HW4

2024-09-27

```
library(data.table)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

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

years <- 1985:2023

ndbc_list <- list()

for (year in years) {
  path <- paste0(file_root, year, tail)
```

```
  header <- scan(path, what = 'character', nlines = 1)
  skip <- ifelse(year >= 2007, 2, 1)
  ndbc <- fread(path, header = FALSE, skip = skip)
  num_cols <- ncol(ndbc)
  if (length(header) > num_cols) {
    header <- header[1:num_cols]
  } else if (length(header) < num_cols) {
    header <- c(header, paste0("V", (length(header) + 1):num_cols))
  }

  colnames(ndbc) <- header
    if ("YY" %in% colnames(ndbc)) {
    ndbc$YY <- ifelse(ndbc$YY < 100, ifelse(ndbc$YY > 20, 1900 + ndbc$YY, 2000 + ndbc$YY), ndbc$YY)
  }

  if ("YY" %in% colnames(ndbc) & "MM" %in% colnames(ndbc) & "DD" %in% colnames(ndbc) & "hh" %in% colnam
    ndbc$Date <- ymd_hms(paste(ndbc$YY, ndbc$MM, ndbc$DD, ndbc$hh, ndbc$mm))
  }


  ndbc_list[[as.character(year)]] <- ndbc
}
```

```
## Warning in fread(path, header = FALSE, skip = skip): Stopped early on line
## 5114. Expected 16 fields but found 17. Consider fill=TRUE and comment.char=.
## First discarded non-empty line: <<2000 08 01 00 78 4.3 5.1 0.58 8.33 5.36 999
## 1022.9 17.3 17.5 15.0 99.0 99.00>>
```

```
ndbc_list <- rbindlist(ndbc_list, fill = TRUE)
ndbc_list <- ndbc_list %>%
  mutate(Year = coalesce(as.numeric(YY), as.numeric(YYYY), as.numeric(`#YY`))) %>%
  select(-YY, -YYYY, -`#YY`) %>%
  select(Year, everything())
ndbc_list <- ndbc_list %>%
  mutate(Wind_Direction = coalesce(WD, WDIR)) %>%
  select(-WD, -WDIR)
ndbc_list <- ndbc_list %>%
  mutate(Pressure = coalesce(BAR, PRES)) %>%
  select(-BAR, -PRES)
if (all(c("Year", "MM", "DD", "hh", "mm") %in% colnames(ndbc_list))) {
  ndbc_list[, date := ymd_hms(paste(Year, MM, DD, hh, mm))]
}
```

```
## Warning: 411397 failed to parse.
```

```
str(ndbc_list)
```

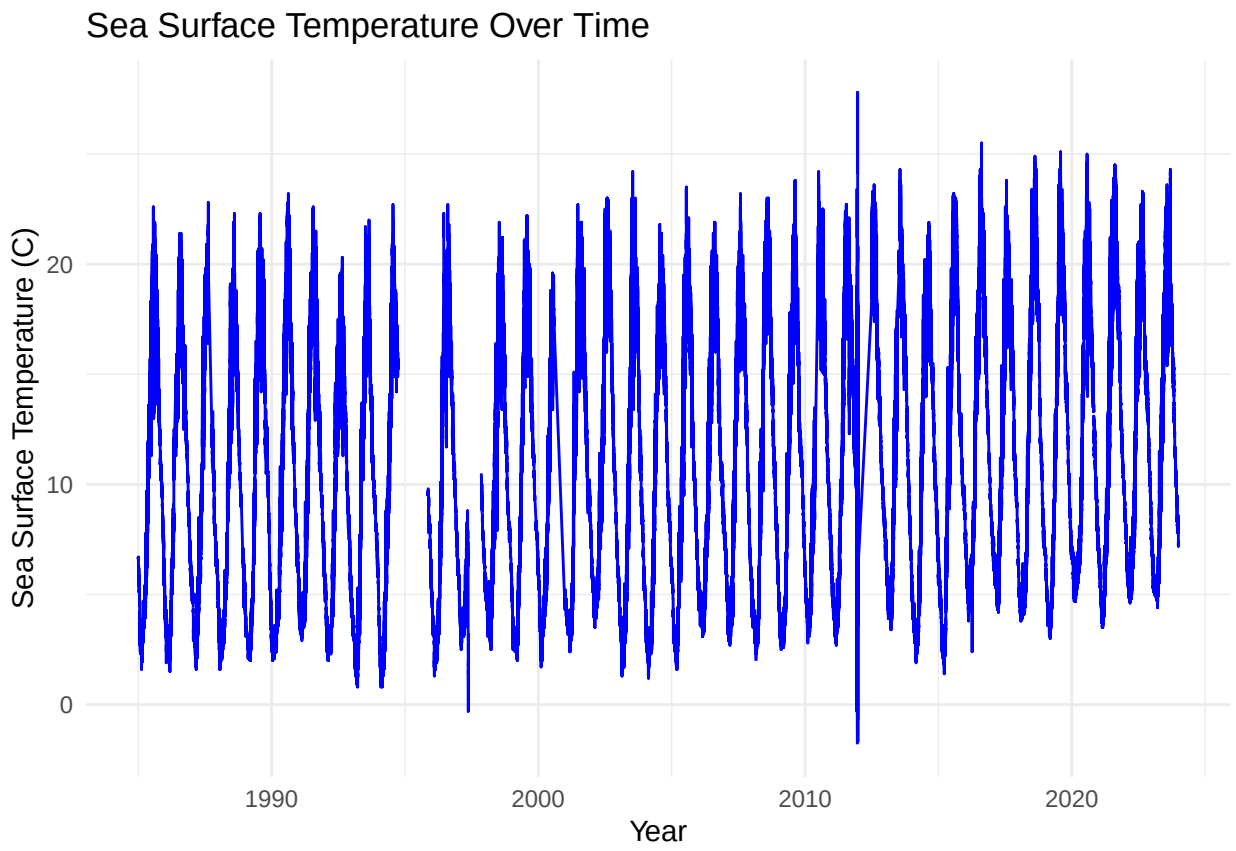
```
## Classes 'data.table' and 'data.frame':   462301 obs. of  19 variables:
##  $ Year          : num  1985 1985 1985 1985 1985 ...
##  $ MM            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ DD            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ hh            : int  0 1 2 3 4 5 6 7 8 9 ...
```

```
##  $ WSPD          : num  4 4 4 4 4 4 4 4 6 7 ...
##  $ GST           : num  5 5 5 5 5 5 6 5 6 8 ...
##  $ WVHT          : num  99 99 99 99 99 99 99 99 99 99 ...
##  $ DPD           : num  99 99 99 99 99 99 99 99 99 99 ...
##  $ APD           : num  99 99 99 99 99 99 99 99 99 99 ...
##  $ MWD           : int  999 999 999 999 999 999 999 999 999 999 ...
##  $ 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  999 999 999 999 999 999 999 999 999 999 ...
##  $ VIS           : num  99 99 99 99 99 99 99 99 99 99 ...
##  $ TIDE          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ mm            : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Wind_Direction: int  60 80 100 100 110 90 60 30 40 40 ...
##  $ Pressure      : num  1030 1030 1030 1029 1029 ...
##  $ date          : POSIXct, format: NA NA ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

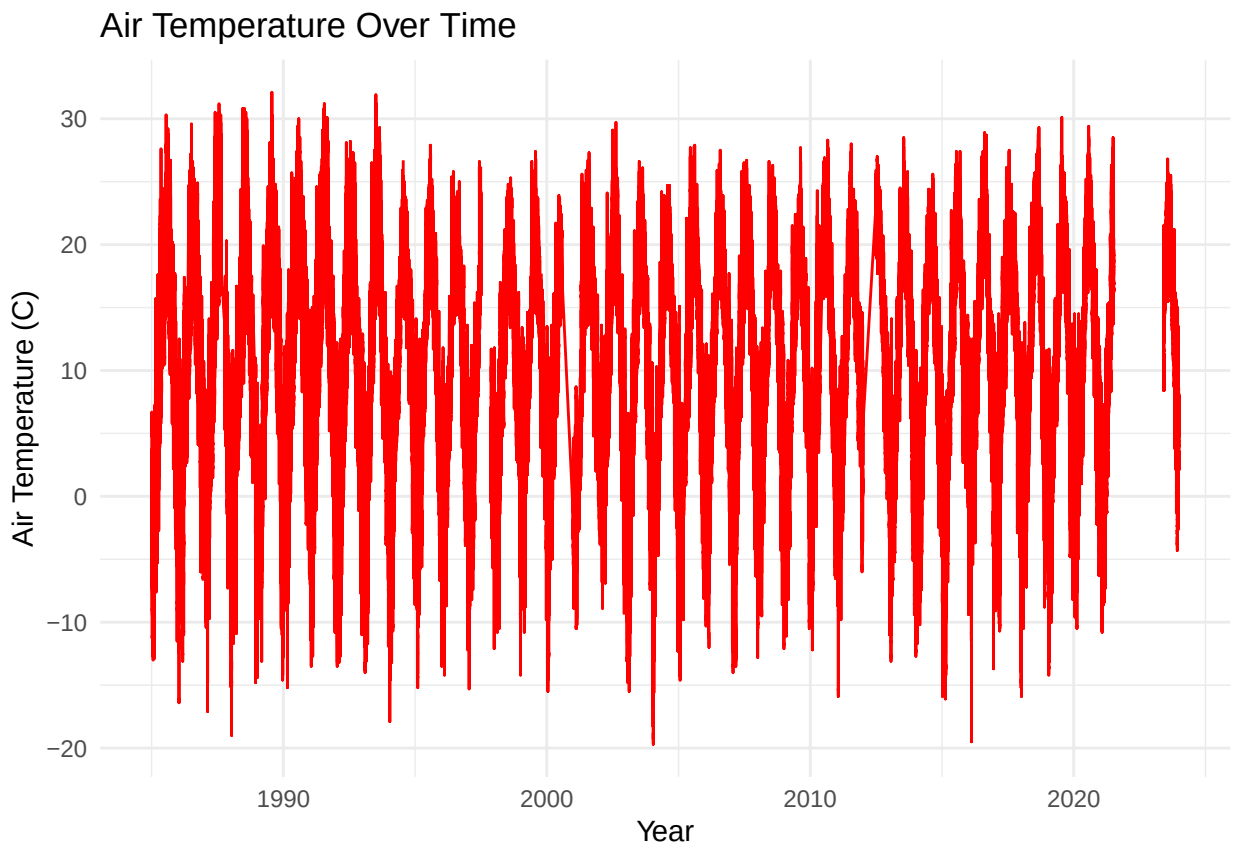
```
ndbc_list <- ndbc_list %>%
  mutate(date = ifelse(complete.cases(Year, MM, DD, hh, mm),
                       make_datetime(year = Year, month = MM, day = DD, hour = hh, min = mm),
                       as.POSIXct(NA)))
ndbc_list$date <- make_datetime(
  year = ifelse(is.na(ndbc_list$Year), 2000, ndbc_list$Year),
  month = ifelse(is.na(ndbc_list$MM), 1, ndbc_list$MM),
  day = ifelse(is.na(ndbc_list$DD), 1, ndbc_list$DD),
  hour = ifelse(is.na(ndbc_list$hh), 0, ndbc_list$hh),
  min = ifelse(is.na(ndbc_list$mm), 0, ndbc_list$mm)
)
```

```
#(b)
ndbc_list[ndbc_list == 999] <- NA
#I don't think it's appropriate because in this data, 999 may mean that it is beyond the forecast range
```

```
#(c)
ndbc_list$date <- as.Date(ndbc_list$date)

library(ggplot2)
library(dplyr)

ggplot(ndbc_list, aes(x = date, y = WTMP)) +
  geom_line(color = "blue") +
  labs(title = "Sea Surface Temperature Over Time", x = "Year", y = "Sea Surface Temperature (C)") +
  theme_minimal()
```




```
#Sea surface temperature over time: This visualization helps us see the long-term warming trend of the

ggplot(ndbc_list, aes(x = date, y = ATMP)) +
  geom_line(color = "red") +
  labs(title = "Air Temperature Over Time", x = "Year", y = "Air Temperature (C)") +
  theme_minimal()
```




```
#Rising air temperatures above bodies of water are an indicator of global warming. Looking at trends in
```