Modeling II

Introduction

- Instructions

  - Download Tutorial 11 Zip
  - Unzip Folder
  - Required Packages
    - `library(tidyverse)`

    - `library(modelr)`
  - Open .Rmd File and Knit

- Daily Spanish River Data

  - W = Max Water Temperature
  - A = Max Air Temperature
  - L = River Identifier (31 Rivers)

Introduction

- Questions About RMarkdown

  - What Does the Following Code Do When Knitted?

    `` `r length(unique(DATA$L))` ``

  - What Does the following Code Chunk Option Do When Knitted?

    ```
    echo=F
    ```
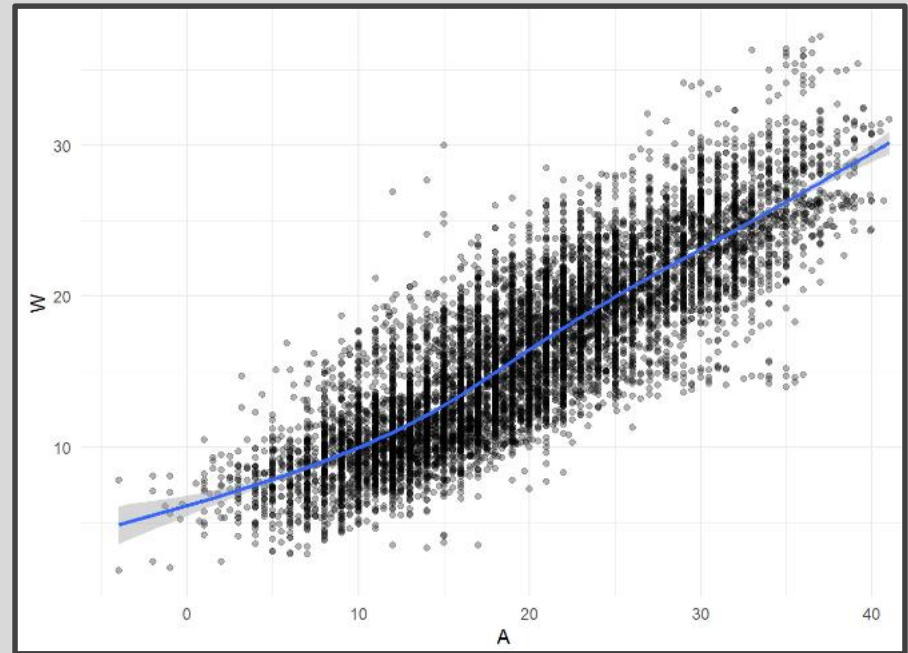
Introduction

- Goal: Build a Model to Predict Max Water Temp Given Max Air Temp

  - What Do You Know About the Relationship of These Variables?

  - Who Would Care About this Relationship?

  - Why Would Someone Want to Predict the Max Water Temp?

  - Why Would this Model Be Useful?

Part 1: Examining the Relationship

- Run Chunk 1

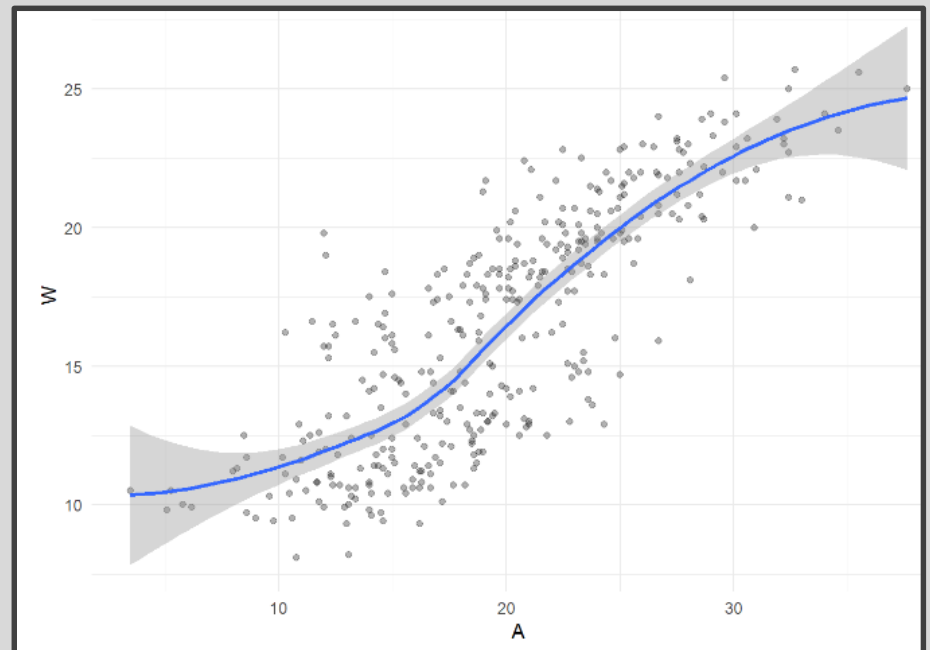 - What Do You Notice About the Overall Relationship?

  - Do You Think This Relationship is the Same for All Locations?

  - Why?   message=F

Part 1: Examining the Relationship

- Run Chunk 2

  - Location is a Numeric Variable
  - What Do You Notice About the Relationship for L==103?

- What do You Notice Now?

Part 1: Examining the Relationship

- Chunk 2 Discussion

  - What are the Differences in the Relationship Between W and A for the Various Locations?

  - Why do You Think These Differences Exist?

  - How do You Suggest We Handle the Differences?
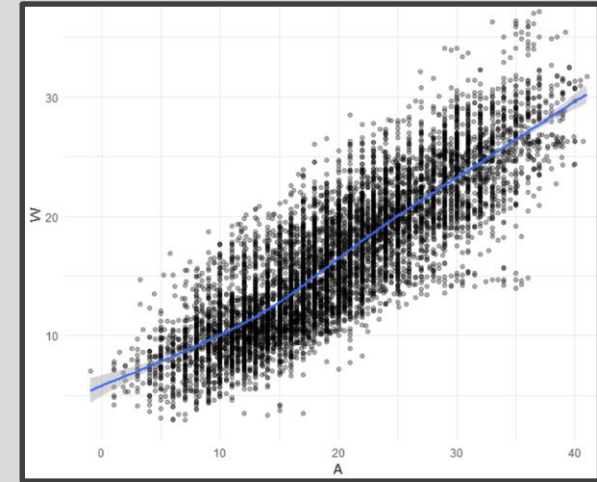
Part 1: Examining the Relationship

- Chunk 3

  - Randomly Samples 3 Locations

  - Plant Your Seed and Run Code

  - Usage:
    - anti_join()
    - semi_join()

  - Why Don't We Handpick the Three Locations?

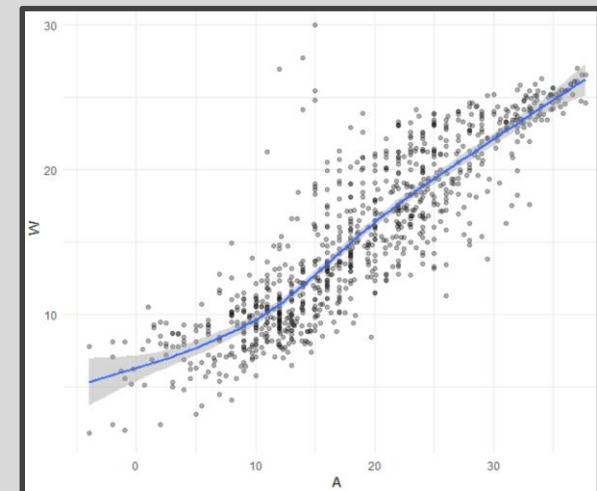# Part 1: Examining the Relationship

- Run Chunk 4

  - Train Plot

  - Test Plot

Part 2:
Linear Model

- Linear Model

$$W = a + bA + \varepsilon$$

  - Simplest Relationship that is Easily Explained

  - For every 1 Degree Change in A, W changes by *b* Degrees

  - When A=0 Degrees, the Expected Water Temperature is *a* Degrees

Part 2:
Linear Model

- Run Chunk 1

    - Fits Linear Model to Train Data
    - What is Your Intercept?
    - What is Your Slope?

- Run Chunk 2

    - Saves Predictions to Train/Test

    `add_predictions(MODEL,var="NAME")`

- Run Chunk 3

    - Saves Residuals to Train/Test

    `add_residuals(MODEL,var="NAME")`

Part 3:
Polynomial Model

- Polynomial Model

$$W = a + b_1 A + b_2 A^2 + \cdots + b_k A^k + \varepsilon$$

- "Feature Engineering"

- Generalized Additive Model

- Geom_smooth() Fits a GAM when Fitting a Curve

- Useful for Approximating Nonlinear Relationships

- Dependent on Degree "k"

- Goal: Choose Best "k"

Part 3: Polynomial Model

- Formula Object in R

  - Special Notation
  - Helpful Table:

Symbol	Example	Meaning
+	+X	include this variable
−	−X	delete this variable
:	X:Z	include the interaction between these variables
*	X*Y	include these variables and the interactions between them
|	X | Z	conditioning: include x given z
^	(X + Z + W)^3	include these variables and all interactions up to three way
I	I(X*Z)	as is: include a new variable consisting of these variables multiplied
1	X − 1	intercept: delete the intercept (regress through the origin)

- We will Use the I() Function to Create New Variables Based Off Variables We Have

Part 3:
Polynomial Model

- Run Chunk 1

  - Fits 2nd Degree Polynomial
  - Fits 3rd Degree Polynomial
  - Fits 4th Degree Polynomial

- Run Chunk 2

  - Obtains Predictions Under the Different Polynomial Models

Part 3: Polynomial Model

- Chunk 3

  - Code Needs Modification
  - Highlight Code

```
TRAIN4 =TRAIN3 %>%
    add_predictions(poly2mod,var="poly2pred") %>%
    add_predictions(poly3mod,var="poly3pred") %>%
    add_predictions(poly4mod,var="poly4pred")

TEST4 =TEST3 %>%
    add_predictions(poly2mod,var="poly2pred") %>%
    add_predictions(poly3mod,var="poly3pred") %>%
    add_predictions(poly4mod,var="poly4pred")
```

- TRAIN3 -> TRAIN4 and etc.
- Use Ctrl+F (Find and Replace)
  - 'predictions' -> 'residuals'
  - 'pred"' -> 'res"'

Q predictions	Next	Prev	All	residuals	Replace	All

☑ In selection ☐ Match case ☐ Whole word ☐ Regex ☑ Wrap

- Run Chunk 3 After Modifying