Supplement of Nat. Hazards Earth Syst. Sci., 22, 1055–1072, 2022
https://doi.org/10.5194/nhess-22-1055-2022-supplement

Natural Hazards and Earth System Sciences
Open Access
EGU

Supplement of

Regional county-level housing inventory predictions and the effects on hurricane risk

Caroline J. Williams et al.

Correspondence to: Rachel A. Davidson (rdavidso@udel.edu)

Supplemental Section

This document serves as a supplement to the "Regional county-level housing inventory predictions and effects on hurricane risk" publication by Williams et al. (2022). All supporting data, code, and documentation is available on the DesignSafe-CI Data Depot under Project PRJ-3303 (Williams and Davidson, 2022). Any questions can be directed to the corresponding author, Rachel Davidson (rdavidso@udel.edu).

Supporting Publication

Williams, C. J., Davidson, R. A., Nozick, L. K., Trainor, J. E., Millea, M., and Kruse, J. L.: Regional county-level housing inventory predictions and the effects on hurricane risk, Nat. Hazards Earth Syst. Sci., 2022.

Supporting Published Data

Williams, C. and Davidson, R.: Regional county-level housing inventory predictions and the effects on hurricane risk using long-short term memory (LSTM) methods and applied to the southeastern United States (US), DesignSafe-CI, 2022.

Table of Contents

S1 Data Inputs

All supporting data used as input for this work is available in the Data folder within Project PRJ-3303 of the DesignSafe-CI Data Depot (Williams and Davidson, 2022).

S1.1 Data organization

S1.1.1 Source data

All downloaded source data for each variable is available in the Data/Downloaded_Data folder. See Sect. S1.2 for details pertaining to each data source. Some files within the Downloaded_Data folder were manually modified from the source file and saved as a new file, where modifications were mostly made to the naming or organization of column headings. These adjusted files have file names corresponding to the modifications; See Sect. S1.2 for details. Note that the data accessed via APIs were not saved in this folder.

State, county, 2010 census tracts, study area, and coastline boundary shapefiles are available in the Data/Boundaries folder. Data/Boundaries/counties_south/co_south_names.csv has a list of the GEOIDs for the 1,000 counties included in this study.

Data references in Section S1 of this document use bracketed citations based on the reference number in Sect. S1.4

S1.1.2 Configured data files

The data compiled for all variables was saved in separate csv files by year. The first three columns of each csv file always consist of:

1. "GEOID": the GEOID for all counties in the study region in that year
2. {variable name}: the value for each county in that year. The header for this column varies depending on the variable. (See Sect. S1.1.3.2 below for the corresponding column heading for each variable)
3. "year": the year corresponding to the data in column 2.

Occasionally, additional columns are provided (such as population density values in units of people per mi^2 instead of people per km^2), but the order of the first three columns always remains the same. Additionally, the number of rows in these columns does not always equal the total number of counties in the study area in 2019 (1,000 counties) because the number of counties varied between 1970 and 2019 (see Step 2 of Sect. S1.2).

S1.1.3 File and folder structure

S1.1.3.1 File naming

Each data file was named using the following scheme:
0_{short theme name}_{short variable name}_{year}_{date last updated in YYYMMDD format}.csv
Example: 0_econ_gdp_2010_20201204.csv
See Sect. S1.1.3.2 below for the list of theme names and variable names.

S1.1.3.2 Data folders

All csv files were saved in the following folder structure:
Data/{long theme name}/0_{short theme name}_{short variable name}/county/{csv file name}

Example: Data/Economic_Impact/0_econ_gdp/county/0_econ_gdp_2010_20201204.csv

where "fname1_co", "fname2_co", and "fdate_co" are available in Data/2021-04-15_DBI_file_name_scheme_co_dif_rate.csv. All columns within this file provide the following information:

1. "fname1_co": "{long theme name}/0_{short theme name}_{variable name}/"
2. "fname2_co": "0_{short theme name}_{short variable name}"
3. "fdate_co": "{date last updated in YYYMMDD format}.csv"
4. "colname_co": column name in the second column of data file (see Sect. S1.1.2)
5. "title_co": variable name
6. "units_co": variable units
7. 1970: "1" indicates data is available in 1970 and the data is saved in a data file. "0" indicates not available.

…

56. 2019: "1" indicates data is available in 2019 and the data is saved in a data file. "0" indicates not available.

S1.2 Data management procedure

Step 1. Collect and organize raw data from sources.

This step describes the process of sourcing data and saving data files using the organization scheme described in Sect. S1.1. Each section below provides the location of the source data, the downloaded or accessed source files, the name of the R script file that creates the associated data file, and the list of data files created. Note that there are more data files created than those listed in the compiled csv file described in Sect. S1.1.3.2. This is because some data files are repeated concepts of each other or are merely an avenue to compute a final

data point. For example, one set of data files were created with the number of married women, another set of data files include the number of married men, but neither of these files were used in the data analysis; they were merely used as a means to compute the percent married population.

a. Annual population estimates
    i. Source: US Census' Population and Housing Unit Estimates https://www.census.gov/programs-surveys/popest/data/data-sets.html [3,4]
    ii. Source files (available in Data/Downloaded_Data/CensusAnnualPopulationEstimates):
        1. co-asr-7079.csv: county-level annual population estimates for 1970s downloaded directly from Census website
        2. co-asr-7079_modified.csv: manually added column headers
        3. pe-02.xls: county-level annual population estimates from 1980s downloaded directly from Census website
        4. pe-02_modified: manually adjusted column headers
        5. 1990s county-level annual population estimates from Census API: https://api.census.gov/data/1990/pep/int_charagegroups
        6. 2000s county-level annual population estimates from Census API (had to split API request in two links): https://api.census.gov/data/2000/pep/int_charage
    iii. R script: pop-estimate.R
    iv. Data files created:
        1. Population
        2. Population identified as white
        3. Population not identified as white
        4. Percent of population identified as white
        5. Percent of population not identified as white
b. Annual housing unit estimates
    i. Source: US Census' Population and Housing Unit Estimates https://www.census.gov/programs-surveys/popest/data/data-sets.html [3]
    ii. Source files (available in Data/Downloaded_Data/CensusAnnualCountyHousingUnitEstimates):
        1. hu-est00int-02-{state FIPS code}.xls: county-level annual housing unit estimates for 2000s
        2. hu_est_co_0009_{state FIPS code}.csv: manually adjusted column headers and saved as csv
    iii. R script: hu-estimates.R
    iv. Data files created:
        1. Number of housing units
c. Decadal education attainment estimates

i. Source: US Census Bureau's Educational attainment for adults aged 25 and older for the U.S., States, and counties, 1970-2018 https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/ [5]
ii. Source files (available in Downloaded_Data/Education):
    1. Education_1970_2018_Original.csv: original data
    2. Education_1970_2018_Modified.csv: manually modified headers and added percentages to reflect % high school diploma, % some college, % college diploma, % grad diploma
iii. R script: education_census-estimates.R
iv. Data files created:
    1. Percent of population aged 25 and older with a high school diploma
    2. Percent of population aged 25 and older with a college degree

d. Other decadal values
    i. Source: IPUMS NHGIS https://data2.nhgis.org/main [6]
        1. This data requires a limited access license, which is available for free by registration here: https://uma.pop.umn.edu/nhgis/user/new
        2. The NHGIS data is not available for redistribution without permission nor for for-profit uses (https://www.nhgis.org/citation-and-use-nhgis-data)
    ii. Source files: Downloaded_Data/NHGIS_Decennial_Census_Data (downloaded in batches of zip files, which dictates the nhgisXXX sections below). There are many files available in this folder. Below are the files used for this project.
        1. nhgis0001
            a. nhgis0001_ts_nominal_county.csv (sourced from NHGIS time series data set)
                i. Total population: 1970, 1980, 1900, 2000 decadal values saved over annual estimates because there are more counties available in decadal data than annual estimates. Code: A00.
                ii. Total housing units: 2000 values saved over annual estimates because there are more counties available in decadal data than annual estimates. Code: A41.
                iii. Number of occupied housing units. Code: A43.
                iv. Number of vacant housing units. Code AT1.
                v. Number of owner-occupied housing units. Code B37.
                vi. Number of renter-occupied housing units. Code B37.
            b. nhgis0001_ts_nominal_county_modified.csv: manually removed unneeded columns and column headers
            c. nhgis0001_ts_nominal_county_codebook.txt: associated metadata
        2. nhgis0003

    a. nhgis0003_ts_nominal_county.csv (sourced from NHGIS time series data set)
        i. Includes number of male and females currently married in 1980, 1990, 2000 (includes married – not separated and married separated) for people ages 15 and older. Code: BM4.
    b. nhgis0003_ts_nominal_county_modified.csv: manually removed unneeded columns and column headers
    c. nhgis0003_ts_nominal_county_codebook.txt: associated metadata
3. nhgis0004
    a. nhgis0004_ds120_1990_county.csv (1990 Census: STF 1)
        i. Number of one-unit detached housing units in 1990. Source code: NH030A, NHGIS code: GAF.
    b. nhgis0004_ds120_1990_county_modified.csv: manually removed unneeded columns and column headers
    c. nhgis0004_ds120_1990_county_codebook.txt: associated metadata
    d. nhgis0004_ds151_2000_county.csv (2000 Census: SF 3a)
        i. Number of one-unit detached housing units in 2000. Source code: NH41, NHGIS code: ETH.
    e. nhgis0004_ds151_2000_county_modified.csv: manually removed unneeded columns and column headers
    f. nhgis0004_ds151_2000_county_codebook.txt: associated metadata
4. nhgis0005
    a. nhgis0005_ts_nominal_county.csv (sourced from NHGIS time series data set)
        i. Average household size for all occupied housing units. Calculates the average as a weighted average of 1, 2, 3, 4, 5, 6+ -person households for 1990 and 2000. Code: CV5.
    b. nhgis0005_ts_nominal_county_modified.csv: manually removed unneeded columns and column headers
    c. nhgis0005_ts_nominal_county_codebook.txt: associated metadata
5. nhgis0006
    a. nhgis0006_ts_nominal_county.csv (sourced from NHGIS time series data set)
        i. Median household income for 1980, 1990, 2000. Code: B79.
    b. nhgis0006_ts_nominal_county_modified.csv: manually removed unneeded columns and column headers
    c. nhgis0006_ts_nominal_county_codebook.txt: associated metadata
6. nhgis0007
    a. nhgis0007_ds120_1990_county.csv (1990 Census: STF 1)
        i. Median property value of specified owner-occupied housing units in 1990. Source code: NH23B. NHGIS code: EST.

   b. nhgis0007_ds120_1990_county_modified.csv: manually removed unneeded columns and column headers
   c. nhgis0007_ds120_1990_county_codebook.txt: associated metadata
   d. nhgis0007_ds151_2000_county.csv (2000 Census: STF 3a)
      i. Median property value of specified owner-occupied housing units in 2000. Source code: NH076A. NHGIS code: GB7.
   e. nhgis0007_ds151_2000_county_modified.csv: manually removed unneeded columns and column headers
   f. nhgis0007_ds151_2000_county_codebook.txt: associated metadata
7. nhgis0009
   a. nhgis0009_ds151_2000_county.csv (2000 Census: STF 3a)
      i. Migration in 1995 for persons 5 years and over. Considers moving from Puerto Rico as moving out of state. Calculated moving within same state but different county as (total moved, minus moved out of state, minus moved out of country, minus moved in same county). Source code: NP024A, NP024B, NP024C, NP024D, NP024E, NP024H. NHGIS code: GJG, GJH, GHI, GHH, GHK, GWO.
      ii. Median year moved in in 1995 of occupied units. Source code: NH039A. NHGIS code: GAP.
   b. nhgis0009_ds151_2000_county_modified.csv: manually removed unneeded columns and column headers
   c. nhgis0009_ds151_2000_county_codebook.txt: associated metadata
8. nhgis0010
   a. nhgis0010_ds151_2000_county.csv (2000 Census: STF 3a)
      i. Aggregate paid property taxes in 2000. Source code: NHCT021A. NHGIS code GEQ.
      ii. Aggregate property values in 2000. Source code: NH079A. NHGIS code: GCB.
   b. nhgis0010_ds151_2000_county_modified.csv: manually removed unneeded columns and column headers
   c. nhgis0010_ds151_2000_county_codebook.txt: associated metadata
9. nhgis0011
   a. nhgis0010_ds151_2000_county.csv (2000 Census: STF 3a)
      i. Median year structure was built of all housing units in 2000. Source code: NH035A. NHGIS code: GAK.
   b. nhgis0011_ds151_2000_county_modified.csv: manually removed unneeded columns and column headers
   c. nhgis0011_ds151_2000_county_codebook.txt: associated metadata

10. nhgis0013
    a. nhgis0013_ts_nominal_county.csv (sourced from NHGIS time series data set)
        i. Median age of population in 1980, 1990, 2000. Code AR9.
    b. nhgis0013_ts_nominal_county_modified.csv: manually removed unneeded columns and column headers
    c. nhgis0013_ts_nominal_county_codebook.txt: associated metadata

iii. R script: decennial_to_csv.R

iv. Data files created:

1. Population
2. Number of housing units
3. Number of occupied housing units
4. Number of vacant housing units
5. Percent of vacant housing units out of all housing units
6. Number of owner-occupied housing units
7. Number of renter-occupied housing units
8. Percent of owner-occupied housing units out of all occupied housing units
9. Number of married males
10. Number of married females
11. Number of one-unit detached housing units
12. Percent of one-unit detached housing units out of all housing units
13. Number of one-unit detached housing units per capita
14. Average household size
15. Median household income
16. Median property value
17. Median household tenure
18. Population that did not move
19. Population that moved within the county
20. Population that moved to a different county within the state
21. Population that moved to a different state
22. Population that moved to a different country
23. Aggregate paid property taxes
24. Aggregate property values
25. Average property tax rate
26. Median year housing structure was built
27. Median age of population`

e. Developed Land cover

i. Source: Multi-Resolution Land Characteristics Consortium (MRLC) https://www.mrlc.gov/data [7]

ii. Source files: Data/Downloaded_Data/NLCD_{YEAR}_Land_Cover_L48_20190424
    1. Data available in 2001, 2004, 2006, 2008, 2011, 2013, 2016 and values are linearly interpolated for in-between years 200 and 2019.
    2. See "NLCD_procedure.docx" document in Data/Downloaded_Data folder for processing procedures.
    3. This procedure uses GIS methods to calculate the percent developed area within a county. "Developed" includes land cover areas designated as "developed, open space" (code: 21), "developed, low intensity" (code 22), "developed, medium intensity" (code 23), "developed, high intensity" (code 24).
iii. R script: nlcd_processing.R
iv. Data files created:
    1. Percent of county area designated as open space
    2. Percent of county area designated as low intensity development
    3. Percent of county area designated as medium intensity development
    4. Percent of county area designated as high intensity development
    5. Percent of county area developed

f. Non-Buildable Area & Proximity to Coast
    i. Source: This data set consists of a series of national and state protected area, water areas, wetlands, parks, or other assumed no-build areas available on national and state websites as of Summer 2020 [8-61]. Steep slopes and poor soils were not included in the assumed non-buildable area.
    ii. Source files: See "Non-Buildable-Area Method.xlsx" document in Data/Downloaded_Data/Non-Buildable-Area folder. Rows with "Yes" in the "Included in Non-Buildable-Land Variable" column are included in the analysis.
        1. These files were imported into GIS software, clipped to the study area, and merged together to form a single non-buildable area layer for each state, except Texas, which only has non-buildable area by county.
            a. Texas non-buildable area could not be merged into one layer because the wetland file is very large (and is provided by NFWS as 3 separate files). Thus, the non-buildable area was processed by county. For counties that had two wetland files overlapping one county, we paid special attention ensure data-merging was done correctly to combine data from both wetland shapefiles. These counties are listed in the "Texas_Wetland_Overlap_Counties.xlsx" and the shapefile regions are in the "Texas Wetland Region Combine" folder located in file in Data/Downloaded_Data/Non-Buildable-Area/48000_land_wetland1_2019_20200813.shp.

2. Final shapefiles per state are available in Data/Land_Classification/0_land_nobuild _2020_shapefile_20201112 folder
3. Percent buildable area was calculated using the R script listed below and saved by state in csv files located in Data/Land_Classification/0_land_nobuild_2020_csv_20201201
4. csv files for all counties in study area are then created for each year in analysis and saved in Data/Land_Classification/0_land_nobuild. Therefore, we assume that the non-buildable areas remain constant over the entire analysis period (1970-2019).

iii. R script: Nobuild.R

iv. Data files created:

1. Percent of county area assumed to be "non-buildable"

g. ACS & other 2009-2019 values

i. Source: NOAA Shoreline [1], US Census Bureau American Community Survey (ACS) API [62,63], US Bureau of Economic Analysis [64], Bureau of Labor Statistics [65], US Census Bureau Building Permits Survey [66], TIGER Shapefiles [2]

ii. Source files:

1. Proximity-to-coast: Data/Downloaded_Data/us_medium_shoreline. Assumed that the distance of the centroid of a county to the coast remains constant over the entire study period.
2. ACS: tidycensus R package https://cran.r-project.org/web/packages/tidycensus/tidycensus.pdf. Data available 2009/2010-2019
3. GDP Data: CAGDP1 GDP current-dollar values for all industries. Values listed in thousands of dollars. Original files available in Data/Downloaded_Data/CAGDP1_GDP_CurrentDollars. The BEA combined independent cities in Virginia with surrounding big counties, which is addressed in the "Step 2: Adjust to 2019 boundaries" section below. Data available annually since 2001
4. Employment data: number of employed people (annual averages) for all industries and types of employers. Original files are available in Data/Downloaded_Data/BLS_QCEW. Data available annually since 1990.
5. Building permit data: available in Data/Downloaded_Data/Building_Permits file and only building permits for one-unit houses are included in the analysis. Data available annually since 1990.

iii. R script: pre_raw_values_county_calcs

iv. Data files created:

1. Distance of county centroid to coastline
2. Number of housing units

3. Number of one-unit detached housing units
4. Number of occupied housing units
5. Number of vacant housing units
6. Number of owner-occupied housing units
7. Number of renter-occupied housing units
8. Average household size of all housing units
9. Average household size of owner-occupied housing units
10. Average household size of renter-occupied housing units
11. Median year residential structure was built
12. Median household tenure of all housing units
13. Median household tenure of owner-occupied housing units
14. Median household tenure of renter-occupied housing units
15. Population identified as white
16. Median age of population
17. Married female population
18. Married male population
19. Population
20. Median property value
21. Aggregated property value
22. Median income
23. Median household income
24. Population aged 25 or older
25. Population with high school degree
26. Population with some college experience
27. Population with college degree
28. Population with a graduate degree
29. Population that did not move
30. Population that moved within the county
31. Population that moved to a different county within the state
32. Population that moved to a different state
33. Population that moved to a different country
34. Aggregate property taxes paid
35. GINI coefficient
36. Percent of population aged 25 and older with a high school diploma
37. Percent of population aged 25 and older with a college degree
38. Average property tax rate
39. GDP
40. Average annual employment

41. Number of one-unit building permits

Step 2. Adjust to 2018 boundaries

a. A list of counties with either boundary changes or name changes for the counties of our study area since 1970 is available in the "dbi_county_boundary_changes_1970_2020.xlsx" in the Data/ Downloaded_Data folder. Sourced from https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2010.html [67]
b. 2019 boundaries are used for this analysis. Since 1970, there were 8 Virginian counties that were joined with existing counties. There were also 3 new counties in Virginia in the 70s. Miami-Dade County's name and FIPS code also changed in the 2000s.
c. csv files named 2020-12-03_DBI_file_name_name_scheme_{calc type}.csv in the Data folder are used to identify which variables should add two counties together, which should calculate the weighted average based on population, and which should calculate the weighted average based on the number of housing units.
d. R script: raw_values_county_adjustment.R
e. Data files created: This document does not create new files but rather modifies the created data files so that data across all years between 1970 and 2019 reflect 2019 boundaries.

Step 3. Calculate relevant variables (percent, density, per capita, etc.)

a. This step calculates ratio values, such as population density, vacancy rate, GDP per capita and other variables.
b. R script: calculated_values_post_county_adjustment.R
c. Data files created:
    a. Population density
    b. Number of housing units per square kilometer
    c. Percent of one-unit detached housing units of total housing units
    d. Population not identified as white
    e. Percent of population identified as white
    f. Percent of population not identified as white
    g. Percent married population
    h. Average paid property tax
    i. Percent vacant housing units of total housing units
    j. Percent owner-occupied housing units of occupied housing units
    k. Number of one-unit detached housing units per capita
    l. Percent of population that did not move
    m. Percent of population that moved within the county
    n. Percent of population that moved to a different county within the state

o. Percent of population that moved to a different state
p. Percent of population that moved to a different country
q. GDP per capita
r. Average annual employment per capita
s. Number of building permits per number of one-unit detached housing units

Step 4. Combine values for all variables, all years, and all counties in one csv

a. This step combines all variables for all years in all counties that were collected in Steps 1 through 3 into one csv file saved in the Data/Variable Combine folder.
b. The Data/2021-04-15_DBI_file_name_scheme_co.csv file in the Data folder shows which years each variable is available, where a "1" indicates it is available and a "0" indicates it's not available.
c. R script: DBI_Var_Combine.R

Step 5. Calculate linear interpolations for decadal values

a. This step uses the file created in Step 4 and performs annual linear interpolations for all variables that are only available on a decadal basis and creates a new variable_combine csv located at Data/Variable_Combine/2021-04-15_1970_2019_co/2021-04-15_DBI_file_name_scheme_co_vars_interpolation.csv.
b. R script: DBI_linear_interp.R

Step 6. Calculate annual difference and percent change values for all variables

a. This step calculates the annual difference and annual percent change values for all variables using the csv file created in Step 5 to produce 2021-04-15_DBI_co_vars_interpolation_dif_rate.csv located in the Data/Variable_Combine/2021-04-15_1970_2019_co/ folder. These values are calculated using the following equations
    i. Annual difference, *d*:

$$d_t = v_t - v_{t-1}$$

    ii. Annual percent change, *r*:

$$r_t = \frac{v_t - v_{t-1}}{v_{t-1}} * 100$$

b. R script: DBI_annual_dif_rate.R

Step 7. Final Adjustments

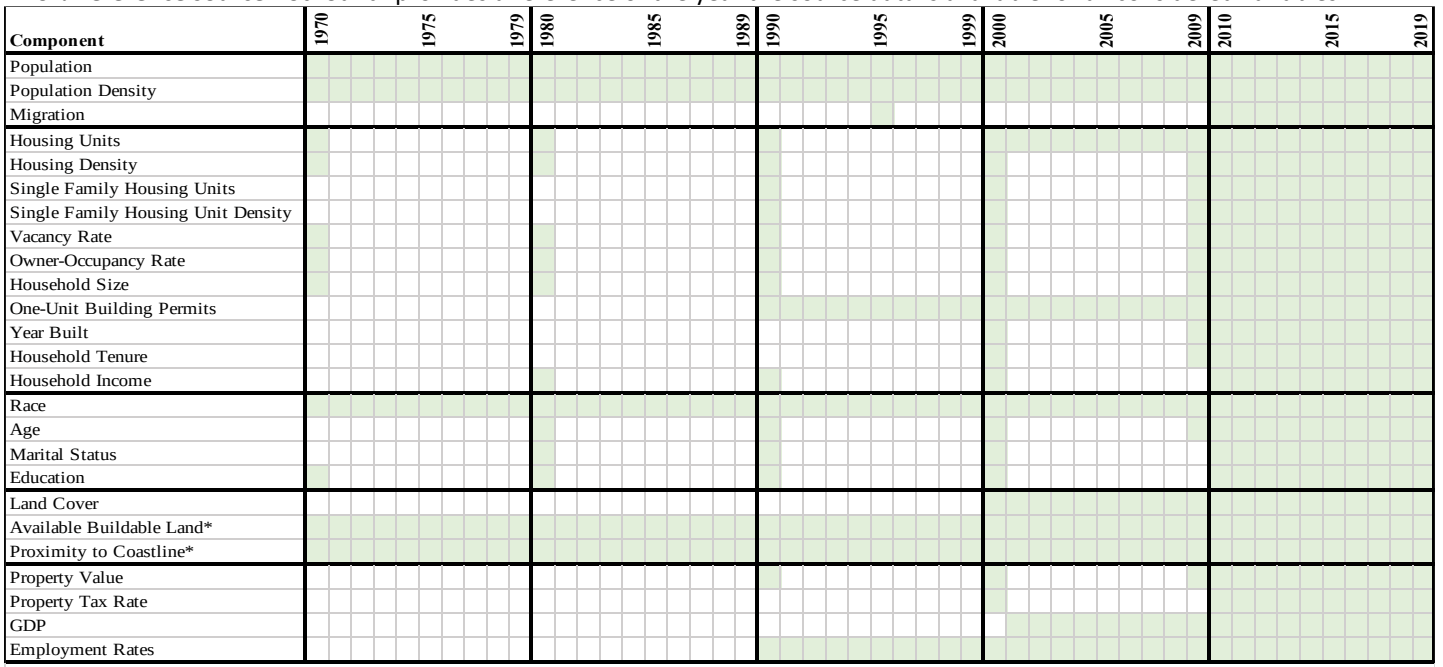
a. Upon a detailed review of the compiled data, a series of lingering items were addressed. The resulting file is 2021-04-15_DBI_co_vars_interpolation_dif_rate_adj.csv located in the Data/Variable_Combine/2021-04-15_1970_2019_co/ folder. A list of the final adjustments are as follows:
    i. The variables related to population, number of housing units, race, and education are not available for the 3 new Virginia counties established in the 1970s (GEOID: 51683, 51685, 51735) in 1970. The proportional values based on data in 1980 were applied to assign estimated 1970 values. For example, the total population of Prince William County (GEOID: 51153), Manassas City (GEOID: 51683), and Manassas Park (GEOID: 51685) in 1980 was first summed, then the proportion of population in the three jurisdictions was applied to 1970 population values in Prince William County to estimate the population in the areas of current-day Prince William County, Manassas City and Manassas Park in 1970.
    ii. The same methodology was applied to Bristol City, VA (GEOID: 51520) to calculate variables related to the number of one-unit detached housing as well as median property values in 1990. 2000 values for Bristol City (GEOID: 51520) and Washington County, VA (GEOID: 51191) were sourced to estimate the number of one-unit detached homes and median property values in 1990 in Bristol City.
    iii. The variables listed in part a and b above are then calculated for inter-decadal values using an annual linear interpolations (Step 5 above), then the annual difference and annual percent change (Step 6 above) is calculated.
    iv. The inv_avgproptax_rate variable for Loving County, TX (GEOID: 48301) in 2001 was replaced with a zero because the original value for this county in 2000 is zero, causing an infinity value for average property tax rate annual percent change in 2001.
    v. The dem_col_pct_rate variable for Loving County, TX (GEOID: 48301), Clay County, TX (GEOID: 48077), Roberts County, TX (GEOID: 48393) is replaced with a zero instead of a NA. The dem_col_pct values are zero in 1970 for these counties.
    vi. All building-permit related variables with NAs are replaced by zeros because it was assumed that an NA value would correlated to zero.
    vii. If race percent is negative, it's replaced with a zero value. This happens 108 times out of 50,000 values.
    viii. 'land_nobuild_dif', 'land_nobuild_rate', 'prox_coast_dif', 'prox_coast_rate' are all removed because the non-buildable area and proximity to coastline data is assumed to be unchanging over the time period analyzed.
b. This procedure results in no infinity values within the data.
c. R script: DBI_all_vars_final_adj.R

Step 8. Forward-Fill and Final Data File

a. Out of the 118 variables over 50 years for 1,000 counties, there were 488 instances of isolated NaN values (where data is available in years previous and after, yet a NaN value is given for a given year). These values were largely from variables sourced from the ACS in years 2009-2019 in counties with low population. To address these NaN values, a "forward fill" approach was used. The forward fill approach takes values from the year prior to the NaN value as the value that replaces the NaN value.
b. The resulting file, and final data file, is Data/Variable_Combine/2021-04-15_1970_2019_co/2021-04-15_DBI_co_vars_interpolation_dif_rate_adj_ffill.csv.
c. Python script: DBI_all_vars_final_ffill.ipynb

S1.3 Data Availability Over Time (Figure S1)

Error! Reference source not found. provides a reference of the year the source data is available for all considered variables.



*Assumed to not change over all years

Figure S1. Source data availability over time

S1.4 Data References (Table S1)

Table S1. Data References

Ref. Number	Name of Data Set	Year Accessed	Folder Saved	Link
1	NOAA Medium Resolution Shoreline	2020	Data/Boundaries/coastline_south	https://shoreline.noaa.gov/data/datasheets/medres.html
2	US Census Bureau TIGER/Line Shapefiles	2020	Data/Boundaries/counties_south/counties_south.shp Data/Boundaries/states_south/states_south.shp Data/Boundaries/qgis_studyareaboundary.shp	https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2019.html
3	US Census Bureau Population and Housing Unit Estimates	2020	Data/Downloaded_Data/CensusAnnualPopulationEstimates Data/Downloaded_Data/CensusAnnualCountyHousingUnitEstimates	https://www.census.gov/programs-surveys/popest/data/data-sets.html
4	US Census Bureau Population Estimates APIs	2020	N/A - API	https://www.census.gov/data/developers/data-sets/popest-popproj/popest.2000-2010_Intercensals.html
5	Educational attainment for the U.S., States, and counties, 1970-2019	2020	Data/Downloaded_Data/Education	https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
6	IPUMS National Historical GIS	2020	Data/Downloaded_Data/NHGIS_Decennial_Census_Data	https://data2.nhgis.org/main
7	Multi-Resolution Land Characteristics (MRLC): National Land Cover Database 2001, 2004, 2006, 2008, 2011, 2013, 2016	2020	Data/Downloaded_Data/NLCD_2001_Land_Cover_L48_20190424 Data/Downloaded_Data/NLCD_2004_Land_Cover_L48_20190424 Data/Downloaded_Data/NLCD_2006_Land_Cover_L48_20190424 Data/Downloaded_Data/NLCD_2008_Land_Cover_L48_20190424 Data/Downloaded_Data/NLCD_2011_Land_Cover_L48_20190424 Data/Downloaded_Data/NLCD_2013_Land_Cover_L48_20190424 Data/Downloaded_Data/NLCD_2016_Land_Cover_L48_20190424	https://www.mrlc.gov/data
8	Homeland Infrastructure Foundation-Level Data (HIFLD) Open Data: National Forests for National Forest System (NFS)	2020	Data/Downloaded_Data/Non-Buildable-Area/0_land_park1_2020_20200712.shp	https://hifld-geoplatform.opendata.arcgis.com/datasets/national-forests/explore
9	HIFLD Open Data: CBRA Boundaries from US Fish & Wildlife Services (USFWS)	2020	Data/Downloaded_Data/Non-Buildable-Area/0_land_protected1_2020_20200712.shp	https://hifld-geoplatform.opendata.arcgis.com/datasets/coastal-barrier-resources-system-cbrs-boundaries/explore
10	Protected Areas Database (PAD-US)	2020	Data/Downloaded_Data/Non-Buildable-Area/0_land_protected3_2018_20200915.shp	https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/science/pad-us-data-download?qt-science_center_objects=0#qt-science_center_objects

Ref. Number	Name of Data Set	Year Accessed	Folder Saved	Link
11	Homeland Infrastructure Foundation-Level Data (HIFLD) Open Data: Water Areas for National Flood Hazard Layers (NFHL)	2020	Data/Downloaded_Data/Non-Buildable-Area/0_land_water1_2020_20200712.shp	https://hifld-geoplatform.opendata.arcgis.com/datasets/geoplatform::lomrs/about
12	Homeland Infrastructure Foundation-Level Data (HIFLD) Open Data: Lakes for Inland Electronic Navigational charts (IENCs)	2020	Data/Downloaded_Data/Non-Buildable-Area/0_land_water2_2020_20200712.shp	https://hifld-geoplatform.opendata.arcgis.com/datasets/lakes-usace-ienc/explore?showTable=true
13	US Fish and Wildlife Services, National Wetlands Inventory	2020	Data/Downloaded_Data/Non-Buildable-Area/01000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/10000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/11000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/12000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/13000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/22000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/24000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/28000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/37000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/45000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/48000_land_wetland_2019_20200813.shp Data/Downloaded_Data/Non-Buildable-Area/51000_land_wetland_2019_20200813.shp	https://www.fws.gov/wetlands/Data/State-Downloads.html
14	FirstMap OpenData: Delaware State Forest	2020	Data/Downloaded_Data/Non-Buildable-Area/10000_land_park2_2017_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/c24b9bfe215040629739dd8a9c877e35/info/metadata/metadata.xml?format=default&output=html
15	FirstMap OpenData: Delaware Nature Area	2020	Data/Downloaded_Data/Non-Buildable-Area/10000_land_park3_2020_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/520e84a0e9dc424eb3ea75fdde00b896/info/metadata/metadata.xml?format=default&output=html
16	FirstMap OpenData: Public Protected Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/10000_land_protected2_2019_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/8fb7b97cc9484bd8bc30795f7f551b35/info/metadata/metadata.xml?format=default&output=html
17	FirstMap OpenData: Agricultural Preservation Districts	2020	Data/Downloaded_Data/Non-Buildable-Area/10000_land_protected5_2019_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/910ddf3ceb164b118bae8791c576f8ed/info/metadata/metadata.xml?format=default&output=html

Ref. Number	Name of Data Set	Year Accessed	Folder Saved	Link
18	FirstMap OpenData: Delaware Forest Land Preservation and Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/10000_land_protected6_2019_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/60e650ca5b86456bb6ce8ce679647ac3/info/metadata/metadata.xml?format=default&output=html
19	FirstMap OpenData: Delaware Forestry Conservation Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/10000_land_protected7_2020_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/25ab4ea7756f4a80af5512dc90cc37ea/info/metadata/metadata.xml?format=default&output=html
20	Open Data DC: District Government Lands (Owned, Operated, and/or Maintained)	2020	Data/Downloaded_Data/Non-Buildable-Area/11000_land_park2_2020_20200817.shp	https://www.arcgis.com/sharing/rest/content/items/c8cacaa7f74944d19a9941bc7bdee1a2/info/metadata/metadata.xml?format=default&output=html
21	Open Data DC: Recreation Fields	2020	Data/Downloaded_Data/Non-Buildable-Area/11000_land_park3_2020_20200817.shp	https://www.arcgis.com/sharing/rest/content/items/c8cacaa7f74944d19a9941bc7bdee1a2/info/metadata/metadata.xml?format=default&output=html
22	Open Data DC: Reservation Boundaries	2020	Data/Downloaded_Data/Non-Buildable-Area/11000_land_protected1_2020_20200817.shp	https://www.arcgis.com/sharing/rest/content/items/0ac4302b2e354fad986f07199e73a19e/info/metadata/metadata.xml?format=default&output=html
23	Open Data DC: Waterbodies	2020	Data/Downloaded_Data/Non-Buildable-Area/11000_land_water1_2020_20200816.shp	https://www.arcgis.com/sharing/rest/content/items/db65ff0038ed4270acb1435d931201cf/info/metadata/metadata.xml?format=default&output=html
24	MyFlorida GeoData: State Forests	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_park1_2020_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/993324f794a24944b38865b4085cb09b/info/metadata/metadata.xml?format=default&output=html
25	MyFlorida GeoData: State Park Boundaries	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_park3_2019_20200714.shp	https://www.arcgis.com/sharing/rest/content/items/bd190cf3d3934fbd9529dfe1c8c8772c/info/metadata/metadata.xml?format=default&output=html
26	MyFlorida GeoData: State Owned Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_park5_2020_20200716.shp	https://www.arcgis.com/sharing/rest/content/items/4f4a9b9d7f494eda9f1262831da2931b/info/metadata/metadata.xml?format=default&output=html
27	MyFlorida GeoData: Conservation Owned Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_protected2_2018_20200716.shp	https://www.arcgis.com/sharing/rest/content/items/47478e38d2634ae0a0e01ffc717e286d/info/metadata/metadata.xml?format=default&output=html
28	MyFlorida GeoData: Conservation Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_protected3_2018_20200716.shp	https://www.arcgis.com/sharing/rest/content/items/657c9b9a51ae48dc8ce1fe99f7b5881e/info/metadata/metadata.xml?format=default&output=html
29	MyFlorida GeoData: Conservation Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_protected4_2019_20200716.shp	https://www.arcgis.com/home/item.html?id=f303ab5da96c4cf89e1dd22bdb3e0594
30	MyFlorida GeoData: Mitigation Banks	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_protected5_2020_20200716.shp	https://www.arcgis.com/sharing/rest/content/items/63be3554c59a4fcc8af4f34c64cb43b4/info/metadata/metadata.xml?format=default&output=html
31	MyFlorida GeoData: CWA Surface Water Class Boundaries (Class I & II)	2020	Data/Downloaded_Data/Non-Buildable-Area/12000_land_water3_2020_20200716.shp	https://www.arcgis.com/sharing/rest/content/items/ddad454b0fa345e4b93ea37ed73add61/info/metadata/metadata.xml?format=default&output=html
32	Georgia Data Clearinghouse: DNR Managed Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/13000_land_park2_2017_20200716.shp	https://data.georgiaspatial.org/data/statewide/dnr/fed_lands/dnr20a.html

Ref. Number	Name of Data Set	Year Accessed	Folder Saved	Link
33	Georgia Data Clearinghouse: Other State Lands not owned by DNR	2020	Data/Downloaded_Data/Non-Buildable-Area/13000_land_parks4_2014_20200716.shp	https://data.georgiaspatial.org/data/statewide/dnr/fed_lands/other_state14.html
34	Georgia Data Clearinghouse: Conservation Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/13000_land_protected2_2019_20200716.shp	https://data.georgiaspatial.org/data/statewide/dnr/fed_lands/ConservationLands_2019.html
35	Georgia Data Clearinghouse: Natural Resource Conservation Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/13000_land_protected3_2018_20200716.shp	https://data.georgiaspatial.org/data/statewide/dnr/fed_lands/nrcs18b.html
36	Lousiana Department of Administration: State Land and Building Systems (SLABS)	2020	Data/Downloaded_Data/Non-Buildable-Area/22000_land_parks1_2020_20200808.shp	https://www.doa.la.gov/doa/osl/gis-data/
37	iMAP Maryland: Maryland Federal Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_park1_2019_20200807.shp	https://www.arcgis.com/sharing/rest/content/items/e8c94de89d6f4319bef28897c6a89581/info/metadata/metadata.xml?format=default&output=html
38	iMAP Maryland: Local Protected Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected2_2020_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/77c20913139d4534b1aa5c72f18d1cde/info/metadata/metadata.xml?format=default&output=html
39	iMAP Maryland: Federal Protected Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected3_2019_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/47defd42c35e4edab5a7e2b836b7bdce/info/metadata/metadata.xml?format=default&output=html
40	iMAP Maryland: Private Conservation Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected6_2020_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/8e7dc14259cb407bb88e2d29312e3395/info/metadata/metadata.xml?format=default&output=html
41	iMAP Maryland: DNR Owned Properties and Conservation Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected7_2019_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/a5de633b6ebf450b94dcb08b519c3414/info/metadata/metadata.xml?format=default&output=html
42	iMAP Maryland: Agricultural Land Preservation Foundation Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected8_2020_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/fe11b829c9a0477b8a370c063b05e29e/info/metadata/metadata.xml?format=default&output=html
43	iMAP Maryland: Coastal and Estuarine Land Conservation Program	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected9_2019_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/7abfc7862f0c4d10ac7ca61951570eda/info/metadata/metadata.xml?format=default&output=html
44	iMAP Maryland: Transfer of Development Rights (TDR) and Purchase Development Rights	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected10_2020_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/524e46dfb56b46729c23dcffae5c1587/info/metadata/metadata.xml?format=default&output=html
45	iMAP Maryland: Environmental Trust Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected11_2020_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/4a6a516aebcc4992ab1164bf5171c277/info/metadata/metadata.xml?format=default&output=html
46	iMAP Maryland: Forest Conservation Act Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_protected12_2019_20200808.shp	https://www.arcgis.com/sharing/rest/content/items/16c99455af424ae699c670473e35d2fa/info/metadata/metadata.xml?format=default&output=html

Ref. Number	Name of Data Set	Year Accessed	Folder Saved	Link
47	iMAP Maryland: Waterbodies - Lakes	2020	Data/Downloaded_Data/Non-Buildable-Area/24000_land_water1_2019_20200807.shp	https://www.arcgis.com/sharing/rest/content/items/71a539948a5b4751b3b676aec5368ed5/info/metadata/metadata.xml?format=default&output=html
48	Mississippi Automated Resource Information System (MARIS): State Parks	2020	Data/Downloaded_Data/Non-Buildable-Area/28000_land_park1_2019_20200806.shp	https://www.maris.state.ms.us/METADATA/PDF/MS_State_Parks_meta.pdf
49	Mississippi Geospatial Data Catalog: Marsh	2020	Data/Downloaded_Data/Non-Buildable-Area/28000_land_water1_2010_20200916.shp	https://www.arcgis.com/home/item.html?id=026fa9940e894a829774879a610da460
50	Mississippi Geospatial Data Catalog: Waterbodies (Small) - Inland	2020	Data/Downloaded_Data/Non-Buildable-Area/28000_land_water2_2020_20200806.shp	https://www.arcgis.com/home/item.html?id=026fa9940e894a829774879a610da460
51	Mississippi Geospatial Data Catalog: Waterbodies (Large) - Inland	2020	Data/Downloaded_Data/Non-Buildable-Area/28000_land_water3_2020_20200806.shp	https://www.arcgis.com/home/item.html?id=026fa9940e894a829774879a610da460
52	Mississippi Geospatial Data Catalog: Waterbodies (Small) - Coastal	2020	Data/Downloaded_Data/Non-Buildable-Area/28000_land_water4_2020_20200806.shp	https://www.arcgis.com/home/item.html?id=026fa9940e894a829774879a610da460
53	NC OneMap: State Owned Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/37000_land_park1_2020_20200623.shp	https://www.arcgis.com/sharing/rest/content/items/5f59713d0f8b4183985485deb5886349/info/metadata/metadata.xml?format=default&output=html
54	NC OneMap: Game lands	2020	Data/Downloaded_Data/Non-Buildable-Area/37000_land_park2_2020_20200623.shp	https://www.arcgis.com/sharing/rest/content/items/e5ddff9b96204c6181be7c022e61d946/info/metadata/metadata.xml?format=default&output=html
55	NC OneMap: Surface Water	2020	Data/Downloaded_Data/Non-Buildable-Area/37000_land_wetland1_2019_20200813.shp	https://www.arcgis.com/sharing/rest/content/items/c2f641dfffb74852835f6c733cc4d31a/info/metadata/metadata.xml?format=default&output=html
56	The Texas General Lands Office (GLO): State Agency Lands	2020	Data/Downloaded_Data/Non-Buildable-Area/48000_land_park1_2018_20200812.shp	https://www.glo.texas.gov/land/land-management/gis/index.html
57	Texas Parks and Wildlife (TPWD): State Park Boundaries	2020	Data/Downloaded_Data/Non-Buildable-Area/48000_land_park2_2016_20200812.shp	https://tpwd.texas.gov/gis/#data
58	Texas Tech University Center for Geospatial Technology: Lakes	2020	Data/Downloaded_Data/Non-Buildable-Area/48000_land_water1_2012_20200814.shp	https://www.depts.ttu.edu/geospatial/center/Data/TxStateLayers/Tx_Hydrology/_Metadata/Tx_Lakes_HighDetail_TIGER_Metadata.html
59	ArcGIS: State Parks	2020	Data/Downloaded_Data/Non-Buildable-Area/51000_land_park1_2015_20200814.shp	https://www.arcgis.com/home/item.html?id=6842cfbf87264b1c98f881bf01b0b7dc
60	Virginia Department of Conservation and Recreation: Conservation lands not including easements	2020	Data/Downloaded_Data/Non-Buildable-Area/51000_land_protected2_2020_20200814.shp	https://www.dcr.virginia.gov/natural-heritage/document/conslands.pdf
61	Virginia Department of Conservation and Recreation: Conservation Easements	2020	Data/Downloaded_Data/Non-Buildable-Area/51000_land_protected3_2020_20200814.shp	https://www.dcr.virginia.gov/natural-heritage/document/easements.pdf

Ref. Number	Name of Data Set	Year Accessed	Folder Saved	Link
62	American Community Survey 5-Year Data (2009-2019) API	2020	N/A - API	https://www.census.gov/data/developers/data-sets/acs-5year.html
63	tidycensus R library	2020	N/A - R Library	https://cran.r-project.org/web/packages/tidycensus/tidycensus.pdf
64	US Bureau of Economic Affairs Regional Data: GDP and Personal Income	2020	Data/Downloaded_Data/CAGDP1_GDP_CurrentDollars	https://apps.bea.gov/itable/iTableHTML.cfm?ReqID=70&step=1
65	US Bureau of Labor Statistics Quarterly Census of Employment and Wages	2020	Data/Downloaded_Data/BLS_QCEW	https://www.bls.gov/cew/downloadable-data-files.htm
66	US Census Bureau Building Permits Survey	2020	Data/Downloaded_Data/Building_Permits	https://www.census.gov/construction/bps/
67	US Census Bureau Substantial Changes to Counties and County Equivalent Entities: 1970-Present	2020	Data/Downloaded_Data/dbi_county_boundary_changes_1970_2020.xlsx	https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2010.html
68	Texas Department of Insurance: Hurricane Harvey Data Call Data through September 30	2021	Data/Downloaded_Data/Hurricane_Harvey_Insurance_Payout_Data	https://www.tdi.texas.gov/reports/documents/harvey-dc-04252019.pdf

S2 Models

All scripts used in this study to develop the three candidate models are available in the Models folder within Project PRJ-3303 of the DesignSafe-CI Data Depot (Williams and Davidson, 2022). Outputs from these scripts are saved in the Outputs folder.

S2.1 Linear Trend Models

The Python script used to develop the linear trend models is available in the Models/HousingInventoryPrediction_AnnualCounties_Linear.ipynb file.

S2.1.1 Inputs

Users are required to specify which test they want to run (Tests A, B, or C) in the second cell of the python notebook.

The modeling requires the following packages:

1. Pandas (Pandas Development Team, 2020)
2. NumPy (Harris et al., 2020)
3. Scikit-learn (Buitinck et al., 2011)

Input data is available in the Data folder of the project

S2.1.2 Model Parameters

None to note.

S2.1.3 Outputs

The code saves a csv with the $RMSE_r$, $E[|H|]$, and $s_{|H|}$ (see Eq. 3 and Eq. 4 in Sect. 5 of the publication) for each input and output combination for each test. This csv is located at Outputs/HousingInventoryPrediction_AnnualCounties_Linear_Results.csv.

S2.2 Autoregressive Integrated Moving Average (ARIMA) Models

The Python script used to develop the linear trend models is available in the Models/HousingInventoryPrediction_AnnualCounties_ARIMA.ipynb file.

S2.2.1 Inputs

Users are required to specify which test they want to run (Tests A, B, or C) in the second cell of the python notebook.

The modeling requires the following packages:

1. Pandas (Pandas Development Team, 2020)
2. NumPy (Harris et al., 2020)
3. Scikit-learn (Buitinck et al., 2011)
4. Statsmodels (Seabold and Perktold, 2010)

Input data is available in the Data folder of the project

S2.2.2 Model Parameters

Values of 0, 1, and 2 were tested for p, d, and q parameters, resulting in eight ARIMA models tested for each input/output combination.

S2.2.3 Outputs

The code saves a csv with the $RMSE_r$, $E[|H|]$, and $s_{|H|}$ (see Eq. 3 and Eq. 4 in Sect. 5 of the publication) for each p, d, and q combination over all input and output combinations for each test. This csv is located at Outputs/HousingInventoryPrediction_AnnualCounties_ARIMA_Results.csv.

S2.3 Long-Short Term Memory (LSTM) Models

The Python script used to develop the linear trend models is available in the Models/HousingInventoryPrediction_AnnualCounties_LSTM.ipynb file.

S2.3.1 Inputs

Users are required to specify which test they want to run (Tests A, B, C, D, E, F, G, H, or I) in the second cell of the python notebook.

The modeling requires the following packages:

1. Pandas (Pandas Development Team, 2020)
2. NumPy (Harris et al., 2020)
3. Scikit-learn (Buitinck et al., 2011)
4. Tensorflow (Martín Abadi et al., 2015)

Input data is available in the Data folder of the project

S2.3.2 Model Parameters (Table S2)

Table S2 provides a list of hyperparameters used for all LSTM models.

Table S2. LSTM Hyperparameters

Hyperparameter Type	Hyperparameter
Training / testing split	80 / 20
Random state	314
Activation method	Relu
Loss metrics	Mean squared error
Optimizer	Adam
Learning rate	1e-3
Decay rate	1e-6
Batch size	64
Dropout rate	0.2
Number of units in each LSTM layer	384
Number of units in final dense layer	32
Number of LSTM layers	3
Number of dense layers	1

S2.3.3 Outputs

The code saves a csv with the $RMSE_r$, $E[|H|]$, and $s_{|H|}$ (see Eq. 3 and Eq. 4 in Sect. 5 of the publication), along with other error metrics and hyperparameters, over all input and output combinations for each test. This csv is located at Outputs/HousingInventoryPrediction_AnnualCounties_LSTM_Results.csv. The best-performing LSTM models for Tests A, B, C, D, E, and F are also saved in the Outputs/LSTM_Models folder as an HDF5 file (note that not all models are currently saved in this folder to save storage space; only the best models for each test is provided in the project. Contact the corresponding author if interested in the additional HDF5 files).

S2.4 Model References

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G.: Scikit-learn: Machine Learning in Python, JMLR, 12, 2825–2830, 2011.

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C.,

and Oliphant, T. E.: Array programming with NumPy, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/, 2015.

Pandas Development Team: pandas-dev/pandas: Pandas, Zenodo, https://doi.org/10.5281/zenodo.3509134, 2020.

Seabold, S. and Perktold, J.: statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference, 2010.

Williams, C. and Davidson, R.: Regional county-level housing inventory predictions and the effects on hurricane risk using long-short term memory (LSTM) methods and applied to the southeastern United States (US), DesignSafe-CI, 2022.

S3 Analysis

All files used for the analysis of the work are available in the Analysis folder within Project PRJ-3303 of the DesignSafe-CI Data Depot (Williams and Davidson, 2022).

S3.1 Best Test Result

The Python script used to determine the input/output combination with the lowest errors for each test is available in the Analysis/HousingInventoryPrediction_AnnualCounties_BestTest.ipynb file. This script finds the test with the lowest multiplied value of $RMSE_r$, $E[|H|]$, and $s_{|H|}$ (see Eq. 3 and Eq. 4 in Sect. 5 of the publication) and publishes it in the resulting csv file saved here: Analysis/HousingInventoryPrediction_AnnualCounties_CombinedBest_Results.csv

S3.2 Figures

The Python script used to visualize the modeling errors and predictions is available in the Analysis/HousingInventoryPrediction_AnnualCounties_Figures.ipynb file.
This script first provides a map of the study area, available as Figure 1. This script then produces the modeling errors for the best univariate models in Tests A, B, and C, used to create Figure 3. Modeling errors and predictions are then computed for the recommended REACH20 model and saved as Outputs/HousingInventoryPrediction_AnnualCounties_LSTM_REACH20_Errors.csv and Outputs/HousingInventoryPrediction_AnnualCounties_LSTM_REACH20_Predictions respectively. Figures 4 and 5 use the REACH20 model error data, while Figs. 6 through 10 use the REACH20 prediction data. The code concludes with an application of Hurricane Harvey loss data to the projected housing rates, utilizing data from the Texas Department of Insurance (2019) [68], to produce Figs. 11 and 12. All figures are saved in the Analysis/Figures folder. Note that the final labeling of most figures used in the supporting publication was performed manually.