

November 23, 2016

Text Mining for Bill Gates and Melinda Gates Foundation data

```
##
## 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 files/story.csv")
```

```
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 ...
## $ 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 "
Cure
PSP.org\\n \\nMy husbandâ€™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)
```

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

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

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

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

```
summary(newscorpus)
```

##	##	Text	Types	Tokens	Sentences
##	text1	306	797	21	
##	text2	380	1072	35	
##	text3	216	413	16	
##	text4	214	409	17	
##	text5	164	280	8	
##	text6	422	1043	37	
##	text7	324	696	22	
##	text8	156	272	11	
##	text9	234	471	6	
##	text10	123	208	9	
##	text11	269	502	23	
##	text12	322	691	19	
##	text13	109	181	8	
##	text14	253	475	14	
##	text15	140	276	10	
##	text16	146	245	8	
##	text17	320	785	36	
##	text18	280	615	15	
##	text19	147	228	15	
##	text20	158	271	12	
##	text21	268	621	25	
##	text22	97	162	5	
##	text23	155	326	10	
##	text24	304	704	24	
##	text25	375	974	31	
##	text26	258	542	11	
##	text27	161	280	9	
##	text28	234	447	17	
##	text29	281	647	15	
##	text30	131	210	8	
##	text31	76	122	5	
##	text32	53	75	5	
##	text33	339	793	31	
##	text34	106	180	8	
##	text35	227	433	13	
##	text36	265	525	23	
##	text37	173	312	8	
##	text38	145	273	6	
##	text39	145	273	6	
##	text40	148	243	14	
##	text41	230	424	20	
##	text42	210	414	6	
##	text43	235	505	5	
##	text44	164	313	5	
##	text45	152	268	16	
##	text46	232	425	18	
##	text47	133	222	7	
##	text48	39	61	2	
##	text49	43	86	2	
##	text50	240	574	21	

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 poverty; 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 allowed me to see how truly blessed my family and many other families here in America are compared to other parts of the world. In the process, I was astonished to discover that about 1.5 billion people in our world live on less than \$1.25 per day. How does someone even 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 pretend

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 25:40 which helps orphans and underprivileged children in South Africa. We were drawn to this organization, first and foremost because of the organization's name 25:40, drawn from Matthew 25:40 - "Truly I tell you, whatever you did for one of the least of these brothers and sisters of mine, you did for me." 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 organization.

Being a dancer for 18 years, I have much passion and love for dancing. God opened up a door to dance ministry a couple of years ago and through some prayer I decided to do a dance workshop over one summer to help benefit 25:40. It was such a blessing for me to be involved with dance again and at the same time such an amazing opportunity for me to use my time, talents, and passions to help these children in South Africa who so desperately needed our help.

During the workshop I was also able to talk about some of the children that 25:40 serves. How some of these children lived in small huts with no floors or had to walk every day long distances to collect water for their family. The children would ask me if I was making these stories up since it was so hard 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 made a great impact on those children's hearts and also mine as well.

I feel very blessed to have had the opportunity to give back and help children in South Africa through 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 amazing 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 time, 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 amazed to see that not only will other's lives be helped and impacted because of it, but our 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!

"

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

```
help(tokenize)
newscorpus<- toLower(newscorpus, keepAcronyms = FALSE)
cleancorpus <- tokenize(newscorpus,
                        removeNumbers=TRUE,
                        removePunct = TRUE,
                        removeSeparators=TRUE,
                        removeTwitter=FALSE,
                        verbose=TRUE)
```

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

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

```
## ...total elapsed: 0.002000000000000067 seconds.
```



```
## Finished tokenizing and cleaning 823 texts.
```

```
head(clean_corpus, 1)
```

file:///localhost/Users/arnavsomani/Desktop/BusinessAnalysis11111111111111111111.html

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

##	[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]	"them"	"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]	"the"	"children"	"come"
##	[469]	"to"	"the"	"workshop"
##	[472]	"emptying"	"their"	"own"
##	[475]	"piggy"	"banks"	"to"
##	[478]	"give"	"back"	"after"

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


```
## 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   1    1
## text2         0      1 11   0   0    0
## text3         0      1 13   0   0    0
```

Displaying most frequent terms in dfm

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

##	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	become	someone	together	mission
##	175	175	173	173	172
##	volunteers	center	free	things	months
##	171	171	171	170	170
##	still	began	gave	told	high
##	170	169	169	168	168
##	continue	inspired	felt	son	important

##	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

```
swlist = c("n", "t", "ni", "a", "s", "just", "go", "nthe", "â")
dfm.stem<- dfm(clean_corpus, toLower = TRUE,
               ignoredFeatures = c(swlist, stopwords("english")),
               verbose=TRUE,
               stem=F)
```

```
##
## ... lowercasing
```

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

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

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

##


```
## ...tokenizing texts
```

```
## ...total elapsed: 0.175000000000001 seconds.
```

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

```
## ...total elapsed: 0.03000000000000011 seconds.
```

```
## ...creating ngrams
```

```
## ...total elapsed: 1.851 seconds.
```

```
## ...replacing names
```

```
## ...total elapsed: 0.00100000000000122 seconds.
```

```
## Finished tokenizing and cleaning 823 texts.
```

```
dfm.bigram<- dfm(cleancorpus, toLower = TRUE,
  ignoredFeatures = c(swlist, stopwords("english")),
  verbose=TRUE,
  stem=FALSE)
```

```
##
## ... lowercasing
```

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

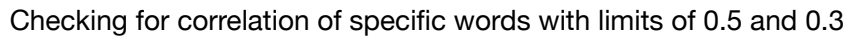
```
##
## ... indexing features:
```

```
## 124,676 feature types
```

##

...

```
## removed 81,058 features, from 182 supplied (glob) feature types
```

22/51

```
## $giving
## numeric(0)
##
## $children
## 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
##           wayne           dyer           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
##           neonatal           nicu           connectivity
##           0.57           0.56           0.56
##           activist           diverse           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
##           0.52           0.51           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)
```

file://localhost/Users/arnavsomani/Desktop/BusinessAnalysis111111111111111111.html

## \$world						
##	craig		kielburger			williamson
##	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		jamal			rahman
##	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
##	www.theneonatalwomb.org		confidential			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	æan
##	0.42	0.39	0.37	0.35	0.32	0.31
##						
## \$people						
##	numeric(0)					
##						
## \$life						
##	began		doctor			swollen
##	0.38		0.38			0.38
##	face		finally			skin
##	0.37		0.36			0.36
##	appointment		anymore			take
##	0.36		0.35			0.35
##	allergic		scratching			youtu.be
##	0.35		0.35			0.35
##	went		severe			time

##	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
##	æbestâ	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
##	ærareâ	subsided		rashes
##	0.32		0.32	0.32
##	diets	recommends		stool
##	0.32		0.32	0.32
##	saliva	urine		parasitic
##	0.32		0.32	0.32
##	fungal	overgrowth		sears
##	0.32		0.32	0.32
##	ladder	appeal	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


```
## 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-04)
## 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-04)
##
```

```
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 18 (approx. per word bound = -7.168, relative change = 5.735e-04)
##
```

```
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 19 (approx. per word bound = -7.164, relative change = 5.757e-04)
##
```

```
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 20 (approx. per word bound = -7.160, relative change = 5.755e-04)
## 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
```

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



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

```
## [4,] "childrennfre"  
## [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)
```

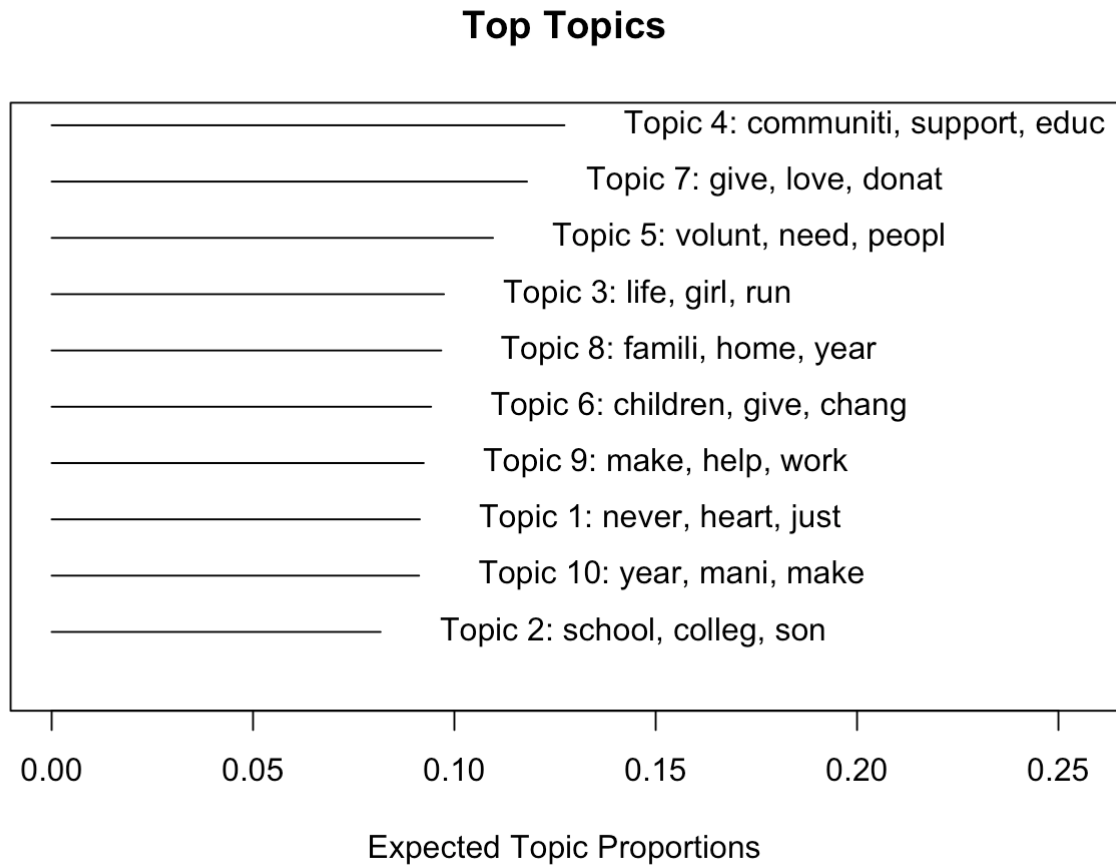
##

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 practicing, going to violin lessons, working on theory homework, and rehearsing with orchestras and chamber groups. \n\nMusic provided me with lifelong lessons in discipline, hard work, 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 parents were able to afford.\n\nSadly, not everyone can and for those that cannot, public funding for music education has been on the wane for years. It's up to private organizations 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 Chicago. \n\nThrough its programs, students get top notch music instruction from the best 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. It is truly remarkable to hear from students the ways in which Merit has touched their lives. \n\nI support Merit because music education builds character and community -- and its dividends are long lasting and far reaching. In a time where hope and joy are sometimes 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 benefits 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, chamber music filled the Great Hall with familiar holiday tunes. It lifted my spirit and mesmerized the great many commuters that were either passing by or awaiting for their scheduled 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 musicians (middle/high school aged kids) that performed that day. Above the stage there was a banner that read 'Merit School of Music'. That name stuck with me since then. Months later, I decided to google it. To my amazement this school (www.meritmusic.org) has been providing free or low cost music education (with free/affordable instrument rental) for children (pre-K to pre-college) in Chicago area for nearly 4 decades. Their mission has been 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 both my daughter and I. The fun-filled summer camp allowed her to not only improve her music 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 dedication 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 rigorous) curriculum. I myself have become part of the MPA (Merit Parent Association). As a parent, I value music education for the youth of our society in a nurturing, encouraging environment. With funding cuts in many area schools, quality music education is hard to find. Merit School of Music (www.meritmusic.org) filled that void and I'm grateful for that, 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 offering tuition-free conservatory program (for qualified and dedicated young talents regardless of family background). I believe Merit School of Music is one of Chicago area's greatest gems and they deserve our accolade and our financial support.

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

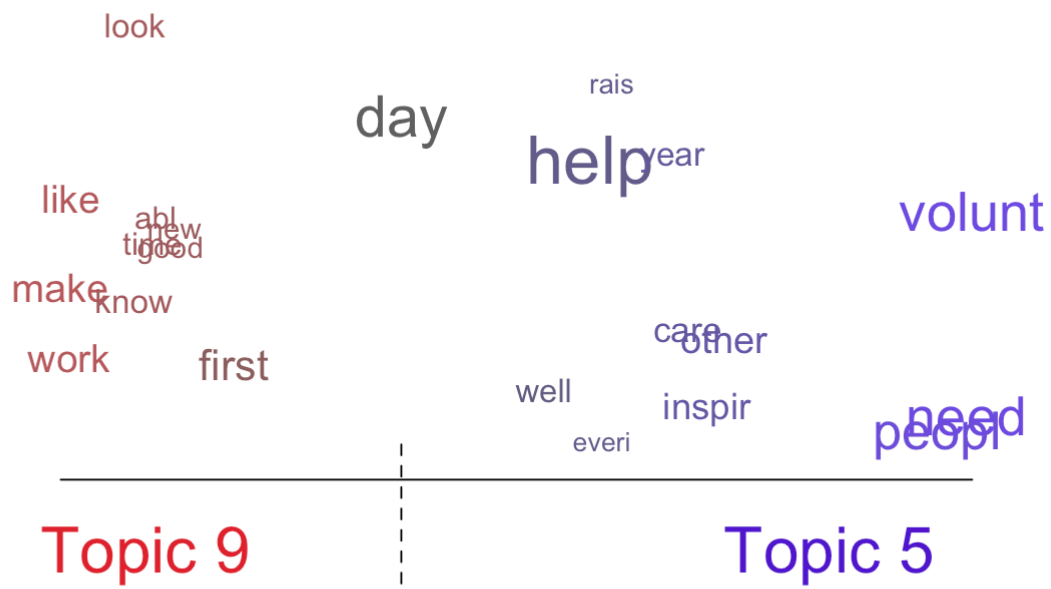


e, girl, run, one, know, back, someth, way, made, mother, got, week, still,
felt, say, name, move, stori, alway, knew

chool, colleg, son, high, back, dream, new, place, take, went, later, right,
without, citi, ask, train, like, offer, attend, boy

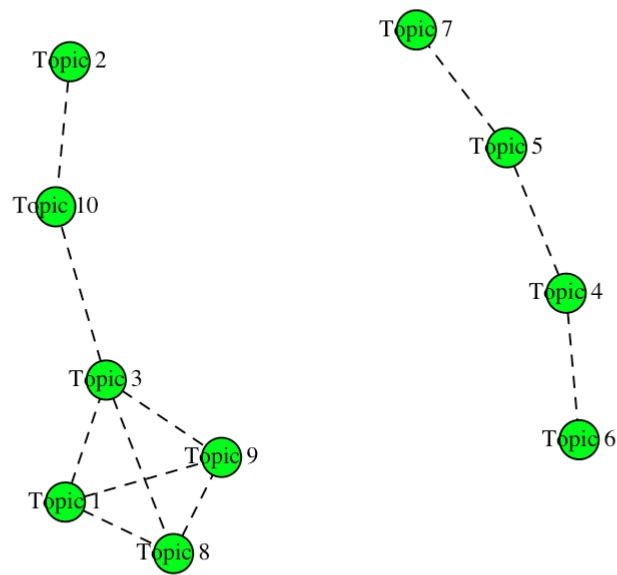
give, love, donat, help, can, time, will, organ, way, see, peopl, dog, one,
life, work, fund, nonprofit, live, rescu, project

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



Correlation of topics

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



Topics 7, 5, 4 and 6 are interrelated. 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")
```

```
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 ...
## $ 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 "Cure
  PSP.org\\n \\nMy husband&acircs 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)
```

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

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

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

```
require(quantda)
```

```
summary(storyold)
```


##

##

##

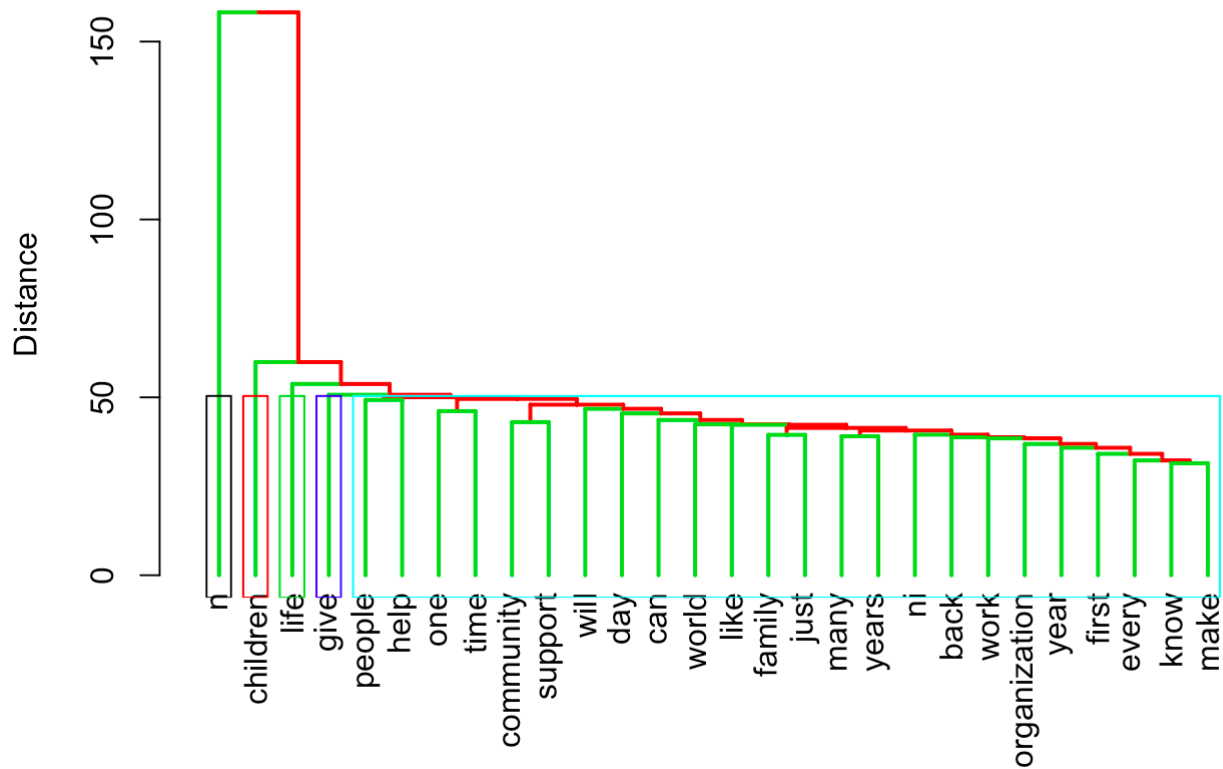
```
## Created: Wed Nov 23 17:00:19 2016
```

Notes:


```
##      ... created a 823 x 16942 sparse dfm
##      ... complete.
## Elapsed time: 0.362 seconds.
```

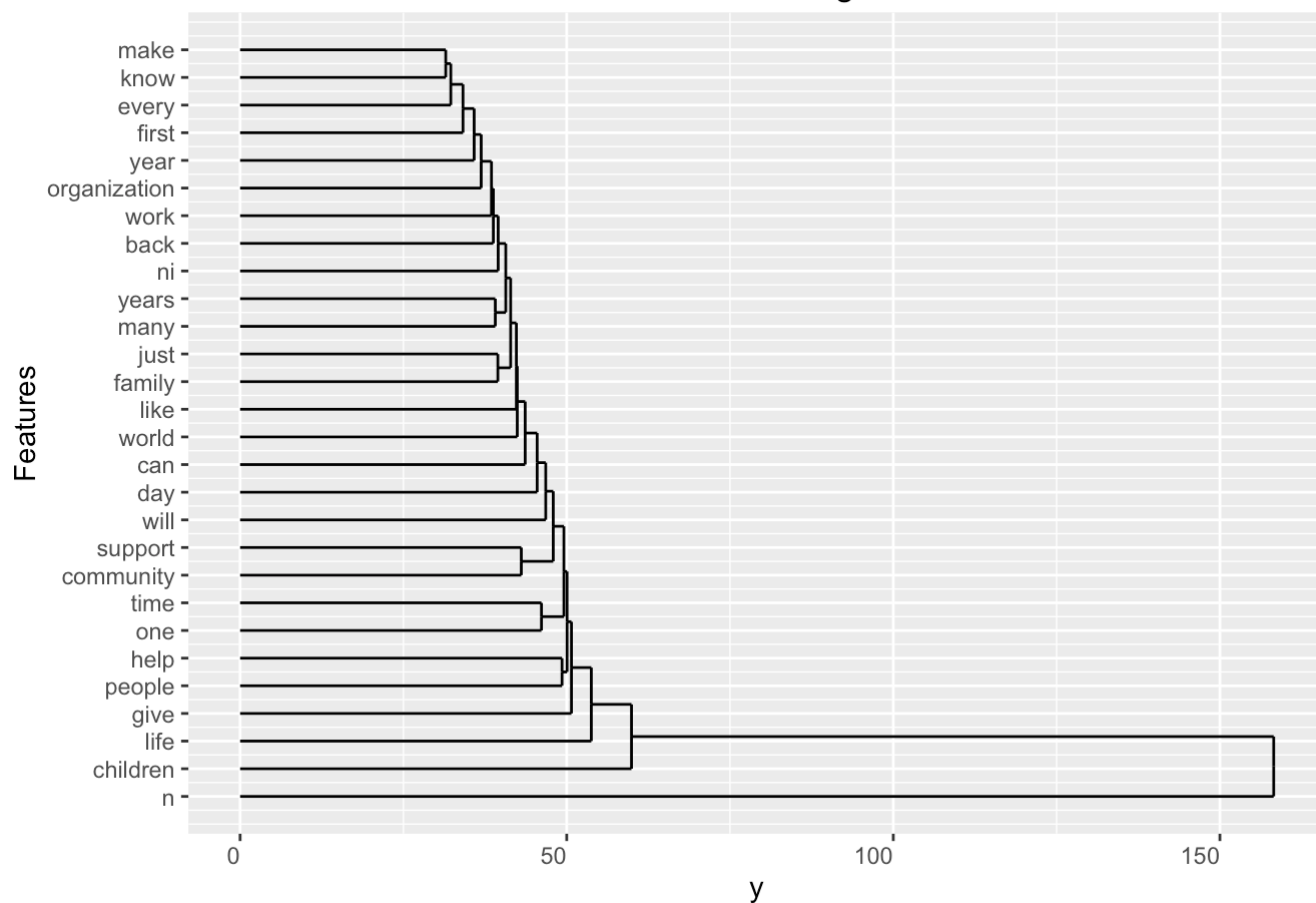
##	n	give	â	time	life
##	3148	878	872	869	866
##	one	help	people	children	can
##	841	832	781	758	735
##	s	years	many	community	will
##	679	669	653	636	628
##	support	ni	family	day	like
##	591	562	555	552	535
##	work	just	school	us	back
##	529	528	526	523	515
##	world	year	first	giving	every
##	487	466	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

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)
```



```
json <- createJSON(phi = news_for_LDA$phi,
                   theta = news_for_LDA$theta,
                   doc.length = news_for_LDA$doc.length,
                   vocab = news_for_LDA$vocab,
                   term.frequency = news_for_LDA$term.frequency)

#serVis(json, out.dir = 'vis1', open.browser = TRUE)
```

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.