Stat 133-Assignment #1

June 16, 2014

You will find the file "assignment-1.r" included in this folder. It contains an outline for the code you are to submit. For each question, we have implemented some of the code. Your job is to complete the code where indicated so that the functions (and variables) perform as requested. In questions where we ask you to supply a variable, we have commented out the variable name (using the "#" symbol). This will allow your code to run properly when testing other functions. You will need to delete the "#" in front of variable names after creating them in order to run those lines of code.

The data file related to the first two problems comes from Phil Spector (a former professor at Berkeley). It contains the low, average, and high temperatures in Berkeley for every day in January from 2005-2011. We have imported the data for you from Phil's website using the command "read.table" which stores the data as a data frame.

Problem 1. Functions

(a) The dataset gives temperature in degrees Farenheit, but suppose you would like to convert to degrees Celcius. The conversion is given by

$$^\circ C = 5/9 \cdot (^\circ F - 32).$$

Implement a function that converts vectors of $^\circ F$ to $^\circ C$ (you don't actually need to run your function on the data).

(b) Suppose you are interested in the statistic

$$S_{ij} = \frac{x_{ij}^{high} - x_{ij}^{low}}{\bar{x}_{ij}}$$

where $x_{ij}^{high}, x_{ij}^{low}, \bar{x}_{ij}$ are the high, low, and average temperatures of year j on day i respectively. Write a function that takes a data frame (like our weather data) and two numeric values (one for day and one for year) as arguments and returns S_{ij} (again, you don't actually need to run your function on the data). Your function should work for any data frame with the same variables as **temperature.data** and dates within the specified range.

Problem 2. Data Frames

(a) Subset the data corresponding for all observations from the year 2010. This subset should be stored as a data frame with the name **subset.2010**. For this subset, find the spread (i.e. max - min) of temperatures

for each day and store this vector as the variable **temp.differences**. What was the maximum spread in temperatures? Store this value as **max.difference**. On what day (just the number given by the "day" variable) did this occur? Store this value as **max.difference.day**.

(b) Use R's "quantile" function to split the data frame into two subsets: days for which the daily high temperature was above the 65^{th} percentile and days for which the daily high temperature was below the 65^{th} percentile. Calculate the mean of the daily low temperature for both of these subsets. Store these means as **mean.low.above** and **mean.low.below.**

Problem 3. More Data Frames

Included in the assignment-1 folder you will find the assignement-1-3.Rda file, which we have loaded into your workspace. It contains contains two variables: **observed.animals** and **animal.key**. The former is a vector of 500 animals (14 unique). The latter is a data frame giving the diet (carnivore or herbivore) and type (mammal, reptile, or bird) for each unique animal. Create the following vectors:

1. **observed.diets**, a vector of length 500 that gives the diet corresponding to each of **observed.animals**
2. **observed.types**, a vector of length 500 that gives the type corresponding to each of **observed.animals**

Use these vectors to calculate the number of **observed.animals** that are both carnivores and mammals. Store this value as the vector **n.carnivore.mammals**.