Quiz 1

Name:

1.

You have the following dataframe saved as the R object `our_course`:

```
##   class_session      date
## 1             1  Aug. 24
## 2             2  Aug. 31
## 3             3 Sept. 14
## 4             4 Sept. 21
## 5             5 Sept. 28
```

How would you print the first two rows at your R console?

    a. head(our_course)
    b. our_course[1:2]
    c. our_course[1:2,]
    d. read.csv(our_course, nrows = 2, header = TRUE)

> **Answer c.** Explanation: `our_course[1:2, ]` subsets out the first two rows (the numbers before the comma in the square brackets) and all of the columns (the blank space after the comma in the square brackets). Answer a. would print out the first few rows of the dataframe, but the default for `head` is to print out 6 rows, not 2 (`head(our_course, n = 2)` would have been correct, though). Answer b. does not work because `our_course` is a dataframe (2-D), not a vector (1-D), so you need two elements inside the square brackets to properly index it (i.e., you need a comma). Answer d. does not work because `read.csv()` reads in the data– you don't want to read in the data, you already have it saved as the R object `our_course`. `read.csv()` needs a file name for a flat file and won't work if you try to put in a dataframe as the first argument to the function.

2.

If you enter the command `getwd()` at your R console, what is printed out?

    a. Your current working directory.
    b. The code that makes up the function `getwd()`.
    c. Your home directory.
    d. Your root directory.

> **Answer a.** Explanation: If you typed `getwd` without the parentheses, you would get information about the code for the `getwd` function (although, because it's an Internal function, you wouldn't get loads of information). Answers c. or d. would only be right if you were working in, respectively, your home or root directory, but since these are only conditionally correct answers, a. is the right answer here.

3.

You have the R package `ggplot2` saved on your computer from using it in an earlier R session. You start a new R session, and you need to use some functions from this package. What code do you need to run?

    a. `install.packages("ggplot2")`
    b. `library(ggplot2)`
    c. Nothing, you're ready to go since it's already on your computer.
    d. `library()`

> **Answer b.** Explanation: Since you already have the package saved on your computer, you do not need to reinstall it, which makes Answer a. incorrect. If you just started a new R session, then the library will not be loaded in your current R session, so Answer c. is not correct. Answer d. is incorrect because, if you run `library()` without anything in the parentheses, R will give you a list of all of the available packages you could load, but to load the one you want, you need to specify which package you want to load.

4.

Refer to the `our_course` dataframe shown in the first question. What is the only code that would **not** work if you want to print out the first four values in the `date` column?

    a. head(our_course[, 2], 4)
    b. our_course[1:4, "date"]
    c. our_course[1:4, 2]
    d. our_course[1:4, date]

> **Answer d.** Explanation: Answer a. would work because it prints the first four rows (`head()` function with option `n = 4`) of the second column (`[ , 2]'`) of `our_course` (`date` is the second column, which is what we want). Answer b. works because it is indexing out the first four rows (`1:4` before the comma in the square brackets) and the column named `"date"` (part after the comma in the square brackets) from `our_course`. Answer c. would work because it's indexing the first four rows and second column (`[1:4, 2]`) of the `our_course` dataframe. Answer d. would only work if you had previously defined the `date` object to be the character `"date"` or the number 2 (i.e., `date <- "date"` or `date <- 2`). Otherwise, R looks for an object called `date` to use for this part of the index and can't find one.

5.

Say you have a .csv file that looks like this when you open it in a text editor:

```
This is a file of data on course
meetings for R Programming
class_session, date
1, "Aug. 24"
2, "Aug. 31"
3, "Sept. 14"
```

Which options would you need to use in the `read.csv()` function to read this file into R?

    a. `header = FALSE, nrows = 2`
    b. `header = TRUE, skip = 2`

c. `header = TRUE, nrows = 2`

d. `skip = 2, nrows = 2`

Answer b. Explanation: Answer b. is right because it skips two lines (the ones where I've just written some general information that I don't want R to read) and then assigns the first line after that to be the column names for the rest of the data. Answer a. is wrong because R would try to read in two rows starting from the very first line and also would assume there's not a row to give the column names. I think it would probably end up giving you two rows with one column, where each value in the column is the whole text of the line (e.g., "This is a file of data on course" would be the value in the first entry for the one column). Answer c. is wrong because R would start reading from the very first row, and so would be trying to read in some of that text about the file that doesn't make up part of the data. Answer d. is wrong because it does not let R know that the first line you're reading in is a header, plus it only reads in two of the lines you want to read in.

6.

What command would you use to list all the files in your current working directory?

a. `getwd()`

b. `setwd("~")`

c. `list.files()`

d. `ls()`

Answer c. Explanation: Answer a. would print your working directory pathname. Answer b. would re-set your working directory to your home directory. Answer d. would list all of the objects in your current R session, not the files in your working directory.

7.

Refer again to the example dataframe from Question 1. How could you find out how many observations are in this dataframe?

a. `nrow(our_course[1, ])`

b. `dim(our_course)`

c. `length(our_course)`

d. `ncol(our_course[1, ])`

Answer b. Explanation: `dim(our_course)` will give you more information than you need– it will give you the number of rows (observations) and columns– but it is the right answer because it is the only one that will give you the number of observations in the full dataset. Answer a. would give you `1`, because it will only count the number of rows in the part of the dataframe you are indexing, which is just the first row (`our_course[1, ]`). Answer c is wrong because, if you use `length()` on a dataframe, it will give you the number of columns, not rows (note: `length()` could work for getting the number of observations if you have a *vector* rather than a *dataframe*. For example, `length(our_course[ , 1])`). Answer d. would give you the number of columns for the first row of the data. Observations are rows, while variables are columns, so this would work if you were trying to get the number of variables, but will not give you the number of observations.

8.

Where do files need to be for you to directly read them into your R session (i.e., read them in with only their file name, like `read.csv("FileName.csv", header = TRUE)`)?

   a. My GitHub repository
   b. Your home directory
   c. Your current working directory.
   d. Your desktop

Answer c. As we discussed in class on Monday, you will need to use longer pathname to read in any file that's not in your working directory. Answers b. or d. would only be correct if your working directory were, respectively, your home directory or your working directory. To read a file from GitHub, you would need to use the `repmis` package and use a different command than `read.csv()` (also, you would need to use the full web address for the raw file on GitHub).