Predicting Sakura Florescence in Japan with a Dual-Model Approach*

Estimating Temperature and Forecasting Bloom Duration for Cherry Blossoms

Shaotong (Max) Li

November 17, 2024

   This paper presents a dual-model approach to predict the duration of cherry blossom blooms in Japan: The first model predicts bloom duration using temperature, while the second model estimates temperature based on latitude, month, and date. By analyzing historical and modern data, we provide insights into how latitude and seasonal timing affect temperature and bloom duration. Our findings indicate that this dual-model approach enhances our understanding of climatic effects on cherry blossom florescence, highlighting potential impacts of climate change on one of Japan's most cherished natural events.

Table of contents

---

*Code and data are available at: Sakura_Florescence_Prediction.

1 Introduction

Cherry blossoms, or sakura, are a powerful symbol of Japan, representing both the fleeting beauty of nature and a cherished cultural heritage (Wikipedia contributors 2024). Each spring, the blooming of cherry blossoms is celebrated across the country, marking a time of renewal and festivity (Wikipedia contributors 2024). However, the timing and duration of these blooms are highly sensitive to climatic conditions, which makes predicting sakura florescence increasingly important in the face of climate change (Ocko 2024). Understanding how environmental factors, such as temperature, latitude, and seasonal timing, influence bloom duration can help forecast these events and provide insight into broader ecological changes (Ocko 2024).

This paper aims to address the challenge of predicting cherry blossom bloom duration by using a dual-model approach. The first model predicts the duration of cherry blossom blooms using temperature. The second model estimates temperature based on environmental variables such as latitude, month, and date, providing a detailed understanding of local climatic conditions. By combining these two models, we seek to fill the gap in understanding how temperature fluctuations driven by geographic and temporal factors affect sakura florescence.

The estimand in this study is the predicted duration of cherry blossom (sakura florescence). Specifically, Model 1 predicts the duration of the bloom using temperature, while Model 2 estimates temperature as a function of latitude, month, and date. This dual-model structure

allows us to precisely quantify the effect of temperature as a mediating factor influenced by geographical and temporal variables on bloom duration. This approach allows us to quantify the relationship between environmental conditions and the duration of cherry blossom blooms.

We utilized both historical and modern phenological data to ensure the robustness of our predictions. Our analysis highlights how latitude and seasonal factors contribute to temperature variation, which in turn affects the timing and duration of the bloom. The results demonstrate the potential for these models to improve the accuracy of bloom forecasts, particularly in the context of changing climatic patterns. This research contributes valuable knowledge to the fields of phenology and climate science, and provides practical implications for tourism, cultural events, and ecological conservation in Japan.

The remainder of this paper is structured as follows: Section 2 discusses the data sources and preprocessing methods. Section 3 details the dual-model approach, including the temperature estimation and bloom prediction models. Section 4 presents the results, followed by a discussion in Section 5 on the implications of our findings. Finally, Section 6 concludes with insights into future research directions and the broader impact of climate change on cherry blossoms.

2 Data

In this project, we used R(R Core Team 2023) and several R packages for data processing, analysis, and visualization. Specifically, tidyverse (Wickham et al. 2023c), arrow(Richardson et al. 2023), here(Müller 2023), ggplot2(Wickham et al. 2023b), patchwork(Pedersen 2023), sf(Pebesma et al. 2023), rnaturalearth(South 2023a), and rnaturalearthdata(South 2023b) were used for data processing, geospatial analysis, and visualization. ggthemes(Arnold 2023) was utilized to apply thematic elements to plots, while dplyr(Wickham et al. 2023a) was key for data manipulation tasks. For dynamic report generation, knitr(Xie 2023) and kableExtra(Zhu 2023) were used, providing enhanced formatting for outputs. Together, these packages enabled efficient data cleaning, analysis, and visualization throughout the study.

2.1 Overview

The data used in this study includes three primary datasets: historical cherry blossom bloom records, modern bloom records, and temperature data. The historical dataset contains records of cherry blossom bloom dates across multiple decades for the Kyoto region, providing a long-term view of bloom trends in this specific area of Japan. The modern dataset contains recent bloom dates from various regions across Japan, reflecting current climatic conditions and bloom dynamics. Lastly, the temperature dataset includes temperature readings for various locations in Japan, which are essential for estimating temperature-related effects on bloom duration.

To ensure comparability, only bloom records that met our defined geographic and seasonal criteria were retained. This filtering process ensured that only reliable and relevant data

points were used in the final analysis, thus maintaining the accuracy and robustness of our predictive models. Finally, the cleaned datasets were merged based on common identifiers such as location and time to create a comprehensive dataset for modeling sakura florescence.

The final combined dataset, after merging and filtering, consisted of 10,456 rows with multiple key attributes, including latitude, month, date, and temperature, serving as predictors for bloom duration. By integrating both historical and modern data, our analysis captures both long-term trends and recent shifts in bloom timing, providing insights into the effects of climate variability and change on cherry blossom phenology.

2.2 Measurement

The transformation of real-world phenomena into dataset entries involved detailed documentation of cherry blossom bloom dynamics. Historical sakura data was derived from records such as diary entries and observations from meteorological agencies, specifically focusing on the Kyoto region. Bloom dates were reconstructed based on these historical accounts dating back to 812 CE, supplemented by temperature data that provided context for climatic influences over centuries.

Modern cherry blossom data was collected by the Japan Meteorological Agency (JMA), including flowering dates and meteorological information across Japan since 1953. This dataset encompasses latitude, longitude, bloom dates, and temperature data, offering a comprehensive view of bloom dynamics under current climatic conditions. Temperature readings were matched with bloom records using geographic and temporal identifiers to ensure consistent analysis.

Data quality checks included removing incomplete records, standardizing date formats, and verifying bloom dates against multiple sources. This ensured that the final dataset was robust and suitable for modeling, enabling reliable predictions of bloom duration based on environmental factors.

2.3 Outcome Variables
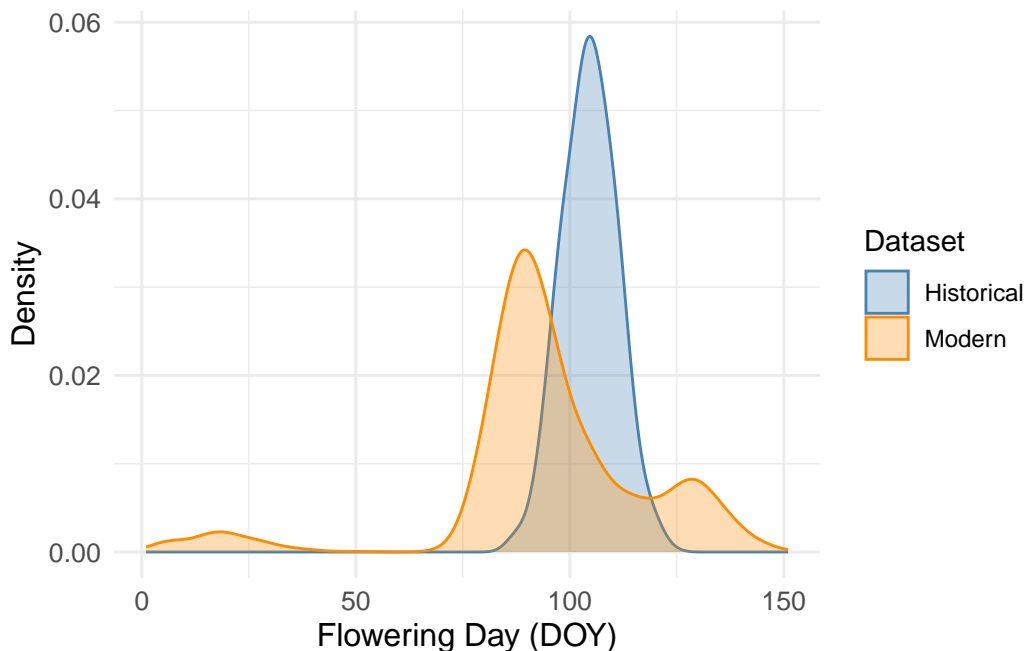
2.3.1 Florescence

Figure 1: Sakura Florescence: Historical vs Modern Data

The variable 'florescence' represents the duration of cherry blossom blooming, defined as the number of days during which cherry blossoms are in full bloom. In the historical dataset, which focuses on the Kyoto region, florescence data shows a narrower distribution, indicating relatively stable bloom timings over centuries. In contrast, the modern dataset includes bloom data from multiple regions across Japan, resulting in a wider distribution that reflects greater variability due to regional climatic differences.

Figure 1 illustrates the density distribution of the day of year (DOY) for full bloom, comparing historical and modern datasets. The historical data (Kyoto) exhibits a peak around a consistent time frame, whereas the modern data shows more dispersed bloom times, likely driven by regional climatic variations and recent temperature changes.

As the outcome variable in our modeling approach, florescence is crucial for understanding bloom dynamics. It allows us to assess the impact of temperature and other environmental factors on the timing and length of cherry blossom blooms, providing insights into the effects of climate variability.

The distribution of florescence in the historical data shows relatively stable bloom timings, with less variability compared to the modern dataset. This stability is largely due to the

consistent climatic conditions in the Kyoto region over centuries. In contrast, the modern data presents more pronounced regional variations in bloom duration, influenced by diverse environmental conditions across Japan. As the outcome variable in our modeling approach, florescence serves as the key measure of bloom dynamics, enabling us to evaluate how temperature and other environmental factors affect the timing and duration of cherry blossom blooms in both historical and modern contexts.

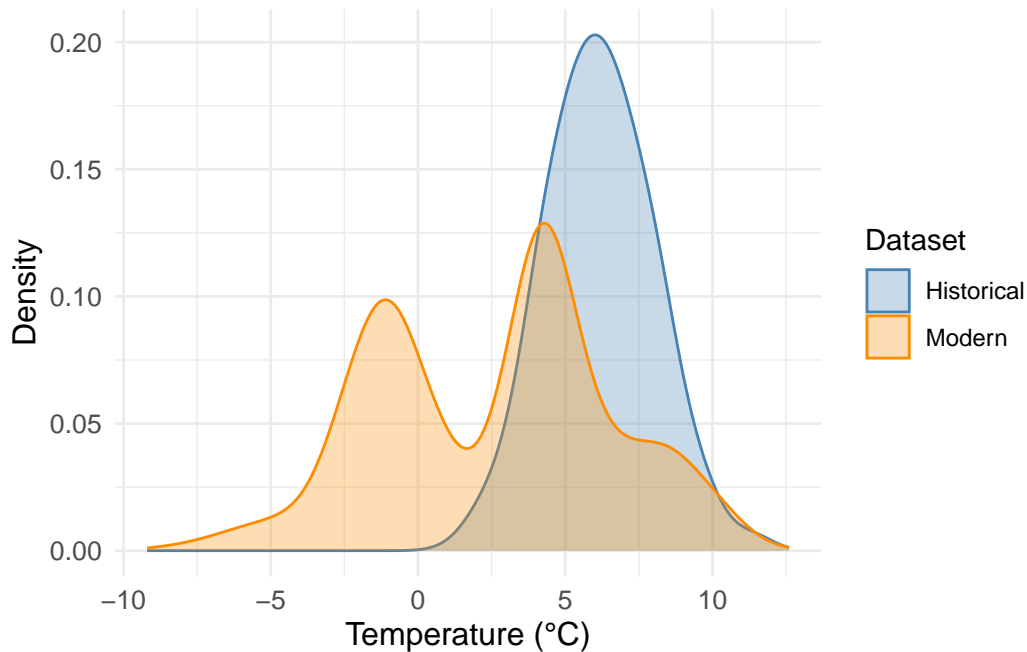## 2.4 Predictor Variables

2.4.1 Temperature

Figure 2: Temperature: Historical vs Modern Data

Temperature plays a crucial role in determining the timing and duration of cherry blossom blooms. Figure 2 shows the density distribution of temperatures in historical and modern datasets, highlighting significant differences between the two periods. The historical dataset, primarily focused on the Kyoto region, shows a relatively narrow temperature range centered around moderate values, while the modern dataset presents a broader range with lower temperatures being more prevalent. This difference in temperature distributions likely reflects the increased regional variability and recent climatic changes observed in Japan.
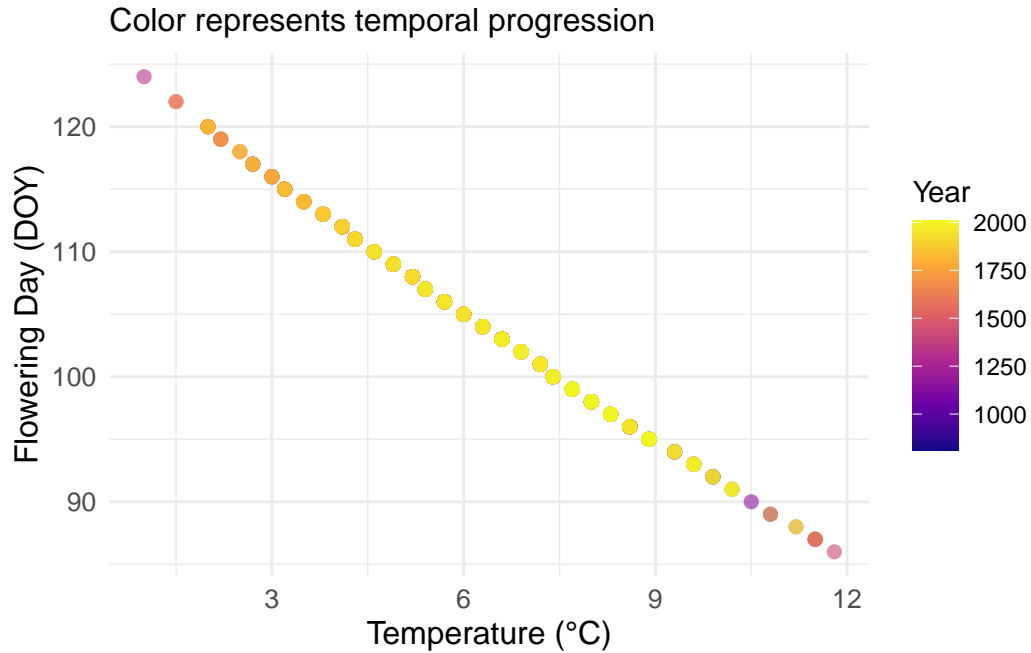
Figure 3: Temperature VS Sakura Florescence

The strong linear relationship between temperature and the day of year (DOY) for full bloom is evident in Figure 3. As temperatures increase, the DOY for blooming decreases, indicating that higher temperatures lead to earlier blooming. This relationship is consistent across both historical and modern datasets, underscoring the critical impact of temperature on bloom timing. The temporal progression in Figure 3, represented by the color gradient, illustrates how this relationship has persisted over time, even as overall climatic conditions have changed.

In our modeling approach, temperature serves as a key predictor for florescence, allowing us to quantify how fluctuations in temperature directly affect the timing and duration of cherry blossom blooms. By examining the historical and modern temperature distributions, we can better understand the impacts of climate variability and identify trends that are crucial for predicting future bloom behavior.

2.4.2 Latitude

Figure 4: Latitude Distribution

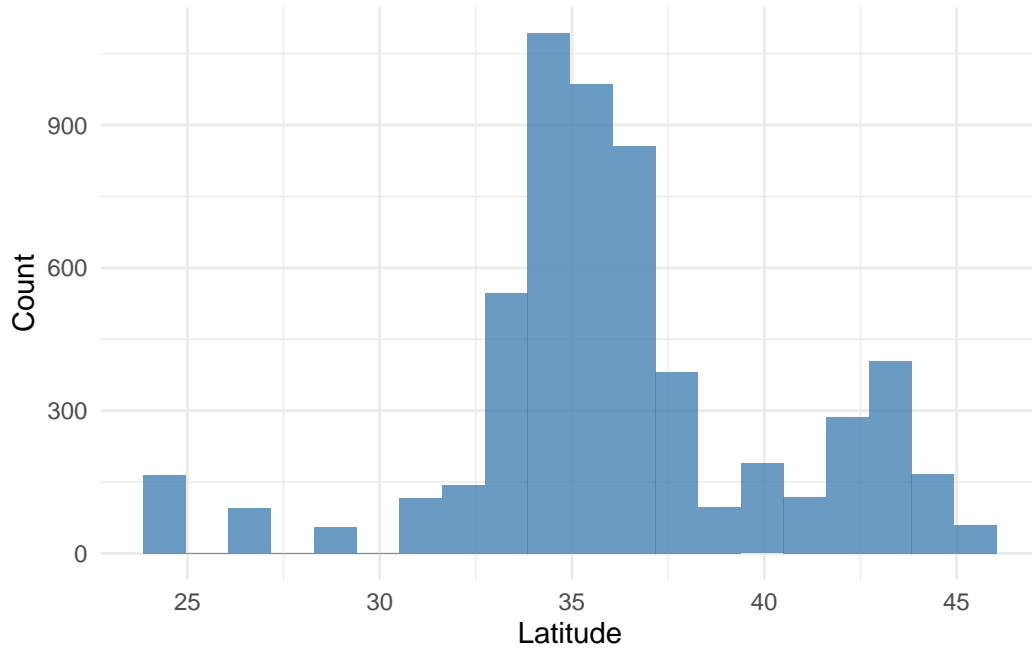Figure 4 shows the distribution of latitudes for the locations included in the dataset. The majority of observations are concentrated between latitudes 34° and 38° N, which represents the regions most commonly associated with cherry blossom observations in Japan. There is a smaller number of observations at lower and higher latitudes, reflecting the geographic range of sakura coverage in the country.
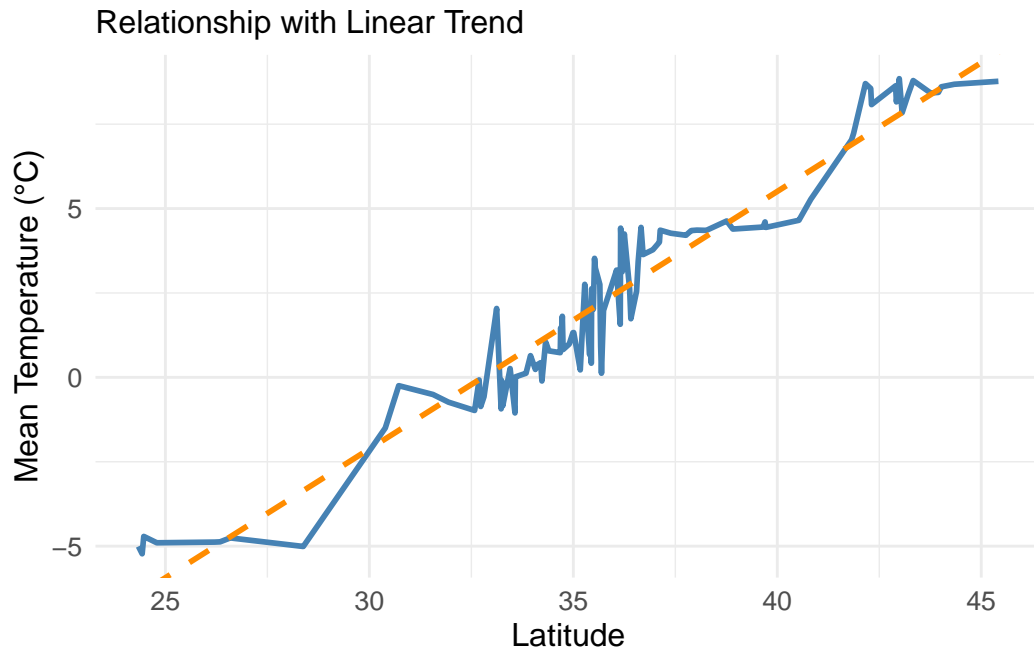
Figure 5: Latitude VS Temperature

Figure 5 illustrates the relationship between latitude and mean temperature, showing a clear positive trend. As latitude increases, so does the average temperature. The dashed orange line represents the linear trend, indicating a consistent increase in temperature as we move towards the northern regions of Japan. This relationship is essential for understanding how geographic position influences local climatic conditions, which in turn affects bloom timing.

Figure 6: Average Temperatures Across Japan

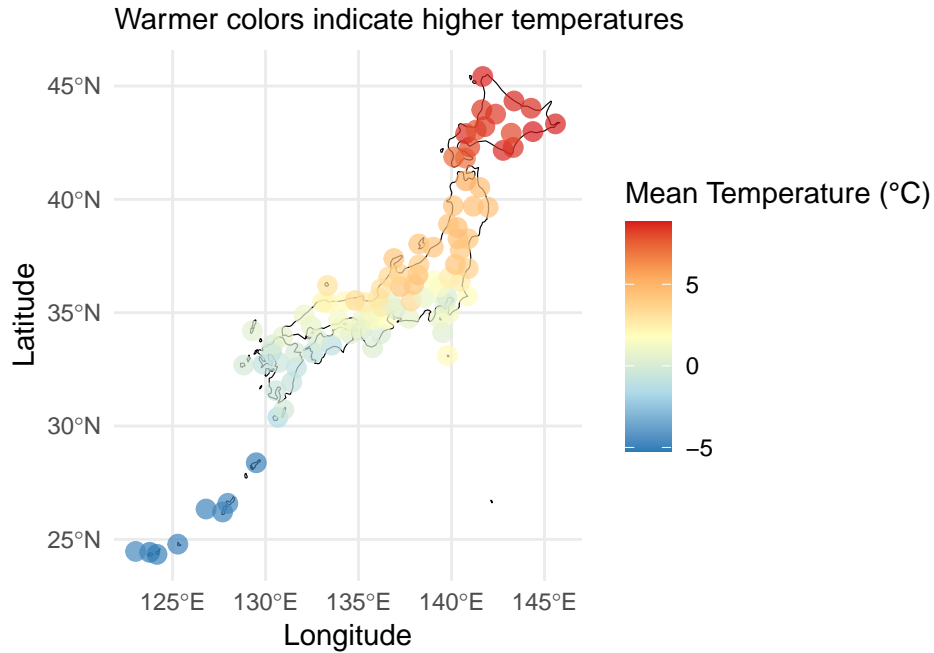Figure 6 presents a geographic distribution of mean temperatures across Japan, with warmer colors indicating higher temperatures. This visualization highlights the spatial temperature variability across the country, where northern regions tend to have higher average temperatures compared to southern regions. The color gradient helps to illustrate how temperature conditions vary as a function of both latitude and longitude.

In summary, latitude has a significant influence on temperature, which directly impacts the timing of cherry blossom blooms. The relationship between latitude and temperature is well-captured by Figures 4, 5, and 6, which collectively demonstrate that regions at higher latitudes tend to experience higher average temperatures, contributing to earlier bloom periods. These insights are crucial for understanding the geographic factors that affect sakura florescence and for improving the accuracy of predictive models.

2.4.3 Time of Blossom
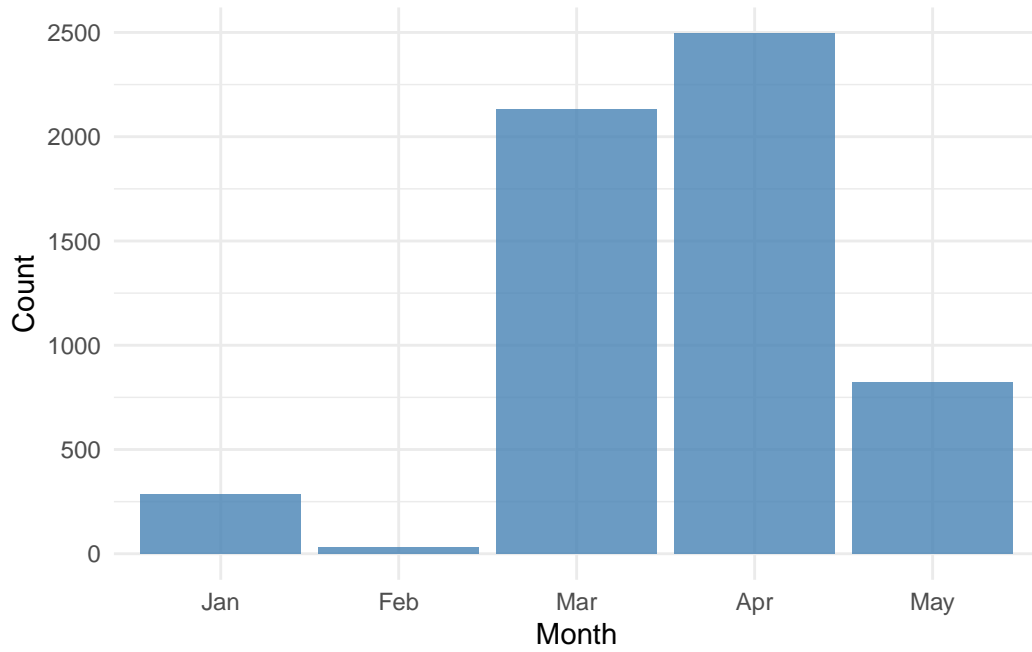
Figure 7: Monthly Distribution of Flower Date

Figure 7 shows the monthly distribution of flower dates, with the majority of cherry blossoms blooming in March and April. This aligns with the typical cherry blossom season in Japan, which occurs in early spring. There are fewer occurrences of blooms in January, February, and May, which indicates that bloom timings outside of this window are uncommon, reflecting the strong seasonal pattern of sakura flowering.
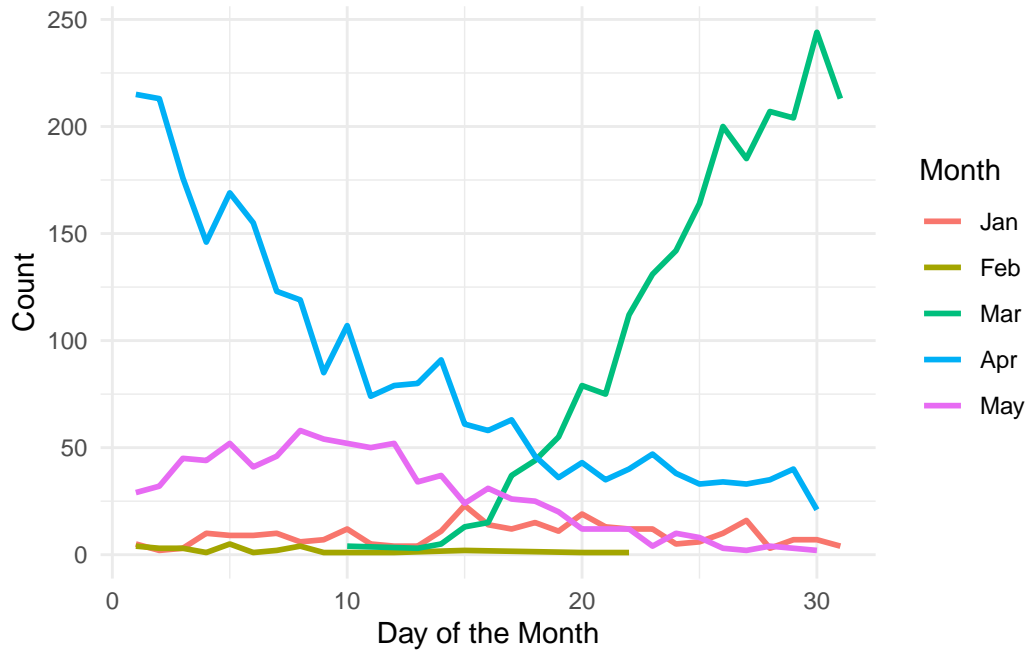
Figure 8: Flower Date Distribution by Month and Day

Figure 8 presents a more detailed view of the flowering dates by month and day. The peak bloom periods are concentrated towards the end of March and throughout April, with the highest counts occurring in late April. This detailed distribution highlights the specific bloom dates, emphasizing how the bulk of flowering happens within a narrow time frame in early spring, largely driven by favorable temperature conditions.
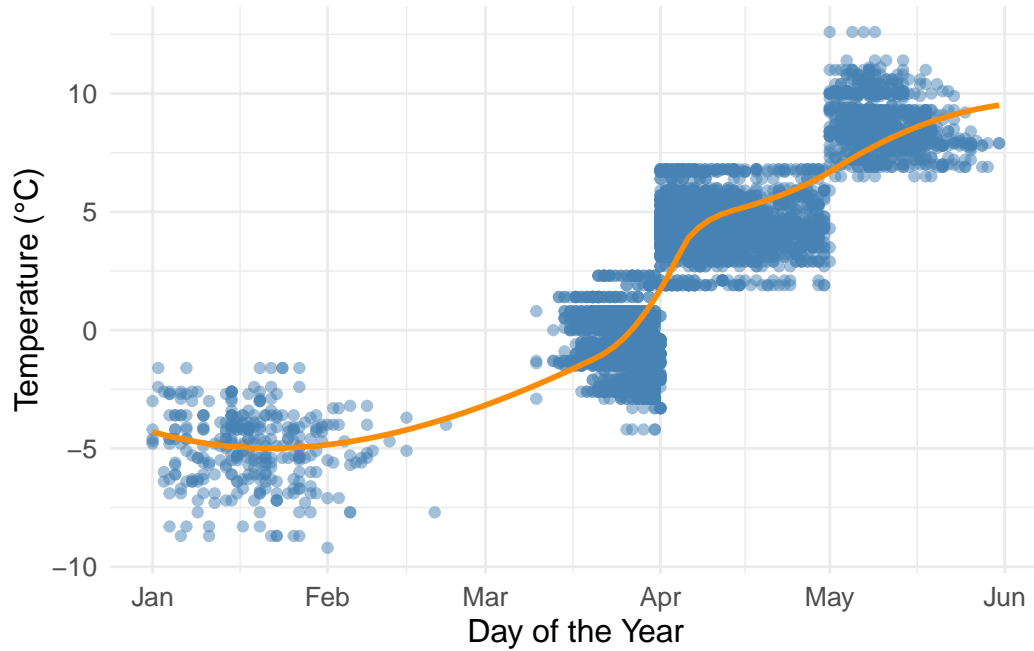
Figure 9: Temperature Distribution by Date

Figure 9 shows the relationship between day of the year (DOY) for blooming and temperature. The trend line suggests a potential linear relationship, where later dates are associated with higher temperatures. This trend indicates that as the year progresses, temperatures rise, which influences the timing of cherry blossom blooms. The scatter plot further emphasizes the variation in temperatures experienced during different bloom periods, illustrating how the blooming date affects the observed temperature.

Figure 10: Average Flowering Dates Across Japan

Figure 10 provides a geographic overview of average flowering dates across Japan. The color gradient represents the mean flowering DOY, with darker points indicating later flowering dates. This map illustrates how flowering dates vary geographically, with earlier blooms occurring in warmer, southern regions and later blooms in cooler, northern regions. This geographic variation in bloom dates underscores the influence of temperature gradients across Japan.

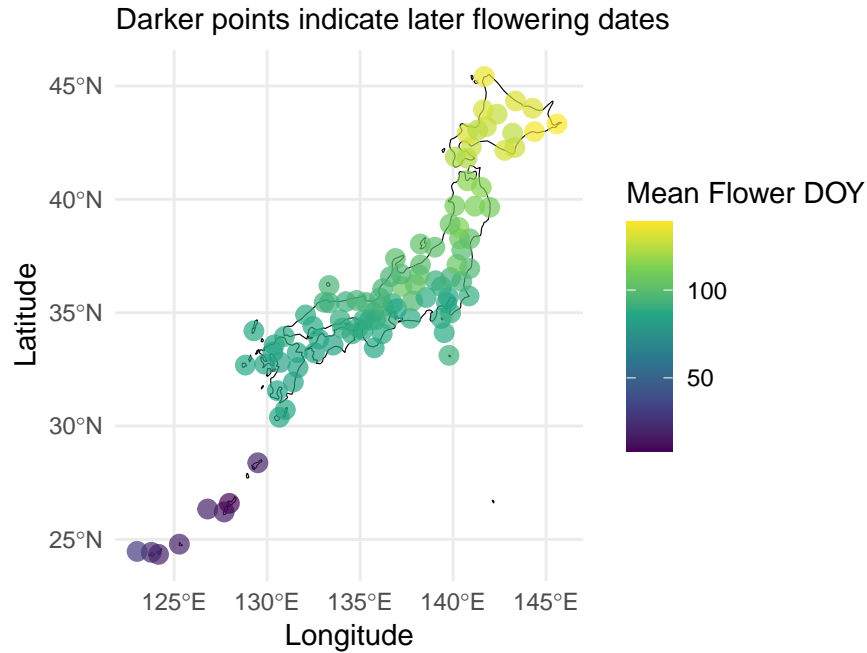In summary, the timing of cherry blossom blooms significantly affects temperature patterns. The majority of cherry blossoms bloom in March and April, which in turn correlates with observed temperature trends. Geographic variation also plays a significant role, with southern regions experiencing earlier blooms compared to northern regions. These insights are essential for understanding the temporal dynamics of sakura florescence and for enhancing the accuracy of predictive models that aim to understand temperature dynamics under varying bloom conditions. These insights are essential for understanding the temporal dynamics of sakura florescence and for enhancing the accuracy of predictive models that aim to forecast bloom timing under varying climatic conditions.

2.5 Excluded Variables

2.5.1 Year
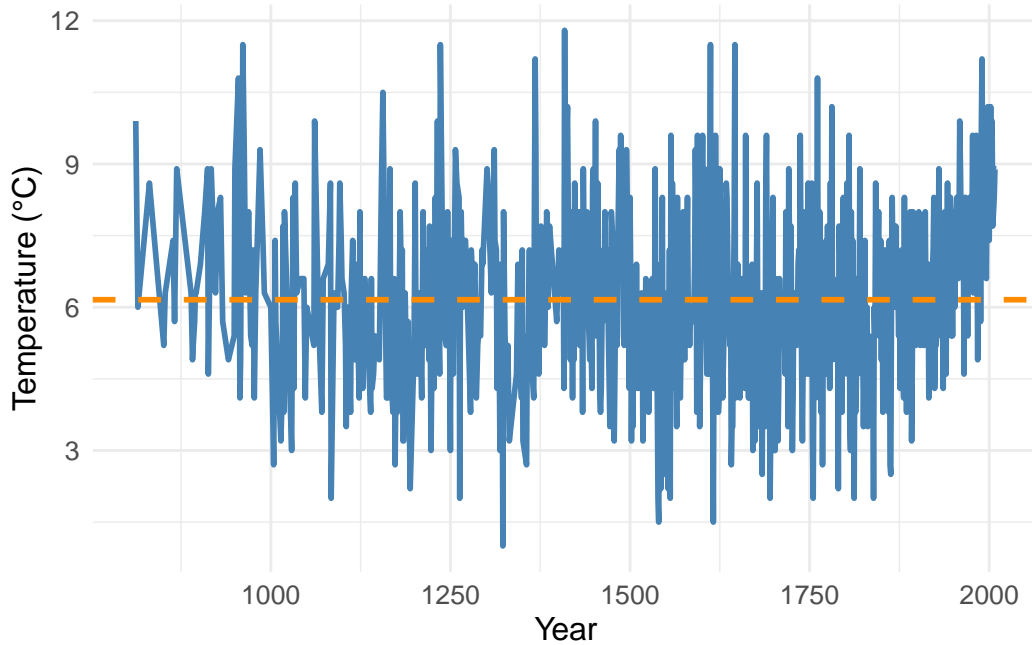
Figure 11: Temperature Trend Over Years

Figure 11 depicts the temperature trend over several centuries, highlighting the distribution of temperatures across different years. The orange dashed line represents the average temperature, providing a baseline for comparison. The plot shows considerable variation in temperatures from year to year, with no consistent linear trend. This lack of a clear relationship between temperature and year suggests that, over the long term, temperature fluctuations have been influenced by a complex interplay of factors beyond a simple time progression.

While there are periods of both high and low temperature variability, the data does not support a strong linear trend correlating temperature with time. This finding emphasizes that year-to-year temperature changes are not solely driven by temporal progression, but rather are subject to regional and potentially global climatic influences. Understanding these temperature trends is crucial for predicting future cherry blossom bloom timings, as temperature is a key determinant in bloom dynamics.

2.5.2 Longitude

Figure 12: Longitude Distribution

Figure 12 shows the distribution of longitudes for the locations included in the dataset. The majority of the observations are concentrated between 135°E and 141°E, representing regions across central and eastern Japan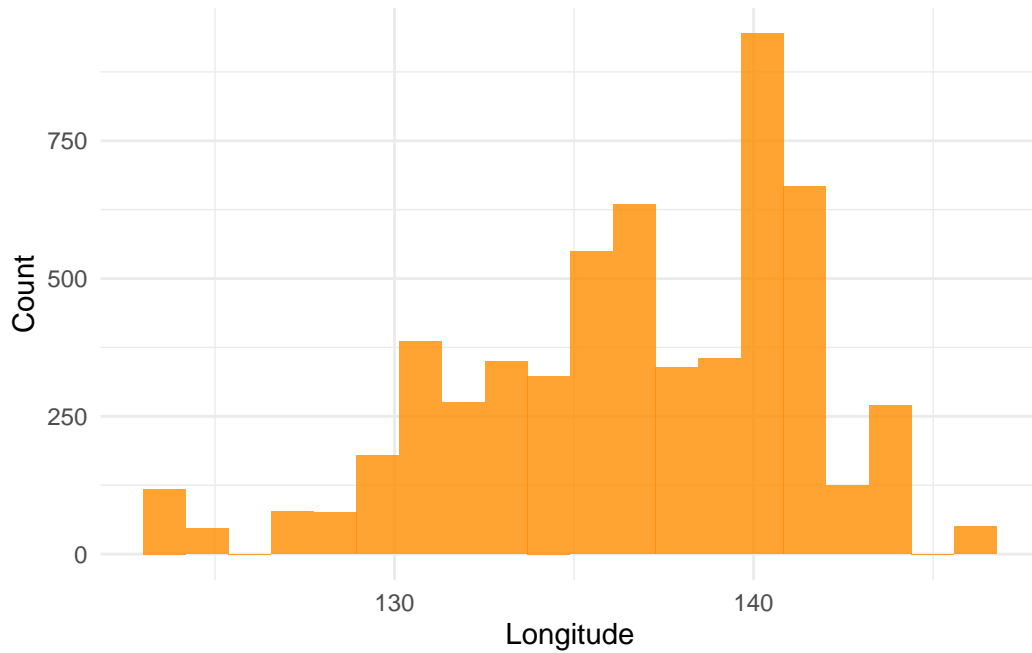. This distribution highlights the geographic focus of our study, which includes areas where cherry blossoms are commonly observed and where temperature data is readily available.
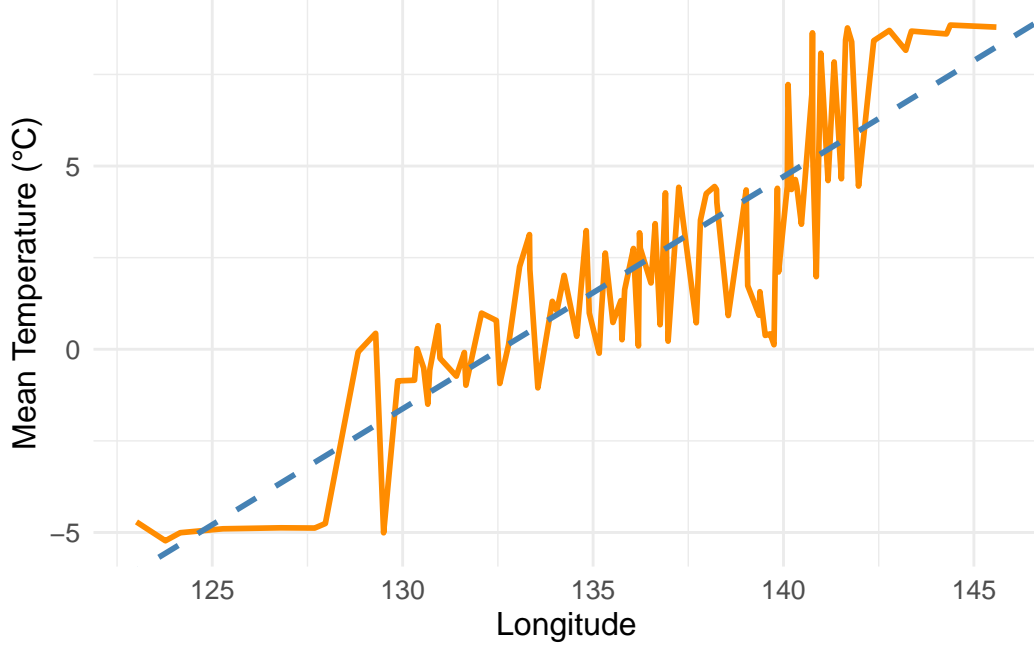
Figure 13: Temperature VS Longitude

Figure 13 illustrates the relationship between longitude and mean temperature. The dashed blue line represents the linear trend, indicating a positive correlation between longitude and temperature. However, compared to latitude, this relationship is weaker and exhibits more variability. The positive trend suggests that regions further east tend to have slightly higher average temperatures, though the variability indicates that other geographic and climatic factors are also at play.

In summary, while longitude does show some relationship with temperature, it is not as strong as the correlation observed with latitude. This weaker relationship and greater variability led us to exclude longitude as a predictor in Model 2. The rationale for this decision will be discussed in detail in Section 3, where we explain the selection of the most relevant variables for predicting temperature and bloom dynamics.

3 Model

3.1 Alternative Model

$$\text{Alternative Model: mean\_temp\_month} = \beta_1 \cdot \text{latitude} + \beta_2 \cdot \text{longitude} + \beta_0 \qquad (1)$$

where:

- latitude $\in [20, 50]$, representing geographical latitude in degrees (°N).

- longitude $\in [120, 150]$, representing geographical longitude in degrees (°E).

- β_1 and β_2 are coefficients for latitude and longitude, respectively.

- β_0 is the intercept term.

- Both latitude and longitude are numerical variables.

As shown in equation~1,the alternative model was developed to investigate whether latitude and longitude can simultaneously maintain a linear relationship with mean temperature. Specifically, this model aims to evaluate if both geographic coordinates contribute meaningfully to temperature estimation, thereby affecting the prediction accuracy for bloom timing.

Table 1: Alternative Model Summary

Variable	Model Summary	
	Estimate	P-Value
latitude	0.778101	<2e-16
longitude	-0.003319	0.806
R-squared	0.672200	
Adjusted R-squared	0.672100	

Table 1 provides the summary of the alternative model, showing the estimates and p-values for both latitude and longitude. The results indicate that while latitude has a strong and statistically significant relationship with temperature, longitude does not contribute significantly to the model (p-value of 0.806). This suggests that latitude is the more important predictor of mean temperature, which aligns with our earlier analysis showing the stronger linear relationship between latitude and temperature compared to longitude.

The findings from this alternative model indicate that incorporating longitude as a predictor does not substantially improve temperature estimation. Consequently, we excluded longitude in our final temperature estimation model (Model 2), focusing solely on latitude, month, and date as predictors. This decision is based on both statistical performance and model simplicity, aiming to create a model that is interpretable without sacrificing predictive accuracy.

3.2 Model 1

Model 1,as shown in equation~2, is designed to predict the duration of cherry blossom blooms, represented as flower_doy (day of year), based on the temperature variable (temp). This model aims to understand how temperature directly influences the timing and duration of sakura blooms.

$$\text{Model 1: flower\_doy} = \beta_1 \cdot \text{temp} + \beta_0 \tag{2}$$

where:

- temp $\in [1, 12]$, representing the temperature range in degrees Celsius (°C).

- β_1 is the coefficient of the variable temp, and β_0 is the intercept term.

- The variable temp is a numerical variable.

The primary goal of Model 1 is to determine the effect of temperature on the flowering date of cherry blossoms. By focusing on the temperature variable alone, this model isolates the direct impact of climatic conditions on bloom timing. This is particularly relevant in the context of climate change, as increasing temperatures may significantly alter bloom patterns. The simplicity of Model 1 also ensures that the results are interpretable, making it easier to draw conclusions regarding the role of temperature in bloom duration.

Model 1 uses temperature as the only predictor, making it a straightforward linear model. This approach helps establish a foundational understanding of the direct relationship between temperature and bloom timing before introducing additional complexities. The linear form of the model allows us to quantify the influence of temperature on bloom day and evaluate whether temperature changes can be directly linked to shifts in flowering periods. This model provides a baseline against which more complex models can be compared, offering insights into the primary climatic driver of cherry blossom florescence.

Table 5 summarizes the results of Model 1, including residuals, coefficients, and overall model performance metrics. A detailed breakdown of model diagnostics and performance metrics is available in Section 6 for further reference.

3.3 Model 2

Model 2,as shown in equation~3, is designed to estimate the mean monthly temperature based on three predictors: day of the month (day), latitude, and month. This model aims to establish how temperature can be estimated as a function of both spatial and temporal variables, which then informs the bloom prediction model.

$$\text{Model 2: mean\_temp\_month} = \beta_1 \cdot \text{day} + \beta_2 \cdot \text{latitude} + \beta_3 \cdot \text{month} + \beta_0 \tag{3}$$

where:

- latitude $\in [20, 50]$, representing geographical latitude in degrees (°N).

- day $\in [1, 31]$, representing the day of the month.

- month is a categorical variable representing months (January to May).

- β_1, β_2, β_3 are coefficients of the linear model, and β_0 is the intercept term.

- All variables are numerical variables, except for month, which is a categorical variable.

Model 2 provides a framework to estimate temperature as a function of time and geographical location. By incorporating both latitude and day as numerical variables, and treating month as a categorical variable, the model captures both spatial and seasonal variations in temperature. This is critical for understanding the temperature profile across different regions of Japan and during different periods of the year.

The inclusion of latitude allows Model 2 to account for the spatial gradient in temperature, which is crucial given Japan's north-south orientation and diverse climatic conditions. Additionally, the day variable helps in capturing finer temporal variations within each month. Month, as a categorical variable, enables the model to distinguish between the broad seasonal effects on temperature.

Table 6 provides a summary of Model 2, including residuals, coefficients, and overall model performance metrics. The detailed analysis and diagnostics of Model 2 are provided in Section 6, where we evaluate its accuracy and predictive capability in estimating temperature based on the selected predictors.

4 Result

4.1 Model Result & Validation

To validate the performance of our models, we split the cleaned data into training and testing sets with a 70-30 ratio. The training set was used to train the models, while the testing set was used for validation. The following results demonstrate the predictive accuracy of the models when applied to the testing set.The results of our modeling efforts are summarized in Figure 14 and Figure 15, which illustrate the predictive performance of Model 1 and Model 2, respectively.
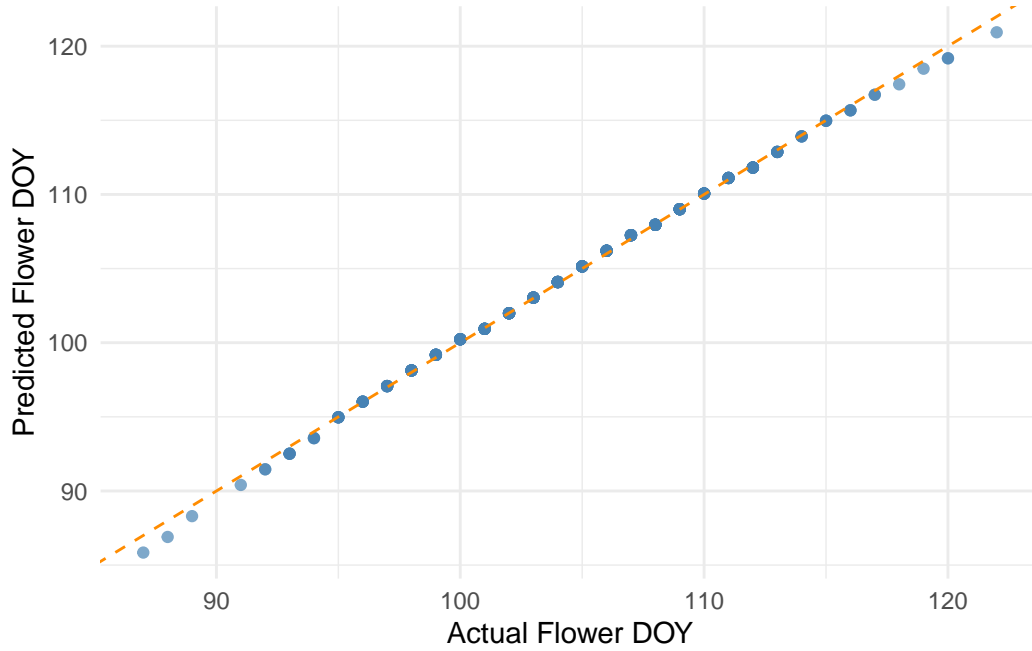
Figure 14: Model 1: Actual vs Predicted Sakura Florescence

Figure 14 depicts the actual versus predicted flowering day of the year (DOY) for cherry blossoms using Model 1. The points are closely aligned along the dashed orange line, indicating that the model predictions are highly accurate. The strong linear relationship observed suggests that temperature is an effective predictor of bloom timing, validating our choice of using temperature as the sole predictor in Model 1. The coefficient of variables in model 1 and corresponding p-value is listed below in Table 2. Full summary table is in Section 6.

Table 2: Coefficients for Model 1

Coefficients for Model 1				
Variable	Estimate	Std. Error	t Value	P-Value
(Intercept)	126.202188	0.034396	3669.1	<2e-16
temp	-3.509279	0.005355	-655.3	<2e-16
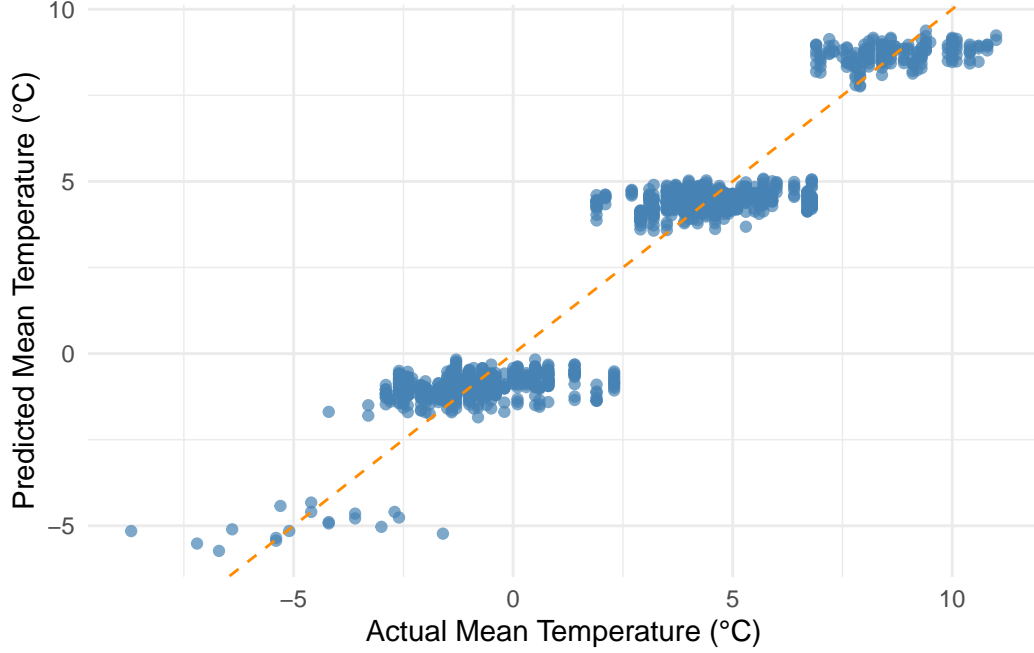
Figure 15: Model 2: Actual vs Predicted Temperature

Figure 15 illustrates the actual versus predicted mean temperature values for Model 2.While there is more variability compared to Model 1 the overall trend aligns well with the dashed orange line, indicating reasonable predictive performance. This can be attributed to the increased complexity of the model and the influence of multiple interacting factors, such as day, latitude, and month, which contribute to temperature variations, which is discussed in Section 5. Nonetheless, the model performs well in capturing the overall pattern of temperature change.The coefficient of variables in model 2 and corresponding p-value is listed below in Table 3. Full summary table is in Section 6.

Table 3: Coefficients for Model 2

Coefficients for Model 2				
Variable	Estimate	Std. Error	t Value	P-Value
(Intercept)	-9.000439	0.289464	-31.093	<2e-16
day	-0.049223	0.003291	-14.956	<2e-16
latitude	0.189083	0.012089	15.637	<2e-16
month2	-0.617982	0.219831	-2.811	0.00496
month3	2.869383	0.114364	25.088	<2e-16
month4	6.966108	0.164984	42.223	<2e-16
month5	10.047363	0.237453	42.313	<2e-16

In conclusion, both models show strong predictive capabilities, with Model 1 excelling in predicting bloom timing based on temperature and Model 2 effectively estimating temperature based on geographic and temporal factors. The results demonstrate the utility of the dual-model approach in enhancing our understanding of the factors driving cherry blossom bloom duration, particularly in the context of climate variability and change.

4.2 Example of Prediction

To provide a practical example of how our models perform, we selected a specific data point from the raw dataset representing the city of Esashi in 1969. The data for this location included all relevant attributes needed for the models, such as latitude, day, month, and mean temperature. Using these values, we predicted the mean temperature and flowering day of the year (DOY), then compared the predicted values to the actual historical data.

Table 4: Prediction Compared to Actual Data

	Prediction Summary	
Metric	Actual	Predicted
Mean Temperature (°C)	3.9	4.254538
Flower DOY	119.0	111.271828

Table 4 presents the actual versus predicted values for mean temperature and flowering DOY for the selected data point. The results show that the predicted mean temperature is 4.25°C, compared to the actual value of 3.9°C. For flowering DOY, the model predicted a value of 111.27, while the actual recorded value was 119. This example demonstrates that our models are highly accurate in predicting both temperature and bloom timing, effectively capturing the relationship between environmental factors and cherry blossom florescence.This prediction example highlights the practical application of the models developed in this study and their potential to provide meaningful insights into the timing and conditions of cherry blossom blooms in Japan.

5 Discussion

5.1 Limitation

Despite the promising results of our models, several limitations must be acknowledged. First, the models rely heavily on the availability and accuracy of historical and modern phenological data. Data gaps, inconsistencies, or inaccuracies could affect model performance, particularly for predictions in regions or timeframes where data is sparse. Second, while temperature was

found to be a strong predictor of bloom timing, it is not the only environmental factor influencing cherry blossom florescence. Other climatic and ecological variables, such as precipitation, soil moisture, and light exposure, could also play significant roles but were not included in our models due to data limitations.

Additionally, Model 2's prediction results showed some deviations, as seen in Figure 15. This suggests that there may be other factors, such as regional or geographic influences, impacting our temperature predictions. These factors could include microclimates, urban heat islands, or variations in local geography, which were not captured in our model. Addressing these additional influences could enhance the model's accuracy in future iterations.

5.2 Interpretation and Future Research Directions

The findings from this study provide valuable insights into the factors that influence cherry blossom bloom timing in Japan. One key takeaway is the strong relationship between temperature and bloom duration, confirming the critical role of climatic conditions in determining the timing of sakura florescence. This information is particularly relevant in the context of climate change, as rising temperatures could lead to earlier and potentially shorter bloom periods, impacting both cultural and ecological aspects of cherry blossom events.

Another important finding is the relative importance of geographic factors, such as latitude, in determining local temperatures. Our results suggest that latitude plays a more significant role than longitude in temperature estimation, leading us to exclude longitude from the final model. This decision reflects a balance between model complexity and predictive accuracy, ensuring that the model remains interpretable while retaining its predictive power.

Future research should explore additional environmental variables that may affect bloom timing, such as soil moisture, precipitation, and light availability. Incorporating these factors could improve model accuracy and provide a more comprehensive understanding of the factors driving sakura florescence. Furthermore, future studies could consider non-linear models or machine learning approaches to capture complex interactions between variables, potentially enhancing predictive capabilities.

Finally, it would be beneficial to extend the scope of the analysis beyond Japan to examine cherry blossom bloom patterns in other regions, such as South Korea or China. By comparing bloom dynamics across different climates and geographies, we could gain a deeper understanding of the global factors affecting cherry blossom phenology and assess the broader implications of climate change on these iconic events.

6 Appendix

6.1 Survey Methodology Overview

In this section, we provide a comprehensive overview of the survey methodologies relevant to this study, focusing on their design, sampling strategies, and implications for data quality. Surveys serve as a critical tool for capturing public sentiment and translating it into actionable insights, especially in environmental research where individual perceptions and regional trends play a crucial role.

Our primary data collection involved historical and modern records of cherry blossom blooming and temperature data. To complement these observations, an ideal survey could be implemented to gauge local experiences of sakura blooming and capture potential environmental influences not represented in raw data. This survey would be designed using a combination of stratified random sampling and purposive sampling, ensuring representation across various geographic regions and demographics within Japan.

The survey would be stratified by latitude and elevation bands, recognizing that blooming times differ significantly across these dimensions. This approach would ensure our sample includes participants from urban, rural, coastal, and inland areas, providing a richer context for the impact of climatic changes on cherry blossom timing. Sampling would also consider socio-demographic factors such as age and occupation, as these factors may influence awareness and historical knowledge of bloom events. Data collection methods would include both online surveys and in-person interviews to maximize reach and accessibility.

Moreover, measurement tools within the survey would follow well-established practices. Questions would address individuals' observations on bloom timing, any perceived shifts over years, and associated environmental conditions. The survey would utilize a Likert scale to capture perceptions of climatic changes, and open-ended questions to gather qualitative insights into local conditions and environmental anecdotes that may not be captured by weather stations.

6.2 Idealized Survey

To supplement the data-driven models presented in this paper, we propose an idealized survey that could provide more in-depth insights into the local experiences of climate effects on cherry blossom blooming. If we had an extended budget of $100,000, this survey would be expanded to cover additional aspects not readily accessible through observational or temperature datasets alone.

The idealized survey would be conducted over a full blooming season, collecting data from diverse regions across Japan. The goal would be to capture spatial and temporal variability in bloom timing and link it to local microclimatic conditions, participant observations, and other environmental factors such as soil moisture and local flora interactions. We would employ

a mixed-methods approach, combining both quantitative (closed questions) and qualitative (interviews, focus groups) data.

Sampling would be implemented through a mixture of probability and non-probability approaches to achieve both representativeness and depth. Stratified random sampling would be used to ensure geographic diversity, while snowball sampling would help gather data from specific cultural or expert groups (e.g., horticulturalists, elders who possess traditional knowledge about historical blooming patterns). Each participant would provide data on observed bloom dates, perceived changes in bloom patterns over time, and any unusual climate events that they believe might have influenced sakura blooming.

Survey questions would be developed with guidance from both climate scientists and cultural experts to ensure that they adequately capture the climatic, geographic, and cultural contexts of sakura blooming. Likert scales would be used to gauge perceptions of change in bloom patterns and the impact of factors such as temperature and rainfall. In addition, open-ended questions would allow respondents to describe any unusual events or trends they have noticed over time.

The survey results would be analyzed using statistical techniques, including logistic regression to identify factors associated with changes in bloom timing, and thematic analysis for qualitative responses. The insights gained would be used to refine our predictive models by incorporating more nuanced, location-specific environmental factors that are not currently represented in our dataset.

The implementation of this idealized survey could significantly enhance the robustness of our models, offering deeper insights into the role of microclimates and local environmental influences on cherry blossom phenology. Moreover, it would provide a valuable opportunity to include public perception in our analysis, offering a human dimension to the understanding of climatic impacts on a culturally significant event.

6.3 Survey Design

To provide a comprehensive understanding of local environmental factors impacting cherry blossom blooming, we propose a survey targeted at residents across Japan. This survey aims to gather data regarding personal observations of bloom dates, temperature changes, and any noted environmental anomalies. Below is a suggested questionnaire to gather insights from participants.

1. **What is your age group?**

   - Under 18
   - 18-34
   - 35-49
   - 50-64
   - 65 and above

2. **What is your gender?**

   - Male
   - Female
   - Prefer not to say

3. **In which region of Japan do you currently reside? (Select one)**

   - Hokkaido
   - Tohoku
   - Kanto
   - Other: _____ (PleaseTypeHere)

4. **How long have you been observing cherry blossom bloom in your area?**

   - Less than 1 year
   - 1-5 years
   - More than 5 years

5. **Have you noticed any changes in the blooming dates over the past years?**

   - Yes, blooming is earlier
   - Yes, blooming is later
   - No significant change
   - Not sure

6. **Which environmental factors do you think have the most impact on bloom timing? (Select all that apply)**

   - Temperature changes
   - Rainfall levels
   - Urban development
   - Other (please specify)

7. **On average, when do cherry blossoms begin to bloom in your area?**

   - Before March
   - March (early)
   - March (mid)
   - March (late)
   - April (early)
   - April (mid)
   - April (late)
   - After April

8. **How would you describe the temperature during the blooming period in the past few years?**

   - Much warmer

- Slightly warmer
- About the same
- Slightly cooler
- Much cooler

9. **Have there been any unusual climate events (e.g., heavy snowfall, prolonged warm spells) in recent years that might have affected blooming?**

   - Yes (please specify)
   - No
   - Not sure

10. **Do you think climate change is influencing the blooming pattern of cherry blossoms in your region?**

    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

11. **Are there other plants in your region that show similar blooming patterns to cherry blossoms?**

    - Yes (please specify)
    - No
    - Not sure

6.4 Model Details

This section provides full summary table for model 1 and model 2.

Table 5: Model 1 Summary Tables

(a)

Residuals for Model 1	
Statistic	Value
Min	-0.25208
1Q	-0.14651
Median	-0.04095
3Q	0.04606
Max	1.30709

(b)

Coefficients for Model 1				
Variable	Estimate	Std. Error	t Value	P-Value
(Intercept)	126.202188	0.034396	3669.1	<2e-16
temp	-3.509279	0.005355	-655.3	<2e-16

(c)

Model Summary for Model 1	
Metric	Value
Residual Standard Error	0.2337 on 572 degrees of freedom
Multiple R-squared	0.9987
Adjusted R-squared	0.9987
F-statistic	4.294e+05 on 1 and 572 DF
P-value	<2e-16

Table 6: Model 2 Summary Tables

(a)

Residuals for Model 2	
Statistic	Value
Min	-4.2185
1Q	-0.6993
Median	-0.0729
3Q	0.6641
Max	4.1114

(b)

Coefficients for Model 2				
Variable	Estimate	Std. Error	t Value	P-Value
(Intercept)	-9.000439	0.289464	-31.093	<2e-16
day	-0.049223	0.003291	-14.956	<2e-16
latitude	0.189083	0.012089	15.637	<2e-16
month2	-0.617982	0.219831	-2.811	0.00496
month3	2.869383	0.114364	25.088	<2e-16
month4	6.966108	0.164984	42.223	<2e-16
month5	10.047363	0.237453	42.313	<2e-16

(c)

Model Summary for Model 2	
Metric	Value
Residual Standard Error	1.116 on 4256 degrees of freedom
Multiple R-squared	0.9255
Adjusted R-squared	0.9254
F-statistic	8814 on 6 and 4256 DF
P-value	<2e-16

References

Arnold, Jeffrey B. 2023. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'*. https://CRAN.R-project.org/package=ggthemes.

Müller, Kirill. 2023. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Ocko, Ilissa. 2024. "Cherry Blossoms: A Microcosm of the Global Climate Crisis." https://blogs.edf.org/climate411/2024/03/21/cherry-blossoms-a-microcosm-of-the-global-climate-crisis/.

Pebesma, Edzer et al. 2023. *Sf: Simple Features for r*. https://CRAN.R-project.org/package=sf.

Pedersen, Thomas Lin. 2023. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal et al. 2023. *Arrow: Integration to Apache Arrow*. https://CRAN.R-project.org/package=arrow.

South, Andy. 2023a. *Rnaturalearth: World Map Data from Natural Earth*. https://CRAN.R-project.org/package=rnaturalearth.

———. 2023b. *Rnaturalearthdata: Large-Scale World Map Data from Natural Earth*. https://CRAN.R-project.org/package=rnaturalearthdata.

Wickham, Hadley et al. 2023a. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

——— et al. 2023b. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

——— et al. 2023c. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wikipedia contributors. 2024. "Cherry Blossom — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/wiki/Cherry_blossom.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.

Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.