Practicing Functions with Gapminder

Marie Højlund Christiansen

31 Oktober, 2022

Contents

This exercise is based on the Gapminder dataset provided by Data Carpentry (https://raw.githubusercontent.com/swcarpentry/r-novice-gapminder/gh-pages/_episodes_rmd/data/gapminder_data.csv)

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.

I begin by loading the dataset into R:

```
gapminder_data <- read.csv('https://raw.githubusercontent.com/swcarpentry/r-novice-gapminder/gh-pages/_e
```

Then, I define a defensive function that calculates the Gross Domestic Product of a nation from the data available in the gapminder dataset and use it to calculate the GDP of Denmark in years 1967, 1977, 1987, 1997, 2007, and 2017:

```
# Defining function
# Function takes a dataset and multiplies the population column with the GDP per capita column
# Adding arguments to the function so I can extract the GDP per given years and given country
calcGDP <- function(dat, year=NULL, country=NULL) {
  if(!is.null(year)) {
    dat <- dat[dat$year %in% year, ]
  }
  if(!is.null(country)) {
    dat <- dat[dat$country %in% country, ]
  }
  gdp <- dat$pop * dat$gdpPercap
  return(gdp)

  new <- cbind(dat, gdp=gdp)
}

# Using the function to extract GDP in Denmark per given years
calcGDP(gapminder_data, year = c(1967, 1977, 1987, 1997, 2007, 2017), country = "Denmark")
```

```
## [1]  77116977700 103920280028 128771236166 157476118456 192906627081
```

I have now calculated the the GDP in Denmark in the years 1967, 1977, 1987, 1997, 2007 and 2017.

2) Write a script

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.

```
# Assigning threshold values to objects to use in for loop
# Creating an object containing countries beginning with 'B' to use in for loop
lowerThreshold <- 50
upperThreshold <- 70
CountriesWithB <- grep("^B", unique(gapminder_data$country), value = TRUE)

# Creating for loop
# Prints whether the life expectancy in the countries in the 'CountriesWithB'-object is smaller than 50
for (iCountry in CountriesWithB) {
  tmp <- mean(gapminder_data[gapminder_data$country == iCountry, "lifeExp"])

  if (tmp < lowerThreshold) {
    cat("Average Life Expectancy in", iCountry, "is less than", lowerThreshold, "\n")
  } else if (tmp > lowerThreshold && tmp < upperThreshold) {
     cat("Average Life Expectancy in", iCountry, "is between", lowerThreshold, "and", upperThreshold, "
  } else {
     cat("Average Life Expectancy in", iCountry, "is greater than", upperThreshold, "\n")
  }
  rm(tmp)
}
```

```
## Average Life Expectancy in Bahrain is between 50 and 70
## Average Life Expectancy in Bangladesh is less than 50
## Average Life Expectancy in Belgium is greater than 70
## Average Life Expectancy in Benin is less than 50
## Average Life Expectancy in Bolivia is between 50 and 70
## Average Life Expectancy in Bosnia and Herzegovina is between 50 and 70
## Average Life Expectancy in Botswana is between 50 and 70
## Average Life Expectancy in Brazil is between 50 and 70
## Average Life Expectancy in Bulgaria is between 50 and 70
## Average Life Expectancy in Burkina Faso is less than 50
## Average Life Expectancy in Burundi is less than 50
```

I have now run 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. Above, we see a list over the countries beginning with "B", letting us know whether the life expectancy in these countries is under 50, between 50 and 70 or over 70.