complete.R

Cliff

Sun May 24 15:41:38 2015

complete <- function(directory, id = 1:332) {  
   
 browser()  
   
 # Write a function that reads a directory full of files and reports the number of completely   
 # observed cases in each data file. The function should return a data frame where the first   
 # column is the name of the file and the second column is the number of complete cases. A prototype   
 # of this function follows   
   
 ## 'directory' is a character vector of length 1 indicating  
 ## the location of the CSV files  
   
 ## 'id' is an integer vector indicating the monitor ID numbers  
 ## to be used  
   
 ## Return a data frame of the form:  
 ## id nobs  
 ## 1 117  
 ## 2 1041  
 ## ...  
 ## where 'id' is the monitor ID number and 'nobs' is the  
 ## number of complete cases  
   
 #Must return a data frame so I might as well create an empty one  
 #NOT Efficient See http://stackoverflow.com/questions/13442461/populating-a-data-frame-in-r-in-a-loop  
 df <- data.frame(id= integer(0), nobs= integer(0))  
   
 #Get the requested files and start a for loop  
 files <- list.files(directory, full.names=T)  
 requestedfiles <- files[id]#replace id  
 numoffiles <- length(requestedfiles)  
   
 for (i in 1:numoffiles){  
 #read the file  
 curfiletxt <- read.csv(requestedfiles[i])  
 #count the completed records  
 countcomplete <- sum(complete.cases(curfiletxt))  
 #build the list  
 #print(paste(i, "value of i counter", sep = " "))  
 #print(paste(countcomplete, "value of complete records", sep = " "))  
 df[i, 1] <- i  
 df[i, 2] <- countcomplete  
   
 #Might need to try this  
 #dataframe <- rbind(dataframe, data.frame(id=id, nobs=nobs))  
 countcomplete <- 0  
 curfiletxt <-data.frame()  
 }  
 print(df)  
}  
  
#Test data here: https://d396qusza40orc.cloudfront.net/rprog%2Fdoc%2Fcomplete-demo.html  
# complete("specdata", c(2, 4, 8, 10, 12))  
# complete("specdata", 30:25)