HUDM 5026 - Introduction to Data Analysis and Graphics in R
HW 04 – Multivariate Plotting

Instructions.

- Use R Markdown to create an html document with the homework tasks.
- You are encouraged to discuss problems with classmates, but all work you submit must be your own.
- As always, any plots should have appropriate axis and overall labels.

1 Data

Data for this HW assignment come from a randomized experiment to study the efficacy of acupuncture for treating headaches. Results of the trial were published in the British Medical Journal in 2004. You may view the paper at the following link: `http://www.bmj.com/content/328/7442/744.full`. The data set includes 301 cases, 140 control (no acupuncture) and 161 treated (acupuncture). Participants were randomly assigned to groups. Variable names and descriptions are as follows:

- `age`; age in years
- `sex`; male = 0, female = 1
- `migraine`; diagnosis of migraines = 1, diagnosis of tension-type headaches = 0
- `chronicity`; number of years of headache disorder at baseline
- `acupuncturist`; ID for acupuncture provider
- `group`; acupuncture treatment group = 1, control group = 0
- `pk1`; headache severity rating at baseline
- `pk5`; headache severity rating 1 year later

Homework problems:

1. Import the data using `read_csv()` and call it `acu`. Note that the data have a header row.
2. Use base R to calculate the mean and sd for each variable in the data set using `apply()`.
3. Use the `by()` function to calculate the mean `pk5` score by treatment group. Are you comfortable making a conclusion about the efficacy of acupuncture based on these means? Why or why not?
4. Use base R to create a new variable called `diff` by calculating the difference score (post - pre) for headache severity rating.

5. Use base R to create a new variable called `remission` that is a 1 if a person's score dropped by 10 points or more and a 0 if not.
6. Use tidyverse functions to replicate (3) - (5), calling the variables `diff2` and `remission2`.
7. Create a scatterplot of baseline headache severity rating (horizontal axis) vs 1-year headache severity rating (vertical axis) using base R.
8. Use **ggplot2** to replicate (7).
9. In either base R or ggplot2, take your pick, incorporate information on treatment group and sex into the plot. Consider using color, point character, point size, or other graphical elements. Be sure to include a legend.
10. Try to replicate what you did in (9) with the other plotting platform. I.e., if you used base R, now try to replicate using ggplot (or vice versa).