2023 CITY SURVEY DETAILED METHODOLOGY

You can find additional results and information about this and prior City Surveys at sf.gov/citysurvey.

San Francisco City Survey

Every two years, the City and County of San Francisco's Controller's Office conducts the City Survey to assess residents' perceptions of, use of, and satisfaction with City services. The survey's goals are also to gauge attitudes around issues facing the City. The 17th City Survey was conducted in 2019. The 18th survey cycle was delayed due to the COVID-19 pandemic; contractors InterEthnica and EMC Research carried out data collection for the 18th City Survey from October through December of 2022.

Residents answered questions about how frequently they use city services and infrastructure like the library, Muni, public safety, and street cleanliness, and rated their satisfaction with those service areas. This year residents also responded to questions about the City's response to the COVID-19 public health emergency.

The City works to maintain continuity across time in the survey but has allowed methodology and content to shift over the years where necessary to maintain relevance and capture the most representative sample of City residents. The following document details the methodology of the current City Survey, how results are calculated and reported, and how the City verifies these findings through statistical analysis.

2023 Methodology and Survey Implementation

Data Collection Strategy

From 1996 to 2013 the City Survey was administered by mail. In 2017 and 2019 the City updated the methodology to a telephone survey using a random digit dialing (RDD) methodology, plus a very small online component. Over the recent years, RDD and other traditional, telephone-based data collection methods have met new challenges, such as declining telephone response rates, and increased inaccuracy of random dialing. To offset these challenges and to generate results more representative of San Francisco residents, the 2023 survey adopted a two-phase data collection strategy. The first phase collected data from a representative sample of adult San Francisco residents through a combination of telephone and online interviewing. The contact information used for reaching out to the representative sample came from a variety of sources, including consumer data and public data sources like assessor data and voter registration records.

The second phase of the research supplemented the data collected through the representative survey with intercept surveying, conducted in person at selected locations, to ensure that demographics that are traditionally harder to reach via other methods were adequately represented in the survey. Specifically, this methodology was designed to include more limited-English proficient (LEP) respondents. Intercept locations were chosen based on the likelihood of the presence of priority populations.

The City Survey collected 2,530 responses; 2,000 of the completed surveys were collected through the phone and online modes, and 530 through intercept interviewing.

Data Collection

To ensure the City Survey questionnaire was understandable to respondents and would result in high-quality data, the contractors conducted a data collection pre-test in mid-October 2022. A total of 30 pretest surveys were conducted; ten in each mode (intercept, telephone, and online).

Survey responses were collected across three modes: telephone, web, and intercept. Surveys were conducted in English, Spanish, Chinese and Filipino; with Vietnamese and Russian versions available for intercept interviewing and self-administration online. The first phase of data collection, which included the telephone and online modes, was completed within a two-week period, from late-October to early-November. InterEthnica and EMC reviewed the demographics from this dataset and adjusted the intercept targets based on this assessment to target specific populations. Demographic targets for intercept interviewing informed the selection of interview locations and times. These intercept surveys were conducted from late-October to mid-December.

Telephone Mode Data Collection: Trained professional interviewers contacted participants via landline and cell phone and asked them to participate in the survey. Interviewers manually dialed identified cell phone numbers as required by law. Interviewers were specifically trained on the survey content, including a review of appropriate prompts, challenging pronunciations, and other administration details. Interviewers administered telephone surveys in English, Spanish, Chinese, and Filipino, with bilingual interviewers used as much as possible to dial targeted neighborhoods to boost response rates among non-English speakers. Respondents completed 800 telephone surveys over a two-week data collection period.

Online Mode Data Collection: The survey implementation firms used targeted text message and email invitations to drive participants to the online survey. Access was limited only to those who were invited to participate by using unique links for each invited participant, and each participant could only complete the survey one time. The online survey was available in all survey languages, and participants were able to choose their preferred language. Respondents completed 1,200 online surveys: 600 via text invites and 600 via email invites.

Intercept Mode Data Collection: City Survey also used intercept interviews to target hard to reach populations, where interviewers engaged respondents in person and used paper survey questionnaires to record responses. The online/phone interviews began prior to intercepts and the City and contractors used periodic data check-ins to identify which audiences had not been reached in sufficient numbers by phone and email. This data helped identify the audience for the intercept surveys and informed where intercept surveying took place. Intercept interviewers were bilingual and specifically trained on the survey content, including a review of appropriate prompts, challenging pronunciations, and other administration details. Further, surveyors were trained in the following: neutral data gathering, cultural humility, COVID-19 safety, public engagement, and outreach safety. Interviewers conducted intercepts in spaces that were most likely to reach a diversity of users within neighborhoods that represented that target audience as much as possible. Respondents completed 530 intercept surveys with a specific target on residents who are native speakers of Chinese (any dialect), Spanish, and Tagalog/Filipino as well as residents who identify as Black or African American.

The order of questions was adapted by mode. Intercept surveys presented the survey questions in the same order on every survey, while online and telephone modes randomized certain survey questions to mitigate potential order effects. Sections A through F of the survey (covering resident attitudes around

utilities, street conditions, parks and recreation, library system, transportation, and public safety) were randomized. The overall City rating remained at the beginning of each survey, and sections pertaining to most important problem in the city, COVID-19 recovery, and demographics were not randomized and remained at the end.

After all data collection was completed, all intercept survey data was entered into the digital survey dataset. Surveyors performed quality control checks against the original paper surveys to ensure accuracy. Upon the completion of data entry and quality checks, data from the telephone, online and intercept modes were merged to form a single dataset for weighting and analysis purposes.

Representativeness and Weighting Approach

The target number of completed interviews from specific demographic subgroups were determined based on the US Census Bureau's ACS (American Community Survey) 2020 5-year estimates. Per the Census Bureau website, "The 5-year estimates from the ACS are 'period' estimates that represent data collected over a period of time. The primary advantage of using multiyear estimates is the increased statistical reliability of the data for less populated areas and small population subgroups."

The updated methodology combining various data collection modes for the 18th City Survey was designed to reach a broader cross-section of San Francisco residents than earlier methodologies. More representative surveying helps to minimize the need for drastic weighting on the final survey data. This helps mitigate risk to data accuracy associated with significant weighting of smaller subgroups.

The survey implementation firms reviewed demographic data daily during the telephone and online data collection window and adjusted targets as needed before telephone dialing / online survey inviting began each day. Upon the completion of data collection through these modes, InterEthnica and EMC Research reviewed the demographics of this dataset and adjusted the intercept targets to ensure enough completed interviews were collected from harder-to-reach populations, where the representative survey did not deliver an adequate number of responses.

To closely approximate the demographic makeup of San Francisco residents with greater accuracy, the final survey data was weighted to the ACS 2020 5-year estimates on age, gender, and ethnicity. Self-identified versions of these variables were used for weighting purposes, and divergences from ACS targets are due to missing data with some respondents refusing to respond certain demographic questions.

Nested weighting targets were generated automatically to ensure gender and ethnic groups reflected accurate age breakdowns. An iterative raking process was used to determine weights to achieve nested weighting targets. The combination of these strategies for ensuring representativeness have helped to keep the weighting scheme simple and ensure that no one respondent carried too significant a weight value compared with others in the dataset.

City Survey Main Outcomes and Calculations

Most outcomes in the City Survey are ratings that use a five-point grading scale (A+ equals five points and F equals one point). The City translates responses to the point scale and averages responses to create a mean score. The full survey instrument from the 2023 City Survey is posted online and can be found through the City Survey landing page sf.gov/citysurvey.

The three main ways the City reports the results are:

- a weighted mean from this scale,
- the corresponding letter grade, or
- the weighted percent of respondents who graded a service an A or B.

Weights are used in the reporting to control for any differences between the demographics of the respondent sample and the overall population of San Francisco.

Results in this report for previous years may not match the exact number from earlier publications. For each survey publication, the City modifies where and how weights are used in calculating outcomes depending on what is most appropriate for the most recent data. This year the City increased the use of weights across the reporting in response to more nuanced weight calculations described above. In addition, the exact calculation for reporting the percent of respondents grading a service an A or B has been slightly adjusted this year. Finally, while most reported service ratings are now based on a single, overarching question, some are based on the average of multiple rating questions. The exact questions used to calculate ratings have changed over time, though the City has worked to keep year over year comparisons are similar as possible and notes where any changes could impact a rating or interpretation.

Calculating Ratings

There are two ways the City Survey asks residents to rate service areas: by rating subcomponents of a service or by providing an overall rating for the service. Historically, most service areas had multiple subcomponents, such as cleanliness or employee courtesy. When only the subcomponent questions appeared on the survey, the overall rating is a calculated average of the subcomponents. The City has added a single overarching question asking about the respondents' perception of that service to most services. Once the survey includes a question that directly asks residents to provide an overall rating, the response to that question is reported as the overall rating. Most service area ratings have been using these overarching questions since 2017.

- Library ratings were an average of subcomponent ratings through 2015. 2017 on uses the overall rating.
- Parks ratings were an average of subcomponent ratings through 2009. 2011 on uses the overall rating.
- Muni ratings were an average of subcomponent ratings through 2015. 2017 on uses the overall rating.
- Street and Sidewalks average rating in 2023 is an average of street and sidewalk cleanliness, street condition, and sidewalk and curb ramp condition.
  - For 2011-2017, the average includes ratings for street lighting.

- In prior years, the Streets and Sidewalks rating was combined with the Utilities rating for an overall infrastructure rating, but Streets and Sidewalks ratings are now separate from Utilities ratings and the data reported has been separated.
  - Utilities ratings are an average of the water rating and sewer rating. CleanPowerSF and Hetch Hetchy power are included in the overall rating for 2023.
  - The Safety rating is an average of each resident's grade of their feeling of safety in their neighborhood during the day and at night.
  - The Police ratings are new in 2023 and are an average of two questions asking residents to rate the quality of Police services in their neighborhood and their trust in San Francisco police officers.

Because all shifts from subcomponent calculated ratings to overall ratings happened several survey iterations ago, the City does not always note on data visualizations where that change happened. If that shift caused a major change in the rating, that is noted and reported.

Weighted means

The City uses weighted averages to report the overall grade for each service. The survey implementation firm calculated the weight for each respondent (detailed above). The City uses that weight to calculate a weighted mean ratings, and uses the methodology detailed in the table below to translate the rating into a letter grade.

Numeric to Letter Grades		
Letter Grade	Lower Mean	Upper Mean
A+	5.00	5.00
A	4.67	4.99
A-	4.33	4.66
B+	4.00	4.32
B+	3.67	3.99
B-	3.33	3.66
C+	3.00	3.32
C	2.67	2.99
C-	2.33	2.66
D+	2.00	2.32
D	1.67	1.99
D-	1.33	1.66
F	1.00	1.32

Where ratings are based off of two or more subcomponent questions, the City first averages the score for each respondent, using equal weights for each question, then calculates the weighted mean for the sample. This follows historical precedent from prior City Surveys and is used both for service areas where the questions remain unchanged and for new ones such as Police ratings.

Weights are used both when calculating the rating of the entire sample or a subgroup (such as ratings by gender, income, or race).

Percent reporting A or B

In addition to weighted means and letter grades, the City also reports out on the percent of respondents who rate a service an A or a B. This provides a proxy for the proportion of respondents who are relatively satisfied with a service area. Weights are used in calculating the percent.

In 2023 the City changed the way this measure is calculated to more closely match how it believes it is likely to be interpreted. Historically, when the overall rating for a service area was the combined average of that service area's subcomponents, the average of the subcomponent ratings is used to calculate the proportion rating a service an A or a B. Functionally, this means that if a respondent rated one subcomponent an A but others a C, the grade could average to a B- or above and would therefore still be marked as a positive within this methodology. The exception has always been safety, where the calculation required all subcomponents to be either an A or a B for a positive rating. This year the City used this same methodology, where all subcomponents must be an A or a B to be counted in the rating, for all service areas. This includes recalculating the percent rating a service an A or a B for all past years, so that changes over time are all comparable.

This change impacts service areas where the rating is based on several subcomponent questions and may lower the percent of respondents who appear to rate that area positively when compared to prior year reports.

2023 Mode Changes and Statistical Checks

Impact of methodology changes on the interpretation of results

This year, the City updated the surveying methodology again to include online and in-person options. These changes took into account changes in how people communicate and prefer to answer questions, and allowed the City to reach a more representative sample of San Francisco residents, including younger respondents and non-white respondents who were sometimes underrepresented in previous surveys.

There are several mechanisms through which the mode changes in 2023 could impact results. Mode may impact the selection of residents that surveyors reach and impact which residents elect to respond to the survey. The mode itself may also influence residents' responses. While there may be unobservable differences in respondents that the City cannot determine the impact of, it is possible to control for observable demographic differences in the survey population and for the mode of response to verify whether ratings and responses in 2023 are comparable to prior years.

This year, 1,200 residents responded online, 800 by telephone, and 530 via in-person surveys on the street. Respondents over 65 were more likely to respond by telephone or in-person. White respondents were more likely to respond online, while Asian and Pacific Islander, and Hispanic or Latino respondents were less likely to have responded online.

Changing the methodology and sampling can also change how residents rate services. On average, respondents online rated services somewhat lower, while in-person responses are somewhat higher.

These changes improve the quality of the survey results and will ensure more accurate results in future City Surveys. They also make it harder to interpret changes in survey ratings between 2019 and 2023.

Finally, the respondent ethnicity question was asked differently in the 2023 survey. While previous surveys allowed participants to select multiple responses, the 2023 survey accepted a single response, thus collecting the primary ethnic identity of respondents rather than capturing intersectional identities. Because the ACS publishes ethnicity data as mutually exclusive categories, this approach did not lead to issues with regards to weighting. Additionally, intercept and telephone interviewers did not report respondents having difficulty answering the question due to only a single response being accepted. Around 8% of respondents refused to answer this question, which is a standard refusal rate for this question. However, it is important to note that the question was framed differently for this survey than in prior surveys. The City plans to revert to a multiple select question in future surveys in alignment with internal recommendations.

Statistical checks to verify survey results

With a sample size of 2,530, outcomes generally have a ±1.95 percentage point margin of error at a 95% confidence interval, calculated by the survey implementation firms. Prior to reporting changes over time or differences in rating by subgroup, the City runs a series of checks, both to confirm statistical significance generally, but also to check on whether that finding could be due to mode or respondents demographic changes rather than a real result. To do this the City runs a series of weighted linear regressions on the outcome in question, controlling for year or the subpopulation as relevant, for mode (both intercept and online), and for a set of main resident demographics. If the finding is no longer significant after these controls are added, it is generally removed as a finding in the report. If the significance decreases or is in question, it is noted explicitly in the findings.