

STAT 3280 Fall 2021 HW6

Due by the end of Oct 31, eastern time. Submit your homework by sending it to our GTA, Ruizhong Miao (rm9dd@virginia.edu), with the subject “STAT 3280-HW6: 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 two pages (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 can use any software or packages for this homework.**

1. (11 pts) In the folder “statisticians”, you can find the data about statisticians’ publications in 4 journals during 2003-2012. Look at the ReadMe file in the folder to understand the data set. You can also explore the paper at <https://arxiv.org/abs/1410.2840> for a detailed analysis if you want. We will explore the data set in class and test basic visualizations. You are supposed to further extend the analysis as below:
 1. Visualize the abstract word cloud over three period (2003-2005, 2006-2008, 2009-2012). Do you observe any trend? (2 pts)
 2. Ego network exploring: among three famous statisticians “Peter_J_Bickel”, “Jian-qing_Fan”, “David_Dunson”, who is the one with the largest number of collaborators? What are their numbers of collaborators? (1 pts)
 3. Pick one of the three persons for the following analysis:
 - (a) Visualize the person’s collaboration network change over the three periods above (the persons and all of his collaborators only, as an induced network). Comment or visualize certain expansion pattern. Make sure you visualize the three networks in the three periods in the same reasonable layout, even though they have different nodes. You need to figure out a way to do this. (Network visualization over time 2 pts, reasonable pattern summary or visualization 2 pts)
 - (b) Visualize the word cloud from his/her articles in the data set over the three periods. Is the trend in for this person similar to the global trend? (Visualization 2 pts, comparison 1 pt)
 - (c) Use Google Scholar, find all the paper titles of this person. Again, visual the word cloud from the titles over time, for the person’s whole academic career. Does the period of 2003-2012 seems to match what you observe from (c)? (Hint:

you do not have to segment the data by each year, just check the word cloud before 2003, and the cloud after 2012.) (1 pt)