buildingcodes replication code

Overview

This repository includes replication code for "Mandated vs. Voluntary Adaptation to Natural Disasters: The Case of U.S. Wildfires" by Patrick Baylis and Judson Boomhower.

The analysis in this paper is built on assessment data collected by Zillow and made available through the now-discontinued ZTRAX program. Because of licensing restrictions, we are not able to provide the assessment data that we use in the paper, nor are we able to provide the final analysis data (since the units of observation are individual homes from the assessment data). Still, in the interest of providing maximizing transparency and making replication possible for users with access to confidential assessment data, this repository includes the the full codebase for building the analysis data. We also include the collected set of damage inspection data that we use to identify destroyed homes in the assessment data. Researchers with access to equivalent assessment data (from, e.g., CoreLogic or similar) should be able to replicate our analysis.

Code

The full replication code for the empirical analysis in the paper is contained in the `code/` folder as a series of R scripts, which should be run in the order specified by the filenames. The code relies on a number of different datasets, including the previously described assessment data and damage inspection data, as well as a number of other supporting datasets.

Users with access to equivalent assessment data should—in principle—be able to run the full build and analyses. In practice, they may have to modify some scripts to accommodate assessment data from other sources, since variable names and inclusion criteria may differ. We provide a brief description of the function of each script below:

- `globals.R`: Global variables and functions used in the analysis
- `01_DINS_Import.R`: Import and clean CAL FIRE Damage Inspection (DINS) data
- `01_Import_Clean_CA_Older.R`: Import and clean older California damage inspection data from incidents prior to the DINS program
- `01_Import_Clean_Other_States.R`: Import and clean damage inspection data from other states
- `01_Import_Clean_SD_Fires.R`: Import and clean damage inspection data from San Diego County fires
- `02a_ZTrax_Current.R`: Import and clean current ZTRAX data for incidents 2016 and later
- `02a_ZTrax_Historical.R`: Import and clean historical ZTRAX data for incidents prior to 2016
- `03a_cleanup_APN_formatting.R`: Clean up assessor parcel numbers (APNs) for DINS homes in incidents 2016 and later
- `03a_cleanup_APN_SD_older.R`: Clean up APNs for older San Diego County homes
- `03b_cleanup_APNs_Historical.R`: Clean up APNs for DINS homes in incidents prior to 2016
- `03b_cleanup_APNs_OtherStates.R`: Clean up APNs for homes in other states
- `03b_cleanup_APNs_older_CA.R`: Clean up APNs for older California homes
- `03c_dupe_UAPN_ZTRAX_CA.R`: Identify and remove duplicate homes in California ZTRAX data
- `03d_Merge_ZTRAX_Damage.R`: Merge ZTRAX data with damage inspection data
- `03e_Merge_to_LotLines.R`: Merge data with parcel boundary data
- `03f_Merge_Lots_Bldgs.R`: Merge data with building footprint data
- `04a_AssignLocations_and_GroundTruth.R`: Assign locations to homes and ground truth destroyed homes
- `04b_SubsetBurnedAreas.R`: Subset data to homes in burned areas
- `06_assign_spatial.R`: Identify home locations with respect FHSZ, SRA, LANDFIRE variables, and other spatial data
- `07a_find_neighbors.R`: Identify nearest neighbors for each home with respect to both centroid-to-centroid and wall-to-wall distances
- `07b_setup_neighbors.R`: Create a dataset of nearest neighbors for each home
- `10_NearMap_Maps.R`: Download and clean NearMap images
- `10_NearMap_QC_Geocoding.R`: Check quality of geocoding and damage indicator using NearMap images

- `20-main-data-setup.R`: Create the main dataset for analysis
- `21-main-descriptive.R`: Run descriptive statistics
- `22-main-regs.R`: Run main regression analyses
- `23-save-public-damage-data.R`: Save the cleaned damage inspection data for public use
- `40-compute-sra-lra-buffer.R`: Compute SRA and LRA buffers for hedonic analysis
- `41-draw-statewide-sample.R`: Draw a statewide sample of homes for hedonic analysis and cost-effectiveness analysis
- `41b-load-statewide-geocoding.R`: Load geocoding data for the statewide sample
- `42-add-statewide-vars.R`: Add variables to the statewide sample
- `45-load-home-sales-from-ztrax.R`: Load home sales data from ZTRAX
- `46-clean-homes-sales.R`: Clean home sales data
- `47-add-characteristics-to-sales.R`: Add characteristics to home sales data
- `52-hedonic-price-regs.R`: Run hedonic price regressions
- `53-hedonic-quantity-regs.R`: Run hedonic quantity regressions
- `60-estimate-baseline-mitigation.R`: Use DINS data to estimate mitigation in absence of building codes
- `61-compute-rebuild-costs-hazus.R`: Compute rebuilding costs using HAZUS data
- `62-statewide-costeff-setup.R`: Set up data for cost-effectiveness analysis
- `63-compute-cost-effectiveness-and-policy-counterfactuals.R`: Compute cost-effectiveness and policy counterfactuals
- `70-compare-samples.R`: Compare sample used in main analysis to broader set of homes in at-risk areas

Damage inspection data

To facilitate replication or additional projects by other users, we separately extract the full set of damaged homes we use in the analysis. Because we draw these homes from publicly available damage inspection records and remove any variables drawn from ZTRAX, we are able to make this dataset public. The cleaned version of the collected damage inspection data we use is contained in the `damage-data-public/` folder in this repository. This folder includes two files.

The first, `damaged-homes-public.csv`, is a table of homes that were identified as damaged or destroyed in the damage inspection data along with their unformatted assessor parcel number (which we use to merge to the assessment data), a variable indicated whether they were damaged or destroyed (`damaged`), and the relevant wildfire incident. Note that for this public-facing dataset, we report our `destroyed` variable from the paper as `damaged` here to limit user confusion, since this variable is a binary indicator of whether a home was damaged or not. In practice, as we discuss in the paper, most homes that are damaged to any degree by fire must be significantly or completely rebuilt, so the distinction between damaged and destroyed is not particularly meaningful.

Users should be able to use the combination of state-county FIPS and the assessor parcel number to link to assessment data and obtain addresses and geolocations (we omit these from this dataset since providing a complete set of addresses and geolocation would require us to draw from the variables obtained through ZTRAX). We restrict the data to the exact homes we identify as destroyed in the main analysis in the paper (readers can verify this by comparing the per-fire counts of destroyed homes in Appendix Table A.5 to this dataset).

The second file, `damaged-homes-perimeters.gpkg`, is a geopackage of incident perimeters that match to the incidents included in the dataset above.

Non-public datasets used in the paper

The following datasets are used in the paper but cannot be shared publicly for licensing reasons. We indicate each dataset with the name of the folder we store it in within our server's private data folder (`/data3/buildingcodes-jpe-data/input/private` in our code).

- `ZTRAX/`: Assessment data from Zillow's ZTRAX program (previously described)

- `ParcelBoundaryFiles/`: Parcel boundary files for counties with wildfire incidents
- `ESRI/`: Geocoding output obtained using ESRI's ArcGIS Geocoding Service

Public datasets used in the paper

The following datasets are publicly available datasets we use in the paper but do not include in this repository due to size. We mirror the copies we use (see below list) and also include links to where researchers can find them online. We indicate each dataset with the name of the folder we store it in within our server's data folder (`/data3/buildingcodes-jpe-data/input/public` in our code).

- `CA-HPI.csv`: Home price index data for California (https://fred.stlouisfed.org/series/CASTHPI)
- `ca_places/`: CA Geographic Boundaries (https://data.ca.gov/dataset/ca-geographic-boundaries)
- `CensusTiger/`: Census TIGER Shapefiles (https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html)
- `DamageData/`: Post-fire damage data from CAL FIRE DINS (https://data.ca.gov/dataset/cal-fire-damage-inspection-dins-data) and other sources (various)
- `FirePerimeters/`: Fire perimeter data from FRAP(https://www.fire.ca.gov/what-we-do/fire-resource-assessment-program), NIFC (https://www.nifc.gov/fire-information/maps), MTBS (https://www.mtbs.gov/direct-download), and other sources (various)
- `HAZUS/`: FEMA HAZUS data on rebuilding costs by Census block (https://www.fema.gov/flood-maps/products-tools/hazus)
- `LANDFIRE/`: LANDFIRE data on aspect, elevation, slope, and Anderson 13 fuel model (https://www.landfire.gov/)
- `Microsoft_Buildings/`: Microsoft Building Footprints from 2018 and 2021 (https://github.com/microsoft/USBuildingFootprints)
- `NolteEtAl/`: Appendix Tables from Nolte et al. 2024 (https://le.uwpress.org/content/100/1/200.abstract)
- `RiskToStructures/`: Risk to Structures raster from the Wildfire Risk to Communities project (https://www.fs.usda.gov/rds/archive/Catalog/RDS-2020-0016)
- `simplemaps_uscities_basicv1.72/`: Simplemaps US Cities Database (https://simplemaps.com/data/us-cities)
- `SRAandFHSZ/`: Collection of SRA and FHSZ shapefiles to identify homes in building code areas (CAL FIRE)

Below are the two Zenodo mirrors that include the versions of the public datasets we use:

- https://zenodo.org/records/14948185
- https://zenodo.org/records/14948225