Climate Change Analysis from Buoy Station 44013

Anna, Bruce, and Jenna

9/25/2020

Introduction

The National Oceanic and Atmospheric Administration (NOAA) studies the sky and the ocean to preserve our planet. Their mission is "to understand and predict changes in climate, weather, oceans, and coasts, to share that knowledge and information with others, and to conserve and manage coastal and marine ecosystems and resources." NOAA's vision of the future is to protect resilient and healthy ecosystems, communities, and economies in the face of change.

Research Question

The purpose of this project is to investigate evidence of global warming from weather buoy data. Meteorological data has been collected from NOAA National Data Buoy Center Station 44013, which is located about 11 miles offshore from Boston, MA. The specific question being asked is as follows: Can data from a single weather buoy provide evidence to support the conclusion that global warming is a real phenomenon?

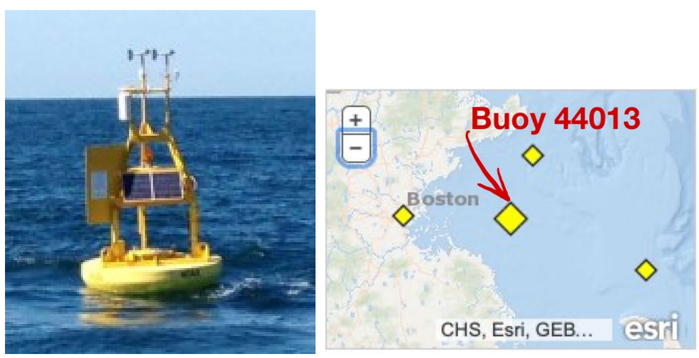

Figure 1: Weather Buoy Station 44013

Approach

While there are several different approaches one could take to look for evidence of global warming or climate change, we chose to focus on air temperature as our outcome variable of interest. We are interested in analyzing air temperature patterns to determine whether there is a significant upward trend over the course of several years. Specifically, we chose to analyze the average air temperature for each month from 2000 to 2018. Our primary analysis consisted of running a series of 12 linear regressions–one for each month–with average monthly air temperature as the outcome variable and the year as the predictor. We can confirm our findings on global warming by observing the coefficients of the predictor for each fitted regression model.

Code Organization

The R code is organized into several chunks with comments for clarity. To start, the following R packages are loaded: tidyverse, stringr, rstanarm, lubridate, and gridExtra. Then, the code reads the meteorological data for a specified range of years from the NOAA website using a series of urls. This produces a separate data frame for each year in the specified range of 2000-2018.

The next goal is to create a single data frame containing the variables of interest for each of every year in the range. This is accomplished by creating a loop within a loop of code:

- The inner loop: Contains the columns that specified years, months, and air temperatures to create one smaller dataset. It then filters out the NAs[1] which is uncollected data.
- The outer loop: Creates a new variable for average temperature (AvgTMP) by taking the mean of the air temperature (ATMP) for every month of every year within the set.

Finally, a loop was added to run a series of linear regressions to analyze the change in average temperatures over time. For each month across every year in the range, plots of the data points with regression lines are created using ggplot2. This was the final step once a tidy data frame was created with only the outcome and predictor variables.

Conclusions

Results

By observing the coefficients of the predictor (year) for each fitted regression model, we were able to determine the general trend of the air temperature for each month over the course of 18 years. Plots of the data and regression lines for each month are shown in Figure 2. Slopes of each regression line are shown in Table 1. Overall, we observed a slight general rise in temperatures, with the average slope of the regression lines equal to 0.028. Only one month, March, showed a decrease in average temperatures over the course of our 18 year range.

[1] In the data, there were values with "NA" that designated an unknown or untracked value. This is because NOAA has the occasional lack of funding which leaves empty spaces in their data. So, we had to ignore those data points in our research.

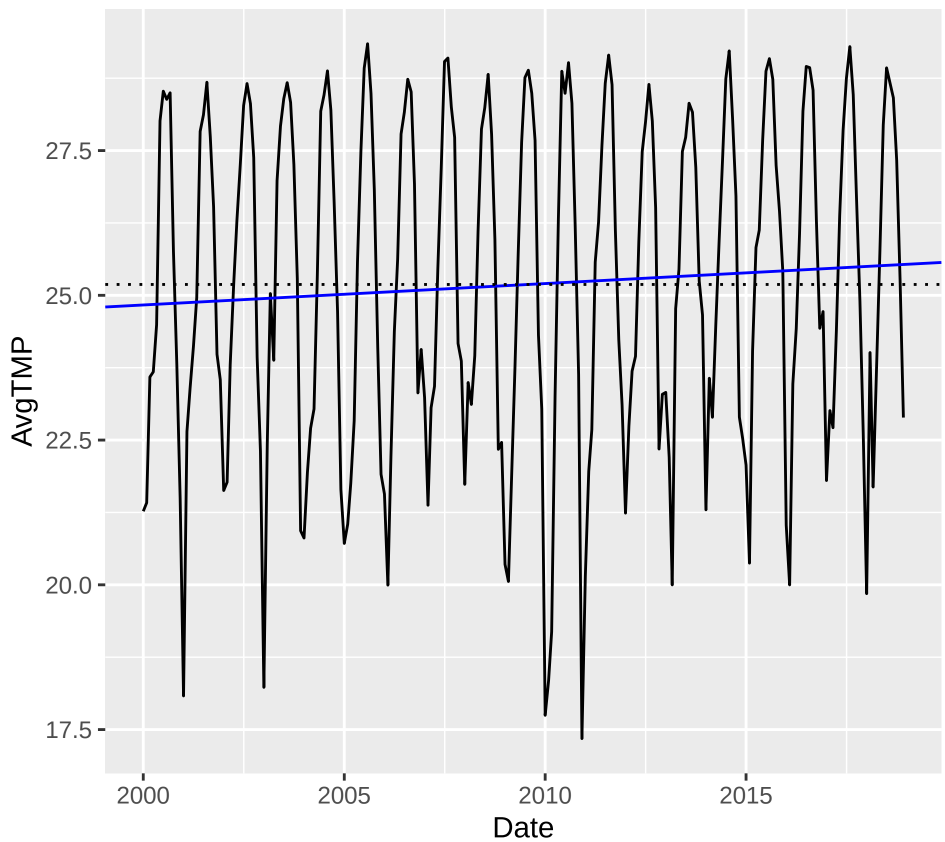

Figure 2: Plot and regression line of average temperatures for each month from 2000 to 2018. Dashed line indicates overall average temperature. Solid line represents the regression equation.

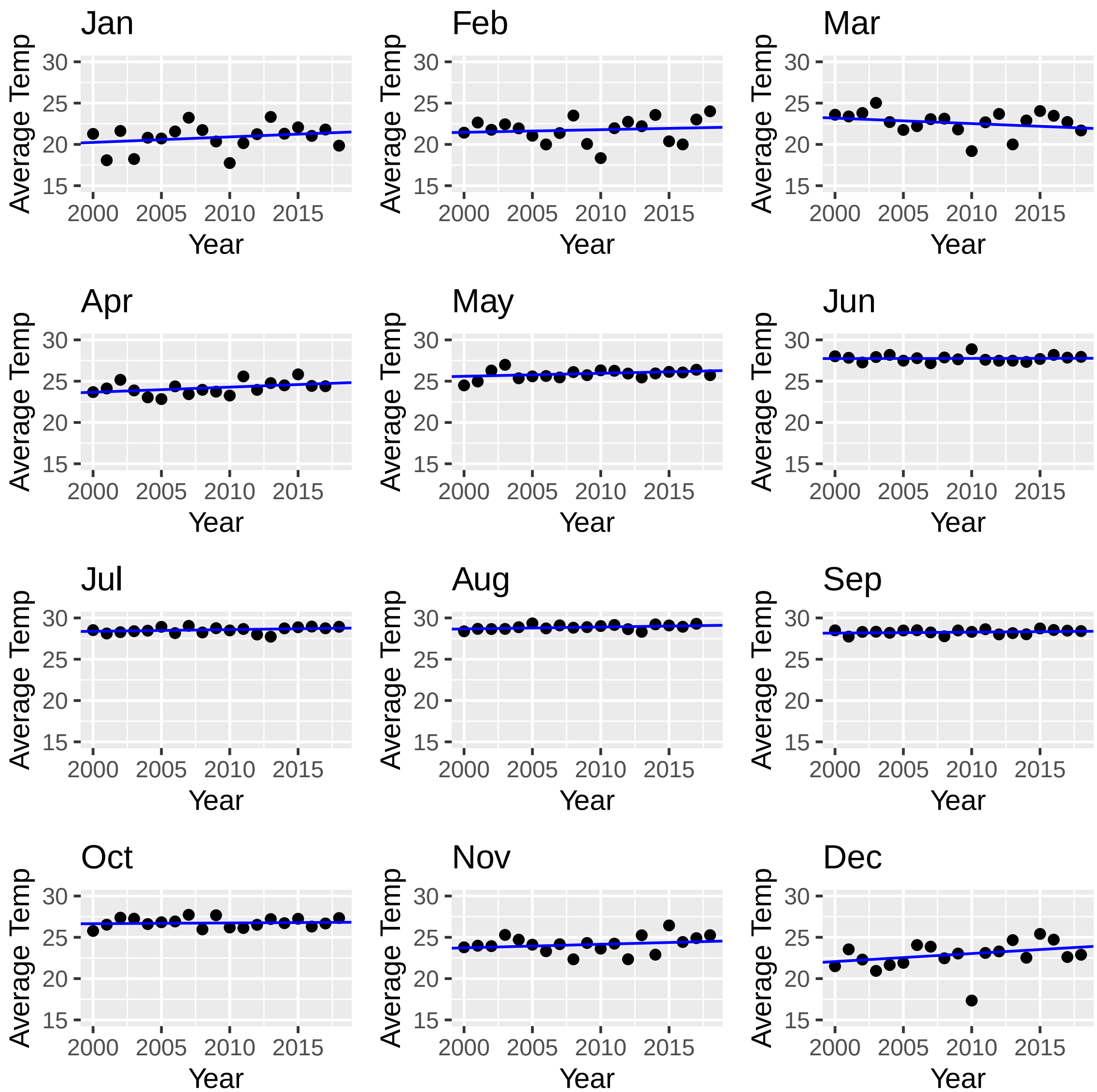

Figure 3: Plots of the linear regression line separated for each month from 2000 to 2018.

Table 1: Slopes of the regression lines for every month from 2000 to 2018.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Slope	0.067	0.032	-0.066	0.062	0.036	0.002	0.021	0.023	0.012	0.009	0.043	0.097

Discussion

The results stated above suggest a general rise in temperatures off the coast of Boston between 2000 and 2018. The average slope of the regression lines is small, but we believe that even a small rise in temperatures can have significant environmental impact. While it is tempting to claim that this positive linear trend is evidence of global warming, there are limitations to consider.

First, we must be careful not to extrapolate when interpreting our results. Because our data was compiled from a single weather buoy, we can only draw conclusions about temperature trends at this geographic location. To make any claims about evidence of climate change on a global scale is beyond the scope of the analyses presented here.

Second, we have not analyzed any confidence intervals or significance levels for our regressions. Including p-values and/or confidence intervals intervals would give us a better idea of whether this observed rise in temperatures is a legitimate phenomenon or due to random chance.

In addition to limitations, there are future analyses to consider. In examining each month separately, we have in essence taken twelve subsets of our data with which to observe trends in temperature. Another approach would be to combine these twelve subsets into a single data frame, which could then be used to run a single linear regression. Future analyses could include this model to explain temperature trends in a more cohesive manner.

Finally, we included a narrow range of years to analyze. Future analyses may seek to broaden this range in order to observe meteorological trends over a more extended period of time, perhaps dating back to the 1970s.

References

1. Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
2. 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/.
3. Goodrich B, Gabry J, Ali I & Brilleman S. (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1 https://mc-stan.org/rstanarm.
4. Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr
5. Hao Zhu (2020). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.2.1. https://CRAN.R-project.org/package=kableExtra
6. Simon Urbanek (2013). png: Read and write PNG images. R package version 0.1-7. https://CRAN.R-project.org/package=png
7. US Department of Commerce, National Oceanic and Atmospheric Administration. "NDBC Station History Page." NDBC, 8 Nov. 1996, www.ndbc.noaa.gov/station_history.php?station=44013.
8. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
9. Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.29.