Homework 1

Statistical Computing, 36-350

Due Wednesday Sept 9, 2015

Your homework must be submitted in R Markdown format. We will not (indeed, cannot) grade homeworks in other formats. Your responses must be supported by both textual explanations and the code you generate to produce your result. (Just examining your various objects in the "Environment" section of R Studio is insufficient—you must use scripted commands.)

1. Reading, exploring and editing data in R. The data set available at http://www.stat.cmu.edu/ ~ryantibs/statcomp/homework/strike.txt consists of annual observations on the level of strike volume (days lost due to industrial disputes per 1000 wage salary earners) and their covariates in 18 OECD countries from 1951-1985. Details can be found at http://lib.stat.cmu.edu/datasets/strikes.

a. First, we need to load the data set into R using

Note: it is important here (and the same for all future homeworks) that you pass the full url to the read.table() function; do not save the strike.txt file locally to your computer first.

- b. How many rows and columns does strikedat have? (If you do not have 625 rows and 7 columns, something is wrong; check the previous part to see what might have gone wrong in the previous part.)
- c. What are the names of the columns of strikedat?
- d. What is the value of row 123, column 4 of strikedat?
- e. Display the last 15 entries of the second column of strikedat in its entirety.
- f. Explain what this command does, by running it on your data and examining the object. (You may find the display functions head() and tail() useful here.)

- g. The column named lefwingprop contains a percentage (between 0 and 100). Create a new column in the data frame called leftwingprop.scaled that contains the actual proportion (between 0 and 1). Display the first five rows of this dataset. (You may find head() helpful here.)
- h. Using this new column, create a line plot of leftwingprop.scaled for country 1 (hint: use the columns named natcode) where the y axis is the proportion and the x axis is year. Is there an apparent trend over time?
- i. Instead of appending columns, create a new data frame strikedat.fix that takes the original dataset and *replaces* the columns for unemployment and leftwingprop with proportions. Display the first five rows of this new dataset.

2. Syntax and class-typing.

a. For each of the following commands, either explain why they should be errors, or explain the non-erroneous result.

```
vector1 <- c("5", "12", "7", "42")
max(vector1)
sort(vector1)
sum(vector1)</pre>
```

b. For the next series of commands, either explain their results, or why they should produce errors.

```
vector2 <- c("5",9,12)
vector2[2] + vector2[3]
dataframe3 <- data.frame(v1="5",v2=10,v3=12)
dataframe3[1,2] + dataframe3[1,3]
list5 <- list(v1="6", v2=42, v3="49", v4=126, v5 = "25")
list5[[3]]+list5[[4]]
list5[3]+list5[4]
```

3. Working with functions and operators.

- a. The colon operator will create a sequence of integers in order, as in 1:5, which gives back a vector containing the numbers 1 through 5. This is a special case of the function seq(). Using the help command ?seq to learn about the function, design an expression using seq()that will give you the equivalent of (1:5000)*2. Also design an expression to create the sequence of numbers from 1 to 10000 in increments of 315. Design another that will give you a sequence between 1 and 10000 that is exactly 40 numbers in length.
- b. The function rep() repeats a vector some number of times. Explain the difference between rep(1:10, times=5) and rep(1:10, each=5).