

Stat 133, Fall '05
Final Project
Due 5 p.m. Tuesday 13 Dec

Dynamic Interactive Documents and the Communication of Statistical Analyses

The main goals of this project are to give you practice with the data analysis process, which includes the following aspects:

1. Analyze data and capture your thought process in doing this, including trying out various alternative approaches to the data analysis. A set of guided investigations will be provided for 6 different data sets. Each group chooses one of these to analyze. No more than 3 groups may study the same problem.
2. Back up your analysis with a computer experiment such as a Monte Carlo simulation, cross-validation of a model, or bootstrapping the distribution of an estimate. Suggestions for conducting the computer experiment will be provided with each guided investigation.
3. Summarize your results in a report that presents your findings. This report will include graphics, tables of numbers, and text.
4. Dynamically generate different versions of your report based on user input. That is, the document is to be generated programmatically where at least one of the graphics and one of the tables are to be generated based on input parameters.
5. Add an interactive component to the document that allows the reader to conduct a what-if investigation of your analysis. An interactive component may be a slider, select box, or button that when modified would recalculate part of the document by changing a plot or table.

We will study three different approaches for creating these interactive and dynamic documents. Depending on the data analysis project one approach may be more suitable than another. Groups may choose between the available techniques, but at least two groups must work on any one technique. All of these techniques will be introduced in class. The basic approach for each techniques follows:

1. Excel - This project is for Windows users only. The document that you will write will be a user's guide/lab manual, and the text, images, and tables will be placed in an Excel spreadsheet. That is, you will use the RDCOM packages and write a function in R to dynamically generate an Excel workbook. Then in the final stage of the project, you will add an interactive component to the workbook such as a slider that when moved will call R to change the spreadsheet.
2. Web server - This project entails creating an HTML form that takes user input via radio buttons, check boxes, and buttons, and processes this input in Perl to create a call to R. Within R an HTML page will be created that is then displayed on the web. A template for the Perl code will be provided.
3. HTML widget and gui - Like the previous project, you will use R to create an HTML page dynamically based on input parameters. The interactive component will be created using Gtk from within R with the RGtk package. This interactive component will be added into the HTML page. Note that the HTML page with the embedded interactive component can only be displayed in a linux/uninx environment. However, the dynamically generated web page can be created and displayed in linux, mac, and windows, and the same is true for the interactive component as a stand alone widget but it requires installation of several pieces of software.

The projects will be structured as follows:

- You may work in groups of 3 or 4 – no more, no less.
- The project will be divided into three stages with intermediate deadlines.
 - Stage 1: Complete the data analysis. Due Nov 30
 - Stage 2: Write and dynamically generate report. Due Dec 6
 - Stage 3: Add an interactive component. Due Dec 13.
- Each member of the group must contribute to all stages of the project.
- Before completing each stage, your group must meet for 15 minutes with Professor Nolan for feedback. Sign up sheets will be posted.