

STAT 3280 Spring 2021 HW2

Due by the end of Sep 21, eastern time. Submit your homework by sending it to our GTA, Ruizhong Miao (rm9dd@virginia.edu), with the subject “STAT 3280-HW2: names”, where the “names” should be replaced by your last name(s) of the group. Each group only has to submit it once. Make sure you include everyone’s **name AND computing id on the first page**. **Missing any part of these will result in missing grades.** Please use a separate page for each problem. And the answer to each problem cannot be longer than one page (with reasonable font size, line space, margins etc.). You can explain how you did it in R by submitting your code with detailed explanations, but only include this part in an appendix. The GTA will not be guaranteed to look at your appendix, so make sure you explains things clearly in your main text. Notice that you are working on a visualization task. So, for each problem, make sure that your **plain language explanations should not exceed 1/2 of the paper in total** for each problem. The main results would be your figures. **You are required to use any software or packages for this homework**

Total points: 10.

1. (4 pts) The file “face-data.txt” contains the pixel values of four 64×64 images of faces, captured from different angles. For each image, the pixel values are stored in one column, in the order of vectorizing a 64×64 matrix (4096 pixel values). So, the data are stored as a 4096×4 matrix. Visualizing each of the 64×64 matrix will give one such image (actually, this is one of the most basic ways to store an image digitally). Visualize the four images in R using the **matrix visualization methods** covered in the lecture. Potentially, you may need some rotation or flipping to show the images in the correct orientation (clearly, it does not make sense if you have them upside down). Show the four faces in a **4×1** layout. You can pick up whatever coloring theme you want. (2 pt for correctly show each all the faces with clearness, 1 pt for correct orientation, 1 pt for correct layout). We will also use this data set for later sections.
2. (6 pts) An older teacher (in her 40’s) applied for a job in her local school system. She was not hired and did not even receive an interview. She claimed to be the victim of **age discrimination** and believed that the school preferred to hire younger teachers. If true, this would be against the law in the United States. Her lawyers obtained a data set “TeacherHires.csv” on applicants for teaching positions from the school files. You are supposed to do **exploratory visualization to analyze the data and give certain insights into the problem**. Although this has been partly processed, it is still very messy. I would do some necessary cleaning and exploration with you in class. However, you will need to clean it further to ensure it is usable for your own analysis. Conduct an exploratory analysis: **is there evidence of age discrimination in the interviewing and/or hiring of teachers?** Use your visualization analysis to answer this question. It is an open problem, clearly. Give your answer and justify your conclusion with your

visualization analysis. (2 pts for the basic presentation of the facts from the data, 2 pts for effective and clear visualization, 2 pts for insightful and convincing analysis.)

For those of you who have difficulty starting the analysis, one suggestion is: start with asking simple and concrete questions, then figure out how to answer that question, and based on it, raise your next question with your final target in mind. You can explore any variables in the dataset that interest you or that might be important. For example, if hired teachers have higher GPA scores, this could be important in your analysis. If you prefer to treat age as “continuous”, that is fine, too. You do not have to include your data cleaning step in your submission, but it will be crucial to ensure your work quality.