Package 'ndi'

August 15, 2022

```
Title Neighborhood Deprivation Indices Version 0.1.1
```

Date 2022-08-14

Maintainer Ian D. Buller <ian.buller@nih.gov>

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; 2010-2020). Using data from the ACS-5 (2009-2020), 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>s, (2) compute spatial the Educational Isolation Index (EI) based on Bravo et al. (2021) <doi:10.3390/ijerph18179384>,and (3) retrieve the aspatial Gini Index based on Gini (1921) <doi:10.2307/2223319>.

License Apache License (>= 2.0)

Encoding UTF-8 **RoxygenNote** 7.2.1 **Depends** R (>= 3.5.0)

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

Suggests ggplot2, testthat, tigris, R.rsp, spelling

VignetteBuilder R.rsp

Language en-US

URL https://github.com/idblr/ndi

2 ndi-package

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 2022-08-15 14:20:02 UTC

R topics documented:

ndi-package . anthopolos																		
bravo																		
gini																		
messer powell_wiley																		

Index 14

ndi-package

The ndi Package: Neighborhood Deprivation Indices

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 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.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; 2010-2020). Using data from the ACS-5 (2009-2020), 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, and (3) retrieve the aspatial Gini Index based on Gini (1921) doi:10.2307/2223319.

Key content of the 'ndi' package include:

Metrics of Socio-Economic Deprivation and Disparity

anthopolos 3

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

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

gini Retrieves the aspatial Gini Index based on Gini (1921) doi:10.2307/2223319.

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

powell_wiley Computes the aspatial Neighborhood Deprivation Index (NDI) 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.17496632.2009.05333.x.

Dependencies

The 'ndi' package relies heavily upon tidycensus and 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 by fictionalizing, adding percent 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 or Slotman et al. (2022) doi:10.1016/j.dib.2022.108002, respectively, but the package maintainer worked directly with the 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 or EI included in Anthopolos et al. (2011) doi:10.1016/j.sste.2011.06.002 or Bravo et al. (2021) doi:10.3390/ijerph18179384, 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@nih.gov>

anthopolos

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

Description

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

Usage

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

4 anthopolos

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 between 2009 and 2020 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 get_acs 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., 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 available for 2009 through 2020 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"

anthopolos 5

• 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 <code>get_acs</code> 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 that is 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 that is 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)
```

6 bravo

h		٠,	_
v	ra	v	u

Educational Isolation Index based on Bravo et al. (2021)

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 geo = "tract" (the default) or counties geo = "county".
year	Numeric. The year to compute the estimate. The default is 2020 and the years between 2009 and 2020 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 get_acs 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., entire U.S. or a single state) based on Bravo et al. (2021) doi: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 <code>get_acs</code> function to obtain U.S. Census Bureau 5-year American Community Survey characteristics used for the geospatial computation. The yearly estimates available for 2009 through 2020 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"

gini 7

Use the internal state and county arguments within the <code>get_acs</code> 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 that is 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 that is 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)
```

gini

Gini Index based on Gini (1921)

Description

Retrieve the Gini Index values.

8 gini

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 between 2009 and 2020 are currently available.
quiet	Logical. If TRUE, will display messages about potential missing census information
	Arguments passed to get_acs 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., 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 estimate of the Gini Index (ACS: B19083). The estimate are available for 2009 through 2020 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).

messer 9

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)
```

messer

Neighborhood Deprivation Index based on Messer et al. (2006)

Description

Compute the Neighborhood Deprivation Index (Messer) values.

Usage

```
messer(geo = "tract", year = 2020, imp = FALSE, 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 between 2010 and 2020 are currently available.
imp	Logical. If TRUE, will impute missing census characteristics within the internal principal. If FALSE (the default), will not impute.
quiet	Logical. If TRUE, will display messages about potential missing census information and proportion of variance explained by principal component analysis. The default is FALSE.
•••	Arguments passed to get_acs 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.

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 2010 and after when all census characteristics became available. The eight characteristics are:

10 messer

- 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 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).

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

```
## 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 11

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, imp = FALSE, 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 (2010-2020 currently available).
imp	Logical. If TRUE, will impute missing census characteristics within the internal principal using median values of variables. If FALSE (the default), will not impute.
quiet	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.
	Arguments passed to get_acs 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 and Slotman et al. (2022) doi:10.1016/j.dib.2022.108002.

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

12 powell_wiley

- 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 <code>get_acs</code> 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) <code>doi:10.1080/17445647.2020.1750066</code> and Slotman et al. (2022) <code>doi:10.1016/j.dib.2022.108002</code> 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).

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, the 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

```
## 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)
```

powell_wiley 13

```
# 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

```
* package
ndi-package, 2

anthopolos, 3, 3

bravo, 3, 6

get_acs, 4-12
gini, 3, 7

Matrix, 3
messer, 3, 9

ndi (ndi-package), 2
ndi-package, 2

powell_wiley, 3, 11
principal, 9-12
psych, 3

sf, 3

tidycensus, 3
```