UC Berkeley
Python class
AY250

Homework 6

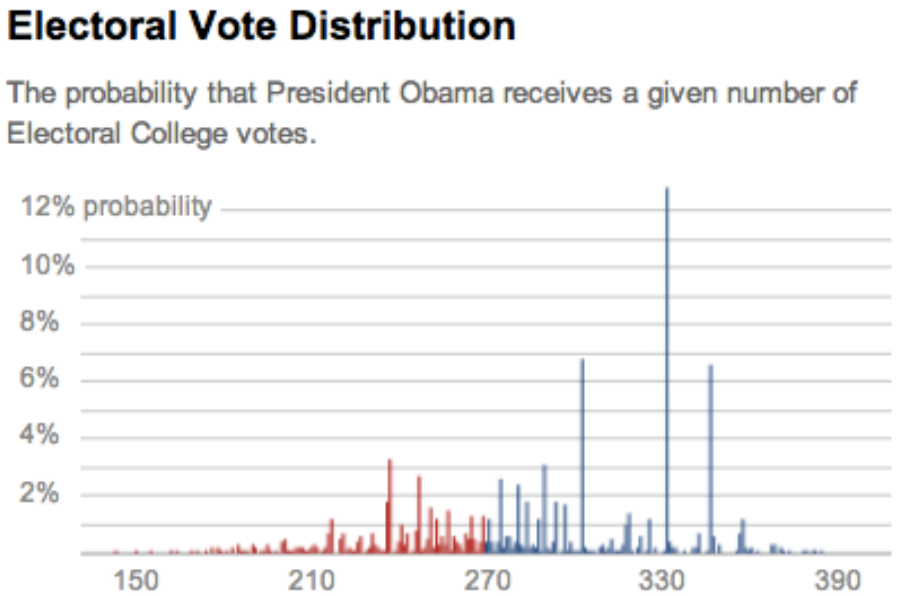
Data(de)basing the 2012 Election

In this assignment we will create a database to analyze prediction data for the **2012 Presidential Election**

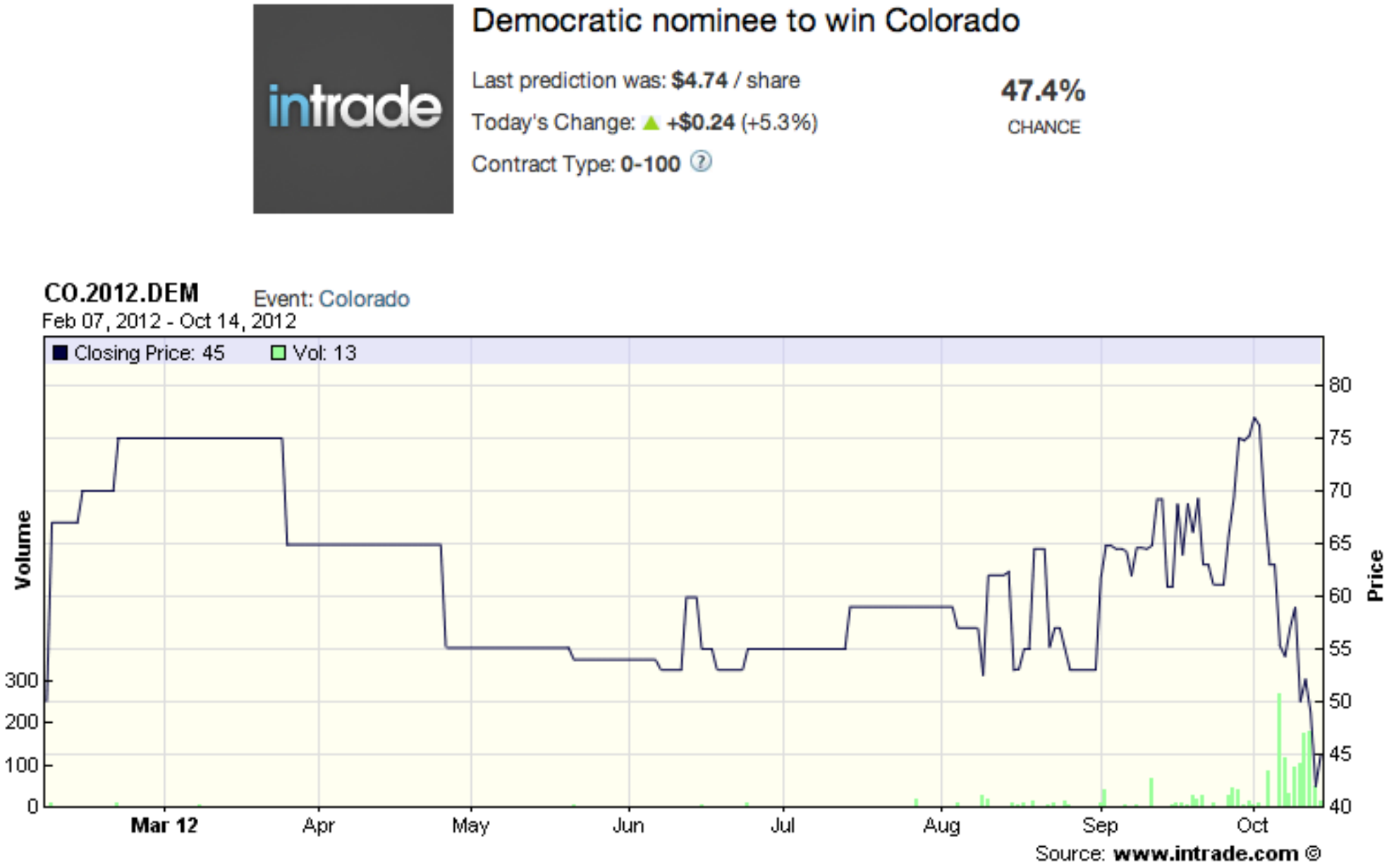
The goal is to pull all of the prediction data for each state and the general election itself into a local database to determine:

a) how closely does the general election price match a prediction calculated from the individual state prices?

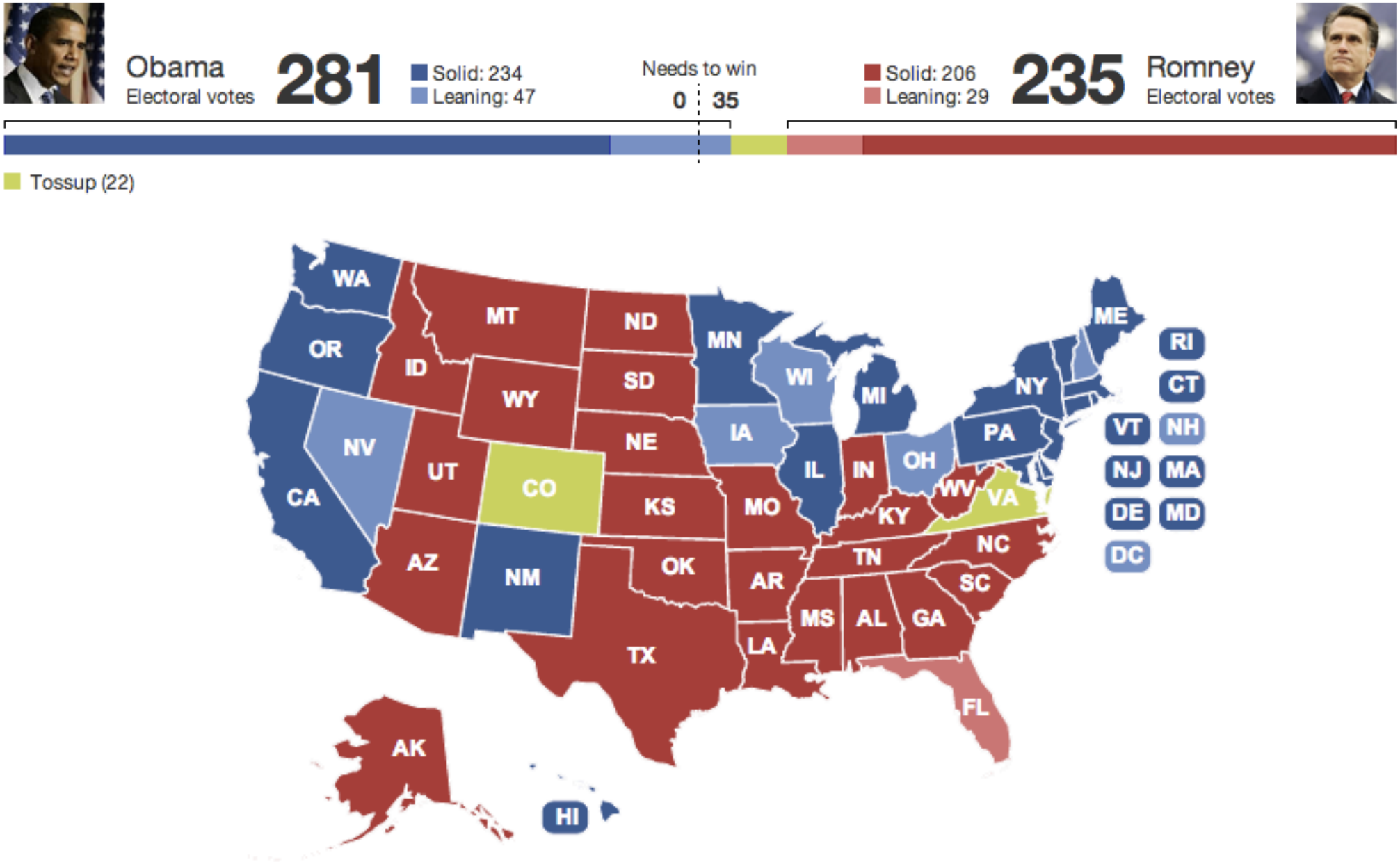
b) how closely do the intrade predictions match Nate Silver's?



http://fivethirtyeight.blogs.nytimes.com/



Interpret price as probability in percent



http://www.intrade.com/v4/misc/electoral-map/#

a) create a table called "election" and populate it with state name, electoral votes, and intrade contract id

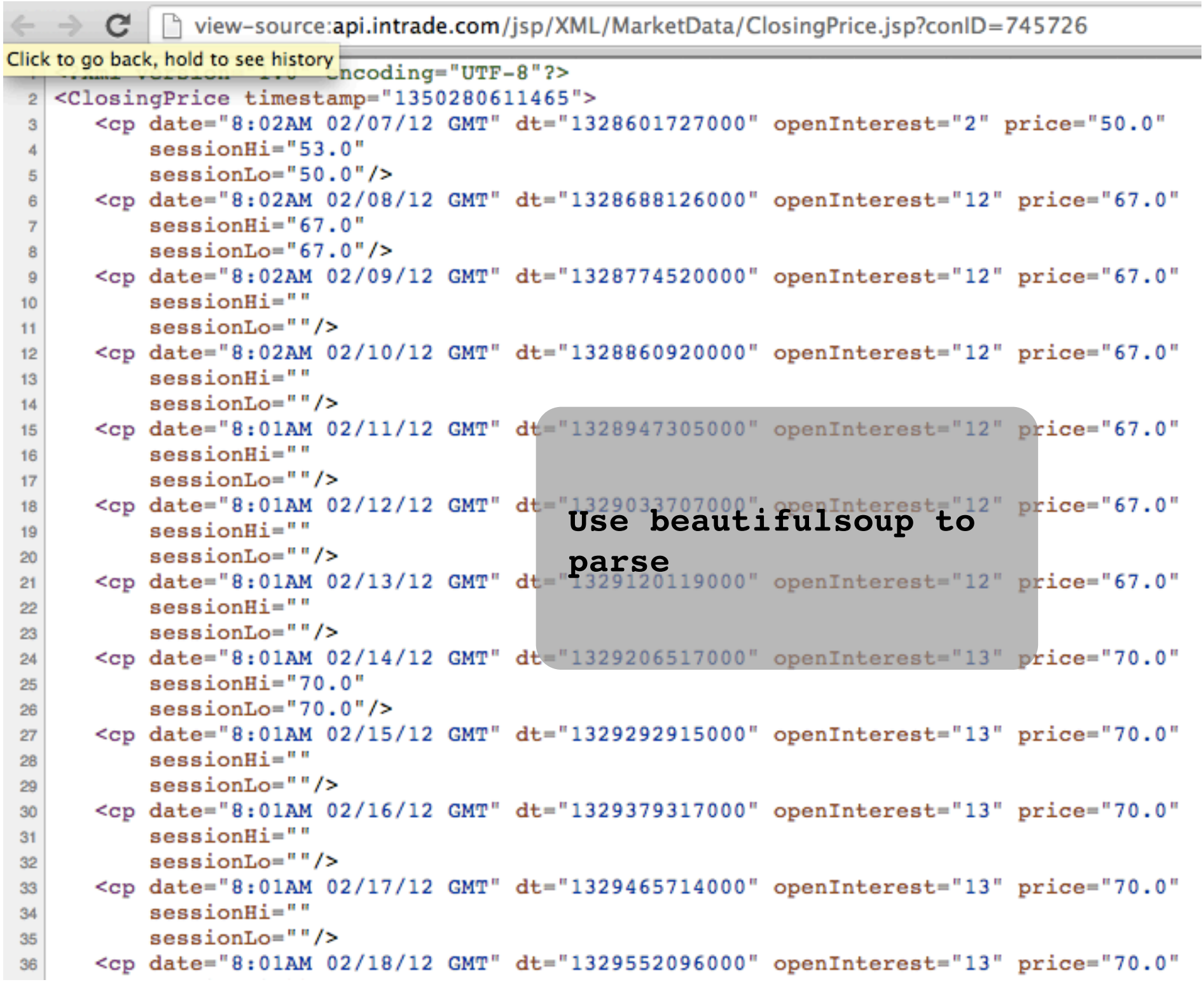
State	contractID	electoral
Colorado	745726	9
California	745723	55
...	...	...
general	743474	

see http://en.wikipedia.org/wiki/Electoral_College_(United_States)#Current_electoral_vote_distribution

b) create a table called "prediction" and populate it with all of the intrade closing data for each state and for the general election itself

http://api.intrade.com/jsp/XML/MarketData/ClosingPrice.jsp?conID=745726

see http://www.intrade.com/v4/misc/help/api/


view-source:api.intrade.com/jsp/XML/MarketData/ClosingPrice.jsp?conID=745726
Click to go back, hold to see history
encoding="UTF-8"?>
<ClosingPrice timestamp="1350280611465">
<cp date="8:02AM 02/07/12 GMT" dt="1328601727000" openInterest="2" price="50.0"
sessionHi="53.0"
sessionLo="50.0"/>
<cp date="8:02AM 02/08/12 GMT" dt="1328688126000" openInterest="12" price="67.0"
sessionHi="67.0"
sessionLo="67.0"/>
<cp date="8:02AM 02/09/12 GMT" dt="1328774520000" openInterest="12" price="67.0"
sessionHi=""
sessionLo=""/>
<cp date="8:02AM 02/10/12 GMT" dt="1328860920000" openInterest="12" price="67.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/11/12 GMT" dt="1328947305000" openInterest="12" price="67.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/12/12 GMT" dt="1329033707000" openInterest="12" price="67.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/13/12 GMT" dt="1329120119000" openInterest="12" price="67.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/14/12 GMT" dt="1329206517000" openInterest="13" price="70.0"
sessionHi="70.0"
sessionLo="70.0"/>
<cp date="8:01AM 02/15/12 GMT" dt="1329292915000" openInterest="13" price="70.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/16/12 GMT" dt="1329379317000" openInterest="13" price="70.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/17/12 GMT" dt="1329465714000" openInterest="13" price="70.0"
sessionHi=""
sessionLo=""/>
<cp date="8:01AM 02/18/12 GMT" dt="1329552096000" openInterest="13" price="70.0"
Use beautifulsoup to parse

c) simulate 10,000 elections using the prediction probabilities today. Plot the histogram of electoral college votes for Obama. Compare the fraction of elections won by Obama to today's general election probability in intrade and at fivethirtyeight.
Repeat this for predictions on July 1, Sept 1, and Oct 1. You should store your simulation results in another db table.

d) discuss why your simulated predictions from the electoral college might be different than the general election/538 predictions.

e) try to alter the simulations in c) such that any state where p<40% or p>60% is fixed for the current favorite. Try weighting by trade volume.