2. Compare M-F Names by Last Year

Shawn Behrend

2022-05-06

This code compares every name in the Top 1000 CSV file for males with every name in the female, then compiles those names and their corresponding rank.

#input what years will be analyzed, based on available data files  
decade1<-c(1880:1889)  
decade2<-decade1+10  
decade3<-decade2+10  
decade4<-decade3+10  
decade5<-decade4+10  
decade6<-decade5+10  
decade7<-decade6+10  
decade8<-decade7+10  
decade9<-decade8+10  
decade10<-decade9+10  
decade11<-decade10+10  
decade12<-decade11+10  
decade13<-decade12+10  
decade14<-c(2010:2020)  
#grab first and last year for use in file names  
all\_years<-decade14  
first\_year<-min(all\_years)  
last\_year<-max(all\_years)  
  
# input names to be searched on  
comp\_names\_df<-read.csv(paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/",last\_year,"\_m\_names\_top1000.txt"), fileEncoding="UTF-8-BOM",colClasses=c("character"))  
comp\_names<-comp\_names\_df[[2]]  
  
# begin find\_name loop  
#choose number of searchnames  
inames<-c(1:length(comp\_names))  
#establish first index finding searchname  
iname<-1  
  
# set up a data frame full of "NA" to store the desired name and sex with each year and rank from the data files  
 name\_rank\_f<-data.frame(matrix(NA,length(comp\_names),5))  
 name\_rank\_m<-data.frame(matrix(NA,length(comp\_names),5))  
  
comp\_file\_f <- read.csv(paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/",last\_year,"\_f\_names\_top1000.txt"),colClasses=c(Name="character",Sex="character"))  
comp\_file\_m <- read.csv(paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/",last\_year,"\_m\_names\_top1000.txt"),colClasses=c(Name="character",Sex="character"))  
   
 #start loop through names  
 for (iname in inames) {  
#establish searchname  
 searchname<-comp\_names[iname]  
 find\_name\_f<-c("NA",searchname,"F","NA","NA","NA")  
 find\_name\_m<-c("NA",searchname,"M","NA","NA","NA")  
# Look for row that search name is in using grep.   
# If length of yob\_file\_sex$Name is greater than zero, replace dummy row with this row #  
#----female name files---  
 if(length(grep(paste("^",searchname,"$", sep=""),comp\_file\_f$Name))>0) {  
 #save row info in new variable  
 find\_name\_f<-c(comp\_file\_f[(which(comp\_file\_f$Name==searchname)),])  
 }  
 # Use index to save desired name data to the data frame  
 name\_rank\_f[iname,]<-c(searchname,find\_name\_f[3],find\_name\_f[4],find\_name\_f[5],  
 find\_name\_f[6])  
#----male name files---  
 if(length(grep(paste("^",searchname,"$", sep=""),comp\_file\_m$Name))>0) {  
 #save row info in new variable  
 find\_name\_m<-c(comp\_file\_m[(which(comp\_file\_m$Name==searchname)),])  
 }  
 # Use index to save desired name data to the data frame. Use female year to know which should be eliminated.  
 name\_rank\_m[iname,]<-c(searchname,find\_name\_m[3],find\_name\_m[4],find\_name\_m[5],find\_name\_f[6])  
 iname=iname+1  
 }  
 #add column names to data frame  
 names(name\_rank\_f)<-c("Name","Sex","Occurrance","Rank","Year")  
 names(name\_rank\_m)<-c("Name","Sex","Occurrance","Rank","Year")  
   
 #save names to data file  
 write.csv(x=name\_rank\_f,file=paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/comp\_gn\_names\_F\_",last\_year,".csv"))  
 write.csv(x=name\_rank\_m,file=paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/comp\_gn\_names\_M\_",last\_year,".csv"))

An excerpt of the matching names and corresponding ranks that were found for males and females is shown here. If the name did not appear on both the male and female list, then the occurrence, rank and year show "NA". As is seen here, Noah was found on both lists, with rank 876 on the female list and rank 2 on the male list.

head(name\_rank\_f,10)

## Name Sex Occurrance Rank Year  
## 1 Liam F NA NA NA  
## 2 Noah F 305 876 2020  
## 3 Oliver F NA NA NA  
## 4 Elijah F NA NA NA  
## 5 William F NA NA NA  
## 6 James F NA NA NA  
## 7 Benjamin F NA NA NA  
## 8 Lucas F NA NA NA  
## 9 Henry F NA NA NA  
## 10 Alexander F NA NA NA

head(name\_rank\_m,10)

## Name Sex Occurrance Rank Year  
## 1 Liam M 19659 1 NA  
## 2 Noah M 18252 2 2020  
## 3 Oliver M 14147 3 NA  
## 4 Elijah M 13034 4 NA  
## 5 William M 12541 5 NA  
## 6 James M 12250 6 NA  
## 7 Benjamin M 12136 7 NA  
## 8 Lucas M 11281 8 NA  
## 9 Henry M 10705 9 NA  
## 10 Alexander M 10151 10 NA

Before this file can be used for the next step in the process, the CSV file must be brought into Microsoft Excel, and all NA lines will be removed.