Statistics of the textmeta Objects

```
## ##############################################################################
## ##############################################################################
## ################################## AT ##########################################
## ##############################################################################
## ##############################################################################

## number of scraped unique URLs: 5377 (22.04 %)

##  - successful: 3231 (60.09 %)

## table of scraped types of content:

##                ges successful percent
## article       3301       3231   97.88
## discussion       5          0    0.00
## event          475          0    0.00
## image          326          0    0.00
## organization   125          0    0.00
## other          541          0    0.00
## product          4          0    0.00
## profile          8          0    0.00
## video          136          0    0.00

## number of invalid UTF8 texts: 0

## number of invalid UTF8 texts after recoding: 0

## number of remaining observations: 3231

## number of associated tweets: 3780

## number of associated politicians: 65 (79.27 % of considered politicians)

## number of observations after merging (filtering to considered politicians): 3780

## Object of class "textmeta":
##  number of observations in text: 3780
##  meta: 3780 observations of 115 variables
##  range of date: 2012-02-27 till 2020-07-07

## Top 10 tweeted languages of media sources:

##
##   de  und   en   es   nl   cs   da   eu   fr   lt
## 3553  114  104    2    2    1    1    1    1    1

## keep de texts!

## Object of class "textmeta":
##  number of observations in text: 3553
##  meta: 3553 observations of 115 variables
##  range of date: 2013-08-07 till 2020-07-07

## check articles on UTF8:
```

```
## next step
```

```
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 3553
## find out vocabularies...
##  done:
##   0
##   3553 next step
## calculate counts...
##  done:
##   0
##   3553 success
## top ten words:
```

```
##         the        mehr          of oesterreich          to         abs
##        7449        4582        4150        3895        3655        3528
##       wurde           a         and    menschen
##        3336        3212        3111        3043
```

```
## keep 13871 words as vocabulary
## object after cleaning and several preprocessing steps:
```

```
## Object of class "textmeta":
##  number of observations in text: 3504
##  meta: 3504 observations of 115 variables
##  range of date: 2013-08-07 till 2020-07-07
```

```
## subsampling of texts with more than 500 tokens: 225
## texts in final preprocessed corpus by party:
```

```
##
##   FPÖ GRÜNE  NEOS   ÖVP   SPÖ
##   656   691  1130    75  1001
```

```
## Top 10 tweeted media sources:
```

```
##
##   derstandard           orf        kurier     diepresse           ots
##           606           429           315           274           246
##         krone          oe24      facebook         heute kleinezeitung
##           195            91            89            75            75
## $FPÖ
## Top 5 url_core
##     krone diepresse      oe24       ots    kurier
##       109        67        65        54        52
##
## $GRÜNE
## Top 5 url_core
## derstandard         orf      kurier         ots   diepresse
##         140         112          57          44          40
##
## $NEOS
```

```
## Top 5 url_core
## derstandard      kurier   diepresse         orf         ots
##         262         136         122         113          39
##
## $ÖVP
## Top 5 url_core
##   diepresse       krone         ots   citizengo derstandard
##          14           6           6           5           5
##
## $SPÖ
## Top 5 url_core
## derstandard         orf         ots      kurier       krone
##         181         173         103          69          39
```

```
## ##########################################################################
## ##########################################################################
## ################################### CH ###################################
## ##########################################################################
## ##########################################################################
##
## number of scraped unique URLs: 10408 (18.37 %)
##  - successful: 6325 (60.77 %)
## table of scraped types of content:
```

```
##               ges successful percent
## article      7229       6325   87.49
## discussion      5          0    0.00
## event         377          0    0.00
## image          58          0    0.00
## organization  519          0    0.00
## other         971          0    0.00
## product         5          0    0.00
## profile        20          0    0.00
## video         232          0    0.00
```

```
## number of invalid UTF8 texts: 0
## number of invalid UTF8 texts after recoding: 0
## number of remaining observations: 6325
## number of associated tweets: 8171
## number of associated politicians: 123 (86.01 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 8039
```

```
## Object of class "textmeta":
##  number of observations in text: 8039
##  meta: 8039 observations of 115 variables
##  range of date: 2011-10-04 till 2020-07-07
```

```
## Top 10 tweeted languages of media sources:
```

```
##
##   de   fr   en  und   it   nl   lt   ca   et   es
## 6739  798  360   67   40    7    6    5    4    3
```

```
## keep de texts!
```

```
## Object of class "textmeta":
##  number of observations in text: 6739
##  meta: 6739 observations of 115 variables
```

```
##  range of date: 2013-02-08 till 2020-07-07

## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 6739
## find out vocabularies...
##  done:
##   0
##   6739 next step
## calculate counts...
##  done:
##   0
##   6739 success
## top ten words:

##   schweiz      mehr       sei schweizer       the      sagt   prozent bundesrat
##     14211     10755      7184      6277      6252      6178      5770      5760
##     wurde   franken
##      5342      5021

## keep 23371 words as vocabulary
## object after cleaning and several preprocessing steps:

## Object of class "textmeta":
##  number of observations in text: 6699
##  meta: 6699 observations of 115 variables
##  range of date: 2013-02-08 till 2020-07-07

## subsampling of texts with more than 500 tokens: 732
## texts in final preprocessed corpus by party:

##
##  BDP  CVP  EVP  FDP  GLP  GPS   SP  SVP
##  263  793   48  417 1024 1022 1831 1341

## Top 10 tweeted media sources:

##
##   tagesanzeiger           blick aargauerzeitung             srf           20min
##            1260             811             500             378             277
##          watson           sp-ps         derbund             nau       bazonline
##             203             168             139             137             135
## $BDP
## Top 5 url_core
## tagesanzeiger         blick           bdp           srf         watson
##            65            57            23            18             17
##
## $CVP
## Top 5 url_core
##           blick   tagesanzeiger aargauerzeitung             srf             cvp
##             138              98              73              63              52
##
```

```
## $EVP
## Top 5 url_core
## aargauerzeitung             srf           blick   bernerzeitung     ideaschweiz
##              24               7               4               2               2
##
## $FDP
## Top 5 url_core
##           blick   tagesanzeiger             srf aargauerzeitung           20min
##              70              54              27              25              23
##
## $GLP
## Top 5 url_core
##   tagesanzeiger aargauerzeitung           blick             srf           20min
##             224              74              74              67              40
##
## $GPS
## Top 5 url_core
##   tagesanzeiger           blick aargauerzeitung          watson          gruene
##             243              81              77              52              51
##
## $SP
## Top 5 url_core
##   tagesanzeiger           blick          sp-ps aargauerzeitung             srf
##             427             185             168             120              97
##
## $SVP
## Top 5 url_core
##           blick   tagesanzeiger           20min aargauerzeitung       bazonline
##             202             147             116              95              95
## ################################################################################
## ################################################################################
## ################################### DK #########################################
## ################################################################################
## ################################################################################
##
## do not consider party "Nye Borgerlige" because of small amount of shared texts (7)!
## number of scraped unique URLs: 10987 (26.97 %)
##  - successful: 8748 (79.62 %)
## table of scraped types of content:
##               ges successful percent
## article      8855       8748   98.79
## discussion    177          0    0.00
## event         386          0    0.00
## image          30          0    0.00
## organization  275          0    0.00
## other         389          0    0.00
## product        12          0    0.00
## profile        17          0    0.00
## video         105          0    0.00
## number of invalid UTF8 texts: 0
## number of invalid UTF8 texts after recoding: 0
## number of remaining observations: 8748
```

```
## number of associated tweets: 11374
## number of associated politicians: 100 (63.69 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 11153

## Object of class "textmeta":
##  number of observations in text: 11153
##  meta: 11153 observations of 115 variables
##  range of date: 2011-03-25 till 2020-07-07

## Top 10 tweeted languages of media sources:

##
##    da    en   und    no    sv    es    de    in    fr    nl
## 10441   404   156    98    23     7     5     5     4     2

## keep da texts!

## Object of class "textmeta":
##  number of observations in text: 10441
##  meta: 10441 observations of 115 variables
##  range of date: 2013-08-24 till 2020-07-07

## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 10441
## find out vocabularies...
##  done:
##   0
##   10441 next step
## calculate counts...
##  done:
##   0
##   10441 success
## top ten words:

##     kan      så   siger     ved danmark      år    mere     the      få   flere
##   37268   29975   26808   16130   15638   13959   13116   12760   12193   12151

## keep 22681 words as vocabulary
## object after cleaning and several preprocessing steps:

## Object of class "textmeta":
##  number of observations in text: 10397
##  meta: 10397 observations of 115 variables
##  range of date: 2013-08-24 till 2020-07-07

## subsampling of texts with more than 500 tokens: 943
## texts in final preprocessed corpus by party:

##
##   DF   EL   KF   LA   RV    S   SF  UFG    V
## 1057 1521  415   70 1245 2140 1306  743 1944

## Top 10 tweeted media sources:
```

```
##
##              dr      berlingske       politiken jyllands-posten             tv2
##            1800            1271            1182            1016             888
##        altinget     information    ekstrabladet              bt     ditoverblik
##             520             402             364             322             233
## $DF
## Top 5 url_core
##     ditoverblik              dr jyllands-posten             tv2      berlingske
##             233             107             107              73              72
##
## $EL
## Top 5 url_core
##              dr       politiken jyllands-posten      berlingske     information
##             340             198             143             109              99
##
## $KF
## Top 5 url_core
##             tv2              dr      berlingske jyllands-posten        altinget
##             113              80              70              34              18
##
## $LA
## Top 5 url_core
##      berlingske jyllands-posten              dr             tv2        altinget
##              19              11               8               6               4
##
## $RV
## Top 5 url_core
##              dr       politiken      berlingske             tv2 jyllands-posten
##             294             210             177              89              79
##
## $S
## Top 5 url_core
##              dr      berlingske jyllands-posten             tv2       politiken
##             382             256             203             202             195
##
## $SF
## Top 5 url_core
##              dr       politiken             tv2        altinget jyllands-posten
##             236             192             126             103             103
##
## $UFG
## Top 5 url_core
##       politiken              dr      berlingske     information jyllands-posten
##             188             112              83              64              41
##
## $V
## Top 5 url_core
##      berlingske jyllands-posten              dr             tv2       politiken
##             406             295             241             159             154
## ################################################################################
## ################################################################################
## ################################### ESP ########################################
## ################################################################################
```

```
## ##########################################################################
##
## number of scraped unique URLs: 37739 (32.27 %)
##  - successful: 30016 (79.54 %)
## table of scraped types of content:

##                ges successful percent
## article      32600      30016   92.07
## discussion       7          0    0.00
## event          236          0    0.00
## image          263          0    0.00
## organization   302          0    0.00
## other         1750          0    0.00
## product          7          0    0.00
## profile         28          0    0.00
## video          536          0    0.00

## number of invalid UTF8 texts: 1
## error while recoding!
## number of invalid UTF8 texts after recoding: 1
## number of remaining observations: 30015
## number of associated tweets: 38427
## number of associated politicians: 281 (98.25 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 32533

## Object of class "textmeta":
##  number of observations in text: 32533
##  meta: 32533 observations of 115 variables
##  range of date: 2011-05-15 till 2020-07-07

## Top 10 tweeted languages of media sources:

##
##    es    ca   und    en    pt    eu    fr    it    de    ht
## 29686  2072   343   196   117    46    36    13     3     3

## keep es texts!

## Object of class "textmeta":
##  number of observations in text: 29686
##  meta: 29686 observations of 115 variables
##  range of date: 2011-05-15 till 2020-07-07

## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 29686
## find out vocabularies...
##  done:
##   0
##   29686 next step
## calculate counts...
##  done:
```

```
##   0
##   29686 success
## top ten words:

## gobierno   españa       si       pp      ser     años  sánchez  partido
##    77451    49791    49071    35424    35353    33428    32272    29043
## personas      dos
##    28296    28239

## keep 49422 words as vocabulary
## object after cleaning and several preprocessing steps:

## Object of class "textmeta":
##  number of observations in text: 29594
##  meta: 29594 observations of 115 variables
##  range of date: 2011-05-15 till 2020-07-07

## subsampling of texts with more than 500 tokens: 5567
## texts in final preprocessed corpus by party:

##
##                  Cs             EAJ-PNV ECP-GUAYEM EL CANVI             EH Bildu
##                1541                  60                 210                  178
##           ERC-CATSÍ               ERC-S         JxCat-JUNTS                   PP
##                 245                 300                  63                 9027
##            PSC-PSOE          PsdeG-PSOE               PSOE                   UP
##                1634                 462                7341                 2989
##                 VOX
##                5636

## Top 10 tweeted media sources:

##
##          elpais             abc         elmundo        okdiario libertaddigital
##            2852            2634            2522            1241            1012
##       cadenaser     europapress       elespanol  elconfidencial         publico
##             926             906             888             844             843
## $Cs
## Top 5 url_core
##        elmundo            abc      elespanol elconfidencial    europapress
##            356            159            131            119             93
##
## $`EAJ-PNV`
## Top 5 url_core
##              facebook               elpais               eaj-pnv
##                    15                    8                    5
## jokinbildarratzsorron           europapress
##                     4                    3
##
## $`ECP-GUAYEM EL CANVI`
## Top 5 url_core
## lavanguardia  elperiodico       elpais     eldiario  cuartopoder
##           43           35           30           18           14
##
## $`EH Bildu`
## Top 5 url_core
##        naiz     publico cuartopoder        eitb    eldiario
```

```
##          80          13           8           8           7
##
## $`ERC-CATSÍ`
## Top 5 url_core
##   elnacional      publico lavanguardia  elperiodico     diario16
##           29           24           21           18           16
##
## $`ERC-S`
## Top 5 url_core
## lavanguardia      publico  elperiodico       elpais     eldiario
##           66           30           25           19           12
##
## $`JxCat-JUNTS`
## Top 5 url_core
##   lavanguardia    elperiodico         elpais     elnacional elsiglodeuropa
##             12              7              5              4              3
##
## $PP
## Top 5 url_core
##             abc         elmundo        okdiario libertaddigital         larazon
##            1758            1385             468             408             323
##
## $`PSC-PSOE`
## Top 5 url_core
##       elpais  elperiodico lavanguardia    elespanol    cadenaser
##          312          239          163          114           73
##
## $`PsdeG-PSOE`
## Top 5 url_core
##         elpais lavozdegalicia      cadenaser       eldiario     farodevigo
##             80             40             35             30             28
##
## $PSOE
## Top 5 url_core
##      elpais   cadenaser        psoe    eldiario europapress
##        1572         598         305         261         236
##
## $UP
## Top 5 url_core
##       publico        elpais      eldiario  lavanguardia elsaltodiario
##           522           341           262           143           116
##
## $VOX
## Top 5 url_core
##        okdiario             abc libertaddigital         elmundo  elconfidencial
##             751             624             584             470             223
## ##############################################################################
## ##############################################################################
## ################################## FR ##########################################
## ##############################################################################
## ##############################################################################
##
## number of scraped unique URLs: 48829 (24.2 %)
```

```
##  - successful: 34770 (71.21 %)
## table of scraped types of content:

##                ges successful percent
## article      35430      34770   98.14
## discussion      58          0    0.00
## event         1245          0    0.00
## image          469          0    0.00
## job              1          0    0.00
## organization  1623          0    0.00
## other         3505          0    0.00
## product         16          0    0.00
## profile         73          0    0.00
## video         1818          0    0.00

## number of invalid UTF8 texts: 11
## error while recoding!
## number of invalid UTF8 texts after recoding: 11
## number of remaining observations: 34759
## number of associated tweets: 47628
## number of associated politicians: 502 (96.91 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 42716

## Object of class "textmeta":
##  number of observations in text: 42716
##  meta: 42716 observations of 115 variables
##  range of date: 2011-09-20 till 2020-07-07

## Top 10 tweeted languages of media sources:

##
##    fr   und    en    es    ca    de    pt    it    ro    no
## 41503   736   310    38    36    19    14    12     9     5

## keep fr texts!

## Object of class "textmeta":
##  number of observations in text: 41503
##  meta: 41503 observations of 115 variables
##  range of date: 2011-09-20 till 2020-07-07

## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 14 success
## number of articles: 41489
## find out vocabularies...
##  done:
##   0
##   41489 next step
## calculate counts...
##  done:
##   0
##   41489 success
```

```
## top ten words:

##      a   plus   cest france    dun   dune  comme   être  aussi   fait
## 257913 166608  93984  80505  73459  73272  71426  67244  61673  56769

## keep 59455 words as vocabulary
## object after cleaning and several preprocessing steps:

## Object of class "textmeta":
##  number of observations in text: 40986
##  meta: 40986 observations of 115 variables
##  range of date: 2011-09-20 till 2020-07-07

## subsampling of texts with more than 500 tokens: 8020
## texts in final preprocessed corpus by party:

##
##  Agir   DVG    FI    LR  LREM MoDem   PCF    PS    RN   UDI
##   695   541  3562  6563 22176  2074   940  2960  1398   594

## Top 10 tweeted media sources:

##
##      lemonde     lefigaro   leparisien francetvinfo        lejdd     lesechos
##         3634         2697         2228         1934         1411         1334
##   liberation ouest-france     facebook        bfmtv
##         1215         1174         1093         1064
## $Agir
## Top 5 url_core
## huffingtonpost       lefigaro     leparisien   francetvinfo        lemonde
##            101             85             69             38             37
##
## $DVG
## Top 5 url_core
## andrechassaigne        humanite    francetvinfo      liberation    alainbruneel
##             318              30              27              11               7
##
## $FI
## Top 5 url_core
## lafranceinsoumise           lemonde         melenchon        liberation
##               701               265               250               189
##      francetvinfo
##               169
##
## $LR
## Top 5 url_core
##          lefigaro         leparisien virginiedubymuller            lemonde
##              1042               397               394               291
##       francetvinfo
##               255
##
## $LREM
## Top 5 url_core
##      lemonde   leparisien     lefigaro francetvinfo     lesechos
##         2274         1174         1095         1035          970
##
## $MoDem
```

```
## Top 5 url_core
##            lemonde           lefigaro         leparisien       francetvinfo
##                170                137                100                 81
## mouvementdemocrate
##                 79
##
## $PCF
## Top 5 url_core
##      lemonde   leparisien francetvinfo   liberation        bfmtv
##          165           96           70           49           34
##
## $PS
## Top 5 url_core
##        lemonde claudegoasguen     liberation     leparisien       facebook
##            328            325            213            162            134
##
## $RN
## Top 5 url_core
## rassemblementnational              lefigaro      valeursactuelles
##                   176                   131                   116
##          francetvinfo                 bfmtv
##                    98                    65
##
## $UDI
## Top 5 url_core
##     lefigaro      lemonde   leparisien ouest-france francetvinfo
##           55           44           44           37           35
```

```
## ##############################################################################
## ##############################################################################
## ################################### GER ######################################
## ##############################################################################
## ##############################################################################
##
## number of scraped unique URLs: 66612 (25.6 %)
##  - successful: 49388 (74.14 %)
## table of scraped types of content:
```

```
##                ges successful percent
## article      49789      49210   98.84
## discussion      32          0    0.00
## event         2969          0    0.00
## faq              1          0    0.00
## image          750          5    0.67
## job              1          0    0.00
## organization  1348          0    0.00
## other         5405          0    0.00
## product         15          0    0.00
## profile        351          0    0.00
## video          941          0    0.00
```

```
## number of invalid UTF8 texts: 1
## error while recoding!
## number of invalid UTF8 texts after recoding: 1
## number of remaining observations: 49387
```

```
## number of associated tweets: 54536
## number of associated politicians: 477 (90.34 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 75030

## Object of class "textmeta":
##  number of observations in text: 75030
##  meta: 75030 observations of 115 variables
##  range of date: 2011-06-20 till 2020-08-03

## Top 10 tweeted languages of media sources:

##
##    de   und    en    fr    nl    da    es    tr    no    sv
## 71808  1372  1345   117    81    58    49    36    34    25

## keep de texts!

## Object of class "textmeta":
##  number of observations in text: 71808
##  meta: 71808 observations of 115 variables
##  range of date: 2011-06-20 till 2020-08-03

## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 71808
## find out vocabularies...
##  done:
##   0
##   71808 next step
## calculate counts...
##  done:
##   0
##   71808 success
## top ten words:

##        mehr         the deutschland         sei     prozent    menschen
##      133604      110544      109304       96717       86896       77331
##     muessen        euro       sagte          of
##       71881       71458       69386       64407

## keep 116920 words as vocabulary
## object after cleaning and several preprocessing steps:

## Object of class "textmeta":
##  number of observations in text: 71516
##  meta: 71516 observations of 115 variables
##  range of date: 2011-06-20 till 2020-08-03

## subsampling of texts with more than 500 tokens: 11667
## texts in final preprocessed corpus by party:

##
##   AfD   CDU   CSU   FDP GRÜNE LINKE   SPD
```

```
## 16204  7303   764 10560 13701 13990  9286

## Top 10 tweeted media sources:

##
##      spiegel         welt sueddeutsche   tagesschau tagesspiegel          faz
##         9756         7517         3500         2889         2882         2591
##         bild        focus handelsblatt    bundestag
##         2569         1762         1565         1544
## $AfD
## Top 5 url_core
##          welt          bild jungefreiheit         focus      facebook
##          3376          1264          1071          1042           975
##
## $CDU
## Top 5 url_core
##     cducsu       welt    spiegel tagesschau       bild
##       1037        900        556        472        445
##
## $CSU
## Top 5 url_core
##    welt    bild spiegel  cducsu     faz
##     176     105      91      56      43
##
## $FDP
## Top 5 url_core
##         welt      spiegel          faz          fdp handelsblatt
##         1512         1190          858          699          497
##
## $GRÜNE
## Top 5 url_core
##      spiegel sueddeutsche tagesspiegel   tagesschau          taz
##         2401         1169          749          704          671
##
## $LINKE
## Top 5 url_core
##      spiegel   tagesschau sueddeutsche  ulla-jelpke          taz
##         2295          635          632          595          559
##
## $SPD
## Top 5 url_core
##      spiegel sueddeutsche   tagesschau tagesspiegel         welt
##         2458          921          514          481          437

## ##########################################################################
## ##########################################################################
## ################################### IT ###################################
## ##########################################################################
## ##########################################################################

## Warning in readRDS(file.path("data", i, "botdata.rds")): strings not
## representable in native encoding will be translated to UTF-8

## Warning in readRDS(file.path("data", i, "botdata.rds")): input string '{"og_vocabulary":[{"uri":"http
##  The Ash Center invites you to a public discussion about empowering citizens through direct democracy
## </p>
```

```
##
## <p>
##  <img height="90" width="177" alt="Yes Europe Lab" title="Yes Europe Lab" class="media-element file-
```

```
## do not consider party "Independente" because of lack of interpretation!
## number of scraped unique URLs: 50227 (24.8 %)
##  - successful: 22424 (44.65 %)
## table of scraped types of content:
```

```
##                ges successful percent
## article      23673      22424   94.72
## discussion     358          0    0.00
## event         3424          0    0.00
## faq              2          0    0.00
## image         2517          0    0.00
## organization   683          0    0.00
## other         5607          0    0.00
## product         12          0    0.00
## profile         72          0    0.00
## video         4929          0    0.00
```

```
## number of invalid UTF8 texts: 1
## number of invalid UTF8 texts after recoding: 0
## number of remaining observations: 22424
## number of associated tweets: 26643
## number of associated politicians: 309 (67.17 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 24197
```

```
## Object of class "textmeta":
##  number of observations in text: 24197
##  meta: 24197 observations of 115 variables
##  range of date: 2012-02-10 till 2020-07-08
```

```
## Top 10 tweeted languages of media sources:
```

```
##
##    it    en   und    es    fr    pt    ca    de    in    ro
## 22730   904   336    91    47    25    15    11    10     9
```

```
## keep it texts!
```

```
## Object of class "textmeta":
##  number of observations in text: 22730
##  meta: 22730 observations of 115 variables
##  range of date: 2012-02-10 till 2020-07-07
```

```
## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 22730
## find out vocabularies...
##  done:
##   0
```

```
##   22730 next step
## calculate counts...
##  done:
##   0
##   22730 success
## top ten words:

## governo     the   stato    solo    anni  essere  lavoro   fatto    fare   legge
##   30034   29867   29627   24255   23553   23506   19700   19527   18728   18142

## keep 42159 words as vocabulary
## object after cleaning and several preprocessing steps:

## Object of class "textmeta":
##  number of observations in text: 22555
##  meta: 22555 observations of 115 variables
##  range of date: 2012-02-10 till 2020-07-07

## subsampling of texts with more than 500 tokens: 2409
## texts in final preprocessed corpus by party:

##
##  FdI   FI   IV   LN  M5S   PD
## 1211 2584 2854 2488 9617 3976

## Top 10 tweeted media sources:

##
##          facebook ilblogdellestelle        repubblica ilfattoquotidiano
##              5227              2459              1541               843
##    huffingtonpost              ansa        ilgiornale          corriere
##               595               586               513               404
##      ilmessaggero       ilsole24ore
##               304               301
## $FdI
## Top 5 url_core
##     secoloditalia        ilgiornale      ilmessaggero          facebook
##               160               131                84                55
## lavocedelpatriota
##                50
##
## $FI
## Top 5 url_core
##      facebook    repubblica          ansa    ilgiornale armandocesaro
##           414           163           125           114           112
##
## $IV
## Top 5 url_core
##       facebook     repubblica huffingtonpost      globalist    matteorenzi
##            782            341            124            123             86
##
## $LN
## Top 5 url_core
##     facebook   ilgiornale         ansa   repubblica ilmessaggero
##          304          253          106          105           84
##
## $M5S
```

```
## Top 5 url_core
##          facebook ilblogdellestelle ilfattoquotidiano     fabianadadone
##              2628              2428               687               255
##        repubblica
##               247
##
## $PD
## Top 5 url_core
##           facebook         repubblica     huffingtonpost               ansa
##               1044                638                222                116
## partitodemocratico
##                112

## ##########################################################################
## ##########################################################################
## ################################## NL ######################################
## ##########################################################################
## ##########################################################################
##
## number of scraped unique URLs: 28509 (30.42 %)
##  - successful: 17736 (62.21 %)
## table of scraped types of content:

##                ges successful percent
## article      18634      17736   95.18
## discussion      41          0    0.00
## event         1316          0    0.00
## image           56          0    0.00
## organization  2275          0    0.00
## other         2516          0    0.00
## product         21          0    0.00
## profile         45          0    0.00
## video          318          0    0.00

## number of invalid UTF8 texts: 0
## number of invalid UTF8 texts after recoding: 0
## number of remaining observations: 17736
## number of associated tweets: 23382
## number of associated politicians: 135 (95.74 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 23017

## Object of class "textmeta":
##  number of observations in text: 23017
##  meta: 23017 observations of 115 variables
##  range of date: 2011-08-19 till 2020-07-08

## Top 10 tweeted languages of media sources:

##
##    nl    en   und    da    de    fr    es    cy    it    no
## 21497  1056   302    31    28    24    15    12     8     8

## keep nl texts!

## Object of class "textmeta":
##  number of observations in text: 21497
##  meta: 21497 observations of 115 variables
##  range of date: 2011-09-12 till 2020-07-08
```

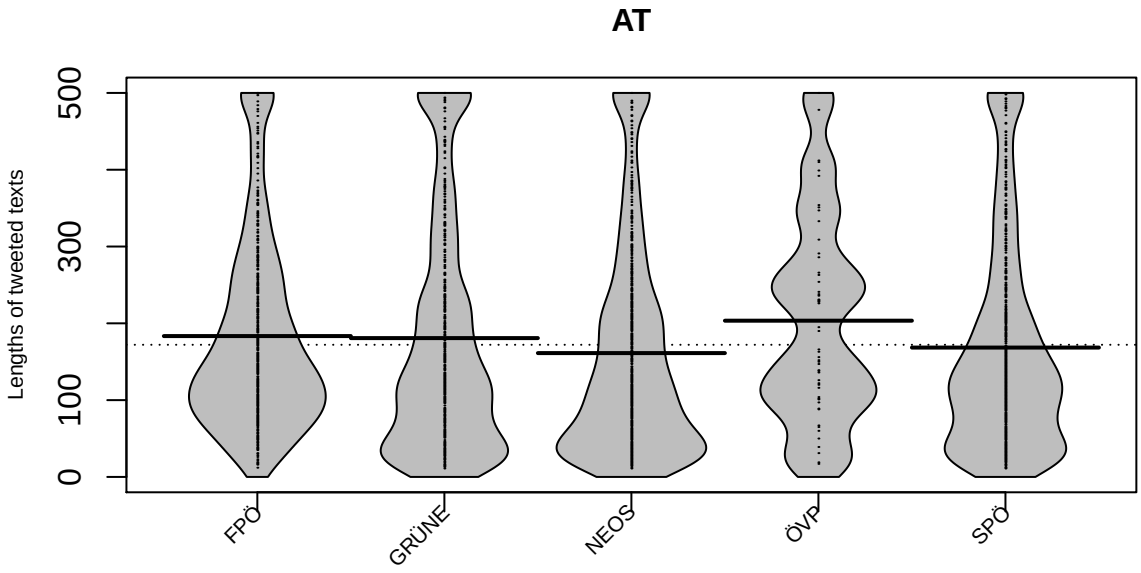
```
## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 0 success
## number of articles: 21497
## find out vocabularies...
##  done:
##   0
##   21497 next step
## calculate counts...
##  done:
##   0
##   21497 success
## top ten words:
##       the        we      jaar    mensen        to nederland       and    moeten
##     58787     51027     29023     28252     25355     25152     24801     24381
##       wel      gaat
##     23380     21367
## keep 38935 words as vocabulary
## object after cleaning and several preprocessing steps:
## Object of class "textmeta":
##  number of observations in text: 20911
##  meta: 20911 observations of 115 variables
##  range of date: 2011-09-12 till 2020-07-08
## subsampling of texts with more than 500 tokens: 2233
## texts in final preprocessed corpus by party:
##
## 50PLUS    CDA     CU    D66   DENK    FvD     GL   PvdA   PvdD    PVV    SGP
##    501   2115   1100   3114    351    242   2667   1951   1042   3013    204
##     SP    VVD
##   3292   1905
## Top 10 tweeted media sources:
##
##        nos   dpgmedia volkskrant        nrc  rtlnieuws         nu  telegraaf
##       3628       2008       1340       1303        972        933        528
##         sp      trouw   facebook
##        421        396        354
## $`50PLUS`
## Top 5 url_core
## 50pluspartij     dpgmedia          nos    telegraaf   volkskrant
##           64           54           46           31           30
##
## $CDA
## Top 5 url_core
##           nos      dpgmedia renepetersoss           nu    volkskrant
##           283           179           148           105            98
```

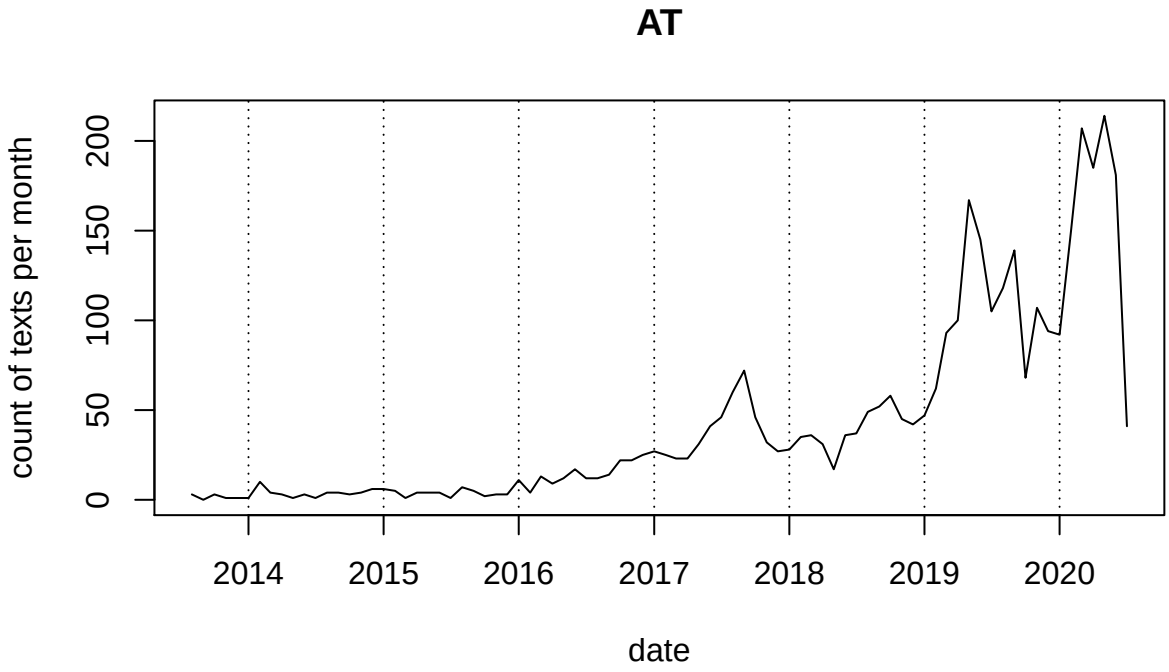
```
##
## $CU
## Top 5 url_core
##          nos christenunie     dpgmedia           nu   volkskrant
##          213          212           88           68           58
##
## $D66
## Top 5 url_core
##        nos   dpgmedia        nrc volkskrant         nu
##        744        313        279        217        176
##
## $DENK
## Top 5 url_core
##        nos   dpgmedia volkskrant        nrc         nu
##         74         27         27         25         18
##
## $FvD
## Top 5 url_core
## dagelijksestandaard                 nos    elsevierweekblad            dpgmedia
##                  58                  20                  17                  16
##  ongehoordnederland
##                  14
##
## $GL
## Top 5 url_core
##        nos        nrc volkskrant   dpgmedia groenlinks
##        424        253        240        228        214
##
## $PvdA
## Top 5 url_core
##        nos   dpgmedia       pvda        nrc volkskrant
##        413        195        167        125        125
##
## $PvdD
## Top 5 url_core
## partijvoordedieren                nos           dpgmedia         volkskrant
##                224                169                 74                 63
##                 nu
##                 53
##
## $PVV
## Top 5 url_core
##  dpgmedia       nos telegraaf rtlnieuws       tpo
##       338       306       263       227       222
##
## $SGP
## Top 5 url_core
##        nos         rd         nu volkskrant        nrc
##         72         21         10          9          8
##
## $SP
## Top 5 url_core
##        nos         sp   dpgmedia volkskrant        nrc
##        541        417        332        268        244
```

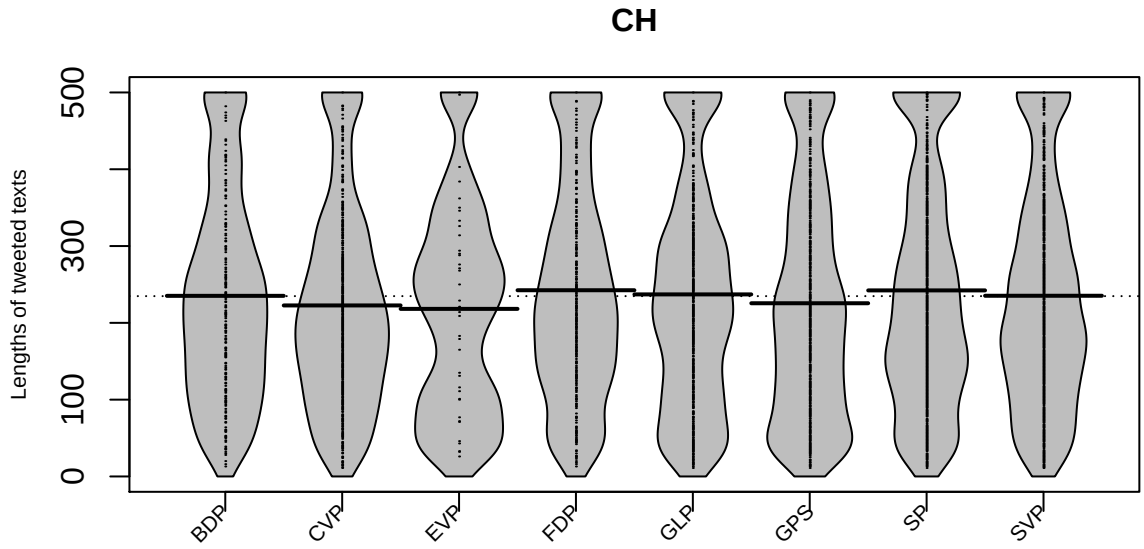
```
##
## $VVD
## Top 5 url_core
##        nos   dpgmedia         nu        vvd volkskrant
##        323        158         99         94         89

## ##########################################################################
## ##########################################################################
## ################################## UK ######################################
## ##########################################################################
## ##########################################################################
##
## number of scraped unique URLs: 70336 (23.51 %)
##  - successful: 50751 (72.16 %)
## table of scraped types of content:

##                ges successful percent
## article      51424      50751   98.69
## discussion      39          0    0.00
## event         1846          0    0.00
## faq              2          0    0.00
## image          160          0    0.00
## job            182          0    0.00
## organization  1560          0    0.00
## other         6760          0    0.00
## product         66          0    0.00
## profile        119          0    0.00
## video          905          0    0.00

## number of invalid UTF8 texts: 0
## number of invalid UTF8 texts after recoding: 0
## number of remaining observations: 50751
## number of associated tweets: 72072
## number of associated politicians: 576 (97.79 % of considered politicians)
## number of observations after merging (filtering to considered politicians): 70422

## Object of class "textmeta":
##  number of observations in text: 70422
##  meta: 70422 observations of 115 variables
##  range of date: 2011-06-16 till 2020-07-08

## Top 10 tweeted languages of media sources:

##
##    en   und    fr    es    de    et    ca    in    cy    pt
## 69149  1103    43    13    12    12    10    10     9     8

## keep en texts!

## Object of class "textmeta":
##  number of observations in text: 69149
##  meta: 69149 observations of 115 variables
##  range of date: 2011-06-16 till 2020-07-08

## check articles on UTF8: next step
## change to lowercase: next step
## remove punctuation: next step
## remove numbers: next step
```

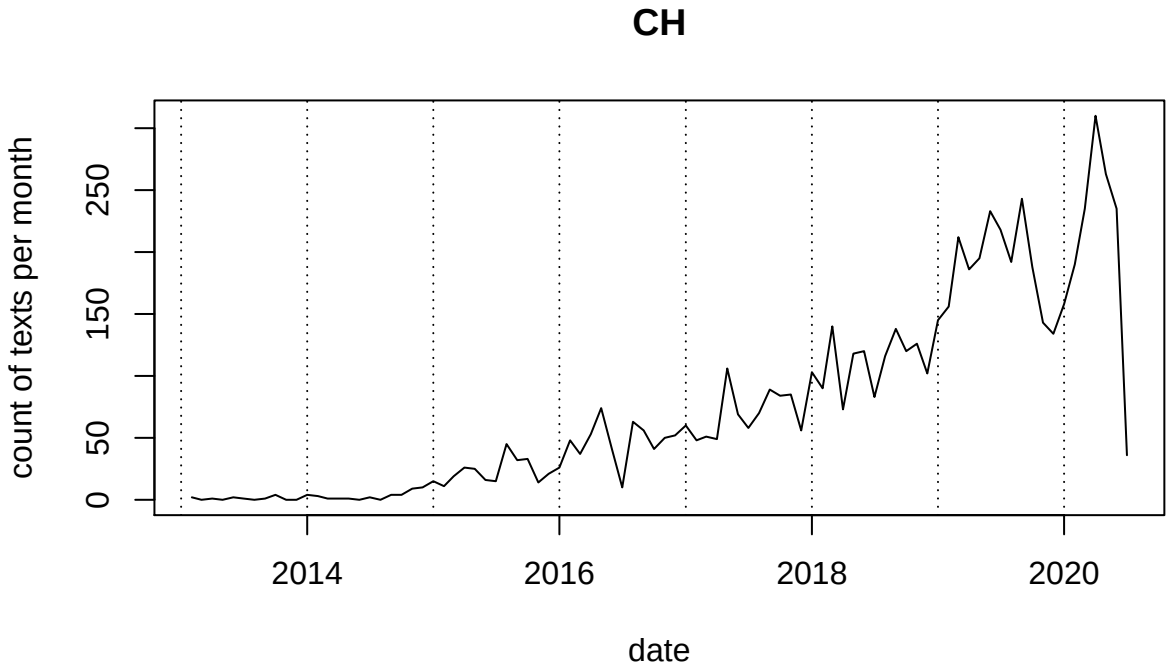
```
## remove stopwords: next step
## remove redundant whitespace: next step
## tokenize: next step
## remove empty articles: 1 success
## number of articles: 69148
## find out vocabularies...
##  done:
##   0
##   69148 next step
## calculate counts...
##  done:
##   0
##   69148 success
## top ten words:
##       will       said     people        can government         uk        new
##     337151     191821     191695     177148     166192     151263     115512
##       also    support        one
##     112554     106927     106654
## keep 47790 words as vocabulary
## object after cleaning and several preprocessing steps:
## Object of class "textmeta":
##  number of observations in text: 68672
##  meta: 68672 observations of 115 variables
##  range of date: 2011-06-16 till 2020-07-08
## subsampling of texts with more than 500 tokens: 15574
## texts in final preprocessed corpus by party:
##
##      Conservative               DUP            Labour Liberal Democrats
##             28869               279             32196              2223
## Scottish National         Sinn Féin
##              5032               550
## Top 10 tweeted media sources:
##
## theguardian         gov         bbc      mirror   telegraph    thetimes
##        9305        7410        4245        2908        2649        2218
## independent  parliament      labour   dailymail
##        2185        1144        1101        1092
## $Conservative
## Top 5 url_core
##       gov telegraph       bbc  thetimes dailymail
##      5771      2351      1533      1277       838
##
## $DUP
## Top 5 url_core
## belfasttelegraph              bbc       newsletter           forbes
##               41               40               29               15
##              gov
##               15
##
## $Labour
```

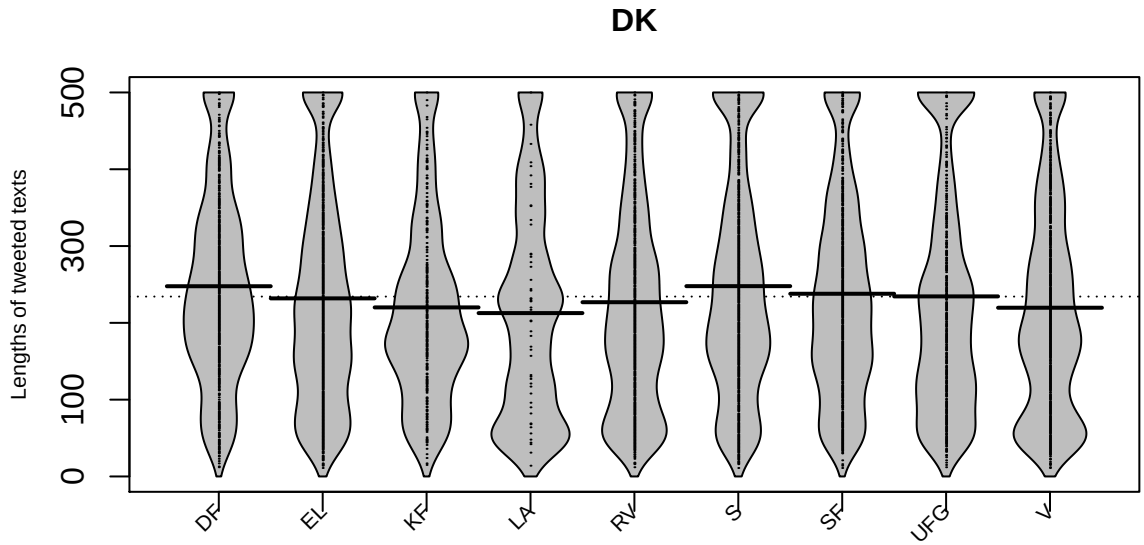
```
## Top 5 url_core
## theguardian      mirror         bbc independent         gov
##        7707        2696        2084        1684        1299
##
## $`Liberal Democrats`
## Top 5 url_core
## theguardian         bbc independent    thetimes      mirror
##         380         236         163          82          68
##
## $`Scottish National`
## Top 5 url_core
##    thenational    theguardian            bbc            gov heraldscotland
##             548            434            335            251            201
##
## $`Sinn Féin`
## Top 5 url_core
##    sinnfein  irishtimes theguardian         bbc    facebook
##         292          31          22          17          17
```

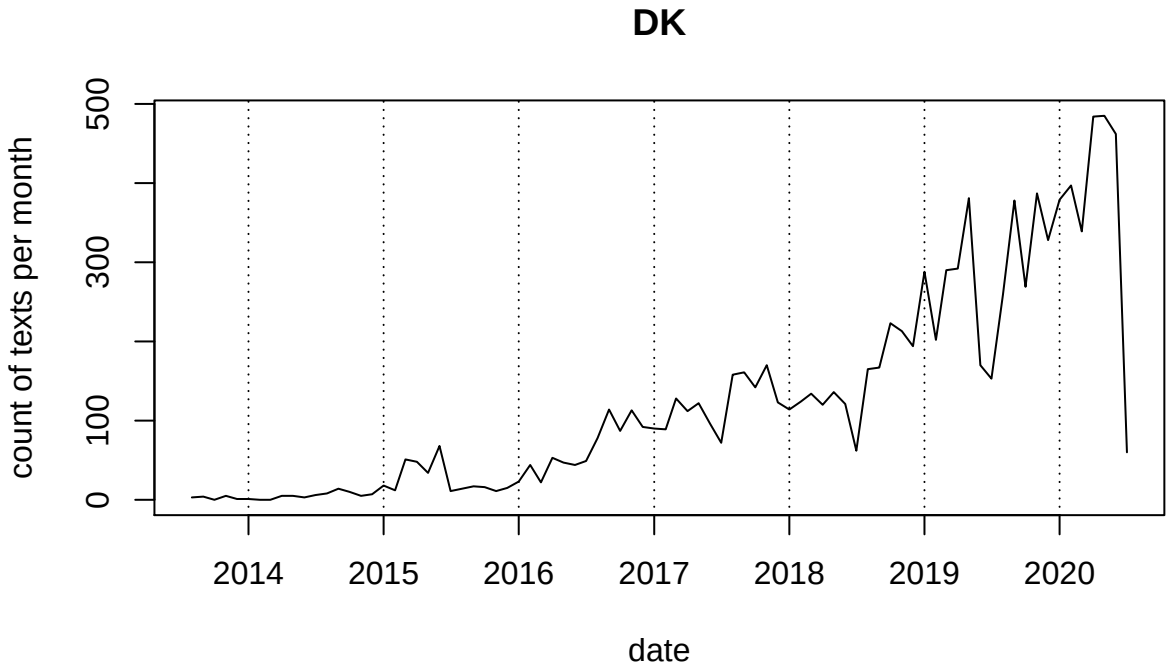


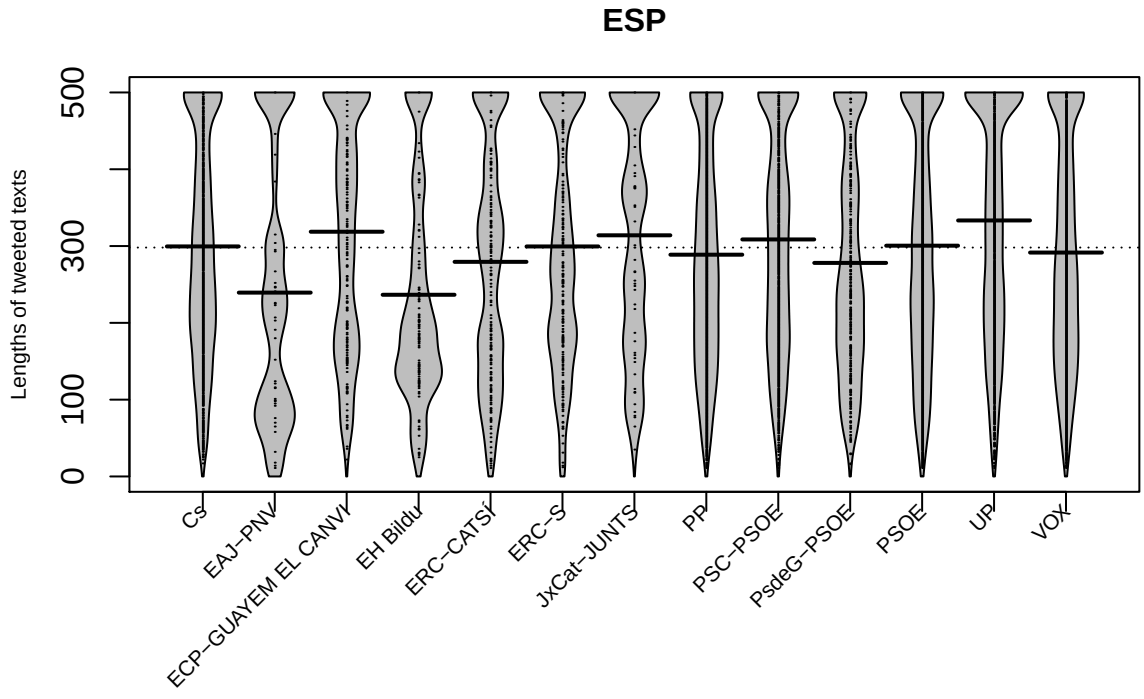

AT
count of texts per month
0
50
100
150
200
2014
2015
2016
2017
2018
2019
2020
date

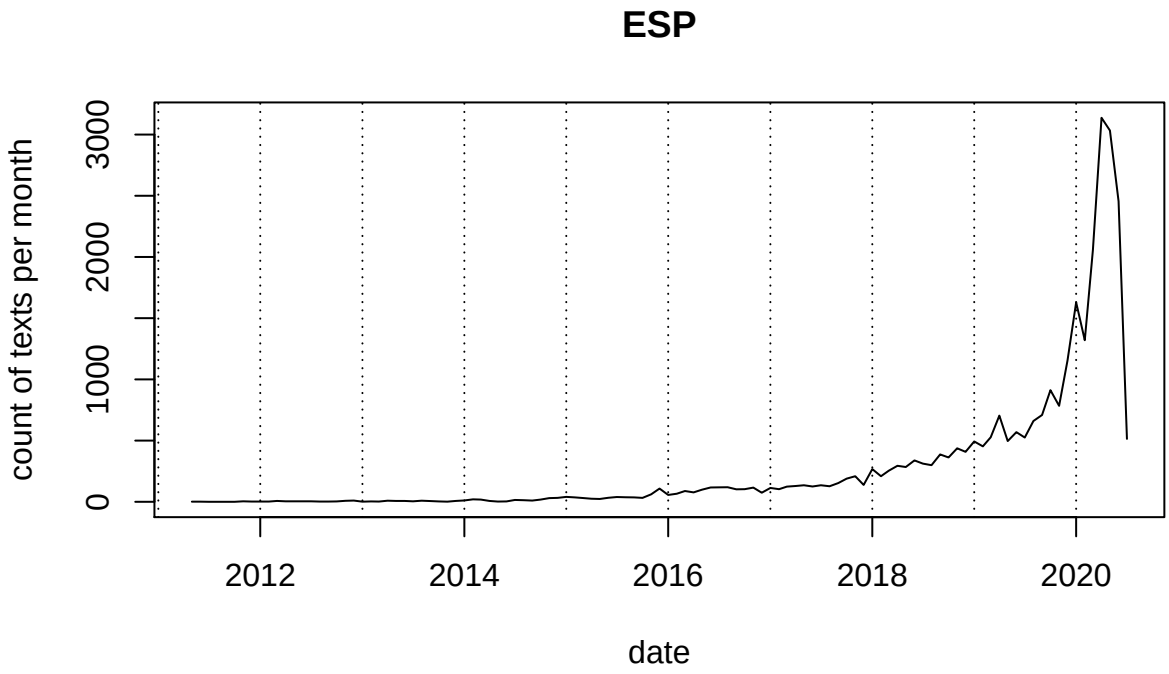

CH
Lengths of tweeted texts
0
100
300
500
BDP
CVP
EVP
FDP
GLP
GPS
SP
SVP

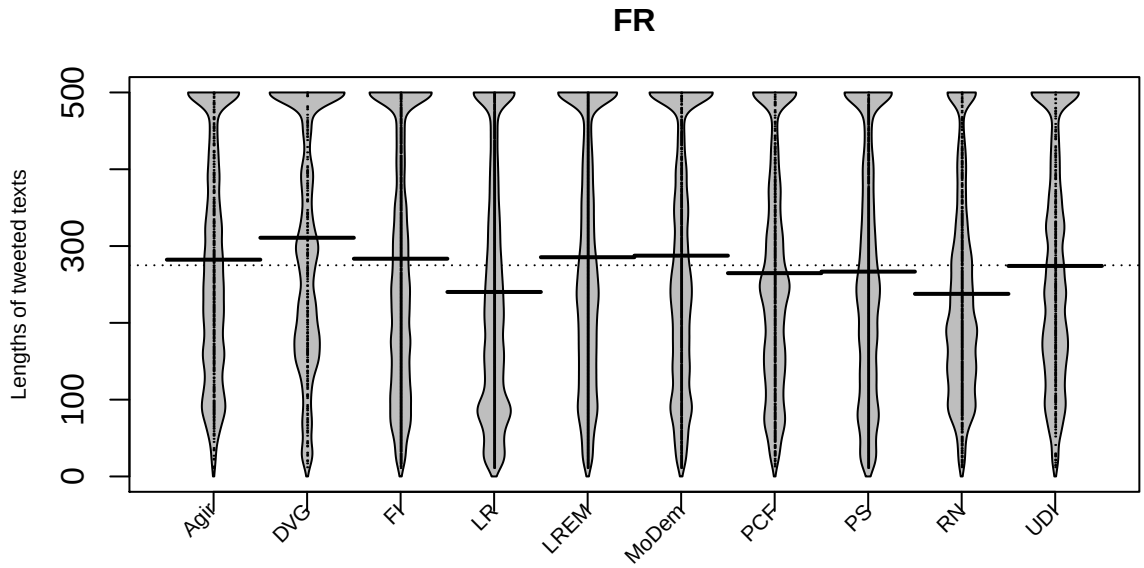

CH
count of texts per month
0
50
150
250
2014
2016
2018
2020
date

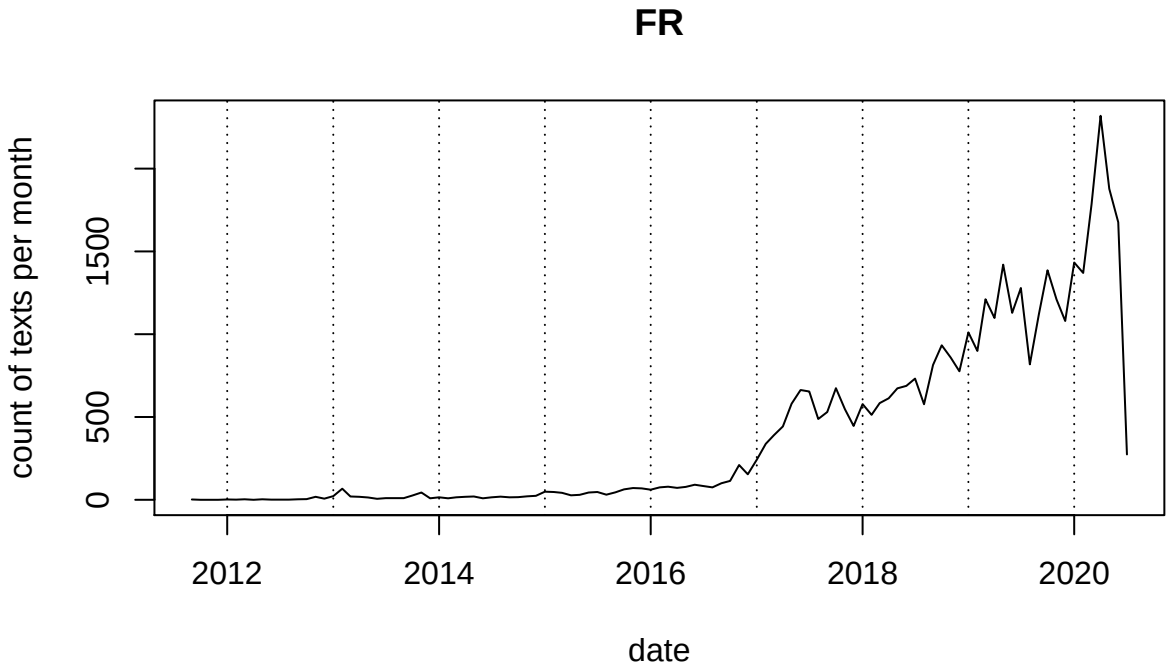

DK
Lengths of tweeted texts
0
100
300
500
DF
EL
KF
LA
RV
S
SF
UFG
V

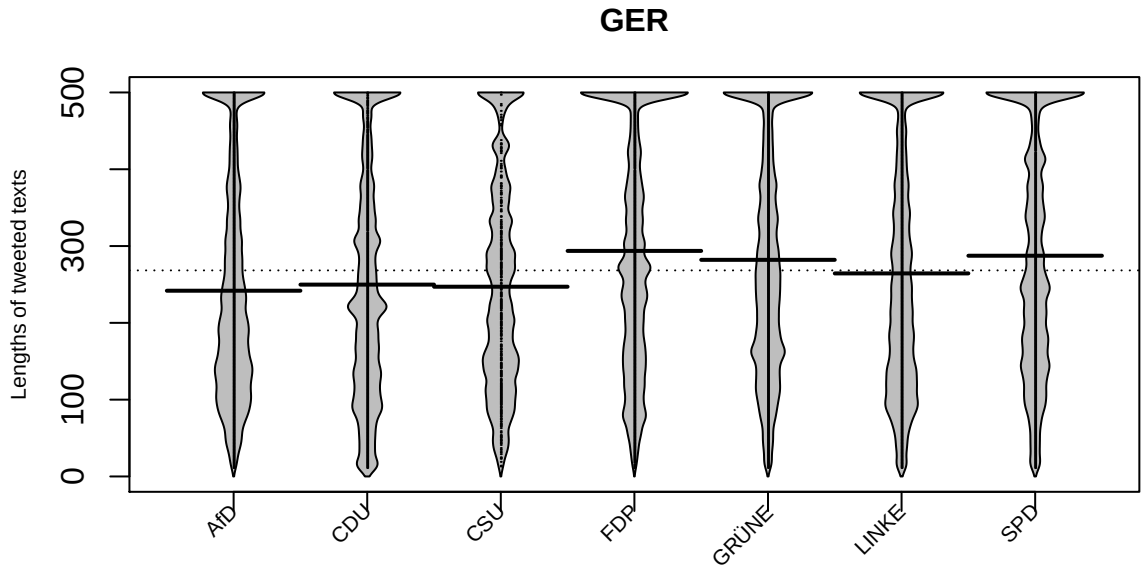

DK
count of texts per month
0
100
300
500
2014
2015
2016
2017
2018
2019
2020
date

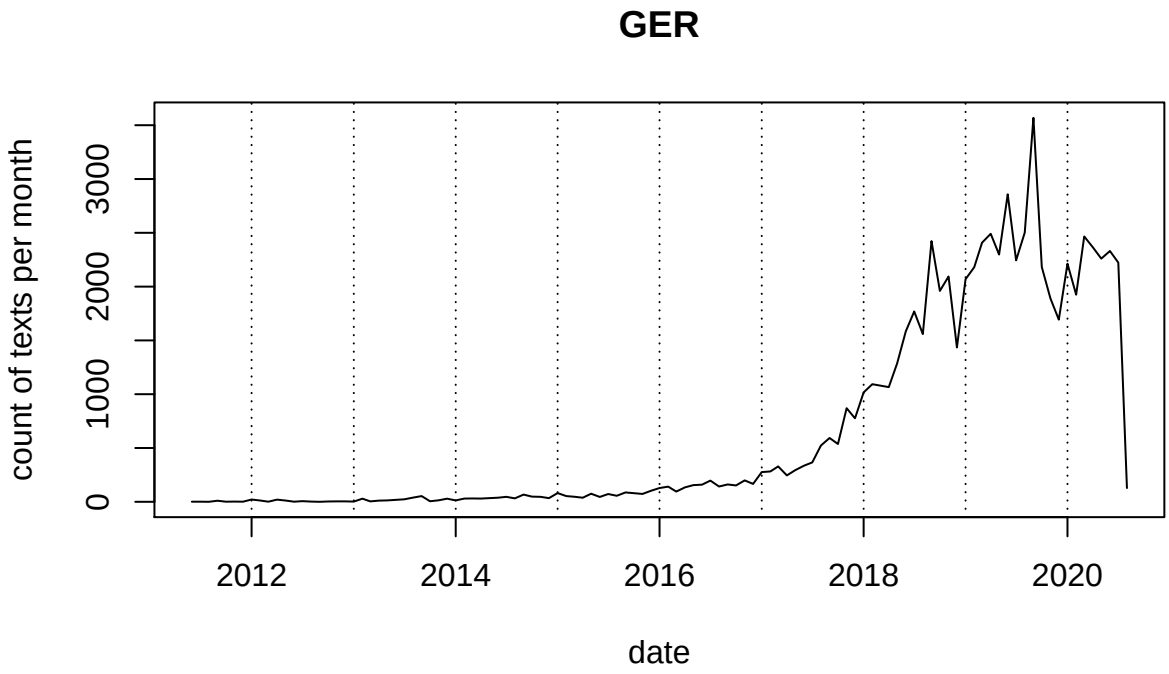

ESP
Lengths of tweeted texts
0
100
300
500
Cs
EAJ−PNV
ECP−GUAYEM EL CANVI
EH Bildu
ERC−CATSÍ
ERC−S
JxCat−JUNTS
PP
PSC−PSOE
PsdeG−PSOE
PSOE
UP
VOX

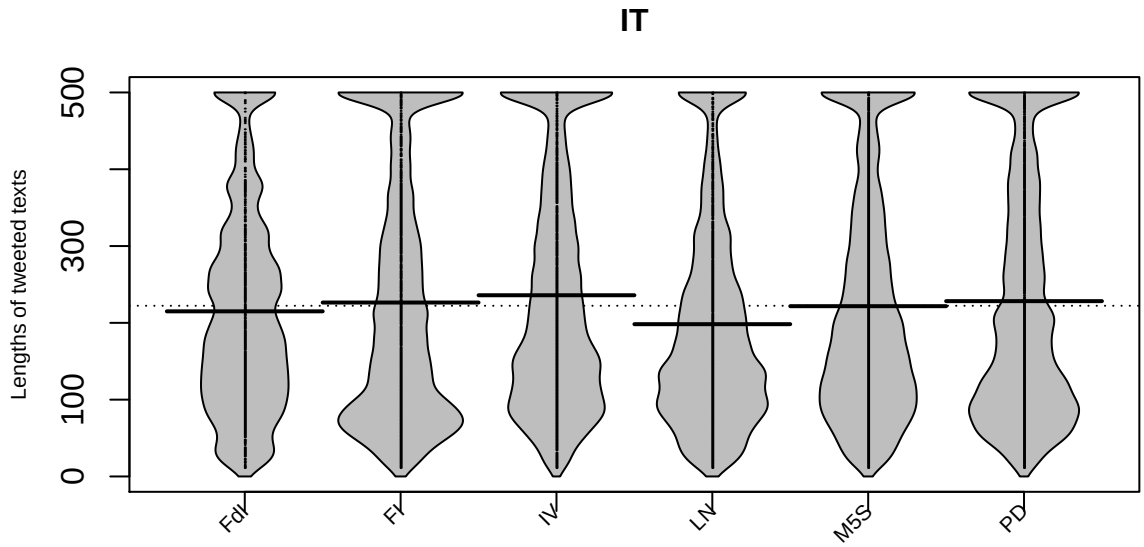

ESP
count of texts per month
0
1000
2000
3000
2012
2014
2016
2018
2020
date

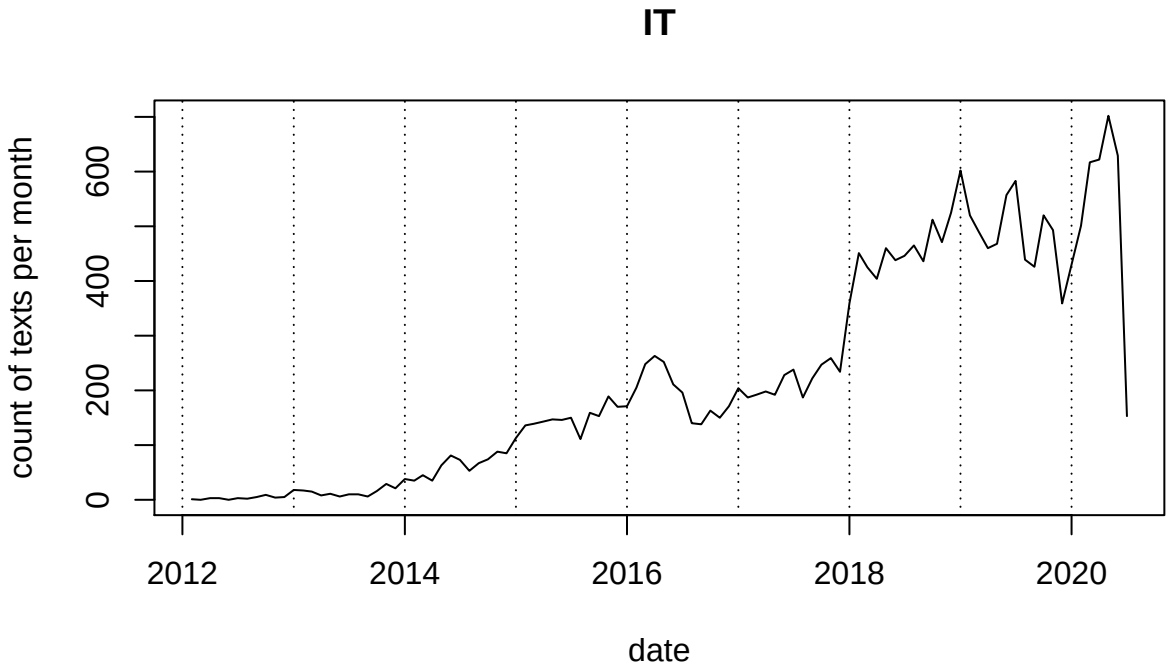

FR
Lengths of tweeted texts
0
100
300
500
Agir
DVG
FI
LR
LREM
MoDem
PCF
PS
RN
UDI

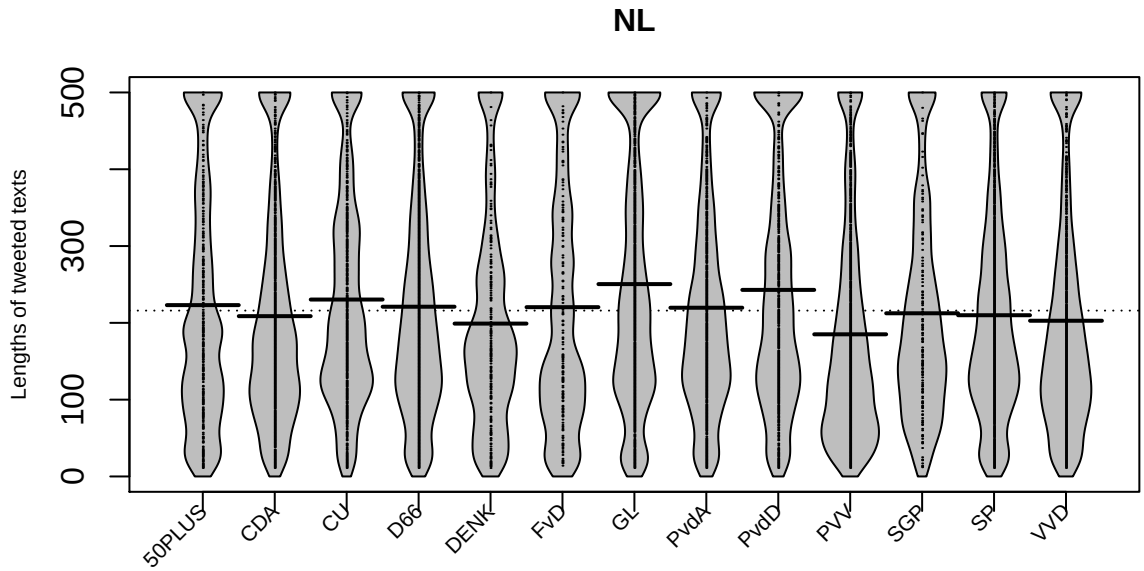

FR
count of texts per month
0
500
1500
2012
2014
2016
2018
2020
date

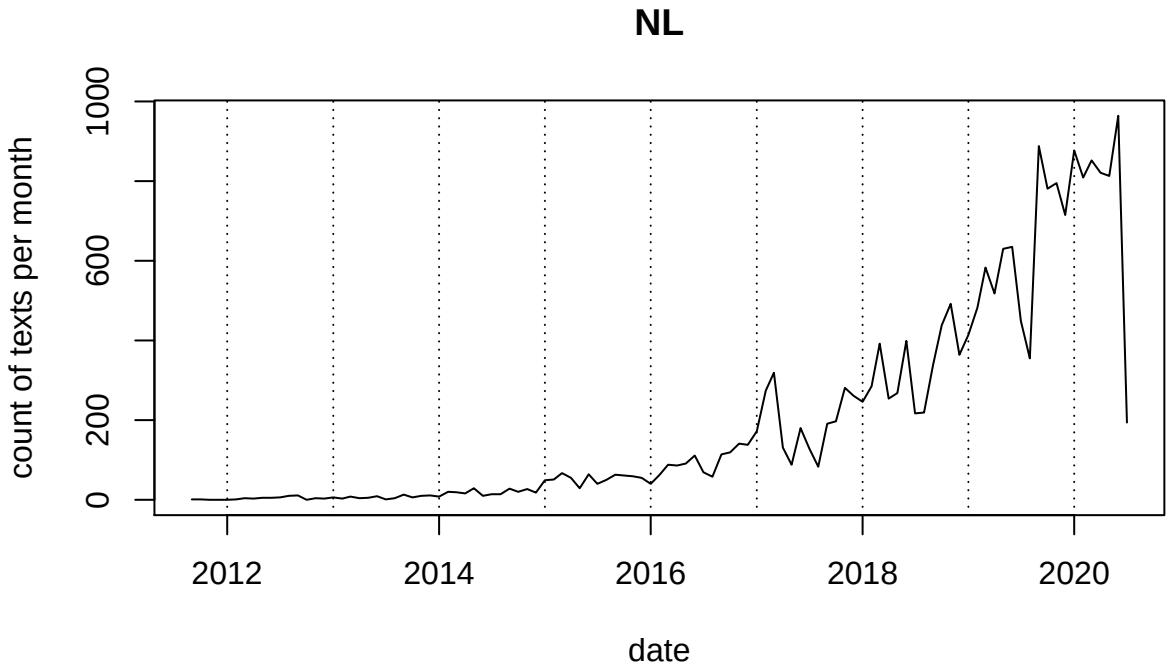

GER
Lengths of tweeted texts
0
100
300
500
AfD
CDU
CSU
FDP
GRÜNE
LINKE
SPD

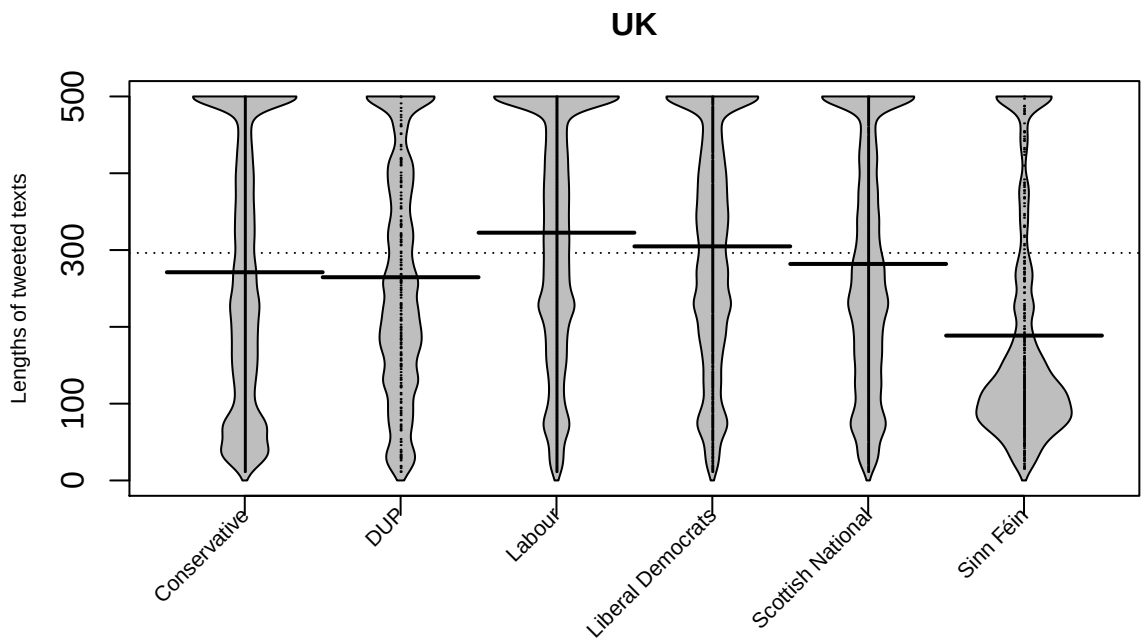

GER
count of texts per month
0
1000
2000
3000
2012
2014
2016
2018
2020
date

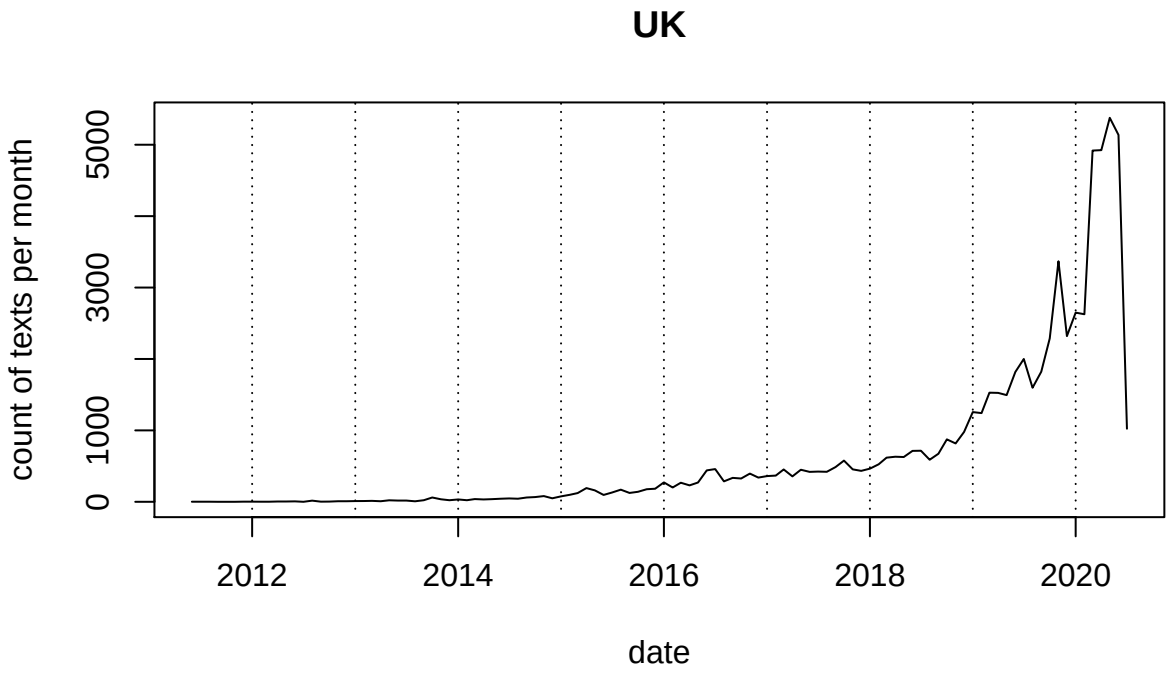

IT
Lengths of tweeted texts
0
100
300
500
FdI
FI
IV
LN
M5S
PD


IT
count of texts per month
0
200
400
600
2012
2014
2016
2018
2020
date


NL
Lengths of tweeted texts
0
100
300
500
50PLUS
CDA
CU
D66
DENK
FvD
GL
PvdA
PvdD
PVV
SGP
SP
VVD


NL
count of texts per month
0
200
600
1000
2012
2014
2016
2018
2020
date


UK
Lengths of tweeted texts
0
100
300
500
Conservative
DUP
Labour
Liberal Democrats
Scottish National
Sinn Féin


UK
count of texts per month
0
1000
3000
5000
2012
2014
2016
2018
2020
date