Buoy_project

JingwenXu

2020/9/25

####Report

##Research Objective: Is there evidence of global warming in the data collected by a single weather buoy in the NOAA National Data Buoy Center? The Boston Buoy Data Analysis project requires you to complete a project in its entirety from acquiring data to presenting results. As you work, keep the goals and vision for the completed work in perspective.

##My understanding of the question:

In my opinion, to find the evidence of global warming, we need to study the trend of certain temperature characteristics along with the year. Firstly, we can simply plot the annual averages of ATMP and WTMP which reflect the central tendency of data. Secondly, I think it is reasonable to analyze the annual minimum ATMP. We usually regard the abnormal higher temperature than before in cold winter as the sign of global warming. So if the annual minimum of ATMP is increasing, we may find the evidence of global warming. For the last temperature characteristic, I choose the difference between annual average of ATMP and WTMP as my index. Because according to the meteorological research, one of the phenomenons of global warming is the getting smaller difference between ATMP and WTMP which can result in environmental damage.

##My approach and how I organized my work:

I use R programming to complete my whole research from importing the data to plotting the trend of annual temperature characteristics. And I organize my work as following steps(containing the codes and plots):

#(a)Import the data from NOAA and clean the data.

```
library(stringr)
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 forcats 0.5.0
## v readr   1.3.1

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

```
library(lubridate)
```

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

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

```r
### make URLs

url1 <- "http://www.ndbc.noaa.gov/view_text_file.php?filename=mlrf1h"
url2 <- ".txt.gz&dir=data/historical/stdmet/"

years <- c(1987:2016)

urls <- str_c(url1, years, url2, sep = "")

filenames <- str_c("mr", years, sep = "")

###  Read the data from the website

N <- length(urls)

for (i in 1:N){
suppressMessages(assign(filenames[i], read.table(urls[i], header = TRUE,fill = T)))
file <- get(filenames[i])
}
# put '19' in front of 2 digit years so that all the year format is "YYYY"
for (i in 1:12){
file <- get(filenames[i])
file$YY <- file$YY +1900
assign(filenames[i],file)
}
# check that all columns are included
# remove the last column containing large numbers of "NA"
for (i in 14:18){
file <- get(filenames[i])
assign(filenames[i],file[,1:16])
}
#remove the "mm" column and the last excess column
for (i in 19:30){
file <- get(filenames[i])
assign(filenames[i],file[,c(1:4,6:17)])
}
#Combine all the data frame
for (i in 1:30){
file <- get(filenames[i])
colnames(file) <- c("YYYY", "MM", "DD", "hh", "WD", "WSPD", "GST", "WVHT", "DPD", "APD", "MWD", "BAR",

if(i==1){
MR <- file
}
else{
MR <- rbind.data.frame(MR, file)
}
}
MR <- MR[,c(1:7,12:14)]
summary(MR)
```

```
##       YYYY            MM               DD              hh
##  Min.   :1987   Min.   : 1.000   Min.   : 1.00   Min.   : 0.00
##  1st Qu.:1995   1st Qu.: 4.000   1st Qu.: 8.00   1st Qu.: 6.00
```

```
##  Median :2002   Median : 7.000   Median :16.00   Median :12.00
##  Mean   :2002   Mean   : 6.533   Mean   :15.74   Mean   :11.51
##  3rd Qu.:2009   3rd Qu.:10.000   3rd Qu.:23.00   3rd Qu.:18.00
##  Max.   :2016   Max.   :12.000   Max.   :31.00   Max.   :23.00
##        WD              WSPD             GST              BAR
##  Min.   :  0.0   Min.   : 0.000   Min.   : 0.000   Min.   : 982.3
##  1st Qu.: 81.0   1st Qu.: 3.800   1st Qu.: 4.300   1st Qu.:1014.9
##  Median :118.0   Median : 5.800   Median : 6.500   Median :1017.1
##  Mean   :148.8   Mean   : 6.944   Mean   : 7.955   Mean   :1556.4
##  3rd Qu.:187.0   3rd Qu.: 8.100   3rd Qu.: 9.000   3rd Qu.:1019.5
##  Max.   :999.0   Max.   :99.000   Max.   :99.000   Max.   :9999.0
##       ATMP            WTMP
##  Min.   :  1.8   Min.   : 15.90
##  1st Qu.: 23.5   1st Qu.: 24.70
##  Median : 25.6   Median : 26.60
##  Mean   : 31.0   Mean   : 35.75
##  3rd Qu.: 28.0   3rd Qu.: 29.10
##  Max.   :999.0   Max.   :999.00
```

```
#We find that there are some extreme values 999 in ATMP and WTMP which should be removed
MR <- filter(MR,MR$ATMP<100&MR$WTMP<100)
MR <- MR %>%mutate(DATETIME = make_datetime(YYYY,MM,DD,hh))
MR<-MR[,5:11]
MR<-MR[,c(7,1:6)]
```

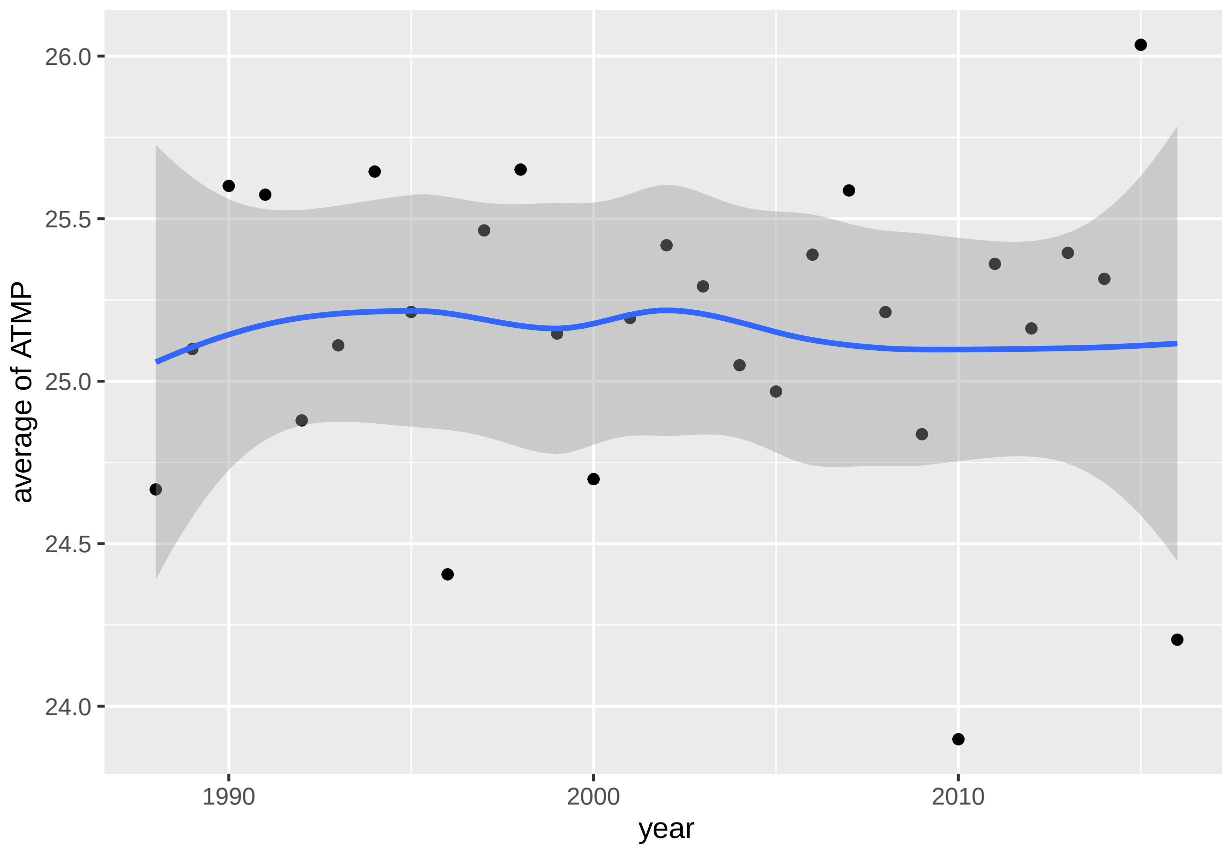
#(b)Use lubridate to transform the date-time data into posix numbers and plot the annual temperature characteristics.

```
library(lubridate)
time<-ymd_hms(MR$DATETIME)
y1<-c()
for(i in 1988:2016){
y1[i-1987]<-mean(subset(MR$ATMP,year(MR$DATETIME)==i))
}
y1
```

```
##  [1] 24.66698 25.09872 25.60080 25.57377 24.87931 25.11022 25.64485 25.21294
##  [9] 24.40576 25.46380 25.65094 25.14648 24.69871 25.19432 25.41811 25.29163
## [17] 25.04914 24.96824 25.38914 25.58666 25.21260 24.83668 23.89854 25.36097
## [25] 25.16222 25.39490 25.31492 26.03482 24.20475
```

```
x<-c(1988:2016)
data1<-data.frame(x,y1)
ggplot(data=data1,mapping=aes(x,y1))+
geom_point()+geom_smooth()+labs(x="year",y="average of ATMP")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



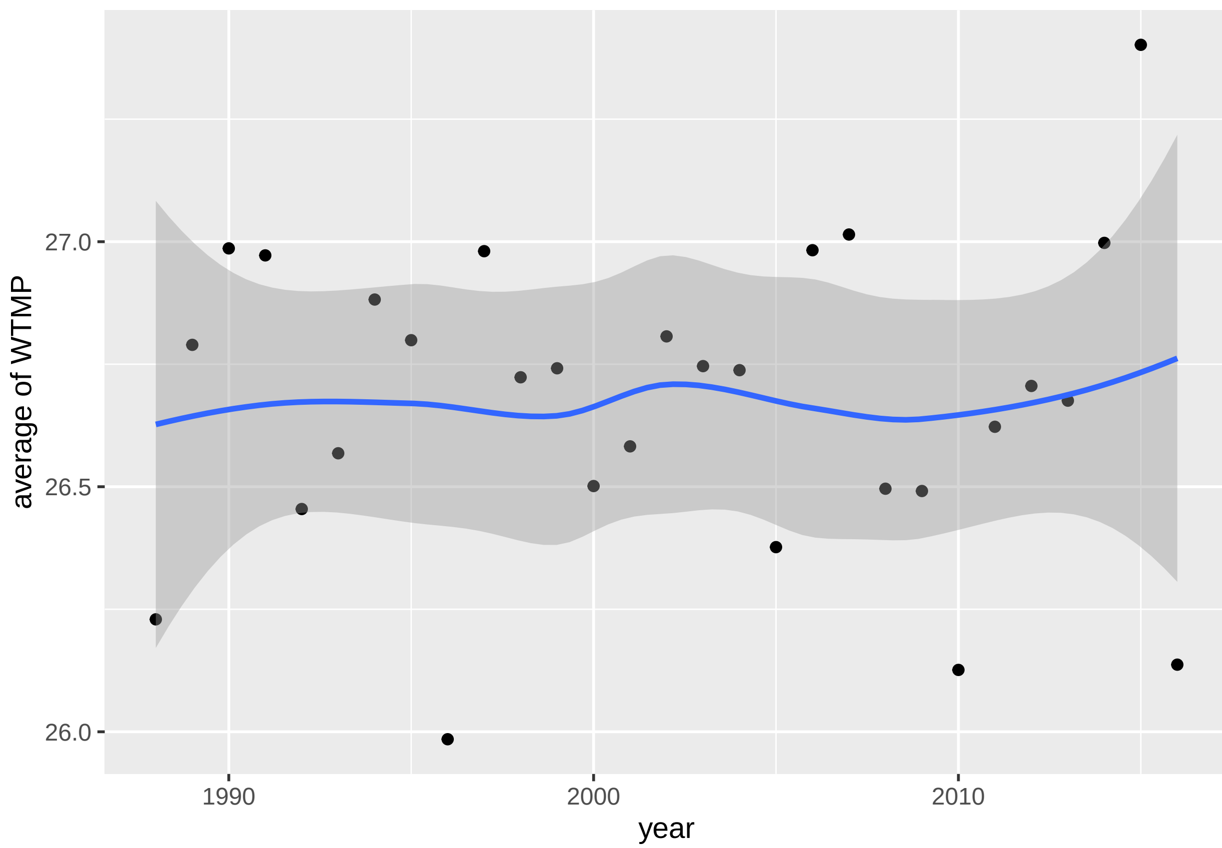
###In above plot, the annual averages of ATMP fluctuate randomly around 25-25.25 celsius.

```
x<-c(1988:2016)
y2<-c()
for(i in 1988:2016){
y2[i-1987]<-mean(subset(MR$WTMP,year(MR$DATETIME)==i))
}
y2
```

```
##  [1] 26.22939 26.78952 26.98638 26.97205 26.45451 26.56829 26.88198 26.79903
##  [9] 25.98483 26.98064 26.72332 26.74169 26.50143 26.58240 26.80680 26.74612
## [17] 26.73784 26.37680 26.98260 27.01463 26.49599 26.49128 26.12623 26.62236
## [25] 26.70564 26.67567 26.99757 27.40170 26.13697
```

```
data2<-data.frame(x,y2)
ggplot(data=data2,mapping=aes(x,y2))+
geom_point()+geom_smooth()+labs(x="year",y="average of WTMP")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



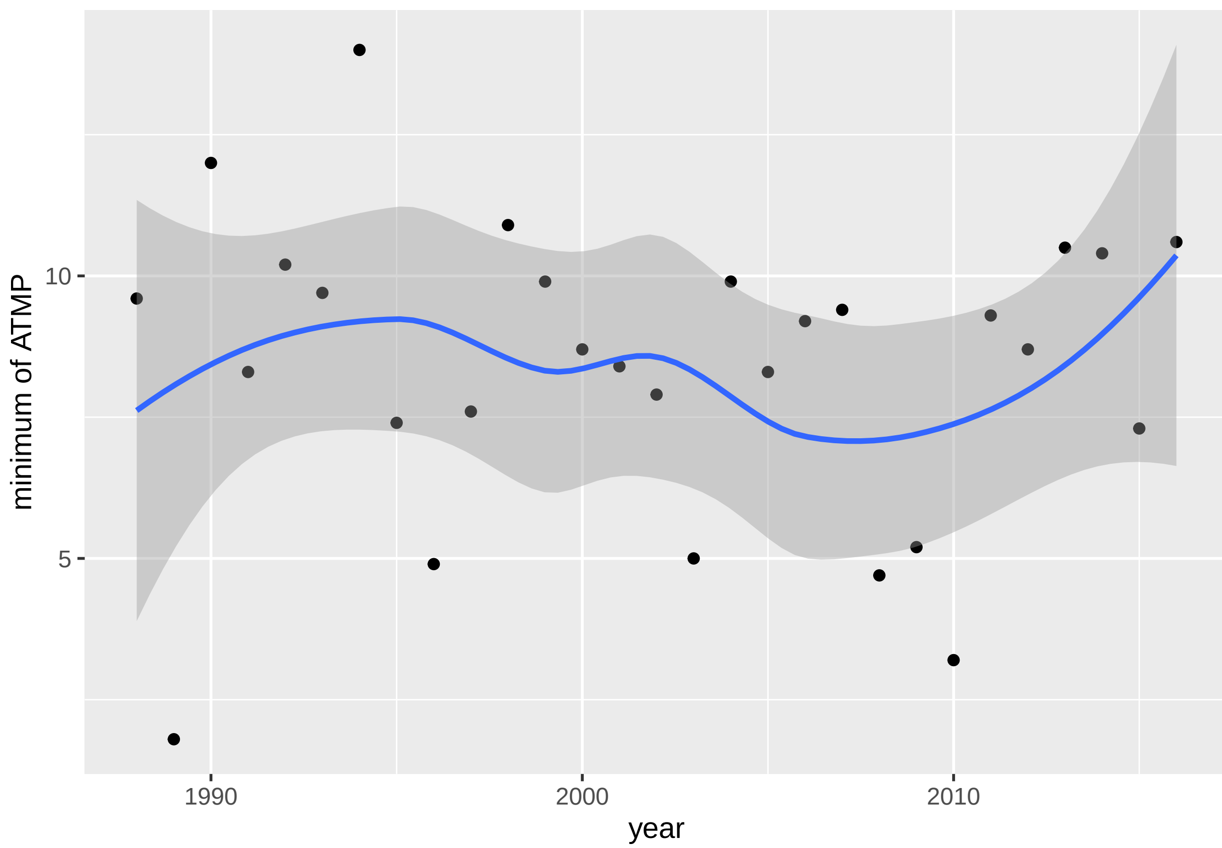
###In above plot, the annual averages of WTMP flctuate around 26.5-26.75 celsius.

```
x<-c(1988:2016)
y3<-c()
for(i in 1988:2016){
y3[i-1987]<-min(subset(MR$ATMP,year(MR$DATETIME)==i))
}
y3
```

```
##  [1]  9.6  1.8 12.0  8.3 10.2  9.7 14.0  7.4  4.9  7.6 10.9  9.9  8.7  8.4  7.9
## [16]  5.0  9.9  8.3  9.2  9.4  4.7  5.2  3.2  9.3  8.7 10.5 10.4  7.3 10.6
```

```
data3<-data.frame(x,y3)
ggplot(data=data3,mapping=aes(x,y3))+
geom_point()+geom_smooth()+labs(x="year",y="minimum of ATMP")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



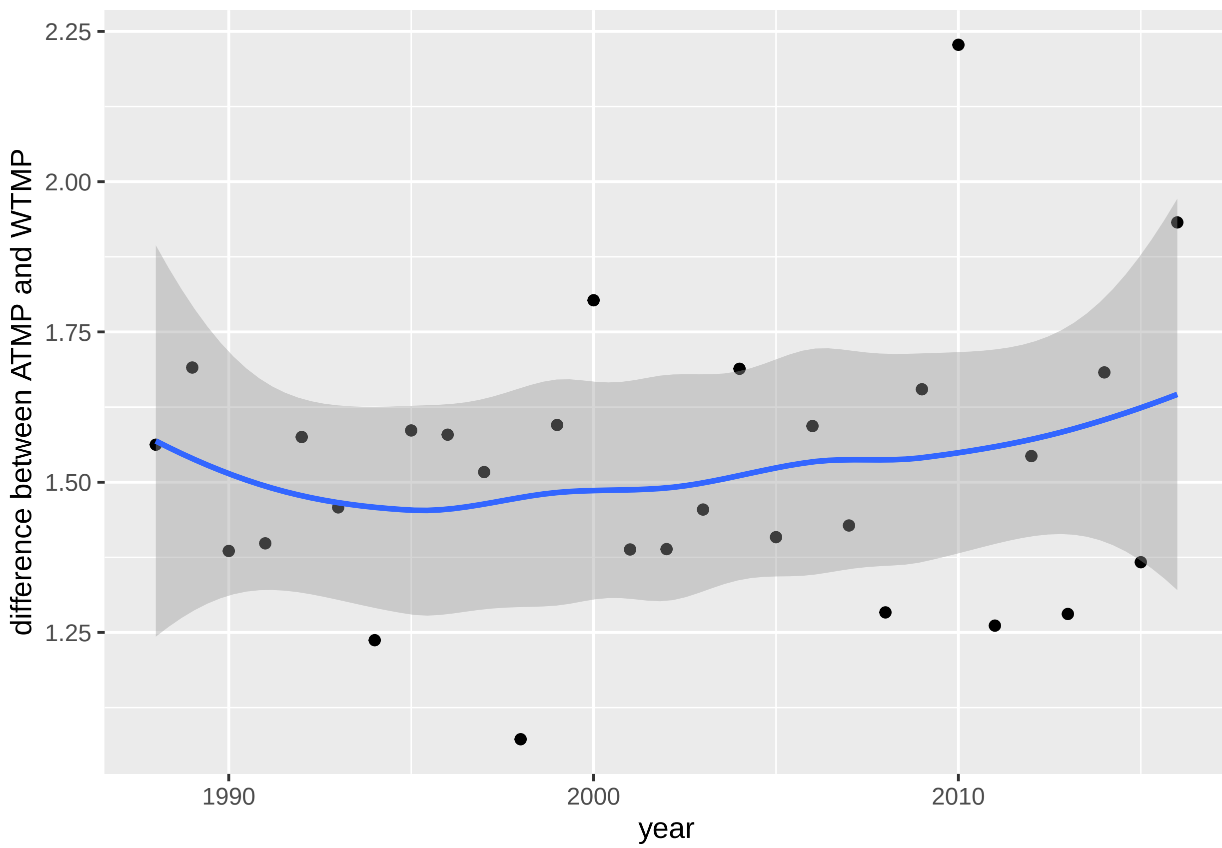
###In above plot, the annual minimum of ATMP has the increasing tendency since 2010.

```
x<-c(1988:2016)
y4<-c()
for(i in 1988:2016){
y4[i-1987]<-y2[i-1987]-y1[i-1987]
}
y4
```

```
##  [1] 1.562408 1.690800 1.385575 1.398285 1.575206 1.458073 1.237131 1.586095
##  [9] 1.579067 1.516843 1.072383 1.595207 1.802716 1.388078 1.388688 1.454496
## [17] 1.688705 1.408557 1.593460 1.427967 1.283392 1.654601 2.227691 1.261390
## [25] 1.543420 1.280769 1.682650 1.366888 1.932220
```

```
data4<-data.frame(x,y4)
ggplot(data=data4,mapping=aes(x,y4))+
geom_point()+geom_smooth()+labs(x="year",y="difference between ATMP and WTMP")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



###In above plot, although there is an slightly increasing difference between annual averages of ATMP and WTMP, the values of 2011,2013 and 2015 are much lower than the tendency line.

##My conclusions:

We can acknowledge little about the evidence from the 1st and 2nd figures which are both regular wave. While due to the increasing annual minimum of ATMP representing the coming of "warm winter", I think figure "year-minimum of ATMP" can be as the evidence of global warming. Besides, maybe because of the environmental policy and measures applied by governments, there is an slightly increasing difference between annual averages of ATMP and WTMP. But the low values of 2011, 2013 and 2015 should be a warning of global warming.

##Possible improvements about the research

If we would like to improve the accuracy and credibility of this research, we can combine more factors about the climate and consider the interactions between these factors. Or we can make plots using more samples such as the monthly or daily observations instead of yearly observations.