

# Data and the State

PUBPOL 2130 / INFO 3130



## Course Intro

Lecture 1, Tuesday Jan 21

# Welcome to “Data and the State”

- New class!
- New prof!
- New everything! Be patient please 😊
- Goals of this class: comfort and literacy with Python **tools** that allow parsing, curation, and visualization of geospatial data; readings to build the **theory**; understanding of **applications** in government and NGOs.

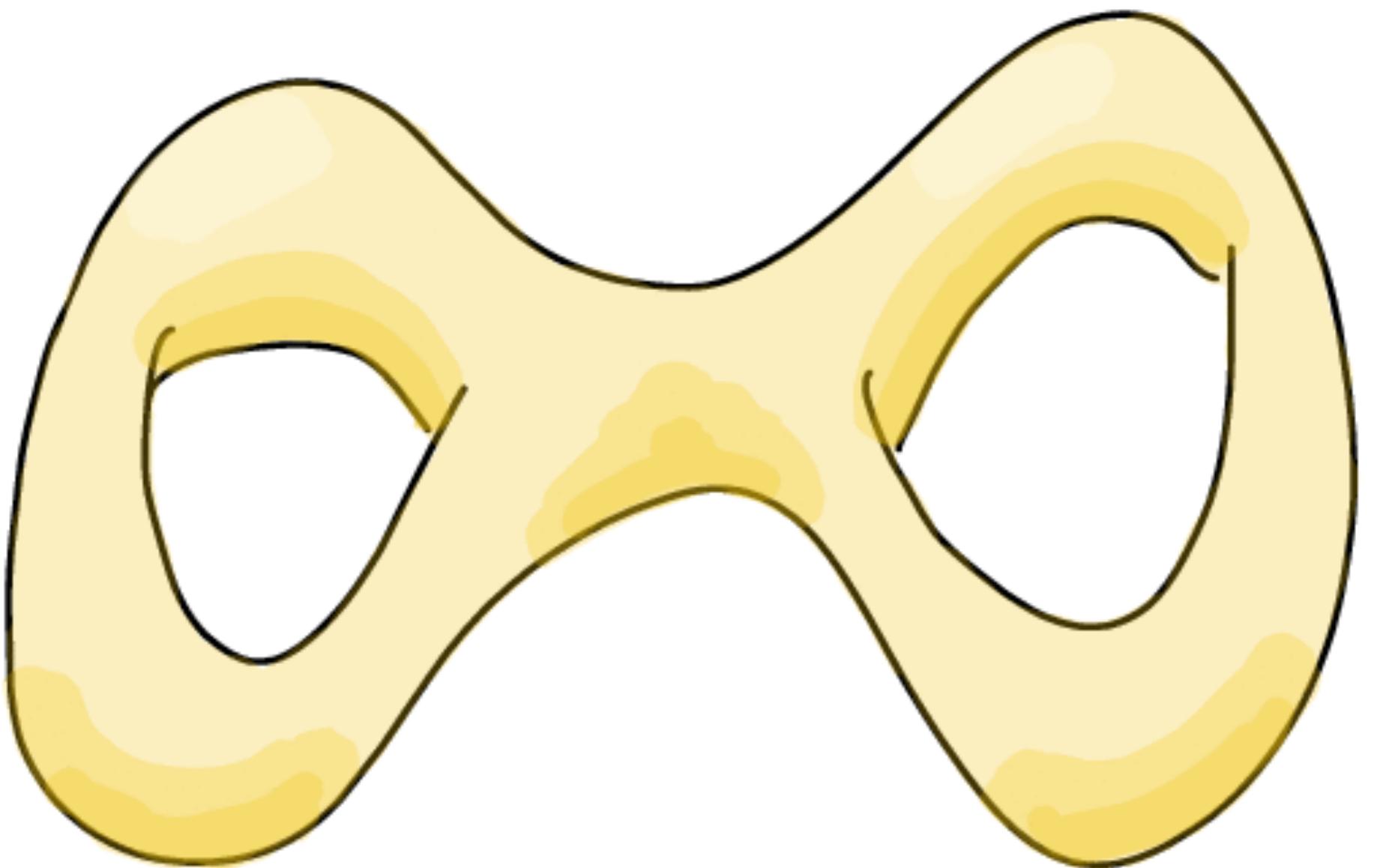


# Moon Duchin

Professor of Mathematics and Public Policy



- **Math** research: geometry, topology, dynamical systems, group theory
- **Elections** research: gerrymandering, ranked choice voting, fairness
- Usually teach a mix of math classes and **interdisciplinary classes** on topics like History of Math, Models in Social Context, Classification.
- **Public service**: consulting for democracy reform, expert witness in state and federal court.



*Tedusa*

CC-BY-NC-SA

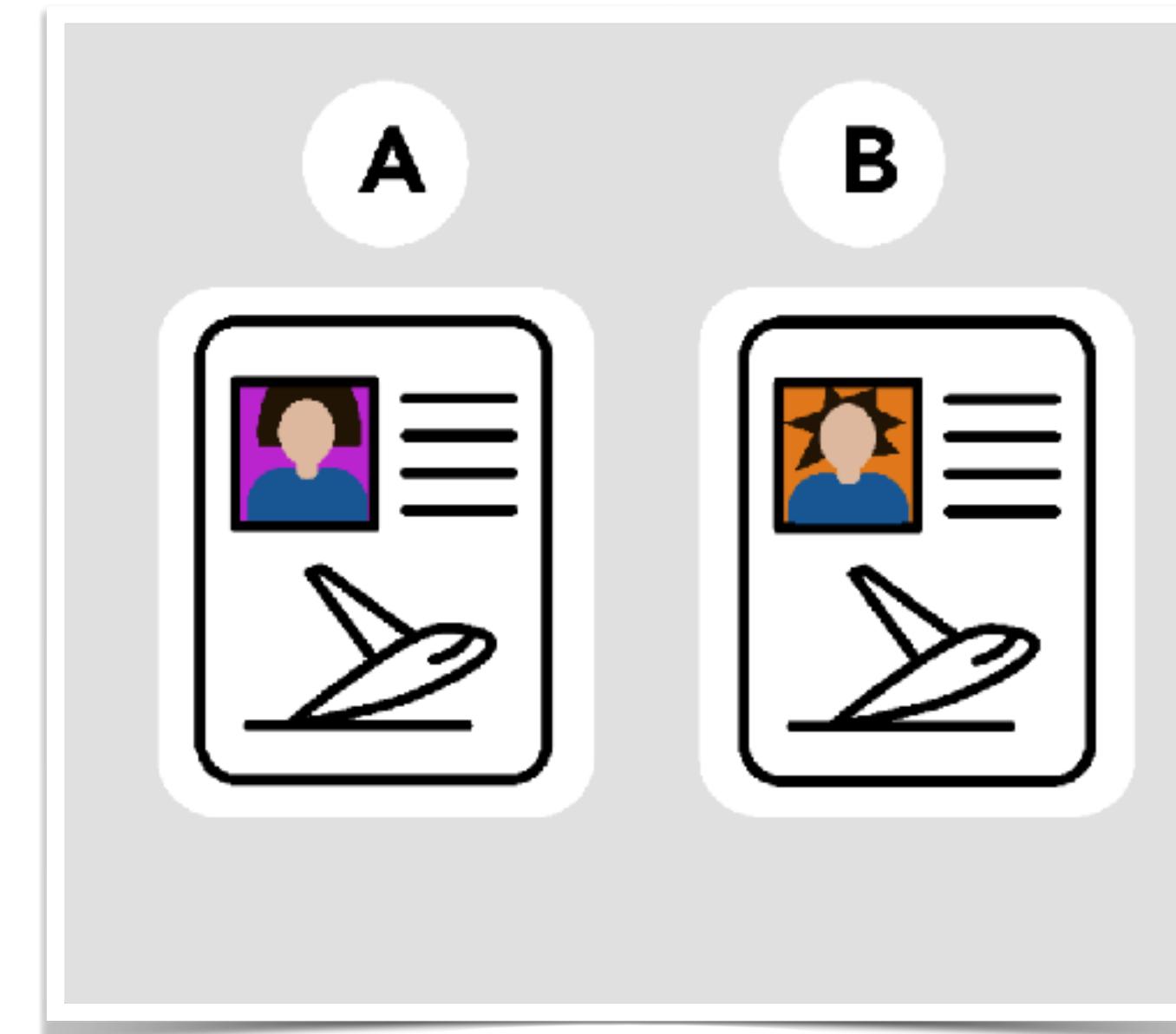
# Jennah Gosciak

PhD student in Information Science

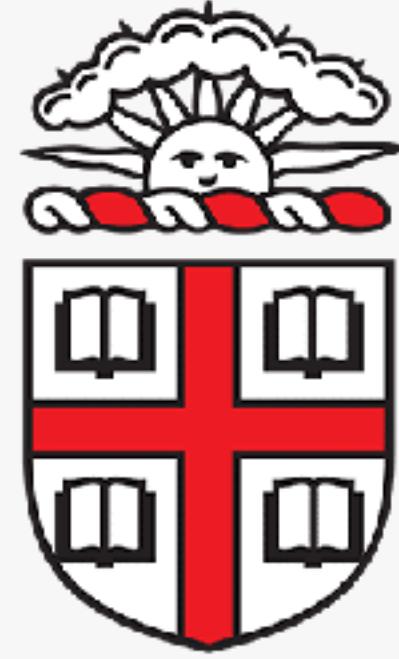


- Developing better tools for **causal inference** and data adaptive experimental design
- Applying causal inference and machine learning (ML) methods to policy areas like **housing and health**
- Studying the **fairness and equity** implications of ML in social science and policy settings

*Data-Adaptive  
Experimentation  
to Find Contexts  
with the Most and  
Least  
Discrimination*

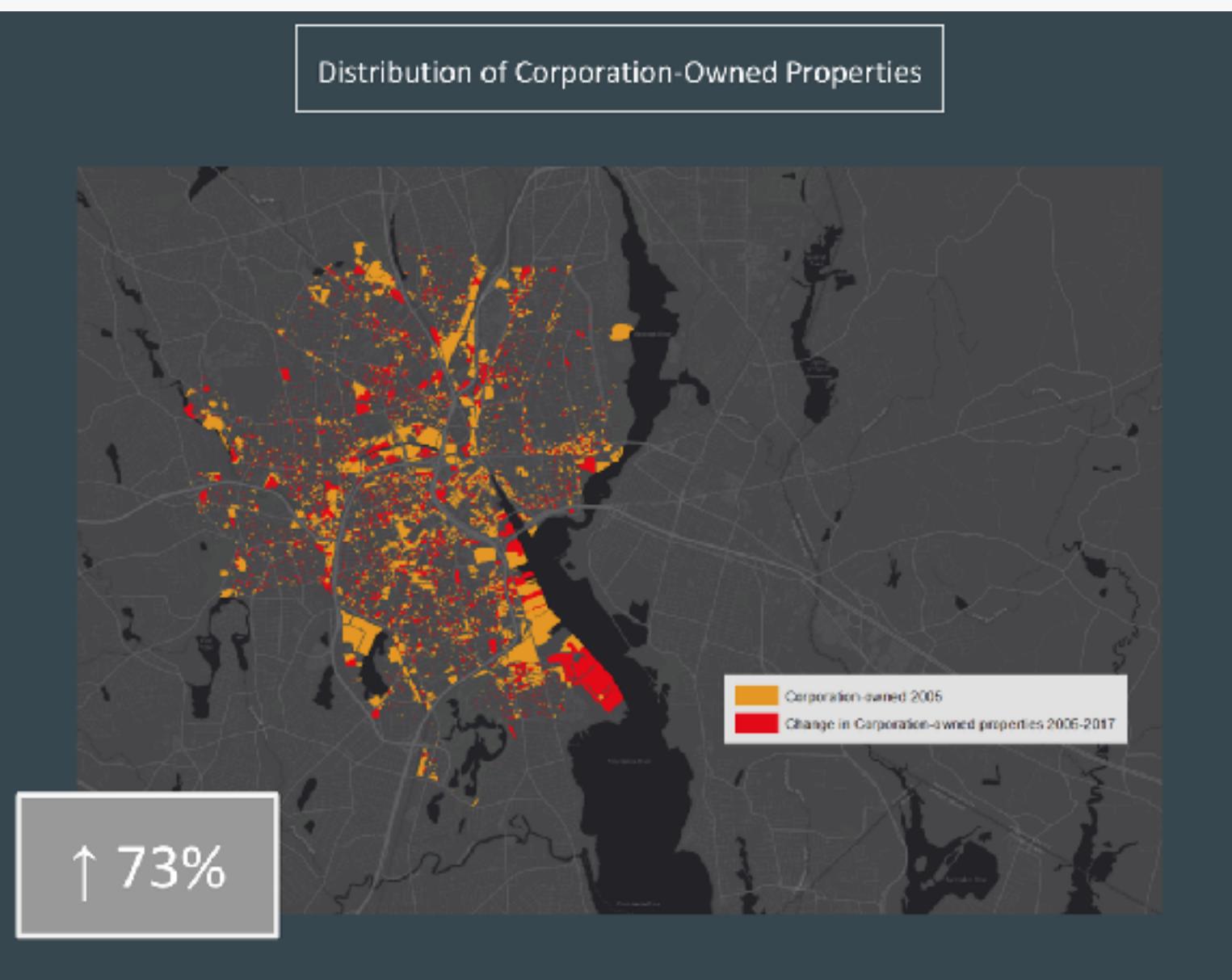


*Lessons from  
Planning Theory:  
A Critical  
Pragmatism  
Approach for  
Algorithmic  
Fairness*

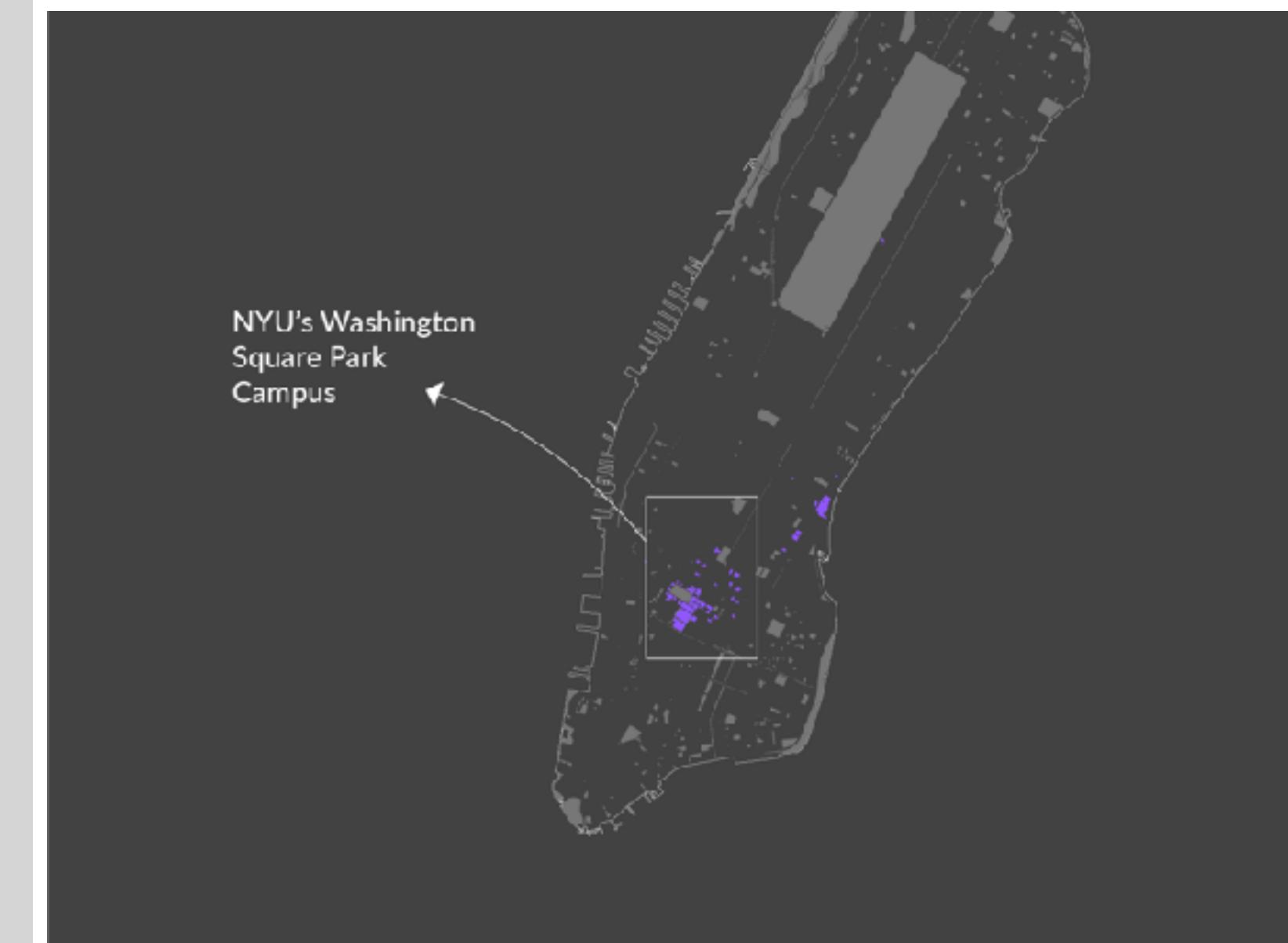
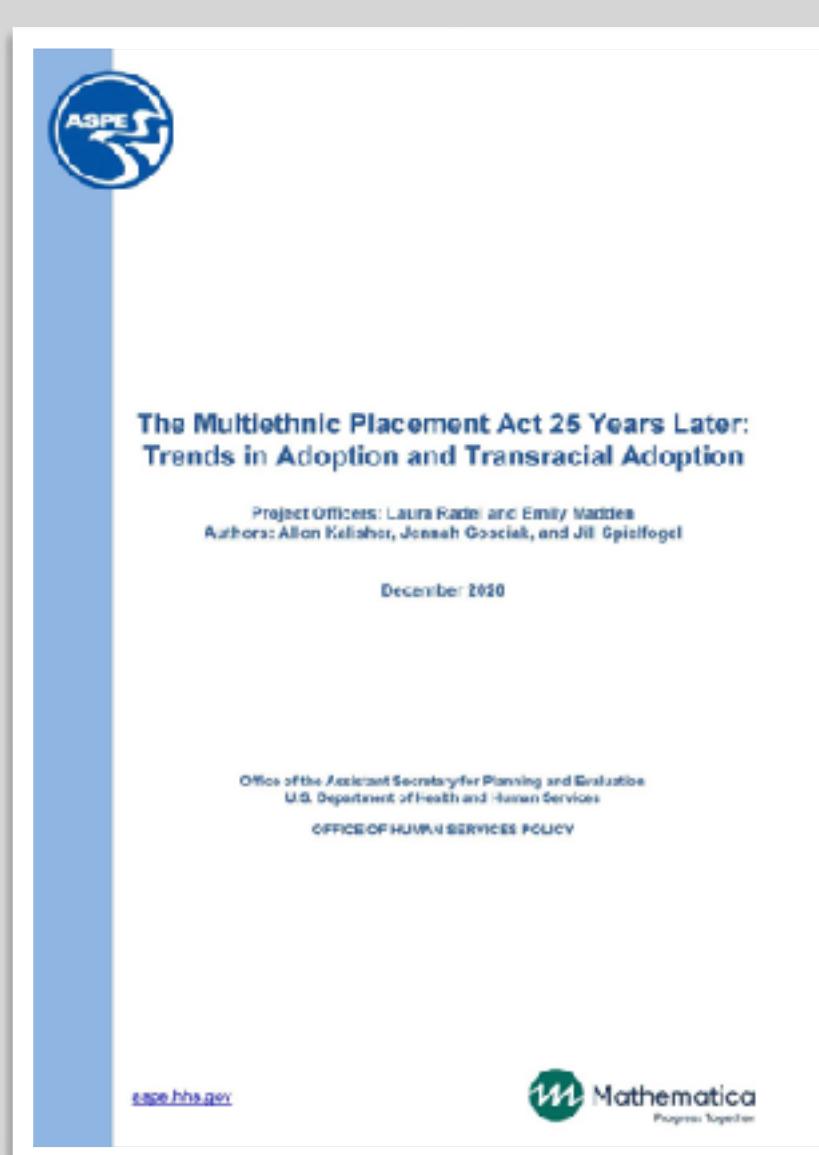
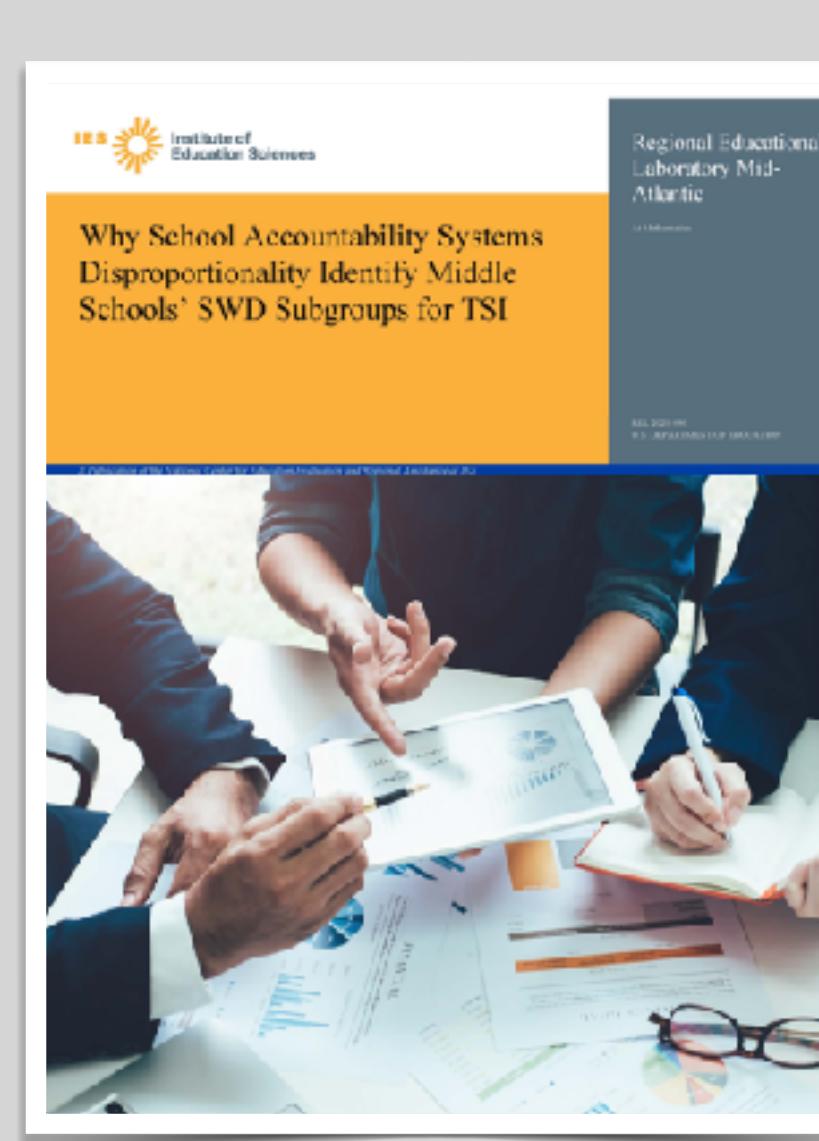


# BROWN UNIVERSITY

*Studying land ownership,  
helped developed course on  
housing justice*



*Data programming for policy*



NYU's Washington  
Square Park  
Campus



**NYU**

# WAGNER

*Urban planning*

# Tanya Popli

Masters student in Public Administration (MPA)

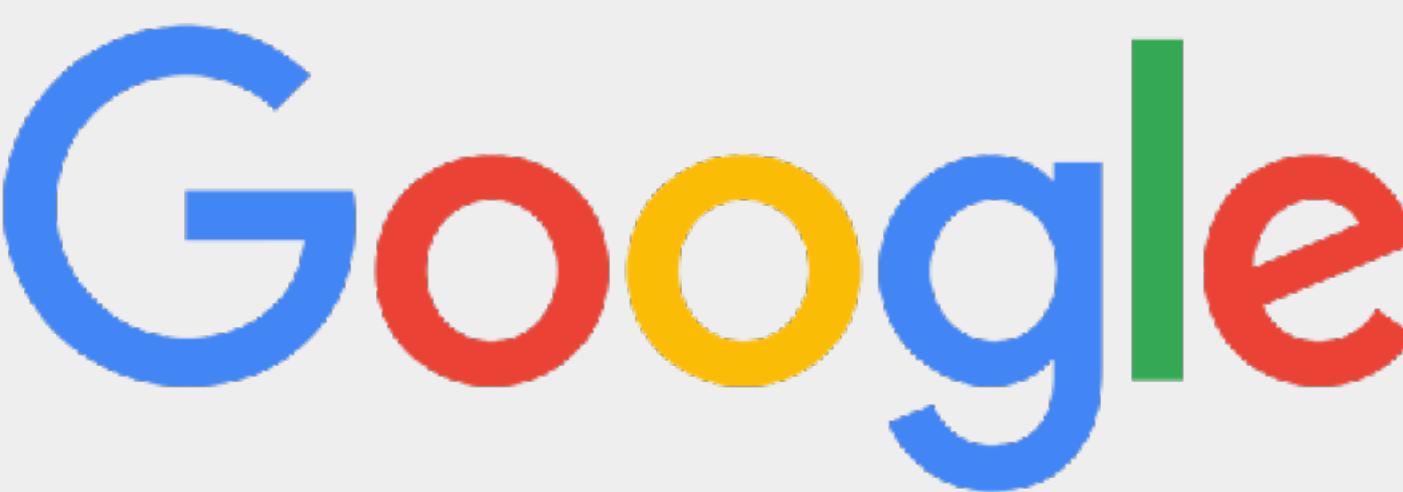


- Concentration in **Science, Technology and Infrastructure**
- Minor in **Data Science**, Certification in **Systems Thinking**
- Aiming to research **Systems Thinking in AI** and **Content Moderation**

**Fun Fact: Undergraduate degree in Mechanical & Automotive Engineering!**



Data Scientist at Mu Sigma  
developing ML models for  
client use cases



Data Analyst at Google Ads,  
Trust & Safety

#### Update to Adult content policy (March 2021)

Google provides translated versions of our Help Center, though they are not meant to change the content of our policies. The English version is the official language we use to enforce our policies. To view this article in a different language, use the language dropdown at the bottom of the page.

In March 2021, the Google Ads [Adult content policy](#) will be updated to restrict sexually suggestive live streaming, sexually suggestive live chat, sexually suggestive role playing games, and general dating with sexually suggestive elements and themes. These content categories are still allowed, but will serve on a restricted basis. Advertisers promoting this content may notice an impact to their campaigns.

Additionally, we'll be making editorial changes to the policy for improved clarity and understanding. These changes will not negatively impact advertiser campaigns and will include:

- Changing the title of the policy to "Sexual content." All references to the "Adult content" policy within Google Ads policies will be updated to "Sexual content."
- Breaking out current "Adult" and "Non-family safe" categories into more nuanced content categories.
- Updating titles, descriptions, and examples for the more nuanced categories.

We will begin enforcing the policy update on March 29, 2021, with full enforcement ramping up over approximately four weeks.

The updated Sexual content policy will read as follows:

#### Update to Google Ads sexually explicit content policy (December 2021)

Google provides translated versions of our Help Center, though they are not meant to change the content of our policies. The English version is the official language we use to enforce our policies. To view this article in a different language, use the language dropdown at the bottom of the page.

In December 2021, Google Ads sexually explicit content policy violations will be updated to egregious policy violations. Egregious policy violations result in account suspension upon detection and without prior warning. We will begin enforcing the policy update on December 1, 2021, with full enforcement ramping up over approximately 4 weeks.

We take violations of the sexually explicit content policy very seriously and will consider them egregious according to the timeline set out in this change log. If we find violations of this policy, we will suspend your Google Ads accounts upon detection and without prior warning, and you will not be allowed to advertise with us again.

Please review this policy update to determine whether or not any of your ads fall in scope of the policy, and if so, remove those ads before December 1, 2021.

#### Compensated sexual acts policy (February 2021)

Google provides translated versions of our Help Center, though they are not meant to change the content of our policies. The English version is the official language we use to enforce our policies. To view this article in a different language, use the language dropdown at the bottom of the page.

In February 2021, the Google Ads [Adult content policy](#) will be updated. All prohibited adult content will move to the [Inappropriate content policy](#). Additionally, we will prohibit compensated dating or sexual arrangements where one participant is expected to provide energy, gifts, financial support, mentorship, or other valuable benefits to another participant such as "Sugar" dating.

We will begin enforcing the policy update on February 11, 2021, with full enforcement ramping up over approximately 5 weeks.

The following categories will move from the [Adult content policy](#) into the [Inappropriate content policy](#):

- Sexually explicit content
- Child sexual abuse imagery
- Mail-order brides
- Adult themes in family content

Content that may be interpreted as promoting a sexual act in exchange for compensation will be moved into a new Compensated sexual acts category, and an additional example will be added.



Policy Enforcement  
Management

#### What is a pre-teen supervised experience on YouTube?

A pre-teen supervised experience on YouTube is a parent-managed version of regular YouTube and YouTube Music for children under 13 (or [relevant age in their country/region](#)).

With a supervised account, parents select a [content setting](#) that limits the videos and music children under 13 can find and play. Supervised accounts also change the features they can use, the default account settings, and the ads they see. Learn more about [creating a supervised account](#).

#### Elections misinformation policies

On June 2, 2023, we updated how this policy applies to past US election outcomes. Learn more in our blog [↗](#).

Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes certain types of misinformation that can cause real-world harm, like certain types of technically manipulated content, and content interfering with democratic processes.





# Some nitty gritty details of how class will run

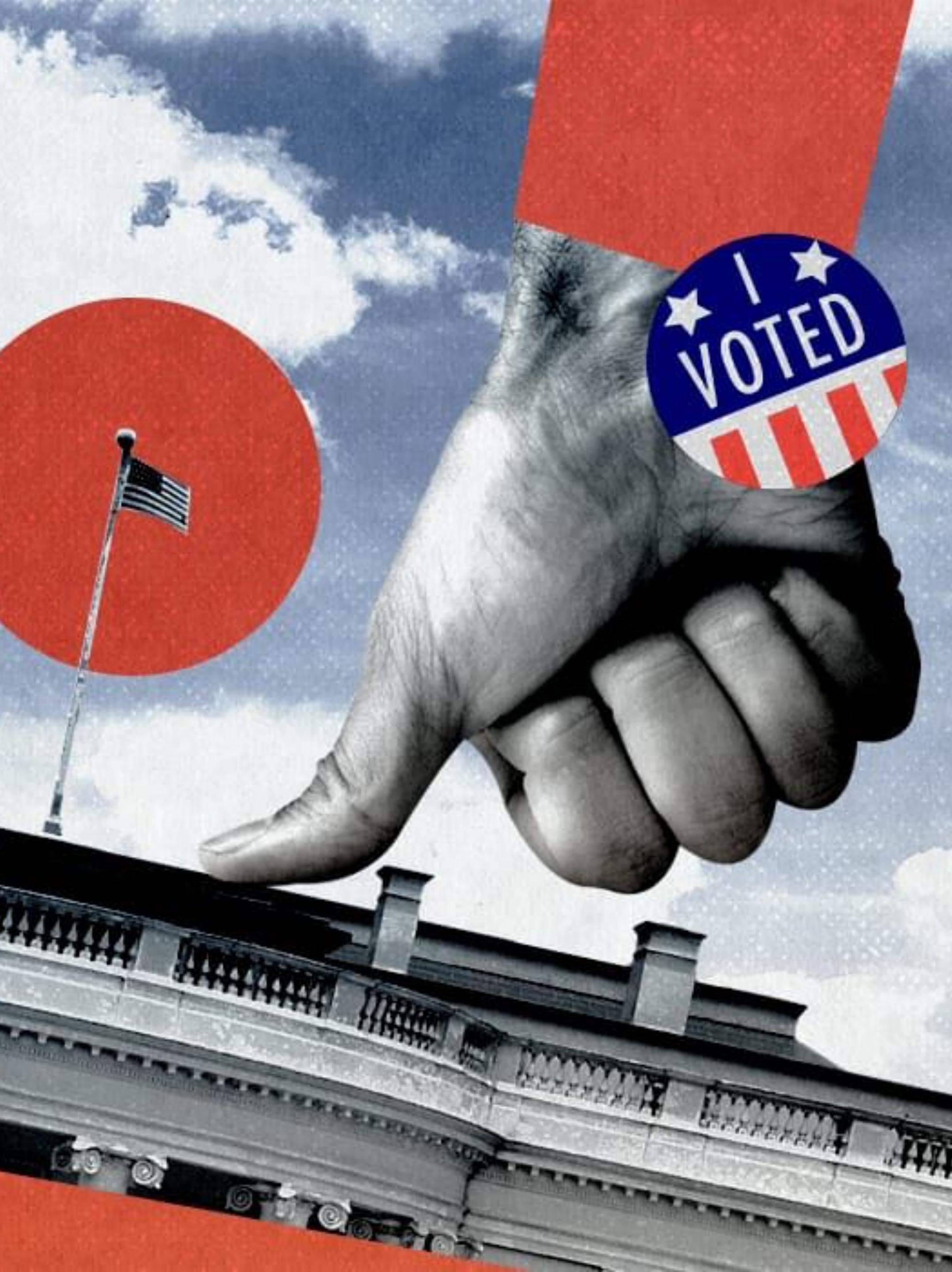
- Lectures here (Hollister 110) Tuesday / Thursday 1:25-2:40
- Lab/discussion sections 50 minutes Friday morning — **required**
  - Warren B02 — Jennah
  - HEB 101 — Tanya
- You will turn in a small data artifact on Gradescope at the end of each lab meeting, produced by making a small modification to what's in that week's notebook
- HW is due Tuesdays at 1:25 (beginning of class) sharp — expect to deliver a data product based on a larger modification of the notebook, together with a question about the reading.
- (This week's reading: 9 pages from the introduction of **Seeing Like a State**)

# Some nitty gritty details of how class will run

- 4 point grading scale for everything – 75% score at the end of the term will earn an A-
- Two short in-class exams (~ Feb 14, Mar 21) and a final project
- Project groups of 4-5 people. Syllabus contains some timing details.
- We will use Canvas and Gradescope (linked).
- Important: **UDL**, Universal Design for Learning. Major commitment to accessibility, within the confines of administrability for a class of this size.
- Course norms: be respectful, be kind, contribute your skillset and receive tools from others.



# Seeing Like the States



# Legibility

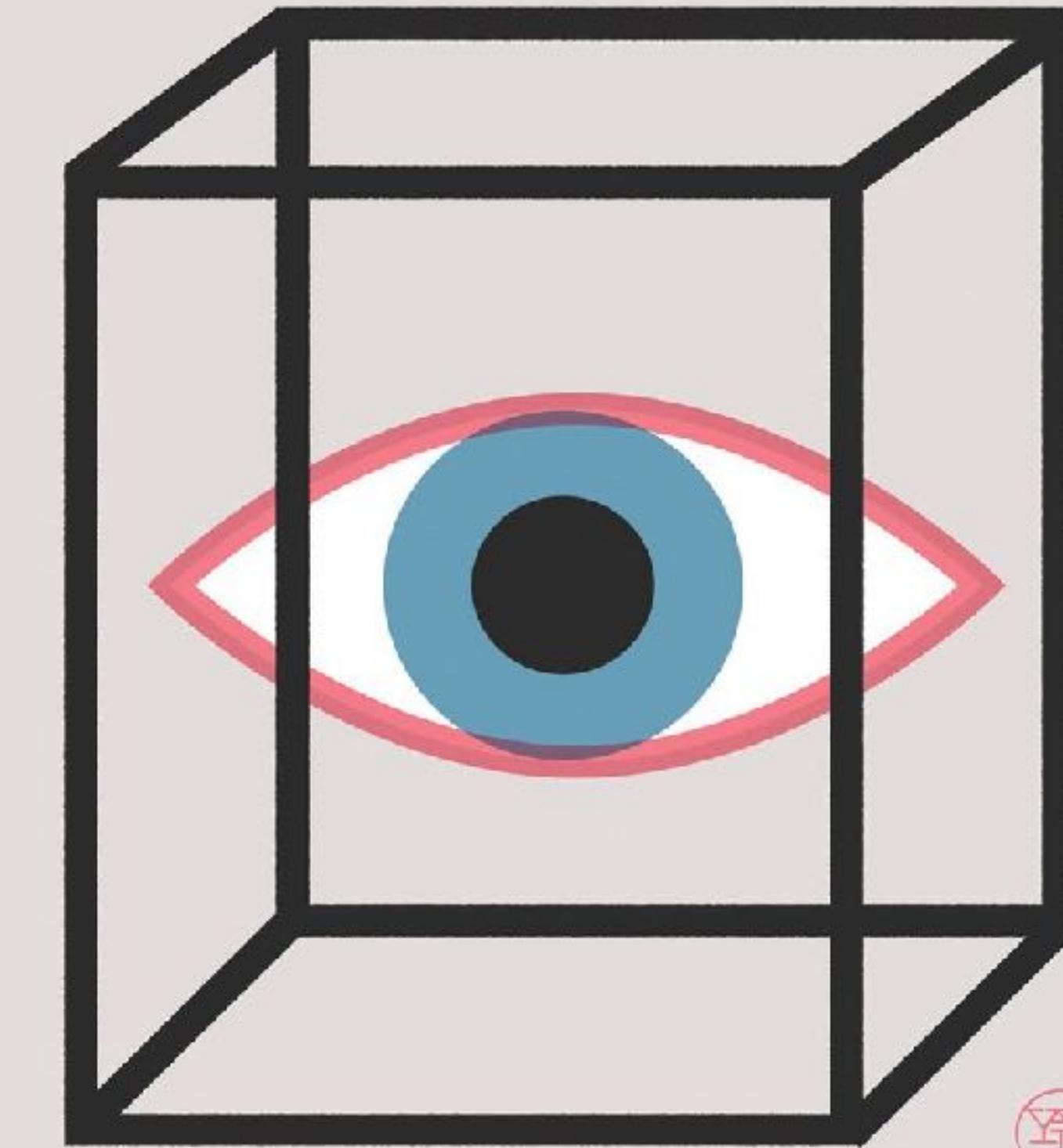
## 2 Introduction

The more I examined these efforts at sedentarization, the more I came to see them as a state's attempt to make a society legible, to arrange the population in ways that simplified the classic state functions of taxation, conscription, and prevention of rebellion. Having begun to think in these terms, I began to see legibility as a central problem in statecraft. The premodern state was, in many crucial respects, partially blind; it knew precious little about its subjects, their wealth, their landholdings and yields, their location, their very identity. It lacked anything like a detailed "map" of its terrain and its people. It lacked, for the most part, a measure, a metric, that would allow it to "translate" what it knew into a common standard necessary for a synoptic view. As a result, its interventions were often crude and self-defeating.

## Seeing Like a State

How Certain Schemes to Improve the Human Condition Have Failed

James C. Scott



# Census and Classification

TABLE II

### *U.S. Census Race Categories, 1790-1840*

1790	1800	1810
Free White Males, Free White Females	Free White Males, Free White Females	Free White Males, Free White Females
All Other Free Persons	All Other Free Persons, except Indians Not Taxed	All Other Free Persons, except Indians Not Taxed
Slaves	Slaves	Slaves
1820	1830	18-
Free White Males, Free White Females	Free White Persons	Free White Pe
Free Colored Persons	Free Colored Persons	Free Colored P
All Other Persons, except Indians Not Taxed	Slaves	Slaves
Slaves		

Source: United States Bureau of the Census.

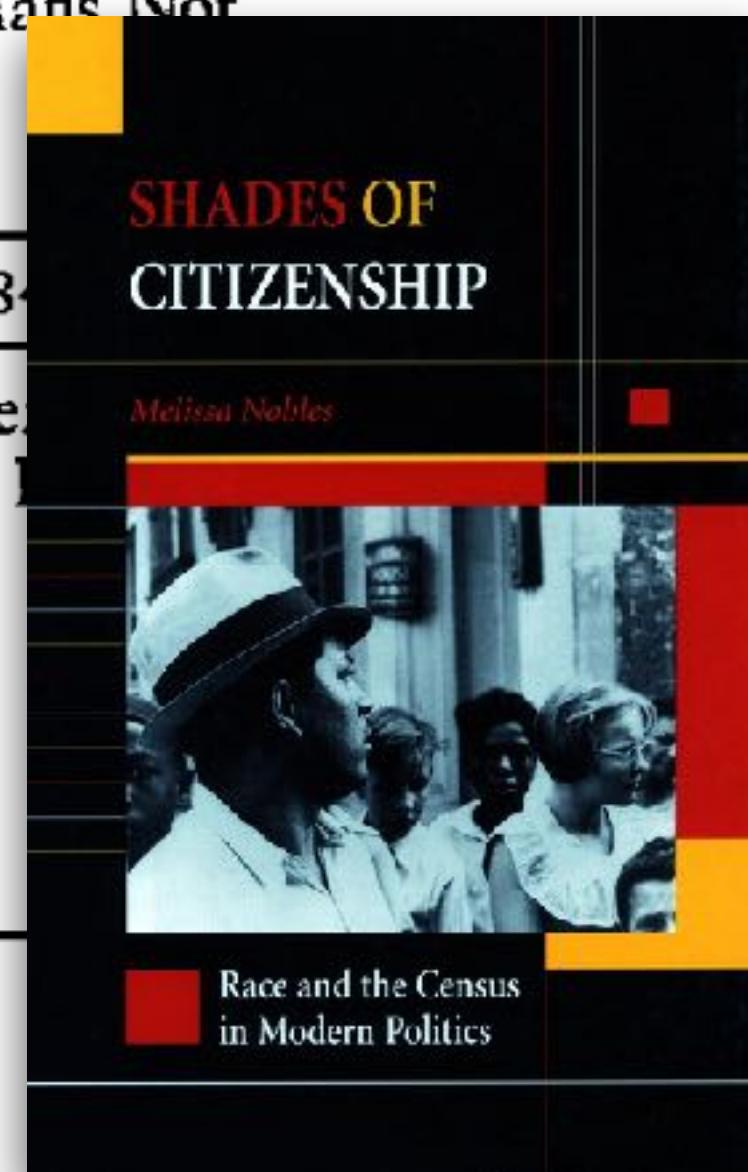


TABLE 2

1850-1920—“The Mulatto Category” and Race Science							
1850	1860 <sup>a</sup>	1870	1880	1890	1900	1910	1920
<i>b</i>	<i>b</i>	White	White	White	White	White	White
Black	Black	Black	Black	Black	Black	Black	Black
Mulatto	Mulatto (Indian)	Mulatto Chinese Indian	Mulatto Chinese Indian	Mulatto Quadroon Octroon Chinese Japanese Indian	Chinese Japanese Indian	Mulatto Chinese Indian Other (+ write in)	Mulatto Indian Chinese Japanese Filipino Hindu Korean Other (+ write in)
1930-1960 “The One-Drop Rule”				1970 Post Civil Rights	1980-1990 Post Statistical Directive #15		
1930	1940	1950	1960	1970	1980	1990	
White	White	White	White	White	White	White	White
Negro	Negro	Negro	Negro	Negro or Black	Negro or Black	Black or Negro	Black or Negro
Mexican	Indian	Indian	American-	Indian	Japanese	Japanese	Indian
Indian	Chinese	Japanese	Indian	(Amer.)	Chinese	Chinese	(Amer.)
Chinese	Japanese	Chinese	Japanese	Japanese	Filipino	Eskimo	
Japanese	Filipino	Filipino	Chinese	Chinese	Korean	Aleut	
Filipino	Hindu	(Other	Filipino	Chinese	Vietnamese	Chinese	
Hindu	Korean	race—	Hawaiian	Filipino	Indian	Filipino	
Korean	(Other	spell out)	Part-	Hawaiian	Hawaiian	Hawaiian	
(Other	races,			Korean	(Amer.)	Hawaiian	
races, spell out	spell out			Other	Asian	Korean	
in full)	in full)			(print race)	Indian	Vietnamese	
					Hawaiian	Japanese	
					Guama- nian	Asian	
					Samoan	Indian	
					Eskimo	Samoan	
					Aleut	Guama- nian	
					Other	Other API	
					(specify)	(Asian or Pacific)	

# Who Names Race?

Office of Management and Budget (OMB)

DIRECTIVE NO. 15

Race and Ethnic Standards for Federal Statistics  
and Administrative Reporting

**(as adopted on May 12, 1977)**

This Directive provides standard classifications for record keeping, collection, and presentation of data on race and ethnicity in Federal program administrative reporting and statistical activities. These classifications should not be interpreted as being scientific or anthropological in nature, nor should they be viewed as determinants of eligibility for participation in any Federal program. They have been developed in response to needs expressed by both the executive branch and the Congress to provide for the collection and use of compatible, nonduplicated, exchangeable racial and ethnic data by Federal agencies.

## **1. Definitions**

The basic racial and ethnic categories for Federal statistics and program administrative reporting are defined as follows:

- a. American Indian or Alaskan Native. A person having origins in any of the original peoples of North America, and who maintains cultural identification through tribal affiliation or community recognition.
- b. Asian or Pacific Islander. A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.
- c. Black. A person having origins in any of the black racial groups of Africa.
- d. Hispanic. A person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race.
- e. White. A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

## **2. Utilization for Record keeping and Reporting**

To provide flexibility, it is preferable to collect data on race and ethnicity separately. If separate race and ethnic categories are used, the minimum designations are:

- a. Race:  
-- American Indian or Alaskan Native

# and How Does It Matter?

**The decision to allow respondents to choose multiple racial categories...**

- (a) was declared in OMBSD15 update in 1997,
- (b) cost millions to implement,
- (c) was opposed by some civil rights groups, and contested with a march in Washington,
- (d) increased # of races from 4 to 63 (so race/ethnicity has 126 combos)
- (e) is reflected in all subsequent research that filters through these datasets.

## How the 2020 census asked about Hispanic origin and race

→ NOTE: Please answer BOTH Question 6 about Hispanic origin and Question 7 about race. For this census, Hispanic origins are not races.

### 6. Are you of Hispanic, Latino, or Spanish origin?

- No, not of Hispanic, Latino, or Spanish origin
- Yes, Mexican, Mexican Am., Chicano
- Yes, Puerto Rican
- Yes, Cuban
- Yes, another Hispanic, Latino, or Spanish origin – Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc. ↗

### 7. What is your race?

Mark  one or more boxes AND print origins.

- White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc. ↗

- Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc. ↗

- American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc. ↗

- |   |                                     |  |
|---|-------------------------------------|--|
| <input type="checkbox"/> Chinese  | <input type="checkbox"/> Vietnamese | <input type="checkbox"/> Native Hawaiian   |
| <input type="checkbox"/> Filipino   | <input type="checkbox"/> Korean     | <input type="checkbox"/> Samoan  |
| <input type="checkbox"/> Asian Indian   | <input type="checkbox"/> Japanese   | <input type="checkbox"/> Chamorro  |
| <input type="checkbox"/> Other Asian –<br>Print, for example,<br>Pakistani, Cambodian,<br>Hmong, etc. ↗ |                                     | <input type="checkbox"/> Other Pacific Islander –<br>Print, for example,<br>Tongan, Fijian,<br>Marshallese, etc. ↗<br><div style="border: 1px solid #ccc; width: 100%; height: 1.2em; margin-top: 5px;"></div> |

- Some other race – Print race or origin. ↗

## Decennial Census

**P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE**

2020: DEC Redistricting Data (PL 94-171) ▾ Universe: Total population

## **Local Census**

#### HISPANIC OR LATINO, AND NOT HISPANIC

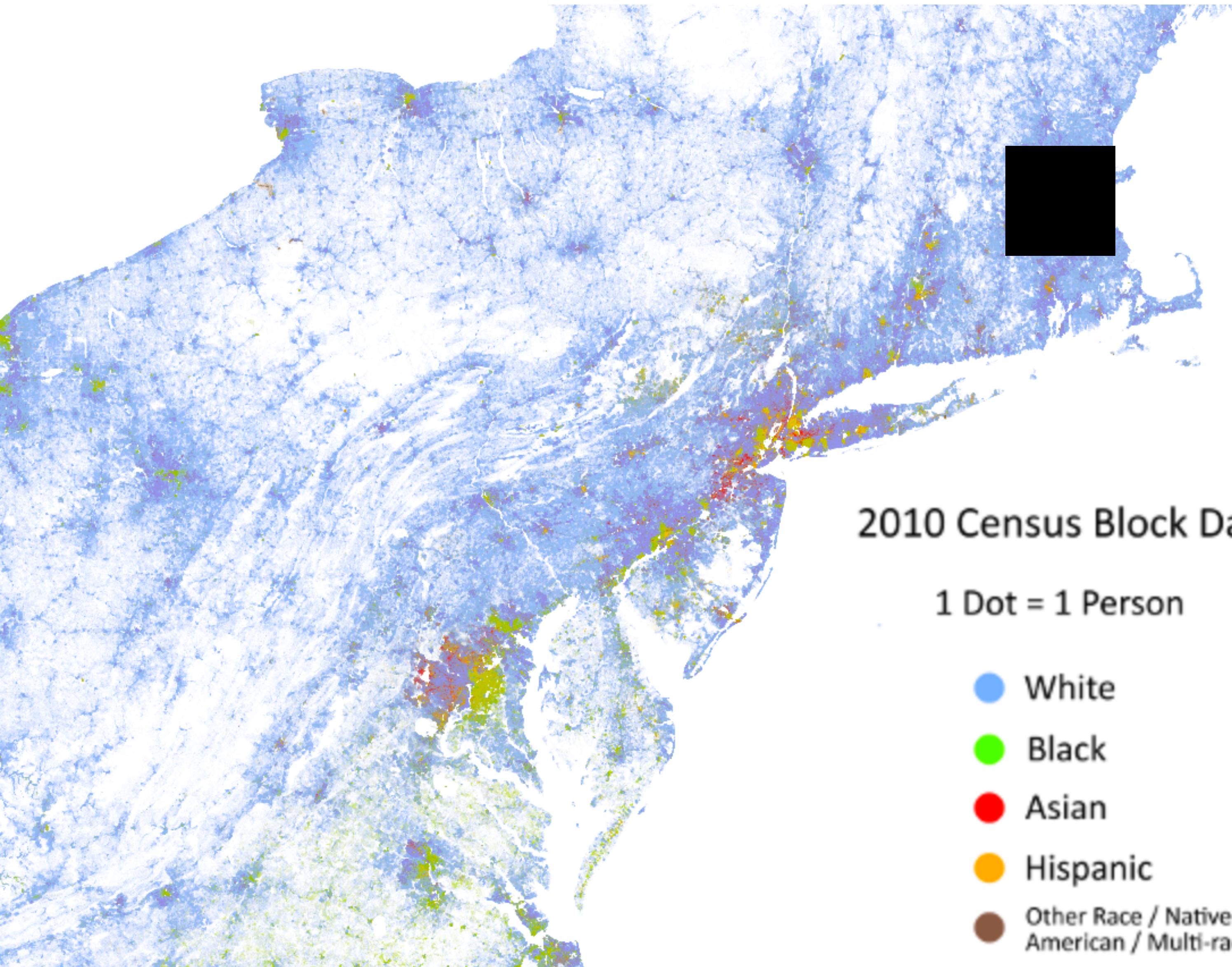
2022-23 NC Redistricting Data (PL 94-171) ▾ Universe: Total population

C OR LATINO, AND NOT HISPANIC OR LATINO BY RACE

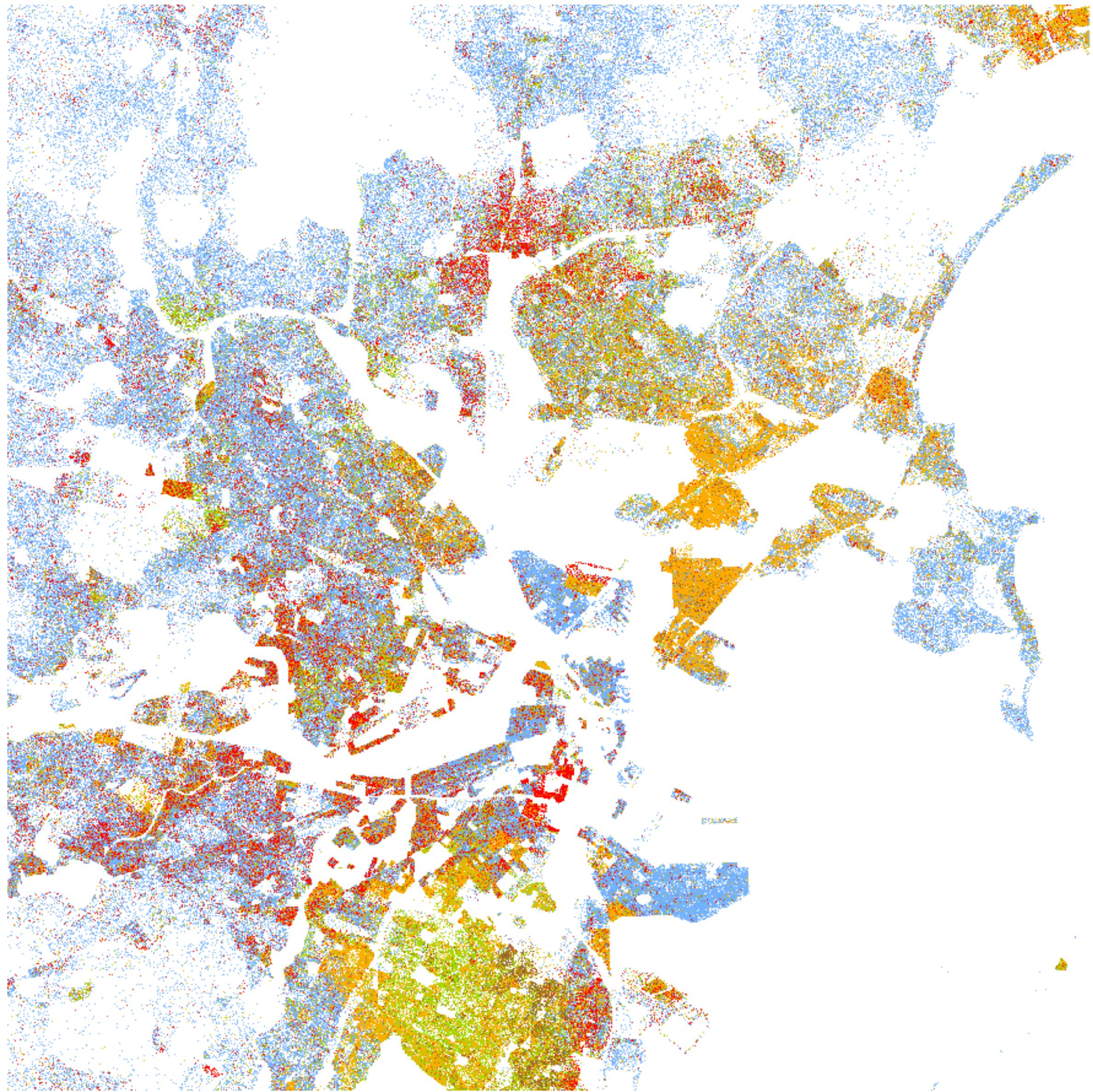
Meeting Date (PT 3M 12/31)  | In years: Total population

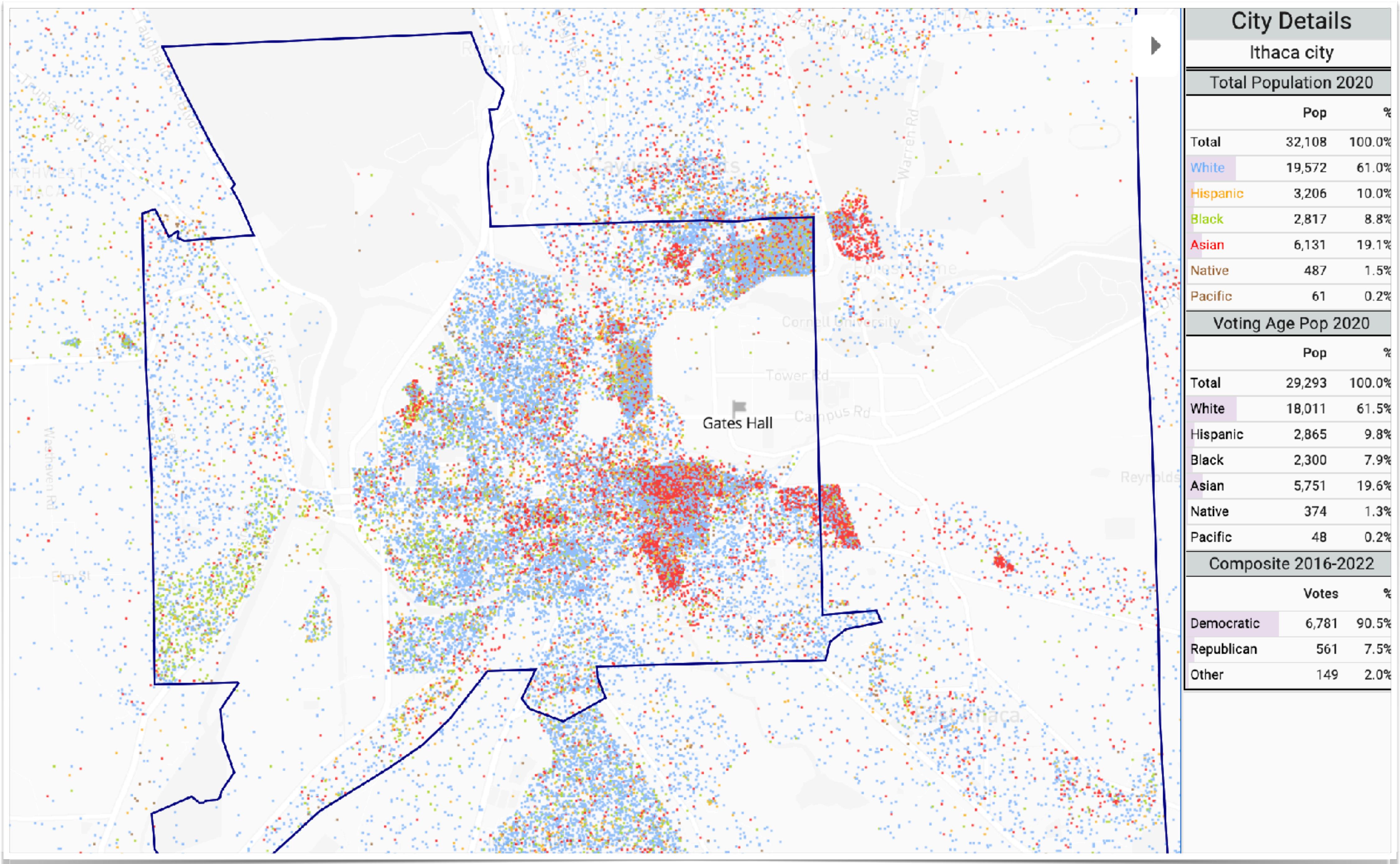
#### WHITE OR LATINO, AND NOT HISPANIC OR - ATION BY RACE

Using Data: Pupil's "I" and "Me" Metacognition and Self-Regulation









what are the stakes?

# What Are Census Counts For?

**apportionment and political representation**

“according to their respective Numbers”

**funding formulas**

Census study: “This paper finds that 132 programs used Census Bureau data to distribute more than \$675 billion in funds during fiscal year 2015.”

**race/ethnicity enables audits across government functions**

undercount, unemployment, voter turnout, lending, .....

**Table 1: Federal Assistance Distributed Using Census Bureau Data in Fiscal Year 2015**

CFDA number <sup>1</sup>	Program Name	Federal Executive Department or Agency <sup>2</sup>	Fiscal Year 2015 Funds
93.778	Medical Assistance Program	HHS	\$311,805,244,413
10.551	Supplemental Nutrition Assistance Program	USDA	\$71,035,786,000 *
93.774 (part)	Medicare Part B Physicians Fee Schedule Services	HHS	\$70,300,000,000 **
20.205	Highway Planning and Construction	DOT	\$38,479,013,855
84.063	Federal Pell Grant Program	ED	\$29,916,694,438
10.555	National School Lunch Program	USDA	\$18,915,944,292
93.558	Temporary Assistance for Needy Families	HHS	\$17,225,738,021
14.871	Section 8 Housing Choice Voucher	HUD	\$15,761,488,440 *
84.010	Title 1 Grants to Local Educational Agencies	ED	\$14,253,154,251
84.027	Special Education Grants to States	ED	\$11,382,885,850
93.600	Head Start	HHS	\$8,538,887,781

# Colonial Legacies

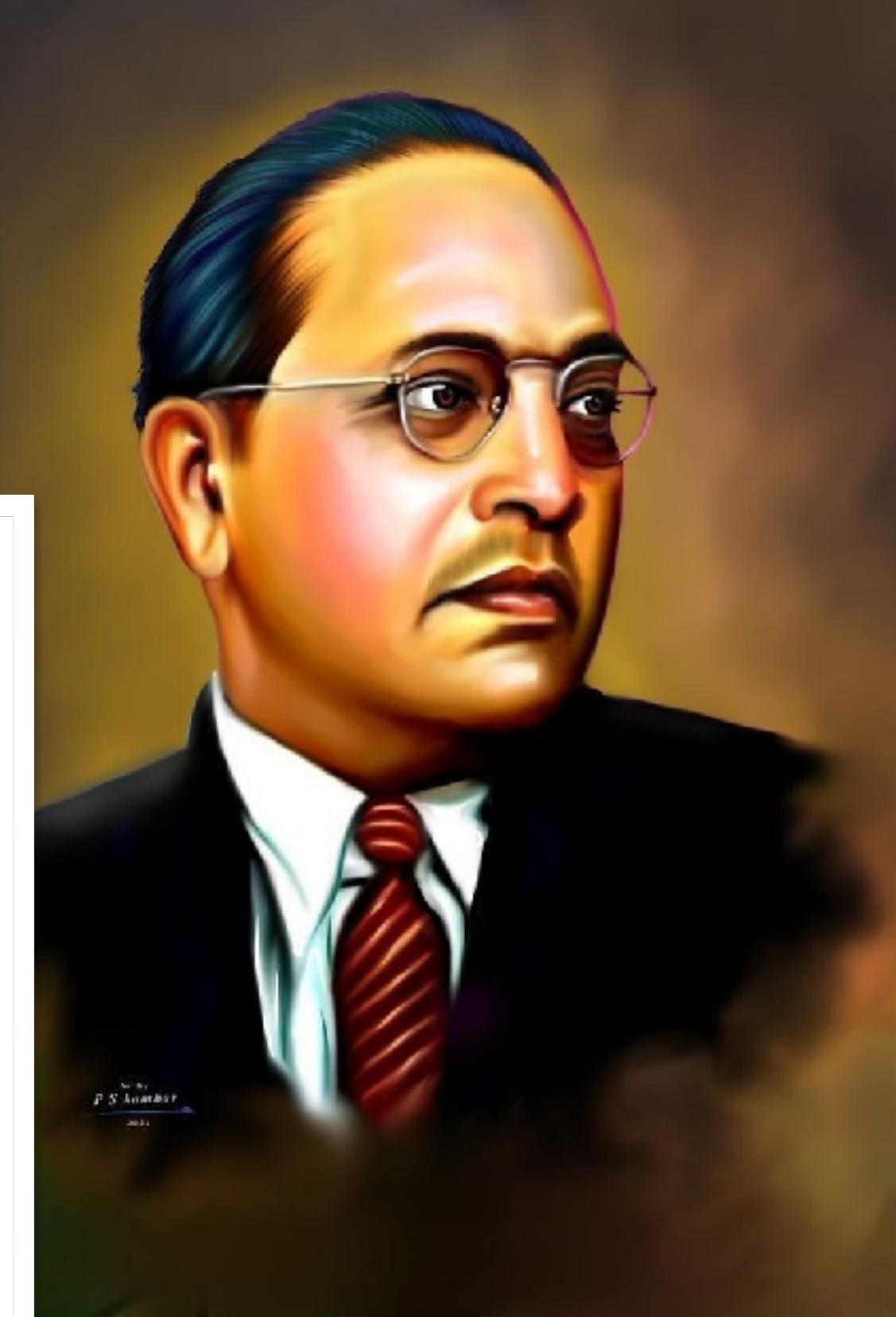
B.R. Ambedkar writes

From Mi

What is the total population of the Untouchables? It is bound to be the first thing about them is

of the Untouchables was. The table below gives the population of the Untouchables, province by province, as found by the Census Commissioner of 1911.<sup>a</sup>

Province	Total population in million	Population of Depressed Classes in million	Total Seats	Seats for the Depressed Classes
Madras	39.8	6.3	120	2
Bombay	19.5	0.6	113	1
Bengal	45.0	9.9	127	1
United Provinces	47.0	10.1	120	1
Punjab	19.5	1.7	85	-
Bihar and Orissa	32.4	9.3	100	1
Central Provinces	12.0	3.7	72	1
Assam	6.0	0.3	54	-
	221.2	41.9	791	7

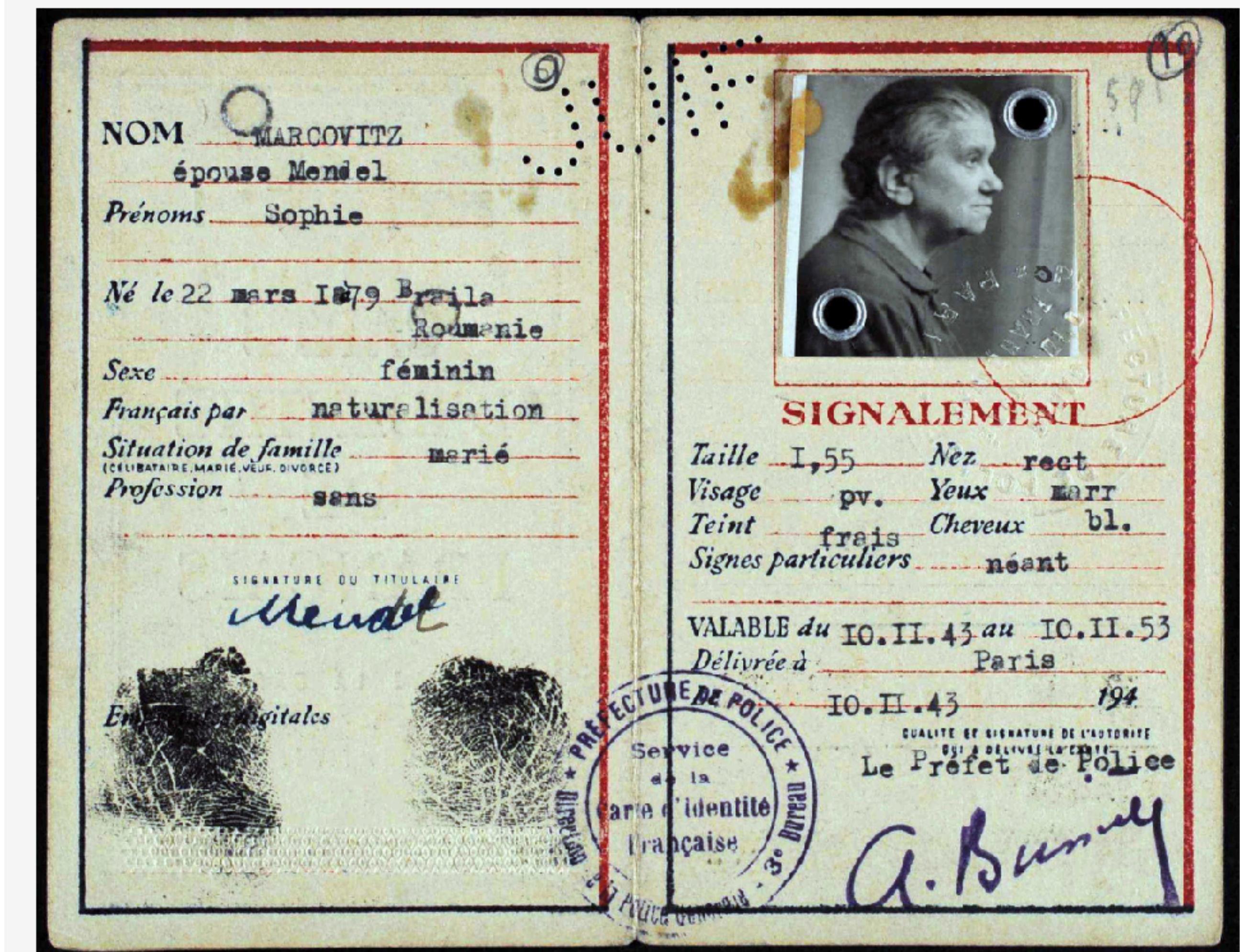


# Conflict Legacies

Today it is illegal for any government entity in **France** to ask for race or ethnicity data

Which means there are no official stats on employment disparity, incarceration, wealth, etc

Why?



source: Yad Vashem

# “At Home”: Redlining and More

**1954:** Brown v. Board of Ed mandates integration, speeds civil right movement

**1957:** Tuskegee, Alabama redrawn by Leg.

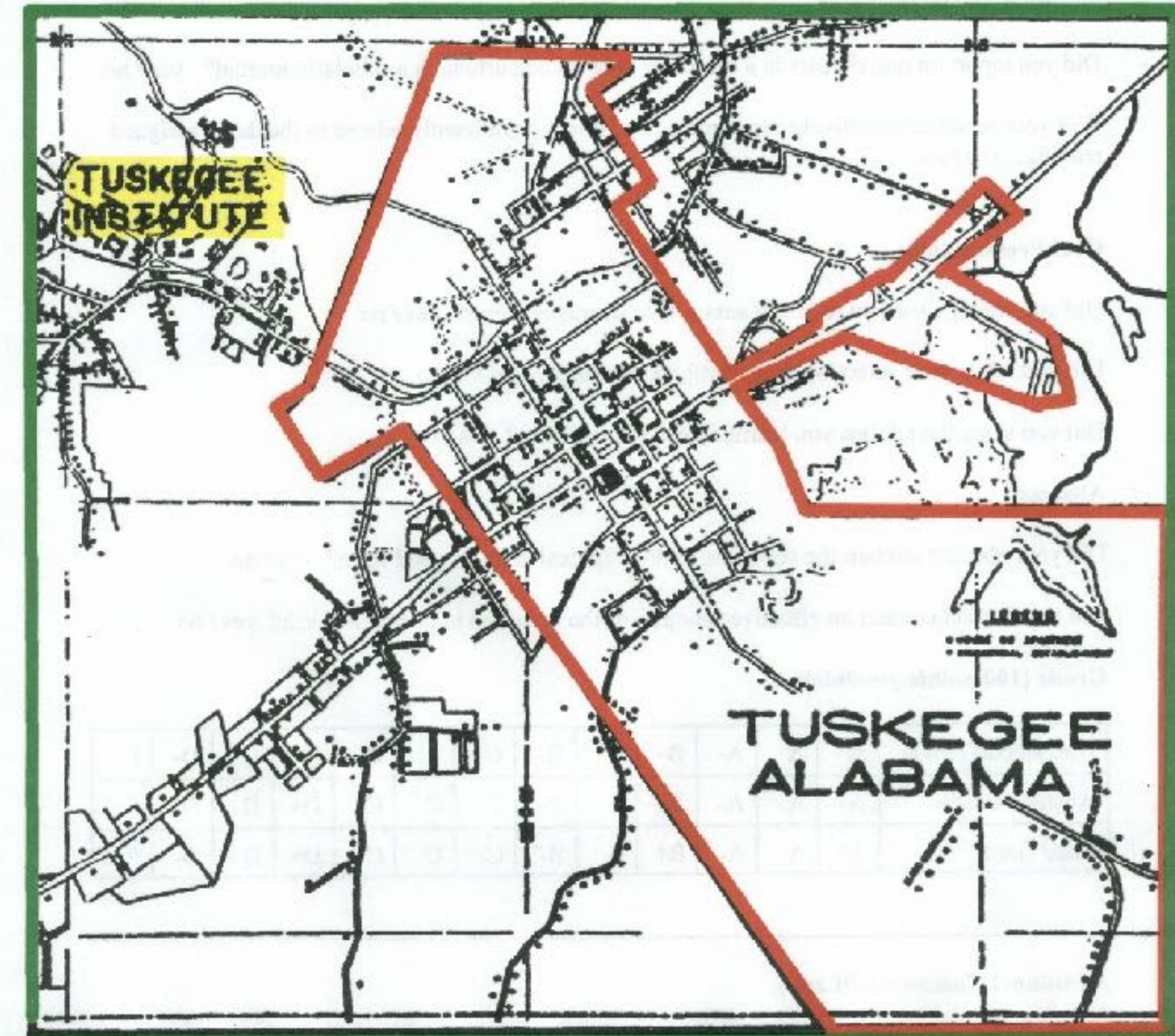
Old (**square**) – 80% Black

New (**28-sided polygon**) – 100% White

Supreme Court intervenes in Gomillion v. Lightfoot  
**(1960)**

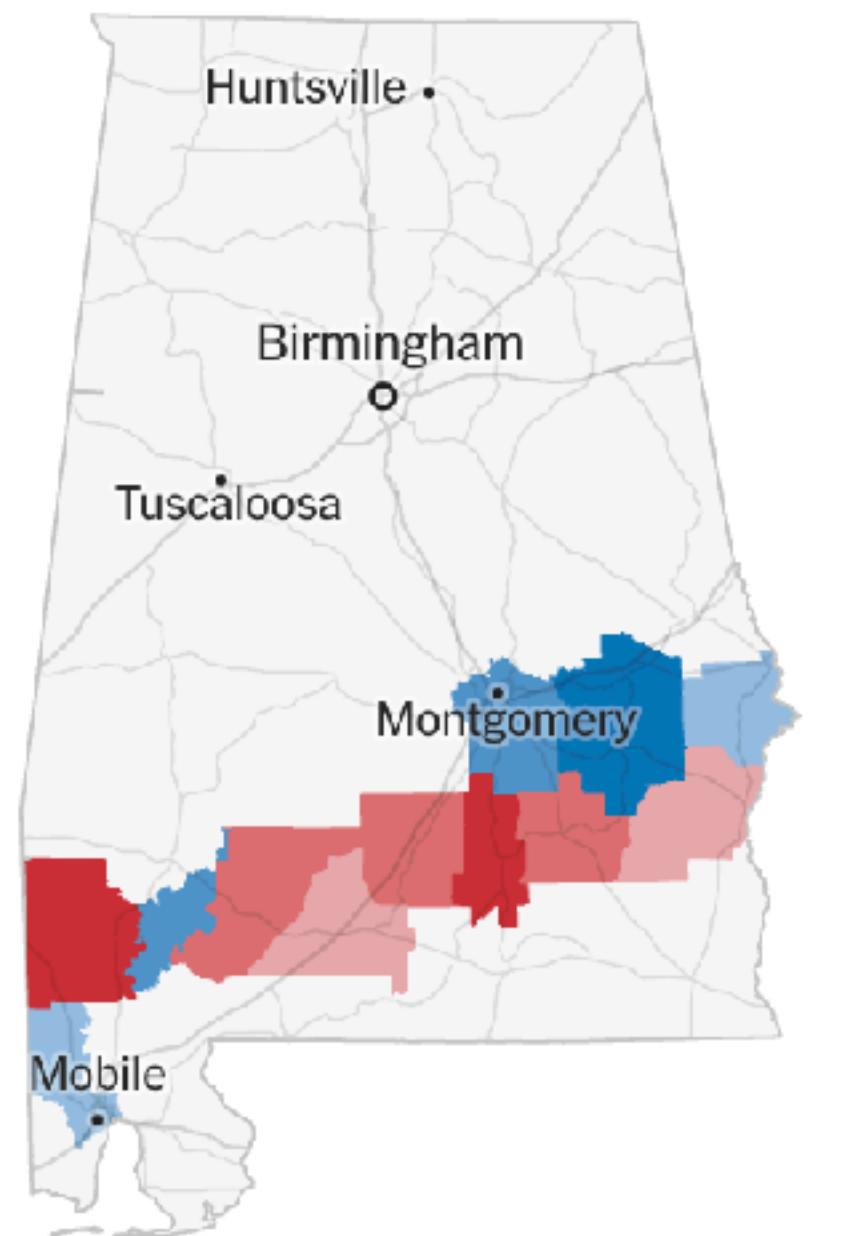
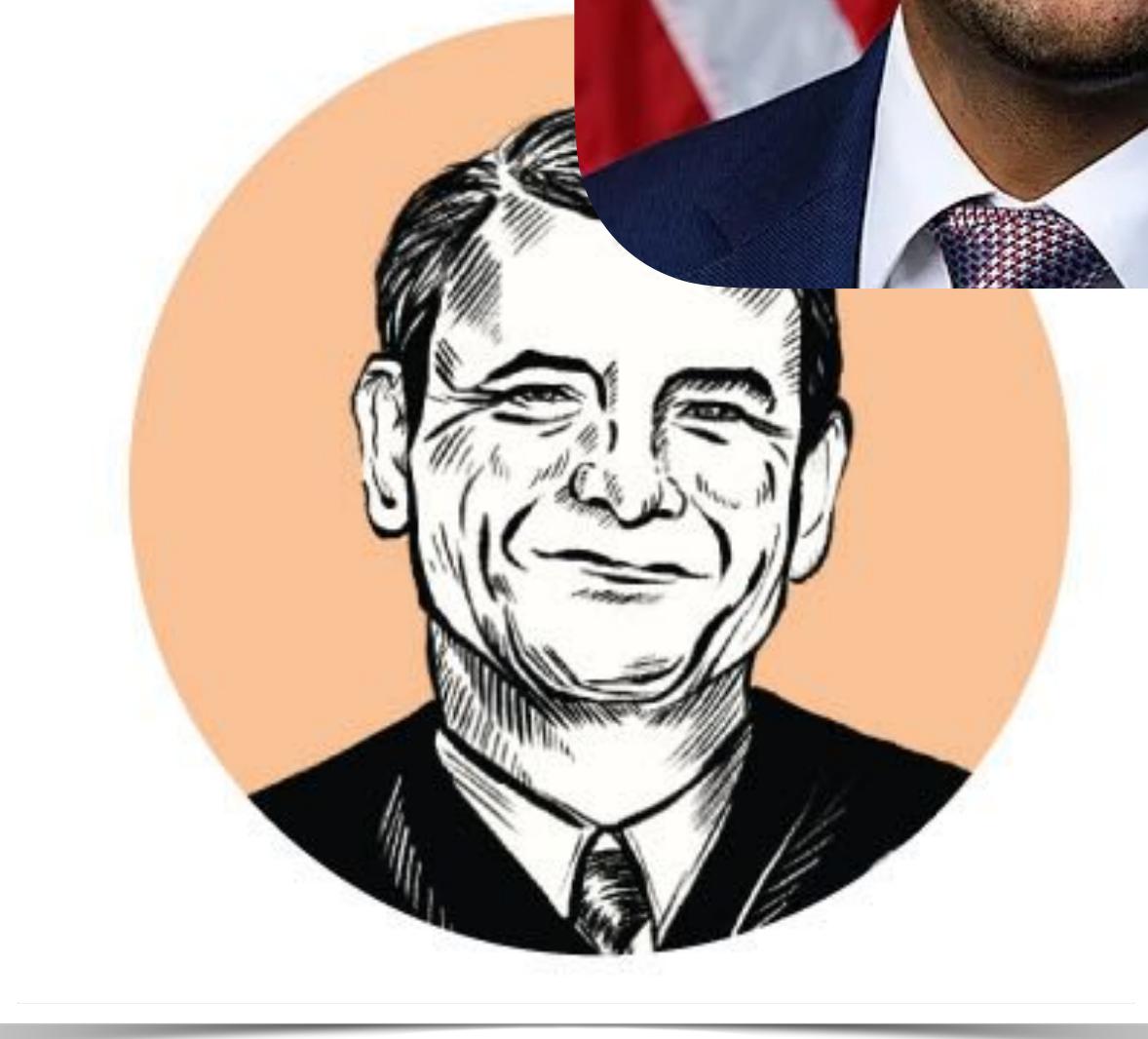
## APPENDIX TO OPINION OF THE COURT.

### CHART SHOWING TUSKEGEE, ALABAMA, BEFORE AND AFTER ACT 140



(The entire area of the square comprised the City prior to Act 140. The irregular black-bordered figure within the square represents the post-enactment city.)

# My Experience



Dem. 60 70% Rep. 60 70%

