Spam Detection

Helee Rana

7/26/2020

## Example 2: Spam Detection by Text mining

require(tidyverse, quietly = TRUE)

## ── Attaching packages ──────────────────────────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.2.1 ✓ purrr 0.3.3  
## ✓ tibble 2.1.3 ✓ dplyr 0.8.4  
## ✓ tidyr 1.0.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.4.0

## ── Conflicts ─────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

sms\_data=read\_csv("/Users/heleerana/Desktop/DATA MINING Prof. YEGIN/sms\_data.csv")

## Parsed with column specification:  
## cols(  
## type = col\_character(),  
## text = col\_character()  
## )

sms\_data

## # A tibble: 5,559 x 2  
## type text   
## <chr> <chr>   
## 1 ham Hope you are having a good week. Just checking in   
## 2 ham K..give back my thanks.   
## 3 ham Am also doing in cbe only. But have to pay.   
## 4 spam complimentary 4 STAR Ibiza Holiday or £10,000 cash needs your URGENT c…  
## 5 spam okmail: Dear Dave this is your final notice to collect your 4\* Tenerif…  
## 6 ham Aiya we discuss later lar... Pick u up at 4 is it?   
## 7 ham Are you this much buzy   
## 8 ham Please ask mummy to call father   
## 9 spam Marvel Mobile Play the official Ultimate Spider-man game (£4.50) on ur…  
## 10 ham fyi I'm at usf now, swing by the room whenever   
## # … with 5,549 more rows

##install.packages("quanteda")  
##install.packages("quanteda.textmodels")

require(quanteda)

## Loading required package: quanteda

## Package version: 2.0.1

## 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

require(quanteda.textmodels)

## Loading required package: quanteda.textmodels

my\_corpus=corpus(sms\_data$text)  
summary(my\_corpus)

## Corpus consisting of 5559 documents, showing 100 documents:  
##   
## Text Types Tokens Sentences  
## text1 11 11 2  
## text2 6 8 1  
## text3 11 12 2  
## text4 25 25 3  
## text5 30 31 3  
## text6 14 16 2  
## text7 5 5 1  
## text8 6 6 1  
## text9 32 34 2  
## text10 11 11 1  
## text11 20 20 1  
## text12 4 6 1  
## text13 31 35 5  
## text14 4 5 1  
## text15 13 13 1  
## text16 8 10 1  
## text17 27 33 3  
## text18 15 17 2  
## text19 15 25 2  
## text20 25 26 1  
## text21 10 11 2  
## text22 13 13 1  
## text23 12 13 2  
## text24 41 47 3  
## text25 11 12 2  
## text26 9 10 1  
## text27 10 10 1  
## text28 8 8 1  
## text29 7 7 1  
## text30 3 5 1  
## text31 8 11 1  
## text32 30 49 4  
## text33 5 5 1  
## text34 18 20 4  
## text35 34 39 4  
## text36 17 21 3  
## text37 21 23 1  
## text38 12 12 1  
## text39 30 36 4  
## text40 31 39 4  
## text41 17 17 2  
## text42 9 9 2  
## text43 7 7 1  
## text44 26 28 5  
## text45 8 8 1  
## text46 6 6 1  
## text47 6 7 1  
## text48 8 8 1  
## text49 16 17 2  
## text50 15 16 2  
## text51 9 9 1  
## text52 15 16 2  
## text53 27 31 1  
## text54 27 27 1  
## text55 13 13 1  
## text56 23 23 1  
## text57 6 6 1  
## text58 5 7 2  
## text59 6 6 2  
## text60 18 20 2  
## text61 24 24 2  
## text62 17 18 2  
## text63 39 46 2  
## text64 67 110 10  
## text65 22 23 1  
## text66 52 70 4  
## text67 21 26 1  
## text68 20 24 2  
## text69 23 33 4  
## text70 34 41 6  
## text71 8 8 1  
## text72 15 15 2  
## text73 11 12 1  
## text74 12 20 3  
## text75 7 7 1  
## text76 6 6 1  
## text77 10 11 2  
## text78 25 28 4  
## text79 8 8 1  
## text80 5 5 1  
## text81 25 30 4  
## text82 26 28 5  
## text83 33 39 3  
## text84 11 11 2  
## text85 8 8 1  
## text86 6 6 1  
## text87 5 7 1  
## text88 15 16 1  
## text89 24 40 4  
## text90 12 13 2  
## text91 31 38 6  
## text92 4 6 1  
## text93 20 21 2  
## text94 8 8 1  
## text95 13 16 3  
## text96 8 8 2  
## text97 8 8 1  
## text98 8 8 1  
## text99 5 5 1  
## text100 28 33 4

head(my\_corpus)

## Corpus consisting of 6 documents.  
## text1 :  
## "Hope you are having a good week. Just checking in"  
##   
## text2 :  
## "K..give back my thanks."  
##   
## text3 :  
## "Am also doing in cbe only. But have to pay."  
##   
## text4 :  
## "complimentary 4 STAR Ibiza Holiday or £10,000 cash needs you..."  
##   
## text5 :  
## "okmail: Dear Dave this is your final notice to collect your ..."  
##   
## text6 :  
## "Aiya we discuss later lar... Pick u up at 4 is it?"

my\_dfm=dfm(my\_corpus)  
my\_dfm

## Document-feature matrix of: 5,559 documents, 9,291 features (99.8% sparse).  
## features  
## docs hope you are having a good week . just checking  
## text1 1 1 1 1 1 1 1 1 1 1  
## text2 0 0 0 0 0 0 0 3 0 0  
## text3 0 0 0 0 0 0 0 2 0 0  
## text4 0 0 0 0 0 0 0 1 0 0  
## text5 0 0 0 0 0 0 0 1 0 0  
## text6 0 0 0 0 0 0 0 3 0 0  
## [ reached max\_ndoc ... 5,553 more documents, reached max\_nfeat ... 9,281 more features ]

my\_dfm = dfm\_weight(my\_dfm,scheme = "boolean")  
docvars(my\_dfm,field = "type")=sms\_data$type

my\_dfm$type

## [1] "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "spam" "ham"   
## [11] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [21] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [31] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham"   
## [41] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [51] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [61] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [71] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [81] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [91] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [101] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [111] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [121] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [131] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [141] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [151] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [161] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham"   
## [171] "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [181] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham"   
## [191] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam"  
## [201] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [211] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [221] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [231] "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [241] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [251] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [261] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [271] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [281] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [291] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [301] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [311] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [321] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [331] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [341] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [351] "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "spam" "spam"  
## [361] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam"  
## [371] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [381] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [391] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [401] "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [411] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [421] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [431] "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham"   
## [441] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [451] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [461] "spam" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham"   
## [471] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [481] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [491] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [501] "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [511] "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "spam" "spam"  
## [521] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [531] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [541] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [551] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham"   
## [561] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [571] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [581] "spam" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "spam"  
## [591] "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [601] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [611] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham"   
## [621] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [631] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [641] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [651] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham"   
## [661] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [671] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [681] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [691] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [701] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [711] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [721] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham"   
## [731] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [741] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [751] "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [761] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham"   
## [771] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [781] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam"  
## [791] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [801] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [811] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [821] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [831] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [841] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [851] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [861] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [871] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [881] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [891] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [901] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [911] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [921] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [931] "ham" "ham" "spam" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [941] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [951] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [961] "spam" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [971] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [981] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [991] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1001] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1011] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1021] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham"   
## [1031] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1041] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1051] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [1061] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1071] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1081] "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1091] "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham"   
## [1101] "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham"   
## [1111] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1121] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1131] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1141] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1151] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1161] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1171] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1181] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1191] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam"  
## [1201] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1211] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1221] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1231] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1241] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1251] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1261] "spam" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "spam"  
## [1271] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [1281] "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham"   
## [1291] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1301] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [1311] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1321] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [1331] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham"   
## [1341] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1351] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1361] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [1371] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1381] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1391] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [1401] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1411] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1421] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1431] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1441] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1451] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1461] "spam" "ham" "spam" "ham" "ham" "spam" "spam" "ham" "ham" "ham"   
## [1471] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1481] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [1491] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1501] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1511] "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1521] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1531] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1541] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1551] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1561] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1571] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1581] "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham"   
## [1591] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1601] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1611] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1621] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1631] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1641] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1651] "ham" "spam" "spam" "ham" "ham" "ham" "spam" "ham" "spam" "ham"   
## [1661] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1671] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1681] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1691] "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1701] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1711] "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [1721] "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham"   
## [1731] "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1741] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham"   
## [1751] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1761] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1771] "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1781] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1791] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1801] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1811] "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [1821] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1831] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1841] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam"  
## [1851] "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam"  
## [1861] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [1871] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1881] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1891] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1901] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [1911] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1921] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1931] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1941] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [1951] "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [1961] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [1971] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam"  
## [1981] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham"   
## [1991] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2001] "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2011] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2021] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham"   
## [2031] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "spam"  
## [2041] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2051] "spam" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [2061] "ham" "ham" "ham" "spam" "spam" "spam" "ham" "spam" "ham" "ham"   
## [2071] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [2081] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2091] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [2101] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2111] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [2121] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [2131] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2141] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2151] "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [2161] "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham"   
## [2171] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2181] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2191] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [2201] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2211] "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [2221] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2231] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2241] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2251] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2261] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2271] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham"   
## [2281] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [2291] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [2301] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2311] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [2321] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham"   
## [2331] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2341] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2351] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2361] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [2371] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2381] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2391] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2401] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2411] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2421] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2431] "ham" "spam" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham"   
## [2441] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2451] "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [2461] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2471] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2481] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [2491] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [2501] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2511] "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [2521] "spam" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2531] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2541] "spam" "ham" "spam" "spam" "ham" "spam" "spam" "ham" "ham" "ham"   
## [2551] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2561] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2571] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [2581] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [2591] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2601] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham"   
## [2611] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [2621] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2631] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2641] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2651] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham"   
## [2661] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [2671] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "spam"  
## [2681] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2691] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2701] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [2711] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2721] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2731] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2741] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2751] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham"   
## [2761] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2771] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2781] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham"   
## [2791] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2801] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "spam" "ham"   
## [2811] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [2821] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2831] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [2841] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2851] "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham"   
## [2861] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2871] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2881] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam"  
## [2891] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham"   
## [2901] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam"  
## [2911] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [2921] "ham" "spam" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam"  
## [2931] "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham"   
## [2941] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [2951] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "spam"  
## [2961] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2971] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2981] "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [2991] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3001] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3011] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "spam"  
## [3021] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3031] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [3041] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham"   
## [3051] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3061] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3071] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3081] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3091] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3101] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3111] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam"  
## [3121] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3131] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "spam"  
## [3141] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3151] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3161] "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3171] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3181] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3191] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [3201] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3211] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3221] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3231] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3241] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3251] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3261] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3271] "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham" "ham"   
## [3281] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3291] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3301] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam"  
## [3311] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3321] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3331] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3341] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3351] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [3361] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3371] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham"   
## [3381] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3391] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham"   
## [3401] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3411] "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3421] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham"   
## [3431] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [3441] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3451] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [3461] "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [3471] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3481] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3491] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham"   
## [3501] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [3511] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3521] "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham"   
## [3531] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [3541] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [3551] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [3561] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [3571] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [3581] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3591] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3601] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [3611] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3621] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3631] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "spam"  
## [3641] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3651] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "spam" "ham"   
## [3661] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3671] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [3681] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3691] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3701] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [3711] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [3721] "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3731] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham"   
## [3741] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3751] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3761] "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3771] "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "spam" "ham"   
## [3781] "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3791] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "spam"  
## [3801] "ham" "ham" "spam" "spam" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3811] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3821] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [3831] "spam" "spam" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [3841] "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3851] "spam" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3861] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3871] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [3881] "spam" "ham" "spam" "spam" "ham" "spam" "ham" "spam" "ham" "ham"   
## [3891] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3901] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3911] "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [3921] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3931] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam"  
## [3941] "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [3951] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3961] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [3971] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3981] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [3991] "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4001] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4011] "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "spam"  
## [4021] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "spam"  
## [4031] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4041] "spam" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "spam" "ham"   
## [4051] "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4061] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4071] "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4081] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham"   
## [4091] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4101] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4111] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4121] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4131] "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4141] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [4151] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4161] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4171] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4181] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4191] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4201] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4211] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4221] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4231] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4241] "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam"  
## [4251] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4261] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [4271] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [4281] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4291] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [4301] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4311] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4321] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4331] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4341] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4351] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [4361] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [4371] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4381] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4391] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4401] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4411] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4421] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4431] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4441] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4451] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4461] "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [4471] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [4481] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4491] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [4501] "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4511] "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham"   
## [4521] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4531] "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham"   
## [4541] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham"   
## [4551] "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [4561] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham"   
## [4571] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham"   
## [4581] "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [4591] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4601] "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "spam" "ham"   
## [4611] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [4621] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4631] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4641] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4651] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4661] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam" "spam"  
## [4671] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4681] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4691] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4701] "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham" "ham" "ham"   
## [4711] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4721] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4731] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4741] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4751] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4761] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4771] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam" "ham"   
## [4781] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4791] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4801] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4811] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4821] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4831] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4841] "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [4851] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham"   
## [4861] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4871] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [4881] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4891] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [4901] "spam" "spam" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "spam"  
## [4911] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam"  
## [4921] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4931] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [4941] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [4951] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4961] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [4971] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [4981] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [4991] "ham" "ham" "spam" "ham" "spam" "ham" "spam" "ham" "ham" "ham"   
## [5001] "spam" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [5011] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [5021] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "spam"  
## [5031] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5041] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5051] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5061] "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham"   
## [5071] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [5081] "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [5091] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [5101] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5111] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5121] "ham" "spam" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [5131] "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham" "ham"   
## [5141] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5151] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5161] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [5171] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5181] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5191] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham"   
## [5201] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "ham" "spam" "ham"   
## [5211] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5221] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5231] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5241] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [5251] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam" "ham" "ham"   
## [5261] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [5271] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5281] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5291] "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "spam"  
## [5301] "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5311] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5321] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam"  
## [5331] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5341] "ham" "ham" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [5351] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5361] "ham" "spam" "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham"   
## [5371] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [5381] "ham" "ham" "spam" "ham" "ham" "spam" "ham" "ham" "spam" "ham"   
## [5391] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5401] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [5411] "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "spam"  
## [5421] "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham" "ham" "ham"   
## [5431] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [5441] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5451] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5461] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5471] "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham"   
## [5481] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "spam"  
## [5491] "ham" "ham" "spam" "spam" "ham" "ham" "spam" "ham" "ham" "ham"   
## [5501] "spam" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [5511] "ham" "spam" "spam" "spam" "ham" "ham" "ham" "ham" "ham" "spam"  
## [5521] "ham" "spam" "ham" "ham" "ham" "spam" "ham" "ham" "ham" "ham"   
## [5531] "ham" "ham" "spam" "ham" "ham" "ham" "ham" "ham" "spam" "ham"   
## [5541] "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham" "ham"   
## [5551] "ham" "spam" "ham" "ham" "ham" "ham" "spam" "spam" "ham"

docvars(my\_dfm,field = "training")=runif(nrow(sms\_data))>.3

training\_dfm = dfm\_subset(my\_dfm,training == TRUE)  
testing\_dfm = dfm\_subset(my\_dfm,training == FALSE)

my\_model = textmodel\_nb(training\_dfm,training\_dfm$type,smooth = 1)

require(caret)

## Loading required package: caret

## Loading required package: lattice

##   
## Attaching package: 'caret'

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

predictions = predict(my\_model,testing\_dfm)  
confusionMatrix(predictions , as.factor(testing\_dfm$type))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction ham spam  
## ham 1382 18  
## spam 27 196  
##   
## Accuracy : 0.9723   
## 95% CI : (0.9631, 0.9797)  
## No Information Rate : 0.8681   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.881   
##   
## Mcnemar's Test P-Value : 0.233   
##   
## Sensitivity : 0.9808   
## Specificity : 0.9159   
## Pos Pred Value : 0.9871   
## Neg Pred Value : 0.8789   
## Prevalence : 0.8681   
## Detection Rate : 0.8515   
## Detection Prevalence : 0.8626   
## Balanced Accuracy : 0.9484   
##   
## 'Positive' Class : ham   
##