

Nonlinear Regression & Application to Leukemia Incidence Data

Alejandro Gonzalez

Advisor: Michael Pearce

UNIVERSITY *of* WASHINGTON



What is Nonlinear Regression?

- **Linear Regression:**
 - Fits a “straight line” model between response variable and covariate(s).
 - Picks the line with smallest sum of squared residuals
- **Nonlinear Regression:**
 - Can break the assumptions of LR
 - Can fit models that aren't straight lines
 - Very flexible

Leukemia Incidence Rates

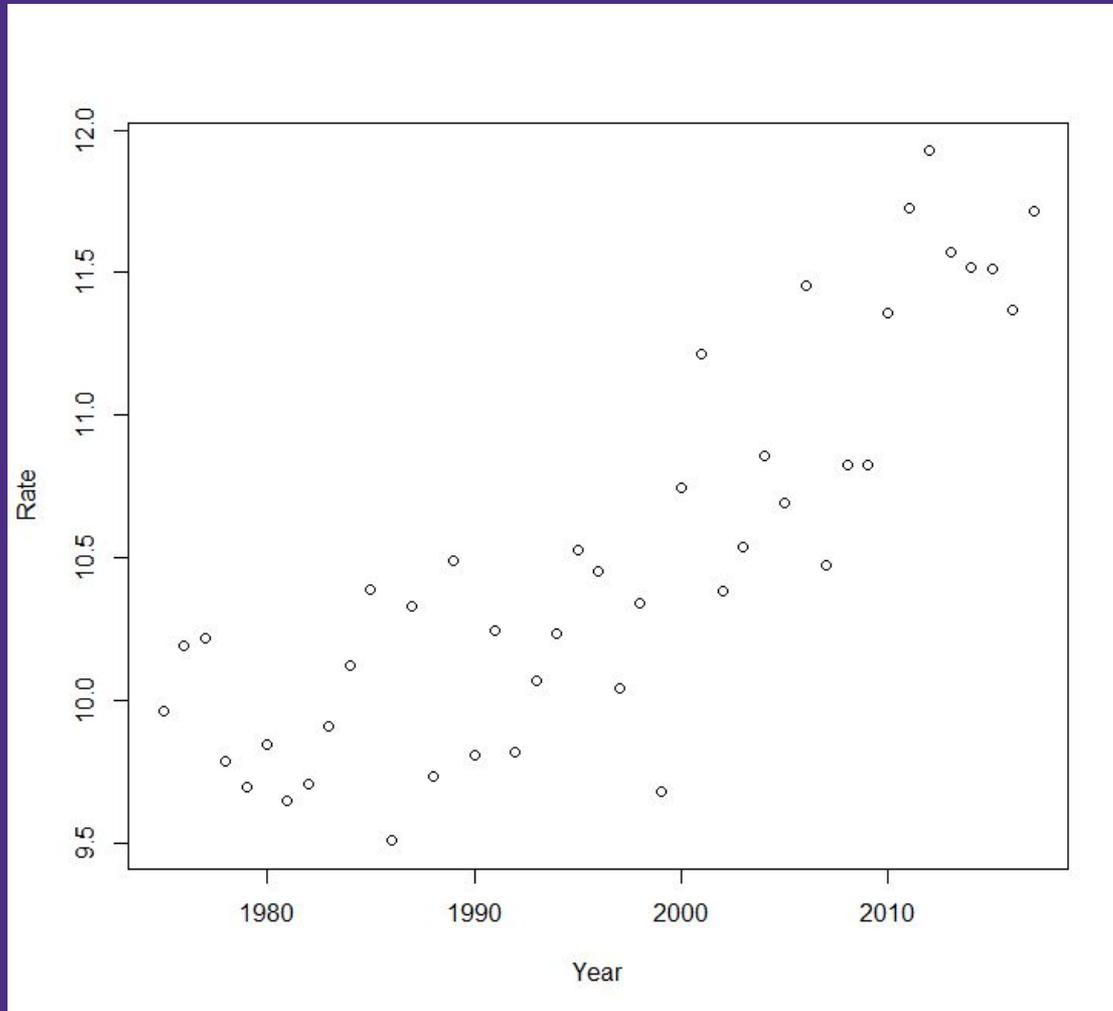
- **Investigating Female Incidence Rates**
 - On Year, Age, & Ethnicity
 - 1975-2017
- **Exploratory Data Analysis**
 - Emphasis on model comparisons
 - Leukemia incidence changes over time in females
- **UW Undergraduate Symposium**



NATIONAL CANCER INSTITUTE

Surveillance, Epidemiology, and End Results Program

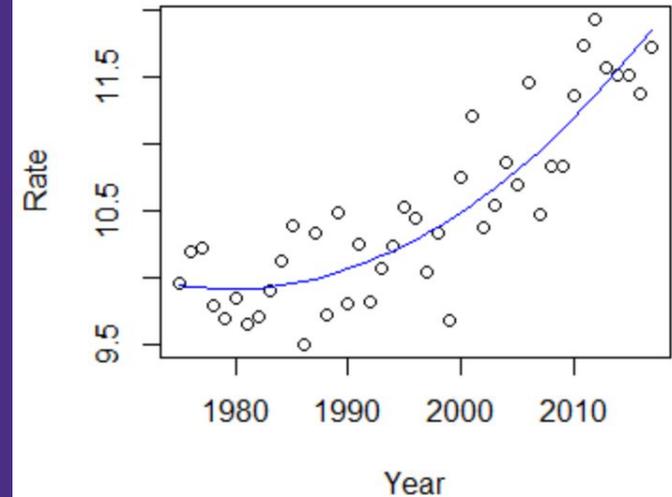
What Type of Model Would You Use for this Data?



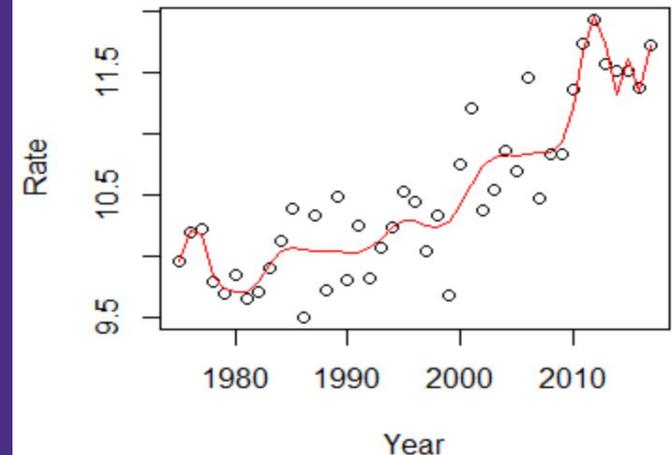
Polynomial Regression

- **Extension of Linear Regression**
 - Quadratics, Cubics, Quartics...
 - Accounts for changes in relationship
 - ANOVA: Quadratic best
- **Use:**
 - Polynomial Growth
 - Population Dynamics
 - Bounded Functions

2 Degree Polynomial

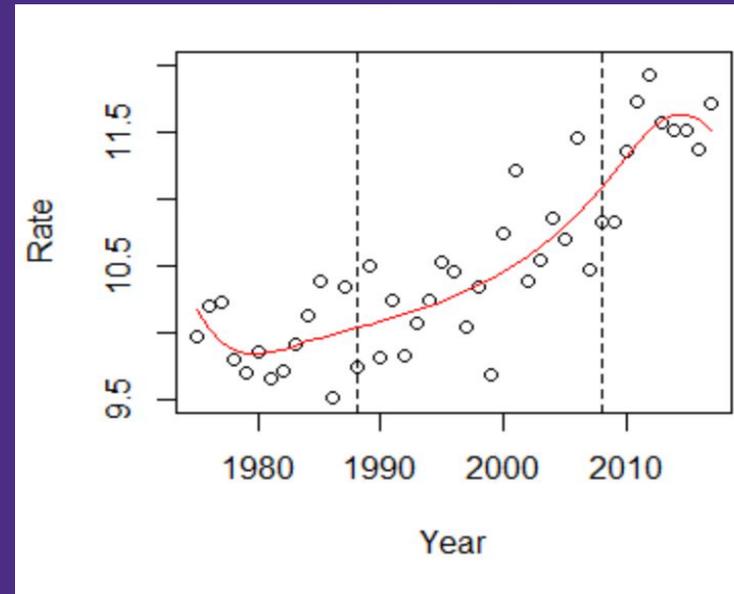


20 Degree Polynomial



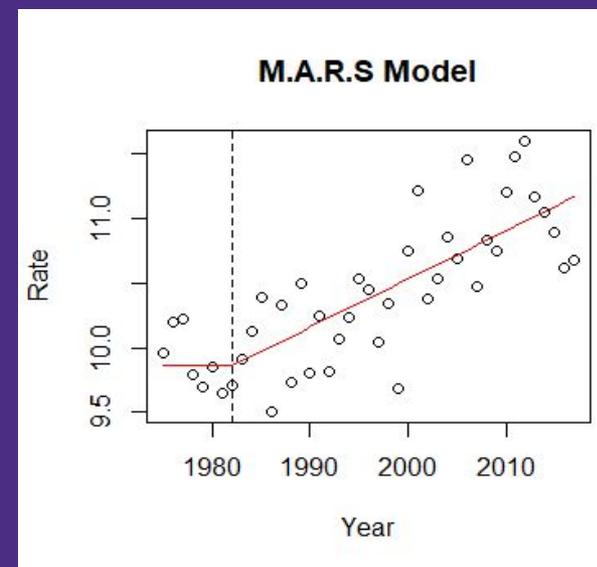
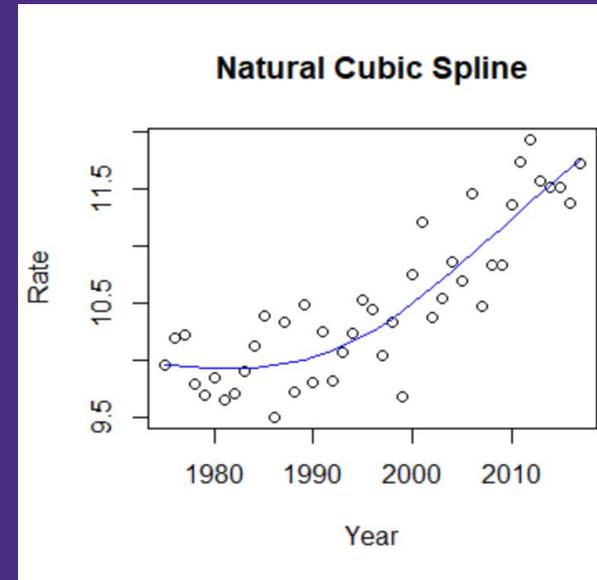
The Family of Splines

- **General Concept**
 1. **Divide covariates into sections**
 2. **Fit separate model in each section**
 3. **Optional: Add constraints**
- **Key Ideas:**
 - a. **Knots**
 - The place where we divide
 - More knots = more flexibility
 - Statistician vs. computer
 - b. **Constraints**
 - Different constraints lead to different model types

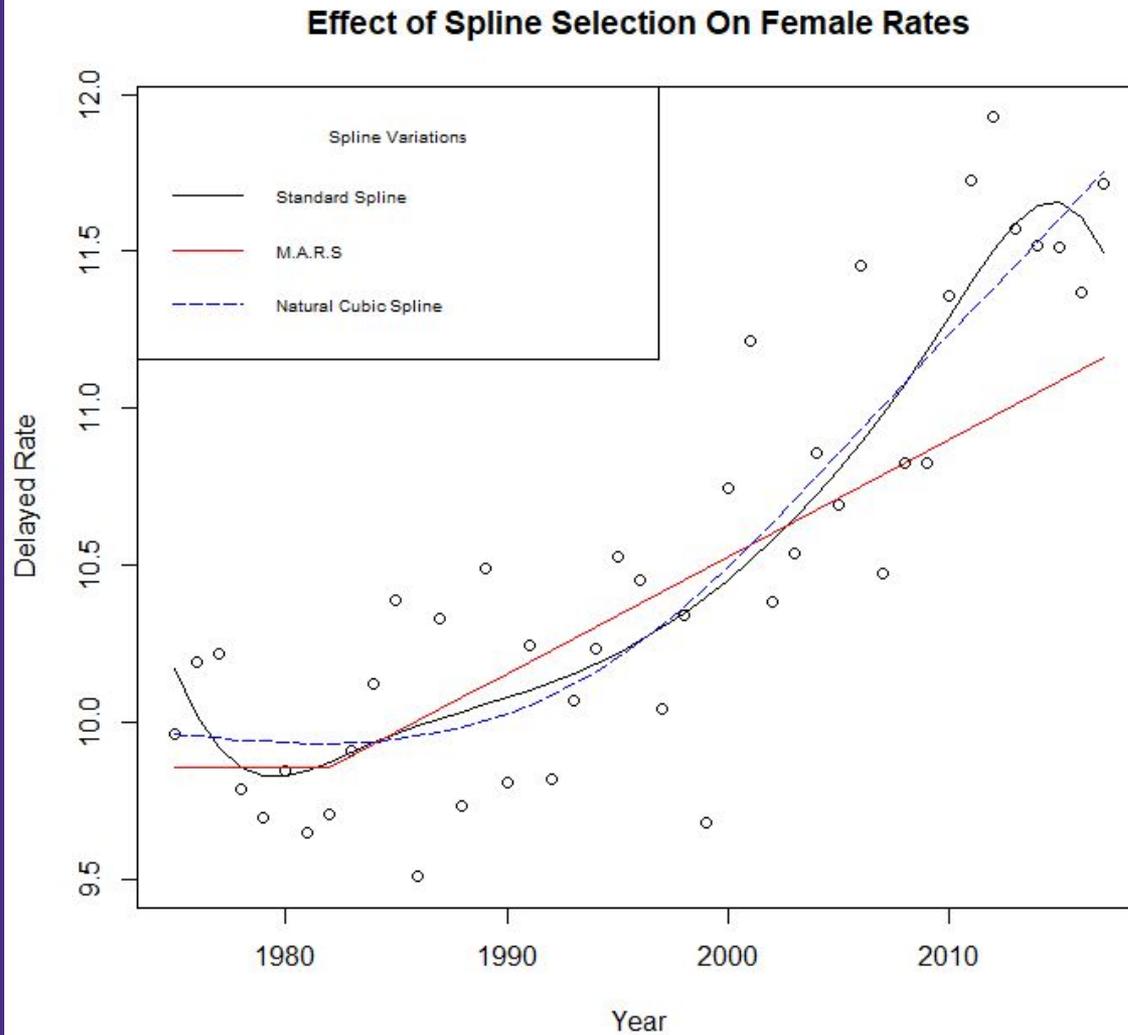


Examples of Spline Models

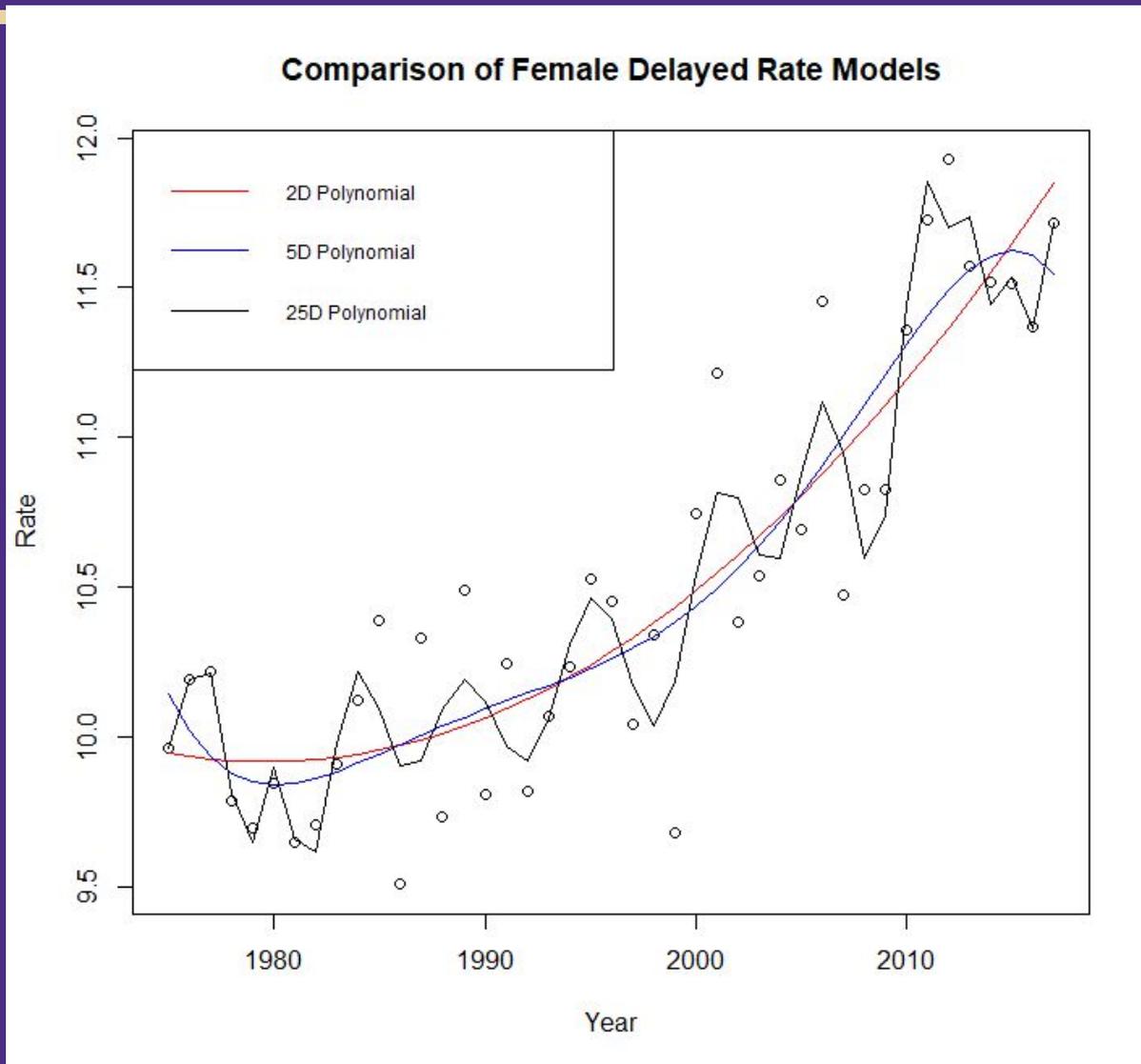
- Natural Spline
 - Constraints straight lines at the end
 - More predictable
- Smoothing Spline (My Favorite)
 - Stable, Gradual, and Sturdy
 - Punishes fast increases & “roughness”
- M.A.R.S Spline
 - Picks the knots points for you
 - Only straight lines



Cross Comparisons



Cross Comparisons



Choosing a model

- **Context Knowledge**
 - Does much make sense?
- **Statistical Tests**
 - ANOVA, AIC, BIC, Residual Plots
- **Cross Validation**
 - LOOCV, GCV, Forward Selection

Questions?



Thank you!

W