



U.S. Census Bureau

American FactFinder

CONFIDENTIALITY

The Census Bureau has modified or suppressed some data on this site to protect confidentiality. [Title 13 United States Code](#), Section 9, prohibits the Census Bureau from publishing results in which an individual's or business' data can be identified.

The Census Bureau's internal Disclosure Review Board sets the confidentiality rules for all data releases. A checklist approach is used to ensure that all potential risks to the confidentiality of the data are considered and addressed. For more information on how the Census Bureau protects the confidentiality of data, please explore the following links.

- [Disclosure Avoidance Procedures](#)
- [Suppression](#)
- [Noise Infusion](#)
- [Data Swapping](#)
- [Protection of Microdata Files](#)
- [Synthetic Data](#)

Questions about confidentiality may be addressed to:

POL.Policy.Office@census.gov.

Title 13, United States Code: Title 13 of the United States Code authorizes the Census Bureau to conduct censuses and surveys. Section 9 of the same Title requires that any information collected from the public under the authority of Title 13 be maintained as confidential. Section 214 of Title 13 and Sections 3559 and 3571 of Title 18 of the United States Code provide for the imposition of penalties of up to five years in prison and up to \$250,000 in fines for wrongful disclosure of confidential census information.

Disclosure Avoidance: Disclosure avoidance is the process for protecting the confidentiality of data. A disclosure of data occurs when someone can use published statistical information to identify either an individual or business that has provided information under a pledge of confidentiality. Using disclosure avoidance procedures, the Census Bureau modifies or removes the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual or business, the Census Bureau has taken steps to disguise or suppress the original data while making sure the results are still useful. The techniques used by the Census Bureau to protect confidentiality in tabulations vary, depending on the type of data.

Suppression: Suppression is a method of disclosure avoidance used to protect individuals' confidentiality by not showing (suppressing) the cell values in tables of aggregate data for cases where only a few individuals or businesses are represented or dominate the cell value. The cells that are not shown are called primary suppressions. To make sure the primary suppressions cannot be closely estimated by subtracting the other cells in the table from the marginal totals, additional cells are also suppressed. These additional suppressed cells are called complementary or secondary suppressions. Values for cells that are not

suppressed remain unchanged. Before the Census Bureau releases data, computer programs check published tables for both primary and complementary disclosures. Suppression was used for the 1980 Census of Population and Housing and is now used for a number of economic surveys and censuses.

Example -- With Disclosure

Number of Firms and Receipts by Industry

Industry	Number of Firms	Receipts (\$1,000)
001	100	1,000
0011	99	990
0012	1	10*
002	200	10,000
0021	188	9,000
0022	12	1,000*

NOTE: * Indicates cells in which data may be identifiable.

Example -- Same Table Without Disclosure, Protected by Suppression

Number of Firms and Receipts by Industry

Industry	Number of Firms	Receipts (\$1,000)
001	100	1,000
0011	99	D
0012	1	D
002	200	10,000
0021	188	D
0022	12	D

NOTE: D indicates data withheld to limit disclosure. Total receipts for Industries 0011 and 0021 are suppressed so that the data for the primary suppressions in Industries 0012 and 0022 cannot be derived by subtraction.

Noise Infusion: Noise infusion is a method of disclosure avoidance in which values for each firm are perturbed prior to table creation by applying a random noise multiplier to the magnitude data (i.e., characteristics such as receipts, payroll and number of employees) for each company. Because noise infusion does not need the complementary suppression that cell suppression requires, many more statistically valid cells can be shown. Disclosure protection is accomplished in a manner that results in a relatively small change in the vast majority of cell values. For tabulations that are not based on a probability sample, each published cell value has an associated noise flag, indicating the relative amount of perturbation in the cell. For sample-based tabulations, the estimated relative standard error for a published cell includes both the estimated sampling

error and the amount of perturbation in the estimated cell value due to noise. In all tabulations using noise, cells in the table may be suppressed for additional protection from disclosure or because the quality of the data does not meet publication standards. Though some of these suppressed cells may be derived by subtraction, the results are not official and may differ substantially from the true estimate.

Example -- Same Table Without Disclosure, Protected by Noise

Number of Firms and Receipts by Industry

Industry	Number of Firms	Receipts (\$1,000)
001	100	1,004
0011	99	995
0012	1	D
002	200	9,948
0021	188	8,937
0022	12	D

NOTE: D indicates data withheld to limit disclosure. The receipts values for Industries 0012 and 0022 are suppressed because the original data were at risk of disclosure. Estimates may be derived by subtraction, but are not official and may differ substantially from the true estimates or values.

Data Swapping: Data swapping is a method of disclosure avoidance designed to protect confidentiality in tables of frequency data (the number or percent of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases when creating a table. A sample of households is selected and matched on a set of selected key variables with households in nearby geographic areas that have similar characteristics (such as the same number of adults and same number of children). Because the swap often occurs within a neighboring area, there is usually no effect on the marginal totals for the area or for totals that include data from multiple areas. Because of data swapping, users should not assume that tables with cells having a value of one or two reveal information about specific individuals. Data swapping procedures were first used in the 1990 Census, and have been used in each subsequent Census as well as the American Community Survey. For a description of the disclosure avoidance procedures used in the economic census and surveys see the discussions on suppression and noise infusion.

Protection of Microdata Files: The Census Bureau sometimes releases microdata files which contain data from the censuses of the United States population and household surveys, which it conducts. These files contain individuals' responses that represent only samples of the population and have had all individual identifiers (such as name and address) removed from the records. In addition, to protect confidentiality, the Census Bureau may modify distinguishing characteristics (such as high levels of income), and restrict geographic identifiers (such as the name of a city) so that populations are composed of at least 100,000 people. This is done to protect the identity of individuals.

Synthetic Data: The generation of synthetic data is another approach utilized by

the Census Bureau to avoid disclosure of confidential data about individuals or businesses. Synthetic data is data generated from a number of statistical models to simulate the values that are similar to and have relationships consistent with the real data. Synthetic data produces results that allow for the release of more information while still addressing strict disclosure avoidance guidelines. Therefore, researchers and other data users have significantly more relevant information that they can use in evaluating issues or making policy decisions. The Census Bureau is already using partial synthetic data to release more information about group quarters (GQ) data from the American Community Survey and will use it to protect GQ data from Census 2010.

Source: U.S. Census Bureau. Last Revised: March 17, 2009