UC Berkeley
Python class
AY250

# Homework 5

## Data(de)basing the 2012 Election

In this assignment we will create a database to analyze event prediction data pertaining to three ongoing elections. We will look at the **2012 Republican Presidential Nominee**, the **2012 Presidential Election**, and **2012 Republican Vice-Presidential (VP) Nominee**

a) Create a database table called "races" and populate it each race's name, election date, and data URL (below)

b) Download prediction data for the three races from intrade.com for all named candidates:

http://www.intrade.com/v4/markets/?eventId=84328
http://www.intrade.com/v4/markets/?eventId=84326
http://www.intrade.com/v4/markets/?eventId=90482

- Use BeautifulSoup to automate this process. To access the prediction data table, click on an event's page and then follow the "Advanced charts" link. The "Download values" link here provides the csv file.

https://data.intrade.com/graphing/jsp/closingPricesForm.jsp?contractId=743554&tradeURL=https://www.intrade.com

**Show:**
- (•) Closing prices
- ( ) Time and Sales

**Time Period**
Lifetime

**Chart Size**
Large
[x] Show Volume Bars

**Chart Type**
Closing Price

**Price Band**
None

**Moving Averages**
None 10
None 25

**Technical Indicators**
None
None
None
None

Update Chart

Download values

## Closing prices chart

# Homework 5

## Databasing the 2012 Election

c) Create another table in your database called "candidates" and write a program to automatically populate it with biographical information about the candidates (use BeautifulSoup to parse wikipedia pages). Include at least home town, home state, party affiliation, birth date, and a link to a local file containing a photographic portrait of the candidate.

d) Create a database table called "predictions" and populate it with the prediction data (from b). You should include date, price, and volume. For each row of prediction entry, create a row in the table with two additional foreign key columns indicating to which candidate and race the prediction data is related.

# Homework 5

e) Use your database to plot as a function of time the probability of a candidate with home state north or south of the Mason-Dixon Line winning each race.

f) In an efficient market P(Obama wins) = 1 - Sum P(repub wins) for all time. For the presidential election, find a few dates over the past year where this is farthest from being true. What was happening on those dates? What can you say about the efficiency of such trading markets?

Hints:

- Steps a-d need not be done in the order given. It might be easier to create the tables first then populate the DB as you crawl through the web.
- For f) it may be advantageous to interpolate sparse trading data to a finer grid so you can more easily add data across candidates. You might play with pandas, which is a good package for timeseries manipulation.
- You might also play with sqlalchemy, instead of using sqlite3 modules.