Package 'geodaData'

May 27, 2020

Title Spatial Analysis Datasets for Teaching

Version 0.1.0

Description Stores small spatial datasets used to teach basic spatial analysis concepts. Datasets are based off of the 'GeoDa' software workbook and data site <https://geodacenter.github.io/data-and-lab/> developed by Luc Anselin and team at the University of Chicago. Datasets are stored as 'sf' objects.

Depends R (>= 3.3.0)

License CC0

URL https://github.com/spatialanalysis/geodaData

BugReports https://github.com/spatialanalysis/geodaData/issues

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

Suggests sf

NeedsCompilation no

Author Angela Li [aut, cre] (<https://orcid.org/0000-0002-8956-419X>), Luc Anselin [ctb] (Creator of original spatial datasets)

Maintainer Angela Li <ali6@uchicago.edu>

Repository CRAN

Date/Publication 2020-05-27 09:20:02 UTC

R topics documented:

`chicago_comm` *Chicago Community Areas (2010).*

Description

Population in Chicago community areas in 2010.

Usage

```
chicago_comm
```

Format

An sf data frame with 77 rows, 4 variables, and a geometry column:

community Community name

area_num_1 Community ID

NID Community ID (repeated)

POP2010 Population in 2010

geometry MULTIPOLYGON

Details

Sf object, unprojected. EPSG 4326: WGS84.

Source

https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-cauq-8yn6

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(chicago_comm)

  plot(chicago_comm["community"])
}
```

clev_pts *Cleveland Home Sales (2015).*

Description

Location and sales price of home sales in a core area of Cleveland, OH for the fourth quarter of 2015.

Usage

```
clev_pts
```

Format

An sf data frame with 205 rows, 9 variables, and a geometry column:

unique_id unique parcel id

parcel unique parcel number

x point latitude

y point longitude

sale_price price paid for the house ($)

tract10int License plate number and sometimes a description (state, color). Some entries did not include a plate number.

quarter quarter of sale (4th for all)

year1 year of sale (2015 for all)

yrquarter year and quarter of sale (4th quarter of 2015 for all)

geometry POINT

Details

Sf object, units in ft. EPSG 3734: NAD83 / Ohio North (ftUS).

Source

Cuyahoga County Fiscal Office. https://geodacenter.github.io/data-and-lab//clev_sls_154_core/

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(clev_pts)

  plot(clev_pts["unique_id"])
}
```

commpop *Chicago Population Change (2000-2010).*

Description

Change in population in Chicago community areas from 2000 to 2010.

Usage

```
commpop
```

Format

An sf data frame with 77 rows, 8 variables, and a geometry column:

community Community name

NID Community ID

POP2010 Population in 2010

POP2000 Population in 2000

POPCH Population change, count

POPPERCH Population percent change

popplus 1 if area has positive population change (17 observations)

popneg 1 if area has negative population change (60 observations)

geometry MULTIPOLYGON

Details

Sf object, unprojected. EPSG 4326: WGS84.

Source

https://www.chicago.gov/city/en/depts/dcd/supp_info/community_area_2000and2010censuspopulationcompa
html

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(commpop)

  plot(commpop["community"])
}
```

guerry *Guerry "Moral Statistics" (1830s).*

Description

Classic social science foundational study by Andre-Michel Guerry on crime, suicide, literacy and other "moral statistics" in 1830s France. Data from the R package Guerry (Michael Friendly and Stephane Dray).

Usage

```
guerry
```

Format

An sf data frame with 85 rows, 23 variables, and a geometry column:

variable Description

dept, code_de Department ID: Standard numbers for the departments

region Region of France ('N'='North', 'S'='South', 'E'='East', 'W'='West', 'C'='Central'). Corsica is coded as NA.

dprtmnt Department name: Departments are named according to usage in 1830, but without accents. A factor with levels Ain Aisne Allier ... Vosges Yonne

crm_prs Population per Crime against persons.

crm_prp Population per Crime against property.

litercy Percent of military conscripts who can read and write.

donatns Donations to the poor.

infants Population per illegitimate birth.

suicids Population per suicide.

maincty Size of principal city ('1:Sm', '2:Med', '3:Lg'), used as a surrogate for population density. Large refers to the top 10, small to the bottom 10; all the rest are classed Medium.

wealth Per capita tax on personal property. A ranked index based on taxes on personal and movable property per inhabitant.

commerc Commerce and Industry, measured by the rank of the number of patents / population.

clergy Distribution of clergy, measured by the rank of the number of Catholic priests in active service population.

crim_prn Crimes against parents, measured by the rank of the ratio of crimes against parents to all crimes – Average for the years 1825-1830.

infntcd Infanticides per capita. A ranked ratio of number of infanticides to population – Average for the years 1825-1830.

dntn_cl Donations to the clergy. A ranked ratio of the number of bequests and donations inter vivios to population – Average for the years 1815-1824.

lottery Per capita wager on Royal Lottery. Ranked ratio of the proceeds bet on the royal lottery to population — Average for the years 1822-1826.

desertn Military desertion, ratio of number of young soldiers accused of desertion to the force of the military contingent, minus the deficit produced by the insufficiency of available billets – Average of the years 1825-1827.

instrct Instruction. Ranks recorded from Guerry's map of Instruction. Note: this is inversely related to Literacy.

prsttts Number of prostitutes registered in Paris from 1816 to 1834, classified by the department of their birth

distanc Distance to Paris (km). Distance of each department centroid to the centroid of the Seine (Paris).

area Area (1000 km^2).

pop1831 Population in 1831, in 1000s.

geometry MULTIPOLYGON

Details

Sf object, units in m. EPSG 27572: NTF (Paris) / Lambert zone II.

Source

- Angeville, A. (1836). Essai sur la Statistique de la Population française Paris: F. Doufour.
- Guerry, A.-M. (1833). Essai sur la statistique morale de la France Paris: Crochard. English translation: Hugh P. Whitt and Victor W. Reinking, Lewiston, N.Y. : Edwin Mellen Press, 2002.
- Parent-Duchatelet, A. (1836). De la prostitution dans la ville de Paris, 3rd ed, 1857, p. 32, 36

https://geodacenter.github.io/data-and-lab/Guerry/

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(guerry)

  plot(guerry["CODE_DE"])
}
```

---

ncovr *Homicides & Socio-Economics (1960-90).*

Description

Homicides and selected socio-economic characteristics for continental U.S. counties. Data for four decennial census years: 1960, 1970, 1980 and 1990.

Usage

```
ncovr
```

Format

An sf data frame with 3085 rows, 69 variables, and a geometry column:

name county name

state_name state name

state_fips state fips code (character)

cnty_fips county fips code (character)

fips combined state and county fips code (character)

stfips state fips code (numeric)

cofips county fips code (numeric)

fipsno fips code as numeric variable

south dummy variable for Southern counties (South = 1)

hr homicide rate per 100,000 (1960, 1970, 1980, 1990)

hc homicide count, three year average centered on 1960, 1970, 1980, 1990

po county population, 1960, 1970, 1980, 1990

rd resource deprivation 1960, 1970, 1980, 1990 (principal component, see Codebook for details)

ps population structure 1960, 1970, 1980, 1990 (principal component, see Codebook for details)

ue unemployment rate 1960, 1970, 1980, 1990

dv divorce rate 1960, 1970, 1980, 1990 (percent males over 14 divorced)

ma median age 1960, 1970, 1980, 1990

pol log of population 1960, 1970, 1980, 1990

dnl log of population density 1960, 1970, 1980, 1990

mfil log of median family income 1960, 1970, 1980, 1990

fp percent families below poverty 1960, 1970, 1980, 1990 (see Codebook for details)

blk percent black 1960, 1970, 1980, 1990

gi Gini index of family income inequality 1960, 1970, 1980, 1990

fh percent female headed households 1960, 1970, 1980, 1990

geometry MULTIPOLYGON

Details

Sf object, unprojected. EPSG 4326: WGS84.

Source

S. Messner, L. Anselin, D. Hawkins, G. Deane, S. Tolnay, R. Baller (2000). An Atlas of the Spatial Patterning of County-Level Homicide, 1960-1990. Pittsburgh, PA, National Consortium on Violence Research (NCOVR). https://geodacenter.github.io/data-and-lab/ncovr/

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(ncovr)

  plot(ncovr["NAME"])
}
```

---

nyc *Rental Housing and Demographics in NYC (2000s), non-spatial.*

Description

Demographic and housing data for New York City's 55 sub-boroughs (2000s).

Usage

```
nyc
```

Format

A data frame with 55 rows and 34 variables:

CODE sub-borough code, 1XX Bronx, 2XX Brooklyn, 3XX Manhattan, 4XX Queens, 5XX Staten Island

FORHIS06 percentage of hispanic population, not born in US, 2006

FORHIS07 percentage of hispanic population, not born in US, 2007

FORHIS08 percentage of hispanic population, not born in US, 2008

FORHIS09 percentage of hispanic population, not born in US, 2009

FORWH06 percentage of white population, not born in US, 2006

FORWH07 percentage of white population, not born in US, 2007

FORWH08 percentage of white population, not born in US, 2008

FORWH09 percentage of white population, not born in US, 2009

HHSIZ1990 average number of people per household, 1990

HHSIZ00 average number of people per household, 2000

HHSIZ02 average number of people per household, 2002

HHSIZ05 average number of people per household, 2005

HHSIZ08 average number of people per household, 2008

KIDS2000 percentage households w kids under 18, 2000

KIDS2005 percentage households w kids under 18, 2005

KIDS2006 percentage households w kids under 18, 2006

KIDS2007 percentage households w kids under 18, 2007

KIDS2008 percentage households w kids under 18, 2008

KIDS2009 percentage households w kids under 18, 2009

NAME name of borough, one of five

RENT2002 median monthly contract rent, 2002

RENT2005 median monthly contract rent, 2005

RENT2008 median monthly contract rent, 2008

RENTPCT02 percentage of housing stock that is market rate rental units, 2002

RENTPCT05 percentage of housing stock that is market rate rental units, 2005

RENTPCT08 percentage of housing stock that is market rate rental units, 2008

SUBBOROUGH name of sub-borough

PUBAST90 percentage of households receiving public assistance, 1990

PUBAST00 percentage of households receiving public assistance, 2000

YRHOM02 average number of years living in current residence, 2002

YRHOM05 average number of years living in current residence, 2005

YRHOM08 average number of years living in current residence, 2008

bor_subb sub-borough code, repeated

Details

Dataframe, no spatial components.

Source

https://geodacenter.github.io/data-and-lab/nyc/

`nyc_sf` *Rental Housing and Demographics in NYC (2000s).*

Description

Demographic and housing data for New York City's 55 sub-boroughs (2000s).

Usage

```
nyc_sf
```

Format

An sf data frame with 55 rows, 34 variables, and a geometry column:

forhis06-09 percentage of hispanic population, not born in US

forwh06-09 percentage of white population, not born in US

hhsiz1990 average number of people per household

hhsiz00 average number of people per household

hhsiz02-05-08 average number of people per household

kids2000, kids2005-2009 percentage households w kids under 18

rent2002,2005,2008 median monthly contract rent

rentpct02,05,08 percentage of housing stock that is market rate rental units

pubast90,00 percentage of households receiving public assistance

yrhom02,05,08 average number of years living in current residence

geometry MULTIPOLYGON

Details

Sf object, units in ft. EPSG 2263: NAD83 / New York Long Island (ftUS).

Source

https://geodacenter.github.io/data-and-lab/nyc/

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(nyc_sf)

  plot(nyc_sf["bor_subb"])
}
```

---

ohio_lung *Ohio Lung Cancer Mortality (1960s-80s).*

Description

Ohio lung cancer data for 1968, 1978 and 1988.

Usage

```
ohio_lung
```

Format

An sf data frame with 88 rows, 42 variables, and a geometry column:

county_id Sequential county ID (alphabetic order)

name County name

fipsno Fips code as numeric

lg_ryy Lung cancer cases for gender G (M or F) and race R (W or B) in year yy (1968, 1978, 1988)

popg_ryy Population at risk for gender G (M or F) and race R (W or B) in year yy (1968, 1978, 1988)

l_gyy Total male and female lung cancer cases for each year

pop_gyy Total population at risk by gender

geometry POLYGON

Details

Sf object, units in m. EPSG 32617: WGS 84 / UTM Zone 17N.

Source

https://geodacenter.github.io/data-and-lab/ohiolung/

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(ohio_lung)

  plot(ohio_lung["FIPSNO"])
}
```

---

vehicle_pts *Abandoned Vehicles (2016).*

Description

Point locations of abandoned vehicles in Chicago in September 2016.

Usage

```
vehicle_pts
```

Format

An sf data frame with 2635 rows, 10 variables, and a geometry column:

CreationDt Date created

Address Address of abandoned vehicle

ZIPCode Zip code of abandoned vehicle

X Projected X, EPSG 32616

Y Projected Y, EPSG 32616

Ward Ward ID

PoliceD Police district ID

Comm Community area ID

Latitude Latitude of vehicle

Longitude Longitude of vehicle

geometry POINT

Details

Sf object, unprojected. EPSG 4326: WGS84.

Source

https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Abandoned-Vehicles/3c9v-pnva

Examples

```
if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(vehicle_pts)

  plot(vehicle_pts["CreationDt"])
}
```

Index