User Guide and Walkthrough to Shiny App for Ecological Inference of Racially Polarized Voting

This user guide will walk you through using the Ecological Inference App to estimate the level of racial vote polarization in a particular election. Before you start, you will need to choose an election between two candidates to analyze. In order to demonstrate racially polarized voting, you will need to examine a number of elections in this manner. Included below is a step-by-step walkthrough of using the app with some example data.

Finding and Formatting Inputs

Example csv available to download.

1. First, create a new spreadsheet. Each row in the spreadsheet will correspond to an individual precinct, and each column will specify pieces of information to input about the precinct. The columns of your spreadsheet should be labeled with the following headers: precinct_ID, precinct_vote_total, percent_[candidate1], percent_[candidate2], percent_[minority]_voters.
2. Next, you will need to find the data to populate the sheet. The first column, precinct_ID, should list a unique identifier for each precinct. You should be able to find a list, depending on the race, through the office of your secretary of state or county clerk. Some offices make this information available online, while others may require phone calls or written requests. You will also need to look up or request a shapefile of precincts (a digitized map showing their boundaries) to complete step 4.
3. The next three columns—precinct_vote_total, percent_[candidate1], and percent_[candidate2]—require election results by precinct. precinct_vote_total should contain the total number of ballots cast at each individual precinct for the election you are examining. (Note that turnout for the precinct as a whole may be higher than turnout for a particular race. For instance, some voters who cast ballots for the state legislature may not specify any choice for the school board election.) percent_[candidate1] and percent_[candidate2] will either be specified in the data you are given, or can be calculated by dividing each candidate's votes at a precinct by the total votes cast for that race at that precinct. Note that whenever a column reports percents, use decimals ranging from zero to one to report percentages.
4. The final column, percent_[minority]_voters, is an estimate of the demographics of each precinct. Although the Census collects detailed information about race by location, that information does not generally map cleanly onto voting precincts. Census data is available in shapefiles at https://www.census.gov/geo/maps-data/data/tiger-data.html. You will need to use a mapping application such as QGIS to load this data and overlay the precinct shapefiles.
    a. You will need to use the mapping application's tools to calculate racial makeup of each precinct. Some census tracts will have area in more than one precinct. In those cases, you can use the mapping application to calculate the area of the

precinct, and allocate its population proportionally to the precincts with which it overlaps.

   b. Precincts and census tracts have different total populations. Because your final output should be the percent [minority] population of each precinct, you will need to calculate the precinct's minority population and divide by the precinct's total population.
   c. Note that census data includes all residents of a particular area, even those who are too young to vote, noncitizens, or unregistered. Experts often use additional regressions to adjust their estimates of minority voters at each precinct.

5. Save or export your spreadsheet as a .csv (Comma Separated Value) file so it can be uploaded to the Shiny App.

Notes on Data Entry and Formatting

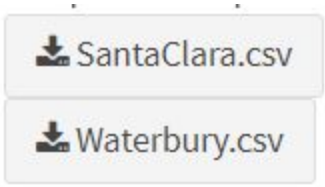
1. The process of manually entering data points can be tedious, and if you are not careful, lead to mistakes. Check over your input multiple times to make sure the correct numbers align with their respective columns and precincts.
2. You can run multiple analyses from the same CSV by creating additional columns. You might want to do this if you are looking for polarization of more than one group against the rest of voters. You will have to run these calculations one group at a time (for instance, by testing first the polarization of Asian versus non-Asian voters, and then the polarization of Hispanic versus non-Hispanic voters.

Using the App

1. On the Shiny App, upload the CSV file by clicking "Browse…" and finding the file. The drop-down menus should now load with your column names as options to select.
2. Under "Number of candidates," choose the number of candidates you will be considering.
3. Under "Candidate 1 data," select "percent_[candidate1]."
4. Under "Name of candidate 1," write that candidate's name.
5. Repeat steps 3-4 for each candidate.
6. Under "Demographic variable," select "percent_[minority]_voters."
7. Under "Name of demographic group," write the name of the race you are considering.
8. Under "Total votes cast," select "precinct_vote_total."
9. For the "Homogeneous precincts threshold," adjust the slider to the percentage of precincts you want to designate as extreme for purposes of the calculation. This setting exists because the most homogenous precincts are most useful for inferring racial vote polarization. You can begin by using 5% as a default value.
10. Under "Source for elections data," write out source used to obtain the election data
11. Under "Source for demographic data," write out source used to obtain the demographic data
12. Click "Run." You can consult the expert report template for further information on the analysis and guidance on its presentation.

Sample runs with Santa Clara 2014 data

1. Download spreadsheet of sample data by clicking one of the buttons provided:



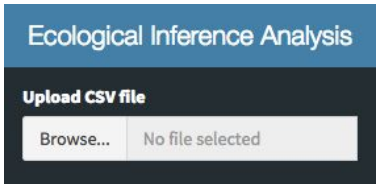
   For this example, we will be using the Santa Clara sample data.

2. Open the spreadsheet to look at the data. You should see twenty-one columns, of which the first five are displayed here:

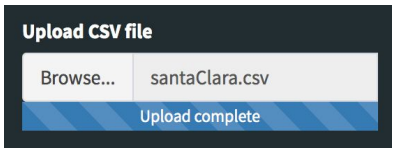
	A	B	C	D	E
1	precinct	total_turnout	total2	pct_for_hardy2	pct_for_kolstad2
2	4201	712	658	0.2462006	0.4316109
3	4208	400	380	0.3289474	0.4315789
4	4211	609	549	0.2513661	0.3843352
5	4215	623	573	0.2582897	0.3717277
6	4217	399	359	0.2813371	0.4066852

   The first column labels the precincts. In this data from Santa Clara, separate contests for multiple seats occurred simultaneously. Column names ending in a number contain information about contests for specific seats. Here, those numbers are 2 and 5. Note that precinct turnout for the races (total2 and total5) differ slightly from each other as well as from the precinct's total turnout (total_turnout). This is because different numbers of people marked their ballots for each contest at these precincts. total_turnout should be the highest number, because it includes every person who voted in at least one race at that precinct. In contrast, demographic estimates remain the same across contests because those numbers come from census data rather than election day results. Note that all percentages are given as decimals between zero and one.

3. Upload the spreadsheet to the Shiny app using the "Browse" button in the upper left hand corner.

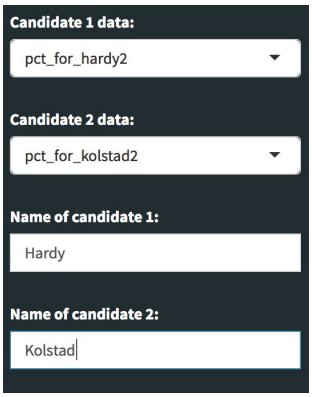


4. Once your spreadsheet is uploaded, the app should create a set of drop-down menus below your uploaded file. The options in these menus come from the column names of your original spreadsheet. You will need to select the right column name for each field. When the csv is done uploading you should see:

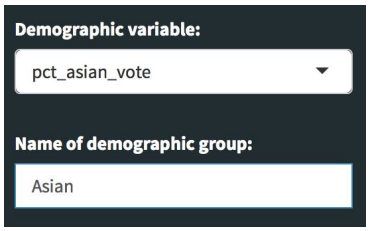


a. In the first section, there is a drop down menu to choose how many candidates you would like to look at. For this example, we will choose 2.

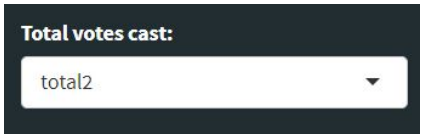
b. In the next section, there is a drop down menu for each candidate. For “Candidate 1 data,” select the title of the column containing percentage of votes for candidate 1, and in “Candidate 2 data”, select the column with percentages for candidate 2. Type out under “Name” of each candidate the way you want their name to appear in figures and tables.



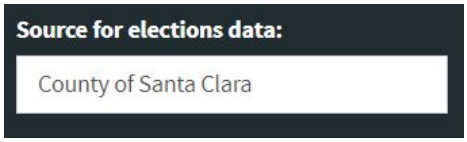
c. In the same section, under the pull down for “Demographic variable” select the column title corresponding to the percent makeup per precinct of the demographic group to be analyzed. Type out the name as you want it to appear in figures in the “Name of demographic group” bar.



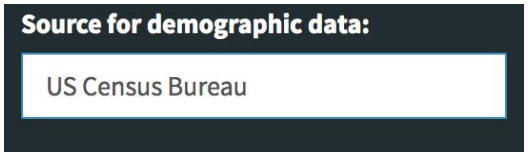
d. The last pull down bar is to select the total voter turnout for the election you are analyzing. In this case we select “total2” because this is the total voter turnout per precinct for the seat the candidates we are analyzing ran for.



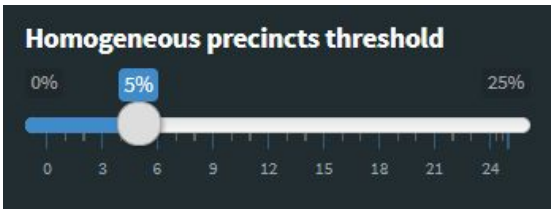
e. In the next entry, “Source for election data,” type out the source for election data (the percentages for each candidate) so that resulting figures properly document where the data came from. This data is from https://www.sccgov.org/sites/rov/resources/pages/pasteresults.aspx



f. Do the same for the demographic data. The data for this example is from the Census.

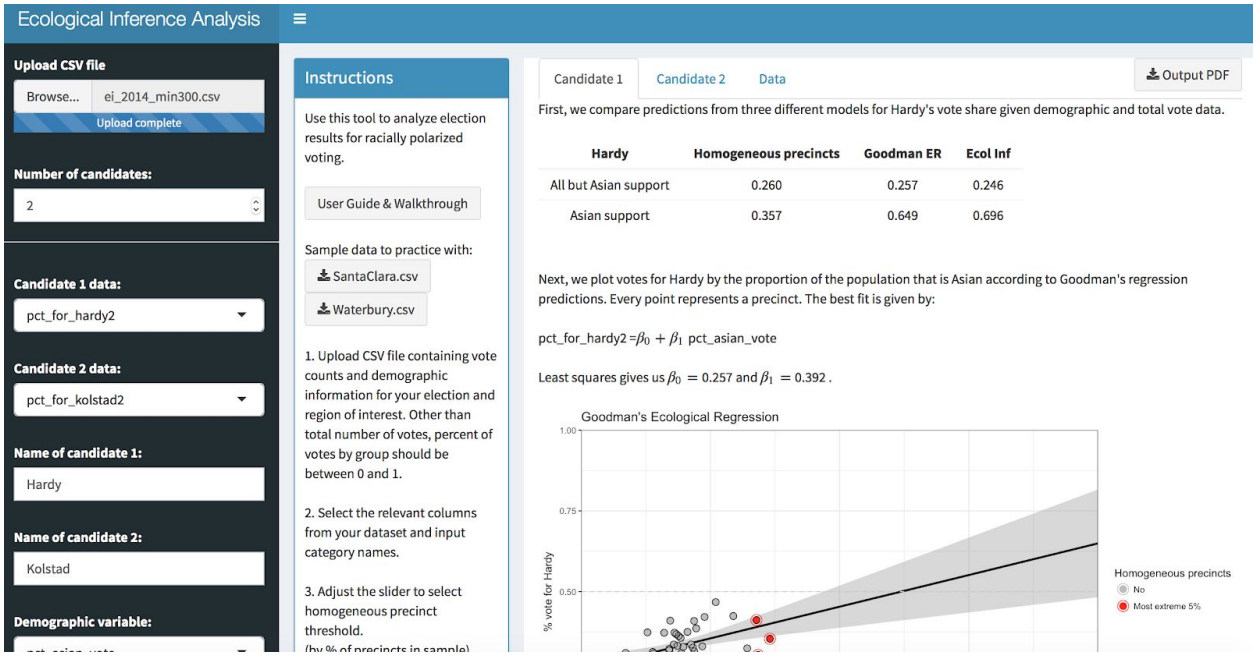


g. Leave the homogeneous precincts threshold slider at its default value of 5%. In step 8, you will experiment with the effects of changing this value.



5. Hit Run, and the app will conduct an EI analysis on your data as entered. This may take some time depending on the size of your data. While the program is running you should see some blue loading bars on the right section of the screen:

6. Results from your analysis will display when analysis is done. Your screen should look like this:



7. The results are organized by tab in the upper right hand corner. The “Candidate 1 Figures” tab shows results for the first candidate and under the “Candidate 2 Figures” is the same analysis but for the second candidate. The data tab on the right shows what the data input looks like. This should match the csv that you uploaded and looks like this:

Candidate 1 | Candidate 2 | Data | Output PDF

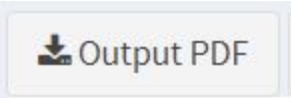
precinct	total_turnout	total2	pct_for_hardy2	pct_for_kolstad2	pct_for_nadeem2	total5	pct_for_caserta5	pct_not_
4201	712	658	0.25	0.43	0.32	642	0.45	
4208	400	380	0.33	0.43	0.24	379	0.46	
4211	609	549	0.25	0.38	0.36	542	0.49	
4215	623	573	0.26	0.37	0.37	558	0.43	
4217	399	359	0.28	0.41	0.31	355	0.43	
4221	509	472	0.26	0.40	0.34	476	0.47	
4222	518	453	0.29	0.38	0.32	448	0.44	
4226	643	571	0.27	0.39	0.34	562	0.42	
4229	695	621	0.27	0.41	0.32	615	0.36	
4238	489	428	0.34	0.42	0.25	420	0.42	
4241	633	578	0.41	0.37	0.22	568	0.39	
4242	593	531	0.37	0.35	0.28	517	0.36	
4247	412	369	0.41	0.38	0.21	363	0.36	
4249	623	581	0.33	0.39	0.28	576	0.39	
4251	515	452	0.37	0.34	0.28	450	0.41	
4256	492	413	0.32	0.34	0.33	396	0.30	
4257	333	304	0.42	0.38	0.20	304	0.39	
4258	342	308	0.47	0.32	0.21	304	0.35	

8. Experiment with adjusting the homogeneous precincts threshold slider. This setting affects the number of points used in the “Homogenous precincts” result. The program will select the percentage of points you specify from each end of the demographic spread (e.g. the 5% least and most Asian precincts). If you run the program with the

slider set at higher percentages, you will see more points highlighted red and used to fit the homogenous precincts line. The program needs at least two points to make the calculation.

9. Also try experimenting with using different columns under the “Demographic Variable” pull down menu. For example, you could select “pct_e_asian” or “pct_ind_asian” with the labels “East Asian” and “Indian” to run the same analysis on more specific sub groups.

10. To download a LaTeX pdf of the results generated here, make sure you have run the program with your desired settings and click the output pdf button above the results:



You can’t copy and paste figures directly from the pdf, but you can easily screenshot portions to put into other documents.