CSE 416, SECTION 1

Project Discussion
Part 1

1

Session Objectives

This is the start of your requirements analysis phase of the project

- Understand issues and terminology used in the analysis of potential changes to the Voting Rights Act
- Understand the top-level goals of your project
- Understand some of the data requirements to support your analysis

2

 CSE416-S01 3

Reading

- Computational Redistricting and the Voting Rights Act (link in References section of the class home page)

3

 CSE416-S01 4

Overall Project Goals

- Understand the implications of the potential Supreme Court dismissal of Section 2 of the Voting Rights Act (VRA)
- Assess the impact on a preclearance state and a non-preclearance state
- Approach
  - Develop a probability distribution of minority representation under the existing VRA
  - Compare with the probability distribution of minority representation under a restructured VRA (i.e., race-blind districting)
  - Apply data analysis techniques to the available data
  - Generate data summaries and visualizations that summarize the results of the analysis

4

 CSE416-S01 5

Project Technical Goals

- Build a large, complex system that requires teamwork to integrate multiple components
- Use multiple programming styles and languages (JavaScript, Python, and Java)
- Incorporate
  - Requirements analysis (application domain knowledge)
  - User interface design
  - Data base design and development
  - Algorithmic analysis and mathematical thinking
- Use Data Analysis techniques in performing analysis

Project technology is too complex for 1 or 2 students to build the system

5

 CSE416-S01 6

Software Engineering Data Analysis Project

- Data gathering from multiple data sources (note: expect data irregularities such as missing data, inconsistent data, typos, etc.
- Data cleaning to improve data accuracy, where possible
- Data generation – simulate elections under current and potential VRA to generate probability distribution estimates
- Data analysis – analyze the data to uncover trends
- Data visualization - visualize information contained in the data

6

 CSE416-S01 7

What Will You Do

- Download data from multiple data sources
- Design and build a database
- Develop/modify election simulation SW to be run on a SBU high performance computer (HPC)
- Extract summary data from the HPC to include in your DB
- Build a Web GUI that allows a user to obtain meaningful information from the DB
  - Tabular data
  - Visualizations
  - Geographic displays
- Build server SW that connects the GUI to the data

7

 CSE416-S01 8

How Do You Measure Racial Fairness?

- Comparison of the VRA-constrained ensemble with the race-blind ensemble
- Racial/ethnic comparison of enacted/proposed plan with expected distribution (simple count of population percentages)
- Comparison of enacted/proposed plan with both ensembles

8

 CSE416-S01 9

How Do You Display Results of Your Analysis?

- Splits (bar charts)
- Map displays (e.g., choropleth)
- Bubble charts
- Ecological Inference
- Seat share / vote share curves
- Box & whisker plots

You will use MCMC ensembles to generate some of the data to be displayed

9

 CSE416-S01 10

How Do You Generate An Ensemble

- Use MGGG software
- For each random plan
  - Simulate the election using previous election data for each precinct in a district (simple assumption - 2024 presidential votes in a precent predicts 2026 Congressional votes)
  - Sum up the precinct votes in a district (aka, node votes in a sub-graph)
  - Determine the winners in each district
- For the ensemble
  - Calculate district splits

Use both the standard ReCom (GerryChain) and the constrained ReCom (for VRA)

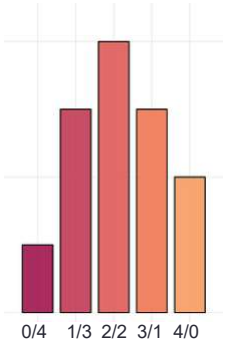
10

 CSE416-S01 11

Ensemble Party Splits

- Party Splits – possible distribution of Republican and Democratic district wins
- Example – Iowa has 4 Congressional districts, so splits are
  - 0/4 (Republican/Democratic)
  - 1/3 (Republican/Democratic)
  - 2/2 (Republican/Democratic)
  - 3/1 (Republican/Democratic)
  - 4/0 (Republican/Democratic)

Height of each bar corresponds to the number of times that split occurred in the ensemble



You will also calculate racial splits (# times a racial group exceeded 50% of district population)

11

 CSE416-S01 12

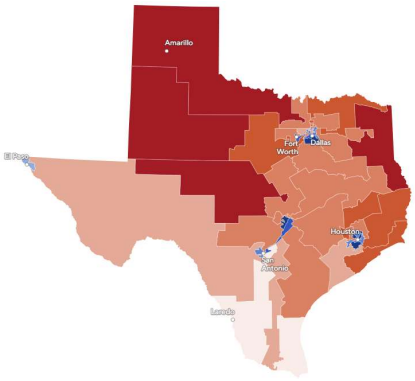
Display Splits

- You can compare (VRA and race-blind) splits
  - Side by side
  - Stacked
- But not in a pie chart (doesn't communicate results as well for this)

12

 CSE416-S01 13

Choropleth Displays



- Some use cases will require choropleth maps or district color maps (*e.g., minority population)
- 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

13

 CSE416-S01 14

Racially Polarized Voting

- Racially polarized voting is a key element of VRA legal cases
- Provides statistical evidence that the political power of minority voters has been compromised
- Must include
  - Minority group votes as a bloc
  - White majority votes as a bloc

Part of the "Gingles test" – Thornburg v. Gingles (1986)

- Racially polarized voting serves as evidence of a discriminatory effect

You will show visualizations (bubble charts and ecological inference) that identify racially polarized voting

14

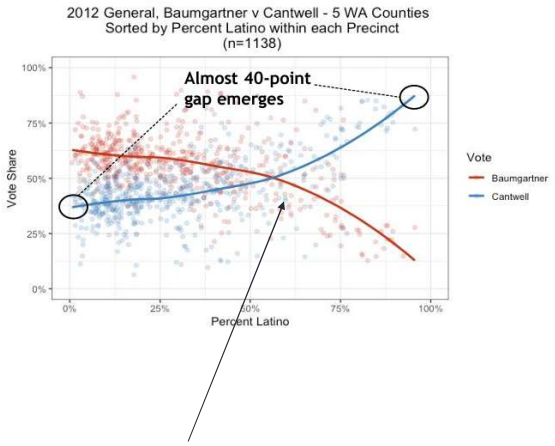
 CSE416-S01 15

Bubble Charts

- Display past voting pattern (e.g., 2024 Presidential voting in a state)
- Bubbles represent
  - Red dots for Republican vote percentage in a precinct
  - Blue dots for Democratic vote percentage in a precinct

  Bubble x-axis location based on population demographic in the precinct

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



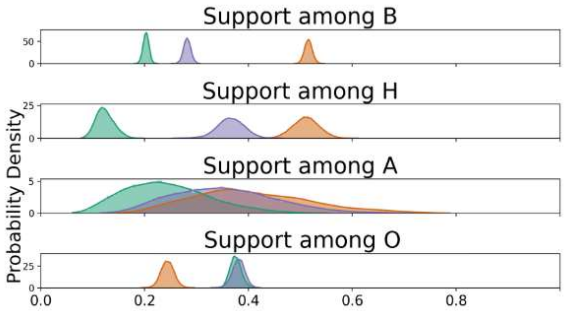
Note the non-linear regression lines

15

 CSE416-S01 16

Ecological Inference (EI)

- Statistical data analysis
- Estimates individual-level voting behavior using aggregate data, typically voting patterns of racial/language groups
- Produces probability voting curves
- Use PyEI software

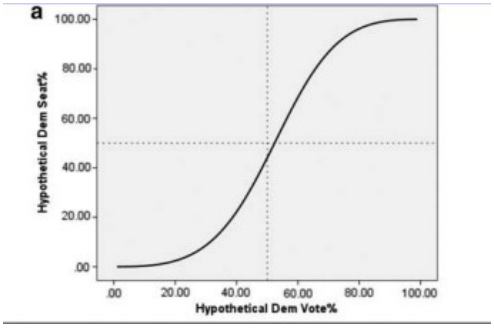


16

 CSE416-S01 17

Political Fairness - Seat vs. Vote Symmetry

- Plot vote share vs seat share to identify potential unfairness in current plan
- Notion that voting is equal when a 50% vote share results in a 50% seat share
- A shift in the curve indicates gerrymandering
- To calculate, change your vote prediction calculation to generate points to plot
- Helpful to compare multiple district plans



Shen code is a good starting point

Image: Election Law Journal, McDonald & Best

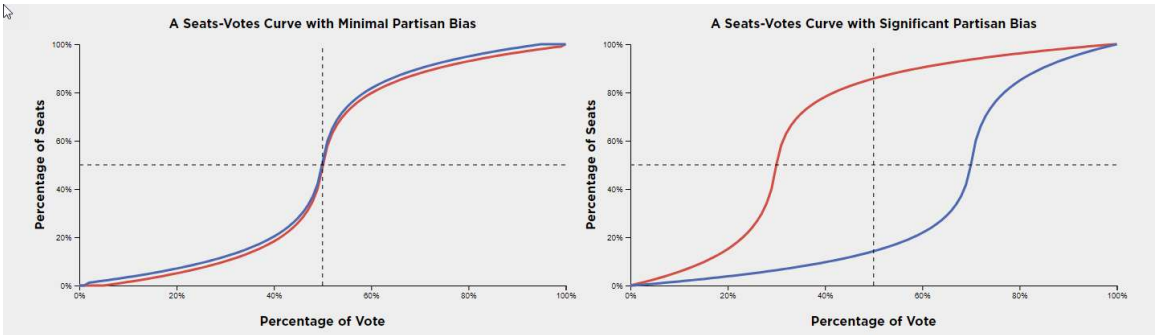
17

 CSE416-S01 18

Partisan Bias

Interesting to compare the current district plan to ensemble results

- Deviation from 50% seat share at 50% vote share



How do you generate a vote/seat share curve for an ensemble?

18

 CSE416-S01 19

Box & Whisker Analysis

- Used in many redistricting court cases to evaluate a plan
- For each random district plan
  - Calculate the population percentage for each district using the group of interest (e.g., African American)
  - Sort the districts in increasing order of the group population percentage
  - Repeat above steps for all random district plans
  - Store data
- For each district group (e.g., lowest percentage of the population)
  - Calculate maximum, minimum, 1st quartile and 3rd quartile
  - Calculate the box (1st quartile to 3rd quartile) and whiskers (max, min)
  - Plot
- Repeat the process for selected district plans (e.g., enacted)

19

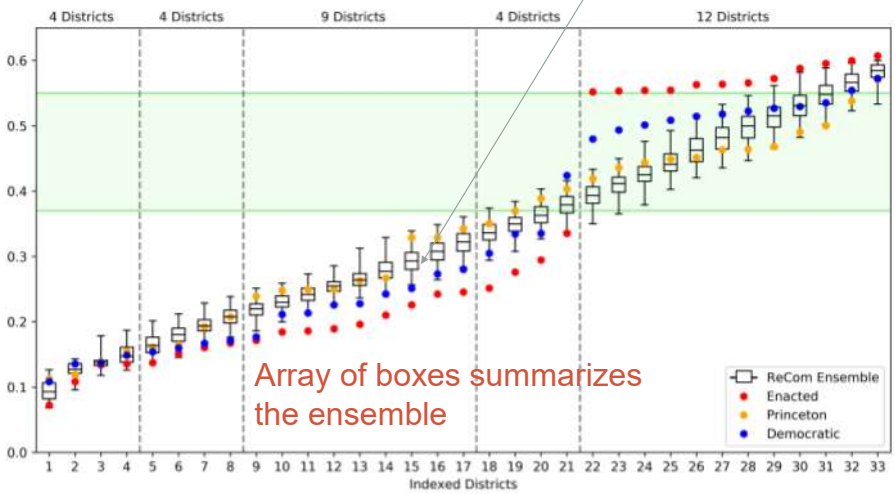
 CSE416-S01 20

Box & Whisker Plot

Average for the district

Example is from the Virginia court case

Y-axis is either % of a minority group or % of votes for a political party



Array of boxes summarizes the ensemble

20

Which Groups Do We Measure?

- Measure a group if they have a reasonable likelihood of electing a representative (i.e., population more than 400,000)
- Possible groups
  - African American
  - Hispanic
  - Asian American

21

Use Cases

- Requirements Analysis – “Requirement analysis in software engineering is the critical initial process of identifying, understanding, documenting, and verifying stakeholder needs to define the precise requirements for a software product. ” – Google AI
- Next few sessions define the needs of a user of your project
- Requirements defined in use cases - you will assist in defining

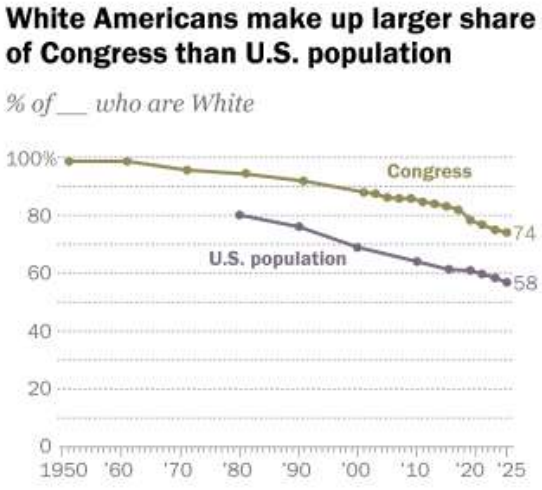
GUI-1. Display map of US on splash page (required)
The project splash page will contain a map of the US (mainland 48 states) that takes up much of the page. States will be highlighted with thicker border lines when the project has additional data for that state (e.g., voting equipment data). The page will also include drop-down menus so that the user can select a state to view more details (e.g., choropleth maps).

22

 CSE416-S01 23

Possible Background Displays

- 16 US Senators are from racial or ethnic minorities (7 Hispanic, 5 Black, 3 Asian, 1 Native American)
- House
  - Black – 14%
  - Native Americans – 1%
  - Hispanic – 11% (vs. 20% total)
  - Asian American – 4% (vs 6% total)

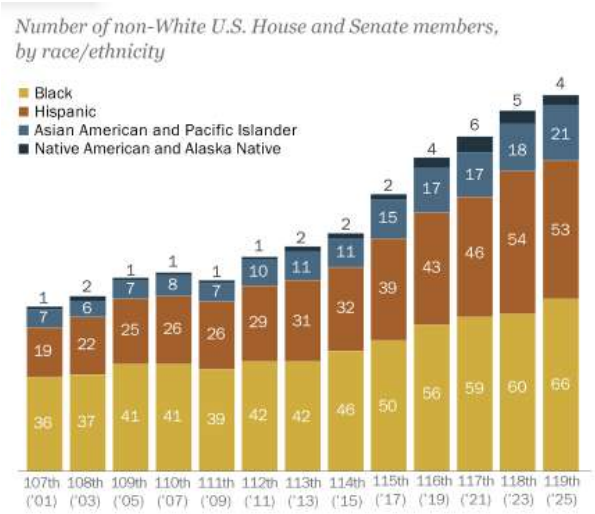


23

 CSE416-S01 24

Demographic Composition of US Congress

- Determine "fairness" of existing racial composition of US Congress
- Project "fairness of changes that might result from a gutting of VRA



24

User Interface Preliminaries

- Select mapping API for your client
  - Leaflet
  - Google maps
  - Others
- Select GUI builder tool
  - Drag & drop functionality
  - Templates
- Select client management system (e.g., React)

As we cover the details of the requirements, you will plan to organize your GUI

Figma TA session Wednesday at 7PM

25

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

26

 CSE416-S01 27

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

27

 CSE416-S01 28

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
```


28

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

29

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

30

What Skills Do You Need?

TAs were selected based on experience with these technologies

- Programming (Java, JavaScript, Python)
- Client/server interaction (e.g., Spring, JAX-RS)
- Data analysis
- Performance analysis
- GUI Map system integration
- Client data display
- Client framework (e.g., React)
- DB
- And more

Free SW libraries are available

Almost impossible for a team to have all these skills

31

How Should Your Team Organize

- Team leader vs. democratic organization
- Coverage of all needed skills (e.g., data gathering and processing, server, DB, GUI)
- How do you define interfaces (what tool will you use)
- How do you review work progress (what project management tool will you use)

What project work can you do now?

32