Using tsdl Library

The Time Series Data Library is no longer hosted on DataMarket, according to Prof. Rob Hyndmans, but an alternative has been provided. The data is still accessible via the `tsdl` R package.

1. Installing `tsdl`

There are two main approaches to install this package.

Approach 1

Executing the chunk below.

```
install.packages("devtools", repos = "http://cran.us.r-project.org") # this package provides the function `install_github()` as used in the next line
devtools::install_github("FinYang/tsdl") # install the `tsdl` package from github
```

Once you have successfully installed the `tsdl` package, there is no need to install it again. You will only need to load the package with `library(tsdl)` each time you use it.

Approach 2

If Approach 1 does not work for you, try downloading the package file directly and installing it locally.

1. Download the package at https://api.github.com/repos/FinYang/tsdl/tarball/HEAD (https://api.github.com/repos/FinYang/tsdl/tarball/HEAD)
2. A file named `FinYang-tsdl-56e0915.tar.gz` should be downloaded.
3. Open R Studio and click **Packages** in the lower right-hand panel, then click **Install**.
4. Change **Install From** from **Repository(CRAN)** to **Package Archive File(.tgz; .tar.gz)**.
5. Click **Browse** and choose **FinYang-tsdl-56e0915.tar.gz** which is stored somewhere in your computer's directory.
6. Click **Install**.

2. Working with `tsdl`

There are 648 univariate time series covering various subjects. One can use the function `attr()` with arguments `source` , `description` and `subject` to get the basic information of where the data is obtained, additional description of it, as well as the subject of each time series. See below for a summary of the subjects.

```
library(tsdl)
tsdl
```

```
## Time Series Data Library: 648 time series
##
##                         Frequency
## Subject                  0.1 0.25   1   4   5   6  12  13  52 365 Total
##   Agriculture              0    0  37   0   0   0   3   0   0   0    40
##   Chemistry                0    0   8   0   0   0   0   0   0   0     8
##   Computing                0    0   6   0   0   0   0   0   0   0     6
##   Crime                    0    0   1   0   0   0   2   1   0   0     4
##   Demography               1    0   9   2   0   0   3   0   0   2    17
##   Ecology                  0    0  23   0   0   0   0   0   0   0    23
##   Finance                  0    0  23   5   0   0  20   0   2   1    51
##   Health                   0    0   8   0   0   0   6   0   1   0    15
##   Hydrology                0    0  42   0   0   0  78   1   0   6   127
##   Industry                 0    0   9   0   0   0   2   0   1   0    12
##   Labour market            0    0   3   4   0   0  17   0   0   0    24
##   Macroeconomic            0    0  18  33   0   0   5   0   0   0    56
##   Meteorology              0    0  18   0   0   0  17   0   0  12    47
##   Microeconomic            0    0  27   1   0   0   7   0   1   0    36
##   Miscellaneous            0    0   4   0   1   1   3   0   1   0    10
##   Physics                  0    0  12   0   0   0   4   0   0   0    16
##   Production               0    0   4  14   0   0  28   1   1   0    48
##   Sales                    0    0  10   3   0   0  24   0   9   0    46
##   Sport                    0    1   1   0   0   0   0   0   0   0     2
##   Transport and tourism    0    0   1   1   0   0  12   0   0   0    14
##   Tree-rings               0    0  34   0   0   0   1   0   0   0    35
##   Utilities                0    0   2   1   0   0   8   0   0   0    11
##   Total                    1    1 300  64   1   1 240   3  16  21   648
```

Please keep in mind the order of data on different subjects are not saved exactly as in the above summary table. For example, if we are interested in the subjects of the first 5 time series, a for loop will do the job:

```
first_few = 5
for(i in 1:first_few) cat("i =", i, attr(tsdl[[i]], "subject"), "\n")
```

```
## i = 1 Macroeconomic
## i = 2 Industry
## i = 3 Utilities
## i = 4 Sales
## i = 5 Macroeconomic
```

Increasing `first_few` and running the above code would get us the index of the data on the specific subject we are interested in.

Suppose we want to use the 9th time series, indexed by 9 in the data set. On execution of the following chunk, we know this time series, having 108 observations, is a monthly car sales record that was taken in Quebec from 1960 to 1968. The original source of the data is from *Statistical Methods for Forecasting, by Abraham, B. and Ledolter, J. (1983).*

```
# length of the data
length(tsdl[[9]])
```

```
## [1] 108
```

```
# subject of the data
attr(tsdl[[9]], "subject")
```

```
## [1] "Sales"
```

```
# source of the data
attr(tsdl[[9]], "source")
```

```
## [1] "Abraham & Ledolter (1983)"
```

```
# description of the data
attr(tsdl[[9]], "description")
```

```
## [1] "Monthly car sales in Quebec 1960-1968"
```

In addition to the above commands, this Introduction to the tsdl package (https://pkg.yangzhuoranyang.com/tsdl/articles/tsdl.html (https://pkg.yangzhuoranyang.com/tsdl/articles/tsdl.html)) written by Yangzhuoran Yang demonstrates the usage of the `subset()` function to extract a specific time series by frequency, subject, or both.

Enjoy your data analysis!