3. GN M-F Names By Year

Shawn Behrend

2022-05-07

After sorting the comp\_gn\_names files for males and females for the last year in the decade and removing all names that are NA for that decade, this code runs and makes a separate CSV for every year in the decade, using only the names from the last year in the decade.

#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)  
#dont write over comp\_gn file for last year in decade  
count\_allyearsminus1<-length(all\_years)-1  
allyearsminus1<-all\_years[1:count\_allyearsminus1]  
  
# input names to be searched on  
gn\_names\_df<-read.csv(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"),fileEncoding="UTF-8-BOM",colClasses=c(Name="character",Sex="character"))  
gn\_names<-gn\_names\_df[[2]]  
#choose number of searchnames  
inames<-c(1:length(gn\_names))  
for (year in allyearsminus1) {  
# begin find\_name loop  
#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(gn\_names),5))  
 name\_rank\_m<-data.frame(matrix(NA,length(gn\_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,"/",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,"/",year,"\_m\_names\_top1000.txt"),  
 colClasses=c(Name="character",Sex="character"))  
   
 #start loop through names  
 for (iname in inames) {  
 #establish searchname  
 searchname<-gn\_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 comp\_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\_",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\_",year,".csv"))  
 }

An excerpt of the output of this file is shown for year 2019. Even though the name Noah shows as "NA" in this year in the female list, it did appear in other years for both males and females.

head(name\_rank\_f,10)

## Name Sex Occurrance Rank Year  
## 1 Noah F NA NA NA  
## 2 Logan F 993 322 2019  
## 3 Carter F 610 513 2019  
## 4 Dylan F 728 436 2019  
## 5 Ezra F 326 830 2019  
## 6 Ryan F 773 399 2019  
## 7 Cameron F 562 544 2019  
## 8 Hunter F 308 866 2019  
## 9 Angel F 781 393 2019  
## 10 Jordan F 768 403 2019

head(name\_rank\_m,10)

## Name Sex Occurrance Rank Year  
## 1 Noah M 19097 2 NA  
## 2 Logan M 10520 16 2019  
## 3 Carter M 8652 30 2019  
## 4 Dylan M 7530 39 2019  
## 5 Ezra M 6521 49 2019  
## 6 Ryan M 6104 54 2019  
## 7 Cameron M 5236 67 2019  
## 8 Hunter M 5369 66 2019  
## 9 Angel M 4895 72 2019  
## 10 Jordan M 4317 89 2019