# Week 1 Quiz

Help Center

The **hard deadline** for this quiz is **Fri 18 Dec 2015 4:30 PM PST**.

☐ **In accordance with the Coursera Honor Code, I (Hugo Soares) 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

How many properties are worth $1,000,000 or more?

- ○ 53
- ○ 159
- ○ 24
- ○ 2076

## 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 no missing values.
- ○ Numeric values in tidy data can not represent categories.

- Each variable in a tidy data set has been transformed to be interpretable.
- Tidy data has one variable per column.

## 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)

- 36534720
- NA
- 184585
- 0

## 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?

- 100
- 181
- 17
- 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?

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

☐ **In accordance with the Coursera Honor Code, I (Hugo Soares) 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!