CSE 416, SECTION 1

Project Discussion
Part 2

1

Session Objectives

This is part of your requirements analysis phase of the project

- Understand issues and terminology used in the analysis of election issues in the US political process
- Understand the top-level goals of your project
- Understand some of the data requirements to support your analysis

2

Voting Drop Boxes

- Secure, locked receptacles
- Operated by election officials
- Placed in well-lit, high traffic areas under video surveillance
- Rare voting fraud (but many claims of voter fraud using drop boxes)
- Varying state policies

Another EAVS data use case

3

Provisional Ballots

- A ballot might be rejected at a voting location because the voter
  - Was not on the voter registration list
  - Voted in the wrong precinct
  - Was challenged by a poll watcher
  - Other
- HAVA requires a provisional ballot to be cast in that case
- The ballot is counted if it has the potential determine an election winner
- Issue – Does the voter always know when his/her ballot is not immediately accepted – Not when it is mailed

A use case will look for differing rates of rejection

4

Voter Registration …

- All US states (except North Dakota) require eligible voters to register before they can vote
- Eligibility
  - 18 years of age or older
  - Citizens
- Some states
  - Encourage registration (e.g., opt-out at DMV)
  - Provide optional registration in a political party
- Registration period varies, as some states
  - Allow Election Day registration
  - Have a defined period prior to Election Day

EAVS use cases will compare voter registration among states

5

… Voter Registration

- Felony disenfranchisement
  - Convicted felons cannot register to vote
  - State policies vary (some not at all, others during incarceration, others during parole, others post parole)
  - Factor in 2000 Florida election
- Registration drives – not legal in all states; some limitations
- Poll books
  - List of eligible voters
  - Maintenance policies vary
  - Content varies (e.g., preferred political party)
  - Data is public (with limitations)

6

Voting Data

- Registration rate
  - Number of people registered to vote divided by the number of people eligible to vote
  - Calculation requires multiple datasets (registered voters from EAVS, eligible voters from Census Bureau)
- Turnout – actual voters divided by eligible voters

Additional use cases

7

Citizenship Requirement

- Varies among states, some states require
  - Potential voter to assert he/she is a citizen (providing false information is a felony)
  - Database cross-referencing
  - Documentary evidence (e.g., passport)
- A recent Presidential Executive Order (currently being challenged in the courts) amends the federal form to require documentary evidence of citizenship

8

Voter Registration Lists

- State requirement due to the HAVA Act
- Lists are statewide (not local)
- States are required to update the list with data within the state
- No consideration of registered voters who move interstate
- Voter lists can be in error due to
  - Delays in coordinating with other statewide DBs (e.g., death records)
  - Clerical errors (e.g., wrong zip code)
  - Out of state moves

9

Voter Registration Data

- When available, registration data will contain name, address, registered political party, and other info (details vary by state)
- Some states set access limits (e.g., resident, fee for access)

```
"SOS_VOTERID","COUNTY_NUMBER","COUNTY_ID","LAST_NAME","FIRST_NAME","MID
"OH0014703381","18","517217","MORRIS","THOMAS","O","","1967-02-20","198
"OH0014703420","18","517276","SOFRANEC","DIANE","L","","1963-07-10","19
"OH0014703421","18","517297","TWINING","STEVEN","R","","1955-04-10","19
"OH0014704502","18","519195","ROCCO","ANDREA","F","","1965-03-31","1983
"OH0014704518","18","519228","ANDERSON","DEBORAH","L","","1969-11-01","
"OH0014705519","18","520975","BERTSCHLER","JOHN","J","","1948-07-04","1
"OH0014705535","18","521002","CARLETON","GARY","P","","1943-04-07","196
"OH0014705550","18","521003","CARLETON","MAUREEN","FRANCES","","1943-10
"OH0014706718","18","523141","WALKO","GREG","P","","1954-11-09","1900-0
"OH0014706720","18","523123","TSITINIDIS","NIKOLAOS","E","","1941-03-14
"OH0014706721","18","523124","TSITINIDIS","OLYMPIA","N","","1950-05-11"
"OH0014706729","18","523145","WASSERBAUER","CATHERINE","M","","1962-07-
"OH0014706733","18","523148","WASSERBAUER","THOMAS","JOHN","JR","1962-0
```

10

Registration Data Use Case

- One of your use cases will be to analyze the registration data for accuracy
- You will need to
  - Select a state
  - Download a state registration file
  - Identify errors in the voter's address
  - Quantify the percentage of errors
- If your state's data is not immediately available, contact the TAs

Data verification use case will depend on data verification tool

Voter Registration Data	Team 1
Arkansas	
Florida	Bulls
Indiana	
Massachusetts	
New York	
North Carolina	
Ohio	
Oklahoma	
Vermont	

11

Data Requirements

- 48 states for EAVS (2016-2024)
- Boundary data (e.g., county, town, etc.)
- Demographic data (e.g., population, population by demographic group, etc.)

Boundary data may not be as easy as it appears

Voter age population is usually the best to use in this sort of analysis

You may need census block data if some town/county data is missing

12

How is Geographic Data Organized

- Think of each area as a large polygon (but sometimes it might be a multi-polygon)
- Boundary data of interest
  - State
  - Town
  - County
  - Census block
- Usually, m census blocks form a precinct, n precincts form a county, and k counties form a state

Some EAVS data is organized by town – might make analysis more difficult

13

Shapefiles

- Geospatial vector data format
- Developed and maintained by ESRI
- Introduced in early 1990s
- Collection of files
  - Usually stored as a zip file
  - Mandatory files (.shp, .shx, and .dbf) and other files
- Represents points, lines, polygons
- Formatted as fixed length header, followed by one or more variable length records

Dominant format for geographic data due to the market dominance of ESRI

14

GeoJSON

Be alert for MultiPolygon data

- Open standard format for representing simple geometric features
- Based on JSON
- Types – Point, LineString, Polygon, MultiPolygon
- Supported by Leaflet, Google Maps, et al
- Position information expressed as longitude, latitude

Become familiar with conversion SW


```
{
"type":"FeatureCollection",
"name":"precincts",
"description":"Minnesota Congressional District 1
"title":"Minnesota Congressional District 1 Votin
"publisher":"Office of the Minnesota Secretary of
"date":"July 1,2019",
"features":[
{"type":"Feature","properties":{"Precinct":"Amboy
Earth","CountyID":"7","CongDist":"1","MNSenDist":
[[[-94.1585,43.8916],[-94.1651,43.8915],[-94.1651
[-94.1657,43.8879],[-94.1665,43.8879],[-94.1665,4
[-94.1664,43.8868],[-94.1664,43.8862],[-94.1582,4
[-94.1583,43.8856],[-94.1585,43.8856],[-94.1585,4
[-94.159,43.8848],[-94.159,43.8849],[-94.1585,43.
[-94.1577,43.8861],[-94.1575,43.8861],[-94.1575,4
[-94.157,43.8842],[-94.157,43.8843],[-94.1574,43.
[-94.1537,43.8828],[-94.153,43.8829],[-94.153,43.
[-94.1529,43.8862],[-94.1529,43.8867],[-94.153,43
[-94.1485,43.8903],[-94.157,43.8902],[-94.157,43.
[-94.153,43.8887],[-94.153,43.8884],[-94.1536,43.
{"type":"Feature","properties":{"Precinct":"Beauf
Earth","CountyID":"7","CongDist":"1","MNSenDist":
[[[-93.8884,44.0222],[-93.9085,44.0221],[-93.9286
[-94.0084,43.964],[-94.0084,43.9349],[-93.9685,43
```


15

Non-Geographic Data

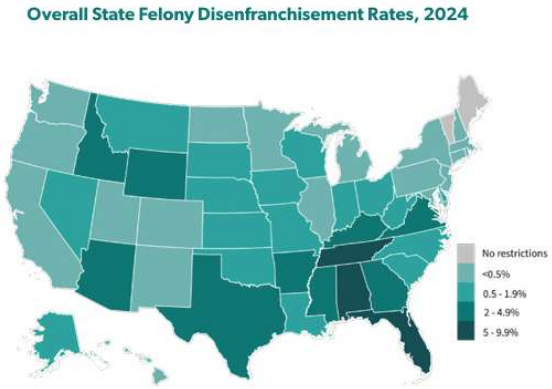
- Population data
  - Total population
  - Voting age population (VAP)
  - Citizen voting age population (CVAP)
- Demographic data
  - Racial/ethnic

If you cannot get data by town/county, you may need to sum up contained census blocks

Consistently use one category of population data – VAP is best

16

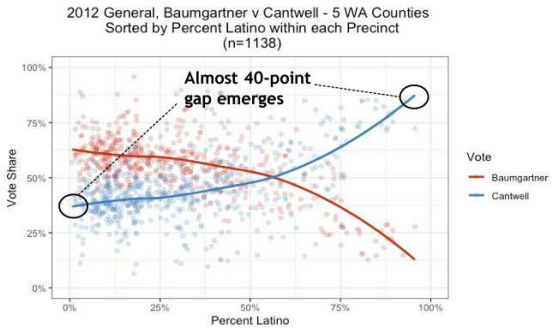
Choropleth Displays



- Many EAVS use cases will require choropleth maps
- Choropleth display – thematic map that uses color to represent different data values
- Visual aspects
  - Bins (usually equally spaced) define range of values for a color
  - Number of bins usually 5-10
  - Colors for sequential data are a single hue (varying lightness)
  - Legend defines the bins

17

Bubble Charts



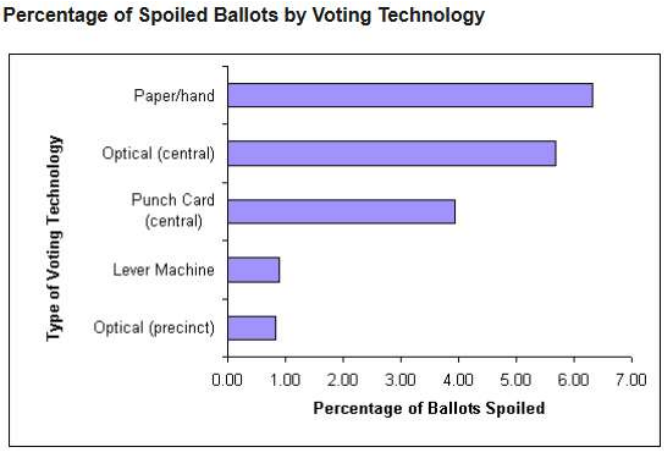
- Required use cases
- Bubble charts represents data with 3 or 4 variables
- Example – dominant demographic group (Census Block)
  - Latitude
  - Longitude
  - Dominant group (color code)
  - % population (size of bubble)

Bubbles can be overlaid on a map to show demographic data by location (e.g., county)

18

Bar Charts

- Graphical representation of data using bars of varying heights or lengths
- Usually for analysis of data within one state (e.g., % of rejected ballots by cause)



19

Sources of Data

- Project Web site suggests many sources
- For example,
  - Redistricting Data Hub
  - Harvard
  - MIT
  - US Census Bureau
  - Open Elections
- Easier sources of data (including some consolidation) are available
- Choose a data source that provides data at a level you need for your states

20

Stuff to do After Week 3

- Get familiar with a GUI builder (TA session was recorded)
  - Pick one
  - Build a splash page (with US map)
- EAVS DB
  - Pick a schema design tool
  - Build a partial schema based on requirements discussed in class
  - Check the Voting Equipment DB (https://verifiedvoting.org/equipmentdb/) for voting equipment found in your states
- Read textbook Chapter 7 to better understand requirements
- Pick your states
- Use Leaflet (or alternate) for including maps in your GUI
- Start using a client-side visualization library (e.g., D3) and build
  - Bubble chart
  - Bar chart
- Download recent ACS data for demographics (EAVS region and census block)

21

Have You Satisfied the Session Objectives?

- Understand issues and terminology used in the analysis of election issues in the US political process
- Understand the top-level goals of your project
- Understand some of the data requirements to support your analysis

22