pollutantmean <- function(directory = "~/r", pollutant = "sulfate", id = 1:332) {

folderlocation <- dir()

generaldf <- do.call(rbind, lapply(folderlocation, read.csv))

snidf <- generaldf[, 2:4]

sidf <- generaldf[, c(2, 4)]

nidf <- generaldf[, 3:4]

nidsort <- nidf[is.element(nidf$ID, id),]

sidsort <- sidf[is.element(sidf$ID, id),]

summary(nidsort)

onlyn <- nidsort[, 1, drop=FALSE]

onlys <- sidsort[, 1, drop=FALSE]

summary(onlyn)

nmeanable <- na.omit(onlyn)

smeanable <- na.omit(onlys)

if (pollutant == "sulfate") {

colMeans(smeanable)

}

if (pollutant == "nitrate" ){

colMeans(nmeanable) }

}

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

folderlocation <- dir()

generaldf <- do.call(rbind, lapply(folderlocation, read.csv))

snidf <- generaldf[, 2:4]

#for (i in id)

#{

# snidsort <- snidf[is.element(snidf$ID, i),]

# tnumber <- complete.cases((snidsort))

# print(paste("The Number Of Complete Cases For", i, "is", sum(tnumber)))

#}

j <- NULL

for (i in id)

{

snidsort <- snidf[is.element(snidf$ID, i),]

tnumber <- complete.cases((snidsort))

k <- c(i, sum(tnumber))

j <- rbind(j, data.frame(i, sum(tnumber)))

}

print(paste("the size of dims for j is", dim.data.frame(j)))

colnames(j) <- c("id", "nobs")

print.data.frame(j)

}

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

folderlocation <- dir()

generaldf <- do.call(rbind, lapply(folderlocation, read.csv))

snidf <- generaldf[, 2:4]

j <- NULL

id <- 1:332

for (i in id)

{

snidsort <- snidf[is.element(snidf$ID, i),]

tnumber <- complete.cases((snidsort))

k <- c(i, sum(tnumber))

j <- rbind(j, data.frame(i, sum(tnumber)))

}

colnames(j) <- c("id", "nobs")

#print.data.frame(j)

corrlistdf <- subset(j, j[,2] > threshold)

corrnumbers <- corrlistdf[["id"]]

cr <- character()

for (i in corrnumbers) {

snidsort <- snidf[is.element(snidf$ID, i),]

u <- snidsort$sulfate

v <- snidsort$nitrate

p <- cor(u, v, use = "complete")

cr <- append(cr, p)

}

ncr <- as.numeric(cr)

length(ncr)