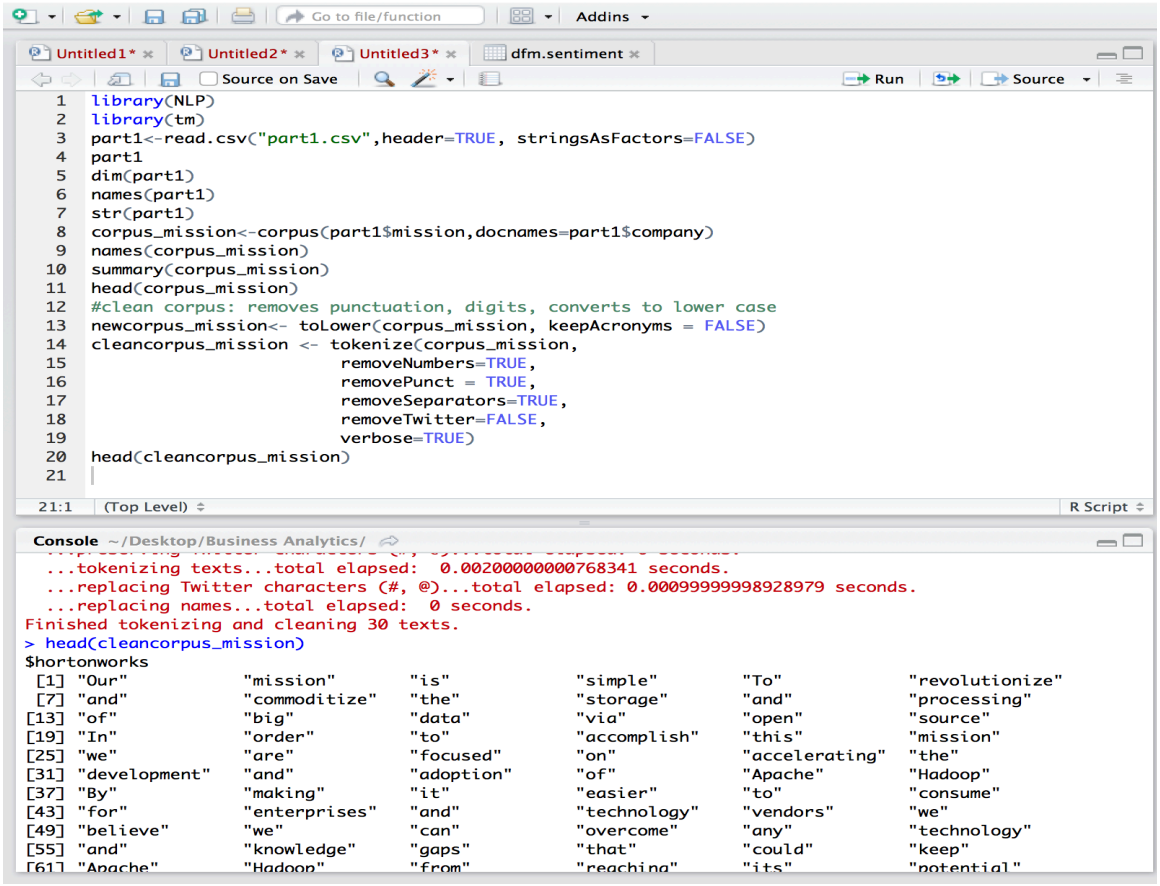
Business Analytics Assignment 7

PART 1:

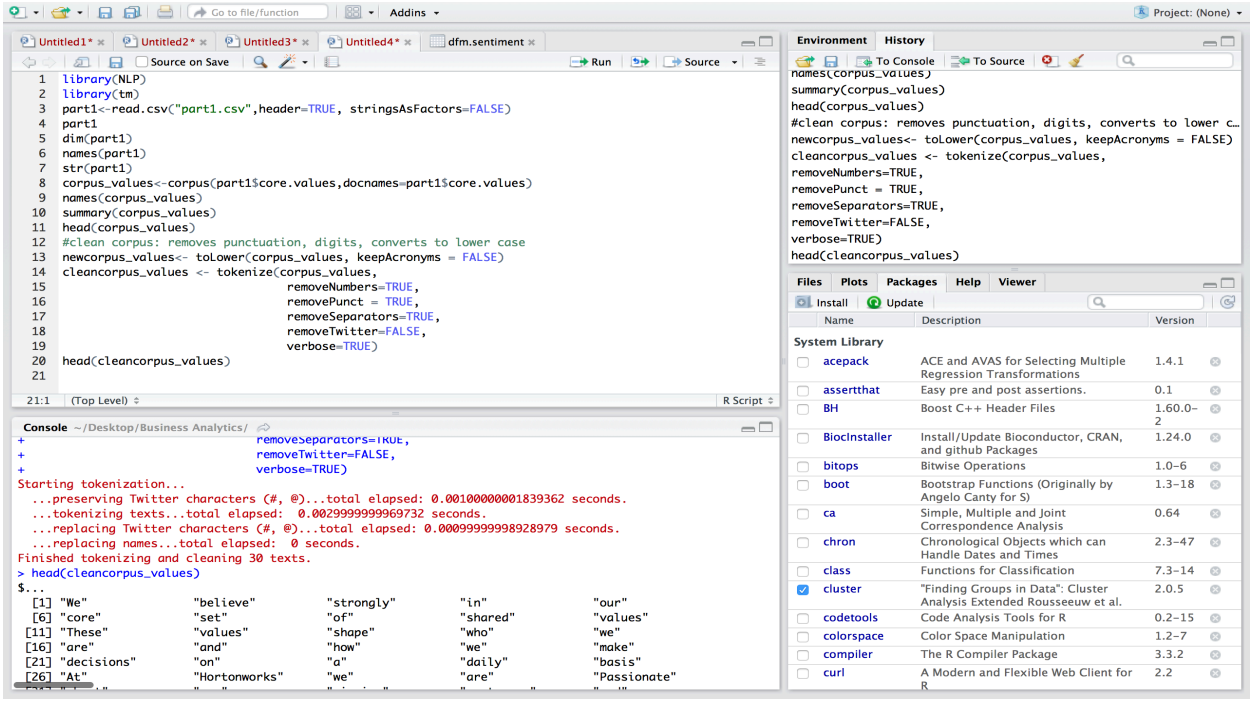
1. Find out the list of successful data-driven companies and select 30 organizations.

- I found the list of top data-driven companies from http://www.forbes.com/sites/louiscolumbus/2016/05/05/the-best-big-data-and-business-analytics-companies-to-work-for-in-2016/ - 43c758974434

2. Create a corpus of their mission statements.



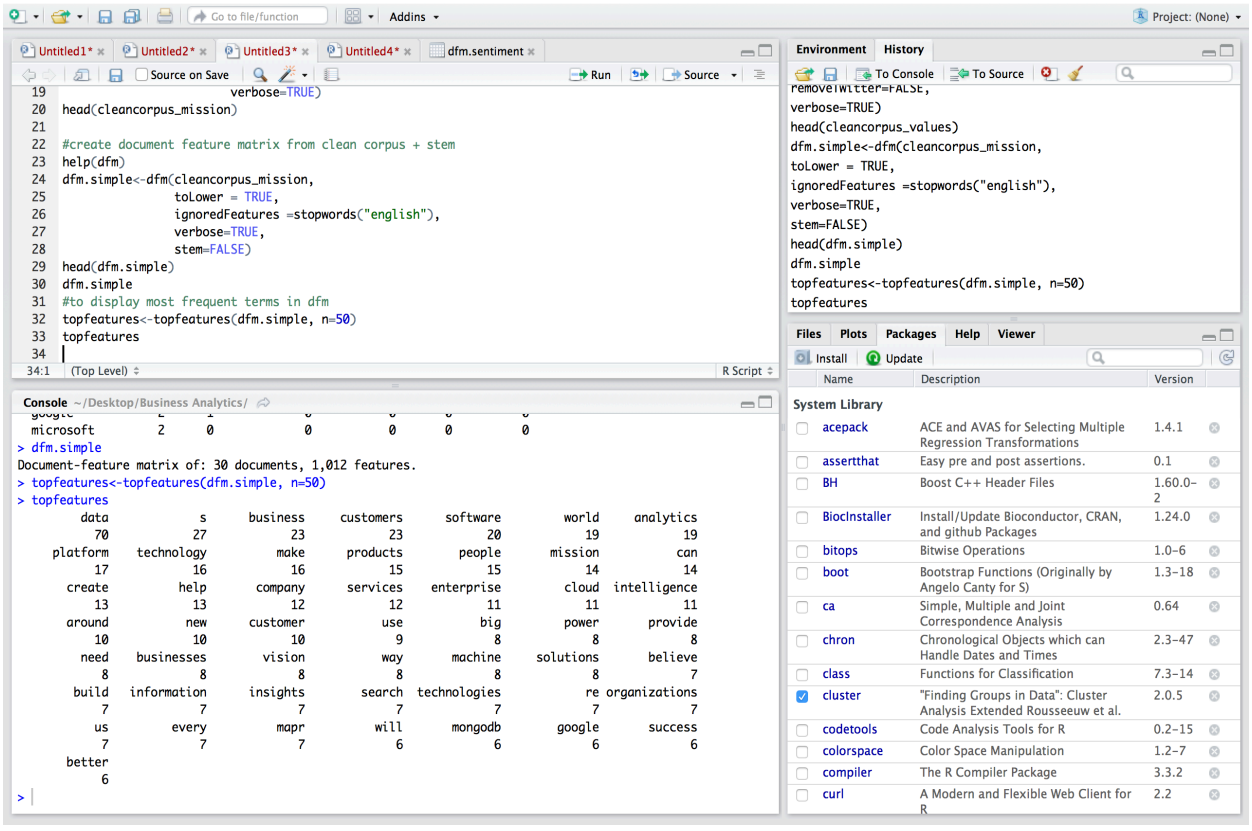
3. Create a corpus of their core values.



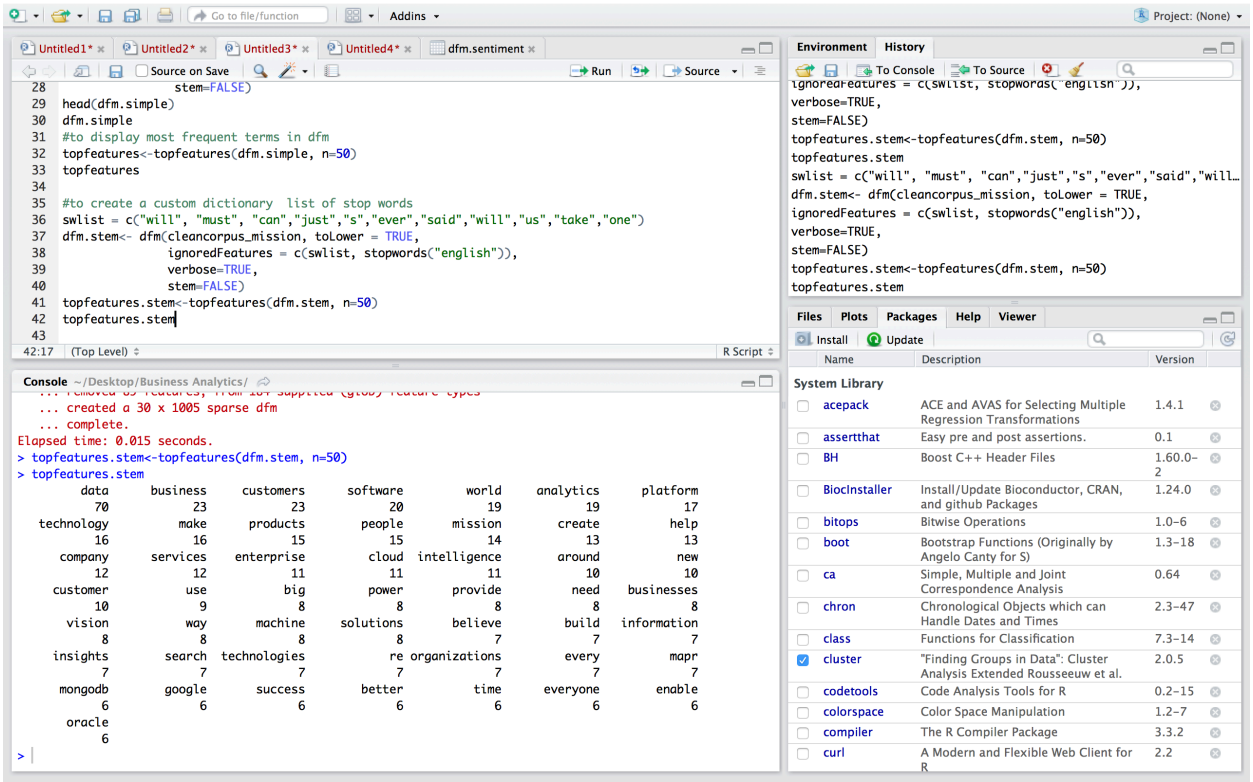
4. Analyze the corpus and provide insight on how to structure a firm for data analysis readiness.

- Performed frequency analysis to find out the most frequent text usage

MISSION:



Frequency analysis using stop word list



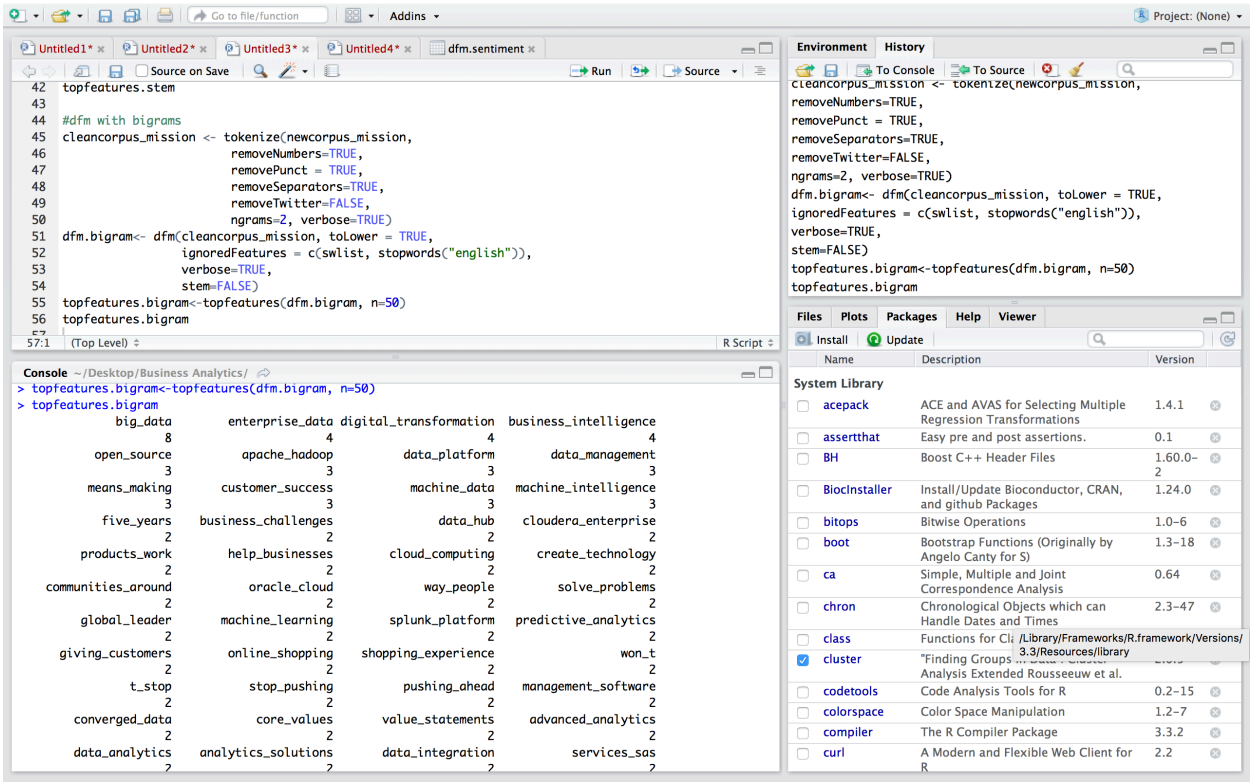
```
                stem=FALSE)
head(dfm.simple)
dfm.simple
#to display most frequent terms in dfm
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures

#to create a custom dictionary  list of stop words
swlist = c("will", "must", "can","just","s","ever","said","Will","us","take","one")
dfm.stem<- dfm(cleancorpus_mission, toLower = TRUE,
               ignoredFeatures = c(swlist, stopwords("english")),
               verbose=TRUE,
               stem=FALSE)
topfeatures.stem<-topfeatures(dfm.stem, n=50)
topfeatures.stem
```

```
   ... created a 30 x 1005 sparse dfm
   ... complete.
Elapsed time: 0.015 seconds.
> topfeatures.stem<-topfeatures(dfm.stem, n=50)
> topfeatures.stem
        data     business    customers     software        world    analytics     platform
          70           23           23           20           19           19           17
  technology         make     products       people      mission       create         help
          16           16           15           15           14           13           13
     company     services   enterprise        cloud intelligence       around          new
          12           12           11           11           11           10           10
    customer          use          big        power      provide         need   businesses
          10            9            8            8            8            8            8
      vision          way      machine    solutions      believe        build  information
           8            8            8            8            7            7            7
    insights       search   technologies           re organizations        every        mapr
           7            7            7            7            7            7            7
     mongodb       google      success       better         time     everyone       enable
           6            6            6            6            6            6            6
      oracle
           6
>
```

```
ignoredFeatures = c(swlist, stopwords("english")),
verbose=TRUE,
stem=FALSE)
topfeatures.stem<-topfeatures(dfm.stem, n=50)
topfeatures.stem
swlist = c("will", "must", "can","just","s","ever","said","will…
dfm.stem<- dfm(cleancorpus_mission, toLower = TRUE,
ignoredFeatures = c(swlist, stopwords("english")),
verbose=TRUE,
stem=FALSE)
topfeatures.stem<-topfeatures(dfm.stem, n=50)
topfeatures.stem
```

	Name	Description	Version
System Library			
☐	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
☐	assertthat	Easy pre and post assertions.	0.1
☐	BH	Boost C++ Header Files	1.60.0-2
☐	BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
☐	bitops	Bitwise Operations	1.0-6
☐	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18
☐	ca	Simple, Multiple and Joint Correspondence Analysis	0.64
☐	chron	Chronological Objects which can Handle Dates and Times	2.3-47
☐	class	Functions for Classification	7.3-14
☑	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.5
☐	codetools	Code Analysis Tools for R	0.2-15
☐	colorspace	Color Space Manipulation	1.2-7
☐	compiler	The R Compiler Package	3.3.2
☐	curl	A Modern and Flexible Web Client for R	2.2

Frequency analysis using bigrams



```
topfeatures.stem

#dfm with bigrams
cleancorpus_mission <- tokenize(newcorpus_mission,
                        removeNumbers=TRUE,
                        removePunct = TRUE,
                        removeSeparators=TRUE,
                        removeTwitter=FALSE,
                        ngrams=2, verbose=TRUE)
dfm.bigram<- dfm(cleancorpus_mission, toLower = TRUE,
                 ignoredFeatures = c(swlist, stopwords("english")),
                 verbose=TRUE,
                 stem=FALSE)
topfeatures.bigram<-topfeatures(dfm.bigram, n=50)
topfeatures.bigram
```

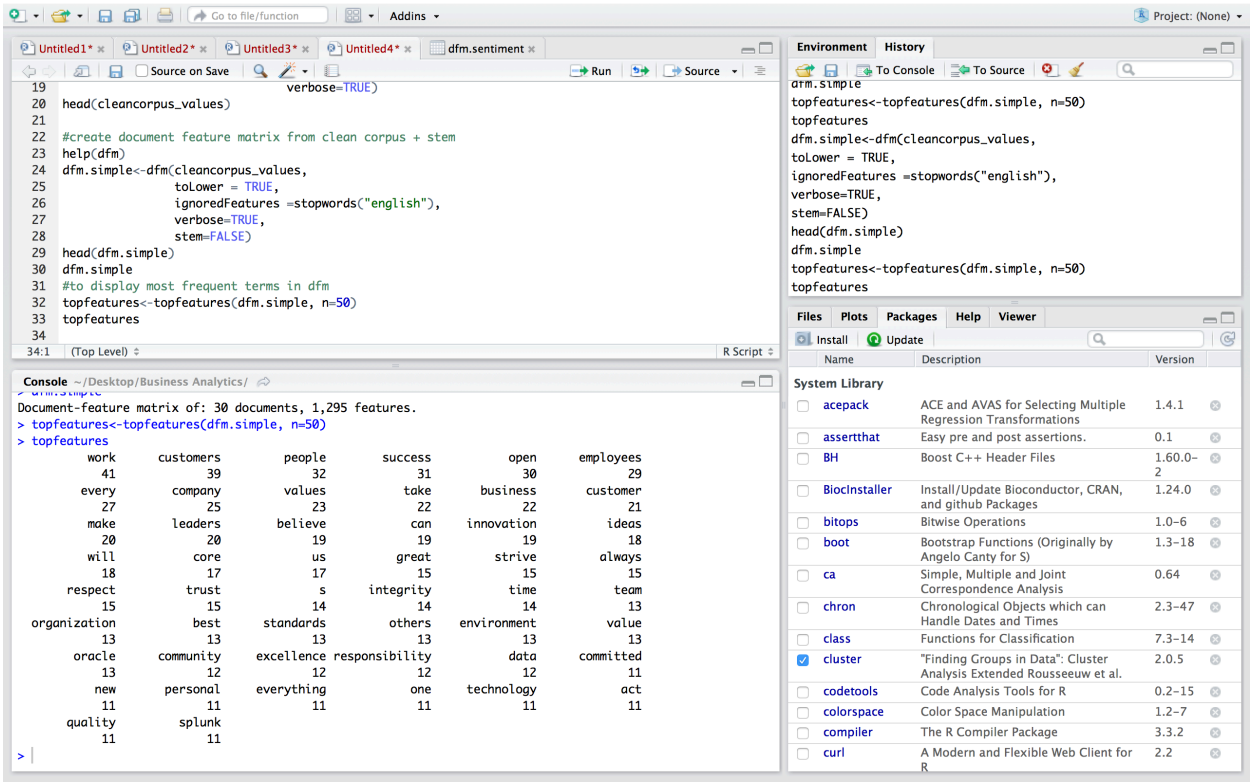
```
> topfeatures.bigram<-topfeatures(dfm.bigram, n=50)
> topfeatures.bigram
             big_data        enterprise_data digital_transformation  business_intelligence
                    8                      4                      4                      4
         open_source          apache_hadoop          data_platform       data_management
                    3                      3                      3                      3
        means_making       customer_success           machine_data machine_intelligence
                    3                      3                      3                      3
           five_years business_challenges              data_hub    cloudera_enterprise
                    2                      2                      2                      2
       products_work       help_businesses       cloud_computing     create_technology
                    2                      2                      2                      2
  communities_around         oracle_cloud           way_people       solve_problems
                    2                      2                      2                      2
       global_leader     machine_learning       splunk_platform predictive_analytics
                    2                      2                      2                      2
   giving_customers       online_shopping    shopping_experience                  won_t
                    2                      2                      2                      2
             t_stop         stop_pushing         pushing_ahead    management_software
                    2                      2                      2                      2
      converged_data          core_values      value_statements    advanced_analytics
                    2                      2                      2                      2
      data_analytics   analytics_solutions      data_integration          services_sas
                    2                      2                      2                      2
```

```
cleancorpus_mission <- tokenize(newcorpus_mission,
removeNumbers=TRUE,
removePunct = TRUE,
removeSeparators=TRUE,
removeTwitter=FALSE,
ngrams=2, verbose=TRUE)
dfm.bigram<- dfm(cleancorpus_mission, toLower = TRUE,
ignoredFeatures = c(swlist, stopwords("english")),
verbose=TRUE,
stem=FALSE)
topfeatures.bigram<-topfeatures(dfm.bigram, n=50)
topfeatures.bigram
```

	Name	Description	Version
System Library			
☐	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
☐	assertthat	Easy pre and post assertions.	0.1
☐	BH	Boost C++ Header Files	1.60.0-2
☐	BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
☐	bitops	Bitwise Operations	1.0-6
☐	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18
☐	ca	Simple, Multiple and Joint Correspondence Analysis	0.64
☐	chron	Chronological Objects which can Handle Dates and Times	2.3-47
☐	class	Functions for Cl…	
☑	cluster	"Finding Groups … Analysis Extended Rousseeuw et al.	
☐	codetools	Code Analysis Tools for R	0.2-15
☐	colorspace	Color Space Manipulation	1.2-7
☐	compiler	The R Compiler Package	3.3.2
☐	curl	A Modern and Flexible Web Client for R	2.2

/Library/Frameworks/R.framework/Versions/3.3/Resources/library

CORE VALUES:



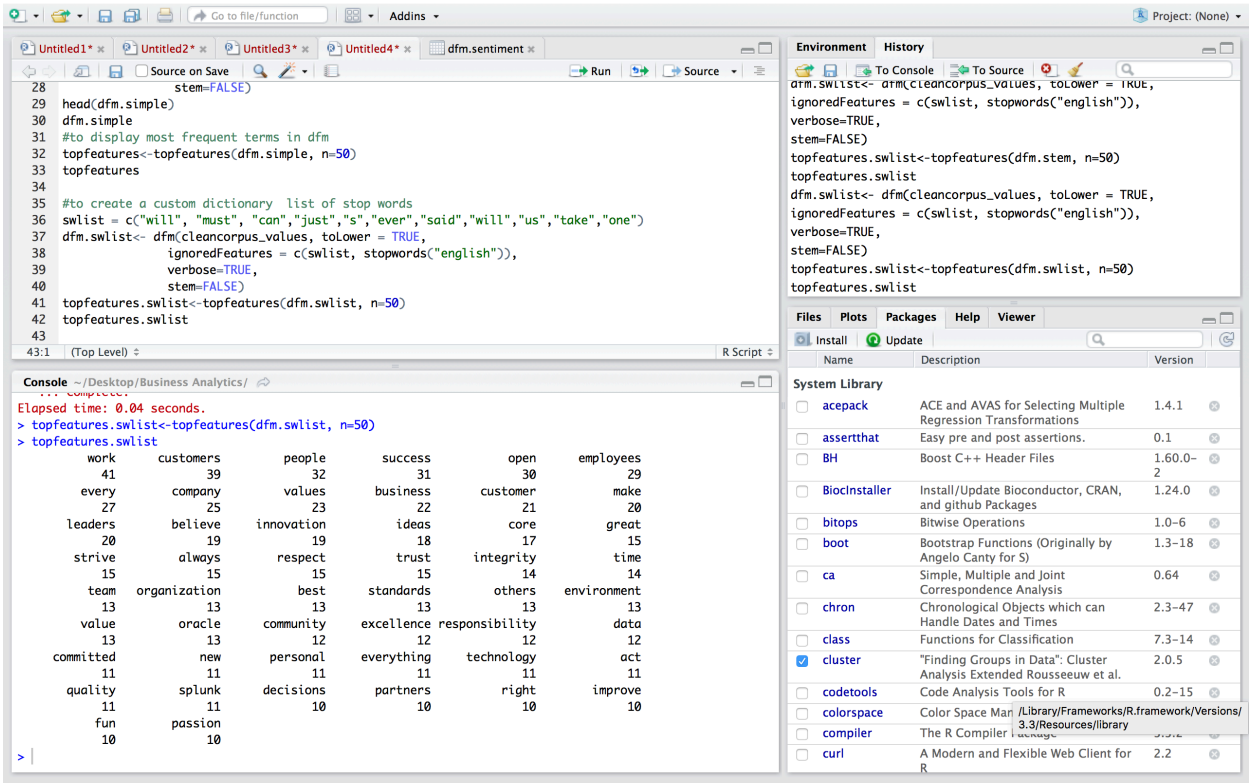
```
                                verbose=TRUE)
head(cleancorpus_values)

#create document feature matrix from clean corpus + stem
help(dfm)
dfm.simple<-dfm(cleancorpus_values,
                toLower = TRUE,
                ignoredFeatures =stopwords("english"),
                verbose=TRUE,
                stem=FALSE)
head(dfm.simple)
dfm.simple
#to display most frequent terms in dfm
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures
```

```
Document-feature matrix of: 30 documents, 1,295 features.
> topfeatures<-topfeatures(dfm.simple, n=50)
> topfeatures
          work      customers         people        success           open      employees
            41             39             32             31             30             29
         every        company         values           take       business       customer
            27             25             23             22             22             21
          make        leaders        believe            can     innovation          ideas
            20             20             19             19             19             18
          will           core             us          great         strive         always
            18             17             17             15             15             15
       respect          trust              s      integrity           time           team
            15             15             14             14             14             13
  organization           best      standards         others    environment          value
            13             13             13             13             13             13
        oracle      community     excellence responsibility           data      committed
            13             12             12             12             12             11
           new       personal     everything            one     technology            act
            11             11             11             11             11             11
       quality         splunk
            11             11
>
```

```
dfm.simple
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures
dfm.simple<-dfm(cleancorpus_values,
toLower = TRUE,
ignoredFeatures =stopwords("english"),
verbose=TRUE,
stem=FALSE)
head(dfm.simple)
dfm.simple
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures
```

	Name	Description	Version
System Library			
☐	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
☐	assertthat	Easy pre and post assertions.	0.1
☐	BH	Boost C++ Header Files	1.60.0-2
☐	BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
☐	bitops	Bitwise Operations	1.0-6
☐	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18
☐	ca	Simple, Multiple and Joint Correspondence Analysis	0.64
☐	chron	Chronological Objects which can Handle Dates and Times	2.3-47
☐	class	Functions for Classification	7.3-14
☑	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.5
☐	codetools	Code Analysis Tools for R	0.2-15
☐	colorspace	Color Space Manipulation	1.2-7
☐	compiler	The R Compiler Package	3.3.2
☐	curl	A Modern and Flexible Web Client for R	2.2

Frequency Analysis using stop word list



```
                stem=FALSE)
head(dfm.simple)
dfm.simple
#to display most frequent terms in dfm
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures

#to create a custom dictionary  list of stop words
swlist = c("will", "must", "can","just","s","ever","said","will","us","take","one")
dfm.swlist<- dfm(cleancorpus_values, toLower = TRUE,
                 ignoredFeatures = c(swlist, stopwords("english")),
                 verbose=TRUE,
                 stem=FALSE)
topfeatures.swlist<-topfeatures(dfm.swlist, n=50)
topfeatures.swlist

```

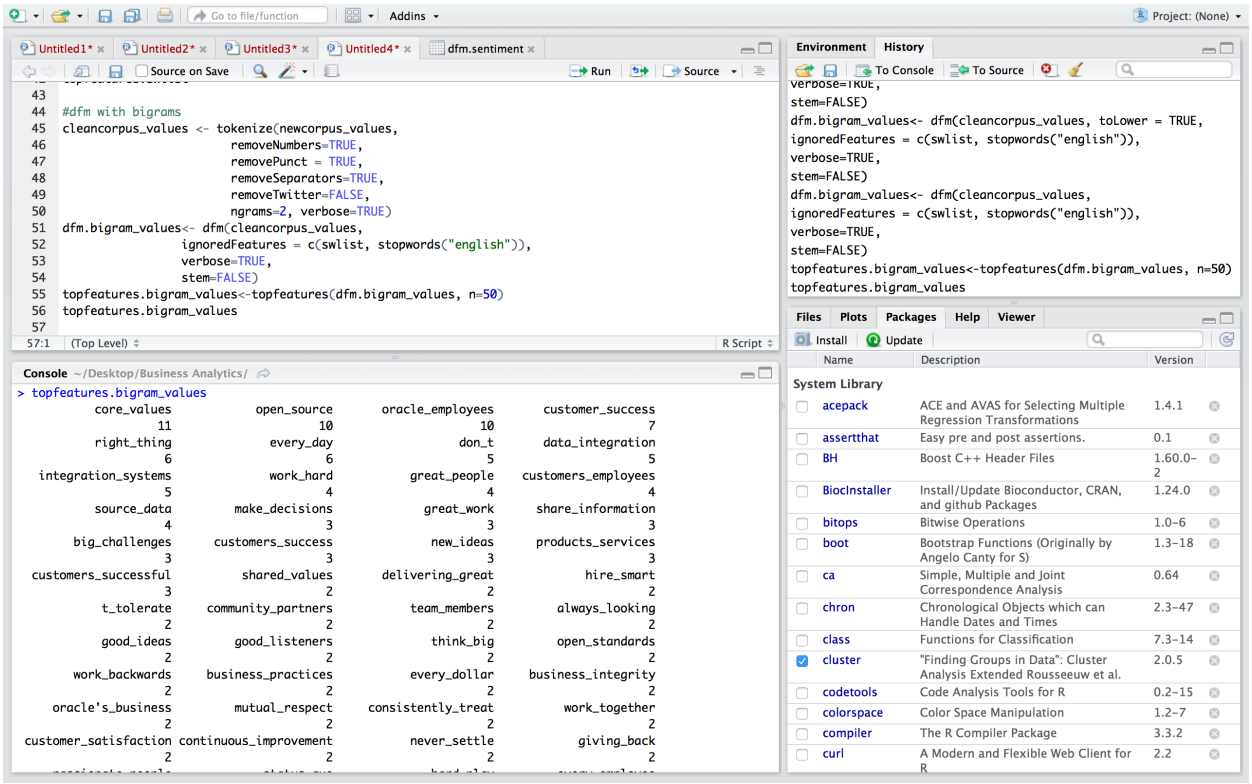
```
Elapsed time: 0.04 seconds.
> topfeatures.swlist<-topfeatures(dfm.swlist, n=50)
> topfeatures.swlist
          work      customers         people        success           open      employees
            41             39             32             31             30             29
         every        company         values       business       customer           make
            27             25             23             22             21             20
       leaders        believe     innovation          ideas           core          great
            20             19             19             18             17             15
        strive         always        respect          trust      integrity           time
            15             15             15             15             14             14
          team   organization           best      standards         others    environment
            13             13             13             13             13             13
         value         oracle      community     excellence responsibility           data
            13             13             12             12             12             12
     committed            new       personal     everything     technology            act
            11             11             11             11             11             11
       quality         splunk      decisions       partners          right        improve
            11             11             10             10             10             10
           fun        passion
            10             10
>
```

```
dfm.swlist<- dfm(cleancorpus_values, toLower = TRUE,
ignoredFeatures = c(swlist, stopwords("english")),
verbose=TRUE,
stem=FALSE)
topfeatures.swlist<-topfeatures(dfm.stem, n=50)
topfeatures.swlist
dfm.swlist<- dfm(cleancorpus_values, toLower = TRUE,
ignoredFeatures = c(swlist, stopwords("english")),
verbose=TRUE,
stem=FALSE)
topfeatures.swlist<-topfeatures(dfm.swlist, n=50)
topfeatures.swlist
```

	Name	Description	Version
System Library			
☐	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
☐	assertthat	Easy pre and post assertions.	0.1
☐	BH	Boost C++ Header Files	1.60.0-2
☐	BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
☐	bitops	Bitwise Operations	1.0-6
☐	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18
☐	ca	Simple, Multiple and Joint Correspondence Analysis	0.64
☐	chron	Chronological Objects which can Handle Dates and Times	2.3-47
☐	class	Functions for Classification	7.3-14
☑	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.5
☐	codetools	Code Analysis Tools for R	0.2-15
☐	colorspace	Color Space Man	
☐	compiler	The R Compiler	
☐	curl	A Modern and Flexible Web Client for R	2.2

/Library/Frameworks/R.framework/Versions/3.3/Resources/library

Frequency analysis using bigrams



INSIGHT:

Mission Statement:

Most of the data driven companies aim to provide excellent big data solutions, business intelligence, digital transformations and analytical solutions.

Core Values:

From the analysis we come to know that most companies are looking out for individuals who:

Work hard, make great decisions, share information, take big challenges, good listeners and help them to successfully deliver the product to the customers.

5. Are there any other data-driven approaches you would recommend the CEO to implement?

- According to me, the CEO should also take into consideration the tools and technologies that the top data driven companies use to provide analytical solutions and business insights.

PART 2:

1. Search for “Donald Trump Speech Transcript” and select three speeches of your choice.

- I selected the following three speech transcripts of Donald Trump for my analysis:
- **Transcript1:**

http://thehill.com/blogs/pundits-blog/campaign/294817-transcript-of-donald-trumps-speech-on-national-security-in

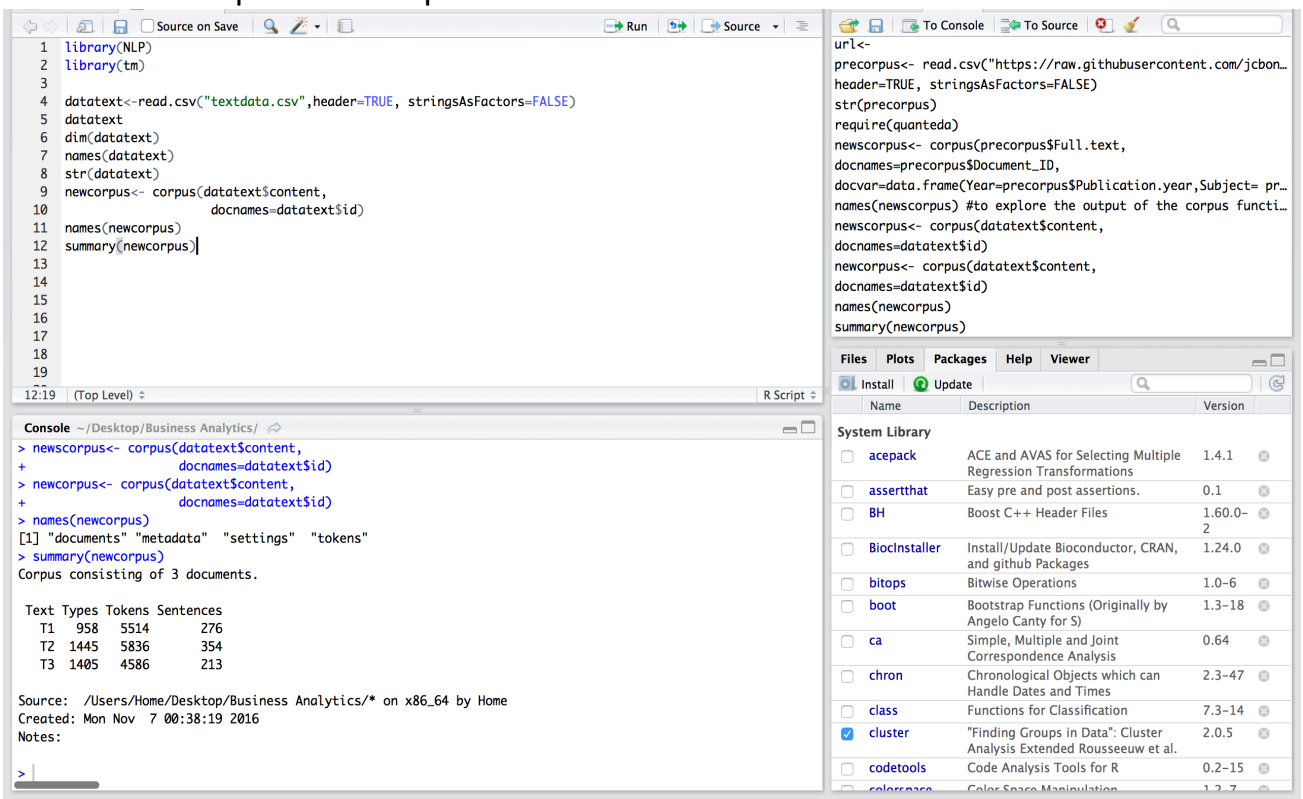
- **Transcript 2:**

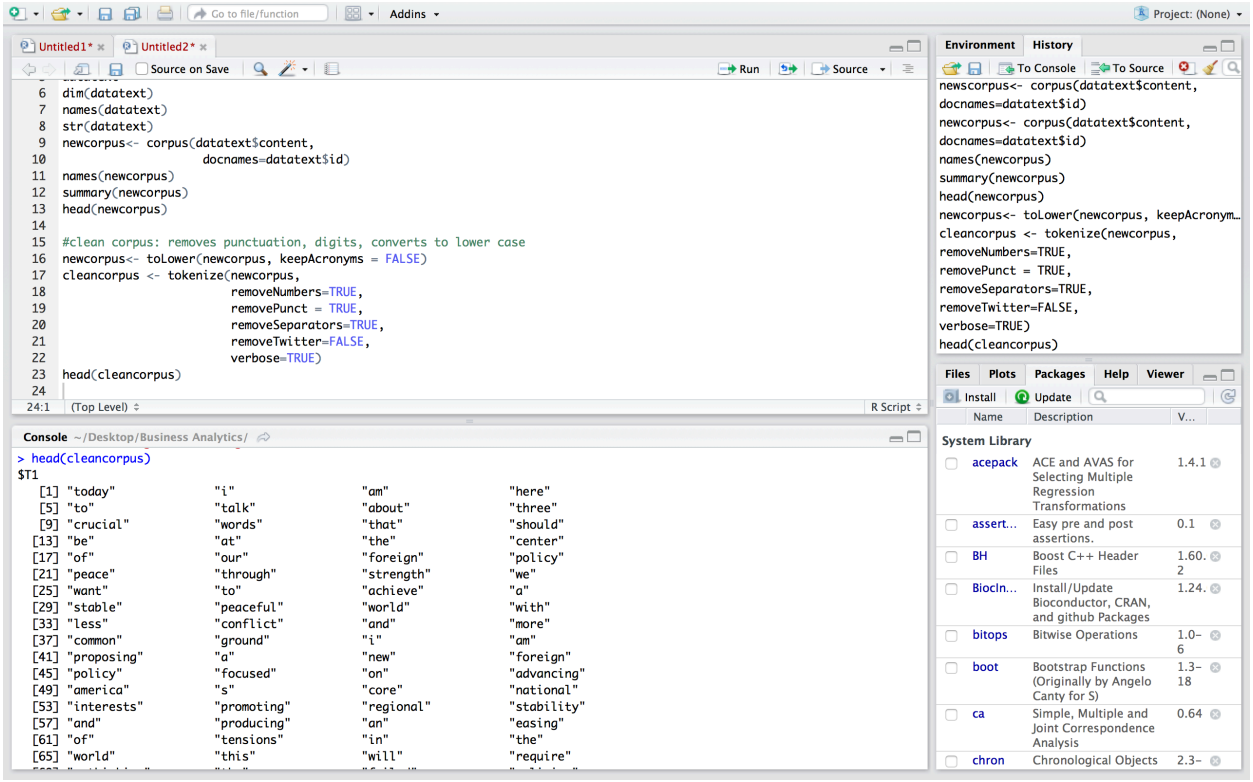
http://www.vox.com/2016/7/21/12253426/donald-trump-acceptance-speech-transcript-republican-nomination-transcript

- **Transcript 3:**

http://heavy.com/news/2016/08/read-donald-trump-full-transcript-speech-foreign-policy-address-remarks-prepared-august-15/

2. Create a corpus for the speeches.





```
dim(datatext)
names(datatext)
str(datatext)
newcorpus<- corpus(datatext$content,
                   docnames=datatext$id)
names(newcorpus)
summary(newcorpus)
head(newcorpus)

#clean corpus: removes punctuation, digits, converts to lower case
newcorpus<- toLower(newcorpus, keepAcronyms = FALSE)
cleancorpus <- tokenize(newcorpus,
                        removeNumbers=TRUE,
                        removePunct = TRUE,
                        removeSeparators=TRUE,
                        removeTwitter=FALSE,
                        verbose=TRUE)
head(cleancorpus)
```

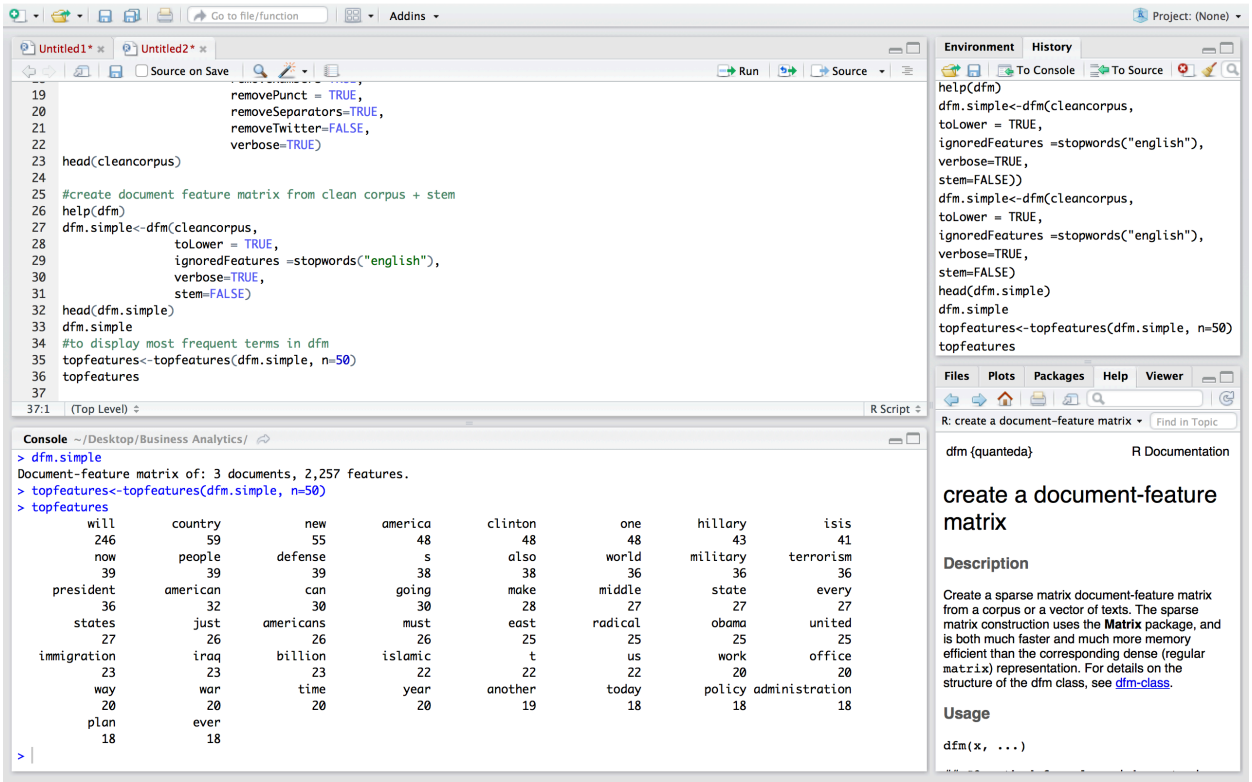
```
> head(cleancorpus)
$T1
   [1] "today"          "i"              "am"             "here"
   [5] "to"             "talk"           "about"          "three"
   [9] "crucial"        "words"          "that"           "should"
  [13] "be"             "at"             "the"            "center"
  [17] "of"             "our"            "foreign"        "policy"
  [21] "peace"          "through"        "strength"       "we"
  [25] "want"           "to"             "achieve"        "a"
  [29] "stable"         "peaceful"       "world"          "with"
  [33] "less"           "conflict"       "and"            "more"
  [37] "common"         "ground"         "i"              "am"
  [41] "proposing"      "a"              "new"            "foreign"
  [45] "policy"         "focused"        "on"             "advancing"
  [49] "america"        "s"              "core"           "national"
  [53] "interests"      "promoting"      "regional"       "stability"
  [57] "and"            "producing"      "an"             "easing"
  [61] "of"             "tensions"       "in"             "the"
  [65] "world"          "this"           "will"           "require"
```

```
newscorpus<- corpus(datatext$content,
docnames=datatext$id)
newcorpus<- corpus(datatext$content,
docnames=datatext$id)
names(newcorpus)
summary(newcorpus)
head(newcorpus)
newcorpus<- toLower(newcorpus, keepAcronym…
cleancorpus <- tokenize(newcorpus,
removeNumbers=TRUE,
removePunct = TRUE,
removeSeparators=TRUE,
removeTwitter=FALSE,
verbose=TRUE)
head(cleancorpus)
```

Name	Description	V...
System Library		
acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
assert...	Easy pre and post assertions.	0.1
BH	Boost C++ Header Files	1.60.2
BiocIn...	Install/Update Bioconductor, CRAN, and github Packages	1.24.
bitops	Bitwise Operations	1.0-6
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18
ca	Simple, Multiple and Joint Correspondence Analysis	0.64
chron	Chronological Objects	2.3-

3. Complete a frequency analysis of the word usage

- Frequency analysis without stemming:



```
                        removePunct = TRUE,
                        removeSeparators=TRUE,
                        removeTwitter=FALSE,
                        verbose=TRUE)
head(cleancorpus)

#create document feature matrix from clean corpus + stem
help(dfm)
dfm.simple<-dfm(cleancorpus,
                toLower = TRUE,
                ignoredFeatures =stopwords("english"),
                verbose=TRUE,
                stem=FALSE)
head(dfm.simple)
dfm.simple
#to display most frequent terms in dfm
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures
```

```
> dfm.simple
Document-feature matrix of: 3 documents, 2,257 features.
> topfeatures<-topfeatures(dfm.simple, n=50)
> topfeatures
          will        country            new        america        clinton            one        hillary           isis
           246             59             55             48             48             48             43             41
           now         people        defense              s           also          world       military      terrorism
            39             39             39             38             38             36             36             36
     president       american            can          going           make         middle          state          every
            36             32             30             30             28             27             27             27
        states           just      americans           must           east        radical          obama         united
            27             26             26             26             25             25             25             25
   immigration           iraq        billion        islamic              t             us           work         office
            23             23             23             22             22             22             20             20
           way            war           time           year        another          today         policy administration
            20             20             20             20             19             18             18             18
          plan           ever
            18             18
>
```

```
help(dfm)
dfm.simple<-dfm(cleancorpus,
toLower = TRUE,
ignoredFeatures =stopwords("english"),
verbose=TRUE,
stem=FALSE))
dfm.simple<-dfm(cleancorpus,
toLower = TRUE,
ignoredFeatures =stopwords("english"),
verbose=TRUE,
stem=FALSE)
head(dfm.simple)
dfm.simple
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures
```

dfm {quanteda} R Documentation

create a document-feature matrix

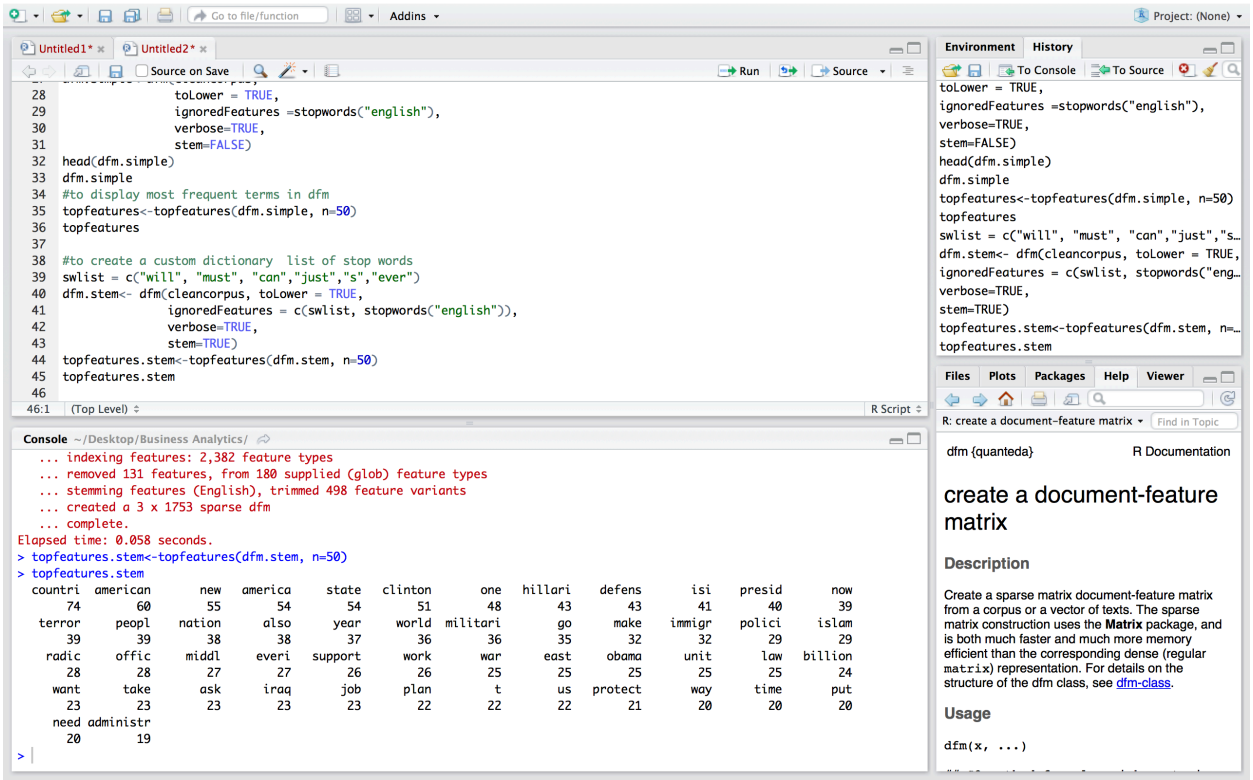
Description

Create a sparse matrix document-feature matrix from a corpus or a vector of texts. The sparse matrix construction uses the **Matrix** package, and is both much faster and much more memory efficient than the corresponding dense (regular `matrix`) representation. For details on the structure of the dfm class, see dfm-class.

Usage

```
dfm(x, ...)
```

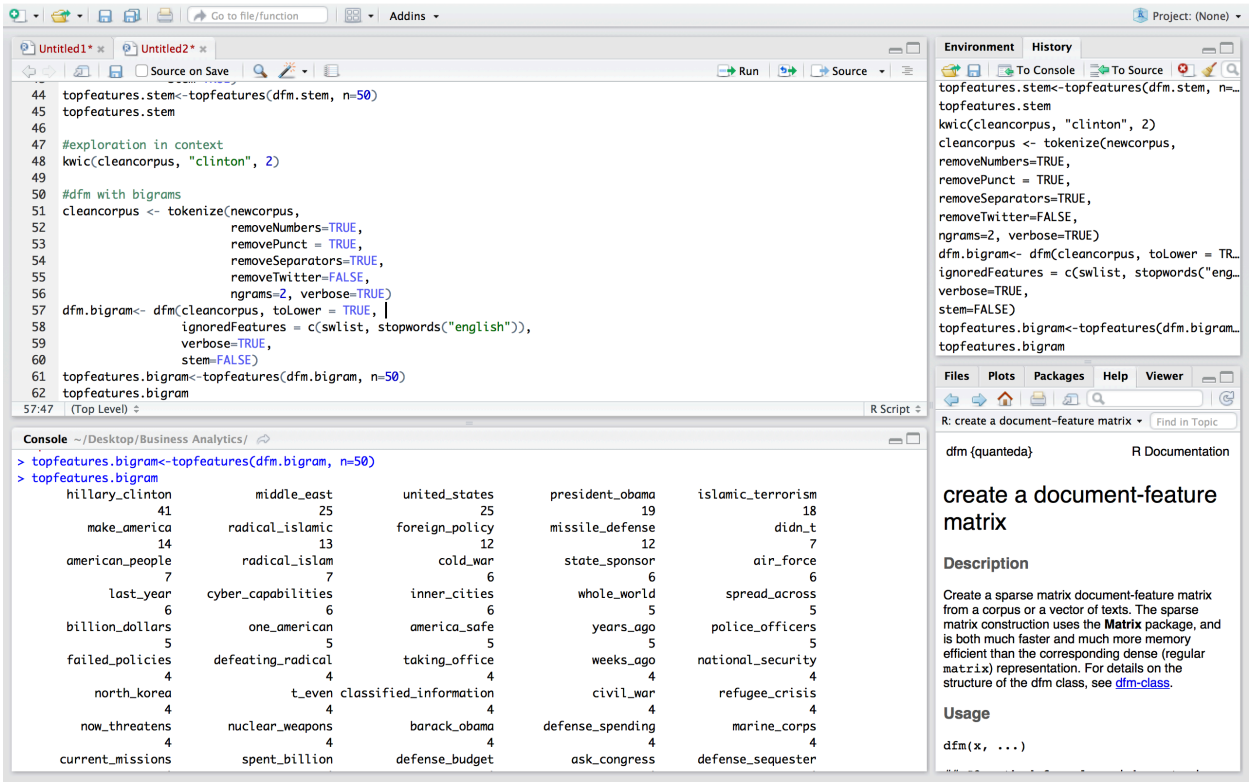
- Frequency analysis with stemming



```
                toLower = TRUE,
                ignoredFeatures =stopwords("english"),
                verbose=TRUE,
                stem=FALSE)
head(dfm.simple)
dfm.simple
#to display most frequent terms in dfm
topfeatures<-topfeatures(dfm.simple, n=50)
topfeatures

#to create a custom dictionary  list of stop words
swlist = c("will", "must", "can","just","s","ever")
dfm.stem<- dfm(cleancorpus, toLower = TRUE,
               ignoredFeatures = c(swlist, stopwords("english")),
               verbose=TRUE,
               stem=TRUE)
topfeatures.stem<-topfeatures(dfm.stem, n=50)
topfeatures.stem
```

```
   ... indexing features: 2,382 feature types
   ... removed 131 features, from 180 supplied (glob) feature types
   ... stemming features (English), trimmed 498 feature variants
   ... created a 3 x 1753 sparse dfm
   ... complete.
Elapsed time: 0.058 seconds.
> topfeatures.stem<-topfeatures(dfm.stem, n=50)
> topfeatures.stem
  countri  american       new   america     state   clinton       one   hillari    defens       isi    presid       now
       74        60        55        54        54        51        48        43        43        41        40        39
   terror     peopl    nation      also      year     world  militari        go      make    immigr    polici     islam
       39        39        38        38        37        36        36        35        32        32        29        29
    radic     offic     middl     everi   support      work       war      east     obama      unit       law   billion
       28        28        27        27        26        26        25        25        25        25        25        24
    want      take       ask      iraq       job      plan         t        us   protect       way      time       put
       23        23        23        23        23        22        22        22        21        20        20        20
    need administr
       20        19
```

- Frequency analysis using Bigrams

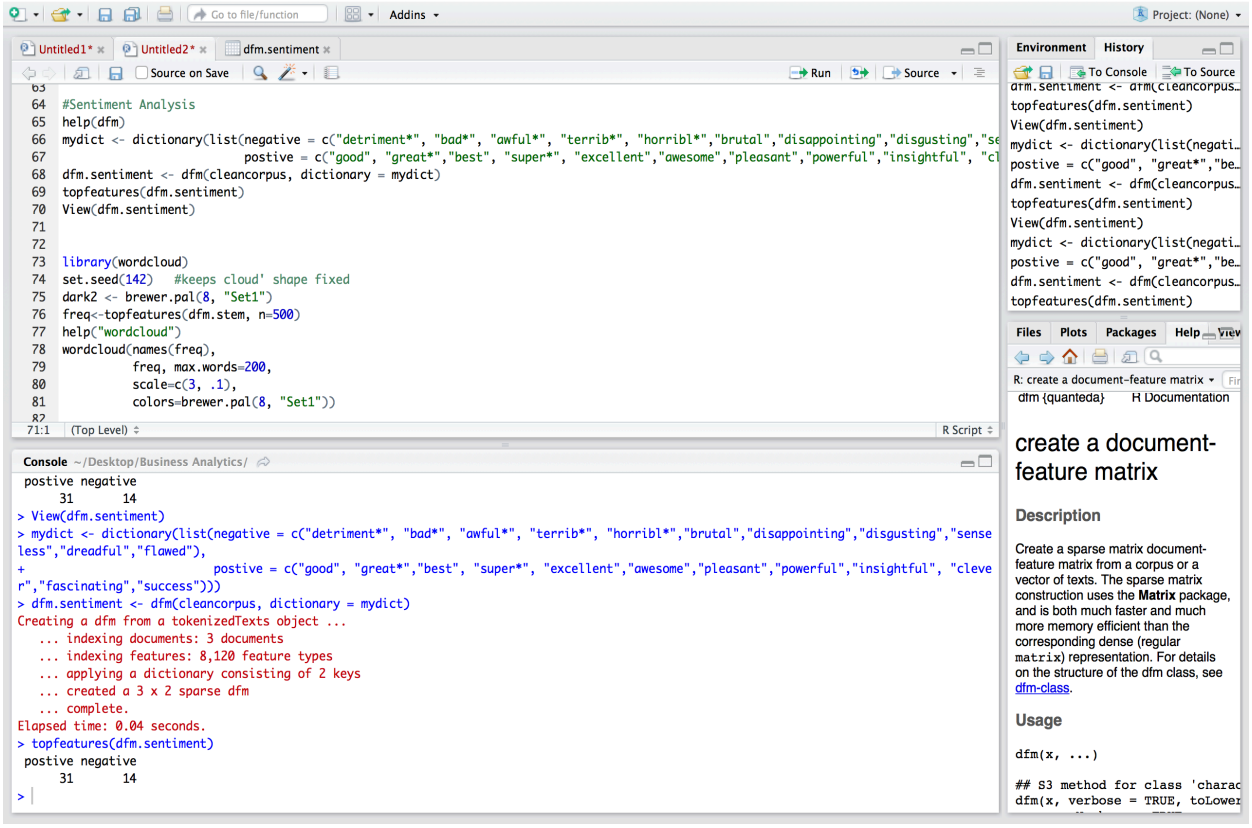


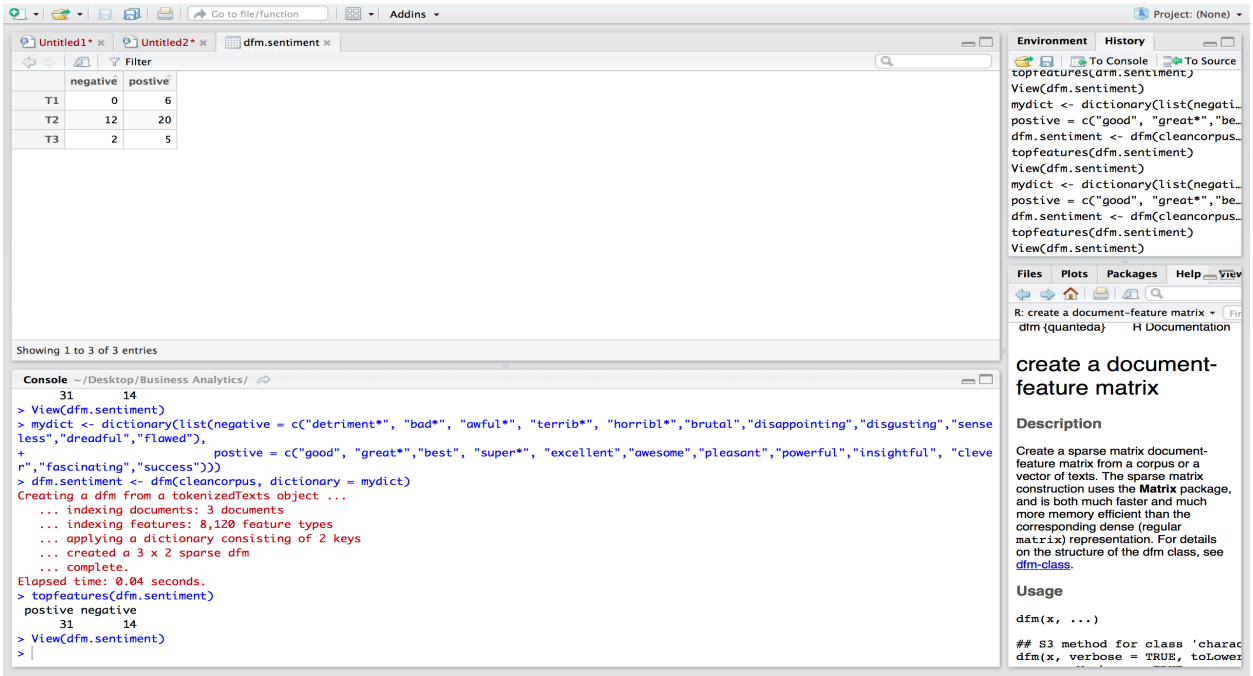
```
topfeatures.stem<-topfeatures(dfm.stem, n=50)
topfeatures.stem

#exploration in context
kwic(cleancorpus, "clinton", 2)

#dfm with bigrams
cleancorpus <- tokenize(newcorpus,
                        removeNumbers=TRUE,
                        removePunct = TRUE,
                        removeSeparators=TRUE,
                        removeTwitter=FALSE,
                        ngrams=2, verbose=TRUE)
dfm.bigram<- dfm(cleancorpus, toLower = TRUE,
                 ignoredFeatures = c(swlist, stopwords("english")),
                 verbose=TRUE,
                 stem=FALSE)
topfeatures.bigram<-topfeatures(dfm.bigram, n=50)
topfeatures.bigram
```

```
> topfeatures.bigram<-topfeatures(dfm.bigram, n=50)
> topfeatures.bigram
      hillary_clinton          middle_east         united_states      president_obama     islamic_terrorism
                   41                   25                   25                   19                   18
        make_america      radical_islamic      foreign_policy      missile_defense              didn_t
                   14                   13                   12                   12                    7
     american_people       radical_islam            cold_war      state_sponsor           air_force
                    7                    7                    6                    6                    6
           last_year   cyber_capabilities        inner_cities       whole_world       spread_across
                    6                    6                    6                    5                    5
      billion_dollars       one_american       america_safe        years_ago      police_officers
                    5                    5                    5                    5                    5
     failed_policies   defeating_radical      taking_office        weeks_ago   national_security
                    4                    4                    4                    4                    4
         north_korea          t_even classified_information        civil_war      refugee_crisis
                    4                    4                    4                    4                    4
       now_threatens      nuclear_weapons       barack_obama   defense_spending       marine_corps
                    4                    4                    4                    4                    4
    current_missions        spent_billion     defense_budget        ask_congress   defense_sequester
```

4. Complete a sentiment analysis.

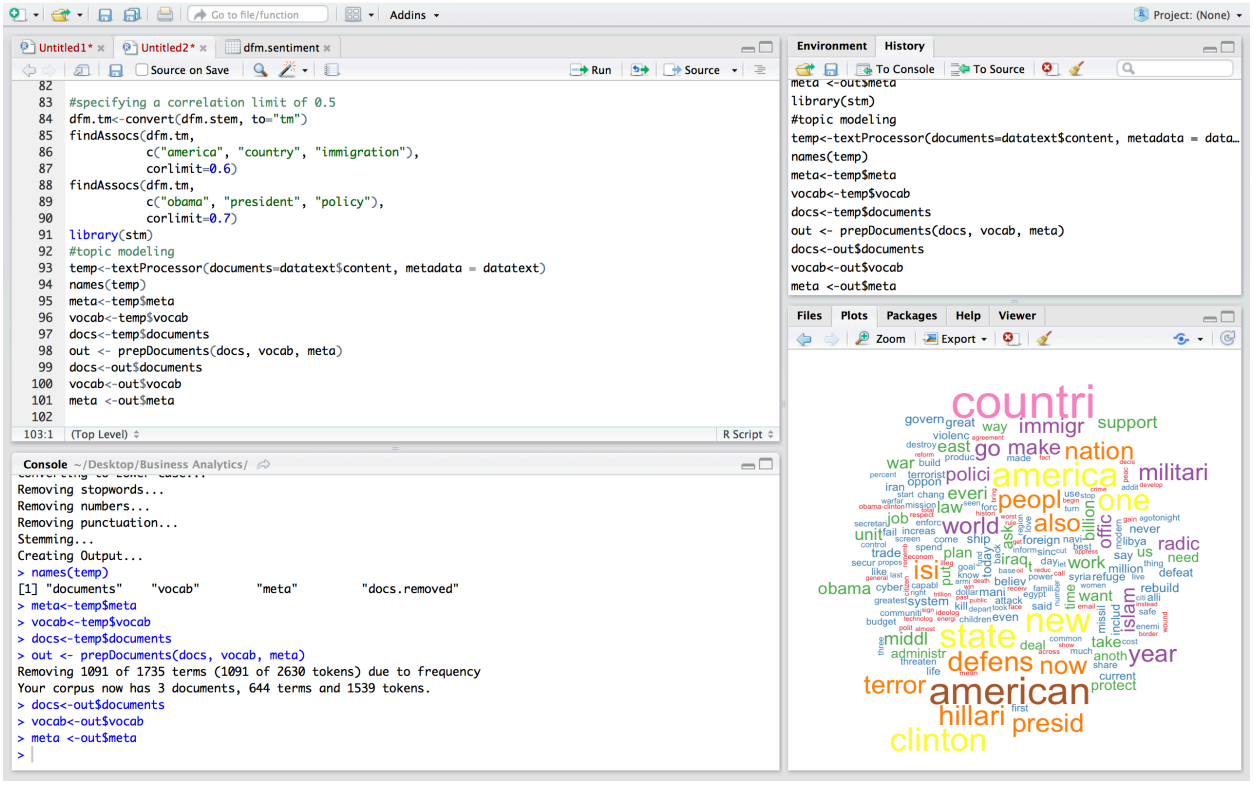
- Created a dictionary with positive and negative words to perform sentiment analysis.

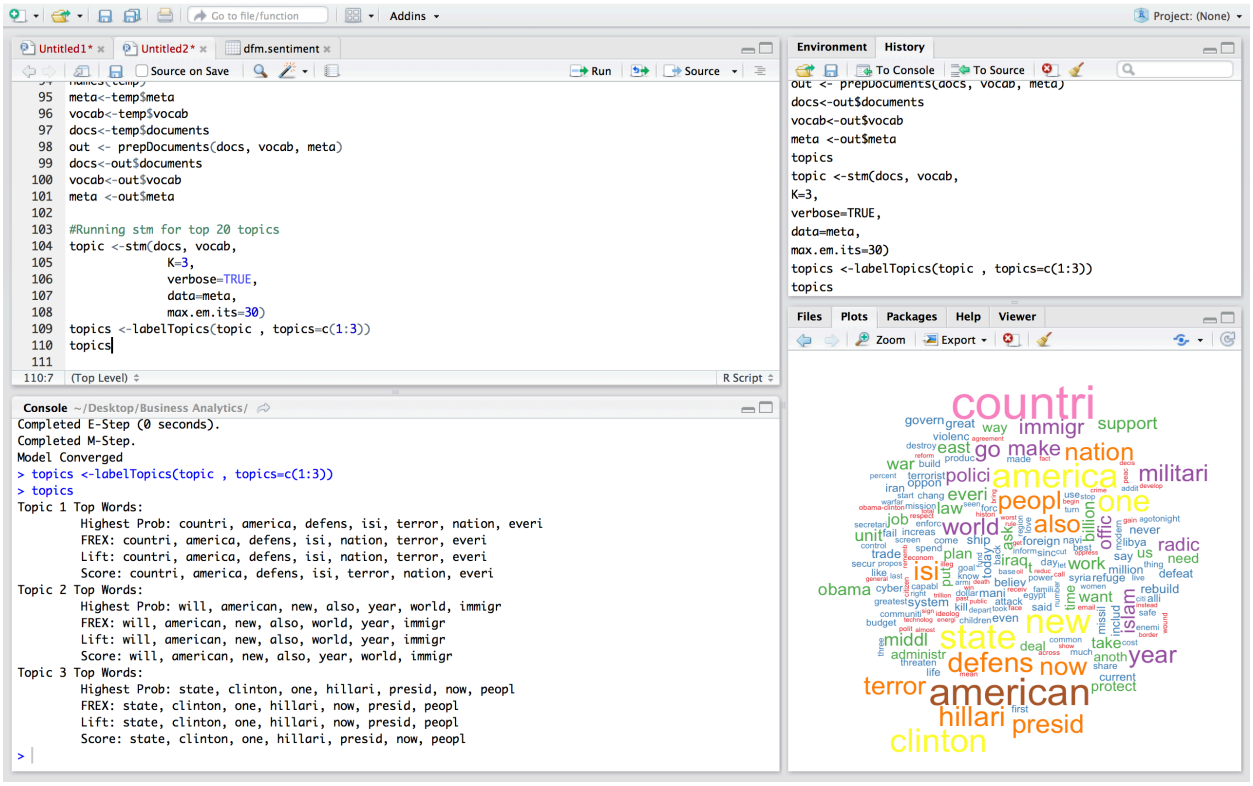


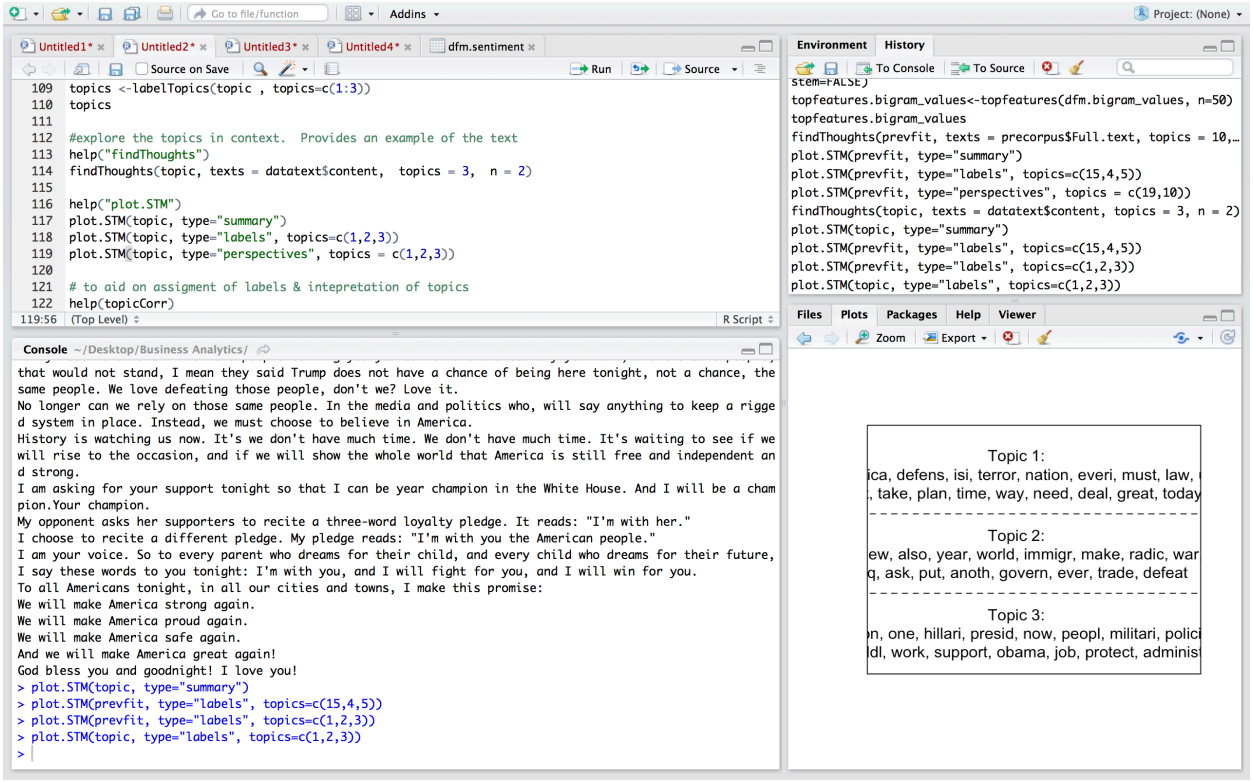


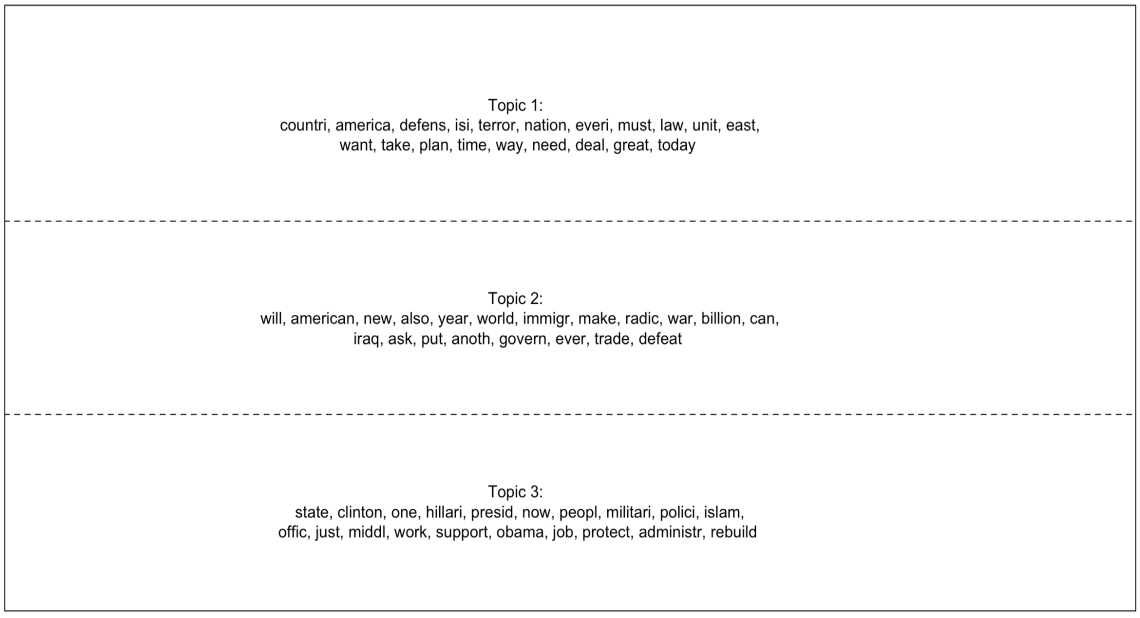
5. What are the common topics in the corpus?

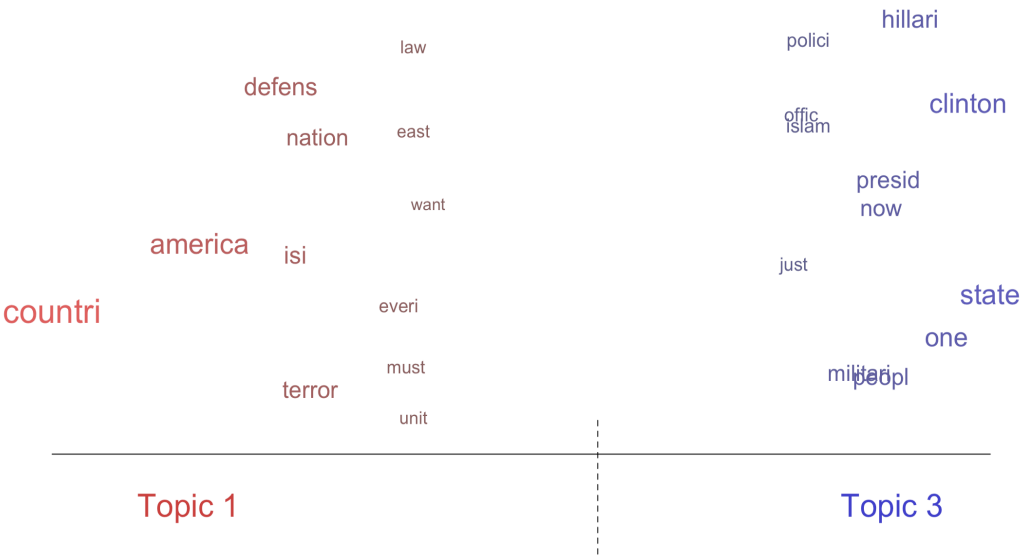
- To identify common topics, performed topic modelling.

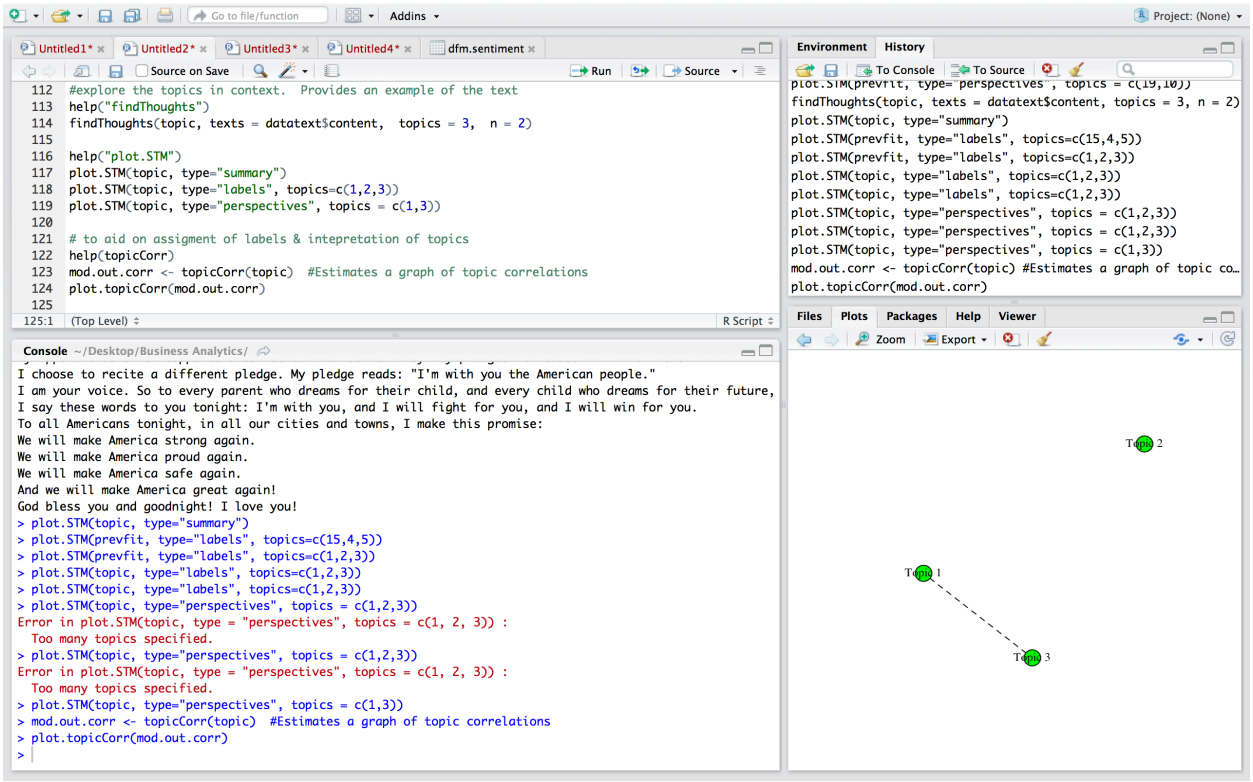





Untitled1*
Untitled2*
Untitled3*
Untitled4*
dfm.sentiment
topics <-labelTopics(topic , topics=c(1:3))
topics
#explore the topics in context. Provides an example of the text
help("findThoughts")
findThoughts(topic, texts = datatext$content, topics = 3, n = 2)
help("plot.STM")
plot.STM(topic, type="summary")
plot.STM(topic, type="labels", topics=c(1,2,3))
plot.STM(topic, type="perspectives", topics = c(1,2,3))
# to aid on assigment of labels & intepretation of topics
help(topicCorr)
Console ~/Desktop/Business Analytics/
that would not stand, I mean they said Trump does not have a chance of being here tonight, not a chance, the same people. We love defeating those people, don't we? Love it.
No longer can we rely on those same people. In the media and politics who, will say anything to keep a rigged system in place. Instead, we must choose to believe in America.
History is watching us now. It's we don't have much time. We don't have much time. It's waiting to see if we will rise to the occasion, and if we will show the whole world that America is still free and independent and strong.
I am asking for your support tonight so that I can be year champion in the White House. And I will be a champion.Your champion.
My opponent asks her supporters to recite a three-word loyalty pledge. It reads: "I'm with her."
I choose to recite a different pledge. My pledge reads: "I'm with you the American people."
I am your voice. So to every parent who dreams for their child, and every child who dreams for their future, I say these words to you tonight: I'm with you, and I will fight for you, and I will win for you.
To all Americans tonight, in all our cities and towns, I make this promise:
We will make America strong again.
We will make America proud again.
We will make America safe again.
And we will make America great again!
God bless you and goodnight! I love you!
> plot.STM(topic, type="summary")
> plot.STM(prevfit, type="labels", topics=c(15,4,5))
> plot.STM(prevfit, type="labels", topics=c(1,2,3))
> plot.STM(topic, type="labels", topics=c(1,2,3))
Environment
History
topfeatures.bigram_values<-topfeatures(dfm.bigram_values, n=50)
topfeatures.bigram_values
findThoughts(prevfit, texts = precorpus$Full.text, topics = 10,…
plot.STM(prevfit, type="summary")
plot.STM(prevfit, type="labels", topics=c(15,4,5))
plot.STM(prevfit, type="perspectives", topics = c(19,10))
findThoughts(topic, texts = datatext$content, topics = 3, n = 2)
plot.STM(topic, type="summary")
plot.STM(prevfit, type="labels", topics=c(15,4,5))
plot.STM(prevfit, type="labels", topics=c(1,2,3))
plot.STM(topic, type="labels", topics=c(1,2,3))
Files
Plots
Packages
Help
Viewer
Zoom
Export
Topic 1:
ica, defens, isi, terror, nation, everi, must, law,
, take, plan, time, way, need, deal, great, today
Topic 2:
ew, also, year, world, immigr, make, radic, war
q, ask, put, anoth, govern, ever, trade, defeat
Topic 3:
n, one, hillari, presid, now, peopl, militari, polici
dl, work, support, obama, job, protect, adminis


Topic 1:
countri, america, defens, isi, terror, nation, everi, must, law, unit, east,
want, take, plan, time, way, need, deal, great, today
Topic 2:
will, american, new, also, year, world, immigr, make, radic, war, billion, can,
iraq, ask, put, anoth, govern, ever, trade, defeat
Topic 3:
state, clinton, one, hillari, presid, now, peopl, militari, polici, islam,
offic, just, middl, work, support, obama, job, protect, administr, rebuild


law
defens
nation
east
want
america
isi
countri
everi
must
terror
unit
Topic 1
hillari
polici
clinton
offic
islam
presid
now
just
state
one
militari
peopl
Topic 3


Addins
Project: (None)
Untitled1*
Untitled2*
Untitled3*
Untitled4*
dfm.sentiment
Source on Save
Run
Source
112 #explore the topics in context. Provides an example of the text
113 help("findThoughts")
114 findThoughts(topic, texts = datatext$content, topics = 3, n = 2)
115
116 help("plot.STM")
117 plot.STM(topic, type="summary")
118 plot.STM(topic, type="labels", topics=c(1,2,3))
119 plot.STM(topic, type="perspectives", topics = c(1,3))
120
121 # to aid on assigment of labels & intepretation of topics
122 help(topicCorr)
123 mod.out.corr <- topicCorr(topic) #Estimates a graph of topic correlations
124 plot.topicCorr(mod.out.corr)
125
125:1 (Top Level)
R Script
Console ~/Desktop/Business Analytics/
I choose to recite a different pledge. My pledge reads: "I'm with you the American people."
I am your voice. So to every parent who dreams for their child, and every child who dreams for their future,
I say these words to you tonight: I'm with you, and I will fight for you, and I will win for you.
To all Americans tonight, in all our cities and towns, I make this promise:
We will make America strong again.
We will make America proud again.
We will make America safe again.
And we will make America great again!
God bless you and goodnight! I love you!
> plot.STM(topic, type="summary")
> plot.STM(prevfit, type="labels", topics=c(15,4,5))
> plot.STM(prevfit, type="labels", topics=c(1,2,3))
> plot.STM(topic, type="labels", topics=c(1,2,3))
> plot.STM(topic, type="labels", topics=c(1,2,3))
> plot.STM(topic, type="perspectives", topics = c(1,2,3))
Error in plot.STM(topic, type = "perspectives", topics = c(1, 2, 3)) :
Too many topics specified.
> plot.STM(topic, type="perspectives", topics = c(1,2,3))
Error in plot.STM(topic, type = "perspectives", topics = c(1, 2, 3)) :
Too many topics specified.
> plot.STM(topic, type="perspectives", topics = c(1,3))
> mod.out.corr <- topicCorr(topic) #Estimates a graph of topic correlations
> plot.topicCorr(mod.out.corr)
>
Environment
History
To Console
To Source
findThoughts(topic, texts = datatext$content, topics = 3, n = 2)
plot.STM(topic, type="summary")
plot.STM(prevfit, type="labels", topics=c(15,4,5))
plot.STM(prevfit, type="labels", topics=c(1,2,3))
plot.STM(topic, type="labels", topics=c(1,2,3))
plot.STM(topic, type="labels", topics=c(1,2,3))
plot.STM(topic, type="perspectives", topics = c(1,2,3))
plot.STM(topic, type="perspectives", topics = c(1,2,3))
plot.STM(topic, type="perspectives", topics = c(1,3))
mod.out.corr <- topicCorr(topic) #Estimates a graph of topic co…
plot.topicCorr(mod.out.corr)
Files
Plots
Packages
Help
Viewer
Zoom
Export
Topic 2
Topic 1
Topic 3

6. Write a memo style report summarizing on Trump's linguistic effectiveness.

Trump's linguistic effectiveness

- Result from frequency Analysis: The most frequent words found in his transcripts are 'will', 'country', 'new', 'america', 'clinton', 'hillary', 'isis', 'people', 'defense', 'radic' etc.
- By performing sentiment analysis and topic modeling, we can understand that trump has always been talking of developing a new America.
- He has always expressed his views on terrorism and means to fight against it.
- From the three transcripts, we understand that he loves talking about his opponent and the drawbacks of having a democratic government.
- He also mentions his views of bringing back jobs to the locals and improving American economy.
- His usage of words, both negative and positive, are very strong and effective.
- Frequent topics of discussion mainly include: