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 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

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

2

Reading

- Chapter 7, Data Analysis for Political Science, Robert Kelly, 2025 – hard copy distributed to each team in class

3

Project Goals

- Build a web-based system that allow for the analysis of data concerning Federal elections
  - Voter registration data
  - Voting equipment data
  - Geographical voting jurisdiction data
  - Demographic data
- Analysis includes
  - Data accuracy
  - Indications of voting process errors
  - Indications of voting process unfairness (racial, political, etc.)

4

Project Topic in the News

- Wall St Journal – 8/18/25
- Wasington Post – 7/16/25

DOJ hits states with broad requests for voter rolls, election data

Election clerks in both parties, already facing harassment and lawsuits over Trump's false 2020 election claims, worry about efforts to examine voting machines.

July 16, 2025

10 min Summary 2,557

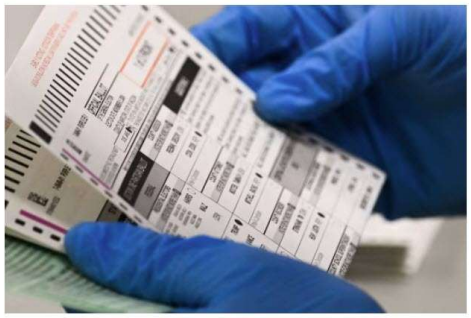
Trump Says He Wants to Get Rid of Mail-In Voting

President continues his on-again, off-again battle with mailed ballots in the lead up to tight midterm elections

By John McCormick Follow and Victoria Albert Follow

Updated Aug. 18, 2025 3:38 pm ET

Share Resize 699 Listen (3 min)



An election worker opening mailed-in ballots last year in Phoenix. PHOTO: PATRICK T. FALLON/AGENCE FRANCE-PRESSE/GETTY IMAGES

President Trump's back-and-forth relationship with mail-in voting has taken its latest turn, with the Republican again lambasting the practice and suggesting he will push

5

Software Engineering Data Analysis Project

- Typical issues with data irregularities –missing data, inconsistent data, typos, etc.
- Data cleaning – Improve data accuracy, where possible
- Multiple data sources – integrate data from different sources
- Analysis of data – analyze the data to uncover trends and visualize information contained in the data

6

What Will You Do

- Examine multiple data sources (e.g., Census, State Secretary of State, Election Assistance Commission (EAC), etc.)
- Download data
- Design and build a database (maybe multiple DBs)
- Build a Web GUI that allows a user to obtain meaningful information from the DB
  - Tabular data
  - Visualizations
  - Geographic displays

7

U.S. Voting Data Analysis

- Began in earnest around 2010
- Increasing analysis activity through this year
- Primarily used today to change the voting process to provide an advantage to a political party (e.g., Republicans in Texas and Democrats in California)

How do we analyze available data to accomplish these goals

8

Project State Allocation

- Your will analyze and display summary data for all mainland states and detailed data for a few states
- Mainland data available in the EAC/EAVS dataset
- You will include more detailed data for a few states you select – according to some set of state features
- Examples
  - Detailed voter registration data
  - Detailed voter equipment data
  - Demographic data (e.g., racial)
  - Political data
  - Geographic data

Your options to select states will be available by week 3

9

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).

10

 CSE416 – Software Engineering 11

Example of Election Data Analysis

- 2000 US Presidential election
- Initial results – 270 electoral votes required to win
  - George W. Bush – 246 electoral votes
  - Al Gore – 267 electoral votes
  - Florida was the one remaining state– a virtual tie
- Data analysis results
  - One county with unusual number of votes for a third-party candidate
  - One county with an unusual number of overvotes (multiple votes for president)
  - Voting irregularities among African American population

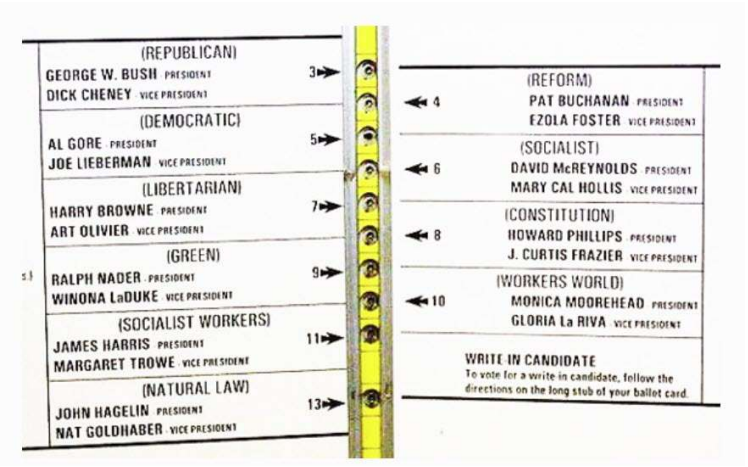
Data errors greatly exceeded the vote margin

11

 CSE416 – Software Engineering 12

2000 Presidential Data Analysis

- Ballot design in Broward County contributed to voter errors
- Not all ballot holes were punched clear through

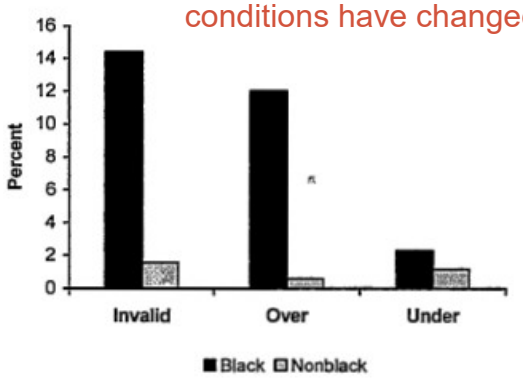


12

Florida 2000 Voting Data Analysis

- 2001 report of the US Commission on Civil Rights
- Findings
  - Significant voter disenfranchisement
  - 14.4% black voters - ballots rejected, 1.6% white voters – ballots rejected
- African-American counties had
  - Old equipment
  - Long lines
  - No supervisor to resolve issues

You will do similar data analysis to see if conditions have changed



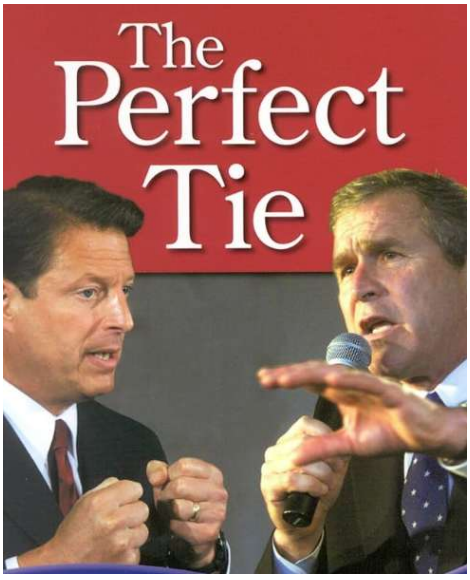
Ballot rejection rate by race

13

2000 Presidential Election Result

- Months of legal challenges and data analysis
- Recount effort ran out of time without determining a clear winner
- Supreme Court stopped the recount, resulting in Bush declared as the winner
- Subsequent data analysis made recommendations to lessen the chances of a reoccurrence

Many of these voting problems remain today



14

2000 Election Data Analysis

- 2000 Florida Presidential election highlights the major (conflicting) goals of people who analyze political data
  - Identify flaws in the election process resulting in new laws and procedures that improve the democratic structure of elections (e.g., flaws in punch card ballots)
  - Identify features of the election system that favor one political party (e.g., African-American voter registration) and change the voting process to favor that political party

Ecological inference
mathematical technique used
to analyze data

15

Voting Process

- Voter registration
- Military and overseas voting
- Early in-person voting
- Mail voting
- In-person voting
- Vote authorization
- Provisional ballots
- Vote count/recount

16

Legal Framework

- US Constitution - "The Times, Places and Manner of holding Elections for Senators and Representatives, shall be prescribed in each State by the Legislature thereof."
- National Voter Registration Act of 1993 –
  - Voter registration standard
  - States must offer mail voter registration
  - States must maintain voter registration lists
- Help America Vote Act (HAVA)
  - Created the Election Assistance Commission (EAC)
  - Replaced punch card and lever systems
  - Set up a mechanism for certification and testing of voting equipment
  - Required statewide voter registration lists
  - Required provisional voting

17

1. Voting Process – Voter Registration

- Potential voters must register to vote
- Voter registration lists available from states (contains name, address, political party, and more)
- Other registration data available from the EAC
- Currently, more than 211M people registered to vote

We will discuss registration further in Part 2

State	Availability	Price Estimate
Alabama	Open	$37,000
Alaska	Open	$20
Arizona	Mixed	$516
Arkansas	Open	$0
California	Mixed	$15,000
Colorado	Open	$50
Connecticut	Open	$300
Delaware	Mixed	$25
District of Columbia	Open	$2
Florida	Open	$0
Georgia	Open	$250
Hawaii	Mixed	$500
Idaho	Open	$20
Illinois	Restricted	$500
Indiana	Open	$0
Iowa	Mixed	$1,100

18

EAVS Dataset

- The EAVS dataset
  - Distributed by the EAC
  - Contains every states voting process data
  - Described in a codebook
- Download files contain detailed information about voting in multiple years
- Data organization is described in a codebook

EAVS Codebook

A5a	Numeric	A5a Mail New Reg
A5b	Numeric	A5b In-person New Reg
A5c	Numeric	A5c Online New Reg
A5d	Numeric	A5d Automatic New Reg
A5e	Numeric	A5e DMV New Reg
A5f	Numeric	A5f NVRA Mandated New Reg
A5g	Numeric	A5g Disabilities Agency New Reg
A5h	Numeric	A5h Armed Forces New Reg
A5i	Numeric	A5i Non-NVRA Mandated New Reg
A5j	Numeric	A5j Registration Drives New Reg
A5k	Numeric	A5k Polling Places New Reg

You will create a database containing most of the information in multiple EAVS files

19

2/3/4. Early Voting

- Early voting increasing
- Pressure to lessen or eliminate mail voting
- Voter turnout issues
- Ballot rejection issues

C9a	Numeric	C9a Total Mail Ballots Rejected
C9b	Numeric	C9b Mail Ballots Rejected Because Late
C9c	Numeric	C9c Mail Ballots Rejected Because Missing V
C9d	Numeric	C9d Mail Ballots Rejected Because Missing V
C9e	Numeric	C9e Mail Ballots Rejected Because Non-Mat
C9f	Numeric	C9f Mail Ballots Rejected Because Unofficial
C9g	Numeric	C9g Mail Ballots Rejected Because Ballot Mis
C9h	Numeric	C9h Mail Ballots Rejected Because No Secre
C9i	Numeric	C9i Mail Ballots Rejected Because Multiple E
C9j	Numeric	C9j Mail Ballots Rejected Because Envelope
C9k	Numeric	C9k Mail Ballots Rejected Because No Postm
C9l	Numeric	C9l Mail Ballots Rejected Because No Reside
C9m	Numeric	C9m Mail Ballots Rejected Because Voter De
C9n	Numeric	C9n Mail Ballots Rejected Because Voter Alr
C9o	Numeric	C9o Mail Ballots Rejected Because Missing D
C9p	Numeric	C9p Mail Ballots Rejected Because Voter No

20

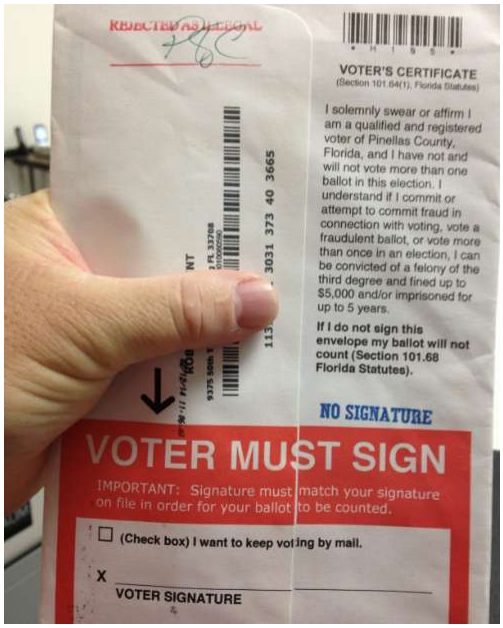
5. In-Person Voting

- States responsible for selection of voting equipment
- Issues
  - Outdated voting equipment
  - Sufficient voting equipment
  - Distribution of voting equipment for political/demographic reasons (e.g., Florida 2000)

21

6. Vote Authorization

- Reasons for rejected ballots
  - Questionable signature
  - Insufficient documentation
  - Challenged by poll watchers
  - Voter not in registration list
- When an in-person ballot is rejected, voter should be able to cast a provisional ballot that, if needed, is verified during ballot curing
- Process varies among states



22

 CSE416 – Software Engineering 23

7. Vote Counting

- Most vote counting begins once the polls close
- Some states count mail ballots before polls close
- Many states accept ballots postmarked by Election Day
- Issue of local vote counting vs regional vote counting

23

 CSE416 – Software Engineering 24

8. Recount

- Recount performed, if necessary
- Usually if the vote difference is less than 1%
- Raises the importance of voting equipment that provides paper backup

Paper ballots lead to inaccurate recounts

24

Project Tasks

Aawab can provide some assistance in your schema design – use Piazza

- You will construct a database from the EAVS materials (2016-2024) for all states
- DB will be used to respond to user queries (e.g., trends in mail voting, % of disallowed ballots, etc.)

EAVS Reports and Materials

- 2024
- 2022
- 2020
- 2018
- 2016

Which EAVS data items will you need?
Need to have an adaptive view towards your DB

Will you use a No-SQL or relational DB?

Need to check year-to-year data consistency

25

Voting Equipment Categories

- Optical scan – device that scans a paper ballot marked by the voter
- Direct Recording Electronic (DRE) – interactive device to record votes and transmit vote totals
- Voter Verified Paper Audit Trail (VVPAT) – device generates a paper trail that can be used in the event of a recount
- Ballot Marking Device – voting machine used to record votes on physical ballots. Often used to meet requirement for accessible voting equipment

Ballot Marking Device

You will analyze age of devices by location (e.g., minority)

26

Project Tasks

- Your DB will include general data (from EAVS) for all states, along with more detailed data (from other datasets) for a few states you select
- You will generate detailed equipment data for the states you select based on publicly available information
  - Age
  - Underlying OS
  - Certification
  - Scan rate
  - Reliability

Let the TAs know of any info that is not available, and they will attempt to contact company

27

Project Tasks

- Some use cases will require you to detail the voting equipment used in various locations
- Example – EAVS provides 3 fields for scanner equipment in Kentucky; Laurel County reports 25 ES&S DS200 scanners and 1 ES&S DS450 scanner

EAVS Codebook

F6a	Numeric	F6a Scanner
F6b_1	Numeric	F6b_1 Make and Model 1
F6b_1c	String	F6b_1 Other Make Model 1
F6c_1	Numeric	F6c_1 Number Deployed 1
F6b_2	Numeric	F6b_2 Make and Model 2
F6b_2c	String	F6b_2 Other Make Model 2
F6c_2	Numeric	F6c_2 Number Deployed 2
F6b_3	Numeric	F6b_3 Make and Model 3
F6b_3c	String	F6b_3 Other Make Model 3
F6c_3	Numeric	F6c_3 Number Deployed 3

As you find equipment details, post in Piazza

EAVS Data

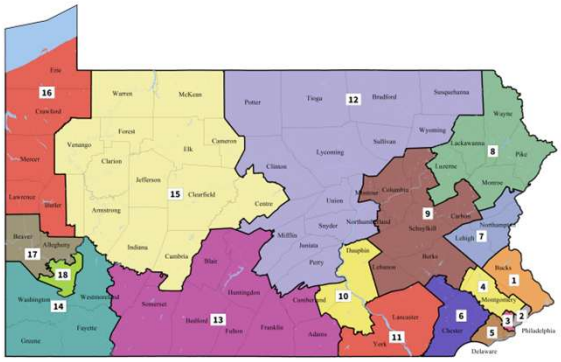
Jurisdiction_Name	St	Stat	F6a	F6b_1	F6b_1othe	F6c_1	F6b_2	F6b_2othe	F6c_2
LARUE COUNTY	KE	KY	Yes	Verity Scan (Hart)		5			
LAUREL COUNTY	KE	KY	Yes	DS200 (ES&S)		25	DS450 (ES		1
LAWRENCE COUNTY	KE	KY	Yes	Verity Scan (Hart)		10			

28

Data Display

- Every state has various geometric districts (e.g., county, precinct, town, congressional district, census block)
- EAVS aggregates data in differing geometric areas (e.g., county, town)
- You will show data values on a map (e.g., choropleth display)

You will display geometry for all states



29

Project Technical Goals

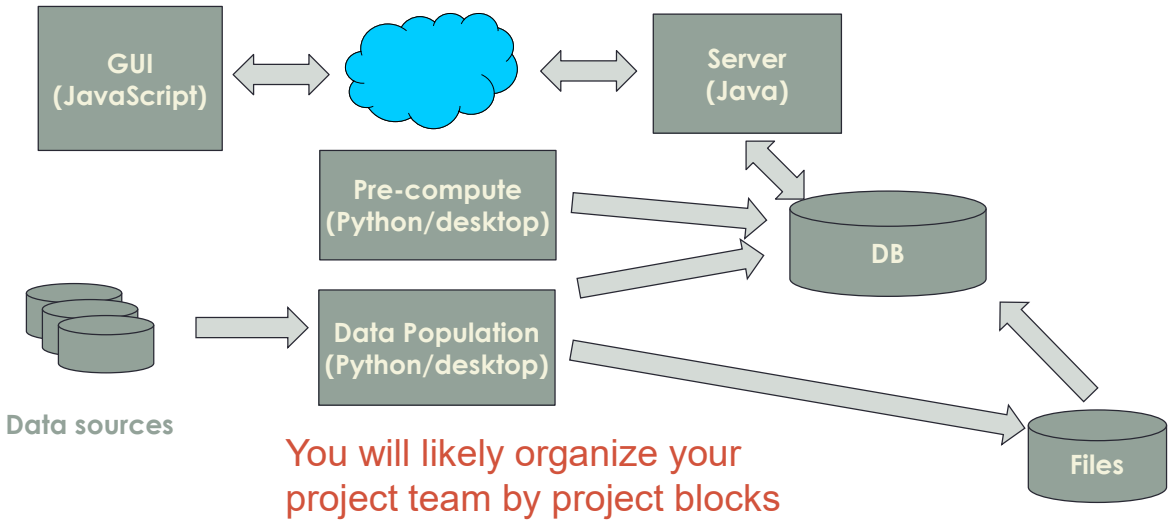
- Build a large, complex system that requires teamwork to integrate multiple components
- Use multiple programming styles and languages (JavaScript, Python, and Java)
- Incorporate
  - Requirements analysis (application domain knowledge)
  - User interface design
  - Data base design and development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

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

30

 31

Top-Level System Architecture



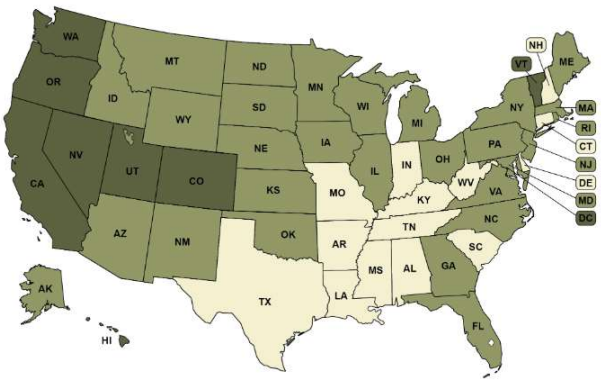
You will likely organize your project team by project blocks

31

 CSE416 – Software Engineering 32

Geographic Considerations

- You will need the geographic boundaries of all mainland states (to display choropleth maps of voting values, such as mail voting)
- You will also need geographic boundaries of voting districts within your selected states (usually found in Redistricting Data Hub)



32

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

33

What Skills Do You Need?

- 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

TAs were selected based on experience with these technologies

Free SW libraries are available

Almost impossible for a team to have all these skills

34

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?

35

How Can You Get Started?

- Talk with your team about differing roles of team members
- Set up team organization (meeting schedule, task tracking tool, virtual meeting technology, etc.)
- Read textbook Chapter 7 to better understand requirements
- Get familiar with a GUI builder (TA session was recorded)
- Download 2024 EAVS data and codebook
- Compare 2024 EAVS data with prior years
- Begin think about DB design and your preference for MongoDB vs. relational DB. Use EAVS data to think about DB design
- Start your state selection process (review state table page)

36

Have You Satisfied the Objectives?

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

37