

List of Tables from ERD

Voter

Already exists as voter_roll

Election

Existing table pvrdr_vph comes close, just need to break registrant info into Election Record table

Election Record

Includes information on individual participation in elections

Election Results

Regarding the county-level outcome of presidential elections. Data can be pulled from [here](#).

Census

Regarding the aggregate demographic (we'll need to choose some interesting dimensions) makeup of counties. Data found [here](#).

List of Project Questions & Methodologies

1. How do "Consistent Voters" differ demographically from the overall population?
 - a. Define "consistent". These are registrants who have no NULLs in their Election Record table stretching back how many election cycles?
 - b. Join this list of registrantids back to the voter table in order to populate demographic info (gender, age, placeofbirth).
 - c. Compare the aggregate demographic data with aggregates for total state population.
 - d. Conclusion will be something along the lines of: consistent voters are OLDER, more MALE, and fewer FOREIGN BORN than the CA population overall. Is this an interesting question?
2. How do "New Voters" differ?
 - a. New voters will be null in all but the last election.
 - b. Same methodology as above, same type of conclusion
3. What was the effect of crossover voters on the most recent election?
 - a. Start with Election Record table, create a subset of voters from the most recent election.
 - b. Join this to the Voter table in order to add partycode & countycode associated with each registrantid.
 - c. Group by countycode to generate a sum (# of registered democrats etc who voted in that election) compare to totalvotesD/R from Election Results table.

- d. Possible conclusion: a list of the counties where Election Results % D vs R looked most unlike the registration profile for that area.
- 4. Other ideas:
 - a. What are the most/least politically engaged counties and what is unique about them? Generate aggregate registered voters % of county population and active voters % of county population. For “most” and “least” actives compare aggregate demographic population with state overall.
 - b. Do any counties stand out as predominantly (male/female, old/young, foreign/native-born)? How are election outcomes here different?
 - c.

To-Dos

1. Pull in two new datasets: Census and Election Results. Clean em up.
2. Generate an “age” field within Voter record by in order to easily compare Voter records (contain DOB) with Census info (contains age composition).
3. Field countycode will be the foreign key linking Voter to Census to Election Results. Examine the existing nomenclature in these tables and create a transformation that normalizes it against the classifications in Voter table.