

Article title: Leptocheliidae Diversity: Indicators of Environmental Change in Threatened Habitats

Journal: Biodiversity and Conservation

Authors: Anna Stępień¹, Hanieh Saeedi^{2,3}, Magdalena Błazewicz¹

¹ University of Lodz, Faculty of Biology and Environmental Protection, Lodz, Poland

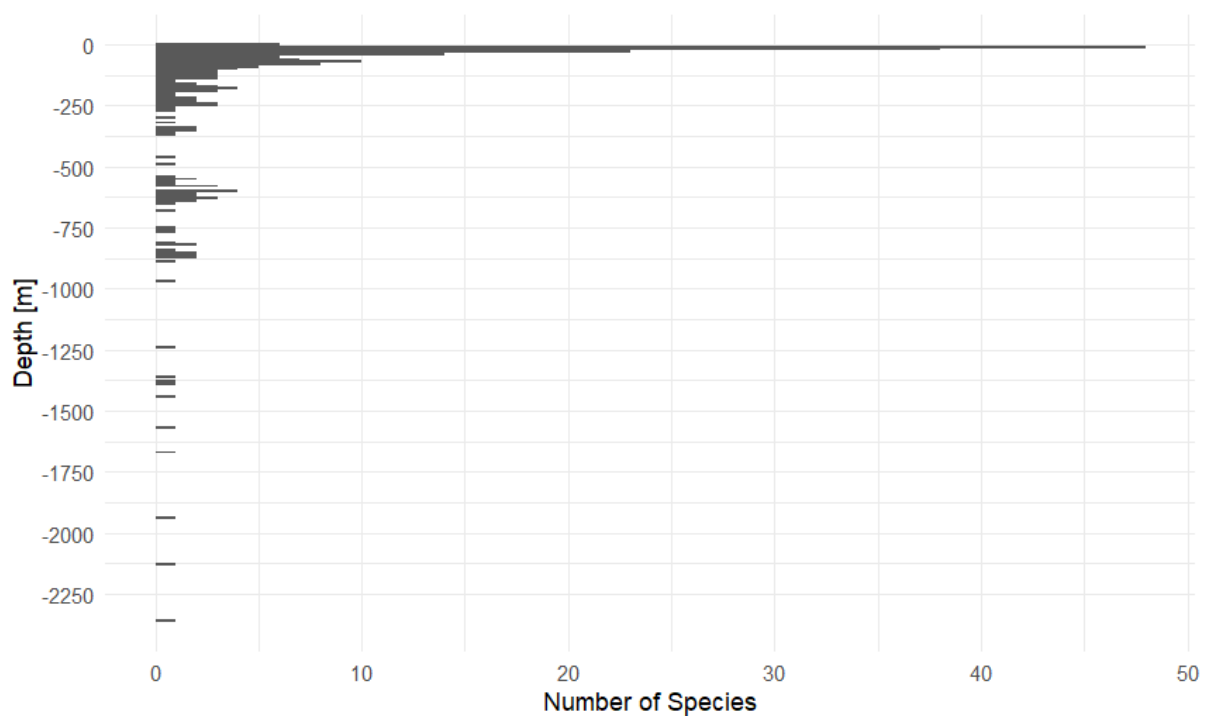
² Department of Marine Zoology, Senckenberg Research Institute and Natural History Museum, Frankfurt, Germany

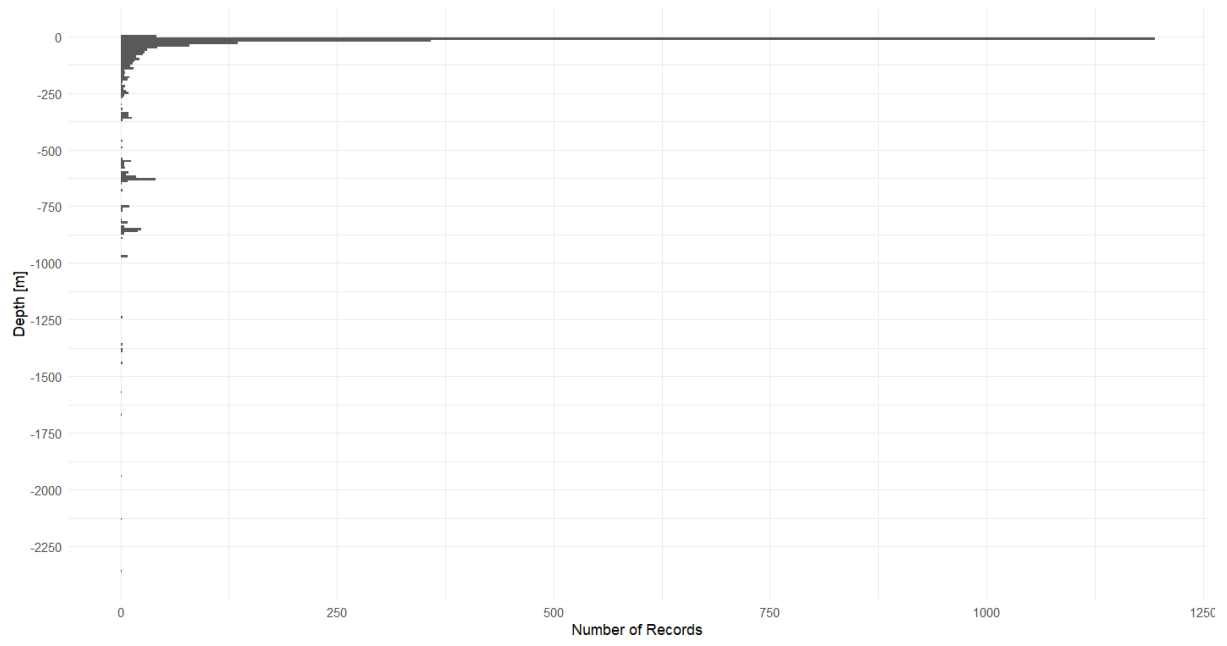
³ Institute for Ecology, Diversity and Evolution, Goethe University Frankfurt, Frankfurt, Germany

Corresponding author: anna.stepien@biol.uni.lodz.pl

Supplementary file 2. Calculation details of species richness and distribution of Leptochelidae members: A. number of species plotted against depth; B. detailed result of Generalized Linear Models; detailed result of Generalized Additive Models

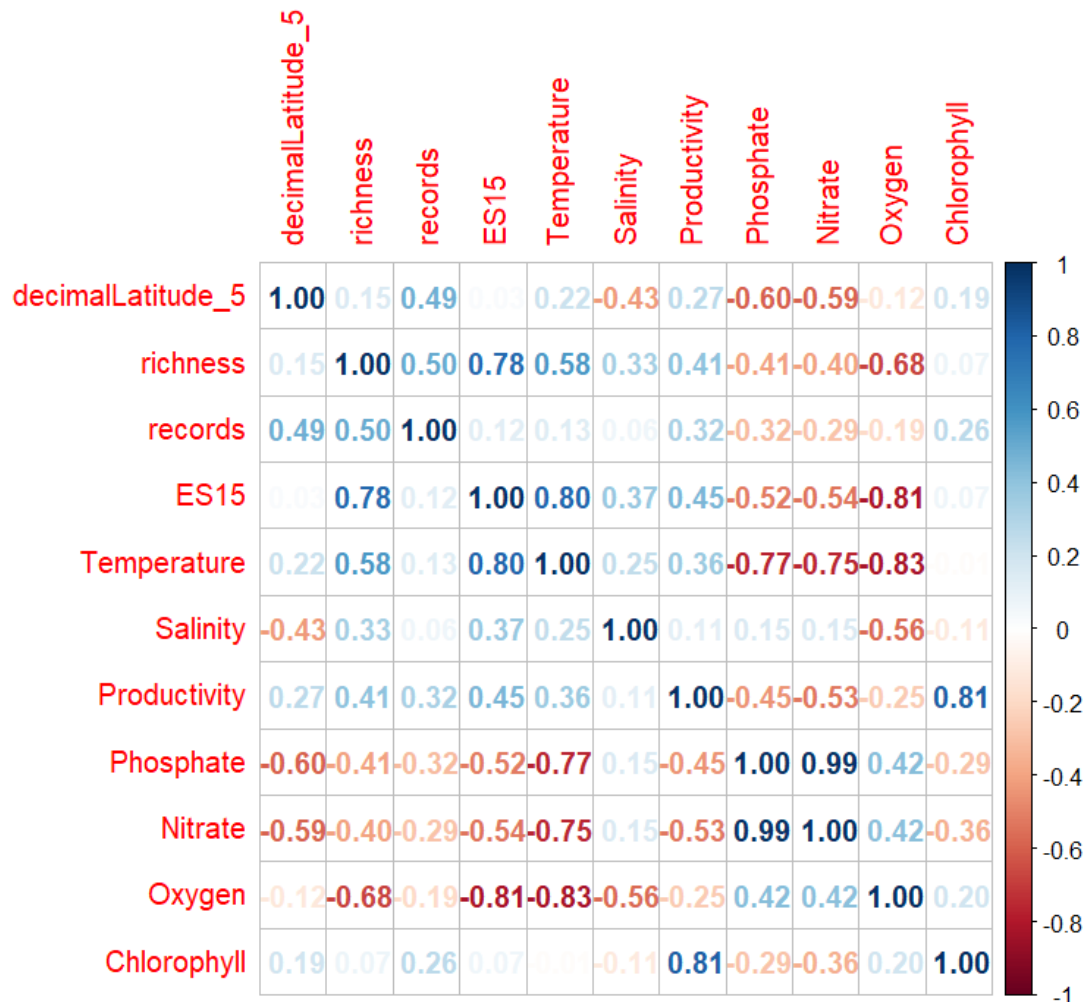
A.





B

GLM FOR RICHNESS PER LATITUDE



```
> GLM.nums.intercept <- glm(richness ~ 1, family = "poisson", data = data)
> summary(GLM.nums.intercept)
```

```
call:
glm(formula = richness ~ 1, family = "poisson", data = data)
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.27461    0.05955   38.2    <2e-16 ***
```

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

```
(Dispersion parameter for poisson family taken to be 1)
```

```
Null deviance: 200.99  on 28  degrees of freedom
Residual deviance: 200.99  on 28  degrees of freedom
AIC: 308.93
```

```
Number of Fisher Scoring iterations: 5
```

```
> GLM.nums.records <- glm(richness ~ records, family = "poisson", data = data)
> summary(GLM.nums.records)
```

```
Call:
glm(formula = richness ~ records, family = "poisson", data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.9191022	0.0866946	22.136	< 2e-16 ***
records	0.0021369	0.0003117	6.856	7.1e-12 ***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 200.99  on 28  degrees of freedom
Residual deviance: 157.88  on 27  degrees of freedom
AIC: 267.82
```

Number of Fisher Scoring iterations: 5

```
> GLM.nums.latitude <- glm(richness ~ records + decimalLatitude_5, family =
"poisson", data = data)
> summary(GLM.nums.latitude)
```

```
Call:
glm(formula = richness ~ records + decimalLatitude_5, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.8653351	0.0951142	19.612	< 2e-16 ***
records	0.0024978	0.0003945	6.332	2.42e-10 ***
decimalLatitude_5	-0.0028285	0.0018361	-1.541	0.123

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 200.99  on 28  degrees of freedom
Residual deviance: 155.45  on 26  degrees of freedom
AIC: 267.39
```

Number of Fisher Scoring iterations: 5

```
> GLM.nums.temp <- glm(richness ~ records + Temperature, family = "poisson"
, data = data)
> summary(GLM.nums.temp)
```

```
Call:
glm(formula = richness ~ records + Temperature, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.5696042	0.2207831	2.580	0.00988 **
records	0.0025058	0.0003421	7.324	2.40e-13 ***
Temperature	0.0751937	0.0095177	7.900	2.78e-15 ***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 200.992  on 28  degrees of freedom
Residual deviance:  79.803  on 26  degrees of freedom
AIC: 191.74
```

Number of Fisher Scoring iterations: 5

```
> GLM.nums.salinity <- glm(richness ~ records + Salinity, family = "poisson",
", data = data)
> summary(GLM.nums.salinity)
```

```
Call:
glm(formula = richness ~ records + Salinity, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.6485982	1.2250200	-2.162	0.030612	*
records	0.0019615	0.0003082	6.364	1.96e-10	***
Salinity	0.1343371	0.0353501	3.800	0.000145	***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 200.99 on 28 degrees of freedom
Residual deviance: 126.82 on 26 degrees of freedom
AIC: 238.76
```

Number of Fisher Scoring iterations: 5

```
> GLM.nums.productivity <- glm(richness ~ records + Productivity, family =
"poisson", data = data)
> summary(GLM.nums.productivity)
```

```
Call:
glm(formula = richness ~ records + Productivity, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.735e+00	1.039e-01	16.693	< 2e-16	***
records	1.741e-03	3.332e-04	5.224	1.75e-07	***
Productivity	3.449e+01	9.916e+00	3.478	0.000504	***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 200.99 on 28 degrees of freedom
Residual deviance: 146.44 on 26 degrees of freedom
AIC: 258.38
```

Number of Fisher Scoring iterations: 5

```
> GLM.nums.phosphate <- glm(richness ~ records + Phosphate, family = "poisson",
", data = data)
> summary(GLM.nums.phosphate)
```

```
Call:
glm(formula = richness ~ records + Phosphate, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.3665539	0.1244784	19.012	< 2e-16	***
records	0.0017252	0.0003244	5.318	1.05e-07	***
Phosphate	-0.5753545	0.1332812	-4.317	1.58e-05	***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 200.99 on 28 degrees of freedom
Residual deviance: 136.75 on 26 degrees of freedom
AIC: 248.69
```

Number of Fisher Scoring iterations: 5

```
> GLM.nums.nitrate <- glm(richness ~ records + Nitrate, family = "poisson",  
data = data)  
> summary(GLM.nums.nitrate)
```

```
Call:  
glm(formula = richness ~ records + Nitrate, family = "poisson",  
data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.254019	0.106613	21.142	< 2e-16 ***
records	0.001805	0.000324	5.569	2.56e-08 ***
Nitrate	-0.038677	0.008714	-4.439	9.06e-06 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 200.99 on 28 degrees of freedom
Residual deviance: 134.90 on 26 degrees of freedom
AIC: 246.85

Number of Fisher Scoring iterations: 5

```
> GLM.nums.oxygen <- glm(richness ~ records + Oxygen, family = "poisson", d  
ata = data)  
> summary(GLM.nums.oxygen)
```

```
Call:  
glm(formula = richness ~ records + Oxygen, family = "poisson",  
data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.9747877	0.3456714	14.392	< 2e-16 ***
records	0.0018657	0.0003251	5.738	9.56e-09 ***
Oxygen	-0.0132792	0.0015622	-8.500	< 2e-16 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 200.992 on 28 degrees of freedom
Residual deviance: 69.477 on 26 degrees of freedom
AIC: 181.42

Number of Fisher Scoring iterations: 5

```
> GLM.nums.chlorophyll <- glm(richness ~ records + Chlorophyll, family = "p  
oisson", data = data)  
> summary(GLM.nums.chlorophyll)
```

```
Call:  
glm(formula = richness ~ records + Chlorophyll, family = "poisson",  
data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.9671612	0.1067982	18.419	< 2e-16 ***
records	0.0022060	0.0003245	6.798	1.06e-11 ***
chlorophyll	-0.1909153	0.2526953	-0.756	0.45

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

(Dispersion parameter for poisson family taken to be 1)

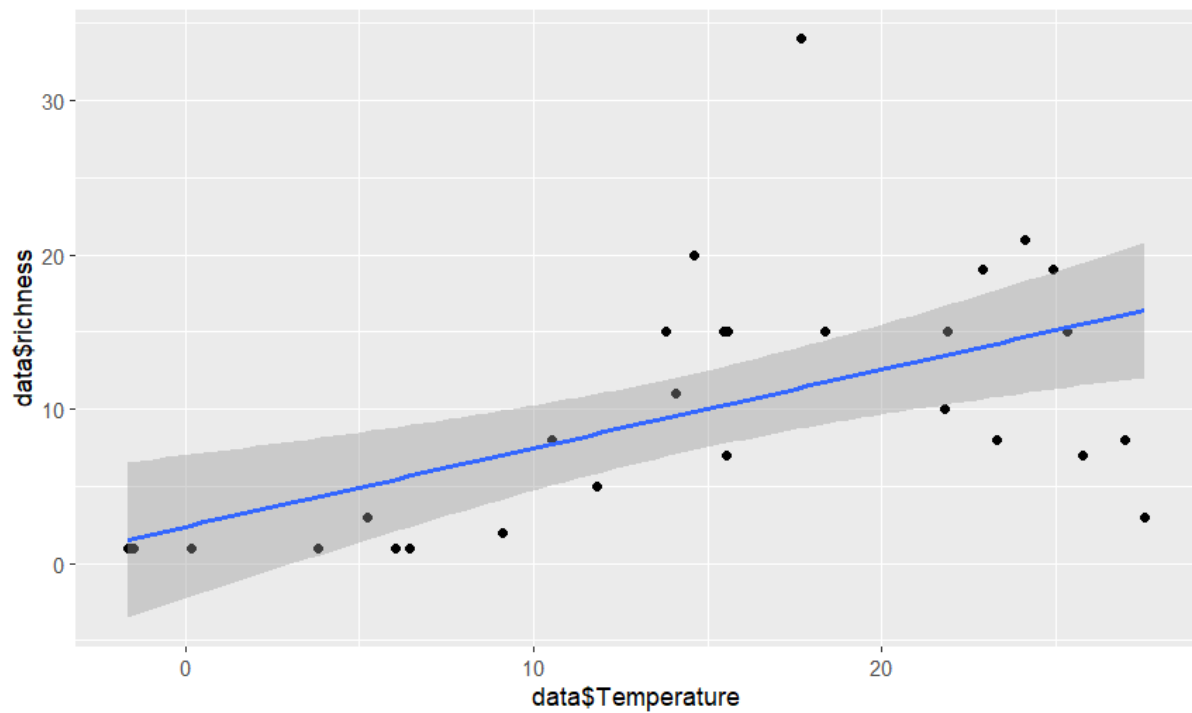
Null deviance: 200.99 on 28 degrees of freedom

Residual deviance: 157.30 on 26 degrees of freedom
AIC: 269.24

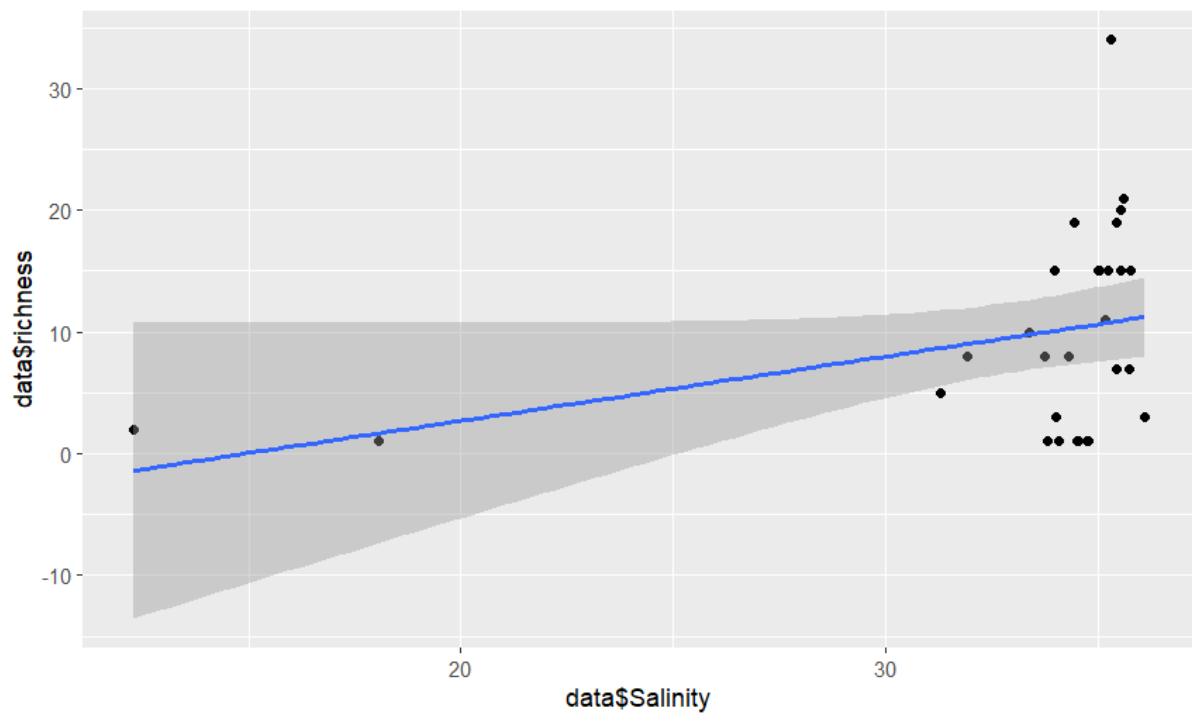
Number of Fisher Scoring iterations: 5

```
> GLM.nums.models <- list(GLM.nums.intercept = GLM.nums.intercept,  
  GLM.nums.records = GLM.nums.records,  
  GLM.nums.latitude = GLM.nums.latitude,  
  GLM.nums.temp = GLM.nums.temp,  
  GLM.nums.salinity = GLM.nums.salinity,  
  GLM.nums.productivity = GLM.nums.productivity,  
  GLM.nums.phosphate = GLM.nums.phosphate,  
  GLM.nums.nitrate = GLM.nums.nitrate,  
  GLM.nums.oxygen = GLM.nums.oxygen,  
  GLM.nums.chlorophyll = GLM.nums.chlorophyll)  
> GLM.nums.aic.df <- data.frame(Model = names(GLM.nums.models),  
  AIC = sapply(GLM.nums.models, function(x) AICc(x)),  
  akaike.weights(sapply(GLM.nums.models, function(x) AICc(x))))  
  
> GLM.nums.aic.df <- GLM.nums.aic.df[order(GLM.nums.aic.df$AIC),]  
> GLM.nums.aic.df$Cumulative.Weight <- cumsum(GLM.nums.aic.df$weights)  
  
> qplot(data$Temperature, data$richness) + geom_smooth(method = "glm")  
`geom_smooth()` using formula = 'y ~ x'
```

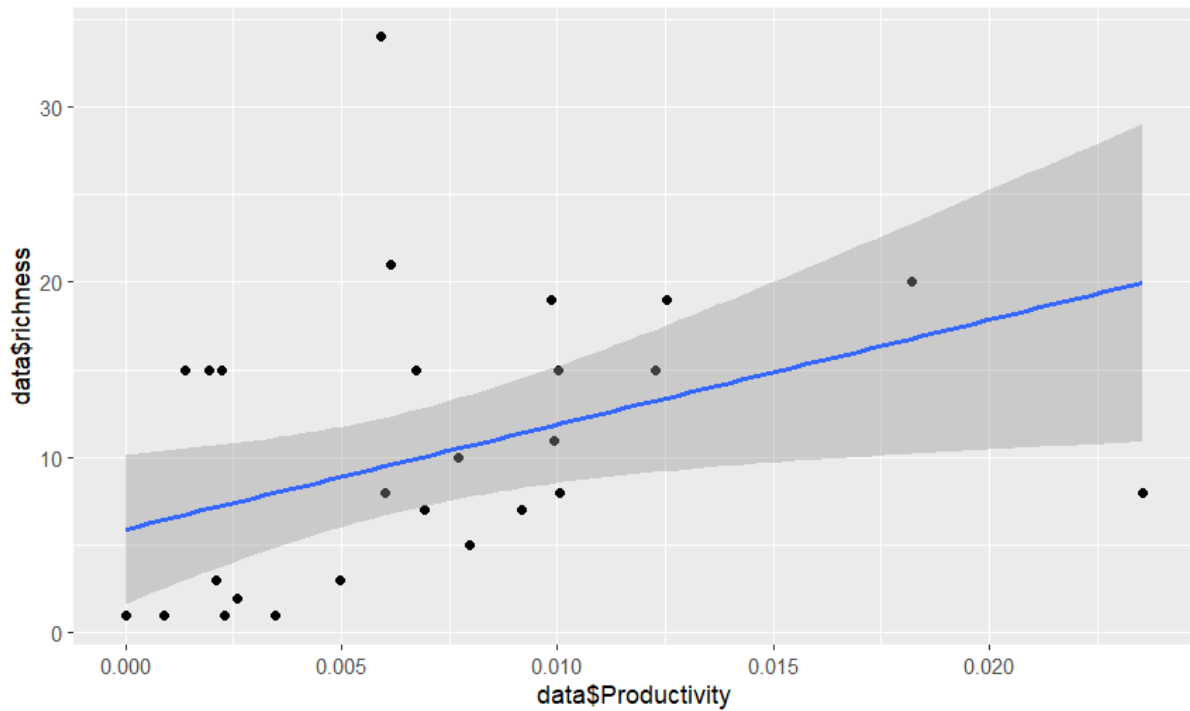
Model	AIC	deltaAIC	rel.LL	weights	Cumulative.Weight
GLM.nums.oxygen	182.378776085212	0	1	0.994305932846329	0.994305932846329
GLM.nums.temp	192.704016397897	10.3252403126847	0.00572667522661835	0.00569406715331072	0.999999999999964
				3.51343662941835e-	
GLM.nums.salinity	239.721378512393	57.342602427181	3.5335569399256e-13	13	0.999999999999991
				6.17126372971456e-	
GLM.nums.nitrate	247.80510092397	65.4263248387582	6.20660455283468e-15	15	0.999999999999998
				2.45178661848794e-	
GLM.nums.phosphate	249.651274213056	67.2724981278443	2.46582720417787e-15	15	1
				1.92672280436413e-	
GLM.nums.productivity	259.343607505087	76.9648314198747	1.93775652011714e-17	17	1
				2.20890260645859e-	
GLM.nums.records	268.280597270592	85.9018211853802	2.22155227429381e-19	19	1
GLM.nums.latitude	268.353155356048	85.9743792708362	2.14240093304535e-19	2.1302019582625e-19	1
				8.47121488266072e-	
GLM.nums.chlorophyll	270.197411260688	87.8186351754759	8.51972677907168e-20	20	1
				3.05068724230928e-	
GLM.nums.intercept	309.081386831396	126.702610746184	3.06815753736508e-28	28	1



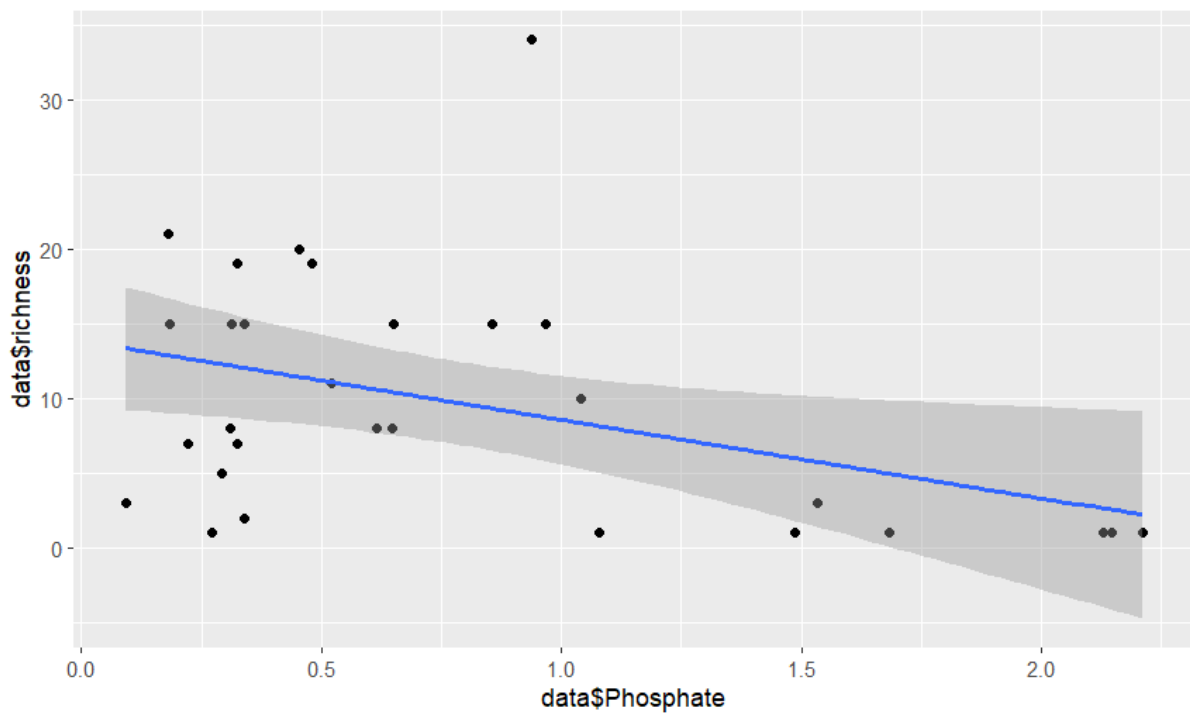
```
> qplot(data$Salinity, data$richness) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



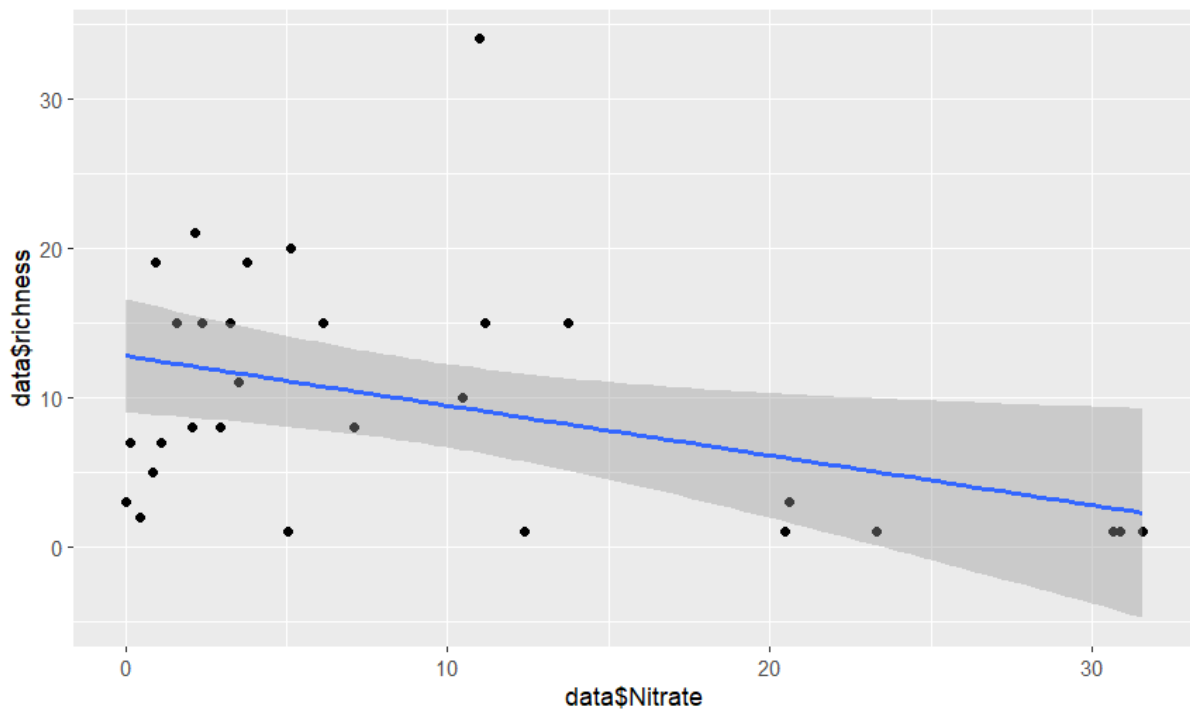
```
> qplot(data$Productivity, data$richness) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```

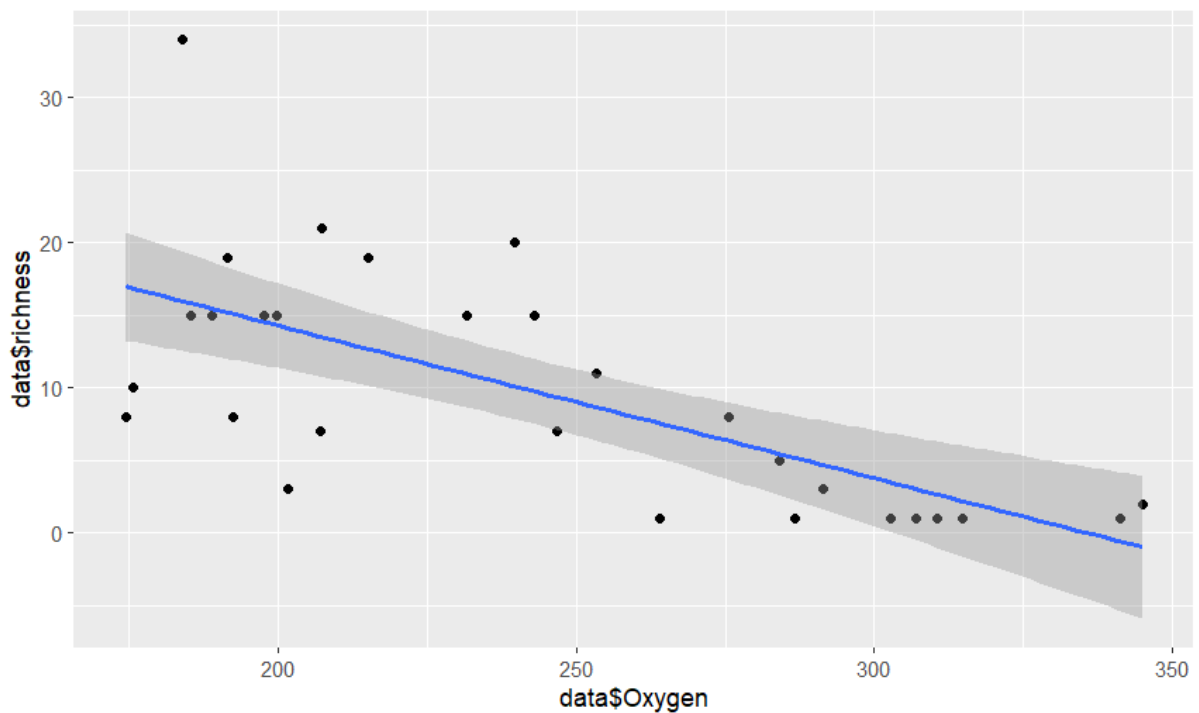
```
> qplot(data$Phosphate, data$richness) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



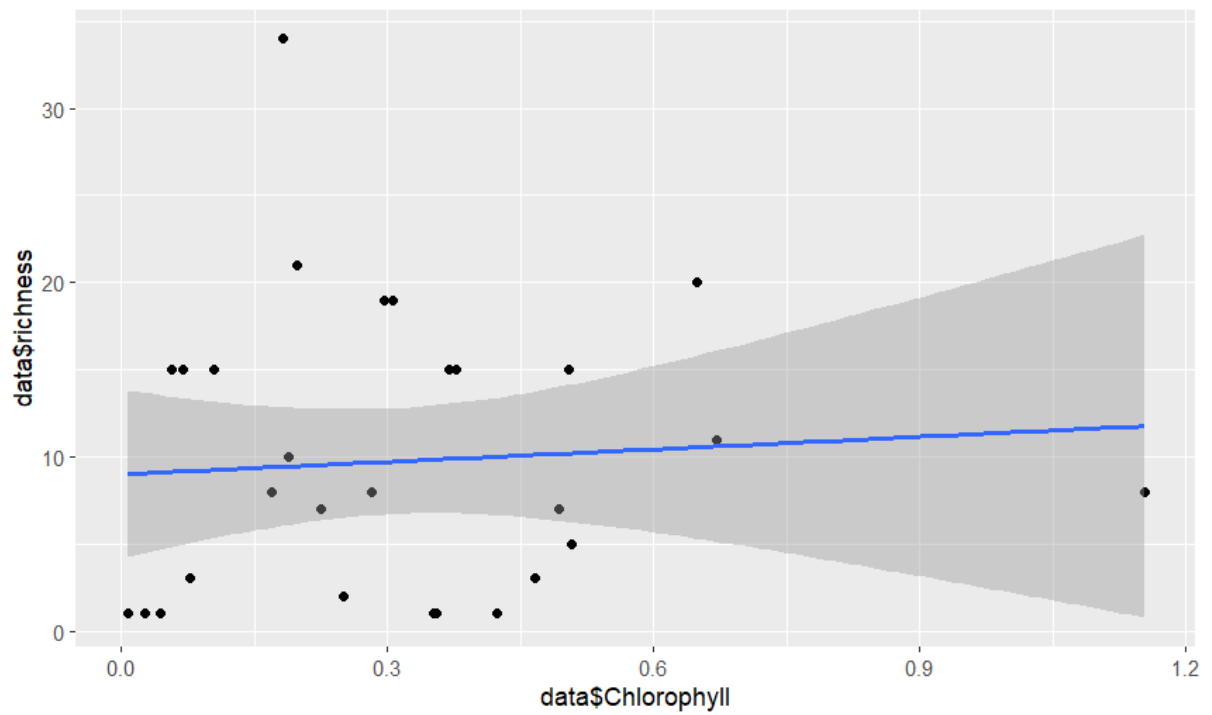
```
> qplot(data$Nitrate, data$richness) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



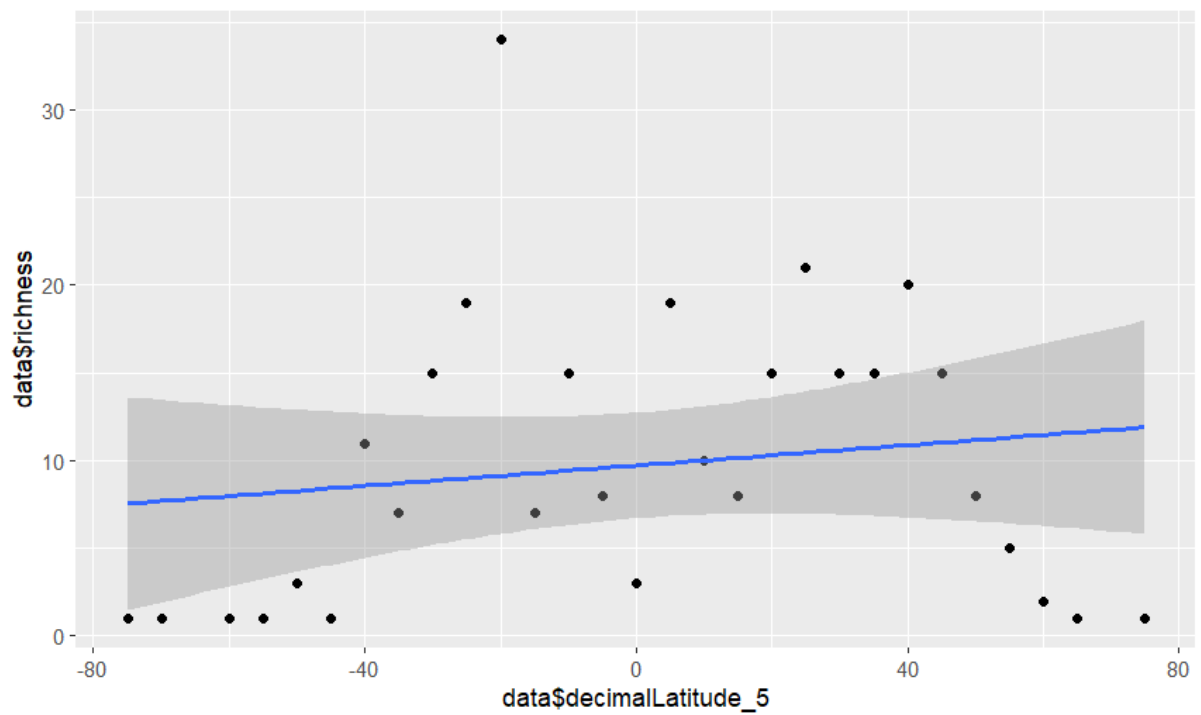
```
> qplot(data$Oxygen, data$richness) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



```
> qplot(data$Chlorophyll, data$richness) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



```
> qplot(data$decimalLatitude_5, data$richness) + geom_smooth(method = "glm")
)
geom_smooth() `using formula = 'y ~ x'`
```



GLM FOR ESTIMATED RICHNESS (ES15)

```
> GLM.ES15.intercept <- glm(as.integer(data$ES15) ~ 1, family = "poisson",
data = data)
> summary(GLM.ES15.intercept)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ 1, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.50790	0.08737	17.26	<2e-16 ***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 58.19  on 28  degrees of freedom
Residual deviance: 58.19  on 28  degrees of freedom
AIC: 150.87
```

Number of Fisher Scoring iterations: 5

```
> GLM.ES15.temp <- glm(as.integer(data$ES15) ~ Temperature, family = "poisson",
data = data)
> summary(GLM.ES15.temp)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Temperature, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.39875	0.23659	1.685	0.0919 .
Temperature	0.06558	0.01150	5.703	1.18e-08 ***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 58.190  on 28  degrees of freedom
Residual deviance: 20.271  on 27  degrees of freedom
AIC: 114.95
```

Number of Fisher Scoring iterations: 4

```
> GLM.ES15.salinity <- glm(as.integer(data$ES15) ~ Salinity, family = "poisson",
data = data)
> summary(GLM.ES15.salinity)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Salinity, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.61284	1.27971	-1.260	0.2076
Salinity	0.09174	0.03704	2.477	0.0133 *

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 58.190  on 28  degrees of freedom
Residual deviance: 47.311  on 27  degrees of freedom
AIC: 141.99
```

Number of Fisher Scoring iterations: 5

```
> GLM.ES15.productivity <- glm(as.integer(data$ES15) ~ Productivity, family
= "poisson", data = data)
> summary(GLM.ES15.productivity)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Productivity, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.2008	0.1396	8.602	< 2e-16 ***
Productivity	43.2400	13.7880	3.136	0.00171 **

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 58.190 on 28 degrees of freedom
Residual deviance: 49.182 on 27 degrees of freedom
AIC: 143.86

Number of Fisher Scoring iterations: 5

```
> GLM.ES15.phosphate <- glm(as.integer(data$ES15) ~ Phosphate, family = "poisson", data = data)
> summary(GLM.ES15.phosphate)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Phosphate, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.9468	0.1357	14.346	< 2e-16 ***
Phosphate	-0.6553	0.1801	-3.639	0.000274 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 58.190 on 28 degrees of freedom
Residual deviance: 42.198 on 27 degrees of freedom
AIC: 136.88

Number of Fisher Scoring iterations: 5

```
> GLM.ES15.nitrate <- glm(as.integer(data$ES15) ~ Nitrate, family = "poisson", data = data)
> summary(GLM.ES15.nitrate)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Nitrate, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.84296	0.11234	16.405	< 2e-16 ***
Nitrate	-0.04584	0.01213	-3.778	0.000158 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 58.190 on 28 degrees of freedom
Residual deviance: 40.321 on 27 degrees of freedom
AIC: 135

Number of Fisher Scoring iterations: 4

```
> GLM.ES15.oxygen <- glm(as.integer(data$ES15) ~ Oxygen, family = "poisson"
, data = data)
> summary(GLM.ES15.oxygen)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Oxygen, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.26091	0.46503	9.163	< 2e-16 ***
Oxygen	-0.01203	0.00211	-5.699	1.2e-08 ***

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 58.190  on 28  degrees of freedom
Residual deviance: 18.929  on 27  degrees of freedom
AIC: 113.61
```

Number of Fisher Scoring iterations: 4

```
> GLM.ES15.chlorophyll <- glm(as.integer(data$ES15) ~ Chlorophyll, family =
"poisson", data = data)
> summary(GLM.ES15.chlorophyll)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ Chlorophyll, family = "poisson",
    data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.4766	0.1399	10.556	<2e-16 ***
Chlorophyll	0.1017	0.3513	0.289	0.772

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 58.190  on 28  degrees of freedom
Residual deviance: 58.107  on 27  degrees of freedom
AIC: 152.79
```

Number of Fisher Scoring iterations: 5

```
> GLM.ES15.latitude <- glm(as.integer(data$ES15) ~ decimalLatitude_5, famil
y = "poisson", data = data)
> summary(GLM.ES15.latitude)
```

```
Call:
glm(formula = as.integer(data$ES15) ~ decimalLatitude_5, family = "poisson"
, data = data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.508e+00	8.737e-02	17.259	<2e-16 ***
decimalLatitude_5	9.457e-05	2.046e-03	0.046	0.963

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 58.190  on 28  degrees of freedom
Residual deviance: 58.188  on 27  degrees of freedom
AIC: 152.87
```

Number of Fisher Scoring iterations: 5

```

> GLM.ES15.models <- list(GLM.ES15.intercept = GLM.ES15.intercept,
  GLM.ES15.latitude = GLM.ES15.latitude,
  GLM.ES15.temp = GLM.ES15.temp,
  GLM.ES15.salinity = GLM.ES15.salinity,
  GLM.ES15.productivity = GLM.ES15.productivity,
  GLM.ES15.phosphate = GLM.ES15.phosphate,
  GLM.ES15.nitrate = GLM.ES15.nitrate,
  GLM.ES15.oxygen = GLM.ES15.oxygen,
  GLM.ES15.chlorophyll = GLM.ES15.chlorophyll)
> GLM.ES15.aic.df <- data.frame(Model = names(GLM.ES15.models),
AIC = sapply(GLM.ES15.models, function(x) AICc(x)),
  akaike.weights(sapply(GLM.ES15.models, function(x) AICc(x))))

> GLM.ES15.aic.df <- GLM.ES15.aic.df[order(GLM.ES15.aic.df$AIC),]
> GLM.ES15.aic.df$Cumulative.Weight <- cumsum(GLM.ES15.aic.df$weights)

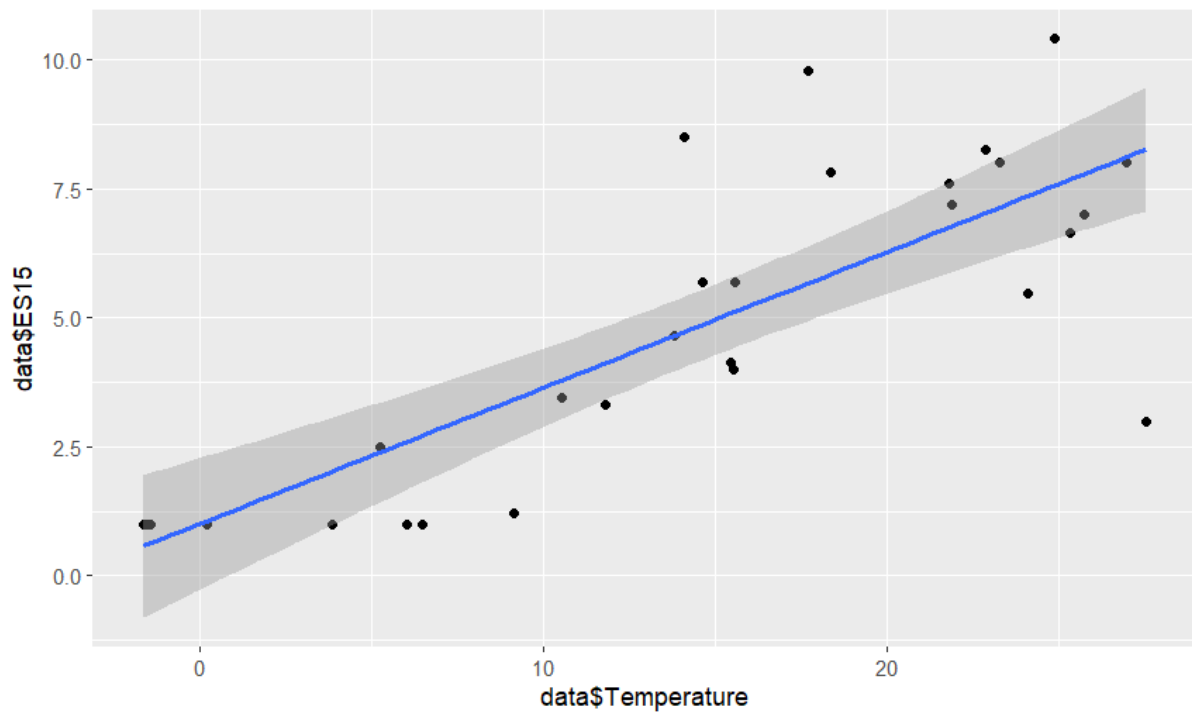
```

Model	AIC	deltaAIC	rel.LL	weights	Cumulative.Weight
GLM.ES15.oxygen	114.070190146627	0	1	0.661714107260068	0.661714107260068
GLM.ES15.temp	115.412201624122	1.34201147749503	0.511194191354967	0.338264407968985	0.999978515229053
GLM.ES15.nitrate	135.462007611575	21.3918174649481	2.26373642270366e-05	1.49794632602145e-05	0.999993494692313
GLM.ES15.phosphate	137.338391592772	23.2682014461448	8.85878688569289e-06	5.86198425547347e-06	0.999999356676569
GLM.ES15.salinity	142.452108871624	28.3819187249975	6.869808894546e-07	4.54584945970179e-07	0.999999811261515
GLM.ES15.productivity	144.322509995695	30.2523198490682	2.69644901136344e-07	1.78427835032665e-07	0.99999998968935
GLM.ES15.intercept	151.01717818199	36.9469880353636	9.4855711647959e-09	6.27673625516477e-09	0.999999995966086
GLM.ES15.chlorophyll	153.247777725213	39.1775875785865	3.10953142504982e-09	2.05762081092397e-09	0.999999998023707
GLM.ES15.latitude	153.328432593532	39.2582424469049	2.98662687012181e-09	1.97629313308159e-09	1

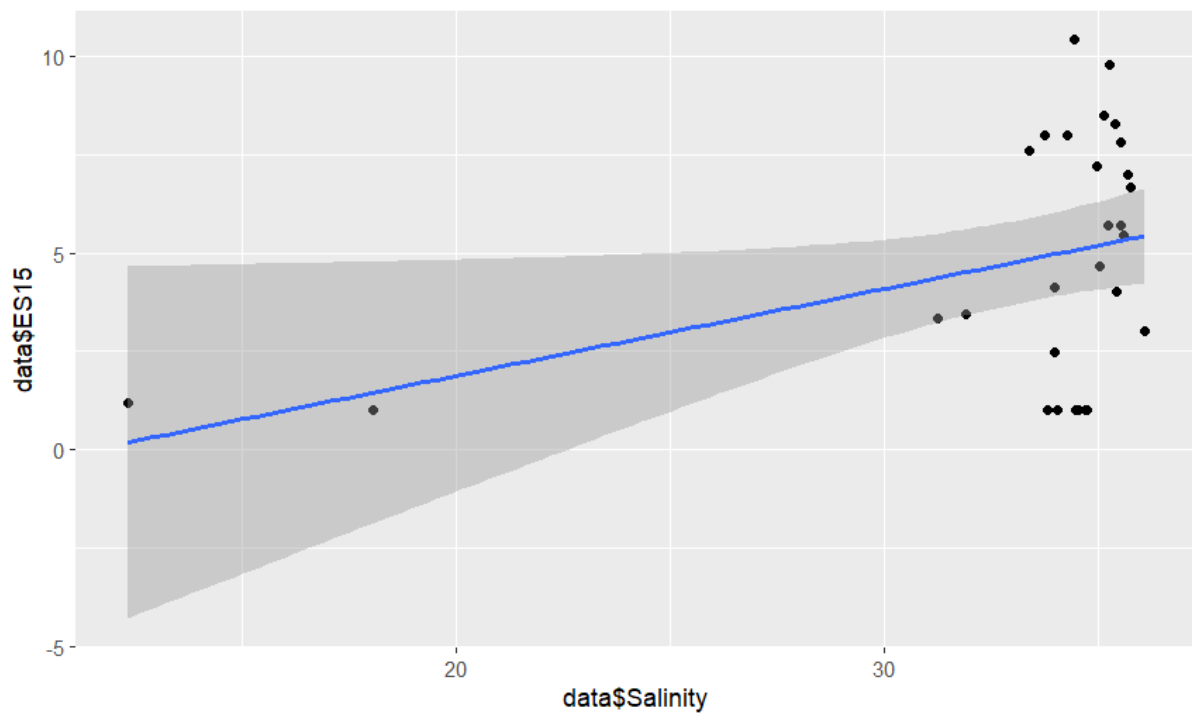
```

> qplot(data$Temperature, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'

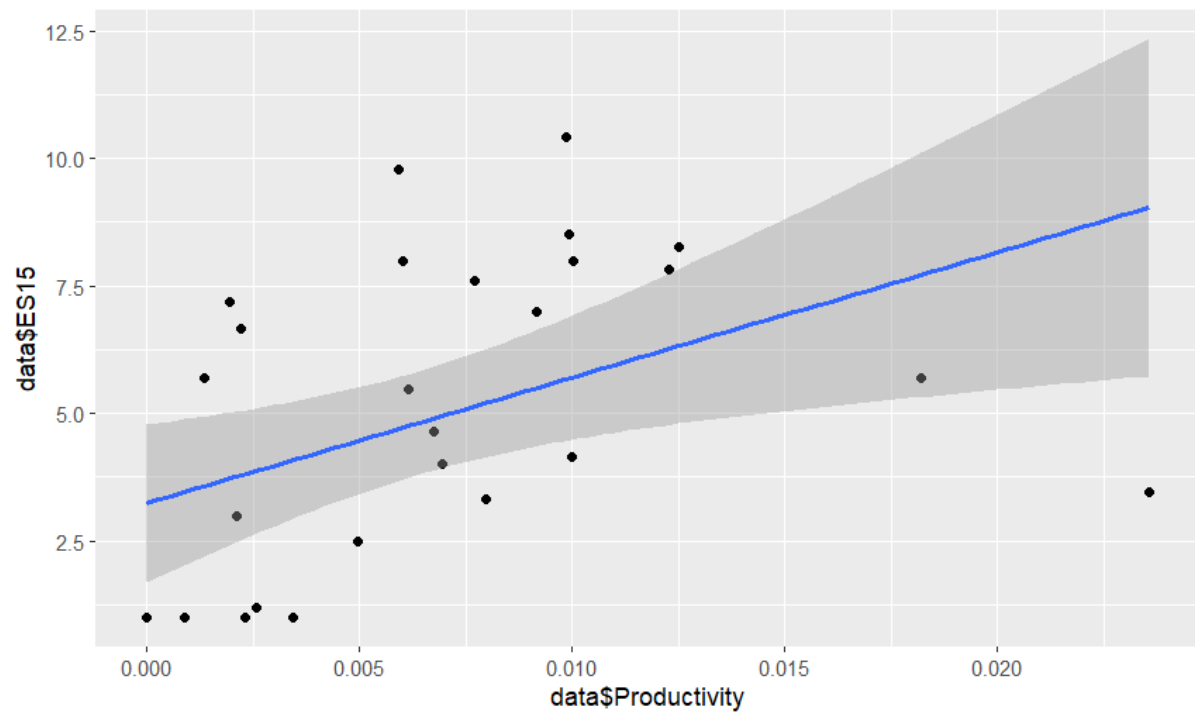
```



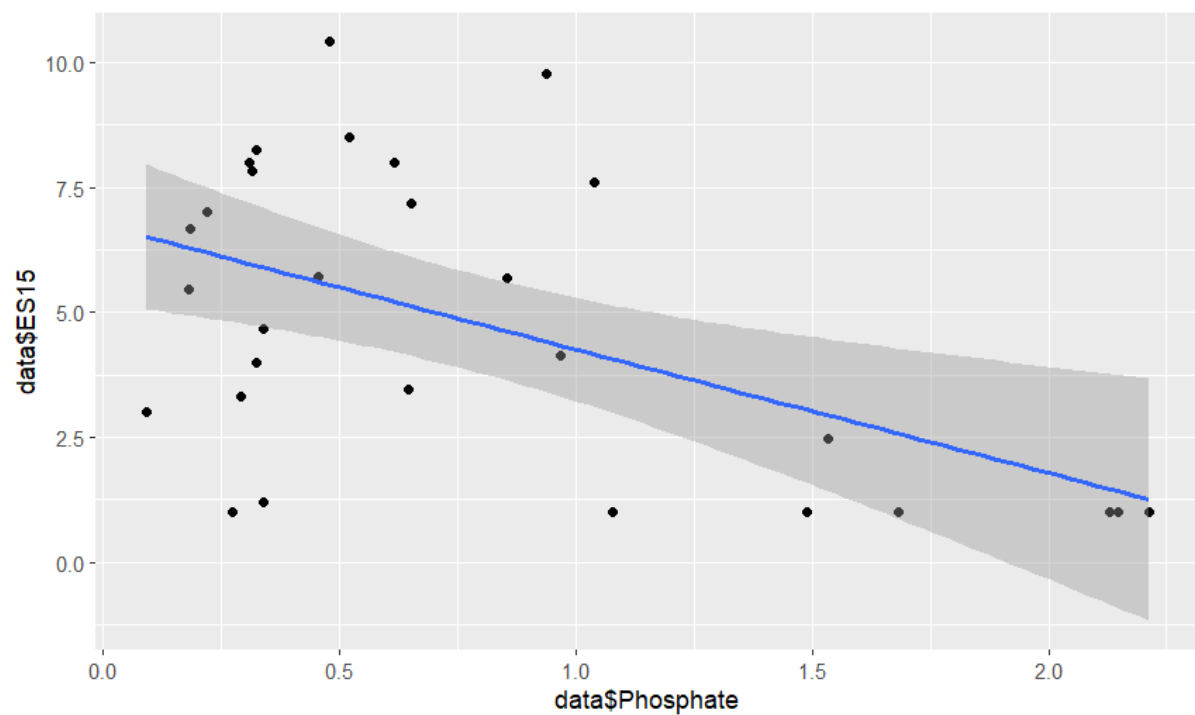
```
> qplot(data$Salinity, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



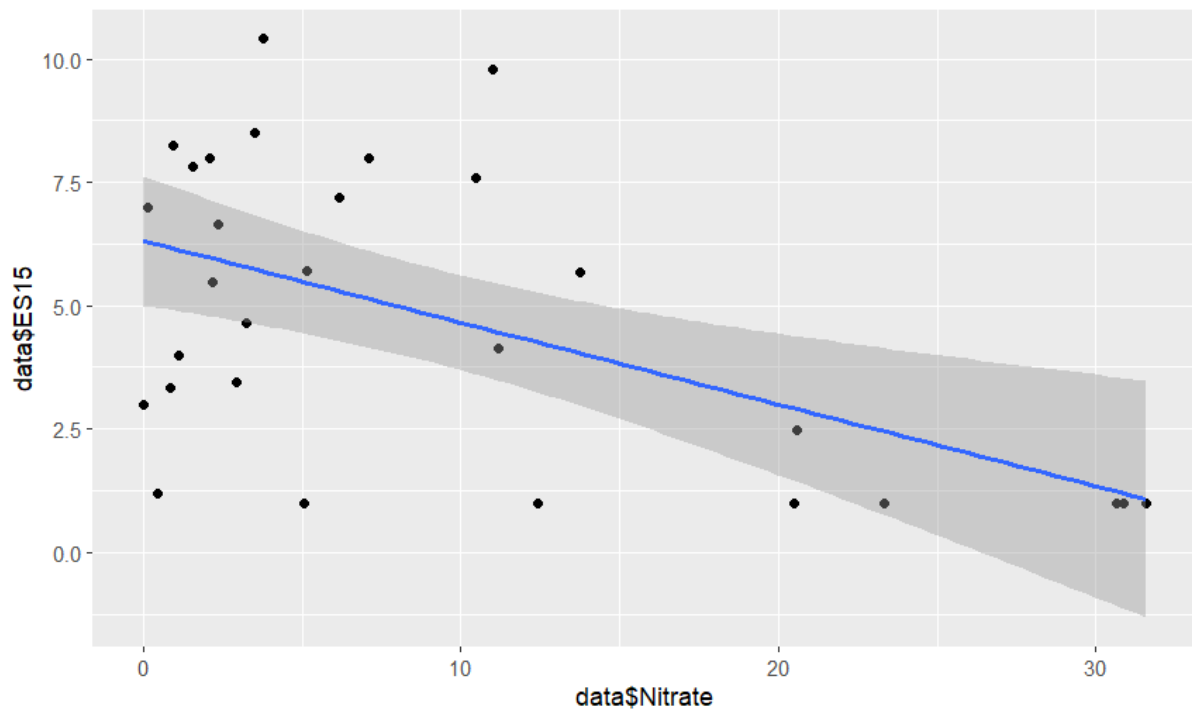
```
> qplot(data$Productivity, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```

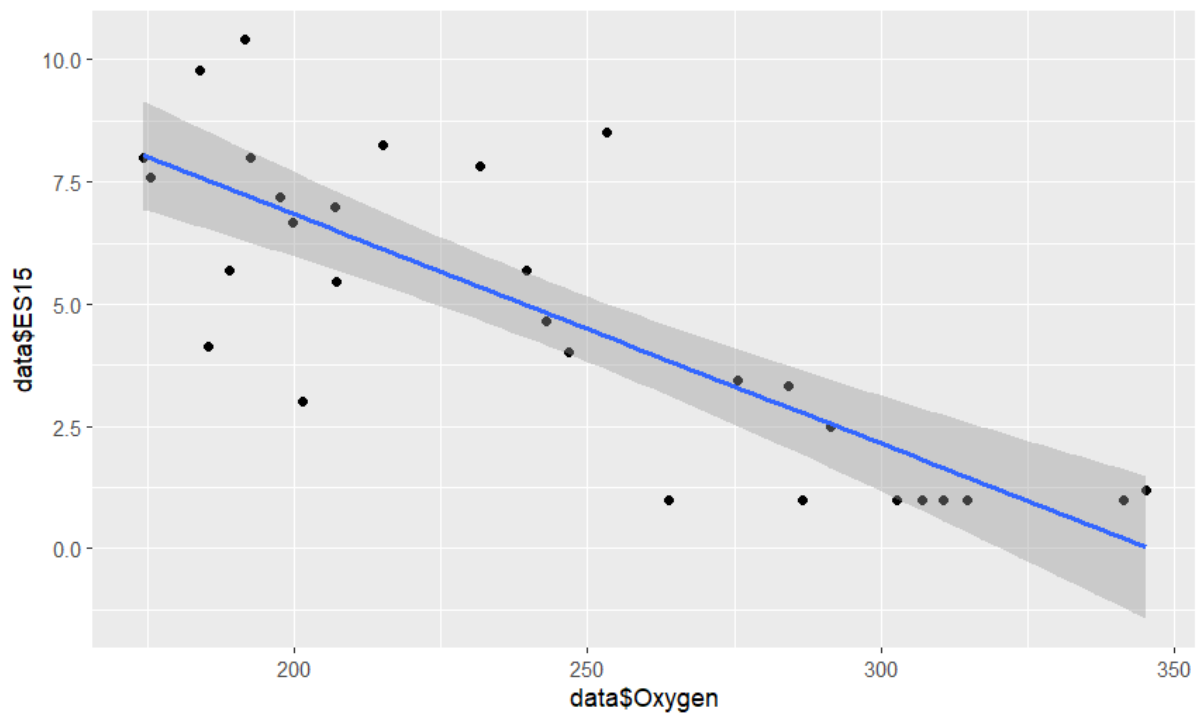
```
> qplot(data$Phosphate, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



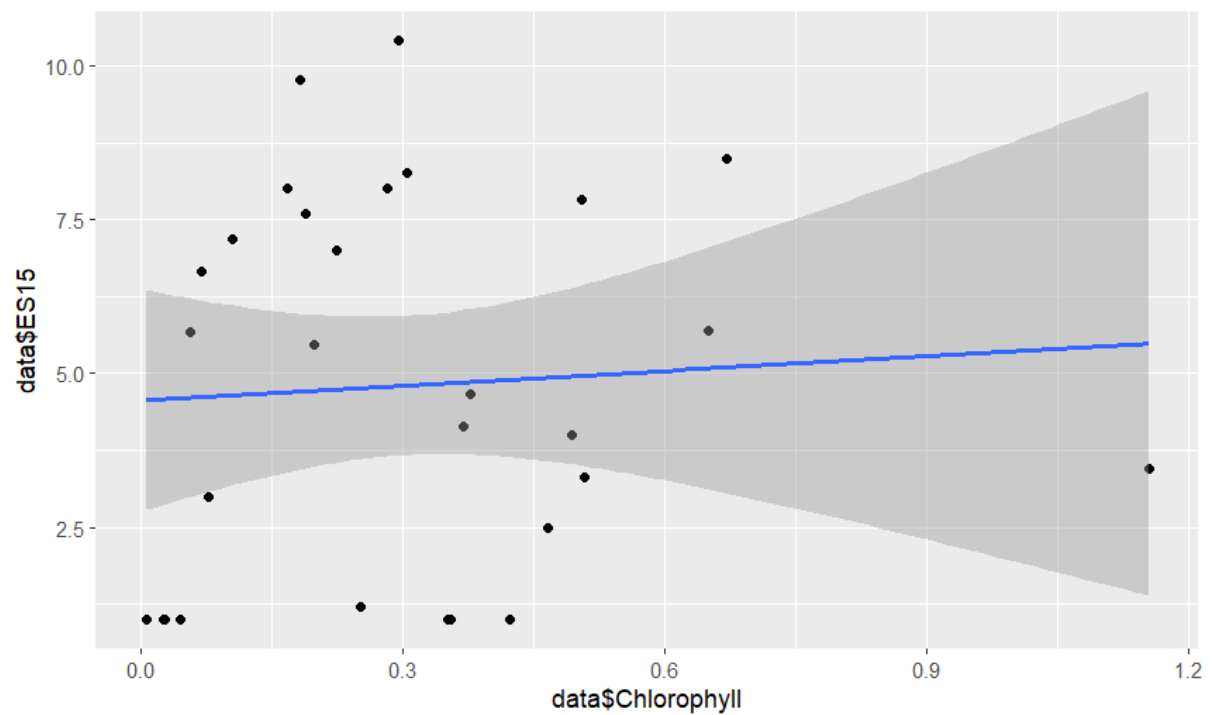
```
> qplot(data$Nitrate, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



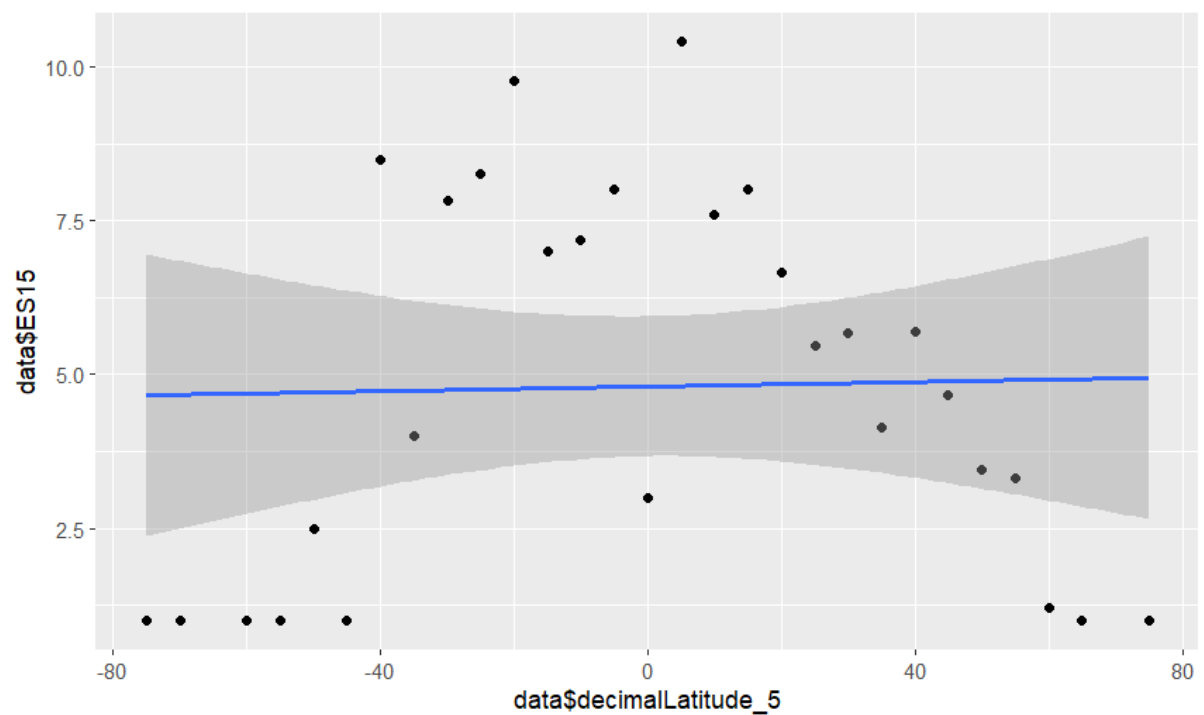
```
> qplot(data$Oxygen, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



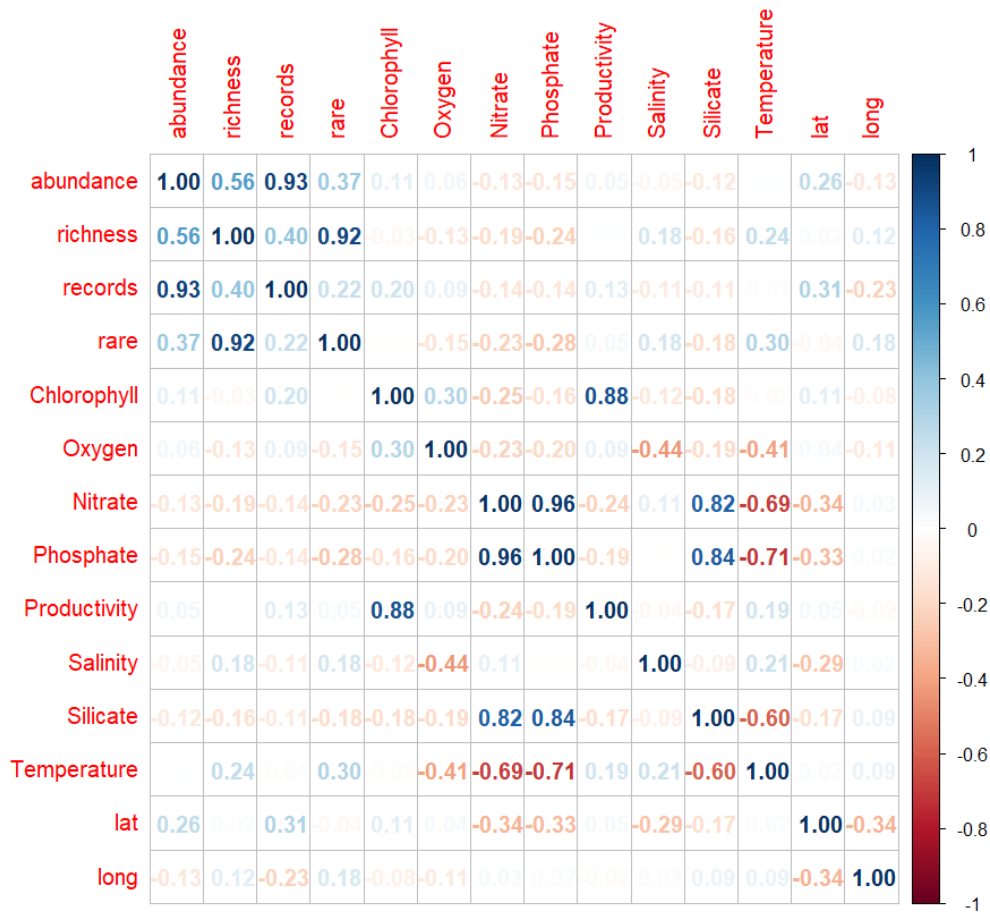
```
> qplot(data$Chlorophyll, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



```
> qplot(data$decimalLatitude_5, data$ES15) + geom_smooth(method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```

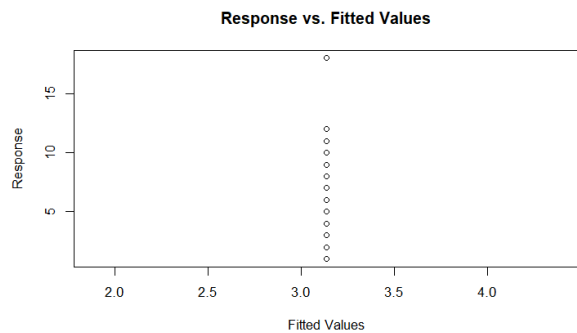
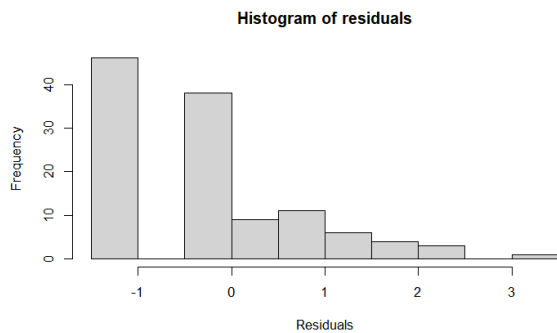
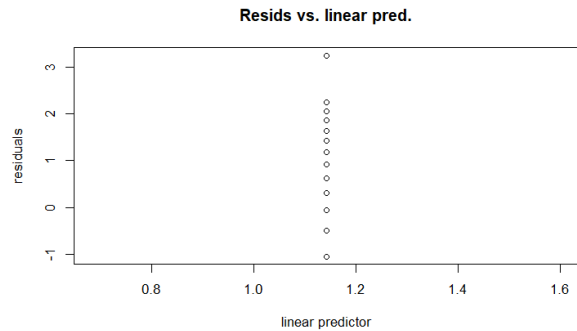
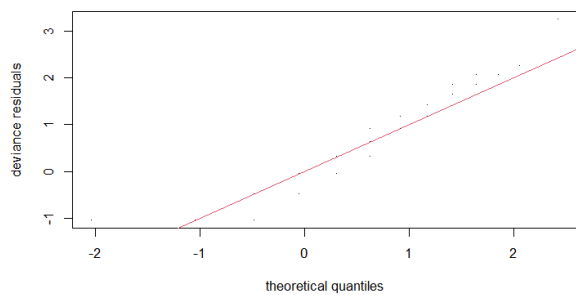


GAM FOR RICHNESS PER HEXAGON 3



```
gam.nums.intercept <- gam(richness ~ 1, data = data, family = "nb", method
= "RELM", select = TRUE)
> par(mfrow = c(2,2))
> gam.check(gam.nums.intercept)
```

Method: REML Optimizer: outer newton
full convergence after 4 iterations.
Gradient range [-7.376129e-09,-7.376129e-09]
(score 259.524 & scale 1).
Hessian positive definite, eigenvalue range [19.21506,19.21506].
Model rank = 1 / 1



```
> summary(gam.nums.intercept)
```

```
Family: Negative Binomial(2.691)
Link function: log
```

```
Formula:
richness ~ 1
```

```
Parametric coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.1428      0.0765   14.94  <2e-16 ***
```

```
---
```

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

```
R-sq.(adj) =    0   Deviance explained = 1.12e-08%
```

```
-REML = 259.52  scale est. = 1          n = 118
```

```
> gam.nums.records <- gam(richness ~ s(records), data = data, family = "nb"
, method = "RELM", select = TRUE)
```

```
> gam.check(gam.nums.records)
```

```
Method: REML   Optimizer: outer newton
```

```
full convergence after 6 iterations.
```

```
Gradient range [-5.968948e-05,1.308381e-05]
```

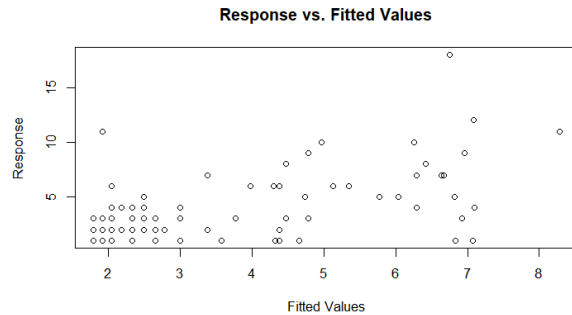
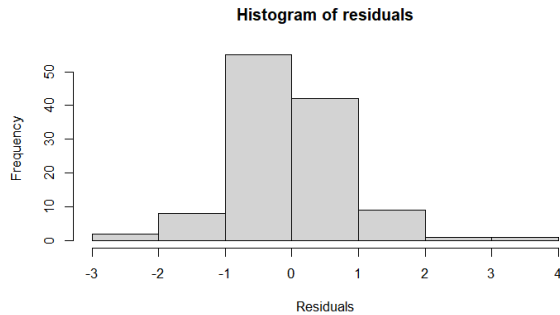
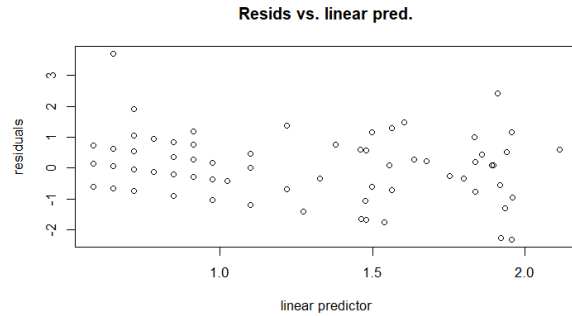
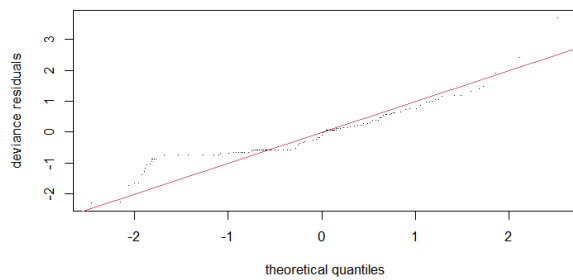
```
(score 238.8067 & scale 1).
```

```
Hessian positive definite, eigenvalue range [3.755091e-05,5.590091].
```

```
Model rank = 10 / 10
```

```
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(records) 9.00 4.92   1.08   0.86
```



```
> summary(gam.nums.records)
```

```
Family: Negative Binomial(7.96)
Link function: log
```

```
Formula:
richness ~ s(records)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.0119     0.0665   15.22  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(records)  4.923     9   72.7  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

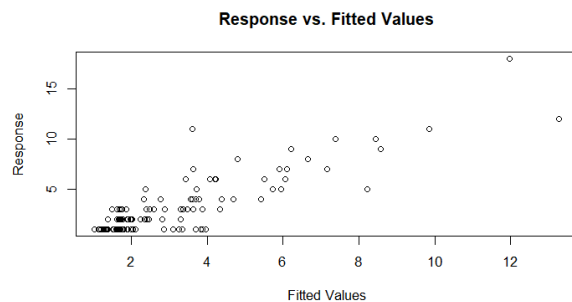
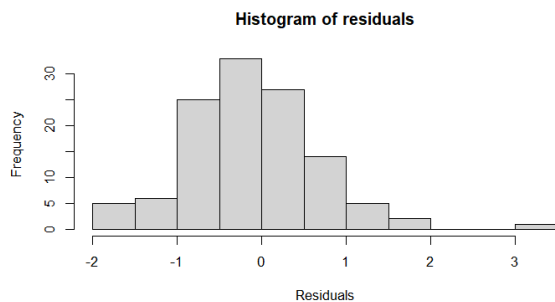
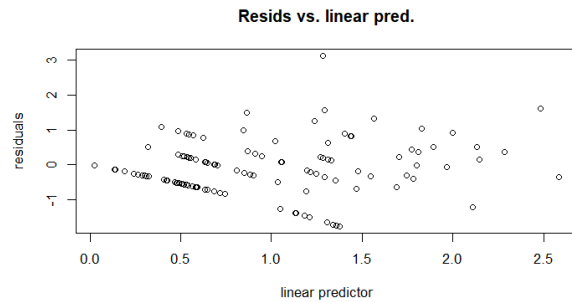
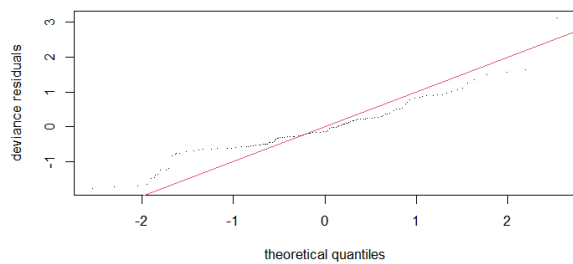
```
R-sq.(adj) = 0.382   Deviance explained = 46.4%
-REML = 238.81   Scale est. = 1         n = 118
```

```
> gam.nums.latlong <- gam(richness ~s(records) + s(lat, long, bs = "sos"),
data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.latlong)
```

```
Method: REML   Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-0.0001954683,-2.617608e-06]
(score 229.8657 & scale 1).
Hessian positive definite, eigenvalue range [9.192234e-05,2.887111].
Model rank = 59 / 59
```

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(records)	9.00	5.09	0.98	0.46
s(lat,long)	49.00	12.77	1.01	0.62



```
> summary(gam.nums.latlong)
```

Family: Negative Binomial(84550.54)
Link function: log

Formula:
richness ~ s(records) + s(lat, long, bs = "sos")

Parametric coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.96411 0.05976 16.13 <2e-16 ***

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

Approximate significance of smooth terms:
 edf Ref.df Chi.sq p-value
s(records) 5.088 9 79.59 <2e-16 ***
s(lat,long) 12.771 49 54.02 <2e-16 ***

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

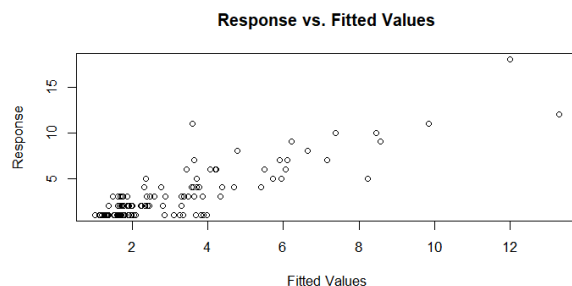
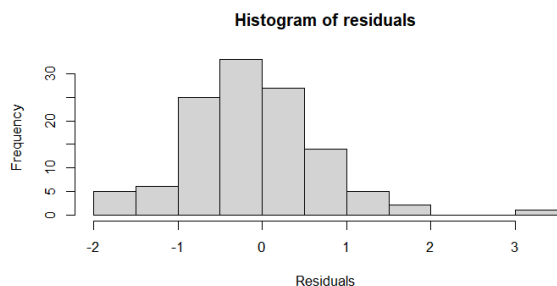
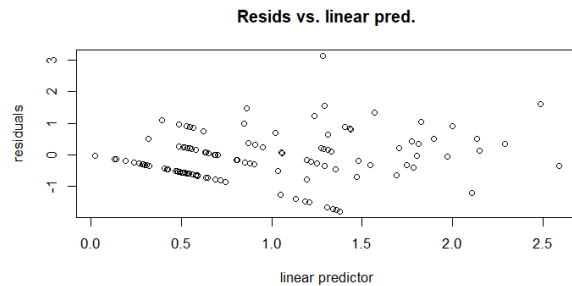
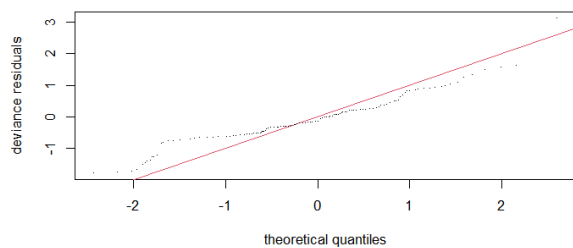
R-sq.(adj) = 0.679 Deviance explained = 71.5%
-REML = 229.87 Scale est. = 1 n = 118

```
> gam.nums.chlorophyll1 <- gam(richness ~s(records) + s(lat, long, bs = "so  
s") + s(Chlorophyll1), data = data, family = "nb", select = TRUE)  
> gam.check(gam.nums.chlorophyll1)
```

Method: REML Optimizer: outer newton
full convergence after 15 iterations.
Gradient range [-3.29772e-05,-2.192878e-07]
(score 229.8656 & scale 1).
Hessian positive definite, eigenvalue range [2.991768e-06,2.887146].
Model rank = 68 / 68

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(records)	9.00e+00	5.09e+00	0.98	0.46
s(lat,long)	4.90e+01	1.28e+01	1.01	0.65
s(Chlorophyll1)	9.00e+00	2.33e-05	1.16	0.96



```
> summary(gam.nums.chlorophyll)
```

```
Family: Negative Binomial(513948.5)
Link function: log
```

```
Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Chlorophyll)
```

```
Parametric coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.96411	0.05976	16.13	<2e-16 ***

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

```
Approximate significance of smooth terms:
```

	edf	Ref.df	Chi.sq	p-value
s(records)	5.088e+00	9	79.59	<2e-16 ***
s(lat,long)	1.277e+01	49	54.02	<2e-16 ***
s(Chlorophyll)	2.332e-05	9	0.00	0.82

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

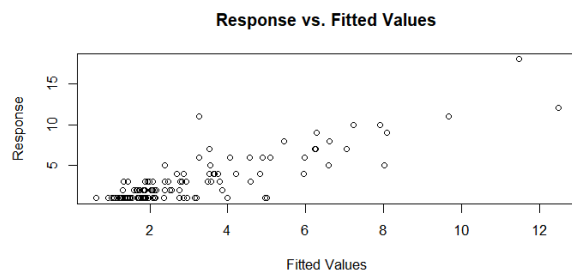
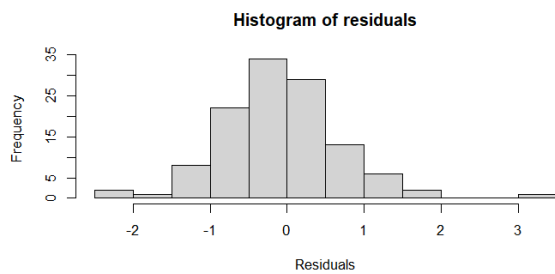
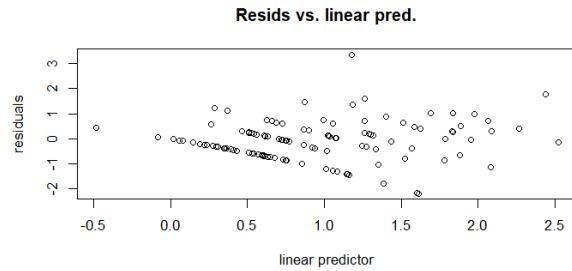
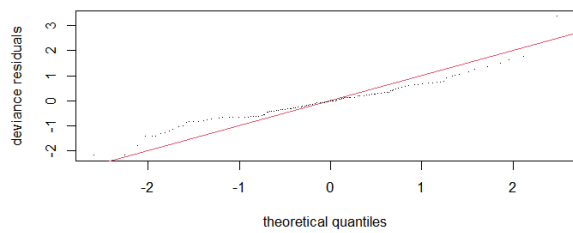
```
R-sq.(adj) = 0.679   Deviance explained = 71.5%
-REML = 229.87   Scale est. = 1           n = 118
```

```
> gam.nums.oxygen <- gam(richness ~s(records) + s(lat, long, bs = "sos") +
s(Oxygen), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.oxygen)
```

```
Method: REML   Optimizer: outer newton
full convergence after 12 iterations.
Gradient range [-0.0007639314,1.048351e-05]
(score 227.3283 & scale 1).
Hessian positive definite, eigenvalue range [1.511103e-05,1.682518].
Model rank = 68 / 68
```

```
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(records)	9.00	5.04	1.06	0.75
s(lat,long)	49.00	7.91	1.01	0.56
s(Oxygen)	9.00	2.15	0.93	0.24



```
> summary(gam.nums.oxygen)
```

```
Family: Negative Binomial(12954.429)
Link function: log
```

```
Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Oxygen)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.95550    0.06049   15.79  <2e-16 ***
---

```

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

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq  p-value
s(records)    5.038     9  87.12  < 2e-16 ***
s(lat,long)   7.909    49  25.73  6.92e-05 ***
s(Oxygen)     2.151     9  10.91  0.000654 ***
---

```

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

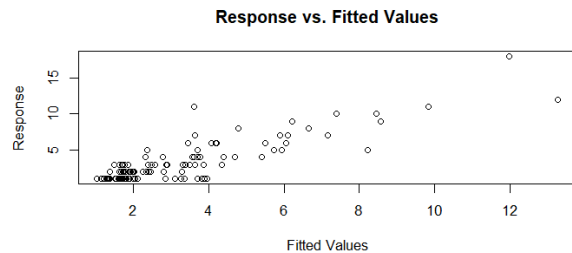
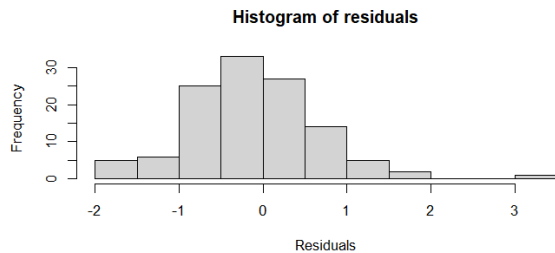
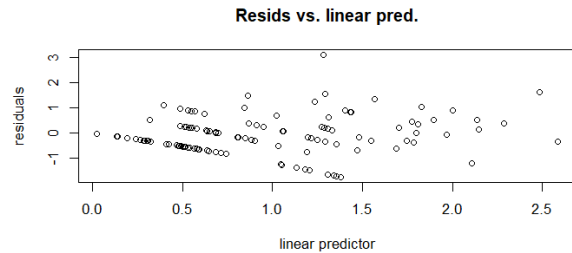
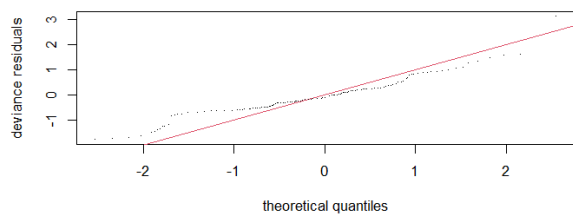
```
R-sq.(adj) = 0.663   Deviance explained = 70.1%
-REML = 227.33   Scale est. = 1           n = 118
```

```
> gam.nums.nitrate <- gam(richness ~s(records) + s(lat, long, bs = "sos")
+ s(Nitrate), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.nitrate)
```

```
Method: REML   Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-0.0002568898,-2.573382e-06]
(score 229.8657 & scale 1).
Hessian positive definite, eigenvalue range [6.116087e-06,2.887094].
Model rank = 68 / 68
```

```
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(records)	9.00e+00	5.09e+00	0.98	0.47
s(lat,long)	4.90e+01	1.28e+01	1.01	0.59
s(Nitrate)	9.00e+00	7.52e-05	1.02	0.57



```
> summary(gam.nums.nitrate)
```

Family: Negative Binomial(66230.325)
Link function: log

Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Nitrate)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.96411	0.05976	16.13	<2e-16 ***

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

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(records)	5.088e+00	9	79.59	<2e-16 ***
s(lat,long)	1.277e+01	49	54.02	<2e-16 ***
s(Nitrate)	7.516e-05	9	0.00	0.411

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

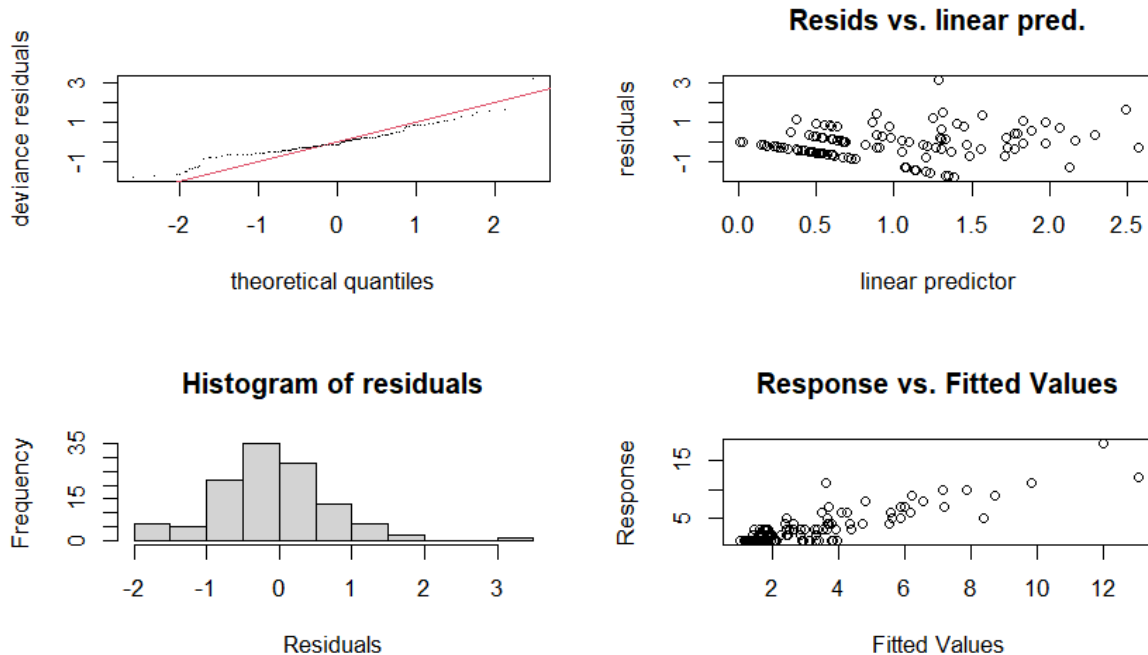
R-sq.(adj) = 0.679 Deviance explained = 71.5%
-REML = 229.87 Scale est. = 1 n = 118

```
> gam.nums.phosphate <- gam(richness ~s(records) + s(lat, long, bs = "sos")
+ s(Phosphate), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.phosphate)
```

Method: REML Optimizer: outer newton
full convergence after 14 iterations.
Gradient range [-0.0001326657,3.234901e-06]
(score 229.7988 & scale 1).
Hessian positive definite, eigenvalue range [3.137372e-06,2.780264].
Model rank = 68 / 68

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(records)	9.000	5.061	0.98	0.46
s(lat,long)	49.000	12.425	1.01	0.62
s(Phosphate)	9.000	0.403	1.07	0.79



```
> summary(gam.nums.phosphate)
```

```
Family: Negative Binomial(128882.696)
Link function: log
```

```
Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Phosphate)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.96320    0.05984   16.1    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(records)    5.0613     9  78.919 <2e-16 ***
s(lat,long)   12.4248    49  52.667 <2e-16 ***
s(Phosphate)  0.4028     9   0.632  0.188
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.675   Deviance explained = 71.5%
-REML = 229.8   Scale est. = 1           n = 118
```

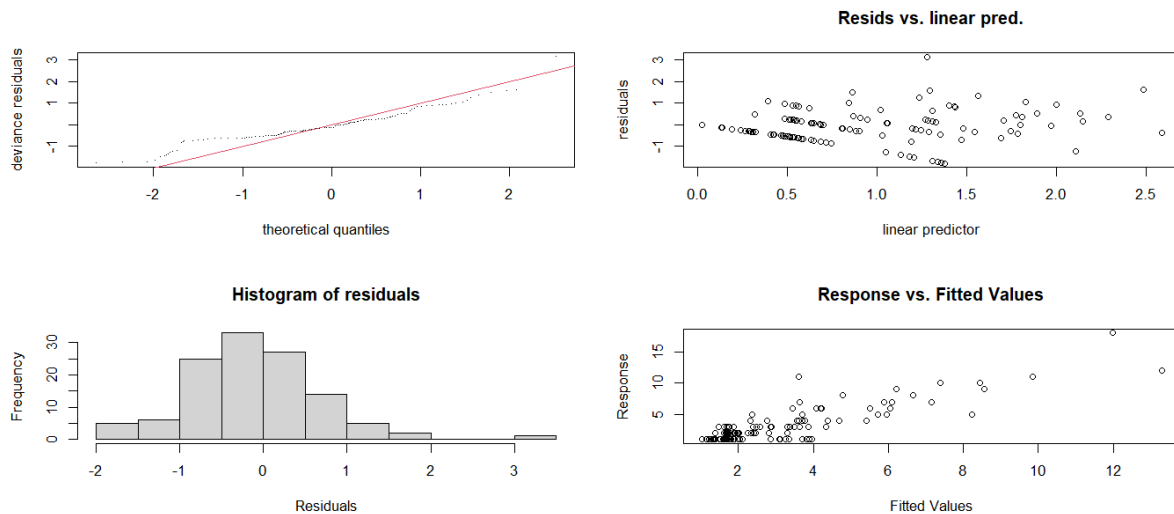
```
> gam.nums.productivity <- gam(richness ~s(records) + s(lat, long, bs = "s
os") + s(Productivity), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.productivity)
```

```
Method: REML   Optimizer: outer newton
full convergence after 24 iterations.
Gradient range [-1.125279e-05,4.762344e-07]
(score 229.8656 & scale 1).
Hessian positive definite, eigenvalue range [1.488824e-06,2.887155].
Model rank = 68 / 68
```

```
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(records)    9.00e+00  5.09e+00   0.98   0.46
```

```
s(lat,long)      4.90e+01 1.28e+01    1.01    0.61
s(Productivity)  9.00e+00 1.22e-05    1.11    0.89
```



```
> summary(gam.nums.productivity)
```

```
Family: Negative Binomial(1392446.109)
Link function: log
```

```
Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Productivity)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.96411    0.05976   16.13  <2e-16 ***
---
```

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

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(records)    5.088e+00     9  79.59 <2e-16 ***
s(lat,long)    1.277e+01    49  54.03 <2e-16 ***
s(Productivity) 1.223e-05     9   0.00  0.695
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

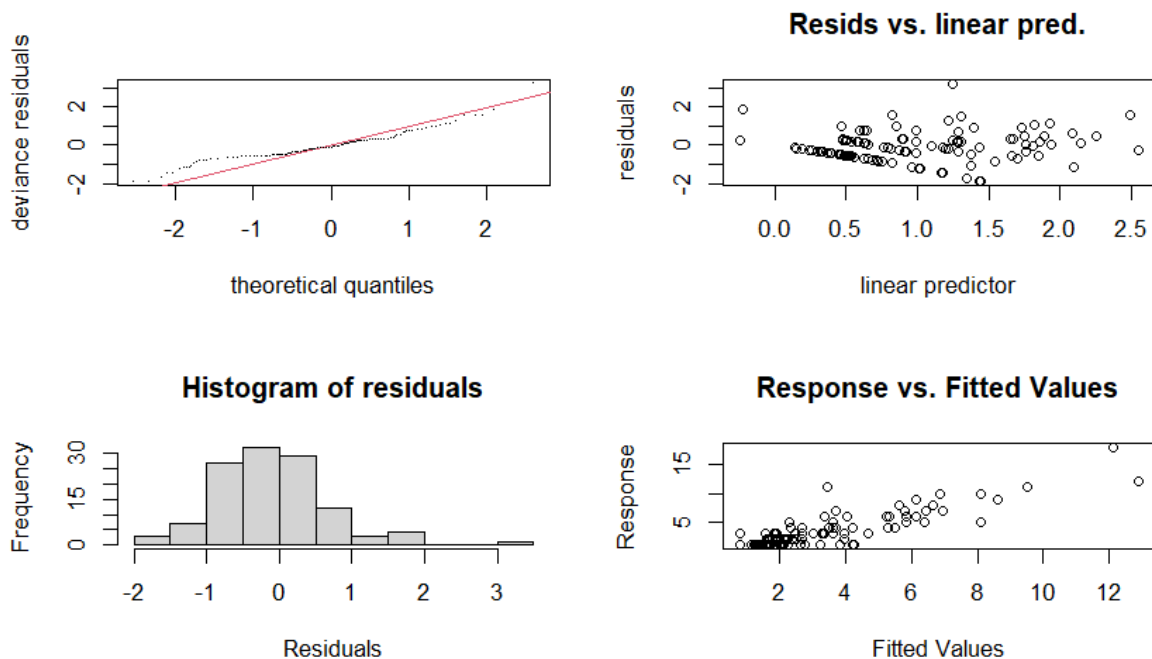
```
R-sq.(adj) = 0.679   Deviance explained = 71.5%
-REML = 229.87   Scale est. = 1           n = 118
```

```
> gam.nums.salinity <- gam(richness ~s(records) + s(lat, long, bs = "sos")
+ s(Salinity), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.salinity)
```

```
Method: REML   Optimizer: outer newton
full convergence after 15 iterations.
Gradient range [-9.584003e-05,-6.088454e-07]
(score 227.213 & scale 1).
Hessian positive definite, eigenvalue range [3.335754e-05,2.521786].
Model rank = 68 / 68
```

```
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(records)    9.00  4.88    0.99    0.50
s(lat,long)   49.00 10.80    1.00    0.53
s(Salinity)   9.00  0.88    1.03    0.61
```



```
> summary(gam.nums.salinity)
```

```
Family: Negative Binomial(398338.09)
Link function: log
```

```
Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Salinity)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.95780    0.06018   15.91  <2e-16 ***
---
```

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

```
Approximate significance of smooth terms:
```

```
              edf Ref.df Chi.sq p-value
s(records)    4.8785     9  80.984  <2e-16 ***
s(lat,long)  10.7985    49  43.325  <2e-16 ***
s(Salinity)   0.8797     9   6.759   0.0032 **
---
```

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

```
R-sq.(adj) =  0.69   Deviance explained = 71.8%
-REML = 227.21  Scale est. = 1          n = 118
```

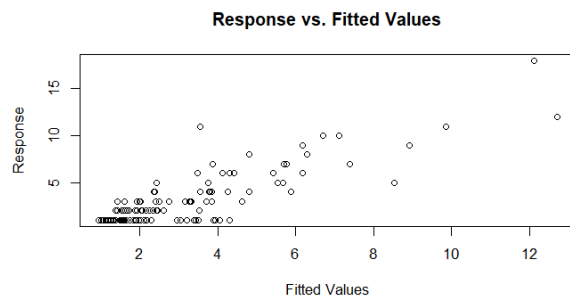
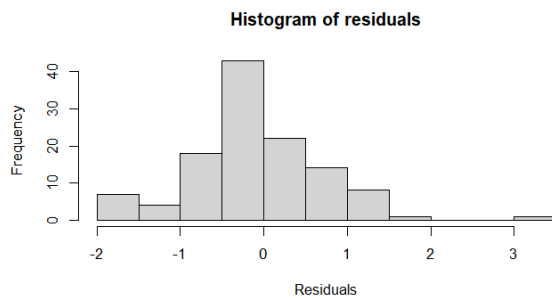
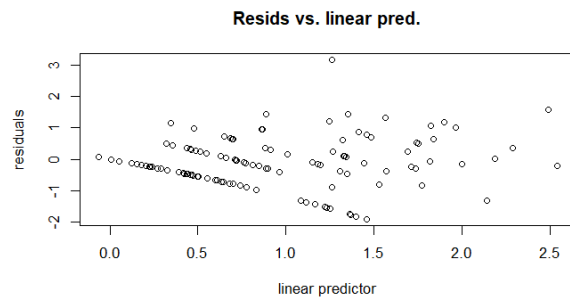
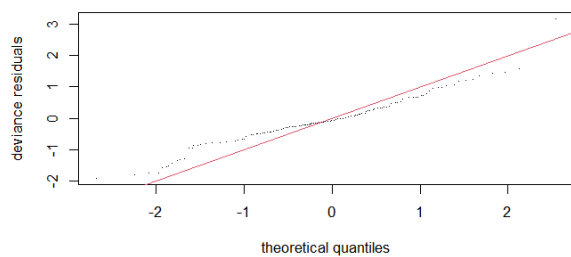
```
> gam.nums.temp <- gam(richness ~s(records) + s(lat, long, bs = "sos") + s
(Temperature), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.temp)
```

```
Method: REML   Optimizer: outer newton
full convergence after 14 iterations.
Gradient range [-0.000657695,-1.97659e-05]
(score 229.213 & scale 1).
Hessian positive definite, eigenvalue range [2.917431e-05,1.837214].
Model rank = 68 / 68
```

```
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(records)    9.000  5.097   1.00   0.52
```

```
s(lat,long)    49.000 10.272    1.00    0.55
s(Temperature) 9.000  0.772    1.15    0.97
```



```
> summary(gam.nums.temp)
```

```
Family: Negative Binomial(8265.905)
Link function: log
```

```
Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Temperature)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.96126    0.06017   15.97  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq  p-value
s(records)    5.097     9  83.669  < 2e-16 ***
s(lat,long)   10.272    49  34.382  5.58e-06 ***
s(Temperature) 0.772     9   3.229   0.0204 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

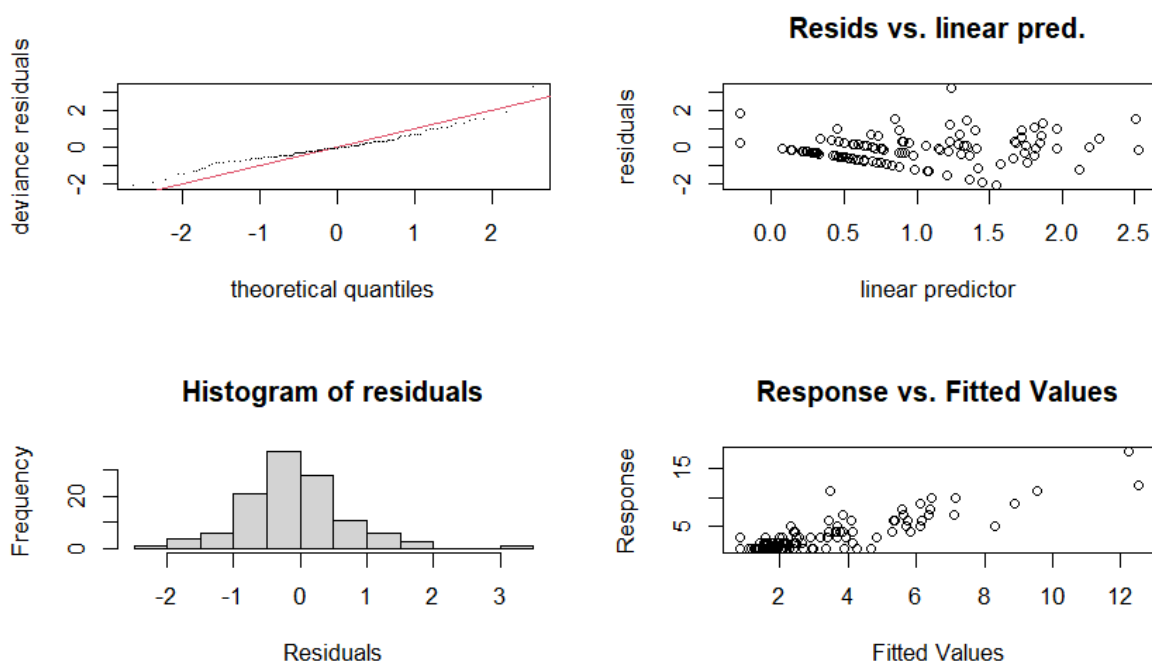
```
R-sq.(adj) = 0.66   Deviance explained = 69.8%
-REML = 229.21   Scale est. = 1           n = 118
```

```
> gam.nums.all_env <- gam(richness ~s(records) + s(lat, long, bs = "sos")
+ s(Temperature) + s(Salinity) + s(Productivity) + s(Phosphate) + s(Nitrate)
) + s(Oxygen) + s(Chlorophyll), data = data, family = "nb", select = TRUE)
> gam.check(gam.nums.all_env)
```

```
Method: REML   Optimizer: outer newton
full convergence after 22 iterations.
Gradient range [-0.0001046477,1.670109e-05]
(score 226.8872 & scale 1).
Hessian positive definite, eigenvalue range [1.132851e-06,1.836405].
Model rank = 122 / 122
```

```
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(records)	9.00e+00	4.88e+00	1.01	0.57
s(lat, long)	4.90e+01	9.27e+00	1.00	0.54
s(Temperature)	9.00e+00	6.68e-01	1.13	0.97
s(Salinity)	9.00e+00	8.67e-01	1.05	0.70
s(Productivity)	9.00e+00	8.51e-06	1.05	0.76
s(Phosphate)	9.00e+00	1.38e-05	1.00	0.50
s(Nitrate)	9.00e+00	8.44e-06	1.09	0.87
s(Oxygen)	9.00e+00	4.40e-04	0.91	0.14
s(Chlorophyll)	9.00e+00	7.09e-06	1.09	0.84



```
> summary(gam.nums.all_env)
```

Family: Negative Binomial(6977704.81)
Link function: log

Formula:
richness ~ s(records) + s(lat, long, bs = "sos") + s(Temperature) +
 s(Salinity) + s(Productivity) + s(Phosphate) + s(Nitrate) +
 s(Oxygen) + s(Chlorophyll)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.95633	0.06036	15.84	<2e-16 ***

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

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(records)	4.877e+00	9	4951.046	1.18e-05 ***
s(lat, long)	9.266e+00	49	32.448	8.18e-06 ***
s(Temperature)	6.678e-01	9	1.943	0.05612 .
s(Salinity)	8.671e-01	9	6.065	0.00524 **
s(Productivity)	8.508e-06	9	0.000	0.83305
s(Phosphate)	1.377e-05	9	0.000	0.71924
s(Nitrate)	8.441e-06	9	0.000	0.98910
s(Oxygen)	4.401e-04	9	0.000	0.31572
s(Chlorophyll)	7.088e-06	9	0.000	0.80909

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

```

R-sq.(adj) = 0.677   Deviance explained = 70.9%
-REML = 226.89   Scale est. = 1           n = 118
> gam.nums.models <- list(gam.nums.intercept = gam.nums.intercept,
                           gam.nums.records = gam.nums.records,
                           gam.nums.latlong = gam.nums.latlong,
                           gam.nums.temp = gam.nums.temp,
                           gam.nums.salinity = gam.nums.salinity,
                           gam.nums.productivity = gam.nums.productivity,
                           gam.nums.phosphate = gam.nums.phosphate,
                           gam.nums.nitrate = gam.nums.nitrate,
                           gam.nums.oxygen = gam.nums.oxygen,
                           gam.nums.chlorophyll = gam.nums.chlorophyll,
                           gam.nums.all_env = gam.nums.all_env)
> gam.nums.aic.df <- data.frame(Model = names(gam.nums.models),
                                AIC = sapply(gam.nums.models, function(x) x$aic),
                                akaike.weights(sapply(gam.nums.models, function(x) x$aic)))

> gam.nums.aic.df <- gam.nums.aic.df[order(gam.nums.aic.df$AIC),]
> gam.nums.aic.df$Cumulative.Weight <- cumsum(gam.nums.aic.df$weights)

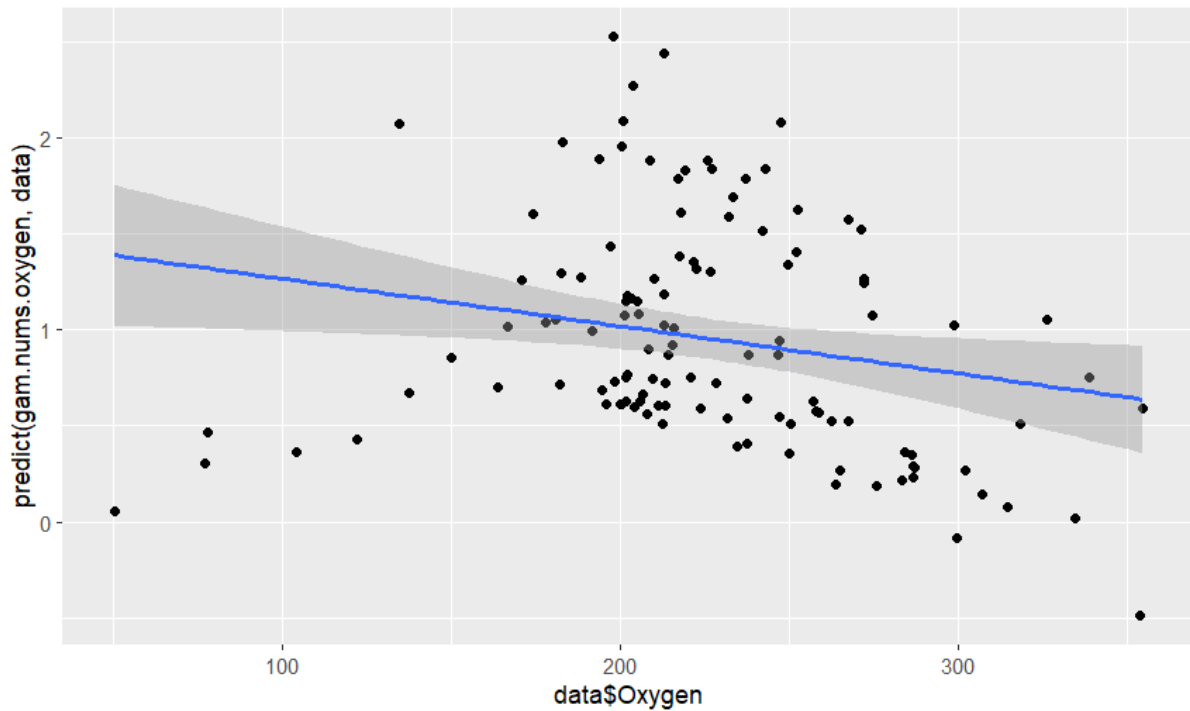
```

Model	AIC	deltaAIC	rel.LL	weights	Cumulative.Weight
gam.nums.salinity	429.103214983686	0	1	0.301522767734659	0.301522767734659
gam.nums.all_env	429.685238337545	0.582023353859029	0.74750694938117	0.225390364278302	0.5269131320129
gam.nums.oxygen	430.435486019208	1.33227103552173	0.513689892408814	0.154889198116425	0.6818023301293
gam.nums.productivity	432.424519397958	3.32130441427205	0.190015010533175	0.0572938518870932	0.7390961820164
gam.nums.chlorophyll	432.424677738356	3.32146275467051	0.189999967602418	0.0572893161009767	0.7963854981174
gam.nums.latlong	432.425719965032	3.32250498134624	0.189900981878737	0.057259469651606	0.8536449677690
gam.nums.nitrate	432.426144907457	3.32292992377069	0.189860637673058	0.0572473049550476	0.9108922727241
gam.nums.phosphate	432.668047990355	3.56483300666923	0.168231124607254	0.0507255143106934	0.9616177870348
gam.nums.temp	433.225718314645	4.12250333095898	0.127294539957768	0.0383822020055764	0.9999999890403
gam.nums.records	463.363491628354	34.2602766446676	3.63475759223444e-08	1.09596216925509e-08	1
gam.nums.intercept	517.744896850765	88.6416818670786	5.64551218697191e-20	1.70225045989552e-20	1

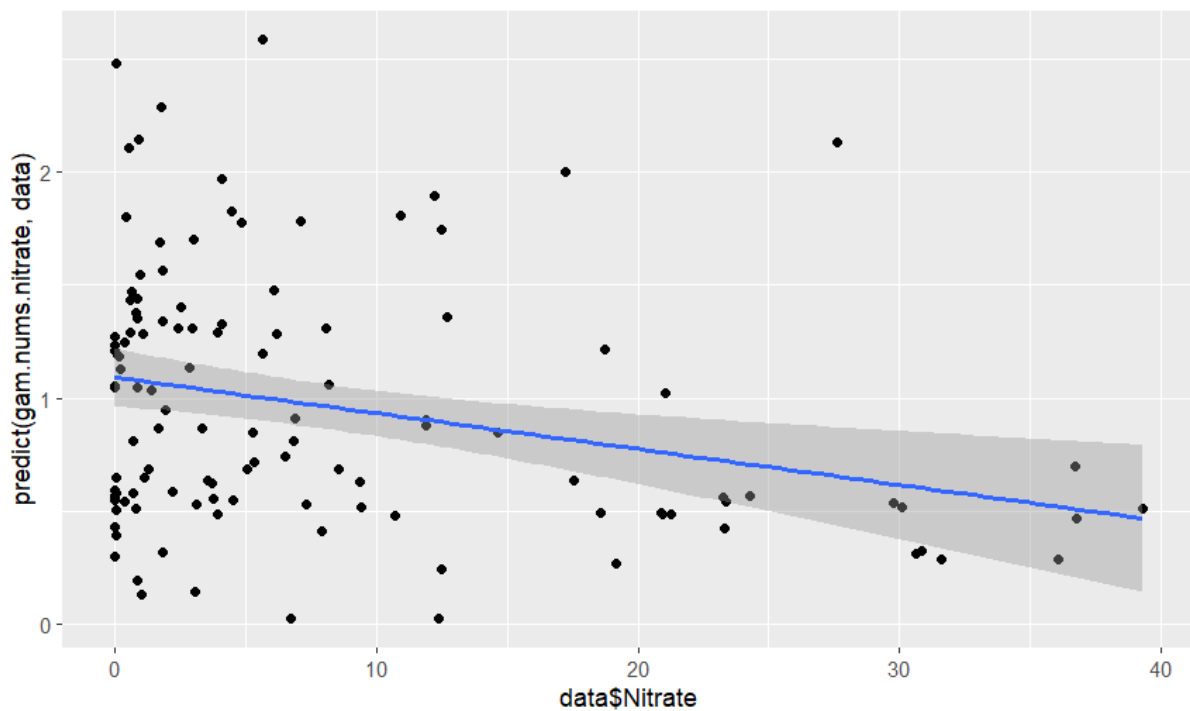
```

> gplot(data$Oxygen, predict(gam.nums.oxygen, data)) + geom_smooth(method =
"glm")
`geom_smooth()` using formula = 'y ~ x'

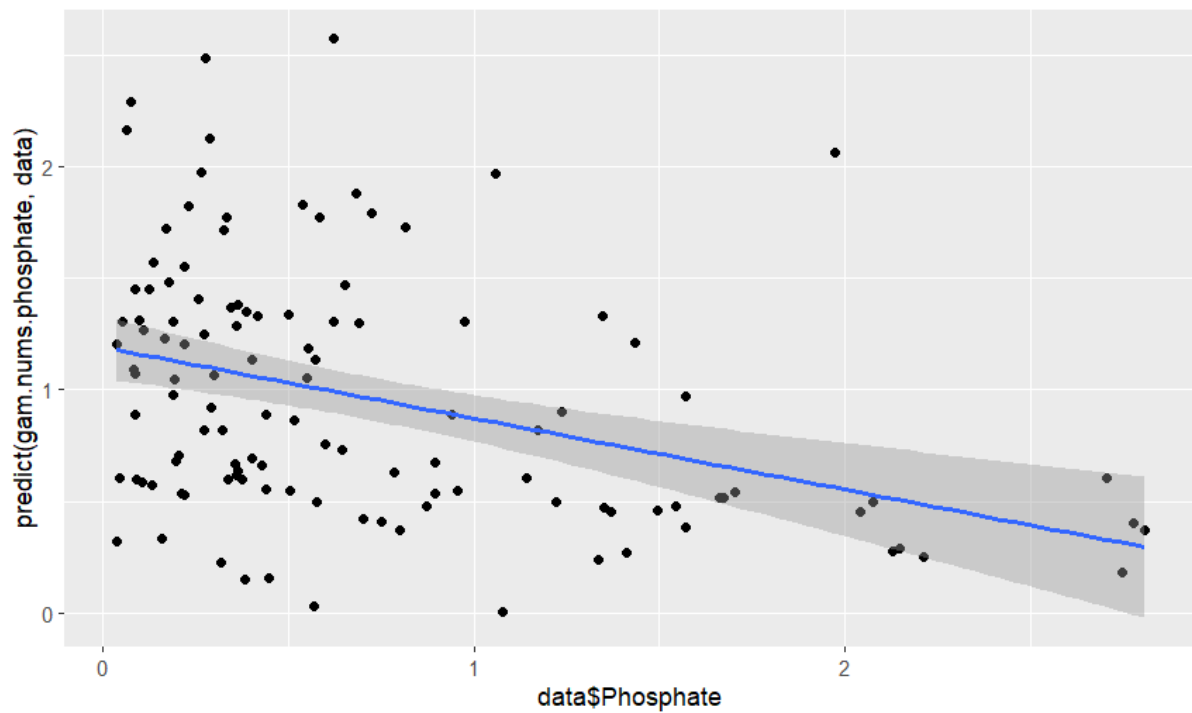
```

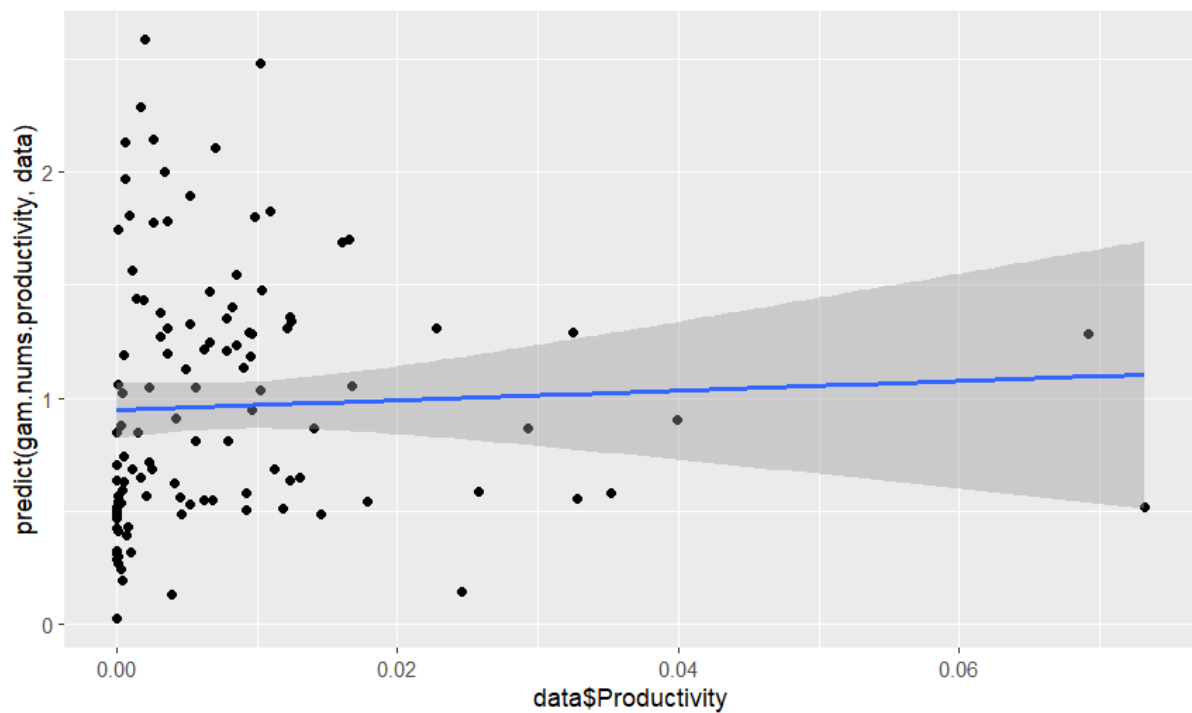
```
> qplot(data$Nitrate, predict(gam.nums.nitrate, data)) + geom_smooth(method
= "glm")
`geom_smooth()` using formula = 'y ~ x'
```



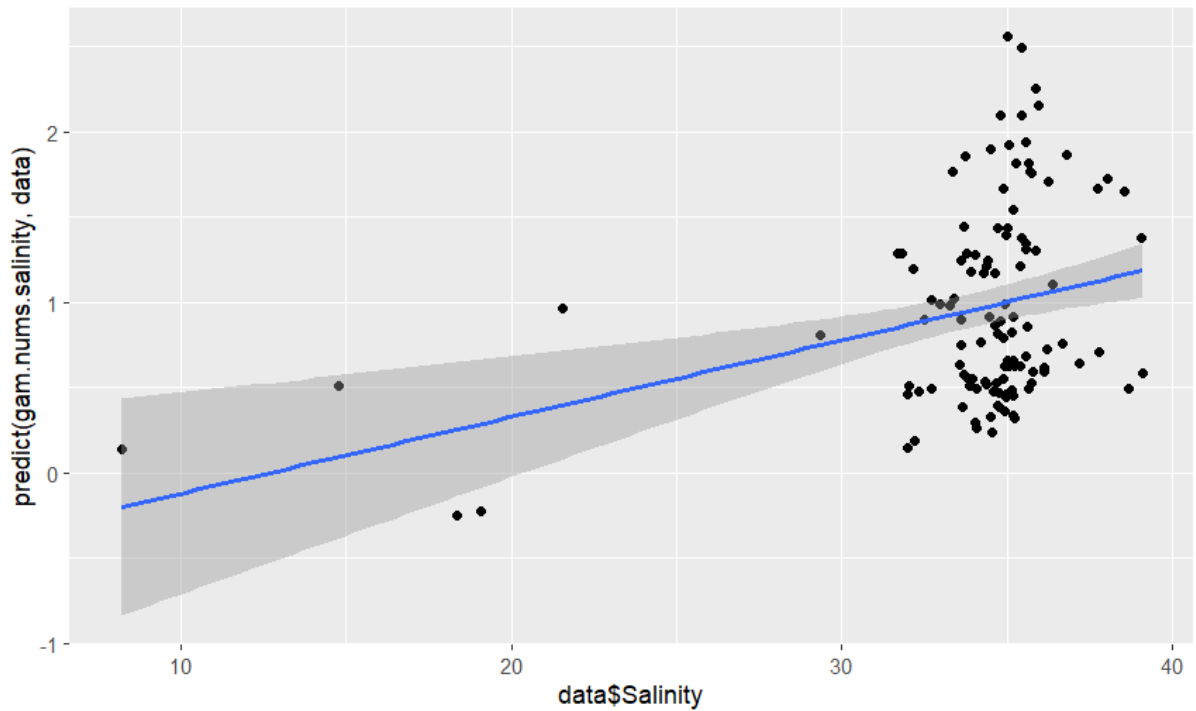
```
> qplot(data$Phosphate, predict(gam.nums.phosphate, data)) + geom_smooth(me
thod = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



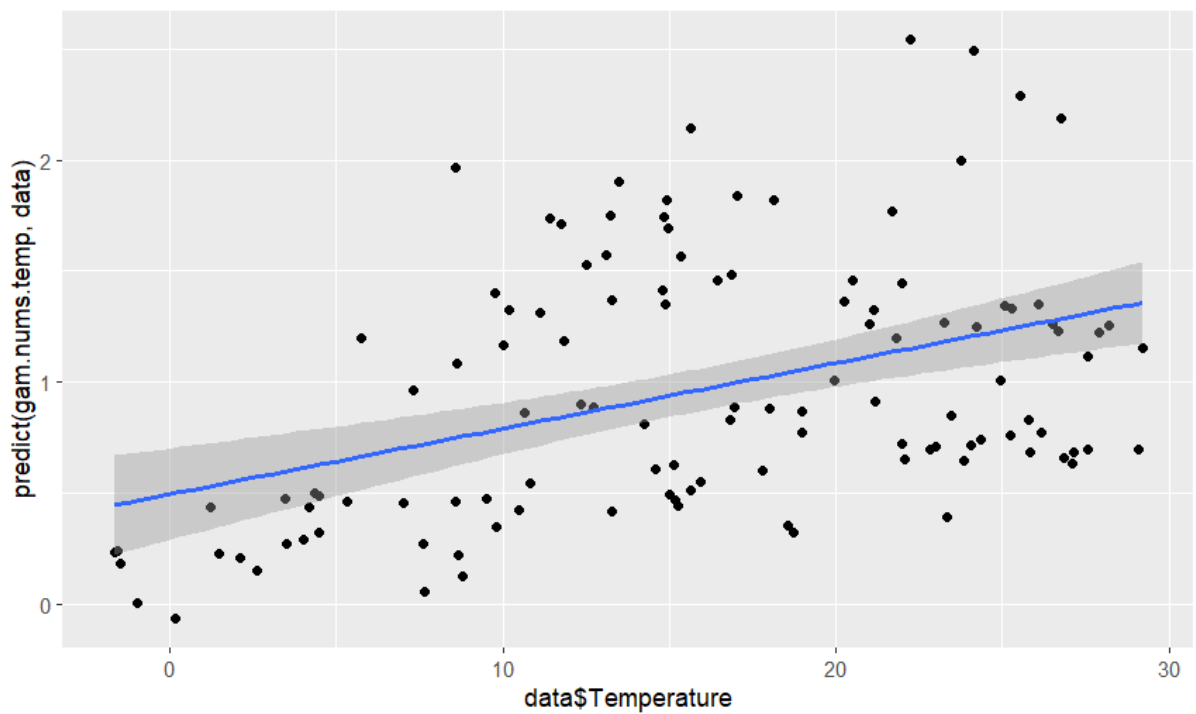
```
> qplot(data$Productivity, predict(gam.nums.productivity, data)) + geom_smooth(
  method = "glm")
# Aesthetics:
#   geom_smooth() using formula = 'y ~ x'
```



```
> qplot(data$Salinity, predict(gam.nums.salinity, data)) + geom_smooth(meth
  od = "glm")
```



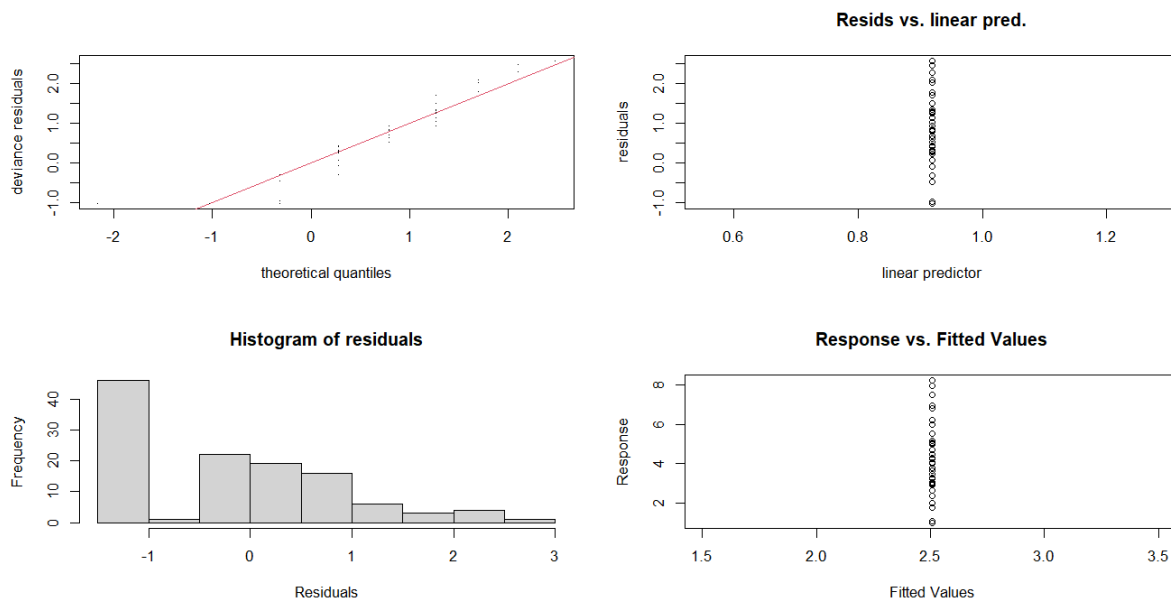
```
`geom_smooth()` using formula = 'y ~ x'
> qplot(data$Temperature, predict(gam.nums.temp, data)) + geom_smooth(metho
d = "glm")
```



```
> gam.ES15.intercept <- gam(rare ~ 1, data = data, family = "nb", method =
"RELM", select = TRUE)
> gam.check(gam.ES15.intercept)
```

Method: REML Optimizer: outer newton
full convergence after 6 iterations.

Gradient range [-0.0002657806,-0.0002657806]
(score 218.982 & scale 1).
Hessian positive definite, eigenvalue range [1.119268,1.119268].
Model rank = 1 / 1



```
> summary(gam.ES15.intercept)
```

Family: Negative Binomial(17.026)
Link function: log

Formula:
rare ~ 1

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.91931	0.06227	14.76	<2e-16 ***

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

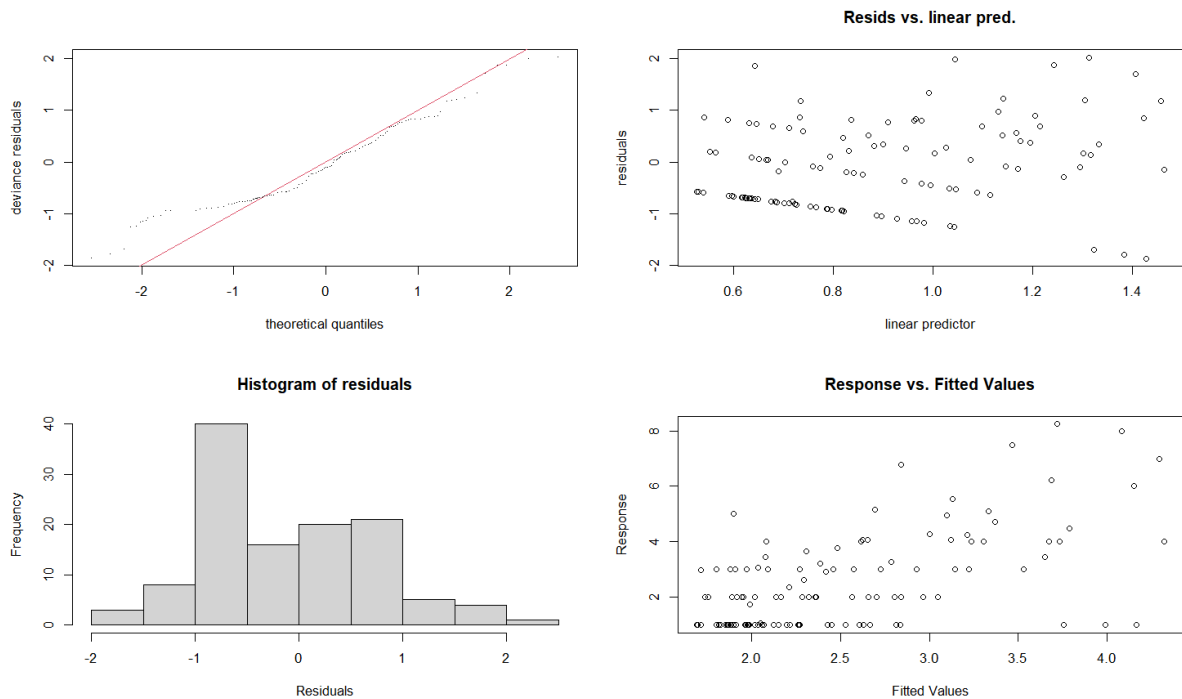
R-sq.(adj) = 0 Deviance explained = 1.69e-08%
-REML = 218.98 Scale est. = 1 n = 118

```
> gam.ES15.latlong <- gam(rare ~ s(lat, long, bs = "sos"), data = data, fa
mily = "nb", select = TRUE)
> gam.check(gam.ES15.latlong)
```

Method: REML Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-5.212414e-05,-2.817568e-05]
(score 214.3255 & scale 1).
Hessian positive definite, eigenvalue range [5.208304e-05,2.338949].
Model rank = 50 / 50

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(lat,long)	49.0	11.2	0.99	0.5



```
> summary(gam.ES15.latlong)
```

```
Family: Negative Binomial(303619.374)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos")
```

```
Parametric coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.88595    0.05952   14.89  <2e-16 ***
```

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

```
Approximate significance of smooth terms:
```

```
              edf Ref.df Chi.sq  p-value
s(lat,long)  11.22    49  32.02 7.17e-05 ***
```

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

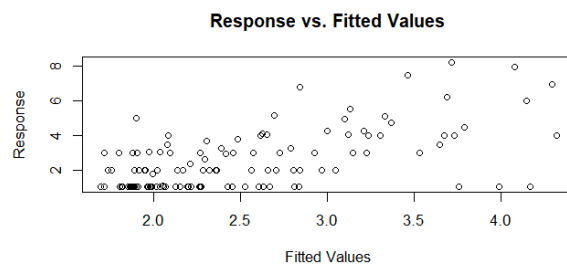
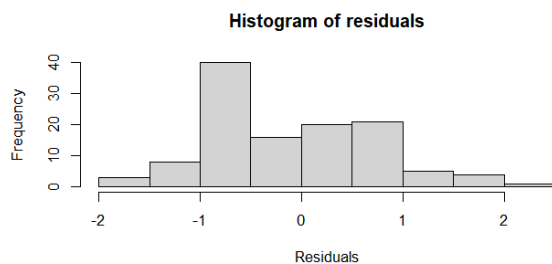
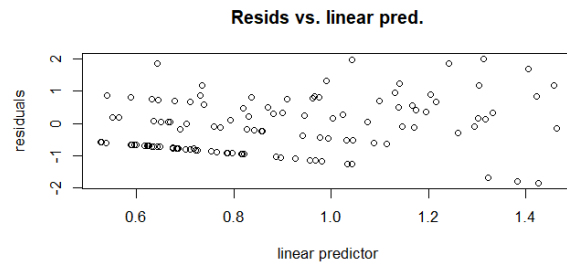
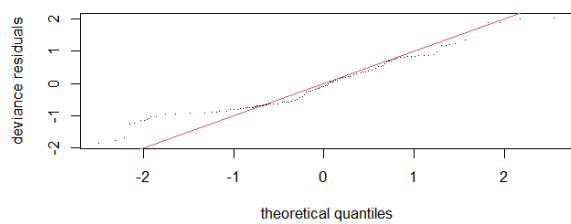
```
R-sq.(adj) = 0.257   Deviance explained = 34.1%
-REML = 214.33   Scale est. = 1           n = 118
```

```
> gam.ES15.chlorophyll <- gam(rare ~ s(lat, long, bs = "sos") + s(Chlorophyll), data = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.chlorophyll)
```

```
Method: REML   Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-4.789816e-05,-4.49323e-07]
(score 214.3255 & scale 1).
Hessian positive definite, eigenvalue range [4.493226e-07,2.33895].
Model rank = 59 / 59
```

```
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(lat,long)   4.90e+01 1.12e+01  0.99  0.40
s(Chlorophyll) 9.00e+00 3.53e-06  0.99  0.46
```



```
> summary(gam.ES15.chlorophyll)
```

```
Family: Negative Binomial(330029.266)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos") + s(Chlorophyll)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.88595    0.05952   14.89  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

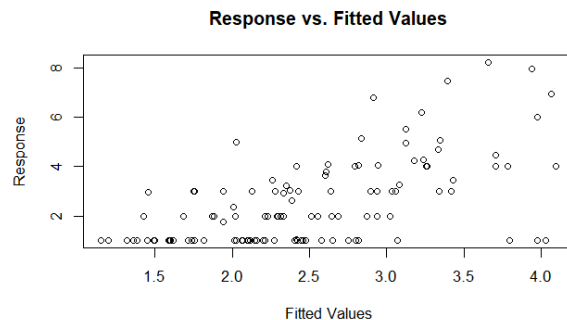
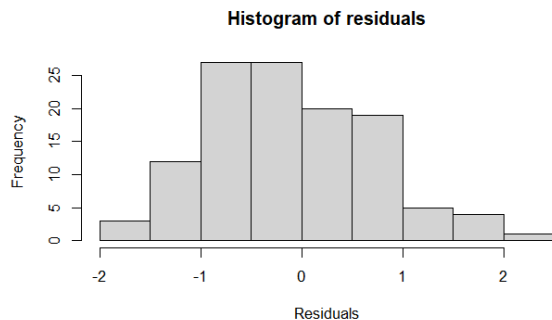
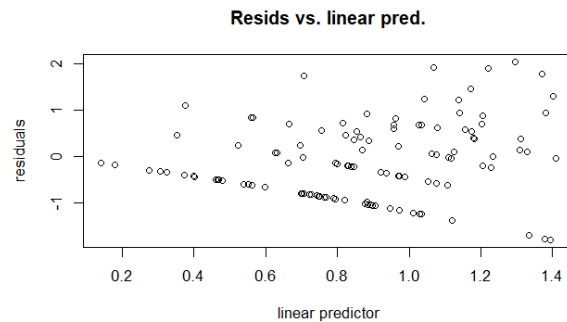
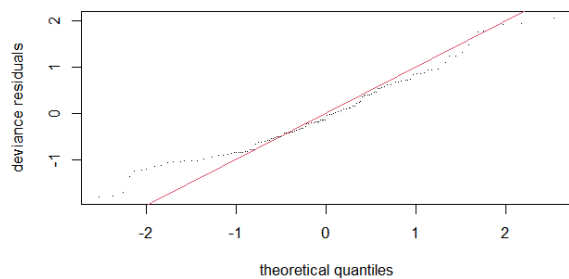
```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq  p-value
s(lat,long)   1.122e+01   49  32.02 7.17e-05 ***
s(Chlorophyll) 3.535e-06    9   0.00  0.602
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.257   Deviance explained = 34.1%
-REML = 214.33   Scale est. = 1           n = 118
> gam.ES15.oxygen <- gam(rare ~ s(lat, long, bs = "sos") + s(Oxygen), data
= data, family = "nb", select = TRUE)
> gam.check(gam.ES15.oxygen)
```

```
Method: REML   Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-9.971601e-05,1.922623e-06]
(score 212.5824 & scale 1).
Hessian positive definite, eigenvalue range [2.817139e-06,1.480182].
Model rank = 59 / 59
```

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(lat,long)	49.00	8.28	1.01	0.57
s(Oxygen)	9.00	2.02	0.99	0.52



```
> summary(gam.ES15.oxygen)
```

```
Family: Negative Binomial(162847.213)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos") + s(Oxygen)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.87889    0.06003   14.64  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(lat,long)  8.285     49 18.567  0.00469 **
s(Oxygen)    2.015      9  7.519  0.00541 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

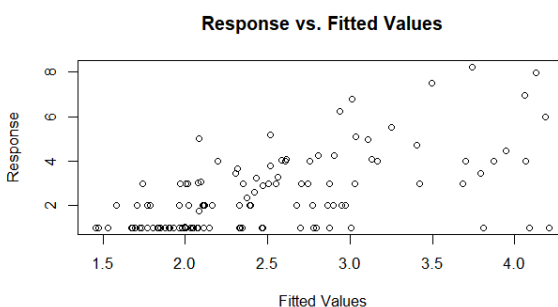
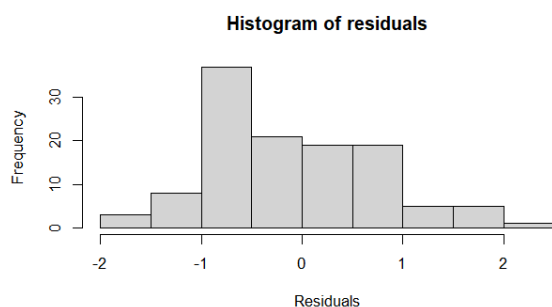
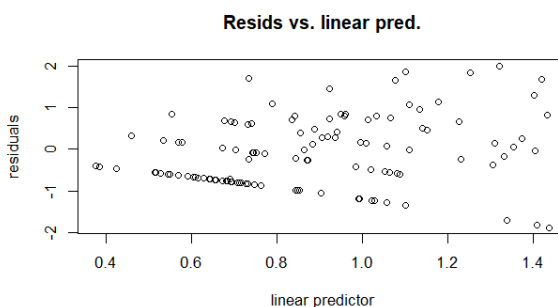
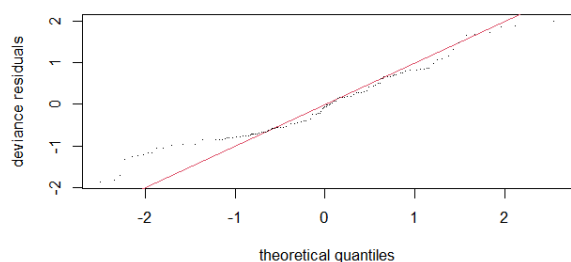
```
R-sq.(adj) = 0.257   Deviance explained = 34.7%
-REML = 212.58   Scale est. = 1           n = 118
```

```
> gam.ES15.nitrate <- gam(rare ~ s(lat, long, bs = "sos") + s(Nitrate), da
ta = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.nitrate)
```

```
Method: REML   Optimizer: outer newton
full convergence after 15 iterations.
Gradient range [-0.0001240115,-8.831426e-06]
(score 213.7073 & scale 1).
Hessian positive definite, eigenvalue range [8.660147e-06,1.818596].
Model rank = 59 / 59
```

```
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(lat,long)	49.000	9.851	0.99	0.44
s(Nitrate)	9.000	0.732	1.06	0.85



```
> summary(gam.ES15.nitrate)
```

```
Family: Negative Binomial(1546714.636)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos") + s(Nitrate)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.8850      0.0596   14.85  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq  p-value
s(lat,long)  9.8515     49 25.342 0.000583 ***
s(Nitrate)   0.7321      9  2.579 0.049997 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

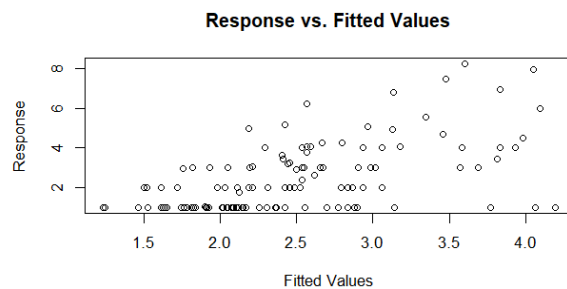
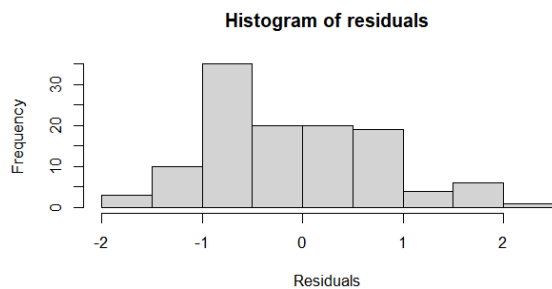
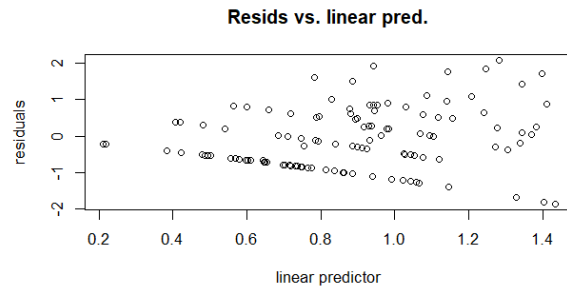
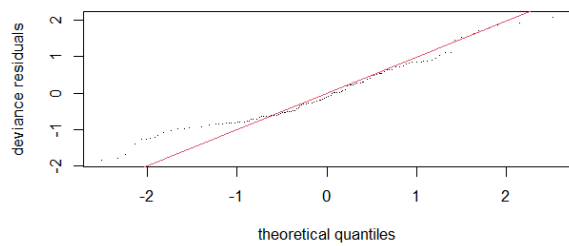
```
R-sq.(adj) = 0.242   Deviance explained = 33.4%
-REML = 213.71   Scale est. = 1           n = 118
```

```
> gam.ES15.phosphate <- gam(rare ~ s(lat, long, bs = "sos") + s(Phosphate)
, data = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.phosphate)
```

```
Method: REML   Optimizer: outer newton
full convergence after 16 iterations.
Gradient range [-5.195265e-05,-4.82767e-06]
(score 212.7507 & scale 1).
Hessian positive definite, eigenvalue range [4.731737e-06,1.326941].
Model rank = 59 / 59
```

```
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(lat,long)	49.000	8.460	0.98	0.44
s(Phosphate)	9.000	0.849	0.98	0.48



```
> summary(gam.ES15.phosphate)
```

Family: Negative Binomial(2592855.529)
Link function: log

Formula:
rare ~ s(lat, long, bs = "sos") + s(Phosphate)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.88277	0.05985	14.75	<2e-16 ***

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

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(lat,long)	8.4603	49	20.58	0.00206 **
s(Phosphate)	0.8488	9	5.35	0.00882 **

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

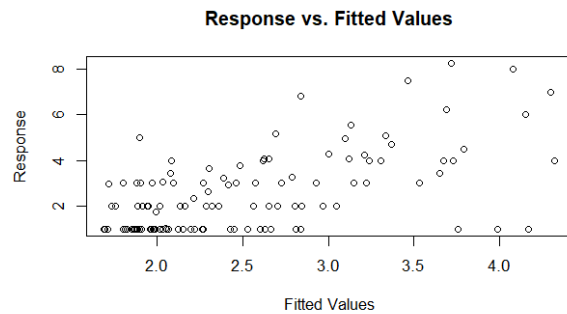
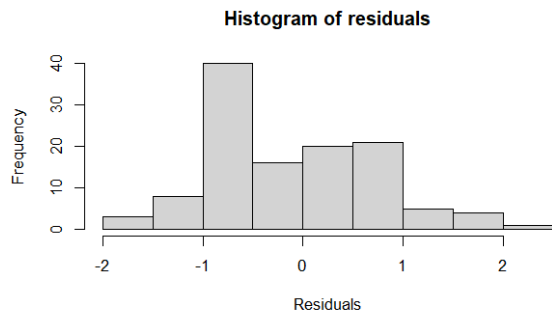
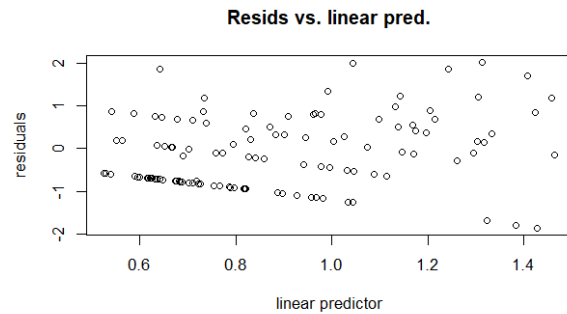
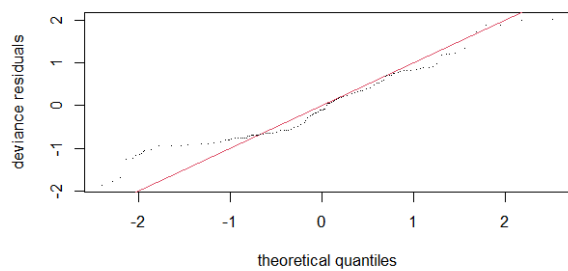
R-sq.(adj) = 0.232 Deviance explained = 32.1%
-REML = 212.75 Scale est. = 1 n = 118

```
> gam.ES15.productivity <- gam(rare ~ s(lat, long, bs = "sos") + s(Productivity), data = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.productivity)
```

Method: REML Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-5.027033e-05,-4.33051e-07]
(score 214.3255 & scale 1).
Hessian positive definite, eigenvalue range [4.33051e-07,2.338949].
Model rank = 59 / 59

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(lat,long)	4.90e+01	1.12e+01	0.99	0.43
s(Productivity)	9.00e+00	3.53e-06	0.96	0.34



```
> summary(gam.ES15.productivity)
```

```
Family: Negative Binomial(314780.572)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos") + s(Productivity)
```

```
Parametric coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.88595	0.05952	14.89	<2e-16 ***

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

```
Approximate significance of smooth terms:
```

	edf	Ref.df	Chi.sq	p-value
s(lat,long)	1.122e+01	49	32.02	7.17e-05 ***
s(Productivity)	3.526e-06	9	0.00	0.566

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

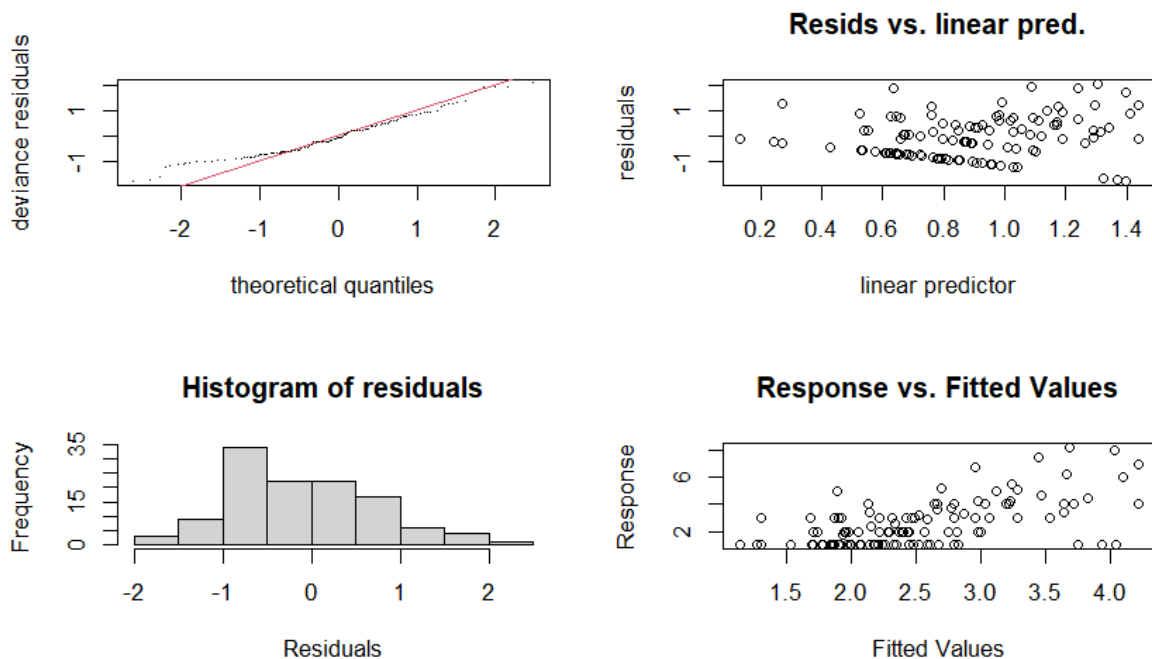
```
R-sq.(adj) = 0.257   Deviance explained = 34.1%
-REML = 214.33   Scale est. = 1           n = 118
```

```
> gam.ES15.salinity <- gam(rare ~ s(lat, long, bs = "sos") + s(salinity),
data = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.salinity)
```

```
Method: REML   Optimizer: outer newton
full convergence after 12 iterations.
Gradient range [-7.799358e-05,-1.405975e-06]
(score 213.5497 & scale 1).
Hessian positive definite, eigenvalue range [7.797331e-05,2.14952].
Model rank = 59 / 59
```

```
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(lat,long)	49.000	10.359	0.99	0.47
s(Salinity)	9.000	0.918	1.07	0.83



```
> summary(gam.ES15.salinity)
```

```
Family: Negative Binomial(230453.494)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos") + s(Salinity)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.88218    0.05979   14.76  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq  p-value
s(lat,long) 10.359     49 28.384 0.000177 ***
s(Salinity)  0.918      9  2.989 0.043851 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

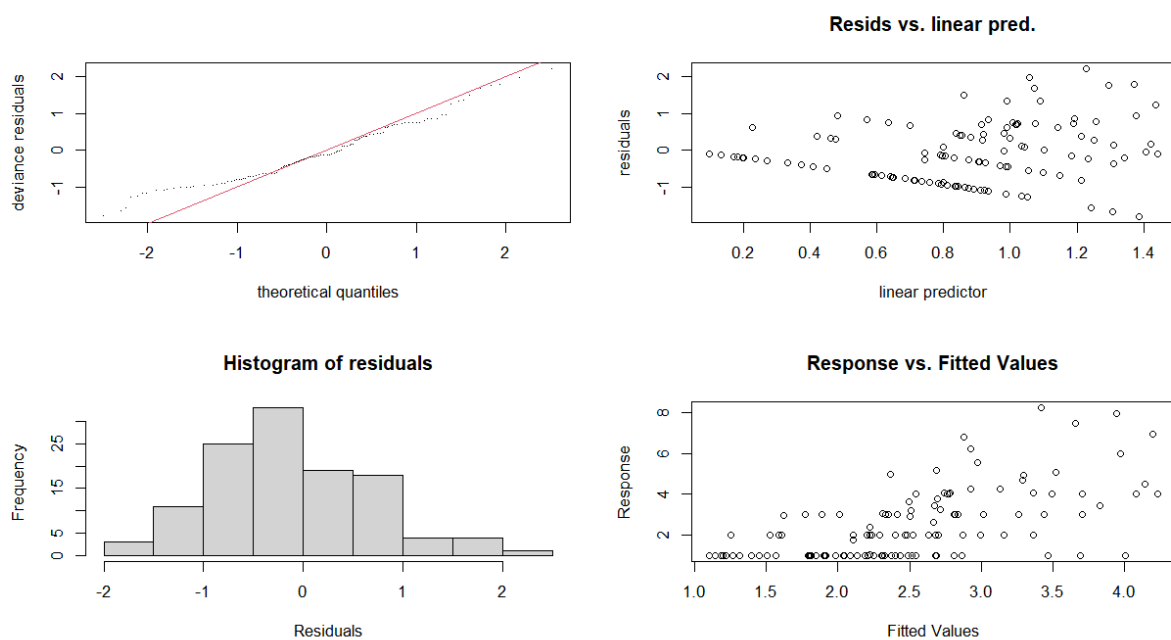
```
R-sq.(adj) = 0.268   Deviance explained = 35.3%
-REML = 213.55   Scale est. = 1           n = 118
```

```
> gam.ES15.temp <- gam(rare ~ s(lat, long, bs = "sos") + s(Temperature), d
ata = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.temp)
```

```
Method: REML   Optimizer: outer newton
full convergence after 13 iterations.
Gradient range [-4.740982e-05,-6.58478e-07]
(score 211.0202 & scale 1).
Hessian positive definite, eigenvalue range [6.584772e-07,1.448353].
Model rank = 59 / 59
```

```
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(lat,long)	49.00	8.19	1.01	0.53
s(Temperature)	9.00	2.07	1.14	0.92



```
> summary(gam.ES15.temp)
```

```
Family: Negative Binomial(420406.477)
Link function: log
```

```
Formula:
rare ~ s(lat, long, bs = "sos") + s(Temperature)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.87192    0.06052   14.41  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Ref.df Chi.sq  p-value
s(lat,long)    8.191    49  18.08 0.005615 **
s(Temperature)  2.074     9  10.82 0.000723 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.282   Deviance explained = 37.6%
-REML = 211.02   Scale est. = 1           n = 118
```

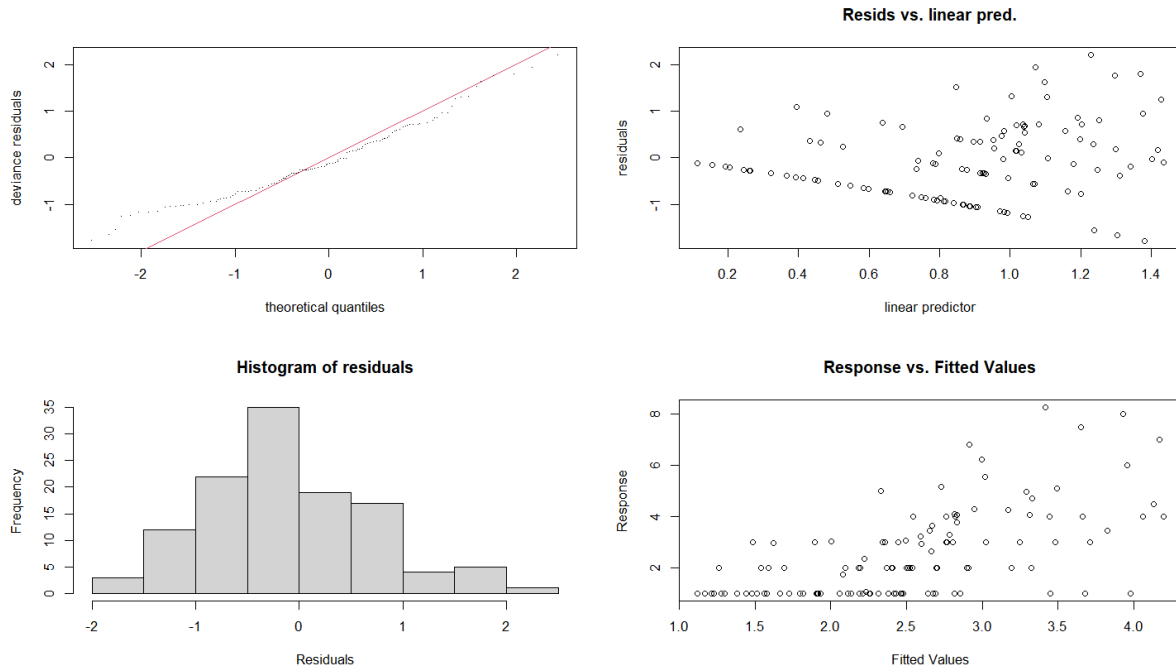
```
> gam.ES15.all_env <- gam(rare ~ s(lat, long, bs = "sos") + s(Temperature)
+ s(Salinity) + s(Productivity) + s(Phosphate) + s(Nitrate) + s(Oxygen) + s
(Chlorophyll), data = data, family = "nb", select = TRUE)
> gam.check(gam.ES15.all_env)
```

```
Method: REML   Optimizer: outer newton
full convergence after 14 iterations.
Gradient range [-1.339882e-05,-1.624943e-07]
(score 210.853 & scale 1).
Hessian positive definite, eigenvalue range [1.624944e-07,1.43296].
Model rank = 113 / 113
```

```
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(lat,long)    4.90e+01 8.09e+00   1.01   0.61
s(Temperature)  9.00e+00 2.03e+00   1.12   0.89
```

s(Salinity)	9.00e+00	5.40e-01	1.10	0.90
s(Productivity)	9.00e+00	2.24e-05	1.00	0.53
s(Phosphate)	9.00e+00	5.13e-05	0.93	0.28
s(Nitrate)	9.00e+00	2.02e-05	1.06	0.73
s(Oxygen)	9.00e+00	3.58e-05	1.02	0.58
s(Chlorophyll)	9.00e+00	2.80e-05	1.05	0.70



```
> summary(gam.ES15.all_env)
```

Family: Negative Binomial(1568534.957)
Link function: log

Formula:
rare ~ s(lat, long, bs = "sos") + s(Temperature) + s(Salinity) +
s(Productivity) + s(Phosphate) + s(Nitrate) + s(Oxygen) +
s(Chlorophyll)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.87093	0.06055	14.38	<2e-16 ***

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

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(lat,long)	8.089e+00	49	17.688	0.00607 **
s(Temperature)	2.033e+00	9	9.618	0.00151 **
s(Salinity)	5.396e-01	9	1.089	0.13796
s(Productivity)	2.238e-05	9	0.000	0.88890
s(Phosphate)	5.125e-05	9	0.000	0.42456
s(Nitrate)	2.024e-05	9	0.000	0.90441
s(Oxygen)	3.584e-05	9	0.000	0.57043
s(Chlorophyll)	2.805e-05	9	0.000	0.85828

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

R-sq.(adj) = 0.289 Deviance explained = 38.6%
-REML = 210.85 Scale est. = 1 n = 118

```
> gam.ES15.models <- list(gam.ES15.intercept = gam.ES15.intercept,  
gam.ES15.latlong = gam.ES15.latlong,  
gam.ES15.temp = gam.ES15.temp,  
gam.ES15.salinity = gam.ES15.salinity,
```

```

gam.ES15.productivity = gam.ES15.productivity,
gam.ES15.phosphate = gam.ES15.phosphate,
gam.ES15.nitrate = gam.ES15.nitrate,
gam.ES15.oxygen = gam.ES15.oxygen,
gam.ES15.chlorophyll = gam.ES15.chlorophyll,
gam.ES15.all_env = gam.ES15.all_env)
> gam.ES15.aic.df <- data.frame(Model = names(gam.ES15.models),
                                AIC = sapply(gam.ES15.models, function(x) x$aic),
                                akaike.weights(sapply(gam.ES15.models, function(x) x$aic)))
> gam.ES15.aic.df <- gam.ES15.aic.df[order(gam.ES15.aic.df$AIC),]
> gam.ES15.aic.df$Cumulative.Weight <- cumsum(gam.ES15.aic.df$weights)

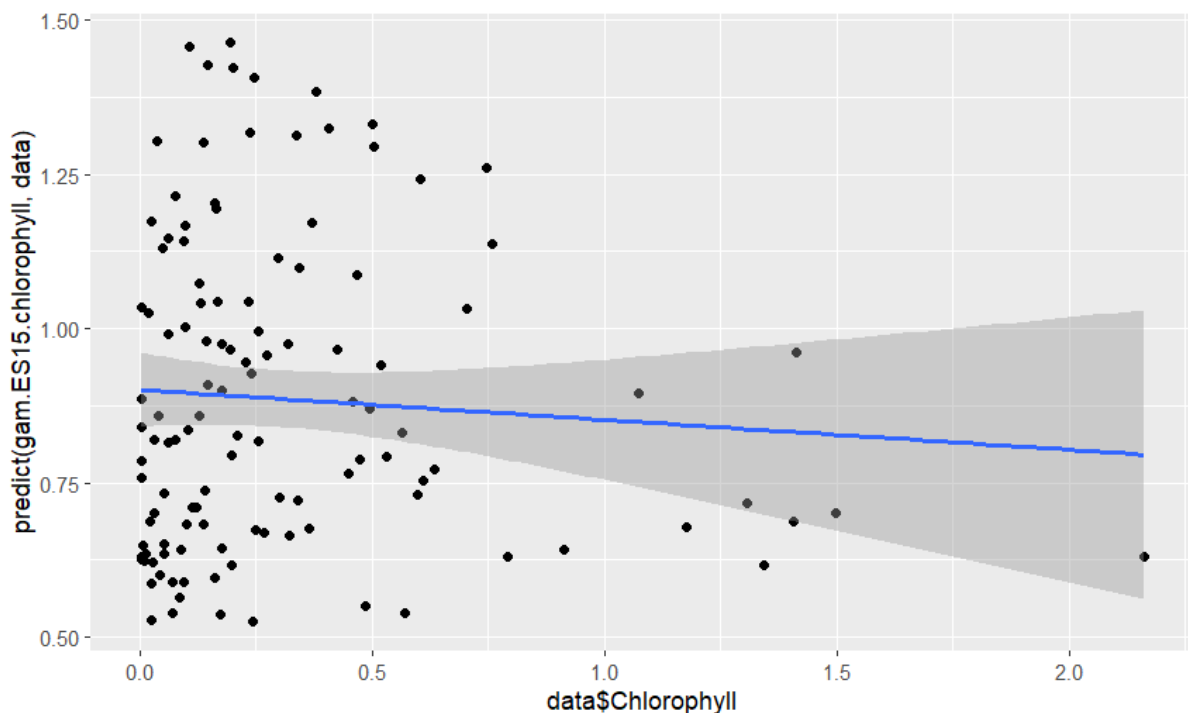
```

Model	AIC	deltaAIC	rel.LL	weights	Cumulative.Weight
gam.ES15.all_env	411.002224914798	0	1	0.445212599493378	0.445212599493378
gam.ES15.temp	411.411594388329	0.40936947353174	0.814904185247584	0.36280561065211	0.808018210145488
gam.ES15.oxygen	415.008329163882	4.00610424908467	0.134922852811878	0.0600693540314386	0.868087564176927
gam.ES15.salinity	416.248916549444	5.24669163464608	0.0725596847893213	0.0323044858834739	0.900392050060401
gam.ES15.phosphate	416.360315676546	5.3580907617482	0.0686286371032305	0.0305543339244169	0.930946383984817
gam.ES15.nitrate	417.238173377111	6.23594846231339	0.0442467113016546	0.0196991933576427	0.95064557734246
gam.ES15.chlorophyll	417.59854971625	6.59632480145285	0.0369510061940321	0.016451053521541	0.967096630864001
gam.ES15.productivity	417.598559841532	6.59633492673441	0.0369508191248348	0.0164509702359773	0.983547601099978
gam.ES15.latlong	417.598564426182	6.59633951138471	0.0369507344216399	0.0164509325250478	0.999998533625022
gam.ES15.intercept	436.249252610872	25.2470276960742	3.2936511119754e-06	1.46637497338682e-06	1

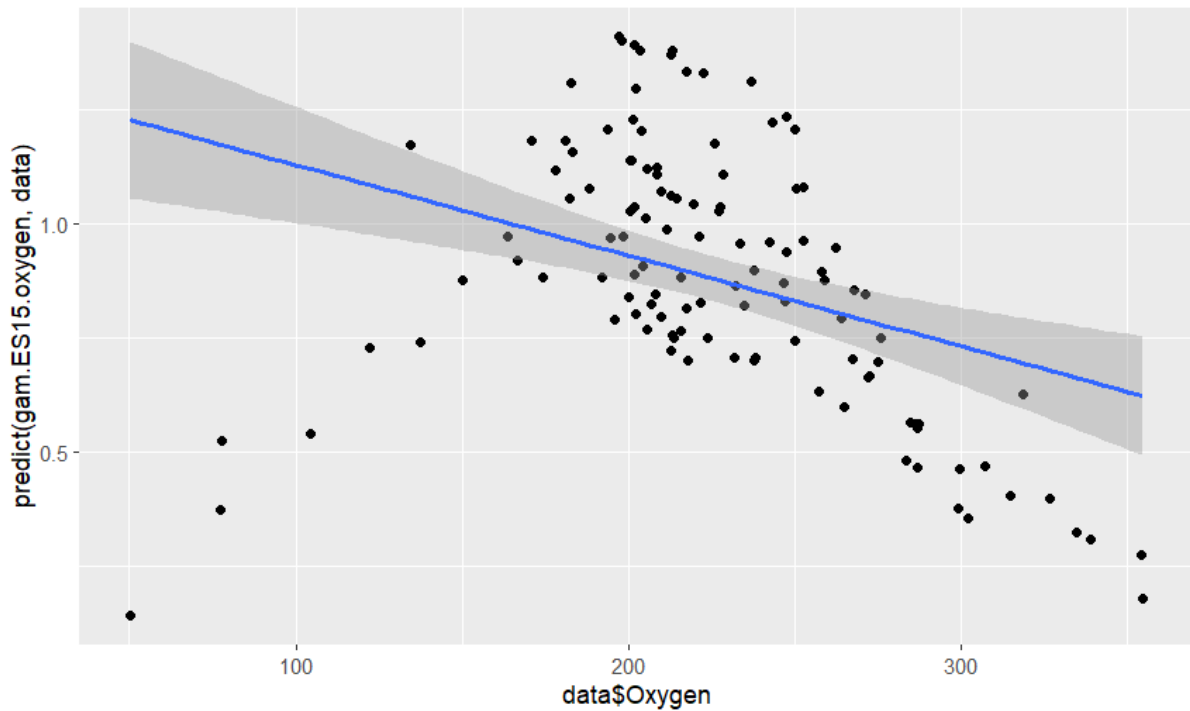
```

> qplot(data$Chlorophyll, predict(gam.ES15.chlorophyll, data)) + geom_smooth(
  h(method = "glm")
  `geom_smooth()` using formula = 'y ~ x'

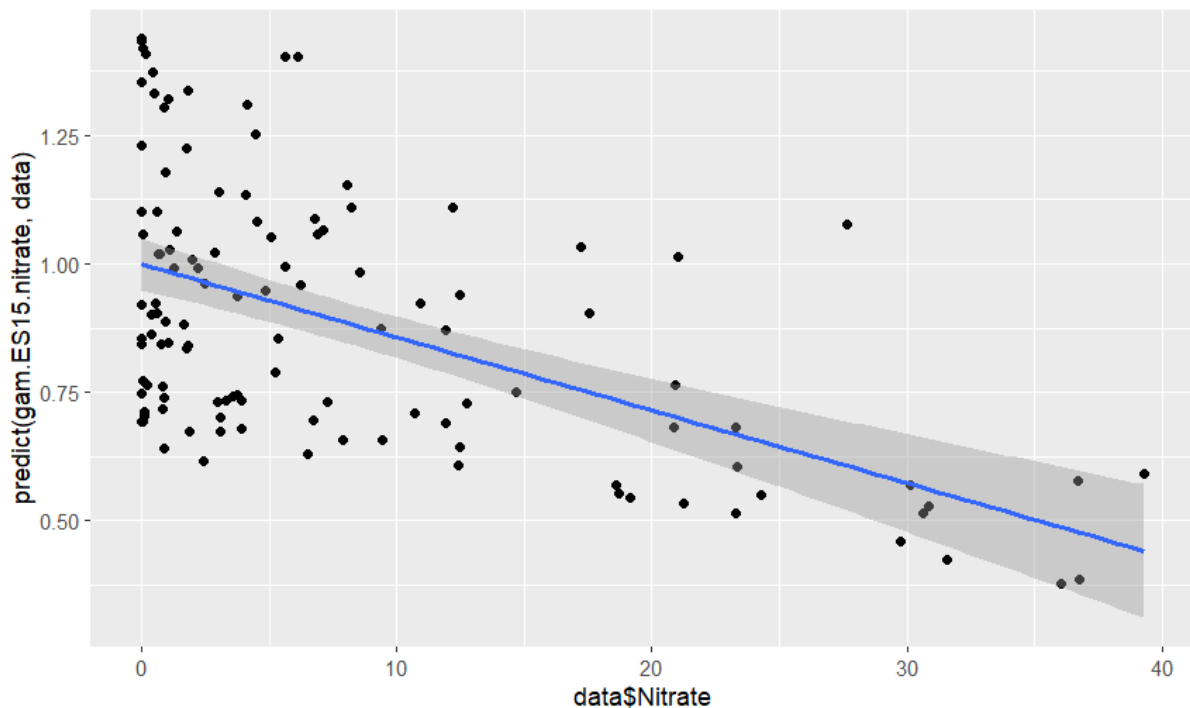
```



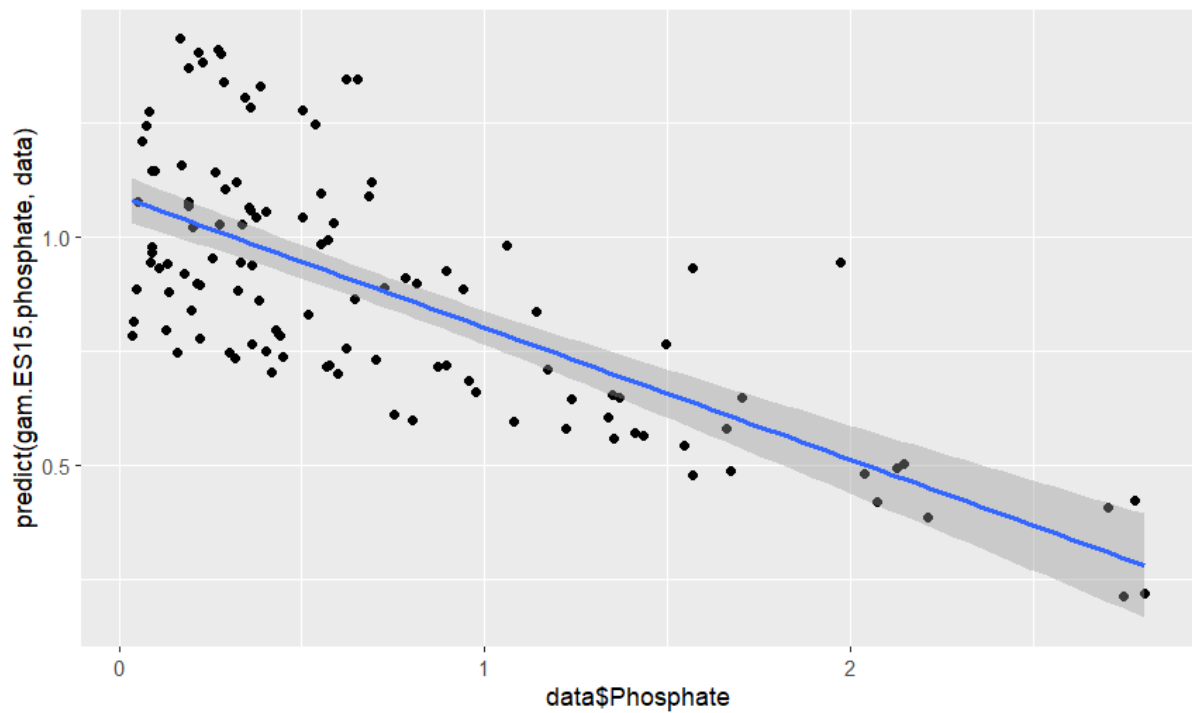
```
> qplot(data$Oxygen, predict(gam.ES15.oxygen, data)) + geom_smooth(method =
"glm")
`geom_smooth()` using formula = 'y ~ x'
```



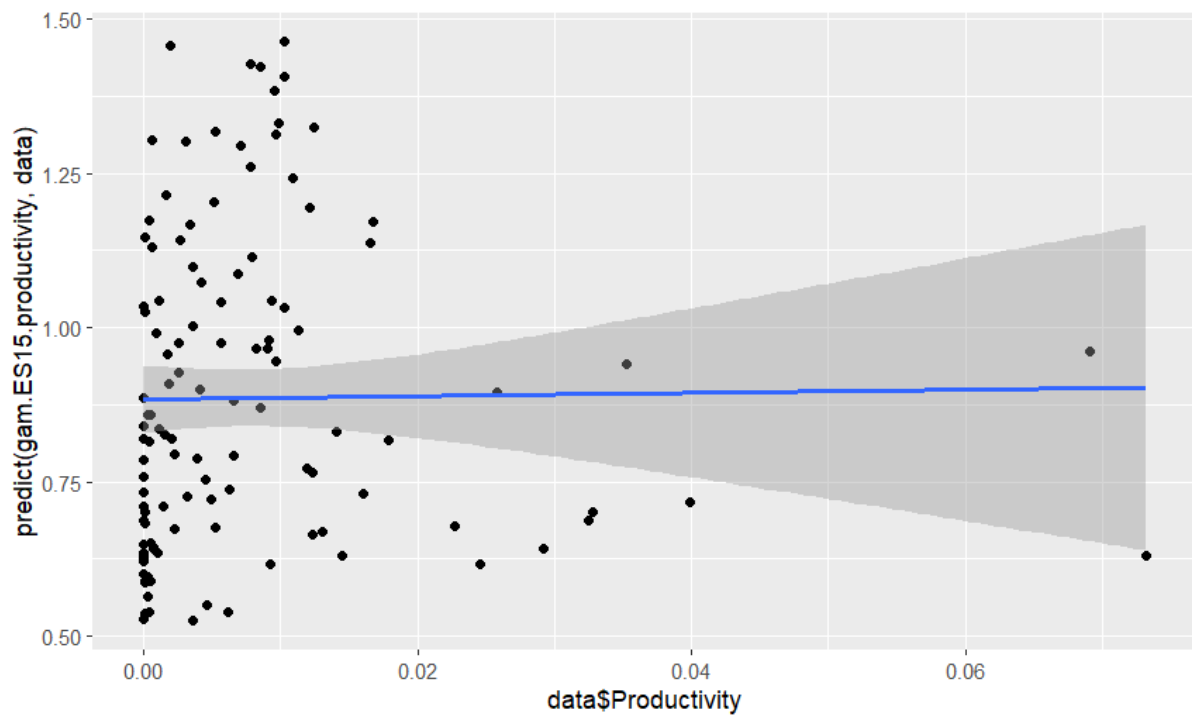
```
> qplot(data$Nitrate, predict(gam.ES15.nitrate, data)) + geom_smooth(method
= "glm")
`geom_smooth()` using formula = 'y ~ x'
```



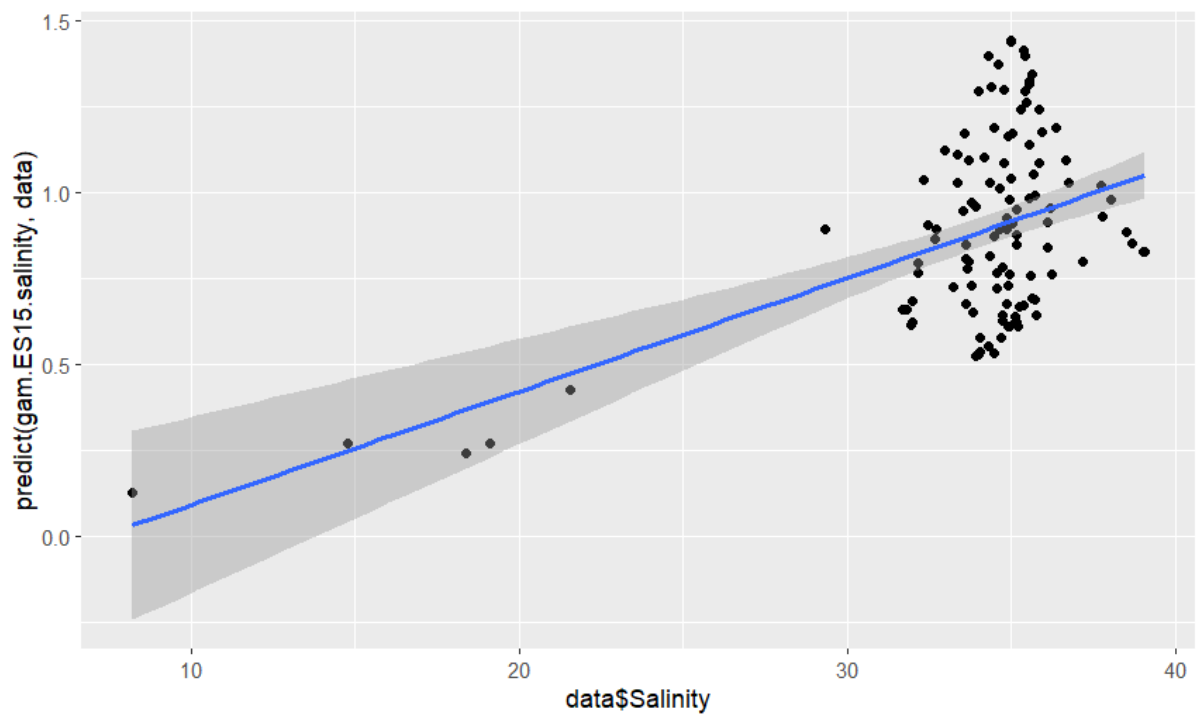
```
> qplot(data$Phosphate, predict(gam.ES15.phosphate, data)) + geom_smooth(me
thod = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



```
> qplot(data$Productivity, predict(gam.ES15.productivity, data)) + geom_smooth(
  method = "glm")
`geom_smooth()` using formula = 'y ~ x'
```



```
> qplot(data$Salinity, predict(gam.ES15.salinity, data)) + geom_smooth(meth
  od = "glm")
`geom_smooth()` using formula = 'y ~ x'
```

```
> qplot(data$Temperature, predict(gam.ES15.temp, data)) + geom_smooth(metho  
d = "glm")  
`geom_smooth()` using formula = 'y ~ x'
```

