Atlanta Temperatures

Using the 20 years of daily high temperature data for Atlanta (July through October), I built and used an exponential smoothing model to help make a judgment of whether the unofficial end of summer has gotten later over the 20 years.

Methodology

First, I loaded any necessary libraries. Then I loaded the data and did some high level exploratory analysis. I noticed the the column header had a "x" in from of each year

```
#load data and libraries

library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```r
library(lubridate)



mydata <- read.table("temps.txt", header = TRUE, sep = "\t")
names(mydata)
```

```
 [1] "DAY"   "X1996" "X1997" "X1998" "X1999" "X2000" "X2001" "X2002" "X2003"
[10] "X2004" "X2005" "X2006" "X2007" "X2008" "X2009" "X2010" "X2011" "X2012"
[19] "X2013" "X2014" "X2015"
```

After examining the data, I decided to reshape the data to better analyze it. Instead of having a column for each year, I reshaped the data so that each column represents a unique data point. I did this by adding a full date column that combined the year, month and day. Before doing so, I also removed the "X" in front of the year column. I saved the new data set as temps_long

```r
#reshape data to have a full yyyy-mm-dd columm to better iterate and analyze data
newdata <- mydata %>%
  pivot_longer(-DAY, names_to = "year", values_to = "tmax") %>%
  mutate(
    year = as.integer(str_replace_all(year, "\\D", "")),   # Removes X in years
    date = dmy(paste(DAY, year))
  )

head(newdata)
```

```
# A tibble: 6 x 4
  DAY    year  tmax date
  <chr> <int> <int> <date>
1 1-Jul  1996    98 1996-07-01
2 1-Jul  1997    86 1997-07-01
3 1-Jul  1998    91 1998-07-01
4 1-Jul  1999    84 1999-07-01
5 1-Jul  2000    89 2000-07-01
6 1-Jul  2001    84 2001-07-01
```

Then to calculate the end of summer, I grouped by all temperature within a year and took the days above 85F. Then I took the latest date in that year. Lastly, the mutate function converts that hottest day into a day of the year number (what day is it out of 365), allowing for easier comparison later on in the analysis.

```r
# Define "end of summer" (EOS)
eos <- newdata %>%
  group_by(year) %>%
  filter(tmax > 85) %>%                # keep only hot days
  summarize(last_hot = max(date)) %>%  # last hot day in Jul-Oct
  mutate(doy = yday(last_hot))
```

Then I created a time series using the end of summer day. Then the Holt Winters Model exponential smoothing was used, with seasonality (gamma) set to False we only have one value per year. The y Axis is the End of Summer value mapped to its respective day of the year (out of 365).

```r
# Build time series of EOS
y <- ts(eos$doy, start = min(eos$year), frequency = 1)

# Fit HoltWinters model
fit_hw <- HoltWinters(y, gamma = FALSE)



# View  and plot results ---
print(fit_hw)
```

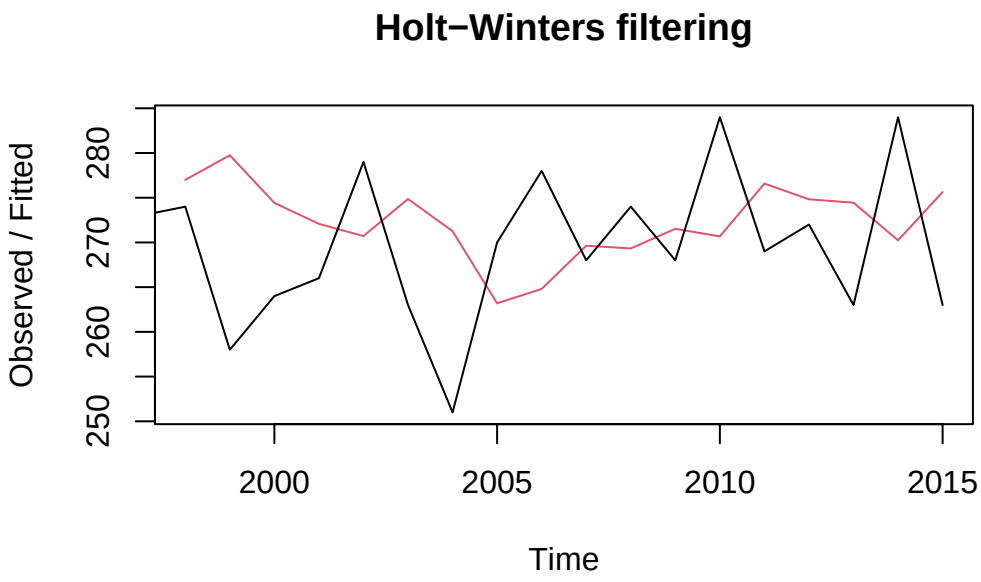
```
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = y, gamma = FALSE)

Smoothing parameters:
 alpha: 0.3373883
 beta : 0.2382403
 gamma: FALSE

Coefficients:
         [,1]
a 271.3748265
b  -0.2563092
```

```r
plot(fit_hw)
```



Results

Based on the HoltWinters Exponential model, it seems like starting in 2005, there is a slight upward trend in End of Summer temperatures. The alpha value is 0.33 and the beta value is 0.23m

Analysis

Since the alpha value is closer to 0, it means more weight is given to observations in the more distant past. Since the beta is greater than 0, it also means the the end of summer is trending later. The graph doesn't show a clear change in the end of summer over the past 20 years as there are a lot of fluctuations, but from 2005 onward, we can see a slight upward slop in the red line, which is inline with what we are observing with the beta.

For more accurate analysis, we would need to study a broader range of dates. Since according to the National Oceanic and Atmospheric Administration, Earth's temperature has rise by an average of 0.11F per decade since 1850, which is 2F in total.