Statistics 585X - Project Description - Spring 2014

The purpose of this project is to provide an opportunity for class members to show off the data technology skills that you have learned in this class. Select a topic that is of interest you, where there is data publicly available in a fairly unprocessed form. This is targeted to be an individual project, but with special exceptions two people may work together. In the case of a two person team, the individual contributions need to be explained. Here are some examples:

- Did flight patterns changed after 9/11? For this topic we would pull different subsets of data from the airline database, using SQL.
- How do player match statistics compare with their success in tennis tournaments? Would use web scraping to pull data from grand slam tennis tournaments, merge with data pulled from other statistics sites.
- What is the usual range of great white sharks. Where do the sharks roam? (http://www.ocearch.org/)
- What are people tweeting about this week?
- How do the movement patterns vary from one team to another in professional basketball?

It can also be a programming activity, for example:

- Draft a Shiny app for teaching elementary statistics using R.
- Provide an interface to a large biological (eg RNA-Seq) data set than enables the exploring of the effect of dispersion estimation on statistical significance.

The key pieces that should be seen in the project work are:

- *Data technologies:* Data needs to be collated from one or more sources, using web scraping, or pulling from a data base, ...
- *Reproducible research:* need to use knitr to produce the report, all code needed to accomplish task needs to be in the report.

Deadlines

Wednesday Mar 26: Draft of plan for the project to the instructors, including topic, questions of interest, motivation, what data is available, and how the data is to be collected, collated, merged, processed. This should be a markdown document, loaded up to your own github repo, with the link emailed to the instructors.

Wednesday Apr 9: Draft of report due. Also uploaded to github.

Wednesday Apr 23: Report due, also on github.

Starting Wednesday Apr 23: Presentation of about 10 minutes, allowing for some questions, of project to the class