Satellite Data for Air Quality Environmental Justice and Equity Applications

Part 3: Interactive Exercises for using Satellite and Demographic Data

Part 3 – Trainers

Gaige Kerr

Senior Research Scientist
George Washington University

Sarah Strode

Assoc. Research Scientist
NASA, Morgan State U.

Carl Malings

Assistant Research Scientist
NASA, Morgan State U.

Part 3 Objectives

By the end of Part 3, participants will be able to:

- Import relevant air quality datasets into EJSCREEN and use EJSCREEN to investigate and compare air quality with other environmental and demographic datasets.
- Pair appropriate satellite datasets for environmental indicators (air quality) with demographic information using Python.

Review of Prior Knowledge

- Multiple NASA and ESA satellite instruments collect data relevant to air quality.
  - MODIS, VIIRS, OMI, TROPOMI, GOES, TEMPO, MAIA
- Satellite and instrument properties affect what data they can provide.
  - Geostationary v. polar orbits (impacts when are where data are collected)
  - Multispectral v. hyperspectral (impacts what kind of data are collected)
- Satellites often retrieve column geophysical quantities (e.g., AOD).
- Satellite data can be combined with models and ground-based measurements to create Level 4 data products which estimate surface-level air quality.
- NASA provides free online tools to visualize, access, and analyze satellite data:
  - Worldview
  - Earthdata Search
  - Giovanni

How to Ask Questions

- Please put your questions in the Questions box and we will address them at the end of the webinar.
- Feel free to enter your questions as we go. We will try to get to all of the questions during the Q&A session after the webinar.
- The remainder of the questions will be answered in the Q&A document, which will be posted to the training website about a week after the training.

Part 3:

Interactive Exercises for using Satellite & Demographic Data

Introduction to EJScreen

EJScreen is an environmental justice screening and mapping tool and uses standard and nationally-consistent data to highlight places that may have higher environmental burdens and/or vulnerable populations.

A few examples of what EJScreen supports include:
- Informing outreach and engagement practices
- As an initial screen for voluntary programs, enhanced outreach in permitting, and prioritizing enforcement work
- Developing retrospective reports of EPA work
- Enhancing place-based activities

EJSCREEN: Environmental Justice Screening and Mapping Tool

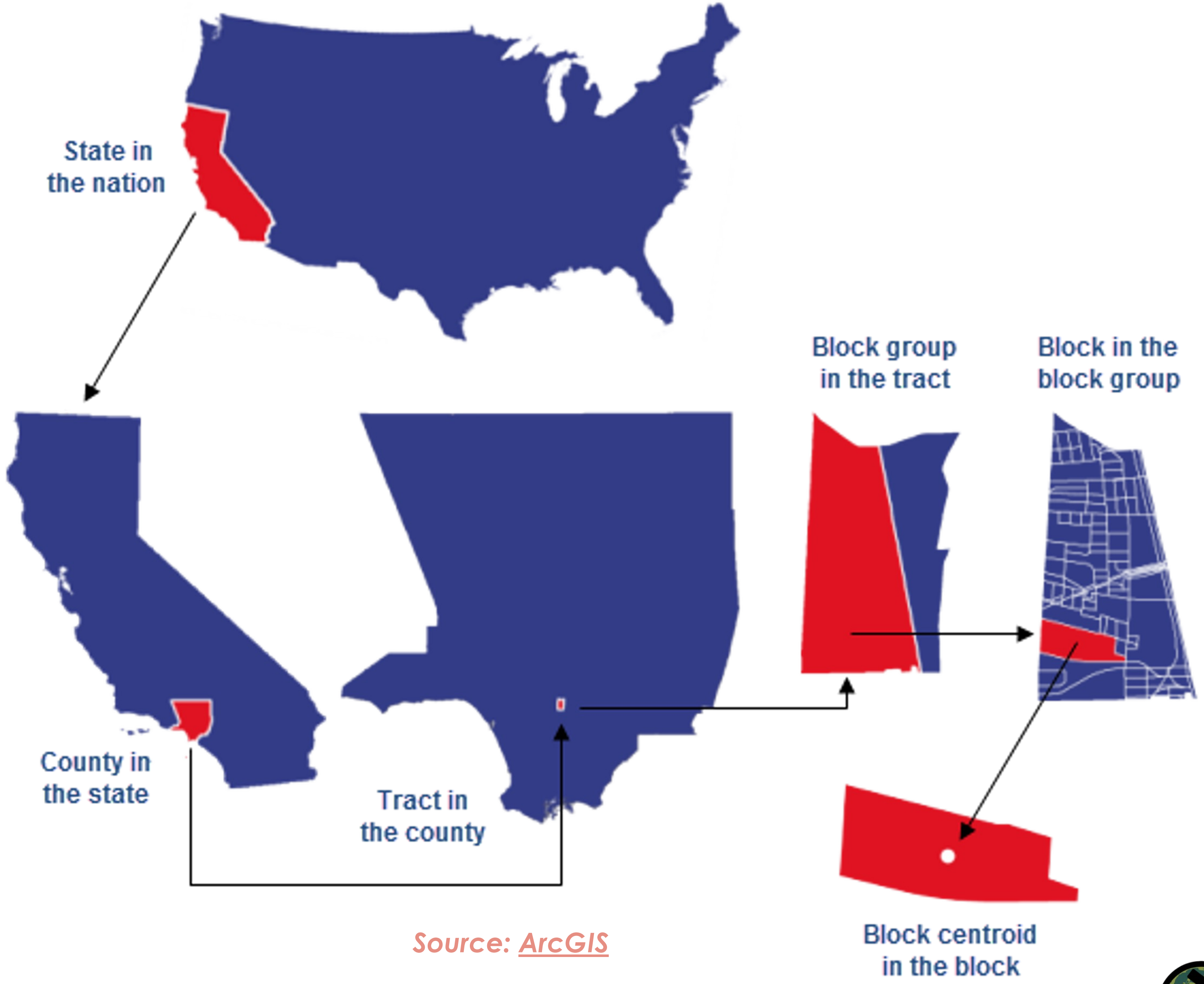
Launch the EJSCREEN Tool

Explore EPA's environmental justice screening and mapping tool

Source: Environmental Protection Agency

Introduction to EJScreen

The standard unit of analysis in EJScreen is the Census "block group." A block group is an area defined by the U.S. Census Bureau that usually has in the range of 600-3,000 inhabitants.



Source: ArcGIS

Introduction to EJScreen

Currently, EJScreen includes 13 environmental indicators and 7 socioeconomic indicators, which are combined to form "EJ indices".

🟥 **Fine particulate matter ($PM_{2.5}$)**	🟥 **Ozone (O_3)**	🟥 **Diesel PM**	🟥 **Lead paint**
🟦 **Traffic proximity and volume**	🟦 **Superfund proximity**	🟦 **RMP facility proximity**	🟦 **Hazardous waste proximity**
🟦 **Underground storage tank proximity**	🟦 **Wastewater discharge proximity**	🟩 **Cancer risk**	🟩 **Respiratory hazard index**
🟩 **Toxic releases to air**			

🟥 **Potential exposure** 🟦 **Proximity** 🟩 **Hazard/risk**

Introduction to EJScreen

Currently, EJScreen includes 13 environmental indicators and 7 socioeconomic indicators, which are be combined to form “EJ indices”.

$$\text{EJ index} = (\text{Environmental indicator percentile}) \bullet (\text{Demographic index})$$

$$\text{Demographic index} = \frac{\%\ \text{low income} + \%\ \text{People of Color}}{2}$$

Introduction to EJScreen

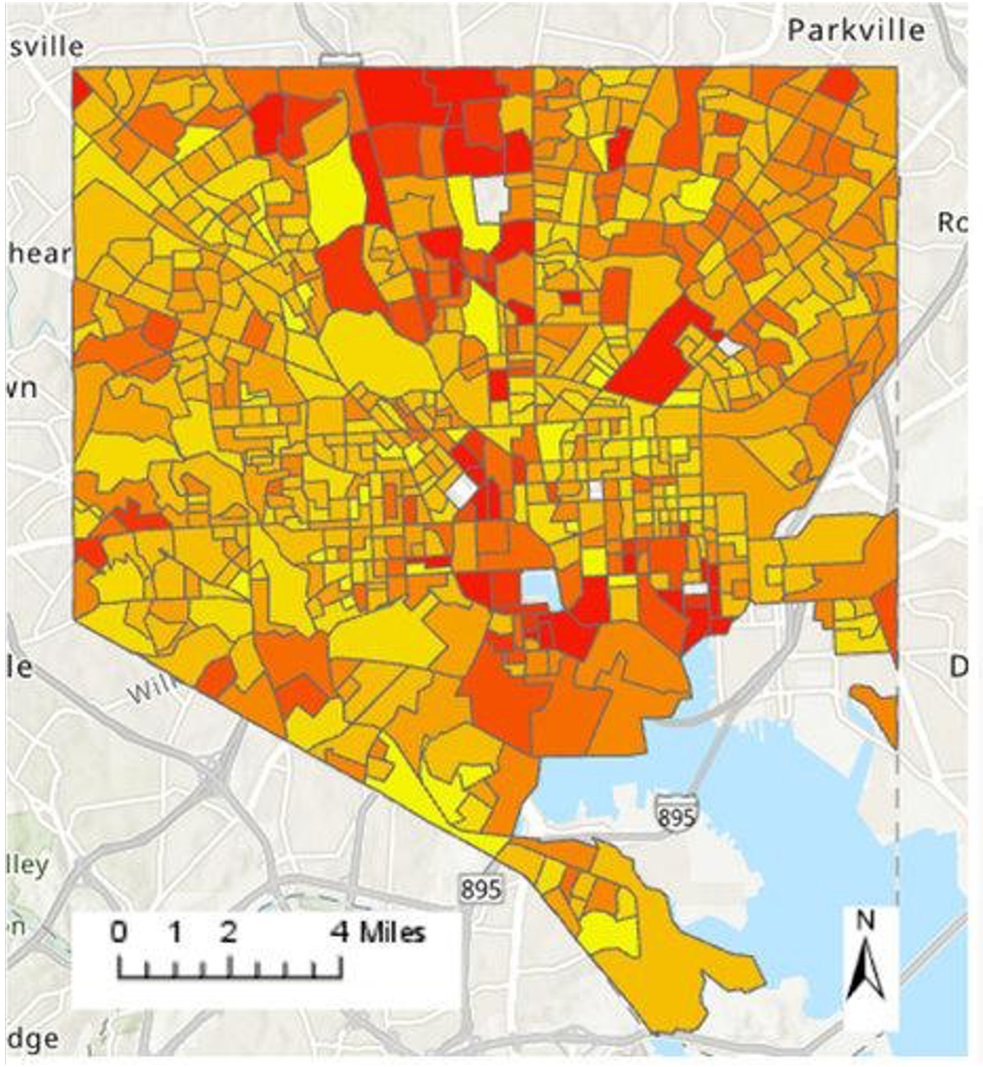
- EJScreen's data inputs represent a mixture of observed versus modeled or estimated products.
- The time periods represented by different data inputs vary, and some of the inputs are updated less frequently than others.

Environmental Indicator	Year	Source
Fine particulate matter ($PM_{2.5}$)	2018	EPA Community Multiscale Air Quality (CMAQ) model and monitor data
Ozone (O_3)	2018	EPA Community Multiscale Air Quality (CMAQ) model and monitor data
Traffic proximity and volume (count of vehicles on major roads, divided by distance in meters)	2019	Federal Highway Administration Highway Performance Monitoring System
Lead paint indicator	2016-2020	U.S. Census Bureau's American Community Survey 5-year summary
Diesel PM	2017	EPA Air Toxics Update

Source: Environmental Protection Agency

Potential EJScreen Limitations

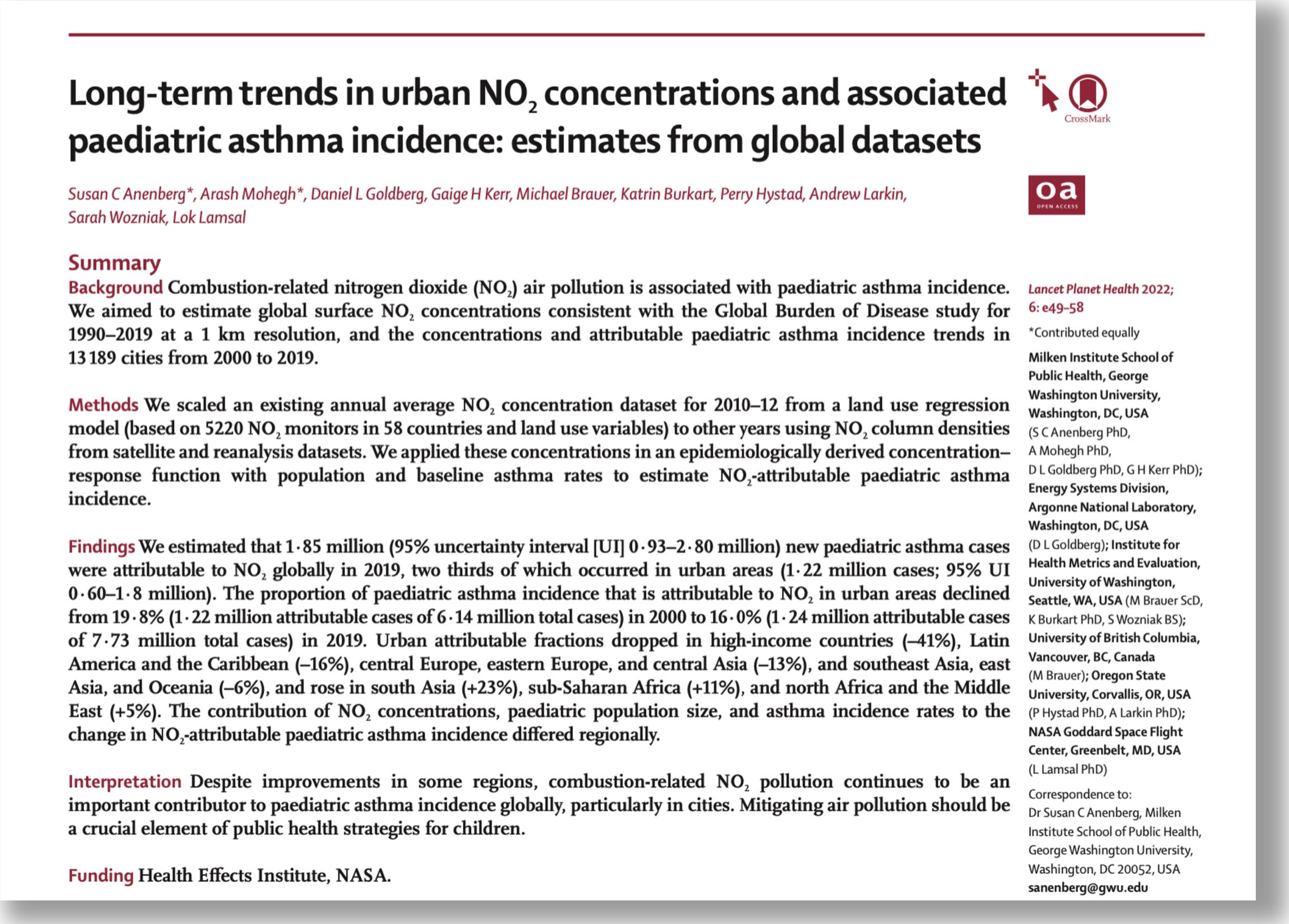
- Indices represent screening-level proxies for risk and exposure, not actual risk or exposure.
- Percentiles put indices in common units but don't indicate whether risks are equal or comparable.
- Tool is only available for the U.S.
- Data inputs may have coarse resolution or be out of date.
- Several important environmental impacts and demographic indicators are not included in the tool.



Source: Dukes et al. (2020 Environ Res Lett)

How to Incorporate Satellite Data in EJScreen

- A custom dataset estimating surface-level nitrogen dioxide (NO_2) concentrations can be directly integrated into EJScreen.
- This dataset combines satellite-derived NO_2 from NASA's Ozone Monitoring Instrument with a land-use regression model.



Long-term trends in urban NO_2 concentrations and associated paediatric asthma incidence: estimates from global datasets

Susan C Anenberg*, Arash Mohegh*, Daniel L Goldberg, Gaige H Kerr, Michael Brauer, Katrin Burkart, Perry Hystad, Andrew Larkin, Sarah Wozniak, Lok Lamsal

Summary

Background Combustion-related nitrogen dioxide (NO_2) air pollution is associated with paediatric asthma incidence. We aimed to estimate global surface NO_2 concentrations consistent with the Global Burden of Disease study for 1990–2019 at a 1 km resolution, and the concentrations and attributable paediatric asthma incidence trends in 13 189 cities from 2000 to 2019.

Methods We scaled an existing annual average NO_2 concentration dataset for 2010–12 from a land use regression model (based on 5220 NO_2 monitors in 58 countries and land use variables) to other years using NO_2 column densities from satellite and reanalysis datasets. We applied these concentrations in an epidemiologically derived concentration–response function with population and baseline asthma rates to estimate NO_2-attributable paediatric asthma incidence.

Findings We estimated that 1·85 million (95% uncertainty interval [UI] 0·93–2·80 million) new paediatric asthma cases were attributable to NO_2 globally in 2019, two thirds of which occurred in urban areas (1·22 million cases; 95% UI 0·60–1·8 million). The proportion of paediatric asthma incidence that is attributable to NO_2 in urban areas declined from 19·8% (1·22 million attributable cases of 6·14 million total cases) in 2000 to 16·0% (1·24 million attributable cases of 7·73 million total cases) in 2019. Urban attributable fractions dropped in high-income countries (–41%), Latin America and the Caribbean (–16%), central Europe, eastern Europe, and central Asia (–13%), and southeast Asia, east Asia, and Oceania (–6%), and rose in south Asia (+23%), sub-Saharan Africa (+11%), and north Africa and the Middle East (+5%). The contribution of NO_2 concentrations, paediatric population size, and asthma incidence rates to the change in NO_2-attributable paediatric asthma incidence differed regionally.

Interpretation Despite improvements in some regions, combustion-related NO_2 pollution continues to be an important contributor to paediatric asthma incidence globally, particularly in cities. Mitigating air pollution should be a crucial element of public health strategies for children.

Funding Health Effects Institute, NASA.

Lancet Planet Health 2022; 6: e49–58

*Contributed equally

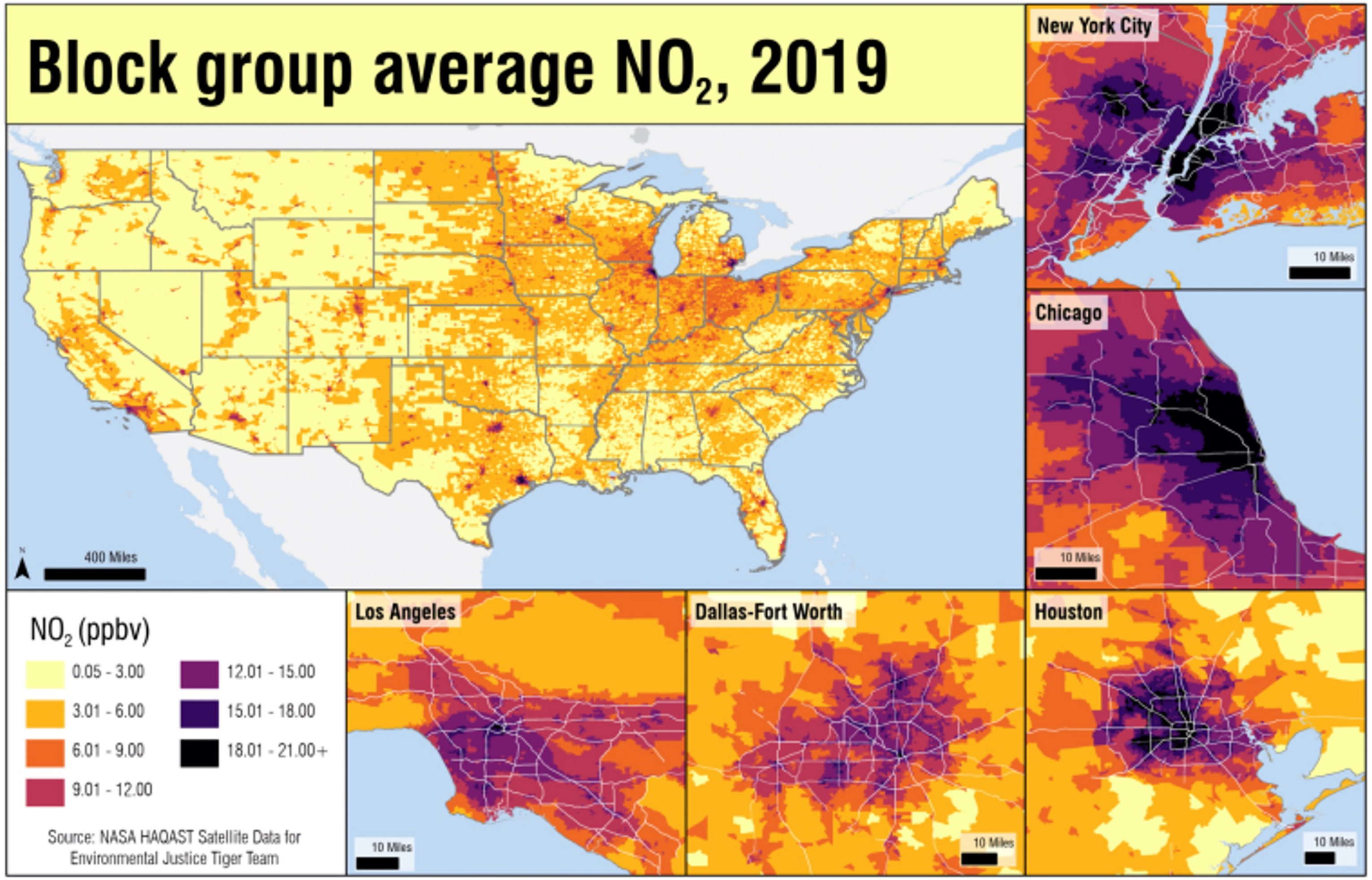
Milken Institute School of Public Health, George Washington University, Washington, DC, USA (S C Anenberg PhD, A Mohegh PhD, D L Goldberg PhD, G H Kerr PhD); **Energy Systems Division, Argonne National Laboratory, Washington, DC, USA** (D L Goldberg); **Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA** (M Brauer ScD, K Burkart PhD, S Wozniak BS); **University of British Columbia, Vancouver, BC, Canada** (M Brauer); **Oregon State University, Corvallis, OR, USA** (P Hystad PhD, A Larkin PhD); **NASA Goddard Space Flight Center, Greenbelt, MD, USA** (L Lamsal PhD)

Correspondence to:
Dr Susan C Anenberg, Milken Institute School of Public Health, George Washington University, Washington, DC 20052, USA
sanenberg@gwu.edu

Source: Anenberg, Mohegh et al. (2022 Lancet Planet Health)

How to Incorporate Satellite Data in EJScreen

- A custom dataset estimating surface-level nitrogen dioxide (NO_2) concentrations can be directly integrated into EJScreen.
- This dataset combines satellite-derived NO_2 from NASA's Ozone Monitoring Instrument with a land-use regression model.



Source: NASA HAQAST

EJScreen Demonstration

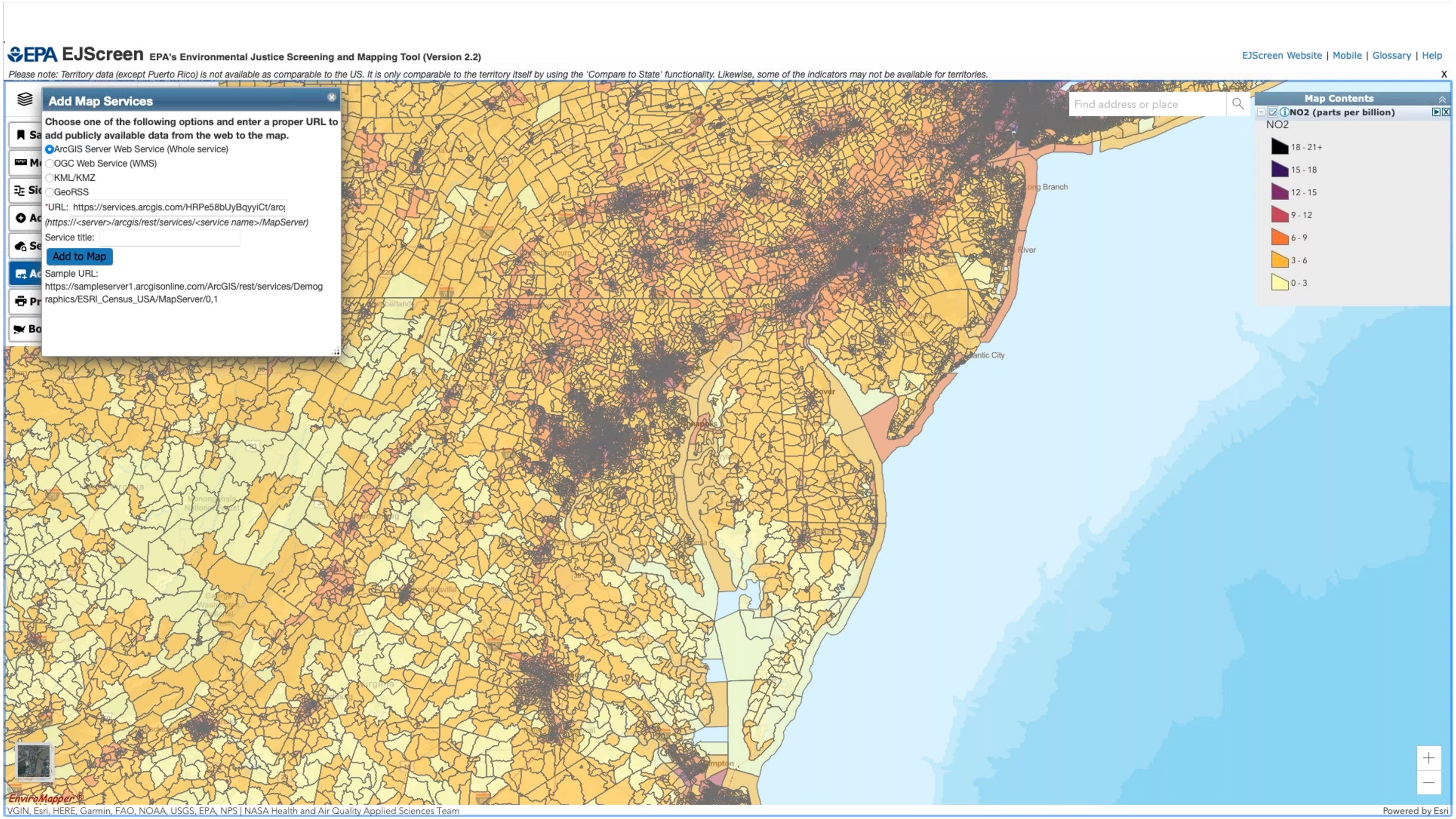
Link to the EPA EJScreen tool:

https://ejscreen.epa.gov/mapper/

The custom NO_2 dataset to be added into EJScreen:

https://services.arcgis.com/HRPe58bUyBqyyiCt/arcgis/rest/services/US_NO2_Block_Groups/FeatureServer

EJScreen Demonstration: adding custom datasets



EPA EJScreen EPA's Environmental Justice Screening and Mapping Tool (Version 2.2)

Please note: Territory data (except Puerto Rico) is not available as comparable to the US. It is only comparable to the territory itself by using the 'Compare to State' functionality. Likewise, some of the indicators may not be available for territories.

Add Map Services

Choose one of the following options and enter a proper URL to add publicly available data from the web to the map.

● ArcGIS Server Web Service (Whole service)
○ OGC Web Service (WMS)
○ KML/KMZ
○ GeoRSS

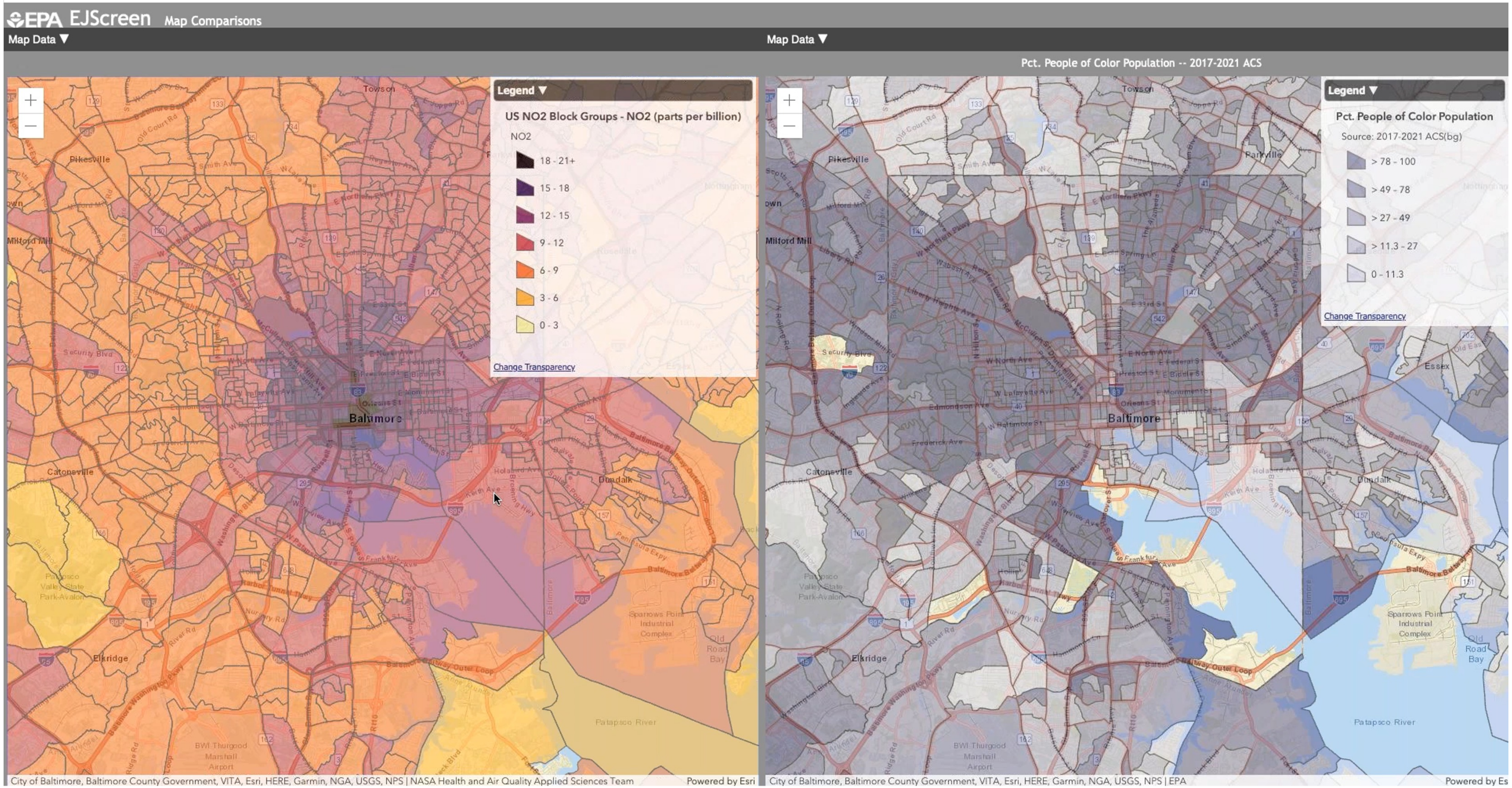
*URL: https://services.arcgis.com/HRPe58bUyBqyyiCt/arc

(https://<server>/arcgis/rest/services/<service name>/MapServer)

Service title:

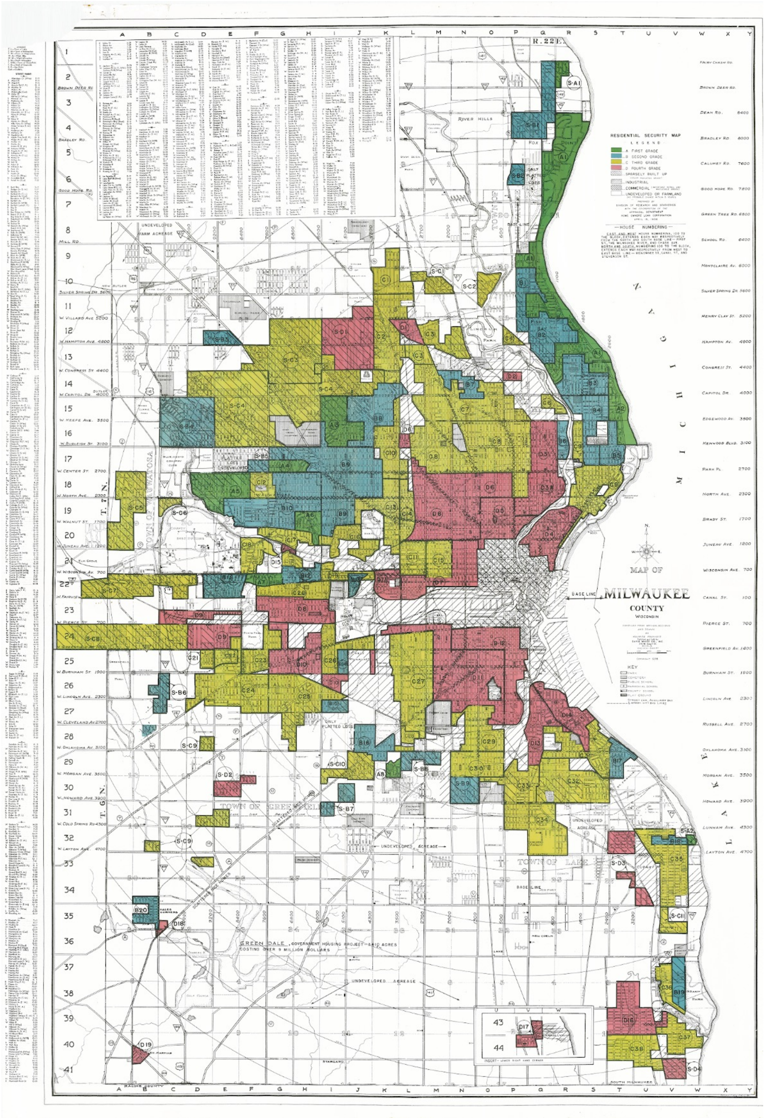
Sample URL: https://sampleserver1.arcgisonline.com/ArcGIS/rest/services/Demographics/ESRI_Census_USA/MapServer/0,1

EJScreen Demonstration: side-by-side comparisons



Independent Exploration with EJScreen

- Which part(s) of your chosen city or area have high NO_2?
- What industries or NO_X sources might lead to these patterns?
- Does NO_2 vary with race, educational attainment, and/or income?
- What factors might be relevant in explaining the colocation of marginalized population groups with high NO_2 (e.g., redlining, industry)?
- Are there known NO_X sources that aren’t apparent in the map?



Source: Mapping Inequality

Independent Exploration with EJScreen

Take some time now to explore EJScreen on your own.

Zoom to an area of interest to you and try to explore the following questions:

- Which part(s) of your chosen city or area have high NO_2?
- What industries or NO_X sources might lead to these patterns?
- Does NO_2 vary with race, educational attainment, and/or income?
- What factors might be relevant in explaining the colocation of marginalized population groups with high NO_2 (e.g., redlining, industry)?
- Are there known NO_X sources that aren't apparent in the map?

The training will resume at 14:15 EDT (UTC-4).

EJScreen- Versus Python-Based Analyses

EJSCREEN

✅ Accessible to those with limited knowledge of computer programming

✅ Avoids having to find and reformat disparate data sources

❌ Doesn’t allow calculation of zonal statistics

Python

✅ Freely-available, open source software

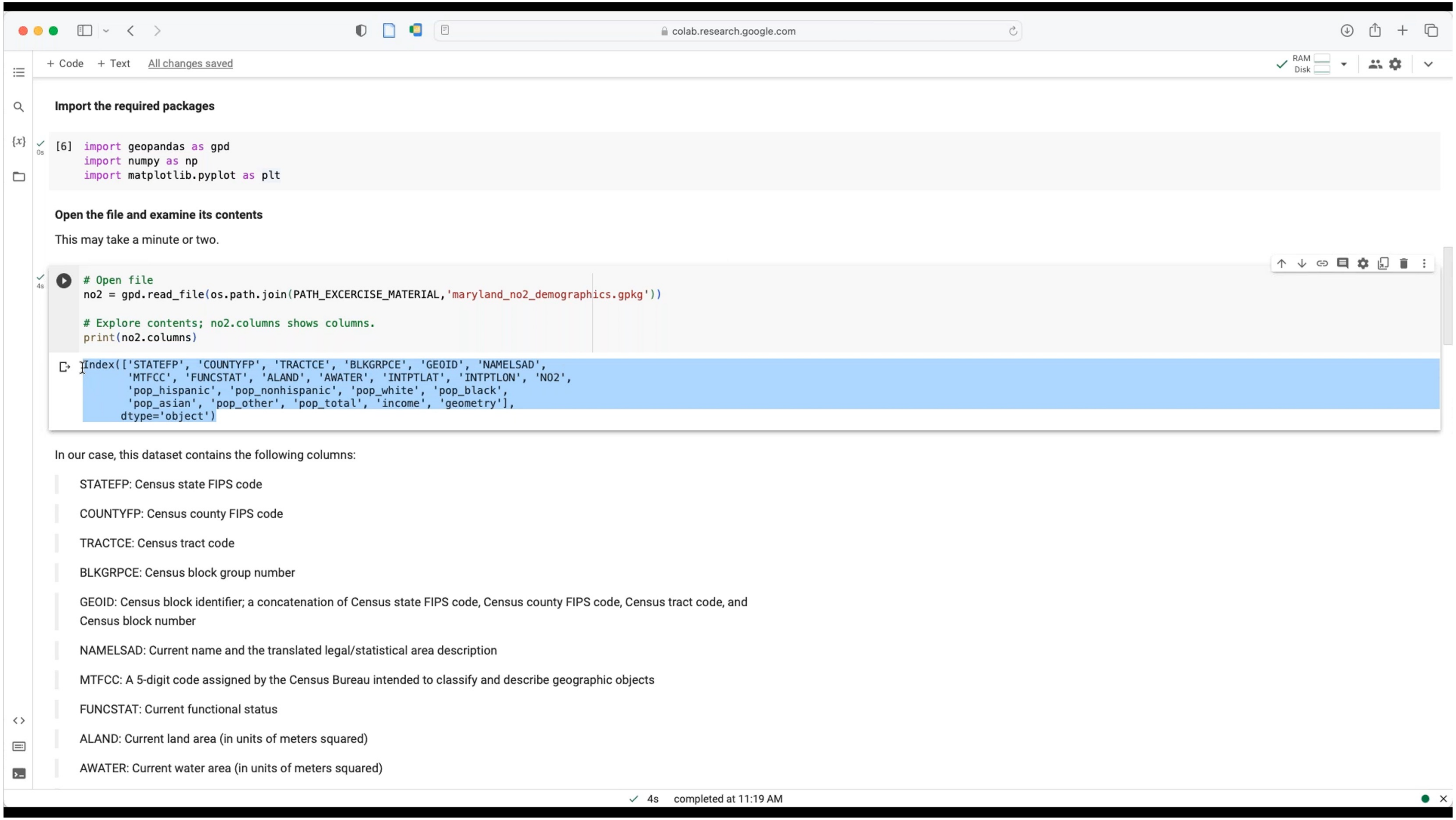
✅ Allows for custom mapping and statistical calculations

❌ Requires prerequisite knowledge on coding and acquisition of data inputs

Python Demonstration: Google Colab setup

- Follow these instructions to setup the Python example in your Google Colab.

Python Demonstration: loading data using geopandas



Import the required packages

```
[6] import geopandas as gpd
    import numpy as np
    import matplotlib.pyplot as plt
```

Open the file and examine its contents

This may take a minute or two.

```
# Open file
no2 = gpd.read_file(os.path.join(PATH_EXCERCISE_MATERIAL,'maryland_no2_demographics.gpkg'))

# Explore contents; no2.columns shows columns.
print(no2.columns)
```

```
Index(['STATEFP', 'COUNTYFP', 'TRACTCE', 'BLKGRPCE', 'GEOID', 'NAMELSAD',
       'MTFCC', 'FUNCSTAT', 'ALAND', 'AWATER', 'INTPTLAT', 'INTPTLON', 'NO2',
       'pop_hispanic', 'pop_nonhispanic', 'pop_white', 'pop_black',
       'pop_asian', 'pop_other', 'pop_total', 'income', 'geometry'],
      dtype='object')
```

In our case, this dataset contains the following columns:

> STATEFP: Census state FIPS code
>
> COUNTYFP: Census county FIPS code
>
> TRACTCE: Census tract code
>
> BLKGRPCE: Census block group number
>
> GEOID: Census block identifier; a concatenation of Census state FIPS code, Census county FIPS code, Census tract code, and Census block number
>
> NAMELSAD: Current name and the translated legal/statistical area description
>
> MTFCC: A 5-digit code assigned by the Census Bureau intended to classify and describe geographic objects
>
> FUNCSTAT: Current functional status
>
> ALAND: Current land area (in units of meters squared)
>
> AWATER: Current water area (in units of meters squared)

✓ 4s completed at 11:19 AM

Population-Weighted Averages

-

POPULATION OF SUBGROUP IN CENSUS BLOCK GROUP B

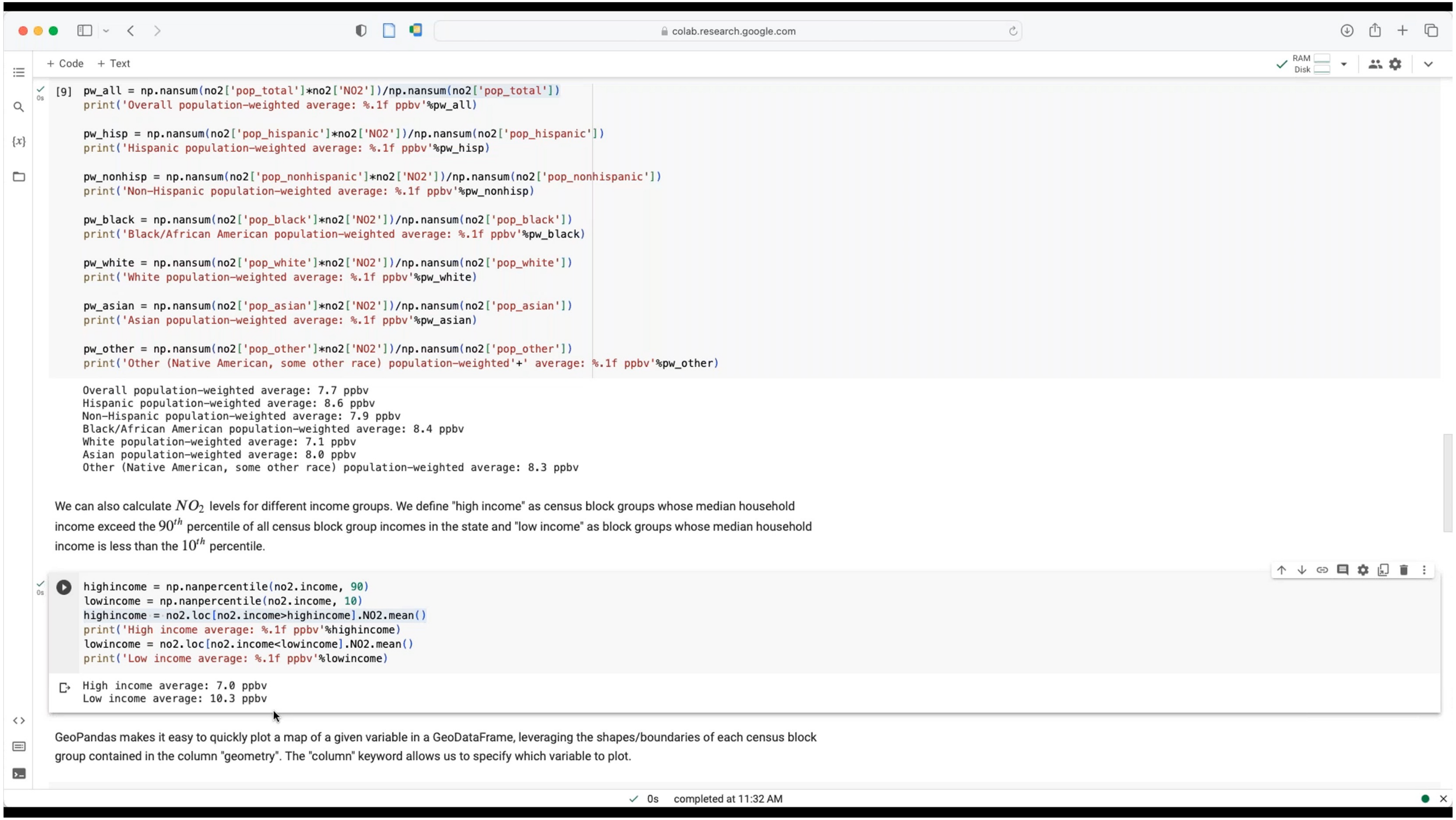
NO_2 CONCENTRATION IN CENSUS BLOCK GROUP B

$$= \frac{\sum_{b=1}^{k} pop_b \bullet NO_{2,\, b}}{\sum_{b=1}^{k} pop_b}$$

SUM OVER ALL CENSUS BLOCK GROUPS B

POPULATION OF SUBGROUP IN CENSUS BLOCK GROUP B

Python Demonstration: statistical analysis



```
pw_all = np.nansum(no2['pop_total']*no2['NO2'])/np.nansum(no2['pop_total'])
print('Overall population-weighted average: %.1f ppbv'%pw_all)

pw_hisp = np.nansum(no2['pop_hispanic']*no2['NO2'])/np.nansum(no2['pop_hispanic'])
print('Hispanic population-weighted average: %.1f ppbv'%pw_hisp)

pw_nonhisp = np.nansum(no2['pop_nonhispanic']*no2['NO2'])/np.nansum(no2['pop_nonhispanic'])
print('Non-Hispanic population-weighted average: %.1f ppbv'%pw_nonhisp)

pw_black = np.nansum(no2['pop_black']*no2['NO2'])/np.nansum(no2['pop_black'])
print('Black/African American population-weighted average: %.1f ppbv'%pw_black)

pw_white = np.nansum(no2['pop_white']*no2['NO2'])/np.nansum(no2['pop_white'])
print('White population-weighted average: %.1f ppbv'%pw_white)

pw_asian = np.nansum(no2['pop_asian']*no2['NO2'])/np.nansum(no2['pop_asian'])
print('Asian population-weighted average: %.1f ppbv'%pw_asian)

pw_other = np.nansum(no2['pop_other']*no2['NO2'])/np.nansum(no2['pop_other'])
print('Other (Native American, some other race) population-weighted'+' average: %.1f ppbv'%pw_other)
```

```
Overall population-weighted average: 7.7 ppbv
Hispanic population-weighted average: 8.6 ppbv
Non-Hispanic population-weighted average: 7.9 ppbv
Black/African American population-weighted average: 8.4 ppbv
White population-weighted average: 7.1 ppbv
Asian population-weighted average: 8.0 ppbv
Other (Native American, some other race) population-weighted average: 8.3 ppbv
```

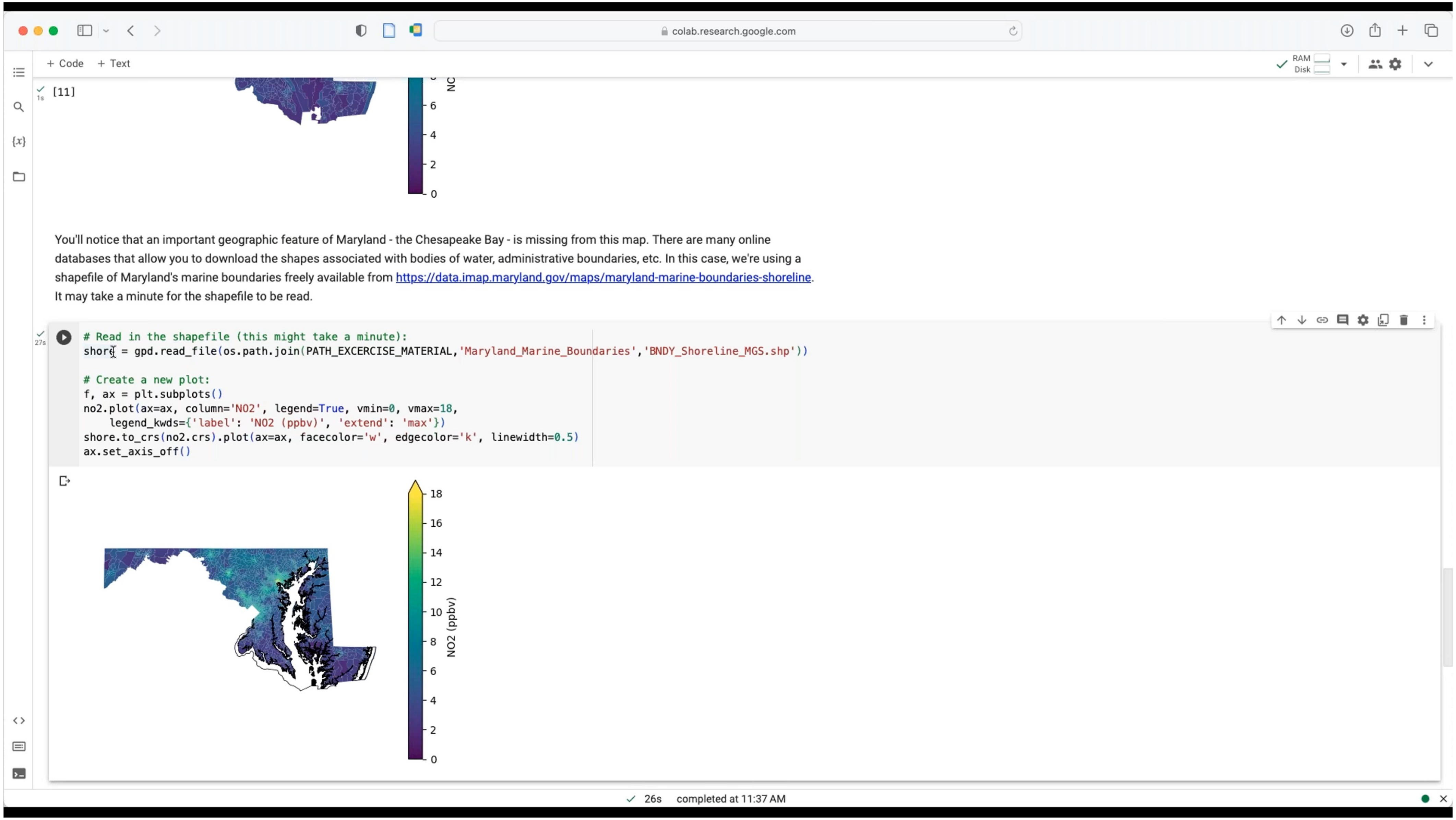
We can also calculate NO_2 levels for different income groups. We define "high income" as census block groups whose median household income exceed the 90^{th} percentile of all census block group incomes in the state and "low income" as block groups whose median household income is less than the 10^{th} percentile.

```
highincome = np.nanpercentile(no2.income, 90)
lowincome = np.nanpercentile(no2.income, 10)
highincome = no2.loc[no2.income>highincome].NO2.mean()
print('High income average: %.1f ppbv'%highincome)
lowincome = no2.loc[no2.income<lowincome].NO2.mean()
print('Low income average: %.1f ppbv'%lowincome)
```

```
High income average: 7.0 ppbv
Low income average: 10.3 ppbv
```

GeoPandas makes it easy to quickly plot a map of a given variable in a GeoDataFrame, leveraging the shapes/boundaries of each census block group contained in the column "geometry". The "column" keyword allows us to specify which variable to plot.

Python Demonstration: plotting maps using matplotlib



You'll notice that an important geographic feature of Maryland - the Chesapeake Bay - is missing from this map. There are many online databases that allow you to download the shapes associated with bodies of water, administrative boundaries, etc. In this case, we're using a shapefile of Maryland's marine boundaries freely available from https://data.imap.maryland.gov/maps/maryland-marine-boundaries-shoreline. It may take a minute for the shapefile to be read.

```
# Read in the shapefile (this might take a minute):
shore = gpd.read_file(os.path.join(PATH_EXCERCISE_MATERIAL,'Maryland_Marine_Boundaries','BNDY_Shoreline_MGS.shp'))

# Create a new plot:
f, ax = plt.subplots()
no2.plot(ax=ax, column='NO2', legend=True, vmin=0, vmax=18,
    legend_kwds={'label': 'NO2 (ppbv)', 'extend': 'max'})
shore.to_crs(no2.crs).plot(ax=ax, facecolor='w', edgecolor='k', linewidth=0.5)
ax.set_axis_off()
```

✓ 26s completed at 11:37 AM

Part 3:
Summary

Summary

- Tools for environmental justice applications:
  - EJScreen is an environmental justice screening and mapping tool
    - Accessible to those with limited knowledge of computer programming
    - Use without having to find and reformat data
  - Python
    - Requires knowledge of coding
    - Allows customization of mapping and statistical calculations
    - Freely available and open source

Satellite Data for Air Quality Environmental Justice and Equity Applications

Summary

Training Summary

- Remote sensing data are a valuable resource for environmental justice applications.
- Example applications include air quality, green space, lights at night, drought, heat and energy.
- Combining satellite remote sensing data with socio-economic information can provide evidence of disparities, inequality, and environmental injustice.
- Benefits of remote sensing data:
  - Extensive spatial coverage
  - Available in regions without ground-based observations
- Important considerations include temporal and spatial resolution, conversion to nose-level pollutants
- Demonstrated tools for analyzing remote sensing data for EJ applications:
  - EJScreen
  - Python

Homework and Certificates

- **Homework:**
  - One homework assignment
  - Opens on Sept. 6, 2023 (today)
  - Access from the training webpage
  - Answers must be submitted via Google Forms
  - **Due by Sept. 20, 2023**

- **Certificate of Completion:**
  - Attend all three live webinars (attendance is recorded automatically)
  - Complete the homework assignment by the deadline
  - You will receive a certificate via email approximately two months after completion of the course.

Acknowledgements

Gaige Kerr

Milken Institute School of Public Health

Susan Anenberg

Milken Institute School of Public Health

Shobhana Gupta

Associate Program Manager, NASA Equity and Environmental Justice

Ufuoma Ovienmhada

Massachusetts Institute of Technology

Qian Xiao

Univ. of Texas Health Science Center at Houston

Tanya Kreutzer Sayyed

University of Maryland, Baltimore County

Daniel Carrión

Yale Center on Climate Change and Health

Danielle Wood

Massachusetts Institute of Technology

Contact Information

Trainers:

- Melanie Follette-Cook
  - melanie.cook@nasa.gov
- Susan Anenberg
  - sanenberg@email.gwu.edu
- Qian Xiao
  - Qian.Xiao@uth.tmc.edu
- Carl Malings
  - carl.a.malings@nasa.gov
- Gaige Kerr
  - gaigekerr@email.gwu.edu
- Sarah Strode
  - sarah.a.strode@nasa.gov

- ARSET Website
- Follow us on Twitter!
  - @NASAARSET
- ARSET YouTube

Visit our Sister Programs:

- DEVELOP
- SERVIR

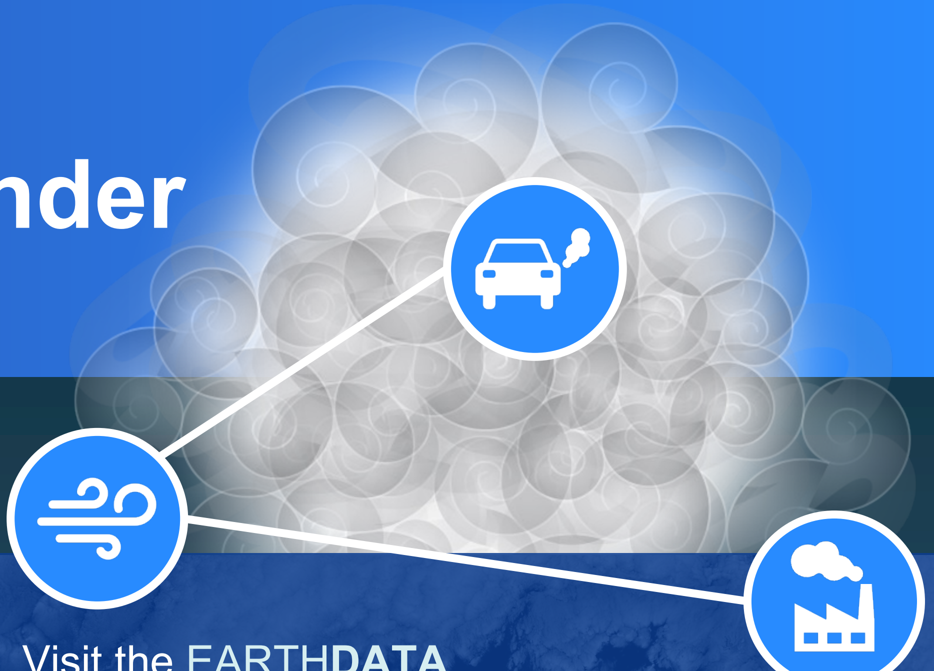
Resources

- ARSET health and air quality trainings
- CDC Environmental Justice Dashboard
- EJScreen
- EPA National Ambient Air Quality Standards
- George Washington University TROPOMI data visualization website
- NASA Earthdata
  - NASA Air Quality Data Pathfinder
  - NASA Environmental Justice Backgrounder
- NASA Giovanni
- NASA HAQAST SD4EJ
- NASA SEDAC (Socioeconomic Data and Applications Center)
- NASA Worldview
- NOAA Aerosol Watch

EARTH**DATA** Offers

The Air Quality Data Pathfinder

for Your Research & Applications



Air pollution is one of the largest global environmental and health threats. NASA provides data resources to better understand the movement of pollutants and the impact of events leading to poor air quality. This Pathfinder helps you access, and leverage data acquired from NASA's satellite, airborne, and ground-based missions and campaigns.

Available Data Types:

- Aerosols
- Trace Gases (e.g., Nitrogen Dioxide, Sulfur Dioxide, Carbon Monoxide, etc.)
- Weather (e.g., Air Temperature, Clouds, Precipitation, etc.)
- Land Surface (e.g., Soil Moisture, Surface Reflectance, Topography, etc.)
- Human Dimensions

Data are from satellites, airborne and ground-based platforms, and models, including:

- AIRS
- AMSR2
- GPM
- MODIS
- OLI/TIRS
- OMI
- OMPS
- SMAP
- TROPOMI
- VIIRS
- GEOS
- MERRA-2

Visit the EARTH**DATA**

Air Quality Data Pathfinder

for more information:

- Commonly Used Datasets for Air Quality Research and Applications
- Tools for Using Data
- Resources for Applying and Connecting NASA Data
- GIS Resources
- Tips for Getting Help and Connecting with NASA experts
- Tutorials and more!

APPLIED REMOTE SENSING TRAINING PROGRAM

TRAIN - EMPOWER - ADVANCE

Thank You!