



NONLINEAR MODELING IN R WITH GAMs

Interpreting GAM outputs

Noam Ross

Senior Research Scientist, EcoHealth Alliance



GAM Summaries

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

```
summary(mod_hwy)
```



GAM Summaries (2)

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 ' ' 1
```

GAM Summaries (5)

```
summary(mod_hwy)
```

Approximate significance of smooth terms:

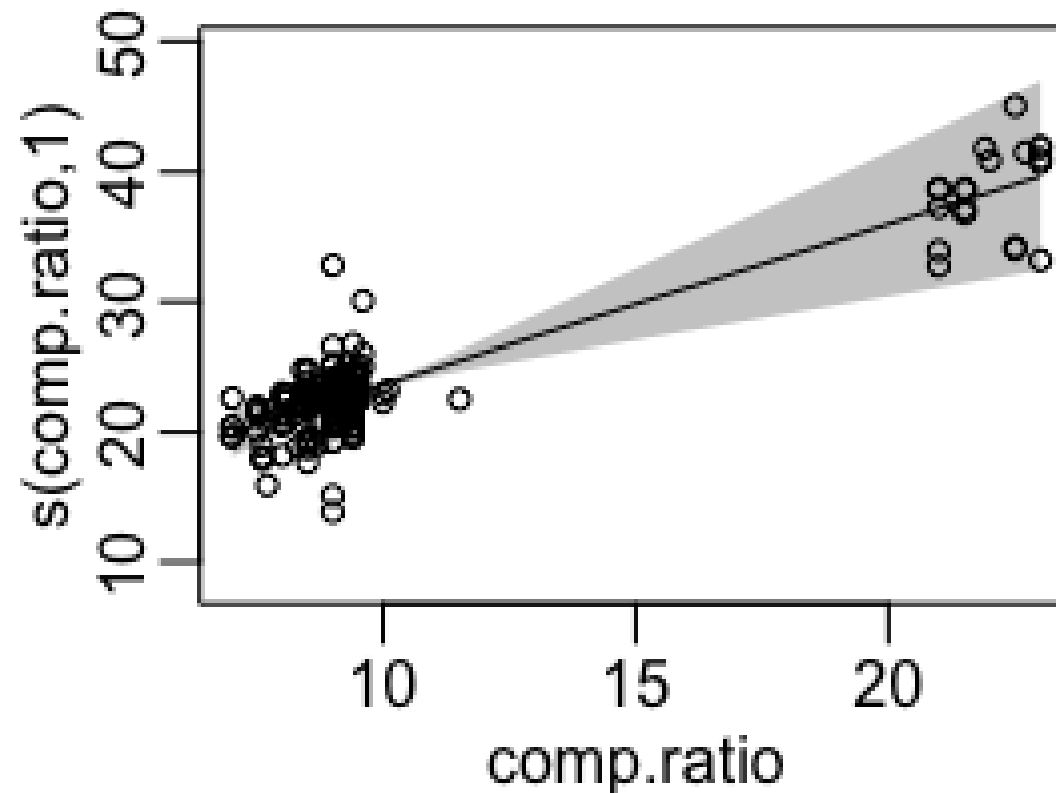
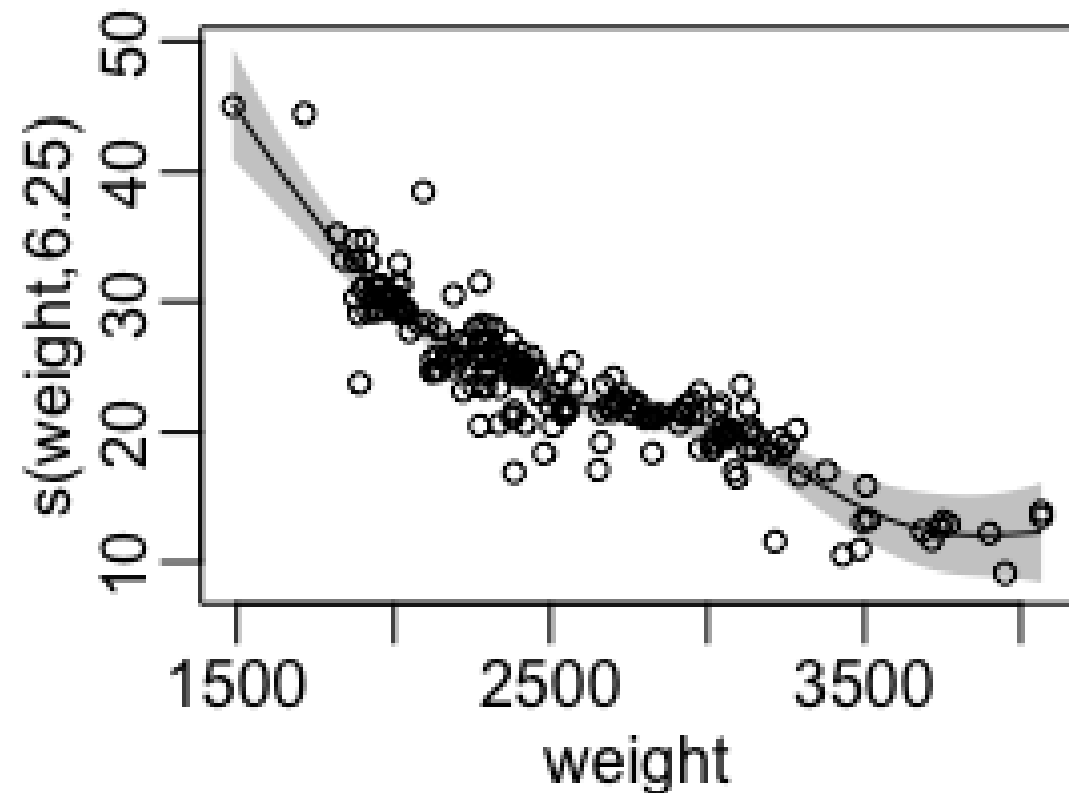
	edf	Ref.df	F	p-value	
s(weight)	6.254	7.439	20.909	< 2e-16	***
s(rpm)	7.499	8.285	8.534	2.07e-09	***
s(price)	2.681	3.421	1.678	0.155	
s(comp.ratio)	1.000	1.001	18.923	2.22e-05	***
s(width)	1.001	1.001	0.357	0.551	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0

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	

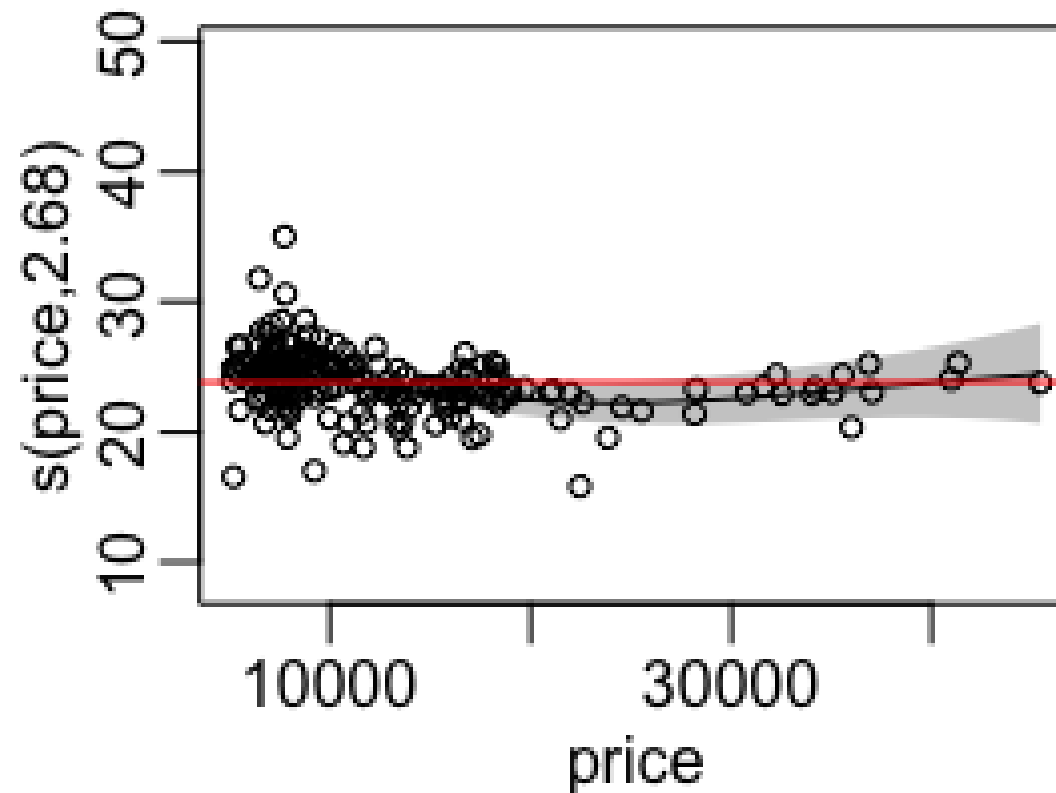
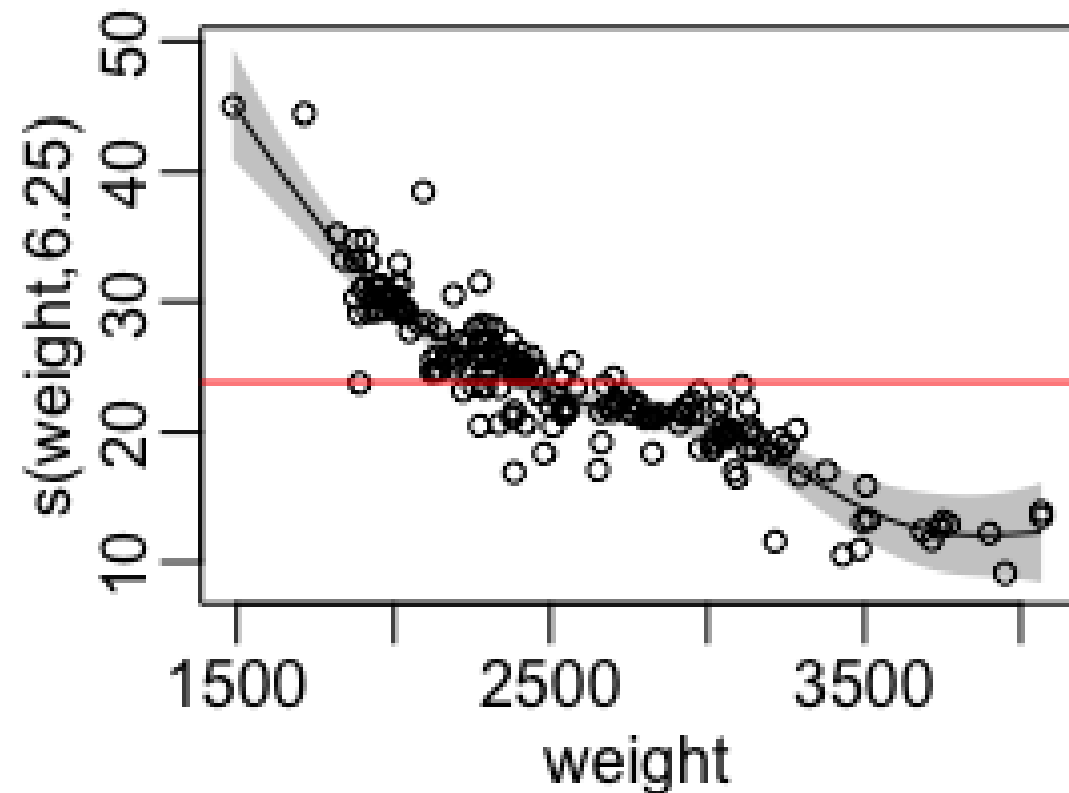
```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Significance of Smooth Terms (2)

Approximate significance of smooth terms:

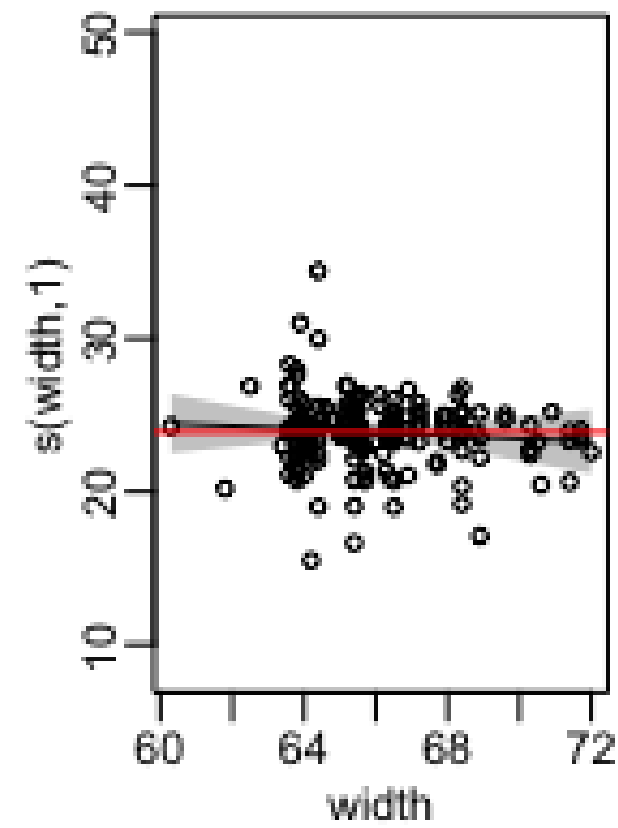
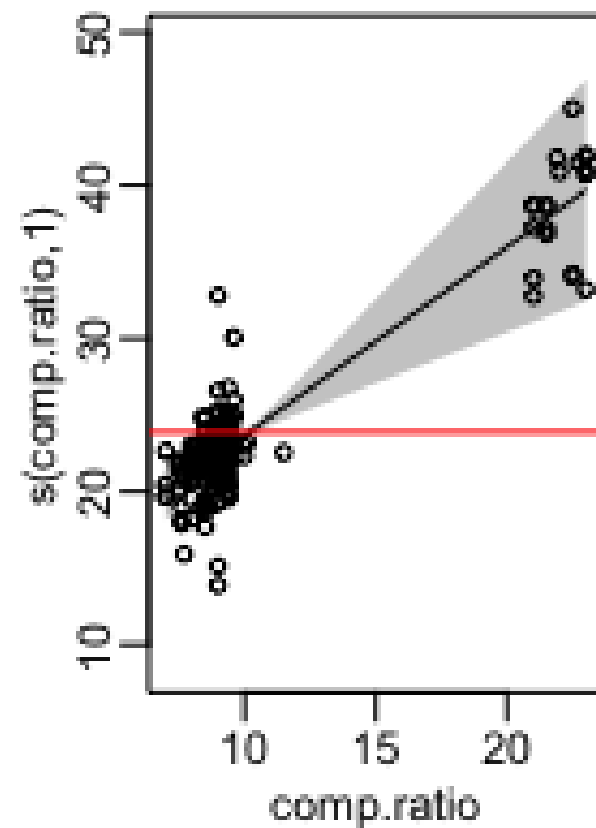
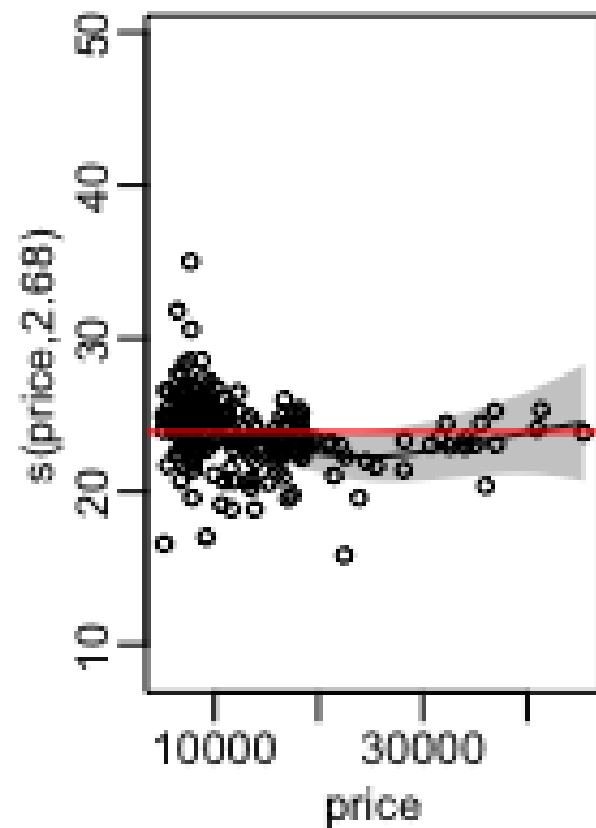
	edf	Ref.df	F	p-value		
s(weight)	6.254	7.439	20.909	< 2e-16	***	<--
s(rpm)	7.499	8.285	8.534	2.07e-09	***	
s(price)	2.681	3.421	1.678	0.155		<--
s(comp.ratio)	1.000	1.001	18.923	2.22e-05	***	
s(width)	1.001	1.001	0.357	0.551		



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	<--





NONLINEAR MODELING IN R WITH GAMs

Let's practice!



NONLINEAR MODELING IN R WITH GAMs

Visualizing GAMs

Noam Ross

Senior Research Scientist, EcoHealth Alliance



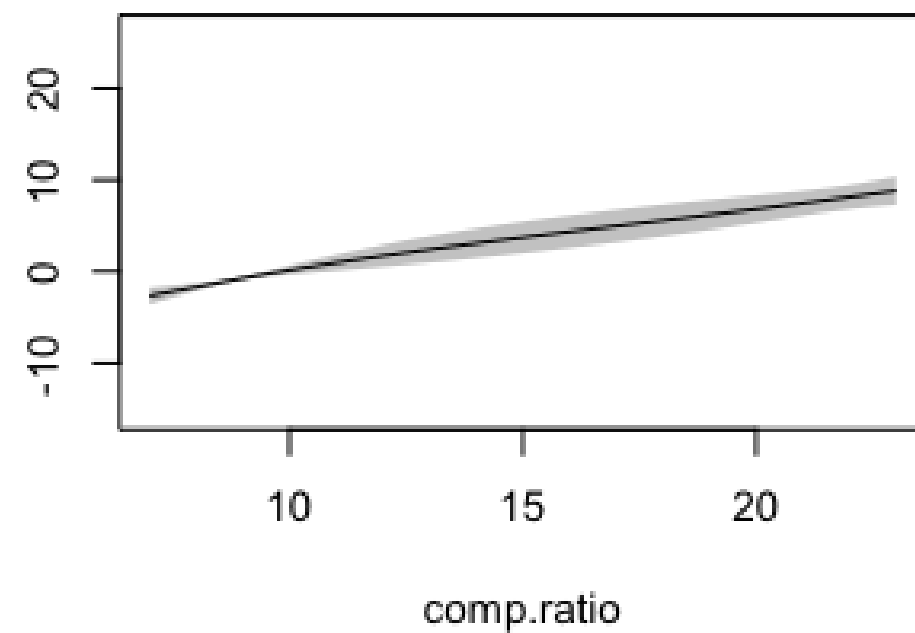
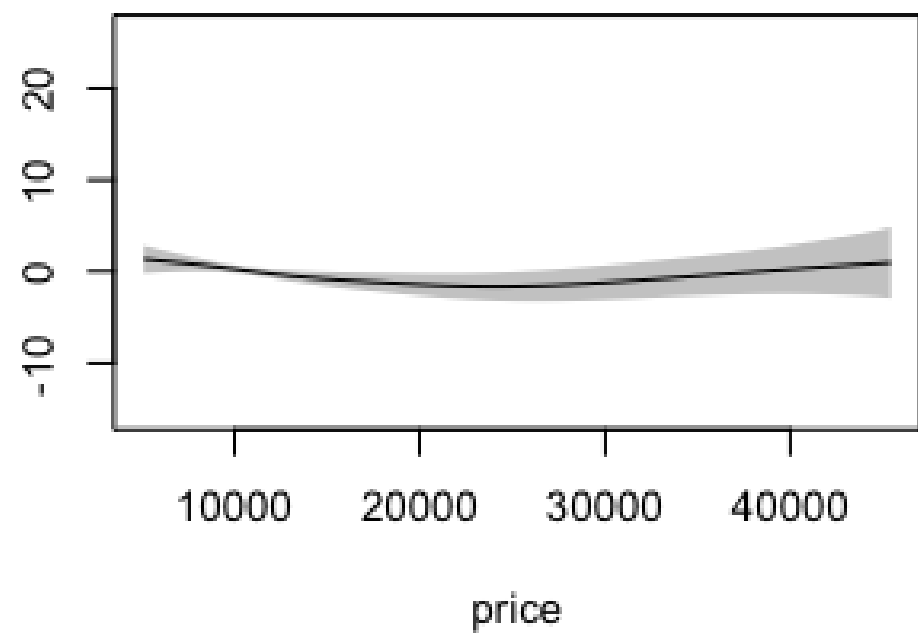
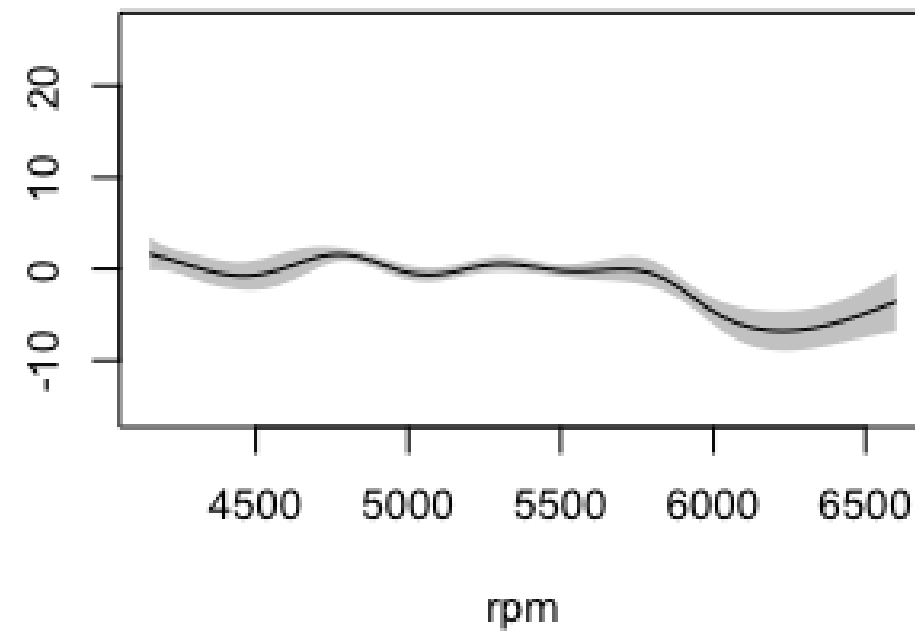
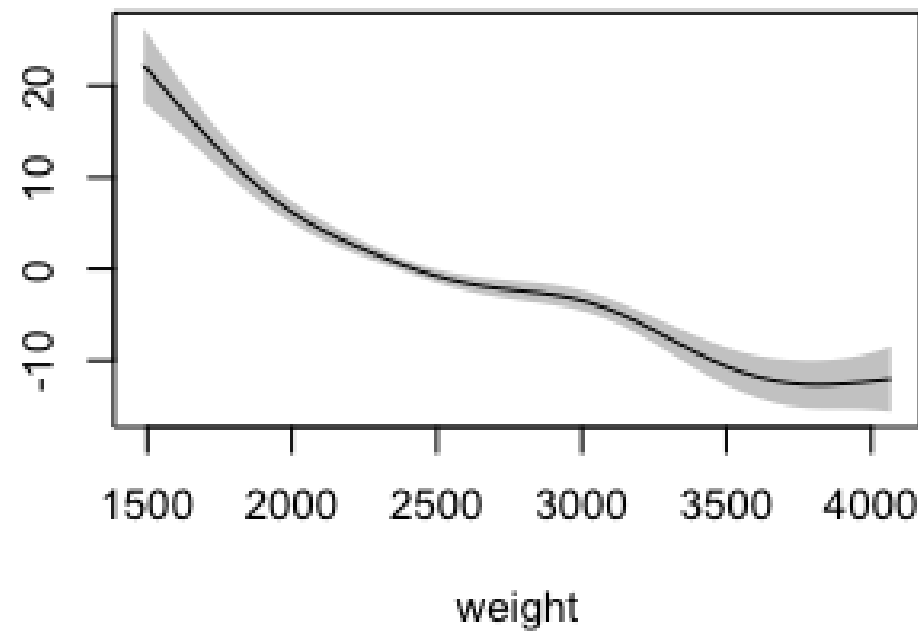
The Plot Command

```
plot(gam_model)
```

```
?plot.gam
```

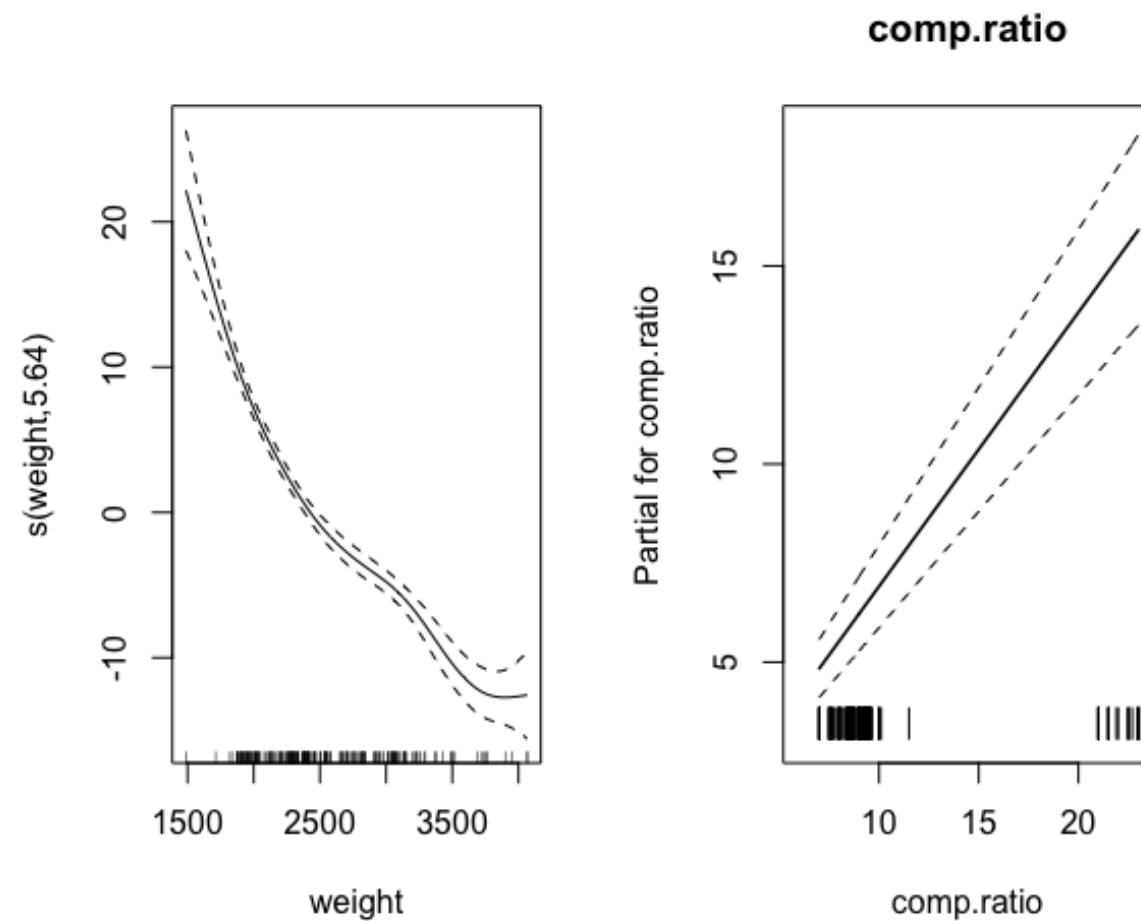


Partial Effect Plots



Selecting partial effects

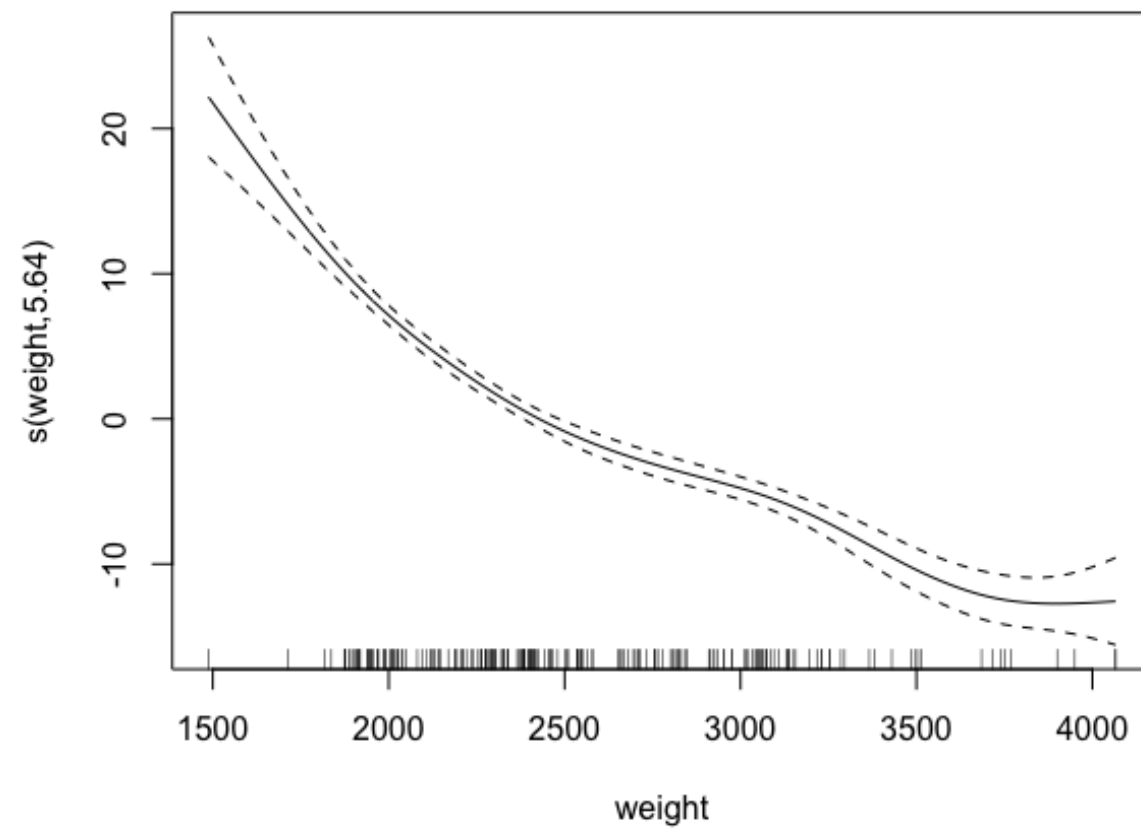
```
plot(gam_model, select = c(2, 3))  
  
plot(gam_model, pages = 1)  
  
plot(gam_model, pages = 1, all.terms = TRUE)
```





Showing data on the plots

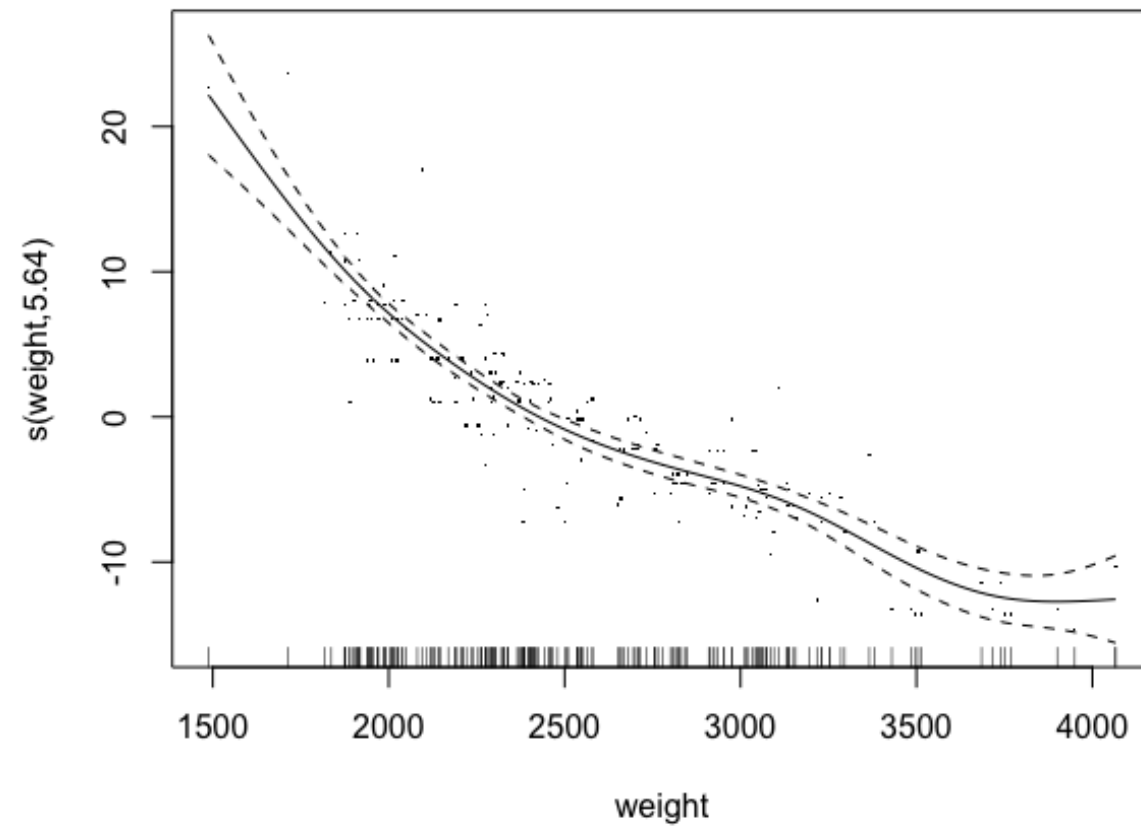
```
plot(gam_model, rug = TRUE)
```





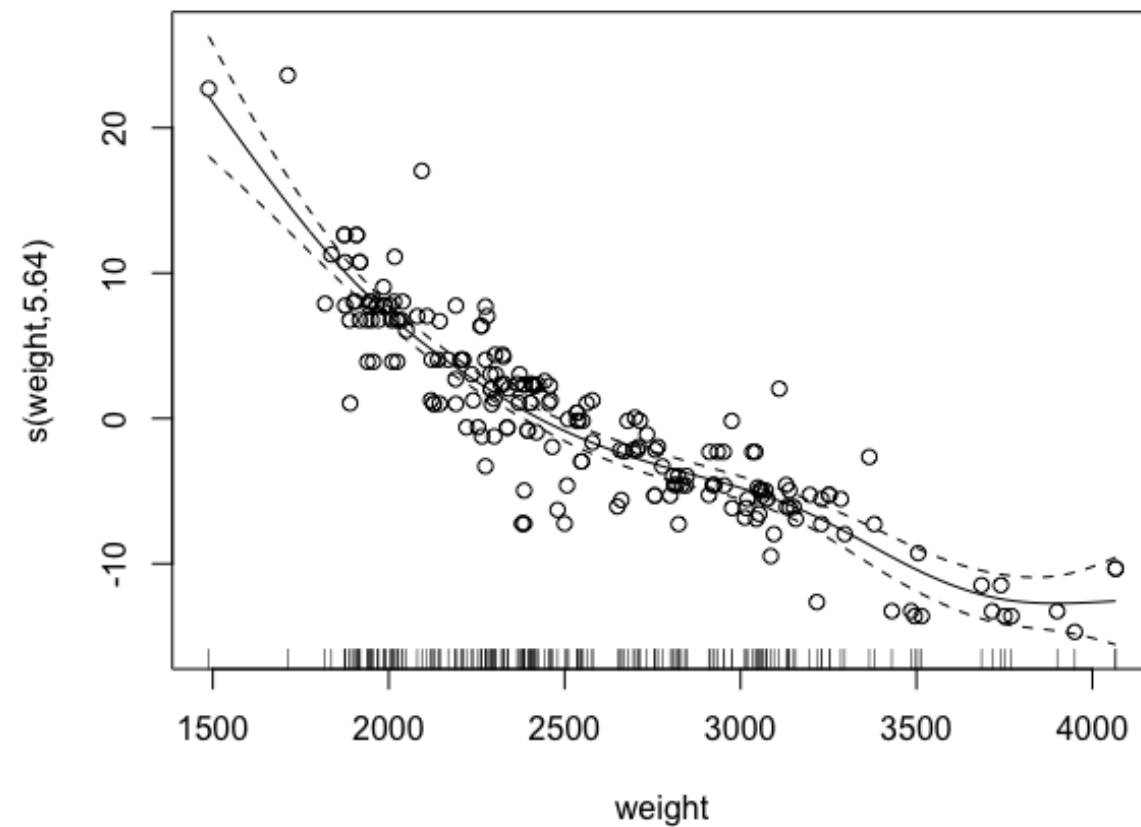
Showing data on the plots (2)

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



Showing data on the plots (3)

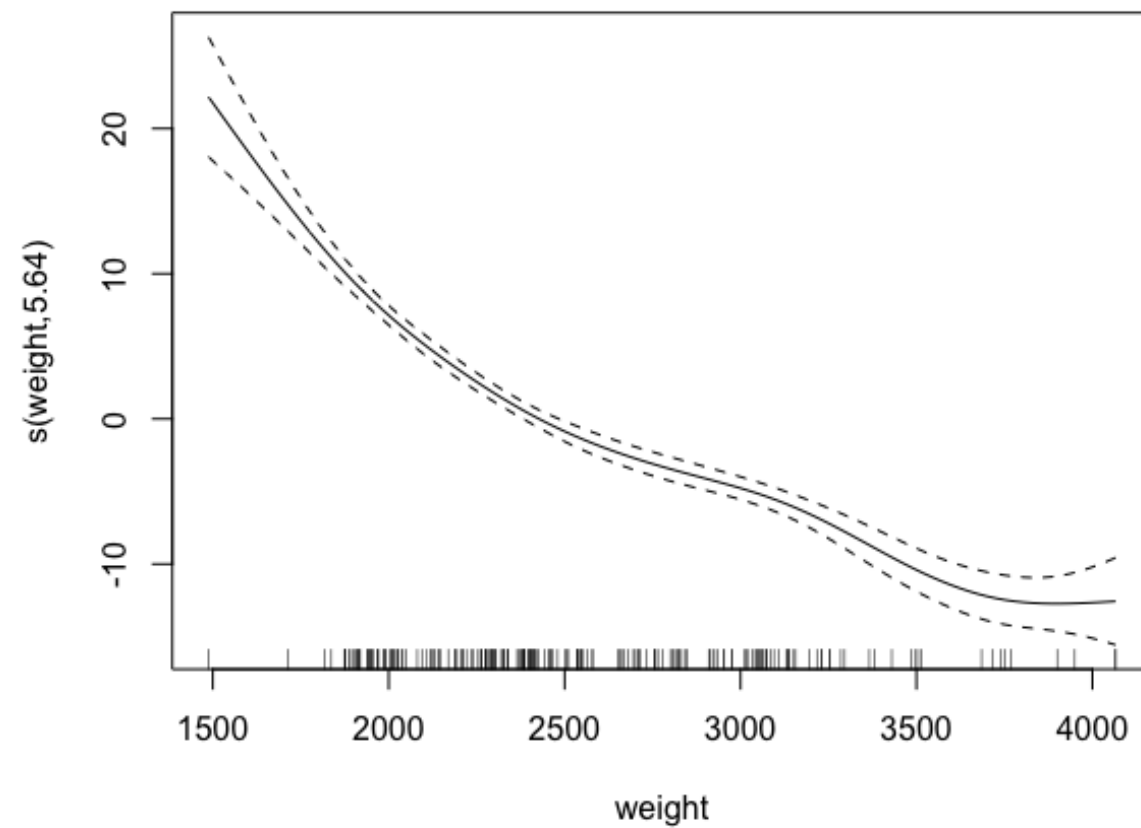
```
plot(gam_model, rug = TRUE, residuals = TRUE,  
     pch = 1, cex = 1)
```





Showing Standard Errors

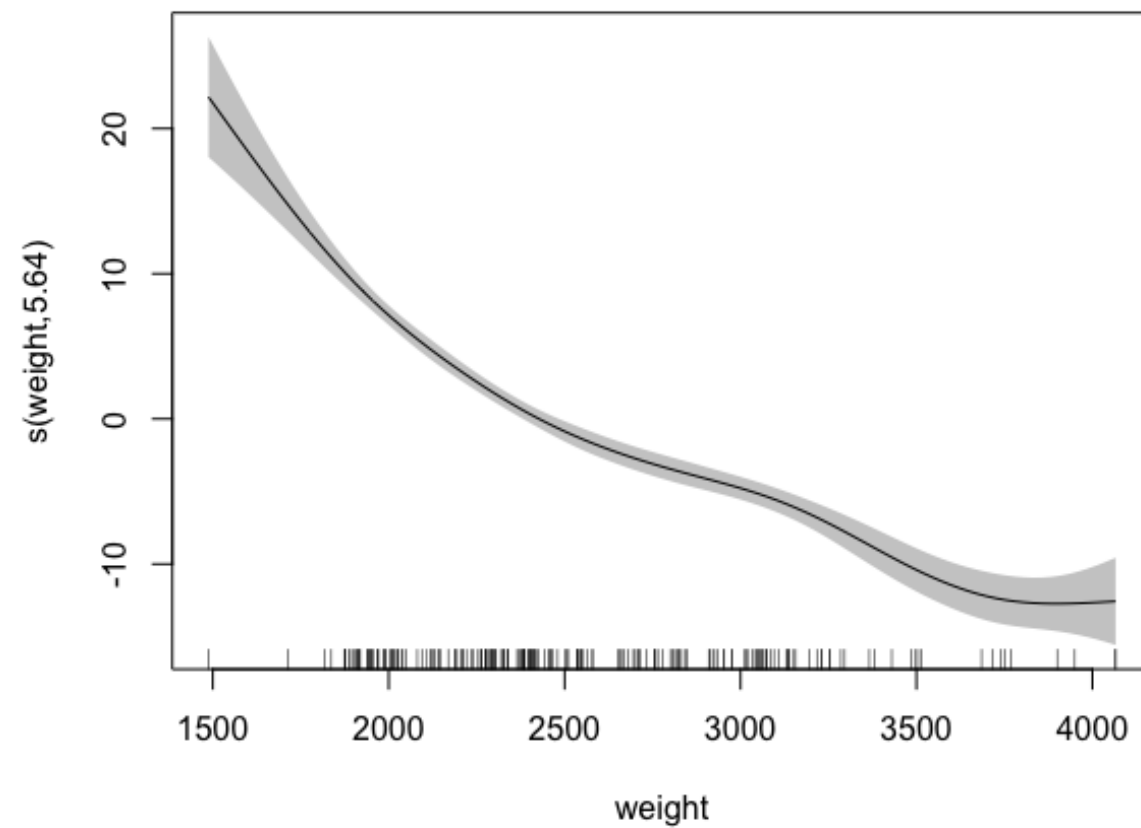
```
plot(gam_model, se = TRUE)
```





Showing Standard Errors (2)

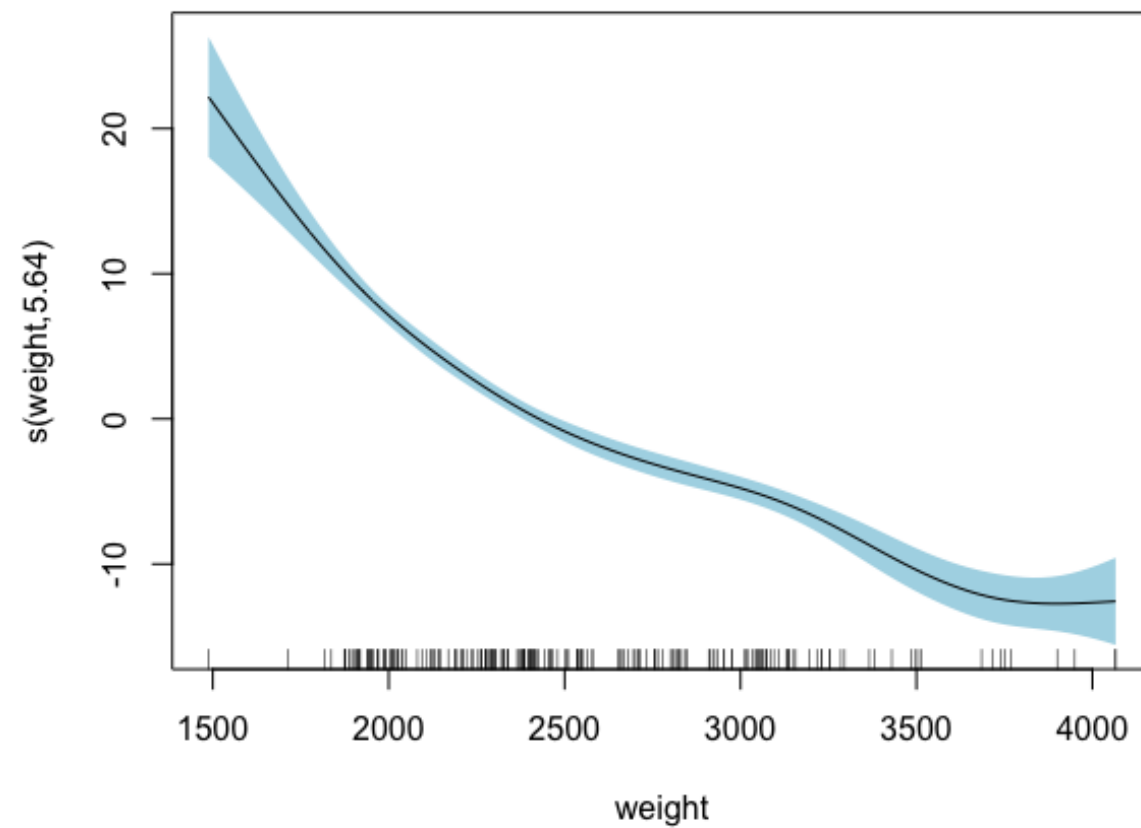
```
plot(gam_model, shade = TRUE)
```





Showing Standard Errors

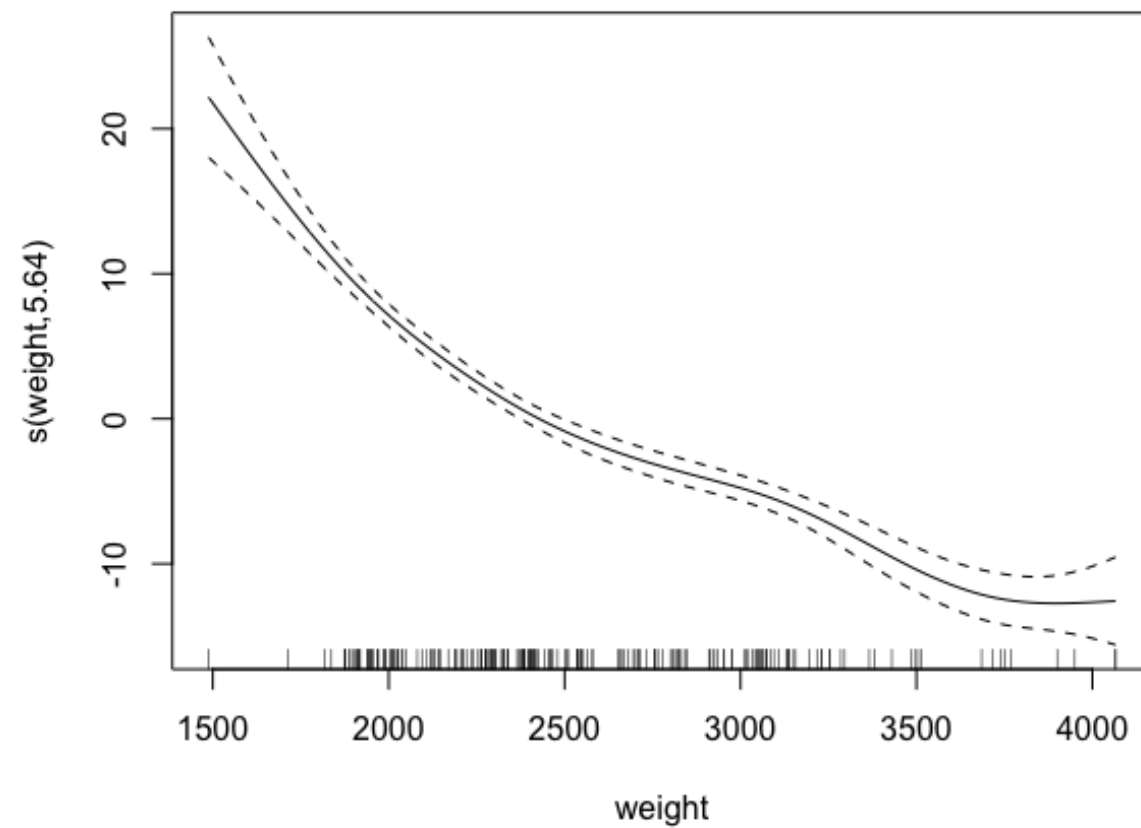
```
plot(gam_model, shade = TRUE, shade.col = "lightblue")
```





Transforming Standard Errors

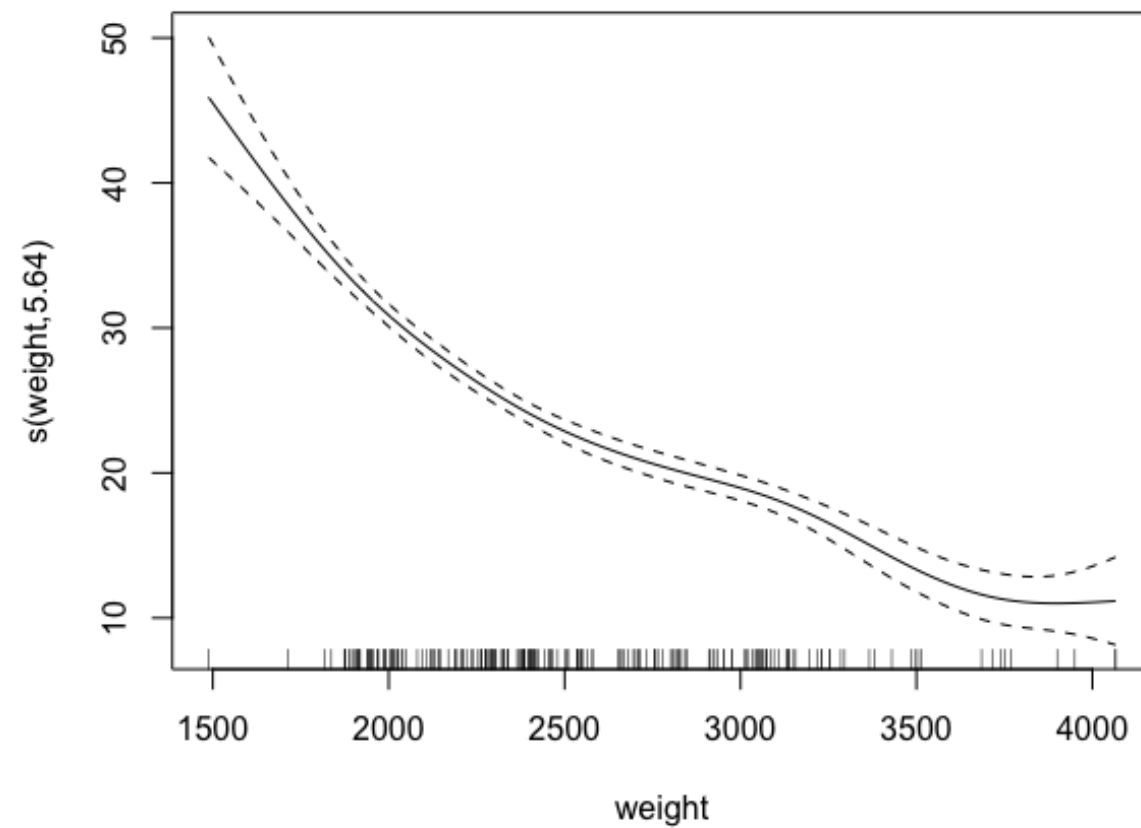
```
plot(gam_model, seWithMean = TRUE)
```





Transforming Standard Errors (2)

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





NONLINEAR MODELING IN R WITH GAMs

Now lets make some plots!



NONLINEAR MODELING IN R WITH GAMs

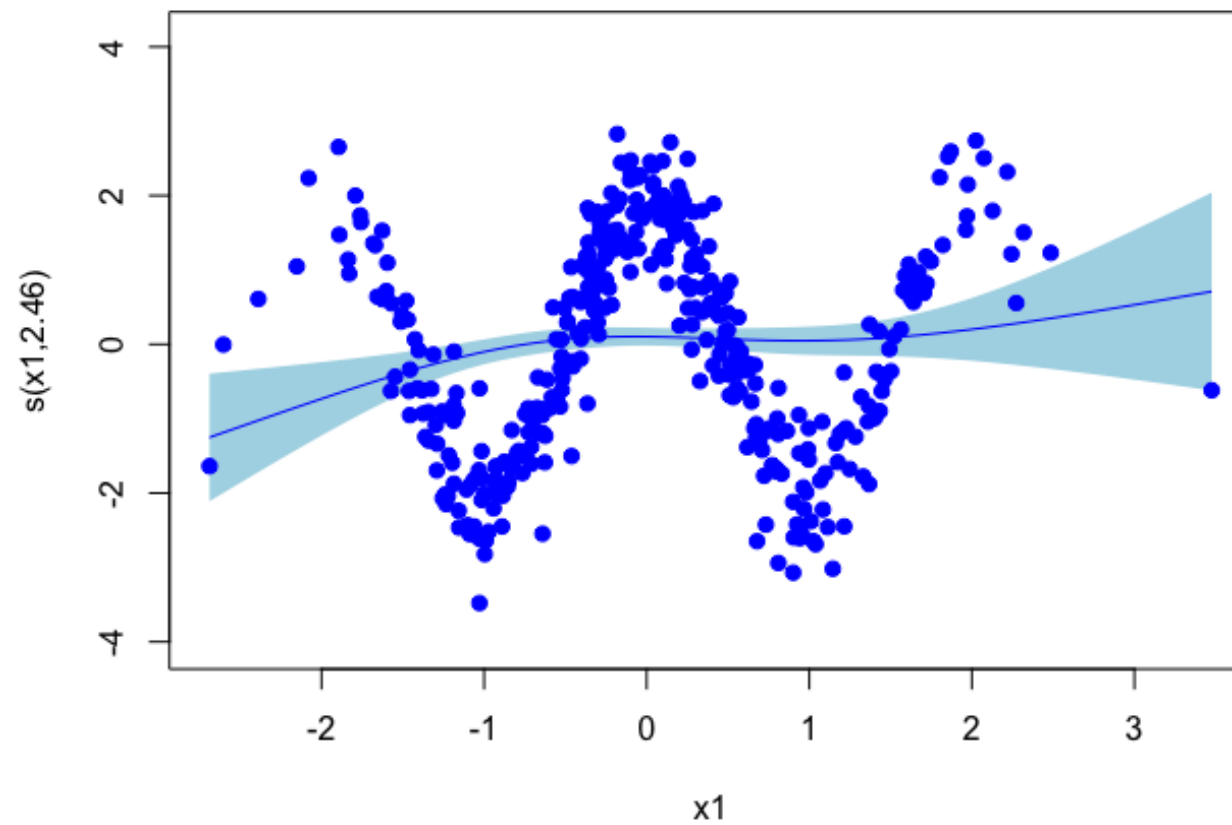
Model checking with `gam.check()`

Noam Ross

Senior Research Scientist, EcoHealth Alliance

Pitfall One: Inadequate Basis Number

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



Running gam.check

```
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 *

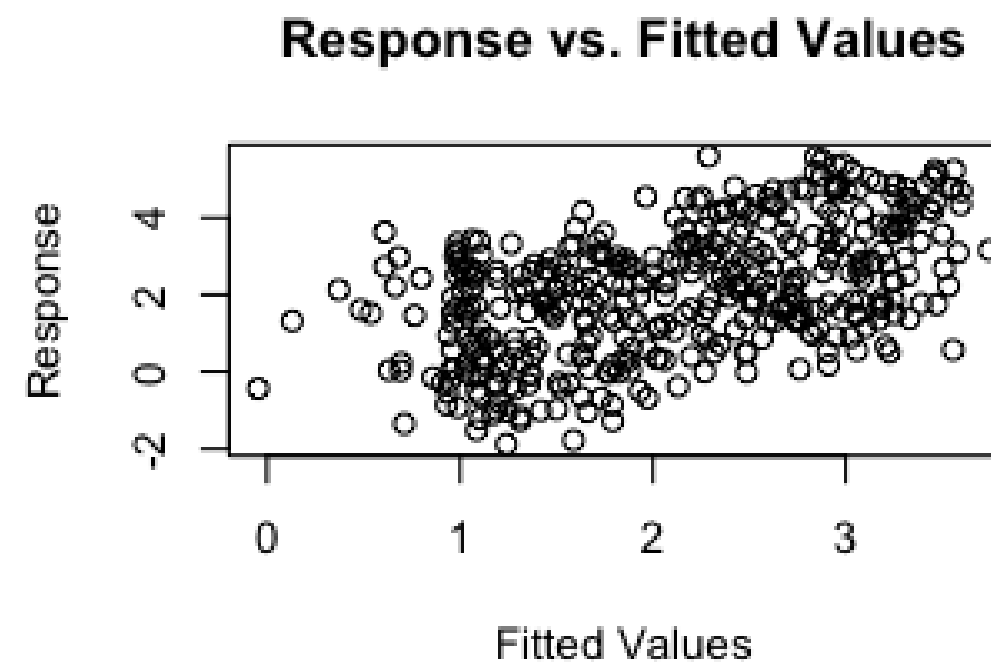
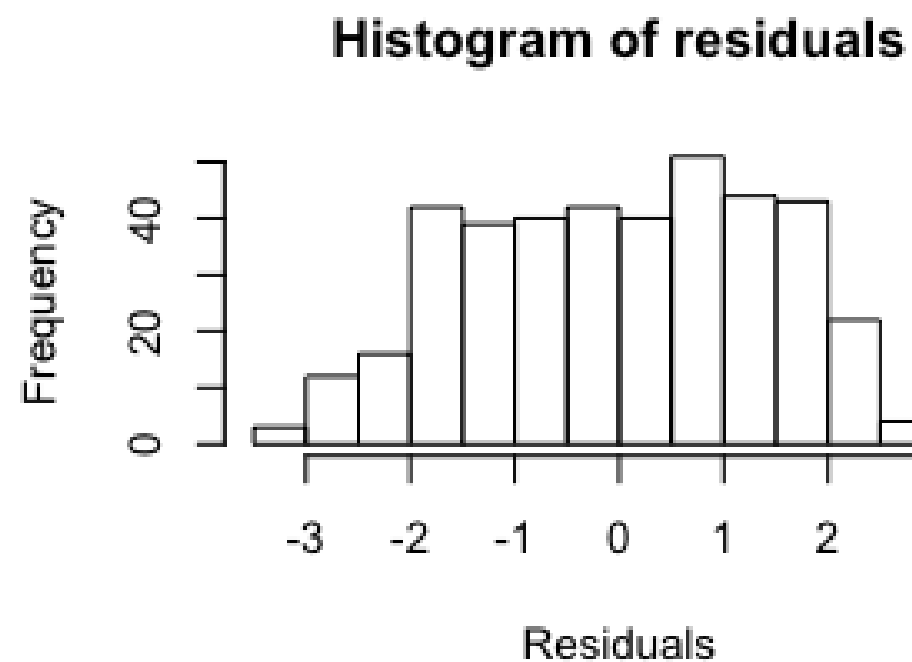
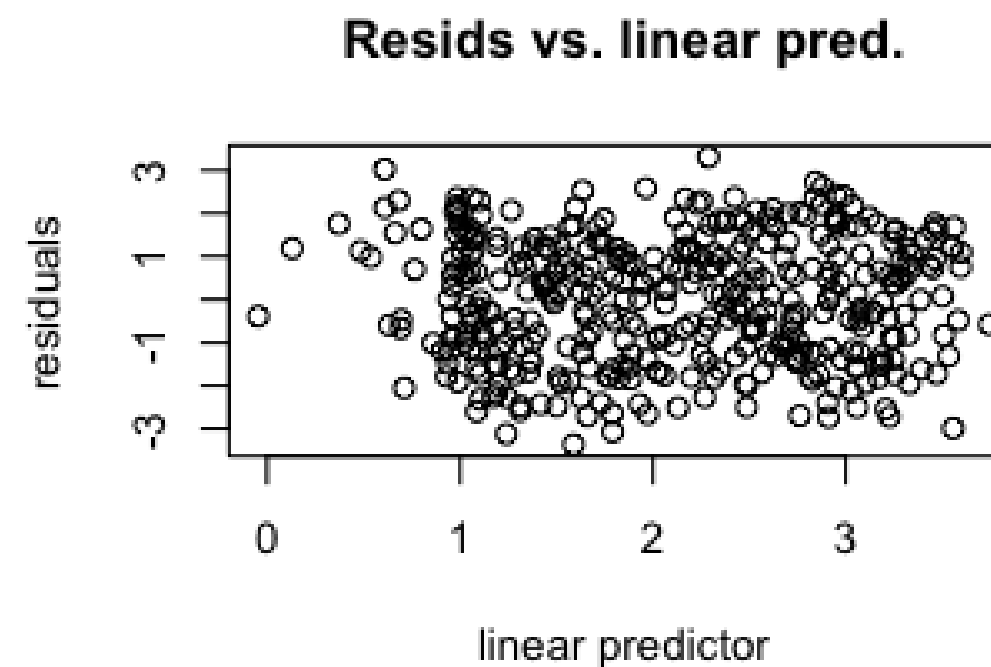
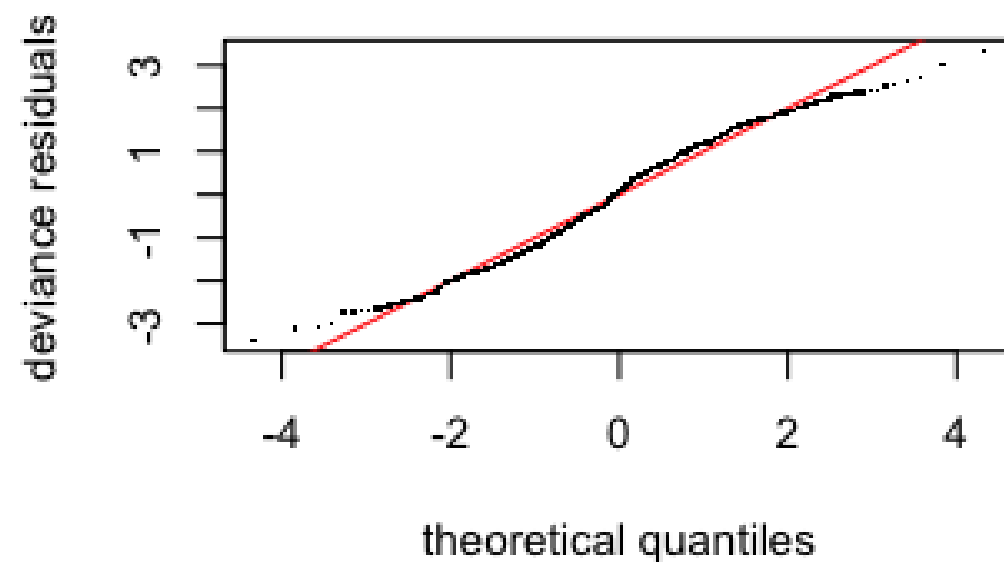
...

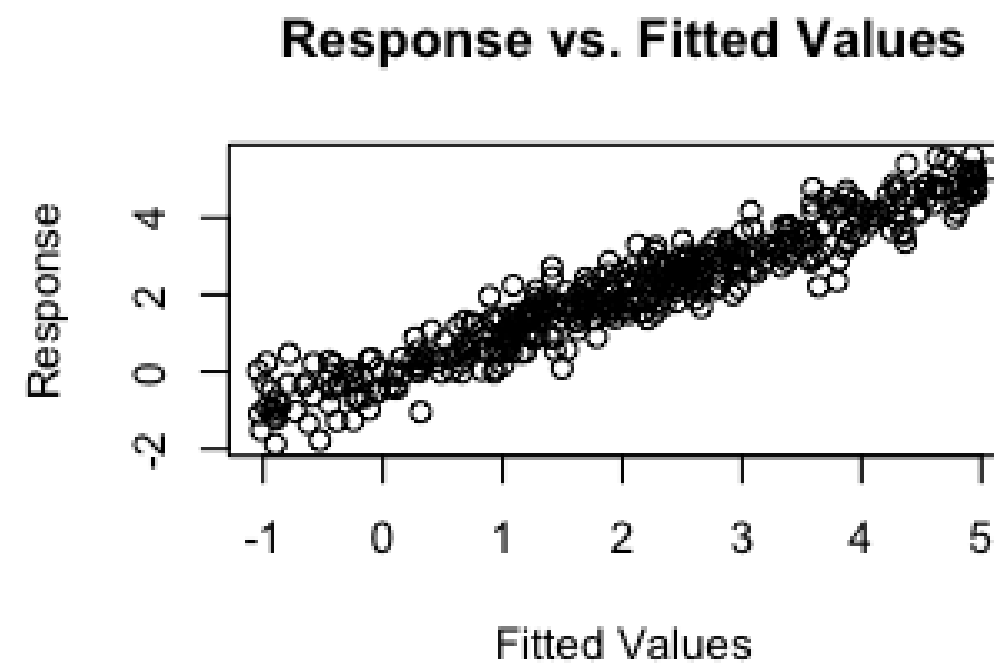
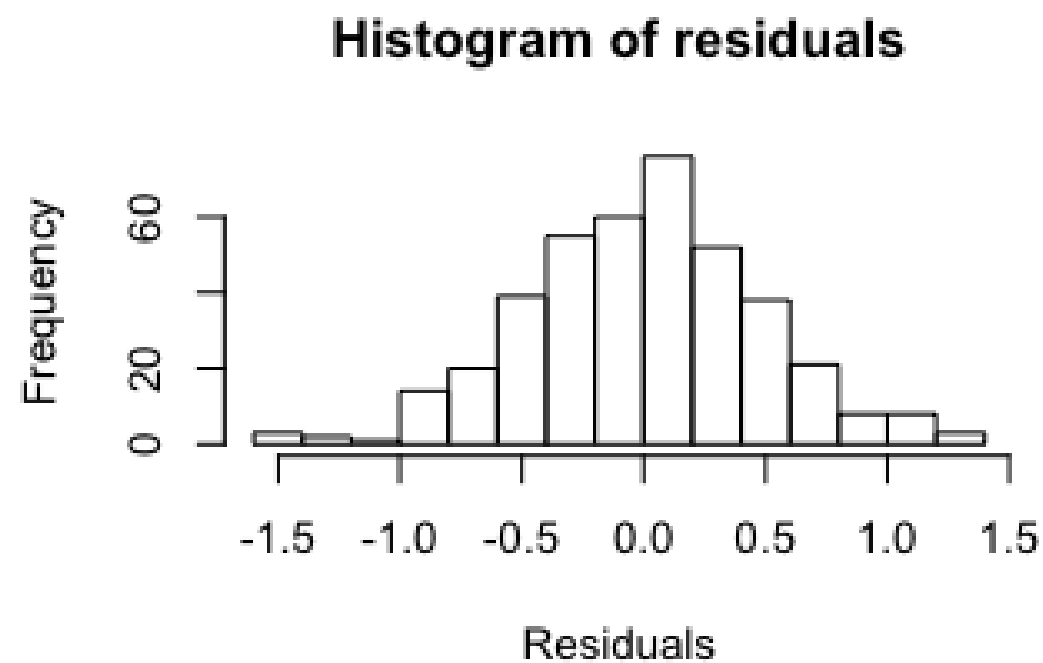
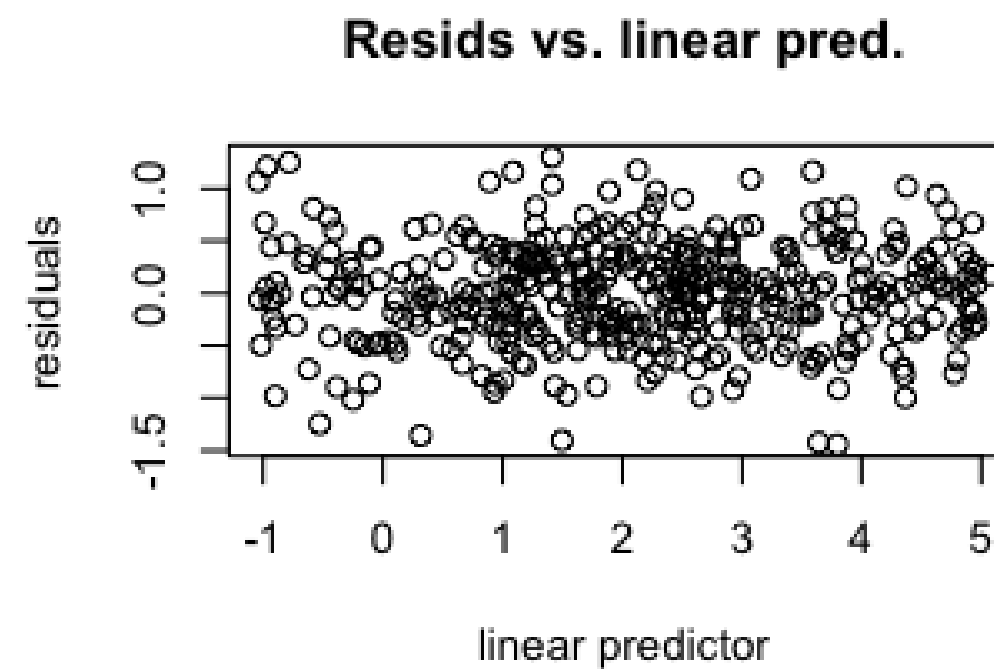
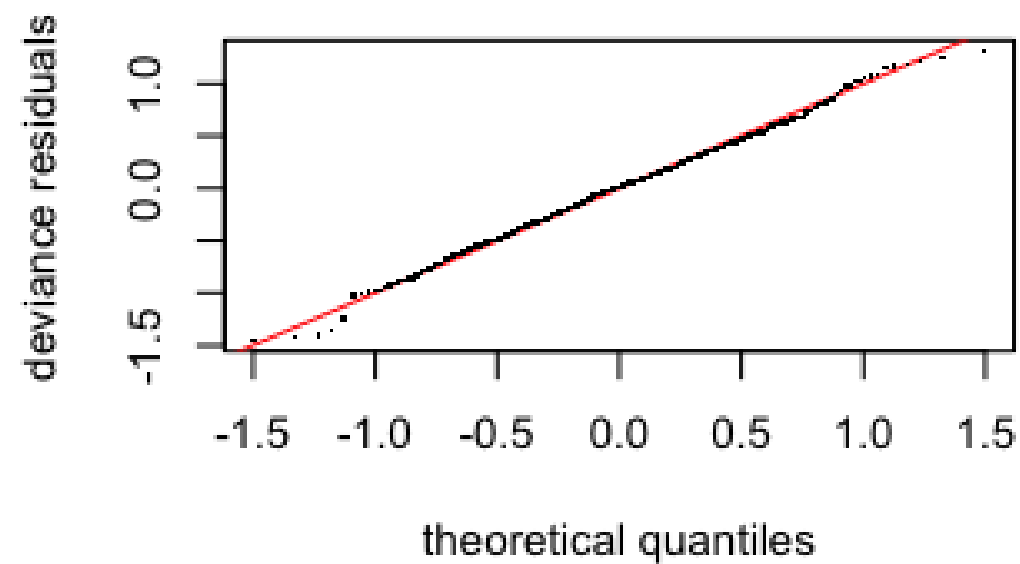


Running gam.check (3)

```
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  
...
```







NONLINEAR MODELING IN R WITH GAMs

Let's check some models

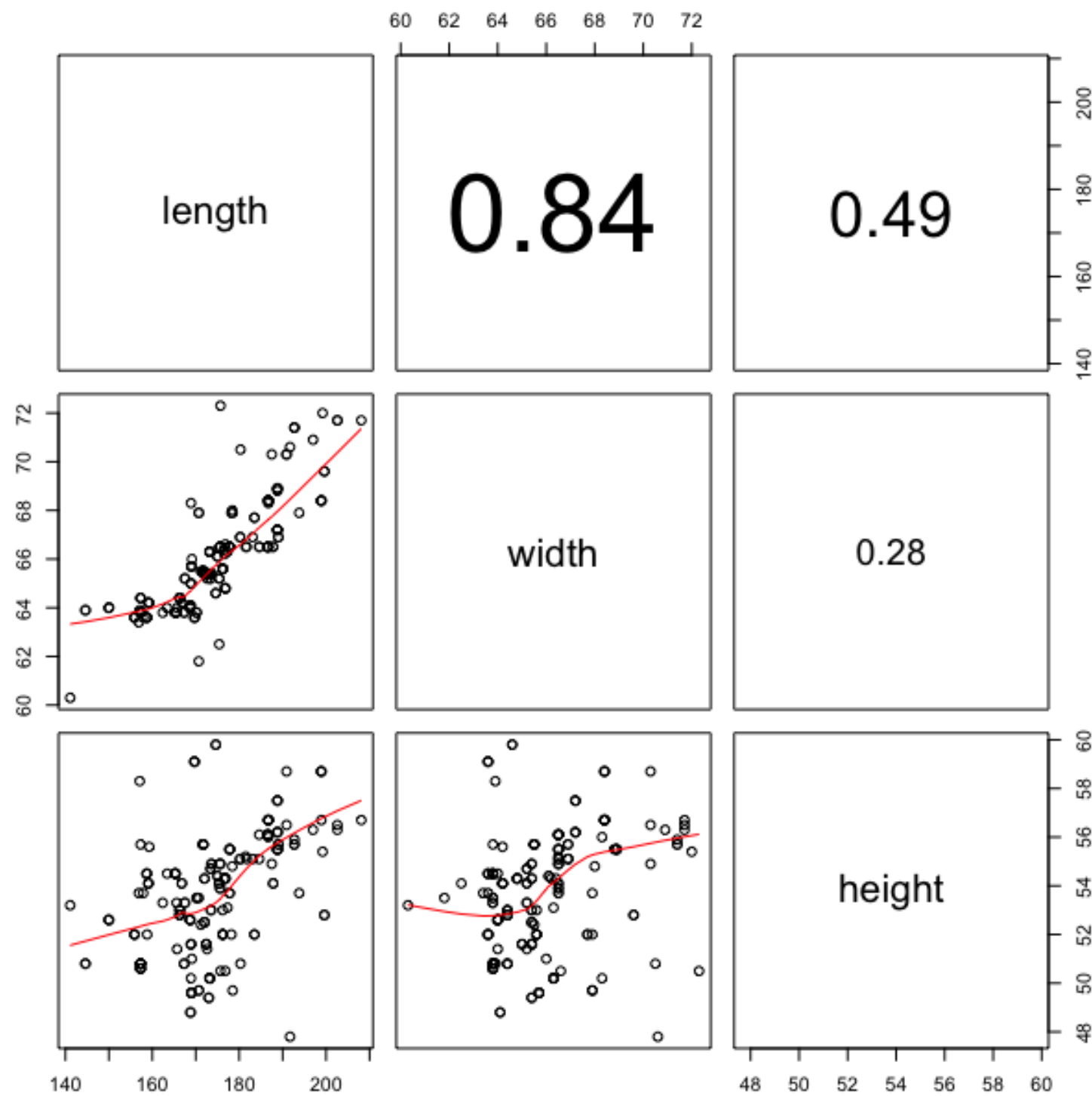


NONLINEAR MODELING IN R WITH GAMs

Checking concavity

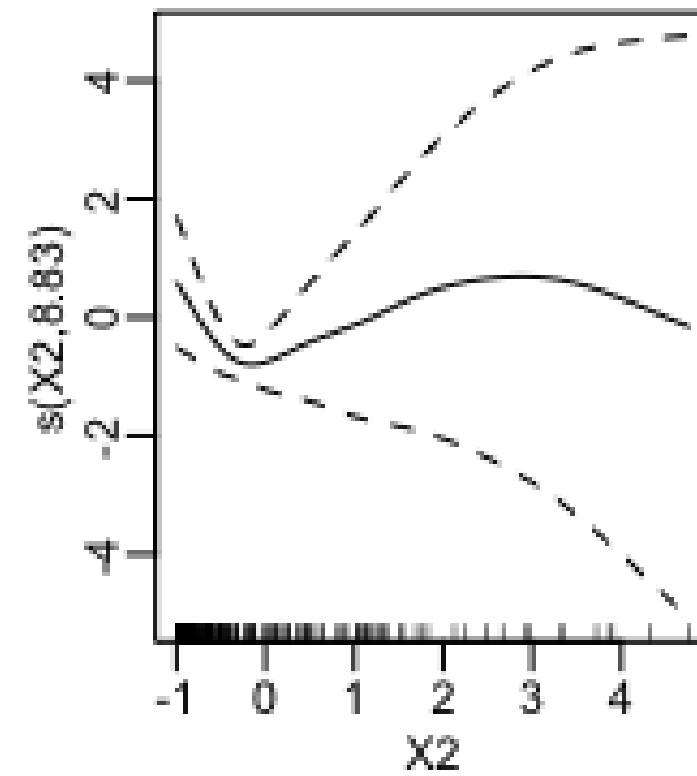
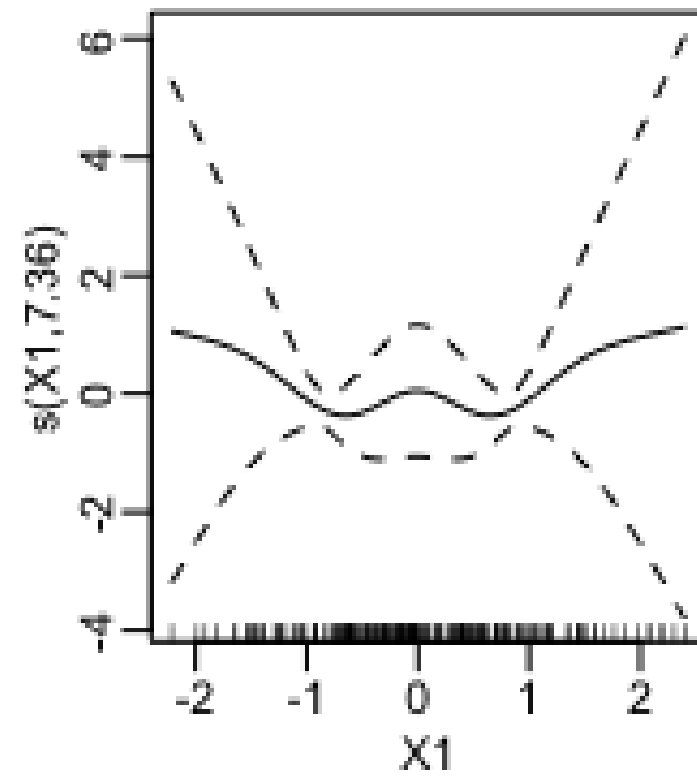
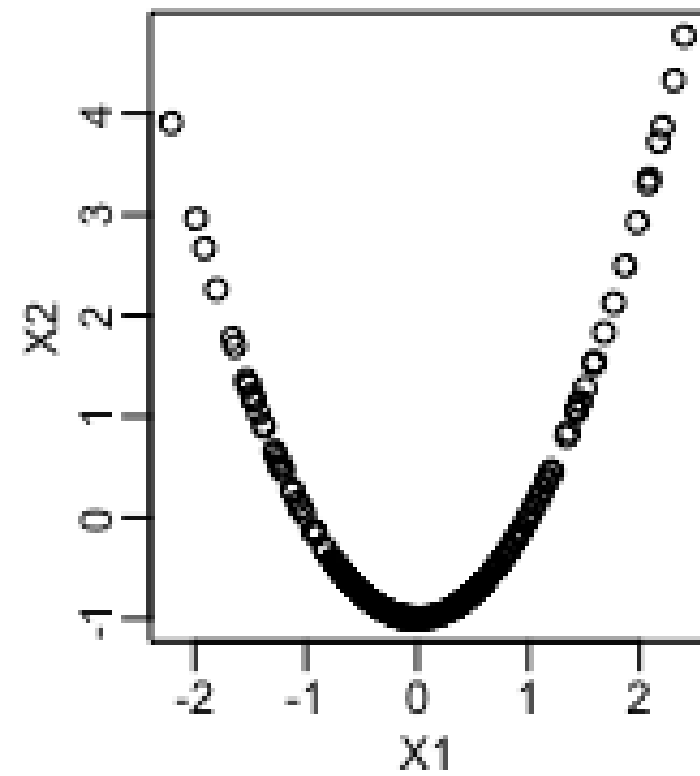
Noam Ross

Senior Research Scientist, EcoHealth Alliance

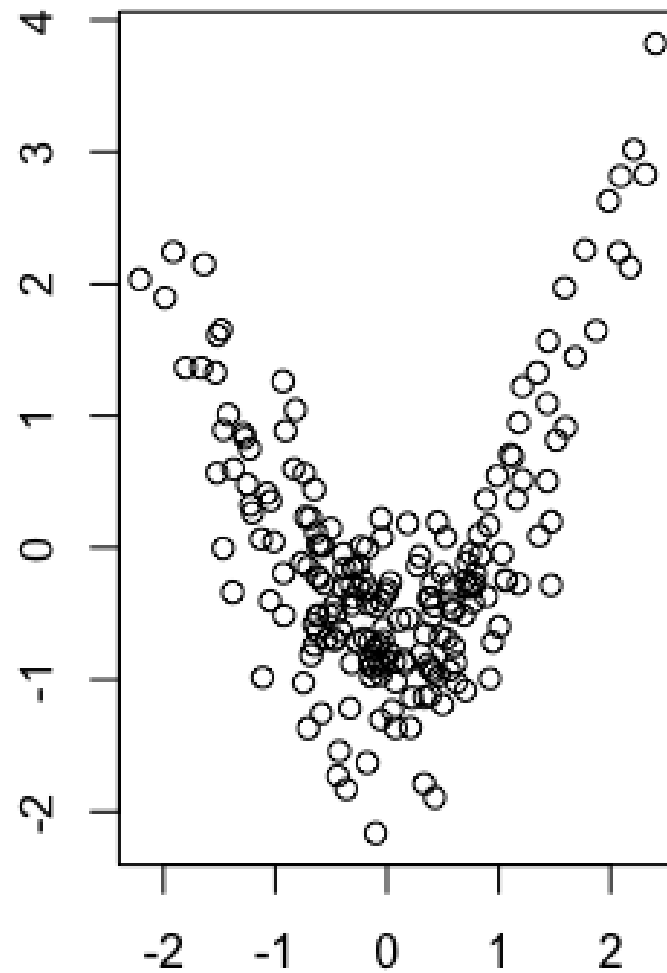




Concurvity



The concavity() 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
```

	para	s (X1)	s (X2)
para	1	0.00	0.00
s (X1)	0	1.00	0.57
s (X2)	0	0.22	1.00

```
$estimate
```

	para	s (X1)	s (X2)
para	1	0.00	0.0
s (X1)	0	1.00	0.6
s (X2)	0	0.28	1.0



NONLINEAR MODELING IN R WITH GAMs

Let's practice!