


Feedback — Weekly Quiz 2

You submitted this quiz on **Fri 1 Feb 2013 10:42 AM CET**. You got a score of **10.00** out of **10.00**.


Question 1

In the text of the final write-up of a data analysis, how should the analyses be reported?

Your Answer	Score	Explanation
 Analyses should be reported in an order to convey the story being told with the data analysis.	✓ 1.00	
Total	1.00 / 1.00	

Question 2

Open a connection to the old version of my blog: <http://simplystatistics.tumblr.com/> , read the first 150 lines of the file and assign them to a vector simplyStats. Apply the nchar() function to simplyStats to count the characters in each element of simplyStats. How many characters long are the lines 2, 45, and 122?

Your Answer	Score	Explanation
 918, 5, 24	✓ 1.00	
Total	1.00 / 1.00	

Question 3

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://dl.dropbox.com/u/7710864/data/csv_hid/ss06hid.csv or here



<https://spark-public.s3.amazonaws.com/dataanalysis/ss06hid.csv>

and load the data into R. You will use this data for the next several questions. The code book, describing the variable names is here:

<https://dl.dropbox.com/u/7710864/data/PUMSDataDict06.pdf> or here:



<https://spark-public.s3.amazonaws.com/dataanalysis/PUMSDataDict06.pdf>

How many housing units in this survey were worth more than \$1,000,000?

Your Answer	Score	Explanation
 53	 1.00	
Total	1.00 / 1.00	



Question 4

Use the data you loaded from Question 3. Consider the variable FES. Which of the "tidy data" principles does this variable violate?

Your Answer	Score	Explanation
 Tidy data has one variable per column.	 1.00	
Total	1.00 / 1.00	

Question 5

Use the data you loaded from Question 3. How many households have 3 bedrooms and 4 total rooms? How many households have 2 bedrooms and 5 total rooms? How many households have 2 bedrooms and 7 total rooms?



Your Answer	Score	Explanation
 148, 386, 49	 1.00	
Total	1.00 / 1.00	

Question 6

Use the data from Question 3. Create a logical vector that identifies the households on greater than 10 acres who sold more than \$10,000 worth of agriculture products. Assign that logical vector to the variable agricultureLogical. Apply the which() function like this to identify the rows of the data frame where the logical vector is TRUE.

```
which(agricultureLogical)
```

What are the first 3 values that result?

Your Answer	Score	Explanation
 125, 238,262	 1.00	
Total	1.00 / 1.00	

Question 7

Use the data from Question 3. Create a logical vector that identifies the households on greater than 10 acres who sold more than \$10,000 worth of agriculture products. Assign that logical vector to the variable agricultureLogical. Apply the which() function like this to identify the rows



of the data frame where the logical vector is TRUE and assign it to the variable indexes.

```
indexes = which(agricultureLogical)
```

If your data frame for the complete data is called `dataFrame` you can create a data frame with only the above subset with the command:



```
subsetDataFrame = dataFrame[indexes,]
```

Note that we are subsetting this way because the NA values in the variables will cause problems if you subset directly with the logical statement. How many households in the `subsetDataFrame` have a missing value for the mortgage status (MRGX) variable?

Your Answer	Score	Explanation
 8	 1.00	
Total	1.00 / 1.00	



Question 8

Use the data from Question 3. Apply `strsplit()` to split all the names of the data frame on the characters "wgt". What is the value of the 123 element of the resulting list?

Your Answer	Score	Explanation
 "" "15"	 1.00	
Total	1.00 / 1.00	

Question 9

What are the 0% and 100% quantiles of the variable YBL? Is there anything wrong with these values? Hint: you may need to use the `na.rm` parameter.

Your Answer	Score	Explanation
 -1, 25, Something wrong.	 1.00	
Total	1.00 / 1.00	

Question 10

In addition to the data from Question 3, the American Community Survey also collects data about populations. Using `download.file()`, ownload the population record data from:

https://dl.dropbox.com/u/7710864/data/csv_hid/ss06pid.csv



or here

<https://spark-public.s3.amazonaws.com/dataanalysis/ss06pid.csv>

Load the data into R. Assign the housing data from Question 3 to a data frame `housingData` and the population data from above to a data frame `populationData`.

Use the merge command to merge these data sets based only on the common identifier "SERIALNO". What is the dimension of the resulting data set?

[OPTIONAL] For fun, you might look at the data and see what happened when they merged.

Your Answer	Score	Explanation
 number of rows = 15451, number of columns = 426	 1.00	
Total	1.00 / 1.00	