Week 4 - Quiz 1

Raphael Carvalho

02/08/2019

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". Apply strsplit() to split all the names of the data frame on the characters "wgtp". What is the value of the 123 element of the resulting list?

```
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
download.file(fileURL, "./w4_q1.csv")

q1 <- read.csv("./w4_q1.csv", sep = ",")
strsplit(colnames(q1), "wgtp")[123]
```

```
## [[1]]
## [1] ""     "15"
```

[x] "" "15"

[ ] "wgtp" "15"

[ ] "15"

[ ] "w" "15"

2. Load the Gross Domestic Product data for the 190 ranked countries in this data set: "https://d396qusza40orc.cloudfront.net/getdata%2Fdata% 2FGDP.csv". Remove the commas from the GDP numbers in millions of dollars and average them. What is the average? Original data sources: "http://data.worldbank.org/data-catalog/GDP-ranking-table".

```
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv"
download.file(fileURL, "./w4_q2.csv")

q2 <- read.csv("./w4_q2.csv", nrow = 190, skip = 4)[, c(1, 2, 4, 5)]
colnames(q2) <- c("countryCode", "Rank", "countryNames", "Total")
q2$Total <- as.integer(gsub(",", "", q2$Total))
mean(q2$Total, na.rm = TRUE)
```

```
## [1] 377652.4
```

[ ] 381668.9

[x] 377652.4

[ ] 379596.5

[ ] 387854.4

3. In the data set from Question 2 what is a regular expression that would allow you to count the number of countries whose name begins with "United"? Addume that the variable with the country names in it is named countryNames. How many countries begin with United?

In the data set from Question 2 what is a regular expression that would allow you to count the number of countries whose name begins with "United"? Assume that the variable with the country names in it is named countryNames. How many countries begin with United?

```
length(as.character(q2$countryNames[grep("United$", q2$countryNames)])) == 3
```

```
## [1] FALSE
```

```
length(as.character(q2$countryNames[grep("^United", q2$countryNames)])) == 3
```

```
## [1] TRUE
```

```
length(as.character(q2$countryNames[grep("^United", q2$countryNames)])) == 4
```

```
## [1] FALSE
```

```
length(as.character(q2$countryNames[grep("*United", q2$countryNames)])) == 2
```

```
## [1] FALSE
```

[ ] grep("United$",countryNames), 3

[ ] grep("^United",countryNames), 4

[x] grep("^United",countryNames), 3

[ ] grep("*United",countryNames), 2

4. 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. Of the countries for which the end of the fiscal year is available, how many end in June? Original data sources: "http://data.worldbank.org/data-catalog/GDP-ranking-table"; "http://data.worldbank.org/data-catalog/ed-stats"

```
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv"
download.file(fileURL, "./w4_q4_1.csv")
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv"
download.file(fileURL, "./w4_q4_2.csv")

q4_1 <- fread("./w4_q4_1.csv", skip = 5, nrows = 190, select = c(1, 2, 4, 5), col.names = c("CountryCod
q4_2 <- fread("./w4_q4_2.csv")

q4_merge <- merge(q4_1, q4_2, by = "CountryCode")
q4_merge
```

```
##     CountryCode Rank              Economy      Total
##  1:         ABW  161                Aruba      2,584
##  2:         AFG  105          Afghanistan     20,497
##  3:         AGO   60               Angola    114,147
##  4:         ALB  125              Albania     12,648
```

```
##   5:        ARE   32 United Arab Emirates  348,595
## ---
## 185:       YEM   90        Yemen, Rep.  35,646
## 186:       ZAF   28        South Africa 384,313
## 187:       ZAR  112   Congo, Dem. Rep.   17,204
## 188:       ZMB  104           Zambia  20,678
## 189:       ZWE  134         Zimbabwe   9,802
##                          Long Name       Income Group
##   1:                        Aruba High income: nonOECD
##   2:     Islamic State of Afghanistan          Low income
##   3:     People's Republic of Angola  Lower middle income
##   4:           Republic of Albania  Upper middle income
##   5:         United Arab Emirates High income: nonOECD
## ---
## 185:           Republic of Yemen  Lower middle income
## 186:       Republic of South Africa  Upper middle income
## 187: Democratic Republic of the Congo          Low income
## 188:           Republic of Zambia          Low income
## 189:         Republic of Zimbabwe          Low income
##                          Region Lending category Other groups
##   1:  Latin America & Caribbean
##   2:            South Asia        IDA       HIPC
##   3:        Sub-Saharan Africa        IDA
##   4:      Europe & Central Asia       IBRD
##   5: Middle East & North Africa
## ---
## 185: Middle East & North Africa        IDA
## 186:        Sub-Saharan Africa       IBRD
## 187:        Sub-Saharan Africa        IDA       HIPC
## 188:        Sub-Saharan Africa        IDA       HIPC
## 189:        Sub-Saharan Africa       Blend
##          Currency Unit Latest population census  Latest household survey
##   1:     Aruban florin            2000
##   2:     Afghan afghani            1979            MICS, 2003
##   3:     Angolan kwanza            1970 MICS, 2001, MIS, 2006/07
##   4:       Albanian lek            2001            MICS, 2005
##   5:      U.A.E. dirham            2005
## ---
## 185:       Yemeni rial            2004            MICS, 2006
## 186: South African rand            2001            DHS, 2003
## 187:     Congolese franc            1984            DHS 2007
## 188:     Zambian kwacha            2000            DHS, 2007
## 189:     Zimbabwe dollar            2002          DHS, 2005/06
##                                                            Special Notes
##   1:
##   2: Fiscal year end: March 20; reporting period for national accounts data: FY.
##   3:
##   4:
##   5:
## ---
## 185:
## 186: Fiscal year end: March 31; reporting period for national accounts data: CY.
## 187:
## 188:
```

```
## 189:  Fiscal year end: June 30; reporting period for national accounts data: CY.
##       National accounts base year National accounts reference year
##   1:                         1995                              NA
##   2:                    2002/2003                              NA
##   3:                         1997                              NA
##   4:                                                          1996
##   5:                         1995                              NA
##  ---
## 185:                         1990                              NA
## 186:                         2000                              NA
## 187:                         1987                              NA
## 188:                         1994                              NA
## 189:                         1990                              NA
##       System of National Accounts SNA price valuation
##   1:                          NA
##   2:                          NA                  VAB
##   3:                          NA                  VAP
##   4:                        1993                  VAB
##   5:                          NA                  VAB
##  ---
## 185:                          NA                  VAP
## 186:                        1993                  VAB
## 187:                        1993                  VAB
## 188:                          NA                  VAB
## 189:                          NA                  VAB
##       Alternative conversion factor PPP survey year
##   1:                                              NA
##   2:                                              NA
##   3:                       1991-96              2005
##   4:                                            2005
##   5:                                              NA
##  ---
## 185:                       1990-96              2005
## 186:                                            2005
## 187:                       1999-01              2005
## 188:                       1990-92              2005
## 189:                     1991, 1998            2005
##       Balance of Payments Manual in use External debt Reporting status
##   1:
##   2:                                                            Actual
##   3:                                  BPM5                      Actual
##   4:                                  BPM5                      Actual
##   5:                                  BPM4
##  ---
## 185:                                  BPM5                      Actual
## 186:                                  BPM5                 Preliminary
## 187:                                  BPM5                    Estimate
## 188:                                  BPM5                 Preliminary
## 189:                                  BPM5                      Actual
##       System of trade Government Accounting concept
##   1:         Special
##   2:         General                  Consolidated
##   3:         Special
##   4:         General                  Consolidated
```

4

```
##   5:         General               Consolidated
## ---
## 185:         General                  Budgetary
## 186:         General               Consolidated
## 187:         Special               Consolidated
## 188:         General                  Budgetary
## 189:         General               Consolidated
##      IMF data dissemination standard
##   1:
##   2:                           GDDS
##   3:                           GDDS
##   4:                           GDDS
##   5:                           GDDS
## ---
## 185:                           GDDS
## 186:                           SDDS
## 187:                           GDDS
## 188:                           GDDS
## 189:                           GDDS
##      Source of most recent Income and expenditure data
##   1:
##   2:
##   3:                                        IHS, 2000
##   4:                                       LSMS, 2005
##   5:
## ---
## 185:                                       ES/BS, 2005
## 186:                                       ES/BS, 2000
## 187:                                     1-2-3, 2005-06
## 188:                                       IHS, 2004-05
## 189:
##      Vital registration complete Latest agricultural census
##   1:
##   2:
##   3:                                                 1964-65
##   4:                        Yes                          1998
##   5:                                                     1998
## ---
## 185:                                                     2002
## 186:                                                     2000
## 187:                                                     1990
## 188:                                                     1990
## 189:                                                     1960
##      Latest industrial data Latest trade data Latest water withdrawal data
##   1:                     NA              2008                           NA
##   2:                     NA              2008                         2000
##   3:                     NA              1991                         2000
##   4:                   2005              2008                         2000
##   5:                     NA              2008                         2005
## ---
## 185:                   2005              2008                         2000
## 186:                   2005              2008                         2000
## 187:                     NA              1986                         2000
## 188:                     NA              2008                         2000
```

```
## 189:                      1995            2008                          2002
##       2-alpha code WB-2 code       Table Name          Short Name
##   1:          AW       AW              Aruba                Aruba
##   2:          AF       AF        Afghanistan          Afghanistan
##   3:          AO       AO             Angola               Angola
##   4:          AL       AL            Albania              Albania
##   5:          AE       AE United Arab Emirates United Arab Emirates
##  ---
## 185:          YE       RY        Yemen, Rep.                Yemen
## 186:          ZA       ZA       South Africa         South Africa
## 187:          CD       ZR   Congo, Dem. Rep.      Dem. Rep. Congo
## 188:          ZM       ZM             Zambia               Zambia
## 189:          ZW       ZW           Zimbabwe             Zimbabwe
```

```r
june <- grep("Fiscal year end: June", q4_merge$`Special Notes`)
NROW(june)
```

```
## [1] 13
```

[ ] 15

[x] 13

[ ] 16

[ ] 7

5. You can use the quantmod (http://www.quantmod.com/) package to get historical stock prices for publicly traded companies on the NASDAQ and NYSE. Use the following code to download data on Amazon's stock price and get the times the data was sampled. How many values were collected in 2012?

```r
library(quantmod)
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:data.table':
##
##     first, last
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
## Loading required package: TTR
```

```
## Version 0.4-0 included new data defaults. See ?getSymbols.
```

```
amzn = getSymbols("AMZN",auto.assign=FALSE)
```

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.
```

```
sampleTimes = index(amzn)
```

How many values were collected on Mondays in 2012?

```
amzn2012 <- sampleTimes[grep("^2012", sampleTimes)]
NROW(amzn2012)
```

```
## [1] 250
```

```
NROW(amzn2012[weekdays(amzn2012) == "Segunda Feira"])
```

```
## [1] 47
```

[ ] 250, 51

[ ] 252, 50

[x] 250, 47

[ ] 251, 47