Buoy

Simu Huang

2020/9/25

Import data

The first step is to download the data from 2000 to 2019 on the NDBC. in these data, the variable types are not exactly the same from year to year, we need to delete some columns, such as the variable tide and the duplicate variable mm. Then, we connect the year's data together to get a data frame that contains all the data.

```
library(data.table)
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::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```
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
```

```
setwd("D:\\Documents\\R Project\\615\\buoy project\\data")

#2000
Buoy2000<-read.csv("buoy2000.txt",sep="",header = T)
Buoy2000$TIDE<-0
Buoy2000$mm<-0
colnames(Buoy2000)[1]<-"X.YY"
```

```
colnames(Buoy2000)[5]<-"WDIR"
colnames(Buoy2000)[12]<-"PRES"

#2001
Buoy2001<-read.csv("buoy2001.txt",sep="",header = T)
Buoy2001$mm<-0
colnames(Buoy2001)[1]<-"X.YY"
colnames(Buoy2001)[5]<-"WDIR"
colnames(Buoy2001)[12]<-"PRES"


#2002
Buoy2002<-read.csv("buoy2002.txt",sep="",header = T)
Buoy2002$mm<-0
colnames(Buoy2002)[1]<-"X.YY"
colnames(Buoy2002)[5]<-"WDIR"
colnames(Buoy2002)[12]<-"PRES"

#2003
Buoy2003<-read.csv("buoy2003.txt",sep="",header = T)
Buoy2003$mm<-0
colnames(Buoy2003)[1]<-"X.YY"
colnames(Buoy2003)[5]<-"WDIR"
colnames(Buoy2003)[12]<-"PRES"

#2004
Buoy2004<-read.csv("buoy2004.txt",sep="",header = T)
Buoy2004$mm<-0
colnames(Buoy2004)[1]<-"X.YY"
colnames(Buoy2004)[5]<-"WDIR"
colnames(Buoy2004)[12]<-"PRES"

#2005
Buoy2005<-read.csv("buoy2005.txt",sep="",header = T)
mm2005<-Buoy2005$mm
Buoy2005$mm<-NULL
Buoy2005$mm<-mm2005
colnames(Buoy2005)[1]<-"X.YY"
colnames(Buoy2005)[5]<-"WDIR"
colnames(Buoy2005)[12]<-"PRES"

#2006
Buoy2006<-read.csv("buoy2006.txt",sep="",header = T)
mm2006<-Buoy2006$mm
Buoy2006$mm<-NULL
Buoy2006$mm<-mm2006
colnames(Buoy2006)[1]<-"X.YY"
colnames(Buoy2006)[5]<-"WDIR"
colnames(Buoy2006)[12]<-"PRES"

#2007
Buoy2007<-read.csv("buoy2007.txt",sep="",header = T)
mm2007<-Buoy2007$mm
```

```
Buoy2007$mm<-NULL
Buoy2007$mm<-mm2007
Buoy2007 = Buoy2007[-c(1),]

#2008
Buoy2008<-read.csv("buoy2008.txt",sep="",header = T)
mm2008<-Buoy2008$mm
Buoy2008$mm<-NULL
Buoy2008$mm<-mm2008
Buoy2008 = Buoy2008[-c(1),]


#2009
Buoy2009<-read.csv("buoy2009.txt",sep="",header = T)
mm2009<-Buoy2009$mm
Buoy2009$mm<-NULL
Buoy2009$mm<-mm2009
Buoy2009 = Buoy2009[-c(1),]

#2010
Buoy2010<-read.csv("buoy2010.txt",sep="",header = T)
mm2010<-Buoy2010$mm
Buoy2010$mm<-NULL
Buoy2010$mm<-mm2010
Buoy2010 = Buoy2010[-c(1),]

#2011
Buoy2011<-read.csv("buoy2011.txt",sep="",header = T)
mm2011<-Buoy2011$mm
Buoy2011$mm<-NULL
Buoy2011$mm<-mm2011
Buoy2011 = Buoy2011[-c(1),]

#2012
Buoy2012<-read.csv("buoy2012.txt",sep="",header = T)
mm2012<-Buoy2012$mm
Buoy2012$mm<-NULL
Buoy2012$mm<-mm2012
Buoy2012 = Buoy2012[-c(1),]

#2013
Buoy2013<-read.csv("buoy2013.txt",sep="",header = T)
mm2013<-Buoy2013$mm
Buoy2013$mm<-NULL
Buoy2013$mm<-mm2013
Buoy2013 = Buoy2013[-c(1),]

#2014
Buoy2014<-read.csv("buoy2014.txt",sep="",header = T)
mm2014<-Buoy2014$mm
Buoy2014$mm<-NULL
Buoy2014$mm<-mm2014
Buoy2014 = Buoy2014[-c(1),]
```

```
#2015
Buoy2015<-read.csv("buoy2015.txt",sep="",header = T)
mm2015<-Buoy2015$mm
Buoy2015$mm<-NULL
Buoy2015$mm<-mm2015
Buoy2015 = Buoy2015[-c(1),]

#2016
Buoy2016<-read.csv("buoy2016.txt",sep="",header = T)
mm2016<-Buoy2016$mm
Buoy2016$mm<-NULL
Buoy2016$mm<-mm2016
Buoy2016 = Buoy2016[-c(1),]

#2017
Buoy2017<-read.csv("buoy2017.txt",sep="",header = T)
mm2017<-Buoy2017$mm
Buoy2017$mm<-NULL
Buoy2017$mm<-mm2017
Buoy2017 = Buoy2017[-c(1),]

#2018
Buoy2018<-read.csv("buoy2018.txt",sep="",header = T)
mm2018<-Buoy2018$mm
Buoy2018$mm<-NULL
Buoy2018$mm<-mm2018
Buoy2018 = Buoy2018[-c(1),]

#2019
Buoy2019<-read.csv("buoy2019.txt",sep="",header = T)
mm2019<-Buoy2019$mm
Buoy2019$mm<-NULL
Buoy2019$mm<-mm2019
Buoy2019 = Buoy2019[-c(1),]

#Combine

Buoy<-rbind(Buoy2000,Buoy2001,Buoy2002,Buoy2003,Buoy2004,Buoy2005,Buoy2006,Buoy2007,Buoy2008,Buoy2009,Bu
```

There are some outliers, like temperatures above 50 degrees and some observations of 999. These are obviously not practical. Therefore, we only choose the observations meeting our requirements to create a new date set.

```
Buoy[Buoy==99|Buoy==999|Buoy==9999]<-0

Buoy <- Buoy [, -11]
Buoy <- Buoy [, -15:-17]

#remove outliers
a <- Buoy %>% select(X.YY, MM, DD, hh)

a2 <- as.data.frame(lapply(a,as.numeric))
```

```
D_date<- make_datetime(year=a2$X.YY, month=a2$MM, day=a2$DD, hour=a2$hh)
Buoy$D_date <- D_date
cols <- colnames(Buoy)
new_cols <- c(cols[1], cols[length(cols)], cols[2:(length(cols) - 1)])
Buoy <- Buoy[, new_cols]
Buoy <- Buoy[,c(-1,-3:-5)]
Buoy <- filter(Buoy, DEWP < 999 & ATMP < 50 & WTMP < 50)
Buoy$D_date <- as.POSIXct(Buoy$D_date)
Buoy[,2:11]<-lapply(Buoy[,2:11],as.numeric)

save(Buoy, file = "Buoy.Rdata")
```

We think it is possible to compare the change in temperature according to the average monthly and yearly temperature. In this step, we reorder the data by date, month and year and calculate the mean value of these data, because in this way, We can use the average daily temperature to conclude the temperature trend, and the average monthly temperature and the average annual temperature can eliminate the influence of seasons and extreme weather on the results.

```
Buoy <- filter(Buoy, DEWP < 999 & ATMP < 50 & WTMP < 50)
Buoy$D_date <- as.POSIXct(Buoy$D_date)
Buoy[,2:11]<-lapply(Buoy[,2:11],as.numeric)

Buoy2 <- Buoy %>% group_by(date(D_date)) %>% summarize(mean(WDIR),mean(WSPD),mean(GST),mean(WVHT),mean(
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Buoy3 <- Buoy %>% group_by(year(D_date),month(D_date)) %>% summarize(mean(WDIR),mean(WSPD),mean(GST),mea
```

```
## `summarise()` regrouping output by 'year(D_date)' (override with `.groups` argument)
```

```
Buoy4 <- Buoy %>% group_by(year(D_date)) %>% summarize(mean(WDIR),mean(WSPD),mean(GST),mean(WVHT),mean(
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
b <- Buoy3 %>% select(`year(D_date)`,`month(D_date)`)
b2 <- as.data.frame(lapply(b,as.numeric))
year_month <- make_date(year = b$`year(D_date)`,month = b$`month(D_date)`)
Buoy3$year_month<-year_month
```

#ATMP As the simple regression model, we can find that for each year of time increase, the predicted annual increase in air temperature is 0.099, which means that the global warming is indeed happening. And we find that the relationship between averaged wind speed over the an eight-minute period and the dominant wave period is not obvious so we delete these two variable. The two factors that best reflect change are the gust speed(GST) and the wind speed (WSPD).

#WTMP In this model, we can find that for each year of time increase, the predicted annual increase in sea surface temperature is 0.1155, which also indicates that the global warming is indeed happening. And we find that the relationship between averaged wind speed over the an eight minute period, wind direction, and the dominant wave period is not obvious so we delete these three variable. The two factors that best reflect change is the wind speed (WSPD).

```
#ATMP
fit_in <- lm (Buoy4$`mean(ATMP)` ~ Buoy4$`year(D_date)`)
```

```
fit_year <- lm ( Buoy4$`mean(ATMP)` ~ Buoy4$`mean(WDIR)`+ Buoy4$`mean(WSPD)`+Buoy4$`mean(GST)`+ Buoy4$`
fit_month <- lm ( Buoy3$`mean(ATMP)` ~ Buoy3$`mean(WDIR)`+ Buoy3$`mean(WSPD)`+Buoy3$`mean(GST)`+ Buoy3$
fit_day <- lm ( Buoy2$`mean(ATMP)` ~ Buoy2$`mean(WDIR)`+ Buoy2$`mean(WSPD)`+Buoy2$`mean(GST)`+ Buoy2$`m
```

```
summary(fit_in)
```

```
##
## Call:
## lm(formula = Buoy4$`mean(ATMP)` ~ Buoy4$`year(D_date)`)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.5612  -0.7517  -0.1068   1.0110   5.7799
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -186.68592  304.87564  -0.612    0.548
## Buoy4$`year(D_date)`    0.09854    0.15175   0.649    0.525
##
## Residual standard error: 3.735 on 17 degrees of freedom
## Multiple R-squared:  0.0242, Adjusted R-squared:  -0.0332
## F-statistic: 0.4217 on 1 and 17 DF,  p-value: 0.5248
```

```
summary(fit_year)
```

```
##
## Call:
## lm(formula = Buoy4$`mean(ATMP)` ~ Buoy4$`mean(WDIR)` + Buoy4$`mean(WSPD)` +
##     Buoy4$`mean(GST)` + Buoy4$`mean(WVHT)` + Buoy4$`mean(APD)`,
##     data = Buoy4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3133 -0.9351 -0.4616  0.8453  2.4544
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         61.098170  19.284352   3.168  0.00741 **
## Buoy4$`mean(WDIR)`   0.002227   0.024630   0.090  0.92932
## Buoy4$`mean(WSPD)`  20.970441  12.552585   1.671  0.11869
## Buoy4$`mean(GST)`  -21.752416   9.381993  -2.319  0.03735 *
## Buoy4$`mean(WVHT)`   4.936915   2.805382   1.760  0.10194
## Buoy4$`mean(APD)`   -5.054957   2.972266  -1.701  0.11278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.599 on 13 degrees of freedom
## Multiple R-squared:  0.8632, Adjusted R-squared:  0.8106
## F-statistic:  16.4 on 5 and 13 DF,  p-value: 3.255e-05
```

```
summary(fit_month)
```

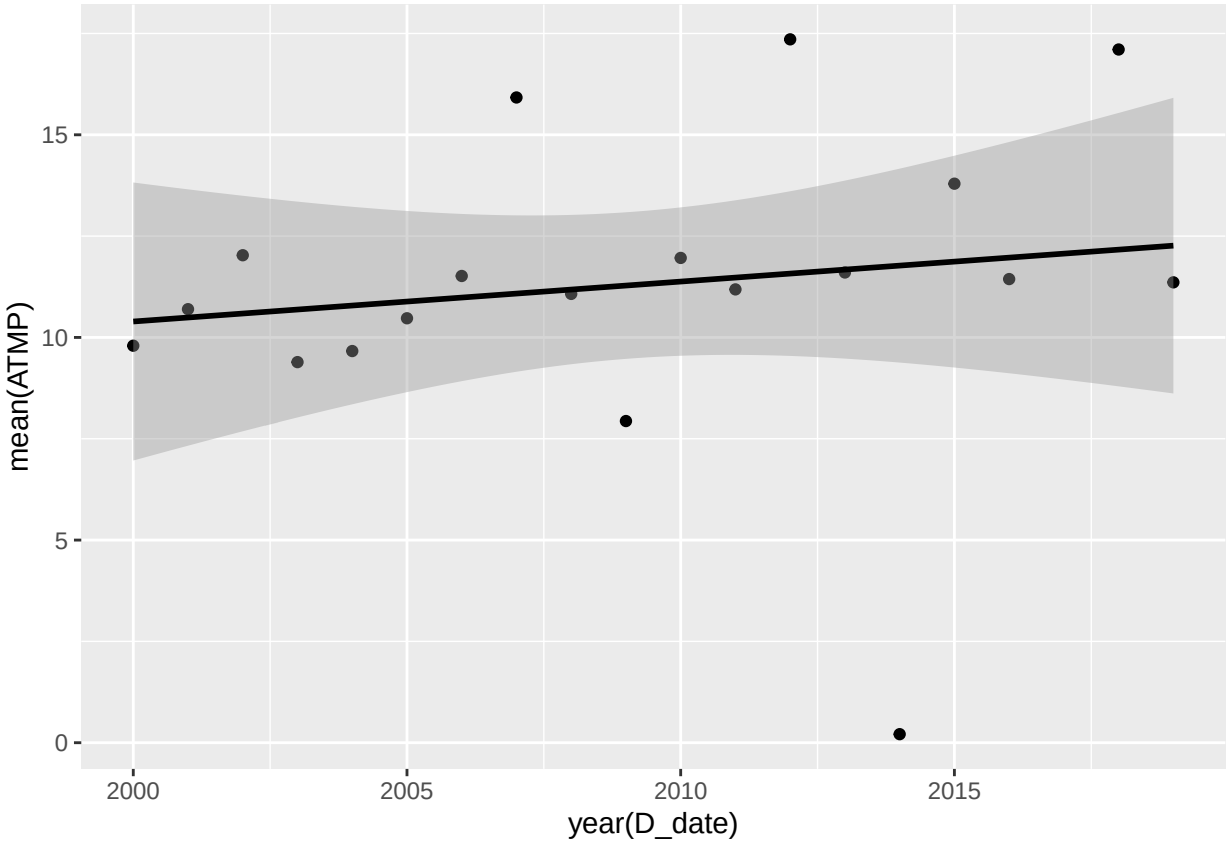
```
##
## Call:
```

```
## lm(formula = Buoy3$`mean(ATMP)` ~ Buoy3$`mean(WDIR)` + Buoy3$`mean(WSPD)` +
##     Buoy3$`mean(GST)` + Buoy3$`mean(WVHT)` + Buoy3$`mean(APD)`,
##     data = Buoy3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.9708  -3.0833   0.2779   3.6329  17.4774
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.139121   4.983003   0.429   0.6682
## Buoy3$`mean(WDIR)`    0.015718   0.008287   1.897   0.0594 .
## Buoy3$`mean(WSPD)`   14.446028   2.514120   5.746 3.61e-08 ***
## Buoy3$`mean(GST)`   -13.352001   1.921131  -6.950 5.75e-11 ***
## Buoy3$`mean(WVHT)`   -3.649850   0.787801  -4.633 6.70e-06 ***
## Buoy3$`mean(APD)`     3.867423   0.829642   4.662 5.92e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.785 on 189 degrees of freedom
## Multiple R-squared:  0.6345, Adjusted R-squared:  0.6248
## F-statistic: 65.62 on 5 and 189 DF,  p-value: < 2.2e-16
```

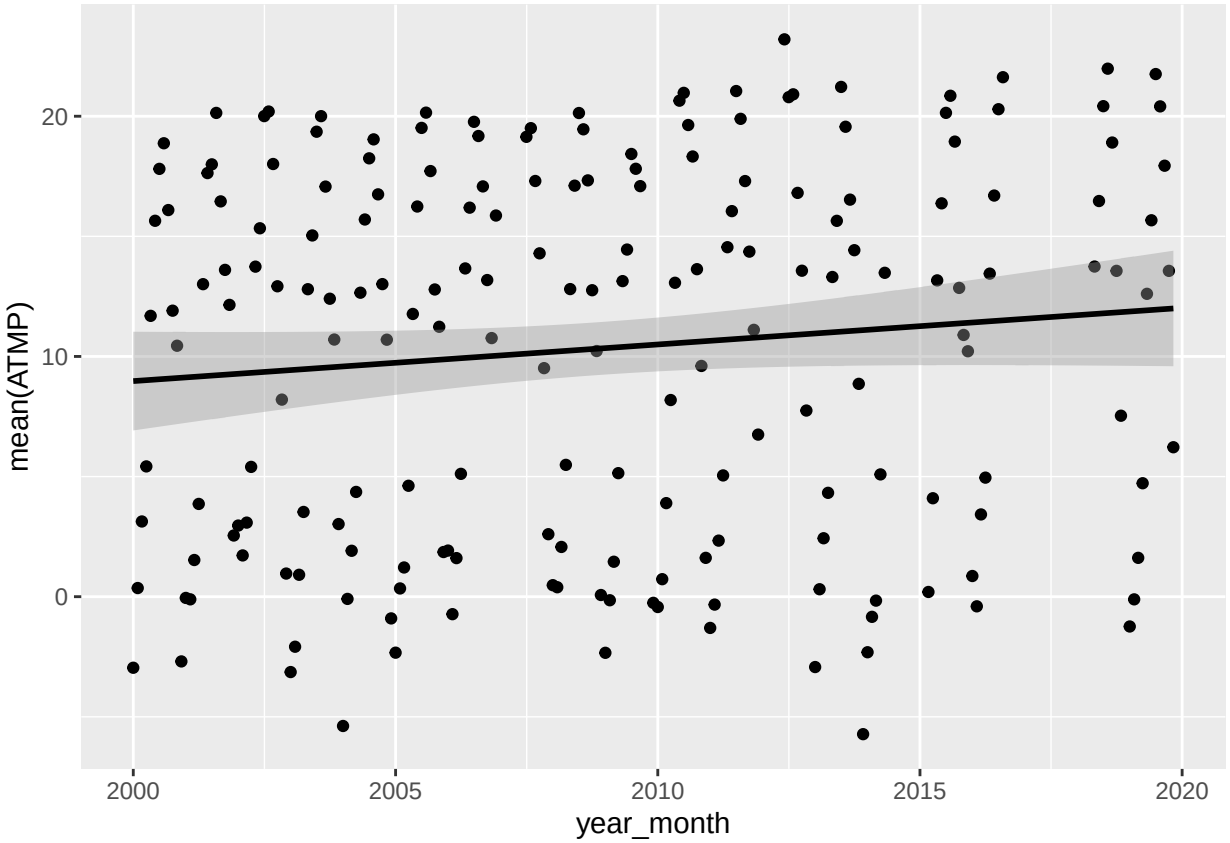
```
summary(fit_day)
```

```
##
## Call:
## lm(formula = Buoy2$`mean(ATMP)` ~ Buoy2$`mean(WDIR)` + Buoy2$`mean(WSPD)` +
##     Buoy2$`mean(GST)` + Buoy2$`mean(WVHT)` + Buoy2$`mean(APD)`,
##     data = Buoy2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.663  -5.304   1.159   5.332  37.159
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          13.918794   0.634095  21.951  < 2e-16 ***
## Buoy2$`mean(WDIR)` -0.006175   0.001278  -4.832 1.39e-06 ***
## Buoy2$`mean(WSPD)`   2.642824   0.240083  11.008  < 2e-16 ***
## Buoy2$`mean(GST)`   -3.149606   0.189065 -16.659  < 2e-16 ***
## Buoy2$`mean(WVHT)` -1.164670   0.106278 -10.959  < 2e-16 ***
## Buoy2$`mean(APD)`    1.202390   0.110128  10.918  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.797 on 4809 degrees of freedom
## Multiple R-squared:  0.3128, Adjusted R-squared:  0.3121
## F-statistic: 437.9 on 5 and 4809 DF,  p-value: < 2.2e-16
```

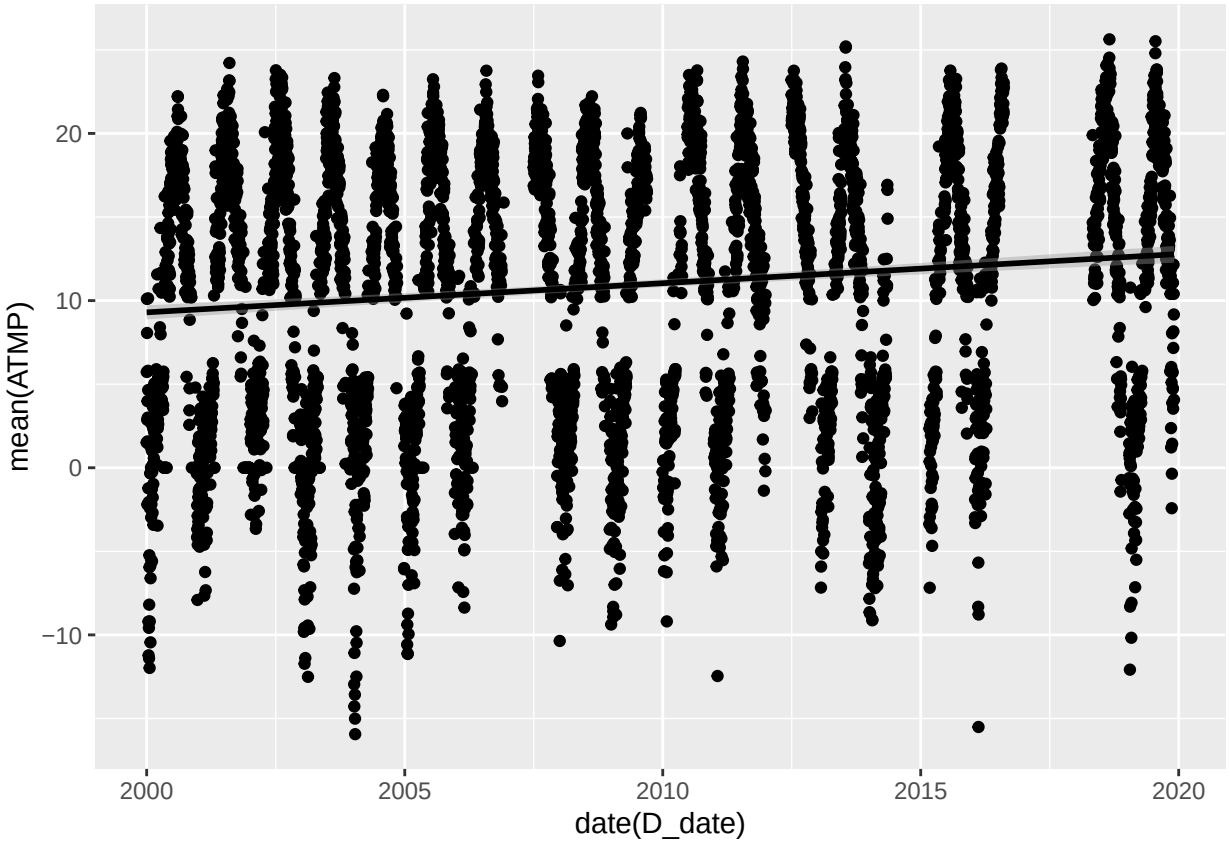
```
ggplot(mapping = aes(x= `year(D_date)`, y= `mean(ATMP)`),data = Buoy4)+ geom_point() + geom_smooth(metho
```




```
ggplot(mapping = aes(x= year_month, y= `mean(ATMP)`),data = Buoy3)+ geom_point() + geom_smooth(method =
```




```
ggplot(mapping = aes(x= `date(D_date)`, y= `mean(ATMP)`),data = Buoy2)+ geom_point() + geom_smooth(metho
```




```
#WTMP
fit_in_1 <- lm (Buoy4$`mean(WTMP)` ~ Buoy4$`year(D_date)`)
fit_year_1 <- lm ( Buoy4$`mean(WTMP)` ~  Buoy4$`mean(WSPD)`+Buoy4$`mean(GST)`+ Buoy4$`mean(WVHT)`+ Buoy
fit_month_1 <- lm ( Buoy3$`mean(WTMP)` ~  Buoy3$`mean(WSPD)`+Buoy3$`mean(GST)`+ Buoy3$`mean(WVHT)`+ Buo
fit_day_1 <- lm ( Buoy2$`mean(WTMP)` ~  Buoy2$`mean(WSPD)`+Buoy2$`mean(GST)`+ Buoy2$`mean(WVHT)`+ Buoy2
```

```
summary(fit_in_1)
```

```
##
## Call:
## lm(formula = Buoy4$`mean(WTMP)` ~ Buoy4$`year(D_date)`)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8836 -0.6753 -0.2132  0.6987  4.8491
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -219.9554   241.0469  -0.913    0.374
## Buoy4$`year(D_date)`    0.1155     0.1200   0.963    0.349
##
## Residual standard error: 2.953 on 17 degrees of freedom
## Multiple R-squared:  0.05168,    Adjusted R-squared:  -0.004101
## F-statistic: 0.9265 on 1 and 17 DF,  p-value: 0.3493
```

```r
summary(fit_year_1)
```

```
##
## Call:
## lm(formula = Buoy4$`mean(WTMP)` ~ Buoy4$`mean(WSPD)` + Buoy4$`mean(GST)` +
##     Buoy4$`mean(WVHT)` + Buoy4$`mean(APD)`, data = Buoy4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8620 -0.8884 -0.4051  0.5711  2.6702
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          50.880     17.319   2.938   0.0108 *
## Buoy4$`mean(WSPD)`   14.629     11.515   1.270   0.2246
## Buoy4$`mean(GST)`   -15.603      8.620  -1.810   0.0918 .
## Buoy4$`mean(WVHT)`    3.709      2.547   1.456   0.1674
## Buoy4$`mean(APD)`    -3.758      2.712  -1.386   0.1875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.474 on 14 degrees of freedom
## Multiple R-squared:  0.8054, Adjusted R-squared:  0.7498
## F-statistic: 14.49 on 4 and 14 DF,  p-value: 7.012e-05
```

```r
summary(fit_month_1)
```

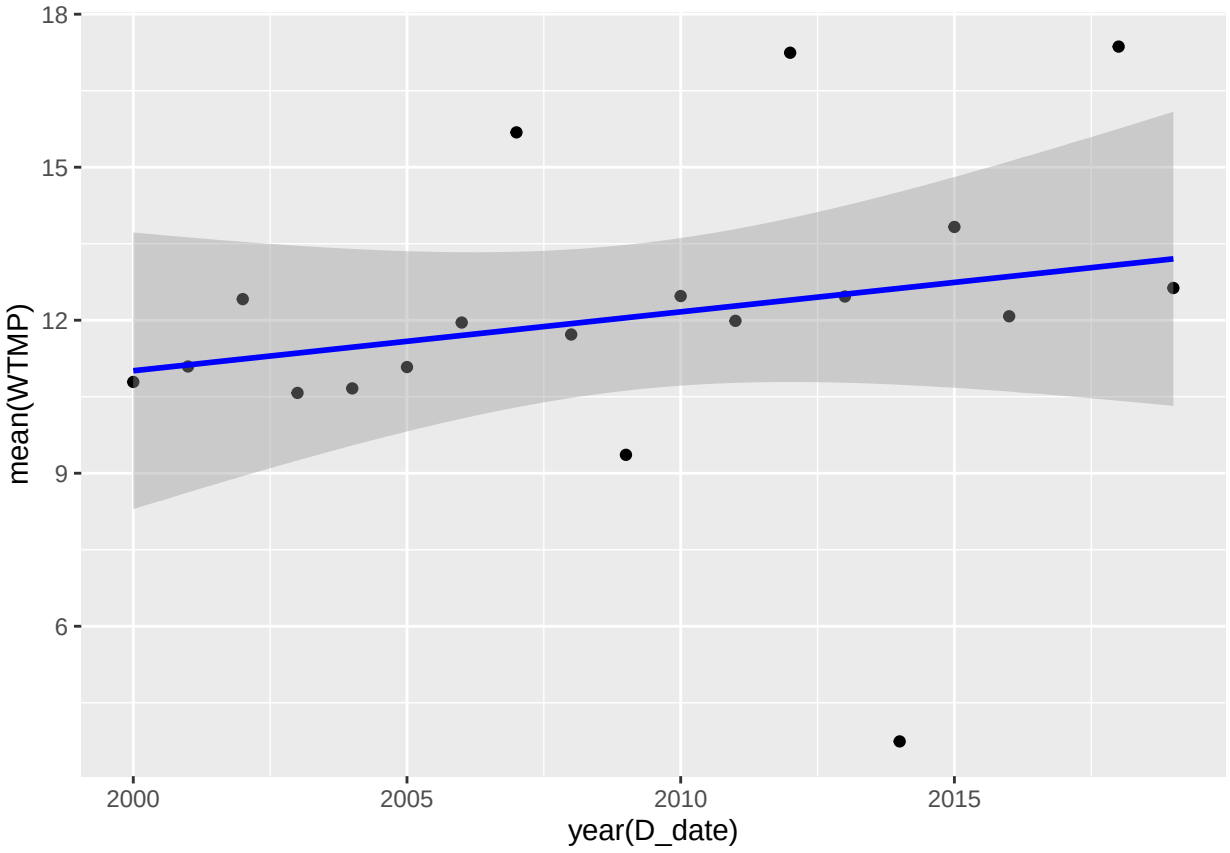
```
##
## Call:
## lm(formula = Buoy3$`mean(WTMP)` ~ Buoy3$`mean(WSPD)` + Buoy3$`mean(GST)` +
##     Buoy3$`mean(WVHT)` + Buoy3$`mean(APD)`, data = Buoy3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6670  -3.3264   0.4011   3.5464   8.6637
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         10.4339     4.3943   2.374  0.01857 *
## Buoy3$`mean(WSPD)`   5.5248     2.3091   2.393  0.01770 *
## Buoy3$`mean(GST)`   -5.7000     1.7544  -3.249  0.00137 **
## Buoy3$`mean(WVHT)`  -2.1256     0.7252  -2.931  0.00379 **
## Buoy3$`mean(APD)`    2.2083     0.7638   2.891  0.00429 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.43 on 190 degrees of freedom
## Multiple R-squared:  0.4636, Adjusted R-squared:  0.4523
## F-statistic: 41.06 on 4 and 190 DF,  p-value: < 2.2e-16
```

```r
summary(fit_day_1)
```

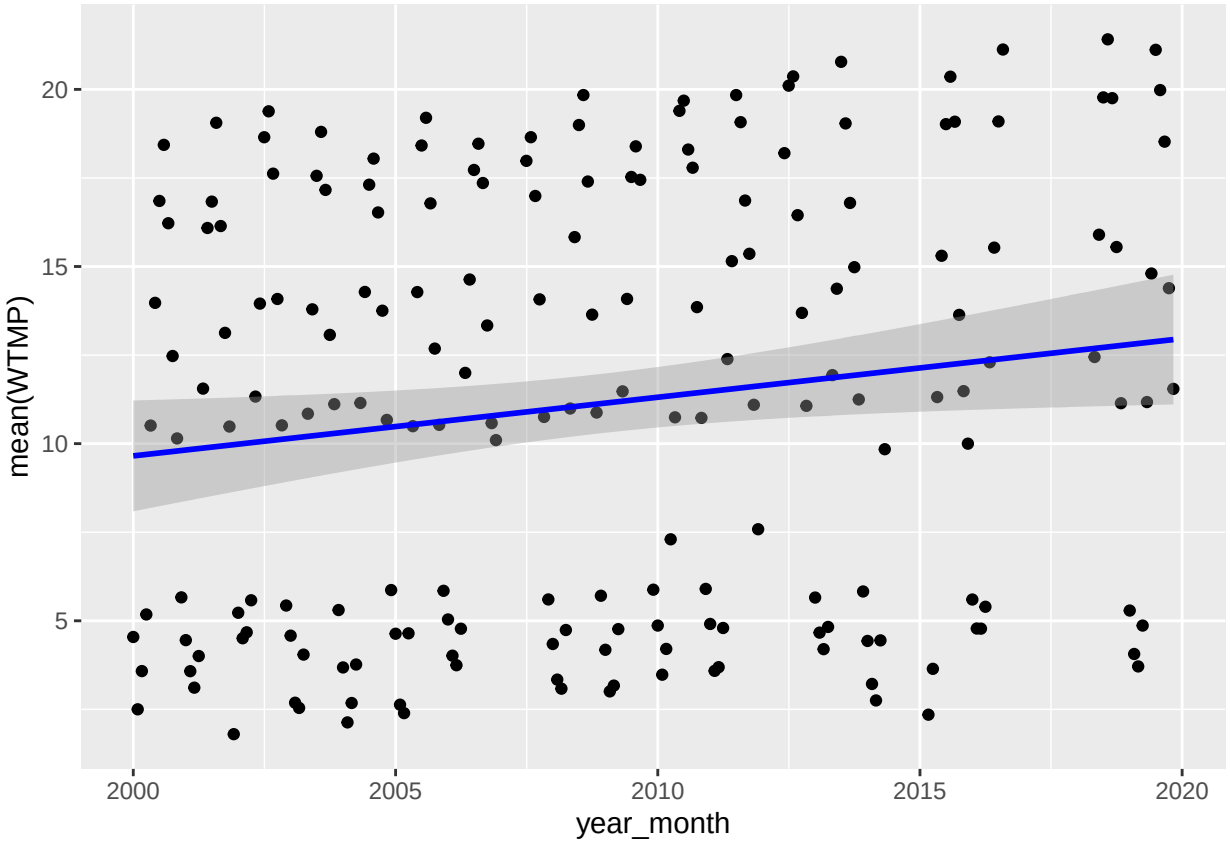
```
##
## Call:
## lm(formula = Buoy2$`mean(WTMP)` ~ Buoy2$`mean(WSPD)` + Buoy2$`mean(GST)` +
```

```
##     Buoy2$`mean(WVHT)` + Buoy2$`mean(APD)`, data = Buoy2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.1775  -4.5423   0.7824   4.6511  20.6362
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        12.18223    0.50858  23.953  < 2e-16 ***
## Buoy2$`mean(WSPD)`  0.94192    0.19928   4.727 2.35e-06 ***
## Buoy2$`mean(GST)`  -1.39681    0.15657  -8.921  < 2e-16 ***
## Buoy2$`mean(WVHT)` -0.88526    0.08814 -10.044  < 2e-16 ***
## Buoy2$`mean(APD)`   0.91045    0.09135   9.967  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.645 on 4810 degrees of freedom
## Multiple R-squared:  0.1989, Adjusted R-squared:  0.1983
## F-statistic: 298.6 on 4 and 4810 DF,  p-value: < 2.2e-16
```

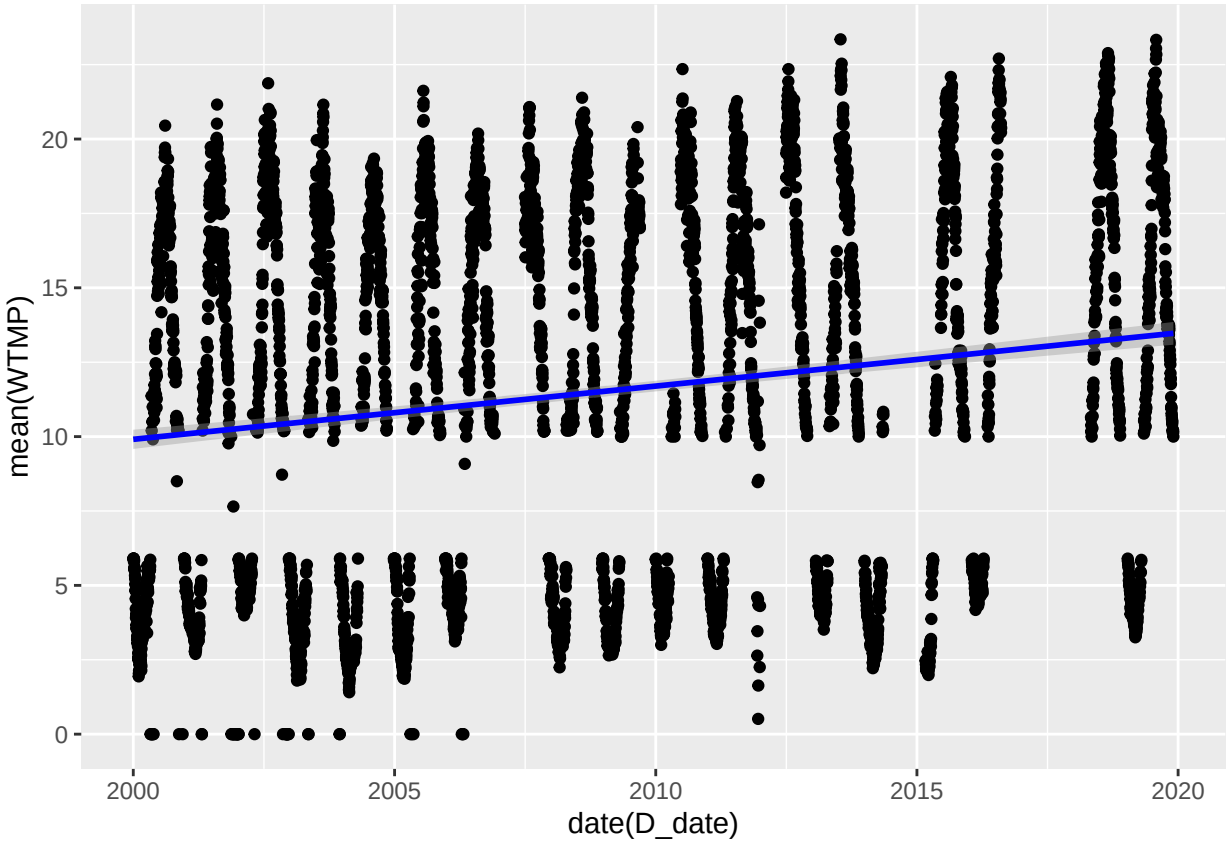
```
ggplot(mapping = aes(x= `year(D_date)`, y= `mean(WTMP)`),data = Buoy4)+ geom_point() + geom_smooth(metho
```




```
ggplot(mapping = aes(x= year_month, y= `mean(WTMP)`),data = Buoy3)+ geom_point() + geom_smooth(method =
```




```
ggplot(mapping = aes(x= `date(D_date)`, y= `mean(WTMP)`),data = Buoy2)+ geom_point() + geom_smooth(metho
```



Conclusion

In conclusion, we could find that the global warming is happening base on the analysis of the data of the weather buoy from 2000 to 2019. Although there are several factors related to the global warming, in our data sets, the factor WSPD (Wind speed) is closely related to the air temperature and sea surface temperature. However, the data is only collected from Boston area, it cannot represent the global trend.

Reference

Packages used:“ggplot2”,“tidyverse”,“lubridate”,“data.table”