

Data 583 Project

Yizhe Qu

April 9, 2019

```
source("C:/Users/yizhe/Desktop/data_583/project/seismictimingsfull.R")
seis <- seismictimingsfull
head(seis)
```

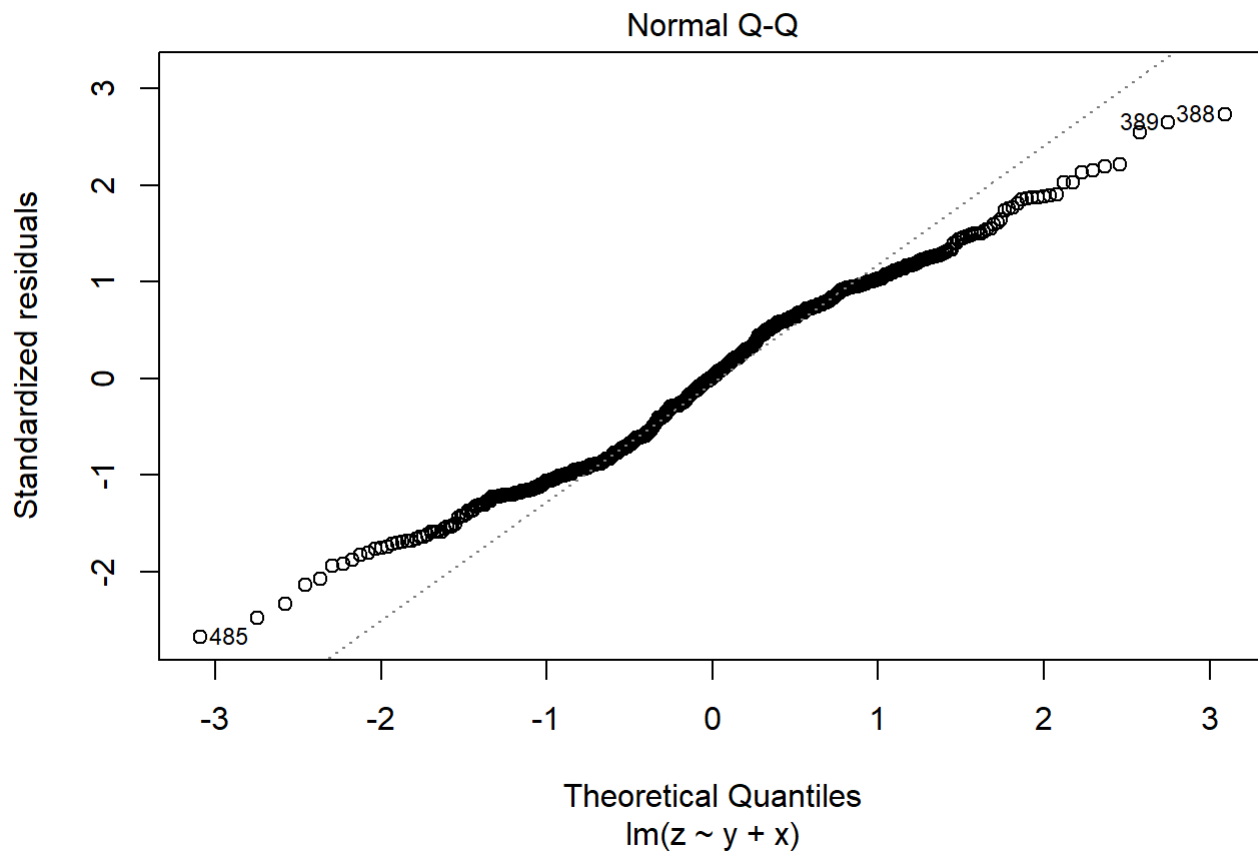
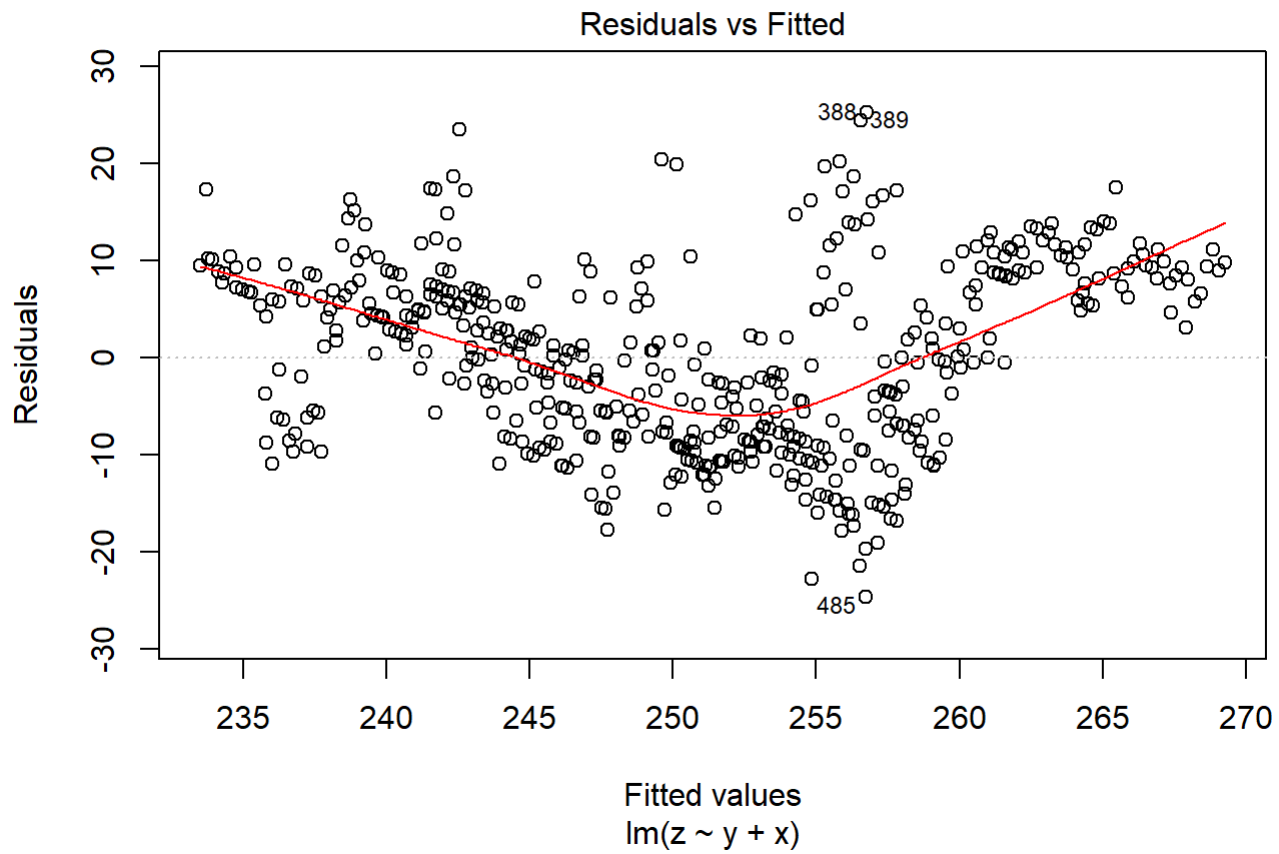
```
##      x y  z
## 1 0.125 8 261
## 2 0.250 8 263
## 3 0.375 8 266
## 4 0.500 8 259
## 5 0.625 8 258
## 6 0.750 8 253
```

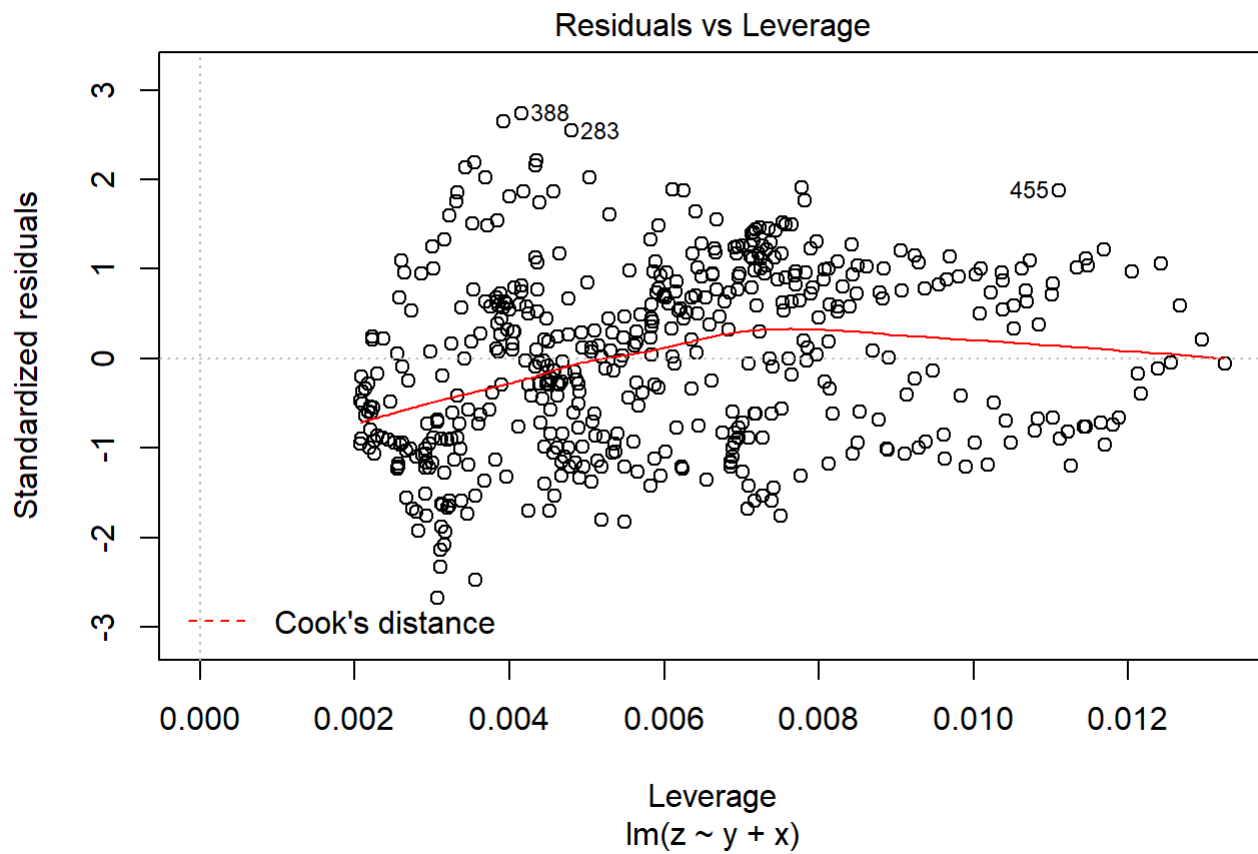
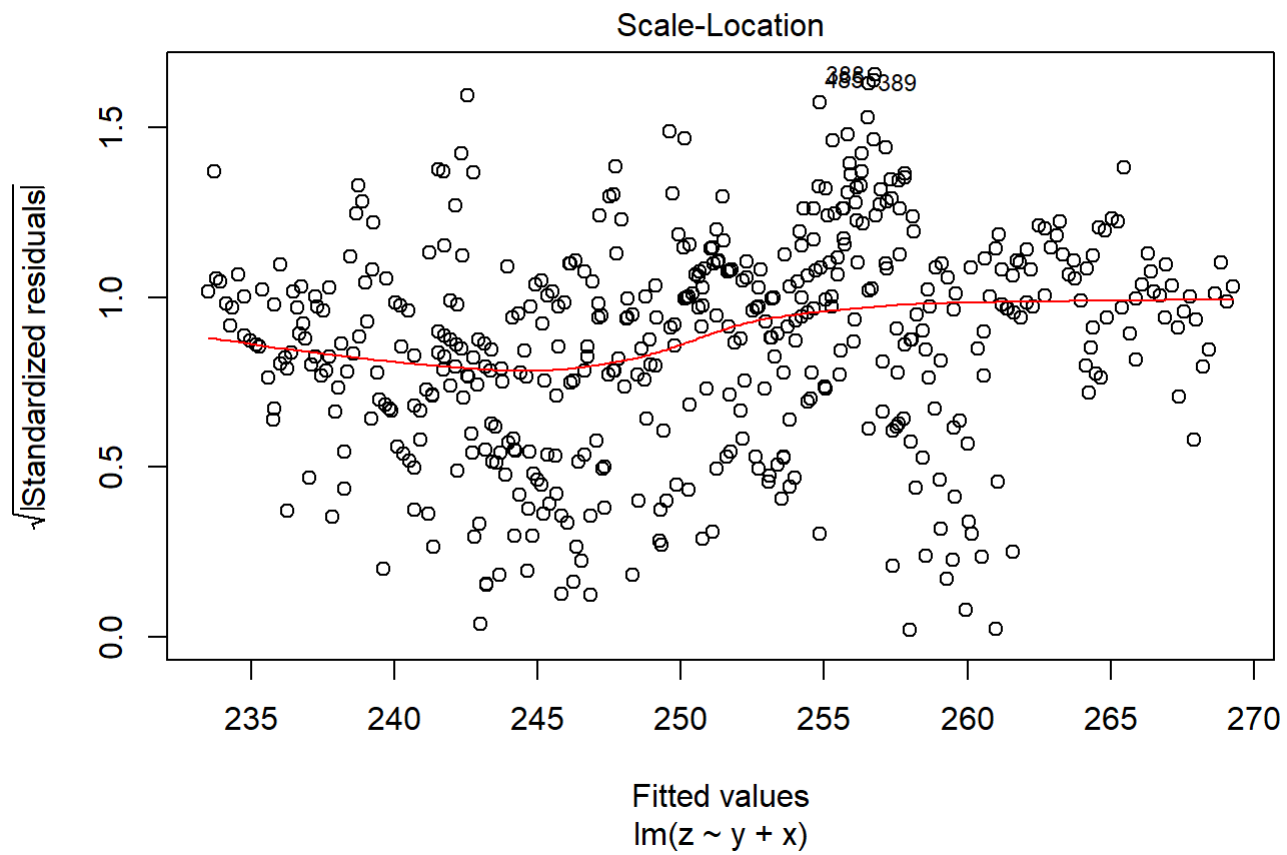
Multiple Linear Regression

```
seis.lm <- lm(z~y+x, data=seis)
summary(seis.lm)
```

```
##
## Call:
## lm(formula = z ~ y + x, data = seis)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.7271  -8.0735   0.1006   7.2143  25.2359
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  275.8787     1.3981  197.327  <2e-16 ***
## y            -1.7250     0.1968   -8.764  <2e-16 ***
## x            -4.0290     0.1907  -21.131  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.239 on 501 degrees of freedom
## Multiple R-squared:  0.4716, Adjusted R-squared:  0.4695
## F-statistic: 223.6 on 2 and 501 DF,  p-value: < 2.2e-16
```

```
plot(seis.lm)
```



```
AIC(seis.lm)
```

```
## [1] 3676.541
```

Bivariate Spline Regression (with equally spaced knots)

```
library(mgcv)
```

```
## Warning: package 'mgcv' was built under R version 3.5.3
```

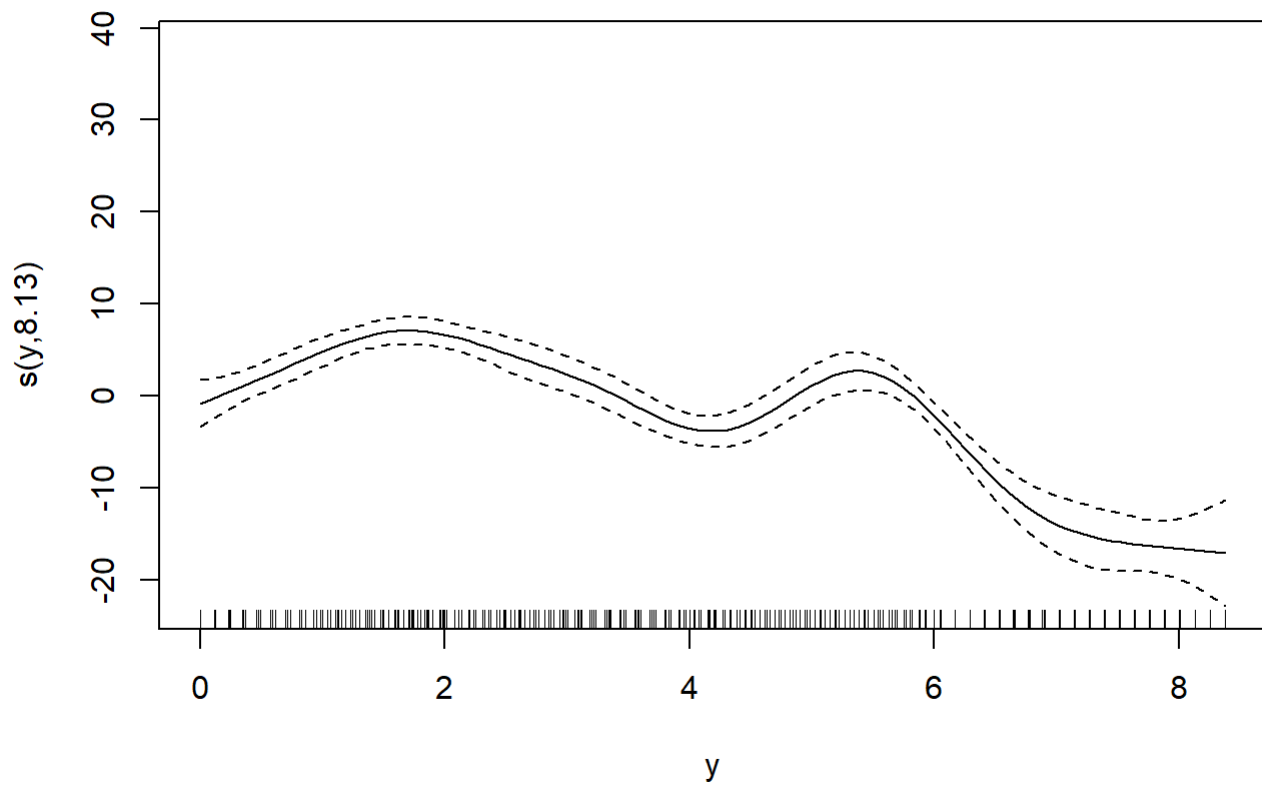
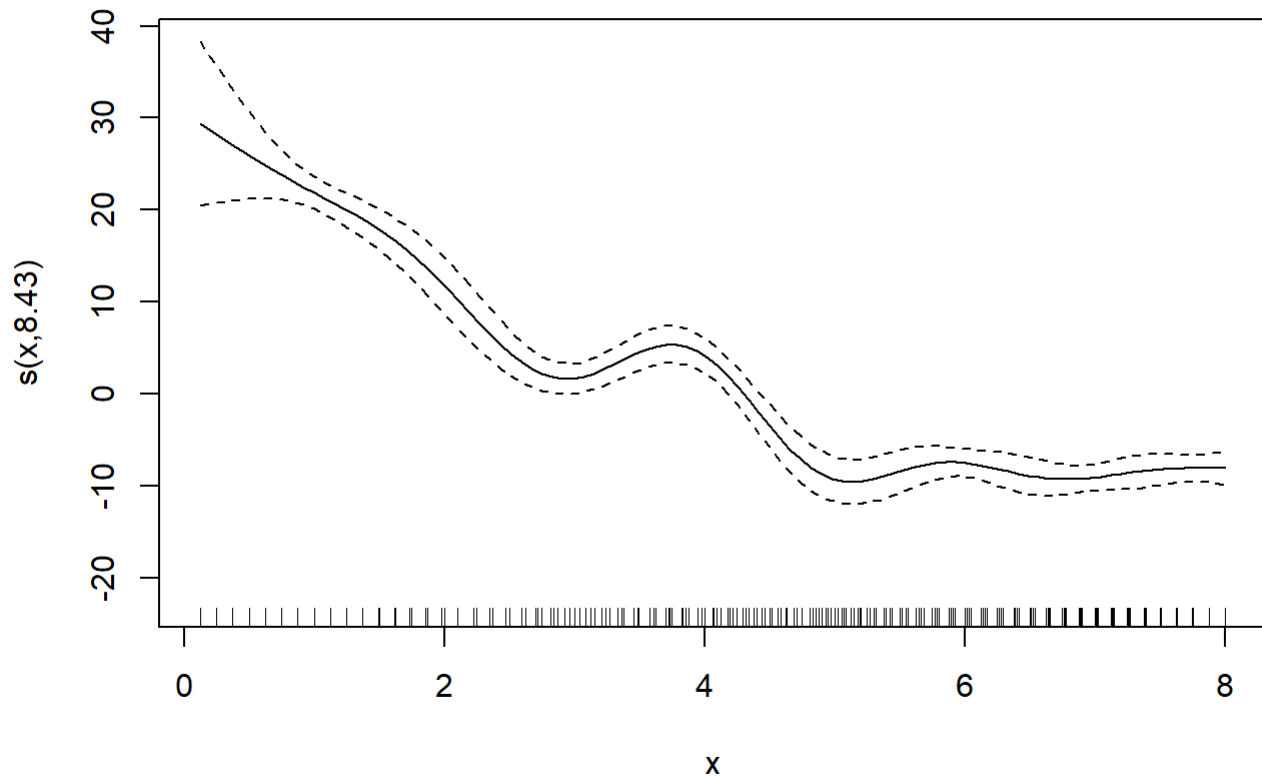
```
## Loading required package: nlme
```

```
## This is mgcv 1.8-28. For overview type 'help("mgcv-package")'.
```

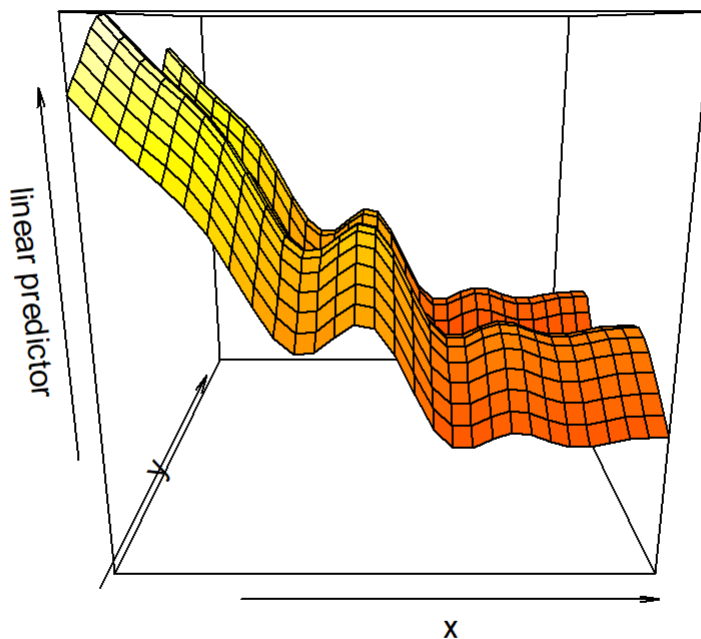
```
number.knots <- 10
spacings_x <- seq(from=min(seis$x),to=max(seis$x),length=number.knots+2)[2:(number.knots+1)]
spacings_y <- seq(from=min(seis$y),to=max(seis$y),length=number.knots+2)[2:(number.knots+1)]
seis.sr <- gam(z ~ s(x) + s(y),knots=list(spacings_x,spacings_y), data = seis)
summary(seis.sr)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## z ~ s(x) + s(y)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 250.6250    0.3218   778.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df    F p-value
## s(x)  8.433   8.898 97.09  <2e-16 ***
## s(y)  8.129   8.774 33.15  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.676   Deviance explained = 68.6%
## GCV = 54.062   Scale est. = 52.178    n = 504
```

```
plot.gam(seis.sr)
```



```
vis.gam(seis.sr)
```



```
AIC(seis.sr)
```

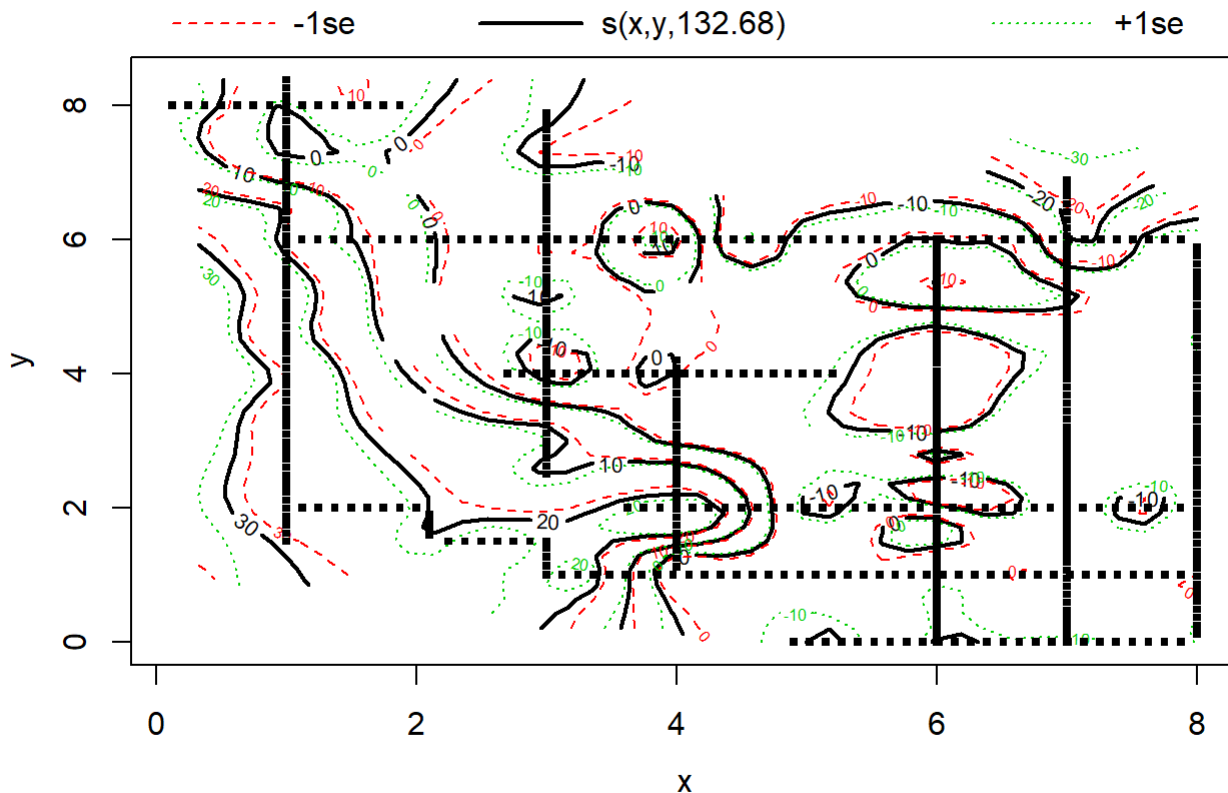
```
## [1] 3442.686
```

Thin-plate Splines

```
# thin plate spline, k=150
seis.tps1 <- gam(z~s(x,y,k=150), data=seis)
summary(seis.tps1)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## z ~ s(x, y, k = 150)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 250.6250    0.1299   1929  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df    F p-value
## s(x,y) 132.7  145.5 62.18  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.947   Deviance explained = 96.1%
## GCV = 11.575   Scale est. = 8.5048    n = 504
```

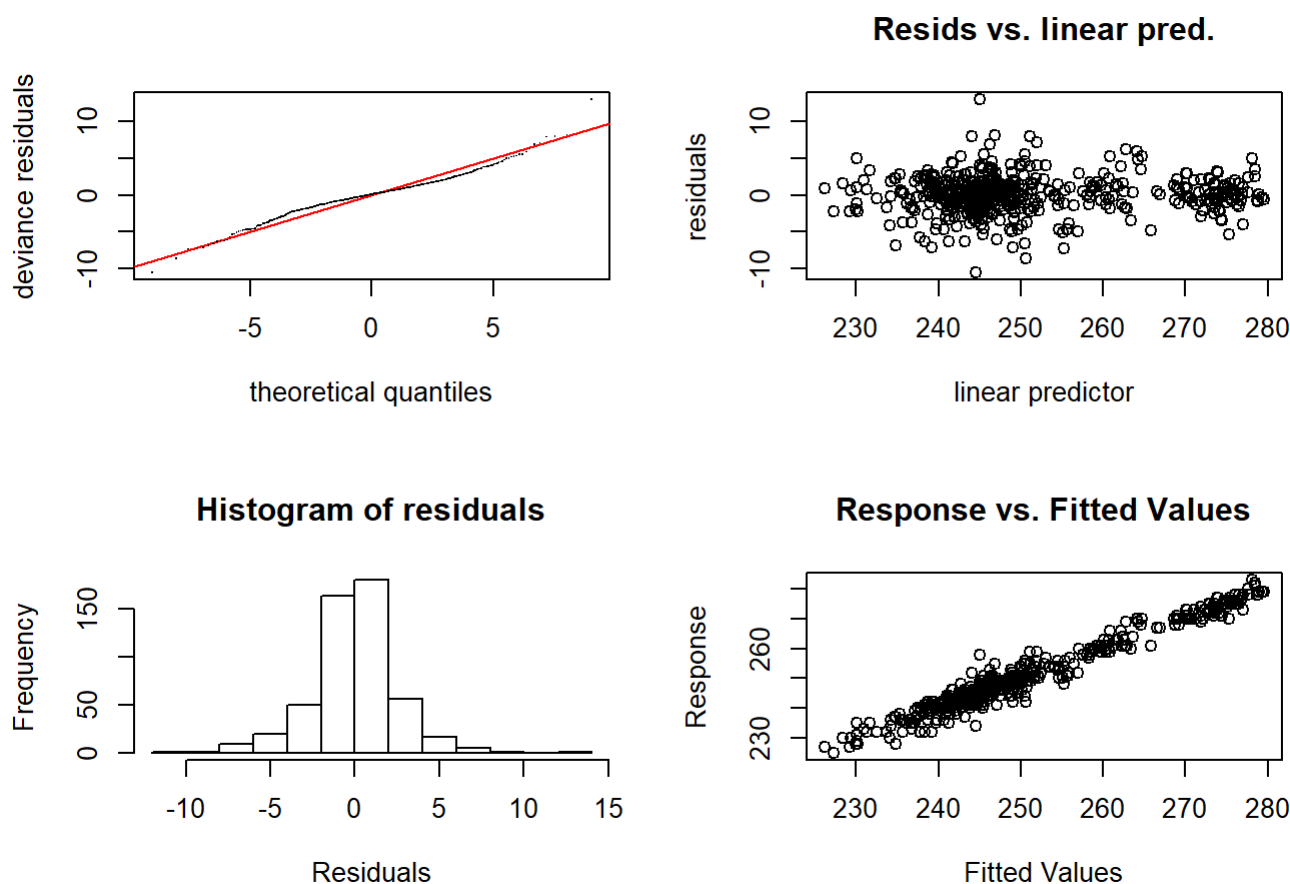
```
plot(seis.tpsl, cex=4)
```




```
AIC(seis.tps1)
```

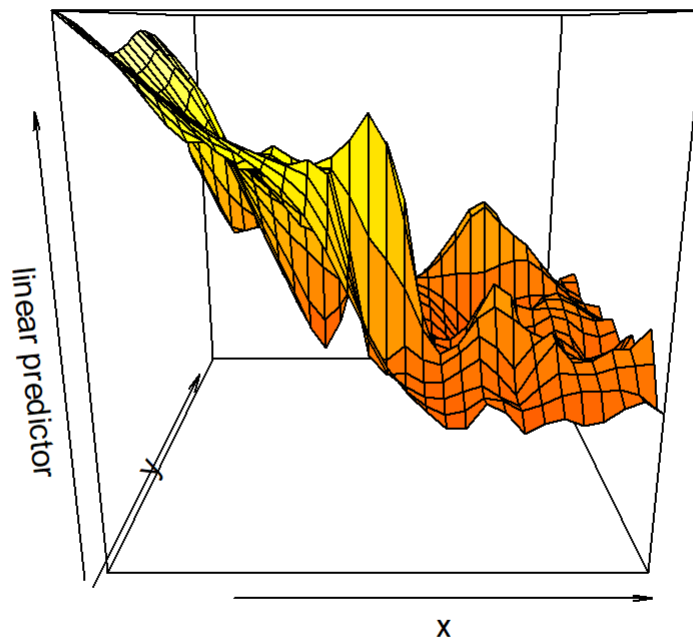
```
## [1] 2623.19
```

```
gam.check(seis.tps1)
```



```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 5 iterations by steepest
## descent step failure.
## The RMS GCV score gradient at convergence was 4.119849e-06 .
## The Hessian was positive definite.
## Model rank = 150 / 150
##
## 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(x,y) 149 133   1.11   0.99
```

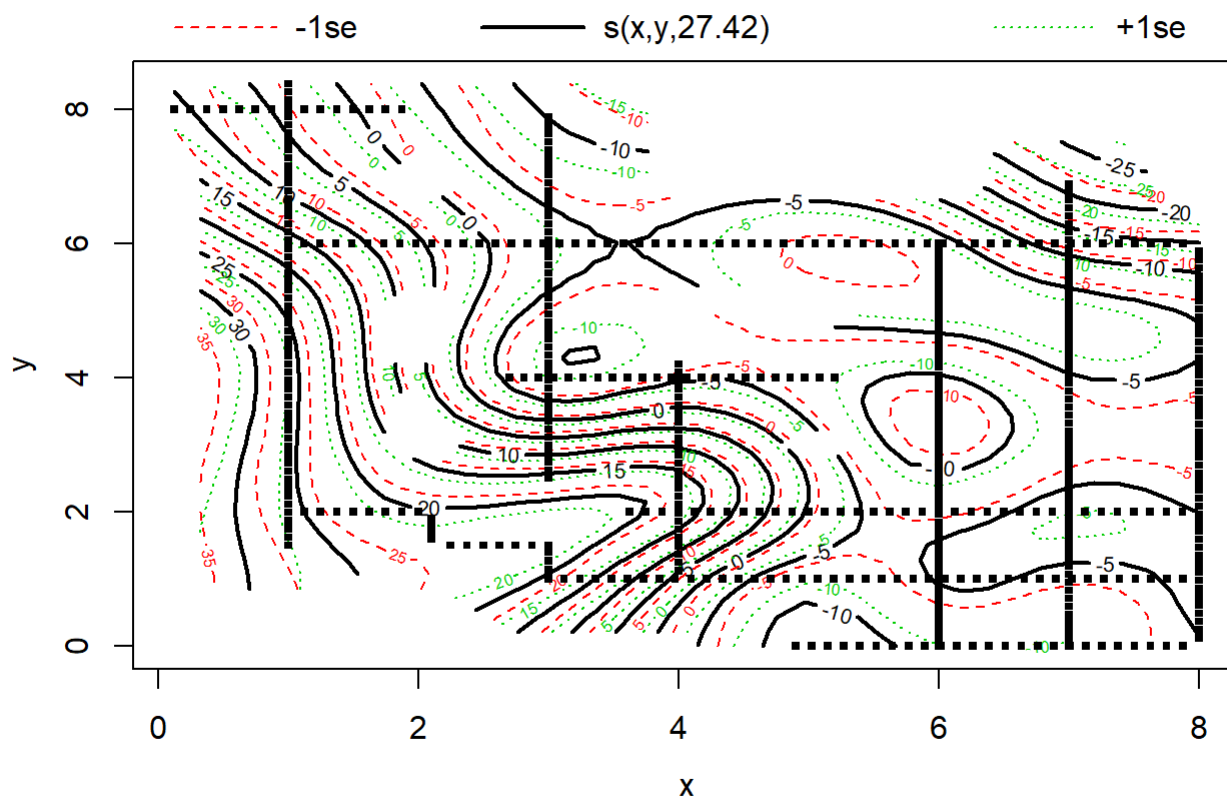
```
vis.gam(seis.tps1)
```



```
# thin plate spline, k=-1
seis.tpsl <- gam(z~s(x,y,k=-1), data=seis)
summary(seis.tpsl)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## z ~ s(x, y, k = -1)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 250.6250    0.2582   970.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df    F p-value
## s(x,y) 27.42  28.84 66.29 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.791   Deviance explained = 80.3%
## GCV = 35.621   Scale est. = 33.612    n = 504
```

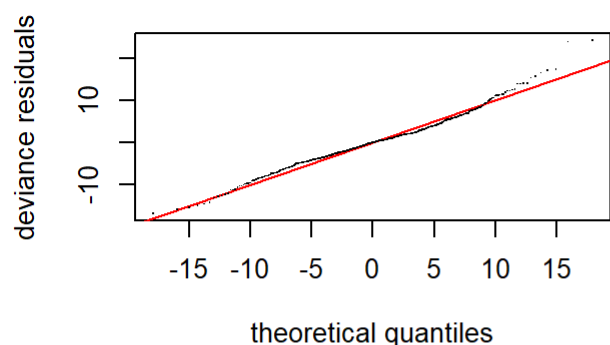
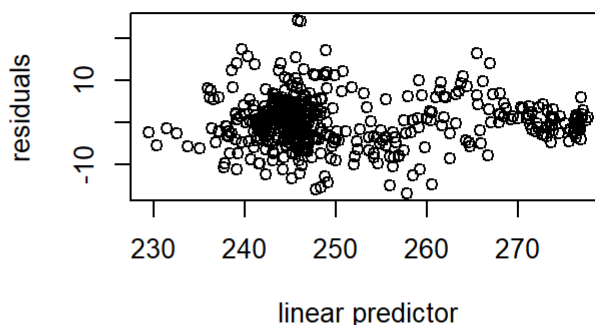
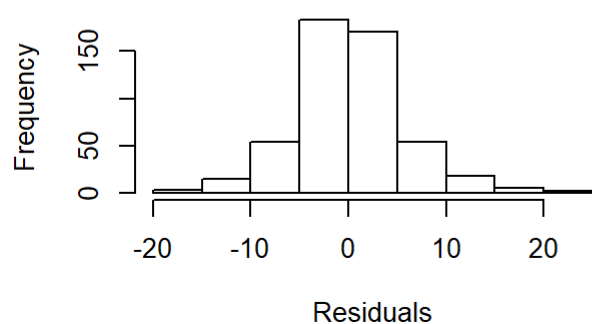
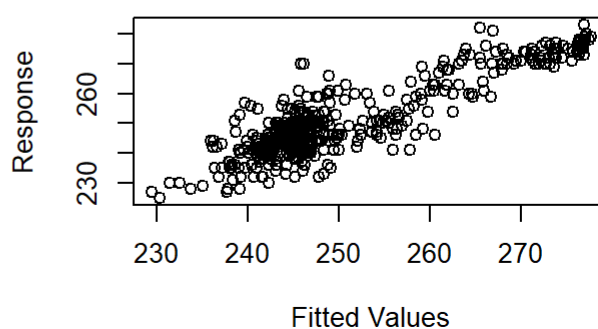
```
plot(seis.tpsl, cex=4)
```



```
AIC(seis.tpsl)
```

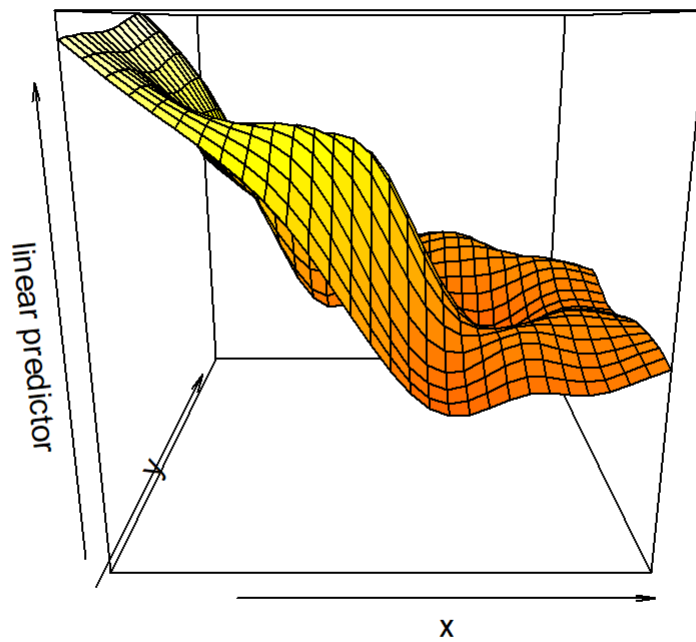
```
## [1] 3231.378
```

```
gam.check(seis.tpsl)
```

**Resids vs. linear pred.****Histogram of residuals****Response vs. Fitted Values**

```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 5 iterations.
## The RMS GCV score gradient at convergence was 2.844238e-05 .
## The Hessian was positive definite.
## Model rank = 30 / 30
##
## 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(x,y) 29.0 27.4   0.37 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vis.gam(seis.tps1)
```

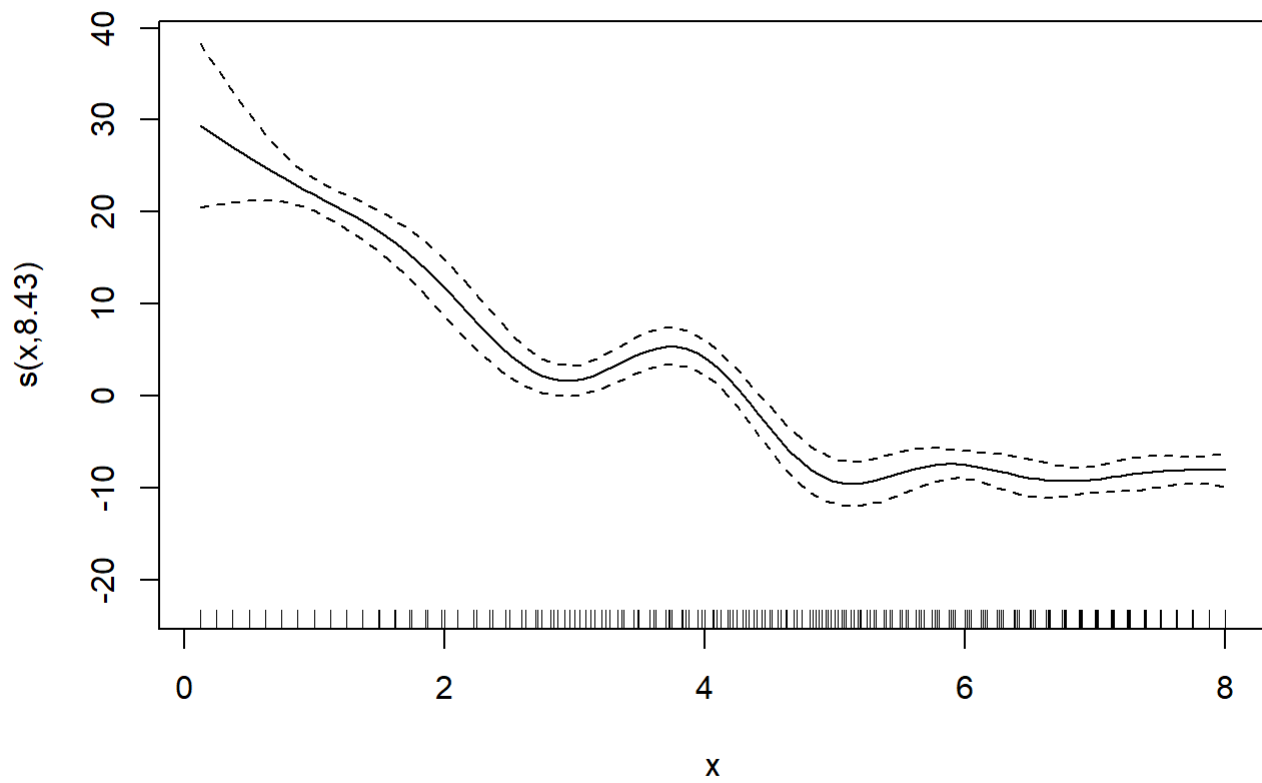
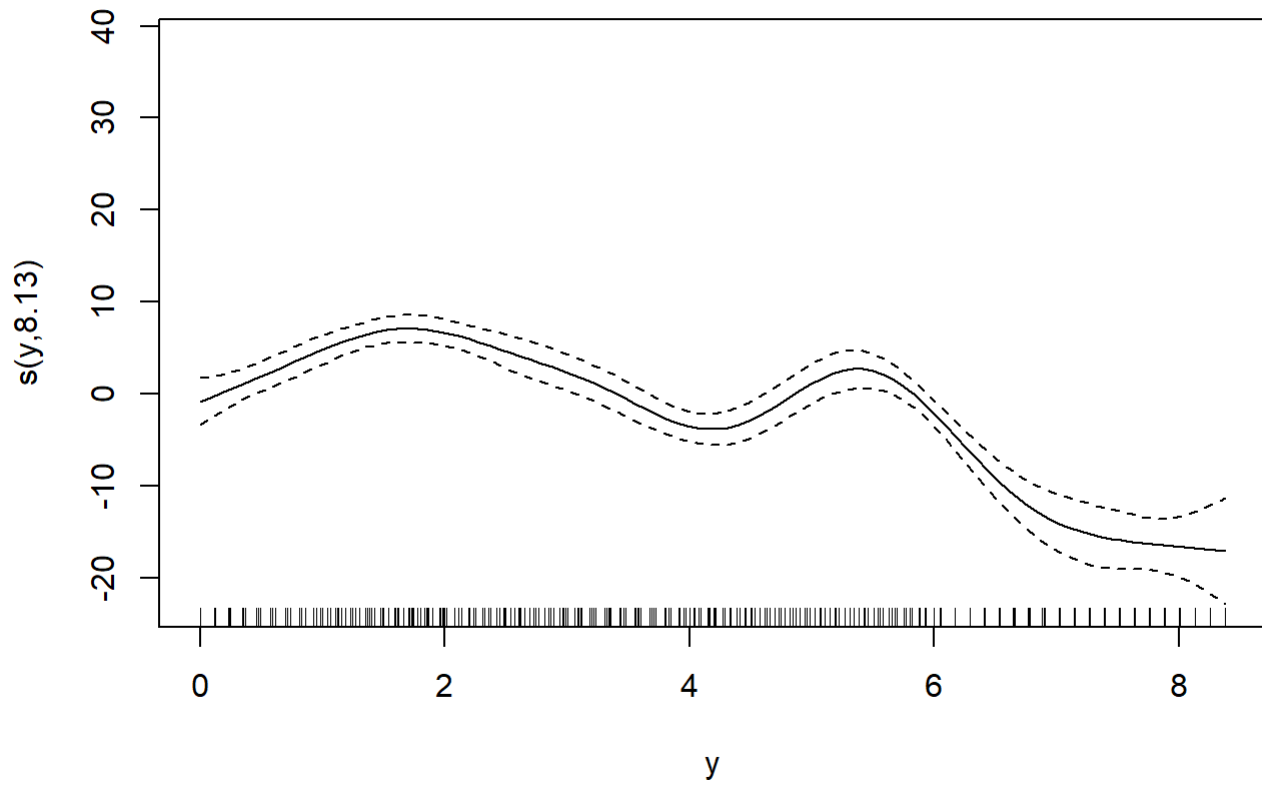


Generalized Additive Model

```
library(mgcv)
# normal family
seis.gam.normal <- gam(z ~ s(y) + s(x), data = seis)
summary(seis.gam.normal)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## z ~ s(y) + s(x)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 250.6250    0.3218   778.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df    F p-value
## s(y)  8.129   8.774 33.15  <2e-16 ***
## s(x)  8.433   8.898 97.09  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.676   Deviance explained = 68.6%
## GCV = 54.062   Scale est. = 52.178     n = 504
```

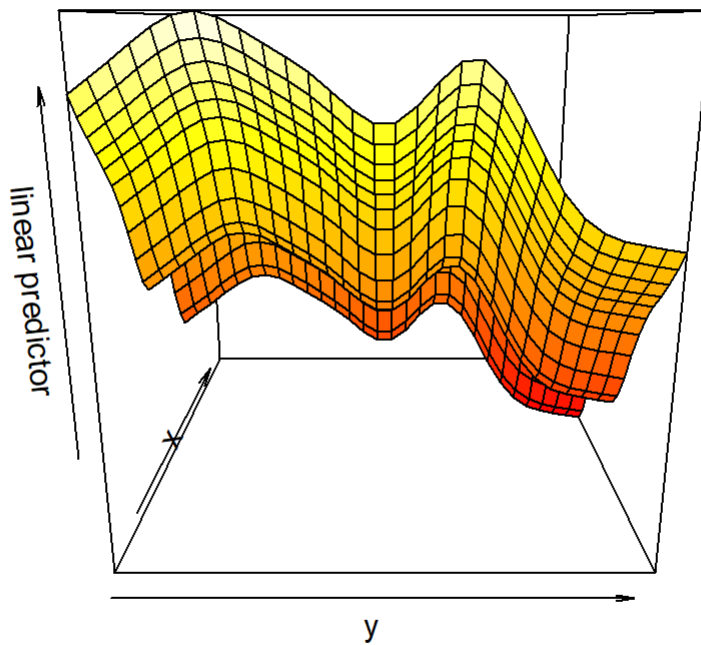
```
plot(seis.gam.normal)
```



```
AIC(seis.gam.normal)
```

```
## [1] 3442.686
```

```
vis.gam(seis.gam.normal)
```

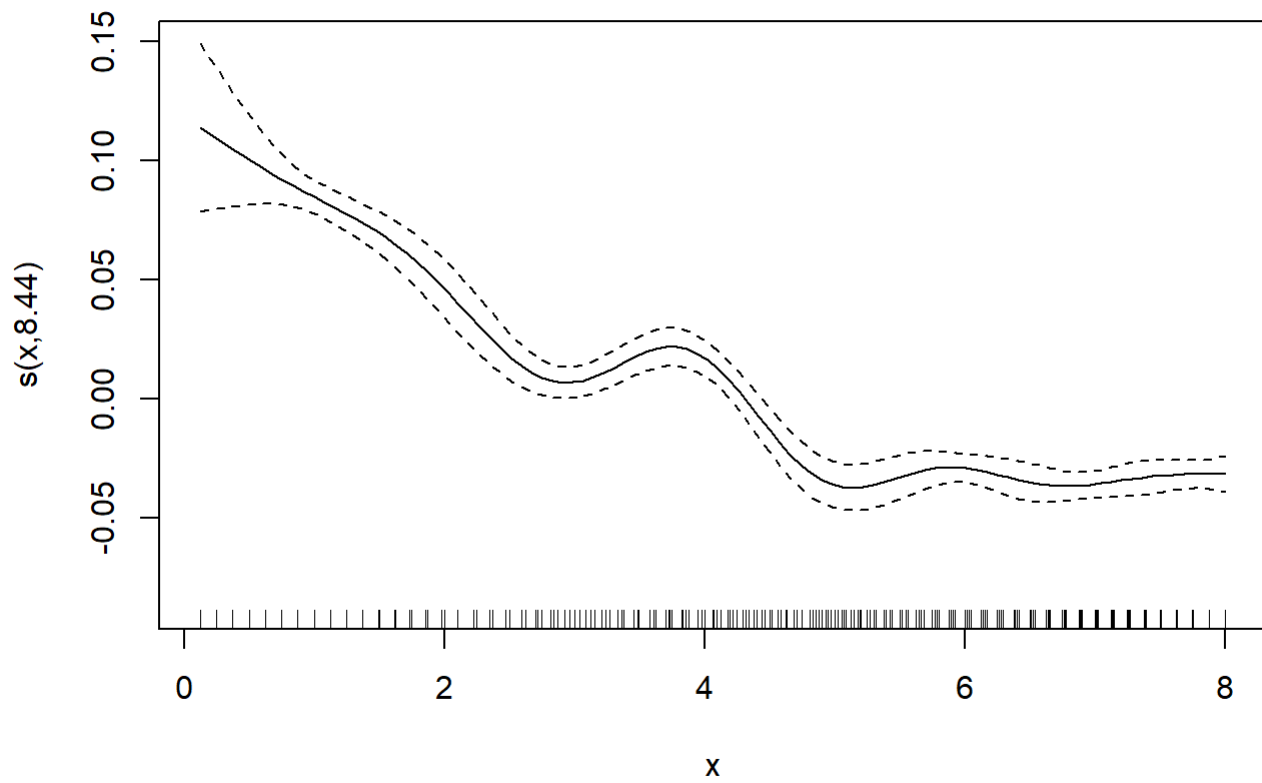
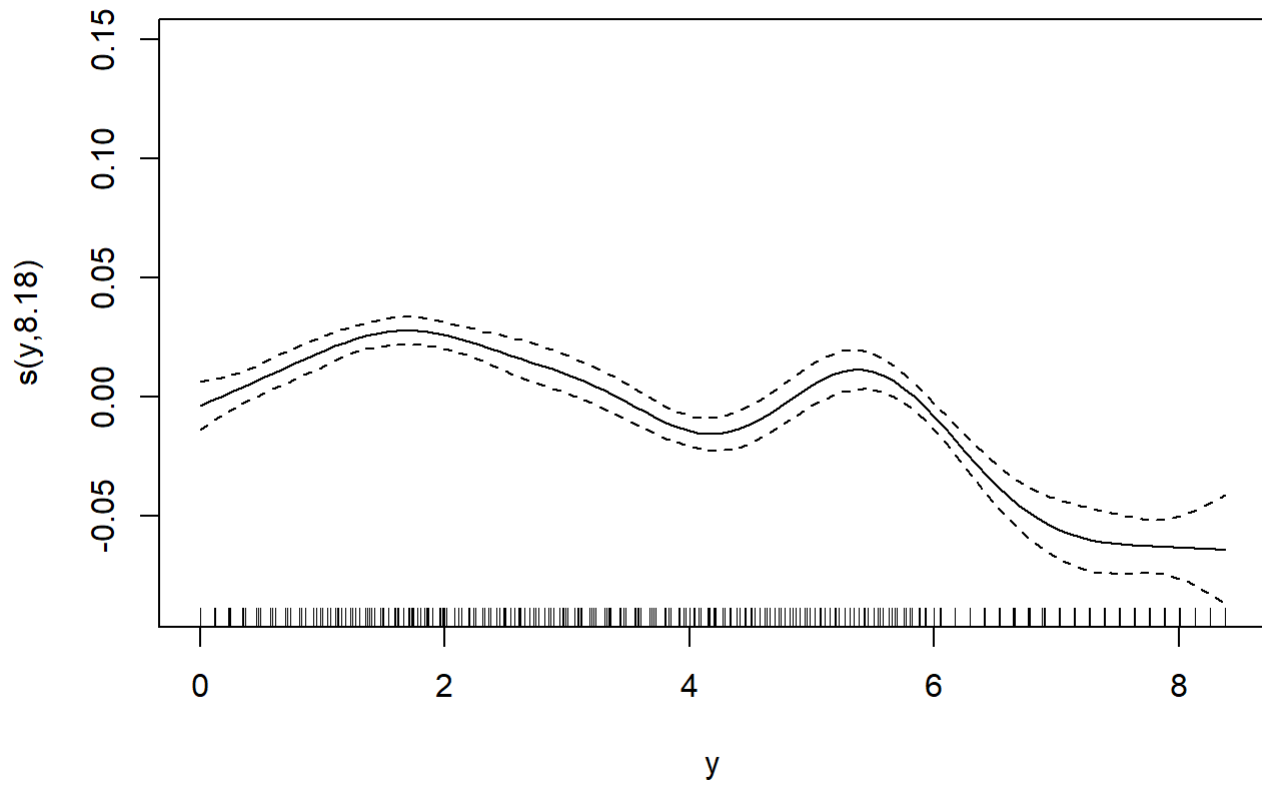


```
# gamma family  
seis.gam.gamma <- gam(z ~ s(y) + s(x), data = seis, family = Gamma(link = "log"))  
summary(seis.gam.gamma)
```



```
##
## Family: Gamma
## Link function: log
##
## Formula:
## z ~ s(y) + s(x)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.523104   0.001273   4338   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df      F p-value
## s(y) 8.179   8.798 32.22   <2e-16 ***
## s(x) 8.437   8.900 94.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.678   Deviance explained = 68.4%
## GCV = 0.00084565   Scale est. = 0.00081696   n = 504
```

```
plot(seis.gam.gamma)
```



```
AIC(seis.gam.gamma)
```

```
## [1] 3432.615
```

```
vis.gam(seis.gam.gamma)
```

