Applying Multiple Linear Regression to the Timing of Spring Migration in Hummingbirds

Sister Megan Thérèse Tunink

Benedictine College

January 26, 2019

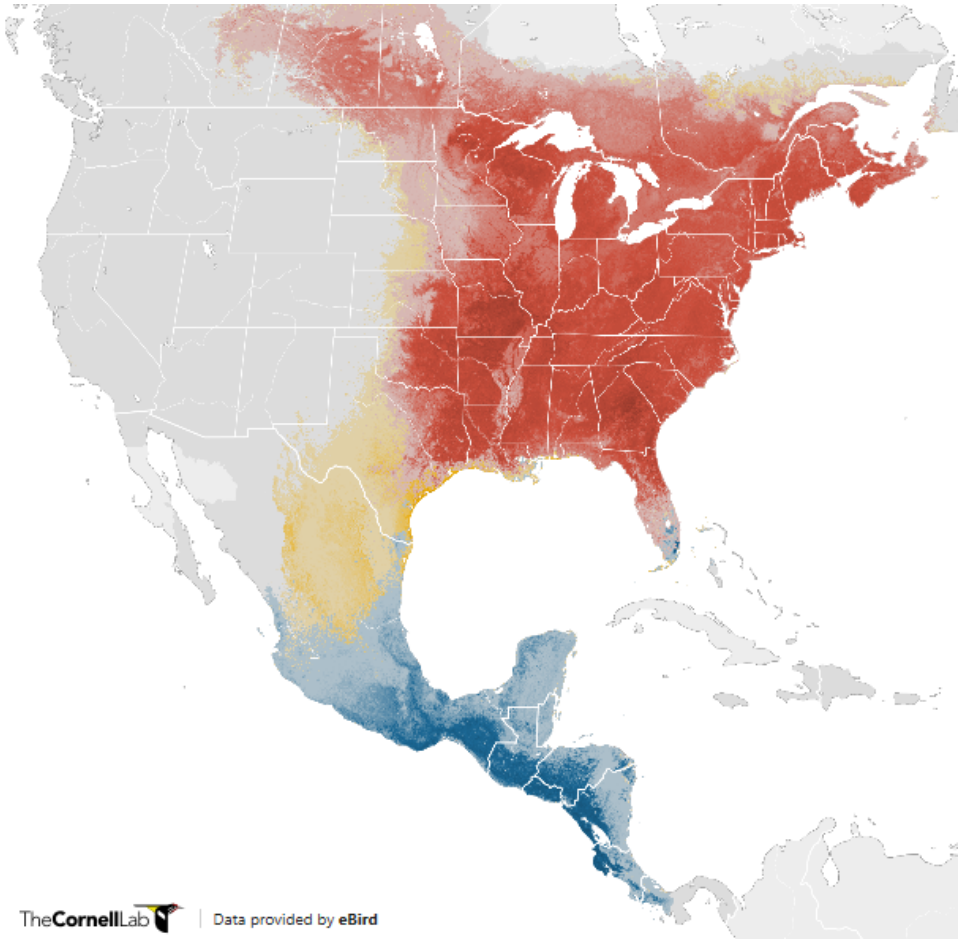
Overview

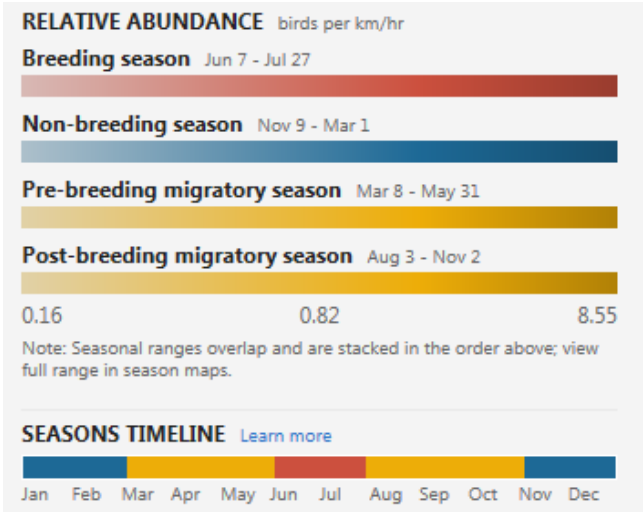
Purpose of this Research

What external cues do hummingbirds use to determine the timing of their spring migration?

Can we create a model that uses external cues to predict arrival date in a given location?

The Ruby-throated Hummingbird





Simple Regression

- Population regression line

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- The least squares regression

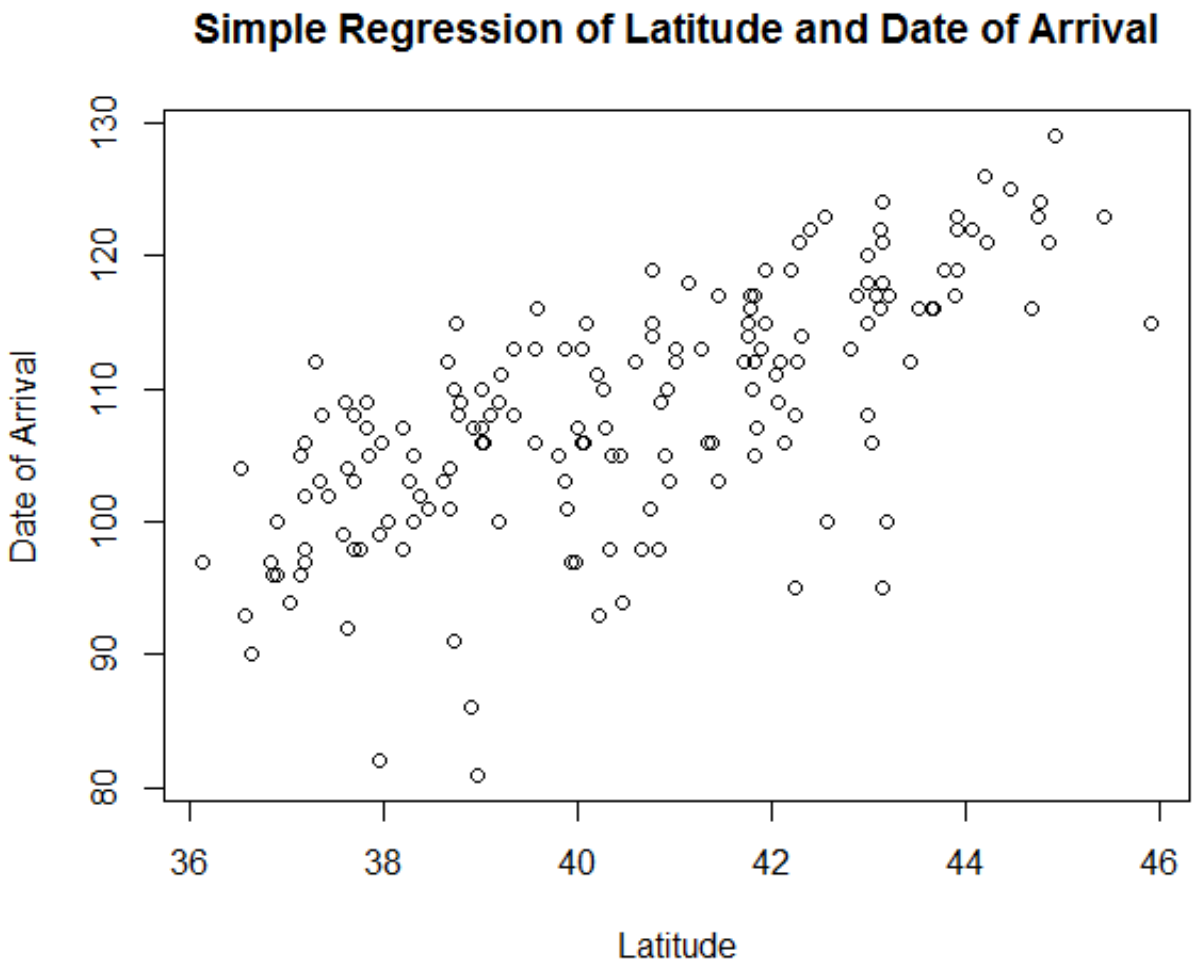
$$\hat{Y}_i = b_0 + b_1 X_i$$

estimates the population regression line

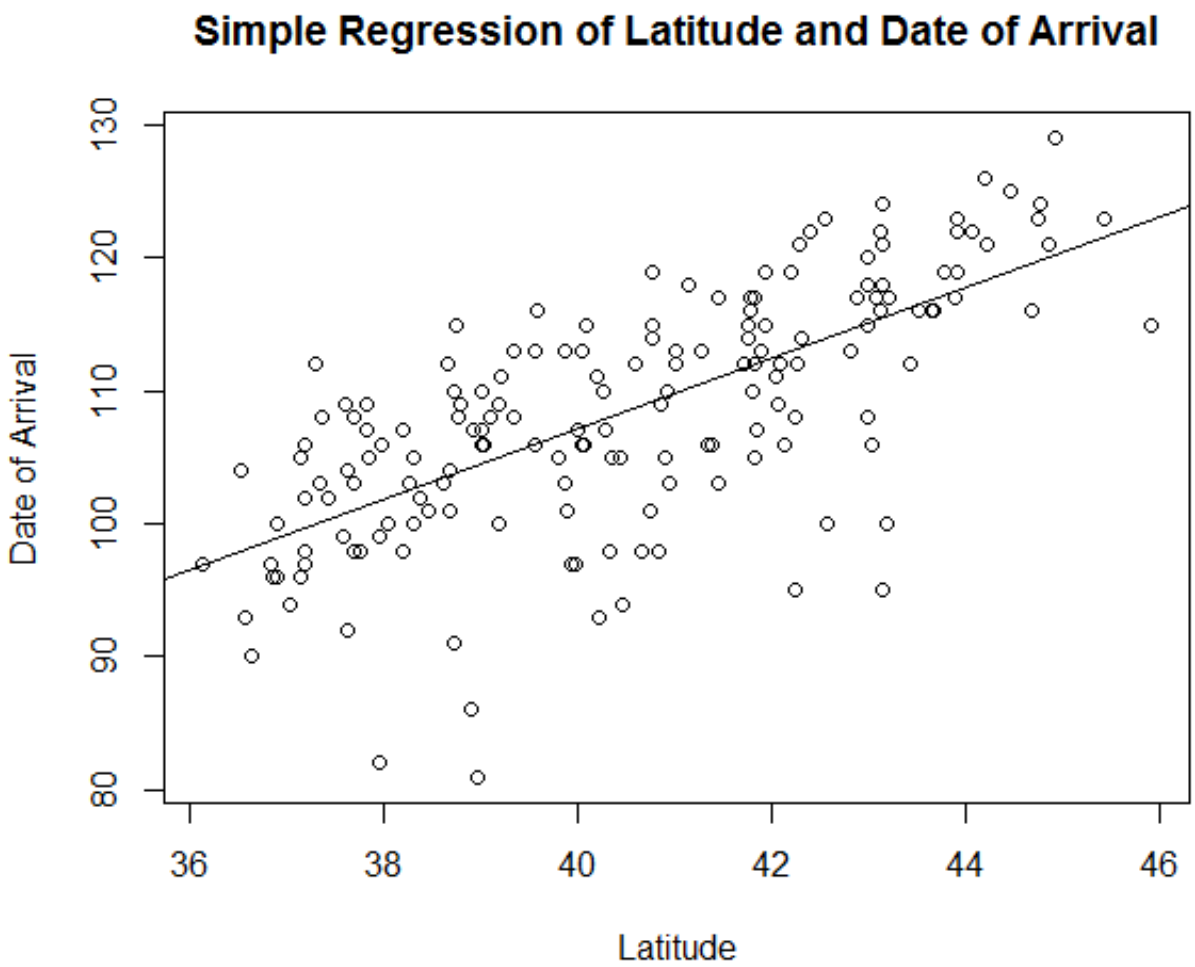
- Minimizes

$$MSE = \frac{1}{N} \sum_i \left(Y_i - \hat{Y}_i \right)^2$$

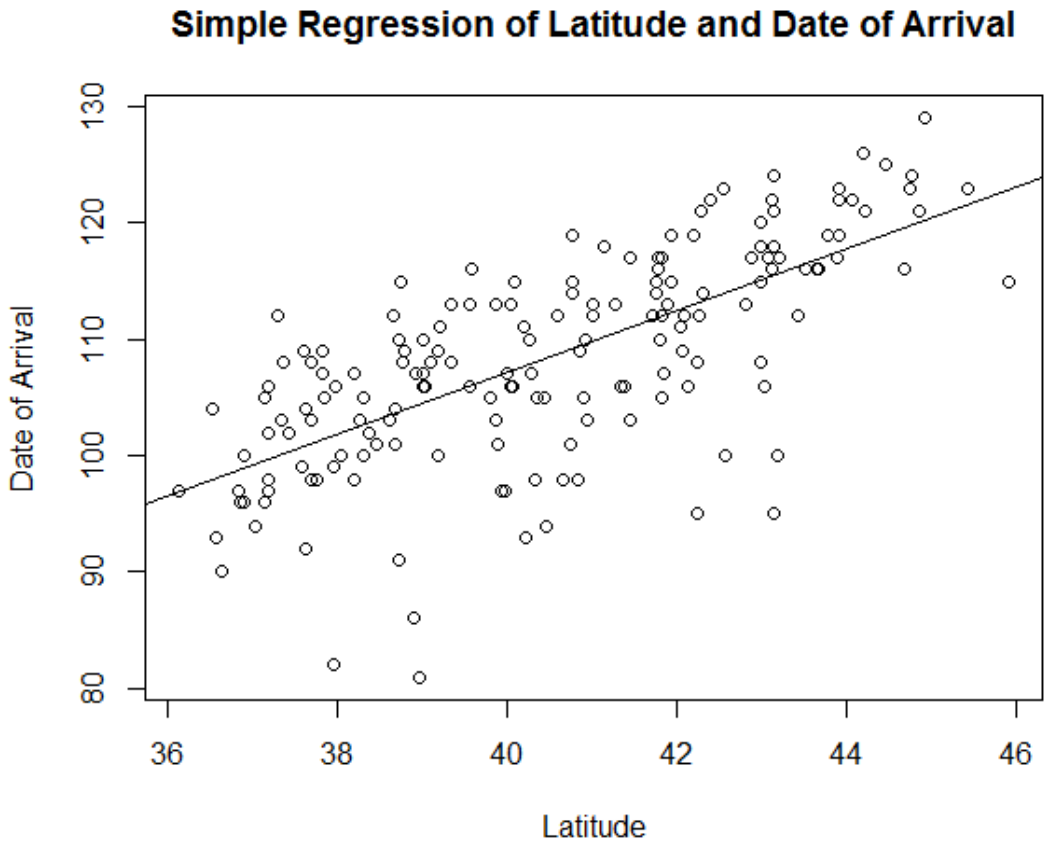
Simple Regression Example



Simple Regression Example



Simple Regression Example



$$\hat{Y}_i = 0.84 + 2.66X_i$$

$$R^2 = 0.476$$

Multiple Regression

- Population regression line

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i,$$

- The least squares regression

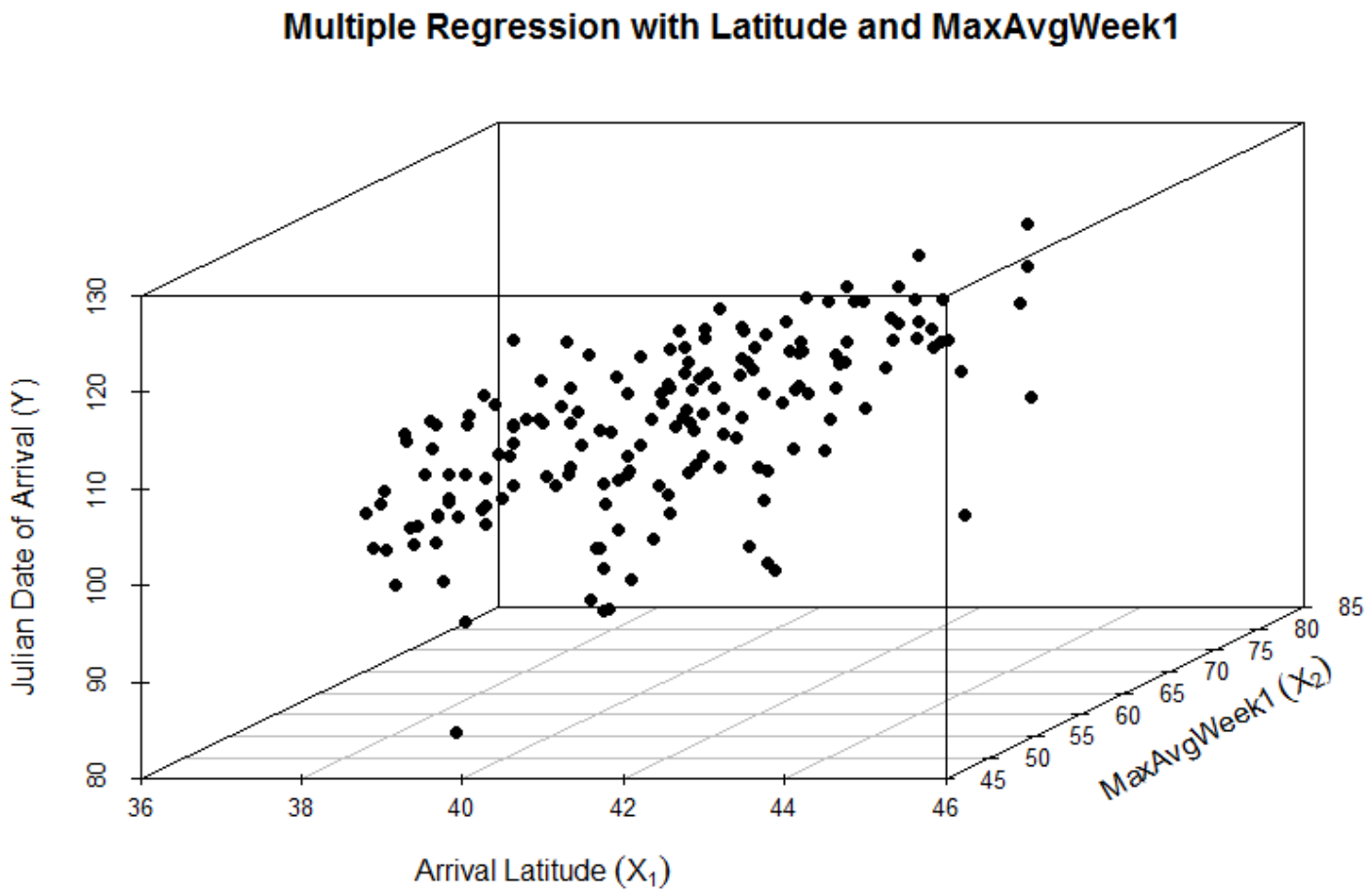
$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_p X_{ip}$$

  estimates the population regression line

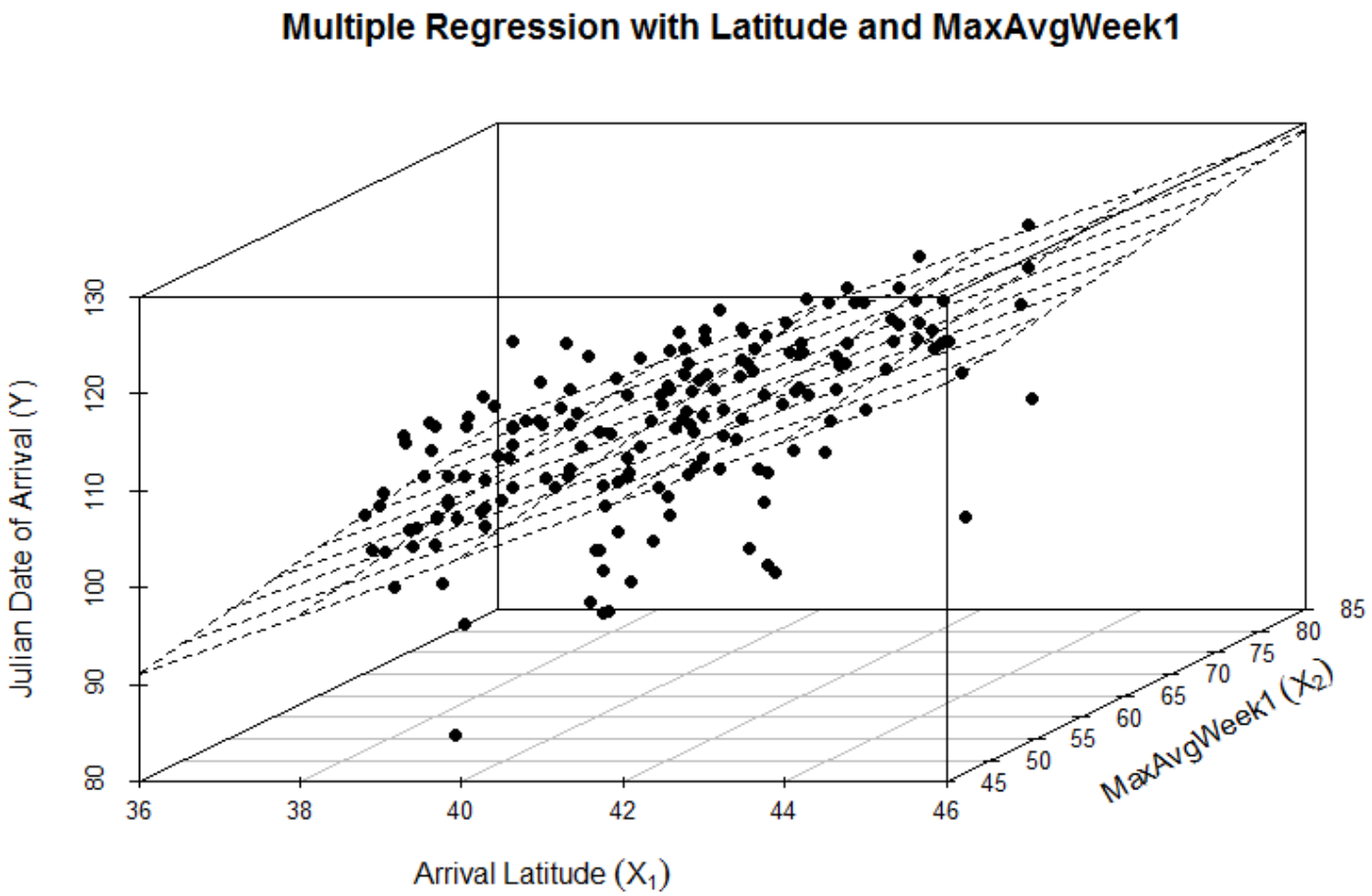
- Minimizes

$$MSE = \frac{1}{N} \sum_i \left(Y_i - \hat{Y}_i \right)^2$$

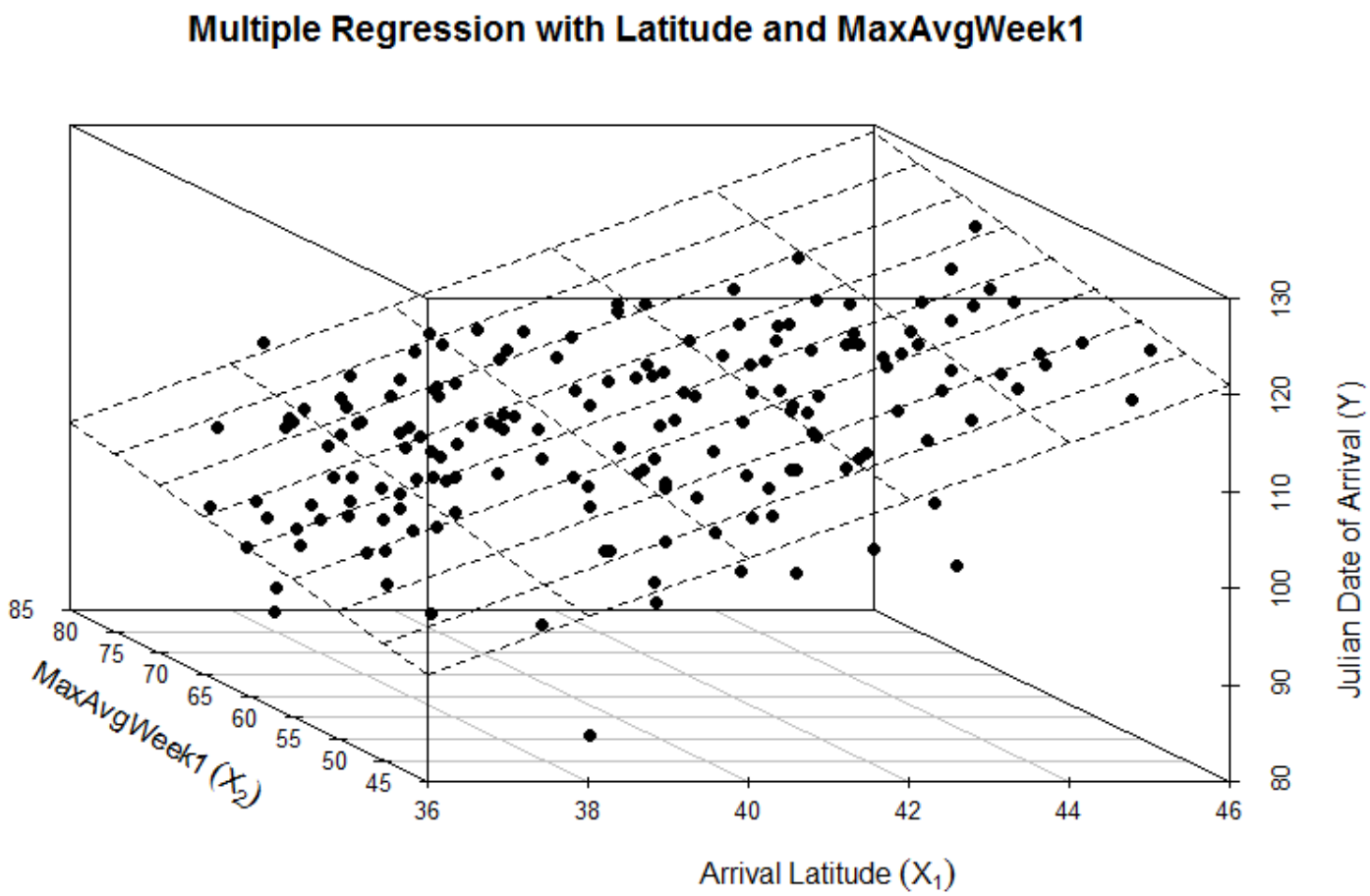
Multiple Regression Example



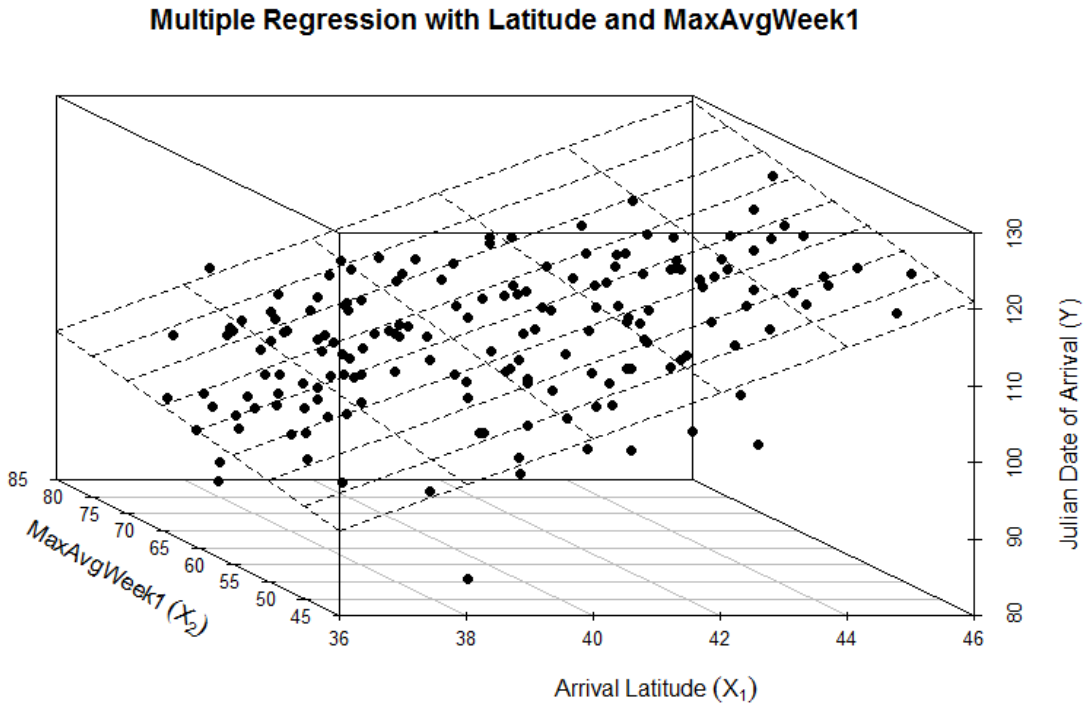
Multiple Regression Example



Multiple Regression Example



Multiple Regression Example



$$\hat{Y}_i = -25.84 + 2.99X_{i1} + 0.21X_{i2}$$

$$R^2 = 0.495 \qquad R^2_{\text{adjusted}} = 0.489$$

How do you create a model?

- Many different methods
- Stepwise Regression
  - Adding and removing variables based on if they significantly contribute to the model

The Model

$$\hat{Y}_i = -41.39 + 3.17X_{i1} + 0.32X_{i2} + 0.14X_{i3}$$

where

X_{i1} = Arrival Latitude

X_{i2} = Average Minimum Temperature in the Week before Arrival Week

X_{i3} = Average Maximum Temperature in Arrival Week

The Model

$$\hat{Y}_i = -41.39 + 3.17X_{i1} + 0.32X_{i2} + 0.14X_{i3}$$

where

X_{i1} = Arrival Latitude

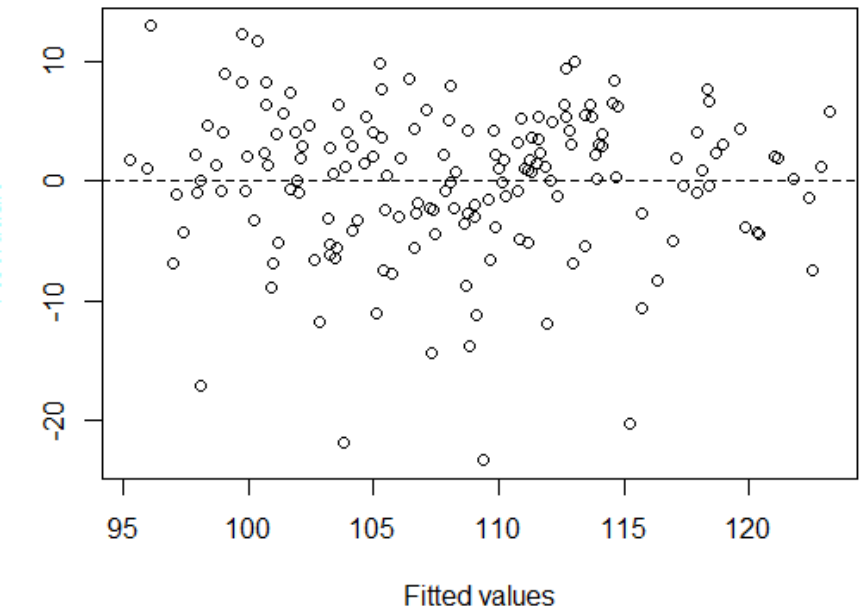
X_{i2} = Average Minimum Temperature in the Week before Arrival Week

X_{i3} = Average Maximum Temperature in Arrival Week

$$R^2 = 0.542 \qquad R^2_{\text{adjusted}} = 0.534$$

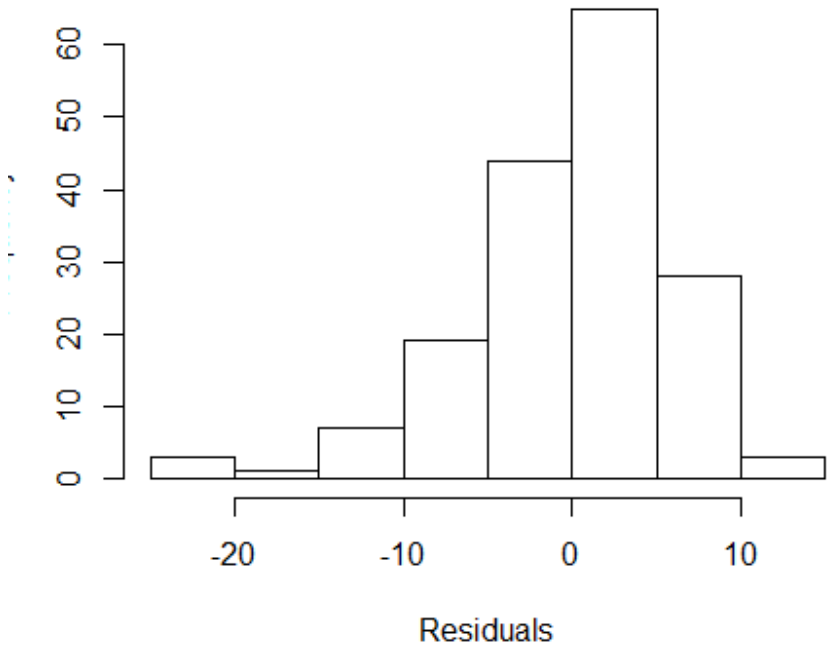
Assumptions of Multiple Regression

1. The errors, ϵ_i, are independent.
2. The errors, ϵ_i are normally distributed.
3. The errors, ϵ_i have equal variances.



Assumptions of Multiple Regression

1. The errors, ϵ_i, are independent.
2. The errors, ϵ_i are normally distributed.
3. The errors, ϵ_i have equal variances.



Future Research Directions

- Data transformations
- Build multiple models using different methods and compare them
- Add variable(s) for conditions in Central America

Sampling of Sources

- eBird. 2012. eBird: An online database of bird distribution and abundance [web application]. eBird, Ithaca, New York. Available: http://www.ebird.org. (Accessed: Mar. 2018).
- National Oceanic and Atmospheric Administration. Daily Summaries Map. https://gis.ncdc.noaa.gov/maps/ncei/summaries/daily (Mar. 2018).
- Myers, J. L., and A. D. Well. 2003. Research Design and Statistical Analysis.Mahwah (NJ): Lawrence Erlbaum Associates
- The Penn State University. 2018. STAT 501 Online Course: Regression Methods.
- Wherry, R. J. 1931. "A new formula for predicting the shrinkage of the coefficient of multiple correlation." *The annals of mathematical statistics 2*(4), 440-457.

Thank You!