**Import Dataset from MongoDB**

> library(mongolite)

**Import collection “HomeTemp” from database “bigdata\_final\_zeyan\_liow”**

> m <- mongo(collection = "HomeTemp", db = "bigdata\_final\_zeyan\_liow", url = "mongodb://localhost")

> hTemp <- m$find(limit = 1000000, skip = 0, fields = '{ "\_id" : true, "temp" : true }')

**Import collection “HomePower” from database “bigdata\_final\_zeyan\_liow”**

> m1 <- mongo(collection = "HomePower", db = "bigdata\_final\_zeyan\_liow", url = "mongodb://localhost")

> hPower <- m1$find(limit = 1000000, skip = 0, fields = '{ "\_id" : true, "power" : true }')

**Import collection “Dublin” from database “bigdata\_final\_zeyan\_liow”**

> m2 <- mongo(collection = "Dublin", db = "bigdata\_final\_zeyan\_liow", url = "mongodb://localhost")

> dublin <- m2$find(limit = 1000000, skip = 0, fields = '{ "\_id" : false, "date" : true, "i" : true, "" : true, "it emp" : true, "temp" : true, "iwb" : true, "wetb" : true, "dewpt" : true, "vappr" : true, "rhum" : true, "msl" : true, "iwdsp" : true, "wdsp" : true, "iwddir" : true, "wddir" : true, "ww" : true, "w" : true, "sun" : true, "vi s" : true, "clht" : true, "clamt" : true }')

**Create a function call is.POSIXct**

is.POSIXct <- function(x) inherits(x, "POSIXct")

**Change any variables in Dataset “hTemp” & “hPower” from type “character” to “POSIXct”**

> hTemp <- lapply(hTemp, function(x) if(is.character(x)) as.POSIXct(x,format="%Y-%m-%d %H:%M:%S") else x)

> hPower <- lapply(hPower, function(x) if(is.character(x)) as.POSIXct(x,format="%Y-%m-%d %H:%M:%S") else x)

**Change column names in Dataset “hTemp” & “hPower”**

> names(hTemp) <- c("date", "hmTemp")

> names(hPower) <- c("date", "hmPow")

**This will automatically merge datasets together by checking if it they have same column name and column type**

> merge1 <- merge(hTemp,hPower)

> data <- merge(merge1,dublin)

**Insert month, day and hour into dataset “data”**

month <- (format(as.Date(data$date), "%m"))

> month <- (format(as.Date(data$date), "%m"))

> month <- as.numeric(month)

> data <- cbind(data, month)

> day <- (format(as.Date(data$date), "%d"))

> day <- as.numeric(day)

> data <- cbind(data, day)

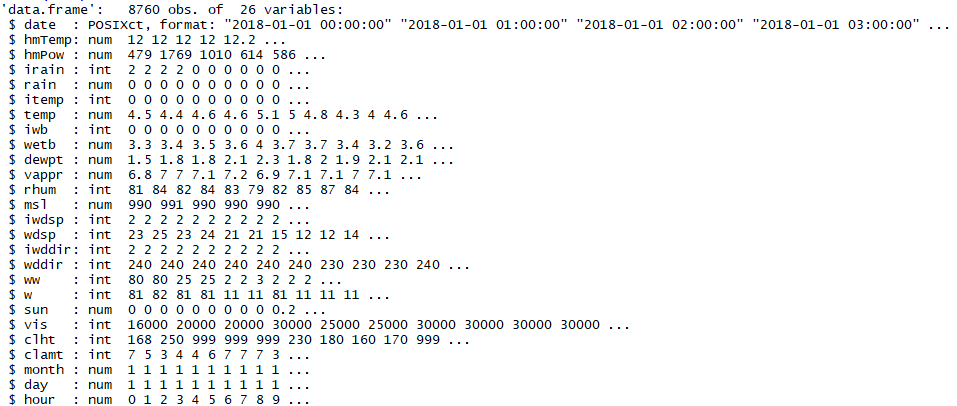
> hour <- (format(as.POSIXct(strptime(data$date,"%Y-%m-%d %H:%M:%S",tz="")) ,format = "%H"))

> hour <- as.numeric(hour)

> data <- cbind(data, hour)

**Data exploration**

> str(data)



**Check for missing data and outliers**

> length(which(is.na(data)==TRUE))

[1] 0

> summary(data)

