Package 'IRISS'

September 27, 2022

Type Package

Title IRISS

Version 1.0

Author AURIN

Maintainer Michael Rigby <rigby.m@unimelb.edu.au>, Angela Ryan <angela.ryan@unimelb.edu.au>, German Gonzalez <german.gonzalez@unimelb.edu.au>

Description The goal of the demonstrator is to create a first pilot of WP3 using the Australian Survey of Social Attitudes (AuSSA) as a proxy for HILDA, while our application to access HILDA is pending. When access to HILDA is granted, the code and the learnings from this demonstrator will inform the development of the GeoSocial service.

License GLP-3

Encoding UTF-8

LazyData true

Imports arsenal,dataverse,sjlabelled,ows4R,utils,httr,sf,tidyverse,config

Depends arsenal,dataverse,sjlabelled,ows4R,utils,httr,sf,tidyverse,config

RoxygenNote 7.2.1

R topics documented:

Index   **12**

---

addPOA   *addPOA*

Description

Only used during testing - randomly assigns postcodes to data to assist in testing joins on spatial
identifiers.

Usage

```
addPOA(POA, path)
```

Arguments

POA	The postcode dataset.
path	The location of the stata dataset

Value

True if the file is storage correctly

---

checkJoinRuleList   *checkJoinRuleList*

Description

checks join rule list against global variables defining years of spatial and survey datasets. Throws
warning if datasets will not be included. To sit within the joinDataDefined function.

Usage

```
checkJoinRuleList(join_rule_list)
```

Arguments

join_rule_list   a named list defining the join rules for years in the intersected dataset

checkPostcodeColumn *checkPostcodeColumn*

Description

Checks that the join column (postcodes) contain valid values.

Usage

```
checkPostcodeColumn(dataset, postcode_column_name)
```

Arguments

dataset   the dataset

postcode_column_name

                  the postcode column (add as an object not a string)

Value

True if all the values are valid

checkVariableNames *checkVariableNames*

Description

Checks that join dataset variable names meet the critera for a stata dataset. The critera is: characters must be equal or less than 32, only letters, numbers and underscore (no punctuation), and must start with letter or underscore.

Usage

```
checkVariableNames(join_dataset)
```

Arguments

join_dataset   The dataset that is intended to be joined to a stata dataset.

Value

if the variable names are valid. Prints a message describing problem and specific variable that is the problem if invalid.

---

checkVariableNames_duplicates

checkVariableNames_duplicates

Description

Checks that the two datasets don't contain any duplicate variable names.

Usage

```
checkVariableNames_duplicates(dataset1, dataset2)
```

Arguments

dataset1	a dataframe to join
dataset2	a dataframe to join

Value

True if none of the variable names are duplicates, otherwise false if there is overlap

---

CreateFolders   *CreateFolders*

Description

This function generates the folders the script needs to run

Usage

```
CreateFolders(path = getwd())
```

Arguments

name	the path

Value

True if the folders are created correctly.

downloadAurinData *getAurinData*

Description

Downloads data from the ADP server

Usage

```
downloadAurinData(id, file)
```

Arguments

id	Unique identifier from the data catalogue https://data.aurin.org.au
file	path string to where the file should be stored

Value

True if the download is correct

downloadDataverseData *getDataverseData*

Description

Function takes inputs doi (get from dataverse listing of desired data) and the year of the data

Usage

```
downloadDataverseData(id, file)
```

Arguments

id	unique doi (get from dataverse listing of desired data)
file	where the file would be storage

Value

True if the download is correct

GenerateLog   *GenerateLog*

Description

GenerateLog

Usage

```
GenerateLog(name)
```

Arguments

name   the name of the log file

Value

True if the log file is generated

---

getAurinData   *getAurinData*

Description

Reads global variables 'spatialDatasets', 'joinData_year', and 'joinData_name' and downloads the data from the Aurin Servers to the listed file location

Usage

```
getAurinData()
```

Value

True if the download is correct

---

getDataverseData   *getDataverseData*

Description

Reads global variables 'surveyDatasets', 'survey_year', and 'survey_name' and downloads the data from the ADA Servers to the listed file location and adds Post code column

Usage

```
getDataverseData()
```

Value

True if the download is correct

---

joinDataDefined	*joinDataDefined*

Description

Joins datasets on postcode and by a rule for years

Usage

```
joinDataDefined(survey_dataset, join_dataset, output_path, join_rules)
```

Arguments

`survey_dataset`	a stata dataframe
`join_dataset`	a dataset
`output_path`	the location and name of the integrated dta file
`join_rules`	a named list that links years between datasets that should be joined

Value

True

---

joinDataSimple	*joinDataSimple*

Description

Joins datasets on postcode with option to also join by matching the added 'data-year' column

Usage

```
joinDataSimple(survey_dataset, join_dataset, output_path, by_data_year = FALSE)
```

Arguments

`survey_dataset`	a stata dataframe
`join_dataset`	a dataset
`output_path`	the location and name of the integrated dta file
`by_data_year`	logical variable indicating if data should also be joined by matching year variable

Value

a joined dataset.

LoadCredentials *LoadCredentials*

Description

Loads the credentials for AURIN ADP and ADA dataverse and set in the global environment.

Usage

```
LoadCredentials(config_file)
```

Arguments

config_file   The path to the configuration file

Value

True if the connection credentials are valid

LoadDataOptions *LoadDataOptions*

Description

This function reads csv files and loads matrixes of available data reference strings and storage locations

Usage

```
LoadDataOptions(surveyParameters, spatialParameters)
```

Arguments

surveyParameters
                csv files that contain the list survey datasets
spatialParameters
                csv files that contain the list of spatial datasets

Value

True

LogDataInputs *Log data inputs*

Description

this function takes inputs (stored as global variables) and records them as a message for the log file

Usage

```
LogDataInputs()
```

Value

True

---

readSpatialDataset *readSpatialDataset*

Description

Reads one or more gml datasets, and combines them if necessary. If more than one dataset is read and combined, a column 'data_year' is added with the datasets year that row belongs to.

Usage

```
readSpatialDataset(folder, data_name, year)
```

Arguments

folder	The path to the folder where the spatial datasets are store
data_name	The name of the dataset
year	Either a string containing a single year, or a list of strings containing multiple years

Value

A dataframe with geometries removed

readSurveyDataset   *readSurveyDataset*

Description

Reads one or more stata datasets, and combines them if necessary. If more than one dataset is read and combined, a column 'data_year' is added with the datasets year that row belongs to.

Usage

```
readSurveyDataset(folder, data_name, year)
```

Arguments

folder	The path to the folder where the stata datasets are store
data_name	The name of the dataset
year	Either a string containing a single year, or a list of strings containing multiple years

Value

A dataframe with stata labels intact.

TestAurinConnection   *TestAurinConnection*

Description

Tests connection to AURIN ADP client. Requires AURIN credentials to be loaded into the system environment.

Usage

```
TestAurinConnection(log = TRUE)
```

Value

True if the connection is correct

TestDataverseConnection

TestDataverseConnection

Description

Tests connection to ADA dataverse. Requires dataverse token to be loaded in the system environment.

Usage

```
TestDataverseConnection()
```

Value

True if the connection is correct

Index