High resolution population grid for the entire United States

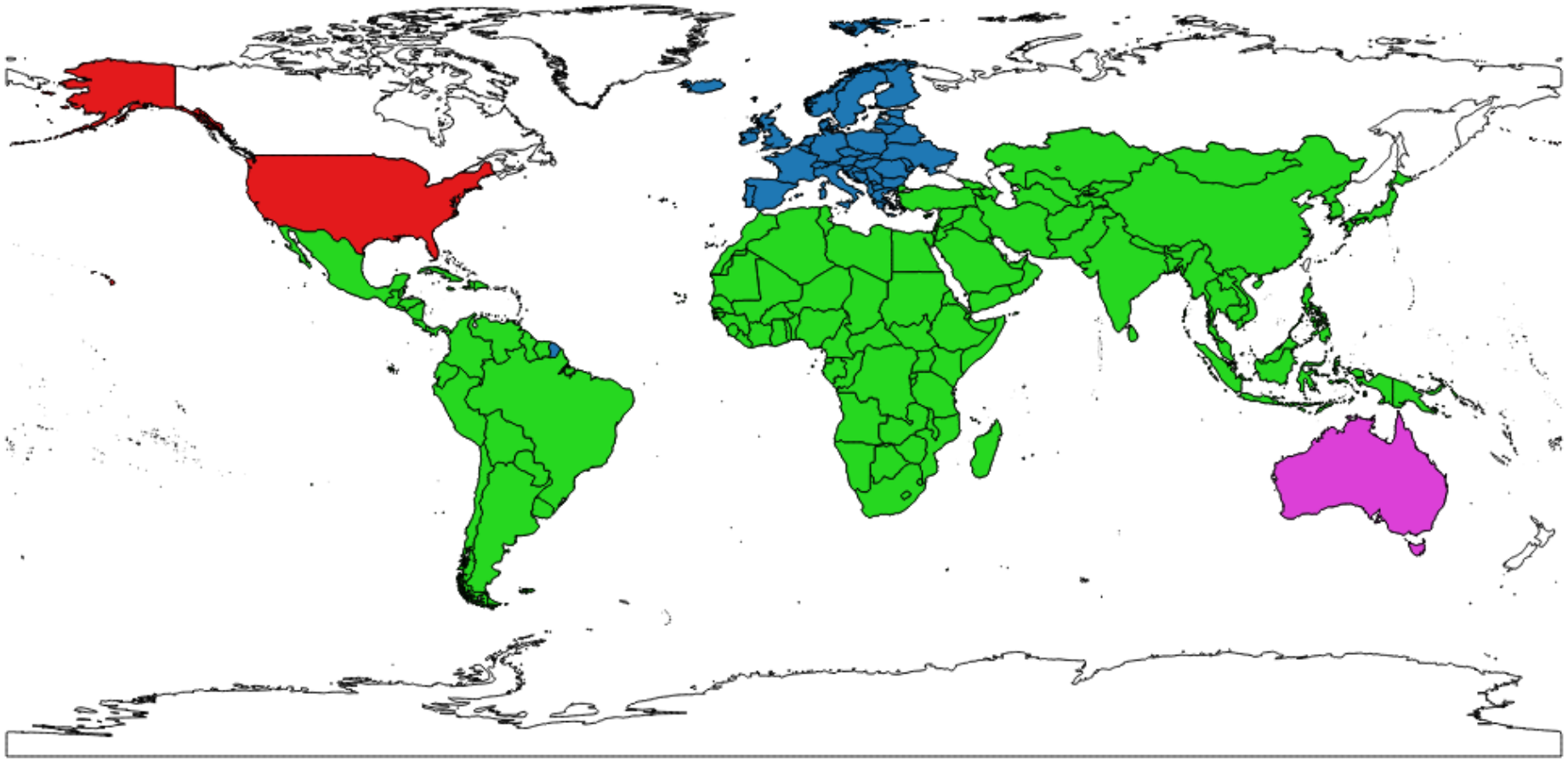
Anna Dmowska, Tomasz Stepinski

dmowskaa@ucmail.uc.edu, stepintz@ucmail.uc.edu

Application of population grid

- Availability of high resolution population distribution data is a sought after information in **academic research** as well as in **practical applications for policy and buisness purposes.**
- Applications include:
  - Administrative decision making process
  - Response to natural disasters
  - Public health
  - Human pressure on the environment

Existing population grids



Worldpop (100 m)
European grid (100 m)
Austrialian grid (1 km)
SEDAC (30" ~ 1km)
LandScan (90 m)
SocScape grid (90 m)

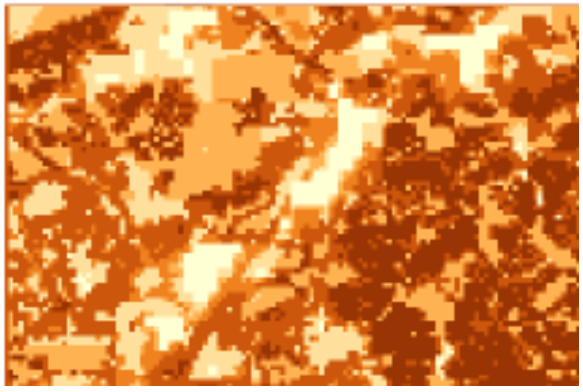
Our projects aims at making high resolution population grids available for the entire U.S.

Existing population grids for the entire U.S.

SEDAC

LandScan -USA

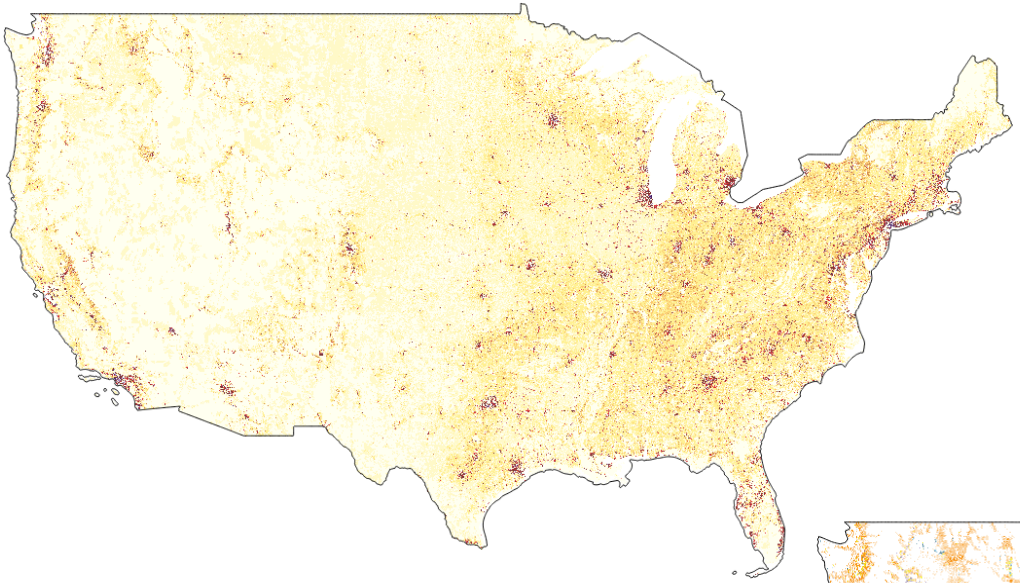
SocScape -USA



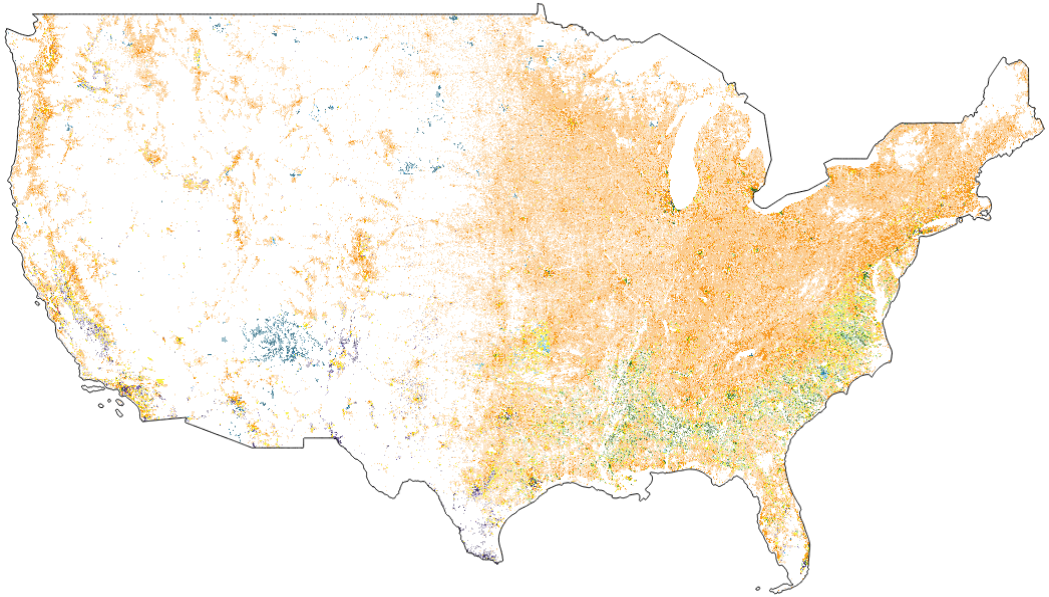
SEDAC	LandScan USA	SocScape USA
Socioeconomic Data and Application Center at Columbia University	The Oak Ridge National Laboratory	Space Informatics Lab, University of Cincinnati
1 km spatial resolution (MSA – 250 m)	90 m spatial resolution	90 m spatial resolution
Simple areal weightening interpolation from census blocks	Disaggregates census block using dasymetric modeling	Sharpening using dasymetric modeling
Available form SEDAC webpages	Not available in public domain	Available trough web-based application - SocScape

Population grids for the entire U.S.

Population grid



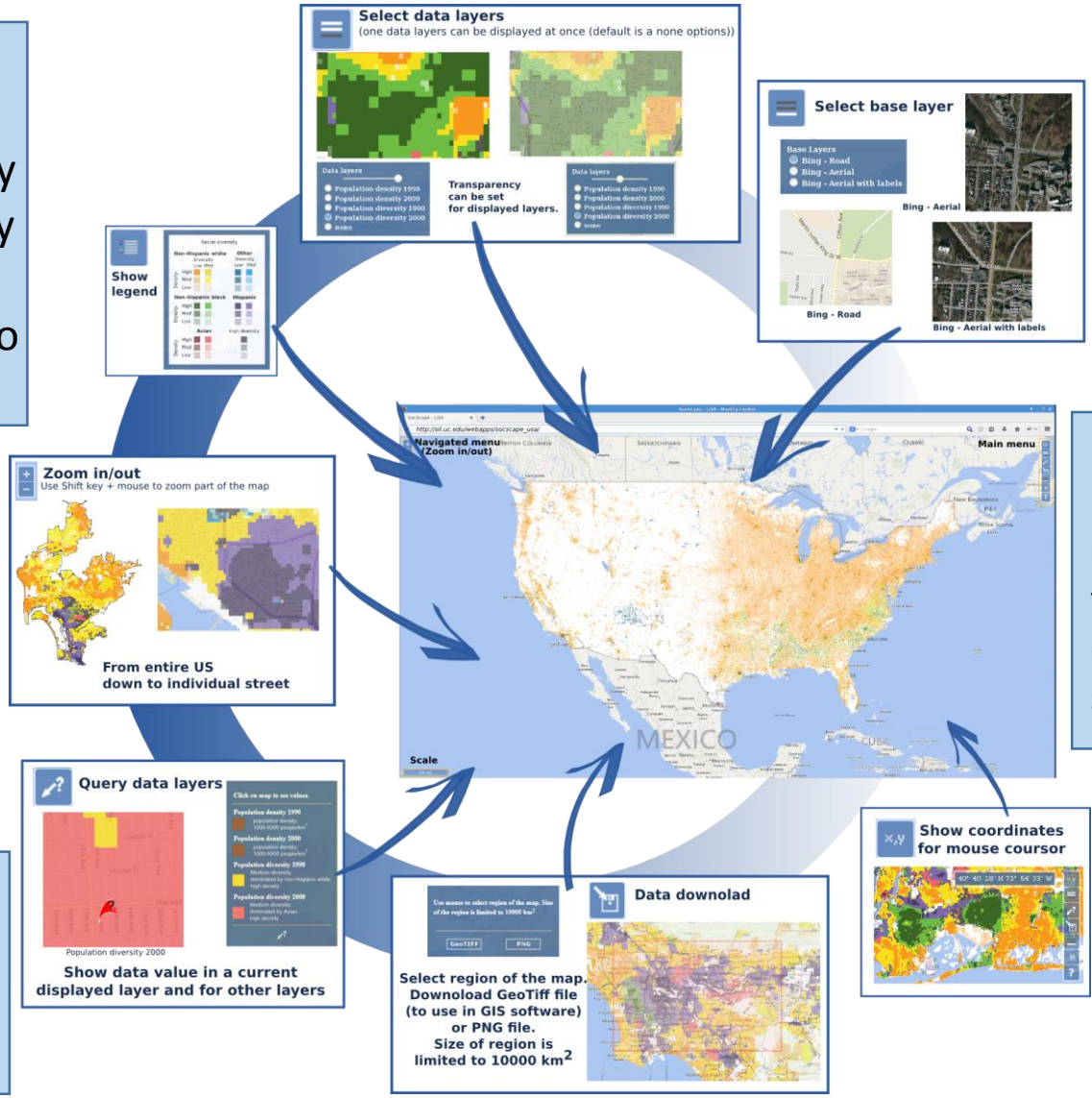
Racial diversity grid



SocScape - exploring population grids online

http://sil.uc.edu/webapps/socscape_usa/

Fast and intuitive exploration of population density and racial diversity in different scales (from U.S. down to the street)

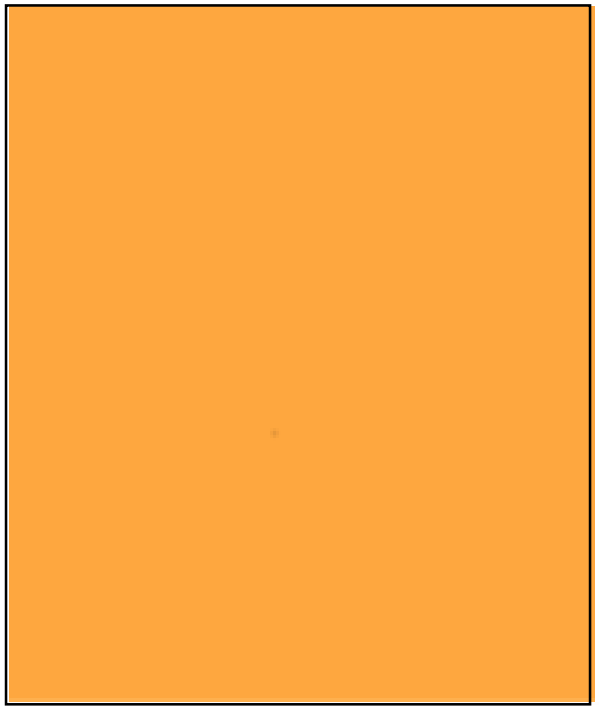


Downloading population, racial diversity grids for future analysis in external GIS software

Detecting change of spatial dynamics of population density and racial diversity.

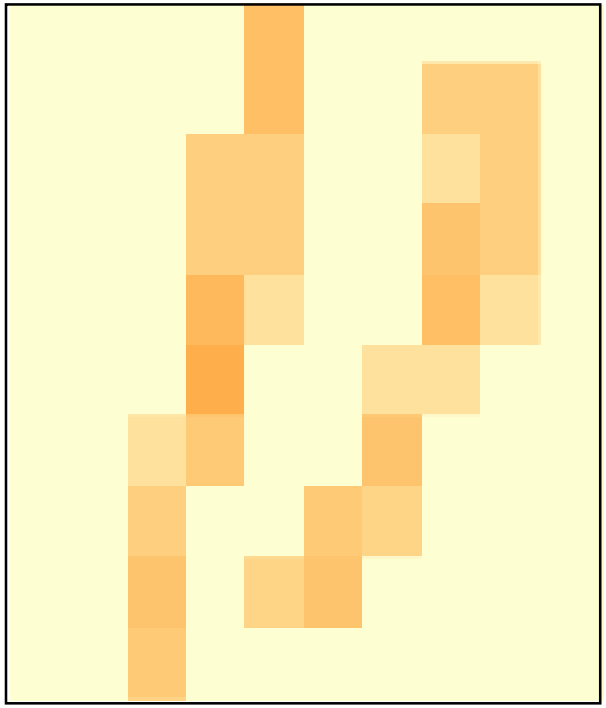
Dasymetric model

Aggregated population data



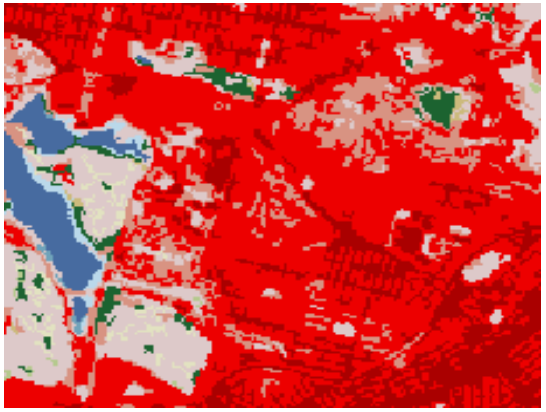
Ancillary data

Dasymetric model

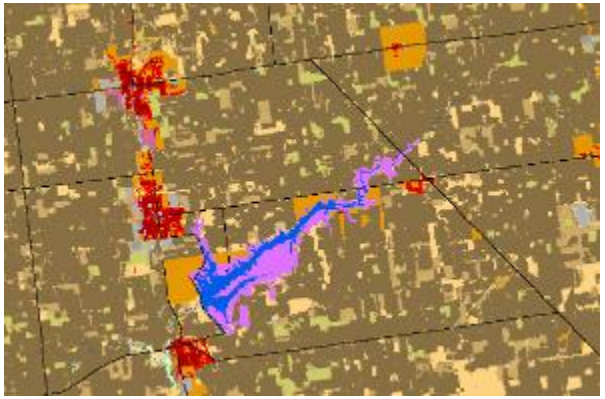


Dasymetric modeling refers to a process of disaggregating (sharpening) spatial data to a finer unit of analysis, using additional (or ancillary) data to help refine locations of population or other phenomena (Mennis 2003).

Dasymetric model - ancillary data



Land cover map from satellite imaginary



Land use map

Land cover data - available for the entire United States in uniform fashion and quality.

Land use data – advantage over LC data, until recently not available for entire U.S

Street network

Parcel data

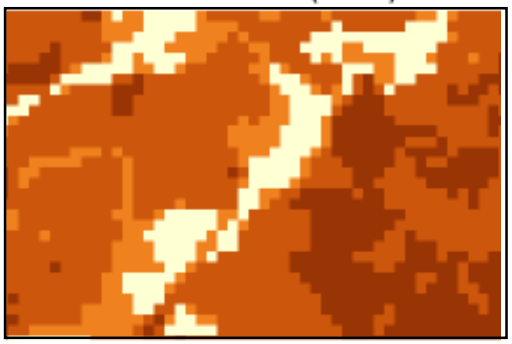
Available for local areas and combining them together is time consuming and may lead to a product of questionable quality.

Dasymetric model based on SEDAC grid

INPUT DATA

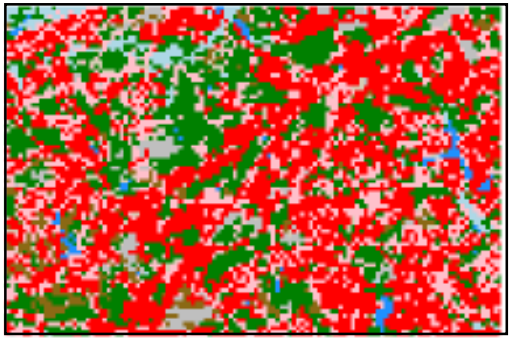
US-SEDAC (30")

MSA-SEDAC (7.5")



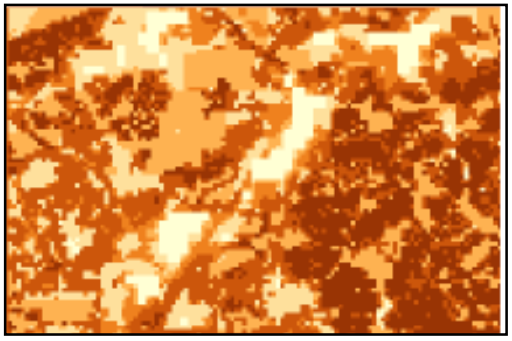
ANCILLARY DATA

NLCD (1")



HI-RES MAP

US-HI-RES (3")



Disaggregate existing 30" (~1km) SEDAC population grids

Dasymetric modeling with National Land Cover Dataset as ancillary data.

High resolution population map

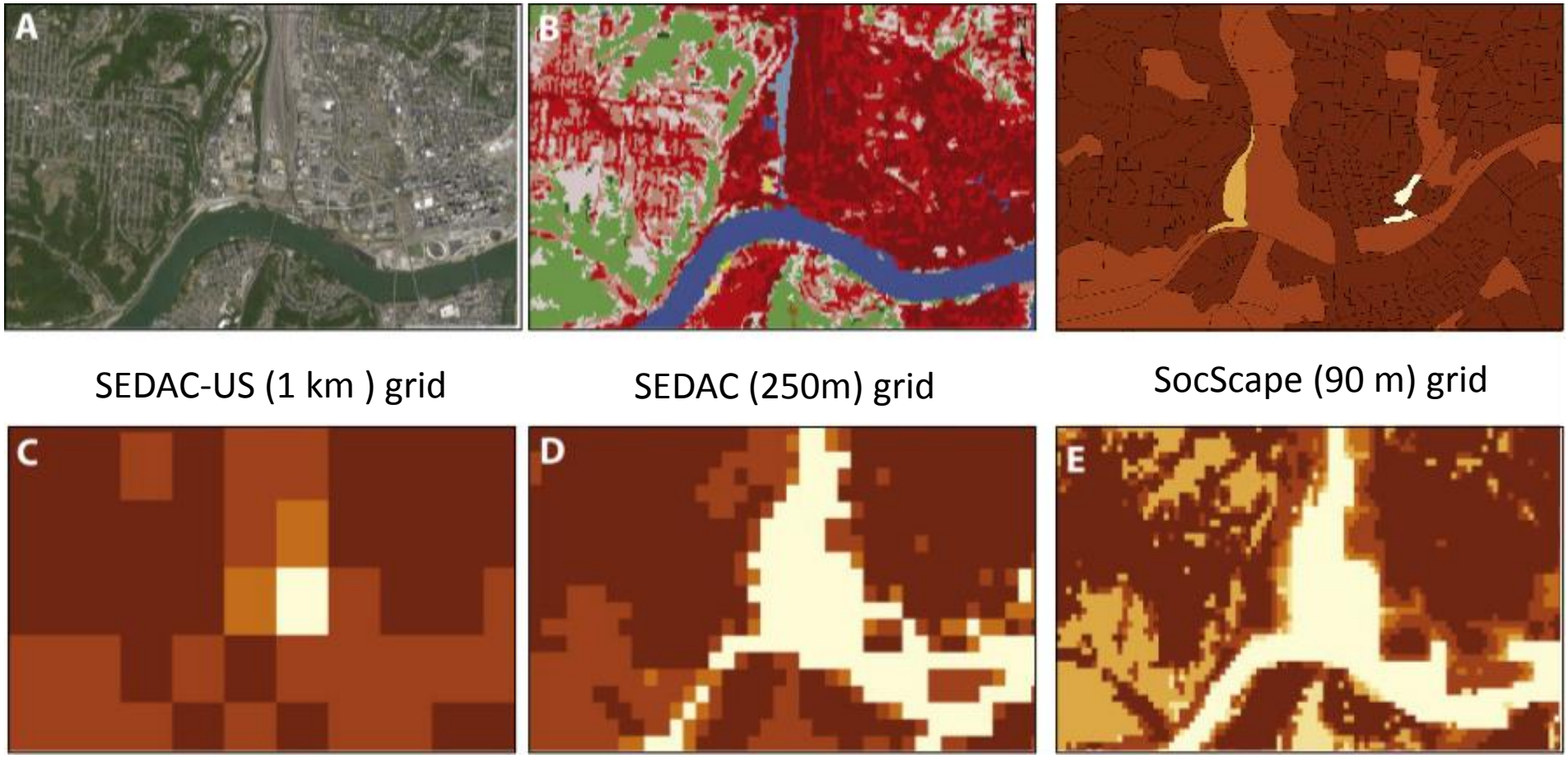
Advantage: Less compuationally demanding than disaggragating directly from census blocks. The computational efficiency comes from working exclusively with grids

Dasymetric model based on SEDAC grid

Satellite image (Google maps)

Land cover map (NLCD2001)

U.S. Census block group



SEDAC-US (1 km) grid

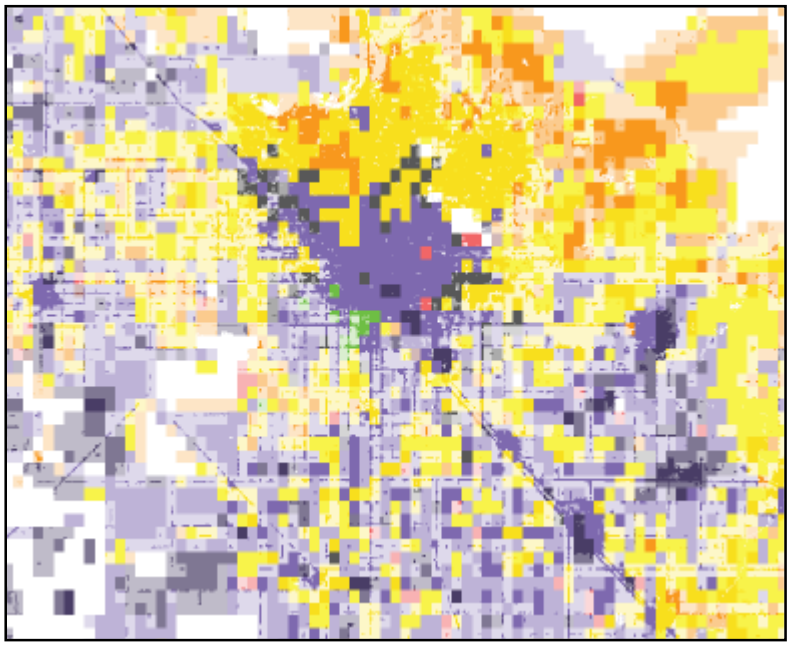
SEDAC (250m) grid

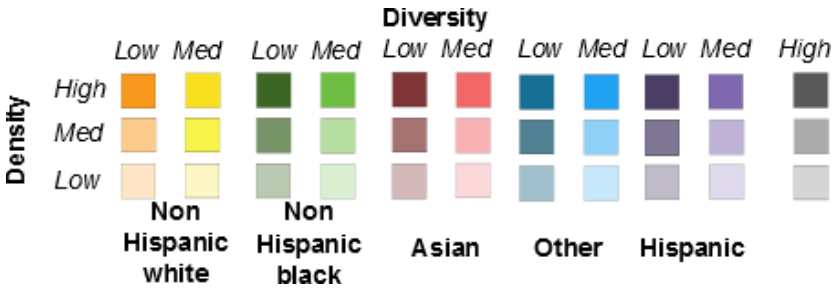
SocScape (90 m) grid

Downtown Cincinnati, Ohio.

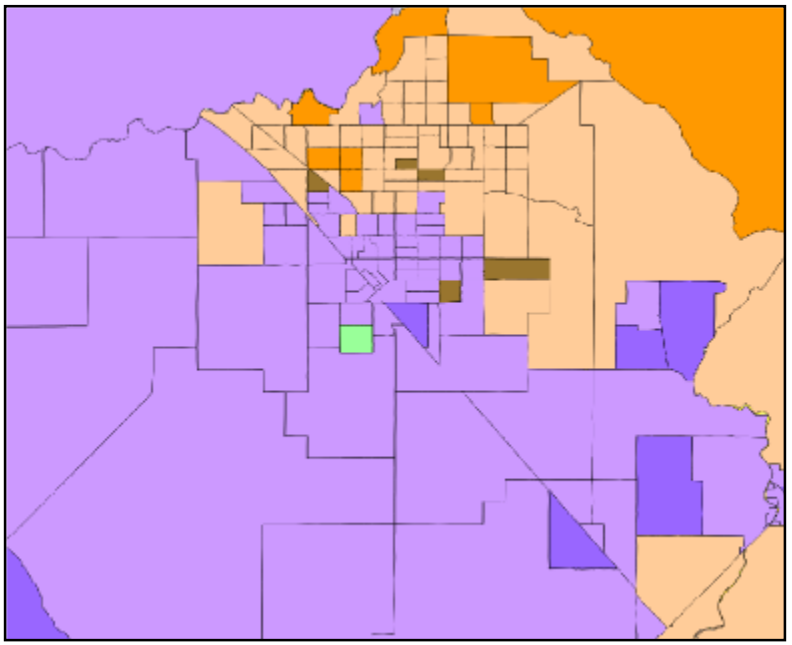
Racial diversity map

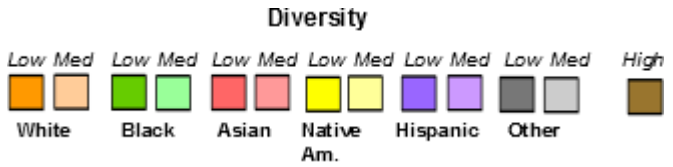
SocScape





Mixed Metro





Fresno,CA

SEDAC vs. Block-based approach

SocScape grid based on SEDAC

SocScape grid based block-level data

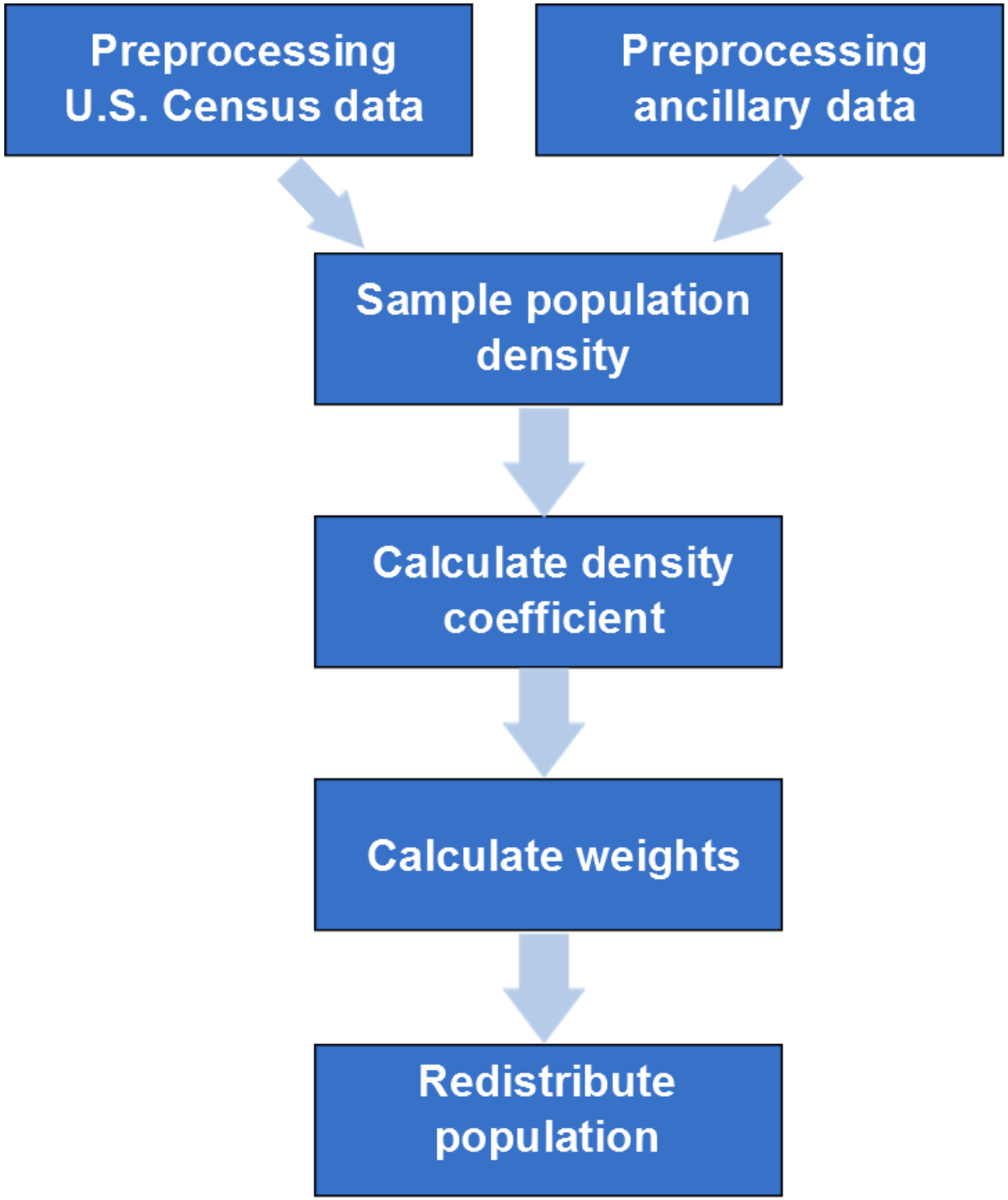
Our work on developing demographic grids for the entire United States is continuing as we switch from sharpening the SEDAC grids to a direct disaggregation of census blocks into 30 m resolution grids.

SEDAC vs. Block-based approach

SEDAC grid	U.S. Census Block level data
Available only for 1990 and 2000	Available for each Decenial Census
Available only for selected variables	Much more variables to analyze
Contains many errors	No errors
Coarser spatial resolution in urban areas	Better spatial resolution in urban areas
Less computational expensive	More computational expensive

The price of performing dasymetric modeling based on block level data is a significantly larger computational cost.
However, by using some innovations in a dasymetric modeling procedure and by having access to sufficient computational resources we have been able to keep computational cost under control.

Dasymetric model based on block level data



Software

GRASS GIS v.7.0

Python

Python scripts for GRASS GIS

All computations were performed using Python scripts written for GRASS GIS 7.0 software which is especially well adapted to work with large volume of data.

Performance

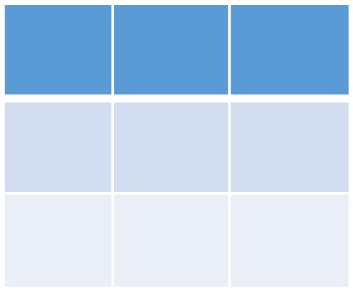
Data	Size of files
Nb. of blocks in U.S.	10 965 243
Nb of blocks per state	6426 for DC to 912 889 for TX
Size of shapefiles for states	39MB - 4037MB
Overall size of shapefiles	39GB
Size of output map	139 GB
Size of output map in cells	16 832 104 560 (no-null: 8 651 173 750)

Processing steps	Calculation time
Preprocessing data	17 h 40 min
Calculating weights	41 h
Calculate dasymetric model	7 h
Overall time	65 h 40 min

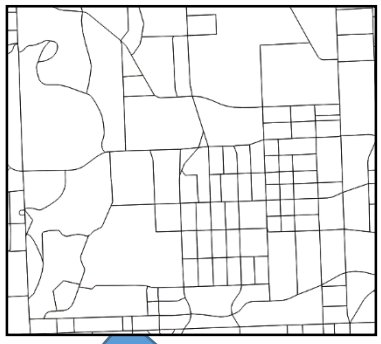
All calculation was done using a PC computer with Intel 3.4GHz, 4-cores processor and 16 GB of memory running the Linux system.

Preprocessing U.S Census data

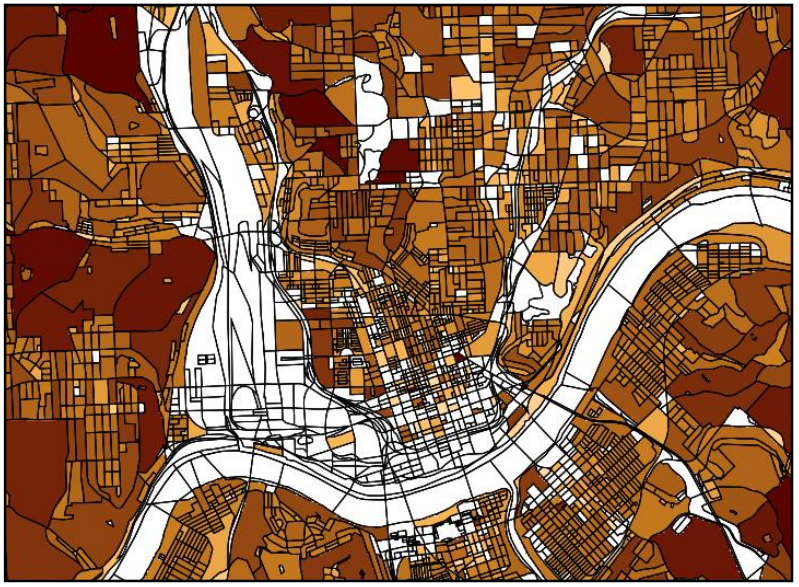
Tabular data



Shapefile geometry

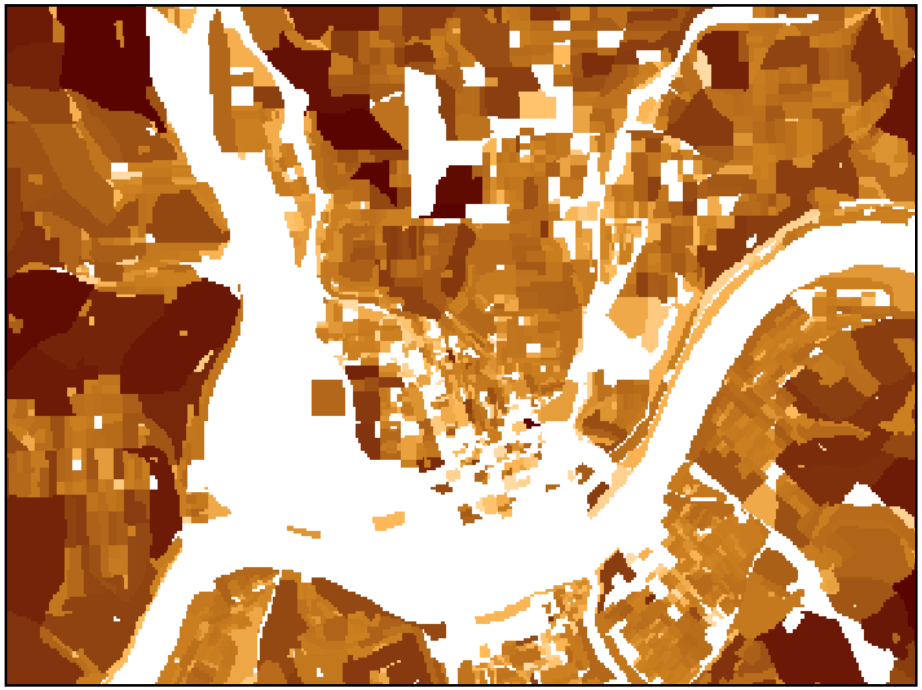


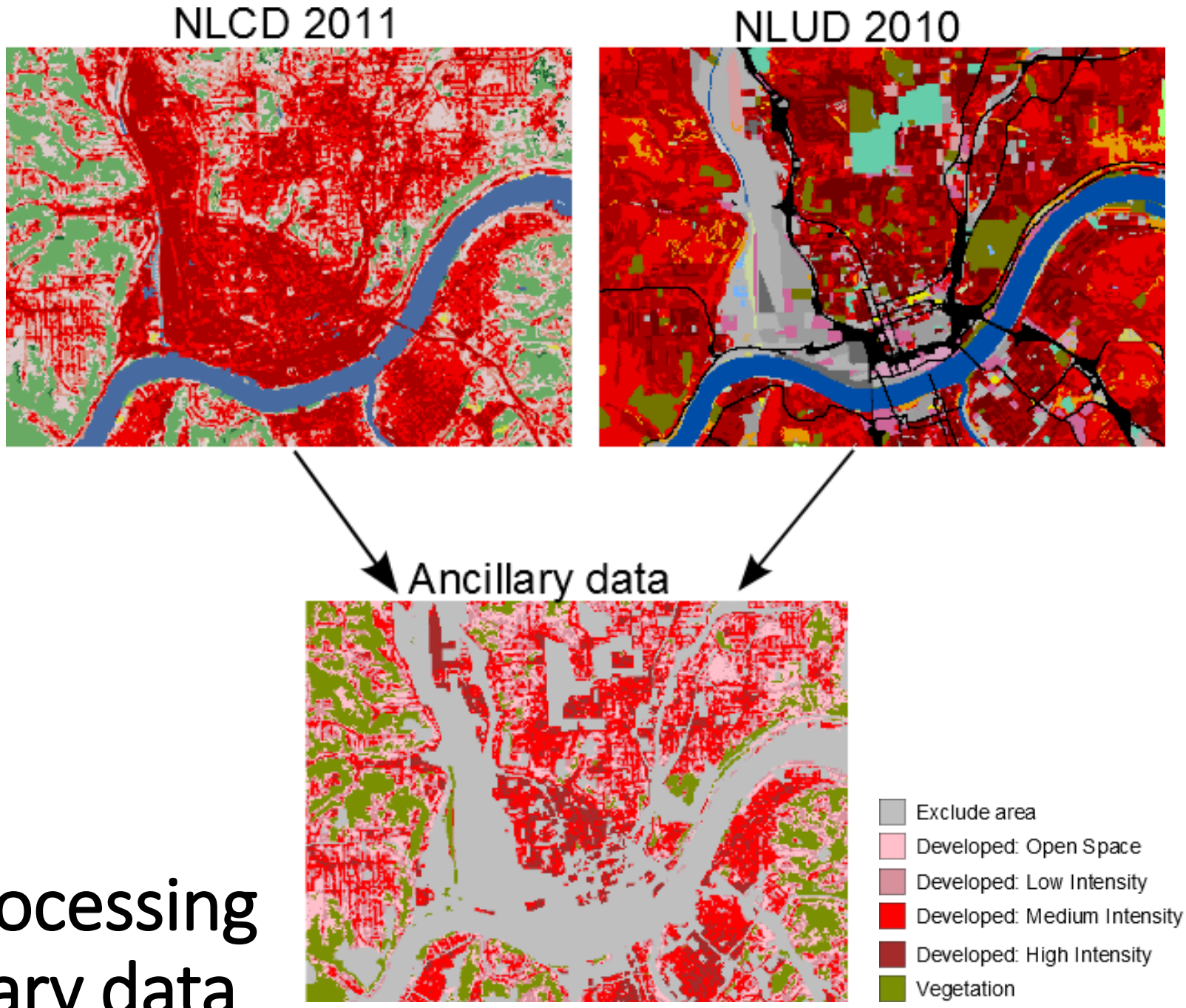
Attribute vector file



Convert to raster

Raster block-level data

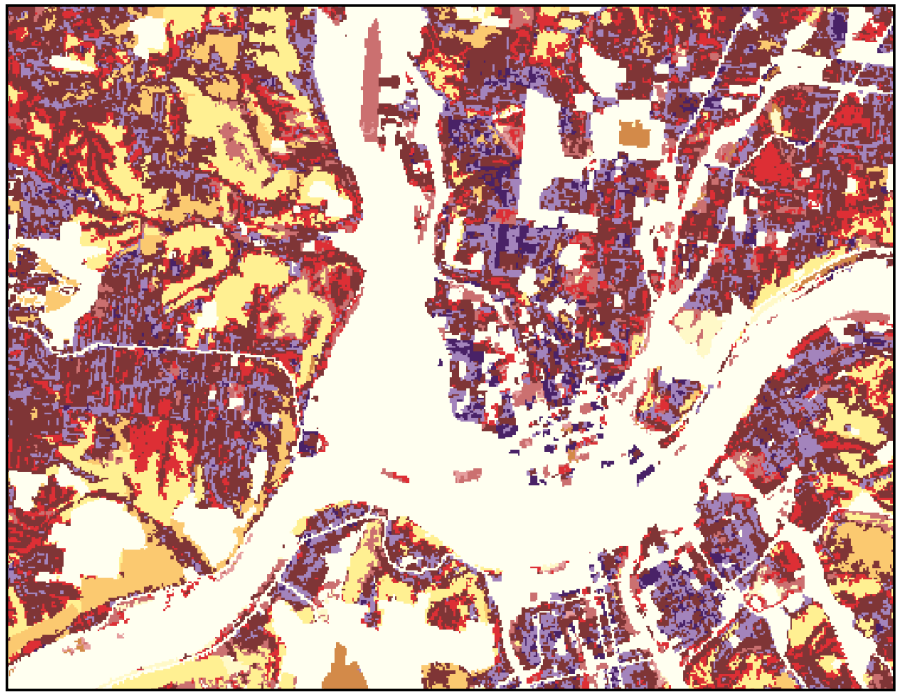




Preprocessing ancillary data

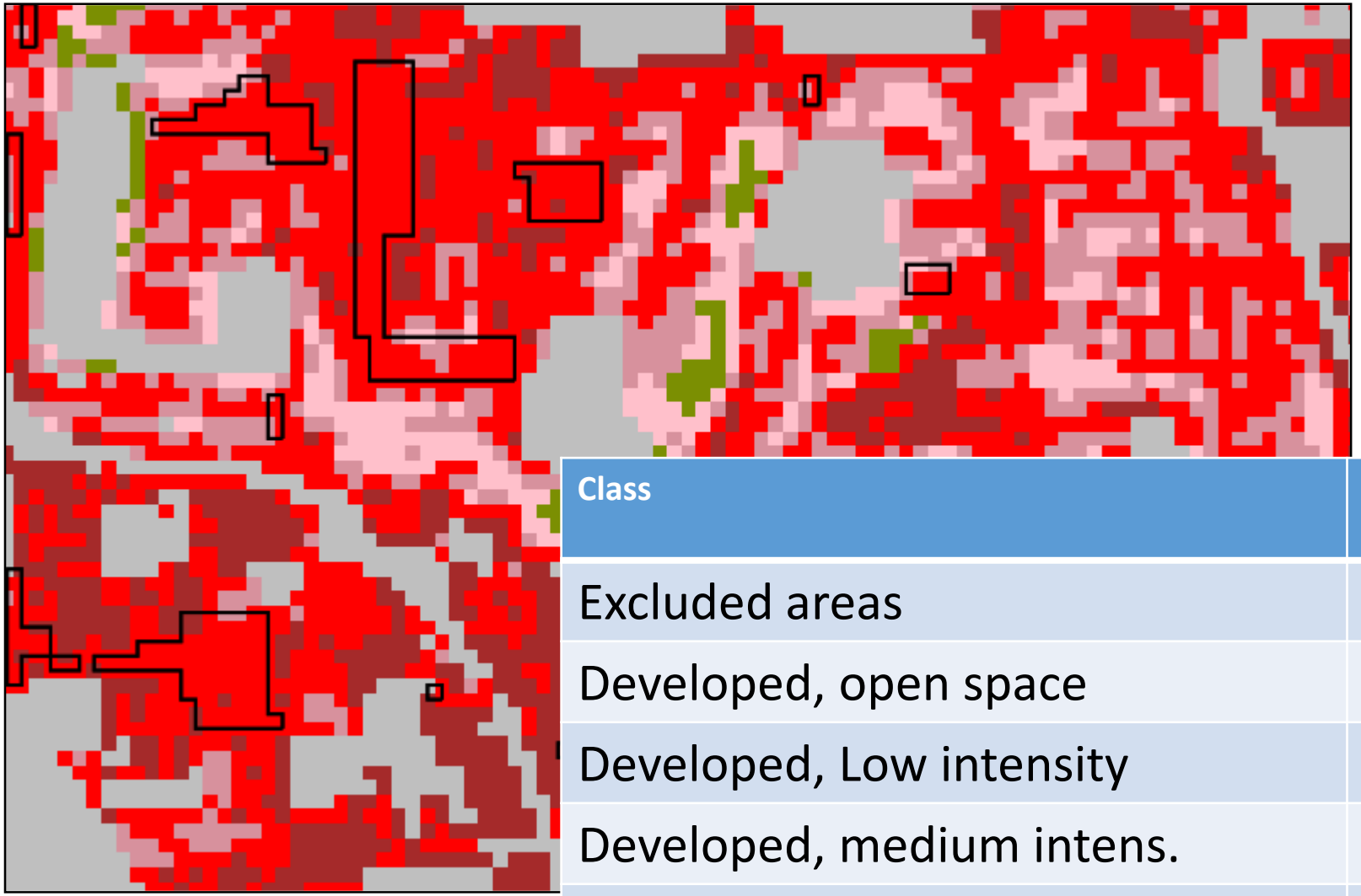
Dasymetric model

Hi-res population grid (30 m)



- The population in each (rasterized) block is redistributed to its cells using block-specific weights.
- The weights are assigned based on two factors:
  - relative density of population for each class,
  - the area of each block occupied by each class (Mennis, 2003).
- Population in each cell = Nb. of people in block × weights

Sample population density



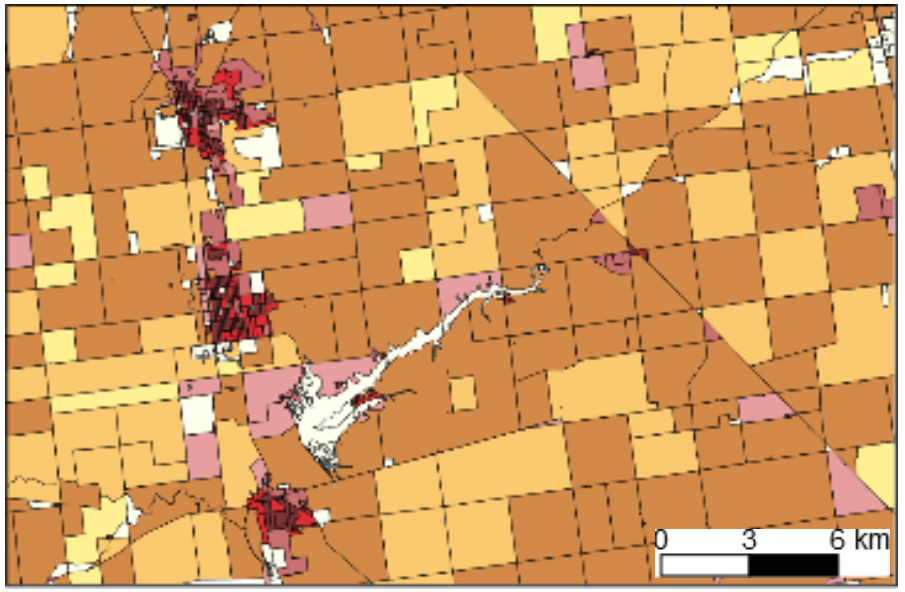
Class	Threshold [%]
Excluded areas	100
Developed, open space	90
Developed, Low intensity	90
Developed, medium intens.	90
Developed, high intensity	90
Vegetation	95

Sample population density

Class	#blocks	Population	Area [km^2]	Pop. Dens. [pop/km^2]	Relative density [%]
Excluded areas	1,695,801	0	253,595	0	0
Developed, open space	52,836	733,893	978	751	4.27
Developed, Low intensity	272,698	5,899,513	2,795	2210	12.56
Developed, medium intens.	161,698	8,257,974	1,767	4672	26.56
Developed, high intensity	35,400	3,109,105	312	9953	56.59
Vegetation	2,058,828	11,809,104	3,571,846	3	0.02

Example of map

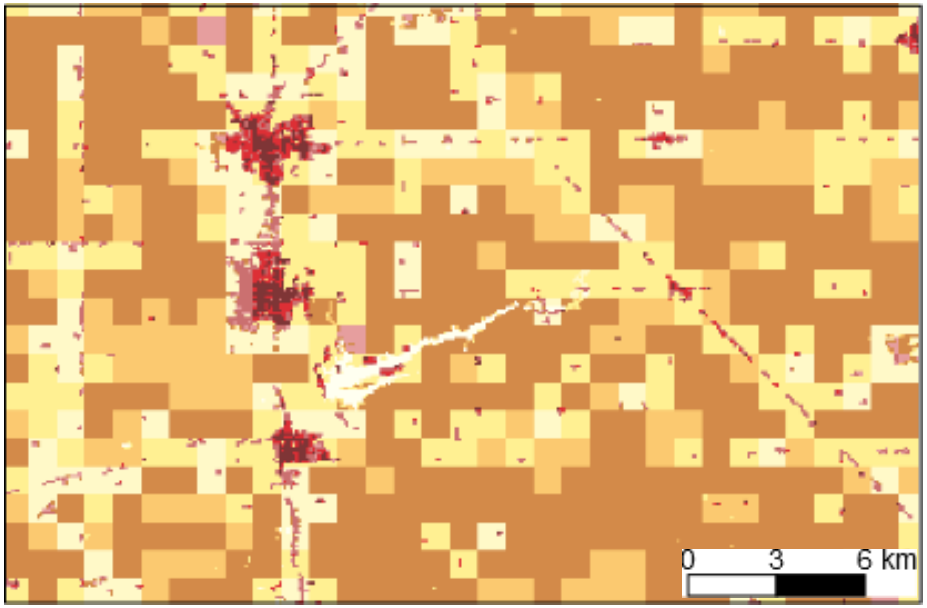
Block-level population density



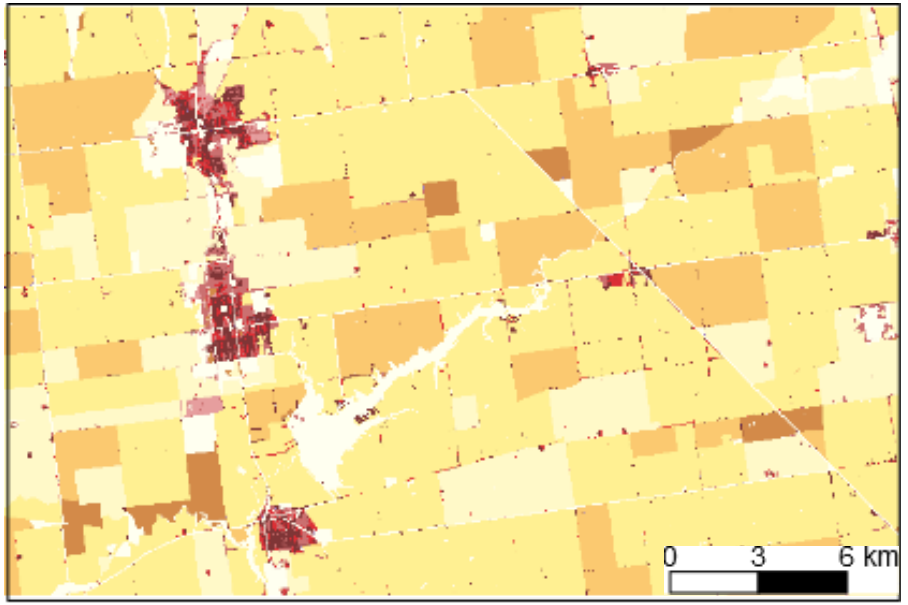
People/km^2

0	5-10	100-500	5000-10000
0-1	10-50	500-1000	> 10000
1-5	50-100	1000-5000	

SocScape grid based on SEDAC

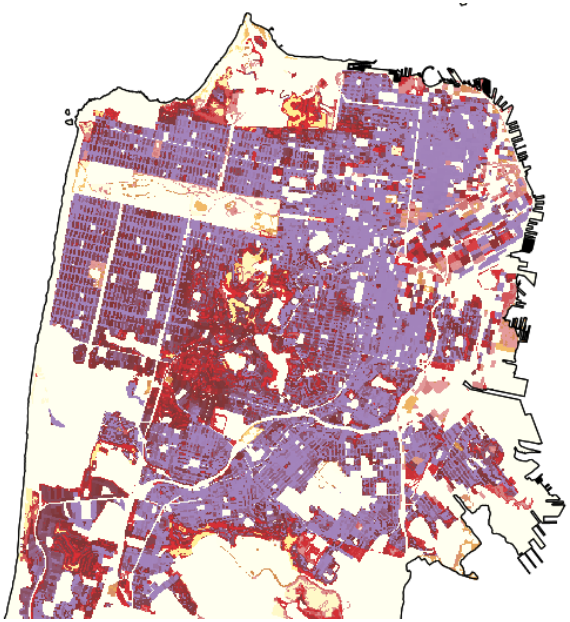


SocScape grid based on block-data

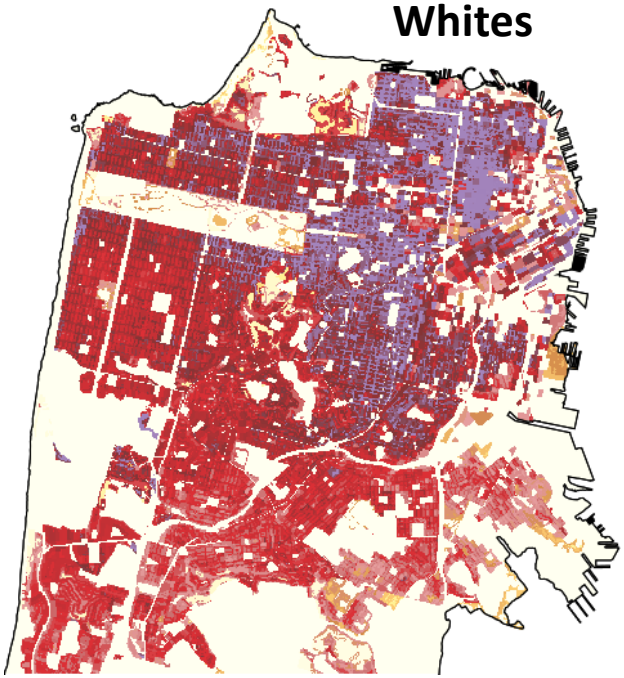


Fort Loramie-Minster-New Bremen, OH

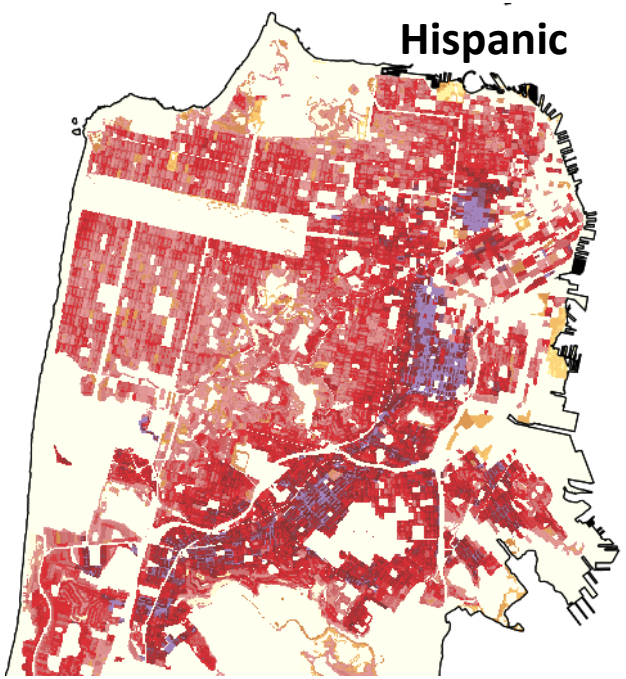
Distribution of racial/ethnicity



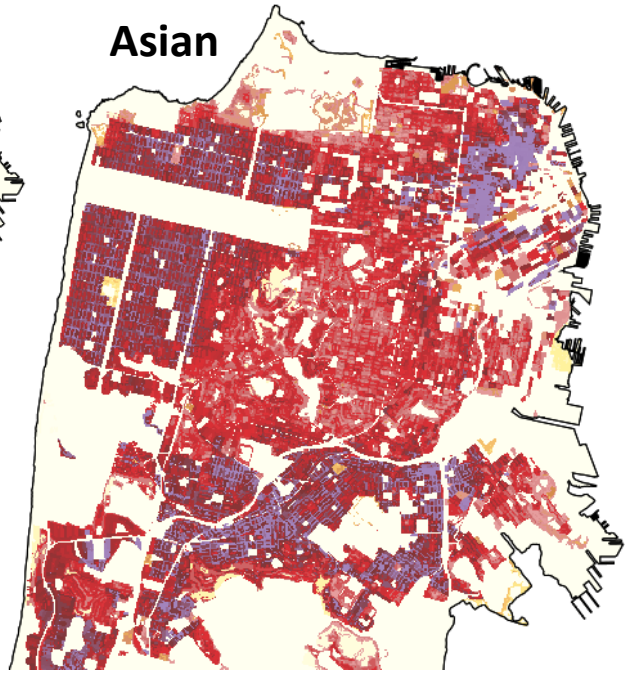
Total population



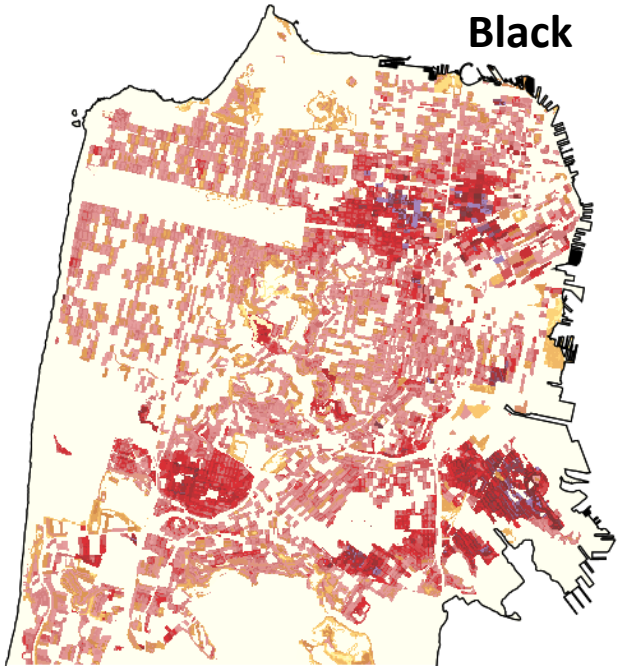
Whites



Hispanic



Asian



Black

People/km^2

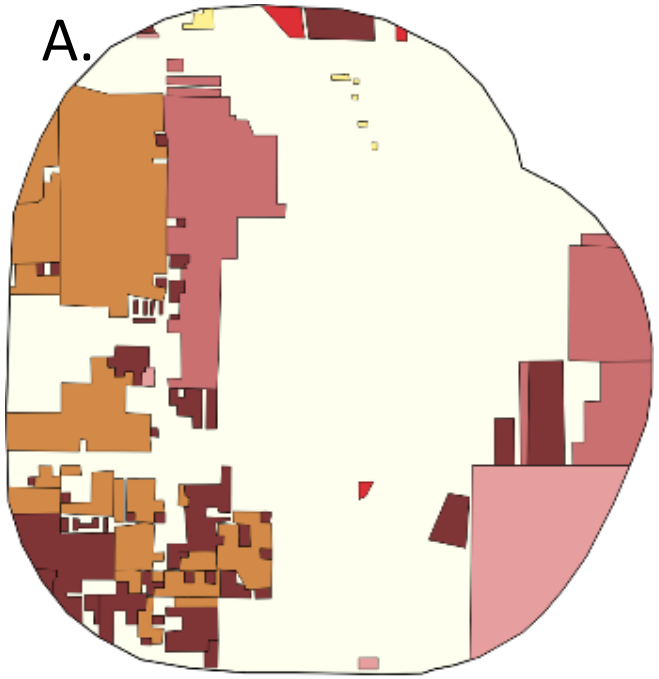
0	5-10	100-500	5000-10000
0-1	10-50	500-1000	> 10000
1-5	50-100	1000-5000	

Comparison to other case study

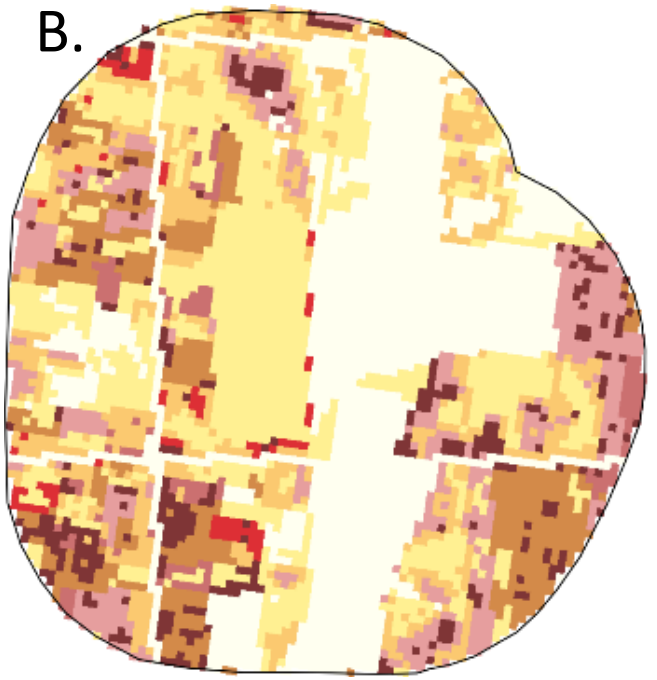
Population map based on:

A. Parcel data
B. Land cover/land use
C. Land cover data

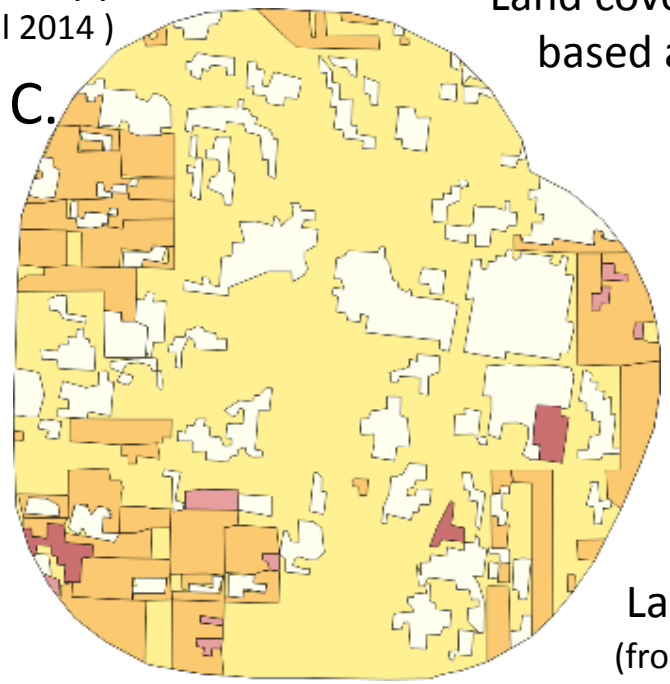
Gainesville, Alachua County, Florida



Parcel data based approach
(from P. Jia et.al 2014)



Land cover/land use based approach



Land cover data
(from P. Jia et.al 2014)

Conclusions

- Applying well-known dasymeric methods to „big data”
- Improvement to previous grid:
  - Using land use map as ancillary data
  - Disaggregating population counts directy from blocks instead of coarse population grids.
  - Developed efficient algorithms
- Calculate also other demographic grids (i.e race/ethnicity) using weights establish to population grids.