NOAA Buoy Data Analysis

Rachel Feingold

2/27/2017

Using at least 20 years of NOAA buoy data (pick any buoy you like), estimate the long term linear trend in air temperature. For the linear regression model air temperature vs time, the errors (the Epsilons) are assumed to be independent and identically distributed. We examine them to see if they are normally distributed and if theres is constant variance and their mean = 0. Since this data is a time series, we know that the correlated observation over time are not independent, we want to see if the relationship is sinusoidal. This can be predicted because air temperature correlated with the yearly cyclical seasons.

I picked buoy 44007 off the coast of Portland, ME

This is reading one year (2016) of buoy data. 1. Read table without header because you need the first row of data included in the document. This function creates a data frame from the web file. 2. Need to read the column names, set the column names as a character vector, "header"" 3. Create buoy data frame from the web file, without the header 4. Need to reattach the header, which are the column names ** SO we will be looking at the buoy data frame

```
a <- read.table("http://www.ndbc.noaa.gov/view_text_file.php?filename=44007h2016.txt.gz&dir=data/histori
fyl <- "http://www.ndbc.noaa.gov/view_text_file.php?filename=44007h2016.txt.gz&dir=data/historical/stdme
header <- scan(fyl,nlines=1,what=character())
buoy <- read.table(fyl,skip=2,header=FALSE)

colnames(buoy)<-header
head(buoy)
```

```
##     #YY MM DD hh mm WDIR WSPD  GST WVHT   DPD  APD MWD   PRES ATMP WTMP
## 1 2015 12 31 23 50  249  3.3  4.9 0.84 10.00 6.72  90 1013.8  5.0  7.2
## 2 2016  1  1  0 50  255  9.1 10.6 1.06 10.00 6.91  98 1013.3  5.2  7.2
## 3 2016  1  1  1 50  250  8.8 10.4 1.02 10.00 5.27 106 1013.4  5.1  7.3
## 4 2016  1  1  2 50  253  7.9  9.7 1.06  9.09 5.37  88 1013.3  4.6  7.2
## 5 2016  1  1  3 50  256  8.9 10.4 1.02  9.09 5.00  98 1013.1  4.3  7.2
## 6 2016  1  1  4 50  248  8.2 10.2 0.97 10.00 5.25  90 1012.9  4.2  7.3
##   DEWP VIS TIDE
## 1  1.5  99   99
## 2  1.3  99   99
## 3  1.3  99   99
## 4  1.5  99   99
## 5  1.0  99   99
## 6  0.7  99   99
```

Now we want to look at the data overa period of years (1982-1999). After 1999, 17 variables are observed, limitation of this loop technique to extract data.

```
str1 = "http://www.ndbc.noaa.gov/view_text_file.php?filename="
str2 = ".txt.gz&dir=data/historical/stdmet/"
buoynum <- "44007"

# years to read since data layout was the same for these years
y1 <- 1982
y2 <- 1999

data <- NULL
```

```
while(y1<=y2){
  # make the file name
  fyl <- paste0(str1,buoynum,"h",y1,str2)

  # read the column names # scan first line so it becomes the header line
  header <- scan(fyl,nlines=1,what=character())

  # read the data
  buoy <- read.table(fyl,skip=2,header=FALSE)

  # re-attach the column names
  colnames(buoy)<-header

  # re-assign 999 as NA
  buoy$ATMP[buoy$ATMP==999]<-NA

  # add the data to the variable # data is where we are colleccting the observations
  data<-c(data,buoy$ATMP)

  # write the year to the screen
  cat(paste(y1))

  # increment the year
  y1<-y1+1
}
```
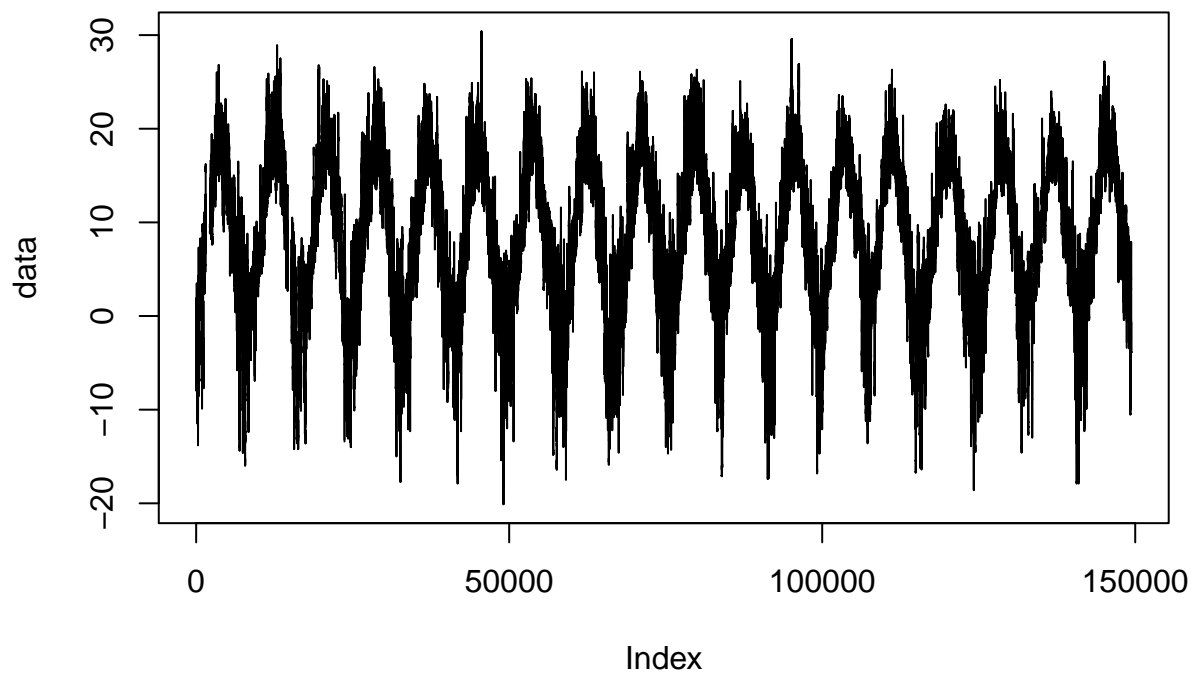
```
## 1982198319841985198619871988198919901991199219931994199519961997199819 99
```

```
# built in plots
plot(data,type="l")
```

```
mm<-c(1:length(data))

# linear regression
reg1<-lm(data~mm)

# output linear regression
summary(reg1)
```

```
##
## Call:
## lm(formula = data ~ mm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.2131   -5.3731    0.4259    6.6708   22.3036
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.881e+00  4.127e-02   190.96   <2e-16 ***
## mm          4.730e-06  4.768e-07     9.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.824 on 147247 degrees of freedom
##   (2125 observations deleted due to missingness)
## Multiple R-squared:  0.0006678,  Adjusted R-squared:  0.000661
## F-statistic:  98.4 on 1 and 147247 DF,  p-value: < 2.2e-16
```

What are your findings? The sinusoidal regression between time and air temperature is consistent with our predictions. On earth we experience seasons, cyclical periods of warming and colding. This constant and consistent change is reflected in the regression.

Ggplot time series plot with a time axis that you have produced with lubridate: