Week 1 Quiz

Back to Week 1

4/5 points earned (80%)

Quiz passed!

1 / 1 points

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
- 53

Correct

- 2076
- 24

1 / 1 points

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

Tidy data has one variable per column.

Correct

Tidy data has one observation per row.

---

1 / 1 points

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)

338924

36534720

Correct

0

184585

---

1 / 1 points

4.

Read the XML data on Baltimore restaurants from here:

https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml

How many restaurants have zipcode 21231?

28

130

127

Correct

100

0 / 1 points

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?

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

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

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

This should not be selected

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

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

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