

# Documentation: State variable longitudinal data-set [1980 – 2014]

Author: Michael Daly

Date: 2/12/2016

## About:

The State demographic panel data-set is an unbalanced panel of select variables at the state-level for the period from 1980 to 2014. This document applies to both the Excel (state\_series\_1980-2014.xlsx) and the Stata (state\_series\_1980-2014.dta) versions of this data-set.

## Important note regarding variable consistency:

Effort was taken to ensure that variables remain comparable across years. In cases where a variable experienced a definitional change, the variable is presented in two separate series. These instances are noted in the variable definitions along with links to resources that document the change and propose methods for bridging the separate series.

## Variables in this data-set were sourced from:

The Bureau of Labor Statistics Local Area Statistics Project

United States Census Bureau Small Area Income and Poverty Estimates

United States Census Bureau Population and Housing Estimates

## Overview of variables contained in the data-set:

Variable	Description
ageunder18	Percent of population under the age of 18
age18_24	Percent of population between 18 and 24 years old
age25_34	Percent of population between 25 and 34 years old
age35_44	Percent of population between 35 and 44 years old
age45_54	Percent of population between 45 and 54 years old
age55_64	Percent of population between 55 and 64 years old
age65	Percent of population over the age 65 or older
female	Percent of population that is female
fips	State FIPS code
median_hhinc	Estimate of median household income
per_am_ind_77	Percent of population that is American Indian (OMB 1977)
per_am_ind_97	Percent of population that is American Indian (OMB 1997)
per_asian_77	Percent of population that is Asian or Pacific Islander (OMB 1977)
per_asian_97	Percent of population that is Asian or Pacific Islander (OMB 1997)
per_black_77	Percent of population that is Black (OMB 1977)
per_black_97	Percent of population that is Black (OMB 1997)
per_hispanic_77	Percent of population that is Hispanic (OMB 1977)
per_hispanic_97	Percent of population that is Hispanic (OMB 1997)
per_mixed_97	Percent of population that is of two or more races (OMB 1997)
per_white_77	Percent of population that is White (OMB 1977)
per_white_97	Percent of population that is White (OMB 1997)
percent_poc0_5	Estimate of children under 5 years old in poverty
percent_pov	Estimate of the percentage of the population in poverty
percent_pov0_17	Estimate of children under 18 years old in poverty
percent_pov5_17	Estimate of related children ages 5 to 17 in families in poverty
state_code	State US Postal Service abbreviation
total_pop	State population
unemployment_rate	Unemployment rate
year	Year

### Variable definitions:

#### *“ageunder18”*

- a. Description: Percentage of the total state population under the age of 18 years. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

#### *“age18\_24”*

- a. Description: Percentage of the total state population aged 18 to 24 years. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

#### *“age25\_34”*

- a. Description: Percentage of the total state population aged 25 to 34 years. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

#### *“age35\_44”*

- a. Description: Percentage of the total state population aged 35 to 44 years. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

*“age45\_54”*

- a. Description: Percentage of the total state population aged 45 to 54 years.  
This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

*“age55\_64”*

- a. Description: Percentage of the total state population aged 55 to 64 years.  
This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

*“age65”*

- a. Description: Percentage of the total state population 65 years of age or older.  
This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

*“female”*

- a. Description: Percentage of the total state population that is female. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

*“fips”*

- a. Description: Federal Information Processing Standard state code. This variable uniquely identifies each state.
- b. Type: Two character string
- c. Missing observations: None
- d. Values: 01 through 56, continental 50 states exclude codes 03, 07, 14, 43 and 52

<b>Fips code:</b>	<b>State:</b>	<b>Fips code:</b>	<b>State:</b>
01	AL	30	MT
02	AK	31	NE
04	AZ	32	NV
05	AR	33	NH
06	CA	34	NJ
08	CO	35	NM
09	CT	36	NY
10	DE	37	NC
11	DC	38	ND
12	FL	39	OH
13	GA	40	OK
15	HI	41	OR
16	ID	42	PA
17	IL	44	RI
18	IN	45	SC
19	IA	46	SD
20	KS	47	TN
21	KY	48	TX
22	LA	49	UT
23	ME	50	VT
24	MD	51	VA
25	MA	53	WA
26	MI	54	WV
27	MN	55	WI
28	MS	56	WY
29	MO		

*“median\_hhinc”*

- a. Description: Estimate of the median household income. This data was obtained from the United States Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) for states. The estimates are not direct counts from enumerations. Instead, for states SAIPE models income and poverty estimates by combining multiple data sources.  
For a detailed overview as well as a cautionary note on comparing years see: <https://www.census.gov/did/www/saipe/index.html>
- b. Type: Numeric
- c. Missing observations: 663 / 1785. Data is available for 1989, 1993, and 1995 through 2014.
- d. Values. Positive real numbers

*“per\_am\_ind\_77”*

- a. Description: Percentage of the population that self-classified as American Indian. Prior to 2000 this was based on the 1977 Standard set by the Office of Management and Budget (OMB). See (Lee, 2001) for proposed methods of bridging the two series. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1977 OMB Standard is available here: <http://wonder.cdc.gov/wonder/help/populations/bridged-race/directive15.html>
- b. Type: Numeric
- c. Missing observations: 816/1785. This series is available for the period from 1981 to 1999.
- d. Values: Positive real numbers

*“per\_am\_ind\_97”*

- a. Description: Percentage of the population that self-classified as American Indian. From 2000 this is based on the 1997 Standard set by the Office of Management and Budget (OMB). See (Lee, 2001) for proposed methods of bridging the two series. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1997 OMB Standard is available here: [https://www.whitehouse.gov/omb/fedreg\\_1997standards](https://www.whitehouse.gov/omb/fedreg_1997standards)
- b. Type: Numeric
- c. Missing observations: 1020/1785. This series is available for the period from 2000 to 2014.
- d. Values: Positive real numbers

*“per\_asian\_77”*

- a. Description: Percentage of the population that self-classified as Asian or Pacific Islander. Prior to 2000 this was based on the 1977 Standard set by the Office of Management and Budget (OMB). See (Lee, 2001) for proposed methods of bridging the two series. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1977 OMB Standard is available here: <http://wonder.cdc.gov/wonder/help/populations/bridged-race/directive15.html>
- b. Type: Numeric
- c. Missing observations: 816/1785. This series is available for the period from 1981 to 1999.
- d. Values: Positive real numbers

*“per\_asian\_97”*

- a. Description: Percentage of the population that self-classified as either Asian or Pacific Islander. From 2000 this is based on the 1997 Standard set by the Office of Management and Budget (OMB). See (Lee, 2001) for proposed methods of bridging the two series. The 1997 standard separately defines Asian and Pacific Islander, for ease of comparison to the earlier standard these two categories were combined. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1997 OMB Standard is available here: [https://www.whitehouse.gov/omb/fedreg\\_1997standards](https://www.whitehouse.gov/omb/fedreg_1997standards)
- b. Type: Numeric
- c. Missing observations: 1020/1785. This series is available for the period from 2000 to 2014.
- d. Values: Positive real numbers

*“per\_black\_77”*

- a. Description: Percentage of the population that self-classified as Black. Prior to 2000 this was based on the 1977 Standard set by the Office of Management and Budget (OMB). See (Lee, 2001) for proposed methods of bridging the two series. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1977 OMB Standard is available here: <http://wonder.cdc.gov/wonder/help/populations/bridged-race/directive15.html>
- b. Type: Numeric
- c. Missing observations: 816/1785. This series is available for the period from 1981 to 1999.
- d. Values: Positive real numbers

*“per\_black\_97”*

- a. Description: Percentage of the population that self-classified as Black. From 2000 this is based on the 1997 Standard set by the Office of Management and Budget (OMB). This is a derived variable. See (Lee, 2001) for proposed methods of bridging the two series. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1997 OMB Standard is available here: [https://www.whitehouse.gov/omb/fedreg\\_1997standards](https://www.whitehouse.gov/omb/fedreg_1997standards)
- b. Type: Numeric
- c. Missing observations: 1020/1785. This series is available for the period from 2000 to 2014.
- d. Values: Positive real numbers

*“per\_hispanic\_77”*

- a. Description: Percentage of the population that identify as Hispanic. Hispanics can be of any race. There is a discontinuity in the trend of this variable between 1999 and 2000 which is likely the result of the change from the 1977 OMB standard to the 1997 OMB standard regarding racial classification. As such, this variable is presented in two separate series to emphasize that pre-2000 data may not be directly comparable to data from 2000 onwards. See (Lee, 2001) for proposed methods of bridging the two series. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric
- c. Missing observations: 816/1785. This series is available for the period from 1981 to 1999.
- d. Values: Positive real numbers

*“per\_hispanic\_97”*

- a. Description: Percentage of the population that identify as Hispanic. Hispanics can be of any race. There is a discontinuity in the trend of this variable between 1999 and 2000 which is likely the result of the change from the 1977 OMB standard to the 1997 OMB standard regarding racial classification. As such, this variable is presented in two separate series to emphasize that data from 2000 onwards may not be directly comparable to earlier years of data. See (Lee, 2001) for proposed methods of bridging the two series. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable.
- b. Type: Numeric
- c. Missing observations: 1020/1785. This series is available for the period from 2000 to 2014.



- d. Values: Positive real numbers

*“per\_mixed\_97”*

- a. Description: Percentage of the population that self-classified as two or more races. The 1997 OMB standard regarding racial classification allowed respondents to classify themselves as being of two or more races. Prior to 2000, respondents could only classify themselves as belonging to a single race. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. See (Lee, 2001) for proposed methods of bridging the two series. The 1997 OMB Standard is available here:  
[https://www.whitehouse.gov/omb/fedreg\\_1997standards](https://www.whitehouse.gov/omb/fedreg_1997standards)
- b. Type: Numeric
- c. Missing observations: 1020/1785. This series is available for the period from 2000 to 2014.
- d. Values: Positive real numbers

*“per\_white\_77”*

- a. Description: Percentage of the population that self-classified as White. Prior to 2000 this was based on the 1977 Standard set by the Office of Management and Budget (OMB). See (Lee, 2001) for proposed methods of bridging the two series. This is a derived variable. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1977 OMB Standard is available here:  
<http://wonder.cdc.gov/wonder/help/populations/bridged-race/directive15.html>
- b. Type: Numeric
- c. Missing observations: 816/1785. This series is available for the period from 1981 to 1999.
- d. Values: Positive real numbers

*“per\_white\_97”*

- a. Description: Percentage of the population that self-classified as White. From 2000 this is based on the 1997 Standard set by the Office of Management and Budget (OMB). This is a derived variable. See (Lee, 2001) for proposed methods of bridging the two series. Refer to the variable description for *“total\_pop”* for information regarding the source data used to construct this variable. The 1997 OMB Standard is available here:  
[https://www.whitehouse.gov/omb/fedreg\\_1997standards](https://www.whitehouse.gov/omb/fedreg_1997standards)
- b. Type: Numeric

- c. Missing observations: 1020/1785. This series is available for the period from 2000 to 2014.
- d. Values: Positive real numbers

*“percent\_pov”*

- a. Description: Estimate of the percent of a state’s population in poverty. This data was obtained from the United States Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) for States. The estimates are not direct counts from enumerations. Instead, for states SAIPE models income and poverty estimates by combining multiple data sources.  
For a detailed overview as well as a cautionary note on comparing years see: <https://www.census.gov/did/www/saipe/index.html>
- b. Type: Numeric
- c. Missing observations: 663 / 1785. Data is available for 1989, 1993, and 1995 through 2014.
- d. Values. Positive real numbers

*“percent\_pov0\_5”*

- a. Description: Estimate of the percent of children under 5 in poverty. This data was obtained from the United States Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) for states. The estimates are not direct counts from enumerations. Instead, for states SAIPE models income and poverty estimates by combining multiple data sources.  
For a detailed overview as well as a cautionary note on comparing years see: <https://www.census.gov/did/www/saipe/index.html>
- b. Type: Numeric
- c. Missing observations: 663 / 1785. Data is available for 1989, 1993, and 1995 through 2014.
- d. Values. Positive real numbers

*“percent\_pov0\_17”*

- a. Description: Estimate of the percent of children under 18 in poverty. This data was obtained from the United States Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) for states. The estimates are not direct counts from enumerations. Instead, for states SAIPE models income and poverty estimates by combining multiple data sources.  
For a detailed overview as well as a cautionary note on comparing years see: <https://www.census.gov/did/www/saipe/index.html>
- b. Type: Numeric
- c. Missing observations: 663 / 1785. Data is available for 1989, 1993, and 1995 through 2014.

- d. Values. Positive real numbers

*“percent\_pov\_5\_17”*

- a. Description: Estimate of the percent of related children ages 5 to 17 in families in poverty. This data was obtained from the United States Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) for states. The estimates are not direct counts from enumerations. Instead, for states SAIPE models income and poverty estimates by combining multiple data sources. For a detailed overview as well as a cautionary note on comparing years see: <https://www.census.gov/did/www/saipe/index.html>
- b. Type: Numeric
- c. Missing observations: 663 / 1785. Data is available for 1989, 1993, and 1995 through 2014.
- d. Values. Positive real numbers

*“state\_code”*

- a. Description: State US Postal Service abbreviation.
- b. Type: Two character string
- c. Missing observations: None
- d. Values: 51 (50 States and the District of Columbia). Refer to the variable description for *“fips”* for more details regarding these codes.

*“total\_pop”*

- a. Description: State level total population count. Census years (1980, 1990, 2000, and 2010) utilize the April 1 census count. Other years utilize the mid-year estimates for the state population. Population counts are for the resident population for years 1980 to 2011 and for the civilian population for the period 2012 to 2014.<sup>1</sup> These estimates are prepared by the United States Census Bureau and are available along with additional years of estimates here: <https://www.census.gov/popest/data/historical/>
- b. Type: Numeric (float)
- c. Missing observations: None
- d. Values: Positive real numbers

---

<sup>1</sup> The *Civilian Population* includes all U.S. residents not in the active-duty military. The *Resident Population* includes all persons who are “usually resident” in a specified geographic area (United States Census Bureau, 2013).

*“unemployment\_rate”*

- a. Description: Annual average rate of unemployment. This data was obtained from the Bureau of Labor Statistics Local Area Unemployment Statistics program. For more details regarding the definition of this variable see: <http://www.bls.gov/lau/>
- b. Type: Numeric
- c. Missing observations: None
- d. Values: Positive real numbers

*“year”*

- a. Description: Identifies the data year.
- b. Type: Four digit numeric
- c. Missing observations: None
- d. Values: 1980 through 2014

## References

- Lee, S. M. (2001, March). *Population Reference Bureau*. Retrieved February 10, 2016, from Using the New Racial Categories in the 2000 Census: [http://www.prb.org/pdf/census2000\\_usingnewracialprofiles.pdf](http://www.prb.org/pdf/census2000_usingnewracialprofiles.pdf)
- United States Census Bureau. (2013, July 17). *Population Estimates Terms and Definitions*. Retrieved February 10, 2016, from United States Census Bureau: <https://www.census.gov/popest/about/terms.html>