

# TEST\_06

- **Guide**
  - **1. Data Wrangling Problem**
  - **2. Visualization Problem(Using ggplot2)**
- 

- 필요 패키지 : `dplyr`, `ggplot2`, `tidyr`, `gridExtra`
  - 필요 데이터 : `disease.csv`
- 

## Guide

- 시험시간 : 13:30 ~ 15:00 (총90분)
- R markdown을 이용하여 생성한 html파일 모두 제출
- Dropbox에 제출시 html 파일과 rmd 파일 모두 업로드 해야함
  - 공유된 폴더안에 시험날짜 폴더를 만들고 (예 : 20211125 )
  - 관련 데이터는 Data폴더 를 만들어 제출
  - html,rmd파일은 Rmd폴더 에 학번\_이름으로 작성 제출 (예 : 202101234\_홍길동)
- 문제에서 주어진 결과와 같게 작성
  - ex) 작성방식 : 1.1번 문제라고 적고 R Chunk를 생성하여 그 안에 해당 code 기입
- 업로드 및 수정시간이 모두 기록 됨에 유의
- 코드공유 적발시 0점 처리
- 문제에서 출력한대로 출력(전체데이터 출력은 지양)
- 위의 유의사항을 어길 시 감점
- 시험시간을 넘겨서 제출하면 0점 처리



## o. Packages

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(gridExtra)
```

## Data Description

- id: unique identifier
- gender: “Male”, “Female” or “Other”
- age: age of the patient
- hypertension: 0 if the patient doesn’t have hypertension, 1 if the patient has hypertension
- heart\_disease: 0 if the patient doesn’t have any heart diseases, 1 if the patient has a heart disease
- ever\_married: “No” or “Yes”
- work\_type: order is “children”, “Never\_worked”, “Private”, “Self-employed”, “Govt\_job”
- Residence\_type: “Rural” or “Urban”
- avg\_glucose\_level: average glucose level in blood
- bmi: body mass index
- smoking\_status: order is “never smoked”, “formerly smoked”, “smokes”, “Unknown”
  - “Unknown” in smoking\_status means that the information is unavailable for this patient

- stroke: 1 if the patient had a stroke or 0 if not

# 1. Data Wrangling Problem

## 1.1 disease 를 불러오고 속성을 확인하시오.[5점]

- head()로 출력 할 것

```
##      id gender age hypertension heart_disease ever_married      work_
type
## 1  9046   Male  67              0              1          Yes      Pri
vate
## 2 51676 Female  61              0              0          Yes Self-empl
oyed
## 3 31112   Male  80              0              1          Yes      Pri
vate
## 4 60182 Female  49              0              0          Yes      Pri
vate
## 5  1665 Female  79              1              0          Yes Self-empl
oyed
## 6 56669   Male  81              0              0          Yes      Pri
vate
##      Residence_type avg_glucose_level  bmi  smoking_status stroke
## 1           Urban          228.69 36.6  formerly smoked        1
## 2           Rural          202.21 N/A   never smoked          1
## 3           Rural          105.92 32.5   never smoked          1
## 4           Urban          171.23 34.4           smokes          1
## 5           Rural          174.12  24   never smoked          1
## 6           Urban          186.21  29  formerly smoked          1
```

## 1.2 Data Descripton 을 보고 변수 값과 속성을 변경 후 str()을 이용 해 확인하시오. [20점]

- gender에서 Other는 제외하고 age는 40세 이상만 filtering 하시오[5점]
- age\_group열을 factor를 활용하여 생성하시오.[5점]
  - age\_group을 age 변수에 따라 적절히 변경하시오.(“40~49age”, ... “80~89age”까지 있음)
  - ex)age가 40세~49세이면 age\_group은 “40~49age”가 되게 하시오
- work\_type과 smoking\_status를 Data Description을 참고하여 변경하시오.[5점]
- bmi 열을 적절히 변경하시오.[5점]

```
## 'data.frame':    2940 obs. of  13 variables:
## $ id           : int  9046 51676 31112 60182 1665 56669 53882 1
0434 27419 60491 ...
## $ gender       : chr  "Male" "Female" "Male" "Female" ...
## $ age          : num  67 61 80 49 79 81 74 69 59 78 ...
## $ hypertension : int  0 0 0 0 1 0 1 0 0 0 ...
## $ heart_disease : int  1 0 1 0 0 0 1 0 0 0 ...
## $ ever_married  : chr  "Yes" "Yes" "Yes" "Yes" ...
## $ work_type     : Factor w/ 5 levels "children","Never_worked",.
.: 3 4 3 3 4 3 3 3 3 3 ...
## $ Residence_type : chr  "Urban" "Rural" "Rural" "Urban" ...
## $ avg_glucose_level: num  229 202 106 171 174 ...
## $ bmi          : num  36.6 NA 32.5 34.4 24 29 27.4 22.8 NA 24.2
...
## $ smoking_status : Factor w/ 4 levels "never smoked",...: 2 1 1 3
1 2 1 1 4 4 ...
## $ stroke        : int  1 1 1 1 1 1 1 1 1 1 ...
## $ age_group      : Factor w/ 5 levels "40~49age","50~59age",...: 3
3 5 1 4 5 4 3 2 4 ...
```

### 1.3 뇌졸중 환자와 그렇지 않은 사람에 대한 다음 2가지 통계량을 구하시오. [15점]

- 성별로 bmi와 glucose 변수에 대한 평균을 구하시오.[5점]
- 나이 그룹별 수를 구하고 비율을 구하시오.[10점]

```
## # A tibble: 4 x 4
## # Groups:   stroke [2]
##   stroke gender bmi_mean glucose_mean
##   <int> <chr>    <dbl>         <dbl>
## 1      0 Female    30.8          110.
## 2      0 Male     31.1          115.
## 3      1 Female    30.2          127.
## 4      1 Male     30.8          143.
```

```
## # A tibble: 10 x 4
## # Groups:   stroke [2]
##   stroke age_group    cnt  ratio
##   <int> <fct>      <int> <dbl>
## 1      0 40~49age     718 0.266
## 2      0 50~59age     786 0.291
## 3      0 60~69age     574 0.213
## 4      0 70~79age     475 0.176
## 5      0 80~89age     146 0.0541
## 6      1 40~49age      12 0.0498
## 7      1 50~59age      48 0.199
## 8      1 60~69age      47 0.195
## 9      1 70~79age      94 0.390
## 10     1 80~89age      40 0.166
```

## 1.4 데이터를 long format 형태로 바꾸고 다음과 같이 출력하시오. [10점]

- long format으로 바꾼 데이터는 long\_data로 저장할 것 [5점]
- 뇌졸중 여부, 성별, factor 별 value의 평균을 구하시오 [5점]

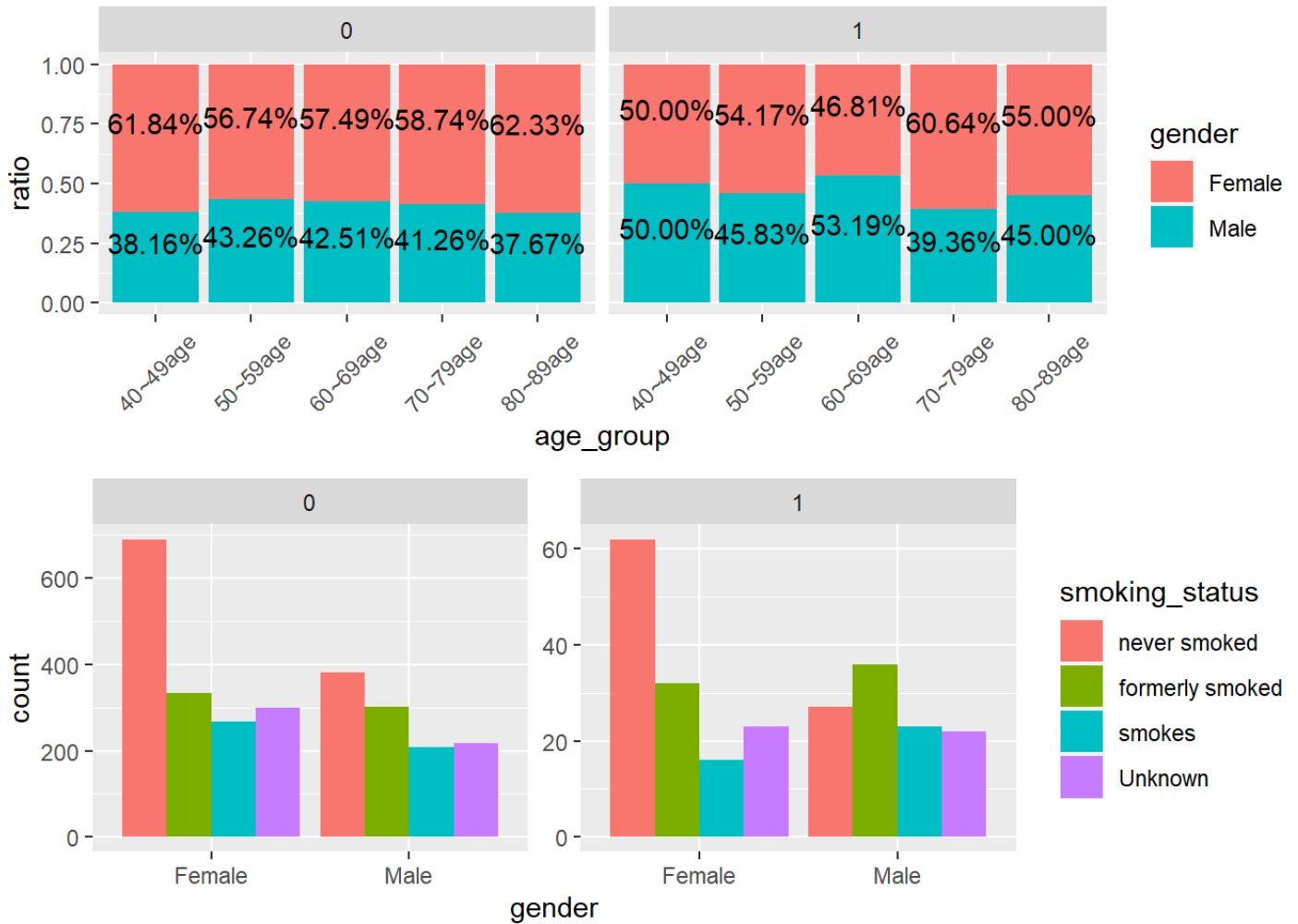
```
## # A tibble: 8 x 4
## # Groups:   stroke, gender [4]
##   stroke gender factor          val
##   <int> <chr>   <chr>      <dbl>
## 1      0 Female avg_glucose_level 110.
## 2      0 Female bmi          30.8
## 3      0 Male   avg_glucose_level 115.
## 4      0 Male   bmi          31.1
## 5      1 Female avg_glucose_level 127.
## 6      1 Female bmi          30.2
## 7      1 Male   avg_glucose_level 143.
## 8      1 Male   bmi          30.8
```

## 2. Visualization Problem(Using ggplot2)

### 2.1 뇌졸중 환자와 그렇지 않은 사람에 대한 그림을 그리시오 [30점]

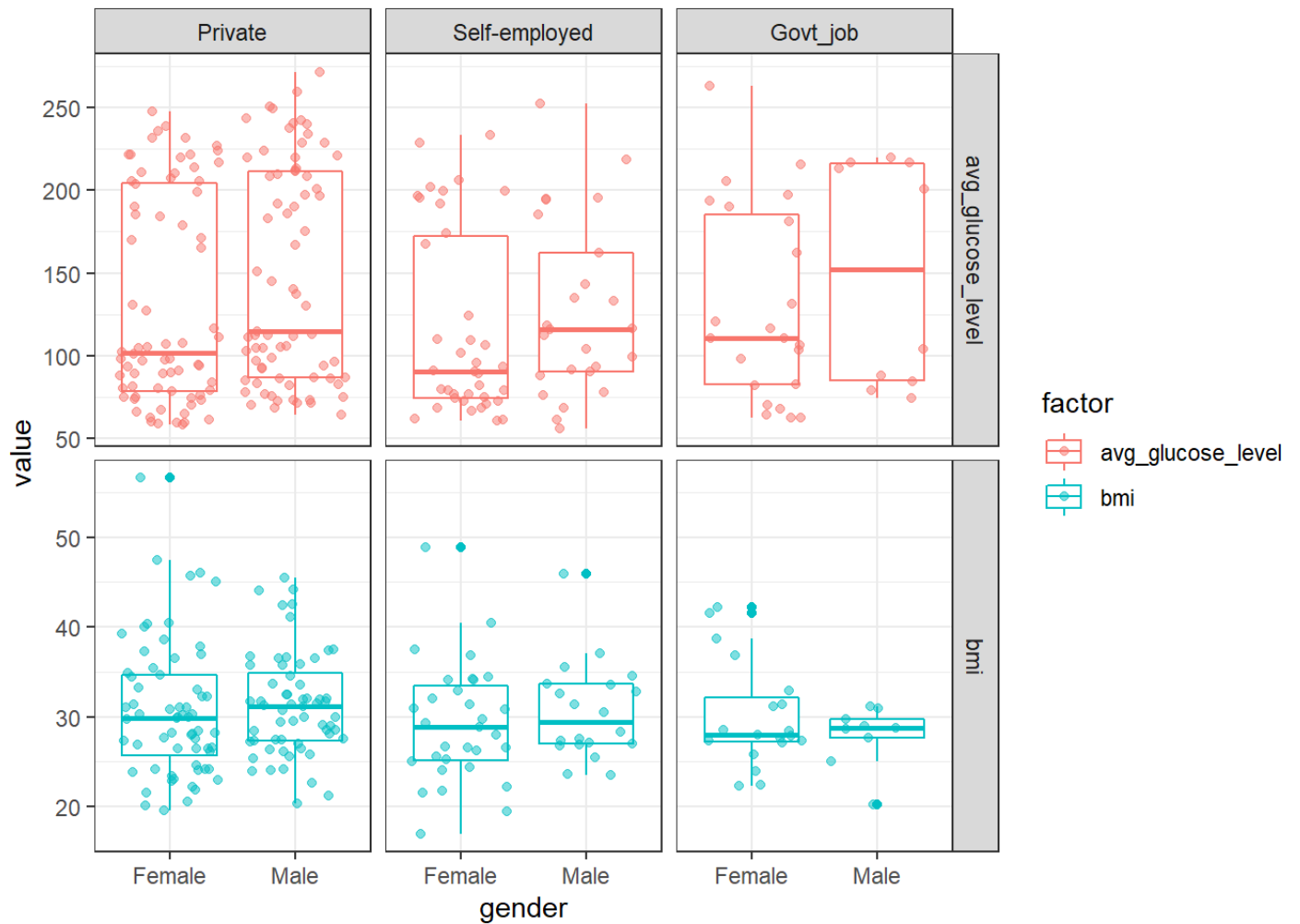
- plot1 부분
  - 뇌졸중 여부, 나이 그룹, 성별 count와 비율을 구하시오[5점]
    - head()로 출력할 것
  - 다음과 같이 geom\_bar와 geom\_text를 활용하여 다음과 같이 그리시오[10점]
- plot2 부분
  - 다음과 같이 geom\_bar를 활용하여 그림을 그리시오 [10점]
- gridExtra 패키지 내 함수를 이용해 한 플롯에 그릴 것.[5점]

```
## # A tibble: 6 x 5
## # Groups:   stroke, age_group [3]
##   stroke age_group gender      n ratio
##   <int> <fct>      <chr> <int> <dbl>
## 1      0 40~49age  Female   444 0.618
## 2      0 40~49age  Male     274 0.382
## 3      0 50~59age  Female   446 0.567
## 4      0 50~59age  Male     340 0.433
## 5      0 60~69age  Female   330 0.575
## 6      0 60~69age  Male     244 0.425
```



## 2.2 뇌졸중 환자에 대해 성별, 직업별, factor별 수에 대한 정보를 다음과 같이 그리시오. [20점]

- 뇌졸중 환자만 filtering 하여 그리시오 [5점]
- work\_type과 factor 별로 그림을 그리시오 [5점]
- 기본 그림[10점]



© 2021 Advice, All Rights Reserved

No part of this contents may be reproduced, copied, modified or adapted, without the prior written consent of the author, unless otherwise indicated for stand-alone materials.