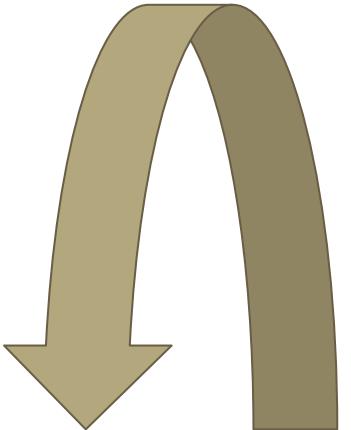


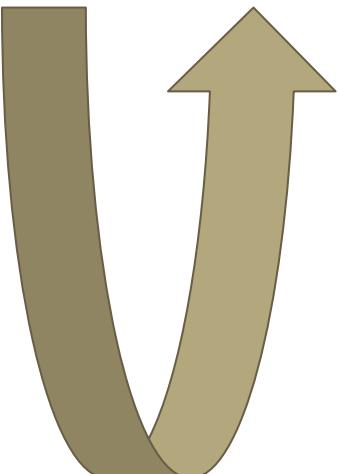
Case-studies of Consciousness & Public: Text and data mining eighteenth century based on ESTC & ECCO

COMHIS Collective

DR2 Conference 2017, Turin



Research questions



Initial data

ECCO

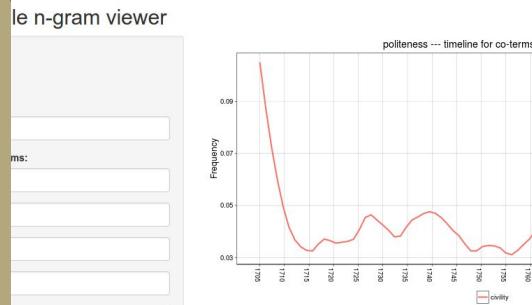
ESTC

Evolving set of analysis and processing tools

APIs

```
{  
  "docFreq" : 6322  
  "totalTermFreq"  
  "collocations" :  
    "conversation"  
    "politeness" :  
    "style" : 0.00  
    "eloquence" :  
    "obliging" : 0  
    "genteel" : 0.00  
}
```

Applications



Scripts

```
21   return(publications_yearly)  
22 }  
23 }  
24  
25 <- get_relative_hits_yearly <- function(hits_subset)  
26   averages_yearly <- hits_subset$yearly  
27   averages_yearly$total_titles <- hits_all$titles  
28   averages_yearly$frequency <- averages_yearly$titles  
29   return(averages_yearly)  
30 }  
31  
32  
33 <- get_query_set_with_freqs <- function(query_file)  
34   pub_year <- read.csv(query_file, header = FALSE)  
35   dataset <- read.csv("dataset")  
36   filtered_dataset <- filter(dataset, pub_year$year == pub_year$year)  
37   query_titles_yearly <- get_publications_yearly(query_file)  
38   query_titles_yearly$averages <-  
39   get_relative_hits_yearly(query_titles_yearly)  
40   query_filename <- basename(query_file)  
41   query_file <- paste(query_filename, ".r", sep = "")  
42   write(query_file, query_file)
```

Etc...

Project goals

Research publications

Keywords: history publishing, short-title catalogue

How to Cite: Lahti, L., Ilomäki, N. & Tolonen, M., (2015). A Quantitative History in the English Short-Title Catalogue (ESTC), 1470–1800. *L25*(2), pp.87–116. DOI: <http://doi.org/10.18352/lq.10112>

Public tools, APIs, code

This repository Search Pull requests Issues

rOpenGov / bibliographica

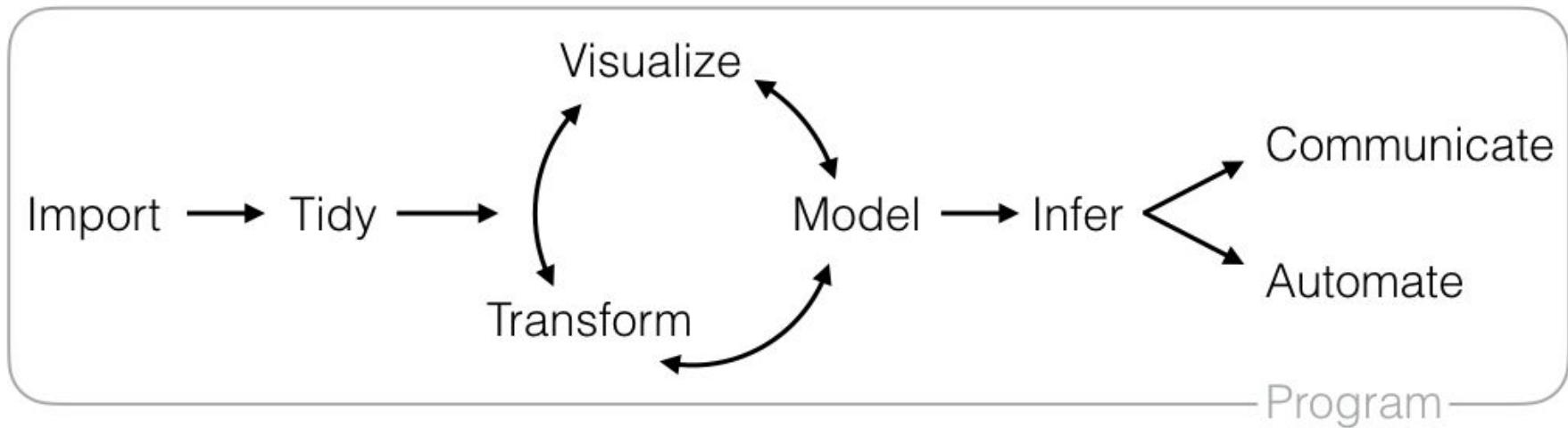
Code Issues 2 Pull requests 0 Projects 0 Wiki

Refined data

publisher	latitude	longitude	publication_id
printed for J. Cooke; and sold by the booksellers of B...	51.50853	-0.125740	London
re-printed at the Old Post Office in Fish shamble street	53.33306	-6.248890	Dublin
Sold at the Bible and Heart, in Cornhill	51.50853	-0.125740	London
printed for J. Bew	51.50853	-0.125740	London
NA	37.56596	14.282404	NA
Printed by J. Evans, Long-lane, West-smithfield, Lon...	51.50853	-0.125740	London
printed for Thomas Ewing, and William Smith	53.33306	-6.248890	Dublin
sold by J. Williams, and all the booksellers in town a...	51.50853	-0.125740	London

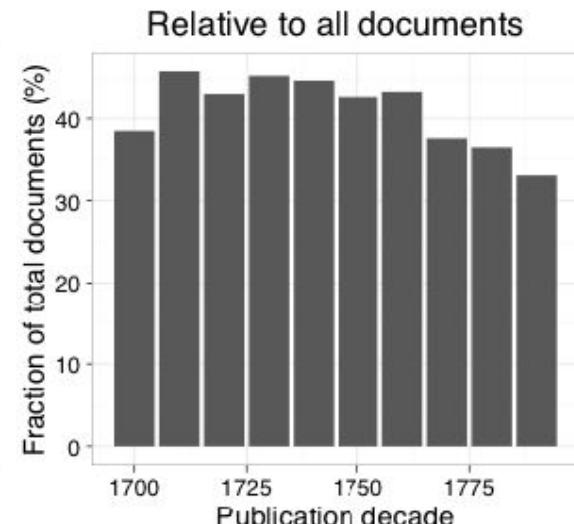
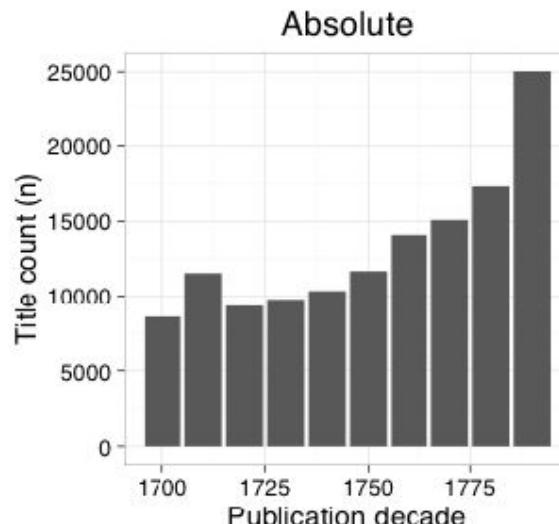
Cleaning up ESTC

- 80 % of statistical analysis is tidying up of the data. Often neglected yet implicitly assumed by many tools.

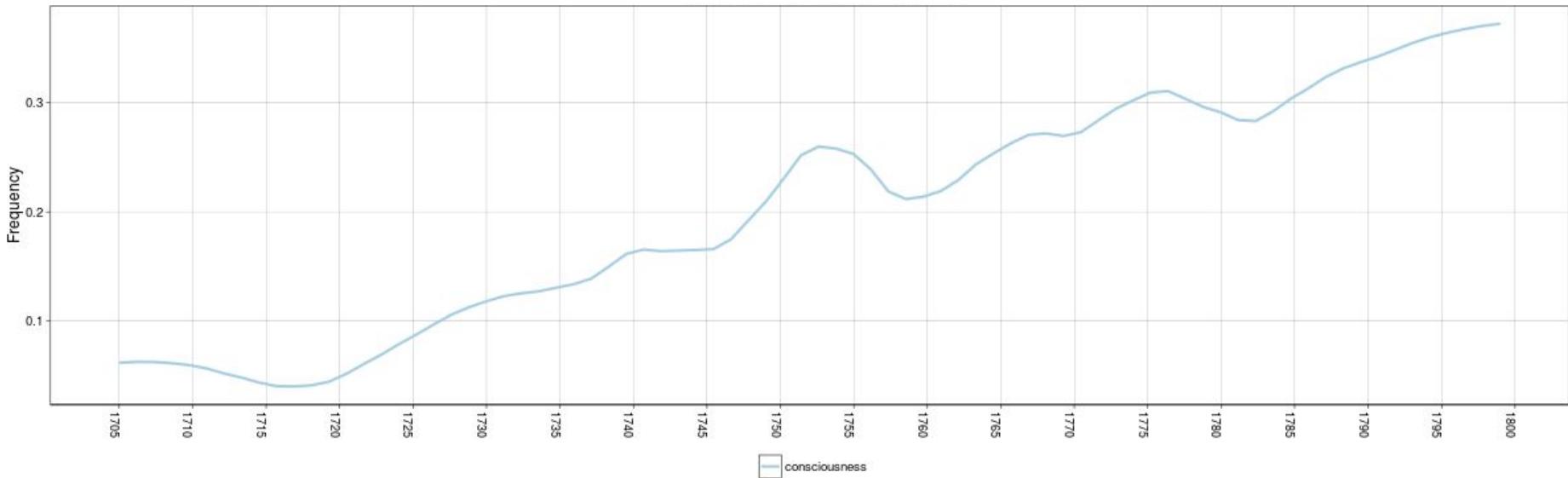


ECCO contains only a subset of the works in ESTC

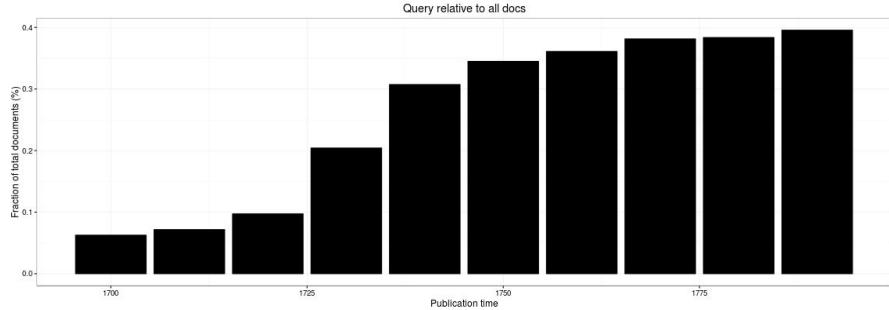
- ECCO titles refer to 54% of the 18th century English records in ESTC



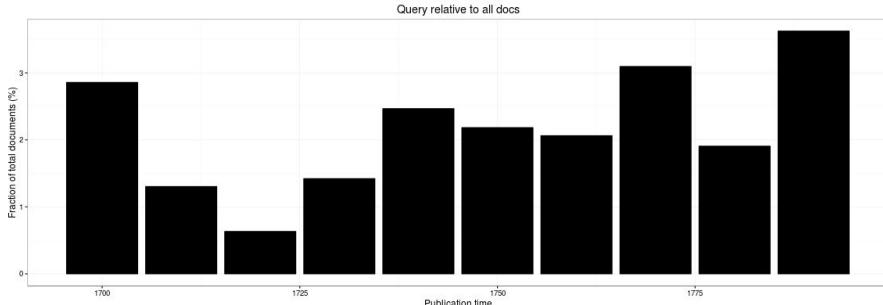
“Consciousness”



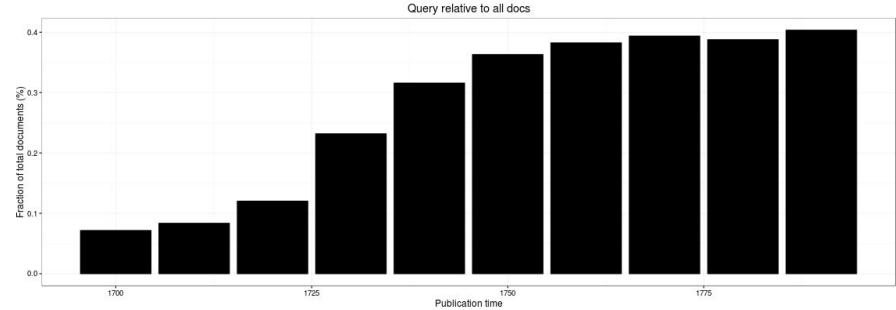
Frequency growth of “consciousness” (Locke’s works excluded)



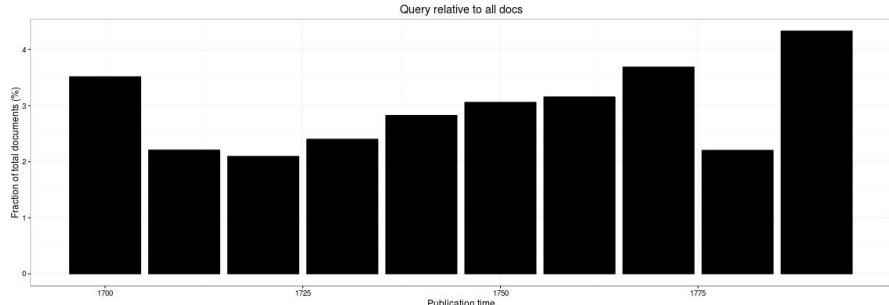
“Consciousness” among philosophers, no
Locke



Frequency growth of “consciousness” (Locke’s works included)



“Consciousness” among philosophers,
with Locke



Terms occurring in paragraphs that contain “consciousness”

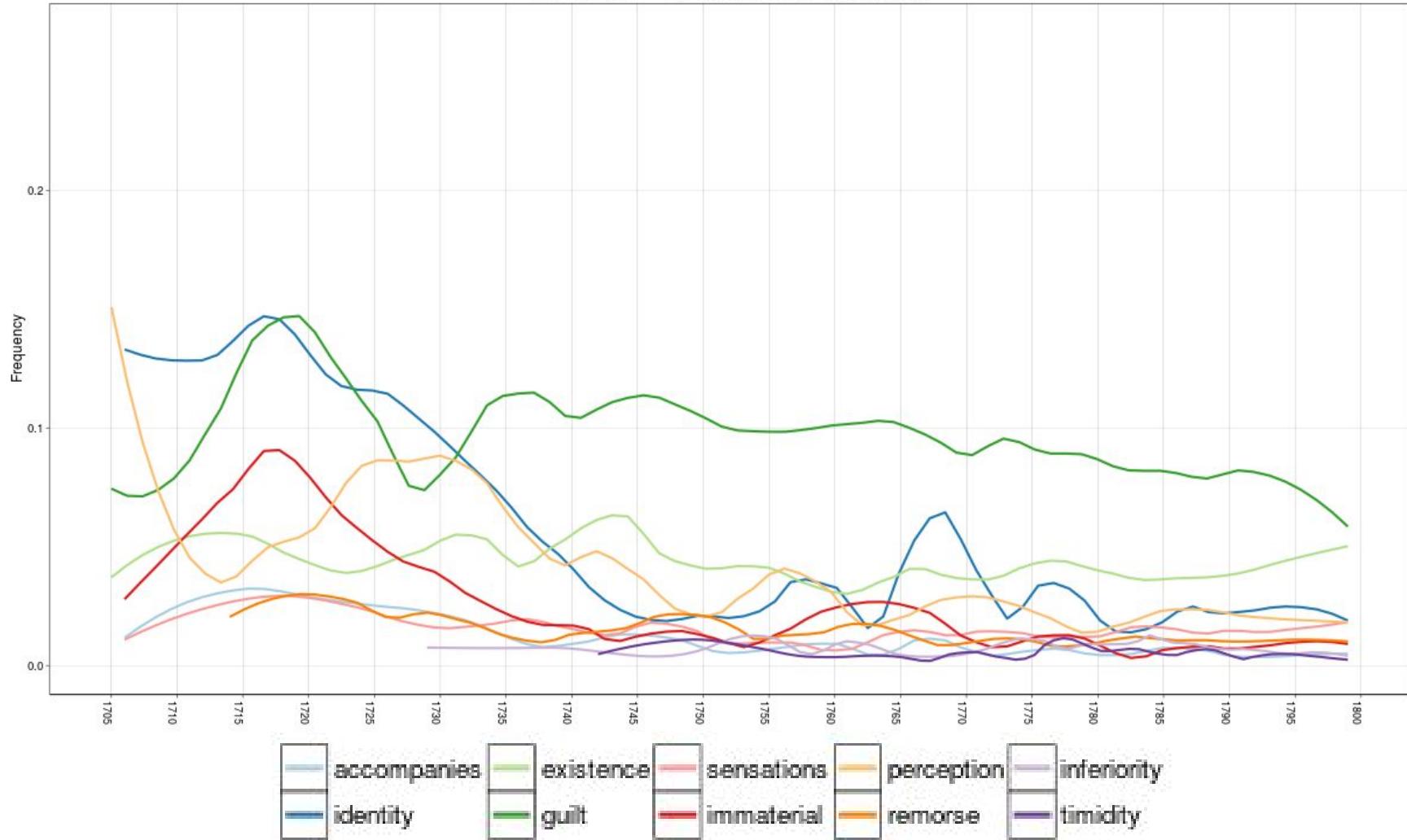
1700-1749

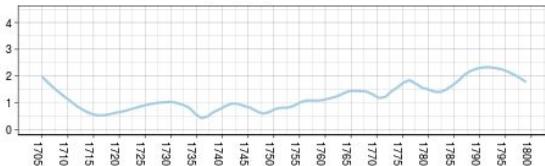
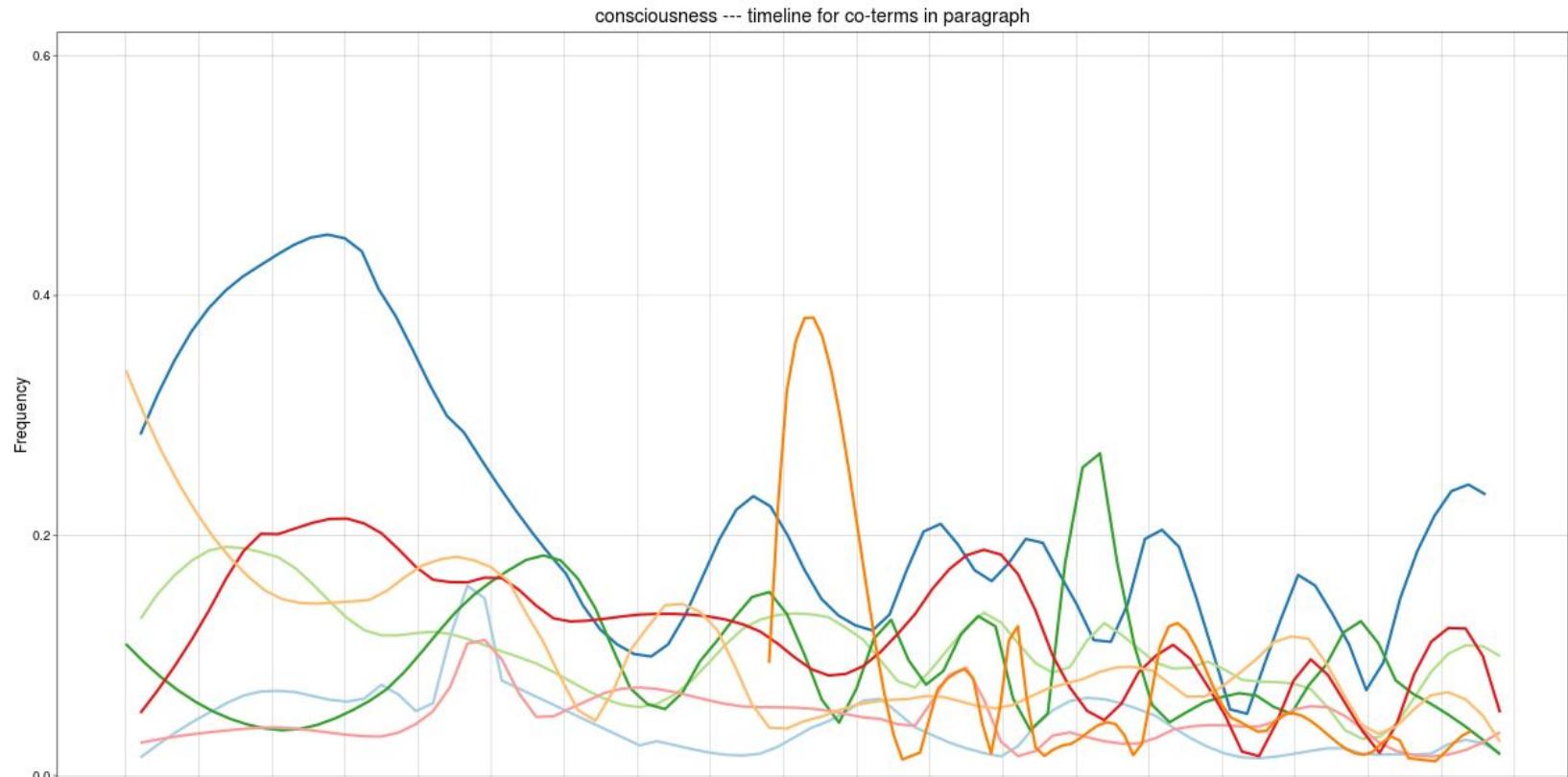
identity	17,09%
perception	5,62%
immaterial	5,60%
perceptions	5,16%
conscious	4,28%
waking	2,58%
subflance	2,34%
sensitive	2,16%
forgetfulness	2,01%
knowledg	1,94%
atom	1,79%
dreaming	1,78%
substances	1,71%
suppositions	1,69%
innate	1,63%
atheist	1,62%
thinking	1,60%
remembred	1,53%
intelligent	1,51%
atheists	1,46%

1750-1799

identity	7,46%
perceptions	2,39%
perception	2,35%
inferiority	2,02%
immaterial	1,84%
conscious	1,82%
sensations	1,19%
rectitude	1,08%
sensation	1,06%
existence	1,03%
betrays	0,96%
guilt	0,94%
accompanies	0,93%
exult	0,90%
affectation	0,89%
unworthiness	0,88%
implanted	0,87%
complacency	0,87%
oufnefs	0,86%
inability	0,82%

consciousness --- timeline for co-terms in paragraph

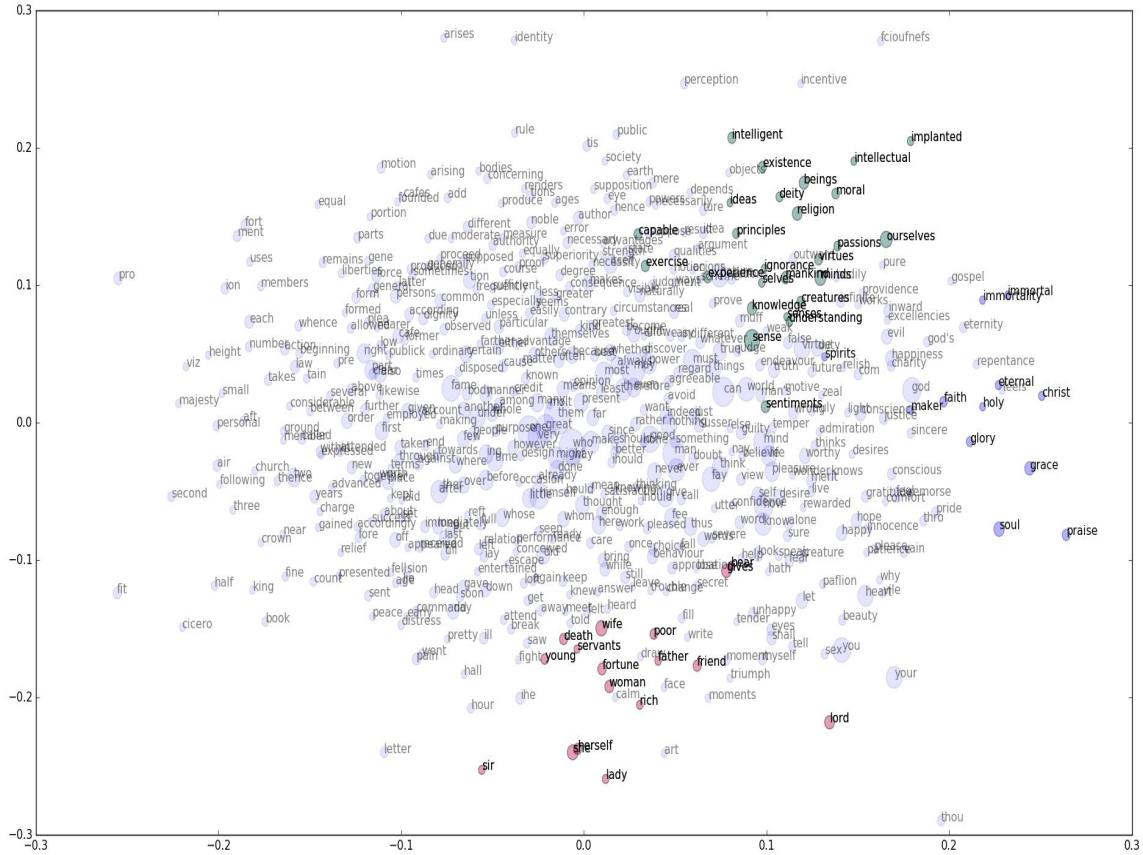




Every dot corresponds to a word occurring in the subcorpus: ‘paragraphs containing “consciousness” in 1750’.

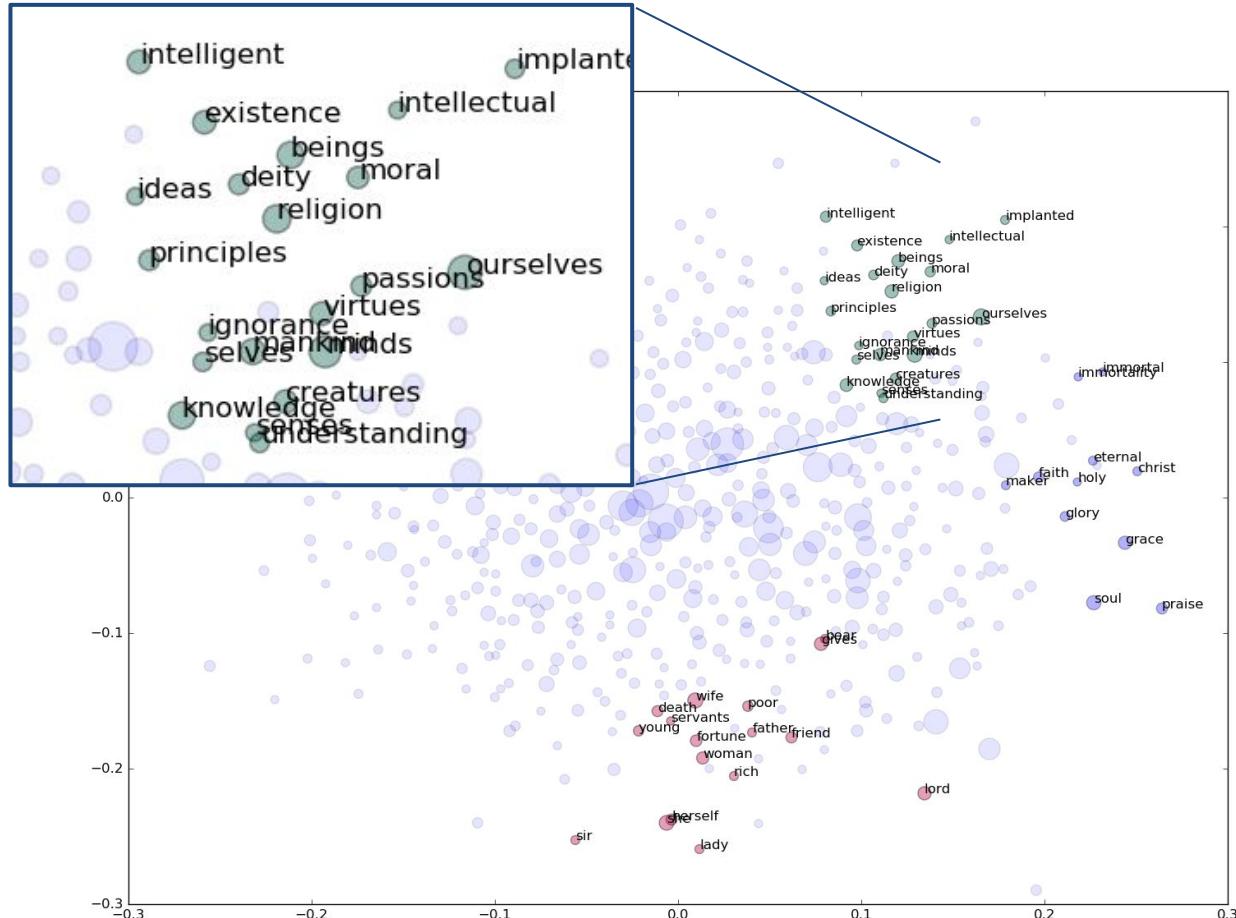
Distance between two dots is an approximation of the similarities between the contexts in which these words occur.

Sizes of the dots indicate relative frequencies of the words in the subcorpus.

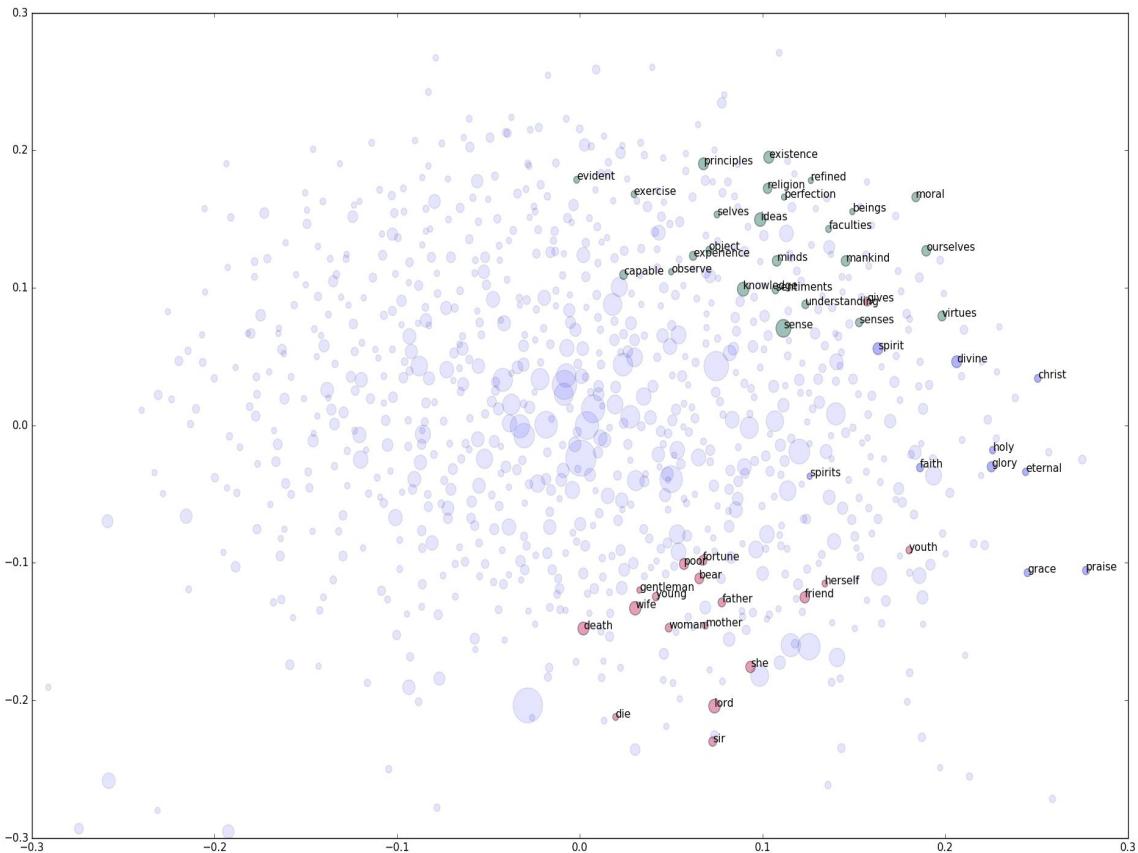


This allows to identify semantic fields.

Tracking the contents and densities of these fields over time can be used to make observations of the contextual changes within the subcorpus.



1780



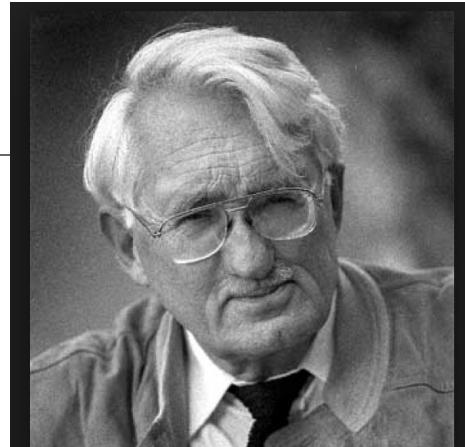
```
}, {
  "matches" : ".; 7, 40. .I aniw.er, This is. oneMean wifely appointed by God for our RefÂ'r-matnic ; but -hat  
whatever it prq;ee4d frqA,it ftppofes nothing like his Liberty: For, whe-ther it arise in a vicious Man, from his.A  
tingcontrary to his :moral Sense or Confcienc ,: (aSense that God has implanted in every 'Man,iby which we  
hneefarly approve; good AAtions in <b>ourselves</b> apd others, and disapprove .il-ones) or froam a  
<b>Consciousness</b> that henhasinw.arr'd the Resentment, Diaprobabataqn, .nd-Âfontempt of others, ( the necessary  
Efte& ofS;elfdefence in them ) which no Man can en,:dure the Want, of, even in Thought, he- must folle Consequence be  
afham'd, grjiev', and haveRemoerf ;'and this Remorse must follo, w.with-put supposing the Agent free in his Sense  
ofthe Word. , } ; 'He concludesi withz faying, \'C ~If thze few4: plain Observations had been adverted .to,,y' Mr.  
'  
  "pubDateStart" : 17340000,  
  "ESTCID" : "T169583",  
  "score" : 2  
, {  
  "matches" : "not so much from national prejudice, as fromna <b>consciousness</b> of the purity of your  
own:intentions in your conduct towards this re-public ;:from having felt a general friendship:to her: cause, and not  
perceiving a sufficientground of complaint on which her resent-ment can be founded. But you are not tolearn that  
jealousy is one of the strongest and:blindest of the human <b>passions</b>; and I believeyou will be convinced that  
the facts hintedat in my Letter, viewed through the mist ofjealousy that had constantly surrounded theleaders in the  
revolution, could not fail ofproducing effects similat to what' we nowdeplore.No!:! ",  
  "pubDateStart" : 17990000,  
  "ESTCID" : "N055117",  
  "score" : 2  
, {  
  "matches" : "We cannot easily divest ourselvesof the <b>consciousness</b> of our native dignity to such a  
degree, asnot to perceive upon the slightest refleaiion, that those fatis-fadions which the brute may enjoy as  
perfectly as the man, andthe ideot, as those of the best capacities, cannot be worthy ofan eager and intense  
perfuit.The transitory nature of most kinds of pleasure is anothercircumstance that ought to moderate our attachment  
to them:their duration is, in general, to short, as to bear no proportionto the time we must spend in order to  
obtain them. ... The immoderate perfuit of themmust be a toil and fatigue, instead of a pleasure ; and, if we arenot  
sparing in the use of them, they must very soon becomefamiliar, unanimated, and insipid.Nor can we long receive any  
considerable fatisfa&ion fromthe most exquisite entertainments of a refined and elegant taute,if we devote  
<b>ourselves</b> to them merely as matter of amusement, and without a view to some superior and more substantial  
utility.Though we are by nature firongly attached to pleasure, yet thehuman mind cannot but resent such an excessive  
application toamusements of any k'nd, as renders them the principal businessof life. ",  
  "pubDateStart" : 17610000,  
  "ESTCID" : "P6032",  
  "score" : 2  
, {
```

By choosing a set of terms in one of the groups, we get directly to the actual texts, i.e the paragraphs containing “consciousness” and at least one of the chosen co-terms.

Spheres of “public” in eighteenth-century Britain

Research question

- Evaluating the **Habermasian discussion** of “public” based on basic assumption of distributional semantics
- Differentiating the **uses of public** (finance, politics, morals and religion); with particular **focus on the personified uses of public in 1790s**
- sketching **Enlightenments** based on distances of authors’ uses of public



Studies of “public” in 18th-century Britain often focus on the Spectator

ESTC Report

Subcorpus: spectator

Search string: /publick*/

Search fields: All fields

Min API hits: 1

Years: 1,700 - 1,799

Publication place: All

Language:

Summary	Books/Pamphlets	Author titlecount	Top places	Top publishers	Top hits per title
Top titles by titlecount	Top authors	Timeline for titlecount	Timeline for paper consumption		

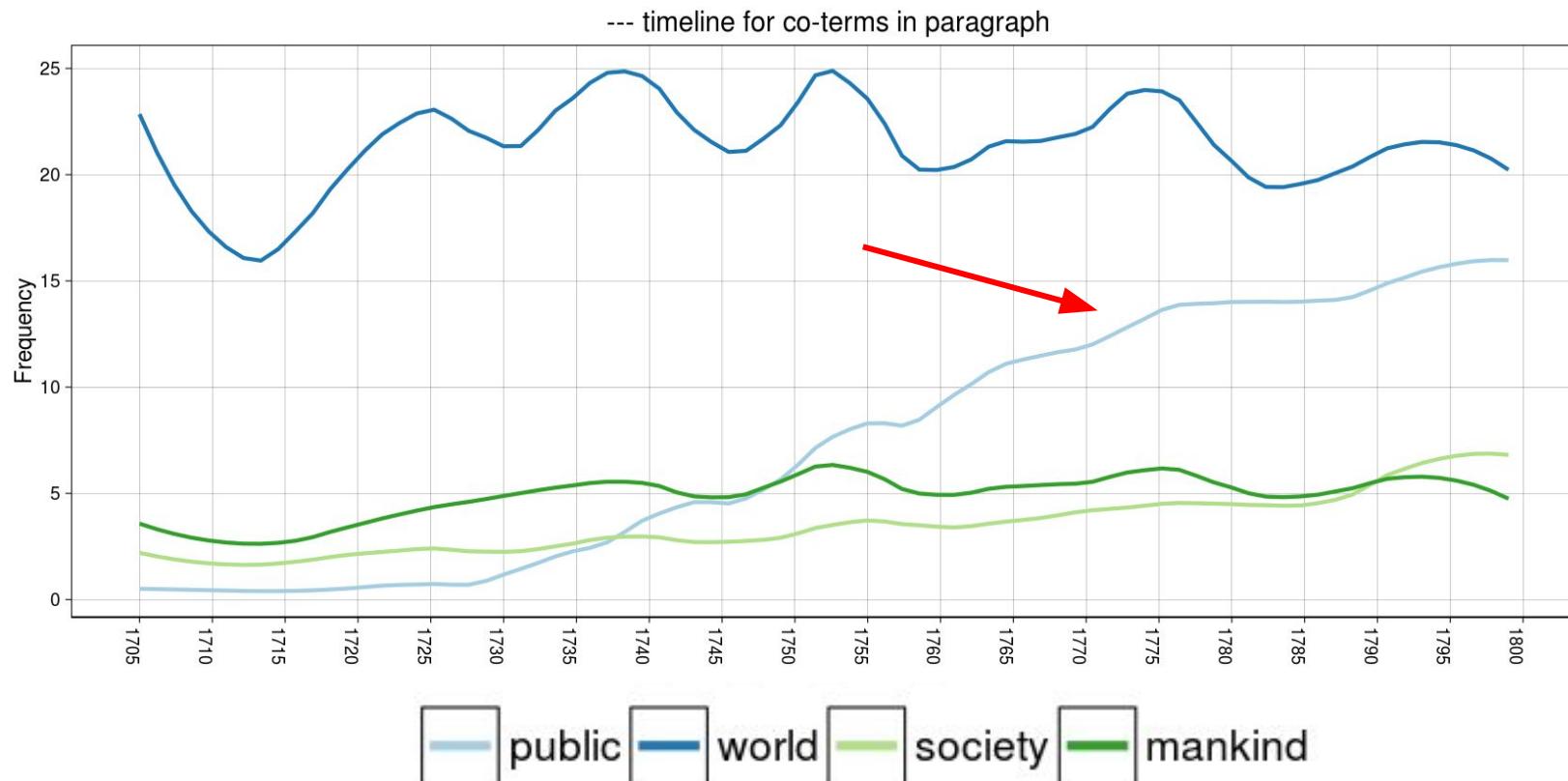
Search summary

Search string	/publick*/
Years	1700 - 1799
Publication place	All
Language	any
Time segment	10
Document type	All
ECCO API query hits	117790
Unique hit IDs in ESTC or subcorpus	137 (0.12 %)
Unique hit IDs after filtering for year, place, etc.	137 (0.12 %)

input idfile name

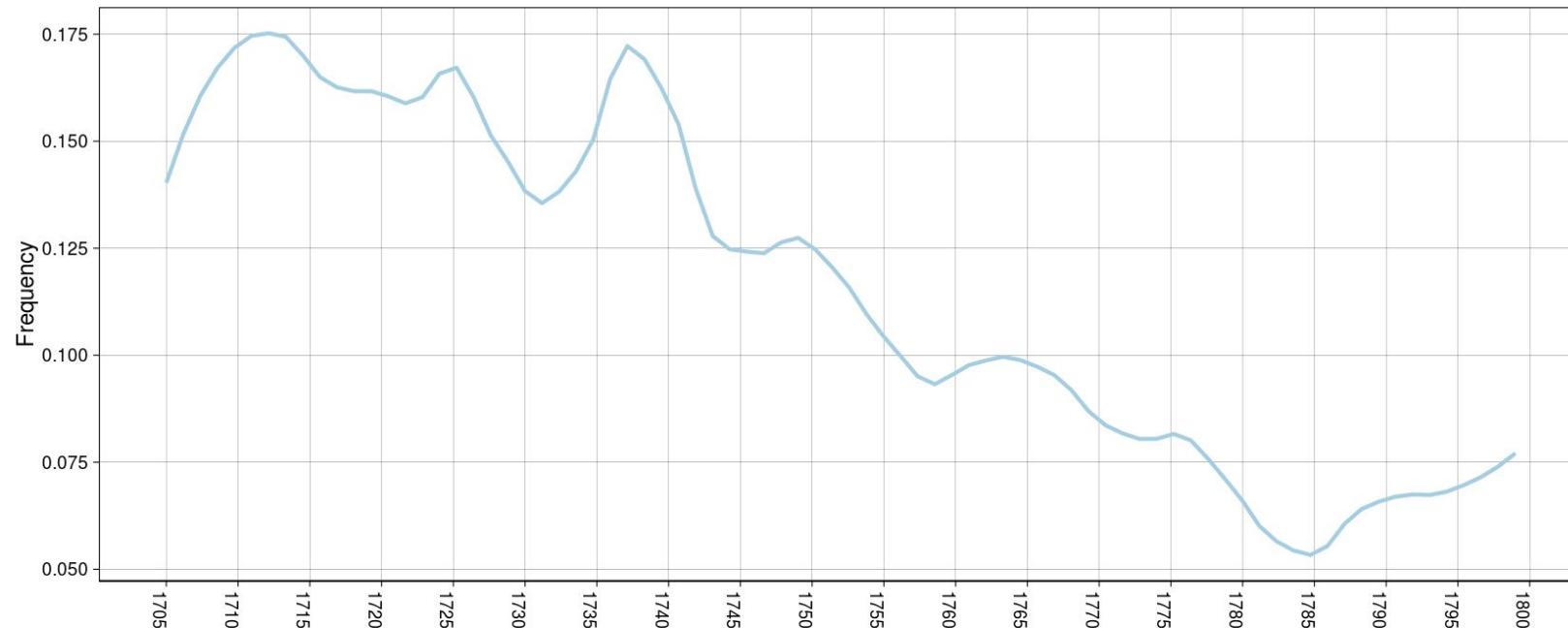
Save subcorpus

“public”, “world”, “society”, “mankind”



Declining religious 2grams in ECCO regarding “public”

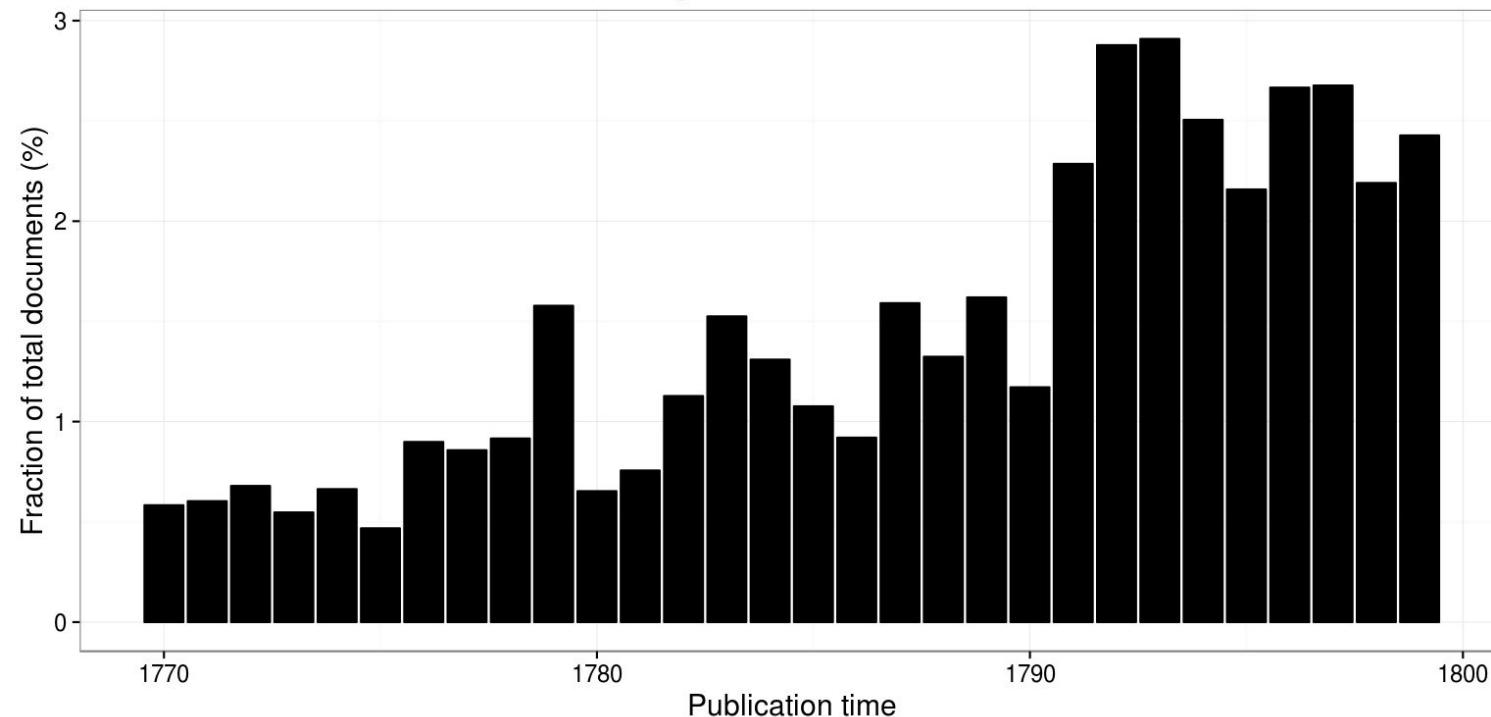
--- timeline for co-terms in paragraph



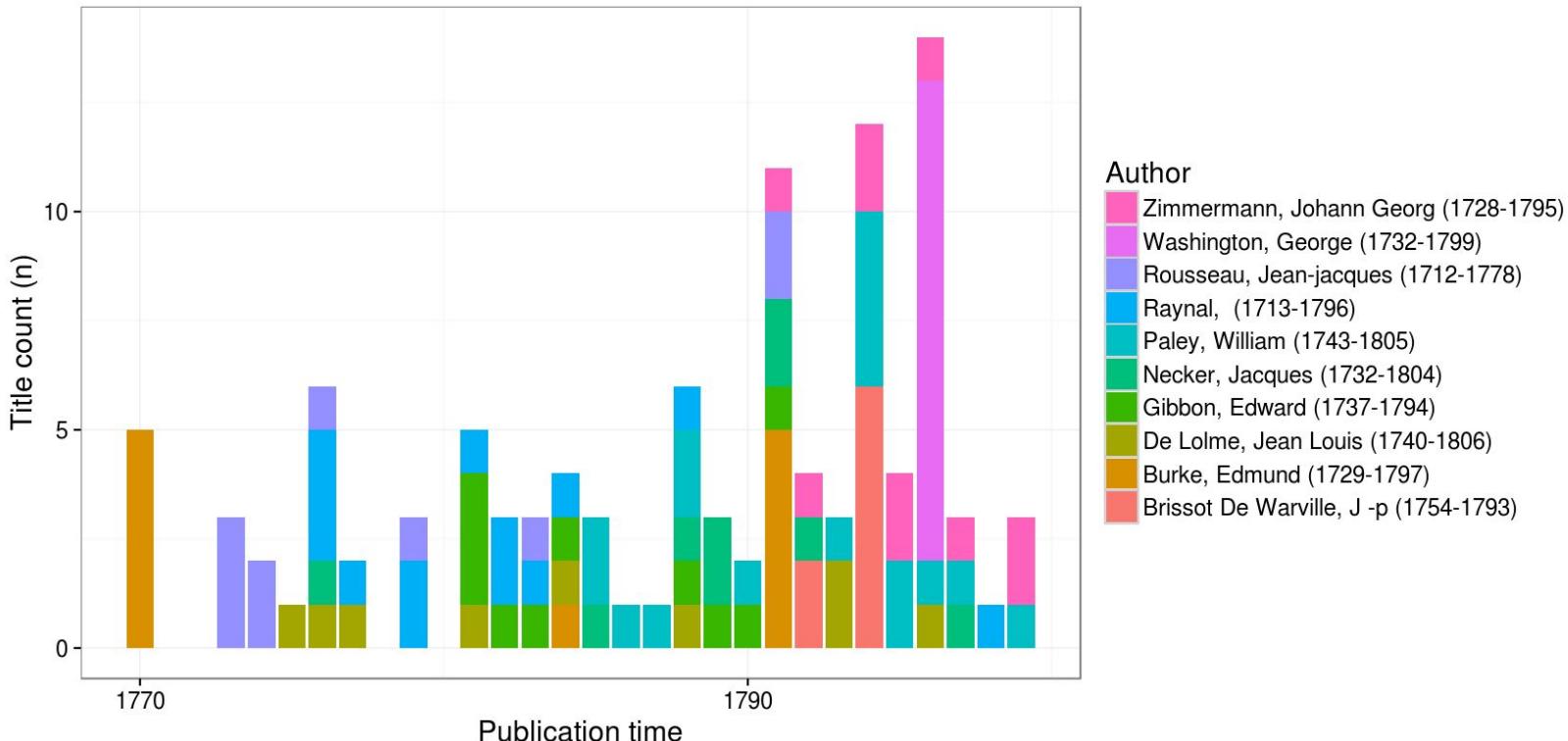
"public impieties" "publick impieties" "public synagogues" "publick synagogues" "public worflipand" "publick worflipand" "public worshipping" "publick worshipping" "public prayersin" "publick prayersin" "public mass" "publick mass" "public synod" "publick synod" "public pray" "publick pray" "public chappels" "publick chappels" "public sin" "publick sin" "public sins" "publick sins" "public baptism" "publick baptism" "public prophecy" "publick prophecy" "public congregation" "publick congregation" "public conscience" "publick conscience" "public repentance" "publick repentance" "public prayers" "publick prayers" "public prayer" "publick prayer"

Public opinion, 1770-1799 (relative to all documents)

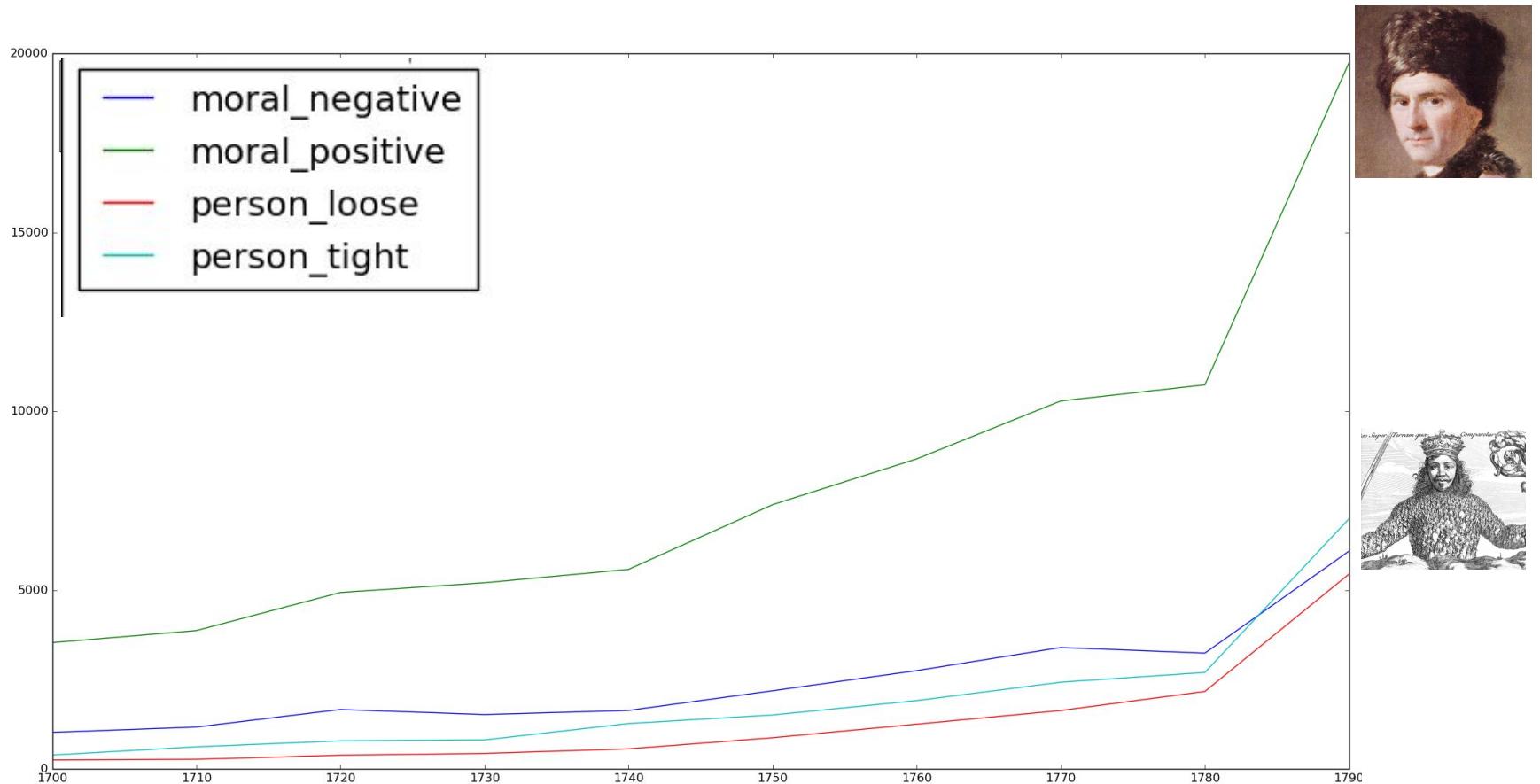
Query relative to all docs



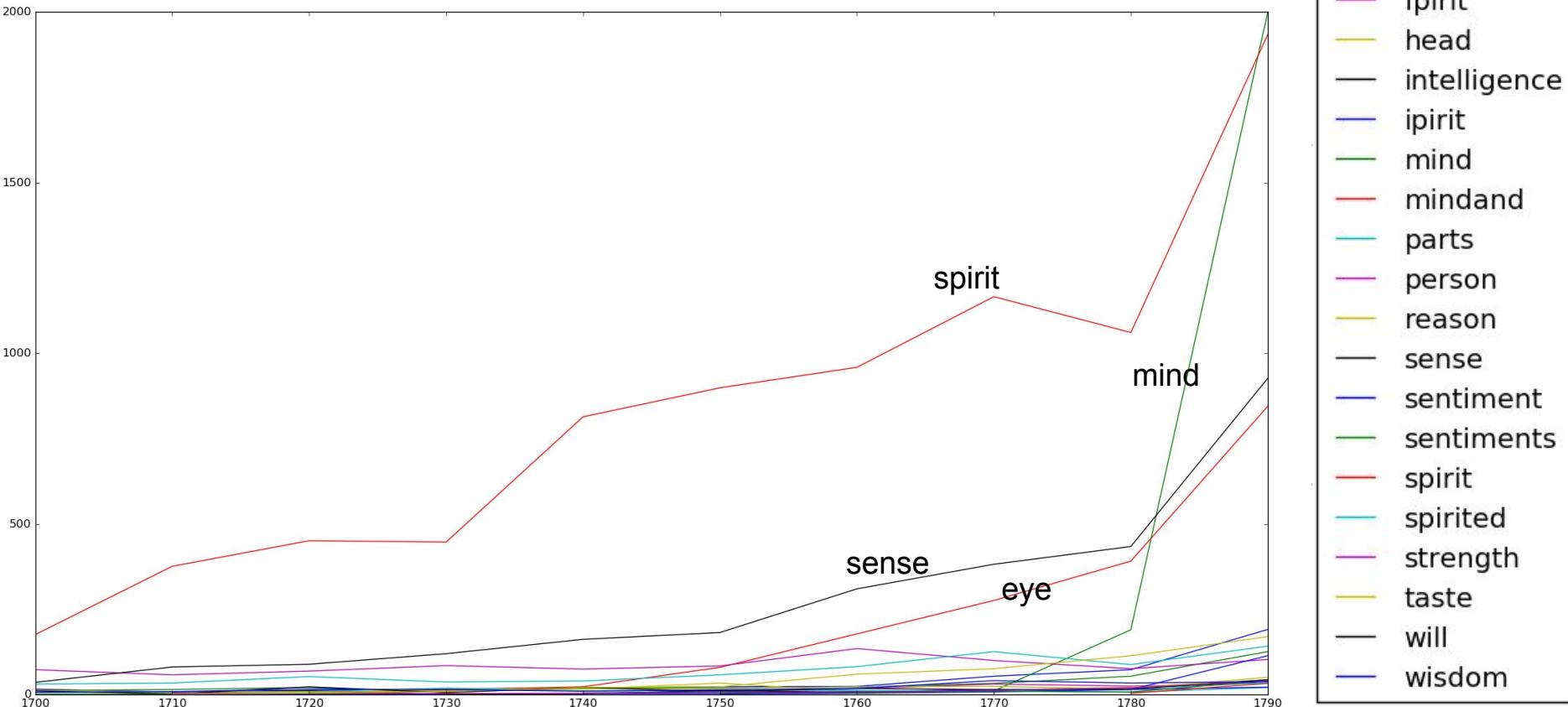
Public opinion, min 5 hits / title, 1770-1799 (authors)



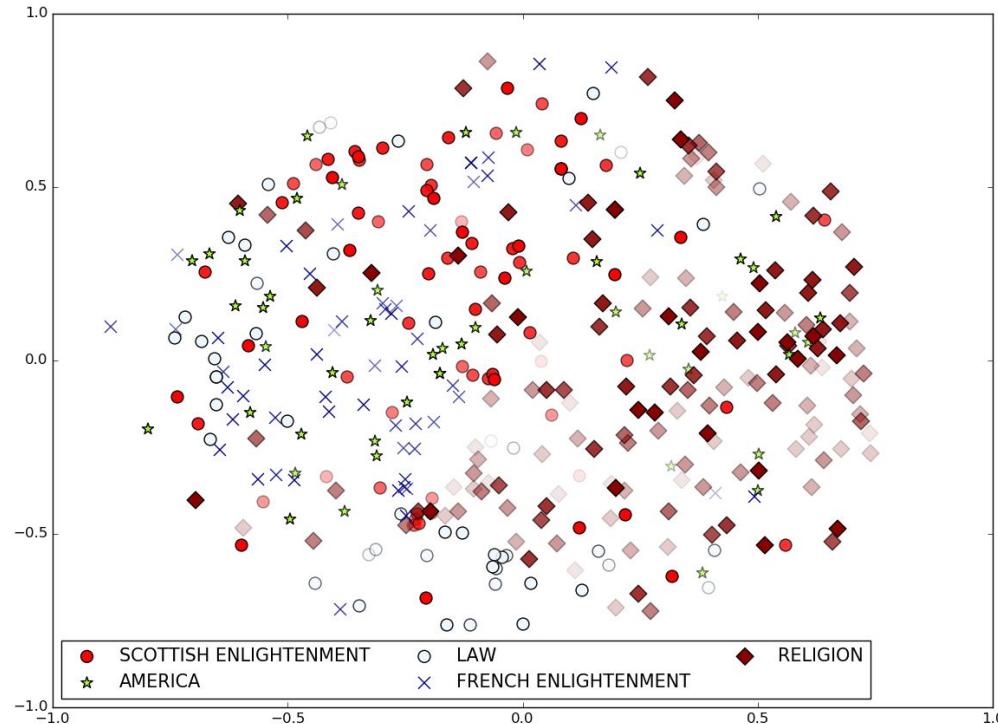
Personified “public” and moral attributes (2grams)



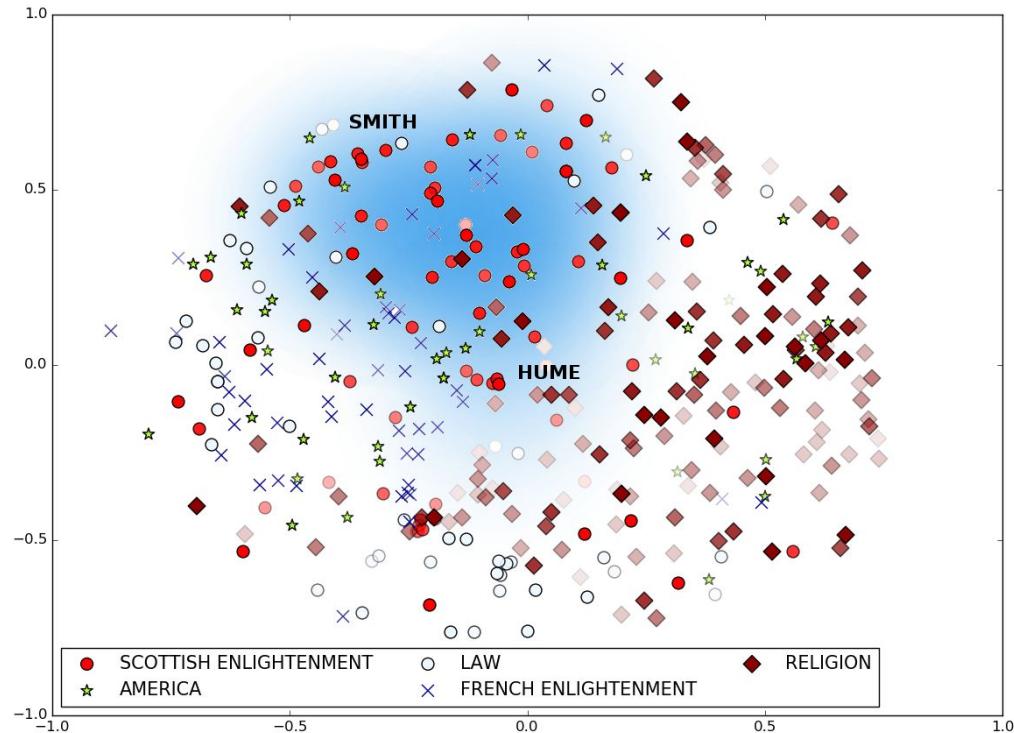
Personified “public” in more detail



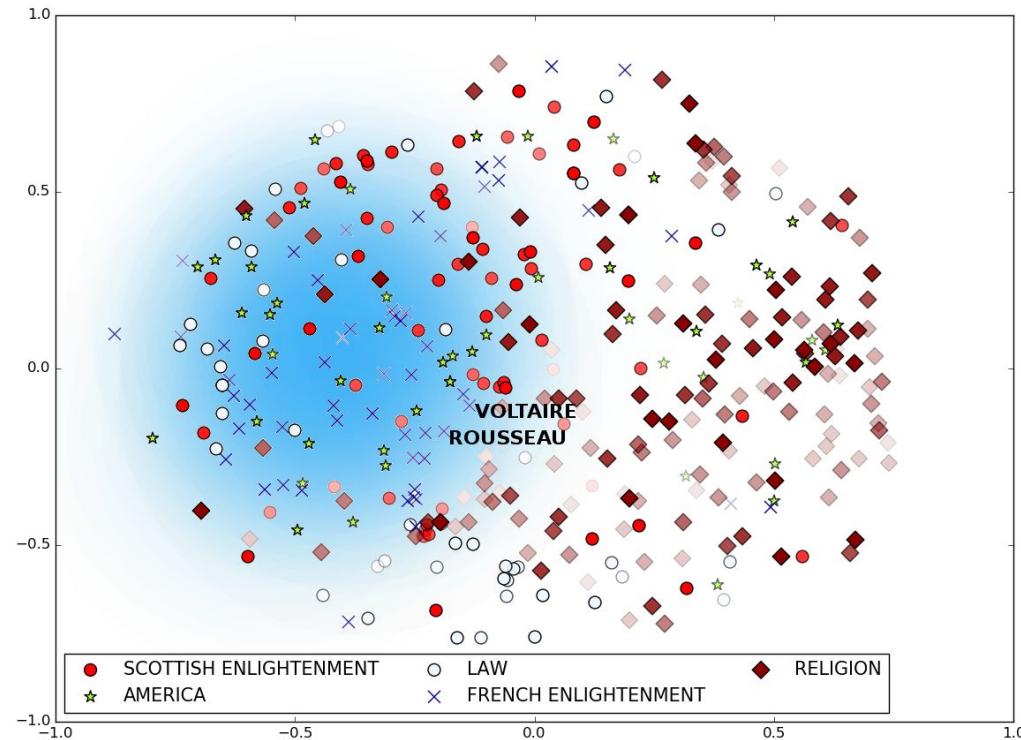
Clustering the uses of “public” in eighteenth century



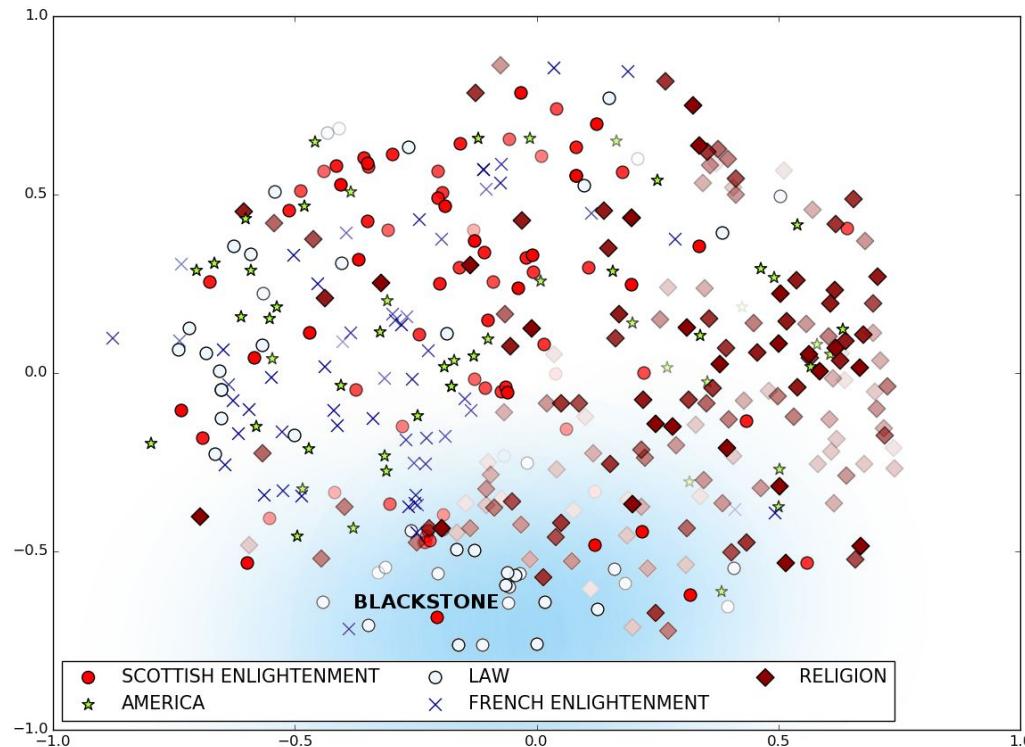
Scottish Enlightenment uses of “public”



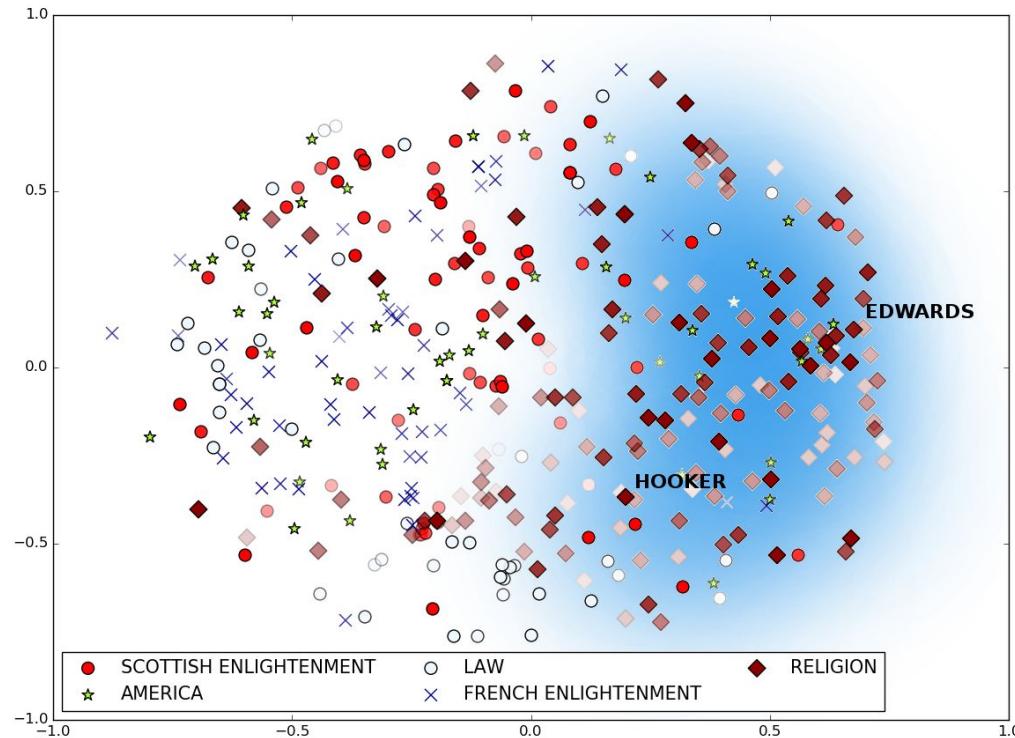
French Enlightenment uses of “public”



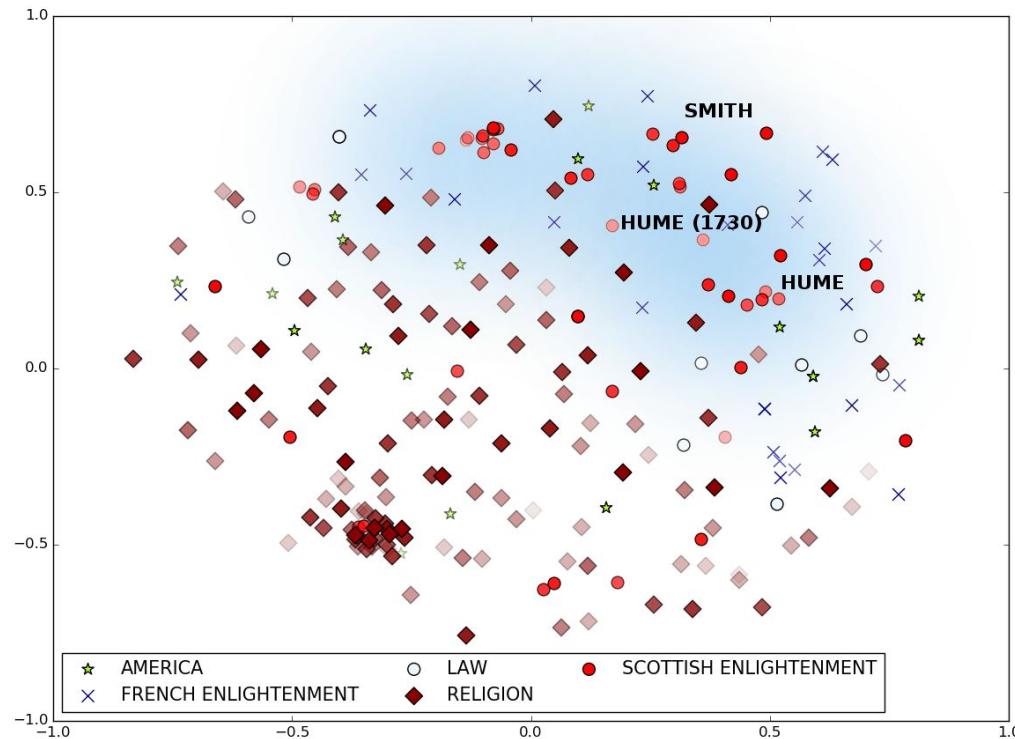
Law cluster of uses of “public”



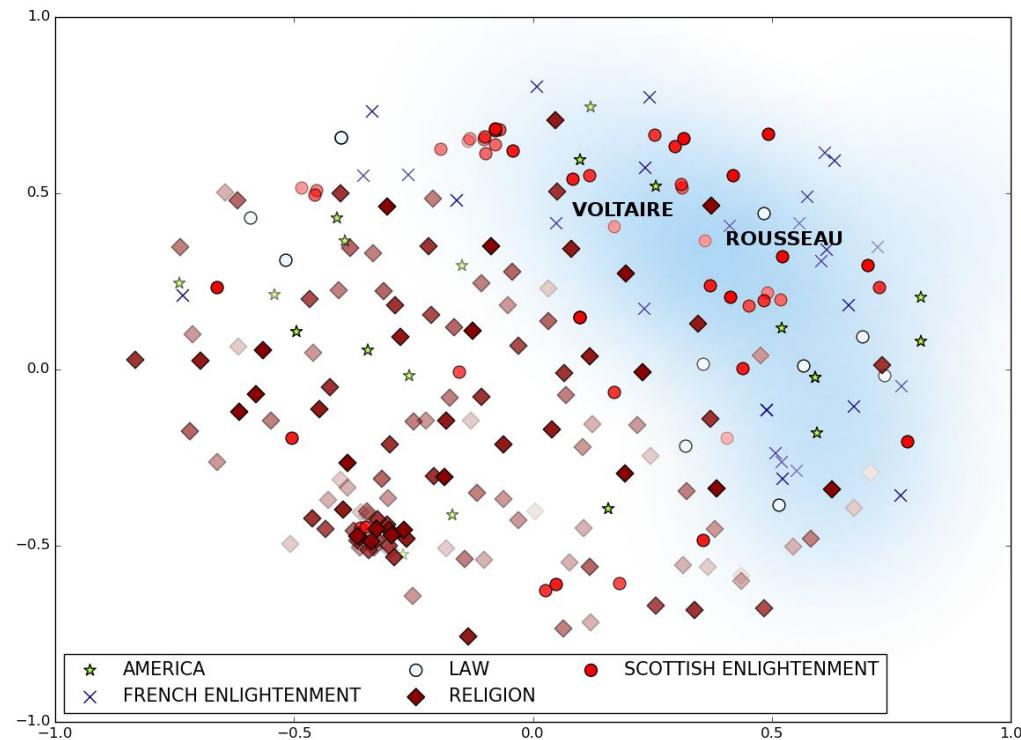
Religious cluster of uses of “public”



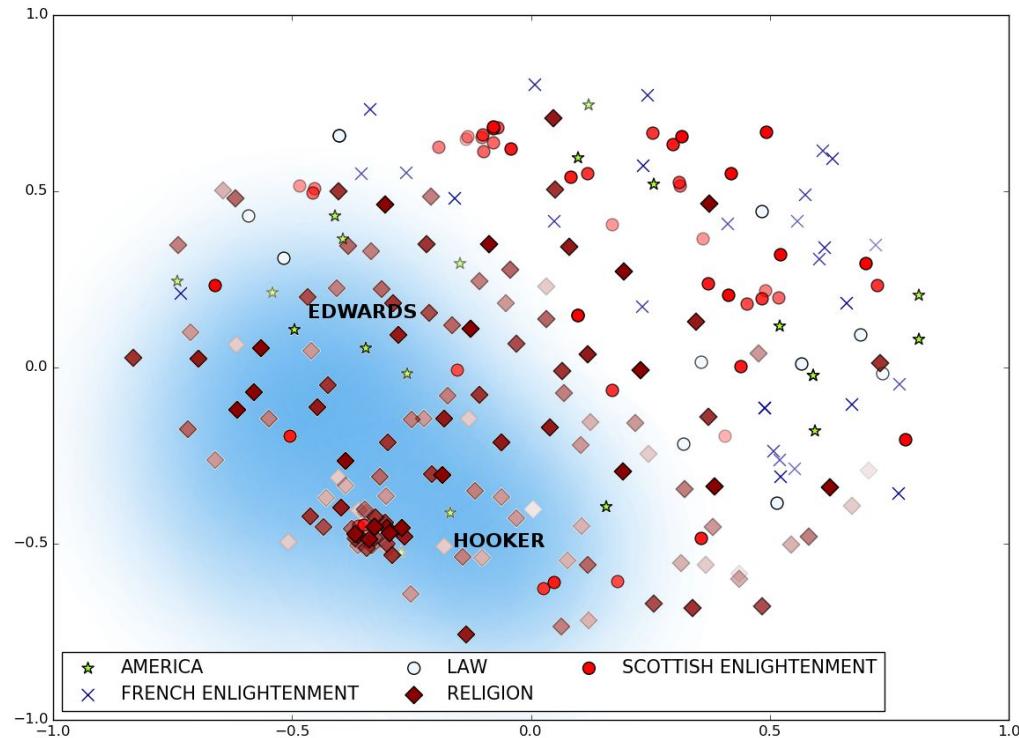
Scottish Enlightenment uses of “moral”



French Enlightenment uses of “moral”



Religious uses of “moral”



extras

COMHIS Collective

The [COMHIS consortium](#) identifies overlooked moments of transformation in public discourse in Finland 1640–1910 by blending historical and computational approaches. The research will reflect on how social change and public discourse are intertwined, and how cultural, institutional, legal and technological changes are reflected both in publication metadata and the textual content of the publications. Particular attention is paid on transparent and reproducible data analysis approaches, published as an [open data analytical ecosystem](#). The newspaper materials digitised by the National Library of Finland up until 1910 are freely available through the [digi.kansalliskirjasto.fi](#) service.

The consortium members include Finnish national library (Centre for Preservation and Digitisation), University of Helsinki (Faculty of Humanities) and University of Turku (Departments of Cultural History and Information Technology).