Homework 1: Data Visualization

January 26, 2022

Answer the following questions in a .pdf or .docx, explaining all of your answers and putting any tables and figures in the document as necessary. When data is called for to answer applied questions, I will provide it in bblearn. Turn in your R code that created all of the tables and figures separately, and be sure that it runs from source in such a way that it loads the data and performs all the tests without me fiddling with it. Make sure to document your R source code using # comments if you want partial credit.

For these questions, we will visualize some data from Gelman, Hill and Vehtari (GHVevals.csv) on beauty assessments of college instructors to see if we can draw any tentative conclusions via eyeballing the data. I have also provided a .txt codebook that briefly describes the data. Please label figures intelligibly and not with raw variable names.

(Quick tip: To make these histograms and bar graphs work, you need to tell R that we have categorical variables by using the `factor(variablename)` command when you call the variable in ggplot. `minority`, for example, takes a value of 1 when it is a racial minority instructor and 0 when it is a non-minority and we need to communicate it is a not a continuous number to R by saying `factor(minority)` when we call that variable—otherwise some of fills and graphs won't run right.)

1. Let's get a rough sense of how average evaluation score is distributed across all of the instructor observations in our dataset. Create a histogram of course evaluations with bins with a .1 score size. Is the course evaluation data approximately normally distributed? Skewed in one direction or the other? Does the average evaluation generally center around an objectively "central" 3?

2. Now I want to make a histogram of course evaluations like question 1, but faceted by gender in adjacent graphs. Eyeballing the graphs, discuss what you see. Are there differences in gender these histograms? What do the graphs suggest about this sample? What would you tell an administrator assessing course eval scores in tenure packets for promotion based just on this data?

3. I want to know whether older faculty are assessed as more or less beautiful by the average student. Using the data provided, create a plot of the age of the instructor and their average beauty across six student raters (`btystdave`). Glancing at your finished plot, is there are relationship? Describe and discuss it.

4. I want to see if the average beauty assessment of the instructor has any influence on their course evaluations. Please plot the variables measuring these two things in the data together and then display and discuss the results. What is your overall assessment of any patterns, and what do they mean in plain language? What would you recommend to school administrators who actually cared about the validity of course evaluations?

5. We are curious if gender and beauty play an interactive role in course evaluations. You do not need to interpret the plot, as interpretations of interactions in graphs are difficult, but please create a plot of course evaluations and instructor beauty assessment whose observations are colored by their gender. Feel free to speculate on any trends you see in the data but these are not graded.

6. Let's pretend this is a random sample of instructors (it is not, there are duplicate observations of professors). Assess the distribution of racial minority instructor observations among the sample of instructor observations we have in the data. Please produce a stacked bar plot that shows the distribution tenure track and non-tenure track instructors and shades those bars by racial minority count. Based on this data, what is your recommendation to diversity-minded administrators seeking to hire in the academy? Is your recommendation different for hiring behavior for non-tenure track and tenure track positions?

7. Continuing our previous line of inquiry, let's look at gender balance in this sample across tenure track and non-tenure track instructors. Please produce a stacked bar plot that shows the distribution tenure track and non-tenure track instructors and shades those bars by gender count. Based on this data, what is your recommendation to gender balance-minded administrators seeking to hire in the academy? Is your recommendation different for hiring behavior for non-tenure track and tenure track positions?