Investigating U.S. Food Insecurity Through Data

Conrad Linus Muhirwe Sharon Wanyana Ryann Tompkins
Alex Arevalo

2025-10-21

Table of contents

1. Administration

App Title: Investigating U.S. Food Insecurity Through Data

Team Members: - Conrad Linus Muhirwe - Sharon Wanyana - Ryann Tompkins - Alex Arevalo

Course: DATA-613: Data Science Practicum

2. Project Overview (Topic and Data)

This project examines the prevalence and distribution of household food insecurity across U.S. states (and, where feasible, counties) and key demographic groups. We link geographic patterns to socioeconomic conditions (income, poverty, employment, education) and local food environment characteristics, and surface insights through an interactive Shiny application designed for policy and program stakeholders.

Objectives

1. Describe levels and trends in food insecurity across places and populations (2023–2025).
2. Explain variation using socioeconomic drivers and food-environment context.
3. Contextualize results with safety-net program reach (SNAP/WIC participation).
4. Communicate findings via an accessible, exportable, and update-ready Shiny tool.

2.1 Problem Context

Food insecurity is a persistent, multidimensional challenge in the United States with direct implications for public health, educational attainment, labor productivity, and long-run economic mobility. Beyond its immediate effects on diet quality and well-being, food insecurity concentrates among low-income households, certain racial/ethnic groups, families with children, and communities facing limited access to stable employment or affordable food outlets. These disparities mirror broader structural inequities income volatility, housing cost burdens, gaps in educational opportunity, and uneven access to social safety-net programs—making food insecurity both an outcome and a signal of wider socioeconomic distress.

This project situates food insecurity within that broader context and pursues three goals.

- Quantify the prevalence and geographic distribution of food insecurity across U.S. states and key demographic groups to reveal patterns that may be obscured in national averages.

- Examine how food insecurity co-varies with socioeconomic drivers household income and poverty, employment conditions, and educational attainment while acknowledging potential confounders such as regional price levels and urban rural differences.
- Translate findings into an accessible, decision support tool: an interactive Shiny application that maps prevalence, visualizes trends over time, and enables users to explore relationships between food insecurity and its determinants.

The intended users include policy makers, program administrators, advocacy organizations, and researchers who require timely, interpretable evidence to target interventions (e.g., SNAP outreach, school meal programs, minimum-wage or workforce initiatives) and to monitor equity impacts. By combining transparent methods with interactive visualization, the project aims to move beyond static reporting supporting data-informed prioritization, cross state learning, and continuous evaluation of strategies to reduce food insecurity and its downstream social and economic costs.

2.2 Literature Review

National monitoring by the USDA Economic Research Service (ERS) shows that food insecurity rose in 2023 to 13.5% of U.S. households, reversing pre-pandemic declines and underscoring ongoing economic stressors (U.S. Department of Agriculture, Economic Research Service [USDA ERS], 2024). These official estimates establish the baseline for state comparisons and subgroup analyses in this project. (USDA ERS, 2024.)

Local estimates repeatedly document spatial concentration of food insecurity in rural areas and the South. Feeding America's Map the Meal Gap reports that a disproportionately large share of counties with the highest food insecurity rates are rural and located in the South, highlighting persistent regional differences that area level covariates (e.g., rurality, retail food access) can help explain (Feeding America, 2024).

Extensive syntheses show systematic disparities by income, race/ethnicity, and family structure, and link food insecurity to adverse health outcomes across the life course (Gundersen & Ziliak, 2015). This literature supports our use of ACS covariates (poverty, income, employment, educational attainment) to model exposure risk and interpret subgroup trends in the app (e.g., households with children, single-parent households, and racially/ethnically minoritized populations). (Gundersen & Ziliak, 2015.)

Credible quasi experimental evidence indicates that SNAP participation causally reduces food insecurity. Using partial identification methods to address selection and misclassification, Gundersen, Kreider, and Pepper (2017) estimate at least a six-percentage-point reduction in food insecurity among households with children attributable to SNAP (Gundersen et al., 2017). This finding motivates our inclusion of SNAP participation rates as a key covariate in state-level models, recognizing that program access and uptake are critical mediators of food insecurity risk.

2.3 Proposed Data Sources

The project will integrate multiple authoritative datasets to analyze food insecurity across U.S. states and demographic groups.
All sources will be harmonized by **state or county FIPS codes** and aligned for **2023–2025**.

Source	Provider	Purpose	Key Variables	Geographic Level / Years
Household Food Security Data	USDA Economic Research Service (ERS)	Primary outcome measure — annual prevalence of food insecurity among U.S. households	% food insecure households, % very low food security	State-level, 2023–2025
American Community Survey (ACS)	U.S. Census Bureau	Socioeconomic and demographic covariates	Poverty rate, median household income, unemployment, education, race/ethnicity, household type	State & County, 2023–2025 (1-year/5-year estimates)
Food Environment Atlas (FEA)	USDA	Contextual indicators of local food access and environment	Food retail density, low-access areas, rural–urban classification, transportation access	County-level, most recent update (aligned to ACS year)
SNAP & WIC Participation Data	USDA Food and Nutrition Service (FNS)	Program coverage indicators to assess reach relative to need	Participation rate, number of beneficiaries	State-level, 2023–2025

3 Use Case and Actor Description

The primary use case for this Shiny application is to support public policy analysis and research on food insecurity trends across the United States. The application makes complex socioeconomic and food insecurity data accessible, interactive, and interpretable for decision makers, researchers, and practitioners.

The intended actor is a Public Policy Analyst or Nonprofit Researcher working in food policy, community development, or social welfare. These professionals often need to identify geographic disparities, assess high risk populations, and communicate evidence based insights to stakeholders.

Through the app, users will be able to:

- Dynamically explore national, state, and county-level food insecurity data.
- Examine relationships between food insecurity and socioeconomic indicators such as income, employment, and education.
- Generate customizable visuals and descriptive analyses to inform reports, presentations, and policy briefs.

This use case ensures that the application provides both exploratory and analytical functionality—supporting evidence-based decision-making at federal, state, and local levels.

3.1 Actor Questions of Interest

The Shiny app is designed to help the actor answer key research and policy questions such as:

- How does food insecurity vary across U.S. states and regions over time?
- Are there seasonal or long-term trends that can inform future policy focus?
- What is the relationship between food insecurity and socioeconomic variables (e.g., income, unemployment, education)?
- Which factors appear to be the strongest predictors of food insecurity?
- How do rural vs. urban areas differ in food insecurity rates?
- Which demographic groups such as households with children or racial/ethnic minorities are most affected?
- Can we observe measurable shifts in food insecurity following policy interventions (e.g., SNAP expansions or COVID-era relief programs)?

3.2 Anticipated Workflow for the Actor

The app's design follows the logical steps of a policy analyst's exploratory workflow:

1. Select Scope – Choose a geographic level (national, state, or county) and a time frame (e.g., 2023–2025).
2. Filter Variables – Select socioeconomic indicators (e.g., poverty rate, income, unemployment, education).
3. Visualize Data – Generate interactive charts, heatmaps, and choropleth maps showing temporal and geographic trends.
4. Explore Relationships – View summary statistics, correlations, and regression results dynamically.
5. Generate Insights & Export – Download customized tables, figures, or reports for communication and decision-support.
6. Iterate – Adjust filters or parameters to test new hypotheses or compare across states and years.

4. Ethical Review

Ethical conduct is fundamental to the integrity, credibility, and societal value of this project. The analysis of food insecurity an issue with profound social implications demands careful adherence to ethical principles at every stage of the data life cycle. This section reviews the ethical considerations associated with data acquisition, storage, processing, analysis, visualization, and dissemination, and aligns them with the American Statistical Association (ASA) Ethical Guidelines for Statistical Practice (2023).

Ethical Considerations Across the Data Life Cycle

1. **Data Acquisition.** The project exclusively utilizes publicly available, de-identified, and non-sensitive datasets from authoritative sources, including the USDA Economic Research Service (ERS), the U.S. Census Bureau's American Community Survey (ACS), and Feeding America's Map the Meal Gap. Each source operates under open-data or Creative Commons licensing, ensuring compliance with data-sharing regulations. To maintain transparency, all data sources will be clearly cited and accompanied by metadata documentation. The team will avoid using proprietary or confidential datasets, and no personally identifiable information (PII) will be collected or stored.

2. **Data Storage and Management.** Data integrity and security will be preserved through organized storage within the project's GitHub repository. Proper directory structure and version control will support traceability and reproducibility. Only processed and approved datasets will be uploaded; raw data will remain unaltered in secured folders. Team members will follow the principle of minimum necessary access, limiting modifications to relevant files and documenting all changes through meaningful commit messages.

3. **Data Cleaning and Analysis.** All data transformations and analytical decisions will be documented to ensure full transparency and reproducibility. Analytical integrity will be maintained by avoiding any manipulation of variables or statistical outputs that could bias findings. The team will adhere to the ASA's principles of objectivity and methodological rigor, ensuring that analytical choices are justified by theory and sound statistical reasoning. Where models are used, assumptions and limitations will be clearly articulated. Correlations and regressions will be interpreted cautiously to avoid implying causality from observational data.

4. **Data Visualization and Interpretation.** Visual ethics are critical in communicating findings responsibly. All graphical outputs in the Shiny application will adhere to data visualization best practices maintaining correct scaling, labeled axes, colorblind-accessible palettes, and proportionate representations of quantities. The team will avoid exaggerating differences or omitting uncertainty intervals that may mislead interpretation. Language used in labels, captions, and explanatory text will remain neutral, respectful, and non-stigmatizing, particularly when describing vulnerable populations or geographic areas with higher food insecurity.

5. **Dissemination and Communication.** The final Shiny application is designed for educational and analytical use, supporting evidence-based policymaking without advancing a particular political agenda. Results will be accompanied by explanatory notes and disclaimers emphasizing limitations and potential sources of error. The team will encourage users to interpret findings contextually, recognizing the socioeconomic complexities underlying food insecurity.

5. Application Concept

5.1 General Layout of the App to Support the Actor Workflow

The Shiny application will serve as an interactive dashboard that enables policy analysts, researchers, and nonprofit professionals to explore and interpret patterns of food insecurity across the United States. The interface will be clean, accessible, and structured around the actor's decision-making workflow—progressing logically from overview → exploration → analysis.

The app will contain three primary tabs, each representing a stage in the analytic process:

1. **Overview Tab** Presents a national snapshot of food insecurity with key performance indicators (KPIs) such as national averages, regional differences, and year-over-year changes. Includes definitions, metadata, and data-source links.

2. **Exploration Tab** Provides interactive visualization tools (maps, line charts, bar plots) for examining trends at the state and county levels. Filters and controls allow comparison by time, region, or demographic subgroup.

3. **Analysis Tab** Offers statistical and modeling features, including correlation matrices, regression summaries, and group comparisons, to examine relationships between food insecurity and socioeconomic variables.

All tabs will maintain a consistent layout with sidebar input controls and a main display panel for dynamic outputs. The design emphasizes logical navigation and usability, guiding the actor through an intuitive data-driven workflow.

5.2 Planned Options for User Data Selection and Manipulation

The app will allow users to tailor analyses through interactive inputs that dynamically update visualizations and tables. Key controls include:

- Geographic Filters - Dropdown menus for selecting specific states, regions, or counties.
- Temporal Filters - Sliders or dropdowns to set year ranges (e.g., 2010 – 2024).
- Variable Selection - Checkboxes or multi-select inputs to include/exclude socioeconomic indicators such as median income, unemployment, education, and poverty rate.
- Transformations - Options to apply log transformations or normalization to improve comparability across scales.
- Subgroup Focus - Filters to isolate demographic or regional groups (e.g., rural vs. urban, households with children).

These tools will update outputs in real time, allowing flexible exploration from multiple analytical perspectives without requiring users to write code.

5.3 Planned Options for User Exploratory Numerical / Graphical Analysis

The Exploration Tab will include a range of descriptive and graphical tools to facilitate insight generation:

- Interactive Maps (Choropleths) - Show food-insecurity rates by state or county with color gradients for geographic comparison.

- Trend Analysis - Line and bar plots illustrating temporal changes in food insecurity and related socioeconomic factors.
- Comparative Visuals - Multi-panel or faceted charts to compare multiple states or demographic groups simultaneously.
- Summary Tables - Interactive data tables built with {DT} for sorting, searching, and exporting selected metrics.
- Dynamic Tooltips - Hover-over elements displaying contextual statistics such as exact rates, rankings, or summary values.

Visualizations will be developed using {ggplot2}, {plotly}, and {leaflet}, ensuring interactivity, consistency in color schemes, and accessibility for all users.

5.4 Planned Options for User Statistical Models and Tests

The Analysis Tab will provide lightweight, transparent tools for exploratory statistical modeling and inference:

1. Correlation Analysis
   - Compute Pearson and Spearman correlation coefficients between food insecurity and socioeconomic indicators.
   - Display results as interactive correlation matrices or heatmaps.
2. Regression Modeling
   - Allow users to specify dependent and independent variables for simple or multiple linear regression.
   - Present model summaries (coefficients, R^2, p-values) with diagnostic plots and interpretation notes.
3. Group Comparisons
   - Conduct two-sample t-tests comparing mean food-insecurity rates across categories (e.g., rural vs. urban, high- vs. low-income states).
   - Report p-values, confidence intervals, and short text interpretations.
4. Predictive Exploration
   - Offer reactive model fitting to generate predicted food-insecurity rates given socioeconomic inputs.

All analyses will be fully transparent, with clear documentation of assumptions, limitations, and non-causal interpretation. Users will have the option to download model results, tables, and figures for inclusion in reports or policy briefs.

6. Collaboration and Branching Plan

Our team has selected the Shared Repository Workflow Model to manage collaboration and version control throughout the development of our Shiny application. This model provides an efficient structure for coordinated teamwork, allowing all members to contribute directly to a single shared GitHub repository while maintaining accountability and version stability.

The main branch will serve as the production-ready version of the project. It will remain protected and contain only tested and verified code. Each team member will create feature branches for specific tasks, such as data preparation, UI design, or statistical modeling. Branches will follow clear naming conventions (e.g., feature-data-cleaning, ui-development, analysis-module) to ensure clarity and traceability.

Team members will commit changes frequently with concise and descriptive messages to document progress. Once a feature or task is complete, members will push their branch to GitHub and open a Pull Request (PR). Another team member will review the PR to ensure consistency with project standards before merging into the main branch. This peer-review process helps prevent errors, maintain code quality, and promote shared understanding of all components in the project.

We will use GitHub Issues to assign and track tasks, Discussions or project boards to coordinate activities, and weekly meetings to synchronize milestones and deliverables. In cases of merge conflicts, we will use GitHub's built-in conflict resolution tools and coordinate through team communication channels to ensure smooth integration.

This workflow model supports continuous collaboration, transparent progress tracking, and reliable version control while enabling all contributors to work efficiently and independently. It balances flexibility with control, ensuring that our final app is developed cohesively and collaboratively within an organized structure.

6.1 Team Roles and Responsibilities

Team Member	Course Level	Primary Responsibilities
Conrad Linus Muhirwe	DATA-613	Lead on statistical modeling and analysis (regression, correlation), data integration, reproducibility documentation, and GitHub coordination.
Sharon Wanyana	DATA-613	Lead on modeling, literature and ethical review, validation of assumptions, interpretation of model outputs, and report writing.

Team Member	Course Level	Primary Responsibilities
Ryann Tompkins	DATA-413	UI/UX development, app layout and styling, and exploratory data visualization using `{ggplot2}` and `{plotly}`.
Alex Arevalo	DATA-413	Data acquisition, cleaning, and transformation, creation of interactive tables and choropleth maps, and repository organization.

6.2 Schedule and Milestones

Milestone	Target Date	Deliverable / Outcome
Data acquisition and cleaning	Oct 25 2025	Verified `.rds` datasets stored in `/data` folder; cleaning scripts documented.
App skeleton (UI structure & tabs)	Oct 30 2025	Functional navigation across Overview, Exploration, and Analysis tabs.
Progress Report submission	Nov 11 2025	Updated `.qmd` and HTML with preliminary visuals and workflow.
Statistical modeling integration	Nov 20 2025	Regression and correlation modules functional with outputs.
App refinement & user testing	Dec 3 2025	Fully interactive dashboard tested across user scenarios.
Vignette drafting & demo preparation	Dec 8 2025	Completed vignette, rehearsal of oral demonstration.
Final submission	Dec 9 2025	Final GitHub repo, HTML deliverables, and class demonstration.

REFERENCES

- Feeding America. (2024). Map the Meal Gap 2024: Executive summary. https://www.feedingamerica.org/research/map-the-meal-gap/overall-executive-summary
- Gundersen, C., & Ziliak, J. P. (2015). Food insecurity and health outcomes. Health Affairs, 34(11), 1830–1839. https://www.healthaffairs.org/doi/10.1377/hlthaff.2015.0645
- Gundersen, C., Kreider, B., & Pepper, J. (2017). Partial identification methods for evaluating food assistance programs: A case study of the causal impact of SNAP on food insecurity. American Journal of Agricultural Economics, 99(4). https://onlinelibrary.wiley.com/doi/10.1093/ajae/aax026
- U.S. Department of Agriculture, Economic Research Service. (2024). Household food security in the United States in 2023 (ERR-337). https://www.ers.usda.gov/publications/pub-details?pubid=109895