CODEBOOK

KaidenFrizu

Functions

datafunctions.R

`Dataextract()` A function that extracts the csv file through Google Drive Code into the local directory.

Syntax:

- `Dataextract(code, filename = "", overwrite = FALSE, na.assign = FALSE)`

Arguments:

- `code` - Refers to the code found in the shareable link of the file in Data Drop. The format of the shareable link would be `https://drive.google.com/file/d/##########################/view?usp=sharing` whereas the string of `#`s correspond to the code required.
- `filename` - A string argument that inputs the filename of the .csv file to be downloaded.
- `overwrite` - It is a logical argument whereas setting to `TRUE` will overwrite the existing file with the same `filename`.
- `na.assign` - A logical argument whether to set all blank cells with `NA`.

`Datasampling()` A function that returns a list of 2 data frames namely `TrainData` and `TestData`. This function is used for data modeling.

Syntax:

- `Datasampling(df, percent = 0.7, seednum = NA)`

Arguments:

- `df` - Refers to the data frame to be passed in the function.
- `percent` - a number between 0 to 1 that determines the percentage of the data that will be considered as train data. The remaining items will be considered test data. Default is 0.7.
- `seednum` - a number to be passed on `set.seed()` for reproduction. Default is `NA` for a random number.

---

filecontrol.R

`checkfolder()` A simple function that checks if the folder exists.

Syntax:

- `checkfolder(foldername)`

Argument:

- `foldername` - a string that checks its name as a folder whether it exists in the directory.

---

plotfunctions.R

`ggplot_missmap()` A simplified function derived from `ggplot_raster()` where it highlights missing data in the data frame.

Syntax:

- `ggplot_missmap(df, title = "", savefile = "")`

Arguments:

- `df` - Refers to the data frame to be passed in the function.
- `title` - A string argument that labels the graph title.
- `savefile` - A string argument that names the exported graph. The default is `""` where the graph will not be exported.

`ggplot_histogram()` A simplified function of `geom_histogram()` that outputs a histogram from the passed arguments.

Syntax:

- `ggplot_histogram(df, xaxis, title = "", Legend = NULL, pos = "identity", binnum = 30, xlabel = "", ylabel = "Number of Cases", savefile = "")`

Arguments:

- `df` - Refers to the data frame to be passed in the function.
- `xaxis` - A (column) vector that would be considered as x variable.

`title` - A string argument that labels the graph title.

- `Legend` - A (column) vector with the same length of the vector passed in `x-axis` that automatically groups the data by its corresponding grouping variable. Default is `NULL` where the data will not be grouped.
- `pos` - A variable passed to `geom_histogram(position)` where it determines the type of histogram to return. Possible values are `"identity"`, `"stack"`, and `"dodge"`.
- `binnum` - Refers to the number of bins to be shown in the histogram. Default is 30.
- `xlabel` - A string argument that labels the x-axis title.
- `ylabel` - A string argument that labels the y-axis title.
- `savefile` - A string argument that names the exported graph. The default is `""` where the graph will not be exported.

`ggplot_tsa()` A graph derived from `geom_line()` where it plots the curve of COVID-19 cases.

Syntax:

- `ggplot_tsa(df, title = "", dates, csum, xlabel, ylabel, savefile = "")`

Arguments:

- `df` - Refers to the data frame to be passed in the function.
- `title` - A string argument that labels the graph title.
- `dates` - A (column) vector that contains each unique days that at least occurred a COVID-19 case.
- `csum` - A (column) vector which has the same length of the vector passed in `dates` which contains the cumulative number of cases occurred in a specific date.
- `xlabel` - A string argument that labels the x-axis title.
- `ylabel` - A string argument that labels the y-axis title.
- `savefile` - A string argument that names the exported graph. The default is `""` where the graph will not be exported.

Values

`DownloadDate` - Contains the date and time where the csv file is downloaded. If file already exists, it returns the the `Date Created` of the file locally.

`urllink` - Outputs the direct link url of the csv file.