This page features MathJax technology to render mathematical formulae. If you are using a screen reader, please visit [MathPlayer](http://www.dessci.com/en/products/mathplayer/) to download the plugin for your browser. Please note that this is an Internet Explorer-only plugin at this time.

**Getting and Cleaning Data**

**Top Navigation Bar**

* [Courses](https://www.coursera.org/courses)
* Harold Simeon

[jhu](https://www.coursera.org/jhu)

Signature Track

[**Getting and Cleaning Data**](https://class.coursera.org/getdata-032/)by Jeff Leek, PhD, Roger D. Peng, PhD, Brian Caffo, PhD

[](https://class.coursera.org/getdata-032)

Top of Form

Search this course Search

Bottom of Form

**Side Navigation Bar**

* **COURSE**
* [**Announcements**](https://class.coursera.org/getdata-032/class/index)
* [**Video Lectures**](https://class.coursera.org/getdata-032/lecture/index)
* **EXERCISES**
* [**Quizzes**](https://class.coursera.org/getdata-032/quiz/index)
* [**swirl Programming Assignment**](https://class.coursera.org/getdata-032/assignment/index)
* [**Course Project**](https://class.coursera.org/getdata-032/human_grading/index)
* **ABOUT THE COURSE**
* [**Syllabus**](https://class.coursera.org/getdata-032/wiki/syllabus)
* [**About the instructor**](https://class.coursera.org/getdata-032/wiki/About_the_Instructor)
* **COMMUNITY**
* [**Discussion Forums**](https://class.coursera.org/getdata-032/forum/index)
* [**DSS Community Site opens in new browser tab**](http://datasciencespecialization.github.io/)
* [Help Center](https://accounts.coursera.org/i/zendesk/courserahelp?return_to=https://learner.coursera.help/hc/)

**Week 4 Quiz \*\*Please Note: No Grace Period\*\***[Help Center](https://accounts.coursera.org/i/zendesk/courserahelp?return_to=https://learner.coursera.help/hc/articles/201523125-Quizzes)

The **due date** for this quiz is **Sun 4 Oct 2015 4:30 PM PDT**.

Top of Form

In accordance with the Coursera Honor Code, I (Harold Simeon) certify that the answers here are my own work.

**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() 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?

"wgtp"

"wgt" "15"

"w" "15"

"" "15"

**Question 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

293700.3

377652.4

381668.9

387854.4

**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?

grep("\*United",countryNames), 5

grep("^United",countryNames), 4

grep("United$",countryNames), 3

grep("^United",countryNames), 3

**Question 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>

15

8

7

13

**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.

library(quantmod)

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

sampleTimes = index(amzn)

How many values were collected in 2012? How many values were collected on Mondays in 2012?

251,51

251, 47

250, 47

252, 47

In accordance with the Coursera Honor Code, I (Harold Simeon) certify that the answers here are my own work.

      

You cannot submit your work until you agree to the Honor Code. Thanks!

Bottom of Form