BuoyReport

Rong Li, id U73933267

Understanding of the question

We need to analysis 20-years of data collected by a single weather buoy in the NOAA National Data Buoy Center. The raw data concludes date, wind direction, wave height, air temperature (ATMP), water temperature(WTMP), dew point and so on. What we should do is mining the information from this data and find out whether there is an evidence of global warming.

My approach

First, I use R to clean and select the raw data. Since temperature is the most important signature of globle warming, I include the air temperature(ATMP) and water temprature(WTMP) at first. Then I consider which of the other columns related to temperature. Obviously, the seasonal change shouldn't be neglected. So the dates are include, too.
Secondly, I use R to analysis the data. I do some exploratory research to get familiar with this data， but there is no obvious trend. So I use the ARIMA model in time series to analysis the data. The function auto.arima() helps me to find appropriate peremeters rapidly. After I test the model, I can draw the conclusion.

How I organized the work

Part1: Processing the data. In the Raw_data.R, I put the code about processing the raw data in it.
Part2: Data analysis. The data_analysis.R includes the exploratory research and time series analysis.

Part1: Processing raw data

In this part, I access to the data of 1999-2018 online. Then I select ATMP, WTMP, date, and clean up the abnormal data. Finally I put the annual data together, turn them into a dataframe which I will use later.

The packages I use:

```
library(tidyverse)
```

```
## — Attaching packages ————————————————————— tidyverse 1.3.0 —
```

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

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

```r
library(stringr)
library(ggplot2)
library(lubridate)
```

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

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

```r
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
library(forecast)
library(urca)
```

Download the data of 1999-2018 online and save them as “mr1999-mr2018”. Change some files’ column names to keep consistent.

```
url_1 <- "http://www.ndbc.noaa.gov/view_text_file.php?filename=mlrf1h"
url_2 <- ".txt.gz&dir=data/historical/stdmet/"
years <- c(1999:2018)
urls <- str_c(url_1, years, url_2, sep = "")
filenames <- str_c("mr", years, sep = "")


# 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
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))),
  )

}
```

Screen out data at 13 o’clock.

```
i<-1999
repeat {
  assign(paste("mr",as.character(i),sep=""),get(paste("mr",as.character(i),sep=""))
[which(get(paste("mr",as.character(i),sep=""))$hh == 13), ])
  i=i+1
  if(i>2018)
  {break}
}
```

Put the data together into a dataframe called MRC and select “YYYY”, “MM”, “DD”, “ATMP”, “WTMP”.

```
mr1999$TIDE <- NA
n <- 20
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){
    MRC <- file
  }else{
    MRC <- rbind.data.frame(MRC, file)
  }

}

MRC<-MRC[c(1,2,3,13,14)]
```

Delete the abnormal data.

```
MRC$ATMP[which(MRC$ATMP>=100)]=NA
MRC$WTMP[which(MRC$WTMP>=100)]=NA
MRC=na.omit(MRC)
```

At the end of the part, we have obtained a dataframe called MRC which we can use later.

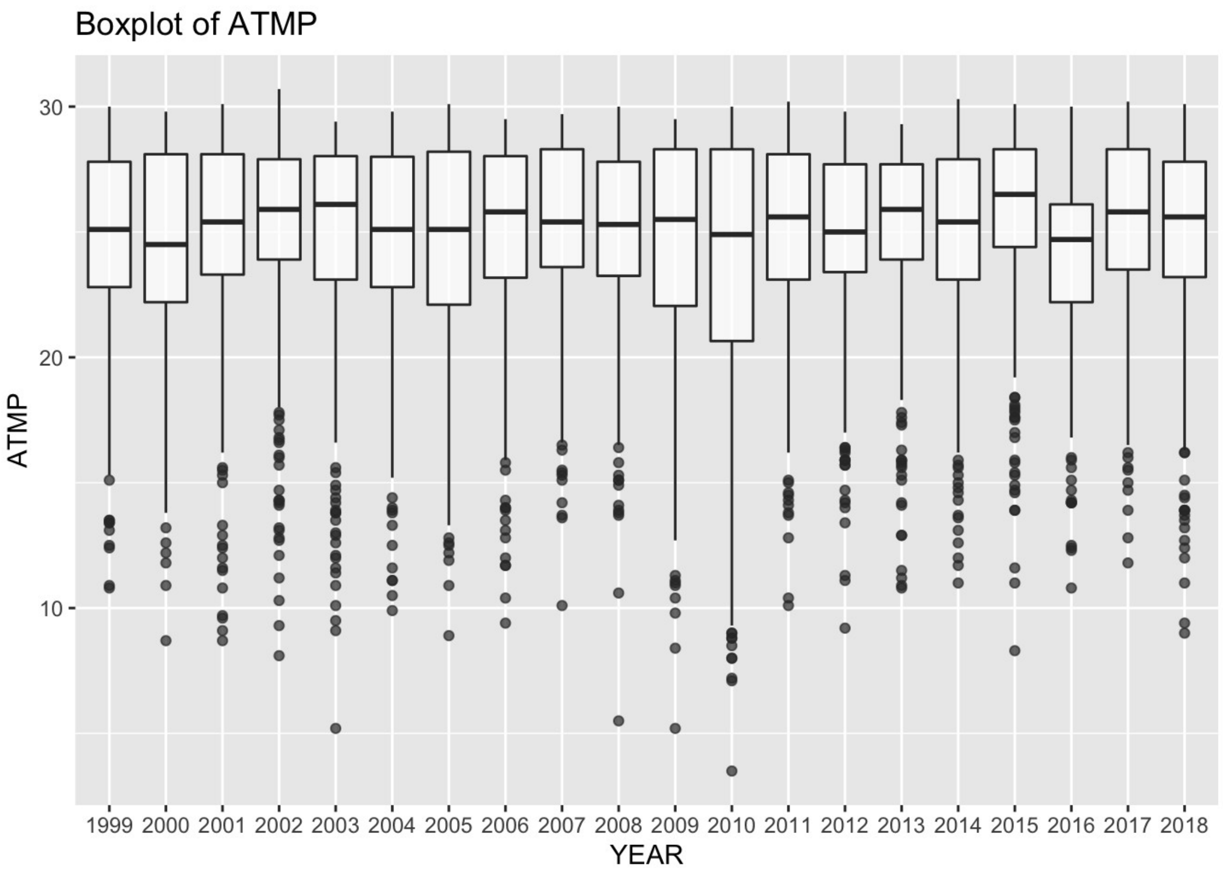
Part2: Data analysis

In this part, I use MRC to further my research. I do some exploratory research such as ploting the boxplot of ATMP versus year and the boxplot of WTMP versus year to get a general understanding of the data. I also go deeper in it and change the data into time series. After testing the ARIMA model I use, I draw a conclusion that there is no obvious evidence of global warming.

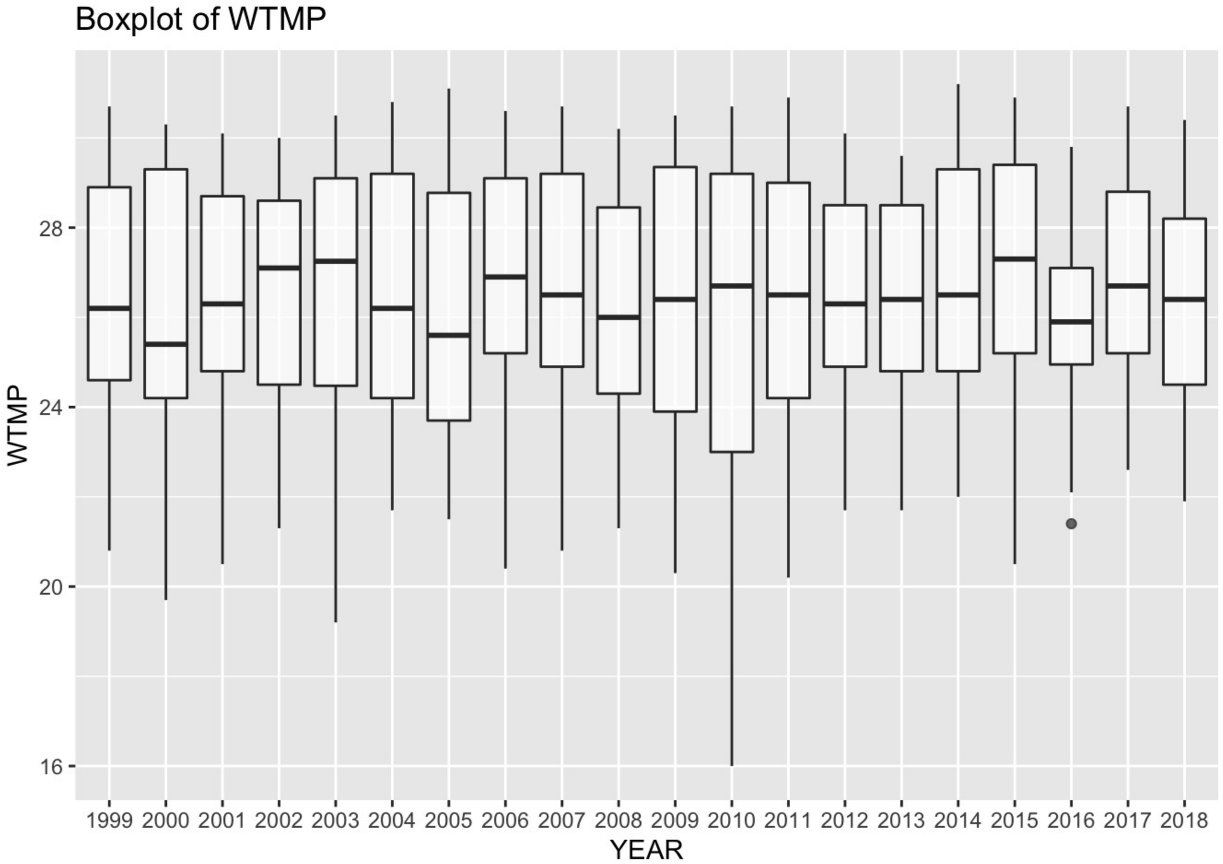
Boxplot

```
MRC$factor <- as.factor(MRC$YYYY)

#Plot the boxplot of ATMP versus Year
ggplot(MRC, aes(x = factor, y = ATMP)) +
  geom_boxplot(alpha=0.7) +
  scale_y_continuous(name = "ATMP")+
  scale_x_discrete(name = "YEAR") +
  ggtitle("Boxplot of ATMP")
```



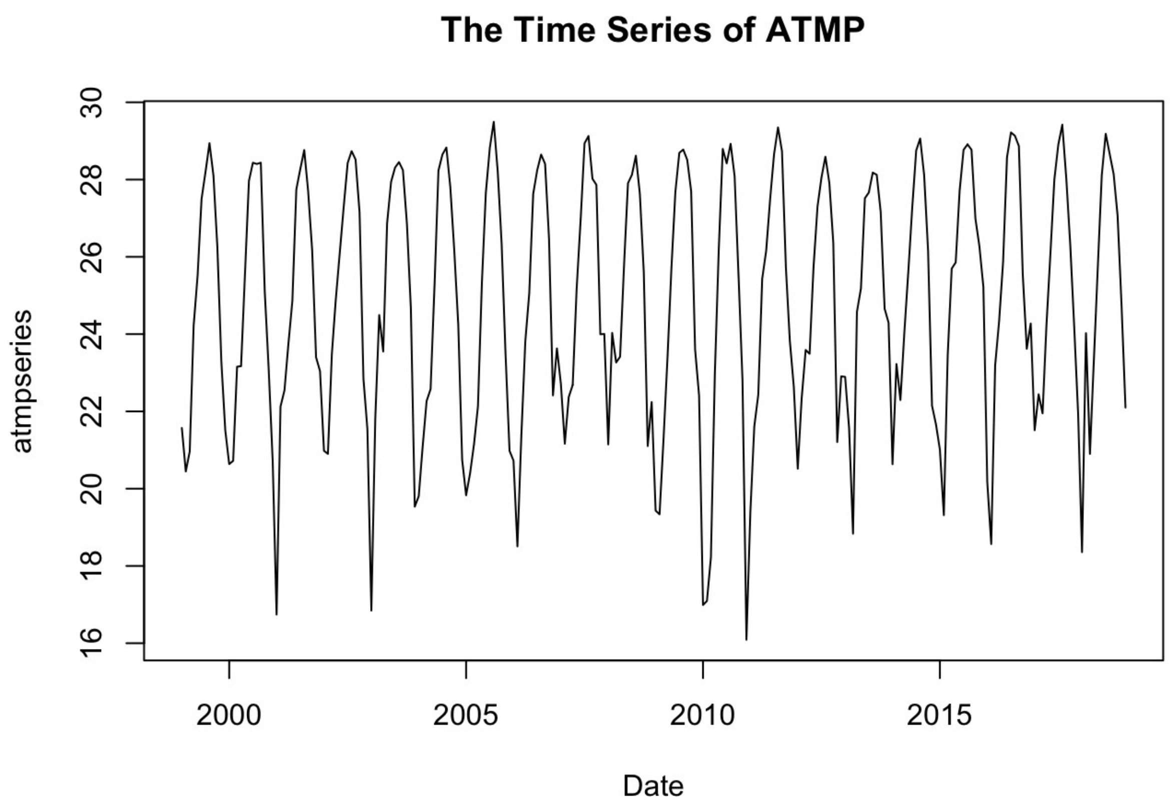

```
#Plot the boxplot of WTMP versus Year
ggplot(MRC, aes(x = factor, y = WTMP)) +
  geom_boxplot(alpha=0.7) +
  scale_y_continuous(name = "WTMP")+
  scale_x_discrete(name = "YEAR") +
  ggtitle("Boxplot of WTMP")
```



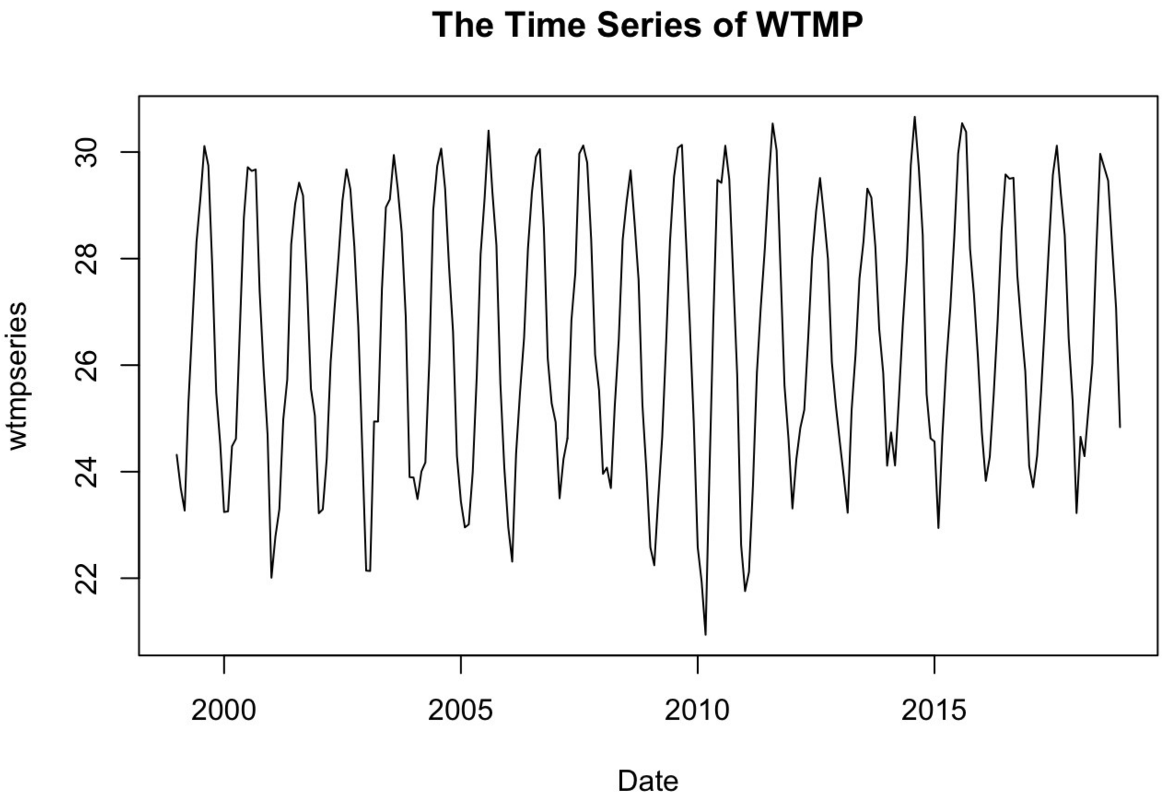
From the picture above, we can’t draw conclusions because there is no obvious trend. So we need to go deeper.

Calculate the monthly average of ATMP&WTMP every year and form time series

```
ameans <- array()
wmeans <- array()
z <- 1
for (i in 1999:2018){
  for (j in 1:12){
    ameans[z] <- mean(subset(MRC, YYYY == i)$ATMP[which(subset(MRC, YYYY == i)$MM =
= j)])
    wmeans[z] <- mean(subset(MRC, YYYY == i)$WTMP[which(subset(MRC, YYYY == i)$MM =
= j)])
    z <- z+1
    if (z > 5){
      if (is.na(ameans[z-2])){
        ameans[z-2] <- (ameans[z-3]+ameans[z-1])/2
      }
      if (is.na(wmeans[z-2])){
        wmeans[z-2] <- (wmeans[z-3]+wmeans[z-1])/2
      }
    }
  }
}
atmpseries <- ts(ameans, frequency=12, start=c(1999,1))
wtmpseries <- ts(wmeans, frequency=12, start=c(1999,1))
plot(atmpseries, main = "The Time Series of ATMP", xlab = "Date")
```



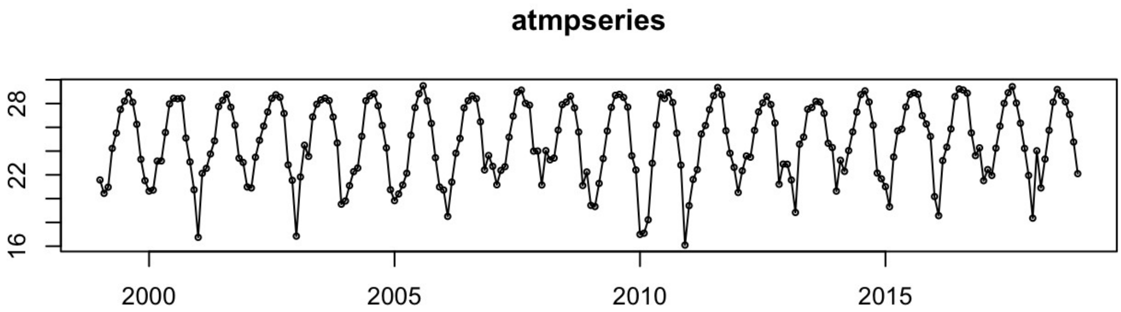
```
plot(wtmpseries, main = "The Time Series of WTMP", xlab = "Date")
```

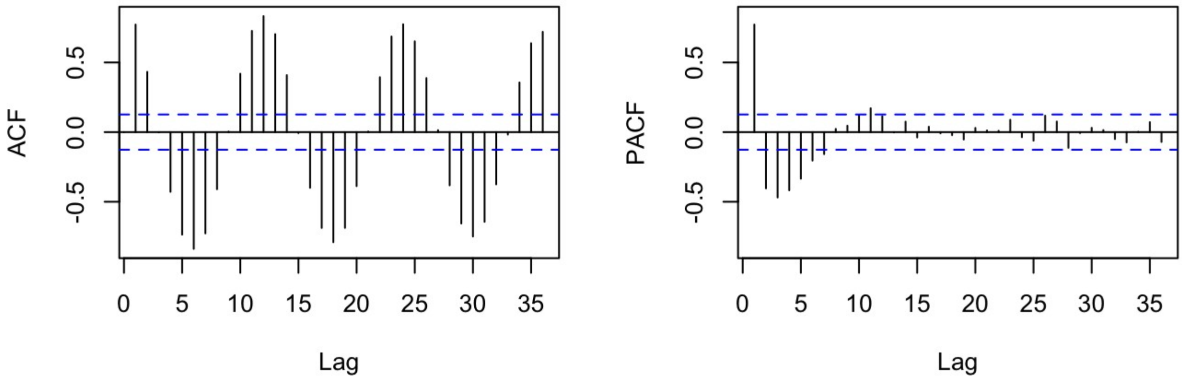


The pictures indicate that there should be seasonal trend in the temperature. Since the two pictures are very alike which means the trends of ATMP and WTMP are very similar, I only need to analysis one of them. In this case, I choose to analysis the ATMP.

####Display the time series of ATMP.

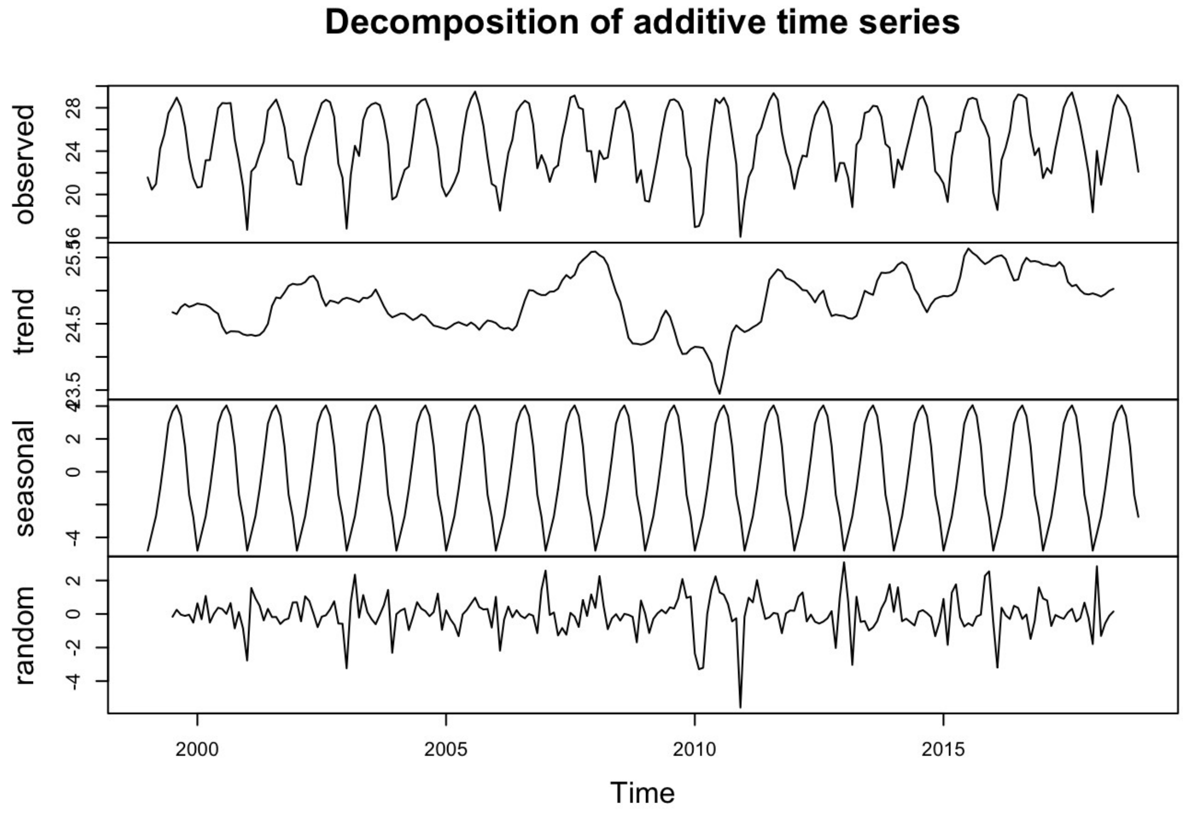
```
tsdisplay(atmpseries)
```





Decompose the time series

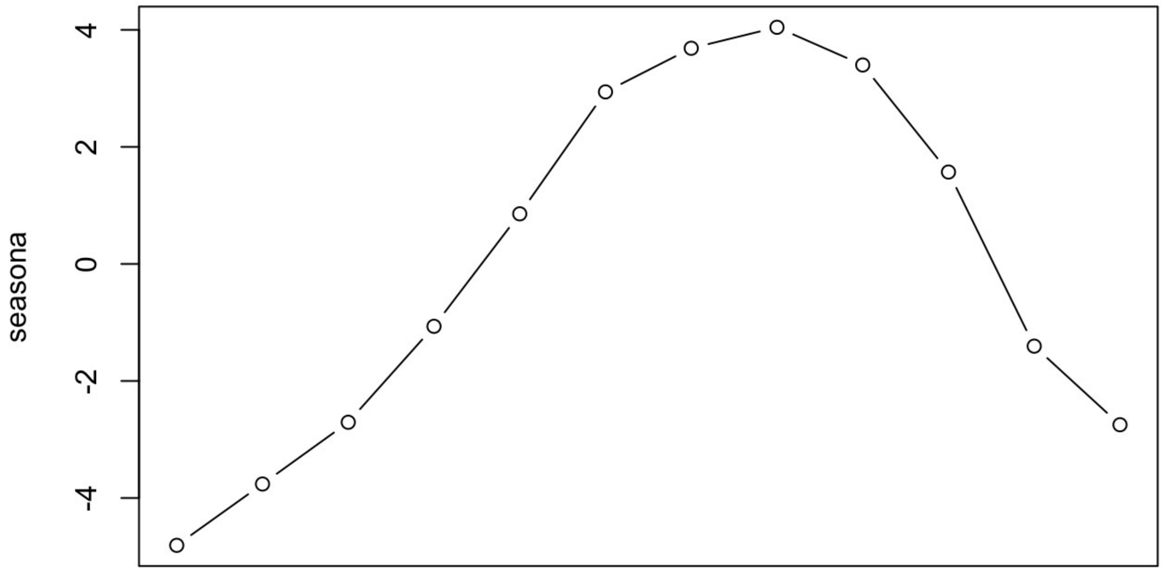
```
dca <- decompose(atmpseries)
plot(dca)
```



In the graph, the seasonal effection is apperant and the trend is unclear.

Find out the seasonal trend of the data.

```
seasona<-dca$figure
plot(seasona,type = "b",xaxt="n",xlab = "")
```



Use function auto.arima() to get the peremeter of ARIMA model.

```
a<-auto.arima(atmpseries)
print(a)
```

```
## Series: atmpseries
## ARIMA(0,0,1)(2,1,0)[12] with drift
##
## Coefficients:
##          ma1     sar1     sar2   drift
##       0.2279  -0.5206  -0.2393  0.0019
## s.e.  0.0636   0.0657   0.0667  0.0055
##
## sigma^2 estimated as 1.941:  log likelihood=-398.99
## AIC=807.99   AICc=808.26   BIC=825.14
```

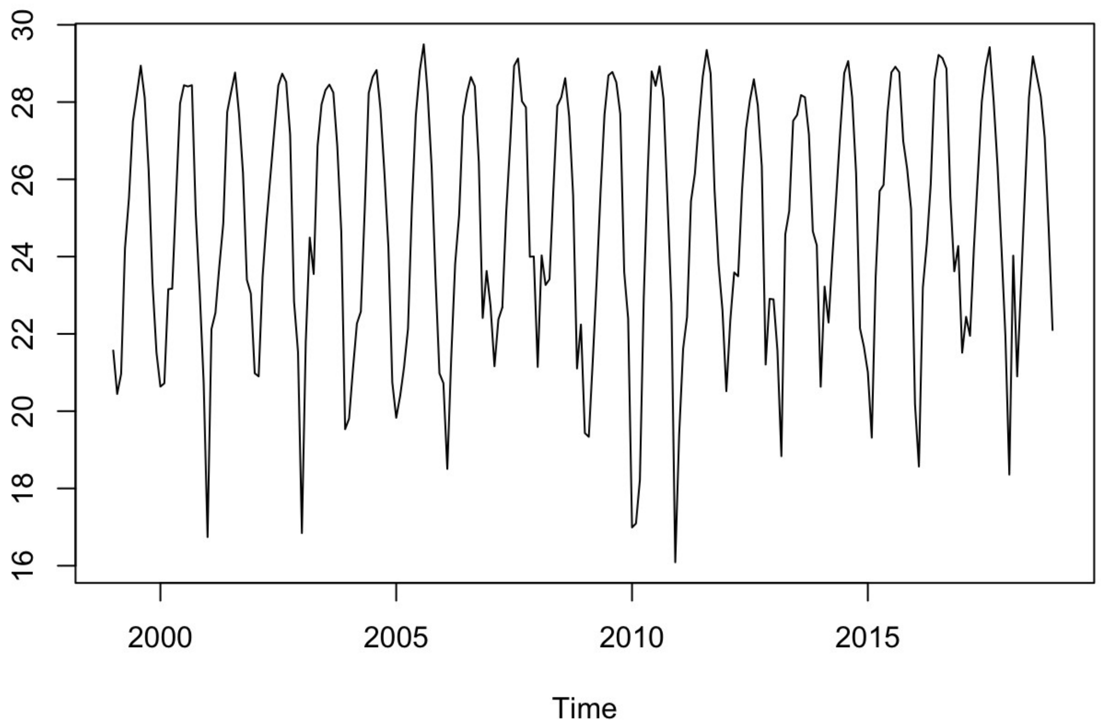
The model is ARIMA(0, 0, 1)x(2, 1, 0).

Fit the model and plot the predicted.

```
fit<-arima(atmpseries,order = c(0, 0, 1),seasonal = list(order=c(2, 1, 0),period=1
2))
print(fit)
```

```
##
## Call:
## arima(x = atmpseries, order = c(0, 0, 1), seasonal = list(order = c(2, 1, 0),
##     period = 12))
##
## Coefficients:
##          ma1     sar1     sar2
##       0.2286  -0.5200  -0.2386
## s.e.  0.0635   0.0657   0.0667
##
## sigma^2 estimated as 1.908:  log likelihood = -399.05,  aic = 806.11
```

```
fore=predict(fit)
ts.plot(atmpseries, fore$pred, col=c(1,2,4,4), lty=c(1,1,2,2))
```



Test the model

```
ur.df(atmpseries)
```

```
##
## ###############################################################
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## ###############################################################
##
## The value of the test statistic is: -0.8381
```

The value of the test statistic is -0.8381 < 5%, which refers the series is stable.
We can draw the conclusion that there is no obvious evidence of global warming.

My conlusions

Through exploratory research and time series analysis, there is no conclusive evidence of global warming in the data.

My reference

[data scource]http://www.ndbc.noaa.gov/view_text_file.php?filename=mlrf1h1999.txt.gz&dir=data/historical/stdmet/ (http://www.ndbc.noaa.gov/view_text_file.php?filename=mlrf1h1999.txt.gz&dir=data/historical/stdmet/) [Methods in time series] https://blog.csdn.net/jiabiao1602/article/details/43153139 (https://blog.csdn.net/jiabiao1602/article/details/43153139) [R packages] https://cran.r-project.org/web/packages/citation/index.html (https://cran.r-project.org/web/packages/citation/index.html)