Using importData tools

Contents

Repo location

https://gitlab.com/Fair_lab/text_manipulation

Intro

Lack of consistency sharing data might lead to delays on analysis (see Wickham manuscript included here). The function import_tidyData_with_Dictionary aims to simplify the process if a few simple rules are followed.

import_tidyData_with_Dictionary is a function for Matlab that requires 2 mandatory csv files: a dictionary (Dictionary_filename) where each row has the variables' description which are the headres of the file with data (tidyData_filename).

This function expects the dictionary to have the following 5 mandatory columns:

1. Variable_name
2. Variable_type: only allowed values are alphanumeric or numeric
3. Description: Would be nice of you indicate what the variable means
4. Range: Expected range in values (it is fine if you provide no info)
5. Units: Self-descriptive (it is fine if you provide no info)

Data provided in the tidyData_filename must have as headers the variables indicated in the first columns of the Dictionary

Usage

You can give it a try using the provided example files and running the following commands on your Matlab session:

```
Dictionary_filename='Dictionary_for_random_data.csv';
tidyData_filename='random_data.csv';
[tidyData, Dictionary] = import_tidyData_with_Dictionary(Dictionary_filename,tidyData_filename);
```

Show the size of the outputs (ie, tidyData and Dictionary)

```
whos tidyData Dictionary
```

```
Name              Size            Bytes  Class    Attributes

Dictionary         8x5             6207  table
tidyData         129x8           126615  table
```

Show the content of the dictionary

```
Dictionary
```

```
Dictionary =

  8×5 table

    Variable_name     Variable_type     Description     Range     Units
    _____     _____     _____     _____     _____

    "SubjectID"       'alphanumeric'     ' '            ' '       ' '
    "Sex"             'alphanumeric'     ' '            ' '       ' '
    "test"            'alphanumeric'     ' '            ' '       ' '
    "Freq"            'numeric'          ' '            ' '       'Hz'
    "AMP"             'numeric'          ' '            ' '       'pA'
    "Rise Time"       'numeric'          ' '            ' '       'ms'
    "Decay Time"      'numeric'          ' '            ' '       'ms'
    "AREA"            'numeric'          ' '            ' '       'fC'
```

Show the first elements of the actual data

```
tidyData(1:10,:)
```

```
ans =

  10×8 table

    SubjectID    Sex      test       Freq        AMP       Rise_Time    Decay_Time     AREA
    _____    ___    _____    _____    _____    _____    _____    _____

    '31287'      'M'    'PU IPSC'   [0.1068]    [1.1928]    [2.1415]     [0.7839]      [0.9184]
    '31503'      'F'    'PU IPSC'   [2.4957]    [6.1997]    [2.2471]     [1.9798]      [1.5533]
    '32041'      'F'    'PU IPSC'   [5.4677]    [1.5826]    [0.3239]     [2.9823]      [0.1011]
    '32050'      'F'    'PU IPSC'   [3.5536]    [5.8354]    [0.9491]     [0.7901]      [0.4555]
    '32228'      'M'    'PU IPSC'   [0.5190]    [0.4047]    [0.4674]     [0.0294]      [1.0455]
    '32207'      'M'    'PU IPSC'   [0.3884]    [0.3260]    [1.1966]     [2.7089]      [0.5228]
    '32269'      'M'    'PU IPSC'   [0.7380]    [0.6823]    [1.0176]     [2.1150]      [0.8611]
    '31418'      'M'    'PU IPSC'   [0.0653]    [1.9187]    [1.3155]     [2.2006]      [3.5560]
    '31348'      'F'    'PU IPSC'   [1.2265]    [0.9818]    [0.1772]     [4.7574]      [2.0087]
    '31970'      'M'    'PU IPSC'   [3.3117]    [0.4047]    [8.7378]     [0.2701]      [0.5576]
```

More details

As you can see, the file Dictionary_for_random_data has the 5 mandatory columns that describe the data that will be imported. The second column of the dictionary (Variable_type)is used to format the data contained in random_data.csv. This Dictionary also

indicates that the data has 8 variables:

1. SubjectID,

2. Sex,

3. test,

4. Freq,

5. AMP,

6. Rise Time,

7. Decay Time, and

8. AREA.

As expected, the file random_data.csv has 8 columns, each corresponding to the variables described in the dictionary.

Post usage recomendations

Do not forget to double check and compare the content of the outputs tidyData and Dictionary with the files random_data.csv and Dictionary_for_random_data.csv respectively

Credits

Developer: Oscar Miranda-Dominguez

First line of code: Jan 24, 2019

Published with MATLAB® R2017b