

Database Structure & Variable Types

Haotian Shi*

January 16, 2026

This project processes four primary datasets related to federal spending, demographics, municipal finance, and household financial health. The data is aggregated at three geographic levels: **State**, **County** and **Congressional District (118th)**.

1 Geographic Variables

1.1 County Level Files (*_county.xlsx)

- `fips` (Integer): Standard unique identifier for US counties (Federal Information Processing Standards).
- `county` (String): County name (clean, lowercase, without “County”/“Parish” suffix).
- `state` (String): Full state name (lowercase).

1.2 State Level Files (*_state.xlsx)

- `state` (String): Full state name (lowercase).

Note: Data is aggregated by summing county-level values for each state.

*University of Maryland Robert H. Smith School of Business.

1.3 Congressional District Level Files (*.congress.xlsx)

- cd_118 (String): Unique identifier for the 118th Congressional District.
 - Format: SS-DD (State Abbreviation - District Number).
 - Example: AK-00 (Alaska At-large), CA-12 (California 12th).
- state_str (String): State Abbreviation (e.g., AL, NY).

2 Dataset Descriptions

2.1 Census Data (ACS)

Source: American Community Survey (ACS) 5-Year Estimates, U.S. Census Bureau

Description: The American Community Survey is an ongoing annual survey conducted by the U.S. Census Bureau that provides vital information on a yearly basis about the nation and its people. The 5-year estimates combine data collected over a 60-month period to produce reliable estimates for small geographic areas, including counties. This dataset covers demographic characteristics, educational attainment, household income distribution, and poverty status across all U.S. counties.

Data URL: <https://data.census.gov/>

File Output: acs_county.xlsx, acs_congress.xlsx, acs_state.xlsx

Variables:

- **Demographics:**

- Total_population, Male, Female
- **Age Groups:** Under_5_years, ..., 85_years_and_over
- **Race:** Race_Total, White, Black, Asian, Hispanic, etc.

- **Education:**

- Less_than_9th_grade, High_school_graduate, Bachelor_degree, Graduate_or_professional_etc.

- Summary: High_school_graduate_or_higher, Bachelor_degree_or_higher
- **Income:**
 - **Households by Income Bracket:** Less_than_10000, ..., 200000_or_more
 - **Summary Stats:** Median_household_income, Mean_household_income
- **Poverty:**
 - Below_poverty_level, At_or_above_poverty_level

2.2 Government Spending (Municipal Finance)

Source: Annual Survey of State and Local Government Finances, U.S. Census Bureau; processed via Reason Foundation's Government Finance Database

Description: This dataset provides comprehensive financial statistics for state and local governments, including data on revenues, expenditures, debt, and assets. The Annual Survey of State and Local Government Finances is conducted by the U.S. Census Bureau and covers the financial activities of all state and local governments in the United States. The data enables analysis of government fiscal health, debt burdens, pension liabilities, and overall financial position at the county level.

Coverage: County-level, fiscal year 2022

Data URL: <https://govfinance.reason.org/>

File Output: gov_county.xlsx, gov_congress.xlsx, gov_state.xlsx

Variables (Aggregated):

- **Assets & Liabilities:** Total_Assets, Total_Liabilities, Current_Assets, Current_Liabilities, Non-Current_Liabilities
- **Debt & Pension:** Bonds,_Loans_&_Notes (Outstanding debt), Net_Pension_Liability, Net_OPEB_Liability
- **Position & Flow:** Net_Position, Revenue, Expenses, Free_Cash_Flow
- **Others:** Compensated_Absences

Note: Variables ending in _per_capita and Ratio variables (Current_Ratio, Debt_Ratio) are **excluded** from aggregated sums (State/Congress level) to avoid statistical errors.

2.3 Contract Data (Federal Awards)

Source: USAspending.gov, U.S. Department of the Treasury

Description: USAspending.gov is the official open data source of federal spending information, mandated by the Federal Funding Accountability and Transparency Act (FFATA). This dataset tracks federal contract awards, including prime contracts and subaward flows between geographic areas. It enables analysis of how federal procurement dollars flow into and out of counties through direct awards and subcontracting relationships. The data captures both the place of performance and the flow of funds through the federal contracting hierarchy.

Coverage: County/Congressional District/State level

Data URL: <https://usaspending.gov/>

File Output: contract_county.xlsx, contract_congress.xlsx, contract_state.xlsx

Variables:

- **Direct Obligations:**

- fed_act_obl: Federal Action Obligation (Direct federal spending in the county)
- fed_act_obl_indirect: Indirect obligation amounts

- **Subawards:**

- subaward_amount_in: Amount received as sub-awardee
- subaward_amount_out: Amount distributed as prime contractor to sub-awardees
- subaward_amount_net_inflow: Net subaward flow (in – out)

- **Economic Context:**

- employment: Employment count
- residents: Resident population count

2.4 FINRA (National Financial Capability Study)

Source: FINRA Investor Education Foundation, National Financial Capability Study (NFCs)

Description: The National Financial Capability Study is a comprehensive survey conducted by the FINRA Investor Education Foundation to assess Americans' financial capability. The study measures financial literacy, financial behaviors, and financial well-being across the United States. It provides insights into how households manage their finances, their access to traditional vs. alternative financial services, their debt burdens, and their self-assessed financial knowledge. The county-level indices are constructed from individual survey responses aggregated to geographic areas.

Coverage: County/Congressional District/State level, 2021

Data URL: <https://www.usfinancialcapability.org/>

File Output: finra_county.xlsx, finra_congress.xlsx, finra_state.xlsx

Variables:

- **financial_constraint:** An index measuring the degree of household financial constraint, constructed from indicators of liquidity stress, difficulty covering expenses, and lack of emergency savings.
- **fc_norm:** A normalized version of the financial constraint index.
- **alternative_financing:** An index capturing households' reliance on nontraditional or high-cost credit sources (e.g., payday loans, pawn shops). Higher values indicate greater use.
- **af_norm:** A normalized version of the alternative financing index.
- **financial_literacy:** An index measuring objective financial knowledge based on correct responses to questions on interest compounding, inflation, bond prices, mortgages, etc.
- **f1_norm:** A normalized version of the financial literacy index.
- **satisfied:** Indicator (0/1). 1 if respondent reports high satisfaction with overall personal financial condition.
- **risk_averse:** Indicator (0/1). 1 if respondent reports low willingness to take financial investment risk.

- `too_much_debt`: Indicator (0/1). 1 if respondent strongly agrees they currently have too much debt.
- `high_fin_knowledge`: Indicator (0/1). 1 if respondent self-assesses their overall financial knowledge as high.