Back

Week 1 Quiz

Quiz, 5 questions

**5/5 points (100%)**

**Congratulations! You passed!**

Next Item

Question 1

Correct

1 / 1

point

**1. 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?



31



47



53

**Correct**



164

Question 2

Correct

1 / 1

point

**2. 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?



Each tidy data table contains information about only one type of observation.



Tidy data has one observation per row.



Each variable in a tidy data set has been transformed to be interpretable.



Tidy data has one variable per column.

**Correct**

Question 3

Correct

1 / 1

point

**3. 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:



1

dat

What is the value of:



1

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

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



0



36534720

**Correct**



338924



184585

Question 4

Correct

1 / 1

point

**4. 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



130



127

**Correct**



28

Question 5

Correct

1 / 1

point

**5. 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



1

DT

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



1

pwgtp15

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



mean(DT$pwgtp15,by=DT$SEX)



tapply(DT$pwgtp15,DT$SEX,mean)



DT[,mean(pwgtp15),by=SEX]

**Correct**



rowMeans(DT)[DT$SEX==1]; rowMeans(DT)[DT$SEX==2]



mean(DT[DT$SEX==1,]$pwgtp15); mean(DT[DT$SEX==2,]$pwgtp15)



sapply(split(DT$pwgtp15,DT$SEX),mean)