Impacts of Climate Extremes on Vegetation

By: Andrea Newlands, Linda Sylvester and Jacqueline Shulack

Background

Phenology is the study of seasonal cycles in natural phenomena. Local environmental factors, such as temperature, sunlight, and precipitation often drive the timing of various events such as first bloom or first leaf. The trends and changes in this timing can be an indicator of climate change. Temperature in particular has been extensively studied as a contributor to earlier first leaf and first bloom dates. We would like to investigate how temperature and precipitation extremes, as indicated by values outside of normal ranges, impact the timing of these vegetative behaviors.

Motivation

We are interested in exploring how climate change is influencing ecosystems. We propose investigating how the lengths and timings of plant phenophases are changing across the Northeastern and Southeastern United States as they relate to temperature and precipitation extremes.

Goals:

1. Determine temporal trends of annual first leaf/bloom dates

2. Determine temporal trends in phenophase length (i.e., first bloom to end bloom)

3. Determine temporal trends in annual/seasonal temperature and precipitation extremes

4. Characterize any association between phenophases and climate extremes

Deepnote Notebooks:

- Steps taken to clean and manipulate the data can be found here.
- For information/visualization on phenophases and their relationship to year and climate index, please visit our notebook.

Data Sources

US Climate Extremes Index (CEI)

Short description:
Summarizes maximum and minimum temperature, and daily precipitation to provide a single number for each variable indicating the percentage of land that is in extreme conditions (either below or above normal for that area). The index is available for the continental United States and nine subregions. Annual, seasonal (Spring, Summer, Fall, Winter), and cold/warm season time frames are available. Time periods run from 1910-2022. We will download the CEI temperature and precipitation values.

Estimated size:
2,352 records

Format:
CSV

Location:
https://www.ncei.noaa.gov/access/monitoring/cei/graph

Access Method:
Download csv *by choosing Region (Northeast), Period (Annual, Spring, Summer, Fall, Winter, Cold Season, and Warm Season), and Indicator (Extremes in Maximum Temperature (Step 1), Extremes in Minimum Temperature (Step 2), and Extremes in Days with/without Precip (Step 5) and then Access Data).*

USA-NPN National Phenology Network

Short description:
The USA National Phenology Network contains observational data for a large range of plants and animals across the US. THe data are collected from scientific observations as well as citizen scientists. We will be looking at individual plant phenometrics using a subset of their data which includes species id info, location of the observed plant, and the dates of the plant phenophases, such as the date of the first leaves to grow after winter and the end of the flowering stage. The time period ranges from 1950 to 2022.

Estimated size:
350,264 records

Format:
CSV

Location:
https://data.usanpn.org/observations/

Access Method:
Download csv *by choosing Data Type:Individual Phenometrics; Start Date: 1/1/1950; End Date: 12/31/2022; States: AL, CT, DE, FL, GA, MA, MD, ME, NC, NH, NJ, NY, PA, RI, SC, VA, VT; Species: select all; Phenophase Categories: Flowers, Leaves, Fruits, Pollen cones, Needles, Seed cones; Additional Output Fields: Plant Nickname, Phenophase Category*

Data Cleaning & Manipulation Methods - Climate Data

Reading In Data

Our first dataset was climate data, which contained CSV files for each region, season, and climate index. Data was read in for one region at a time. For each region, the CSV's were put into a list of 4 lists, one for each season which contained the 3 climate indexes, then a for loop was used to read each in. Each CSV file had the same format containing three columns, year, much above normal, and much below normal. The unique feature was the names of the CSV files which were in the following format, region_climate index name_season. When reading in each file, the file name was split by _ and the climate index name and season were used to distinguish columns in the final DataFrame. For example, the climate index name was added to the much above normal and much below normal columns to distinguish individual climate index names for each season.

Combining Seasonal and Regional Data

After each seasonal list was read in, the DataFrames were horizontally concatenated to assign a season column using the season part of the CSV file name. After all season DataFrames were read in, they were vertically concatenated to form a finished DataFrame for each region. There were duplicated columns for the year that were dropped using Pandas duplicated function where any columns that were not duplicates were kept. After we had a DataFrame for each region, they were concatenated together to form our final climate DataFrame. (See section 1)

Data Cleaning & Manipulation Methods - Phenological Data

Reading in Data

We filtered to only columns and phenophases we were interested in and shortened the phenophase names using split within a lambda function. The USA-NPN recognizes the historical and current terminology of phenophase observations, so the same phenophases were referenced by different names depending on the year. The historical terms “First leaf”, “First bloom”, “Full bloom”, and “End bloom” were kept in our analyses so the analogous modern terms “Breaking leaf buds”, “Open flowers”, “Full flowering”, and “End of flowering” were changed to stay consistent. The data contained “First Yes (Month/Day/Year)” as individual columns when referring to a phenophase observation. These columns were combined into one date using Pandas to_datetime. A lambda function was used to strip only the month and date out of the full date as the first date column. A region column was created using Pandas where and isin. If the state was not in the Northeast region list of states then it was assigned Southeast. Data were plotted for each observation on a map of the Eastern United States to see the distribution for each genus. Due to the majority of the data being lilac (genus Syringa) data in the Northeast, the DataFrame was then filtered for this data. (See section 2.2)

Pivoting Data

A pivot table was constructed to transform the data for phenophases and the first dates associated with those phenophases to a first date and First_Yes_DOY for each phenophase. The columns were renamed and the multi-level index was dropped. A season column was added to this data based on the first bloom date using Pandas apply and a lambda function. This function was used to assign Spring as the season if the bloom date was null and for all other dates, the month day was split into month only which was looked up in a dictionary of months and seasons to return the correct season.(See section 2.3)

Data Cleaning & Manipulation Methods - Final DataFrame

Final Variables Used:

Climate Data:
- **Region:** Northeast
- **Time frame:** Seasonal (Spring, Summer, Fall, Winter)
- **Variables:**
  - Maximum Temperature
  - Minimum Temperature
  - Days With/Without Precipitation

Phenology Data:
- **Region:** Northeast
- **Plant Species:** Lilac (Genus: Syringa)
- **Variables:**
  - First Leaf date
  - First Bloom date
  - Full Bloom date
  - End Bloom date

Final DataFrame

To get the final DataFrame used for analysis, the climate data were merged with the phenology data using an inner join on the Year, Region, and Season columns. The season the phenological observation of "first bloom" occurred is the season that was chosen to join on. This allowed us to investigate the climate of the main growing season of the plant. (See section 3)

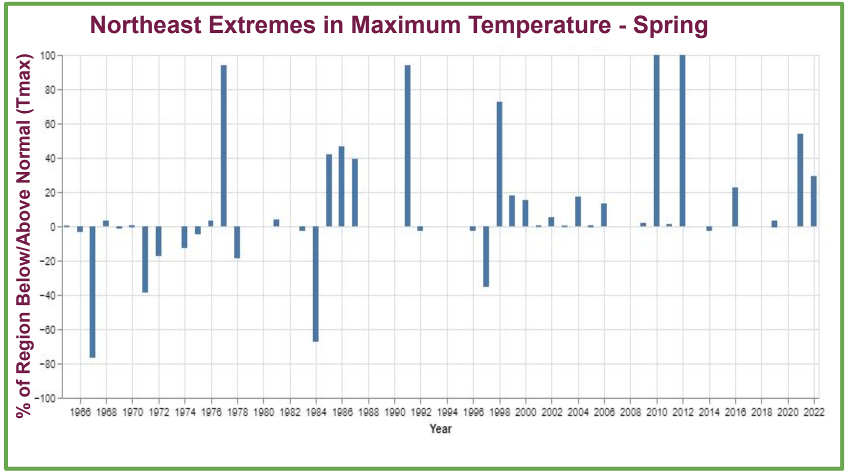
Missing Values

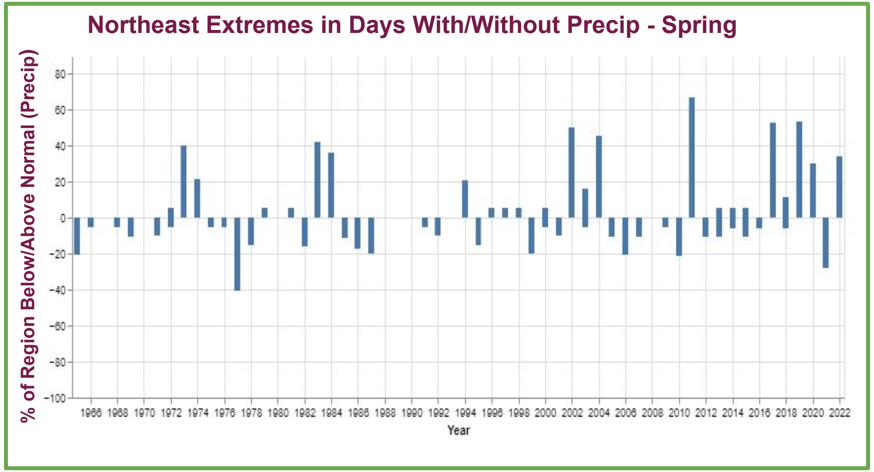
Each row in the final DataFrame represents one plant and its observations in that year, along with the spring climate indices. Not all plants had all phenophases observed. Due to how we pivoted the phenology data there are missing values within each of the phenophases, but these rows cannot be dropped because they would eliminate other phenophase dates. They were handled by only looking at the data we have for each phenophase.

Exploratory Data Analysis

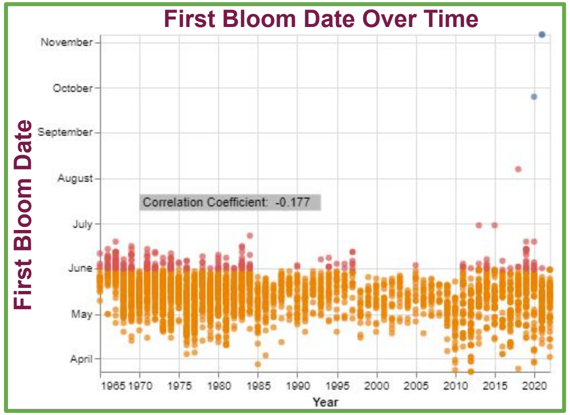
There is a trend of higher percentages of the land in the Northeast having both maximum temperatures and precipitation higher than normal, indicated by values greater than zero. There has also been a drastic decrease over time in the percentage of land with maximum temperatures lower than normal, represented by any values below zero. To simplify and prevent repetition, the variables 'maxT much above normal' and 'precip much above normal' may be used for further analysis, while the below normal measurements will not be used as ‘maximum temperature below normal’ has not occurred in any percentage of the land since that late 1990’s and only a small percentage of the land has had ‘precipitation much below normal’.

Higher temperatures can lead to higher evaporation and thus higher precipitation, so these variables may be correlated, however the plants may respond differently to each outcome. For example, a plant may be able to tolerate higher temperatures but not increased rainfall or vice versa. (See section 5.1.2)





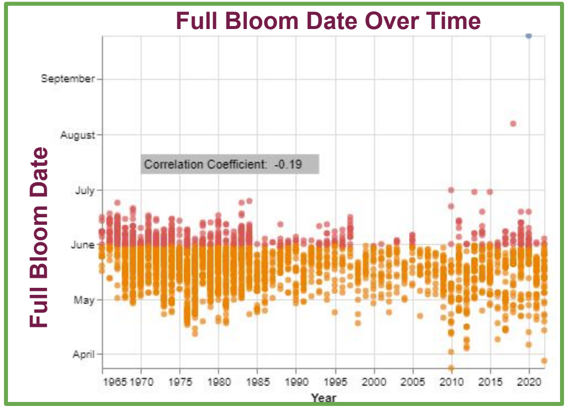
Analysis on Combined Dataset

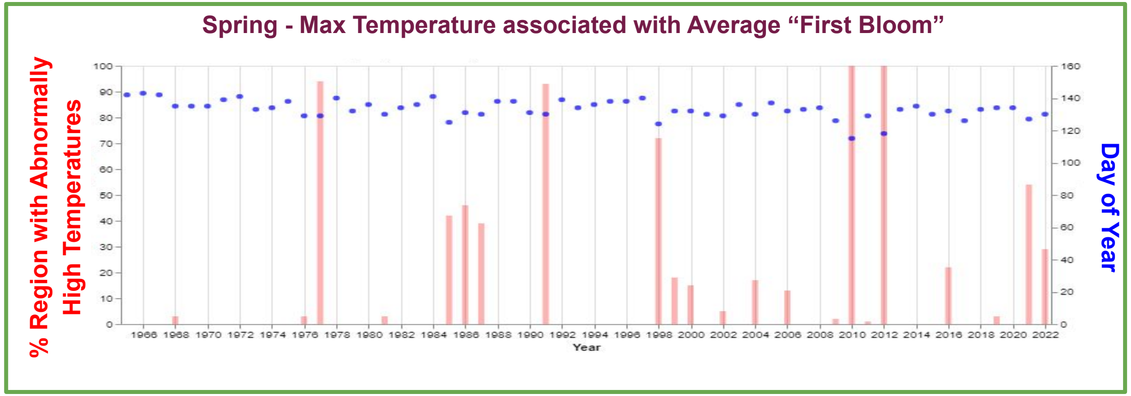


The two charts to the left show the change in first bloom date and full bloom date over time. The orange points represent observations in the spring, while red points are for summer observations and blue points are for fall observations. There appears to be a slight trend towards earlier bloom and full bloom phenophases. This is quantified with the correlation coefficient of -0.177 for first bloom and -0.195 for full bloom. (See 5.2.1)

Our investigation into possible reasons why this might be occurring are to begin looking at the maximum temperature and precipitation extremes.

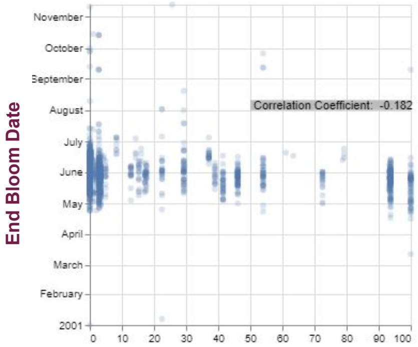
When we plot the maximum temperature springtime extremes (shown below) we can see patterns of early first bloom dates and years with extreme daytime spring temperatures. (See 5.3.3)





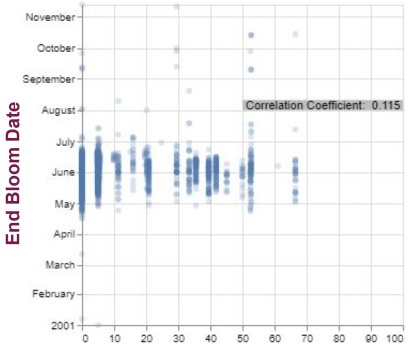
Analysis on Combined Dataset

End Bloom vs Abnormally High Temperatures



% of Region w/ Abnormally High Temperatures

End Bloom vs Abnormally High Precipitation



# of Days w/ Abnormally High Precipitation

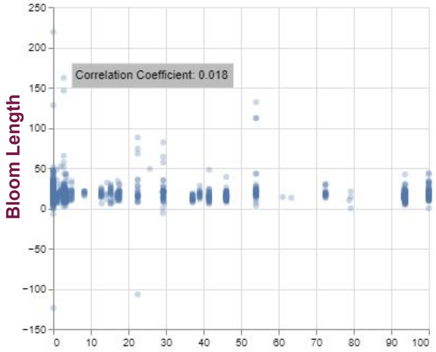
We found that end bloom dates were negatively correlated with abnormally high temperatures but positively correlated with abnormally high precipitation. One possible reason for this could be that high temperatures could stress plants and shorten their lifespans. The increased precipitation may be within a plants tolerable range and instead help extend their lifespan.

In contrast, we found that first bloom was somewhat correlated with abnormally high temperatures (-0.22) but had a rather weak correlation with abnormally high precipitation (-0.08). (See 5.3.1)

Bloom length is the number of days between first bloom and end bloom. First bloom and end bloom are correlated relatively similar with abnormally high temperatures, so it is not surprising that there is not much of a correlation between high temperatures and bloom length.

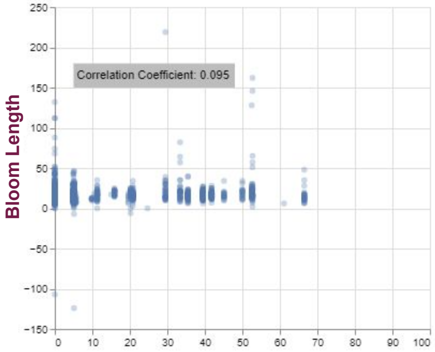
First bloom does not have a correlation with precipitation (-0.015), so combined with end bloom's weak correlation with precipitation, it is also not surprising that there is a weak correlation between bloom length and precipitation.

Bloom Length vs Abnormally High Temperatures



% of Region w/ Abnormally High Temperatures

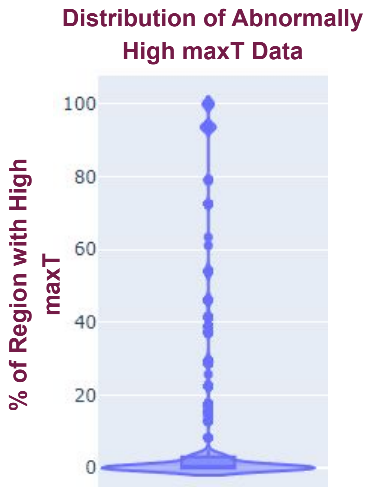
Bloom Length vs Abnormally High Precipitation



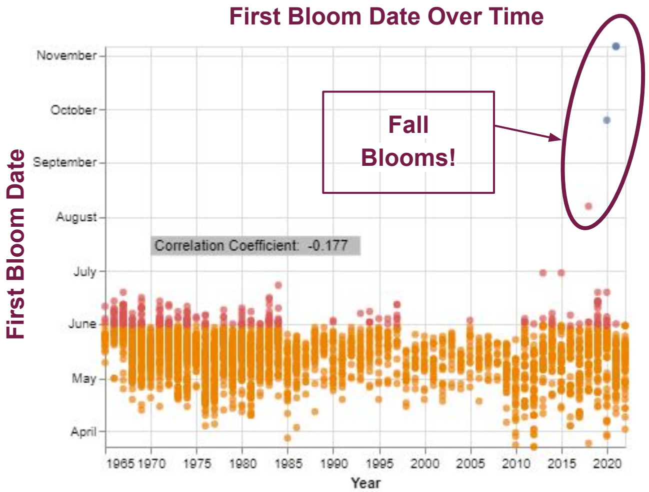
# of Days w/ Abnormally High Precipitation

Weaknesses in the Data

Trying to visualize correlations between phenophase dates and the climate indexes proved to be a challenge in which we accepted defeat. There is a massive amount of data with climate indexes of 0, while the abnormal readings are few and far between. This led us to look into the possibility of scaling or sampling. Unfortunately, scaling does not fit the situation, and we fear the data above 0 is too interspersed to effectively support sampling. Because abnormal indexes have become more common in recent years than earlier years, future data collection could increase the reliability of data and enhance our ability to transform it.



In addition, errors in data recording impact the reliability of the data. For example, some observed individuals had multiple first blooms but only a single end bloom while others had end blooms that occurred prior to first bloom. This could impact the calculation of bloom length. The duplicate errors were somewhat mitigated by using the earliest date for each phenophase. Negative bloom dates from either entry errors or the result of taking the minimum date per phenophase were not removed in order to show and acknowledge this weakness.



We noticed the later blooming dates, but these might not necessarily be an error…With dry, warm fall weather, some plants, such as lilacs, have been known to bloom again!

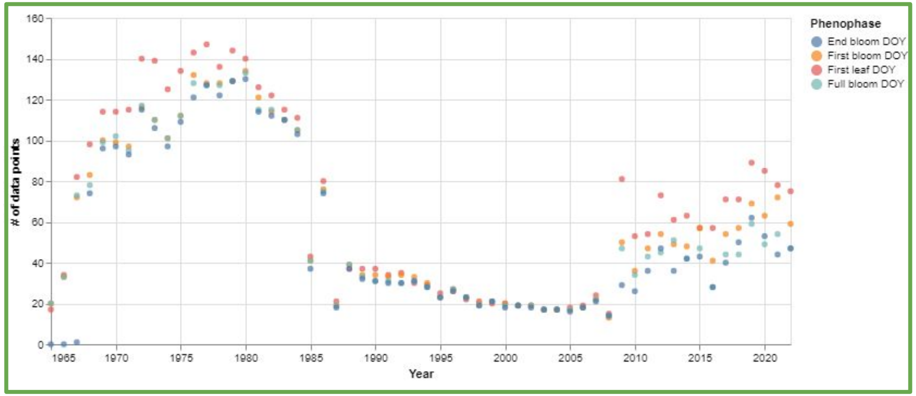
https://yaleclimateconnections.org/2020/12/what-causes-flowers-to-bloom-in-the-fall/

Weaknesses in the Data

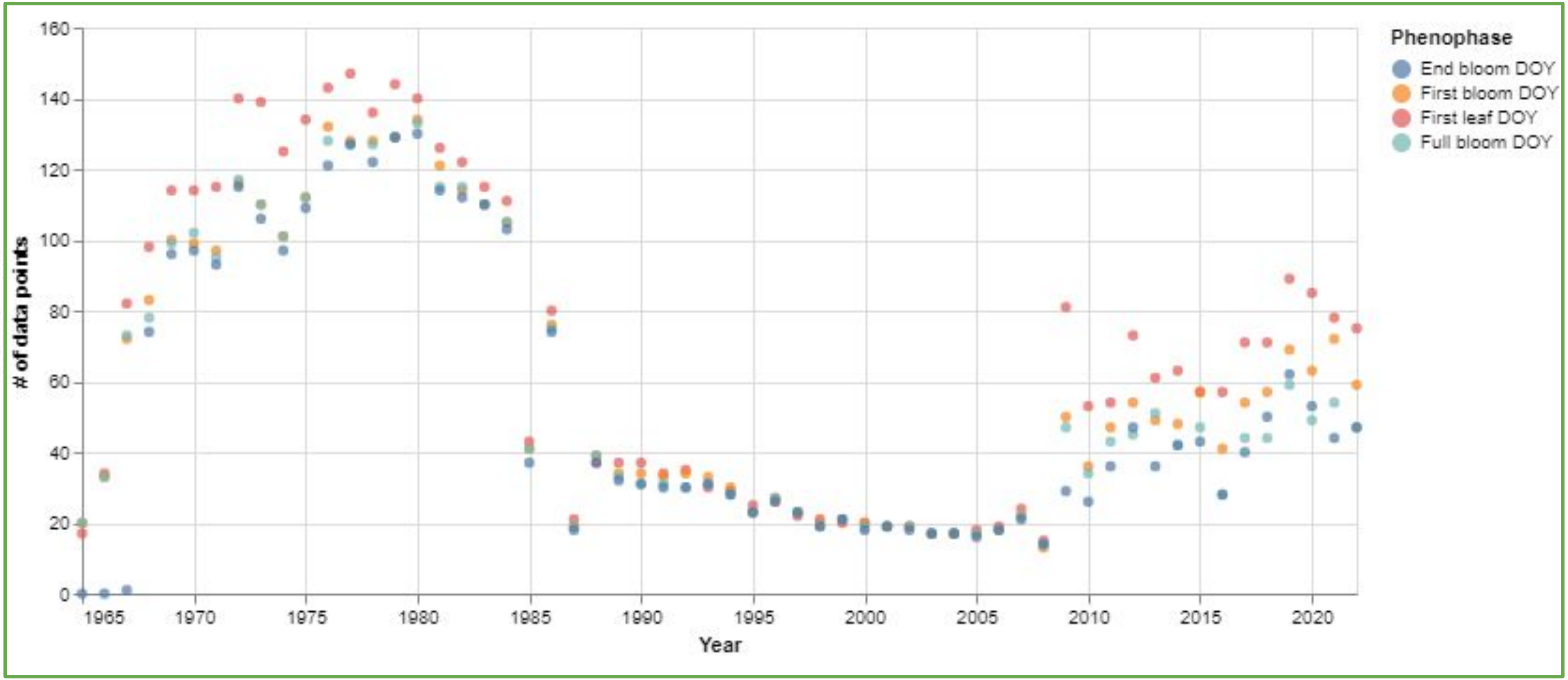
Distribution of Lilac (Genus Syringa) & Honeysuckle (Genus Lonicera) in the Northeastern & Southeastern US with Phenological Observations

In creating the above map, it became obvious that **our analysis should focus on the observation-rich Northeast and focus on Lilacs.** Note the point in the middle of the US. The point is listed as being in ME but the coordinates are in Kansas. It is possible that coordinates were not available and the data was just populated with the center of the country instead. Whatever the reason, it is a weakness in the reliability in the data quality.

Collection of Lilac (Genus Syringa) Phenological Observations in the Northeast



There is a large gap from ~1985 to ~2008 where only about 20-40 individuals were observed each year. Official observations of lilac and honeysuckle species began in the US in the late 1960s to help with agricultural forecasts. That program ended, for most site locations, in the late 1980s. In 2009, Nature's Notebook, a program that consists of professional and citizen scientists, restarted and expanded the effort to observe both plant and animal phenologies (Rosemartin, et al., 2015). Data during these years may not be representative of the entire population and are more likely to be biased towards those individuals. (See Section 4)



Phenophase
End bloom DOY
First bloom DOY
First leaf DOY
Full bloom DOY
# of data points
Year
0
20
40
60
80
100
120
140
160
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020

Statement of Work

Andrea:

- Phenological data acquisition
- Data analysis/exploration
- Final report
- Project management

Linda:

- Climate data acquisition
- Data analysis/exploration
- Final report
- Meeting setup

Jacqueline:

- Data cleaning/manipulation
- Data analysis/exploration
- Project report outline
- Final report

Our team communicated exceptionally well. We were all communicative and available from beginning to end and were open with our thoughts and feedback. We all contributed helpful ideas that furthered our project and allowed us to investigate the data in more ways than if we had worked alone. Future collaboration could be improved through additional meetings that were pre planned to fit everyone's schedules and provide ample time for adjustments.

Next Steps:

Climate data was joined using the season of the first bloom. Given the correlation between increased precipitation and end bloom, which is later in the season than first bloom, it may be interesting to look at correlations between the phenophases and the *prior* season.

In addition, future analyses should attempt to find more appropriate ways to manipulate the indexes to allow for clearer visualizations when investigating the correlations with phenophases.

Finally, expanding the data set to include more regions or species may reveal additional trends.

References:

Rosemartin, A., Denny, E., Weltzin, J. *et al.* Lilac and honeysuckle phenology data 1956–2014. *Sci Data* 2, 150038 (2015). https://doi.org/10.1038/sdata.2015.38