

Week 1 Quiz

[Help](#)

The **due date** for this quiz is **Sun 11 May 2014 4:30 PM PDT**.

- ☐ In accordance with the Coursera Honor Code, I (Gene Cho) 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%2FPUMSDDataDict06.pdf>

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

- ☐ 24
- ☐ 53
- ☐ 164
- ☐ 25

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?

- ☐ Tidy data has one observation per row.
- ☐ Tidy data has variable values that are internally consistent.
- ☐ Each tidy data table contains information about only one type of observation.
- ☐ Tidy data has one variable per column.

Question 3

Download the Excel spreadsheet on Natural Gas Acquisition 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>)

- ☐ 36534720
- ☐ 338924
- ☐ 33544718
- ☐ 154339

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?

- ☐ 130
- ☐ 28
- ☐ 156
- ☐ 127

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?

- ☐ `mean(DT[DT$SEX==1,]$pwgtp15); mean(DT[DT$SEX==2,]$pwgtp15)`
- ☐ `tapply(DT$pwgtp15,DT$SEX,mean)`
- ☐ `sapply(split(DT$pwgtp15,DT$SEX),mean)`
- ☐ `DT[,mean(pwgtp15),by=SEX]`
- ☐ `mean(DT$pwgtp15,by=DT$SEX)`
- ☐ `rowMeans(DT)[DT$SEX==1]; rowMeans(DT)[DT$SEX==2]`

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

[Submit Answers](#)[Save Answers](#)

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