Package 'UScancer'

February 19, 2015

Type Package

Title Create US cancer datasets from SEER, IARC, and US Census data

Version 0.1-2

Date 2014-08-08

Author Jonathan Lee

Maintainer Jonathan Lee <`jonathan@utstat.utoronto.ca`>

Depends R (>= 3.0.0), rgdal

Suggests diseasemapping

Description This package contains functions to read cancer data from SEER (http://seer.cancer.gov/) and IARC (http://www.iarc.fr) to create datasets at the county level based on US census information.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2014-08-10 00:25:33

R topics documented:

`UScancer-package` *Create US cancer datasets from SEER, IARC, and US Census data*

Description

This package contains functions to read cancer data from SEER (http://seer.cancer.gov/) and IARC (http://www.iarc.fr) to create datasets at the county level based on US census information.

Details

Package:	UScancer
Type:	Package
Version:	0.1-1
Date:	2013-08-21
License:	GPL-2

Due to size constraints, US census and cancer data are not included in this package, but are necessary for the package to run. This data is freely available for download on the internet.

SEER cancer data can be downloaded (after completing a signed SEER Research Data Agreement form) at http://seer.cancer.gov/data/access.html.

The 2010 US Census Demographic Profile data in Shapefile format can be downloaded from http://www2.census.gov/geo/tiger/TIGER2010DP1/County_2010Census_DP1.zip.

Conversion from FIPS codes to county names is done through a lookup table included in this package – see the `fipslookup` dataset.

Conversion from demographic profile data element to its full name is done through a lookup table included in this package – see the `dplookup` dataset.

Conversion from cancer sites (types) to ICD-9 cancer codes are done through the `siteLookup`.

Author(s)

Jonathan Lee

Maintainer: Jonathan Lee <jonathan@utstat.utoronto.ca>

Examples

```
data(fipslookup)
data(dplookup)

## Not run:

censusData <- readCensus(state="Kentucky",fips=fipslookup,codes=dplookup)
seerData <- readSEER(year=2005,state="Kentucky",cancer="Larynx",fips=fipslookup)
```

```
library(diseasemapping)
larynxRates <- cancerRates("USA", year=1998:2002,site="Larynx")
kentucky <- getSMR(censusData,larynxRates,seerData,regoinCode="County")

breaks = c(0,1,seq(2, ceiling(max(kentucky$SMR)),by=2))
thecol = terrain.colors(length(breaks)-1)
plot(kentucky, col = thecol[cut(kentucky$SMR, breaks,include.lowest=TRUE)] )

## End(Not run)
```

`dplookup` *Census demographics profile lookup table*

Description

A lookup table for US census demographic profile data element to its full name. Only contains the codes for sex and age groups.

Usage

```
data(dplookup)
```

Format

A `data.frame` object with 39 rows with the following 2 variables.

`code`	Demographic profile data element code
`desc`	A simplified description of the code. i.e. Male ages 5 to 9 is M5_9

Source

Original data from `http://gis.drcog.org/datacatalog/sites/default/.../Census_Table_Descriptions.xls`.

Examples

```
data(dplookup)
```

`fipslookup` *FIPS lookup table*

Description

Lookup table for 5-digit FIPS codes to state and county name.

Usage

```
data(fipslookup)
```

Format

A `data.frame` object with 3227 rows and the following 2 variables.

`state`	State name
`county`	County name

Row names are the 5-digit FIPS code (state+county). State FIPS codes with only 1 digit are appended by a leading 0.

Details

This is meant to be used with the `readSEER` function to be passed in as the `fips` argument.

Source

Original file taken from `http://www.schooldata.com/pdfs/US_FIPS_Codes.xls` with updates done manually. New counties (as of 2010) were added, along with FIPS codes for Hawaii, Alaska, and Puerto Rico.

Examples

```
data(fipslookup)
```

`readCensus` *Read US census*

Description

Read US census demographic profile data. The raw data can be downloaded in Shapefile format from `http://www2.census.gov/geo/tiger/TIGER2010DP1/County_2010Census_DP1.zip` to be used with this function.

Usage

```
readCensus(dsn="",layer="County_2010Census_DP1",state,fips,codes)
```

Arguments

`dsn`	The data source name, usually a path pointing to the folder the county shapefile is in. Defaults to current directory.
`layer`	Layer name, usually the filename without the extension. Defaults to the 2010 Census data by county (County_2010_Census_DP1).
`state`	A character vector specifying which states to extract. If not specified, defaults to all states.
`fips`	A `data.frame` object with 5-digit FIPS codes as row numbers and two columns (state,county) with names of the state and county matching the FIPS code. A complete object is included in this package, see the `fipslookup` dataset.
`codes`	A lookup table for demographic profile data element to its full name. One such lookup table for age and sex demographic groups is provided with this package, see the `dplookup` dataset.

Value

A `SpatialPolygonsDataFrame`. The `data` slot contains renamed columns of all the age and sex groups along with a 5 digit state+county FIPS column. This format is compatible for use with the `getSMR` function from the `diseasemapping` package.

Author(s)

Jonathan Lee `<jonathan@utstat.utoronto.ca>`

See Also

`dplookup`

Examples

```
data(dplookup)
data(fipslookup)

## Not run:
usPop <- readCensus(dsn="/",layer="County_2010Census_DP1",state="Kentucky",
                     fips=fipslookup,codes=dplookup)
## End(Not run)
```

`readSEER` *Read SEER cancer data (ASCII)*

Description

Reads an ASCII text version of SEER data for one or more given years, states and cancer sites. Data is freely available (after completing a signed SEER Research Data Agreement form) at `http://seer.cancer.gov/data/access.html`.

Usage

```
readSEER(file,year,state,cancer,site,fips)
```

Arguments

`file`	A vector of filenames specifying where to find the SEER data. For example, for respiratory cancers between 1973-2010, one could specify the following path from the default directory structure of SEER data `"SEER_1973_2010_TEXTDATA/incidence/yr1973_201`
`year`	A numeric vector specifying which years of the data to look at. If not specified, defaults to all years.
`state`	A character vector specifying which states to extract. Should be one or more of "Georgia", "California", "New Jersey", "Kentucky", "Louisiana", "Connecticut", "Michigan", "Hawaii", "Iowa", "New Mexico", "Utah", "Washington". If not specified, defaults to all states.
`cancer`	A character vector specifying which cancers. This is passed to `siteLookup` to convert to a regular expression matching the proper site codes. Should be one or more of cancers from `listCancers`. If not specified, `site` is used instead. If neither are specified, then defaults to all cancers.
`site`	A character vector of regular expressions to match cancer site codes on. This takes precidence over `cancer`. If neither are specified, then defaults to all cancers.
`fips`	A `data.frame` object with 5-digit FIPS codes as row numbers and two columns (state,county) with names of the state and county matching the FIPS code. A complete object is included in this package, see the `fipslookup` dataset.

Value

Returns a `data.frame` of cancer cases with sex, age of diagnosis, year of diagnosis, site, state, county, and 5-digit state+county FIPS code. This format is compatible for use with the `getSMR` function from the `diseasemapping` package.

Author(s)

Jonathan Lee `<jonathan@utstat.utoronto.ca>`

See Also

`fipslookup`

Examples

```
data(fipslookup)

## Not run:
#file should point to the proper TXT file containing the cancer data you are interested in
seerData <- readSEER(file="RESPIR.TXT",year=2005,state="Kentucky",cancer="Larynx",fips=fipslookup)
## End(Not run)
```

`siteLookup` *Converts a name description to ICD matching pattern*

Description

Given a cancer name (one from `listCancers`), a regular expression pattern is returned to match all ICD-09 site codes that are associated with that cancer name. This regex expression can be passed to `readSEER` as the `site` argument. Alternatively, this function is called internally by `readSEER` if only a cancer name is given.

Usage

```
siteLookup(cancer)
listCancers()
```

Arguments

`cancer`	Character string of a cancer name. Accepted names listed by `listCancers`.

Value

`siteLookup` returns a regular pattern expression used to match ICD-09 codes.

`listCancers` returns a vector of accepted cancer names.

Author(s)

Jonathan Lee `<jonathan@utstat.utoronto.ca>`

See Also

`readSEER`

Examples

```
listCancers()
siteLookup("Larynx")
```

Index