Assessing Flood Risk to Support Environmental Justice: Data and Tools from NASA's SEDAC

Greg Yetman, Kytt MacManus, Alex de Sherbinin and Juan Martinez

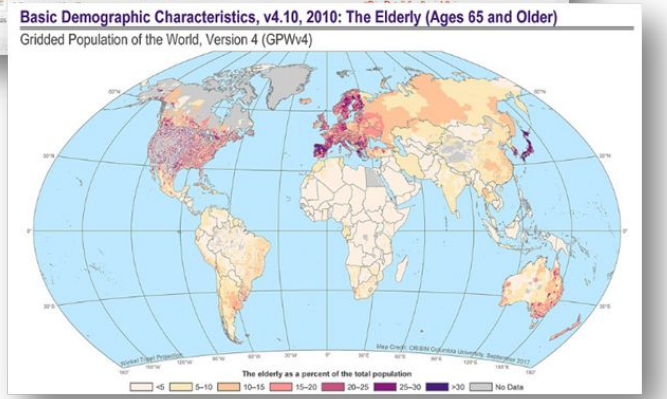
NASA Socioeconomic Data and Applications Center (SEDAC)
Managed by CIESIN at the Columbia Climate School

COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK

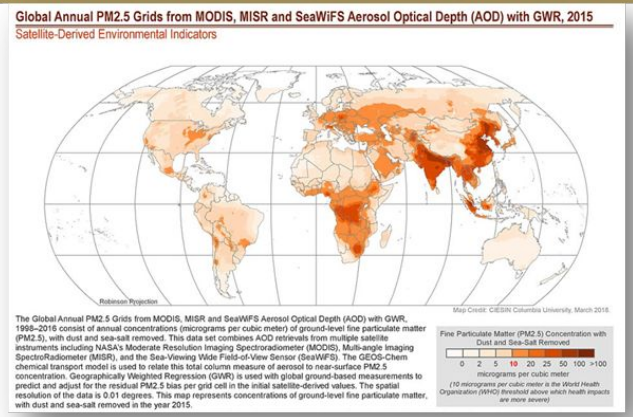
What is SEDAC?

- SEDAC is one of 12 NASA Distributed Active Archive Centers under the Earth Science Data and Information System (ESDIS)
- SEDAC has been managed by the Center for International Earth Science Information Network (CIESIN) at Columbia University since 1998
- CIESIN is based at Columbia's Lamont Campus in Palisades NY
- SEDAC has a staff of social and natural scientists, GIS specialists, data scientists, data management experts, programmers, and systems engineers

What do we do?



- SEDAC data provide the ground level context for NASA's remote sensing data
- Focus on human dimensions of environmental change
- Big emphasis on data integration
- Direct support to scientists, applied and operational users, decision makers, and policy communities
- Strong links to geospatial data community



Presentation Outline

- What is Environmental Justice
- SEDAC's efforts to advance equal opportunities
  - City University of New York (CUNY) Lehman College
  - NASA Applied Remote Sensing Training Program (ARSET)
  - Group on Earth Observations Human Planet
- SEDAC Datasets Tools and Applications
  - Low Elevation Coastal Zone, v3
    - POPGRID Viewer demonstration
    - Key statistics for the world and the US
  - U.S. Social Vulnerability Index Grids with Population, Landscape, And Climate Estimates (PLACE), v4
    - Social vulnerability in the 5km coastal buffer
  - NY State Flood Information Decision Support System Mapper demonstration
  - Global Gridded Relative Deprivation Index (GRDI), v1
- Q and A

What is Environmental Justice?

Environmental Protection Agency (EPA) definition: environmental justice is the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies.

Related Concept

Climate justice is an acknowledgement that climate change can result in poor and marginalized communities having a disproportionate exposure and vulnerability to environmental impacts, particularly in developing countries.

How does NASA contribute

Providing scientific data and analysis across domains with direct implications for Environmental and Climate Justice:

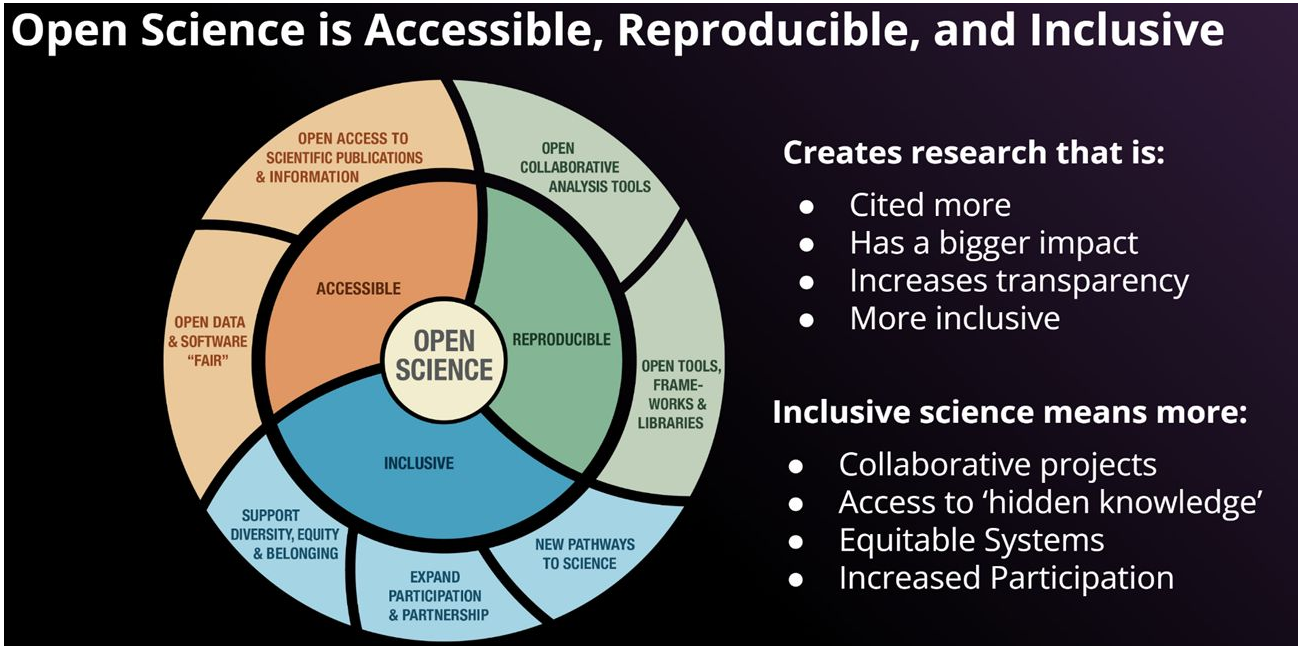
1. Air Quality
2. Disaster Preparedness and Response
3. Extreme Heat
4. Water Availability
5. Flooding

For more information see: https://www.earthdata.nasa.gov/topics/human-dimensions/social-behavior/environmental-justice

Environmental Justice and Open Science

NASA Open Science Initiatives like the Transform to Open Science (TOPS) mission seek to inspire and empower scientists, researchers, and communities to embrace open science as a catalyst for positive change.

https://github.com/nasa/Transform-to-Open-Science



SEDAC's Efforts to Advance Equal Opportunities

Columbia Climate School
Center for International Earth Science Information Network

Advancing Data for Street-Level Flood Vulnerability: Characterizing Vulnerability of Buildings

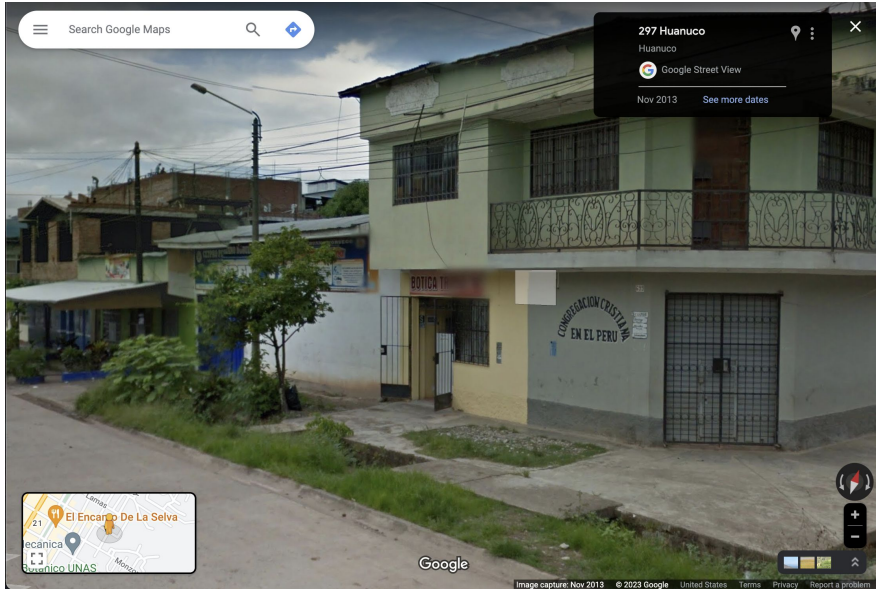
CUNY Lehman Professor, Yuri Gorokhovich leads students as they present research results at the 2022 American Geophysical Union's annual meeting.

Sample of locations from cities in (primarily) low-income countries

Data set that will be released by SEDAC in 2024: Built infrastructure using sampling strategies

Sampling from Google StreetView to Characterize Vulnerability

- For areas without detailed infrastructure inventories, can we estimate vulnerability from samples?
- A random sample of urban areas can be used to estimate prevalence of vulnerabilities and to characterize different land use categories.



Approach

1. Survey design
2. Select sample sites
3. Setup Mechanical Turk
4. Data collection and analysis

Survey Design/Site Selection

- Define variables
- Standardize through codebook
- Develop template task script of survey variables
- Select sites (random locations)

Variable	Entry Options
Building Material	Masonry, cinder block, wood, brick, steel
Roof Type	Corrugated metal, tile, thatched or pal leaves, tar, gravel, concrete, mixed
Street Type	Asphalt, cobble or cement, dirt or gravel, potholed
Land Use	Residential, Commercial, Natural, Agricultural, mixed
Structure Type	Detached, semi-detached, attached, not applicable
Occupancy status	Occupied, vacant, cannot determine occupancy
Sill Height (lowest entry point)	Ground level, low (1-6”), medium (7-12”), high (12-18”)
Building Condition	Very poor, poor, fair, good with minor defects, very good
Street Topography	Flat or low slope, medium slope, steep slope
Floors	Number of floors
Drains	Number of drains (street)

What is mechanical turking?

Crowdsourcing method to break down tasks into simpler components, then distribute these tasks to many individuals.

localturk

Local Turk implements Amazon's Mechanical Turk API on your own machine.

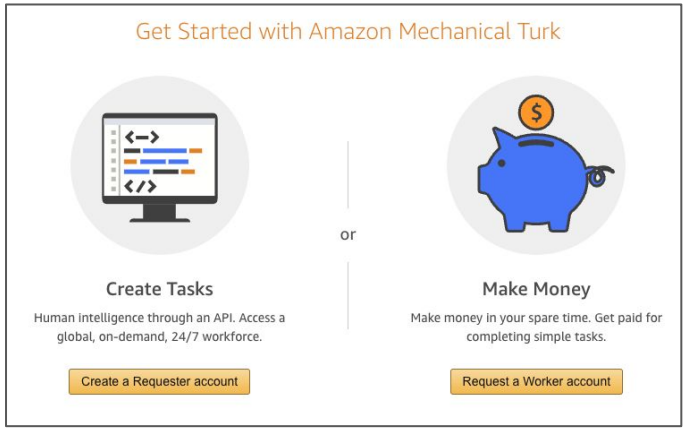
It's handy if you want to:

1. Develop a Mechanical Turk template
2. Do some repetitive tasks on your own, without involving Turkers.

You could use it, for instance, to generate test and training data for a Machine Learning algorithm.

Why use it?

Data-collection via many repetitive tasks can be done rapidly and remotely at a low cost.



Set Up Mechanical Turk

- Export remote sensing optical images for each point and load in web accessible folder (e.g., Amazon Web Services “bucket”)
- Install Local Turk on local or accessible machine
  - Alternatively, use Amazon paid service.
- Run module with template task script

Building / Structure Information

Select a Sill Height

○ None, Ground Level ○ Low, 1-6" ○ Medium, 7-12" ○ High, 12-18" ○ Not Applicable

☐ Detached ☐ Semi-detached ☐ Attached ☐ Not Applicable

1 Number of Floors

Building Condition, Status and Material

○ Very Poor ○ Poor ○ Fair ○ Good with Minor Defects ○ Very Good

☐ Under Construction ☐ Masonry or Cinder Block ☐ Wood Construction ☐ Brick ☐ Steel

○ Occupied ○ Vacant ○ Cannot determine occupancy

Roof Type

☐ Corrugated Metal ☐ Tile ☐ Thatched or Palm Leaves ☐ Tar ☐ Gravel ☐ Other: Additional Notes

Land Use

☐ Residential ☐ Commercial ☐ Agricultural ☐ Industrial ☐ Natural ☐ Other:

Street Information

○ Flat or Low Slope ○ Medium Slope ○ Steep Slope

0 Number of Drains Visible

☐ Paved Street (Asphalt) ☐ Cobble or Cement Blocks ☐ Dirt or Gravel ☐ Potholes ☐ Other: Additional Notes

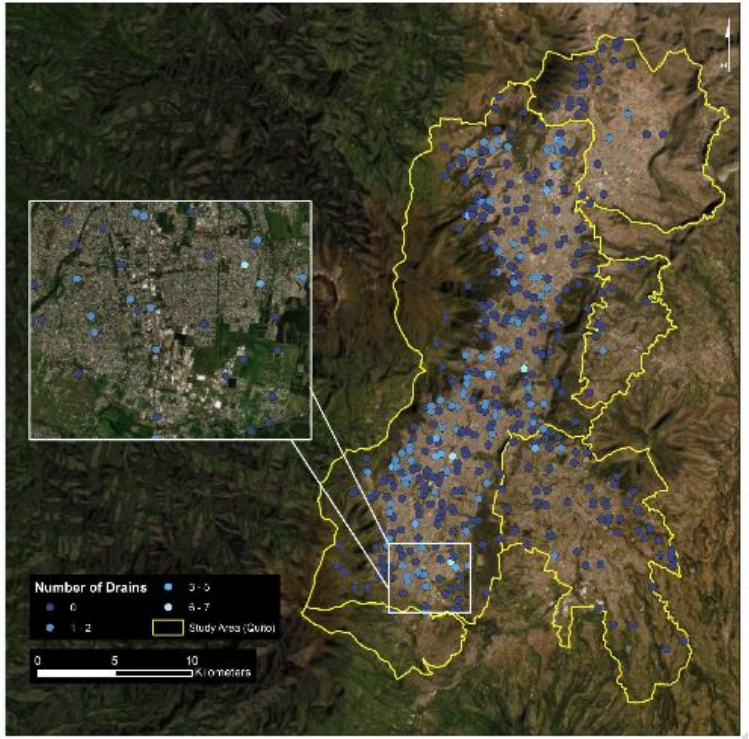
Add any notes here

☐ Google StreetView Not Available

Submit

Data Collection and Analysis

- Use Mechanical Turk to collect data for each location
- Evaluate data collected for each variable
  - Review consistency across analysts
- Visualize results in context



Results and Conclusions

- Flood vulnerability variables for hundreds of buildings can be collected in just a few hours.
- Uncertainty of output should be measured.
  - Easy for different analysts to interpret instructions differently; examples are useful
  - Not always consistent between analysts
- Ongoing work: using ML/AI to categorize vulnerability for large areas

Variable	Kappa (κ)
Building Material	0.94
Roof Type	0.83
Street Type	0.87
Land Use	0.88
Structure Type	0.64
Occupancy Status	0.80
Sill Height	0.66
Building Condition	0.54
Street Topography	0.79

Data set that will be released by SEDAC in 2024: Built infrastructure using sampling strategies

Great partnership with great outcomes

Team of students from Lehman College helped SEDAC to achieve HBCU/MSI spending goal

Paper published in IEEE journal
https://doi.org/10.1109/OJCS.2022.3166887

Two students hired by NASA MSFC; one by Census Bureau

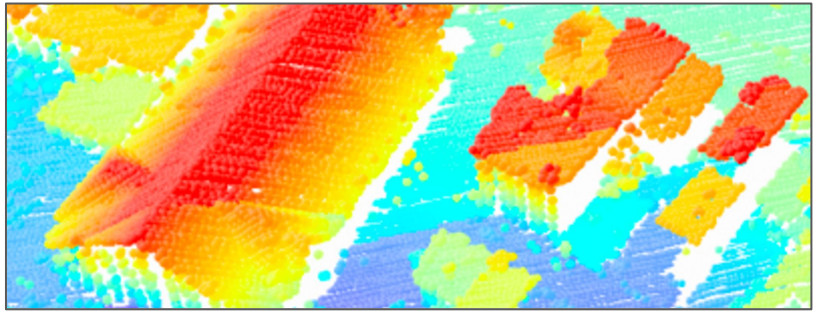
LEHMAN COLLEGE

Supporting the NASA Applied Remote Sensing Training Program (ARSET)

SEDAC has contributed modules to several ARSET training sessions:

- Earth Observations Toolkit for Sustainable Cities and Human Settlements
- Transforming Earth Observation (EO) Data into Building Infrastructure Data Sets for Disaster Risk Modeling
- Satellite Remote Sensing for Measuring Urban Heat Islands and Constructing Heat Vulnerability Indices

ARSET trainings are multilingual:
http://appliedsciences.nasa.gov/arset



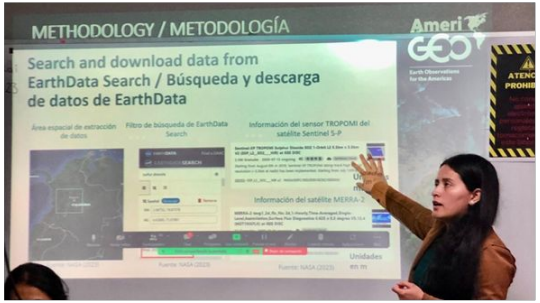
GEO GROUP ON EARTH OBSERVATIONS

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
Acreditación Institucional de Alta Calidad

HUMAN PLANET

Human Planet Course

- District University of Francisco Jose de Caldas in Bogota, Colombia
- Approximately 20 students. Most lectures in Spanish with Zoom translation
- Agenda:
  - **April 14** Introducción a python (G. Yetman)
  - **April 21** ARSET enfoque en "Introducción a las Cuadrículas de Población y su Integración con Datos de Teledetección para el Desarrollo Sostenible y la Gestión de Desastres" / Gridded Population (J. Martinez)
  - **April 28** Introducción a python 2 (G. Yetman)
  - **May 5** Risk Assessment and Sustainable Development (M. Mendoza), Exposure Development with EO data (G. Esquivias), and Tools for Risk Analysis and Visualization (Z. Hu)
  - **May 12** Agrupación de estadísticas espaciales y regresión de estadísticas espaciales / statistics and regression (K. MacManus)
  - **May 19** Drone (A. Gaughan, K. MacManus)
  - **May 26** ARSET enfoque en "Introducción a los Datos de Luce Nocturnas de NASA Black Marble" (J. Martinez)
  - **June 2** Global Relative Deprivation Index (J. Martinez)
  - **June 9** Conclusión del curso, resumen, Sesión de propuesta de proyecto
  - **June-July** Desarrollo de proyectos (por ejemplo, registros, horarios de oficina)



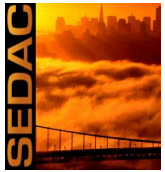
AmeriGEO 2023, “New Generation” Youth Day
11 August 2023, San Juan, Costa Rica

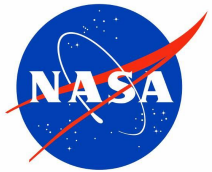
NASA

SEDAC Data and Tools

Columbia Climate School
Center for International Earth Science Information Network

Flood exposure: The LECZ data set





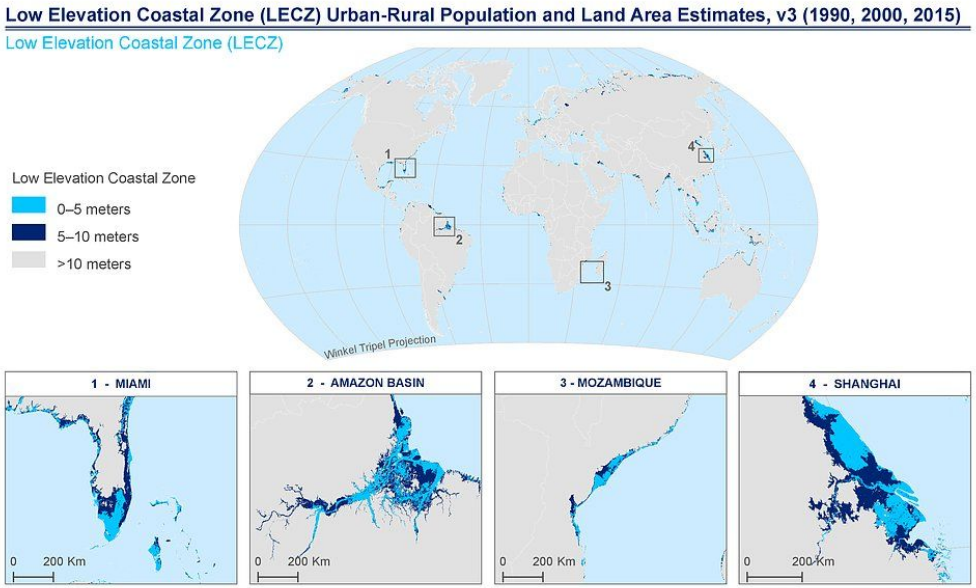
COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK

Low Elevation Coastal Zone, v3 (LECZv3) – Evaluating exposure

How Sensitive are SLR Exposure Estimates to Input Data Choices?

LECZv3 includes an Evaluation of Population, Settlement, and Elevation Data in Low Elevation Coastal Zones

https://doi.org/10.7927/d1x1-d702



Map Credit: CIESIN Columbia University, October 2021.

The Urban-Rural Population and Land Area Estimates, Version 3 is part of the Low Elevation Coastal Zone (LECZ) collection. The data set contains land areas with urban, quasi-urban, rural, and total population (counts) for 234 countries and other recognized territories over a 25 year period (1990–2015). This map represents the spatial extent of the Low Elevation Coastal Zone, by classifying land areas as the 5 meter LECZ (≤5m above sea level), 10 meter LECZ (≤10m above sea level), and non-coastal areas of any elevation (outside of LECZs), as estimated by the Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT-DEM).

Center for International Earth Science Information Network
Earth Institute | Columbia University

Data Source: Center for International Earth Science Information Network - CIESIN - Columbia University, and CUNY Institute for Demographic Research - CIDR - City University of New York. 2021. Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 3. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/d1x1-d702.

© 2021. The Trustees of Columbia University in the City of New York

This document is licensed under a Creative Commons Attribution 4.0 International License. https://creativecommons.org/licenses/by/4.0/

Columbia Climate School
Center for International Earth Science Information Network

CUNY Institute for Demographic Research

Elevation Data

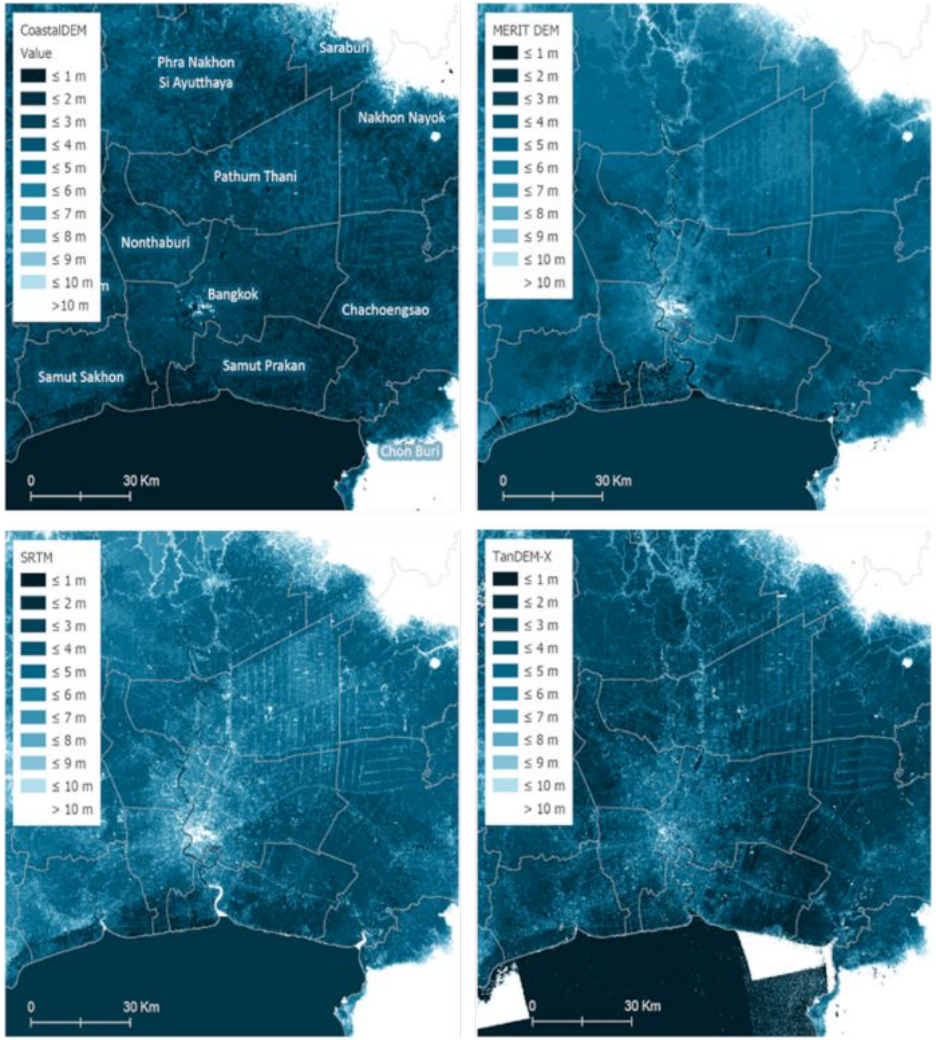
Many sources to choose From, we examined a few

1. Shuttle Radar Topography Mission (**SRTM**)

2. The Multi-Error-Removed Improved-Terrain DEM (**MERIT DEM**), “separated absolute bias, stripe noise, speckle noise and tree height bias using multiple satellite data sets and filtering techniques” to improve on the SRTM baseline

3. **CoastalDEM** - utilizes neural networks with an array of spatial covariates to improve on SRTM

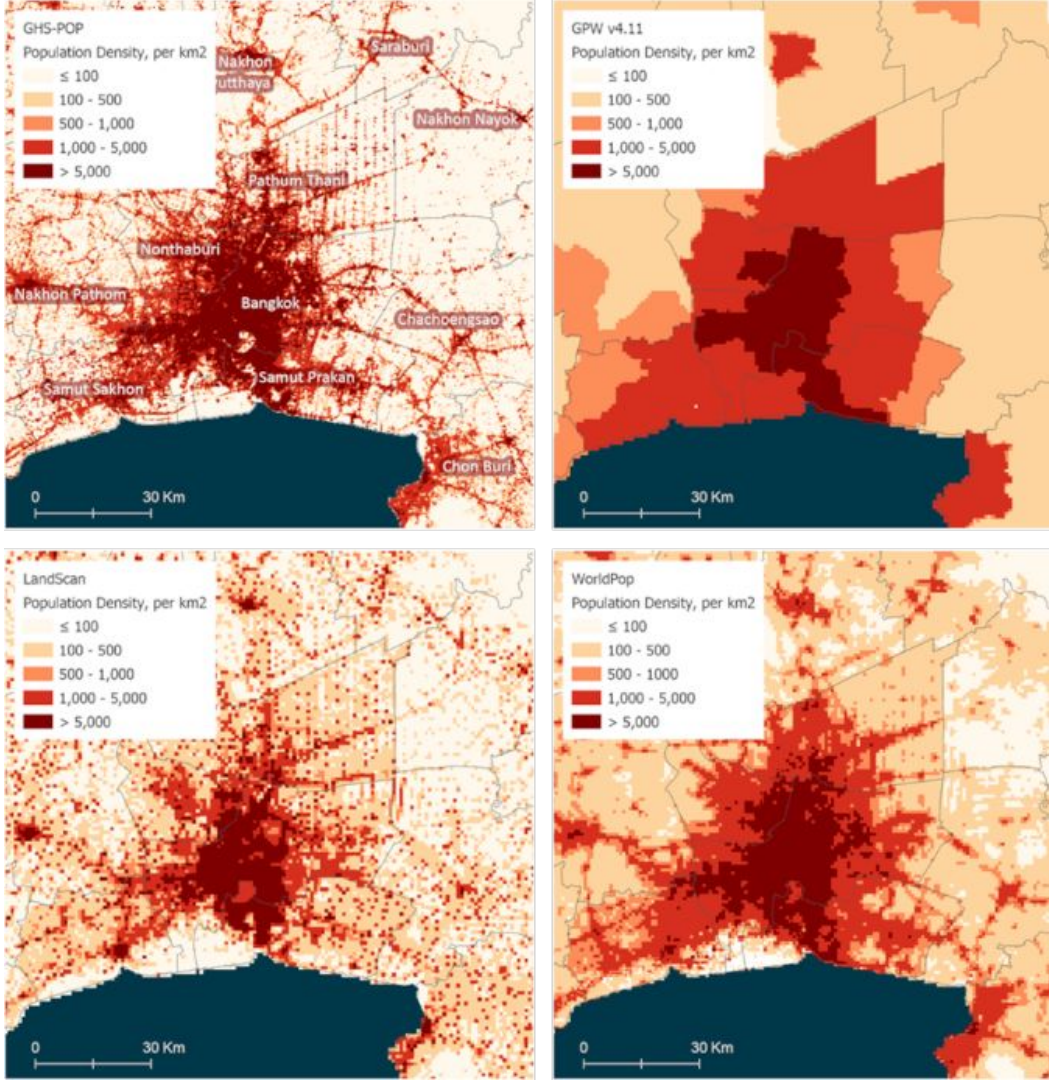
4. The TerraSAR-X add-on for Digital Elevation Measurements (**TanDEM-X**) departs from an SRTM baseline and provides a synthetic-aperture radar (SAR) based estimate of elevations globally.



Population Data

Even more options

1. SEDAC's Gridded Population of the World (GPW)
2. JRC's Global Human Settlement Layer-Population (GHS-POP)
3. ORNL's LandScan
4. WorldPop's population grid



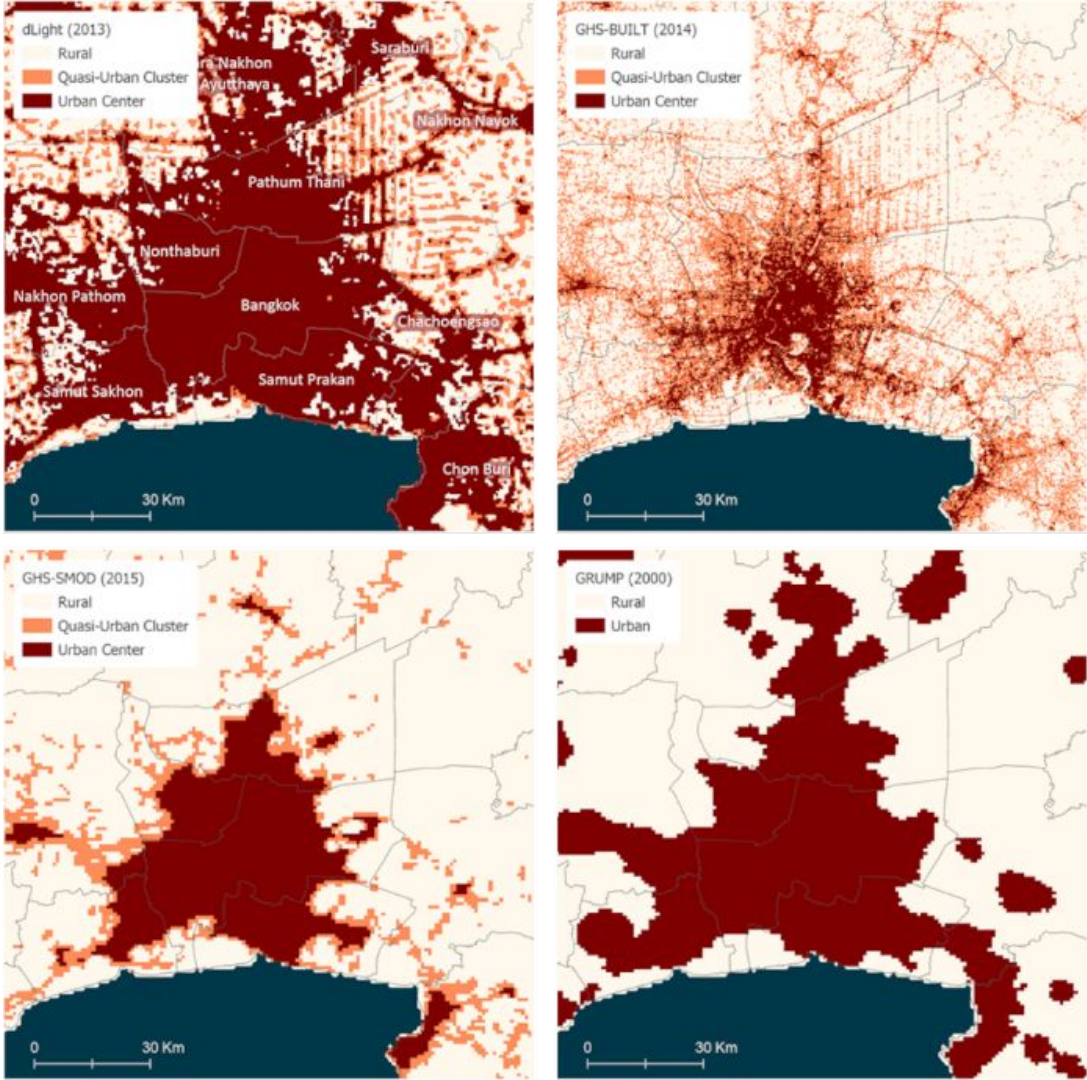
Settlement Data

1. GRUMP - DMSP-OLS nighttime lights derived Urban and Rural Extents

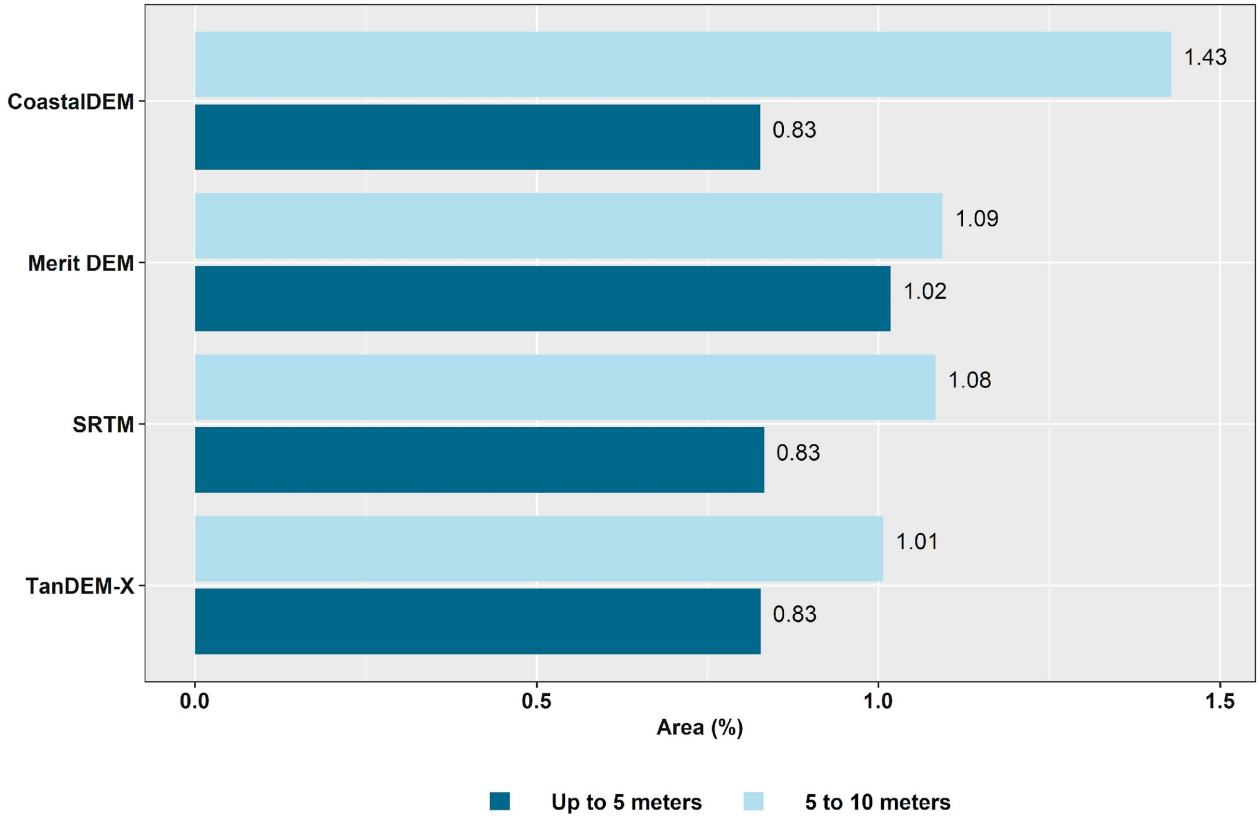
2. dLIGHT - DMSP-OLS and VIIRS nighttime lights derived time series

3. GHS-BUILT - thresholds based on percentage built-up

4. GHS-SMOD



Results in a Nutshell – Land Area

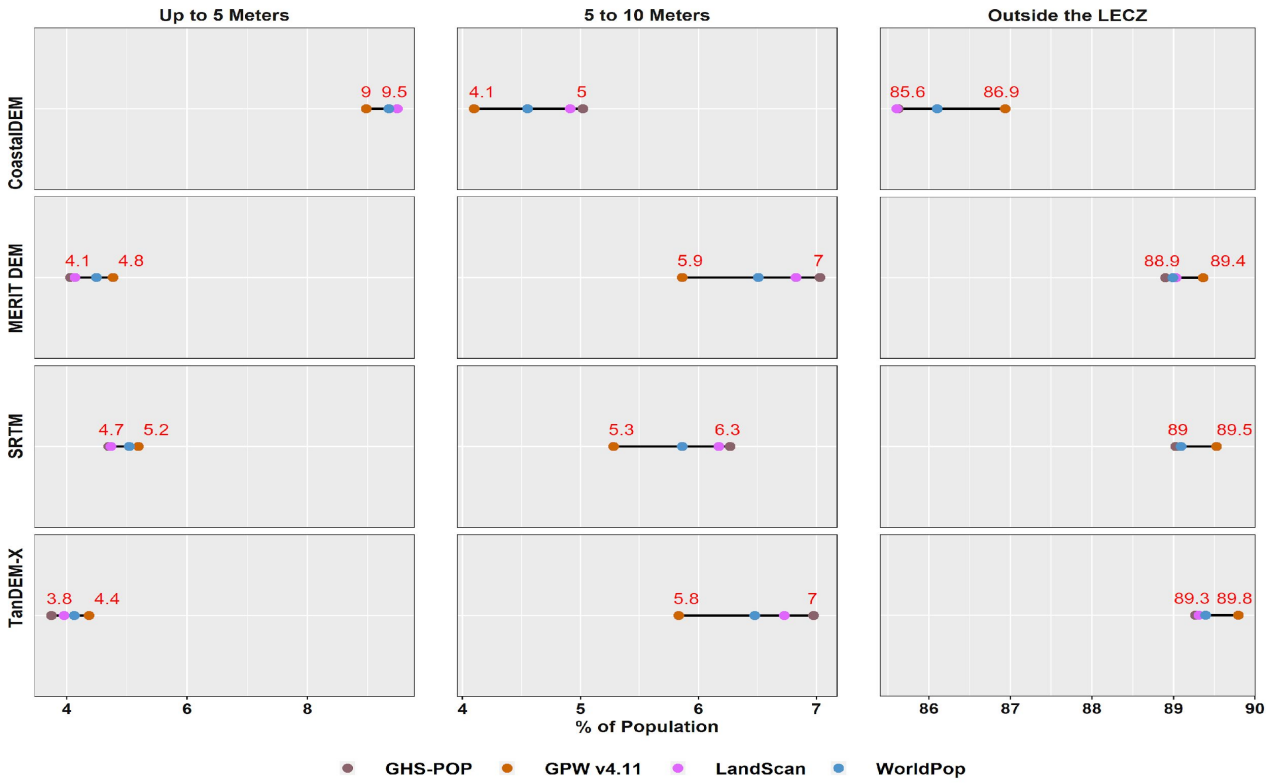


CoastalDEM assigns more area at 0 – 5m

MERIT assigns more area at 5-10m

Estimates are sensitive to choice of DEM

Results in a Nutshell – Population



GPW consistently estimates the lowest percentages in LECZs

GHS-POP consistently estimates the highest percentages in LECZs

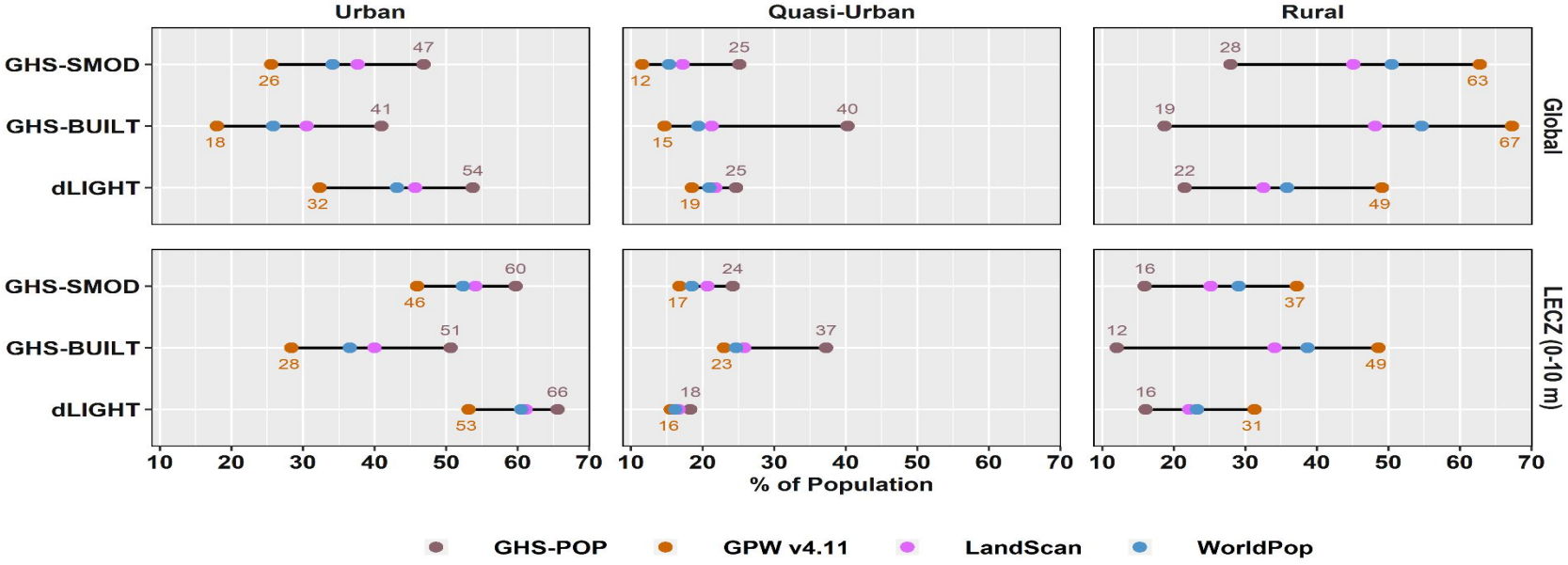
Estimates are sensitive to Population

Results in a Nutshell – Urban Continuum

GHS-POP concentrates more people in urban and quasi-urban

GPW concentrates more people in rural

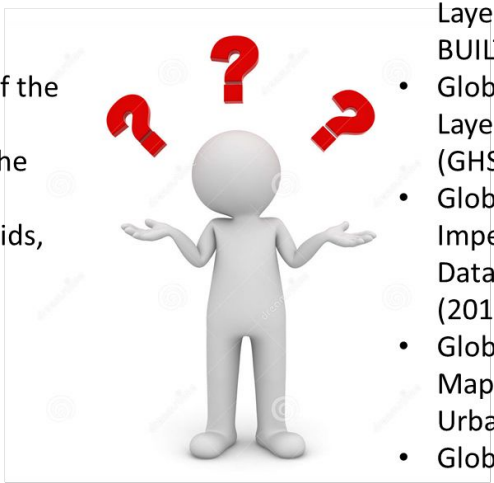
Settlement estimates are highly sensitive to data source



POPGRID can help demystify data choices for everyone

Population Grids

- Global Human Settlement Layer – Population (GHS-POP)
- Global Rural Urban Mapping Project (GRUMP)
- Gridded Population of the World, v4
- History Database of the Global Environment (HYDE) Population Grids, version 3.1
- LandScan Global Population database
- World Population Estimate
- WorldPop



Settlement Layers

- Global Human Built-up And Settlement Extent (HBASE) Dataset From Landsat, v1
- Global Human Settlement Layer – Built Up (GHS-BUILT)
- Global Human Settlement Layer – Settlement Model (GHS-SMOD)
- Global Man-made Impervious Surface (GMIS) Dataset From Landsat, v1 (2010)
- Global Rural Urban Mapping Project (GRUMP): Urban Extents
- Global Urban Footprint

POPGRID
DATA COLLABORATIVE

Fitness for use

Concept of **“relative data quality”** (Tayi & Ballou 1998)

Assess the **appropriateness** of a given dataset for an **intended purpose**

Guide user community in making **informed decisions** by better understanding:

Spatial, thematic and temporal **accuracy** in relation to the intended use, driven by…

(1) Input population data properties
(2) Modeling assumptions behind products
(3) Ancillary data

Slide courtesy of Stefan Leyk, University of Colorado

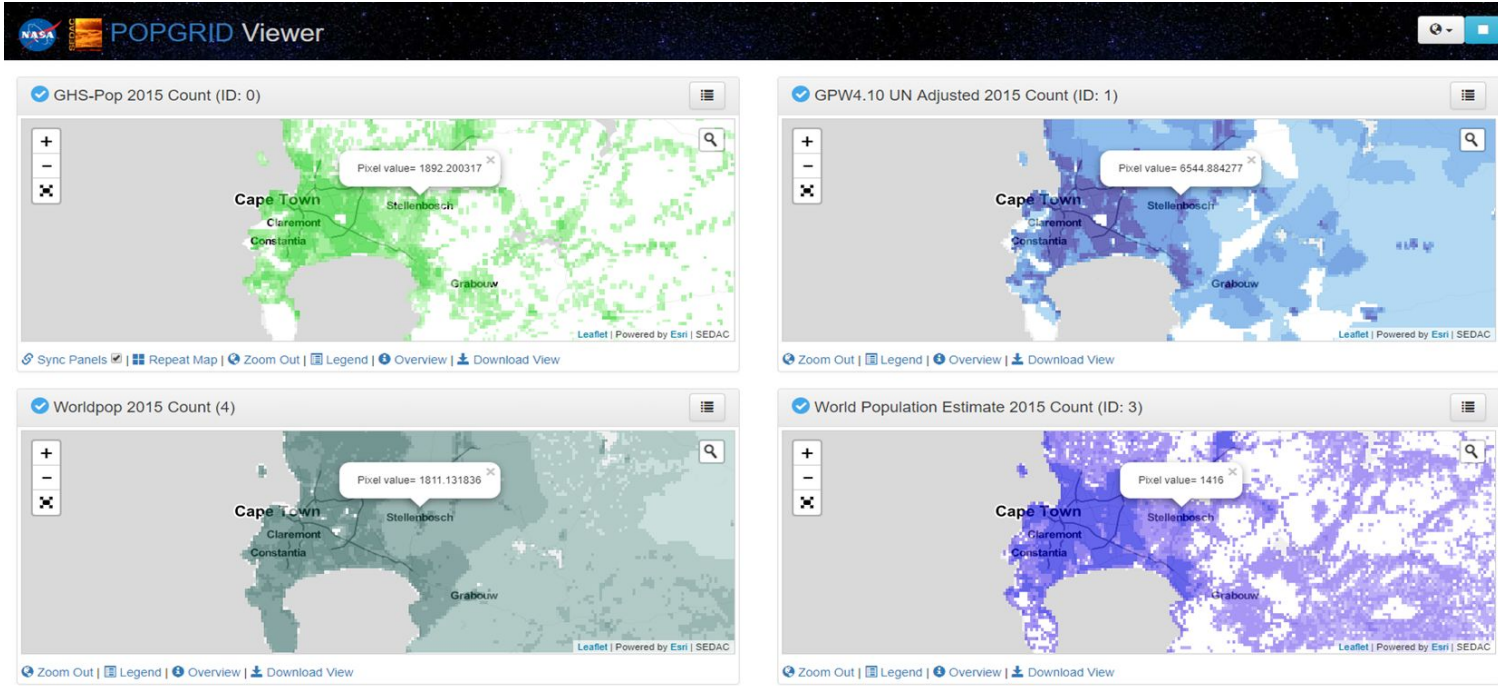
SEDAC's POPGRID Viewer allows users to visualize and compare population grids and settlement layers for GPW and many other sources

Quickly estimate population for an area of interest from multiple sources

Recent Earthdata article

Data Tool in Focus: POPGRID Viewer

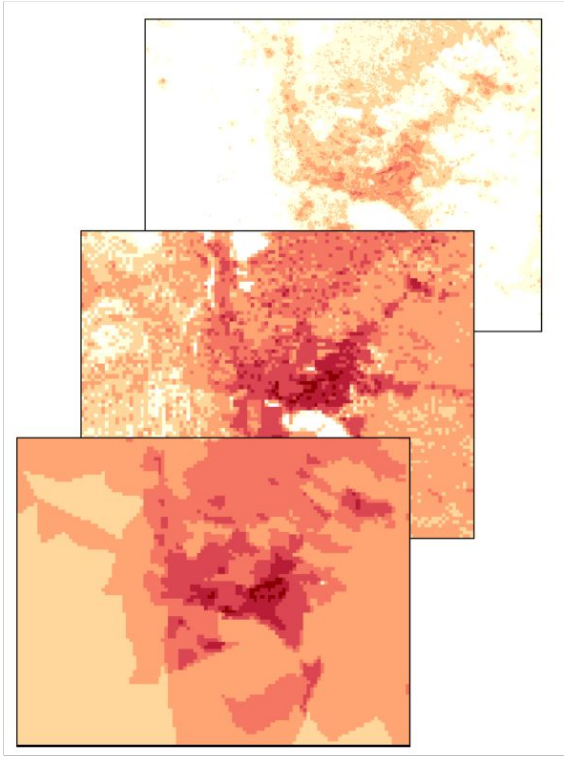
https://www.earthdata.nasa.gov/learn/articles/popgrid-viewer-sedac



https://sedac.ciesin.columbia.edu/mapping/popgrid/

The spatial allocation of population:

A review of large-scale gridded population data products and their fitness for use



The POPGRID Data Collaborative

Stefan Leyk[1*], Andrea E. Gaughan[2, 10], Susana B. Adamo[3], Alex de Sherbinin[3], Deborah Balk[9], Sergio Freire[5], Amy Rose[4], Forrest R. Stevens[2], Brian Blankespoor[8], Charlie Frye[7], Joshua Comenetz[6], Alessandro Sorichetta[10], Kytt MacManus[3], Linda Pistolesi[3], Marc Levy[3], Andrew J Tatem[10], Martino Pesaresi[5]

[1]*Department of Geography, University of Colorado Boulder, Boulder, CO 80309, U.S.A.*
[2]*Department of Geography and Geosciences, University of Louisville, KY, 40292, U.S.A.*
[3]*CIESIN, Columbia University, Palisades, NY, 10964, U.S.A.*
[4]*Human Dynamics Group, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, U.S.A.*
[5]*European Commission, Joint Research Centre (JRC), Ispra, Italy*
[6]*U.S. Census Bureau, Washington, D.C., 20233, U.S.A.*
[7]*Environmental Systems Research Institute, Redlands, CA 92373*
[8]*Development Research Group, World Bank, Washington, D.C. 20433, U.S.A.*
[9]*Institute for Demographic Research, Marxe School of Public and International Affairs, Baruch College, City University of New York, 10010, USA*
[10]*WorldPop, School of Geography and Environmental Sciences, University of Southampton, Southampton, SO17 1BJ, UK*

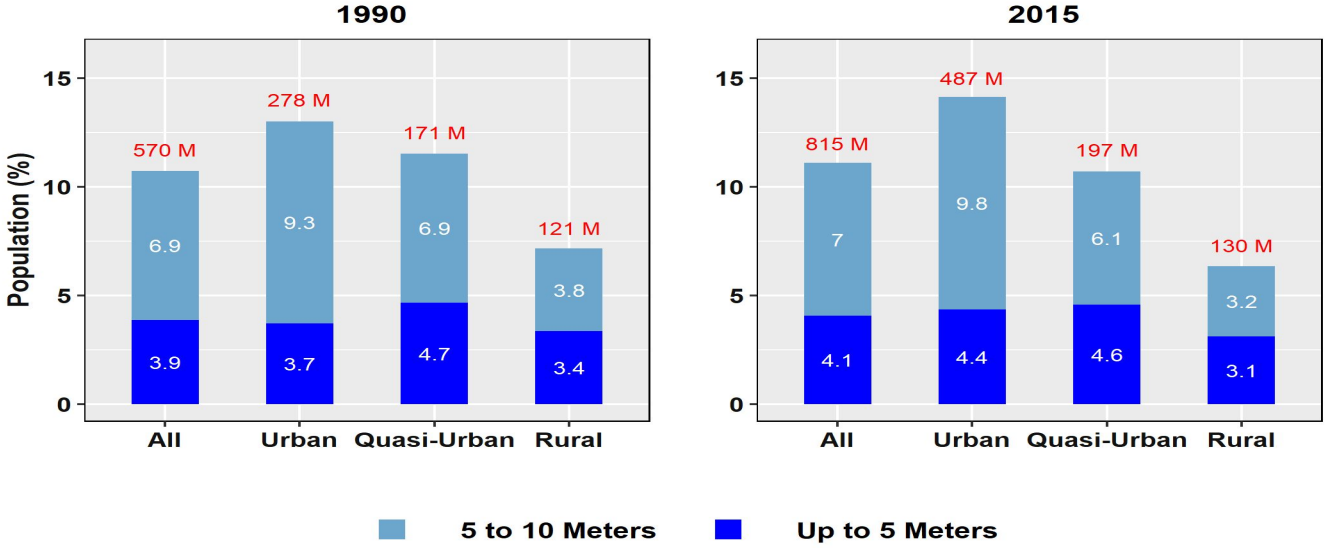
https://doi.org/10.5194/essd-11-1385-2019

Slide courtesy of Stefan Leyk, University of Colorado

Results in a Nutshell – Changes over Time

Urban areas have experienced the greatest increase in population

Urban areas within the LECZ have grown even faster



Data choices can lead to differences in estimates

- Large differences in estimates of potential SLR and coastal hazards
- While high agreement for urban centers and rural areas, but less so for the harder-to-classify areas (towns, peri-urban, sub-urban places), etc

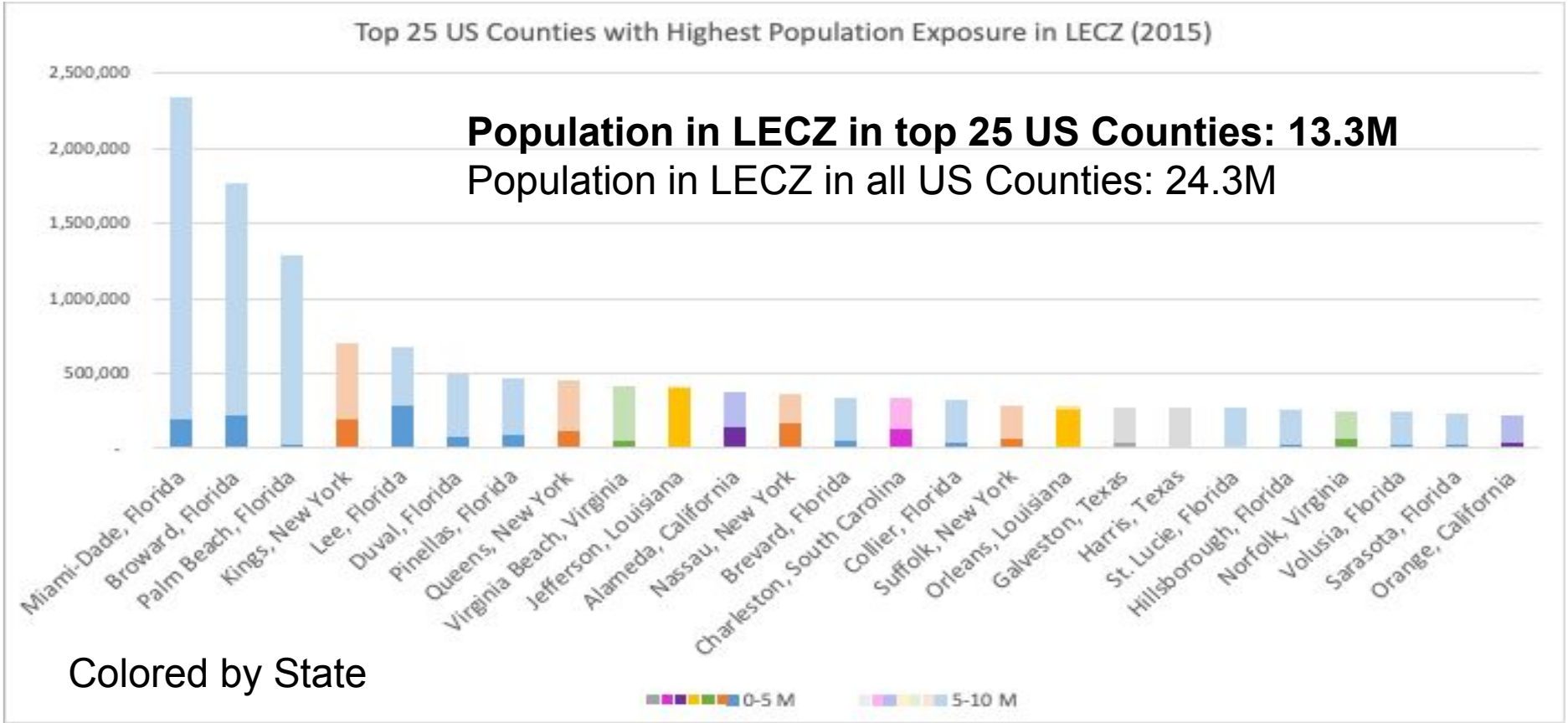
Consistency in estimation

- Despite important differences, every source we evaluated shows that LECZs
  - are disproportionately urban
  - urban population in the LECZ is growing at a rate faster than we see outside of the LECZ

Fitness for use matters

- Depends on respective use cases
- Change over time?
- Better local data?

Top 25 US Counties with Highest Population Exposure in 0-10M (2015)

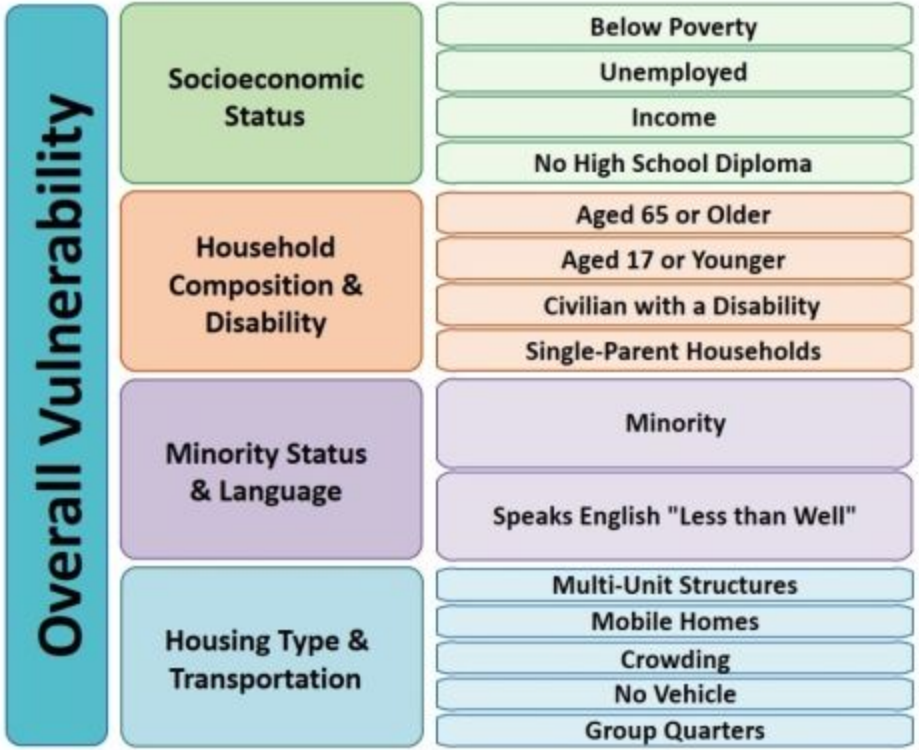


Moving from exposure to vulnerability

COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK

Moving from Flood Risk to Characterizing Vulnerability

Social Vulnerability Indexes look at factors like social and economic status, health, education, and housing to determine which areas have populations that would be particularly vulnerable to a flood (or other event) if it occurred, or that lack adequate resources to develop resilience from hazards.

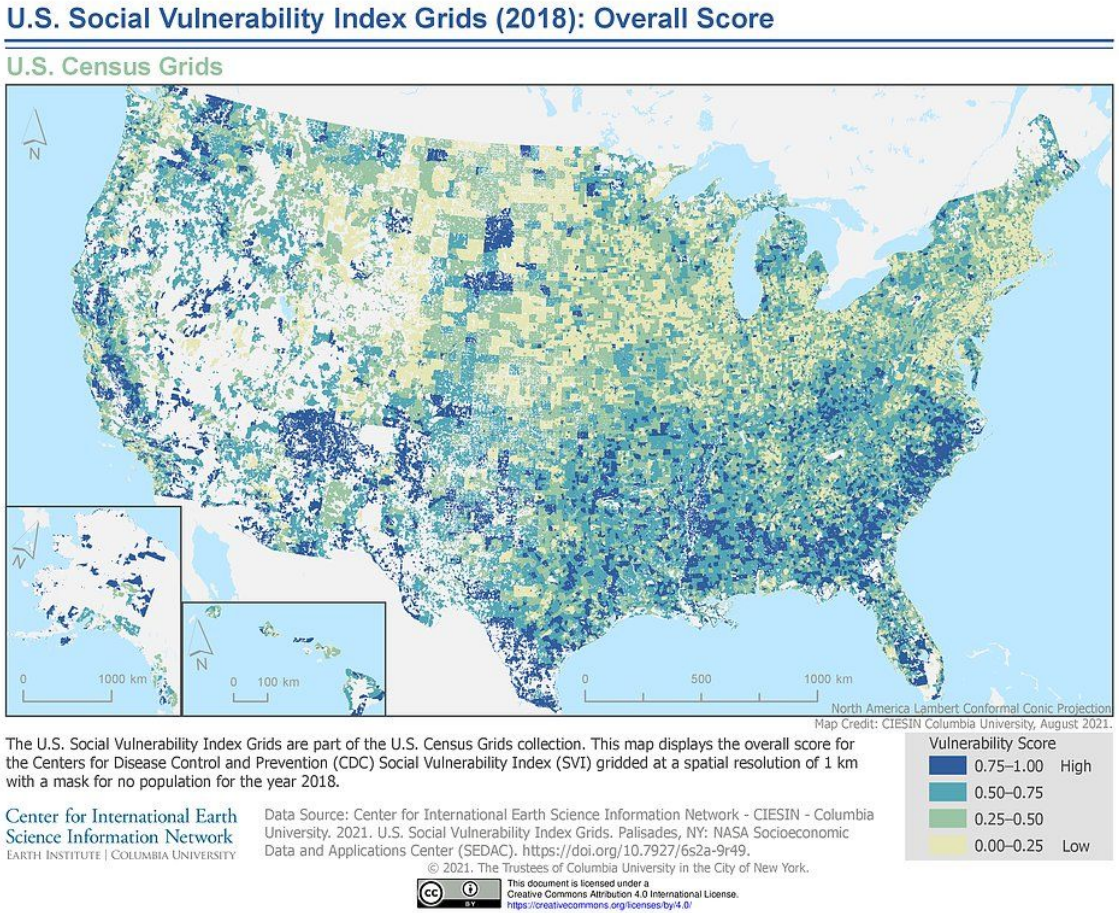


https://www.earthdata.nasa.gov/learn/articles/sedac-social-vulnerability-dataset

SEDAC Dataset: U.S. Social Vulnerability Index Grids

The U.S. Social Vulnerability Index Grids use data from the Center for Disease Control (CDC) for the years 2000, 2010, 2014, 2016, and 2018

https://doi.org/10.7927/6s2a-9r49



Columbia Climate School
Center for International Earth Science Information Network

Centers for Disease Control and Prevention

Population, Landscape, And Climate Estimates (PLACE), v4

Population, Landscape And Climate Estimates, v4: Coastal Zones

National Aggregates of Geospatial Data Collection (NAGDC)

The Population, Landscape, And Climate Estimates, Version 4 (PLACE IV) data set is part of the National Aggregates of Geospatial Data Collection (NAGDC). This map depicts coastal proximity zones (regions within 5, 10, 100, and 200 kilometers of a coast) created by generating buffers around a linear shoreline derived from the Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11 polygon layer and clipping the buffers to the coastline so that they included only land area.

https://doi.org/10.7927/rht8-jv78

Columbia Climate School
Center for International Earth Science Information Network

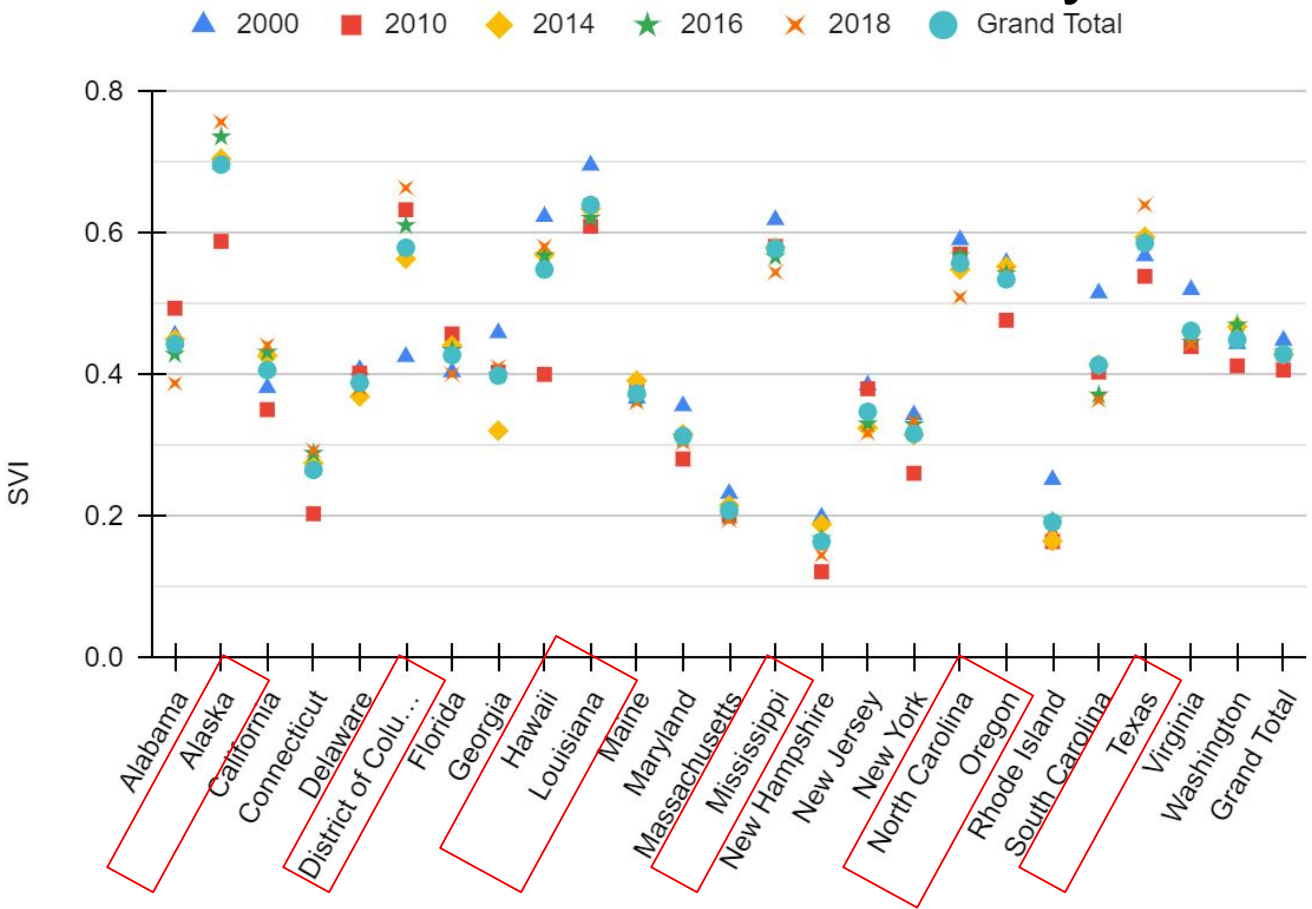
Columbia Climate School
Center for International Earth Science Information Network

Data Source: Center for International Earth Science Information Network - CIESIN - Columbia University. 2022. National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 4 (PLACE IV). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/rht8-jv78.

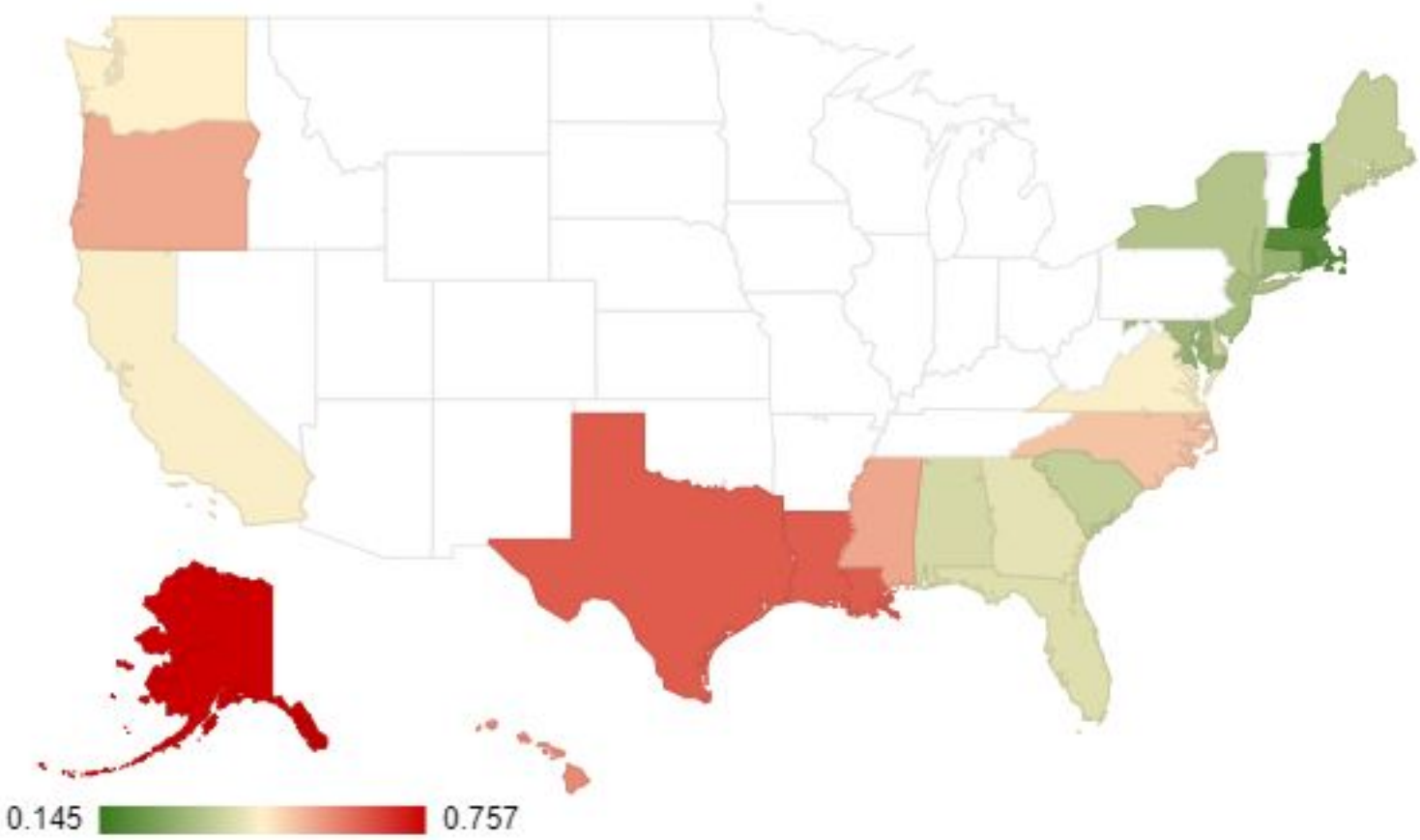
© 2022. The Trustees of Columbia University in the City of New York.

This document is licensed under a Creative Commons Attribution 4.0 International License.
https://creativecommons.org/licenses/by/4.0/

SVI in the 5km Coastal Zone by State



SVI in the 5km Coastal Zone by State



Similar analyses can be done for riparian flood zones

Check for updates

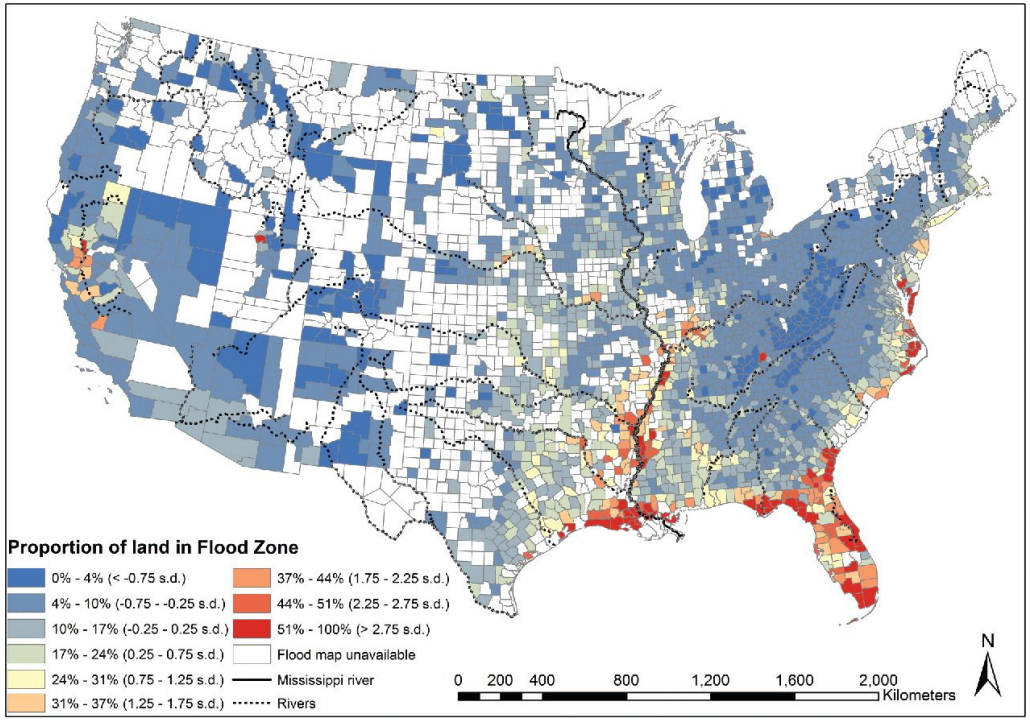
Changes in Exposure to Flood Hazards in the United States

Yi Qiang,* Nina S. N. Lam,† Heng Cai,† and Lei Zou†

*Department of Geography, University of Hawaii–Manoa

†Department of Environmental Sciences, Louisiana State University

https://doi.org/10.1080/24694452.2017.1320214



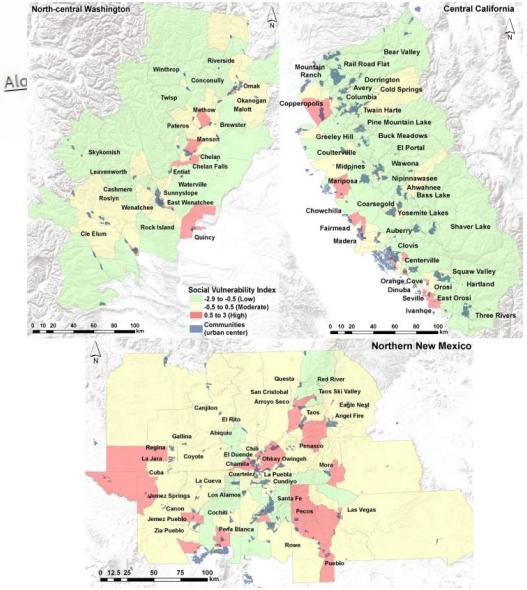
Similar analyses can be done for areas at risk of fires

ELSEVIER
Landscape and Urban Planning
Volume 189, September 2019, Pages 99-116

Research Paper

Social vulnerability to large wildfires in the western USA

Palaiologos Palaiologou [a], Ala
Michelle A. Day [d]



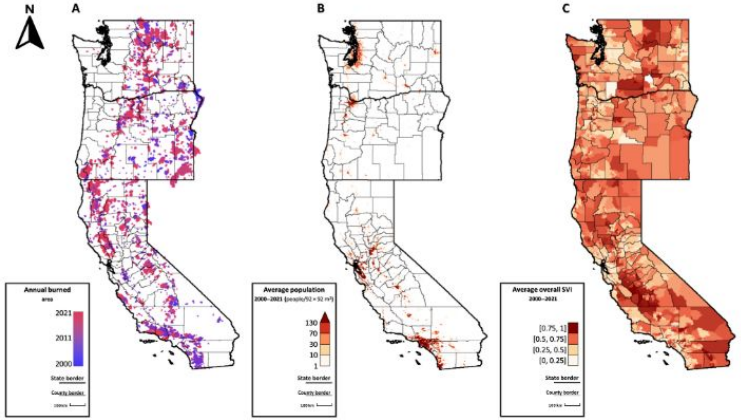
Limitations of using data by census tract

ScienceAdvances
HOME > SCIENCE ADVANCES > VOL. 9, NO. 38 > SOCIAL VULNERABILITY OF THE PEOPLE EXPOSED TO WILDFIRES IN U.S. WEST COAST STATES
Current Issue | First release papers | Archive | About
RESEARCH ARTICLE | ENVIRONMENTAL STUDIES

Social vulnerability of the people exposed to wildfires in U.S. West Coast states

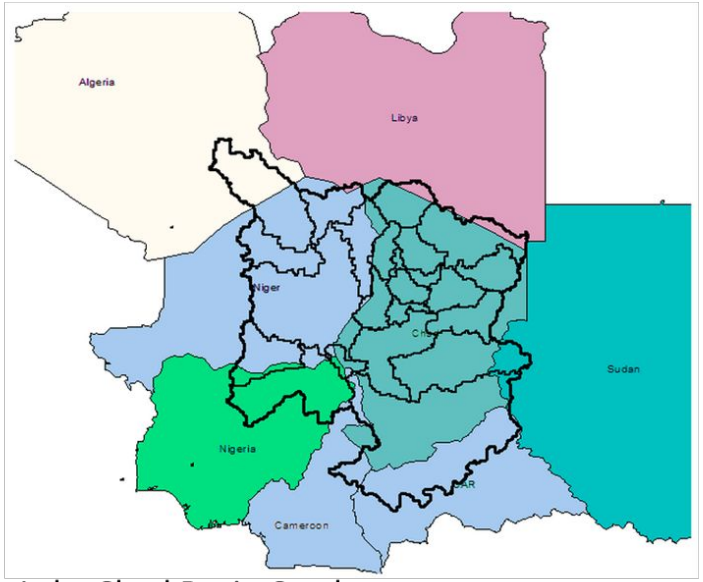
ARASH MODARESI RAD, JOHN T. ABATZOGLOU, ERICA FLEISHMAN, MIRANDA H. MOCKRIN, VOLKER C. RADELOFF, YAVAR POURMOHAMAD, MEGAN CATTAU, J. MICHAEL JOHNSON, PHILIP HIGUERA, [...], AND MOJTABA SADEGH +1 authors
Authors Info & Affiliations

SCIENCE ADVANCES · 20 Sep 2023 · Vol 9, Issue 38 · DOI: 10.1126/sciadv.adh4615



Importance of gridded population data

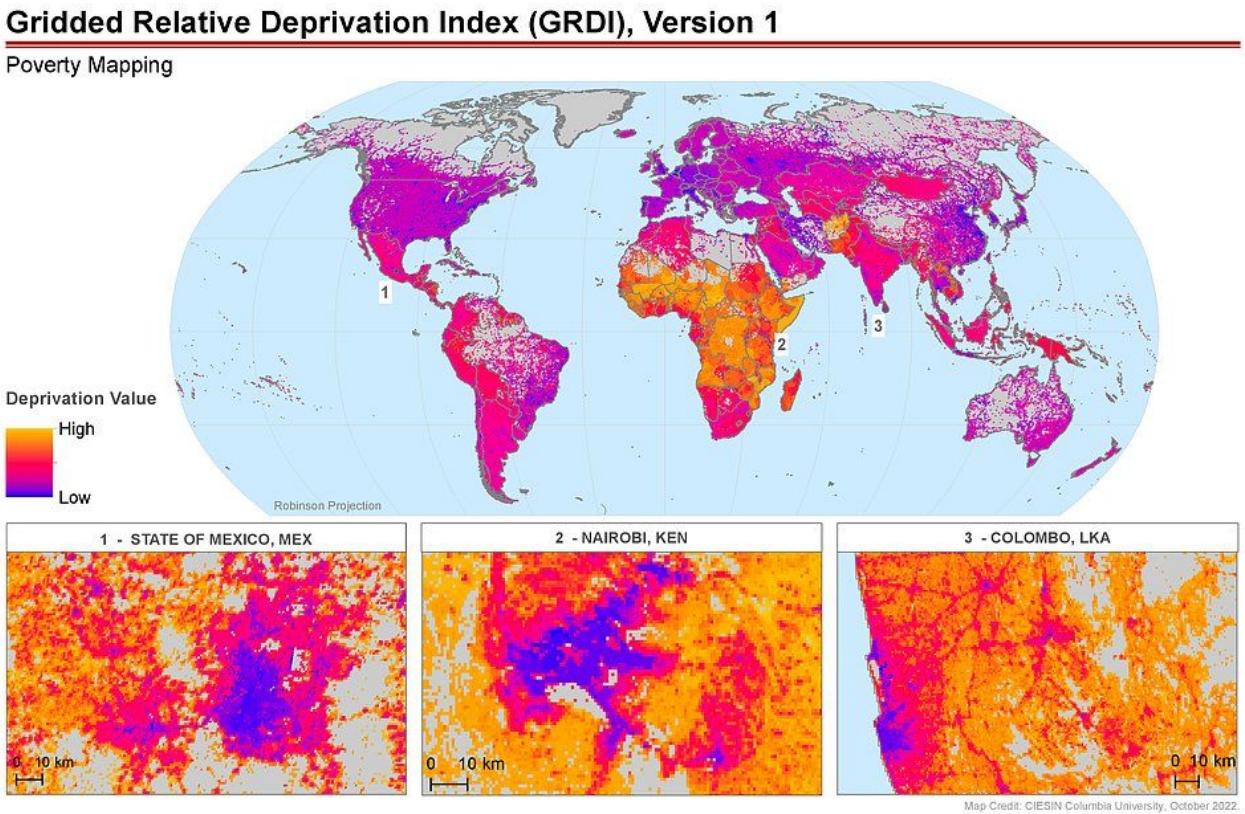
- Easily integrates with remote sensing data
- Allows population data to be aggregated to non-administrative geographies
- Census can be hard to come by
- Allows users to compare units of same shape and size



Lake Chad Basin Catchments

http://projects.inweh.unu.edu/inweh/display.php?ID=1282

Global Gridded Relative Deprivation Index (GRDI), v1



The Global Gridded Relative Deprivation Index (GRDI), Version 1 (GRDIv1) data set is part of the Poverty Mapping collection. The GRDIv1 provides a global index value that characterizes the levels of multidimensional deprivation in each pixel, where a value of 100 represents the highest level of deprivation and a value of 0 the lowest. All insets are indexed respectively. The GRDIv1 uses a variety of sociodemographic and satellite data inputs that were spatially harmonized, indexed, and weighted to provide an index that represents relative levels of multidimensional deprivation and poverty. This map displays deprivation values at a spatial resolution of ~1 km.

Columbia Climate School Center for International Earth Science Information Network

Data Source: Center for International Earth Science Information Network - CIESIN - Columbia University. 2022. Global Gridded Relative Deprivation Index (GRDI), Version 1. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/3xxe-ap97.

© 2022. The Trustees of Columbia University in the City of New York.
This document is licensed under a Creative Commons Attribution 4.0 International License.
https://creativecommons.org/licenses/by/4.0/

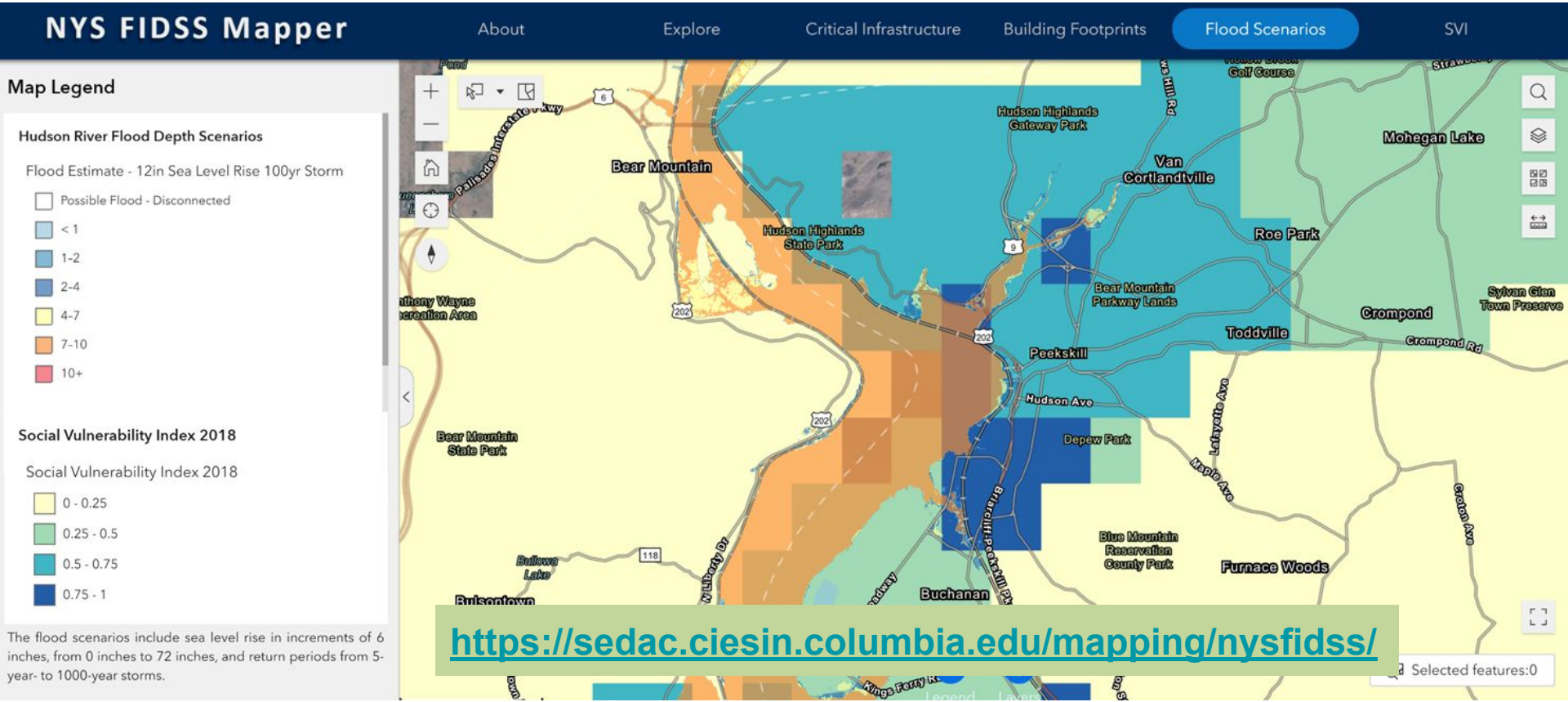
https://doi.org/10.7927/3xxe-ap97

Columbia Climate School
Center for International Earth Science Information Network

Additional Tools to Support Flood Analysis and forward EJ

COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK

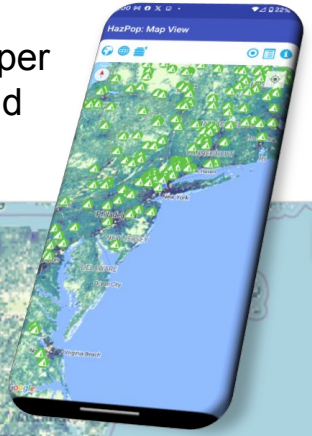
NY State Flood Information Decision Support System Demo
Comparing SVI with High Resolution SLR/Flood Scenarios

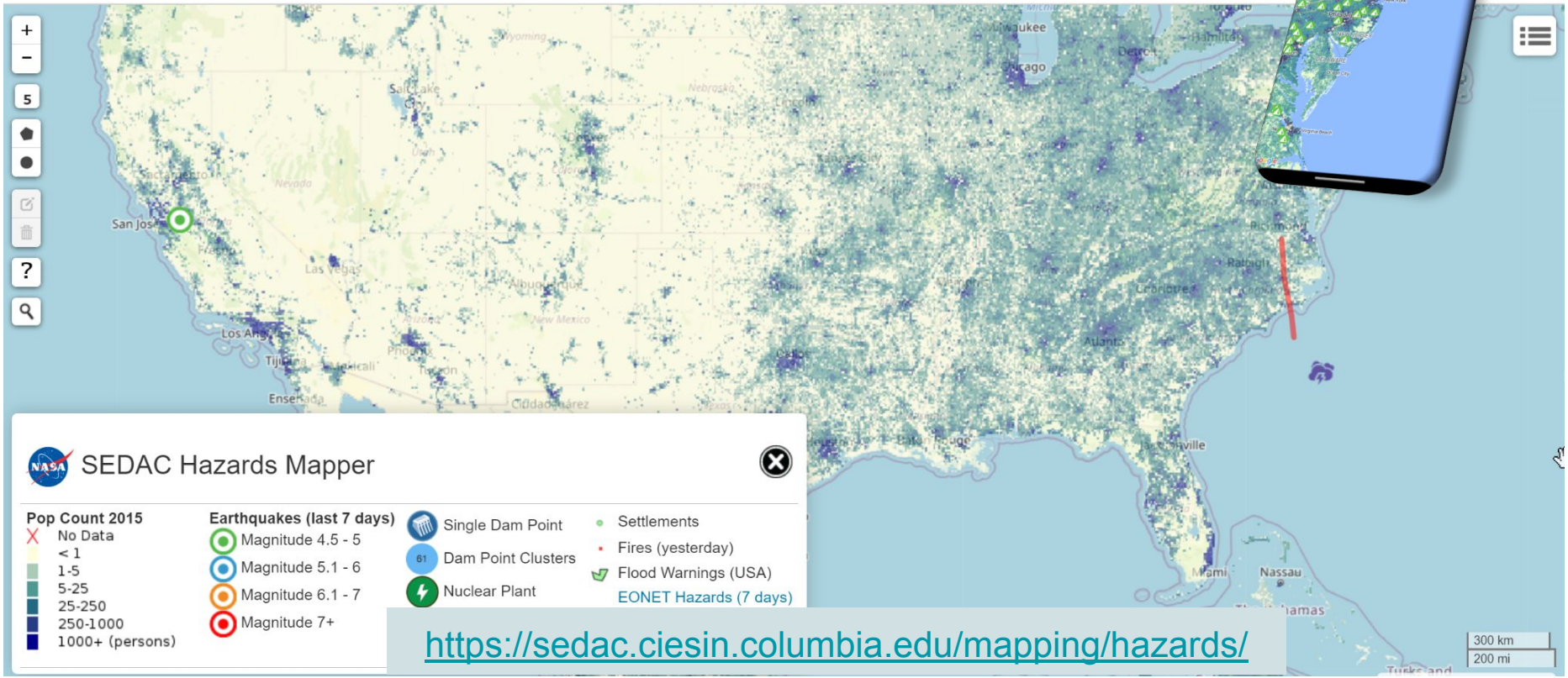


https://sedac.ciesin.columbia.edu/mapping/nysfidss/

SEDAC Hazards Mapper

HazPop Mapper for iPhone and Android





https://sedac.ciesin.columbia.edu/mapping/hazards/

Thanks!

NASA Socioeconomic Data and Applications Center (SEDAC)

COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK