

Homework:

The benefit of doing data manipulation in R is that there's a record of every step you've taken. Please work independently, and write an R script to do the following:

(Before you get started, download the files "admissions.csv" and "ptdata.csv" and save them in your working directory.)

1. Read in "admissions.csv", clean and format the data, and add the column weekdayname, as we did in class.
2. Read in "ptdata.csv", clean and format the data, and add the column agegroup, as we did in class.
3. Merge the two data frames into a new one (name it whatever you like). The final data frame should contain the following columns (you may rename them if you like, and you may also include other columns):
 1. id as integer
 2. sex as factor
 3. age as numeric with range 1 to 99
 4. agegroup as factor (5 year age bins, from 0 to 90)
 5. emerg as factor
 6. admdate as date
 7. weekdayname as factor
 8. los as numeric

Clarification on step 3 (Sept 16): the merge should leave you with information for each hospital admission. Don't include information for patients who were not admitted to hospital.

(Technically it would be cleaner to have three different scripts for these three cleaning/processing steps, but for convenience here we're including all steps in a single file.)

- Add enough comments so that someone who is familiar with R will be able to follow the logic of what you're doing.
- Check that the script runs correctly.
- Do not include extraneous commands like str or table—the only things in the file should be the commands necessary to clean the data; the assumption is that you've already checked that the input data files have the expected format.
- Give it a name of the form "YourName_datacleaning.R".