Unsupervised Classification Model: Structural Topic Models and Dictionary Model (Sentiment Analysis)

Miras Tolepbergen

2023-10-24

Structural Topic Models

```
install.packages("stm", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'stm' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'stm'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Miras\AppData\Local\R\win-library\4.3\00LOCK\stm\libs\x64\stm.dll
to
## C:\Users\Miras\AppData\Local\R\win-library\4.3\stm\libs\x64\stm.dll:
Permission
## denied
```

```
## Warning: restored 'stm'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp2RdsUR\downloaded_packages
```

```
install.packages("syuzhet", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'syuzhet' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp2RdsUR\downloaded_packages
```

```
install.packages("reshape2", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'reshape2' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'reshape2'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Miras\AppData\Local\R\win-
library\4.3\00LOCK\reshape2\libs\x64\reshape2.dll
## to
## C:\Users\Miras\AppData\Local\R\win-
library\4.3\reshape2\libs\x64\reshape2.dll:
## Permission denied

## Warning: restored 'reshape2'

##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp2RdsUR\downloaded_packages

rm(list=ls(all=TRUE))
setwd("C:/Users/Miras/Desktop/u Milan/1st year classes/Big Data
Analystics/Labs/Lab1")
getwd()

## [1] "C:/Users/Miras/Desktop/u Milan/1st year classes/Big Data
Analystics/Labs/Lab1"

library(quanteda)

## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "replValueSp"; definition not updated

## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "xMatrix"; definition not updated

## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "mMatrix"; definition not updated

## Package version: 3.3.1
## Unicode version: 13.0
## ICU version: 69.1

## Parallel computing: 4 of 4 threads used.

## See https://quanteda.io for tutorials and examples.

library(readtext)

##
## Attaching package: 'readtext'
```

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

```
library(ggplot2)
library(stm)
```

```
## stm v1.3.6.1 successfully loaded. See ?stm for help.
##  Papers, resources, and other materials at structuraltopicmodel.com
```

```
library(dplyr)
```

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

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(syuzhet)
library(reshape2)
```

```
trump <- readRDS("C:/Users/Miras/Desktop/u Milan/1st year classes/Big Data
Analystics/Labs/Lab1/Trump2018.rds")
glimpse(trump)
```

```
## Rows: 3,565
## Columns: 5
## $ ID         <chr> "@realDonaldTrump", "@realDonaldTrump",
"@realDonaldTrump",…
## $ Time       <chr> " 2018-01-01 02:36", " 2018-01-01 03:06", " 2018-01-01
06:0…
## $ Tweet.URL  <chr> "
https://twitter.com/realDonaldTrump/status/94753695146433…
## $ Tweet.Text <chr> " My deepest condolences to the victims of the terrible
sho…
## $ date       <date> 2018-01-01, 2018-01-01, 2018-01-01, 2018-01-01, 2018-
01-01…
```

```
str(trump)
```

```
## 'data.frame':    3565 obs. of  5 variables:
##  $ ID        : chr  "@realDonaldTrump" "@realDonaldTrump"
"@realDonaldTrump" "@realDonaldTrump" ...
##  $ Time      : chr  " 2018-01-01 02:36" " 2018-01-01 03:06" " 2018-01-01
06:00" " 2018-01-01 06:18" ...
##  $ Tweet.URL : chr  "
https://twitter.com/realDonaldTrump/status/947536951464333318" "
```

```
https://twitter.com/realDonaldTrump/status/947544600918372353" "
https://twitter.com/realDonaldTrump/status/947588263103139841" "
https://twitter.com/realDonaldTrump/status/947592785519173637" ...
##  $ Tweet.Text: chr  " My deepest condolences to the victims of the
terrible shooting in Douglas County @DCSheriff, and their familie"|
__truncated__ " What a year it’s been, and we're just getting started.
Together, we are MAKING AMERICA GREAT AGAIN! Happy New "| __truncated__ "
Iran, the Number One State of Sponsored Terror with numerous violations of
Human Rights occurring on an hourly"| __truncated__ " As our Country rapidly
grows stronger and smarter, I want to wish all of my friends, supporters,
enemies, hate"| __truncated__ ...
##  $ date      : Date, format: "2018-01-01" "2018-01-01" ...

trump$date <- as.numeric(trump$date)

date <- trump$date

trump$Tweet.Text <- iconv(trump$Tweet.Text, "", "UTF-8")
trump$Text2 <- trump$Tweet.Text

myCorpus <- corpus(trump, text_field = "Tweet.Text")
head(summary(myCorpus))

##    Text Types Tokens Sentences                ID              Time
## 1 text1    31     33         3 @realDonaldTrump  2018-01-01 02:36
## 2 text2    24     27         4 @realDonaldTrump  2018-01-01 03:06
## 3 text3    34     37         2 @realDonaldTrump  2018-01-01 06:00
## 4 text4    41     49         1 @realDonaldTrump  2018-01-01 06:18
## 5 text5    18     19         2 @realDonaldTrump  2018-01-01 07:43
## 6 text6    52     59         3 @realDonaldTrump  2018-01-01 20:12
##                                                        Tweet.URL  date
## 1  https://twitter.com/realDonaldTrump/status/947536951464333318 17532
## 2  https://twitter.com/realDonaldTrump/status/947544600918372353 17532
## 3  https://twitter.com/realDonaldTrump/status/947588263103139841 17532
## 4  https://twitter.com/realDonaldTrump/status/947592785519173637 17532
## 5  https://twitter.com/realDonaldTrump/status/947614110082043904 17532
## 6  https://twitter.com/realDonaldTrump/status/947802588174577664 17532
##
Text2
## 1
My deepest condolences to the victims of the terrible shooting in Douglas
County @DCSheriff, and their families. We love our police and law enforcement
- God Bless them all! #LESM
## 2
What a year it’s been, and we're just getting started. Together, we are
MAKING AMERICA GREAT AGAIN! Happy New Year!! https://t.co/qsMNyN1UJG
## 3
Iran, the Number One State of Sponsored Terror with numerous violations of
Human Rights occurring on an hourly basis, has now closed down the Internet
so that peaceful demonstrators cannot communicate. Not good!
```

```
## 4                                                                  As our Country
rapidly grows stronger and smarter, I want to wish all of my friends,
supporters, enemies, haters, and even the very dishonest Fake News Media, a
Happy and Healthy New Year. 2018 will be a great year for America!
## 5
HAPPY NEW YEAR! We are MAKING AMERICA GREAT AGAIN, and much faster than
anyone thought possible!
## 6  The United States has foolishly given Pakistan more than 33 billion
dollars in aid over the last 15 years, and they have given us nothing but
lies &amp;amp; deceit, thinking of our leaders as fools. They give safe haven to
the terrorists we hunt in Afghanistan, with little help. No more!
```

```
tok <- tokens(myCorpus , remove_punct = TRUE, remove_numbers=TRUE,
remove_symbols = TRUE, split_hyphens = TRUE, remove_separators = TRUE,
remove_url = TRUE)
tok <- tokens_remove(tok, stopwords("en"))
tok <- tokens_wordstem (tok)
myDfm <- dfm(tok)

myDfm [ntoken(myDfm ) == 0,]
```

```
## Document-feature matrix of: 56 documents, 6,346 features (100.00% sparse)
and 5 docvars.
##           features
## docs       deepest condol victim terribl shoot dougla counti @dcsheriff
famili
##   text556        0      0      0       0     0      0      0          0
0
##   text655        0      0      0       0     0      0      0          0
0
##   text1086       0      0      0       0     0      0      0          0
0
##   text1131       0      0      0       0     0      0      0          0
0
##   text1201       0      0      0       0     0      0      0          0
0
##   text1305       0      0      0       0     0      0      0          0
0
##           features
## docs       love
##   text556     0
##   text655     0
##   text1086    0
##   text1131    0
##   text1201    0
##   text1305    0
## [ reached max_ndoc ... 50 more documents, reached max_nfeat ... 6,336 more
features ]
```

```
myDfm<- myDfm[ntoken(myDfm) > 0,]

head(docvars(myCorpus))

##                  ID              Time
## 1 @realDonaldTrump  2018-01-01 02:36
## 2 @realDonaldTrump  2018-01-01 03:06
## 3 @realDonaldTrump  2018-01-01 06:00
## 4 @realDonaldTrump  2018-01-01 06:18
## 5 @realDonaldTrump  2018-01-01 07:43
## 6 @realDonaldTrump  2018-01-01 20:12
##                                                                Tweet.URL  date
## 1  https://twitter.com/realDonaldTrump/status/947536951464333318 17532
## 2  https://twitter.com/realDonaldTrump/status/947544600918372353 17532
## 3  https://twitter.com/realDonaldTrump/status/947588263103139841 17532
## 4  https://twitter.com/realDonaldTrump/status/947592785519173637 17532
## 5  https://twitter.com/realDonaldTrump/status/947614110082043904 17532
## 6  https://twitter.com/realDonaldTrump/status/947802588174577664 17532
##
Text2
## 1
My deepest condolences to the victims of the terrible shooting in Douglas
County @DCSheriff, and their families. We love our police and law enforcement
- God Bless them all! #LESM
## 2
What a year it’s been, and we're just getting started. Together, we are
MAKING AMERICA GREAT AGAIN! Happy New Year!! https://t.co/qsMNyN1UJG
## 3
Iran, the Number One State of Sponsored Terror with numerous violations of
Human Rights occurring on an hourly basis, has now closed down the Internet
so that peaceful demonstrators cannot communicate. Not good!
## 4                                                                As our Country
rapidly grows stronger and smarter, I want to wish all of my friends,
supporters, enemies, haters, and even the very dishonest Fake News Media, a
Happy and Healthy New Year. 2018 will be a great year for America!
## 5
HAPPY NEW YEAR! We are MAKING AMERICA GREAT AGAIN, and much faster than
anyone thought possible!
## 6  The United States has foolishly given Pakistan more than 33 billion
dollars in aid over the last 15 years, and they have given us nothing but
lies &amp;amp; deceit, thinking of our leaders as fools. They give safe haven to
the terrorists we hunt in Afghanistan, with little help. No more!

head(docvars(myDfm))

##                  ID              Time
## 1 @realDonaldTrump  2018-01-01 02:36
## 2 @realDonaldTrump  2018-01-01 03:06
## 3 @realDonaldTrump  2018-01-01 06:00
## 4 @realDonaldTrump  2018-01-01 06:18
```

```
## 5 @realDonaldTrump  2018-01-01 07:43
## 6 @realDonaldTrump  2018-01-01 20:12
##                                                    Tweet.URL  date
## 1  https://twitter.com/realDonaldTrump/status/947536951464333318 17532
## 2  https://twitter.com/realDonaldTrump/status/947544600918372353 17532
## 3  https://twitter.com/realDonaldTrump/status/947588263103139841 17532
## 4  https://twitter.com/realDonaldTrump/status/947592785519173637 17532
## 5  https://twitter.com/realDonaldTrump/status/947614110082043904 17532
## 6  https://twitter.com/realDonaldTrump/status/947802588174577664 17532
##
Text2
## 1
My deepest condolences to the victims of the terrible shooting in Douglas
County @DCSheriff, and their families. We love our police and law enforcement
- God Bless them all! #LESM
## 2
What a year it’s been, and we're just getting started. Together, we are
MAKING AMERICA GREAT AGAIN! Happy New Year!! https://t.co/qsMNyN1UJG
## 3
Iran, the Number One State of Sponsored Terror with numerous violations of
Human Rights occurring on an hourly basis, has now closed down the Internet
so that peaceful demonstrators cannot communicate. Not good!
## 4                                                          As our Country
rapidly grows stronger and smarter, I want to wish all of my friends,
supporters, enemies, haters, and even the very dishonest Fake News Media, a
Happy and Healthy New Year. 2018 will be a great year for America!
## 5
HAPPY NEW YEAR! We are MAKING AMERICA GREAT AGAIN, and much faster than
anyone thought possible!
## 6  The United States has foolishly given Pakistan more than 33 billion
dollars in aid over the last 15 years, and they have given us nothing but
lies &amp;amp; deceit, thinking of our leaders as fools. They give safe haven to
the terrorists we hunt in Afghanistan, with little help. No more!
```

```
DfmStm <- convert(myDfm, to = "stm", docvars = docvars(myDfm))
str(DfmStm)
```

```
## List of 3
##  $ documents:List of 3509
##   ..$ text1   : int [1:2, 1:16] 100 1 284 1 1202 1 1685 1 1778 1 ...
##   ..$ text2   : int [1:2, 1:11] 860 1 2745 1 2823 1 2891 1 3279 1 ...
##   ..$ text3   : int [1:2, 1:20] 1111 1 1572 1 1644 1 1963 1 2786 1 ...
##   ..$ text4   : int [1:2, 1:21] 860 1 1779 1 2084 1 2299 1 2375 1 ...
##   ..$ text5   : int [1:2, 1:11] 860 1 911 1 2496 1 2823 1 2891 1 ...
##   ..$ text6   : int [1:2, 1:26] 785 1 807 1 871 1 1176 1 1908 1 ...
##   ..$ text7   : int [1:2, 1:27] 759 1 843 1 871 1 1458 1 1895 1 ...
##   ..$ text8   : int [1:2, 1:10] 2147 1 2597 1 2823 1 3519 1 3938 1 ...
##   ..$ text9   : int [1:2, 1:23] 733 1 1167 1 1316 1 1764 1 2552 1 ...
##   ..$ text10  : int [1:2, 1:34] 699 1 723 1 733 1 799 1 807 1 ...
##   ..$ text11  : int [1:2, 1:19] 323 1 871 1 1245 2 1268 1 1298 1 ...
```

```
##   ..$ text12  : int [1:2, 1:10] 1167 1 1174 1 1238 1 1648 1 1847 1 ...
##   ..$ text13  : int [1:2, 1:22] 1129 1 1167 1 1872 1 2569 1 2587 1 ...
##   ..$ text14  : int [1:2, 1:16] 1041 1 1151 1 1635 1 1900 1 2786 1 ...
##   ..$ text15  : int [1:2, 1:26] 765 1 1453 1 1698 1 2463 1 2503 1 ...
##   ..$ text16  : int [1:2, 1:27] 920 1 2147 1 2416 1 2469 1 2761 1 ...
##   ..$ text17  : int [1:2, 1:17] 739 1 1857 2 1954 1 1960 1 2473 1 ...
##   ..$ text18  : int [1:2, 1:19] 291 1 540 1 862 1 1370 1 1394 1 ...
##   ..$ text19  : int [1:2, 1:19] 707 1 1159 1 1395 1 1698 1 2628 1 ...
##   ..$ text20  : int [1:2, 1:24] 845 1 935 1 1176 1 1779 1 2139 1 ...
##   ..$ text21  : int [1:2, 1:18] 2697 1 3276 2 3316 1 3605 1 3674 1 ...
##   ..$ text22  : int [1:2, 1:25] 871 1 1169 1 1347 3 1974 1 1996 1 ...
##   ..$ text23  : int [1:2, 1:20] 871 2 895 1 1049 1 1067 1 1412 1 ...
##   ..$ text24  : int [1:2, 1:3] 533 1 5662 1 6044 1
##   ..$ text25  : int [1:2, 1:23] 410 1 717 2 801 1 871 1 988 1 ...
##   ..$ text26  : int [1:2, 1:15] 939 1 1061 1 1764 1 2797 1 2823 1 ...
##   ..$ text27  : int [1:2, 1:15] 1142 1 1514 1 1523 1 1890 1 1900 1 ...
##   ..$ text28  : int [1:2, 1:23] 662 1 1238 1 1472 1 1648 1 1700 1 ...
##   ..$ text29  : int [1:2, 1:26] 759 1 845 1 898 1 1156 1 1174 1 ...
##   ..$ text30  : int [1:2, 1:3] 860 1 2823 1 3674 1
##   ..$ text31  : int [1:2, 1:24] 1636 2 1882 1 1960 1 2248 1 2643 1 ...
##   ..$ text32  : int [1:2, 1:20] 722 1 841 1 863 1 1621 1 1779 1 ...
##   ..$ text33  : int [1:2, 1:24] 908 1 1138 1 1638 1 2026 1 2429 1 ...
##   ..$ text34  : int [1:2, 1:7] 1120 1 3425 1 4027 1 4370 1 4392 1 ...
##   ..$ text35  : int [1:2, 1:9] 1167 1 1697 1 1734 1 1800 1 1847 1 ...
##   ..$ text36  : int [1:2, 1:28] 871 2 1324 1 1374 1 1444 1 1457 1 ...
##   ..$ text37  : int [1:2, 1:3] 860 1 2823 1 3677 1
##   ..$ text38  : int [1:2, 1:26] 715 1 742 1 1030 2 1239 2 2144 1 ...
##   ..$ text39  : int [1:2, 1:26] 898 1 1095 1 1170 1 1240 1 1343 1 ...
##   ..$ text40  : int [1:2, 1:24] 798 1 860 1 1803 1 2139 1 2162 1 ...
##   ..$ text41  : int [1:2, 1:24] 1239 1 1567 1 1610 2 2247 1 2380 1 ...
##   ..$ text42  : int [1:2, 1:10] 1091 1 2201 1 2479 1 3431 1 3514 1 ...
##   ..$ text43  : int [1:2, 1:3] 2786 1 3073 1 4634 1
##   ..$ text44  : int [1:2, 1:28] 841 1 1067 1 1091 1 1127 1 1239 1 ...
##   ..$ text45  : int [1:2, 1:19] 789 1 863 1 2524 1 2569 1 2786 1 ...
##   ..$ text46  : int [1:2, 1:25] 697 1 838 1 1061 1 1176 1 1288 1 ...
##   ..$ text47  : int [1:2, 1:27] 863 1 1610 1 1960 1 2471 1 3002 1 ...
##   ..$ text48  : int [1:2, 1:26] 744 1 845 1 997 1 1345 1 1393 1 ...
##   ..$ text49  : int [1:2, 1:10] 2569 1 2735 2 4479 1 4597 1 5191 1 ...
##   ..$ text50  : int [1:2, 1:28] 871 1 1245 1 1322 1 1353 1 1367 1 ...
##   ..$ text51  : int [1:2, 1:20] 895 1 1030 1 1239 1 1890 1 2073 1 ...
##   ..$ text52  : int [1:2, 1:17] 759 1 1581 2 2005 1 2471 1 2602 1 ...
##   ..$ text53  : int [1:2, 1:23] 642 1 871 1 904 1 911 1 1048 2 ...
##   ..$ text54  : int [1:2, 1:21] 1146 1 1167 1 1779 1 1803 2 2139 1 ...
##   ..$ text55  : int [1:2, 1:19] 717 1 721 1 1567 2 1712 1 1890 1 ...
##   ..$ text56  : int [1:2, 1:25] 860 1 1051 1 1298 1 1457 2 1506 1 ...
##   ..$ text57  : int [1:2, 1:20] 717 1 721 1 1567 1 1715 1 1890 1 ...
##   ..$ text58  : int [1:2, 1:26] 860 1 871 1 1051 1 1093 1 1298 1 ...
##   ..$ text59  : int [1:2, 1:22] 125 1 323 1 789 1 863 1 1572 1 ...
##   ..$ text60  : int [1:2, 1:20] 7 1 312 1 631 1 690 1 753 1 ...
##   ..$ text61  : int [1:2, 1:10] 7 1 862 1 1890 2 1950 1 1951 1 ...
```

```
##   ..$ text62  : int [1:2, 1:27] 862 1 863 1 871 2 910 1 1646 1 ...
##   ..$ text63  : int [1:2, 1:11] 845 1 863 1 901 1 1779 1 2488 1 ...
##   ..$ text64  : int [1:2, 1:22] 98 1 863 1 1132 1 1214 1 2301 1 ...
##   ..$ text65  : int [1:2, 1:16] 733 1 2741 1 2823 1 2996 3 3019 1 ...
##   ..$ text66  : int [1:2, 1:20] 804 1 1174 1 1245 1 1444 1 1857 1 ...
##   ..$ text67  : int [1:2, 1:22] 1061 1 1394 1 1536 1 1993 1 2314 2 ...
##   ..$ text68  : int [1:2, 1:10] 526 1 1146 1 1529 1 1847 1 4646 1 ...
##   ..$ text69  : int [1:2, 1:13] 940 1 1245 1 1557 1 1779 1 1857 1 ...
##   ..$ text70  : int [1:2, 1:10] 104 1 338 1 803 1 1769 1 3405 1 ...
##   ..$ text71  : int [1:2, 1:18] 98 1 358 1 493 1 799 1 864 1 ...
##   ..$ text72  : int [1:2, 1:19] 692 1 841 1 850 2 1305 1 1405 1 ...
##   ..$ text73  : int [1:2, 1:24] 1030 1 1610 1 2028 1 2080 1 2460 1 ...
##   ..$ text74  : int [1:2, 1:24] 863 1 871 2 1610 2 1734 1 1742 1 ...
##   ..$ text75  : int [1:2, 1:19] 248 1 725 1 806 1 860 1 1132 1 ...
##   ..$ text76  : int [1:2, 1:17] 625 1 837 1 1688 1 2342 1 2673 1 ...
##   ..$ text77  : int [1:2, 1:14] 1 1 940 1 1245 1 1779 1 1857 1 ...
##   ..$ text78  : int [1:2, 1:20] 817 1 860 1 895 1 1176 1 1298 1 ...
##   ..$ text79  : int [1:2, 1:22] 817 2 895 1 1061 1 1167 1 1388 1 ...
##   ..$ text80  : int [1:2, 1:27] 323 1 845 1 1368 1 1567 1 1821 1 ...
##   ..$ text81  : int [1:2, 1:12] 2228 1 2378 1 2400 1 2520 1 2786 1 ...
##   ..$ text82  : int [1:2, 1:20] 709 1 733 2 759 1 1067 1 1368 1 ...
##   ..$ text83  : int [1:2, 1:27] 668 1 745 1 895 1 1176 1 1238 1 ...
##   ..$ text84  : int [1:2, 1:5] 323 1 3084 1 3100 1 3631 1 6214 2
##   ..$ text85  : int [1:2, 1:19] 1067 1 2051 1 2578 1 2622 3 2745 1 ...
##   ..$ text86  : int [1:2, 1:8] 733 1 871 1 1849 1 2823 1 3329 1 ...
##   ..$ text87  : int [1:2, 1:22] 84 1 749 1 1182 1 1370 1 1618 1 ...
##   ..$ text88  : int [1:2, 1:13] 323 1 746 1 1779 1 2145 1 2378 1 ...
##   ..$ text89  : int [1:2, 1:12] 940 1 1129 1 1343 2 2981 1 2997 1 ...
##   ..$ text90  : int [1:2, 1:21] 1521 2 1914 1 2611 1 2891 1 3604 1 ...
##   ..$ text91  : int [1:2, 1:21] 1030 1 1196 1 1597 1 1639 2 1914 1 ...
##   ..$ text92  : int [1:2, 1:28] 863 1 871 1 1245 1 1324 1 1444 1 ...
##   ..$ text93  : int [1:2, 1:25] 759 1 1067 1 1151 1 1167 1 1176 1 ...
##   ..$ text94  : int [1:2, 1:29] 871 1 1066 1 1167 1 1182 1 1363 1 ...
##   ..$ text95  : int [1:2, 1:22] 1067 1 1107 1 1779 2 1940 1 1954 1 ...
##   ..$ text96  : int [1:2, 1:23] 860 1 1066 1 1167 1 1857 1 1954 1 ...
##   ..$ text97  : int [1:2, 1:11] 1167 1 1857 2 3472 2 3658 1 3776 1 ...
##   ..$ text98  : int [1:2, 1:19] 1394 1 1779 1 1857 1 1895 2 1960 1 ...
##   ..$ text99  : int [1:2, 1:21] 912 1 1779 1 1954 1 1983 1 2697 1 ...
##   .. [list output truncated]
##  $ vocab    : chr [1:6346] "#1" "#243navybday" "#24th" "#500day" ...
##  $ meta     :'data.frame':   3509 obs. of  5 variables:
##   ..$ ID       : chr [1:3509] "@realDonaldTrump" "@realDonaldTrump"
"@realDonaldTrump" "@realDonaldTrump" ...
##   ..$ Time     : chr [1:3509] " 2018-01-01 02:36" " 2018-01-01 03:06" "
2018-01-01 06:00" " 2018-01-01 06:18" ...
##   ..$ Tweet.URL: chr [1:3509] "
https://twitter.com/realDonaldTrump/status/947536951464333318" "
https://twitter.com/realDonaldTrump/status/947544600918372353" "
https://twitter.com/realDonaldTrump/status/947588263103139841" "
https://twitter.com/realDonaldTrump/status/947592785519173637" ...
```

```
##   ..$ date     : num [1:3509] 17532 17532 17532 17532 17532 ...
##   ..$ Text2    : chr [1:3509] " My deepest condolences to the victims of
the terrible shooting in Douglas County @DCSheriff, and their familie"|
__truncated__ " What a year it’s been, and we're just getting started.
Together, we are MAKING AMERICA GREAT AGAIN! Happy New "| __truncated__ "
Iran, the Number One State of Sponsored Terror with numerous violations of
Human Rights occurring on an hourly"| __truncated__ " As our Country rapidly
grows stronger and smarter, I want to wish all of my friends, supporters,
enemies, hate"| __truncated__ ...
```

```
str(DfmStm$meta)
```

```
## 'data.frame':    3509 obs. of  5 variables:
##  $ ID       : chr  "@realDonaldTrump" "@realDonaldTrump"
"@realDonaldTrump" "@realDonaldTrump" ...
##  $ Time     : chr  " 2018-01-01 02:36" " 2018-01-01 03:06" " 2018-01-01
06:00" " 2018-01-01 06:18" ...
##  $ Tweet.URL: chr  "
https://twitter.com/realDonaldTrump/status/947536951464333318" "
https://twitter.com/realDonaldTrump/status/947544600918372353" "
https://twitter.com/realDonaldTrump/status/947588263103139841" "
https://twitter.com/realDonaldTrump/status/947592785519173637" ...
##  $ date     : num  17532 17532 17532 17532 17532 ...
##  $ Text2    : chr  " My deepest condolences to the victims of the terrible
shooting in Douglas County @DCSheriff, and their familie"| __truncated__ "
What a year it’s been, and we're just getting started. Together, we are
MAKING AMERICA GREAT AGAIN! Happy New "| __truncated__ " Iran, the Number One
State of Sponsored Terror with numerous violations of Human Rights occurring
on an hourly"| __truncated__ " As our Country rapidly grows stronger and
smarter, I want to wish all of my friends, supporters, enemies, hate"|
__truncated__ ...
```

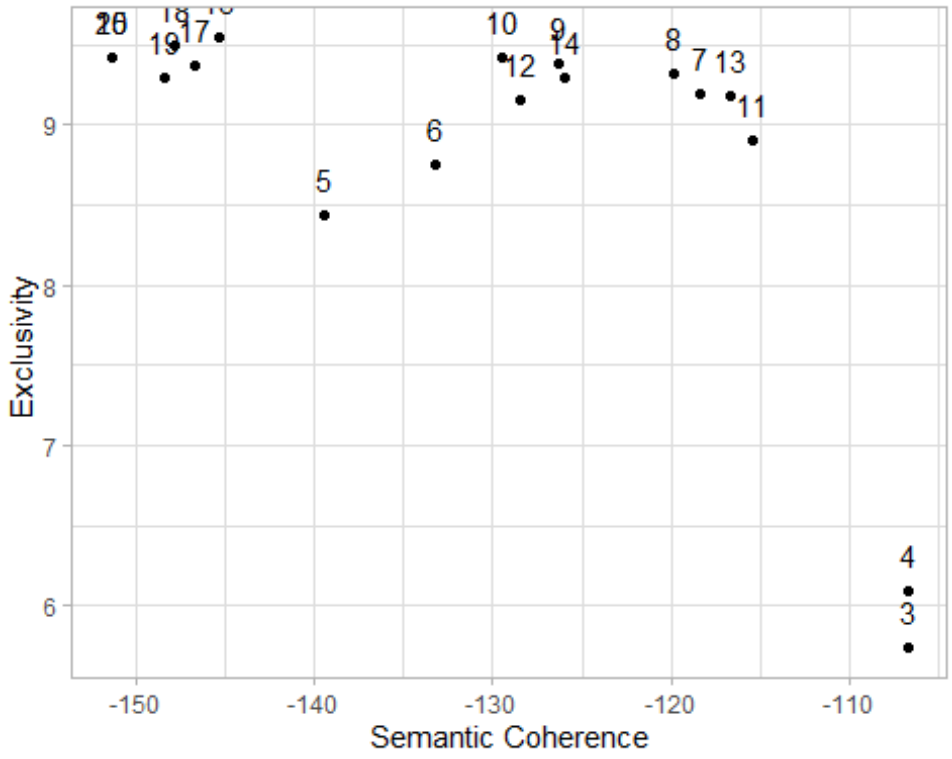
```
set.seed(02138)
K <-c(3:20)
system.time(storage  <- searchK(DfmStm $documents, DfmStm $vocab, K,
max.em.its = 250,
                                prevalence = ~ s(date), data = DfmStm$meta,
init.type = "Spectral")) #non-linear relationship is expected
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ...
##   Recovering initialization...
##      ..........................................................
## Initialization complete.
## ##    user  system elapsed
## 1613.53  177.76 2584.77
```

```
results <- data.frame(Coherence=unlist(storage$results$semcoh),
Exclusivity=unlist(storage$results$exclus), K=unlist(storage$results$K))
results
```

```
##    Coherence Exclusivity  K
## 1  -106.7946    5.742436  3
## 2  -106.7946    6.099761  4
## 3  -139.3932    8.443715  5
## 4  -133.2038    8.749674  6
## 5  -118.4191    9.191041  7
## 6  -119.8715    9.315203  8
## 7  -126.3211    9.385999  9
## 8  -129.5197    9.416078 10
## 9  -115.4150    8.900017 11
## 10 -128.5008    9.154457 12
## 11 -116.7163    9.181902 13
## 12 -126.0269    9.301428 14
## 13 -151.3616    9.423738 15
## 14 -145.3786    9.548823 16
## 15 -146.6465    9.368827 17
## 16 -147.8201    9.499575 18
## 17 -148.4001    9.299864 19
## 18 -151.2954    9.421346 20
```

```
ggplot(results , aes(x=Coherence, y=Exclusivity)) + geom_point() +
  geom_text(label=results$K, vjust=-1) +
  ylab(label="Exclusivity ") +  xlab("Semantic Coherence")  +
  theme_light()
```




```
str(DfmStm)
```

```
## List of 3
##  $ documents:List of 3509
##   ..$ text1   : int [1:2, 1:16] 100 1 284 1 1202 1 1685 1 1778 1 ...
##   ..$ text2   : int [1:2, 1:11] 860 1 2745 1 2823 1 2891 1 3279 1 ...
##   ..$ text3   : int [1:2, 1:20] 1111 1 1572 1 1644 1 1963 1 2786 1 ...
##   ..$ text4   : int [1:2, 1:21] 860 1 1779 1 2084 1 2299 1 2375 1 ...
##   ..$ text5   : int [1:2, 1:11] 860 1 911 1 2496 1 2823 1 2891 1 ...
##   ..$ text6   : int [1:2, 1:26] 785 1 807 1 871 1 1176 1 1908 1 ...
##   ..$ text7   : int [1:2, 1:27] 759 1 843 1 871 1 1458 1 1895 1 ...
##   ..$ text8   : int [1:2, 1:10] 2147 1 2597 1 2823 1 3519 1 3938 1 ...
##   ..$ text9   : int [1:2, 1:23] 733 1 1167 1 1316 1 1764 1 2552 1 ...
##   ..$ text10  : int [1:2, 1:34] 699 1 723 1 733 1 799 1 807 1 ...
##   ..$ text11  : int [1:2, 1:19] 323 1 871 1 1245 2 1268 1 1298 1 ...
##   ..$ text12  : int [1:2, 1:10] 1167 1 1174 1 1238 1 1648 1 1847 1 ...
##   ..$ text13  : int [1:2, 1:22] 1129 1 1167 1 1872 1 2569 1 2587 1 ...
##   ..$ text14  : int [1:2, 1:16] 1041 1 1151 1 1635 1 1900 1 2786 1 ...
##   ..$ text15  : int [1:2, 1:26] 765 1 1453 1 1698 1 2463 1 2503 1 ...
##   ..$ text16  : int [1:2, 1:27] 920 1 2147 1 2416 1 2469 1 2761 1 ...
##   ..$ text17  : int [1:2, 1:17] 739 1 1857 2 1954 1 1960 1 2473 1 ...
##   ..$ text18  : int [1:2, 1:19] 291 1 540 1 862 1 1370 1 1394 1 ...
##   ..$ text19  : int [1:2, 1:19] 707 1 1159 1 1395 1 1698 1 2628 1 ...
##   ..$ text20  : int [1:2, 1:24] 845 1 935 1 1176 1 1779 1 2139 1 ...
##   ..$ text21  : int [1:2, 1:18] 2697 1 3276 2 3316 1 3605 1 3674 1 ...
##   ..$ text22  : int [1:2, 1:25] 871 1 1169 1 1347 3 1974 1 1996 1 ...
##   ..$ text23  : int [1:2, 1:20] 871 2 895 1 1049 1 1067 1 1412 1 ...
```

```
##   ..$ text24  : int [1:2, 1:3] 533 1 5662 1 6044 1
##   ..$ text25  : int [1:2, 1:23] 410 1 717 2 801 1 871 1 988 1 ...
##   ..$ text26  : int [1:2, 1:15] 939 1 1061 1 1764 1 2797 1 2823 1 ...
##   ..$ text27  : int [1:2, 1:15] 1142 1 1514 1 1523 1 1890 1 1900 1 ...
##   ..$ text28  : int [1:2, 1:23] 662 1 1238 1 1472 1 1648 1 1700 1 ...
##   ..$ text29  : int [1:2, 1:26] 759 1 845 1 898 1 1156 1 1174 1 ...
##   ..$ text30  : int [1:2, 1:3] 860 1 2823 1 3674 1
##   ..$ text31  : int [1:2, 1:24] 1636 2 1882 1 1960 1 2248 1 2643 1 ...
##   ..$ text32  : int [1:2, 1:20] 722 1 841 1 863 1 1621 1 1779 1 ...
##   ..$ text33  : int [1:2, 1:24] 908 1 1138 1 1638 1 2026 1 2429 1 ...
##   ..$ text34  : int [1:2, 1:7] 1120 1 3425 1 4027 1 4370 1 4392 1 ...
##   ..$ text35  : int [1:2, 1:9] 1167 1 1697 1 1734 1 1800 1 1847 1 ...
##   ..$ text36  : int [1:2, 1:28] 871 2 1324 1 1374 1 1444 1 1457 1 ...
##   ..$ text37  : int [1:2, 1:3] 860 1 2823 1 3677 1
##   ..$ text38  : int [1:2, 1:26] 715 1 742 1 1030 2 1239 2 2144 1 ...
##   ..$ text39  : int [1:2, 1:26] 898 1 1095 1 1170 1 1240 1 1343 1 ...
##   ..$ text40  : int [1:2, 1:24] 798 1 860 1 1803 1 2139 1 2162 1 ...
##   ..$ text41  : int [1:2, 1:24] 1239 1 1567 1 1610 2 2247 1 2380 1 ...
##   ..$ text42  : int [1:2, 1:10] 1091 1 2201 1 2479 1 3431 1 3514 1 ...
##   ..$ text43  : int [1:2, 1:3] 2786 1 3073 1 4634 1
##   ..$ text44  : int [1:2, 1:28] 841 1 1067 1 1091 1 1127 1 1239 1 ...
##   ..$ text45  : int [1:2, 1:19] 789 1 863 1 2524 1 2569 1 2786 1 ...
##   ..$ text46  : int [1:2, 1:25] 697 1 838 1 1061 1 1176 1 1288 1 ...
##   ..$ text47  : int [1:2, 1:27] 863 1 1610 1 1960 1 2471 1 3002 1 ...
##   ..$ text48  : int [1:2, 1:26] 744 1 845 1 997 1 1345 1 1393 1 ...
##   ..$ text49  : int [1:2, 1:10] 2569 1 2735 2 4479 1 4597 1 5191 1 ...
##   ..$ text50  : int [1:2, 1:28] 871 1 1245 1 1322 1 1353 1 1367 1 ...
##   ..$ text51  : int [1:2, 1:20] 895 1 1030 1 1239 1 1890 1 2073 1 ...
##   ..$ text52  : int [1:2, 1:17] 759 1 1581 2 2005 1 2471 1 2602 1 ...
##   ..$ text53  : int [1:2, 1:23] 642 1 871 1 904 1 911 1 1048 2 ...
##   ..$ text54  : int [1:2, 1:21] 1146 1 1167 1 1779 1 1803 2 2139 1 ...
##   ..$ text55  : int [1:2, 1:19] 717 1 721 1 1567 2 1712 1 1890 1 ...
##   ..$ text56  : int [1:2, 1:25] 860 1 1051 1 1298 1 1457 2 1506 1 ...
##   ..$ text57  : int [1:2, 1:20] 717 1 721 1 1567 1 1715 1 1890 1 ...
##   ..$ text58  : int [1:2, 1:26] 860 1 871 1 1051 1 1093 1 1298 1 ...
##   ..$ text59  : int [1:2, 1:22] 125 1 323 1 789 1 863 1 1572 1 ...
##   ..$ text60  : int [1:2, 1:20] 7 1 312 1 631 1 690 1 753 1 ...
##   ..$ text61  : int [1:2, 1:10] 7 1 862 1 1890 2 1950 1 1951 1 ...
##   ..$ text62  : int [1:2, 1:27] 862 1 863 1 871 2 910 1 1646 1 ...
##   ..$ text63  : int [1:2, 1:11] 845 1 863 1 901 1 1779 1 2488 1 ...
##   ..$ text64  : int [1:2, 1:22] 98 1 863 1 1132 1 1214 1 2301 1 ...
##   ..$ text65  : int [1:2, 1:16] 733 1 2741 1 2823 1 2996 3 3019 1 ...
##   ..$ text66  : int [1:2, 1:20] 804 1 1174 1 1245 1 1444 1 1857 1 ...
##   ..$ text67  : int [1:2, 1:22] 1061 1 1394 1 1536 1 1993 1 2314 2 ...
##   ..$ text68  : int [1:2, 1:10] 526 1 1146 1 1529 1 1847 1 4646 1 ...
##   ..$ text69  : int [1:2, 1:13] 940 1 1245 1 1557 1 1779 1 1857 1 ...
##   ..$ text70  : int [1:2, 1:10] 104 1 338 1 803 1 1769 1 3405 1 ...
##   ..$ text71  : int [1:2, 1:18] 98 1 358 1 493 1 799 1 864 1 ...
##   ..$ text72  : int [1:2, 1:19] 692 1 841 1 850 2 1305 1 1405 1 ...
##   ..$ text73  : int [1:2, 1:24] 1030 1 1610 1 2028 1 2080 1 2460 1 ...
```

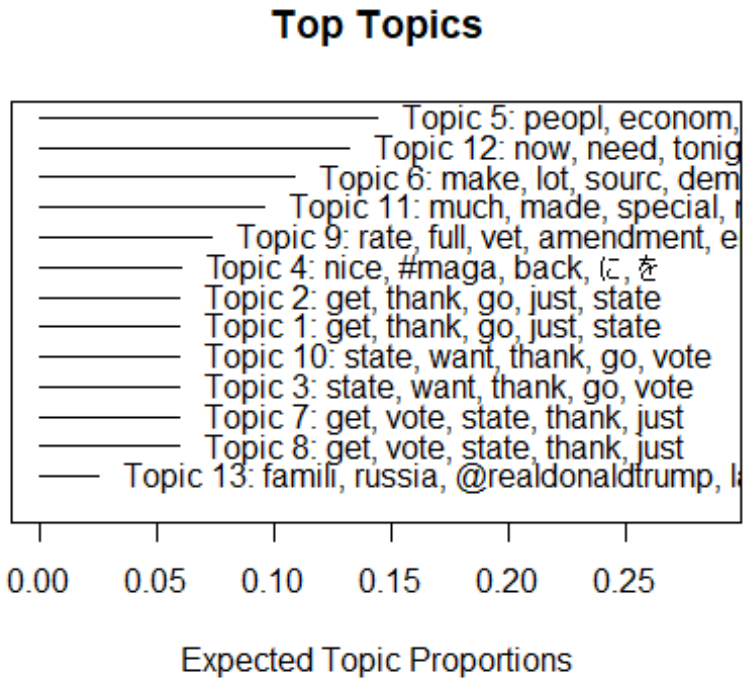
```
##   ..$ text74  : int [1:2, 1:24] 863 1 871 2 1610 2 1734 1 1742 1 ...
##   ..$ text75  : int [1:2, 1:19] 248 1 725 1 806 1 860 1 1132 1 ...
##   ..$ text76  : int [1:2, 1:17] 625 1 837 1 1688 1 2342 1 2673 1 ...
##   ..$ text77  : int [1:2, 1:14] 1 1 940 1 1245 1 1779 1 1857 1 ...
##   ..$ text78  : int [1:2, 1:20] 817 1 860 1 895 1 1176 1 1298 1 ...
##   ..$ text79  : int [1:2, 1:22] 817 2 895 1 1061 1 1167 1 1388 1 ...
##   ..$ text80  : int [1:2, 1:27] 323 1 845 1 1368 1 1567 1 1821 1 ...
##   ..$ text81  : int [1:2, 1:12] 2228 1 2378 1 2400 1 2520 1 2786 1 ...
##   ..$ text82  : int [1:2, 1:20] 709 1 733 2 759 1 1067 1 1368 1 ...
##   ..$ text83  : int [1:2, 1:27] 668 1 745 1 895 1 1176 1 1238 1 ...
##   ..$ text84  : int [1:2, 1:5] 323 1 3084 1 3100 1 3631 1 6214 2
##   ..$ text85  : int [1:2, 1:19] 1067 1 2051 1 2578 1 2622 3 2745 1 ...
##   ..$ text86  : int [1:2, 1:8] 733 1 871 1 1849 1 2823 1 3329 1 ...
##   ..$ text87  : int [1:2, 1:22] 84 1 749 1 1182 1 1370 1 1618 1 ...
##   ..$ text88  : int [1:2, 1:13] 323 1 746 1 1779 1 2145 1 2378 1 ...
##   ..$ text89  : int [1:2, 1:12] 940 1 1129 1 1343 2 2981 1 2997 1 ...
##   ..$ text90  : int [1:2, 1:21] 1521 2 1914 1 2611 1 2891 1 3604 1 ...
##   ..$ text91  : int [1:2, 1:21] 1030 1 1196 1 1597 1 1639 2 1914 1 ...
##   ..$ text92  : int [1:2, 1:28] 863 1 871 1 1245 1 1324 1 1444 1 ...
##   ..$ text93  : int [1:2, 1:25] 759 1 1067 1 1151 1 1167 1 1176 1 ...
##   ..$ text94  : int [1:2, 1:29] 871 1 1066 1 1167 1 1182 1 1363 1 ...
##   ..$ text95  : int [1:2, 1:22] 1067 1 1107 1 1779 2 1940 1 1954 1 ...
##   ..$ text96  : int [1:2, 1:23] 860 1 1066 1 1167 1 1857 1 1954 1 ...
##   ..$ text97  : int [1:2, 1:11] 1167 1 1857 2 3472 2 3658 1 3776 1 ...
##   ..$ text98  : int [1:2, 1:19] 1394 1 1779 1 1857 1 1895 2 1960 1 ...
##   ..$ text99  : int [1:2, 1:21] 912 1 1779 1 1954 1 1983 1 2697 1 ...
##   .. [list output truncated]
##  $ vocab    : chr [1:6346] "#1" "#243navybday" "#24th" "#500day" ...
##  $ meta     :'data.frame':   3509 obs. of  5 variables:
##   ..$ ID       : chr [1:3509] "@realDonaldTrump" "@realDonaldTrump"
"@realDonaldTrump" "@realDonaldTrump" ...
##   ..$ Time     : chr [1:3509] " 2018-01-01 02:36" " 2018-01-01 03:06" "
2018-01-01 06:00" " 2018-01-01 06:18" ...
##   ..$ Tweet.URL: chr [1:3509] "
https://twitter.com/realDonaldTrump/status/947536951464333318" "
https://twitter.com/realDonaldTrump/status/947544600918372353" "
https://twitter.com/realDonaldTrump/status/947588263103139841" "
https://twitter.com/realDonaldTrump/status/947592785519173637" ...
##   ..$ date     : num [1:3509] 17532 17532 17532 17532 17532 ...
##   ..$ Text2    : chr [1:3509] " My deepest condolences to the victims of
the terrible shooting in Douglas County @DCSheriff, and their familie"|
__truncated__ " What a year it’s been, and we're just getting started.
Together, we are MAKING AMERICA GREAT AGAIN! Happy New "| __truncated__ "
Iran, the Number One State of Sponsored Terror with numerous violations of
Human Rights occurring on an hourly"| __truncated__ " As our Country rapidly
grows stronger and smarter, I want to wish all of my friends, supporters,
enemies, hate"| __truncated__ ...
```

```
################## let's try K = 13 number of topics

system.time(stmFitted <- stm(DfmStm $documents, DfmStm $vocab, K = 13,
max.em.its = 250,
                             prevalence = ~ s(date),  data = DfmStm $meta,
init.type = "Spectral"))

## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      .............
##   Recovering initialization...
##      ............................................................
## Initialization complete.
##
.....................................................................................
.....................
## Completed E-Step (2 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -7.517)
##
.....................................................................................
.....................
## Completed E-Step (1 seconds).
## Completed M-Step.
## Model Converged

##    user  system elapsed
##   26.39    8.09   44.47

labelTopics(stmFitted, n=7) # 7 features for each topic

## Topic 1 Top Words:
##       Highest Prob: great, countri, get, state, just, vote, thank
##       FREX: get, thank, go, just, state, trade, vote
##       Lift: #autismawarenessday, rage, korea', startl, botch, slime,
amendmen
##       Score: great, get, vote, thank, want, countri, go
## Topic 2 Top Words:
##       Highest Prob: great, countri, get, state, just, vote, thank
##       FREX: get, thank, go, just, state, trade, vote
##       Lift: #lightitupblu, rage, korea', startl, botch, slime, amendmen
##       Score: great, get, vote, thank, want, countri, go
## Topic 3 Top Words:
##       Highest Prob: great, countri, state, get, just, want, vote
##       FREX: state, want, thank, go, vote, get, just
##       Lift: #madeinamerica, everybody', piggi, @wrbethesda, clay,
louisiana, @clayhiggins18
##       Score: great, vote, want, get, thank, countri, state
## Topic 4 Top Words:
##       Highest Prob: american, new, back, good, world, #maga, nice
```

```
##       FREX: nice, #maga, back, に, を, world, particip
##       Lift: jeffress, #sbagala, @sbalist, finland, ま, は, 問題
##       Score: american, back, world, good, new, #maga, nice
## Topic 5 Top Words:
##       Highest Prob: peopl, trump, great, presid, rt, amp, work
##       FREX: peopl, econom, done, donald, trump, iran, presid
##       Lift: #september11th, premis, incident, teamwork, asa,
@asahutchinson, arkansa
##       Score: trump, peopl, great, presid, rt, done, amp
## Topic 6 Top Words:
##       Highest Prob: make, dem, total, hunt, witch, sourc, report
##       FREX: make, lot, sourc, dem, rebuild, african, usmca
##       Lift: constitut, @senatordol, easter, @uscg, #prouddeplor, birthday,
thanksgiving
##       Score: make, dem, sourc, lot, total, rebuild, witch
## Topic 7 Top Words:
##       Highest Prob: great, get, countri, state, just, vote, thank
##       FREX: get, vote, state, thank, just, go, want
##       Lift: locations, reckon, poli, @debbiestabenow, paus, data,
@braun4indiana
##       Score: great, vote, get, thank, want, go, state
## Topic 8 Top Words:
##       Highest Prob: great, get, countri, state, just, vote, thank
##       FREX: get, vote, state, thank, just, go, want
##       Lift: polling, reckon, poli, @debbiestabenow, paus, data,
@braun4indiana
##       Score: great, vote, get, thank, want, go, state
## Topic 9 Top Words:
##       Highest Prob: great, peopl, rt, trump, rate, love, amp
##       FREX: rate, full, vet, amendment, endorsement, love, second
##       Lift: @vdem, vdot, prevail, bee, samantha, active, barricad
##       Score: trump, peopl, rate, great, love, rt, full
## Topic 10 Top Words:
##       Highest Prob: great, countri, state, get, just, want, vote
##       FREX: state, want, thank, go, vote, get, just
##       Lift: showcas, everybody’, piggi, @wrbethesda, clay, louisiana,
@clayhiggins18
##       Score: great, vote, want, get, thank, countri, state
## Topic 11 Top Words:
##       Highest Prob: much, made, noth, elect, success, love, special
##       FREX: much, made, special, nuclear, success, noth, current
##       Lift: wholeheart, much, yountvill, @ainsleyearhardt, nuke, special,
employment
##       Score: much, made, special, success, noth, love, elect
## Topic 12 Top Words:
##       Highest Prob: now, democrat, need, border, amp, wall, secur
##       FREX: now, need, tonight, week, receiv, wall, money
##       Lift: now, □□, □□, enabl, heroin, synthet, □□
##       Score: now, need, democrat, border, amp, wall, senat
```

```
## Topic 13 Top Words:
##       Highest Prob: @realdonaldtrump, republican, russia, famili,
administr, start, trump
##       FREX: famili, russia, @realdonaldtrump, lawyer, thought, god,
republican
##       Lift: lawyer, arizona, change, vega, thought,
#makeamericagreatagain, god
##       Score: @realdonaldtrump, russia, famili, republican, administr,
trump, start
```

```
plot(stmFitted, type = "summary", labeltype = c("frex"), n=5)  # topic 10 is
the most frequent one
```



#topics 7 and 8 seem to overlap, as well as topics 2 and 1. Should number of K topics be reduced?

```
################## let's try K = 11

system.time(stmFitted <- stm(DfmStm $documents, DfmStm $vocab, K = 11,
max.em.its = 250,
                             prevalence = ~ s(date),  data = DfmStm $meta,
init.type = "Spectral"))

## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ..........
```

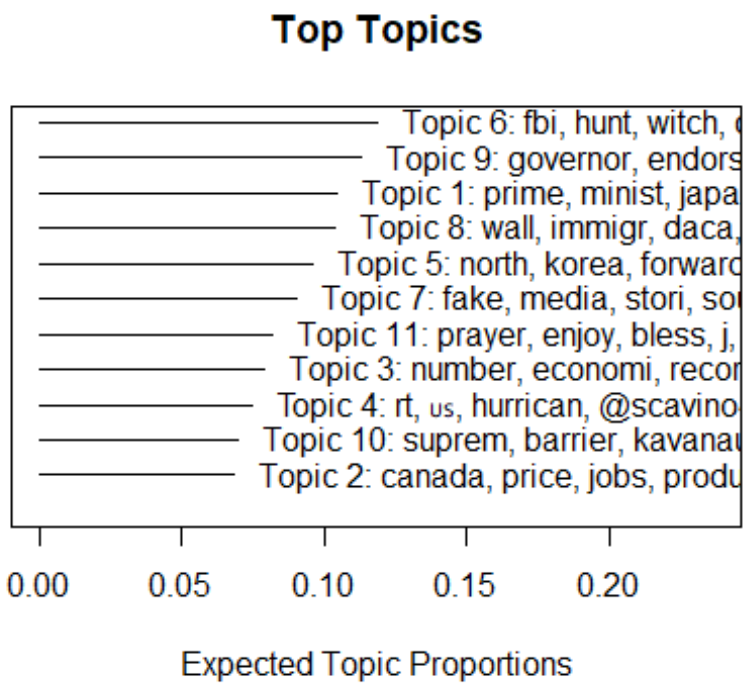
```
##   Recovering initialization...
##      ..........................................................
## Initialization complete.
##

##    user  system elapsed
##  303.94    9.19  394.12
```

```
labelTopics(stmFitted, n=7) # 7 features for each topic
```

```
## Topic 1 Top Words:
##       Highest Prob: great, thank, today, american, nation, america, honor
##       FREX: prime, minist, japan, today, honor, behalf, thank
##       Lift: #243navybday, #500day, #afghanistan, #armedforcesday,
#autismawarenessday, #confirmgina, #endtraffick
##       Score: honor, thank, @whitehous, today, american, minist, nation
## Topic 2 Top Words:
##       Highest Prob: back, deal, mani, come, amp, tax, make
##       FREX: canada, price, jobs, product, nafta, back, deal
##       Lift: #taxcutsandjobsact, apple, auto, bet, business, declin,
journey
##       Score: deal, tax, compani, back, jobs, mexico, canada
## Topic 3 Top Words:
##       Highest Prob: trump, presid, year, ever, job, better, number
##       FREX: number, economi, record, sinc, unemploy, hit, poll
##       Lift: consum, #trumptim, @anna_giaritelli, @barackobama,
@breitbartnew, @cnbc, @cortesstev
##       Score: unemploy, economi, lowest, record, ever, number, market
## Topic 4 Top Words:
##       Highest Prob: rt, @realdonaldtrump, □□, hurrican, @scavino45, local,
amp
##       FREX: rt, □□, hurrican, @scavino45, storm, #florenc, flood
##       Lift: #fakenew, #florencenc, #hurrican, #makeamericagreatagain,
#youthsport, @coreystewartva, @femaespanol
##       Score: rt, □□, @realdonaldtrump, #florenc, hurrican, storm, @fema
## Topic 5 Top Words:
##       Highest Prob: look, see, time, meet, north, good, great
##       FREX: north, korea, forward, kim, iran, jong, summit
##       Lift: tiger, #september11th, asia, assad, assembl, bore, button
##       Score: korea, north, kim, jong, un, meet, forward
## Topic 6 Top Words:
##       Highest Prob: fbi, hunt, witch, trump, collus, campaign, amp
##       FREX: fbi, hunt, witch, collus, hillari, russian, clinton
##       Lift: #spygate, @greggjarrett, @judicialwatch, agenci, appar, basic,
burr
##       Score: witch, hunt, fbi, collus, mueller, hillari, comey
## Topic 7 Top Words:
##       Highest Prob: new, fake, media, peopl, even, call, amp
##       FREX: fake, media, stori, sourc, cnn, fals, quot
##       Lift: #ny22, #remembering41, @amykrem, @bretbaier, @carlbernstein,
```

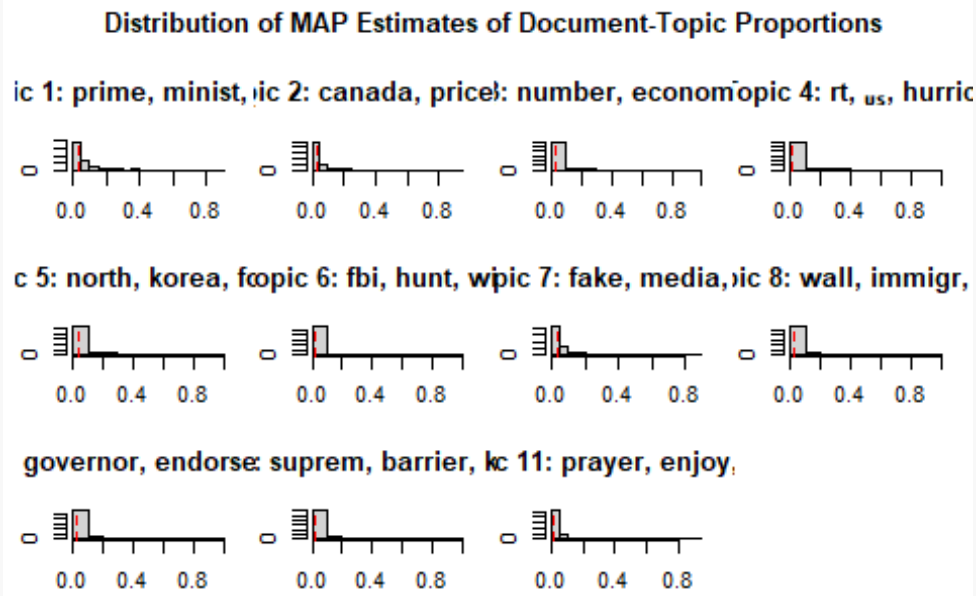
```
@claudiatenney, @katiepavlich
##       Score: fake, media, stori, new, book, cnn, sourc
## Topic 8 Top Words:
##       Highest Prob: border, democrat, want, secur, wall, must, countri
##       FREX: wall, immigr, daca, ice, secur, legisl, shutdown
##       Lift: #buildthewal, #changethelaw, #changethelaws, #itsallpolit,
#schumershutdown, #weneedmorerepublicansin18, @jackposobiec
##       Score: border, wall, democrat, secur, daca, immigr, law
## Topic 9 Top Words:
##       Highest Prob: great, vote, strong, love, senat, crime, job
##       FREX: governor, endorsement, vet, congressman, amendment, 2nd,
fantast
##       Lift: #magaralli, #az08, #broward, #farmbil, #fortifyfl,
#jobsnotmobs, #ndsen
##       Score: endorsement, vote, vet, governor, border, crime, congressman
## Topic 10 Top Words:
##       Highest Prob: trade, state, countri, billion, pay, unite, dollar
##       FREX: suprem, barrier, kavanaugh, nato, union, treat, european
##       Lift: #finland, #madeinamerica, #nato, #natosummit2018,
#pledgetoamericaswork, #scotus, @secazar
##       Score: trade, tariff, billion, china, pay, dollar, european
## Topic 11 Top Words:
##       Highest Prob: first, famili, live, hous, white, @foxnew, trump
##       FREX: prayer, enjoy, bless, j, shoot, ladi, victim
##       Lift: @usmc, @usnationalguard, @wvgovernor, adept, admiral, al,
annual
##       Score: prayer, god, melania, bless, victim, @flotus, ladi
```

```
plot(stmFitted, type = "summary", labeltype = c("frex"), n=5)  # topic 10 is
the most frequent one
```



#topics do not seem to overlap and contain unique terms

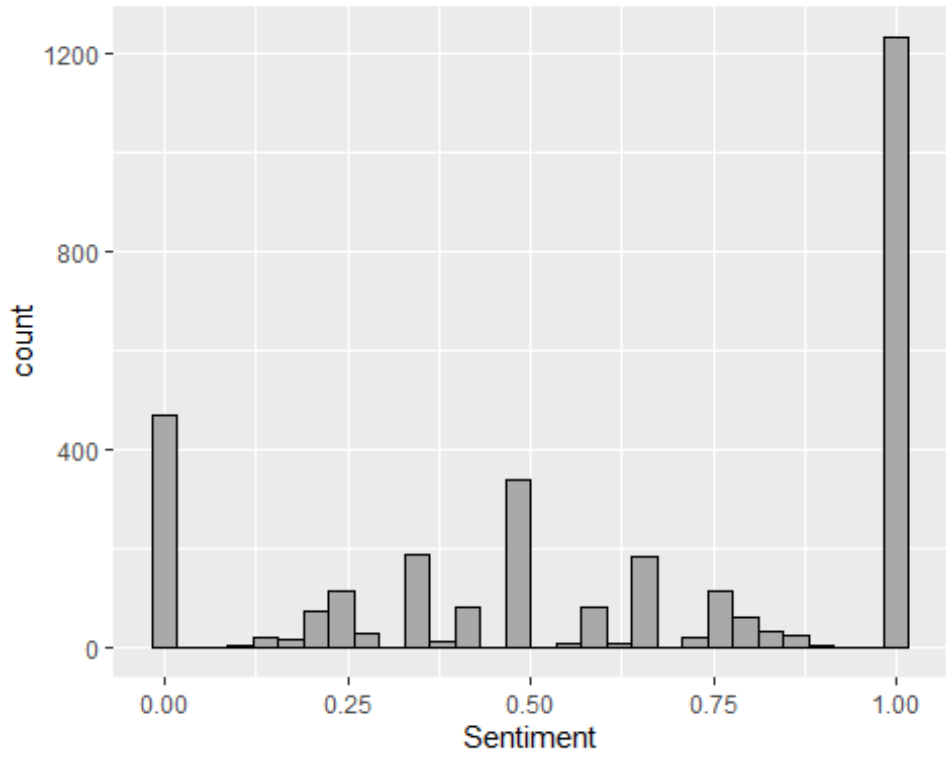
```
plot(stmFitted, type = "hist", labeltype = c("frex")) # Here topic 1 appears
as more "evenly" distributed across documents than topic 4 for example
```



Sentiment Analysis: applying sentiments dictionary

```
lengths(data_dictionary_LSD2015)

##     negative     positive neg_positive neg_negative
##         2858         1709         1721         2860

head(data_dictionary_LSD2015)

## Dictionary object with 4 key entries.
## - [negative]:
##   - a lie, abandon*, abas*, abattoir*, abdicat*, aberra*, abhor*, abject*,
abnormal*, abolish*, abominab*, abominat*, abrasiv*, absent*, abstrus*,
absurd*, abus*, accident*, accost*, accursed* [ ... and 2,838 more ]
## - [positive]:
##   - ability*, abound*, absolv*, absorbent*, absorption*, abundanc*,
abundant*, acced*, accentuat*, accept*, accessib*, acclaim*, acclamation*,
accolad*, accommodat*, accomplish*, accord, accordan*, accorded*, accords [
... and 1,689 more ]
## - [neg_positive]:
##   - best not, better not, no damag*, no no, not ability*, not able, not
```

```
abound*, not absolv*, not absorbent*, not absorption*, not abundanc*, not
abundant*, not acced*, not accentuat*, not accept*, not accessib*, not
acclaim*, not acclamation*, not accolad*, not accommodat* [ ... and 1,701
more ]
## - [neg_negative]:
##   - not a lie, not abandon*, not abas*, not abattoir*, not abdicat*, not
aberra*, not abhor*, not abject*, not abnormal*, not abolish*, not abominab*,
not abominat*, not abrasiv*, not absent*, not abstrus*, not absurd*, not
abus*, not accident*, not accost*, not accursed* [ ... and 2,840 more ]

is.dictionary(data_dictionary_LSD2015)

## [1] TRUE

sentiment <- dfm_lookup(myDfm  , dictionary = data_dictionary_LSD2015[1:2])
sentiment

## Document-feature matrix of: 3,509 documents, 2 features (35.54% sparse)
and 5 docvars.
##        features
## docs    negative positive
##   text1        2        2
##   text2        0        2
##   text3        2        2
##   text4        4        7
##   text5        0        2
##   text6        4        4
## [ reached max_ndoc ... 3,503 more documents ]

Dictionary <-convert(sentiment , to="data.frame")
str(Dictionary )

## 'data.frame':    3509 obs. of  3 variables:
##  $ doc_id  : chr  "text1" "text2" "text3" "text4" ...
##  $ negative: num  2 0 2 4 0 4 3 0 4 3 ...
##  $ positive: num  2 2 2 7 2 4 4 1 1 1 ...

Dictionary$Sentiment <- Dictionary$posit/(Dictionary$negat+Dictionary$posit)
str(Dictionary )

## 'data.frame':    3509 obs. of  4 variables:
##  $ doc_id   : chr  "text1" "text2" "text3" "text4" ...
##  $ negative : num  2 0 2 4 0 4 3 0 4 3 ...
##  $ positive : num  2 2 2 7 2 4 4 1 1 1 ...
##  $ Sentiment: num  0.5 1 0.5 0.636 1 ...

ggplot(Dictionary, aes(x=Sentiment))+
  geom_histogram(color="black", fill="darkgrey")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 395 rows containing non-finite values (`stat_bin()`).
```



#Trump seems to be more positive than not over his tweets