

Your Data Building Blocks for Accurately Researching Populations in Today's Markets

Since 2001, STI: PopStats™ quarterly population estimates have revolutionized the market research industry's conventional view of population estimates by achieving unmatched levels of accuracy by:

- Uniquely leveraging U.S. Postal Service's ZIP + 4® records
- Creating an innovative "bottom-up" methodology
- Updating estimates quarterly versus annually
- Adding new data variables in response to clients' specific requests
- Providing a "building blocks" philosophy that unleashes unlimited research potential
- Extending a responsive customer service approach for questions and enhancements

With the launch of PopStats data, for the first time, companies viewed population estimates as a highly accurate, dependable, and essential component of market research, instead of data that added only marginal value to their research. The data has become even more valuable with the addition of over 1,300 population and demographic variables, including neighborhood segmentation and workplace estimates.

No matter what research goals are driving a company forward — from locating new high-growth areas before competitors, to consolidating store networks, to accelerating growth plans — PopStats data brings immense value to critical business decisions. What's more, companies also gain confidence in their research, knowing they have the most accurate and current insight — including market growth and decline, demographic variations, seasonal population fluctuations, mortgage risk differences, income changes, and much more.

Today market researchers in a wide range of industries — including retail, healthcare, real estate development, telecommunications, and economic development — rely on PopStats data to gain in-depth and dependable insight on where people live and work across the U.S. With PopStats data fueling their data engines, companies are enjoying the confidence to make more informed business decisions regarding markets, locations, and consumers.



What Would You Like To Build Today?

PopStats Data's Innovation Overcomes Traditional Population Data Challenges

PopStats™ data's unique approach to population estimates includes four primary innovations over traditional population data methodologies.

1. ZIP + 4 and 2020 Census vs. Census-Only Source Data. STI was the first data provider to envision a way to leverage ZIP + 4 postal data to improve accuracy and timeliness. ZIP + 4 targets areas as small as a specific group of houses — typically four to 12 — or a building. It is possible to literally see structures come online as they are built and occupied. The ZIP + 4 level leads to more accurate population estimates for five reasons: (1) it is extremely detailed, (2) it contains over 28 million records, (3) it includes all major population centers, (4) it can be manipulated statistically, and (5) it is easily consolidated into any geography.

2. Bottom-Up vs. Top-Down Methodology. The decennial U.S. Census is based on a traditional top-down construction, which takes macro-level data and extrapolates it down to a micro-level: from U.S., to state, to county, to tract, to block-group. The Census Bureau's national-to-local direction was copied by demographers to generate population estimates at the block-group level. However, there are significant problems with this direction. Foremost, macro-level data is unsuitable for use at a micro-level, like block groups, which are greatly influenced by singular local events, such as a new apartment complex or a building demolition. To compensate, demographers developed population-spreading techniques that broad-stroke areas of growth and decline at the sub-county level. This is an improvement, but still retains limitations. Namely, it can mask block-group level growth or decline. PopStats data delivers a more accurate population count on a micro level by starting at the ZIP + 4 level, then moving up ("bottom-up") to the block group, tract, county, then state levels.

3. Quarterly vs. Annual Updates. Traditional population data is updated only once every 12 months (typically in May or June). As a result, the data is chronically out-of-date. This puts researchers at a serious disadvantage. STI created the industry's first population data to be updated on a quarterly basis — every January, April, July, and October. Today we provide the industry's leading quarterly updated population data.

4. Expanding vs. Static Variables. Unlike many population data providers, STI continually expands into new data territory. PopStats data launched in 2001 with 21 variables and by 2010 had over 1,200. For example some added variables include mortgage risk, home values, employment, five- and ten-year forecasts, and much more. New variables are based on a combination of STI innovation and client data requests. For example, a leading grocery store chain requested seasonal data, a QSR (quick serve restaurant) client requested transient (i.e., hotel, motel, and RV park) population counts, and a drug store chain requested Puerto Rico population counts. Every new data variable is available to all PopStats clients.

PopStats Data's Revolutionary Population Estimating Methodology

The PopStats™ model is a collection of models that calculate the quarterly population estimates. The methodology consists of the following three steps.

STEP 1 — Estimate Households. STI's research has shown that a unique and quantifiable relationship exists between USPS (United States Postal Service) data and U.S. Census Bureau household counts. Due to this relationship, STI can model population shifts quickly and accurately using a proprietary technique leveraging the correlation between the two. The process is initiated by base-lining the ZIP + 4 data and its associated statistics as they existed in April 2020. Then, as new ZIP + 4 data is provided (new data and statistics are delivered monthly) we can model and derive a USPS growth factor for every ZIP + 4 in the country. This application occurs via our proprietary model that uses this information as well as other pertinent factors to generate a current estimate. To limit bias in the data due to extraneous figures, such as errors in the raw data, PopStats methodology includes automated processes for overcoming any and all anomalies, including ZIP + 4 inaccuracies, data smoothing issues, conversions (lofts), and overrides.

STEP 2 — Estimate Household Populations. A variety of U.S. Census Bureau and private studies have shown that the relationship of persons-to-households remains relatively stable over time. STI takes the Census 2020 persons-per-household-per-block group figures, and adjusts the ratio to reflect any changes in the county estimated persons-per-household generated by the U.S. Census Bureau. These new figures are then applied to the estimated households to derive an estimated household population.

STEP 3 — Apply Controls. To further ensure accuracy and limit bias in our estimates, STI uses a series of checks-and-balances to validate the results. One of these steps is to compare our estimates to the U.S. Census Bureau's annual population estimates released every Spring. If any major discrepancies occur between the two numbers, the model applies a set of heuristics to determine the most probable population figure. We also consult with multiple state and federal agencies whose data is independently gathered and calculated. In addition, selected cities throughout the U.S. are field-surveyed to further validate our model's results.

Methodology for Key PopStats Data Variables "Break Outs"

Once the base population has been estimated, the PopStats model "breaks out" several demographic estimates, such as age and sex, race and ethnicity, group quarters, incomes, and housing values. Many more data variables are available in the ever-expanding PopStats product.

Age and Gender. Age and gender are determined through a traditional cohort survival analysis. This sub-model to the main model looks at each age distribution within a race category and applies the appropriate birth and survival rates as determined by the NCHS (National Center for Health Statistics). These results are then balanced back to the base population using an iterative approach. In addition, information from the NCES (National Center for Education Statistics) is applied to validate the age distribution of school-age children. U.S. Census estimates are used to validate all other age ranges.

Race (Ethnicity). Race is calculated using a ratio analysis of April 2020 observed and annual U.S. Census estimates. In areas of high growth we use race information gathered by the FFIEC (Federal Financial Institutions Examination Council, which collects information from financial institutions concerning loans and race issues. It is a reasonable source for understanding race percentages in high-growth areas. As a final check for race, our model also consults with NCES race data for elementary school children and checks NCES data against our figures.

Income Estimates. Income estimates are based on a two-step process. First, household incomes at the county level are estimated using a blend of information from the IRS's Survey of Income, U.S. Census Bureau's March CPS (Current Population Survey) income estimates, and personal income estimates from the BEA (Bureau of Economic Analysis). Once the county estimate is derived, we estimate the block-group-level. This is done in two parts. First, we separate existing households from new-growth households, because our research has found that in high-growth areas existing households are not a good indicator for determining the income of the new households entering the area. Therefore, we use a typical income growth approach that resembles the growth of county income. Then we add to that a separate income growth for new households modeled on the FFIEC's mortgage data transactions.

Group Quarters. Group quarters are a collection of unrelated people where no one individual can claim "head of household," such as college students and military personnel. Generally speaking, group quarters data can be divided into three categories: colleges, military bases, and institutions (i.e., state homes, hospitals, and prisons). We estimate each category individually, then combine them for a total estimate. College student dormitory information is derived from the NCES annual college survey. Military group quarters are determined based on a direct data feed received from the DOD (Department of Defense) Manpower Data Center. Institutionalized persons are estimated using historical trends from the U.S. Census.

Housing Values. Housing Values are determined in a fashion similar to income estimates. Housing and associated values that existed as of April 2020 are updated using data from the OFHEO (Office of Federal Housing Enterprise Oversight). Our model performs a detailed analysis of same-home selling prices that occur over time. We use the resulting growth factors and apply them to existing April 2020 owner-occupied homes. New home values (homes built after April 2020) are determined by ratio analysis of the FFIEC's mortgage values and actual selling prices.

Data Sources for STI: PopStats:

United States Census Bureau
United States Postal Service (USPS)
United States Department of Defense (DMDC)
National Center for Education Statistics (NCES)
National Center for Health Statistics (NCHS)
Federal Financial Institutions Examination Council (FFIEC)
Internal Revenue Service (IRS)
Bureau of Economic Analysis (BEA)
Bureau of Labor Statistics (BLS)
Office of Federal Housing Enterprise Oversight (OFHEO)
Department of Defense (DOD)