#Class: Week 02
#Course: Big Data and Social Analysis
#Semester: Spring 2021
#Lesson: R, Vector, and Object
#Instructor: Chung-pei Pien
#Organization: ICI, NCCU

Student Information --------
#Chinese Name: 辛鐘成
#English First Name: Jongsung Shin
#UID: 110ZU1038
#E-mail: sjongsung97@gmail.com

Questions --------
#Please answer the following questions. Remember, No Comments, No Points!!!!!!!

#Question 1: (3 points)
#From 115-117 terms of US house, which term has the largest number of lawmakers?
length(birth_115) # 456
length(birth_116) # 451
length(birth_117) # 447

I used 'length' to know how many elements in each term.
It is seen that the number of lawmakers has been decreased in the following term.

#Question 2: (9 points)
#From 115-117 terms of US house, which term has worse gender inequality performance?
a <- length(gender_115[gender_115 == 'M']) # 363
b <- length(gender_115[gender_115 == 'F']) # 93
a/b # 3.90

c <- length(gender_116[gender_116 == 'M']) # 345
d <- length(gender_116[gender_116 == 'F']) # 106
c/d # 3.25

e <- length(gender_117[gender_117 == 'M']) # 320
f <- length(gender_117[gender_117 == 'F']) # 127
e/f # 2.52

Obviously, male ratio is higher than female based on the results.
I assigned an alphabet variable to gender in each term to calculate easily
It is seen that the gender ratio has been narrowed in the following term.

```
#Question 3: (9 points)
#From 115-117 terms of US house, which term's age is oldest?

year_115 <- substr(birth_115,1,4)
year_115 <- as.numeric(year_115)
age_115 <- 2022 - year_115
mean(age_115) # 63.50
median(age_115) # 64
max(age_115) # 93

year_116 <- substr(birth_116,1,4)
year_116 <- as.numeric(year_116)
age_116 <- 2022 - year_116
mean(age_116) # 60.90
median(age_116) # 61
max(age_116) #89

year_117 <- substr(birth_117,1,nchar(birth_117)-6)
year_117 <- as.numeric(year_117)
age_117 <- 2022 - year_117
mean(age_117) # 59.52
median(age_117) # 60
max(age_117) #89

# I wasn't sure if the question asked me to get the mean or the max age. Therefore, I got it all.
# I used 'nchar' for the last one to show another way to get the result.
# According to the data I got, term 115 has the highest mean and age at 63.50 and 93, followed by term 116 and 117.
# It is seen that their average age and the max age has been decreased since the term 115.
```

#Question 4: (9 points)
#Please tell me the control party of 115-117 term of US House
unique(party_115) # R, D
unique(party_116) # R, D, I
unique(party_117) # R, D

table(party_115)
sort(table(party_115), decreasing = TRUE) # R:252 / D: 204

table(party_116)
sort(table(party_116), decreasing = TRUE) # D:241 / R:208 / I:2

table(party_117)
sort(table(party_117), decreasing = TRUE) # D:230 / R: 217

I found a special party "I" in the term 116 by using 'unique'. I wonder if there is other party in other term. But only the term 116 has it.
I used table to see the proportion clearly and sort it in descending order.
Except the term 115 when republican party controlled the party, democratic party has controlled the term 116 and 117.
It is seen that, from the term 116, democratic party tend to be more popular than republican party.

#Now you have 30 points! Question 5 is a bonus!
#Question 5: (5 points)
#Please do research to tell me why gender inequality becomes worse or better from 115-117 terms of US House.

According to the data above, the gap ratio between male and female has been narrowed every term, which means the gender equality becomes better.
White House has recognized that the problem of longstanding gender discrimination and the systemic barriers to full participation for women. They want it to be fixed for a better society.
#
https://www.whitehouse.gov/briefing-room/statements-releases/2021/10/22/fact-sheet-national-strategy-on-gender-equity-and-equality/