Interpreting GAM outputs

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Noam Ross
Senior Research Scientist, EcoHealth Alliance

datacamp

GAM summaries

```
mod_hwy <- gam(hw.mpg ~ s(weight) + s(rpm) +
               s(price) + s(comp.ratio) +
               s(width) + fuel + cylinders,
               data = mpg, method = "REML")
```

```
summary(mod_hwy)
```

```
Family: gaussian
Link function: identity

Formula:
hw.mpg ~ s(weight) + s(rpm) + s(price) + s(comp.ratio) +
  s(width) + fuel

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.873      3.531   6.760 1.89e-10 ***
fuelgas        7.571      3.922   1.931   0.0551 .
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(weight)     6.254  7.439 20.909  < 2e-16 ***
s(rpm)        7.499  8.285  8.534 2.07e-09 ***
s(price)      2.681  3.421  1.678    0.155
s(comp.ratio) 1.000  1.001 18.923 2.22e-05 ***
s(width)      1.001  1.001  0.357    0.551
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.89   Deviance explained = 90.1%
REML = 464.81  Scale est. = 5.171     n = 199
```

GAM summaries (3)

```
summary(mod_hwy)
```

```
Family: gaussian
Link function: identity

Formula:
hw.mpg ~ s(weight) + s(rpm) + s(price) +
  s(comp.ratio) + s(width) + fuel
...
```

GAM summaries (4)

```
summary(mod_hwy)
```

```
...
Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.873      3.531   6.760 1.89e-10 ***
fuelgas        7.571      3.922   1.931   0.0551 .
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

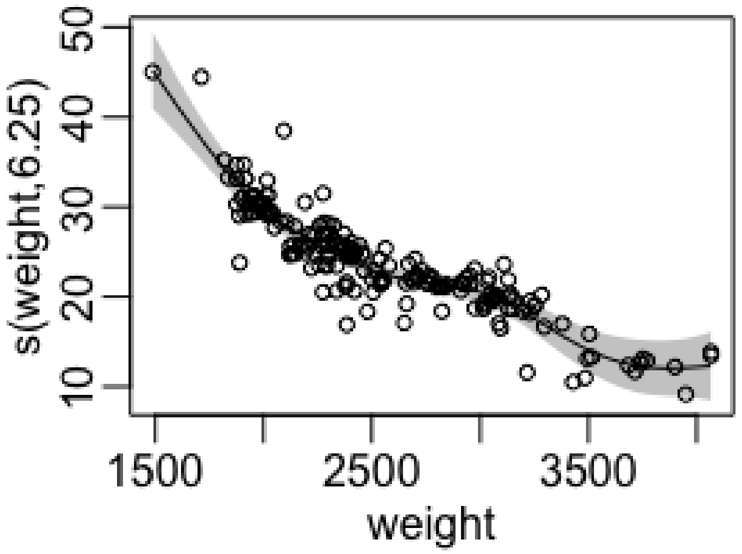
GAM summaries (5)

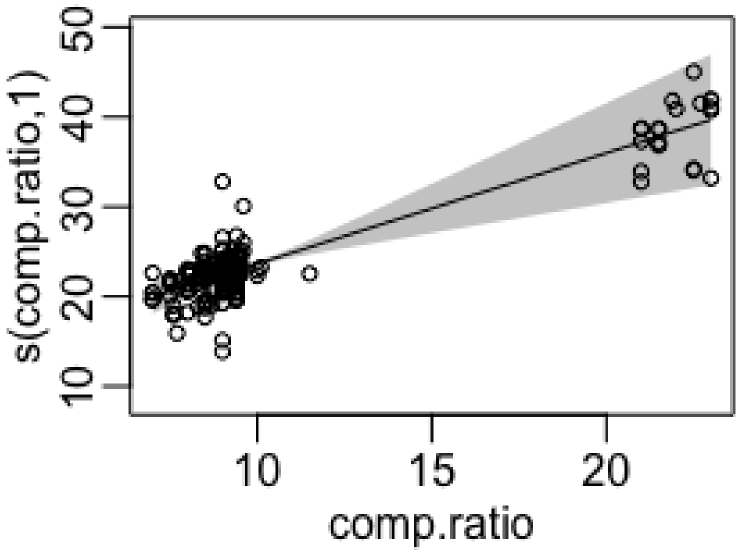
```
summary(mod_hwy)
```

```
...
Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(weight)     6.254  7.439 20.909  < 2e-16 ***
s(rpm)        7.499  8.285  8.534 2.07e-09 ***
s(price)      2.681  3.421  1.678    0.155
s(comp.ratio) 1.000  1.001 18.923 2.22e-05 ***
s(width)      1.001  1.001  0.357    0.551
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
```

Effective degrees of freedom

```
Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(weight)     6.254  7.439 20.909  < 2e-16 ***  <--
s(rpm)        7.499  8.285  8.534 2.07e-09 ***
s(price)      2.681  3.421  1.678    0.155
s(comp.ratio) 1.000  1.001 18.923 2.22e-05 ***  <--
s(width)      1.001  1.001  0.357    0.551
```



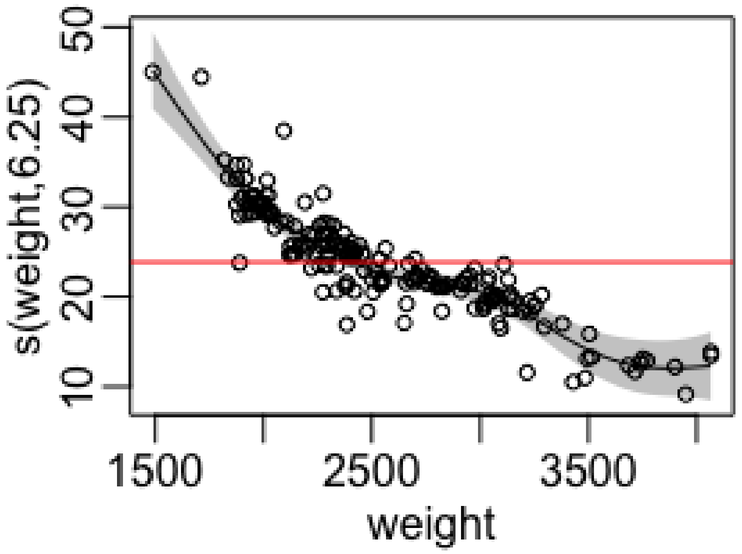


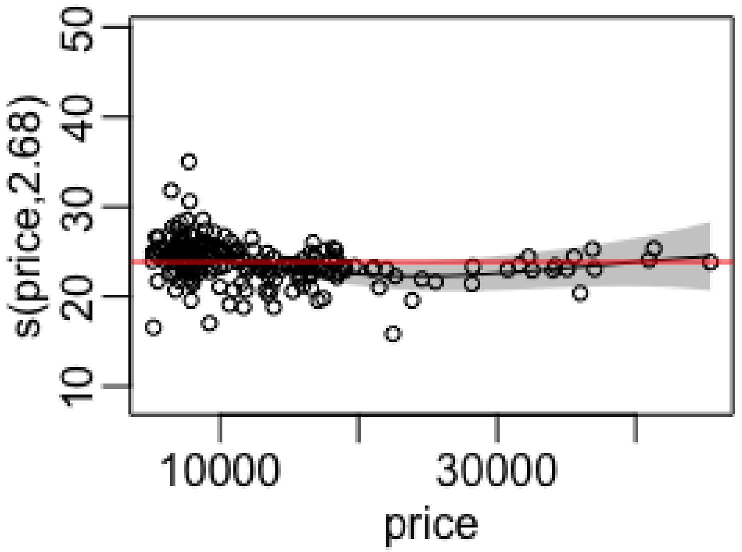
Significance of smooth terms

```
Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(weight)     6.254  7.439 20.909  < 2e-16 ***
s(rpm)        7.499  8.285  8.534 2.07e-09 ***
s(price)      2.681  3.421  1.678    0.155
s(comp.ratio) 1.000  1.001 18.923 2.22e-05 ***
s(width)      1.001  1.001  0.357    0.551
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significance of smooth terms (2)

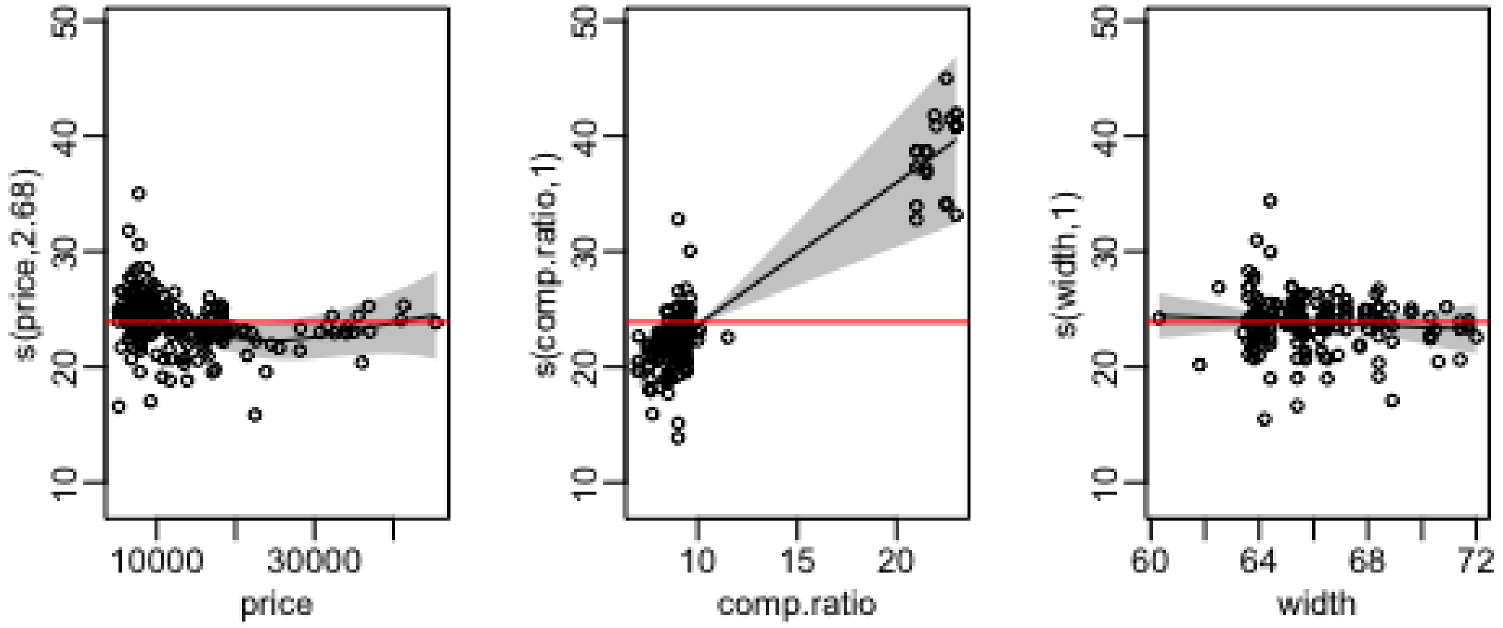
```
Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(weight)     6.254  7.439 20.909  < 2e-16 ***  <--
s(rpm)        7.499  8.285  8.534 2.07e-09 ***
s(price)      2.681  3.421  1.678    0.155      <--
s(comp.ratio) 1.000  1.001 18.923 2.22e-05 ***
s(width)      1.001  1.001  0.357    0.551
```





Significance and effective degrees of freedom

```
Approximate significance of smooth terms:
                 edf Ref.df      F  p-value
s(weight)      6.254  7.439 20.909  < 2e-16 ***
s(rpm)         7.499  8.285  8.534 2.07e-09 ***
s(price)       2.681  3.421  1.678    0.155      <--
s(comp.ratio) 1.000  1.001 18.923 2.22e-05 ***  <--
s(width)       1.001  1.001  0.357    0.551      <--
```



Let's practice!

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

The plot command

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Noam Ross
Senior Research Scientist, EcoHealth Alliance

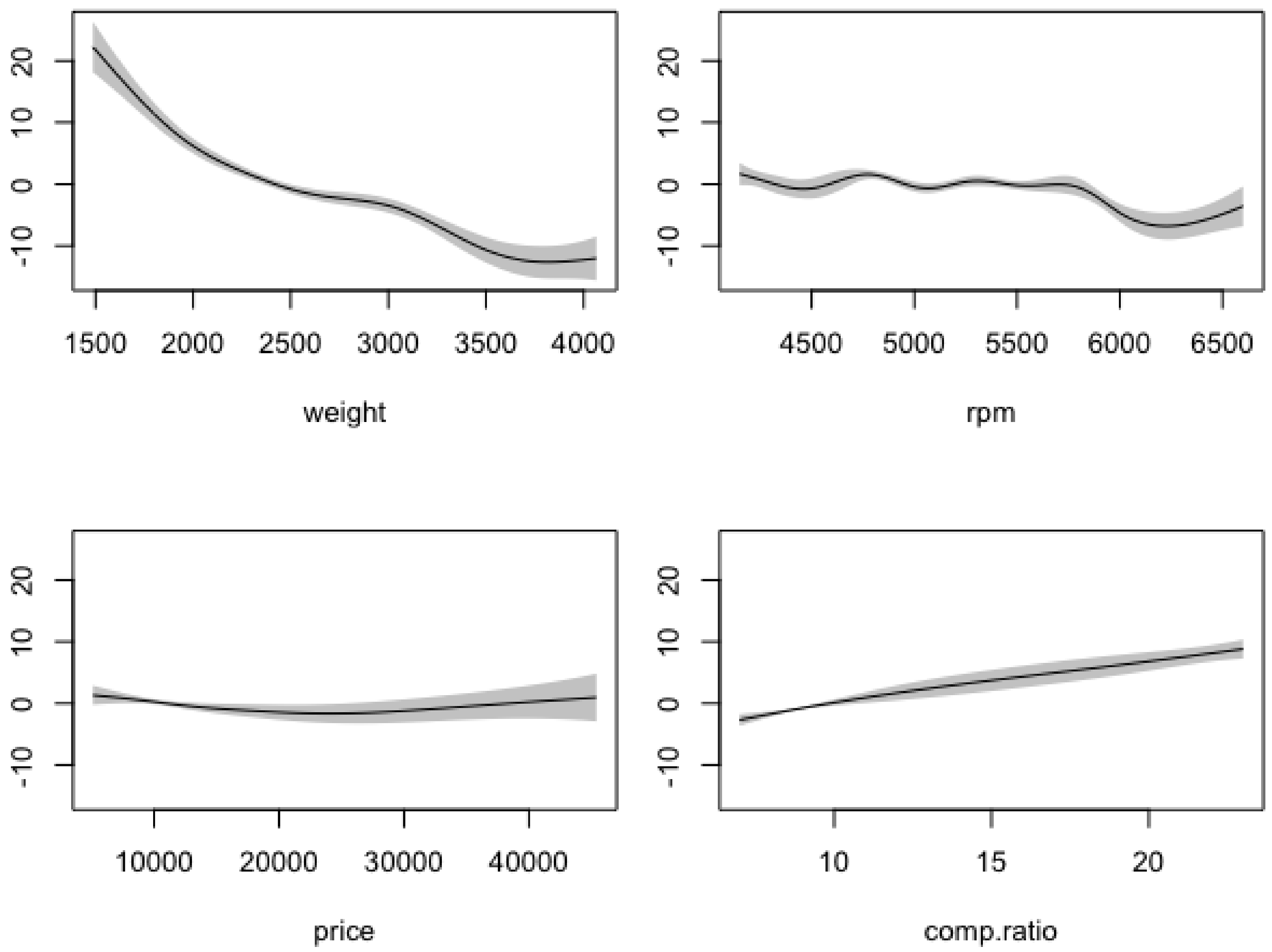
datacamp

The plot command

```
plot(gam_model)
```

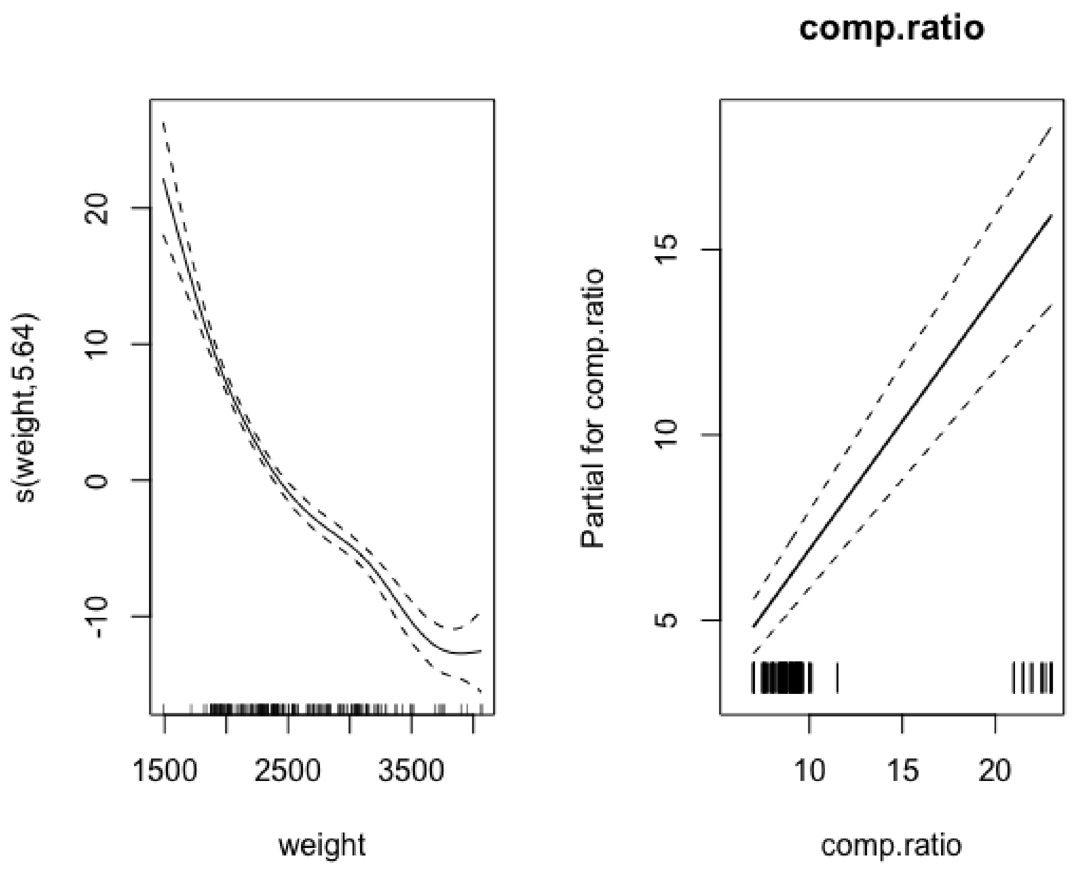
```
?plot.gam
```

Partial Effect Plots



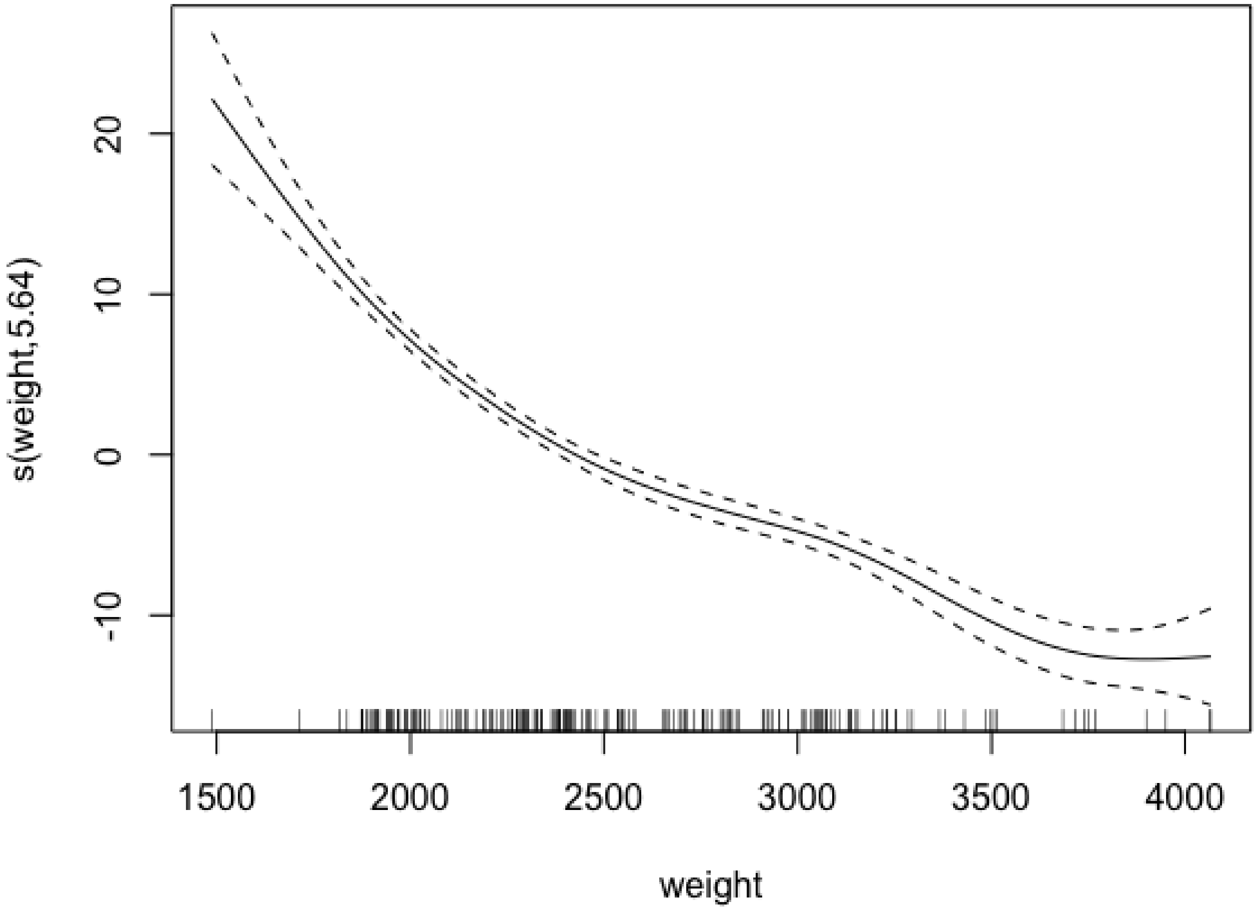
Selecting partial effects

```
plot(gam_model, select = c(2, 3))
plot(gam_model, pages = 1)
plot(gam_model, pages = 1, all.terms = TRUE)
```



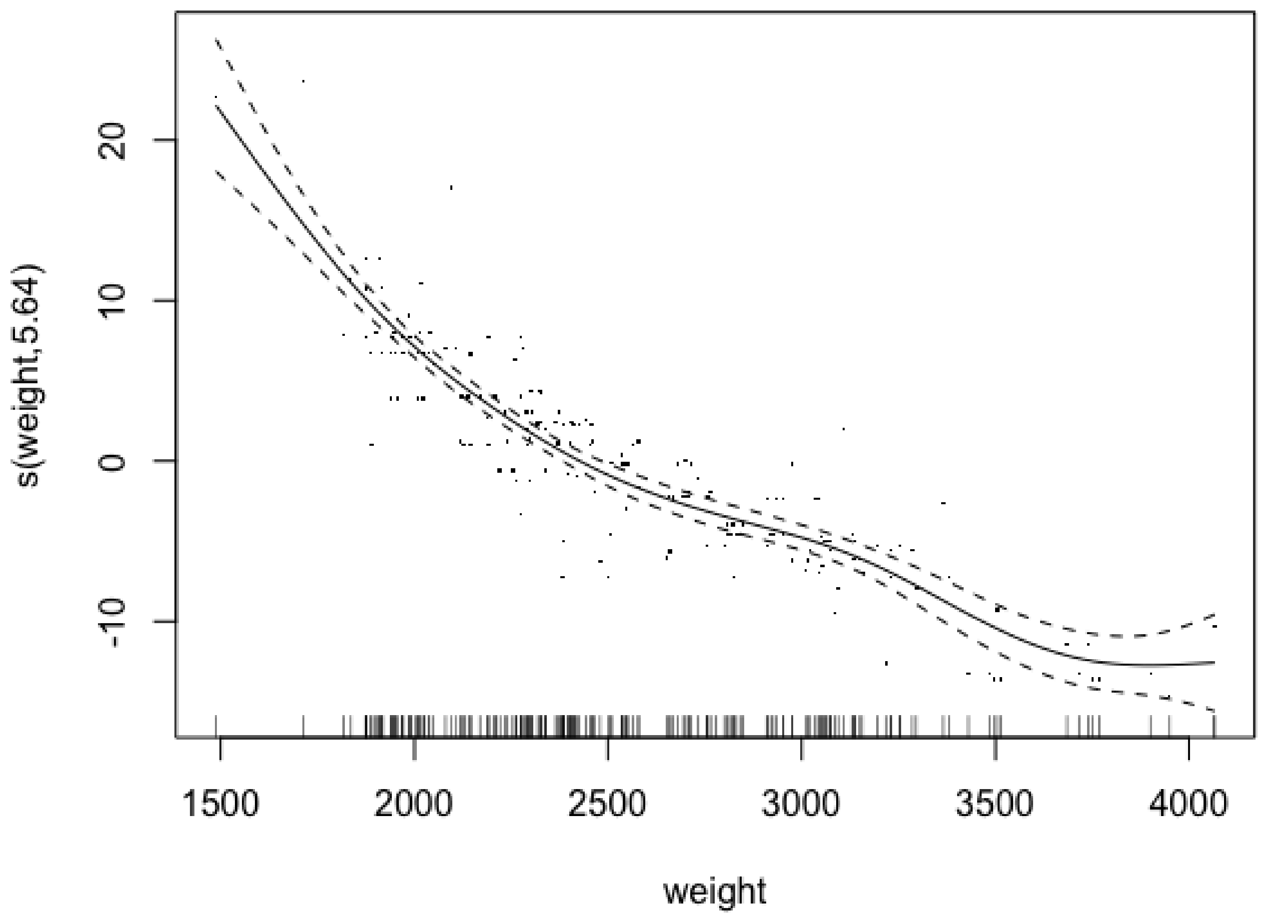
Showing data on the plots

```
plot(gam_model, rug = TRUE)
```



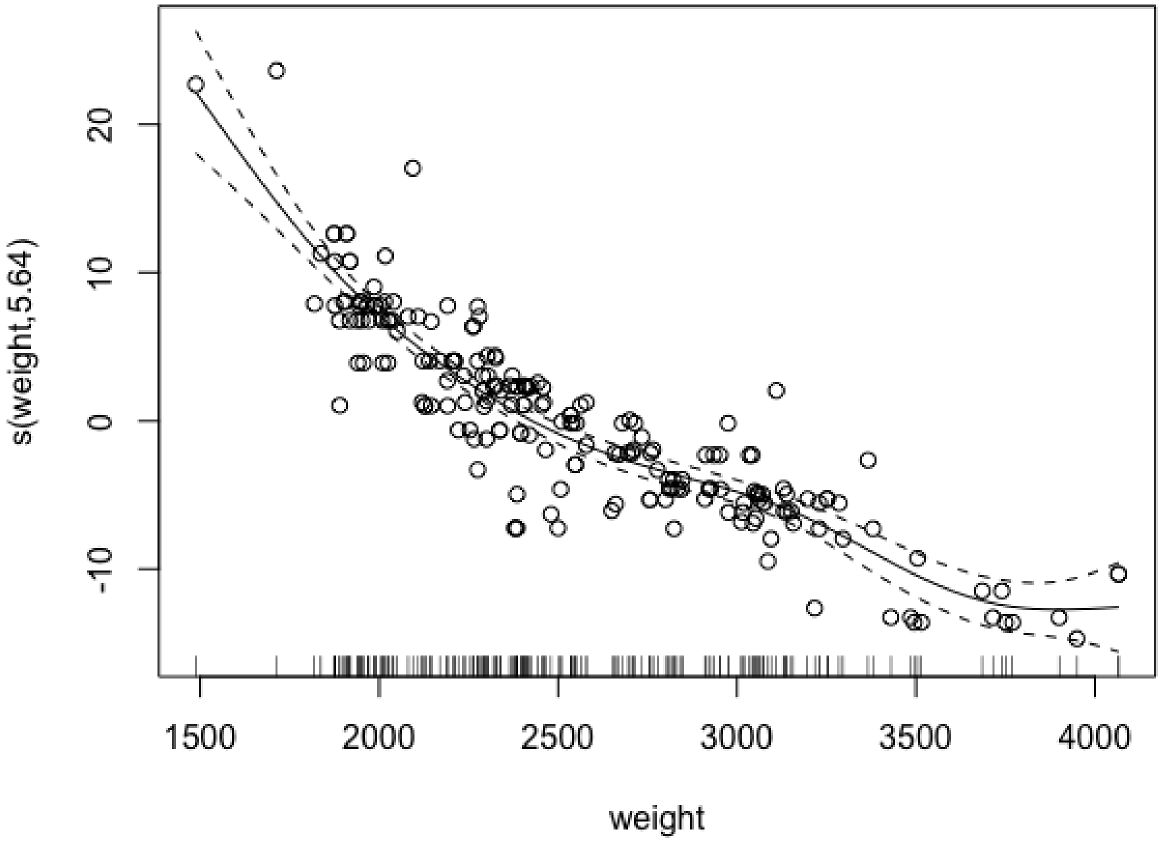
Showing data on the plots (2)

```
plot(gam_model, residuals = TRUE)
```



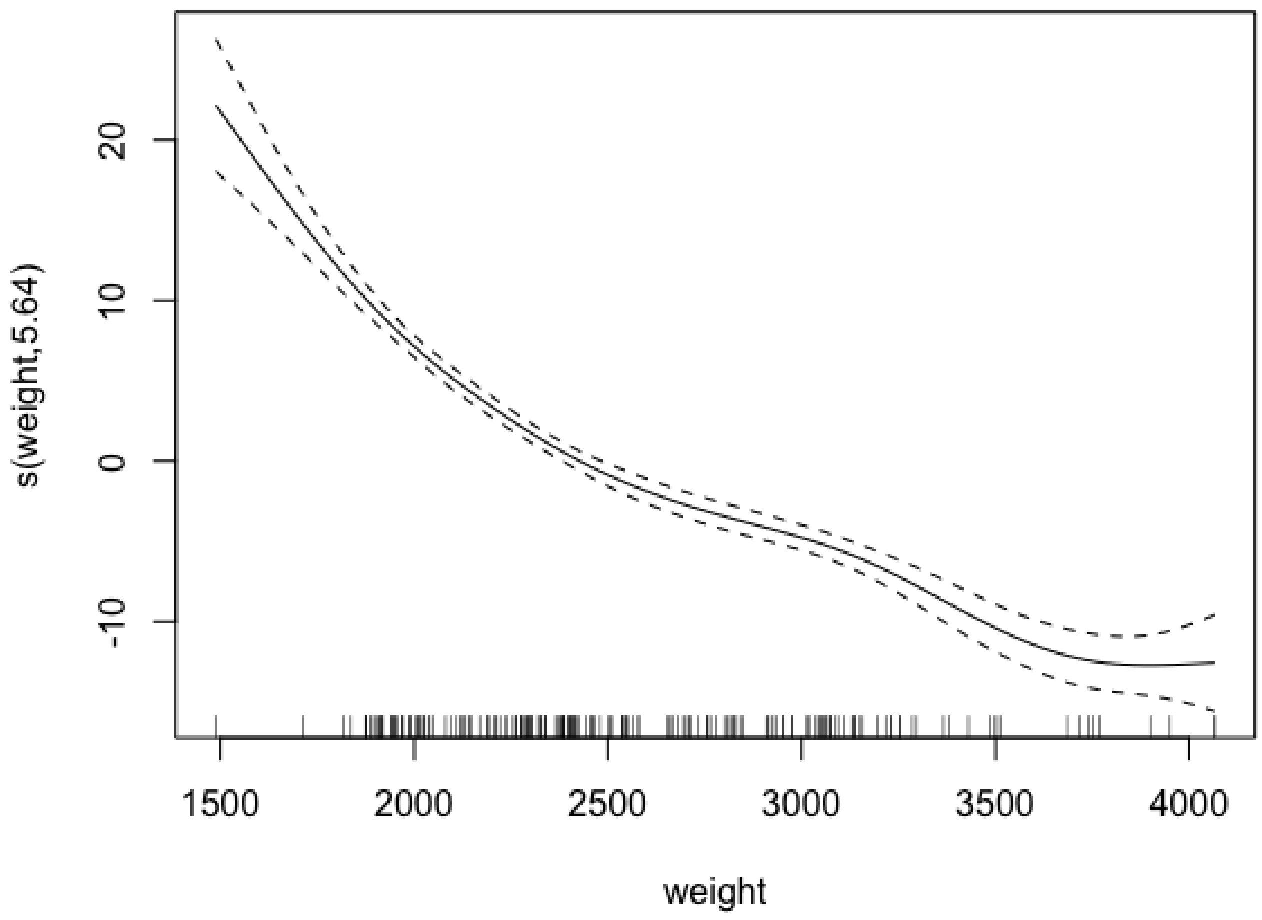
Showing data on the plots (3)

```
plot(gam_model, rug = TRUE, residuals = TRUE,
     pch = 1, cex = 1)
```



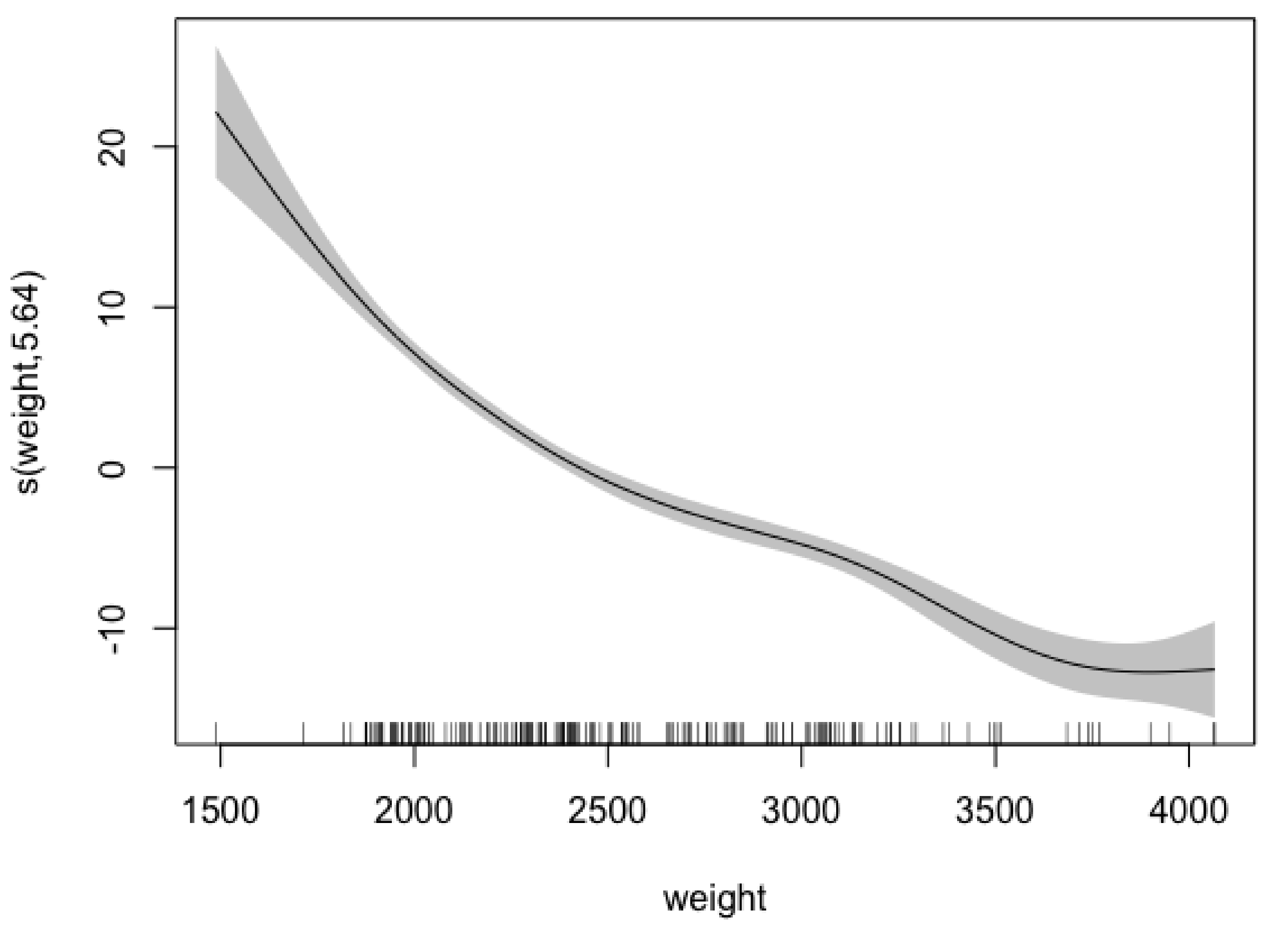
Showing standard errors

```
plot(gam_model, se = TRUE)
```



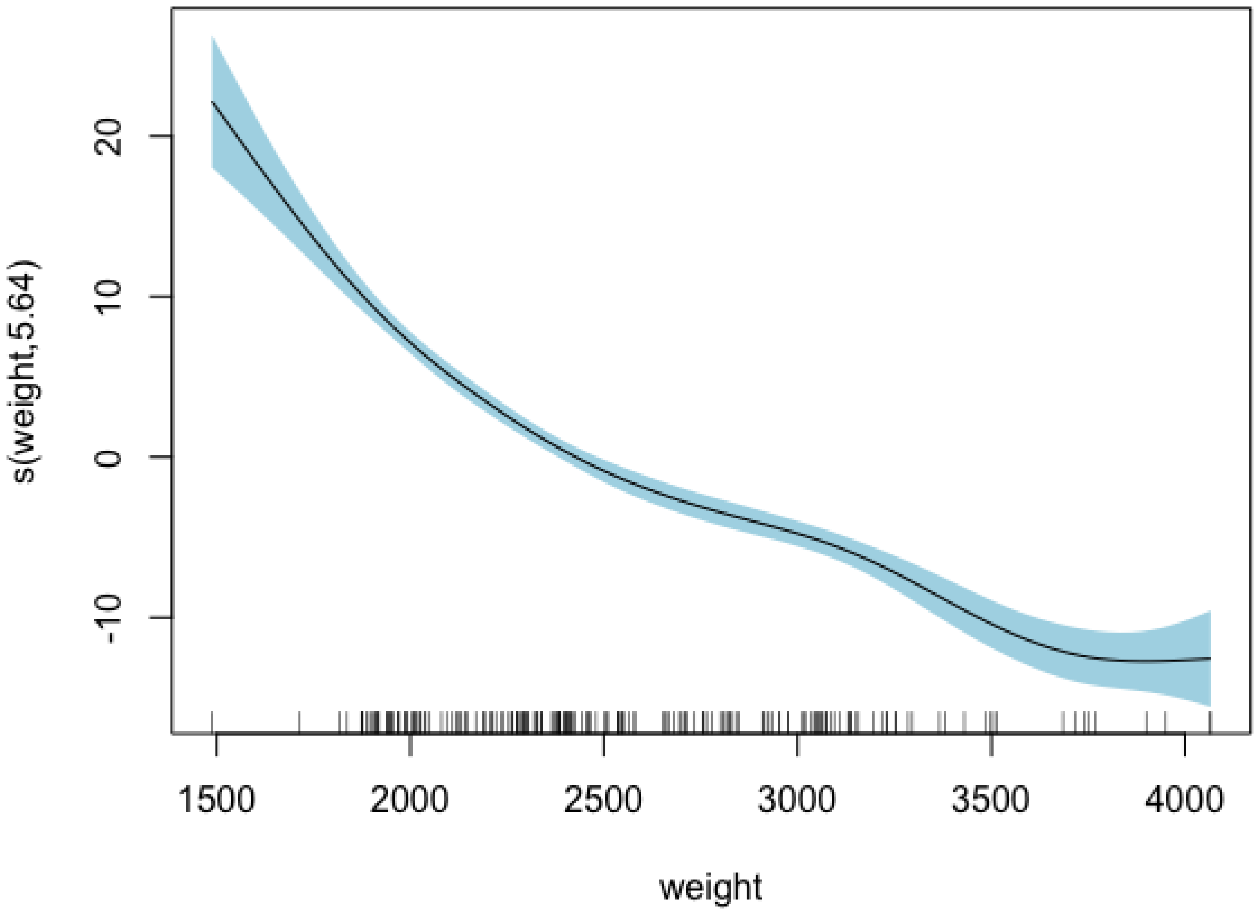
Showing standard errors (2)

```
plot(gam_model, shade = TRUE)
```



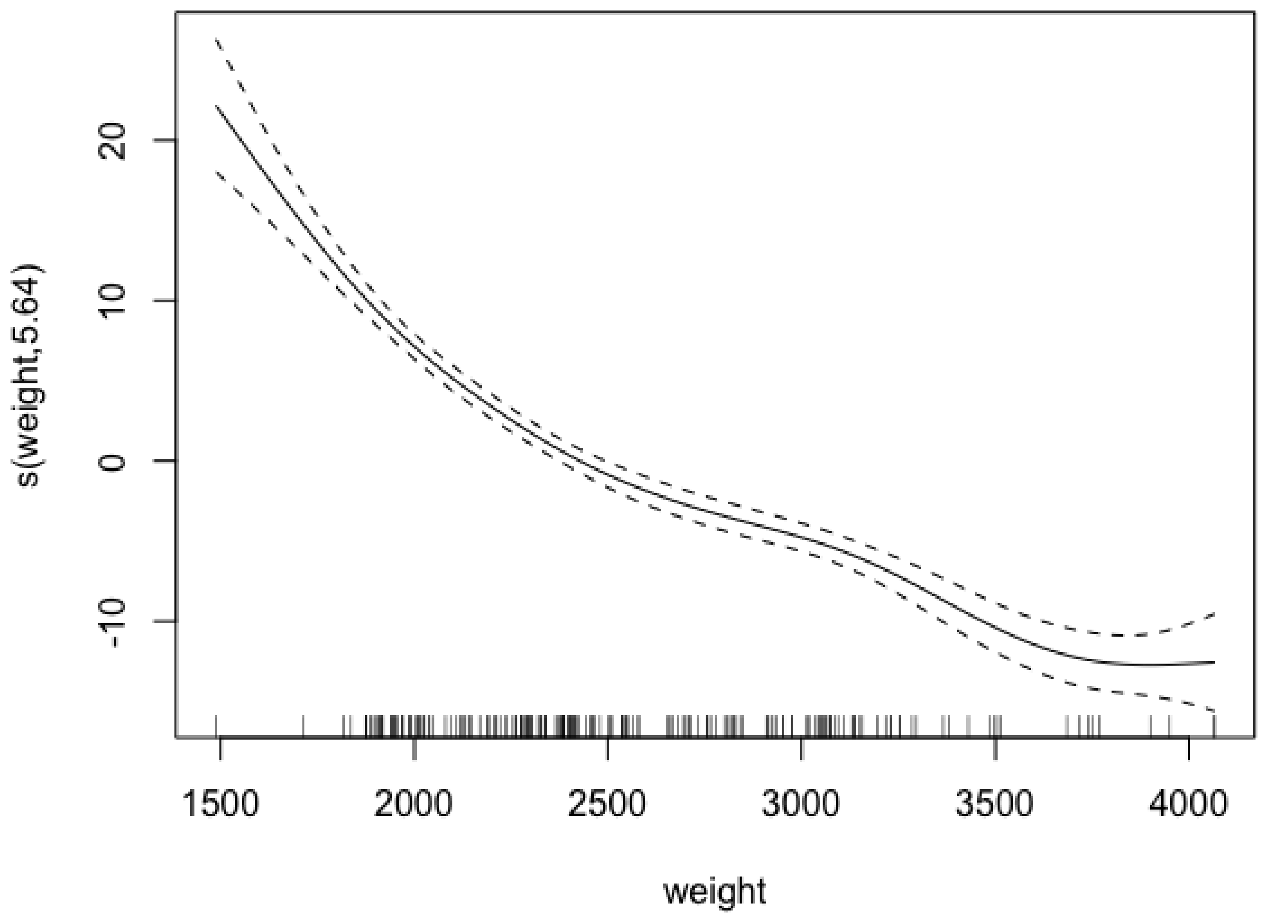
Showing standard errors

```
plot(gam_model, shade = TRUE, shade.col = "lightblue")
```



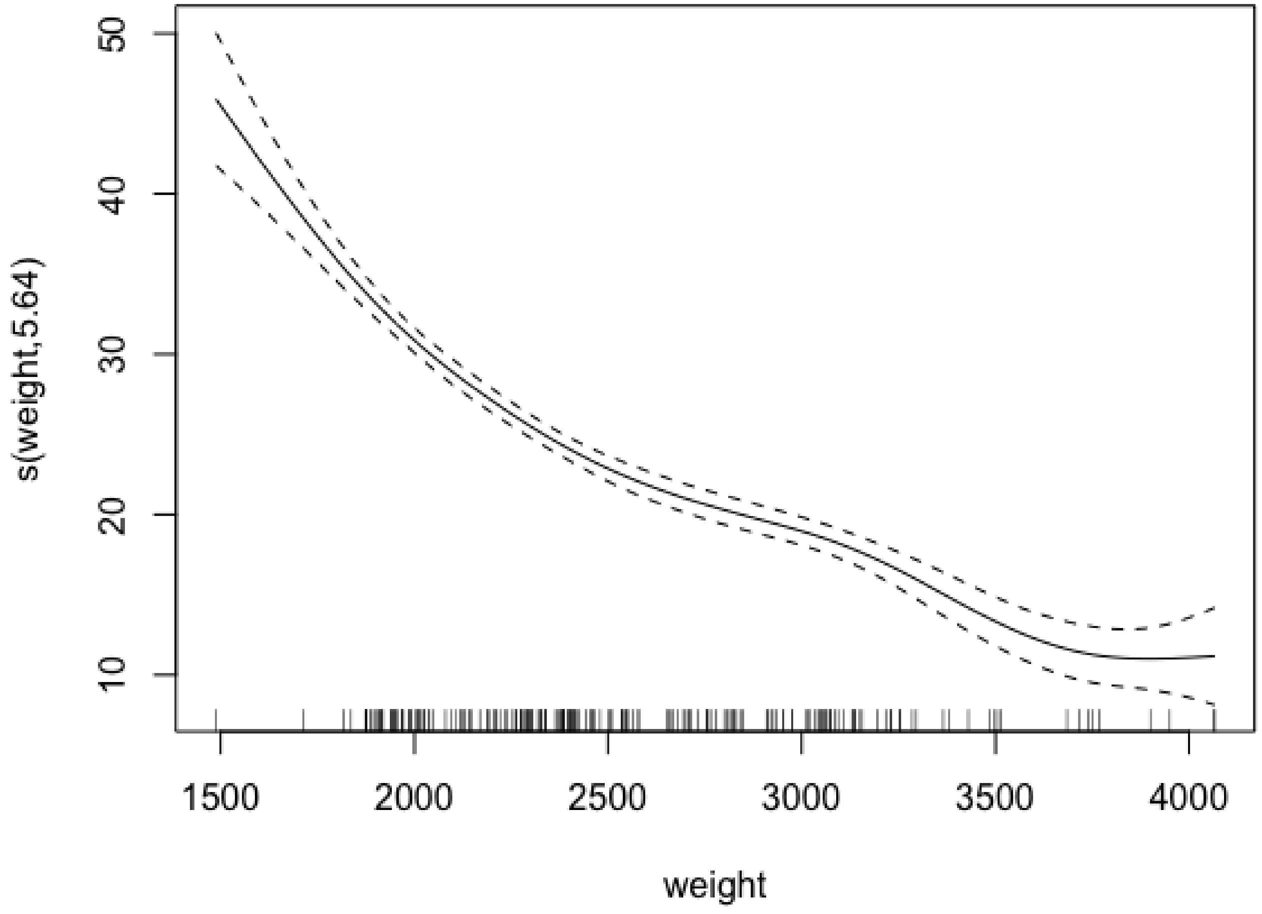
Transforming standard errors

```
plot(gam_model, seWithMean = TRUE)
```



Transforming standard errors (2)

```
plot(gam_model, seWithMean = TRUE, shift = coef(gam_model)[1])
```



Now lets make some plots!

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Model checking with gam.check()

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

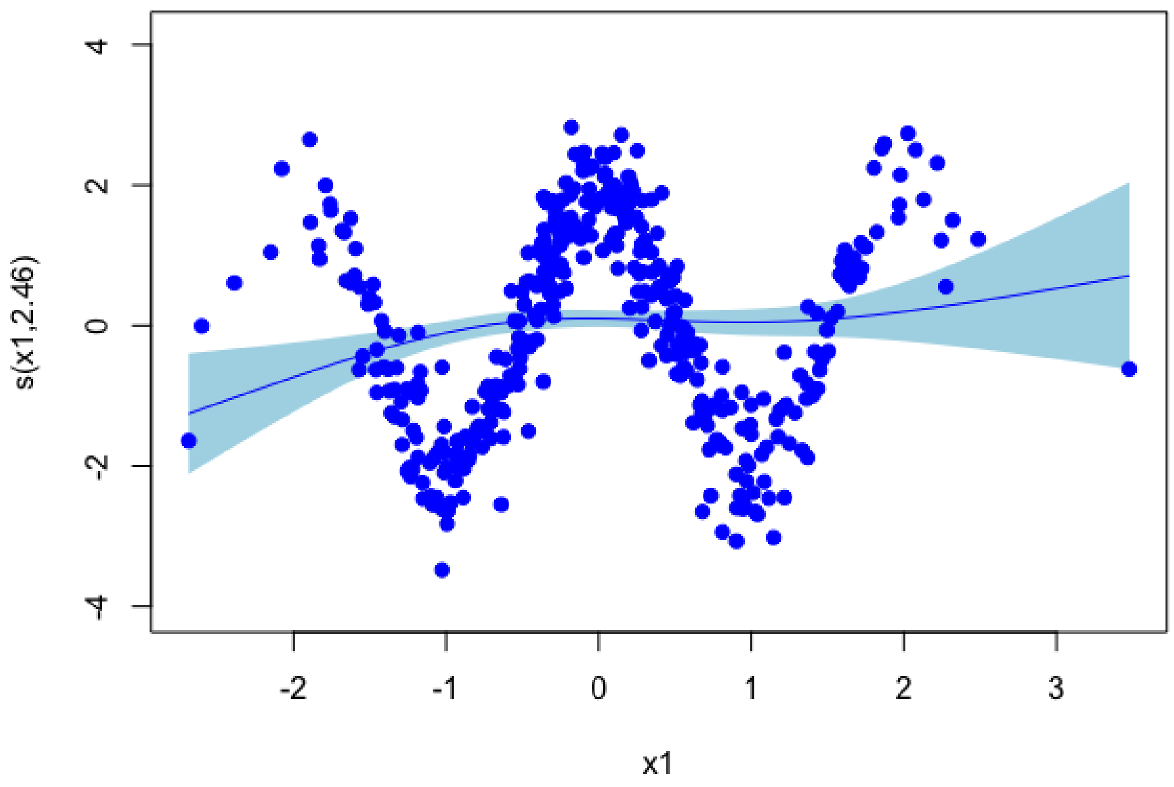
Noam Ross

Senior Research Scientist, EcoHealth Alliance

datacamp

Pitfall one: inadequate basis number

```
mod <- gam(y ~ s(x1, k = 4) + s(x2, k = 4),
           data = check_data, method = "REML")
```



```
gam.check(mod)
```

```
Method: REML   Optimizer: outer newton
full convergence after 9 iterations.
Gradient range [-0.0001467222,0.00171085]
(score 784.6012 & scale 2.868607).
Hessian positive definite, eigenvalue range [0.00014,198.5]
Model rank =  7 / 7

Basis dimension (k) checking results. Low p-value
(k-index<1) may indicate that k is too low, especially
if edf is close to k'.

        k'  edf k-index p-value
s(x1) 3.00 1.00    0.35  <2e-16 ***
s(x2) 3.00 2.88    1.00    0.52
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Running gam.check (2)

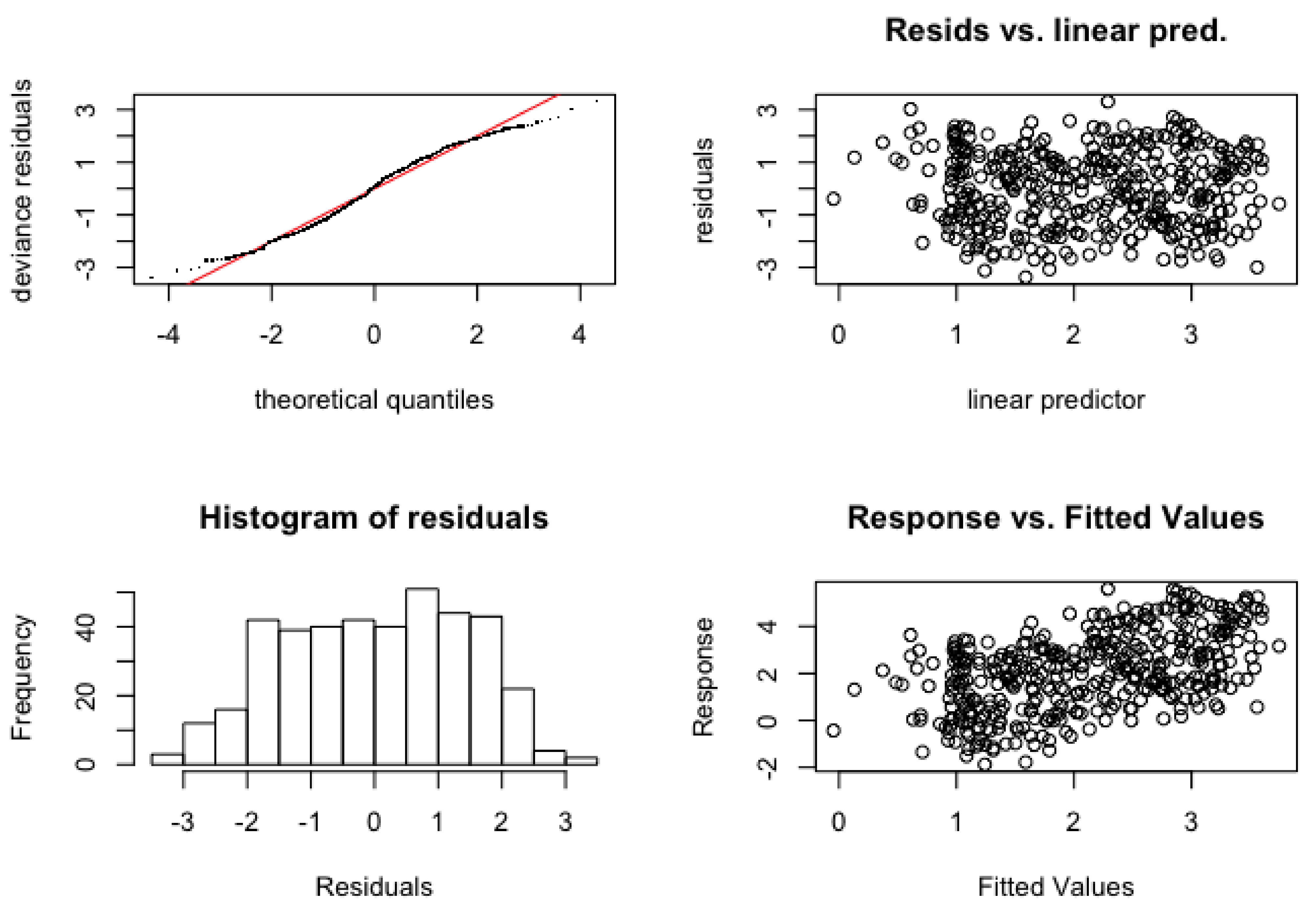
```
mod <- gam(y ~ s(x1, k = 12) + s(x2, k = 4),
                data = dat, method = "REML")
gam.check(mod)
```

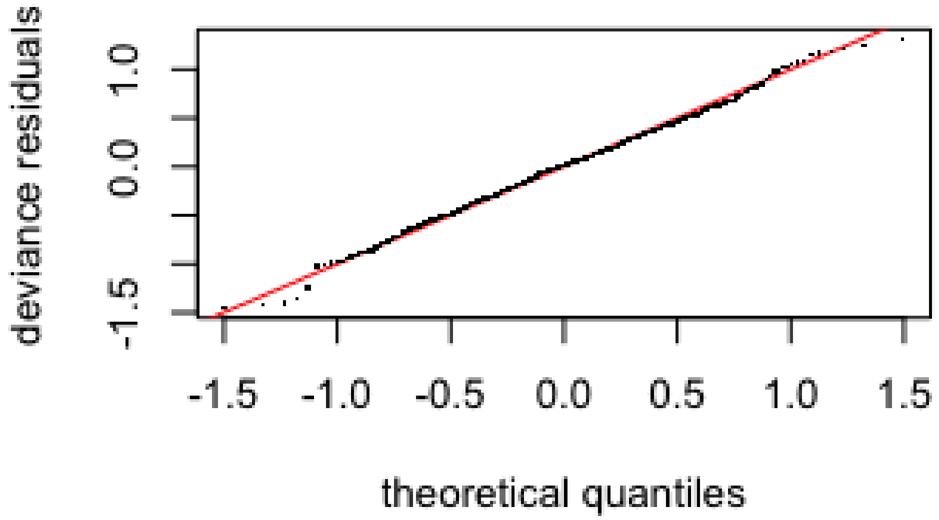
```
...

         k'   edf k-index p-value
s(x1) 11.00 10.85    1.05   0.830
s(x2)  3.00  2.98    0.89   0.015 *

...
```

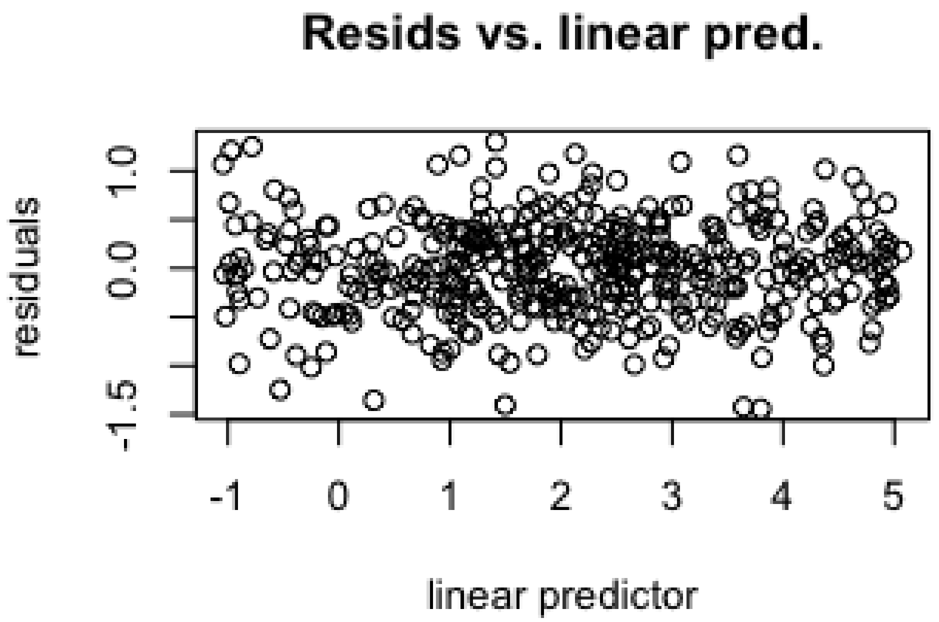
Running gam.check (3)

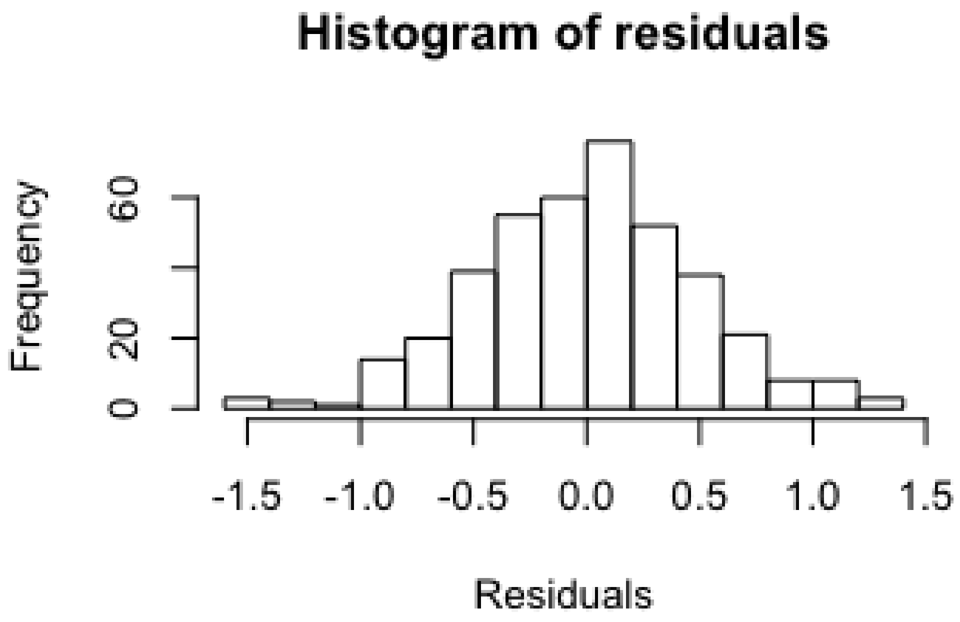
```
mod <- gam(y ~ s(x1, k = 12) + s(x2, k = 12),
            data = dat, method = "REML")
gam.check(mod)
```

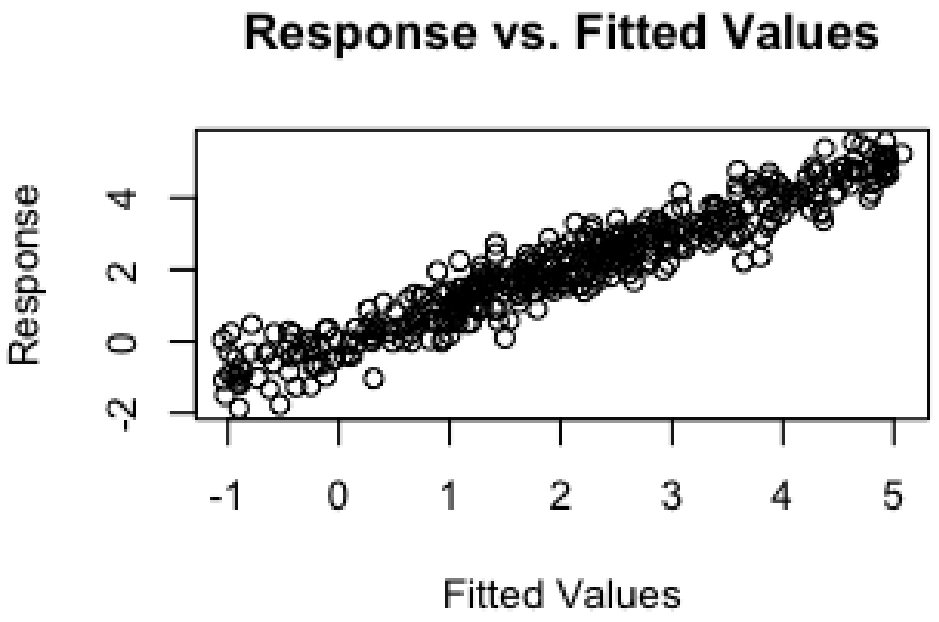
```
...
         k'   edf k-index p-value
s(x1) 11.00 10.86    1.08    0.94
s(x2) 11.00  7.78    0.94    0.12
...
```


Resids vs. linear pred.
deviance residuals
theoretical quantiles
residuals
linear predictor
Histogram of residuals
Frequency
Residuals
Response vs. Fitted Values
Response
Fitted Values


deviance residuals
-1.5
0.0
1.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
theoretical quantiles


Resids vs. linear pred.
residuals
-1.5
0.0
1.0
-1
0
1
2
3
4
5
linear predictor


Histogram of residuals
Frequency
0
20
60
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Residuals


Response vs. Fitted Values
Response
-2
0
2
4
-1
0
1
2
3
4
5
Fitted Values

Let's check some models

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

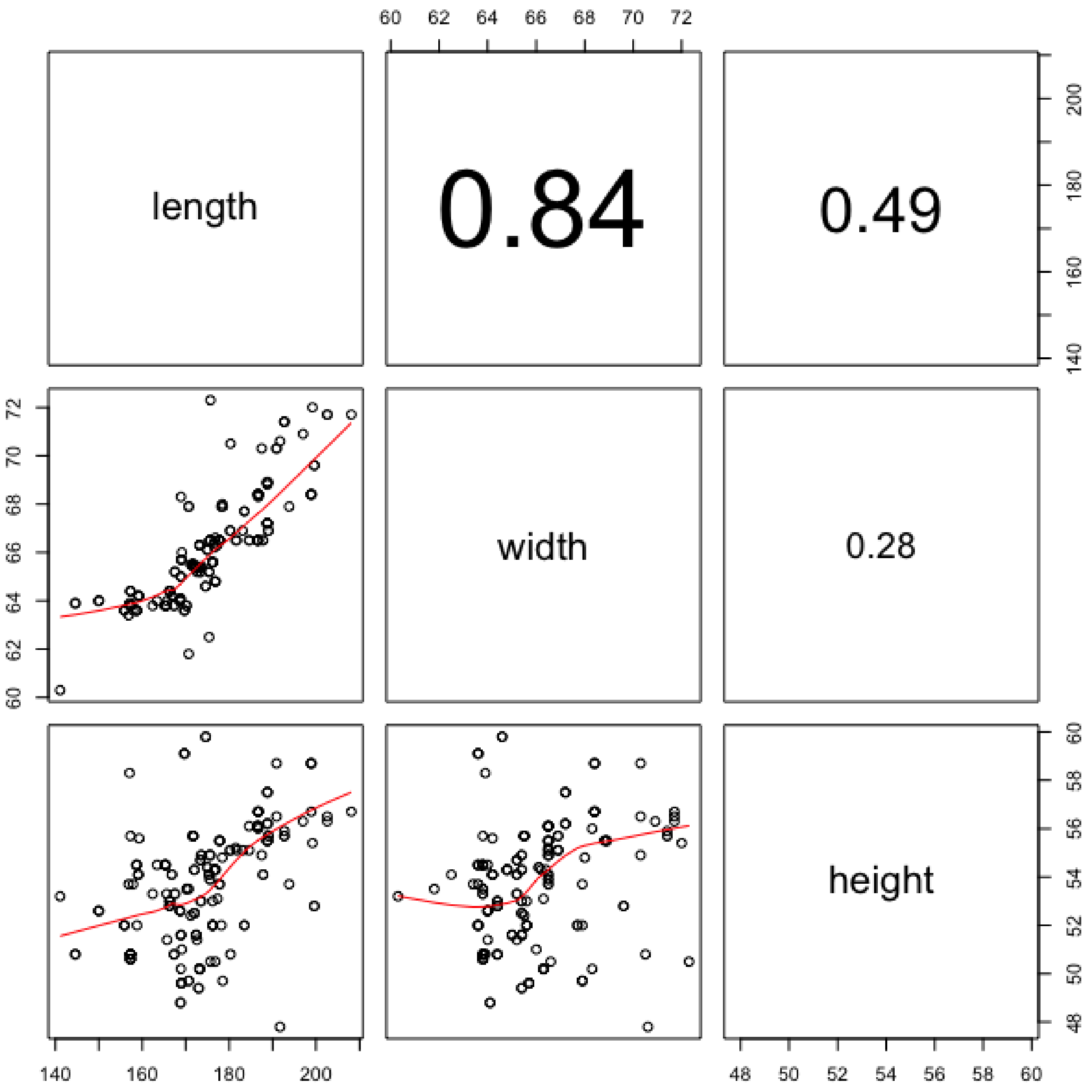
Checking concurvity

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

Noam Ross

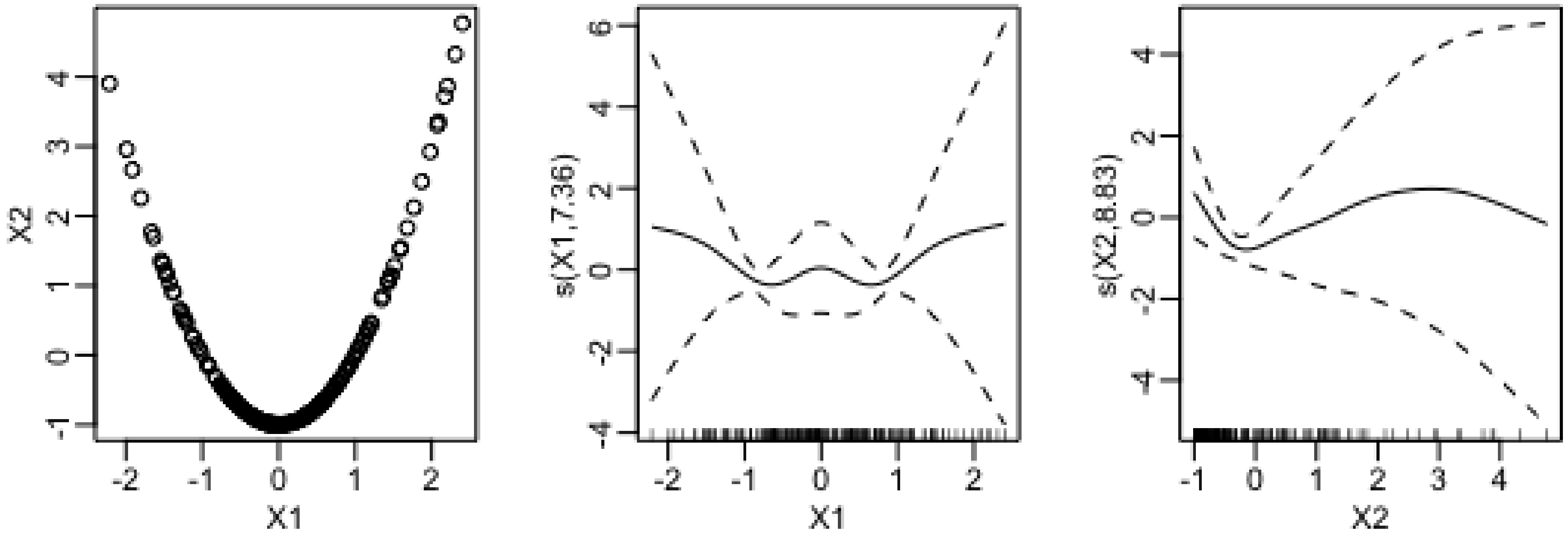
Senior Research Scientist, EcoHealth Alliance

datacamp

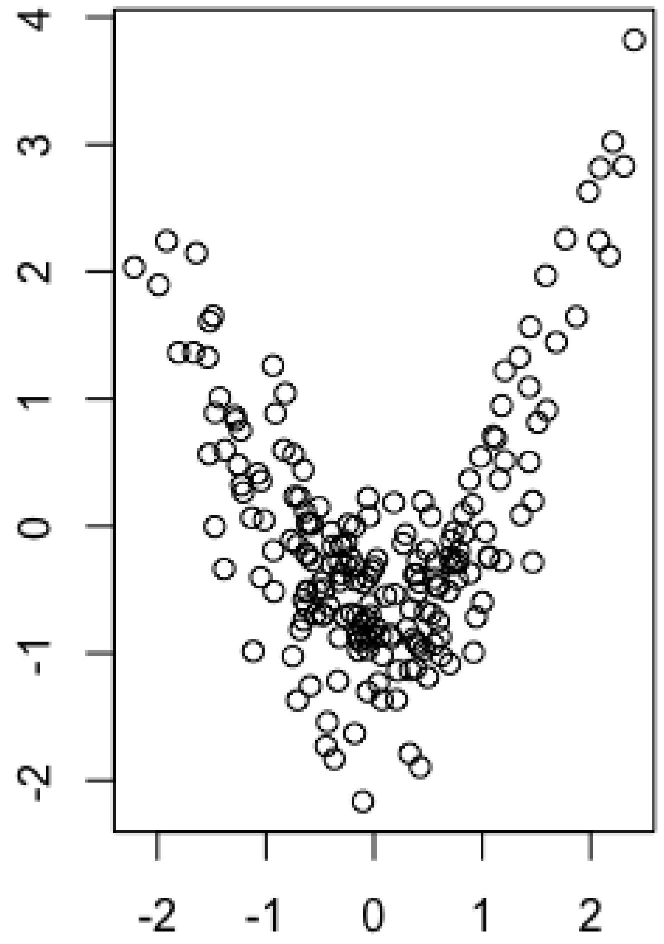


length
0.84
0.49
width
0.28
height
60 62 64 66 68 70 72
140 160 180 200
48 50 52 54 56 58 60

Concurvity



The concurvity() function




```
concurvity(m1, full = TRUE)
```

```
         para s(X1) s(X2)
worst       0  0.84  0.84
observed    0  0.22  0.57
estimate    0  0.28  0.60
```

Pairwise concurvities

```
concurvity(model, full = FALSE)
```

```
$worst
      para s(X1) s(X2)
para     1  0.00  0.00
s(X1)    0  1.00  0.84
s(X2)    0  0.84  1.00

$observed                   |  $estimate
      para s(X1) s(X2)      |        para s(X1) s(X2)
para     1  0.00  0.00      |  para     1  0.00   0.0
s(X1)    0  1.00  0.57      |  s(X1)    0  1.00   0.6
s(X2)    0  0.22  1.00      |  s(X2)    0  0.28   1.0
```

Let's practice!

NONLINEAR MODELING WITH GENERALIZED ADDITIVE MODELS (GAMS) IN R

datacamp