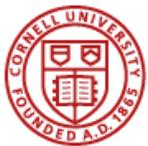


- Demonstration data availability
- Proposed 2020 data products

Jan Vink

Cornell Program on Applied Demographics

[jkv3@cornell.edu](mailto:jkv3@cornell.edu)



Cornell University



# Demonstration data products data

---

- At the Census
  - <https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/?#>
  - Data format as SF1, 1 geographic table and 15 segments with 1 or more tables within
    - Geographic table is comma separated, segment tables are fixed width
    - No headers
    - Tables link with unique LOGRECNO
  - Detailed code book is missing, but can be reconstructed from available documentation

# Demonstration data products data

- At NHGIS
  - <https://www.nhgis.org/differentially-private-2010-census-data>
  - Separate files for different levels of geographies
  - Demonstration data and SF1 combined
  - Select number of tables
  - CSV files



The screenshot shows the IPUMS NHGIS website interface. At the top right, it says 'IPUMS.ORG'. Below that, there are tabs for 'NHGIS' and 'GEOMARKER', with 'NATIONAL HISTORICAL GIS' written below. A navigation bar includes links for 'HOME | SELECT DATA | MY DATA | FAQ | HELP | LOG IN'. A row of six small images is displayed below the navigation bar. The main content area is titled 'DIFFERENTIALLY PRIVATE 2010 CENSUS DATA'. The text reads: 'To protect the confidentiality of 2020 Census respondents, the U.S. Census Bureau plans to use a framework termed "differential privacy". The IPUMS team is concerned about the potential impacts of this approach on the scope and quality of census data products.' Below this, it states: 'To help data users assess the impacts of differential privacy, the Census Bureau has released 2010 demonstration data products, which supply differentially private versions of 2010 data. To facilitate comparisons between the original and differentially private 2010 data, we provide both versions joined together in a simplified format through the links below.'

# NHGIS DP data

---

- Features of the NHGIS Demonstration Files:
  - They have separated the data for different geographic summary levels into individual files
  - The data files include standard NHGIS "GISJOIN" identifiers and NHGIS variable codes for both the original and demonstration data
  - The ZIP archives contain human-readable codebooks describing the contents of the data files

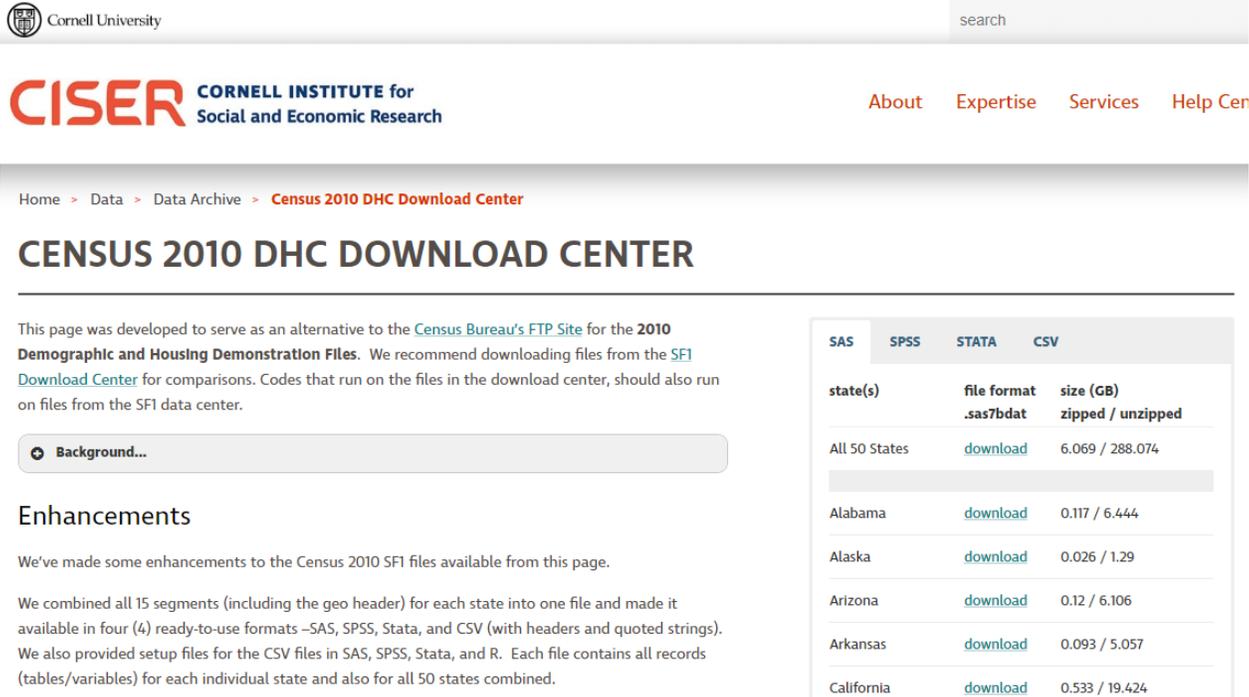
# NHGIS DP data

---

- Features of the NHGIS Demonstration Files:
  - "Wide" files place each data variable in a separate column. "Long" files place each data variable on a separate row (so each row corresponds to a single geographic unit and single data variable).
  - Coverage:
    - Geographic Levels: 11 levels selected from the "National" Demographic and Housing Demonstration File and 3 levels selected from the "State" Demographic and Housing Demonstration File
    - Tables: All tables from the 1st (P1-P22) and 14th (H1-H13) segments of the Demographic and Housing Demonstration File

# Demonstration data products data

- At Cornell
  - <https://ciser.cornell.edu/data/data-archive/census-2010-dhc-download-center/>
  - Separate files for each state in several formats
  - All geographic information and data tables in this single file
  - SF1 data in similar format available



The screenshot shows the Cornell University CISER website. The header includes the Cornell University logo and a search bar. The main navigation menu has links for 'About', 'Expertise', 'Services', and 'Help Cen'. The breadcrumb trail is 'Home > Data > Data Archive > Census 2010 DHC Download Center'. The page title is 'CENSUS 2010 DHC DOWNLOAD CENTER'. The main content area contains a paragraph explaining the page's purpose as an alternative to the Census Bureau's FTP site, a 'Background...' button, and an 'Enhancements' section. The 'Enhancements' section describes how 15 segments for each state were combined into one file and made available in four formats: SAS, SPSS, Stata, and CSV. A table on the right side of the page lists download options for various states.

Home > Data > Data Archive > **Census 2010 DHC Download Center**

## CENSUS 2010 DHC DOWNLOAD CENTER

This page was developed to serve as an alternative to the [Census Bureau's FTP Site](#) for the **2010 Demographic and Housing Demonstration Files**. We recommend downloading files from the [SF1 Download Center](#) for comparisons. Codes that run on the files in the download center, should also run on files from the SF1 data center.

[Background...](#)

### Enhancements

We've made some enhancements to the Census 2010 SF1 files available from this page.

We combined all 15 segments (including the geo header) for each state into one file and made it available in four (4) ready-to-use formats --SAS, SPSS, Stata, and CSV (with headers and quoted strings). We also provided setup files for the CSV files in SAS, SPSS, Stata, and R. Each file contains all records (tables/variables) for each individual state and also for all 50 states combined.

SAS	SPSS	STATA	CSV
state(s)	file format	size (GB)	zipped / unzipped
All 50 States	<a href="#">download</a>	6.069 / 288.074	
Alabama	<a href="#">download</a>	0.117 / 6.444	
Alaska	<a href="#">download</a>	0.026 / 1.29	
Arizona	<a href="#">download</a>	0.12 / 6.106	
Arkansas	<a href="#">download</a>	0.093 / 5.057	
California	<a href="#">download</a>	0.533 / 19.424	

# Cornell DHC download center

- The field names (2686 fields) are as defined in: [2010 Demonstration Data Products Technical Document](#)

Table number	Table contents	Data dictionary reference name	Segment	Max size
P3	<b>RACE [8]</b>			
	Universe: Total population			
	Total:	P0030001	01	9
	White alone	P0030002	01	9
	Black or African American alone	P0030003	01	9
	American Indian and Alaska Native alone	P0030004	01	9
	Asian alone	P0030005	01	9
	Native Hawaiian and Other Pacific Islander alone	P0030006	01	9
	Some Other Race alone	P0030007	01	9
Two or More Races	P0030008	01	9	
P4	<b>HISPANIC OR LATINO ORIGIN [3]</b>			
	Universe: Total population			
	Total:	P0040001	01	9
	Not Hispanic or Latino	P0040002	01	9
Hispanic or Latino	P0040003	01	9	

# Proposed 2020 data products

# Overview of the Proposed 2020 Census Data Products Plan

- The current disclosure avoidance system either supports or is being expanded to support the following 2020 Census data products
  - Apportionment Product
  - Redistricting File (P.L. 94-171)
  - Demographic Profile
  - Demographic and Housing Characteristics File (DHC)
  - Congressional District Demographic and Housing Characteristics File
- We are developing other formal privacy systems for public data products that include the tables below that meet user needs yet maintain data confidentiality
  - Detailed race and Hispanic origin tables and some family/household tables included in Summary File 1 in 2010
  - Summary File 2
  - American Indian and Alaska Native Summary File
  - Public Use Microdata Sample (PUMS) File
- We are not proposing to produce the following products for 2020
  - Summary of Population and Housing Characteristics (CPH-1)
  - Population and Housing Unit Counts Report Series (CPH-2)

## Apportionment Product

- The Apportionment Product will be the first release of the 2020 Census and will consist of the resident population of the 50 states plus the overseas federal employees (military and civilian) and their dependents living with them, who are included in their home states

## Redistricting File (P.L. 94-171)

- Public Law 94-171 directs the Census Bureau to provide data to the governors and legislative leadership in each of the 50 states for redistricting purposes. This product will be the first file released that will include demographic and housing characteristics about detailed geographic areas

## Demographic Profile

- This product will provide critical demographic and housing characteristics about local communities as soon after the release of the Redistricting file as possible

## Demographic and Housing Characteristics File (DHC)

- The DHC will include many of the demographic and housing tables previously included in Summary File 1

## Congressional District Demographic and Housing Characteristics File

- Provides a re-tabulation of the Demographic and Housing Characteristics File for the new congressional districts

# Demographic and Housing Characteristics File (DHC)

- The DHC will include many of the demographic and housing tables previously included in Summary File 1
- **Subjects will include:**
  - Age
  - Sex
  - Race
  - Hispanic or Latino origin
  - Household type
  - Family type
  - Relationship to householder
  - Some subjects provided for major OMB race/ethnicity groups
  - Group quarters population
  - Housing occupancy
  - Housing tenure
- **Geography:** Various; data will be available for all geographies at and above the lowest level they are released for. In many cases this will be at the Census Block level.
- The DHC will not include tables that provide counts by detailed race, Hispanic origin, tribes, and population by household/family types due to challenges associated with the use of differential privacy for privacy protection
- We are developing other solutions for data products that include these tables

# Data Available from the Redistricting File and the Demographic and Housing Characteristics File (DHC)

Census Block (\* = provided for major race and Hispanic origin groups)

- Total Population
- Urban Rural Designation
- Sex
- Race
- Hispanic or Latino origin
- Sex by 5 year and other age groups\*
- Sex by Single year-of-age under 20
- Relationship to Householder
- Households by Householder Characteristics\*
- Household Type by Relationship\*
- Households by presence of people of specific ages\*
- Housing Tenure\*
- Household Size
- Group Quarters population by sex and age for major types
- Median Age for various groups\*

## Census Tract

- Population by Single year-of-age up to 100 by Sex
- Tenure by Race by Hispanic or Latino origin
- Tenure by Presence of Own Children

## County

- Household Size by Household Type by Presence of Own Children
- Household Type by Age of Householder
- Household Type by presence of people of specific ages

## State

- Group Quarters Population by Sex by Age by Major Group Quarters type

# Census Data Products Crosswalk

---

- <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html>
- Excel file that contains
  - Table layouts for proposed 2020 PL, DP and DHC tables
  - A crosswalk from the SF1 table ID's to the 2020 DHC table ID's

# Table ID and availability

Table prefix	2010 SF1	2020 DHC
P, H	All geographies	Available for all geographies
PCT, HCT	Tracts, Places, MCD, Legislative Districts, ZCTA, School Districts	Tracts, <i>other geographies not yet decided, possibly with threshold</i>
PCO	Counties, Legislative Districts	Counties, <i>other geographies not yet decided</i>
PST	N/A	State and Nation (new in 2020)

Table postfix		Table postfix	
A	White alone	I	Non Hispanic White alone
B	Black or African American alone	J	Non Hispanic Black or African American alone
C	American Indian and Alaska Native alone	K	Non Hispanic American Indian and Alaska Native alone
D	Asian alone	L	Non Hispanic Asian alone
E	Native Hawaiian and Other Pacific Islander alone	M	Non Hispanic Native Hawaiian and Other Pacific Islander alone
F	Some Other Race alone	N	Non Hispanic Some Other Race alone
G	Two or more races	O	Non Hispanic Two or more races
H	Hispanic or Latino		

Not in 2020 proposed products

# 2010 SF1 – 2020 DHC Crosswalk (counts)

2010 table	Post fix	2020 proposed tables				Eliminated
		P, H	PCT, HCT	PCO	PST	
P, H		31		12		20 (15)
P, H	A-I	4*9		5*9		9*9
PCT, HCT			5	4	1	18
PCT	A-I			2*9		3*9
PCT	A-O					1*15
PCO				10		
PCO	A-I			1*9		
Allocation tables						11

# Conclusions

---

- For race
  - Detailed race groups in publications to be announced
  - Multi-race groups not in DHC, but in PL94
- For age
  - Single year of age only available for total population (PCT), not longer by race
  - Age groups (23) available at block level, also for A-I race groups BUT not J-O
    - A-I adds up to a total  $\geq$  Total population
  - Age of household population only available in PCO table

# Conclusions

---

- For households
  - Many cross tabulations from block level moved to County level
  - Several tables eliminated
  - Some tables collapse children age groups
  - Some tables add detail on same-sex couples
- For families
  - Most tables eliminated
  - Some household type and relationship tables contain some family information

# Conclusions

---

- For Group Quarters
  - Many tables moved up from block and tract to county and state
- For housing units
  - Household size by tenure eliminated