# CSE 416, SECTION 1

## Project Overview

1




## Project Teams

- Teams –all teams filled
- Please sit with your team for all class sessions

| Team | Team members |
|---|---|
| Astros | Kamael Li, Jiaxin Xie, Supriya Gurung, and Collin Chan |
| Braves | Ayden Budhoo, Kobe Du, Akash Hongal, and Sahil Parikh |
| Cubs | Cherril Bhansali, Vivian Zhu, Sumon Ahasan, and Jason Yeung |
| Dodgers | Joel Cruz, Tae Kim, Ryan Miklos, and Jaglin Parmar |
| Giants | Justin Cheng, Jerry Lin, Joseph Lin, and Danny Li |
| Mariners | Joseph Wu, Alvin Shin, Tyler Wu, and Jaehyeon Park |
| Mets | Pooja Ginjupalli, Scott Burgert, Yu Ying Chen, and Mihir Naik |
| Nationals | Eric Digiacomo, Elvin Ly, Jasmine Ngo, and Anna Louie |
| Orioles | |
| Padres | Amy Li, Arsh Maklai, Dylan Kirchner, and Chaeeun Kyung |
| Pirates | Zhenbin Lin, Stephen Chang, James Lee, and Vincent Dong |
| Rockies | Hannah Lee, Christine Song, Meha Dhyani, and Xinyue Su |
| Royals | Benjamin Tang, Steven Yan, Chunqiang Chen, and Leo Zhao |
| Tigers | Karen Zhao, Bo Kim, Junsung Hwang, and Kevin Darby |
| Twins | Seongwoo Hong, Yize Dai, Jake Alok, and Brandon Koo |
| Yankees | Dylan Nicastro, Jade Chen, Jake Alessi, and Thomas Wilk |

2

## Reading

- MGGG Web site
  - https://mggg.org
  - Start to become familiar with GerryChain and maup

3

## Overall Project Goals

- Analyze the impact of potential revisions to the Voting Rights Act (VRA) resulting from a current Supreme Court case
  - Generate ensembles for 2 states simulating “race blind” redistricting
  - Generate ensembles for 2 states simulating VRA redistricting
  - Compare ensembles for each state
- Determine to what extent your states exhibit racially polarized voting (multiple techniques)
- Determine if any state exhibits gerrymandering

4

## Session Objectives

- Understand issues and terminology used in US congressional redistricting and voting analysis
- Understand the ensemble generation process
- Understand data requirements to support analysis of district plans

We will explore the project functionality in more detail in the next 1-3 class sessions

5

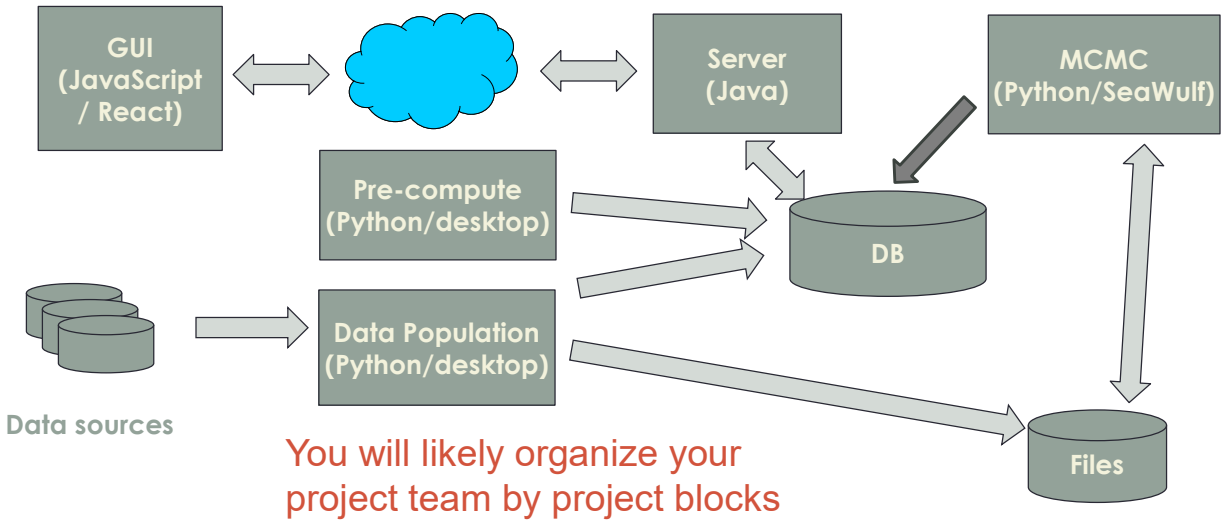
## Project Technical Goals

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

6

 7

## Top-Level System Architecture




You will likely organize your project team by project blocks

7

 CSE416 – Software Engineering 8

## History of Voting in the United States

- Right to vote is central idea in US democracy…
  - … but very limited at start of Republic (1789)
- Voting rights were left for states to decide
  - White males with land; some enforced religious tests – only Christians
  - Slowly, voting rights expanded to White men without land

Article I, Section 4, Clause 1:

> *The Times, Places and Manner of holding Elections for Senators and Representatives, shall be prescribed in each State by the Legislature thereof; but the Congress may at any time by Law make or alter such Regulations, except as to the Places of chusing Senators.*

8

# Voting Rights

- What were reasons for limiting voting rights?
  - … given the *Constitution* states people were created equal.
  - Fear of tyrannous majorities
  - James Madison:
    - "The right of suffrage is a fundamental Article in Republican Constitutions. The regulation of it is, at the same time, a task of peculiar delicacy. Allow the right [to vote] exclusively to property [owners], and the rights of persons may be oppressed... . Extend it equally to all, and the rights of property [owners] ...may be overruled by a majority without property...."

9



# Civil War Era

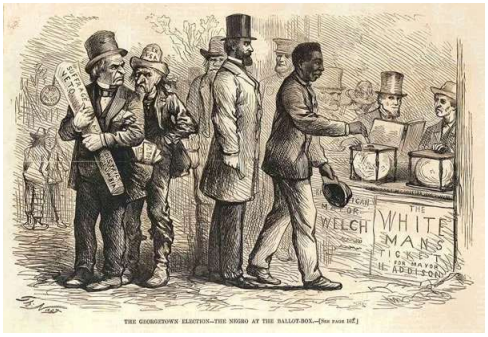
- Civil War - 1861-1865
- Prior to the war, enslaved people did not have the right to vote
- War began as a conflict to preserve the Union
- Evolved as a fight over the issue of slavery
- Resulted in the Emancipation Proclamation and the 13th Amendment (1865)

10

# Reconstruction Era

- North enforces US *Constitution* in the South after the Civil War (1867-1877)
  - African Americans were elected to offices in South (e.g., state senate, town mayors, sheriffs)
  - Racial tensions increased as white southerners formed racist organizations (e.g., KKK)




11

# Reconstruction & Voting Rights

- 1868: ratification of the 14th Amendment
  - Grants full citizenship rights (e.g., voting) to all men born or naturalized in US
- 1870: ratification of the 15th Amendment
  - Voting could not be denied on the basis of *race**
  - Native Americans, however, were not recognized as full citizens until 1924 with the Snyder Act

12

## Jim Crow Era

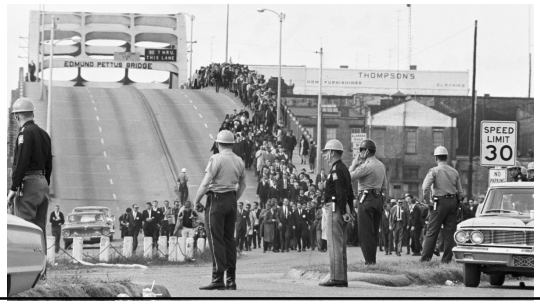
- 1877-1964
- 1877 - North withdrew forces
- Restrictive voter laws passed
  - Revised state constitutions
  - poll taxes, literacy tests, etc.
- SCOTUS Case: *Plessy v. Ferguson* (1896)
  - "separate but equal"; upheld racial segregation

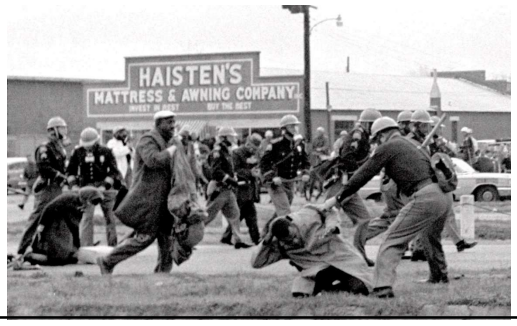

13

## Civil Rights Movement

- Occurs during the 1950s and 1960s
- Major inflection point:
  - March 7th, 1965 – march from Selma to Montgomery violently disrupted by Alabama state troopers
  - Violence captured on national television

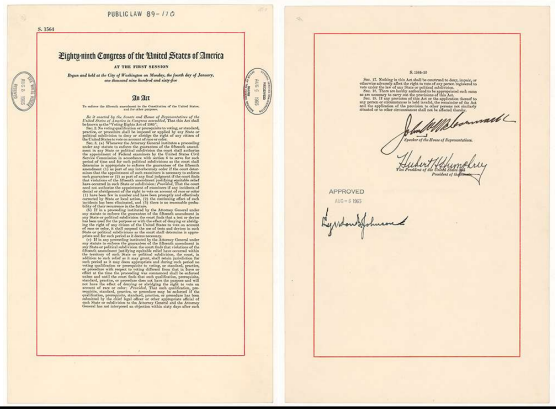



14

# Voting Rights Act …

- Congress passes VRA in 1965
  - Multiple targets:
    - Prohibit states (and local governments) from passing laws to discriminate the right to vote


PUBLIC LAW 89-110

S. 1564

Eighty-ninth Congress of the United States of America

AT THE FIRST SESSION

An Act

APPROVED

15



# …Voting Rights Act …

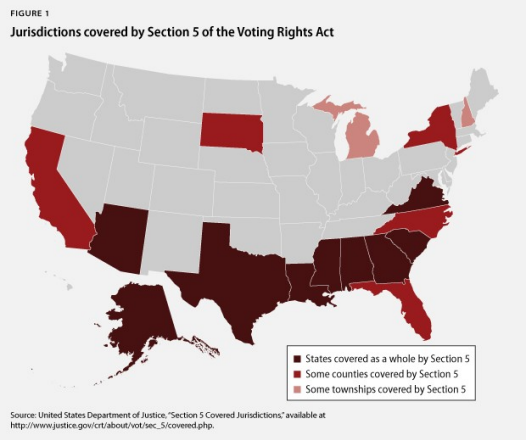
- Major provisions in the bill:
- "AN ACT To enforce the Fifteenth Amendment to the Constitution of the United States, and for other purposes."
- Section 2: "No voting qualification or prerequisite to voting, or standard, practice, or procedure shall be imposed or applied by any State or political subdivision to deny or abridge the right of any citizen of the United States to vote on account of race or color."

16

 CSE416 – Software Engineering 17

# … Voting Rights Act

- Defined "preclearance" voting regions
  - Section 5: Any changes to election practices (voting laws, redistricting, etc.) need to be approved by DOJ
- VRA expanded three times:
- 1975: language minorities
- 1982: voters with disabilities
- 1993: voter registration at DMV locations

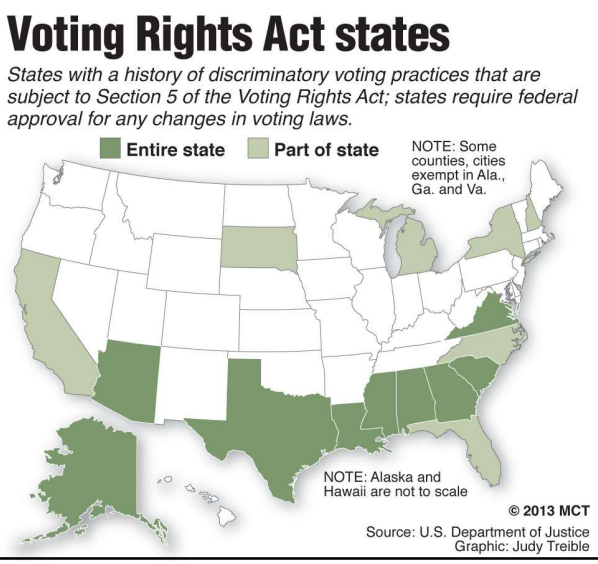



17

 CSE416 – Software Engineering 18

# *Shelby County v. Holder*, 2013

- Unconstitutional to use "coverage formula in Section 4(b) of the VRA
  - (Mostly) Southern states do not need preclearance from the DoJ for new voting changes




18

# Voting Rights Act Today

- Possible gutting of VRA because of a current Supreme Court case (Louisiana v. Callais)


19

# Louisiana v. Callais …

- Current Supreme Court case – oral arguments in October 2025
- Louisianna Population
  - 62.3% white
  - 32.6% black
  - 2.1% Asian
  - 7.8% Hispanic (can identify as white, black, Asian, etc.)
- Louisianna Congressional delegation
  - 6 representatives
  - New district plan in 2022 with 1 black district,
  - Changed in 2024 to 2 majority-black districts with 2 black representatives (Troy Carter and Cleo Fields)

Louisiana v. Callais …

20

## Louisiana v. Callais …

- Does the creation of a second majority-Black district, even to comply with the VRA, violate the 14th or 15th Amendments
- Arguments
  - Challengers (Non-Black Voters): Argue that drawing districts based on race is inherently discriminatory and unconstitutional, even with good intentions, citing Shaw v. Reno
  - Defenders (Black Voters & State): Argue the second district is necessary to ensure Black voters have equal political power, as required by the VRA, and that the state must remedy past dilution
- Eventual ruling may weaken VRA

Arguments from Google AI

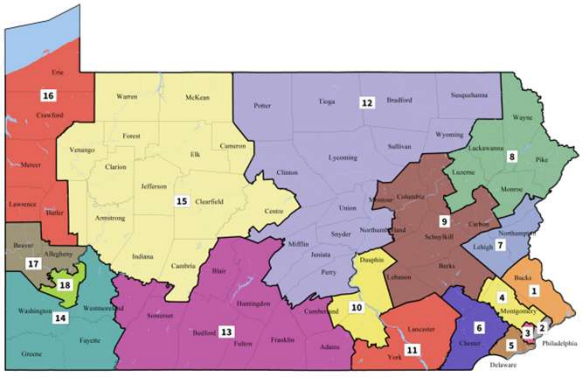
21

## Background Info

- Every state has one or more congressional districts, proportional to the state population
- Population is recalculated after a US Census.
- District boundaries must be recalculated if the number of representatives change or population shifts

Population in districts within a state must be almost equal



Redistricting now possible without a new census

22

## Graph Partitioning View of Redistricting

- Redistricting problem is a classic problem in computer science and applied math
- Think of geographic regions in a state as nodes in a graph
- A district plan is a partition of the graph into n sub-graphs, where n is the number of congressional districts
- Constraints on the partitioning are imposed by law (court decisions, state constitutions, Voting Rights Act, etc.)

One of the quality measures is whether a given redistricting appears to be a random sample from among all possible partitions

23

## Analysis Approach

- For both the race blind district generation algorithm and standard ReCom algorithm
  - Generate a random sample of district plans
  - Apply fairness measures (e.g., Republican/Democratic splits) to the random sample
- Compare fairness measures
- Compare measure of current and proposed plans to the ensemble of random plans using the fairness measures

24

## Graph View of the Project

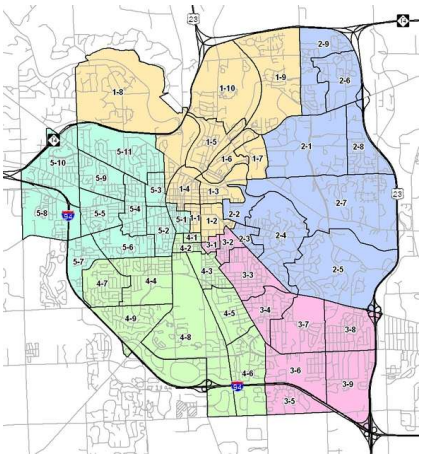
- Each team will work with 2 states
- Each graph (i.e., state) will have approximately 5,000 - 25,000 nodes
- Each node will have approximately 1-10 edges
- Each graph will require (depending on state size) 6-52 partitions (i.e., districts) for the baseline ensemble

25

## Precincts (sometimes known as Wards)

- Lowest level voting division
- Contained in one polling place
- Data usually available for voting totals

*Sometimes multiple precincts share a polling place*

26

 CSE416 – Software Engineering 27

## System Background

- Current redistricting approach leads to many unfair practices (Gerrymandering)
- Some US states have a history of denying equal voting access to minority groups (e.g., African American)
- Unfair approaches usually involve "packing" minorities or political opponents into a small number of districts, thereby minimizing their overall representation

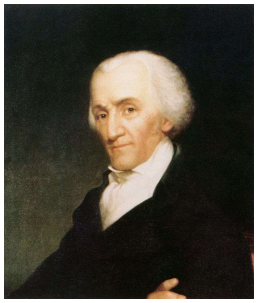
*"I propose that we draw the maps to give a partisan advantage to 10 Republicans and three Democrats because I do not believe it's possible to draw a map with 11 Republicans and two Democrats." – Chairman of NC House redistricting committee*

27

 CSE416 – Software Engineering 28

## What is a Gerrymander?

- The name refers to a voting district that might resemble a salamander
- Named after Elbridge Gerry, 5th VP of US

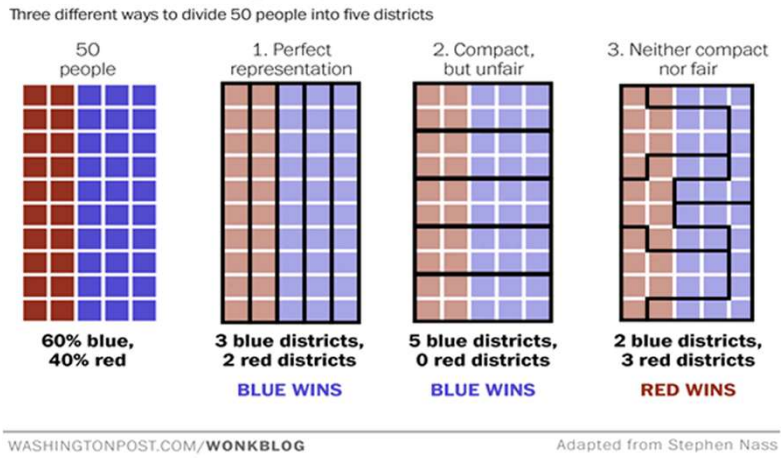



28

# How Gerrymandering Works

- With the same statewide vote share, congressional seat share can vary widely due to districting
- Example – 50 voters, 5 districts

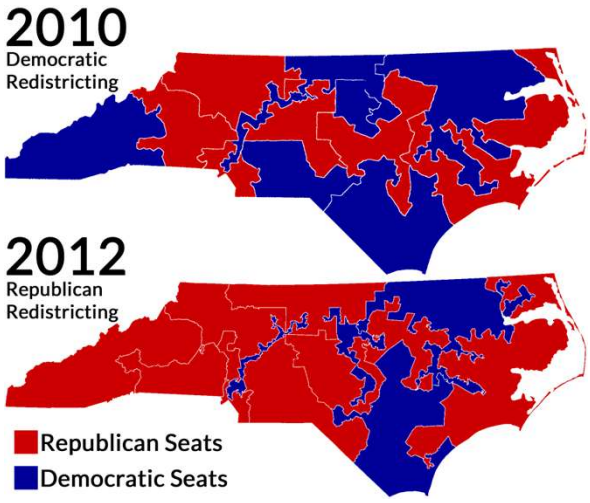



29

# Why is Gerrymandering a Hot Topic?

- Gerrymandering is a practice intended to establish an advantage for a particular party or group by manipulating district boundaries
- Usually features “packing” (grouping categories of voters into a handful of districts) and “cracking” (splitting categories of voters into many districts)
- Occurring since the early days of the US
- Used aggressively in 2010, resulting in congressional dysfunction
- Continued after 2020 census




Definition from Wikipedia

30

## Gerrymandering Enabled By

- Mapping software and big data analysis tools (e.g., voter ID analysis)
- Supreme Court decisions
  - Shelby County v. Holder (2013) – struck down parts of the Voting Rights Act, specifically requirements for pre-clearance of voting changes for certain states and regions
  - Rucho v. Common Cause (2019) – ruled that questions of partisan gerrymandering were non-justiciable
- Post-2020 state supreme court decisions demonstrated inconsistent results

31

## Consequences of Current Gerrymandering

- Many congressional seats are not competitive (only 36 of 435 races in 2022 decided by a margin of 5% or less)
- Members of congress are more concerned with a primary battle than an election battle
- Primaries tend to emphasize extreme candidates
- Congressional representatives represent their party's position more than the wishes of their constituents
- Extremes of each party dominate, instead of the middle

32

## Plan Analysis Techniques

- How do we evaluate a proposed district plan to determine if it is "fair" ?
- Note: historically unfair district plans
  - Underrepresent the voting importance of minority groups (e.g., African-Americans)
  - Underrepresent the voters of a party not responsible for redistricting
  - Generate unlikely safe districts for incumbents

In most states, the party in power is responsible for the district plan

33

## Markov Chain Monte Carlo (MCMC)

- Markov Chain – A sequence of possible events in which the probability of each event depends only on the previous event
- Markov Chain Monte Carlo – a class of algorithms for sampling from a probability distribution
  - Step by step algorithms in which each step is random
  - After a sufficient number of steps, the state is a random sample from the underlying probability distribution

Think of a Markov Chain as a random walk

34

## MGGG ReCom Algorithm - Geographic View

- Start with any district plan
- Combine the districts
- Split the combined district into two randomly selected districts such that the new districts are nearly equal in population and contiguous
- Continue for many iterations
- Resulting district plan is considered random

Code is available at MGGG

Constraints do not alter the randomness of a resulting district plan

35

## MGGG ReCom Algorithm – Graph View

1. Start with an initial valid partition (districting plan) of the dual graph.
2. Select two adjacent sub-graphs from the current plan.
3. Merge the two chosen sub-graphs to form a single, larger sub-graph.
4. Generate a random spanning tree within this merged sub-graph.
5. Identify an edge in the spanning tree that can be removed to split the sub-graph into two new, connected components.
6. Check if the new components satisfy population equality constraints (e.g., within 1% or 2% of the target).
7. If acceptable, keep the new plan; if not, reject and return to Step 2.
8. Repeat the process for a large number of iterations (e.g., 10,000+).

36

## Spanning Tree Properties

- Vertex and Edge Count: A spanning tree has the same number of vertices (\(n\)) as the original graph and exactly \(n-1\) edges.
- Connectedness and Cycles: It is a connected graph with no cycles.
- Minimally Connected: It is a minimal subgraph that connects all vertices; removing any edge makes the graph disconnected.
- Multiple Spanning Trees: A connected graph can have more than one spanning tree.
- Minimum Weight: In a weighted graph, the MST is the spanning tree with the lowest total edge weight, less than or equal to any other possible spanning tree.

37

## Random Ensembles

- By rerunning the Recom algorithm many times (e.g., 10,000), we generate an ensemble of plans that can estimate of the probability distribution of district plans in a state
- Ensembles are typically used to analyze the percentage of a category of voter (e.g., African Americans) in districts
- Can also be used to study the possible range of district voting (e.g., Rep/Dem splits) in an election

You will reuse and modify MGGG code that implements the MCMC algorithm and the VRA constrained version

38

## Further Constrained ReCom

- "Computational Redistricting and the Voting Rights Act" paper
  - Defines modifications to ReCom to accommodate the VRA
  - Link to paper shown as Reference #1 on the CSE416 home page
  - Modified ReCom will generate random district plans that conform to VRA

You will follow the analysis of the paper, but use some simplifications (e.g., election simulation)

39

## High Level View of the Plan Analysis

- Build a robust system to
  - Calculate a set of 5,000 district plans (ensemble) for each state selected by the team (using MGGG software)
  - Calculate a set of 5,000 VRA constrained district plans (ensemble) (using the same states)
  - Estimate the votes in each plan based on historical data
  - Estimate the set of African-American districts in each
  - Calculate fairness results (e.g., political) for each of the ensembles
  - Display fairness results
  - Compare with proposed and enacted plans
  - Compare and display the comparison of the two ensembles for each state

40

## Project Requirements Analysis

- You will use detailed requirements (use cases)
- Requirements will evolve over the first 3-5 weeks of the project
  - Top-Level functional requirements provided in project overview
  - You develop more requirements as you design your GUI
  - Requirements analysis assessed during your GUI Review
  - Requirements aggregated into a master use-case list that evolves during the semester

41

## What Data is Needed?

- Geospatial boundary data
  - Precincts
  - Existing Congressional districts
  - Proposed Congressional districts if one of your states is currently undergoing redistricting (e.g., California)
  - Possibly census blocks (including demographic data)
- Election results data
- Population data
- And more

Start gathering your data now

42

# State Selection

- Your system will analyze 2 states of your choosing with at least
  - One preclearance state
  - One non-preclearance state with a minority population greater than 20%
- You can change any state during the semester if it is available
- Send me an e-mail with your team's choices. Be sure to include alternate selections in case your first choice is already filled

**CSE416**
**State Allocation for Project Teams**

The table below contains the states whose data will be used by each team for analysis. Each tea should select 2 states, one non-preclearance state and one preclearance state. A maximum of 3 t will be permitted for any state. Florida is not officially a preclearance state, but we give it that designation because of previous voting issues (e.g., 2000 Presidential).

| Non-preclearance States | Team 1 | Team 2 | Team 3 |
|---|---|---|---|
| Arkansas | | | |
| California | | | |
| Colorado | | | |
| Delaware | | | |
| Illinois | | | |
| Iowa | | | |
| Maryland | | | |
| Massachusetts | | | |
| New York | | | |
| Oklahoma | | | |
| Oregon | | | |
| **Preclearance States** | | | |
| Alabama | | | |
| Arizona | | | |
| Florida | | | |
| Georgia | | | |

43



# Sources of Data

- Project Web site suggests many sources of data
- The most accurate data source could be
  - Redistricting Data Hub
  - US Census Bureau
  - State Election Office
  - US Government repository of region borders
- Easier sources of data (including some consolidation) are available
- Sometimes difficult to locate the best source of data

44

## What Skills Do You Need?

- Programming (Java, JavaScript, Python)
- Client/server interaction (e.g., Spring, JAX-RS)
- Graph algorithms (e.g., spanning tree)
- Data serialization (migration of data –client/server/SeaWulf)
- Performance analysis (parallel speedup measurement)
- Map system integration
- Client data display
- Client framework (e.g., React)
- DB
- And more

Free SW libraries are available for everything you need

Almost impossible for a team to have all these skills

45



## What Can Your Team Do Now?

- Select states
- Gather best sources of data for those states
- Develop client software for geographic display of US and your selected states (Leaflet?)
- Integrate and display district plan boundary data into your states
- Read the referenced paper
- Download and understand MGGG code
- Select a starting plan for your MGGG code in the format needed by the code

46

## Have You Satisfied the Objectives?

- Understand issues and terminology used in US congressional redistricting and voting analysis
- Understand approach of FRA to generation of Congressional district plans
- Understand some of the data requirements to support analysis of district plans

47