PUC-Rio, Brazil & Linguateca
Linguateca & University of Oslo, Norway

Gender depiction in Portuguese
Distant reading Brazilian and Portuguese literature

Cláudia Freitas
claudiafreitas@puc-rio.

Diana Santos
d.s.m.santos@ilos.uio.no

Presentation

Introduction
Two aims with this study
Two goals with this presentation
Disclaimer

The material
Distribution in time
Genre description

Basic results
List of characterizing expressions
Relative characterisation
Gender and authorship

Discussion

Two aims

► Study characterizing descriptions of people in Portuguese
  - ► in fiction
  - ► and in other texts by the same authors

Two aims

► Study characterizing descriptions of people in Portuguese
  - ► in fiction
  - ► and in other texts by the same authors
► See how they are distributed when depicting feminine and masculine characters

Hopes

▶ Our study may be a way to uncover bias, and also a way to find especially interesting descriptions that may have changed with time, but mainly it is a first step in providing the linguistic means for further, and more literary inspired, studies.

▶ It is a first step about gender in Portuguese (literature), that can be improved and criticized, and hopefully foster more research on the subject with a distant reading approach.

▶ We identify what was possible to find using rather simple descriptions, which may be interpreted in different ways and with different biases by different researchers.

Goals of this presentation

▶ to present the methodology and the resources,

▶ to bring one other language to computational literary studies: Portuguese

Disclaimer

We have no strong opinions as to gender, and no agenda

- ▶ Gender is a social construct, so no need for gender to be binary

- ▶ We are aware that "cultural analytics" works have been blamed for naively reinforcing binarity in gender studies by e.g. Mandell (2019)

- ▶ But we are also aware that most of the literature we survey and study precludes these ideas and analyses, that is, in those works and those times, there were two sexes and two genders, something which is very easy to trace given (binary) morphological gender in Portuguese

- ▶ By making our data available we give other researchers the opportunity to explore other classification schemes where e.g. different genders are taken into consideration

---

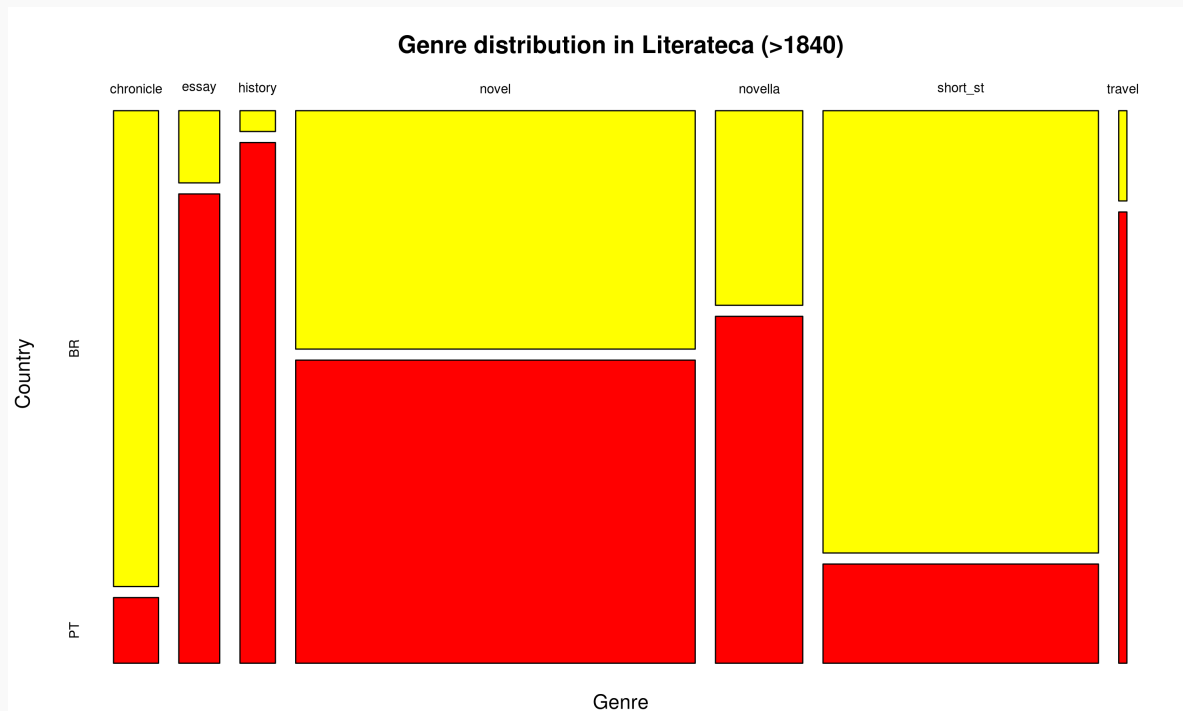# Corpora of literature in Portuguese

We use in this study **Literateca**

- ▶ literary works and/or works written by literary authors

- ▶ including Portuguese and Brazilian authors

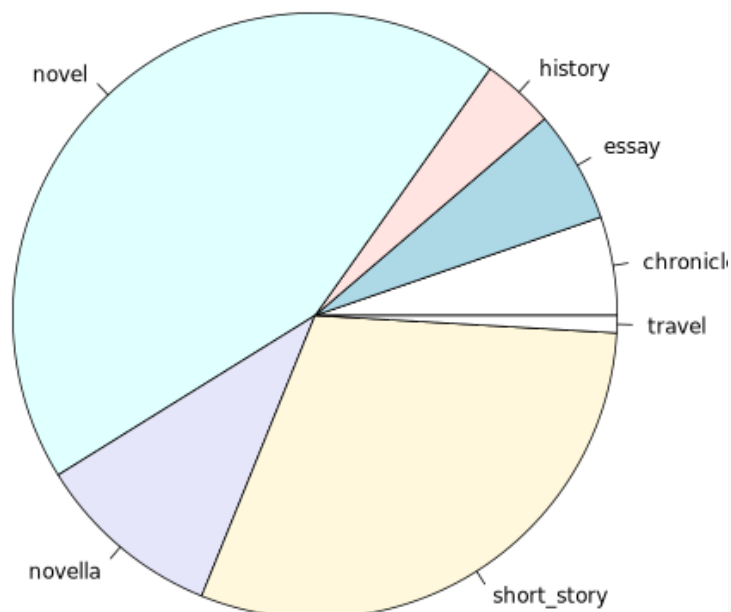- ▶ imbalanced according to country, to time, to author gender
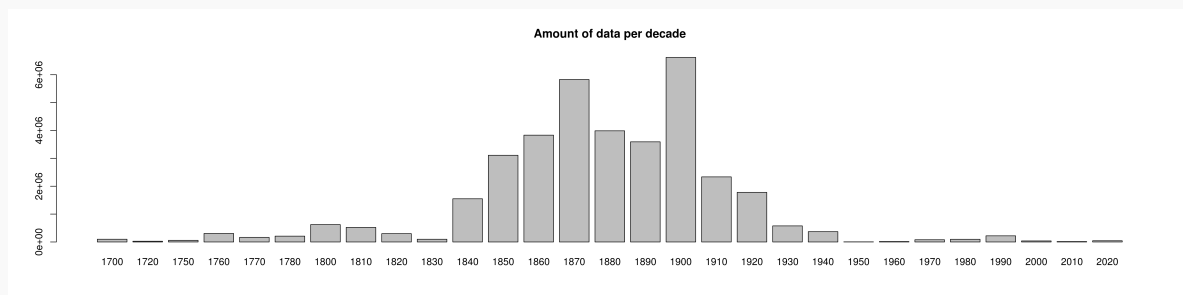
Genre distribution per works

Genre distribution in Literateca (>1840)

Genre distribution per words

1. people?: one reviewer questioned the conflation of (fictional) people with real people, and argued that descriptions of these two kinds of humans should not be confused. However, they are descriptions by the same people (authors). Anyway, we also study separately descriptions in fiction.
   - ► We are also aware that literature is not a true description of society, but that literary authors do have an important influence in the minds of the readers of their time, namely the elites in Portugal and Brazil.
2. gender vs. sex: while we say we are interested in gender, we use mainly morphological gender of (proper) nouns as identifiable categories.

List of characterizing expressions

A lexicon of characterizing expressions

Circa 4,500 ways of characterizing a person, divided in 4 categories, plus one other category. Classified out of context.

Number of depicting word (types)

social	1391
appearance	672
emotional	514
character	1578
other	326
total	4481

Search patterns for characterisarion

We developed 29 complex search patterns which we also make available

In the whole material
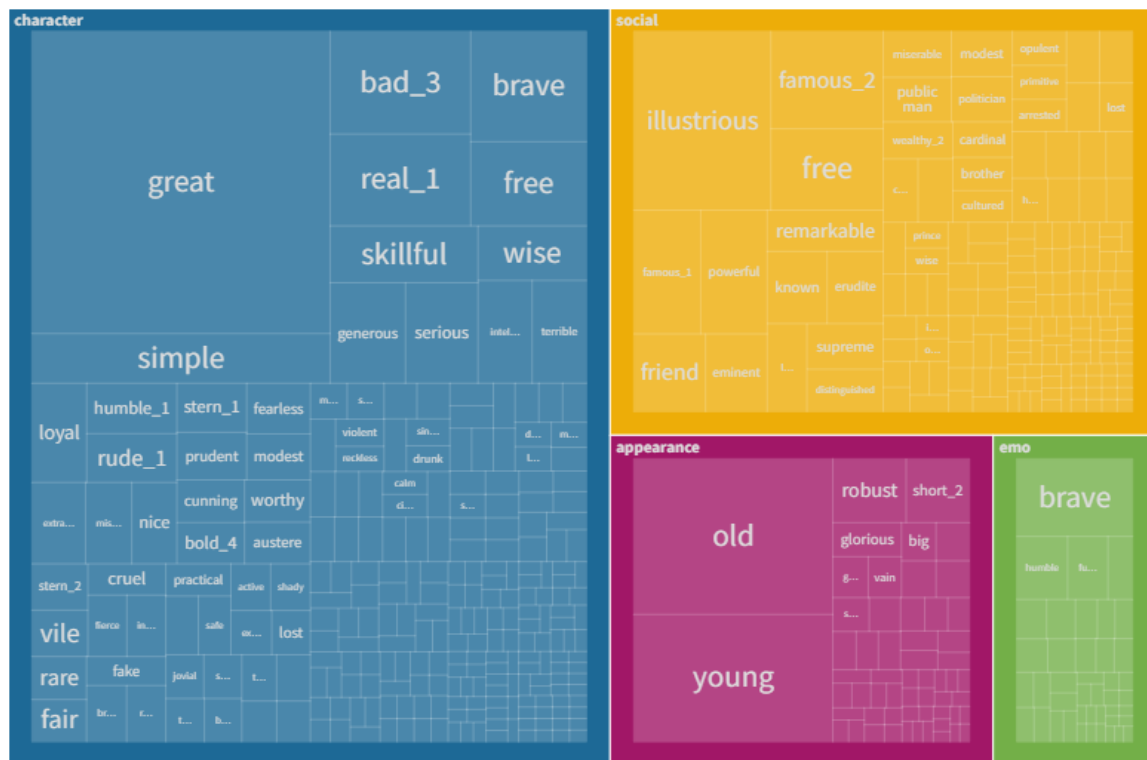
Masculine

Feminine
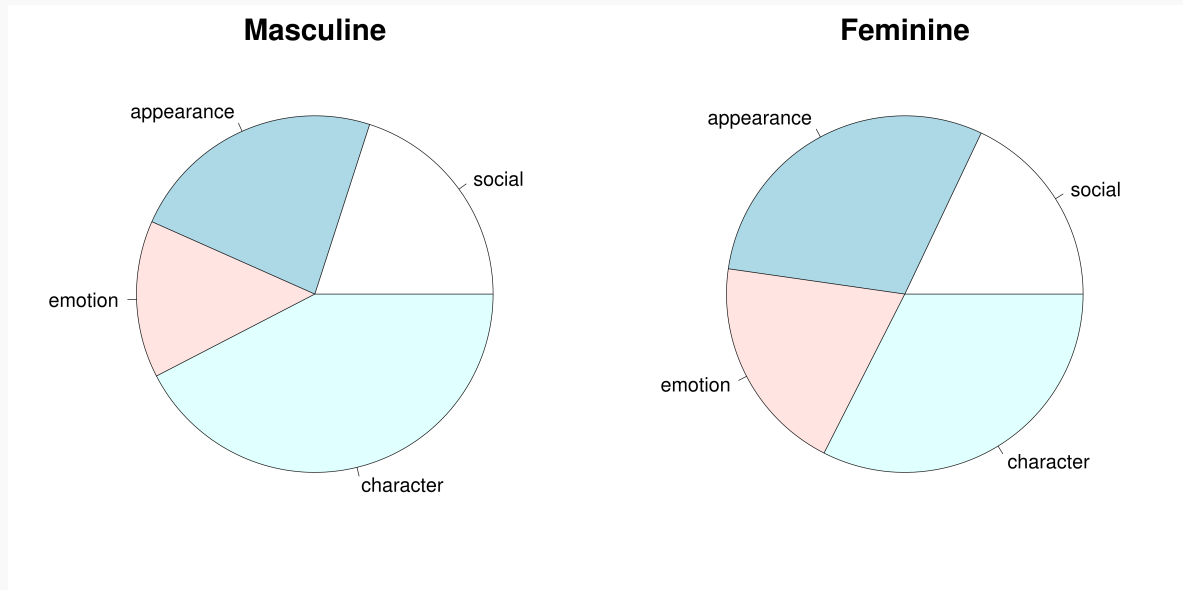
Women are more frequently described by their appearance; men by their character.

But is this all? Lexical preferences
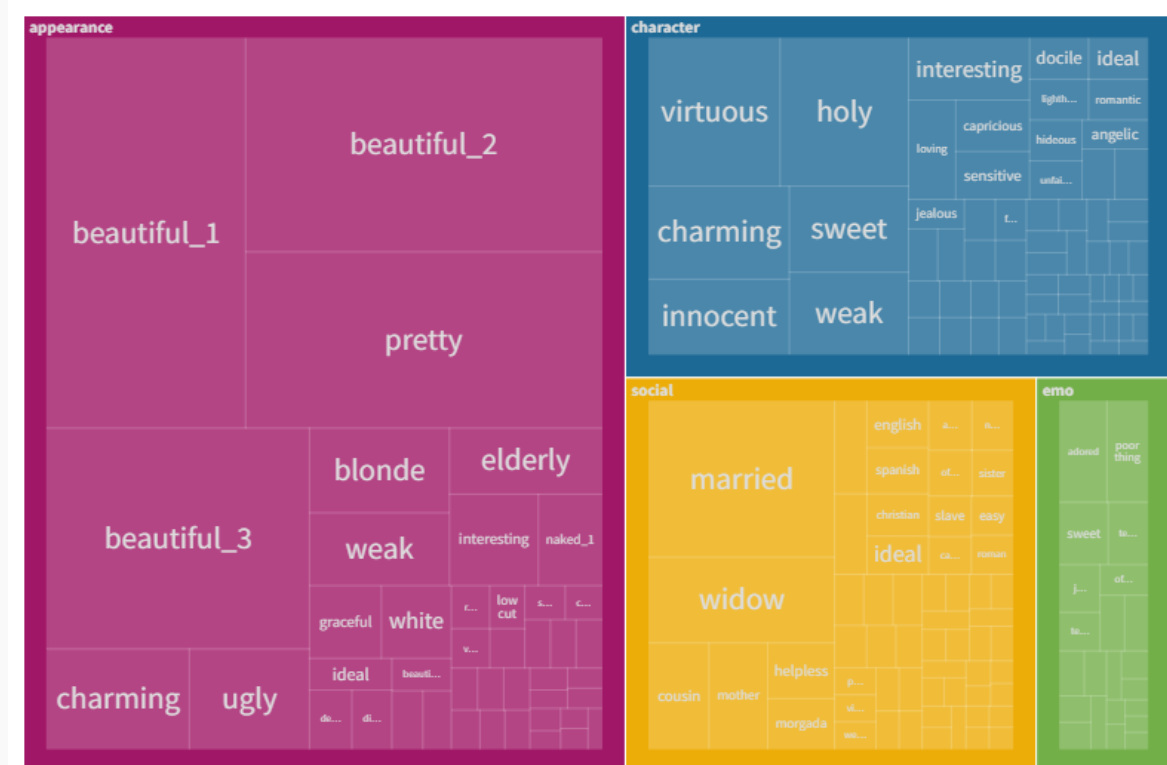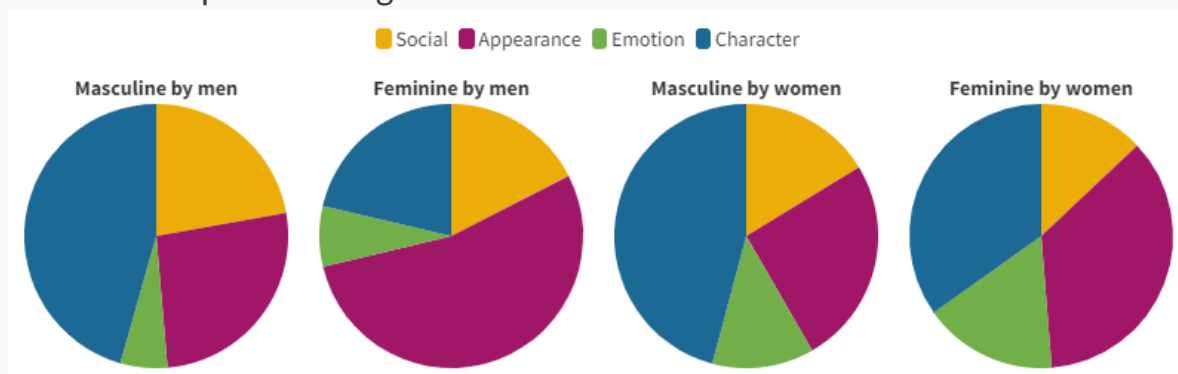
But is this all? Lexical preferences
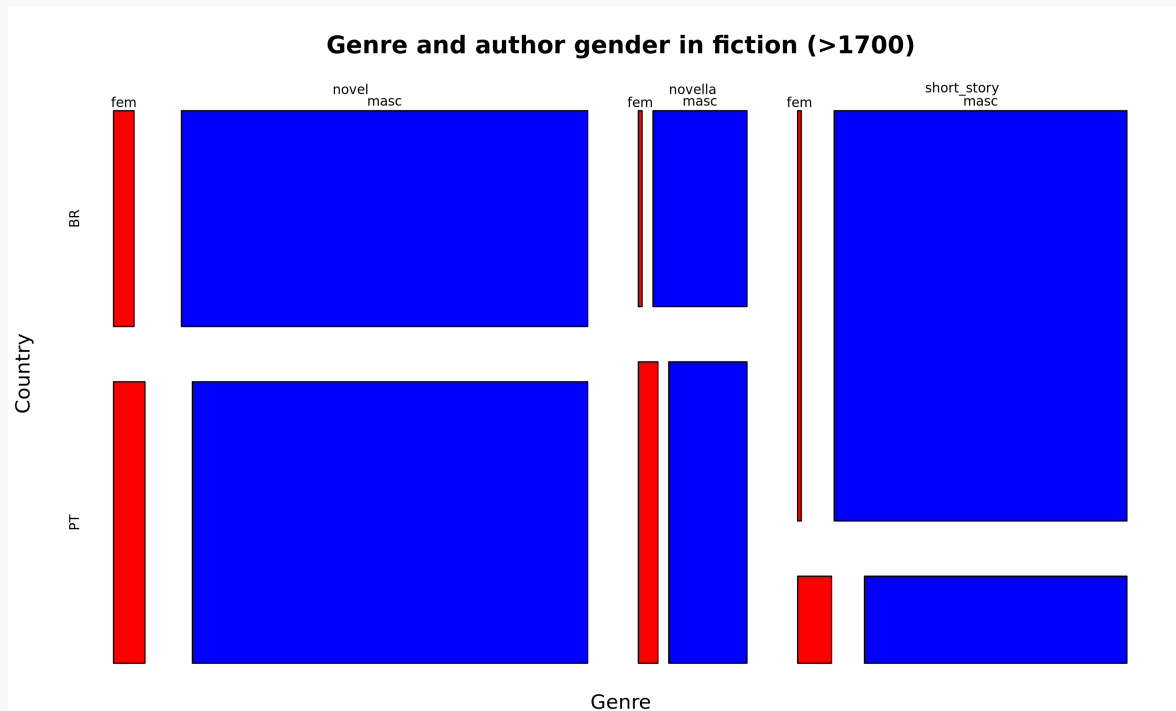
Gender and authorship

Preferences per author gender.

When we aggregate all authors, we mainly see the masculine view.
However, we have far too few women writers to be able to generalize.

Only fiction

Genre and author gender in fiction (>1700)

Discussion of the material

1. we have not included drama nor poetry, for two reasons: fear of different syntactic structures and therefore poorer parser performance, and gross imbalance across varieties (most came from Portugal)

2. we have included other text written by the same authors when it was available (literature scholars had done the digitizing work) (but we also did studies only on fiction)

3. we removed translated material and material from varieties other than Brazilian and Portuguese

4. we used 1700 as a cutoff point, except when the studies compared the two varieties, in which case we used 1840 (Brazil became independent in 1822)
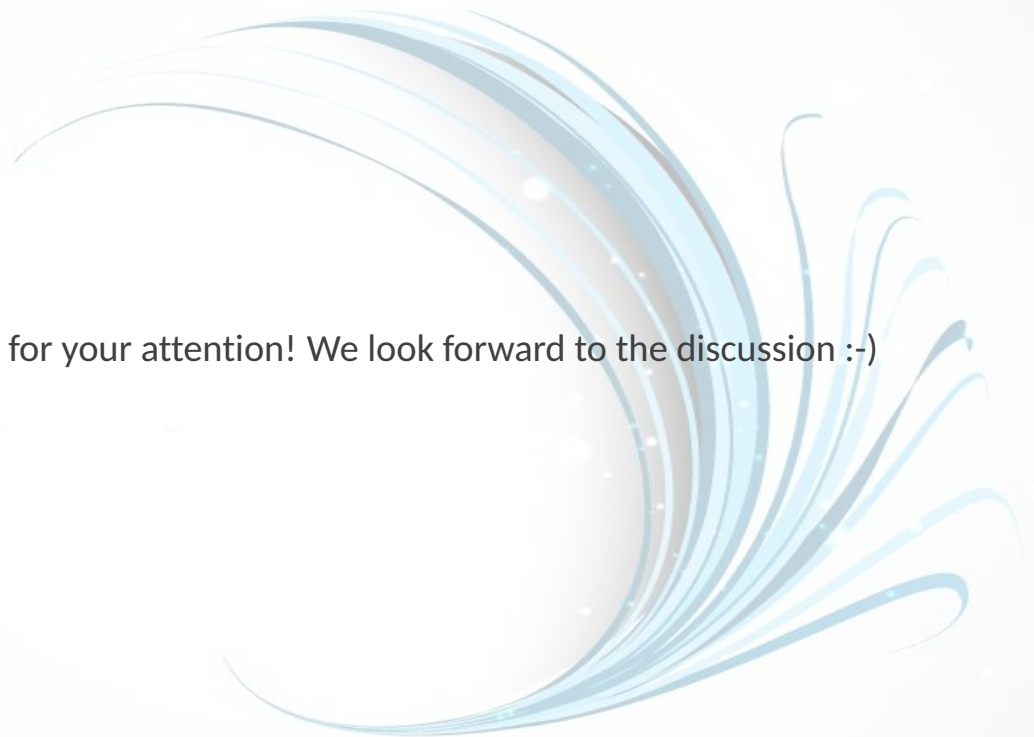
All these decisions are arguable, but readers can redo the work with other subsets, since the material (and the commands) are available.
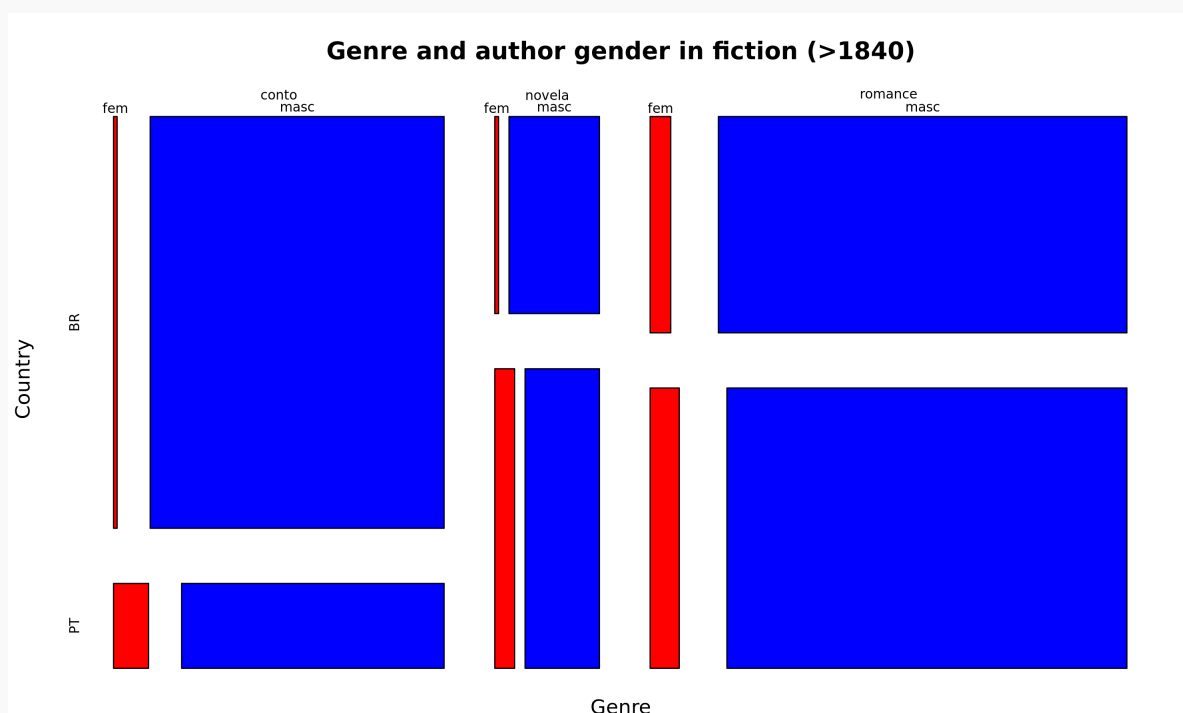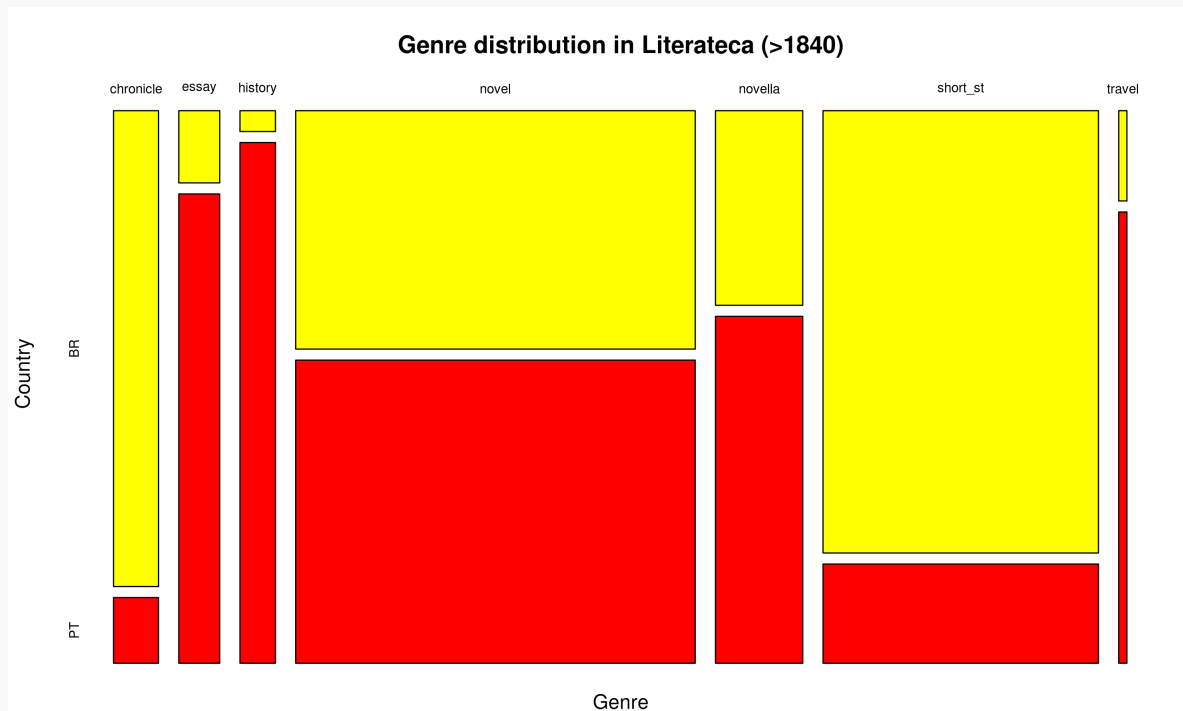
Some other themes that we find interesting to pursue, in future work:

▶ Are there authors specially mysogenous in lusophone literature?

▶ Are there intrinsically gender-related adjetives that should be looked into more specifically? Such as *varonil*, *feminino*, *masculino*, *efeminado*, *mulheril*, ...

▶ Can one identify gender orientation using adjectives/characterizing words?

▶ Have there been clear changes in depicting masculine and feminine characters through time?

▶ Are there obvious differences in gender depiction between translated and original literature?

▶ Are there other depicting categories, like age, or civil status, that should be studied separately?

▶ What novelty can implied descriptions (eg, sad eyes) offer?

Thank you for your attention! We look forward to the discussion :-)

After 1840

Genre distribution in Literateca (>1840)

Fiction after 1840

Genre and author gender in fiction (>1840)

Different phases and objects in DH

Physical objects

pdf

text

corpus

database

visualisation