Absentee Ballot Curing Project
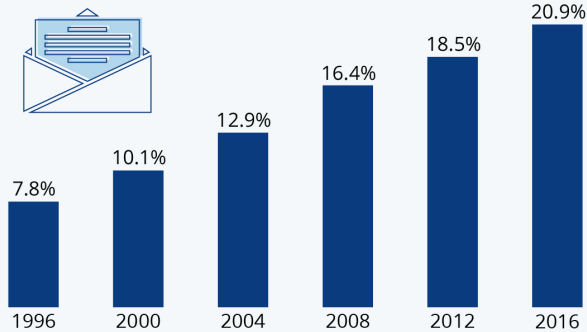
• • •

Andrew Zhang, Brice Halder, Isaac Frumkin, and Nicholas Bowen

Background

Voting By Mail Has Gotten More Common In The U.S.

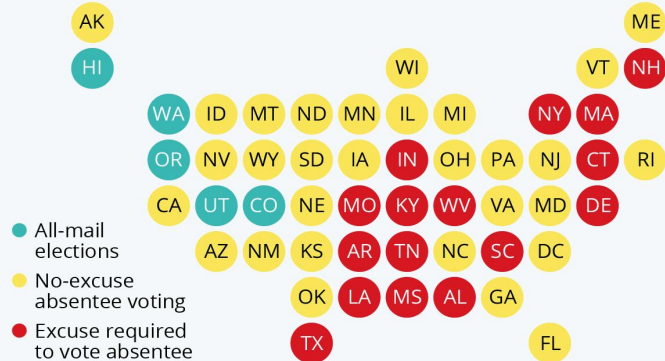Share of U.S. voters who cast their ballot by mail in presidential elections

- 1996: 7.8%
- 2000: 10.1%
- 2004: 12.9%
- 2008: 16.4%
- 2012: 18.5%
- 2016: 20.9%

Source: U.S. Census Bureau

statista

The Current State(s) of Mail-in Voting

Absentee voting rules in U.S. states and the District of Columbia

- All-mail elections
- No-excuse absentee voting
- Excuse required to vote absentee

Source: National Conference of State Legislatures

statista

65,642,049

Mail ballots returned in the November 2020 Presidential Election

Disparities in Mail Ballot Rejections

By Race:

	White	Black	Hispanic	Asian	Other
% of All Rejected	47.1%	35.8%	2.5%	3.3%	9.7%
% of All Returned Ballots	63.1%	26.8%	1.3%	1.5%	7.3%

By Age:

|	18-29	30-44	45-59	60+
% of All Rejected	15.9%	17.7%	25.4%	39.5%
% of All Returned Ballots	7.7%	11.6%	22.4%	58.2%

Non-white voters and younger voters are more likely to have mail-in ballots rejected

Ballot Curing

Ballot curing is the process of correcting a ballot that was rejected due to certain issues that prevent it from being counted in its current form.

Common Issues That Can Be Cured

- Missing Signature
- Invalid Signature
- Additional ID Needed
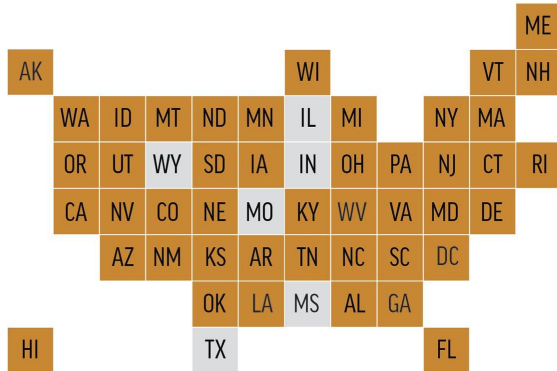- Submitted provisional ballot
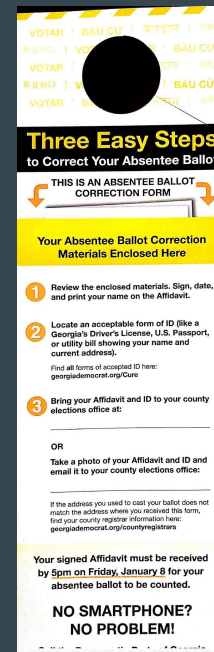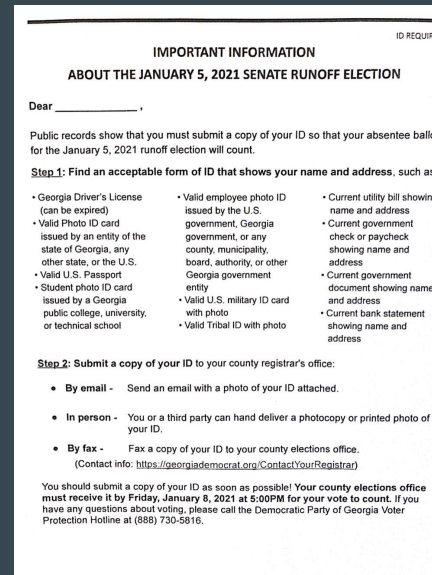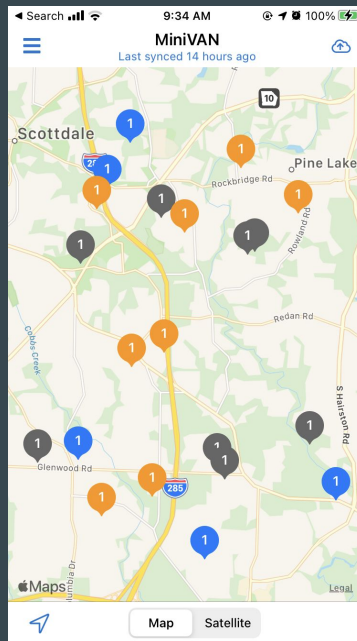
Stakeholders

Tracking mail-in ballots

Most states allow voters to check online if their early votes are received by election officials. If a mail-in ballot is invalidated due to damage or a signature discrepancy, how it gets fixed depends on where you live.

States allowing voters to *track their ballots* after they are returned:

States *requiring a voter be notified* if a signature problem arises on a mail-in ballot:

AZ CA CO FL GA
HI IL IA MA MN
MT NV NJ OH OR
RI UT WA

Sources: AP reports; National Conference of State Legislatures

AP
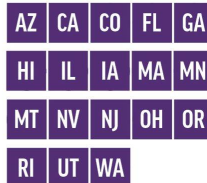
Our Project

Design and implement a software system that will improve the **efficacy** and **transparency** of the ballot curing process

Progression

Market Research	Database Design	Database Insertion & Standardization	Analysis	Frontend & Backend
● Interviewed SOS, Board of Election, and grassroots non-profit offices.	● Designed database architecture ● Determined appropriate schemas for relations	● Script for inserting voting files into database ● Created method for standardizing dissimilar data points/templates across states	● Basic queries to find relevant information about voting statistics ● More specific queries to segment voting statistics by group	● Wireframe and develop frontend with HTML and JS ● Use Flask to execute queries and run scripts to automatically download data dumps

Interviews With State, Party, and Non-Profit Officials and Research

Purpose of Interviews

One of our first actions as a group was to reach out to various state, party, and non-profit organizations in an attempt to discuss with various groups our ideas as well as gain a further idea of the current "lay of the land" in regards to curing ballots.

Ended up being able to interview individuals involved with the Colorado Democratic Party, North Carolina Democratic Party, Colorado Secretary of State's Office, Maryland State Board of Elections, and Common Cause.

Karin Ascenio - Colorado Democratic Party (Volunteer Coordinator)

- Most organizations involved in curing do so in a partisan manner
- The counties send letters out to those who ballots need curing and actually conduct the curing process
- The Democratic Party in North Carolina has a tool called votebuilder that they use to get voting data and find out whose ballots need cured
- Txt2Cure was a tool used in Colorado and especially useful for those that lived in rural areas

Seth Morris - NC Democratic Party (Voter Protection Director)

- Use VoteBuilder (NGP Van) to generate data for curing efforts
  - 501(c)(4)'s can also purchase access
- Data team spent a long time manually filtering
- Legislation around ballot curing subject to change
- Each County Board of Elections handles elections differently
  - Some instances of voters not being notified about rejection
  - Sometimes county systems were wrong

Bruce Norikane - CO Democratic Party (Tech Director)

- Colorado Democratic party has a system to automate the retrieving of ballot information 8 times a day
- Uploads the data into Votebuilder after small amount of processing
- Ballot curing is a labor intensive process that happens more often in smaller districts
- Txt2Cure is an outside system tried by Colorado this year to allow electronic curing of ballots

Colorado Secretary of State Office

- Ballot cures are processed at the county level
- Txt2Cure is offered through a company called GlobalMobil...the data for the cure is sent to the company, who then converts it to a pdf and uploads it to a SFPT serve the counties can access
- Important to keep in mind accessibility for any software that may work with the public
- Security is extremely important anytime you are working with voting data

Izzy Bronstein - Common Cause (National Campaigns Manager)

- National data comes from Movement Cooperative (who buys from Catalist or TargetSmart)
  - Models to predict voter characteristics such as internet access, number of times voted, nationality of parents, and likelihood of owning a gun
  - VAN also gets data from Target Smart
- Catalist and Target Smart have fostered relationships with county election officials across the country to get most up-to-date information
  - County administrators often have better/more updated info than SOS
- Potential for an open-source tool that works like Twilio

Nikki Charlson - MD State Board of Elections (Deputy Admin)

- Maryland is very centralized at the state level with ballot information and procedures for elections - **Top Down** hierarchy
  - Counties access Maryland state database
  - Local election boards follow directives from the state level
- Counties are required to inform voters if their ballots were rejected
- Potential access to voter files for $125

Ideation

1) Ballot Curing Database: provides info to a user of the program about whose ballots need cured. Allows further targeting by geographic location, gender, race, etc. depending on state data

2) Ballot Curing Dashboard: the purpose of this would provide info to a user of the website in the form of various statistics about mail voting/ballot curing, including what areas tend to have more rejections, higher rate of curing, slower time to cure, etc.

3) Contact Voters Whose Ballots Needed Curing : uses the data to find the voters whose ballots need cured and then contact them through either text message or email.

Project Requirements

- As a maintainer of the project, I need scripts for downloading ballot data multiple times a day during an election cycle for each state
- As a maintainer, use the downloaded CSV file and convert it into MySQL database automatically
- As a user, I want to sort through the data to get specific results that I want to see.
- As a user, I want to compare old data with new data over time.
- As a user, I want to download the queries conducted so I can use it for ballot curing.

Database Design

Project Architecture & Schema
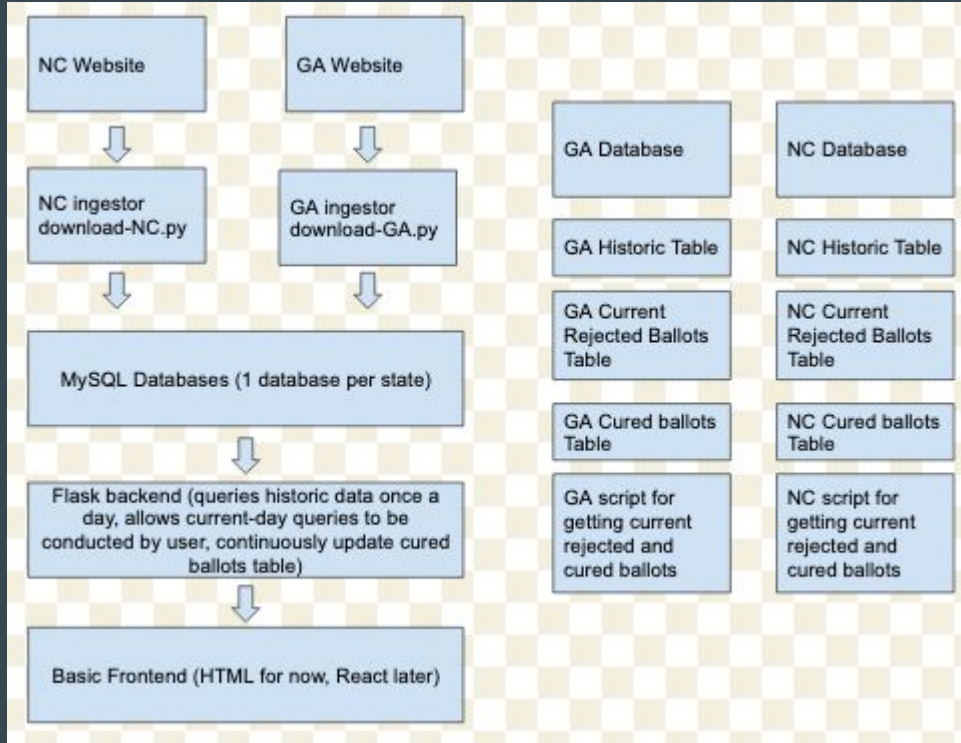
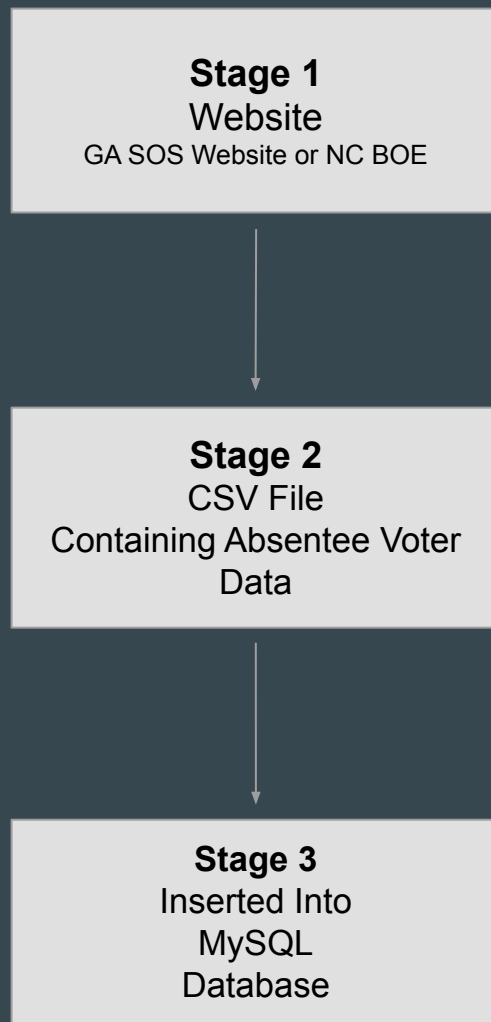| Attribute Name	Data Type	Description
id	INT NOT NULL AUTO_INCREMENT	Primary key
proc_date	DATETIME	Datetime entry was updated
county	VARCHAR(25)	County name/description
voter_reg_num	INT	Voter registration number
first_name	VARCHAR(50)	First name
middle_name	VARCHAR(50)	Middle Name
last_name	VARCHAR(50)	Last name
race	VARCHAR(50)	Race
ethnicity	VARCHAR(50)	Ethnicity
gender	VARCHAR(10)	Gender/sex
age	INT	Age
street_address	VARCHAR(255)	Street Address
city	VARCHAR(50)	City
state	VARCHAR(10)	State
zip	VARCHAR(10)	Zip code
election_dt	DATETIME	Election date
party_code	VARCHAR(10)	Party of voter
precinct	VARCHAR(50)	Precinct name
cong_dist	VARCHAR(50)	Congressional district name
st_house	VARCHAR(50)	State House district name
st_senate	VARCHAR(50)	State Senate district name
ballot_style	VARCHAR(50)	Ballot style (Mailed, in-person, etc)
ballot_req_dt	DATETIME	Ballot request date
ballot_send_dt	DATETIME	Ballot send date
ballot_ret_dt	DATETIME	Ballot return date
ballot_issue	VARCHAR(255)	Issue with ballot
ballot_rtn_status	VARCHAR(50)	Ballot return status

Database Insertion & Standardization

Database Insertion

The data is first contained on either the Georgia Secretary of State Website or North Carolina Board of Education. A zip file is then downloaded from those websites using either Selenium or HTTP Requests, respectively.

Then processing of the zip file is done to unzip it, store the CSV files appropriately, and then destroy the old zip file.

Finally, the data is then processed into MySQL by connecting the Python script to the database, then running an insert SQL command to take the data from the CSV and put it into the actual database.

Stage 1
Website
GA SOS Website or NC BOE

Stage 2
CSV File
Containing Absentee Voter Data

Stage 3
Inserted Into
MySQL
Database

Database Insertion - Maintaining Unique Indices

- Averaging 1 hr to insert
  new day of data after ~17 days

- Faster to re-create table daily
  - avg 8-10 mins

- Only works for states that keep
  historical data in CSV
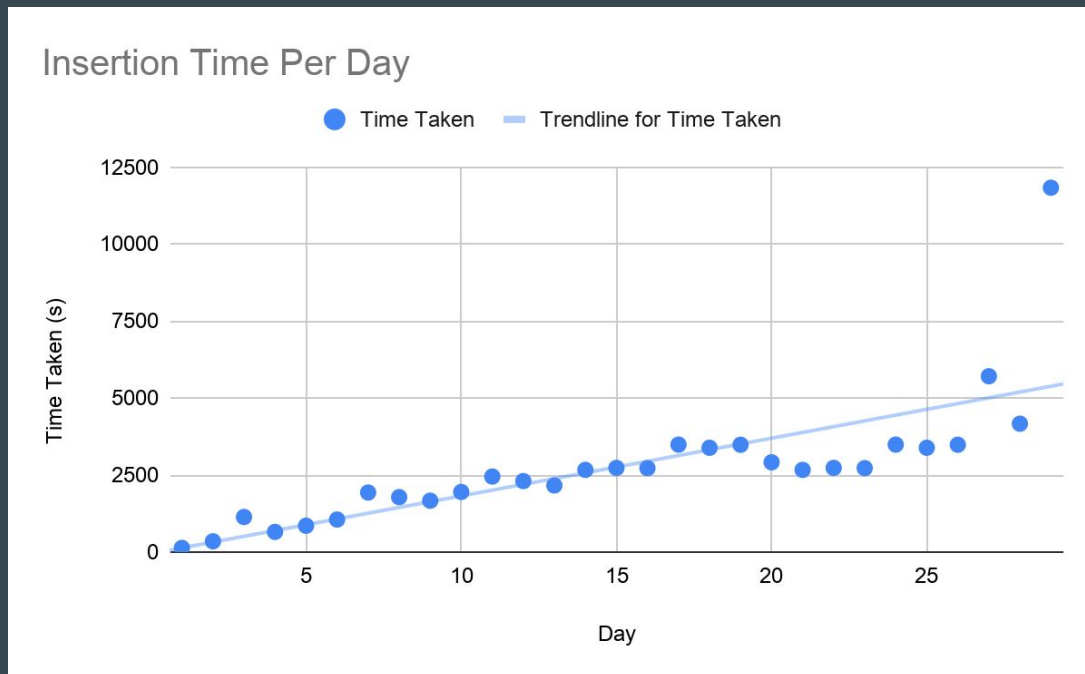
GA - NC Comparison

		Georgia	North Carolina	Notes
1	Race, ethnicity, age, political party data	✕	✓	
2	Separate ballot issue and ballot status	✓	✕	• Georgia's ballot issues were often non-standardized
3	Daily data dump contains cumulative ballot info	✓	✕	• North Carolina's data contains final day results
4	Download via direct link to file	✕	✓	• Georgia required using Selenium

Standardization
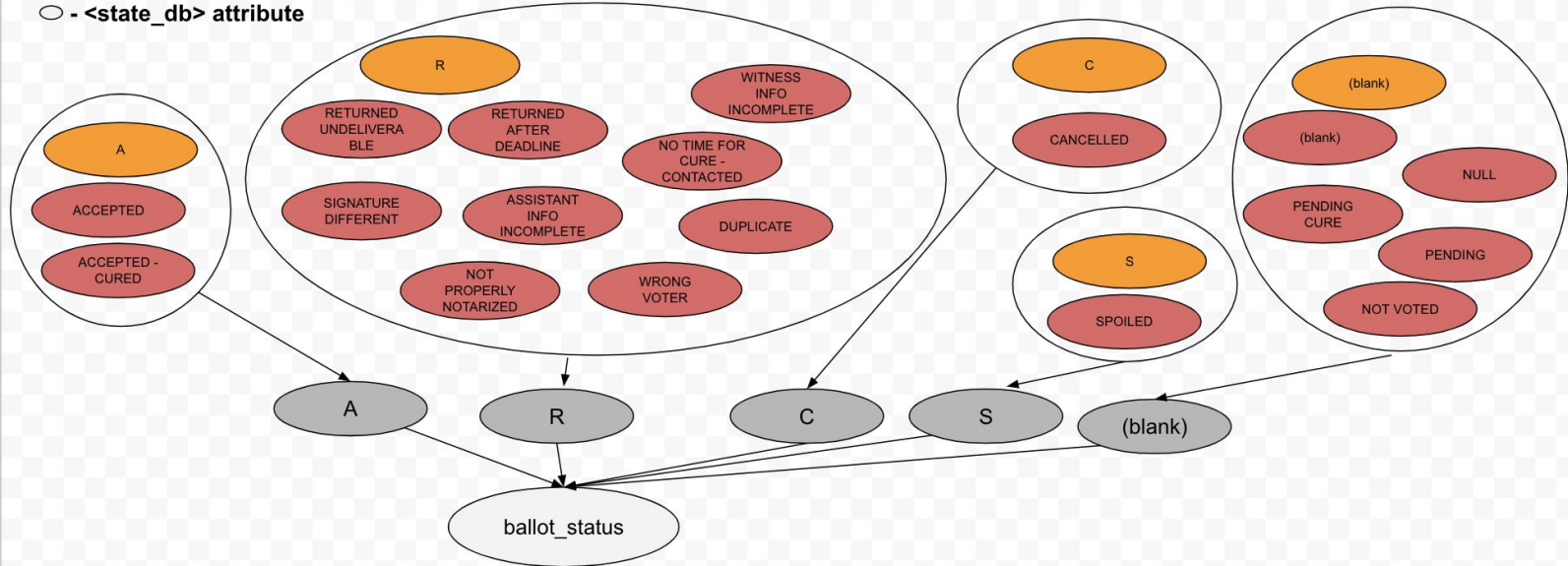
Analysis

Query Results (GA Jan 5 Runoff)

- Total processed **absentee ballots**: 3,459,712
- Total processed **voters**: 3,331,476
    - Total accepted ballots: 3,158,599
    - Total rejected ballots: 6,542
    - Total canceled ballots: 150,467
- Average Entry Per Voter: 1.0385
    - Our prior architecture (add each day of election) resulted in about 29 entries per voter

Query Rejection Results (GA Jan 5 Runoff)

Top 10 counties for rejected ballots

Types of Issues for Rejected Ballots

ballot_issue	COUNT(*)
Ballot Received after Deadline	3555
Invalid Signature	1659
Missing Signature	1232
Ineligible Elector	51
MIDR - ID not Provided	45

county	total
DEKALB	852
GWINNETT	737
COBB	682
HENRY	390
CHATHAM	330
FULTON	256
CHEROKEE	245
FORSYTH	235
CLAYTON	216
FAYETTE	155

Cured Queries (GA Jan 5 Runoff)

As of Dec 28, 2020

- 119 ballots cured
- 1008 ballots still rejected

Anticipating ~**3,000 cured** by the end (3,376 non-cured*)

How querying for cured ballots works:

1. From rejected table, query each rejected ballot in today's snapshot to see if there are any newly accepted ballots
2. From today's data, query for all rejected ballots and add to rejected table

U.S. Elections Project

Preliminary Cured Queries (NC Nov 3 General Election)

Ballots cured overall

- 7,890 Ballots Cured

Ballots cured by race

```
+--------------------------------------+----------+
| race                                 | COUNT(*) |
+--------------------------------------+----------+
| WHITE                                |     5136 |
| BLACK or AFRICAN AMERICAN            |     1675 |
| UNDESIGNATED                         |      659 |
| OTHER                                |      198 |
| ASIAN                                |      134 |
| TWO or MORE RACES                    |       52 |
| INDIAN AMERICAN or ALASKA NATIVE     |       35 |
| NATIVE HAWAIIAN or PACIFIC ISLANDER  |        1 |
+--------------------------------------+----------+
```

Cured by ethnicity

```
+---------------------------+----------+
| ethnicity                 | COUNT(*) |
+---------------------------+----------+
| NOT HISPANIC or NOT LATINO|     5623 |
| UNDESIGNATED              |     2107 |
| HISPANIC or LATINO        |      160 |
+---------------------------+----------+
```

Total cured by county

```
+--------------+----------+
| county       | COUNT(*) |
+--------------+----------+
| WAKE         |      615 |
| GUILFORD     |      507 |
| BUNCOMBE     |      461 |
| BRUNSWICK    |      401 |
| UNION        |      352 |
| DURHAM       |      302 |
| MECKLENBURG  |      289 |
| NEW HANOVER  |      235 |
| CUMBERLAND   |      234 |
| CABARRUS     |      216 |
```

The Rest of the Semester

• • •

Currently we are at an inflection point. We have spent most of our time so far either working directly with the data or doing research. Now, we are beginning to pivot to working with what we actually do with the data once we have it in a database with our web app.
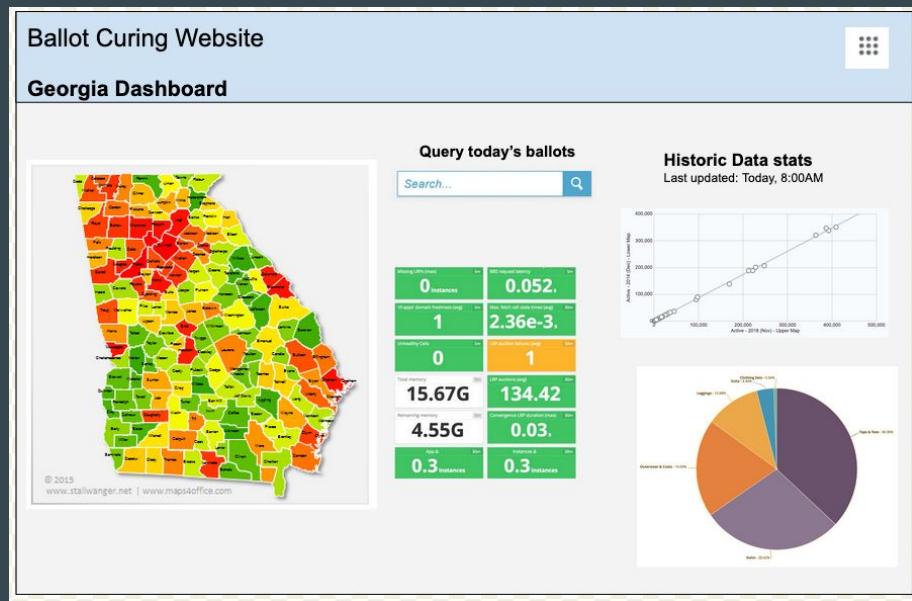
Frontend

Basic HTML + JS

- Allow users to query our NC and GA databases
- Give overview of states and counties on curing efforts
- Allow users to download query data

Backend

Flask

- Accept queries from frontend for queries
- Automatically run scripts daily to download data dumps

Rough wireframe

Current Challenges

- Requesting data from other states taking a long time
- MySQL ingest issues causing some data to be added in different column
- Different states record different attributes (NC has quite a lot
- Many states are bottom-up, so statewide data may not be most updated or correct