

FMB819: R을 이용한 데이터분석

Introduction

2025-02-13

Welcome to FMB 819!



데이터 분석을 통해 답하고자 하는 질문

...continued

R이란?

데이터 처리 (Data Wrangling)

```
1 # load gapminder package
2 library(gapminder)
3 # load the dataset from the gapminder package
4 data(gapminder, package = "gapminder")
5 # show first 4 lines of this dataframe
6 head(gapminder, n = 4)
```

```
# A tibble: 4 × 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.
```

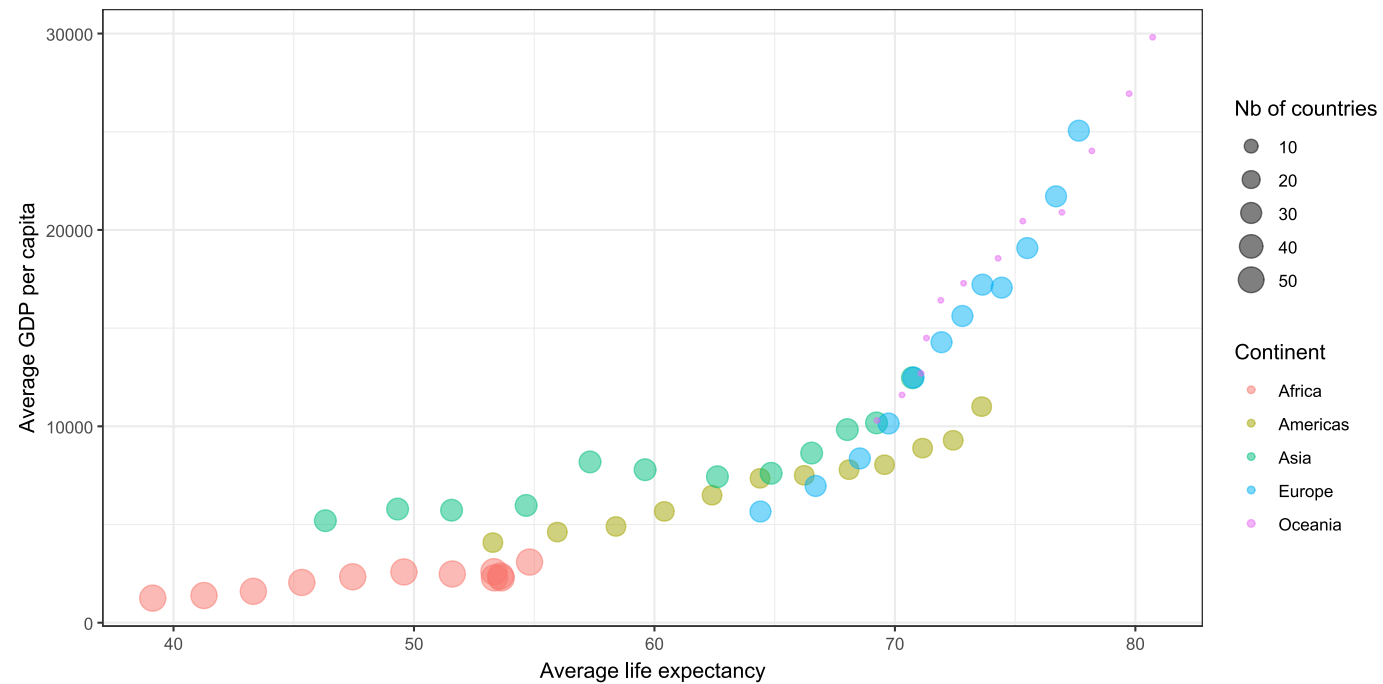
...continued

```
1 library(dplyr)
2 # compute the required statistics
3 gapminder_dplyr <- gapminder %>%
4   group_by(continent, year) %>%
5   summarise(count = n(),
6             mean_lifeexp = mean(lifeExp),
7             mean_gdppercap = mean(gdpPercap))
8 # show first 5 lines of the new data
9 head(gapminder_dplyr, n = 5)
```

```
# A tibble: 5 × 5
# Groups:   continent [1]
  continent year count mean_lifeexp mean_gdppercap
  <fct>      <int> <int>      <dbl>      <dbl>
1 Africa    1952    52        39.1       1253.
2 Africa    1957    52        41.3       1385.
3 Africa    1962    52        43.3       1598.
4 Africa    1967    52        45.3       2050.
5 Africa    1972    52        47.5       2340.
```

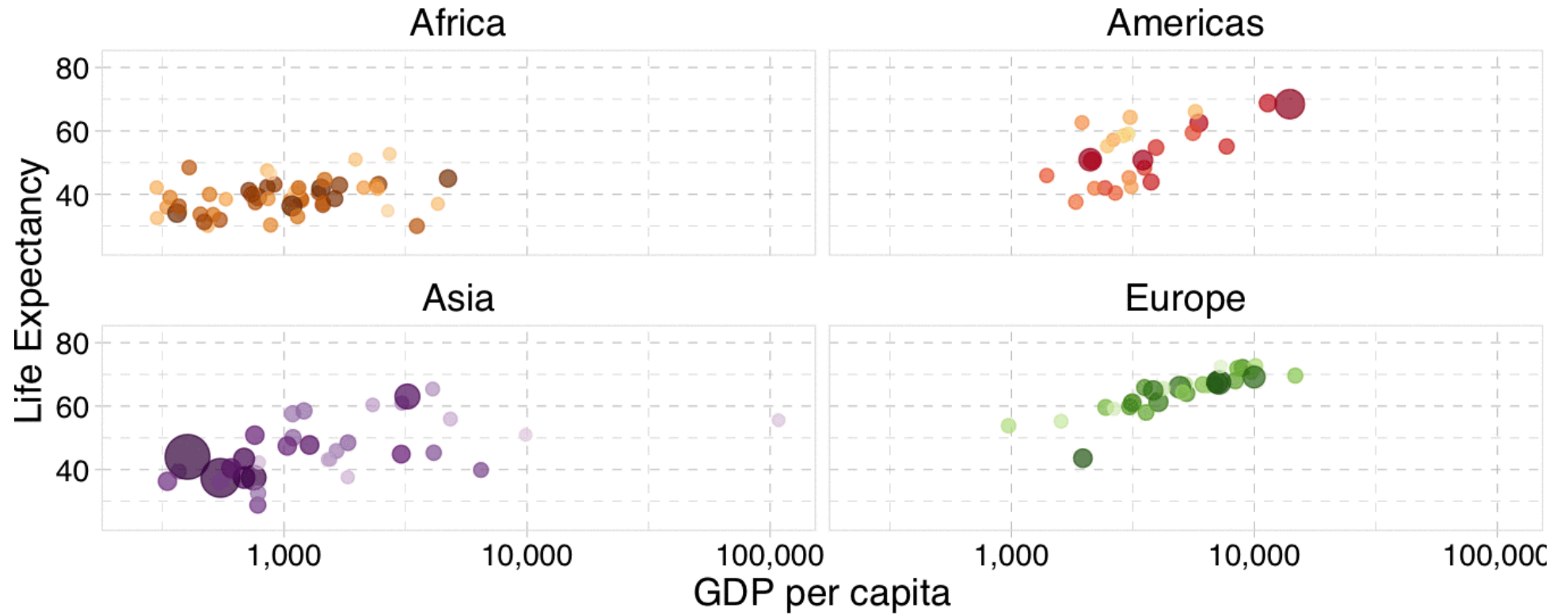
Visualisation

```
1 ggplot(data = gapminder_dplyr,  
2       mapping = aes(x = mean_lifeexp,  
3                     y = mean_gdppercap,  
4                     color = continent,  
5                     size = count)) +  
6   geom_point(alpha = 1/2) +  
7   labs(x = "Average life expectancy",  
8        y = "Average GDP per capita",  
9        color = "Continent",  
10       size = "Nb of countries") +  
11   theme_bw()
```



Animated Plotting

Year: 1952



Gapminder

출처: [Ed Rubin](#)

R 설치 방법: Windows 환경 기준

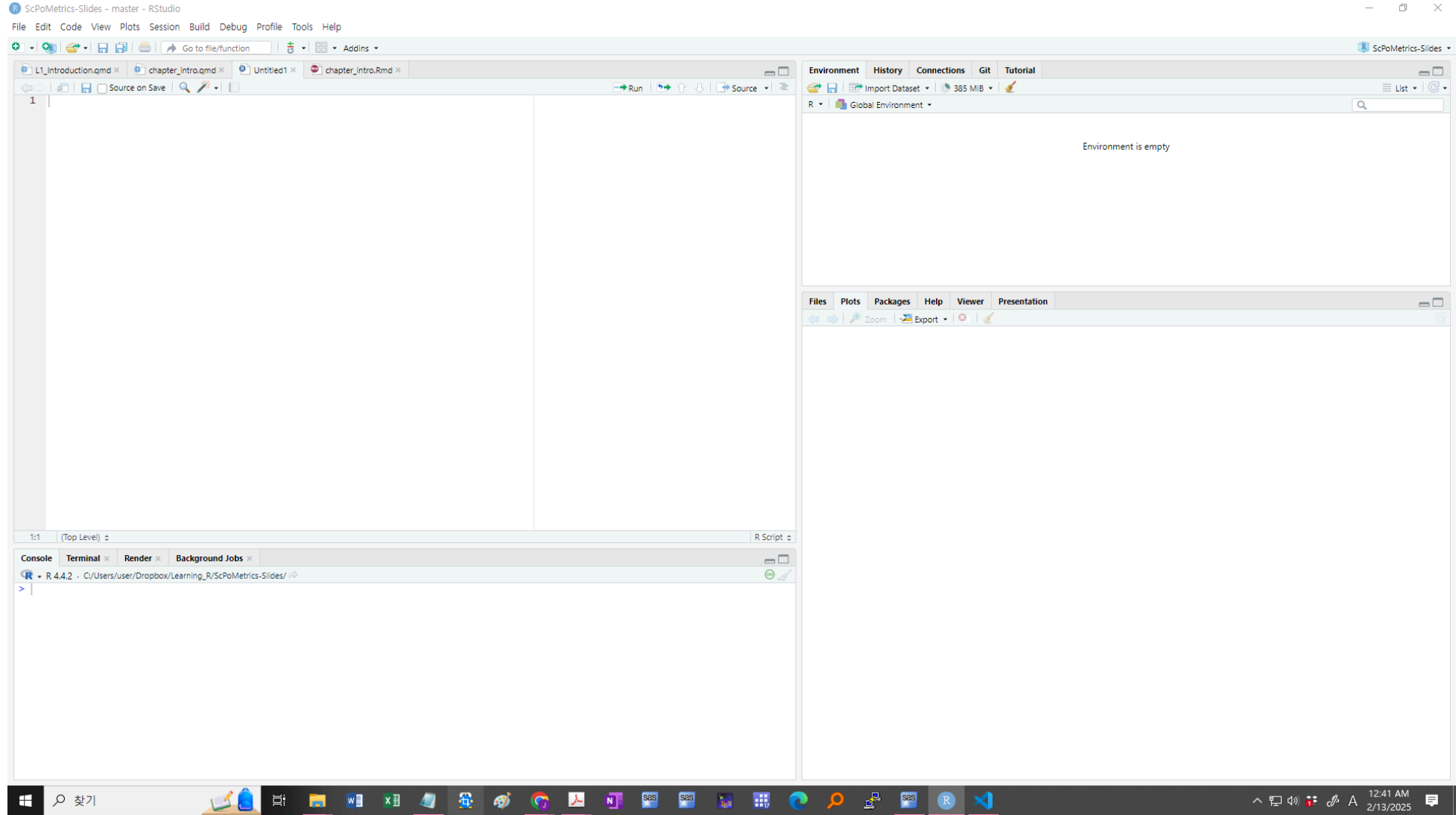
RStudio 설치 방법

R 프로그래밍 언어를 위한 통합 개발 환경(IDE, Integrated Development Environment)

설치 확인

설치가 정상적으로 완료되었는지 확인하기 위해 R과 RStudio 실행.

RStudio Layout



RStudio 환경

RStudio를 실행하면 여러 창(윈도우)으로 구성되어 있음.

콘솔과 프롬프트

R 스크립트 (Script)

[1] 15

Task 1



도움말 찾는 방법

R built-in help:

```
1 ?log #? in front of function
2 help(lm) # help() is equivalent
3 ??plot # get all help on keyword "plot"
```

In practice:



Jesse Maegan

@kierisi

Following



My **#rstats** learning path:

1. Install R
2. Install RStudio
3. Google "How do I [THING I WANT TO DO] in R?"

Repeat step 3 ad infinitum.

7:19 AM - 18 Aug 2017

Learning R



패키지 (Packages)

벡터 (Vectors)

```
1 c(1, 3, 5, 7, 8, 9)
```

```
[1] 1 3 5 7 8 9
```

```
1 (v <- c(42, "Statistics", TRUE))
```

```
[1] "42"      "Statistics" "TRUE"
```

* 소괄호: 객체 값을 바로 출력

```
1 1:10
```

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

```
1 v[c(1,3)]
```

```
[1] "42"      "TRUE"
```

data.frame

data.frame은 표 형식(tabular)의 데이터를 나타냄. 엑셀 스프레드시트와 유사.

```
1 example_data = data.frame(x = c(1, 3, 5, 7),  
2                           y = c(rep("Hello", 3), "Goodbye"),  
3                           z = c("one", 2, "three", 4))  
4 example_data
```

	x	y	z
1	1	Hello	one
2	3	Hello	2
3	5	Hello	three
4	7	Goodbye	4

data.frame ...continued

Dataframe을 설명하는 데 유용한 함수

```
1 str(example_data) # `str` 함수는 R 객체의 구조를 설명
```

```
'data.frame':  4 obs. of  3 variables:  
 $ x: num  1 3 5 7  
 $ y: chr  "Hello" "Hello" "Hello" "Goodbye"  
 $ z: chr  "one" "2" "three" "4"
```

```
1 names(example_data) # column names
```

```
[1] "x" "y" "z"
```

```
1 nrow(example_data) # number of rows
```

```
[1] 4
```

```
1 ncol(example_data) # number of columns
```

```
[1] 3
```

Task 2



데이터프레임 열 (column) 접근하기

```
1 first5 <- murders[1:5, ] # 처음 5개 주만 선택  
2 first5$state # $ 연산자로 추출
```

```
[1] "Alabama" "Alaska" "Arizona" "Arkansas" "California"
```

```
1 first5[, "state"] # 열 이름으로 추출
```

```
[1] "Alabama" "Alaska" "Arizona" "Arkansas" "California"
```

```
1 first5[, 1] # 첫 번째 열 가져오기
```

```
[1] "Alabama" "Alaska" "Arizona" "Arkansas" "California"
```

```
1 class(murders)
```

```
[1] "data.frame"
```

```
1 typeof(murders)
```

```
[1] "list"
```


data.frames: subset

```
1 # 총기 살인 사건이 500건 이상인 주만 선택하고 "state"와 "total" 변수만 유지
2 murders[murders$total > 500, c("state", "total")]
```

```
      state total
5 California 1257
10  Florida  669
33  New York  517
44    Texas  805
```

```
1 # 캘리포니아와 텍사스만 선택하고 "state"와 "total" 변수만 유지
2 murders[murders$state %in% c("California", "Texas"), c("state", "total")]
```

```
      state total
5 California 1257
44    Texas  805
```

* `subset` 명령어 사용 가능 (종종 더 직관적)

```
1 subset(murders, total > 500, select = c(state, total))
2 subset(murders, state %in% c("California", "Texas"), select = c(state, total))
```

Task 3

```
1 murders$total_percap = (murders$total / murders$population) * 10000
```

`murders` 객체를 클릭하여 새 변수를 확인해 보시오.

강의 관련 정보

성적 산출

수업 정책

Be nice. Be honest. Don't cheat.

숙제는 늦게 제출하면 안됨: 특별한 상황이 아닌 이상, 수업 끝나기 전에 모두 제출

부정행위 및 타인의 과제 무단 활용을 금지: 이유 막론하고 C나 F를 받게 됨

그룹으로 협력: 협력 장려, 다만 숙제나 프로젝트는 개인이 작성하여 제출

과제 제출 방법

Quarto Document로 작성, RPub에 Publish하여, 웹페이지 링크를 LMS에 제출.

1 새 Quarto 문서 생성

- RStudio에서 **File** → **New File** → **Quarto Document...** 선택
- **title**과 **포맷 (HTML)** 선택 후 **Create** 버튼 클릭
- 생성된 `.qmd` 파일에서 문서를 작성

2 과제 템플릿 예시

- 아래 YAML를 항상 포함할 것

```
---
title: "과제 제목"
author: "이름 (학번)"
date: "2025-02-13 01:59:54"
format: html_document
execute:
  echo: true
  include: true
---
```

3. Quarto 문서 렌더링

4. RPub에 게시하기

강의 계획서

주제 1: *Introduction*

주제 2 & 3: *Tidying, Visualising and Summarising Data*

주제 4: *Simple Linear Regression*

주제 5: *Introduction to Causality* (Optional)

주제 6: *Multiple Linear Regression*

주제 7: *Linear Regression Extensions* (Optional)

주제 8: *Sampling*

주제 9: *Confidence Intervals and Hypothesis Testing*

주제 10: *Statistical Inference*

기타 유용한 자료

통계/ 계량경제

R의 활용

SEE YOU NEXT WEEK!