Week 4 Quiz

Ken Wood

7/17/2020

```
# Question #1

# 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?

download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv","hid.csv")
hid <- read.csv("hid.csv")
colnames <- names(hid)
strsplit(colnames, "wgtp") [[123]]
```

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

```
# Question #2

# Load the Gross Domestic Product data for the 190 ranked countries.
# Remove the commas from the GDP numbers in millions of dollars and average them.
# What is the average?

download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv","FGDP.csv")
fgdp <- read.csv("FGDP.csv",skip=4,nrows=190)
fgdp <- subset(fgdp,select=c(X.4))
fgdp_no_commas <- as.numeric(gsub(",","",fgdp$X.4))
mean(fgdp_no_commas)
```

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

```
# Question #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"?
# Assume that the variable with the country names in it is named countryNames.
# How many countries begin with United?

fgdp <- read.csv("FGDP.csv",skip=4,nrows=190)
fgdp <- subset(fgdp,select=c(X.3))
colnames(fgdp) <- "countrynames"
grep("^United",fgdp$countrynames)
```

```
## Warning in grep("^United", fgdp$countrynames): input string 99 is invalid in
## this locale
```

```
## Warning in grep("^United", fgdp$countrynames): input string 186 is invalid in
## this locale
```

```
## [1]  1  6 32
```

```
# Question #4 - Load the Gross Domestic Product data for the 190 ranked countries
# and load the educational data. Match the data based on the country shortcode.
```

```r
# Of the countries for which the end of the fiscal year is available, how many
# end in June?

download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv","FEDSTATS_Cou

fgdp <- read.csv("FGDP.csv",skip=4,nrows=190)
fgdp <- subset(fgdp, select=-c(X.2,X.5:X.9))
colnames(fgdp) <- c("CountryCode","Rank","Country","GDP")
country_stats <- read.csv("FEDSTATS_Country.csv")
merged_df <- merge(fgdp,country_stats,by="CountryCode")
sum(grepl("Fiscal year end: June", merged_df$Special.Notes))
```

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

```r
# Question #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?

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

## Loading required package: TTR

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

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

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
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.
```

```r
sampleTimes = index(amzn)
dates_2012 <- sampleTimes[sampleTimes>="2012-01-01" &  sampleTimes<="2012-12-31"]
length(dates_2012)
```

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

```r
 # How many values were collected on Mondays in 2012?

length(dates_2012[wday(dates_2012)==2])
```

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