Integrating Multilevel Data to Assess Massachusetts Food Vulnerability

Qian Zhao

Department of Mathematics and Statistics

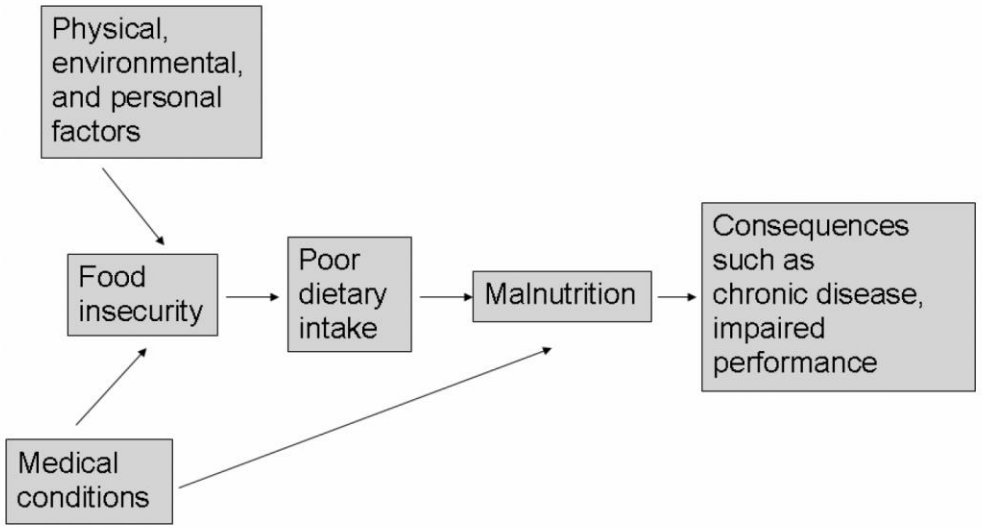
Chaitra Gopalappa

Department of Mechanical and Industrial Engineering

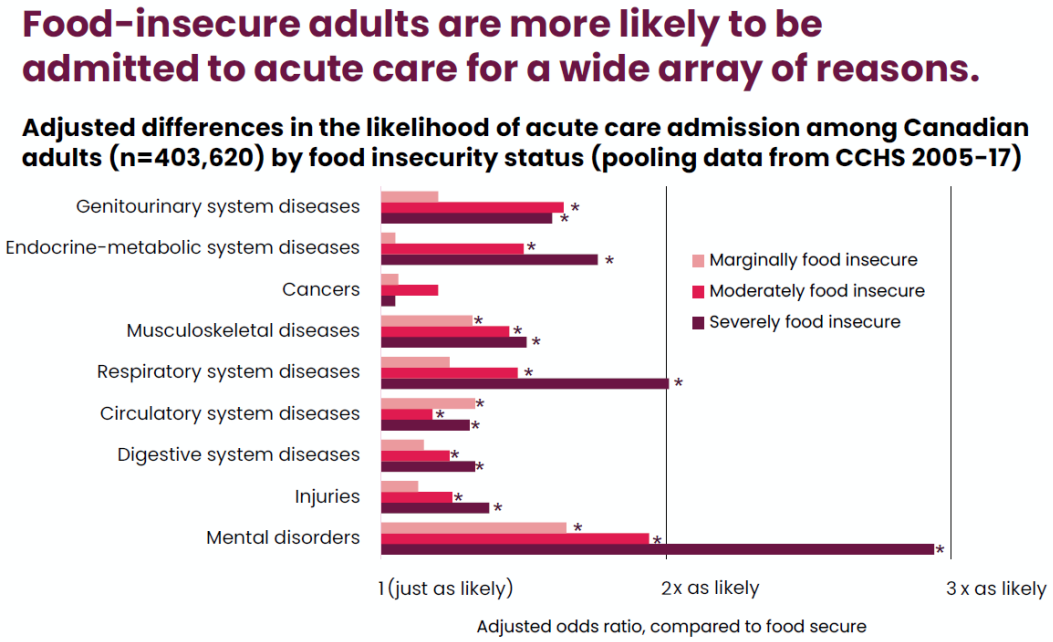
University of Massachusetts, Amherst

Joint Statistical Meetings, August 2025

Food insecurity is associated with multiple adverse health outcomes



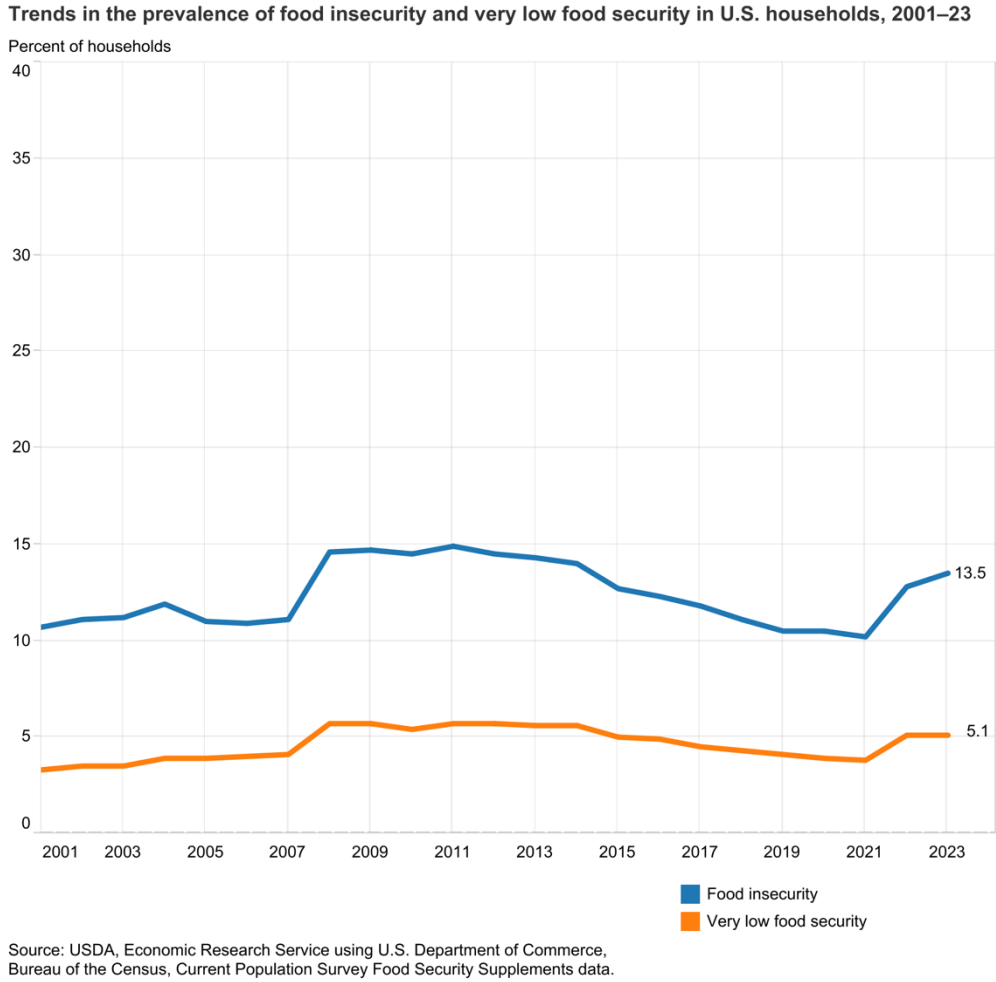
Food Insecurity and Hunger in the United States: An Assessment of the Measure. National Research Council (2006)



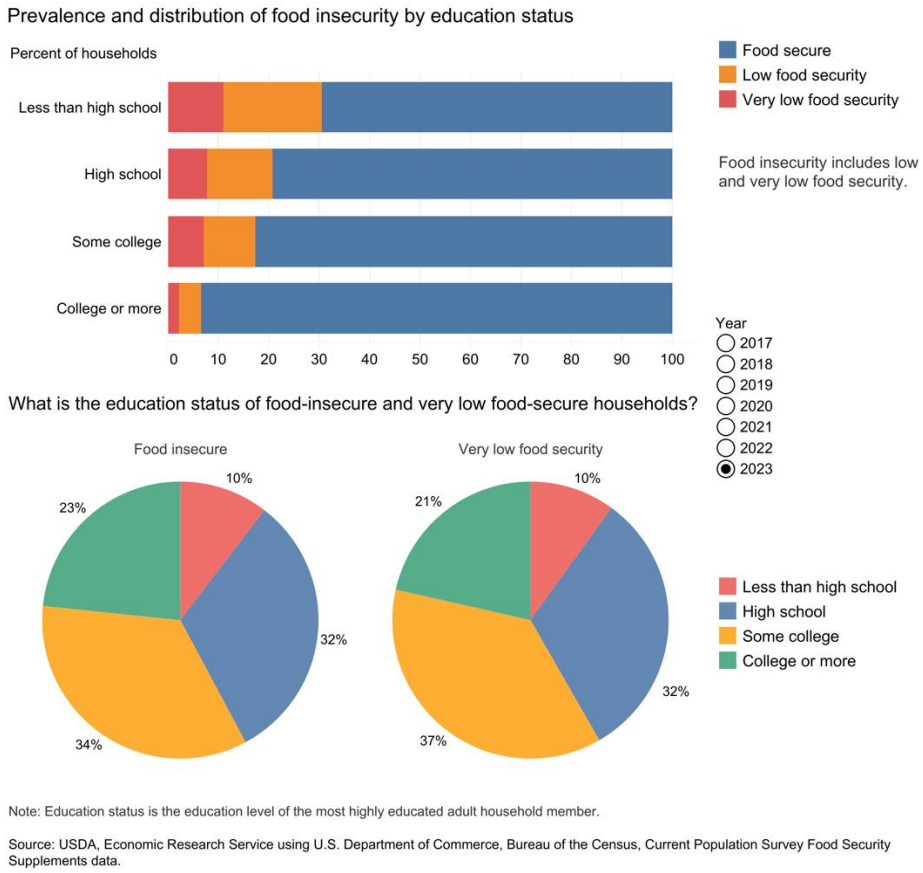
Website from Food Insecurity Policy Research, Canadian Institute of Health Research

18 million US households experienced food insecurity in some time during 2023

- Food insecure households were uncertain they would be able to acquire sufficient food to meet the needs of all household members.
- Explore the USDA food insecurity interactive visualization (2 min), what do you observe?
  - What did you learn from the visualization that you didn't know before?
  - What factors are associated with higher risk of food insecurity?
  - Which state has higher food insecurity?



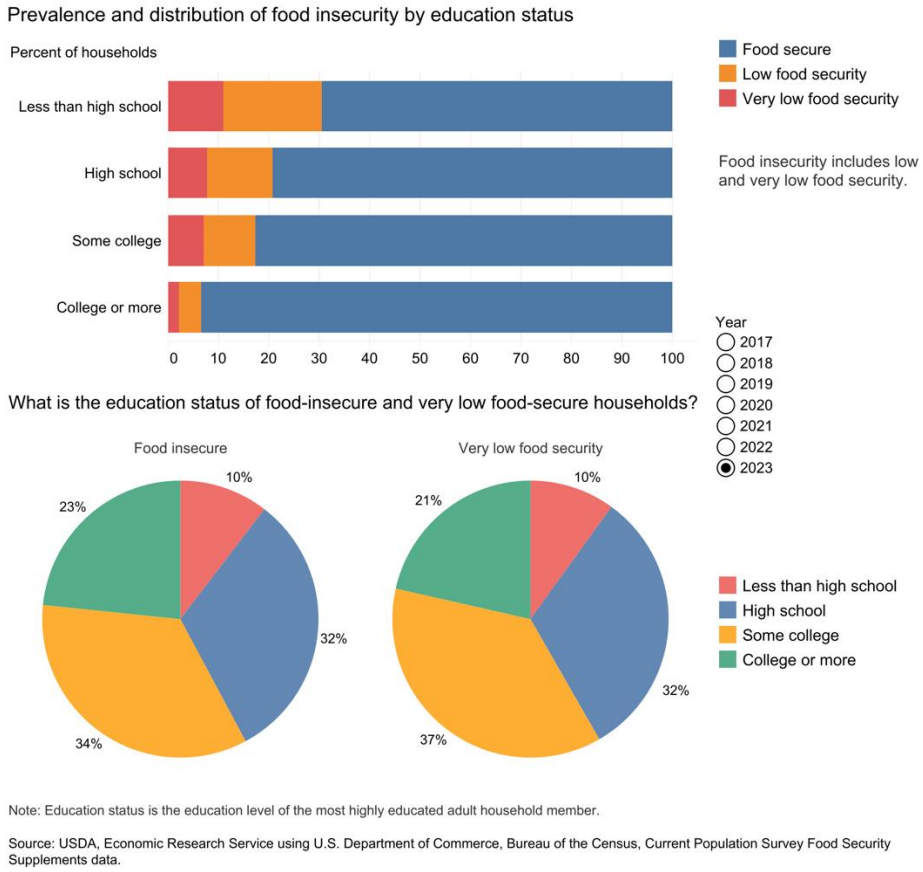
Individual-level factors influence food insecurity

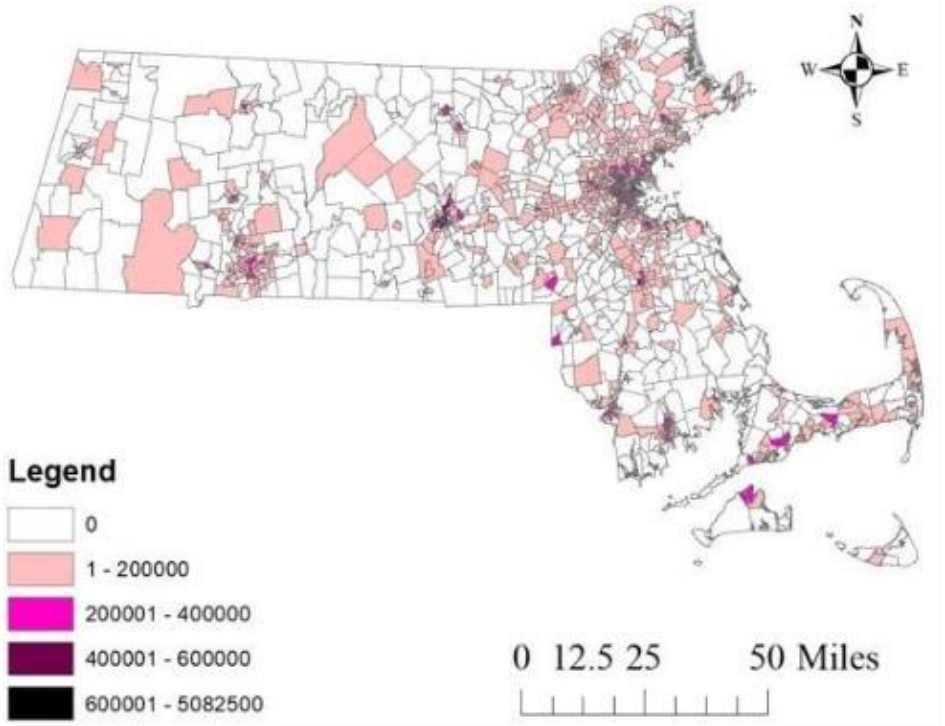


Note: Education status is the education level of the most highly educated adult household member.

Source: USDA, Economic Research Service using U.S. Department of Commerce, Bureau of the Census, Current Population Survey Food Security Supplements data.

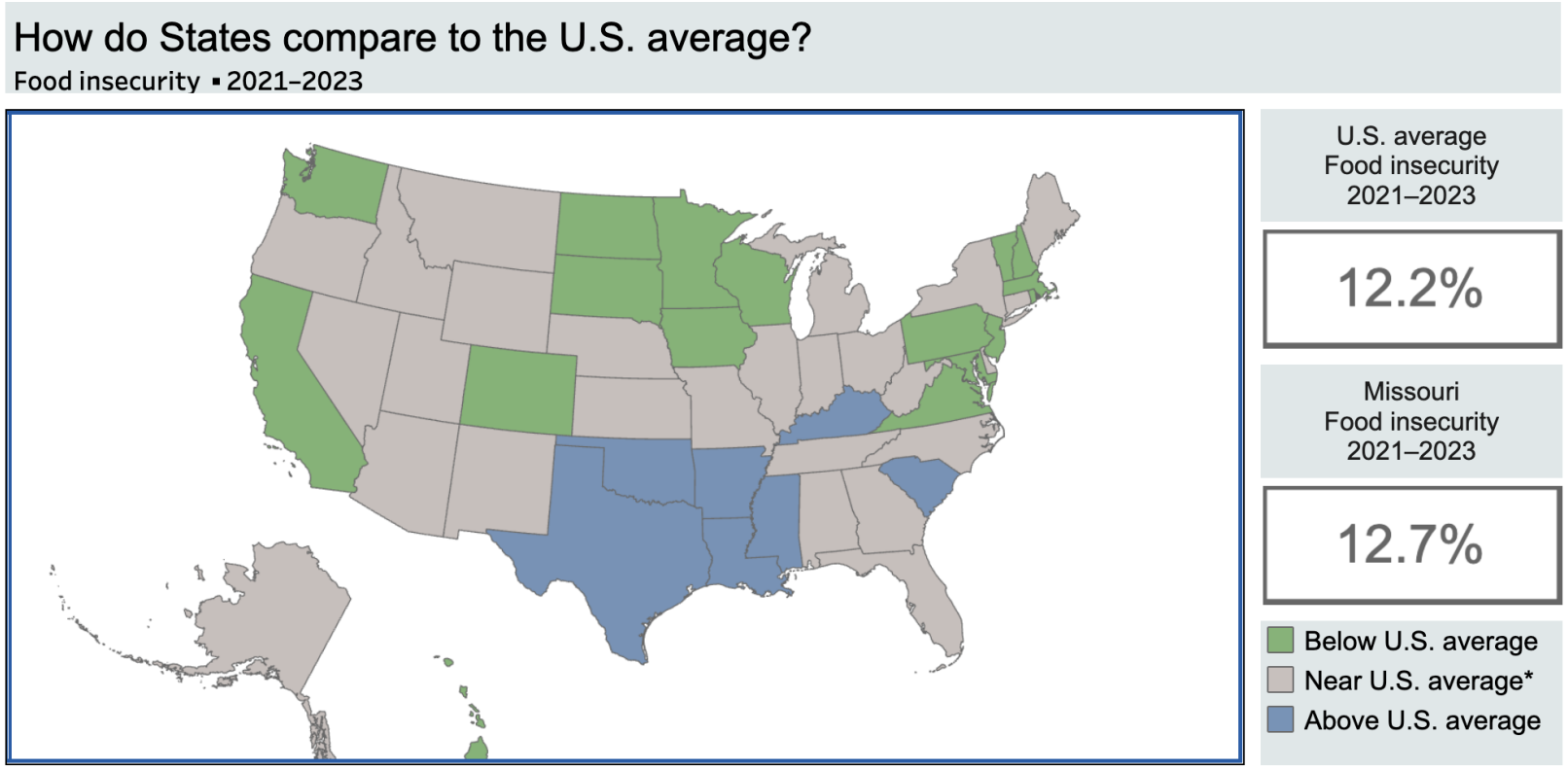
Individual-level factors influence food insecurity, and so do **community-level factors**





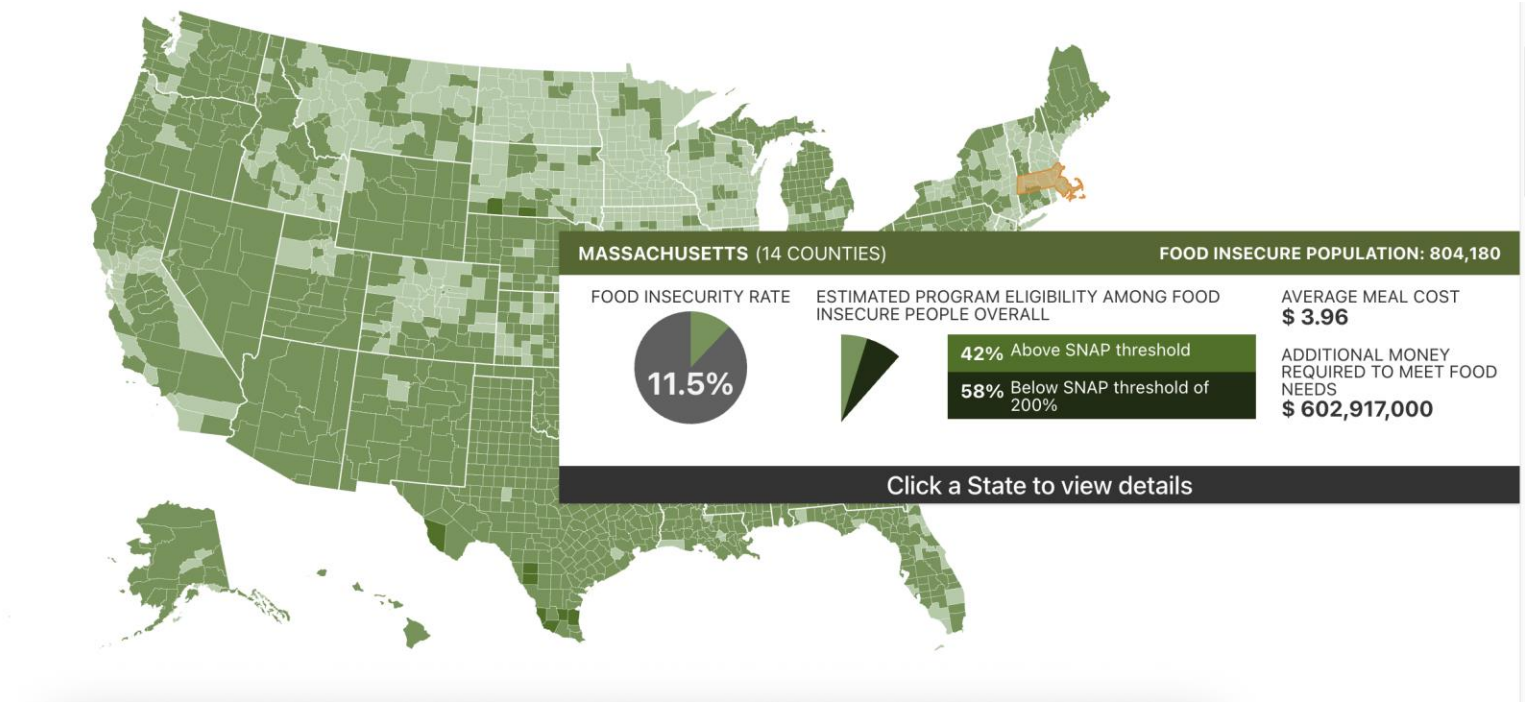
Total square footage of supermarkets within 30 min public transit
Christofa et al. Measuring Food Access to Improve Public Health (No. Report No. 23-042). Massachusetts Dept. of Transportation (2023)

State-level data is reported yearly, what about counties?



Source: USDA, Economics Research Service using U.S. Department of Commerce, Bureau of Census, Current Population Survey Food Security Supplements Data

County-level food insecurity is estimated with linear models



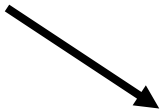
Mind the Meal Gap 2025 Food Insecurity Report and Technical Appendix (using 2023 data)

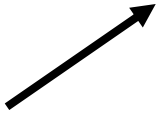
$$FI_{st} = \alpha + \beta_{UN}UN_{st} + \beta_{POV}POV_{st} + \beta_{MI}MI_{st} + \beta_{HISP}HISP_{st} + \beta_{BLACK}BLACK_{st} + \beta_{OWN}OWN_{st} + \beta_{DSBL}DSBL_{st} + \mu_t + \upsilon_s + \varepsilon_{st} \quad (1)$$

$$FI^*_c = \hat{\alpha} + \widehat{\beta_{UN}} UN_c + \widehat{\beta_{POV}} POV_c + \widehat{\beta_{MI}} MI_c + \widehat{\beta_{HISP}} HISP_c + \widehat{\beta_{BLACK}} BLACK_c + \widehat{\beta_{OWN}} OWN_C + \widehat{\beta_{DSBL}} DSBL_C + \widehat{\mu_{2022}} + \hat{v_s} \quad (2)$$

This work combines **individual-level model** and **synthetic population** to estimate county-level food insecurity

1		P(individual living in a food insecure household given income, difficulty, age)
2		P_{MA}(income, difficulty, age)
3		P_{county}(income) P_{county}(difficulty) P_{county}(age)

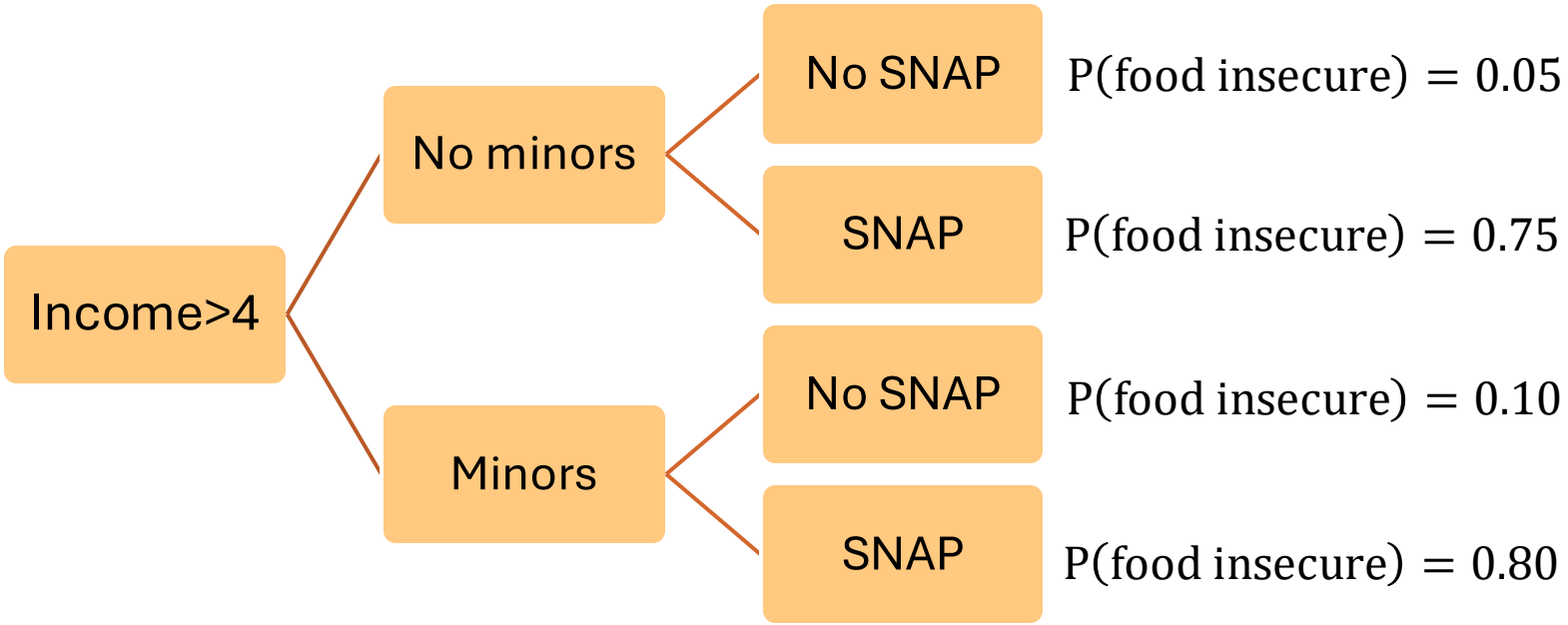




P_{county}(individual living in a food insecure household)

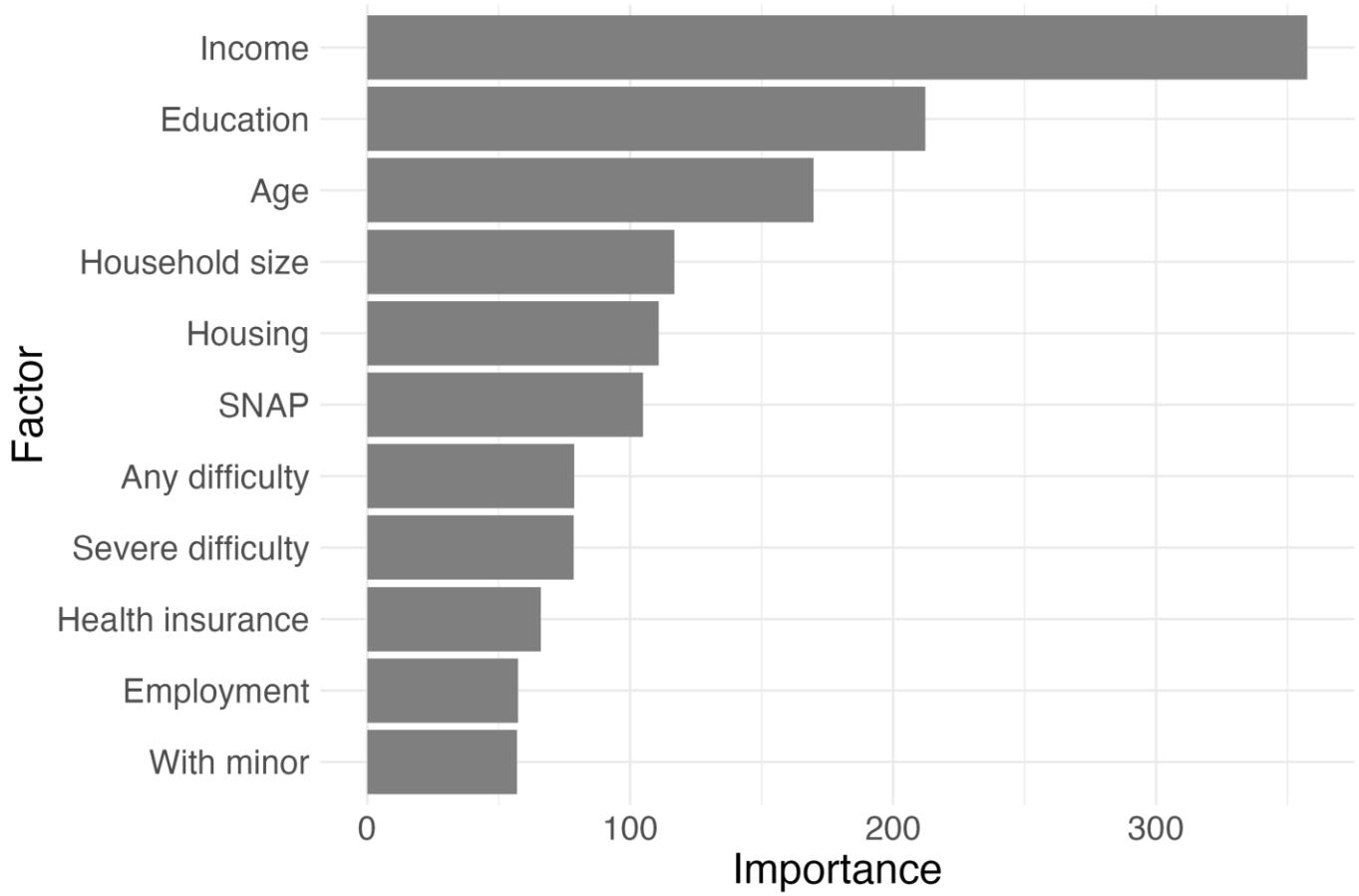
P_{county}(income, difficulty, age)

1. **Probability forest** estimates an individual's likelihood to be food insecure



Malley J.D. et al. Probability machines: consistent probability estimation using nonparametric learning machines (2012)

Income, education and age are the most important variables in the fitted model



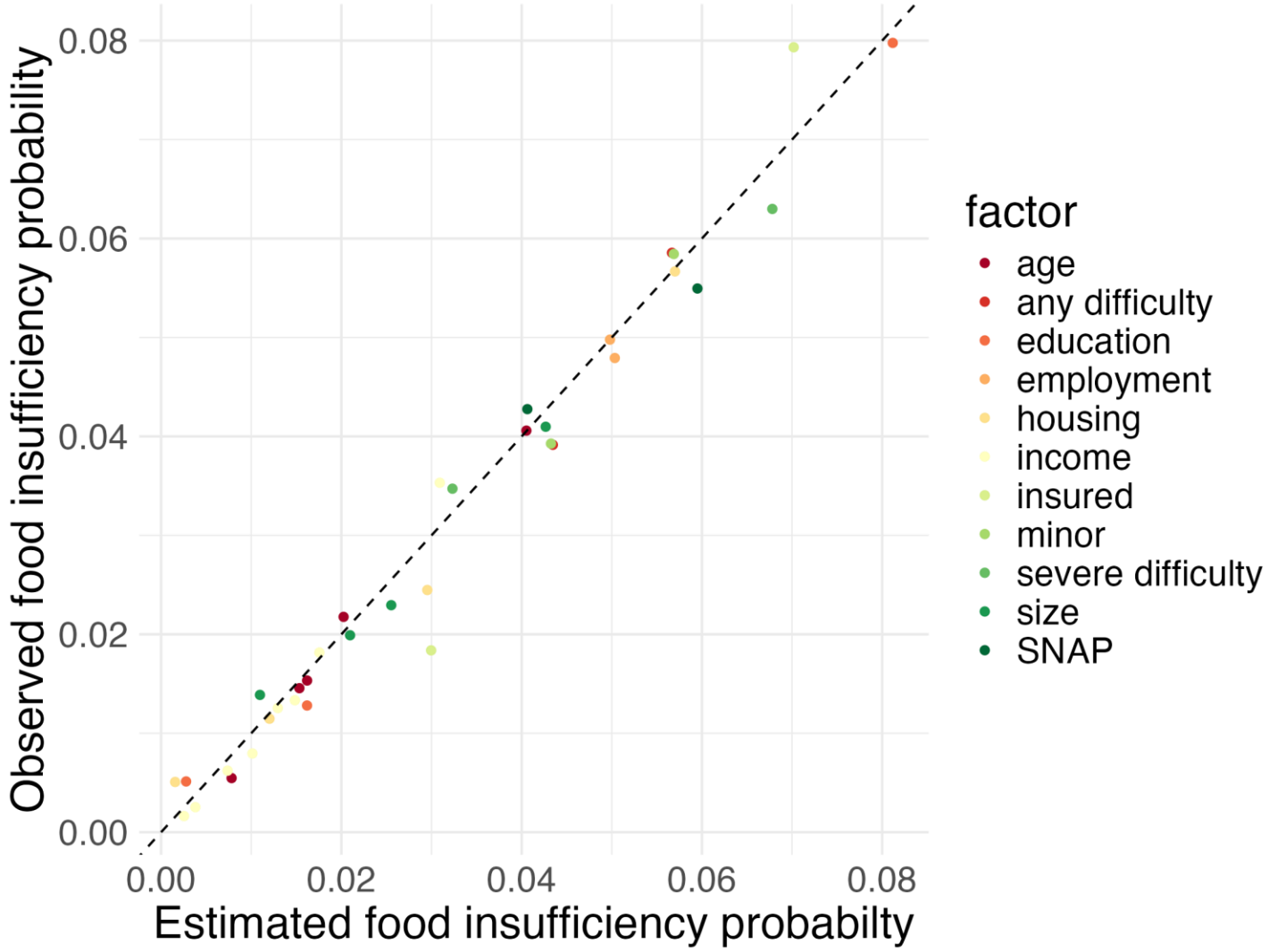
We use PULSE household survey public data, with ~18K observations from Massachusetts in 2023 (link to website)

2. Estimate **state-level joint distribution** of individual factors using a probabilistic graphical model

- Step 1: Apply **one-shot encoding** to each categorical variable
- Step 2: Estimate **precision matrix** using graphical LASSO with cross validated regularization parameter
- Step 3: construct **cliques** based on estimated edges
- Step 4: Compute maximum likelihood estimates assuming an **exponential family** model

$$p(x) = \exp\left(\sum_{c \in \text{Cliques}} \theta_c \; I_c(x_c) \; - \; \phi(\theta) \right)$$

Estimated state level food insecurity aligns with observed frequency for each factor

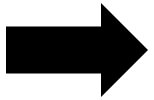


3. Estimate **county-level joint distribution** of individual factors using iterative proportional fitting

Household income	Age		State total
	18 - 24	25 – 34	
<$24,999	a	b	$a+b$
$25,000 - $34,999	c	d	$c+d$
State total	$a+c$	$b+d$	

Household income	County total
<$24,999	m
$25,000 - $34,999	n

IPF

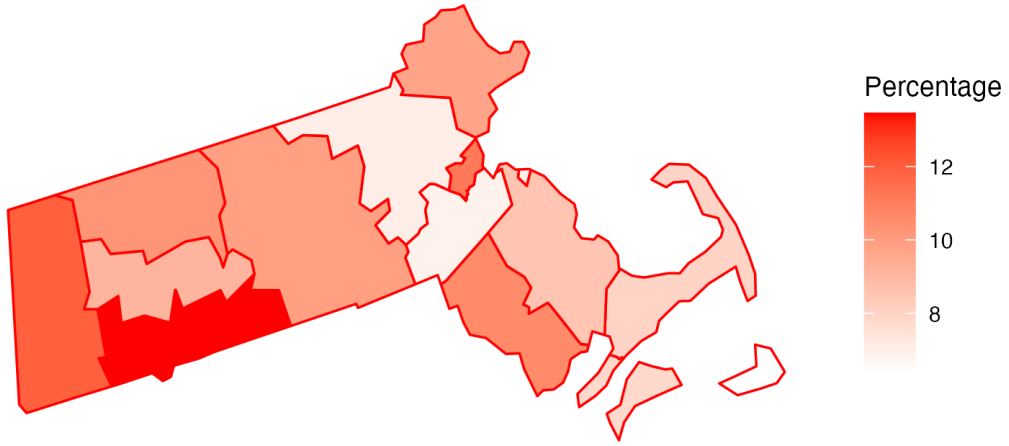


Household income	Age		County total
	18 - 24	25 – 34	
<$24,999	$\frac{a}{a+b}m$	$\frac{b}{a+b}m$	m
$25,000 - $34,999	$\frac{c}{c+d}n$	$\frac{d}{c+d}n$	n
County total	$\frac{a}{a+b}m + \frac{c}{c+d}n$	$\frac{b}{a+b}m + \frac{d}{c+d}n$	

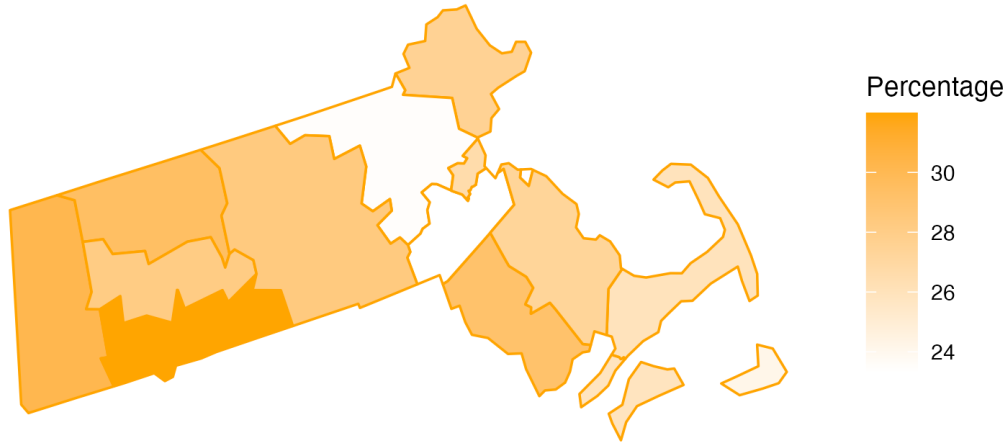
Lancaster et al. The importance of a household living budget in the context of measuring economic vulnerability: a census curated data enterprise use case demonstration (2023)

4. Combining individual probability and synthetic county population to estimate county level food insecurity

Estimated food insecurity 2023 (severe)

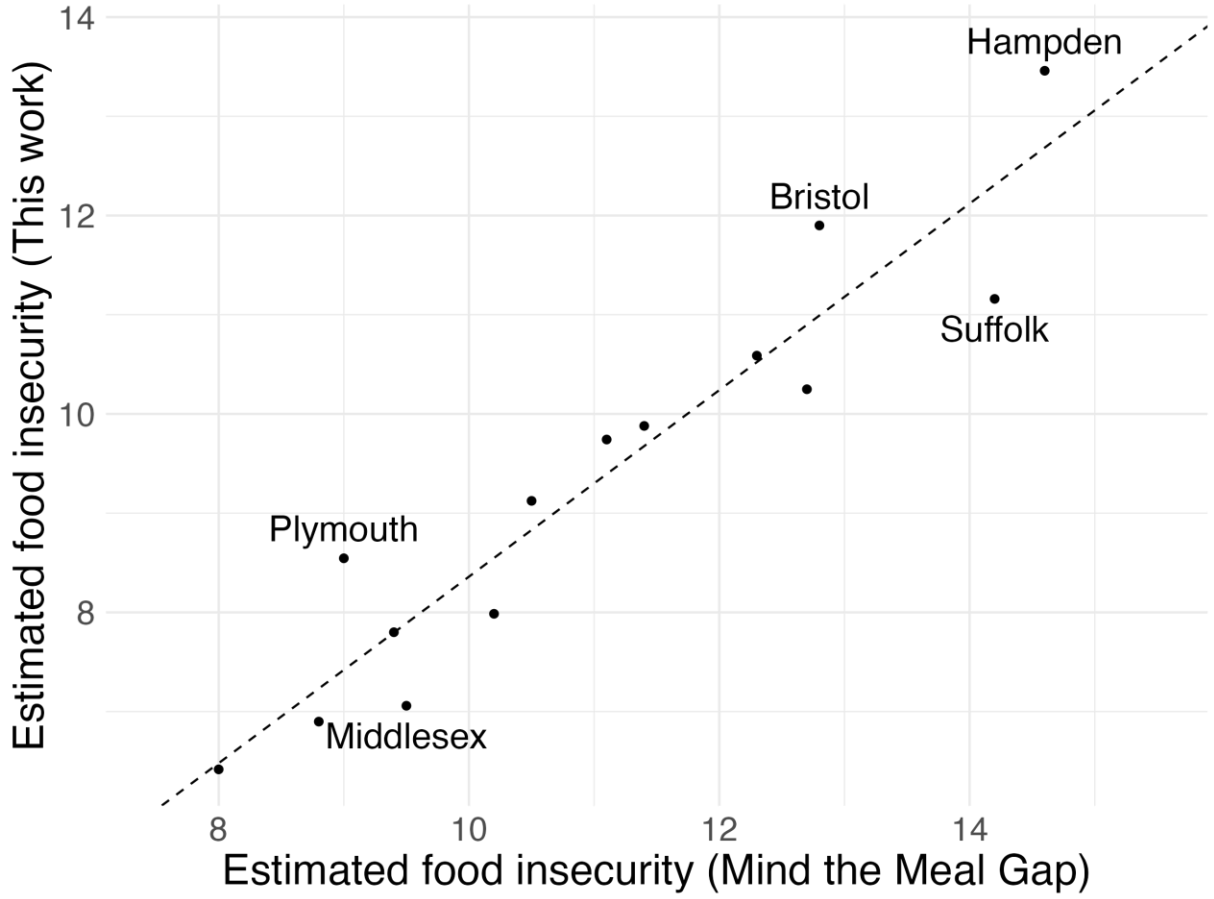


Estimated food insecurity 2023 (low)



Severe food insecurity: sometime / often not enough food to eat
Low food insecurity: Enough, but not always the kind of food we want to eat

Our estimates mostly align with Mind the Meal Gap estimates



Dashed line has slope 0.94 and intercept -1.04.

Summary

- This work combines **individual-level model** and **synthetic population** to estimate county-level food insecurity
- Future work includes incorporating **community characteristics** in the model and considering longitudinal model of food insecurity

Thank you!

qianzhao@umass.edu