## **Homework 2: News Analytics**

## Step 1. Fit a topic model to the existing news data archive

```
#import archive data
news = read.csv("~/Desktop/NewsArticles.csv")
# clean and reprocess the text
library("tm")
## Loading required package: NLP
corp.original = VCorpus(VectorSource(news$content))
corp = tm_map(corp.original, removePunctuation)
corp = tm map(corp, removeNumbers)
corp = tm_map(corp, content_transformer(removeWords), stopwords("SMART"),lazy
=TRUE)
corp = tm map(corp, content transformer(tolower),lazy=TRUE)
corp = tm map(corp, content transformer(stemDocument),lazy=TRUE)
corp = tm_map(corp, stripWhitespace)
# generate a document-term matrix
dtm = DocumentTermMatrix(corp)
dtms = removeSparseTerms(dtm, .995)
dtm matrix = as.matrix(dtms)
# train a topic model
library(topicmodels)
terms = rowSums(dtm_matrix) != 0 #set parameters
dtm matrix = dtm matrix[terms,]
lda_result = LDA(dtm_matrix, 10, method="Gibbs") #fit LDA model
LDA = terms(lda_result, 20)
# assign names to categories
colnames(LDA) = c("Real Estate", "Careers", "Finance",
                  "Government", "Economy", "Healtcare", "Politics",
                  "Earnings", "Energy", "Social Media")
print(LDA)
##
         Real Estate Careers
                                 Finance
                                            Government Economy
                                                                  Healtcare
                     "compani"
                                 "oil"
                                             "fund"
                                                        "percent" "secur"
    [1,] "market"
##
   [2,] "fed"
                     "apple"
                                 "citi"
                                             "financi"
                                                        "price"
                                                                  "case"
                                 "car"
                                             "bank"
                                                        "month"
   [3,] "rate"
                     "technolog"
                                                                  "report"
                                                        "year"
## [4,] "stock"
                     "servic"
                                 "state"
                                             "money"
                                                                  "rule"
## [5,] "year"
                     "read"
                                  "energi"
                                             "invest"
                                                        "home"
                                                                  "court"
## [6,] "read"
                     "time"
                                 "year"
                                             "tax"
                                                        "increas" "file"
```

```
[7,] "economi"
                      "googl"
                                   "china"
                                              "manag"
                                                          "rate"
                                                                    "inform"
    [8,] "interest"
                      "make"
                                   "vehicl"
                                              "vear"
                                                          "report"
                                                                     "public"
##
## [9,] "bank"
                                              "pay"
                                                          "sale"
                                                                    "law"
                      "user"
                                   "read"
## [10,] "investor"
                                                          "growth"
                      "twitter"
                                   "product"
                                              "loan"
                                                                     "govern"
## [11,] "bond"
                                                                    "regul"
                      "media"
                                   "industri" "credit"
                                                          "sinc"
## [12,] "move"
                                   "gas"
                                                                     "feder"
                      "facebook"
                                              "make"
                                                          "averag"
## [13,] "time"
                      "mobil"
                                   "accord"
                                              "debt"
                                                          "week"
                                                                    "issu"
## [14,] "econom"
                      "busi"
                                   "countri"
                                              "investor"
                                                          "expect"
                                                                    "statement"
## [15,] "polici"
                      "product"
                                   "area"
                                              "retir"
                                                          "rise"
                                                                    "read"
## [16,] "expect"
                      "launch"
                                   "unite"
                                              "firm"
                                                          "read"
                                                                    "investig"
                                                          "job"
## [17,] "trade"
                      "devic"
                                                                     "comment"
                                   "water"
                                              "save"
## [18,] "reserv"
                      "network"
                                   "power"
                                              "incom"
                                                          "consum"
                                                                    "depart"
## [19,] "global"
                      "app"
                                              "return"
                                                          "spend"
                                                                     "york"
                                   "nation"
## [20,] "central"
                      "custom"
                                   "airlin"
                                              "percent"
                                                          "market"
                                                                    "ani"
##
         Politics
                       Earnings
                                   Energy
                                             Social Media
                       "peop1"
                                   "compani"
                                             "retail"
##
    [1,] "presid"
##
    [2,] "obama"
                       "plan"
                                   "billion" "store"
##
    [3,] "state"
                       "health"
                                   "share"
                                             "year"
    [4,] "hous"
                       "work"
                                   "million" "million"
##
##
    [5,] "republican"
                       "cost"
                                   "vear"
                                             "sale"
    [6,] "govern"
                       "insur"
                                   "percent" "day"
##
    [7,] "tax"
                                             "food"
##
                       "worker"
                                   "earn"
    [8,] "support"
                       "year"
                                   "revenu"
                                             "read"
##
                       "job"
                                   "analyst" "brand"
    [9,] "democrat"
                       "care"
## [10,] "cut"
                                   "quarter"
                                             "game"
## [11,] "bill"
                       "state"
                                   "report"
                                             "time"
## [12,] "senat"
                       "studi"
                                   "expect"
                                             "make"
## [13,] "congress"
                       "pay"
                                   "stock"
                                             "busi"
                       "mani"
                                   "busi"
## [14,] "trump"
                                             "custom"
## [15,] "vote"
                       "read"
                                   "deal"
                                             "open"
## [16,] "white"
                       "famili"
                                   "trade"
                                             "consum"
## [17,] "budget"
                       "student"
                                   "profit"
                                             "shop"
## [18,] "polit"
                                   "ceo"
                       "benefit"
                                             "team"
## [19,] "american"
                       "colleg"
                                   "sale"
                                             "includ"
## [20,] "nation"
                       "american" "street"
                                             "onlin"
```

## **Step 2. Retrieve new articles**

```
# extract article URLs and text
library(rvest)
## Loading required package: xml2
library(stringr)
```

```
article.urls=NULL
# loop for scraping all article URLs from the first 10 pages
for (i in (1:10)) {
 urls = read_html(paste("http://www.cnbc.com/us-news/?page=", i, sep="")) %>
%
    html_nodes(".bigHeader .headline a") %>%
    html_attr("href")
 article.urls = cbind(article.urls, urls)
}
article.urls = article.urls[-(24:26)] # remove non-text articles/videos
head(article.urls, 40)
## [1] "/2017/04/26/macau-market-improving-but-las-vegas-sands-misses-on-fir
st-quarter-earnings.html"
## [2] "/2017/04/26/grail-adds-kleiner-perkins-co-founder-brook-byers-to-boa
rd.html"
## [3] "/2017/04/26/big-tech-thursday-could-be-the-catalyst-for-new-highs.ht
ml"
## [4] "/2017/04/26/north-korea-is-a-direct-threat-to-our-national-security-
panetta-says.html"
## [5] "/2017/04/26/shkreli-seeks-document-hand-over-to-bar-evidence-from-us
e-at-trial.html"
## [6] "/2017/04/26/bets-on-a-stock-market-rally-using-options-just-hit-reco
rd-levels.html"
## [7] "/2017/04/26/trump-aims-to-pressure-north-korea-through-tighter-sanct
ions-and-diplomacy.html"
## [8] "/2017/04/26/ron-howards-new-work-of-genius-einstein-with-a-more-huma
n-twist.html"
## [9] "/2017/04/26/paypal-earnings-1q-2017.html"
## [10] "/2017/04/26/despite-market-highs-its-not-too-late-to-get-into-the-ga
me-experts-say.html"
## [11] "/2017/04/26/the-ugly-truth-in-twitters-earnings-is-that-its-shrinkin
g-into-irrelevance.html"
## [12] "/2017/04/26/what-happened-the-last-time-companies-got-a-break-on-ove
rseas-profits.html"
## [13] "/2017/04/26/retail-stocks-move-higher-after-mnuchin-shoots-down-cont
roversial-tax-proposal.html"
## [14] "/2017/04/26/these-10-retailers-could-be-next-on-activist-investors-h
it-list.html"
## [15] "/2017/04/26/fcc-chairman-to-propose-reversing-net-neutrality-rules.h
## [16] "/2017/04/26/mnuchin-says-trump-has-no-intention-of-releasing-tax-ret
## [17] "/2017/04/26/nintendo-has-made-a-huge-bet-on-the-switch-console-and-t
omorrow-well-see-how-well-its-selling.html"
## [18] "/2017/04/26/heres-what-was-missing-from-trumps-tax-plan-that-wall-st
reet-really-wanted.html"
## [19] "/2017/04/26/kevin-oleary-drops-out-of-race-to-lead-canadas-conservat
```

```
ive-party.html"
## [20] "/2017/04/26/did-chuxing-raising-5-billion-at-50-billion-valuation.ht
m1"
## [21] "/2017/04/26/you-can-now-have-your-own-designer-pair-of-mud-covered-j
eans.html"
## [22] "/2017/04/25/goldman-has-figured-out-the-trick-for-making-money-off-a
mazon.html"
## [23] "/2017/04/25/buffalo-wild-wings-and-its-activist-investors-are-fighti
ng-over-this-stock-chart.html"
## [24] "/2017/04/26/macau-market-improving-but-las-vegas-sands-misses-on-fir
st-quarter-earnings.html"
## [25] "/2017/04/26/grail-adds-kleiner-perkins-co-founder-brook-byers-to-boa
rd.html"
## [26] "/2017/04/26/big-tech-thursday-could-be-the-catalyst-for-new-highs.ht
ml"
## [27] "/2017/04/26/north-korea-is-a-direct-threat-to-our-national-security-
panetta-says.html"
## [28] "/2017/04/26/former-yahoo-interim-ceo-explains-how-twitter-can-avoid-
his-old-companys-fate.html"
## [29] "/2017/04/26/ann-coulter-cancels-her-speech-at-uc-berkeley.html"
## [30] "/2017/04/26/us-admiral-at-hearing-takes-the-blame-for-the-vinson-loc
ation-flap.html"
## [31] "/2017/04/26/ceo-of-struggling-disk-drive-maker-says-business-is-stab
le-as-shares-plunge-16.html"
## [32] "/2017/04/26/house-freedom-caucus-announces-it-will-support-amended-o
bamacare-replacement-bill.html"
## [33] "/2017/04/26/commerce-secretary-ross-to-announce-plan-to-probe-import
ed-aluminum-source.html"
## [34] "/2017/04/26/fiat-chrysler-ceo-says-jeep-ram-are-strong-enough-brands
-to-follow-in-ferraris-footsteps.html"
## [35] "/2017/04/26/once-secretive-hedge-fund-legend-ray-dalio-opens-up-join
s-twitter.html"
## [36] "/2017/04/26/five-ways-the-trump-economy-is-off-to-a-record-setting-p
ace.html"
## [37] "/2017/04/26/watch-treasurys-mnuchin-and-white-houses-cohn-talk-trump
-tax-plan.html"
## [38] "/2017/04/26/jpmorgan-says-the-economy-barely-grew-in-the-first-quart
er.html"
## [39] "/2017/04/26/the-key-to-finding-alien-life-will-be-locating-new-quack
ers.html"
## [40] "/2017/04/26/steve-wynn-were-seeing-a-resurgence-in-china.html"
article.text=NULL
# loop for collecting and cleaning the textual data from each article
for (i in 1:length(article.urls)) {
  text = read_html(paste("http://www.cnbc.com", article.urls[i], sep="")) %>%
    html_nodes("p") %>%
    html text()
```

```
text = text[-1] # remove first element in vector
  text = gsub("^\\s+|\\s+$", "", text, fixed = FALSE) # strip leading and tra
iling whitespace
  text = gsub('"', '', text, fixed = FALSE) # replace all quotations
  text = paste(text, collapse=" ") # compress text into one block
  article.text = cbind(article.text, text)
}
length(article.text) # total number of articles scraped
### [1] 257
# complete text of first 4 articles
article.text[1:4]
```

## [1] "• Macau growth accelerating for third consecutive quarter • Las V egas operations turn in best showing since 2008 • Management optimistic abo ut eventually getting nod to enter Japan Las Vegas Sands Corp. late Wednesday reported first-quarter earnings that were slightly below expectations althoug h it indicated Las Vegas operations were strong and the Macau market was impr oving. Also, management was optimistic on the earnings call about the chances of getting approval to eventually open a casino resort in the Japanese market . On the earnings miss, the stock dipped just over 1 percent in after-hours t rading but started to recover after the upbeat earnings call. After a challen ging period the Macau market is growing again — and its growth rate has been accelerating for three consecutive quarters, said Chairman and CEO Sheldon Ad elson on the company's earnings call. At the same time, the Las Vegas operati ons delivered its best quarter since 2008, Adelson said. The company's Veneti an Las Vegas and Palazzo properties, including the Sands Expo and Convention Center, posted better results. On a GAAP basis, earnings per share grew by 50 percent in the first quarter. Even so, this marked the second consecutive qua rter the company's EPS missed analyst expectations, according to Thomson Reut ers. Net revenue in the quarter was up 14 percent. Here's how the company did in the quarter compared with analysts' consensus estimates: • EPS: 60 cents a share vs. 61 cents a share expected by analysts polled by Thomson Reuters. • Net revenue: \$3.11 billion, or generally inline with expectations. Overall, adjusted property EBITDA, or earnings before interest, taxes, depreciation an d amortization, increased about 25 percent in the quarter compared with a yea r ago and were above Street consensus. By segment, the Macau operations grew adjusted property EBITDA by 20 percent on a year-on-year basis in the quarter . It was driven by 17 percent growth in mass gaming revenues. The Las Vegas o perating properties showed adjusted property EBITDA up by 40 percent compared with the year-ago period. It said contributing to the strong results were the Venetian Las Vegas and Palazzo properties, as well as record convention and g roup meeting business. EBITDA is considered an important metric of operating profit performance for casino companies because it essentially neutralizes th e impact of the company's capital structure and deducts interest expense, tax es and depreciation charges. Meantime, Adelson was asked on the earnings call about the U.S.-based casino giant's growth plans in the Asian market. Besides the company's Macau operations, it also has a presence already in the Singapo

re market. Japan is what everybody is talking about, Adelson said. We have be en informed by people in the know in Japan that LVS (or Las Vegas Sands) is by far, not just marginally, but by far ahead of the other competition to get the integrated resort approval. Last year, Japan took steps to legalize Las V egas-style casino resort gaming, which could allow for casino and hotel space as well as convention facilities and retail space. The Japanese government ha sn't formally selected the companies for casino development but Las Vegas San ds and its competitor MGM Resorts have openly indicated they are indeed inter ested. The casino mogul said there's a possibility the Japanese government co uld allow the company more than one so-called integrated resort property. To be clear, nothing is finalized and he questioned whether the Japanese would e ven allow a foreign entity to have two casino locations. As for other Asian m arkets, Adelson said the company is looking at Vietnam, Thailand and South Ko rea. He added that the South American market was attractive too."

## [2] "Grail, a start-up that's raising over \$1 billion to work on early can cer detection, has just brought on veteran venture capitalist Brook Byers to its board of directors, according to an SEC filing. The filing on Wednesday a lso revealed that Grail's total amount raised in its current round has increa sed to \$973 million from \$914 million in mid-March. Byers, a co-founder of Si licon Valley venture firm Kleiner Perkins Caufield & Byers, previously invest ed in companies including GHDX , Foundation Medicine, Pacific Biosciences and Electronic Arts. He is no longer an active partner at the firm. Grail's aim i s to develop so-called liquid biopsies for spotting cancer in the earliest st ages by detecting circulating fragments of tumor DNA in the blood. The compan y, led by former Google executive Jeff Huber, was originally an offshoot of I llumina, which develops supercomputers for genomic sequencing. Huber, who led the teams in charge of Google Maps and its ad software, joined Grail after lo sing his wife to colon cancer. Grail now sits just shy of its stated goal to raise over \$1 billion in the round. The company is using the capital for rese arch and development, as well as for large-scale clinical trials to prove tha t its technology is both sensitive and accurate. Ken Drazan, Grail's chief bu siness officer, confirmed that Byers joined the board on his own -- not on be half of Kleiner Perkins -- and said the company has not yet closed its Series B funding round. Byers didn't immediately respond to a request for comment. D razan said the additional came from new investors, such as family offices, pr ofessional investors and crossover investors. Byers' son Blake Byers, who wor ks for Alphabet's venture arm GV, is backing a rival start-up. GV recently pa rticipated Freenome's \$65 million funding round that was led by Andreessen H orowitz through its new bio fund."

## [3] "Thursday's blast of high-profile earnings, including big tech bellwet hers, could help set the stage for new stock market highs in the very near fu ture. There's a rush of earnings Thursday morning, including Union Pacific, B ristol-Myers Squibb, Ford, American Airlines, Comcast, Dow Chemical, Raytheon and Deutsche Bank, to name a few. But some of the tech names that have led th

e market's gains report after the closing bell and could be big catalysts int o Friday's trading. That includes Amazon.com, up 21 percent year to date, and Alphabet, which has risen 12 percent. Microsoft, Intel and Starbucks also rep ort after the close. There are risks to that scenario, including the fact tha t Congress still needs to pass a resolution to keep the government funded by Friday, though it is expected to happen. There are also wild cards, like the worrisome tensions with North Korea. Another factor that markets are warily w atching is the string of softening data reports. Thursday data includes March durable goods orders, weekly jobless claims and advanced economic indicators, all at 8:30 a.m. ET. Economists have been cutting forecasts for first-quarter GDP, expected Friday, but are hopeful the second quarter will bounce back. JP Morgan lowered its forecast to 0.4 percent for the first quarter on Wednesday . Stocks slid into the close Wednesday, in a sell-the-news move after Preside nt Donald Trump's much anticipated tax plan. The proposal contained a juicy c orporate tax cut but was viewed as short on details and likely to be seriousl y modified. Nonetheless, it is an opening round in what is expected to be mon ths of horse trading on taxes. People are somewhat optimistic, but they wante d more detail, said Doron Barness, head of global equities trading at Oppenhe imer. He said what the market is really focusing on are the nearby highs, lik e 2,400 on the S&P 500. That index closed Wednesday at 2,387, down 1 point. T he Nasdaq was just under its high, set Tuesday, closing at 6,025, and the Dow was off 21 at 20,975. The Russell 2000 closed at a new all-time high of 1,419 . We've had some soft economic data, but people are ignoring that. You have N orth Korea stuff going on. People are ignoring that. Those are the kind of th ings that could bring the market down if more economic data comes out that's weaker and there's something geopolitical. But absent that, I don't see how t he market doesn't go higher. People are not going to sell this market, said B arness. As of Thursday, half the S&P 500 will have reported earnings, and so far profits are expected to be up about 11.8 percent from last year's first q uarter, according to Thomson Reuters. The companies that have reported so far , saw profits rise an average of 14.4 percent and beat expectations 77 percen t of the time. Obviously, it's all been about earnings, and obviously large-c ap earnings have been coming in quite strong, and that's giving people a lot of optimism. You have the VIX back under 11 so it's risk-on, said Steve DeSan ctis, Jefferies equity strategist. The VIX is the CBOE's Volatility Index, a measure of market fear since it is calculated based on the puts and calls on the S&P 500. It was at the low level of 10.85 on Wednesday, up 0.8 percent. B arness said if the S&P does get close to 2,400, it could hover there for a wh ile. What usually happens is we'll toy with this level, back and forth, he sa id. If we go through, it's going to have velocity. He said that would force s hort covering, meaning traders who have short positions would have to buy as the prices go up. There's a lot of euphoria going around. I hate to be the sk unk at the garden party, but I'm just a bit more skeptical here. We've had a pretty big move in three days, said DeSanctis. Bond yields, which move invers ely to price, also slid Wednesday amid concerns that the tax plan could widen the federal budget deficit, which would ultimately make yields rise. In late trading, the 10-year yield drifted to 2.30 percent, an important technical ar ea that the market had held above for months — until last week. Who is going

to be right? The bond market is telling you there's sluggish growth ahead, an d the stock market is not telling you that, said DeSanctis. UBS economists on Wednesday said they were sticking with their view that first-quarter growth i s tracking at 0.9 percent, after government retail sales revisions showed a w eaker-than-expected consumer in the first quarter. But they did note that Thu rsday's data on retail and wholesale inventories and the advanced merchandise trade, as well as durables could influence their forecast for GDP ahead of Fr iday's report. Disclosure: Comcast is the owner of NBCUniversal, the parent c ompany of CNBC and CNBC.com."

## [4] "North Korea is a rogue nation with an unstable leader and the ability to develop missiles and miniaturized nuclear weapons, former Defense Secretar y Leon Panetta told CNBC on Wednesday. They really are a direct threat to our national security. But it's also a complicated issue. That's the reason most recent presidents haven't pulled the trigger, he said in an interview with Cl osing Bell. It's going to take a lot of continuing work because of the conseq uences of direct military action. On Wednesday afternoon, all 100 senators me t with Secretary of State Rex Tillerson, Secretary of Defense Jim Mattis, Dir ector of National Intelligence Dan Coats and Marine Corps. Gen. Joseph Dunfor d, chairman of the Joint Chiefs of Staff, for a classified briefing on the m atter. After the meeting, the top administration officials said President Don ald Trump wants to pressure North Korea through tighter sanctions and diploma tic channels, but the United States is prepared to defend itself and allies. Panetta believes the best course of action is for the U.S. to increase its pr esence in the area, support allies and continue to put pressure on China to t ry to get North Korea to the negotiating table. He called national security T rump's greatest strength. He's got a great national security team. He's been willing to listen to them. And he's willing to change positions if necessary on China, on Syria, on NATO and some other areas, he said. Meanwhile, Trump's weak point is dealing with Congress and getting any major legislation passed, Panetta said. - CNBC's Jacob Pramuk contributed to this report. Playing Shar e this video... Watch Next..."

## Step 3. Classify news articles using your topic model

```
# create a new document term matrix
news.corp = VCorpus(VectorSource(article.text))

# specify dictionary when creating the dtm for the new articles,
# which will limit the dtm it creates to only the words that also appeared in
the archive
dic = Terms(dtms)
new_dtm = DocumentTermMatrix(news.corp, control=list(dictionary = dic))
new_dtm = new_dtm[rowSums(as.matrix(new_dtm))!=0,]
topic_probabilities = posterior(lda_result, new_dtm)
```

```
# rename columns as categories
colnames(topic probabilities$topics) = colnames(LDA)
head(topic probabilities$topics, 10)
##
      Real Estate
                    Careers
                               Finance Government
                                                     Economy Healtcare
      0.10101010 0.08080808 0.10101010 0.06060606 0.14141414 0.07070707
## 1
## 2
      0.06896552 0.16091954 0.08045977 0.11494253 0.05747126 0.13793103
## 3
      0.26760563 0.09154930 0.10563380 0.04225352 0.11267606 0.03521127
## 4
      0.08602151 0.09677419 0.09677419 0.05376344 0.06451613 0.06451613
## 5
      0.09395973 0.09395973 0.06711409 0.06040268 0.03355705 0.25503356
## 6
      0.13684211 0.06315789 0.09473684 0.11578947 0.09473684 0.06315789
      0.08571429 0.07619048 0.15238095 0.06666667 0.05714286 0.07619048
## 7
## 8
      0.11956522 0.11956522 0.09782609 0.09782609 0.05434783 0.06521739
## 9
      0.10144928 0.10144928 0.08695652 0.07246377 0.11594203 0.07246377
## 10 0.09523810 0.07142857 0.08333333 0.09523810 0.10714286 0.05952381
        Politics
                  Earnings
                               Energy Social Media
##
## 1 0.08080808 0.05050505 0.18181818
                                        0.13131313
## 2 0.09195402 0.12643678 0.10344828
                                        0.05747126
## 3 0.08450704 0.03521127 0.13380282
                                        0.09154930
## 4 0.20430108 0.11827957 0.12903226
                                        0.08602151
## 5 0.14093960 0.10067114 0.06711409
                                        0.08724832
## 6 0.13684211 0.06315789 0.09473684
                                        0.13684211
## 7 0.21904762 0.07619048 0.09523810
                                        0.09523810
## 8 0.11956522 0.08695652 0.05434783
                                        0.18478261
## 9 0.07246377 0.07246377 0.23188406
                                        0.07246377
## 10 0.17857143 0.07142857 0.17857143
                                        0.05952381
# generate a vector that assigns to each document
# the topic for which it has the highest probability of appearing
categories=NULL
for (i in 1:length(article.urls)) {
 max = which.max(topic_probabilities$topics[i,])
 cat = colnames(topic_probabilities$topics)[max]
 categories = rbind(categories, cat)
}
colnames(categories) = "Category" # rename column
rownames(categories) = c(1:length(article.urls)) # number documents/articles
head(categories, 40) # categories for the first 40 articles
##
      Category
## 1
     "Energy"
## 2 "Careers"
## 3
     "Real Estate"
## 4 "Politics"
## 5
     "Healtcare"
## 6 "Real Estate"
## 7 "Politics"
```

```
## 8 "Social Media"
## 9 "Energy"
## 10 "Politics"
## 11 "Real Estate"
## 12 "Real Estate"
## 13 "Politics"
## 14 "Real Estate"
## 15 "Politics"
## 16 "Healtcare"
## 17 "Real Estate"
## 18 "Real Estate"
## 19 "Politics"
## 20 "Energy"
## 21 "Social Media"
## 22 "Energy"
## 23 "Healtcare"
## 24 "Real Estate"
## 25 "Careers"
## 26 "Real Estate"
## 27 "Politics"
## 28 "Careers"
## 29 "Real Estate"
## 30 "Finance"
## 31 "Energy"
## 32 "Politics"
## 33 "Finance"
## 34 "Energy"
## 35 "Careers"
## 36 "Real Estate"
## 37 "Politics"
## 38 "Economy"
## 39 "Careers"
## 40 "Energy"
# in a table, print the contents of any ten news articles
# and the corresponding categories
news.content=NULL
news.category=NULL
random_news = sample(1:length(article.urls), 10, replace=F) # random article
generator
for (i in random_news) {
  # get contents of selected articles
  news.content = rbind(news.content, article.text[i])
  # get categories of selected articles
  news.category = rbind(news.category, categories[i])
}
# combine contents and categories into one table
news.table = cbind(news.category, news.content)
```

```
news.table = as.data.frame(news.table)
colnames(news.table) = c("Category", "Content")
View(news.table)
```

Randomly generated articles and their respective categories below:

	Category <sup>‡</sup>	Content
1	Finance	Goldman Sachs gave its clients the best way to trade A
2	Politics	A senior administration official told CNBC that Preside
3	Energy	The U.S. military has started moving parts of the contr
4	Politics	North Korea is a rogue nation with an unstable leader
5	Politics	North Korea is a rogue nation with an unstable leader
6	Social Media	Alphabet's effort to catch Amazon's cloud is taking ce
7	Energy	· Economic pressure from Chinese won't be enough to
8	Careers	Grail, a start-up that's raising over \$1 billion to work o
9	Finance	The battle between Buffalo Wild Wings and activist inve
10	Politics	Former acting attorney general Sally Yates is schedule

	Category <sup>‡</sup>	Content
1	Politics	North Korea is a rogue nation with an unstable leader
2	Politics	U.S. lawmakers want to leave briefings on North Korea
3	Economy	Should President Donald Trump choose to replace Fed
4	Social Media	If you're a guy, Nordstrom now wants to help you look
5	Politics	U.S. President Donald Trump will meet with Australian
6	Earnings	Thursday's blast of high-profile earnings, including big
7	Economy	Relief over the French election sparked a global rally in
8	Government	When a cadre of manufacturing executives came to the
9	Careers	Funding to keep the federal government up and runnin
10	Energy	The threats seem to come almost daily now out of Nor

	Category <sup>‡</sup>	Content
1	Careers	AT&T and T-Mobile may be declaring victory among th
2	Real Estate	President Trump appears to have backed off of his pla
3	Earnings	· Macau growth accelerating for third consecutive qua
4	Finance	· Macau growth accelerating for third consecutive qua
5	Finance	Word that Yahoo CEO Marissa Mayer could pocket up t
6	Careers	When Franklin D. Roosevelt was inaugurated in 1933,
7	Politics	President Donald Trump has reason to brag about the
8	Politics	President Donald Trump again pushed for changes at t
9	Government	Grail, a start-up that's raising over \$1 billion to work o
10	Careers	If you're a guy, Nordstrom now wants to help you look