**Feedback — Week 1 Quiz** [**Help**](https://class.coursera.org/getdata-003/help/quizzes?url=https%3A%2F%2Fclass.coursera.org%2Fgetdata-003%2Fquiz%2Ffeedback%3Fsubmission_id%3D2041%26sig_reload%3D1)

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Thu 8 May 2014 6:46 AM EDT**. You got a score of **15.00** out of **15.00**.

Top of Form

**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>  
  
How many housing units in this survey were worth more than $1,000,000?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| 2076 |  |  |  |
| 164 |  |  |  |
| 53 | Correct | 3.00 |  |
| 159 |  |  |  |
| Total |  | 3.00 / 3.00 |  |

**Question 2**

Use the data you loaded from Question 1. Consider the variable FES in the code book. Which of the "tidy data" principles does this variable violate?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Each variable in a tidy data set has been transformed to be interpretable. |  |  |  |
| Tidy data has one variable per column. | Correct | 3.00 |  |
| Tidy data has no missing values. |  |  |  |
| Tidy data has one observation per row. |  |  |  |
| Total |  | 3.00 / 3.00 |  |

**Question 3**

Download the Excel spreadsheet on Natural Gas Aquisition Program here:   
  
<https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FDATA.gov_NGAP.xlsx>  
  
Read rows 18-23 and columns 7-15 into R and assign the result to a variable called:

dat

What is the value of:

sum(dat$Zip\*dat$Ext,na.rm=T)

(original data source: <http://catalog.data.gov/dataset/natural-gas-acquisition-program>)

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| 338924 |  |  |  |
| 0 |  |  |  |
| 36534720 | Correct | 3.00 |  |
| 184585 |  |  |  |
| Total |  | 3.00 / 3.00 |  |

**Question 4**

Read the XML data on Baltimore restaurants from here:   
  
<https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml>  
  
How many restaurants have zipcode 21231?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| 17 |  |  |  |
| 156 |  |  |  |
| 127 | Correct | 3.00 |  |
| 181 |  |  |  |
| Total |  | 3.00 / 3.00 |  |

**Question 5**

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%2Fss06pid.csv>   
  
using the fread() command load the data into an R object

DT

Which of the following is the fastest way to calculate the average value of the variable

pwgtp15

broken down by sex using the data.table package?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| DT[,mean(pwgtp15),by=SEX] |  |  |  |
| sapply(split(DT$pwgtp15,DT$SEX),mean) | Correct | 3.00 |  |
| rowMeans(DT)[DT$SEX==1]; rowMeans(DT)[DT$SEX==2] |  |  |  |
| mean(DT[DT$SEX==1,]$pwgtp15); mean(DT[DT$SEX==2,]$pwgtp15) |  |  |  |
| mean(DT$pwgtp15,by=DT$SEX) |  |  |  |
| tapply(DT$pwgtp15,DT$SEX,mean) |  |  |  |
| Total |  | 3.00 / 3.00 |  |

Bottom of Form