# Week 1 Quiz

NOTE DE LA SOUMISSION LA PLUS RÉCENTE
100%

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?

1 / 1 point

- ◯ 31
- ◯ 47
- ◯ 164
- ◉ 53

✓ **Correct**

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?

1 / 1 point

- ◯ Tidy data has one observation per row.
- ◉ Tidy data has one variable per column.
- ◯ 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.

✓ **Correct**

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)

1 / 1 point

- ◉ 36534720
- ◯ NA
- ◯ 0
- ◯ 154339

✓ **Correct**

4. Read the XML data on Baltimore restaurants from here:

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

How many restaurants have zipcode 21231?

1 / 1 point

- ◯ 17
- ◯ 181
- ◉ 127
- ◯ 100

✓ **Correct**

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?

1 / 1 point

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

✓ **Correct**