4. Years in Decade Summary Table

Shawn Behrend

2022-05-07

This code reads each GN file per decade to create a summary table of all applicable parameters for this study.

#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)

Note that this code must be run separately per decade. Currently the code was set up for decade 14 (2010-2020), shown below.

#grab first and last year for use in file names  
all\_years<-decade14  
first\_year<-min(all\_years)  
last\_year<-max(all\_years)  
#establish first index for data frame based on years chosen for analysis  
firstyearminus1<-all\_years[1]-1  
#--- MAKE TABLE  
# set up a data frame full of "NA" to store the desired info by year from the data files  
years\_per\_decade\_table<-data.frame(matrix(NA,length(all\_years),13))  
names(years\_per\_decade\_table)<-c("Year","# of GN Names","Total Names","Ratio of GN Names to Total Names","F occurrances of GN names","Total F occurrances","Ratio of F GN Occur to Total F Occur","M occurrances of GN names","Total M occurrances","Ratio of M GN Occur to Total M Occur","Total GN Occurrances","Total Occurrances","Ratio of GN Occurrances to Total Occurrances")  
# start years loop  
for (year in all\_years) {  
# years per decade table  
# read in the gn names files for the year  
#female gn files  
year\_info\_gn\_f\_df<-read.csv(paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/comp\_gn\_names\_F\_",year,".csv"),fileEncoding="UTF-8-BOM",colClasses=c(Name="character",Sex="character"))  
year\_info\_gn\_f\_names<-year\_info\_gn\_f\_df[[2]]  
year\_info\_gn\_f\_occur<-year\_info\_gn\_f\_df[[4]]  
sum\_year\_info\_gn\_f\_occur<-sum(year\_info\_gn\_f\_occur,na.rm=TRUE)  
count\_year\_info\_gn\_f\_names<-length(year\_info\_gn\_f\_names)  
# male gn files  
year\_info\_gn\_m\_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\_",year,".csv"),fileEncoding="UTF-8-BOM",colClasses=c(Name="character",Sex="character"))  
year\_info\_gn\_m\_names<-year\_info\_gn\_m\_df[[2]]  
year\_info\_gn\_m\_occur<-year\_info\_gn\_m\_df[[4]]  
sum\_year\_info\_gn\_m\_occur<-sum(year\_info\_gn\_m\_occur,na.rm=TRUE)  
count\_year\_info\_gn\_m\_names<-length(year\_info\_gn\_m\_names)  
# sum of gn occurrances  
sum\_gn\_occur<-sum\_year\_info\_gn\_f\_occur+sum\_year\_info\_gn\_m\_occur  
# read in the total names files for the year  
# female all names  
year\_info\_all\_f\_df<-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"),fileEncoding="UTF-8-BOM",colClasses=c(Name="character",Sex="character"))  
year\_info\_all\_f\_names<-year\_info\_all\_f\_df[[2]]  
year\_info\_all\_f\_occur<-year\_info\_all\_f\_df[[4]]  
sum\_year\_info\_all\_f\_occur<-sum(year\_info\_all\_f\_occur,na.rm=TRUE)  
count\_year\_info\_all\_f\_names<-length(year\_info\_all\_f\_names)  
# male all names  
year\_info\_all\_m\_df<-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"),fileEncoding="UTF-8-BOM",colClasses=c(Name="character",Sex="character"))  
year\_info\_all\_m\_names<-year\_info\_all\_m\_df[[2]]  
year\_info\_all\_m\_occur<-year\_info\_all\_m\_df[[4]]  
sum\_year\_info\_all\_m\_occur<-sum(year\_info\_all\_m\_occur,na.rm=TRUE)  
count\_year\_info\_all\_m\_names<-length(year\_info\_all\_m\_names)  
# sum of all names and occur  
count\_year\_info\_all\_names<-sum(count\_year\_info\_all\_m\_names,count\_year\_info\_all\_f\_names)  
# sum of all occurrences  
sum\_all\_occur<-sum\_year\_info\_all\_f\_occur+sum\_year\_info\_all\_m\_occur  
#ratios  
ratio\_gn\_and\_all\_names<-count\_year\_info\_gn\_f\_names/count\_year\_info\_all\_names  
ratio\_f\_gn\_and\_all\_occur<-sum\_year\_info\_gn\_f\_occur/sum\_year\_info\_all\_f\_occur  
ratio\_m\_gn\_and\_all\_occur<-sum\_year\_info\_gn\_m\_occur/sum\_year\_info\_all\_m\_occur  
ratio\_gn\_and\_all\_occur<-sum\_gn\_occur/sum\_all\_occur  
# assemble row in table  
i<-year-firstyearminus1  
years\_per\_decade\_table[i,]<-c(year,count\_year\_info\_gn\_m\_names,count\_year\_info\_all\_names,ratio\_gn\_and\_all\_names,sum\_year\_info\_gn\_f\_occur,sum\_year\_info\_all\_f\_occur,ratio\_f\_gn\_and\_all\_occur,sum\_year\_info\_gn\_m\_occur,sum\_year\_info\_all\_m\_occur,ratio\_m\_gn\_and\_all\_occur,sum\_gn\_occur,sum\_all\_occur,ratio\_gn\_and\_all\_occur)  
}  
#save to data file  
write.csv(x=years\_per\_decade\_table,file=paste0("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/",first\_year,"\_",last\_year,"/",first\_year,"\_",last\_year,"\_summary\_table.csv"))

An excerpt of the first few rows from the Years per Decade summary table for Decade 14 are shown below.

head(years\_per\_decade\_table,20)

## Year # of GN Names Total Names Ratio of GN Names to Total Names  
## 1 2010 85 2000 0.0425  
## 2 2011 85 2000 0.0425  
## 3 2012 85 2000 0.0425  
## 4 2013 85 2000 0.0425  
## 5 2014 85 2000 0.0425  
## 6 2015 85 2000 0.0425  
## 7 2016 85 2000 0.0425  
## 8 2017 85 2000 0.0425  
## 9 2018 85 2000 0.0425  
## 10 2019 85 2000 0.0425  
## 11 2020 85 2000 0.0425  
## F occurrances of GN names Total F occurrances  
## 1 73973 1307926  
## 2 73607 1295259  
## 3 75797 1294652  
## 4 80402 1296592  
## 5 85416 1325930  
## 6 89910 1322802  
## 7 93287 1312169  
## 8 91657 1275695  
## 9 92724 1257343  
## 10 93043 1235481  
## 11 93268 1180696  
## Ratio of F GN Occur to Total F Occur M occurrances of GN names  
## 1 0.05655748 149710  
## 2 0.05682802 149875  
## 3 0.05854623 151182  
## 4 0.06201025 152187  
## 5 0.06441969 160444  
## 6 0.06796936 162211  
## 7 0.07109374 155867  
## 8 0.07184868 155443  
## 9 0.07374599 150056  
## 10 0.07530913 146688  
## 11 0.07899408 142672  
## Total M occurrances Ratio of M GN Occur to Total M Occur  
## 1 1615800 0.09265379  
## 2 1599083 0.09372559  
## 3 1594004 0.09484418  
## 4 1592997 0.09553502  
## 5 1616173 0.09927403  
## 6 1609215 0.10080132  
## 7 1585894 0.09828337  
## 8 1538982 0.10100378  
## 9 1501581 0.09993201  
## 10 1473990 0.09951764  
## 11 1404385 0.10159038  
## Total GN Occurrances Total Occurrances  
## 1 223683 2923726  
## 2 223482 2894342  
## 3 226979 2888656  
## 4 232589 2889589  
## 5 245860 2942103  
## 6 252121 2932017  
## 7 249154 2898063  
## 8 247100 2814677  
## 9 242780 2758924  
## 10 239731 2709471  
## 11 235940 2585081  
## Ratio of GN Occurrances to Total Occurrances  
## 1 0.07650614  
## 2 0.07721340  
## 3 0.07857599  
## 4 0.08049207  
## 5 0.08356608  
## 6 0.08598893  
## 7 0.08597260  
## 8 0.08778982  
## 9 0.08799807  
## 10 0.08847889  
## 11 0.09126987