02-mi-segundo-script-paquetes-lectura-datos.R

Joshua

Mon Aug 27 21:12:58 2018

# P: Que es lo primero que hago?  
# R: Instalar paquetes!  
# Esto es una vez por instalacion de R (o formateo de PC XD)  
# install.packages(c("prophet", "tidyverse", "readxl", "writexl"))  
  
# P: Que #$%#$ es tidyverse?  
# R: Es un pack de paquetes  
  
# P: Como se usa/carga?  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.4.4

## -- Attaching packages -------------------

## v ggplot2 3.0.0.9000 v purrr 0.2.5   
## v tibble 1.4.2 v dplyr 0.7.6   
## v tidyr 0.8.1 v stringr 1.3.1   
## v readr 1.1.1 v forcats 0.3.0

## Warning: package 'tibble' was built under R version 3.4.4

## Warning: package 'tidyr' was built under R version 3.4.4

## Warning: package 'purrr' was built under R version 3.4.4

## Warning: package 'dplyr' was built under R version 3.4.4

## Warning: package 'stringr' was built under R version 3.4.4

## Warning: package 'forcats' was built under R version 3.4.4

## -- Conflicts --- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# P: Y donde estan los datos???  
# install.packages("gapminder"), o manual :´(  
library(gapminder) # cargamos el paquete que contiene los datos  
data(gapminder) # ponemos en el enviorment los datos  
gapminder # llamamos o imprimimos los datos

## # A tibble: 1,704 x 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Afghanistan Asia 1952 28.8 8425333 779.  
## 2 Afghanistan Asia 1957 30.3 9240934 821.  
## 3 Afghanistan Asia 1962 32.0 10267083 853.  
## 4 Afghanistan Asia 1967 34.0 11537966 836.  
## 5 Afghanistan Asia 1972 36.1 13079460 740.  
## 6 Afghanistan Asia 1977 38.4 14880372 786.  
## 7 Afghanistan Asia 1982 39.9 12881816 978.  
## 8 Afghanistan Asia 1987 40.8 13867957 852.  
## 9 Afghanistan Asia 1992 41.7 16317921 649.  
## 10 Afghanistan Asia 1997 41.8 22227415 635.  
## # ... with 1,694 more rows

# Entiendo el mundo -------------------------------------------------------  
  
# funcion clever para ojear  
glimpse(gapminder)

## Observations: 1,704  
## Variables: 6  
## $ country <fct> Afghanistan, Afghanistan, Afghanistan, Afghanistan, ...  
## $ continent <fct> Asia, Asia, Asia, Asia, Asia, Asia, Asia, Asia, Asia...  
## $ year <int> 1952, 1957, 1962, 1967, 1972, 1977, 1982, 1987, 1992...  
## $ lifeExp <dbl> 28.801, 30.332, 31.997, 34.020, 36.088, 38.438, 39.8...  
## $ pop <int> 8425333, 9240934, 10267083, 11537966, 13079460, 1488...  
## $ gdpPercap <dbl> 779.4453, 820.8530, 853.1007, 836.1971, 739.9811, 78...

# P: Cuantos paises hay?  
count(gapminder, country)

## # A tibble: 142 x 2  
## country n  
## <fct> <int>  
## 1 Afghanistan 12  
## 2 Albania 12  
## 3 Algeria 12  
## 4 Angola 12  
## 5 Argentina 12  
## 6 Australia 12  
## 7 Austria 12  
## 8 Bahrain 12  
## 9 Bangladesh 12  
## 10 Belgium 12  
## # ... with 132 more rows

# V2 (con el simbolito raro pipe, CTRL+SHIFT+M)  
# se lee como \_luego\_  
gapminder %>%   
 count(country)

## # A tibble: 142 x 2  
## country n  
## <fct> <int>  
## 1 Afghanistan 12  
## 2 Albania 12  
## 3 Algeria 12  
## 4 Angola 12  
## 5 Argentina 12  
## 6 Australia 12  
## 7 Austria 12  
## 8 Bahrain 12  
## 9 Bangladesh 12  
## 10 Belgium 12  
## # ... with 132 more rows

# P: Para que me sirve el %>% %>% %>% ?  
sqrt(2)

## [1] 1.414214

2 %>% sqrt()

## [1] 1.414214

tan(exp(sin(log(sqrt(2)))))

## [1] 5.95761

2 %>%   
 sqrt() %>%   
 log() %>%   
 sin() %>%   
 exp() %>%   
 tan()

## [1] 5.95761

gapminder %>%   
 filter(continent == "Asia") %>%   
 arrange(year) %>%   
 group\_by(country) %>%   
 mutate(crecimiento\_anual = (pop - lag(pop))/pop) %>%   
 filter(country == "Japan")

## Warning: package 'bindrcpp' was built under R version 3.4.4

## # A tibble: 12 x 7  
## # Groups: country [1]  
## country continent year lifeExp pop gdpPercap crecimiento\_anual  
## <fct> <fct> <int> <dbl> <int> <dbl> <dbl>  
## 1 Japan Asia 1952 63.0 86459025 3217. NA   
## 2 Japan Asia 1957 65.5 91563009 4318. 0.0557   
## 3 Japan Asia 1962 68.7 95831757 6577. 0.0445   
## 4 Japan Asia 1967 71.4 100825279 9848. 0.0495   
## 5 Japan Asia 1972 73.4 107188273 14779. 0.0594   
## 6 Japan Asia 1977 75.4 113872473 16610. 0.0587   
## 7 Japan Asia 1982 77.1 118454974 19384. 0.0387   
## 8 Japan Asia 1987 78.7 122091325 22376. 0.0298   
## 9 Japan Asia 1992 79.4 124329269 26825. 0.0180   
## 10 Japan Asia 1997 80.7 125956499 28817. 0.0129   
## 11 Japan Asia 2002 82 127065841 28605. 0.00873  
## 12 Japan Asia 2007 82.6 127467972 31656. 0.00315

filter(mutate(group\_by(arrange(filter(gapminder, continent == "Asia"), year),country), crecimiento\_anual = (pop - lag(pop))/pop), country == "Japan")

## # A tibble: 12 x 7  
## # Groups: country [1]  
## country continent year lifeExp pop gdpPercap crecimiento\_anual  
## <fct> <fct> <int> <dbl> <int> <dbl> <dbl>  
## 1 Japan Asia 1952 63.0 86459025 3217. NA   
## 2 Japan Asia 1957 65.5 91563009 4318. 0.0557   
## 3 Japan Asia 1962 68.7 95831757 6577. 0.0445   
## 4 Japan Asia 1967 71.4 100825279 9848. 0.0495   
## 5 Japan Asia 1972 73.4 107188273 14779. 0.0594   
## 6 Japan Asia 1977 75.4 113872473 16610. 0.0587   
## 7 Japan Asia 1982 77.1 118454974 19384. 0.0387   
## 8 Japan Asia 1987 78.7 122091325 22376. 0.0298   
## 9 Japan Asia 1992 79.4 124329269 26825. 0.0180   
## 10 Japan Asia 1997 80.7 125956499 28817. 0.0129   
## 11 Japan Asia 2002 82 127065841 28605. 0.00873  
## 12 Japan Asia 2007 82.6 127467972 31656. 0.00315

# P: Cada pais tiene la misma cantidad de registros?  
gapminder %>%   
 count(country) %>%   
 count(n)

## # A tibble: 1 x 2  
## n nn  
## <int> <int>  
## 1 12 142

# R: Efectivamente todas tienen la misma cantidad blabla  
  
  
# P: Cual es el pais que ha tenido menor   
# poblacion en la historia de los paises  
# around the world? y diga el año  
#   
# hint:   
x <- c(2, 3, 1)   
x

## [1] 2 3 1

x == min(x)

## [1] FALSE FALSE TRUE

# R:  
gapminder %>%   
 arrange(pop) %>%   
 head(1)

## # A tibble: 1 x 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Sao Tome and Principe Africa 1952 46.5 60011 880.

gapminder %>%   
 filter(pop == min(pop))

## # A tibble: 1 x 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Sao Tome and Principe Africa 1952 46.5 60011 880.

# P: Que pais ha tenido la mayor poblacion  
# en promedio de en los datos registrados?  
gapminder %>%   
 group\_by(country) %>%   
 summarise(  
 promedio\_pop = mean(pop),  
 esperanza\_vida\_max = max(lifeExp)  
 ) %>%   
 # filter(promedio\_pop == max(promedio\_pop))  
 arrange(desc(promedio\_pop))

## # A tibble: 142 x 3  
## country promedio\_pop esperanza\_vida\_max  
## <fct> <dbl> <dbl>  
## 1 China 958160052. 73.0  
## 2 India 701130740. 64.7  
## 3 United States 228211232. 78.2  
## 4 Indonesia 148322833. 70.6  
## 5 Brazil 122312127. 72.4  
## 6 Japan 111758808 82.6  
## 7 Pakistan 93683386. 65.5  
## 8 Bangladesh 90755395. 64.1  
## 9 Germany 77547043. 79.4  
## 10 Nigeria 73708018. 47.5  
## # ... with 132 more rows

# hint  
x <- c(1:10)  
x

## [1] 1 2 3 4 5 6 7 8 9 10

sqrt(x)

## [1] 1.000000 1.414214 1.732051 2.000000 2.236068 2.449490 2.645751  
## [8] 2.828427 3.000000 3.162278

mean(x)

## [1] 5.5

max(x)

## [1] 10