

# Project 1

*Oct. 13, 2011, 2:25 pm*

## Task 1:

**(Independent)** Choose a problem related to the work we've done so far in the course. This is somewhat open-ended, so we'll give you several options.

- i. Perform the same sort of analyses we did for the senate, but apply them to the House for some session;
- ii. Perform the same analyses for the senate for two consecutive years, and then create a meaningful comparison of the two. What will you do with senators whose terms end? With freshman senators? How will you compare one senator's performance in one year to his/her performance in the next? You'll need to decide all these before you start.
- iii. Attack some political question using the methods we've discussed so far. For example, are the Blue Dog democrats more similar to (a) other democrats or (b) other southerners?
- iv. Apply these methods to some completely different domain. For instance, maybe you have a list of 300 kinds of mushrooms, each with yes/no descriptions of 50 properties (grows on trees? is red? has spots? etc.) You'd like to cluster them into groups, and see whether the resulting clusters have biological significance.

The most difficult aspect of this option could very well be getting the data into Excel. The Senate data was in XML format, which made it easy to import. If you choose this option, make sure you include in your description (due October 3rd) where you plan to pull the data from. The TAs will be available to help you import your data; but you should be mindful of how "well-formed" it is.

## Handin

## Task 2:

By **6:00 pm on 10/2** email `cs0931tas@cs.brown.edu` a **very** brief description of your intended project, describing your goal, your data source, and your plan, each in one or two sentences (see an example [here](#)). We will help you do the sanity check and if anything alarming shows up we will contact you and give suggestions so that everyone proposes a feasible project.

By **2:25 pm on 10/4** email `cs0931tas@cs.brown.edu` a more detailed description of your intended project, and how you intend to go about completing it. This should include any needed resources (“I’m going to need the four buttons that move rows and columns around” or “I need help with downloading 500 XML files from webpages”) and potential roadblocks.

By **2:25 pm on 10/13**, email the TAs a zip file containing all work associated with your project.

This should include:

- (a) A file named “README” which can be a text file/Word document/PDF file that contains:
  - A description of your project which makes it clear which spreadsheet shows the final results of your data gathering and manipulation.
  - An short description of how you manipulated your data. Make sure to reference any intermediate spreadsheets that the TAs should look at. You will want to keep intermediate spreadsheets so that your grader can follow the different steps you took in coming to your solution.
  - An analysis and reflection of your findings, including a description of what you could have done differently and any problems you ran into.
- (b) A link to a web page that displays your results in some attractive form (which may have a lot of overlap with the first item)
- (c) Any Excel spreadsheets or other things you used in producing your results.