March 14th

Assessing Uncertainty in Indoor Radon Exposure Estimates: Implications for Radiation Epidemiology

Heidi Hanson

Computational Sciences and Engineering Division

U.S. DEPARTMENT OF ENERGY

ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY

Two primary datasets

Utah

- Point level data for approximately 60k residences
- Pre-mitigation
- Basement or bottom level

Pennsylvania

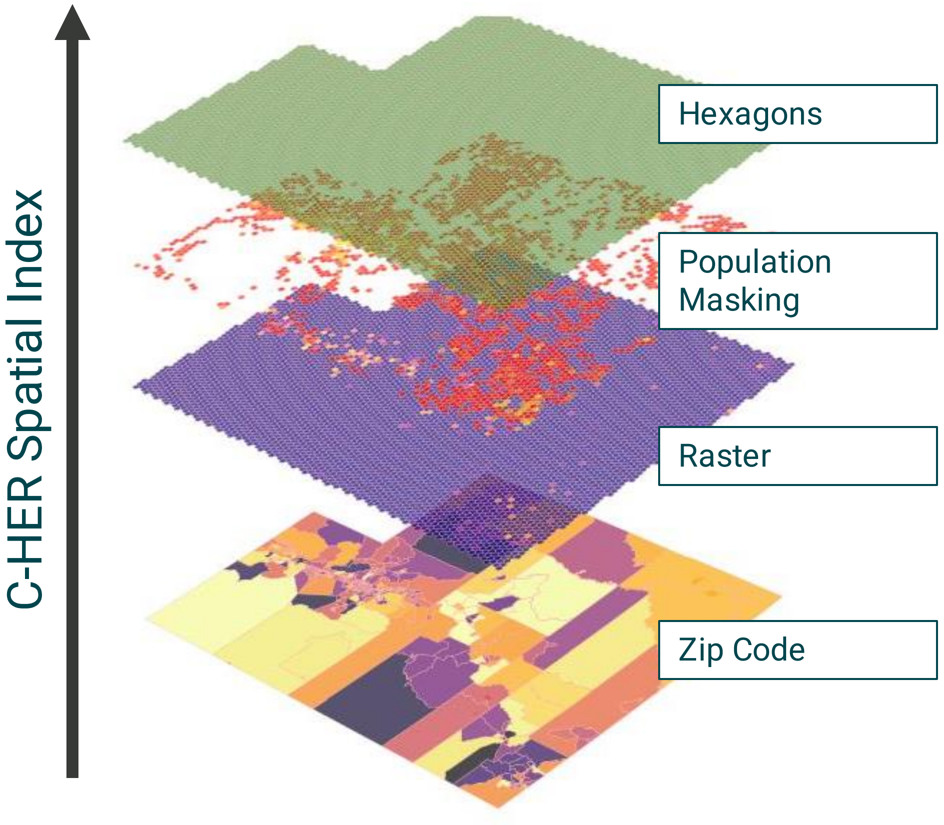
- Individual level estimates with zip code for approximately 720k residences
- Pre-mitigation
- Basement or bottom level

Error in Indoor Radon Exposure Assessment

Difference between the observed exposure and the true exposure

Residential Radon Exposure		
Concentration	Exposure	Organ dose

Uncertainty in Exposure Modeling: Indoor Radon Concentration



Temporal Variation

- Fluctuations in weather, meteorological factors, and ventilation affect radon concentrations.

Spatial Variation

- Variation in soil, bedrock and water table.
- Housing characteristics: Basement, age of structure, quality of structure.

Measurement Variation

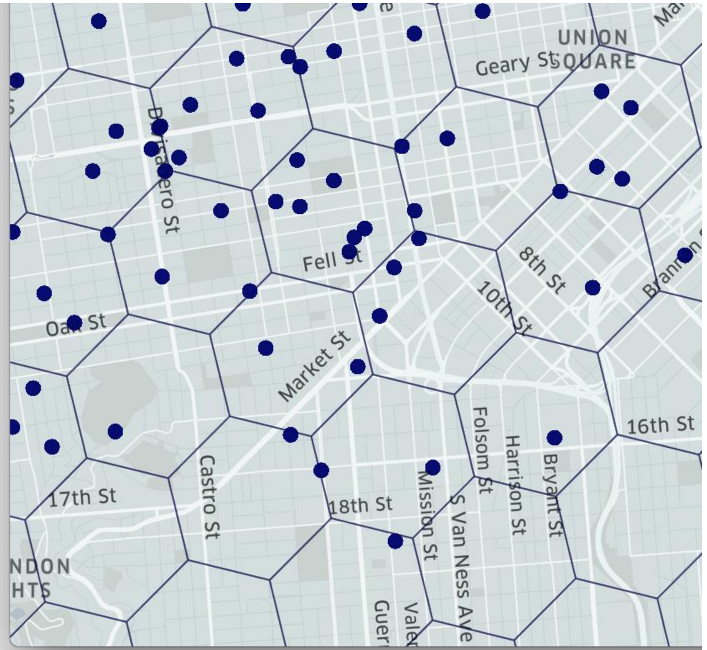
- Alpha-track detector, activated charcoal detector, electret ion chamber, electronic integrating device, continuous radon monitor, etc.
- Short/long term
- Adherence to standards

Geoprocessing of data

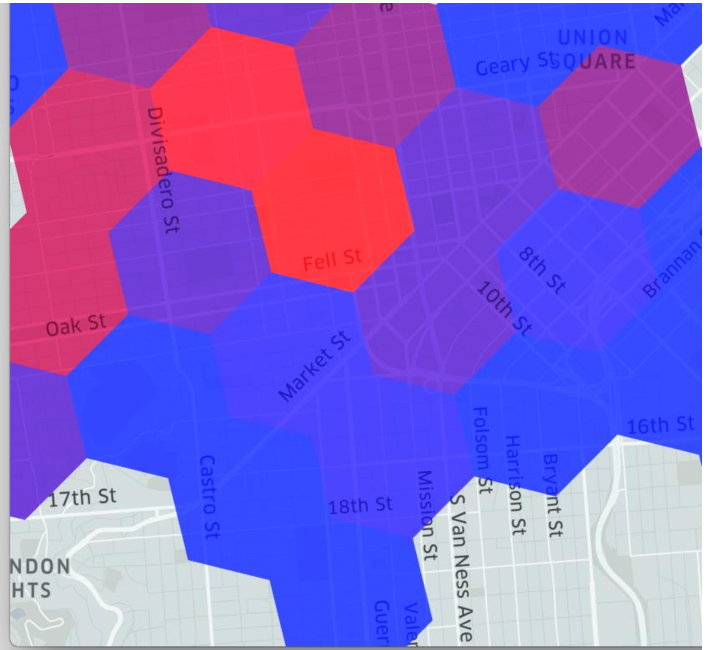
- Aggregating data leads to loss of detailed insights and important information.

Stacking Environmental Datasets at small spatial scales with Uber H3 Hex

Coordinate information



Overlay with hexagon “fishnet”



Aggregate to new polygons

A spatial join can be used to map all datasets to the same spatially linked polygon. All environmental datasets can be stacked to create multi-exposure measures for a region.

https://www.uber.com/en-FR/blog/h3/

Population Masking – A clever way to aggregate

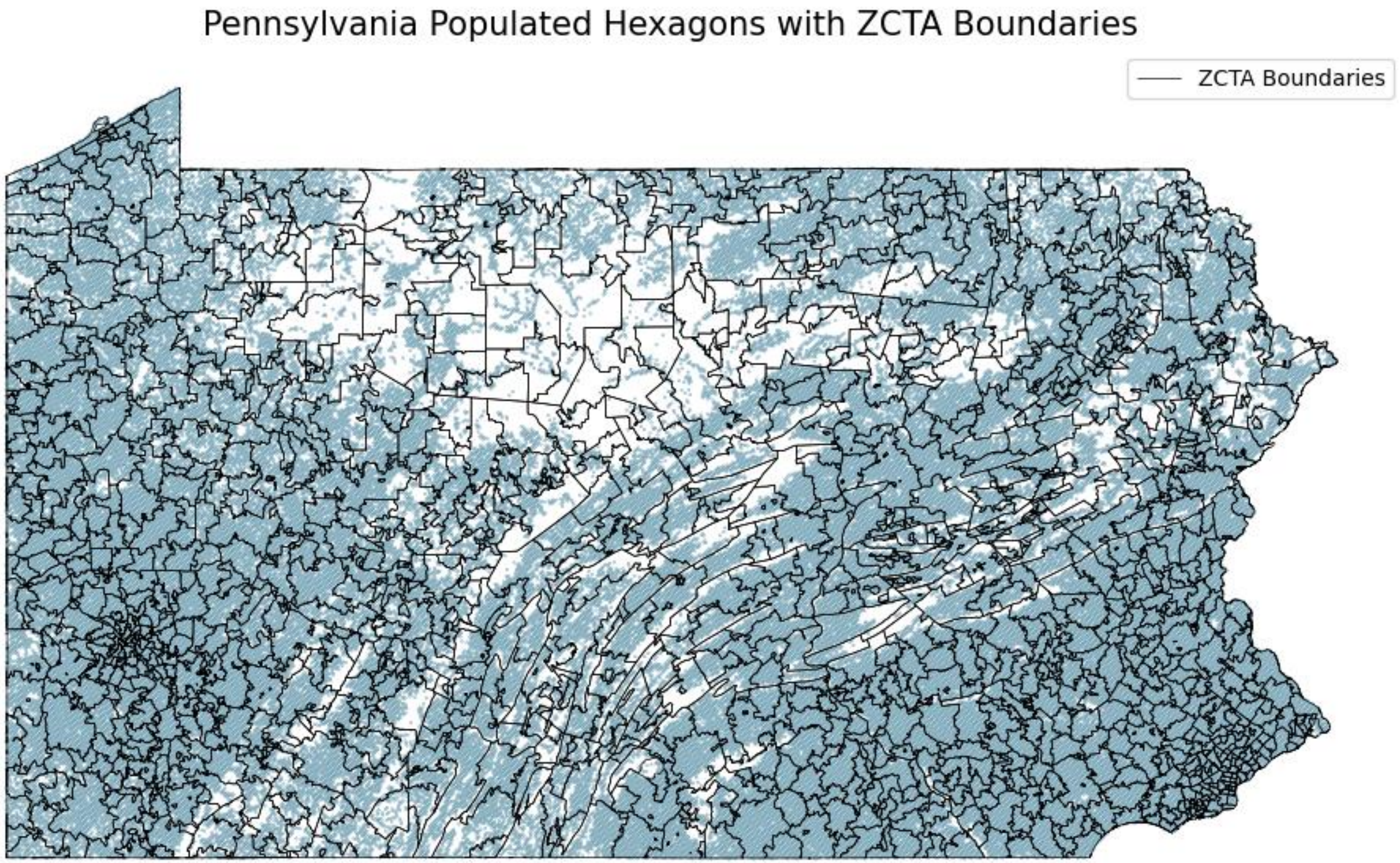
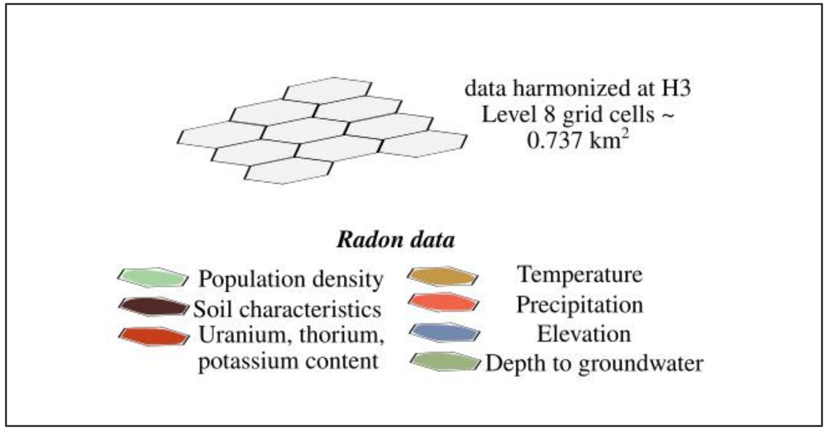


Image: Jeremy Logan

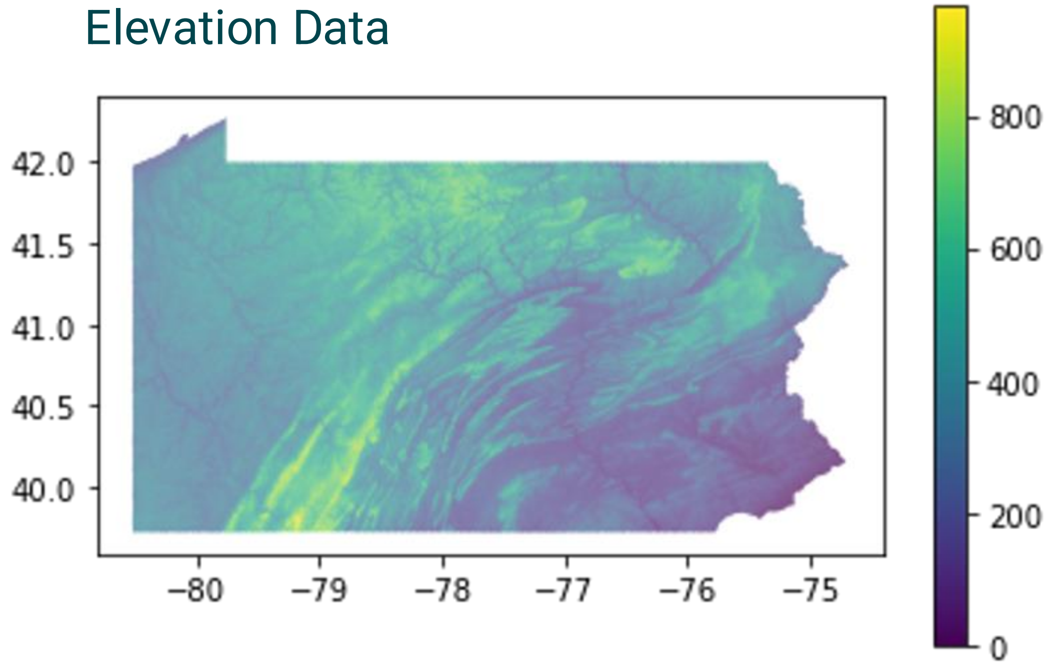


Rose, A., Weber, E., Moehl, J., Laverdiere, M., Yang, H., Whitehead, M., Trombley, N., Sims, K., Whitlock, C., & Bhaduri, B. (2019). *LandScan USA 2018* [Data set]. Oak Ridge National Laboratory. https://doi.org/10.48690/1523375

Traditional to estimating indoor radon exposure

- Aggregating to the zip code or county level leads to aggregation bias, loss of information, inaccurate estimates.
- Removal of variance on both sides of the equation.
- Artificially inflated model fit statistics.

Elevation Data



Elevation Data: Zip Code

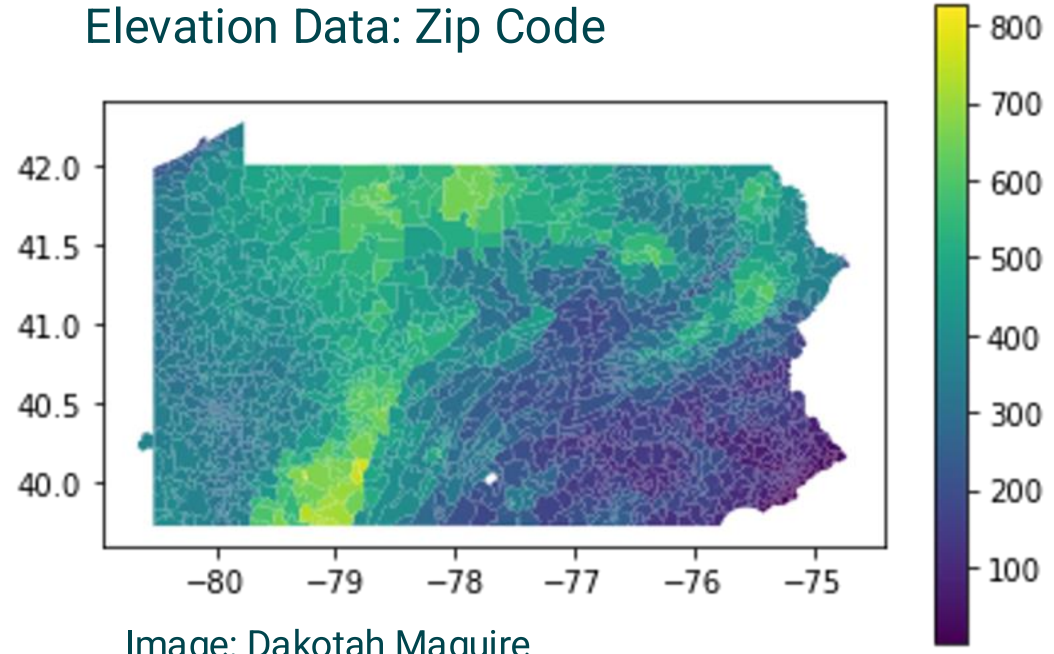
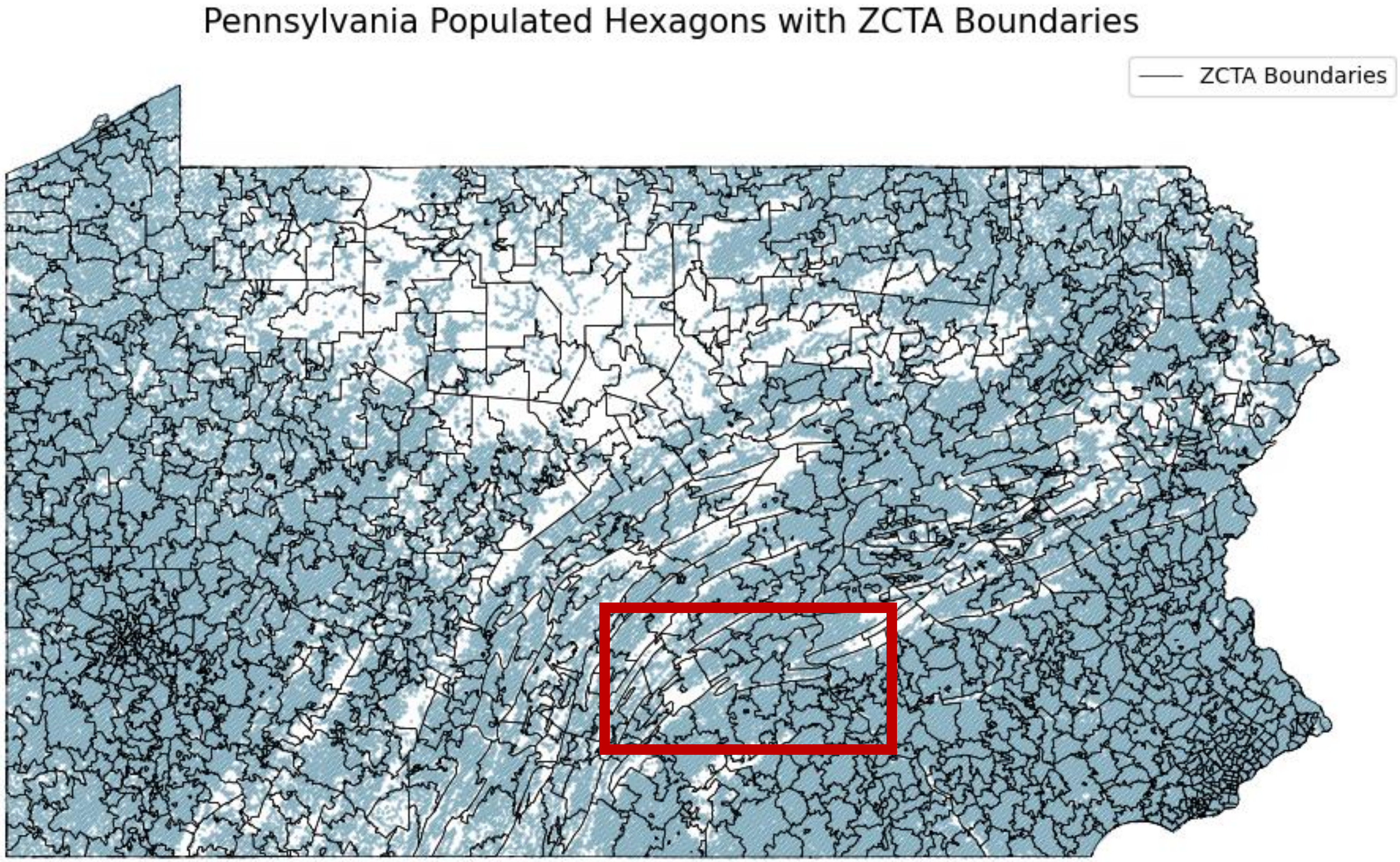
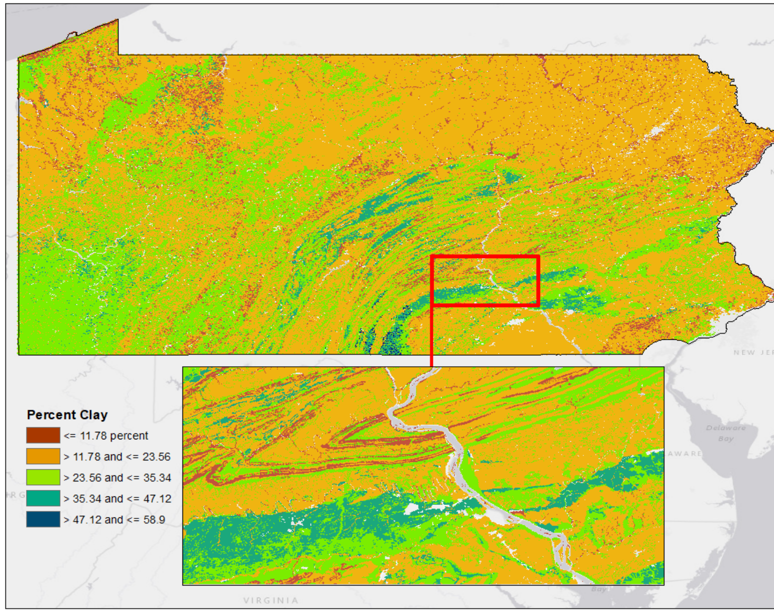


Image: Dakotah Maguire

Population Masking – A clever way to aggregate





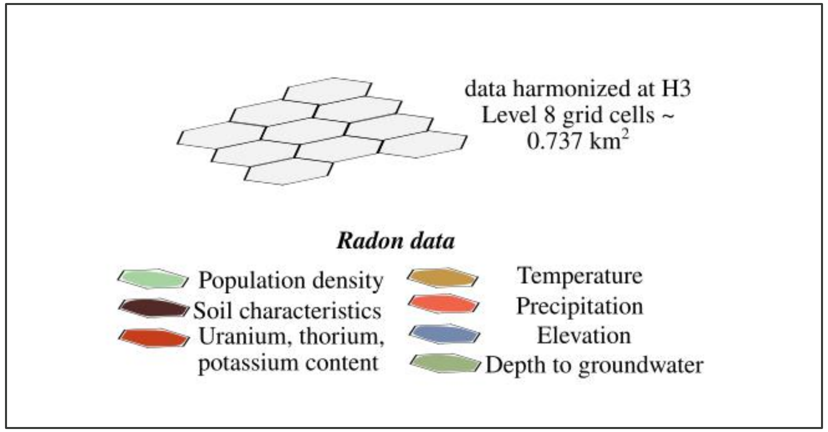


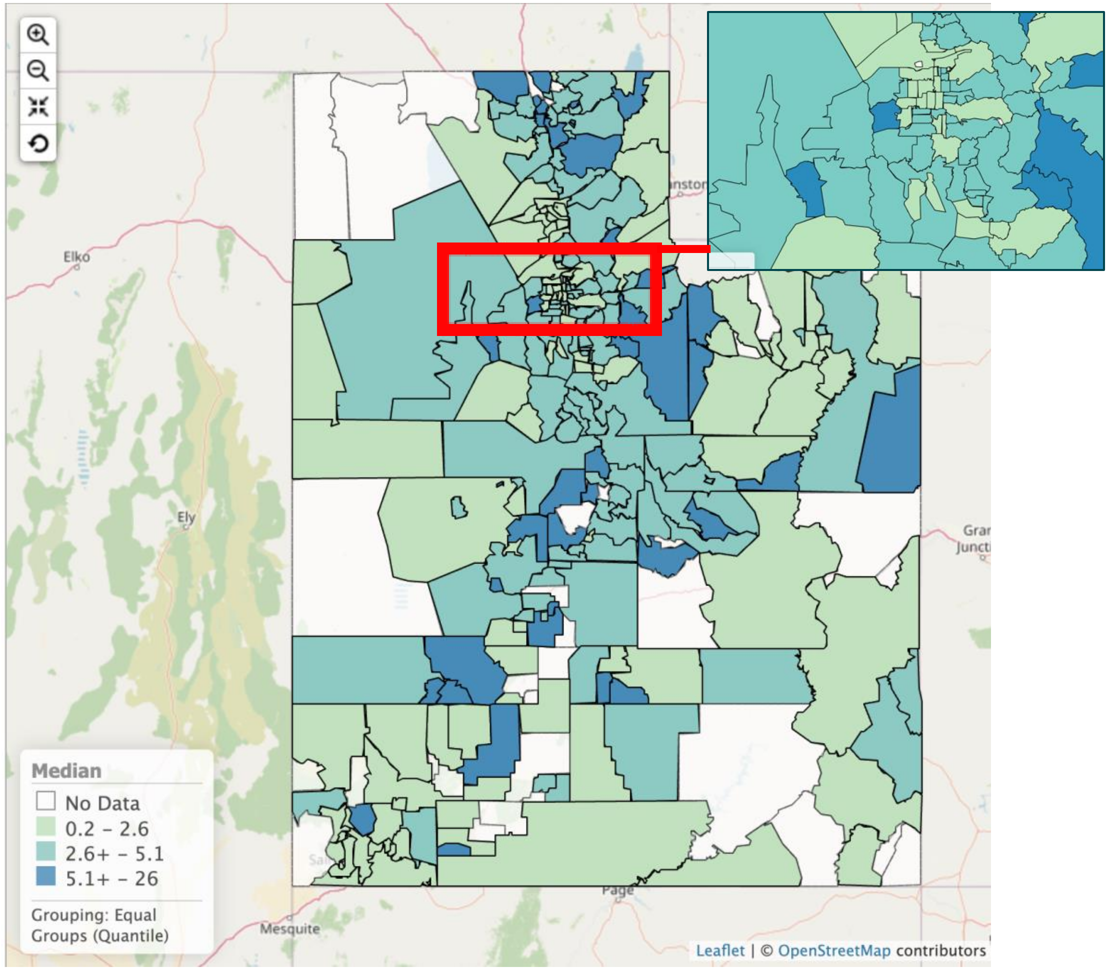
Image: Jeremy Logan, Dakotah Maguire, and Zach Fox

Rose, A., Weber, E., Moehl, J., Laverdiere, M., Yang, H., Whitehead, M., Trombley, N., Sims, K., Whitlock, C., & Bhaduri, B. (2019). *LandScan USA 2018* [Data set]. Oak Ridge National Laboratory. https://doi.org/10.48690/1523375

Measuring indoor radon exposure at the zip code vs. hexagon

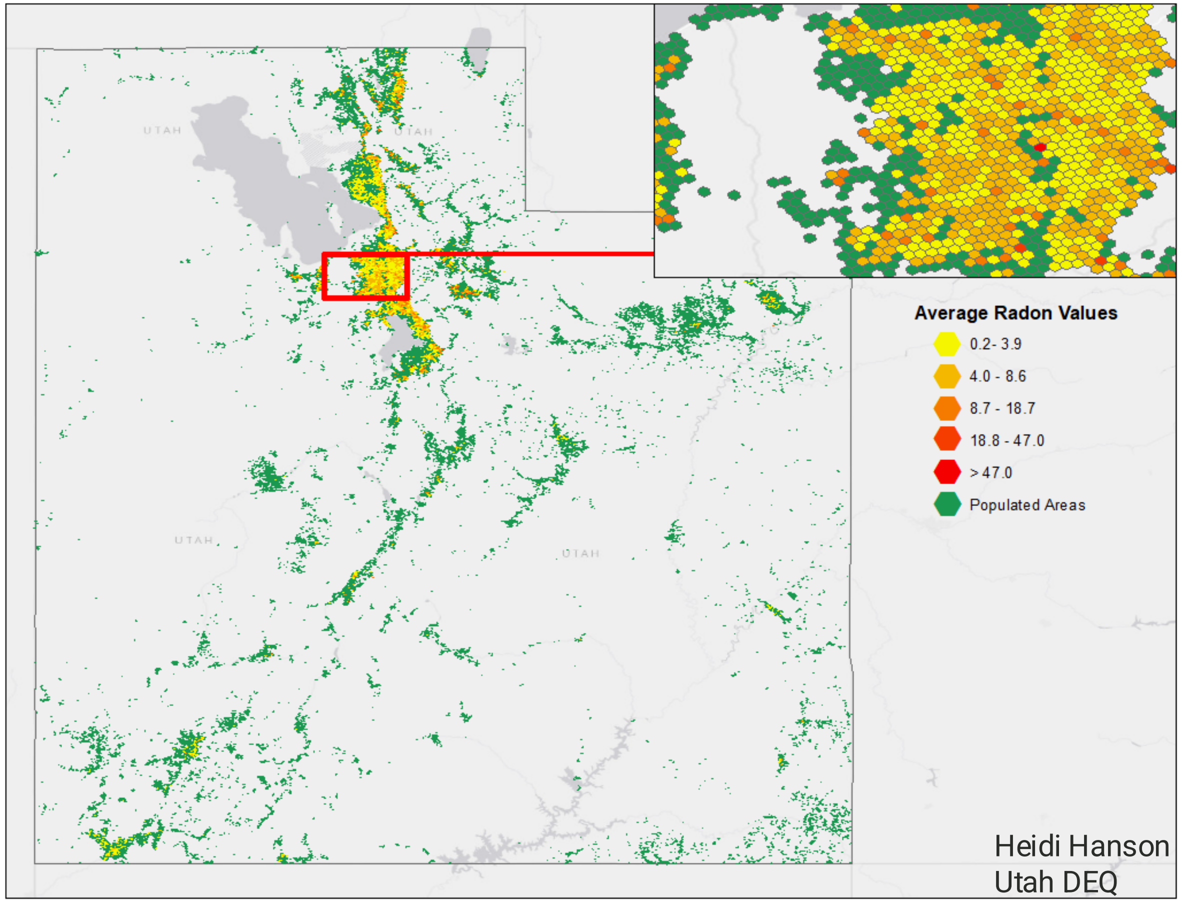
Median Radon Levels

Zip Code



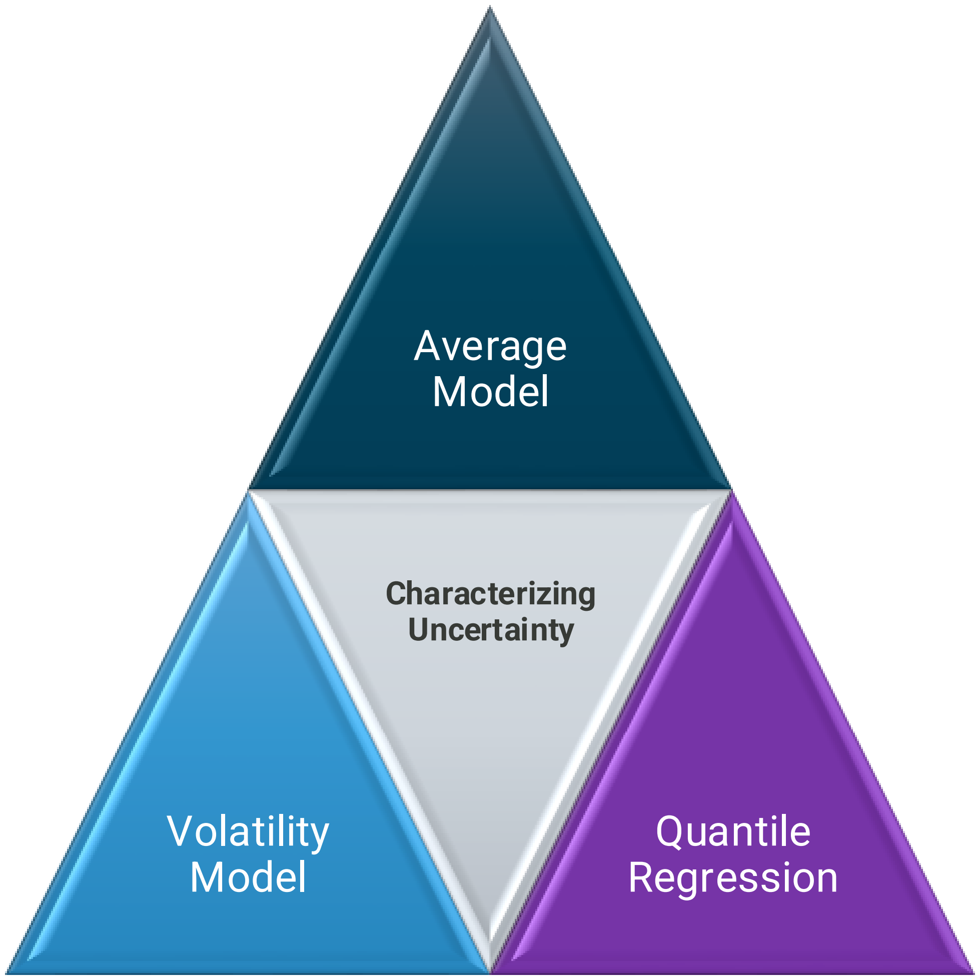
Mean Radon Levels

Uber Hex Res 8



https://epht.health.utah.gov/epht-view/query/result/radon/RadonTestResult/Count.html

Quantifying uncertainty and identifying factors that increase uncertainty in estimates.



- Aggregation isn't great – but it is often necessary to protect privacy.

- When it is necessary, how can we characterize the uncertainty that we are seeing in our model?
  - Not random
  - Varies by space and time – but how?

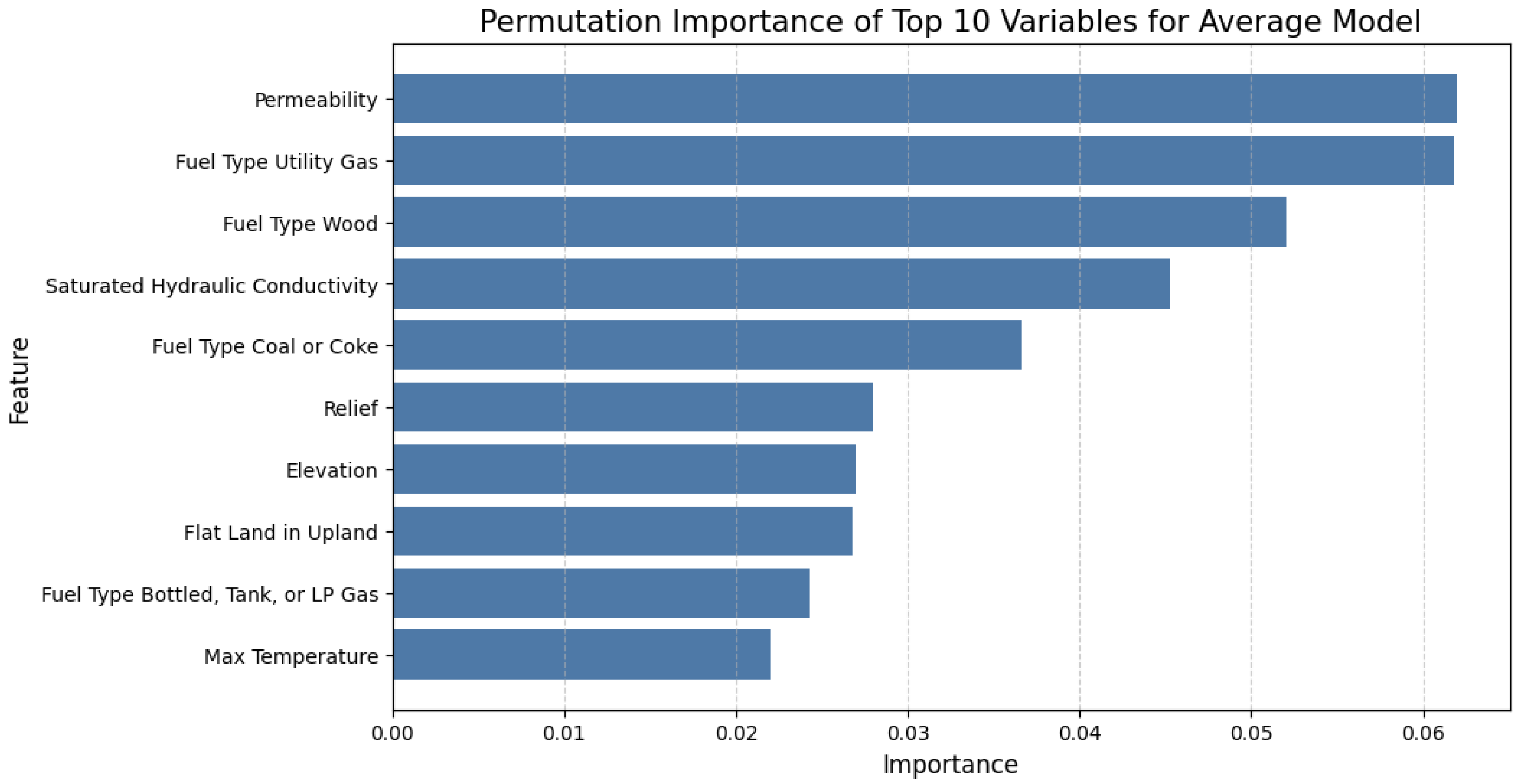
- Approach 1: Triangulating Evidence.
  - Modeling Average Radon Concentration: Random Forest

  - Modeling Volatility of Estimates: Random Forest with variance of measures as the dependent and independent variables.

  - Quantile Regression: Estimating the distribution of radon concentration (50%, 75%, 95%)

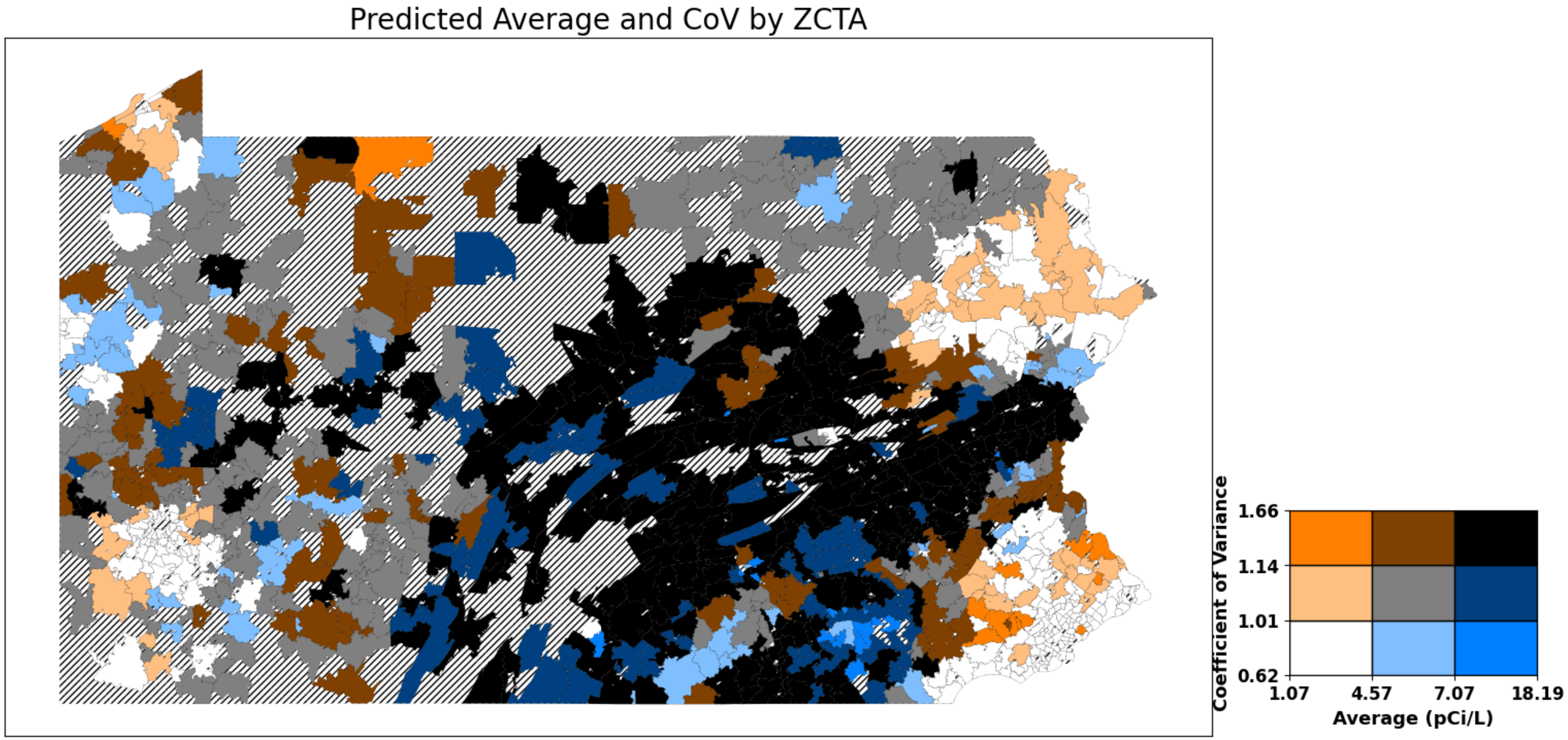
Average Model at the ZCTA Level

Variable Importance and Model Fit



	ZCTA Average		Individual Model (Exact location is unknown)	
	5-fold CV	Group 5-fold (ZCTA) CV	5-fold CV	Group 5-fold (ZCTA) CV
RMSE	2.67 (0.19)	3.17 (0.14)	7.80 (0.15)	7.86 (0.30)
R^2	0.67 (0.022)	0.53 (0.021)	0.12 (0.0020)	0.10 (0.0079)
MAPE	20.68 (0.42)	27.71 (0.81)	166 (1.59)	167 (2.70)

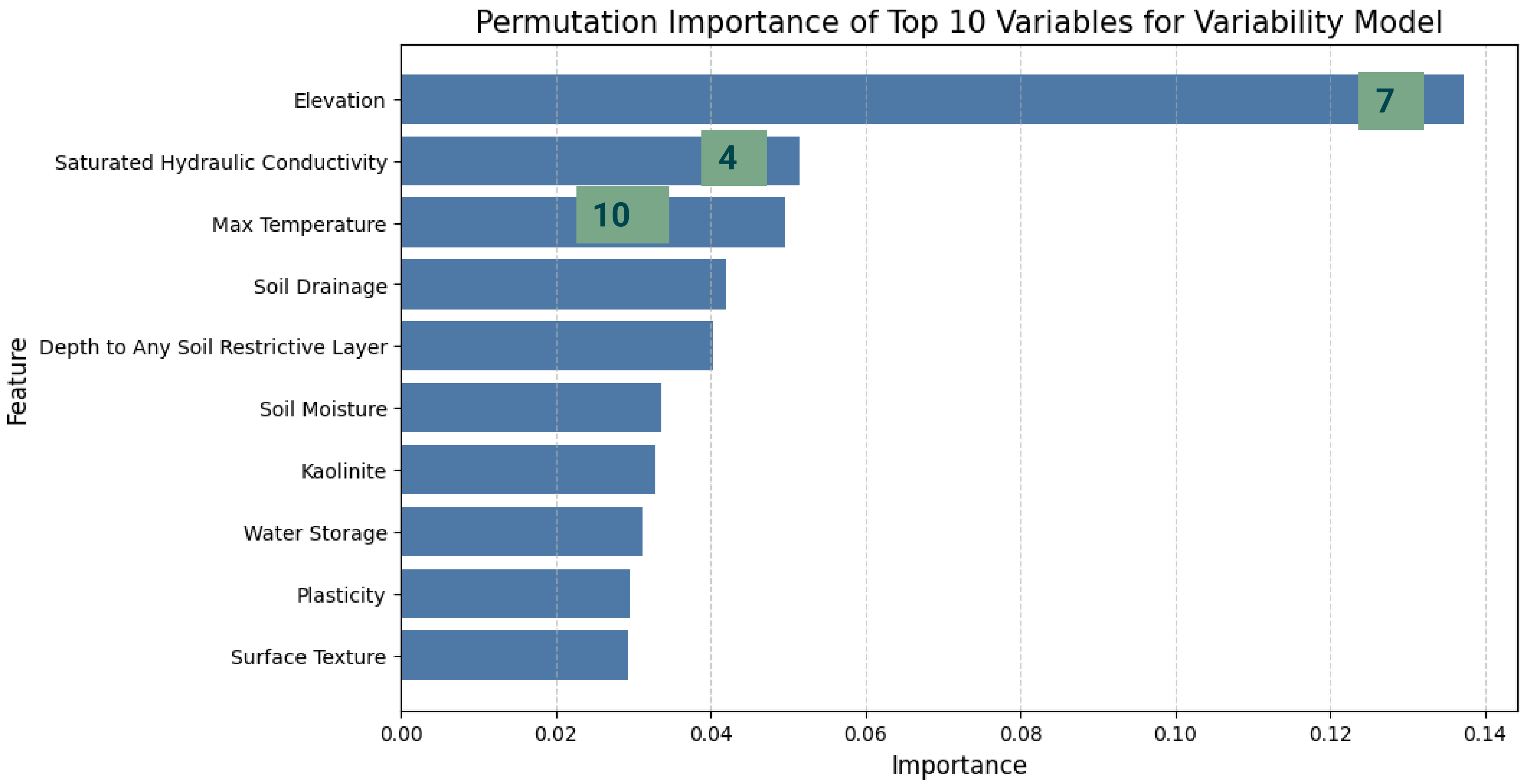
2D plot of Average and Variance of Radon Concentration



Average and CoV of radon concentration in January by ZCTAs in Pennsylvania. The hatched area illustrate where there are less than three observations

Volatility Model at the ZCTA Level

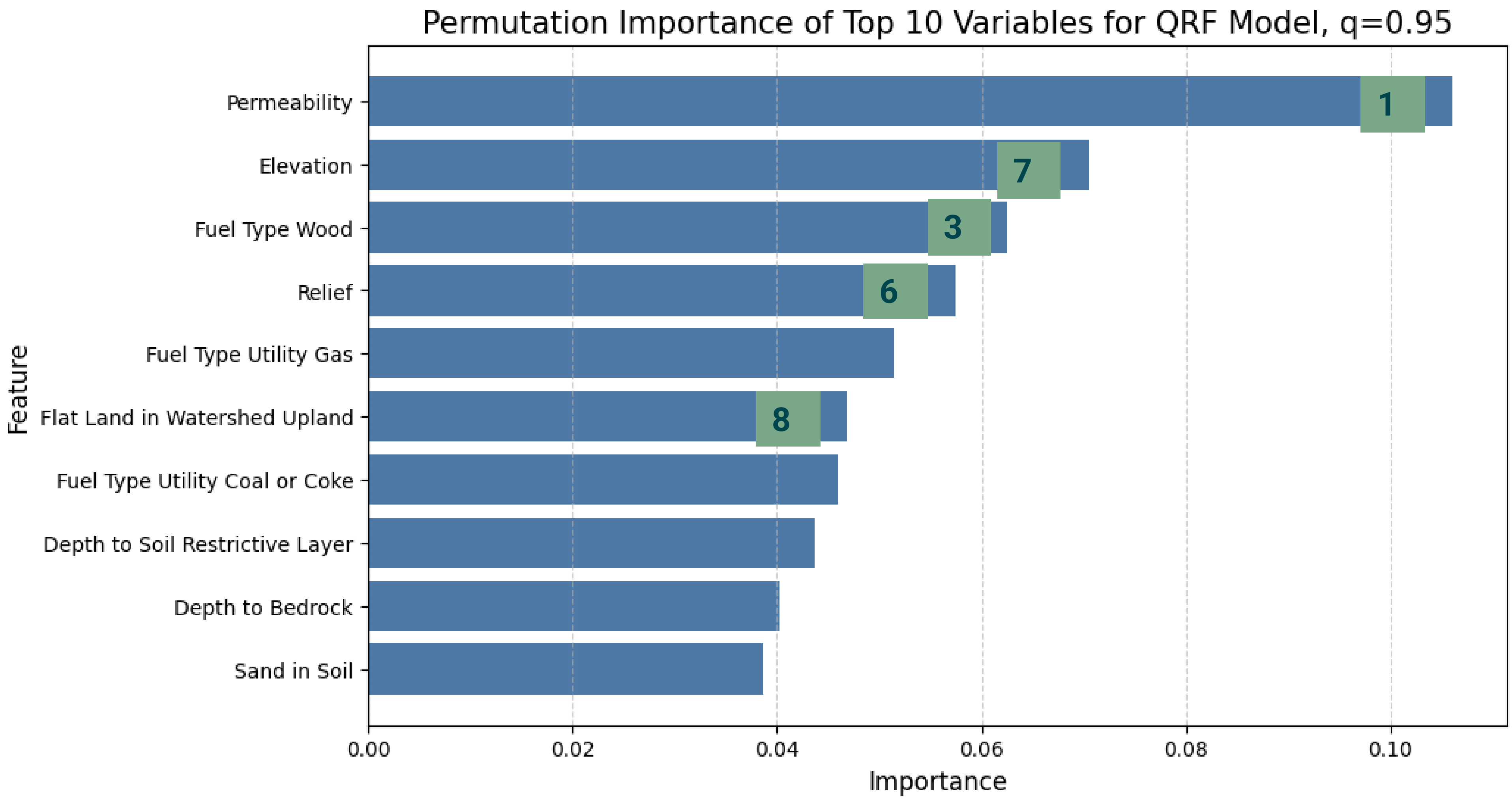
Variable Importance



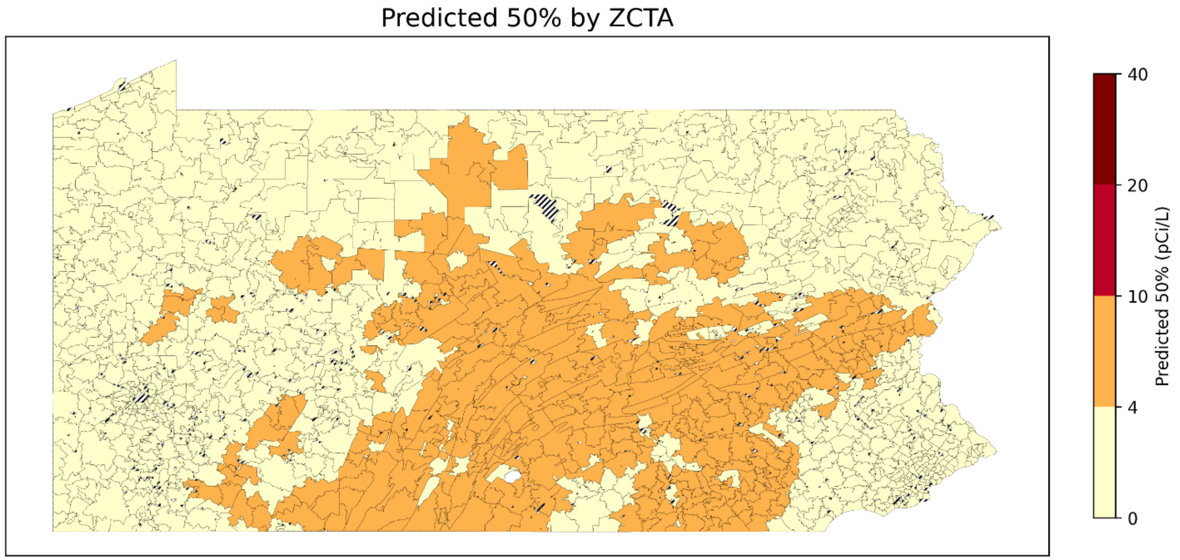
Heechan Lee, Dakotah Maguire, Jeremy Logan, Shaheen Dewji, Greeshma Agastha, Heidi Hanson

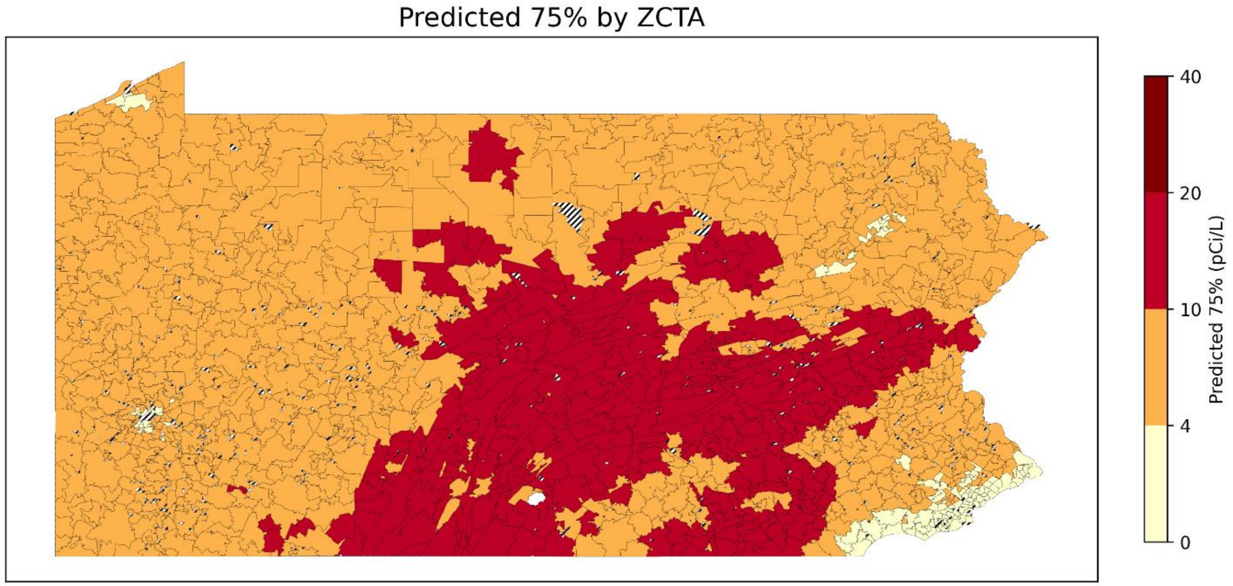
Quantile Regression Forest at the ZCTA Level

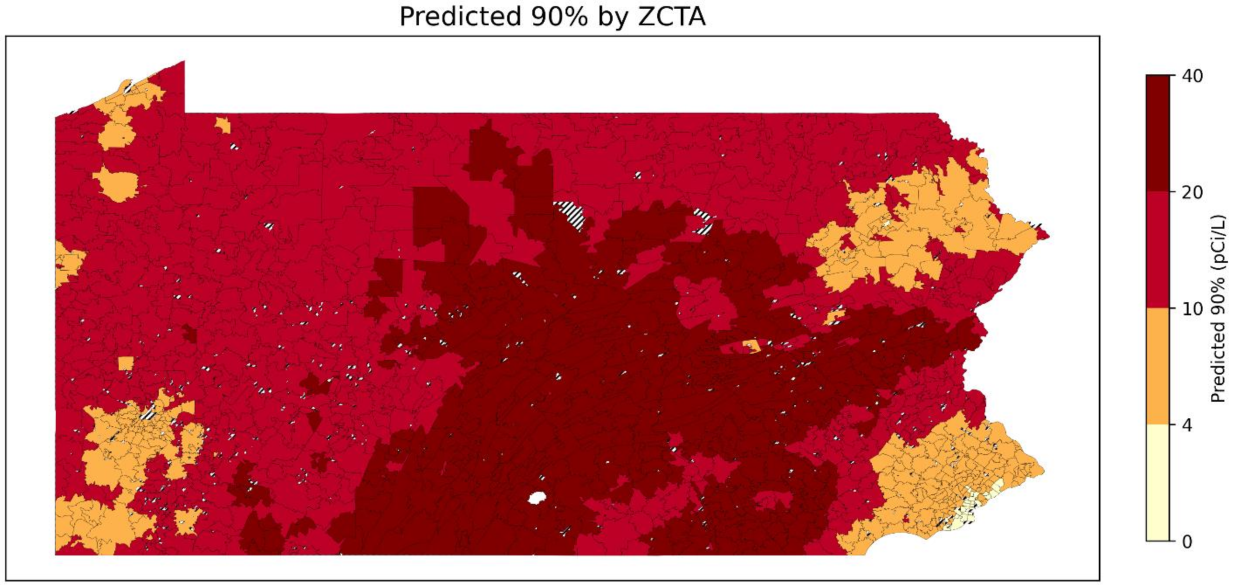
Variable Importance



Quantile Regression Results







OAK RIDGE National Laboratory

Heechan Lee, Dakotah Maguire, Jeremy Logan, Shaheen Dewji, Greeshma Agastha, Heidi Hanson

Error in Indoor Radon Exposure Assessment

Difference between the observed exposure and the true exposure

Residential Radon Exposure		
Concentration	Exposure	Organ dose

Longitudinal Residential History Data for Surveillance, Epidemiology, and End-Results (SEER) Registries

Over 25 years of LexisNexis Residential History data linked to cancer incidence

- 11 SEER registries have been linked (3.2 million individuals diagnosed from 2005 - 2022)
- High quality data from 1995 – 2020
- 83% are geocoded to the point location
- Our team is linking radon, air pollution, and toxic release estimate to the addresses

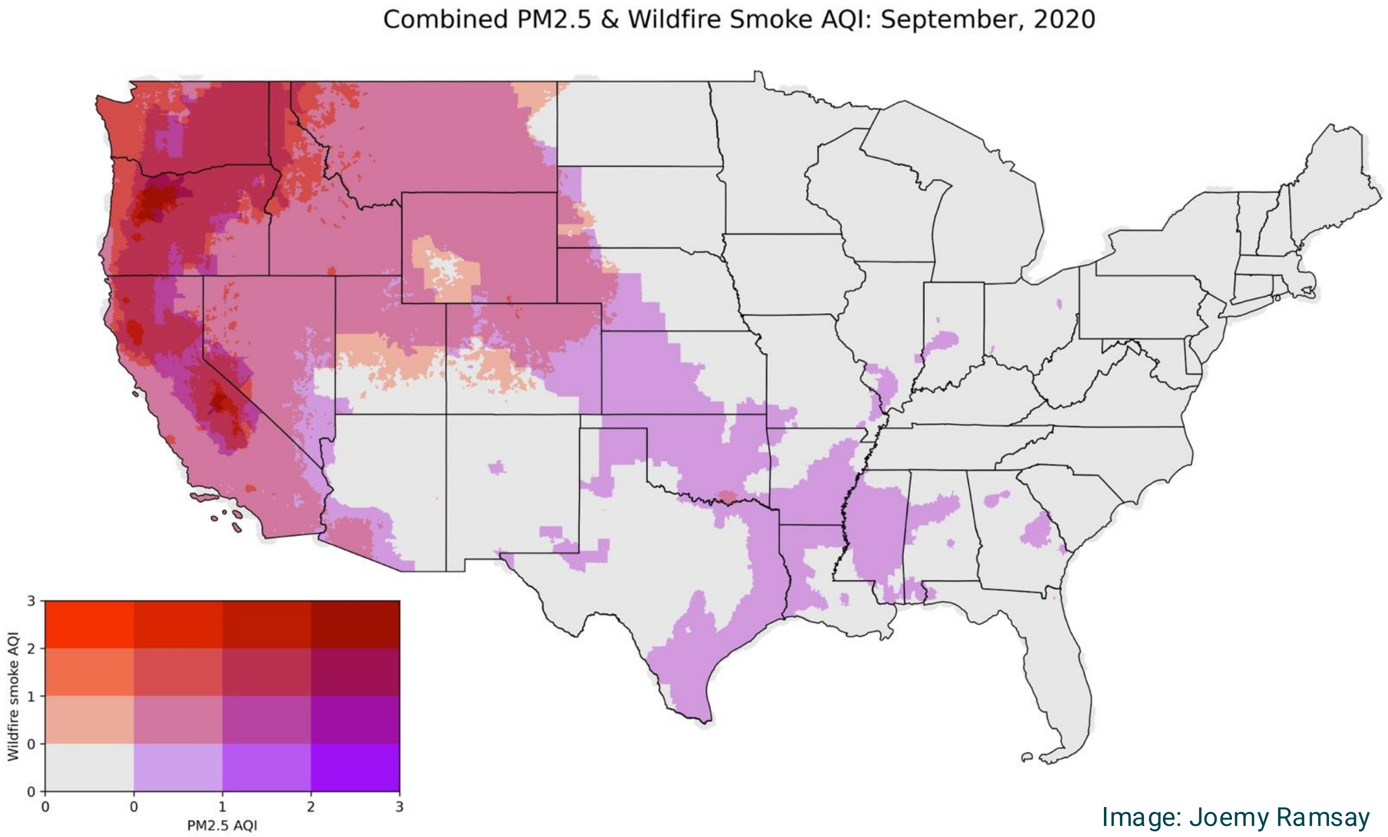
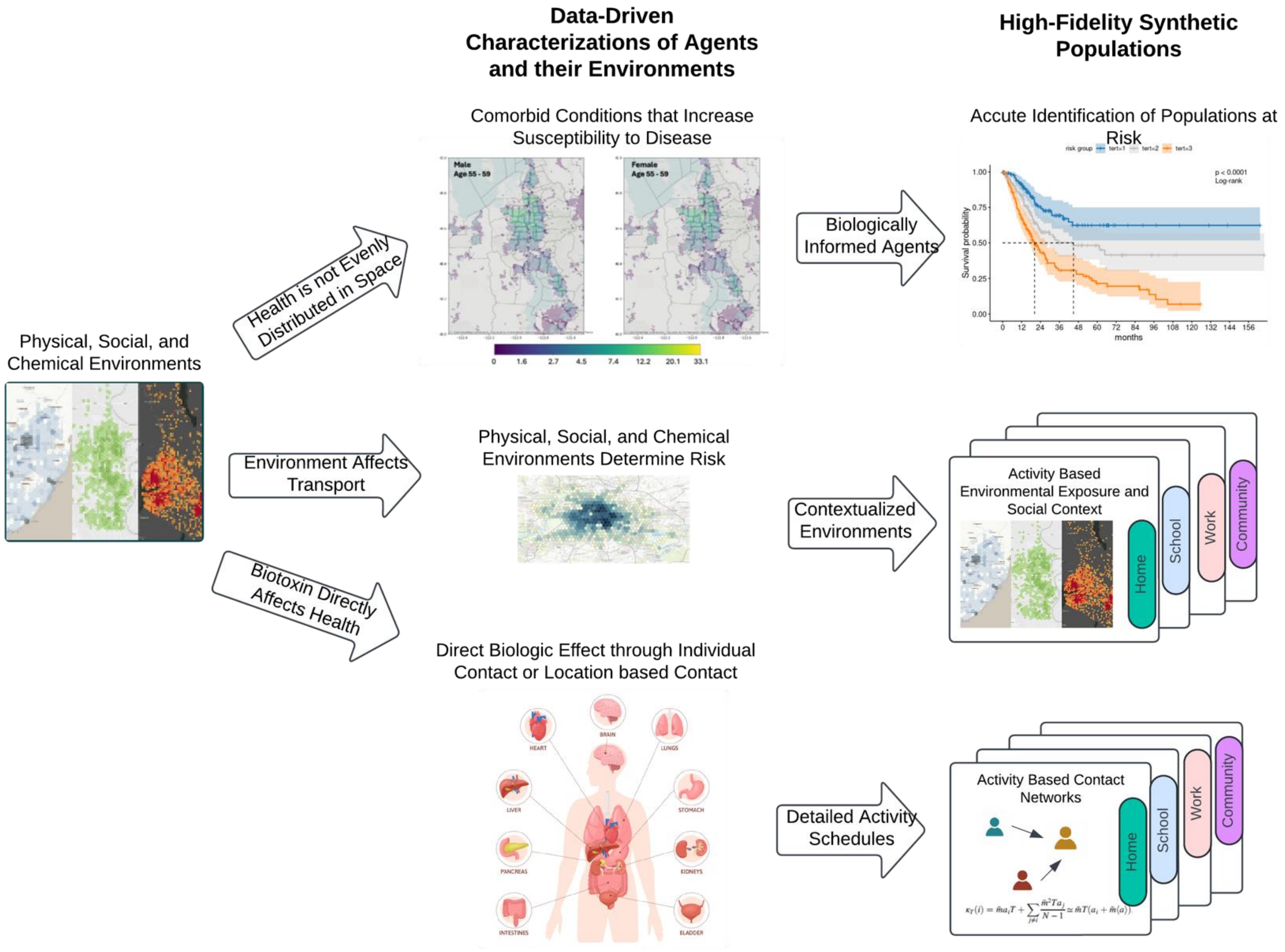


Image: Joemy Ramsay

EHRLICH tools being designed to simulate population movement and generate a distribution of exposure profiles that can be used to quantify uncertainty of exposures over a specified period of time

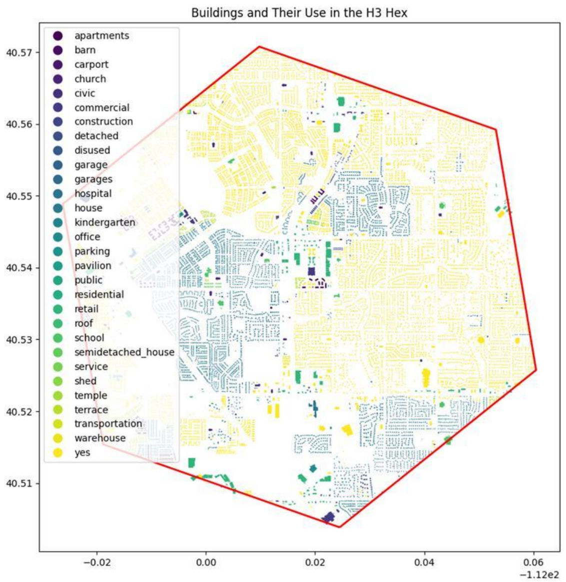


Heidi Hanson, ORNL

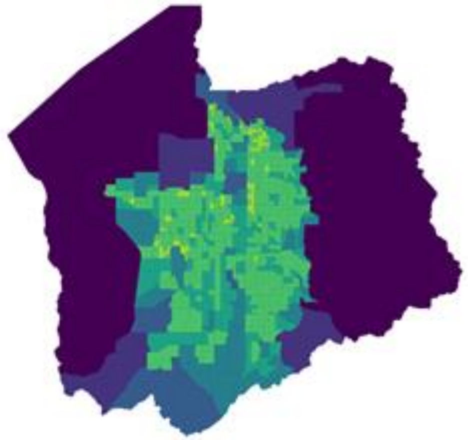
Weekday Activity Schedules

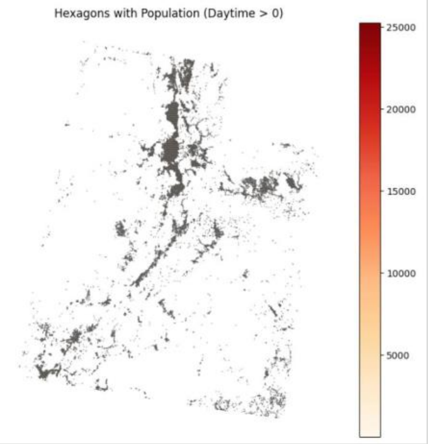
Create a **pool of synthetic daily activity schedules for each weekday** from NHTS based on

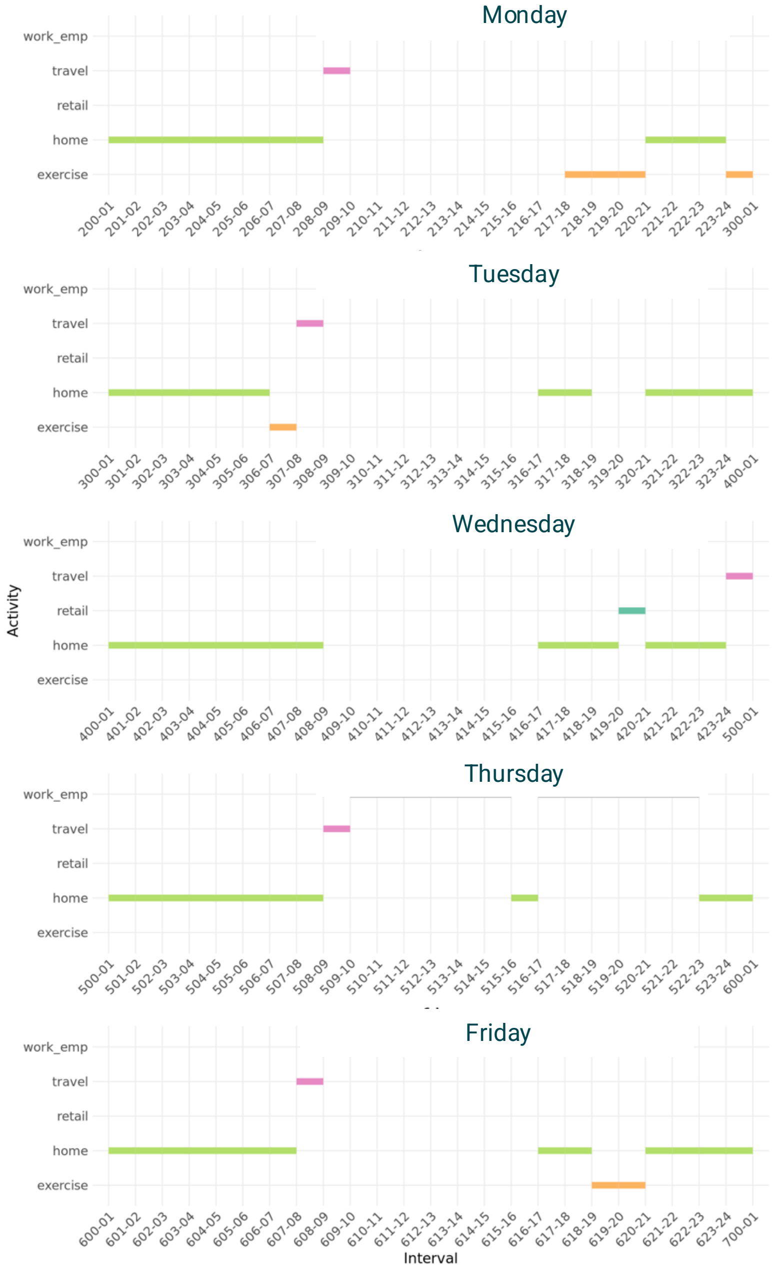
- **Geographic context**
- **Primary role during week**
- **Occupation category (workers)**



Synthetic Population of Salt Lake County, Utah







Joe Tucillo, ORNL

- Worker in professional/scientific/technical services
  - MSA with 1m+ pop and no rail
  - Home BG pop density: [10k – 25k)

PLACING A PULSE ON POPULATION HEALTH

Federated Learning

Synthetic Data Generation

Privacy Preserving AI

Identify existing patterns and anomoly detection

Predict patterns relating to environmental exposures

High-performance computing for modeling and simulating health outcomes

C-HER
CENTRALIZED HEALTH AND ENVIRONMENTAL REPOSITORY

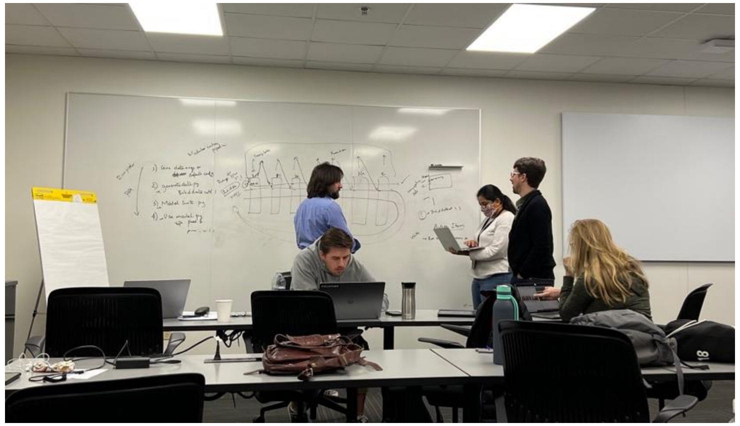
MOSSAIC

National Cancer Institute
Lynne Penberthy (PI)
Elizabeth Hsu (Technical Lead)
Valentina Petkov
Serban Negoita
Ola Adeyemi
Sylkk Ansah
Sarah Bonds

IMS
Linda Coyle
Jennifer Stevens
Scott Depuy
Rusty Sheilds
Gary Beverungen

Los Alamos National Laboratory
Jamaludin Mohd Yusof
Sayera Dhaubhadel
Tanmoy Bhattacharya

Oak Ridge National Laboratory
Heidi Hanson (PI)
John Gounley (Technical Lead)
Shelaine Curd (Program Manager)
Georgia Tourassi
Joe Lake
Adam Spannaus
Dakota Murdock
Zachary Fox
Patrycja Krawczuck
Dakotah Maguire
Jordan Miller
Mayanka Chandra Shekar
Noah Schaefferkoetter
Sajal Dash
Isaac Lyngaas
Abhishek Shivanna
Robert Bridges
Christopher Stanley
Vandy Tombs
Christoph Metzner



EHRLICH

Oak Ridge National Laboratory
Heidi Hanson (PI)
John Gounley (Technical Lead)
Shelaine Curd (Program Manager)
Adam Spannaus
Zachary Fox
Patrycja Krawczuck
Dakotah Maguire
Mayanka Chandra Shekar
Robert Bridges
Christopher Stanley
Vandy Tombs
Sudip Seal
James Nutaro
Sifat Moon
Christoph Metzner

Los Alamos National Laboratory
Jamal Mohd-Yusof
Cristina Garcia Cardona

Argonne National Laboratory
Rick Stevens
Thomas Brettin

University of Utah
James VanDerslice
Joemy Ramsay

University of Arizona
Nirav Merchant
Ravi Tandon

University of Southern California
Rima Habre

Duke
Amanda Randles

University of Chicago/Morehouse School of Medicine
Lilly Immergluck