

Readme: State Income Distribution Data*

Kevin Rinz and John Voorheis

This version: July 10, 2023
Current version available [here](#)

This document describes data released alongside the paper "Re-examining Regional Income Convergence: A Distributional Approach," available [here](#).¹ Three data files provide measures related to the analysis presented in that paper for all states.

Measures are generated using the universe of 1040 tax returns with positive values of the income concepts in question. Percentiles are constructed to represent one percent of people (rather than tax returns) in each group. Each record is assigned its tax unit's value of each income measure.

Files are available [here](#). The files are

- `state_curves.csv`: This file provides ordinates of absolute and relative regional income divergence curves (see Section 4.2.1) for each year from 1998–2019 and for all states. In addition to curves for the full population of each state, it provides curves for subgroups defined by gender (male, female) and race/ethnicity (non-Hispanic white alone; non-Hispanic Black alone; non-Hispanic Asian alone, including Native Hawaiians and other Pacific Islanders; and Hispanic). Only curves for which every ordinate is based on at least 30 observations are reported.
- `state_indexes.csv`: This file contains aggregations of the absolute and relative regional income divergence curves using various Gini weights (see Section 4.2.2). These regional income divergence indexes were produced from the estimates in `state_curves.csv` after they were disclosed, so they are only available for groups for which absolute and relative regional income divergence curves are available.
- `nat_dists.csv`: This file provides information about the full-population national income distribution in each year. Estimates from `state_curves.csv` can be combined with estimates in this file to produce absolute measures of income in dollars or percentiles of the national income distribution.

The tables below describe the variables available in each file and explain what their values represent.

* All estimates have been reviewed to ensure that no confidential information is disclosed. Disclosure Review Board release authorization number CBDRB-FY23-0136. Rinz: kevin.rinz@census.gov; Voorheis: john.l.voorheis@census.gov.

¹The content of this document also appears in Appendix B of that paper.

Table 1: Data dictionaries for state_curves.csv

Variable name and description	Value	Value definition
file: how are subgroups within states defined?	all gender racegrp	Estimates for the full population regardless of race/gender Estimates for groups defined by gender Estimates for groups defined by race/ethnicity
inctype: what income concept is used?	AGI WSI	Estimates are based on tax unit adjusted gross income Estimates are based on tax unit wage and salary income
year: what year are estimates for?	1998–2019	Estimates are based on data from this tax year
state: what state are estimates for?	postal codes for 50 states and DC	Estimates are based on data from this state
groupid: what subgroup are estimates for?	0 1 2 4 6	For file = all, 0 = full population For file = gender, 1 = male For file = racegrp, 1 = non-Hispanic white For file = gender, 2 = female For file = racegrp, 2 = non-Hispanic Black For file = racegrp, 4 = non-Hispanic Asian (including native Hawaiian and other Pacific islanders) For file = racegrp, 6 = Hispanic
pct: what percentile are estimates for?	1–100	Estimates are based on data from this percentile
rridc: what is the average of $F_N(q_n(p)) - p$ for people in this year-state-group-percentile?	[min,max]	Average value of the relative regional income divergence curve (RRIDC) within state-group-percentile
aridc: what is the average of $q_n(p) - q_N(p)$ for people in this year-state-group-percentile?	[min,max]	Average value of the absolute regional income divergence curve (ARIDC) within state-group-percentile in nominal dollars

Table 2: Data dictionary for state_indexes.csv

Variable name and description	Value	Value definition
file: how are subgroups within states defined?	all gender racegrp	Estimates for the full population regardless of race/gender Estimates for groups defined by gender Estimates for groups defined by race/ethnicity
inctype: what income concept is used?	AGI WSI	Estimates are based on tax unit adjusted gross income Estimates are based on tax unit wage and salary income
year: what year are estimates for?	1998–2019	Estimates are based on data from this tax year
state: what state are estimates for?	postal codes for 50 states and DC	Estimates are based on data from this state
groupid: what subgroup are estimates for?	0 1 2 4 6	For file = all, 0 = full population For file = gender, 1 = male For file = racegrp, 1 = non-Hispanic white For file = gender, 2 = female For file = racegrp, 2 = non-Hispanic Black For file = racegrp, 4 = non-Hispanic Asian (including native Hawaiian and other Pacific islanders) For file = racegrp, 6 = Hispanic
nu: what value of v is used to create the Gini weights used to construct these estimates?	1–5	Gini weights are constructed using this value of v
adi: what is the average value of the ARIDC within this year-state-group?	[min,max]	Average value of the ARIDC, in thousands of nominal dollars
rdi: what is the average value of the ARIDC within this year-state-group?	[min,max]	Average value of the RRIDC

Table 3: Data dictionary for nat_dists

Variable name and description	Value	Value definition
measure: what income concept is used?	AGI WSI	Estimates are based on tax unit adjusted gross income Estimates are based on tax unit wage and salary income
year: what year are estimates for?	1998–2019	Estimates are based on data from this tax year
pct: what percentile are estimates for?	1–100	Estimates are based on data from this percentile
meaninc: what is the average income within this percentile?	[min,max]	Average income within this percentile in nominal dollars
cutoff: what is the maximum income value within this percentile?	[min,max]	Income value y such that pct % of people have income less than or equal to y , in nominal dollars