

Estimations of Quantitative Measures: StringBase Corpus

Chris Bentz

27/09/2021

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(entropy)
library(ggExtra)
library(gsubfn)
```

```
## Loading required package: proto
```

```
#library(devtools)
#install_github("dimalik/Hrate")
library(Hrate)
```

List files

Create list with all the files in the directory “corpus”.

```
# file list for stringBase texts
file.list <- list.files(path = "~/Github/StringBase/corpus/original",
                        recursive = T, full.names = T)
head(file.list)
```

```
## [1] "/home/chris/Github/StringBase/corpus/original/animal/animal_bhg_0001.txt"
## [2] "/home/chris/Github/StringBase/corpus/original/animal/animal_bhg_0002.txt"
## [3] "/home/chris/Github/StringBase/corpus/original/animal/animal_bhg_0003.txt"
## [4] "/home/chris/Github/StringBase/corpus/original/animal/animal_bhg_0004.txt"
## [5] "/home/chris/Github/StringBase/corpus/original/animal/animal_bhg_0005.txt"
## [6] "/home/chris/Github/StringBase/corpus/original/animal/animal_bhg_0006.txt"
```

```
length(file.list)
```

```
## [1] 183
```

Character entropy calculation per file

```
# set counter
counter = 0
# set the maximal number of units (n) to be used for analysis
n = 100
# initialize dataframe to append results to
estimations.df <- data.frame(filename = character(0), subcorpus = character(0),
                             code = character(0), huni.chars = numeric(0),
                             huni.strings = numeric(0), hrate.chars = numeric(0),
                             hrate.strings = numeric(0), ttr.chars = numeric(0),
                             ttr.strings = numeric(0))

# start time
start_time <- Sys.time()
for (file in file.list)
{
  # basic processing
  # loading textfile
  textfile <- scan(file, what = "char", quote = "", comment.char = "",
                  encoding = "UTF-8", sep = "\n", skip = 7)
  # remove annotations marked by '<>'
  textfile <- gsub("<.*>", "", textfile)
  # print(head(textfile))
  # get filename
  filename <- basename(file)
  print(filename) # for visual inspection
  # get subcorpus category
  subcorpus <- sub("_.*", "", filename)
  # print(subcorpus) # for visual inspection
  # get the three letter identification code + the running number
  code <- substring(substring(filename, regexpr("_", filename) + 1), 1, 8)
  # print(code) # for visual inspection
  # Split into individual characters/signs
  # distinguish between ancient languages and other files
  if(subcorpus == "ancient"){
    # remove tabs and parentheses
    textfile <- str_replace_all(textfile, c("\\t" = "", "\\(" = "", "\\)" = "", "\\]" = "", "\\[" = "",
    #textfile <- gsubfn(".", list("\t" = "", "(" = "", ")" = "", "]" = "",
    #      "[" = "" ), textfile)
    # split textfile into strings of utf-8 character which represent
    # ancient signs. We here use hypens and white spaces as delimiters.
    chars <- unlist(strsplit(textfile, "\\-|\\s|\\{|\\}|\\}"))
    chars <- chars[1:n] # use only maximally n units
    chars <- chars[!is.na(chars)] # remove NAs for vectors which are already shorter
    # than n
    chars <- chars[chars != ""] # remove empty strings from character vector
    # Split the textfile into strings delimited by white spaces. The output of strsplit()
    # is a list, so it needs to be "unlisted" to get a vector.
    strings <- unlist(strsplit(textfile, "\\s|\\{|\\}|\\}"))
    strings <- strings[1:n] # use only maximally n units
    strings <- strings[!is.na(strings)] # remove NAs for vectors which are already
    strings <- strings[strings != ""] # remove empty strings from character vector
```

```

# shorter than n
} else {
# remove tabs and parentheses
textfile <- str_replace_all(textfile, c("\\\\t" = "", "\\(" = "", "\\)" = "",
                                     "\\]" = "", "\\[" = "", "\\}" = "", "\\{" = ""))
#textfile <- gsubfn(".", list("\t" = "", "(" = "", ")" = "", "]" = "",
#                             "[" = "", "{" = "", "}" = ""), textfile)
# split the textfile into individual utf-8 characters. Note that white spaces are
# counted as utf-8 characters here.
chars <- unlist(strsplit(textfile, ""))
chars <- chars[1:n] # use only maximally n units
chars <- chars[!is.na(chars)] # remove NAs for vectors which are already shorter
# than n
# chars <- chars[chars != " "] # remove white spaces from character vector
# Split the textfile into strings delimited by white spaces. The output of strsplit()
# is a list, so it needs to be "unlisted" to get a vector.
strings <- unlist(strsplit(textfile, " "))
strings <- strings[1:n] # use only maximally n units
strings <- strings[!is.na(strings)] # remove NAs for vectors which are already
# shorter than n
}

# unigram entropy estimation
# calculate unigram entropy for characters
chars.df <- as.data.frame(table(chars))
# print(chars.df)
huni.chars <- entropy(chars.df$Freq, method = "ML", unit = "log2")
# calculate unigram entropy for units (Maximum Likelihood method)
strings.df <- as.data.frame(table(strings))
# print(units.df)
huni.strings <- entropy(strings.df$Freq, method = "ML", unit = "log2")

# entropy rate estimation
# the values chosen for max.length and every.word will crucially
# impact processing time. In case of "max.length = NULL" the length is n units
hrate.chars <- get.estimate(text = chars, every.word = 1, max.length = NULL)
hrate.strings <- get.estimate(text = strings, every.word = 1, max.length = NULL)

# calculate type-token ratio (ttr)
ttr.chars <- nrow(chars.df)/sum(chars.df$Freq)
ttr.strings <- nrow(strings.df)/sum(strings.df$Freq)

# calculate repetition measure according to Sproat (2014)
# the overall number of repetitions is the sum of frequency counts minus 1.
R <- sum(chars.df$Freq-1)
# calculate the number of adjacent repetitions
r = 0
if (length(chars) > 1){
  for (i in 1:(length(chars)-1)){
    if (chars[i] == chars[i+1]){
      r = r + 1
    } else {
      r = r + 0
    }
  }
}

```

```

    }
  }
  # calculate the repetition measure
  rm.chars <- r/R
} else {
  rm.chars <- "NA"
}

# append results to dataframe
local.df <- data.frame(filename, subcorpus, code, huni.chars,
                        hrate.chars, huni.strings, hrate.strings,
                        ttr.chars, ttr.strings, rm.chars)
estimations.df <- rbind(estimations.df, local.df)
# counter
counter <- counter + 1
# print(counter)
}

```

```

## [1] "animal_bhg_0001.txt"
## [1] "animal_bhg_0002.txt"
## [1] "animal_bhg_0003.txt"
## [1] "animal_bhg_0004.txt"
## [1] "animal_bhg_0005.txt"
## [1] "animal_bhg_0006.txt"
## [1] "animal_bhg_0007.txt"
## [1] "animal_bhg_0008.txt"
## [1] "animal_bhg_0009.txt"
## [1] "animal_bhg_0010.txt"
## [1] "animal_cad_0001.txt"
## [1] "animal_cad_0002.txt"
## [1] "animal_cav_0001.txt"
## [1] "animal_cav_0002.txt"
## [1] "animal_cav_0003.txt"
## [1] "animal_cav_0004.txt"
## [1] "animal_cav_0005.txt"
## [1] "animal_cav_0006.txt"
## [1] "animal_cav_0007.txt"
## [1] "animal_cav_0008.txt"
## [1] "animal_cav_0009.txt"
## [1] "animal_cth_0001.txt"
## [1] "animal_cth_0002.txt"
## [1] "animal_cth_0003.txt"
## [1] "animal_cth_0004.txt"
## [1] "animal_cth_0005.txt"
## [1] "animal_cth_0006.txt"
## [1] "animal_cth_0007.txt"
## [1] "animal_cth_0008.txt"
## [1] "animal_cth_0009.txt"
## [1] "animal_cth_0010.txt"
## [1] "animal_cth_0011.txt"
## [1] "animal_zfi_0001.txt"
## [1] "natural_dna_0001.txt"
## [1] "natural_dna_0002.txt"
## [1] "natural_dna_0003.txt"

```

```
## [1] "natural_dna_0004.txt"
## [1] "natural_dna_0005.txt"
## [1] "natural_dna_0006.txt"
## [1] "natural_dna_0007.txt"
## [1] "natural_dna_0008.txt"
## [1] "natural_dna_0009.txt"
## [1] "natural_dna_0010.txt"
## [1] "natural_dna_0011.txt"
## [1] "natural_dna_0012.txt"
## [1] "natural_dna_0013.txt"
## [1] "natural_dna_0014.txt"
## [1] "natural_dna_0015.txt"
## [1] "natural_dna_0016.txt"
## [1] "natural_dna_0017.txt"
## [1] "natural_dna_0018.txt"
## [1] "natural_dna_0019.txt"
## [1] "natural_dna_0020.txt"
## [1] "natural_dna_0021.txt"
## [1] "natural_dna_0022.txt"
## [1] "natural_dna_0023.txt"
## [1] "natural_dna_0024.txt"
## [1] "natural_dna_0025.txt"
## [1] "natural_dna_0026.txt"
## [1] "natural_dna_0027.txt"
## [1] "natural_dna_0028.txt"
## [1] "natural_dna_0029.txt"
## [1] "natural_dna_0030.txt"
## [1] "nonwriting_moc_0001.txt"
## [1] "nonwriting_moc_0002.txt"
## [1] "nonwriting_moc_0003.txt"
## [1] "nonwriting_tsl_0001.txt"
## [1] "nonwriting_wsy_0001.txt"
## [1] "paleolithic_aur_0001.txt"
## [1] "paleolithic_bla_0001.txt"
## [1] "paleolithic_bla_0002.txt"
## [1] "paleolithic_bla_0003.txt"
## [1] "paleolithic_bla_0004.txt"
## [1] "paleolithic_bla_0006.txt"
## [1] "paleolithic_bla_0007.txt"
## [1] "paleolithic_bla_0008.txt"
## [1] "paleolithic_bla_0009.txt"
## [1] "paleolithic_bla_0010.txt"
## [1] "paleolithic_bla_0011.txt"
## [1] "paleolithic_bla_0013.txt"
## [1] "paleolithic_bla_0014.txt"
## [1] "paleolithic_brs_0003.txt"
## [1] "paleolithic_brs_0004.txt"
## [1] "paleolithic_cas_0001.txt"
## [1] "paleolithic_cas_0003.txt"
## [1] "paleolithic_cas_0006.txt"
## [1] "paleolithic_cas_0007.txt"
## [1] "paleolithic_cas_0008.txt"
## [1] "paleolithic_cel_0002.txt"
## [1] "paleolithic_cel_0003.txt"
```

```
## [1] "paleolithic_gdr_0001.txt"
## [1] "paleolithic_gdr_0002.txt"
## [1] "paleolithic_gdr_0003.txt"
## [1] "paleolithic_hfc_0012.txt"
## [1] "paleolithic_hss_0002.txt"
## [1] "paleolithic_laf_0004.txt"
## [1] "paleolithic_laf_0005.txt"
## [1] "paleolithic_laf_0007.txt"
## [1] "paleolithic_laf_0008.txt"
## [1] "paleolithic_laf_0011.txt"
## [1] "paleolithic_laf_0012.txt"
## [1] "paleolithic_laf_0013.txt"
## [1] "paleolithic_laf_0014.txt"
## [1] "paleolithic_laf_0015.txt"
## [1] "paleolithic_laf_0016.txt"
## [1] "paleolithic_laf_0018.txt"
## [1] "paleolithic_laf_0019.txt"
## [1] "paleolithic_lar_0001.txt"
## [1] "paleolithic_lar_0002.txt"
## [1] "paleolithic_pal_0001.txt"
## [1] "paleolithic_sgr_0001.txt"
## [1] "paleolithic_tdc_0001.txt"
## [1] "paleolithic_tri_0001.txt"
## [1] "paleolithic_vhc_0090.txt"
## [1] "paleolithic_vhc_0145.txt"
## [1] "paleolithic_vhc_0161.txt"
## [1] "paleolithic_vhc_0165.txt"
## [1] "unclassified_pic_0001.txt"
## [1] "unclassified_rrg_0001.txt"
## [1] "unclassified_rrg_0002.txt"
## [1] "unclassified_voy_0001.txt"
## [1] "ancient_sum_0001.txt"
## [1] "ancient_sum_0002.txt"
## [1] "ancient_sum_0003.txt"
## [1] "ancient_sum_0004.txt"
## [1] "ancient_sum_0005.txt"
## [1] "ancient_sum_0006.txt"
## [1] "ancient_sum_0007.txt"
## [1] "ancient_sum_0008.txt"
## [1] "ancient_sum_0009.txt"
## [1] "ancient_sum_0010.txt"
## [1] "heraldics_bla_0001.txt"
## [1] "heraldics_bla_0002.txt"
## [1] "writing_aii_0001.txt"
## [1] "writing_arb_0001.txt"
## [1] "writing_azj_0001.txt"
## [1] "writing_azj_0002.txt"
## [1] "writing_ben_0001.txt"
## [1] "writing_blt_0001.txt"
## [1] "writing_bod_0001.txt"
## [1] "writing_bos_0001.txt"
## [1] "writing_bos_0002.txt"
## [1] "writing_chr_0001.txt"
## [1] "writing_cmn_0001.txt"
```

```
## [1] "writing_cmn_0002.txt"
## [1] "writing_csw_0001.txt"
## [1] "writing_div_0001.txt"
## [1] "writing_ell_0001.txt"
## [1] "writing_eng_0001.txt"
## [1] "writing_epo_0001.txt"
## [1] "writing_eus_0001.txt"
## [1] "writing_gaz_0001.txt"
## [1] "writing_guj_0001.txt"
## [1] "writing_heb_0001.txt"
## [1] "writing_hin_0001.txt"
## [1] "writing_hye_0001.txt"
## [1] "writing_ibt_0001.txt"
## [1] "writing_iii_0001.txt"
## [1] "writing_ike_0001.txt"
## [1] "writing_jav_0001.txt"
## [1] "writing_jav_0002.txt"
## [1] "writing_jpn_0001.txt"
## [1] "writing_kal_0001.txt"
## [1] "writing_kan_0001.txt"
## [1] "writing_kat_0001.txt"
## [1] "writing_khm_0001.txt"
## [1] "writing_kkh_0001.txt"
## [1] "writing_kor_0001.txt"
## [1] "writing_lao_0001.txt"
## [1] "writing_lug_0001.txt"
## [1] "writing_mal_0001.txt"
## [1] "writing_mya_0001.txt"
## [1] "writing_pan_0001.txt"
## [1] "writing_rus_0001.txt"
## [1] "writing_sin_0001.txt"
## [1] "writing_tam_0001.txt"
## [1] "writing_tel_0001.txt"
## [1] "writing_tgl_0001.txt"
## [1] "writing_tha_0001.txt"
## [1] "writing_tir_0001.txt"
## [1] "writing_vai_0001.txt"
## [1] "writing_zgh_0001.txt"
## [1] "writing_zul_0001.txt"
```

```
end_time <- Sys.time()
end_time - start_time
```

```
## Time difference of 28.18249 secs
```

```
estimations.df
```

##	filename	subcorpus	code	huni.chars	hrate.chars
## 1	animal_bhg_0001.txt	animal	bhg_0001	3.4947514	2.7743133
## 2	animal_bhg_0002.txt	animal	bhg_0002	2.9883959	2.8259567
## 3	animal_bhg_0003.txt	animal	bhg_0003	3.4717829	2.9226335
## 4	animal_bhg_0004.txt	animal	bhg_0004	3.0610425	2.6829183
## 5	animal_bhg_0005.txt	animal	bhg_0005	3.4641481	2.9507039
## 6	animal_bhg_0006.txt	animal	bhg_0006	2.5542540	2.7308654
## 7	animal_bhg_0007.txt	animal	bhg_0007	3.3843958	2.9726295
## 8	animal_bhg_0008.txt	animal	bhg_0008	3.2706877	2.9004744

## 9	animal_bhg_0009.txt	animal bhg_0009	3.1697116	2.6953485
## 10	animal_bhg_0010.txt	animal bhg_0010	3.1864896	2.8111131
## 11	animal_cad_0001.txt	animal cad_0001	2.1252128	1.8330585
## 12	animal_cad_0002.txt	animal cad_0002	1.9099011	1.4055566
## 13	animal_cav_0001.txt	animal cav_0001	3.2782777	2.8599520
## 14	animal_cav_0002.txt	animal cav_0002	3.3203304	2.7167563
## 15	animal_cav_0003.txt	animal cav_0003	3.2007938	2.7582866
## 16	animal_cav_0004.txt	animal cav_0004	2.7277807	2.0993541
## 17	animal_cav_0005.txt	animal cav_0005	2.8320581	2.6104753
## 18	animal_cav_0006.txt	animal cav_0006	2.3737183	2.4397871
## 19	animal_cav_0007.txt	animal cav_0007	2.6991983	1.9755894
## 20	animal_cav_0008.txt	animal cav_0008	3.0678500	2.6795165
## 21	animal_cav_0009.txt	animal cav_0009	3.2362941	2.7715679
## 22	animal_cth_0001.txt	animal cth_0001	2.7534322	2.2302114
## 23	animal_cth_0002.txt	animal cth_0002	3.1250167	2.7284740
## 24	animal_cth_0003.txt	animal cth_0003	2.7057917	2.4211062
## 25	animal_cth_0004.txt	animal cth_0004	2.7219830	2.3663868
## 26	animal_cth_0005.txt	animal cth_0005	3.0451615	2.5633773
## 27	animal_cth_0006.txt	animal cth_0006	2.5980159	2.3893199
## 28	animal_cth_0007.txt	animal cth_0007	2.8958513	2.4964281
## 29	animal_cth_0008.txt	animal cth_0008	3.0852531	2.5169808
## 30	animal_cth_0009.txt	animal cth_0009	2.7465062	2.4917414
## 31	animal_cth_0010.txt	animal cth_0010	2.9742465	2.3871399
## 32	animal_cth_0011.txt	animal cth_0011	2.9345943	2.6414098
## 33	animal_zfi_0001.txt	animal zfi_0001	2.6297493	2.4663452
## 34	natural_dna_0001.txt	natural dna_0001	1.8895491	1.5933784
## 35	natural_dna_0002.txt	natural dna_0002	1.9354070	1.5427758
## 36	natural_dna_0003.txt	natural dna_0003	1.9349850	1.5398654
## 37	natural_dna_0004.txt	natural dna_0004	1.9747933	1.5723663
## 38	natural_dna_0005.txt	natural dna_0005	1.9527993	1.5930424
## 39	natural_dna_0006.txt	natural dna_0006	1.8599938	1.5559014
## 40	natural_dna_0007.txt	natural dna_0007	1.8825380	1.4900496
## 41	natural_dna_0008.txt	natural dna_0008	1.5120154	1.0888722
## 42	natural_dna_0009.txt	natural dna_0009	1.9346489	1.5844600
## 43	natural_dna_0010.txt	natural dna_0010	1.9543556	1.6116174
## 44	natural_dna_0011.txt	natural dna_0011	1.9924830	1.5456395
## 45	natural_dna_0012.txt	natural dna_0012	1.8081904	1.4240895
## 46	natural_dna_0013.txt	natural dna_0013	1.9456687	1.5214048
## 47	natural_dna_0014.txt	natural dna_0014	1.8797028	1.5776384
## 48	natural_dna_0015.txt	natural dna_0015	1.8729388	1.6049957
## 49	natural_dna_0016.txt	natural dna_0016	1.6836145	1.2251730
## 50	natural_dna_0017.txt	natural dna_0017	1.4597554	0.4745727
## 51	natural_dna_0018.txt	natural dna_0018	1.9200295	1.5912382
## 52	natural_dna_0019.txt	natural dna_0019	1.4573993	0.9422011
## 53	natural_dna_0020.txt	natural dna_0020	1.4577175	0.7480833
## 54	natural_dna_0021.txt	natural dna_0021	1.5593654	1.1767620
## 55	natural_dna_0022.txt	natural dna_0022	1.6957083	1.4328030
## 56	natural_dna_0023.txt	natural dna_0023	1.5125120	0.9406015
## 57	natural_dna_0024.txt	natural dna_0024	1.8551300	1.5332971
## 58	natural_dna_0025.txt	natural dna_0025	1.9907269	1.5720881
## 59	natural_dna_0026.txt	natural dna_0026	1.9897170	1.5881855
## 60	natural_dna_0027.txt	natural dna_0027	1.8708107	1.5532241
## 61	natural_dna_0028.txt	natural dna_0028	1.9907269	1.5933961
## 62	natural_dna_0029.txt	natural dna_0029	1.8797591	1.5657802

## 63	natural_dna_0030.txt	natural dna_0030	1.7166112	1.5058083
## 64	nonwriting_moc_0001.txt	nonwriting moc_0001	1.7752863	2.2417427
## 65	nonwriting_moc_0002.txt	nonwriting moc_0002	1.6819385	2.2511064
## 66	nonwriting_moc_0003.txt	nonwriting moc_0003	1.7613849	2.2203262
## 67	nonwriting_tsl_0001.txt	nonwriting tsl_0001	4.2485495	2.5940437
## 68	nonwriting_wsy_0001.txt	nonwriting wsy_0001	3.5635531	1.7960634
## 69	paleolithic_aur_0001.txt	paleolithic aur_0001	0.5916728	0.8024226
## 70	paleolithic_bla_0001.txt	paleolithic bla_0001	1.6031365	1.0649105
## 71	paleolithic_bla_0002.txt	paleolithic bla_0002	0.4999160	1.1441641
## 72	paleolithic_bla_0003.txt	paleolithic bla_0003	0.9182958	0.9485351
## 73	paleolithic_bla_0004.txt	paleolithic bla_0004	0.5435644	0.8109395
## 74	paleolithic_bla_0006.txt	paleolithic bla_0006	0.6840384	1.1203569
## 75	paleolithic_bla_0007.txt	paleolithic bla_0007	1.2686749	1.0239274
## 76	paleolithic_bla_0008.txt	paleolithic bla_0008	1.5000000	1.1462406
## 77	paleolithic_bla_0009.txt	paleolithic bla_0009	0.9182958	0.5974938
## 78	paleolithic_bla_0010.txt	paleolithic bla_0010	1.0000000	0.5000000
## 79	paleolithic_bla_0011.txt	paleolithic bla_0011	0.8112781	0.6981203
## 80	paleolithic_bla_0013.txt	paleolithic bla_0013	1.0000000	0.5000000
## 81	paleolithic_bla_0014.txt	paleolithic bla_0014	0.5787946	0.9559518
## 82	paleolithic_brs_0003.txt	paleolithic brs_0003	0.6500224	0.7542675
## 83	paleolithic_brs_0004.txt	paleolithic brs_0004	0.5032583	0.7603804
## 84	paleolithic_cas_0001.txt	paleolithic cas_0001	1.2738722	1.1131769
## 85	paleolithic_cas_0003.txt	paleolithic cas_0003	0.8112781	0.7954871
## 86	paleolithic_cas_0006.txt	paleolithic cas_0006	0.7219281	0.7240224
## 87	paleolithic_cas_0007.txt	paleolithic cas_0007	0.9182958	0.5974938
## 88	paleolithic_cas_0008.txt	paleolithic cas_0008	1.0000000	0.5000000
## 89	paleolithic_cel_0002.txt	paleolithic cel_0002	0.6500224	0.7542675
## 90	paleolithic_cel_0003.txt	paleolithic cel_0003	1.0000000	0.5000000
## 91	paleolithic_gdr_0001.txt	paleolithic gdr_0001	0.5916728	0.7578517
## 92	paleolithic_gdr_0002.txt	paleolithic gdr_0002	0.5435644	0.7652069
## 93	paleolithic_gdr_0003.txt	paleolithic gdr_0003	1.2177968	1.1179581
## 94	paleolithic_hfc_0012.txt	paleolithic hfc_0012	0.9182958	0.5974938
## 95	paleolithic_hss_0002.txt	paleolithic hss_0002	1.5000000	1.1524935
## 96	paleolithic_laf_0004.txt	paleolithic laf_0004	0.6193822	0.8335619
## 97	paleolithic_laf_0005.txt	paleolithic laf_0005	0.3712323	0.7274908
## 98	paleolithic_laf_0007.txt	paleolithic laf_0007	0.4394970	0.7503656
## 99	paleolithic_laf_0008.txt	paleolithic laf_0008	0.7219281	0.7240224
## 100	paleolithic_laf_0011.txt	paleolithic laf_0011	1.3293889	1.4435938
## 101	paleolithic_laf_0012.txt	paleolithic laf_0012	1.5566567	1.1157954
## 102	paleolithic_laf_0013.txt	paleolithic laf_0013	1.0000000	0.5000000
## 103	paleolithic_laf_0014.txt	paleolithic laf_0014	1.0000000	0.5000000
## 104	paleolithic_laf_0015.txt	paleolithic laf_0015	1.0000000	0.5000000
## 105	paleolithic_laf_0016.txt	paleolithic laf_0016	1.1246581	1.3105689
## 106	paleolithic_laf_0018.txt	paleolithic laf_0018	1.1772189	1.4835599
## 107	paleolithic_laf_0019.txt	paleolithic laf_0019	0.2055925	0.5774563
## 108	paleolithic_lar_0001.txt	paleolithic lar_0001	0.8112781	0.6981203
## 109	paleolithic_lar_0002.txt	paleolithic lar_0002	0.5842388	1.2484055
## 110	paleolithic_pal_0001.txt	paleolithic pal_0001	1.8431582	1.2361269
## 111	paleolithic_sgr_0001.txt	paleolithic sgr_0001	1.5385015	2.1010853
## 112	paleolithic_tdc_0001.txt	paleolithic tdc_0001	1.4219115	1.3862322
## 113	paleolithic_tri_0001.txt	paleolithic tri_0001	1.5000000	1.1462406
## 114	paleolithic_vhc_0090.txt	paleolithic vhc_0090	1.4353714	1.2901549
## 115	paleolithic_vhc_0145.txt	paleolithic vhc_0145	1.5166023	1.4251726
## 116	paleolithic_vhc_0161.txt	paleolithic vhc_0161	1.3718335	1.6202455

## 117	paleolithic_vhc_0165.txt	paleolithic_vhc_0165	1.4185644	1.1957809
## 118	unclassified_pic_0001.txt	unclassified_pic_0001	4.1835049	2.5881374
## 119	unclassified_rrg_0001.txt	unclassified_rrg_0001	3.3733607	2.5996039
## 120	unclassified_rrg_0002.txt	unclassified_rrg_0002	3.2146701	2.5025849
## 121	unclassified_voy_0001.txt	unclassified_voy_0001	3.6473770	2.4852381
## 122	ancient_sum_0001.txt	ancient_sum_0001	2.9488376	2.2774736
## 123	ancient_sum_0002.txt	ancient_sum_0002	4.8987512	2.2170598
## 124	ancient_sum_0003.txt	ancient_sum_0003	4.9107015	3.6931184
## 125	ancient_sum_0004.txt	ancient_sum_0004	5.2053880	3.6827233
## 126	ancient_sum_0005.txt	ancient_sum_0005	5.6193165	3.9489304
## 127	ancient_sum_0006.txt	ancient_sum_0006	4.3748952	3.0168381
## 128	ancient_sum_0007.txt	ancient_sum_0007	3.9414876	3.1183933
## 129	ancient_sum_0008.txt	ancient_sum_0008	3.8913810	3.0108193
## 130	ancient_sum_0009.txt	ancient_sum_0009	3.5727603	2.7976392
## 131	ancient_sum_0010.txt	ancient_sum_0010	3.9083249	2.9571902
## 132	heraldics_bla_0001.txt	heraldics_bla_0001	4.2680472	3.0455810
## 133	heraldics_bla_0002.txt	heraldics_bla_0002	4.3415935	3.0460646
## 134	writing_aai_0001.txt	writing_aai_0001	4.0060773	2.9553428
## 135	writing_arb_0001.txt	writing_arb_0001	4.1492855	3.0444697
## 136	writing_azj_0001.txt	writing_azj_0001	4.3351602	3.0125972
## 137	writing_azj_0002.txt	writing_azj_0002	4.3351602	3.0125972
## 138	writing_ben_0001.txt	writing_ben_0001	4.5406707	3.3796527
## 139	writing_blt_0001.txt	writing_blt_0001	4.3723434	3.3739103
## 140	writing_bod_0001.txt	writing_bod_0001	4.2318518	3.8125316
## 141	writing_bos_0001.txt	writing_bos_0001	4.4344373	3.0964981
## 142	writing_bos_0002.txt	writing_bos_0002	4.4010891	3.0191758
## 143	writing_chr_0001.txt	writing_chr_0001	4.4947817	3.1888884
## 144	writing_cmn_0001.txt	writing_cmn_0001	5.9036961	4.4177092
## 145	writing_cmn_0002.txt	writing_cmn_0002	5.9036961	4.4177092
## 146	writing_csw_0001.txt	writing_csw_0001	4.2372249	3.1356321
## 147	writing_div_0001.txt	writing_div_0001	4.5676562	3.0704228
## 148	writing_ell_0001.txt	writing_ell_0001	4.9182427	3.4696208
## 149	writing_eng_0001.txt	writing_eng_0001	4.0012025	2.7416123
## 150	writing_epo_0001.txt	writing_epo_0001	3.9445137	2.8145753
## 151	writing_eus_0001.txt	writing_eus_0001	4.6560419	3.2446058
## 152	writing_gaz_0001.txt	writing_gaz_0001	4.1366637	3.0021000
## 153	writing_guj_0001.txt	writing_guj_0001	4.3101563	3.2055055
## 154	writing_heb_0001.txt	writing_heb_0001	4.0404504	3.0017605
## 155	writing_hin_0001.txt	writing_hin_0001	4.3956235	3.3618022
## 156	writing_hye_0001.txt	writing_hye_0001	4.4083192	3.1568048
## 157	writing_ibt_0001.txt	writing_ibt_0001	4.2859808	2.9061997
## 158	writing_iii_0001.txt	writing_iii_0001	5.5649530	4.0496639
## 159	writing_ike_0001.txt	writing_ike_0001	4.7619154	3.2065111
## 160	writing_jav_0001.txt	writing_jav_0001	4.3289268	3.3624811
## 161	writing_jav_0002.txt	writing_jav_0002	4.3267592	3.0347559
## 162	writing_jpn_0001.txt	writing_jpn_0001	5.7634652	4.3412628
## 163	writing_kal_0001.txt	writing_kal_0001	4.0962174	2.8184375
## 164	writing_kan_0001.txt	writing_kan_0001	4.3052968	3.3097899
## 165	writing_kat_0001.txt	writing_kat_0001	3.9727394	2.7816244
## 166	writing_khm_0001.txt	writing_khm_0001	4.8750159	3.6042375
## 167	writing_kkh_0001.txt	writing_kkh_0001	4.6700154	3.8898727
## 168	writing_kor_0001.txt	writing_kor_0001	4.9926966	4.0289612
## 169	writing_lao_0001.txt	writing_lao_0001	4.9245620	3.3360444
## 170	writing_lug_0001.txt	writing_lug_0001	4.3278871	2.8854610

## 171	writing_mal_0001.txt	writing mal_0001	4.3869626	3.4773093
## 172	writing_mya_0001.txt	writing mya_0001	4.7332697	3.4472310
## 173	writing_pan_0001.txt	writing pan_0001	4.3330460	3.1107259
## 174	writing_rus_0001.txt	writing rus_0001	4.4936689	3.1559659
## 175	writing_sin_0001.txt	writing sin_0001	4.2359526	3.1139618
## 176	writing_tam_0001.txt	writing tam_0001	4.1058464	3.3176428
## 177	writing_tel_0001.txt	writing tel_0001	4.3371250	3.2983929
## 178	writing_tgl_0001.txt	writing tgl_0001	3.8247202	2.7833387
## 179	writing_tha_0001.txt	writing tha_0001	4.9600782	3.4183192
## 180	writing_tir_0001.txt	writing tir_0001	4.6826832	3.4408378
## 181	writing_vai_0001.txt	writing vai_0001	4.2504735	3.5187710
## 182	writing_zgh_0001.txt	writing zgh_0001	3.7692126	2.8707716
## 183	writing_zul_0001.txt	writing zul_0001	4.0110554	2.6953365
##	huni.strings	hrate.strings	ttr.chars	ttr.strings
## 1	4.0179224	1.5120346	0.18000000	0.18000000
## 2	4.4862385	2.0351474	0.18000000	0.2637363
## 3	4.3309521	2.1719603	0.21000000	0.24000000
## 4	4.0057908	2.3787317	0.15000000	0.23000000
## 5	4.2971514	2.6230538	0.21000000	0.4098361
## 6	4.3554343	2.4289625	0.15000000	0.26000000
## 7	5.4651132	3.6081115	0.19000000	0.51000000
## 8	5.2413078	3.3284878	0.18000000	0.45000000
## 9	3.6814882	2.1913025	0.16000000	0.25000000
## 10	4.5216571	2.4353193	0.17000000	0.28000000
## 11	2.3527458	1.6692846	0.16666667	0.4285714
## 12	4.0874628	2.8433884	0.04000000	1.00000000
## 13	5.1135006	3.2416322	0.18000000	0.42000000
## 14	4.9849805	3.2629824	0.15000000	0.38000000
## 15	3.6340280	2.3742294	0.16000000	0.3050847
## 16	2.7500000	1.7369413	0.43478261	0.87500000
## 17	3.7784017	2.5906332	0.11000000	0.3214286
## 18	1.8670721	1.2436300	0.07000000	0.1025641
## 19	2.1867043	1.4958226	0.25000000	0.4545455
## 20	4.4050013	2.9113461	0.14000000	0.4814815
## 21	4.4068211	2.8508970	0.16000000	0.5897436
## 22	3.0163115	1.8862512	0.09000000	0.1388889
## 23	3.7503565	2.4583080	0.20000000	0.23000000
## 24	3.4593490	1.9799942	0.12000000	0.1326531
## 25	3.2464991	2.1427511	0.12000000	0.13000000
## 26	4.8570132	3.0210492	0.15000000	0.37000000
## 27	4.6362728	2.8191248	0.11000000	0.30000000
## 28	4.8417553	3.1410194	0.14000000	0.36000000
## 29	4.7190812	2.9572547	0.14000000	0.34000000
## 30	4.9023339	3.3259938	0.13000000	0.38000000
## 31	4.1307229	2.3795656	0.12000000	0.22000000
## 32	4.6603365	2.8996348	0.17000000	0.32000000
## 33	2.9842342	1.6417823	0.29032258	0.4705882
## 34	6.6438562	5.2476499	0.04000000	1.00000000
## 35	6.6438562	5.2476499	0.04000000	1.00000000
## 36	5.8932697	4.5647268	0.04000000	0.83000000
## 37	6.6438562	5.2476499	0.04000000	1.00000000
## 38	6.6438562	5.2476499	0.04000000	1.00000000
## 39	6.6438562	5.2476499	0.04000000	1.00000000
## 40	6.6438562	5.2476499	0.04000000	1.00000000

## 41	6.6438562	5.2476499	0.04000000	1.0000000	0.60416667
## 42	6.6438562	5.2476499	0.04000000	1.0000000	0.22916667
## 43	6.6438562	5.2476499	0.04000000	1.0000000	0.32291667
## 44	6.6238562	5.2147251	0.04000000	0.9900000	0.20833333
## 45	6.6438562	5.2476499	0.04000000	1.0000000	0.30208333
## 46	6.6438562	5.2476499	0.04000000	1.0000000	0.25000000
## 47	6.6438562	5.2476499	0.04000000	1.0000000	0.36458333
## 48	6.6438562	5.2476499	0.04000000	1.0000000	0.34375000
## 49	6.6438562	5.2476499	0.04000000	1.0000000	0.44791667
## 50	6.6438562	5.2476499	0.03000000	1.0000000	0.56701031
## 51	6.6438562	5.2476499	0.04000000	1.0000000	0.33333333
## 52	6.6438562	5.2476499	0.03000000	1.0000000	0.52577320
## 53	6.6038562	5.1900419	0.03000000	0.9800000	0.58762887
## 54	6.6438562	5.2476499	0.04000000	1.0000000	0.57291667
## 55	6.5638562	5.1795506	0.04000000	0.9700000	0.36458333
## 56	5.4217419	4.1156371	0.04000000	0.7500000	0.60416667
## 57	6.6438562	5.2476499	0.04000000	1.0000000	0.37500000
## 58	6.6438562	5.2476499	0.04000000	1.0000000	0.27083333
## 59	6.6438562	5.2476499	0.04000000	1.0000000	0.26041667
## 60	6.6438562	5.2476499	0.04000000	1.0000000	0.34375000
## 61	6.6438562	5.2476499	0.04000000	1.0000000	0.34375000
## 62	6.6438562	5.2476499	0.04000000	1.0000000	0.31250000
## 63	6.6438562	5.2476499	0.04000000	1.0000000	0.51041667
## 64	4.0062117	2.7756067	0.04000000	0.2200000	0.23958333
## 65	3.9998562	2.6906843	0.04000000	0.2000000	0.21875000
## 66	3.8736199	2.6242165	0.04000000	0.3396226	0.19791667
## 67	4.9847696	3.6100050	0.25000000	0.8947368	0.00000000
## 68	3.1754684	2.3517762	0.16000000	0.1400000	0.00000000
## 69	1.5849625	0.8616542	0.14285714	1.0000000	0.83333333
## 70	2.5849625	1.5819755	0.04761905	1.0000000	0.92500000
## 71	2.2877837	1.5896539	0.02000000	0.4615385	0.79591837
## 72	0.6500224	0.7542675	0.13333333	0.3333333	0.38461538
## 73	0.9182958	0.5974938	0.12500000	0.6666667	0.85714286
## 74	2.2810361	1.4978326	0.04545455	0.6666667	0.66666667
## 75	2.8073549	1.7570297	0.03896104	1.0000000	0.87837838
## 76	1.5849625	0.8616542	0.75000000	1.0000000	0.00000000
## 77	1.0000000	0.5000000	0.66666667	1.0000000	1.00000000
## 78	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 79	1.0000000	0.5000000	0.50000000	1.0000000	1.00000000
## 80	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 81	2.3219281	1.3813781	0.06896552	1.0000000	0.77777778
## 82	1.0000000	0.5000000	0.33333333	1.0000000	1.00000000
## 83	1.0000000	0.5000000	0.22222222	1.0000000	1.00000000
## 84	2.3219281	1.3813781	0.09677419	1.0000000	0.82142857
## 85	1.5849625	0.8616542	0.25000000	1.0000000	0.66666667
## 86	1.0000000	0.5000000	0.40000000	1.0000000	1.00000000
## 87	1.0000000	0.5000000	0.66666667	1.0000000	1.00000000
## 88	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 89	1.0000000	0.5000000	0.33333333	1.0000000	1.00000000
## 90	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 91	1.0000000	0.5000000	0.28571429	1.0000000	1.00000000
## 92	1.0000000	0.5000000	0.25000000	1.0000000	1.00000000
## 93	1.5849625	0.8616542	0.12000000	1.0000000	0.95454545
## 94	1.0000000	0.5000000	0.66666667	1.0000000	1.00000000

## 95	1.5849625	0.8616542	0.37500000	1.0000000	0.80000000
## 96	1.5849625	0.8616542	0.15384615	1.0000000	0.81818182
## 97	1.0000000	0.5000000	0.14285714	1.0000000	1.00000000
## 98	1.0000000	0.5000000	0.18181818	1.0000000	1.00000000
## 99	1.0000000	0.5000000	0.40000000	1.0000000	1.00000000
## 100	2.9219281	1.9291061	0.09302326	0.8000000	0.64102564
## 101	1.5849625	0.8616542	0.42857143	1.0000000	0.75000000
## 102	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 103	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 104	1.0000000	0.5000000	1.00000000	1.0000000	NaN
## 105	1.3567796	1.1984280	0.11111111	0.4000000	0.37500000
## 106	3.0000000	1.9124010	0.13333333	1.0000000	0.61538462
## 107	0.9182958	0.5974938	0.03225806	0.6666667	0.96666667
## 108	1.0000000	0.5000000	0.50000000	1.0000000	1.00000000
## 109	3.4441010	2.4854745	0.02000000	0.3823529	0.74489796
## 110	4.9815381	3.5581023	0.05000000	0.4400000	0.92631579
## 111	4.3958597	3.0749255	0.10000000	0.5652174	0.52222222
## 112	3.1219281	2.0630097	0.04347826	0.9000000	0.77272727
## 113	1.5849625	0.8616542	0.75000000	1.0000000	0.00000000
## 114	1.5000000	0.9481203	0.27272727	0.7500000	0.62500000
## 115	3.1219281	2.0291061	0.07017544	0.9000000	0.71698113
## 116	2.6174925	1.9350872	0.04838710	0.6666667	0.67796610
## 117	1.5849625	0.8616542	0.18750000	1.0000000	0.92307692
## 118	4.3338845	3.6649628	0.23000000	0.3000000	0.02597403
## 119	5.8006721	4.3037961	0.14000000	0.7000000	0.17441860
## 120	5.5196323	3.8855909	0.12000000	0.6100000	0.03409091
## 121	6.3012096	4.7689574	0.16000000	0.8500000	0.02380952
## 122	2.5249672	2.1600750	0.27835052	0.2164948	0.50000000
## 123	4.5850551	2.1986132	0.37234043	0.4807692	0.00000000
## 124	4.3636085	3.2763273	0.72916667	0.7500000	0.46153846
## 125	4.4331536	3.2135721	0.68253968	0.7428571	0.15000000
## 126	4.6940194	3.1601654	0.73684211	0.9310345	0.00000000
## 127	3.7898981	2.5036245	0.64102564	0.5666667	0.50000000
## 128	3.6719281	2.9421215	0.55769231	0.6000000	0.60869565
## 129	3.1061450	2.4595314	0.62790698	0.5312500	0.81250000
## 130	3.2763292	2.6543992	0.34020619	0.3163265	0.54687500
## 131	3.3954618	2.6352867	0.54716981	0.5500000	0.70833333
## 132	5.6273688	4.2790844	0.26000000	0.6100000	0.01351351
## 133	5.1563470	3.6993431	0.28000000	0.5100000	0.02777778
## 134	6.2763073	4.7847044	0.22000000	0.8500000	0.01282051
## 135	6.2661118	4.7489203	0.27000000	0.8400000	0.00000000
## 136	6.1287584	4.8193807	0.29000000	0.8200000	0.01408451
## 137	6.1287584	4.8193807	0.29000000	0.8200000	0.01408451
## 138	6.2051132	4.7308257	0.32000000	0.8300000	0.00000000
## 139	5.7739190	3.8832425	0.37000000	0.6200000	0.03174603
## 140	6.6438562	5.2476499	0.30000000	1.0000000	0.00000000
## 141	5.7959163	4.3200846	0.30000000	0.7100000	0.00000000
## 142	5.7959163	4.3200846	0.28000000	0.7100000	0.00000000
## 143	5.8208186	4.4285531	0.33000000	0.7200000	0.00000000
## 144	4.7978598	4.0544617	0.71000000	0.6000000	0.03448276
## 145	5.9319217	5.2476499	0.71000000	0.8000000	0.03448276
## 146	5.4641847	4.0939634	0.30000000	0.6200000	0.00000000
## 147	6.4038562	5.0766127	0.32000000	0.9000000	0.00000000
## 148	5.9200508	4.3911112	0.41000000	0.7200000	0.01694915

## 149	5.4955253	4.1263581	0.22000000	0.6200000	0.01282051
## 150	5.3180255	4.1766798	0.22000000	0.6000000	0.00000000
## 151	5.9916634	4.5247943	0.32000000	0.7700000	0.04411765
## 152	5.9000508	4.4668227	0.27000000	0.7300000	0.20547945
## 153	6.0877598	4.5928673	0.27000000	0.8000000	0.01369863
## 154	6.2487584	4.6635112	0.23000000	0.8400000	0.01298701
## 155	5.8810494	4.3951502	0.30000000	0.7200000	0.00000000
## 156	6.1575643	4.6470197	0.33000000	0.8100000	0.00000000
## 157	5.7351760	4.0244517	0.28000000	0.6400000	0.00000000
## 158	6.0257208	5.2476499	0.60000000	0.8200000	0.00000000
## 159	6.0575643	4.7733732	0.33000000	0.8300000	0.00000000
## 160	6.1638562	5.2476499	0.28000000	0.8600000	0.00000000
## 161	5.8895970	4.3228156	0.30000000	0.7200000	0.00000000
## 162	5.8835770	5.2476499	0.67000000	0.7900000	0.00000000
## 163	6.2687584	4.7218313	0.23000000	0.8300000	0.20779221
## 164	6.2451132	4.7587364	0.27000000	0.8400000	0.02739726
## 165	5.9667026	4.6400216	0.22000000	0.8000000	0.00000000
## 166	6.4073413	5.0618025	0.38000000	0.9200000	0.03225806
## 167	6.5638562	5.2179962	0.36000000	0.9700000	0.00000000
## 168	6.3887584	4.9498310	0.56000000	0.8800000	0.00000000
## 169	5.2174058	4.4475022	0.38000000	0.6400000	0.00000000
## 170	6.1963073	4.6902496	0.27000000	0.8400000	0.01369863
## 171	6.4438562	4.9671411	0.29000000	0.9100000	0.00000000
## 172	6.3197925	5.0475007	0.32000000	0.8800000	0.00000000
## 173	5.9824665	4.4587915	0.29000000	0.7300000	0.00000000
## 174	5.9614679	4.5983421	0.32000000	0.7600000	0.00000000
## 175	6.2500154	5.0073595	0.23000000	0.8400000	0.00000000
## 176	6.4763073	5.1528926	0.23000000	0.9200000	0.01298701
## 177	6.5763073	5.1530114	0.26000000	0.9700000	0.01351351
## 178	5.1695761	3.7637882	0.26000000	0.5400000	0.00000000
## 179	6.2722436	4.8395126	0.38000000	0.8800000	0.01612903
## 180	6.1687584	4.5792755	0.40000000	0.7900000	0.00000000
## 181	5.6574042	3.9752288	0.40000000	0.6100000	0.00000000
## 182	5.4287584	4.1719340	0.19000000	0.6400000	0.04938272
## 183	6.4277598	4.9143854	0.21000000	0.9100000	0.00000000

Safe outputs to file

```
write.csv(estimations.df, "/home/chris/Github/StringBase/code/Tables/output_stringBase.csv", row.names = FALSE)
```