

Project 4: Your State / Putting it all together

Goals:

- Same as goals of Project 1, but with more depth, and one of the following:
 - Collect shapefile and/or PDFs for a state, including digitizing, or
 - Continue digitizing from Project 2
- Learn how to process your data and combine spatial data with election results
- Document your data so it can be understood, expanded upon, and used by others
- Contribute to redistricting research, education, and public mapping with your data

Things you are not expected to do:

- Contact counties
- Finish data collection and processing for the entirety of your state

Project 4 Day 1: Creating an Action Plan and Gathering Data

Part I. Pre-planning (10:00-10:15am)

Zoom logistics: “All hands.”

1. Discuss and decide: do you want to continue *searching* for and possibly digitizing new data (building on Project 1), or *digitize* existing data (building on Project 2)?
2. We may combine or split cohorts based on preferences (i.e., if 5 participants in two cohorts want to digitize and 4 want to continue searching, we’ll combine those 5 people into one cohort and combine the remaining participants into another).

Part II. Creating an Action Plan (11:00-11:15am)

Zoom logistics: Cohorts. One person should **facilitate**, in this case meaning making sure everyone’s ideas are heard. Another person should **take notes**.

1. With your cohort, review the data you have and identify an area (county, urban area, entire state) for which you think you can collect and digitize data by mid-afternoon
 - a. Come up with a list of counties, a region, etc.
 - b. Choose a county/region for which you can access election data for at least one recent year. Election data can be from [MEDSL](#) (the MIT Election Data + Science Lab), from your state, or from your county/counties.
 - c. Check in with Ruth and Heather about your project goals – warning: we may try to narrow your goal!
2. Set up a structure for data collection and digitizing in your cohort’s Google Sheet.
 - a. E.g., specify the Searcher in the Project 1 sheet

Part III. Data collection (Project 1 track) and Digitizing (Project 2 track) (11:15am-12:30pm)

Communication: At your discretion -- feel free to use Slack or Zoom to communicate, but otherwise, work independently.

If you chose the Project 1 track, continue filling out the [Project 1 sheet](#) for your state's counties as you go.

If you chose the Project 2 track, begin digitizing.

Part IV. Digitizing (2:00-3:30pm)

Communication: At your discretion -- feel free to use Slack or Zoom to communicate, but otherwise, work independently.

If you chose the Project 1 track, now's the time to switch to digitizing what you found, as necessary. If you finish early, you can help others in your cohort, return to searching for a bit and, depending on what you find, do another round of digitizing.

If you chose the Project 2 track, keep digitizing! If you finish early, you can add another county (check with Ruth or Heather).

Take note of any exceptions, oddities, and so forth in the Notes column of your Google Sheet (Project 1 track) or a separate document you upload with your county shapefile (Project 2 track). These will be important, so don't consider them an afterthought.

Part V. Downloading and formatting election results (3:30pm-4:30pm)

1. Download (tabular) election results by precinct for your county/region/state, using [MEDSL](#), the websites you filled in during the search process, and doing additional searches as necessary. Start with 2018, 2016, or 2014 general election results as available. Complete this section for one election for one year, for the sake of time. If you finish early, add another election and/or another year. Note in the [Project 1](#) spreadsheet what election results are available and where (in the state search and/or counties sheets, depending on what data you find)
2. Using the guidelines below and the sample figures, format your election results so that they are based on political party and year. This may entail looking up which party a candidate was in: Ballotpedia is a good source for this. E.g., Congressperson Hernandez and Congressperson Jones might both be Democrats who ran in the 2016 Congressional

election, so you'll report them both as USH16D. If a third party candidate received less than 5% of the vote, you can remove them – again, Ballotpedia is a good resource.

a. Our election data attribute names follow the convention: *Office Code + Last Two Digits of Year + Party Code*. Office and party codes are listed below. A column containing votes for Hillary Clinton in the 2016 general election, for example, would be named **PRES16D**. We do not, as of yet, have a standard for primary or non-partisan elections.

b. Office Codes:

- **PRES**: President
- **SEN**: Senate
- **USH**: United States House
- **GOV**: Governor
- **AUD**: Auditor
- **COMP**: Comptroller
- **LG**: Lieutenant Governor
- **TRE**: Treasurer
- **AG**: Attorney General
- **SOS**: Secretary of States
- **SSEN**: State Senate
- **SH**: State House

c. Party Codes

- **D**: Democratic Party
- **R**: Republican Party
- **L**: Libertarian Party
- **G**: Green Party
- **I**: Independent
- **U**: Unaffiliated

The following figures illustrate what you might find (top) and what you should convert it into (bottom):

Precinct	Candidate	Office	Party	Votes
Precinct A	Josh Hawley	US Senate	Republican	100
Precinct A	Claire McCaskill	US Senate	Democrat	200
Precinct A	Robert Vroman	US House District 1	Republican	105
Precinct A	Lacy Clay	US House District 1	Democrat	195
Precinct A	Sandra McDowell	Auditor	Republican	102
Precinct A	Nicole Galloway	Auditor	Democrat	198
Precinct B	Josh Hawley	US Senate	Republican	300
Precinct B	Claire McCaskill	US Senate	Democrat	150
Precinct B	Ann Wagner	US House District 2	Republican	290
Precinct B	Cort VanOstren	US House District 2	Democrat	160
Precinct B	Sandra McDowell	Auditor	Republican	250
Precinct B	Nicole Galloway	Auditor	Democrat	200

Precinct	SEN18R	SEN18D	USH18R	USH18D	AUD18R	AUD18D
Precinct A	100	200	105	195	102	198
Precinct B	300	150	290	160	250	200

- Now that your election results are formatted, combine them into a single spreadsheet.

Project 4 Day 2: Election Data, Cleaning, and Merging

Times for today are estimates: some parts might take more or less time. Please make your own copy of this worksheet now (beginning of the day) and use it to take notes as you go -- combined with the edits to the project sheet yesterday, they'll be vital for creating summary documentation at the end of the day!

Communication: At your discretion -- feel free to use Slack or Zoom to communicate, but otherwise, work independently.

Part I. Matching and cleaning (10:00am-12:30pm)

In this part, we're answering the question: do your precinct names in the election results table match the precinct names in the shapefile? If the answer is "no," that's what this section is about: working with the data to address discrepancies. If you choose to use python/pandas for this part see this [pandas reference guide](#) on the github.

- Using Python or Excel, compare the precinct names from the precinct file and the election results. Check whether they match.
- If they all match, you're incredibly lucky and can help someone else. But more likely, they don't match in at least some places.
- If they don't match, do some sleuthing to figure out the story:
 - A precinct might have been split in the shapefile or election results, such as Precinct 14 becoming Precinct 14a and Precinct 14b. If that's the case, merge the precincts (but record it!). The variation in these doesn't matter that much for our purposes.
 - A precinct might be named in one file for the polling location. Using google maps and county information, try to identify what precincts match which polling locations, replacing polling locations with precinct names as you find them
- ...and then correct unmatched precincts. Keep your original work, but you might do one or more of the following:
 - Add a column to your precinct shapefile with new precinct names. This would include the originals and renamed ones.
 - Merge precincts.
 - Take note of any changes you make!

Part II. Joining in QGIS or Python (1:45pm-2:30pm)

1. Is there a shared key -- a common column in the Precinct shapefile and Election tabular data -- on which you can join results? If not, create one.
 - a. Examples of a shared key are precinct name, precinct number, and county FIPS with precinct name/number.
2. Join your election results to the precinct shapefile in QGIS/Python, based on personal preference.
3. Check your work: look up county/state totals of election results on the county/state's election website.
 - a. Do they match your results? If not, figure out why. There might be, for instance, absentee votes or county-wide precincts. Once you've figured this out, use MAUP or other tools to address it. List and describe any discrepancies in this worksheet.
 - b. ...and if it's too complicated or time-consuming to figure out (this might be the case, and it's fine!), make a note of it in this worksheet. Explain where the discrepancies are and what you think could address them.

Part III. Adding demographic data using the Census and MAUP (2:30pm-3:45pm)

1. Using the [National Historic GIS](#), download Census data for the following:
 - a. Population by race/ethnicity for your region by Census blocks: 2010 Census, P5 (Hispanic or Latino Origin by Race)
 - b. Voting age population by race/ethnicity for your region by Census blocks: 2010 Census, P11 (Hispanic or Latino, and Not Hispanic or Latino by Race for the Population 18 Years and Over)
 - c. Census blocks (shapefile, under GIS files)
 - d. When you get to Review and Submit, adjust Geographic Extents as appropriate for your state, county, region, etc. You might have to download more data than you need (e.g., data for a whole county when you only need part of it): this is ok!
 - e. Download and unzip the data when it's ready (shouldn't take more than a few minutes to be processed by NHGIS)
2. Because Census datasets on race/ethnicity have so many categories, we created a smaller set of summary categories. Create a single spreadsheet from the downloaded tables that has the following categories and data for your area, by renaming and summing existing Census categories as necessary:
 - **TOTPOP**: Total population
 - **NH_WHITE**: White, non-Hispanic alone population
 - **NH_BLACK**: Black, non-Hispanic alone population
 - **NH_AMIN**: American Indian and Alaska Native, non-Hispanic alone population
 - **NH_ASIAN**: Asian, non-Hispanic alone population
 - **NH_NHPI**: Native Hawaiian and Pacific Islander, non-Hispanic alone population
 - **NH_OTHER**: Other race, non-Hispanic alone population
 - **NH_2MORE**: Non-Hispanic population of two or more races
 - **HISP**: Total Hispanic/Latino population
 - **H_WHITE**: White, Hispanic alone population

- **H_BLACK:** Black, Hispanic alone population
 - **H_AMIN:** American Indian and Alaska Native, Hispanic alone population
 - **H_ASIAN:** Asian, Hispanic alone population
 - **H_NHPI:** Native Hawaiian and Pacific Islander, Hispanic alone population
 - **H_OTHER:** Other race, Hispanic alone population
 - **H_2MORE:** Hispanic population of two or more races
 - **VAP:** Total voting age population
 - **HVAP:** Hispanic voting age population
 - **WVAP:** White, non-Hispanic voting age population
 - **BVAP:** Black, non-Hispanic voting age population
 - **AMINVAP:** American Indian and Alaska Native, non-Hispanic voting age population
 - **ASIANVAP:** Asian, non-Hispanic voting age population
 - **NHPIVAP:** Native Hawaiian and Pacific Islander, non-Hispanic voting age population
 - **OTHERVAP:** Other race, non-Hispanic voting age population
 - **2MOREVAP:** Voting age population of two or more races
3. Use Python/QGIS to join the new sheet to the shapefile, drawing on what you already know about joining.
 4. Open a Python environment (e.g. Colab) and use MAUP to aggregate the block data up to the precinct level. Refer to the [MAUP exercise you did earlier](#) or the [MAUP documentation](#) for details. Save your shapefile when you're done. Open up your new shapefile in QGIS and make sure all the right fields are there.

Part II. Review and summarize your documentation and prepare your presentation.
(5:00pm-6:00pm, eve of Juneteenth)

Communication: At your discretion for Questions 1-3 -- feel free to use Slack or Zoom to communicate, but otherwise, work independently. For questions 4-6, Zoom with your cohort.

1. Take a few minutes to review what you did from:
 - a. The "Notes" column of your cohort's spreadsheet.
 - b. The notes you took yesterday in this Worksheet.
 - c. The comments in your Python scripts.
2. Summarize and organize the information from the previous question into the following sections (in this document! Feel free to reformat the sections below):
 - a. Data Collection
 - b. Digitizing
 - c. Joining Election results to Precinct shapefiles
 - d. Downloading and processing Census data
 - e. Joining Census data to precinct shapefiles using MAUP
3. Think back: were there any roadblocks you ran into that aren't already mentioned in Question 2? Take note of roadblocks you ran into and how, if at all, you were able to address them.

4. Review your summaries with your cohort, asking and addressing questions as they emerge.
5. Presentations will be tomorrow, which is also Juneteenth. Take a few minutes to talk about Juneteenth and then reflect on or look at some of your group's data. How might this data to be used to support Black voting rights/Black liberation? Be specific.
6. Now, as a cohort, turn your answers to question 2, 3, and 5 into a slide show and presentation. There should be at least one slide for each part of question 2 and at least one slide for question 5. Focus on accomplishments, challenges, quirks, and lessons learned: you won't be teaching each other how to repeat your process in the slide show, but rather sharing what you did. Presentations should be about 20 minutes, after which there will be time for questions.

Day 3: Presentations

Zoom logistics: Each person should **participate**, presenting one section of their cohort's slide show. Instructors will facilitate a Q&A after each presentation.

Each cohort presents, with Q&A after each presentation.

Part IV. Next steps

We'll discuss what'll happen with the work participants have done, and introduce opportunities to stay involved.