Statistics 5014: Homework 8

Due In Class November 1, 9am

2017-10-25

For each assignment, turn in by the due date/time. Late assignments must be arranged prior to submission. In every case, assignments are to be typed neatly using proper English in Markdown.

This week, we spoke about text mining and sentiment analysis. Most of the material came from http://tidytextmining.com/. While this is not the only way to mine textual data, it fits nicely into the tidy process we used in our search for Reproducible Analysis.

Problem 1 (GitHub 2 pts + 2 Style pts)

This week we will change gears a little on GitHub. Today we start making homework something we can use to highlight our professionalism and abilities. You will continue to retrieve the lectures and homeworks from my GitHub (anyway you like), but you will:

1. create a new GitHub Repository (online)
2. invite me as a collaborator (settings –> collaborators)
3. setup an Rproject pointing to this new repository
4. create a new RNotebook file within the project folder
5. save the notebook HW7_lastname_firstname

Note that this homework will be graded for professionalism as well as your ability to solve the problems (passing is ≥ 7. Remember, reproducibility requires weaving text (appropriate for a reader), code (commented) and necessary output as a compendium for what was done.

Problem 2: Class data

Load munge and create a story on the class dataset: survey_data.txt in this class repo.

Problem 3: Case study

In the http://tidytextmining.com writeup, there are 3 case studies. Your task is to make a new one. You can take any one (or more) of these and change the dataset, combine them to do something new, etc. Up to you!!

Problem 4

Get an account at arc.vt.edu. Make sure you can log in.

Problem 5

Push your homework to YOUR GitHub.

This is YOUR GitHub Repository. Save what you want, there is a limit to how big a repo can be, so be a little sparing on what you put up there. Please save the file you would like me to read as: HW#_lastname_firstname.html. I am not going to go through your Markdown to try to figure out what you were doing.

Optional preperation for next class:

Read up on regular expressions.