<https://github.com/Xiaodan/Coursera-R-Programming/blob/master/week2/pollutantmean.R>

|  |
| --- |
| setwd("~/Desktop/Online Coursera/Coursera-R-Programming/week2/") |
|  | #getwd() |
|  | #list.files() |
|  |  |
|  | pollutantmean <- function(directory, pollutant = "sulfate", id = 1:332) { |
|  | ## 'directory' is a character vector of length 1 indicating |
|  | ## the location of the CSV files |
|  |  |
|  | ## 'pollutant' is a character vector of length 1 indicating |
|  | ## the name of the pollutant for which we will calculate the |
|  | ## mean; either "sulfate" or "nitrate". |
|  |  |
|  | ## 'id' is an integer vector indicating the monitor ID numbers |
|  | ## to be used |
|  |  |
|  | ## Return the mean of the pollutant across all monitors list |
|  | ## in the 'id' vector (ignoring NA values) |
|  |  |
|  | # set working directory |
|  | if(grep("specdata", directory) == 1) { |
|  | directory <- ("./specdata/") |
|  | } |
|  | # initialize a vector to hold the pollutant data |
|  | mean\_vector <- c() |
|  | # find all files in the specdata folder |
|  | all\_files <- as.character( list.files(directory) ) |
|  | file\_paths <- paste(directory, all\_files, sep="") |
|  | for(i in id) { |
|  | current\_file <- read.csv(file\_paths[i], header=T, sep=",") |
|  | head(current\_file) |
|  | pollutant |
|  | na\_removed <- current\_file[!is.na(current\_file[, pollutant]), pollutant] |
|  | mean\_vector <- c(mean\_vector, na\_removed) |
|  | } |
|  | result <- mean(mean\_vector) |
|  | return(round(result, 3)) |
|  | } |
|  |  |
|  | # tests |
|  | pollutantmean("specdata", "sulfate", 1:10) == 4.064 |
|  | pollutantmean("specdata", "nitrate", 70:72) == 1.706 |
|  | pollutantmean("specdata", "nitrate", 23) == 1.281 |

<https://github.com/Xiaodan/Coursera-R-Programming/blob/master/week2/complete.R>

|  |
| --- |
| complete <- function(directory, id = 1:332) { |
|  | ## '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 |
|  |  |
|  | # set working directory |
|  | if(grep("specdata", directory) == 1) { |
|  | directory <- ("./specdata/") |
|  | } |
|  | # get the length of id vector |
|  | id\_len <- length(id) |
|  | complete\_data <- rep(0, id\_len) |
|  | # find all files in the specdata folder |
|  | all\_files <- as.character( list.files(directory) ) |
|  | file\_paths <- paste(directory, all\_files, sep="") |
|  | j <- 1 |
|  | for (i in id) { |
|  | current\_file <- read.csv(file\_paths[i], header=T, sep=",") |
|  | complete\_data[j] <- sum(complete.cases(current\_file)) |
|  | j <- j + 1 |
|  | } |
|  | result <- data.frame(id = id, nobs = complete\_data) |
|  | return(result) |
|  | } |
|  |  |
|  | # tests |
|  | complete("specdata", 1) |
|  | complete("specdata", c(2, 4, 8, 10, 12)) |
|  | complete("specdata", 30:25) |
|  | complete("specdata", 3) |

<https://github.com/Xiaodan/Coursera-R-Programming/blob/master/week2/corr.R>

|  |
| --- |
| setwd("~/Desktop/Online Coursera/Coursera-R-Programming/week2/") |
|  |  |
|  | corr <- function(directory, threshold = 0) { |
|  | ## 'directory' is a character vector of length 1 indicating |
|  | ## the location of the CSV files |
|  |  |
|  | ## 'threshold' is a numeric vector of length 1 indicating the |
|  | ## number of completely observed observations (on all |
|  | ## variables) required to compute the correlation between |
|  | ## nitrate and sulfate; the default is 0 |
|  |  |
|  | ## Return a numeric vector of correlations |
|  |  |
|  | if(grep("specdata", directory) == 1) { |
|  | directory <- ("./specdata/") |
|  | } |
|  | # get the complete table |
|  | complete\_table <- complete("specdata", 1:332) |
|  | nobs <- complete\_table$nobs |
|  | # find the valid ids |
|  | ids <- complete\_table$id[nobs > threshold] |
|  | # get the length of ids vector |
|  | id\_len <- length(ids) |
|  | corr\_vector <- rep(0, id\_len) |
|  | # find all files in the specdata folder |
|  | all\_files <- as.character( list.files(directory) ) |
|  | file\_paths <- paste(directory, all\_files, sep="") |
|  | j <- 1 |
|  | for(i in ids) { |
|  | current\_file <- read.csv(file\_paths[i], header=T, sep=",") |
|  | corr\_vector[j] <- cor(current\_file$sulfate, current\_file$nitrate, use="complete.obs") |
|  | j <- j + 1 |
|  | } |
|  | result <- corr\_vector |
|  | return(result) |
|  | } |
|  |  |
|  | # tests |
|  | cr <- corr("specdata", 150) |
|  | head(cr) |
|  | cr <- corr("specdata", 400) |
|  | head(cr) |
|  | cr <- corr("specdata", 5000) |
|  | summary(cr) |