NOTES ON OHIO

KYLE DROPP

1. Files

- "oh_geo.csv" - This is the Census voting district file. Dimensions: 11029 by 17.
- "oh_steve_2008.csv" - This is the vote total file sent by SDA. It is modified and only contains 2008 vote totals. Dimensions: 11107 by 41.
- "oh_geo_merged_final.csv" - This is the Census voting file with vote totals merged. Dimensions: 11029 by 57.
- "oh_vote_merged_final.csv" - This is the vote file with merged Census voting districts. Dimensions: 11107 by 57.

2. Overview

- Matching summary
  - I matched 11,009 out of 11,029 precincts (99.8 percent) in the Census voting district file with entries in the 2008 vote total file.
  - I matched 11,006 out of 11,107 precincts (99.1 percent) in the 2008 vote file with precincts in the Census voting district file.
- Important merging details
  - The key variable in both files is called "precinct_code". It is a concatenation of the three digit county code and three letter precinct code for each county. It is in the following format: "county number + space + three letter code".
  - The attached files include 2008 vote totals merged with 2010 Census voting districts. I can finish the time series component at a later date.
  - After deleting all entries that did not contain 2008 vote totals, there remained approximately 1,000 cases where multiple rows represented a single voting precinct. These cases were of two types: first, one row contained 2008 Presidential and Attorney General vote totals and the second row contained only U.S. House totals. These two rows were summed and merged into a single row. Second, in many instances, the two rows representing a single precinct were identical in every respect. This is likely due to a problem with data entry. Alternatively, this may be caused by

a first row containing 2008 and 2006 vote totals and a second row containing 2008 and 2004 vote totals. For this second set of problem cases, I deleted the second row, leaving a single 2008 vote total entry.