Practica III

Val Huerta

10/22/2019

Librerias

```
#lIBRERIAS
library(knitr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(skimr) # Beautiful Summarize
```

```
##
## Attaching package: 'skimr'
```

```
## The following object is masked from 'package:knitr':
##
##     kable
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(magrittr) # Pipe operators
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(corrplot) # Correlations
```

```
## corrplot 0.84 loaded
```

```
library(ggcorrplot)  # Correlations
library(PerformanceAnalytics) # Correlations
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend
```

```r
library(leaps) # Model selection
library(caret) # Cross Validation
```

```
## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(bestglm) # Cross Validation
library(glmnet) # Regularization
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loading required package: foreach

##
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':
##
##     accumulate, when

## Loaded glmnet 2.0-18
```

```r
library(gam) #GAM
```

```
## Loading required package: splines
## Loaded gam 1.16.1
```

```
library(rsample) #Para el train/test
```

Read Data

```
library(ISLR)
day<- read.csv("day.csv")
```

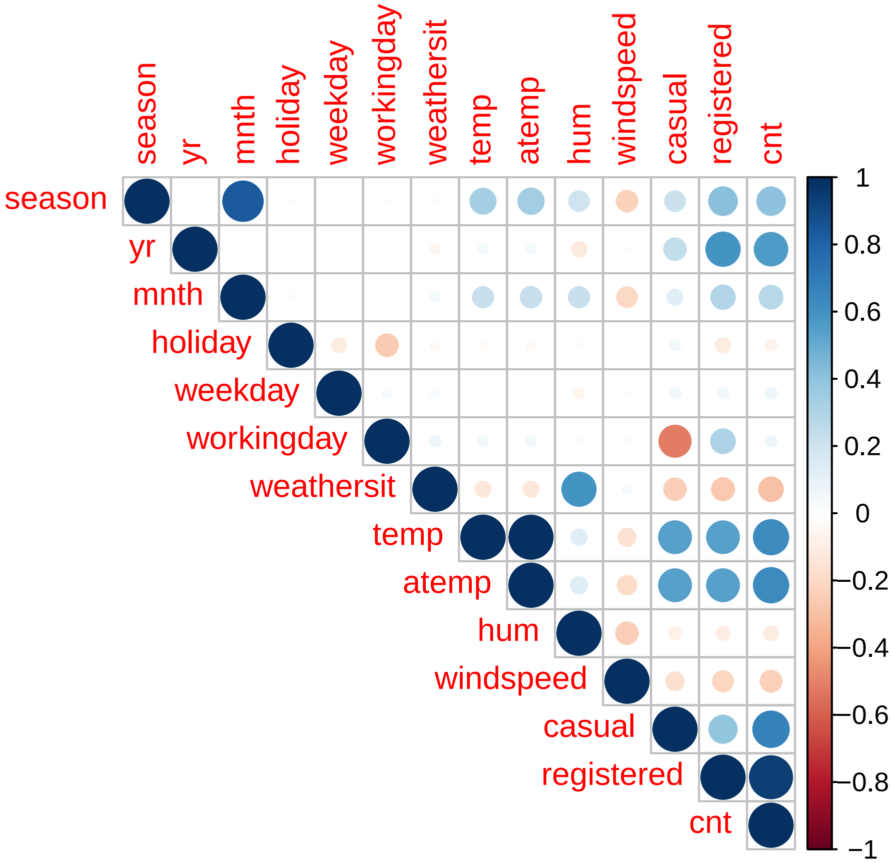
Summarize Data

```
skim(day)
```

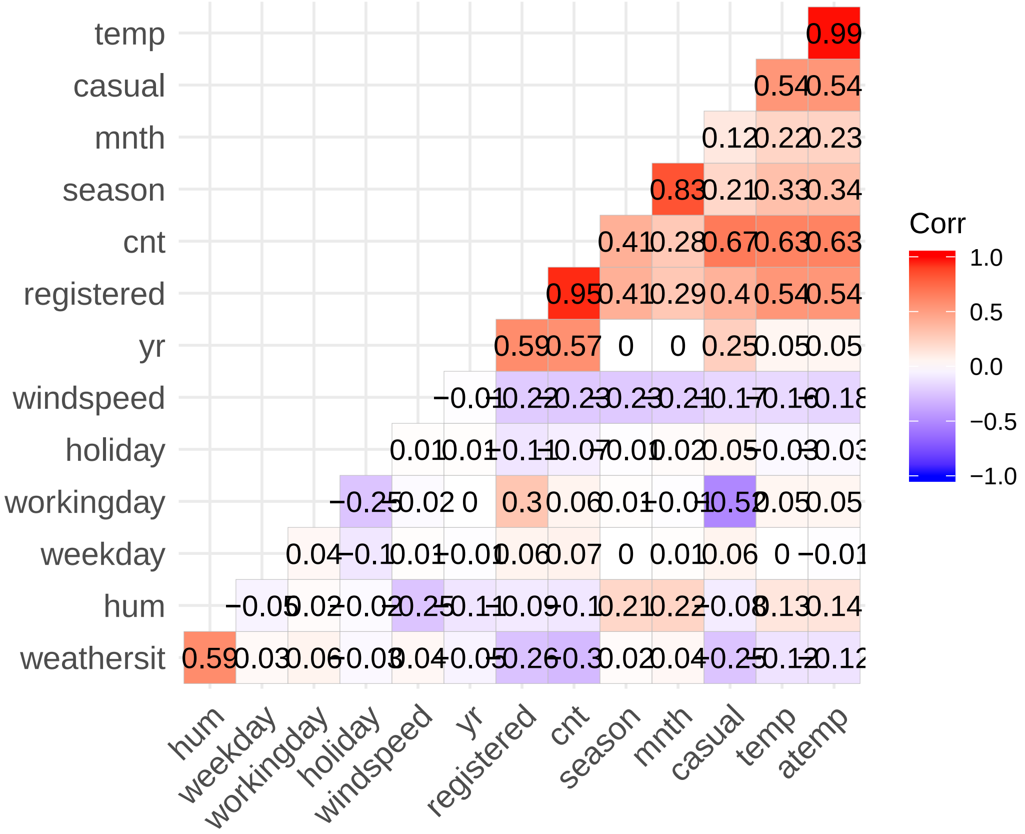
```
## Skim summary statistics
##  n obs: 731
##  n variables: 16
##
## -- Variable type:factor --------------------------------------------------------
##  variable missing complete   n n_unique                     top_counts
##    dteday       0      731 731      731 201: 1, 201: 1, 201: 1, 201: 1
##  ordered
##    FALSE
##
## -- Variable type:integer -------------------------------------------------------
##    variable missing complete   n     mean      sd p0    p25  p50    p75
##      casual       0      731 731  848.18   686.62  2  315.5  713 1096
##         cnt       0      731 731 4504.35  1937.21 22 3152   4548 5956
##     holiday       0      731 731    0.029    0.17  0    0      0    0
##     instant       0      731 731  366      211.17  1  183.5  366  548.5
##        mnth       0      731 731    6.52     3.45  1    4      7   10
##  registered       0      731 731 3656.17  1560.26 20 2497   3662 4776.5
##      season       0      731 731    2.5      1.11  1    2      3    3
##  weathersit       0      731 731    1.4      0.54  1    1      1    2
##     weekday       0      731 731    3        2     0    1      3    5
##  workingday       0      731 731    0.68     0.47  0    0      1    1
##          yr       0      731 731    0.5      0.5   0    0      1    1
##  p100     hist
##  3410
##  8714
##     1
##   731
##    12
##  6946
##     4
##     3
##     6
##     1
##     1
##
## -- Variable type:numeric -------------------------------------------------------
##   variable missing complete   n mean    sd    p0  p25  p50  p75 p100
##      atemp       0      731 731 0.47 0.16  0.079 0.34 0.49 0.61 0.84
##        hum       0      731 731 0.63 0.14  0     0.52 0.63 0.73 0.97
##       temp       0      731 731 0.5  0.18  0.059 0.34 0.5  0.66 0.86
##  windspeed       0      731 731 0.19 0.077 0.022 0.13 0.18 0.23 0.51
##      hist
##
##
##
##
```

Correlaciones

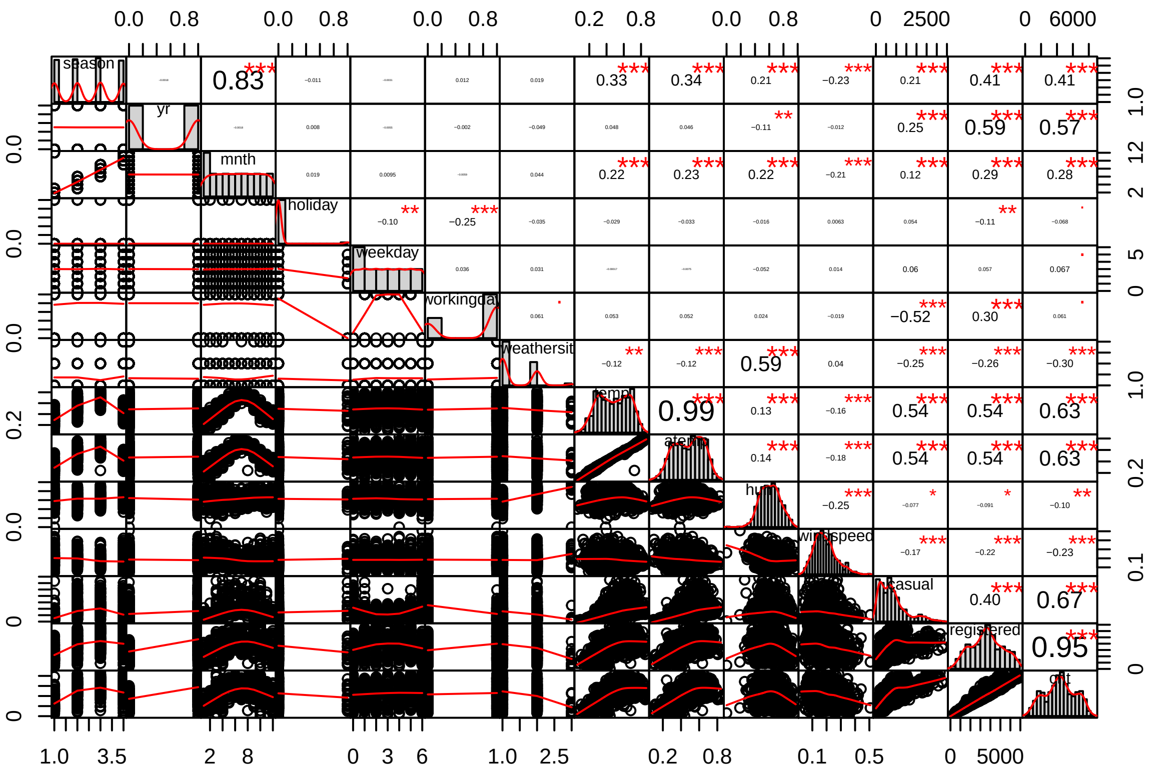
```
#AQUI MIRO LAS CORRELACIONES y quito la variable instant y dteday porque es una variable factor.

#PARA VER PROBLEMAS DE MULTICOLINEALIDAD, dependiendo de su correlacion si es lineal o no.
#El rojo significa que son valores negativos.

# Variables excluidas:
factores <- c("instant","dteday")


# Correlaciones
#corplot solo grafica la correlacion no me la calcula por eso pone luego cor.
corrplot(cor(day%>%
               select_at(vars(-factores)),
             use = "complete.obs"),
         method = "circle",type = "upper")
```



```r
# Other Correlations

ggcorrplot(cor(day %>%
                 select_at(vars(-factores)),
              use = "complete.obs"),
              hc.order = TRUE,
              type = "lower",   lab = TRUE)
```



```
# Other Correlations
#Las estrellas en rojo lo que indica es que si es distinta de 0, es decir cuales estan correlacionadas.
chart.Correlation(day %>%
                    select_at(vars(-factores)),
                 histogram=TRUE, pch=19)
```



Grados de libertad

```
#Aqui estoy sacando los grados de libertad de cada variable junto con el CV.
#Unicamente los calculo para las variables que no son categoricas ni dumbies.

DOFtemp <- smooth.spline(day$temp,day$cnt, cv=TRUE)
```

```
## Warning in smooth.spline(day$temp, day$cnt, cv = TRUE): cross-validation
## with non-unique 'x' values seems doubtful
```

```
DOFatemp <- smooth.spline(day$atemp,day$cnt, cv=TRUE)
```

```
## Warning in smooth.spline(day$atemp, day$cnt, cv = TRUE): cross-validation
## with non-unique 'x' values seems doubtful
```

```
DOFhum <- smooth.spline(day$hum,day$cnt, cv=TRUE)
```

```
## Warning in smooth.spline(day$hum, day$cnt, cv = TRUE): cross-validation
## with non-unique 'x' values seems doubtful
```

```
DOFwindspeed <- smooth.spline(day$windspeed,day$cnt, cv=TRUE)
```

```
## Warning in smooth.spline(day$windspeed, day$cnt, cv = TRUE): cross-
## validation with non-unique 'x' values seems doubtful
```

```
DOFcasual <- smooth.spline(day$casual,day$cnt, cv=TRUE)
```

```
## Warning in smooth.spline(day$casual, day$cnt, cv = TRUE): cross-validation
## with non-unique 'x' values seems doubtful
```

```r
DOFregistered <- smooth.spline(day$registered, day$cnt,cv=TRUE)
```

```
## Warning in smooth.spline(day$registered, day$cnt, cv = TRUE): cross-
## validation with non-unique 'x' values seems doubtful
```

```r
DOFtemp$df
```

```
## [1] 9.103704
```

```r
DOFatemp$df
```

```
## [1] 8.805497
```

```r
DOFhum$df
```

```
## [1] 4.548876
```

```r
DOFwindspeed$df
```

```
## [1] 6.007664
```

```r
DOFcasual$df
```

```
## [1] 11.27571
```

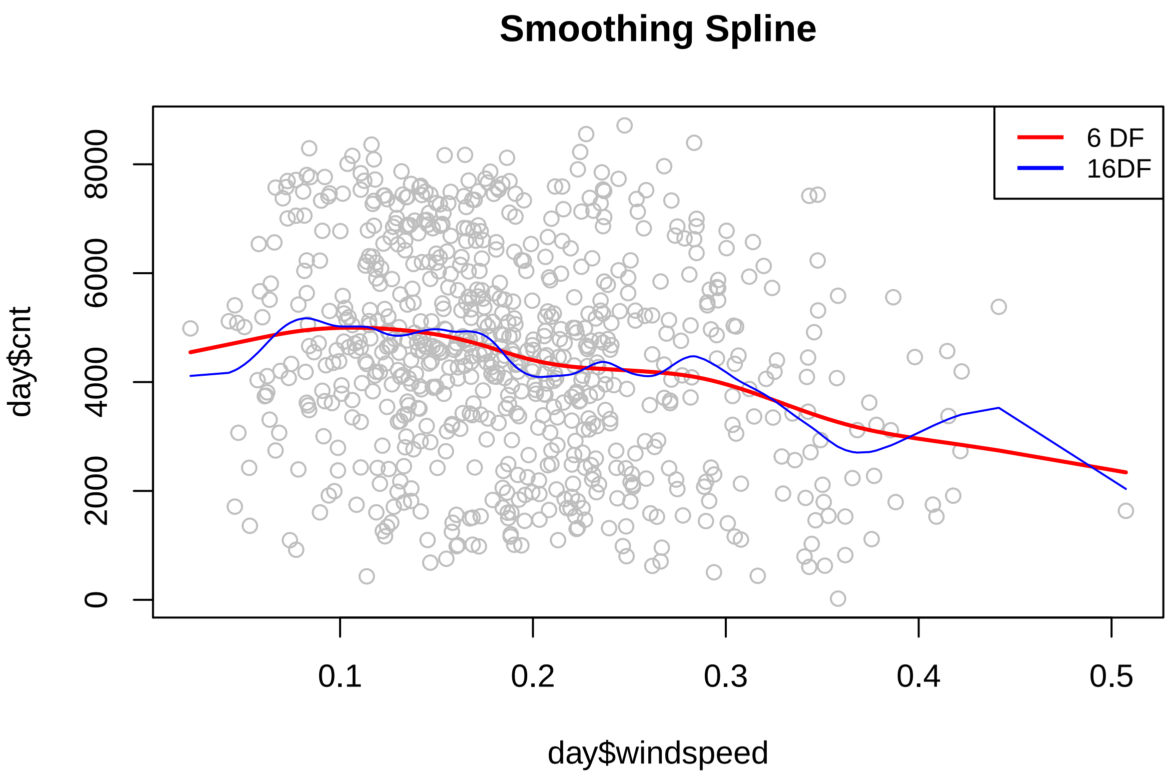
```r
DOFregistered$df
```

```
## [1] 12.95976
```

```r
#Ejemplo gráfico cogiendo la variable "windspeed". Utilizando para comparar con
#16 grados de libertad

plot(day$windspeed,day$cnt, xlim=day$windspeedLims, col='gray')
title('Smoothing Spline')
DOFwindspeed <- smooth.spline(day$windspeed,day$cnt, cv=TRUE)
```

```
## Warning in smooth.spline(day$windspeed, day$cnt, cv = TRUE): cross-
## validation with non-unique 'x' values seems doubtful
```

```r
DOFwindspeed2 <- smooth.spline(day$windspeed,day$cnt, df=16)
lines(DOFwindspeed, col='red', lwd=2)
lines(DOFwindspeed2, col='blue', lwd=1)
legend('topright', legend=c('6 DF', '16DF'),
       col=c('red','blue'), lty=1, lwd=2, cex=0.8)
```



#

Cambio a factor

```
#Procedemos al cambio de las variables categoricas a factor para poder incluirlas en el modelo.
#Son las siguientes:

day$season <- as.factor(day$season)
day$weekday <- as.factor(day$weekday)
day$weathersit <- as.factor(day$weathersit)
day$mnth <- as.factor(day$mnth)

#Las DUMBIES no hay que cambiarlas a factor pero son: holiday, season y workingday
```

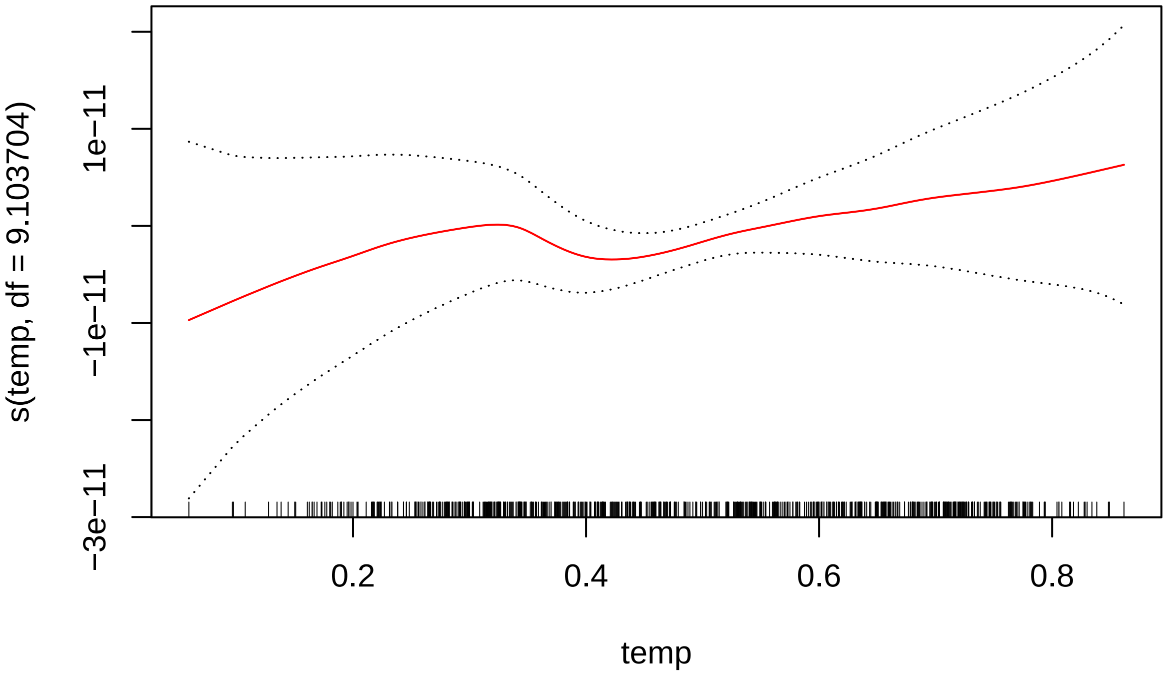
Modelo GAM

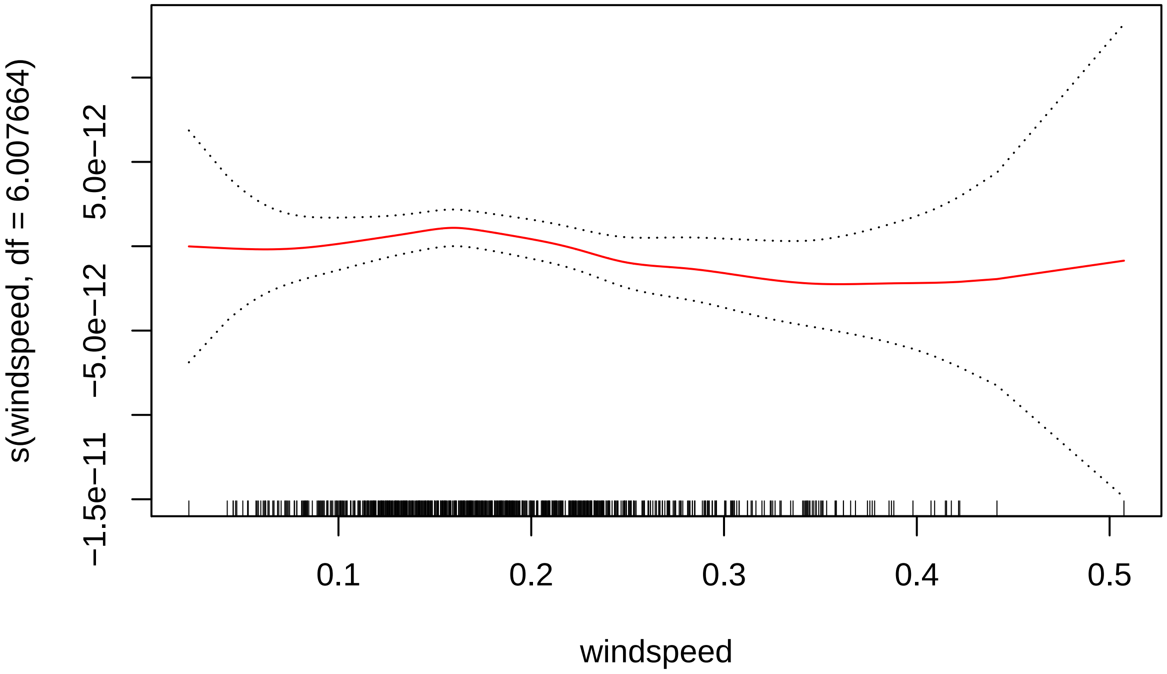
```
#A continuacion vamos a realizar los pertinentes modelos con GAM.

gam1 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, df=4.5
            data=day)
```

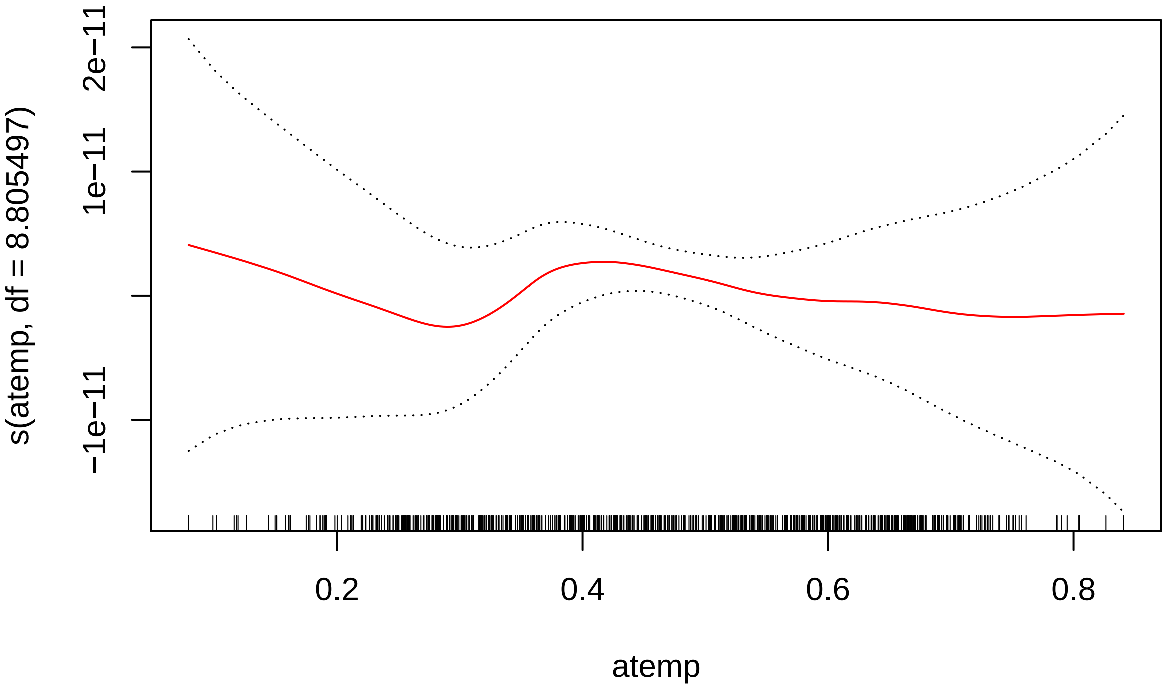
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

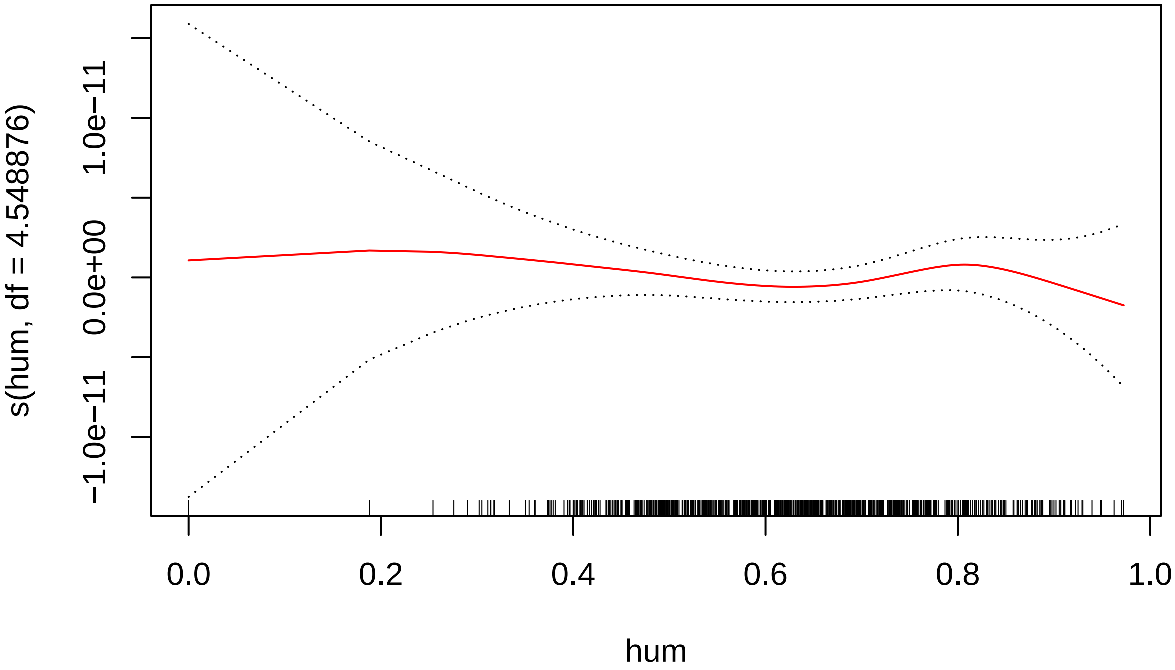
```
plot(gam1, se=TRUE, col='red')
```

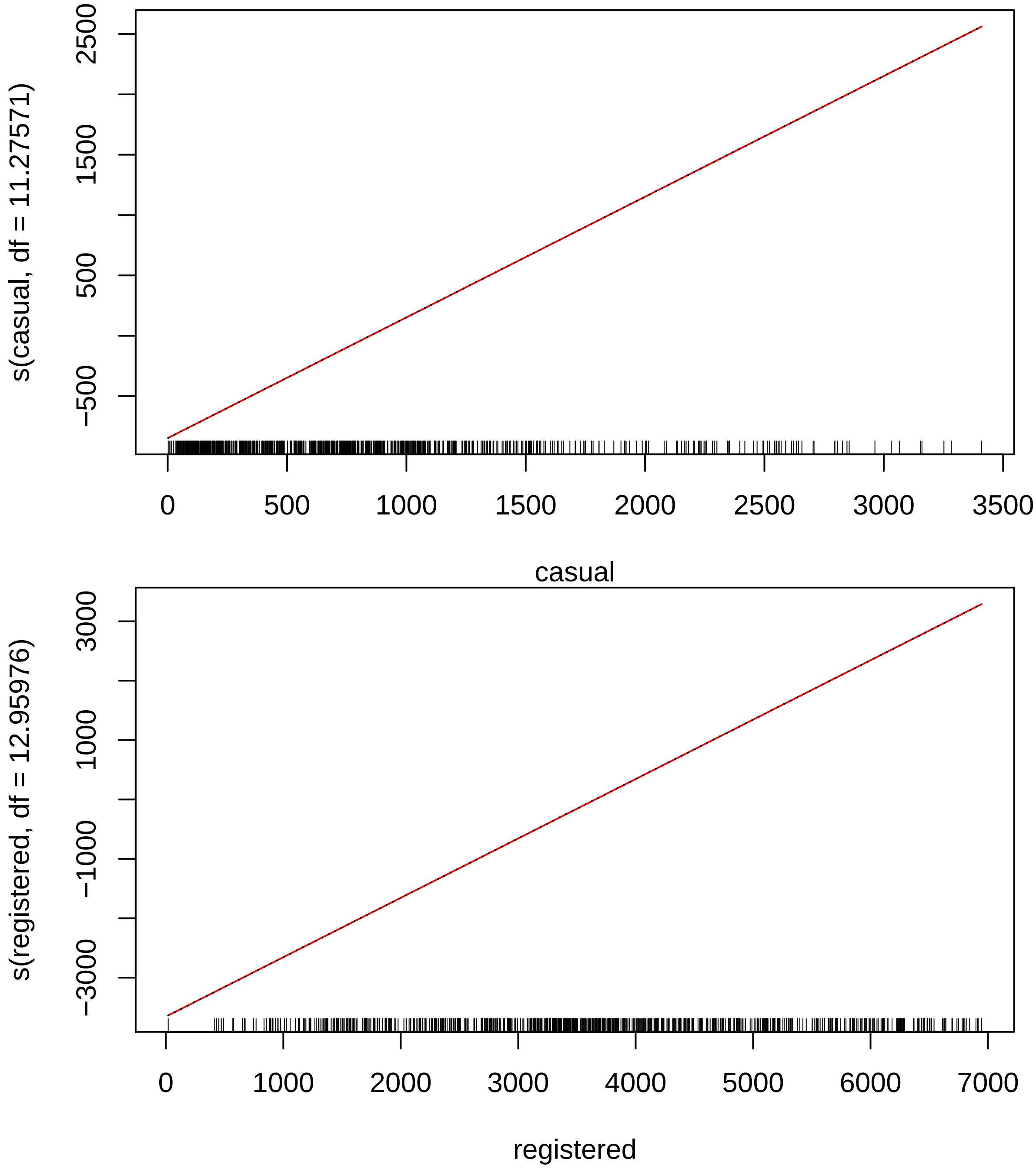
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

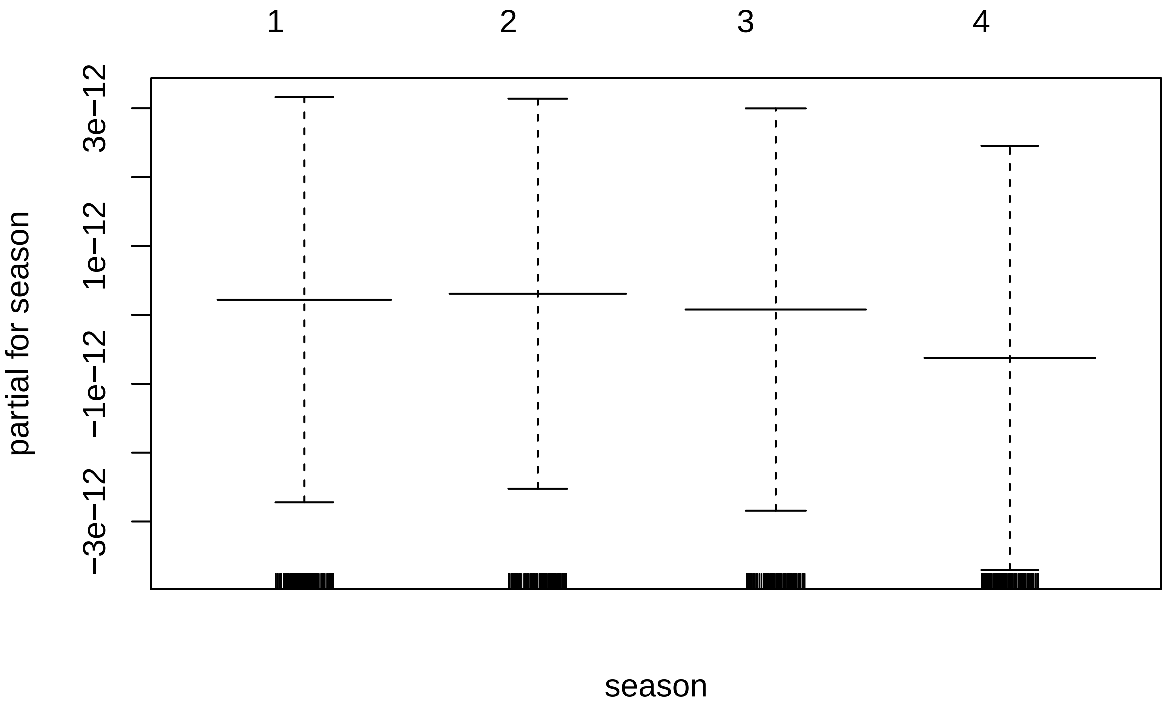

s(temp, df = 9.103704)
1e−11
−1e−11
−3e−11
0.2
0.4
0.6
0.8
temp

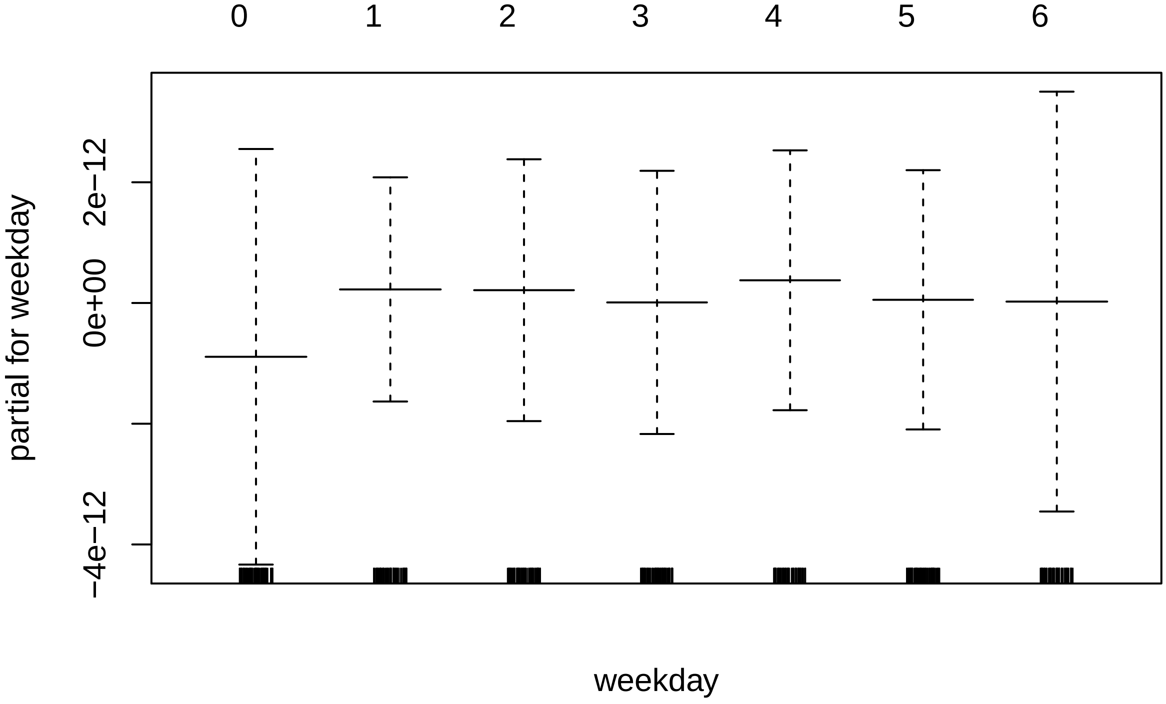

s(windspeed, df = 6.007664)
5.0e−12
−5.0e−12
−1.5e−11
0.1
0.2
0.3
0.4
0.5
windspeed

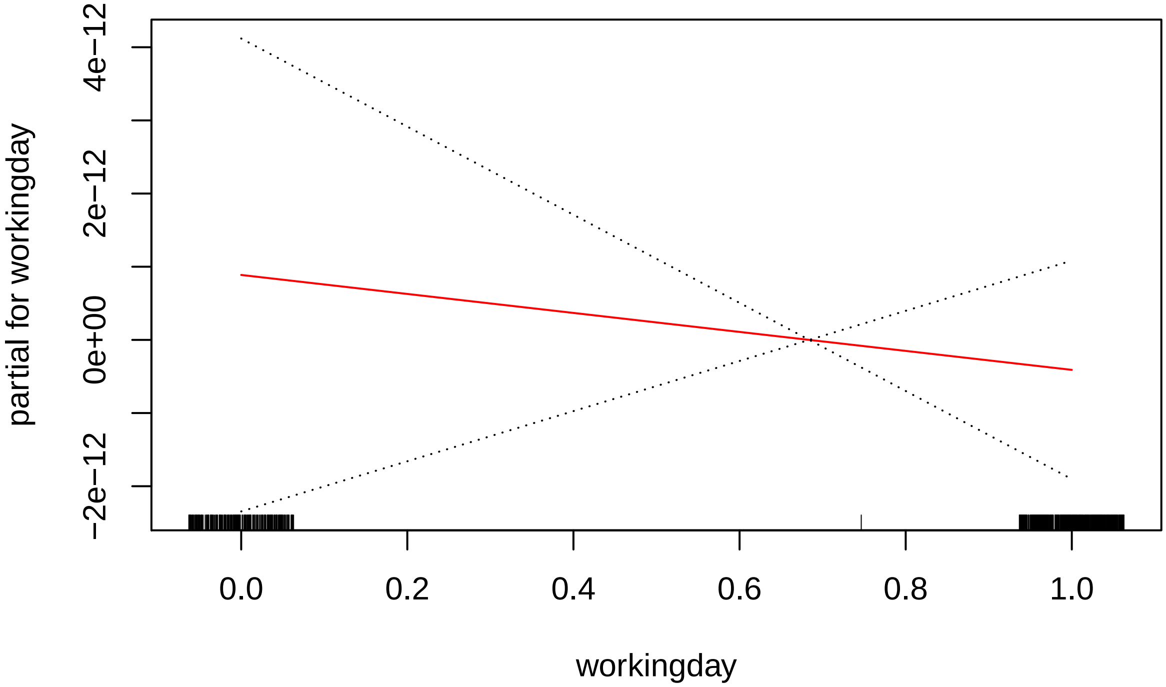

s(atemp, df = 8.805497)
2e−11
1e−11
−1e−11
0.2
0.4
0.6
0.8
atemp

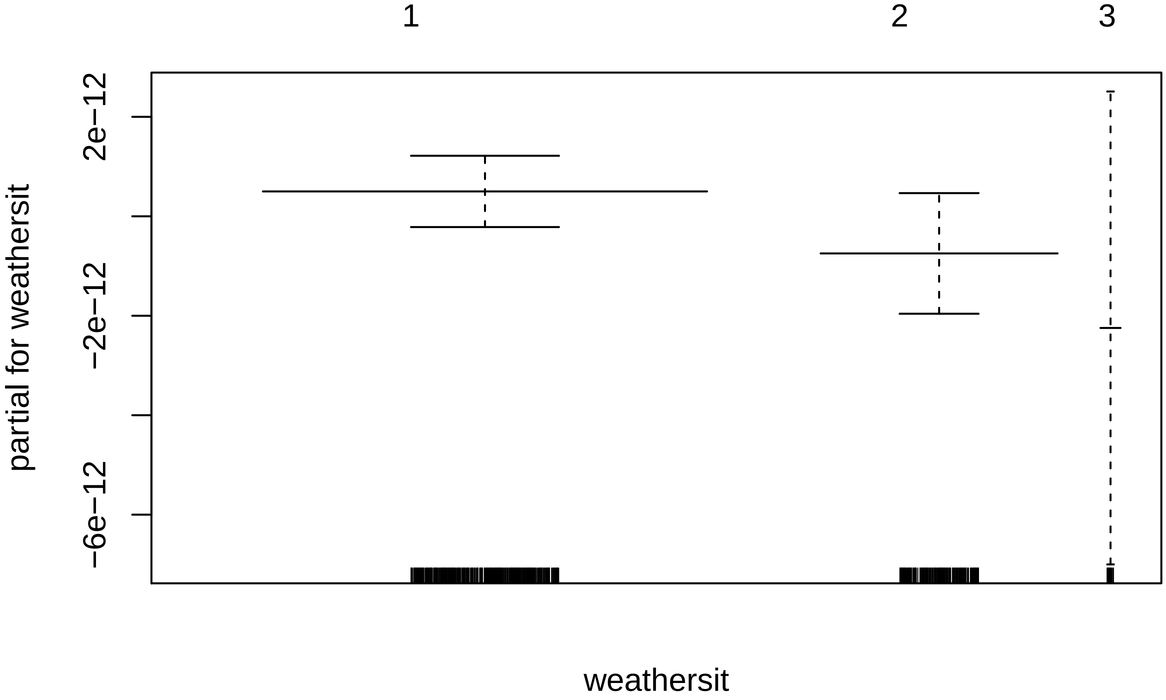

s(hum, df = 4.548876)
1.0e−11
0.0e+00
−1.0e−11
0.0
0.2
0.4
0.6
0.8
1.0
hum

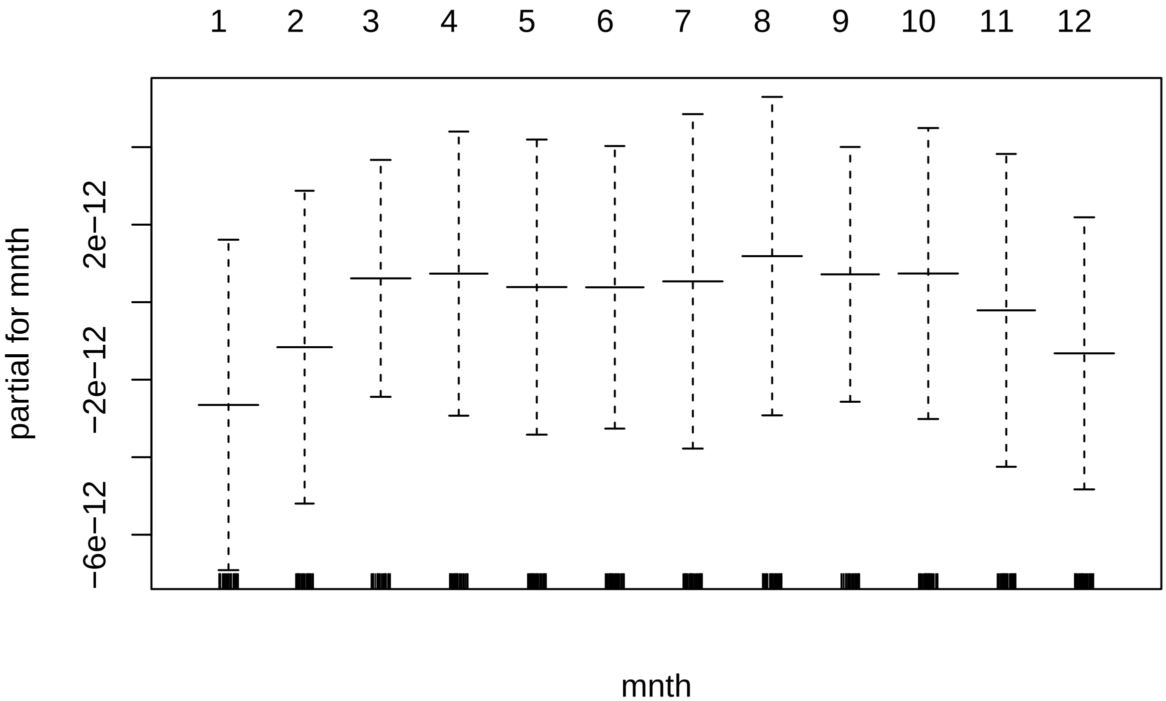

s(casual, df = 11.27571)
−500
500
1500
2500
0
500
1000
1500
2000
2500
3000
3500
casual
s(registered, df = 12.95976)
−3000
−1000
1000
3000
0
1000
2000
3000
4000
5000
6000
7000
registered

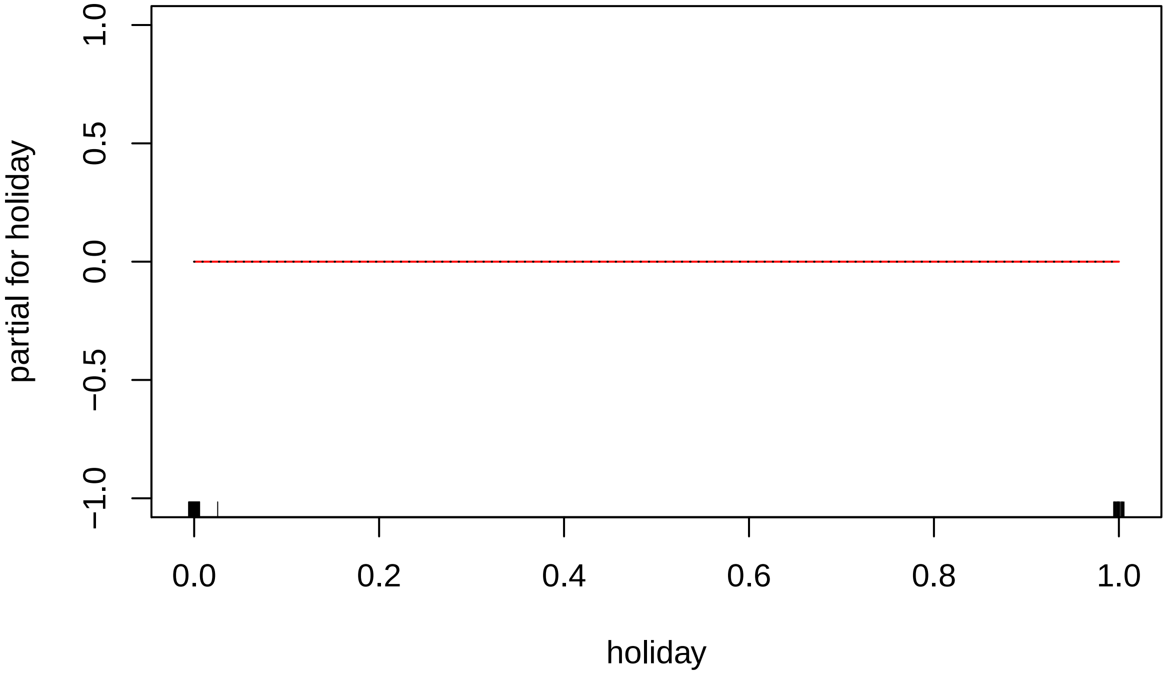

1
2
3
4
3e−12
1e−12
−1e−12
−3e−12
partial for season
season

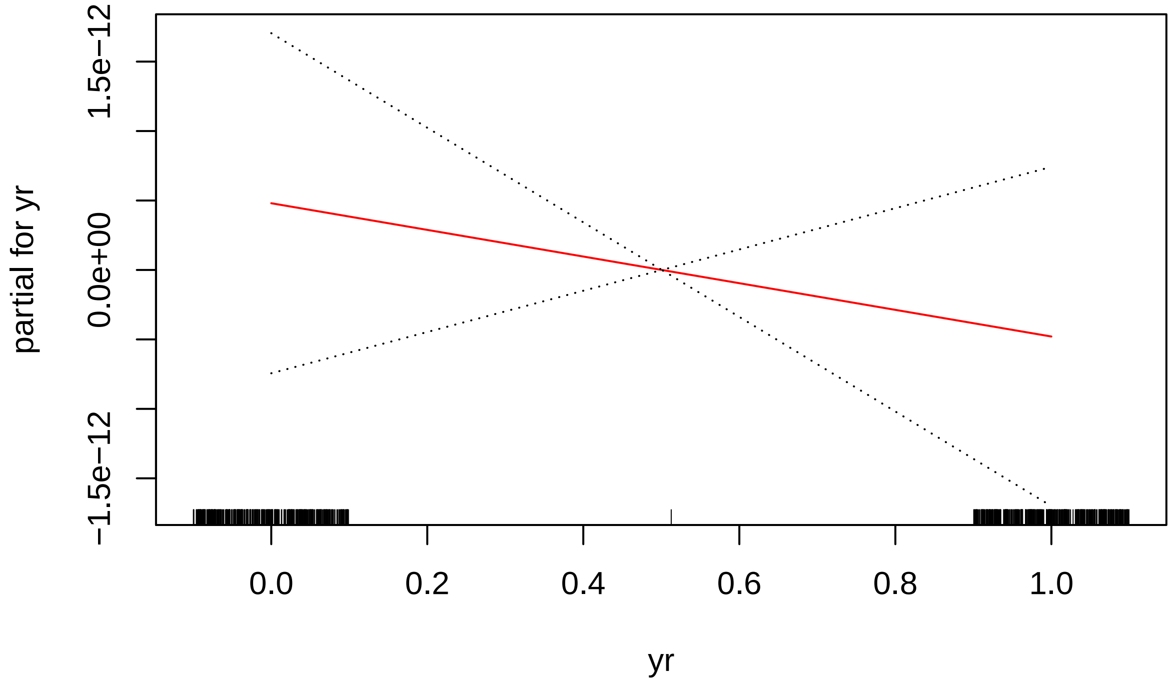

0
1
2
3
4
5
6
2e−12
0e+00
−4e−12
partial for weekday
weekday


partial for workingday
4e−12
2e−12
0e+00
−2e−12
0.0
0.2
0.4
0.6
0.8
1.0
workingday


1
2
3
partial for weathersit
2e−12
−2e−12
−6e−12
weathersit


1
2
3
4
5
6
7
8
9
10
11
12
partial for mnth
2e−12
−2e−12
−6e−12
mnth


partial for holiday
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0
holiday



```r
summary(gam1)
```

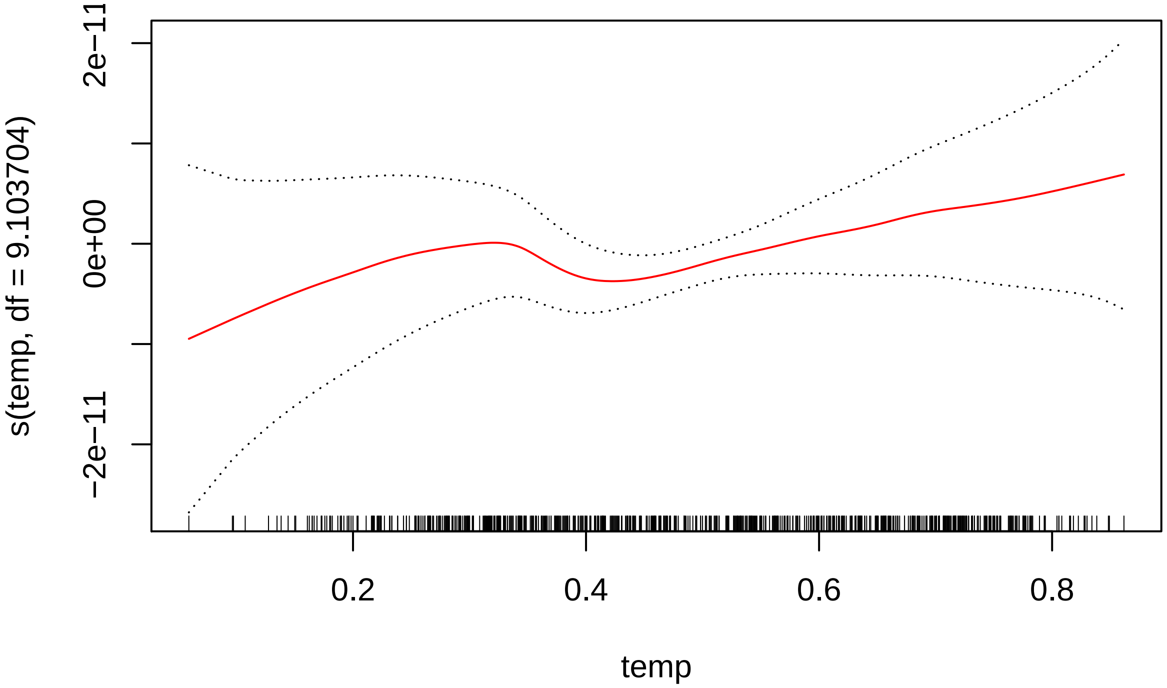
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable

##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday + workingday + weathersit + mnth + holiday + yr,
##     data = day)
## Deviance Residuals:
##        Min         1Q     Median         3Q        Max
## -7.935e-11 -2.728e-12 -4.547e-13  1.819e-12  2.018e-10
##
## (Dispersion Parameter for gaussian family taken to be 0)
##
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 0 on 653.2976 degrees of freedom
## AIC: -34989.29
##
## Number of Local Scoring Iterations: 1
##
## Anova for Parametric Effects
##                                Df     Sum Sq    Mean Sq    F value Pr(>F)
## s(temp, df = 9.103704)        1.0 1078688585 1078688585 1.2519e+31 <2e-16
## s(windspeed, df = 6.007664)   1.0   51536710   51536710 5.9812e+29 <2e-16
## s(atemp, df = 8.805497)       1.0    4387703    4387703 5.0923e+28 <2e-16
## s(hum, df = 4.548876)         1.0  136071493  136071493 1.5792e+30 <2e-16
## s(casual, df = 11.27571)      1.0  324226292  324226292 3.7629e+30 <2e-16
## s(registered, df = 12.95976)  1.0 1144624609 1144624609 1.3284e+31 <2e-16
## season                        3.0          0          0 3.5190e-01 0.7878
## weekday                       6.0          0          0 8.7700e-02 0.9975
## workingday                     1.0          0          0 8.5200e-02 0.7705
## weathersit                     2.0          0          0 1.3026e+00 0.2725
## mnth                          11.0          0          0 3.8810e-01 0.9609
## yr                             1.0          0          0 6.1590e-01 0.4328
## Residuals                    653.3          0          0
##
## s(temp, df = 9.103704)       ***
## s(windspeed, df = 6.007664)  ***
## s(atemp, df = 8.805497)      ***
## s(hum, df = 4.548876)        ***
## s(casual, df = 11.27571)     ***
## s(registered, df = 12.95976) ***
## season
## weekday
## workingday
## weathersit
## mnth
## yr
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                              Npar Df  Npar F   Pr(F)
## (Intercept)
## s(temp, df = 9.103704)           8.1 2.07137 0.03577 *
## s(windspeed, df = 6.007664)      5.0 0.87862 0.49508
## s(atemp, df = 8.805497)          7.8 2.47375 0.01277 *
## s(hum, df = 4.548876)            3.5 1.33475 0.25872
## s(casual, df = 11.27571)        10.3 1.64932 0.08679 .
## s(registered, df = 12.95976)    12.0 1.13852 0.32543
## season
## weekday
## workingday
## weathersit
## mnth
## holiday
## yr
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

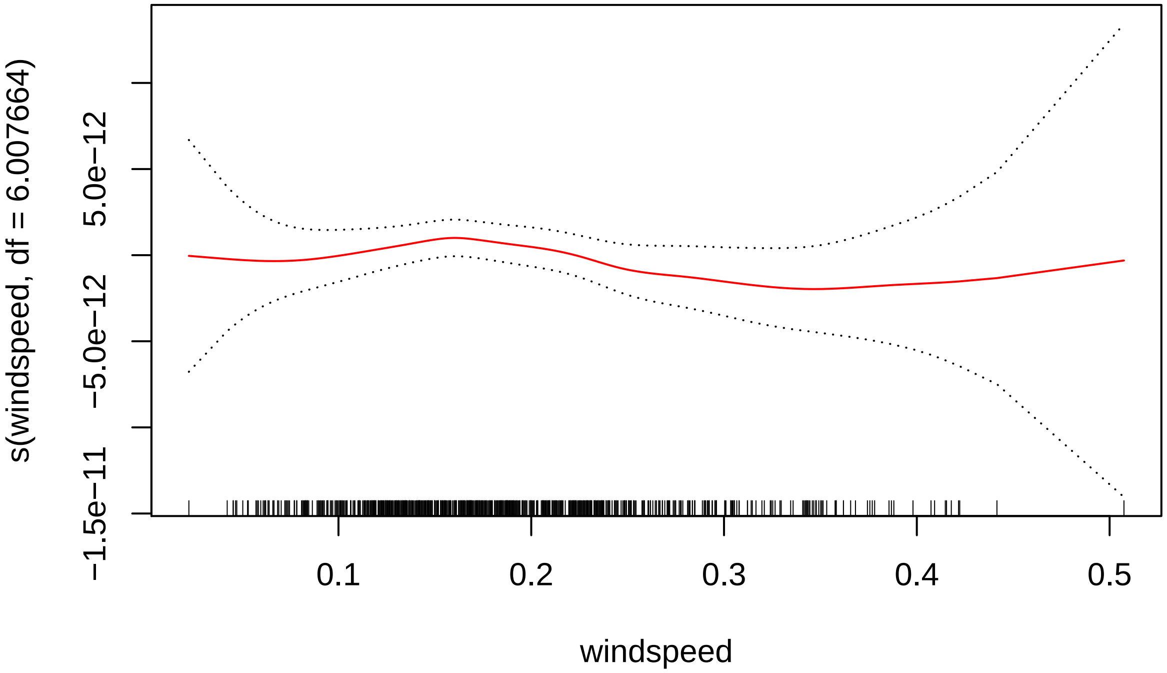
```
#Ahora voy a realizar mas modelos GAM quitando las variables menos significativas
  #Sin mnth, weathersit, holiday
gam2 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, df=4.5
```

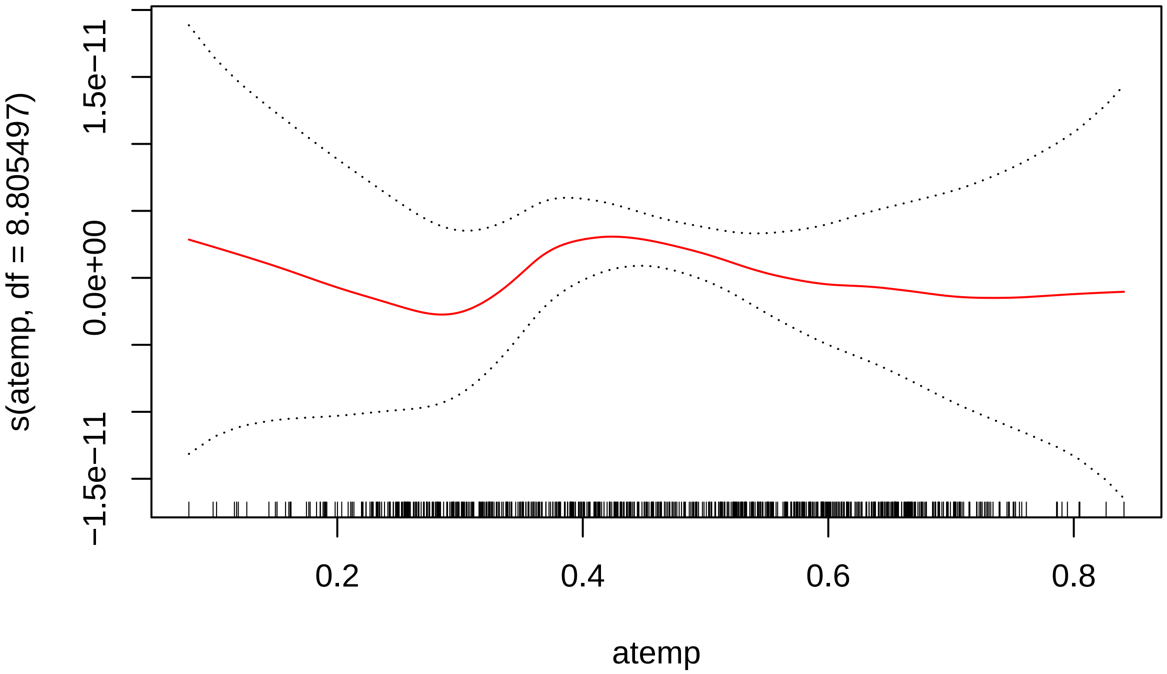
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

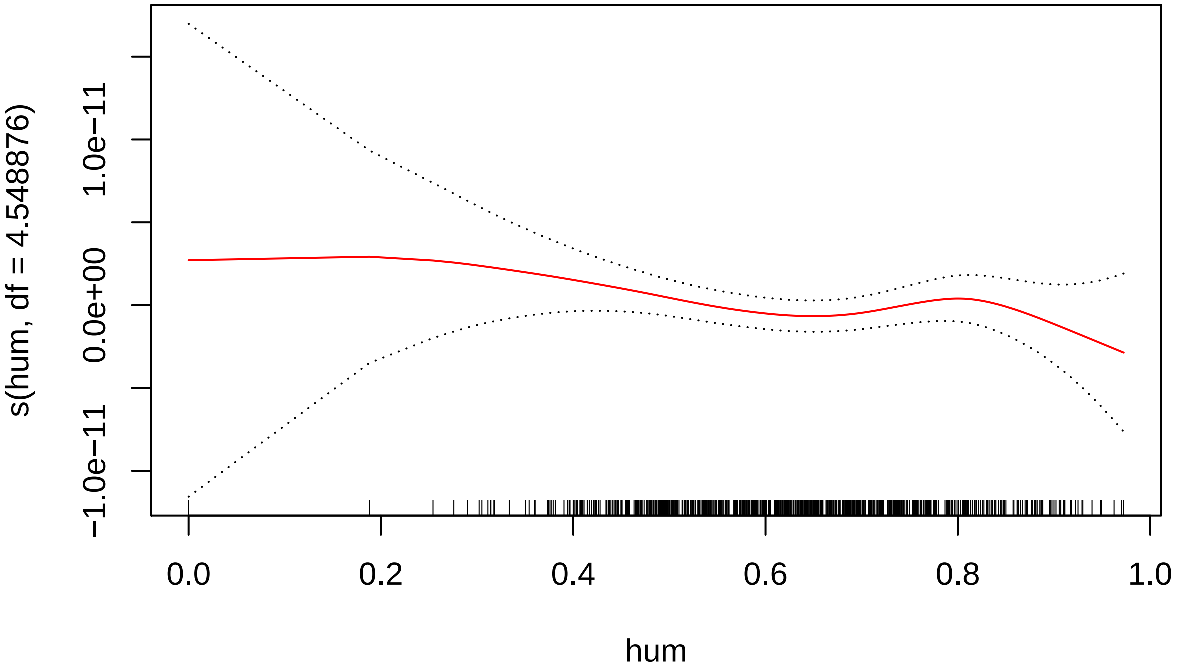
```
plot(gam2, se=TRUE, col='red')
```

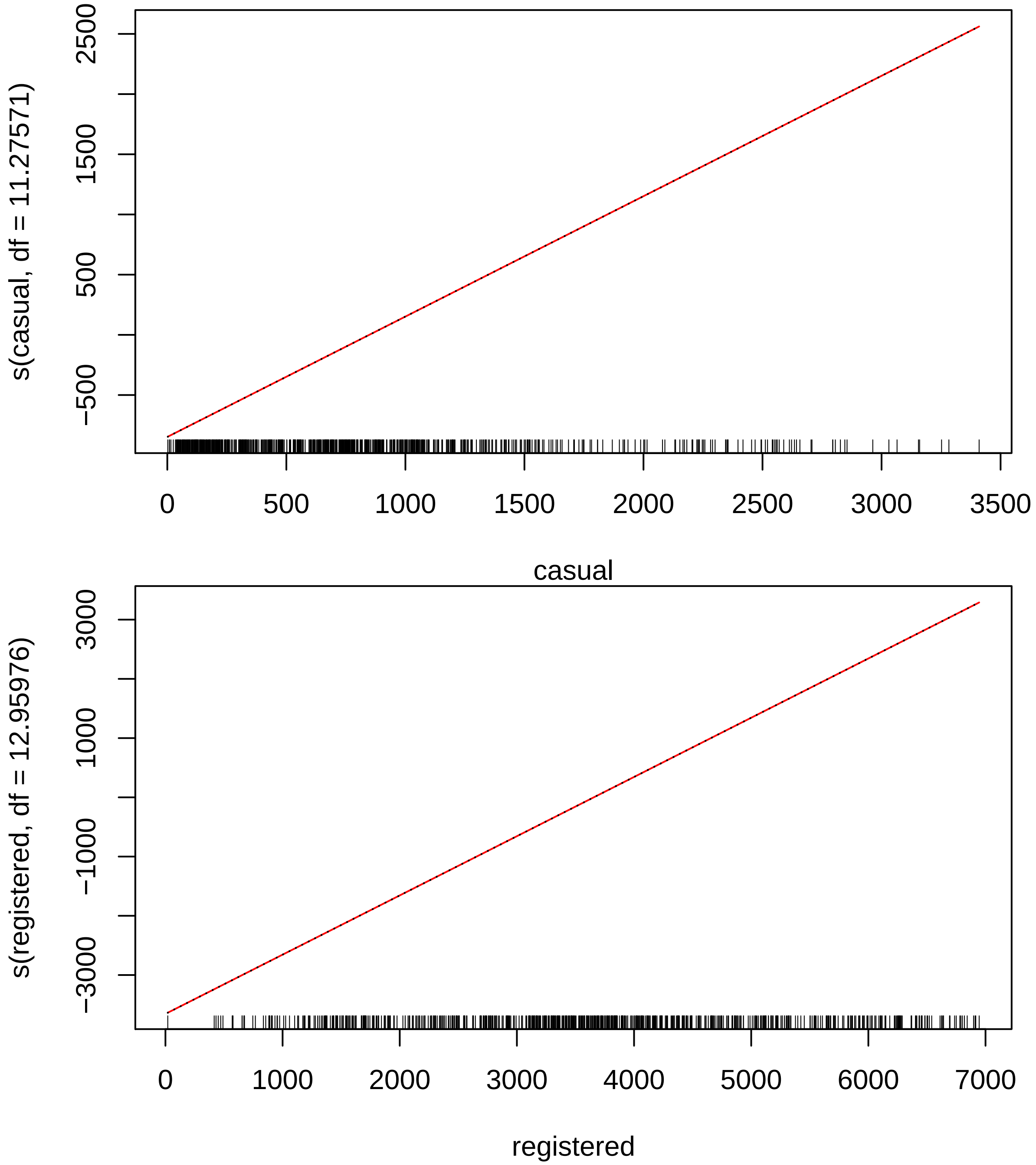
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

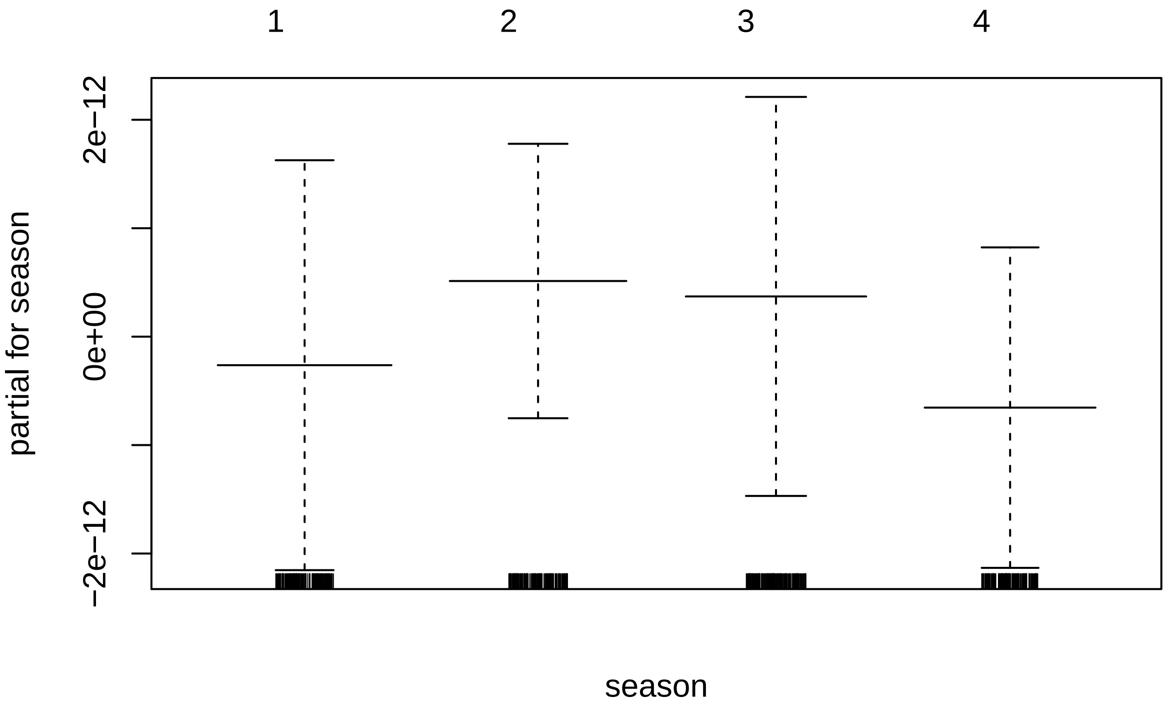

s(temp, df = 9.103704)
2e−11
0e+00
−2e−11
0.2
0.4
0.6
0.8
temp

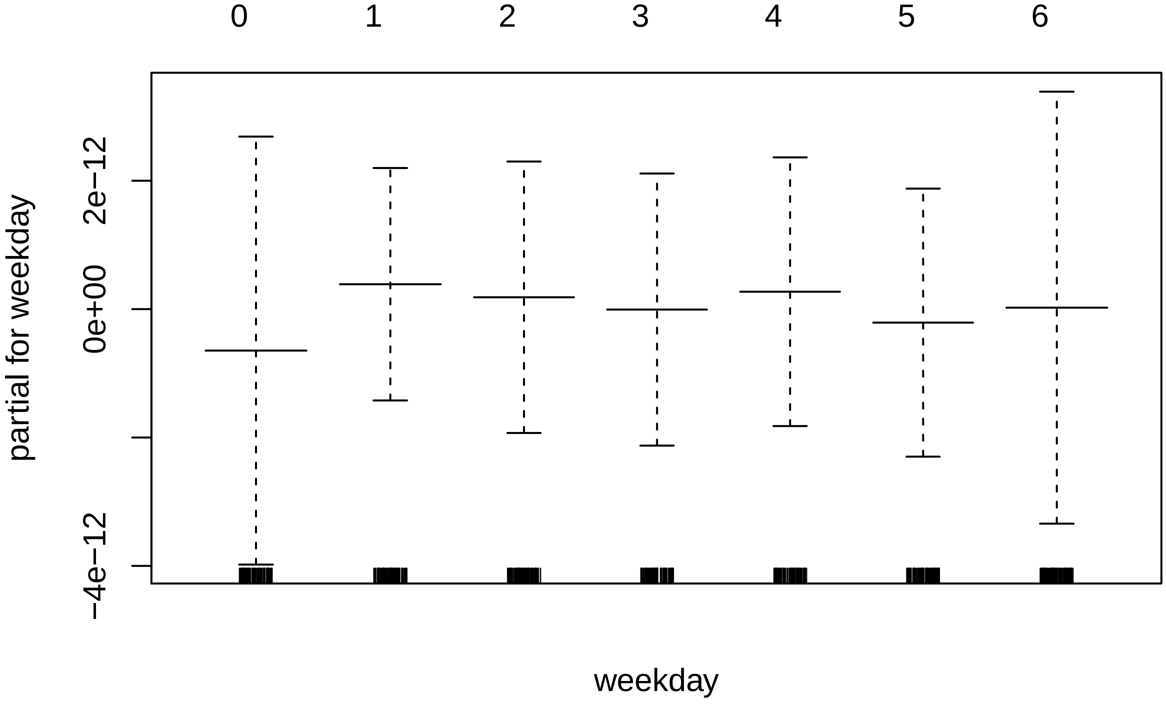

s(windspeed, df = 6.007664)
5.0e−12
−5.0e−12
−1.5e−11
0.1
0.2
0.3
0.4
0.5
windspeed

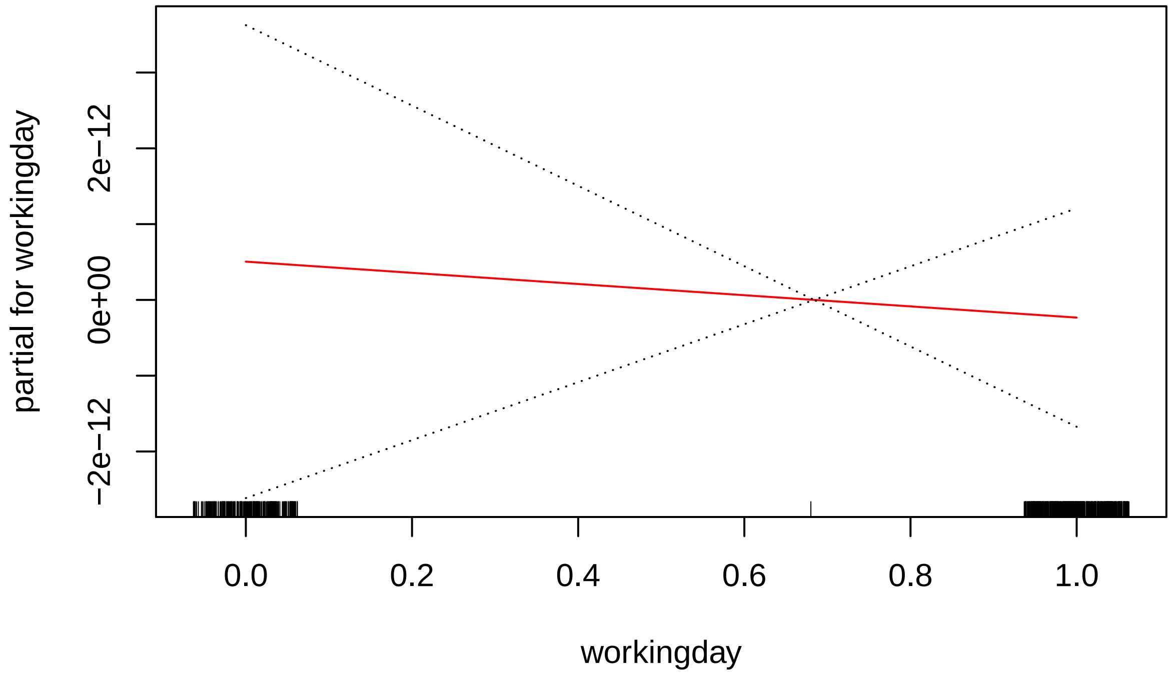

s(atemp, df = 8.805497)
−1.5e−11
0.0e+00
1.5e−11
0.2
0.4
0.6
0.8
atemp

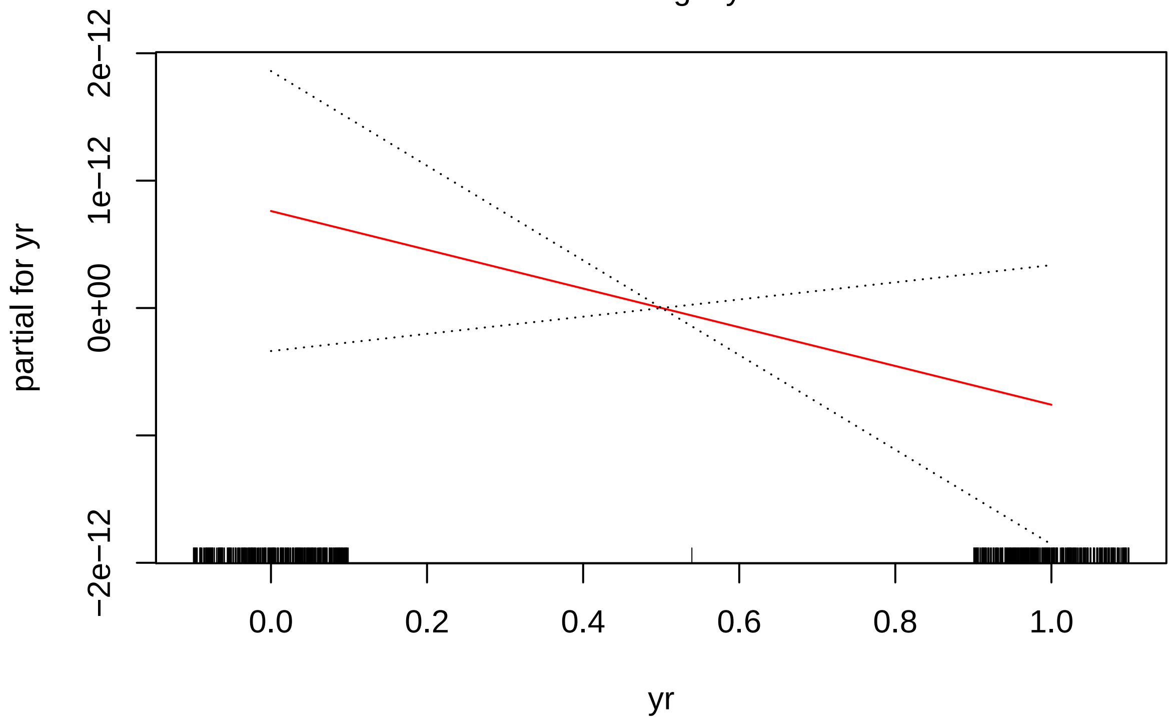

s(hum, df = 4.548876)
−1.0e−11
0.0e+00
1.0e−11
0.0
0.2
0.4
0.6
0.8
1.0
hum


s(casual, df = 11.27571)
−500
500
1500
2500
0
500
1000
1500
2000
2500
3000
3500
casual
s(registered, df = 12.95976)
−3000
−1000
1000
3000
0
1000
2000
3000
4000
5000
6000
7000
registered


1
2
3
4
2e−12
0e+00
−2e−12
partial for season
season


0
1
2
3
4
5
6
2e−12
0e+00
−4e−12
partial for weekday
weekday





```r
summary(gam2)
```

```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable

##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday + workingday + yr, data = day)
## Deviance Residuals:
```

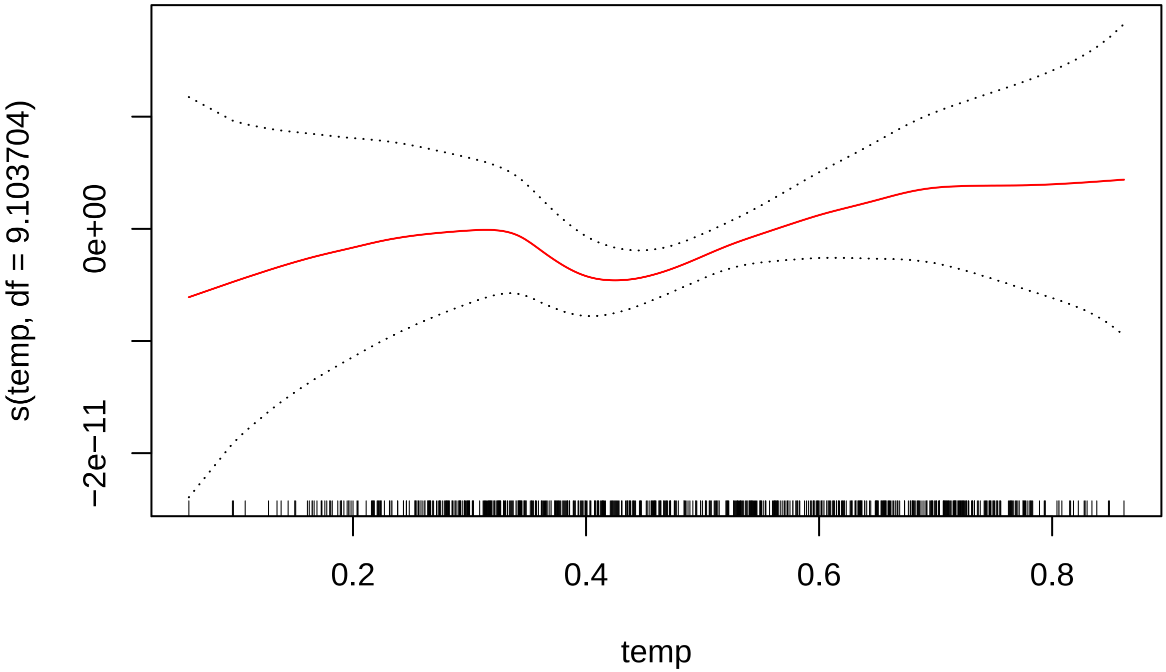
```
##        Min         1Q     Median         3Q        Max 
## -7.935e-11 -2.728e-12  0.000e+00  1.819e-12  2.010e-10 
## 
## (Dispersion Parameter for gaussian family taken to be 0)
## 
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 0 on 666.2976 degrees of freedom
## AIC: -35028.87 
## 
## Number of Local Scoring Iterations: 1 
## 
## Anova for Parametric Effects
##                                Df     Sum Sq    Mean Sq    F value Pr(>F)
## s(temp, df = 9.103704)        1.0 1078688585 1078688585 1.3007e+31 <2e-16
## s(windspeed, df = 6.007664)   1.0   51536710   51536710 6.2146e+29 <2e-16
## s(atemp, df = 8.805497)       1.0    4387703    4387703 5.2910e+28 <2e-16
## s(hum, df = 4.548876)         1.0  136071493  136071493 1.6408e+30 <2e-16
## s(casual, df = 11.27571)      1.0  324226292  324226292 3.9097e+30 <2e-16
## s(registered, df = 12.95976)  1.0 1144624609 1144624609 1.3803e+31 <2e-16
## season                        3.0          0          0 3.6730e-01 0.7767
## weekday                       6.0          0          0 1.1750e-01 0.9943
## workingday                    1.0          0          0 4.0000e-03 0.9498
## yr                            1.0          0          0 1.9198e+00 0.1663
## Residuals                   666.3          0          0                  
##                                 
## s(temp, df = 9.103704)       ***
## s(windspeed, df = 6.007664)  ***
## s(atemp, df = 8.805497)      ***
## s(hum, df = 4.548876)        ***
## s(casual, df = 11.27571)     ***
## s(registered, df = 12.95976) ***
## season                          
## weekday                         
## workingday                      
## yr                              
## Residuals                       
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Anova for Nonparametric Effects
##                              Npar Df  Npar F    Pr(F)   
## (Intercept)                                             
## s(temp, df = 9.103704)           8.1 2.78586 0.004679 **
## s(windspeed, df = 6.007664)      5.0 0.91679 0.469560   
## s(atemp, df = 8.805497)          7.8 3.14694 0.001842 **
## s(hum, df = 4.548876)            3.5 1.62127 0.174263   
## s(casual, df = 11.27571)        10.3 1.68968 0.077037 . 
## s(registered, df = 12.95976)    12.0 1.16473 0.304975   
## season                                                  
## weekday                                                 
## workingday                                              
## yr                                                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

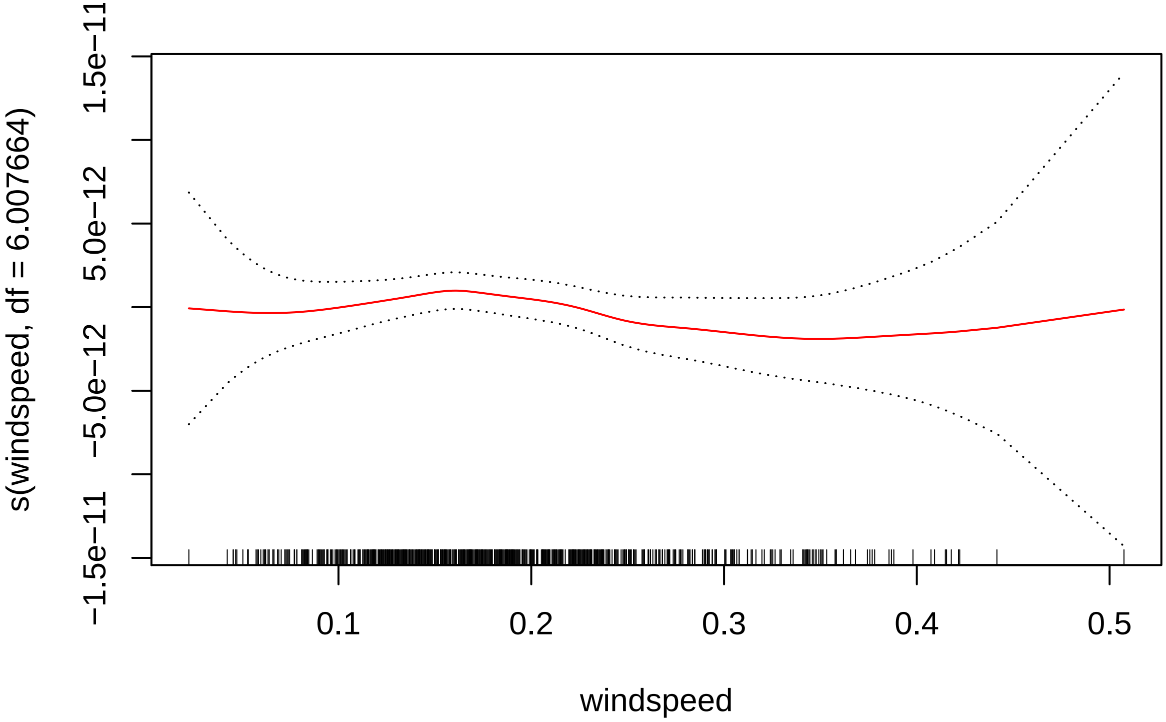
```
#Sin workingday,yr
gam3 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, df=4.54
```

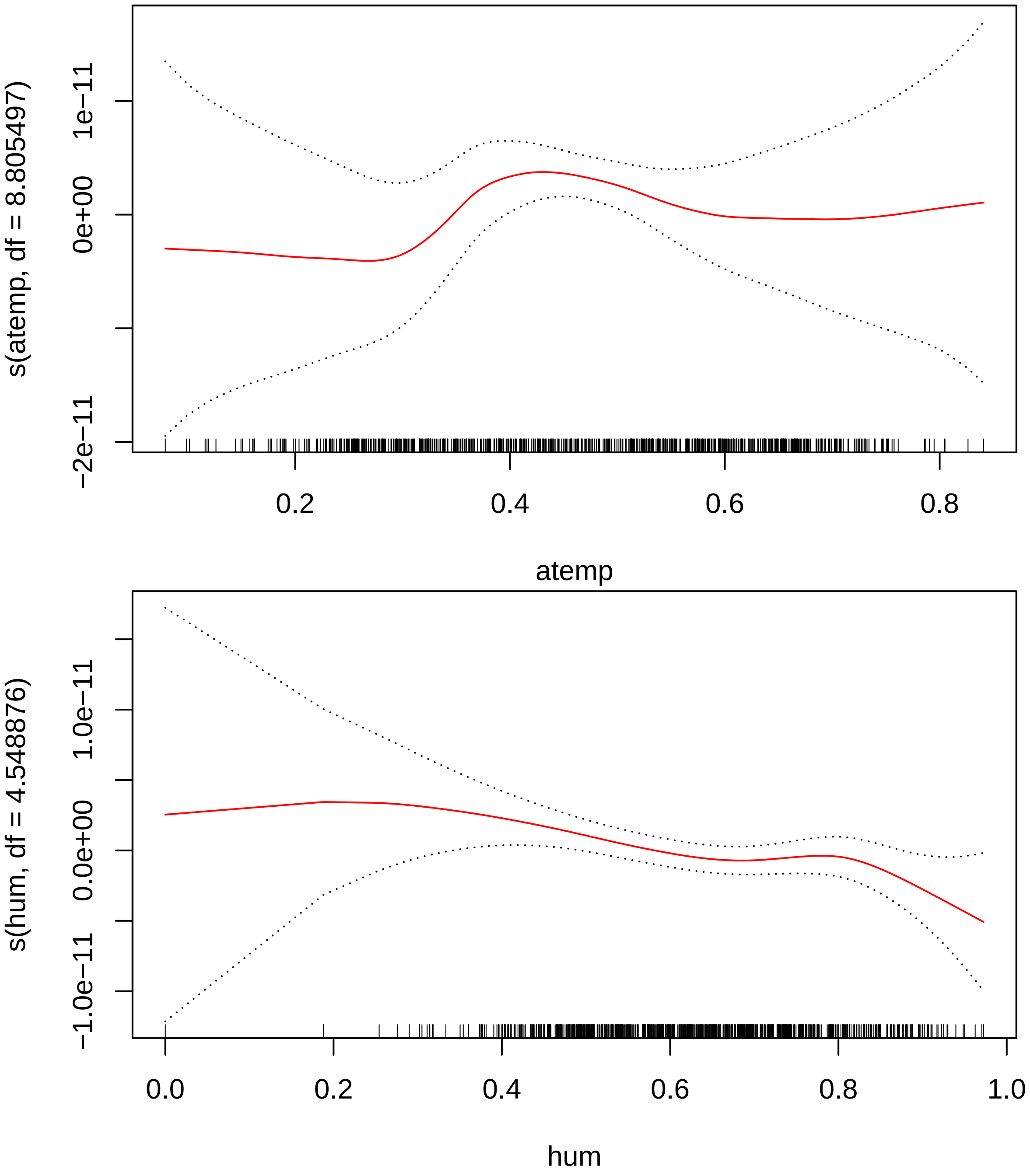
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

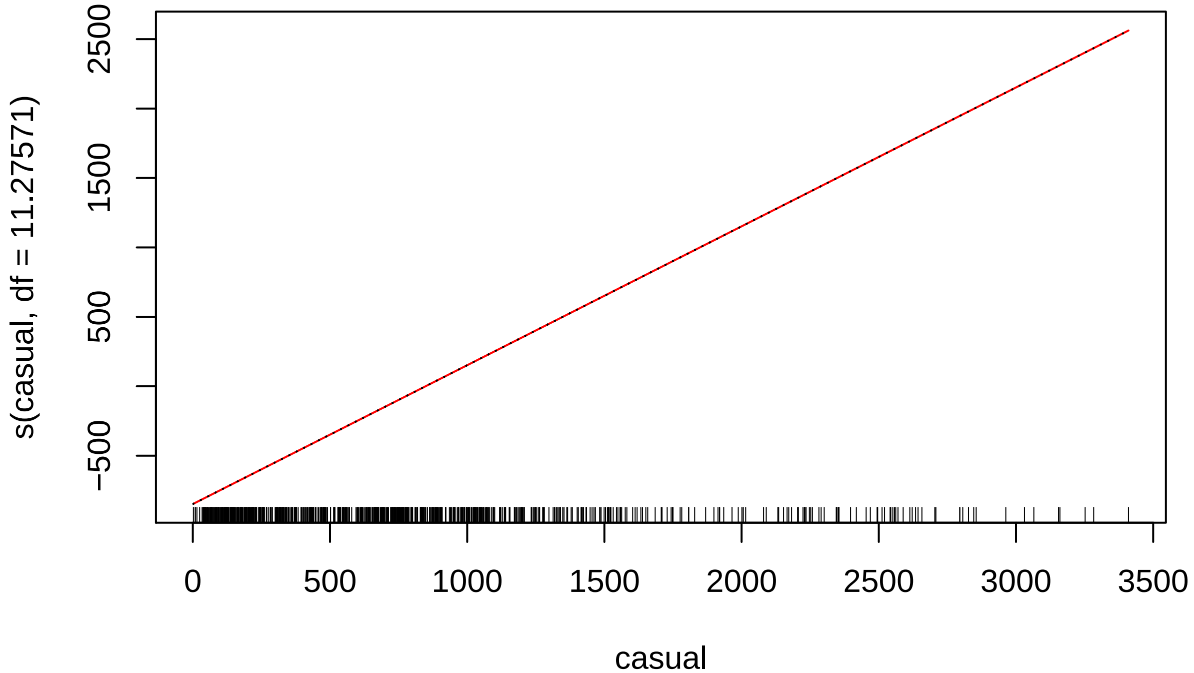
```
plot(gam3, se=TRUE, col='red')
```

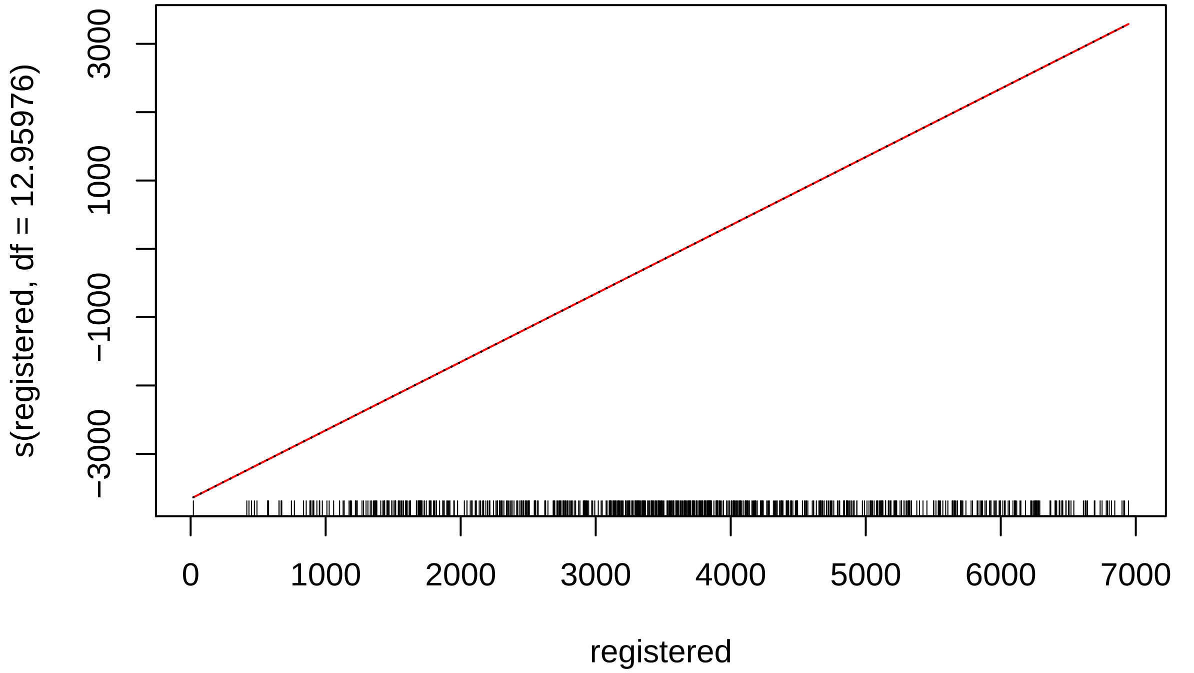
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

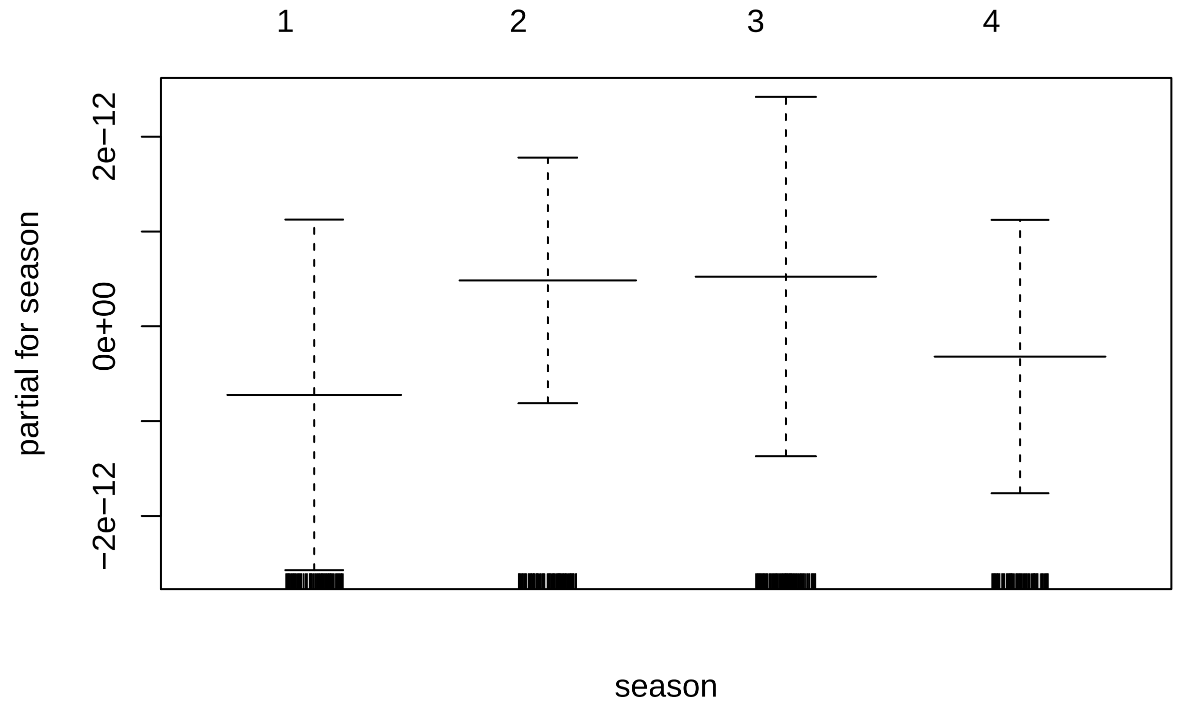


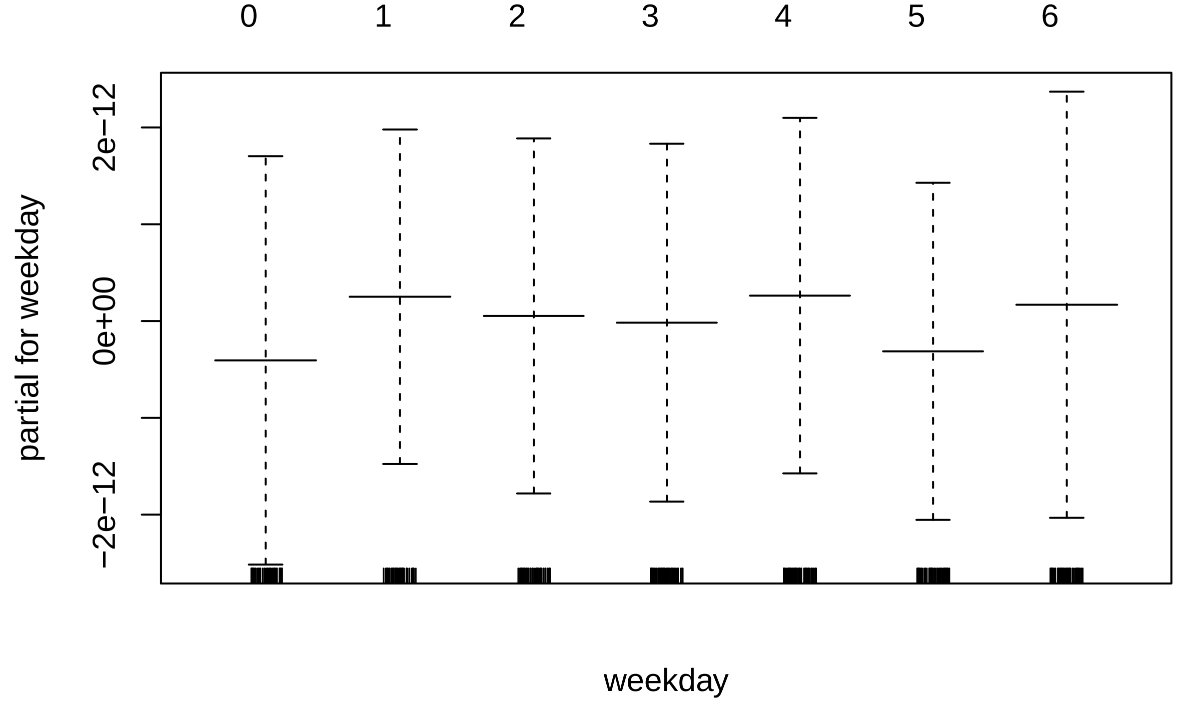



s(atemp, df = 8.805497)
1e−11
0e+00
−2e−11
0.2
0.4
0.6
0.8
atemp
s(hum, df = 4.548876)
1.0e−11
0.0e+00
−1.0e−11
0.0
0.2
0.4
0.6
0.8
1.0
hum


s(casual, df = 11.27571)
−500
500
1500
2500
0
500
1000
1500
2000
2500
3000
3500
casual


s(registered, df = 12.95976)
−3000
−1000
1000
3000
0
1000
2000
3000
4000
5000
6000
7000
registered





```r
summary(gam3)
```

```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

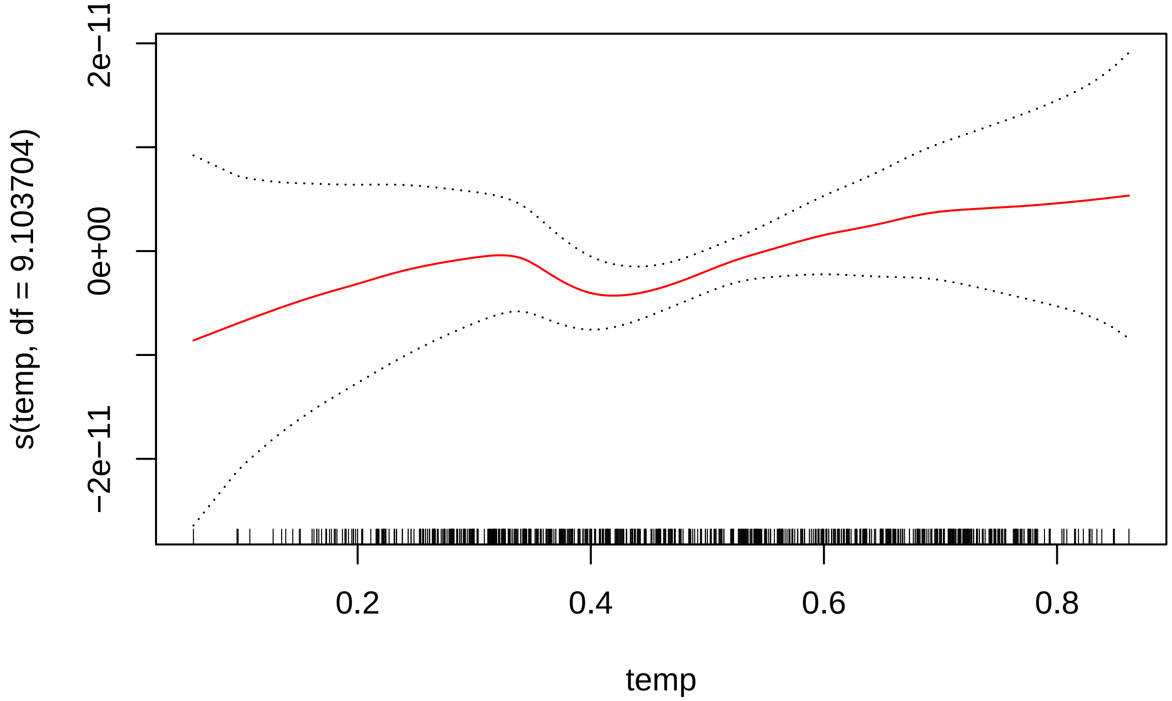
```
##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday, data = day)
## Deviance Residuals:
##        Min         1Q     Median         3Q        Max
## -7.321e-11 -2.728e-12 -4.547e-13  1.819e-12  2.126e-10
##
## (Dispersion Parameter for gaussian family taken to be 0)
##
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 0 on 668.2976 degrees of freedom
## AIC: -34985.94
##
## Number of Local Scoring Iterations: 1
##
## Anova for Parametric Effects
##                                  Df     Sum Sq    Mean Sq    F value Pr(>F)
## s(temp, df = 9.103704)          1.0 1078688585 1078688585 1.2235e+31 <2e-16
## s(windspeed, df = 6.007664)     1.0   51536710   51536710 5.8457e+29 <2e-16
## s(atemp, df = 8.805497)         1.0    4387703    4387703 4.9769e+28 <2e-16
## s(hum, df = 4.548876)           1.0  136071493  136071493 1.5434e+30 <2e-16
## s(casual, df = 11.27571)        1.0  324226292  324226292 3.6776e+30 <2e-16
## s(registered, df = 12.95976)    1.0 1144624609 1144624609 1.2983e+31 <2e-16
## season                          3.0          0          0 3.1690e-01 0.8132
## weekday                         6.0          0          0 7.9300e-02 0.9981
## Residuals                     668.3          0          0
##
## s(temp, df = 9.103704)        ***
## s(windspeed, df = 6.007664)   ***
## s(atemp, df = 8.805497)       ***
## s(hum, df = 4.548876)         ***
## s(casual, df = 11.27571)      ***
## s(registered, df = 12.95976) ***
## season
## weekday
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                               Npar Df Npar F     Pr(F)
## (Intercept)
## s(temp, df = 9.103704)            8.1 3.4153 0.0006916 ***
## s(windspeed, df = 6.007664)       5.0 0.8532 0.5125055
## s(atemp, df = 8.805497)           7.8 4.2134 7.203e-05 ***
## s(hum, df = 4.548876)             3.5 1.5106 0.2033569
## s(casual, df = 11.27571)         10.3 1.7270 0.0689307 .
## s(registered, df = 12.95976)     12.0 1.2635 0.2359781
## season
## weekday
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

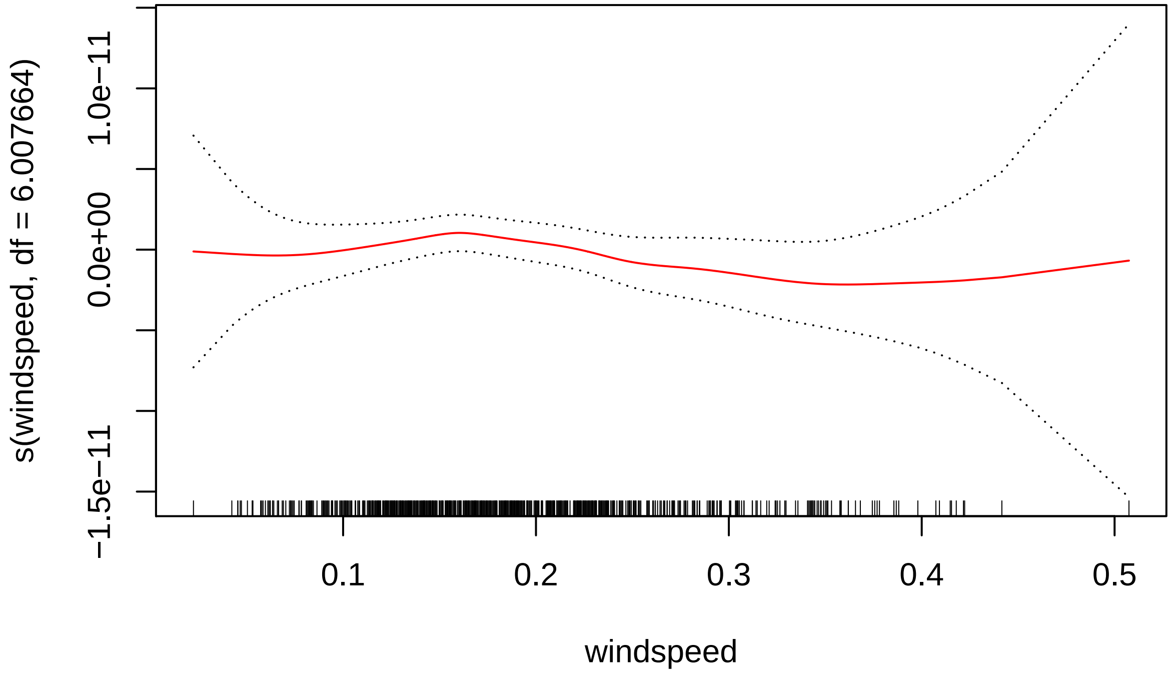
```
#Sin season y weekday
gam4 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, df=4.5
```

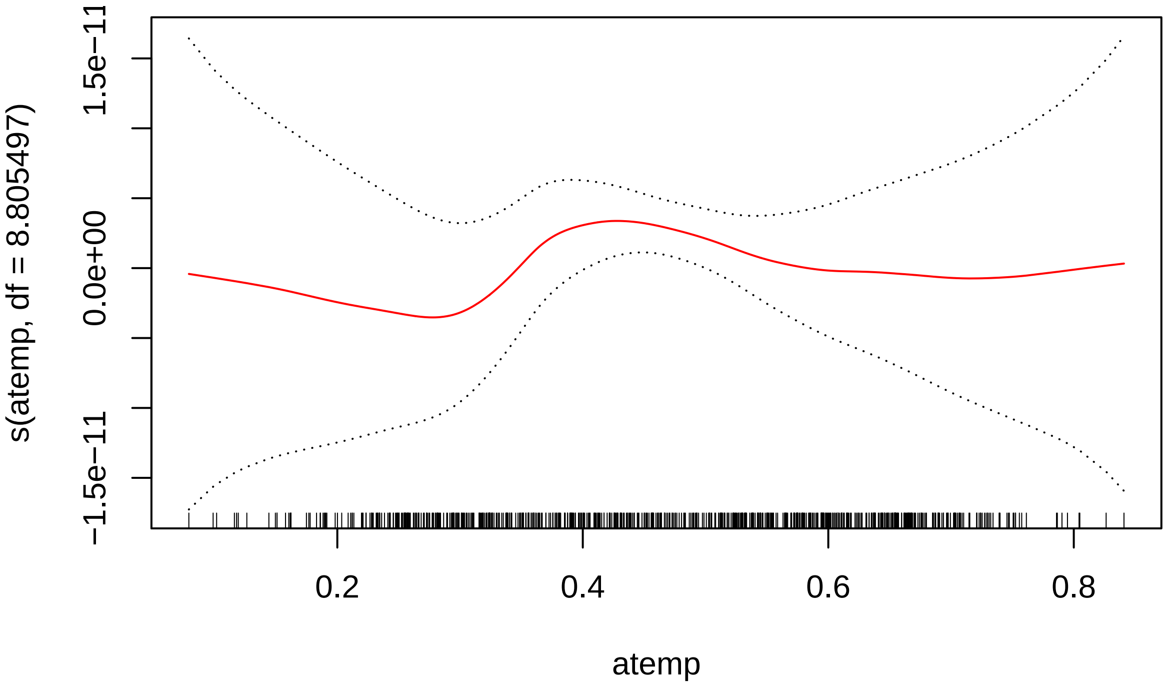
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

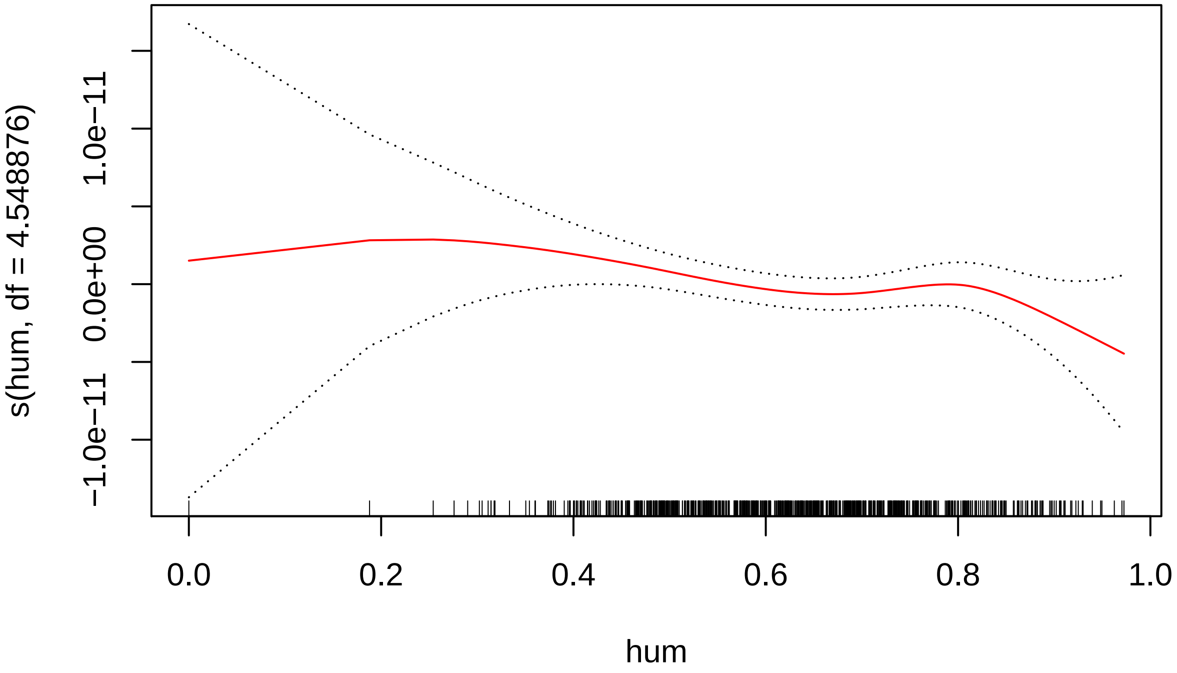
```r
plot(gam4, se=TRUE, col='red')
```

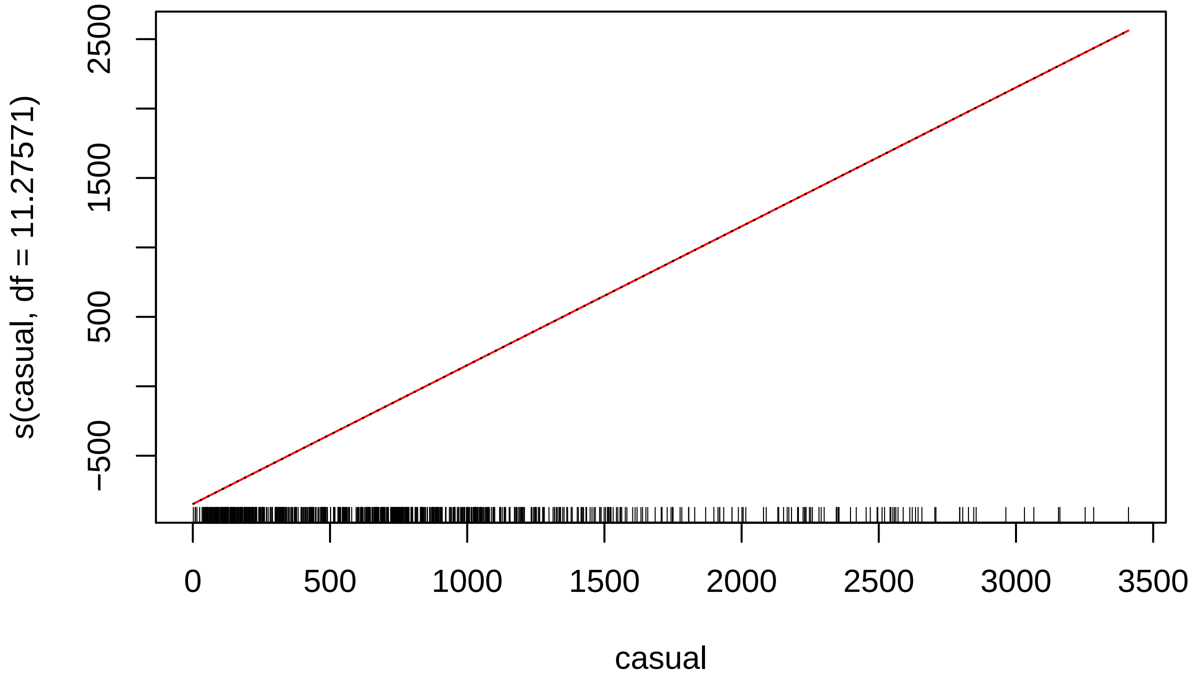
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

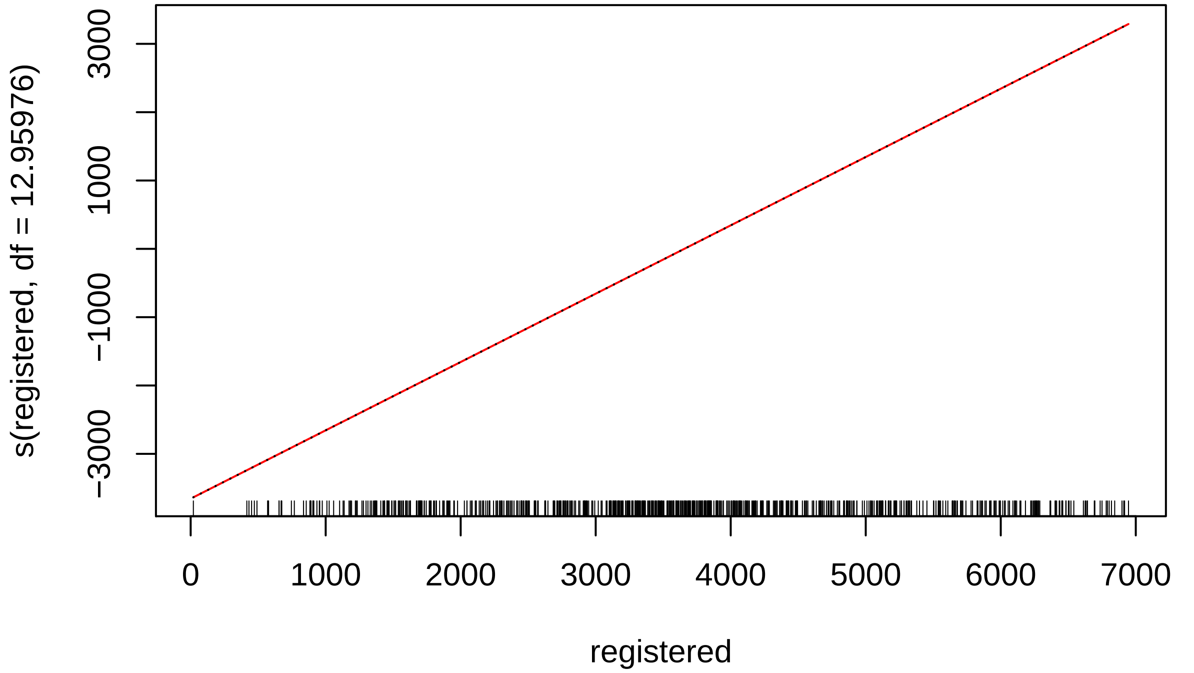





s(atemp, df = 8.805497)
1.5e−11
0.0e+00
−1.5e−11
0.2
0.4
0.6
0.8
atemp


s(hum, df = 4.548876)
1.0e−11
0.0e+00
−1.0e−11
0.0
0.2
0.4
0.6
0.8
1.0
hum





```r
summary(gam4)
```

```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

```
##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976), data = day)
## Deviance Residuals:
##        Min         1Q     Median         3Q        Max
## -7.367e-11 -2.728e-12  0.000e+00  1.819e-12  2.256e-10
```

```
##
## (Dispersion Parameter for gaussian family taken to be 0)
##
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 0 on 677.2976 degrees of freedom
## AIC: -34940.33
##
## Number of Local Scoring Iterations: 1
##
## Anova for Parametric Effects
##                                 Df     Sum Sq    Mean Sq    F value
## s(temp, df = 9.103704)         1.0 1078688585 1078688585 1.1367e+31
## s(windspeed, df = 6.007664)    1.0   51536710   51536710 5.4307e+29
## s(atemp, df = 8.805497)        1.0    4387703    4387703 4.6236e+28
## s(hum, df = 4.548876)          1.0  136071493  136071493 1.4339e+30
## s(casual, df = 11.27571)       1.0  324226292  324226292 3.4165e+30
## s(registered, df = 12.95976)   1.0 1144624609 1144624609 1.2062e+31
## Residuals                    677.3          0          0
##                                 Pr(>F)
## s(temp, df = 9.103704)       < 2.2e-16 ***
## s(windspeed, df = 6.007664)  < 2.2e-16 ***
## s(atemp, df = 8.805497)      < 2.2e-16 ***
## s(hum, df = 4.548876)        < 2.2e-16 ***
## s(casual, df = 11.27571)     < 2.2e-16 ***
## s(registered, df = 12.95976) < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                              Npar Df Npar F     Pr(F)
## (Intercept)
## s(temp, df = 9.103704)           8.1 2.7076 0.0058866 **
## s(windspeed, df = 6.007664)      5.0 0.8898 0.4875272
## s(atemp, df = 8.805497)          7.8 3.7267 0.0003217 ***
## s(hum, df = 4.548876)            3.5 1.6142 0.1759883
## s(casual, df = 11.27571)        10.3 1.6160 0.0955072 .
## s(registered, df = 12.95976)    12.0 1.3014 0.2127985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#ANOVA

```
#ANOVA
  #Realizamos el test anova para comparar los 4 modelos que hemos propuesto anteriormente
  #Podemos comprobar que el que menor residuo tiene es el modelo 1 por lo que va a ser
  #el modelo con el que vamos a trabajar.
anova(gam1, gam2, gam3, gam4, test='F')
```

```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

```
## Analysis of Deviance Table
##
## Model 1: cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday + workingday + weathersit + mnth + holiday + yr
## Model 2: cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday + workingday + yr
## Model 3: cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday
## Model 4: cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976)
##   Resid. Df Resid. Dev  Df    Deviance       F    Pr(>F)
## 1     653.3 5.6291e-20
## 2     666.3 5.5255e-20 -13  1.0357e-21
## 3     668.3 5.8918e-20  -2 -3.6634e-21 21.2582 1.137e-09 ***
## 4     677.3 6.4275e-20  -9 -5.3565e-21  6.9074 1.532e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

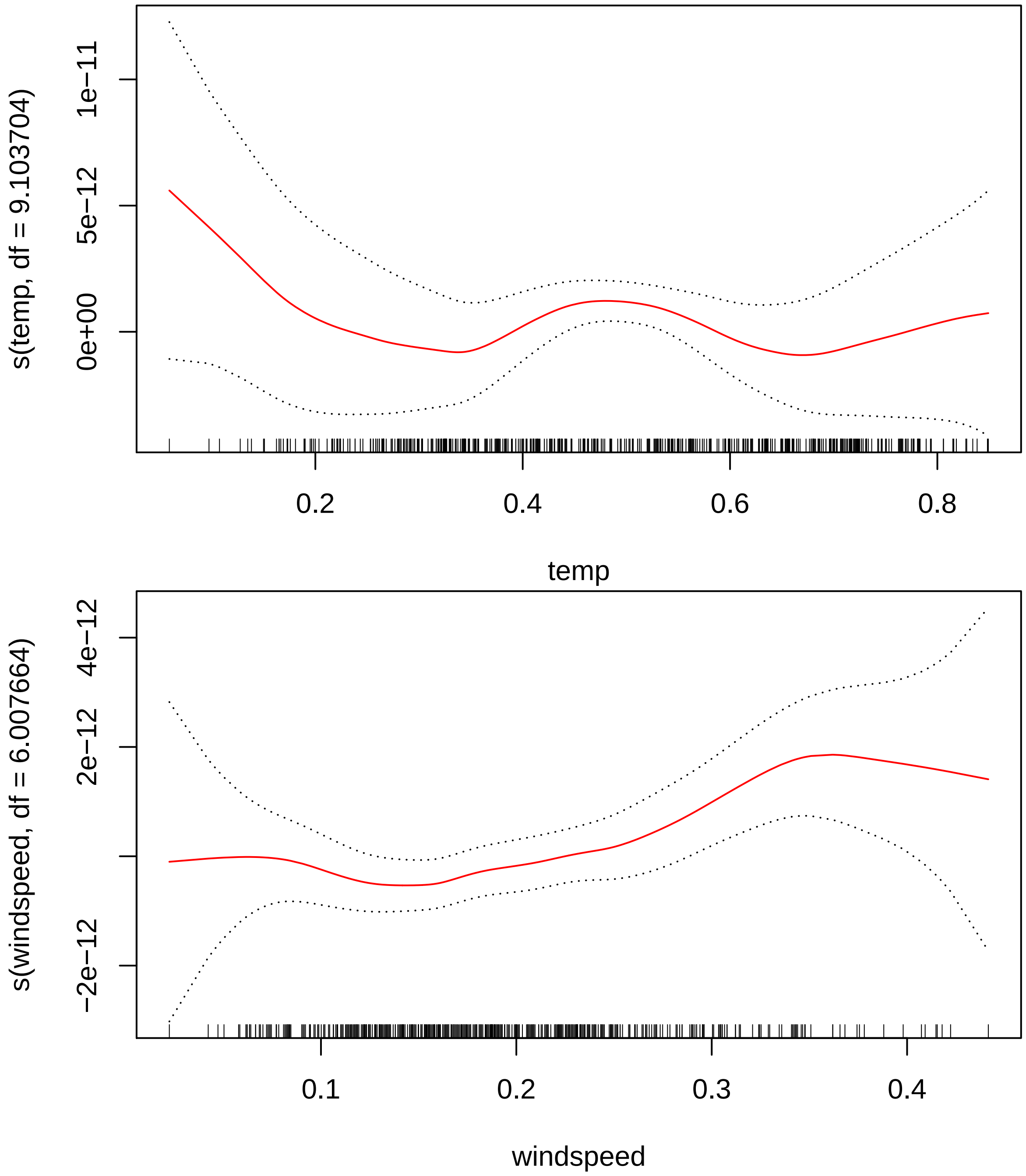
#CROSS VALIDATION

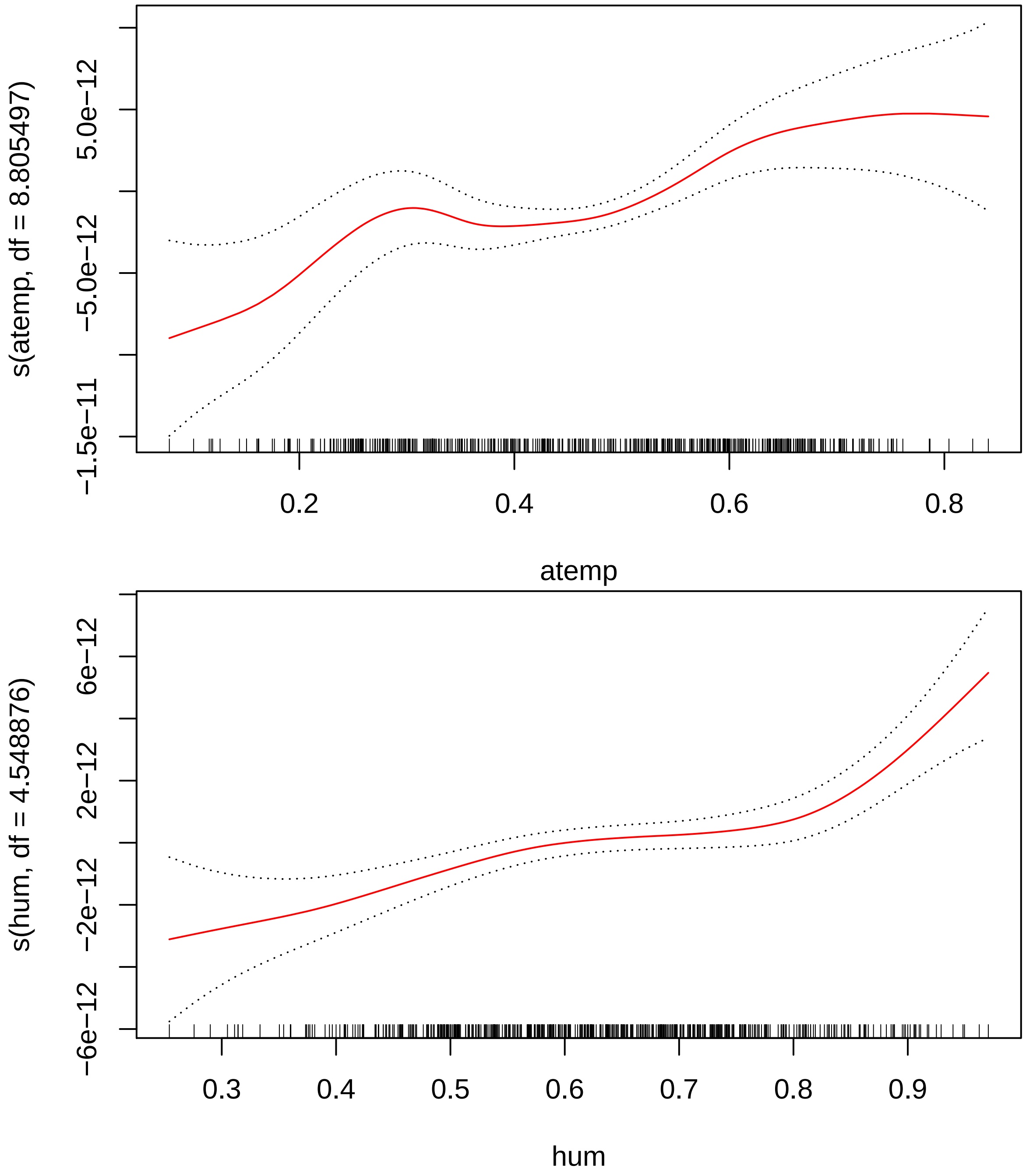
```
#Una vez escogido el modelo, vamos a proceder a dividir nuestra base de datos en
#train y test para predecir.

set.seed(123)
day_split <- initial_split(day, prop =.7, strata = "cnt")
day_train <- training(day_split)
day_test <- testing(day_split)

#Tenemos la base de datos dividida en 70/30, y vamos a proceder a introducir nuestro modelo
#en el test para saber como predice.

gam_train <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, d
```

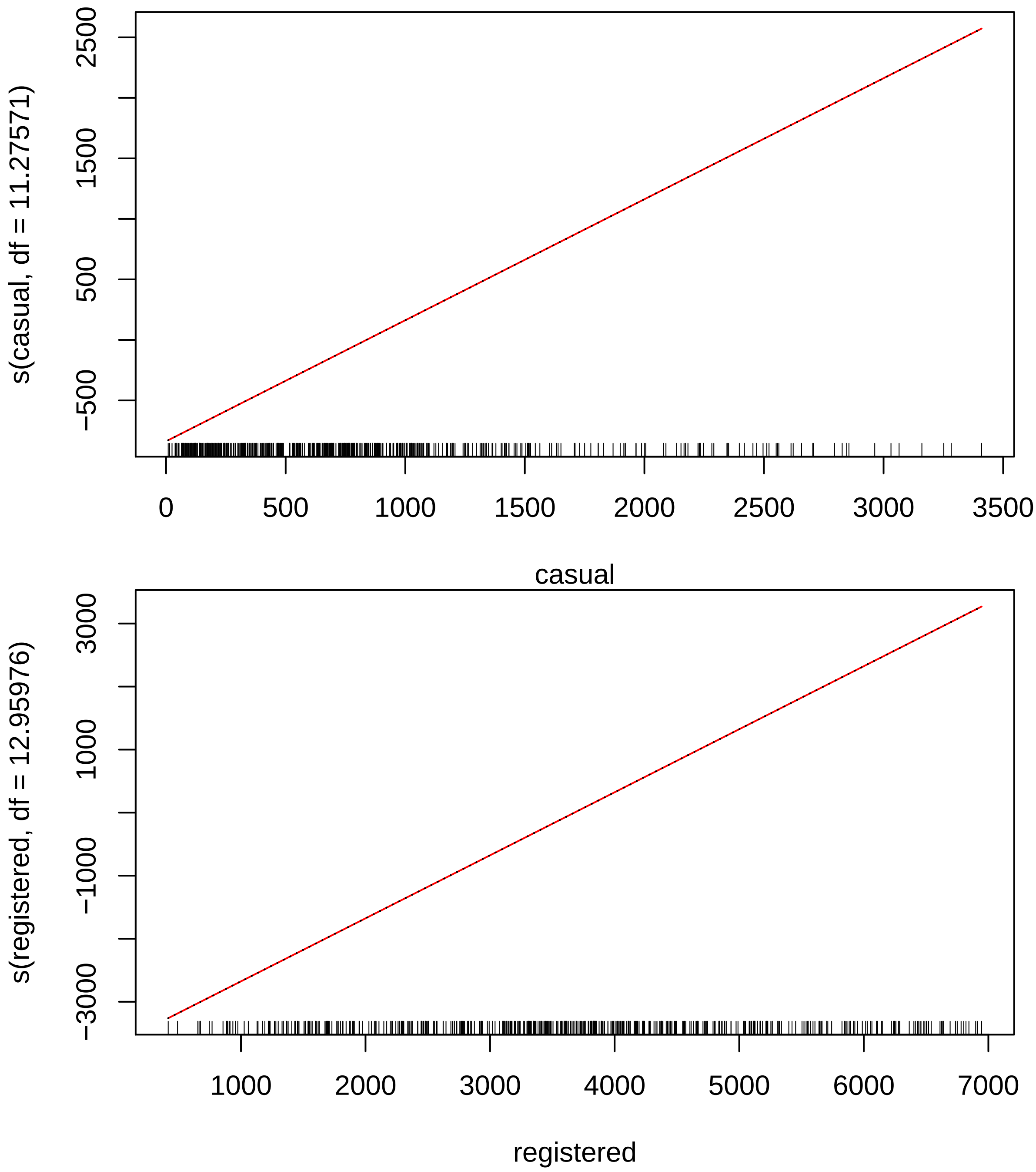
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

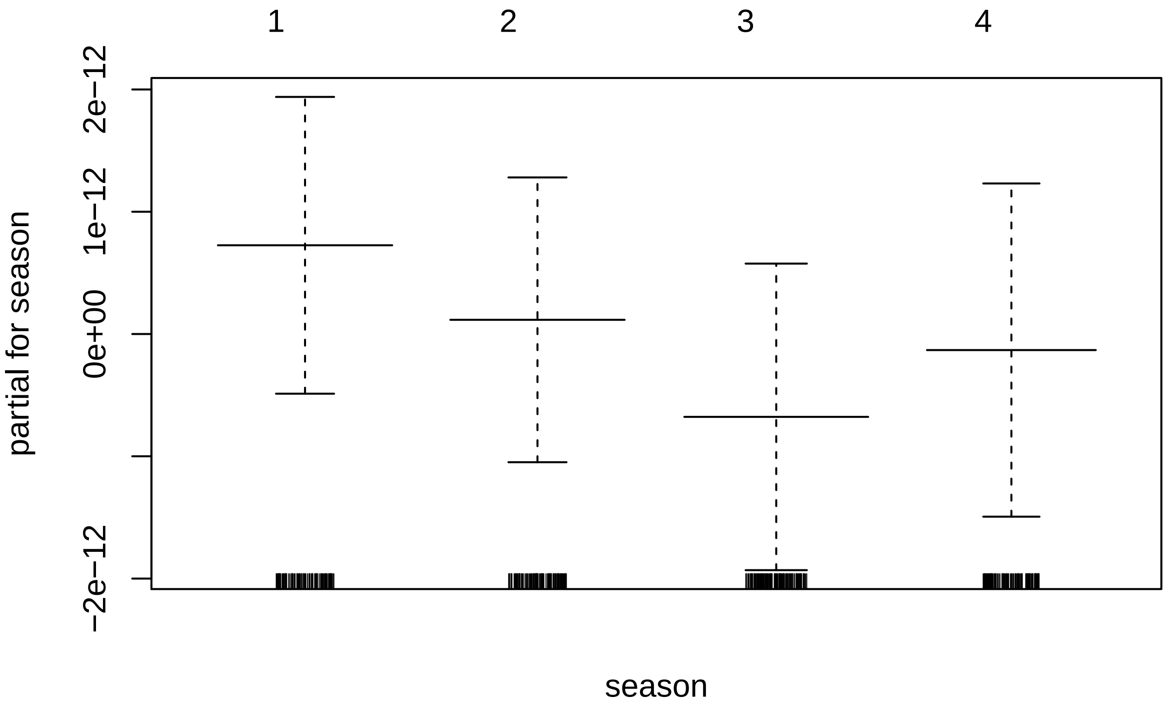
```
plot(gam_train, se=TRUE, col='red')
```

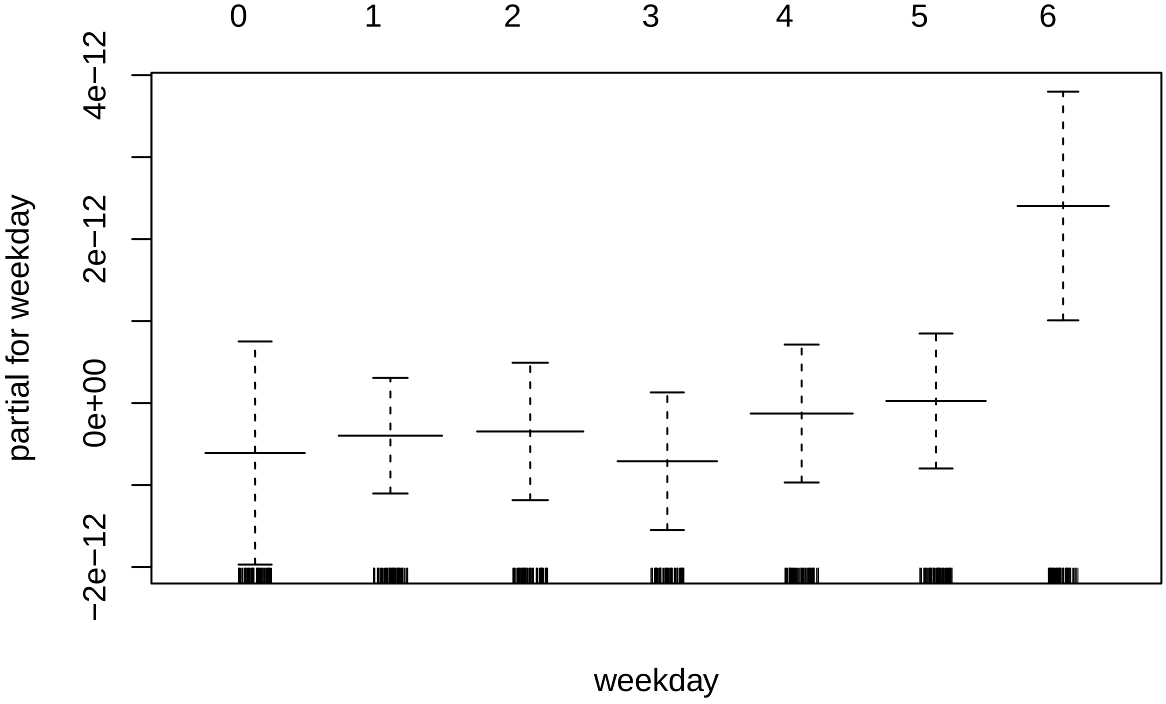
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

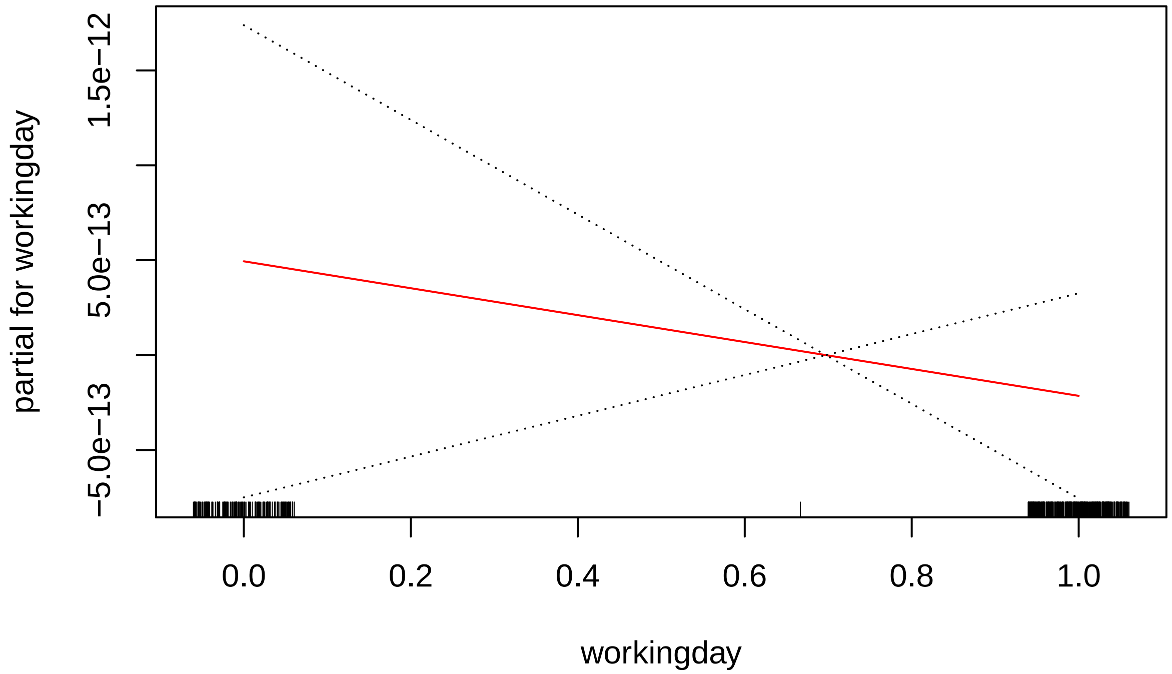

s(temp, df = 9.103704)
1e−11
5e−12
0e+00
0.2
0.4
0.6
0.8
temp
s(windspeed, df = 6.007664)
4e−12
2e−12
−2e−12
0.1
0.2
0.3
0.4
windspeed

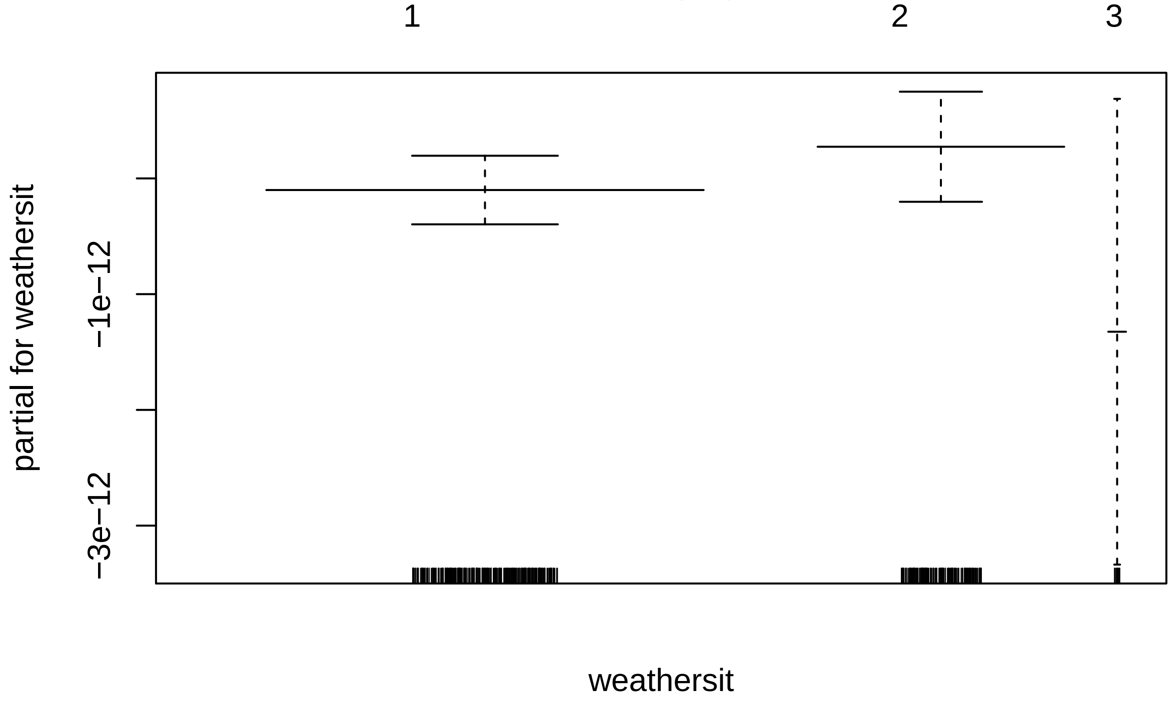

s(atemp, df = 8.805497)
−1.5e−11
−5.0e−12
5.0e−12
0.2
0.4
0.6
0.8
atemp
s(hum, df = 4.548876)
−6e−12
−2e−12
2e−12
6e−12
0.3
0.4
0.5
0.6
0.7
0.8
0.9
hum

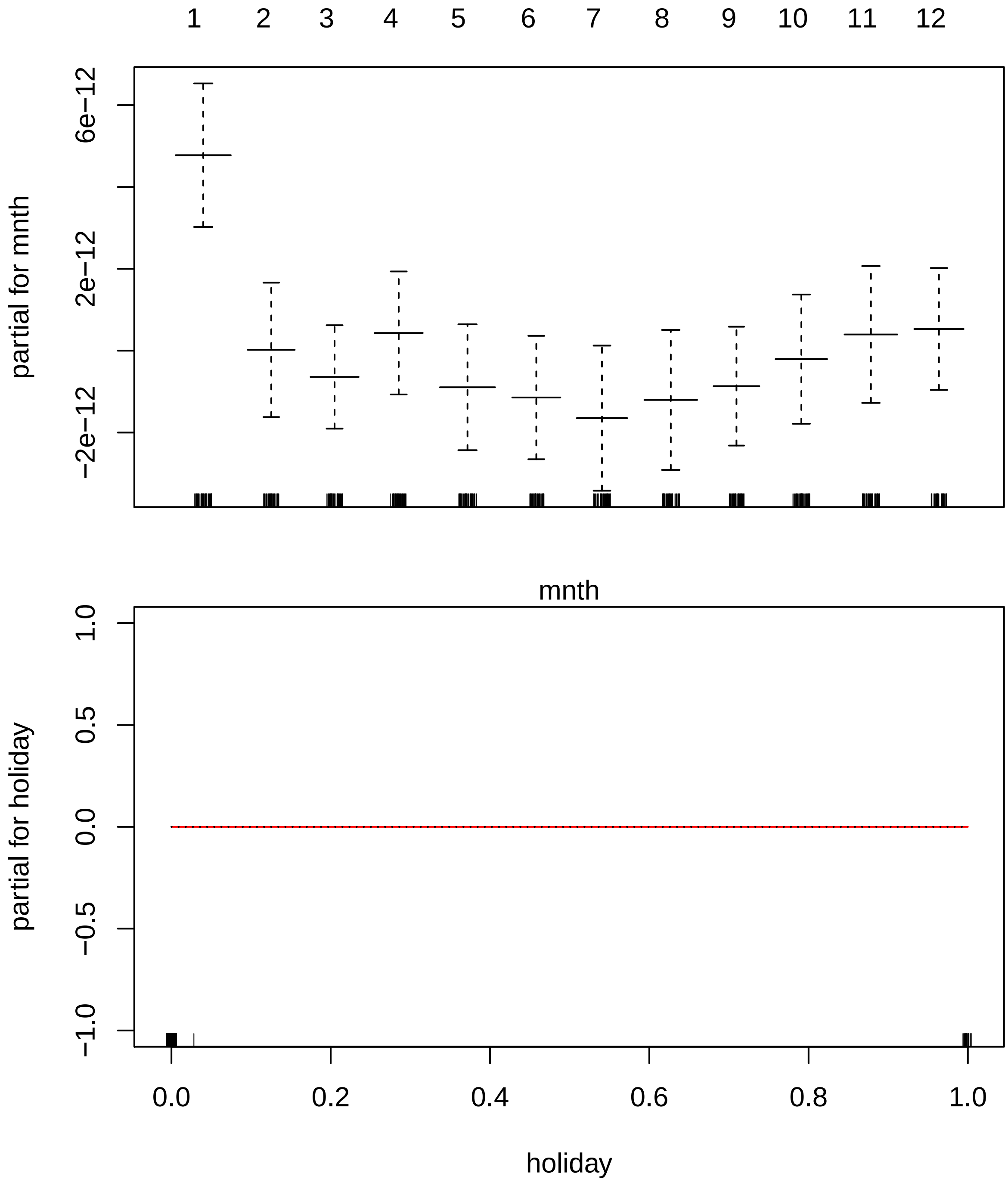

s(casual, df = 11.27571)
2500
1500
500
−500
0
500
1000
1500
2000
2500
3000
3500
casual
s(registered, df = 12.95976)
3000
1000
−1000
−3000
1000
2000
3000
4000
5000
6000
7000
registered

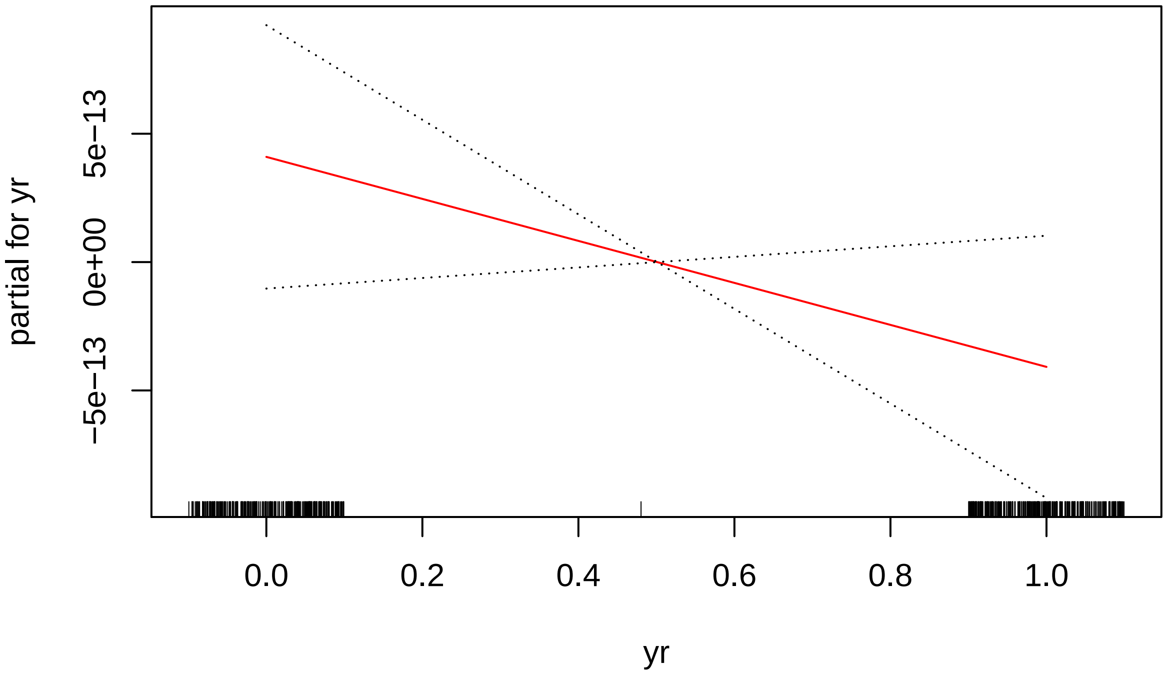

1
2
3
4
partial for season
2e−12
1e−12
0e+00
−2e−12
season


0
1
2
3
4
5
6
partial for weekday
4e−12
2e−12
0e+00
−2e−12
weekday


partial for workingday
1.5e−12
5.0e−13
−5.0e−13
0.0
0.2
0.4
0.6
0.8
1.0
workingday


1
2
3
partial for weathersit
−1e−12
−3e−12
weathersit


1
2
3
4
5
6
7
8
9
10
11
12
6e−12
2e−12
−2e−12
partial for mnth
mnth
1.0
0.5
0.0
−0.5
−1.0
partial for holiday
0.0
0.2
0.4
0.6
0.8
1.0
holiday



```r
summary(gam_train)
```

```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

```
##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + s(casual,
##     df = 11.27571) + s(registered, df = 12.95976) + season +
##     weekday + workingday + weathersit + mnth + holiday + yr,
##     data = day_train)
## Deviance Residuals:
##        Min         1Q     Median         3Q        Max
## -4.025e-11 -1.023e-12  0.000e+00  1.364e-12  8.015e-12
##
## (Dispersion Parameter for gaussian family taken to be 0)
##
##     Null Deviance: 1902143029 on 514 degrees of freedom
## Residual Deviance: 0 on 437.3005 degrees of freedom
## AIC: -25765.1
##
## Number of Local Scoring Iterations: 1
##
## Anova for Parametric Effects
##                                 Df    Sum Sq   Mean Sq    F value
## s(temp, df = 9.103704)         1.0 784713398 784713398 8.2480e+31
## s(windspeed, df = 6.007664)    1.0  30406192  30406192 3.1959e+30
## s(atemp, df = 8.805497)        1.0    226042    226042 2.3759e+28
## s(hum, df = 4.548876)          1.0 118758619 118758619 1.2482e+31
## s(casual, df = 11.27571)       1.0 175037047 175037047 1.8398e+31
## s(registered, df = 12.95976)   1.0 793001731 793001731 8.3351e+31
## season                         3.0         0         0 4.0335e+00
## weekday                        6.0         0         0 7.2506e+00
## workingday                     1.0         0         0 1.4731e+00
## weathersit                     2.0         0         0 2.9706e+00
## mnth                          11.0         0         0 6.0129e+00
## yr                             1.0         0         0 2.5529e+00
## Residuals                    437.3         0         0
##                                  Pr(>F)
## s(temp, df = 9.103704)       < 2.2e-16 ***
## s(windspeed, df = 6.007664)  < 2.2e-16 ***
## s(atemp, df = 8.805497)      < 2.2e-16 ***
## s(hum, df = 4.548876)        < 2.2e-16 ***
## s(casual, df = 11.27571)     < 2.2e-16 ***
## s(registered, df = 12.95976) < 2.2e-16 ***
## season                        0.007558 **
## weekday                      2.145e-07 ***
## workingday                    0.225508
## weathersit                    0.052309 .
## mnth                         3.580e-09 ***
## yr                            0.110812
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                              Npar Df Npar F     Pr(F)
## (Intercept)
## s(temp, df = 9.103704)           8.1 3.8993 0.0001706 ***
## s(windspeed, df = 6.007664)      5.0 1.7703 0.1174986
## s(atemp, df = 8.805497)          7.8 7.2264 7.357e-09 ***
## s(hum, df = 4.548876)            3.5 4.1444 0.0039459 **
## s(casual, df = 11.27571)        10.3 0.9732 0.4671411
## s(registered, df = 12.95976)    12.0 6.2883 2.864e-10 ***
## season
## weekday
## workingday
## weathersit
## mnth
## holiday
## yr
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#PREDICCION

```
#Vamos a predecir para saber el error. Vemos que es practicamente 0 por lo que
#voy a realizar otro modelo sin las variables casual y register.
predict_modelo_gam <- predict(gam1,day_test)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading
```

```
test_error_gam <- mean((predict_modelo_gam - day_test$cnt)^2)
test_error_gam
```

```
## [1] 2.701973e-23
```

– MODELO 2 –

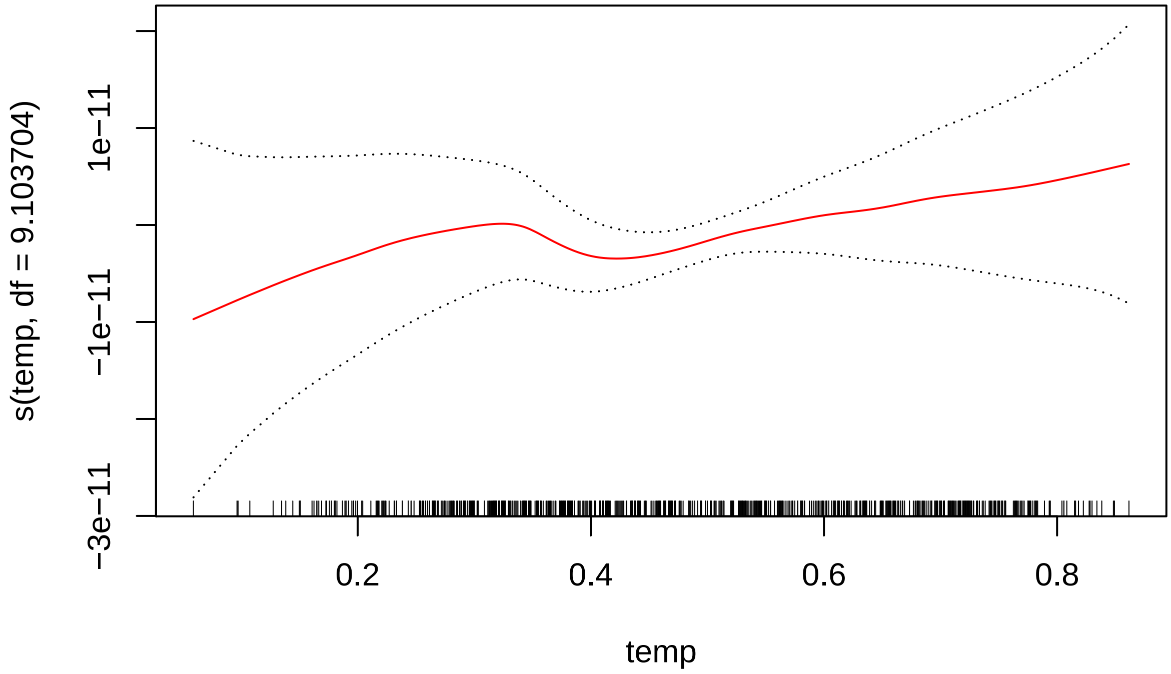
GAM

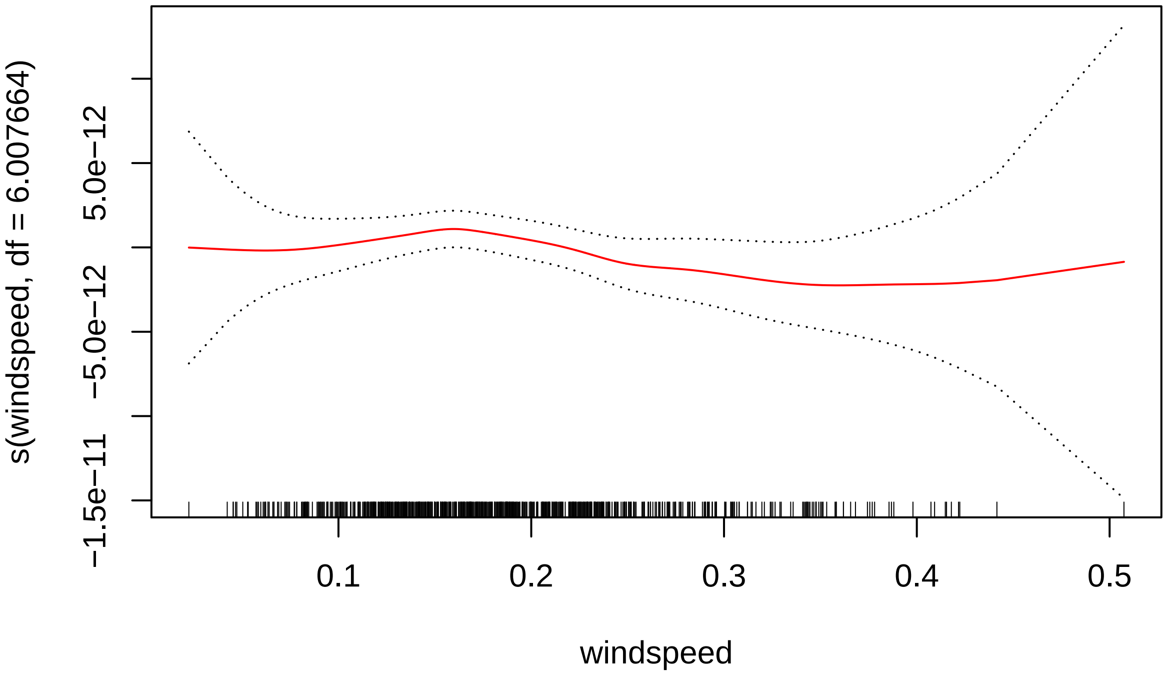
```
#Realizamos los posibles modelos primero sin las variables casual y register
gam1.2 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, df=4
            data=day)
```

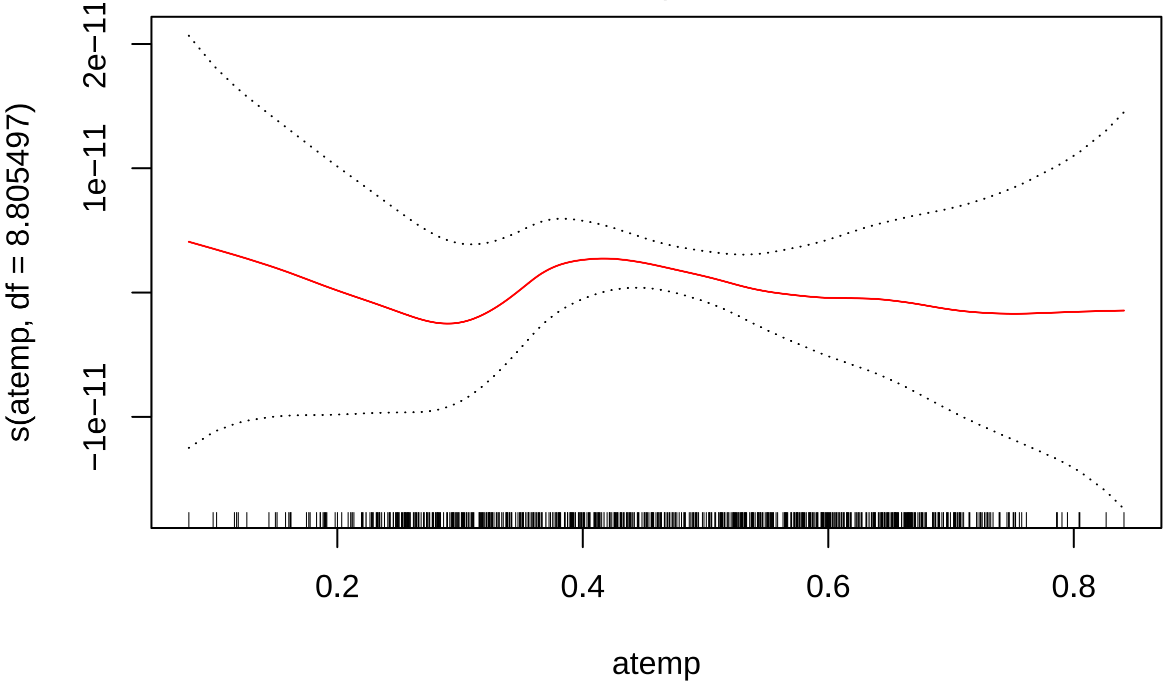
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

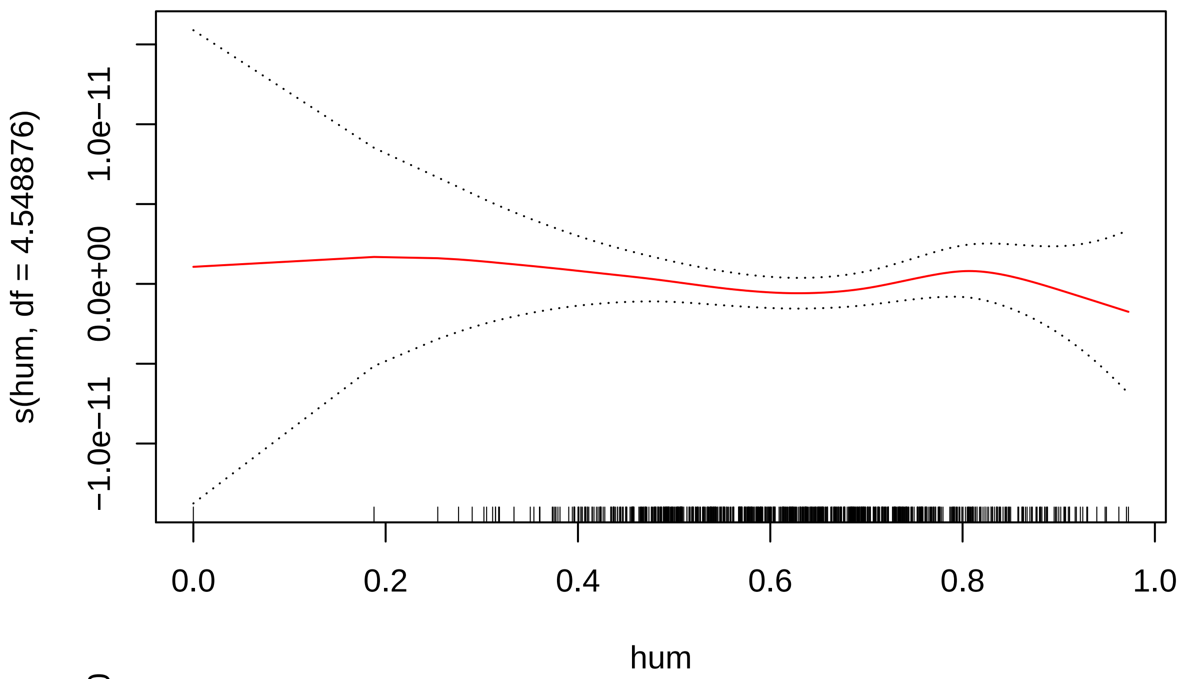
```
plot(gam1, se=TRUE, col='red')
```

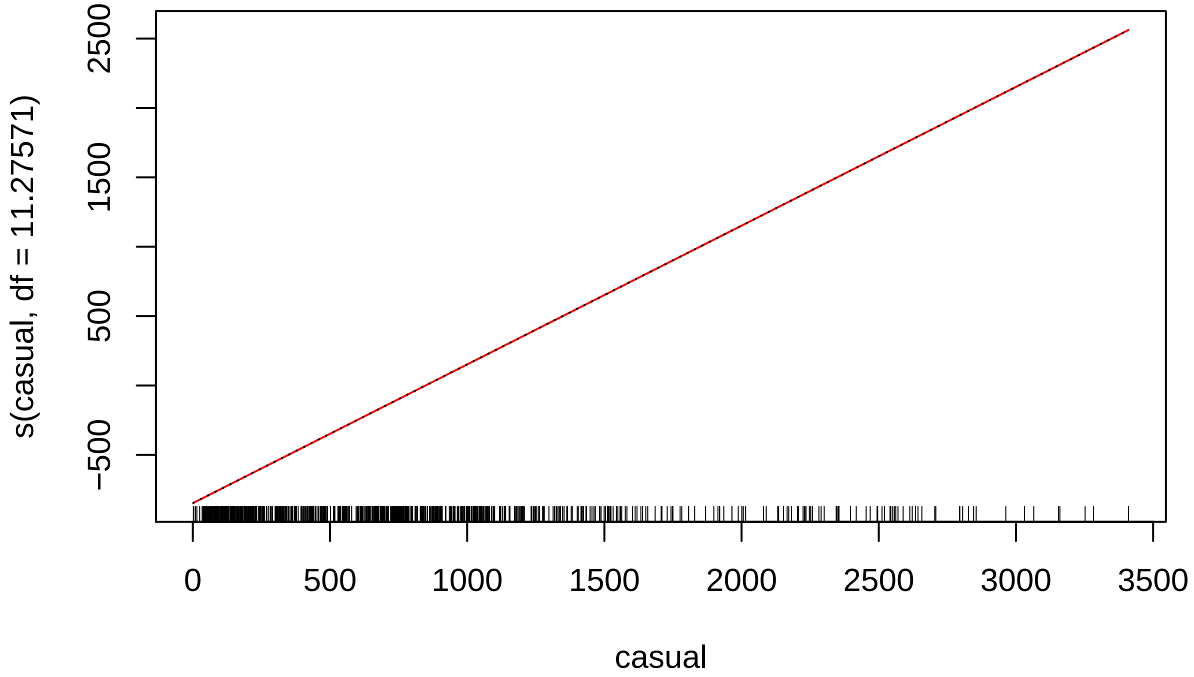
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

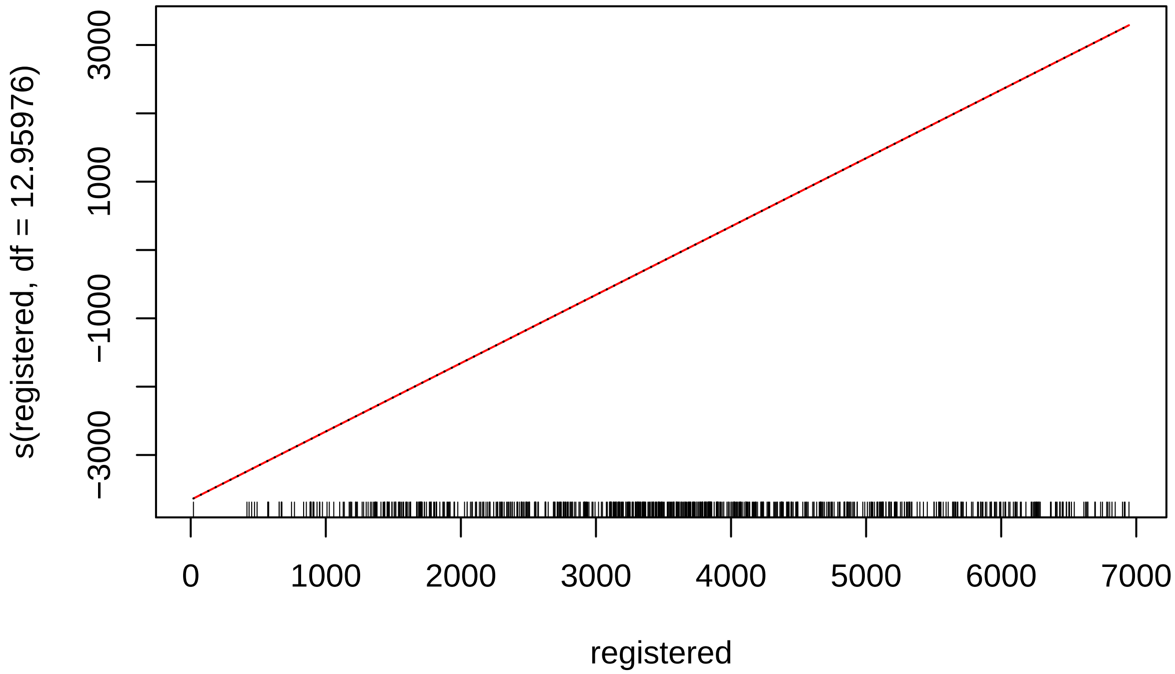


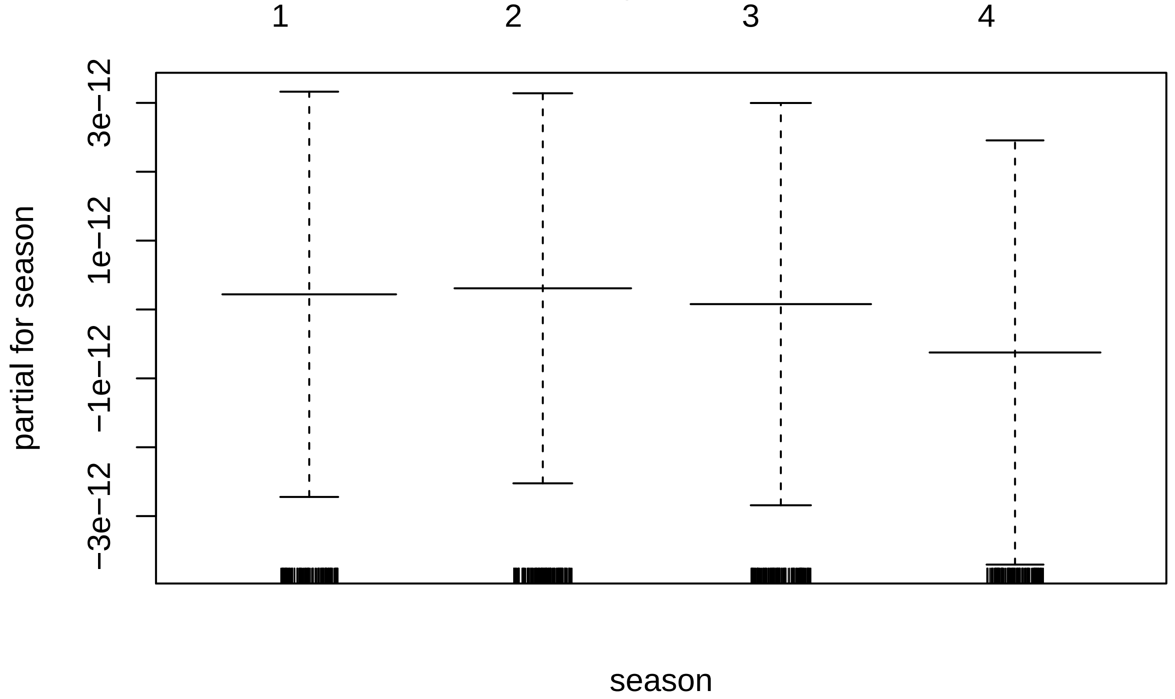

s(windspeed, df = 6.007664)
5.0e−12
−5.0e−12
−1.5e−11
0.1
0.2
0.3
0.4
0.5
windspeed

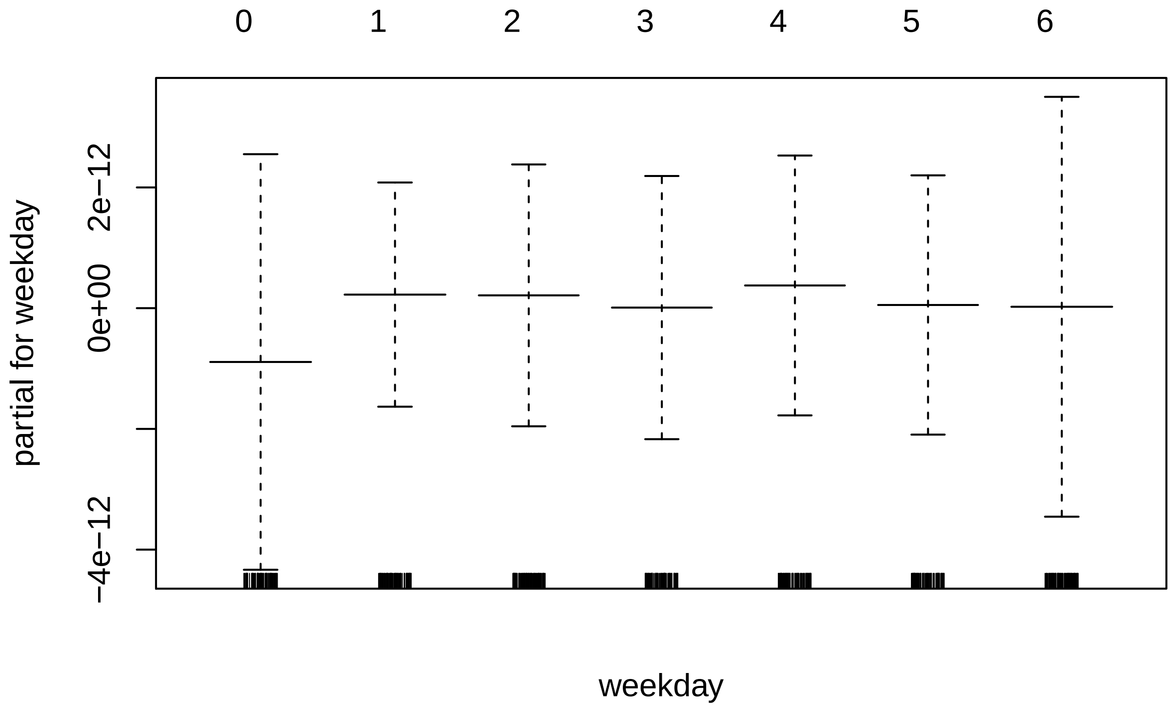

s(atemp, df = 8.805497)
2e−11
1e−11
−1e−11
0.2
0.4
0.6
0.8
atemp

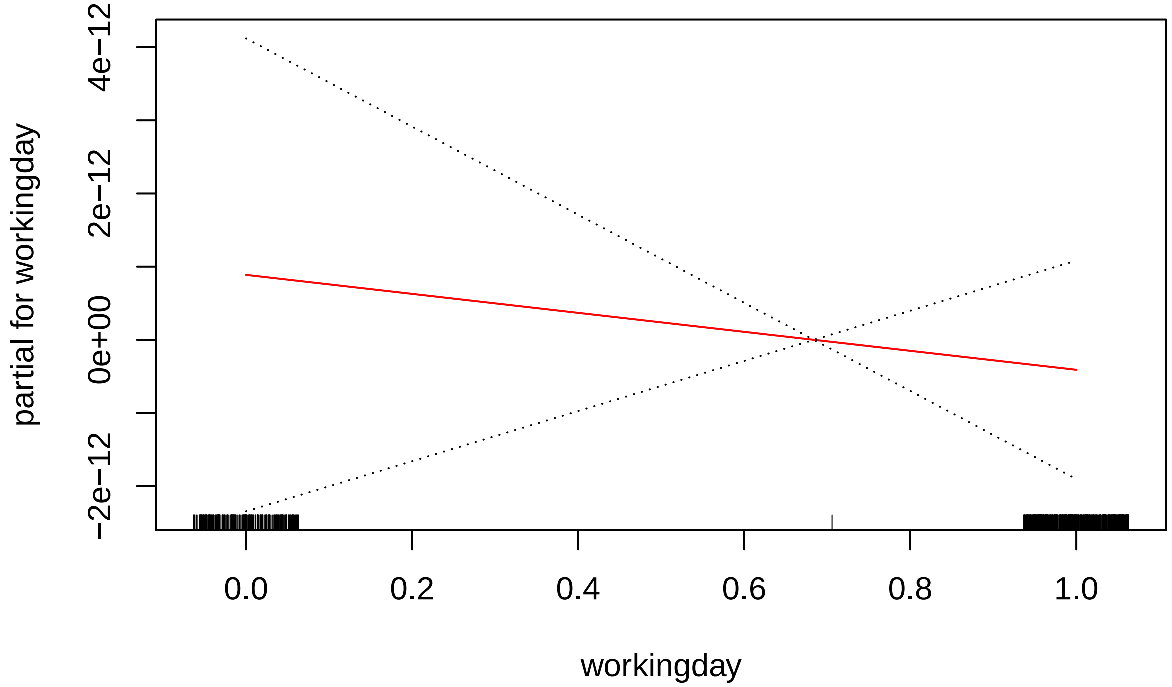

s(hum, df = 4.548876)
−1.0e−11
0.0e+00
1.0e−11
0.0
0.2
0.4
0.6
0.8
1.0
hum

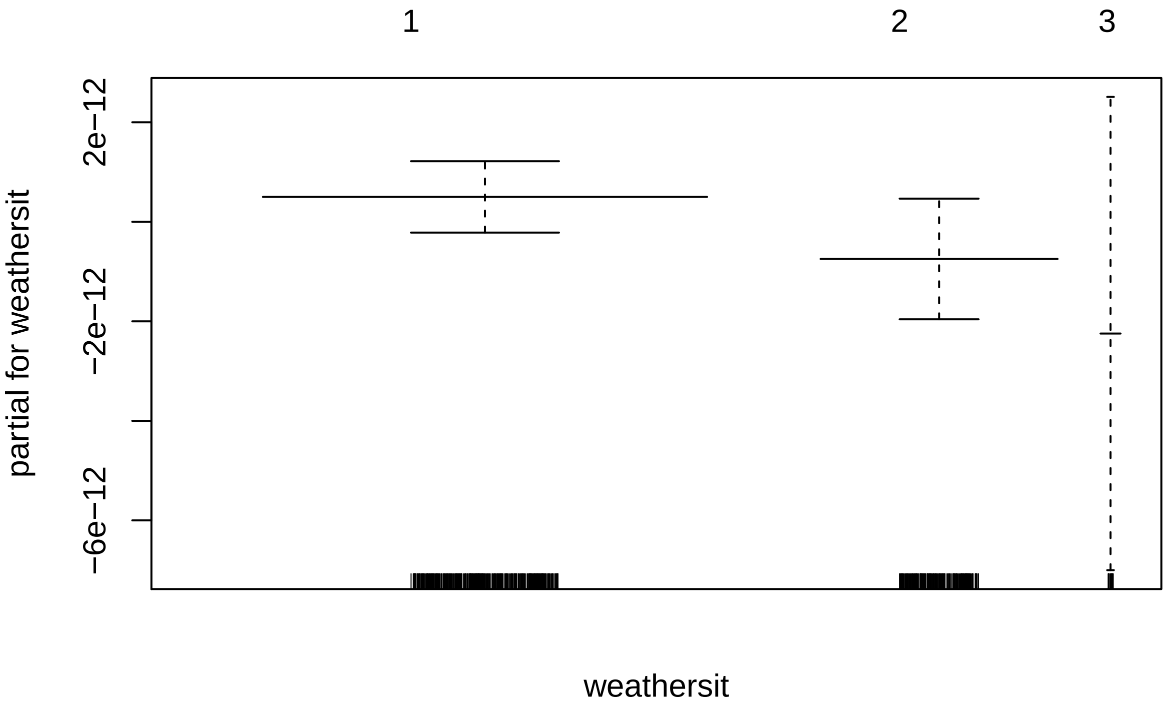

s(casual, df = 11.27571)
−500
500
1500
2500
0
500
1000
1500
2000
2500
3000
3500
casual

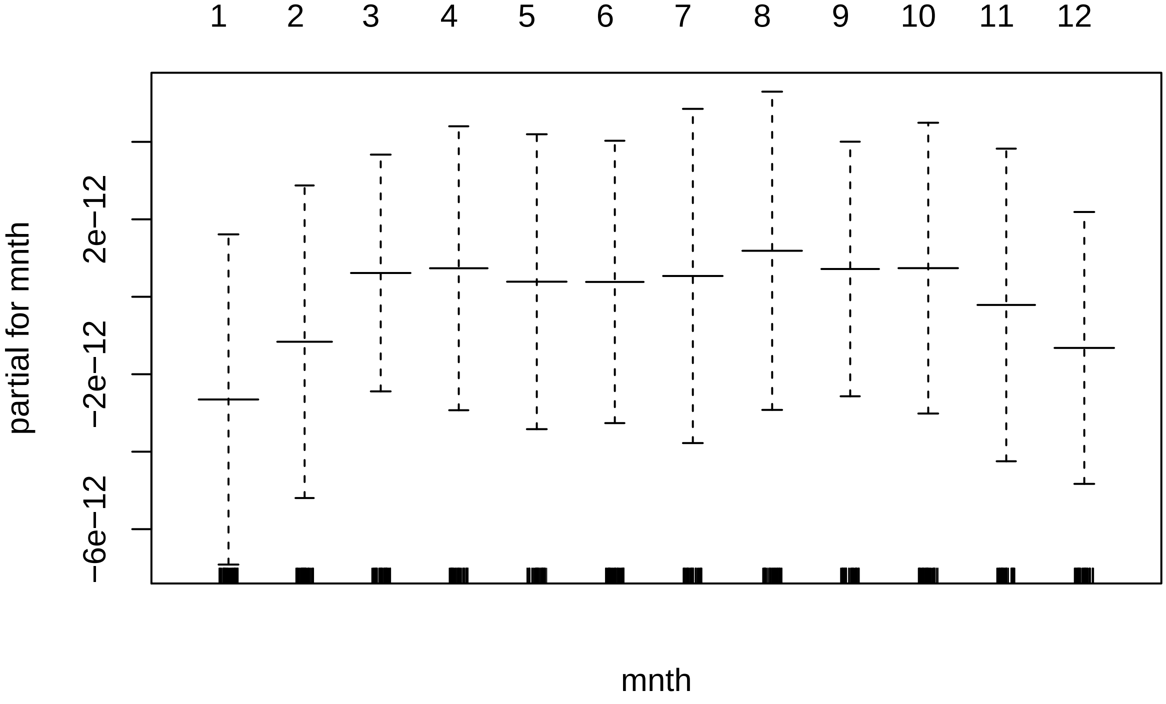

s(registered, df = 12.95976)
3000
1000
−1000
−3000
0
1000
2000
3000
4000
5000
6000
7000
registered

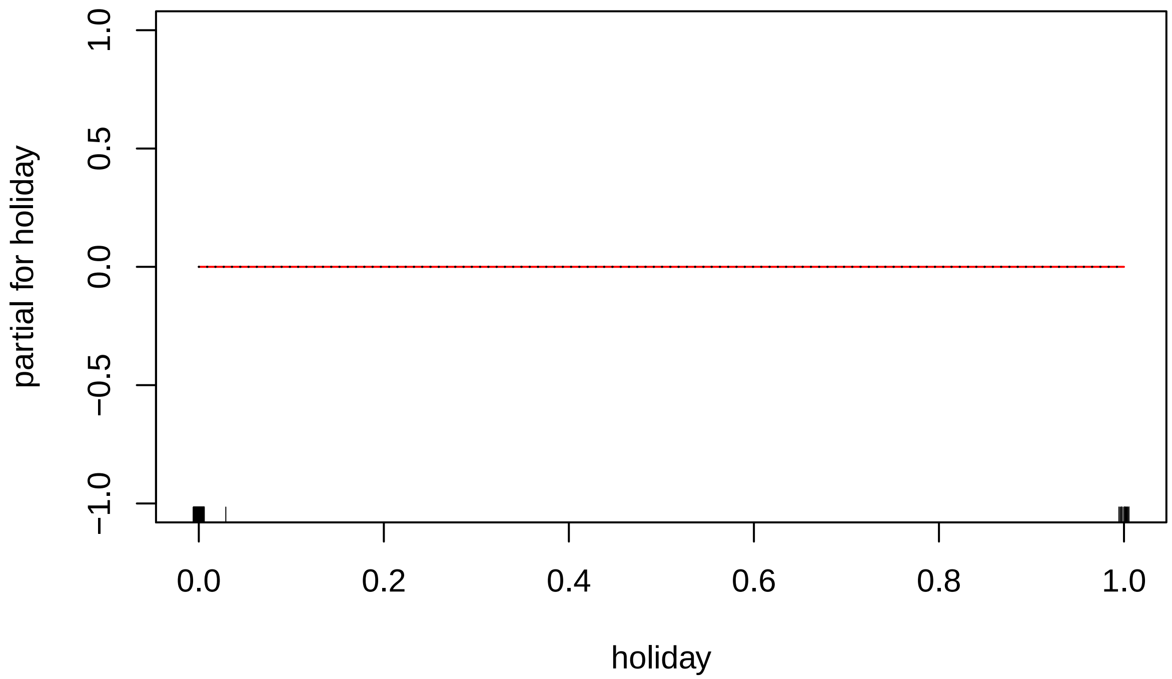

1
2
3
4
partial for season
3e−12
1e−12
−1e−12
−3e−12
season

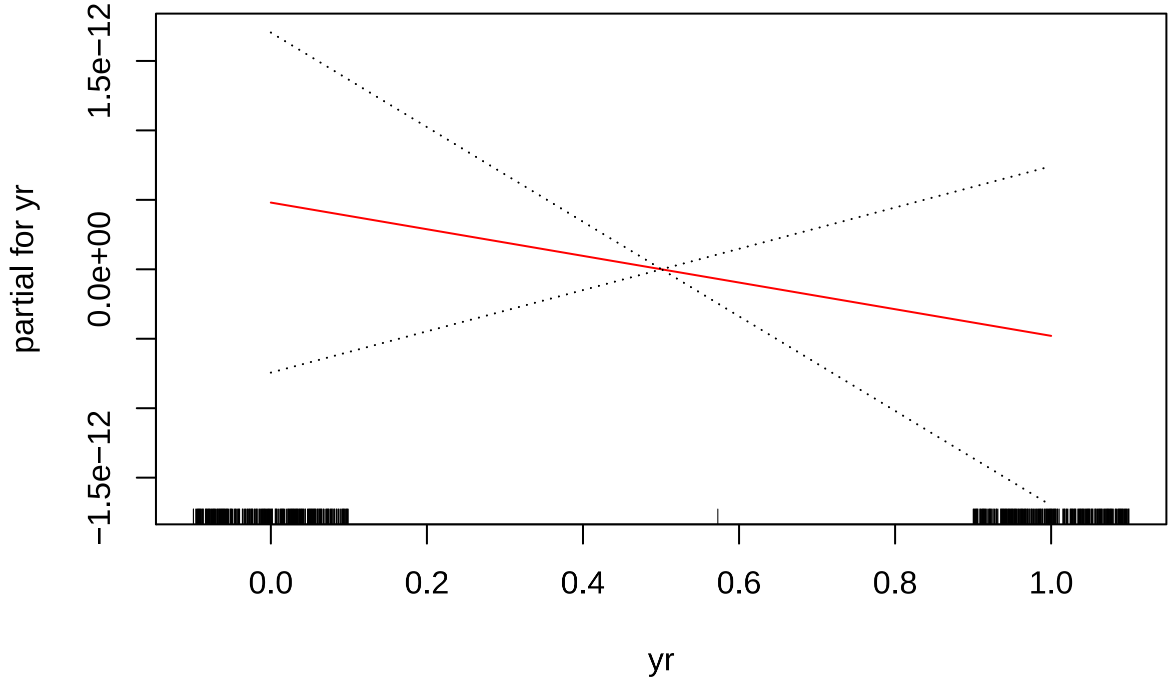

0
1
2
3
4
5
6
partial for weekday
2e−12
0e+00
−4e−12
weekday


partial for workingday
4e−12
2e−12
0e+00
−2e−12
0.0
0.2
0.4
0.6
0.8
1.0
workingday


1
2
3
partial for weathersit
2e−12
−2e−12
−6e−12
weathersit


1
2
3
4
5
6
7
8
9
10
11
12
partial for mnth
2e−12
−2e−12
−6e−12
mnth





```r
summary(gam1.2)
```

```
##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + weekday +
##     workingday + weathersit + mnth + holiday + yr, data = day)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3114.68  -330.88    42.92   423.07  2154.85
##
## (Dispersion Parameter for gaussian family taken to be 479515)
```

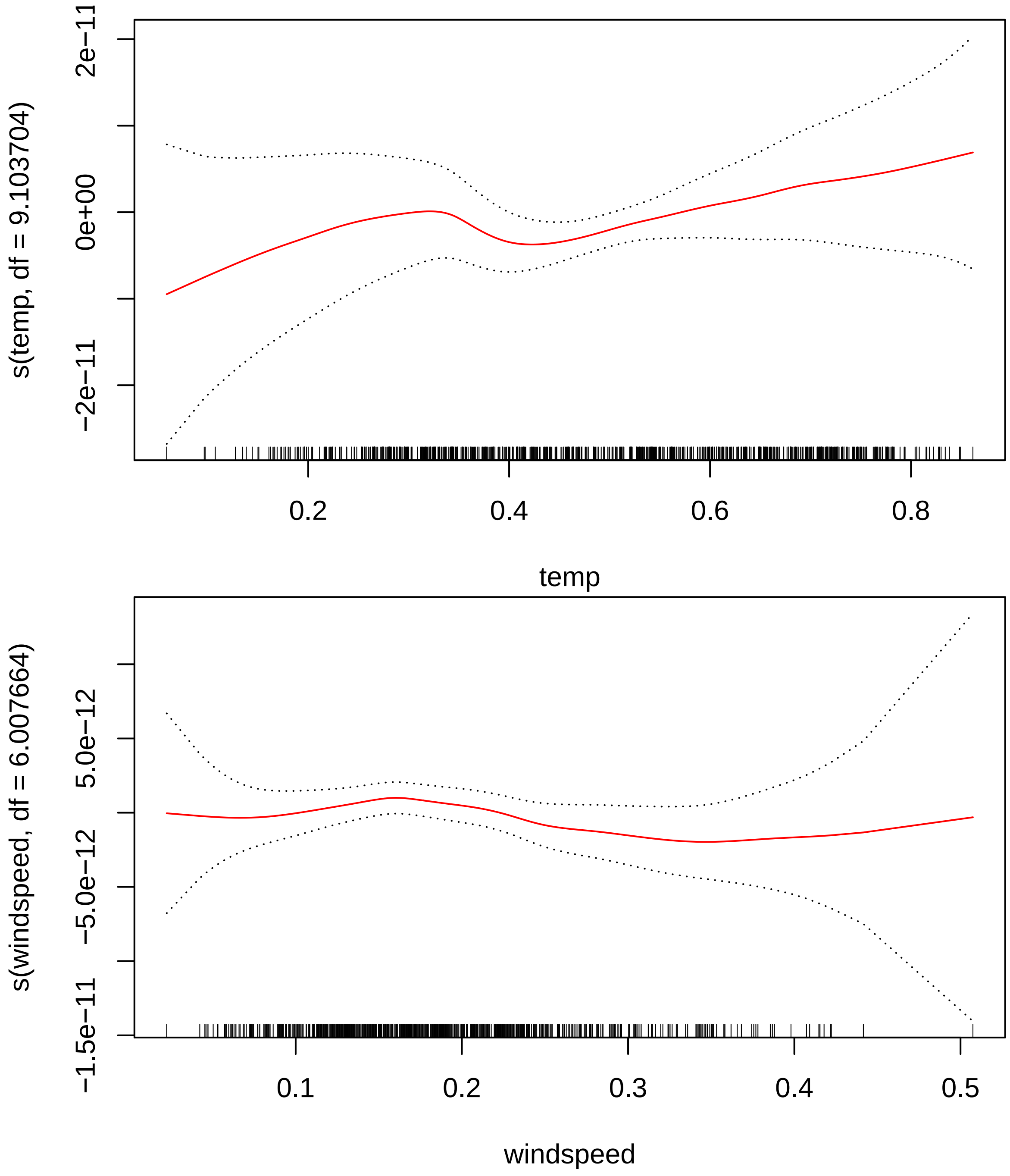
```
##
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 326326105 on 680.5337 degrees of freedom
## AIC: 11687
##
## Number of Local Scoring Iterations: 16
##
## Anova for Parametric Effects
##                                 Df     Sum Sq    Mean Sq   F value
## s(temp, df = 9.103704)        1.00 1028707877 1028707877 2145.3093
## s(windspeed, df = 6.007664)   1.00   59263290   59263290  123.5901
## s(atemp, df = 8.805497)       1.00      59461      59461    0.1240
## s(hum, df = 4.548876)         1.00  214861672  214861672  448.0813
## weekday                       6.00   13950686    2325114    4.8489
## workingday                    1.00    4493000    4493000    9.3699
## weathersit                    2.00   36367346   18183673   37.9210
## mnth                         11.00   83098967    7554452   15.7544
## yr                            1.00  683063628  683063628 1424.4887
## Residuals                   680.53  326326105     479515
##                                Pr(>F)
## s(temp, df = 9.103704)      < 2.2e-16 ***
## s(windspeed, df = 6.007664) < 2.2e-16 ***
## s(atemp, df = 8.805497)      0.724843
## s(hum, df = 4.548876)       < 2.2e-16 ***
## weekday                     7.330e-05 ***
## workingday                   0.002293 **
## weathersit                  2.431e-16 ***
## mnth                        < 2.2e-16 ***
## yr                          < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                             Npar Df Npar F     Pr(F)
## (Intercept)
## s(temp, df = 9.103704)          8.1 39.429 < 2.2e-16 ***
## s(windspeed, df = 6.007664)     5.0  5.989 1.939e-05 ***
## s(atemp, df = 8.805497)         7.8  5.716 6.155e-07 ***
## s(hum, df = 4.548876)           3.5  6.646 7.004e-05 ***
## weekday
## workingday
## weathersit
## mnth
## holiday
## yr
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

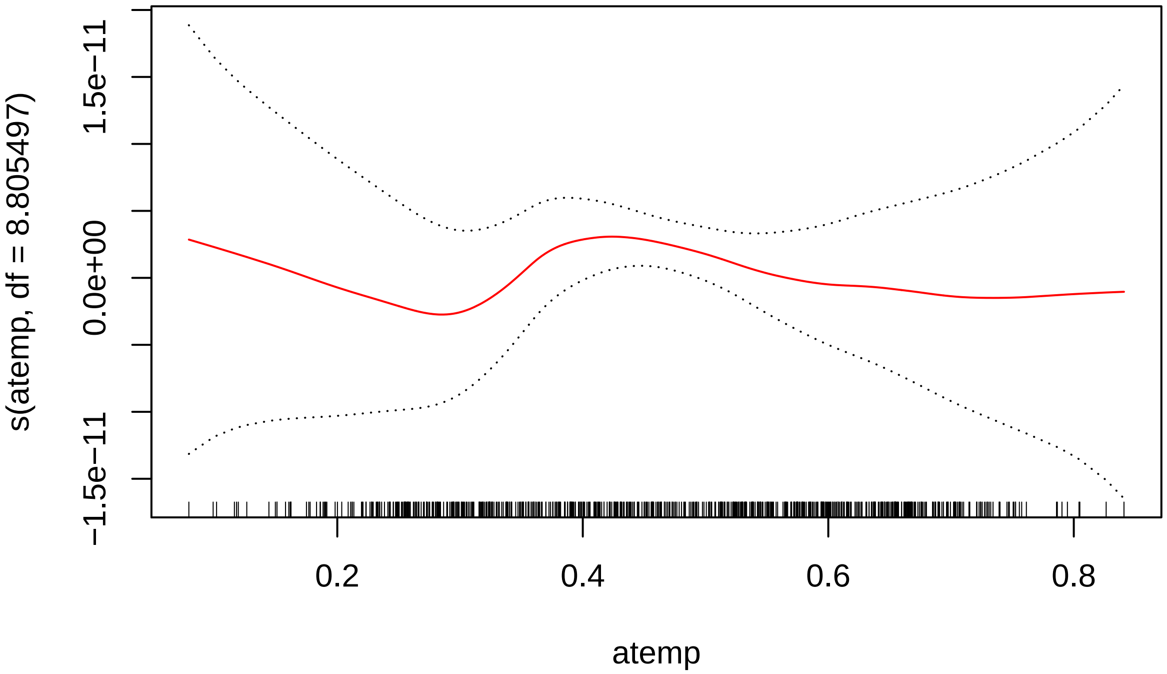
```
#En este gam lo realizamos quitando weathersit.
gam1.2.2 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, df
```

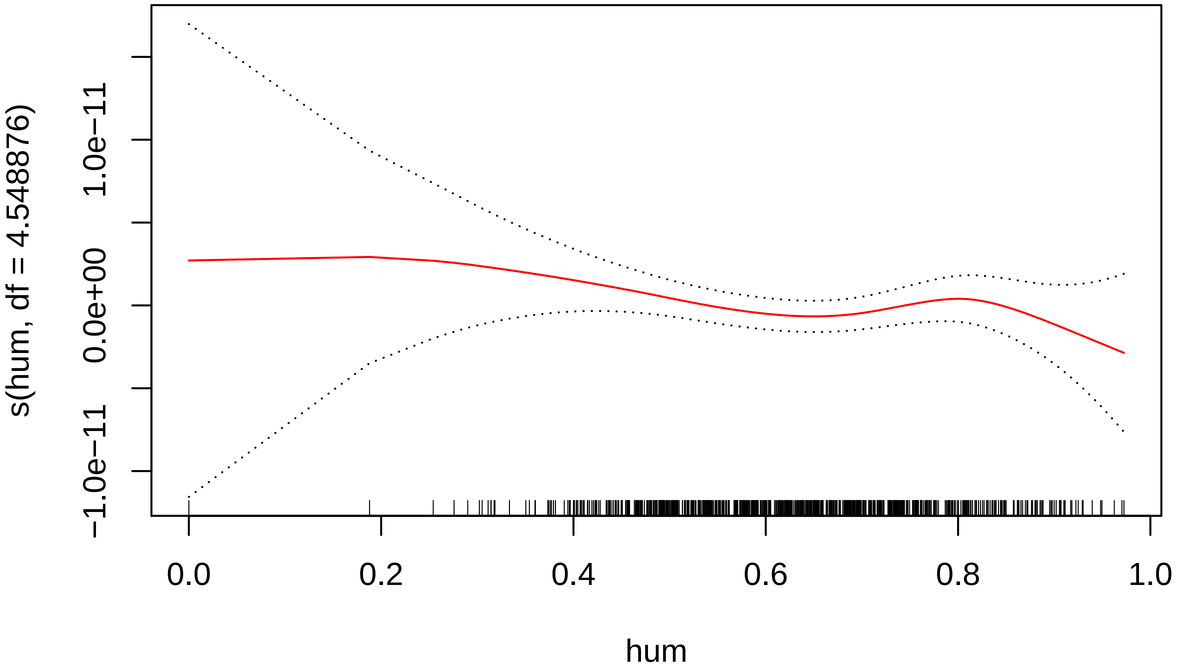
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

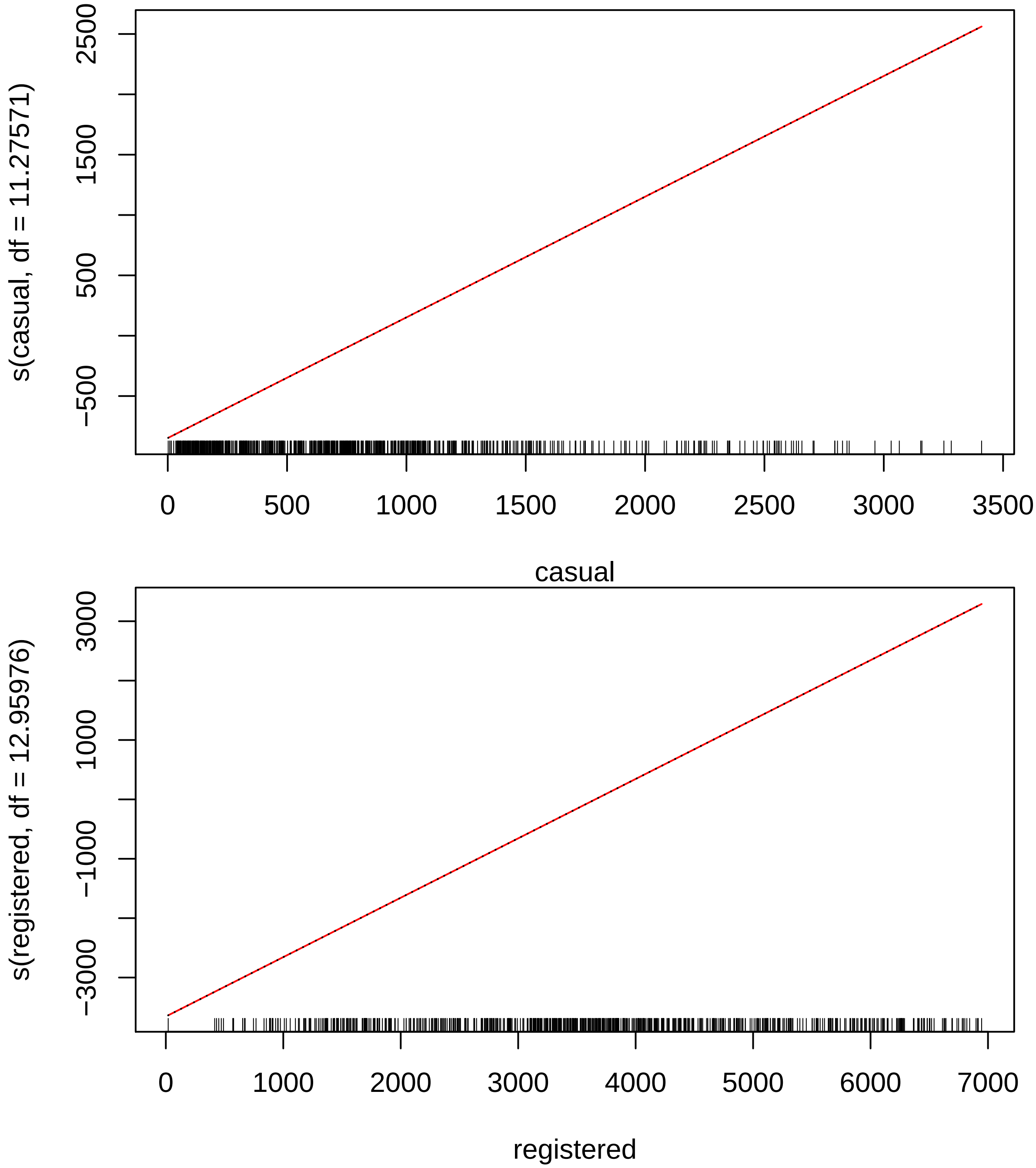
```
plot(gam2, se=TRUE, col='red')
```

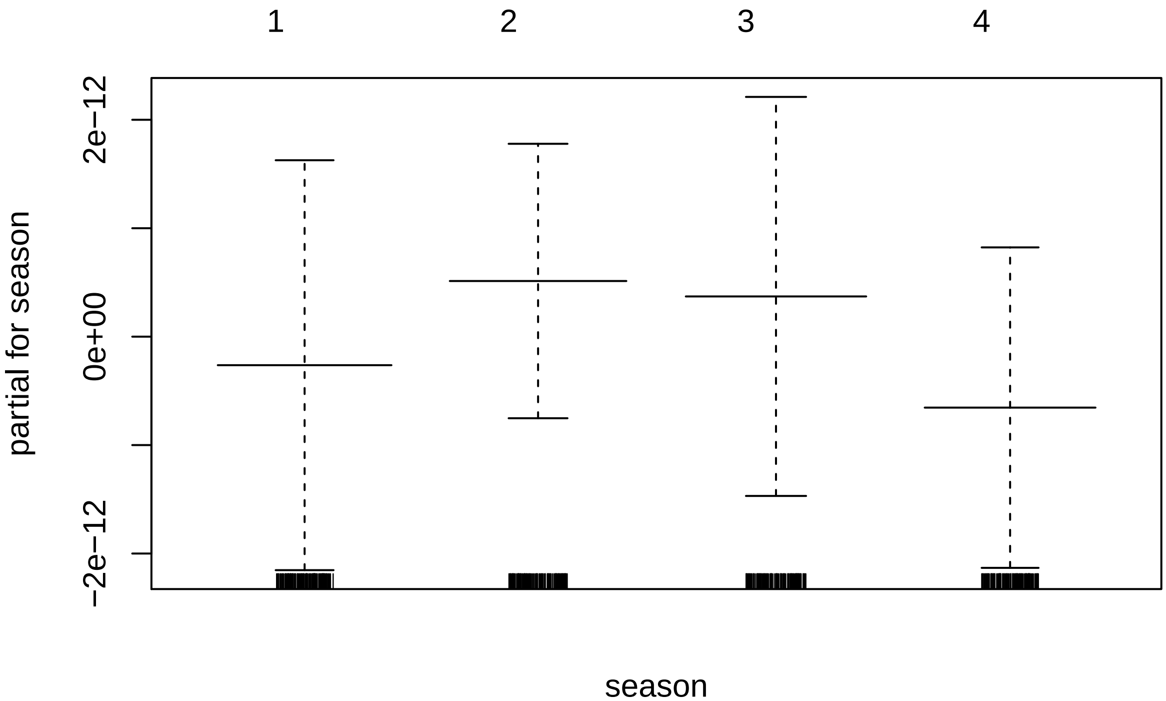
```
## Warning in anova.lm(object.lm, ...): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

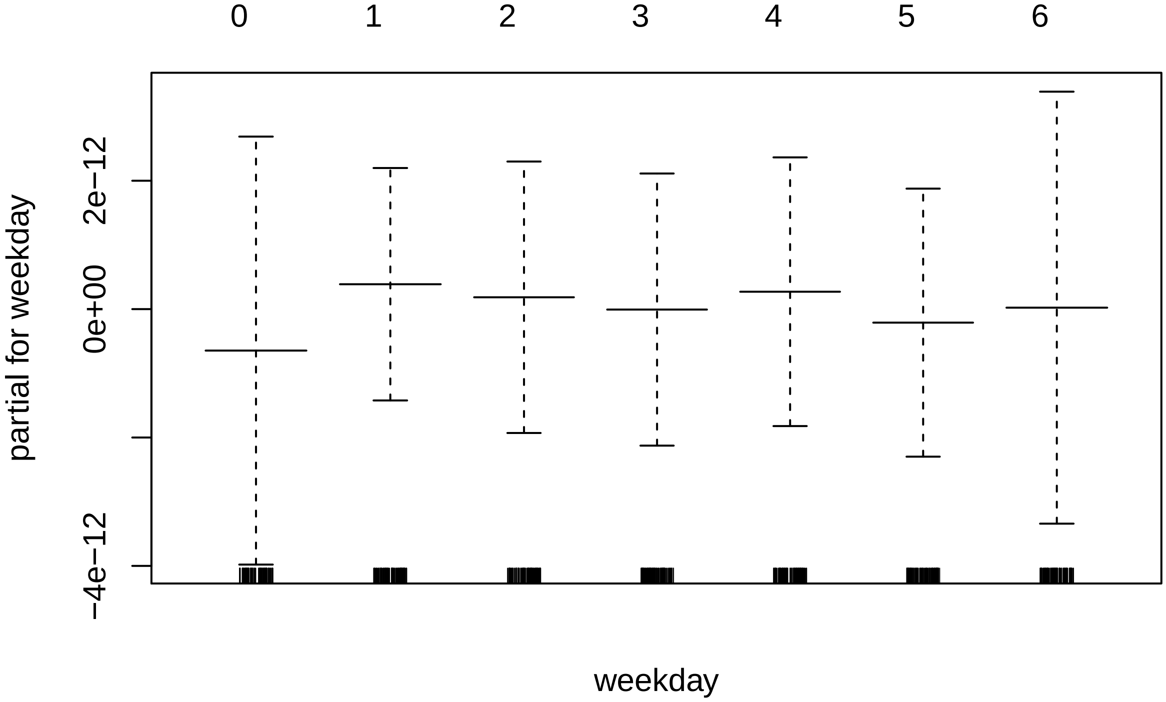


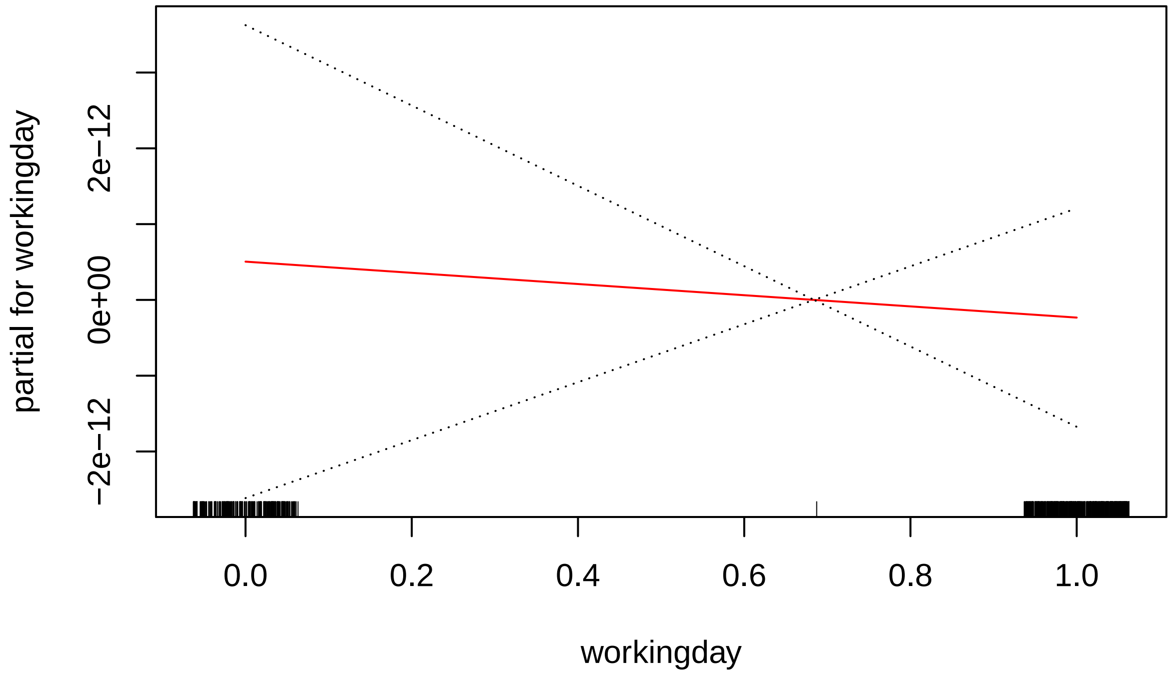

s(atemp, df = 8.805497)
−1.5e−11
0.0e+00
1.5e−11
0.2
0.4
0.6
0.8
atemp

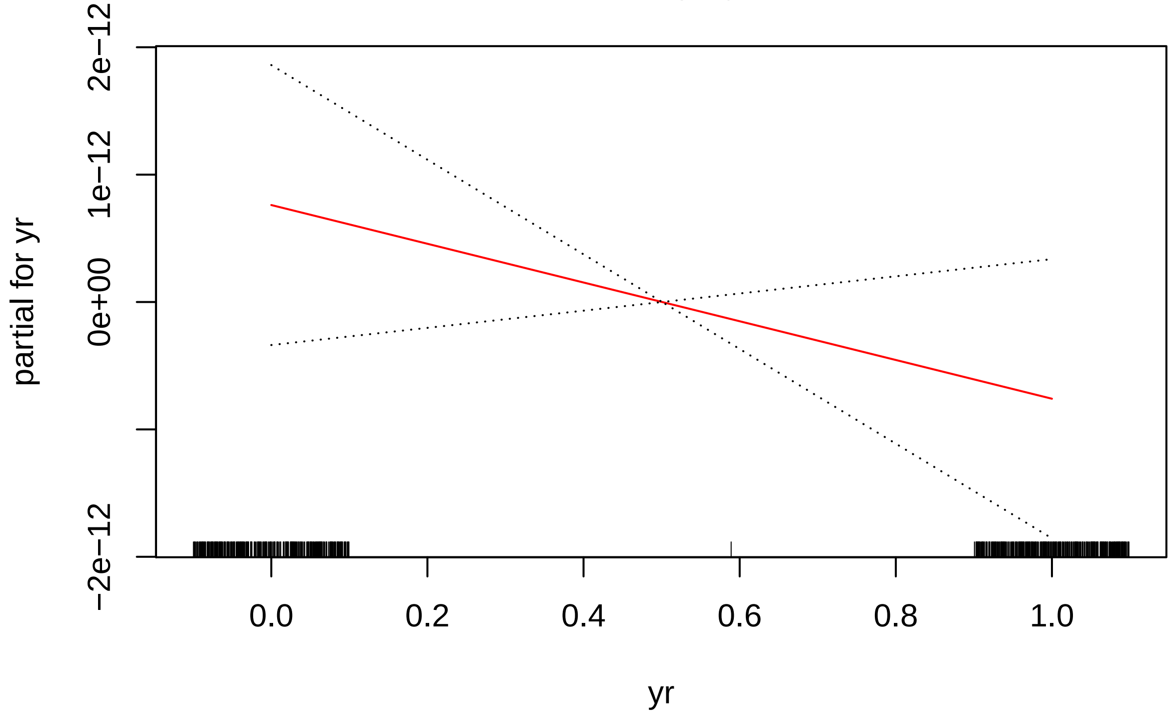

s(hum, df = 4.548876)
−1.0e−11
0.0e+00
1.0e−11
0.0
0.2
0.4
0.6
0.8
1.0
hum


s(casual, df = 11.27571)
−500
500
1500
2500
0
500
1000
1500
2000
2500
3000
3500
casual
s(registered, df = 12.95976)
−3000
−1000
1000
3000
0
1000
2000
3000
4000
5000
6000
7000
registered


1
2
3
4
partial for season
2e−12
0e+00
−2e−12
season


0
1
2
3
4
5
6
partial for weekday
2e−12
0e+00
−4e−12
weekday





```
summary(gam1.2.2)
```

```
##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + weekday +
##     workingday + yr, data = day)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3226.33  -465.80    31.24   511.46  1555.69
##
```

```
## (Dispersion Parameter for gaussian family taken to be 598085)
##
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 414792137 on 693.5337 degrees of freedom
## AIC: 11836.35
##
## Number of Local Scoring Iterations: 16
##
## Anova for Parametric Effects
##                                 Df     Sum Sq    Mean Sq   F value
## s(temp, df = 9.103704)        1.00 1000006569 1000006569 1672.0141
## s(windspeed, df = 6.007664)   1.00   55580836   55580836   92.9313
## s(atemp, df = 8.805497)       1.00     778020     778020    1.3009
## s(hum, df = 4.548876)         1.00  228139940  228139940  381.4507
## weekday                       6.00   13775358    2295893    3.8387
## workingday                    1.00    5242316    5242316    8.7652
## yr                            1.00  656242171  656242171 1097.2390
## Residuals                   693.53  414792137     598085
##                                Pr(>F)
## s(temp, df = 9.103704)      < 2.2e-16 ***
## s(windspeed, df = 6.007664) < 2.2e-16 ***
## s(atemp, df = 8.805497)     0.2544513
## s(hum, df = 4.548876)       < 2.2e-16 ***
## weekday                     0.0008946 ***
## workingday                  0.0031752 **
## yr                          < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                             Npar Df Npar F     Pr(F)
## (Intercept)
## s(temp, df = 9.103704)          8.1 87.577 < 2.2e-16 ***
## s(windspeed, df = 6.007664)     5.0  4.664 0.0003366 ***
## s(atemp, df = 8.805497)         7.8 12.266 3.331e-16 ***
## s(hum, df = 4.548876)           3.5 23.054 2.220e-16 ***
## weekday
## workingday
## yr
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Procedemos a hacer el ANOVA para saber cual de los dos modelos es mejor teniendo en cuenta
#el residuo que tiene uno, el que menor residuo tenga será el que escojamos. En nuestro caso,
#el mejor modelo es el gam.1.2

anova(gam1.2,gam1.2.2, test="F")
```

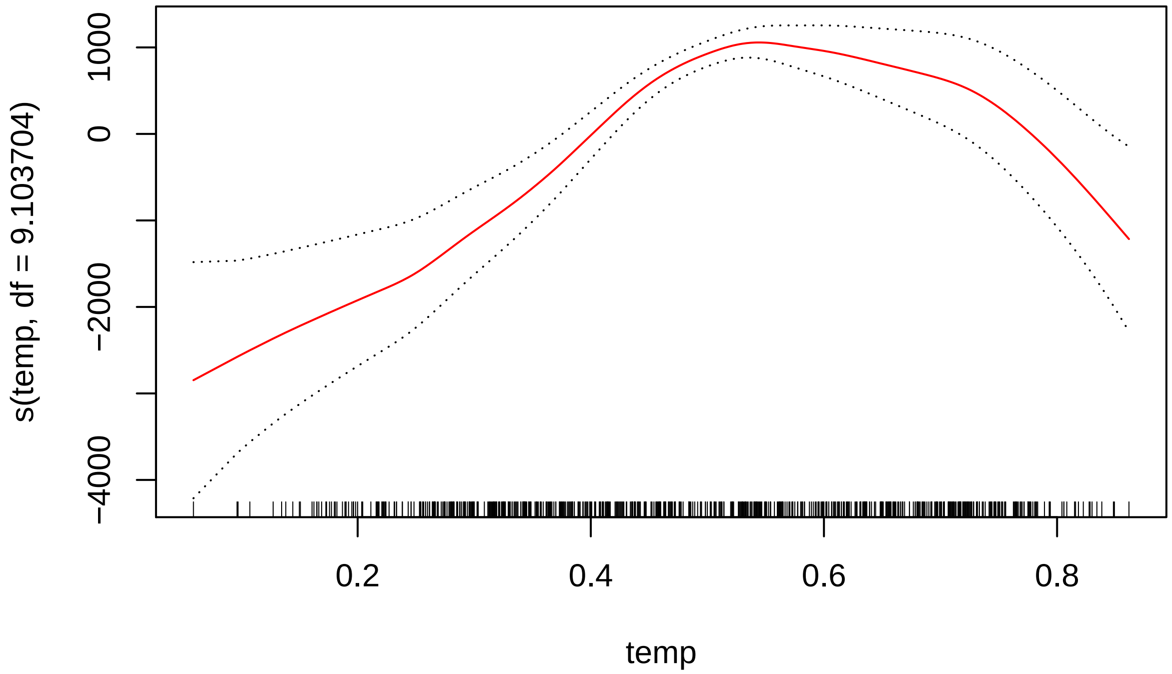
```
## Analysis of Deviance Table
##
## Model 1: cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + weekday +
##     workingday + weathersit + mnth + holiday + yr
## Model 2: cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + weekday +
##     workingday + yr
##   Resid. Df Resid. Dev  Df  Deviance      F    Pr(>F)
## 1    680.53  326326105
## 2    693.53  414792137 -13 -88466032 14.192 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

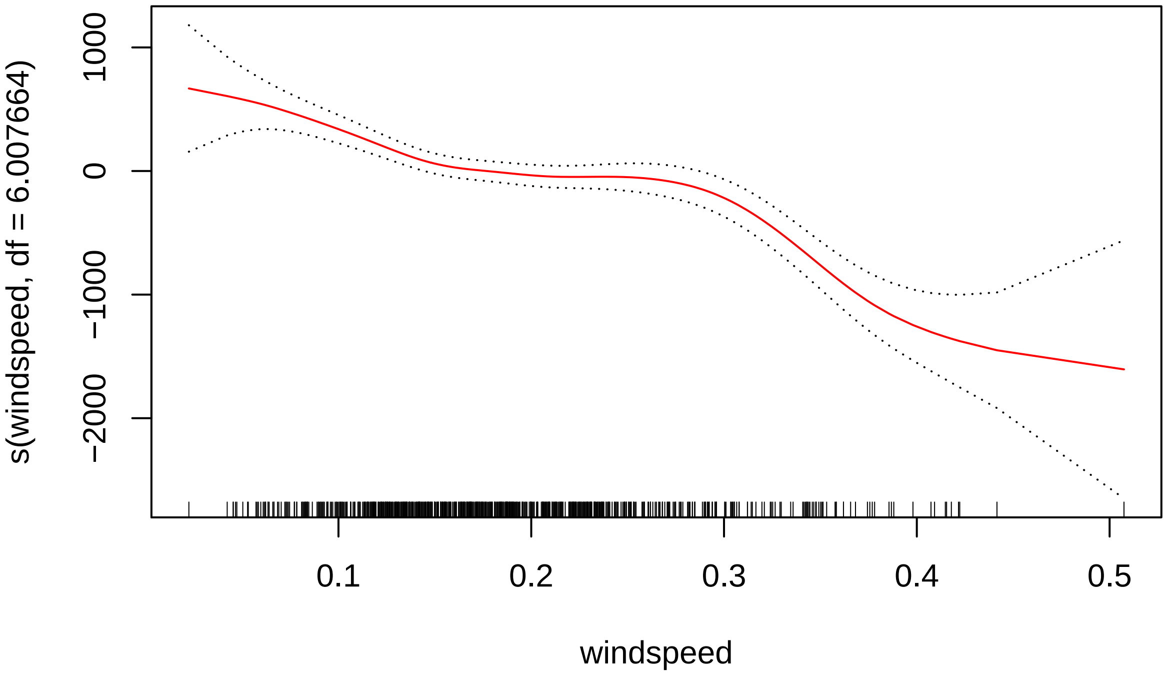
CROSS VALIDATION 2

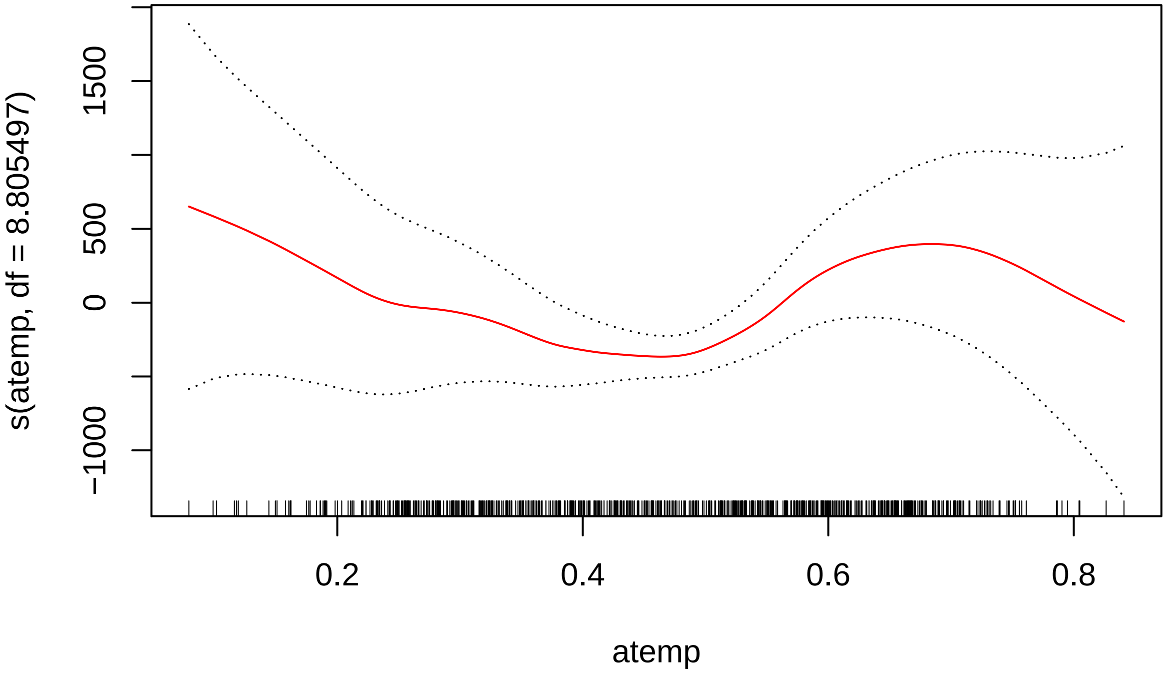
```
#Una vez escogido el modelo, vamos a proceder a dividir nuestra base de datos en
#train y test para predecir.

set.seed(123)
day_split2 <- initial_split(day, prop =.7, strata = "cnt")
day_train2 <- training(day_split2)
day_test2 <- testing(day_split2)

#Tenemos la base de datos dividida en 70/30, y vamos a proceder a introducir nuestro modelo
#en el test para saber como predice.

gam_train2 <- gam(cnt~ s(temp, df=9.103704) + s(windspeed, df=6.007664)+ s(atemp, df=8.805497)+ s(hum, 
            data=day)
```

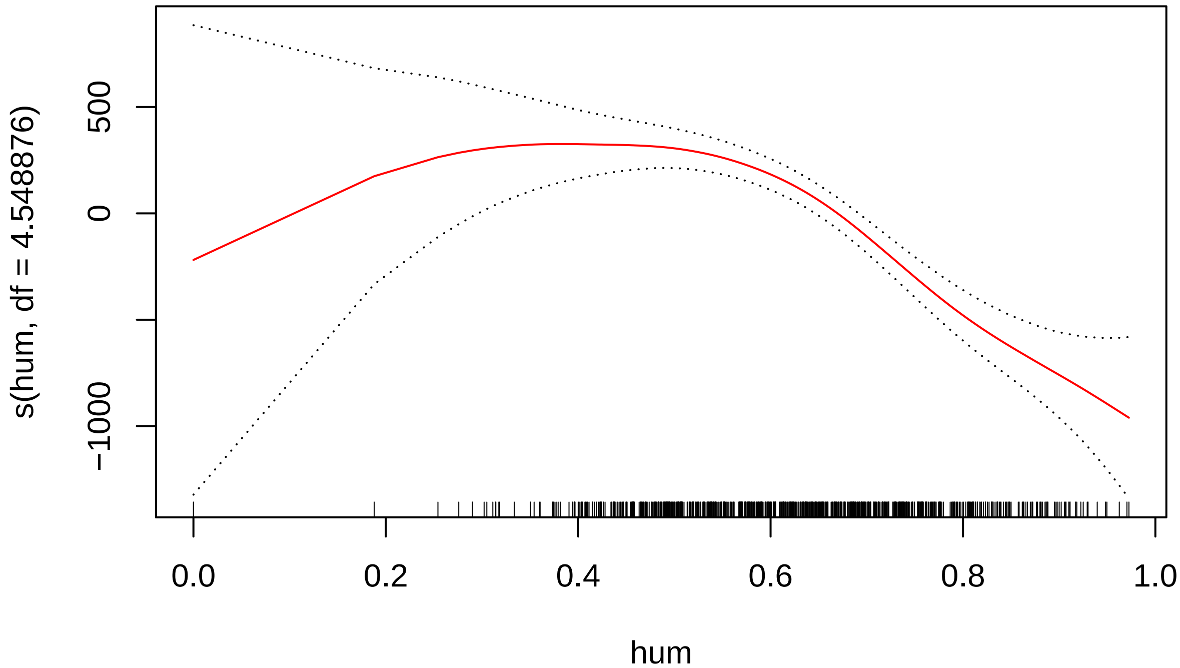
```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts
## argument ignored
```

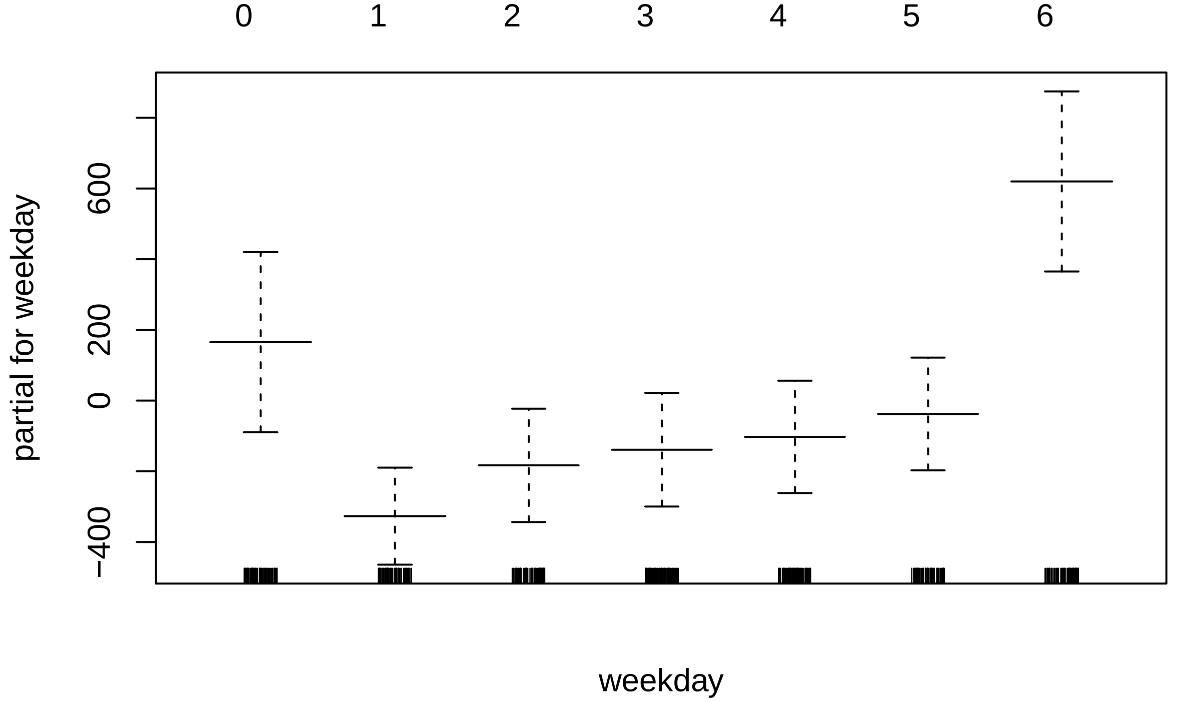
```
plot(gam_train2, se=TRUE, col='red')
```

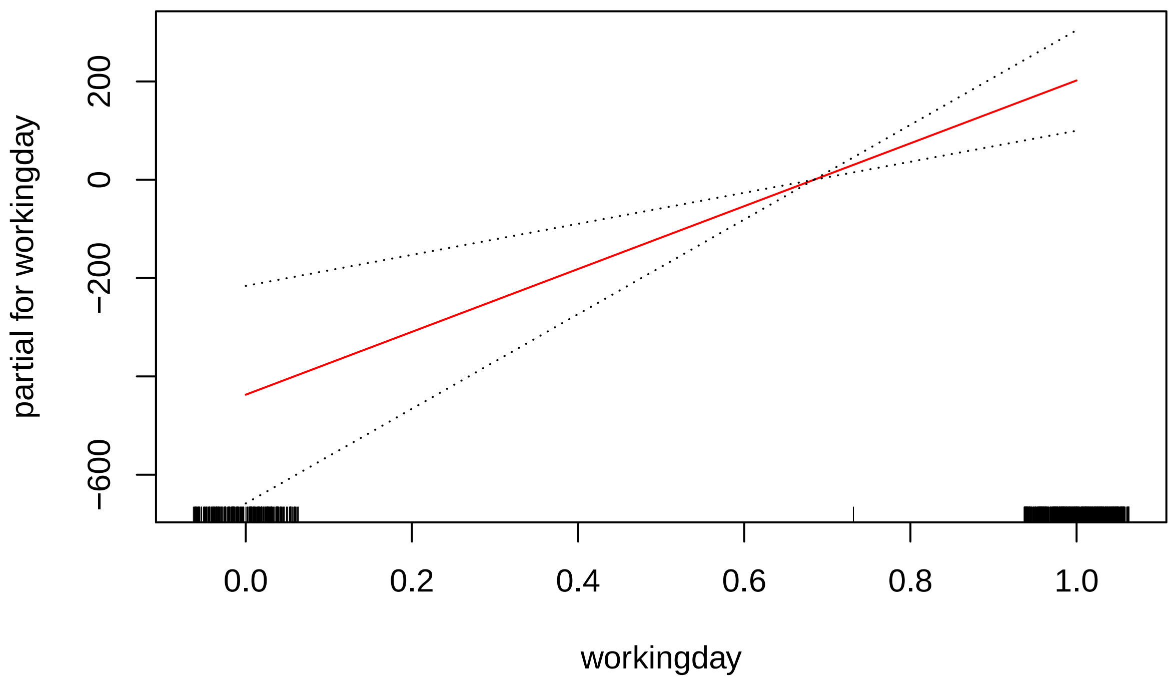


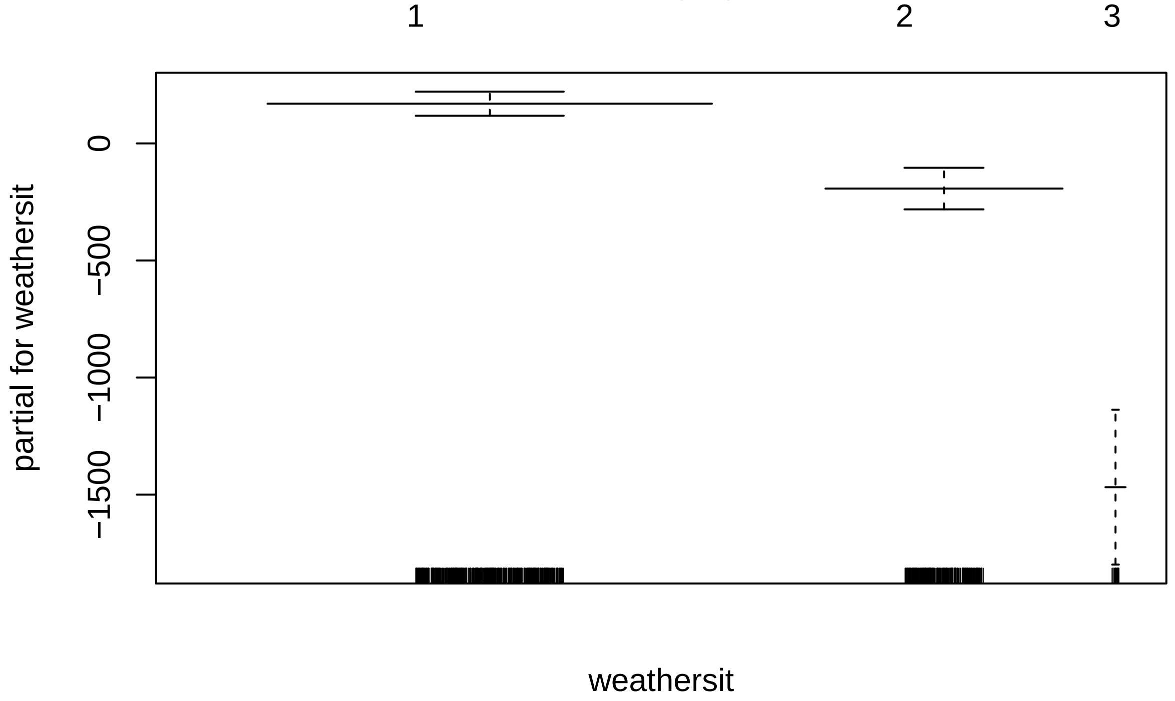

s(windspeed, df = 6.007664)
1000
0
−1000
−2000
0.1
0.2
0.3
0.4
0.5
windspeed

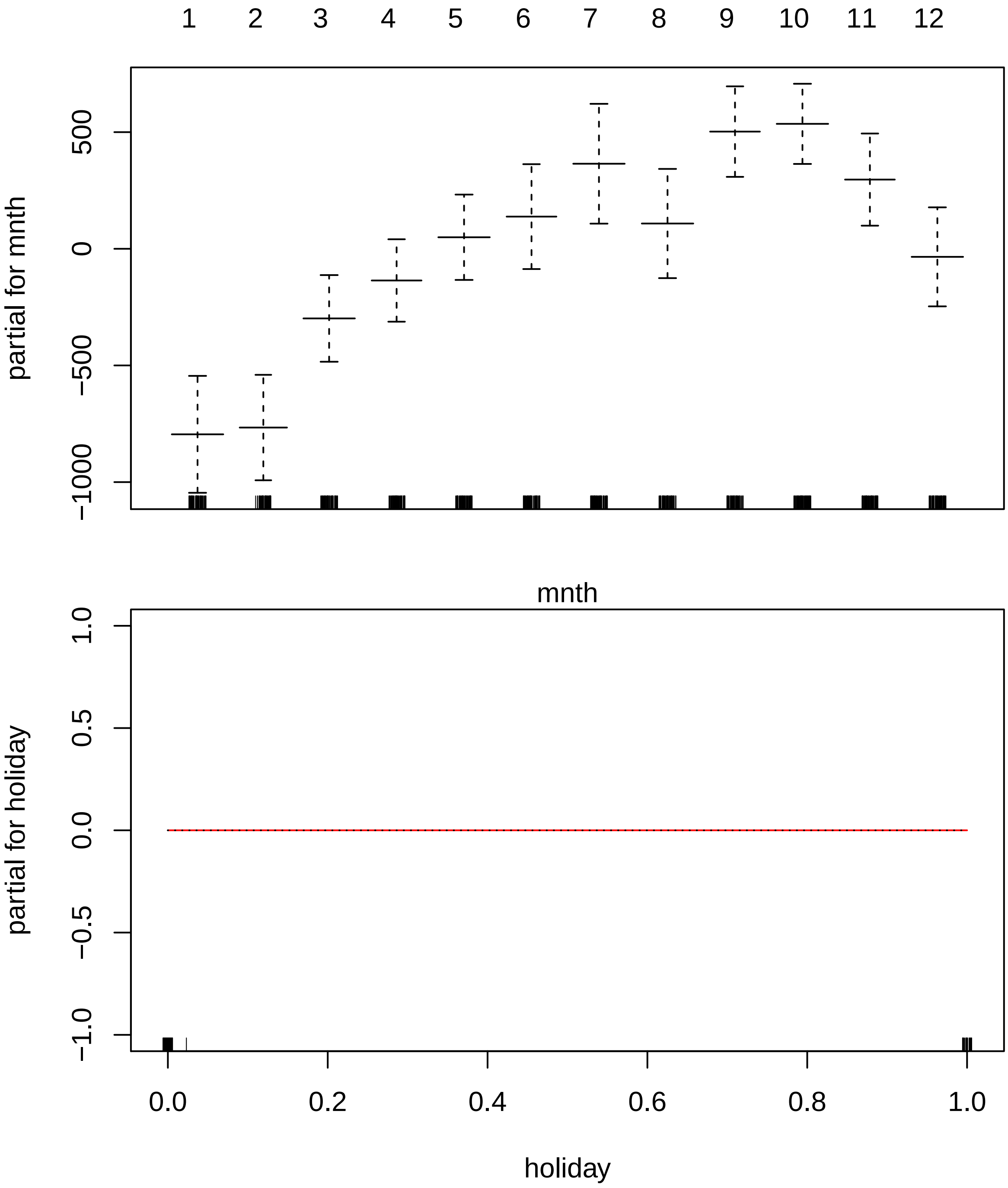

s(atemp, df = 8.805497)
1500
500
0
−1000
0.2
0.4
0.6
0.8
atemp

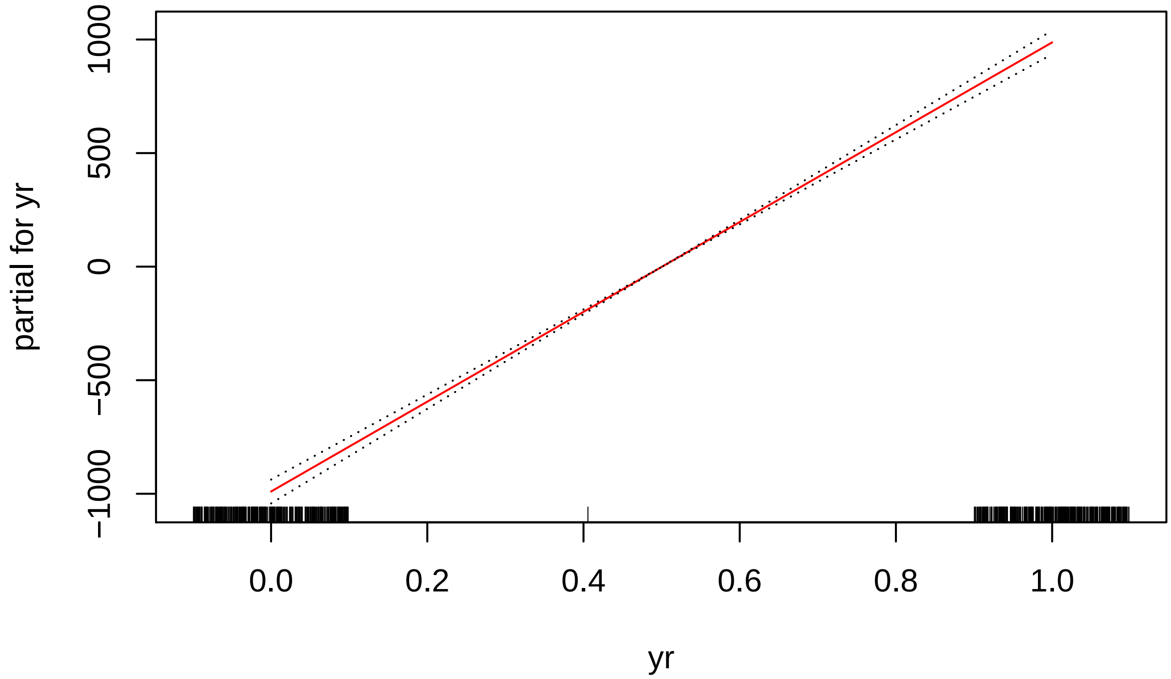

s(hum, df = 4.548876)
500
0
−1000
0.0
0.2
0.4
0.6
0.8
1.0
hum


0
1
2
3
4
5
6
partial for weekday
600
200
0
−400
weekday


partial for workingday
200
0
−200
−600
0.0
0.2
0.4
0.6
0.8
1.0
workingday


1
2
3
partial for weathersit
0
−500
−1000
−1500
weathersit


1
2
3
4
5
6
7
8
9
10
11
12
500
0
−500
−1000
partial for mnth
mnth
1.0
0.5
0.0
−0.5
−1.0
partial for holiday
0.0
0.2
0.4
0.6
0.8
1.0
holiday



```r
summary(gam_train2)
```

```
##
## Call: gam(formula = cnt ~ s(temp, df = 9.103704) + s(windspeed, df = 6.007664) +
##     s(atemp, df = 8.805497) + s(hum, df = 4.548876) + weekday +
##     workingday + weathersit + mnth + holiday + yr, data = day)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3114.68  -330.88    42.92   423.07  2154.85
##
## (Dispersion Parameter for gaussian family taken to be 479515)
##
##     Null Deviance: 2739535392 on 730 degrees of freedom
## Residual Deviance: 326326105 on 680.5337 degrees of freedom
## AIC: 11687
##
## Number of Local Scoring Iterations: 16
##
## Anova for Parametric Effects
##                                 Df     Sum Sq    Mean Sq   F value
## s(temp, df = 9.103704)        1.00 1028707877 1028707877 2145.3093
## s(windspeed, df = 6.007664)   1.00   59263290   59263290  123.5901
## s(atemp, df = 8.805497)       1.00      59461      59461    0.1240
## s(hum, df = 4.548876)         1.00  214861672  214861672  448.0813
## weekday                       6.00   13950686    2325114    4.8489
## workingday                    1.00    4493000    4493000    9.3699
## weathersit                    2.00   36367346   18183673   37.9210
## mnth                         11.00   83098967    7554452   15.7544
## yr                            1.00  683063628  683063628 1424.4887
## Residuals                   680.53  326326105     479515
##                                Pr(>F)
## s(temp, df = 9.103704)      < 2.2e-16 ***
## s(windspeed, df = 6.007664) < 2.2e-16 ***
## s(atemp, df = 8.805497)      0.724843
## s(hum, df = 4.548876)       < 2.2e-16 ***
## weekday                     7.330e-05 ***
## workingday                   0.002293 **
## weathersit                  2.431e-16 ***
## mnth                        < 2.2e-16 ***
## yr                          < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                             Npar Df Npar F     Pr(F)
## (Intercept)
## s(temp, df = 9.103704)          8.1 39.429 < 2.2e-16 ***
## s(windspeed, df = 6.007664)     5.0  5.989 1.939e-05 ***
## s(atemp, df = 8.805497)         7.8  5.716 6.155e-07 ***
## s(hum, df = 4.548876)           3.5  6.646 7.004e-05 ***
## weekday
## workingday
## weathersit
## mnth
## holiday
## yr
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Prediccion 2

```r
#Vamos a predecir para saber el error. Vemos que es practicamente 0 por lo que
#voy a realizar otro modelo sin las variables casual y register.
predict_modelo_gam2 <- predict(gam1.2,day_test)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading
```

```r
test_error_gam2 <- mean((predict_modelo_gam2 - day_test$cnt)^2)
test_error_gam2
```

```
## [1] 574417
```

#Error final

```r
sqrt(test_error_gam2)
```

```
## [1] 757.903
```

```r
#Tras la realizacion de los dos modelos, concluimos que las variables casual y register
#no son necesarias ya que la suma de ambas es el resultado de cnt.
#Por lo tanto, centrandonos en el segundo modelo, aplicando los test pertinentes, tenemos
#un error de 757.903 que teniendo en cuenta que la media de registros esta al rededor de 4000,
#es muy buen error
```