Modeling III

Introduction

- Instructions
  - Download Tutorial
  - Unzip Folder
  - Required Packages
    - library(tidyverse)
    - library(modelr)
    - library(xtable)
  - Open .Rmd File and Knit
- Within R, Run all Code Chunks for Parts 1,2, and 3 (This was Covered in Previous Lecture)

Part 4:
Logistic Model

- Logistic Model

$$W = l + \frac{h}{1+e^{a-bA}} + \varepsilon$$

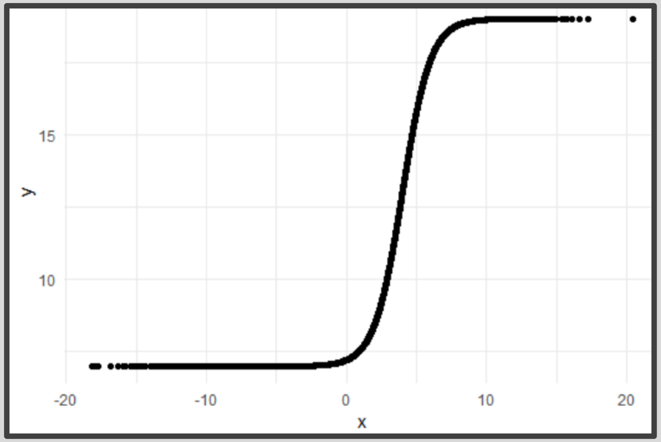
  - “Smart” Model Based On Physical Relationship Between A and W

  - Four Parameters
    - Controls the Shape of the Relationship
    - l and h
    - a and b

  - What Shape Do You Think This Function Makes?
    - Idea: Precalculus

Part 4:
Logistic Model

- Run Chunk 1
  - Plant that Seed
  - Example Model



  - Parameter Investigation
    - What Does 7 Represent?
    - What Does 12 Represent?
    - What Does 4 Represent?
    - What Does 1 Represent?

Part 4: Logistic Model

- Run Chunk 2
  - Creation of Modeling Function
  - Creation of MSE Function Specific to this Model
- Run Chunk 3
  - Use optim() Function With Smart Starting Values Based on Understanding of The Model
  - Finds Estimates Based on Minimization of MSE

Part 4: Logistic Model

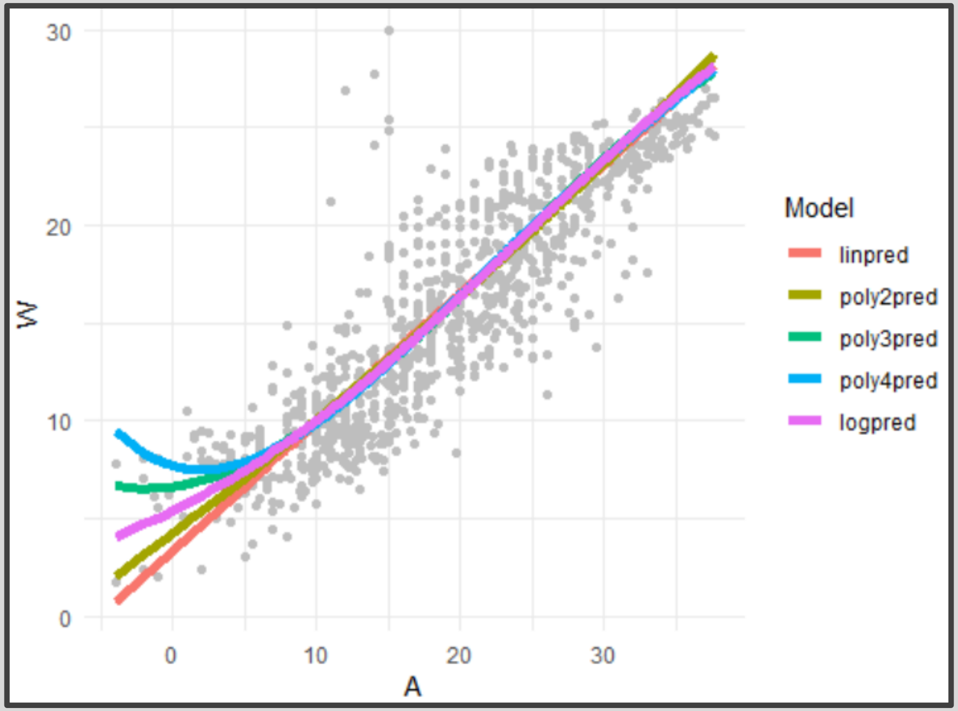
- Run Chunk 4
  - Use Logistic Model Function and Estimated Parameters from optim() to Obtain
    - Predictions
    - Residuals

Intermission

- Run Code Chunk
  - save.image() = Used to Save Workspace into .Rdata File
  - load() = Used to Load Workspace from .Rdata File
  - .Rdata = File Extension of R Workspace File (All Objects in Global Environment)

Part 5: Evaluation by Visualization

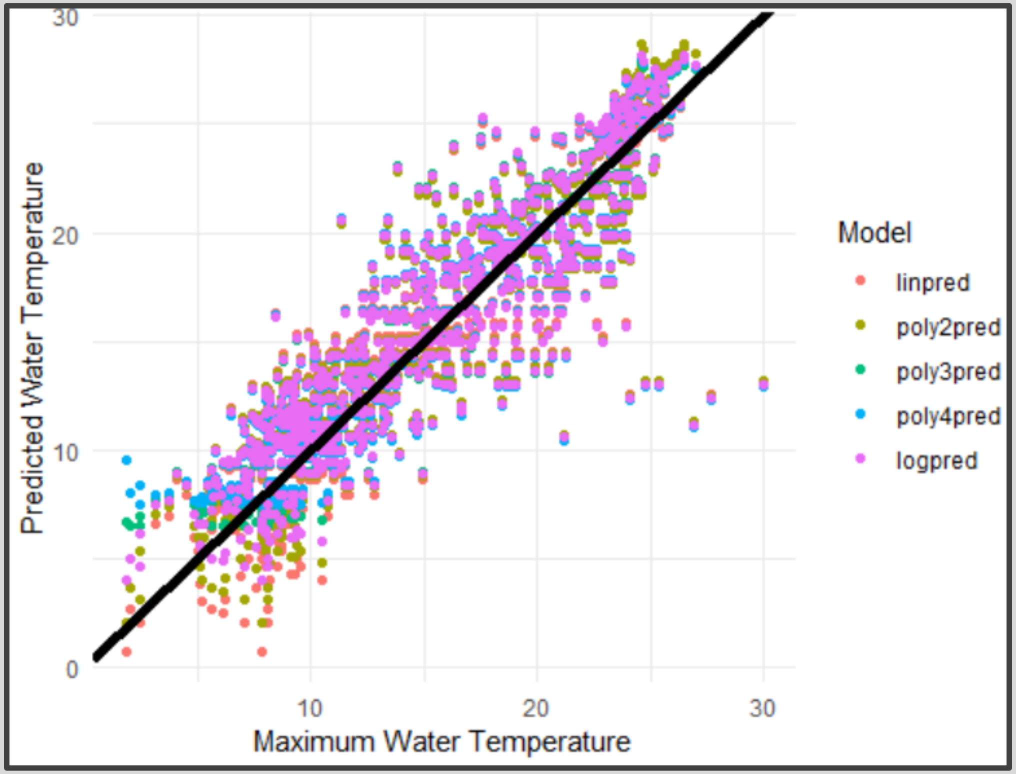
- Run Chunk 1
  - Plots of Different Models
  - What Can We Say About the Different Models?



  - Which Model Would You Use?

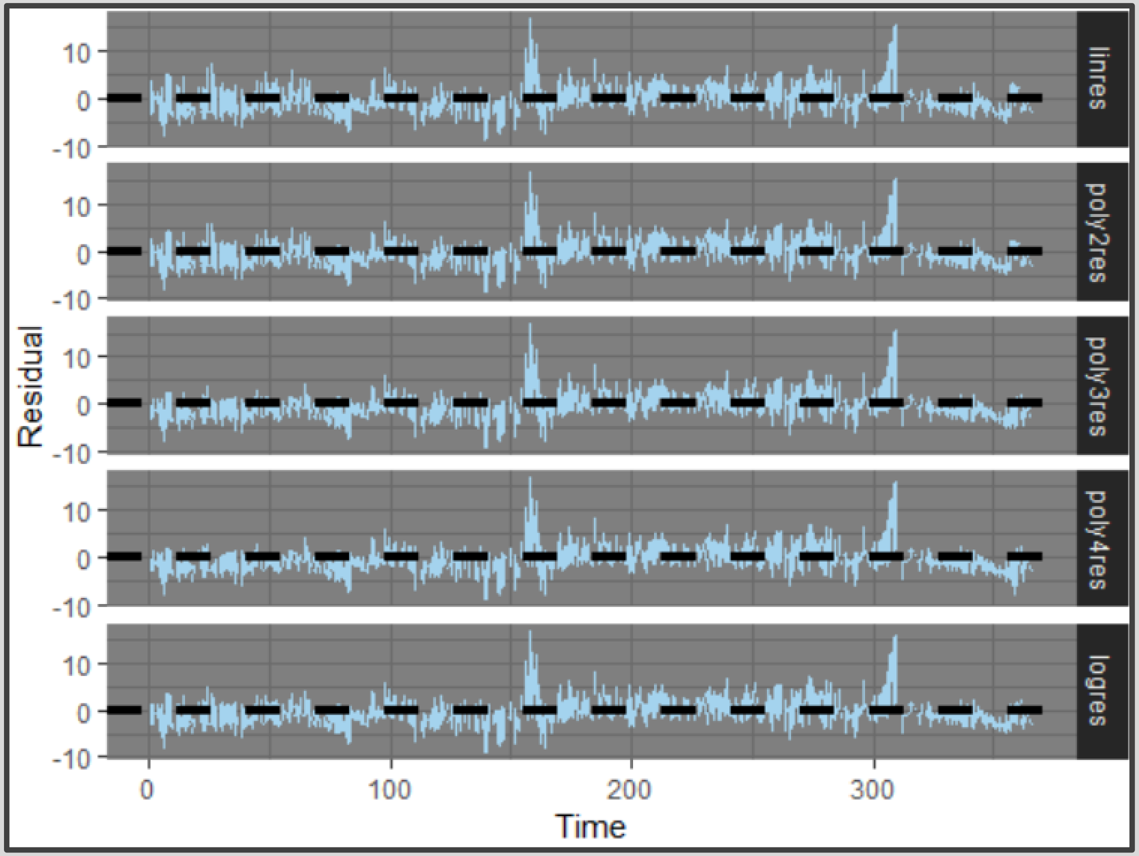
Part 5: Evaluation by Visualization

- Run Chunk 2
  - Comparing Predictions vs Actual Maximum Water Temperatures
  - Models Give Similar Predictions



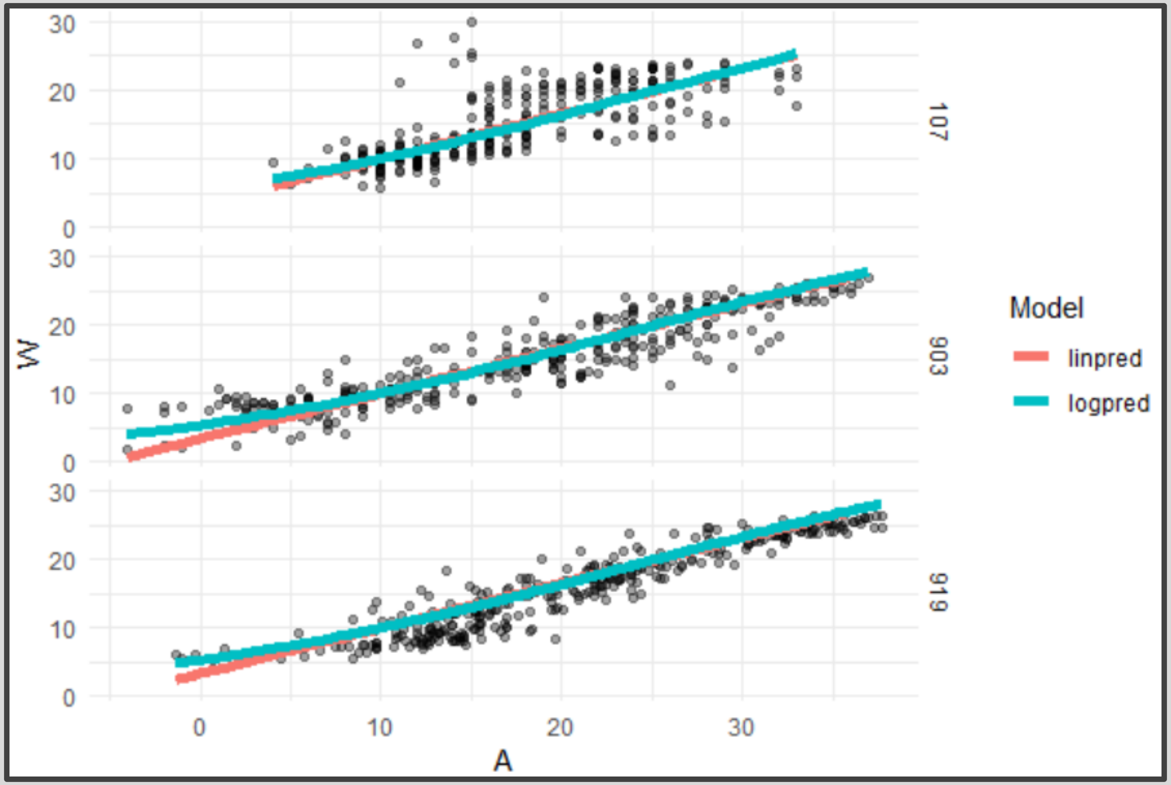
Part 5: Evaluation by Visualization

- Run Chunk 3
  - Shows Residuals Under the 4 Models Plotted Over Time
  - What is the Problem?



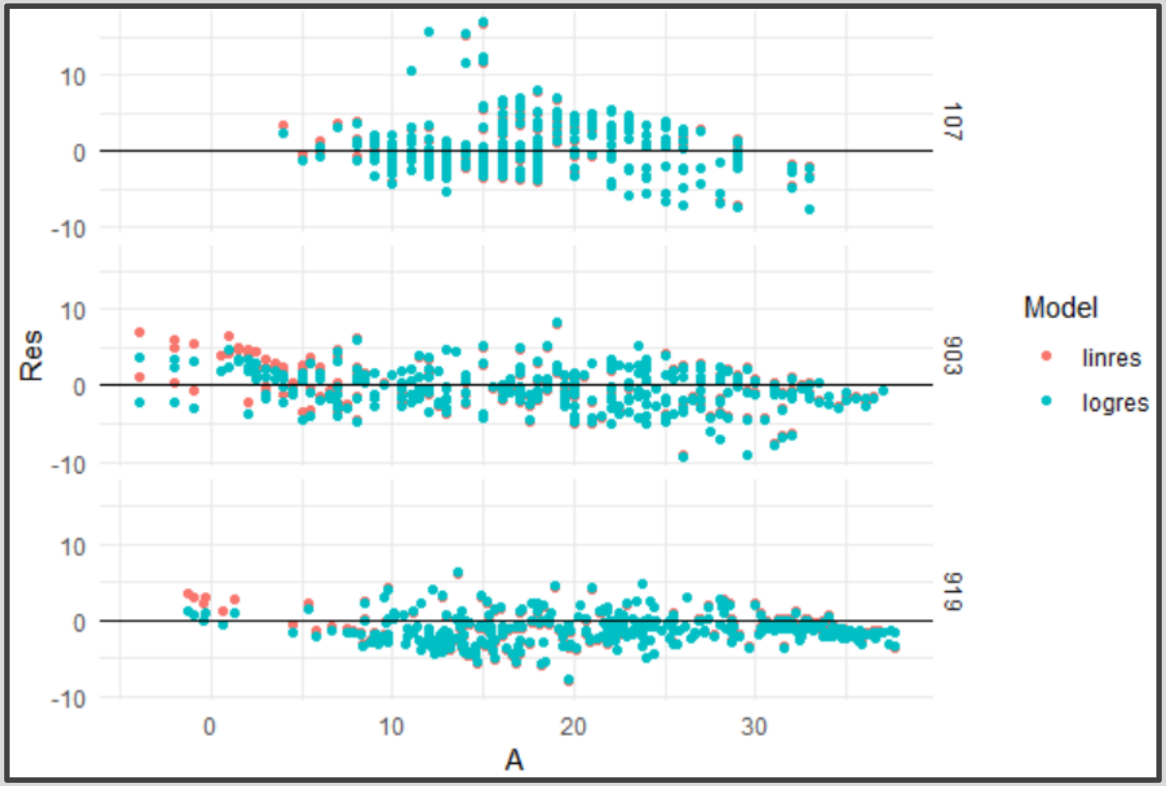
Part 5: Evaluation by Visualization

- Run Chunk 4
  - Evaluate Models For the Three Locations Separately



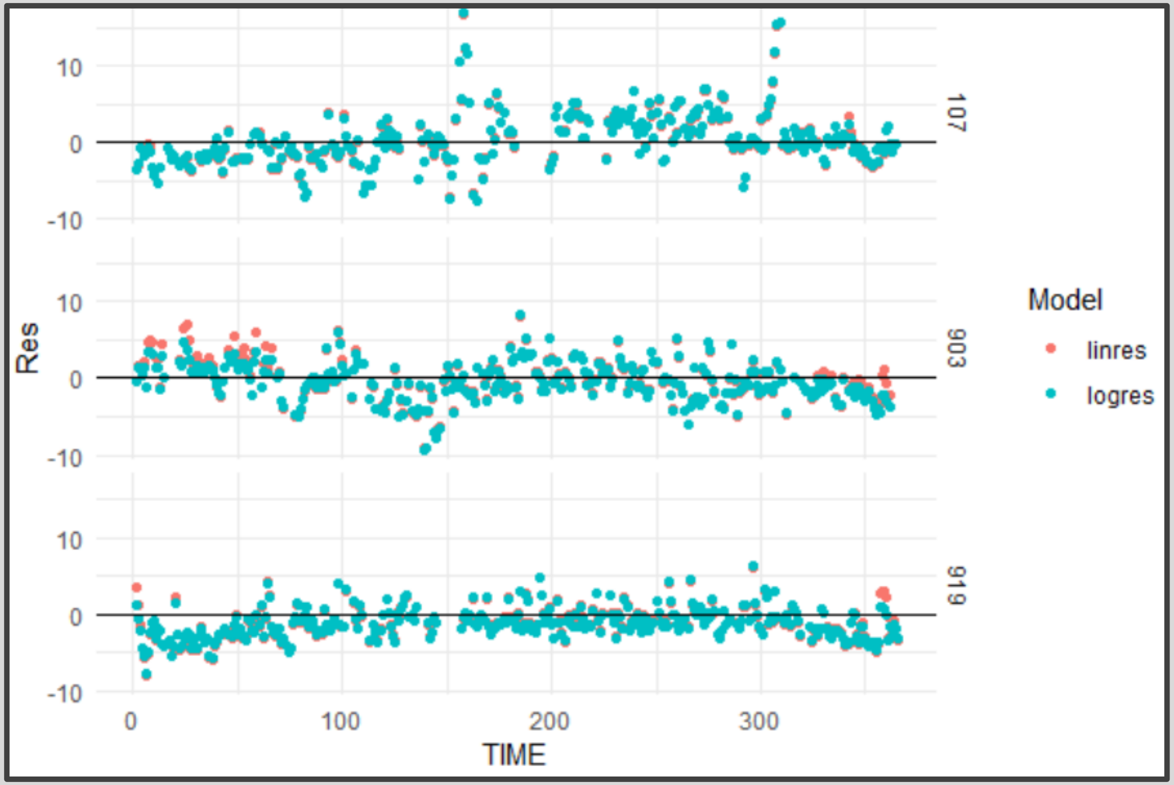
Part 5: Evaluation by Visualization

- Run Chunk 5
  - Evaluate Error For the Three Locations Separately (by A)



Part 5:
Evaluation by Visualization

- Run Chunk 6
  - Evaluate Error For the Three Locations Separately (by Time)



Part 6: Evaluation by Numerical Summary

- Run Chunk 1
  - Mean Bias
    $$\mathrm{MB} = \frac{1}{N}\sum \hat{\varepsilon}_k$$
  - Mean Absolute Error
    $$\mathrm{MAE} = \frac{1}{N}\sum |\hat{\varepsilon}_k|$$
  - Root Mean Squared Error
    $$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum \hat{\varepsilon}_k^2}$$
  - MB, MAE, and RMSE are in Degrees Celsius

Part 6: Evaluation by Numerical Summary

- Summarizing Table
  - Evaluate MB, MAE, and RMSE on Test Data to Choose Best Model Going Forward
  - Sketch of Table We Want

Model	MB	MAE	RMSE
Linear			
Poly(2)			
Poly(3)			
Poly(4)			
Logistic			

  - Before Writing Code, Have a Plan for the Output

Part 6:
Evaluation by
Numerical
Summary

- Chunk 2
  - Run Line-By-Line
  - Think About Ways to Quickly Apply All 3 Functions to All Residuals
- Run Chunk 3
  - Combine rename(), gather(), group_by(), and summarize()
- Chunk 4
  - Change eval=F to eval=T and Knit the File (What is Seen?)

Part 6: Evaluation by Numerical Summary

- My Results Based on My Seed

Model	MB	MAE	RMSE
<fct>	<dbl>	<dbl>	<dbl>
Linear	-0.350	2.18	2.87
Poly(2)	-0.387	2.17	2.86
Poly(3)	-0.466	2.11	2.82
Poly(4)	-0.492	2.10	2.83
Logistic	-0.426	2.13	2.83

- When Results Are This Close, Always Consider the Most Simple Model

Closing

Disperse and Make Reasonable Decisions