

FREQUENTLY OCCURRING SURNAMES IN THE 2010 CENSUS

Joshua Comenetz

October 2016

INTRODUCTION

The 2010 Census Surname Table is the third in the U.S. Census Bureau's series of data products listing the most common surnames (last names) reported in the decennial U.S. census. The Census 2000 product was based on very different methods from the 1990 Census product, while the 2010 product is based on the methodology and programming used in 2000 with some revisions. This document first summarizes the methodological changes between 2000 and 2010, then describes the 2010 results and compares them with the 2000 surname list.

This report documents both the overall frequency of surnames (last names), as well as some of the basic demographic characteristics that are associated with surnames. The presentation of data in this report focuses on summarized aggregates of counts and characteristics associated with surnames, and the data do not in any way identify any specific individuals.

METHODOLOGY CHANGES

The 2010 surname project began as a replication of the 2000 project, using the same programming. This section describes ways in which the 2010 methods differed from the 2000 methodology that is described in Word et al. (2007). For the complete methodology employed in the 2010 project, please refer to both Word et al. (2007) and this document.

Creation of the 2010 product started with raw data from the 2010 Census: the surnames provided by census respondents. These data were processed by the 2000 programming, applying edits described in Word et al. (2007), resulting in the initial 2010 draft product. The initial 2010 draft names list was extensively reviewed by a group of Census Bureau staff. Their findings were consolidated into a new set of edits, comprising updates or corrections needed to fix problems that arose in 2010 but not in 2000, or to fix problems that were partially addressed in the 2000 programming but were more widespread or complex in the 2010 data set. New programming, to add to the 2000 programs, was written to implement the new set of edits. Applying the new programming to the initial draft product resulted in a revised draft. The revised draft was reviewed to see if all of the new set of edits were addressed, and also compared with the 2000 product to see if changes in name frequency (name count, or number of times each name was reported in the census) over the decade made sense in light of general demographic trends. Several rounds of programming and review resulted in the publicly available 2010 surname product.

Edits applied in both 2000 and 2010 with variation

In the 2000 project, first names were extensively examined as part of the overall review of reported names. First names were used in 2010 only to the extent that the 2000 programming uses them as a component of its edits. We did not otherwise employ first names in the 2010 project.

In 2000, manual review (“eyeballing”) was conducted for names with a frequency of at least 300. Names with lower frequencies were not examined. In 2010, partly to compensate for the reduced role of first names, we conducted a much larger manual review. All names with a frequency of at least 10 were scanned by two or more reviewers, with an eye toward obvious erroneous entries (e.g., I LIVE ALONE), extraneous characters (SMITH--- rather than SMITH), spelling errors (JACKOSN rather than JACKSON), inclusion of suffixes (GONZALEZJR rather than GONZALEZ), and any other potential errors.

Edits and a correction newly applied in the 2010 product

Most edits applied in the Census 2000 names product were carried over to the 2010 product. Based on the manual review team’s work, we developed a list of additional corrections necessary to fix problems seen in the 2010 data. These included deletion of punctuation, spaces, suffixes, prefixes, and non-names, along with correction of obvious spelling errors, as summarized below.

One set of edits related to punctuation and spaces:

- All punctuation marks and spaces that were part of surnames were deleted. SMITH-JONES and SMITH JONES were changed to SMITHJONES. O’HARA and O HARA became OHARA.
- Many names on the raw list also contained extraneous punctuation. The punctuation was deleted, so, for example, SMITH--- became SMITH.
- All reported surnames consisting of two letters separated by a space (A A, C D) were deleted from the list.
- Names with more than two letters that contained extraneous spaces were edited to remove the spaces, e.g., S M I TH became SMITH.

Another group of edits removed suffixes and prefixes:

- JUNIOR, SENIOR, and THIRD were removed from the ends of names.
- JR and III were deleted from both the beginning and end of names.
- SR was deleted from the end of name only if the name had at least seven alphabetic characters, because several shorter names do end in SR while longer names ending in SR tended to include suffixes (e.g., NASR was kept, WILSONSR was changed to WILSON).
- II was deleted from the beginning or end of a name only if it was separated by a space, because a number of names do begin or end with II (NISHII was not changed, JONES II and II JONES were both changed to JONES).

- Similarly, J R and S R were deleted from the beginning or end of a name if they were separated from the rest of a name by a space (GONZALEZ S R became GONZALEZ).
- IV was deleted from the end of a name if separated by a space.

A third set of edits deleted and corrected names. The 2000 programming contains deletion and correction lists that were maintained and expanded for the 2010 project. For example:

- Except for NG, all 2-letter names that did not contain a vowel or Y were deleted (AA, AB, and BY were kept; DD and BC were deleted).
- Obvious non-names were deleted: for example, ABCDEFG, ADULT FEMALE, DECLINE TO STAT, DONT KNOW, NO NOMBRE, OOOOO, SAME AS ABOVE. The goal was to delete only entries that, like these, were very unlikely to be genuine surnames. Names that looked improbable but not impossible were kept, e.g., ANONYMOUS.
- Several obvious spelling errors were corrected. For example, JACKOSN became JACKSON, RODRIGU Z became RODRIGUEZ, WRGHT became WRIGHT.

Finally, a small error in the classification of names by race in the 2000 programming was discovered and corrected. The 2000 and 2010 products use the same six Hispanic origin and race categories: Non-Hispanic White Alone, Non-Hispanic Black or African American Alone, Non-Hispanic American Indian and Alaska Native Alone, Non-Hispanic Asian and Native Hawaiian and Other Pacific Islander Alone, Non-Hispanic Two or More Races, and Hispanic or Latino origin. However, in the 2000 product the Non-Hispanic Native Hawaiian and Other Pacific Islander Alone group was combined with the Non-Hispanic Two or More Races group, rather than with the Non-Hispanic Asian Alone group. This misclassification affected only about 0.1 percent of people responding to Census 2000. For the 2010 product, we corrected the classification procedure, and people identifying as Non-Hispanic Asian Alone or Non-Hispanic Native Hawaiian and Other Pacific Islander Alone are grouped together.¹

SUMMARY OF RESULTS

Table 1 outlines the national distribution of surnames by frequency in the United States in the 2010 Census. About 6.3 million different surnames were reported by census respondents in 2010. Of these, 11 were reported more than a million times each, while 3.9 million, or 61.9 percent, were reported only once: unique surnames or unique variations of more frequent names (including errors resulting from the transcription of names handwritten on census forms). Surnames were recorded for 295 million people, or 95.5 percent of the population enumerated in the 2010 Census. This coverage is comparable to surname coverage in 2000 (95.9 percent of the population). All of the editing procedures applied to the 2010 Census surnames together affected about 7.7 million individual records, or 2.6% of the 295 million records that contained surnames.

The 50 most common surnames are summarized in Appendix Table A-1, which provides frequency, rank, and distribution by Hispanic origin and race groups. The Hispanic origin and

¹ In the rest of this document, Black is used interchangeably with Non-Hispanic Black or African American Alone; Hispanic is used interchangeably with Hispanic or Latino origin; Asian and Native Hawaiian and Other Pacific Islander is used interchangeably with Non-Hispanic Asian and Native Hawaiian and Other Pacific Islander Alone.

race groups are the same as used in the 2000 product (Word et al., 2007). The top five surnames are the same in both years (SMITH, JOHNSON, WILLIAMS, BROWN, JONES). Below the top five, the most notable change from 2000 to 2010 is the increasing prominence of predominantly Hispanic surnames. In 2000, the top 50 contained 10 surnames that were at least 90 percent Hispanic, and the top-ranked Hispanic surname (GARCIA) was ranked 8. In 2010, 12 of the top 50 were at least 90 percent Hispanic, and GARCIA was ranked 6.

Table 1. Surnames by frequency range and number of people: 2010

Surnames				People with these Surnames		
Frequency range	Number	Cumulative Number	Cumulative Proportion (percent)	Number	Cumulative Number	Cumulative Proportion (percent)
1,000,000+	11	11	0.0	15,505,815	15,505,815	5.3
100,000-999,999	303	314	0.0	66,723,063	82,228,878	27.9
10,000-99,999	3,253	3,567	0.1	84,107,944	166,336,822	56.4
1,000-9,999	21,322	24,889	0.4	61,658,357	227,995,179	77.3
100-999	137,364	162,253	2.6	37,672,049	265,667,228	90.1
50-99	114,079	276,332	4.4	7,939,724	273,606,952	92.8
25-49	177,884	454,216	7.2	6,182,636	279,789,588	94.9
10-24	354,912	809,128	12.8	5,448,514	285,238,102	96.7
5-9	424,454	1,233,582	19.6	2,761,552	287,999,654	97.6
2-4	1,165,587	2,399,169	38.1	3,079,711	291,079,365	98.7
1	3,899,864	6,299,033	100.0	3,899,864	294,979,229	100.0

Source: U.S. Census Bureau, 2010 Census.

The next two tables look only at the 1,000 most frequent surnames. Table 2 shows the 10 names with the highest shares of identification with each Hispanic origin or race group—for example, 87.5 percent of people with the name WASHINGTON identified themselves as Black. There is little overlap among the six lists, highlighting the degree to which surnames vary among racial and ethnic groups in the United States.

Most individual surnames do not reflect the diversity of the population as a whole. In many cases, over 90 percent of people reporting a name identify with just one of the six Hispanic origin or race groups. In contrast, few surnames are equally common in multiple Hispanic origin or race groups. Table A-1 shows something similar: almost all of the top 50 names show Hispanic origin and race patterns very different from the U.S. population as a whole.

**Table 2. Top 1,000 surnames with the largest share in each
Hispanic origin and race group: 2010**

NAME	RANK AMONG ALL SURNAMES	FREQUENCY	OCCURENCES PER 100,000 PEOPLE	PERCENT IN THIS GROUP
------	----------------------------------	-----------	-------------------------------------	-----------------------------

NON-HISPANIC WHITE ALONE

YODER	598	56,410	19.1	97.8
FRIEDMAN	901	38,374	13.0	96.1
KRUEGER	946	36,743	12.5	96.0
SCHWARTZ	345	90,071	30.5	95.9
SCHMITT	967	36,043	12.2	95.7
MUELLER	523	64,191	21.8	95.7
WEISS	587	57,112	19.4	95.6
NOVAK	980	35,461	12.0	95.5
OCONNELL	961	36,194	12.3	95.4
KLEIN	392	81,471	27.6	95.4

NAME	RANK AMONG ALL SURNAMES	FREQUENCY	OCCURRENCES PER 100,000 PEOPLE	PERCENT IN THIS GROUP
------	----------------------------------	-----------	--------------------------------------	-----------------------------

NON-HISPANIC AMERICAN INDIAN AND ALASKA NATIVE ALONE

LOWERY	799	43,278	14.7	4.5
HUNT	169	158,421	53.7	4.2
SAMPSON	886	39,277	13.3	3.9
JACOBS	257	118,614	40.2	3.8
MOSES	887	39,216	13.3	3.2
LUCERO	865	39,986	13.6	2.9
JAMES	85	249,379	84.5	2.6
PROCTOR	977	35,636	12.1	2.4
ASHLEY	900	38,499	13.1	2.4
CUMMINGS	403	79,316	26.9	2.3

NON-HISPANIC BLACK OR AFRICAN AMERICAN ALONE

WASHINGTON	145	177,386	60.1	87.5
JEFFERSON	615	55,179	18.7	74.2
BOOKER	941	36,840	12.5	65.2
BANKS	292	105,833	35.9	54.5
JOSEPH	313	100,959	34.2	54.2
MOSLEY	730	47,963	16.3	53.2
JACKSON	19	708,099	240.1	53.0
CHARLES	548	61,211	20.8	53.0
DORSEY	793	43,631	14.8	52.2
RIVERS	897	38,662	13.1	50.9

NON-HISPANIC TWO OR MORE RACES

ALI	559	60,002	20.3	5.7
KHAN	427	76,171	25.8	5.3
WONG	274	111,371	37.8	5.2
SINGH	260	116,749	39.6	4.8
CHANG	391	81,939	27.8	4.2
CHUNG	805	43,110	14.6	4.0
AHMED	720	48,319	16.4	4.0
WASHINGTON	145	177,386	60.1	3.8
BOOKER	941	36,840	12.5	3.8
JEFFERSON	615	55,179	18.7	3.5

**NON-HISPANIC ASIAN AND NATIVE HAWAIIAN AND OTHER PACIFIC
ISLANDER ALONE**

XIONG	929	37,499	12.7	98.1
ZHANG	472	70,125	23.8	98.1
HUANG	456	72,328	24.5	97.1
TRUONG	903	38,267	13.0	96.9
YANG	290	106,033	36.0	96.8
LI	273	111,786	37.9	96.8
VANG	726	48,036	16.3	96.7
HUYNH	620	54,996	18.6	96.7
VU	962	36,179	12.3	96.6
NGUYEN	38	437,645	148.4	96.5

HISPANIC OR LATINO ORIGIN

BARAJAS	763	45,153	15.3	96.5
ZAVALA	667	51,153	17.3	96.1
VELAZQUEZ	530	63,736	21.6	96.0
AVALOS	940	36,922	12.5	95.8
OROZCO	512	65,904	22.3	95.8
VAZQUEZ	203	138,322	46.9	95.8
JUAREZ	306	101,949	34.6	95.7
MEZA	639	53,230	18.1	95.5
HUERTA	735	47,455	16.1	95.5
IBARRA	507	66,063	22.4	95.3

Source: U.S. Census Bureau, 2010 Census.

Table 3 shows the 15 most rapidly increasing surnames that were among the top 1,000 in both 2000 and 2010, along with their Hispanic origin and race group characteristics. All 15 are predominantly Asian or Hispanic, except for ALI, a name with unusually great Hispanic origin and race diversity. This pattern reflects national population change between 2000 and 2010, when the Hispanic and Asian populations increased much more rapidly than the population as a whole.

Table 3. Fastest growing surnames among the top 1,000: 2000 to 2010

NAME	FREQUENCY IN 2000	FREQUENCY IN 2010	NON-HISPANIC					PERCENT HISPANIC OR LATINO ORIGIN	PERCENT CHANGE IN FREQUENCY
			PERCENT WHITE ALONE	PERCENT BLACK OR AFRICAN AMERICAN ALONE	PERCENT ASIAN AND NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER ALONE	PERCENT AMERICAN INDIAN AND ALASKA NATIVE ALONE	PERCENT TWO OR MORE RACES		
ZHANG	33,202	70,125	1.0	0.2	98.1	0.0	0.6	0.2	111.2
LI	57,786	111,786	1.5	0.2	96.8	0.0	0.9	0.6	93.4
ALI	36,079	60,002	17.3	30.7	42.9	0.5	5.7	2.9	66.3
LIU	47,665	78,370	1.8	0.2	95.6	0.0	1.9	0.5	64.4
KHAN	46,713	76,171	6.9	3.9	81.3	0.5	5.3	2.2	63.1
VAZQUEZ	84,926	138,322	3.5	0.4	0.2	0.1	0.1	95.8	62.9
WANG	67,570	109,883	2.6	0.3	95.2	0.0	1.5	0.4	62.6
HUANG	44,715	72,328	0.9	0.1	97.1	0.0	1.4	0.4	61.8
LIN	49,360	79,508	1.7	0.2	95.9	0.0	1.8	0.5	61.1
SINGH	72,642	116,749	4.3	4.5	82.8	1.2	4.8	2.5	60.7
CHEN	105,544	169,580	1.4	0.3	96.1	0.0	1.6	0.5	60.7
BAUTISTA	37,847	60,264	3.2	0.3	18.2	0.3	1.3	76.8	59.2
VELAZQUEZ	40,030	63,736	3.3	0.4	0.1	0.1	0.1	96.0	59.2
PATEL	145,066	229,973	2.1	0.4	94.8	0.7	1.7	0.4	58.5
WU	45,815	71,721	1.4	0.2	96.4	0.0	1.6	0.5	56.5

Source: U.S. Census Bureau, Census 2000 and 2010 Census.

Note: Hispanic origin and race percent distribution is from the 2010 Census.

The publicly released table of surnames includes all with a frequency of 100 or more. These 162,253 names cover 90 percent of the people with surnames recorded in the 2010 Census (see Table 1 in this document). The proportion of the population covered by the most frequent surnames increased slightly between 2000 and 2010. In 2000, 76.6 percent of the population reported a surname that had a frequency of 1,000 or more, while in 2010 that share increased to 77.3 percent—that is, “surname diversity” declined slightly over the decade.

Recent geographically diverse immigration might be expected to increase, rather than decrease, surname diversity. However, this has not happened, because most immigration is from Latin America and Asia, and Hispanic and Asian surnames are less diverse than the U.S. average. For the population as a whole in 2010, the 10 most frequent surnames accounted for just 4.9 percent of the population (Table 4), while to cover a quarter of the total population required 239 names.

The White population was more diverse than the national average in this respect, but in other groups the most common surnames were reported by a larger share of the population. The highest level of clustering is found in the Hispanic population, where just 26 surnames cover a quarter of the population and 16.3 percent of people reported one of the top 10 names. A similar pattern was seen for the Asian and Native Hawaiian and Other Pacific Islander population and for the Black population.

Table 4. Surname clustering by Hispanic origin and race groups: 2010

Group	Percent of group covered by top 10 surnames	Number of surnames needed to cover 25 percent of group
Total population	4.9	239
White Alone	4.5	319
Black Alone	13.0	43
Asian and Native Hawaiian and Other Pacific Islander Alone	13.4	41
American Indian and Alaska Native Alone	6.6	144
Two or More Races	6.2	196
Hispanic or Latino origin	16.3	26

Source: U.S. Census Bureau, 2010 Census.

ACKNOWLEDGEMENTS

Several rounds of review of the 2010 surname list were completed by William Bredemeyer, Derek Breese, Warren Davis, Delma Frankel, Myoung Ouk Kim, Danielle Kittredge, Julie VanEerden, and Tiffani Whaley. Aaron Dixon wrote programming to assist with quality review. Programming to apply edits and updates to the raw 2010 data was written by Hung Pham and Chandararith (Ricky) Phe. Robert Nunziata reviewed and updated the 2000 programming before the 2010 project began, and he and Brian Barenbaum advised on the 2010 names project. Robert Kominski and Charles Coleman provided useful background based on experience with the 2000 project. The final surname product, including this document, was reviewed by Aaron Dixon, James Fitzsimmons, and Marc Perry.

REFERENCES

Word, David L., Charles D. Coleman, Robert Nunziata and Robert Kominski. 2007. Demographic Aspects of Surnames from Census 2000. Available at: <<http://www2.census.gov/topics/genealogy/2000surnames/surnames.pdf>>. Last accessed September 16, 2016.

APPENDIX

Table A-1. Top 50 surnames: 2010

NAME	RANK	FREQUENCY	PROPORTION PER 100,000		NON-HISPANIC					
			NAME	CUMULATIVE	PERCENT WHITE ALONE	PERCENT BLACK OR AFRICAN AMERICAN ALONE	PERCENT ASIAN AND NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER ALONE	PERCENT AMERICAN INDIAN AND ALASKA NATIVE ALONE	PERCENT TWO OR MORE RACES	PERCENT HISPANIC OR LATINO ORIGIN
TOTAL		294,979,229			64.3	12.0	4.9	0.7	1.8	16.3
SMITH	1	2,442,977	828	828	70.9	23.1	0.5	0.9	2.2	2.4
JOHNSON	2	1,932,812	655	1,483	59.0	34.6	0.5	0.9	2.6	2.4
WILLIAMS	3	1,625,252	551	2,034	45.8	47.7	0.5	0.8	2.8	2.5
BROWN	4	1,437,026	487	2,522	58.0	35.6	0.5	0.9	2.6	2.5
JONES	5	1,425,470	483	3,005	55.2	38.5	0.4	1.0	2.6	2.3
GARCIA	6	1,166,120	395	3,400	5.4	0.5	1.4	0.5	0.3	92.0
MILLER	7	1,161,437	394	3,794	84.1	10.8	0.5	0.7	1.8	2.2
DAVIS	8	1,116,357	378	4,172	62.2	31.6	0.5	0.8	2.5	2.4
RODRIGUEZ	9	1,094,924	371	4,544	4.8	0.5	0.6	0.2	0.2	93.8
MARTINEZ	10	1,060,159	359	4,903	5.3	0.5	0.6	0.5	0.2	92.9
HERNANDEZ	11	1,043,281	354	5,257	3.8	0.4	0.6	0.2	0.2	94.9
LOPEZ	12	874,523	296	5,553	4.9	0.6	1.0	0.4	0.3	92.9
GONZALEZ	13	841,025	285	5,838	4.0	0.4	0.4	0.1	0.1	95.0
WILSON	14	801,882	272	6,110	67.4	26.0	0.6	1.1	2.4	2.6
ANDERSON	15	784,404	266	6,376	75.2	18.9	0.6	0.7	2.1	2.4
THOMAS	16	756,142	256	6,632	52.6	38.8	2.4	1.1	2.6	2.5
TAYLOR	17	751,209	255	6,887	65.4	28.4	0.6	0.8	2.4	2.5
MOORE	18	724,374	246	7,132	66.4	27.7	0.5	0.7	2.3	2.3
JACKSON	19	708,099	240	7,373	39.9	53.0	0.4	1.1	3.1	2.5
MARTIN	20	702,625	238	7,611	74.8	15.8	0.9	1.0	2.0	5.6
LEE	21	693,023	235	7,846	36.0	16.3	42.2	1.0	2.6	1.9
PEREZ	22	681,645	231	8,077	5.0	0.5	1.2	0.2	0.3	93.0
THOMPSON	23	664,644	225	8,302	69.8	23.6	0.6	1.2	2.4	2.5
WHITE	24	660,491	224	8,526	65.5	28.2	0.5	1.1	2.4	2.4
HARRIS	25	624,252	212	8,738	51.4	42.4	0.5	0.7	2.8	2.3
SANCHEZ	26	612,752	208	8,945	5.0	0.5	1.0	0.4	0.2	93.0
CLARK	27	562,679	191	9,136	74.7	19.0	0.5	1.0	2.2	2.6
RAMIREZ	28	557,423	189	9,325	3.9	0.3	0.9	0.2	0.2	94.5
LEWIS	29	531,781	180	9,505	58.2	34.8	0.6	1.2	2.6	2.6
ROBINSON	30	529,821	180	9,685	48.7	44.9	0.5	0.5	2.8	2.6
WALKER	31	523,129	177	9,862	58.7	35.1	0.5	0.9	2.6	2.4
YOUNG	32	484,447	164	10,027	66.3	24.7	3.0	0.8	2.7	2.6
ALLEN	33	482,607	164	10,190	67.6	26.2	0.5	0.9	2.4	2.5
KING	34	465,422	158	10,348	70.2	22.8	1.2	1.0	2.3	2.6
WRIGHT	35	458,980	156	10,504	65.8	28.2	0.5	0.7	2.4	2.4
SCOTT	36	439,530	149	10,653	60.2	32.9	0.5	1.2	2.6	2.6
TORRES	37	437,813	148	10,801	5.4	0.6	1.4	0.2	0.3	92.2
NGUYEN	38	437,645	148	10,949	1.0	0.1	96.5	0.0	1.8	0.6
HILL	39	434,827	147	11,097	64.4	29.1	0.5	1.0	2.5	2.5
FLORES	40	433,969	147	11,244	4.9	0.4	2.1	0.3	0.4	91.9
GREEN	41	430,182	146	11,390	56.8	37.0	0.4	0.6	2.5	2.6
ADAMS	42	427,865	145	11,535	74.0	19.9	0.6	0.8	2.2	2.6
NELSON	43	424,958	144	11,679	77.7	16.0	0.7	1.2	2.0	2.6
BAKER	44	419,586	142	11,821	79.8	14.4	0.6	0.9	2.0	2.3
HALL	45	407,076	138	11,959	72.7	21.6	0.6	0.7	2.2	2.3
RIVERA	46	391,114	133	12,092	5.4	1.0	2.0	0.2	0.4	91.1
CAMPBELL	47	386,157	131	12,223	73.7	20.5	0.5	0.7	2.1	2.5
MITCHELL	48	384,486	130	12,353	61.0	32.5	0.5	1.0	2.5	2.5
CARTER	49	376,966	128	12,481	58.2	35.6	0.5	0.8	2.7	2.3
ROBERTS	50	376,774	128	12,608	77.3	16.7	0.6	0.9	2.2	2.4

Source: U.S. Census Bureau, 2010 Census.

Note: Hispanic origin and race groups are mutually exclusive.