Codebook

This codebook contains the definition of variables used in Steven Liao. "Chinese Foreign Real Estate Investment and Local Voting in U.S. Presidential Elections." *International Studies Quarterly*.

Datasets

- `data-merge-county.RData`: A county-year level dataset used in the main analyses. The dataset contains 99 variables and 12,428 observations.
  - `year`: Year
  - `anticorrupt`: Year $\geq$ 2013? 1 = Yes, 0 = No
  - `placebo_anticorrupt_2012`: Year $\geq$ 2012? 1 = Yes, 0 = No
  - `county_fips`: County FIPS code
  - `county_name`: County name
  - `state_abb`: State abbreviation
  - `state_name`: State name
  - `aland_sqmi_county`: County land area (square miles)
  - `dem_share`: Two-party Democratic vote share
  - `china_undergraduate_county`: Total weighted number of Chinese international undergraduate students
  - `china_undergraduate_county_2012`: Total weighted number of Chinese international undergraduate students in 2012
  - `china_undergraduate_county_sqmi`: Total weighted number of Chinese international undergraduate students per square mile
  - `china_undergraduate_county_sqmi_2012`: Total weighted number of Chinese international undergraduate students per square mile in 2012
  - `cn_ug_sqmi_cat_2012_bi`: Dichotomous version of `china_undergraduate_county_sqmi_2012` with values $<$ the median = "CN-UG-low" and "CN-UG-high" otherwise
  - `cn_ug_sqmi_cat_2012_bi_low`: `cn_ug_sqmi_cat_2012_bi` = "CN-UG-low"? 1 = Yes, 0 = No
  - `cn_ug_sqmi_cat_2012_bi_high`: `cn_ug_sqmi_cat_2012_bi` = "CN-UG-high"? 1 = Yes, 0 = No
  - `cn_ug_sqmi_cat_2012_tri`: Trichotomous version of `china_undergraduate_county_sqmi_2012` based on terciles ("CN-UG-low", "CN-UG-medium, "CN-UG-high")
  - `cn_ug_sqmi_cat_2012_tri_low`: `cn_ug_sqmi_cat_2012_tri` = "CN-UG-low"? 1 = Yes, 0 = No

- cn_ug_sqmi_cat_2012_tri_medium: cn_ug_sqmi_cat_2012_tri = "CN-UG-medium"? 1 = Yes, 0 = No
- cn_ug_sqmi_cat_2012_tri_high: cn_ug_sqmi_cat_2012_tri = "CN-UG-high"? 1 = Yes, 0 = No
- ln_china_undergraduate_sqmi: log (china_undergraduate_county_sqmi + 1)
- ln_cn_ug_sqmi_2012: log(china_undergraduate_county_sqmi_2012 + 1)
- china_graduate_county: Total weighted number of Chinese international graduate students
- china_graduate_county_2012: Total weighted number of Chinese international graduate students in 2012
- china_graduate_county_sqmi: Total weighted number of Chinese international graduate students per square mile
- china_graduate_county_sqmi_2012: Total weighted number of Chinese international graduate students per square mile in 2012
- cn_g_sqmi_cat_2012_bi: Dichotomous version of china_graduate_county_sqmi_2012 with values $<$ the median = "CN-G-low" and "CN-G-high" otherwise
- cn_g_sqmi_cat_2012_bi_low: cn_g_sqmi_cat_2012_bi = "CN-G-low"? 1 = Yes, 0 = No
- cn_g_sqmi_cat_2012_bi_high: cn_g_sqmi_cat_2012_bi = "CN-G-high"? 1 = Yes, 0 = No
- cn_g_sqmi_cat_2012_tri: Trichotomous version of china_graduate_county_sqmi_2012 based on terciles ("CN-G-low", "CN-G-medium", "CN-G-high")
- cn_g_sqmi_cat_2012_tri_low: cn_g_sqmi_cat_2012_tri = "CN-G-low"? 1 = Yes, 0 = No
- cn_g_sqmi_cat_2012_tri_medium: cn_g_sqmi_cat_2012_tri = "CN-G-medium"? 1 = Yes, 0 = No
- cn_g_sqmi_cat_2012_tri_high: cn_g_sqmi_cat_2012_tri = "CN-G-high"? 1 = Yes, 0 = No
- ln_china_graduate_sqmi: log (china_graduate_county_sqmi + 1)
- ln_cn_g_sqmi_2012: log(china_graduate_county_sqmi_2012 + 1)
- india_undergraduate_county: Total weighted number of Indian international undergraduate students
- india_undergraduate_county_2012: Total weighted number of Indian international undergraduate students in 2012
- india_undergraduate_county_sqmi: Total weighted number of Indian international undergraduate students per square mile

- india_undergraduate_county_sqmi_2012: Total weighted number of Indian international undergraduate students per square mile in 2012
- ind_ug_sqmi_cat_2012_bi: Dichotomous version of india_undergraduate_county_sqmi_2012 with values $<$ the median = "IND-UG-low" and "IND-UG-high" otherwise
- ind_ug_sqmi_cat_2012_bi_low: ind_ug_sqmi_cat_2012_bi = "IND-UG-low"? 1 = Yes, 0 = No
- ind_ug_sqmi_cat_2012_bi_high: ind_ug_sqmi_cat_2012_bi = "IND-UG-high'? 1 = Yes, 0 = No
- ind_ug_sqmi_cat_2012_tri: Trichotomous version of india_undergraduate_county_sqmi_2012 based on terciles ("IND-UG-low", "IND-UG-medium", "IND-UG-high")
- ind_ug_sqmi_cat_2012_tri_low: ind_ug_sqmi_cat_2012_tri = "IND-UG-low"? 1 = Yes, 0 = No
- ind_ug_sqmi_cat_2012_tri_medium: ind_ug_sqmi_cat_2012_tri = "IND-UG-medium"? 1 = Yes, 0 = No
- ind_ug_sqmi_cat_2012_tri_high: ind_ug_sqmi_cat_2012_tri = "IND-UG-high"? 1 = Yes, 0 = No
- ln_india_undergraduate_sqmi: log (india_undergraduate_county_sqmi + 1)
- ln_ind_ug_sqmi_2012: log(india_undergraduate_county_sqmi_2012 + 1)
- pop_county_tot: Total county population
- share_pop_county_female: The share of female population
- share_pop_under_5_county: The share of population under 5 years old
- share_pop_5_17_county: The share of population from 5 to 17 years old
- share_pop_18_24_county: The share of population from 18 to 24 years old
- share_pop_25_34_county: The share of population from 25 to 34 years old
- share_pop_35_44_county: The share of population from 35 to 44 years old
- share_pop_45_54_county: The share of population from 45 to 54 years old
- share_pop_55_64_county: The share of population from 55 to 64 years old
- share_pop_65_county: The share of population 65 years old and above
- share_pop_white_county: The share of White population
- share_pop_white_county_2012: The share of White population in 2012
- share_white_2012_cat: Trichotomous version of share_pop_white_county_2012 based on terciles ("low", "medium", "high")
- share_white_2012_cat_low: share_white_2012_cat = "low"? 1 = Yes, 0 = No

- share_white_2012_cat_medium: share_white_2012_cat = "medium"? 1 = Yes, 0 = No
- share_white_2012_cat_high: share_white_2012_cat = "high"? 1 = Yes, 0 = No
- share_pop_black_county: The share of Black population
- share_pop_hispanic_county: The share of Hispanic population
- share_pop_asian_county: The share of Asian population
- share_pop_cn_county: The share of Chinese population
- share_foreign_born_county: The share of foreign-born population
- share_enroll_county: The share of population enrolled in college or above
- share_edu_ba_above_county: The share of population with BA degree or above
- share_edu_ba_above_county_2012: The share of population with BA degree or above in 2012
- share_edu_ba_above_county_2012_tri: Trichotomous version of share_edu_ba_above_county_2012 based on terciles ("low", "medium", "high")
- share_edu_ba_above_county_2012_tri_low: share_edu_ba_above_county_2012 = "low"? 1 = Yes, 0 = No
- share_edu_ba_above_county_2012_tri_medium: share_edu_ba_above_county_2012 = "medium"? 1 = Yes, 0 = No
- share_edu_ba_above_county_2012_tri_high: share_edu_ba_above_county_2012 = "high"? 1 = Yes, 0 = No
- ln_pop_density_county: Population density (log)
- effective_tax_county: Effective tax rate (%)
- ipw_county: Trade Exposure (Imports Per Worker)
- gdp_growth: GDP growth rate (%)
- gdp_growth_2016: GDP growth rate (%) in 2016
- gdp_growth_2020: GDP growth rate (%) in 2020
- employ_rate_county: Employment rate
- employ_rate_county_2012: Employment rate in 2012
- employ_rate_county_2016: Employment rate in 2016
- employ_rate_county_2020: Employment rate in 2020
- median_household_income_county: Median household income ($10,000)
- vacancy_county: Share of vacant houses
- vacancy_county_2012: Share of vacant houses in 2012
- vacancy_county_2012_tri: Trichotomous version of vacancy_county_2012 based on terciles ("low", "medium", "high")

  - `vacancy_county_2012_tri_low`: `vacancy_county_2012_tri` = "low"? 1 = Yes, 0 = No
  - `vacancy_county_2012_tri_medium`: `vacancy_county_2012_tri` = "medium"? 1 = Yes, 0 = No
  - `vacancy_county_2012_tri_high`: `vacancy_county_2012_tri` = "high"? 1 = Yes, 0 = No
  - `share_homeowner_county`: Share of homeowners
  - `share_homeowner_county_2012`: Share of homeowners in 2012
  - `share_homeowner_county_2012_cat`: Trichotomous version of `share_homeowner_county_2012` based on terciles ("low", "medium", "high")
  - `share_homeowner_county_2012_cat_low`: `share_homeowner_county_2012_cat` = "low"? 1 = Yes, 0 = No
  - `share_homeowner_county_2012_cat_medium`: `share_homeowner_county_2012_cat` = "medium"? 1 = Yes, 0 = No
  - `share_homeowner_county_2012_cat_high`: `share_homeowner_county_2012_cat` = "high"? 1 = Yes, 0 = No

- `foia-county-cn-under-2012.RData`: A county level dataset of Chinese international undergraduate students in 2012 used in the analysis "Figure-1.R". The dataset contains 3 variables and 3,220 observations.
  - `county_fips`: County FIPS code
  - `year`: Year
  - `china_undergraduate_county_sqmi_2012`: Total weighted number of Chinese international undergraduate students per square mile in 2012

- `nar-df.RData`: A country-year level dataset used "Figure-A1.R". The dataset contains 5 variables and 55 observations.
  - `cty`: Country
  - `year`: Year
  - `sale`: Annual sales to foreign buyers from the country
  - `sale_total`: Annual sales to all foreign buyers
  - `share`: The country's share of sales among all foreign buyers in the year

- `foia-zcta.RData`: A ZCTA-country-year level dataset used in "Figure-A2.R" and "Figure-A5.R". The dataset contains 7 variables and 3,619,827 observations.
  - `cty`: Student origin country
  - `iso3n`: Student origin country ISO numeric code
  - `zipcode`: ZIP code
  - `zcta`: ZIP Code Tabulation Area code

  - year: Year
  - secondary: Total number of secondary-level students enrolled in institutions located in ZCTA based on FOIA data
  - undergraduate: Total number of undergraduate students enrolled in institutions located in ZCTA based on FOIA data
  - graduate: Total number of graduate students enrolled in institutions located in ZCTA based on FOIA data

- pres.RData: A county-year level dataset used in "Figure-A3.R". The dataset contains 5 variables and 12,437 observations.
  - fips: County FIPS code
  - year: Year
  - state_abb: State abbreviation
  - dem_share: Two-Party Democratic Vote Share
  - rep_share: Two-Party Republican Vote Share

- foia-zcta-cn-under-2012.RData: A ZCTA-level dataset of Chinese international undergraduate students in 2012 used in "Figure-A4.R". The dataset contains 5 variables and 33,112 observations.
  - ZCTA: ZIP Code Tabulation Area code
  - student_level: Student Program Level
  - year: Year
  - wt_total: Total weighted number of Chinese international students in ZCTA constructed from the FOIA data
  - student_origin: Student origin country

- cn-frei-dest.csv: A state-year level dataset used in "Figure-A5.R". The dataset contains 3 variables and 65 observations.
  - year: Year
  - state: State
  - percent: State shares of Chinese FREI (%) from the National Association of Realtors

- data-merge-zcta.RData: A ZCTA-year level dataset used in "Figure-B1.R". The dataset contains 52 variables and 326,646 observations.
  - zcta: ZIP Code Tabulation Area code
  - year: Year
  - anticorrupt: Year $\geq$ 2013? 1 = Yes, 0 = No
  - placebo_anticorrupt_2012: Year $\geq$ 2012? 1 = Yes, 0 = No

- `county_fips`: County FIPS code
- `county_name`: County name
- `state_name`: State name
- `zhvi_mean`: Mean Zillow Home Value Index in ZCTA
- `ln_zhvi_mean`: log(`zhvi_mean`)
- `china_undergraduate`: Total weighted number of Chinese international undergraduate students in ZCTA
- `china_undergraduate_2012`: Total weighted number of Chinese international undergraduate students in ZCTA in 2012
- `china_undergraduate_sqmi`: Total weighted number of Chinese international undergraduate students per square mile in ZCTA
- `china_undergraduate_sqmi_2012`: Total weighted number of Chinese international undergraduate students per square mile in ZCTA in 2012
- `cn_ug_sqmi_bi_2012`: Dichotomous version of `china_undergraduate_sqmi_2012` with values $<$ the median = "CN-UG-low" and "CN-UG-high" otherwise
- `cn_ug_sqmi_bi_2012_low`: `cn_ug_sqmi_bi_2012` = "CN-UG-low"? 1 = Yes, 0 = No
- `cn_ug_sqmi_bi_2012_high`: `cn_ug_sqmi_bi_2012` = "CN-UG-high'? 1 = Yes, 0 = No
- `china_graduate`: Total weighted number of Chinese international graduate students in ZCTA
- `china_graduate_2012`: Total weighted number of Chinese international graduate students in ZCTA in 2012
- `china_graduate_sqmi`: Total weighted number of Chinese international graduate students per square mile in ZCTA
- `china_graduate_sqmi_2012`: Total weighted number of Chinese international graduate students per square mile in ZCTA in 2012
- `cn_g_sqmi_bi_2012`: Dichotomous version of `china_graduate_sqmi_2012` with values $<$ the median = "CN-G-low" and "CN-G-high" otherwise
- `cn_g_sqmi_bi_2012_low`: `cn_g_sqmi_bi_2012` = "CN-G-low"? 1 = Yes, 0 = No
- `cn_g_sqmi_bi_2012_high`: `cn_g_sqmi_bi_2012` = "CN-G-high'? 1 = Yes, 0 = No
- `india_undergraduate`: Total weighted number of Indian international undergraduate students in ZCTA
- `india_undergraduate_2012`: Total weighted number of Indian international undergraduate students in ZCTA in 2012

- `india_undergraduate_sqmi`: Total weighted number of Indian international undergraduate students per square mile in ZCTA
- `india_undergraduate_sqmi_2012`: Total weighted number of Indian international undergraduate students per square mile in ZCTA in 2012
- `ind_ug_sqmi_bi_2012`: Dichotomous version of `india_undergraduate_sqmi_2012` with values $<$ the median = "IND-UG-low" and "IND-UG-high" otherwise
- `ind_ug_sqmi_bi_2012_low`: `ind_ug_sqmi_bi_2012` = "IND-UG-low"? 1 = Yes, 0 = No
- `ind_ug_sqmi_bi_2012_high`: `ind_ug_sqmi_bi_2012` = "IND-UG-high'? 1 = Yes, 0 = No
- `share_pop_zcta_female`: The share of female population
- `share_pop_5_17_zcta`: The share of population from 5 to 17 years old
- `share_pop_18_24_zcta`: The share of population from 18 to 24 years old
- `share_pop_25_34_zcta`: The share of population from 25 to 34 years old
- `share_pop_35_44_zcta`: The share of population from 35 to 44 years old
- `share_pop_45_54_zcta`: The share of population from 45 to 54 years old
- `share_pop_55_64_zcta`: The share of population from 55 to 64 years old
- `share_pop_65_zcta`: The share of population 65 years old and above
- `share_pop_white_zcta`: The share of White population
- `share_pop_black_zcta`: The share of Black population
- `share_pop_hispanic_zcta`: The share of Hispanic population
- `share_pop_asian_zcta`: The share of Asian population
- `share_pop_cn_zcta`: The share of Chinese population
- `share_foreign_born_zcta`: The share of foreign-born population
- `share_edu_ba_above_zcta`: The share of population with BA degree or above
- `share_enroll_zcta`: The share of population enrolled in college or above
- `ln_pop_density_zcta`: Population density (log)
- `effective_tax_zcta`: Effective tax rate (%)
- `ipw_county`: Trade Exposure (Imports Per Worker)
- `employ_rate_zcta`: Employment rate
- `median_household_income_zcta`: Median household income ($10,000)
- `vacancy_zcta`: Share of vacant houses