Vulgar language on the web across World Englishes - Part 2: data processing

Anonymous

2025-04-05

Contents

1 Intro **1**

2 data preparation **2**

3 loading data **2**

4 load kwics **5**
4.1 Lemma 7

5 Clean kwic **9**

6 Remove duplicate rows using dplyr **14**

7 Check KWICs **15**

8 Save KWIC **15**

9 Outro **16**

1 Intro

This document shows an analysis that was performed with the aim of finding differences in swearing across geographically distinct varieties of English around the world based on the GloWbE corpus.

install packages

```
# install packages
install.packages("tidyverse")
install.packages("quanteda")
install.packages("here")
install.packages("udpipe")
install.packages("future")
install.packages("furrr")
install.packages("stringi")
install.packages("parallel")
install.packages("usethis")
install.packages("data.table")
devtools::install_github("jimjam-slam/ggflags")
install.packages("ggflags")
```

load packages and set options

```
# load packages
library(data.table)
library(tidyverse)
library(quanteda)
library(here)
library(stringi)
library(parallel)
library(usethis)
library(data.table)
library(ggflags)
```

2 data preparation

3 loading data

raw

```
# paths to results
cdfs <- list.files(here::here("test"), pattern = "dfc.rda", full.names = T)
# load tables
lapply(cdfs, function(x) {
  x <- readRDS(x)
  }) -> lcdfs
# combine into a single df
data.table::rbindlist(lcdfs)   %>%
  # remove NAs
  dplyr::filter(txt != "") -> cdf
# clean
cdf  %>%
  dplyr::mutate(sfl = stringr::str_remove_all(sfl, ".txt.*")) %>%
  # rename columns
  dplyr::rename(corpus = crp,
                file = sfl,
                subfile = fl,
                rawtext = txt,
                cleantext = ctxt) -> df
# inspect clean dataframe
head(df)
```

```
##       corpus subfile     file
##       <char>  <char>   <char>
## 1: Australia   ##703 w_au_g01
## 2: Australia  ##1000 w_au_g01
## 3: Australia  ##1004 w_au_g01
## 4: Australia  ##1100 w_au_g01
## 5: Australia  ##1400 w_au_g01
## 6: Australia  ##1500 w_au_g01
##
##
## 1:
## 2:
being as a source of stress . Most of us typically perceive stress as a ... <p> In terms of losing weigh
```

```
## 3: ##1004 <h> Question : Can identical twins ever be boy/girl twins ? <p> In some rare cases , identi
up of a girl is XX . An XO baby is outwardly a girl , but her cells only have one copy of the X chromos
sex chromosomes and 1 sex chromosome ( commonly called the X or the Y chromosome ) <p> Each ' set ' of
question : What are chromosomes ? <p> Chromosomes can also be thought of as similar to a ball of wool -
chromosome is also responsible for differences in female identicals . When a female foetus is created ;
20 cells ) . If by chance one identical twin ' silences ' the X chromosome that came from Dad 's sperm
identical . This is due to something called ' chimerism ' - when an individual is composed of 2 genetica
identical twins are chimeras ; the rise in IVF is considered to be a contributing factor . A report has
copy , but some should be available as electronic versions -- it depends on what type of service the res
87 <p> Department of Psychology USA . This e-mail address is being protected from spambots . You need Ja
sex twin sample . The data were examined with reference to psychobiological and evolutionary perspectiv
2 months before the co-twin 's death , 1-2 months following the co-twin 's death and currently . A Grie
2 months following the loss , and currently . Information on physical symptoms was available from the So
twin ( retrospective twin group ) ; a second @ @ @ @ @ @ @ @ @ @ . Consistent with psychobiological the
mail address is being protected from spambots . You need JavaScript enabled to view it <p> Contemporary
the mourner 's memory and emotions . A complementary perspective is offered by archetypal psychology , w
mail address is being protected from spambots . You need JavaScript enabled to view it <p> We @ @ @ @ @
26 . <p> Department UK . This e-mail address is being protected from spambots . You need JavaScript enab
depth about their understanding and experience of twinship . Participants were selected who had a rich
year study was based on individual interviews of over 200 bereaved MZ and DZ adult twins . Its purpose w
analysis . <h> Question : Do you have any resources for twins who do n't get along with each other ? <p>
authored another book Uniting Psychology and Biology : Integrative Perspectives on Human Development wh
## 4:
third of the land mass . Yet when it comes time to doll out the money each year for the federal road bud
## 5:
good-reason Italian food can be found on the Croatian coast . <p> But because it has none of the hype th
season where prices are low , the weather is good , and you will find very few tourists roaming around
point , but no matter where you go in Buenos Aires , will you most likely turn to your travel companion
sized hole in your wallet . <h> Caribbean = Southern Thailand <p> A beautiful Thai beach on the island
## 6:
facing wall ( or North facing wall if in the Southern Hemisphere ) , which is painted black and made of
e glass - this is special glass which has a lower rate of heat transmission , similar in effect to doubl
14 inches for brick , 12-18 for concrete , 8-12 " for adobe or other earth material and at least 6 inche
e glass for a Trombe wall . Do you have data ? Even with the latest , high-transmittance versions of thi
potable @ @ @ @ @ @ @ @ @ @ . am considering all passive solar technologies , may mix with a micro photo
scaled : less water , less heat , less energy consumption . also looking for less complexity in activity
##
##
## 1:
## 2:
## 3: question can identical twins ever be boy girl twins in some rare cases identical twins from an egg
## 4:
## 5:
## 6:
```

create summary table

```
df  %>%
  dplyr::ungroup() %>%
  dplyr::mutate(# count number of words
    words = length(unlist(quanteda::tokenize_fastestword(cleantext))),
    # determine type of data (blog vs general web)
    type = dplyr::case_when(stringr::str_detect(file, "^w_") ~ "General Web",
                            TRUE ~ "Blog")) -> dfw
# save
```

```r
#base::saveRDS(dfw, file = here::here("test", "glowbe.rda"))
# inspect
head(dfw)
```

```
##       corpus subfile     file
##       <char>  <char>   <char>
## 1: Australia   ##703 w_au_g01
## 2: Australia  ##1000 w_au_g01
## 3: Australia  ##1004 w_au_g01
## 4: Australia  ##1100 w_au_g01
## 5: Australia  ##1400 w_au_g01
## 6: Australia  ##1500 w_au_g01
##
##
## 1:
## 2:
being as a source of stress . Most of us typically perceive stress as a ... <p> In terms of losing weigh
## 3: ##1004 <h> Question : Can identical twins ever be boy/girl twins ? <p> In some rare cases , identi
up of a girl is XX . An XO baby is outwardly a girl , but her cells only have one copy of the X chromos
sex chromosomes and 1 sex chromosome ( commonly called the X or the Y chromosome ) <p> Each ' set ' of
question : What are chromosomes ? <p> Chromosomes can also be thought of as similar to a ball of wool -
chromosome is also responsible for differences in female identicals . When a female foetus is created ;
20 cells ) . If by chance one identical twin ' silences ' the X chromosome that came from Dad 's sperm
identical . This is due to something called ' chimerism ' - when an individual is composed of 2 genetica
identical twins are chimeras ; the rise in IVF is considered to be a contributing factor . A report has
copy , but some should be available as electronic versions -- it depends on what type of service the res
87 <p> Department of Psychology USA . This e-mail address is being protected from spambots . You need Ja
sex twin sample . The data were examined with reference to psychobiological and evolutionary perspective
2 months before the co-twin 's death , 1-2 months following the co-twin 's death and currently . A Grie
2 months following the loss , and currently . Information on physical symptoms was available from the So
twin ( retrospective twin group ) ; a second @ @ @ @ @ @ @ @ @ @ . Consistent with psychobiological theo
mail address is being protected from spambots . You need JavaScript enabled to view it <p> Contemporary
the mourner 's memory and emotions . A complementary perspective is offered by archetypal psychology , w
mail address is being protected from spambots . You need JavaScript enabled to view it <p> We @ @ @ @ @
26 . <p> Department UK . This e-mail address is being protected from spambots . You need JavaScript enab
depth about their understanding and experience of twinship . Participants were selected who had a rich l
year study was based on individual interviews of over 200 bereaved MZ and DZ adult twins . Its purpose w
analysis . <h> Question : Do you have any resources for twins who do n't get along with each other ? <p
authored another book Uniting Psychology and Biology : Integrative Perspectives on Human Development whi
## 4:
third of the land mass . Yet when it comes time to doll out the money each year for the federal road bud
## 5:
good-reason Italian food can be found on the Croatian coast . <p> But because it has none of the hype th
season where prices are low , the weather is good , and you will find very few tourists roaming around
point , but no matter where you go in Buenos Aires , will you most likely turn to your travel companion
sized hole in your wallet . <h> Caribbean = Southern Thailand <p> A beautiful Thai beach on the island
## 6:
facing wall ( or North facing wall if in the Southern Hemisphere ) , which is painted black and made of
e glass - this is special glass which has a lower rate of heat transmission , similar in effect to doubl
14 inches for brick , 12-18 for concrete , 8-12 " for adobe or other earth material and at least 6 inche
e glass for a Trombe wall . Do you have data ? Even with the latest , high-transmittance versions of thi
potable @ @ @ @ @ @ @ @ @ @ . am considering all passive solar technologies , may mix with a micro photo
scaled : less water , less heat , less energy consumption . also looking for less complexity in activity
##
```

```
##
## 1:
## 2:
## 3: question can identical twins ever be boy girl twins in some rare cases identical twins from an egg
## 4:
## 5:
## 6:
##         words        type
##         <int>      <char>
## 1: 1766216633 General Web
## 2: 1766216633 General Web
## 3: 1766216633 General Web
## 4: 1766216633 General Web
## 5: 1766216633 General Web
## 6: 1766216633 General Web
```

extract basic information

```
# info table
dfw %>%
  dplyr::group_by(corpus, file, subfile, type) %>%
  dplyr::summarise(tokens = sum(words))  -> glowbe_table
```

```
## `summarise()` has grouped output by 'corpus', 'file', 'subfile'. You can
## override using the `.groups` argument.
```

```
# save
#base::saveRDS(glowbe_table, file = here::here("test", "glowbe_table.rda"))
# inspect
head(glowbe_table)
```

```
## # A tibble: 6 x 5
## # Groups:   corpus, file, subfile [6]
##   corpus    file     subfile  type            tokens
##   <chr>     <chr>    <chr>    <chr>            <dbl>
## 1 Australia w_au_g01 ##1000   General Web 1766216633
## 2 Australia w_au_g01 ##100100 General Web 1766216633
## 3 Australia w_au_g01 ##100101 General Web 1766216633
## 4 Australia w_au_g01 ##100303 General Web 1766216633
## 5 Australia w_au_g01 ##1004   General Web 1766216633
## 6 Australia w_au_g01 ##100400 General Web 1766216633
```

save and reload results

```
dfw <- base::readRDS(file = here::here("test", "glowbe.rda"))
glowbe_table <- base::readRDS(file = here::here("test", "glowbe_table.rda"))
```

4 load kwics

```
# paths to results
kwics <- list.files(here::here("test"), pattern = "_kwic_results.rda", full.names = T)
# inspect
kwics
```

```
##  [1] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Australia_kwic_results
##  [2] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Bangladesh_kwic_result
```

```
##  [3] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Canada_kwic_results.rd
##  [4] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/GBBlog_kwic_results.rd
##  [5] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/GBGeneral_kwic_results
##  [6] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Ghana_kwic_results.rda
##  [7] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/HongKong_kwic_results
##  [8] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/India_kwic_results.rda
##  [9] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Ireland_kwic_results.r
## [10] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Jamaica_kwic_results.r
## [11] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Kenya_kwic_results.rda
## [12] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Malaysia_kwic_results
## [13] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/NewZealand_kwic_resul
## [14] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Nigeria_kwic_results.r
## [15] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Pakistan_kwic_results
## [16] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Philippines_kwic_resul
## [17] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Singapore_kwic_results
## [18] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/SouthAfrica_kwic_resul
## [19] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/SriLanka_kwic_results
## [20] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/Tanzania_kwic_results
## [21] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/USBlog_kwic_results.rd
## [22] "C:/Users/Martin/OneDrive/Dokumente/University/UQ/VulgarityWorldWide/test/USGeneral_kwic_results
```

process kwics and combine into a single dataframe

```
# load tables
lapply(kwics, function(x) {
  x <- readRDS(x)
  }) -> kwicdfs
# combine into a single df
data.table::rbindlist(kwicdfs) -> kwicdf
# inspect
head(kwicdf)
```

```
##    docname                    pre keyword                            post
##     <char>                 <char>  <char>                          <char>
## 1:  text14     but to take a large     hoe            and a shovel also and
## 2:  text19   get within a monkey s    fart      of being a musician working
## 3:  text20      when i first saw a     jug            of water in an indian
## 4:  text20    10 59 am cleaning ur    butt          with water is much more
## 5:  text20      do the same to you    butt smearing is not cleaning washing
## 6:  text20 shelly do you wash your    butt           with your hands do you
##                                                                    pattern
##                                                                     <fctr>
## 1:                                                \\bh[oO0øöóõ][e3€ëéê][zs]*\\b
## 2:                                                  \\bfart(s|z|ing|in|ed)*\\b
## 3:                                                       \\bjug[g]*[s5$§z]*\\b
## 4: \\b[b8ß3][uµüúû][t7+†][t7+†][sz]*(face|head|wit|whipe|hole|h)*[hl]*[sz]*\\b
## 5: \\b[b8ß3][uµüúû][t7+†][t7+†][sz]*(face|head|wit|whipe|hole|h)*[hl]*[sz]*\\b
## 6: \\b[b8ß3][uµüúû][t7+†][t7+†][sz]*(face|head|wit|whipe|hole|h)*[hl]*[sz]*\\b
##       corpus
##       <char>
## 1: Australia
## 2: Australia
## 3: Australia
## 4: Australia
## 5: Australia
```

```
## 6: Australia
```

4.1 Lemma

annotate lemmas

```
kwicdf_annotated <- kwicdf %>%
  dplyr::mutate(lemma = dplyr::case_when(pattern ==  "\\b(dumb|stupid|lazy|worthless|useless|brain|dead
                                         pattern == "\\b(dumb|stupid|lazy|worthless|useless|brain|dead|
    pattern == "\\b[8ß|3][i1!|ïíì]+[a@4äáå]*[t7+†][c¢©(][h](es|ez|ing|ed)*\\b" ~ "bitch",
    pattern == "\\b[8ß|3][a@4äáå][s5$§z][t7+†][a@4äáå]r[d][o]*(s|z)*\\b" ~ "bastard",
    pattern == "\\b[8ß|3][e3€ëéê][a@4äáå]n[e3€ëéê]r(s|z)*\\b" ~ "beaner",
    pattern ==   "\\b[8ß|3][e3€ëéê]ll[e3€ëéê]nd(s|z)*\\b" ~ "bellend",
    pattern ==     "\\b[8ß|3][1!|ïíì]mb[oO0øöóõ](s|z)*\\b" ~ "bimbo",
    pattern == "\\bbl[oO0øöóõ]{2,}d[iy¥](ed)*\\b" ~ "bloody",
    pattern == "\\b[8ß|3][oO0øöóõ]ll[oO0øöóõïíì][xc¢©(k|<{(]+[s5$§z]?\\b" ~ "bollocks",
    pattern == "\\bboner[s]*\\b" ~ "boner",
    pattern == "\\b[8ß|3][oO0øöóõ]n[k|<{(](in|ing)*\\b" ~ "bonk",
    pattern == "\\b[8ß|3][oO0øöóõ]{2,}[b8ß|3][ie]*[s5$§z]?\\b" ~ "boobs",
    pattern == "\\b[8ß|3][u|µüúû]gg[e3€ëéê]r(ing|s|z)*\\b" ~ "bugger",
    pattern == "\\b[b8ß|3][u|µüúû]ll[s5$§z]h[1!|ïíì]+[t7+†]*\\b" ~ "bullshit",
    pattern == "\\b[b8ß3][uµüúû][t7+†][t7+†][sz]*(face|head|wit|whipe|hole|h)*[hl]*[sz]*\\b" ~ "butt(ho
    pattern == "\\b(god)*damn\\b" ~ "damn",
    pattern == "\\bdarki(es)*\\b" ~ "darkie",
    pattern == "\\b(bull)*d[iy]*ke(s|z)*\\b" ~ "dike",
    pattern == "\\bdildo(s|z)*\\b" ~ "dildo",
    pattern == "\\bdork(s|z)*\\b" ~ "dork",
    pattern == "\\beff(ing|in|ed|d)*[-._+ ]*[(you|up|off)]*\\b" ~ "eff",
    pattern == "\\bfann(y|ies)+\\b" ~ "fanny",
    pattern == "\\bfart(s|z|ing|in|ed)*\\b" ~ "fart",
    pattern == "\\bfrig(g|gin|ging|ged|gs)*\\b" ~ "frig",
    pattern == "\\b(cluster|head|mother|motha|mutha|mada|cock|mom|mum|daddy|father|sister|brother)*[f=f]
    pattern == "\\bf[-._+ ](me|you|it|this|that|the|these|those|him|her|us|them)+\\b" ~ "fuck",
    pattern == "\\b[f=f][c¢©(*k<{(uµüúû*]+\\b" ~ "fuck",
    pattern == "\\bgash\\b" ~ "gash",
    pattern == "\\bg[o]*ok(s|z)*\\b" ~ "gook",
    pattern == "\\bidiot(s|z)*\\b" ~ "idiot",
    pattern == "\\bjacka[s5$§z][s5$§z]\\b" ~ "jackass",
    pattern == "\\bjap[zs]*\\b" ~ "jap",

    pattern == "\\bjerk(s|z|in|ing|ed)*\\b" ~ "jerk",
    pattern == "\\bji[s5$§z][s5$§z]+\\b" ~ "jiss",
    pattern == "\\bjug[g]*[s5$§z]*\\b" ~ "jug",
    pattern == "\\b[(dip)]*[s5$§z]h[1!|ïíì]+[t7+†]*(ing|e|in|er|a|ed|ers|az|s|z)*\\b" ~ "shit",
    pattern == "\\bktfo\\b" ~ "online",
    pattern == "\\bstfu\\b" ~ "online",
    pattern == "\\bgtfo\\b" ~ "online",
    pattern == "\\bngaf\\b" ~ "online",
    pattern == "\\bdgaf\\b" ~ "online",
    pattern == "\\bffs\\b" ~ "online",
    pattern == "\\bfml\\b" ~ "online",
    pattern == "\\bomfg\\b" ~ "online",
    pattern == "\\baf\\b" ~ "online",
    pattern == "\\btf\\b" ~ "online",
```

```
    pattern == "\\bwtf\\b" ~ "online",
    pattern == "\\blmao\\b" ~ "online",
    pattern == "\\blmfao\\b" ~ "online",
    pattern == "\\brofl\\b" ~ "online",
    pattern == "\\bch[i]*nk[zs]*\\b" ~ "chink",
    pattern == "\\bcoon[zs]*\\b" ~ "coon",
    pattern == "\\b[bull]*crap(ping|ped|s|z|pin)*\\b" ~ "crap",
    pattern == "\\bcum(ming)*\\b" ~ "cum",
    pattern == "\\bc[oOøöóõ][c¢©]+[(k|<{(|x]+(suck|sak|suk)*[k]*(er|ers|a|az|as)*\\b" ~ "cock",
    pattern == "\\b[kc¢©(][u|µüúû]*nt[zs]*\\b" ~ "cunt",
    pattern == "\\b[d][1!|ïíì][c¢©(][xk|<{(][(head)]*[zs]*\\b" ~ "dick",
    pattern == "\\b[f|=f][a@4äáåe]g[g]*[ioa]*[t]*[zs]*\\b" ~ "fag(got)",

    pattern == "\\bh[oOøöóõ][e3€ëéê][zs]*\\b" ~ "hoe",
    pattern == "\\bh[oOøöóõ]r[e3€ëéê]*[zs]*\\b" ~ "whore",
    pattern == "\\b[k|<{(][i1!ïíìy][k|<{(][e3€ëéê][zs]*\\b" ~ "kike",
    pattern == "\\bn[i1!ïíì]gg[e3€ëéê|@a4äáå][r]*[zs]*\\b" ~ "nigger",
    pattern == "\\bknob[(head)]*[zs]*\\b" ~ "knob",
    pattern == "\\blesbo[sz]*\\b" ~ "lesbo",
    pattern == "\\bming[(a|er)]+(s|z)*\\b" ~ "minger" ,
    pattern == "\\bm[o|u]ron(ic|s|z)*\\b" ~ "moron",
    pattern == "\\bmuff\\b" ~ "muff",
    pattern == "\\bnonce\\b" ~ "nonce",
    pattern == "\\bnympho\\w*\\b" ~ "nympho",
    pattern == "\\bp[e3€ëéê]ck[ae]?[rs]*\\b" ~ "pecker",
    pattern == "\\bp[e3€ëéê]do[philf]*[e]*[zs]*\\b" ~ "pedo",
    pattern == "\\bpik(i|is|ey|ies|eys|eyz|iez|iz)+\\b" ~ "pikey",
    pattern == "\\bpimp(s|ing|in|z|ed)*\\b" ~ "pimp",
    pattern == "\\bp[i1!|ïíì][s5$§z][s5$§]+(in|ing|er|a|ers|erz)*\\b" ~ "piss",
    pattern == "\\bpooft(er|ers|as|az)+\\b" ~ "poofter",
    pattern == "\\bprick[zs]*\\b" ~ "prick",
    pattern == "\\bpuk[e]*(s|z|ing|ed)*\\b" ~ "puke",
    pattern == "\\bp[u|µüúû][s5$§z][s5$§z][@a4äáå]*[y¥][zs]*\\b" ~ "pussy",
    pattern == "\\bqu[e]+[a]*f(s|z|ing|ed)*\\b" ~ "queef",
    pattern == "\\bscr[e]+w(ing|ed|s|z)*\\b" ~ "screw",
    pattern == "\\bshag(ging|gin|ged)*\\b" ~ "shag",
    pattern == "\\bskank[yzs]*\\b" ~ "skank",
    pattern == "\\bslag[zs]*\\b" ~ "slag" ,
    pattern == "\\b[s5$§z]l[uµüúû][t7+†](i|y)*[zs]*\\b" ~ "slut",
    pattern == "\\bsod[d]*[sz]*(ing)*\\b" ~ "sod",
    pattern == "\\bspast(ic|ics|icz)*\\b" ~ "spastic",
    pattern == "\\bretard[zs]*\\b" ~ "retard",
    pattern == "\\btit[t]*(i|ies|ay|ays|ayz)*\\b" ~ "tits",
    pattern == "\\btosser[sz]*\\b" ~ "tosser",
    pattern == "\\btr[a@4äáå]nn(y|ies|iez)+\\b" ~ "tranny",
    pattern == "\\bturd[sz]*\\b" ~ "turd",
    pattern == "\\b[t7+†]w[a@4äáå][t7+†][zs]*\\b" ~ "twat",
    pattern == "\\bw[a@4äáå]n[k|<{(](z|er|ers|ing|az|a|ed)*\\b" ~ "wank",
    pattern == "\\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b" ~ "whore",
    pattern == "\\b[f|=f][-._+ ]*[u|µüúû][-._+ ]*[c¢©(][-._+ ]*[k|<{(][-._+ ]*[s5$§z]*(ing|in|ed)*\\b"
    pattern == "\\b[s5$§z][-._+ ]*[h][-._+ ]*[i1!|ïíì]+[-._+ ]*[t7+†][e3€ëéê]*\\b" ~ "shit",
    pattern == "\\b[a@4äáå][-._+ ]*[s5$§z][-._+ ]*[s5$§z]\\b" ~ "ass",
    pattern == "\\b[f=f][c¢©(*k<{(uµüúû*]+(ing|er|a|ed|ers|az)*\\b" ~ "fuck",
```

```
    pattern == "\\bf\\b" ~ "fuck",
    pattern == "\\beff(ing|in|ed|d)*\\b" ~ "eff",
T ~ pattern))
# inspect
names(table(kwicdf_annotated$lemma))
```

```
##  [1] "arse(hole)" "ass(hole)"  "bastard"    "bitch"      "bloody"
##  [6] "bollocks"   "boner"      "boobs"      "bullshit"   "butt(hole)"
## [11] "chink"      "cock"       "coon"       "crap"       "cum"
## [16] "cunt"       "damn"       "darkie"     "dick"       "dike"
## [21] "dildo"      "dork"       "eff"        "fag(got)"   "fanny"
## [26] "fart"       "frig"       "fuck"       "gash"       "gook"
## [31] "hoe"        "idiot"      "jackass"    "jap"        "jerk"
## [36] "jiss"       "jug"        "kike"       "knob"       "lesbo"
## [41] "minger"     "moron"      "muff"       "nigger"     "nonce"
## [46] "nympho"     "online"     "pecker"     "pedo"       "pikey"
## [51] "pimp"       "piss"       "poofter"    "prick"      "puke"
## [56] "pussy"      "queef"      "retard"     "shag"       "shit"
## [61] "skank"      "slag"       "slut"       "sod"        "spastic"
## [66] "tits"       "tosser"     "tranny"     "turd"       "twat"
## [71] "wank"       "whore"
```

check kwics

```
#kwicdf %>%   dplyr::filter(stringr::str_detect(keyword, "\\bjerk"))
names(table(kwicdf_annotated$keyword))
```

5 Clean kwic

```
kwicdf_clean <- kwicdf_annotated %>%
  # Retain rows where 'lemma' is 'hoe' and 'pre' ends with specified phrases
  filter(!(lemma == "hoe" & !str_detect(pre, "(such a( \\w+)?|other|is a|that( \\w+)?)\\b$")))


kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'keyword' contains specific unwanted patterns
  filter(!str_detect(keyword, "fak(r|er|ing|ers|ed|e)+")) %>%
  filter(!str_detect(keyword, "fk(r|er|ing|ers|ed|e)+")) %>%
  filter(!str_detect(keyword, "\\w\\+f$")) %>%
  filter(!str_detect(keyword, "n\\|f")) %>%
  filter(!str_detect(keyword, "^\\+f$")) %>%
  filter(!str_detect(keyword, "=(x|a|r|v|c|k|f)")) %>%
  filter(!str_detect(keyword, "(x|a|r|v|c|k|f)=")) %>%
  filter(!str_detect(keyword, "feg(i|o)")) %>%
  filter(!str_detect(keyword, "^feek.*")) %>%
  filter(!str_detect(keyword, "(x|a|r|v|c|k|f)\\+")) %>%
  filter(!str_detect(keyword, "\\+(x|a|r|v|c|k|f)"))  %>%
  dplyr::filter(!str_detect(keyword, "^\\$")) %>%
  dplyr::filter(!str_detect(keyword, "\\d{3}")) %>%
  dplyr::filter(!str_detect(keyword, "\\d{2}=")) %>%
  dplyr::filter(!str_detect(keyword, "fag(a|o)+")) %>%
  dplyr::filter(!str_detect(keyword, "fauk\\w*")) %>%
  dplyr::filter(!str_detect(keyword, "hoess")) %>%
```

```
  dplyr::filter(!str_detect(keyword, "z=u")) %>%
  dplyr::filter(!str_detect(keyword, "\\+jug")) %>%
  dplyr::filter(!str_detect(keyword, "2=u")) %>%
  dplyr::filter(!str_detect(keyword, "b=ok")) %>%
  dplyr::filter(!str_detect(keyword, "st=ok")) %>%
  dplyr::filter(!str_detect(keyword, "stat=u")) %>%
  dplyr::filter(!str_detect(keyword, "t=0k")) %>%
  dplyr::filter(!str_detect(keyword, "^feak\\*")) %>%
  dplyr::filter(!str_detect(keyword, "^fc+$")) %>%
  dplyr::filter(!str_detect(keyword, "^f[c]+a.*$")) %>%
  dplyr::filter(!str_detect(keyword, "^fauc.*$")) %>%
  dplyr::filter(!str_detect(keyword, "^f\\$\\d+$"))

kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'lemma' is 'jerk' and 'pre' ends with 'knee' or 'chest'
  filter(!(lemma == "jerk" & str_detect(pre, "(knee|chest)$"))) %>%

  # Retain rows where 'lemma' is 'jerk' and:
  # - 'post' starts with 'off', or
  # - 'pre' ends with 'this', 'that', 'such a', or 'is a'
  filter(!(lemma == "jerk" &
           !(str_detect(post, "^off\\b") | str_detect(pre, "(this|that|such a|is a)\\b$"))))


kwicdf_clean <- kwicdf_clean %>%
  # Retain rows where 'lemma' is 'eff' and 'post' starts with 'you', 'it', 'off', or 'up'
  filter(!(lemma == "eff" & !str_detect(post, "^(you|it|off|up)")))

kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'lemma' is 'f' unless 'post' starts with specified pronouns or similar words or p
  filter(!(keyword == "f" & !str_detect(post, "^(king|me|you|it|this|that|these|those|him|her|us|them)\

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'lemma' is 'crap' and 'pre' ends with specified phrases
  filter(!(lemma == "crap" & str_detect(pre, "(metal|poker|roulette)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'lemma' is 'dyke' and 'pre' does not end with specified phrases
  filter(!(lemma == "dyke" & !str_detect(pre, "\\b(a|like|butch|bull|all|fags|as|fucking|club)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'lemma' is 'faggot' and 'pre' ends with specified phrases
  filter(!(keyword == "faggot" & str_detect(post, "(bearer)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'keyword' is 'fag' and 'pre'  ends with specified phrases
  filter(!(keyword == "fags" & str_detect(pre, "\\b(of|and|buy|the|few|candy)\\b$") | str_detect(post,

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'keyword' is 'fk' and 'pre' ends with specified phrases
  filter(!(keyword == "fk" & str_detect(pre, "(bryne)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
```

```
  # remove rows where 'keyword' is 'jiss' and 'pre' ends with specified phrases
  filter(!(keyword == "jizz" & str_detect(pre, "\\b(dns)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'keyword' is 'fu' and 'pre' ends with specified phrases
  filter(!(keyword == "fu" & str_detect(pre, "(k(ou)ng)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'keyword' is 'knob' and 'post' does not end with specified phrases
  filter(!(lemma == "knob" & !str_detect(post, "^\\b(head[s]?)\\b")))

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'keyword' is 'pecker' and 'pre' ends with specified phrases
  filter(!(keyword == "pecker" & str_detect(pre, "\\b(a|wood|of|so)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # Retain rows where 'lemma' is 'sod' and 'pre' ends with a xxx xxx
  filter(!(lemma == "sod" & !str_detect(pre, "\\ba[n]{0,1}( \\w+)?( \\w+)?$")))

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'keyword' is 'sodd' and 'post' starts with specified phrases
  filter(!(keyword == "sodd" & str_detect(post, "^\\b(s)\\b")))

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'keyword' is 'sodd' and 'pre' ends with specified phrases
  filter(!(keyword == "sods" & str_detect(pre, "\\band\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # Remove rows where 'keyword' is 'tf' and 'pre' does not end with specified phrases
  filter(!(lemma == "tf" & !str_detect(pre, "\\b(how)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # Retain rows where 'keyword' is 'wanka' and 'pre' ends with specified phrases
  filter(!(keyword == "wanka" & !str_detect(pre, "(such a( \\w+)?|other|is a|that( \\w+)?)\\b$")))

kwicdf_clean <- kwicdf_clean %>%
  # remove rows where 'keyword' is 'whor' and 'pre' ends with specified phrases
  filter(!(keyword == "whor" & str_detect(pre, "\\b(casteism)\\b$")))

# inspect
kwicdf_clean[which(kwicdf_clean$keyword == "whrs"),]
```

```
##        docname                                          pre keyword
##         <char>                                       <char>  <char>
##  1:  text50260                       low waist to hip ration    whrs
##  2: text134941                  women have much lower facial    whrs
##  3: text134941               of more aggressive higher facial    whrs
##  4:  text15331                    headlines and claims to the    whrs
##  5:  text15331               initially persisted in using some    whrs
##  6:  text15331                       in another case where the    whrs
##  7:  text15331                          call went out for the    whrs
##  8:  text15331                        fap claim a good quality    whrs
##  9:  text29164      the weathertight homes resolution service    whrs
## 10:  text29165                    reports the purpose of the    whrs
```

```
## 11:  text39061        the weathertight homes resolution services    whrs
## 12:  text39061                            lodge a claim under the    whrs
## 13:  text42373              building surveyors council staff and    whrs
## 14:  text42373                     failure around 95 of eligible    whrs
## 15:  text42373                          history to emerge in the    whrs
## 16:  text42373                            lodge a claim with the    whrs
## 17:  text42865         of weathertight homes resolution services    whrs
## 18:  text42865          a weathertight homes resolution services    whrs
## 19:  text53326                            of lodging the claim or    whrs
## 20:  text53326                            or have been subject to    whrs
## 21:  text53326         the weathertight homes resolution service    whrs
## 22:  text53326           1 weathertight homes resolution service    whrs
## 23:  text53326                 homes resolution service whrs the    whrs
## 24:  text53326                           arrive at a decision the    whrs
## 25:  text53326                              the time of seeking a    whrs
## 26:  text53326                           be quicker with both the    whrs
## 27:  text53326                              cons of the court and    whrs
## 28:  text53326                        noted 1 when following the    whrs
## 29:  text53326                             3 both the court and    whrs
## 30:  text53326         weathertight homes resolution service the    whrs
## 31:  text53326            delivers the potentially lowest return    whrs
## 32:  text53326                           duplex 2 units low cost    whrs
## 33:  text53326           associated claims processes contact the    whrs
## 34:  text53326 resolutions weathertightness issues identified by    whrs
## 35:  text53326                       resolution based solely on a    whrs
## 36:  text53326                              to obtain a low cost    whrs
## 37:  text55320                           second one is that the    whrs
## 38:  text55321                             into the realms of a    whrs
## 39:  text65266        the weathertight homes resolution services    whrs
## 40:  text78538                            or have been subject to    whrs
## 41:  text78539                            or have been subject to    whrs
## 42:  text81418                       damage from water leaks the    whrs
## 43:  text81418                         on the processing of your    whrs
## 44:  text81418                       claim in particular if your    whrs
##        docname                                             pre keyword
##                                          post
##                                         <char>
##  1:   duke university researchers john graham
##  2:                       and do not have the
##  3:             feel more powerful we propose
##  4: weathertight homes resolution service and
##  5:               assessor s reports which it
##  6:         assessor s report had recommended
##  7:                assessor to return to site
##  8:               assessor s report is needed
##  9:                    does not take away the
## 10:                  act is to provide speedy
## 11:                     act 2006 a leaky home
## 12:                   act we can arrange for
## 13:          assessors the experts were asked
## 14:                   claims so far come from
## 15:                  and the courts there has
## 16:                from the time the dwelling
## 17:        decisions where the claim involved
```

```
## 18:         decision where the claim involved
## 19:             application or a claim should
## 20:       claims however properties that have
## 21:                   or with the district or
## 22:                the whrs offers two claims
## 23:             offers two claims processes 1
## 24:                   and wht path is claimed
## 25:             resolution the process may be
## 26:                   and court paths you can
## 27:          processes a the following points
## 28:                      path to arrive at a
## 29:         processes are often intervened by
## 30:                    offers a low cost path
## 31:  assessments affected homeowners who meet
## 32:            assessments are not speedy and
## 33:            37 845 toolong wht resolutions
## 34:      reports can progress to adjudication
## 35:                   report is high risk the
## 36:        assessment report and then contact
## 37:                    assessor does n t work
## 38:     or other weathertightness report such
## 39:                  act defines a leaky home
## 40:          claims from that date properties
## 41:          claims from that date properties
## 42:                    can provide a means of
## 43:               claim in particular if your
## 44:                  claim is either going to
##                                         post
##                                                pattern      corpus  lemma
##                                                 <fctr>      <char> <char>
##  1: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b      Canada  whore
##  2: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b  GB-General  whore
##  3: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b  GB-General  whore
##  4: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
##  5: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
##  6: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
##  7: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
##  8: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
##  9: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 10: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 11: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 12: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 13: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 14: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 15: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 16: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 17: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 18: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 19: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 20: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 21: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 22: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 23: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 24: \\b(cam|man|m)*wh[oOøöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
```

```
## 25: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 26: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 27: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 28: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 29: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 30: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 31: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 32: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 33: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 34: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 35: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 36: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 37: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 38: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 39: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 40: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 41: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 42: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 43: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
## 44: \\b(cam|man|m)*wh[oO0øöóõ]?r[e3€ëéê]*(d|s|z|ing)*\\b New Zealand  whore
##                                                pattern      corpus  lemma
```

```
#names(table(kwicdf_clean$keyword))
```

Bulk exclusion after manual check

```
exclude <- c("af", "arsel", "arzel", "assle", "asslee", "azzl", "ba$turd", "ba$turd$", "bolivares=u", "
kwicdf_clean <- kwicdf_clean %>%
  filter(!keyword %in% exclude)
```

6 Remove duplicate rows using dplyr

```
nrow(kwicdf_clean)
```

```
## [1] 466336
```

```
kwicdf_clean <- kwicdf_clean %>% dplyr::select(-pattern)
kwicdf_clean <- kwicdf_clean %>% dplyr::distinct()
nrow(kwicdf_clean)
```

```
## [1] 385133
```

check kwics

```
kwicdf_clean[which(kwicdf_clean$keyword == "jizz"),]
```

```
##         docname                                 pre keyword
##          <char>                              <char>  <char>
##   1:   text1614                    to be used as a    jizz
##   2:  text10093         arrived in with the cheese    jizz
##   3:  text55615              to 2 million makes me    jizz
##   4:  text61854              s every chance you ll    jizz
##   5:  text62060  hebraphrenia and teased out brain    jizz
##  ---
## 137: text129707 titty photographed with air drying    jizz
```

```
## 138: text154267     of his descendents through his    jizz
## 139: text157739              be the same people to    jizz
## 140: text159743        ever answered the door with    jizz
## 141: text165352           you have a little odumbo    jizz
##                            post     corpus  lemma
##                           <char>     <char> <char>
##   1: spittoon i think maybe that  Australia   jiss
##   2:  leaving some kind of sweet  Australia   jiss
##   3:        in my pants lol also  Australia   jiss
##   4:    your balls inside out at  Australia   jiss
##   5:          which i m going to  Australia   jiss
##  ---
## 137:  and therefore sort of sent US-General   jiss
## 138:    no the adam from genesis US-General   jiss
## 139:      all over this news and US-General   jiss
## 140:       in his beard now that US-General   jiss
## 141:      on it still your knees US-General   jiss
```

7 Check KWICs

```
kwicdf_clean %>% filter(corpus == "Malaysia", lemma == "online")
```

```
##        docname                              pre keyword
##         <char>                           <char>  <char>
##   1:     text1         but i somehow pity them     wtf
##   2:   text289          look up at others haha     wtf
##   3:   text289 halfway through the parade haha     wtf
##   4:   text289    utter any words anymore haha     wtf
##   5:   text290 and my different department lol     wtf
##  ---
## 570: text43977        did a couple tracks with   lmfao
## 571: text44521         night with a nokia c301     wtf
## 572: text44521         then we were shooed off   lmfao
## 573: text44523             i had to go through    lmao
## 574: text44524  bro s girlfriend when suddenly   lmfao
##                               post   corpus  lemma
##                              <char>   <char> <char>
##   1:       are they your money and Malaysia online
##   2:       i guess they use nippon Malaysia online
##   3:  another rojak group not sure Malaysia online
##   4:     we decided to walk around Malaysia online
##   5:         this gadget is my new Malaysia online
##  ---
## 570:         and they are just far Malaysia online
## 571:           and i tried so hard Malaysia online
## 572: because we support shaira mae Malaysia online
## 573:                 i did n t get Malaysia online
## 574:     strangers sat on my table Malaysia online
```

8 Save KWIC

```
base::saveRDS(kwicdf_clean, file = here::here("test", "kwic_results_clean.rda"))
```

9 Outro

```
sessionInfo()
```

```
## R version 4.4.2 (2024-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26100)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: Australia/Brisbane
## tzcode source: internal
##
## attached base packages:
## [1] parallel  stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] ggflags_0.0.4     usethis_3.1.0     stringi_1.8.7     here_1.0.1
##  [5] quanteda_4.2.0    lubridate_1.9.4   forcats_1.0.0     stringr_1.5.1
##  [9] dplyr_1.1.4       purrr_1.0.4       readr_2.1.5       tidyr_1.3.1
## [13] tibble_3.2.1      ggplot2_3.5.1     tidyverse_2.0.0   data.table_1.17.0
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.4        generics_0.1.3    lattice_0.22-6    hms_1.1.3
##  [5] digest_0.6.37     magrittr_2.0.3    evaluate_1.0.3    grid_4.4.2
##  [9] timechange_0.3.0  bookdown_0.42     fastmap_1.2.0     rprojroot_2.0.4
## [13] Matrix_1.7-1      stopwords_2.3     scales_1.3.0      cli_3.6.4
## [17] rlang_1.1.5       munsell_0.5.1     withr_3.0.2       yaml_2.3.10
## [21] tools_4.4.2       tzdb_0.5.0        colorspace_2.1-1  fastmatch_1.1-6
## [25] vctrs_0.6.5       R6_2.6.1          lifecycle_1.0.4   fs_1.6.5
## [29] pkgconfig_2.0.3   pillar_1.10.1     gtable_0.3.6      glue_1.8.0
## [33] Rcpp_1.0.14       xfun_0.52         tidyselect_1.2.1  rstudioapi_0.17.1
## [37] knitr_1.50        htmltools_0.5.8.1 rmarkdown_2.29    compiler_4.4.2
```