Moving Beyond Linearity

AU STAT627

Emil Hvitfeldt

2021-06-14

Moving beyond Linearity

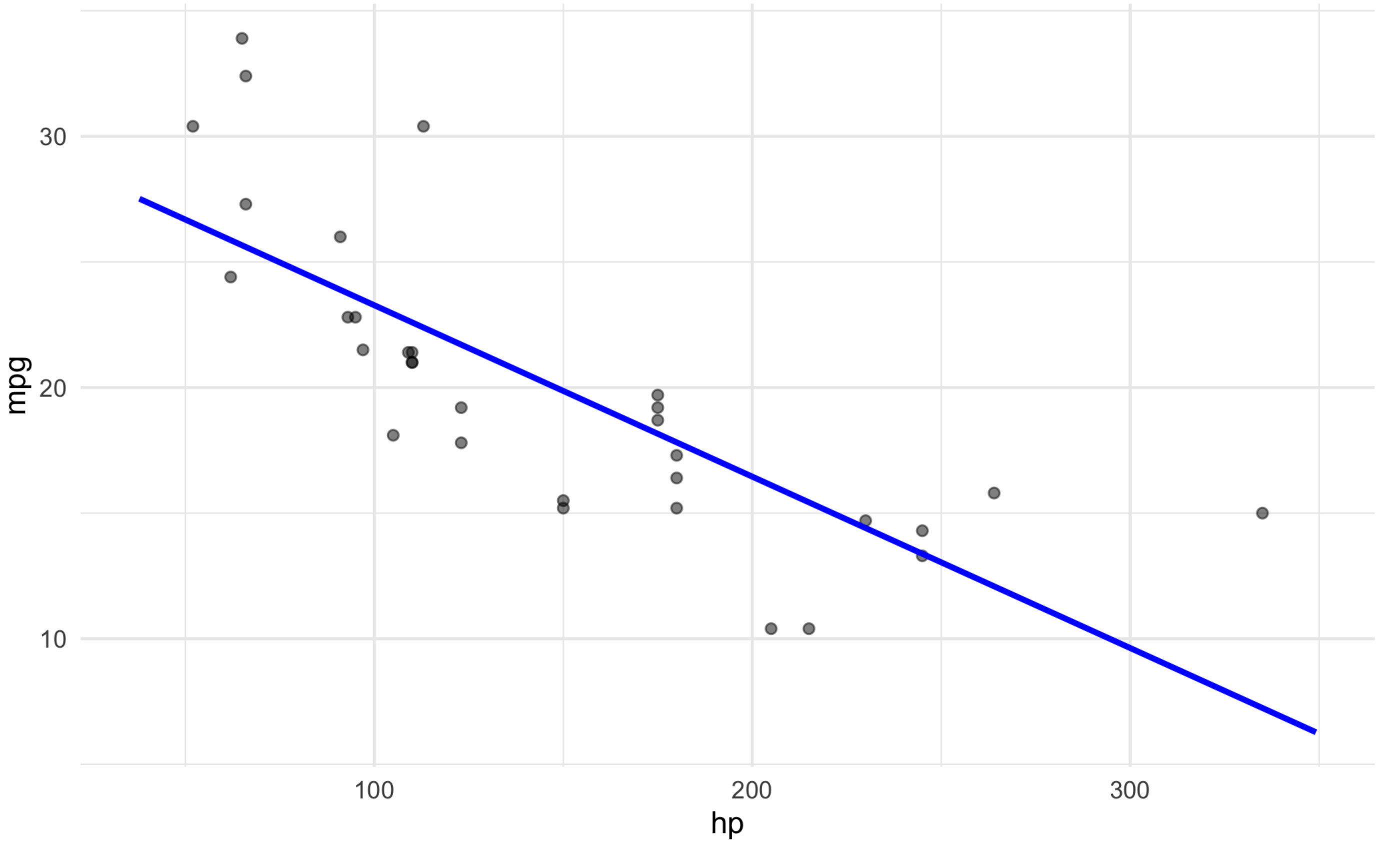
We have so far worked (mostly) with linear models

linear models are great because they are simple to describe, easy to work with in terms of interpretation and inference

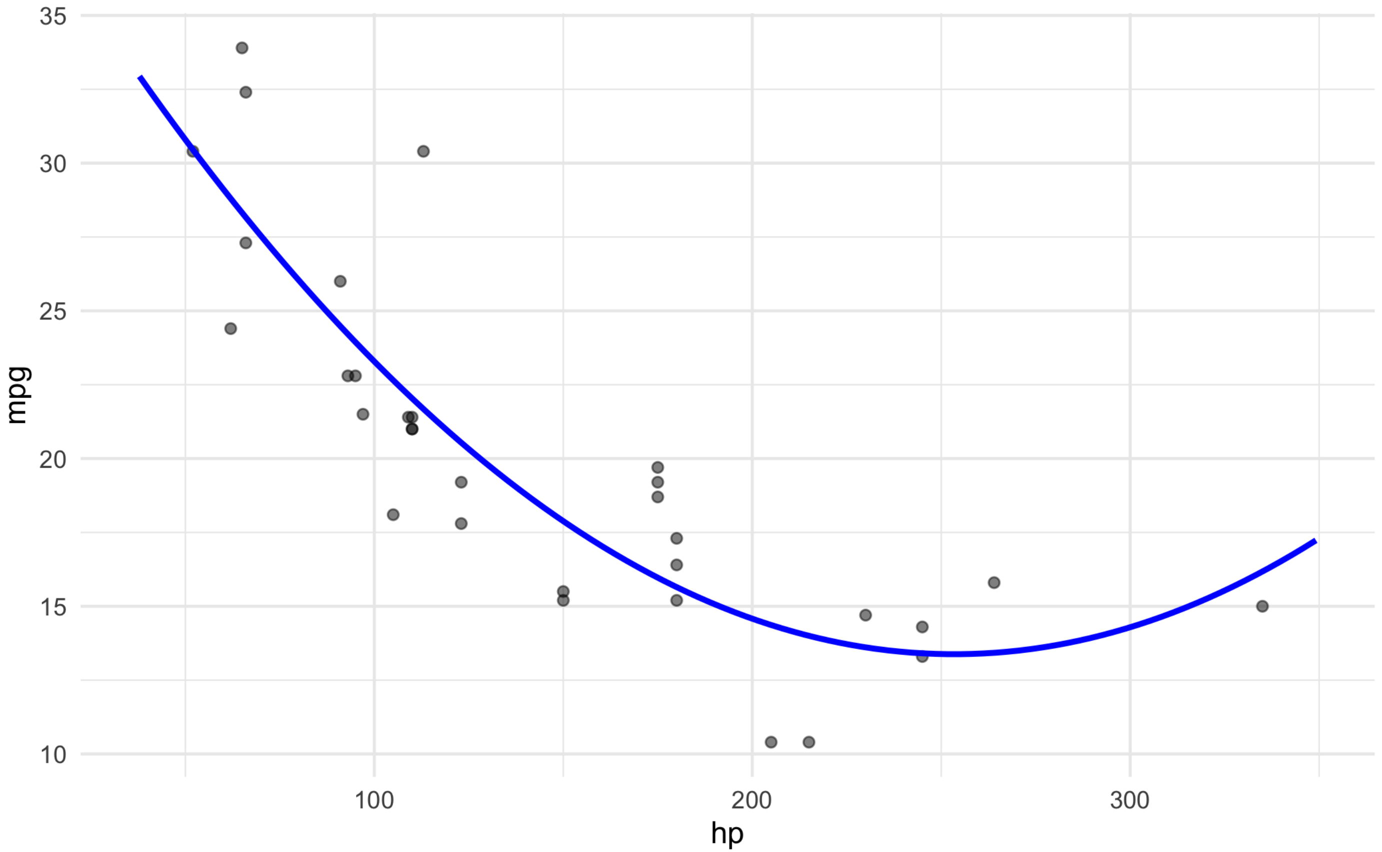
However, the linear assumption is often not satisfied

This week we will see what happens once we slowly relax the linearity assumption

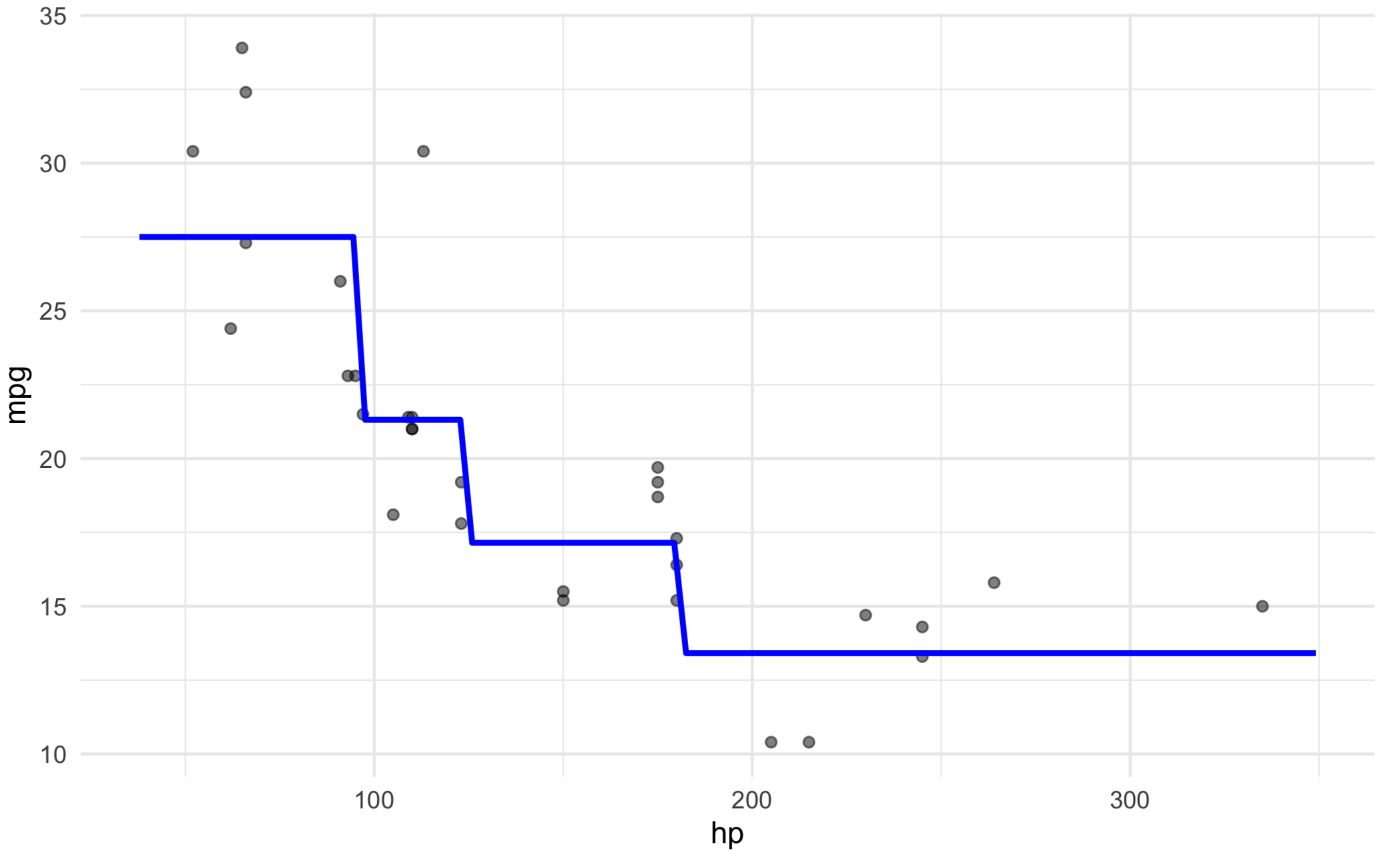
Moving beyond Linearity



Moving beyond Linearity



Moving beyond Linearity



Polynomial regression

Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

2nd degree polynomial regression

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

Polynomial regression

Polynomial regression function with d degrees

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \ldots + \beta_d x_i^d + \epsilon_i$$

Notice how we can treat the polynomial regression as

Polynomial regression

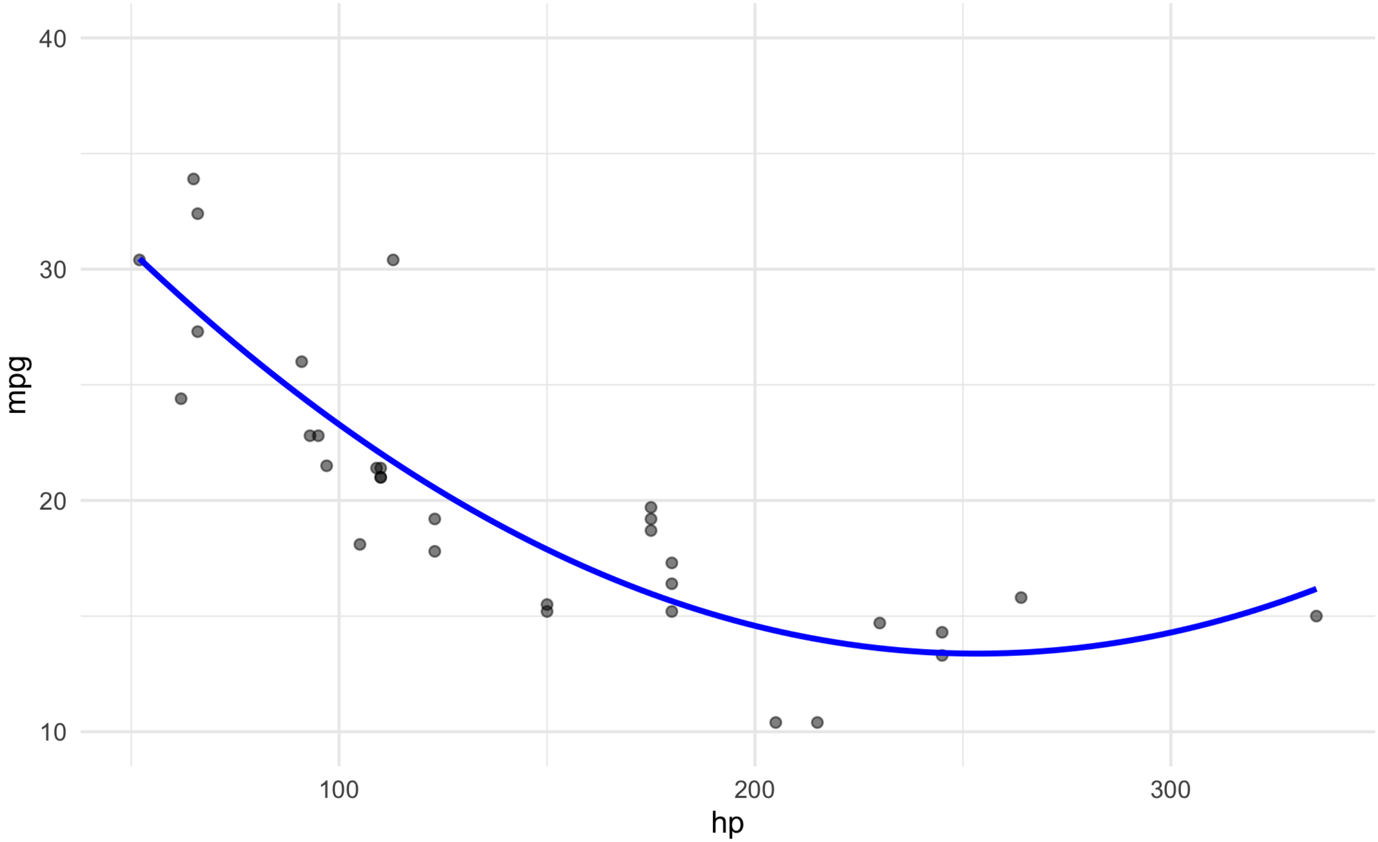
We are not limited to only use 1 variable when doing polynomial regression

Instead of thinking of it as fitting a "polynomial regression" model

Think of it as fitting a linear regression using polynomially expanded variables

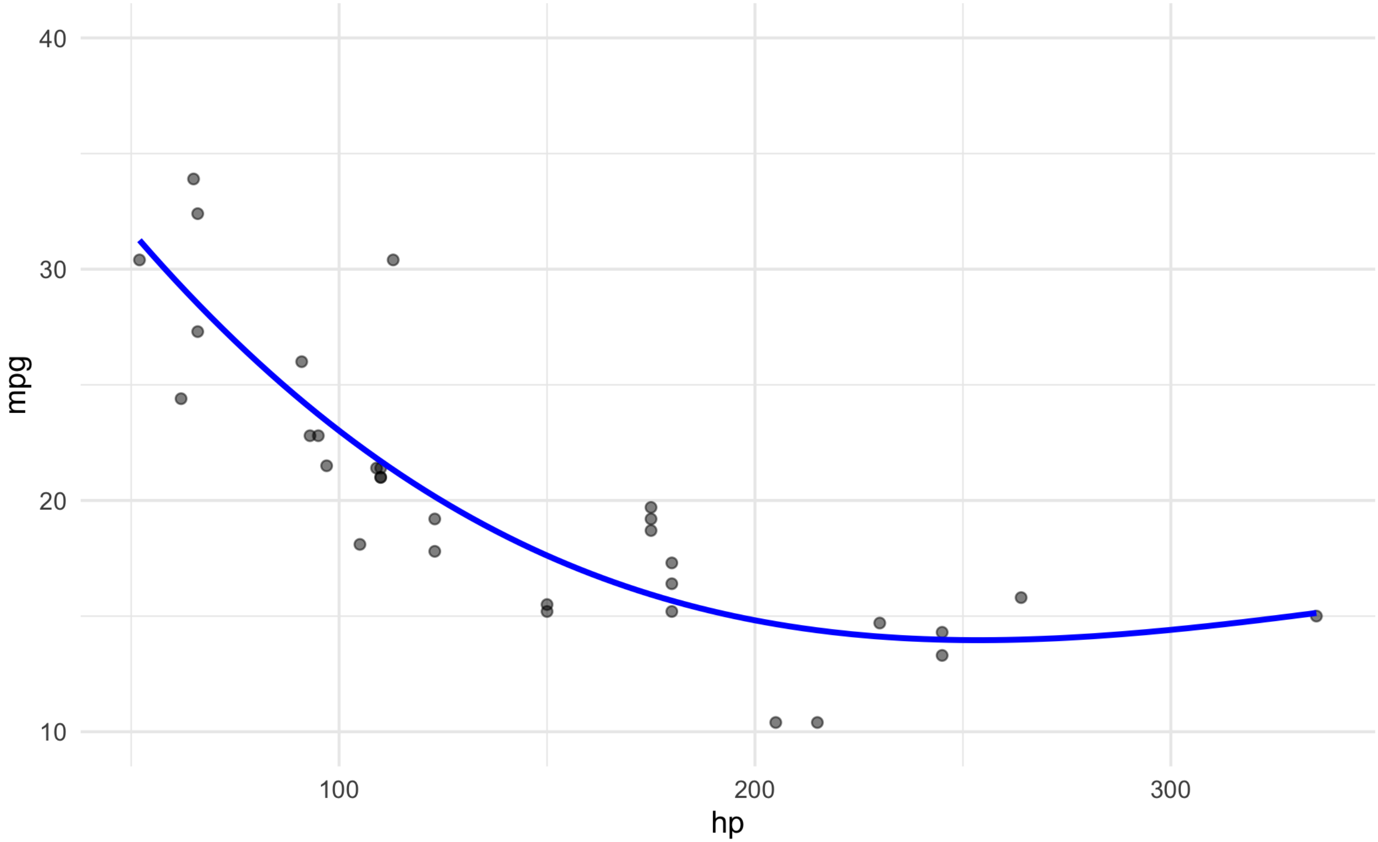
Polynomial regression

2 degrees



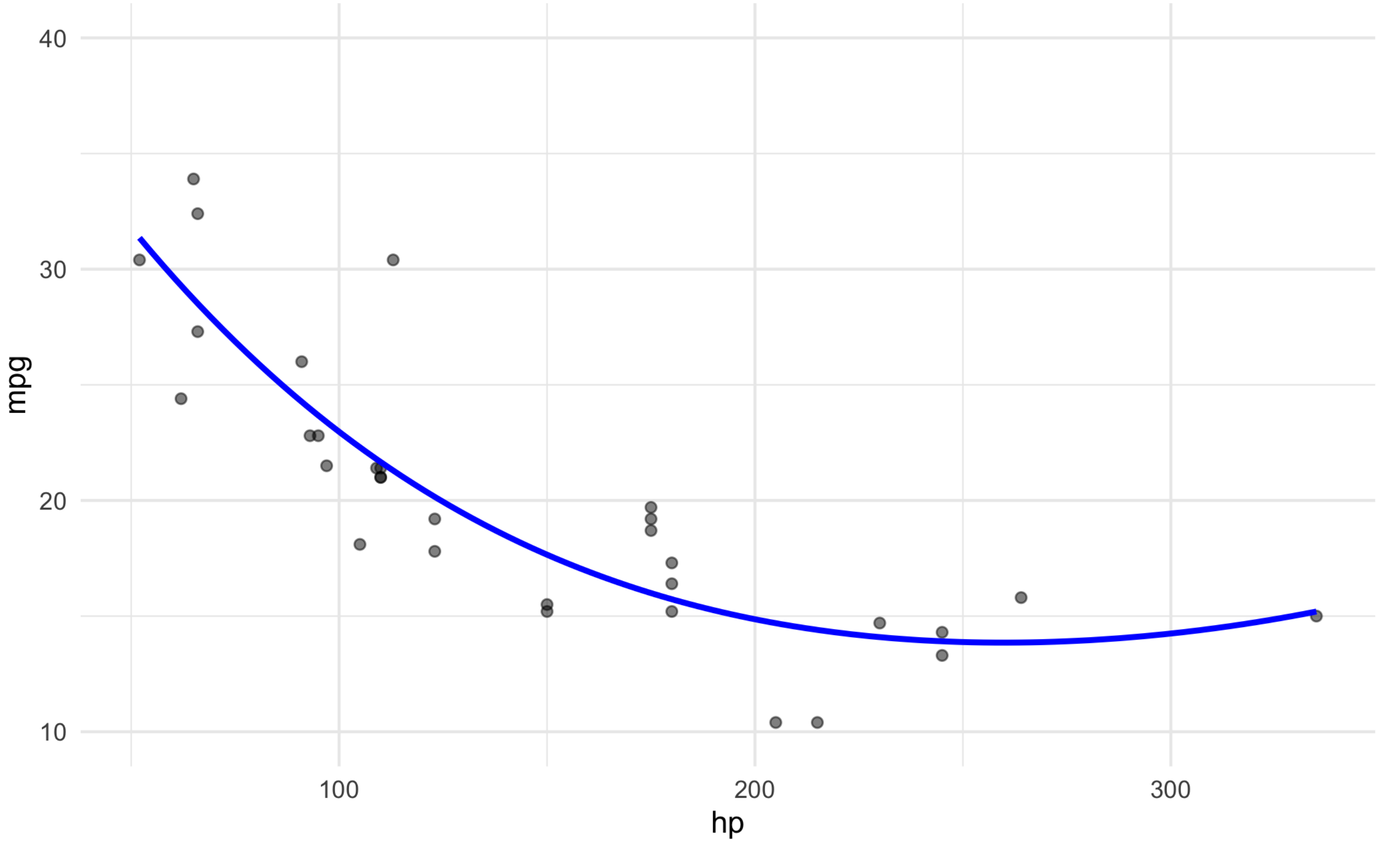
Polynomial regression

3 degrees



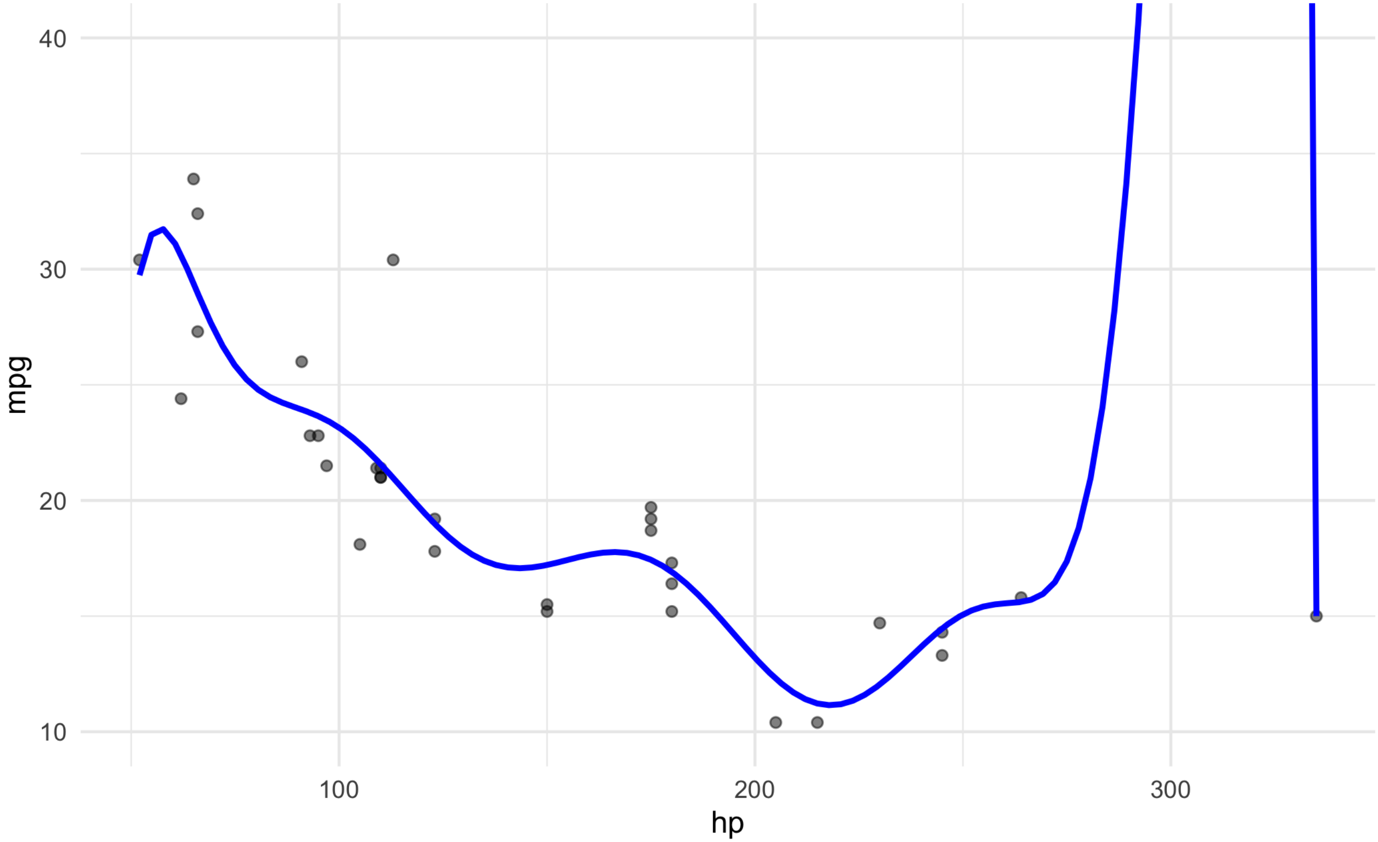
Polynomial regression

4 degrees



Polynomial regression

10 degrees



Step Functions

We can also try to turn continuous variables into categorical variables

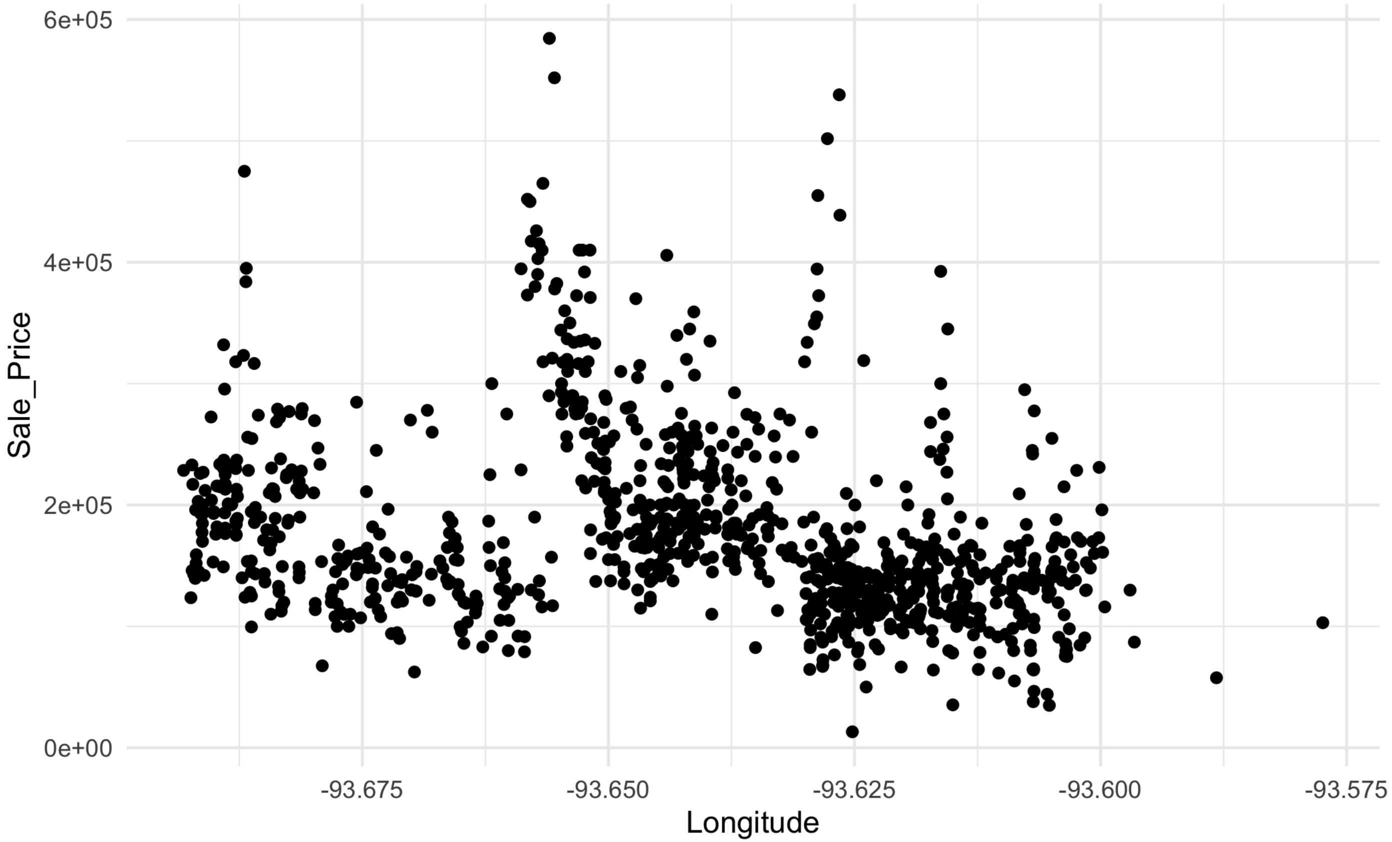
If we have data regarding the ages of people, then we can arrange the groups such as

- under 21
- 21-34
- 35-49
- 50-65
- over 65

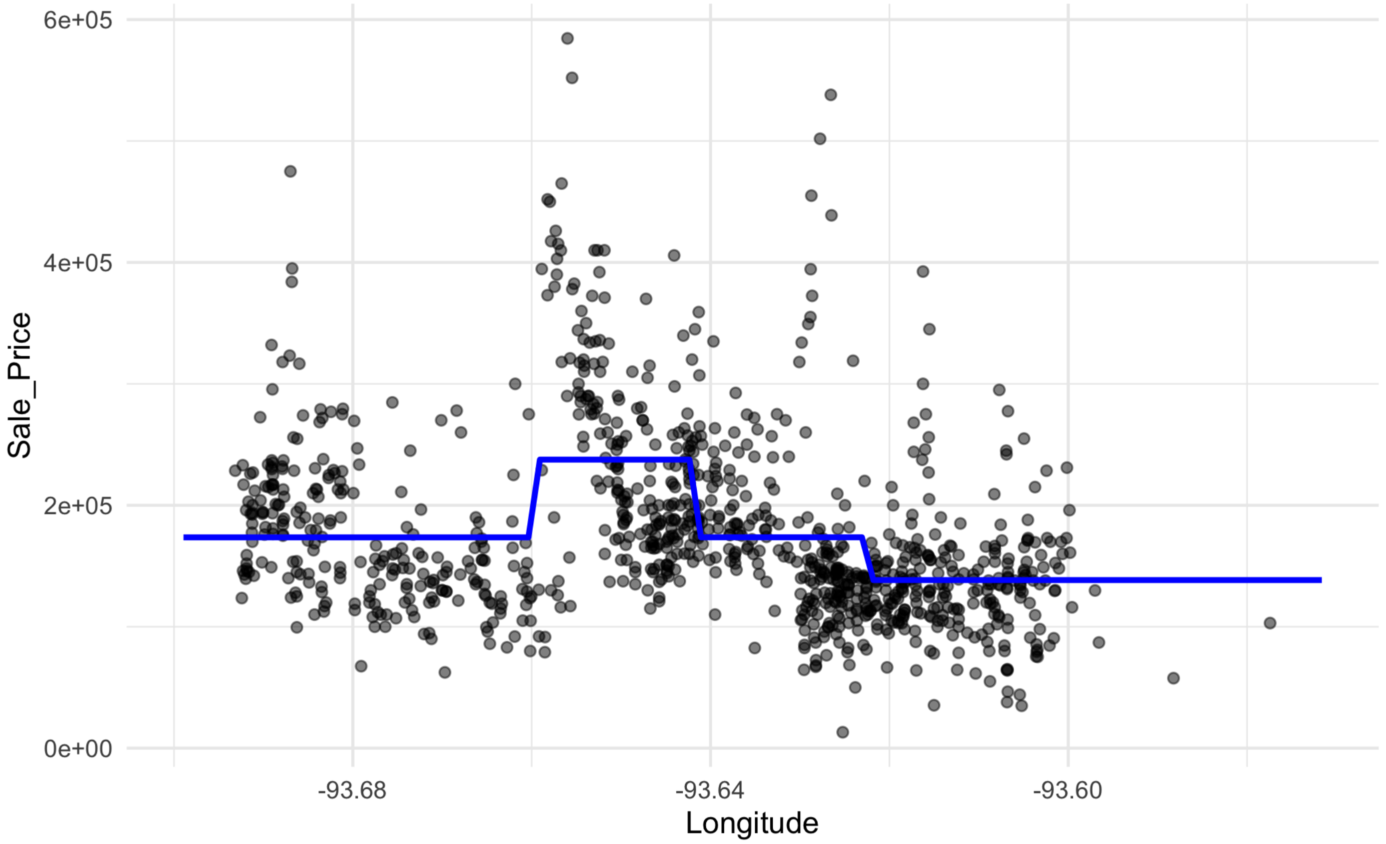
Step Functions

We divide a variable into multiple bins, constructing an ordered categorical variable

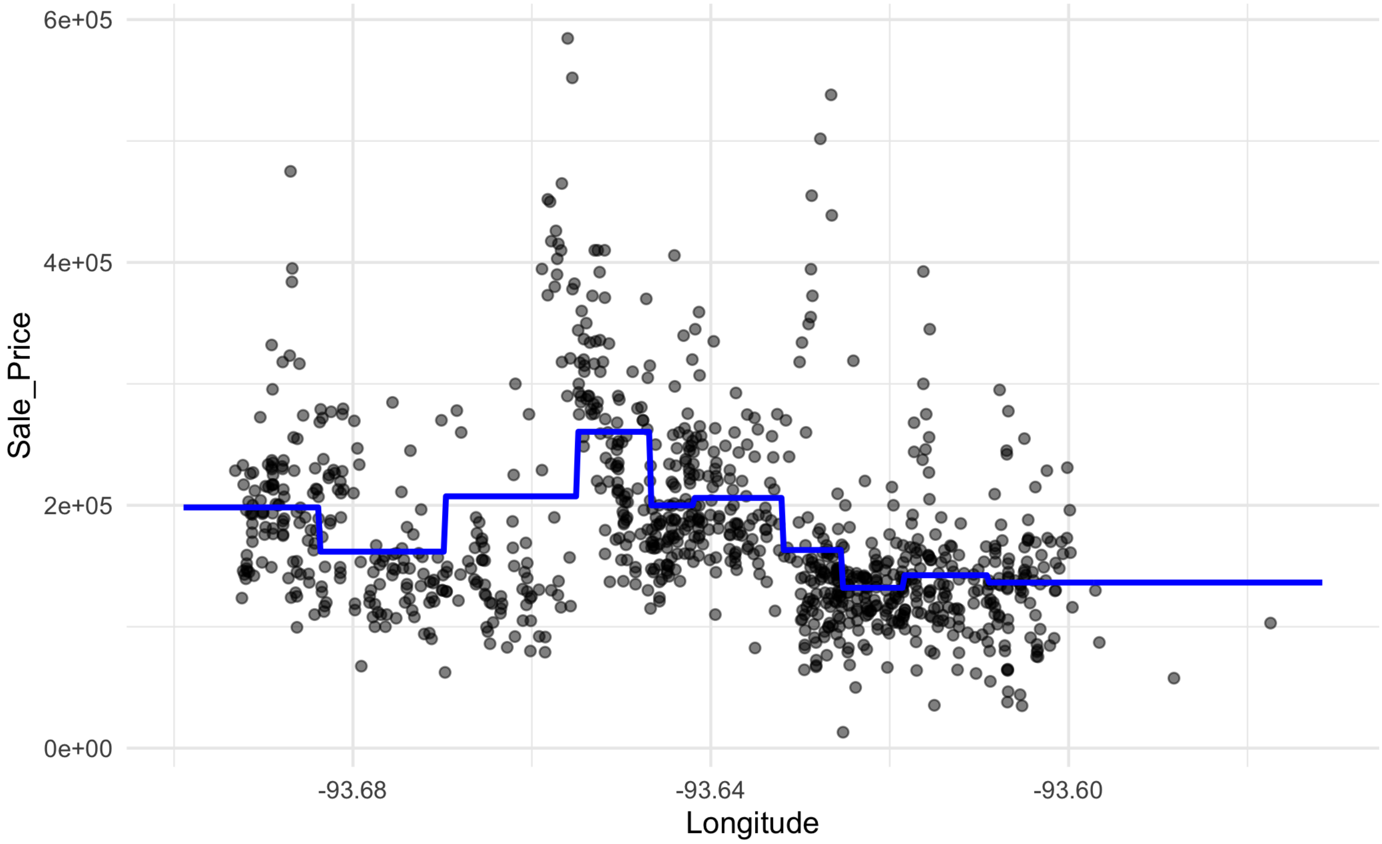
Step Functions



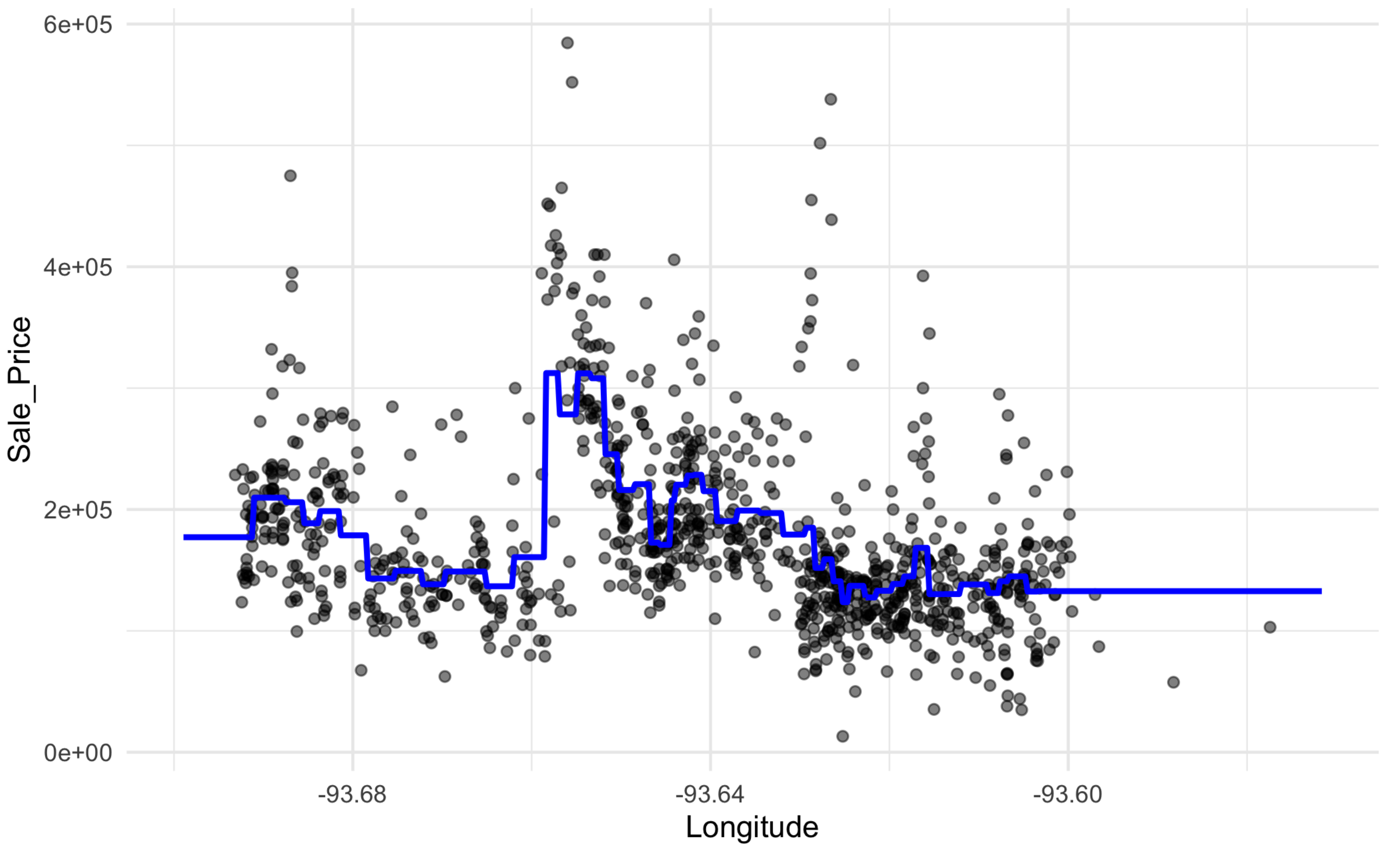
Step Functions



Step Functions



Step Functions



Step Functions

Depending on the number of cuts, you might miss the action of the variable in question

Be wary about using this method if you are going in blind, you end up creating a lot more columns of your data set and your flexibility increase drastically

Basis Functions

Both polynomial and piecewise-constant regression models are special cases of the **basis function** modeling approach

The idea is to have a selection of functions $b_1(X), b_2(X), \ldots, b_K(X)$ that we apply to our predictors

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \beta_3 b_3(x_i) + \ldots + \beta_K b_K(x_i) + \epsilon_i$$

Where $b_1(X), b_2(X), \ldots, b_K(X)$ are fixed and known

Basis Functions

The upside to this approach is that we can take advantage of the linear regression model for calculations along with all the inference tools and tests

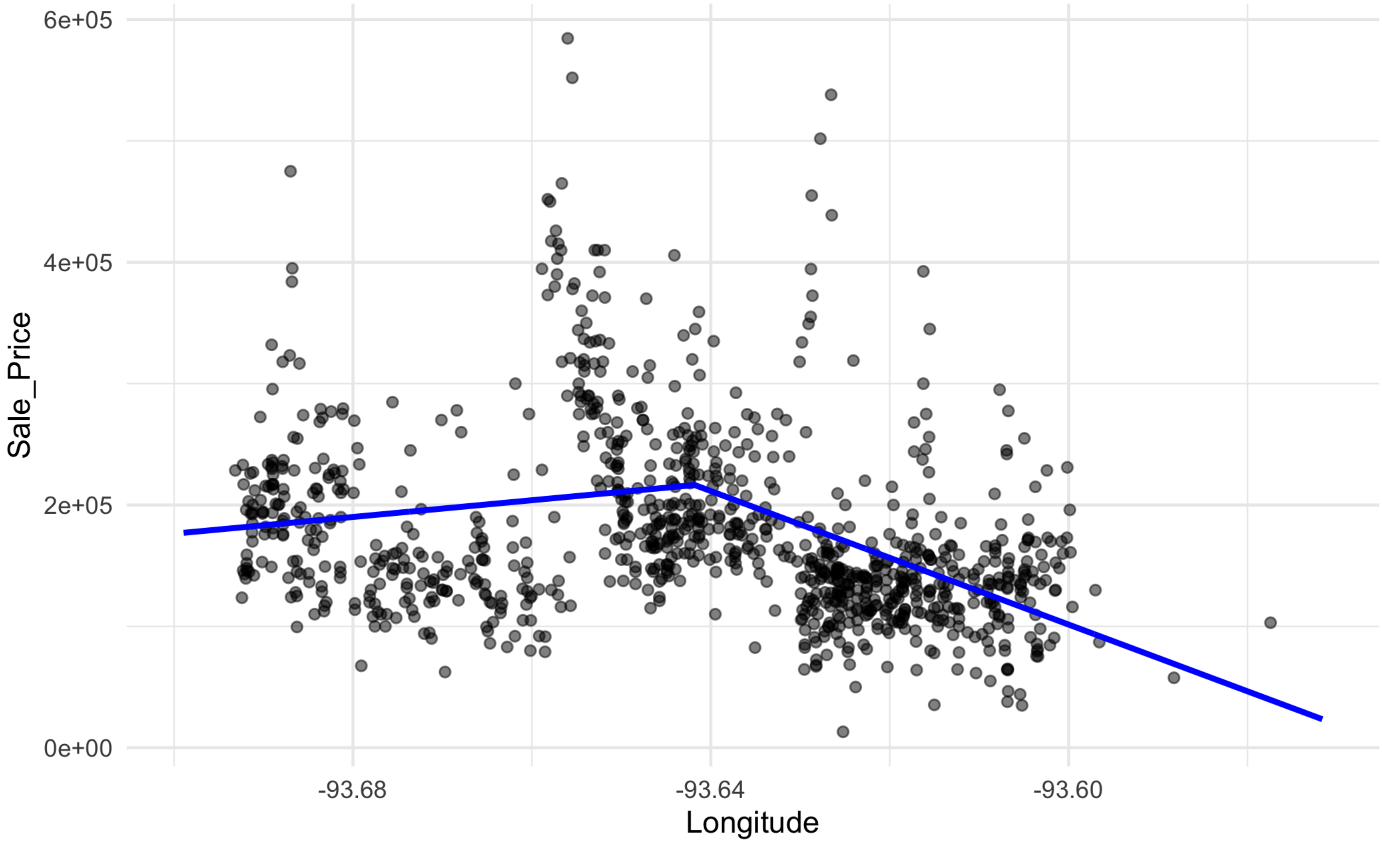
This does not mean that we are limited to using linear regression models when using basis functions

Regression Splines

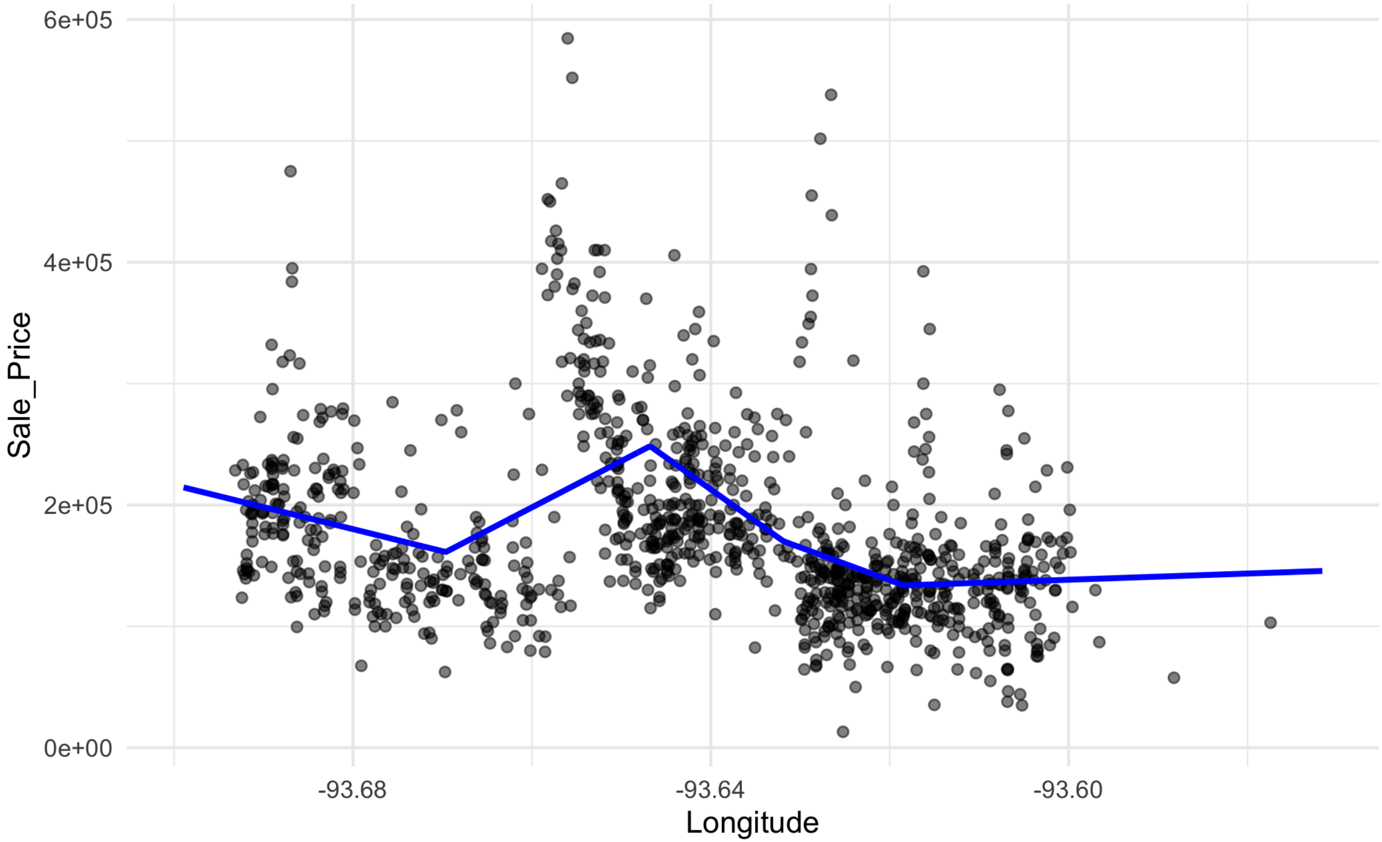
We can combine polynomial expansion and step functions to create **piecewise polynomials**

Instead of fitting 1 polynomial over the whole range of the data, we can fit multiple polynomials in a piecewise manner

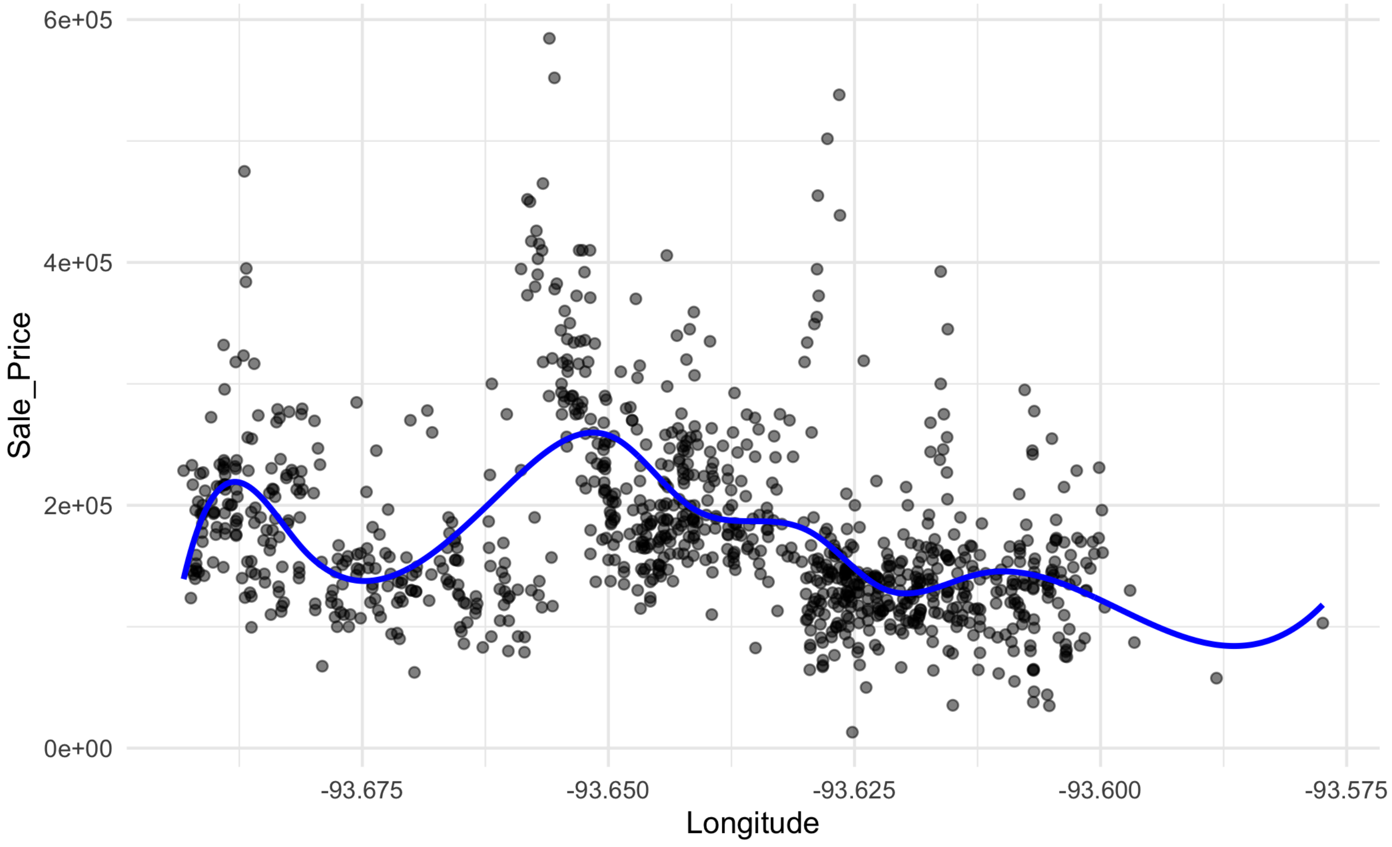
Regression Splines



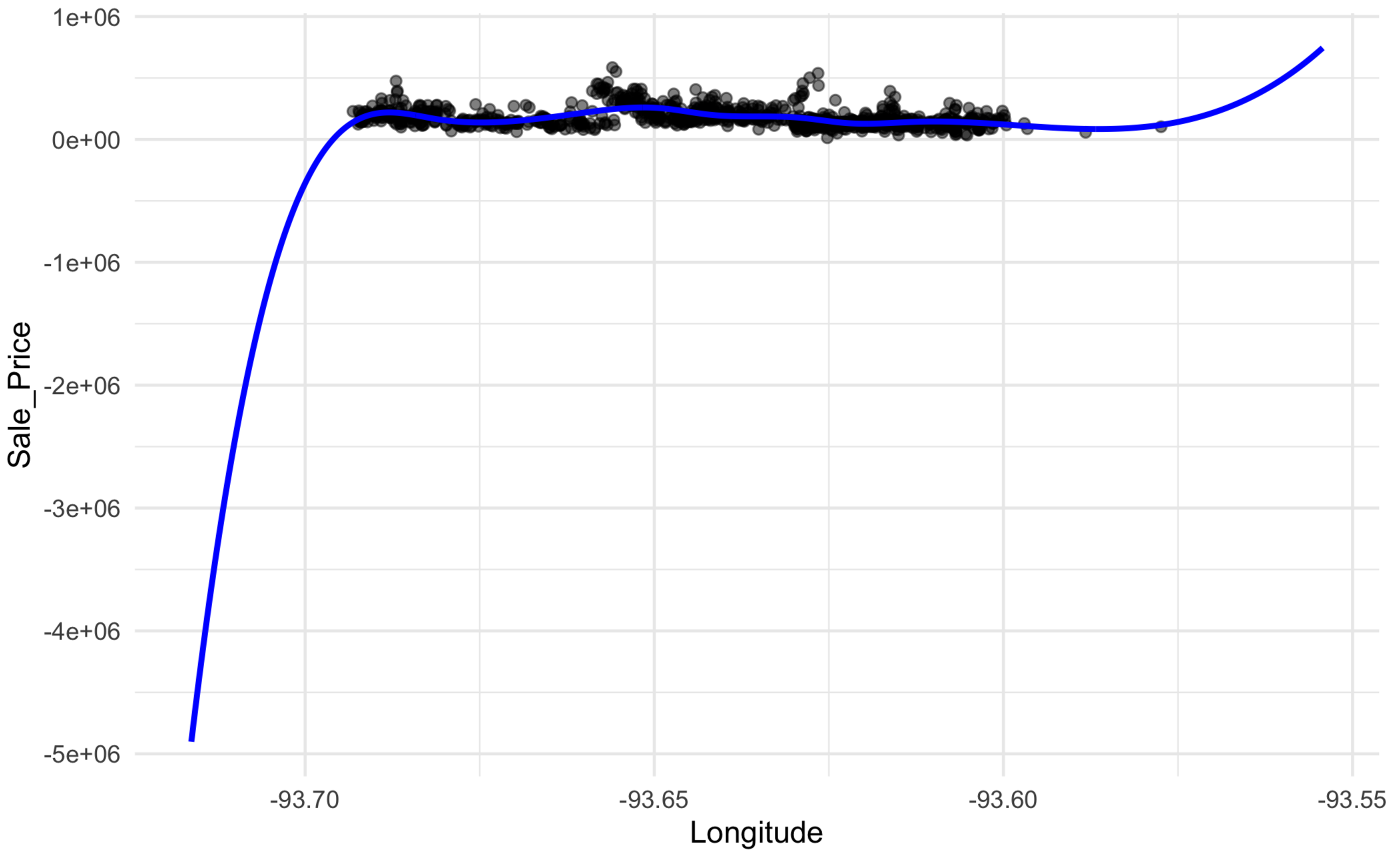
Regression Splines



Regression Splines



Regression Splines

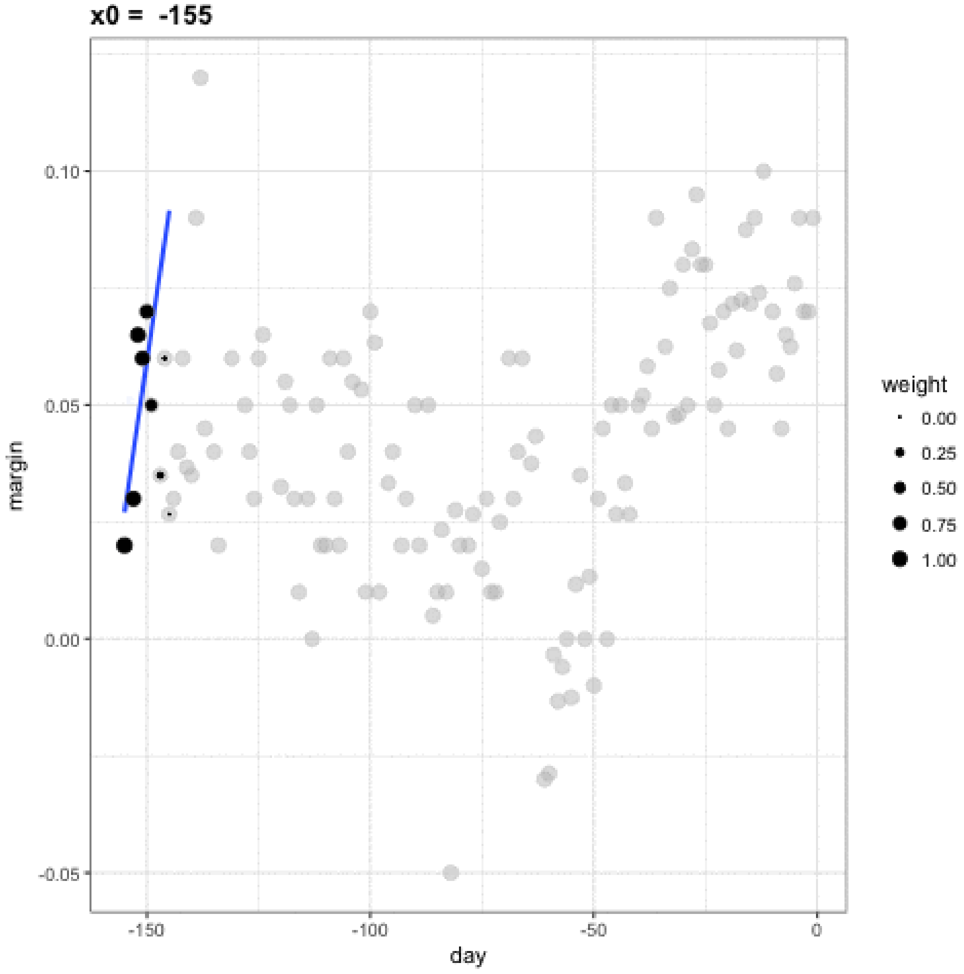


Local Regression

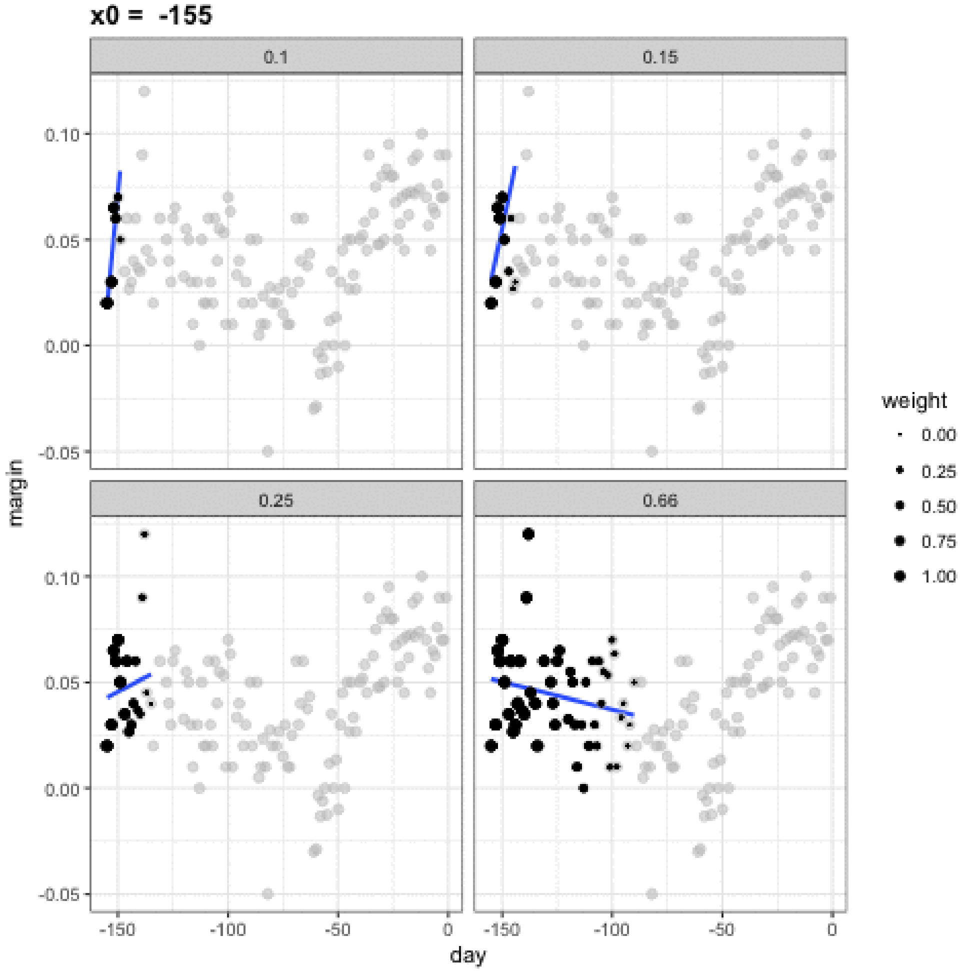
local regression is a method where the modeling is happening locally

namely, the fitted line only takes in information about nearby points

Local Regression



Local Regression



Generalized Additive Models

Generalized Additive Models provide a general framework to extend the linear regression model by allowing non-linear functions of each predictor while maintaining additivity

The standard multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$

is extended by replacing each linear component $\beta_j x_{ij}$ with a smooth linear function $f_j(x_{ij})$

Generalized Additive Models

Given us

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + f_3(x_{i3}) + \ldots + f_p(x_{ip}) + \epsilon_i$$

Since we are keeping the model additive we left with a more interpretable model since we are able to look at the effect of each of the predictors on the response by keeping the other predictors constant