merging.R

riserate

Fri Jul 14 12:48:25 2017

#set the directory where the files are located  
#setwd()  
  
#import the data sets  
d1 <- read.csv2("GT\_EM\_5yr\_US.csv", sep = ",") #This file has row with line "Catagory" you need to manually remove it.  
d2 <- read.csv2("Wiki\_Em.csv", sep = ",")  
d3 <- read.csv("daily\_price.csv", sep =",")  
  
#check the data  
head(d1)

## Week Ed  
## 1 7/8/12 3  
## 2 7/15/12 0  
## 3 7/22/12 0  
## 4 7/29/12 0  
## 5 8/5/12 0  
## 6 8/12/12 3

head(d2)

## Date C EM  
## 1 7/1/15 2304 0  
## 2 7/2/15 2304 0  
## 3 7/3/15 1628 0  
## 4 7/4/15 1230 0  
## 5 7/5/15 1216 0  
## 6 7/6/15 2056 0

head(d3)

## Date Open High Low Close Adj.Close Volume  
## 1 1/16/16 NA NA NA NA NA NA  
## 2 1/17/16 NA NA NA NA NA NA  
## 3 1/18/16 NA NA NA NA NA NA  
## 4 1/19/16 NA NA NA NA NA NA  
## 5 1/20/16 NA NA NA NA NA NA  
## 6 1/21/16 NA NA NA NA NA NA

#need libraries  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

#change column name in d1 to match so we can merge then we can change it back after the merge if needed.  
names(d1)[names(d1) == 'Week'] <- 'Date'  
  
#check the d1 data  
head(d1)

## Date Ed  
## 1 7/8/12 3  
## 2 7/15/12 0  
## 3 7/22/12 0  
## 4 7/29/12 0  
## 5 8/5/12 0  
## 6 8/12/12 3

#format the column to be a date, R thinks they are just characters  
d1$Date <- as.Date(d1$Date,format="%m/%d/%y")  
d2$Date <- as.Date(d2$Date,format="%m/%d/%y")  
d3$Date <- as.Date(d3$Date,format="%m/%d/%y")  
  
#merge the data sets  
mydata <- merge(d1, d2, by.x = "Date", by.y = "Date", all = TRUE, sort = TRUE)  
mydata <- merge(mydata, d3, by.x = "Date", by.y = "Date", all = TRUE, sort = TRUE)  
  
#View the final data set  
head(mydata)

## Date Ed C EM Open High Low Close Adj.Close Volume  
## 1 2012-07-08 3 NA NA NA NA NA NA NA NA  
## 2 2012-07-15 0 NA NA NA NA NA NA NA NA  
## 3 2012-07-22 0 NA NA NA NA NA NA NA NA  
## 4 2012-07-29 0 NA NA NA NA NA NA NA NA  
## 5 2012-08-05 0 NA NA NA NA NA NA NA NA  
## 6 2012-08-12 3 NA NA NA NA NA NA NA NA

tail(mydata)

## Date Ed C EM Open High Low Close Adj.Close Volume  
## 888 2017-07-01 NA 2856 15 NA NA NA NA NA NA  
## 889 2017-07-02 12 3086 15 NA NA NA NA NA NA  
## 890 2017-07-03 NA 3623 17 16.78 17.50 16.78 17.43 17.43 173300  
## 891 2017-07-04 NA 4875 20 NA NA NA NA NA NA  
## 892 2017-07-05 NA 3952 27 17.35 17.93 17.07 17.51 17.51 428600  
## 893 2017-07-06 NA 3876 19 17.57 17.60 16.90 17.02 17.02 334200

#Export new csv file  
write.csv(mydata, file = "merged.csv", row.names = FALSE)