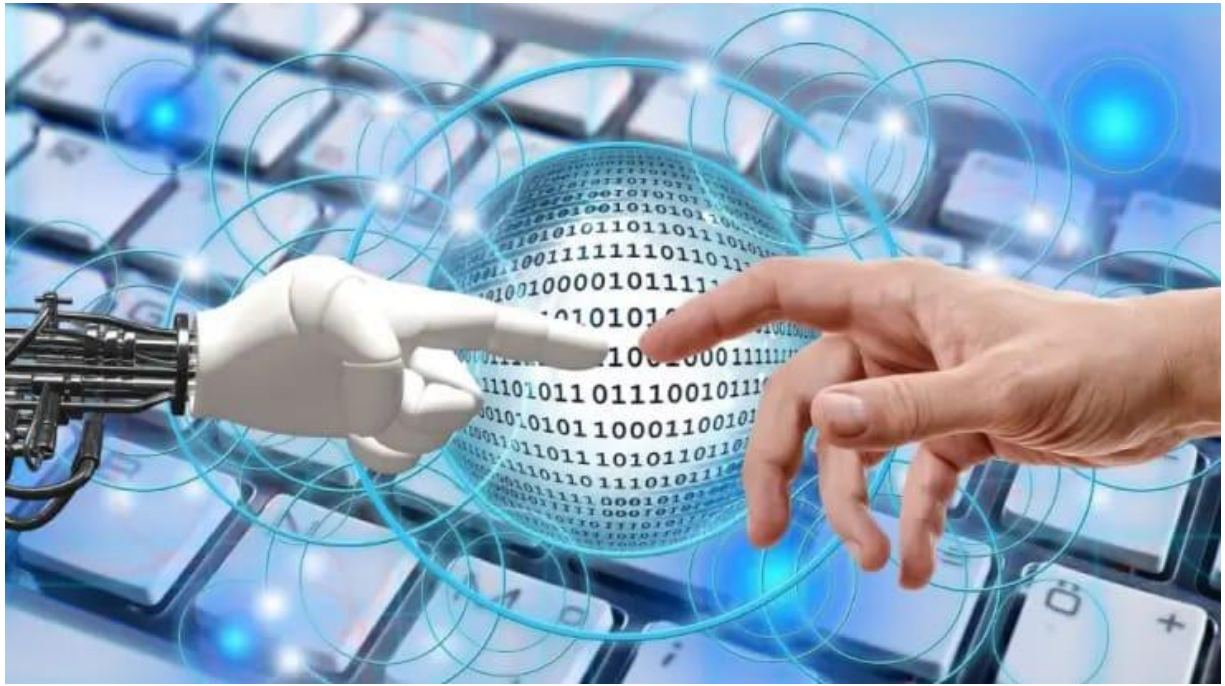
Looking from a Distance

An approach to text analysis with R by the Digital Lab and the JRAAS

Agenda

- Building the research problem
- Diving into technicality
- Outputs, new inputs
- A maelstrom



Lyman Tower Sargent Bibliography

ARUS

Permanent URI for this collection https://cetapsrepository.letras.up.pt/id/cetaps/94026

Welcome to the ARUS Digital Repository of the University of Porto.

With over 20,000 entries compiled by Lyman Tower Sargent, this already is the major resource of critical bibliography on utopia, but we expect it to grow with your contribution. Our aim is to have a complete list of critical sources in any language, add digital objects such as abstracts and full texts, and make the database searchable by subject. For advanced search click here .

For authors of titles listed here, authors of new titles, and volunteers that would like to become contributing editors, please click here.

To report on typos or errors please click here or click "Contact-us" on the bottom menu.

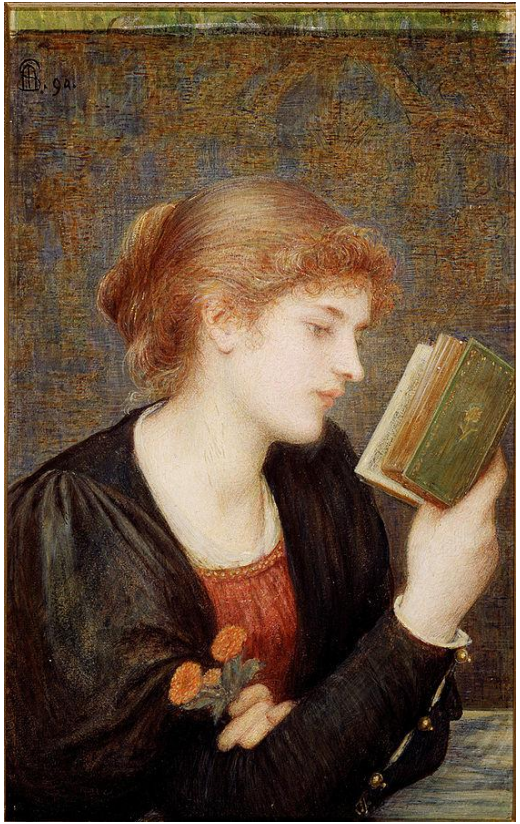
Join the Advanced Research in Utopian Studies (ARUS) community.

News

This September 2024, we added 453 new works to the repository. Thank you to all contributors1

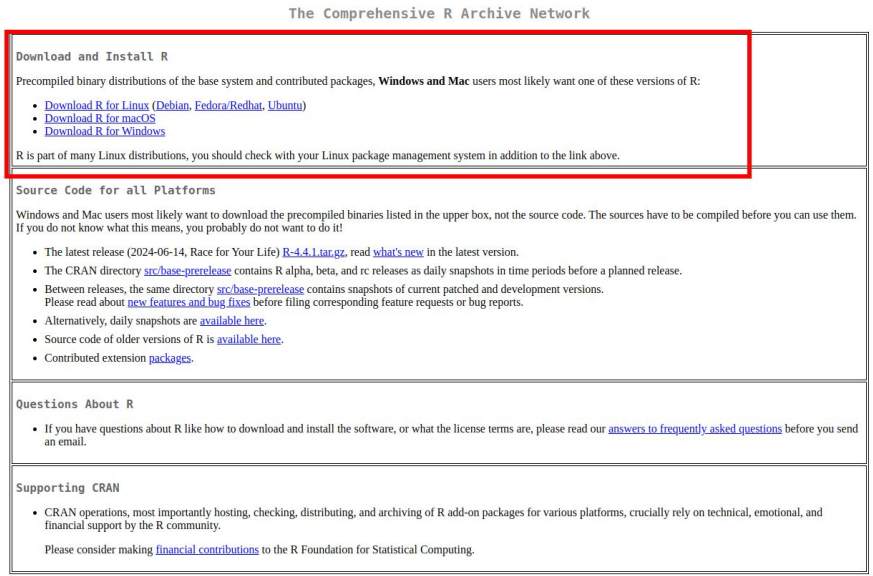
Search

Search this collection... Search



Download R language and R Studio

https://cran.r-project.org/



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux (Debian, Fedora/Redhat, Ubuntu)
- Download R for macOS
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2024-06-14, Race for Your Life) R-4.4.1.tar.gz, read what's new in the latest version.
- The CRAN directory src/base-prerelease contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory src/base-prerelease contains snapshots of current patched and development versions. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Alternatively, daily snapshots are available here.
- Source code of older versions of R is available here.
- Contributed extension packages.

Questions About R

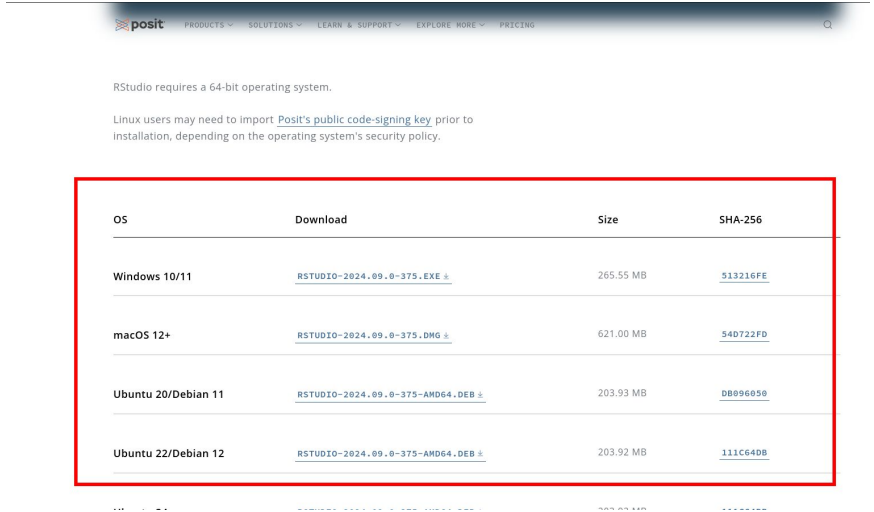
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

Supporting CRAN

- CRAN operations, most importantly hosting, checking, distributing, and archiving of R add-on packages for various platforms, crucially rely on technical, emotional, and financial support by the R community.

  Please consider making financial contributions to the R Foundation for Statistical Computing.

https://posit.co/download/rstudio-desktop/



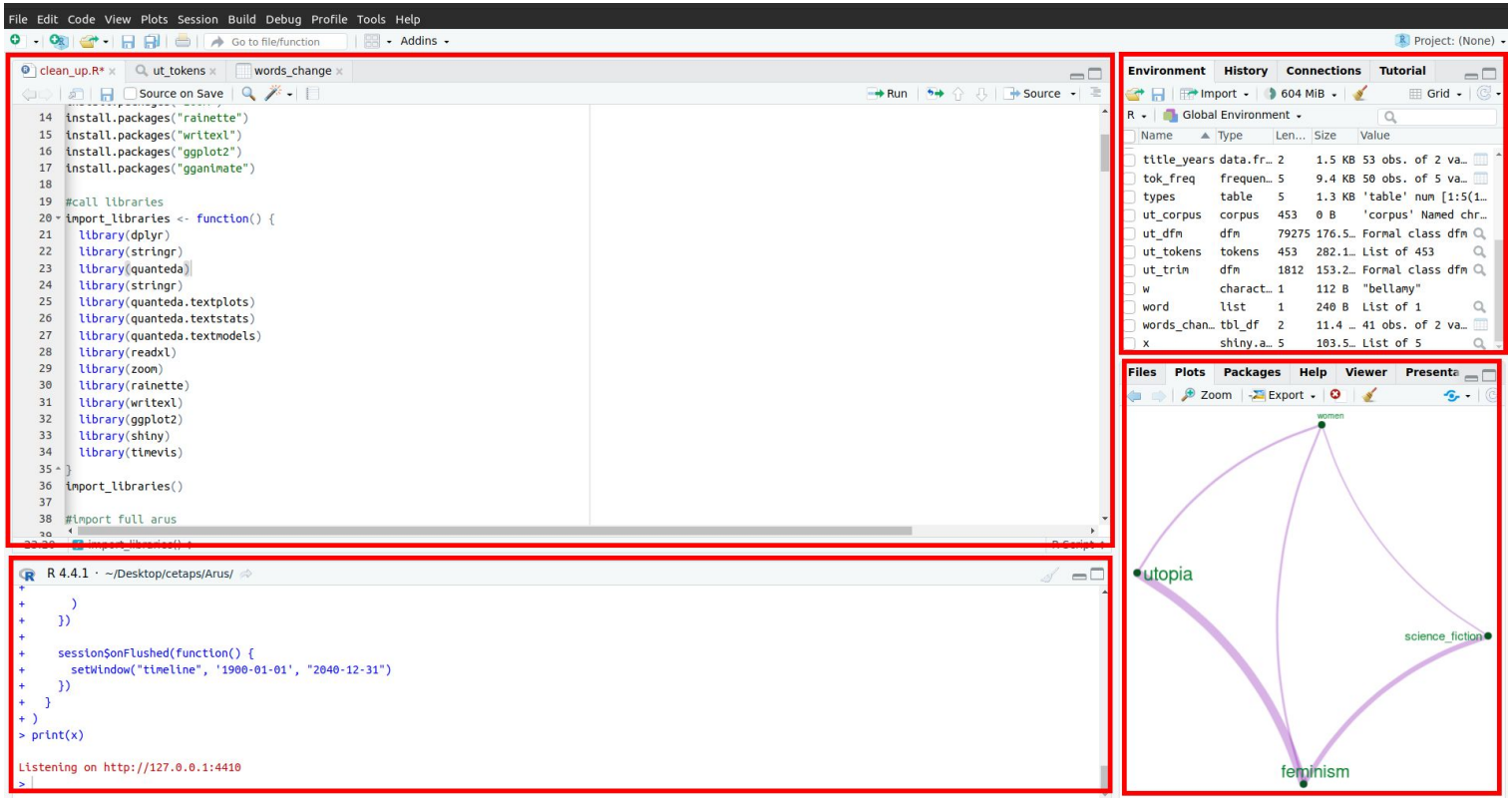
posit PRODUCTS SOLUTIONS LEARN & SUPPORT EXPLORE MORE PRICING

RStudio requires a 64-bit operating system.

Linux users may need to import Posit's public code-signing key prior to installation, depending on the operating system's security policy.

OS	Download	Size	SHA-256
Windows 10/11	RSTUDIO-2024.09.0-375.EXE	265.55 MB	513216FE
macOS 12+	RSTUDIO-2024.09.0-375.DMG	621.00 MB	54D722FD
Ubuntu 20/Debian 11	RSTUDIO-2024.09.0-375-AMD64.DEB	203.93 MB	DB096050
Ubuntu 22/Debian 12	RSTUDIO-2024.09.0-375-AMD64.DEB	203.92 MB	111C64DB

R studio structure



Install packages and import libraries

```
remove(list = ls())
setwd("/home/felipe/Desktop/cetaps/Arus")

#install packages
install.packages("dplyr")
install.packages("stringr")
install.packages("quanteda")
install.packages("stringr")
install.packages("quanteda.textplots")
install.packages("quanteda.textstats")
install.packages("quanteda.textmodels")
install.packages("readxl")
install.packages("zoom")
install.packages("rainette")
install.packages("writexl")
install.packages("ggplot2")
install.packages("gganimate")
```

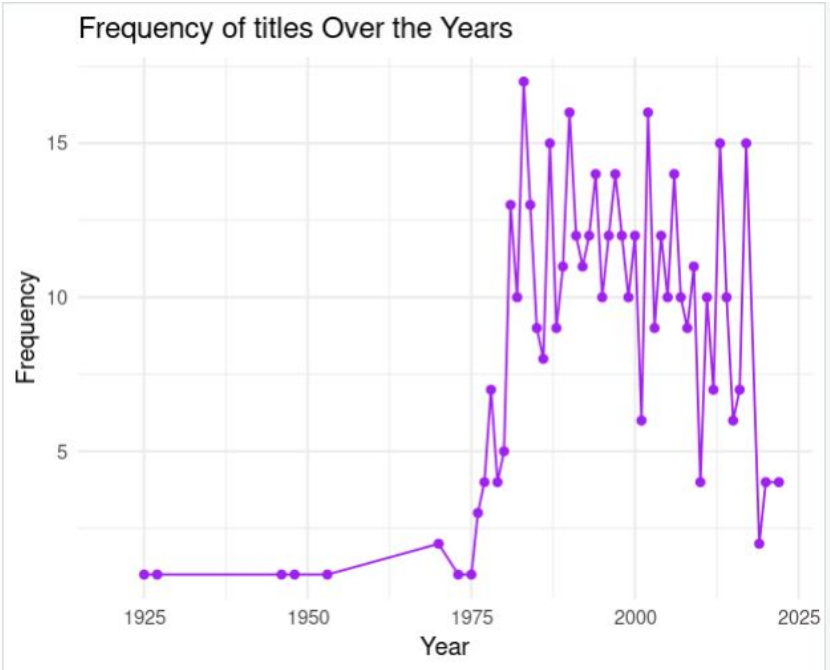
```
#call libraries
import_libraries <- function() {
  library(dplyr)
  library(stringr)
  library(quanteda)
  library(stringr)
  library(quanteda.textplots)
  library(quanteda.textstats)
  library(quanteda.textmodels)
  library(readxl)
  library(zoom)
  library(rainette)
  library(writexl)
  library(ggplot2)
  library(shiny)
  library(timevis)
}
import_libraries()
```

Importing and subsetting the database

- ARUS Digital Repository of the University of Porto
- 29 columns of metadata
- Over 19.000 entries
- Subset of feminist utopias.

Quantile and titles by year

```r
#Creates an object with the quantiles
quant <- as.data.frame(quantile(df$year, na.rm=TRUE, probs = c(0,.1,.2,.3,.4,.5,.6,.7,.8,.9,1)))
View(quant)

#Creates a plot with the quantiles
plot(quant)

  #creates a dataframe with the number of works by year
title_years = as.data.frame(table(df$year))

#changes the years so we can create the chart
title_years$Var1=as.numeric(levels(title_years$Var1))[title_years$Var1]

#creates a line chart with the number of works per year
ggplot(title_years, aes(x = Var1, y = Freq)) +
  geom_line(color="purple") +
  geom_point(color="purple") +
  labs(title = "Frequency of titles Over the Years",
       x = "Year",
       y = "Frequency") +
  theme_minimal() +
  xlim(1920, max(title_years$Var1))
```

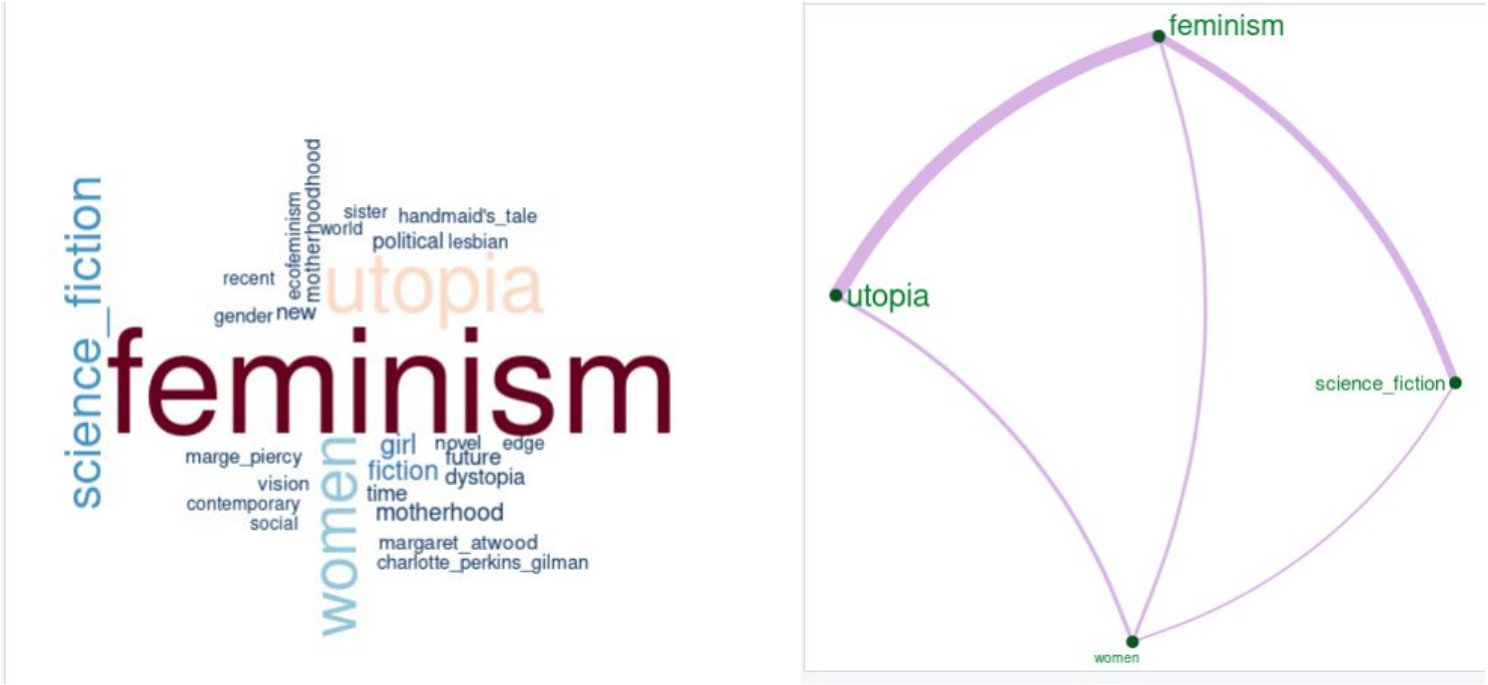


Quanteda

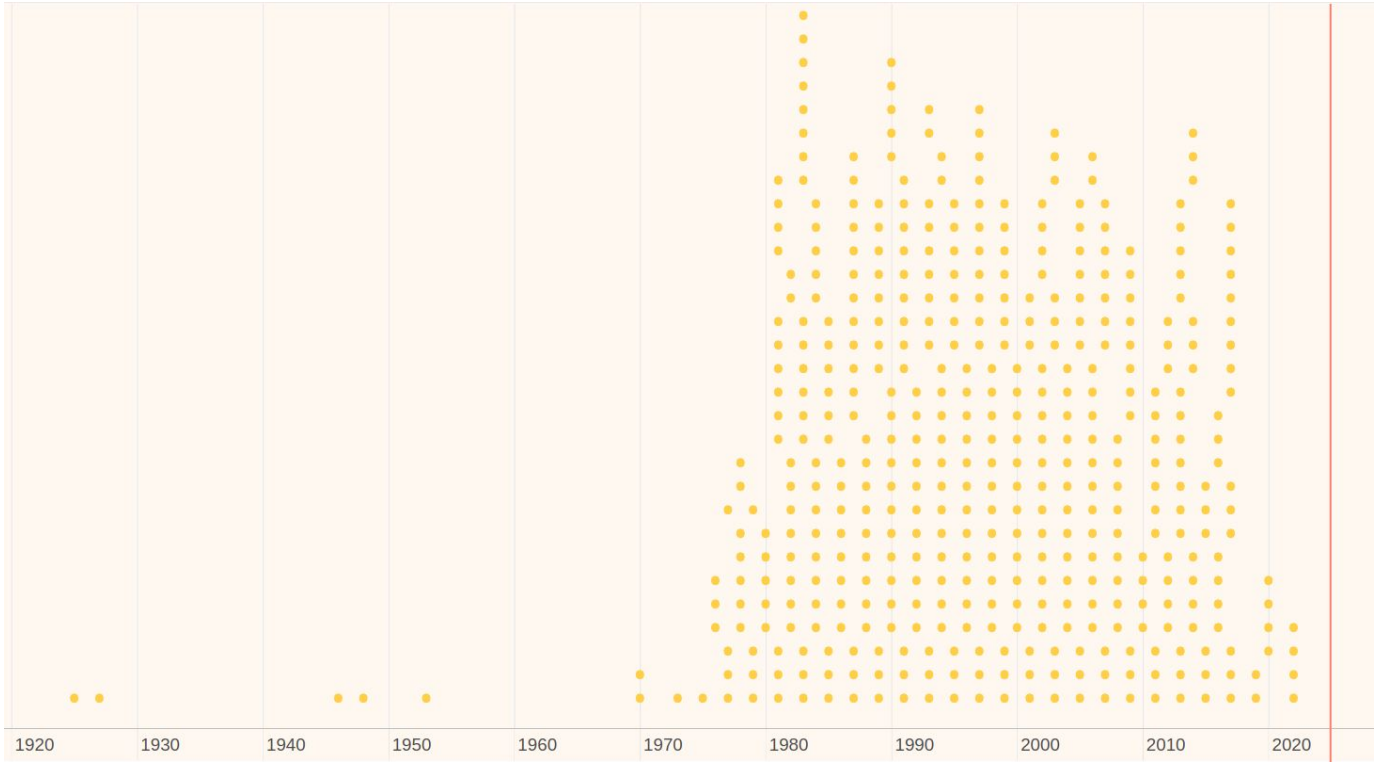
- Corpus
- Tokens
- Dfm

```r
#creates the corpus
ut_corpus <- corpus(df, docid_field = "docid_field", text_field = "text_field")

#creates the tokens
ut_tokens <- tokens(ut_corpus,
                    remove_punct = TRUE,
                    remove_symbols = TRUE,
                    remove_separators = TRUE,
                    include_docvars = TRUE)

#removes the apostrophe
ut_tokens <- tokens_split(
  ut_tokens,
  separator = "'",
  remove_separator = TRUE
)


#removes stopwords and unwanted patterns
ut_tokens <- tokens_select(ut_tokens, pattern = c(stopwords("en"),stopwords("fr"),stopwords("it"), stopwords("de"), "s"), selection = "remov

#textstat_summary(ut_tokens)
```

Word Cloud and Network



Timeline



Thank You!!