REDISTRICTING DATA HUB PRISON GERRYMANDERING PART 2: BENEATH THE HOOD

Dr. Brianna Remster
Dr. Rory Kramer
Villanova University
March 15, 2022
PART 2
AGENDA

Our nationwide project

How we got the data

How we clean the data

How we use the data
CURRENT PROJECT: NATIONWIDE ASSESSMENT WITH NEW MAPS

• 37 states continue to use the Census’s unadjusted data

• 13 states addressing this redistricting cycle

• Research team: Denise Wilson, Chelsea Canal, Gabi Oliveira
IMPACT IN TEXAS (AS OF JAN 2023)*

- ~15% of House districts would be unconstitutional
  - 12 districts would be legally too small
    - Range: Lose between 400-~13,000 people
  - 11 districts would be legally too large
    - Range: Gain between 350-~1300 people

*Reallocation method: state facilities using county of origin and race; Legal size requirement: +/-5%
Racially Unequal Impact in Texas

District Size Change by Race and Ethnicity with Incarcerated Persons Reallocated, TX House

- Weighted average district change post reallocation
- white
- black
- hispanic
- other
GETTING THE DATA: THE EASIEST AND HARDEST PART

• Census data: super easy!

• state redistricting websites or Dave’s Redistricting (usually easy)

• Direct contact with Departments of Correction followed by FOIA/sunshine law requests (ranges from easy to nearly impossible)
  • Best practices: individual addresses with race
  • Second best: aggregate data publicly available
  • Worst practices: asking for $500, IRB approval (!?!?) and/or months of delay to *maybe* fulfill request

• Feds and local jails not in data set
CLEANING THE DATA: THE PROBLEMS

- State data isn’t the same as Census data
  - Block ID codes inconsistent
  - racial/ethnic data options
    - Problems with Latinx—either double counted (NC) or left out entirely (MI)
    - We use best possible racial/ethnic data given constraints for individual states
- Only state facilities
  - Some private facilities did not provide good data to state partners
  - Federal BOP refuses to give data
  - Local jails often lack data (big problem for local redistricting, smaller problem for state level)
PROCESSING THE DATA

• If given specific addresses:
  • Remove exact number of incarcerated from prison location block
    • Differential privacy mini-rant: Activate
  • Use batch Census API to identify blockid, return incarcerated to that block
    • Typically 50-70% match rate after basic error cleaning)
  • If API cannot identify blockid, use the magic of Google Maps
    • Look up address on google maps, confirm it has a dwelling via satellite imagery
    • Find geographic coordinates
    • Use geographic coordinates in Census API
PROCESSING THE DATA

• Our Process if given aggregate data (county/city):
  • Remove exact number of incarcerated from prison location block
    • Differential privacy mini-rant: still valid
  • Identify each individual block’s share of county/city racial population
  • Return that percentage of county/city incarcerated population to each block
    • Imprecise estimates in counties with large populations
    • Imprecision means we UNDERESTIMATE local impact of prison gerrymandering
ANALYZING THE DATA: STATA CODE
THANKFULLY, YOU DON’T HAVE TO DO THAT WORK!

• Cleaned data now/soon available on redistricting hub website
  • Individual blocks
  • Aggregated results
• Word of caution with block level data: Blockid csv -> excel can be a problem
  • Turn it into a string variable if you do use excel
  • Better: don’t use excel!
  • Best: use R and share your code if you do (because I’m terrible at R)
RESULTING DATA: BLOCK LEVEL

<table>
<thead>
<tr>
<th>Block ID</th>
<th>District HC</th>
<th>District SE</th>
<th>Total</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Other</th>
<th>Adjusted C</th>
<th>Adjusted W</th>
<th>Adjusted B</th>
<th>Adjusted LA</th>
<th>Adjusted LB</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.82E+14</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1.106888</td>
<td>1.011819</td>
<td>0</td>
<td>10.05706</td>
<td>0</td>
</tr>
<tr>
<td>4.84E+14</td>
<td>6</td>
<td>1</td>
<td>100</td>
<td>87</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>100.994</td>
<td>87.75837</td>
<td>8.227081</td>
<td>1.006049</td>
<td>4.002524</td>
</tr>
<tr>
<td>4.85E+14</td>
<td>5</td>
<td>1</td>
<td>126</td>
<td>102</td>
<td>0</td>
<td>10</td>
<td>14</td>
<td>126.7875</td>
<td>102.7433</td>
<td>0</td>
<td>10.03763</td>
<td>14.00653</td>
</tr>
<tr>
<td>4.82E+14</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>6.054996</td>
<td>1.004091</td>
<td>3.045416</td>
<td>2.005488</td>
<td>0</td>
</tr>
<tr>
<td>4.83E+14</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>23.16919</td>
<td>13.05897</td>
<td>6.103448</td>
<td>2.006768</td>
<td>0</td>
</tr>
<tr>
<td>4.85E+14</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.84E+14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.84E+14</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.85E+14</td>
<td>5</td>
<td>1</td>
<td>24</td>
<td>2</td>
<td>3</td>
<td>19</td>
<td>0</td>
<td>24.13816</td>
<td>2.017262</td>
<td>3.057416</td>
<td>19.06801</td>
<td>0</td>
</tr>
<tr>
<td>4.81E+14</td>
<td>1</td>
<td>1</td>
<td>39</td>
<td>32</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>39.20591</td>
<td>32.17805</td>
<td>0</td>
<td>5.026197</td>
<td>2.001664</td>
</tr>
<tr>
<td>4.83E+14</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.85E+14</td>
<td>13</td>
<td>1</td>
<td>38</td>
<td>30</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>38.31584</td>
<td>30.21862</td>
<td>4.09206</td>
<td>1.003763</td>
<td>3.0014</td>
</tr>
<tr>
<td>4.83E+14</td>
<td>15</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>6.16497</td>
<td>2.02359</td>
<td>4.14138</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.82E+14</td>
<td>17</td>
<td>1</td>
<td>45</td>
<td>41</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>45.3787</td>
<td>41.35547</td>
<td>1.022095</td>
<td>0</td>
<td>3.00113</td>
</tr>
<tr>
<td>4.85E+14</td>
<td>19</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>8.140334</td>
<td>1.006363</td>
<td>7.133971</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.8E+14</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
RESULTING DATA: DISTRICT LEVEL

• I’m gonna show you a live excel! Exciting!
LESSONS FOR THE FUTURE: EARLY IS ON TIME, ON TIME IS LATE

• Proactively work with CJ system to create better data well before 2030
• Provide incentive for CJ system to create that better data (California approach)
• Expect resistance
• Expect fear of “messy” data
QUESTIONS?

• Can also email at rory.Kramer@villanova.edu or Brianna.remster@villanova.edu

• Thank you!!!