Boston Buoy Data Analysis Report

Danping Liu

9/24/2020

Introduction

The objective for this project is to find the evidence for global warming. From NASA's *Climate change evidence: How do we know?* page, it is said that rising average temperature, warming oceans and extreme climate are significant factors that show global warming. So to prove that global warming is true, we are going to analyze the change in air temperature and water temperature.

Firstly, we import the data we are going use to R. Secondly, we build linear regression models for to check how the monthly temperature change.Thirdly, we build linear regression models to check how the annual average temperature change. Lastly, we make conclusion based on the two aspects above.

Import data

The data we use is NOAA's Boston's air and water temperature from 1987 to 2016. There are 246245 observations in total. The website's data is grouped by year. Some years' data has different number of columns and different format of year. For the year column, we make all year data into four digits. Since we only need year, month, day, hour, air temperature and water temperature here, we can remove the columns that some data does not have.

```r
###Make URLs and read Data
##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=paste0(url1,years,url2)
Dnames=paste0('D',years)
##Read the data from the URLs
for(i in years) assign(Dnames[i-years[1]+1],read_table2(urls[i-years[1]+1]))
coln=colnames(get(Dnames[1]))

###Combine the data into a frame
for(i in years){
  D=get(Dnames[i-years[1]+1])
  ##From Y2000 to Y2016, delete an additional variable of 'TIDE'
  if(i %in% 2000:2016){D=select(D,-TIDE)}
  ##From Y2005 to Y2016, delete an additional variable of 'mm'
  if(i %in% 2005:2016){D=select(D,-mm)}
  ##From Y2007 to Y2016, delete first row of units
  if(i %in% 2007:2016){D=D[-1,]}
  ##Check and unify col names and set data type as 'numeric'
  if(ncol(D)==length(coln)){colnames(D)=coln}
  D=sapply(D, as.numeric)
  ##From Y1987 to Y1999, transfer the Year from 'XX' to '19XX'
  D[,1][D[,1]<100]=D[,1][D[,1]<100]+1900
```

```r
  ##Create and combine to form final data set Buoy
  if(i==years[1]){Buoy=D}
  else{Buoy=rbind.data.frame(Buoy,D)}
}

###Save data
save(Buoy,file='Buoy.Rdata')
```

For better comparison, we remove the NA values and use the temperature at 12pm everyday. As a result, we have 7069 observations for analyzing. The year, month, day, hour, air temperature and water temperature are stored in the data frame as string. For later analyzing, we change the data types for air and water temperature to double. We create a new TIME column combining the year, month, day and hour columns into POXISct form. However, POXISct data cannot join the regression fitting, so we unclass it to get a double type variable representing the number of seconds from 00:00:00 January 1, 1970.

```r
load("Buoy.Rdata")
# Remove NA values
`%notin%` <- Negate(`%in%`)
Buoysub <- filter(Buoy, ATMP%notin%c(999,99,"999","99","999.0","99.0") &
                  WTMP%notin%c(999,99,"999","99","999.0","99.0") & hh=="12")
# Convert string values to double
Buoysub$ATMP <- as.double(Buoysub$ATMP)
Buoysub$WTMP <- as.double(Buoysub$WTMP)
# Convert YY, MM, DD and hh to POSIXct and then to seconds from 1970/1/1
buoyTimeStr <- paste(Buoysub$YY, Buoysub$MM, Buoysub$DD, Buoysub$hh, sep="-")
Buoysub$TIME <- ymd_h(buoyTimeStr)
Buoysub$DATE <- unclass(Buoysub$TIME)
Buoysub$YY <- as.double(Buoysub$YY)
```

Monthly temperature change

Since the temperature varies during different months, we decide to build linear regression models for each month respectively and for annual average. For better understanding, we make a table presenting how monthly air temperature and water temperature is changing from year to year. We also make plots of temperature versus time for every month.

```r
numMonths <- levels(factor(Buoysub$MM))
nameMonths <- c("January", "February", "March", "April", "May", "June", "July", "August",
                "September", "October", "November", "December")
fitA_list <- list()
fitW_list <- list()
coefAPerYear <- list()
coefWPerYear <- list()
for(i in 1:12){
  fitA_list[[i]] <- lm(ATMP~DATE, data=Buoysub[Buoysub$MM==numMonths[i],])
  fitW_list[[i]] <- lm(WTMP~DATE, data=Buoysub[Buoysub$MM==numMonths[i],])
  coefAPerYear <- 60*60*24*365*coef(fitA_list[[i]])[2]
  coefWPerYear <- 60*60*24*365*coef(fitW_list[[i]])[2]
}
Buoymonth <- data.frame(nameMonths, coefAPerYear, coefWPerYear)
```

```
## Warning in data.frame(nameMonths, coefAPerYear, coefWPerYear): row names were
## found from a short variable and have been discarded
```

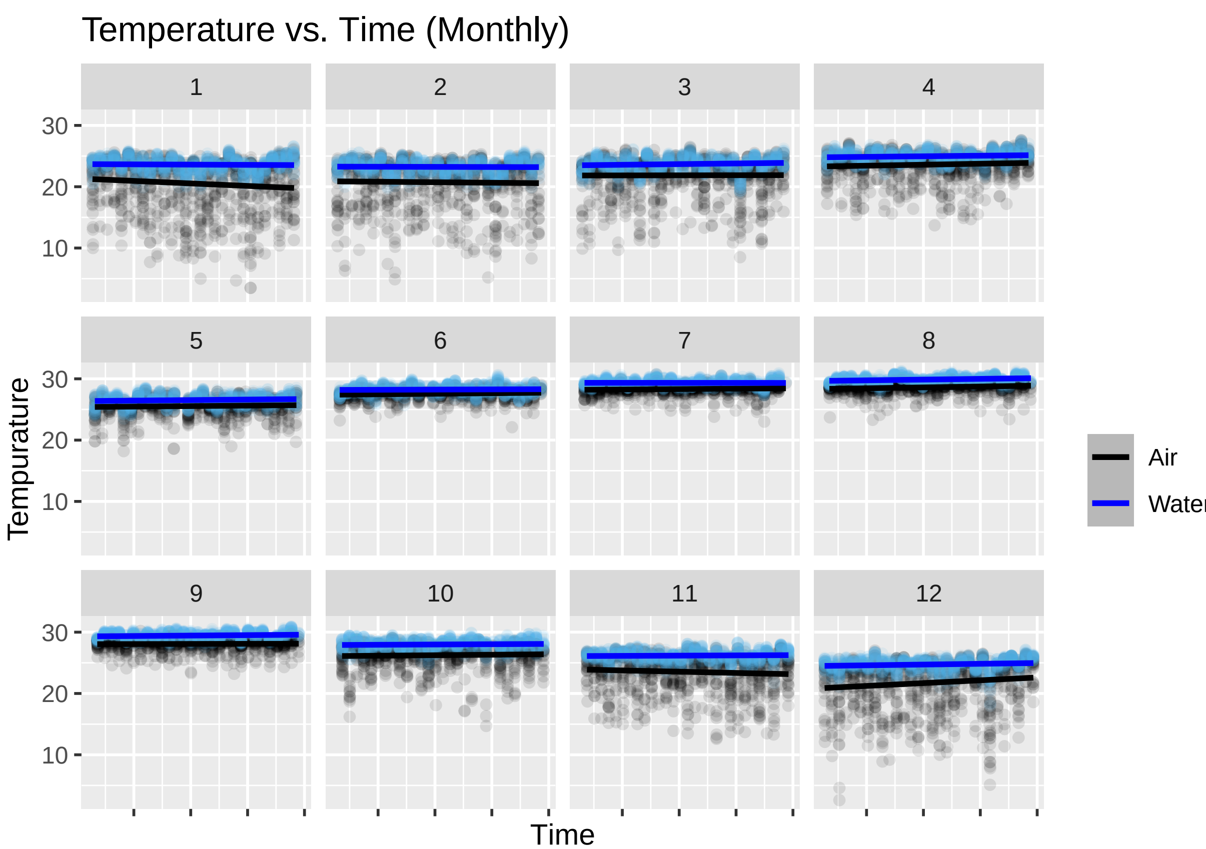
```r
kable(Buoymonth, col.names=c("Month",
                             "Monthly air temperature change every year in celcius degrees",
                             "Monthly water temperature change every year in celcius degrees"))
```

Month	Monthly air temperature change every year in celcius degrees	Monthly water temperature change every year in celcius degrees
January	0.0578672	0.0156709
February	0.0578672	0.0156709
March	0.0578672	0.0156709
April	0.0578672	0.0156709
May	0.0578672	0.0156709
June	0.0578672	0.0156709
July	0.0578672	0.0156709
August	0.0578672	0.0156709
September	0.0578672	0.0156709
October	0.0578672	0.0156709
November	0.0578672	0.0156709
December	0.0578672	0.0156709

```r
plotMonth <- ggplot(data=Buoysub)+
  geom_point(aes(x=DATE,y=ATMP),alpha=0.1,color="black")+
  geom_point(aes(x=DATE,y=WTMP),alpha=0.1, color = "#56B4E9")+
  geom_smooth(aes(x=DATE,y=ATMP,color="Air"),method="lm", formula="y~x")+
  geom_smooth(aes(x=DATE,y=WTMP,color="Water"),method="lm", formula="y~x")+
  scale_color_manual("",breaks=c("Air","Water"),values=c("black","blue"))+
  ggtitle("Temperature vs. Time (Monthly)")+
  xlab("Time")+
  ylab("Tempurature")+
  theme(axis.text.x = element_blank())+
  facet_wrap(~factor(MM))
ggsave("plotMonth.jpg", plot=plotMonth)
```

```
## Saving 6.5 x 4.5 in image
```

```r
plotMonth
```



The coefficients of the linear regression models show the variation of monthly average air temperature and water temperature from 1987 to 2016. In January and February, both the air temperature and water temperature decrease. In March, April, May, June, August, September, October and December, both the air temperature and water temperature increase. In July, the air temperature increases while the water temperature decreases. In November, the air temperature decreases while the water temperature increases. In most cases, the air temperature and the water temperature is increasing from 1987 to 2016. According to National Geographic's artical *Why cold weather doesn't mean climate change is fake*, colder winter could be a signal of extreme weather.

Average annual temperature change

We found that temperature decreases in some months. In this step, we are going to find out if it would make a difference to the annual average temperature. Similar as the last step, we are going to plot an average annual temperature versus year graph for visualization.

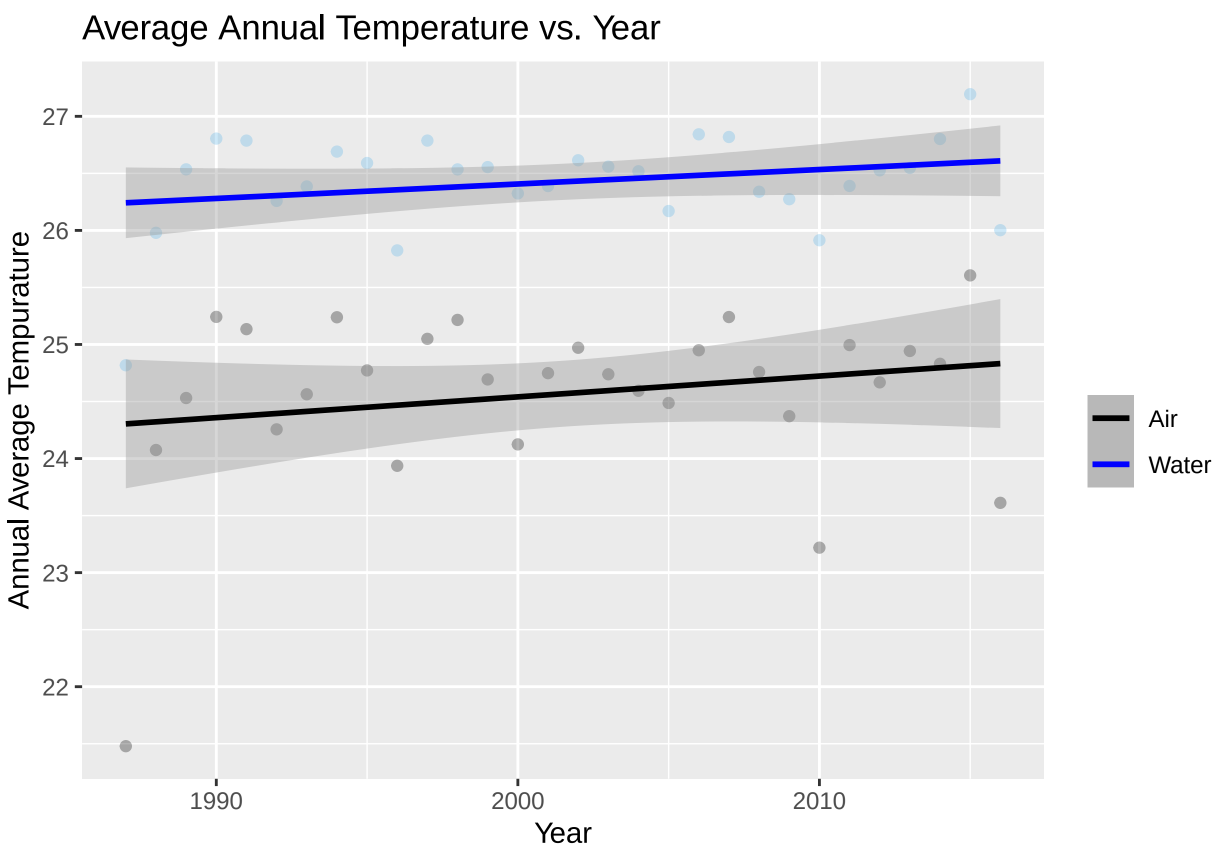
```
avg <- Buoysub%>%group_by(YY)%>%summarize(meanATMP = mean(ATMP), meanWTMP = mean(WTMP))
```

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

```
plotAnnualAvg <- ggplot(data=avg)+
  geom_point(aes(x=YY, y=meanATMP), alpha=0.3, color="black")+
  geom_point(aes(x=YY, y=meanWTMP), alpha=0.3, color="#56B4E9")+
  geom_smooth(aes(x=YY,y=meanATMP,color="Air"),method="lm", formula="y~x")+
  geom_smooth(aes(x=YY,y=meanWTMP,color="Water"),method="lm", formula="y~x")+
  scale_color_manual("",breaks=c("Air","Water"),values=c("black","blue"))+
  ggtitle("Average Annual Temperature vs. Year")+
  xlab("Year")+
  ylab("Annual Average Tempurature")
ggsave("plotAnnualAvg.jpg", plot=plotAnnualAvg)
```

```
## Saving 6.5 x 4.5 in image
```

```r
plotAnnualAvg
```




```r
fitA_annual <- lm(meanATMP~YY, data=avg)
fitW_annual <- lm(meanWTMP~YY, data=avg)
cat(paste0("By average, the annual average air temperature increases ",
           as.character(coef(fitA_annual)[2]),
           " celcius degrees \nand the annual average water temperature increases ",
           as.character(coef(fitW_annual)[2]),
           " celcius degrees every year. ",
           sep=""))
```

```
## By average, the annual average air temperature increases 0.0182281750720344 celcius degrees
## and the annual average water temperature increases 0.0126780717416379 celcius degrees every year.
```

The graph and the coefficient of the linear regression line show that even though temperature decreases in some months, both the air temperature and water temperature is getting warmer in general. The annual average temperatures in 1987, 2010, 2015 and 2016 are also noticeable because of the large difference comparing to the others, which indicates the occurrence of extreme climate. The abrupt change from 2015 to 2016 further demonstrates the unusual climate change.

Conclusion

There are evidence validating the climate change in Boston area. Even though the overall temperature is changing slowly, it is gradually increasing in the main direction. The decreasing temperature in winter also prove the extreme climate change. However, the conclusion is only based on temperature change. To further improve the credibility of global warming, we need to increase our sample size and number of predictors. For example, to prove that the climate change is really global, we need to consider about other places in the world. Moreover, temperature change is not the only phenomenon for global warming. Data such as sea levels and amount of carbon dioxide in the ocean could also help us to make a comprehensive analysis on evidence of global warming.

Reference

NASA. (2018, September 21). *Climate change evidence: How do we know?* Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/evidence/

Gibbens, S. (2019, January 23). *Why cold weather doesn't mean climate change is fake.* National Geographic. https://www.nationalgeographic.com/environment/2019/01/climate-change-colder-winters-global-warming-polar-vortex/


```r
citation("knitr")
```

```
##
## To cite the 'knitr' package in publications use:
##
##   Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report
##   Generation in R. R package version 1.29.
##
##   Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition.
##   Chapman and Hall/CRC. ISBN 978-1498716963
##
##   Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible
##   Research in R. In Victoria Stodden, Friedrich Leisch and Roger D.
##   Peng, editors, Implementing Reproducible Computational Research.
##   Chapman and Hall/CRC. ISBN 978-1466561595
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```r
citation("tidyverse")
```

```
##
##   Wickham et al., (2019). Welcome to the tidyverse. Journal of Open
##   Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Welcome to the {tidyverse}},
##     author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostin
##     year = {2019},
##     journal = {Journal of Open Source Software},
##     volume = {4},
##     number = {43},
##     pages = {1686},
##     doi = {10.21105/joss.01686},
##   }
```

```r
citation("stringr")
```

```
##
## To cite package 'stringr' in publications use:
```

```
##
##   Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for
##   Common String Operations. R package version 1.4.0.
##   https://CRAN.R-project.org/package=stringr
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {stringr: Simple, Consistent Wrappers for Common String Operations},
##     author = {Hadley Wickham},
##     year = {2019},
##     note = {R package version 1.4.0},
##     url = {https://CRAN.R-project.org/package=stringr},
##   }
```

```r
citation("lubridate")
```

```
##
## To cite lubridate in publications use:
##
##   Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy
##   with lubridate. Journal of Statistical Software, 40(3), 1-25. URL
##   http://www.jstatsoft.org/v40/i03/.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Dates and Times Made Easy with {lubridate}},
##     author = {Garrett Grolemund and Hadley Wickham},
##     journal = {Journal of Statistical Software},
##     year = {2011},
##     volume = {40},
##     number = {3},
##     pages = {1--25},
##     url = {http://www.jstatsoft.org/v40/i03/},
##   }
```

```r
citation("dplyr")
```

```
##
## To cite package 'dplyr' in publications use:
##
##   Hadley Wickham, Romain François, Lionel Henry and Kirill Müller
##   (2020). dplyr: A Grammar of Data Manipulation. R package version
##   1.0.2. https://CRAN.R-project.org/package=dplyr
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {dplyr: A Grammar of Data Manipulation},
##     author = {Hadley Wickham and Romain François and Lionel {
##              Henry} and Kirill Müller},
##     year = {2020},
##     note = {R package version 1.0.2},
##     url = {https://CRAN.R-project.org/package=dplyr},
```

```
## }
```

```r
citation("ggplot2")
```

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {https://ggplot2.tidyverse.org},
##   }
```

```r
citation("gridExtra")
```

```
##
## To cite package 'gridExtra' in publications use:
##
##   Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid"
##   Graphics. R package version 2.3.
##   https://CRAN.R-project.org/package=gridExtra
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {gridExtra: Miscellaneous Functions for "Grid" Graphics},
##     author = {Baptiste Auguie},
##     year = {2017},
##     note = {R package version 2.3},
##     url = {https://CRAN.R-project.org/package=gridExtra},
##   }
```