Branch: master ▾ getdata / week3-quiz.R Find file Copy path

Pavel Rebrov week3 Quiz complete 88f82fb on Sep 26, 2014

0 contributors

86 lines (63 sloc) 3.9 KB Raw Blame History

```
### QUESTION 1. Create a logical vector that identifies the households on greater than 10 acres who sold more than $10,000 worth of agricul

fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
destFile <- tempfile()
download.file(fileURL, destFile, method = "curl")

data <- read.csv(destFile, header = T)

## ACR == 3 - lot size of 10 or more acres
## AGS == 6 - $10000+ of Sales of Agricultural sales

## Creating logical vector
agricultureLogical <- with(data, ACR == 3 & AGS ==6)
## first 3 values of which on that vector
which(agricultureLogical)[1:3]

### ANSWER 1. [1] 125 238 262

### QUESTION 2. Use the parameter native=TRUE to read img file with jpeg package. What are the 30th and 80th quantiles of the resulting dat
library(jpeg)

imgURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fjeff.jpg"
imgFile <- tempfile()
download.file(imgURL, imgFile, method = "curl")

## Read file
img <- readJPEG(imgFile, native = TRUE)
## Get quantiles at 30% and 80%
quantile(img, probs = c(0.3, 0.8))

### ANSWER 2. -15259150 -10575416

### QUESTION 3. Match the data based on the country shortcode. How many of the IDs match? Sort the data frame in descending order by GDP ra

gdpURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv"
eduURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv"
gdpFile <- tempfile()
eduFile <- tempfile()

download.file(gdpURL, gdpFile, method = "curl")
download.file(eduURL, eduFile, method = "curl")

## Read 190 ranked countries
gdpData <- read.csv(gdpFile, skip = 5, nrows = 190, stringsAsFactors = F, header = F)
eduData <- read.csv(eduFile, stringsAsFactors = F)

## Subset only needed data, name columns in gdpData and convert GDP Value to numeric
gdpData <- gdpData[, c(1, 2, 4, 5)]
colnames(gdpData) <- c("CountryCode", "Rank", "Country.Name", "GDP.Value")
gdpData$GDP.Value <- as.numeric(gsub(",", "", gdpData$GDP.Value))
```

```
51
52  ## Merge data by country codes
53  matchedData <- merge(gdpData, eduData, by.x = "CountryCode", by.y = "CountryCode")
54  ## Number of matched countries
55  dim(matchedData)[1]
56
57  ## Arrange by GDP rank (descending) and get the name of the 13th country
58  library(plyr)
59  arrange(matchedData, desc(Rank))[13, 3]
60
61  ### ANSWER 3. [1] "St. Kitts and Nevis"
62
63  ### QUESTION 4. What is the average GDP ranking for the "High income: OECD" and "High income: nonOECD" group?
64
65  ## Subset "High income: OECD" and calculate the mean GDP Rank
66  mean(subset(matchedData, Income.Group %in% "High income: OECD", select = c(Rank))$Rank)
67  ## [1] 32.96667
68
69  ## Subset "High income: nonOECD"
70  mean(subset(matchedData, Income.Group %in% "High income: nonOECD", select = c(Rank))$Rank)
71  ## [1] 91.91304
72
73  #### QUESTION 5. Cut the GDP ranking into 5 separate quantile groups. Make a table versus Income.Group. How many countries are Lower middle
74
75  library(Hmisc)
76  ## Cut Ranks into 5 groups and store as factor variable
77  matchedData$Rank.Groups = cut2(matchedData$Rank, g = 5)
78  ## Build a table of Income Groups across Rank Groups
79  table(matchedData$Income.Group, matchedData$Rank.Groups)
80
81  ##                        [  1, 39) [ 39, 77) [ 77,115) [115,154) [154,190]
82  ## High income: nonOECD         4         5         8         5         1
83  ## High income: OECD           18        10         1         1         0
84  ## Low income                   0         1         9        16        11
85  ## Lower middle income          *5*        13        12         8        16
86  ## Upper middle income         11         9         8         8         9
```