

Week 1 Quiz



5/5 questions correct

Quiz passed!

Continue Course (/learn/data-cleaning/lecture/njjbw/reading-from-mysql)

Back to 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/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%2FPUMSDataDic t06.pdf

(https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FPUMSDataDict06.pdf)

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



2076



53



Well done!		
0	164	
0	25	
	2.	
	e data you loaded from Question 1. Consider the variable FES in the book. Which of the "tidy data" principles does this variable violate?	
0	Each variable in a tidy data set has been transformed to be interpretable.	
0	Numeric values in tidy data can not represent categories.	
0	Tidy data has no missing values.	
0	Tidy data has one variable per column.	

Well done!



Download the Excel spreadsheet on Natural Gas Aquisition Program here:

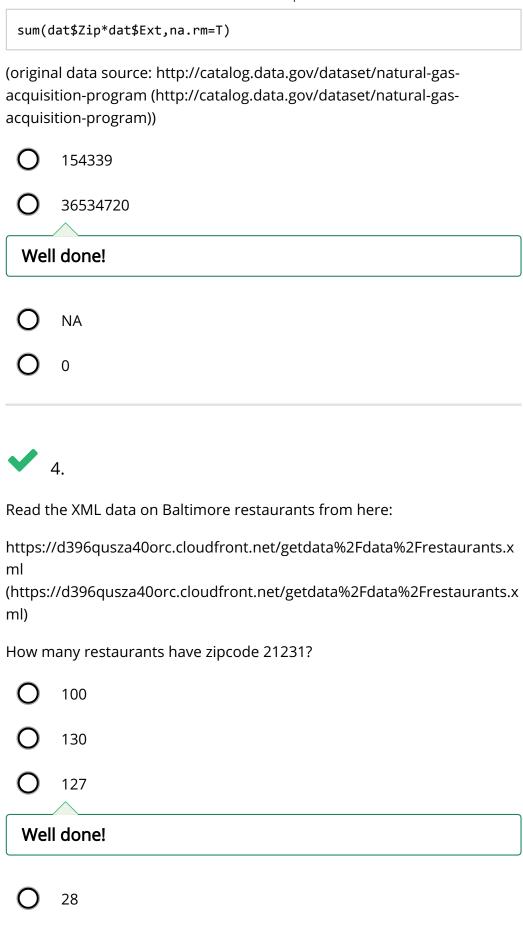
https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FDATA.gov_NG AP.xlsx

(https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FDATA.gov_N GAP.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:





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/getdata%2Fdata%2Fss06pid.csv)

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

asing the neadly command load the data into an it asject		
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?		
0	mean(DT[DT\$SEX==1,]\$pwgtp15); mean(DT[DT\$SEX==2,]\$pwgtp15)	
0	tapply(DT\$pwgtp15,DT\$SEX,mean)	
0	DT[,mean(pwgtp15),by=SEX]	
Well done!		
0	sapply(split(DT\$pwgtp15,DT\$SEX),mean)	
0	mean(DT\$pwgtp15,by=DT\$SEX)	
0	rowMeans(DT)[DT\$SEX==1]; rowMeans(DT)[DT\$SEX==2]	





