USAID DC

Deon, Akua, Gerardo, Sam

4/20/2020

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

#install needed packages (only run this once)  
#install.packages('tm')  
#install.packages('tidytext')  
#install.packages('tidyr')  
#install.packages('dplyr')  
#install.packages('NLP')  
#install.packages('ggplot2')  
#install.packages('wordcloud')  
#install.packages('topicmodels')  
#install.packages('quanteda')  
#install.packages('tm')  
#install.packages('SnowballC')  
#install.packages('corpus')  
#install.packages('textdata')

#Library the packages  
library(tm)

## Loading required package: NLP

library(tidytext)

## Warning: package 'tidytext' was built under R version 3.6.2

library(tidyr)  
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(NLP)  
library(ggplot2)

##   
## Attaching package: 'ggplot2'

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

library(wordcloud)

## Loading required package: RColorBrewer

library(topicmodels)

## Warning: package 'topicmodels' was built under R version 3.6.2

library(quanteda)

## Package version: 2.0.1

## Parallel computing: 2 of 4 threads used.

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

##   
## Attaching package: 'quanteda'

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

## The following objects are masked from 'package:NLP':  
##   
## meta, meta<-

## The following object is masked from 'package:utils':  
##   
## View

library(SnowballC)

## Warning: package 'SnowballC' was built under R version 3.6.2

library(corpus)

## Warning: package 'corpus' was built under R version 3.6.2

#Running the Jaccard Index  
setwd("~/Desktop/USAID Final Project")  
data <- Corpus(DirSource("DCTxtfiles"))  
  
dfmat <- dfm(corpus\_subset(corpus(data)),  
remove\_punct = TRUE, remove = stopwords("english"))  
(tstat1 <- textstat\_simil(dfmat, method = "jaccard", margin = "documents"))

## textstat\_simil object; method = "jaccard"  
## p1.txt p10.txt p11.txt p12.txt p13.txt p14.txt p15.txt p16.txt  
## p1.txt 1.0000 0.0526 0.0577 0.0785 0.0605 0.0613 0.0360 0.0311  
## p10.txt 0.0526 1.0000 0.0406 0.0621 0.0488 0.0609 0.0515 0.0385  
## p11.txt 0.0577 0.0406 1.0000 0.0547 0.0737 0.0677 0.0457 0.0550  
## p12.txt 0.0785 0.0621 0.0547 1.0000 0.0763 0.0672 0.0568 0.0499  
## p13.txt 0.0605 0.0488 0.0737 0.0763 1.0000 0.0890 0.0746 0.0889  
## p14.txt 0.0613 0.0609 0.0677 0.0672 0.0890 1.0000 0.0605 0.0563  
## p15.txt 0.0360 0.0515 0.0457 0.0568 0.0746 0.0605 1.0000 0.2674  
## p16.txt 0.0311 0.0385 0.0550 0.0499 0.0889 0.0563 0.2674 1.0000  
## p17.txt 0.0601 0.0462 0.0636 0.0494 0.0919 0.0986 0.0416 0.0369  
## p18.txt 0.0611 0.0371 0.0657 0.0533 0.0748 0.0947 0.0505 0.0412  
## p19.txt 0.0403 0.0625 0.0555 0.0556 0.0958 0.0592 0.0557 0.0433  
## p2.txt 0.1093 0.0572 0.0527 0.0987 0.0446 0.0531 0.0413 0.0270  
## p20.txt 0.0742 0.0431 0.0870 0.1183 0.1179 0.0832 0.0516 0.0536  
## p21.txt 0.0617 0.0541 0.0478 0.1333 0.0871 0.0657 0.0564 0.0599  
## p22.txt 0.0530 0.0431 0.0691 0.0500 0.0640 0.0544 0.0428 0.0427  
## p23.txt 0.0325 0.0309 0.0270 0.0773 0.0575 0.0365 0.0594 0.0606  
## p24.txt 0.0580 0.0525 0.0702 0.0604 0.1046 0.0850 0.0841 0.0809  
## p25.txt 0.0522 0.0585 0.0457 0.0701 0.0731 0.0648 0.1355 0.1379  
## p26.txt 0.0340 0.0402 0.0515 0.0213 0.0472 0.0295 0.0519 0.0427  
## p27.txt 0.0748 0.0697 0.0801 0.0787 0.0859 0.1370 0.0744 0.0567  
## p28.txt 0.0591 0.0565 0.0854 0.0645 0.0924 0.0688 0.0665 0.0567  
## p29.txt 0.0631 0.0512 0.0859 0.0430 0.0833 0.0686 0.0445 0.0445  
## p3.txt 0.0565 0.0438 0.0712 0.0648 0.1259 0.0855 0.0708 0.0656  
## p30.txt 0.0523 0.0534 0.0731 0.0415 0.0745 0.0874 0.0530 0.0481  
## p4.txt 0.0669 0.0314 0.0669 0.0990 0.0913 0.0724 0.0559 0.0520  
## p5.txt 0.0995 0.0480 0.0623 0.1431 0.0956 0.0705 0.0530 0.0425  
## p6.txt 0.0482 0.0601 0.0377 0.0438 0.0476 0.0454 0.0464 0.0430  
## p7.txt 0.0789 0.0907 0.0517 0.0696 0.0435 0.0668 0.0364 0.0406  
## p8.txt 0.0636 0.0552 0.0892 0.0448 0.0791 0.0731 0.0653 0.0619  
## p9.txt 0.0683 0.0913 0.0728 0.0696 0.0984 0.1086 0.0611 0.0519  
## p17.txt p18.txt p19.txt p2.txt p20.txt p21.txt p22.txt p23.txt  
## p1.txt 0.0601 0.0611 0.0403 0.1093 0.0742 0.0617 0.0530 0.0325  
## p10.txt 0.0462 0.0371 0.0625 0.0572 0.0431 0.0541 0.0431 0.0309  
## p11.txt 0.0636 0.0657 0.0555 0.0527 0.0870 0.0478 0.0691 0.0270  
## p12.txt 0.0494 0.0533 0.0556 0.0987 0.1183 0.1333 0.0500 0.0773  
## p13.txt 0.0919 0.0748 0.0958 0.0446 0.1179 0.0871 0.0640 0.0575  
## p14.txt 0.0986 0.0947 0.0592 0.0531 0.0832 0.0657 0.0544 0.0365  
## p15.txt 0.0416 0.0505 0.0557 0.0413 0.0516 0.0564 0.0428 0.0594  
## p16.txt 0.0369 0.0412 0.0433 0.0270 0.0536 0.0599 0.0427 0.0606  
## p17.txt 1.0000 0.0697 0.0803 0.0625 0.0711 0.0465 0.0362 0.0389  
## p18.txt 0.0697 1.0000 0.0702 0.0469 0.0585 0.0405 0.0607 0.0238  
## p19.txt 0.0803 0.0702 1.0000 0.0406 0.0808 0.0524 0.0602 0.0646  
## p2.txt 0.0625 0.0469 0.0406 1.0000 0.0779 0.0811 0.0528 0.0410  
## p20.txt 0.0711 0.0585 0.0808 0.0779 1.0000 0.1069 0.0637 0.0428  
## p21.txt 0.0465 0.0405 0.0524 0.0811 0.1069 1.0000 0.0603 0.0519  
## p22.txt 0.0362 0.0607 0.0602 0.0528 0.0637 0.0603 1.0000 0.0316  
## p23.txt 0.0389 0.0238 0.0646 0.0410 0.0428 0.0519 0.0316 1.0000  
## p24.txt 0.0808 0.0803 0.0773 0.0643 0.0859 0.0713 0.0656 0.0557  
## p25.txt 0.0383 0.0439 0.0463 0.0558 0.0510 0.0665 0.0449 0.0508  
## p26.txt 0.0639 0.0401 0.0504 0.0504 0.0452 0.0215 0.0475 0.0392  
## p27.txt 0.1340 0.0768 0.0779 0.0868 0.1102 0.0808 0.0672 0.0546  
## p28.txt 0.0610 0.0683 0.0653 0.0509 0.0810 0.0630 0.0814 0.0268  
## p29.txt 0.0988 0.0715 0.0779 0.0515 0.0784 0.0606 0.0604 0.0516  
## p3.txt 0.0614 0.0415 0.0522 0.0647 0.0931 0.0815 0.0569 0.0414  
## p30.txt 0.0829 0.0701 0.0850 0.0637 0.0652 0.0353 0.0541 0.0484  
## p4.txt 0.0675 0.0580 0.0694 0.0771 0.1019 0.0572 0.0672 0.0640  
## p5.txt 0.0564 0.0582 0.0695 0.1129 0.1250 0.1222 0.0675 0.0594  
## p6.txt 0.0494 0.0495 0.0578 0.0492 0.0316 0.0309 0.0254 0.0319  
## p7.txt 0.0637 0.0699 0.0670 0.0759 0.0493 0.0460 0.0502 0.0405  
## p8.txt 0.0919 0.0673 0.0661 0.0609 0.0719 0.0490 0.0663 0.0403  
## p9.txt 0.0764 0.0815 0.0668 0.0684 0.0932 0.0641 0.0565 0.0394  
## p24.txt p25.txt p26.txt p27.txt p28.txt p29.txt p3.txt p30.txt  
## p1.txt 0.0580 0.0522 0.0340 0.0748 0.0591 0.0631 0.0565 0.0523  
## p10.txt 0.0525 0.0585 0.0402 0.0697 0.0565 0.0512 0.0438 0.0534  
## p11.txt 0.0702 0.0457 0.0515 0.0801 0.0854 0.0859 0.0712 0.0731  
## p12.txt 0.0604 0.0701 0.0213 0.0787 0.0645 0.0430 0.0648 0.0415  
## p13.txt 0.1046 0.0731 0.0472 0.0859 0.0924 0.0833 0.1259 0.0745  
## p14.txt 0.0850 0.0648 0.0295 0.1370 0.0688 0.0686 0.0855 0.0874  
## p15.txt 0.0841 0.1355 0.0519 0.0744 0.0665 0.0445 0.0708 0.0530  
## p16.txt 0.0809 0.1379 0.0427 0.0567 0.0567 0.0445 0.0656 0.0481  
## p17.txt 0.0808 0.0383 0.0639 0.1340 0.0610 0.0988 0.0614 0.0829  
## p18.txt 0.0803 0.0439 0.0401 0.0768 0.0683 0.0715 0.0415 0.0701  
## p19.txt 0.0773 0.0463 0.0504 0.0779 0.0653 0.0779 0.0522 0.0850  
## p2.txt 0.0643 0.0558 0.0504 0.0868 0.0509 0.0515 0.0647 0.0637  
## p20.txt 0.0859 0.0510 0.0452 0.1102 0.0810 0.0784 0.0931 0.0652  
## p21.txt 0.0713 0.0665 0.0215 0.0808 0.0630 0.0606 0.0815 0.0353  
## p22.txt 0.0656 0.0449 0.0475 0.0672 0.0814 0.0604 0.0569 0.0541  
## p23.txt 0.0557 0.0508 0.0392 0.0546 0.0268 0.0516 0.0414 0.0484  
## p24.txt 1.0000 0.0660 0.0481 0.1141 0.0857 0.0769 0.0666 0.0741  
## p25.txt 0.0660 1.0000 0.0433 0.0704 0.0586 0.0391 0.0557 0.0476  
## p26.txt 0.0481 0.0433 1.0000 0.0641 0.0606 0.1722 0.0333 0.1620  
## p27.txt 0.1141 0.0704 0.0641 1.0000 0.0922 0.0951 0.1003 0.1064  
## p28.txt 0.0857 0.0586 0.0606 0.0922 1.0000 0.0838 0.0907 0.0622  
## p29.txt 0.0769 0.0391 0.1722 0.0951 0.0838 1.0000 0.0693 0.1893  
## p3.txt 0.0666 0.0557 0.0333 0.1003 0.0907 0.0693 1.0000 0.0575  
## p30.txt 0.0741 0.0476 0.1620 0.1064 0.0622 0.1893 0.0575 1.0000  
## p4.txt 0.0991 0.0400 0.0342 0.0968 0.0726 0.0530 0.0802 0.0732  
## p5.txt 0.0760 0.0617 0.0350 0.0894 0.0806 0.0623 0.0776 0.0493  
## p6.txt 0.0343 0.0941 0.0569 0.0525 0.0492 0.0655 0.0330 0.0572  
## p7.txt 0.0510 0.0574 0.0662 0.0779 0.0511 0.0870 0.0519 0.0897  
## p8.txt 0.0759 0.0575 0.1172 0.0991 0.0990 0.1694 0.0807 0.1553  
## p9.txt 0.0938 0.0426 0.0364 0.1107 0.0867 0.0654 0.0776 0.0631  
## p4.txt p5.txt p6.txt p7.txt p8.txt p9.txt  
## p1.txt 0.0669 0.0995 0.0482 0.0789 0.0636 0.0683  
## p10.txt 0.0314 0.0480 0.0601 0.0907 0.0552 0.0913  
## p11.txt 0.0669 0.0623 0.0377 0.0517 0.0892 0.0728  
## p12.txt 0.0990 0.1431 0.0438 0.0696 0.0448 0.0696  
## p13.txt 0.0913 0.0956 0.0476 0.0435 0.0791 0.0984  
## p14.txt 0.0724 0.0705 0.0454 0.0668 0.0731 0.1086  
## p15.txt 0.0559 0.0530 0.0464 0.0364 0.0653 0.0611  
## p16.txt 0.0520 0.0425 0.0430 0.0406 0.0619 0.0519  
## p17.txt 0.0675 0.0564 0.0494 0.0637 0.0919 0.0764  
## p18.txt 0.0580 0.0582 0.0495 0.0699 0.0673 0.0815  
## p19.txt 0.0694 0.0695 0.0578 0.0670 0.0661 0.0668  
## p2.txt 0.0771 0.1129 0.0492 0.0759 0.0609 0.0684  
## p20.txt 0.1019 0.1250 0.0316 0.0493 0.0719 0.0932  
## p21.txt 0.0572 0.1222 0.0309 0.0460 0.0490 0.0641  
## p22.txt 0.0672 0.0675 0.0254 0.0502 0.0663 0.0565  
## p23.txt 0.0640 0.0594 0.0319 0.0405 0.0403 0.0394  
## p24.txt 0.0991 0.0760 0.0343 0.0510 0.0759 0.0938  
## p25.txt 0.0400 0.0617 0.0941 0.0574 0.0575 0.0426  
## p26.txt 0.0342 0.0350 0.0569 0.0662 0.1172 0.0364  
## p27.txt 0.0968 0.0894 0.0525 0.0779 0.0991 0.1107  
## p28.txt 0.0726 0.0806 0.0492 0.0511 0.0990 0.0867  
## p29.txt 0.0530 0.0623 0.0655 0.0870 0.1694 0.0654  
## p3.txt 0.0802 0.0776 0.0330 0.0519 0.0807 0.0776  
## p30.txt 0.0732 0.0493 0.0572 0.0897 0.1553 0.0631  
## p4.txt 1.0000 0.0838 0.0382 0.0625 0.0607 0.0836  
## p5.txt 0.0838 1.0000 0.0487 0.0665 0.0605 0.0634  
## p6.txt 0.0382 0.0487 1.0000 0.0536 0.0520 0.0316  
## p7.txt 0.0625 0.0665 0.0536 1.0000 0.0874 0.0524  
## p8.txt 0.0607 0.0605 0.0520 0.0874 1.0000 0.0683  
## p9.txt 0.0836 0.0634 0.0316 0.0524 0.0683 1.0000

#Run the Text Analysis  
library(tidytext, pos=15)  
library(SnowballC)  
library(tm, pos=16)  
library(ggplot2)  
  
library(NLP, pos=16)  
library(tm, pos=16)  
library(dplyr)  
  
library(textdata)  
library(topicmodels)  
library(tidyr)  
library(tm)  
library(SnowballC)  
b <- Corpus(DirSource("DCTxtfiles"))  
b <- tm\_map(b, stripWhitespace)  
b <- tm\_map(b, removeNumbers)  
b <- tm\_map(b, removePunctuation)  
b <- tm\_map(b,content\_transformer(tolower))  
b<- tm\_map(b, removeWords, c(stopwords("english"),"]-"))  
  
dtmc<- DocumentTermMatrix(b)  
dtmc

## <<DocumentTermMatrix (documents: 30, terms: 4743)>>  
## Non-/sparse entries: 9578/132712  
## Sparsity : 93%  
## Maximal term length: 21  
## Weighting : term frequency (tf)

termsc <- Terms(dtmc)  
head(terms)

##   
## 1 new("standardGeneric", .Data = function (x, ...)   
## 2 standardGeneric("terms"), generic = structure("terms", package = "stats"),   
## 3 package = "stats", group = list(), valueClass = character(0),   
## 4 signature = "x", default = new("derivedDefaultMethod", .Data = function (x,   
## 5 ...)   
## 6 UseMethod("terms"), target = new("signature", .Data = "ANY",

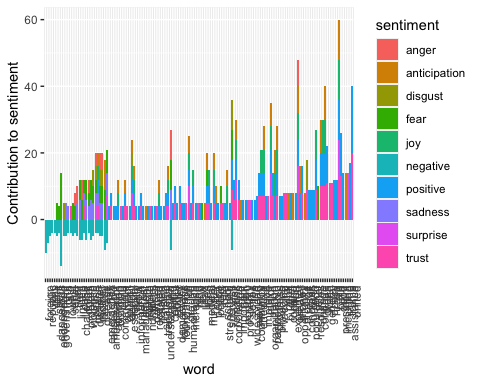
library(tidytext)  
library(dplyr)  
dtm\_tdc <- tidy(dtmc)  
dtm\_tdc

## # A tibble: 9,578 x 3  
## document term count  
## <chr> <chr> <dbl>  
## 1 p1.txt abating 1  
## 2 p1.txt ability 3  
## 3 p1.txt abounding 1  
## 4 p1.txt abroad 1  
## 5 p1.txt academician 1  
## 6 p1.txt acceleration 1  
## 7 p1.txt acceptable 1  
## 8 p1.txt accepted 2  
## 9 p1.txt accessible 2  
## 10 p1.txt according 7  
## # … with 9,568 more rows

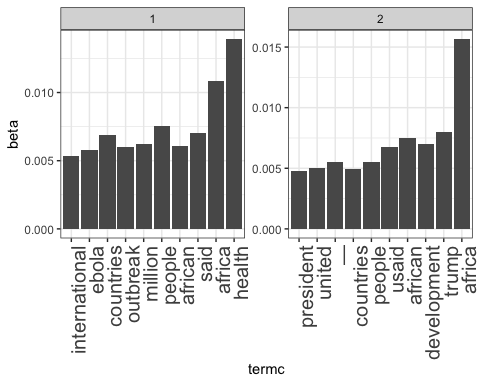
# rename the "term" column to "word"  
colnames(dtm\_tdc) <- c("document", "word", "count")  
library(textdata)  
sent <- get\_sentiments("nrc")  
nrc\_word\_countsc <- dtm\_tdc %>% inner\_join(sent, by="word") %>% count(word,sentiment, sort=TRUE) %>% ungroup()  
nrc\_word\_countsc

## # A tibble: 2,034 x 3  
## word sentiment n  
## <chr> <chr> <int>  
## 1 united positive 20  
## 2 united trust 20  
## 3 assistance positive 17  
## 4 government fear 14  
## 5 government negative 14  
## 6 including positive 14  
## 7 time anticipation 14  
## 8 trump surprise 14  
## 9 president positive 13  
## 10 president trust 13  
## # … with 2,024 more rows

library(ggplot2)  
nrc\_word\_countsc %>%  
filter(n > 3) %>%  
mutate(n = ifelse(sentiment == "negative", -n, n)) %>%  
mutate(word = reorder(word, n)) %>%  
ggplot(aes(word, n, fill = sentiment)) +  
geom\_bar(stat = "identity") +  
theme(axis.text.x = element\_text(angle = 90, hjust = 1)) + ylab("Contribution to sentiment")



library(tidytext)  
library(topicmodels)  
dtm\_ldac <- LDA(dtmc, k = 2, control = list(seed = 1234))  
dtm\_topicsc <- tidy(dtm\_ldac, matrix = "beta")  
library(ggplot2)  
library(dplyr)  
dtm\_top\_termsc <- dtm\_topicsc %>%  
 group\_by(topic) %>%  
top\_n(10, beta) %>%  
ungroup() %>%  
arrange(topic, -beta)  
library(NLP, pos=15)  
library(tidytext, pos=18)  
library(ggplot2, pos=20)  
library(tidyr)  
beta\_spreadc <- dtm\_topicsc %>%  
mutate(topic = paste0("topic", topic)) %>%  
spread(topic, beta) %>%  
filter(topic1 > .001 | topic2 > .001) %>%  
mutate(log\_ratio = log2(topic2 / topic1))  
library(ggplot2)  
theme\_set(theme\_bw())  
dtm\_top\_termsc %>%  
mutate(termc = reorder(term, beta)) %>%  
ggplot(aes(termc, beta)) +  
geom\_bar(stat = "identity") +  
facet\_wrap(~ topic, scales = "free") +  
theme(axis.text.x = element\_text(size = 15, angle = 90, hjust = 1))

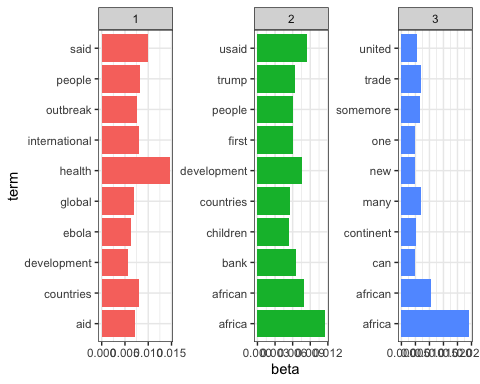


#Create and Clean Corpus  
setwd("~/Desktop/USAID Final Project")  
Combined\_corpus <- Corpus(DirSource("DCTxtfiles"))  
Combined\_corpus <- tm\_map(Combined\_corpus, removePunctuation)  
Combined\_corpus <- tm\_map(Combined\_corpus, tolower)  
Combined\_corpus <- tm\_map(Combined\_corpus, removeWords, stopwords("english"))

#Run LDA analysis to determine the topics  
Combined\_df <- DocumentTermMatrix(Combined\_corpus)  
Combined\_lda <- LDA(Combined\_df, k = 3, control = list(seed = 1234))  
Combined\_topics <- tidy(Combined\_lda, matrix = "beta")  
Combined\_top\_terms <- Combined\_topics %>% group\_by(topic) %>% top\_n(10, beta) %>%arrange(topic, -beta)  
  
#Display the topic model using ggplot  
Combined\_top\_terms %>% mutate(term = reorder(term, beta)) %>% ggplot(aes(term, beta, fill= factor(topic))) + geom\_col(show.legend = FALSE) + facet\_wrap(~topic, scales = "free") +coord\_flip()

## Warning in mutate\_impl(.data, dots, caller\_env()): Unequal factor levels:  
## coercing to character

## Warning in mutate\_impl(.data, dots, caller\_env()): binding character and  
## factor vector, coercing into character vector  
  
## Warning in mutate\_impl(.data, dots, caller\_env()): binding character and  
## factor vector, coercing into character vector  
  
## Warning in mutate\_impl(.data, dots, caller\_env()): binding character and  
## factor vector, coercing into character vector



#read text from each publication articles(in a .text file)  
setwd("~/Desktop/USAID Final Project/DCTxtfiles/")  
txt1 <- readLines("p1.txt")  
txt2 <- readLines("p2.txt")  
txt3 <- readLines("p3.txt")  
txt4 <- readLines("p4.txt")  
txt5 <- readLines("p5.txt")  
txt6 <- readLines("p6.txt")

## Warning in readLines("p6.txt"): incomplete final line found on 'p6.txt'

txt7 <- readLines("p7.txt")  
txt8 <- readLines("p8.txt")  
txt9 <- readLines("p9.txt")  
txt10 <- readLines("p10.txt")  
txt11 <- readLines("p11.txt")  
txt12 <- readLines("p12.txt")  
txt13 <- readLines("p13.txt")  
txt14 <- readLines("p14.txt")  
txt15 <- readLines("p15.txt")  
txt16 <- readLines("p16.txt")  
txt17 <- readLines("p17.txt")  
txt18 <- readLines("p18.txt")  
txt19 <- readLines("p19.txt")

## Warning in readLines("p19.txt"): incomplete final line found on 'p19.txt'

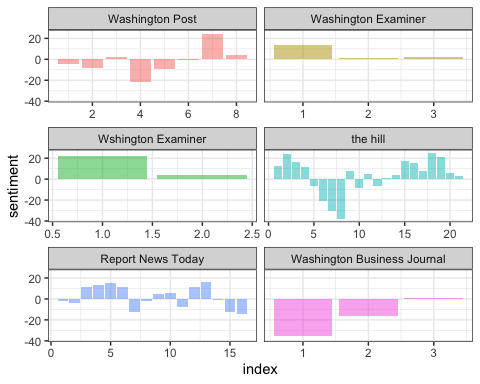
txt20 <- readLines("p20.txt")  
txt21 <- readLines("p21.txt")  
txt22 <- readLines("p22.txt")  
txt23 <- readLines("p23.txt")  
txt24 <- readLines("p24.txt")  
txt25 <- readLines("p25.txt")  
txt26 <- readLines("p26.txt")  
txt27 <- readLines("p27.txt")  
txt28 <- readLines("p28.txt")  
txt29 <- readLines("p29.txt")  
txt30 <- readLines("p30.txt")

titles <- c("Washington Post", "Washington Examiner", "Wshington Examiner", "the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill","the hill", "Report News Today","Report News Today","Report News Today","Report News Today","Report News Today","Report News Today","Report News Today","Report News Today","Report News Today","Report News Today","Report News Today", "Washington Business Journal", "Report News Today")#, #"NBC Washington", "NBC Washington" )  
books <- list(txt1, txt2, txt3, txt4, txt5, txt6, txt7, txt8, txt9, txt10, txt11, txt12, txt13, txt14, txt15, txt16, txt17, txt18, txt19, txt20, txt21, txt22, txt23, txt24, txt25, txt26, txt27, txt28, txt29, txt30)   
  
series <- tibble()

#Sentiment Analysis  
for( i in seq\_along(titles)) {  
 clean <- tibble(chapter = seq\_along(books[i]),  
 text = books[[i]]) %>%  
 unnest\_tokens(word, text) %>%  
 mutate(book = titles[i]) %>%  
 select(book, everything())  
 series <-rbind(series, clean)  
}  
  
series$book <- factor(series$book, levels = rev(unique(titles)))

series %>%   
 group\_by(book) %>%  
 mutate(word\_count = 1:n(),   
 index = word\_count %/% 500+ 1) %>%   
 inner\_join(get\_sentiments("bing")) %>%  
 count(book, index = index , sentiment)%>%  
 ungroup() %>% spread(sentiment, n, fill = 0) %>%  
 mutate(sentiment = positive - negative,book = factor(book, levels = unique(titles))) %>%  
 ggplot(aes(index, sentiment, fill = book)) +geom\_bar(alpha = 0.5, stat = "identity", show.legend = FALSE) + facet\_wrap(~ book, ncol = 2,scales = "free\_x")

## Joining, by = "word"



Everything below here is where the code stops working, I commented it out so that it would not run.