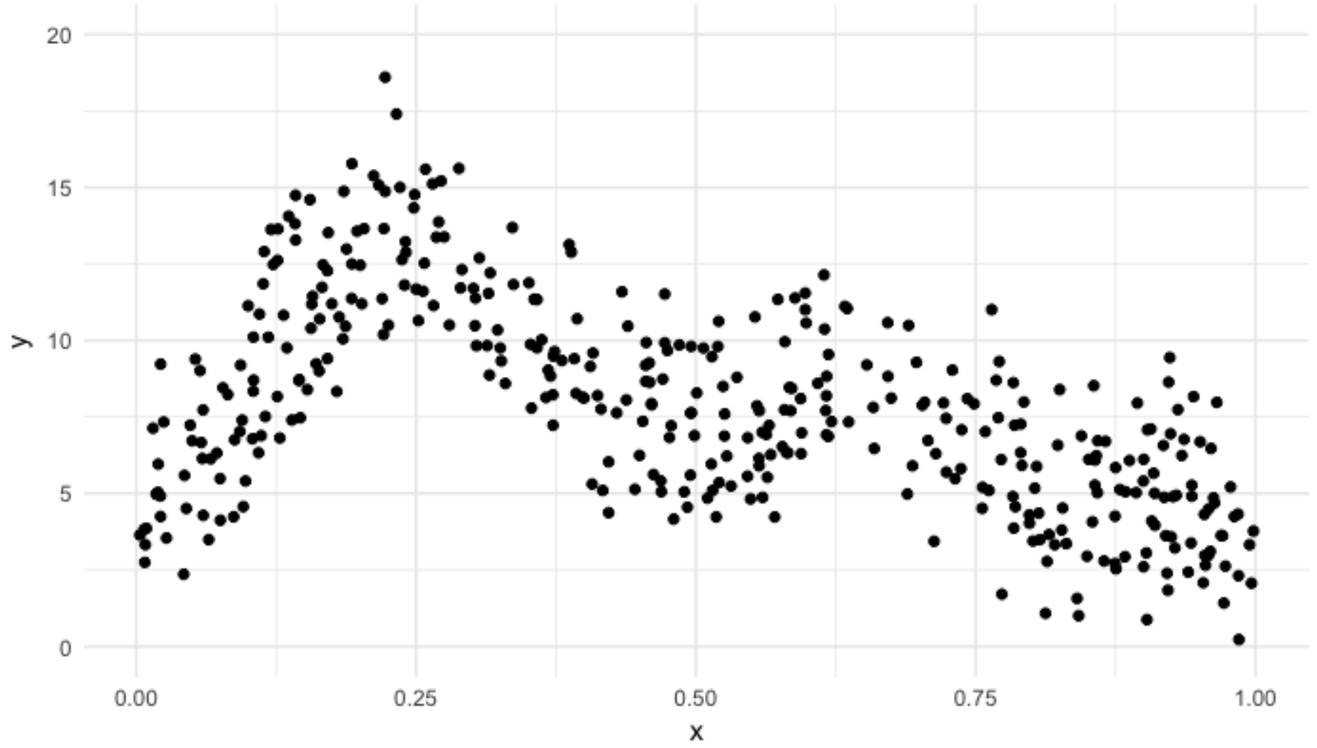
NONLINEAR MODELING IN R WITH GAMS

Introduction to Generalized Additive Models

Noam Ross
Senior Research Scientist, EcoHealth Alliance

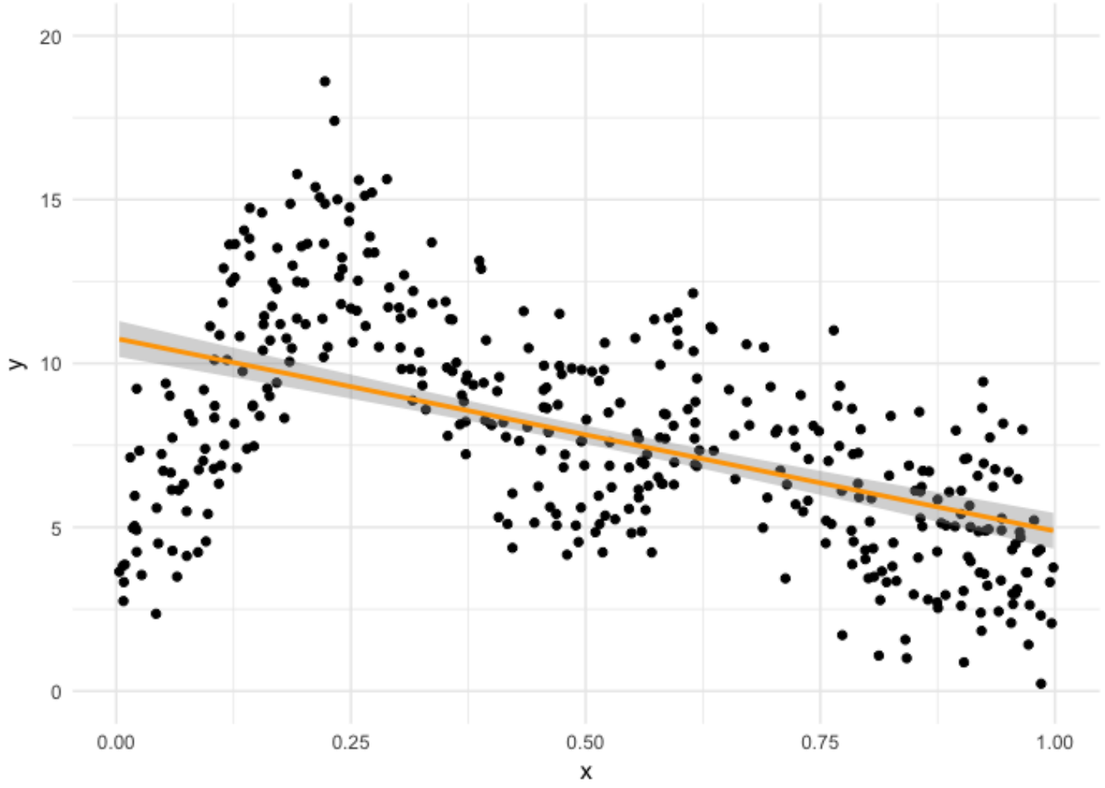
Trade-offs in Model Building

Non-linear Relationships



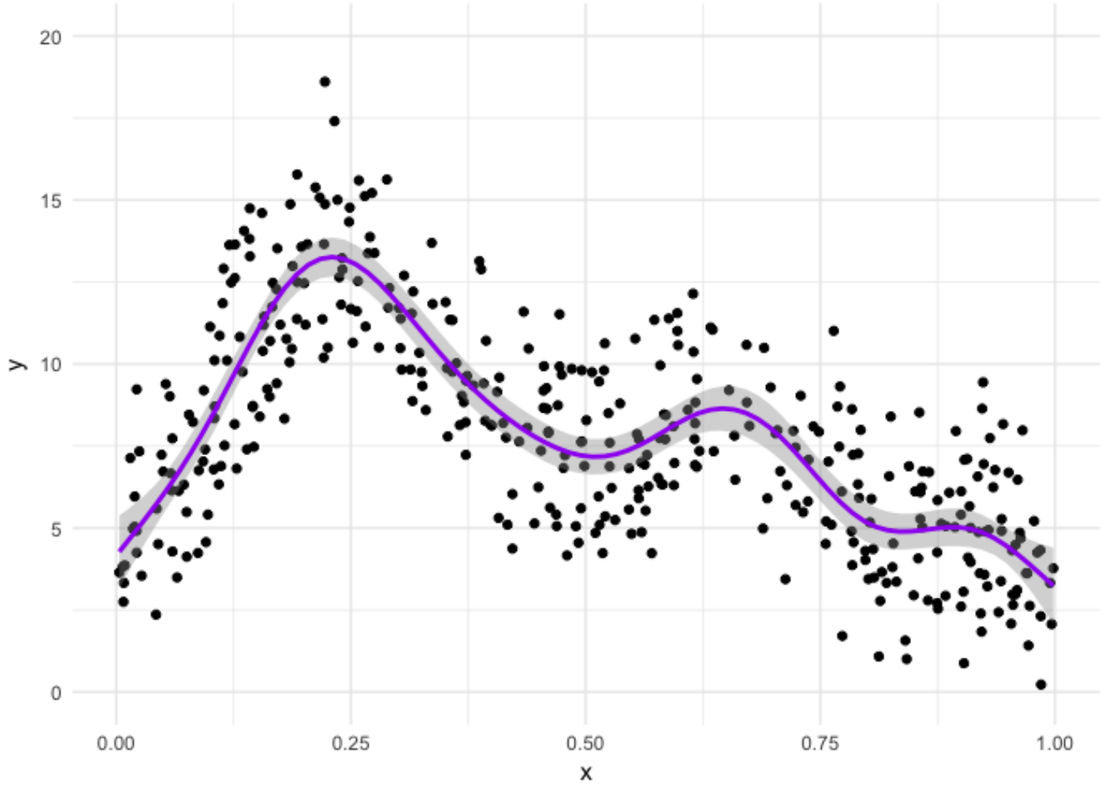
Nonlinear Relationships (2)

```
linear_mod <- lm(y ~ x, data = my_data)
```

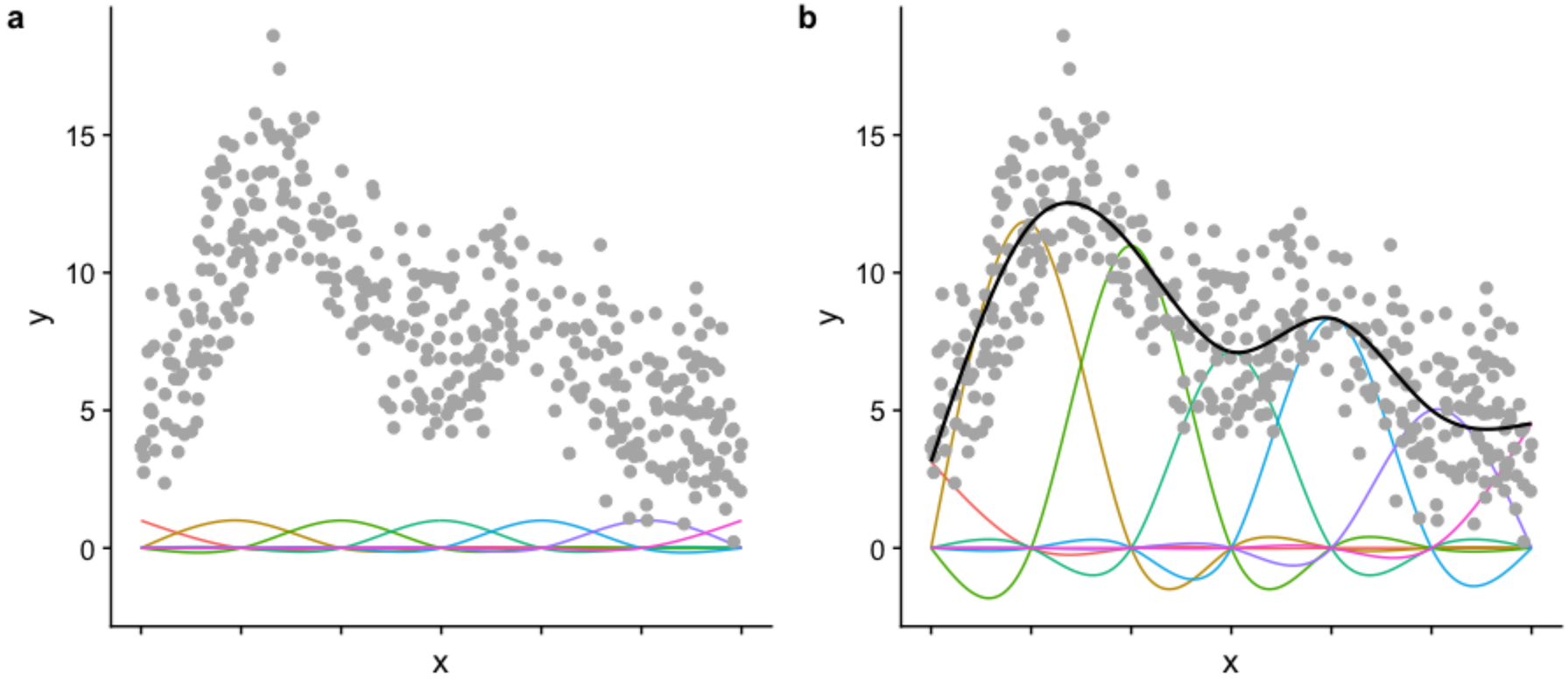


Nonlinear Relationships (3)

```
library(mgcv)
gam_mod <- gam(y ~ s(x), data = my_data)
```



Basis Functions



Basis Functions (2)

```
gam_mod <- gam(y ~ s(x), data = my_data)

coef(gam_mod)

(Intercept)      s(x2).1      s(x2).2
7.814448         5.272290     5.104941

s(x2).3          s(x2).4      s(x2).5
1.271135         1.720561    -1.180613

s(x2).6
-2.676133
```

NONLINEAR MODELING IN R WITH GAMS

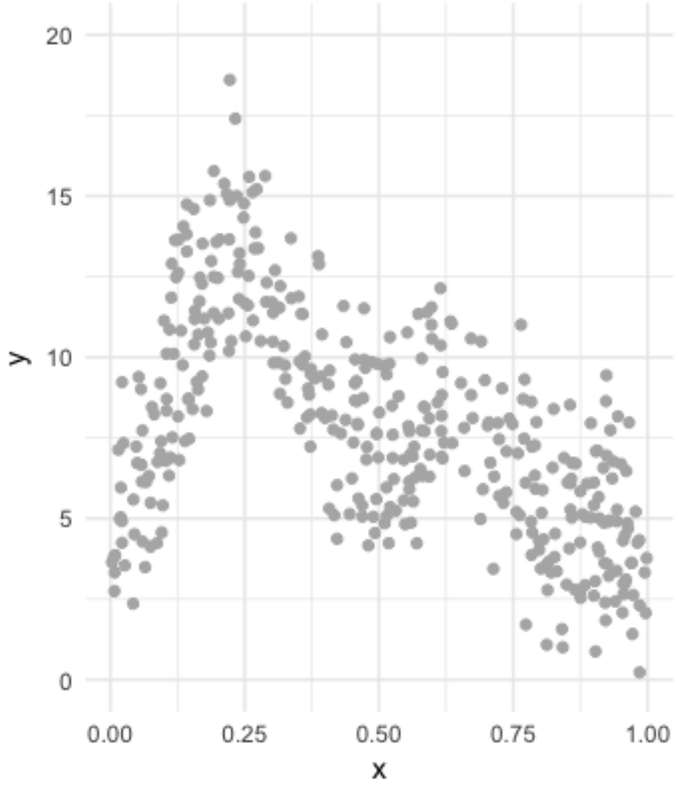
Let's practice!

NONLINEAR MODELING IN R WITH GAMS

Basis Functions and Smoothing

Noam Ross
Senior Research Scientist, EcoHealth Alliance

Getting the right fit

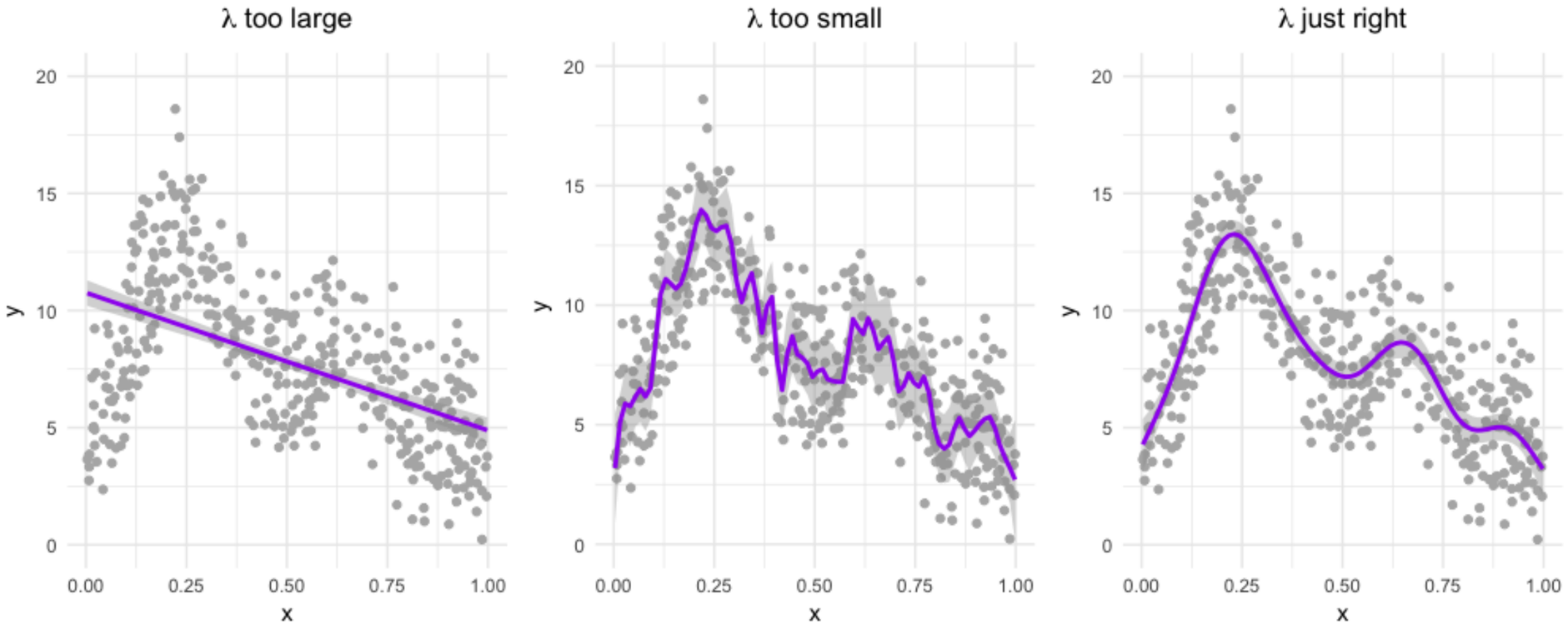


- Close to the data (avoiding under-fitting)
- Not fitting the noise (avoiding over-fitting)

Balancing Wiggliness

$$\text{Fit} = \text{Likelihood} - \lambda \times \text{Wiggliness}$$

Choosing the Right Smoothing Parameter



Smoothing Syntax

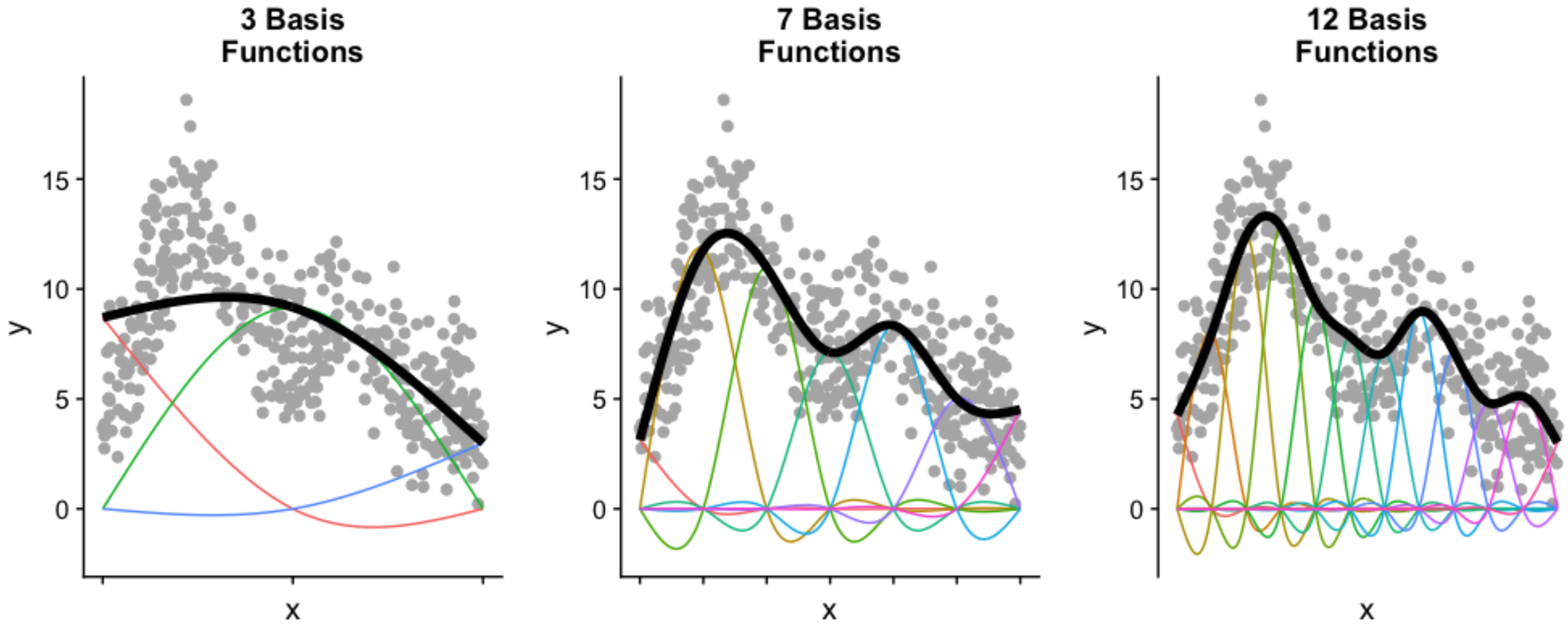
Setting a fixed smoothing parameter

```
gam(y ~ s(x), data = dat, sp = 0.1)
gam(y ~ s(x, sp = 0.1), data = dat)
```

Smoothing via restricted maximum likelihood

```
gam(y ~ s(x), data = dat, method = "REML")
```

Number of basis functions



Basis Function Syntax

Setting number of basis functions

```
gam(y ~ s(x, k = 3), data = dat, method = "REML")

gam(y ~ s(x, k = 10), data = dat, method = "REML")
```

Use the defaults

```
gam(y ~ s(x), data = dat, method = "REML")
```

NONLINEAR MODELING IN R WITH GAMS

Let's practice!

NONLINEAR MODELING IN R WITH GAMS

Multiple Regression with GAMs

Noam Ross
Senior Research Scientist, EcoHealth Alliance

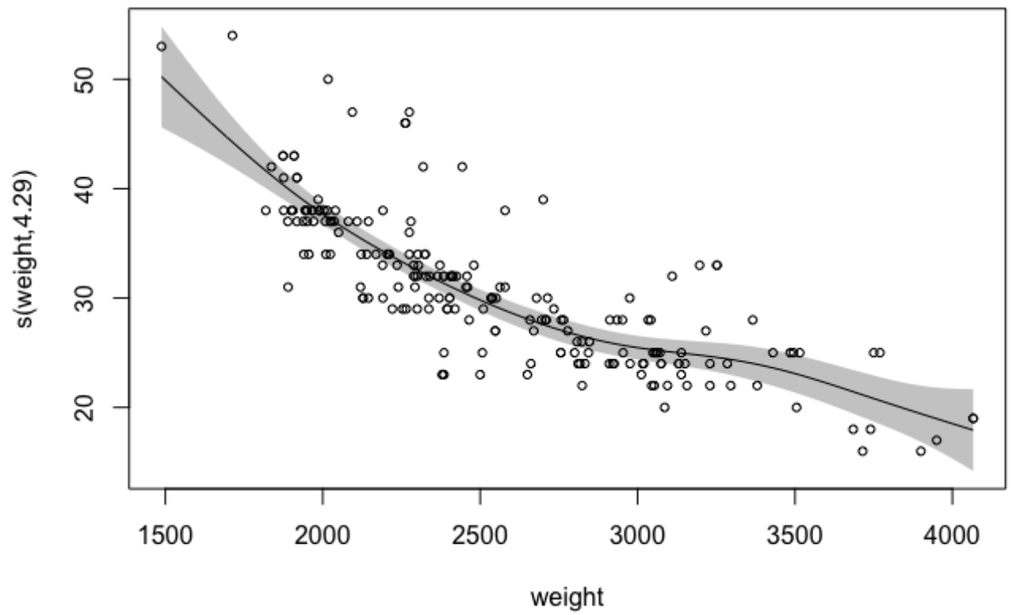
Our Working Dataset: mpg

```
mpg
```

```
    symbol loss           make   fuel aspir doors       style drive eng.loc    wb
1        3   NA   alfa-romero    gas   std   two convertible   rwd   front  88.6
2        3   NA   alfa-romero    gas   std   two convertible   rwd   front  88.6
3        1   NA   alfa-romero    gas   std   two   hatchback   rwd   front  94.5
4        2  164          audi    gas   std  four       sedan   fwd   front  99.8
5        2  164          audi    gas   std  four       sedan   4wd   front  99.4
6        2   NA          audi    gas   std   two       sedan   fwd   front  99.8
7        1  158          audi    gas   std  four       sedan   fwd   front 105.8
8        1   NA          audi    gas   std  four       wagon   fwd   front 105.8
9        1  158          audi    gas turbo  four       sedan   fwd   front 105.8
10       0   NA          audi    gas turbo   two   hatchback   4wd   front  99.5
...
```

Multiple Smooths (1)

```
model <- gam(hw.mpg ~ s(weight), data = mpg,
             method = "REML")
```



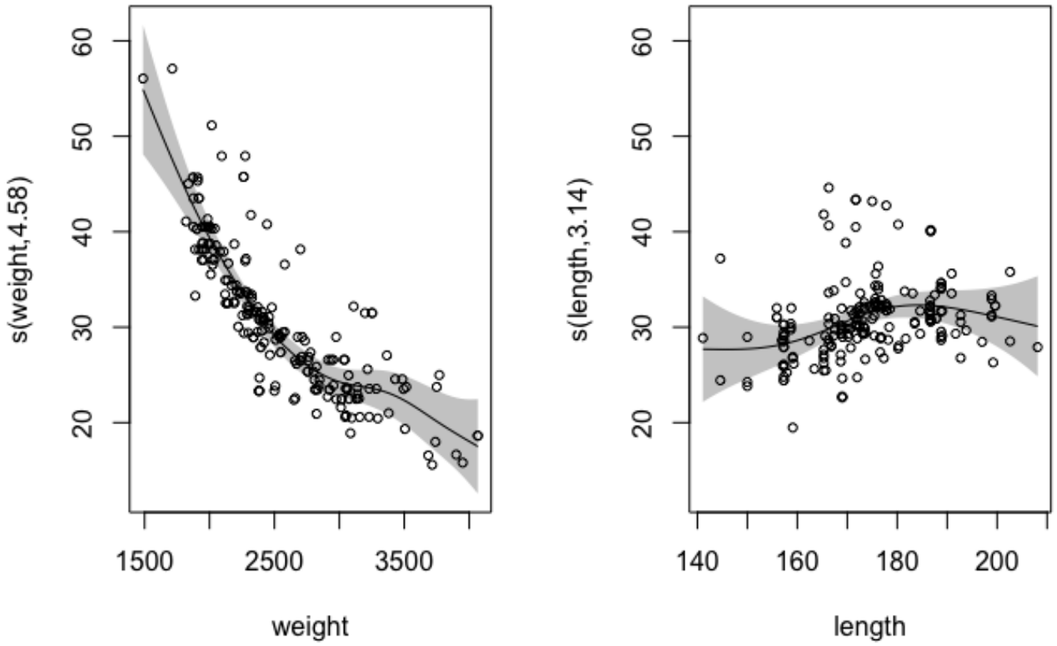
Multiple Smooths (2)

```
model <- gam(hw.mpg ~ s(weight), data = mpg,
             method = "REML")
```

```
model2 <- gam(hw.mpg ~ s(weight) + s(length), data = mpg,
              method = "REML")
```

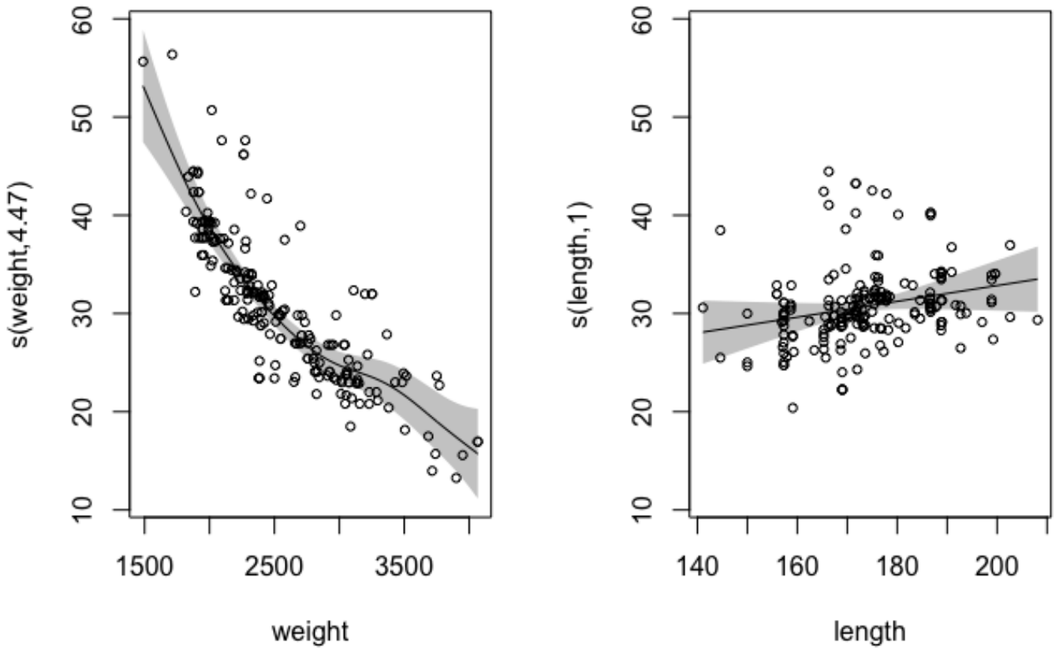
Multiple Smooths (3)

```
model2 <- gam(hw.mpg ~ s(weight) + s(length), data = mpg,
              method = "REML")
```



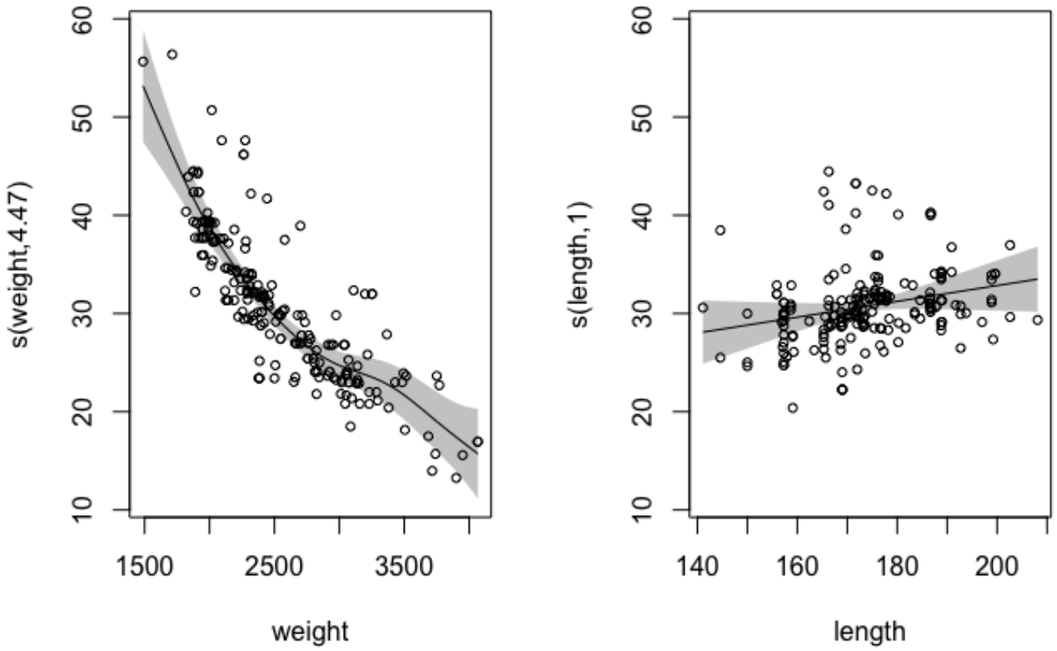
Linear terms

```
model2 <- gam(hw.mpg ~ s(weight) + length, data = mpg,
              method = "REML")
```



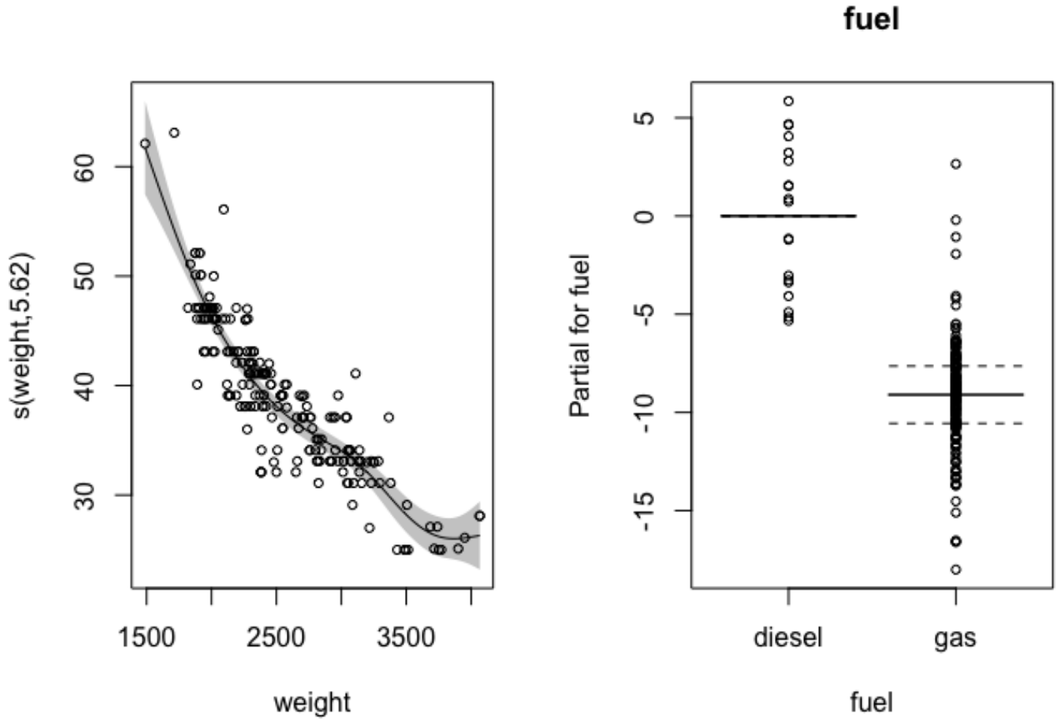
Linear Terms (2)

```
model2b <- gam(hw.mpg ~ s(weight) + s(length, sp = 1000), data = mpg,
               method = "REML")
```



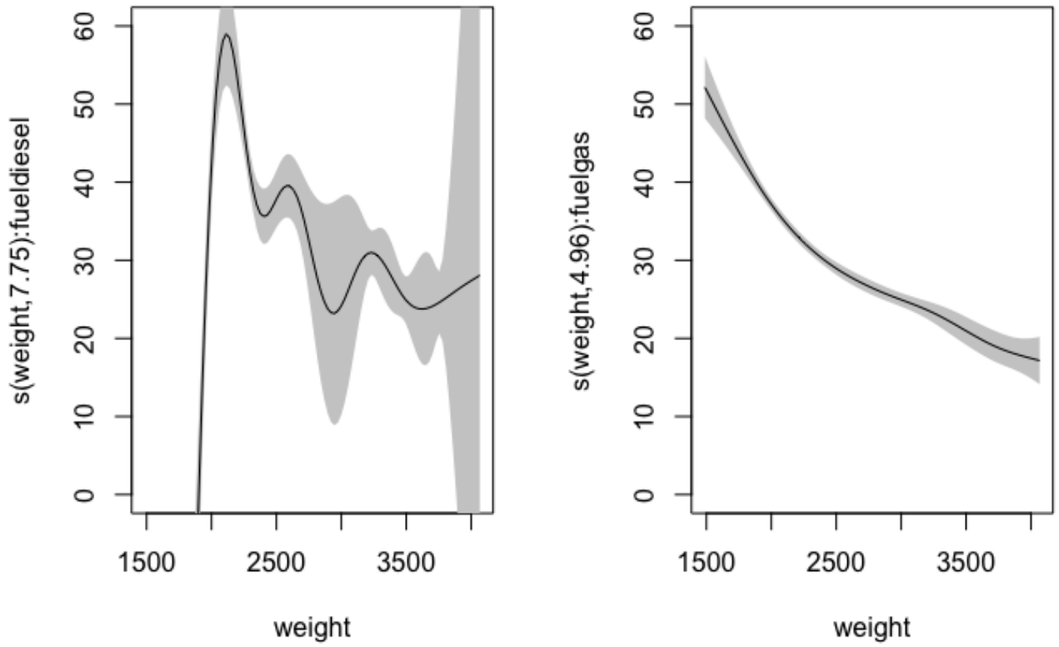
Categorical Terms (1)

```
model3 <- gam(hw.mpg ~ s(weight) + fuel, data = mpg,
              method = "REML")
```



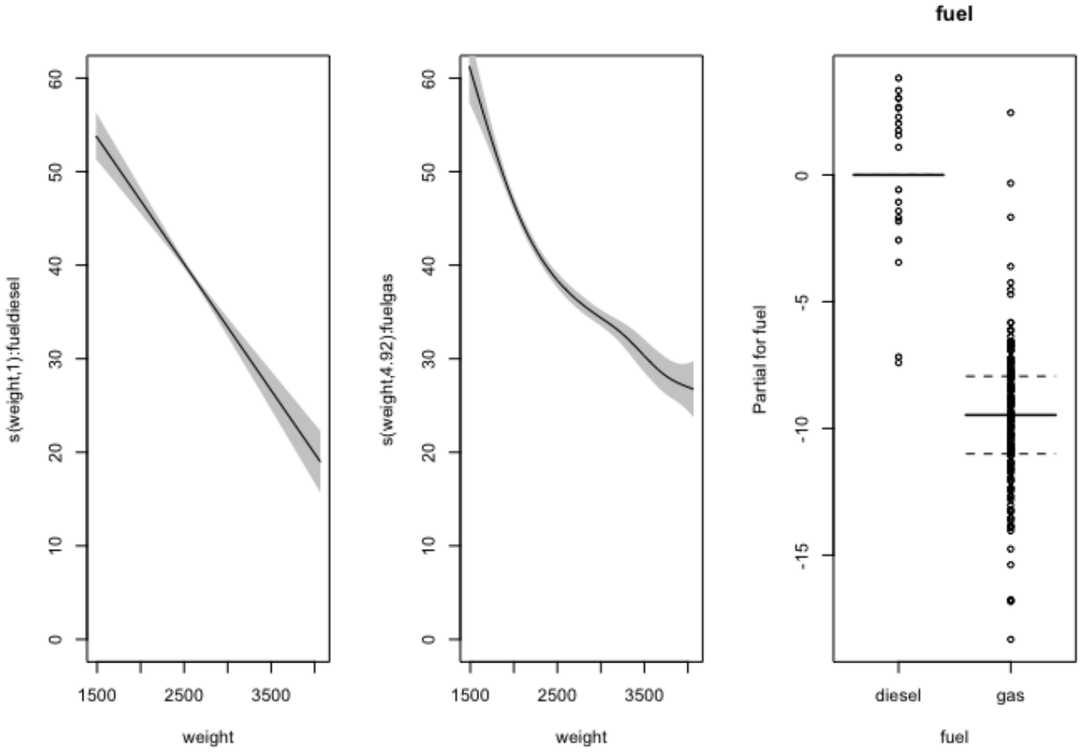
Categorical Terms (2)

```
model4 <- gam(hw.mpg ~ s(weight, by = fuel), data = mpg,
              method = "REML")
```



Categorical Terms (3)

```
model4b <- gam(hw.mpg ~ s(weight, by = fuel) + fuel, data = mpg,
               method = "REML")
```



NONLINEAR MODELING IN R WITH GAMS

Let's practice!