Buoy project

Haoqi Wang

9/25/2020

1.1 Overview

I read 20-years of data NDBC Station 44013, which is a platform to provide quality observations in the marine environment in a safe and sustainable manner to support the understanding of and predictions to changes in weather, climate, oceans and coast.. By downloading data from 44013, I collected the data from 1984 to 2019, which is included many variables,like air temperature, water temperature and so forth. You may know the average global temperature has increased in recorded history. And experts see the trend is accelerating.So around this topic, I did a project and report to indicate whether the global warming exists or not.

1.2 Outline

The outline of this report is as follows. Firstly, I display the process of data cleaning. Next, I show the data analysis via a statistical model. Lastly, we interpret our model and discuss the conclusions.

2.1 import data

I read 20-year of data from NDBC Station 44013. There are 20 files in my local file folder. According to look through the data, I found some of data columns' names are different. In order to make data more clear, I delete the "TIDE" column for 2000, and delete the "mm" column, which is the minutes.

```
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
```

```
library(ggplot2)
library(citation)
#import data
#2000
Buoy2000<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file00.txt"
Buoy2000$TIDE<-0
Buoy2000$mm<-0
colnames(Buoy2000)[1]<-"X.YY"
colnames(Buoy2000)[5]<-"WDIR"
colnames(Buoy2000)[12]<-"PRES"

#2001
Buoy2001<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file01.txt"
Buoy2001$mm<-0
colnames(Buoy2001)[1]<-"X.YY"
colnames(Buoy2001)[5]<-"WDIR"
colnames(Buoy2001)[12]<-"PRES"


#2002
Buoy2002<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file02.txt"
Buoy2002$mm<-0
colnames(Buoy2002)[1]<-"X.YY"
colnames(Buoy2002)[5]<-"WDIR"
colnames(Buoy2002)[12]<-"PRES"

#2003
Buoy2003<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file03.txt"
Buoy2003$mm<-0
colnames(Buoy2003)[1]<-"X.YY"
colnames(Buoy2003)[5]<-"WDIR"
colnames(Buoy2003)[12]<-"PRES"

#2004
Buoy2004<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file04.txt"
Buoy2004$mm<-0
colnames(Buoy2004)[1]<-"X.YY"
colnames(Buoy2004)[5]<-"WDIR"
colnames(Buoy2004)[12]<-"PRES"

#2005
Buoy2005<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file05.txt"
mm2005<-Buoy2005$mm
Buoy2005$mm<-NULL
Buoy2005$mm<-mm2005
colnames(Buoy2005)[1]<-"X.YY"
colnames(Buoy2005)[5]<-"WDIR"
colnames(Buoy2005)[12]<-"PRES"

#2006
```

```r
Buoy2006<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file06.txt"
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("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file07.txt"
mm2007<-Buoy2007$mm
Buoy2007$mm<-NULL
Buoy2007$mm<-mm2007
Buoy2007 = Buoy2007[-c(1),]

#2008
Buoy2008<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file08.txt"
mm2008<-Buoy2008$mm
Buoy2008$mm<-NULL
Buoy2008$mm<-mm2008
Buoy2008 = Buoy2008[-c(1),]


#2009
Buoy2009<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file09.txt"
mm2009<-Buoy2009$mm
Buoy2009$mm<-NULL
Buoy2009$mm<-mm2009
Buoy2009 = Buoy2009[-c(1),]

#2010
Buoy2010<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file10.txt"
mm2010<-Buoy2010$mm
Buoy2010$mm<-NULL
Buoy2010$mm<-mm2010
Buoy2010 = Buoy2010[-c(1),]

#2011
Buoy2011<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file11.txt"
mm2011<-Buoy2011$mm
Buoy2011$mm<-NULL
Buoy2011$mm<-mm2011
Buoy2011 = Buoy2011[-c(1),]

#2012
Buoy2012<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file12.txt"
mm2012<-Buoy2012$mm
Buoy2012$mm<-NULL
Buoy2012$mm<-mm2012
Buoy2012 = Buoy2012[-c(1),]

#2013
Buoy2013<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file13.txt"
```

```
mm2013<-Buoy2013$mm
Buoy2013$mm<-NULL
Buoy2013$mm<-mm2013
Buoy2013 = Buoy2013[-c(1),]

#2014
Buoy2014<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file14.txt"
mm2014<-Buoy2014$mm
Buoy2014$mm<-NULL
Buoy2014$mm<-mm2014
Buoy2014 = Buoy2014[-c(1),]

#2015
Buoy2015<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file15.txt"
mm2015<-Buoy2015$mm
Buoy2015$mm<-NULL
Buoy2015$mm<-mm2015
Buoy2015 = Buoy2015[-c(1),]

#2016
Buoy2016<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file16.txt"
mm2016<-Buoy2016$mm
Buoy2016$mm<-NULL
Buoy2016$mm<-mm2016
Buoy2016 = Buoy2016[-c(1),]

#2017
Buoy2017<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file17.txt"
mm2017<-Buoy2017$mm
Buoy2017$mm<-NULL
Buoy2017$mm<-mm2017
Buoy2017 = Buoy2017[-c(1),]

#2018
Buoy2018<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file18.txt"
mm2018<-Buoy2018$mm
Buoy2018$mm<-NULL
Buoy2018$mm<-mm2018
Buoy2018 = Buoy2018[-c(1),]

#2019
Buoy2019<-read.csv("/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/view_text_file19.txt"
mm2019<-Buoy2019$mm
Buoy2019$mm<-NULL
Buoy2019$mm<-mm2019
Buoy2019 = Buoy2019[-c(1),]
```

2.2 clean data

After importing data, I use fwrite() to combine 20-year data in a csv file. Then we have 18 variables in the data frame. However, there are some columns the values are 0, I delete these columns. Since the data in data frame are not numeric, I transfer the data type. And I combine the year, month, day and hour into one variable in order to use it conveniently by using make_datetime() function. Later, I use filter() to remove the outliers.

```r
#Combine the 20 years data into one file

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


#create a file that can store the new data in local
fwrite(Buoy,"/Users/wanghaoqi/Desktop/2020 Fall/MA615/class3/MA615project1/Buoy.csv")

#clean data
Buoy[Buoy==99|Buoy==999|Buoy==9999]<-0
Buoy<-Buoy[,-11]
Buoy<-Buoy[,-15:-17]

#combine year, month, day and hour into one variable
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)])
Buoy5 <- Buoy[, new_cols]
Buoy6 <- Buoy5[,c(-1,-3:-5)]

#remove the outliers
Buoy6[,2:11]<-lapply(Buoy6[,2:11],as.numeric)
Buoy6<-filter(Buoy6,DEWP!=999&ATMP<50&WTMP<50)
```

2.3 data analysis

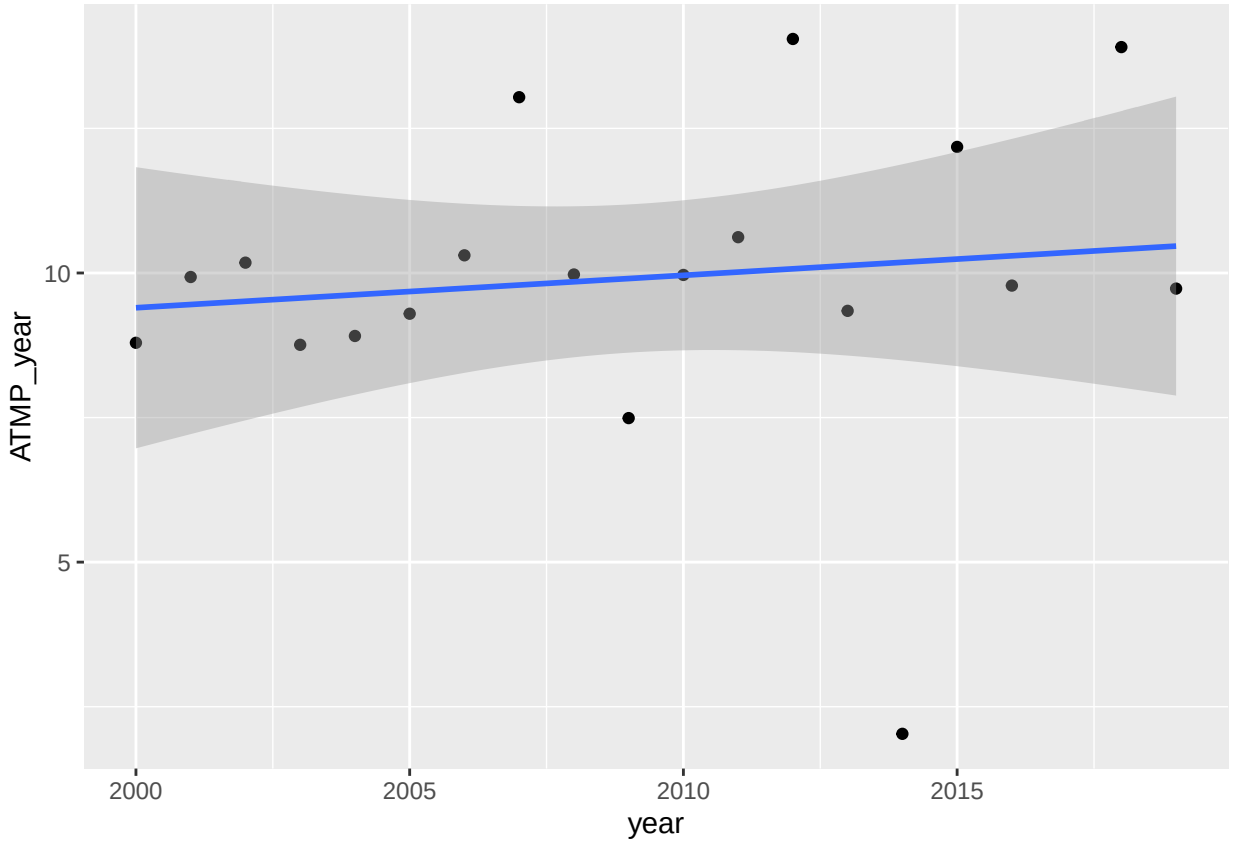
Firstly, due to the huge dataset, I decide to use the mean of every variable to check whether the global temperature gets warmer. So I use group_by() to divide the data by year, month and day, and calculate the mean of the variable.

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

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

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

Secondly, (1)I will display the trend of average air temperature yearly, monthly, and daily, (2)and I show the trend of average water temperature yearly, monthly, and daily. (1-1)In order to find the trend of average air temperature yearly, I use lm() function to find the linear regression and display the relationship by ggplot().

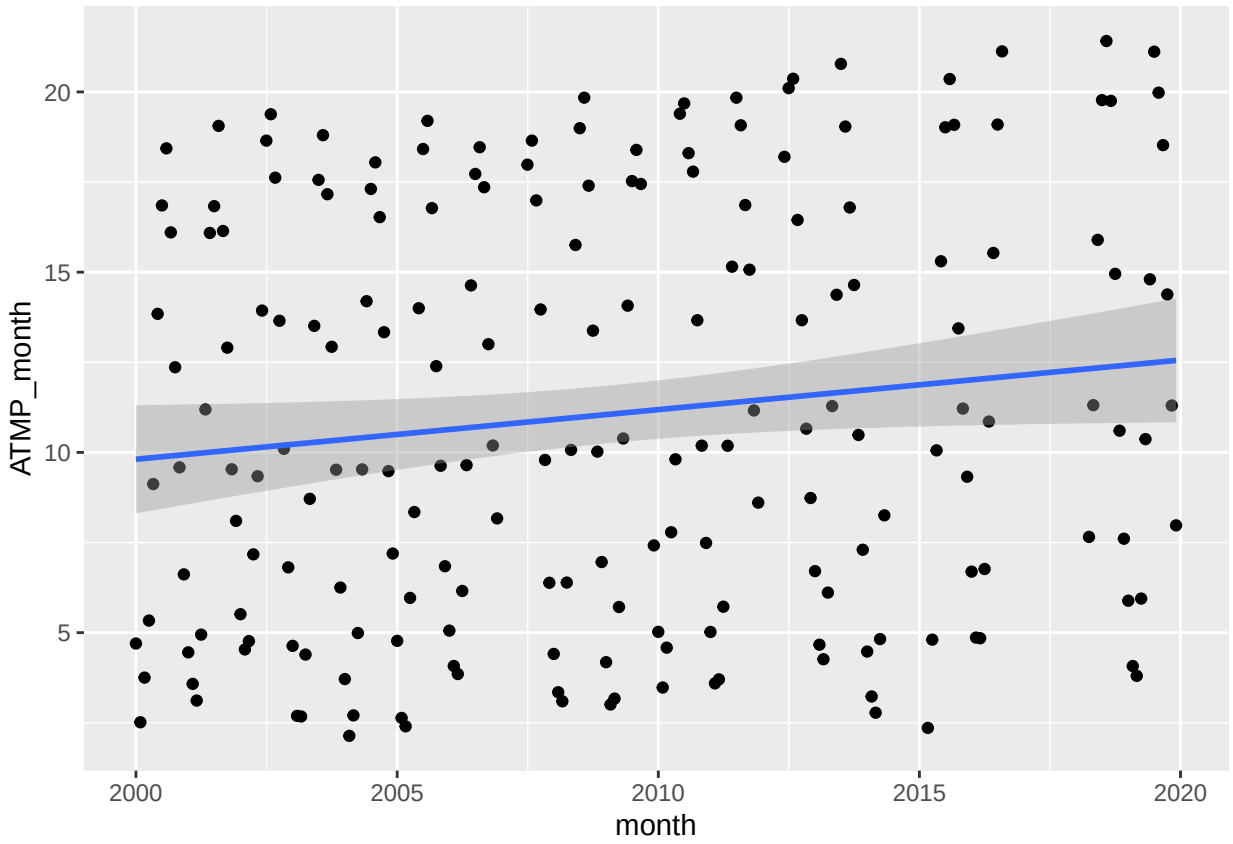
```
##
## Call:
## lm(formula = ATMP_year ~ year, data = fit1_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1529 -0.7235  0.0064  0.6355  3.9745
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -102.82021  216.06412  -0.476    0.640
## year           0.05611    0.10754   0.522    0.609
##
## Residual standard error: 2.647 on 17 degrees of freedom
```

```
## Multiple R-squared:  0.01576,    Adjusted R-squared:  -0.04214
## F-statistic: 0.2722 on 1 and 17 DF,  p-value: 0.6086
```



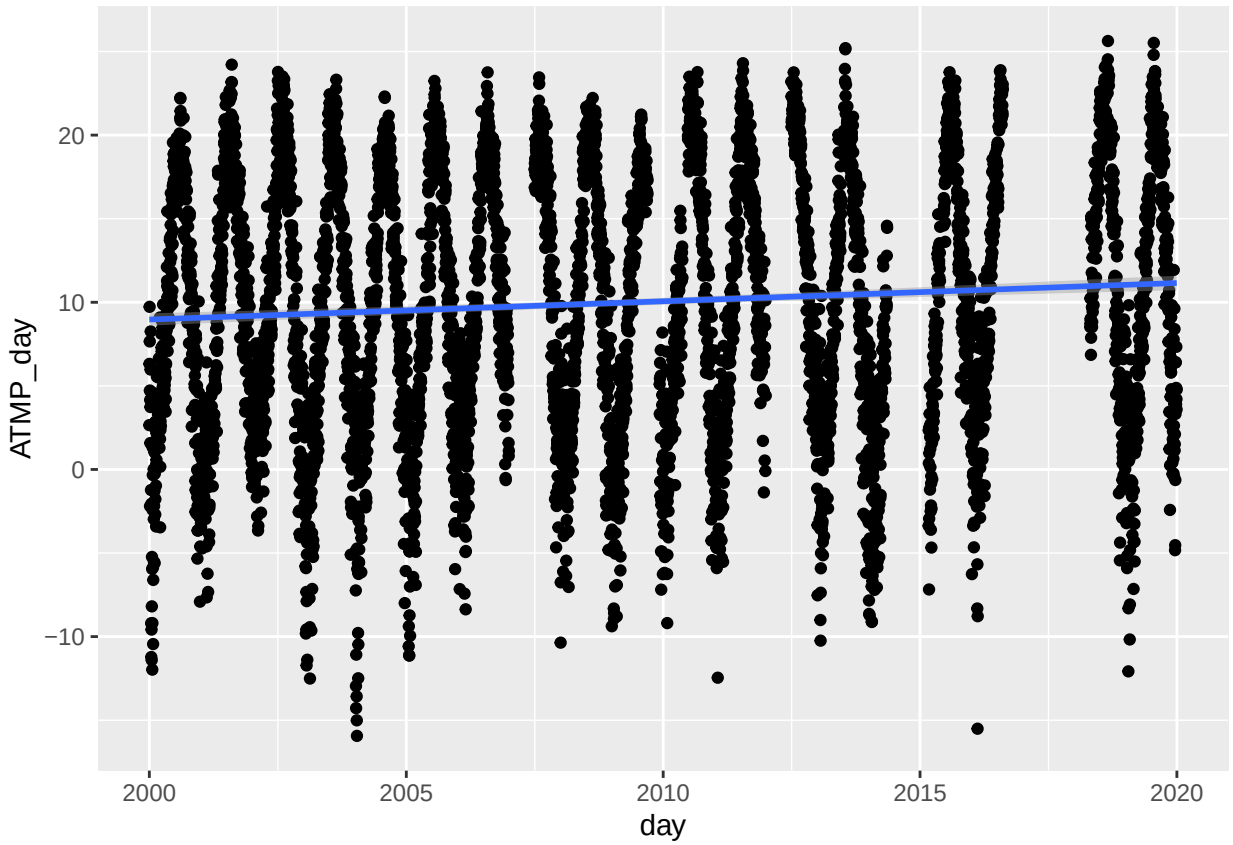
(1-2)In order to find the trend of average air temperature monthly, I use lm() function to find the linear regression and display the relationship by ggplot().

```
##
## Call:
## lm(formula = ATMP_month ~ month, data = fit2_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5393 -5.3314 -0.8873  6.0926  9.2153
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.6817453  2.8004107   2.029   0.0438 *
## month       0.0003768  0.0001942   1.940   0.0538 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.72 on 197 degrees of freedom
## Multiple R-squared:  0.01874,    Adjusted R-squared:  0.01376
## F-statistic: 3.763 on 1 and 197 DF,  p-value: 0.05383
```



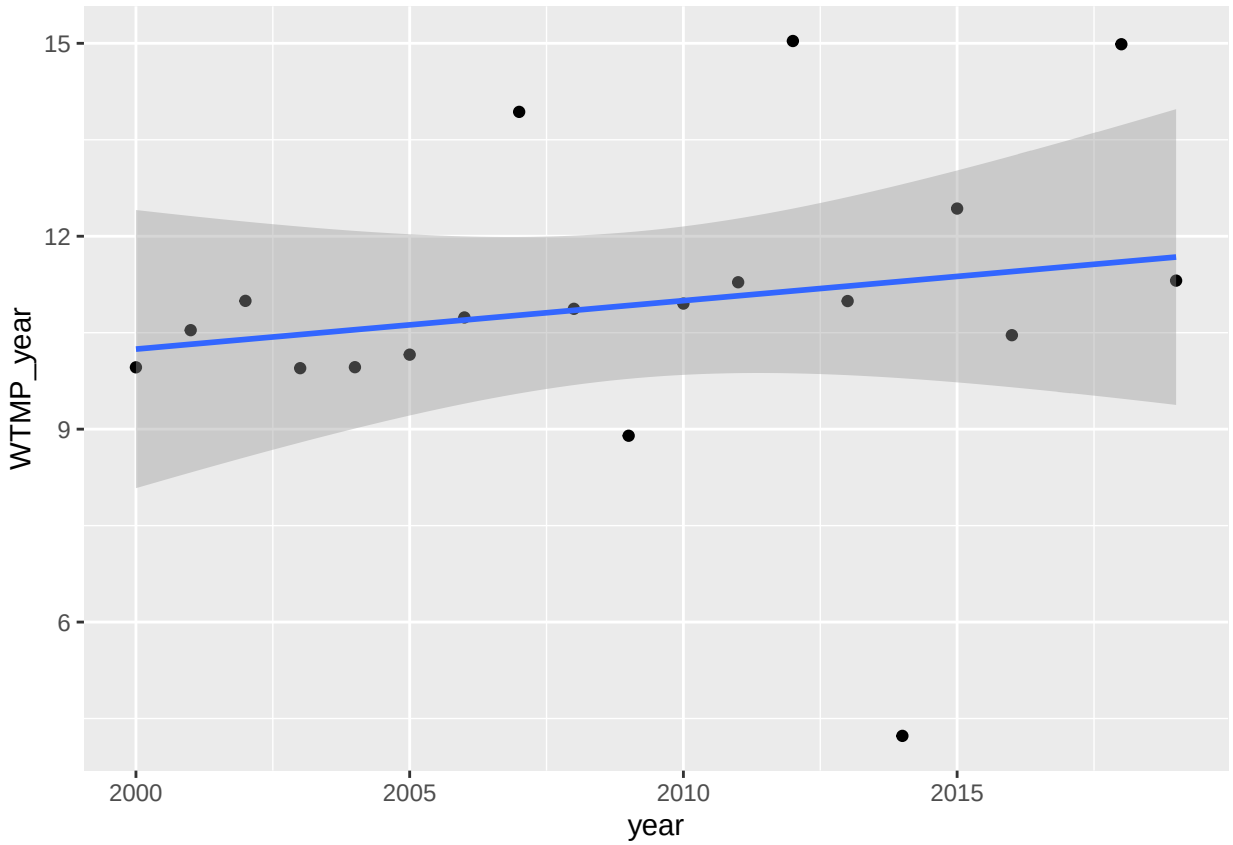
(1-3)In order to find the trend of average air temperature daily, I use lm() function to find the linear regression and display the relationship by ggplot().

```
##
## Call:
## lm(formula = ATMP_day ~ day, data = fit3_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.237  -6.108  -0.053   7.165  15.073
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.6912889  0.6985325   8.147 4.51e-16 ***
## day         0.0002991  0.0000486   6.153 8.10e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.766 on 5824 degrees of freedom
## Multiple R-squared:  0.006459,   Adjusted R-squared:  0.006288
## F-statistic: 37.86 on 1 and 5824 DF,  p-value: 8.098e-10
```



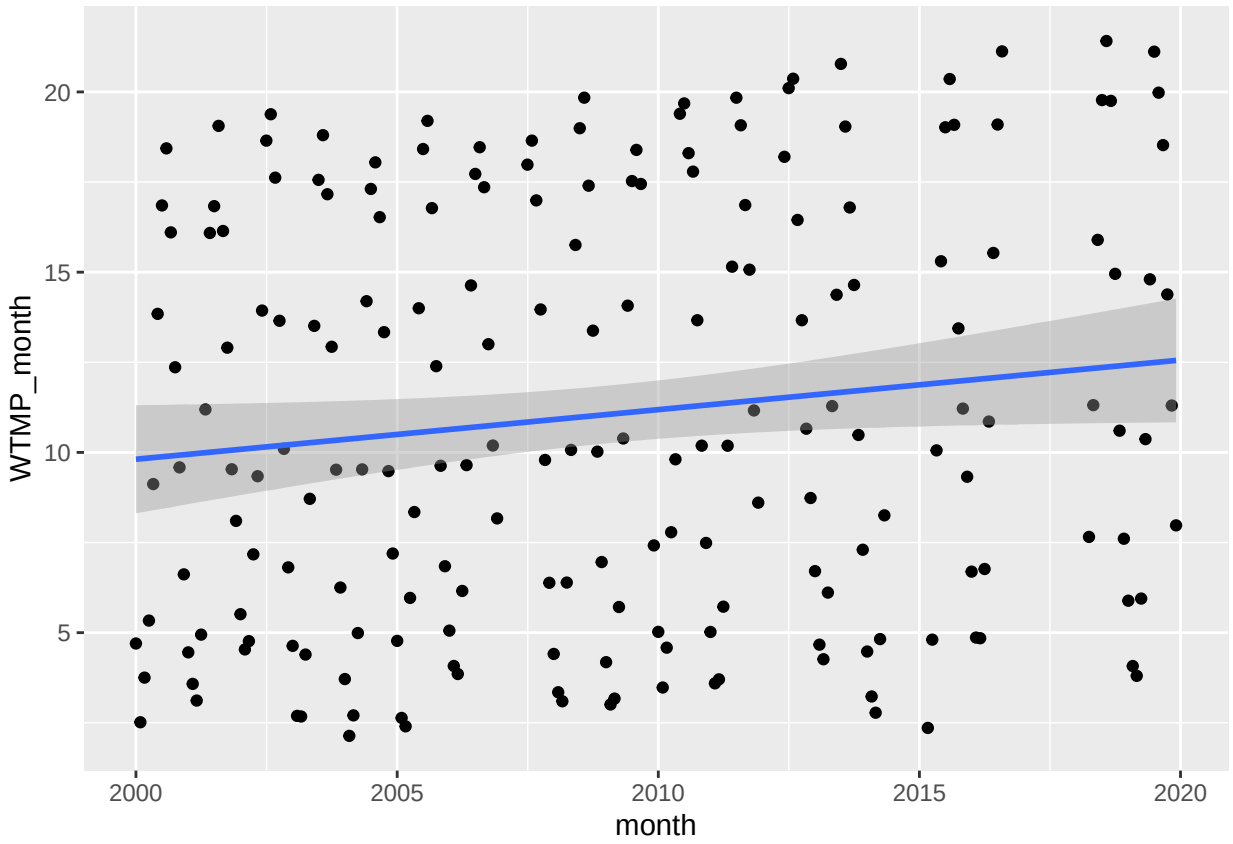
(2-1)In order to find the trend of average water temperature yearly, I use lm() function to find the linear regression and display the relationship by ggplot().

```
##
## Call:
## lm(formula = WTMP_year ~ year, data = fit4_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0694 -0.4929 -0.0456  0.4091  3.8868
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -140.45194  192.23495  -0.731    0.475
## year           0.07535    0.09568   0.787    0.442
##
## Residual standard error: 2.355 on 17 degrees of freedom
## Multiple R-squared:  0.0352, Adjusted R-squared:  -0.02156
## F-statistic: 0.6201 on 1 and 17 DF,  p-value: 0.4418
```



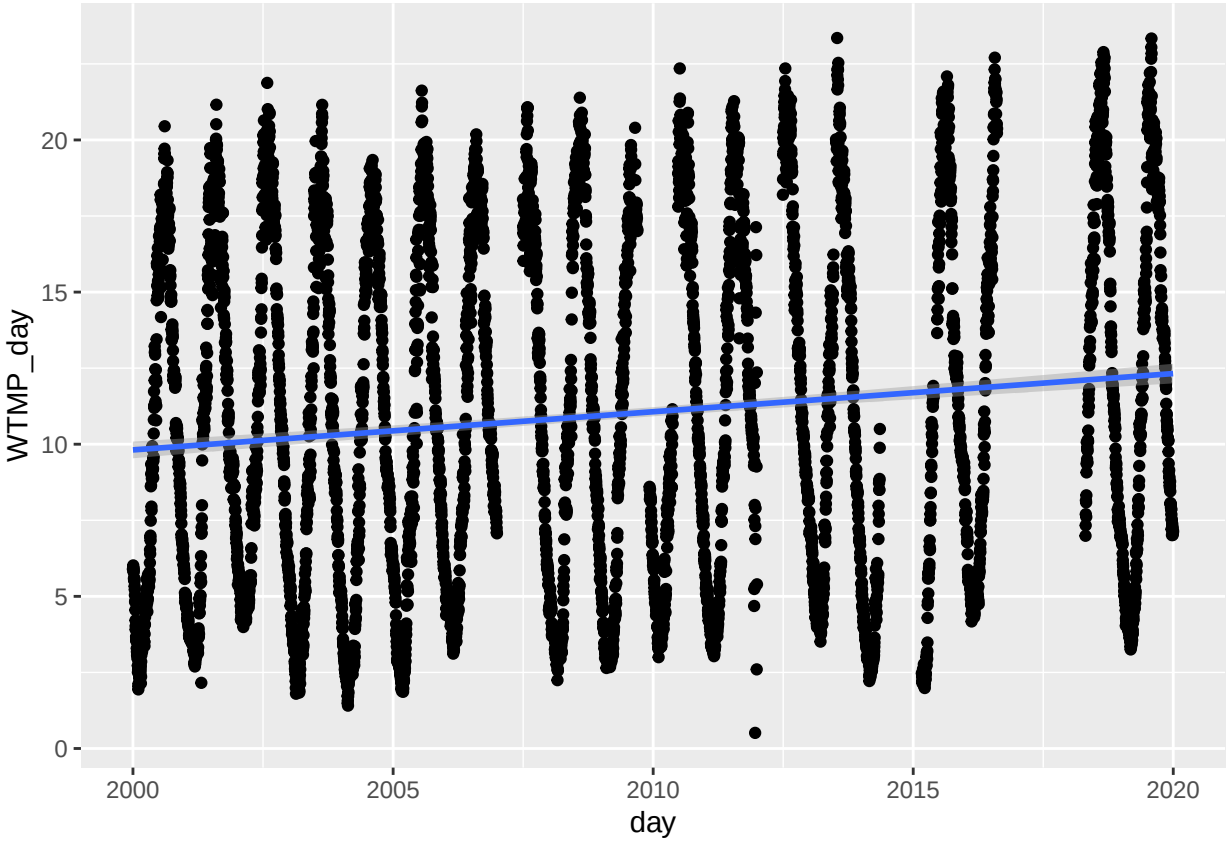
(2-2)In order to find the trend of average water temperature monthly, I use lm() function to find the linear regression and display the relationship by ggplot().

```
##
## Call:
## lm(formula = WTMP_month ~ month, data = fit5_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5393 -5.3314 -0.8873  6.0926  9.2153
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.6817453  2.8004107   2.029   0.0438 *
## month       0.0003768  0.0001942   1.940   0.0538 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.72 on 197 degrees of freedom
## Multiple R-squared:  0.01874,    Adjusted R-squared:  0.01376
## F-statistic: 3.763 on 1 and 197 DF,  p-value: 0.05383
```



(2-3)In order to find the trend of average water temperature daily, I use lm() function to find the linear regression and display the relationship by ggplot().

```
##
## Call:
## lm(formula = WTMP_day ~ day, data = fit6_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.7997  -5.3000  -0.7105   5.5968  11.8392
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.049e+00  5.194e-01  11.647   <2e-16 ***
## day         3.435e-04  3.614e-05   9.505   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.774 on 5824 degrees of freedom
## Multiple R-squared:  0.01528,    Adjusted R-squared:  0.01511
## F-statistic: 90.35 on 1 and 5824 DF,  p-value: < 2.2e-16
```



3 Conclusion

To sum up, the trend of average water temperature(air temperature) yearly, monthly, and daily increase slowly from 2000 to 2020, which may indicate that the global warming exists.

4 Curiosity

For this project, I want to create a polar diagram that can not only show the changes of temperature among years, but also show the changes of temperature among months. And I think I need to consider whether the seasonal change can influence the temperature.

5 Reference


```
## {
##   "title": "data.table: Extension of `data.frame`",
##   "version": "1.13.0",
##   "description": "<p>Fast aggregation of large data (e.g. 100GB in RAM), fast ordered joins, fast ad
##   "creators": [
##     {
##       "name": "Dowle, Matt"
##     },
##     {
##       "name": "Srinivasan, Arun"
##     },
##     {
##       "name": "Gorecki, Jan"
##     },
##     {
##       "name": "Chirico, Michael"
```

```
##     },
##     {
##       "name": "Stetsenko, Pasha"
##     },
##     {
##       "name": "Short, Tom"
##     },
##     {
##       "name": "Lianoglou, Steve"
##     },
##     {
##       "name": "Antonyan, Eduard"
##     },
##     {
##       "name": "Bonsch, Markus"
##     },
##     {
##       "name": "Parsonage, Hugh"
##     },
##     {
##       "name": "Ritchie, Scott"
##     },
##     {
##       "name": "Ren, Kun"
##     },
##     {
##       "name": "Tan, Xianying"
##     },
##     {
##       "name": "Saporta, Rick"
##     },
##     {
##       "name": "Seiskari, Otto"
##     },
##     {
##       "name": "Dong, Xianghui"
##     },
##     {
##       "name": "Lang, Michel"
##     },
##     {
##       "name": "Iwasaki, Watal"
##     },
##     {
##       "name": "Wenchel, Seth"
##     },
##     {
##       "name": "Broman, Karl"
##     },
##     {
##       "name": "Schmidt, Tobias"
##     },
##     {
##       "name": "Arenburg, David"
```

```
##     },
##     {
##       "name": "Smith, Ethan"
##     },
##     {
##       "name": "Cocquemas, Francois"
##     },
##     {
##       "name": "Gomez, Matthieu"
##     },
##     {
##       "name": "Chataignon, Philippe"
##     },
##     {
##       "name": "Blaser, Nello"
##     },
##     {
##       "name": "Selivanov, Dmitry"
##     },
##     {
##       "name": "Riabushenko, Andrey"
##     },
##     {
##       "name": "Lee, Cheng"
##     },
##     {
##       "name": "Groves, Declan"
##     },
##     {
##       "name": "Possenriede, Daniel"
##     },
##     {
##       "name": "Parages, Felipe"
##     },
##     {
##       "name": "Toth, Denes"
##     },
##     {
##       "name": "Yaramaz-David, Mus"
##     },
##     {
##       "name": "Perumal, Ayappan"
##     },
##     {
##       "name": "Sams, James"
##     },
##     {
##       "name": "Morgan, Martin"
##     },
##     {
##       "name": "Quinn, Michael"
##     },
##     {
##       "name": "@javrucebo"
```

```
##     },
##     {
##       "name": "@marc-outins"
##     },
##     {
##       "name": "Storey, Roy"
##     },
##     {
##       "name": "Saraswat, Manish"
##     },
##     {
##       "name": "Jacob, Morgan"
##     },
##     {
##       "name": "Schubmehl, Michael"
##     },
##     {
##       "name": "Vaughan, Davis"
##     },
##     {
##       "name": "Hocking, Toby"
##     },
##     {
##       "name": "Silvestri, Leonardo"
##     },
##     {
##       "name": "Barrett, Tyson"
##     },
##     {
##       "name": "Hester, Jim"
##     },
##     {
##       "name": "Damico, Anthony"
##     },
##     {
##       "name": "Freundt, Sebastian"
##     },
##     {
##       "name": "Simons, David"
##     },
##     {
##       "name": "Sales de Andrade, Elliott"
##     },
##     {
##       "name": "Miller, Cole"
##     },
##     {
##       "name": "Meldgaard, Jens Peder"
##     },
##     {
##       "name": "Tlapak, Vaclav"
##     },
##     {
##       "name": "Ushey, Kevin"
```

```
##     }
##   ]
## }
```

```
## {
##   "title": "lubridate: Make Dealing with Dates a Little Easier",
##   "version": "1.7.9",
##   "description": "<p>Functions to work with date-times and time-spans:\n    fast and user friendly pa
##   "creators": [
##     {
##       "name": "Spinu, Vitalie"
##     },
##     {
##       "name": "Grolemund, Garrett"
##     },
##     {
##       "name": "Wickham, Hadley"
##     },
##     {
##       "name": "Lyttle, Ian"
##     },
##     {
##       "name": "Costigan, Imanuel"
##     },
##     {
##       "name": "Law, Jason"
##     },
##     {
##       "name": "Mitarotonda, Doug"
##     },
##     {
##       "name": "Larmarange, Joseph"
##     },
##     {
##       "name": "Boiser, Jonathan"
##     },
##     {
##       "name": "Lee, Chel Hee"
##     }
##   ]
## }
```

```
## {
##   "title": "dplyr: A Grammar of Data Manipulation",
##   "version": "1.0.2",
##   "description": "<p>A fast, consistent tool for working with data frame\n    like objects, both in m
##   "creators": [
##     {
##       "name": "Wickham, Hadley"
##     },
##     {
##       "name": "François, Romain"
##     },
##     {
##       "name": "\n             Henry, Lionel"
```

```
##     },
##     {
##       "name": "Müller, Kirill"
##     },
##     {
##       "name": "RStudio"
##     }
##   ]
## }
```

```
## {
##   "title": "ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics",
##   "version": "3.3.2",
##   "description": "<p>A system for 'declaratively' creating graphics,\n    based on \"The Grammar of
##   "creators": [
##     {
##       "name": "Wickham, Hadley"
##     },
##     {
##       "name": "Chang, Winston"
##     },
##     {
##       "name": "Henry, Lionel"
##     },
##     {
##       "name": "Pedersen, Thomas Lin"
##     },
##     {
##       "name": "Takahashi, Kohske"
##     },
##     {
##       "name": "Wilke, Claus"
##     },
##     {
##       "name": "Woo, Kara"
##     },
##     {
##       "name": "Yutani, Hiroaki"
##     },
##     {
##       "name": "Dunnington, Dewey"
##     },
##     {
##       "name": "RStudio"
##     }
##   ]
## }
```

```
## {
##   "title": "citation: Software Citation Tools",
##   "version": "0.4.1",
##   "description": "<p>A collection of functions to extract citation information from 'R' packages and
##   "creators": [
##     {
##       "name": "Dietrich, Jan Philipp"
```

```
##     }
##   ]
## }
```