Narrative

Data Preparation and Visualization: For my competition entry I firstly imported the flowering date observations for the Washington DC, Liestal and Kyoto sites, and visualized the cherry blossom bloom over time, using firstly a linear trend and secondly gam smoothing. Climate Data Analysis: Then, flowering day observations matched with climate data for Japan, Kyoto, Liestal, and Washington D.C. was read. In addition to the Kyoto site, many more of the Japanese sites have available historic climate data. A generalized additive model (gam) was fitted to the climate data to understand its relationship with bloom dates. I modelled the response of changes in flowering date to mean monthly maximum temperature for January and February measured at sites as close as possible to the sites of the flowering Prunus trees, in the years of the observations. One complication is the paucity of historic observations for Vancouver (2 years), and the complete lack of observation for New York City, at least as finalized data sets constructed from systematic sampling effort. To overcome this, predictions were made for Vancouver and New York based on the fitted model, and estimates were rounded. Predictions were attached to their respective datasets. A limitation of this method is that incorporating these estimates into the data set to be used to model to predict for the 2024 flowering dates creates circularity for the final predictions. Incorporating Predictions for 2024: Climate data for January and February 2024 was read. These data were sourced from NOAA, JMA, and meteoblue. By the 27th of February, mean monthly maximum temperature values for the sites to that date were combined with forecast temperatures for each site, sourced from AccuWeather, to give mean monthly values for February. Locations were converted into factors. A second gam model, with location as a factor of the February smooth was fitted to predict bloom dates for the five sites in 2024. Predictions were rounded to the nearest integer. Predictions were adjusted regionally for Washington D.C., Vancouver, and New York City, by taking one day from each estimate to bring the estimate from 80% blooms open to 70% blooms open. The prediction for Liestal was adjusted to the regionally used measure of 25% blooms open by taking five days of the estimate. (The number of days used to adjust the estimate was determined from a time-lapse video of cherry blossom, based on the amount of visible blossom colour: https://www.youtube.com/watch?v=uL6XLdZJ35o). Predictions were re-ordered from alphabetical to the order required for the competition and written to a CSV file named "predictions.csv". This modelling reflects the role of maximum day time temperatures in driving changes in flowering time. Higher temperatures are known to advance flowering timing in contrast to lower temperatures which delay flowering. The premise of the competition is that it is difficult to predict the timing of an arbitrarily designated peak by more than 10 days in advance. However, there are demonstrated close relationships with the weather at the time of these peaks, and these weather conditions are also difficult to predict accurately more than 10 days in advance. Because of this, I based my predictions on a model that incorporated the mean monthly maxima of January and February, but I wanted the variables that are closer to the time of flowering to be more influential. In the generalized additive model used for predictions, I constrained the smooth of the January mean monthly maximum temperature, k = 3. The smooth for the February mean monthly maximum temperature was not constrained and the location factor was applied.