Estimation Plots

Chris Bentz

03/01/2023

Load libraries

If the libraries are not installed yet, you need to install them using, for example, the command: install.packages("ggplot2"). For the Hrate package this is different, since it comes from github. The devtools library needs to be installed, and then the install_github() function is used.

```
library(stringr)
library(ggplot2)
library(ggrepel)
library(plyr)
library(ggExtra)
library(ggpubr)
```

```
##
## Attaching package: 'ggpubr'
```

```
## The following object is masked from 'package:plyr':
##
##     mutate
```

Load Data

Load data table with values per text file.

```
# load estimations from stringBase corpus
estimations10.df <- read.csv("~/Github/NaLaFi/results/estimation10chars.csv")
estimations100.df <- read.csv("~/Github/NaLaFi/results/estimation100chars.csv")
estimations1000.df <- read.csv("~/Github/NaLaFi/results/estimation1000chars.csv")
#head(estimations10.df.)
```

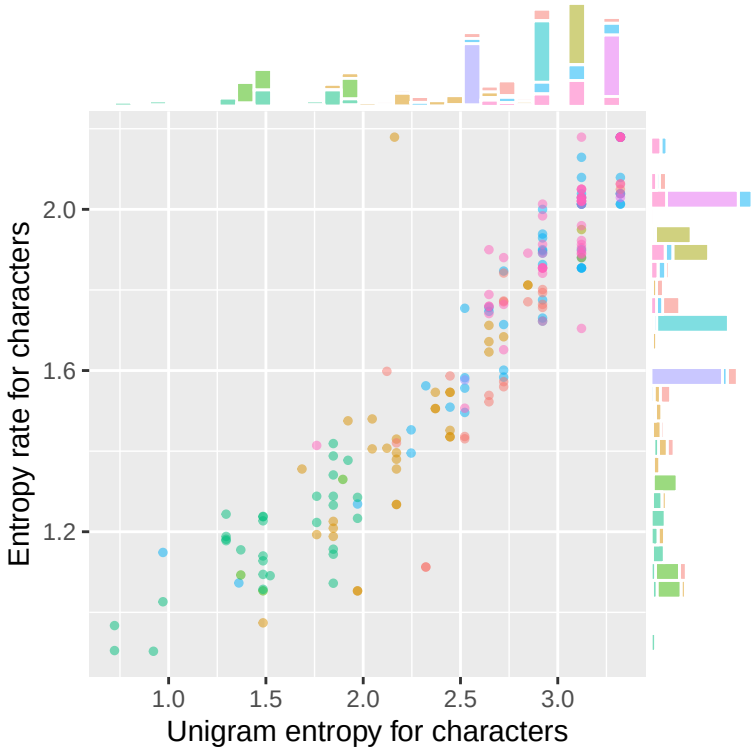
Select subcorpora (if needed).

```
selected <- c("shuffled")
estimations10.df <- estimations10.df[!(estimations10.df$subcorpus %in% selected), ]
```

Scatterplots

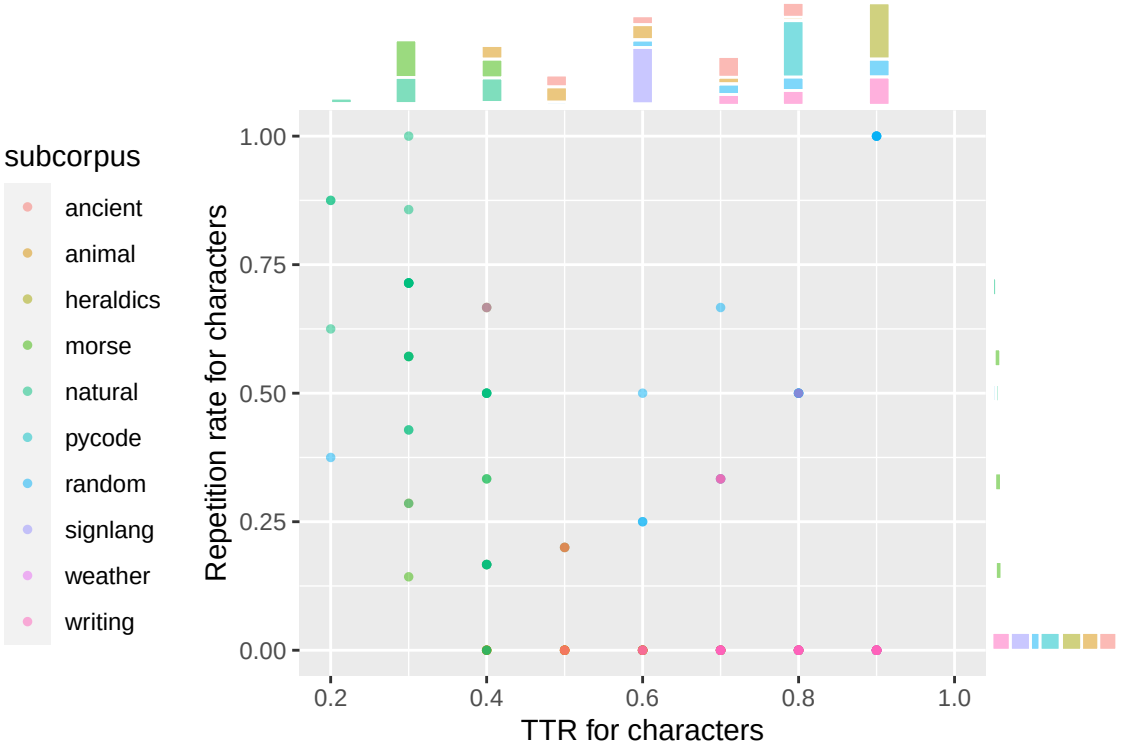
Entropy rate vs. unigram entropy for characters

```
# plot
huni.hrate.chars.plot <- ggplot(estimations10.df,
                                aes(x = huni.chars, y = hrate.chars,
                                    colour = subcorpus)) +
  geom_point(alpha = 0.5, size  = 1) +
  labs(x = "Unigram entropy for characters", y = "Entropy rate for characters") +
  theme(legend.position = "none")
huni.hrate.chars.plot <- ggMarginal(huni.hrate.chars.plot, groupFill = T, type = "histogram", colour =
huni.hrate.chars.plot
```



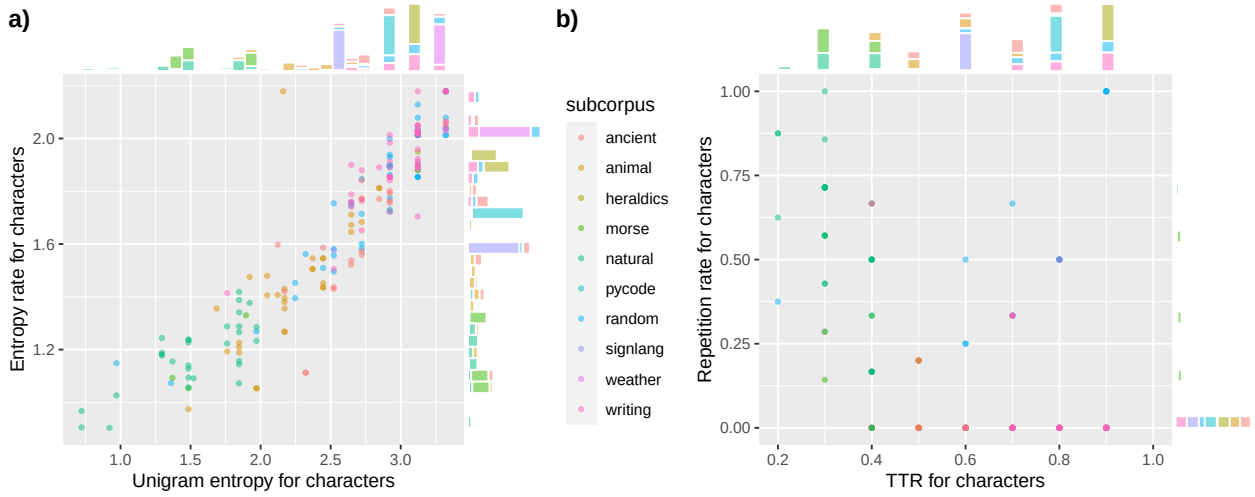
TTR vs. repetition rate for characters

```
ttr.rm.chars.plot <- ggplot(estimations10.df,
                    aes(x = ttr.chars, y = rm.chars,
                    colour = subcorpus)) +
  geom_point(alpha = 0.5, size  = 1) +
  theme(legend.position = "left") +
  labs(x = "TTR for characters", y = "Repetition rate for characters")
ttr.rm.chars.plot <- ggMarginal(ttr.rm.chars.plot, groupFill = T, type = "histogram", colour = "white")
ttr.rm.chars.plot
```



Combined Plots

```
plots.combined <- ggarrange(huni.hrate.chars.plot, ttr.rm.chars.plot,
                  labels = c("a)", "b)"),
                  ncol = 2, nrow = 1, widths = c(1, 1.3) )
plots.combined
```



Safe complete figure to file

```r
ggsave("~/Github/NaLaFi/figures/plots_combined.pdf", plots.combined, width = 10, height = 4, dpi = 300,
       scale = 1, device = cairo_pdf)
```