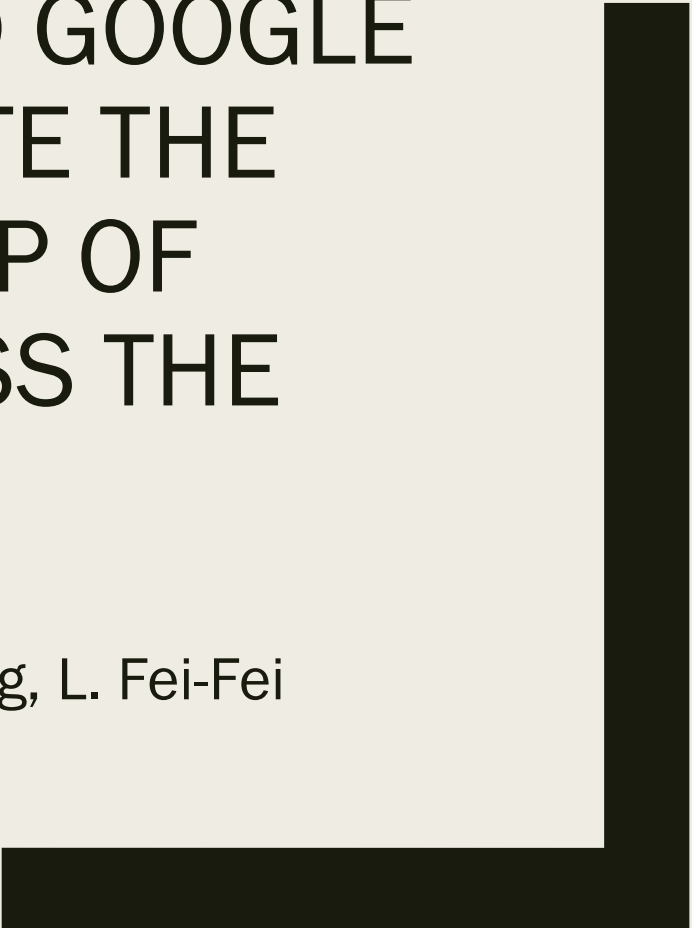


USING DEEP LEARNING AND GOOGLE STREET VIEW TO ESTIMATE THE DEMOGRAPHIC MAKEUP OF NEIGHBORHOODS ACROSS THE UNITED STATES

T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, L. Fei-Fei

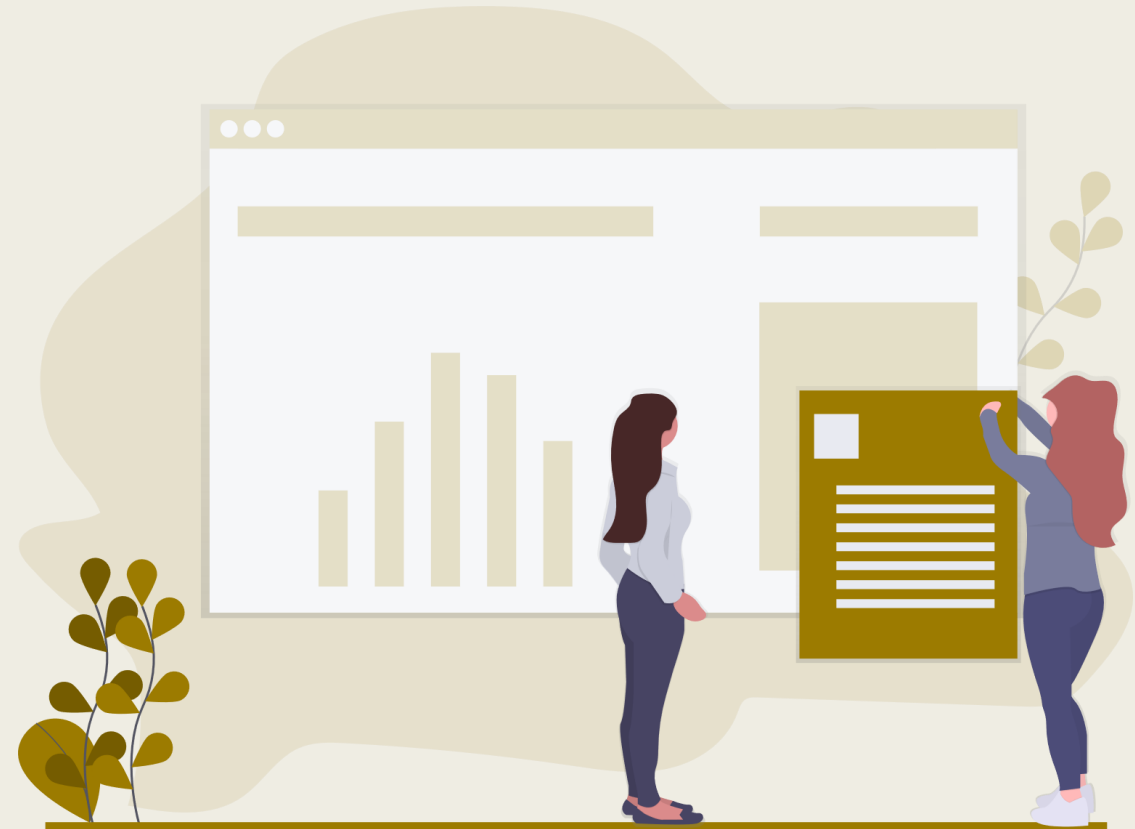
DH-500 Computational Social Media

Maxime Jan



Outline

- Problem solving
- Research question
- Methods and Data
- Results
- Conclusion



Problem solving

- American Community Survey (ACS)
 - *Door-to-door study*
 - *Gathers demographic factors*
 - *\$250 million/year*
 - *Lag between demographic changes and census*
 - *Only interrogates bigger regions*

- Can computational methods help in this task ?



AMERICAN
COMMUNITY
SURVEY

U.S. CENSUS BUREAU

Education of population 18-24	Estimate	Margin of error
Less than high school graduate	3'662'080	± 41'512
High school graduate	9'921'331	± 56'156
College degree	13'168'280	± 67'371
Bachelor's degree or higher	3'621'479	± 38'999

Source : <https://data.census.gov/cedsci/table?q=education&tid=ACSST1Y2019.S1501>

Research question

Income, Race, Education,
and Voting Patterns

Can socioeconomic statistics be inferred
from objective characteristics of images
from a neighborhood ?

Vehicles' characteristics (Make, Model, Year, Price,
etc...)

Research question

Can the vehicles of a neighbourhood predict its average income, education level, race distribution and voting preferences ?



METHODS AND DATA



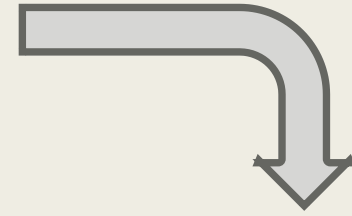
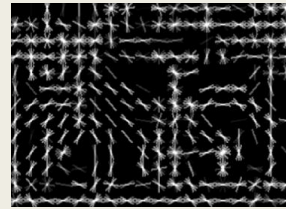
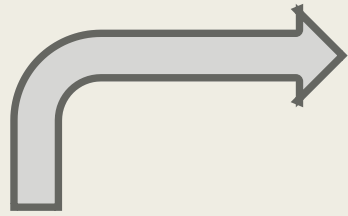
Neighbourhood images

- 200 cities
 - 2 largest cities from each state
 - Next 100 largest cities
- Grid of 20km
- 6 images every 25m
- 50 million images



Car detector

Deformable Part Models

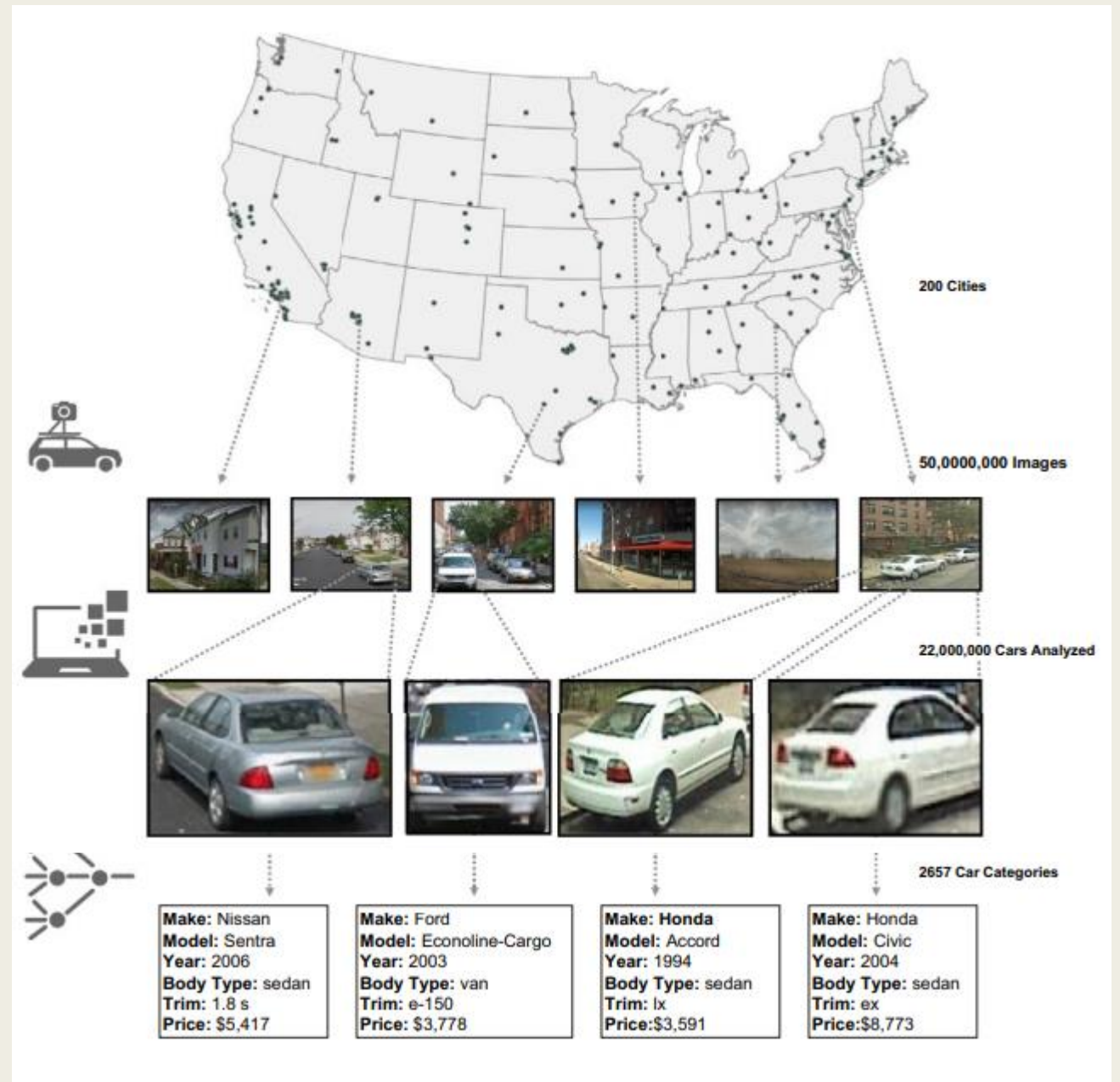


amazon mechanicalturk™
Artificial Artificial Intelligence

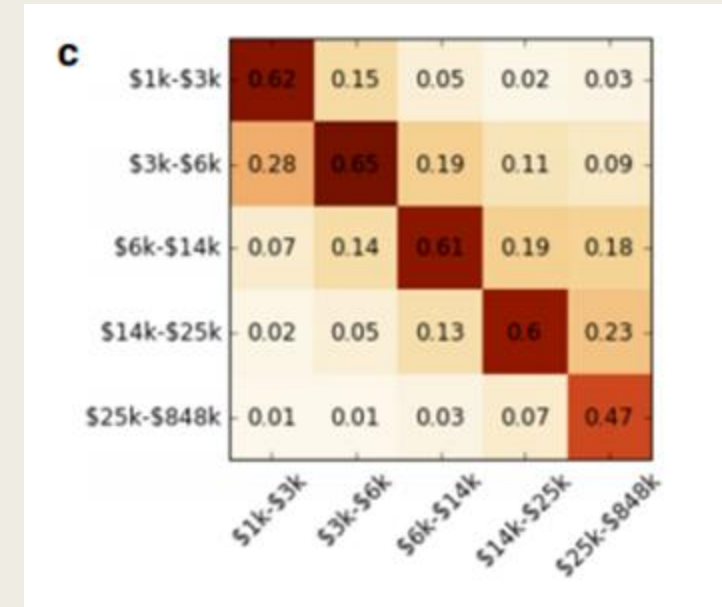
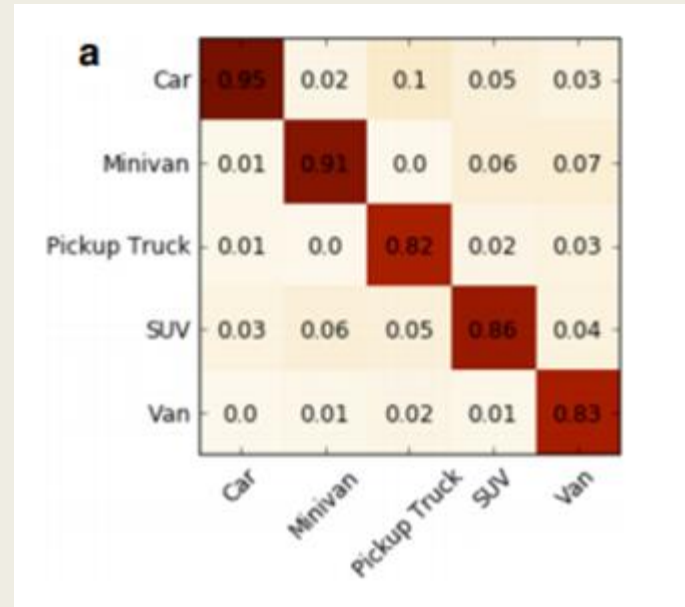


Car classifier

- Annotation by car experts
- >2000h of combined annotations
- CNN for automatic car classification
- 2657 categories



Car classifier



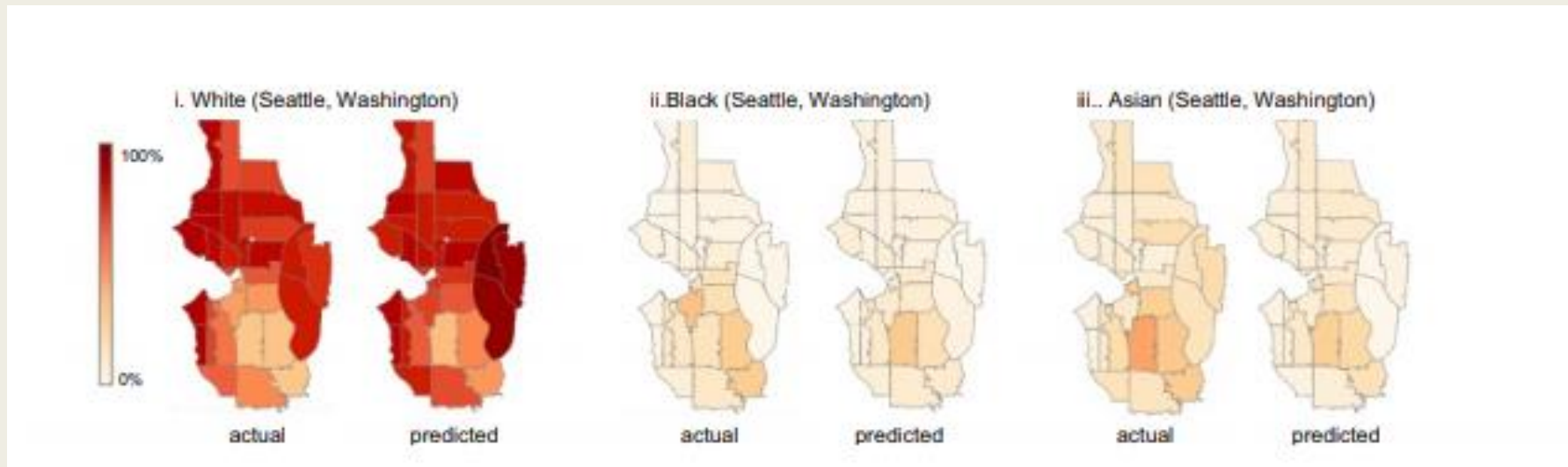
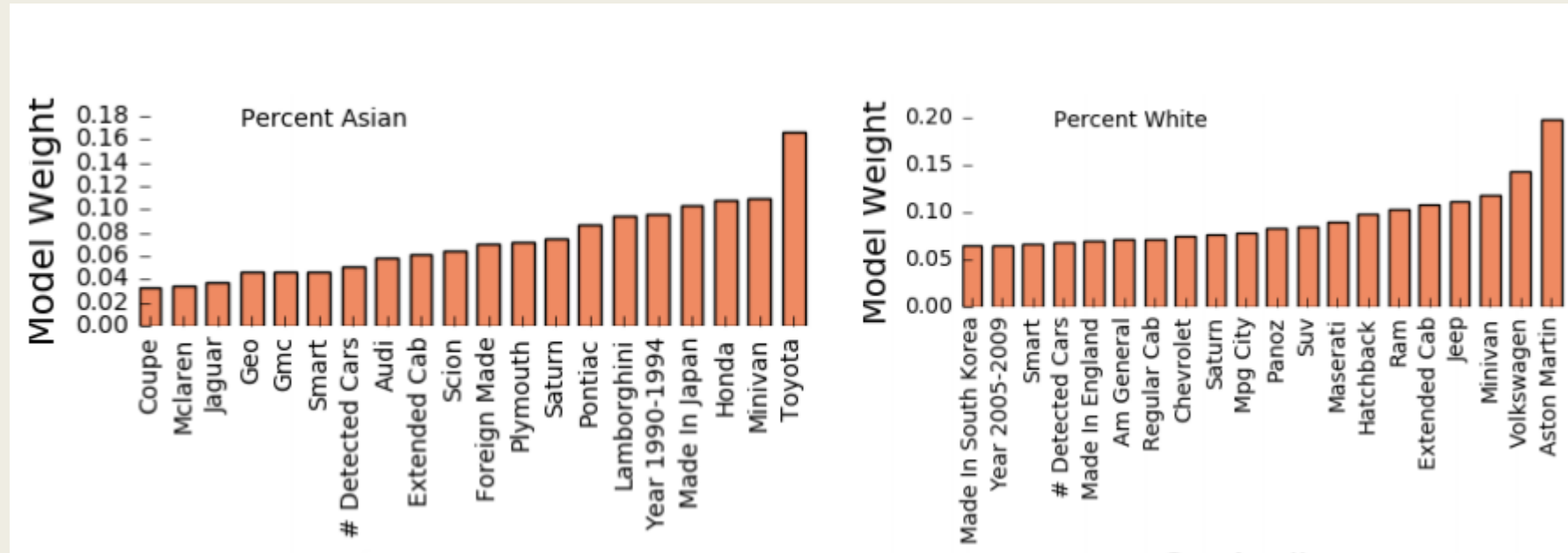
- On the fine-grained level (2'657 classes) the classifier achieves 33.27% accuracy

Final dataset and method

- Data is
 - *divided by county*
 - *labeled with socioeconomics data from the ACS and 2012 voting data*
 - *Grouped in training and test set*
 - Training set : counties starting by A, B or C
 - Test set : all the others
- Logistic regression to predict race and education level
- Ridge regressor to predict income and voter preferences

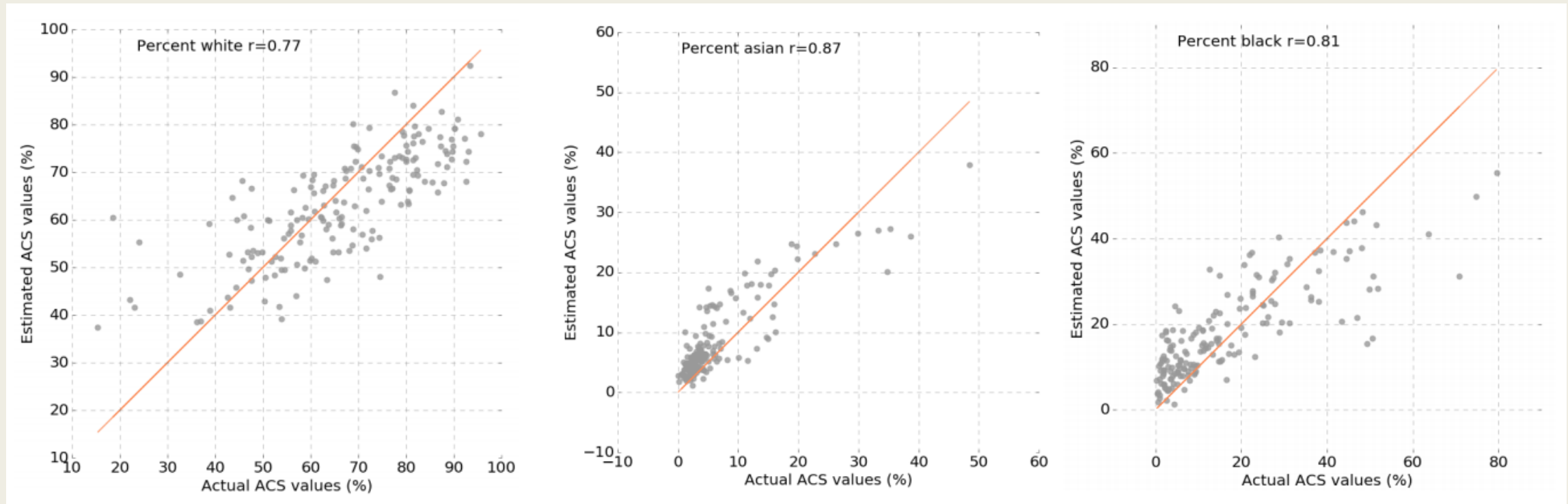
Results

Race



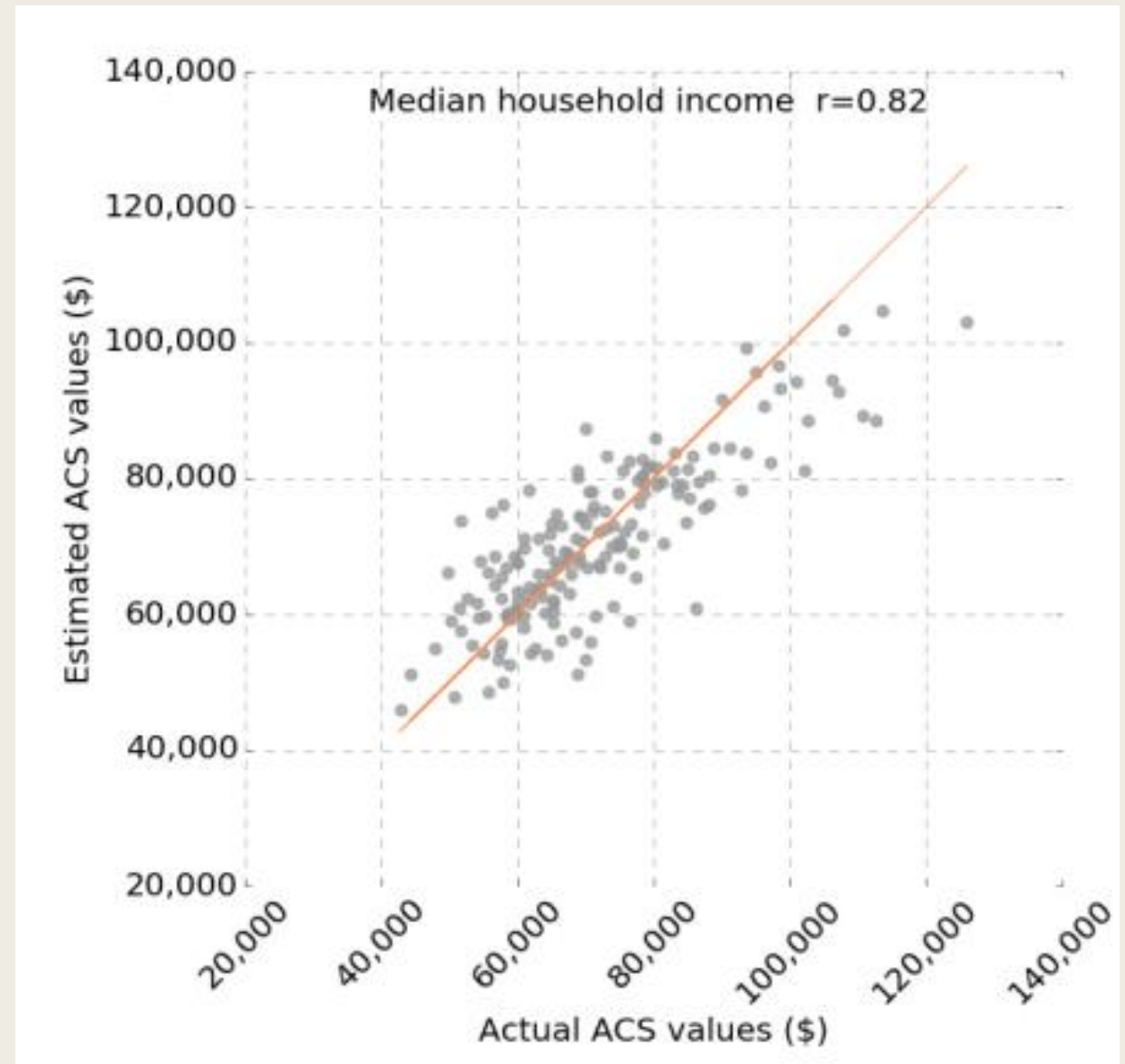
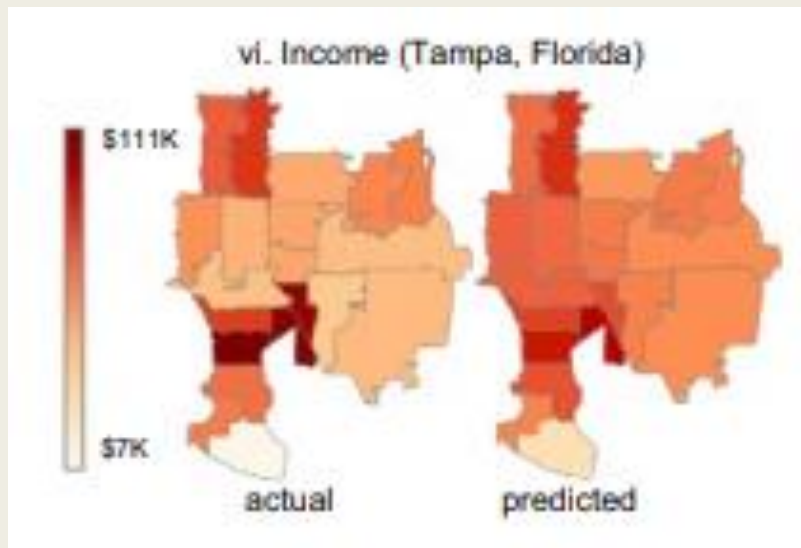
Results

Race



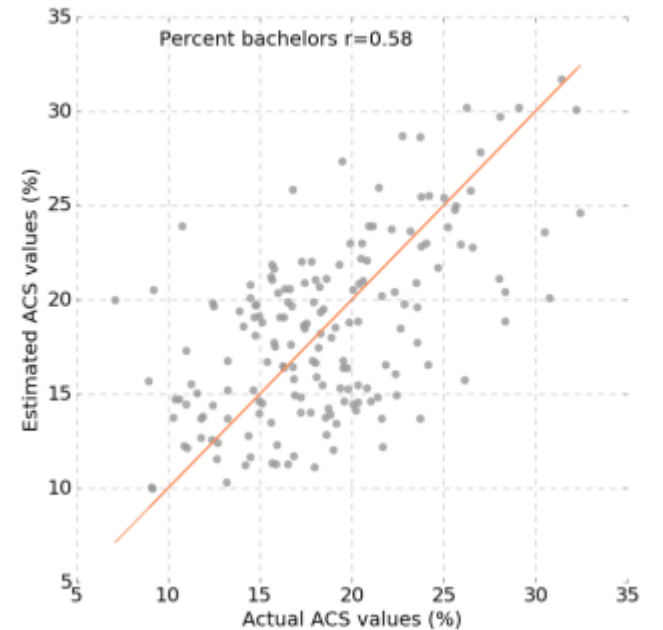
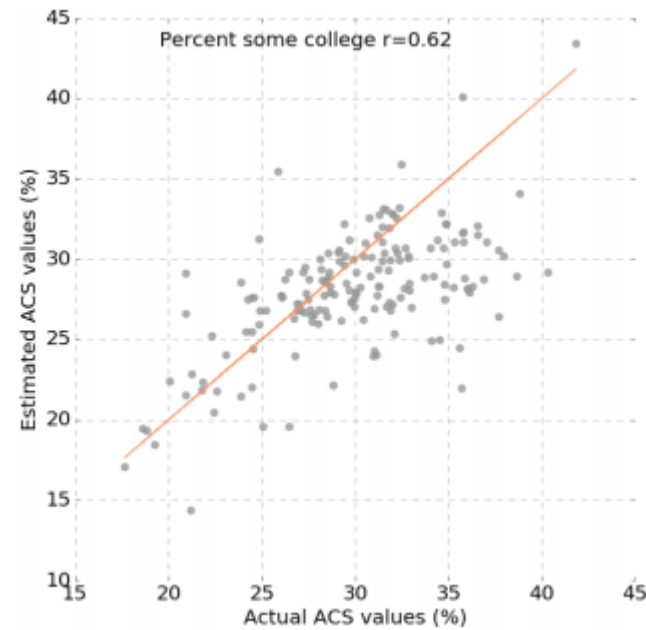
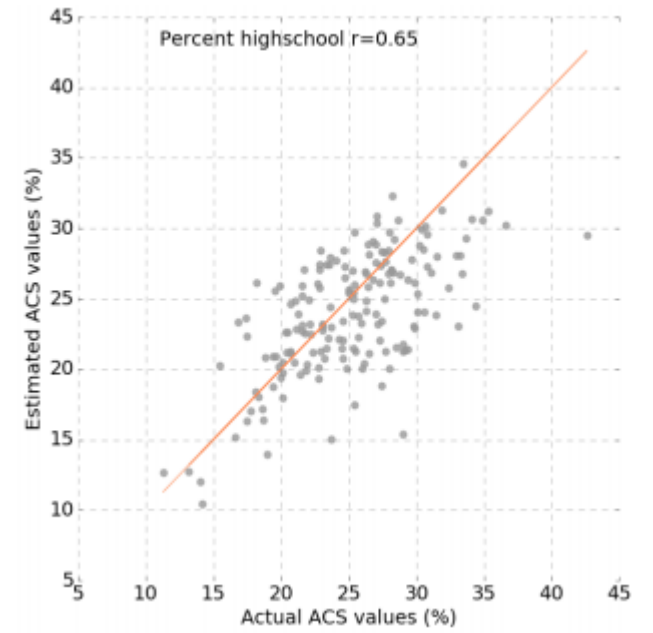
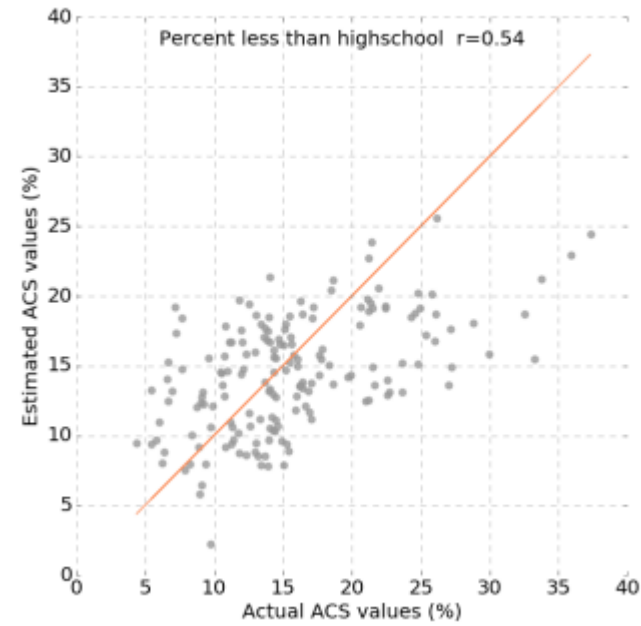
Results

Income



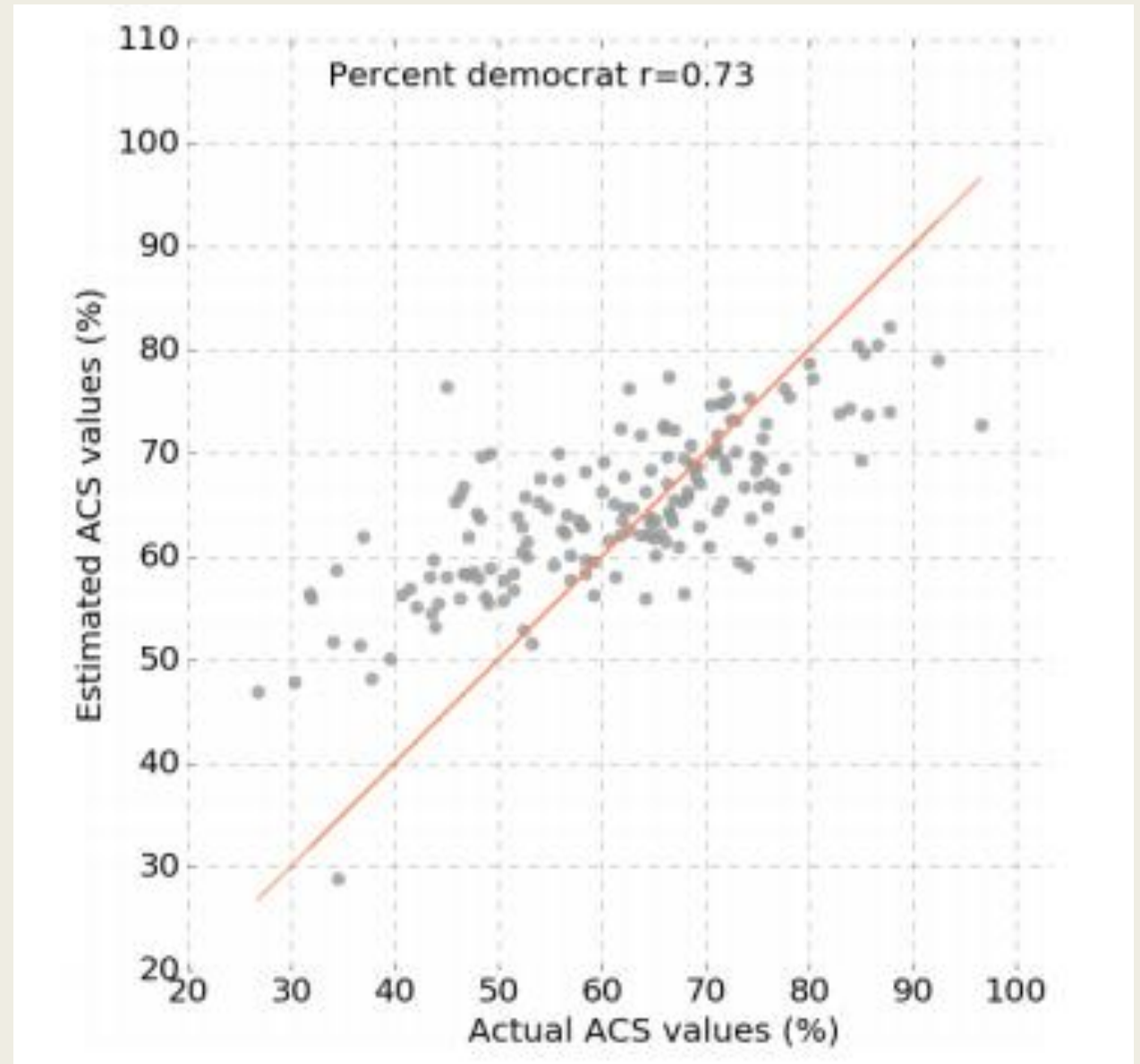
Results

Education



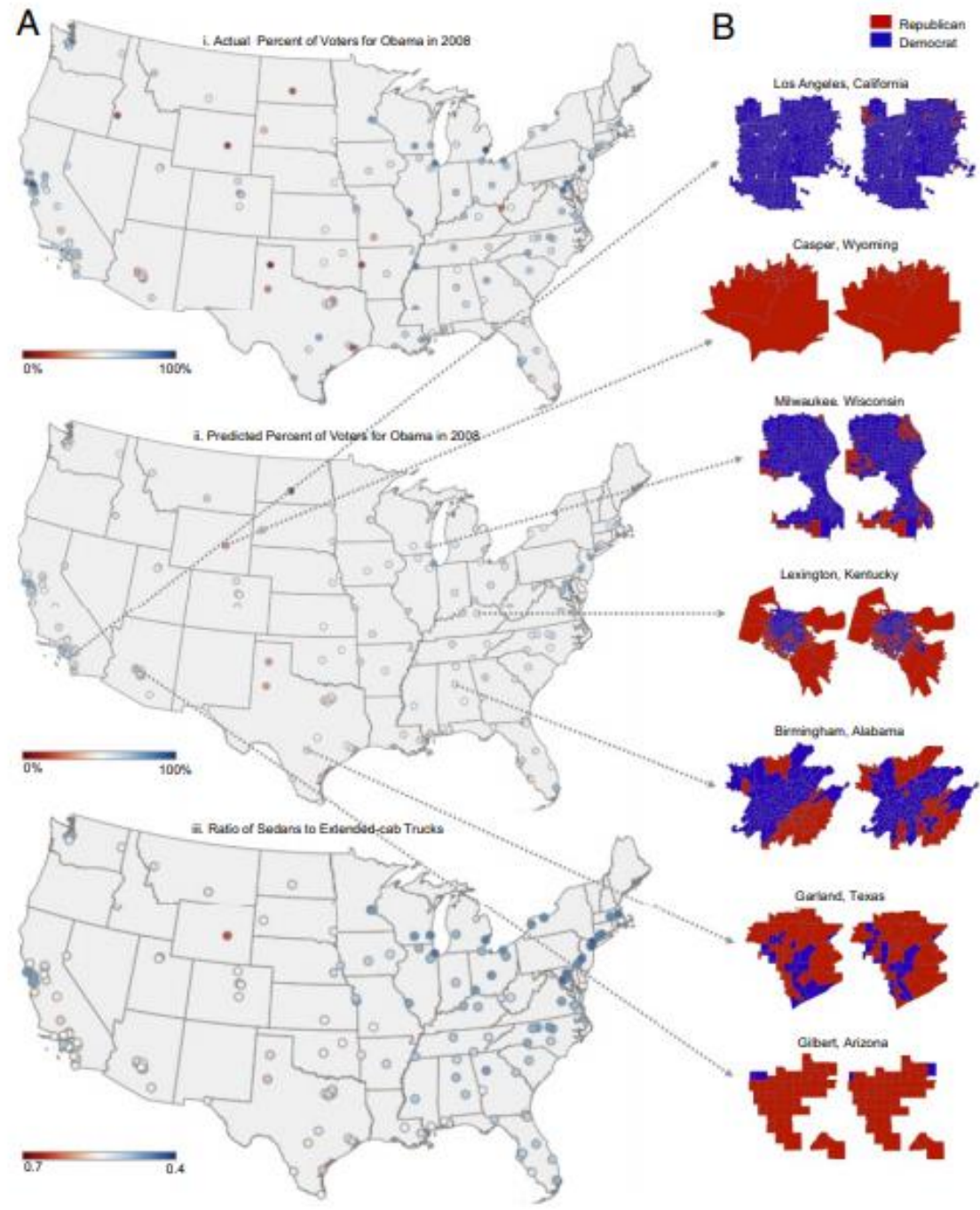
Results

Voter preferences



Results

Voter preferences



Results

Voter preferences



Conclusion

- Results show that Machine Learning algorithms can predict socioeconomic attributes based on cars' data
- Particularly interesting with the apparition of self-driving cars
- Methods could be improved : add new features
- This technique raises ethical issues