Data Dictionary

Stay-at-Home Order Mobility Impact Study

Dataset Overview

Dataset: data/processed/panel_mobility.csv

Purpose: Measure the impact of stay-at-home orders on workplace mobility using staggered difference-in-differences estimation.

This is a state-by-day panel dataset covering early 2020. Each row represents one U.S. state on one day. The dataset is structured to support staggered treatment timing, event-study analysis, and modern difference-in-differences estimators.

Column Schema

The dataset contains nine columns organized to support causal analysis with staggered policy adoption:

Column	Type	Description
`unit`	*string*	State name
`date`	*date*	Observation date
`outcome`	*float*	Workplace mobility (% change from baseline)
`time`	*integer*	Days since start of dataset
`policy_date`	*date / null*	Stay-at-home order start date
`first_treat`	*integer*	Treatment start time index
`ever_treated`	*integer (0/1)*	Ever treated indicator
`treated`	*integer (0/1)*	Post-treatment indicator
`post`	*integer (0/1)*	Alias of treated

Variable Definitions

Outcome Variable

`outcome`

Definition: Percent change in workplace mobility relative to a pre-COVID baseline.

Source: Google COVID-19 Community Mobility Reports

Interpretation:

- 0 = No change from baseline
- Negative values = Reduced workplace presence
- Positive values = Increased workplace presence

This is the primary dependent variable used in all analyses.

Panel Structure

`unit`

Identifies the state (panel unit).

`date`

Calendar date of the observation.

`time`

Numeric day index starting from the first date in the dataset. Used by estimators that require integer time variables.

Treatment Timing

`policy_date`

The date when a statewide stay-at-home order became effective. Null for never-treated states.

`first_treat`

The `time` index corresponding to `policy_date`. This is the key variable used by staggered difference-in-differences estimators to align treatment timing across states.

Treatment Indicators

`ever_treated`

Indicates whether a state adopts a stay-at-home order at any point.

`treated`

Time-varying indicator equal to 1 on and after the policy start date.

`post`

Alias for `treated`, included for readability and diagnostics.

Intended Use

This dataset is designed for:

1. Staggered difference-in-differences estimation
2. Event-study analysis
3. Estimator robustness checks

Minimum required columns for causal estimation:

- unit
- time
- outcome
- first_treat

Other columns exist to improve transparency and validation.

Key Assumptions and Notes

- Treatment adoption varies by state.
- Mobility measures behavioral response, not economic output.
- Estimates represent average within-state changes relative to control states.
- The dataset prioritizes methodological clarity over national representativeness.

Reproducibility

Raw data source: Google COVID-19 Community Mobility Reports

Preprocessing logic: `data/preprocess.py`

Note: The final dataset is fully reproducible from source files.