Introduction to Generalized Additive Models

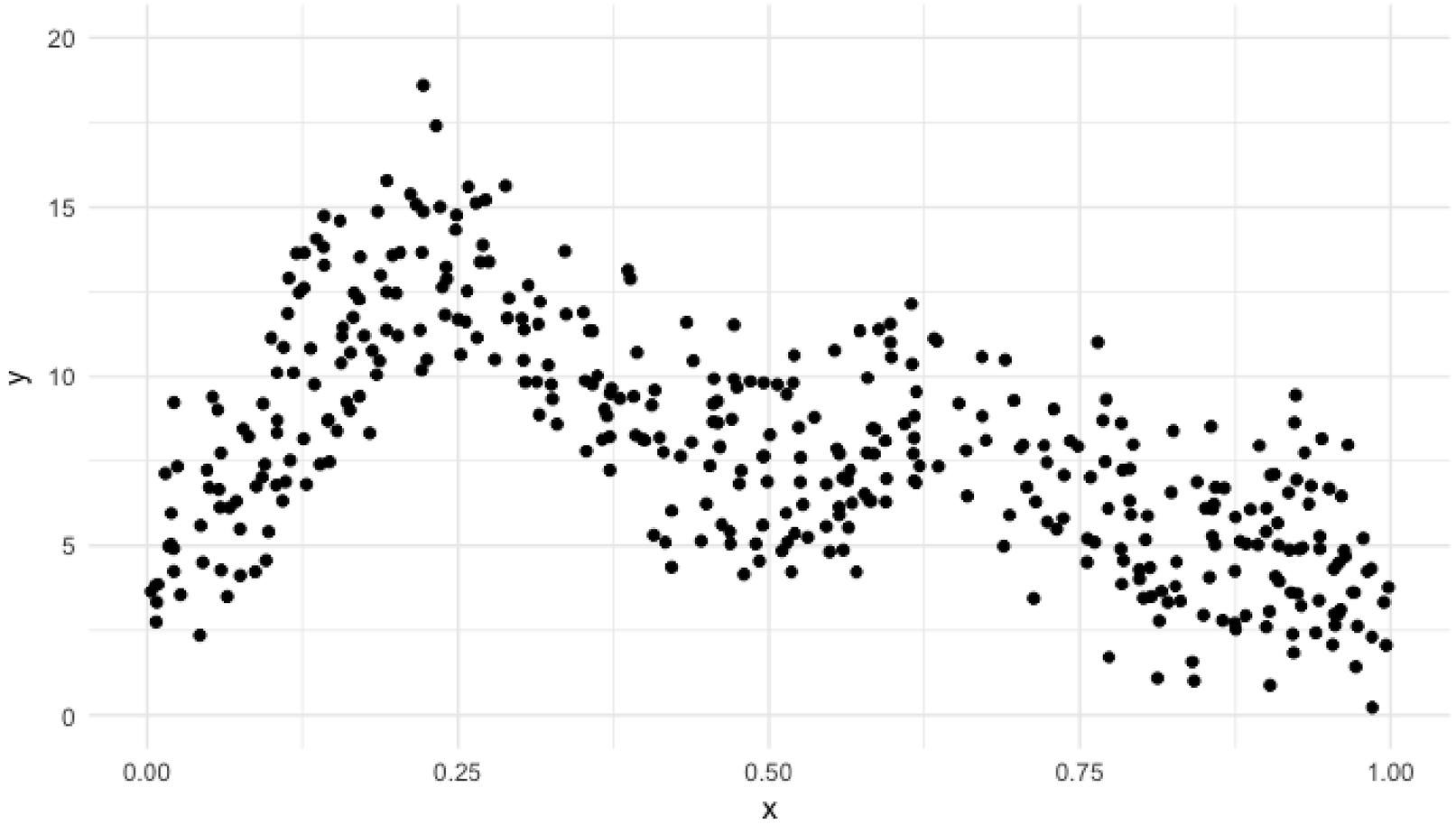
NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Noam Ross
Senior Research Scientist, EcoHealth Alliance

datacamp

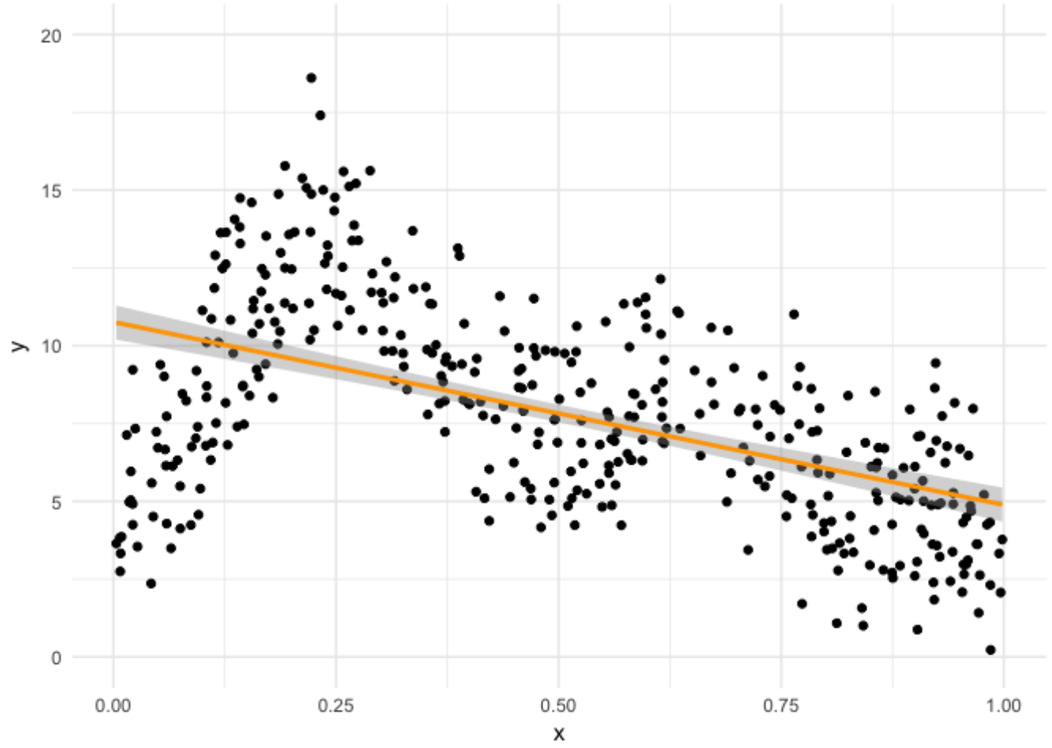
Trade-offs in model building

Non-linear relationships



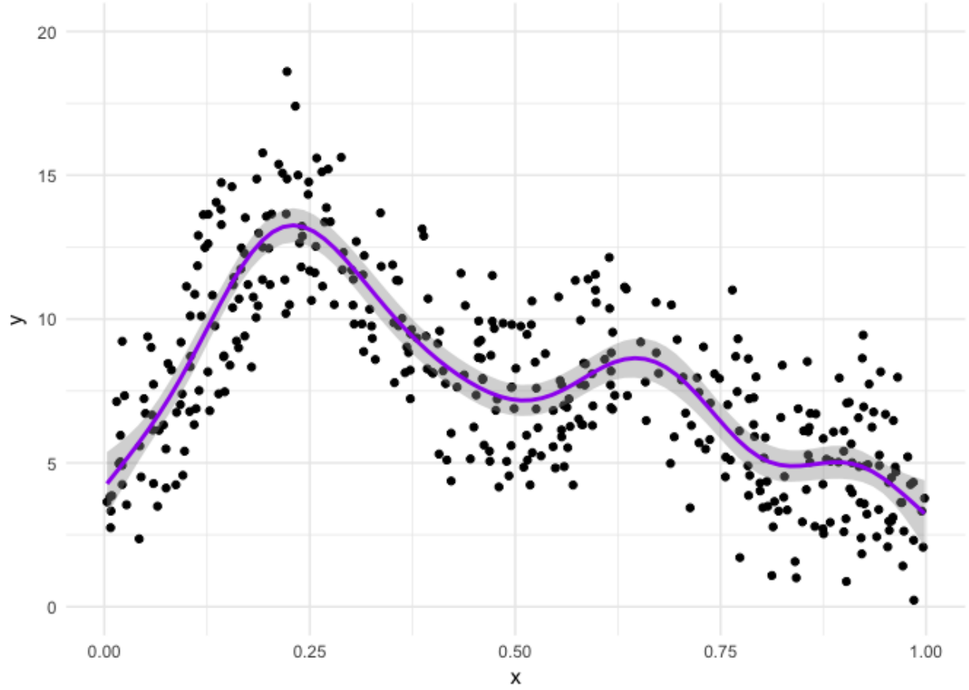
Nonlinear relationships

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

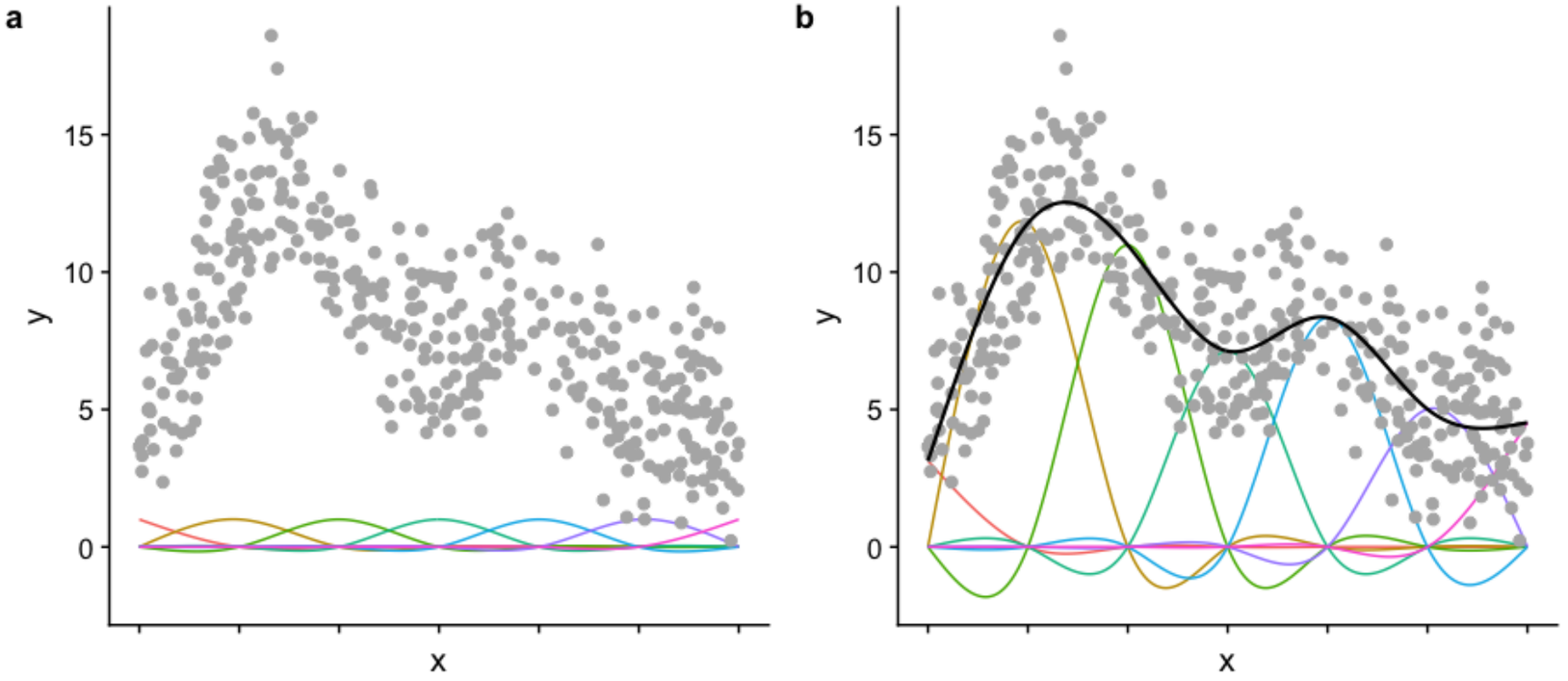


Nonlinear relationships

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



Basis functions



Basis functions

```
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
```

Let's practice!

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

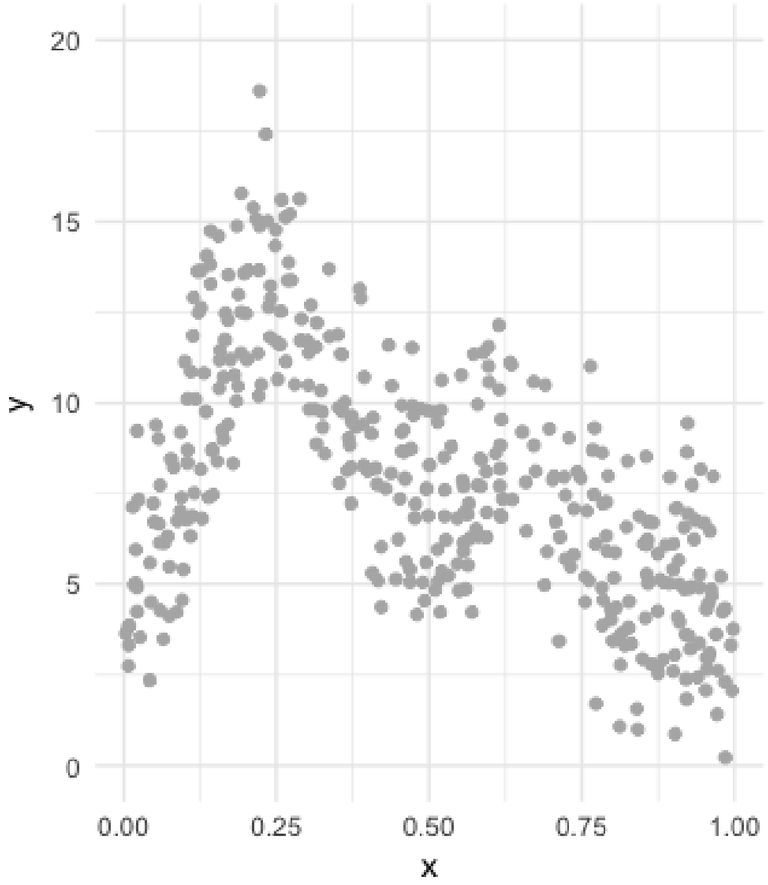
Basis Functions and Smoothing

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Noam Ross

Senior Research Scientist, EcoHealth Alliance

Getting the right fit

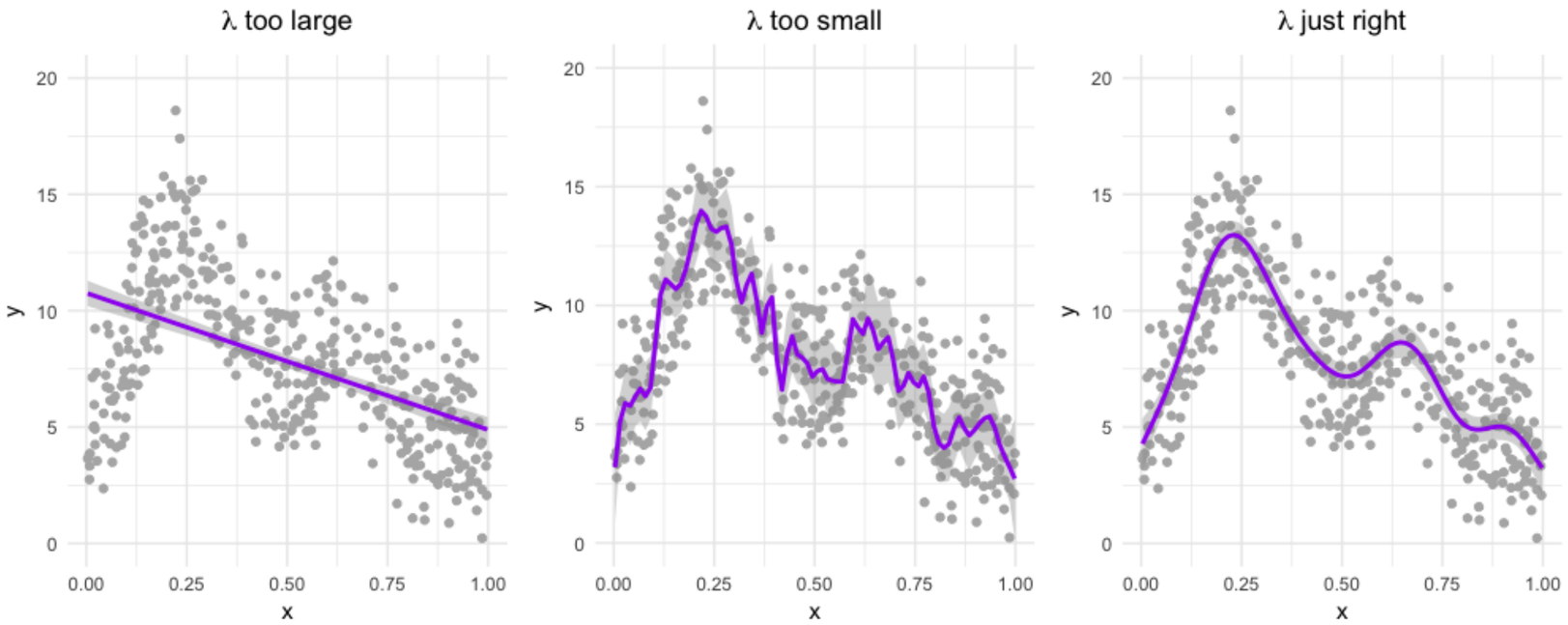


- Close to the data (avoiding under-fitting)
- Not fitting the noise (avoiding over-fitting)

Balancing Wiggliness

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

Choosing the right smoothing parameter



Smoothing syntax

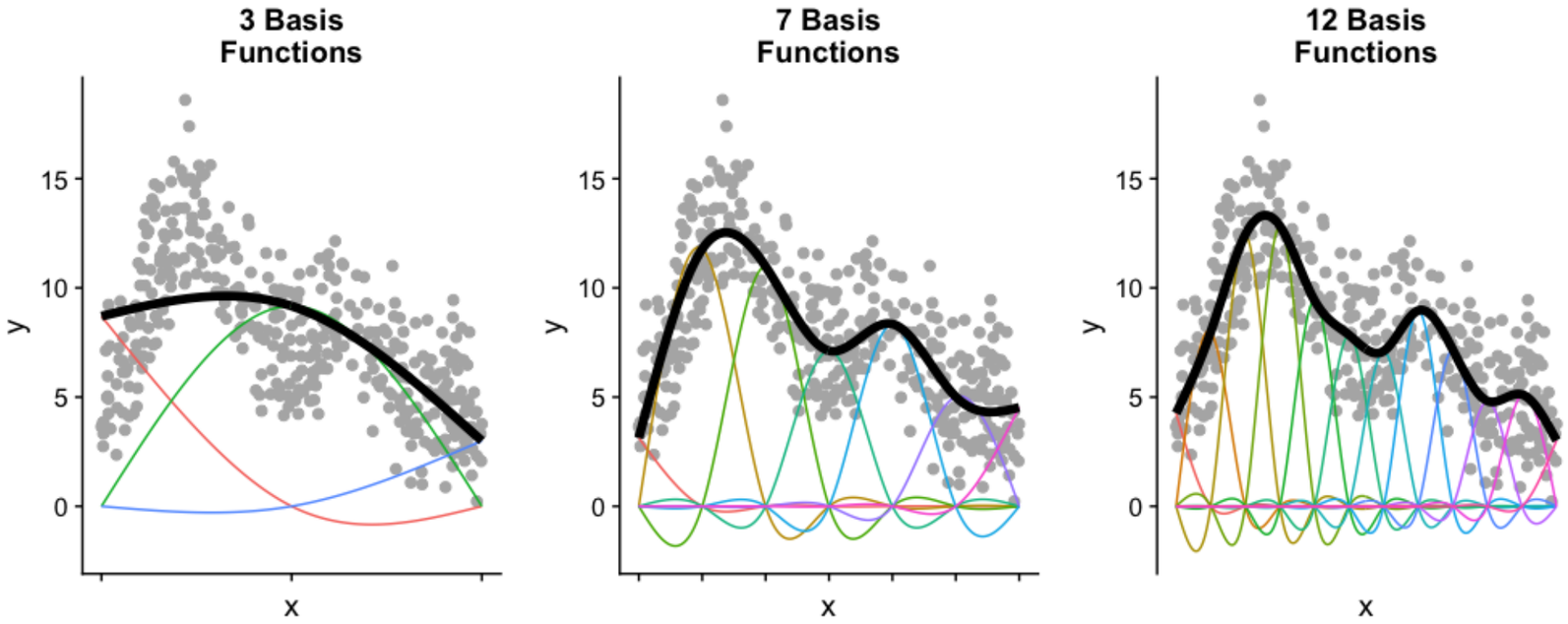
Setting a fixed smoothing parameter

```
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")
```

Let's practice!

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Multivariate GAMs

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Noam Ross

Senior Research Scientist, EcoHealth Alliance

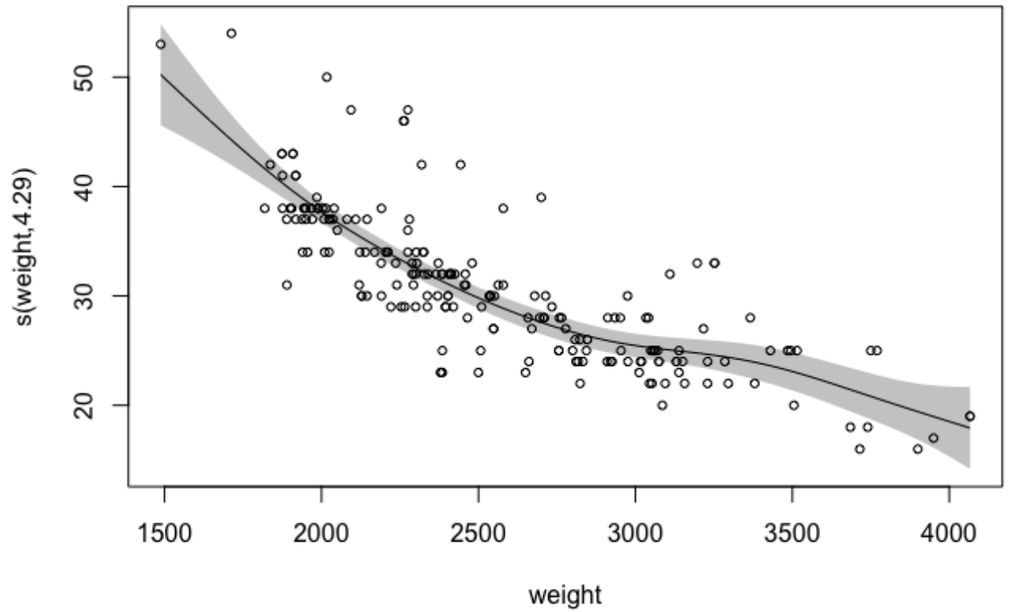
Our working dataset: mpg

```
mpg
```

```
  symbol loss         make  fuel aspir doors ... eng.loc    wb length width
1      3   NA alfa-romero   gas   std   two ...   front  88.6  168.8  64.1
2      3   NA alfa-romero   gas   std   two ...   front  88.6  168.8  64.1
3      1   NA alfa-romero   gas   std   two ...   front  94.5  171.2  65.5
4      2  164         audi   gas   std  four ...   front  99.8  176.6  66.2
5      2  164         audi   gas   std  four ...   front  99.4  176.6  66.4
6      2   NA         audi   gas   std   two ...   front  99.8  177.3  66.3
7      1  158         audi   gas   std  four ...   front 105.8  192.7  71.4
8      1   NA         audi   gas   std  four ...   front 105.8  192.7  71.4
9      1  158         audi   gas turbo  four ...   front 105.8  192.7  71.4
10     0   NA         audi   gas turbo   two ...   front  99.5  178.2  67.9
...
```

Multiple smooths (1)

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



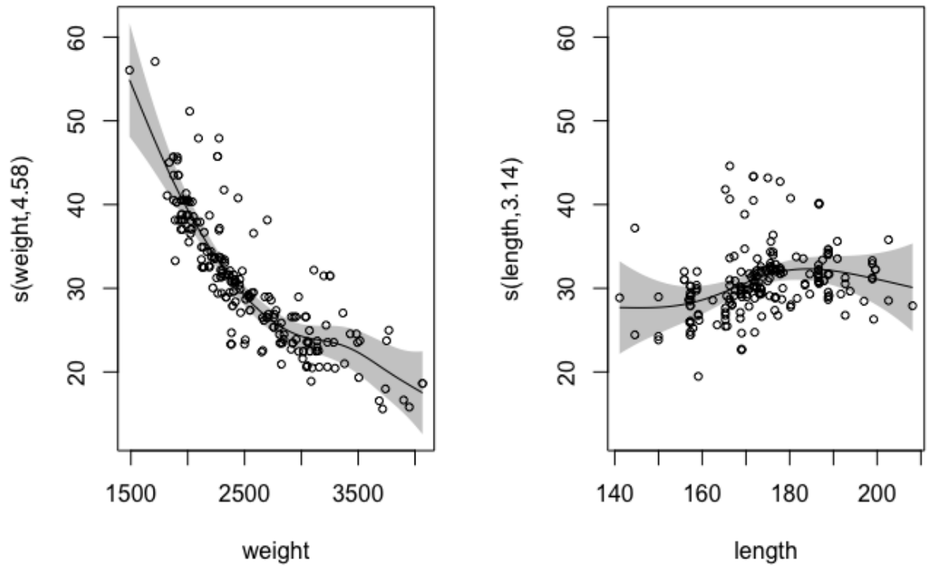
Multiple smooths (2)

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

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

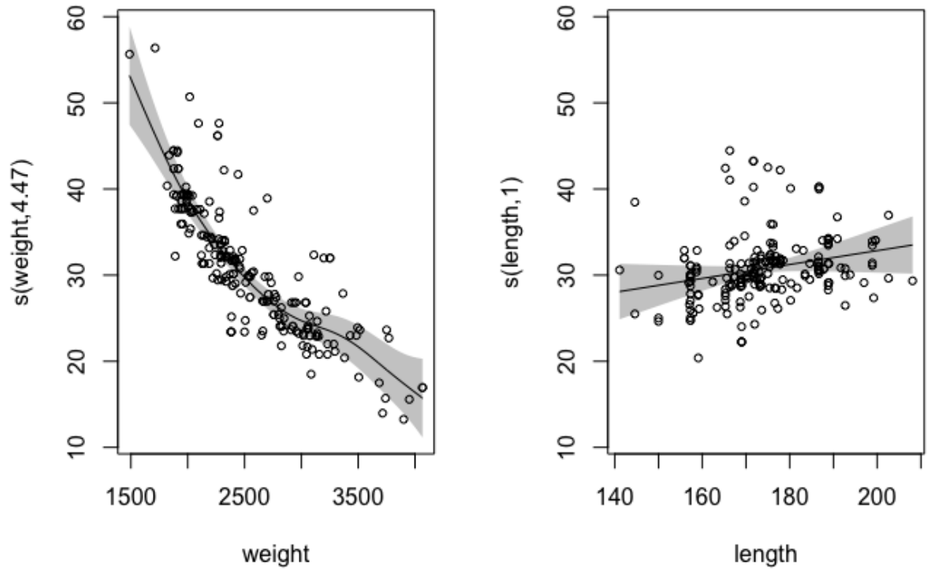
Multiple smooths (3)

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



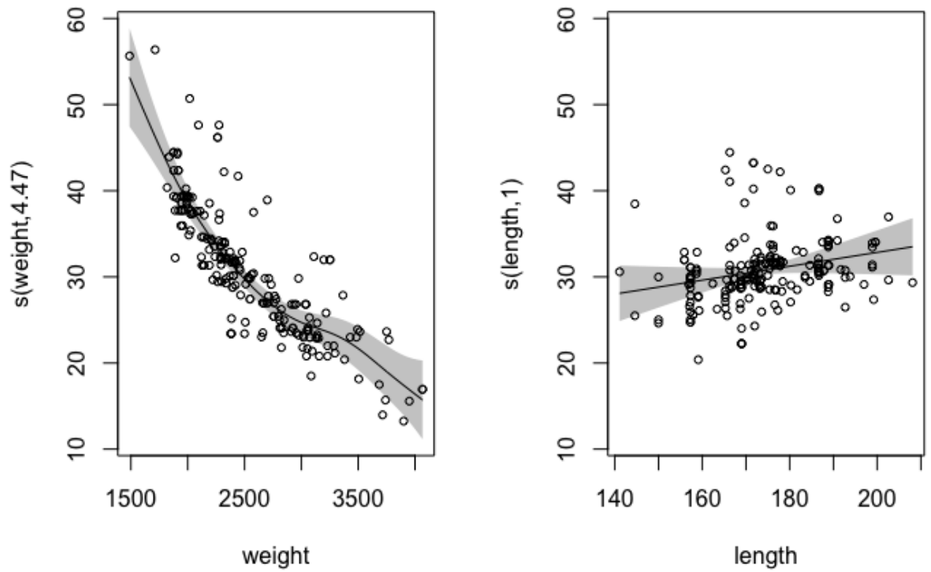
Linear terms

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



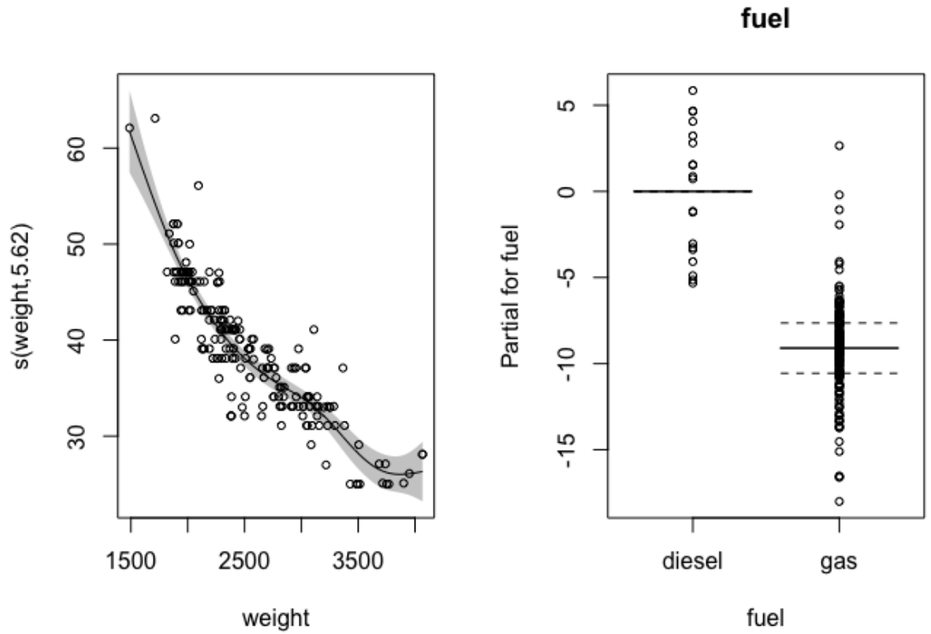
Linear terms (2)

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



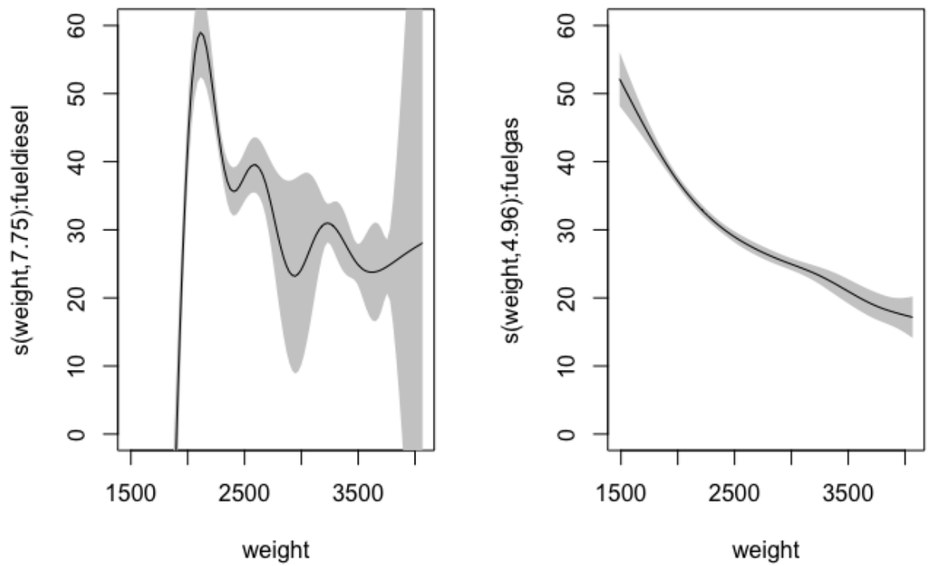
Categorical terms (1)

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



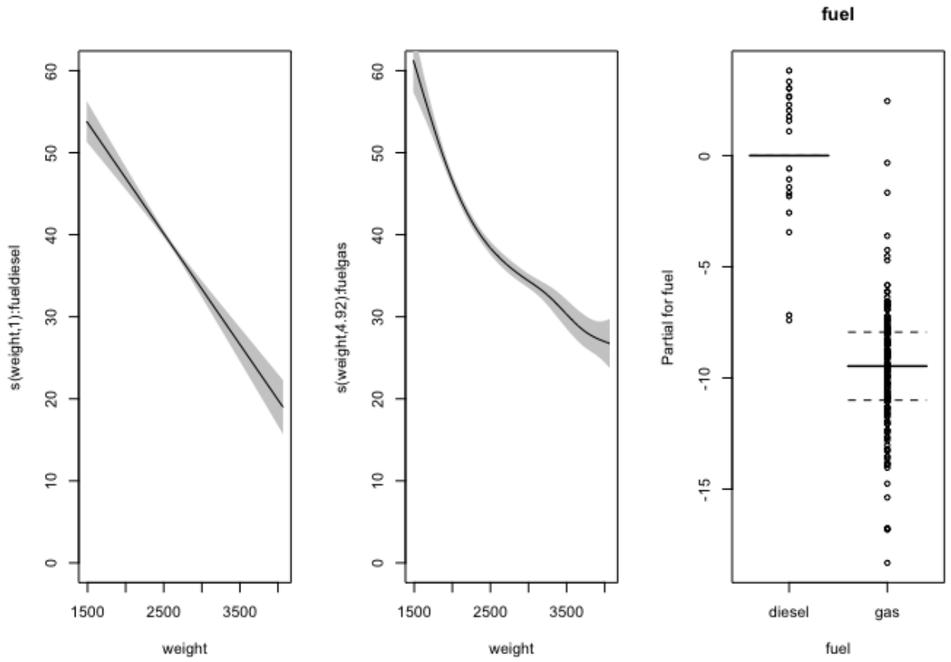
Categorical terms (2)

```
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")
```



Let's practice!

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R