

Daily Time Lags

Mahalia Clark

12/2/2019

Calculating Daily Averages

First I read in the data: data frames of hourly buoy data. I then created a function which takes our hourly data and returns a data frame of daily averages. It also appends columns with the PC averages 0-6 days into the future (so that all the environmental variables are lagged). Finally, I ran the function on our data so we could use daily averages for subsequent analyses.

```
# Load required packages
library(dplyr)
library(ggplot2)
#install.packages("openair")
#install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
library(openair)

# Load data
MB17 <- read.csv("data/AllDepthData/MB17_allDepths.csv", header = T)
MB18 <- read.csv("data/AllDepthData/MB18_allDepths.csv", header = T)
SA17 <- read.csv("data/AllDepthData/SA17_allDepths.csv", header = T)
SA18 <- read.csv("data/AllDepthData/SA18_allDepths.csv", header = T)

# Function that takes a df of hourly data and returns a df of daily averages,
# including columns with daily PC 0-6 days into the future (foolishly named PC1-7).
dailify <- function(df_in, time_header){
  df_out <- df_in %>%
    select(c(time_header, TEMP_0.5m, SPCOND_0.5m, PH_0.5m, ODO_CONC_0.5m,
             CHL_RFU_0.5m, TURB_0.5m, deltaDO, deltaTemp, PC_RFU_0.5m)) %>%
    rename(Timestamp = time_header,
           Temp = TEMP_0.5m,
           SpCond = SPCOND_0.5m,
           pH = PH_0.5m,
           ODO = ODO_CONC_0.5m,
           Chl = CHL_RFU_0.5m,
           Turb = TURB_0.5m,
           DeltaDO = deltaDO,
           DeltaTemp = deltaTemp,
           PC = PC_RFU_0.5m) %>%
    mutate(timestamp = as.POSIXct(strptime(Timestamp, format = "%m/%d/%Y %H:%M", tz = "Etc/GMT-4")),
           date = as.Date(timestamp, tz = "Etc/GMT-4"),
           PC = replace(PC, which(PC<0), 0),
           Chl = replace(Chl, which(Chl<0), 0)) %>%
    group_by(date) %>%
    summarize(TempMean = mean(Temp, na.rm = F),
              SpCondMean = mean(SpCond, na.rm = F),
              pHMean = mean(pH, na.rm = F),
              ODOMean = mean(ODO, na.rm = F),
              ChlMean = mean(Chl, na.rm = F),
```

```

    TurbMean = mean(Turb, na.rm = F),
    DeltaDOMean = mean(DeltaDO, na.rm = F),
    DeltaTempMean = mean(DeltaTemp, na.rm = F),
    PCMean1 = mean(PC, na.rm = F))
df_out$PCMean2 <- lead(df_out$PCMean1)
df_out$PCMean3 <- lead(df_out$PCMean1, 2)
df_out$PCMean4 <- lead(df_out$PCMean1, 3)
df_out$PCMean5 <- lead(df_out$PCMean1, 4)
df_out$PCMean6 <- lead(df_out$PCMean1, 5)
df_out$PCMean7 <- lead(df_out$PCMean1, 6)
return(df_out)
}

# Dailify our data
MB17_daily <- dailify(MB17, "TIMESTAMP")
write.csv(MB17_daily, "MB17_daily.csv")

MB18_daily <- dailify(MB18, "TIMESTAMP")
write.csv(MB18_daily, "MB18_daily.csv")

SA17_daily <- dailify(SA17, "timestamp")
write.csv(SA17_daily, "SA17_daily.csv")

SA18_daily <- dailify(SA18, "timestamp")
write.csv(SA18_daily, "SA18_daily.csv")

```

Plotting the data

Next, I made a function that creates (messy) time plots of the data so we can just quickly visually inspect how PC varies with different environmental factors over time.

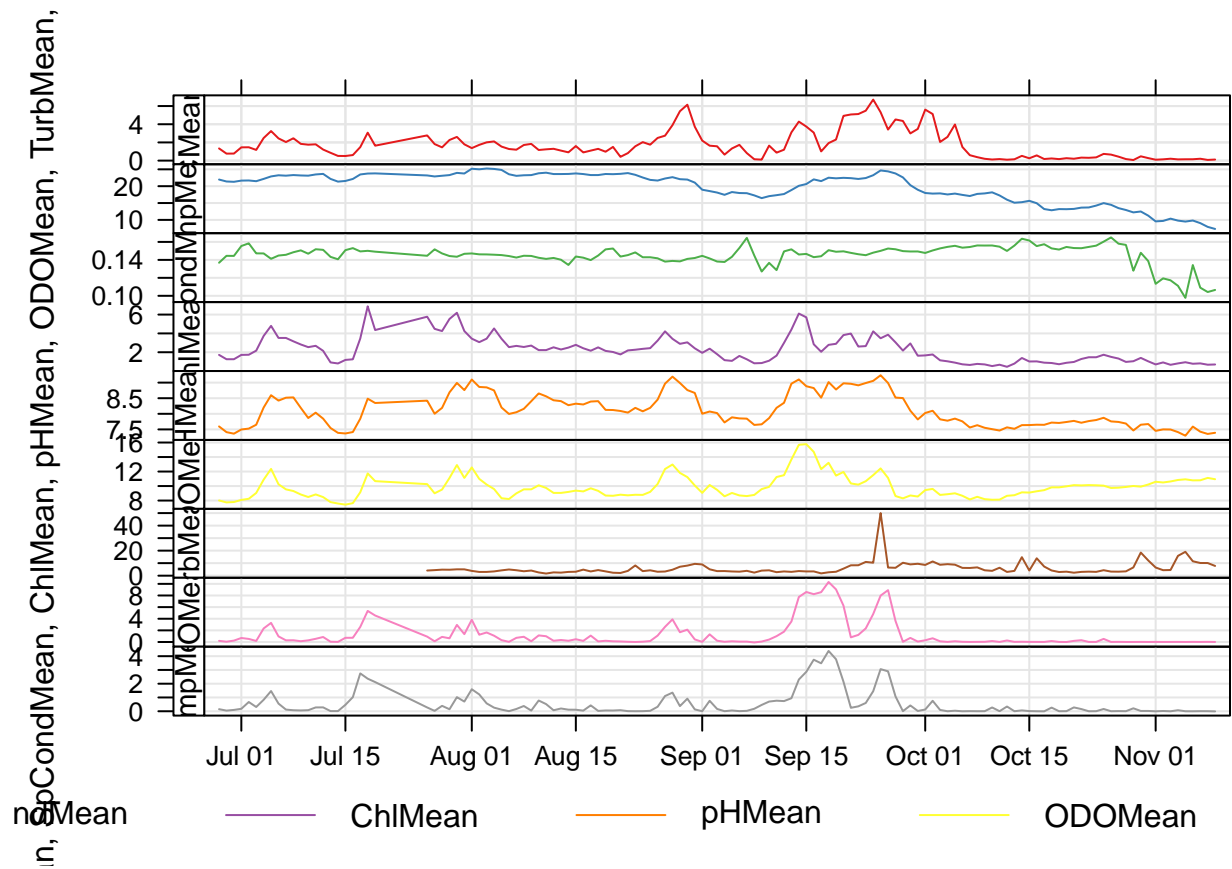
```

#Function that creates a timeplot of the df (with unlagged PC)
plotify <- function(df){
  timePlot(df, pollutant = c("PCMean1",
                             "TempMean",
                             "SpCondMean",
                             "ChlMean",
                             "pHMean",
                             "ODOMean",
                             "TurbMean",
                             "DeltaDOMean",
                             "DeltaTempMean"),

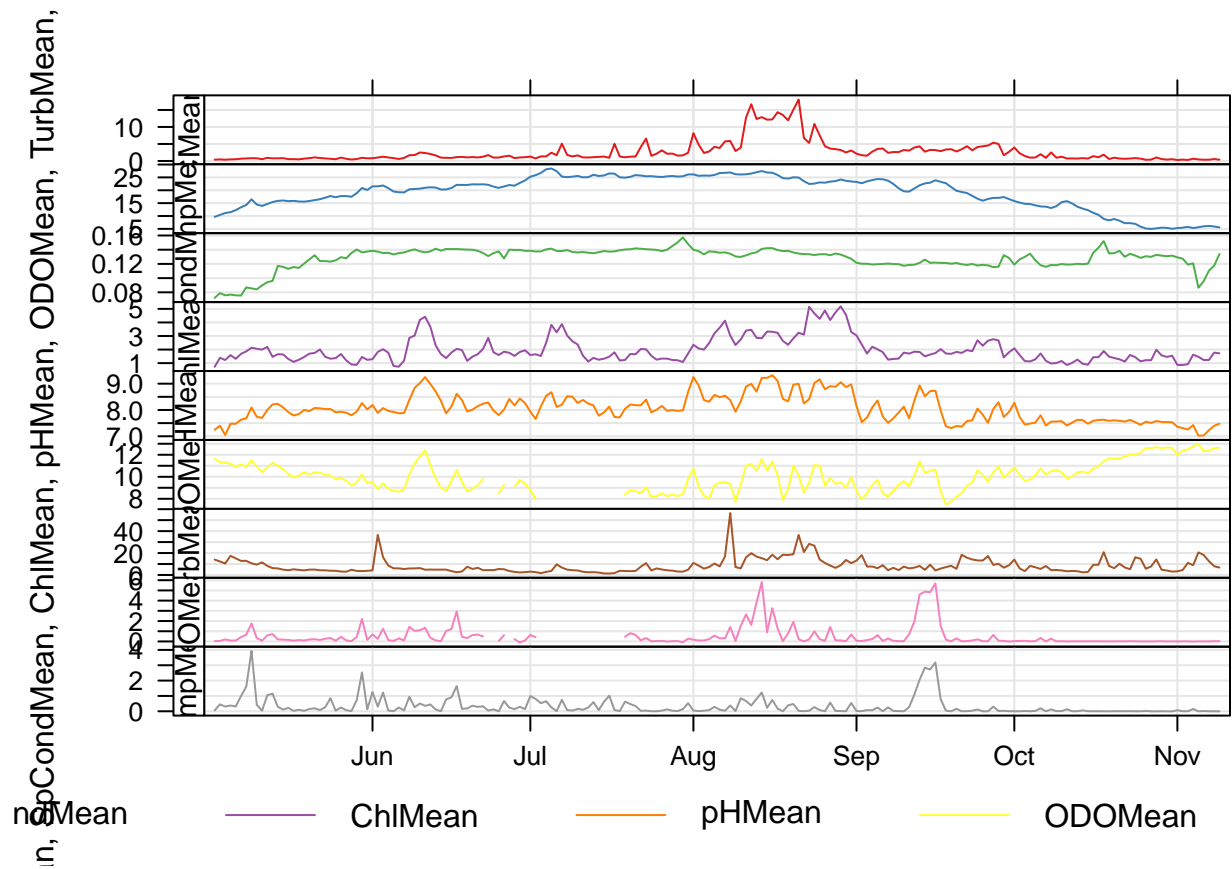
           y.relation = "free")
}

# Plot the data
plotify(MB17_daily)

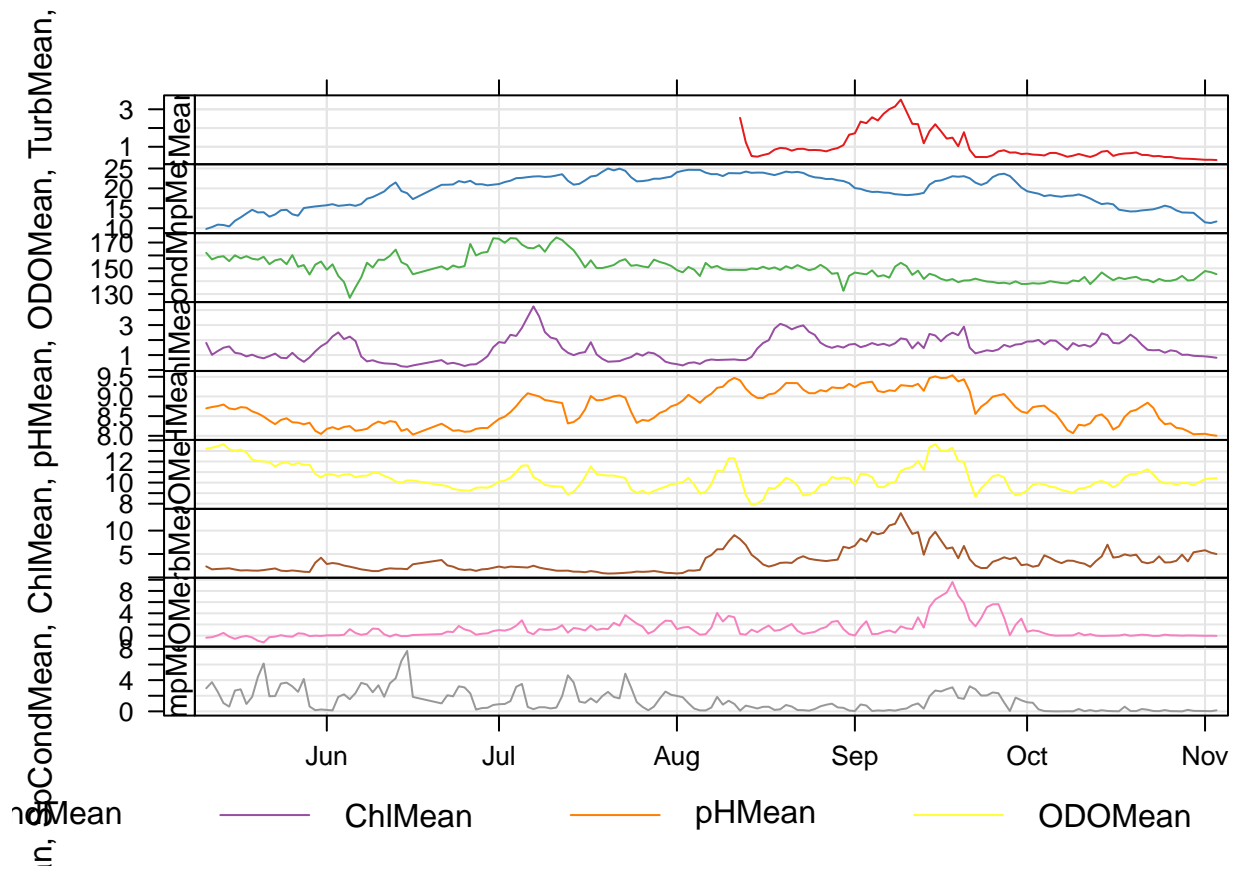
```



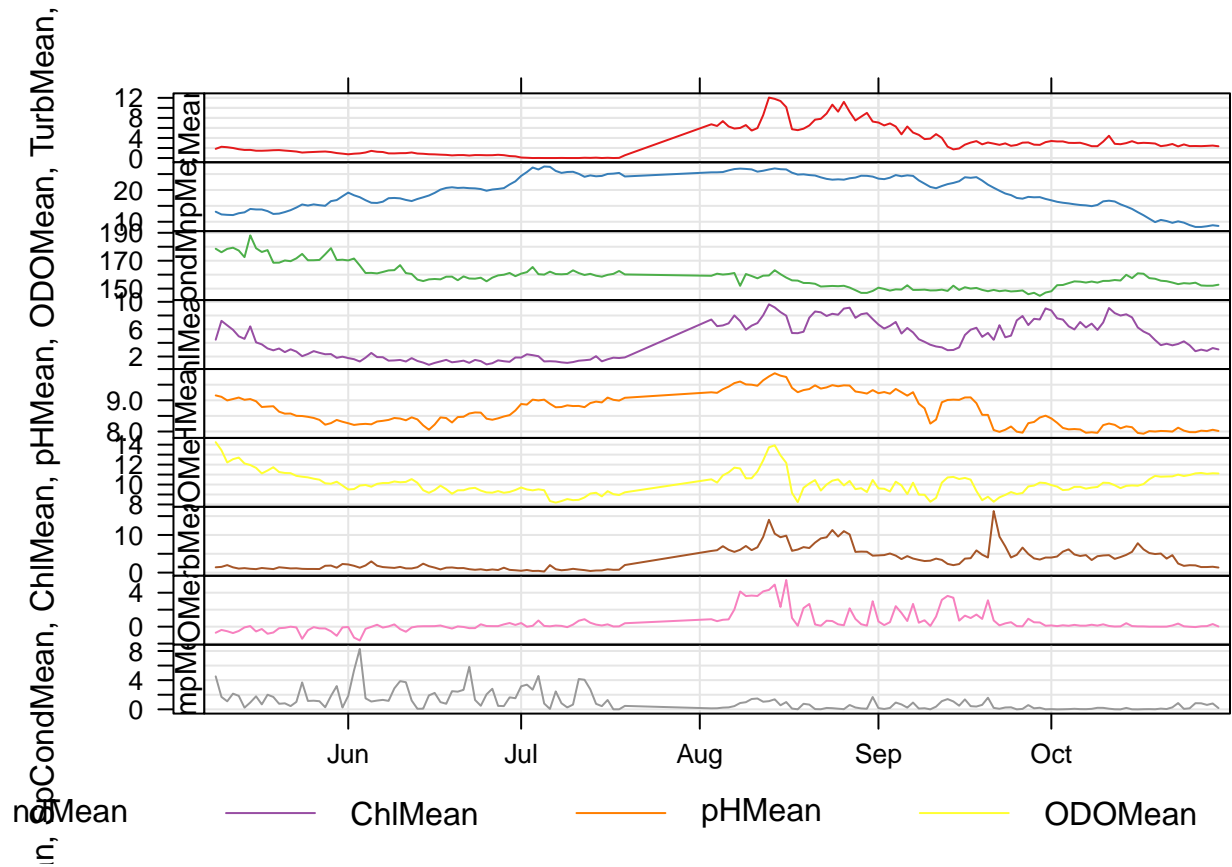
```
plotify(MB18_daily)
```



```
plotify(SA17_daily)
```

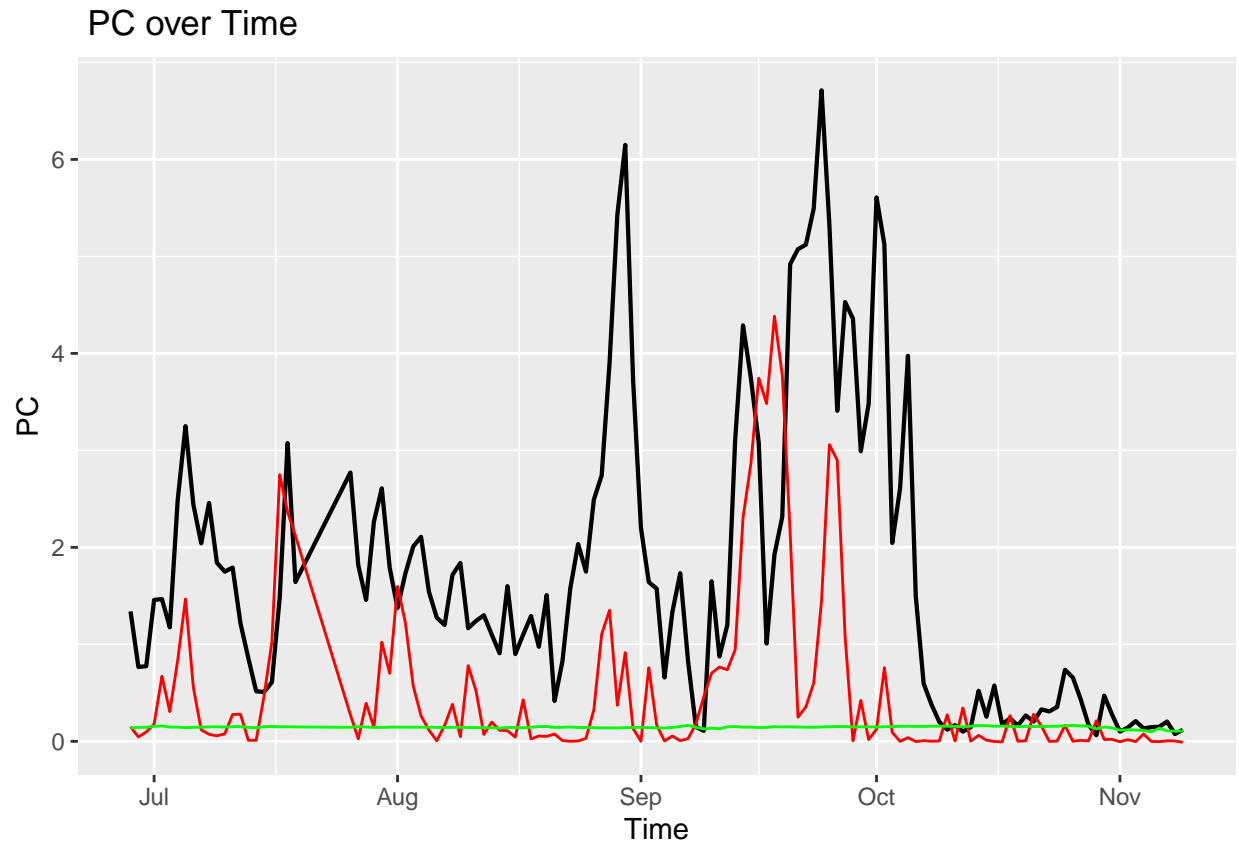


```
plotify(SA18_daily)
```



Prettier Time Plots with ggPlot2?

```
ggplot(MB17_daily, aes( x = date, y = PCMean1)) + geom_line(size = 0.75) + labs(title = " PC over Time")
```

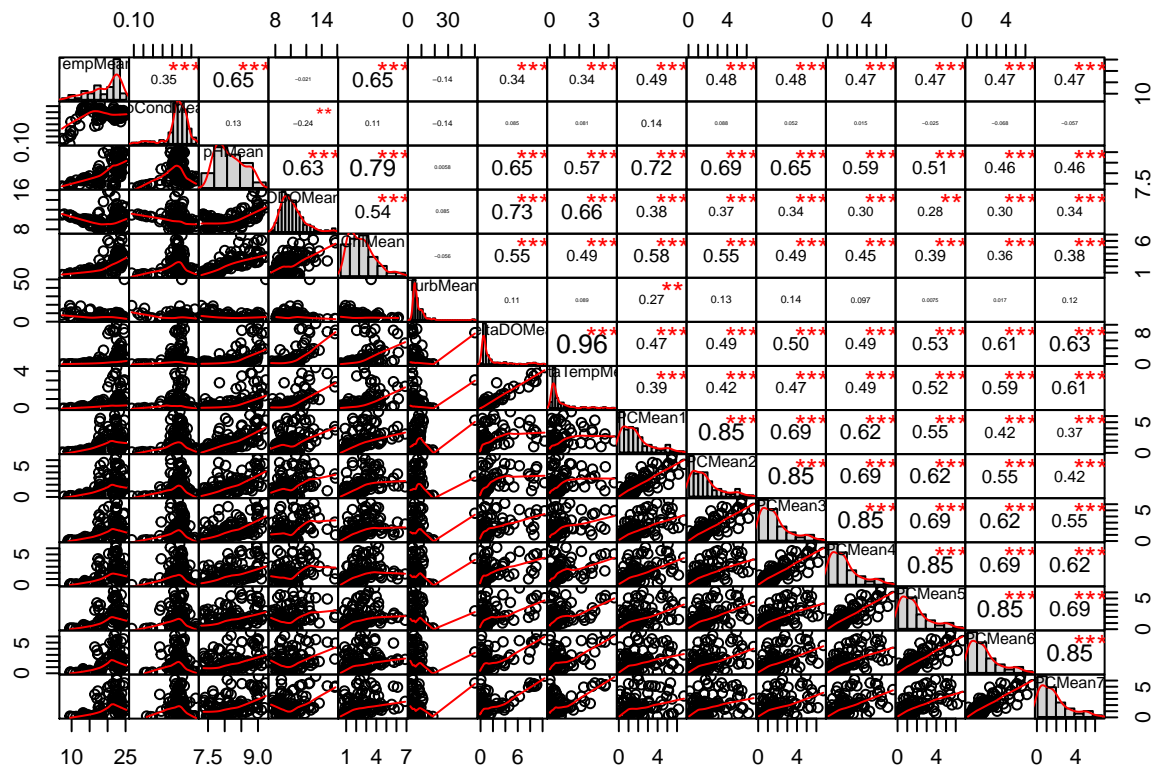


Y-axes are too different. Better to have separate plots.

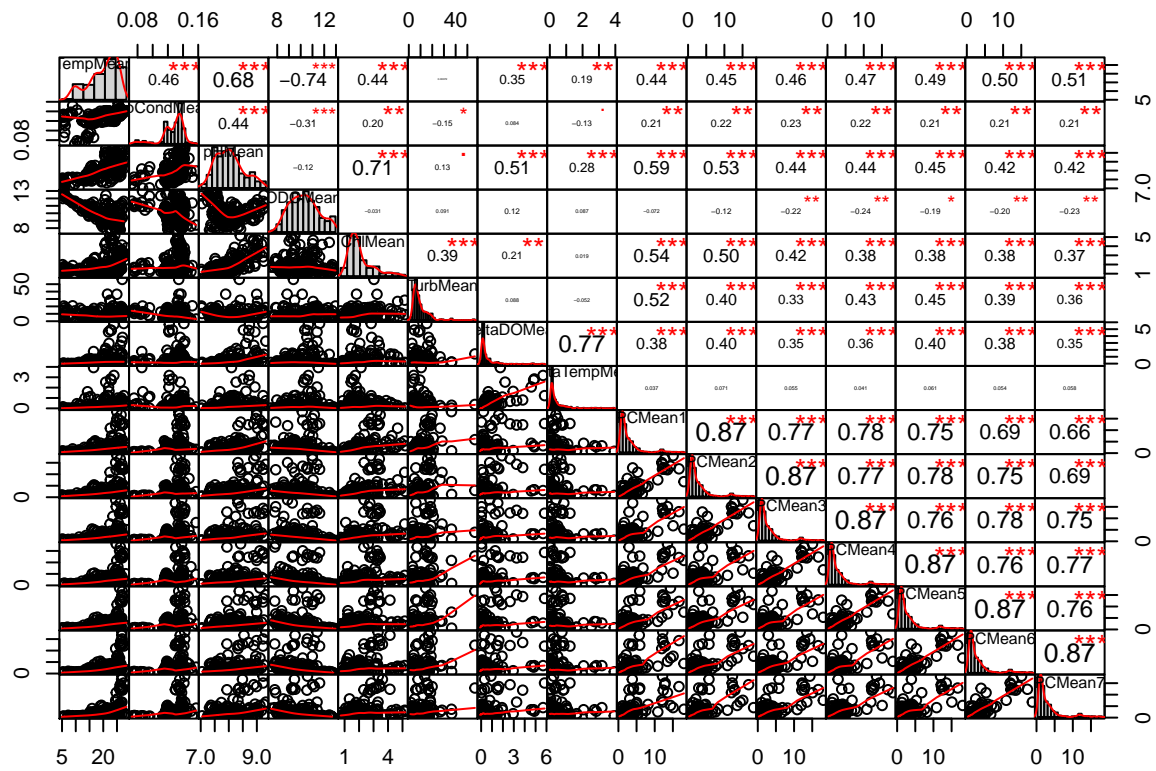
Investigate Correlations

Finally, we were interested in investigating correlations between various environmental factors and different time lags of PC. These tables are a little too crowded, so I will focus on looking at correlations between PC and individual environmental variables that were most important in the Feature Importance results.

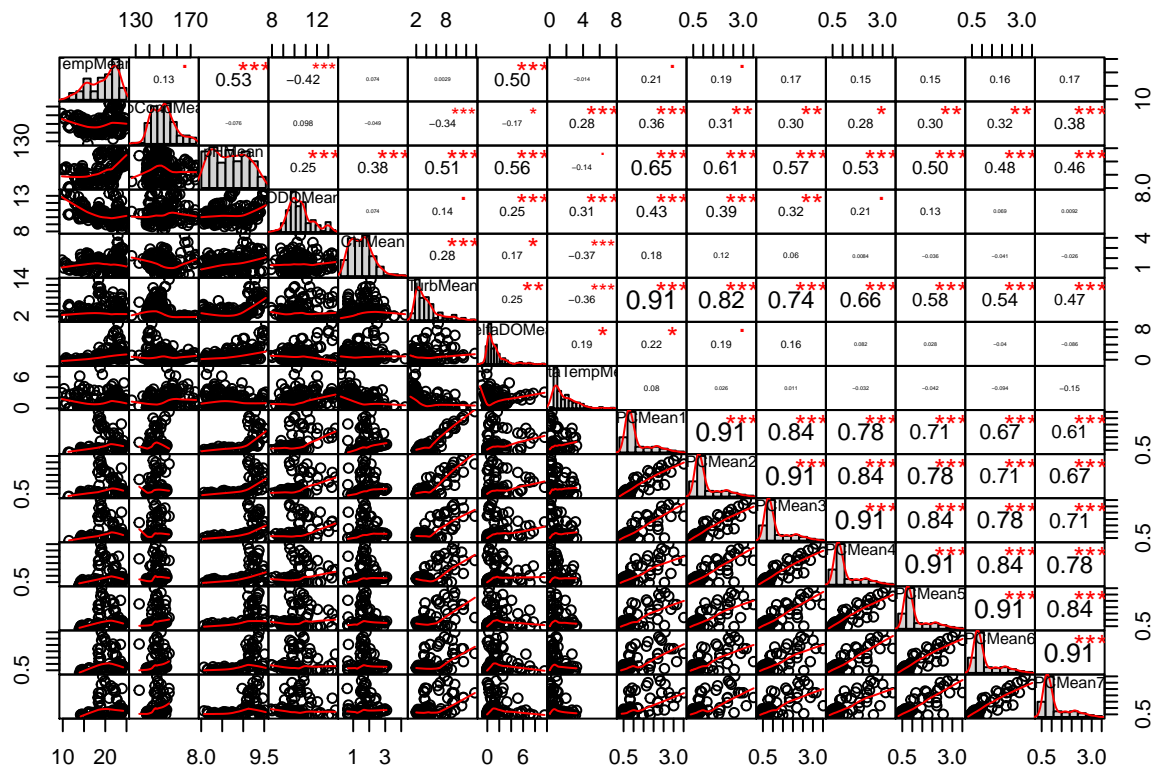
```
# Look at all correlations  
chart.Correlation(MB17_daily[,2:16], histogram = T)
```



```
chart.Correlation(MB18_daily[,2:16], histogram = T)
```

```
chart.Correlation(SA17_daily[,2:16], histogram = T)
```



```
chart.Correlation(SA18_daily[,2:16], histogram = T)
```

