PR4

PR4\_201821479\_황혜린

2018년 10월 4일

# Dataframe

## 1.벡터를 이용해 데이터프레임 만들기

name <- c("Boil","Tom","Ravindra","Bob","Sobia")  
gender <- c("M","M","F","M","F")  
age <- c(17,21,33,12,37)  
marriage <- c(F,T,F,F,T)  
  
customer <- data.frame(name,gender,age,marriage)  
str(customer)

## 'data.frame': 5 obs. of 4 variables:  
## $ name : Factor w/ 5 levels "Bob","Boil","Ravindra",..: 2 5 3 1 4  
## $ gender : Factor w/ 2 levels "F","M": 2 2 1 2 1  
## $ age : num 17 21 33 12 37  
## $ marriage: logi FALSE TRUE FALSE FALSE TRUE

customer <- data.frame(name,gender,age,marriage, stringsAsFactors = F)  
str(customer)

## 'data.frame': 5 obs. of 4 variables:  
## $ name : chr "Boil" "Tom" "Ravindra" "Bob" ...  
## $ gender : chr "M" "M" "F" "M" ...  
## $ age : num 17 21 33 12 37  
## $ marriage: logi FALSE TRUE FALSE FALSE TRUE

str(customer)

## 'data.frame': 5 obs. of 4 variables:  
## $ name : chr "Boil" "Tom" "Ravindra" "Bob" ...  
## $ gender : chr "M" "M" "F" "M" ...  
## $ age : num 17 21 33 12 37  
## $ marriage: logi FALSE TRUE FALSE FALSE TRUE

names(customer)

## [1] "name" "gender" "age" "marriage"

rownames(customer)

## [1] "1" "2" "3" "4" "5"

## 2. Data Frame 변수명 바꾸기

colnames(customer)

## [1] "name" "gender" "age" "marriage"

rownames(customer)

## [1] "1" "2" "3" "4" "5"

colnames(customer) <- c("cust\_name","cust\_gend","cust\_age","cust\_mrg")  
rownames(customer) <- c('a', 'b', 'c', 'd', 'e')  
customer

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE  
## d Bob M 12 FALSE  
## e Sobia F 37 TRUE

## 3. Data Frame 데이터 추출

customer[1,] ; customer["a",]

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE

customer[customer$cust\_name=="Tom",]

## cust\_name cust\_gend cust\_age cust\_mrg  
## b Tom M 21 TRUE

customer[2:5,] ; customer[-1,]

## cust\_name cust\_gend cust\_age cust\_mrg  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE  
## d Bob M 12 FALSE  
## e Sobia F 37 TRUE

## cust\_name cust\_gend cust\_age cust\_mrg  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE  
## d Bob M 12 FALSE  
## e Sobia F 37 TRUE

customer[customer$cust\_name!="Tom",]

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE  
## c Ravindra F 33 FALSE  
## d Bob M 12 FALSE  
## e Sobia F 37 TRUE

customer[c("b","c"),]

## cust\_name cust\_gend cust\_age cust\_mrg  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE

## 4. Data Frame에 데이터 추가

customer$cust\_height <- c("185","165","156","174","155")  
customer["f",] <- list("Jack","M",50,T,"167")  
customer

## cust\_name cust\_gend cust\_age cust\_mrg cust\_height  
## a Boil M 17 FALSE 185  
## b Tom M 21 TRUE 165  
## c Ravindra F 33 FALSE 156  
## d Bob M 12 FALSE 174  
## e Sobia F 37 TRUE 155  
## f Jack M 50 TRUE 167

customer <- cbind(customer,weight=c(80,70,65,48,55,100))  
customer <- rbind(customer,g=list("Merry","F",42,F,"172",60))  
customer<- rbind(customer,h=c("Merry","F",42,F,"172",60))  
customer

## cust\_name cust\_gend cust\_age cust\_mrg cust\_height weight  
## a Boil M 17 FALSE 185 80  
## b Tom M 21 TRUE 165 70  
## c Ravindra F 33 FALSE 156 65  
## d Bob M 12 FALSE 174 48  
## e Sobia F 37 TRUE 155 55  
## f Jack M 50 TRUE 167 100  
## g Merry F 42 FALSE 172 60  
## h Merry F 42 FALSE 172 60

## 5. Data Frame 에 데이터 삭제

customer<-customer[,-5]  
customer<-customer[-7,]  
customer$weight<-NULL

## 6. Data 조건문을 활용해 조작하기

customer[customer$cust\_gend=="M",] #cust\_gend칼럼이 M인row만 출력

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE  
## d Bob M 12 FALSE  
## f Jack M 50 TRUE

customer[customer$cust\_gend!="F",] #cust\_gend칼럼이 F가 아닌 row

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE  
## d Bob M 12 FALSE  
## f Jack M 50 TRUE

nrow(customer[customer$cust\_gend=="M",]) #행의 개수

## [1] 4

customer[customer$cust\_name=="Bob", c("cust\_age","cust\_mrg")] #cust\_name컬럼 중에서 Bob의 cust\_age,cust\_mrg만 출력

## cust\_age cust\_mrg  
## d 12 FALSE

customer[customer$cust\_name=="Tom" | customer$cust\_name=="Ravindra",] #customer컬럼중에서 Tom과 Ravindra만 출력

## cust\_name cust\_gend cust\_age cust\_mrg  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE

customer[customer$cust\_gend=="M" & customer$cust\_age>24,] #cust\_gend컬럼 중에서 M을 나타내고, cust\_age컬럼 중에서 24세 초과인 row출력

## cust\_name cust\_gend cust\_age cust\_mrg  
## f Jack M 50 TRUE

## 7. Data Frame 정렬하기

order(customer$cust\_age)

## [1] 4 1 2 3 5 7 6

customer[order(customer$cust\_age),]

## cust\_name cust\_gend cust\_age cust\_mrg  
## d Bob M 12 FALSE  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE  
## e Sobia F 37 TRUE  
## h Merry F 42 FALSE  
## f Jack M 50 TRUE

order(customer$cust\_age, decreasing=F)

## [1] 4 1 2 3 5 7 6

customer[order(customer$cust\_age, decreasing=F),]

## cust\_name cust\_gend cust\_age cust\_mrg  
## d Bob M 12 FALSE  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE  
## e Sobia F 37 TRUE  
## h Merry F 42 FALSE  
## f Jack M 50 TRUE

## 8. Data frame 기타 함수

head(customer)

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE  
## c Ravindra F 33 FALSE  
## d Bob M 12 FALSE  
## e Sobia F 37 TRUE  
## f Jack M 50 TRUE

head(customer, 2)

## cust\_name cust\_gend cust\_age cust\_mrg  
## a Boil M 17 FALSE  
## b Tom M 21 TRUE

tail(customer, 2)

## cust\_name cust\_gend cust\_age cust\_mrg  
## f Jack M 50 TRUE  
## h Merry F 42 FALSE

# 파일 입출력

## 1. 내장데이터 불러오기

library(MASS)  
head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

str(iris)

## 'data.frame': 150 obs. of 5 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

str(mtcars)

## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : num 0 0 1 1 0 1 0 1 1 1 ...  
## $ am : num 1 1 1 0 0 0 0 0 0 0 ...  
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...  
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...

head(USArrests)

## Murder Assault UrbanPop Rape  
## Alabama 13.2 236 58 21.2  
## Alaska 10.0 263 48 44.5  
## Arizona 8.1 294 80 31.0  
## Arkansas 8.8 190 50 19.5  
## California 9.0 276 91 40.6  
## Colorado 7.9 204 78 38.7

str(USArrests)

## 'data.frame': 50 obs. of 4 variables:  
## $ Murder : num 13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...  
## $ Assault : int 236 263 294 190 276 204 110 238 335 211 ...  
## $ UrbanPop: int 58 48 80 50 91 78 77 72 80 60 ...  
## $ Rape : num 21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...

## 2. file로 저장된 데이터 불러오기

setwd("C:/Users/hyere/Desktop/PR")  
csv<-read.csv("read\_csv.csv") ; csv

## X..연습.테이블.입니다. X X.1 X.2  
## 1 1 Daredevil Hawkeye Loki  
## 2 2 Deadpool Hulk Luke Cage  
## 3 3 Doctor Strange Human Torch .  
## 4 6 Invisible Woman Ms. Marvel  
## 5 5 Iron Man Nightcrawler  
## 6 7 Ghost Rider Jean Grey Psylocke  
## X.3 X.4  
## 1 Punisher Storm  
## 2 Rocket Raccoon Taskmaster  
## 3 Scarlet Witch Thing  
## 4 Silver Surfer Thor  
## 5 N.A. Wolverine  
## 6 Squirrel Girl Barricade

str(csv)

## 'data.frame': 6 obs. of 6 variables:  
## $ X..연습.테이블.입니다.: int 1 2 3 6 5 7  
## $ X : Factor w/ 5 levels "","Daredevil",..: 2 3 4 1 1 5  
## $ X.1 : Factor w/ 6 levels "Hawkeye","Hulk",..: 1 2 3 4 5 6  
## $ X.2 : Factor w/ 6 levels ".","Loki","Luke Cage",..: 2 3 1 4 5 6  
## $ X.3 : Factor w/ 6 levels "N.A.","Punisher",..: 2 3 4 5 1 6  
## $ X.4 : Factor w/ 6 levels "Barricade","Storm",..: 2 3 4 5 6 1

getwd()

## [1] "C:/Users/hyere/Desktop/PR"

csv2 = read.csv("read\_csv.csv", header=F, stringsAsFactors=F) ; csv2

## V1 V2 V3 V4  
## 1 # 연습 테이블 입니다.   
## 2 1 Daredevil Hawkeye Loki  
## 3 2 Deadpool Hulk Luke Cage  
## 4 3 Doctor Strange Human Torch .  
## 5 6 Invisible Woman Ms. Marvel  
## 6 5 Iron Man Nightcrawler  
## 7 7 Ghost Rider Jean Grey Psylocke  
## V5 V6  
## 1   
## 2 Punisher Storm  
## 3 Rocket Raccoon Taskmaster  
## 4 Scarlet Witch Thing  
## 5 Silver Surfer Thor  
## 6 N.A. Wolverine  
## 7 Squirrel Girl Barricade

str(csv2)

## 'data.frame': 7 obs. of 6 variables:  
## $ V1: chr "# 연습 테이블 입니다." "1" "2" "3" ...  
## $ V2: chr "" "Daredevil" "Deadpool" "Doctor Strange" ...  
## $ V3: chr "" "Hawkeye" "Hulk" "Human Torch" ...  
## $ V4: chr "" "Loki" "Luke Cage" "." ...  
## $ V5: chr "" "Punisher" "Rocket Raccoon" "Scarlet Witch" ...  
## $ V6: chr "" "Storm" "Taskmaster" "Thing" ...

csv3 = read.csv("csv\_NA.csv", header=F, stringsAsFactors=F, na.strings=c(".","N.A.","")) ; csv3

## V1  
## 1 癤\xbf#\xec뿰\xec뒿 \xed뀒\xec씠釉\x94 \xec엯\xeb땲\xeb떎.  
## 2 1  
## 3 2  
## 4 3  
## 5 6  
## 6 5  
## 7 7  
## V2 V3 V4 V5 V6  
## 1 <NA> <NA> <NA> <NA> <NA>  
## 2 Daredevil Hawkeye Loki Punisher Storm  
## 3 Deadpool Hulk Luke Cage Rocket Raccoon Taskmaster  
## 4 Doctor Strange Human Torch <NA> Scarlet Witch Thing  
## 5 <NA> Invisible Woman Ms. Marvel Silver Surfer Thor  
## 6 <NA> Iron Man Nightcrawler <NA> Wolverine  
## 7 Ghost Rider Jean Grey Psylocke Squirrel Girl Barricade

str(csv3)

## 'data.frame': 7 obs. of 6 variables:  
## $ V1: chr "癤\xbf#\xec뿰\xec뒿 \xed뀒\xec씠釉\x94 \xec엯\xeb땲\xeb떎." "1" "2" "3" ...  
## $ V2: chr NA "Daredevil" "Deadpool" "Doctor Strange" ...  
## $ V3: chr NA "Hawkeye" "Hulk" "Human Torch" ...  
## $ V4: chr NA "Loki" "Luke Cage" NA ...  
## $ V5: chr NA "Punisher" "Rocket Raccoon" "Scarlet Witch" ...  
## $ V6: chr NA "Storm" "Taskmaster" "Thing" ...

csv4 = read.csv("csv\_NA.csv", header=F, stringsAsFactors=F, encoding="UTF-8") ; csv4

## V1 V2 V3 V4  
## 1 <U+FEFF>#연습 테이블 입니다.   
## 2 1 Daredevil Hawkeye Loki  
## 3 2 Deadpool Hulk Luke Cage  
## 4 3 Doctor Strange Human Torch .  
## 5 6 Invisible Woman Ms. Marvel  
## 6 5 Iron Man Nightcrawler  
## 7 7 Ghost Rider Jean Grey Psylocke  
## V5 V6  
## 1   
## 2 Punisher Storm  
## 3 Rocket Raccoon Taskmaster  
## 4 Scarlet Witch Thing  
## 5 Silver Surfer Thor  
## 6 N.A. Wolverine  
## 7 Squirrel Girl Barricade

str(csv4)

## 'data.frame': 7 obs. of 6 variables:  
## $ V1: chr "<U+FEFF>#연습 테이블 입니다." "1" "2" "3" ...  
## $ V2: chr "" "Daredevil" "Deadpool" "Doctor Strange" ...  
## $ V3: chr "" "Hawkeye" "Hulk" "Human Torch" ...  
## $ V4: chr "" "Loki" "Luke Cage" "." ...  
## $ V5: chr "" "Punisher" "Rocket Raccoon" "Scarlet Witch" ...  
## $ V6: chr "" "Storm" "Taskmaster" "Thing" ...

table = read.table("read\_csv.csv", header=F,sep=",", stringsAsFactors=F)  
table

## V1 V2 V3 V4 V5 V6  
## 1 1 Daredevil Hawkeye Loki Punisher Storm  
## 2 2 Deadpool Hulk Luke Cage Rocket Raccoon Taskmaster  
## 3 3 Doctor Strange Human Torch . Scarlet Witch Thing  
## 4 6 Invisible Woman Ms. Marvel Silver Surfer Thor  
## 5 5 Iron Man Nightcrawler N.A. Wolverine  
## 6 7 Ghost Rider Jean Grey Psylocke Squirrel Girl Barricade

## 3. 웹에서 데이터 읽어오기

store.df <- read.csv("http://goo.gl/QPDdMl")  
head(store.df,2)

## storeNum Year Week p1sales p2sales p1price p2price p1prom p2prom country  
## 1 101 1 1 127 106 2.29 2.29 0 0 US  
## 2 101 1 2 137 105 2.49 2.49 0 0 US

library(XML)  
world\_pop = readHTMLTable("http://www.worldometers.info/world-population/")  
head(world\_pop[[1]])

## Year (July 1) Population Yearly % Change Yearly Change Median Age  
## 1 2018 7,632,819,325 1.09 % 82,557,224 29.9  
## 2 2017 7,550,262,101 1.12 % 83,297,821 29.9  
## 3 2016 7,466,964,280 1.14 % 83,955,460 29.9  
## 4 2015 7,383,008,820 1.19 % 84,967,932 30  
## 5 2010 6,958,169,159 1.24 % 83,201,955 29  
## 6 2005 6,542,159,383 1.26 % 79,430,479 27  
## Fertility Rate Density (P/Km짼) Urban Pop % Urban Population  
## 1 2.51 51 54.9 % 4,186,975,665  
## 2 2.51 51 54.4 % 4,110,778,369  
## 3 2.51 50 54.0 % 4,034,193,153  
## 4 2.52 50 53.6 % 3,957,285,013  
## 5 2.57 47 51.3 % 3,571,272,167  
## 6 2.63 44 48.9 % 3,199,013,076

## 4. 데이터 저장하기

table

## V1 V2 V3 V4 V5 V6  
## 1 1 Daredevil Hawkeye Loki Punisher Storm  
## 2 2 Deadpool Hulk Luke Cage Rocket Raccoon Taskmaster  
## 3 3 Doctor Strange Human Torch . Scarlet Witch Thing  
## 4 6 Invisible Woman Ms. Marvel Silver Surfer Thor  
## 5 5 Iron Man Nightcrawler N.A. Wolverine  
## 6 7 Ghost Rider Jean Grey Psylocke Squirrel Girl Barricade

write.table(table,"PR\_table.csv")  
write.table(table,"PR\_table1.csv",row.names=F)  
write.csv(table,"PR\_table2.csv",row.names=F)

# 연습문제

# 문제1

setwd("C:/Users/hyere/Desktop/PR")  
pubg <- read.csv("PUBG\_Player\_Statistics.csv") ; csv

## X..연습.테이블.입니다. X X.1 X.2  
## 1 1 Daredevil Hawkeye Loki  
## 2 2 Deadpool Hulk Luke Cage  
## 3 3 Doctor Strange Human Torch .  
## 4 6 Invisible Woman Ms. Marvel  
## 5 5 Iron Man Nightcrawler  
## 6 7 Ghost Rider Jean Grey Psylocke  
## X.3 X.4  
## 1 Punisher Storm  
## 2 Rocket Raccoon Taskmaster  
## 3 Scarlet Witch Thing  
## 4 Silver Surfer Thor  
## 5 N.A. Wolverine  
## 6 Squirrel Girl Barricade

str(pubg) #문제 1.1

## 'data.frame': 499 obs. of 7 variables:  
## $ player\_name : Factor w/ 497 levels "0II0ll0","0tan",..: 66 57 292 134 312 100 173 254 451 162 ...  
## $ solo\_RoundsPlayed : int 17 33 5 8 6 16 142 19 44 8 ...  
## $ solo\_AvgDamage : num 255 393 330 752 637 ...  
## $ solo\_AvgKills : num 2.59 3.61 3.6 7 7 3.94 2.24 2 3.18 3.12 ...  
## $ solo\_AvgSurvivalTime: Factor w/ 400 levels "1,001","1,005",..: 135 112 68 260 255 179 119 178 290 133 ...  
## $ solo\_MoveDistance : Factor w/ 497 levels " - ","1,000,122",..: 384 144 105 360 326 476 412 486 143 289 ...  
## $ play\_region : int 2 3 2 1 1 4 1 1 4 3 ...

#문제 1.2) 엑셀파일 solo\_AvgSurvivalTime, solo\_MoveDistance의 숫자들에 ,(콤마)와 빈 값(-)있어 str()을 실행 할 때 숫자가 아닌 문자로 인식하여 제대로 출력되지 않습니다.  
  
  
#문제1.3) 첫번째로는 엑셀파일 solo\_AvgSurvivalTime, solo\_MoveDistance의 숫자들에 콤마를 지워야 합니다. 또한, solo\_MoveDistance의 경우에는 빈 값 까지 제거해주어야합니다.  
#두번째로는 컬럼player\_name은 factor->charactor로, 컬럼 solo\_AvgSurvivalTime, solo\_MoveDistance 는 factor->int으로, play\_region는 int->factor로 전환해주어야 합니다.  
  
  
pubg1 <- read.csv("PUBG\_Player\_Statistics\_fix.csv")  
#수정한 엑셀파일을 pubg1에 지정  
  
pubg1$player\_name <- as.character(pubg1$player\_name)  
pubg1$solo\_AvgSurvivalTime <- as.integer(pubg1$solo\_AvgSurvivalTime)  
pubg1$solo\_MoveDistance <- as.integer(pubg1$solo\_MoveDistance)  
pubg1$play\_region <- as.factor(pubg1$play\_region)  
#자료형을 변환해 주는 과정  
str(pubg1)

## 'data.frame': 499 obs. of 7 variables:  
## $ player\_name : chr "BreakNeck" "Blackwalk" "mercedes\_benz" "DORA" ...  
## $ solo\_RoundsPlayed : int 17 33 5 8 6 16 142 19 44 8 ...  
## $ solo\_AvgDamage : num 255 393 330 752 637 ...  
## $ solo\_AvgKills : num 2.59 3.61 3.6 7 7 3.94 2.24 2 3.18 3.12 ...  
## $ solo\_AvgSurvivalTime: int 1263 1221 1133 1678 1640 1375 1239 1373 652 1260 ...  
## $ solo\_MoveDistance : int 56462 165707 14024 47895 39162 91395 670638 95534 165491 31283 ...  
## $ play\_region : Factor w/ 5 levels "1","2","3","4",..: 2 3 2 1 1 4 1 1 4 3 ...

# 문제2

#play\_region의 각각 값에 대응하는 country를 입력합니다.  
pubg1[pubg1$play\_region == "1", "Country"] = "KOREA"  
pubg1[pubg1$play\_region == "2", "Country"] = "CHINA"  
pubg1[pubg1$play\_region == "3", "Country"] = "JAPAN"  
pubg1[pubg1$play\_region == "4", "Country"] = "USA"  
pubg1[pubg1$play\_region == "5", "Country"] = "OTHERS"  
  
head(pubg1, 15) #확인

## player\_name solo\_RoundsPlayed solo\_AvgDamage solo\_AvgKills  
## 1 BreakNeck 17 255.36 2.59  
## 2 Blackwalk 33 393.04 3.61  
## 3 mercedes\_benz 5 329.76 3.60  
## 4 DORA 8 751.95 7.00  
## 5 n2tstar 6 637.48 7.00  
## 6 coldoxygen 16 411.95 3.94  
## 7 Giken 142 284.57 2.24  
## 8 KoreaNo1 19 238.40 2.00  
## 9 undor 44 339.44 3.18  
## 10 Fordune 8 372.38 3.12  
## 11 PandaTV-Tongk 116 325.38 2.93  
## 12 Benny-\_- 335 144.50 1.18  
## 13 PanDaTV\_Karl 87 436.99 4.00  
## 14 homeboye 72 290.98 2.46  
## 15 YUPPIEE 23 327.68 3.13  
## solo\_AvgSurvivalTime solo\_MoveDistance play\_region Country  
## 1 1263 56462 2 CHINA  
## 2 1221 165707 3 JAPAN  
## 3 1133 14024 2 CHINA  
## 4 1678 47895 1 KOREA  
## 5 1640 39162 1 KOREA  
## 6 1375 91395 4 USA  
## 7 1239 670638 1 KOREA  
## 8 1373 95534 1 KOREA  
## 9 652 165491 4 USA  
## 10 1260 31283 3 JAPAN  
## 11 940 333831 1 KOREA  
## 12 839 692141 2 CHINA  
## 13 1022 465510 1 KOREA  
## 14 777 247990 5 OTHERS  
## 15 1370 111077 1 KOREA

# 문제3

Avgdmg <- tapply(pubg1$solo\_AvgDamage, pubg1$Country, mean)  
#Avgdmg 라는 변수에 국가별 평균 데미지를 할당합니다  
  
Avgkills <- tapply(pubg1$solo\_AvgKills, pubg1$Country, mean)  
#Avgkills라는 변수에 국가별 평균 킬수를 할당합니다  
  
ranking <- Avgkills \* 0.4 + Avgdmg \* 0.6  
#Avgdmg, Avgkills에 각각 가중치를 두어 ranking이란 변수에 할당합니다  
  
sort(ranking, decreasing=T)

## KOREA USA OTHERS CHINA JAPAN   
## 206.2068 184.3741 177.6293 173.9978 173.6385

#ranking이 높은 순서대로 출력합니다