

Statistics 5014: Homework 4

Due Wednesday September 27 IN CLASS Torg 1100, 9 am

2017-09-26

For each assignment, turn in by the due date/time. Late assignments must be arranged prior to submission. In every case, assignments are to be typed neatly using proper English in Markdown.

This week, we spoke about Exploratory Data Analysis and plotting. To begin the homework, we will as usual, start by loading, munging and creating tidy data sets. In this homework, our goal is to create informative (and perhaps pretty) plots showing features or perhaps deficiencies in the data.

Problem 1

Work through the Swirl “R_programming_E” lesson parts 10 and 11.

swirl()

Problem 2

As in the last homework, create a new R Markdown file within the project folder within the “05_R_apply_family” subfolder (file->new->R Markdown->save as).

The filename should be: HW4_lastname_firstname, i.e. for me it would be HW5_Settlage_Bob

You will use this new R Markdown file to solve the following problems:

Problem 3

What are you thoughts for what makes a good figure?

Problem 4

- Create a function that computes the proportion of successes in a vector. Use good programming practices.
- Create a matrix created to simulate 10 flips of a coin with varying degrees of “fairness” as follows:

```
P4b_data <- matrix(rbinom(10, 1, prob = (1:10)/100), nrow = 10,
  ncol = 10)
```

- Use your function in conjunction with apply to compute the proportion of success in P4b_data by column and row. What do you observe? What is going on?
- Fix the above matrix by creating a function whose input is a probability and output is a vector whose elements are the outcomes of 10 flips of a coin. Now create a vector of the desired probabilities. Using the appropriate apply family function, create the matrix we really wanted above. Prove this has worked by using the function created in part a to compute and tabulate the appropriate marginal successes.

Problem 5

Load, munge, explore the data given in Wu and Hamada from the starch experiment. Consider strength as the response. You do not need to form a model or otherwise analyze the dataset, you do need to explore the data and make observations.

<http://www2.isye.gatech.edu/~jeffwu/book/data/starch.dat>

Problem 6

Push your homework and submit a pull request.

When it is time to submit, –ONLY– submit the .Rmd and .pdf solution files. Names should be formatted HW4__lastname__firstname.Rmd

Optional preperation for next class:

Next week we will talk about the dual handling of vectors and matrices in R. No swirl. :)