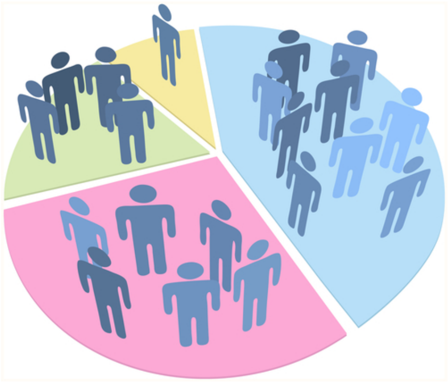
Using AI to study society: Considerations in fairness accountability transparency and ethics

Timnit Gebru

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

United States™
Census
Bureau



$1 billion per year

The Run 2016

Analysis about the next race for the White House by David Catanese

Early Polls Tough to Swallow for Bush

Hillary Clinton's Appeal Survives Scrutiny, Poll Says

Scant Gains for Romney in a Poll of Young Voters

The New York Times | CBS NEWS Poll

Americans' Views on the 2016 Presidential Campaign and the Issues

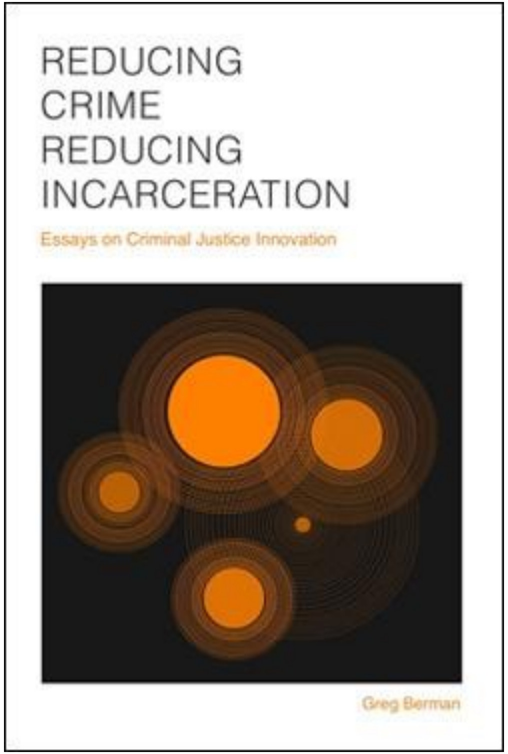
crimereduction

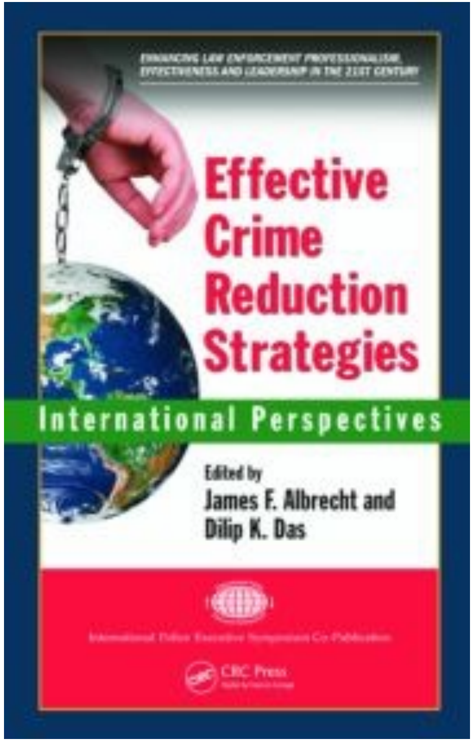
NEIGHBORHOOD WATCH

Taking Back Our Neighborhoods

OperationOurTown.org

Neighborhood Watch





safe & secure

Neighbourhood Watch.

Can we do this with computer vision?

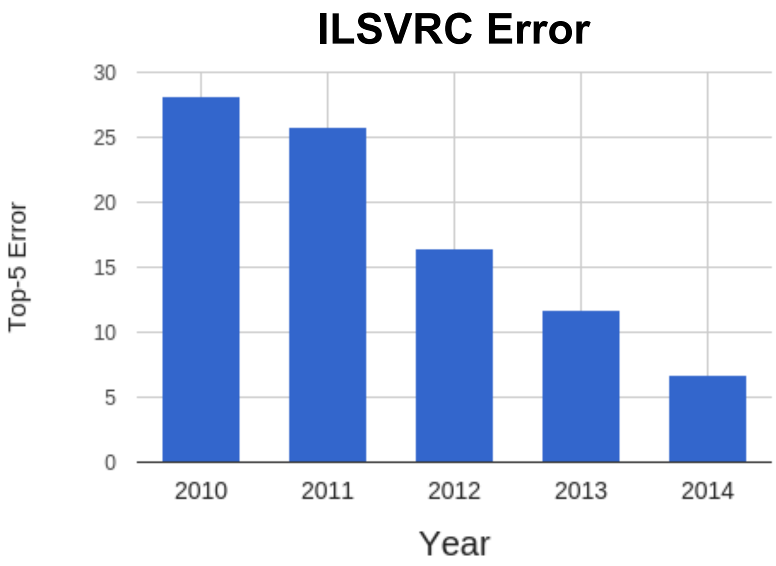
A source of data: Google Street View

2008 Toyota Prius

2008
Toyota (Japanese)
Hatchback
$9,542
45 MPG (highway)
48 MPG (city)
Hybrid

How can we use vision?

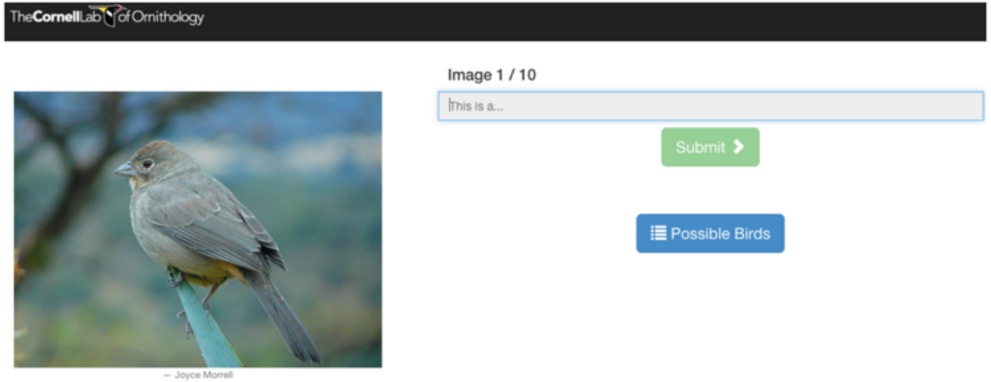
Progress in object recognition



Problem: Training data

Fine-grained datasets are notoriously difficult to collect
Require experts or extremely careful AMT pipelines

CUB-200-2011



NABirds

Problem: Class List

Car types are not well-defined!

2006 Toyota Prius

2007 Toyota Prius

Grouping Classes Together

2,657 classes since 1990

Group Car Classes Without Experts

-Graph based algorithm to group visually similar cars

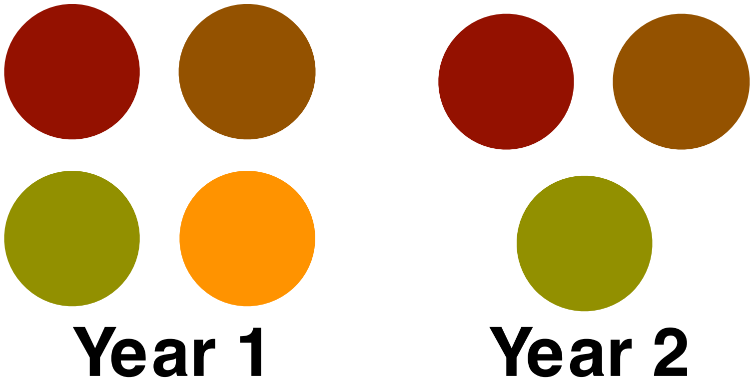
-Only task is to answer whether 2 cars are visually the same

-Use Amazon Mechanical Turk (AMT)

Gebru et al. CHI 2017

Group Car Classes Without Experts

Graph based algorithm to group visually similar cars



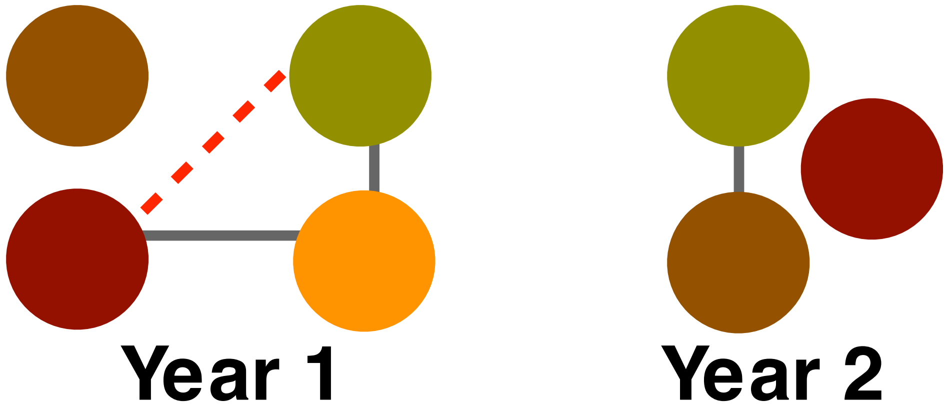
Gebru et al. CHI 2017

Group Car Classes Without Experts

Gebru et al. CHI 2017

Group Car Classes Without Experts

Add an edge between two cars if AMT task returns a yes

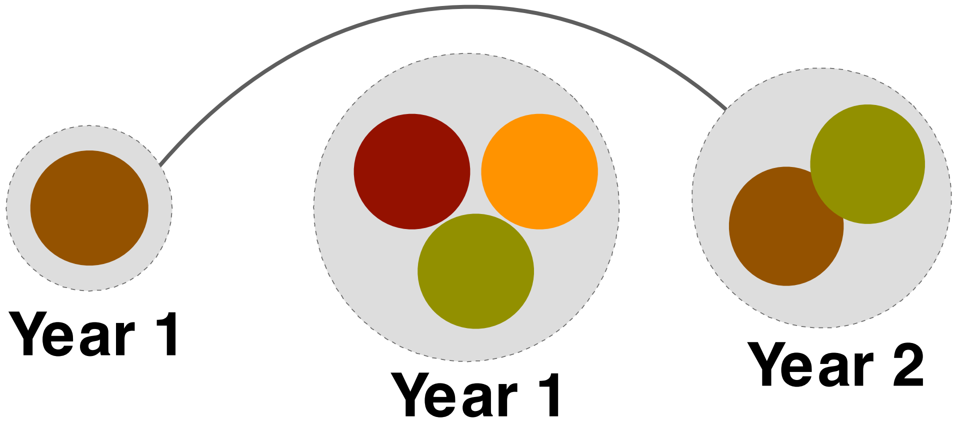


Gebru et al. CHI 2017

Group Car Classes Without Experts

Gebru et al. CHI 2017

Group Car Classes Without Experts



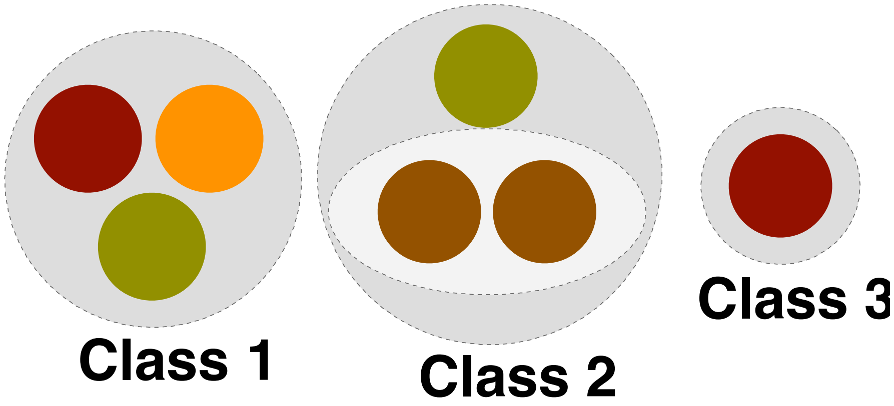
Gebru et al. CHI 2017

Group Car Classes Without Experts

Gebru et al. CHI 2017

Group Car Classes Without Experts

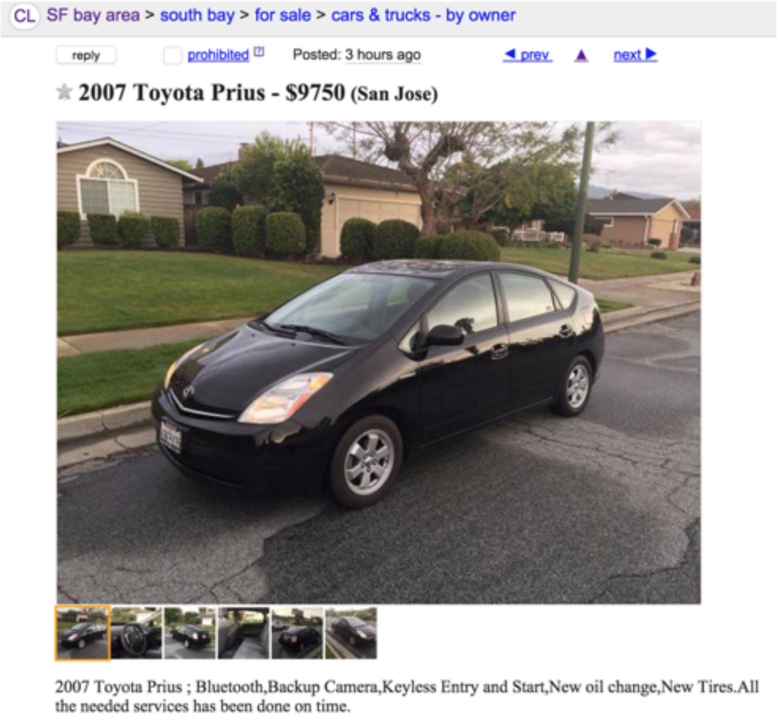
Use connected components to create final class list



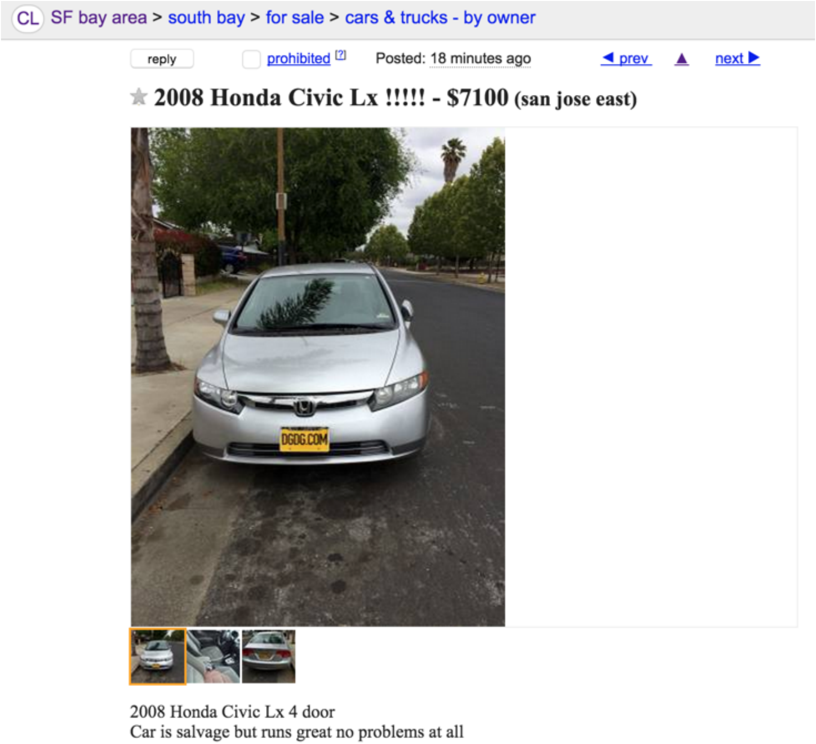
Gebru et al. CHI 2017

Training Data: An Insight

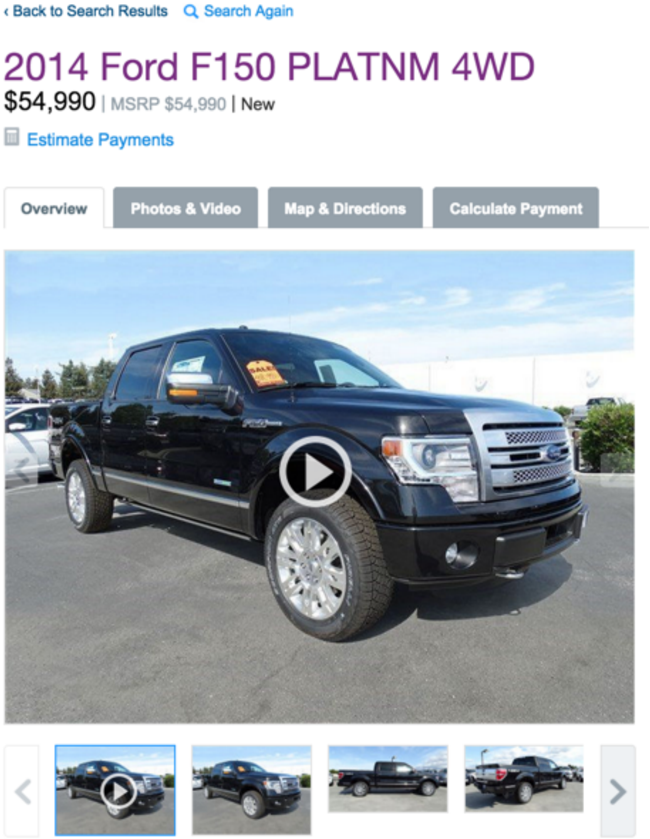
People don’t know all cars, but they know their own car



Gebru et al. CHI 2017



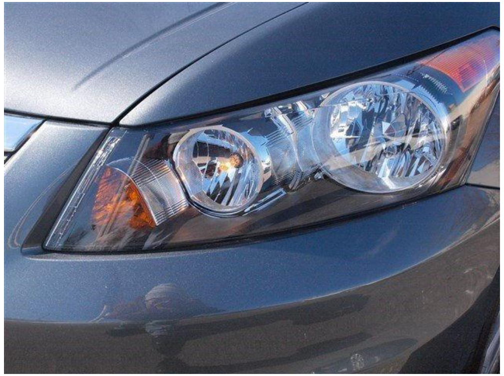
craigslist



cars.com

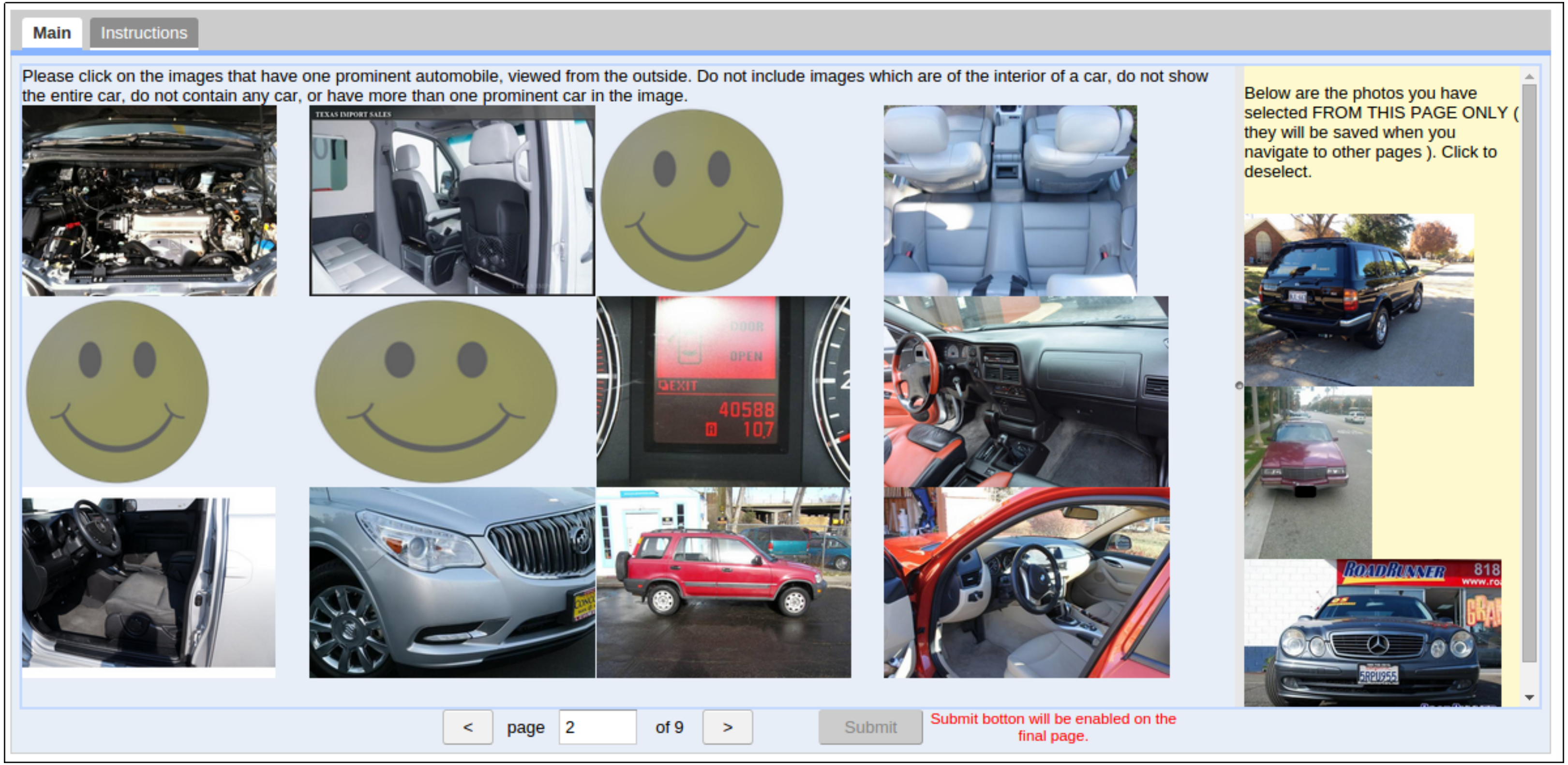
Gebru et al. CHI 2017

Bad: Closeup



Bad: Interior

Good


Main
Instructions
Please click on the images that have one prominent automobile, viewed from the outside. Do not include images which are of the interior of a car, do not show the entire car, do not contain any car, or have more than one prominent car in the image.
TEXAS IMPORT SALES
DOOR
OPEN
EXIT
40588
107
Below are the photos you have selected FROM THIS PAGE ONLY (they will be saved when you navigate to other pages). Click to deselect.
ROADRUNNER
818
5RPU955
<
page
2
of 9
>
Submit
Submit botton will be enabled on the final page.

313,099 labels at ~1/20th the cost of experts

Training Data: Another problem

Still need detection data + real world images

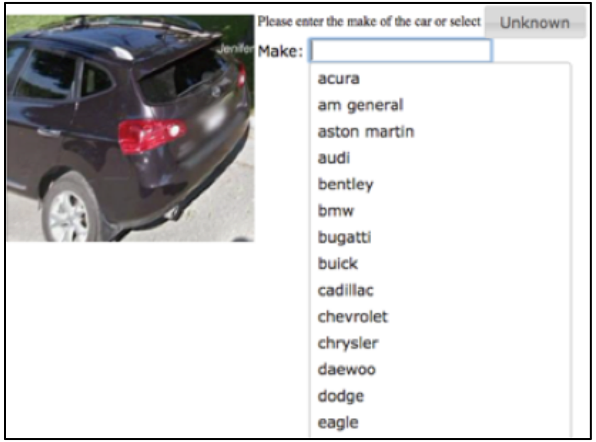
what we have

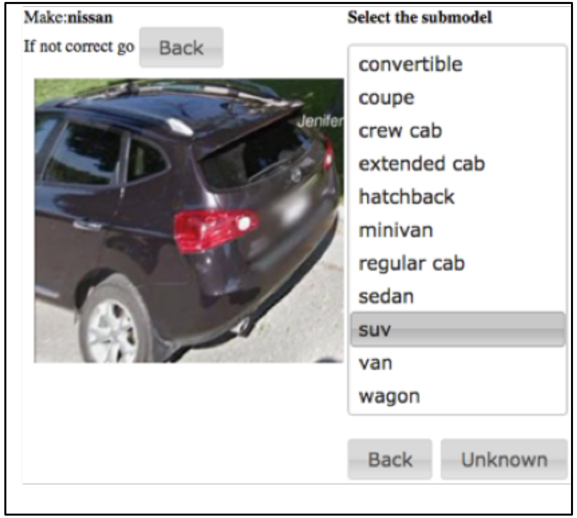
what we want to recognize

Solution: Experts

No way around this

Hierarchical annotation UI





Data Statistics

Attribute	Training	Validation	Test
Street View Images	199,666	39,933	159,732
Product Shot Images	313,099	-	-
Total Images	512,765	39,933	159,732
Street View Bounding Boxes	272,142	54,691	216,808
Product Shot Bounding Boxes	313,099	-	-
Total Bounding Boxes	585,241	54,691	216,808
Street View Category Labels	34,753	6,921	27,888
Product Shot Category Labels	313,099	-	-
Total Category Labels	347,852	6,921	27,888

Gebru et al. AAAI 2017

50 Million Images in 200 Cities

How do we efficiently detect cars?

Detection

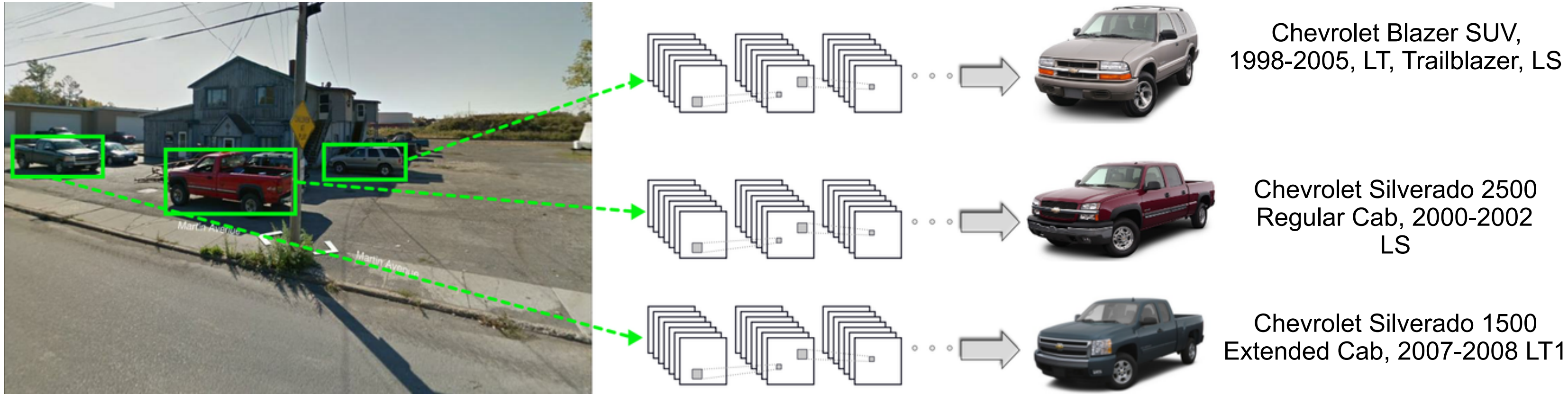
(in 2014)
State-of-the-art: R-CNN
→ 31 GPU years

Previous state-of-the-art: DPM
→ 2 weeks w/200 cpus

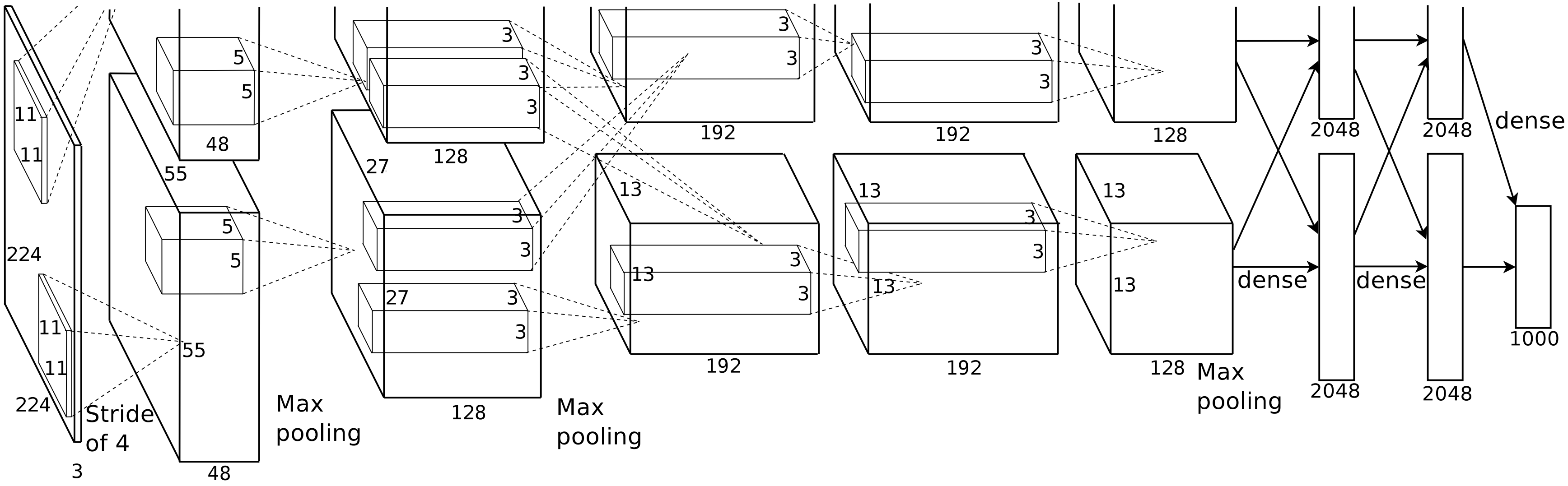
Classification

CNN (we use AlexNet)

- On detected box
- No fine-tuning needed



Classification (AlexNet)



Detected ~21.8M cars (8% of automobiles in the US)

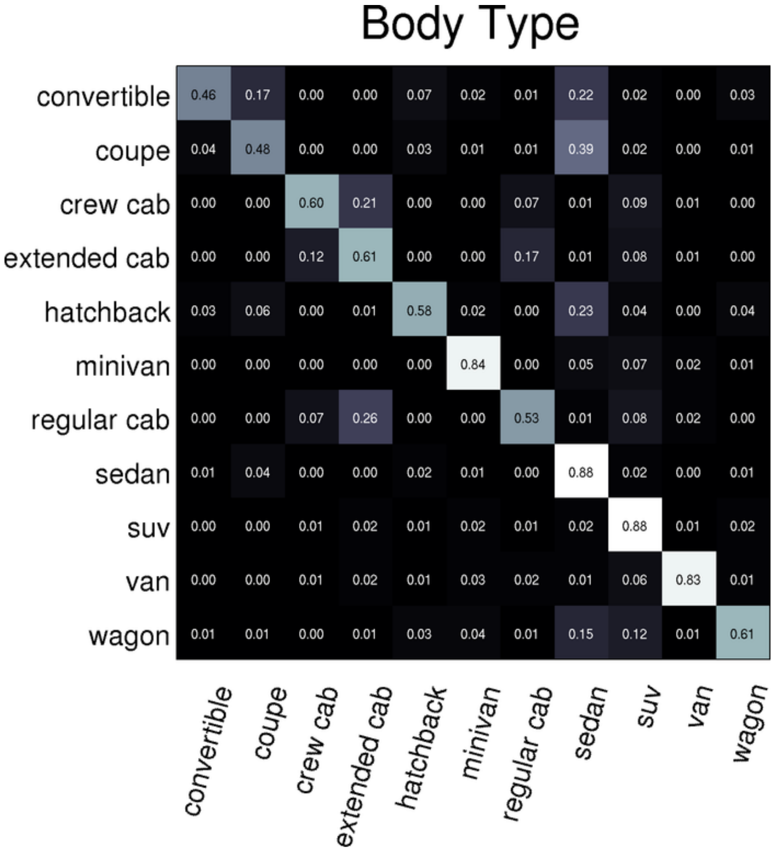
Vision Performance

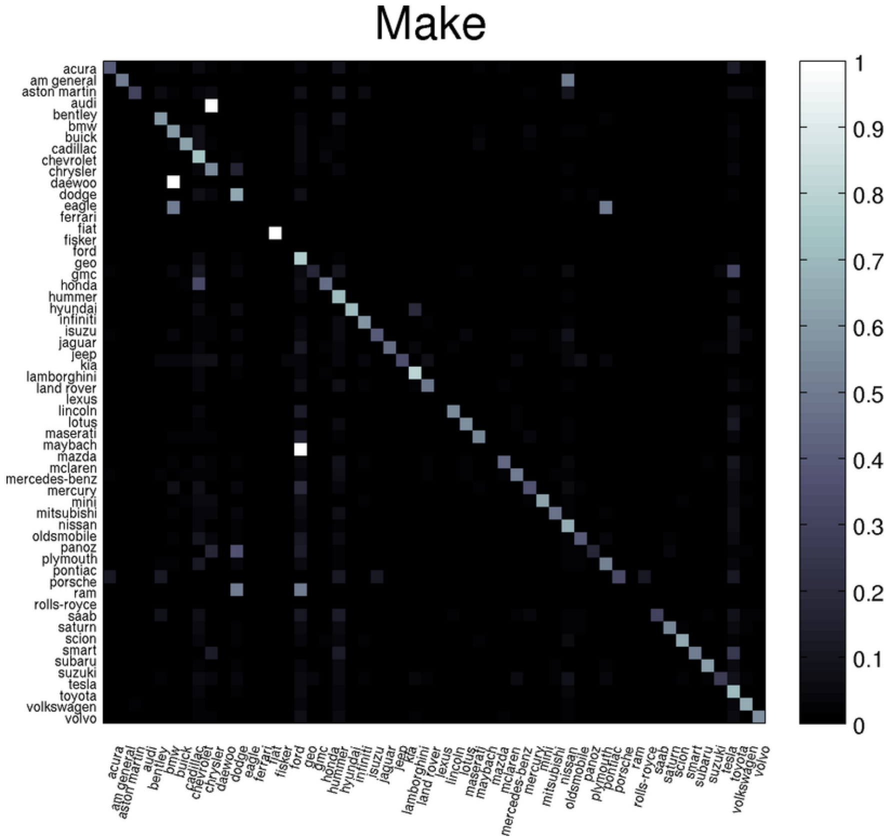
- Detection AP: 65.7
- 2,657-way accuracy: 31.20%
- Model accuracy: 52.27%
- Make accuracy: 66.86%
- Body Type: 78.02%
- Country: 84.35%

2000 Toyota 4Runner Limited

2005 Nissan Sentra 1.8 S

Mistakes are Reasonable





Gebru et al. AAAI 2017

Now let’s use computer vision to study the US

Question 1:

How green is each state?

Methodology

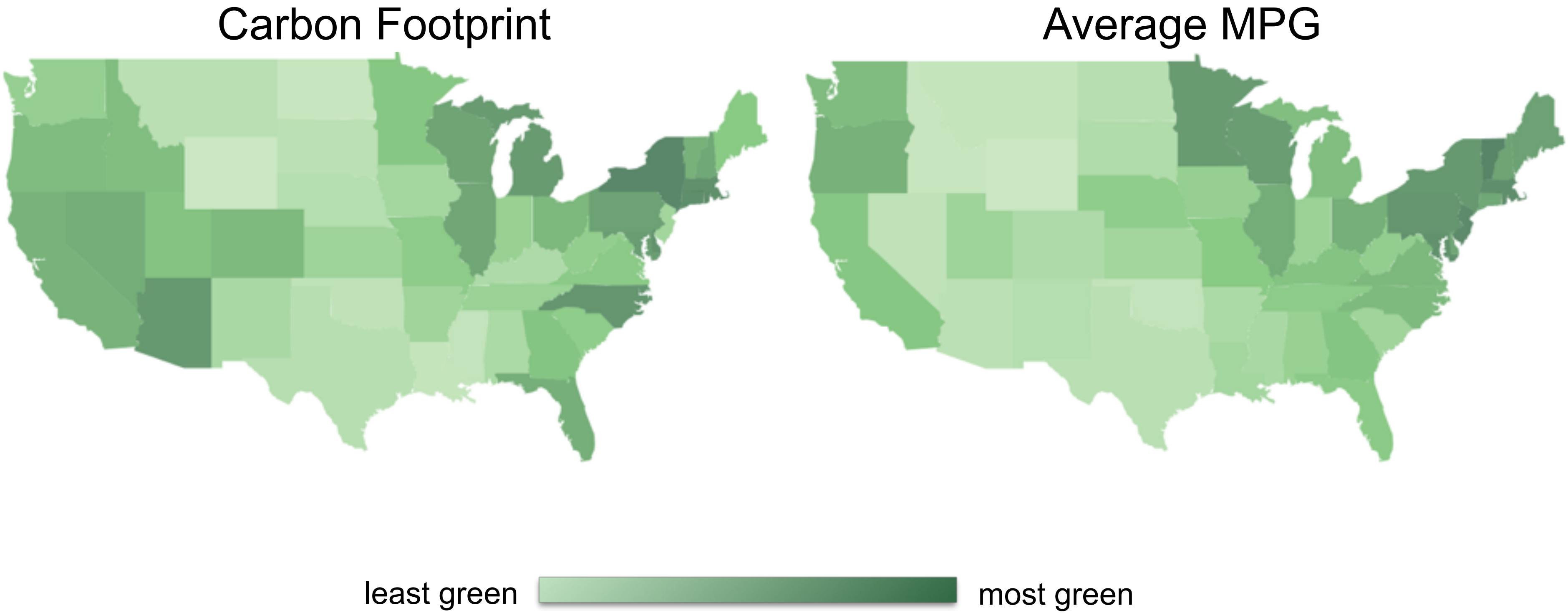
$p(\text{car}|I_i)$

$p(\text{class } c|I_i)$

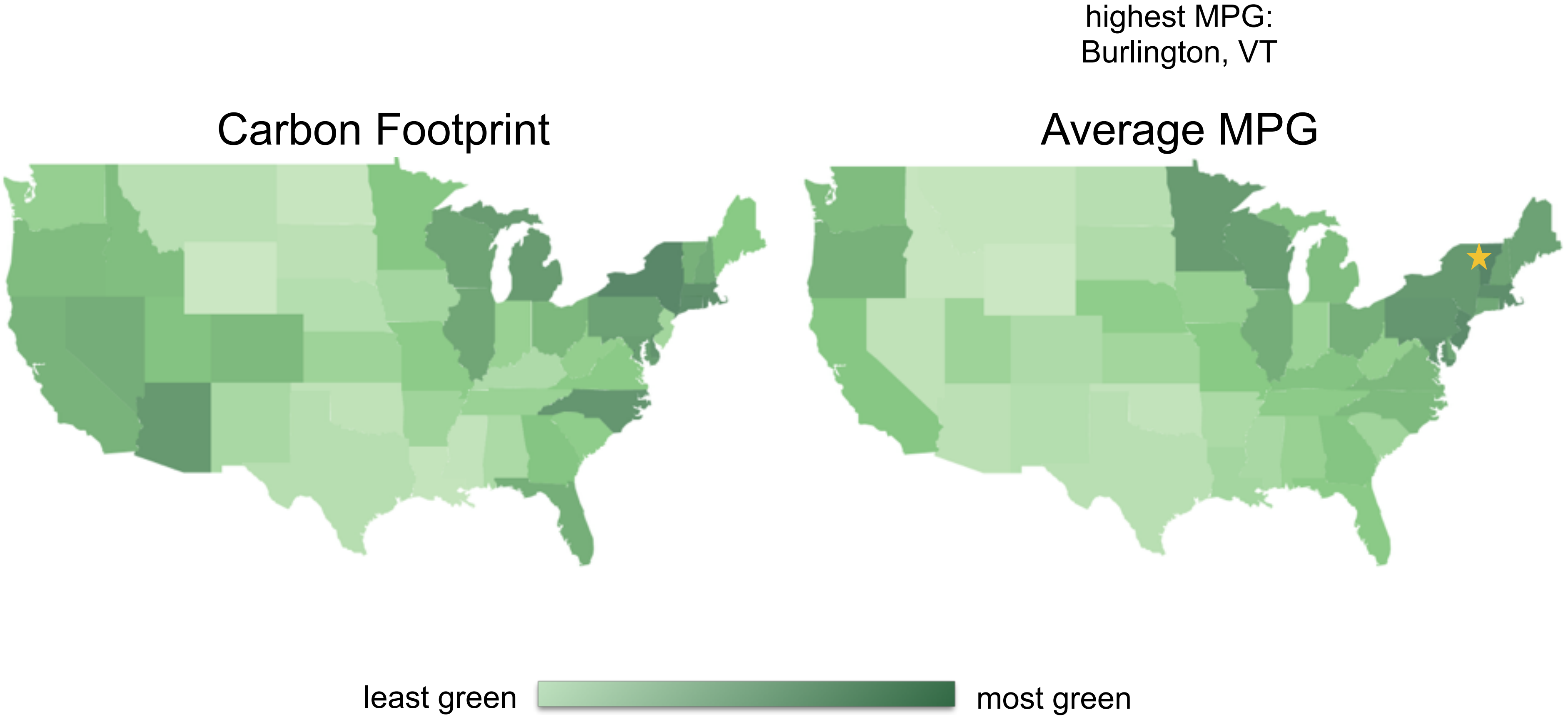
$\text{mpg}(c)$

Expected MPG in each state:

$$\mathbb{E}[\text{mpg}] = \sum_i p(\text{car}|I_i) \sum_c p(\text{class } c|I_i)\text{mpg}(c)$$



Gebru et al. AAAI 2017



Gebru et al. AAAI 2017

100% of power for Vermont city now renewable

Question 2:

Which cities are most segregated?

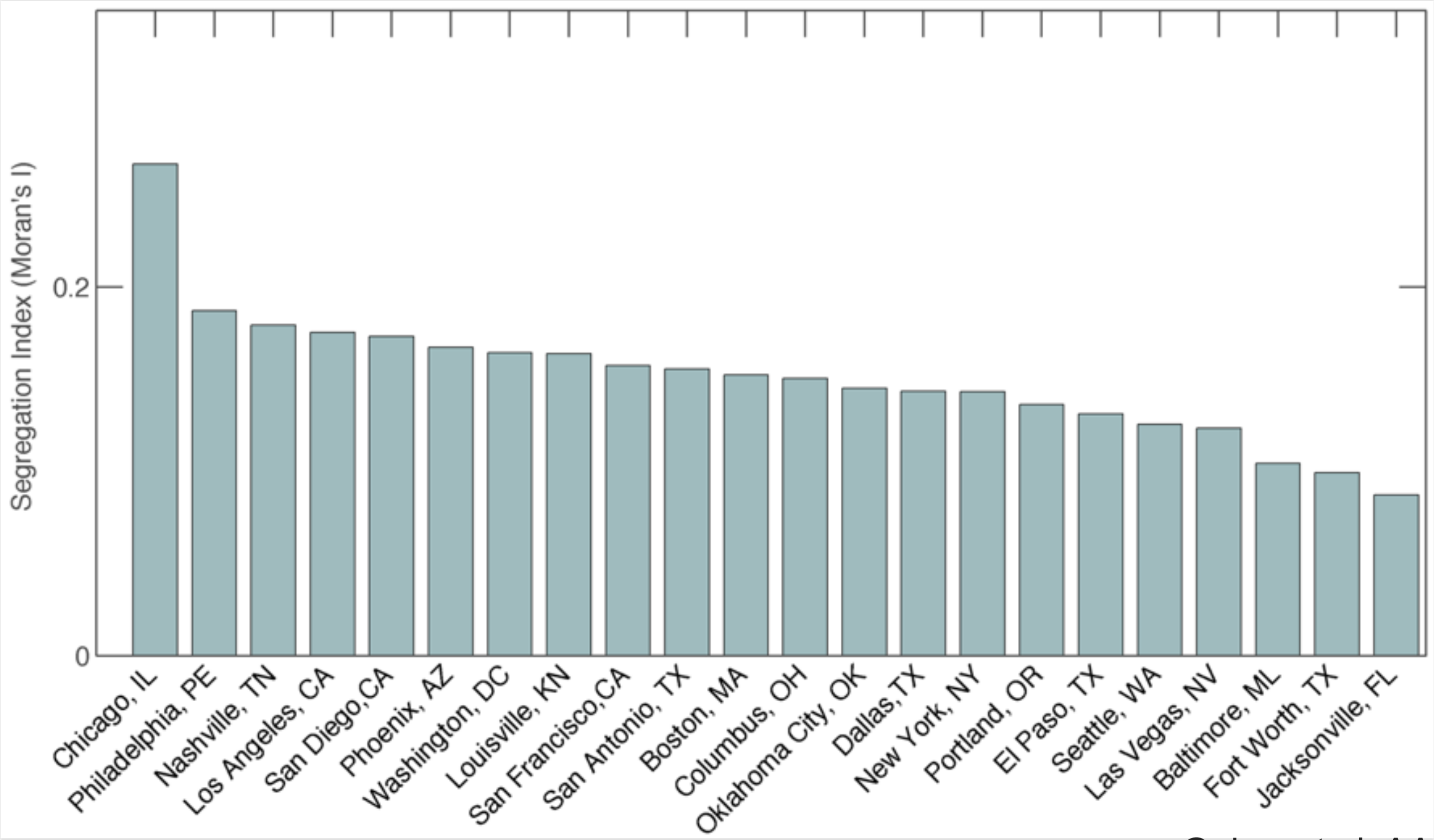
Methodology

- Average car price at each GPS point
- Measure segregation as Moran's I statistic

$$I = \frac{N \sum_{i,j} w_{i,j}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i,j} w_{i,j} \sum_k (x_k - \bar{x})^2}$$

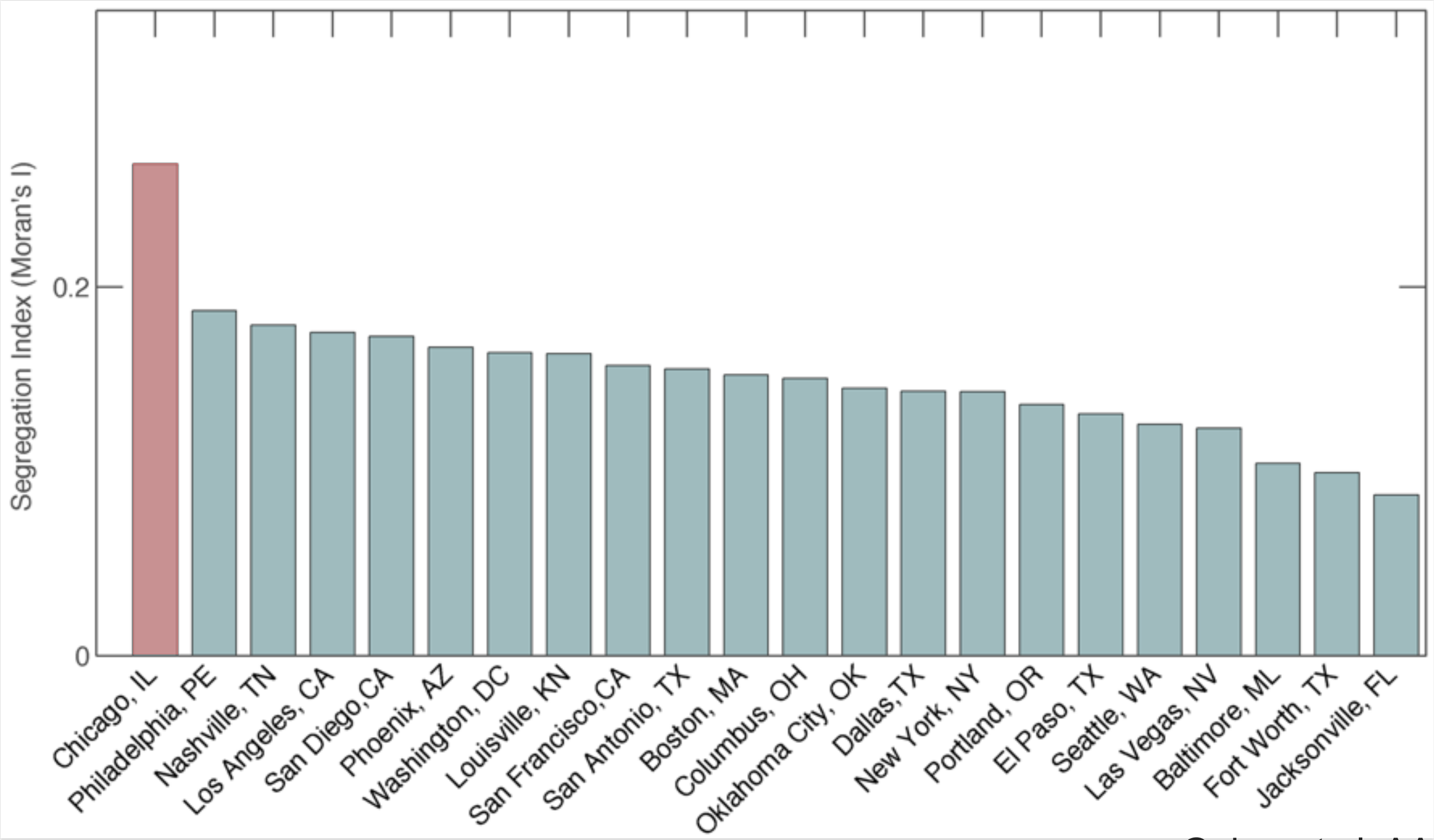
- Higher = more segregated

Results



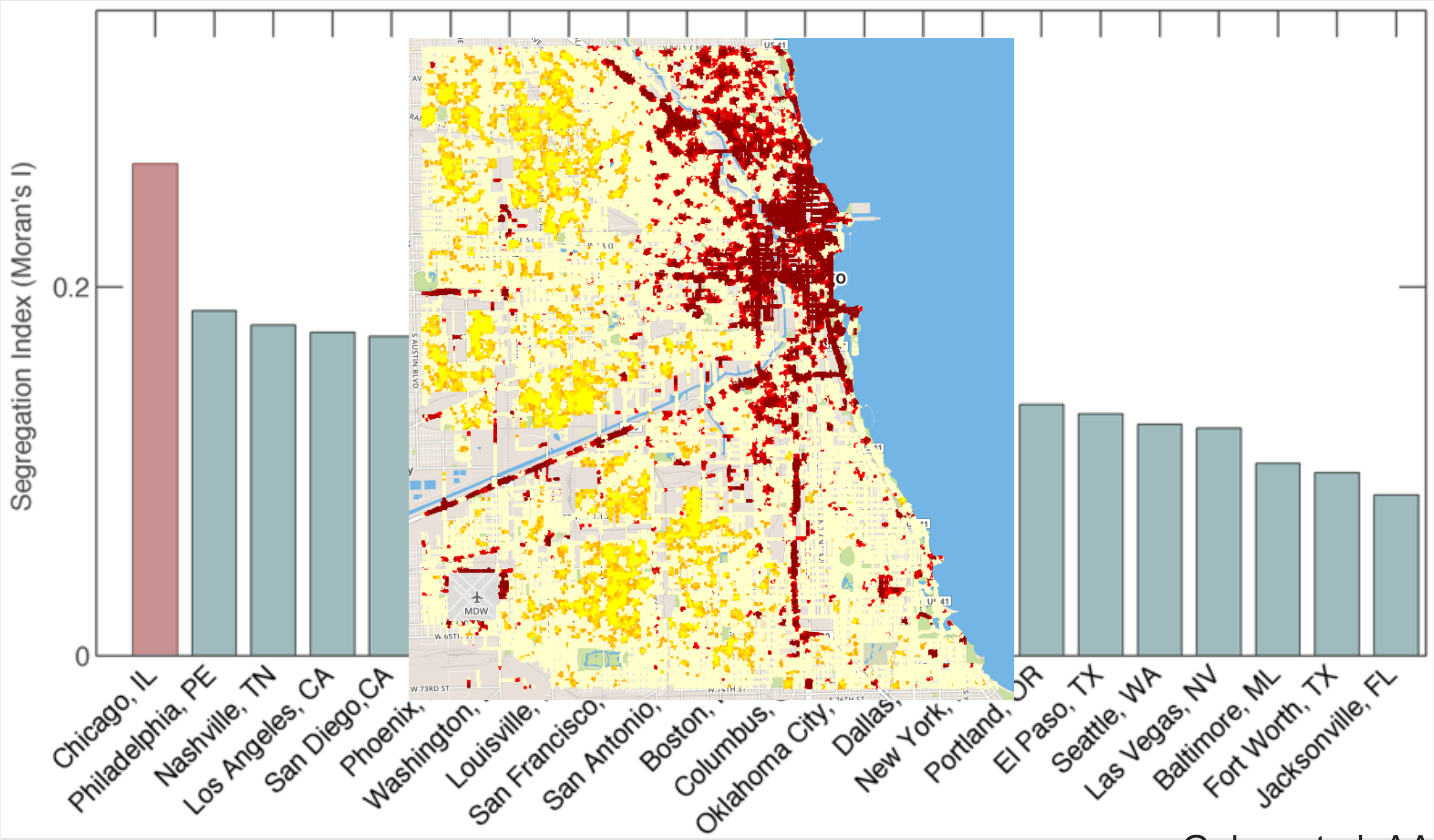
Gebru et al. AAAI 2017

Results



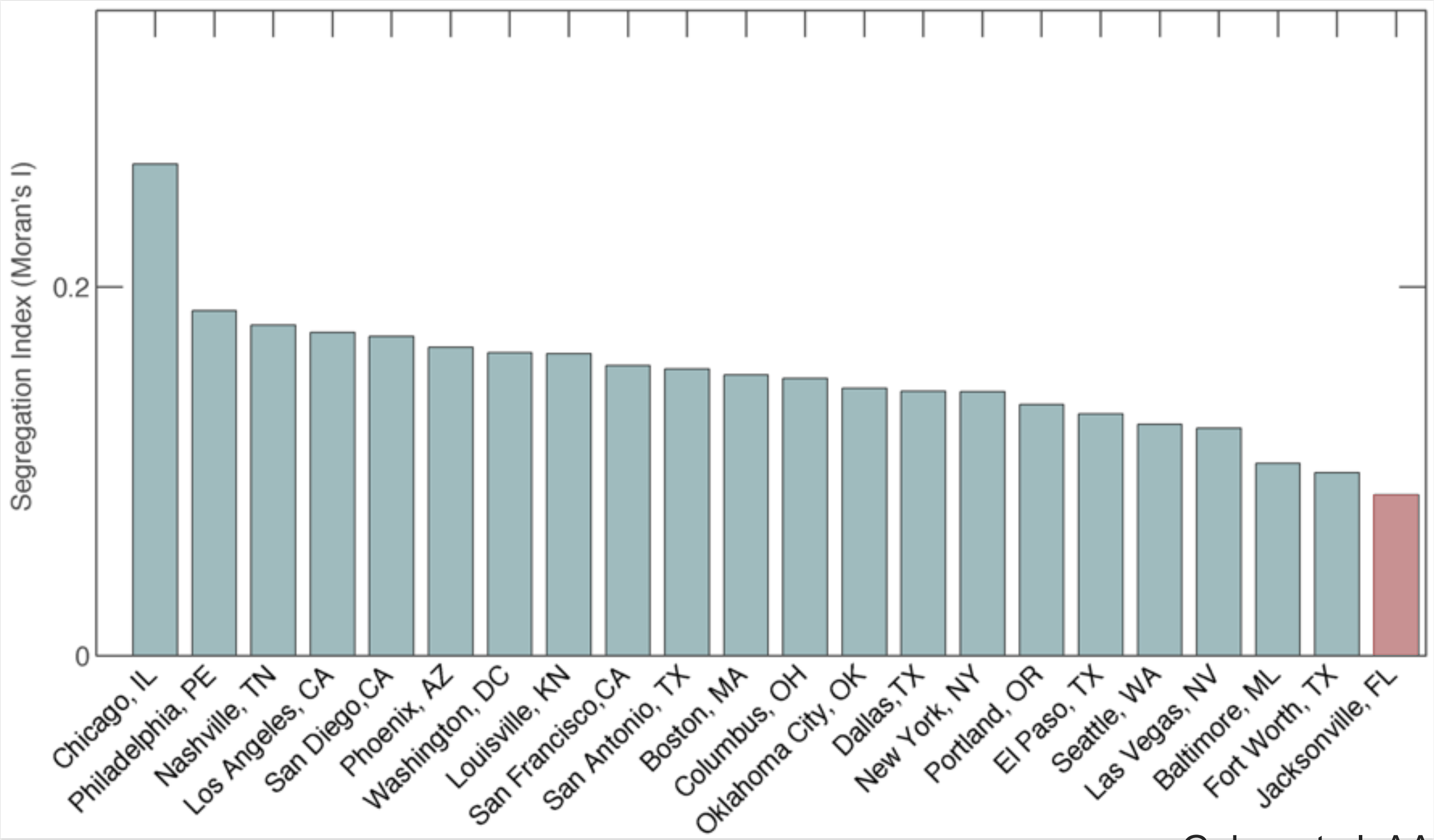
Gebru et al. AAAI 2017

Results



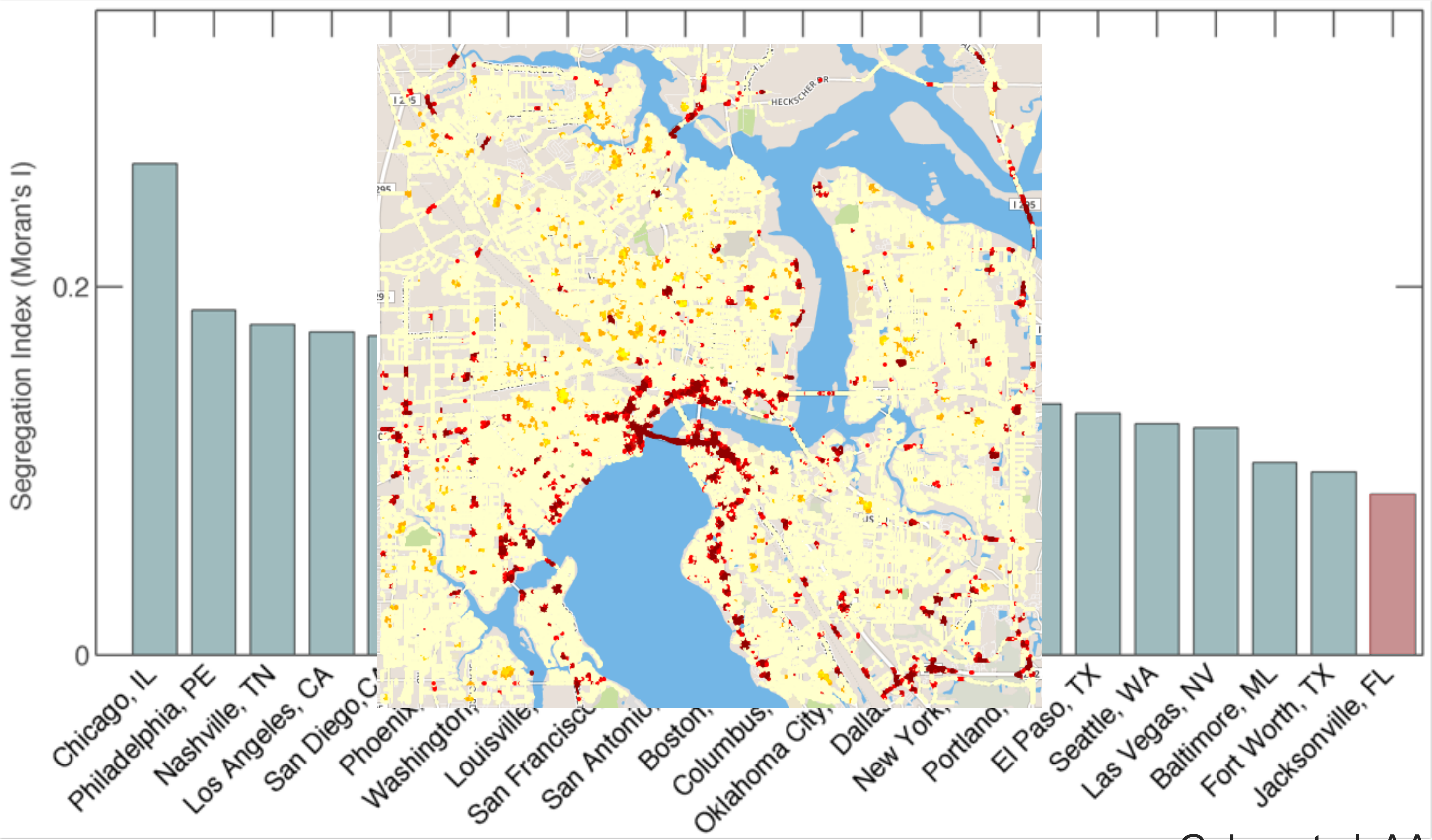
Gebru et al. AAAI 2017

Results



Gebru et al. AAAI 2017

Results



Gebru et al. AAAI 2017

Question 3:

Can we predict income?

Methodology

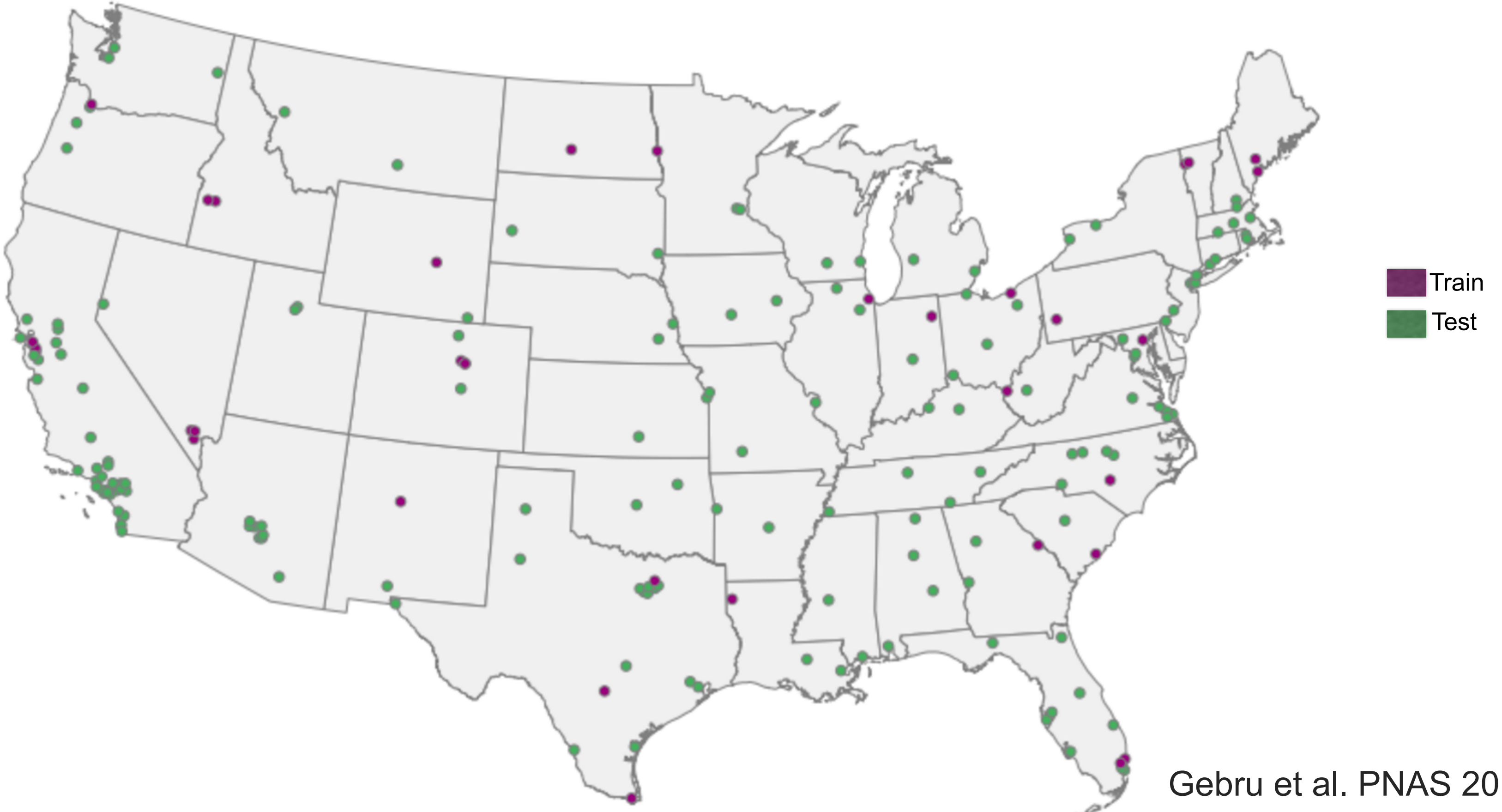
Aggregate 88 car attributes across zip codes

- #Cars/Image
- Average Car Price
- MPG (City/Highway)
- %Hybrid
- %Electric
- %Cars from each country
- %Foreign cars
- %Cars of each body type
- %Cars in year ranges (e.g. 1990-1994)
- %Cars of each make

Use ridge regression to predict income from attributes

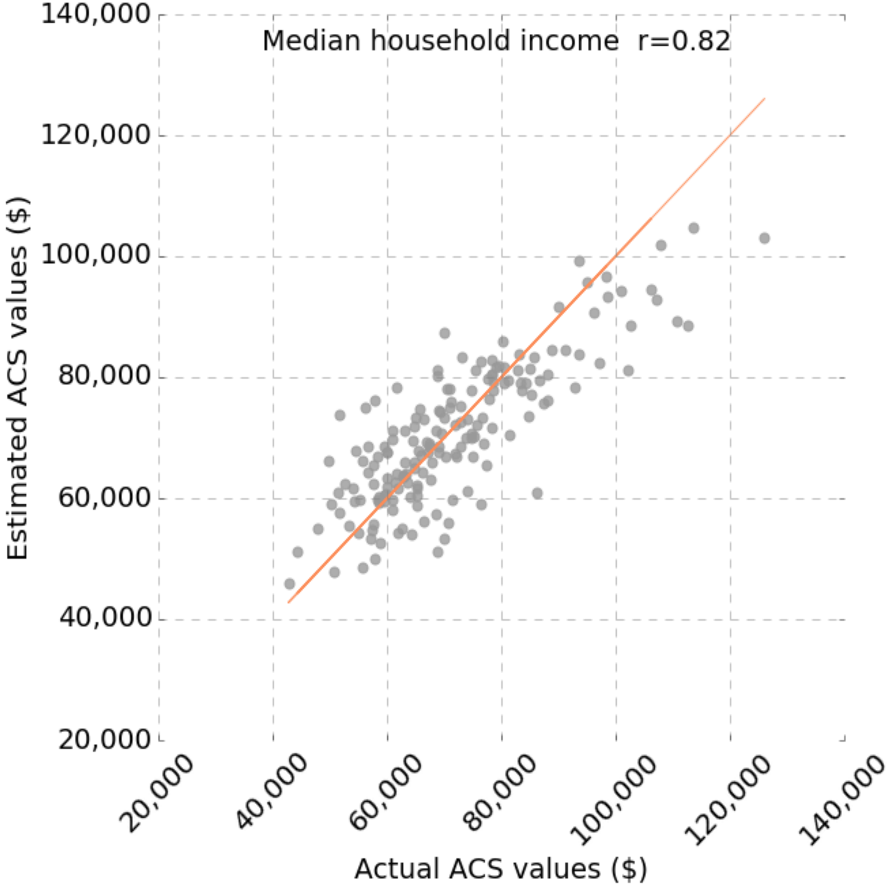
Gebru et al. PNAS 2017

Methodology



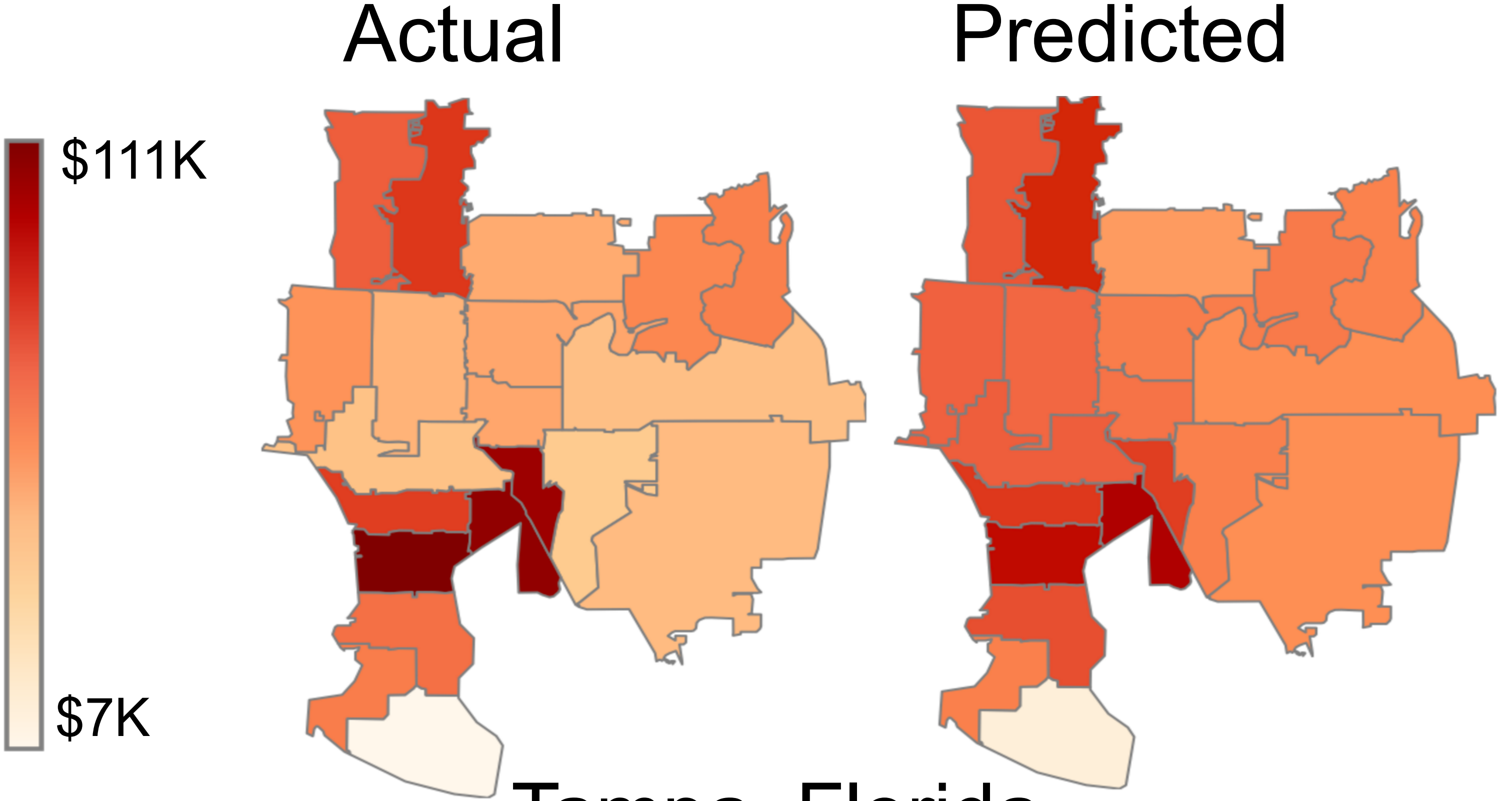
Gebru et al. PNAS 2017

Results



Gebru et al. PNAS 2017

Results



Tampa, Florida

Gebru et al. PNAS 2017

Bonus: What predicts income?

Most positively correlated:

1. %Foreign cars
2. %Japanese cars
3. Average car price
4. %Make: Lexus
5. %German cars

Most negatively correlated:

1. %American cars
2. %Year: 1995-1999
3. %Make: Buick
4. %Make: Oldsmobile
5. %Make: Dodge

Gebru et al. PNAS 2017

Question 4:

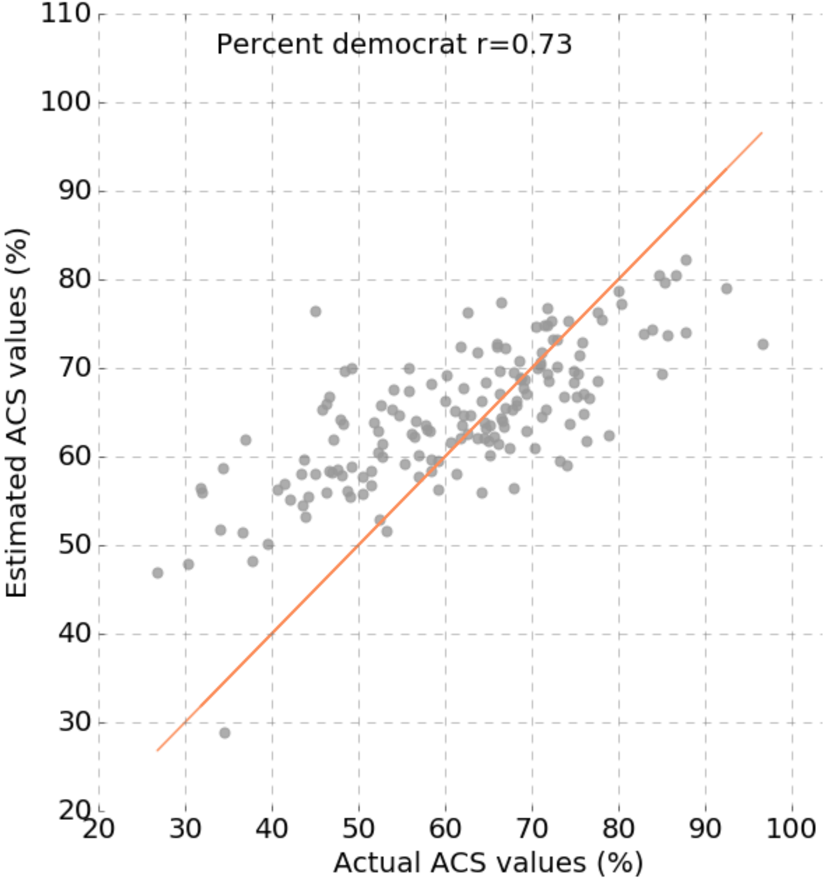
Can we predict voting patterns?

Methodology

Same as income, but precinct level

Data: 2008 U.S. Presidential Election

Results



Gebru et al. PNAS 2017

Results

Actual Percent of Voters for Obama in 2008

Predicted Percent of Voters for Obama in 2008

Gebru et al. PNAS 2017

Results

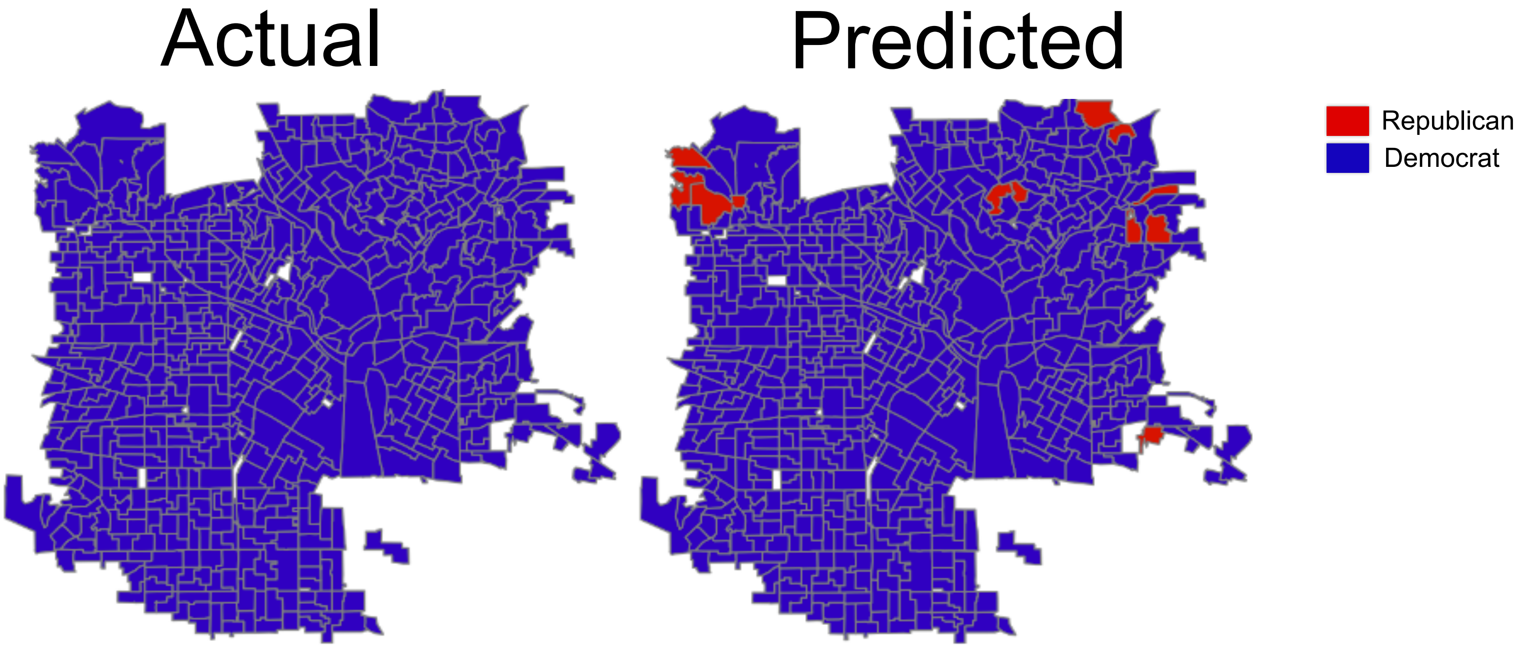
Actual Percent of Voters for Obama in 2008

Ratio of Sedans to Extended-Cab Trucks

Republican
Democrat

Gebru et al. PNAS 2017

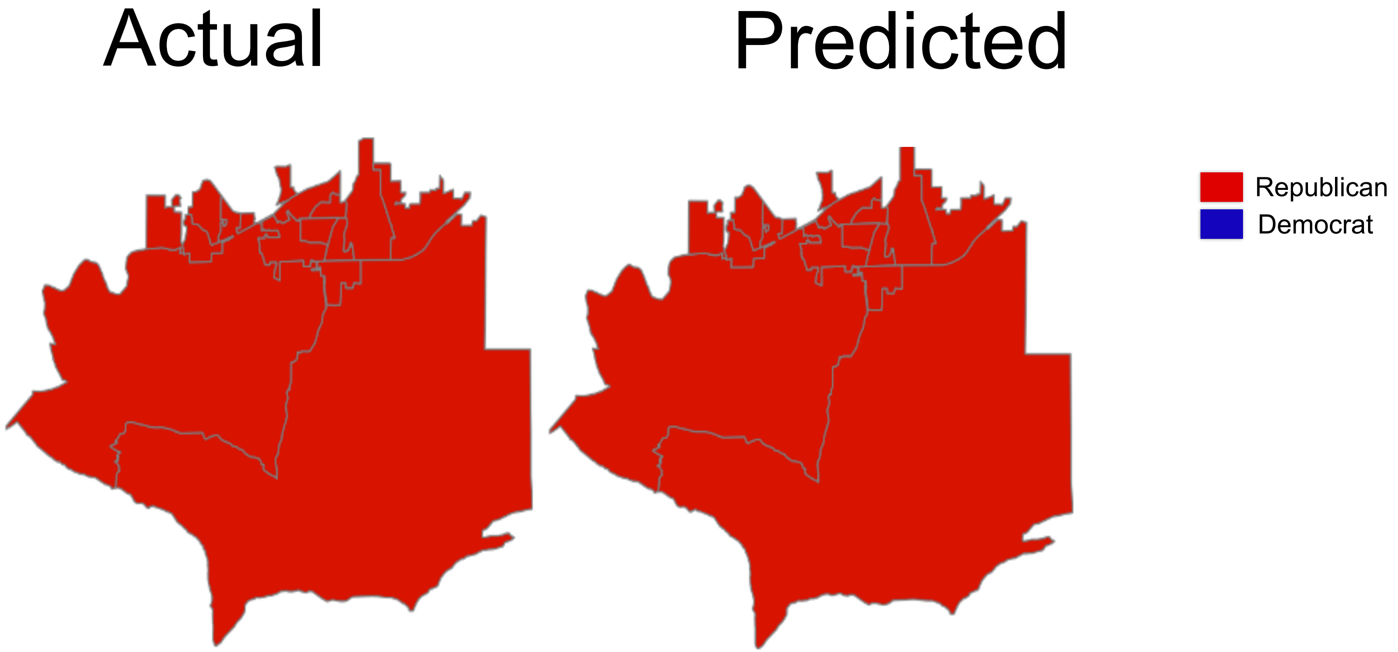
Results



Los Angeles, California

Gebru et al. PNAS 2017

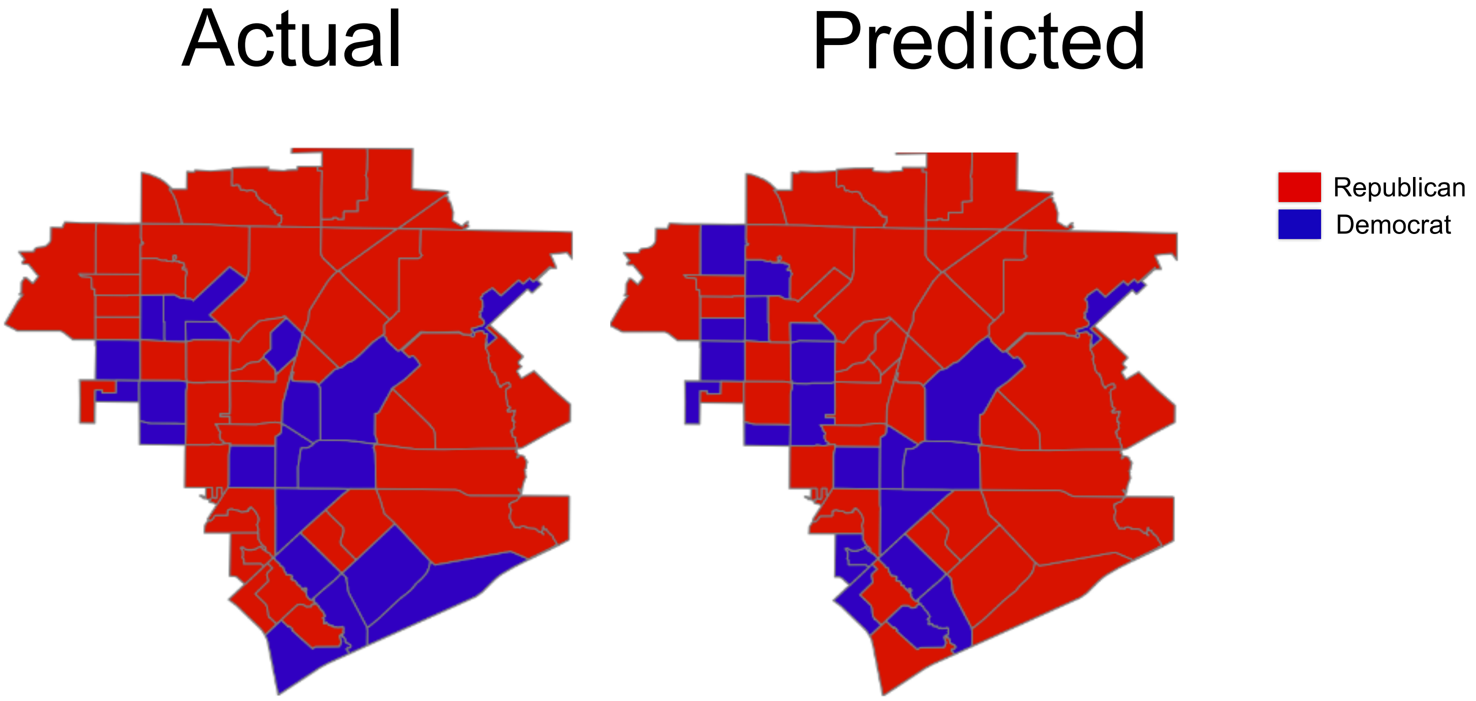
Results



Casper, Wyoming

Gebru et al. PNAS 2017

Results



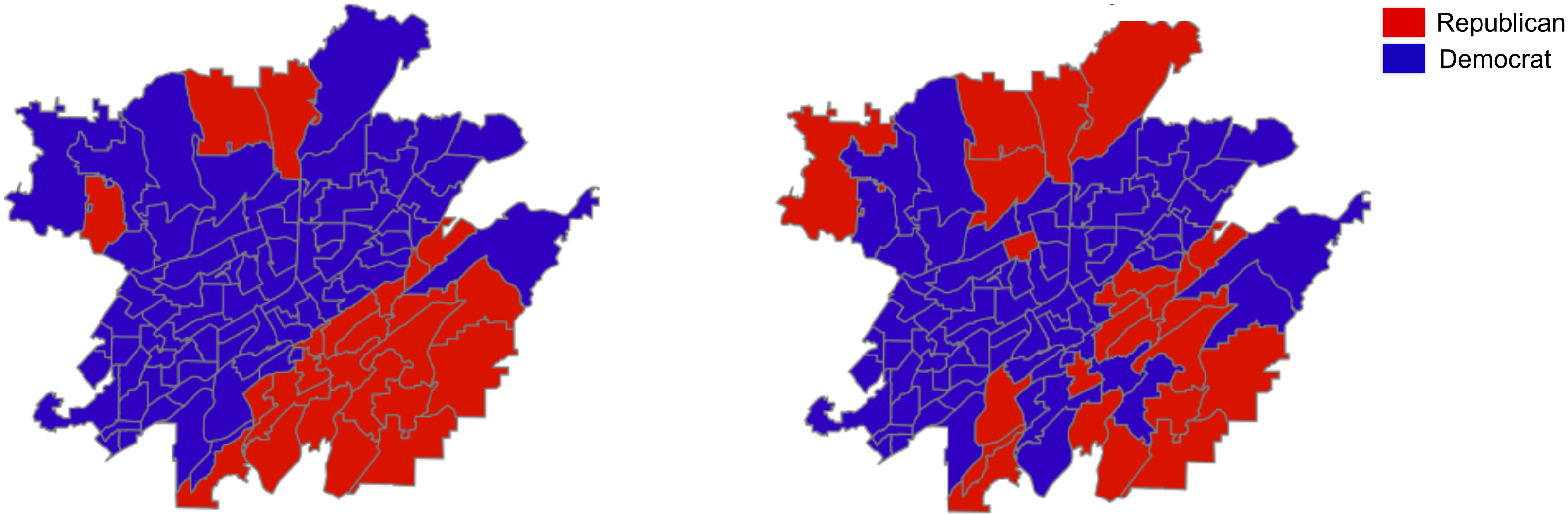
Garland, Texas

Gebru et al. PNAS 2017

Results

Actual

Predicted



Birmingham, Alabama

Gebru et al. PNAS 2017

What predicts %Obama?

Most positively correlated:

1. %Sedan
2. #Cars/Image
3. MPG (highway)
4. %Year: 1995-1999
5. MPG (city)

Most negatively correlated:

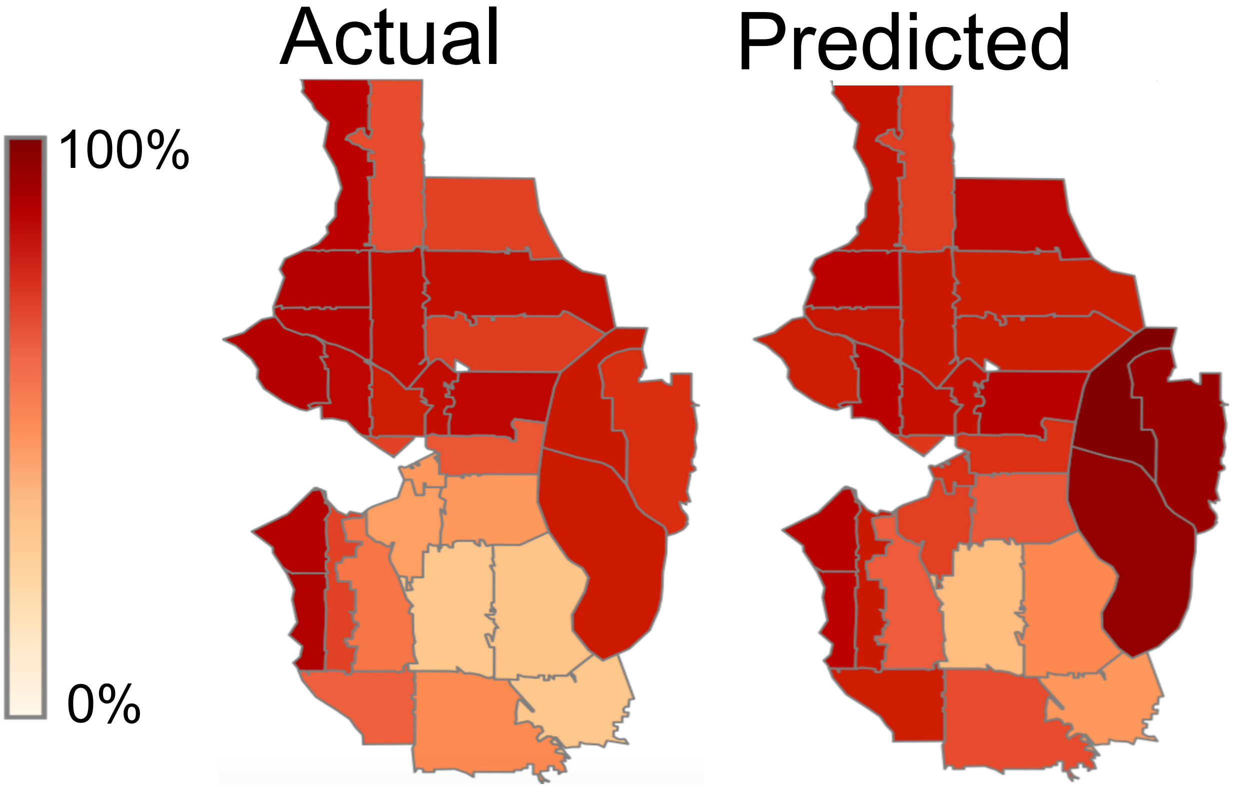
1. %Crew cab (truck)
2. %Extended cab (truck)
3. %Regular cab (truck)
4. Average Price
5. %SUV

Gebru et al. PNAS 2017

Question 5:

Can we predict race?

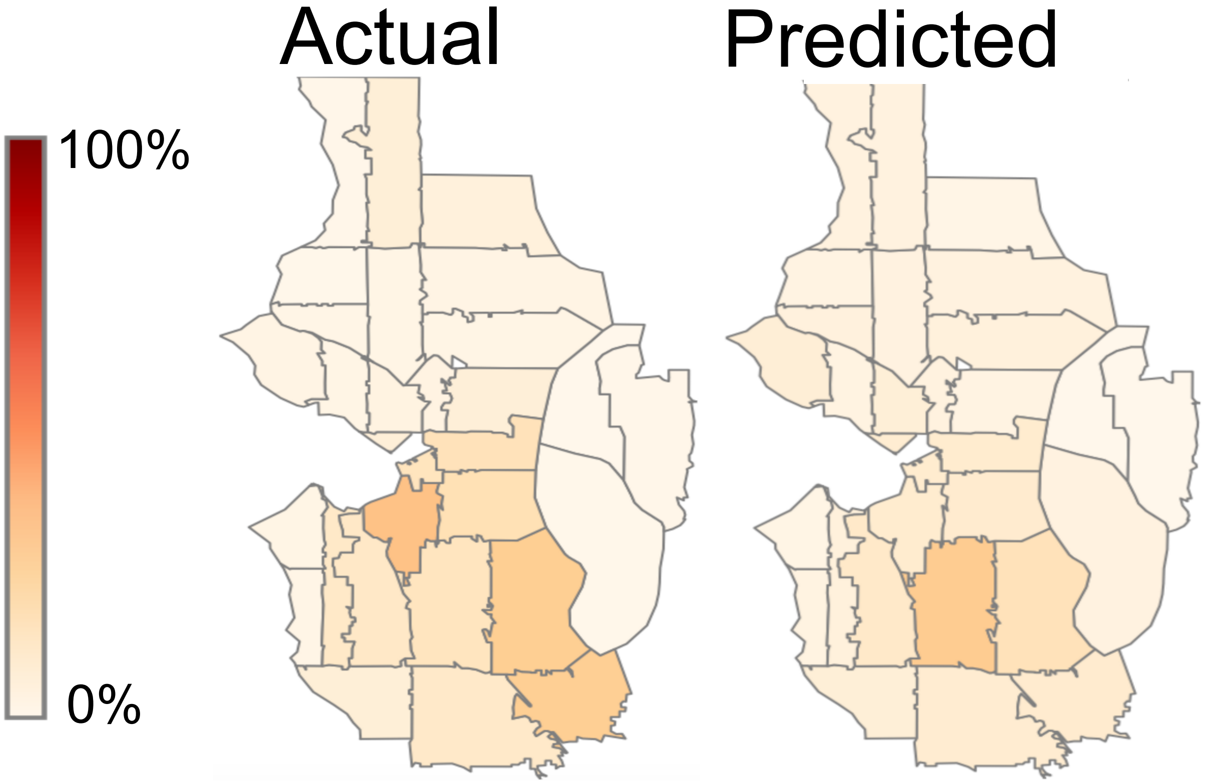
Results (%White)



Seattle, Washington

Gebru et al. PNAS 2017

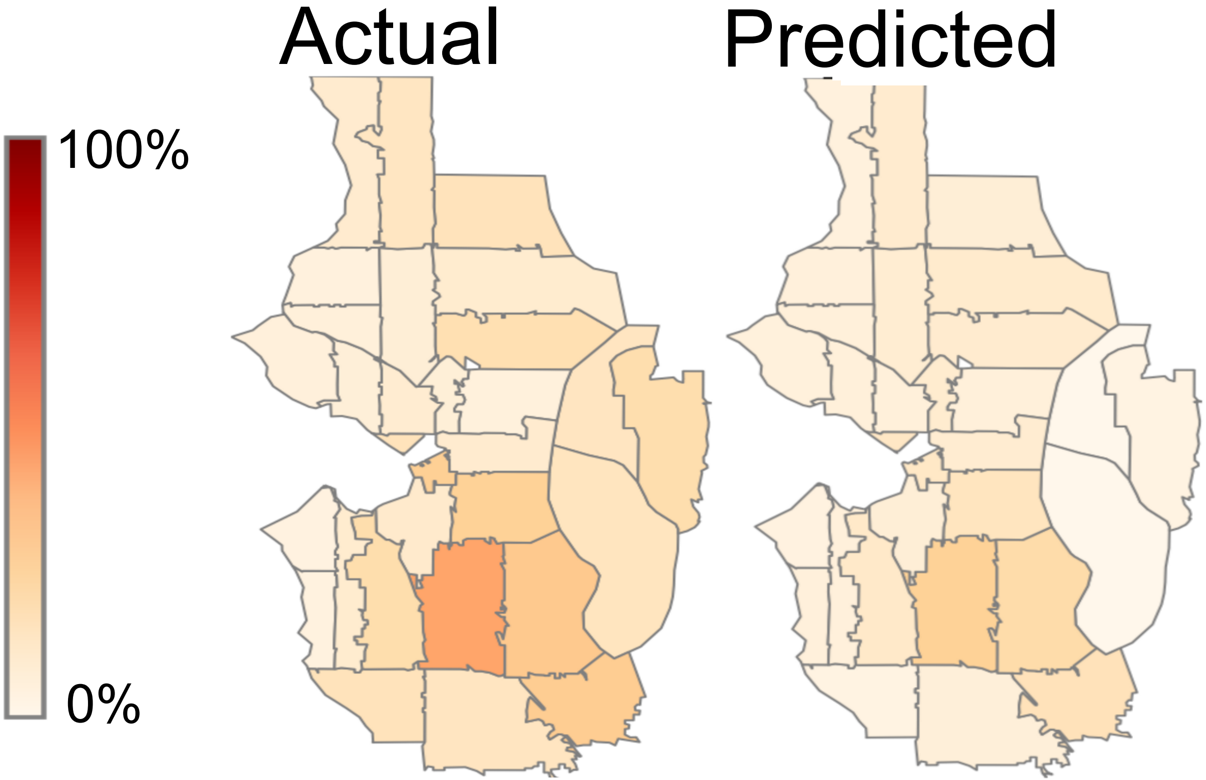
Results (%Black)



Seattle, Washington

Gebru et al. PNAS 2017

Results (%Asian)



Seattle, Washington

Gebru et al. PNAS 2017

What predicts %Asian?

Most positively correlated:

1. %Make: Toyota
2. %Japanese cars
3. %Make: Honda
4. %Foreign cars
5. %Make: Lexus

Most negatively correlated:

1. %American cars
2. %Make: Dodge
3. %Make: Chevrolet
4. %Make: Buick
5. %Make: Ford

Gebru et al. PNAS 2017

What predicts %Black?

Most positively correlated:

1. %Make: Cadillac
2. %Make: Buick
3. %Make: Mercury
4. %Sedan
5. %Make: Chrysler

Most negatively correlated:

1. %Foreign
2. %Extended cab (truck)
3. %Hatchback
4. %Japanese cars
5. %Make: Toyota

Gebru et al. PNAS 2017

What predicts %White?

Most positively correlated:

1. %SUV
2. %Make: Jeep
3. %Make: Subaru
4. Average Car Price
5. %Wagon

Most negatively correlated:

1. %Make: Mercury
2. %Year: 1995-1999
3. %Make: Lincoln
4. %Make: Cadillac
5. %Sedan

Gebru et al. PNAS 2017

What was the bottleneck?

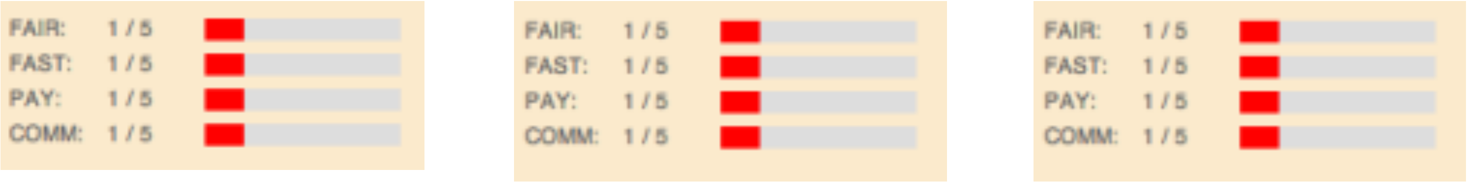
- Using images, we analyzed the relationship between cars and people.

In reality….

These tasks are really hard

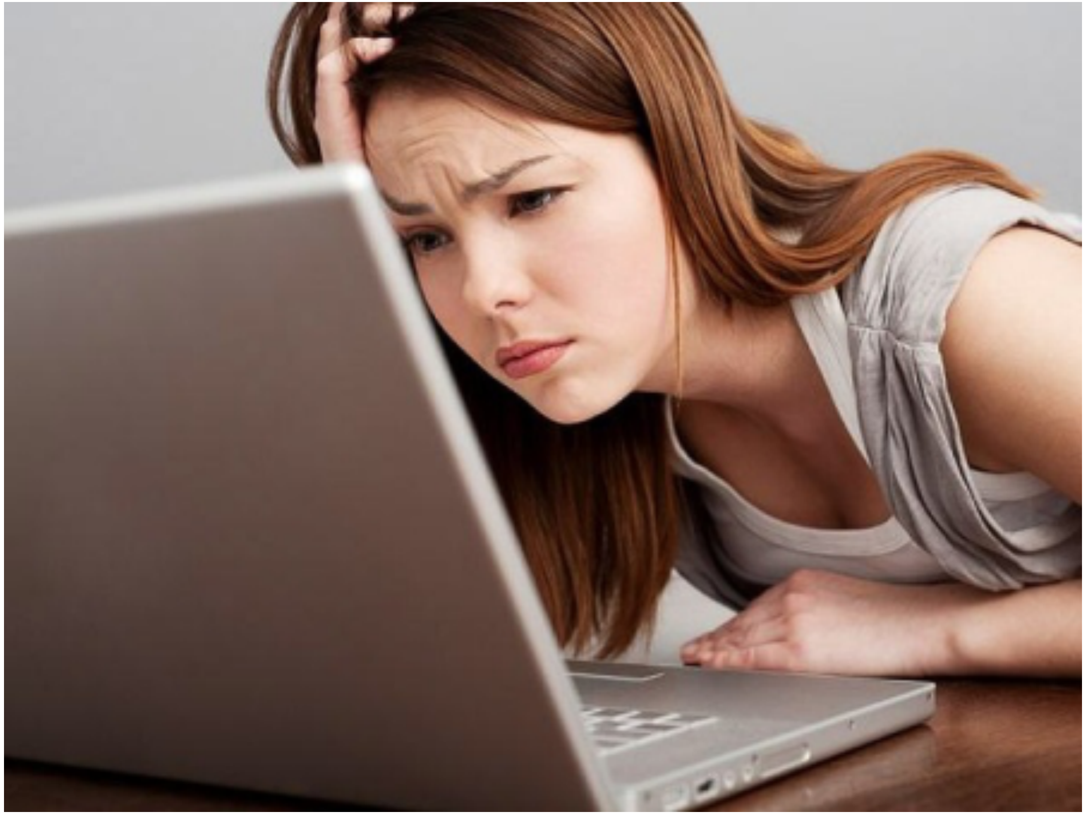
- We spent over 1 year and $35k on the image gathering and annotation process

We got terrible reviews on AMT



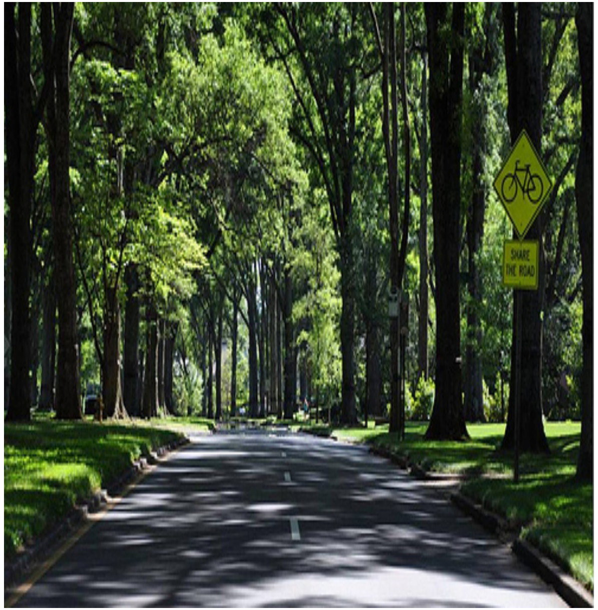
“apparently the virtually indistinguishable cars are super distinguishable to them”

“I paid very close attention to the pictures and let the requester know that I feel these rejects were unfair. Will update when I get communication back.”



Timnit-sourcing

What was the bottleneck?



What was the bottleneck?

What was the bottleneck?

- In each of these scenarios, we would have to annotate Google Street View images with the appropriate objects

What was the bottleneck?

- We would have to hire experts to annotate the data and spend lots of $$$ and time

Predicting things from data….

GOOGLE MAPS

Tech Policy / Privacy

How a Google Street View image of your house predicts your risk of a car accident

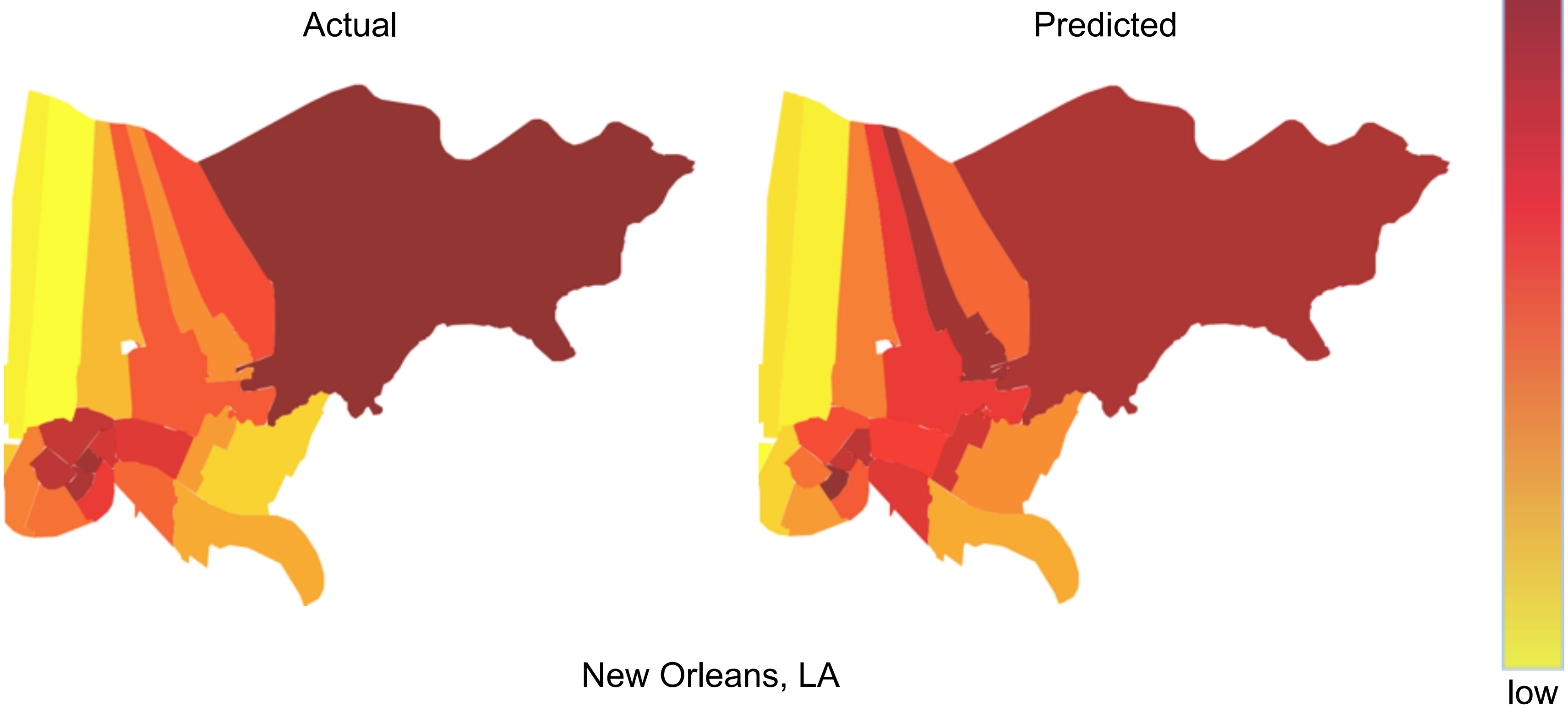
"In 2017 a team of researchers used the images to study the distribution of car types in the US and then used that data to determine the demographic makeup of the country. ***It turns out that the car you drive is a surprisingly reliable proxy for your income level, your education, your occupation, and even the way you vote in elections. "***

"Now a different group has gone even further….have used Street View images of people's houses to determine how likely they are to be involved in a car accident. ***That is valuable information that an insurance company could use to set premiums****"*

Do we want this to be used by insurance companies?

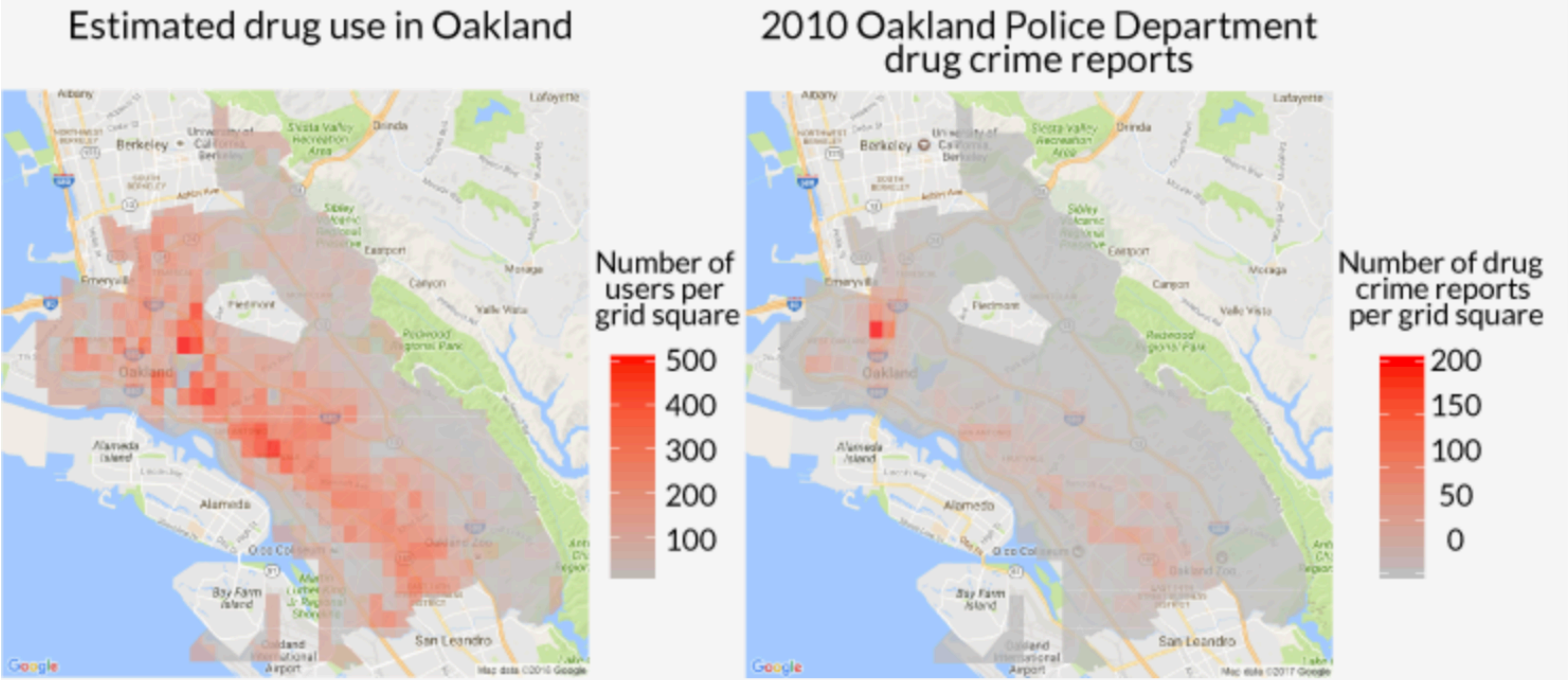
The ground truth is biased

Results (Crimes against People)



New Orleans, LA

Drug use in Oakland is probably fairly widespread (left) based on estimates derived in part from the 2011 National Survey on Drug Use and Health. But police records of drug reports and crimes are concentrated in areas that are largely nonwhite and low-income (right).



K. LUM/HRDAG

HireVue

PRODUCTS + WHY HIREVUE + CUSTOMERS + RESOURCES + COMPANY + SEE A DEMO LOGIN

Extreme Vetting Initiative Objectives (cont.)

Performance Objectives of the Overarching Vetting Contract:

1. Centralizes screening and vetting processes to mitigate case backlog and provide law enforcement and field agents with timely, actionable information;

2. Allows ICE to develop richer case files that provide more value-added information to further investigations or support prosecutions in immigration or federal courts;

3. Allows ICE to perform regular, periodic and/or continuous review and vetting of nonimmigrants for changes in their risk profile after they enter the United States and;

4. Automates at no loss of data quality or veracity any manually-intensive vetting and screening processes that inhibit ICE from properly and thoroughly vetting individuals in a timely fashion.

Home > Israel News

Israel Arrests Palestinian Because Facebook Translated 'Good Morning' to 'Attack Them'

No Arabic-speaking police officer read the post before arresting the man, who works at a construction site in a West Bank settlement

US ADULTS INDEXED

130 MILLION

One in two American adults is in a law enforcement face recognition network used in unregulated searches employing algorithms with unaudited accuracy.

The Perpetual Line Up

(Garvie , Bedoya, Frankle 2016)

Error Rate$_{(1\text{-}PPV)}$ By Female x Skin Type

	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TYPE VI
	1.7%	1.1%	3.3%	0%	23.2%	25.0%
FACE++	11.9%	9.7%	8.2%	13.9%	32.4%	46.5%
IBM	5.1%	7.4%	8.2%	8.3%	33.3%	46.8%

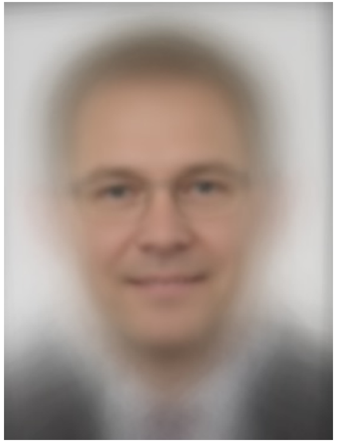
Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

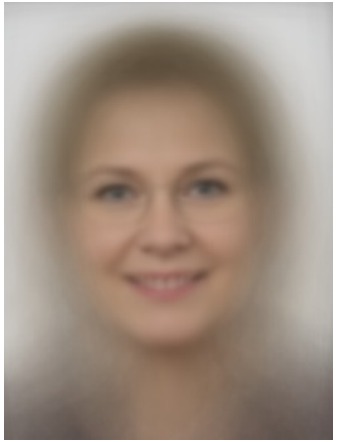
Existing Benchmarks

IJB-A *ADIENCE*

75.4% *Male*

79.6% *Lighter*



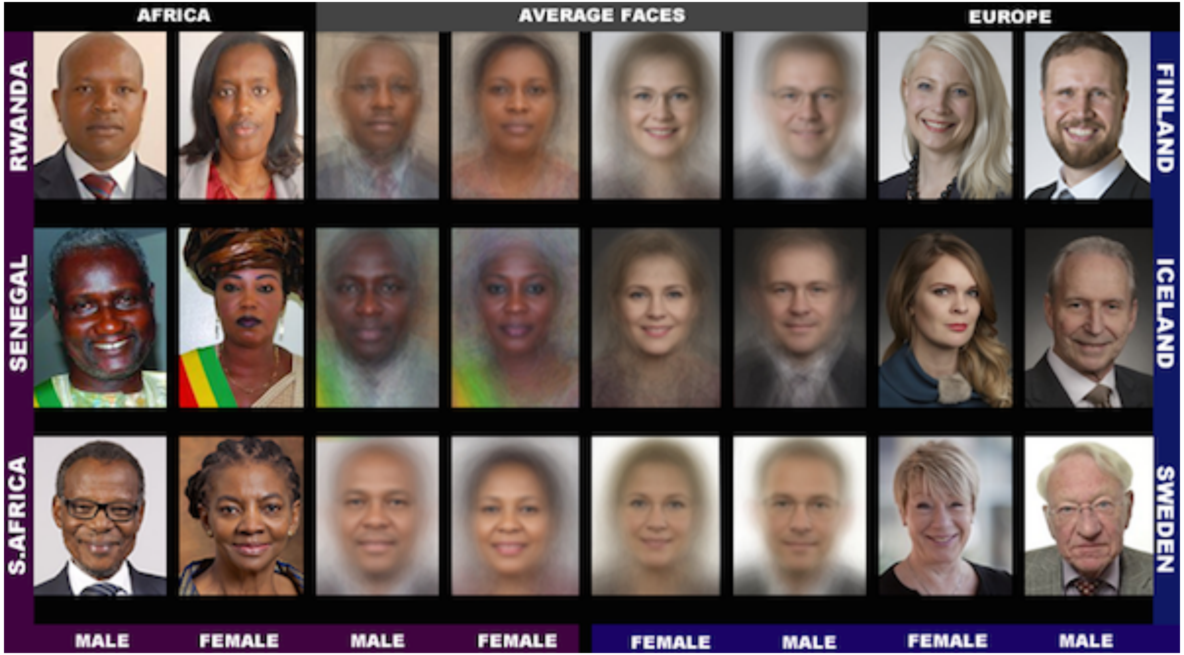


52% *Female*

86.2% *Lighter*

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Pilot Parliaments Benchmark (PPB)



1270 Faces
6 Countries

54.4% Male
53.6% Lighter

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Demographic and Phenotypic Labels

- ***BINARY GENDER LABELS (M/F)***
- ***FITZPATRICK SKIN TYPE LABELS***

SKIN TYPE	one 	two 	three 	four 	five 	six 
Hair	red, blonde	blonde, red, light brown	chestnut, dark blonde	brown, medium brown, dark brown	dark brown	black
Eyes	blue, grey, green	blue, grey, green, hazel	brown, blue, grey, green, hazel	hazel, brown	brown	brown
Skin	very pale white, pale white	pale white	white, light brown	medium brown, dark brown	dark brown	black
Tanning Ability	burns very easily, never tans	burns easily, rarely tans	sometimes burns, gradually tans	hardly ever burn, tans very easily	Rarely burns, tans easily and quickly darkens	Never burns, tans very dark

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Overall Accuracy(PPV) ON PPB

FACE++

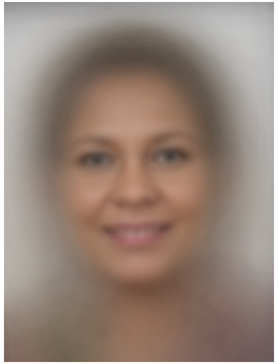
IBM

93.7% | 90% | 87.9%

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Accuracy by Gender

	FEMALE FACES	MALE FACES
	89.3%	97.4%
FACE++	78.7%	99.3%
IBM	79.7%	94.4%



FEMALE

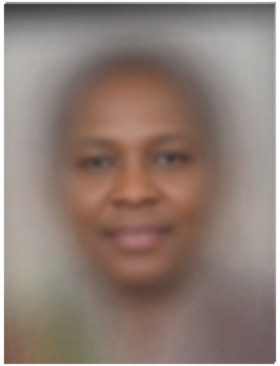
8-21% ERROR GAP

MALE

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

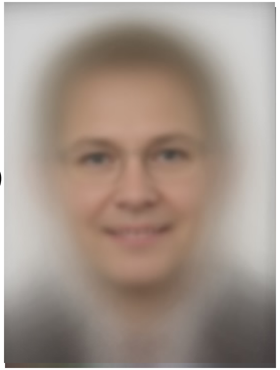
Accuracy by Skin Type

	DARKER FACES	LIGHTER FACES
	87.1%	99.3%
FACE++	83.5%	95.3%
IBM	77.6%	96.8%



DARKER

12-19% ERROR GAP

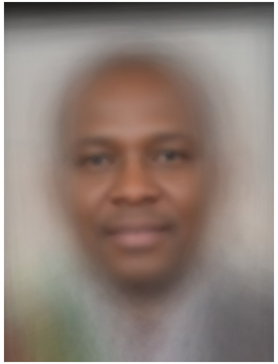


LIGHTER

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

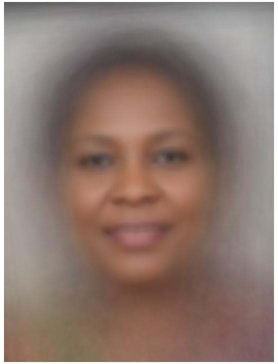
Intersectional Accuracy - MSFT

94%



DARKER MALES

79.2%

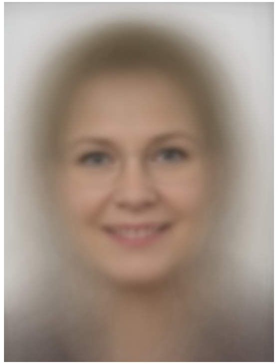


DARKER FEMALES

100%

LIGHTER MALES

98.3%



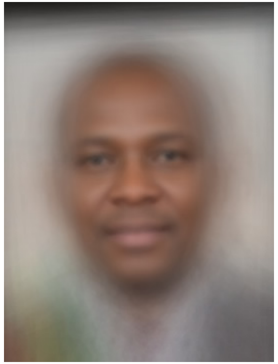
LIGHTER FEMALES

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Intersectional Accuracy - FACE++

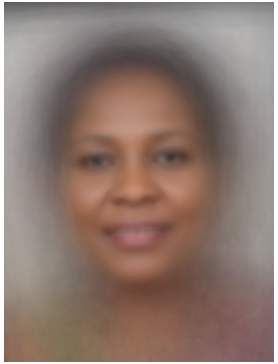
FACE++

99.3%



DARKER MALES

65.5%

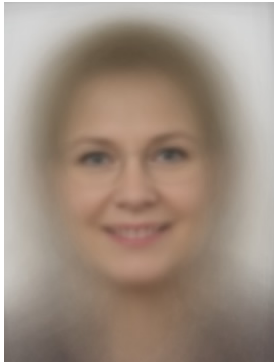


DARKER FEMALES

99.2%

LIGHTER MALES

98.3%

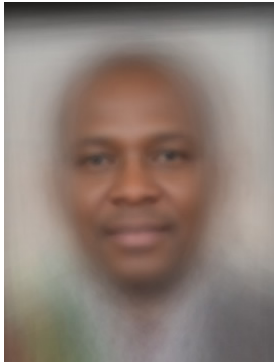
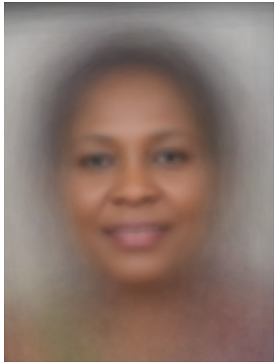
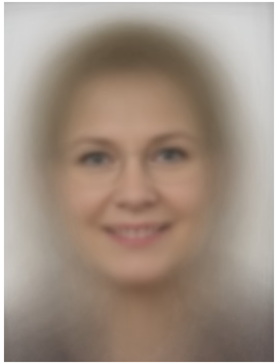


LIGHTER FEMALES

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Intersectional Accuracy - IBM

IBM

88%	65.3%	99.7%	92.9%
			
DARKER MALES	**DARKER FEMALES**	**LIGHTER MALES**	**LIGHTER FEMALES**

Microsoft improves facial recognition technology to perform well across all skin tones, genders

June 26, 2018 | John Roach

DIGITAL

Amazon Rekognition May Finally Be Audited and Ranked Alongside Other Vendors

A more universal test for facial recognition systems is needed

By Lisa Lacy | February 19, 2019

IBM

IBM Research Blog Topics Labs About

AI

IBM Research Releases 'Diversity in Faces' Dataset to Advance Study of Fairness in Facial Recognition Systems

NEWS MANAGEMENT OVERSIGHT DEFENSE TECH CONTRACTING PAY & B

Senators Are Asking Whether Artificial Intelligence Could Violate U.S. Civil Rights Laws

By Dave Gershgorn | Quartz | September 21, 2018 | 5 Comments

Senators Kamala Harris and Cory Booker both signed letters to federal agencies asking about AI bias. J. Scott Applewhite/AP

RELATED

Shutdown Roundup: NTSB Isn't Investigating Accidents, Warren Warns of Decreased Financial Investigations and More

January 25, 2019 | 1 Comment

Democratic Senator Questions Mulvaney's Hatch Act Compliance

September 21, 2018 | 32 Comments

Why Philadelphia Is on the Federal Government's

We can’t ignore social & structural problems

US ADULTS INDEXED

130 MILLION

One in two American adults is in a law enforcement face recognition network used in unregulated searches employing algorithms with unaudited accuracy.

The Perpetual Line Up

(Garvie , Bedoya, Frankle 2016)

Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.

Amazon Is Pushing Facial Technology That a Study Says Could Be Biased

In new tests, Amazon's system had more difficulty identifying the gender of female and darker-skinned faces than similar services from IBM and Microsoft.

Black in AI

Donate

Black in AI (BAI)

Black in AI is a place for sharing ideas, fostering collaborations and discussing initiatives to increase the presence of Black people in the field of Artificial Intelligence. If you are in the field of AI and self-identify as Black, please fill out this Google Form to request to join and we will add you to various platforms that we maintain. We also welcome allies to join our group using the Google form. Allies will be added to our email lists, where we send out group updates and requests for assistance.

Like our Facebook Page and follow us on Twitter to learn about our members and various activities!

Latest Tweets by Black in AI

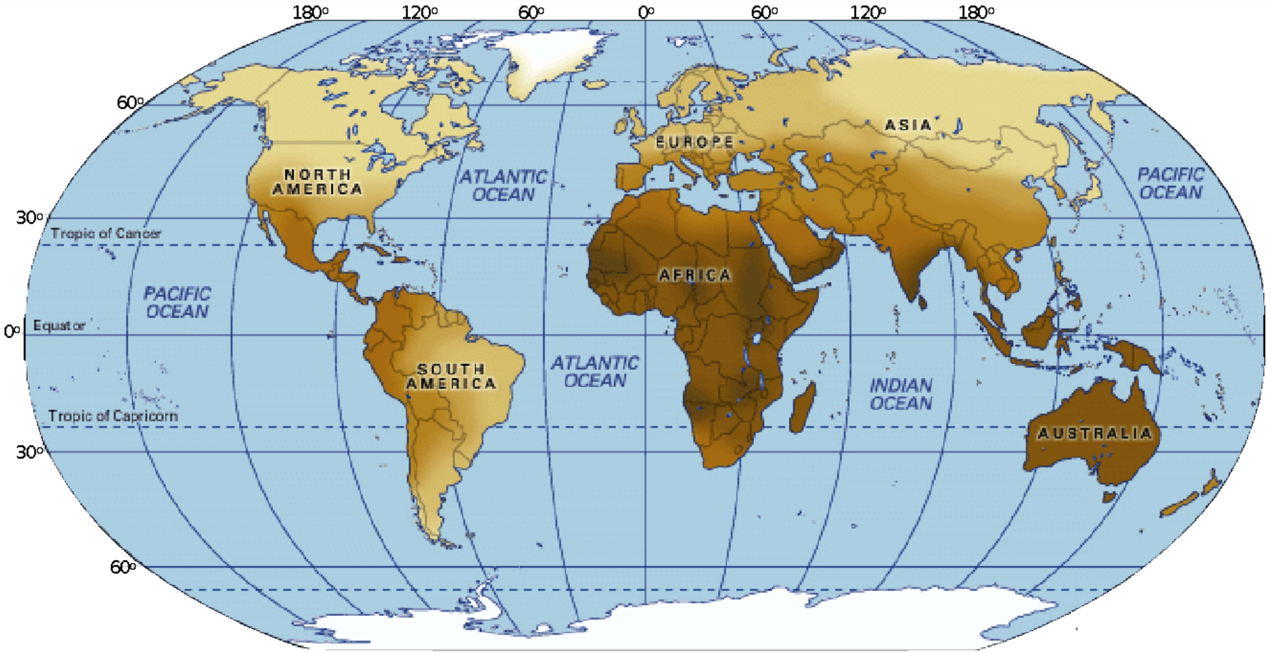
Related Organizations

- Black Girls Code
- Data Science Africa
- Deep Learning Indaba

Workshops

- 2nd Black in AI Workshop, 2018
  in Conjuction with NIPS 2018
- 1st Black in AI Workshop, 2017

The world is incredibly diverse



Darkest

Lightest

Intersectionality Matters

558 F.2d 480

15 Fair Empl.Prac.Cas. 573, 14 Empl. Prac. Dec. P 7692

Emma DeGRAFFENREID et al., Appellants,

v.

GENERAL MOTORS ASSEMBLY DIVISION, ST. LOUIS, et al., Appellees.

No. 76-1599.

United States Court of Appeals, Eighth Circuit.

Submitted March 18, 1977.

Decided July 15, 1977.

We have to analyze tech as something embedded in a sociotechnical system. Not enough to make everything “fair”

Microsoft improves facial recognition technology to perform well across all skin tones, genders

June 26, 2018 | John Roach

f t in

DIGITAL

Amazon Rekognition May Finally Be Audited and Ranked Alongside Other Vendors

A more universal test for facial recognition systems is needed

By Lisa Lacy | February 19, 2019

IBM

IBM Research Blog   Topics   Labs   About

AI

IBM Research Releases 'Diversity in Faces' Dataset to Advance Study of Fairness in Facial Recognition Systems

NEWS   MANAGEMENT   OVERSIGHT   DEFENSE   TECH   CONTRACTING   PAY & B

Senators Are Asking Whether Artificial Intelligence Could Violate U.S. Civil Rights Laws

By Dave Gershgorn | Quartz | September 21, 2018 | 5 Comments

Senators Kamala Harris and Cory Booker both signed letters to federal agencies asking about AI bias. J. Scott Applewhite/AP

RELATED

Shutdown Roundup: NTSB Isn't Investigating Accidents, Warren Warns of Decreased Financial Investigations and More

January 25, 2019 | 1 Comment

Democratic Senator Questions Mulvaney's Hatch Act Compliance

September 21, 2018 | 32 Comments

Why Philadelphia Is on the Federal Government's

New report from Georgetown Center from privacy and security **America Under Watch** detailing how police and ICE use face recognition technology asking for a moratorium

One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority

Science is inherently political.
So is technology. Question the power structure, funding and interests involved

Feminists have long critiqued “the view from nowhere”: the belief that science is about finding objective “truths” without taking people’s lived experiences into account.

According to most feminists and some pragmatists, the acknowledgment of both subject and object as historically and politically situated requires that the subjects and objects of knowledge be placed on a more level playing field. When this is done, objectivity, as a form of responding to the rights and well being of fellow subjects as well as the objects of scientific inquiry, must be considered (Heldke & Kellert, 1995). Objectivity, then, is achieved to the extent that responsibility in inquiry is fulfilled and expanded. It follows that scientists must be held accountable for the results of their projects and that scientists must acknowledge the political nature of their work. Objectivity understood as such implies relationships between people, objects, and inquiry projects as central to its conception (Sullivan, 2001). **-- Replacing the “View from Nowhere”: A Pragmatist-Feminist Science Classroom**

https://wolfweb.unr.edu/homepage/crowther/ejse/stitzlein.pdf

> … researchers are encouraged to take their privileges for granted, even to the point where these become invisible […] ignor[ing] how much labor is done for them, labor that allows them to be flexible, self-determining, and independent.
>
> - *Helmreich 1999*

Google Employees Walk Out To Protest Company's Treatment Of Women

November 1, 2018 · 7:10 AM ET

EMILY SULLIVAN LAUREL WAMSLEY

New York

San Francisco

There are no laws that restrict who can use our APIs & Datasets for what.

Gebru et al, FAT/ML 2018

It is even possible that some algorithms are breaking existing laws (e.g. EEOC)

Gebru et al, FAT/ML 2018

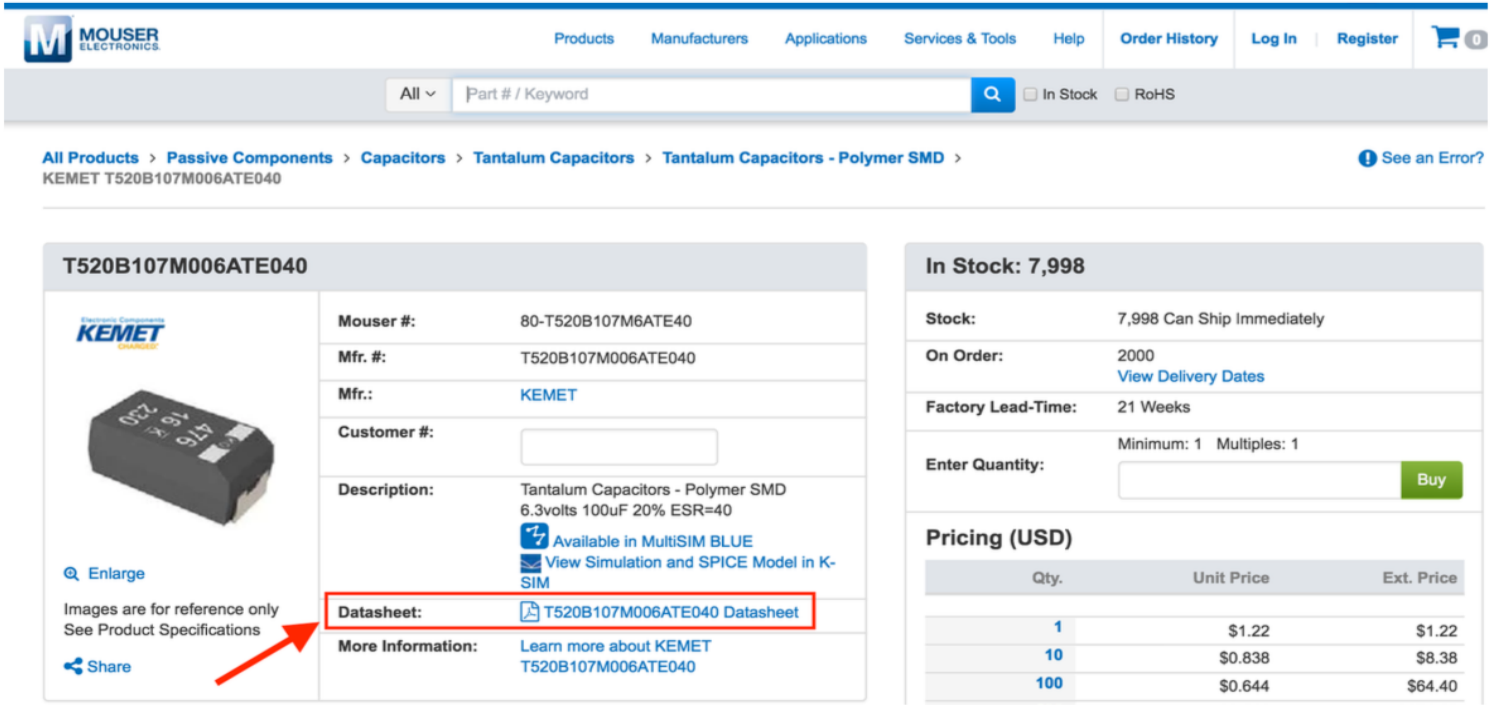
We need standards/documentation

Gebru et al, FAT/ML 2018

Other industries have been there

Gebru et al, FAT/ML 2018

Electronics



Gebru et al, FAT/ML 2018

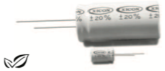
Electronics

XICON

Miniature Aluminum Electrolytic Capacitors — XRL Series

■ FEATURES

- Low impedance characteristics
- Case sizes are smaller than conventional general-purpose capacitors, with very high performance
- Can size larger than 9mm diameter has safety vents on rubber end seal
- RoHS Compliant



RoHS Compliant

■ CHARACTERISTICS

Item	Characteristics													
Operating Temperature Range	-40°C ~ +85°C													
Capacitance Tolerance	±20% at 120Hz, 20°C													
Leakage Current	≤100V	I = 0.01CWV or 3μA whichever is greater after 2 minutes of applied rated DC working voltage at 20°C Where: C = rated capacitance in μF; WV = rated DC working voltage												
	>100V	CWV ≤ 1000 μF: I= 0.03 CWV + 15uA; C= rated capacitance in uF CWV ≥ 1000 μF: I= 0.02 CWV + 25uA; WV= rated DC working voltage in V												
Dissipation Factor (Tan δ, at 20°C 120Hz)	Working voltage (WV)		6.3	10	16	25	35	50	63	100	160	250	350	450
	Tan δ		0.23	0.20	0.16	0.14	0.12	0.10	0.09	0.08	0.12	0.17	0.20	0.25
	For capacitors whose capacitance exceeds 1,000μF, the specification of tan δ is increased by 0.02 for every addition of 1,000μF													
Surge Voltage	Working voltage (WV)		6.3	10	16	25	35	50	63	100	160	250	350	450
	Surge voltage (SV)		8	13	20	32	44	63	79	125	200	300	400	500
Low Temperature Characteristics (Imp. ratio @ 120Hz)	Working voltage (WV)		6.3	10	16	25	35	50	63	100	160	250	350	450
	Z(-25°C)/Z(+20°C)	øD<16	6	4	3	3	2	2	2	2	3	8	12	16
		øD≥16	8	6	4	4	3	3	3	3	3	8	12	16
	Z(-40°C)/Z(+20°C)	øD<16	10	8	6	6	4	3	3	3	4	10	16	20
		øD≥16	18	16	12	10	8	8	6	6	4	10	16	20
Load Test	When returned to +20°C after 2,000 hours application of working voltage at +85°C, the capacitor will meet the following limits: Capacitance change is ≤ ±20% of initial value; tan δ is < 200% of specified value; leakage current is within specified value													
Shelf Life Test	When returned to +20°C after 1,000 hours at +85°C with no voltage applied, the capacitor will meet the following limits: Capacitance change is ≤ ±20% of initial value; tan δ is < 200% of specified value; leakage current is within specified value													

■ PART NUMBERING SYSTEM

1 4 0	–	X R L	1 6 V	1 0 0	–	R C
Prefix		Series	Voltage Actual Value	Capacitance (μF) Actual Value		Suffix RoHS Compliant

■ RIPPLE CURRENT AND FREQUENCY MULTIPLIERS

Capacitance (μF)	Frequency (Hz)				
	60 (50)	120	500	1K	≥10K
<100	0.70	1.0	1.30	1.40	1.50
100 ~ 1000	0.75	1.0	1.20	1.30	1.35
>1000	0.80	1.0	1.10	1.12	1.15

■ RIPPLE CURRENT AND TEMPERATURE MULTIPLIERS

Temperature (°C)	<50	70	85
Multiplier	1.78	1.4	1.0

XICON

XICON PASSIVE COMPONENTS • (800) 628-0544

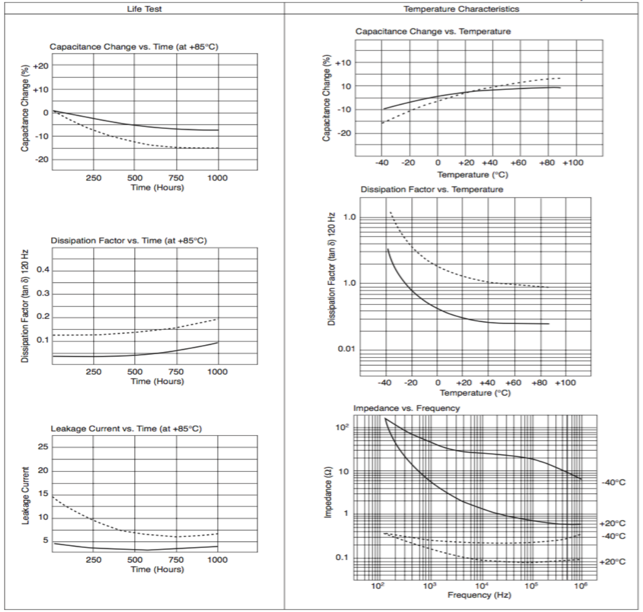
XC-600178 Date Revised: 1/8/07

Specifications are subject to change without notice. No liability or warranty implied by this information. Environmental compliance based on producer documentation.

XICON

Miniature Aluminum Electrolytic Capacitors — XRL Series

■ TYPICAL PERFORMANCE CHARACTERISTICS



XICON

XICON PASSIVE COMPONENTS • (800) 628-0544

XC-600178 Date Revised: 1/8/07

Specifications are subject to change without notice. No liability or warranty implied by this information. Environmental compliance based on producer documentation.

Gebru et al, FAT/ML 2018

Datasheets for Datasets

We need datasheets for APIs, pretrained models and datasets

Gebru et al, FAT/ML 2018, Mitchell et al, FAT* 2019

Datasheets for Datasets

Need to have information regarding standard operating characteristics, recommended usage, how the dataset was gathered etc..

Gebru et al, FAT/ML 2018, Mitchell et al, FAT* 2019

Datasheets for Datasets

E.g. we do not expect Face API to accurately identify the gender of young children.

Gebru et al, FAT/ML 2018, Mitchell et al, FAT* 2019

Datasheets for Datasets

Are there disclaimers in case someone uses this API for something it was not intended for?

E.g. in electronics, disclaimers for use of components in high stakes scenarios like nuclear power plants, life support…

Gebru et al, FAT/ML 2018, Mitchell et al, FAT* 2019

Datasheets for Datasets

What are some of the characteristics of the data it was trained on?

Gebru et al, FAT/ML 2018, Mitchell et al, FAT* 2019

Datasheets for Datasets

E.g. distribution of age, skin types, geography, gender

Gebru et al, FAT/ML 2018, Mitchell et al, FAT* 2019

Datasheets for Datasets

Motivation for Dataset Creation

Why was the dataset created? (e.g., was there a specific task in mind? was there a specific gap that needed to be filled?)
Labeled Faces in the Wild was created to provide images that can be used to study face recognition in the unconstrained setting where image characteristics (such as pose, illumination, resolution, focus), subject demographic makeup (such as age, gender, race) or appearance (such as hairstyle, makeup, clothing) cannot be controlled. The dataset was created for the specific task of pair matching: given a pair of images each containing a face, determine whether or not the images are of the same person.[1]

What (other) tasks could the dataset be used for?
The LFW dataset can be used for the face identification problem. Some researchers have developed protocols to use the images in the LFW dataset for face identification.[2]

Has the dataset been used for any tasks already? If so, where are the results so others can compare (e.g., links to published papers)?
Papers using this dataset and the specified evaluation protocol are listed in http://vis-www.cs.umass.edu/lfw/results.html

Who funded the creation of the dataset?
The building of the LFW database was supported by a United States National Science Foundation CAREER Award.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)? Features/attributes? Is there a label/target associated with instances? If the instances related to people, are subpopulations identified (e.g., by age, gender, etc.) and what is their distribution?
Each instance contains a pair of images that are 250 by 250 pixels in JPEG 2.0 format. Each image is accompanied by a label indicating the name of the person in the image. While subpopulation data was not available at the initial release of the dataset, a subsequent paper[3] reports the distribution of images by age, race and gender. Table 2 lists these results.

Is everything included or does the data rely on external resources? (e.g., websites, tweets, datasets) If external resources, a) are there guarantees that they will exist, and remain constant, over time; b) is there an official archival version; c) are there access restrictions or fees?
Everything is included in the dataset.

Are there recommended data splits and evaluation measures? (e.g., training, development, testing; accuracy or AUC)
The dataset comes with specified train/test splits such that none of the people in the training split are in the test split and vice versa. The data is split into two views, View 1 and View 2. View 1 consists of a training subset (pairsDevTrain.txt) with 1100 pairs of matched and 1100 pairs of mismatched images, and a test subset (pairsDevTest.txt) with 500 pairs of matched and mismatched images. Practitioners can train an algorithm on the training set and test on the test set, repeating as often as necessary. Final

Gebru et al, FAT/ML 2018

Model Cards for Model Reporting

Model Card - Smiling Detection in Images

Model Details

- Developed by researchers at Google and the University of Toronto, 2018, v1.
- Convolutional Neural Net.
- Pretrained for face recognition then fine-tuned with cross-entropy loss for binary smiling classification.

Intended Use

- Intended to be used for fun applications, such as creating cartoon smiles on real images; augmentative applications, such as providing details for people who are blind; or assisting applications such as automatically finding smiling photos.
- Particularly intended for younger audiences.
- Not suitable for emotion detection or determining affect; smiles were annotated based on physical appearance, and not underlying emotions.

Factors

- Based on known problems with computer vision face technology, potential relevant factors include groups for gender, age, race, and Fitzpatrick skin type; hardware factors of camera type and lens type; and environmental factors of lighting and humidity.
- Evaluation factors are gender and age group, as annotated in the publicly available dataset CelebA [36]. Further possible factors not currently available in a public smiling dataset. Gender and age determined by third-party annotators based on visual presentation, following a set of examples of male/female gender and young/old age. Further details available in [36].

Quantitative Analyses

Mitchell et al, FAT* 2019

Automobiles

Gebru et al, FAT/ML 2018

Automobiles

- No stop signs, drivers licenses drunk driving laws, seatbelts etc.
- Lots of accidents
- Crash tests done on male dummies
- Studies show that accidents disproportionately affected women

Gebru et al, FAT/ML 2018

Clinical Trials

- Used to be illegal
- Illegal experimentation on vulnerable populations
- Women were not required to be part of clinical trials until recently
- Study shows that 8-10 drugs that were pulled from circulation between 1997-2001 disproportionately affected women

Gebru et al, FAT/ML 2018

Clinical Trials

It took many years for standards to be placed and we are still suffering consequences from bias in automobile design and clinical trials

Gebru et al, FAT/ML 2018

Concluding Remarks

Societal biases enter when we

- **Formulate what problems to work on**
- Collect training and evaluation data
- Architect our models and loss functions
- Analyze how our models are used

Spatial Apartheid

Spatial Apartheid

Ongoing and Future Work

FGVCx Cassava disease diagnosis

iCassava Challenge 2019

20,000 leaf images
5-class prediction problem

FGVC 2019

Ongoing and Future Work

Disease Incidence

Healthy

Bacterial Blight

Green Mite

Mosaic Disease

Brown Streak

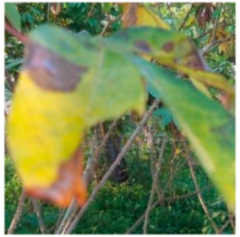
Disease Severity (Mosaic Disease)

Severity-1

Severity-2

Severity-3

Severity-4



Severity-5

Concluding Remarks

Societal biases enter when we

- Formulate what problems to work on
- **Collect training and evaluation data**
- Architect our models and loss functions
- Analyze how our models are used

Concluding Remarks

…while the fair ML literature has largely focused on "de-biasing" methods and viewed the training data as fixed, most of our interviewees report that their teams consider data collection, rather than model development, as the most important place to intervene

Holstein et al FAT* 2019

We need to end "parachute" research which sidelines the work of African scientists

By Moses John Bockarie,l Njala University • January 29, 2019

Some reading Material

- Ali Alkhatib: Anthropological/Artificial Intelligence & the HAI
  https://ali-alkhatib.com/blog/anthropological-intelligence
- Philip Rogaway: The moral character of cryptographic work
- Clare Garvie & Laura Moy: America Under Watch https://www.americaunderwatch.com/
- Stitzlein, Sarah M. "Replacing the 'View from Nowhere': A Pragmatist-Feminist Science Classroom." Electronic Journal of Science Education (2004).

Questions?