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

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

- 2076
- 53

Well done!

- 164
- 25

✔ 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 variable in a tidy data set has been transformed to be interpretable.
- Numeric values in tidy data can not represent categories.
- Tidy data has no missing values.
- Tidy data has one variable per column.

Well done!

✔ 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))

- 154339
- 36534720

Well done!

- NA
- 0

✔ 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)

How many restaurants have zipcode 21231?

- 100
- 130
- 127

Well done!

- 28

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/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?

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

Well done!

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