Problem Set 1

Name:

About this format: This problem set is saved as an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code. When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file). To add a new R chunk click the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

TO submit: Please rename this file LASTNAME\_ProblemSet1 and upload to the assignments folder on Canvas before class on W 1/17

1. Take the following courses in swirl. Some of these will be parallel to and reinforcing of our in-class exercise, and others will expand on what we have learned. As you proceed with each module please pause and use the play() function at relevant stages in each module to answer the associated questions. Please note that swirl is a little buggy, and if you answer a multiple choice incorrectly it may give you an error and exit swirl. If that happens, type swirl() to resume and enter the same user name you used previously, and you should be able to navigate directly back to where you left off.
2. R Programming -> 4. Vectors  
   Copy your assignment of num\_vector in the chunk below. Create a new vector that is num\_vector divided by 3

num\_vect <- c(1,2,3)  
num\_vect

## [1] 1 2 3

1. R Programming -> 5. Missing Values  
   Copy your assignment of my\_data below, then repeat the code that generated my\_data to create my\_data2. Using R logic statements, test whether my\_data and my\_data2 are identical. In hash-tagged lines at the bottom of the chunk explain why or why not.

y <- rnorm(1000)  
z <- rep(NA, 1000)

1. R Programming -> 6. Subsetting  
   Copy your assignment of vect below. Save vect as a new object vect2 with the names c(“first”, “second”, “third”).
2. R Programming -> 7. Matrices and Data Frames  
   Copy your assignment of patients below. Bind the vectors patients and patientnumber together to create a data frame.

patientnumber <- c(1,2,3,4)

1. R Programming -> 8. Logic  
   In the chunk below create a vector called logicvec where the first entry is the expression that 5 is equal to 5, the second entry is the expression that 6 is not equal to 7, the third entry is the expression that 6 is equal to or greater than 7, and the fourth is that 2 or 3 is less than 3. Use logicvec as an argument to the function isTRUE. What is the output, and why? The helpfile for isTRUE (?isTRUE) might be useful.
2. Without running the code, what does the following print? Please explain why.

a <- 1  
b <- 2  
c <- a + b  
b <- 4  
a <- b  
c <- a  
c

Answer:

1. Three of the following lines produce the same result. Without running the code, which one will produce a different result than the others? Please explain why. The helpfile for log (?log) may be helpful.

log(x = 1000, base = 10)  
log10(1000)  
log(base = 10, x = 1000)  
log(10, 1000)

Answer:

1. Use a function to create a new vector that is the entries in numvector arranged in descending order. We have not learned this function, but a combination of google and function documentation should get you there.

numvector <- c(5,2,3,1,6,8)

1. Which elephant weighs more? Convert one’s weight to the units of the other, and store the result in an appropriately-named new variable. In the R chunk below write a command to test whether elephant1 weights more than elephant2 (1 kg ≈ 2.2 lb).

elephant1\_kg <- 3492  
elephant2\_lb <- 7757

1. The National Ecological Observatory Network has 20 replicated flux towers across the United States to measure atmospheric conditions. Automated sensors on the towers continuously collect data on the flow of chemicals between the atmosphere and the surface of the ecosystem (e.g., carbon dioxide, wind speed). In what ways is this dataset “big data”? In what ways is it not big data?

Answer: