Gapminder_exercise

Rikke Uldbæk

10/31/2022

```
#packages
pacman::p_load(tidyverse, ggplot2, gapminder, stringr)
library(gapminder)

#loading data from gapminder package
df <- gapminder
df <- as.data.frame(df)
```

Loading the data and packages

1) Define a Function

Define a defensive function that calculates the Gross Domestic Product of a nation from the data available in the gapminder dataset. You can use the population and GDPpercapita columns for it. Using that function, calculate the GDP of Denmark in the following years: 1967, 1977, 1987, 1997, 2007, and 2017.

```
#Define a function to calculate GDP of a nation for a given year, using the columns "population" and "G
GDP_func <- function(data, chosen_nation, chosen_year) { #function arguments

  #isolate specified year and country
  data <- data[data$country %in% chosen_nation, ]
  data <- data[data$year %in% chosen_year, ]

  #calculate GDP
  GDP <- data$pop * data$gdpPercap
  data$GDP <- GDP

  return(data)
}

# function works fine
GDP_func(df, "Bulgaria", 1967)
```

```
##      country continent year lifeExp     pop gdpPercap         GDP
## 184 Bulgaria    Europe 1967   70.42 8310226  5577.003 46346153671
```

Applying the function on Denmark

```r
#Using the function on Denmark
dk_years <- c(1967, 1977, 1987, 1997, 2007, 2017)

for (i in dk_years){
  if(i %in% df$year){
    dk_GDP <- GDP_func(df, "Denmark", i)
    print(dk_GDP)
  }else{
    print(paste0("data was not found for year: ",i ))
  }
}
```

```
##     country continent year lifeExp     pop gdpPercap         GDP
## 412 Denmark    Europe 1967   72.96 4838800  15937.21 77116977700
##     country continent year lifeExp     pop gdpPercap          GDP
## 414 Denmark    Europe 1977   74.69 5088419   20422.9 103920280028
##     country continent year lifeExp     pop gdpPercap          GDP
## 416 Denmark    Europe 1987    74.8 5127024  25116.18 128771236166
##     country continent year lifeExp     pop gdpPercap          GDP
## 418 Denmark    Europe 1997   76.11 5283663  29804.35 157476118456
##     country continent year lifeExp     pop gdpPercap          GDP
## 420 Denmark    Europe 2007  78.332 5468120  35278.42 192906627081
## [1] "data was not found for year: 2017"
```

2) Make a Loop

Write a script that loops over each country in the gapminder dataset, tests whether the country starts with a 'B' , and prints out whether the life expectancy is smaller than 50, between 50 and 70, or greater than 70. (Hint: remember the grepl function, and review the Control Flow tutorial)

```r
# creating an initial dataframe
b_df <- data.frame()

# loop 1
for(i in unique(df$country)) {
  if(grepl("^B", i)){
    b_df <- rbind(b_df, df[which(df$country %in% i),]) }

}

# loop 2
for(j in unique(b_df$country)){

  new_df <- b_df %>%
    filter(country == j)

  # Calculation of mean life expectancy
  mean <- mean(new_df$lifeExp)

  # Output of each B-country and their mean life expectancy and
  # whether they have a long-, short-, and "medium" life expectancy
   if (mean > 70) {
```

```
     print(paste0(j, " has a long life expectancy above 70 years"))
    } else {
      if (mean < 50) {
       print(paste0(j, " has a short life expectancy below 50 years"))
     } else {
      print(paste0(j, " has a life expectancy between 50 and 70 years of age"))
    }
  }
}
```

```
## [1] "Bahrain has a life expectancy between 50 and 70 years of age"
## [1] "Bangladesh has a short life expectancy below 50 years"
## [1] "Belgium has a long life expectancy above 70 years"
## [1] "Benin has a short life expectancy below 50 years"
## [1] "Bolivia has a life expectancy between 50 and 70 years of age"
## [1] "Bosnia and Herzegovina has a life expectancy between 50 and 70 years of age"
## [1] "Botswana has a life expectancy between 50 and 70 years of age"
## [1] "Brazil has a life expectancy between 50 and 70 years of age"
## [1] "Bulgaria has a life expectancy between 50 and 70 years of age"
## [1] "Burkina Faso has a short life expectancy below 50 years"
## [1] "Burundi has a short life expectancy below 50 years"
```

For some reason, probably my poor coding skills, I could not merge loop 1 and loop 2. However, I solved the exercise anyway, and allowed my self to calculate the mean of life expectancy for each county (starting with B) over the years, to minimize the output and make it more manageable.

3) Optional

I choose not to make this exercise.