w4_quiz

Thi NGUYEN

4/26/2021

Question 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() and load the data into R. 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?

```
# split variable names
var_names <- colnames(acs)
split_names <- strsplit(var_names,"wgtp")
q1 <- split_names[[123]]

length(q1) # 2
```

```
## [1] 2
```

```
sprintf("The value of the 123 element of the resulting list is: '%s', '%s'", q1[1],q1[2])
```

```
## [1] "The value of the 123 element of the resulting list is: '', '15'"
```

Question 2:

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

```
# Reshaping dataset
gdp <- as_tibble(gdp)
gdp <- gdp %>%
   select_if(~!(all(is.na(.)) | all(. == ""))) %>%
   select(-V6) %>%
   slice(1:190) %>%
   rename(Ranking = V2 , CountryCode =V1,CountryName = V4,  GDP= V5) %>%
   mutate(GDP = as.numeric(gsub(",","",GDP)))

# Average the GDP numbers
q2 <- round(mean(gdp$GDP,na.rm = TRUE),3)
sprintf("The average of GDP numbers of 190 countries: %.3f (millions dollar)",q2)
```

```
## [1] "The average of GDP numbers of 190 countries: 377652.421 (millions dollar)"
```

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?

```
countryName <- gdp$CountryName
United_countries <- grep("^United", countryName,value = TRUE)

paste(c("The appropriate regular expression:", "grep('^United', countryName,value = TRUE)"),collapse =
```

```
## [1] "The appropriate regular expression: grep('^United', countryName,value = TRUE)"
```

```
q3 <- length(United_countries)
sprintf("There are %s countries whose name starts with 'United'.",q3)
```

```
## [1] "There are 3 countries whose name starts with 'United'."
```

```
sprintf("They are: %s",paste(United_countries, collapse = ", "))
```

```
## [1] "They are: United States, United Kingdom, United Arab Emirates"
```

Question 4:

Load the educational data. 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?

```
# Join two datasets based on country code
res <- inner_join(gdp, edu_data, by ="CountryCode")

# get variable store fiscal year end information
special_notes <- res$Special.Notes

# use grep to find values
fiscal_year_end_available <- grep("^[Fiscal year end]",special_notes,value=TRUE)
fiscal_year_end_june <- grep("[J][u][n][e]", fiscal_year_end_available, value=TRUE)
q4 <- length(fiscal_year_end_june)
q4_ex <- length(fiscal_year_end_available)

sprintf("Of the %s countries for which the end of the fiscal year is available, there are %s that end i
```

```
## [1] "Of the 30 countries for which the end of the fiscal year is available, there are 13 that end in
```

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.

```
# load Amazon dataset
amzn = getSymbols("AMZN",auto.assign=FALSE)
sampleTimes = index(amzn)

# find entries collected in 2012
data_2012 <- grep("[2][0][1][2]",sampleTimes,value=TRUE)
res1 <- length(data_2012)
sprintf("Values collected in 2012: %s entries",res1)
```

```
## [1] "Values collected in 2012: 250 entries"
```

```
# find entries collected on Mondays in 2012
res2 <- length(which(wday(data_2012,label=TRUE) == "Mon"))
sprintf("Values collected on Mondays in 2012: %s entries", res2)
```

```
## [1] "Values collected on Mondays in 2012: 47 entries"
```