

Package ‘faoswsSeed’

March 10, 2015

Type Package

Title Package to perform the imputation of seed usage for the FAO seed domain.

Version 1.0.0

Date 2015-01-15

Author Joshua M. Brown-
ing <joshua.browning@fao.org>, Michael C. J. Kao <michael.kao@fao.org>

Maintainer Joshua M. Browning <joshua.browning@fao.org>

Description This package provides all the functions for imputation of seed usage.

URL https://github.com/rockclimber112358/sws_seed

License GPL (>= 3)

Imports

Depends faoswsFlag (>= 0.1.1), faoswsUtil (>= 0.1.2), faosws (>= 0.3.2), data.table (>= 1.9.2)

LazyData yes

ZipData no

VignetteBuilder knitr

Suggests knitr, ggplot2

R topics documented:

fillCountrySpecificSeedRate	2
fillGeneralSeedRate	2
getAreaData	3
getCountryGeneralSeedRate	3
getCountrySpecificSeedRate	4
imputeAreaSown	4
imputeSeed	6
imputeSeedDomain	7
saveSeedData	7
seedData	8
Index	9

fillCountrySpecificSeedRate

Fill country specific seed rates

Description

Fill country specific seed rates

Usage

```
fillCountrySpecificSeedRate(data, countryVariable = "geographicAreaM49",
  commodityVariable = "measuredItemCPC",
  countrySpecificData = getCountrySpecificSeedRate())
```

Arguments

data	The seed data.table, typically as produced by a call to getAreaData.
countryVariable	The column name of data that specifies the country code variable. This is needed to join with the table from the database.
commodityVariable	The column name of data that specifies the commodity code variable. This is needed to join with the table from the database.
countrySpecificData	A data.table with data describing seed rates for each country.

Value

No value is returned. Instead, seed rates (from countrySpecificData) are appended onto data.

fillGeneralSeedRate *Fill general seed rate*

Description

This function updates missing values of the Value_seedRate column in data. If the value is missing, it is replaced with the commodity value.

Usage

```
fillGeneralSeedRate(data, generalSeedData = getCountryGeneralSeedRate())
```

Arguments

data	The data.table object containing the seed data, typically as produced via getAreaData.
generalSeedData	A data.table containing seeding data specific to each commodity, typically as produced by getCountryGeneralSeedRate.

Value

No object is returned. Instead, the underlying data object is modified.

getAreaData

Function for obtaining the area harvested/sown data

Description

This function pulls the trade data from the database. The main function pulling the data is faosws::GetData, but additional steps are performed by this function (such as setting up the appropriate pivot, adding variables which are missing from the data as NA's, and setting data with missing flags and 0 values to NA values).

Usage

```
getAreaData(dataContext, areaSownElementCode = "5212",
             areaHarvestedElementCode = "5312", seedElementCode = "5525")
```

Arguments

dataContext	The context for the data, as generated by the SWS. This object can be created via a call like swsContext.datasets[[1]] (assuming the user is running this script on the SWS or after a call to GetTestEnvironment).
areaSownElementCode	The element code providing the dimension which corresponds to the area sown variable in the database.
areaHarvestedElementCode	The element code providing the dimension which corresponds to the area harvested variable in the database.
seedElementCode	The element code providing the dimension which corresponds to the seed variable in the database.

Value

A data.table object containing the data queried from the database.

getCountryGeneralSeedRate

Get Country General Seed Rates

Description

This function grabs data from the default_seed_rate table in the ess schema in the database and applies a few simple operations to it (renames columns, sets keys). These default seed rates are provided at the commodity level, as opposed to the commodity-country level in getCountrySpecificSeedRate().

Usage

```
getCountryGeneralSeedRate()
```

Details

The seed rates that are pulled from the database are assumed to be official values. That is, the observation status flags of all of these observations are set to "".

Value

The data.table object with specific seed rate data.

```
getCountrySpecificSeedRate
```

Function to load the country specific rate data

Description

This function grabs data from the specific_seed_rate table in the ess schema in the database and applies a few simple operations to it (renames columns, converts dimensions to characters, and sets all flags that aren't "E" to ""). These rates are provided at the commodity-country level as opposed to commodity level in getCountrySpecificSeedRate().

Usage

```
getCountrySpecificSeedRate()
```

Details

The seed rates that are pulled from the database are assumed to be official values. That is, the observation status flags of all of these observations are set to "".

Value

The data.table object with specific seed rate data.

```
imputeAreaSown
```

Get the area harvested/sown data

Description

Imputation is currently quite simple: if the area sown variable has no data, then area sown is assumed to be equal to area harvested. If some values exist and some are missing, than an average ratio is computed of area sown to area harvested and this ratio is applied to all missing values.

Usage

```
imputeAreaSown(data, valueAreaSown = "Value_measuredElement_5212",
  valueAreaHarvested = "Value_measuredElement_5312",
  flagObsAreaSown = "flagObservationStatus_measuredElement_5212",
  flagMethodAreaSown = "flagMethod_measuredElement_5212",
  flagObsAreaHarvested = "flagObservationStatus_measuredElement_5312",
  imputedObsFlag = "I", imputedMethodFlag = "e", byKey = NULL,
  imputationParameters = NULL)
```

Arguments

- | | |
|----------------------|--|
| data | The data.table object containing the seed data, usually as created by getArea-Data. |
| valueAreaSown | The column name of data which contains the value of the area sown variable. |
| valueAreaHarvested | The column name of data which contains the value of the area harvested variable. |
| flagObsAreaSown | The column name of data which contains the observation flag for the area sown variable. |
| flagMethodAreaSown | The column name of data which contains the method flag for the area sown variable. |
| flagObsAreaHarvested | The column name of data which contains the observation flag for the area harvested variable. |
| imputedObsFlag | The value to be assigned to the observation status flag for imputed observations. |
| imputedMethodFlag | The value to be assigned to the method flag for imputed observations. |
| byKey | Only used if imputationParameters is NULL (if imputationParameters is not NULL, ensemble imputation is performed and the byKey variable should be specified in this list). This value specifies how the mean seed to harvest ratio should be computed: by country, globally, or in some other way. Should be a column name of data, or NULL (in which case a global mean is used). |
| imputationParameters | A vector containing the parameters to use in the imputation of areaSown variable. If NULL, imputation is done by computing the mean ratio of area harvested to area sown and applying that to all cases. |

Value

No object is returned, instead the underlying data.table object is modified.

imputeSeed

Function to impute the seed.

Description

First, an estimate for the seed value at time t is generated by multiplying the area sown at time $t-1$ by the seed rate (divided by 1,000) at time t . Then, the estimate is inserted into the seed data.table if the seed value at time t is currently missing. The observation flag is updated by aggregating the flags of the seed rate and area sown, and the method flag is updated with the value of imputedFlag (passed as an argument). Lastly, the seedRateValue and seedRateFlag columns are deleted from the data.table.

Usage

```
imputeSeed(data, seedValue = "Value_measuredElement_5525",
  seedMethodFlag = "flagMethod_measuredElement_5525",
  seedObsFlag = "flagObservationStatus_measuredElement_5525",
  areaSownValue = "Value_measuredElement_5212",
  areaSownObsFlag = "flagObservationStatus_measuredElement_5212",
  seedRateValue = "Value_areaSownRatio",
  seedRateFlag = "flagObservationStatus_areaSownRatio", imputedFlag = "i",
  byKey = key(data))
```

Arguments

data	The data.table object containing the seed data, typically as created by getArea-Data.
seedValue	The column name of data which contains the value of the seed variable.
seedMethodFlag	The column name of data which contains the method flag of the seed variable.
seedObsFlag	The column name of data which contains the observation flag of the seed variable.
areaSownValue	The column name of data which contains the value of the area sown variable.
areaSownObsFlag	The column name of data which contains the observation flag of the area sown variable.
seedRateValue	The column name of data which contains the seed rate variable.
seedRateFlag	The column name of data which contains the seed rate flag.
imputedFlag	When the seed value is imputed, what character value should be assigned to the seedMethodFlag?
byKey	A character vector of the column name(s) of data which should be treated as keys for performing the imputation.

Value

No value is returned, instead data is modified by filling in missing values with imputed estimates.

imputeSeedDomain	<i>Impute Seed Domain</i>
------------------	---------------------------

Description

This function is a wrapper function which calls all the individual functions of this package. First, values where the area sown is 0 and the area harvested is non-zero (or vice-versa) are removed. Next, area sown is imputed, and then country specific and general seed rates are appended to the data.table. Lastly, the seed is imputed.

Usage

```
imputeSeedDomain(data, imputationParameters, byKey = NULL)
```

Arguments

data	The data.table object containing the seed data.
imputationParameters	See ?faoswsImputation::defaultImputationParameters. This is a list of arguments specifying how imputation should be done, and it pertains to imputation of the area sown variable. If imputation should be done by a simple mean, set imputationParameters to NULL.
byKey	In the case where imputationParameters is NULL, imputation is done by computing a mean within each group defined by byKey and using that mean for each missing value. byKey = NULL, the default, uses a global mean for each commodity. If imputationParameters is not NULL, this is ignored.

Value

No value is returned, but the data.table object is modified: seed is imputed and additional columns are returned.

saveSeedData	<i>Save Seed Data</i>
--------------	-----------------------

Description

This function takes the seed data dataset and saves it back to the database.

Usage

```
saveSeedData(data)
```

Arguments

data	The data.table object containing the seed data to be written to the database.
------	---

Value

No R objects are returned, as this function's purpose is solely to write to the database.

`seedData`*Example seed data for the vignette.*

Description

The data contain seed data from 8 countries, 140 commodities and spanning the years 1994-2013. The countries were not randomly chosen: some were picked because of the availability of area sown variables (which is generally rather rare).

Usage

```
data(seedData)
```

Format

A data.table object with 8960 rows and 12 variables

Details

The rather cryptic columns correspond to what exists in the SWS:

- `geographicAreaM49`: The M49 codes specifying the country
- `measuredItemCPC`: The codes specifying the commodity
- `timePointYears`: The year of the observation
- `Element_5212`: These three variables correspond to area sown. The value variable contains the data, and the `flagObservationStatus` and `flagMethod` variables contain the flags.
- `Element_5312`: Area Harvested
- `Element_5525`: Seed usage

Index

*Topic **datasets**

seedData, [8](#)

fillCountrySpecificSeedRate, [2](#)

fillGeneralSeedRate, [2](#)

getAreaData, [3](#)

getCountryGeneralSeedRate, [3](#)

getCountrySpecificSeedRate, [4](#)

imputeAreaSown, [4](#)

imputeSeed, [6](#)

imputeSeedDomain, [7](#)

saveSeedData, [7](#)

seedData, [8](#)