

Statistics 6021

Final project

The body of the project – proposal, report, and presentation – will be completed in assigned teams of three or four.

The data for the project will be provided by the “Stack Exchange Data Explorer” <https://data.stackexchange.com/stats/query/new>. You are free to query whatever you like, but the “Posts” table will be provided to you as `nondeletedposts.csv`. It was obtained with the command `select * from Posts;`. There are many other tables such as “Users,” “Closed”, etc. You can also `merge` data sets on columns that they have in common. For all the information you need about this data set, please see <https://meta.stackexchange.com/questions/2677/database-schema-documentation-for-the-public-data-dump-and-sede>.

The subject of the project will not be assigned, instead it will be chosen collaboratively by the members of each team.

If you are unsure about whether the topic of your presentation will “work,” please come talk to me. Each project should feature:

- Clear central analytic goal(s) and/or question(s) to answer – the more interesting and challenging, the better.
- Analytic methods developed in this course.
- A substantial computational component, required for data cleaning prior to analysis or for the actual analysis (or both).

Component 1: Report

Each team will prepare a project report. This report should be no more than 10 pages and should include the following:

- An executive summary that describes the goals/questions of the project, the nature of the data used in the analysis, and the results of the analysis, including any recommendations. (Maximum: 1 page)
- A detailed description of the data, the analysis, and the conclusions reached, making use of graphics as appropriate.
- Correct grammar, clear explanations, and professional presentation.

The audience of the report is a client who is not a statistician and a statistician to whom the client may give the report for a second opinion.

The printed project report is due before class on **Thursday, August 1st at 9am**. Your annotated R code should be submitted to Collab by the beginning of class on the same day.

Component 2: Presentation

Each team will give a 15-minute presentation. This presentation should be designed to be understandable by anyone familiar with the course material, but who has not read the project report. Each team is free to organize who talks about which topics during the presentation. Not everyone need talk, but all team members should be active contributors to the presentation materials.

Presentations will take place on **Thursday, August 1st** and **Thursday, August 2nd** during the regularly-schedule class time in our classroom. You must be present for all presentations unless you have prior permission due to a schedule conflict. Each team should have their presentations be prepared to be **quickly** accessed. The presentation slides should be submitted to Collab by the beginning of class on the same day.

The slides should also be submitted to Collab before **Thursday, August 1st at 9am**.