FYE Stats/Sports Project Requirements

Fall 2018

Requirements

- Work in groups of 2

- You may use data from the internet, data that you've collected yourself, or data that you may have used for another class. However, this project is meant to be different than anything you've done before.

Skills

- RMarkdown

- Data cleaning and (possibly) data manipulation

- Data visualization including histograms, scatter plots, density curves

- Statistical summaries of data, including correlation/correlation matrices, Stein's estimators, mean absolute errors

- Writing and presentation ability

- Summary statistics including means/medians, ranges, min/max, etc

Deadlines

- Revised proposal 11/30: a one page introduction which (i) explores your question(s) of interest, (ii) an explanation of your data, including a description of each variable and the desired sample size and (iii) who your work will generalize too (i.e, why should *other* people be interested in this work). The major challenge is to motivate your peers – why should they care about the project? Hint: why do you care about the project?

- Initial working version of your data 11/30: Use google sheets, as per discussions in class

- Presentations 12/12 - we do these on the last day of class

- Technical reports 12/20 - due at midnight at december 19th. This will be a continuation of your proposals to include your graphics and conclusions, as well as any edits that are suggested during the presentation stage

Presentation details

- Each group member must be present

- Each group member must share evenly in the presentation

- Each group will have 8 minutes to present, plus 3-4 minutes of questions and answers with the audience

- Your presentation grade will be graded based on how well you work as a group and describe your data during their presentation. For example, transitions between graphs, group members, etc should be smooth. Individuals should refer to other materials presented by other members of their group when appropriate.

Project details

- Precise bullet points: why is your project important?
- 3-4 high quality plots. More work may be required for groups of three.
- Summary statistics about your data: what are interesting findings?
- Correlation matrix/evaluation about your data
- Stein's estimators (if applicable)

Evaluation of presentation (10 points each)

- Creativity (in design)
- Visualization details
- Depth/appropriateness of work
- Accuracy of presentation
- Presentation (volume, presence, eye contact, etc)

Technical report details

1. Introduction (1-2 pages).

- Why is your problem important?
- What other work has been done? Hint: use google, but don't copy and paste
- Why are your questions of interest actually interesting?
- What does your data look like? Where does it come from? How was it obtained? What are the variable types and units? What were the steps you took in cleaning the data?

2. Descriptions of visualizations (page length will vary).

A summary of each graphic should contain

- the graphic itself (or an example version of the graphic if the graphic is interactive)
- one paragraph (3-5 sentences) summarizing the key takeaways of graphic and any interesting features / associations / relationships / etc that the user should understand
- one paragraph (3-5 sentences) describing the tools used to create the graphic (e.g. `ggplot` features, other plotting tools, etc), the coding techniques necessary to manipulate the data in order to create the graphic (e.g. subsetting, etc), and any graphical choices that were made (e.g. bandwidth / bin width choices, etc)

2. Statistical work

- Summary statistics and calculations

- Correlation matrices/Stein's estimators/etc

3. Conclusions (1 page). Practical findings and conclusions. Weaknesses in design. Other variables that you'd want to look at?

This is a formal report. Your writing should be professional and free of spelling/grammar errors. Your graphs should be high-quality, with titles, axis labels, legends, etc.

Notes

- You should show **all** code for your technical report, but that your report to the class should only show code where needed (e.g., if you are doing something new)

- Please cite your sources, either using traditional format (Lopez, 2013) or links to URLs.

- Consider the project requirements when thinking about what data you'd like to use. If you have a data set with 2 or 3 variables, it is likely insufficient for meeting the requirements above. However, you can certainly merge existing data sets together (which I can help with).

- Each group should submit a single .Rmd file and a single .html file containing their report. Each group should submit one report only.

- One grade will be awarded to the entire group, although your professor reserves the rate to change group grades if individual participation varies.

- Come up with a clever title!

-