Homework 1

36-350, Statistical Computing

Due at 11:59 pm on Thursday, 5 September 2013

Submit a single plain text file. (Files in other formats won't be graded.) Clearly indicate which parts of your responses go with which problems. Raw R is unacceptable, and will be marked down; you are communicating with human beings, and need to write in a human language.

This homework, like all the others, has a maximum possible score of 100; 90 points on this are for correctness of your answers (as marked), and 10 for over-all clarity and thoroughness.

1. *Working with data* (45 points total) The data set at `http://www.stats.uwo.ca/faculty/braun/data/rnf6080.dat` records hourly rainfall at a certain location in Canada, every day from 1960 to 1980.

    (a) (1) Load the data set into R and make it a data frame called `rain.df`. What command did you use?

    (b) (1) How many rows and columns does `rain.df` have? How do you know? (If there are not 5070 rows and 27 columns, you did something wrong in the first part of the problem.)

    (c) (1) What command would you use to get the names of the columns of `rain.df`? What are those names?

    (d) (1) What command would you use to get the value at row 2, column 4? What is the value?

    (e) (1) What command would you use to display the whole second row? What is the content of that row?

    (f) (3) What does the following command do?

    ```
    names(rain.df) <- c("year","month","day",seq(0,23))
    ```

    (g) (4) Create a new column called `daily`, which is the sum of the 24 hourly columns.

    (h) (8) Give the command you would use to create a histogram of the daily rainfall amounts. Submit this histogram as a separate PDF file, named with your Andrew ID, the assignment number, and `Fig1`.

    (i) (10) Explain why that histogram cannot possibly be right.

    (j) (10) Give the command you would use to fix the data frame.

(k) (5) Create a corrected histogram and submit a PDF file of it, named as before but with `Fig2` instead of `Fig1`. Explain why it is more reasonable than the previous histogram.

2. *Data types* (25 points) Make sure your answers to different parts of this problem are compatible with each other.

(a) (9) For each of the following commands, either explain why they should be errors, or explain the non-erroneous result.

```
x <- c("5","12","7")
max(x)
sort(x)
sum(x)
```

(b) (8) For the next two commands, either explain their results, or why they should produce errors.

```
y <- c("5",7,12)
y[2] + y[3]
```

(c) (8) For the next two commands, either explain their results, or why they should produce errors.

```
z <- data.frame(z1="5",z2=7,z3=12)
z[1,2] + z[1,3]
```

3. *Drawing outside the lines* (20) George is bored in class, so he fills his notepad with random lines. The paper is horizontally ruled, with rules exactly two centimeters apart. George's lines have centers uniformly distributed over the page, and uniformly distributed angles. Each of his lines is exactly a centimeter long. (George is obsessive as well as bored.)

(a) (4) Explain why the vertical distance between the center of one of George's lines and the nearest rule on the paper should be uniformly distributed between 0 and 1cm.

(b) (4) The R command `runif(n=137,min=-1,max=0.9)` returns 137 numbers uniformly distributed between -1 and 0.9. What commands would you use to generate the random centers of 1000 of George's lines[1], and store them in a vector named `centers`?

(c) (4) R measures angles in radians. What commands would you use to generate the angle from vertical of George's lines, and store them in a vector called `orientations`?

(d) (4) If y is the location of the center of a line and a is its angle from the vertical, the line crosses one of the horizontal rules when $y \leq f(a)$. What is the function f? (*Hint*: trigonometry.)

---

[1] George is *very* bored by programming lectures.

(e) (4) What commands would you use to calculate the fraction of George's 1000 lines which cross horizontal rules? What is the fraction?

(f) (5, extra credit) This is a version of a famous math problem. Name the problem, and what the fraction in the last part should be.