‘.oppose() method to the Senecan disputed plays

Paschalis Agapitos

Importing the libraries

```
library(stylo)
```

```
##
## ### stylo version: 0.7.4 ###
##
## If you plan to cite this software (please do!), use the following reference:
##     Eder, M., Rybicki, J. and Kestemont, M. (2016). Stylometry with R:
##     a package for computational text analysis. R Journal 8(1): 107-121.
##     <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>
##
## To get full BibTeX entry, type: citation("stylo")
```

Setting working directory

The first step that we need to take is to set the working directory to the folder where this notebook and our dataset are saved. In our case `Analysis` is our working directory.

```
setwd("../oppose/")
getwd()
```

```
## [1] "/Users/paschalis/Documents/MA_DH/Thesis/seneca_stylometry/analysis/oppose"
```

For `.oppose()` to be applied to the texts there is no need of preprocessing. We just need to create two different corpora: * the `primary_corpus`: it will contain the non-disputed plays of Seneca the Younger * the `secondary_corpus`: + `secondary_set`: it will contain `Octavia` + `secondary_set1`: it will contain `Hercules Oetaeus`

```
primary.corpus.sene <- load.corpus(files = "all", corpus.dir = "primary_set/",
                                   encoding = "UTF-8")

secondary.corpus.oct <- load.corpus(files="all", corpus.dir = "secondary_set/",
                                    encoding = "UTF-8")

tokenized.primary.corpus <- txt.to.words.ext(primary.corpus.sene, corpus.lang = "Latin.corr",
                                             preserve.case = F)

tokenized.secondary.o <- txt.to.words.ext(secondary.corpus.oct, corpus.lang = "Latin.corr",
                                          preserve.case = F)
```

Remove the pronouns

It was decided to remove the pronouns, since some pronouns are connected to the genre of the text.

```
prim.corpus.no.pronouns <- delete.stop.words(tokenized.primary.corpus,
                                             stop.words = stylo.pronouns(corpus.lang = "Latin.corr"))

sec.corpus.oct.no.pronouns <- delete.stop.words(tokenized.secondary.o,
                                                stop.words = stylo.pronouns(corpus.lang = "Latin.corr"))


# the list with the pronouns removed
# stylo.pronouns(corpus.lang = "Latin")
```

Character 4-grams

Extracting the features (character 4-grams)

The final step before proceeding to the methods is to extract the features that we want to use and add them to a table with frequencies. In our case, we want to extract character 4-grams.

```
# Seneca non-disputed plays

prim.w.1grams <- txt.to.features(prim.corpus.no.pronouns,
                                 features = "w",
                                 ngram.size = 1)


freq.features.prim.cor <- make.frequency.list(prim.w.1grams,
                                              head = 1500)

freqs.prim.corpus <- make.table.of.frequencies(prim.w.1grams,
                                               features = freq.features.prim.cor,
                                               relative = T)
```

```
## processing  8  text samples
```

```
##
```

```
## combining frequencies into a table...
```

```
# Octavia

sec.w.1grams.oct <- txt.to.features(sec.corpus.oct.no.pronouns,
                                    features = "w",
                                    ngram.size = 1)

freq.features.sec.corp.oct <- make.frequency.list(sec.w.1grams.oct,
                                                  head = 1500)

freqs.sec.corpus.oct <- make.table.of.frequencies(sec.w.1grams.oct,
                                                  features = freq.features.sec.corp.oct,
                                                  relative = T)
```

```
## processing  1  text samples
```

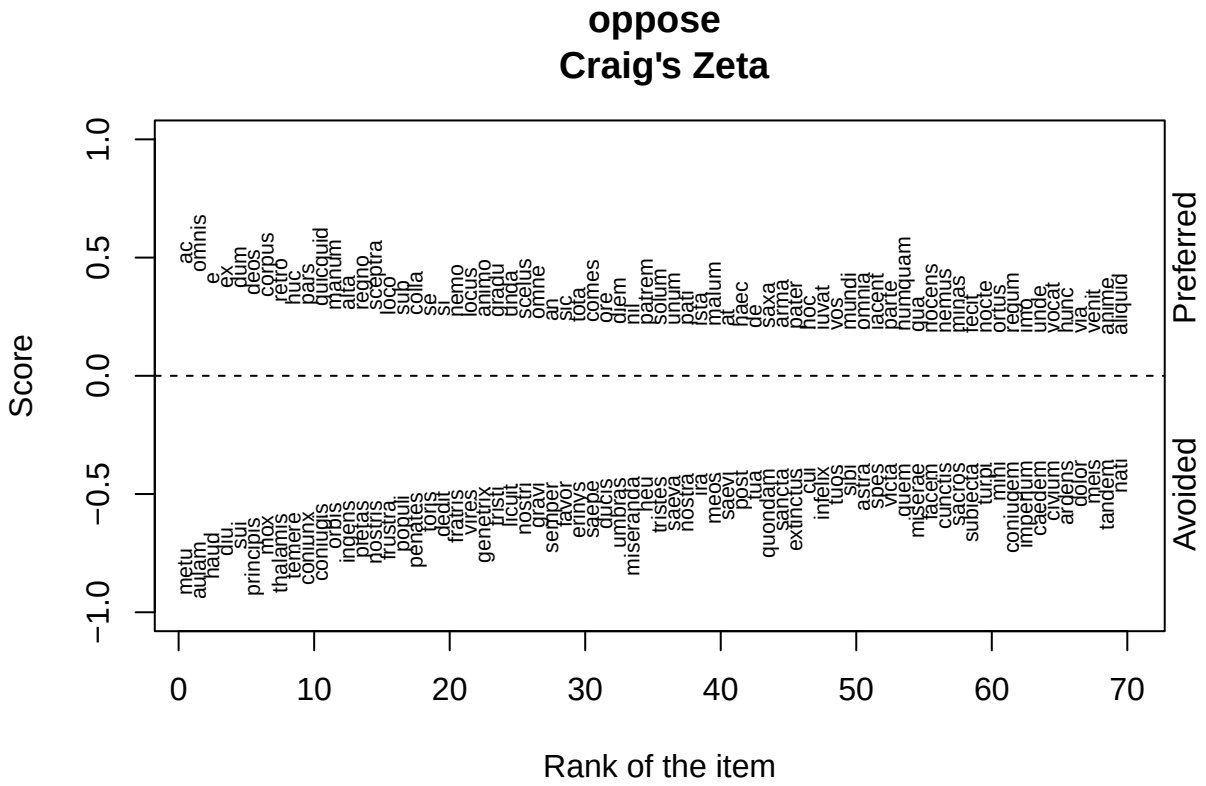
```
##
```

```
## combining frequencies into a table...
# relative=True to compute the relative frequency
```

```
oct_vs_seneca = oppose(gui = T,
                       path = "../oppose/",
                       primary.corpus.dir = "primary_set/",
                       test.corpus.dir = "secondary_set/")
```

```
## The subcorpora will be loaded from text files...
## slicing input text into tokens...
##
## turning words into features, e.g. char n-grams (if applicable)...
## slicing input text into tokens...
##
## turning words into features, e.g. char n-grams (if applicable)...
##
## Slicing the texts into samples...
## sen_ag
##  - text length (in words): 5447
##  - nr. of samples: 10
##  - nr. of words dropped at the end of the text: 447
## sen_her_f
##  - text length (in words): 7496
##  - nr. of samples: 14
##  - nr. of words dropped at the end of the text: 496
## sen_med
##  - text length (in words): 5558
##  - nr. of samples: 11
##  - nr. of words dropped at the end of the text: 58
## sen_oed
##  - text length (in words): 5764
##  - nr. of samples: 11
##  - nr. of words dropped at the end of the text: 264
## sen_phaed
##  - text length (in words): 7061
##  - nr. of samples: 14
##  - nr. of words dropped at the end of the text: 61
## sen_phoen
##  - text length (in words): 4072
```

```
##  - nr. of samples: 8
##  - nr. of words dropped at the end of the text: 72
## sen_thy
##  - text length (in words): 6159
##  - nr. of samples: 12
##  - nr. of words dropped at the end of the text: 159
## sen_tro
##  - text length (in words): 6671
##  - nr. of samples: 13
##  - nr. of words dropped at the end of the text: 171
## sen_oct
##  - text length (in words): 5093
##  - nr. of samples: 10
##  - nr. of words dropped at the end of the text: 93
##
## Extracting distinctive words... (this might take a while)
##
## Primary set...
## Secondary set...
## comparison done!
## slicing input text into tokens...
##
## turning words into features, e.g. char n-grams (if applicable)...
## sen_oct
##  - text length (in words): 5093
##  - nr. of samples: 10
##  - nr. of words dropped at the end of the text: 93
##
```




```
# do the same for Hercules Oetaeus
secondary.corpus.her.o <- load.corpus(files="all", corpus.dir = "secondary_set1/",
                                      encoding = "UTF-8")

tokenized.secondary.her.o <- txt.to.words.ext(secondary.corpus.her.o, corpus.lang = "Latin.corr",
                                              preserve.case = F)
```

Remove the pronouns

It was decided to remove the pronouns, since some pronouns are connected to the genre of the text.

```
sec.corpus.her.o.no.pronouns <- delete.stop.words(tokenized.secondary.her.o,
                                                  stop.words = stylo.pronouns(corpus.lang = "Latin.corr"))


# the list with the pronouns removed
# stylo.pronouns(corpus.lang = "Latin")
```

```
# Hercules Oetaeus

sec.w.1grams.her.o <- txt.to.features(sec.corpus.her.o.no.pronouns,
                                      features = "w",
                                      ngram.size = 1)

freq.features.sec.corp.her.o <- make.frequency.list(sec.w.1grams.her.o,
                                                    head = 1500)

freqs.sec.corpus.her.o <- make.table.of.frequencies(sec.w.1grams.her.o,
                                                    features = freq.features.sec.corp.her.o,
```

```
                                                    relative = T)
```

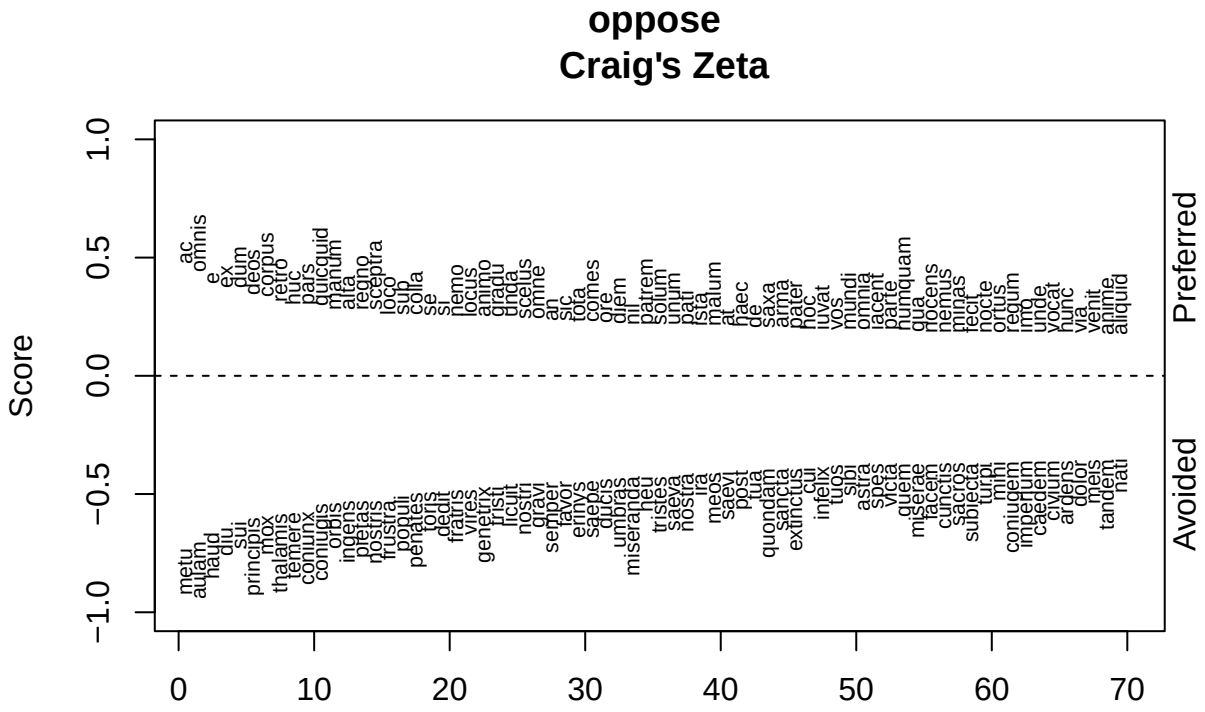
```
## processing  1  text samples
##
## combining frequencies into a table...
```

```
# relative=True to compute the relative frequency
```

```
hero_vs_seneca = oppose(gui = T,
                        path = "../oppose/",
                        primary.corpus.dir = "primary_set/",
                        test.corpus.dir = "secondary_set1/")
```

```
## The subcorpora will be loaded from text files...
## slicing input text into tokens...
##
## turning words into features, e.g. char n-grams (if applicable)...
## slicing input text into tokens...
##
## turning words into features, e.g. char n-grams (if applicable)...
##
## Slicing the texts into samples...
## sen_ag
##  - text length (in words): 5447
##  - nr. of samples: 10
##  - nr. of words dropped at the end of the text: 447
## sen_her_f
##  - text length (in words): 7496
##  - nr. of samples: 14
##  - nr. of words dropped at the end of the text: 496
## sen_med
##  - text length (in words): 5558
##  - nr. of samples: 11
##  - nr. of words dropped at the end of the text: 58
## sen_oed
##  - text length (in words): 5764
##  - nr. of samples: 11
##  - nr. of words dropped at the end of the text: 264
## sen_phaed
##  - text length (in words): 7061
##  - nr. of samples: 14
```

```
##  - nr. of words dropped at the end of the text: 61
## sen_phoen
##  - text length (in words): 4072
##  - nr. of samples: 8
##  - nr. of words dropped at the end of the text: 72
## sen_thy
##  - text length (in words): 6159
##  - nr. of samples: 12
##  - nr. of words dropped at the end of the text: 159
## sen_tro
##  - text length (in words): 6671
##  - nr. of samples: 13
##  - nr. of words dropped at the end of the text: 171
## sen_oct
##  - text length (in words): 5093
##  - nr. of samples: 10
##  - nr. of words dropped at the end of the text: 93
##
## Extracting distinctive words... (this might take a while)
##
## Primary set...
## Secondary set...
## comparison done!
## slicing input text into tokens...
##
## turning words into features, e.g. char n-grams (if applicable)...
## sen_her_o
##  - text length (in words): 11158
##  - nr. of samples: 22
##  - nr. of words dropped at the end of the text: 158
##
```




```
# summary(oct_vs_seneca)
print(oct_vs_seneca$words.preferred.scores)[1:10]
```

```
##        ac      omnis         e        ex       dum      deos     corpus
##  0.4741935  0.4408602  0.3870968  0.3731183  0.3731183  0.3408602  0.3333333
##      retro       huc      pars   quicquid     manum      alta     regno
##  0.3118280  0.3010753  0.3010753  0.2946237  0.2795699  0.2795699  0.2795699
##    sceptra      loco       sub     colla        se        si     nemo
##  0.2795699  0.2580645  0.2580645  0.2580645  0.2548387  0.2526882  0.2473118
##      locus     animo     gradu     unda     scelus      omne        an
##  0.2473118  0.2473118  0.2473118  0.2473118  0.2408602  0.2408602  0.2301075
##        sic      tota     comes      ore      diem       nil    patrem
##  0.2301075  0.2258065  0.2258065  0.2225806  0.2193548  0.2150538  0.2150538
##      solum      unum      pati     ista     malum        at      haec
##  0.2150538  0.2150538  0.2150538  0.2118280  0.2118280  0.2086022  0.2053763
##         de      saxa      arma     pater      hoc     iuvat       vos
##  0.2043011  0.2043011  0.2010753  0.1978495  0.1956989  0.1935484  0.1935484
##      mundi     omnia    iacent     parte   numquam       qua    nocens
##  0.1935484  0.1935484  0.1935484  0.1935484  0.1903226  0.1870968  0.1827957
##      nemus     minas     fecit     nocte     ortus     regum       imo
##  0.1827957  0.1827957  0.1827957  0.1827957  0.1827957  0.1827957  0.1827957
##       unde     vocat      hunc       via     venit     anime   aliquid
##  0.1827957  0.1827957  0.1795699  0.1795699  0.1795699  0.1720430  0.1720430
##      matri     nullus     redit     fortis     rebus      mors    precor
##  0.1720430  0.1720430  0.1720430  0.1720430  0.1720430  0.1688172  0.1688172
##      latus      regni      satis      esse      haut   monstra     moras
##  0.1688172  0.1688172  0.1688172  0.1655914  0.1612903  0.1612903  0.1612903
##     spatia       loca      magno quodcumque     regia    sanguis     solet
##  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903
```

```
##       vias      lucis      nimis      perit     silvis       viam   quisquis
##  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1580645
##     nondum       aras      troia     mortem      natos     placet        rex
##  0.1580645  0.1580645  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376
##   sequitur      tuque      video      perge    phoebus      auras      iugis
##  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376
##      multo       omni     saevus      tecum     terrae    vertice       quid
##  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1494624
##        sit       hinc        tam      magna        bis     misera      movet
##  0.1473118  0.1440860  0.1440860  0.1440860  0.1397849  0.1397849  0.1397849
##       stat      comam      senex       ulla      solus     spolia    cervice
##  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849
##        ire       nube      vides      illuc    populos     praeda      maius
##  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1365591
##     victor       alto     cuncta       mare       date       iuga     quidem
##  0.1365591  0.1365591  0.1365591  0.1333333  0.1290323  0.1290323  0.1290323
##      ditis     effare     fretum     gradum      miser     noctem      posse
##  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323
##       exul   impotens       mora     phoebi      silva        uno       mali
##  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323
##     medium    similis   utrumque     frater      mater      regis      terra
##  0.1290323  0.1290323  0.1290323  0.1258065  0.1258065  0.1258065  0.1258065
##       nova      turba     hector        has   corporis       domo     illinc
##  0.1258065  0.1225806  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##    omnibus      vinci      agros      audax      laeva      seque       tuto
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##      argos     credis      curru       dona       duos      genas      ipsum
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##     mentem     nostro      pedes    penitus      prece      pudet     regium
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##     sacrum      errat      minor     nescit       vada     qualis        tum
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1150538  0.1150538
##      labor       sine      artus    classis    liberos    quaeque      vobis
##  0.1150538  0.1150538  0.1118280  0.1075269  0.1075269  0.1075269  0.1075269
##       feci      iubes      mille      secum     silvas     taurus    vulnere
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##     dubius     facies     greges     iussus     phoebe    agnosco       arva
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##       atra      circa      decem      dubia      durus      funus     pondus
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##     pontum      regio     restat      sedet      super    thalami    trahens
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##      bello      cogit   extremum     fessis    gressus       ipso      modum
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##      notas     passim     petunt    populis     segnis      senis      styga
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##   utrimque      vagus      vasti      ignis     patris      vitae        nox
##  0.1075269  0.1075269  0.1075269  0.1043011  0.1043011  0.1043011  0.1043011
##      habet       igne     umquam     animus
##  0.1043011  0.1043011  0.1043011  0.1010753
## attr(,"description")
## [1] "Zeta scores for the preferred words"
```

```
##        ac     omnis         e        ex       dum      deos    corpus     retro
```

```
## 0.4741935 0.4408602 0.3870968 0.3731183 0.3731183 0.3408602 0.3333333 0.3118280
##       huc      pars
## 0.3010753 0.3010753
```

```
print(hero_vs_seneca$words.preferred.scores)[1:10]
```

```
##         ac      omnis          e         ex        dum       deos     corpus
##  0.4741935  0.4408602  0.3870968  0.3731183  0.3731183  0.3408602  0.3333333
##      retro        huc       pars   quicquid      manum       alta      regno
##  0.3118280  0.3010753  0.3010753  0.2946237  0.2795699  0.2795699  0.2795699
##    sceptra       loco        sub      colla         se         si       nemo
##  0.2795699  0.2580645  0.2580645  0.2580645  0.2548387  0.2526882  0.2473118
##      locus      animo      gradu       unda     scelus       omne         an
##  0.2473118  0.2473118  0.2473118  0.2473118  0.2408602  0.2408602  0.2301075
##        sic       tota      comes        ore       diem        nil     patrem
##  0.2301075  0.2258065  0.2258065  0.2225806  0.2193548  0.2150538  0.2150538
##      solum       unum       pati       ista      malum         at       haec
##  0.2150538  0.2150538  0.2150538  0.2118280  0.2118280  0.2086022  0.2053763
##         de       saxa       arma      pater        hoc      iuvat        vos
##  0.2043011  0.2043011  0.2010753  0.1978495  0.1956989  0.1935484  0.1935484
##      mundi      omnia     iacent      parte    numquam        qua     nocens
##  0.1935484  0.1935484  0.1935484  0.1935484  0.1903226  0.1870968  0.1827957
##      nemus      minas      fecit      nocte      ortus      regum        imo
##  0.1827957  0.1827957  0.1827957  0.1827957  0.1827957  0.1827957  0.1827957
##       unde      vocat       hunc        via      venit      anime    aliquid
##  0.1827957  0.1827957  0.1795699  0.1795699  0.1795699  0.1720430  0.1720430
##      matri     nullus      redit     fortis      rebus       mors     precor
##  0.1720430  0.1720430  0.1720430  0.1720430  0.1720430  0.1688172  0.1688172
##      latus      regni      satis       esse       haut    monstra      moras
##  0.1688172  0.1688172  0.1688172  0.1655914  0.1612903  0.1612903  0.1612903
##     spatia       loca      magno quodcumque      regia    sanguis      solet
##  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903
##       vias      lucis      nimis      perit     silvis       viam   quisquis
##  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1612903  0.1580645
##     nondum       aras      troia     mortem      natos     placet        rex
##  0.1580645  0.1580645  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376
##    sequitur      tuque      video      perge    phoebus      auras      iugis
##  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376
##      multo       omni     saevus      tecum     terrae    vertice       quid
##  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1505376  0.1494624
##        sit       hinc        tam      magna        bis     misera      movet
##  0.1473118  0.1440860  0.1440860  0.1440860  0.1397849  0.1397849  0.1397849
##       stat      comam      senex       ulla      solus     spolia    cervice
##  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849
##        ire       nube      vides      illuc    populos     praeda      maius
##  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1397849  0.1365591
##     victor       alto     cuncta       mare       date       iuga     quidem
##  0.1365591  0.1365591  0.1365591  0.1333333  0.1290323  0.1290323  0.1290323
##      ditis     effare     fretum     gradum      miser     noctem      posse
##  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323
##       exul   impotens       mora     phoebi      silva        uno       mali
##  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323  0.1290323
##     medium    similis   utrumque     frater      mater      regis      terra
##  0.1290323  0.1290323  0.1290323  0.1258065  0.1258065  0.1258065  0.1258065
##       nova      turba     hector        has   corporis       domo     illinc
```

```
##  0.1258065  0.1225806  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##    omnibus      vinci      agros      audax      laeva      seque       tuto
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##      argos     credis      curru       dona       duos      genas      ipsum
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##     mentem     nostro      pedes    penitus      prece      pudet     regium
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796
##     sacrum      errat      minor     nescit       vada     qualis        tum
##  0.1182796  0.1182796  0.1182796  0.1182796  0.1182796  0.1150538  0.1150538
##      labor       sine      artus    classis    liberos    quaeque      vobis
##  0.1150538  0.1150538  0.1118280  0.1075269  0.1075269  0.1075269  0.1075269
##       feci      iubes      mille      secum     silvas     taurus    vulnere
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##     dubius     facies     greges     iussus     phoebe    agnosco       arva
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##       atra      circa      decem      dubia      durus      funus     pondus
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##     pontum      regio     restat      sedet      super    thalami    trahens
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##      bello      cogit   extremum     fessis    gressus       ipso      modum
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##      notas     passim     petunt    populis     segnis      senis      styga
##  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269  0.1075269
##   utrimque      vagus      vasti      ignis     patris      vitae        nox
##  0.1075269  0.1075269  0.1075269  0.1043011  0.1043011  0.1043011  0.1043011
##      habet       igne     umquam     animus
##  0.1043011  0.1043011  0.1043011  0.1010753
## attr(,"description")
## [1] "Zeta scores for the preferred words"
```

```
##        ac     omnis         e        ex       dum      deos    corpus     retro
## 0.4741935 0.4408602 0.3870968 0.3731183 0.3731183 0.3408602 0.3333333 0.3118280
##       huc      pars
## 0.3010753 0.3010753
```