**Business Problem: Political Network and Influence Among Graduates of Top US Universities**

1. What are the connections between graduates of Top US Schools and US Politicians ?
   * How is this network affecting US Federal Government in recent elections?
2. Current election season: what is current real time social media sentiment?
3. How much money do graduates of top schools donate and to which politicians/parties?
   * What are the total dollars donated to politicians and how are dollars split by party?

## **Obtaining and Loading Data**

## ***Challenge: How to obtain a complete data set of LittleSis.org***

## LittleSis makes its entire database freely available through API accessbile with a user key

* Calls are HTTP calls; limited to 10,000/day

**Plan** : Obtain data via loop calls to API or webcrawler

* Unsure of total DB size/time required

**Resolution**: Reach LittleSis support for direct transfer of complete DB dump

**Challenge**: SQL format of MySQL Littesis datadump incompatible with PostgresSQL

**Why Postgres as Datasource?**

Postgres is the database of choice for the following reasons:

* Recent coursework featured Postgres (familiarity)
* Open source , non-proprietary
* Simple table structure and types
* Pre-installed and configured on EC2 AMI for w205

***Challenge: How to migrate, transform and load local 4 GB MySQL DB to EC2/Postgres DB***

**Plan:**

1. generate schema from MySQL
2. export MySQL tables to CSVs
3. migrate CSVs to EC2

**Challenge:** MySQL CSV export format **incompatible** with Postgres

* Entire wikis dedicated to converters

**Discovery of Pentaho converter by Sally**

* Pentaho takes a MySQL .sql and automaically generates :

1. Postgres schema and Postgres-compatible csvs
2. Entire Postgres .sql dump file including schema and data

**Final Build Plan: Build LittleSis database stepwise/modularly:**

1) load schema

2) load data/CSVs generated by Pentaho

**Why build step-wise rather than running Postgres .SQL?**

* Devloper can view/modify schema code
* Understand grants in schema
* Modify schema/grants piece-wise while developing

**Working with LittleSis Tables:**

**Challenge:** understanding the actual data content/locations and it’s propensity to satisfy the project proposal

Tables are highly indexed and require multiple, iterative joins to achieve digestible content

**Challenge**: dealing with big data, inherent tendency to “Boil the Ocean”:

* LittleSis is a vast trove of a political data / connections
* Desire was to use as much data as possible for the project
* Too much data/too little time dilemma
* Siphoning some information required advanced programming
* Example for next/future steps:
  + Lobby groups/PACs: these groups are not tagged to a political party in LittleSis
  + Recursive joins through the “relationship” table may yield a political association
  + Number of iterations not known; political association not guaranteed
* **Solution** : identify all persons (not organizations) classified as ‘elected representative’ or ‘politician’ and examine donations made to these people by graduates of Top Schools

**Why use a REST API as the Serving Layer?**

* REST API lends itself to “walking a graph”of relationships
  + Queries linked to results of previous query
* Similar in concept to following links through web pages
* Similar to LittleSis.org website

**Final Results using REST API request:**

curl http://<hostname>.compute-1.amazonaws.com:8080/topschools

curl http://<hostname>.compute-1.amazonaws.com:8080/donationsummaries/<school or ‘all’>/<year or ‘all’>

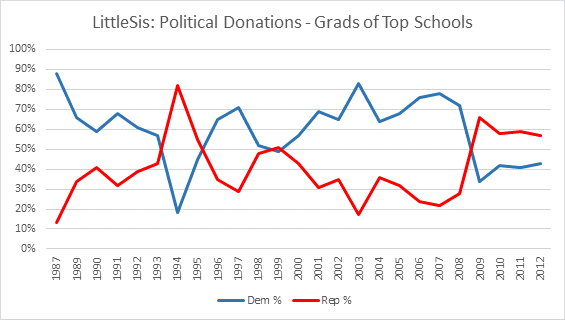
curl [http://<hostname>.compute-1.amazonaws.com:8080/donations/<school or all>](http://ec2-54-87-186-104.compute-1.amazonaws.com:8080/donations/%3cschool%20or%20all%3e)

curl [http://<hostname>.compute-1.amazonaws.com:8080/donations/<gradid](http://ec2-54-87-186-104.compute-1.amazonaws.com:8080/donations/%3cgradid)>

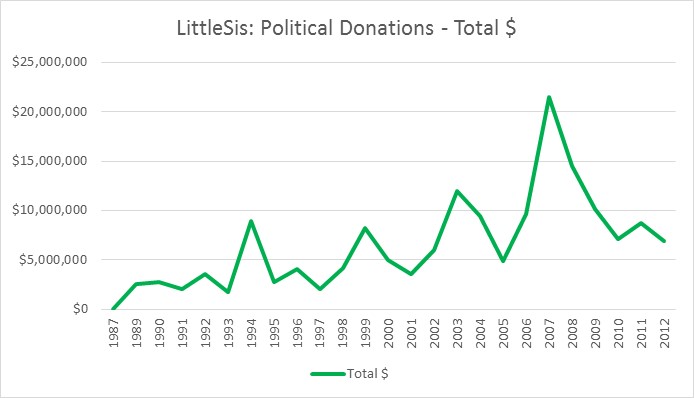
curl [http://<hostname>.compute-1.amazonaws.com:8080/connections/<gradid](http://ec2-54-87-186-104.compute-1.amazonaws.com:8080/connections/%3cgradid)>

**Findings and Discussion (p1. of 3):**

Results for all graduates across all top schools:



* 1994 saw a record breaking Republican sweep of federal government positions
* Post 9/11 and entry into the Iraq war, donations trended higher to Democratic candidates /politicians (2001-2008)
* This changed after Obama became president, after which Republican donations surpassed Democratic donations.
* The record-breaking Repbulican sweep of 2010 elections were a direct result of this support.
* Dollar spike in 2007-2008 (seen above) driven by the Obama Presidency campaign.



**Findings and Discussion (p2. of 3):**

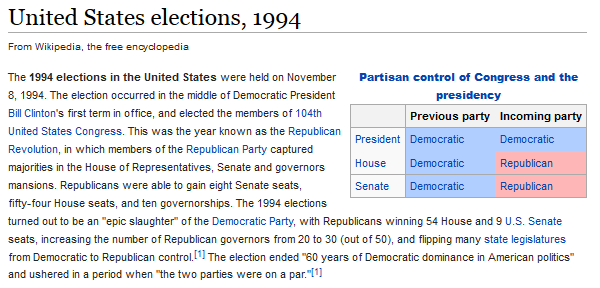
Using the REST API, one can generate all the summary views in the discussion below using the call:

*<hostname>/donationsummaries/<string:topschool>/<string:year>, where year == “all”.*

The REST API can also be used to show top 10 donations for a top school or for a single grad, top 10 connections for a grad, and additional summary data for donation patterns of graduates of top schools along party lines.

Results show that donation patterns at the top schools ***have direct impact on national elections***. While UPenn, UChicago, Princeton, Columbia show consistent democtratic support, other schools see flips between democrat and republican, that mirror the winning parties for those years. In the charts below, there are visible spikes in donations to Republicans during 1994 and 2010 election seasons when republicans had record breaking sweeps in the federal government.

From Wikipedia, record breaking republican election sweeps in 1994 and 2010:





**Findings and Discussion (p3. of 3):**