FIrst.R

Um Ar

2021-05-22

```
# Importing the required libraries
library(quantmod)
```

```
## Warning: package 'quantmod' was built under R version 4.0.5
```

```
## Loading required package: xts
```

```
## Warning: package 'xts' was built under R version 4.0.5
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 4.0.5
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: TTR
```

```
## Warning: package 'TTR' was built under R version 4.0.5
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(magrittr)
```

```
## Warning: package 'magrittr' was built under R version 4.0.3
```

```
library(visdat)
```

```
## Warning: package 'visdat' was built under R version 4.0.5
```

```
library(naniar)
```

```
## Warning: package 'naniar' was built under R version 4.0.5
```

```
library(stringr)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.3
```

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

## The following objects are masked from 'package:xts':
##
##     first, last

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

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

library(quanteda)

## Package version: 2.1.2

## Parallel computing: 2 of 8 threads used.

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

##
## Attaching package: 'quanteda'

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

# library(tensorflow)
# library(keras)
library(reticulate)

## Warning: package 'reticulate' was built under R version 4.0.3

library(purrr)

##
## Attaching package: 'purrr'

## The following object is masked from 'package:magrittr':
##
##     set_names

library(topicmodels)

## Warning: package 'topicmodels' was built under R version 4.0.3

library(tm)

## Warning: package 'tm' was built under R version 4.0.3

## Loading required package: NLP
```

```
## Warning: package 'NLP' was built under R version 4.0.3

##
## Attaching package: 'NLP'

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

##
## Attaching package: 'tm'

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

library(doSNOW)

## Warning: package 'doSNOW' was built under R version 4.0.3

## Loading required package: foreach

## Warning: package 'foreach' was built under R version 4.0.3

##
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':
##
##     accumulate, when

## Loading required package: iterators

## Warning: package 'iterators' was built under R version 4.0.3

## Loading required package: snow

library(dplyr)
library(LDAvis)

## Warning: package 'LDAvis' was built under R version 4.0.3

library(lda)

## Warning: package 'lda' was built under R version 4.0.3

library(pals)

## Warning: package 'pals' was built under R version 4.0.3

library(pacman)

## Warning: package 'pacman' was built under R version 4.0.3

library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 4.0.3

library(igraph)

## Warning: package 'igraph' was built under R version 4.0.3

##
## Attaching package: 'igraph'

## The following objects are masked from 'package:purrr':
##
##     compose, simplify

## The following object is masked from 'package:quanteda':
##
##     as.igraph

## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union

library(servr)

## Warning: package 'servr' was built under R version 4.0.3

library(plotly)

## Warning: package 'plotly' was built under R version 4.0.3

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

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

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:igraph':
##
##     groups

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout

library(webshot)

## Warning: package 'webshot' was built under R version 4.0.3

library(htmlwidgets)

## Warning: package 'htmlwidgets' was built under R version 4.0.3

library(wordcloud)

## Loading required package: RColorBrewer

library(caret)

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift

# Reading in the data
Data <- read.csv(file = "C:/Users/Um Ar/R Projects/UN
MINOR/Datasets/Nai/train.csv",
                 header = T)

# Data2 <- read.csv(file = "C:/Users/Um Ar/R Projects/UN
MINOR/Datasets/data.csv")

# colnames(Data)
# colnames(Data2)

# colnames(Data2)[2] <- "title"
# colnames(Data2)[3] <- "text"
# colnames(Data2)[4] <- "label"

# Data2 <- Data2[,-1]
# Data <- Data[,-1]
```

```
# Data <- Data[,-2]

# FN_Data <- rbind(Data, Data2)


# Data <- FN_Data


# Exploratory Data Analysis
summary(Data)
```

```
##        id           title              author              text          
##  Min.   :    0   Length:20800       Length:20800       Length:20800      
##  1st Qu.: 5200   Class :character   Class :character   Class :character  
##  Median :10400   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :10400                                                           
##  3rd Qu.:15599                                                           
##  Max.   :20799                                                           
##      label       
##  Min.   :0.0000  
##  1st Qu.:0.0000  
##  Median :1.0000  
##  Mean   :0.5006  
##  3rd Qu.:1.0000  
##  Max.   :1.0000
```

```
head(Data)
```

```
##   id
## 1  0
## 2  1
## 3  2
## 4  3
## 5  4
## 6  5
##
title
## 1                                          House Dem Aide: We Didnâ\200\231t
Even See Comeyâ\200\231s Letter Until Jason Chaffetz Tweeted It
## 2
FLYNN: Hillary Clinton, Big Woman on Campus - Breitbart
## 3
Why the Truth Might Get You Fired
## 4                                                                      15
Civilians Killed In Single US Airstrike Have Been Identified
## 5                                  Iranian woman jailed for fictional
unpublished story about woman stoned to death for adultery
## 6 Jackie Mason: Hollywood Would Love Trump if He Bombed North Korea over
Lack of Trans Bathrooms (Exclusive Video) - Breitbart
##               author
```

```
## 1      Darrell Lucus
## 2    Daniel J. Flynn
## 3 Consortiumnews.com
## 4    Jessica Purkiss
## 5     Howard Portnoy
## 6    Daniel Nussbaum
##
text
## 1
House Dem Aide: We Didnâ\200\231t Even See Comeyâ\200\231s Letter Until Jason Chaffetz Tweeted It By Darrell Lucus on October 30, 2016 Subscribe Jason Chaffetz on the stump in American Fork, Utah ( image courtesy Michael Jolley, available under a Creative Commons-BY license) \nWith apologies to Keith Olbermann, there is no doubt who the Worst Person in The World is this weekâ\200“FBI Director James Comey. But according to a House Democratic aide, it looks like we also know who the second-worst person is as well. It turns out that when Comey sent his now-infamous letter announcing that the FBI was looking into emails that may be related to Hillary Clintonâ\200\231s email server, the ranking Democrats on the relevant committees didnâ\200\231t hear about it from Comey. They found out via a tweet from one of the Republican committee chairmen. \nAs we now know, Comey notified the Republican chairmen and Democratic ranking members of the House Intelligence, Judiciary, and Oversight committees that his agency was reviewing emails it had recently discovered in order to see if they contained classified information. Not long after this letter went out, Oversight Committee Chairman Jason Chaffetz set the political world ablaze with this tweet. FBI Dir just informed me, "The FBI has learned of the existence of emails that appear to be pertinent to the investigation." Case reopened \nâ\200” Jason Chaffetz (@jasoninthehouse) October 28, 2016 \nOf course, we now know that this was not the case . Comey was actually saying that it was reviewing the emails in light of â\200œan unrelated caseâ\200\235â\200“which we now know to be Anthony Weinerâ\200\231s sexting with a teenager. But apparently such little things as facts didnâ\200\231t matter to Chaffetz. The Utah Republican had already vowed to initiate a raft of investigations if Hillary winsâ\200“at least two yearsâ\200\231 worth, and possibly an entire termâ\200\231s worth of them. Apparently Chaffetz thought the FBI was already doing his work for himâ\200“resulting in a tweet that briefly roiled the nation before cooler heads realized it was a dud. \nBut according to a senior House Democratic aide, misreading that letter may have been the least of Chaffetzâ\200\231 sins. That aide told Shareblue that his boss and other Democrats didnâ\200\231t even know about Comeyâ\200\231s letter at the timeâ\200“and only found out when they checked Twitter. â\200œDemocratic Ranking Members on the relevant committees didnâ\200\231t receive Comeyâ\200\231s letter until after the Republican Chairmen. In fact, the Democratic Ranking Members didnâ\200\231 receive it until after the Chairman of the Oversight and Government Reform Committee, Jason Chaffetz, tweeted it out and made it public.â\200\235 \nSo letâ\200\231s see if weâ\200\231ve got this right. The FBI director tells Chaffetz and other GOP committee chairmen about a major development in a potentially politically explosive investigation, and neither Chaffetz nor his other colleagues had the courtesy to let their Democratic
```

counterparts know about it. Instead, according to this aide, he made them find out about it on Twitter. \nThere has already been talk on Daily Kos that Comey himself provided advance notice of this letter to Chaffetz and other Republicans, giving them time to turn on the spin machine. That may make for good theater, but there is nothing so far that even suggests this is the case. After all, there is nothing so far that suggests that Comey was anything other than grossly incompetent and tone-deaf. \nWhat it does suggest, however, is that Chaffetz is acting in a way that makes Dan Burton and Darrell Issa look like models of responsibility and bipartisanship. He didnâ\200\231t even have the decency to notify ranking member Elijah Cummings about something this explosive. If that doesnâ\200\231t trample on basic standards of fairness, I donâ\200\231t know what does. \nGranted, itâ\200\231s not likely that Chaffetz will have to answer for this. He sits in a ridiculously Republican district anchored in Provo and Orem; it has a Cook Partisan Voting Index of R+25, and gave Mitt Romney a punishing 78 percent of the vote in 2012. Moreover, the Republican House leadership has given its full support to Chaffetzâ\200\231 planned fishing expedition. But that doesnâ\200\231t mean we canâ\200\231t turn the hot lights on him. After all, he is a textbook example of what the House has become under Republican control. And he is also the Second Worst Person in the World. About Darrell Lucus \nDarrell is a 30-something graduate of the University of North Carolina who considers himself a journalist of the old school. An attempt to turn him into a member of the religious right in college only succeeded in turning him into the religious right's worst nightmare--a charismatic Christian who is an unapologetic liberal. His desire to stand up for those who have been scared into silence only increased when he survived an abusive three-year marriage. You may know him on Daily Kos as Christian Dem in NC . Follow him on Twitter @DarrellLucus or connect with him on Facebook . Click here to buy Darrell a Mello Yello. Connect

2

Ever get the feeling your life circles the roundabout rather than heads in a straight line toward the intended destination? [Hillary Clinton remains the big woman on campus in leafy, liberal Wellesley, Massachusetts. Everywhere else votes her most likely to don her inauguration dress for the remainder of her days the way Miss Havisham forever wore that wedding dress. Speaking of Great Expectations, Hillary Rodham overflowed with them 48 years ago when she first addressed a Wellesley graduating class. The president of the college informed those gathered in 1969 that the students needed â\200œno debate so far as I could ascertain as to who their spokesman was to beâ\200\235 (kind of the like the Democratic primaries in 2016 minus the terms unknown then even at a Seven Sisters school). â\200œI am very glad that Miss Adams made it clear that what I am speaking for today is all of us â\200” the 400 of us,â\200\235 Miss Rodham told her classmates. After appointing herself Edger Bergen to the Charlie McCarthys and Mortimer Snerds in attendance, the bespectacled in granny glasses (awarding her matronly wisdom â\200” or at least John Lennon wisdom) took issue with the previous speaker. Despite becoming the first to win election to a seat in the U. S. Senate since Reconstruction, Edward Brooke came in for criticism for calling for â\200œempathyâ\200\235 for the goals of protestors as he criticized tactics. Though Clinton in her senior thesis on Saul Alinsky lamented â\200œBlack

Power demagoguesâ\200\235 and â\200œelitist arrogance and repressive intoleranceâ\200\235 within the New Left, similar words coming out of a Republican necessitated a brief rebuttal. â\200œTrust,â\200\235 Rodham ironically observed in 1969, â\200œthis is one word that when I asked the class at our rehearsal what it was they wanted me to say for them, everyone came up to me and said â\200\230Talk about trust, talk about the lack of trust both for us and the way we feel about others. Talk about the trust bust.â\200\231 What can you say about it? What can you say about a feeling that permeates a generation and that perhaps is not even understood by those who are distrusted?â\200\235 The â\200œtrust bustâ\200\235 certainly busted Clintonâ\200\231s 2016 plans. She certainly did not even understand that people distrusted her. After Whitewater, Travelgate, the vast conspiracy, Benghazi, and the missing emails, Clinton found herself the distrusted voice on Friday. There was a load of compromising on the road to the broadening of her political horizons. And distrust from the American people â\200” Trump edged her 48 percent to 38 percent on the question immediately prior to Novemberâ\200\231s election â\200” stood as a major reason for the closing of those horizons. Clinton described her vanquisher and his supporters as embracing a â\200œlie,â\200\235 a â\200œcon,â\200\235 â\200œalternative facts,â\200\235 and â\200œa assault on truth and reason. â\200\235 She failed to explain why the American people chose his lies over her truth. â\200œAs the history majors among you here today know all too well, when people in power invent their own facts and attack those who question them, it can mark the beginning of the end of a free society,â\200\235 she offered. â\200œThat is not hyperbole. â\200\235 Like so many people to emerge from the 1960s, Hillary Clinton embarked upon a long, strange trip. From high school Goldwater Girl and Wellesley College Republican president to Democratic politician, Clinton drank in the times and the place that gave her a degree. More significantly, she went from idealist to cynic, as a comparison of her two Wellesley commencement addresses show. Way back when, she lamented that â\200œfor too long our leaders have viewed politics as the art of the possible, and the challenge now is to practice politics as the art of making what appears to be impossible possible. â\200\235 Now, as the big woman on campus but the odd woman out of the White House, she wonders how her current station is even possible. â\200œWhy arenâ\200\231t I 50 points ahead?â\200\235 she asked in September. In May she asks why she isnâ\200\231t president. The woman famously dubbed a â\200œcongenital liarâ\200\235 by Bill Safire concludes that lies did her in â\200” theirs, mind you, not hers. Getting stood up on Election Day, like finding yourself the jilted bride on your wedding day, inspires dangerous delusions.
3 Why the Truth Might Get You Fired October 29, 2016 \nThe tension between intelligence analysts and political policymakers has always been between honest assessments and desired results, with the latter often overwhelming the former, as in the Iraq War, writes Lawrence Davidson. \nBy Lawrence Davidson \nFor those who might wonder why foreign policy makers repeatedly make bad choices, some insight might be drawn from the following analysis. The action here plays out in the United States, but the lessons are probably universal. \nBack in the early spring of 2003, George W. Bush initiated the invasion of Iraq. One of his key public reasons for doing so was the claim that the countryâ\200\231s dictator, Saddam Hussein, was on the verge of

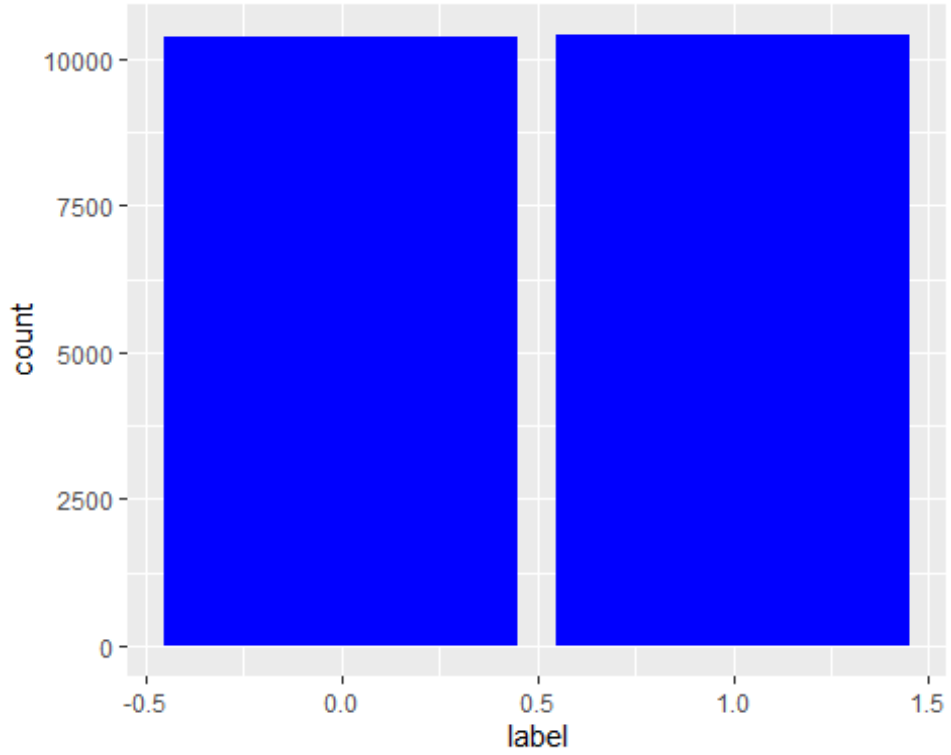
developing nuclear weapons and was hiding other weapons of mass destruction. The real reason went beyond that charge and included a long-range plan for â\200œregime changeâ\200\235 in the Middle East. President George W. Bush and Vice President Dick Cheney receive an Oval Office briefing from CIA Director George Tenet. Also present is Chief of Staff Andy Card (on right). (White House photo) \nFor our purposes, we will concentrate on the belief that Iraq was about to become a hostile nuclear power. Why did President Bush and his close associates accept this scenario so readily? \nThe short answer is Bush wanted, indeed needed, to believe it as a rationale for invading Iraq. At first he had tried to connect Saddam Hussein to the 9/11 attacks on the U.S. Though he never gave up on that stratagem, the lack of evidence made it difficult to rally an American people, already fixated on Afghanistan, to support a war against Baghdad. \nBut the nuclear weapons gambit proved more fruitful, not because there was any hard evidence for the charge, but because supposedly reliable witnesses, in the persons of exiled anti-Saddam Iraqis (many on the U.S. governmentâ\200\231s payroll), kept telling Bush and his advisers that the nuclear story was true. \nWhat we had was a U.S. leadership cadre whose worldview literally demanded a mortally dangerous Iraq, and informants who, in order to precipitate the overthrow of Saddam, were willing to tell the tale of pending atomic weapons. The strong desire to believe the tale of a nuclear Iraq lowered the threshold for proof . Likewise, the repeated assertions by assumed dependable Iraqi sources underpinned a nationwide U.S. campaign generating both fear and war fever. \nSo the U.S. and its allies insisted that the United Nations send in weapons inspectors to scour Iraq for evidence of a nuclear weapons program (as well as chemical and biological weapons). That the inspectors could find no convincing evidence only frustrated the Bush administration and soon forced its hand. \nOn March 19, 2003, Bush launched the invasion of Iraq with the expectation was that, once in occupation of the country, U.S. inspectors would surely find evidence of those nukes (or at least stockpiles of chemical and biological weapons). They did not. Their Iraqi informants had systematically lied to them. \nSocial and Behavioral Sciences to the Rescue? \nThe various U.S. intelligence agencies were thoroughly shaken by this affair, and today, 13 years later, their directors and managers are still trying to sort it out â\200“ specifically, how to tell when they are getting â\200œtrueâ\200\235 intelligence and when they are being lied to. Or, as one intelligence worker has put it, we need â\200œ help to protect us against armies of snake oil salesmen. â\200\235 To that end the CIA et al. are in the market for academic assistance. Ahmed Chalabi, head of the Iraqi National Congress, a key supplier of Iraqi defectors with bogus stories of hidden WMD. \nA â\200œpartnershipâ\200\235 is being forged between the Office of the Director of National Intelligence (ODNI), which serves as the coordinating center for the sixteen independent U.S. intelligence agencies, and the National Academies of Sciences, Engineering and Medicine . The result of this collaboration will be a â\200œ permanent Intelligence Community Studies Boardâ\200\235 to coordinate programs in â\200œsocial and behavioral science research [that] might strengthen national security .â\200\235 \nDespite this effort, it is almost certain that the â\200œsocial and behavioral sciencesâ\200\235 cannot give the spy agencies what they want â\200“ a way of detecting lies that is better than their present standard procedures of

polygraph tests and interrogations. But even if they could, it might well make no difference, because the real problem is not to be found with the liars. It is to be found with the believers. \nThe Believers \nIt is simply not true, as the ODNI leaders seem to assert, that U.S. intelligence agency personnel cannot tell, more often than not, that they are being lied to. This is the case because there are thousands of middle-echelon intelligence workers, desk officers, and specialists who know something closely approaching the truth â\200“ that is, they know pretty well what is going on in places like Afghanistan, Iraq, Syria, Libya, Israel, Palestine and elsewhere. Director of National Intelligence James Clapper (right) talks with President Barack Obama in the Oval Office, with John Brennan and other national security aides present. (Photo credit: Office of Director of National Intelligence) \nTherefore, if someone feeds them â\200œsnake oil,â\200\235 they usually know it. However, having an accurate grasp of things is often to no avail because their superiors â\200“ those who got their appointments by accepting a pre-structured worldview â\200“ have different criterion for what is â\200œtrueâ\200\235 than do the analysts. \nListen to Charles Gaukel, of the National Intelligence Council â\200“ yet another organization that acts as a meeting ground for the 16 intelligence agencies. Referring to the search for a way to avoid getting taken in by lies, Gaukel has declared, â\200œ Weâ\200\231re looking for truth. But weâ\200\231re particularly looking for truth that works. â\200\235 Now what might that mean? \nI can certainly tell you what it means historically. It means that for the power brokers, â\200œtruthâ\200\235 must match up, fit with, their worldview â\200“ their political and ideological precepts. If it does not fit, it does not â\200œwork.â\200\235 So the intelligence specialists who send their usually accurate assessments up the line to the policy makers often hit a roadblock caused by â\200œgroup think,â\200\235 ideological blinkers, and a â\200œwe know betterâ\200\235 attitude. \nOn the other hand, as long as what youâ\200\231re selling the leadership matches up with what they want to believe, you can peddle them anything: imaginary Iraqi nukes, Israel as a Western-style democracy, Saudi Arabia as an indispensable ally, Libya as a liberated country, Bashar al-Assad as the real roadblock to peace in Syria, the Strategic Defense Initiative (SDI) aka Star Wars, a world that is getting colder and not warmer, American exceptionalism in all its glory â\200“ the list is almost endless. \nWhat does this sad tale tell us? If you want to spend millions of dollars on social and behavioral science research to improve the assessment and use of intelligence, forget about the liars. What you want to look for is an antidote to the narrow-mindedness of the believers â\200“ the policymakers who seem not to be able to rise above the ideological presumptions of their class â\200“ presumptions that underpin their self-confidence as they lead us all down slippery slopes. \nIt has happened this way so often, and in so many places, that it is the source of Shakespeareâ\200\231s determination that â\200œwhat is past, is prelude.â\200\235 Our elites play out our destinies as if they have no free will â\200“ no capacity to break with structured ways of seeing. Yet the middle-echelon specialists keep sending their relatively accurate assessments up the ladder of power. Hope springs eternal.
4
Videos 15 Civilians Killed In Single US Airstrike Have Been Identified The

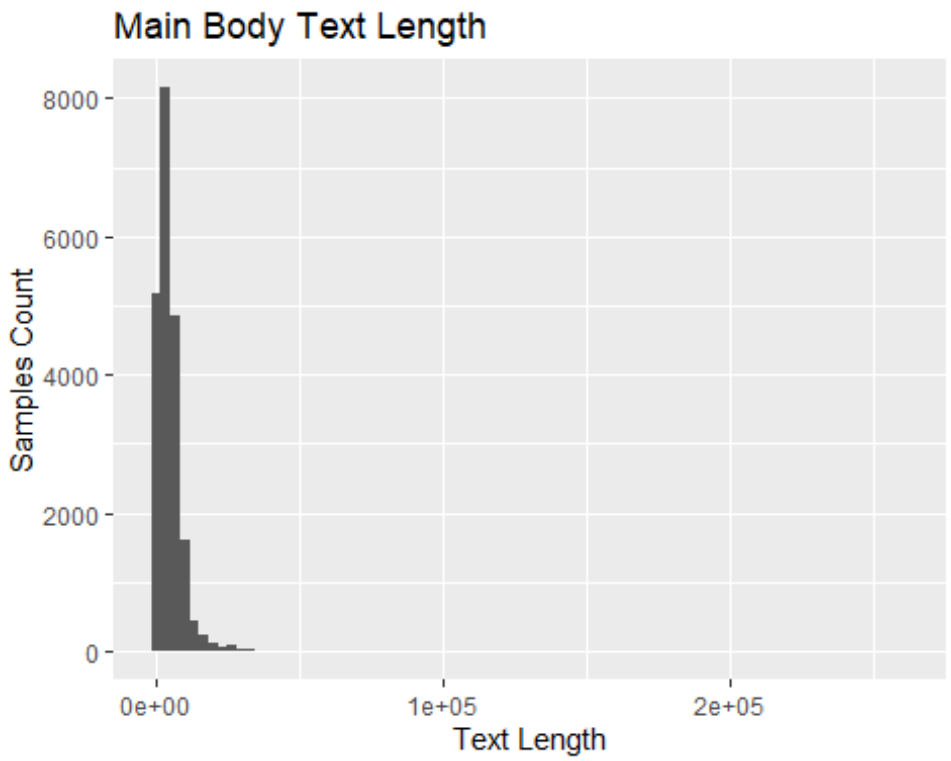
rate at which civilians are being killed by American airstrikes in Afghanistan is now higher than it was in 2014 when the US was engaged in active combat operations. Photo of Hellfire missiles being loaded onto a US military Reaper drone in Afghanistan by Staff Sgt. Brian Ferguson/U.S. Air Force. \nThe Bureau has been able to identify 15 civilians killed in a single US drone strike in Afghanistan last month â\200“ the biggest loss of civilian life in one strike since the attack on the Medecins Sans Frontieres hospital (MSF) last October. \nThe US claimed it had conducted a â\200œcounter-terrorismâ\200\235 strike against Islamic State (IS) fighters when it hit Nangarhar province with missiles on September 28. But the next day the United Nations issued an unusually rapid and strong statement saying the strike had killed 15 civilians and injured 13 others who had gathered at a house to celebrate a tribal elderâ\200\231s return from a pilgrimage to Mecca. \nThe Bureau spoke to a man named Haji Rais who said he was the owner of the house that was targeted. He said 15 people were killed and 19 others injured, and provided their names (listed below). The Bureau was able to independently verify the identities of those who died. \nRaisâ\200\231 son, a headmaster at a local school, was among them. Another man, Abdul Hakim, lost three of his sons in the attack. \nRais said he had no involvement with IS and denied US claims that IS members had visited his house before the strike. He said: â\200œI did not even speak to those sort of people on the phone let alone receiving them in my house.â\200\235 \nThe deaths amount to the biggest confirmed loss of civilian life in a single American strike in Afghanistan since the attack on the MSF hospital in Kunduz last October, which killed at least 42 people. \nThe Nangarhar strike was not the only US attack to kill civilians in September. The Bureauâ\200\231s data indicates that as many as 45 civilians and allied soldiers were killed in four American strikes in Afghanistan and Somalia that month. \nOn September 18 a pair of strikes killed eight Afghan policemen in Tarinkot, the capital of Urozgan provice. US jets reportedly hit a police checkpoint, killing one officer, before returning to target first responders. The use of this tactic â\200“ known as a â\200œdouble-tapâ\200\235 strike â\200“ is controversial because they often hit civilian rescuers. \nThe US told the Bureau it had conducted the strike against individuals firing on and posing a threat to Afghan forces. The email did not directly address the allegations of Afghan policemen being killed. \nAt the end of the month in Somalia, citizens burnt US flags on the streets of the north-central city of Galcayo after it emerged a drone attack may have unintentionally killed 22 Somali soldiers and civilians. The strike occurred on the same day as the one in Nangarhar. \nIn both the Somali and Afghan incidents, the US at first denied that any non-combatants had been killed. It is now investigating both the strikes in Nangarhar and Galcayo. \nThe rate at which civilians are being killed by American airstrikes in Afghanistan is now higher than it was in 2014 when the US was engaged in active combat operations. Name
5
Print \nAn Iranian woman has been sentenced to six years in prison after Iranâ\200\231s Revolutionary Guard searched her home and found a notebook that contained a fictional story sheâ\200\231d written about a woman who was stoned to death, according to the Eurasia Review . \nGolrokh Ebrahimi Iraee, 35, is the wife of political prisoner Arash Sadeghi, 36, who is serving a 19-

```
year prison sentence for being a human rights activist, the publication
reported. \nâ\200œWhen the intelligence unit of the Revolutionary Guards came
to arrest her husband, they raided their apartment â\200“ without a warrant
â\200“ and found drafts of stories that Ebrahimi Iraee had written,â\200\235
the article stated. \nâ\200œOne of the confiscated drafts was a story about
stoning women to death for adultery â\200“ never published, never presented
to anyone,â\200\235 the article stated. â\200œThe narrative followed the
story of a protagonist that watched a movie about stoning of women under
Islamic law for adultery.
## 6
In these trying times, Jackie Mason is the Voice of Reason. [In this
weekâ\200\231s exclusive clip for Breitbart News, Jackie discusses the
looming threat of North Korea, and explains how President Donald Trump could
win the support of the Hollywood left if the U. S. needs to strike first.
â\200œIf he decides to bomb them, the whole country will be behind him,
because everybody will realize he had no choice and that was the only thing
to do,â\200\235 Jackie says. â\200œExcept the Hollywood left. Theyâ\200\231ll
get nauseous. â\200\235 â\200œ[Trump] could win the left over,
theyâ\200\231ll fall in love with him in a minute. If he bombed them for a
better reason,â\200\235 Jackie explains. â\200œLike if they have no
transgender toilets. â\200\235 Jackie also says itâ\200\231s no surprise that
Hollywood celebrities didnâ\200\231t support Trumpâ\200\231s strike on a
Syrian airfield this month. â\200œThey were infuriated,â\200\235 he says.
â\200œBecause it might only save lives. That doesnâ\200\231t mean anything to
them. If it only saved the environment, or climate change! Theyâ\200\231d be
the happiest people in the world. â\200\235 Still, Jackie says heâ\200\231s
got nothing against Hollywood celebs. Theyâ\200\231ve got a tough life in
this country. Watch Jackieâ\200\231s latest clip above.   Follow Daniel
Nussbaum on Twitter: @dznussbaum
##   label
## 1     1
## 2     0
## 3     1
## 4     1
## 5     1
## 6     0

colnames(Data)

## [1] "id"     "title"  "author" "text"   "label"

# Creating a new feature  Text Length
Data$title <- as.character(Data$title)
Data$text <- as.character(Data$text)
Data$title_Text_Length <- nchar(Data$title)
Data$text_Text_Length <- nchar(Data$text)

library(ggplot2)
# Checking the distribution of data
```

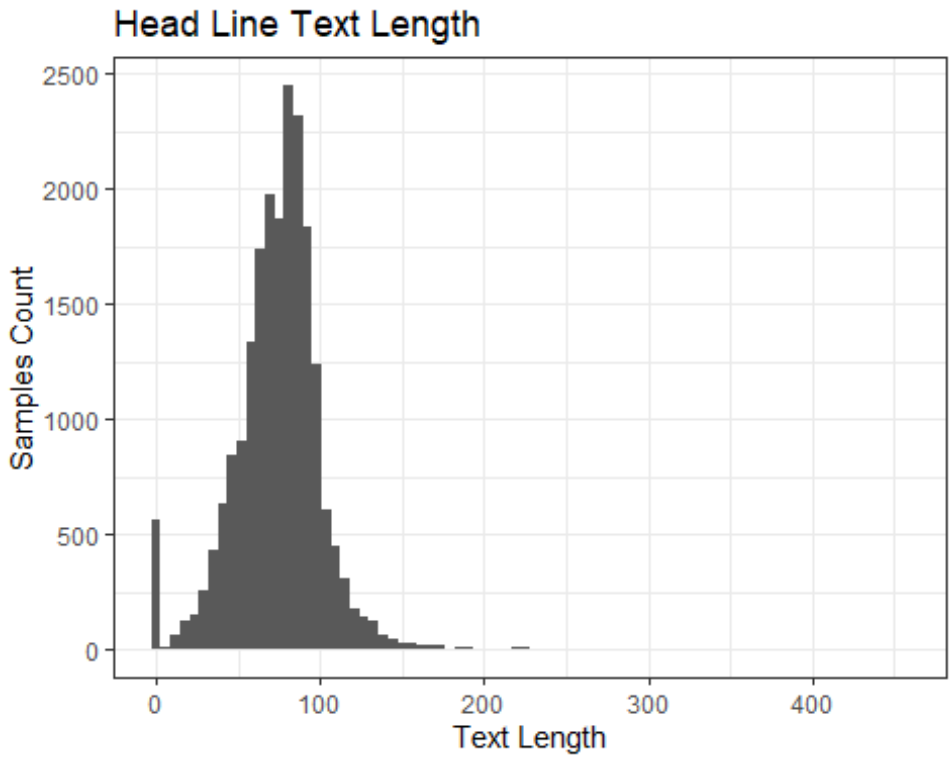
```r
ggplot(Data, aes(label)) +
  geom_bar(fill = "blue")
```



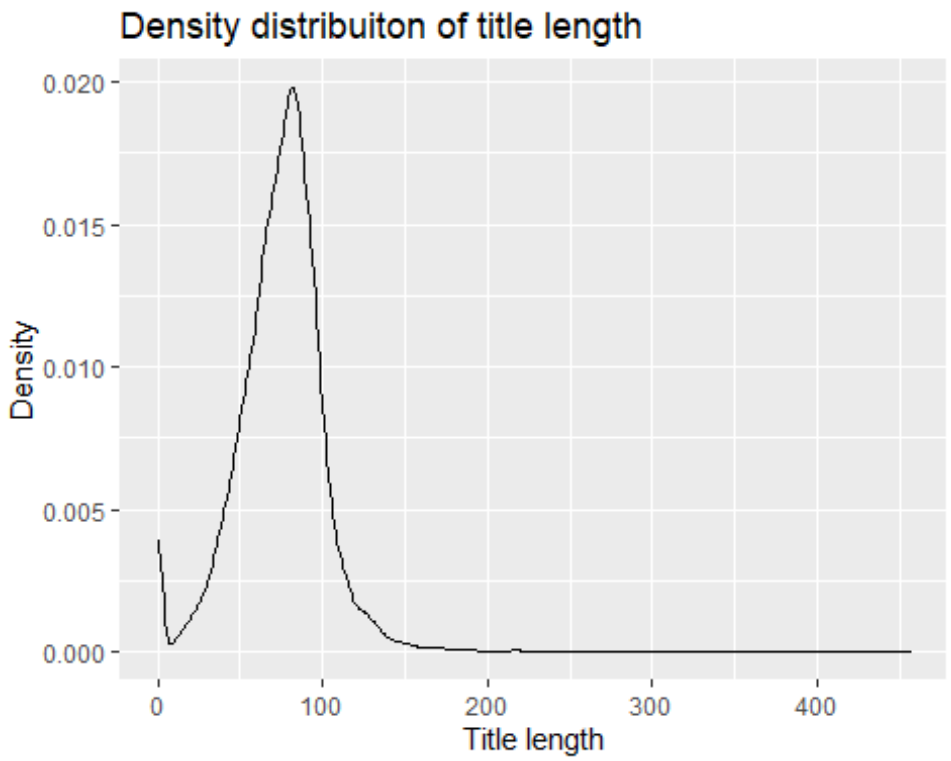

```r
# Text Length of Main body
ggplot(Data, aes(x = text_Text_Length, fill = label)) +
  geom_histogram(bins = 80) +
  labs(y = "Samples Count",
       x = "Text Length",
       title = "Main Body Text Length")
```



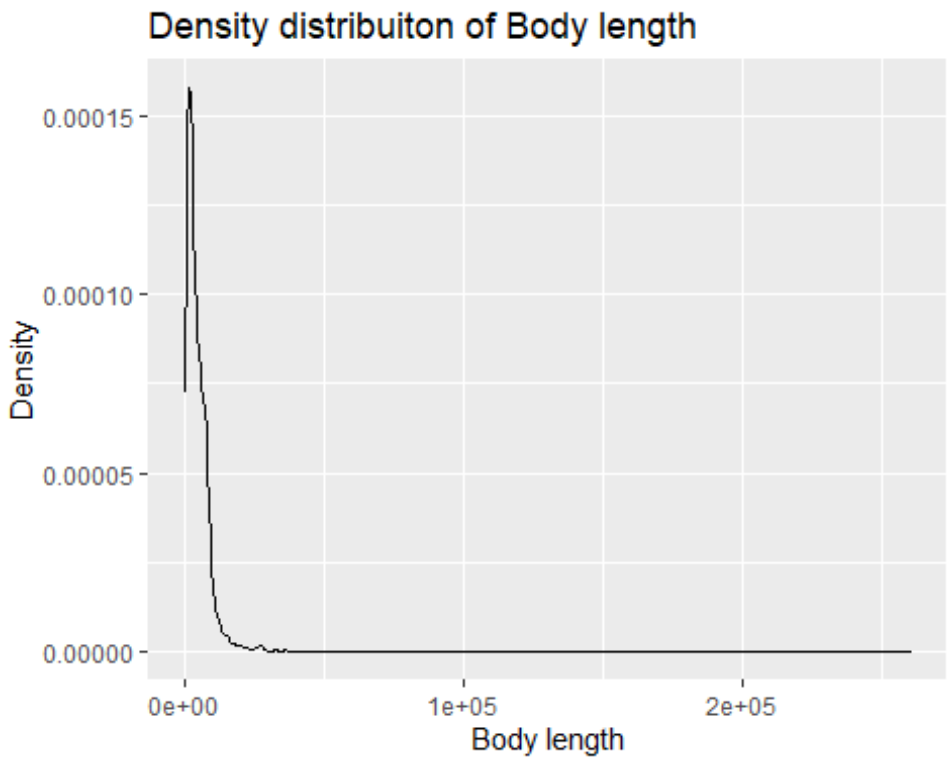

```r
# Text Length of Headlines
ggplot(data = Data, mapping = aes(x = title_Text_Length)) +
  geom_histogram(bins = 80) +
  theme_bw() +
  labs(y = "Samples Count",
       x = "Text Length",
       title = "Head Line Text Length")
```



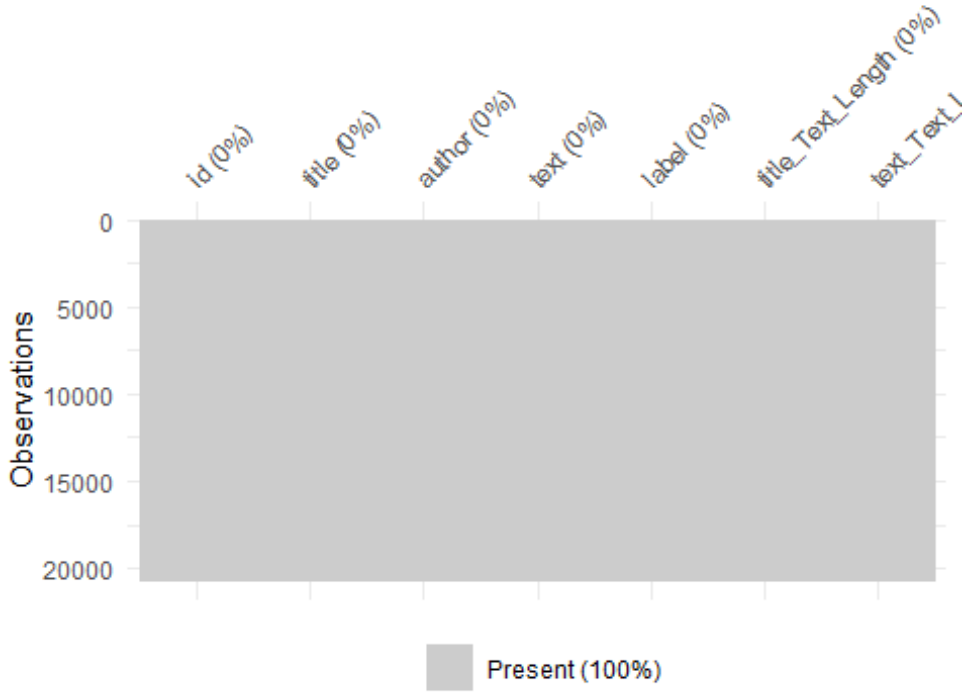

```
# Data$que <- sapply(Data$Headline, function(x)
#  length(unlist(strsplit(as.character(x), "\\?+"))))

# Count of exclamations in fake and real news
# Data %>% group_by(Label) %>% summarise(exclamations=sum(que))

# plotting histogram of title length
ggplot(Data ,aes(x = title_Text_Length, fill = label)) +
  geom_density(alpha=0.5) +
  guides(fill=guide_legend(title="News type")) +
  xlab("Title length") + ylab("Density") + theme() +
  ggtitle("Density distribuiton of title length")
```




```
ggplot(Data ,aes(x = text_Text_Length, fill = label)) +
  geom_density(alpha=0.5) +
  guides(fill=guide_legend(title="News type")) +
  xlab("Body length") + ylab("Density") + theme() +
  ggtitle("Density distribuiton of Body length")
```




```
# write.csv(Data, "FINAL_DATA.csv")

# Checking the missing values
# Heatmap of missing data
vis_miss(Data, cluster = T)
```




```
# Preprocessing
# Preprocessing the text files
corpus <- Corpus(VectorSource(Data$text))
english_words <- readLines("stopwords.txt",
                           encoding = "UTF-8")

processedCorpus <- tm_map(corpus, content_transformer(tolower))

## Warning in tm_map.SimpleCorpus(corpus, content_transformer(tolower)):
## transformation drops documents

processedCorpus <- tm_map(processedCorpus, removeWords, english_words)

## Warning in tm_map.SimpleCorpus(processedCorpus, removeWords,
english_words):
## transformation drops documents

processedCorpus <- tm_map(processedCorpus, removePunctuation,
preserve_intra_word_dashes = TRUE)

## Warning in tm_map.SimpleCorpus(processedCorpus, removePunctuation,
## preserve_intra_word_dashes = TRUE): transformation drops documents

processedCorpus <- tm_map(processedCorpus, removeNumbers)

## Warning in tm_map.SimpleCorpus(processedCorpus, removeNumbers):
transformation
## drops documents
```

```r
processedCorpus <- tm_map(processedCorpus, stemDocument, language = "en")
```

```
## Warning in tm_map.SimpleCorpus(processedCorpus, stemDocument, language =
"en"):
## transformation drops documents
```

```r
processedCorpus <- tm_map(processedCorpus, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(processedCorpus, stripWhitespace):
transformation
## drops documents
```

```r
# Changing the corpus into a document term matrix
minimumFrequency <- 7
DTM <- DocumentTermMatrix(processedCorpus,
                          control = list(bounds = list(global =
c(minimumFrequency, Inf))))

dim(DTM)
```

```
## [1] 20800 30822
```

```r
# Checking for missing values and removing them
raw.sum <- apply(DTM, 1, FUN = sum)
dfm_trimmed <- DTM[raw.sum!=0,]



# Hyperparameters for topic modelling
burnin <- 100
# Iterations
iter <- 100
# Taking every 100 iteration for further use
thin <- 10
# using 10 starting points
nstart <- 1
#Seeds for 10 starts
seed <- list(1103)
best <- T
# number of topicsrm

gc()
```

```
##            used  (Mb) gc trigger    (Mb)   max used    (Mb)
## Ncells  3036989 162.2    5902111   315.3    5902111   315.3
## Vcells 43037766 328.4 1518716063 11586.9 1898376771 14483.5
```

```r
cl <- makeCluster(6, type = "SOCK")
registerDoSNOW(cl)

# TOpic modelling
topicModel <- LDA(dfm_trimmed, k = 50, method = "Gibbs",
```

```
                  control = list(seed = seed, nstart = nstart,
                                 best = best, burnin = burnin, iter = iter,
                                 thin = thin, verbose = 25))

## K = 50; V = 30822; M = 20670
## Sampling 110 iterations!
## Iteration 25 ...
## Iteration 50 ...
## Iteration 75 ...
## Iteration 100 ...
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!
## K = 50; V = 30822; M = 20670
## Sampling 10 iterations!
## Gibbs sampling completed!

# Groping the topics
Topics <- tidy(topicModel, matrix = "beta")
top_terms <- Topics %>% group_by(topic) %>% top_n(20, beta) %>%
  ungroup() %>% arrange(topic, -beta)

# save.image("C:/Users/Um Ar/R Projects/UN MINOR/.RData")

# Taking out the top 5 terms
# TopicsT <- posterior(topicModel, dfm_trimmed)[["topics"]]
tmResult <- posterior(topicModel)
top5termsPerTopic <- terms(topicModel, 5)
```

```r
topicNames <- apply(top5termsPerTopic, 2, paste, collapse=" ")

topicToViz <- 20 # change for topics of interest
# topicToViz <- grep("feel", topicNames)[1] # Or select a topic by a term
contained in its name
# selecting to 80 most probable terms from the topic by sorting the term-
topic-probability vector in decreasing order
top40terms <- sort(tmResult$terms[topicToViz,], decreasing=TRUE)[1:80]
words <- names(top40terms)
# extract the probabilites of each of the 40 terms
probabilities <- sort(tmResult$terms[topicToViz,], decreasing=TRUE)[1:80]
# visualize the terms as wordcloud
is.na(words)
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

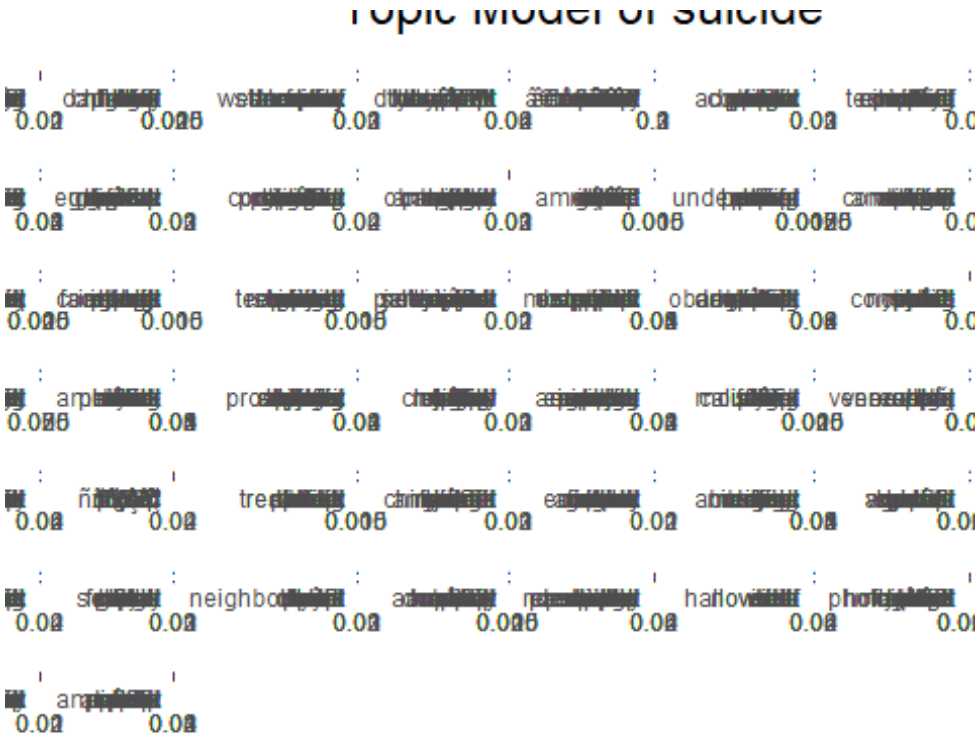
```r
mycolors <- brewer.pal(8, "Dark2")
wordcloud(words, probabilities, random.order = FALSE, color = mycolors)
```

```
## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): colombia could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): expect could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): histori could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): london could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): phelp could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): swimmer could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): spread could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): global could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): emerg could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): control could not be fit on page. It will not be plotted.

## Warning in wordcloud(words, probabilities, random.order = FALSE, color =
## mycolors): antidop could not be fit on page. It will not be plotted.
```

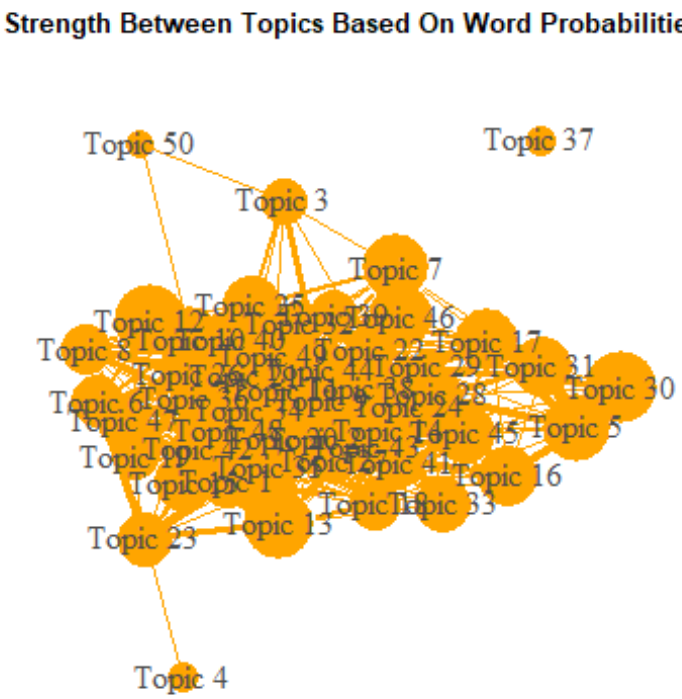

```
# re-rank top topic terms for topic names
beta <- tmResult$terms
topicNames <- apply(lda::top.topic.words(beta, 10, by.score = T),
                    2, paste, collapse = " ")



# Plotting the top terms
top_terms %>% mutate(term = reorder(term, beta)) %>%
  mutate(topic = paste("Topic #", topic)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = F) +
  facet_wrap(~ topic, scales = "free") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 18)) +
```

```
  labs(title = "Topic Model of suicide",
       caption = "Top Terms by Topic") +
  ylab("") +
  xlab("") +
  coord_flip()
```

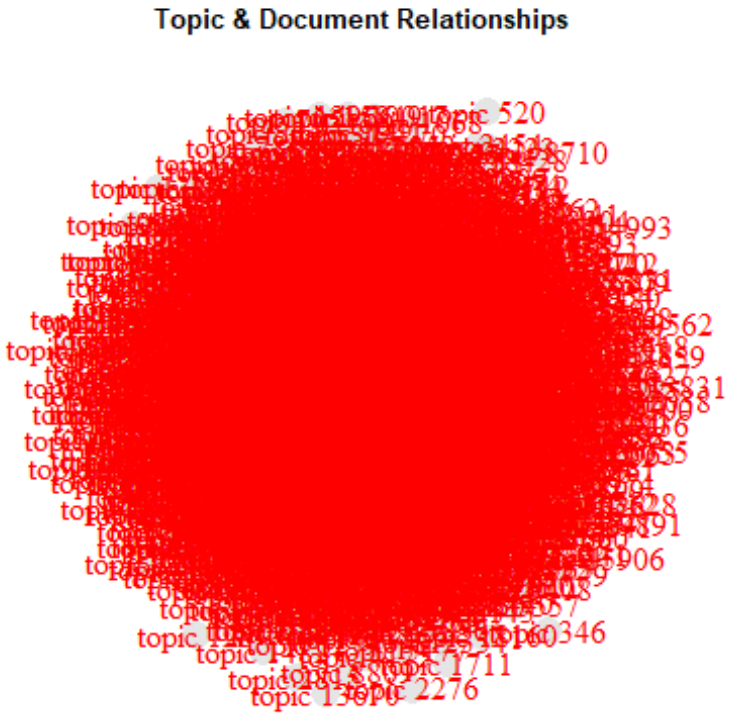


```
# Heatmap of the topics terms
# heatmap(TopicsT)

# Network of the Word distribution over TOpics
post <- posterior(topicModel)

cor_mat <- cor(t(post[["terms"]]))
cor_mat[ cor_mat < .05 ] <- 0
diag(cor_mat) <- 0

graph <- graph.adjacency(cor_mat, weighted=TRUE, mode="lower")
graph <- delete.edges(graph, E(graph)[ weight < 0.05])

E(graph)$edge.width <- E(graph)$weight*20
V(graph)$label <- paste("Topic", V(graph))
V(graph)$size <- colSums(post[["topics"]]) * 0.05

par(mar=c(0, 0, 3, 0))
set.seed(110)
plot.igraph(graph, edge.width = E(graph)$edge.width,
```

```
            edge.color = "orange", vertex.color = "orange",
            vertex.frame.color = NA, vertex.label.color = "grey30")
title("Strength Between Topics Based On Word Probabilities", cex.main=.8)
```




```
# Network of topics over docuents
minval = .1
topic_mat <- posterior(topicModel)[["topics"]]
graph <- graph_from_incidence_matrix(topic_mat, weighted = T)
graph <- delete.edges(graph, E(graph)[weight < minval])

E(graph)$edge.width <- E(graph)$weight*0.2
E(graph)$color <- "blue"
V(graph)$color <- ifelse(grepl("^\\d+$", V(graph)$name), "grey75", "orange")
V(graph)$frame.color <- NA
V(graph)$label <- ifelse(grepl("^\\d+$", V(graph)$name),
                         paste("topic", V(graph)$name),
                         gsub("_", "\n", V(graph)$name))
V(graph)$size <- c(rep(10, nrow(topic_mat)), colSums(topic_mat) * 0.01)
V(graph)$label.color <- ifelse(grepl("^\\d+$", V(graph)$name), "red",
"grey30")

par(mar=c(0, 0, 3, 0))
set.seed(365)
plot.igraph(graph, edge.width = E(graph)$edge.width,
            vertex.color = adjustcolor(V(graph)$color, alpha.f = .4))
title("Topic & Document Relationships", cex.main=.8)
```




```
# Ldavis interactive plot
# topicModel %>%
#   topicmodels2LDAvis() %>%
#  LDAvis::serVis()
# save.image("C:/Users/Um Ar/R Projects/UN MINOR/.RData")
```