

Stat 133, Fall 2014

**Homework 3: Data Frames, Plotting**

**Due Thursday, Sept 25, by 11:55pm on bspace**

**Directions:** Turn in a plain text file containing the R commands you use to solve these problems and make these plots. Put your name in a comment at the top of the file, and add comments throughout to indicate problem numbers. Also submit a .pdf file that has all the plots for this HW. For most of the plots, I've done examples that you have in the lecture code that incorporate these ideas. For some techniques you will need to look in the help for the different functions/arguments, but all the functions you need either appear in lecture code or are suggested below. Style counts, please make the plots look nice (legends reasonably placed, descriptive axis labels and titles, etc.) You may make your plots look nice in other ways, but don't make them too messy.

1. Load in the data on infant birth weights using the command  
`load(url("http://www.stat.berkeley.edu/users/nolan/data/KaiserBabies.rda"))`  
(You will need to have an internet connection.) This will put a data frame called `infants` in your workspace.  
  
Make a histogram of the mothers weights (`infants$wt`). Use the argument `breaks` to make your histogram have EXACTLY 5 bars. (They can be different widths, which might help you get exactly 5 bars.) Change the title and label on the x axis to better labels which don't have a \$ in them. Use the functions `text()` and `arrows()` to put the word `max` in the plotting area of the histogram with an arrow pointing to the maximum value of mothers weight along the x axis.
2. Make a mosaicplot showing the relationship between education level of mother and smoking. Simplify the categories by changing `smoke` to have only three factors: `Never`, `Now`, and `Other` (includes all categories besides `Never` and `Now`). Also simplify education level of mother to have only two factors: `HS or Less` (include "Unknown") and `More than HS` (include "Trade"). Replace the columns in `infants` with these simpler forms. To avoid frustration, you might want to copy the `infants` data frame before changing those vectors.
3. Make a plot with `infants$gestation` on the x axis and `infants$bwt` on the y axis. Make the plot with the `pch` argument equal to 19 and with the points in the plot different colors depending on the different factors of `infants$smoke`. Also change the plotting characters, using two different plotting characters corresponding to the two education levels you made. Use transparent colors and change the size of the characters so that the density of the points is easier to see. Include a legend using the function `legend()`. Add horizontal lines going through the average infant birth weight for women who never smoked and the average infant birth weight for the women who smoke now. Make the colors of the horizontal lines consistent with the legend. Make the title and axes labels more descriptive.