DS 5110 Project - Fall 2018

Overview

The purpose of the project is to gain hands-on experience working on real-life data, in a collaborative team. Although a majority of the data management and processing for the project should be done in R using methods learned in class, you are free to incorporate techniques that have not been covered in lecture.

The project should be done in groups of 3 - 5 class members. The project is worth a total of 300 pts (30% of class grade) and includes a project proposal (50 pts), an in-class group presentation (100 pts), and a final written report (150 pts). Each group should work independently on their project, but may collaborate on technical questions via Piazza. *Make sure to proofread your proposal, slides, and final report – it should appear polished and professional.*

Projects with a very high degree of similarity with other past or current students' work, or with other existing work (e.g., found online), will be considered plagiarism. In such instance, all group members will receive a zero on the project, and will be reported to the university.

Suggestions

You are encouraged to think outside-the-box in your project proposals. A straightforward way to approach the project is to select a dataset, and perform an in-depth analysis of it using methods covered in class. This is the most common type of project. Alternatively, you may choose to develop a tool for the management and processing of a particular genre of data (e.g., multimedia archives, or spatio-temporal data, etc.), or which enables or streamlines a particular aspect of data analysis (e.g, reproducibility, or visualization, or model-building, etc.). You are encouraged to talk to the instructor about potential project ideas before submitting the proposal. All projects must demonstrate results on real-life data.

Group members: due Friday, October 19

Please form groups of 3 - 5 class members who will work together on the project. You may browse class members' mini-posters on Piazza for ideas of what types of data interest your classmates. You may reach out on Piazza when seeking out other project group members.

On the Midterm Exam, there will be a section to fill in the full names of all members of your project group.

Project proposal: due Tuesday, October 30

The project proposal should be submitted as a PDF, uploaded as a ***private*** *note* on Piazza in the *project* folder with the title "[proposal] all group members' last names". Include in the body of the note: (1) the title of the project and (2) a summary of the proposal (you may copy-paste the summary section from the proposal itself). Each group only needs to submit one proposal.

The project proposal should be no more than one page, not including any references or figures. It does *not* need to contain code or be generated from R Markdown. The proposal should contain the following:

1. **Title:** A descriptive title of the project
2. **Authors:** List all group members' full names
3. **Summary:** 2-3 paragraphs summarizing the problem you wish to solve, including a description of the dataset, and a very brief, non-technical description of your proposed methods.
4. **Proposed plan of research:** 2-3 paragraphs describing the methods you will use to solve the problem. These may be processing, visualization, and analytic methods already discussed in class, or it may be

your design for a software tool for working with a particular type of data or solving a common data analysis challenge, etc.

5. **Preliminary results:** 1 paragraph describing any preliminary results you have. This should at least include confirmation that you are able to load the dataset satisfactorily.
6. **References:** Cite any references used in the proposal, including any sources of data and associated publications. Use a consistent format and numbering scheme.

Project presentation: Tuesday, November 27 - December 7

During the last two weeks of classes, each group will make a short in-class presentation of their project. You may use the project report format (below) as a guideline for formatting the presentation. All group members are expected to speak during the in-class presentation.

The slides for the presentation should be submitted by the Sunday before your group's presentation as a PDF, PowerPoint, or Keynote file, uploaded as a ***public*** *note* on Piazza in the *project* folder with the title "[presentation] all group members' last names". Include in the body of the note: (1) the title of the project and (2) a summary of the report (you may copy-paste the summary section from your report or proposal). Each group only needs to submit their slides once.

Project report: due Sunday, December 9

The project proposal should be submitted as a PDF, uploaded as a ***private*** *note* on Piazza in the *project* folder with the title "[project report] all group members' last names". Include in the body of the note: (1) the title of the project and (2) a summary of the report (you may copy-paste the summary section from the report itself). Each group only needs to submit one final report.

The project report should be no more than eight pages, not including references or appendices. The main text should not include code unless relevant for demonstration purposes and does *not* need to be generated from R Markdown. The report should contain the following:

1. **Title:** A descriptive title of the project
2. **Authors:** List all group members' full names
3. **Summary:** Summarize the background of the project (i.e., what is the problem to be solved or the question to be answer), any related work, a description of the data, and a brief, non-technical description of your methods and results.
4. **Methods:** A technical description of the methods you used in the project. This may be a description of the steps of data tidying, transformation, and modeling you performed, or it may be a description of the implementation and functionality of a software tool, whichever is appropriate for your project.
5. **Results:** All projects must demonstrate results on real-life data. Show the most relevant results from your analysis, using appropriate figures or tables. Describe only the most interesting and useful results.
6. **Discussion:** What can we learn from the results of this project? What would you do differently next time, or what would you do to improve it?
7. **Statement of contributions:** List the full names of the authors and how each member contributed to the completion of the project.
8. **References:** Cite any references used in the report, including any sources of data and associated publications. Use a consistent format and numbering scheme
9. **Appendix:** Include relevant code here, along with any other relevant technical details or supplementary figures that do not belong in the other sections.