HW4

2024-09-25

Part a.

```
DatYr <- function(year) {
  paste0("https://www.ndbc.noaa.gov/view_text_file.php?filename=44013h", year, ".txt.gz&dir=data/histor
}

###1: Inputting the 1985 data##
year1 <- "1985"
buoy <- read.table(DatYr(year1), na.strings = "MM", header = TRUE, fill = TRUE, sep = "")

header <- scan(DatYr(year1), what = 'character', nlines = 1)
colnames(buoy) <- header
buoy <- buoy %>%
  add_column(mm = NA, .after = "hh") %>%
  add_column(TIDE = NA, .after = "VIS")

##2: Inputting the years from 1986-2006##
YearAll <- 1986:2006
for (i in YearAll) {
  url <- DatYr(i)
  temp_data <- read.table(DatYr(i), header = TRUE, fill = TRUE, sep = "")
  temp_data <- temp_data %>%
    add_column(mm = NA, .after = "hh") %>%
    add_column(TIDE = NA, .after = "VIS")
  buoy <- bind_rows(buoy, temp_data)
}
buoy$YY<-na.omit(c(buoy$YY, buoy$YYYY))
buoy <- buoy %>%
  select(-YYYY, -TIDE.1, -mm.1)
colnames(buoy)[colnames(buoy) == "YY"] <- "YYYY"

##3: Inputting the years from 2007-2023##
YearAll<-2007:2023
for (i in YearAll) {
  url <- DatYr(i)
  temp_data <- (read.table(DatYr(i), header = FALSE, fill = TRUE, sep = "", skip = 1))
  header=scan(DatYr(i), what = 'character', nlines = 1)
  colnames(temp_data) <- header
  buoy <- bind_rows(buoy, temp_data)
}

##4: Remove all of the observations that contain missing data##
buoy$YYYY <- na.omit(c(buoy$YYYY, buoy$`#YY`))
```

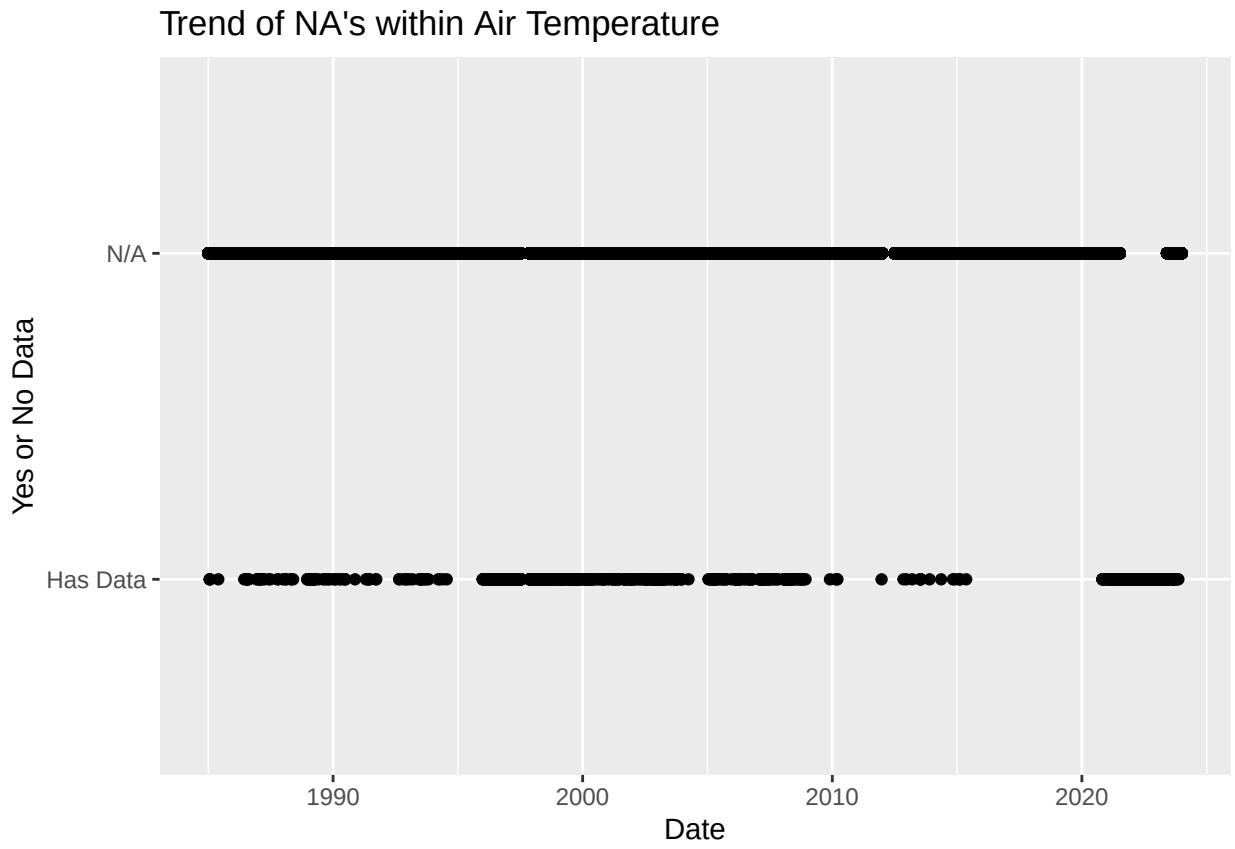
```r
buoy$BAR <- na.omit(c(buoy$BAR, buoy$PRES))
buoy <- buoy %>%
  select(-`#YY`, -PRES)

buoy$WD <- na.omit(c(buoy$WD, buoy$WDIR))
buoy <- buoy %>%
  select(-WDIR)
```

Part B.

```r
##1: Remove TIDE column because it does not serve a purpose ##
buoy = buoy %>% select(-TIDE)

##2: Replace all of the numbers that are unavailable with NA ##
buoy$VIS <- ifelse(buoy$VIS == 99, NA, buoy$VIS)
buoy$DEWP <- ifelse(buoy$DEWP == 999, NA, buoy$DEWP)
buoy$MWD <- ifelse(buoy$MWD == 999, NA, buoy$MWD)
buoy$APD <- ifelse(buoy$APD == 99, NA, buoy$APD)
buoy$DPD <- ifelse(buoy$DPD == 99, NA, buoy$DPD)
buoy$WVHT <- ifelse(buoy$WVHT == 99, NA, buoy$WVHT)
buoy$BAR <- ifelse(buoy$BAR == 999, NA, buoy$BAR)
buoy$ATMP <- ifelse(buoy$ATMP == 999, NA, buoy$ATMP)
buoy$WTMP <- ifelse(buoy$WTMP == 999, NA, buoy$WTMP)
buoy$WSPD <- ifelse(buoy$WSPD == 99, NA, buoy$WSPD)
buoy$BAR <- ifelse(buoy$BAR == 9999, NA, buoy$BAR)

##3: Add a datetime column##
buoy$datetime <- ymd_h(paste(buoy$YYYY, buoy$MM, buoy$DD, buoy$hh, sep = "-"))

cat("It can be beneficial to convert missing or null data points to `NA` in R because many functions ar
```

```
## It can be beneficial to convert missing or null data points to ‘NA‘ in R because many functions are
```

```r
VIScategory <- ifelse(is.na(buoy$ATMP), "Has Data", "N/A")

##4: Find the pattern (or patterns) of the NA's ##
ggplot(buoy, aes(x = datetime, y = VIScategory)) +
  geom_point() +
  labs(x = "Date", y = "Yes or No Data", title = "Trend of NA's within Air Temperature")
```



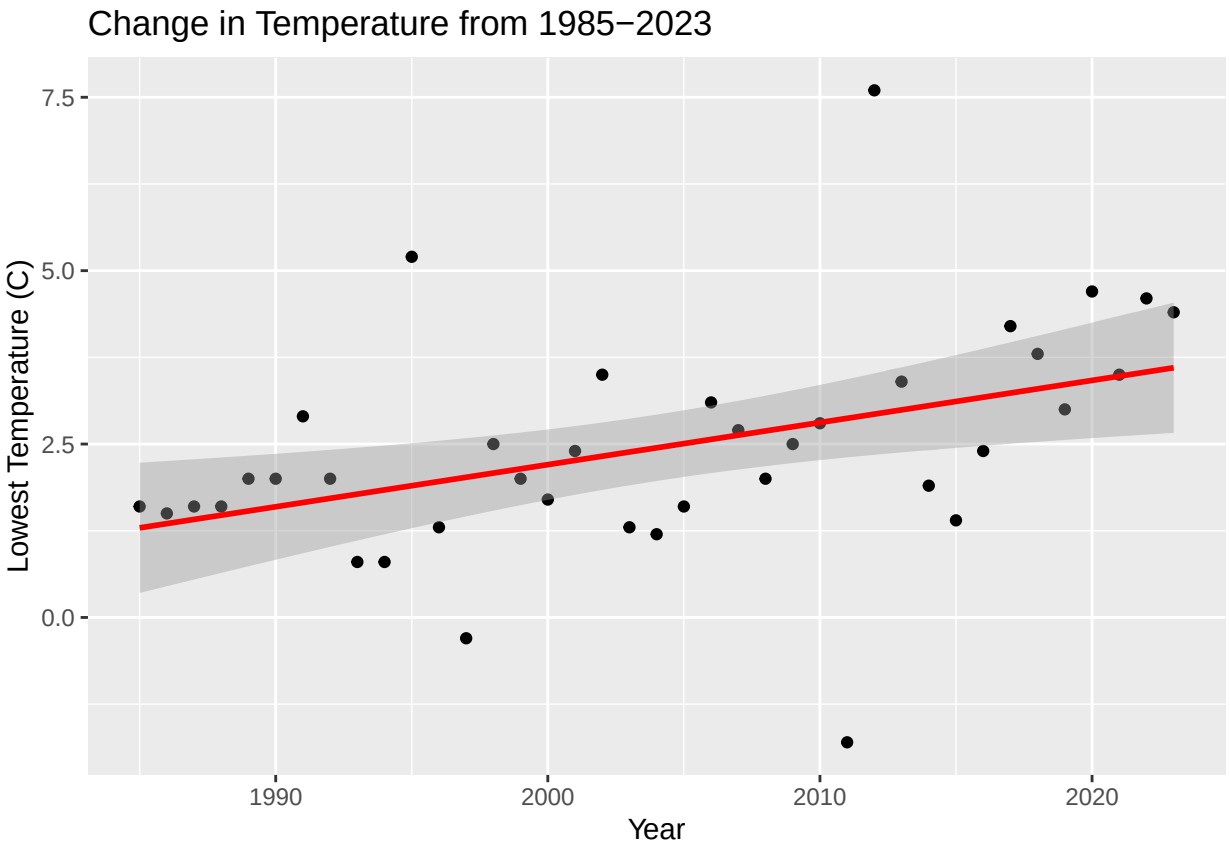

```
cat("\n\nThere appears to be a noticeable trend of increased available ATMP data in the Buoy dataset. Ba
```

```
##
##
## There appears to be a noticeable trend of increased available ATMP data in the Buoy dataset. Based or
```

Part C.

```
## In this instance, the rise in temperature observed in the Buoy data was identified as a useful indica
YrMin <- buoy %>%
  group_by(year = year(datetime)) %>%
  summarise(min = min(WTMP, na.rm = TRUE))

## Set the axis and plot ##
ggplot(YrMin, aes(x = year, y = min)) +
  geom_point() +
  geom_smooth(method = "lm", col = "red")+
  labs(title = "Change in Temperature from 1985-2023",
       x = "Year",
       y = "Lowest Temperature (C)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

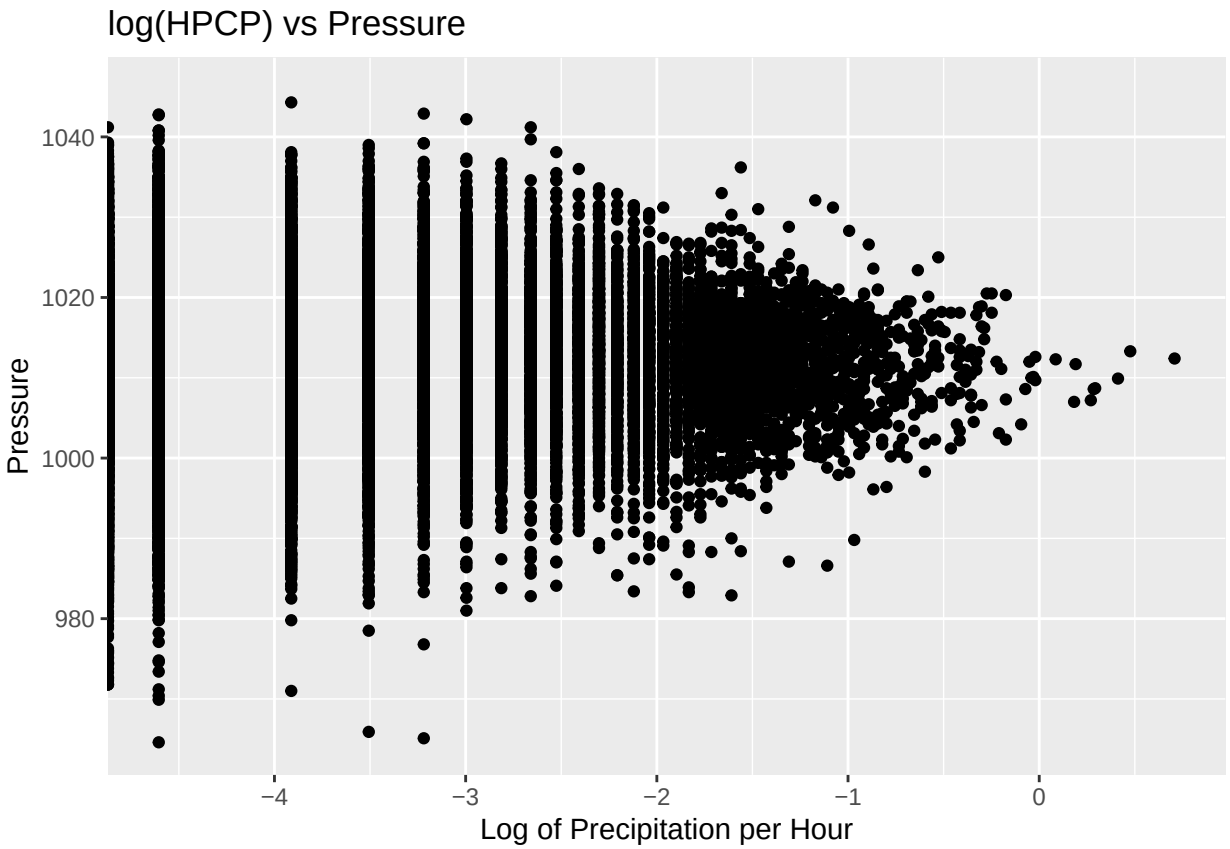



```
cat("A useful statistic to enhance my plot would be to calculate the average temperature over the entir
```

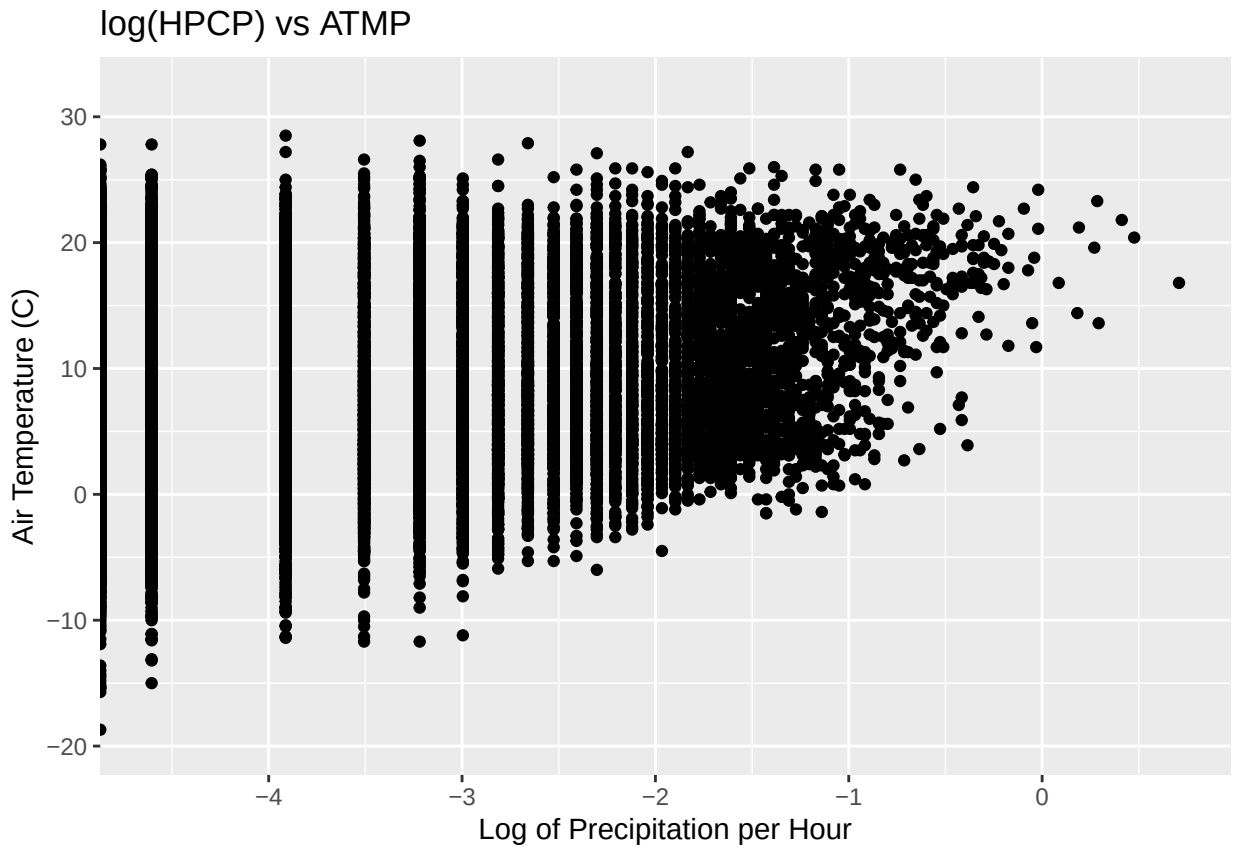
```
## A useful statistic to enhance my plot would be to calculate the average temperature over the entire y
```

Part D.

```
# The goal is to merge the two datasets into a single `NewDat` dataset by aligning them based on simila
rainfall <- read.csv("Rainfall.csv")
rainfall$datetime <- ymd_hm(rainfall$DATE)
rainfall_subset <- rainfall[, c("datetime", "HPCP")]
NewDat <- merge(buoy, rainfall_subset, by = "datetime", all = TRUE)


ggplot(NewDat, aes(x = log(HPCP), y = BAR), pch = 20) +
  geom_point() +
  labs(x = "Log of Precipitation per Hour",
       y = "Pressure",
       title = "log(HPCP) vs Pressure")
```




```
ggplot(NewDat, aes(x = log(HPCP), y = ATMP), pch = 20) +
  geom_point() +
  labs(x = "Log of Precipitation per Hour",
       y = "Air Temperature (C)",
       title = "log(HPCP) vs ATMP")
```




```
## I have selected these two graphs because they provide the best visualization of trends in HPCP along
```

```
cat("I find it difficult to sympathize with TV meteorologists, given their access to vast amounts of wea
```

```
## I find it difficult to sympathize with TV meteorologists, given their access to vast amounts of weath
```