Tulip Analysis

Jackson Curtis

Stat 536

April 21, 2018

Introduction

- Tulips in the Netherlands are a huge driver of economic activity and tourism
- Global warming is especially bad in the Netherlands (flooding and large temperature increase).
- Tulips need a a winter chilling time before they bloom in the spring

Research Questions

1. Is the probability of germination for each chilling time the same across all species? Which species are the same or different?
2. Is there an "ideal" chilling time? Does this time vary by species?
3. What effect will a decrease from 10 to 8 weeks of chilling time have for tulips? Is it the same for each species?

The Data

The experiment examined 12 species of tulips, and assigned groups of thirty bulbs to 7 different chilling times (0-12 weeks).

- 2,520 flowers planted in total
- Response measured whether the tulip germinated
- The twelfth species never germinated, so we removed it from our analysis

Germination Rates

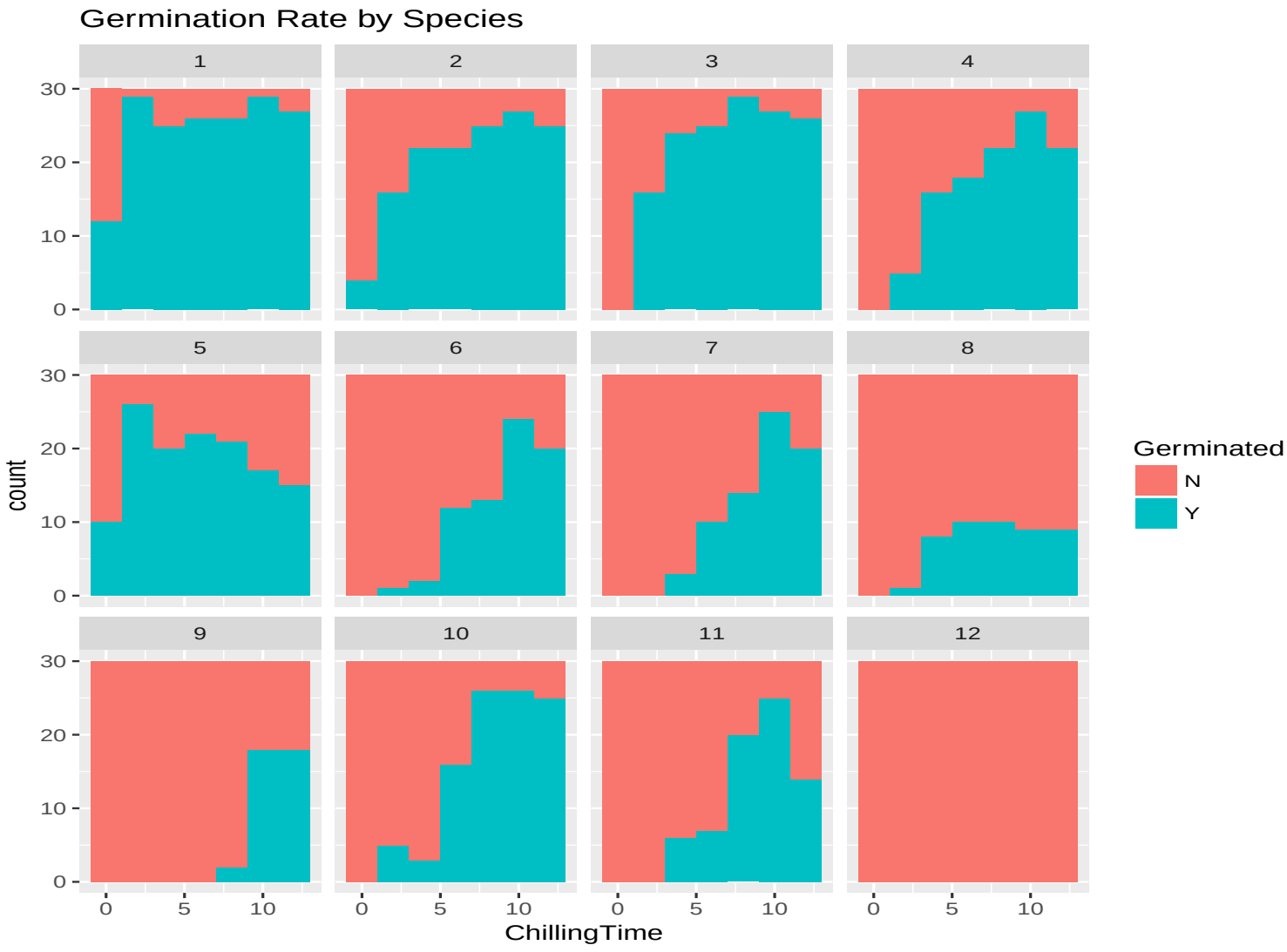
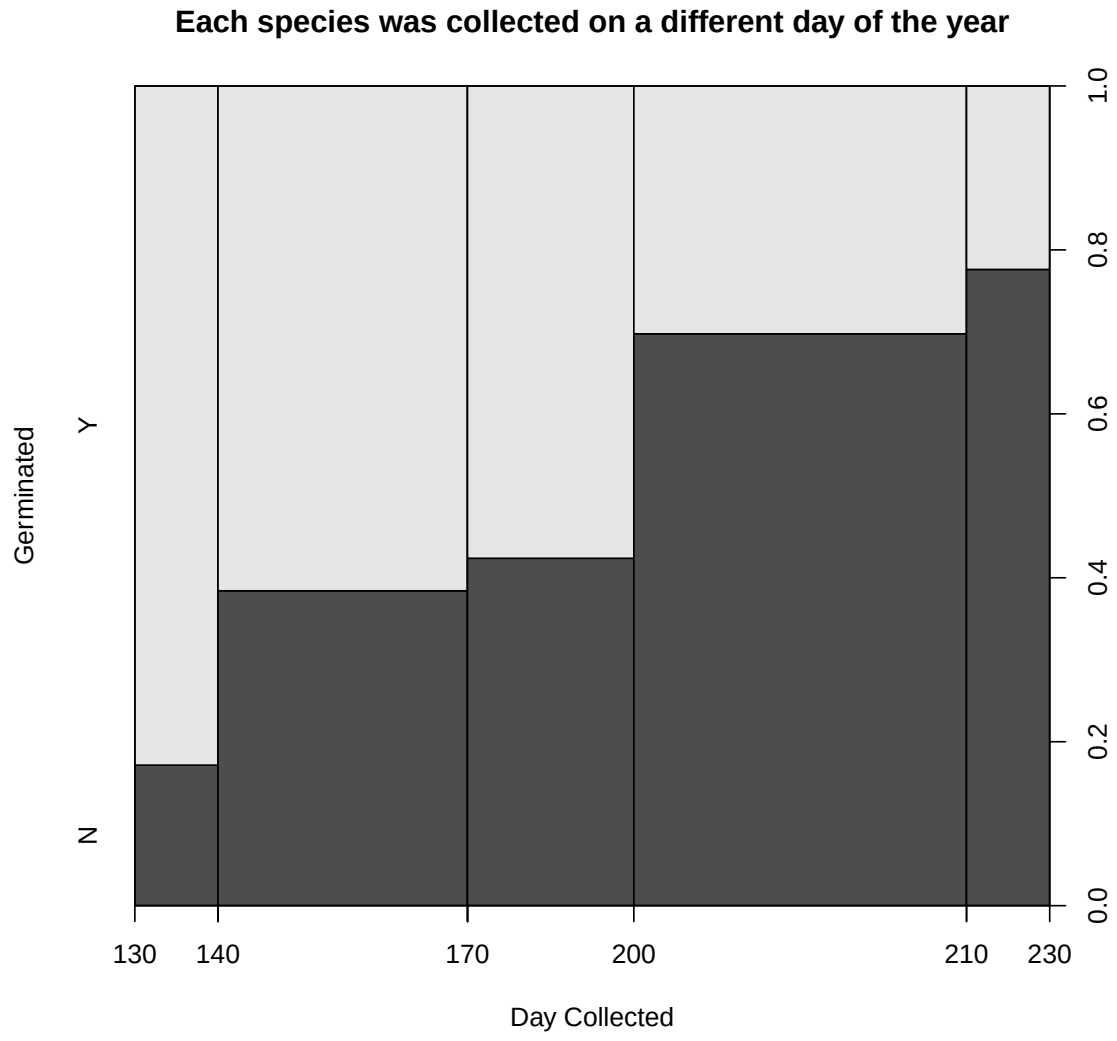


Figure 1: The number of plants that germinated at each chilling time, by species

Confounding Variable?



Model Framework

We have a categorical response and want to model the probability of germination. Logistic regression seems like a natural fit.

- We will have to watch out for non-monotonicity
- Looks like there is a strong case for interactions

Question #1

Is the probability of germination for each chilling time the same across all species? Which species are the same or different?

To answer the first half of this question, we will fit a robust model and compare it to a model that enforces all species to have similar germination probabilities.

Question #1

Our full model:

$$y_i \sim Bern(p_i)$$

$$log(\frac{p_i}{1-p_i}) = X\beta$$

The X-matrix will be made up of an indicator for each species, and an interaction between chilling time and species, but since there appeared to be non-monotonicity, I added a second degree spline (with a knot at 6 weeks) to the chilling time. In total the X-matrix is 2310 x 44. β estimates are provided by maximizing the likelihood numerically.

Question #1

From our full model we can plot the fitted lines over each species:

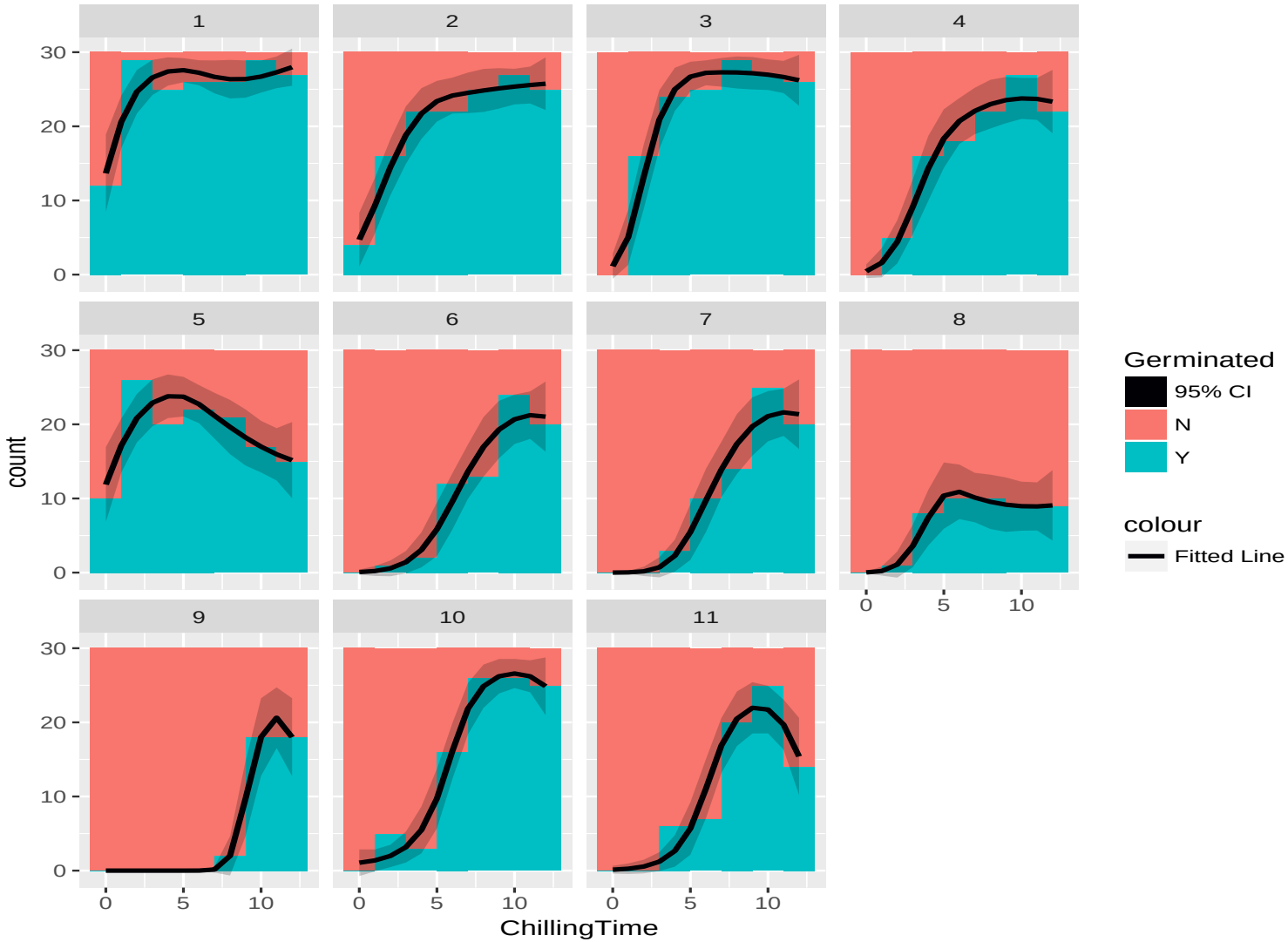


Figure 2: The fitted curves clearly appear to be different across species

Question #1

We can adjust our model to ignore any species effect

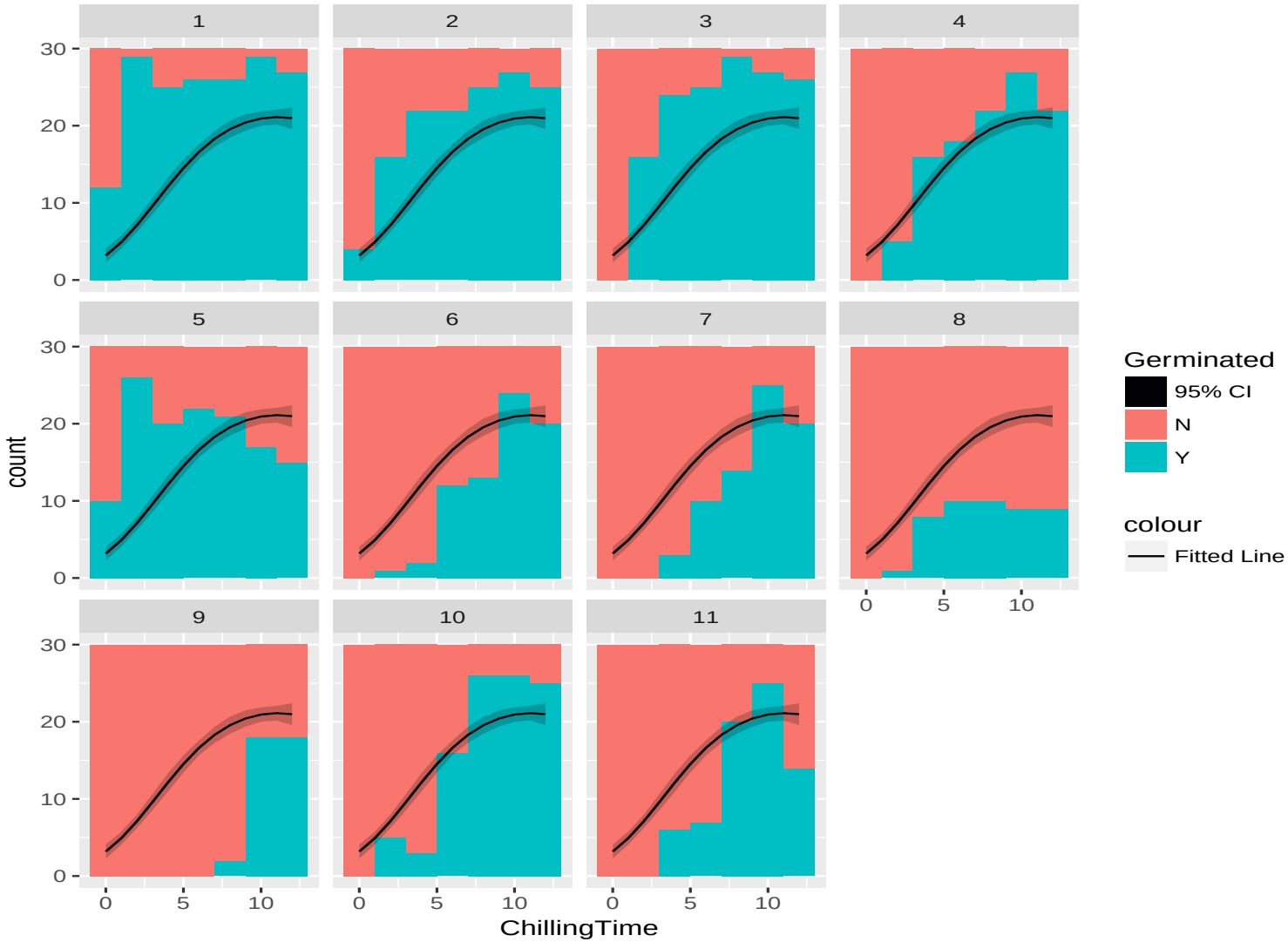


Figure 3: Ignoring species clearly results in a much worse fitting model

Question #1

To answer the follow-up question (which species could we reasonably treat as the same) we can compare the BICs for model fit when we reduce the model by combining categories.
Consulting with a subject matter expert could probably result in better groupings, but the groupings we'll use from here on out is:

2, 3	4, 6, 7, 10	1	5	8	9

Table 1: Species in the same box had very similar predicted curves

Simplified Model Fit

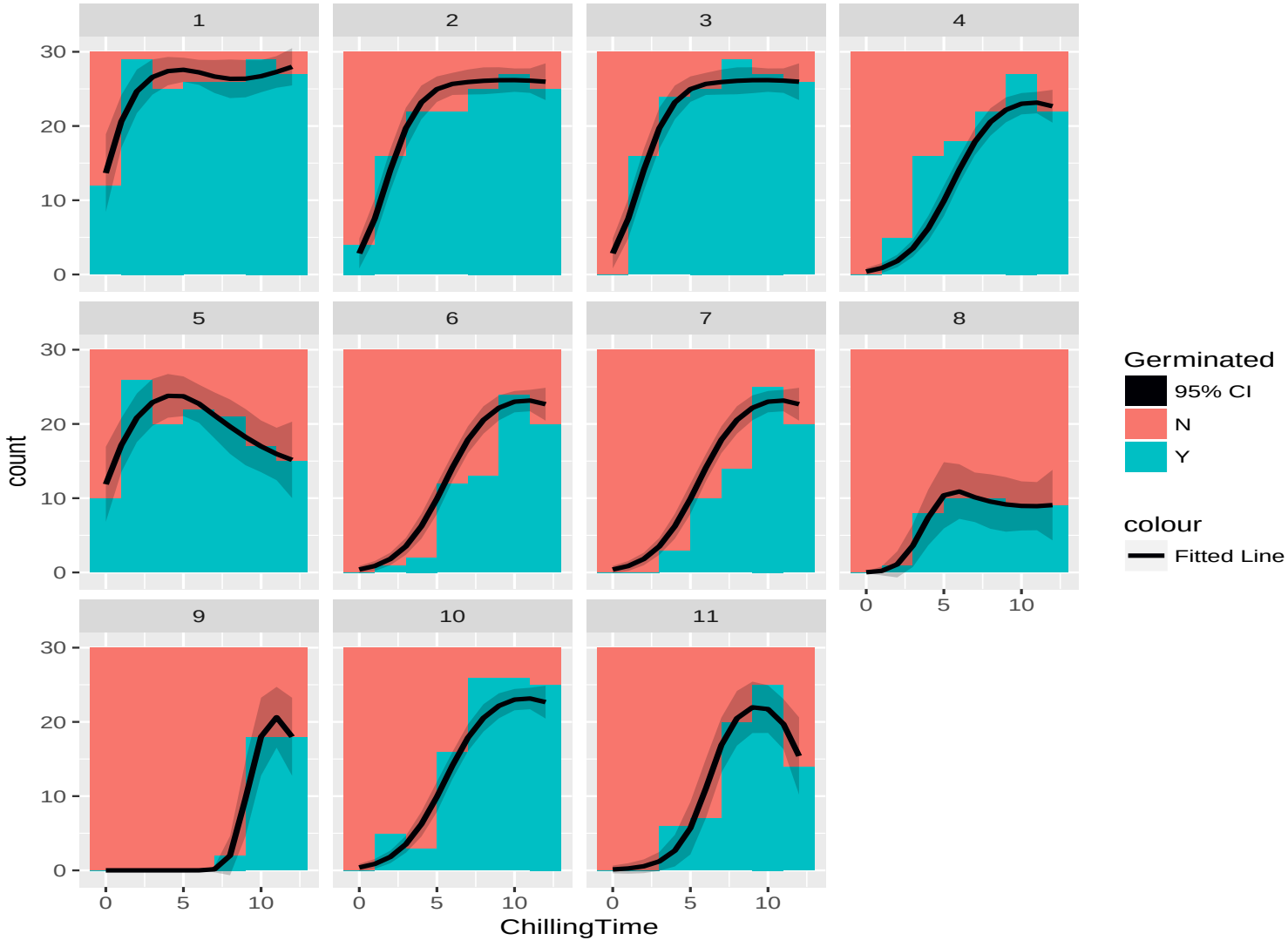


Figure 4: Combined groups have significantly smaller confidence bands

Assessing Model Fit

The Hosmer-Lemeshow Goodness of Fit test assesses whether the probabilities predicted by the model match empirical 0/1 observations. We will put 10/30 observations in a test set and train our model on the remaining 20. A failed test could indicate that our model is overfitting the data.

The Hosmer-Lemeshow test provides a p-value of 0.86, indicating that our predicted probabilities do not seem unreasonable as the probabilities of germination.

Question #2

Is there an "ideal" chilling time? Does this time vary by species?

Question #2

Species	Ideal Chilling Time
1	12
2	10
3	10
4	11
5	4
6	11
7	11
8	6
9	11
10	11
11	9

Table 2: Chilling time where predicted probability of germination is maximized

Question #3

What effect will a decrease from 10 to 8 weeks of chilling time have for tulips? Is it the same for each species?

Question #3

Species	% at 8 weeks	At 10 weeks	Dif	95% Bootstrap CI
1	87.86	89.03	-1.17	(-15.3, 3.1)
2	86.95	87.26	-0.31	(-5.3, 2.7)
3	86.95	87.26	-0.31	(-5.3, 2.7)
4*	68.46	76.67	-8.22	(-12.1, -4.7)
5*	65.45	56.62	8.84	(1.5, 15.8)
6*	68.46	76.67	-8.22	(-12.1, -4.7)
7*	68.46	76.67	-8.22	(-12.1, -4.7)
8	31.82	29.86	1.95	(-4.1, 9.6)
9*	6.62	60.05	-53.43	(-73.2, -34.9)
10*	68.46	76.67	-8.22	(-12.1, -4.7)
11	68.30	72.42	-4.12	(-11.6, 2.4)

Table 3: Predicted percentage that will germinate at 8 and 10 weeks, along with bootstrap confidence intervals of difference. (* for different than 0)

Question #3

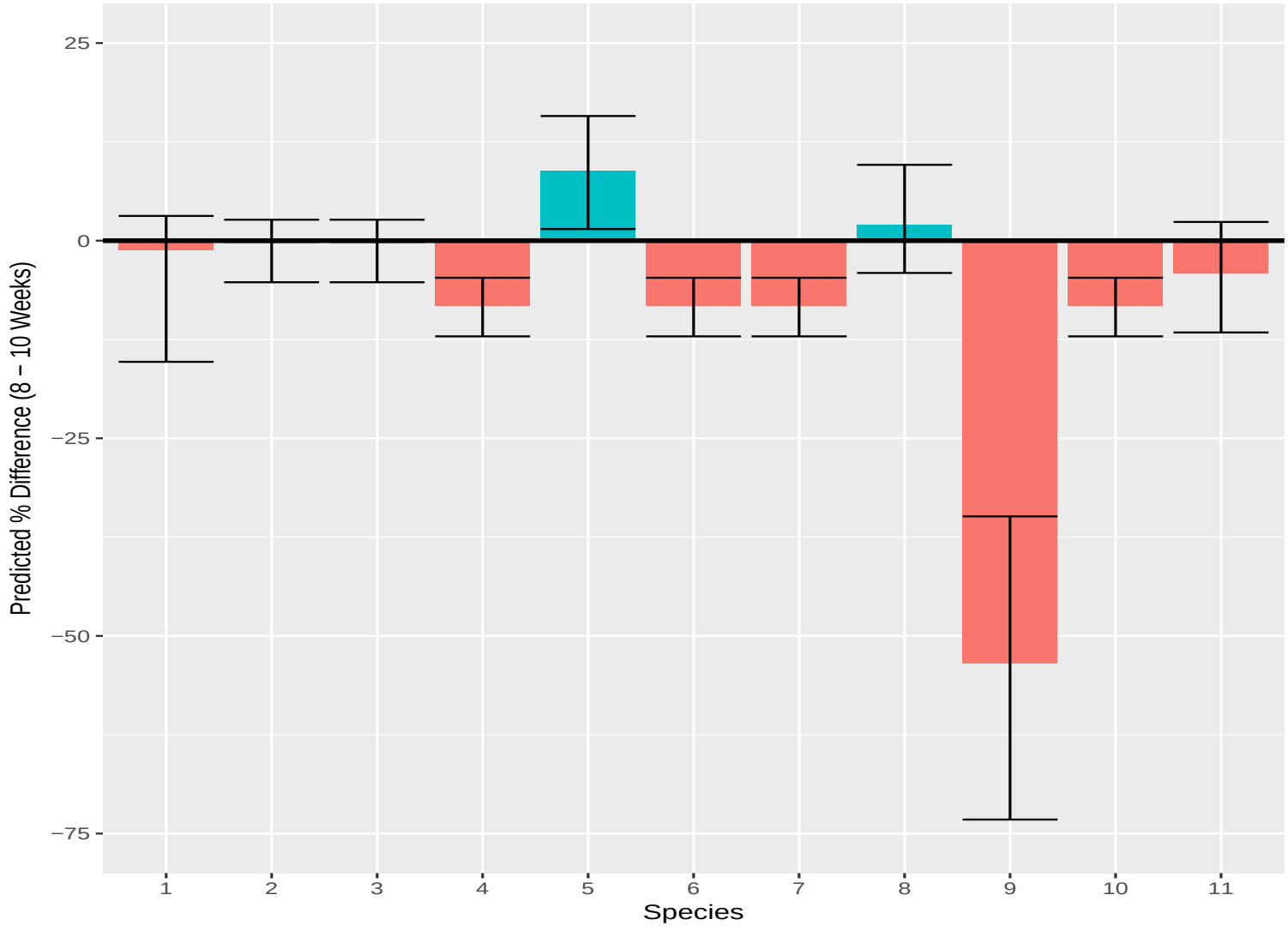


Figure 5: Most species of tulips will be negatively effected with a shortened winter

Conclusion

- Our model accurately describes the relationships between germination rates and chilling time.
- While there is significant differences between species, most benefit from longer chilling times.

Shortcomings and Future Work

Shortcomings:

- Because each species was collected at a single time in a single year, replication may not produce identical results
- Approach to identifying similar species was ad hoc (may be a feature unique to the sample)

Future work:

- Work with scientists on meaningful ways to group species
- Collect new data for species 12.