1. Split F M Names All Years

Shawn Behrend

2022-05-05

This code opens each YOB file per year and splits out the female names from the male names. Each file is saved as a separate CSV file named for the year and the following ending “\_m\_names\_top1000.txt” or “\_f\_names\_top1000.txt”.

#input what years will be analyzed, based on available data files  
allyears<-c(1880:2020)  
  
#establish first index for data frame based on years chosen for analysis  
firstyearminus1<-allyears[1]-1  
#get index for last year to be analyzed, used in plot later  
allyears\_last<-length(allyears)  
  
# set up a data frame full of "NA" to store the desired name and sex with each year and rank from the data files  
year\_rank<-data.frame(matrix(NA,length(allyears),4))  
  
# loop for each year to be searched  
for (year in allyears) {  
 #load names by year file  
 yob\_file <- read.csv(paste0("C:/Users/shawn/Shawn/CSU\_global/MIS581/names/yob",year,".txt"), header=FALSE)  
 #add headers  
 names(yob\_file) <- c("Name","Sex","Occurance")  
  
#which lines are female and male names  
is\_f<-yob\_file$Sex=="F"  
firstf<-min(which(is\_f==TRUE))  
lastf<-max(which(is\_f==TRUE))  
is\_m<-yob\_file$Sex=="M"  
firstm<-min(which(is\_m==TRUE))  
lastm<-max(which(is\_m==TRUE))  
#split file into female and male names  
yob\_file\_f<-yob\_file[firstf:lastf,]  
yob\_file\_m<-yob\_file[firstm:lastm,]  
# add columns/headers for year and rank to the file  
f\_rows<-nrow(yob\_file\_f)  
m\_rows<-nrow(yob\_file\_m)  
# year  
Year<-year  
#rank  
f\_Rank<-c(1:f\_rows)  
m\_Rank<-c(1:m\_rows)  
# add rank and year to dataset  
#cbind  
yob\_file\_f\_rank<-cbind(yob\_file\_f,f\_Rank,Year)  
yob\_file\_m\_rank<-cbind(yob\_file\_m,m\_Rank,Year)  
# subset only top 100 male names  
yob\_file\_m\_1000<-yob\_file\_m\_rank[c(1:1000),,]  
# subset only top 1000 female names  
yob\_file\_f\_1000<-yob\_file\_f\_rank[c(1:1000),,]  
# write file with top 1000 F or top 100 M names  
write.csv(x=yob\_file\_f\_1000,file=paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",year,"\_f\_names\_top1000.txt"))  
write.csv(x=yob\_file\_m\_1000,file=paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",year,"\_m\_names\_top1000.txt"))   
}

A excerpt of the 2020 year of birth (YOB) file showing the first 10 male names and first 10 female names is shown here.

head(yob\_file\_m\_1000,10)

## Name Sex Occurance m\_Rank Year  
## 17361 Liam M 19659 1 2020  
## 17362 Noah M 18252 2 2020  
## 17363 Oliver M 14147 3 2020  
## 17364 Elijah M 13034 4 2020  
## 17365 William M 12541 5 2020  
## 17366 James M 12250 6 2020  
## 17367 Benjamin M 12136 7 2020  
## 17368 Lucas M 11281 8 2020  
## 17369 Henry M 10705 9 2020  
## 17370 Alexander M 10151 10 2020

head(yob\_file\_f\_1000,10)

## Name Sex Occurance f\_Rank Year  
## 1 Olivia F 17535 1 2020  
## 2 Emma F 15581 2 2020  
## 3 Ava F 13084 3 2020  
## 4 Charlotte F 13003 4 2020  
## 5 Sophia F 12976 5 2020  
## 6 Amelia F 12704 6 2020  
## 7 Isabella F 12066 7 2020  
## 8 Mia F 11157 8 2020  
## 9 Evelyn F 9445 9 2020  
## 10 Harper F 8778 10 2020