Week 3 - Quiz 1

Raphael Carvalho

28/07/2019

1. The American Community Survey distributes downloadable data about United States communities. Download the 2006 microdata survey about housing for the state of Idaho using download.file() from here: "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv" and load the data into R. The code book, describing the variable names is here: "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDataDict06.pdf". Create a logical vector that identifies the households on greater than 10 acres who sold more than $10,000 worth of agriculture products. Assign that logical vector to the variable agricultureLogical. Apply the which() function like this to identify the rows of the data frame where the logical vector is TRUE. -> which(agricultureLogical). What are the first 3 values that result?

```r
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
download.file(fileURL, "./w3_q1.csv")

q1 <- read.csv("./w3_q1.csv", stringsAsFactors = FALSE)

which(q1$ACR == 3 & q1$AGS == 6)[1:3]
```

```
## [1] 125 238 262
```

[x] 125, 238, 262

[ ] 236, 238, 262

[ ] 59, 460, 474

[ ] 153, 236, 388

2. Using the jpeg package read in the following picture of your instructor into R "https://d396qusza40orc.cloudfront.net/getdata%2Fjeff.jpg". Use the parameter native-TRUE. What are the 30th and 80th quantiles of the resulting data? (some Linux systems may produce an answer 648 different for the 30th quantile)

```r
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fjeff.jpg"
download.file(fileURL, "./w3_q2.jpg")
q2 <- readJPEG("./w3_q2.jpg", native=TRUE)
quantile(q2, probs = c(0.3, 0.8))
```

```
##       30%       80%
## -15259150 -10575416
```

[ ] 10904118 -594524

[ ] -10904118 -10575416

[ ] -16776430 -15390165

[x] -15259150 -10575416

3. Load the Gross Domestic Product data for the 190 ranked countries in this data set: "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv". Load the educational data from this data set: "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv". Match the data based on the country shortcode. How many of the IDs match? Sort the data frame in descending order by GDP rank (so United States is last). What is the 13th country in the resulting data frame?

```
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv"
download.file(fileURL, "./w3_q3_1.csv")

fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv"
download.file(fileURL, "./w3_q3_2.csv")

q3_1 <- fread("./w3_q3_1.csv", skip = 5, nrows = 190, select = c(1, 2, 4, 5), col.names = c("CountryCod
q3_2 <- fread("./w3_q3_2.csv", stringsAsFactors = FALSE)

dt <- merge(q3_1, q3_2, by='CountryCode') %>% arrange(desc(Rank))
paste(nrow(dt), " matches, 13th country is ", dt$Economy[13])
```

```
## [1] "189  matches, 13th country is  St. Kitts and Nevis"
```

[ ] 234 matches, 13th country is Spain

[ ] 190 matches, 13th country is Spain

[x] 234 matches, 13th country is St. Kitts and Nevis

[ ] 190 matches, 13th country is St. Kitts and Nevis

[ ] 189 matches, 13th country is Spain

[ ] 189 matches, 13th country is St. Kitts and Nevis

4. What is the average gdp RANKING FOR THE "High income: OECD" and "High income: nonOECD" group?

```
dt %>% group_by(`Income Group`) %>% filter("High income: OECD" %in% `Income Group` | "High income: nonO
```

```
## # A tibble: 2 x 2
##   `Income Group`          avg
##   <chr>                 <dbl>
## 1 High income: OECD      33.0
## 2 High income: nonOECD   91.9
```

[ ] 23, 30

[ ] 23.966667, 30.91304

[ ] 133.72973, 32.96667

[ ] 23, 45

[ ] 30, 37

[x] 32.96667, 91.91304

5. Cut the GDP ranking into 5 separate quantile groups. Make a table versus Income.Group. How many countries are lower middle income but among the 38 nations with highest GDP?

```
dt$RankQuantile <- cut(dt$Rank, breaks = 5)
table(dt$RankQuantile, dt$`Income Group`)
```

```
##
##                 High income: nonOECD High income: OECD Low income
##   (0.811,38.8]                     4                18          0
##   (38.8,76.6]                      5                10          1
##   (76.6,114]                       8                 1          9
##   (114,152]                        4                 1         16
##   (152,190]                        2                 0         11
##
##                 Lower middle income Upper middle income
##   (0.811,38.8]                    5                  11
##   (38.8,76.6]                    13                   9
##   (76.6,114]                     12                   8
##   (114,152]                       8                   8
##   (152,190]                      16                   9
```

are Lower middle income but among the 38 nations with highest GDP?

[ ] 3

[x] 5

[ ] 18

[ ] 0