Buoy-Project

Ruxin Liu

9/24/2020

Introduction -your understanding of the question being addressed, your approach,

Methodology

Discussion

Conclusion

```
# Loading required packages
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stringr)
```

Importing the data

```
# The first part of the data URL
url_1 <- "http://www.ndbc.noaa.gov/view_text_file.php?filename=mlrf1h"
# The second part of the data URL
url_2 <- ".txt.gz&dir=data/historical/stdmet/"
# Read the latest available 20 years of data
years <- c(1999:2018)
# The complete 20 URLs
urls <- str_c(url_1, years, url_2, sep = "")
filenames <- str_c("mr", years, sep = "")
```

```
# Load in the 20-year data from the NDBC website
# Year 1999 - 2006
for(i in 1:8){
  suppressMessages(
```

```
    # Fill any missing values with NA:
    assign(filenames[i], read.table(urls[i], header = TRUE, fill = TRUE))
  )

}
# Year 2007 - 2018: Column names started with #, which will cause issue
for(i in 9:20){
  suppressMessages(
    # Fill any missing values with NA and use the same column names as year 2006
    assign(filenames[i], read.table(urls[i], header = FALSE,
                                    fill = TRUE, col.names = colnames(mr2006))),
  )

}
```

Since the buoy data was collected over years, there were updates and changes on how many variables are measured. During the exploring process, it is found that the data from the year 1999 only contained 16 columns(without Tide), the year 2000 to 2004 contained 17 columns, while starting from the year 2005, the data included one extra column mm, which is the minute variable. Since the mm variable only contained values of 0 and the earlier years didn't have this information available, the mm variable will not be used in the analysis and can be removed from the data.

```
# Add Column Tide to the 1999 data
mr1999$TIDE <- NA
# Delete the mm column
n <- length(urls)
for (i in 1:n){
  file <- get(filenames[i])
  colnames(file)[1] <-"YYYY"
  if(ncol(file) == 18){
    file <- subset(file, select = -mm )
  }
  if(i == 1){
    MR <- file
  }else{
    MR <- rbind.data.frame(MR, file)
  }

}
```

Cleaning the data

The goal of this project is to find possible evidence of global warming, therefore, the possibly related variables are Air Temperature (ATMP) and Water Temperature (WTMP). So, the other variables could be removed for now.

```
MR_temp <- subset(MR, select = c(YYYY, MM, DD, ATMP, WTMP))
```

For the temperature variable, there are many missing values indicated by 999.0 (temperature could not go up to 999), which will cause issues when calculating averages and findind maximun values. Therefore, these points should be substituted by NA.

```
MR_temp[MR_temp == 999] <- NA
```

Formating the data into posix numbers with lubridate

```
library(lubridate)
```

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

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

```
# Format the year, month and time variables into one single date variable
MR_temp <- MR_temp %>%
  select(YYYY, MM, DD, ATMP, WTMP) %>%
  mutate(date = make_datetime(YYYY, MM, DD))
# The new variable is POSIX number
is.POSIXt(MR_temp$date)
```

```
## [1] TRUE
```

Filtering the data down to one data point per day (the maximum temperature)

```
library(dplyr)
MR_max <- MR_temp %>%
  group_by(date) %>%
  # Finding the max and exclude NA
  summarize(max_ATMP = max(ATMP, na.rm = TRUE),
            max_WTMP = max(WTMP, na.rm = TRUE))
```

```
## Warning in max(ATMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf
```

```
## Warning in max(ATMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf
```

```
## Warning in max(ATMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf
```

```
## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf
```

```
## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf
```

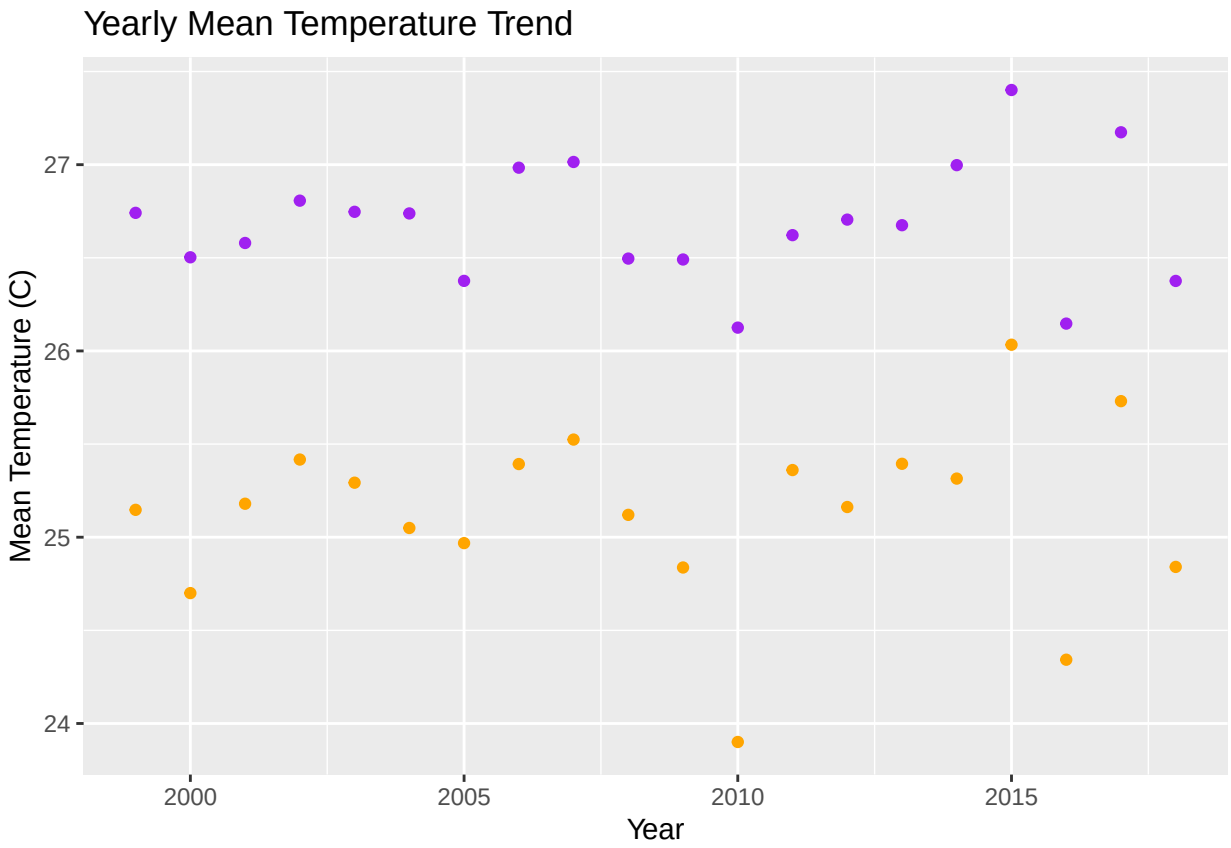
```
## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
```

```
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf

## Warning in max(WTMP, na.rm = TRUE): no non-missing arguments to max; returning -
## Inf
```

```
MR_year_mean <- MR_temp %>%
  group_by(YYYY) %>%
  # Finding the yearly average temperature and exclude NA
  summarize(mean_ATMP = mean(ATMP, na.rm = TRUE),
            mean_WTMP = mean(WTMP, na.rm = TRUE))
```

EDA

```
library(ggplot2)
ggplot(MR_year_mean, aes(YYYY, mean_ATMP)) +
  geom_point(colour = "orange") +
  geom_point(data = MR_year_mean, aes(y = mean_WTMP), colour = 'purple') +
  ggtitle("Yearly Mean Temperature Trend") +
  labs(y ="Mean Temperature (C)", x = "Year")
```


Yearly Mean Temperature Trend
Mean Temperature (C)
27
26
25
24
2000
2005
2010
2015
Year