Programming Project1

Saydaliev

3/31/2020

```
install.packages("data.table")
library('data.table')


pollutantmean<-function(directory, pollutant, id=1:332) {
  fileNames<-paste0(directory, '/', formatC(id, width=3,flag = '0'), ".csv")
  lst<-lapply(fileNames, data.table:: fread)
  dt<-rbinlist(lst)
  if(c(pollutant) %in% names(dt)) {
    return(dt[,lapply(.SD, mean, na.rm=TRUE), .SDcols=pollutant][[1]])
  }
}
#USAGE
pollutantmean(directory = '~/Desktop/specdata', pollutant = 'sulfate', id = 20))
```

Part 2

```
complete <- function(directory,  id = 1:332) {

fileNames <- paste0(directory, '/', formatC(id, width=3, flag="0"), ".csv" )

lst <- lapply(fileNames, data.table::fread)
  dt <- rbindlist(lst)

  return(dt[complete.cases(dt), .(nobs = .N), by = ID])

  #USAGE
  complete(directory = '~/Desktop/specdata', id = 20:30)
```

Part 3

```
corr <- function(directory, threshold = 0) {

  lst <- lapply(file.path(directory, list.files(path = directory, pattern="*.csv")), data.table::fread)
  dt <- rbindlist(lst)

  dt <- dt[complete.cases(dt),]

  dt <- dt[, .(nobs = .N, corr = cor(x = sulfate, y = nitrate)), by = ID][nobs > threshold]
  return(dt[, corr])
}

# Example Usage
corr(directory = '~/Desktop/specdata', threshold = 150)
```