



Building Better Nonprofits

Ganesh, Razib, Matt, Ashley

Project Overview

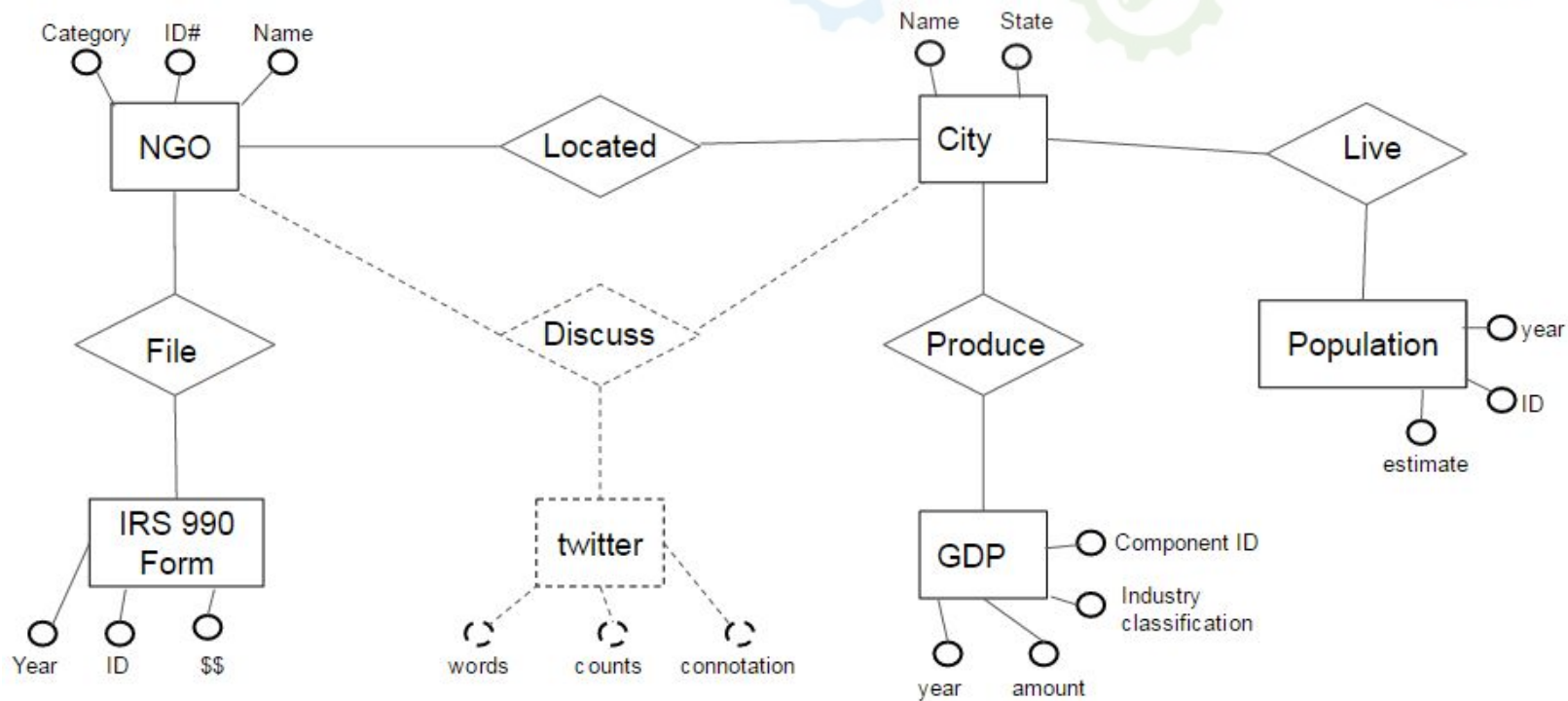
- Focused on building better nonprofits and linking analysis to geographic locations
- Provide resources to enable nonprofits to steer limited resources towards issues for maximum impact instead of employing data scientists within the organizations
- Enable decisions on new or starting locations

Acquisition and Organization

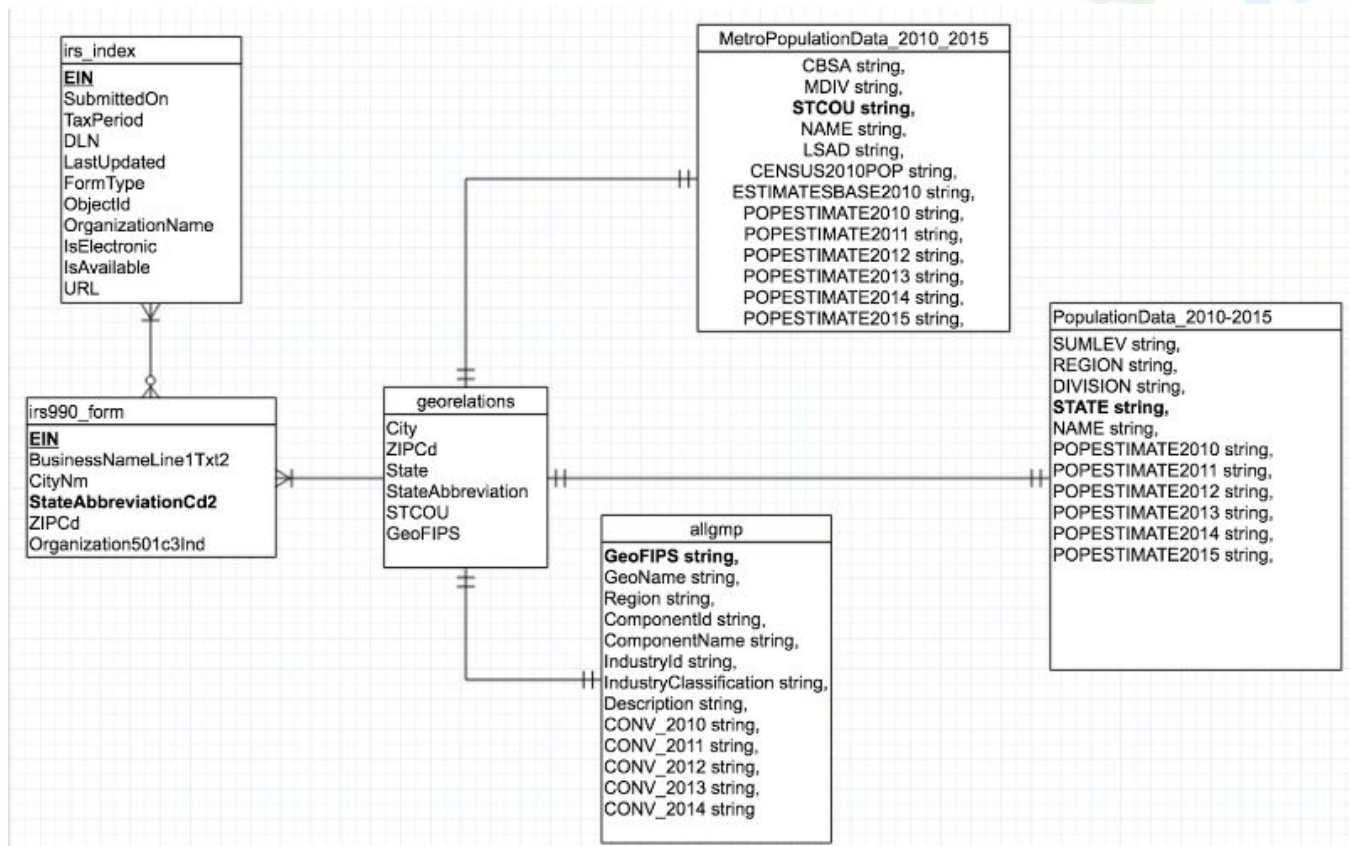
- IRS 990 tax documents
 - Available through AWS
- GDP data
 - Publicly available from US Department of Commerce Bureau of Economic Analysis
- Population data
 - Publicly available from US Census Bureau



ER Diagram First Look

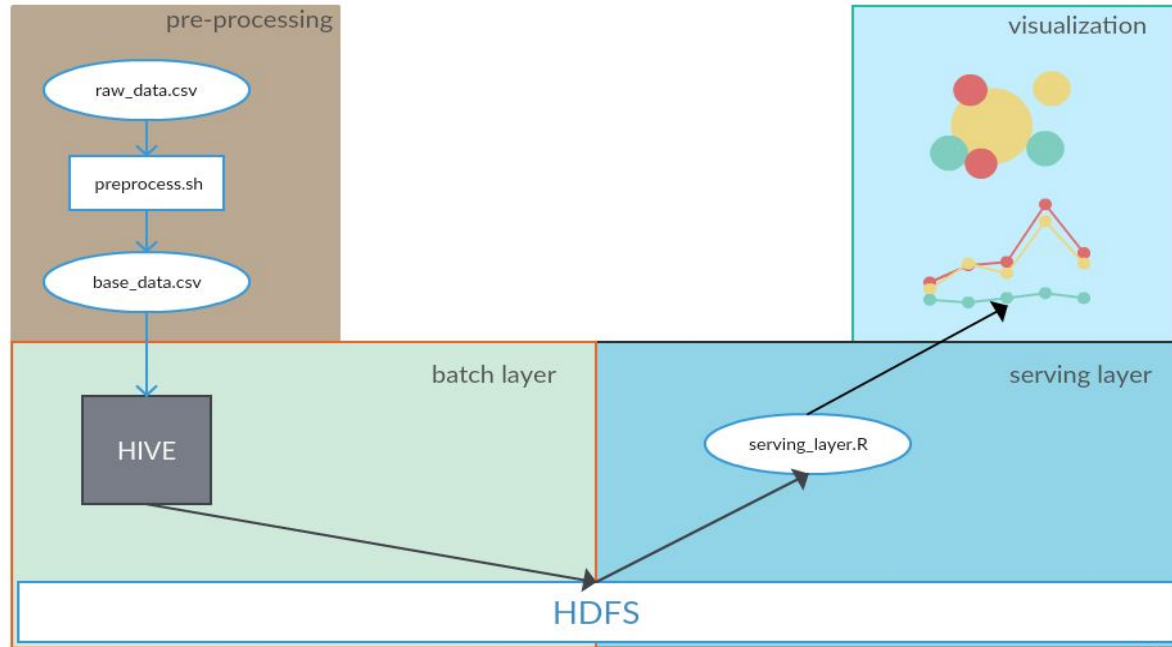


ER Diagram



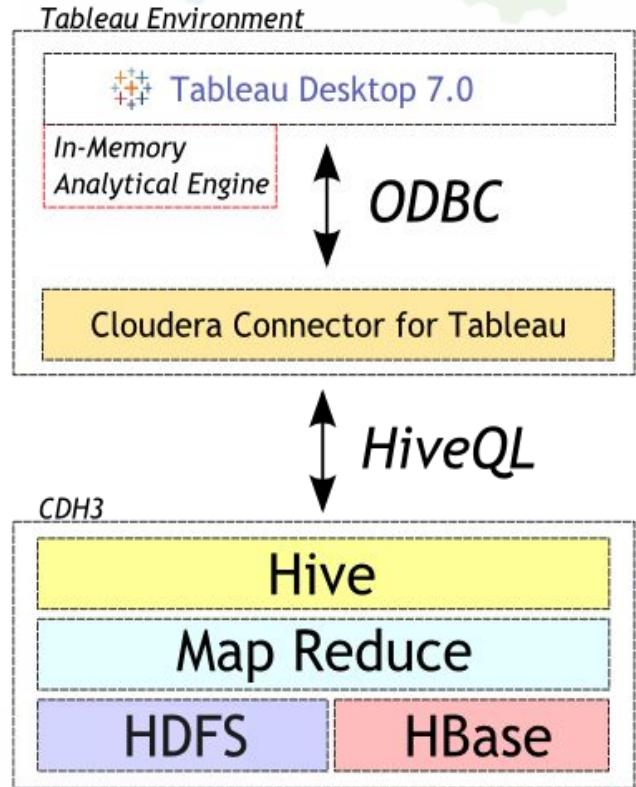
Overall Architecture & Technologies

Lambda Architecture



Architecture & Technologies (cont)

- Pre-processing
 - The raw datasets downloaded need to be cleaned up to create a final CSV that has only the fields that are relevant for the study
 - This means the project team members must have a general idea of what they are looking
- Load the final CSV as table to HDFS
 - SQL (using HIVE layer) is used to create table that will be available to Tableau for analysis and visualization
- HDFS is used as a data warehouse for storing the tables that is used by Tableau
- Batch layer is used to process the master data set. We use PERL script to load pre-processed data from CSV files into HDFS
- The serving layer. In this layer, we use a PERL script that loads the data from HDFS via Hive and inserts it into a HBase table, and hence creating a batch view of the data. This layer also provides query capabilities, necessary in the runtime phase to serve the front-end Tableau.
- Tableau interacts with the serving layer (Hive) by creating SQL based queries to extract data and render a visual representation.



Complexity & Storage

- Data Sets can be
 - Large
 - Incomplete
 - pre-processing.
- Data storage:
 - complex
 - Expensive
 - sharing
- Analysis is time consuming

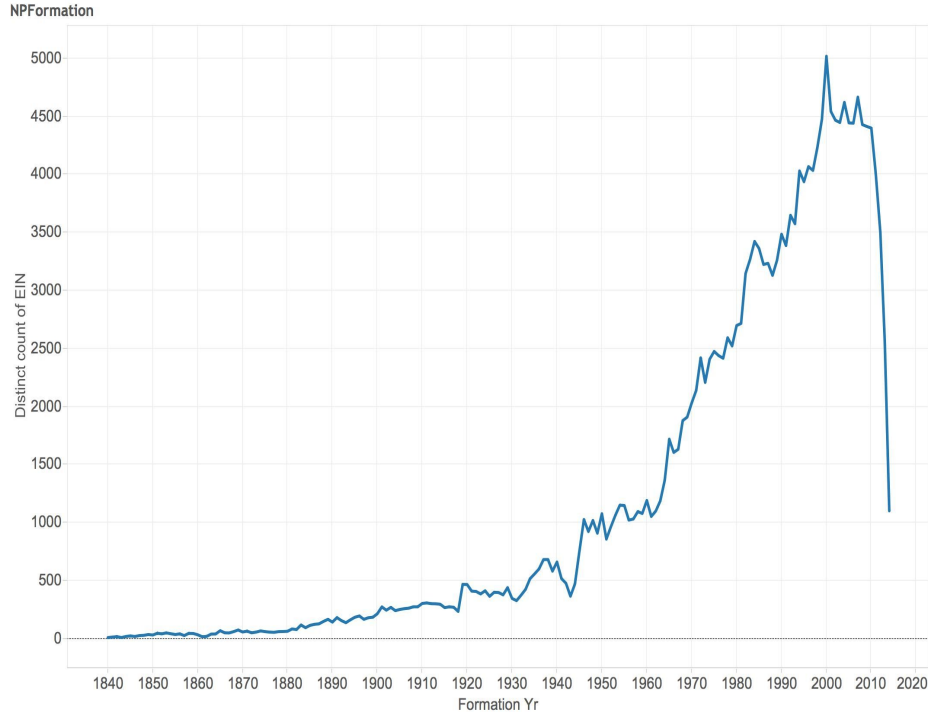


Limitations & Possible Solutions

Limitations	Solutions
Large data volume-- difficult to fully exploit and identify patterns in data	Pre-filter based on specific question
Data history - only available for the years 2013 and 2014.	Manual insertion of old data
Differences in schema	Reconcile schemas



Analysis & Results

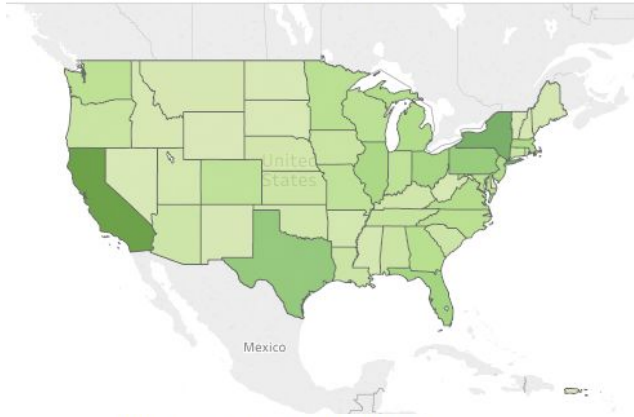


The trend of distinct count of EIN for Formation Yr. The view is filtered on Formation Yr, which ranges from 1840 to 2014.

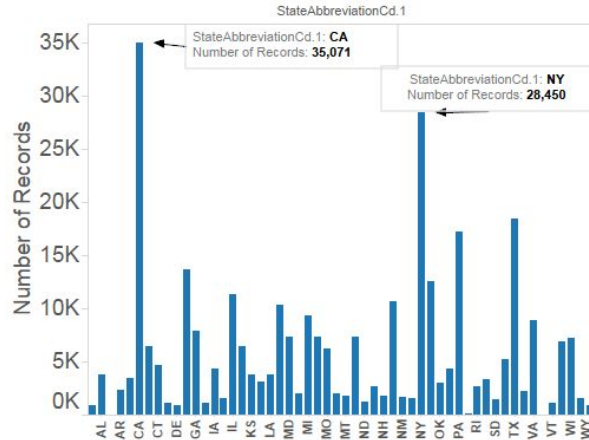
- Nonprofit growth over time
- Small plateau after 2000
- Significant drop after 2010, nonprofits may not be submitting taxes or decrease in growth is real

Analysis and Results

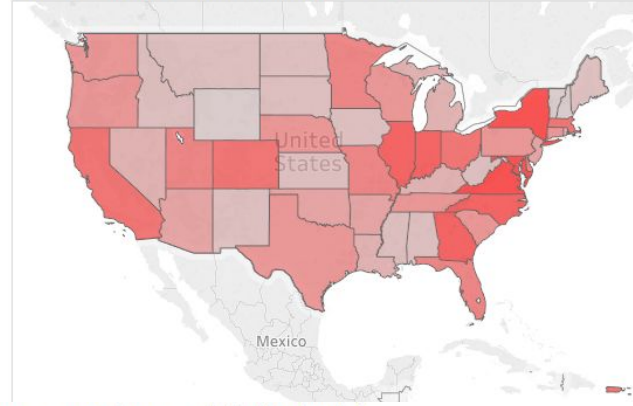
Number of non profits by States (map view)



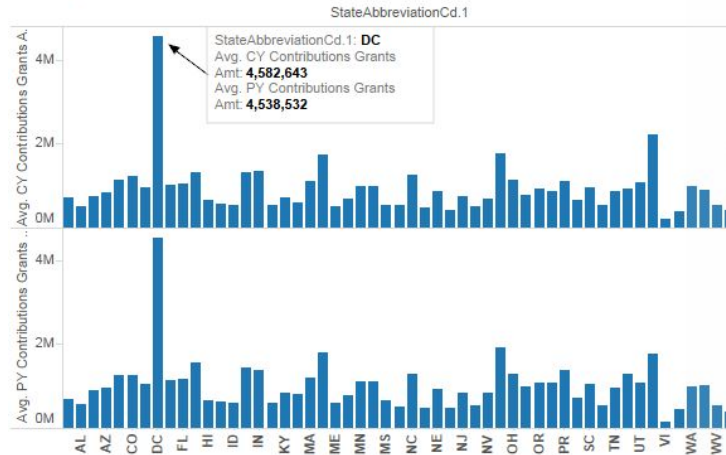
Number of non profits by States



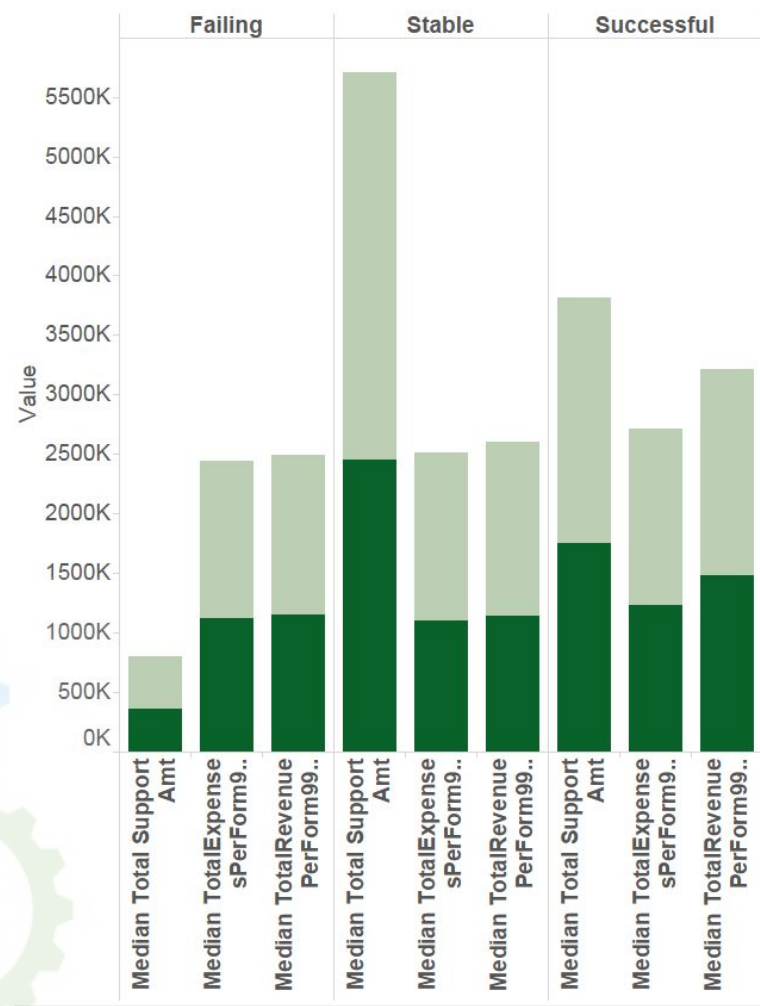
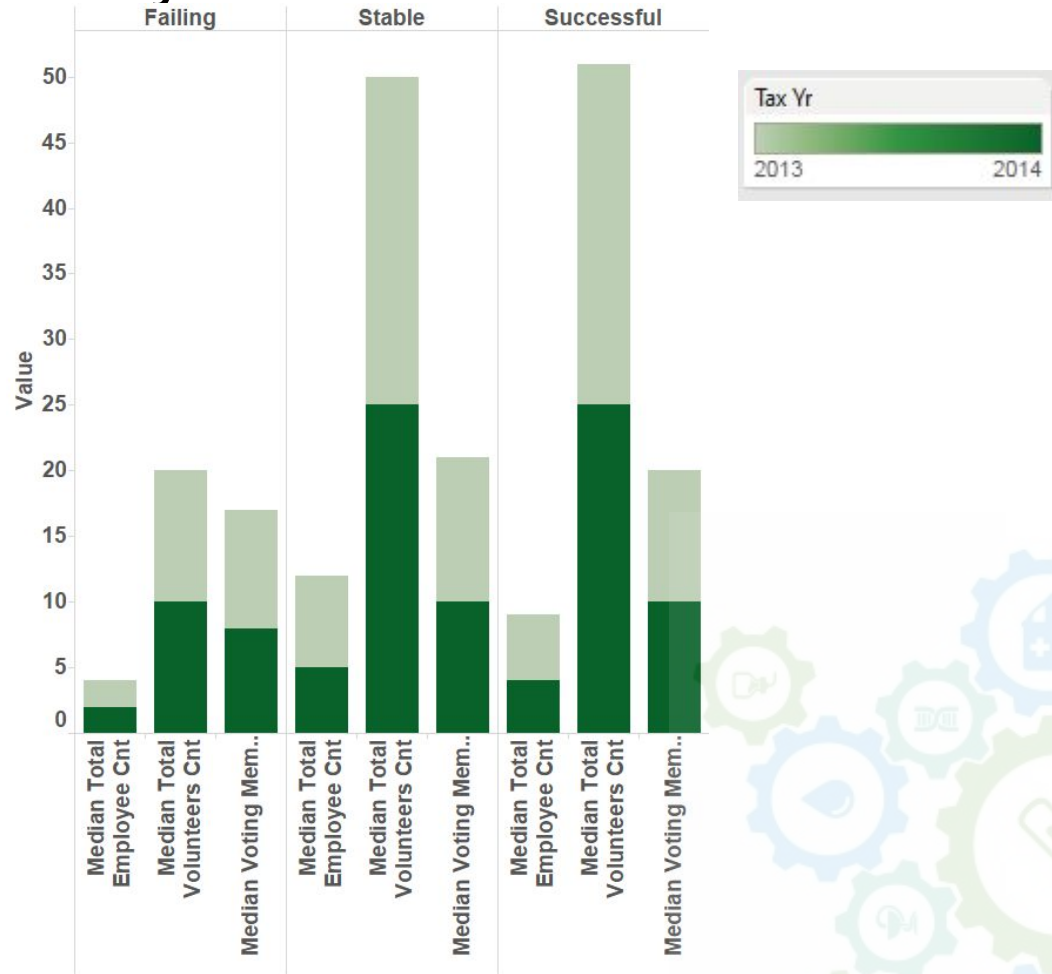
Average current year contribution by state (map view)



Average current year contribution by state

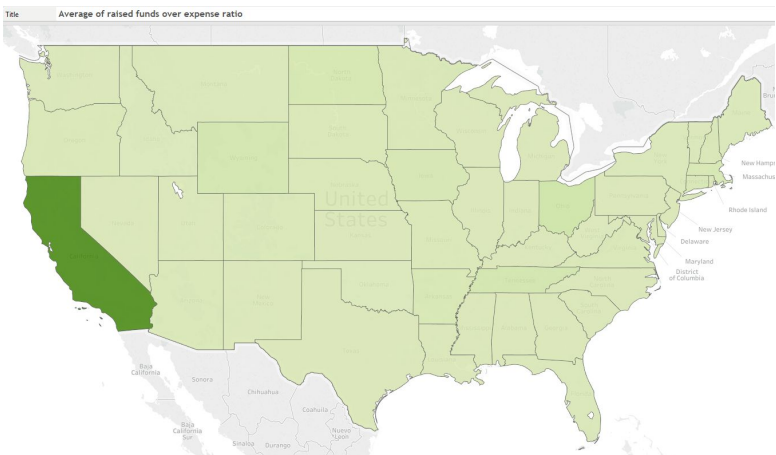


Analysis and Results



Analysis and Results

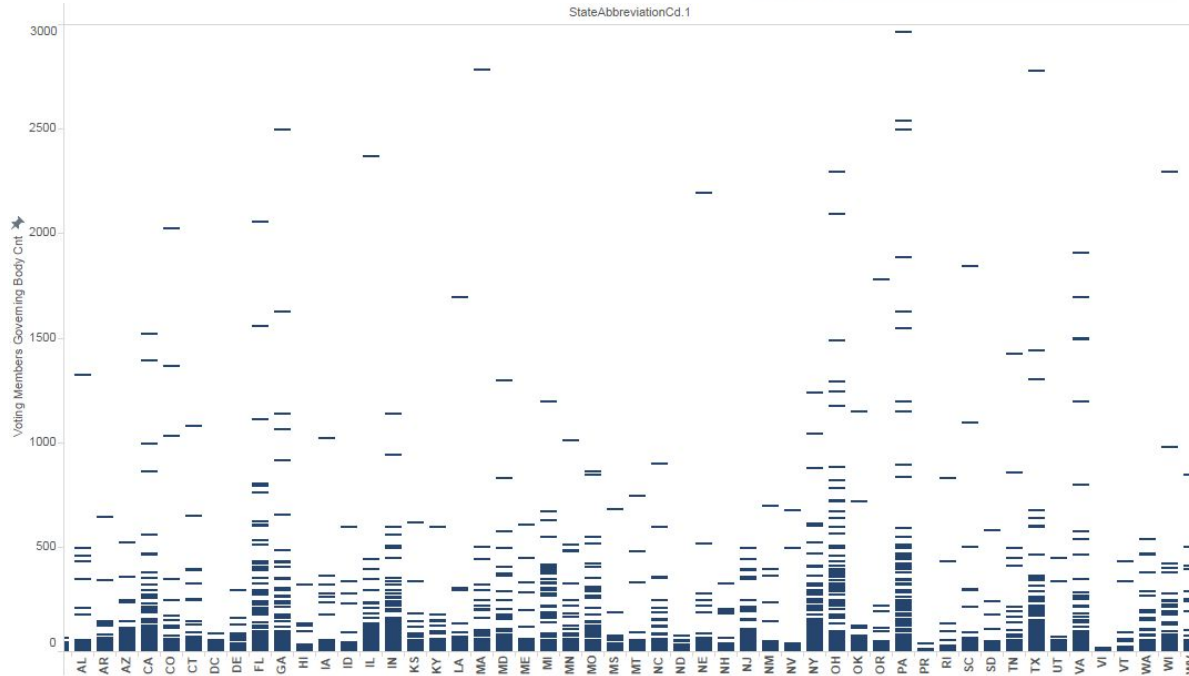
Average of raised funds over expense ratio



Avg. Fundraising Amt

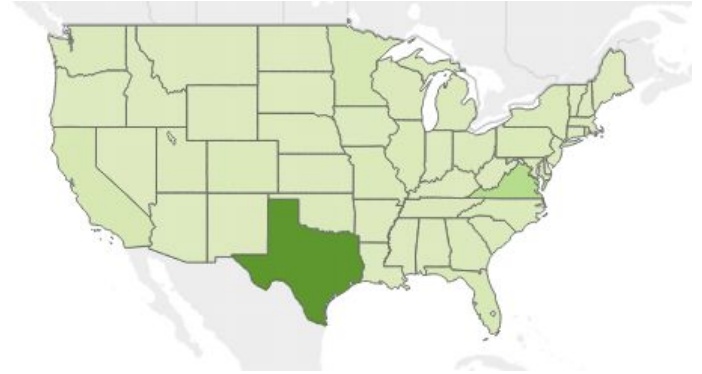
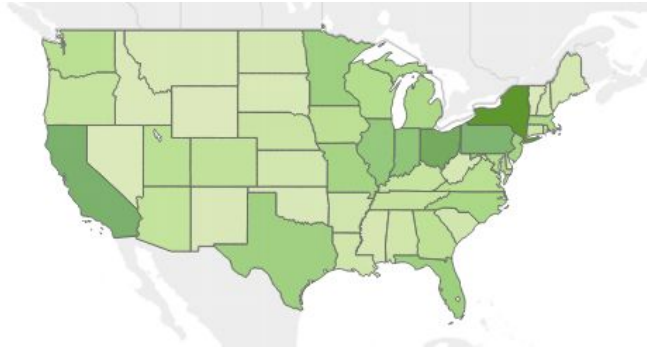


Analysis & Results



Successful Nonprofits tend to have less voting members on their governing board, with possible state differences

Analysis & Results



Employees vs Volunteers for Successful Nonprofits



Future Analysis

- Employees/volunteer ratios and underlying factors
- Multiple-factor correlations
- City level analysis

