



Week 1 Quiz



3/5 questions correct

You haven't passed yet. You need at least 4 questions correct to pass.

Review the material and try again! You have 3 retakes every 8 hours.

[Review Related Lesson \(/learn/data-cleaning/home/week/1\)](/learn/data-cleaning/home/week/1)



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>
([https://d396qusza40orc.cloudfront.net](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv)
[/getdata%2Fdata%2Fss06hid.csv](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](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDDataDict06.pdf)
[/getdata%2Fdata%2FPUMSDDataDict06.pdf](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDDataDict06.pdf)
([https://d396qusza40orc.cloudfront.net](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDDataDict06.pdf)
[/getdata%2Fdata%2FPUMSDDataDict06.pdf](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDDataDict06.pdf))

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

☐ 53

Well done!

☐ 2076

☐ 25

☐ 164

✖ 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?

- ☐ Numeric values in tidy data can not represent categories.

Sorry, that's incorrect.

- ☐ Each variable in a tidy data set has been transformed to be interpretable.
- ☐ Tidy data has one variable per column.
- ☐ Tidy data has variable values that are internally consistent.
-

✖ 3.

Download the Excel spreadsheet on Natural Gas Aquisition Program here:

https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FDATA.gov_NGAP.xlsx
(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> (<http://catalog.data.gov/dataset/natural-gas-acquisition-program>))

☐ 0

Sorry, that's incorrect.

☐ 36534720

☐ 338924

☐ 184585




4.

Read the XML data on Baltimore restaurants from here:

[https://d396qusza40orc.cloudfront.net
/getdata%2Fdata%2Frestaurants.xml](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml)
([https://d396qusza40orc.cloudfront.net
/getdata%2Fdata%2Frestaurants.xml](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml))

How many restaurants have zipcode 21231?

☐ 100☐ 181☐ 130☐ 127
Well done!



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>
([https://d396qusza40orc.cloudfront.net](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06pid.csv)
[/getdata%2Fdata%2Fss06pid.csv](https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06pid.csv))

using the `fread()` command load the data into an R object

DT

The following are ways to calculate the average value of the variable

pwgtp15

broken down by sex. Using the `data.table` package, which will deliver the fastest user time?

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

Well done!

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

