##	209	207	206	204	203
##	best	provide	health	needed	change
##	201	200	200	198	198
##	opportunity	come	amazing	group	college
##	197	196	192	192	192
##	since	live	helping	parents	working
##	191	190	187	187	182
##	got	small	believe	may	went
##	180	179	178	177	175
##	away 175	become 175	someone	together 173	mission
##			173		172
##	volunteers	center	free	things	months
##	171	171	171	170	170
##	still 170	began 169	gave 169	told 168	high 168
##	continue	inspired	felt	son	important
π#	Concinue	THEPTIER	TETC	2011	Impor cant

23/2010			Businessi marysis	
## 168	168	167	167	166
## project	mother	needs	difference	ago
## 165	162	162	161	160
## long	without	last	learned	given
## 159	159	159	158	158
## run	learn	mom	job	living
## 158	157	154	153	152
## programs	hospital	hard	donate	iâ
## 151	149	147	147	147
## making	staff	became	medical	think
## 146	146	146	145	144
## share	lost	days	next	took
## 142	142	142	140	140
## different	person	another	nmy	three
## 139	139	139	138	138
## services	truly	walk	later	called
## 137	136	136	136	135
## disease	animals	future	age	big
## 135	134	132	132	132
## brain	met	ever	husband	helped
## 132	130	130	129	129
## often	keep	right	everyone	put
## 129	129	129	128	127
## worked	thought	house	really	nicu
## 126	125	125	124	124
## impact	volunteering	hours	youth	asked
## 123	123	122	122	122

#### Creating a custom dictionary list of stop words

... indexing features:

##

```
## 17,103 feature types
```

##

```
## ...
```

```
## removed 169 features, from 182 supplied (glob) feature types
```

```
## ... created a 823 x 16934 sparse dfm
## ... complete.
## Elapsed time: 0.509 seconds.
```

```
topfeatures.stem<-topfeatures(dfm.stem, n=50)
topfeatures.stem</pre>
```

##	give	time	life	one	help	
##	878	869	866	841	832	
##	people	children	can	years	many	
##	781	758	735	669	653	
##	community	will	support	family	day	
##	636	628	591	555	552	
##	like	work	school	us	back	
##	535	529	526	523	515	
##	world	year	first	giving	every	
##	487	466	448	446	446	
##	know	make	organization	love	way	
##	444	443	436	426	414	
##	get	also	much	new	now	
##	395	385	374	370	360	
##	others	home	families	need	even	
##	358	354	353	348	345	
##	see	lives	never	able	kids	
##	342	341	340	334	316	
##	hope	program	foundation	made	care	
##	308	302	298	296	296	

#### Creating a Dfm with bigrams

```
## Starting tokenization...
```

```
## ...preserving Twitter characters (#, @)
```

```
## ...total elapsed: 0.00099999999999446 seconds.
```

```
## ...tokenizing texts
## ...total elapsed: 0.17500000000001 seconds.
## ...replacing Twitter characters (#, 0)
## ...total elapsed: 0.030000000000011 seconds.
## ...creating ngrams
## ...total elapsed: 1.851 seconds.
## ...replacing names
## ...total elapsed: 0.0010000000000122 seconds.
## Finished tokenizing and cleaning 823 texts.
dfm.bigram<- dfm(cleancorpus, toLower = TRUE,</pre>
                 ignoredFeatures = c(swlist, stopwords("english")),
                 verbose=TRUE,
                 stem=FALSE)
##
##
      ... lowercasing
##
      ... indexing documents: 823 documents
##
##
##
      ... indexing features:
## 124,676 feature types
##
## ...
## removed 81,058 features, from 182 supplied (glob) feature types
```

11/23/2016

```
topfeatures.bigram<-topfeatures(dfm.bigram, n=50)
topfeatures.bigram</pre>
```

```
##
                  give_back
                                                                      high_school
                                             years_ago
##
                                                                               102
                         154
                                                    125
##
                   every day
                                                                              iâ m
                                             years old
##
                          88
                                                     67
                                                                                67
                   year_old
##
                                               one day
                                                                      giving back
##
                          63
                                                     55
                                                                                53
##
                 first_time
                                           help_others
                                                                            iâ_ve
##
                          51
                                                     49
                                                                                49
##
                 united way
                                           many_people
                                                                      even though
##
                          47
                                                                                42
##
                  two_years
                                         will_continue
                                                                         can_help
##
                          40
                                                                                38
##
                    can make
                                        nicu_community
                                                                    united_states
##
##
                 many_years
                                           years_later
                                                                        make_sure
##
                                                     33
                                                                                32
              neonatal womb
##
                                           raise money
                                                                  brain aneurysm
                          32
                                                     30
##
                                                                                30
##
                 every year
                                             last year
                                                                   mental health
##
                          29
                                                     29
                                                                                28
##
               social media
                                         oxford county
                                                                     giving story
                                                     28
                                                                                27
##
##
                help people
                                             felt like
                                                                      health care
##
                          27
                                                     27
                                                                                26
              special needs
                                                                     someone else
##
                                          every single
##
                          26
                                                                                25
                                                     26
##
                  feel like
                                            will never
                                                                       free minds
                          25
##
##
            wildlife center
                                          better place non-profit organization
##
             giving tuesday
                                            every time
##
                                                                        young man
##
                          24
                                                                                24
##
              middle school
                               nonprofit organization
##
                          2.4
```

#### Sentiment Analysis

```
##
##
      ... indexing documents: 823 documents
##
##
      ... indexing features:
## 124,676 feature types
##
## ...
## applying a dictionary consisting of 2 keys
##
      ... created a 823 x 2 sparse dfm
##
      ... complete.
## Elapsed time: 0.664 seconds.
topfeatures(dfm.sentiment)
##
    postive negative
##
       2846
                 105
```

## **WORD CLOUD**

Craeting a word cloud that displays the high frequency terms

```
library(wordcloud)

## Loading required package: RColorBrewer
```



Checking for correlation of specific words with limits of 0.5 and 0.3

```
## $giving
## numeric(0)
##
##
  $childern
   numeric(0)
##
   $help
##
##
   numeric(0)
##
##
   $community
##
                  teachings
                                               mariner
                                                                            jamal
##
                        0.58
                                                   0.58
                                                                             0.58
##
                      rahman
                                              dr.larry
                                                                           dossey
                        0.58
##
                                                   0.58
                                                                             0.58
                      bernie
##
                                                siegel
                                                                           deepak
                        0.58
##
                                                   0.58
                                                                             0.58
##
                      chopra
                                              marianne
                                                                            dr.oz
##
                        0.58
                                                   0.58
                                                                             0.58
                                                  dyer
##
                       wayne
                                                                             amma
                                                   0.58
##
                        0.58
                                                                             0.58
##
                health-care
                                                healer
                                                                             womb
##
                        0.58
                                                   0.58
                                                                             0.58
##
                      campos
                                                ounces
                                                                           bleeds
                        0.58
##
                                                   0.58
                                                                             0.58
##
                     neonate
                                            fatalities www.theneonatalwomb.org
                        0.58
                                                   0.58
                                                                             0.58
##
                                                  nicu
##
                   neonatal
                                                                    connectivity
                                                   0.56
##
                        0.57
                                                                             0.56
                                               diverse
                   activist
                                                                         weighing
##
                        0.55
                                                   0.54
##
                                                                             0.54
##
                     forming
                                          long-lasting
                                                                             marc
                        0.53
                                                   0.53
                                                                             0.53
##
##
                       craig
                                            kielburger
                                                                 non-traditional
##
                        0.53
                                                   0.53
                                                                             0.53
##
                 caretakers
                                         technological
                                                                     drastically
                        0.53
##
                                                   0.52
                                                                             0.52
##
                  recreated
                                                viable
                                                                     reconnected
                        0.52
                                                   0.52
                                                                             0.52
##
               confidential
                                                 nborn
                                                                            fuels
##
                                                  0.51
                        0.52
                                                                             0.51
##
               complication
                                             gestation
                                                                    transfusions
##
                        0.51
                                                   0.51
##
                                                                             0.51
##
                 williamson
                                              resource
                                                                         champion
                        0.51
                                                   0.50
                                                                             0.50
##
##
                        thru
##
                        0.50
##
## $support
## numeric(0)
```

```
##
  $world
##
                                            kielburger
                                                                      williamson
                       craig
##
                        0.37
                                                   0.37
                                                                             0.36
                   stadiums
##
                                          motivational
                                                                         speakers
                        0.36
##
                                                   0.36
                                                                             0.36
##
              humanitarians
                                               diverse
                                                                         activist
##
                        0.36
                                                   0.34
                                                                             0.33
##
                     forming
                                            anesthesia
                                                                       teachings
##
                        0.33
                                                  0.33
                                                                             0.33
##
                     mariner
                                                                           rahman
                                                 jamal
##
                        0.33
                                                   0.33
                                                                             0.33
##
                   dr.larry
                                                dossey
                                                                           bernie
                        0.33
                                                                             0.33
##
                                                   0.33
##
                      siegel
                                                deepak
                                                                           chopra
                        0.33
                                                  0.33
##
                                                                             0.33
                   marianne
                                                 dr.oz
##
                                                                            wayne
##
                        0.33
                                                  0.33
                                                                             0.33
##
                        dyer
                                                  amma
                                                                     health-care
                        0.33
                                                   0.33
                                                                             0.33
##
##
                      healer
                                                campos
                                                                           ounces
##
                        0.33
                                                   0.33
                                                                             0.33
##
                      bleeds
                                               neonate
                                                                      fatalities
##
                        0.33
                                                   0.33
                                                                             0.33
                                          confidential
   www.theneonatalwomb.org
                                                                         neonatal
##
##
                        0.33
                                                  0.33
                                                                             0.32
                reconnected
                                              start-up
##
                                                                             womb
##
                        0.32
                                                  0.32
                                                                             0.32
##
                      global
                                                 fuels
                                                                   technological
                                                  0.31
                        0.31
                                                                             0.31
##
##
                   weighing
                                              resource
                                                                             thru
                        0.31
                                                  0.30
                                                                             0.30
##
##
## $family
   going home
##
    0.31 0.31
##
  $school
##
##
        high
               students
                           teacher classroom education
                                                                 œan
        0.42
                                          0.35
                                                     0.32
##
                    0.39
                              0.37
                                                                0.31
##
## $people
   numeric(0)
##
##
## $life
##
                  began
                                       doctor
                                                            swollen
##
                    0.38
                                          0.38
                                                                0.38
##
                    face
                                      finally
                                                                skin
                                                                0.36
                    0.37
                                          0.36
##
##
            appointment
                                      anymore
                                                                take
##
                    0.36
                                          0.35
                                                                0.35
##
               allergic
                                   scratching
                                                           youtu.be
##
                    0.35
                                          0.35
                                                                0.35
##
                   went
                                        severe
                                                                time
```

23/2010			Businessi marysis
##	0.34	0.34	0.33
##	thighs	toll	paperwork
##	0.33	0.33	0.33
##	months	cuts	expressed
##	0.32	0.32	0.32
##	doses	medication	dorm
##	0.32	0.32	0.32
##	glance	rash	colonized
##	0.32	0.32	0.32
##	prescribe	corticosteroid	prednisone
##	0.32	0.32	0.32
##	suppress	weaned	reappear
##	0.32	0.32	0.32
##	itching	expansion	bumps
##	0.32	0.32	0.32
##	œbestâ	dermatologist	prescribed
##	0.32	0.32	0.32
##	scales	wean	clueless
##	0.32	0.32	0.32
##	lasted	60mg	mg
##	0.32	0.32	0.32
##	demanded	masking	vents
##	0.32	0.32	0.32
##	prick	dander	astonishment
##	0.32	0.32	0.32
##	plant	indoors	gashes
##	0.32	0.32	0.32
##	dryness	digestible	swelling
##	0.32	0.32	0.32
##	ketogenic	vegetarian	stung
##	0.32	0.32	0.32
##	fasting	functional	conventional
##	0.32	0.32	0.32
##	œrareâ	subsided	rashes
##	0.32	0.32	0.32
##	diets	recommends	stool
##	0.32 saliva	0.32 urine	0.32 parasitic
##	0.32	0.32	0.32
##	fungal	overgrowth	sears
##	0.32	0.32	0.32
##	ladder		worldstarhiphop.com
##	0.32	0.32	0.32
##	averages	worldstarhiphop	pricing
##	0.32	0.32	0.32
##	ein	back	never
##	0.32	0.31	0.31
##	get	couldnâ	body
##	0.31	0.31	0.31
##	took	subside	tested
##	0.31	0.31	0.31
##	rules	laboratory	even
##	0.31	0.31	0.30

## one arms supplements ## 0.30 0.30 0.30

## **Topic Modeling**

```
library(stm)
```

Processing the data for analysis.

```
require(stm)
temp<-textProcessor(storytext, metadata = story)</pre>
```

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

```
names(temp)
```

```
## [1] "documents" "vocab" "meta" "docs.removed"
```

```
meta<-temp$meta
vocab<-temp$vocab
docs<-temp$documents
out <- prepDocuments(docs, vocab, meta)</pre>
```

```
## Removing 7572 of 12928 terms (7572 of 102210 tokens) due to frequency
## Your corpus now has 823 documents, 5356 terms and 94638 tokens.
```

```
docs<-out$documents
vocab<-out$vocab
meta <-out$meta</pre>
```

#### Running stm for top 20 topics

```
## Beginning Initialization.
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -7.226)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -7.224, relative change = 3.314e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -7.222, relative change = 2.321e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -7.220, relative change = 2.579e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -7.218, relative change = 3.234e-04)
## Topic 1: time, never, heart, just, can
## Topic 2: school, new, back, colleg, high
## Topic 3: life, girl, know, run, one
## Topic 4: support, communiti, educ, program, provid
## Topic 5: volunt, need, peopl, day, other
## Topic 6: children, live, chang, hope, give
   Topic 7: give, help, donat, love, can
##
## Topic 8: famili, year, home, one, get
   Topic 9: help, work, make, first, like
##
   Topic 10: year, mani, kid, make, want
##
##
```

```
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -7.215, relative change = 3.952e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -7.212, relative change = 4.671e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -7.208, relative change = 5.252e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -7.204, relative change = 5.411e-04)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 10 (approx. per word bound = -7.200, relative change = 5.397e-0
4)
## Topic 1: never, time, heart, just, can
## Topic 2: school, new, back, colleg, high
## Topic 3: life, girl, know, run, one
## Topic 4: communiti, support, educ, program, provid
## Topic 5: volunt, need, peopl, day, other
   Topic 6: children, live, chang, hope, give
##
## Topic 7: give, help, donat, love, can
## Topic 8: famili, home, year, one, get
   Topic 9: help, work, make, first, like
##
   Topic 10: year, mani, kid, make, want
##
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 11 (approx. per word bound = -7.196, relative change = 5.412e-0
```

```
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 12 (approx. per word bound = -7.192, relative change = 5.409e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 13 (approx. per word bound = -7.188, relative change = 5.409e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 14 (approx. per word bound = -7.185, relative change = 5.428e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 15 (approx. per word bound = -7.181, relative change = 5.467e-0
4)
## Topic 1: never, heart, just, time, can
##
   Topic 2: school, new, back, colleg, high
## Topic 3: life, girl, know, run, one
## Topic 4: communiti, support, educ, program, provid
   Topic 5: volunt, need, peopl, day, help
##
   Topic 6: children, chang, hope, live, give
##
## Topic 7: give, help, donat, love, can
## Topic 8: famili, home, year, one, get
   Topic 9: help, work, make, like, day
##
   Topic 10: year, mani, kid, make, want
##
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 16 (approx. per word bound = -7.177, relative change = 5.536e-0
4)
##
```

```
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 17 (approx. per word bound = -7.173, relative change = 5.638e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 18 (approx. per word bound = -7.168, relative change = 5.735e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 19 (approx. per word bound = -7.164, relative change = 5.757e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 20 (approx. per word bound = -7.160, relative change = 5.755e-0
4)
## Topic 1: never, heart, just, can, time
## Topic 2: school, colleg, back, new, son
## Topic 3: life, girl, run, one, know
## Topic 4: communiti, support, educ, program, provid
## Topic 5: volunt, need, peopl, help, day
## Topic 6: children, chang, hope, give, live
   Topic 7: give, love, donat, help, can
##
## Topic 8: famili, home, year, one, get
## Topic 9: help, make, work, like, day
   Topic 10: year, mani, kid, make, want
##
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 21 (approx. per word bound = -7.156, relative change = 5.787e-0
4)
##
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 22 (approx. per word bound = -7.152, relative change = 5.862e-0
```

BusinessAnalysis

```
## Completed E-Step (0 seconds).

## Completed M-Step.

## Completing Iteration 23 (approx. per word bound = -7.148, relative change = 5.979e-0
4)

##

## Completed E-Step (0 seconds).

## Completed M-Step.

## Completing Iteration 24 (approx. per word bound = -7.143, relative change = 6.102e-0
4)

##

##

## Completed E-Step (0 seconds).

## Completed E-Step (0 seconds).

## Completed M-Step.

## Model Terminated Before Convergence Reached
```

```
topics <-labelTopics(prevfit , topics=c(1:20))
head(topics,3)</pre>
```

```
## $prob
##
          [,1]
                       [,2]
                                   [,3]
                                           [,4]
                                                       [,5]
                                                                  [,6]
                                                                            [,7]
                                   "just"
    [1,] "never"
                       "heart"
                                                                  "time"
                                                                             "call"
##
                                            "can"
                                                       "walk"
    [2,]
                       "colleg"
                                   "son"
                                            "high"
                                                       "back"
                                                                  "dream"
                                                                            "new"
##
          "school"
##
    [3,] "life"
                       girl"
                                   "run"
                                            "one"
                                                       "know"
                                                                  "back"
                                                                             "someth"
                                                                  "provid"
                                                                            "world"
##
    [4,]
          "communiti"
                       "support"
                                   "educ"
                                            "program"
                                                       "student"
##
    [5,] "volunt"
                       "need"
                                   "peopl"
                                           "help"
                                                       "other"
                                                                  "day"
                                                                             "inspir"
    [6,] "children"
                       "give"
                                   "chang"
                                           "hope"
                                                       "live"
                                                                  "young"
                                                                             "differ"
##
                       "love"
                                                       "can"
                                                                  "time"
                                                                             "will"
##
    [7,] "give"
                                   "donat"
                                           "help"
##
    [8,] "famili"
                       "home"
                                   "year"
                                            "one"
                                                       "get"
                                                                  "cancer"
                                                                            "month"
    [9,] "make"
                       "help"
                                   "work"
                                            "like"
                                                       "day"
                                                                  "first"
                                                                             "know"
##
##
   [10,] "year"
                       "mani"
                                   "make"
                                            "kid"
                                                       "want"
                                                                  "live"
                                                                            "learn"
##
## $frex
##
          [,1]
                    [,2]
                               [,3]
                                         [,4]
                                                   [,5]
                                                                [,6]
                                                                           [,7]
    [1,] "hospit"
                                         "doctor"
##
                    "walk"
                               "told"
                                                   "room"
                                                                "tell"
                                                                           "ever"
          "danc"
                    "dream"
                               "eagl"
                                         "hors"
                                                    "club"
                                                                "ride"
                                                                           "colleg"
##
    [2,]
                    girl"
                                         "smile"
                                                                           "got"
##
    [3,] "run"
                               "sister"
                                                    "name"
                                                                "grace"
                    "neonat"
                               "womb"
                                         "educ"
                                                    "leader"
                                                                "women"
                                                                           "global"
##
    [4,] "nicu"
                                                    "chariti"
##
    [5,] "bless"
                    "struggl"
                               "dedic"
                                         "volunt"
                                                                "human"
                                                                           "daughter"
                                         "net"
                                                                           "bulli"
##
    [6,] "haiti"
                    "diseas"
                               "villag"
                                                    "malaria"
                                                                "counsel"
    [7,]
                               "pet"
                                         "cat"
                                                    "adopt"
                                                                "counti"
                                                                           "bird"
##
          "dog"
                    "rescu"
                    "diagnos" "home"
                                         "anoth"
                                                   "diagnosi"
                                                                "toy"
                                                                           "pass"
##
    [8,] "cancer"
                                                    "realli"
##
    [9,] "job"
                    "wildlif"
                               "long"
                                         "forest"
                                                                "enjoy"
                                                                           "good"
##
   [10,] "kid"
                    "camp"
                               "hous"
                                         "hear"
                                                    "art"
                                                                "joy"
                                                                           "state"
##
## $lift
##
          [,1]
                      [,2]
                                             [,3]
                                                                [,4]
                      "compartment"
                                             "consciousnessn"
                                                                "detach"
##
    [1,] "aliveâ"
    [2,] "personn"
                      "âœewodihalehuâ"
                                             "bass"
                                                                "elabor"
##
                                             "swore"
                                                                "togoâ"
##
    [3,1
          "togoles"
                      "supportingnnthank"
##
    [4,] "amma"
                      "babybabi"
                                             "berni"
                                                                "blessednnw"
    [5,] "surpass"
##
                      "contagi"
                                             "festiv"
                                                                "childcar"
                      "âœimprov"
                                             "blah"
                                                                "chainsaw"
##
    [6,] "ahkon"
##
    [7,] "greta"
                      "wic"
                                             "chhairo"
                                                                "euthanasia"
    [8,] "kulturec"
                                                                "criteria"
##
                      "lifebok"
                                             "cancer"
                      "thor"
                                                                "grassi"
##
    [9,] "forest"
                                             "deplor"
                                                                "cmt."
##
  [10,] "notch"
                      "nneach"
                                             "camp"
##
          [,5]
                                                         [,6]
                                                         "hypotherm"
##
    [1,] "httpswwwindiegogocomprojectscureptsdnow"
##
    [2,] "toledo"
                                                         "courtesi"
    [3,] "togomoney"
                                                         "torrenti"
##
                                                         "campo"
##
    [4,] "brazil"
    [5,] "juic"
                                                         "workâ"
##
##
    [6,] "chop"
                                                         "clang"
##
    [7,] "leash"
                                                         "nepali"
    [8,] "lifelin"
##
                                                         "emma"
                                                         "thorâ"
##
    [9,] "piti"
##
   [10,]
          "por"
                                                         "banner"
##
          [,7]
    [1,] "infiltr"
##
##
    [2,] "des"
    [3,] "awhil"
##
```

```
## [4,] "childrennnfre"
## [5,] "invis"
## [6,] "nnbi"
## [7,] "wwwrestorationworksinternationalorg"
## [8,] "custodi"
## [9,] "canyon"
## [10,] "homeâ"
```

Exploring the topics in context. Provides an insight and words contained in aa specific topic

```
help("findThoughts")
findThoughts(prevfit, texts = storytext, topics = 10, n = 2)
```

##

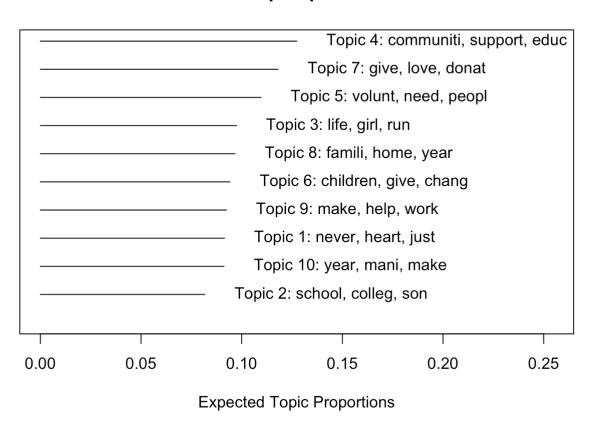
## Topic 10:

## Discipline, Hope, and Joy. That's what music education provides. I know first hand its transformative power -- I spent many, many hours as a kid in Chicago practicin g, going to violin lessons, working on theory homework, and rehearsing with orchestras a nd chamber groups. \n\nMusic provided me with lifelong lessons in discipline, hard wor k, and the rewards it provides. These lessons carried over beyond the concert hall and into my academics and later my professional life. It was a tremendous gift that my par ents were able to afford. \n\nSadly, not everyone can and for those that cannot, public f unding for music education has been on the wane for years. It's up to private organizat ions to step in and fill that gap. The Merit School of Music (www.meritmusic.org) does a phenomenal job of providing access to world class music education to students across \n\nThrough its programs, students get top notch music instruction from the b est teachers in the city. A masterclass with Renee Fleming or Yo-Yo Ma -- can you even imagine how exciting for a young student? Opportunities to go on tour? Amazing. s truly remarkable to hear from students the ways in which Merit has touched their live \n\nI support Merit because music education builds character and community -- and it s dividends are long lasting and far reaching. In a time where hope and joy are sometim es in short supply, music fills that void and Merit is on the front lines preparing the next generation of music performers and equally important, music appreciators.

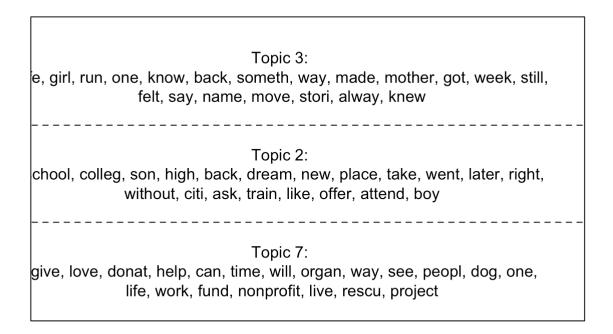
## It is a good feeling to be truly passionate about a cause, a mission, that benef its the community we live in. I found that passion about a year ago. It all started on a regular workday when I passed Union Station's Great Hall in downtown Chicago, catching my normal Metra train home during rush hour. I recall it was the only time I would not mind if there were mass train delays. Because on that early cool December evening, cham ber music filled the Great Hall with familiar holiday tunes. It lifted my spirit and mes merized the great many commuters that were either passing by or awaiting for their sched uled trains. I've passed that hall on a regular basis over some ten years and I've never seen this type of delightful public performances. What amazed me was the young musician s (middle/high school aged kids) that performed that day. Above the stage there was a ba nner that read 'Merit School of Music'. That name stuck with me since then. Months late r, I decided to google it. To my amazement this school (www.meritmusic.org) has been pro viding free or low cost music education (with free/affordable instrument rental) for chi ldren (pre-K to pre-college) in Chicago area for nearly 4 decades. Their mission has bee n to inspire youth to achieve their fullest potential in life through music, focusing on underserved communities. In the following summer, I signed my 4th-grade daughter up for Merit School of Music's summer String Camp. It was truly an enriching experience, for b oth my daughter and I. The fun-filled summer camp allowed her to not only improve her mu sic playing, but also to build friendship with fellow campers from various parts of the city/metro areas. As for me, I was deeply impressed with the integrity, talent, and de dication of Merit's staff and faculty. It was through that experience, I want to be part of Merit family and continue to support this organization in any way I can. This fall, my daughter started her weekly music education at Merit, with a more formal (and rigoro us) curriculum. I myself have become part of the MPA (Merit Parent Association). As a pa rent, I value music education for the youth of our society in a nurturing, encouraging e nvironment. With funding cuts in may area schools, quality music education is hard to fi nd. Merit School of Music (www.meritmusic.org) filled that void and I'm grateful for tha t, as are many other parents. Most importantly I value the opportunities this school has provided to area children whose families may not otherwise have the financial means to seek quality music education. Merit is the only school in the area that insists on offe ring tuition-free conservatory program (for qualified and dedicated young talents regard less of family background). I believe Merit School of Music is one of Chicago area's gre atest gems and they deserve our accolade and our financial support.

```
help("plot.STM")
plot.STM(prevfit, type="summary")
```

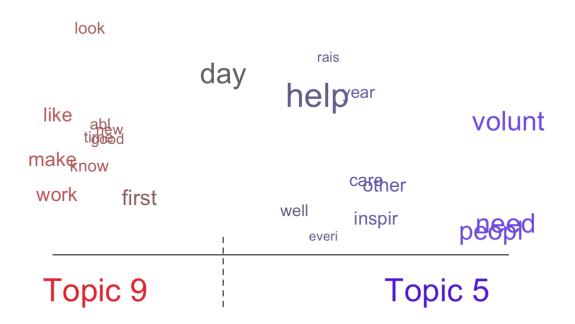
### **Top Topics**



```
plot.STM(prevfit, type="labels", topics=c(3,2,7))
```

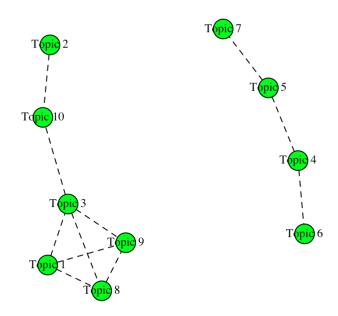


plot.STM(prevfit, type="perspectives", topics = c(9,5))



#### Correlation of topics

```
help(topicCorr)
mod.out.corr <- topicCorr(prevfit)
plot.topicCorr(mod.out.corr)</pre>
```



Topics 7, 5, 4 and 6 are interrrelated. Topics 2, 10, 3, 1, 9 and 8 on the other hand are related to one another.

# **Advance Topic Modelling**

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

story <- read.csv("/Users/arnavsomani/Desktop/NYU COURSE/sem 3/ba/r programming csv files/story.csv")</pre>
```

str(story)

```
## 'data.frame':
                   823 obs. of 13 variables:
## $ Inclusion
                        : logi TRUE TRUE TRUE TRUE TRUE ...
## $ votes total
                         : int 33 1029 20 7 15 29 2 54 66 6 ...
## $ votes gallery .Calc.: int 4 72 2 2 1 3 0 27 22 1 ...
## $ votes local network : int 29 957 18 5 14 26 2 27 44 5 ...
                        : Factor w/ 803 levels "11/1/15 14:32",..: 301 747 726 378 446
## $ createdate
135 21 567 622 458 ...
## $ ipaddress
                       : Factor w/ 777 levels "100.15.14.250",..: 484 27 151 273 258
40 16 119 134 382 ...
                         : Factor w/ 596 levels "01-0360745", "01-0450352",...: 88 225 59
## $ EIN
0 108 108 24 1 232 554 274 ...
## $ Photo
                        : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ URL_in_Story_.Calc. : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ matched_zip : Factor w/ 556 levels "","0","01331",..: 545 278 463 79 79 80
244 327 556 317 ...
## $ Story Text
                        : Factor w/ 810 levels "
PSP.org\\n \\nMy husband's diagnosis of PSP came as a terrible shock.\\nFor a number
"| __truncated__,..: 512 120 30 781 734 638 248 84 92 94 ...
## $ give_reason
                        : int 3 1 2 2 2 1 2 2 2 1 ...
## $ give method
                        : int 2 2 2 6 4 2 2 4 2 2 ...
```

```
storytext<-as.character(story$Story_Text)
class(storytext)</pre>
```

```
## [1] "character"
```

```
storyold<-corpus(storytext)
```

```
library(quanteda)
library(stm)
library(tm)
library(NLP)
library(openNLP)
library(ggplot2)
library(ggdendro)
library(cluster)
library(fpc)
```

```
require(quanteda)
```

```
summary(storyold)
```

```
## Corpus consisting of 823 documents, showing 100 documents.
##
##
       Text Types Tokens Sentences
##
      text1
                306
                        797
                                    21
##
      text2
                380
                      1072
                                    35
##
                                    16
      text3
                216
                        413
##
      text4
                214
                        409
                                    17
##
                        280
                                     8
      text5
                164
##
                                    37
      text6
                422
                      1043
##
      text7
                324
                        696
                                    22
##
                                    11
                156
                        272
      text8
##
                234
                        4\,7\,1
                                     6
      text9
                                     9
##
     text10
                123
                        208
##
     text11
                269
                        502
                                    23
##
     text12
                322
                        691
                                    19
                                     8
##
     text13
                109
                        181
##
                253
                        475
                                    14
     text14
##
                                    10
     text15
                140
                        276
##
                146
                        245
                                     8
     text16
##
                320
                        785
                                    36
     text17
##
                280
                                    15
     text18
                        615
##
                                    15
                147
                        228
     text19
##
                        271
                                    12
     text20
                158
                                    25
##
     text21
                268
                        621
##
     text22
                97
                        162
                                     5
                                    10
##
     text23
                155
                        326
##
     text24
                304
                        704
                                    24
##
     text25
                375
                        974
                                    31
##
                                    11
     text26
                258
                        542
##
     text27
                        280
                                     9
                161
                                    17
##
     text28
                234
                        447
##
     text29
                281
                        647
                                    15
##
     text30
                131
                        210
                                     8
                                     5
##
     text31
                76
                        122
##
     text32
                53
                        75
                                     5
##
     text33
                339
                        793
                                    31
                                     8
##
     text34
                106
                       180
##
     text35
                227
                        433
                                    13
##
                                    23
     text36
                265
                        525
                                     8
##
     text37
                173
                        312
##
     text38
                145
                        273
                                     6
##
     text39
                145
                        273
                                     6
                                    14
##
     text40
                148
                        243
##
     text41
                230
                        424
                                    20
##
     text42
                210
                        414
                                     6
                                     5
##
                235
                       505
     text43
                                     5
##
     text44
                        313
                164
##
     text45
                152
                        268
                                    16
##
     text46
                232
                        425
                                    18
##
     text47
                133
                        222
                                     7
##
                 39
                         61
                                     2
     text48
                                     2
##
     text49
                 43
                        86
##
     text50
                240
                        574
                                    21
```

```
##
               479
                      1278
                                    35
     text51
##
     text52
               139
                       226
                                     4
                                    46
##
               347
                       967
     text53
##
     text54
               360
                       828
                                    31
##
     text55
               413
                       896
                                    38
##
                                    10
     text56
               148
                       259
##
     text57
               171
                       315
                                    18
##
     text58
               209
                       396
                                    17
##
               376
                       846
                                    33
     text59
##
     text60
               127
                       228
                                    12
##
               214
                       394
                                    19
     text61
##
               203
                       403
                                    24
     text62
##
                                    36
     text63
               262
                       567
##
                                     9
               180
                       356
     text64
                                     9
##
     text65
               158
                       260
##
     text66
               148
                       250
                                    13
##
               244
                       525
                                    12
     text67
##
                       397
                                    13
     text68
               191
##
               140
                       207
                                    11
     text69
##
               156
                       257
                                    12
     text70
##
                                     9
                       429
     text71
               215
##
     text72
               193
                       426
                                     9
                                    27
##
               265
                       556
     text73
##
                                    12
     text74
               213
                       409
                                     9
##
     text75
               168
                       309
##
     text76
               304
                       722
                                    14
##
     text77
               140
                       258
                                    10
##
                                    10
     text78
                79
                       131
                                     6
##
     text79
               154
                       262
                                     5
##
     text80
                 54
                         76
                                     3
##
     text81
                         95
                 66
##
     text82
               178
                       351
                                    13
##
     text83
               240
                       499
                                    13
##
     text84
                                     6
               102
                       169
                                    31
##
     text85
               289
                       593
##
     text86
               117
                       194
                                     9
##
     text87
                91
                       150
                                     7
                                     9
##
                       303
     text88
               173
                                    34
##
     text89
               433
                      1108
                                     9
##
     text90
               162
                       302
                                     5
##
     text91
                70
                       102
##
     text92
               165
                       314
                                    11
##
                       171
                                     7
     text93
               112
##
     text94
               289
                       614
                                    26
##
     text95
               229
                       446
                                    12
                                    13
##
     text96
               173
                       310
##
               399
                                    35
     text97
                       961
##
                                    13
     text98
               184
                       355
##
                                    18
     text99
               336
                       620
##
    text100
               284
                       586
                                    28
##
             /Users/arnavsomani/Desktop/* on x86_64 by arnavsomani
## Source:
## Created: Wed Nov 23 17:00:19 2016
## Notes:
```

#### Generating DFM

```
corpus<- toLower(storyold, keepAcronyms = FALSE)</pre>
cleancorpus <- tokenize(corpus,</pre>
                        removeNumbers=TRUE,
                        removePunct = TRUE,
                        removeSeparators=TRUE,
                        removeTwitter=FALSE,
                        verbose=TRUE,
                        stem=F,
                        bigram=T)
## Warning in tokenize.character(corpus, removeNumbers = TRUE, removePunct =
## TRUE, : Arguments stembigram not used.
## Starting tokenization...
## ...preserving Twitter characters (#, 0)
## ...total elapsed: 0.001999999999934 seconds.
## ...tokenizing texts
## ...total elapsed: 0.17200000000004 seconds.
## ...replacing Twitter characters (#, 0)
## ...total elapsed: 0.033000000000013 seconds.
## ...replacing names
## ...total elapsed: 0 seconds.
## Finished tokenizing and cleaning 823 texts.
dfm<- dfm(cleancorpus,
                 toLower = TRUE,
                 ignoredFeatures =stopwords("english"),
                 verbose=TRUE,
                 stem=F)
##
```

... lowercasing

##

```
11/23/2016
                                                   Business Analysis \\
   ##
   ##
          ... indexing documents: 823 documents
   ##
   ##
          ... indexing features:
   ## 17,103 feature types
   ##
   ## ...
   ## removed 161 features, from 174 supplied (glob) feature types
   ##
          ... created a 823 x 16942 sparse dfm
   ##
          ... complete.
   ## Elapsed time: 0.362 seconds.
  Reviewing top features
   topfeatures(dfm, 50)
```

##	n	give	â	time	life
##	3148	878	872	869	866
##	one	help	people	children	can
##	841	832	781	758	735
##	s	years	many	community	will
##	679	669	653	636	628
##	support	ni	family	day	like
##	591	562	555	552	535
##	work	just	school	us	back
##	529	528	526	523	515
##	world	year	first	giving	every
##	487	466	448	446	446
##	know	make	organization	love	t
##	444	443	436	426	424
##	way	get	also	much	new
##	414	395	385	374	370
##	now	others	home	families	need
##	360	358	354	353	348
##	even	see	lives	never	able
##	345	342	341	340	334

# Hierc. Clustering

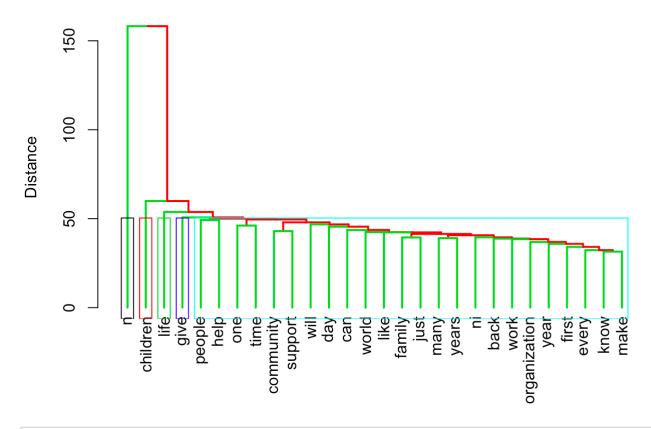
Performing heirarchial clustering and removing sparse terms to achieve an optimum term count of 28.

```
dfm.tm<-convert(dfm, to="tm")</pre>
dfm.tm
## <<DocumentTermMatrix (documents: 823, terms: 16942)>>
## Non-/sparse entries: 113655/13829611
## Sparsity
                       : 99%
## Maximal term length: 102
## Weighting
                       : term frequency (tf)
dtmss <- removeSparseTerms(dfm.tm, 0.67)</pre>
dtmss
## <<DocumentTermMatrix (documents: 823, terms: 28)>>
## Non-/sparse entries: 9760/13284
## Sparsity
## Maximal term length: 12
## Weighting
                       : term frequency (tf)
d.dfm <- dist(t(dtmss), method="euclidian")</pre>
fit <- hclust(d=d.dfm, method="average")</pre>
hcd <- as.dendrogram(fit)</pre>
require(cluster)
k<-5
plot(hcd, ylab = "Distance", horiz = FALSE,
     main = "Five Cluster Dendrogram",
```

edgePar = list(col = 2:3, lwd = 2:2))

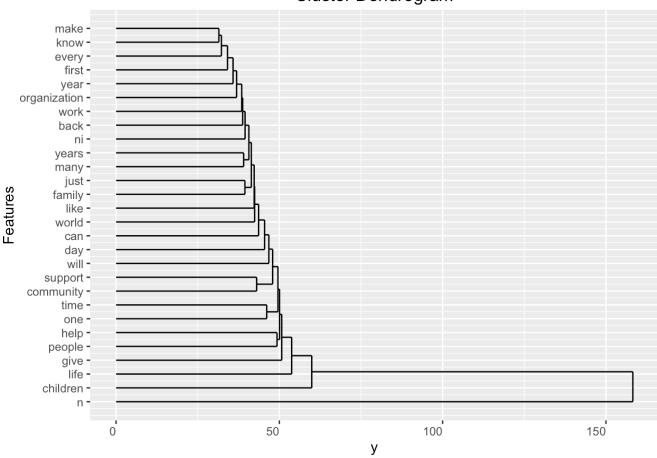
rect.hclust(fit, k=k, border=1:5)

### **Five Cluster Dendrogram**



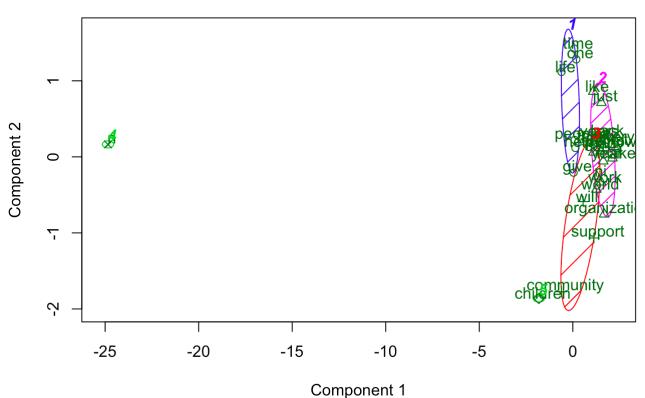
```
ggdendrogram(fit, rotate = TRUE, size = 4, theme_dendro = FALSE, color = "blue") +
xlab("Features") +
ggtitle("Cluster Dendrogram")
```

### Cluster Dendrogram



```
require(fpc)
d <- dist(t(dtmss), method="euclidian")
kfit <- kmeans(d, 5)
clusplot(as.matrix(d), kfit$cluster, color=T, shade=T, labels=2, lines=0)</pre>
```

### CLUSPLOT( as.matrix(d) )



These two components explain 86.77 % of the point variability.

# Pre-processing:

Before fitting a topic model, we need to tokenize the text. We remove punctuation and some common and some irrelevant stop words to get a fairly clean data set. In particular, we use the english stop words from the SMART information retrieval system, available in the R package tm.

```
require(tm)
stop_words <- stopwords("SMART")
stop_words <- c(stop_words,"n", "t", "ni","a","s","just","go","nthe","â")
stop_words <- tolower(stop_words)

storyold <- gsub("'", "", storyold)
storyold <- gsub("[[:punct:]]", " ", storyold)
storyold <- gsub("[[:cntrl:]]", " ", storyold)
storyold <- gsub("^[[:space:]]+", "", storyold)
storyold <- gsub("[[:space:]]+$", "", storyold)
storyold <- gsub("[-a-zA-Z -]", " storyold)
storyold <- tolower(storyold)</pre>
```

```
storyold <- storyold[storyold != ""]</pre>
```

#### Removing spaces

```
doc.list <- strsplit(storyold, "[[:space:]]+")</pre>
```

computing the table of terms:

```
term.table <- table(unlist(doc.list))
term.table <- sort(term.table, decreasing = TRUE)</pre>
```

Remove terms that are stop words or occur fewer than 5 times:

```
del <- names(term.table) %in% stop_words | term.table < 5
term.table <- term.table[!del]
term.table <- term.table[names(term.table) != ""]
vocab <- names(term.table)</pre>
```

Now put the documents into the format required by the Ida package:

```
get.terms <- function(x) {
  index <- match(x, vocab)
  index <- index[!is.na(index)]
  rbind(as.integer(index - 1), as.integer(rep(1, length(index))))
}
documents <- lapply(doc.list, get.terms)</pre>
```

Computing some statistics related to the data set:

The object documents is a list where each element represents one document, according to the specifications of the lda package. After creating this list, we compute a few statistics about the corpus:

```
D <- length(documents)
W <- length(vocab)
doc.length <- sapply(documents, function(x) sum(x[2, ]))
N <- sum(doc.length)
term.frequency <- as.integer(term.table)</pre>
```

# MCMC and model tuning parameters:

Next, we set up a topic model with 10 topics, relatively diffuse priors for the topic-term distributions ( $\eta\eta=0.02$ ) and document-topic distributions ( $\alpha\alpha=0.02$ ), and we set the collapsed Gibbs sampler to run for 3,000 iterations (slightly conservative to ensure convergence). A visual inspection of fit\$log.likelihood shows that the MCMC algorithm has converged after 25 iterations. This block of code takes about 2.98 minutes to run.

```
K <- 10
G <- 3000
alpha <- 0.02
eta <- 0.02</pre>
```

### Fit the model:

```
## Time difference of 1.700927 mins
```

## Visualizing the fitted model with LDAvis

To visualize the result using LDAvis, we'll need estimates of the document-topic distributions, which we denote by the D×K matrix  $\theta$ , and the set of topic-term distributions, which we denote by the K×W matrix  $\phi$ . We estimate the "smoothed" versions of these distributions ("smoothed" means that we've incorporated the effects of the priors into the estimates) by cross-tabulating the latent topic assignments from the last iteration of the collapsed Gibbs sampler with the documents and the terms, respectively, and then adding pseudocounts according to the priors.

## create the JSON object to feed the visualization:

We now call the createJSON() function in LDAvis. This function will return a character string representing a JSON object used to populate the visualization. The createJSON() function computes topic frequencies, inter-topic distances, and projects topics onto a two-dimensional plane to represent their similarity to each other. It also loops through a grid of values of a tuning parameter,  $0 \le \lambda \le 10 \le \lambda \le 1$ , that controls how the terms are ranked for each topic, where terms are listed in decreasing of relevance, where the relevance of term w to topic t is defined as  $\lambda \times p(w|t) + (1-\lambda) \times p(w|t)/p(w)\lambda \times p(w|t) + (1-\lambda) \times p(w|t)/p(w)$ . Values of  $\lambda$  near 1 give high relevance rankings to frequent terms within a given topic, whereas values of  $\lambda$  near zero give high relevance rankings to exclusive terms within a topic. The set of all terms which are ranked among the top-R most relevant terms for each topic are precomputed by the createJSON() function and sent to the browser to be interactively visualized using D3 as part of the JSON object.

### **Key topics:**

Topic 3 discusses about rescuing the Girl child and proving shelter, medical assistance and schooling.

Topic 2 talks about providing funds to the foundation for medical assistance to the Cancer patients and create awareness about mental health.

Topic 7 talks about the how passionate the people are when they involve in community services and satisfaction that people get when they use their money for Volunteering activities. They consider this to be an opportunity to improve the lives of the underprivileged and thereby making the world a better place.

Topic 9 Talks about heart diseases and create awareness about the donations required to take care of hospital expenses.

Topic 5 is about the programs for the youth to provide education and research opportunities for them.

An interactive model for the topic model is as below:

https://cdn.rawgit.com/HarshithaMaithry/BAExam/master/index.html (https://cdn.rawgit.com/HarshithaMaithry/BAExam/master/index.html)

### **Recommendations:**

The analysis of the data shows that the people involve in volunteering activities as it gives a sense of satisfaction and helps to make the world a better place. The people actively fund the non-profit organizations and several foundations that target different problems. Gates and Melinda foundation can capture these aspects of the donors while undertaking any activities.

- Most of the topics talk about providing medical assistance for children, girls, heart patients, cancer patients, disabled children. So the foundation can start a campaign to promote and create awareness about healthcare among the underprivileged.
- The next most talked about topic was promoting education for children especially girls and the disabled. An initiative to encourage people to provide financial assistance to those who cannot afford the expensive education can be undertaken.
- Also, we see that one of the major key point was to promote research opportunities among the youth.
   These social concerns can be the major target points while planning the next Campaign which is to be held in 2016.