**Data Issues**

For this part, please explain the issues that you see with the data, e.g. noise, missing values, etc. Make a detailed list so that when you are ready to conduct your descriptive analysis in the next assignment, you will be able to clean the data accordingly.

Political Contributions:

Description:

We will collect data on contributions to congressional candidates from the 2004-2014 general elections, separated by industry. Data for 2012 and 2014, for election winners, will be obtained using Opensecrets’ getLegislators and candIndustry APIs. Data for 2004-2010 election winners and losers, and for 2012-2014 election losers, will be obtained by scraping Opensecrets.org. Opensecrets gets this data from the FEC.

Issues:

Missing data on some candidates from 2012-2014 (6 total) due to a glitch in Opensecrets’ data collection – they were notified of the glitch but couldn’t fix it in time for our data collection.

Some of the names in the scraped data did not extract cleanly – names with accents presented a problem when brought into Python. This may cause issues when we merge our data.

The scraped data includes all industries from which a candidate received at least $200 in a single contribution, while the data from the API only includes the top 10 industries that donated to the candidate, measured by aggregate contributions.

Election Results:

Description:

Election outcomes (vote counts and percentages) for every House of Representatives and Senate race from 2004-2014. Data for 2004-2012 will be downloaded from FEC.gov as csv files, and 2014 data were scraped from NYtimes.com.

Issues:

FEC 2004-2012 Election data:

FEC does not have 2014 election data available on their website, so we had to find a separate source (NY Times 2014 election data).

No vote count is available for unopposed elections.

Data retrieved from: <http://www.fec.gov/pubrec/electionresults.shtml>

NY Times 2014 election data:

No data are available for uncontested elections – while the candidate may have received nearly 100% of the vote, it might be useful to know the exact percent as well as the number of votes they received.

Some of the names in the scraped data did not extract cleanly – names with accents presented a problem when brought into Python.

Data retrieved from http://elections.nytimes.com/2014/.

Financial Market Data:

Description:

We will collect daily stock market data for the S&P 500 and Dow Jones Industrial Average from 2004 to 2014 from Quandl.com. In addition, we will collect data on 5 sector-specific indices; the Nasdaq Banking Index, Nasdaq Industrial Index, Nasdaq Biotechnology Index, NYSE Amex Oil Index, and the PHLX Gold/Silver Index. S&P and DJIA data will be obtained using the API, while data for the sector-specific indices will be downloaded as csv files.

Issues:

S&P 500: Missing data for Navient Corporation (ticker NAVI); the API did not work for this ticker, as the page for Navient apparently does not exist on Quandl.com. The rest of the data are there, and did not contain any noise.

Dow Jones: Same source as S&P 500, no noise or missing data

Sector Specific Indices:

Nasdaq Banking Index: No issues, downloaded csv from <https://www.quandl.com/data/NASDAQOMX/BANK-NASDAQ-Bank-BANK>

*The NASDAQ Bank Index contains securities of NASDAQ-listed companies classified according to the Industry Classification Benchmark as Banks. They include banks providing a broad range of financial services, including retail banking, loans and money transmissions. On February 5, 1971, the NASDAQ Bank Index began with a base of 100.00.* (from Quandl)

Nasdaq Industrial Index: No issues, downloaded csv from

<https://www.quandl.com/data/NASDAQOMX/INDS-NASDAQ-Industrial-INDS>

*The NASDAQ Industrial Index contains securities of NASDAQ-listed companies not classified in one of the NASDAQ sector indexes. These include firms that are involved in oil and gas productions, oil equipment, services & distribution, chemicals, forestry and paper, industrial metals, mining, construction and materials, aerospace and defense, general industrials, electronic and electrical equipment, industrial engineering, support services, automobiles and parts, beverages, food producers, household goods, leisure goods, personal goods, tobacco, food and drug retailers, general retailers, media, gambling, hotels, recreational services, restaurants and bars, travel & tourism, electricity, gas distribution, water, and multi-utilities. On February 5, 1971, the NASDAQ Industrial Index began with a base of 100.00.* (from Quandl)

Nasdaq Biotechnology Index: No issues, downloaded csv from

<https://www.quandl.com/data/NASDAQOMX/NBI-NASDAQ-Biotechnology-NBI>

*The NASDAQ Biotechnology Index contains securities of NASDAQ-listed companies classified according to the Industry Classification Benchmark as either Biotechnology or Pharmaceuticals which also meet other eligibility criteria. The NASDAQ Biotechnology Index is calculated under a modified capitalization-weighted methodology.The Index began on November 1, 1993 at a base value of 200.00.* (from Quandl)

NYSE Amex Oil Index: No issues, downloaded csv from

<https://www.quandl.com/data/YAHOO/INDEX_XOI-NYSE-AMEX-Oil-Index>

*The NYSE Arca Oil Index, previously*[***AMEX***](https://en.wikipedia.org/wiki/American_Stock_Exchange)***Oil Index****,*[*ticker symbol*](https://en.wikipedia.org/wiki/Ticker_symbol)*XOI, is a price-weighted index of the leading companies involved in the exploration, production, and development of petroleum. It measures the performance of the oil industry through changes in the sum of the prices of component stocks. The index was developed with a base level of 125 as of August 27, 1984.* (from Wikipedia)

PHLX Gold/Silver Sector Index: No issues, downloaded csv from <https://www.quandl.com/data/YAHOO/INDEX_XAU-PHLX-Gold-Silver-Sector-Index>*The PHLX Gold/Silver Sector Index (XAU) is a capitalization-weighted index composed of companies involved in the gold or silver mining industry. The Index began on January 19, 1979 at a base value of 100.00; options commenced trading on December 19, 1983.* (from https://indexes.nasdaqomx.com/Index/Overview/XAU)