Untitled

Sim, Ph.D.

2019뀈 1썡 13씪

Table of Contents

## 0. Setting

#### 0. 강의노트 마지막 페이지에 실습과제를 내드렸는데, 아직 Rmarkdown을 다루지 않았으므로 R을 이용하는 버전으로 바꿔서 아래에 내 드리겠습니다.

#### 1. 이메일에 첨부된 lifeCountry.csv파일을 다운받으세요.

#### 2. C:/LS-DS/classProject라는 폴더를 만드세요.

#### 3. Rstudio를 열어서 File -> New File -> R Script를 하면 새로운 소스 파일이 생성됩니다.

#### 4. 이를 File -> Save를 이용해서 위의 폴더에 classProject1.R이라는 파일로 저장하세요.

#### 5. RStudio를 완전히 닫고 탐색기에서 classProject1.R을 더블클릭하여 엽니다.

setwd("C:/LS-DS/classProject")

## 1. 불러오기

#### 6. 아래 명령을 사용해 파일을 불러옵니다.

dataset <- read.csv("lifeCountry.csv", stringsAsFactors = FALSE)

### 7. library(dplyr)과 library(ggplot2)를 실행합니다.

library(dplyr)  
library(ggplot2)

### 8. 데이터는 총 몇개의 행과 열로 되어 있습니까? (hint: str)

데이터셋 dataset에는 161개의 행과 11개의 열로 구성이 되어있습니다.

library(ggplot2)

# 9. 가장 GDP가 높고 낮은 나라는 어디인가요? 가장 기대수명이 길고 짧은 나라는 어디인가요?

정답: Luxembourg

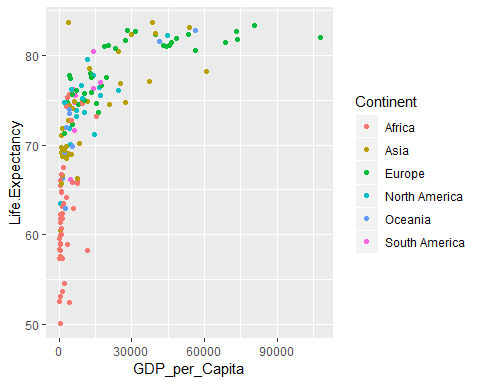
# 10. 대륙별로 GDP와 기대수명의 평균을 구해보세요.

dataset %>% group\_by(Continent) %>% summarise(mean(GDP\_per\_Capita), mean(Life.Expectancy))

## # A tibble: 6 x 3  
## Continent `mean(GDP\_per\_Capita)` `mean(Life.Expectancy)`  
## <chr> <dbl> <dbl>  
## 1 Africa 2618. 62.4  
## 2 Asia 13321. 73.6  
## 3 Europe 28589. 78.4  
## 4 North America 11175. 74.5  
## 5 Oceania 13483. 72.5  
## 6 South America 8957. 74.7

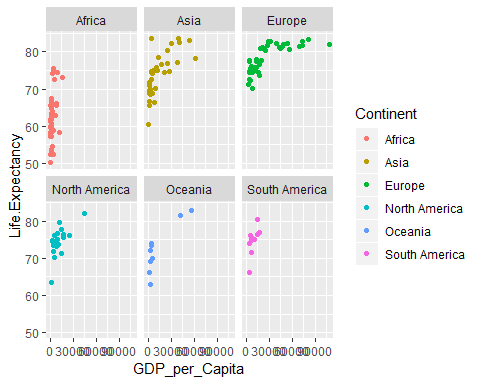
# 11. 각 나라의 GDP를 x축으로, 기대수명을 y축으로 산점도를 그리고 대륙에 따라서 점의 색깔이 달라지게 해보세요.

a <- ggplot(dataset) +   
 geom\_point(aes(x = GDP\_per\_Capita, y = Life.Expectancy, color = Continent))  
a



# 12. facet을 이용해서 대륙별로 GDP와 기대수명에 대한 산점도를 그려보세요.

a + facet\_wrap(~ Continent)



# 13. 국가별로 성별에 따라 기대수명이 다릅니다. mutate함수를 사용해서 ageSexDiff라는 변수를 만들어 보세요.

dataset <- dataset %>% mutate(ageSexDiff = Female - Male)

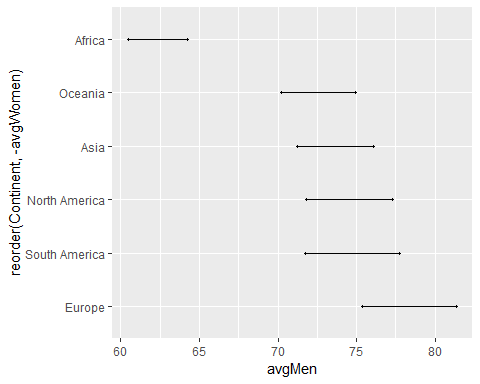
# 14. 어떤 대륙에서 ageSexDiff가 가장 크고 작은가요?

dataset %>%   
 group\_by(Continent) %>%   
 summarise(contiSexDiff = mean(ageSexDiff)) %>%   
 arrange(desc(contiSexDiff))

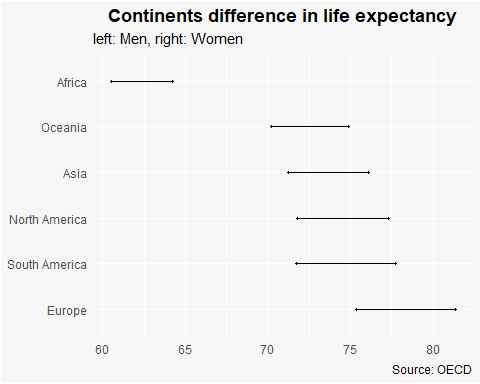
## # A tibble: 6 x 2  
## Continent contiSexDiff  
## <chr> <dbl>  
## 1 Europe 6.00  
## 2 South America 5.97  
## 3 North America 5.52  
## 4 Asia 4.86  
## 5 Oceania 4.71  
## 6 Africa 3.72

# 15. 자유롭게 분석을 시작해보세요.

# Dumbell plot (p25 in M24-ggplot2\_Gallery)  
library(ggalt)  
continentGender <- dataset %>%   
 group\_by(Continent) %>%   
 summarise(avgMen = mean(Male), avgWomen = mean(Female)) %>%   
 arrange(desc(avgWomen))  
b <- ggplot(continentGender,   
 aes(x = avgMen,   
 xend = avgWomen,   
 y= reorder(Continent, -avgWomen),   
 group = Continent)) +  
 geom\_dumbbell()  
b



c <- b +   
 labs(x=NULL, y=NULL, title="Continents difference in life expectancy",  
 subtitle = "left: Men, right: Women",  
 caption = "Source: OECD") +  
 theme(plot.title = element\_text(hjust=0.5, face="bold"),  
 plot.background=element\_rect(fill="#f7f7f7"),  
 panel.background=element\_rect(fill="#f7f7f7"),  
 axis.ticks=element\_blank(),  
 legend.position="top",  
 panel.border=element\_blank())  
c



# 18. classProject1.R을 저장해서 [learningSpoonsR@gmail.com](mailto:learningSpoonsR@gmail.com) 로 보내주세요.