### **Script Directory**

This project analyzes welfare program participation across several states, focusing on the impact of high-profile welfare fraud events. It involves identifying relevant events, cleaning and preparing county-level data, and comparing USDA and state SNAP data for quality control. The analysis includes population weighting, county-level comparisons between treatment and control groups, and event studies to assess the effects of significant events. The project produces detailed outputs, including plots and summary files, to support the analysis.

#### **Theme: Identifying High-Profile Welfare Fraud Events**

This theme links welfare fraud events reported in the media with county-level welfare program data across various states to analyze changes in welfare participation around these events.

**Data Sources:**

1. **Program County-Level Data Sources:**
   * A list of publicly available data sources for SNAP, TANF, WIC, and Medicaid at the county level by state, used to identify data availability for the analysis.
2. **Fraud Story PDFs:**
   * A collection of PDFs documenting significant welfare fraud stories, categorized by the type and number of media sources, to identify events with substantial media coverage.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#### **Theme: Cleaning County-Level Data**

This theme focuses on collecting, formatting, and preparing county-level welfare program data for analysis across multiple states.

**Data Sources:**

1. **Web Scrape:**
   * Python scripts per state that download county-month data from state websites. These scripts are not modified for R Project functionality and do not contribute to any data output.
2. **Raw Data County Compiling:**
   * Cleaned Excel sheets for each state. Most of the cleaning was done manually in Excel, and the data is prepared for uniform formatting.
3. **County Compiling Scripts:**
   * R scripts that finalize the data cleaning process, producing the uniformly formatted and cleaned data ready for analysis.
4. **County Compiling Data Output:**
   * The final cleaned data sources for each state, formatted uniformly and ready for detailed analysis.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#### **Theme: USDA Comparisons**

This theme involves comparing state and county/locality SNAP data between state websites and USDA sources to identify discrepancies and trends.

**Data Sources:**

1. **USDA Locality/County Data:**
   * Tracks SNAP categories by locality for January and July of each year.
2. **USDA State Data:**
   * Tracks SNAP categories by state for every month.

**Scripts:**

1. **State-Level Comparison Script:**
   * Compares state-level aggregates between the state website data and USDA data, identifying discrepancies or trends.
2. **County/Locality-Level Comparison Script:**
   * Compares county/locality-level SNAP data between state sources and USDA data to assess consistency and reliability.

**Plots:**

1. **State-Level Comparison Plots:**
   * Visual representations of the discrepancies or trends found in state-level SNAP data comparisons.
2. **County/Locality-Level Comparison Plots:**
   * Plots visualizing the differences and trends in county/locality-level SNAP data comparisons.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#### **Theme: County Population Weighting**

This theme focuses on adjusting county-level SNAP data by population to ensure accurate comparisons across counties.

**Data Sources:**

1. **Annual Population Estimates:**
   * Annual population estimates for each county provided by the US Census Bureau, stored in the raw data.

**Scripts:**

1. **Population Weighting Script:**
   * This script weights each county’s value for various SNAP categories by its population and saves the updated data over the original .dta file.

**Data Outputs:**

1. **County-Level SNAP Totals:**
   * The final data file containing population-weighted SNAP totals for each county, ensuring that all values are adjusted for population size.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#### **Theme: County Comparison**

This theme focuses on analyzing raw comparison trends between treatment and control groups for county-level SNAP data.

**Data Sources:**

1. **County-Level SNAP Totals:**
   * The population-weighted SNAP totals for each county, serving as the basis for comparison across treatment and control groups.

**Scripts:**

1. **County Comparison Script:**
   * This script creates treatment and control groups, aggregates the data based on these groups, and saves the resulting summary data in both .dta and Excel formats. The summary includes aggregates and denotes treatment and control by locality.

**Outputs:**

1. **Plots:**
   * Visual representations of trends for each category by state, comparing treatment and control groups. These are saved to the **Plots** folder.
2. **Summary Files:**
   * The script produces a summary .dta and Excel file that includes aggregates and denotes treatment and control by locality. These files are saved to the **Data Outputs** folder.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#### **Theme: Event Study**

This theme focuses on conducting event studies to analyze differences between treatment and control groups at a specific event date, using the summary data produced in previous analyses.

**Data Sources:**

1. **Summary Files:**
   * Aggregated and categorized data from the County Comparison Script, which includes treatment and control group designations by locality.

**Scripts:**

1. **Event Study Script:**
   * This script differences out the treatment and control groups at the event date and compares trends over time for each state.

**Outputs:**

1. **Event Study Plots:**
   * The script generates and saves plots comparing the trends between treatment and control groups post-event, saved to the **Event Study Plot** folder for each state.