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 3 Quiz**[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 27 Sep 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>  
  
Create a logical vector that identifies the households on greater than 10 acres who sold more than $10,000 worth of agriculture products. Assign that logical vector to the variable agricultureLogical. Apply the which() function like this to identify the rows of the data frame where the logical vector is TRUE. which(agricultureLogical) What are the first 3 values that result?

125, 238,262

25, 36, 45

236, 238, 262

403, 756, 798

**Question 2**

Using the jpeg package read in the following picture of your instructor into R   
  
<https://d396qusza40orc.cloudfront.net/getdata%2Fjeff.jpg>  
  
Use the parameter native=TRUE. What are the 30th and 80th quantiles of the resulting data? (some Linux systems may produce an answer 638 different for the 30th quantile)

-15259150 -10575416

-10904118 -10575416

-14191406 -10904118

-16776430 -15390165

**Question 3**

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. How many of the IDs match? Sort the data frame in descending order by GDP rank (so United States is last). What is the 13th country in the resulting data frame?   
  
Original data sources:   
<http://data.worldbank.org/data-catalog/GDP-ranking-table>   
<http://data.worldbank.org/data-catalog/ed-stats>

189 matches, 13th country is St. Kitts and Nevis

234 matches, 13th country is Spain

190 matches, 13th country is Spain

190 matches, 13th country is St. Kitts and Nevis

189 matches, 13th country is Spain

234 matches, 13th country is St. Kitts and Nevis

**Question 4**

What is the average GDP ranking for the "High income: OECD" and "High income: nonOECD" group?

32.96667, 91.91304

133.72973, 32.96667

30, 37

23, 45

23.966667, 30.91304

23, 30

**Question 5**

Cut the GDP ranking into 5 separate quantile groups. Make a table versus Income.Group. How many countries are Lower middle income but among the 38 nations with highest GDP?

5

1

18

12

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