REDISTRICTING AND DIFFERENTIAL PRIVACY

Michael McDonald
University of Florida
www.electproject.org ● @ElectProject

Redistricting Experience

- **Since late 1980s consulted to redistricting authorities or was an expert witness in court cases in 15 states, most recently:**
  - Successful challenge of Virginia's congressional districts as racial gerrymander
  - Unsuccessful challenge of Maryland's congressional districts as partisan gerrymander
- **Co-PI of award-winning Public Mapping Project to promote public engagement and transparency in redistricting**
  - Produced DistrictBuilder online mapping tool
- **Currently leading team to produce accurate precinct boundaries tiling the entire country**
- **Authored many scholarly redistricting publications**

US Elections Project

TWO LEGAL QUESTIONS

1) **Equal Population**
2) **Voting Rights Compliance**

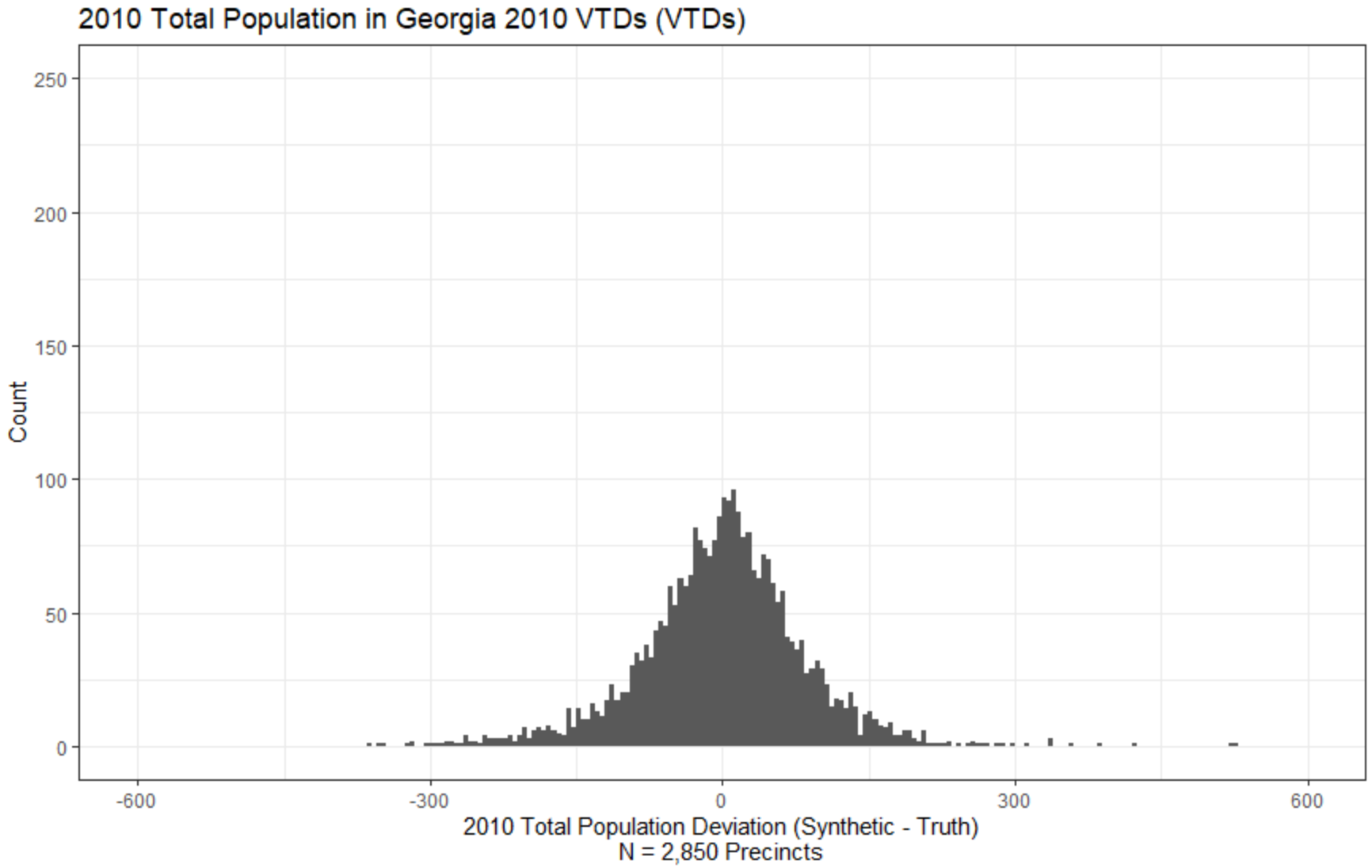
Investigate these questions through a Georgia case study

Population Equality

Equal Population

- Congressional districts must generally have de minimus equal total population deviations of 1 or 9 persons.
  - If a compelling state interest exists, a 1% population deviation is allowed (*Tennant* v *Jefferson County Commission* 2012)
- State legislative districts may have a wider 10% deviation
- Other state and federal laws may apply

US Elections Project


2010 Total Population in Georgia 2010 VTDs (VTDs)
Count
250
200
150
100
50
0
-600
-300
0
300
600
2010 Total Population Deviation (Synthetic - Truth)
N = 2,850 Precincts

Reimagining Equal Population

If States and Courts understand the affect of differential privacy on total population counts, then the strict equal population requirement for congressional districts likely relaxes.

Will synthetic data point estimates satisfy equal population standards?

Are confidence intervals needed?

Recommendation: Once epsilon has been chosen, publish official confidence intervals (or approximations) of population counts

US Elections Project

Voting Rights

The Voting Rights Act § 2

The *Gingles* Test Three Prongs

Thornberg v *Gingles* 478 U.S. 30 (1986)

1. Can a reasonably compact district with 50% or greater minority voting-age population (VAP) be drawn?
2. Is there the presence of racially polarized voting, where sufficient numbers of Whites vote against the minority community's preferred candidate of choice to deny the minority community an opportunity to elect their candidate?
3. The Totality of the Circumstances

US Elections Project

Gingles Test First Prong: Drawing a 50%+ Minority VAP District

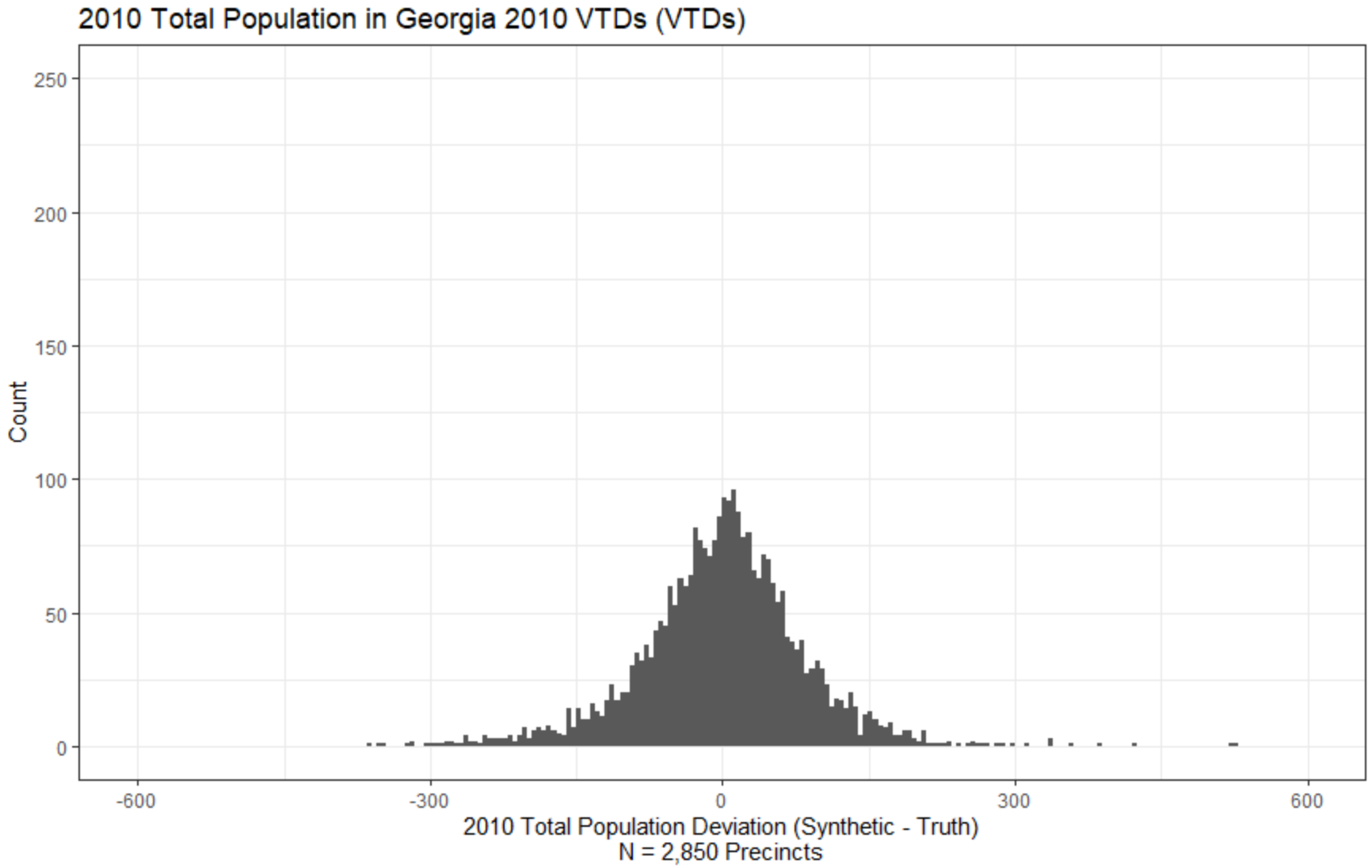
US Elections Project

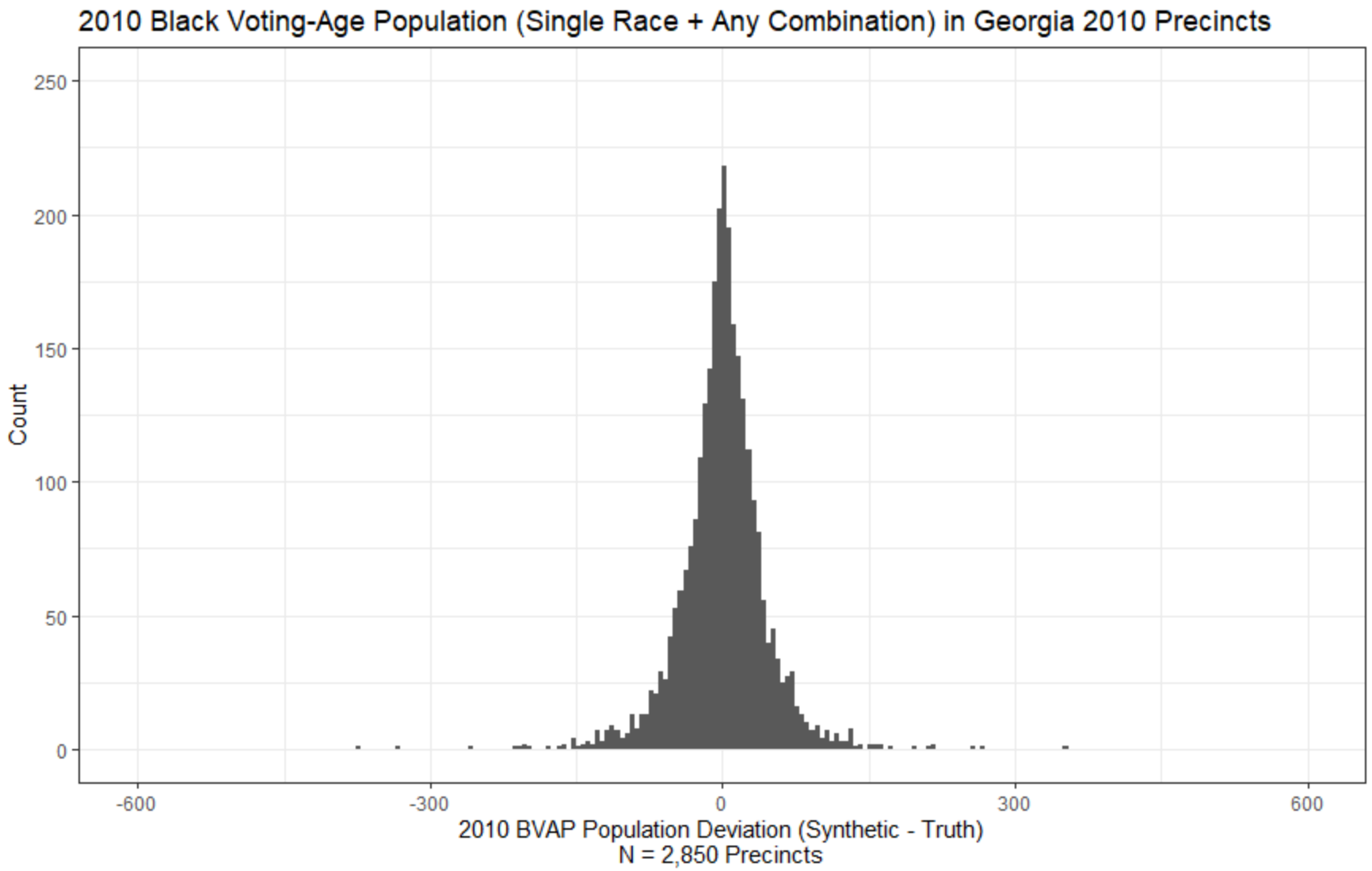
Georgia Statewide Counts

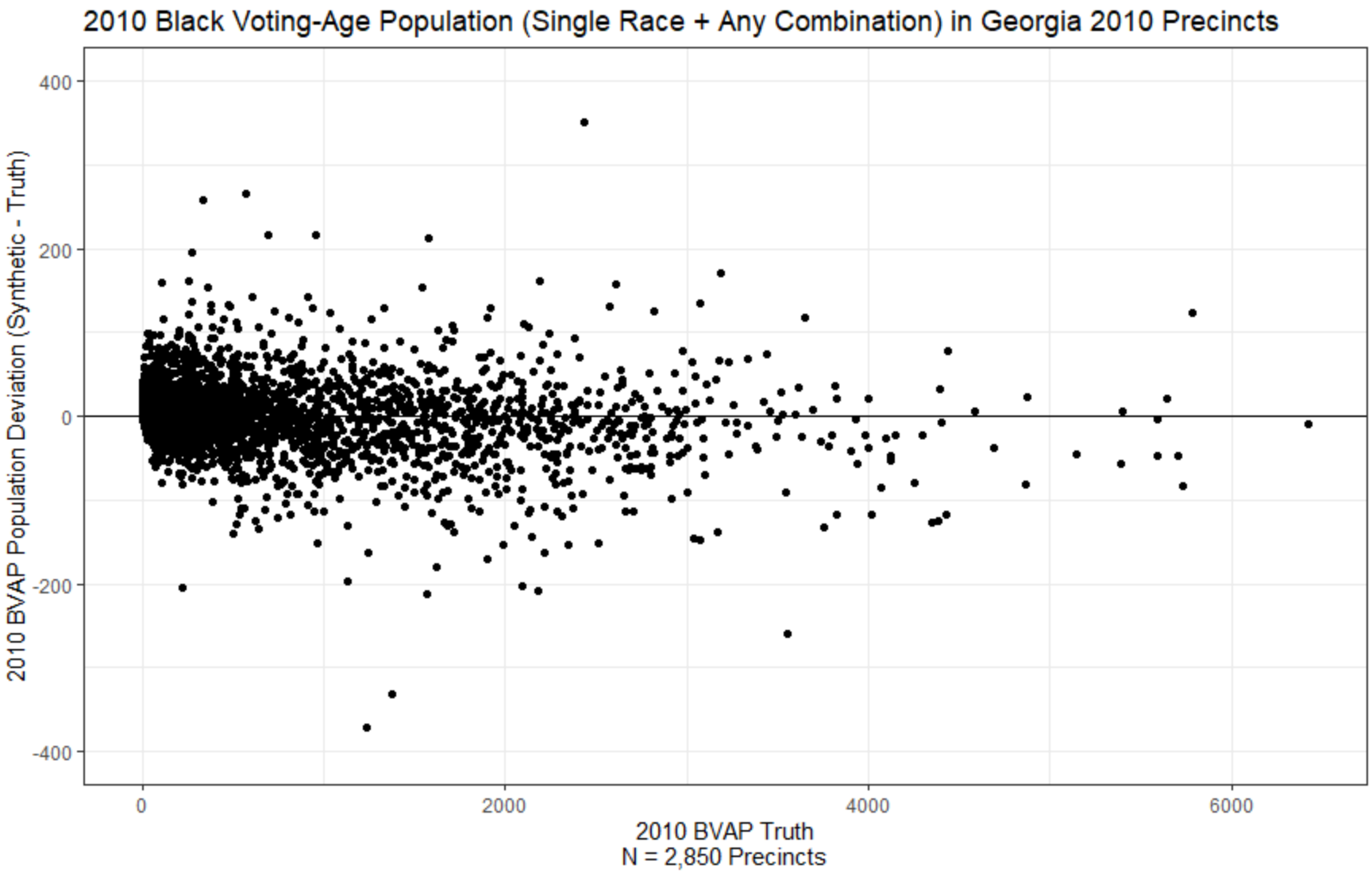
	Synthetic	Truth	Difference
Total Population	9,687,653	9,687,653	0
Non-Hispanic White VAP	4,242,496	4,242,514	-18
Black VAP	2,141,665	2,140,789	+876
Hispanic VAP	538,732	539,002	-270

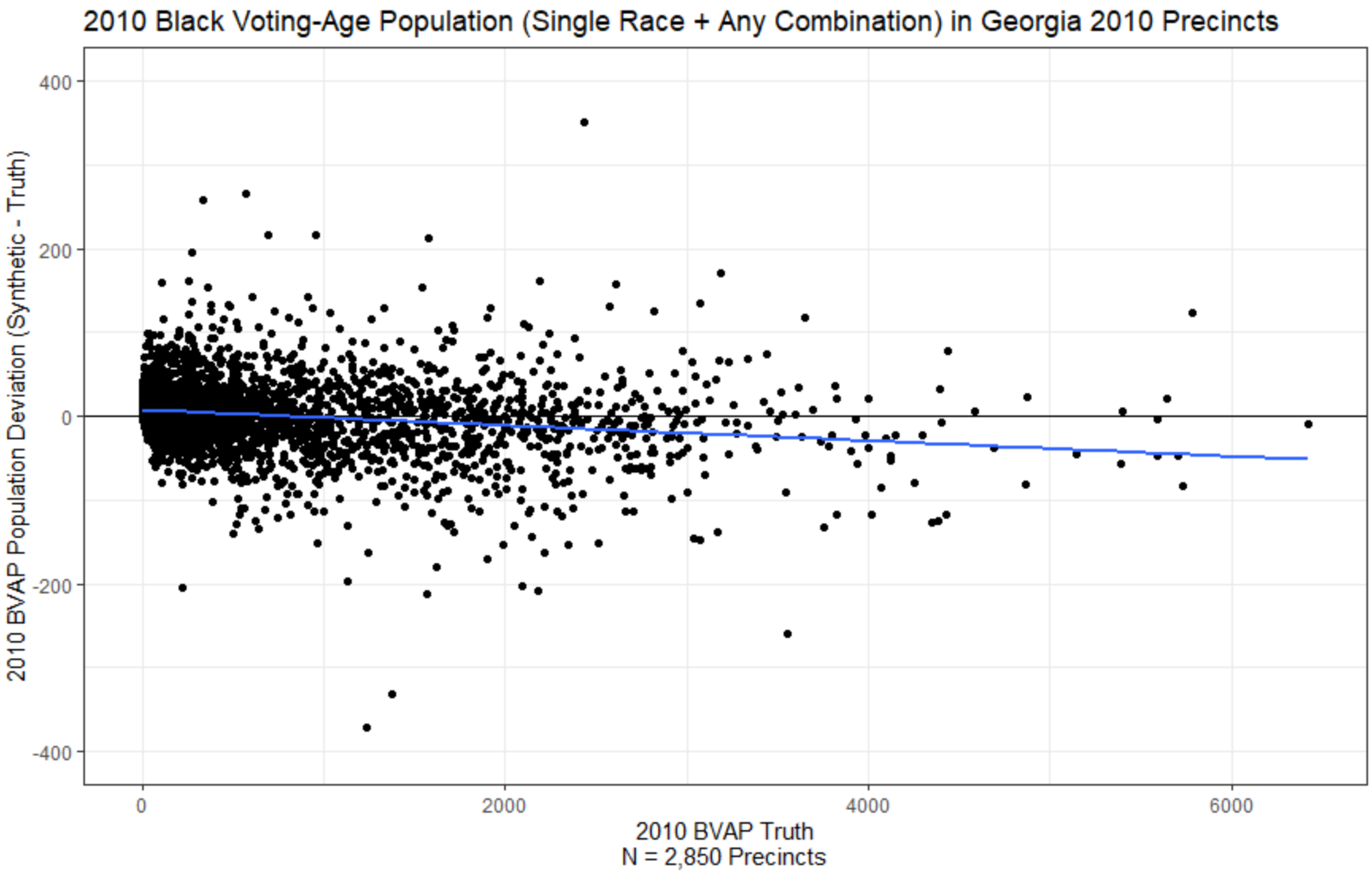
Statewide synthetic total population is constrained to equal the truth, but statewide sub-population totals are not similarly constrained

US Elections Project


2010 Total Population in Georgia 2010 VTDs (VTDs)
Count
250
200
150
100
50
0
-600
-300
0
300
600
2010 Total Population Deviation (Synthetic - Truth)
N = 2,850 Precincts


2010 Black Voting-Age Population (Single Race + Any Combination) in Georgia 2010 Precincts
Count
250
200
150
100
50
0
-600
-300
0
300
600
2010 BVAP Population Deviation (Synthetic - Truth)
N = 2,850 Precincts


2010 Black Voting-Age Population (Single Race + Any Combination) in Georgia 2010 Precincts
2010 BVAP Population Deviation (Synthetic - Truth)
400
200
0
-200
-400
0
2000
4000
6000
2010 BVAP Truth
N = 2,850 Precincts


2010 Black Voting-Age Population (Single Race + Any Combination) in Georgia 2010 Precincts
2010 BVAP Population Deviation (Synthetic - Truth)
400
200
0
-200
-400
0
2000
4000
6000
2010 BVAP Truth
N = 2,850 Precincts

Shifting Blacks from Homogenous Black Communities to Homogeneous White Communities

Synthetic Data Are (Loosely) Bounded

Black VAP, Synthetic minus Truth

Precincts with Black VAP ≥ 1,000: -9,526

Precincts with Black VAP < 1,000: +10,402

(Recall, 876 Black VAP added statewide)

Could affect *Gingles* test first prong in that it may be more difficult to draw a 50%+ Black VAP district.

US Elections Project

Recommendations

Recommendation #1: Spend less of the privacy budget on total population and VAP by race and ethnicity cells

Recommendation #2: Publish official confidence intervals of counts and race and ethnicity proportions for gradated population sizes.

Recommendation #3: Apply negative spatial correlation to differential privacy algorithm to mitigate random chance of positively correlated groupings of population deviations.

US Elections Project

Gingles Test Second Prong: Racially Polarized Voting

Ecological Inference

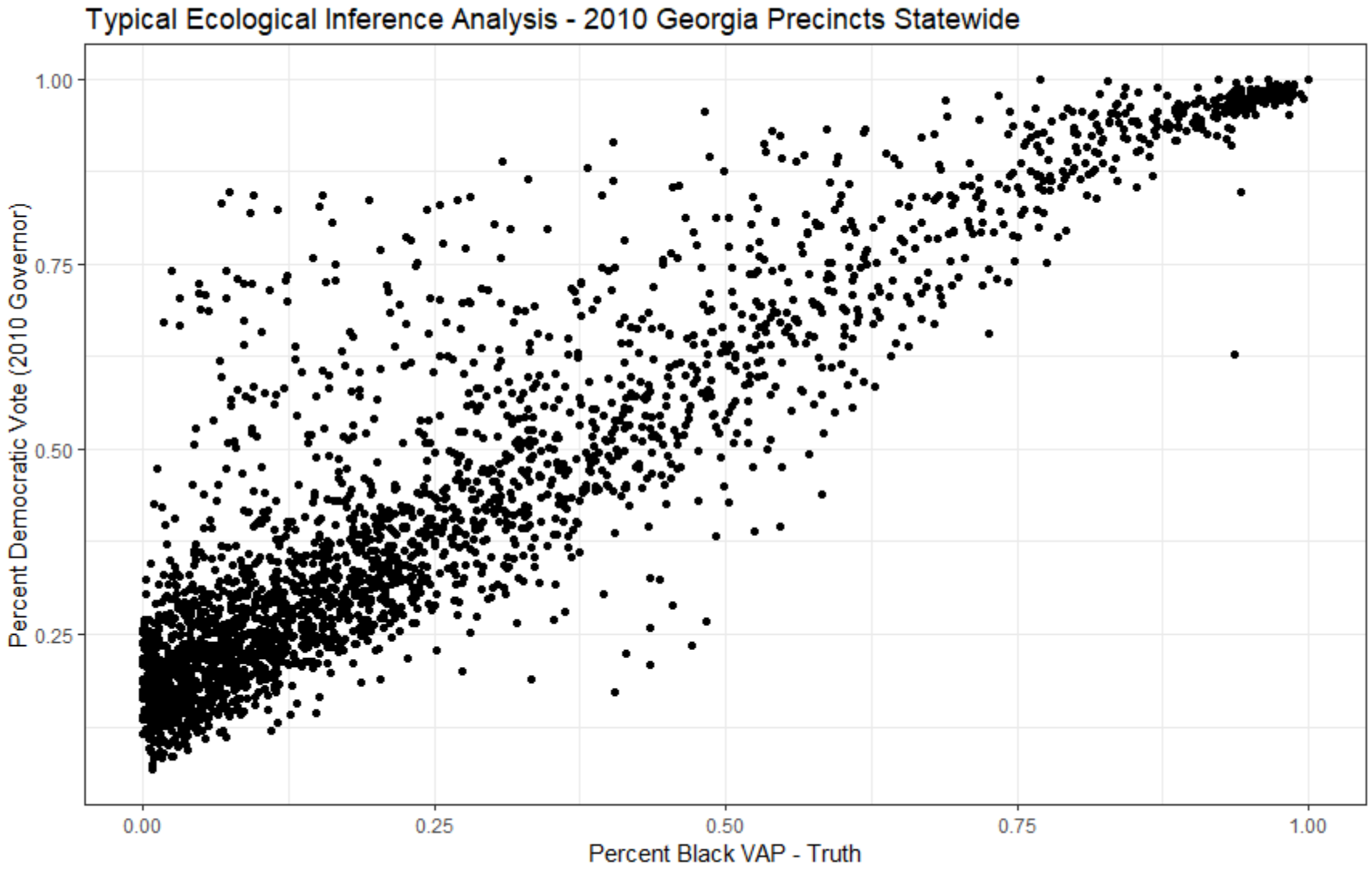
Secret ballot laws protect the confidentiality of individual voters

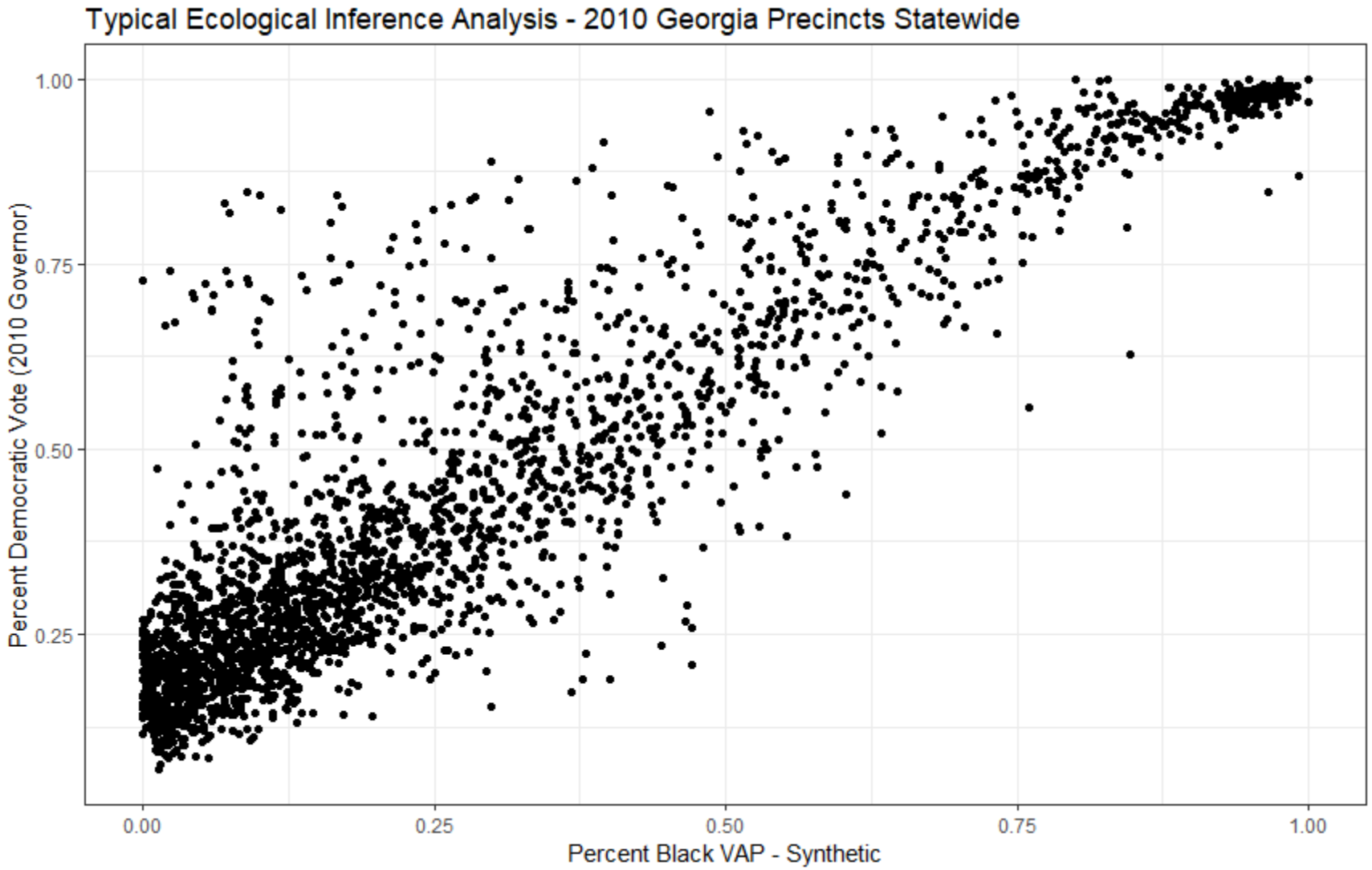
Aggregate election results reported within precincts

When precinct boundaries are known, precinct boundaries can be spatially joined to census geography, so that census population counts can be aggregated within precincts

Estimate individual voting propensities by race from these aggregate data. In the crudest form, a simple correlation

US Elections Project


Typical Ecological Inference Analysis - 2010 Georgia Precincts Statewide
Percent Democratic Vote (2010 Governor)
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
Percent Black VAP - Truth


Typical Ecological Inference Analysis - 2010 Georgia Precincts Statewide
Percent Democratic Vote (2010 Governor)
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
Percent Black VAP - Synthetic

Ecological Inference

Simple Goodman's Regression

$$TwoParty\ Vote_{Minority\ Preferred\ Candidate} = \boldsymbol{\beta_1}\text{BVAP} + \boldsymbol{\beta_2}(1 - \text{BVAP})$$

$\boldsymbol{\beta_1}$ = Black Vote for Minority Preferred Candidate

$\boldsymbol{\beta_2}$ = Non-Black Vote for Minority Preferred Candidate

Two-Stage Goodman's Regression (controls for differential turnout rates)

$$Turnout\ as\ Percentage\ of\ VAP = \boldsymbol{\beta_1}\text{BVAP} + \boldsymbol{\beta_2}(1 - \text{BVAP})$$

$$Candidate's\ Vote\ as\ Percentage\ of\ VAP = \boldsymbol{\beta_3}\text{BVAP} + \boldsymbol{\beta_4}(1 - \text{BVAP})$$

β_3/β_1 = Black Vote for Minority Preferred Candidate

β_4/β_2 = Non-Black Vote for Minority Preferred Candidate

RxC Bayesian method proposed by Gary King and co-authors

US Elections Project

2010 Governor Statewide

2010 Governor Statewide

Goodman's Regression

	Support for Candidate of Choice			Standard Error		
	Truth	*Synthetic*	*Difference (Synthetic - Truth)*	*Truth*	*Synthetic*	*Difference (Synthetic - Truth)*
Black	1.0419	1.0443	0.0023	0.0056	0.0058	0.0002
Non-Black	0.2017	0.1981	-0.0036	0.0030	0.0031	0.0001

Two-Stage Goodman's Regression

	Support for Candidate of Choice		
	Truth	*Synthetic*	*Difference (Synthetic - Truth)*
Black	1.0861	1.0879	0.0018
Non-Black	0.1872	0.1826	-0.0046

RxC EI

	Support for Candidate of Choice			Standard Error		
	Truth	*Synthetic*	*Difference (Synthetic - Truth)*	*Truth*	*Synthetic*	*Difference (Synthetic - Truth)*
Black	0.9510	0.9414	-0.0096	0.0148	0.0156	0.0009
Non-Hispanic White	0.1929	0.1967	0.0038	0.0015	0.0042	0.0026
Other	0.5884	0.5616	-0.0268	0.0789	0.0403	-0.0386

2010 Congressional District 12

2010 CD12

Goodman's Regression

	Support for Candidate of Choice			Standard Error		
	Truth	*Synthetic*	*Difference (Synthetic - Truth)*	*Truth*	*Synthetic*	*Difference (Synthetic - Truth)*
Black	1.0302	1.0503	0.0201	0.0143	0.0149	0.0007
Non-Black	0.2411	0.2250	-0.0161	0.0102	0.0106	0.0004

Two-Stage Goodman's Regression

	Support for Candidate of Choice		
	Truth	*Synthetic*	*Difference (Synthetic - Truth)*
Black	1.0516	1.0640	0.0123
Non-Black	0.2167	0.2043	-0.0123

RxC EI

	Support for Candidate of Choice			Standard Error		
	Truth	*Synthetic*	*Difference (Synthetic - Truth)*	*Truth*	*Synthetic*	*Difference (Synthetic - Truth)*
Black	0.9589	0.9503	-0.0086	0.0205	0.0232	0.0026
Non-Hispanic White	0.2440	0.2245	-0.0194	0.0172	0.0208	0.0036
Other	0.6156	0.6757	0.0601	0.3046	0.2325	-0.0721

Recommendations

Estimates are sensitive to the application of the differential privacy algorithm in the two cases examined.

More complex model specifications tend to be more sensitive to addition of measurement error.

Recommendation #1: Run multiple model specifications to diagnose potential problems

Recommendation #2: If expected simulation variance is known (or an estimate provided), apply multiple imputation methods to check sensitivity of results

US Elections Project

REDISTRICTING AND DIFFERENTIAL PRIVACY

Michael McDonald
University of Florida
www.electproject.org ● @ElectProject