

Metadata from ERB

ERB metadata

Abstract

Estonian National Bibliography is a metadata set that aims to collect information on all publications written in any language in Estonia and all texts written in Estonian in whichever country. In this set, only the publications in Estonian language are used. The dataset has been compiled in digital format since 2002. and aggregates the work of multiple institutions and generations in collecting the publication information.

This dataset presents the Estonian National Bibliography dataset in wide instead of long format used in Marc21, with some of the variables that may be useful for text-mining studies. It includes information about some variables have also been standardized

Information on coding the variables can be found here:

Helpful information on the metadata available can be found here: <http://data.digar.ee/#page5>

The rules followed in adding information on older books can be found here: http://www.elnet.ee/images/pdf/juhendid/vanaraamat_MARC21.pdf

Intro

Converts the dataset into tidy format

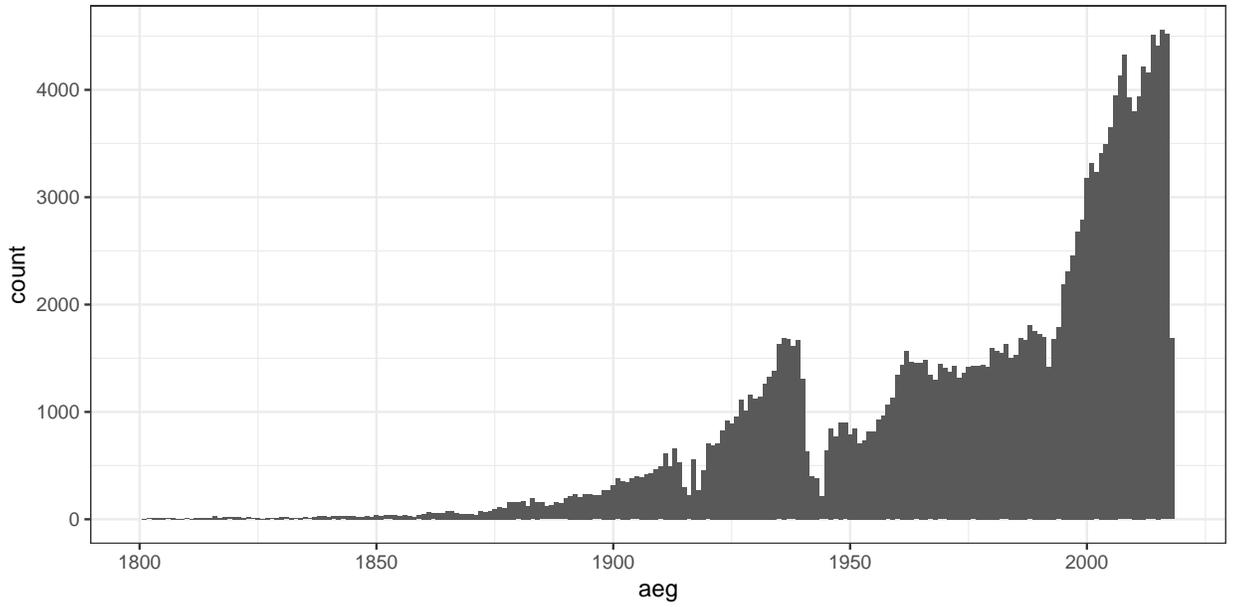
```
## -- Attaching packages ----- tidyverse 1.2.1 --

## √ ggplot2 3.0.0    √ purrr  0.2.5
## √ tibble  1.4.2    √ dplyr  0.7.6
## √ tidyr   0.8.1    √ stringr 1.3.1
## √ readr   1.1.1    √ forcats 0.3.0

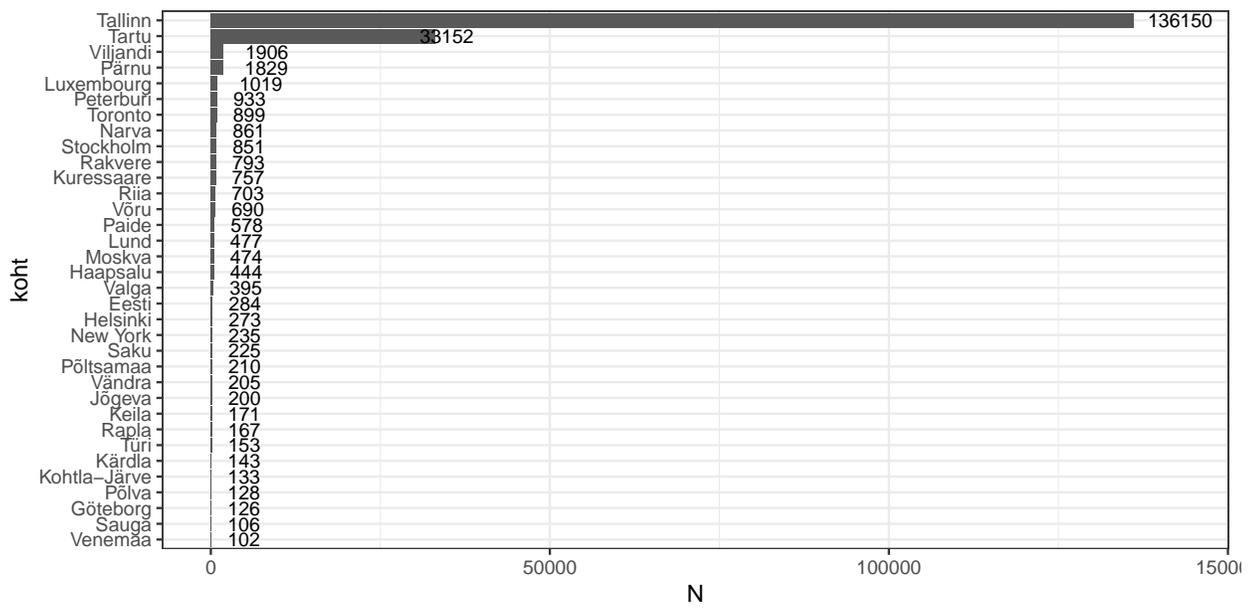
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::between() masks data.table::between()
## x dplyr::filter()  masks stats::filter()
## x dplyr::first()   masks data.table::first()
## x dplyr::lag()     masks stats::lag()
## x dplyr::last()    masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

Summary

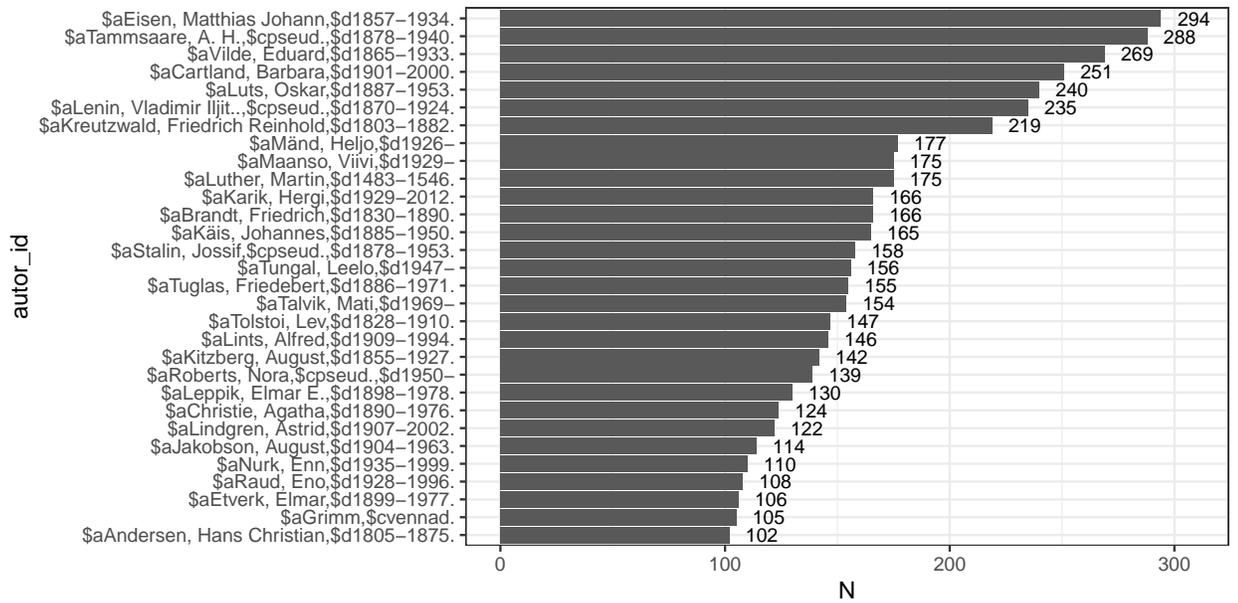
The dataset has altogether information on 193848 printed items. Over time the distribution is the following.



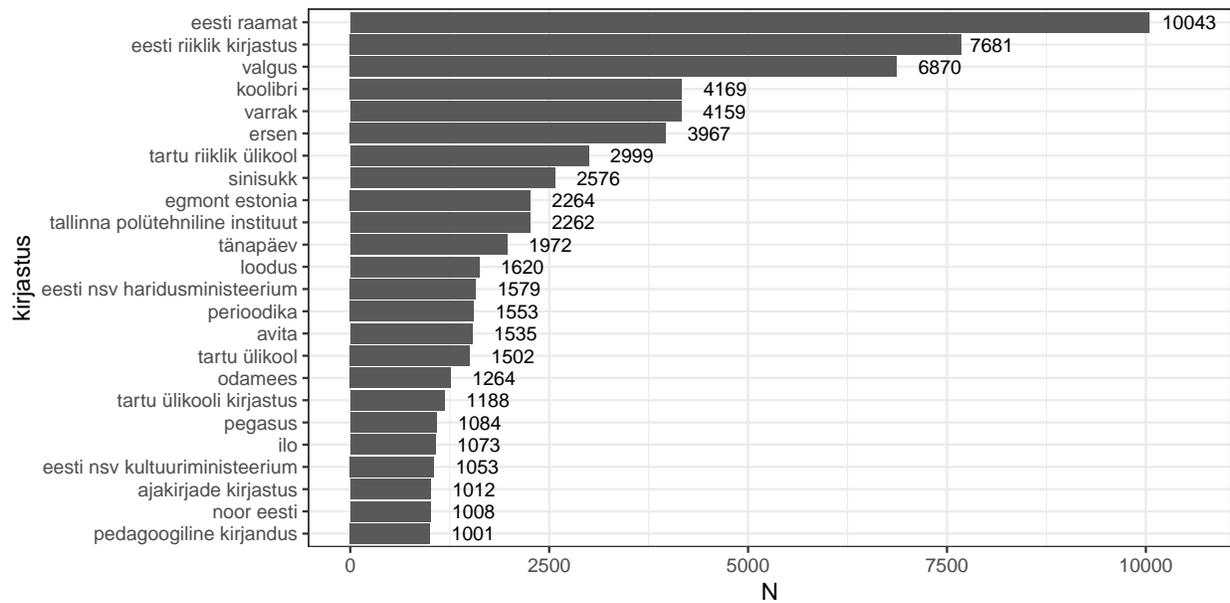
data-1.bb



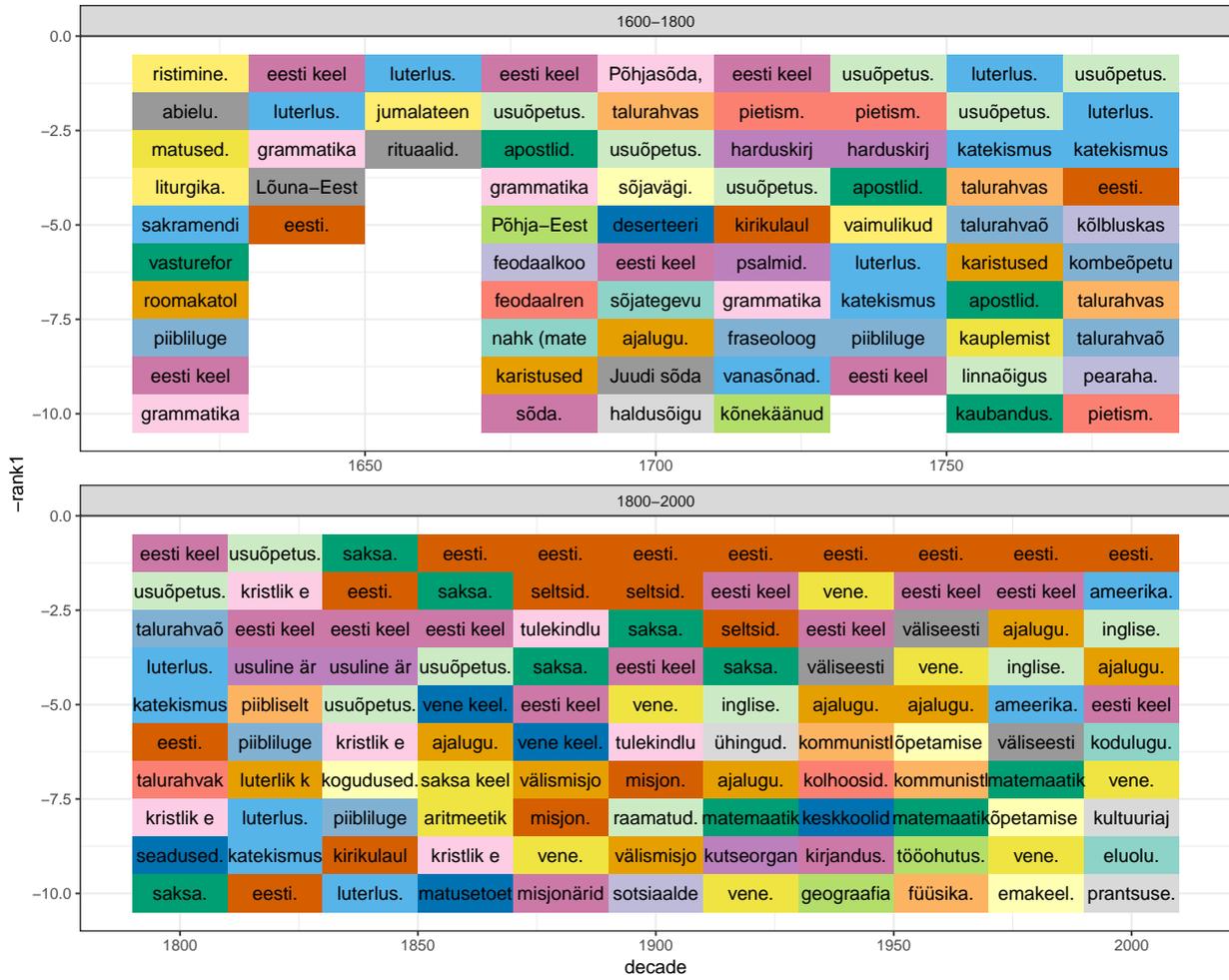
cities-1.bb



cities-2.bb



publishers-1.bb



genres-1.bb