Random week sample generator

Vivien Benert

2023-02-06

# library(dplyr)  
# library(lubridate)

# Function to draw artificial random week from sample period

The function needs a start date and an end date, provided in the format yyyy-mm-dd. Additionally, there is the option to specify whether the sample should be drawn from the whole week or only from weekdays. Default is set to take the whole week as the basis for the random sample, including weekends. If weekends should be excluded (i.e. if the random sample should only inlcude mondays to fridays, set the optional weekends-argument to FALSE).

sample\_week <- function(start.date, end.date, weekend=TRUE) {  
 require(dplyr)  
 require(lubridate)  
 # Create sequence of days for sample period  
 df <- as.data.frame(seq(ymd(start.date), ymd(end.date), by='days'))  
 colnames(df) <- "date"  
   
 # Create weekday and year variable with lubridate   
 df$weekday <- wday(df$date)  
 df$year <- year(df$date)  
 df$week <- week(df$date)  
   
 # Select only weekdays, only if weekends=FALSE  
 if (weekend==FALSE) {  
 x <- 5  
 } else {  
 x <- 7  
 }  
  
 df <- df[which(df$weekday<=x),]  
   
 # Subset years, returns a list of dfs  
 df\_list <- split(df, f=df$year)  
   
 # Create empty data frame  
 sample <- data.frame(matrix(ncol=4, nrow=0))  
 colnames(sample) <- c("date", "weekday", "year", "week")  
  
 for (i in 1:length(df\_list)) {  
 x <- df\_list[[i]]  
 z <- x %>%   
 group\_by(week) %>%   
 sample\_n(1) %>%   
 ungroup()  
 sample <- rbind(sample, z)  
 }  
  
 return(sample)  
}

# Use function by providing start date and end date as yyyy-mm-dd format  
# Don't forget to set.seed if sample needs to be reproducible.  
  
# set.seed(12345)  
  
# sample\_week <- sample\_week('1955-06-08', '2023-01-01')  
# sample\_weekdays <- sample\_week('1989-05-12', '2022-10-30', weekend = FALSE)