setwd("C:/Users/PC/Documents/Projects/College Assignments/RProject")

rdata=read.csv("dataset.csv")

rdata1=as.data.frame(rdata)

View(rdata1)

class(rdata1)

library(zoo)

library(tidyr)

library(dplyr)

#for (i in seq(1,5744))+

# if(rdata1$Air.Quality[i] == "--")

# rdata1$Air.Quality[i]<-TRUE

#for (i in seq(1,5744))

#if(rdata1$O3.Quality[i] == "--")

# rdata1$O3.Quality[i]<-"NA"

#for (i in seq(1,5744))

#if(rdata1$NO2.Quality[i] == "--")

# rdata1$NO2.Quality[i]<-TRUE

#for (i in seq(1,5744))

# if(rdata1$PM10.Quality[i] == "--")

# rdata1$PM10.Quality[i]<-"NA"

# Repacing all the blank values with NA

rdata1$NO2.Quality[rdata1$NO2.Quality=="--"]<-NA

rdata1$O3.Quality[rdata1$O3.Quality=="--"]<-NA

rdata1$PM10.Quality[rdata1$PM10.Quality=="--"]<-NA

rdata1$Air.Quality[rdata1$Air.Quality=="--"]<-NA

# Replacing all the NA values with the previous value

dfair<-as.data.frame(rdata1[,2])

rdata1[,2]<-dfair%>%do(na.locf(.))

dfoxy<-as.data.frame(rdata1[,c(5,6)])

rdata1[,c(5,6)]<-dfoxy%>%do(na.locf(.))

dfoxy1<-as.data.frame(rdata1[,7])

rdata1[,7]<-dfoxy1%>%do(na.locf(.))

dfnitro<-as.data.frame(rdata1[8,9])

rdata1[,c(8,9)]<-dfoxy%>%do(na.locf(.))

dfnitro1<-as.data.frame(rdata1[,10])

rdata1[,10]<-dfnitro1%>%do(na.locf(.))

dfpm<-as.data.frame(rdata1[,c(11,12)])

rdata1[,c(11,12)]<-dfoxy%>%do(na.locf(.))

dfpm1<-as.data.frame(rdata1[,13])

rdata1[,13]<-dfpm1%>%do(na.locf(.))

rdata1$Latitude<-as.numeric(rdata1$Latitude)

# correcting the outliers in latitude column

for (i in seq(1,5744)){

if(rdata1$Latitude[i] > 42)

rdata1$Latitude[i]=rdata1$Latitude[i]/10000

}

# normalizing the PM10.Value column

mx = max(rdata1$PM10.Value, na.rm=TRUE)

print(mx)

mn = min(rdata1$PM10.Value, na.rm=TRUE)

print(mn)

for (i in seq(1,5744)){

rdata1$PM10.Value[i]=round((rdata1$PM10.Value[i]-mn)/(mx-mn),digits=3)

}

#Visualisations

a = data.frame()

b = data.frame()

c = data.frame()

d = data.frame()

e = data.frame()

f = data.frame()

g = data.frame()

h = data.frame()

for(i in seq(1,5744,8)){

a = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(2,5744,8)){

b = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(3,5744,8)){

c = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(4,5744,8)){

d = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(5,5744,8)){

e = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(6,5744,8)){

f = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(7,5744,8)){

g = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

for(i in seq(8,5744,8)){

h = rbind(a,c(rdata1$Station[i],rdata1$Latitude[i],rdata1$Longitude[i],rdata1$NO2.Value[i],rdata1$O3.Value[i]))

}

library(ggplot2)

sp2 = ggplot(rdata1, aes(x=Latitude, y=Longitude, color=NO2.Value)) + geom\_point()

print(sp2)