

HW8

Chrinovic Mukulu

2025-11-03

```
#Question 1
library(readr)
ss06hid <- read_csv("ss06hid.csv")

## Rows: 6496 Columns: 188
## -- Column specification -----
## Delimiter: ","
## chr (1): RT
## dbl (187): SERIALNO, DIVISION, PUMA, REGION, ST, ADJUST, WGTP, NP, TYPE, ACR...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

View(ss06hid)
fileUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
download.file(fileUrl, destfile = "ss06hid.csv",)

## Warning in download.file(fileUrl, destfile = "ss06hid.csv", ): URL
## https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv: cannot open
## destfile 'ss06hid.csv', reason 'Permission denied'

## Warning in download.file(fileUrl, destfile = "ss06hid.csv", ): download had
## nonzero exit status

Housing<- read.csv("ss06hid.csv")

agricultureLogical<- Housing$ACR == 3 & Housing$AGS == 6

which(agricultureLogical)[1:3]

## [1] 125 238 262

#Question 2
library(jpeg)

## Warning: package 'jpeg' was built under R version 4.5.2
```

```

url<- "https://drive.google.com/uc?export=download&id=1OKOr-BZuoL0bVcce8tlM7_GvQyECpjPp"
download.file(url, destfile = "hw8pic.jpg", mode = "wb")
img<- readJPEG("hw8pic.jpg", native = TRUE)

quantile(img, probs = c(0.3,0.8))

##      30%      80%
## -16776945 -6383497

#Question 3

gdp <- read.csv("getdata_data_GDP.csv", skip = 4, stringsAsFactors = FALSE)
edu <- read.csv("getdata_data_EDSTATS_Country.csv", stringsAsFactors = FALSE)

gdp <- gdp[!is.na(gdp$X), ]
names(gdp)[1:5] <- c("CountryCode", "Rank", "Economy", "GDP", "Extra")
merged <- merge(gdp[, c("CountryCode", "Rank", "Economy", "GDP")],
                 edu,
                 by = "CountryCode")

merged$Rank <- as.numeric(merged$Rank)

tapply(merged$Rank, merged$Income.Group, mean, na.rm = TRUE) [
  c("High income: OECD", "High income: nonOECD")
]

##      High income: OECD High income: nonOECD
##            32.96667          91.91304

merged$gdpGroup <- cut(merged$Rank,
                        breaks = quantile(merged$Rank, probs = seq(0,1,0.2), na.rm = TRUE),
                        include.lowest = TRUE)
table(merged$gdpGroup, merged$Income.Group)

## 
##      High income: nonOECD High income: OECD Low income
## [1,38.6]        0           4        18        0
## (38.6,76.2]     0           5        10        1
## (76.2,114]      0           8         1        9
## (114,152]       0           4         1       16
## (152,190]       0           2         0       11
## 
##      Lower middle income Upper middle income
## [1,38.6]           5           11
## (38.6,76.2]        13           9
## (76.2,114]         11           8
## (114,152]          9            8
## (152,190]         16           9

topGroup <- merged$gdpGroup == levels(merged$gdpGroup)[1]

sum(topGroup & merged$Income.Group == "Lower middle income")

```

```
## [1] NA
```