Marilyn Davidson

2/27/2023

Cherry Blossom Bloom Date Prediction Model

The bloom dates of the cherry blossom trees in Kyoto, Japan have been recorded since the year 812 AD. However, a dependable modeling method for predicting future bloom dates still has not been created. I attempted to create an efficient model which could adequately execute this task for cherry blossom trees located in Kyoto, Japan, Washington DC, USA, Liestal, Switzerland, and Vancouver, Canada for the next 10 years. To do this I made a loess regression model in R for each location. I started by extracting historic temperature data for each location and then subsetting that data into just the temperatures recorded between the start of the year and the recorded bloom date for that year and location. Then, I created a sum of the temperatures for each year in hope to emulate a growing degrees days (GDD) function. This function will help to find if there is a particular temperature threshold that is predictive of the bloom dates and has been used to determine the bloom dates of other flowers.

I chose to do a loess model with my GDD function and year as the explanatory variables and the day of the year that the cherry blossoms bloom as the response variable. I determined that a loess function would be the best for my prediction because of the graph created when plotting the years on the x-axis and the sums of each year on the y-axis, which I have displayed below.

Chart, scatter chart

Description automatically generated

You can see that there is a gap of missing data as well as no obvious trends or patterns. I used a smoothing function to create a line of best fit, the function indicated that it used a loess regression model to obtain the formula for the line.

Chart

Description automatically generated

The line and its confidence interval indicates that there will be a lower sum of average temperatures at the time of blooming this year.

As mentioned before, you can see that there is a gap spanning over 20 years in the middle of the data, and the graph does not follow a linear trend. I concluded that a loess regression model was the best for this model because loess regression is about determining local predictions of slope rather than for the entire model by using a combination of linear regression and a time series function. This will help with the variability created by the gap in years of temperature data and nonlinear patterns which we can see by the inflation of the confidence interval in the gap of data, is a serious problem when trying to make a predictive model.

My observations found that this year the 70% of the cherry blossom trees in Kyoto, Japan will bloom on April 9th, 2023. I used the same method of a loess regression model for predicting the bloom dates in Liestal, DC, and Vancouver as well. When predicting Vancouvers bloom dates, I used the DC model because it had the most similar longitude, latitude and altitude as Washington, DC.