Buoy

```
library(data.table)
library(dplyr)
library(lubridate)
library(ggplot2)
library(zoo)
library(tibble)
library(readr)
```

Getting Buoy Data

```
file_root <- "https://www.ndbc.noaa.gov/view_text_file.php?filename=44013h"
tail <- ".txt.gz&dir=data/historical/stdmet/"

load_buoy_data1 <- function(year) {
  path <- paste0(file_root, year, tail)


  if (year < 2007) {
  header <- scan(path, what = 'character', nlines = 1)
  buoy <- read.table(path, fill = TRUE, header = TRUE, sep = "")
  buoy <- add_column(buoy, mm = NA, .after = "hh")
  buoy <- add_column(buoy, TIDE = NA, .after = "VIS")

  } else {
    header <- scan(path, what = 'character', nlines = 1)
    buoy <- fread(path, header = FALSE, skip = 1, fill = TRUE)

    setnames(buoy, header)
  }

  return(buoy)
}

all_data1 <- lapply(1985:2024, load_buoy_data1)

combined_data1 <- rbindlist(all_data1, fill = TRUE)
```

```
load_buoy_data <- function(year) {
  path <- paste0(file_root, year, tail)

  header <- scan(path, what = 'character', nlines = 1)
  num_columns <- length(header)

  if (num_columns == 16) {
    buoy <- read.table(path, fill = TRUE, header = TRUE, sep = "")
    buoy <- add_column(buoy, mm = NA, .after = "hh")
```

```
    buoy <- add_column(buoy, TIDE = NA, .after = "VIS")

  } else if (num_columns == 17) {
    buoy <- read.table(path, fill = TRUE, header = TRUE, sep = "")
    buoy <- add_column(buoy, TIDE = NA, .after = "VIS")

  } else {
    buoy <- fread(path, header = FALSE, skip = 1, fill = TRUE)
    setnames(buoy, header)
  }

  return(buoy)
}
all_data <- lapply(1985:2024, load_buoy_data)
combined_data <- rbindlist(all_data, fill = TRUE)
```

```
combined_data1 <- combined_data1 %>%
  mutate(across(any_of(c("YY", "#YY", "YYYY")), as.character)) %>%
  mutate(
    YY = as.character(YY),
    `#YY` = as.character(`#YY`),
    YYYY = as.character(YYYY)
  )

# Combine year columns safely using coalesce
combined_data1 <- combined_data1 %>%
  mutate(YYYY = coalesce(YYYY, `#YY`, YY))
combined_data1 <- combined_data1 %>%
  mutate(BAR = coalesce(as.numeric(BAR), as.numeric(PRES)),  # Convert BAR and PRES to nu
    WD = coalesce(as.numeric(WD), as.numeric(WDIR)))
```

```
Warning: There were 2 warnings in `mutate()`.
The first warning was:
ℹ In argument: `BAR = coalesce(as.numeric(BAR), as.numeric(PRES))`.
Caused by warning in `list2()`:
! NAs introduced by coercion
ℹ Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

```
combined_data1 <- combined_data1 %>%
  select(-TIDE, -TIDE.1, -mm,- WDIR, -PRES,-`#YY`,-YY)

combined_data1$datetime <- ymd_h(paste(combined_data1$Year, combined_data1$MM, combined_d
```

```
Warning: All formats failed to parse. No formats found.
```

```
combined_data1 <- combined_data1 %>%
  mutate(across(everything(),
                ~ na_if(as.numeric(as.character(.)), 99)) %>%
```

```
                na_if(999) %>%
                na_if(9999)))
```

```
Warning: There were 14 warnings in `mutate()`.
The first warning was:
ℹ In argument: `across(...)`.
Caused by warning in `vec_cast()`:
! NAs introduced by coercion
ℹ Run `dplyr::last_dplyr_warnings()` to see the 13 remaining warnings.
```

```
#summary(combined_data)
str(combined_data1)
```

```
Classes 'data.table' and 'data.frame':  518656 obs. of  18 variables:
 $ MM      : num  1 1 1 1 1 1 1 1 1 1 ...
 $ DD      : num  1 1 1 1 1 1 1 1 1 1 ...
 $ hh      : num  0 1 2 3 4 5 6 7 8 9 ...
 $ WD      : num  60 80 100 100 110 90 60 30 40 40 ...
 $ WSPD    : num  4 4 4 4 4 4 4 6 7 ...
 $ GST     : num  5 5 5 5 5 6 5 6 8 ...
 $ WVHT    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ DPD     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ APD     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ MWD     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ BAR     : num  1030 1030 1030 1029 1029 ...
 $ ATMP    : num  4.7 5.1 5.6 5.8 5.8 5.3 5.5 5.8 5.9 6.2 ...
 $ WTMP    : num  6.7 6.7 6.6 6.7 6.7 6.7 6.7 6.7 6.7 6.7 ...
 $ DEWP    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ VIS     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ YYYY    : num  85 85 85 85 85 85 85 85 85 85 ...
 $ mm.1    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ datetime: num  NA NA NA NA NA NA NA NA NA NA ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
#str(combined_data$datetime)
if (!inherits(combined_data1$datetime, "POSIXct")) {
  combined_data1$datetime <- ymd_h(paste(combined_data1$YYYY, combined_data1$MM, combined
}
```

```
Warning: 18 failed to parse.
```

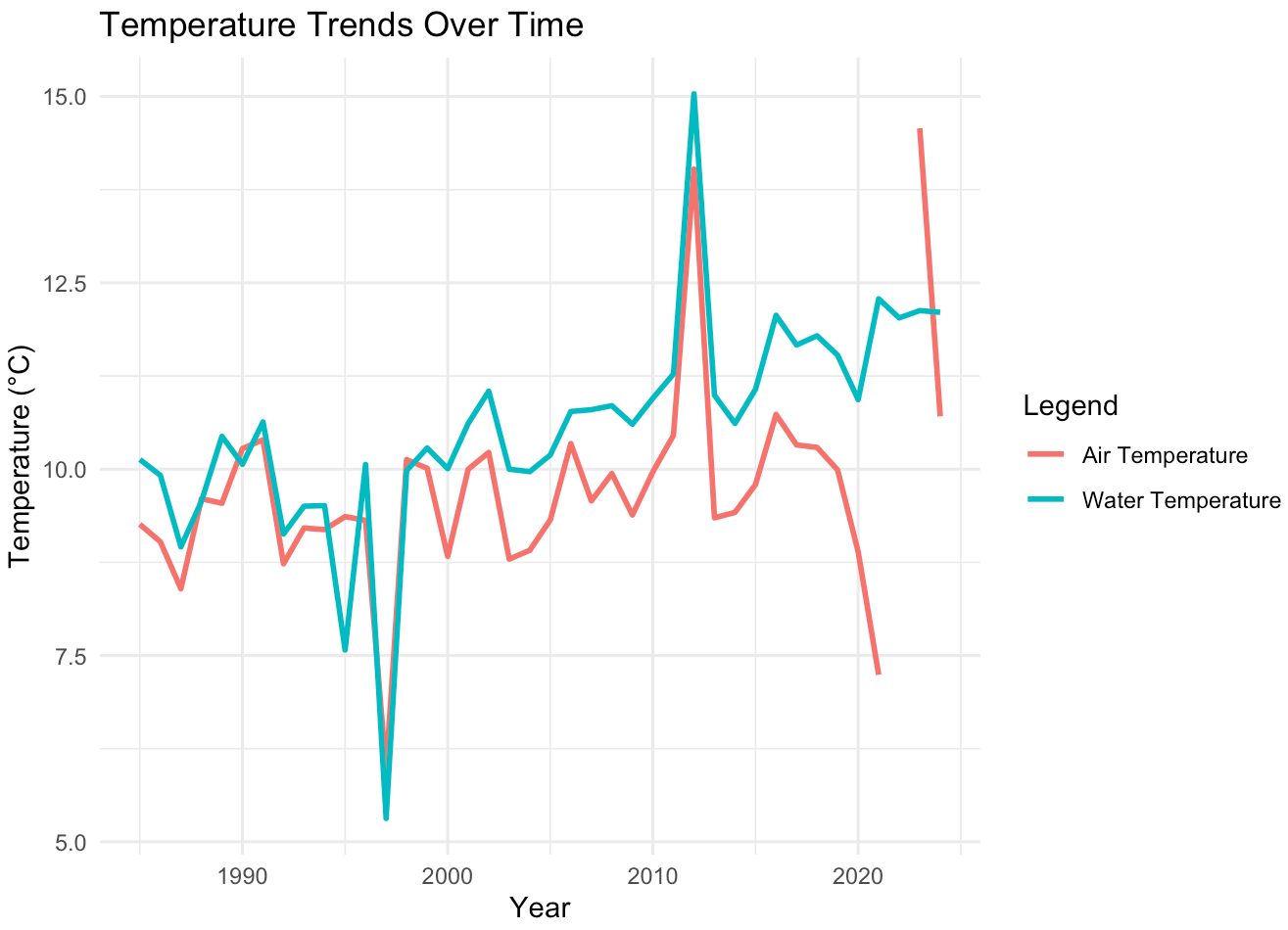
```
combined_data1 <- combined_data1 %>%
  mutate(Year = year(datetime)) %>%
  select(-YYYY)

yearly_avg_temp <- combined_data1 %>%
  group_by(Year) %>%
  summarise(
    avg_air_temp = mean(ATMP, na.rm = TRUE),
```

```
    avg_water_temp = mean(WTMP, na.rm = TRUE)
  )
ggplot(yearly_avg_temp, aes(x = Year)) +
  geom_line(aes(y = avg_air_temp, color = "Air Temperature"), size = 1) +
  geom_line(aes(y = avg_water_temp, color = "Water Temperature"), size = 1) +
  labs(title = "Temperature Trends Over Time",
       x = "Year",
       y = "Temperature (°C)",
       color = "Legend") +
  theme_minimal()
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.

Warning: Removed 1 row containing missing values or values outside the scale range
(`geom_line()`).
Removed 1 row containing missing values or values outside the scale range
(`geom_line()`).
```




```
combi <- combined_data1 %>%
  mutate(Year = lubridate::year(datetime))

na_summary <- combi %>%
```

```
  group_by(Year) %>%
  summarise(
    air_temp_NA_count = sum(is.na(ATMP)),
    water_temp_NA_count = sum(is.na(WTMP)),
    total_rows = n()
  ) %>%
  arrange(desc(air_temp_NA_count), desc(water_temp_NA_count))
na_summary
```

```
# A tibble: 41 × 4
    Year air_temp_NA_count water_temp_NA_count total_rows
   <dbl>             <int>                <int>      <int>
 1  2022             52529                  313      52529
 2  2021             27379                  275      51563
 3  2023             21668                  254      52480
 4  1996               263                  263       8952
 5  1997               187                 1295       5712
 6  1998               173                  172       8784
 7  1999                89                   91       8348
 8  1989                76                   26       7933
 9  2001                60                   61       8760
10  2002                39                   41       8760
# ℹ 31 more rows
```