

The High Resolution Settlement Layer (HRSL)

Greg Yetman

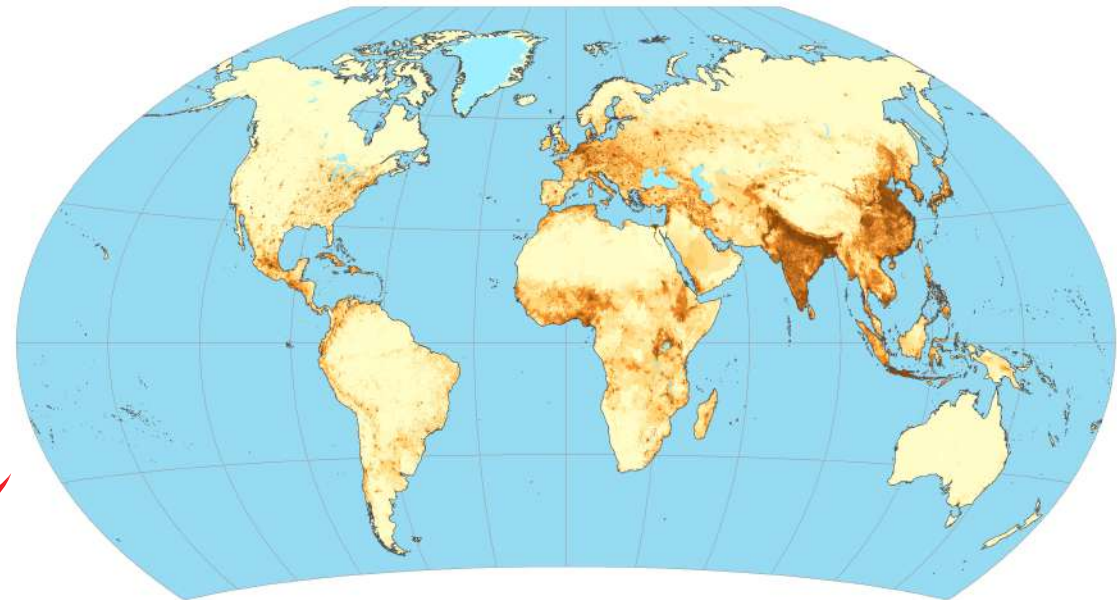
gyetman@ciesin.columbia.edu

HRSL Creation

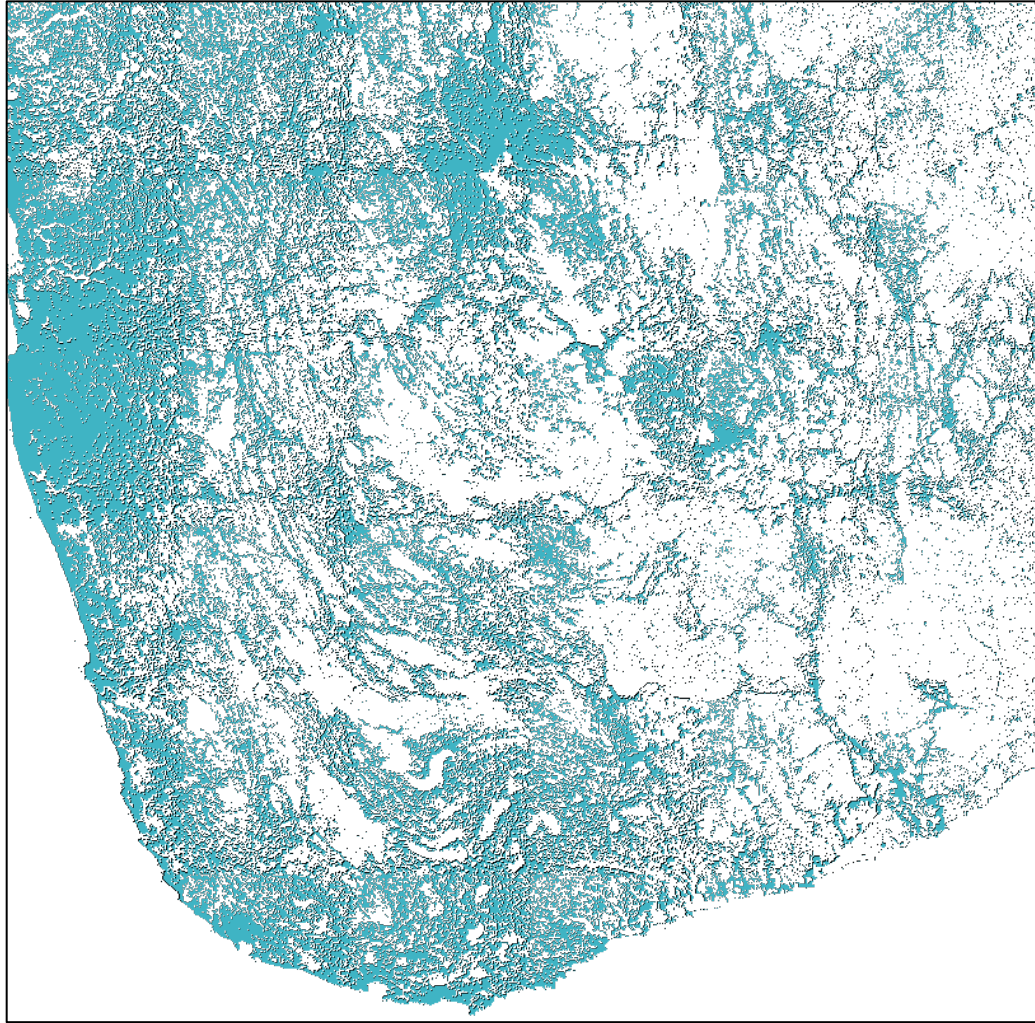
- Initially developed to support Internet.org initiative to provide rural areas access to Internet, with other uses by the public
 - 34 Countries released in CIESIN-Facebook partnership
- High demand from Humanitarian community led to wider release via HDX by Facebook Data for Good group
 - Now over 100 countries
 - Many have age and sex variables available as well
 - Frequent (ongoing) updates planned

Methods

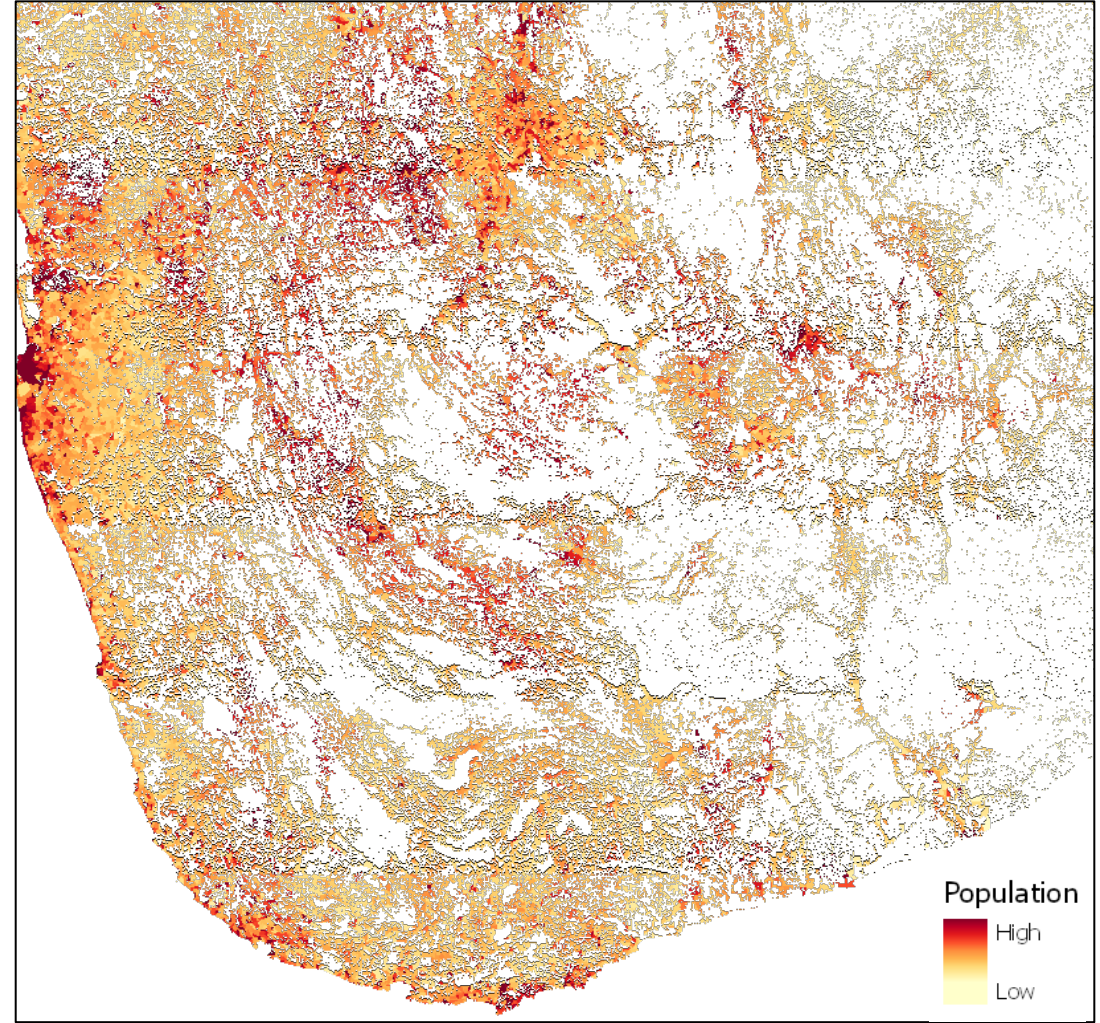
- Uses 50cm Maxar (formerly DigitalGlobe) optical imagery and CIESIN/SEDAC's base data for Gridded Population of the World (GPWv4)
- Identify 30m blocks of pixels that contain buildings using a neural network
- Apportion population counts to 30m blocks using proportional aggregation



Example: Southwestern Sri Lanka

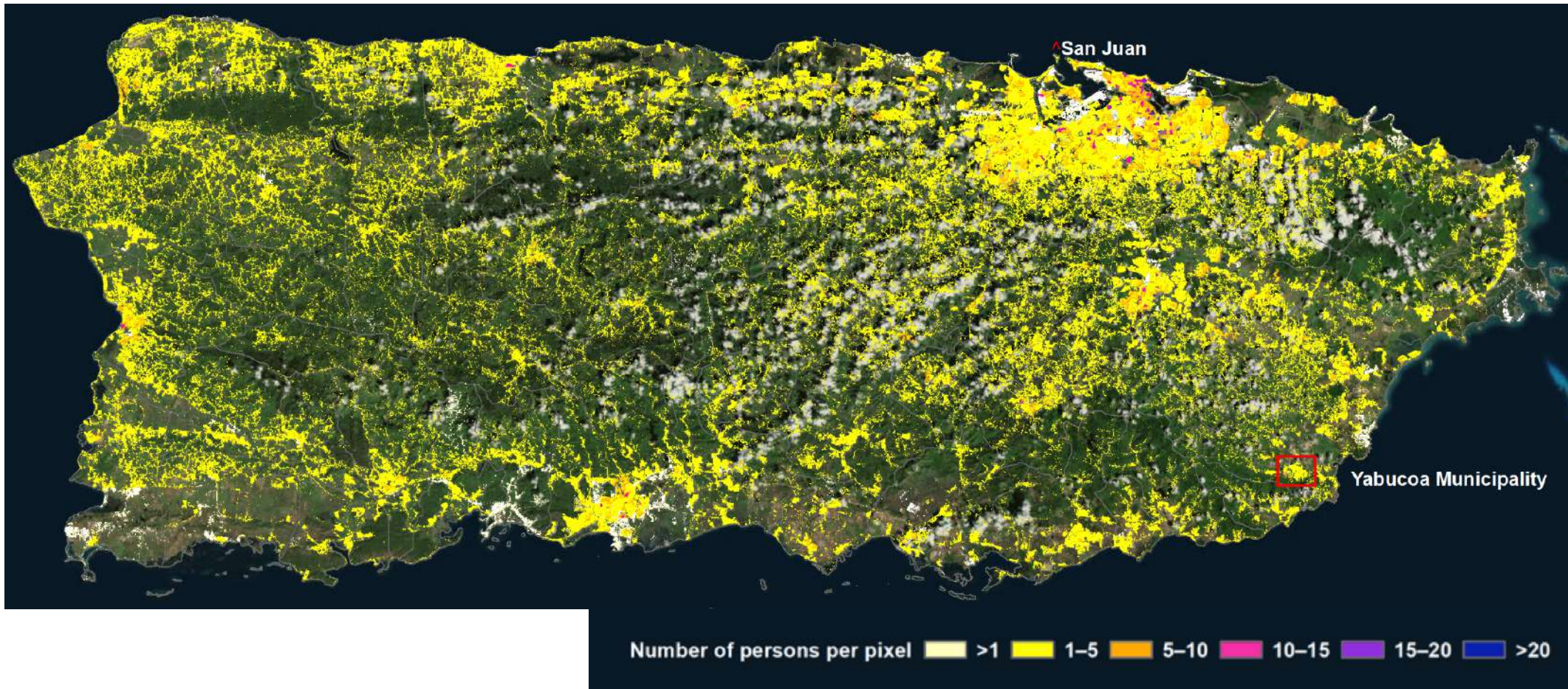


Boolean settlement surface



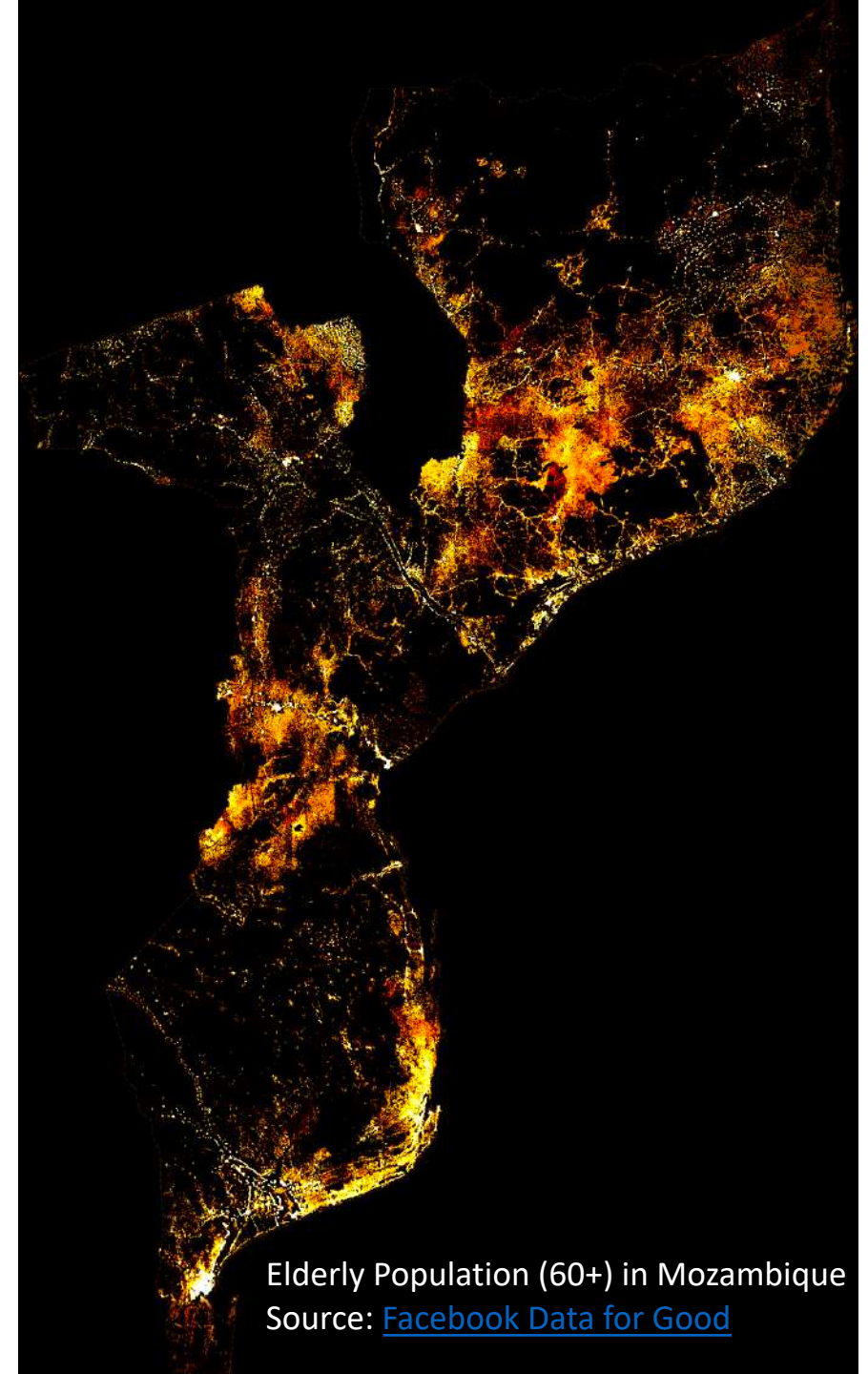
Population counts

Puerto Rico released after Maria



Adding Demographic Data

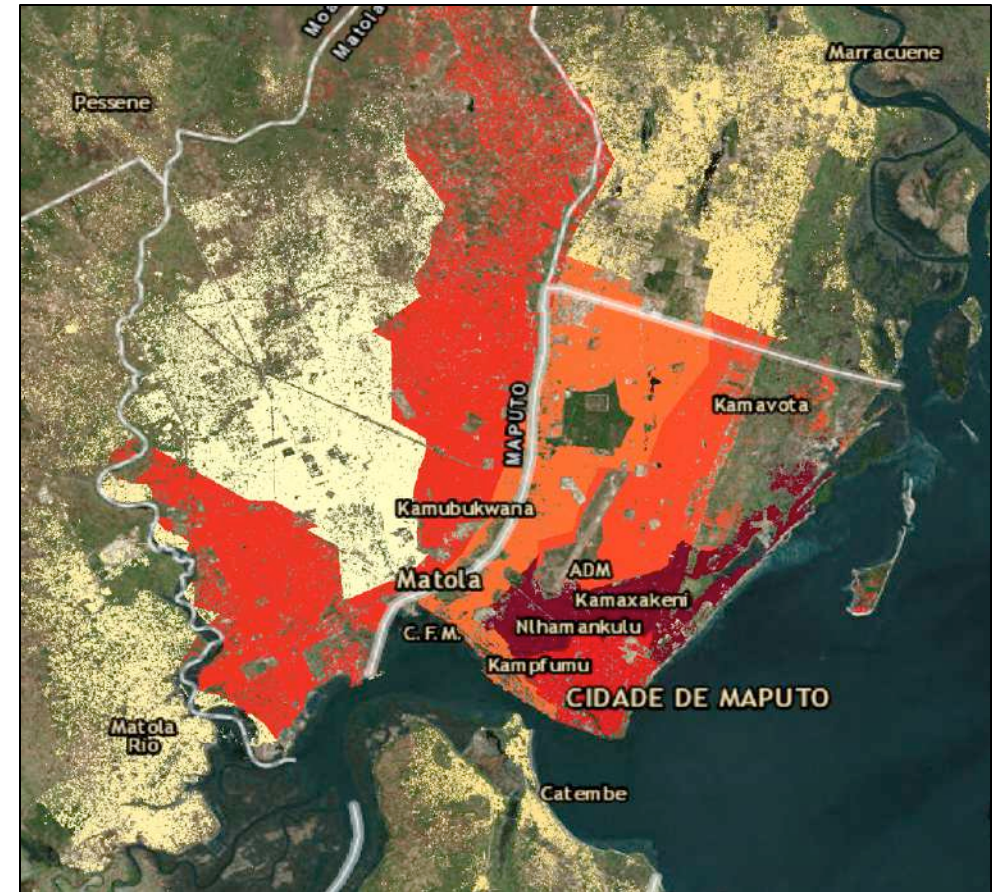
- Census estimates of age and sex for 2010 downscaled to population estimates
 - 5 year age categories by sex
 - Based on GPWv4 data



Elderly Population (60+) in Mozambique
Source: [Facebook Data for Good](#)

Quality Issues & Remedies

- Perpetually cloudy areas
- Census data issues in some countries
 - Out of date
 - Poor boundary location, detail
- Use of alternative urban classification (RADAR-based)
- Possible future improvements
 - Updated estimates from microcensus
 - Predictive population modeling

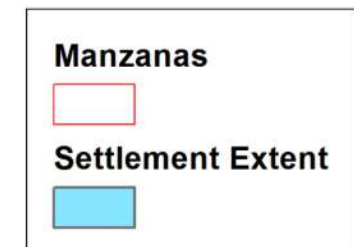


Evaluating Population Distribution Models

- High resolution data aids in developing methods for estimation of population and infrastructure data in areas where detailed data are not available or out of date



INEGI census population counts over settlement classification and high-resolution imagery



Machine Learning & Expanded Data Better Products?

- Increasing compute, decreasing cost
- New methods designed for new data sources
- Greater openness: code, software, methods, data
- Algorithms designed to work well with biased data: tension between results (black box) and understanding (hypothesis testing)
 - Black box model output as inputs for applied programs are less problematic



xkcd: <https://xkcd.com/1838/>

- Amara's Law: We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.

Is the Census approach useful for Gridded Data?

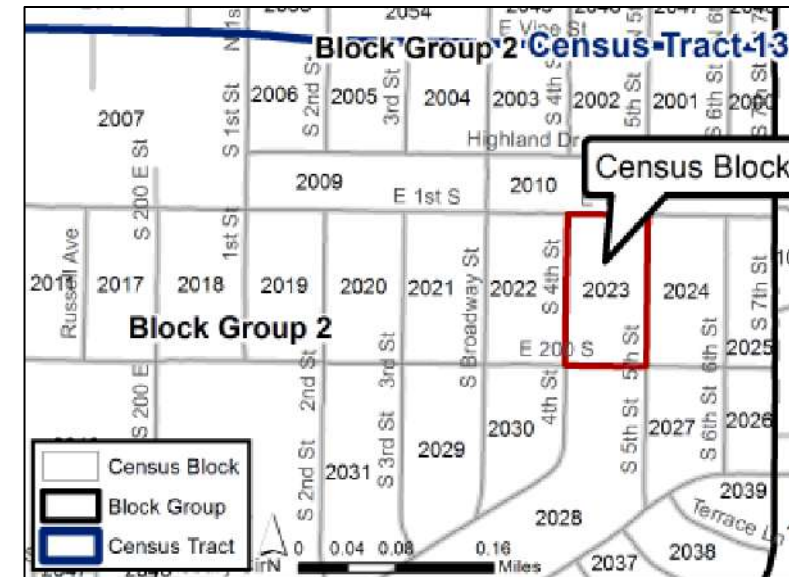
- Basic demographics at a detailed level are very valuable

- U.S. Census data example

- Average population: 34
- Over 11 million geographic units

- Average population 1,200
- 211,000 geographic units

Source: [U.S. Census Bureau](#)

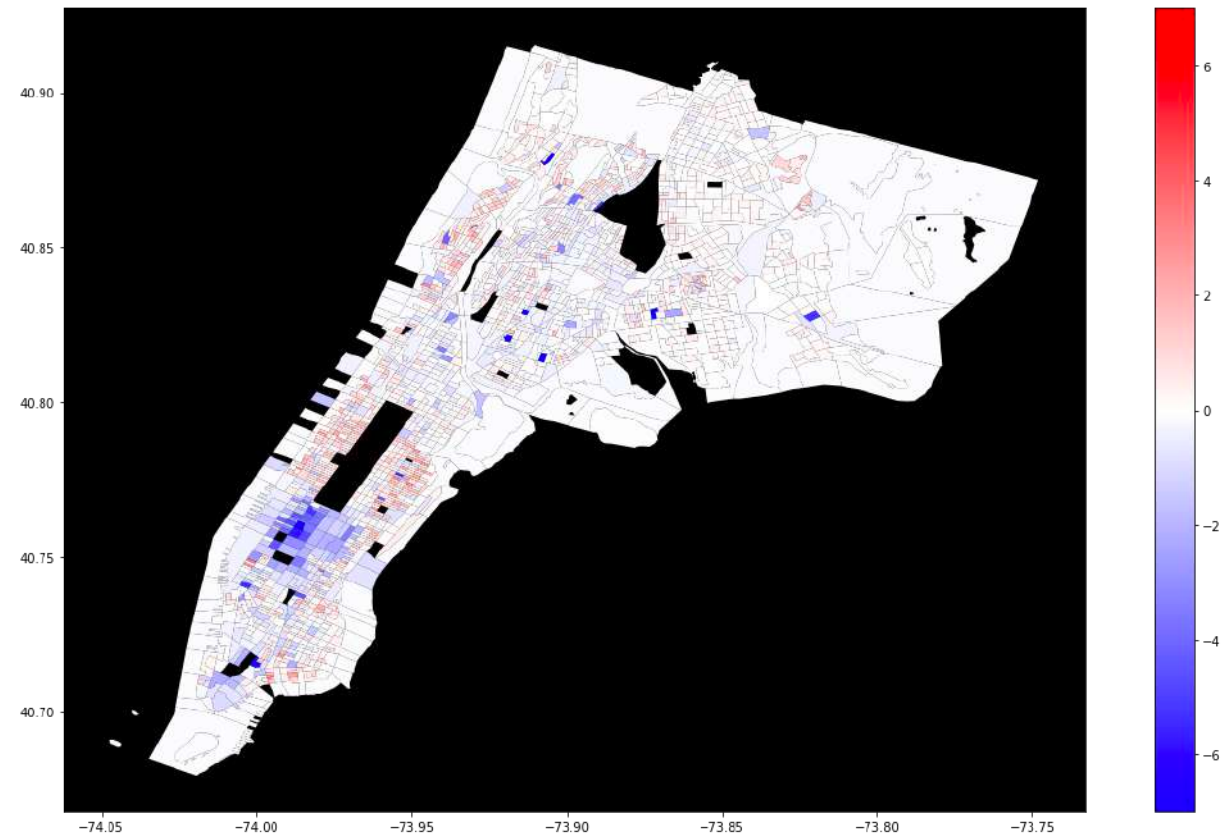


Wither Privacy?

Striking a balance is difficult

- Additional data sources **not** used in HRSL make very-fine grained population estimates possible, even when pseudo-anonymized
 - Social media APIs
 - Mobile device check-in data
- High demand for detailed data in time and space
 - Hazards mitigation & response
 - Health service delivery

Linear Scaled Difference
Census population - summed
mobile device check-ins



Sources: U.S. Census Bureau, SafeGraph

Notes & Acknowledgements

Data are available for download

<https://data.humdata.org/organization/facebook>

Questions? Please reach out!

gyetman@ciesin.columbia.edu

Funding for CIESIN's contributions to the HRSL was supplied by Facebook.

GPWv4 data used were developed with funding from the National Aeronautics and Space Administration under Contract 80GSFC18C0111 for the Socioeconomic Data and Applications Distributed Active Archive Center (DAAC).