Moving Beyond Linearity

Tianhang Zheng

https://tianzheng4.github.io

Moving Beyond Linearity

The truth is usually not linear.

It may be

- polynomials,
- step functions,
- splines,
- local regression, and
- generalized additive models

Polynomial Regression

$$\hat{f}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 + \hat{\beta}_2 x_0^2 + \hat{\beta}_3 x_0^3 + \hat{\beta}_4 x_0^4.$$

Since $\hat{f}(x_0)$ is a linear function of the $\hat{\beta}_\ell$, can get a simple expression for *pointwise-variances* $\text{Var}[\hat{f}(x_0)]$ at any value x_0. In the figure we have computed the fit and pointwise standard errors on a grid of values for x_0. We show $\hat{f}(x_0) \pm 2 \cdot \text{se}[\hat{f}(x_0)]$.

Logistic Regression

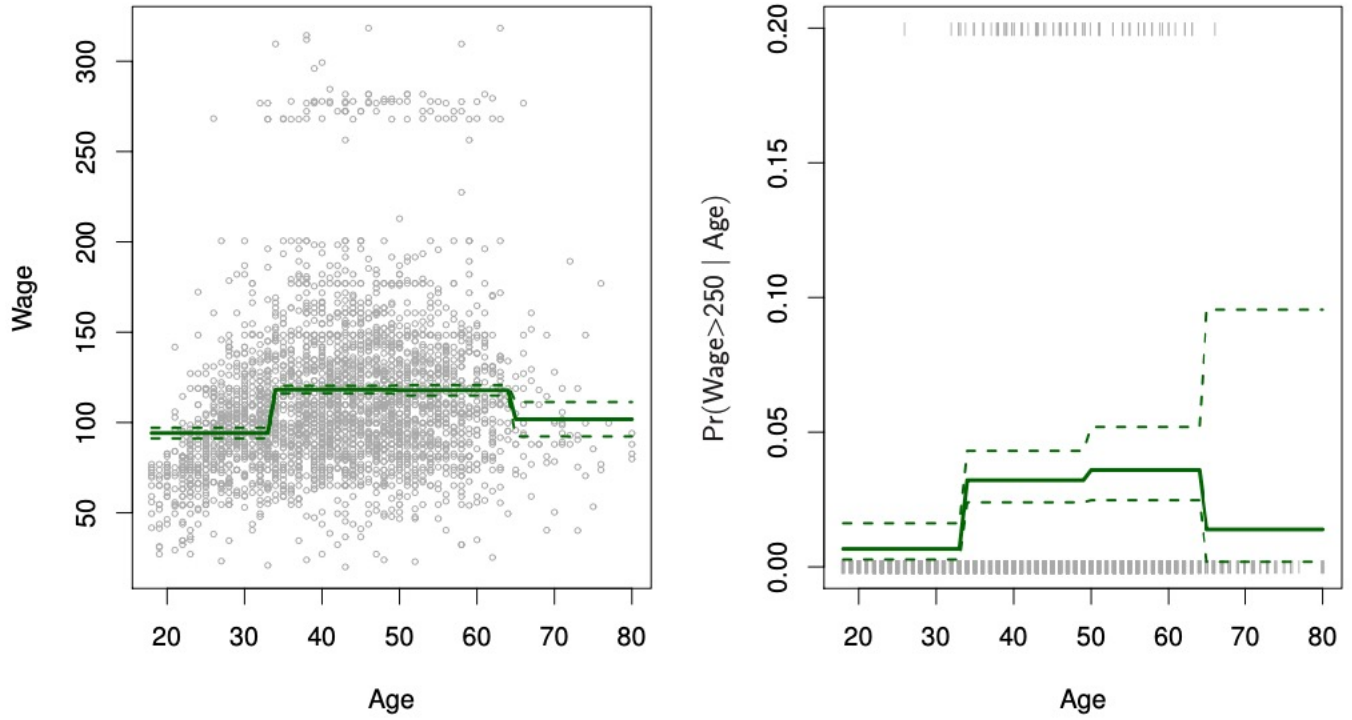
$$\Pr(y_i > 250|x_i) = \frac{\exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_d x_i^d)}{1 + \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_d x_i^d)}.$$

To get confidence intervals, compute upper and lower bounds on *on the logit scale*, and then invert to get on probability scale.

Step Functions

Another way of creating transformations of a variable — cut the variable into distinct regions.

$$C_1(X) = I(X < 35), \quad C_2(X) = I(35 \leq X < 50), \ldots, C_3(X) = I(X \geq 65)$$

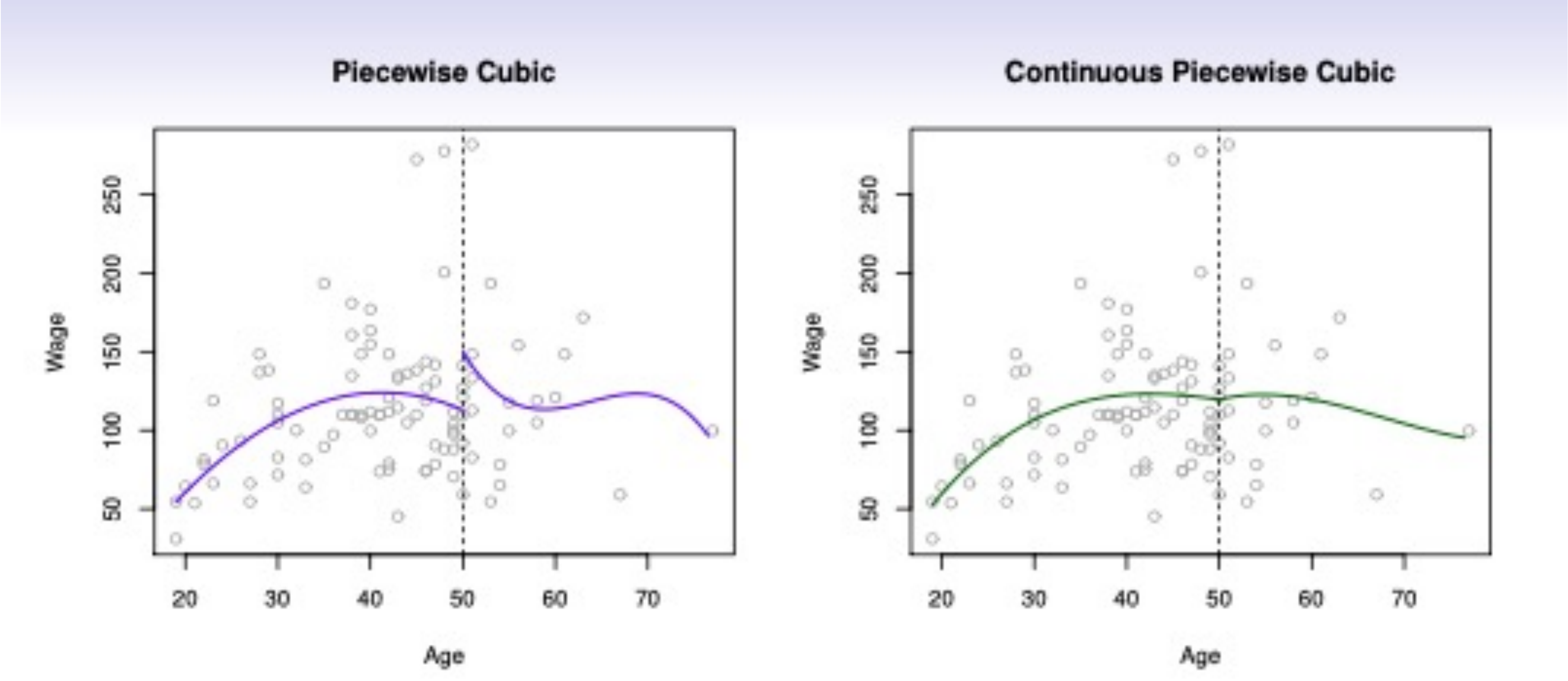


Piecewise Polynomials

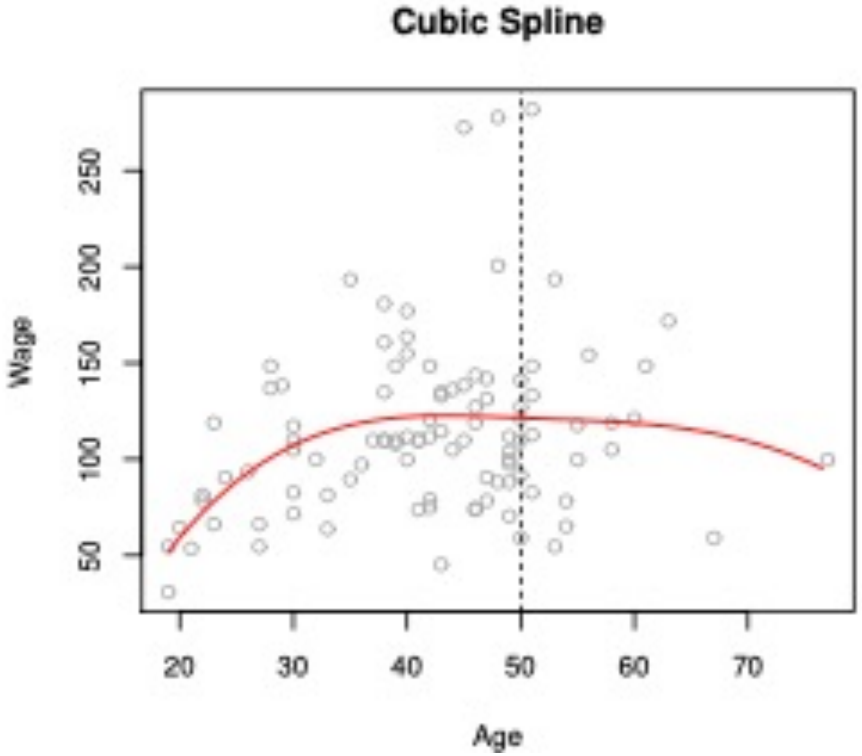
Instead of a single polynomial in X over its whole domain, we can rather use different polynomials in regions defined by knots.

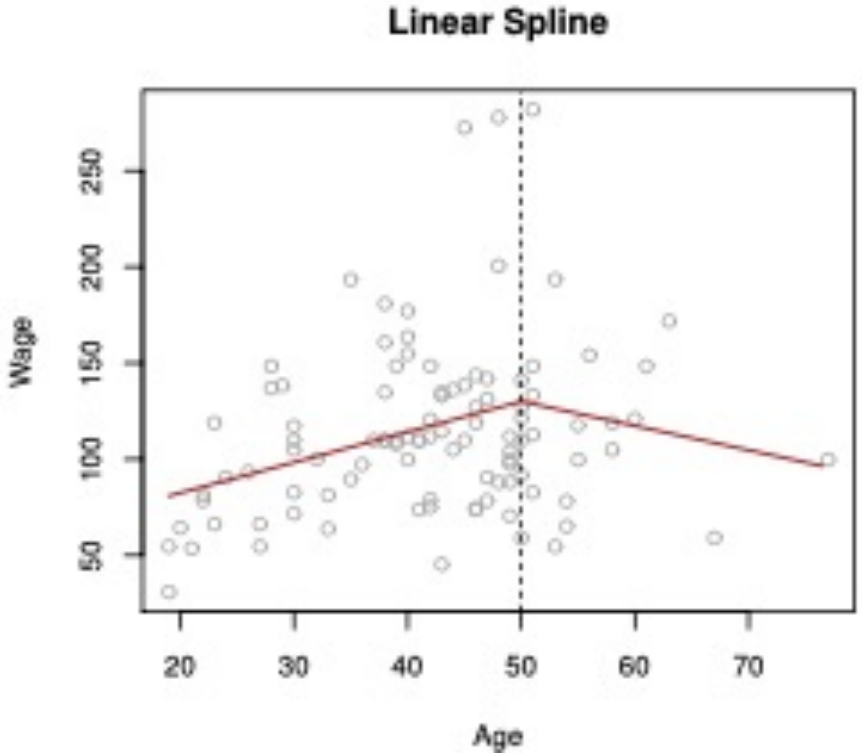
$$y_i = \begin{cases} \beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2 + \beta_{31}x_i^3 + \epsilon_i & \text{if } x_i < c; \\ \beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2 + \beta_{32}x_i^3 + \epsilon_i & \text{if } x_i \geq c. \end{cases}$$

Piecewise Polynomials



Piecewise Polynomials





Linear Splines

A linear spline with knots at ξ_k, k = 1,...,K is a piecewise linear polynomial continuous at each knot. We can represent this model as

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \cdots + \beta_{K+1} b_{K+1}(x_i) + \epsilon_i$$

$$\begin{aligned} b_1(x_i) &= x_i \\ b_{k+1}(x_i) &= (x_i - \xi_k)_+, \quad k = 1, \ldots, K \end{aligned}$$

$$(x_i - \xi_k)_+ = \begin{cases} x_i - \xi_k & \text{if } x_i > \xi_k \\ 0 & \text{otherwise} \end{cases}$$

Cubic Splines

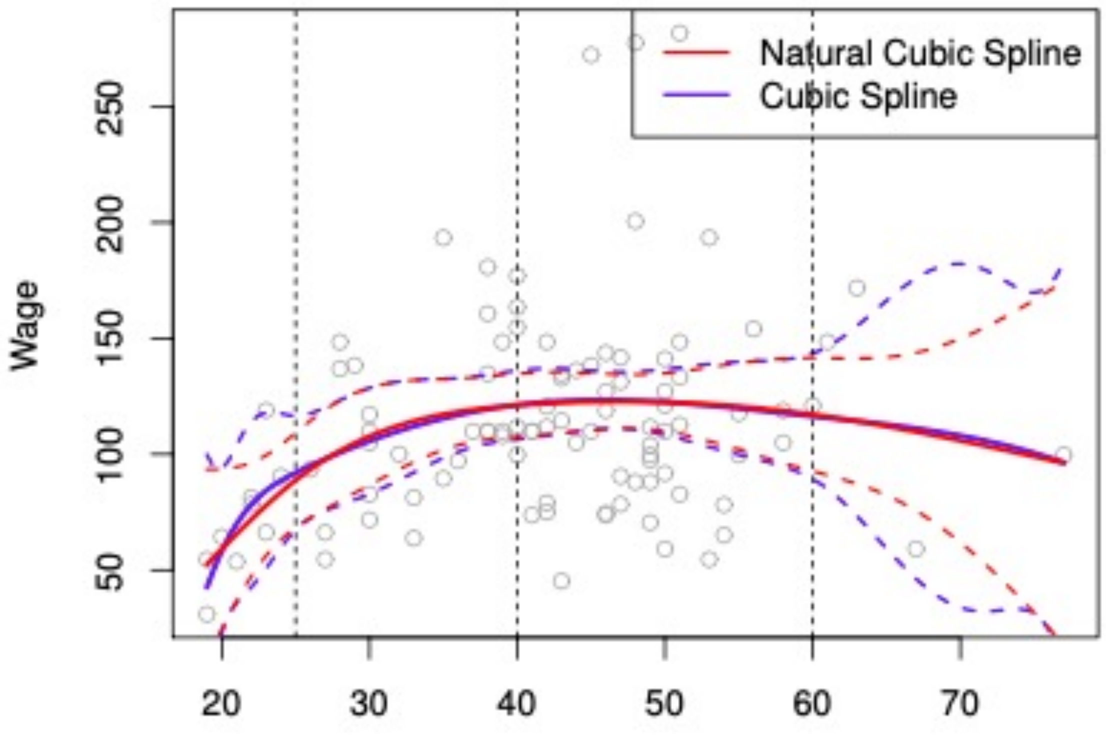
$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \cdots + \beta_{K+3} b_{K+3}(x_i) + \epsilon_i$$

$$\begin{aligned} b_1(x_i) &= x_i \\ b_2(x_i) &= x_i^2 \\ b_3(x_i) &= x_i^3 \\ b_{k+3}(x_i) &= (x_i - \xi_k)_+^3, \quad k = 1, \ldots, K \end{aligned}$$

$$(x_i - \xi_k)_+^3 = \begin{cases} (x_i - \xi_k)^3 & \text{if } x_i > \xi_k \\ 0 & \text{otherwise} \end{cases}$$

Natural Cubic Splines

A natural cubic spline extrapolates linearly beyond the boundary knots, which adds $4 = 2{\times}2$ constraints



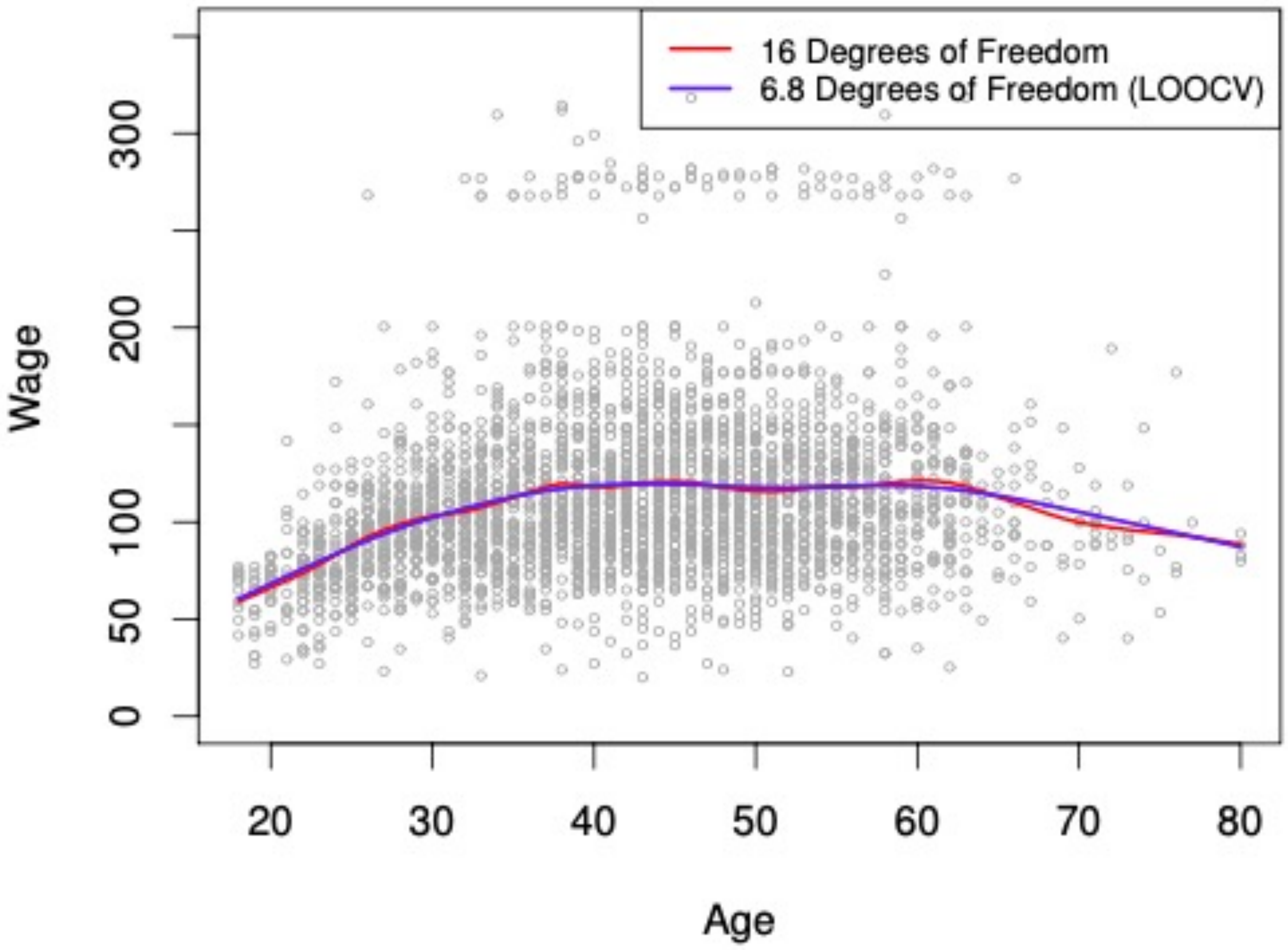
Smoothing Splines

Objective for smoothing splines

$$\underset{g \in \mathcal{S}}{\text{minimize}} \sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int g''(t)^2 dt$$

The second term is a roughness penalty and controls how wiggly g(x) is. It is modulated by the tuning parameter.

Smoothing Splines



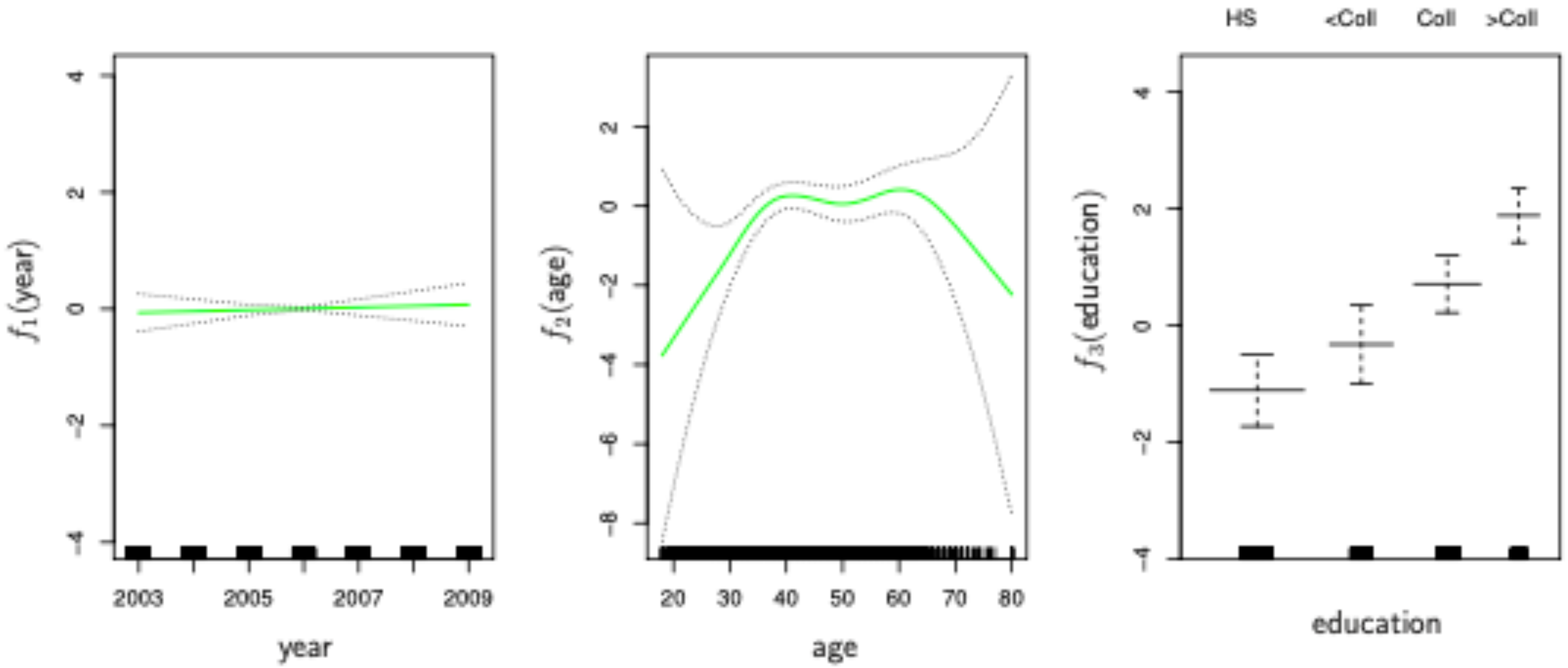
Generalized Additive Models (GAM)

Allows for flexible nonlinearities in several variables, but retains the additive structure of linear models.

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \epsilon_i.$$

GAMs for classification

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p).$$



Tree-based Methods

Tree-based methods can be used for regression and classification

Stratifying or segmenting the predictor space into a number of simple regions.

Since the set of splitting rules used to segment the predictor space can be summarized in a tree, these types of approaches are known as decision-tree methods.

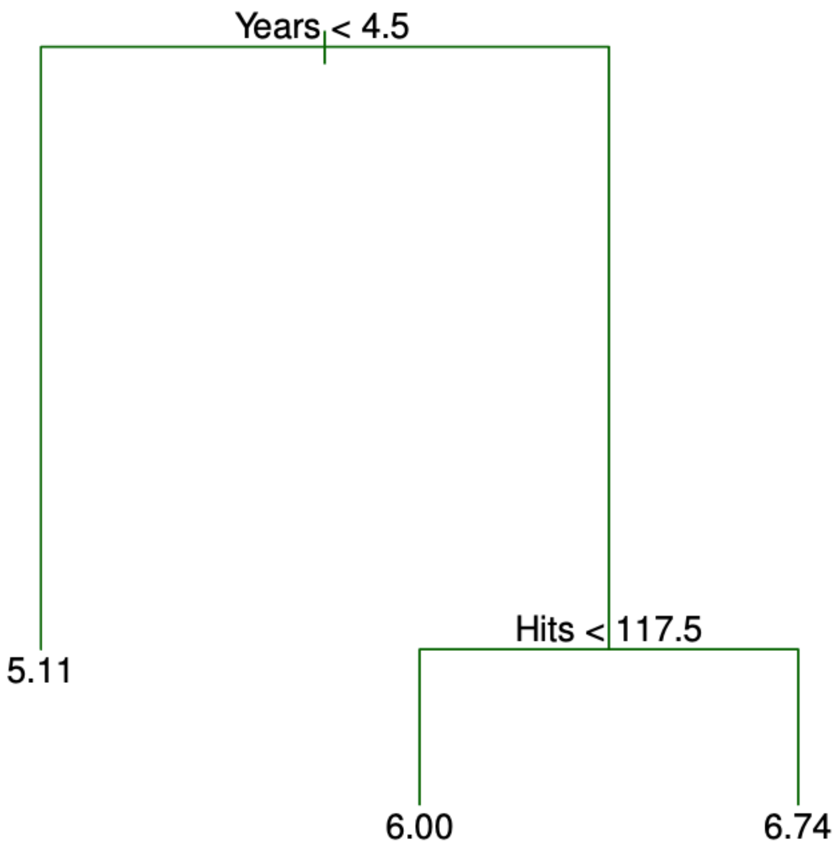
Pros and Cons

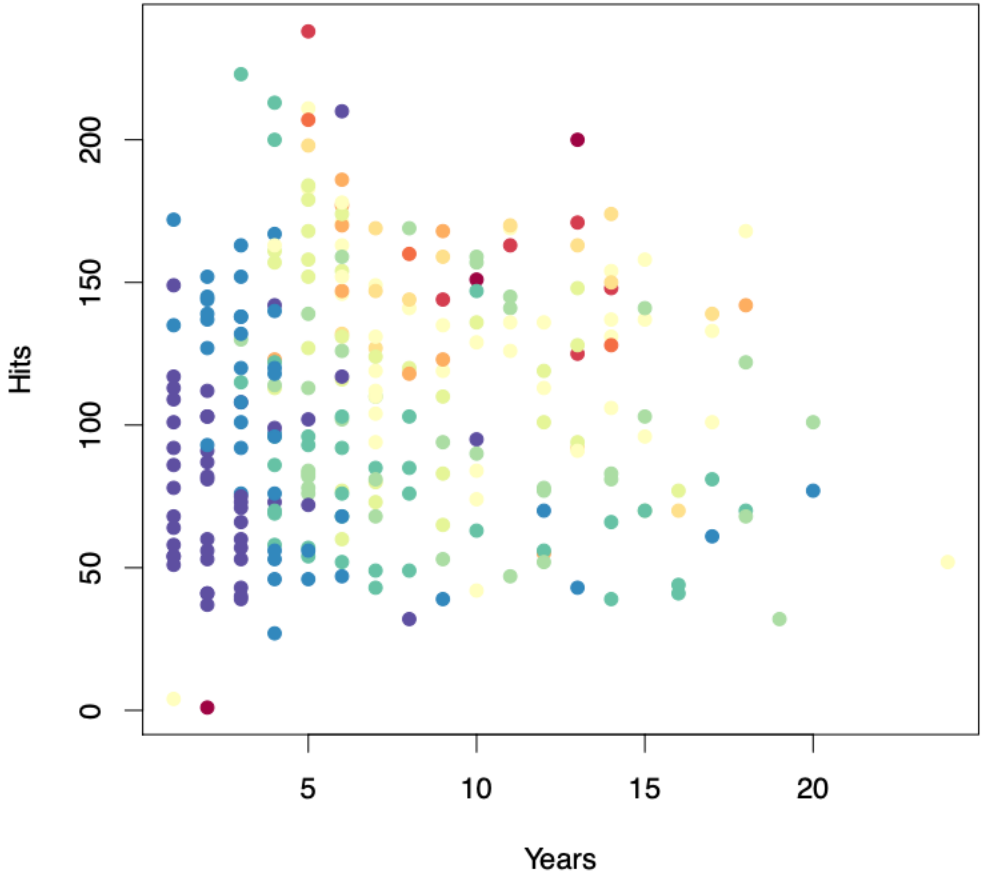
Tree-based methods are simple and useful for interpretation.

We will also discuss methods that can grow multiple trees which are then combined to yield a single consensus prediction

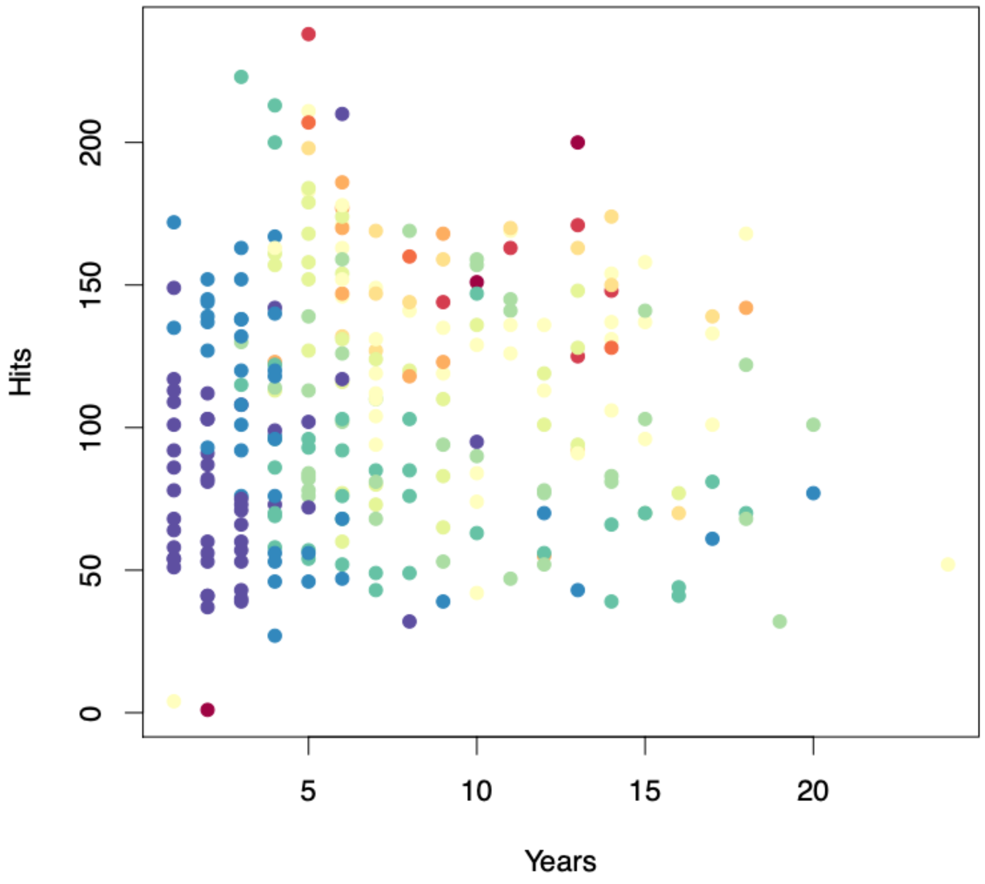
Combining a large number of trees can often result in dramatic improvements in prediction accuracy, at the expense of some loss interpretation.

Decision Tree Example

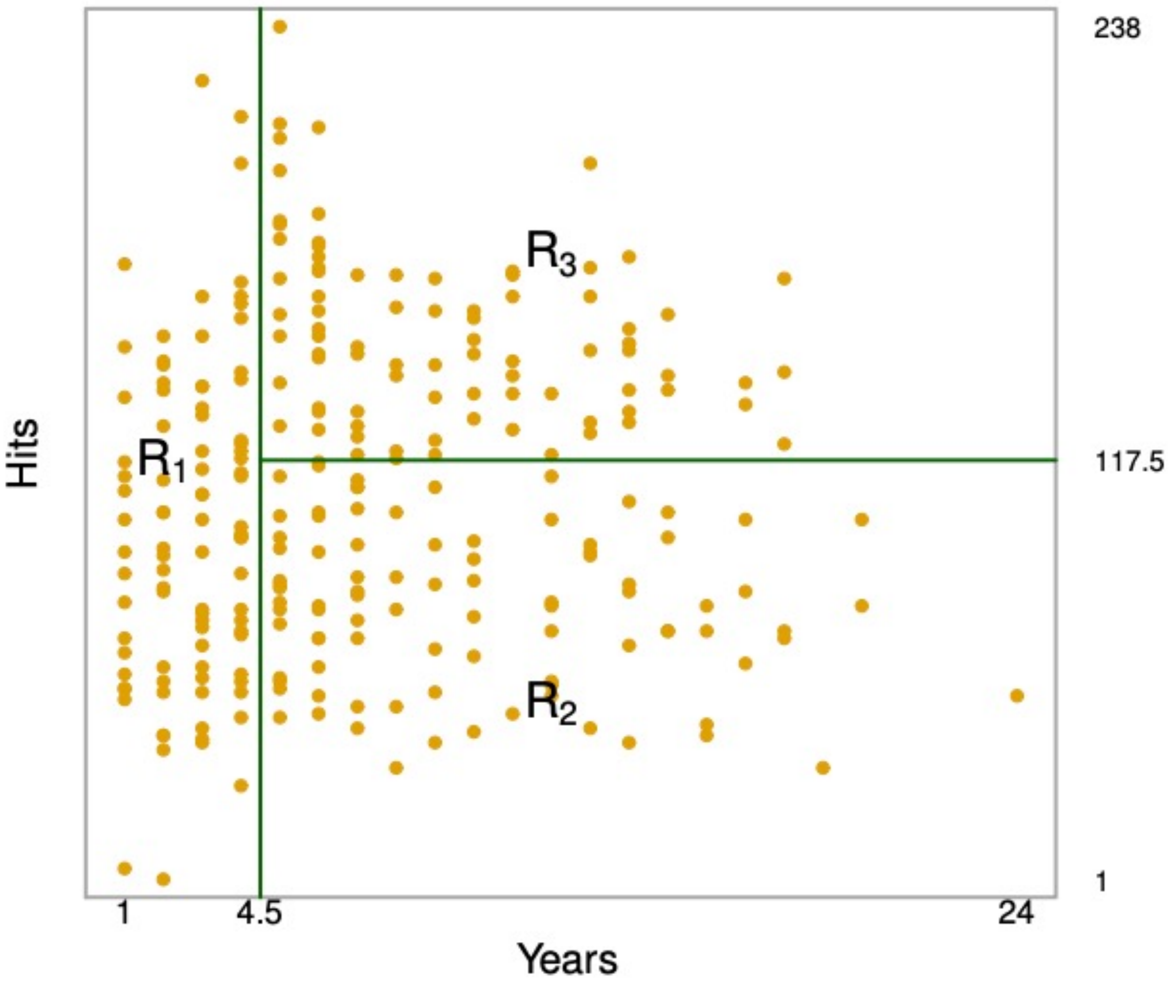




Decision Tree Example



Decision Tree Example



Q & A