**Codebook for Bias-Corrected Income Segregation Dataset, 1970-2014**

Data Version 1.0, Released January 2018

See: Reardon, S.F., Bischoff, K., Owens, A., & Townsend, J.B. (2018). *Has Income Segregation Really Increased? Bias and Bias Correction in Sample-Based Segregation Estimates* (CEPA Working Paper No.18-02). Retrieved from Stanford Center for Education Policy Analysis: <http://cepa.stanford.edu/content/has-income-segregation-really-increased-bias-and-bias-correction-sample-based-segregation-estimates>

|  |  |  |
| --- | --- | --- |
| Var Name | Variable Label | Notes |
| metro | Metro Area (2003 OMB CBSA or Div Code) | Definitions [here](https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/2003/historical-delineation-files/0312mfips.txt) |
| group | Group | hh = households  fam = families  hnc = households without children  hwc = family households with children  w = white families (all, not non-Hispanic)  b = black families (all, not non-Hispanic)  h = Hispanic families |
| mettot | Total N of observations in metro (group-specific) |  |
| year | Census or ACS year | String var, e.g. ACS 2005-09=0509 |
| n\_tracts | N of tracts in metro w/ population | N of tracts underlying income segregation estimate |
| metname | Metro area/Division name (2003 OMB) | Definitions [here](https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/2003/historical-delineation-files/0312mfips.txt) |
| cbsacode | CBSA Code, 2003 OMB | CBSA, not division, code; Definitions [here](https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/2003/historical-delineation-files/0312mfips.txt) |
| cbsaname | CBSA Name, 2003 OMB | Definitions [here](https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/2003/historical-delineation-files/0312mfips.txt) |
| cbsalevl | CBSA Level (1=metro; 2=micro) | All observations are metros (=1) |
| rank2010 | Metro rank by population 2010 | Source: Census 2010 |
| rank2007 | Metro rank by population 2005-09 ACS | Source: ACS 2005-09 |
| rank2000 | Metro rank by population 2000 | Source: Census 2000 |
| O\_ASR16\_metros | Metros in Owens, ASR 2016 (N=100; rank2010<101) | rank2010<101; [Owens 2016](http://journals.sagepub.com/doi/full/10.1177/0003122416642430), |
| RB\_AJS11\_metros | etros in Reardon+Bischoff, AJS 2011 (N=100; rank2000<101) | rank2000<101; [Reardon & Bischoff 2011](http://www.journals.uchicago.edu/doi/full/10.1086/657114) |
| BR\_RSF14\_metros | Metros in Bischoff+Reardon, RSF 2014 (N=116; pop2007>500k & no Cape Coral) | pop2007>500000 & metro!=15980; [Bischoff & Reardon 2014](https://www.russellsage.org/sites/all/files/logan/logan_diversity_chapter7.pdf) |
| RB\_CEPA16\_metros | Metros in Reardon+Bischoff, CEPA, 2016 (N=116; pop2007>500k, no Cape Coral) | pop2007>500000 & metro!=15980; [Reardon & Bischoff 2016](https://cepa.stanford.edu/sites/default/files/the%20continuing%20increase%20in%20income%20segregation%20march2016.pdf) |
| O\_ASR16\_obs | Metro-group-year obs in Owens, ASR 2016 | See [Owens 2016](http://journals.sagepub.com/doi/full/10.1177/0003122416642430) |
| RB\_AJS11\_obs | Metro-group-year obs in Reardon+Bischoff, AJS 2011 | See [Reardon & Bischoff 2011](http://www.journals.uchicago.edu/doi/full/10.1086/657114) |
| BR\_RSF14\_obs | Metro-group-year obs in Bischoff+Reardon, RSF 2014 | See [Bischoff & Reardon 2014](https://www.russellsage.org/sites/all/files/logan/logan_diversity_chapter7.pdf) |
| RB\_CEPA16\_obs | Metro-group-year obs in Reardon+Bischoff, CEPA, 2016 | See [Reardon & Bischoff 2016](https://cepa.stanford.edu/sites/default/files/the%20continuing%20increase%20in%20income%20segregation%20march2016.pdf) |
| pop2010 | Metro population 2010 | Source: Census 2010 |
| pop2007 | Metro population 2005-09 | Source: ACS 2005-09 |
| pop2000 | Metro population 2000 | Source: Census 2000 |
| year\_n | Census or ACS year | numeric, midpoint for ACS (e.g., 2005-09=2007) |
| h10 | Uncorrected H, 10th pctile | Estimated using rankseg v 3.0 |
| hadj10 | Corrected H, 10th pctile | Estimated using rankseg v 3.0 |
| r10 | Uncorrected R, 10th pctile | Estimated using rankseg v 3.0 |
| radj10 | Corrected R, 10th pctile | Estimated using rankseg v 3.0 |
| h50 | Uncorrected H, 50th pctile | Estimated using rankseg v 3.0 |
| hadj50 | Corrected H, 50th pctile | Estimated using rankseg v 3.0 |
| r50 | Uncorrected R, 50th pctile | Estimated using rankseg v 3.0 |
| radj50 | Corrected R, 50th pctile | Estimated using rankseg v 3.0 |
| h90 | Uncorrected H, 90th pctile | Estimated using rankseg v 3.0 |
| hadj90 | Corrected H, 90th pctile | Estimated using rankseg v 3.0 |
| r90 | Uncorrected R, 90th pctile | Estimated using rankseg v 3.0 |
| radj90 | Corrected R, 90th pctile | Estimated using rankseg v 3.0 |
| h | Uncorrected Rank-Order H | Estimated using rankseg v 3.0 |
| hadj | Corrected Rank-Order H | Estimated using rankseg v 3.0 |
| r | Uncorrected Rank-Order R | Estimated using rankseg v 3.0 |
| radj | Corrected Rank-Order R | Estimated using rankseg v 3.0 |
| tmean | Mean tract pop in metro | Mean population per tract in metro area. We recommend limiting analyses to tmean>=200 |
| nmean | Mean tract pop in metro, unweighted | Mean unweighted (sample) population per tract in metro area |
| hrate | Mean tract HU sampling rate in metro | Mean housing unit sampling rate in metro area (not group specific) |
| t1harm | Harmonic mean of tract pop in metro | Harmonic mean of tract population-1 in metro area |
| longsamp | 10k pop in 70-00, used by R+B | Indicator for metros with at least 10k population in 1970-2000, used for race-specific results |

**Notes and Errata**

* Unique observations are metro-year-group
* 1970, 1980, 1990, and 2000 data come from Decennial Censuses; 0509 (2007), 0610 (2008), 0711 (2009), 0812 (2010), 0913 (2011), 1014 (2012), 1115 (2013), and 1216 (2014) data come from the 5-year aggregate American Community Surveys
* Tract-level income data downloaded via Social Explorer (SE), National Historical Geographic Information System (NHGIS), or the Neighborhood Change Database (NCDB)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | hh | fam | hnc | hwc | w | b | h | Sample  Counts |
| 1970 | -- | NHGIS | -- | -- | NHGIS | NHGIS | -- | SE |
| 1980 | SE | NCDB | -- | -- | NCDB | NCDB | NCDB | SE |
| 1990 | SE | SE | SE | NHGIS | NHGIS | NHGIS | NHGIS | SE |
| 2000 | SE | SE | SE | SE | NHGIS | NHGIS | NHGIS | SE |
| 2007 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2008 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2009 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2010 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2011 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2012 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2013 | SE | SE | SE | SE | SE | SE | SE | SE |
| 2014 | SE | SE | SE | SE | SE | SE | SE | SE |

* Sample is metropolitan areas using 2003 OMB definitions for CBSAs and their divisions. N=380 unique metropolitan areas (all metros in the U.S. excluding Puerto Rico)
  + N=379 metros in 1980. Barnstable County (metro=12700) is missing tract level data on unweighted sample counts in both NHGIS and Social Explorer and is therefore dropped
  + N=279 in 1970 because not all metros are tracted. Cape Coral (metro 15980) is missing income data in 1970 and thus excluded from longitudinal analyses of large metro areas
  + N varies for black (group = b) and Hispanic (group = h) families across years. We only estimate income segregation in metros with at least 1,000 families of each group.
* We provide flags indicating the samples of the four papers we replicate in our study (Reardon & Bischoff 2011, 2016; Bischoff & Reardon 2014; Owens 2010)
  + Flags ending in \_metros are metro-level without regard to years and groups in the papers.
  + Flags ending in \_obs are metro-group-year-level, identifying the observations in the papers.
* To replicate Reardon et al. 2018, use flags:
  + Table 1, Figs 5-7, Table C1: if BR\_RSF14\_metros==1
  + Table 2:
    - Panel 1: if BR\_RSF14\_metros==1
    - Panel 2-4: if O\_ASR16\_metros ==1
  + Table C2, Fig C1:
    - Create variable for tmean>=200 in all years

gen sample\_200=1 if tmean>=200 & tmean<.

bys metro group: egen sample\_total=sum(sample\_200)

* + - Use if BR\_RSF14\_metros==1 & sample\_total==3 & longsamp==1
* H10, H50, and H90 are not reported by race or for group = hnc or hwc in Version 1 of this dataset