Assignment 3

1. Created a function named pollutantmean with arguments Directory, Pollutant and ID.

List.files(object,full.name=T) returns the files inside specdata and stores it inside an object called Filesname

DF <- lapply(Filenames[ID],read.csv,header=T) returns a list of dataframe which contains all the .csv files that has been asked to import in ID. If the ID is like “1:10” it will return mean of that pollutant across 1:10.

DF <- do.call("rbind",DF) return a dataframe that binds all the .csv files rowwise.

Return(mean(DF[,Pollutant],na.rm = TRUE)) returns the mean

1. Created a function named CompleteCases with arguments Directory and ID.

CompleteCounts <- function(Filenames) sum(complete.cases(read.csv(Filenames))) is a function that returns sum of the completecases after reading the file.

Filenames <- list.files(Directory, full.names=TRUE)[ID]

returns the files inside specdata and stores it inside an object called Filesname.

data.frame(ID = ID, CompleteCases =unlist(lapply(Filenames, CompleteCounts))) returns a dataframe containing ID and counts of Complete cases.

1. Created a function named Cor with arguments Directory and Threshold.

CC <- CompleteCases(Directory) returns all 332 dataframe containing the complete cases of each .csv file.

CCTM <- CC[CC$CompleteCases>Threshold,]$ID returns all those ID’s that are greater than Threshold.

Filenames <- paste(Directory,"/", formatC(CCTM, width = 3, flag = "0"),".csv", sep = "") return character vector with format like “specdata/003.csv”.

cor <- numeric() initializing a numeric vector.

for (i in 1:length(CCTM)) {

Read = read.csv(Filenames[i])

DF = Read[complete.cases(Read), ]

cor = c(cor, cor(DF$sulfate, DF$nitrate))

}

Returns the correlation coefficient of sulfate and nitrate for every those csv file whose thresholds are met.