CS10 Fa13 Prj1 Proposal

This is a chance for your GSI to hear some of the basic ideas you are planning on pursing on this project.

Feel free to review the spec: http://inst.eecs.berkeley.edu/~cs10/fa13/proj/BitsToInformationToKnowledge.pdf

Partner 1's Name *

Pierce Vollucci

Partner 1's SID *

12345678

Partner 1's Email *

cool@berkeley.edu

Partner 2's Name *

Dan Garcia

Partner 2's SID *

98765432

Partner 2's Email *

awesome@berkeley.edu

Partner 3's Name

Partner 3's SID

Partner 3's Email

Discussion GSI *
Select the GSI whose discussion you attend.

- ○ Ian
- ○ Jon
- ○ Max
- ○ Michael
- ● Peter
- ○ Samir
- ○ Zack

What are 3-5 potential questions you are considering for your project?

Question 1 *

Do Republicans and Democrats always vote with their own?

Question 2 *

What type of issues have votes from: all Republicans, all Democrats, or both?

Question 3 *

Is the style of the name of the item being voted on is highly correlated with the party that votes on it?

Question 4

Question 5

What are 1-3 data sets you are considering looking at to answer these questions?

Paste the URLs in the spaces provided.

Data Set 1 *

http://archive.ics.uci.edu/ml/machine-learning-databases/voting-records/

Data Set 2

Data Set 3

What techniques from what you've learned in lecture and your readings might you use on your data to help answer your questions? *

For question 1, we were hoping to compare trends of who voted for which bill, doing a scatter plot with bills on the independent axis and congressmen on the dependent axis.
For question 2, we were planning on plotting the vote types for each bill labeling one party as 1 and the other party as -1 and see how close to either -1, 0, or 1 each issue got.
For question 3, we would use some of the word analysis techniques we learned in the Spam lab to

draw associations between certain types of words (such as child or war related terms) and correlate them to groups of voters.

Explain how/what are you planning to do to the data set using Snap! or other tools. *

We were hoping to use the tools from the first part of the data lab to help with our scatter plot visualization only with a dot with 2 different costumes, red and blue for the two different parties.
Then we were hopping to use the tools from the spam lab to help with our last question in looking for word correlation, possibly with a list of suspected influential words.

What do you propose to have done by the 'progress report' half-way part? *

We propose to have the visualizations of questions 1 and 2 done and hopefully begin to draw some conclusions on those questions.