# POL 345: Quantitative Analysis and Politics

## Precept 3 (Week 4)

In this precept, you will learn the following new materials:

- Using `abs()` to calculate absolute value

## Predicting Election Results Using Polls

Last year, we built a simple regression model of election prediction using the opinion polls for the New Jersey governor's race where Chris Christie defeated the incumbent Jon Corzine. The dataset **newjersey.csv**, which is available at Blackboard, contains the following variables; `Pollster` is the polling company, `Dates` represents the dates the poll was conducted, `DaysBefore` stands for the number of days the poll was taken before the election, `Undecided` is the proportion of undecided, and `Christie`, `Corzine`, `Daggett`, and `Other` represent the proportion of respondents who expressed support for Christie, Corzine, Daggett, and others in the poll, respectively. The actual election result was as follows: 48.5% (Christie) 44.9% (Corzine), and 5.8% (Daggett).

1. Calculate the percentage of the two-party vote going to Christie and generate a scatter plot of this variable. Using the election results provided above, calculate the final percentage of the two-party vote going to Christie and add as a separate symbol. Make sure to adjust the `xlim` argument as necessary to have time progressing from left (oldest polls) to right (most recent). Add a dashed vertical line for the day of the election – whatever it takes to help your figure aid the reader in deciphering the "story" quickly. What pattern do you observe?

2. We wish to forecast the likely winner from poll results in two ways. First, we base our forecast on all polls in the data set. Second, we use only polls conducted within 100 days of the election. Create a variable for the lead in the polls of Christie over Corzine (using raw—not two-party—support). Produce a scatter plot for the lead in the polls, but display in your plot only those polls that occurred within 180 days of the election. On this scatter plot include two best-fit lines, one for Christie's lead in the polls using all polls in the full dataset, and another best-fit line using just those recent polls contained in the subsetted dataset. Finally, add the actual election result to the plot.

3. Using each of the two best-fit lines obtained in the previous question, forecast the winner and his margin of victory (that is, obtain one set of predictions for each line). Given the election result, which model did better? Note that enclosing a numerical object within `abs()` will calculate its absolute value.