

# Package ‘ndi’

December 1, 2022

**Title** Neighborhood Deprivation Indices

**Version** 0.1.3

**Date** 2022-12-01

**Maintainer** Ian D. Buller <ian.buller@alumni.emory.edu>

**Description** Computes various metrics of socio-economic deprivation and disparity in the United States. Some metrics are considered “spatial” because they consider the values of neighboring (i.e., adjacent) census geographies in their computation, while other metrics are “aspatial” because they only consider the value within each census geography. Two types of aspatial neighborhood deprivation indices (NDI) are available: including: (1) based on Messer et al. (2006) <doi:10.1007/s11524-006-9094-x> and (2) based on Andrews et al. (2020) <doi:10.1080/17445647.2020.1750066> and Slotman et al. (2022) <doi:10.1016/j.dib.2022.108002> who use variables chosen by Roux and Mair (2010) <doi:10.1111/j.1749-6632.2009.05333.x>. Both are a decomposition of multiple demographic characteristics from the U.S. Census Bureau American Community Survey 5-year estimates (ACS-5; 2006-2010 onward). Using data from the ACS-5 (2005-2009 onward), the package can also (1) compute the spatial Racial Isolation Index (RI) based on Anthopolos et al. (2011) <doi:10.1016/j.sste.2011.06.002>, (2) compute the spatial Educational Isolation Index (EI) based on Bravo et al. (2021) <doi:10.3390/ijerph18179384>, (3) compute the aspatial Index of Concentration at the Extremes (ICE) based on Feldman et al. (2015) <doi:10.1136/jech-2015-205728> and Krieger et al. (2016) <doi:10.2105/AJPH.2015.302955>, (4) compute the aspatial Dissimilarity Index based on Duncan & Duncan (1955) <doi:10.2307/2088328>, and (5) retrieve the aspatial Gini Index based on Gini (1921) <doi:10.2307/2223319>.

**License** Apache License (>= 2.0)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.1

**Depends** R (>= 3.5.0)

**Imports** dplyr, MASS, Matrix, psych, sf, stats, stringr, tidycensus, tidyr, utils

**Suggests** ggplot2, testthat, tigris, R.rsp, spelling, usethis  
**VignetteBuilder** R.rsp  
**Language** en-US  
**URL** <https://github.com/idblr/ndi>  
**BugReports** <https://github.com/idblr/ndi/issues>  
**NeedsCompilation** no  
**Author** Ian D. Buller [aut, cre, cph] (<<https://orcid.org/0000-0001-9477-8582>>),  
NCI [cph, fnd]  
**Repository** CRAN  
**Date/Publication** 2022-12-01 15:20:02 UTC

**R topics documented:**

ndi-package . . . . .	2
anthopolos . . . . .	4
bravo . . . . .	6
DCtracts2020 . . . . .	8
duncan . . . . .	9
gini . . . . .	12
krieger . . . . .	13
messer . . . . .	15
powell_wiley . . . . .	17
<b>Index</b>	<b>21</b>

---

ndi-package	<i>The ndi Package: Neighborhood Deprivation Indices</i>
-------------	--

---

**Description**

Computes various metrics of socio-economic deprivation and disparity in the United States based on information available from the U.S. Census Bureau.

**Details**

The 'ndi' package computes various metrics of socio-economic deprivation and disparity in the United States. Some metrics are considered "spatial" because they consider the values of neighboring (i.e., adjacent) census geographies in their computation, while other metrics are "aspatial" because they only consider the value within each census geography. Two types of aspatial neighborhood deprivation indices (NDI) are available: (1) based on Messer et al. (2006) [doi:10.1007/s115240069094x](https://doi.org/10.1007/s115240069094x) and (2) based on Andrews et al. (2020) [doi:10.1080/17445647.2020.1750066](https://doi.org/10.1080/17445647.2020.1750066) and Slotman et al. (2022) [doi:10.1016/j.dib.2022.108002](https://doi.org/10.1016/j.dib.2022.108002) who use variables chosen by Roux and Mair (2010) [doi:10.1111/j.17496632.2009.05333.x](https://doi.org/10.1111/j.17496632.2009.05333.x). Both are a decomposition of multiple demographic

characteristics from the U.S. Census Bureau American Community Survey 5-year estimates (ACS-5; 2006-2010 onward). Using data from the ACS-5 (2005-2009 onward), the package can also (1) compute the spatial Racial Isolation Index (RI) based on Anthopolos et al. (2011) [doi:10.1016/j.sste.2011.06.002](https://doi.org/10.1016/j.sste.2011.06.002), (2) compute the spatial Educational Isolation Index (EI) based on Bravo et al. (2021) [doi:10.3390/ijerph18179384](https://doi.org/10.3390/ijerph18179384), (3) compute the aspatial Index of Concentration at the Extremes (ICE) based on Feldman et al. (2015) [doi:10.1136/jech2015205728](https://doi.org/10.1136/jech2015205728) and Krieger et al. (2016) [doi:10.2105/AJPH.2015.302955](https://doi.org/10.2105/AJPH.2015.302955), (4) compute the aspatial Dissimilarity Index based on Duncan & Duncan (1955) [doi:10.2307/2088328](https://doi.org/10.2307/2088328), and (5) retrieve the aspatial Gini Index based on Gini (1921) [doi:10.2307/2223319](https://doi.org/10.2307/2223319).

Key content of the 'ndi' package include:

### Metrics of Socio-Economic Deprivation and Disparity

**anthopolos** Computes the spatial Racial Isolation Index (RI) based on Anthopolos (2011) [doi:10.1016/j.sste.2011.06.002](https://doi.org/10.1016/j.sste.2011.06.002).

**bravo** Computes the spatial Educational Isolation Index (EI) based on Bravo (2021) [doi:10.3390/ijerph18179384](https://doi.org/10.3390/ijerph18179384).

**duncan** Computes the aspatial Dissimilarity Index based on Duncan & Duncan (1955) [doi:10.2307/2088328](https://doi.org/10.2307/2088328).

**gini** Retrieves the aspatial Gini Index based on Gini (1921) [doi:10.2307/2223319](https://doi.org/10.2307/2223319).

**krieger** Computes the aspatial Index of Concentration at the Extremes based on Feldman et al. (2015) [doi:10.1136/jech2015205728](https://doi.org/10.1136/jech2015205728) and Krieger et al. (2016) [doi:10.2105/AJPH.2015.302955](https://doi.org/10.2105/AJPH.2015.302955).

**messer** Computes the aspatial Neighborhood Deprivation Index (NDI) based on Messer et al. (2006) [doi:10.1007/s115240069094x](https://doi.org/10.1007/s115240069094x).

**powell\_wiley** Computes the aspatial Neighborhood Deprivation Index (NDI) based on Andrews et al. (2020) [doi:10.1080/17445647.2020.1750066](https://doi.org/10.1080/17445647.2020.1750066) and Slotman et al. (2022) [doi:10.1016/j.dib.2022.108002](https://doi.org/10.1016/j.dib.2022.108002) who use variables chosen by Roux and Mair (2010) [doi:10.1111/j.17496632.2009.05333.x](https://doi.org/10.1111/j.17496632.2009.05333.x).

### Pre-formatted U.S. Census Data

**Dctracts2020** A sample dataset containing information about U.S. Census American Community Survey 5-year estimate data for the District of Columbia census tracts (2020). The data are obtained from the **get\_acs** function and formatted for the **messer** and **powell\_wiley** functions input.

### Dependencies

The 'ndi' package relies heavily upon **tidycensus** to retrieve data from the U.S. Census Bureau American Community Survey five-year estimates and the **psych** for computing the neighborhood deprivation indices. The **messer** function builds upon code developed by Hruska et al. (2022) [doi:10.17605/OSF.IO/M2SAV](https://doi.org/10.17605/OSF.IO/M2SAV) by fictionalizing, adding the percent of households earning <\$30,000 per year to the NDI computation, and providing the option for computing the ACS-5 2006-2010 NDI values. There is no code companion to compute NDI included in Andrews et al. (2020) [doi:10.1080/17445647.2020.1750066](https://doi.org/10.1080/17445647.2020.1750066) or Slotman et al. (2022) [doi:10.1016/j.dib.2022.108002](https://doi.org/10.1016/j.dib.2022.108002), but the package author worked directly with the Slotman et al. (2022) [doi:10.1016/j.dib.2022.108002](https://doi.org/10.1016/j.dib.2022.108002) authors to replicate their SAS code in R. The spatial metrics RI and EI rely on the **sf** and **Matrix** packages to compute the geospatial adjacency matrix between census geographies. There is no code companion to compute RI, EI, or DI included in Anthopolos et al. (2011) [doi:10.1016/j.sste.2011.06.002](https://doi.org/10.1016/j.sste.2011.06.002), Bravo et al. (2021) [doi:10.3390/ijerph18179384](https://doi.org/10.3390/ijerph18179384), or Duncan & Duncan (1955) [doi:10.2307/2088328](https://doi.org/10.2307/2088328), respectively.

**Author(s)**

Ian D. Buller

*Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland, USA.*

Maintainer: I.D.B. <ian.buller@alumni.emory.edu>

---

anthopolos

*Racial Isolation Index based on Anthopolos et al. (2011)*


---

**Description**

Compute the Racial Isolation Index (Anthopolos) values for a selected subgroup(s).

**Usage**

```
anthopolos(geo = "tract", year = 2020, subgroup, quiet = FALSE, ...)
```

**Arguments**

geo	Character string specifying the geography of the data either census tracts geo = "tract" (the default) or counties geo = "county".
year	Numeric. The year to compute the estimate. The default is 2020, and the years 2009 onward are currently available.
subgroup	Character string specifying the racial/ethnic subgroup(s). See Details for available choices.
quiet	Logical. If TRUE, will display messages about potential missing census information. The default is FALSE.
...	Arguments passed to <a href="#">get_acs</a> to select state, county, and other arguments for census characteristics

**Details**

This function will compute the Racial Isolation Index (RI) of U.S. census tracts or counties for a specified geographical extent (e.g., the entire U.S. or a single state) based on Anthopolos et al. (2011) [doi:10.1016/j.sste.2011.06.002](#) who originally designed the metric for the racial isolation of non-Hispanic Black individuals. This function provides the computation of RI for any of the U.S. Census Bureau race/ethnicity subgroups (including Hispanic and non-Hispanic individuals).

The function uses the [get\\_acs](#) function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for the geospatial computation. The yearly estimates are available for 2009 onward when ACS-5 data are available but are available from other U.S. Census Bureau surveys. The twenty racial/ethnic subgroups (U.S. Census Bureau definitions) are:

- B03002\_002: not Hispanic or Latino "NHoL"
- B03002\_003: not Hispanic or Latino, white alone "NHoLW"

- B03002\_004: not Hispanic or Latino, Black or African American alone "NHoLB"
- B03002\_005: not Hispanic or Latino, American Indian and Alaska Native alone "NHoLA-  
IAN"
- B03002\_006: not Hispanic or Latino, Asian alone "NHoLA"
- B03002\_007: not Hispanic or Latino, Native Hawaiian and Other Pacific Islander alone  
"NHoLNHOPI"
- B03002\_008: not Hispanic or Latino, Some other race alone "NHoLSOR"
- B03002\_009: not Hispanic or Latino, Two or more races "NHoLTOMR"
- B03002\_010: not Hispanic or Latino, Two races including Some other race "NHoLTRiSOR"
- B03002\_011: not Hispanic or Latino, Two races excluding Some other race, and three or more  
races "NHoLTReSOR"
- B03002\_012: Hispanic or Latino "HoL"
- B03002\_013: Hispanic or Latino, white alone "HoLW"
- B03002\_014: Hispanic or Latino, Black or African American alone "HoLB"
- B03002\_015: Hispanic or Latino, American Indian and Alaska Native alone "HoLAIAN"
- B03002\_016: Hispanic or Latino, Asian alone "HoLA"
- B03002\_017: Hispanic or Latino, Native Hawaiian and Other Pacific Islander alone "HoLNHOPI"
- B03002\_018: Hispanic or Latino, Some other race alone "HoLSOR"
- B03002\_019: Hispanic or Latino, Two or more races "HoLTOMR"
- B03002\_020: Hispanic or Latino, Two races including Some other race "HoLTRiSOR"
- B03002\_021: Hispanic or Latino, Two races excluding Some other race, and three or more  
races "HoLTReSOR"

Use the internal state and county arguments within the `get_acs` function to specify geographic extent of the data output. NOTE: Current version does not correct for edge effects (e.g., census geographies along the specified spatial extent border, coastline, or U.S.-Mexico / U.S.-Canada border) may have few neighboring census geographies, and RI values in these census geographies may be unstable. A stop-gap solution for the former source of edge effect is to compute the RI for neighboring census geographies (i.e., the states bordering a study area of interest) and then use the estimates of the study area of interest.

A census geography (and its neighbors) that has nearly all of its population who identify with the specified race/ethnicity subgroup(s) (e.g., non-Hispanic or Latino, Black or African American alone) will have an RI value close to 1. In contrast, a census geography (and its neighbors) that is nearly none of its population who identify with the specified race/ethnicity subgroup(s) (e.g., not non-Hispanic or Latino, Black or African American alone) will have an RI value close to 0.

## Value

An object of class 'list'. This is a named list with the following components:

`ri` An object of class 'tbl' for the GEOID, name, RI, and raw census values of specified census geographies.

`missing` An object of class 'tbl' of the count and proportion of missingness for each census variable used to compute the RI.

See Also

[get\\_acs](#) for additional arguments for geographic extent selection (i.e., state and county).

Examples

```
## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Tract-level metric (2020)
anthopolos(geo = "tract", state = "GA", year = 2020, subgroup = c("NHoLB", "HoLB"))

# County-level metric (2020)
anthopolos(geo = "county", state = "GA", year = 2020, subgroup = c("NHoLB", "HoLB"))

## End(Not run)
```

---

bravo	<i>Educational Isolation Index based on Bravo et al. (2021)</i>
-------	---

---

Description

Compute the Educational Isolation Index (Bravo) values for selected educational attainment category(ies).

Usage

```
bravo(geo = "tract", year = 2020, subgroup, quiet = FALSE, ...)
```

Arguments

geo	Character string specifying the geography of the data either census tracts <code>geo = "tract"</code> (the default) or counties <code>geo = "county"</code> .
year	Numeric. The year to compute the estimate. The default is 2020, and the years 2009 onward are currently available.
subgroup	Character string specifying the educational attainment category(ies). See Details for available choices.
quiet	Logical. If TRUE, will display messages about potential missing census information. The default is FALSE.
...	Arguments passed to <a href="#">get_acs</a> to select state, county, and other arguments for census characteristics

## Details

This function will compute the Educational Isolation Index (EI) of U.S. census tracts or counties for a specified geographical extent (e.g., the entire U.S. or a single state) based on Bravo et al. (2021) [doi:10.3390/ijerph18179384](https://doi.org/10.3390/ijerph18179384) who originally designed the metric for the educational isolation of individual without a college degree. This function provides the computation of EI for any of the U.S. Census Bureau educational attainment levels.

The function uses the [get\\_acs](#) function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for the geospatial computation. The yearly estimates are available for 2009 onward when ACS-5 data are available but are available from other U.S. Census Bureau surveys. The five educational attainment levels (U.S. Census Bureau definitions) are:

- B06009\_002: Less than high school graduate "LtHS"
- B06009\_003: High school graduate (includes equivalency) "HSGiE"
- B06009\_004: Some college or associate's degree "SCoAD"
- B06009\_005: Bachelor's degree "BD"
- B06009\_006: Graduate or professional degree "GoPD"

Note: If year = 2009, then the ACS-5 data (2005-2009) are from the "B15002" question.

Use the internal state and county arguments within the [get\\_acs](#) function to specify geographic extent of the data output. NOTE: Current version does not correct for edge effects (e.g., census geographies along the specified spatial extent border, coastline, or U.S.-Mexico / U.S.-Canada border) may have few neighboring census geographies, and EI values in these census geographies may be unstable. A stop-gap solution for the former source of edge effect is to compute the EI for neighboring census geographies (i.e., the states bordering a study area of interest) and then use the estimates of the study area of interest.

A census geography (and its neighbors) that has nearly all of its population with the specified educational attainment category (e.g., a Bachelor's degree or more) will have an EI value close to 1. In contrast, a census geography (and its neighbors) that is nearly none of its population with the specified educational attainment category (e.g., less than a Bachelor's degree) will have an EI value close to 0.

## Value

An object of class 'list'. This is a named list with the following components:

**ei** An object of class 'tbl' for the GEOID, name, EI, and raw census values of specified census geographies.

**missing** An object of class 'tbl' of the count and proportion of missingness for each census variable used to compute the EI.

## See Also

[get\\_acs](#) for additional arguments for geographic extent selection (i.e., state and county).

## Examples

```
## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Tract-level metric (2020)
bravo(geo = "tract", state = "GA", year = 2020, subgroup = c("LtHS", "HSGiE"))

# County-level metric (2020)
bravo(geo = "county", state = "GA", year = 2020, subgroup = c("LtHS", "HSGiE"))

## End(Not run)
```

---

DCtracts2020	<i>Formatted U.S. Census American Community Survey 5-year estimate data for DC census tracts (2020) from the 'tidycensus' package</i>
--------------	---

---

## Description

A sample dataset containing information about U.S. Census American Community Survey 5-year estimate data for the District of Columbia census tracts (2020). The data are obtained from the [get\\_acs](#) function and formatted for the [messer](#) and [powell\\_wiley](#) functions input.

## Usage

```
DCtracts2020
```

## Format

A data frame with 206 rows and 23 variables:

**GEOID** census tract ID

**TotalPop** arcsinh-transformed CD3

**OCC** percent males in management, science, and arts occupation

**CWD** percent of crowded housing

**POV** percent of households in poverty

**FHH** percent of female headed households with dependents

**PUB** percent of households on public assistance

**U30** percent of households earning <\$30,000 per year

**EDU** percent earning less than a high school education

**EMP** percent unemployed

**logMedHHInc** median household income (dollars), natural log-transformed

**PctNoIDRZ** percent of households receiving dividends, interest, or rental income, Z-transformed



**PctPubAsstZ** percent of households receiving public assistance, Z-transformed

**logMedHomeVal** median home value (dollars), natural log-transformed

**PctWorkClassZ** percent in a management, business, science, or arts occupation, Z-transformed

**PctFemHeadKidsZ** percent of households that are female headed with any children under 18 years, Z-transformed

**PctNotOwnerOccZ** percent of housing units that are owner occupied, Z-transformed

**PctNoPhoneZ** percent of households without a telephone, Z-transformed

**PctNComPlmbZ** percent of households without complete plumbing facilities, Z-transformed

**PctEducLTHSZ** percent with a high school degree or higher (population 25 years and over), Z-transformed

**PctEducLTBchZ** percent with a college degree or higher (population 25 years and over), Z-transformed

**PctFamBelowPovZ** percent of families with incomes below the poverty level, Z-transformed

**PctUnemplZ** percent unemployed, Z-transformed

## Source

<https://github.com/idblr/ndi/blob/master/README.md>

## Examples

```
head(DCtracts2020)
```

---

duncan

*Dissimilarity Index based on Duncan & Duncan (1955)*

---

## Description

Compute the Dissimilarity Index (Duncan) values for a selected subgroup(s) and U.S. geographies

## Usage

```
duncan(
  geo_large = "county",
  geo_small = "tract",
  year = 2020,
  subgroup,
  subgroup_ref,
  quiet = FALSE,
  ...
)
```

## Arguments

<code>geo_large</code>	Character string specifying the larger geographical unit of the data. The default is counties <code>geo_large = "county"</code> .
<code>geo_small</code>	Character string specifying the smaller geographical unit of the data. The default is census tracts <code>geo_large = "tract"</code> .
<code>year</code>	Numeric. The year to compute the estimate. The default is 2020, and the years 2009 onward are currently available.
<code>subgroup</code>	Character string specifying the racial/ethnic subgroup(s) as the comparison population. See Details for available choices.
<code>subgroup_ref</code>	Character string specifying the racial/ethnic subgroup(s) as the reference population. See Details for available choices.
<code>quiet</code>	Logical. If TRUE, will display messages about potential missing census information. The default is FALSE.
<code>...</code>	Arguments passed to <code>get_acs</code> to select state, county, and other arguments for census characteristics

## Details

This function will compute the Dissimilarity Index (DI) of selected racial/ethnic subgroups and U.S. geographies for a specified geographical extent (e.g., the entire U.S. or a single state) based on Duncan & Duncan (1955) [doi:10.2307/2088328](https://doi.org/10.2307/2088328). This function provides the computation of DI for any of the U.S. Census Bureau race/ethnicity subgroups (including Hispanic and non-Hispanic individuals).

The function uses the `get_acs` function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for the geospatial computation. The yearly estimates are available for 2009 onward when ACS-5 data are available but are available from other U.S. Census Bureau surveys. The twenty racial/ethnic subgroups (U.S. Census Bureau definitions) are:

- B03002\_002: not Hispanic or Latino "NHoL"
- B03002\_003: not Hispanic or Latino, white alone "NHoLW"
- B03002\_004: not Hispanic or Latino, Black or African American alone "NHoLB"
- B03002\_005: not Hispanic or Latino, American Indian and Alaska Native alone "NHoLA-  
IAN"
- B03002\_006: not Hispanic or Latino, Asian alone "NHoLA"
- B03002\_007: not Hispanic or Latino, Native Hawaiian and Other Pacific Islander alone "NHoLNHOPI"
- B03002\_008: not Hispanic or Latino, Some other race alone "NHoLSOR"
- B03002\_009: not Hispanic or Latino, Two or more races "NHoLTOMR"
- B03002\_010: not Hispanic or Latino, Two races including Some other race "NHoLTRiSOR"
- B03002\_011: not Hispanic or Latino, Two races excluding Some other race, and three or more races "NHoLTReSOR"
- B03002\_012: Hispanic or Latino "HoL"
- B03002\_013: Hispanic or Latino, white alone "HoLW"

- B03002\_014: Hispanic or Latino, Black or African American alone "HoLB"
- B03002\_015: Hispanic or Latino, American Indian and Alaska Native alone "HoLAIAN"
- B03002\_016: Hispanic or Latino, Asian alone "HoLA"
- B03002\_017: Hispanic or Latino, Native Hawaiian and Other Pacific Islander alone "HoLNHOPI"
- B03002\_018: Hispanic or Latino, Some other race alone "HoLSOR"
- B03002\_019: Hispanic or Latino, Two or more races "HoLTOMR"
- B03002\_020: Hispanic or Latino, Two races including Some other race "HoLTRiSOR"
- B03002\_021: Hispanic or Latino, Two races excluding Some other race, and three or more races "HoLTReSOR"

Use the internal state and county arguments within the [get\\_acs](#) function to specify geographic extent of the data output.

DI is a measure of the evenness of racial/ethnic residential segregation when comparing smaller geographical areas to larger ones within which the smaller geographical areas are located. The DI metric can range in value from 0 to 1 and represents the proportion of racial/ethnic subgroup members that would have to change their area of residence to achieve an even distribution within the larger geographical area under conditions of maximum segregation.

Larger geographies available include state `geo_large = "state"`, county `geo_large = "county"`, and census tract `geo_large = "tract"` levels. Smaller geographies available include, county `geo_small = "county"`, census tract `geo_small = "tract"`, and census block group `geo_small = "block group"` levels. If a larger geographical area is comprised of only one smaller geographical area (e.g., a U.S. county contains only one census tract), then the DI value returned is NA.

## Value

An object of class 'list'. This is a named list with the following components:

`di` An object of class 'tbl' for the GEOID, name, and DI metric at specified larger census geographies.

`di_data` An object of class 'tbl' for the raw census values at specified smaller census geographies.

`missing` An object of class 'tbl' of the count and proportion of missingness for each census variable used to compute the DI

## See Also

[get\\_acs](#) for additional arguments for geographic extent selection (i.e., state and county).

## Examples

```
## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Dissimilarity Index of non-Hispanic Black vs. non-Hispanic white populations
## of census tracts within counties (2020)
duncan(geo_large = "county", geo_small = "tract", state = "GA", year = 2020,
        subgroup = "NHoLB", subgroup_ref = "NHoLW")
```

```
## End(Not run)
```

---

gini	<i>Gini Index based on Gini (1921)</i>
------	--

---

## Description

Retrieve the Gini Index values.

## Usage

```
gini(geo = "tract", year = 2020, quiet = FALSE, ...)
```

## Arguments

geo	Character string specifying the geography of the data either census tracts geo = "tract" (the default) or counties geo = "county".
year	Numeric. The year to compute the estimate. The default is 2020, and the years 2009 onward are currently available.
quiet	Logical. If TRUE, will display messages about potential missing census information
...	Arguments passed to <a href="#">get_acs</a> to select state, county, and other arguments for census characteristics

## Details

This function will retrieve the Gini Index of U.S. census tracts or counties for a specified geographical extent (e.g., the entire U.S. or a single state) based on Gini (1921) [doi:10.2307/2223319](#).

The function uses the [get\\_acs](#) function to obtain U.S. Census Bureau 5-year American Community Survey estimates of the Gini Index (ACS: B19083). The estimates are available for 2009 onward when ACS-5 data are available but are available from other U.S. Census Bureau surveys.

Use the internal state and county arguments within the [get\\_acs](#) function to specify geographic extent of the data output.

According to the U.S. Census Bureau <https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html>: "The Gini Index is a summary measure of income inequality. The Gini coefficient incorporates the detailed shares data into a single statistic, which summarizes the dispersion of income across the entire income distribution. The Gini coefficient ranges from 0, indicating perfect equality (where everyone receives an equal share), to 1, perfect inequality (where only one recipient or group of recipients receives all the income). The Gini is based on the difference between the Lorenz curve (the observed cumulative income distribution) and the notion of a perfectly equal income distribution."

Value

An object of class 'list'. This is a named list with the following components:

gini An object of class 'tbl' for the GEOID, name, and Gini index of specified census geographies.

missing An object of class 'tbl' of the count and proportion of missingness for the Gini index.

See Also

[get\\_acs](#) for additional arguments for geographic extent selection (i.e., state and county).

Examples

```
## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Tract-level metric (2020)
gini(geo = "tract", state = "GA", year = 2020)

# County-level metric (2020)
gini(geo = "county", state = "GA", year = 2020)

## End(Not run)
```

---

krieger	<i>Index of Concentration at the Extremes based on Feldman et al. (2015) and Krieger et al. (2016)</i>
---------	--

---

Description

Compute the Index of Concentration at the Extremes (Krieger) values.

Usage

```
krieger(geo = "tract", year = 2020, quiet = FALSE, ...)
```

Arguments

geo	Character string specifying the geography of the data either census tracts <code>geo = "tract"</code> (the default) or counties <code>geo = "county"</code> .
year	Numeric. The year to compute the estimate. The default is 2020, and the years 2009 onward are currently available.
quiet	Logical. If TRUE, will display messages about potential missing census information. The default is FALSE.
...	Arguments passed to <a href="#">get_acs</a> to select state, county, and other arguments for census characteristics

## Details

This function will compute the three Index of Concentration at the Extremes (ICE) of U.S. census tracts or counties for a specified geographical extent (e.g., entire U.S. or a single state) based on Feldman et al. (2015) [doi:10.1136/jech2015205728](https://doi.org/10.1136/jech2015205728) and Krieger et al. (2016) [doi:10.2105/AJPH.2015.302955](https://doi.org/10.2105/AJPH.2015.302955). The authors expanded the metric designed by Massey in a chapter of Booth & Crouter (2001) [doi:10.4324/9781410600141](https://doi.org/10.4324/9781410600141) who initially designed the metric for residential segregation. This function computes five ICE metrics:

- Income80th income percentile vs. 20th income percentile
- Educationless than high school vs. four-year college degree or more
- Race/Ethnicitywhite non-Hispanic vs. black non-Hispanic
- Income and race/ethnicity combinedwhite non-Hispanic in 80th income percentile vs. black alone (including Hispanic) in 20th income percentile
- Income and race/ethnicity combinedwhite non-Hispanic in 80th income percentile vs. white non-Hispanic in 20th income percentile

The function uses the [get\\_acs](#) function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for the geospatial computation. The yearly estimates are available for 2009 onward when ACS-5 data are available but are available from other U.S. Census Bureau surveys. The ACS-5 groups used in the computation of the five ICE metrics are:

- B03002: HISPANIC OR LATINO ORIGIN BY RACE
- B15002: SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
- B19001: HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 20XX INFLATION-ADJUSTED DOLLARS)
- B19001B: HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 20XX INFLATION-ADJUSTED DOLLARS) (BLACK OR AFRICAN AMERICAN ALONE HOUSEHOLDER)
- B19001H: HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 20XX INFLATION-ADJUSTED DOLLARS) (WHITE ALONE, NOT HISPANIC OR LATINO HOUSEHOLDER)

Use the internal state and county arguments within the [get\\_acs](#) function to specify geographic extent of the data output.

ICE metrics can range in value from -1 (most deprived) to 1 (most privileged). A value of 0 can thus represent two possibilities: (1) none of the residents are in the most privileged or most deprived categories, or (2) an equal number of persons are in the most privileged and most deprived categories, and in both cases indicates that the area is not dominated by extreme concentrations of either of the two groups.

## Value

An object of class 'list'. This is a named list with the following components:

**ice** An object of class 'tbl' for the GEOID, name, ICE metrics, and raw census values of specified census geographies.

**missing** An object of class 'tbl' of the count and proportion of missingness for each census variable used to compute the ICEs.

See Also

[get\\_acs](#) for additional arguments for geographic extent selection (i.e., state and county).

Examples

```
## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Tract-level metric (2020)
krieger(geo = "tract", state = "GA", year = 2020)

# County-level metric (2020)
krieger(geo = "county", state = "GA", year = 2020)

## End(Not run)
```

---

messer	<i>Neighborhood Deprivation Index based on Messer et al. (2006)</i>
--------	---

---

Description

Compute the Neighborhood Deprivation Index (Messer) values.

Usage

```
messer(
  geo = "tract",
  year = 2020,
  imp = FALSE,
  quiet = FALSE,
  round_output = FALSE,
  df = NULL,
  ...
)
```

Arguments

geo	Character string specifying the geography of the data either census tracts <code>geo = "tract"</code> (the default) or counties <code>geo = "county"</code> .
year	Numeric. The year to compute the estimate. The default is 2020, and the years 2010 onward are currently available.
imp	Logical. If TRUE, will impute missing census characteristics within the internal <a href="#">principal</a> . If FALSE (the default), will not impute.
quiet	Logical. If TRUE, will display messages about potential missing census information and the proportion of variance explained by principal component analysis. The default is FALSE.

<code>round_output</code>	Logical. If TRUE, will round the output of raw census and NDI values from the <a href="#">get_acs</a> at one and four significant digits, respectively. The default is FALSE.
<code>df</code>	Optional. Pass a pre-formatted 'dataframe' or 'tibble' with the desired variables through the function. Bypasses the data obtained by <a href="#">get_acs</a> . The default is NULL. See Details below.
<code>...</code>	Arguments passed to <a href="#">get_acs</a> to select state, county, and other arguments for census characteristics

## Details

This function will compute the Neighborhood Deprivation Index (NDI) of U.S. census tracts or counties for a specified geographical referent (e.g., US-standardized) based on Messer et al. (2006) [doi:10.1007/s115240069094x](https://doi.org/10.1007/s115240069094x).

The function uses the [get\\_acs](#) function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for computation involving a principal component analysis with the [principal](#) function. The yearly estimates are available for 2010 and after when all census characteristics became available. The eight characteristics are:

- C24030: percent males in management, science, and arts occupation
- B25014: percent of crowded housing
- B17017: percent of households in poverty
- B25115: percent of female headed households with dependents
- B19058: percent of households on public assistance
- B19001: percent of households earning <\$30,000 per year
- B06009: percent earning less than a high school education
- B23025: percent unemployed (2011 onward)
- B23001: percent unemployed (2010 only)

Use the internal state and county arguments within the [get\\_acs](#) function to specify the referent for standardizing the NDI (Messer) values. For example, if all U.S. states are specified for the state argument, then the output would be a U.S.-standardized index.

The continuous NDI (Messer) values are z-transformed, i.e., "standardized," and the categorical NDI (Messer) values are quartiles of the standardized continuous NDI (Messer) values.

Check if the proportion of variance explained by the first principal component is high (more than 0.5).

Users can bypass [get\\_acs](#) by specifying a pre-formatted data frame or tibble using the `df` argument. This function will compute an index using the first component of a principal component analysis (PCA) with a Varimax rotation (the default for [principal](#)) and only one factor (note: PCA set-up not unspecified in Messer et al. (2006)). The recommended structure of the data frame or tibble is an ID (e.g., GEOID) in the first feature (column), followed by the variables of interest (in any order) and no additional information (e.g., omit state or county names from the `df` argument input).



**Value**

An object of class 'list'. This is a named list with the following components:

- ndi An object of class 'tbl' for the GEOID, name, NDI (standardized), NDI (quartile), and raw census values of specified census geographies.
- pca An object of class 'principal', returns the output of [principal](#) used to compute the NDI values.
- missing An object of class 'tbl' of the count and proportion of missingness for each census variable used to compute the NDI.

**See Also**

[get\\_acs](#) for additional arguments for geographic referent selection (i.e., state and county).

**Examples**

```
messer(df = DCtracts2020[ , c(1, 3:10)])

## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Tract-level metric (2020)
messer(geo = "tract", state = "GA", year = 2020)

# Impute NDI for tracts (2020) with missing census information (median values)
messer(state = "tract", "GA", year = 2020, imp = TRUE)

# County-level metric (2020)
messer(geo = "county", state = "GA", year = 2020)

## End(Not run)
```

---

powell\_wiley

*Neighborhood Deprivation Index based on Andrews et al. (2020) and Slotman et al. (2022)*

---

**Description**

Compute the Neighborhood Deprivation Index (Powell-Wiley) values.

**Usage**

```
powell_wiley(
  geo = "tract",
  year = 2020,
  imp = FALSE,
```

```

    quiet = FALSE,
    round_output = FALSE,
    df = NULL,
    ...
  )

```

## Arguments

<code>geo</code>	Character string specifying the geography of the data either census tracts <code>geo = "tract"</code> (the default) or counties <code>geo = "county"</code> .
<code>year</code>	Numeric. The year to compute the estimate. The default is 2020, and the years 2010 onward are currently available.
<code>imp</code>	Logical. If TRUE, will impute missing census characteristics within the internal <code>principal</code> using median values of variables. If FALSE (the default), will not impute.
<code>quiet</code>	Logical. If TRUE, will display messages about potential missing census information, standardized Cronbach's alpha, and proportion of variance explained by principal component analysis. The default is FALSE.
<code>round_output</code>	Logical. If TRUE, will round the output of raw census and NDI values from the <code>get_acs</code> at one and four significant digits, respectively. The default is FALSE.
<code>df</code>	Optional. Pass a pre-formatted 'dataframe' or 'tibble' with the desired variables through the function. Bypasses the data obtained by <code>get_acs</code> . The default is NULL. See Details below.
<code>...</code>	Arguments passed to <code>get_acs</code> to select state, county, and other arguments for census characteristics

## Details

This function will compute the Neighborhood Deprivation Index (NDI) of U.S. census tracts or counties for a specified geographical referent (e.g., US-standardized) based on Andrews et al. (2020) [doi:10.1080/17445647.2020.1750066](https://doi.org/10.1080/17445647.2020.1750066) and Slotman et al. (2022) [doi:10.1016/j.dib.2022.108002](https://doi.org/10.1016/j.dib.2022.108002).

The function uses the `get_acs` function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for computation involving a factor analysis with the `principal` function. The yearly estimates are available in 2010 and after when all census characteristics became available. The thirteen characteristics chosen by Roux and Mair (2010) [doi:10.1111/j.1749-6632.2009.05333.x](https://doi.org/10.1111/j.1749-6632.2009.05333.x) are:

- MedHHInc (5B19013): median household income (dollars)
- PctRecvIDR (B19054): percent of households receiving dividends, interest, or rental income
- PctPubAsst (B19058): percent of households receiving public assistance
- MedHomeVal (B25077): median home value (dollars)
- PctMgmtBusScArti (C24060): percent in a management, business, science, or arts occupation
- PctFemHeadKids (B11005): percent of households that are female headed with any children under 18 years
- PctOwnerOcc (DP04): percent of housing units that are owner occupied

- PctNoPhone (DP04): percent of households without a telephone
- PctNComPlm (DP04): percent of households without complete plumbing facilities
- PctEducHSPlus (S1501): percent with a high school degree or higher (population 25 years and over)
- PctEducBchPlus (S1501): percent with a college degree or higher (population 25 years and over)
- PctFamBelowPov (S1702): percent of families with incomes below the poverty level
- PctUnempl (S2301): percent unemployed

Use the internal state and county arguments within the `get_acs` function to specify the referent for standardizing the NDI (Powell-Wiley) values. For example, if all U.S. states are specified for the state argument, then the output would be a U.S.-standardized index. Please note: the NDI (Powell-Wiley) values will not exactly match (but will highly correlate with) those found in Andrews et al. (2020) [doi:10.1080/17445647.2020.1750066](https://doi.org/10.1080/17445647.2020.1750066) and Slotman et al. (2022) [doi:10.1016/j.dib.2022.108002](https://doi.org/10.1016/j.dib.2022.108002) because the two studies used a different statistical platform (i.e., SPSS and SAS, respectively) that intrinsically calculate the principal component analysis differently from R.

The categorical NDI (Powell-Wiley) values are population-weighted quintiles of the continuous NDI (Powell-Wiley) values.

Check if the proportion of variance explained by the first principal component is high (more than 0.5).

Users can bypass `get_acs` by specifying a pre-formatted data frame or tibble using the `df` argument. This function will compute an index using the first component of a principal component analysis (PCA) with a Promax (oblique) rotation and a minimum Eigenvalue of 1, omitting variables with absolute loading score < 0.4. The recommended structure of the data frame or tibble is an ID (e.g., GEOID) in the first feature (column), an estimate of the total population in the second feature (column), followed by the variables of interest (in any order) and no additional information (e.g., omit state or county names from the `df` argument input).

## Value

An object of class 'list'. This is a named list with the following components:

`ndi` An object of class 'tbl' for the GEOID, name, NDI continuous, NDI quintiles, and raw census values of specified census geographies.

`pca` An object of class 'principal', returns the output of `principal` used to compute the NDI values.

`missing` An object of class 'tbl' of the count and proportion of missingness for each census variable used to compute the NDI.

`cronbach` An object of class 'character' or 'numeric' for the results of the Cronbach's alpha calculation. If only one factor is computed, a message is returned. If more than one factor is computed, Cronbach's alpha is calculated and should check that it is >0.7 for respectable internal consistency between factors.

## See Also

`get_acs` for additional arguments for geographic referent selection (i.e., state and county).

**Examples**

```
powell_wiley(df = DCtracts2020[ , -c(3:10)])

## Not run:
# Wrapped in \dontrun{} because these examples require a Census API key.

# Tract-level metric (2020)
powell_wiley(geo = "tract", state = "GA", year = 2020)

# Impute NDI for tracts (2020) with missing census information (median values)
powell_wiley(state = "tract", "GA", year = 2020, imp = TRUE)

# County-level metric (2020)
powell_wiley(geo = "county", state = "GA", year = 2020)

## End(Not run)
```

# Index

- \* **datasets**
  - DCtracts2020, [8](#)
- \* **package**
  - ndi-package, [2](#)
- anthopolos, [3](#), [4](#)
- bravo, [3](#), [6](#)
- DCtracts2020, [3](#), [8](#)
- duncan, [3](#), [9](#)
- get\_acs, [3–8](#), [10–19](#)
- gini, [3](#), [12](#)
- krieger, [3](#), [13](#)
- Matrix, [3](#)
- messer, [3](#), [8](#), [15](#)
- ndi (ndi-package), [2](#)
- ndi-package, [2](#)
- powell\_wiley, [3](#), [8](#), [17](#)
- principal, [15–19](#)
- psych, [3](#)
- sf, [3](#)
- tidycensus, [3](#)