



AN ESRI
TECHNICAL PAPER

NOVEMBER 2024

Methodology statement: 2020 Census

380 New York Street
Redlands, California 92373-8100 usa
909 793 2853
info@esri.com
esri.com



esri

THE
SCIENCE
OF
WHERE®

Copyright © 2024 Esri
All rights reserved.
Printed in the United States of America.

The information contained in this document is the exclusive property of Esri. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by Esri. All requests should be sent to Attention: Contracts and Legal Services Manager, Esri, 380 New York Street, Redlands, CA 92373-8100 USA.

The information contained in this document is subject to change without notice.

Esri, the Esri globe logo, The Science of Where, Tapestry, ArcGIS, [esri.com](https://www.esri.com), and @esri.com are trademarks, service marks, or registered marks of Esri in the United States, the European Community, or certain other jurisdictions. Other companies and products or services mentioned herein may be trademarks, service marks, or registered marks of their respective mark owners.

Table of contents

Introduction 4

Data products..... 4

Esri value added to 2020 DHC data..... 5

Census changes to 2020 DHC data..... 7

Differential privacy..... 7

Additional resources..... 8

Methodology statement: 2020 Census

Introduction The United States Census is a once-a-decade exercise to capture a snapshot of the nation’s population “once, only once, and in the right place.” The national head count is always a complex endeavor; however, the 2020 Census was like no other decennial census. The 2020 Census operations were forced to cope with the COVID-19 pandemic, record-setting wildfire and hurricane seasons, civil unrest in many urban centers, and political challenges around a potential citizenship question and how to count undocumented immigrants for congressional apportionment. Given these unprecedented challenges, the United States Census Bureau still completed the census in a timely manner and accounted for 99.98 percent of all housing units. The 2020 Census was also the first census in the United States to offer online response. This response mode was an enormous success with around four of every five households responding online.

Data products Data from the 2020 Census is primarily released through four data products. These data releases are currently ongoing as the Census Bureau is operating under a revised data product release schedule due to the impacts of the COVID-19 pandemic. Current 2020 Census data products are as follows:

- Apportionment Data, released in April 2021. This product, consisting of the counts of the resident population at the state level, is used to apportion the 435 seats in the U.S. House of Representatives among the 50 states.
- Redistricting Data (P.L. 94-171), released in August 2021. This product consists of census data at the block level and above for six tables: Race; Hispanic or Latino, and Not Hispanic or Latino by Race; Race for the Population 18 Years and Over; Hispanic or Latino, and Not Hispanic or Latino by Race for the Population 18 Years and Over; Group Quarters Population by Group Quarters Type; and Occupancy Status. This product is used by the states to delineate voting districts to be used for the next 10 years.

- The Demographic and Housing Characteristics file (DHC) was released in May 2023. This product replaces the Summary File 1 (SF1) data from the 2010 Census. This data provides basic information on population and housing as well as detailed information on age, sex, race, Hispanic or Latino origin, household type, family type, relationship to householder, group quarters population, housing occupancy, and housing tenure. For information contained in both the DHC and the Redistricting Data, values will match.
- The Detailed Demographic and Housing Characteristics (Detailed DHC) data tables partially replace the SF1 and Summary File 2 (SF2) data from the 2010 Census. These tables are split into three products and are released on a flow basis:
 - Detailed DHC-A—This file covers population by sex by age tables for detailed race and ethnic groups and American Indian and Alaskan Native tribes and villages and was released in September 2023.
 - Detailed DHC-B—This file covers household type and tenure characteristics for the same detailed race and ethnicity and American Indian and Alaskan Native populations covered in DHC-A. This file is planned for release in September 2024.
 - Supplemental DHC (S-DHC)—This file covers detailed information on population and housing in combined data tables with reconciled discrepancies between the population and housing universes resulting from the 2020 Census privacy protections. This file is planned for release in September 2024.¹

Esri value added to 2020 DHC data

The 2020 Census DHC data was released by [Esri Demographics](#) for the United States and the territory of Puerto Rico in October 2023. This product allows data users to access select information from the DHC data down to the block group level through Esri's ArcGIS GeoEnrichment Service, ArcGIS Business Analyst web and mobile apps, ArcGIS Business Analyst Pro, and ArcGIS Community Analyst. Users who are interested in census block-level 2020 data can access those files in ArcGIS [Living Atlas of the World](#) and through [Esri's Redistricting application](#). These ArcGIS Living Atlas layers can be [imported into Business Analyst](#) to satisfy additional workflows.

Esri has added value to the 2020 DHC data in several ways:

- Facilitation of temporal analysis
 - Esri converted select SF1 variables from the 2000 Census and 2010 Census to 2020 geography, enabling temporal analysis for variables such as total population, household population, group quarters population, households, housing units, and average household size. Converting 2000 and 2010 data to 2020 geographic boundaries enables comparisons across three time periods. Converting this data

¹ <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/release/about-2020-data-products.html>

required the creation of a correspondence file from 2000 and 2010 block groups to 2020 block groups built from 2000 and 2010 block to 2020 block group relationships.

To establish these relationships between vintages, Esri incorporates the block-to-block relationship files from the Census Bureau.²

Esri created versions of these files that were required for the 2000 to 2020 block relationships since this file is unavailable from the Census Bureau. These relationship files are used to build out all 11 geographic schemas for select SF1 2000 and 2010 data in 2020 geography.

- User-defined areas
 - ArcGIS Business Analyst and ArcGIS GeoEnrichment Service in ArcGIS Pro and ArcGIS Online allow users to build reports and infographics for any user-defined area. These systems use decennial census block weights in the data apportionment process.
- Standard geographies
 - For the United States, the 2020 DHC data is available in Esri's 14 geographic schemas, including TomTom residential ZIP codes and Nielsen Designated Market Areas (DMAs). Neither schema was part of the Census Bureau's data release. For Puerto Rico, data is available for eight geographic schemas, including TomTom residential ZIP codes.
- Calculated items
 - Esri produces a set of derived statistics, including compound annual growth rates for all totals, vacant housing units, average household size, population density, and Esri's Diversity Index.
- Hispanic by Race and Hispanic by Race for the population under the age of 18
 - The Census Bureau releases tables for the non-Hispanic population by race and non-Hispanic population by race for the population 18 years of age and older. Esri uses the released census data to calculate the residual Hispanic data tables. These include Hispanic population by race and Hispanic population by race for the population under the age of 18.

² <https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html>

Census changes to 2020 DHC data

Every decennial census effort intends to improve on the prior decennial census. In addition to allowing for online response, the Census Bureau implemented subtle but important changes to improve the race and ethnicity questions on the 2020 decennial census questionnaire. These consisted of changes to the wording and examples provided on the form for questions related to race and ethnicity. The word *Negro* was removed, and the choice Guamanian or Chamorro was changed to Chamorro. Most importantly, the write-in instructions for Some Other Race were changed from Print race to Print race or origin. The maximum number of characters processed by the Census Bureau for write-in responses was lengthened from 30 to 200 characters. Write-in responses were processed into a maximum of six categories, which is up from two in the 2010 Census.³

Differential privacy

The most significant departure from the 2010 Census is the Disclosure Avoidance System (DAS) used for the 2020 Census. The Census Bureau is required to keep the collected information confidential for 72 years, and under Title 13, any individual must be protected from being identifiable in published data. In past censuses, the Census Bureau used various forms of disclosure avoidance to ensure privacy protection. These techniques consisted of table suppression and data swapping.

The 2020 Census opens a new era of disclosure avoidance with the implementation of differential privacy. Differential privacy is a formal statistical technique used to add noise to the tabulations to better safeguard individual privacy. The DAS consists of two components: differential privacy and post-processing adjustments. Differential privacy adds statistical noise to the data to protect individual privacy while post-processing is used to adjust the noisy data so that it looks like census data that users are accustomed to receiving. These adjustments ensure that fractions or negative values are removed, components sum to their respective table totals, and impossible or improbable statistics are kept to a minimum. State-level population and housing unit statistics at every geographic level are actual counts and do not have noise added. All other statistics are subjected to noise and should not be treated as actual counts.

There are positive and negative aspects to using differential privacy as part of the 2020 Census DAS. In the past, the magnitude of data swapping in the decennial census was not disclosed. For the 2020 Census, the Census Bureau has openly shared its statistical methods while working with stakeholders to fine-tune the DAS to protect privacy but also produce data that is fit for use. However, census data users are likely to find it difficult to understand the magnitude of noise in differentially privatized data. While the post-processing does fix many of the issues caused by the application of the differential privacy method, the DAS still produces many impossible and improbable results at smaller levels of geography—something that did not exist in prior decennial census releases.

The 2020 Census DAS has inherent limitations that reduce both the number of data tables and the geographic granularity of data availability in the DHC and Detailed DHC. In summary, decennial census data users will have less data overall and less small area data to work with compared to prior censuses.

³ Improvements to the 2020 Census Race and Hispanic Origin Question Designs, Data Processing, and Coding Procedures. <https://www.census.gov/newsroom/blogs/random-samplings/2021/08/improvements-to-2020-census-race-hispanic-origin-question-designs.html>

Additional resources

To better understand how differential privacy impacts 2020 Census data, view additional resources here: <https://differential-privacy-resources-esri.hub.arcgis.com/>. Esri is dedicated to educating users about changes in the 2020 Census data releases and is committed to helping users leverage census data using best practices. As more information and data are released, this document and additional 2020 Census documentation will be updated.

For more information about business data, call 1-800-447-9778.

Esri's Data Development team

Led by chief demographer Kyle Cassal, and economist Douglas Skuta, Esri's Data Development team uses sophisticated quantitative methods to produce small area demographic and socioeconomic data to support informed decision-making. The team builds on a rich history of market intelligence to produce trusted independent estimates and forecasts for the United States based on innovative methodologies that use public and private data sources with the power of ArcGIS. Esri's Data Development team provides more than 7,000 proprietary data items to better understand the characteristics of people and places across multiple statistical and administrative boundaries and custom trade areas.



Esri, the global market leader in geographic information system (GIS) software, location intelligence, and mapping, helps customers unlock the full potential of data to improve operational and business results.

Founded in 1969 in Redlands, California, USA, Esri software is deployed in more than 350,000 organizations globally and in over 200,000 institutions in the Americas, Asia and the Pacific, Europe, Africa, and the Middle East. Esri has partners and local distributors in over 100 countries on six continents, including Fortune 500 companies, government agencies, nonprofits, and universities. With its pioneering commitment to geospatial information technology, Esri engineers the most innovative solutions for digital transformation, the Internet of Things (IoT), and advanced analytics.

Visit us at esri.com.



Contact Esri

380 New York Street
Redlands, California 92373-8100 USA

1 800 447 9778
T 909 793 2853
F 909 793 5953
info@esri.com
esri.com

Offices worldwide
esri.com/locations

For more information, visit
esri.com/data/esri_data.