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?

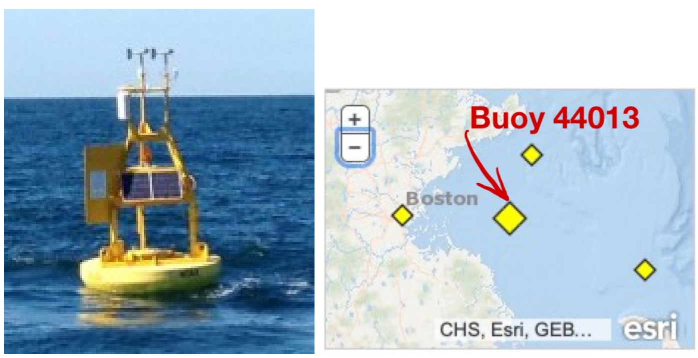

Figure 1: Weather Buoy Station 44013

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 or deny evidence of global warming by observing the coefficients of the predictor for each fitted regression model. We then combined these twelve subsets into a single data frame, which was used to run a single linear regression in order to examine the slope of the regression line for each month over 18 years in a single plot.

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 tidy the data such that a single data frame is produced 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. Next, a linear regression was run to analyze average temperature as a function of Date, for each month over our 18 year range. A plot was created displaying the data, regression line, and overall average temperature.

Then, 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, regression slopes are displayed in a table. Plots of the data points with regression lines are created using ggplot2, and combined into a single figure using grid.arrange().

Finally, citations are created for each R package used in the above code.

Conclusions

Results

By observing the coefficients of the predictor (Date) for the combined regression model, we are able to observe the change in average temperature over the course of 18 years. The plot and regression line for this model is shown in Figure 2. The slope of this regression line is 9.842×10^{-5}. Over the course of our 18 year range, the average temperature increased by 0.68 degrees Celsius.

For each of the twelve separate fitted regression models, by observing the coefficients of the predictor (Year), we were able to determine the general trend of the air temperature for each month, independent from the other months. Plots of the data and regression lines for each month are shown in Figure 3. Slopes of each regression line are shown in Table 1. The average slope of the regression lines is 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.

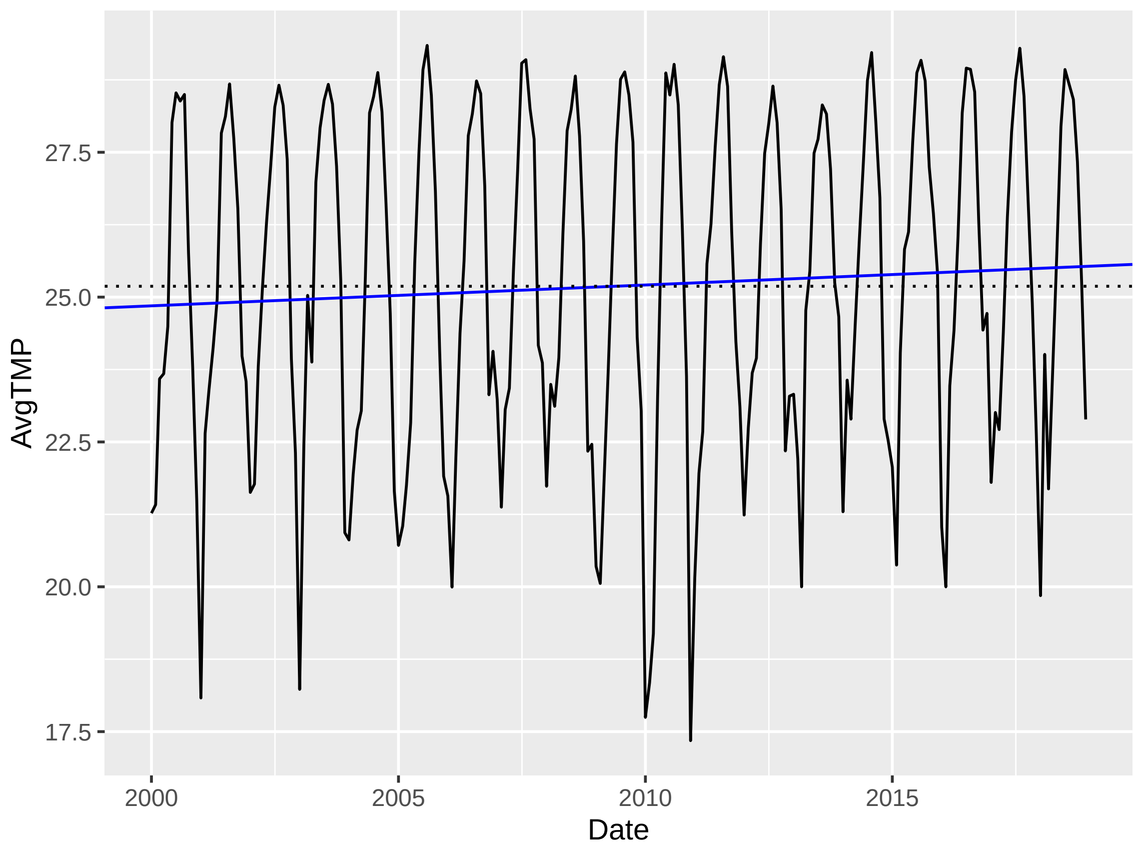

Figure 2: Plot and regression line of average temperatures for each month from 2000 to 2018. The black solid line shows the variation in temperatures from month to month. The dashed line indicates overall average temperature. The blue solid line represents the regression equation. The linear regression line is positive, showing that there is a temperature increase. This temperature increase is 0.68 degrees Celsius.

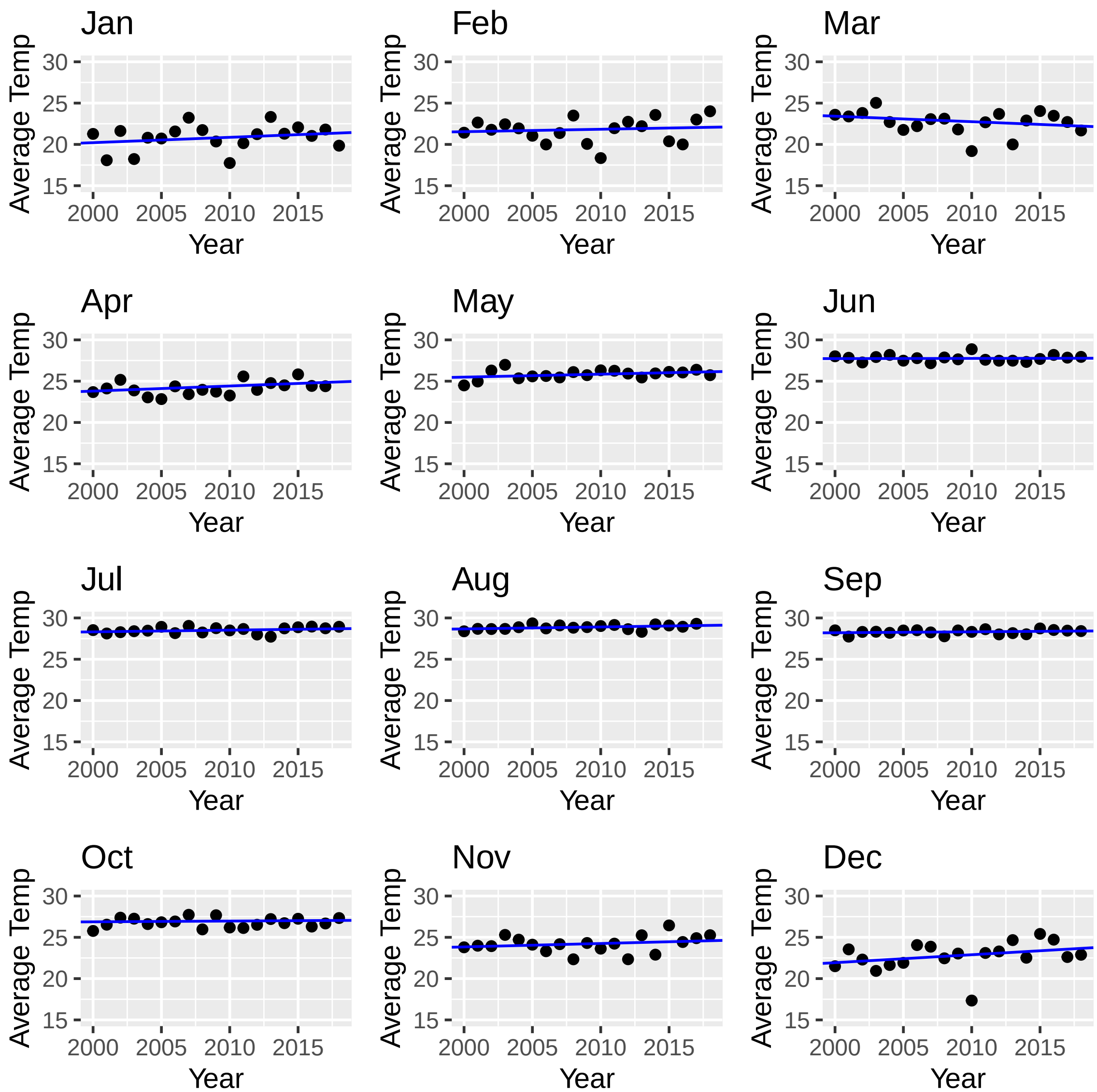

Figure 3: Plots of the linear regression line separated for each month from 2000 to 2018.

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	0.065	0.030	-0.067	0.062	0.036	0.002	0.021	0.024	0.011	0.009	0.041	0.096

Discussion

The results stated above suggest a general rise in temperatures off the coast of Boston between 2000 and 2018. By combining these twelve subsets into a single dataframe, we are able to run a single regression and interpret the slope (see Figure 2). The positive slope is what indicates a rise in temperatures. In addition, the value of the average temperature increases by 0.68 degrees Celsius over the course of our range of dates.

We further analyze this data by looking at regressions for each month separately. In examining each month separately (see Figure 3), we have taken twelve subsets of our data with which to observe trends in temperature. The average slope of these twelve regression lines is small, but we see that 11 out of 12 months have a positive slope. This suggests that our results are likely not due to random chance but instead due to a genuine change in temperature patterns. While it is tempting to claim that this positive linear trend is evidence of global warming (an increase in temperatures on a global scale), there are limitations to consider.

First, we must be careful not to extrapolate when interpreting our results. Due to our data being compiled from a single weather buoy, we can only draw conclusions about temperature trends at this geographic location. It is beyond the scope of the analyses presented here to make any claims about evidence of climate change on a global scale. Second, we have not analyzed any confidence intervals or significance levels for our regressions. Including p-values and/or confidence intervals would give us a better idea of whether this observed rise in temperatures is a legitimate phenomenon with a significant effect size, and not simply due to random chance. 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.

Additional Notes

There were a few techniques that we strived to achieve but did not work in our time frame. First, we attempted to condense the twelve plots in the grid.arrange() function but we were not able to successfully run it in the allotted time. Instead, we simply listed all twelve plots (P01, P02, . . . , P12) in our grid.arrange() which printed exactly what we wanted. Lastly, we attempted to transpose our table of slopes for each month in one step, but we settled on using a multi-step process for the sake of time (violating rules of tidy data for the sake of readability). We resulted with the layout we liked, but if we had more time, we would have condensed five lines of code into one or two to have our code be more efficient.

In addition to these ideas with time constraints, we had ideas about the layout of the report. We wanted to have a whole page dedicated to the twelve plots and the table but when we used /newpage, although it worked, there was a lot of blank space on the page before. So, we decided to split the figure of plots and the table to two separate pages. We also ran into a problem about having the figure and table on the same page as the table would always print on the next page. Again, with the time limit, we decided not to worry about it.

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.