

TEST_04

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

- 필요 패키지 : dplyr , ggplot2
 - 필요 데이터 : olympic.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



0. Packages

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

1. Data Wrangling Problem

1.1 olympic.csv 를 불러오고 속성을 확인하시오.[10점]

- head()로 출력 한 뒤 str()을 확인할 것

##	medal_type	medal_code	medal_date	athlete_short_name
## 1	Gold Medal	1	2021-07-24 00:00:00.0	KIM JD
## 2	Gold Medal	1	2021-07-24 00:00:00.0	AN S
## 3	Silver Medal	2	2021-07-24 00:00:00.0	SCHLOESSER G
## 4	Silver Medal	2	2021-07-24 00:00:00.0	WIJLER S
## 5	Bronze Medal	3	2021-07-24 00:00:00.0	ALVAREZ L
## 6	Bronze Medal	3	2021-07-24 00:00:00.0	VALENCIA A

##	athlete_name	athlete_sex
## 1	KIM Je Deok	X
## 2	AN San	X
## 3	SCHLOESSER Gabriela	X
## 4	WIJLER Steve	X
## 5	ALVAREZ Luis	X
## 6	VALENCIA Alejandra	X

##	athlete_link
## 1	.../.../.../en/results/archery/athlete-profile-n1440267-kim-j e-deok.htm
## 2	.../.../.../en/results/archery/athlete-profile-n1338248- an-san.htm
## 3	.../.../.../en/results/archery/athlete-profile-n1458456-schloesser-ga briela.htm
## 4	.../.../.../en/results/archery/athlete-profile-n1458465-wijler -steve.htm
## 5	.../.../.../en/results/archery/athlete-profile-n1294470-alvare z-luis.htm
## 6	.../.../.../en/results/archery/athlete-profile-n1294484-valencia-ale jandra.htm

##	country_code	discipline_code	event	country	discipli ne
## 1	KOR	ARC Mixed Team	Republic of Korea	Arche ry	
## 2	KOR	ARC Mixed Team	Republic of Korea	Arche ry	
## 3	NED	ARC Mixed Team	Netherlands	Arche ry	
## 4	NED	ARC Mixed Team	Netherlands	Arche ry	
## 5	MEX	ARC Mixed Team	Mexico	Arche ry	
## 6	MEX	ARC Mixed Team	Mexico	Arche ry	

```
## 'data.frame':    2401 obs. of  12 variables:
##  $ medal_type      : chr  "Gold Medal" "Gold Medal" "Silver Medal"
##  "Silver Medal" ...
##  $ medal_code       : int   1 1 2 2 3 3 1 2 3 1 ...
##  $ medal_date       : chr   "2021-07-24 00:00:00.0" "2021-07-24 00:0
##  0:00.0" "2021-07-24 00:00:00.0" "2021-07-24 00:00:00.0" ...
##  $ athlete_short_name: chr   "KIM JD" "AN S" "SCHLOESSER G" "WIJLER S
##  " ...
##  $ athlete_name     : chr   "KIM Je Deok" "AN San" "SCHLOESSER Gabri
##  ela" "WIJLER Steve" ...
##  $ athlete_sex      : chr   "X" "X" "X" "X" ...
##  $ athlete_link      : chr   "../.../en/results/archery/athlete-pro
##  file-n1440267-kim-je-deok.htm" "../.../en/results/archery/athlete-pro
##  file-n1338248-an-san.htm" "../.../en/results/archery/athlete-profile-
##  n1458456-schloesser-gabriela.htm" "../.../en/results/archery/athlete-
##  profile-n1458465-wijler-steve.htm" ...
##  $ country_code      : chr   "KOR" "KOR" "NED" "NED" ...
##  $ discipline_code   : chr   "ARC" "ARC" "ARC" "ARC" ...
##  $ event             : chr   "Mixed Team" "Mixed Team" "Mixed Team" "
##  Mixed Team" ...
##  $ country           : chr   "Republic of Korea" "Republic of Korea"
##  "Netherlands" "Netherlands" ...
##  $ discipline        : chr   "Archery" "Archery" "Archery" "Archery"
##  ...
```

1.2 다음과 같이 데이터를 생성한 뒤 medals이라는 이름으로 저장하시오[10점]

- medal_type에서 메달 이름만 가져오시오.[5점]
 - 힌트: separate() 활용
- 필요한 열만 선택하여 head()로 출력할 것. [5점]

```
##      medal      athlete_name athlete_sex country_code discipline
## 1   Gold      KIM Je Deok           X           KOR      Archery
## 2   Gold      AN San                 X           KOR      Archery
## 3 Silver SCHLOESSER Gabriela         X           NED      Archery
## 4 Silver      WIJLER Steve           X           NED      Archery
## 5 Bronze      ALVAREZ Luis            X           MEX      Archery
## 6 Bronze VALENCIA Alejandra          X           MEX      Archery
```

1.3 대한민국의 종목별 메달 개수를 구하시오 [15점]

- 종목별, 메달 별 통계량 [5점]
- 메달과 count는 정렬할 것[5점]
 - 메달 순서는 Gold, Silver, Bronze가 되도록 하시오[5점]
 - 힌트: factor()를 활용

```
## # A tibble: 13 x 3
## # Groups:   discipline [8]
##   discipline      medal count
##   <chr>          <fct> <int>
## 1 Archery        Gold      9
## 2 Fencing         Gold      4
## 3 Artistic Gymnastics Gold      1
## 4 Fencing         Silver     4
## 5 Judo            Silver     1
## 6 Shooting        Silver     1
## 7 Taekwondo        Silver     1
## 8 Fencing         Bronze     9
## 9 Badminton        Bronze     2
## 10 Judo            Bronze     2
## 11 Taekwondo        Bronze     2
## 12 Artistic Gymnastics Bronze     1
## 13 Modern Pentathlon Bronze     1
```

1.4 국가별 메달 획득 개수를 구하시오[20점]

- 금메달이 가장 많은 국가 순으로 정렬 할 것 [10점]
 - 그 안에서는 Gold, Silver, Bronze 순으로 정렬하시오(결과물과 같이 출력) [5점]
- head()로 다음과 같이 일부만 출력할 것[5점]

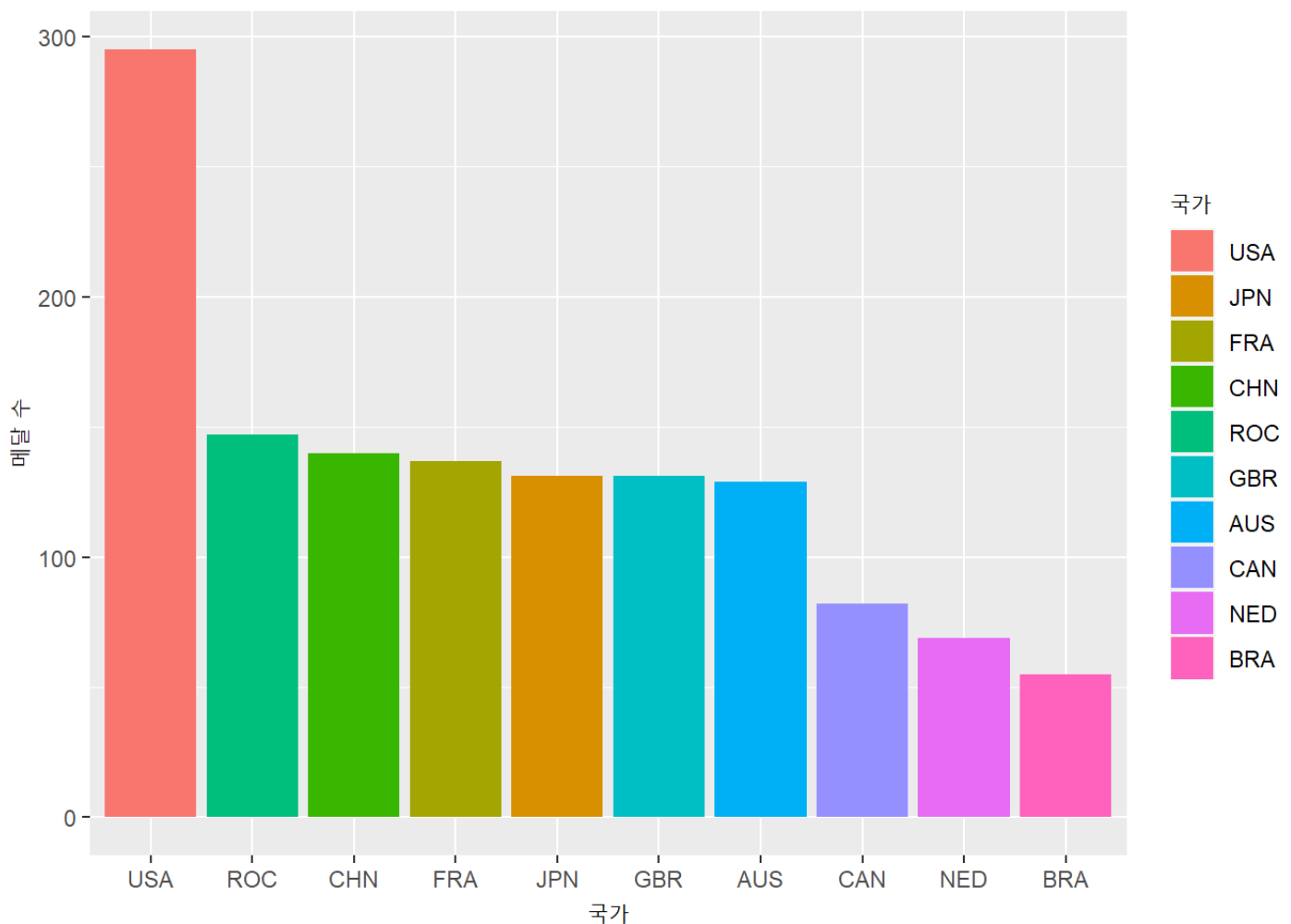
```
## # A tibble: 15 x 3
## # Groups:   country_code [5]
##   country_code medal  count
##   <fct>         <fct> <int>
## 1 USA           Gold    112
## 2 USA           Silver   108
## 3 USA           Bronze    75
## 4 JPN           Gold     68
## 5 JPN           Silver    41
## 6 JPN           Bronze    22
## 7 FRA           Gold     65
## 8 FRA           Silver    41
## 9 FRA           Bronze    31
## 10 CHN          Gold     58
## 11 CHN          Silver    49
## 12 CHN          Bronze    33
## 13 ROC          Gold     40
## 14 ROC          Silver    80
## 15 ROC          Bronze    27
```

2. Visualization Problem(Using ggplot2)

2.1 국가별 전체 메달 개수를 barplot을 활용해 그리시오 [20점]

- 전체 메달 개수 기준 상위 10개국을 대상으로 그래프를 그리시오[5점]
- 전체 메달 수가 높은 순으로 다음과 같이 그리시오[10점]
- x축 제목, y축 제목, legend 제목을 변경하시오 [5점]

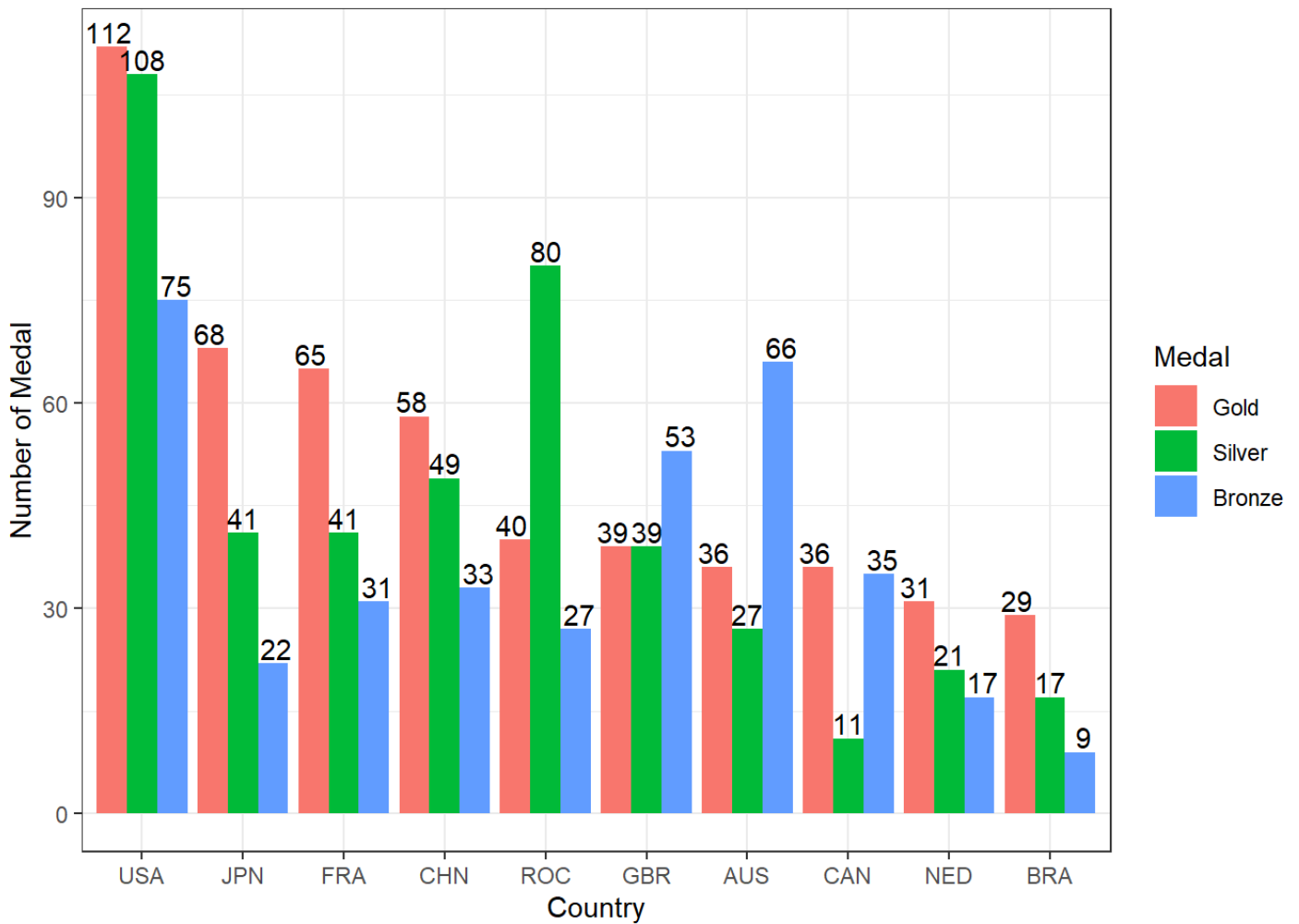
```
## # A tibble: 10 x 2
##   country_code count
##   <fct>          <int>
## 1 USA            295
## 2 JPN            131
## 3 FRA            137
## 4 CHN            140
## 5 ROC            147
## 6 GBR            131
## 7 AUS            129
## 8 CAN             82
## 9 NED             69
## 10 BRA           55
```



2.2 국가별 메달 개수를 barplot을 활용해 그리시오 [25점]

- 전체 메달 개수 기준이 아닌 금메달 개수 기준 상위 10개국을 대상으로 그래프를 그리시오 [5점]
- x축 제목, y축 제목, legend 제목을 변경하시오 [5점]
- 메달 개수를 barplot위에 나타내시오[15점]

◦ 힌트: geom_text를 활용할 것



© 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.