wrangling

Christin Khan

October 22, 2018

# install packages

library(tidyverse) # install.packages("tidyverse")

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

## -- Attaching packages --------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.1 v dplyr 0.7.4  
## v tidyr 0.8.1 v stringr 1.2.0  
## v readr 1.1.1 v forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.4.4

## Warning: package 'tibble' was built under R version 3.4.3

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

## Warning: package 'readr' 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 3.4.3

## 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()

# read in data

## read gapminder csv. Note the readr:: prefix identifies which package it's in  
gapminder <- readr::read\_csv('https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/gapminder.csv')

## Parsed with column specification:  
## cols(  
## country = col\_character(),  
## year = col\_integer(),  
## pop = col\_double(),  
## continent = col\_character(),  
## lifeExp = col\_double(),  
## gdpPercap = col\_double()  
## )

# look at the data

head(gapminder)

## # A tibble: 6 x 6  
## country year pop continent lifeExp gdpPercap  
## <chr> <int> <dbl> <chr> <dbl> <dbl>  
## 1 Afghanistan 1952 8425333 Asia 28.8 779  
## 2 Afghanistan 1957 9240934 Asia 30.3 821  
## 3 Afghanistan 1962 10267083 Asia 32.0 853  
## 4 Afghanistan 1967 11537966 Asia 34.0 836  
## 5 Afghanistan 1972 13079460 Asia 36.1 740  
## 6 Afghanistan 1977 14880372 Asia 38.4 786

tail (gapminder)

## # A tibble: 6 x 6  
## country year pop continent lifeExp gdpPercap  
## <chr> <int> <dbl> <chr> <dbl> <dbl>  
## 1 Zimbabwe 1982 7636524 Africa 60.4 789  
## 2 Zimbabwe 1987 9216418 Africa 62.4 706  
## 3 Zimbabwe 1992 10704340 Africa 60.4 693  
## 4 Zimbabwe 1997 11404948 Africa 46.8 792  
## 5 Zimbabwe 2002 11926563 Africa 40.0 672  
## 6 Zimbabwe 2007 12311143 Africa 43.5 470

summary (gapminder)

## country year pop continent   
## Length:1704 Min. :1952 Min. :6.001e+04 Length:1704   
## Class :character 1st Qu.:1966 1st Qu.:2.794e+06 Class :character   
## Mode :character Median :1980 Median :7.024e+06 Mode :character   
## Mean :1980 Mean :2.960e+07   
## 3rd Qu.:1993 3rd Qu.:1.959e+07   
## Max. :2007 Max. :1.319e+09   
## lifeExp gdpPercap   
## Min. :23.60 Min. : 241.2   
## 1st Qu.:48.20 1st Qu.: 1202.1   
## Median :60.71 Median : 3531.8   
## Mean :59.47 Mean : 7215.3   
## 3rd Qu.:70.85 3rd Qu.: 9325.5   
## Max. :82.60 Max. :113523.1