CourseraAssignment1

Iris Shvartsman

July 30, 2019

Part 1

Create a function named pollutantmean that calculates the mean of a pollutant in a specified list of observations.

```
pollutantmean <- function(directory, pollutant, id=1:332){
  filenames <- list.files(directory, full.names = TRUE)
  #create an empty data frame to populate
  emptydf <- data.frame()

  for(i in id){
    #reading in a temporary data frame for the filenames i called in my function
    df <- read.csv(filenames[i], header = TRUE)
    #combine the temporary df with the emptydf created
    emptydf <- rbind(emptydf, df)
  }
  return(mean(emptydf[,pollutant],na.rm = TRUE))
}


#Example:
pollutantmean(directory= "C:/Users/iriss/Documents/Iris/Courseradatascience/specdata/",pollutant
= "nitrate", 1:10)
```

```
## [1] 0.7976266
```

Part 2

Create a function that reads files and reports the number of observed cases in each file specified.

```
complete <- function(directory, id=1:332){
  filenames <- list.files(directory, full.names = TRUE)
  #create an empty data frame to populate
  emptydf <- data.frame()

  for(i in id){
    df <- na.omit(read.csv(filenames[i], header = TRUE))
    dfobs <- nrow(df)
    emptydf <- rbind(emptydf, data.frame(i, dfobs))
  }
  return(emptydf)
}


#Example:
complete(directory = "C:/Users/iriss/Documents/Iris/Courseradatascience/specdata/", 30:25)
```

```
##    i dfobs
## 1 30   932
## 2 29   711
## 3 28   475
## 4 27   338
## 5 26   586
## 6 25   463
```

Part 3

Write a function that takes a directory of files and a threshold for complete cases, calculates the correlation between sulfate and nitrate for monitor locations where the number of completely observed cases is greater than the threshold. **The function should return a vector**

```
corr <- function(directory, threshold = 0){
  #read in the list of files like in the previous questions:
  filenames <- list.files(directory, full.names = TRUE)
  #create empty vector, instead of df
  vec <- vector(mode = "numeric")

  for (i in 1:length(filenames)){
    tempdf <- read.csv(filenames[i], header = TRUE)
    tempdf <- na.omit(tempdf) #remove NA observations
    #count the number of rows to check if it is greater than the threshold
    cnt <- nrow(tempdf)
    if(cnt>threshold){
      #if the number of rows is greater than the threshold then
      #return the correlation of nitrate and sulfate
      vec <- c(vec, cor(tempdf$nitrate, tempdf$sulfate))
    }
  }

  return(vec)

}


#Example:
c <- corr(directory = "C:/Users/iriss/Documents/Iris/Courseradatascience/specdata/", 150)
```