

Data visualization in R

Homework: In class you've gotten to know some of the main features of the patient admissions data, and practiced making various kinds of tables and graphs to summarize the data. Now pretend you are a consultant/analyst and put together a summary of the data, suitable for an analyst or a technically-oriented person at the client organization.

- You may work alone or with a partner; if you work with a partner, you will both get the same grade, so make sure you're both happy with how you divide up the work.
- Prepare an R markdown file named "YourName_PartnerName_report.rmd" and a pdf of the output named "YourName_PartnerName_report.pdf". [If for some reason you can't get pdf to work, html or Word formats are also fine.]
 - You must submit both the R Markdown and the output!
- Hide all R code snippets and warnings—do show output!
- The files must contain the following
 - A "data dictionary"—a brief summary of what the fields in the **original** data tables mean.
 - Make a separate dictionary for each original data file. Include columns for variable name, variable type, a description in words of what the variable is, and a summary of the values the variable takes. (min? max? number of blank or NA values?)
 - The code to generate a tidy data file from the raw input files (the content may be the same as your script from the in-class exercise—copy it over to your .rmd file).
 - Any tables or graphs that will help the reader understand what's important about the data. (Part of the assignment is to use your judgement about what's important or worth highlighting.)
 - The one table you **must** include is a table of the number of admissions by sex and 5 year age group, formatted with one column for M and one column for F, and age groups down the rows.
 - Each table or figure must have suitable axis labels and legends, and you must also give a brief explanation (one sentence) summarizing the main conclusions to be drawn from it. Make sure that your observations are justified by the data—it's fine to state the obvious but it's not fine to make false assertions.
- The output of the Markdown file must be formatted well. Tables containing numeric data from the original data files must be populated starting from a data frame, not by entering values "by hand". (It's fine to construct your data dictionaries cell-by-cell instead of from a data frame if you like.)
- Submit the two files via Connect by 11:59 pm on Sunday Sept. 25. Each group should submit only one copy of the files.