6:W43: Practicing functions with Gapminder

10/30/2022

Use the gapminder dataset from Week 43 to produce solutions to the three tasks below. Post the .R script or .Rmd and .html in your au##### github repository and link it here: https://github.com/Digital-Methods-HASS/CultDat_LauraWPaaby (https://github.com/Digital-Methods-HASS/CultDat_LauraWPaaby) *OBS: the assignment can be found under* **Homework/Answers.**

Loading in libraries and data:

```
library(gapminder)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──✔
ggplot2 3.3.6      ✔ purrr   0.3.4
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.1      ✔ stringr 1.4.0
## ✔ readr   2.1.2      ✔ forcats 0.5.2 ── Conflicts ────────────────────────────────────
────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(knitr)
```

```
data <- as.data.frame(gapminder)
head(data)
```

```
##       country continent year lifeExp      pop gdpPercap
## 1 Afghanistan      Asia 1952  28.801  8425333  779.4453
## 2 Afghanistan      Asia 1957  30.332  9240934  820.8530
## 3 Afghanistan      Asia 1962  31.997 10267083  853.1007
## 4 Afghanistan      Asia 1967  34.020 11537966  836.1971
## 5 Afghanistan      Asia 1972  36.088 13079460  739.9811
## 6 Afghanistan      Asia 1977  38.438 14880372  786.1134
```

1. Define a defensive function that calculates the Gross Domestic Product (GDP) 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.

The GDP of a nation is found by multiplying the GDP per capita with the size of the population.

```
cal_GDP <- function(df, chosen_country, chosen_year){
  #### MAKING IT DEFENSIVE, BY TAKING CARE OF THE YEARS W NO DATA
  stopifnot(is.numeric(chosen_year))
  stopifnot(is.character(chosen_country))
  #### ISOLATE THE YEAR AND COUNTRY OF CHOICE
  df <- df[df$year %in% chosen_year, ]
  df <- df[df$country %in% chosen_country, ]

  GDP <- df$pop * df$gdpPercap
  return(GDP)
}


##### RUNNING THE FUNCTION FOR ALL YEARS:
years = c(1967, 1977, 1987, 1997, 2007, 2017)

for (i in years){
  if(i %in% data$year){
    gdp <- cal_GDP(data, "Denmark", i)
    print(paste0("The GDP in Denmark was ", gdp, " in ", i, "!"))
  } else {
    print(paste0("There were no data from year ", i))
  }
}
```

```
## [1] "The GDP in Denmark was 77116977699.724 in 1967!"
## [1] "The GDP in Denmark was 103920280027.729 in 1977!"
## [1] "The GDP in Denmark was 128771236166.089 in 1987!"
## [1] "The GDP in Denmark was 157476118455.789 in 1997!"
## [1] "The GDP in Denmark was 192906627080.569 in 2007!"
## [1] "There were no data from year 2017"
```

2. 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)

```
### LOOP TIME BABY
b_df <- data.frame()

## LETS FIRST FIND THE COUNTRIES STARTING W A B:
for(i in unique(data$country)) {
  if(grepl("^B", i)){
    b_df <- rbind(b_df, data[which(data$country %in% i),]) }

}


### NOW I EVALUATE THE LIFEEXP FOR EACH COUNTRY TO ALL TIMES IN ANOTHER LOOP:
for(ii in unique(b_df$country)){

  country_dat <- b_df %>%
    filter(country == ii)

  #### Finding the life expectancy mean over all times
  le_mean <- mean(country_dat$lifeExp)

  #### printing them for each country
   if (le_mean > 70) { ### finding the lifeexps larger than 70
      print(paste0("In ", ii, " they have a mean life expetancy above 70 years from 1
952 to 2007! It is: ", round(le_mean, digits = 2), " years!"))
    } else {
      if (le_mean < 50) { ### finding the lifeexps smaller than 50
       print(paste0("In ", ii, " they have a mean life expetancy below 50 years from
1952 to 2007! It is: ", round(le_mean, digits = 2), " years!"))
     } else {
      print(paste0("In ", ii, " they have a mean life expetancy between 50 and 70 yea
rs from 1952 to 2007! It is: ", round(le_mean, digits = 2), " years!")) ### finding t
he lifeexps in between
    }
  }
}
```

```
## [1] "In Bahrain they have a mean life expetancy between 50 and 70 years from 1952
to 2007! It is: 65.61 years!"
## [1] "In Bangladesh they have a mean life expetancy below 50 years from 1952 to 200
7! It is: 49.83 years!"
## [1] "In Belgium they have a mean life expetancy above 70 years from 1952 to 2007!
It is: 73.64 years!"
## [1] "In Benin they have a mean life expetancy below 50 years from 1952 to 2007! It
is: 48.78 years!"
## [1] "In Bolivia they have a mean life expetancy between 50 and 70 years from 1952
to 2007! It is: 52.5 years!"
## [1] "In Bosnia and Herzegovina they have a mean life expetancy between 50 and 70 y
ears from 1952 to 2007! It is: 67.71 years!"
## [1] "In Botswana they have a mean life expetancy between 50 and 70 years from 1952
to 2007! It is: 54.6 years!"
## [1] "In Brazil they have a mean life expetancy between 50 and 70 years from 1952 t
o 2007! It is: 62.24 years!"
## [1] "In Bulgaria they have a mean life expetancy between 50 and 70 years from 1952
to 2007! It is: 69.74 years!"
## [1] "In Burkina Faso they have a mean life expetancy below 50 years from 1952 to 2
007! It is: 44.69 years!"
## [1] "In Burundi they have a mean life expetancy below 50 years from 1952 to 2007!
It is: 44.82 years!"
```

This prints out the mean life expectancy of the countries across all measured times. The latter loop can likewise be applied, if one wish to find the life expectansies for all countries.