AEMOD: Ejercicio 2 (Fichero College)

Inmaculada Perea Fernández

Febrero 2017

Con el fichero College de la librería ISLR

Proponer un modelo gam para la variable Grad.Rate eligiendo la función que considere adecuada para cada variable predictora.

1. Carga de librerías necesarias

```
library(ISLR)
library(ggplot2)
library(gam)
```

2. Obtención e inspección del conjunto de datos para el estudio

El fichero College tiene las siguientes variables:

- Private: A factor with levels No and Yes indicating private or public university
- Apps: Number of applications received (socicitudes recibidas)
- Accept: Number of applications accepted (solicitudes aceptadas)
- Enroll: Number of new students enrolled (numero de nuevos estuciantes matriculados)
- Top10perc: Pct. new students from top 10% of H.S. class (top 10)
- Top25perc: Pct. new students from top 25% of H.S. class (top 25)
- F.Undergrad: Number of fulltime undergraduates (matriculados a tiempo completo)
- P.Undergrad: Number of parttime undergraduates (matriculados a tiempo parcial)
- Outstate: Out-of-state tuition (coste matricula para no residentes)
- Room.Board: Room and board costs (gastos en alojamiento y pension)
- Books: Estimated book costs (coste libros)
- Personal: Estimated personal spending (gastos personales)
- PhD: Pct. of faculty with Ph.D.'s (profesorado con Phd)
- Terminal: Pct. of faculty with terminal degree
- S.F.Ratio: Student/faculty ratio
- perc.alumni: Pct. alumni who donate
- Expend: Instructional expenditure per student
- Grad.Rate: Graduation rate

```
data(College)
college <- College
attach(college)
dim(college)
```

```
## [1] 777  18
```

```
str(college)
```

```
## 'data.frame':    777 obs. of  18 variables:
##  $ Private    : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
```

```
##  $ Apps       : num  1660 2186 1428 417 193 ...
##  $ Accept     : num  1232 1924 1097 349 146 ...
##  $ Enroll     : num  721 512 336 137 55 158 103 489 227 172 ...
##  $ Top10perc  : num  23 16 22 60 16 38 17 37 30 21 ...
##  $ Top25perc  : num  52 29 50 89 44 62 45 68 63 44 ...
##  $ F.Undergrad: num  2885 2683 1036 510 249 ...
##  $ P.Undergrad: num  537 1227 99 63 869 ...
##  $ Outstate   : num  7440 12280 11250 12960 7560 ...
##  $ Room.Board : num  3300 6450 3750 5450 4120 ...
##  $ Books      : num  450 750 400 450 800 500 500 450 300 660 ...
##  $ Personal   : num  2200 1500 1165 875 1500 ...
##  $ PhD        : num  70 29 53 92 76 67 90 89 79 40 ...
##  $ Terminal   : num  78 30 66 97 72 73 93 100 84 41 ...
##  $ S.F.Ratio  : num  18.1 12.2 12.9 7.7 11.9 9.4 11.5 13.7 11.3 11.5 ...
##  $ perc.alumni: num  12 16 30 37 2 11 26 37 23 15 ...
##  $ Expend     : num  7041 10527 8735 19016 10922 ...
##  $ Grad.Rate  : num  60 56 54 59 15 55 63 73 80 52 ...
```

```r
head(college)
```

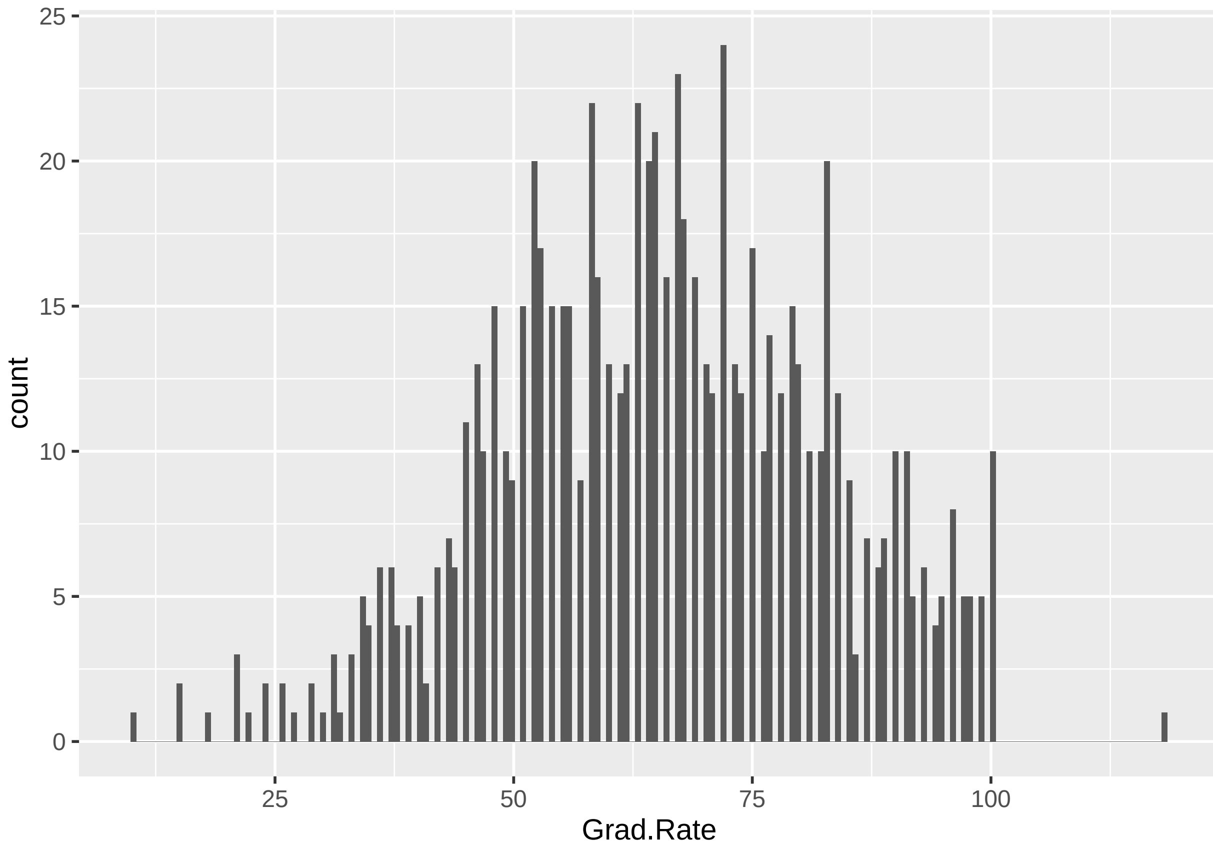
```
##                              Private Apps Accept Enroll Top10perc
## Abilene Christian University     Yes 1660   1232    721        23
## Adelphi University               Yes 2186   1924    512        16
## Adrian College                   Yes 1428   1097    336        22
## Agnes Scott College              Yes  417    349    137        60
## Alaska Pacific University        Yes  193    146     55        16
## Albertson College                Yes  587    479    158        38
##                              Top25perc F.Undergrad P.Undergrad Outstate
## Abilene Christian University        52        2885         537     7440
## Adelphi University                  29        2683        1227    12280
## Adrian College                      50        1036          99    11250
## Agnes Scott College                 89         510          63    12960
## Alaska Pacific University           44         249         869     7560
## Albertson College                   62         678          41    13500
##                              Room.Board Books Personal PhD Terminal
## Abilene Christian University       3300   450     2200  70       78
## Adelphi University                 6450   750     1500  29       30
## Adrian College                     3750   400     1165  53       66
## Agnes Scott College                5450   450      875  92       97
## Alaska Pacific University          4120   800     1500  76       72
## Albertson College                  3335   500      675  67       73
##                              S.F.Ratio perc.alumni Expend Grad.Rate
## Abilene Christian University      18.1          12   7041        60
## Adelphi University                12.2          16  10527        56
## Adrian College                    12.9          30   8735        54
## Agnes Scott College                7.7          37  19016        59
## Alaska Pacific University         11.9           2  10922        15
## Albertson College                  9.4          11   9727        55
```

```r
summary(college)
```

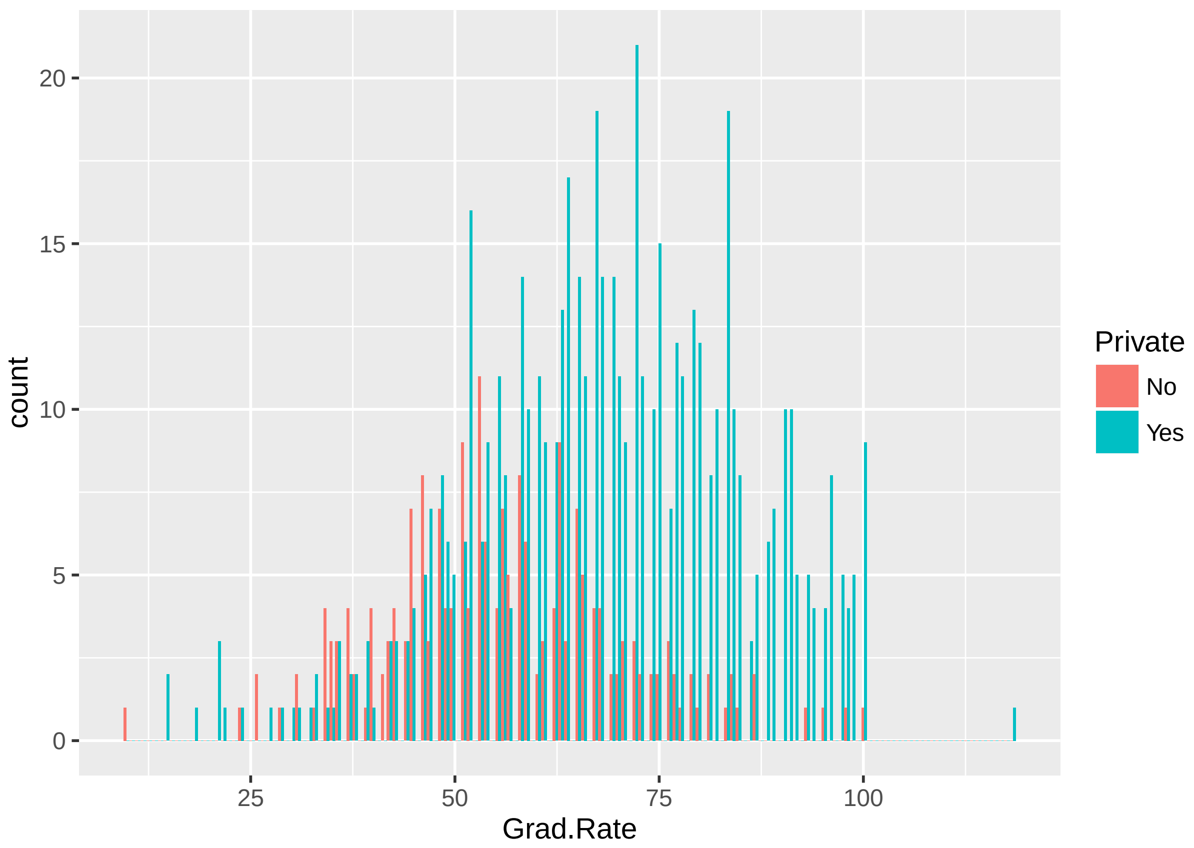
```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00
```

```
##            Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56
##            3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00
##            Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00
##    Top25perc      F.Undergrad     P.Undergrad         Outstate
##   Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##   1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##   Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##   Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
##   Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
##    Room.Board       Books           Personal         PhD
##   Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00
##   1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00
##   Median :4200   Median : 500.0   Median :1200   Median : 75.00
##   Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66
##   3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00
##   Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00
##     Terminal       S.F.Ratio      perc.alumni        Expend
##   Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186
##   1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751
##   Median : 82.0   Median :13.60   Median :21.00   Median : 8377
##   Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660
##   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830
##   Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233
##    Grad.Rate
##   Min.   : 10.00
##   1st Qu.: 53.00
##   Median : 65.00
##   Mean   : 65.46
##   3rd Qu.: 78.00
##   Max.   :118.00
```

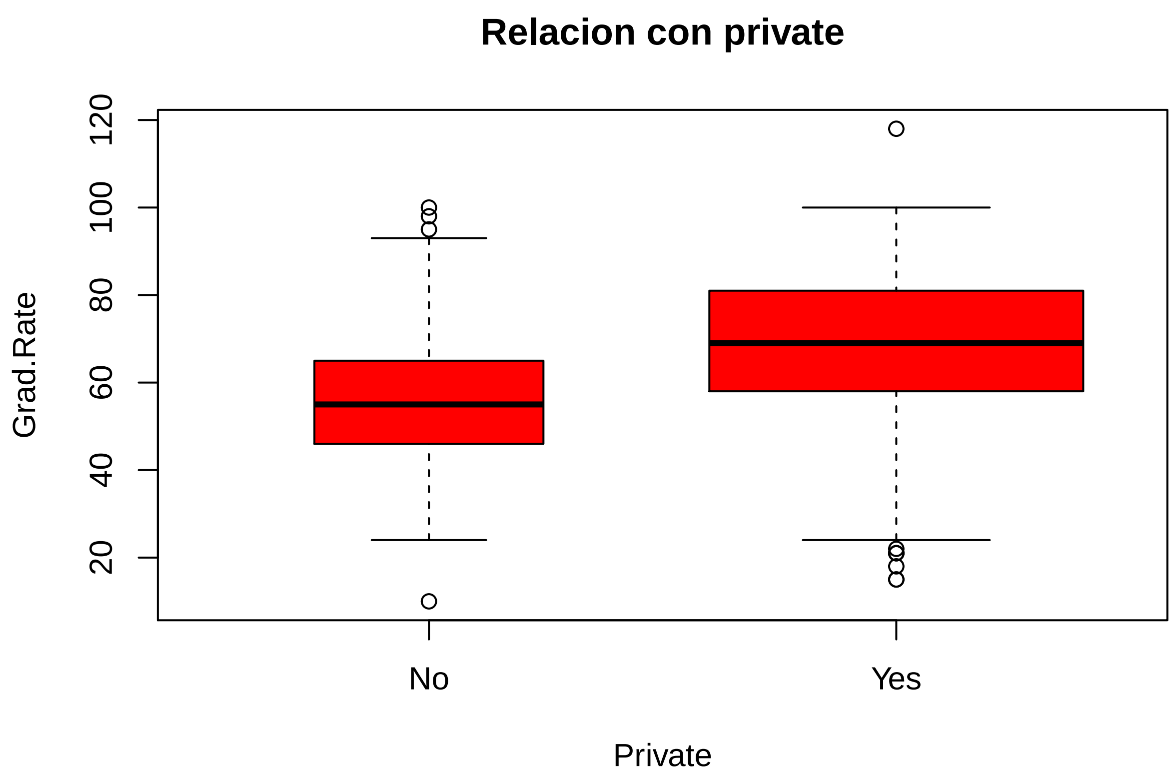
```r
ggplot(college, aes(Grad.Rate)) +  geom_histogram(binwidth=.6, position="dodge")
```




```r
ggplot(college, aes(Grad.Rate, fill = Private)) +
  geom_histogram(binwidth=.7, position="dodge")
```




```r
plot(Private, Grad.Rate, varwidth=T, col="red",
     xlab="Private", ylab="Grad.Rate",
     main="Relacion con private")
```



2.2 Seleccion de variables

El conjunto de datos existe un elevado número de variables, vamos a aplicar alguna de las técnicas de reduccion de la dimensionalidad estudiadas para simplificar el problema.

Utilizamos procedimiento de selección secuencial hacia delante proporcionado por la funcion **Step**, que usa como criterio de selección el AIC.

```
# Partimos del modelo lineal que incluye todas las variables
modeloRL=lm(Grad.Rate~.,data=college)

modelostep_fw=step(modeloRL,direction="forward", trace=FALSE)
min(modelostep_fw$anova$AIC)
```

```
## [1] 3972.977
```

```
modelostep_fw$call
```

```
## lm(formula = Grad.Rate ~ Private + Apps + Accept + Enroll + Top10perc +
##     Top25perc + F.Undergrad + P.Undergrad + Outstate + Room.Board +
##     Books + Personal + PhD + Terminal + S.F.Ratio + perc.alumni +
##     Expend, data = college)
```

```
modelostep_bw=step(modeloRL,direction="backward", trace=FALSE)
min(modelostep_bw$anova$AIC)
```

```
## [1] 3965.382
```

```
modelostep_bw$call
```

```
## lm(formula = Grad.Rate ~ Private + Apps + Top25perc + P.Undergrad +
##     Outstate + Room.Board + Personal + PhD + Terminal + perc.alumni +
##     Expend, data = college)
```

```
modelostep_both=step(modeloRL,direction="both", trace=FALSE)
min(modelostep_both$anova$AIC)
```

```
## [1] 3965.382
```

```
modelostep_both$call
```

```
## lm(formula = Grad.Rate ~ Private + Apps + Top25perc + P.Undergrad +
##     Outstate + Room.Board + Personal + PhD + Terminal + perc.alumni +
##     Expend, data = college)
```

Tanto el procedimiento hacia adelante como el que combina ambas direcciones proporcionan el mismo modelo.

El procedimiento **elimina** las siguientes variables:

- **Accept**: Number of applications accepted (solicitudes aceptadas)
- **Enroll**: Number of new students enrolled (numero de nuevos estuciantes matriculados)
- **Top10perc**: Pct. new students from top 10% of H.S. class (top 10)
- **F.Undergrad**: Number of fulltime undergraduates (matriculados a tiempo completo)
- **Books**: Estimated book costs (coste libros)
- **S.F.Ratio**: Student/faculty ratio

Las variables anteriores se pueden eliminar del estudio, ya que no son significativas para el estudio de la variable objetivo o porque son combinaciones de otras y por tanto ya stán explicadas por esas otras variables.

3. Construcción del modelo aditivo generalizado

Partiré del modelo obtenido tras la selección de varibales realizada en el apartado anterior e iré estudiando la funcion que mejor ajusta cada variable en el modelo aditivo.

```
summary(modelostep_bw)
```

```
##
## Call:
## lm(formula = Grad.Rate ~ Private + Apps + Top25perc + P.Undergrad +
##     Outstate + Room.Board + Personal + PhD + Terminal + perc.alumni +
##     Expend, data = college)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.684  -7.488  -0.282   7.363  53.482
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.4888648  3.3489573  10.000  < 2e-16 ***
## PrivateYes   3.5847682  1.6283712   2.201  0.02800 *
## Apps         0.0008950  0.0001609   5.563 3.67e-08 ***
## Top25perc    0.1697318  0.0321993   5.271 1.76e-07 ***
## P.Undergrad -0.0016749  0.0003631  -4.613 4.65e-06 ***
## Outstate     0.0010061  0.0002257   4.458 9.51e-06 ***
```

```
## Room.Board   0.0018799  0.0005795   3.244  0.00123 **
## Personal    -0.0018516  0.0007485  -2.474  0.01358 *
## PhD          0.0997365  0.0554704   1.798  0.07257 .
## Terminal    -0.0950484  0.0612000  -1.553  0.12082
## perc.alumni  0.2887259  0.0484841   5.955 3.96e-09 ***
## Expend      -0.0003942  0.0001290  -3.055  0.00233 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.73 on 765 degrees of freedom
## Multiple R-squared:  0.4585, Adjusted R-squared:  0.4507
## F-statistic: 58.88 on 11 and 765 DF,  p-value: < 2.2e-16
```

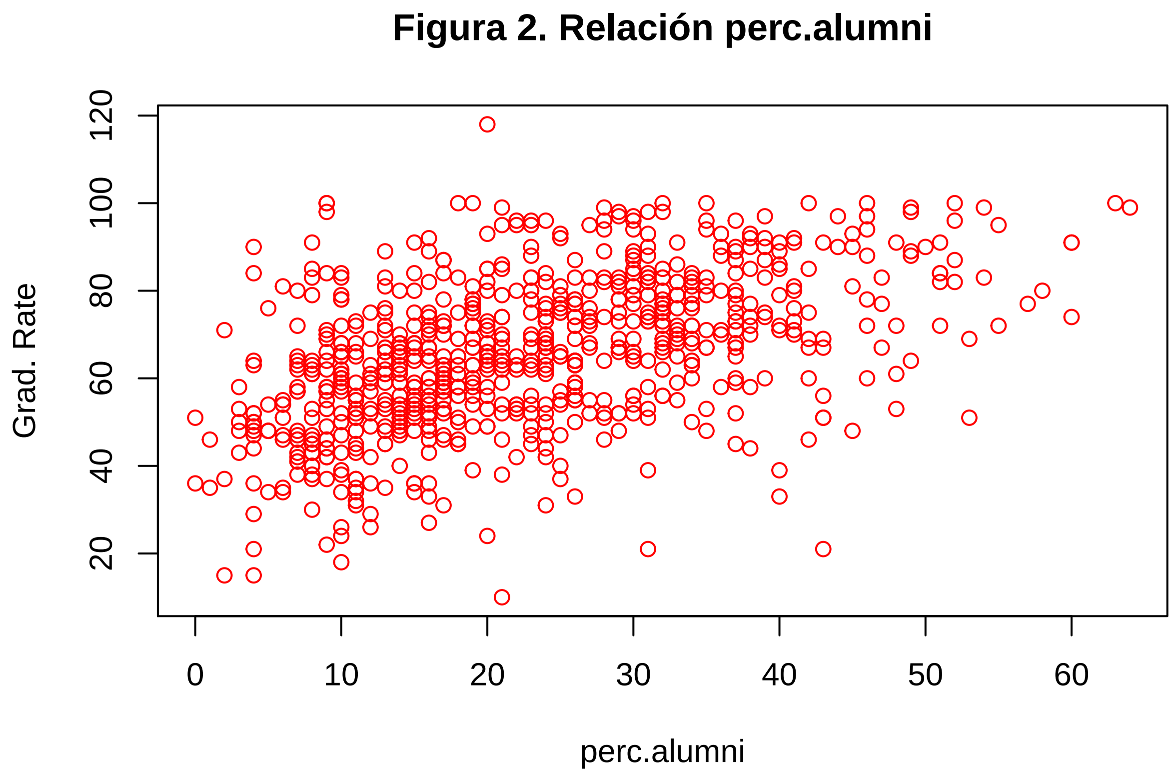
```r
modelostep_bw
```

```
##
## Call:
## lm(formula = Grad.Rate ~ Private + Apps + Top25perc + P.Undergrad +
##     Outstate + Room.Board + Personal + PhD + Terminal + perc.alumni +
##     Expend, data = college)
##
## Coefficients:
## (Intercept)   PrivateYes         Apps    Top25perc  P.Undergrad
##  33.4888648    3.5847682    0.0008950    0.1697318   -0.0016749
##    Outstate   Room.Board     Personal          PhD     Terminal
##   0.0010061    0.0018799   -0.0018516    0.0997365   -0.0950484
## perc.alumni       Expend
##   0.2887259   -0.0003942
```

A la vista de los coeficientes las variables que parecen tener más influencia en la variable respuesta son en este orden: *PrivateYes*, *perc.alumni*, *Top25perc*, *PhD*, *Terminal*, *Room.Board*, *Personal*, *Undergrad*, *Outstate*, *Apps* y *Expend*

3.1 Estudio de la variable *perc.alumni*

```r
plot(perc.alumni, Grad.Rate,
     xlab="perc.alumni",  ylab="Grad. Rate",
     main="Figura 2. Relación perc.alumni",
     col="red")
```



Ajuste con splines suavizados

```
fit_spl.alu=smooth.spline(perc.alumni, Grad.Rate ,cv=TRUE)

spl.alu.1=gam(Grad.Rate ~ Private ,data=college)
spl.alu.2=gam(Grad.Rate ~ Private + perc.alumni ,data=college)
spl.alu.3=gam(Grad.Rate ~ Private + s(perc.alumni, 1), data=college)
spl.alu.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2), data=college)
spl.alu.5=gam(Grad.Rate ~ Private + s(perc.alumni, round(fit_spl.alu$df, 3)), data=college)

anova(spl.alu.1, spl.alu.2, spl.alu.3, spl.alu.4, spl.alu.5, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private
## Model 2: Grad.Rate ~ Private + perc.alumni
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, round(fit_spl.alu$df, 3))
##   Resid. Df Resid. Dev         Df Deviance        F    Pr(>F)
## 1    775.00     203102
## 2    774.00     168939 1.0000e+00    34162 157.5716 < 2.2e-16 ***
## 3    774.00     168939 2.2462e-07        0  44.7355 1.306e-06 ***
## 4    773.00     167787 9.9989e-01     1152   5.3157    0.0214 *
## 5    771.46     167257 1.5410e+00      530   1.5872    0.2094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste polinómico

```
poly.alu.1= gam(Grad.Rate~Private + perc.alumni ,data=college)
poly.alu.2= gam(Grad.Rate~Private + poly(perc.alumni ,2) ,data=college)
poly.alu.3= gam(Grad.Rate~Private + poly(perc.alumni ,3) ,data=college)
poly.alu.4= gam(Grad.Rate~Private + poly(perc.alumni ,4) ,data=college)
poly.alu.5= gam(Grad.Rate~Private + poly(perc.alumni ,5) ,data=college)

anova(poly.alu.1, poly.alu.2, poly.alu.3, poly.alu.4, poly.alu.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + perc.alumni
## Model 2: Grad.Rate ~ Private + poly(perc.alumni, 2)
## Model 3: Grad.Rate ~ Private + poly(perc.alumni, 3)
## Model 4: Grad.Rate ~ Private + poly(perc.alumni, 4)
## Model 5: Grad.Rate ~ Private + poly(perc.alumni, 5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       774     168939
## 2       773     167930  1  1009.56  0.03106 *
## 3       772     167368  1   561.86  0.10770
## 4       771     167248  1   120.02  0.45721
## 5       770     167192  1    55.58  0.61291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(spl.alu.3, spl.alu.4, poly.alu.2)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 1)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2)
## Model 3: Grad.Rate ~ Private + poly(perc.alumni, 2)
##   Resid. Df Resid. Dev         Df Deviance Pr(>Chi)
## 1       774     168939
## 2       773     167787 0.99989137  1152.35  0.02127 *
## 3       773     167930 0.00010841  -142.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
spl.alu.3$aic
```

```
## [1] 6394.731
```

```
spl.alu.4$aic
```

```
## [1] 6391.413
```

```
poly.alu.2$aic
```

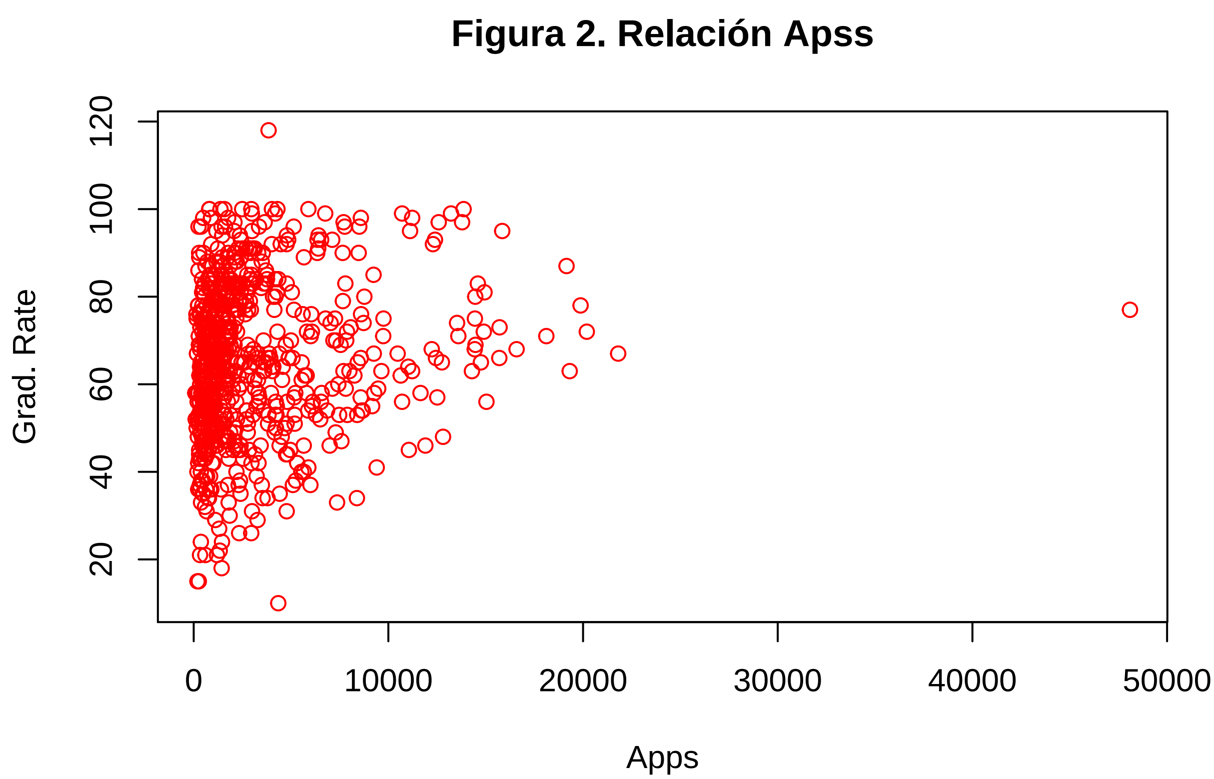
```
## [1] 6392.074
```

Modelo resultante del estudio: gam(Grad.Rate ~ Private + s(perc.alumni, 2), data=college)

Da mejores resultados modelando la variable perc.alumni con splines de grado 2

3.2 Estudio de la variable *Apps*

```
plot(Apps, Grad.Rate,
     xlab="Apps",  ylab="Grad. Rate",
     main="Figura 2. Relación Apss",
     col="red")
```



Ajuste con splines suavizados

```
fit_spl.app=smooth.spline(Apps, Grad.Rate ,cv=TRUE)
fit_spl.app$df
```

```
## [1] 6.582906
```

```
spl.app.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2), data=college)
spl.app.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + Apps ,data=college)
spl.app.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, 1), data=college)
spl.app.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, 2), data=college)
spl.app.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, round(fit_spl.app$df,3)), data=college)

anova(spl.app.1, spl.app.2, spl.app.3, spl.app.4, spl.app.5, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + Apps
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, 1)
```

```
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, round(fit_spl.app$df,
##     3))
##   Resid. Df Resid. Dev       Df Deviance        F    Pr(>F)
## 1    773.00     167787
## 2    772.00     149608 1.000000  18179.2 100.3194 < 2.2e-16 ***
## 3    772.00     149597 0.001568     11.0  38.8249  0.002308 **
## 4    771.00     144871 0.998418   4725.6  26.1191 4.127e-07 ***
## 5    766.42     138885 4.582873   5986.0   7.2079 3.137e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste polinómico

```r
poly.alu.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + Apps, data=college)
poly.alu.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,2) ,data=college)
poly.alu.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,3) ,data=college)
poly.alu.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,4) ,data=college)
poly.alu.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) ,data=college)

anova(poly.alu.1, poly.alu.2, poly.alu.3, poly.alu.4, poly.alu.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + Apps
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       772     149608
## 2       771     145456  1   4151.2 1.678e-06 ***
## 3       770     142982  1   2474.3 0.0002180 ***
## 4       769     140978  1   2004.3 0.0008761 ***
## 5       768     139021  1   1956.6 0.0010101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(spl.app.4, spl.app.5, poly.alu.4, poly.alu.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, 2)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + s(Apps, round(fit_spl.app$df,
##     3))
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 4)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5)
##   Resid. Df Resid. Dev      Df Deviance  Pr(>Chi)
## 1    771.00     144871
## 2    766.42     138885  4.5829   5986.0 2.335e-06 ***
## 3    769.00     140978 -2.5829  -2092.8  0.006023 **
## 4    768.00     139021  1.0000   1956.6  0.001017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
spl.app.4$aic
```

```
## [1] 6281.309
```

```
spl.app.5$aic
```

```
## [1] 6257.688
```

```
poly.alu.4$aic
```

```
## [1] 6264.143
```

```
poly.alu.5$aic
```

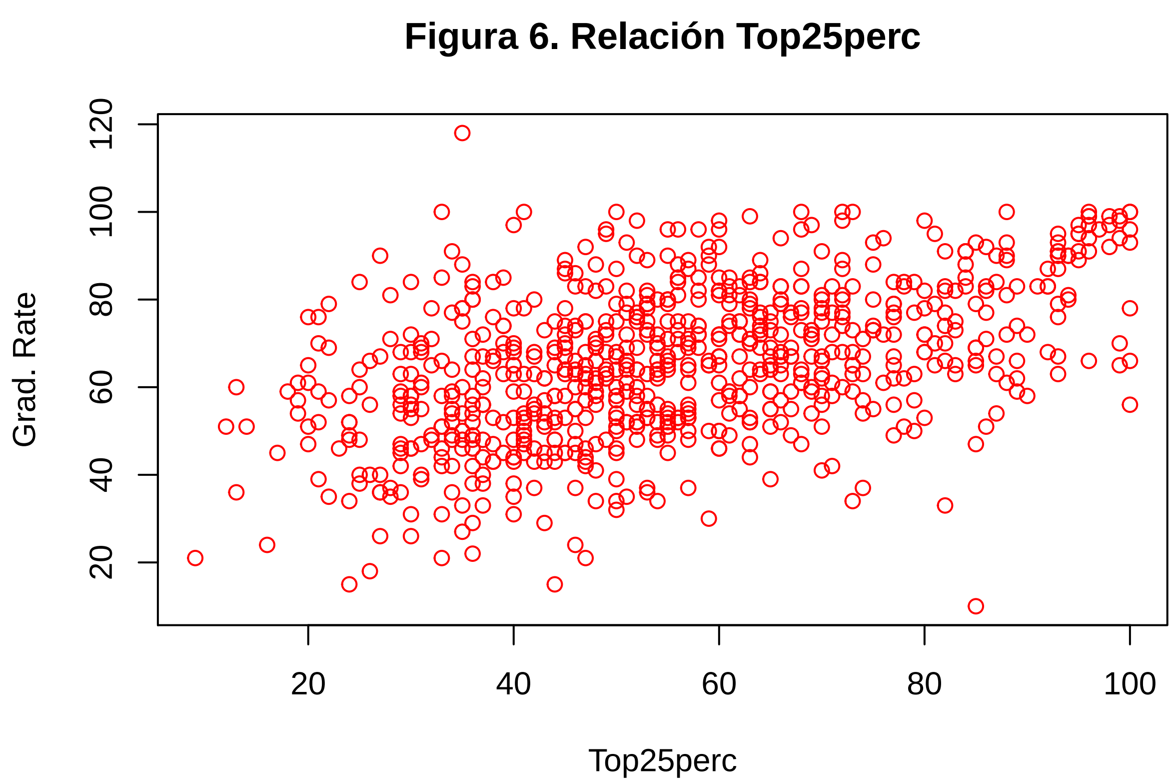
```
## [1] 6255.284
```

La funcion que mejor ajusta la variable *Apps* es el polinomio de orden 5, por tanto, el mejor modelo hata ahora es:

poly.alu.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) ,data=college)

3.3 Estudio de la variable *Top25perc*

```
plot(Top25perc, Grad.Rate,
     xlab="Top25perc",  ylab="Grad. Rate",
     main="Figura 6. Relación Top25perc",
     col="red")
```



Ajuste con splines suavizados

```r
fit_spl.top=smooth.spline(Top25perc, Grad.Rate ,cv=TRUE)
fit_spl.top$df
```

```
## [1] 5.241798
```

```r
spl.top.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5), data=college)

spl.top.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc ,data=college)

spl.top.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + s(Top25perc, 1), data=college)

spl.top.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + s(Top25perc, 2), data=college)

spl.top.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + s(Top25perc,
                          round(fit_spl.top$df,3)), data=college)

anova(spl.top.1, spl.top.2, spl.top.3, spl.top.4, spl.top.5, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + s(Top25perc,
##     1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + s(Top25perc,
##     2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + s(Top25perc,
##     round(fit_spl.top$df, 3))
##   Resid. Df Resid. Dev         Df Deviance       F    Pr(>F)
## 1    768.00     139021
## 2    767.00     132952 1.0000e+00   6069.4 35.1820 4.552e-09 ***
## 3    767.00     132952 6.2001e-07      0.0  3.4077 4.087e-06 ***
## 4    766.00     132741 1.0000e+00    210.8  1.2219   0.26934
## 5    762.76     131588 3.2420e+00   1153.2  2.0619   0.09866 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con polinomios hasta orden 5

```r
poly.top.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc,
                data=college)

poly.top.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + poly(Top25perc ,2),
                data=college)

poly.top.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + poly(Top25perc ,3),
                data=college)

poly.top.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + poly(Top25perc ,4),
                data=college)

poly.top.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + poly(Top25perc ,5),
                data=college)
```

```r
anova(poly.top.1, poly.top.2, poly.top.3, poly.top.4, poly.top.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + poly(Top25perc,
##     2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + poly(Top25perc,
##     3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + poly(Top25perc,
##     4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + poly(Top25perc,
##     5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       767     132952
## 2       766     132946  1     5.53   0.8580
## 3       765     132538  1   407.98   0.1245
## 4       764     132187  1   350.85   0.1543
## 5       763     131910  1   276.99   0.2056
```

```r
spl.top.2$aic
```

```
## [1] 6222.598
```

```r
spl.top.3$aic
```

```
## [1] 6222.598
```

```r
spl.top.5$aic
```

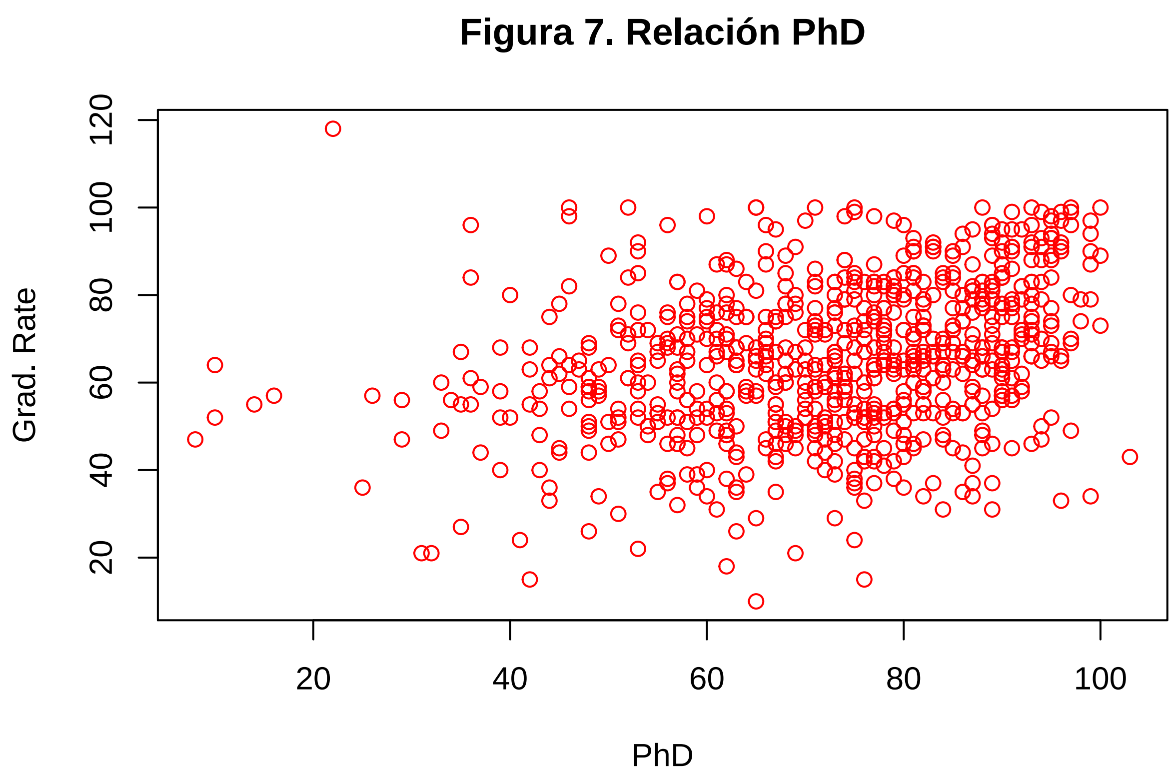
```
## [1] 6223.07
```

El mejor modelo es el que se muestra a continuación, para la variable *Top25perc* la mejor funcion es la lineal no es necesario añadir splines ni polinomios

spl.top.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc ,data=college)

3.4 Estudio de la variable *PhD*

```r
plot(PhD, Grad.Rate,
     xlab="PhD",  ylab="Grad. Rate",
     main="Figura 7. Relación PhD",
     col="red")
```



Ajuste mediante splines suavizados

```
fit_spl.phd=smooth.spline(PhD, Grad.Rate ,cv=TRUE)
fit_spl.phd$df
```

```
## [1] 4.32194
```

```
spl.phd.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc,
              data=college)

spl.phd.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
              PhD, data=college)

spl.phd.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
              s(PhD, 1), data=college)

spl.phd.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
              s(PhD, 2), data=college)

spl.phd.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
              s(PhD, round(fit_spl.phd$df,3)), data=college)

anova(spl.phd.1, spl.phd.2, spl.phd.3, spl.phd.4, spl.phd.5, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
```

```
##     PhD
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(PhD, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(PhD, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(PhD, round(fit_spl.phd$df, 3))
##   Resid. Df Resid. Dev         Df Deviance      F    Pr(>F)
## 1    767.00     132952
## 2    766.00     132728 1.0000e+00   224.11 1.2940    0.2557
## 3    766.00     132728 4.3638e-06     0.00 0.7397 2.784e-05 ***
## 4    765.00     132545 9.9987e-01   182.16 1.0519    0.3054
## 5    762.68     132096 2.3224e+00   449.34 1.1171    0.3335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste polinómico

```r
poly.phd.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                PhD, data=college)

poly.phd.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                poly(PhD ,2), data=college)

poly.phd.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                poly(PhD ,3), data=college)

poly.phd.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                poly(PhD ,4), data=college)

poly.phd.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                poly(PhD ,5), data=college)

anova(poly.phd.1, poly.phd.2, poly.phd.3, poly.phd.4, poly.phd.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     PhD
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(PhD, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(PhD, 3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(PhD, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(PhD, 5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       766     132728
## 2       765     132679  1   48.299   0.5978
## 3       764     132520  1  159.100   0.3383
## 4       763     132276  1  244.725   0.2350
## 5       762     132219  1   56.504   0.5682
```

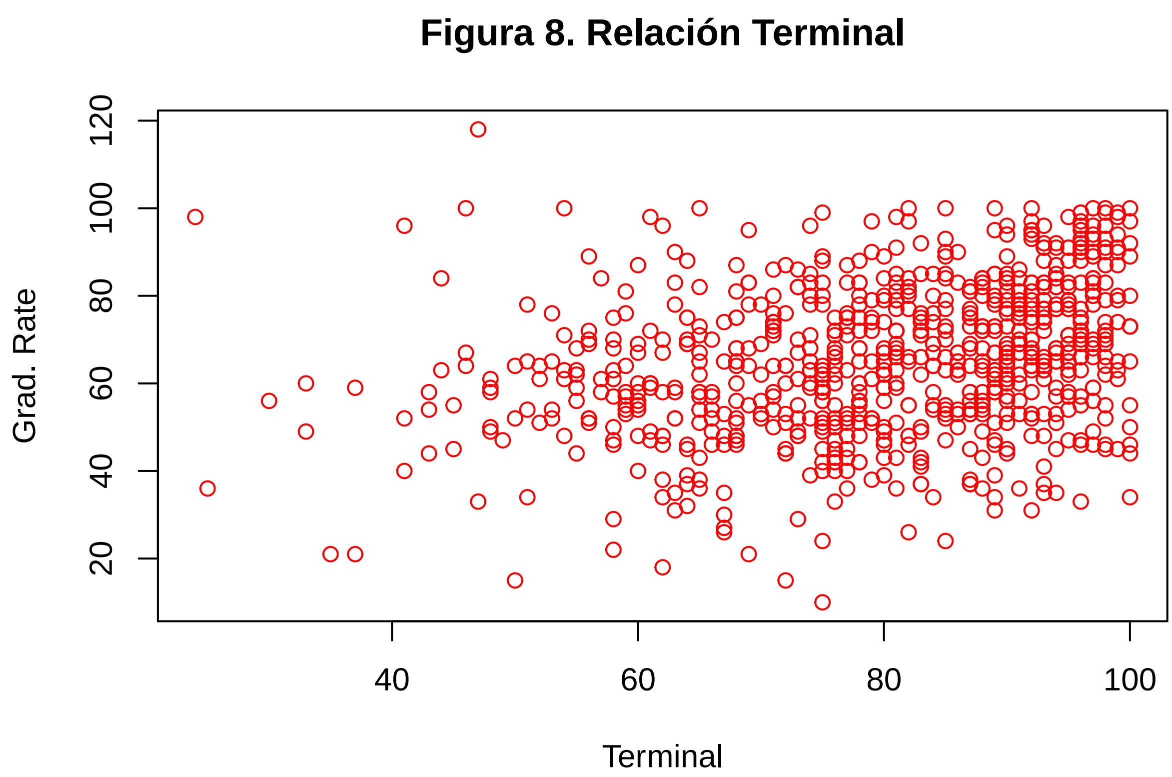
```r
spl.phd.3$aic
```

```
## [1] 6223.287
```

No mejora añadiendo la variable *PhD* al estudio, por tanto la eliminamos del modelo

3.5 Estudio de la variable *Terminal*

```
plot(Terminal, Grad.Rate,
     xlab="Terminal",  ylab="Grad. Rate",
     main="Figura 8. Relación Terminal",
     col="red")
```



Ajuste con splines suavizados

```
fit_spl.terminal=smooth.spline(Terminal, Grad.Rate ,cv=TRUE)
fit_spl.terminal$df
```

```
## [1] 3.248826
```

```
spl.terminal.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc,
                   data=college)

spl.terminal.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   Terminal, data=college)

spl.terminal.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Terminal, 1), data=college)
```

```r
spl.terminal.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Terminal, 2), data=college)

spl.terminal.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Terminal, round(fit_spl.terminal$df,3)), data=college)

anova(spl.terminal.1, spl.terminal.2, spl.terminal.3, spl.terminal.4, spl.terminal.5, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     Terminal
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Terminal, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Terminal, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Terminal, round(fit_spl.terminal$df, 3))
##   Resid. Df Resid. Dev         Df Deviance      F    Pr(>F)
## 1    767.00     132952
## 2    766.00     132938 1.0000e+00   13.674 0.0787    0.7792
## 3    766.00     132938 9.6015e-06    0.001 0.7695 5.729e-05 ***
## 4    765.00     132847 1.0001e+00   91.426 0.5259    0.4686
## 5    763.75     132748 1.2489e+00   98.448 0.4535    0.5435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con polinomios

```r
poly.terminal.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                     Terminal, data=college)

poly.terminal.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                     poly(Terminal, 2), data=college)

poly.terminal.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                     poly(Terminal, 3), data=college)

poly.terminal.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                     poly(Terminal, 4), data=college)

poly.terminal.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                     poly(Terminal, 5), data=college)

anova(poly.terminal.1, poly.terminal.2, poly.terminal.3, poly.terminal.4, poly.terminal.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     Terminal
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Terminal, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Terminal, 3)
```

```
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Terminal, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Terminal, 5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       766     132938
## 2       765     132870  1   68.277   0.5309
## 3       764     132762  1  107.326   0.4320
## 4       763     132762  1    0.419   0.9608
## 5       762     132477  1  284.871   0.2005
```

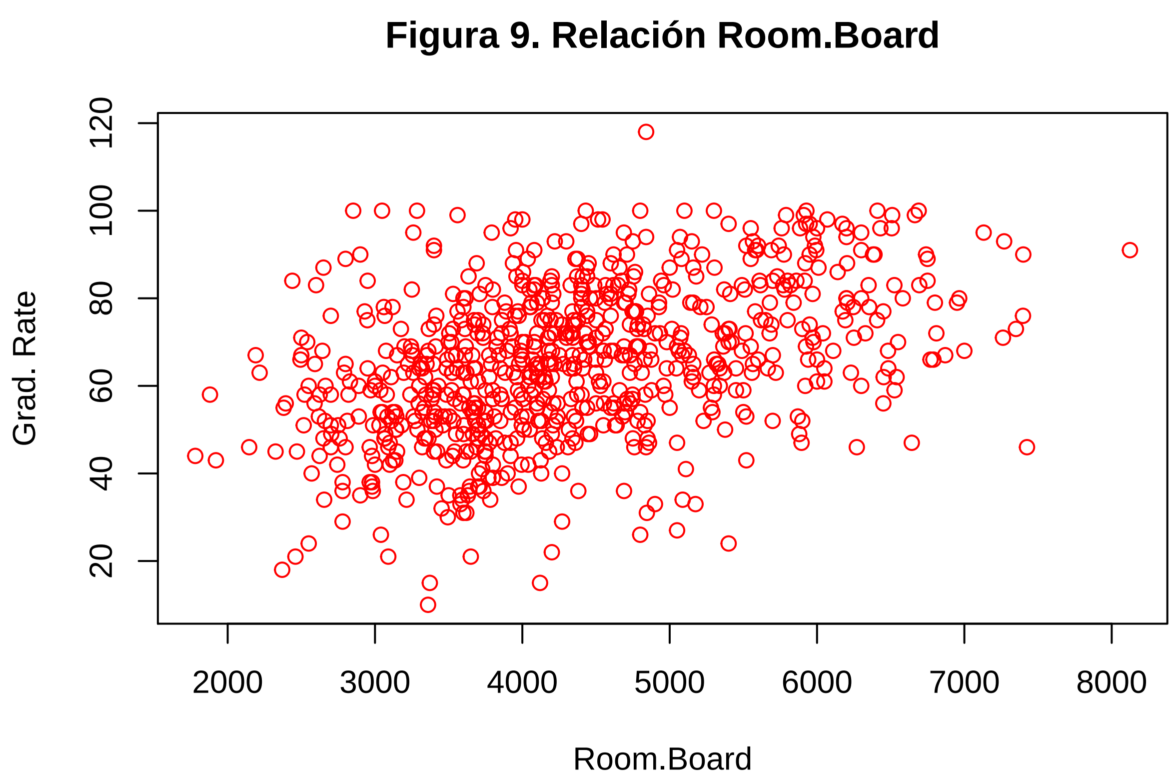
```
spl.terminal.3$aic
```

```
## [1] 6224.518
```

El modelo no mejora introduciendo la variable *Terminal*, por tanto lo eliminamos del modelo

3.6 Estudio de la variable *Room.Board*

```
plot(Room.Board, Grad.Rate,
     xlab="Room.Board",  ylab="Grad. Rate",
     main="Figura 9. Relación Room.Board",
     col="red")
```



Ajuste con splines suavizados

```r
fit_spl.room=smooth.spline(Room.Board, Grad.Rate ,cv=TRUE)
fit_spl.room$df
```

```
## [1] 8.046605
```

```r
spl.room.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc,
               data=college)

spl.room.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               Room.Board, data=college)

spl.room.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               s(Room.Board, 1), data=college)

spl.room.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               s(Room.Board, 2), data=college)

spl.room.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               s(Room.Board, 3), data=college)

spl.room.6=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               s(Room.Board, 4), data=college)

spl.room.7=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               s(Room.Board, 5), data=college)

spl.room.8=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
               s(Room.Board, round(fit_spl.room$df,3)), data=college)


anova(spl.room.1, spl.room.2, spl.room.3, spl.room.4, spl.room.5,
      spl.room.6, spl.room.7, spl.room.8, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     Room.Board
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 3)
## Model 6: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4)
## Model 7: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 5)
## Model 8: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, round(fit_spl.room$df, 3))
##   Resid. Df Resid. Dev         Df Deviance       F    Pr(>F)
## 1    767.00     132952
## 2    766.00     129368 1.0000e+00   3583.9 21.4974 4.171e-06 ***
## 3    766.00     129368 4.0807e-05      0.0  1.9460  0.000195 ***
```

```
## 4    765.00     128769 9.9998e-01    598.9  3.5927  0.058414 .
## 5    764.00     127986 1.0000e+00    782.8  4.6950  0.030560 *
## 6    763.00     127463 9.9997e-01    523.4  3.1397  0.076812 .
## 7    762.00     127131 1.0002e+00    331.3  1.9868  0.159086
## 8    758.95     126529 3.0467e+00    602.4  1.1861  0.314217
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con polinomios de hasta orden 5

```r
poly.room.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                 Room.Board, data=college)

poly.room.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                 poly(Room.Board, 2), data=college)

poly.room.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                 poly(Room.Board, 3), data=college)

poly.room.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                 poly(Room.Board, 4), data=college)

poly.room.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                 poly(Room.Board, 5), data=college)

anova(poly.room.1, poly.room.2, poly.room.3, poly.room.4, poly.room.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     Room.Board
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Room.Board, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Room.Board, 3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Room.Board, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     poly(Room.Board, 5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       766     129368
## 2       765     129366  1     2.01 0.912819
## 3       764     128240  1  1126.07 0.009579 **
## 4       763     127963  1   276.77 0.199010
## 5       762     127847  1   115.70 0.406294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
spl.room.2$aic
```

```
## [1] 6203.366
```

```r
spl.room.3$aic
```

```
## [1] 6203.366
```

```
spl.room.4$aic
```

```
## [1] 6201.76
```

```
spl.room.5$aic
```

```
## [1] 6199.022
```

```
spl.room.6$aic
```

```
## [1] 6197.838
```

```
poly.room.3$aic
```

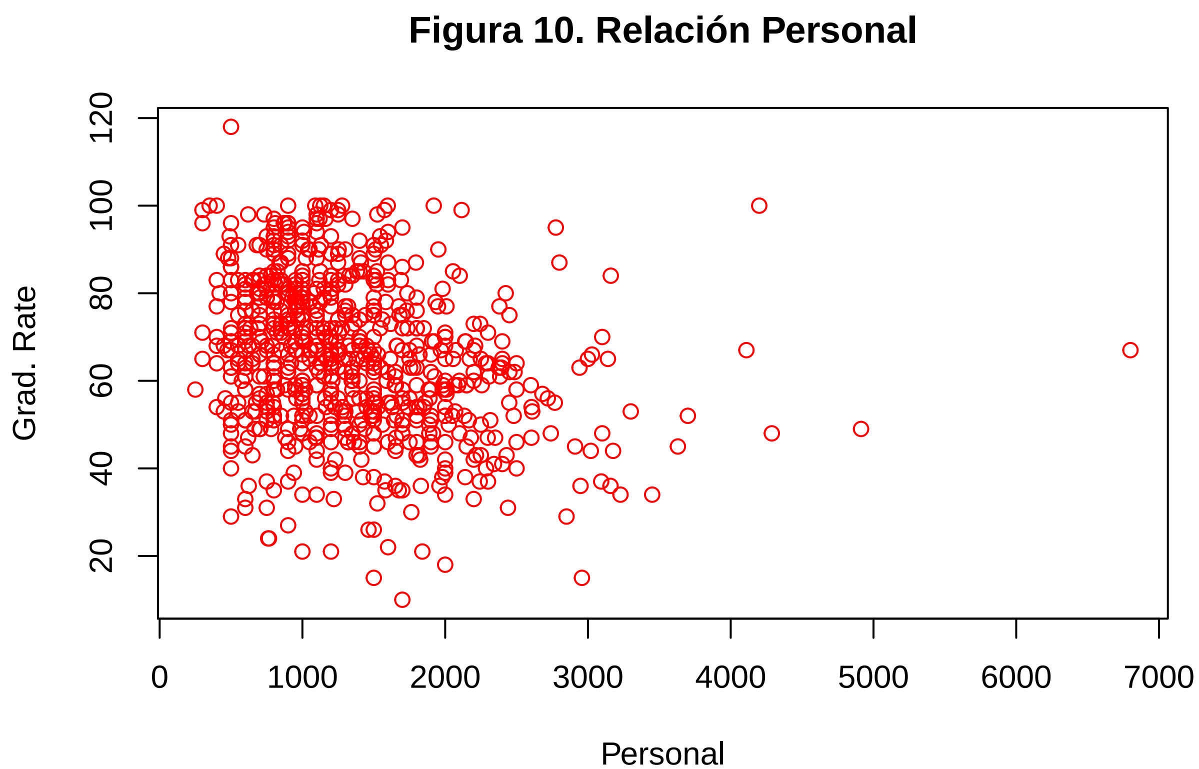
```
## [1] 6200.56
```

Vemos que la inclusión de la variable *Room.Board* mejora considerablemente el modelo y que obtenemos mejor resultado cuando se ajusta con splines de grado 4 de libertad El mejor modelo es:

spl.room.6=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc + s(Room.Board, 4), data=college)

3.7 Estudio de la variable *Personal*

```
plot(Personal, Grad.Rate,
     xlab="Personal",  ylab="Grad. Rate",
     main="Figura 10. Relación Personal",
     col="red")
```



Ajuste con splines suavizados

```
fit_spl.personal=smooth.spline(Personal, Grad.Rate ,cv=TRUE)
fit_spl.personal$df
```

```
## [1] 5.261014
```

```
spl.personal.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4), data=college)

spl.personal.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + Personal ,data=college)

spl.personal.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + s(Personal, 1), data=college)

spl.personal.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + s(Personal, 2), data=college)

spl.personal.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + s(Personal, 3), data=college)

spl.personal.6=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + s(Personal, 4), data=college)

spl.personal.7=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + s(Personal, 5), data=college)

spl.personal.8=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                   s(Room.Board, 4) + s(Personal, round(fit_spl.personal$df,3)), data=college)

anova(spl.personal.1, spl.personal.2, spl.personal.3, spl.personal.4, spl.personal.5,
      spl.personal.6, spl.personal.7, spl.personal.8, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + s(Personal, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + s(Personal, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + s(Personal, 3)
## Model 6: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + s(Personal, 4)
## Model 7: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + s(Personal, 5)
## Model 8: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + s(Personal, round(fit_spl.personal$df,
##     3))
##   Resid. Df Resid. Dev         Df Deviance        F    Pr(>F)
## 1    763.00     127463
```

```
## 2    762.00     124809 1.0000e+00  2653.43  16.2627 6.072e-05 ***
## 3    762.00     124809 3.9167e-07     0.01 134.5497 1.952e-06 ***
## 4    761.00     124422 1.0000e+00   386.80   2.3707    0.1241
## 5    760.00     124139 1.0001e+00   283.79   1.7392    0.1876
## 6    759.00     123898 9.9995e-01   240.45   1.4738    0.2251
## 7    758.00     123686 9.9970e-01   211.94   1.2993    0.2547
## 8    757.74     123633 2.6095e-01    53.14   1.2481    0.1755
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con polinomios de hasta orden 5

```r
poly.personal.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                      s(Room.Board, 4) + Personal, data=college)

poly.personal.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                      s(Room.Board, 4) + poly(Personal ,2) ,data=college)

poly.personal.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                      s(Room.Board, 4) + poly(Personal ,3) ,data=college)

poly.personal.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                      s(Room.Board, 4) + poly(Personal ,4) ,data=college)

poly.personal.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                      s(Room.Board, 4) + poly(Personal ,5) ,data=college)

anova(poly.personal.1, poly.personal.2, poly.personal.3, poly.personal.4, poly.personal.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + poly(Personal, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + poly(Personal, 3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + poly(Personal, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + poly(Personal, 5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       762     124809
## 2       761     124567  1   241.96  0.22383
## 3       760     124550  1    17.05  0.74674
## 4       759     124085  1   464.77  0.09182 .
## 5       758     123956  1   129.81  0.37296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
spl.personal.2$aic
```

```
## [1] 6183.492
```

```r
spl.personal.3$aic
```

```
## [1] 6183.492
```

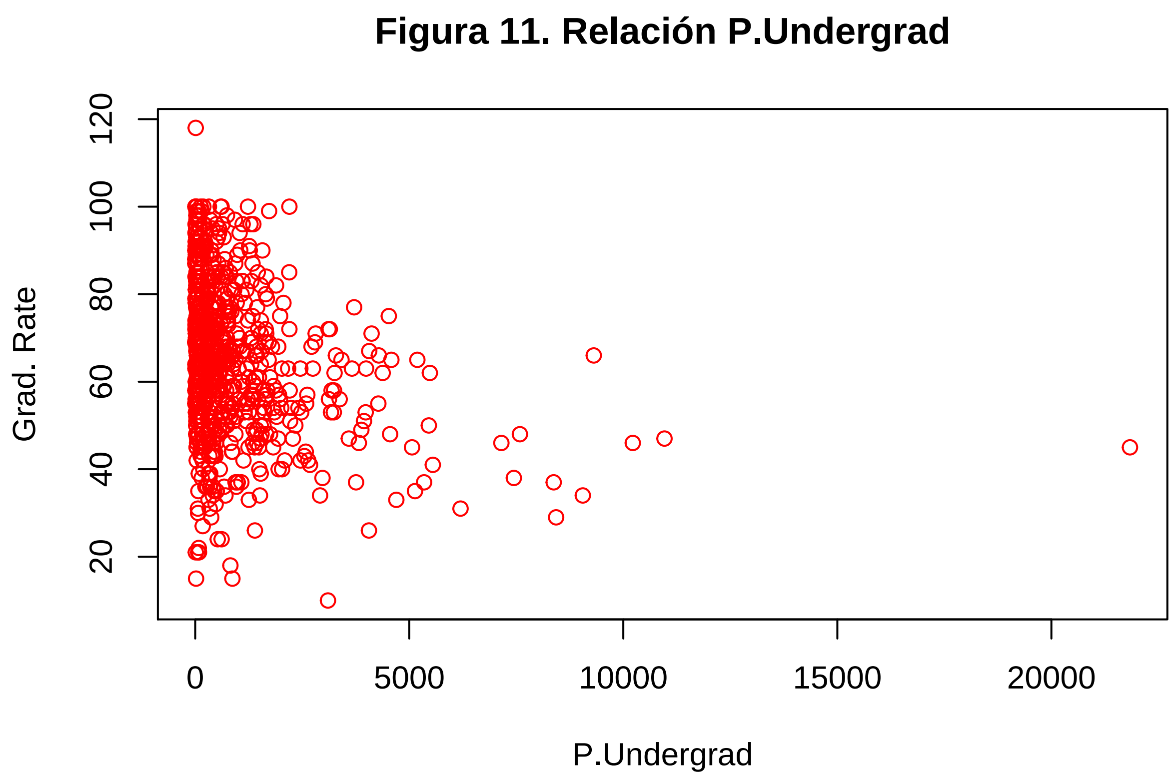
```r
poly.personal.4$aic
```

```
## [1] 6184.973
```

La inclusión de la variable *Personal* mejora el modelo. El mejor modelo obtenido hasta el momento es *spl.personal.2*

3.8 Estudio de la variable *P.Undergrad*

```r
plot(P.Undergrad, Grad.Rate,
     xlab="P.Undergrad",  ylab="Grad. Rate",
     main="Figura 11. Relación P.Undergrad",
     col="red")
```



Ajuste con splines suavizados

```r
fit_spl.under=smooth.spline(P.Undergrad, Grad.Rate ,cv=TRUE)
fit_spl.under$df
```

```
## [1] 13.63541
```

```r
spl.under.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal, data=college)

spl.under.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + P.Undergrad ,data=college)
```

```r
spl.under.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + s(P.Undergrad, 1), data=college)

spl.under.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + s(P.Undergrad, 2), data=college)

spl.under.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + s(P.Undergrad, 3), data=college)

spl.under.6=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + s(P.Undergrad, 4), data=college)

spl.under.7=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + s(P.Undergrad, 5), data=college)

spl.under.8=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + s(P.Undergrad, round(fit_spl.under$df,3)),
                data=college)

anova(spl.under.1, spl.under.2, spl.under.3, spl.under.4, spl.under.5, spl.under.6,
      spl.under.7, spl.under.8, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + P.Undergrad
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + s(P.Undergrad, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + s(P.Undergrad, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + s(P.Undergrad, 3)
## Model 6: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + s(P.Undergrad, 4)
## Model 7: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + s(P.Undergrad, 5)
## Model 8: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + s(P.Undergrad, round(fit_spl.under$df,
##     3))
##   Resid. Df Resid. Dev        Df Deviance       F    Pr(>F)
## 1    762.00     124809
## 2    761.00     120841 1.0000000   3968.1 25.1930 6.487e-07 ***
## 3    761.00     120839 0.0023149      1.7  4.5426   0.00540 **
## 4    760.00     120324 0.9975725    515.9  3.2833   0.07049 .
## 5    759.00     119985 1.0002293    339.1  2.1521   0.14279
## 6    758.00     119632 0.9998733    352.3  2.2368   0.13518
## 7    757.00     119385 1.0002849    246.8  1.5668   0.21108
## 8    748.37     117874 8.6348124   1511.2  1.1111   0.35254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con polinomios de hasta orden 5

```r
poly.under.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                  s(Room.Board, 4) + Personal + Personal, data=college)

poly.under.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                  s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) ,data=college)

poly.under.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                  s(Room.Board, 4) + Personal + poly(P.Undergrad ,3) ,data=college)

poly.under.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                  s(Room.Board, 4) + Personal + poly(P.Undergrad ,4) ,data=college)

poly.under.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                  s(Room.Board, 4) + Personal + poly(P.Undergrad ,5) ,data=college)

anova(poly.under.1, poly.under.2, poly.under.3, poly.under.4, poly.under.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + Personal
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 5)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       762     124809
## 2       760     120397  2   4412.4 8.554e-07 ***
## 3       759     120361  1     36.2   0.63223
## 4       758     119786  1    574.9   0.05638 .
## 5       757     119535  1    250.6   0.20778
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
spl.under.2$aic
```

```
## [1] 6160.388
```

```r
spl.under.3$aic
```

```
## [1] 6160.381
```

```r
spl.under.4$aic
```

```
## [1] 6159.052
```

```r
poly.under.2$aic
```

```
## [1] 6159.525
```

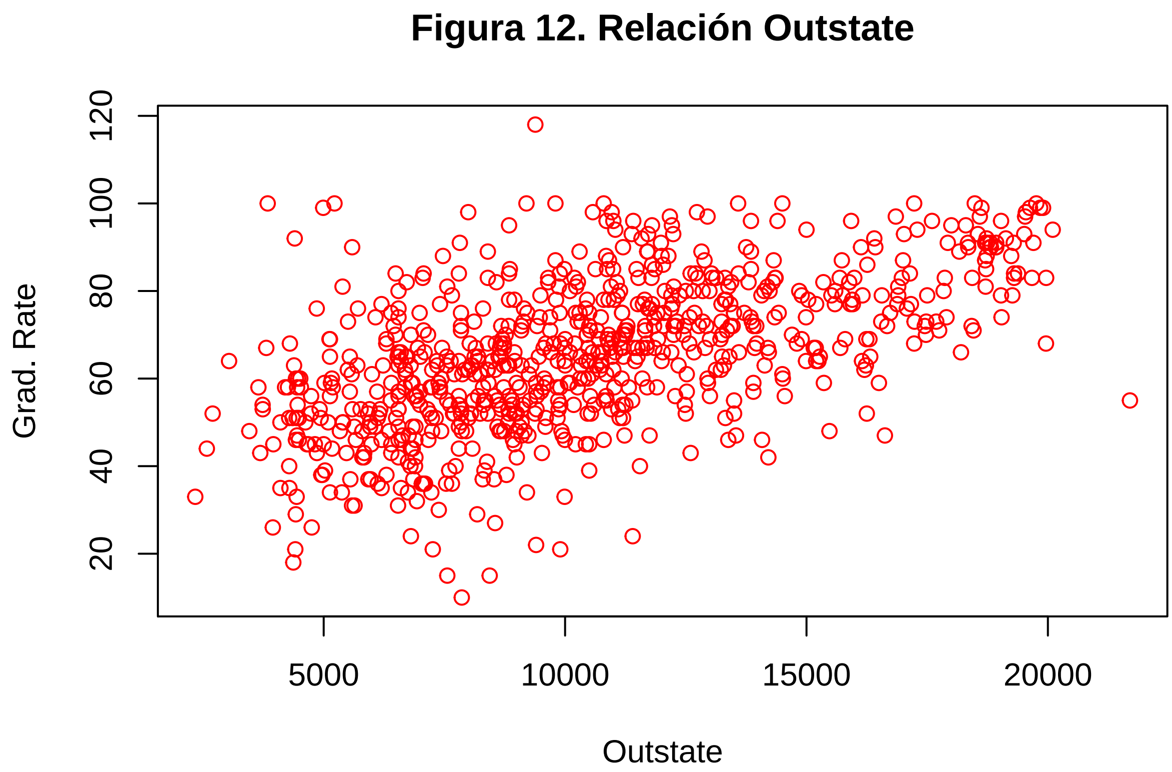
```r
poly.under.4$aic
```

```
## [1] 6159.572
```

El modelo mejora añadiendo la variable *Undergrad* ajustada con polinomios de orden 2

3.9 Esutido de la variable *Outstate*

```
plot(Outstate, Grad.Rate,
     xlab="Outstate",  ylab="Grad. Rate",
     main="Figura 12. Relación Outstate",
     col="red")
```



Ajuste con splines suavizados

```
fit_spl.out=smooth.spline(Outstate, Grad.Rate ,cv=TRUE)
fit_spl.out$df
```

```
## [1] 6.416494
```

```
spl.out.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
              s(Room.Board, 4) + Personal + poly(P.Undergrad ,2),
              data=college)

spl.out.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
              s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + Outstate ,
              data=college)

spl.out.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
              s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 1),
```

```r
             data=college)

spl.out.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
             s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 2),
             data=college)

spl.out.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
             s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 3),
             data=college)

spl.out.6=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
             s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 4),
             data=college)

spl.out.7=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
             s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5),
             data=college)

spl.out.8=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
             s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) +
             s(Outstate, round(fit_spl.out$df,3)),
             data=college)

anova(spl.out.1, spl.out.2, spl.out.3, spl.out.4, spl.out.5,
      spl.out.6, spl.out.7, spl.out.8, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + Outstate
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     3)
## Model 6: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     4)
## Model 7: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5)
## Model 8: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     round(fit_spl.out$df, 3))
##   Resid. Df Resid. Dev          Df Deviance      F    Pr(>F)
## 1    760.00     120397
## 2    759.00     119313  1.0000e+00  1083.66  7.1171 0.0077992 **
## 3    759.00     119313 -9.4302e-07     0.04
```

```
## 4    758.00     117323  9.9995e-01  1990.53 13.0738 0.0003195 ***
## 5    757.00     116316  1.0001e+00  1006.83  6.6119 0.0103188 *
## 6    756.00     115608  9.9987e-01   707.91  4.6499 0.0313761 *
## 7    755.00     115149  1.0004e+00   459.01  3.0134 0.0829703 .
## 8    753.58     114742  1.4155e+00   407.36  1.8901 0.1640045
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con polinomios de hasta orden 5

```r
poly.out.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + Outstate,
                data=college)

poly.out.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + poly(Outstate ,2),
                data=college)

poly.out.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + poly(Outstate ,3),
                data=college)

poly.out.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + poly(Outstate ,4),
                data=college)


poly.out.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + poly(Outstate ,5),
                data=college)

anova(poly.out.1, poly.out.2, poly.out.3, poly.out.4, poly.out.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + Outstate
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + poly(Outstate,
##     2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + poly(Outstate,
##     3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + poly(Outstate,
##     4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + poly(Outstate,
##     5)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       759     119313
## 2       758     117467  1  1846.59 0.0005498 ***
## 3       757     117404  1    63.04 0.5232097
## 4       756     116783  1   620.70 0.0451504 *
## 5       755     116778  1     5.09 0.8560241
```

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

```r
spl.out.2$aic
```

```
## [1] 6154.5
```

```r
spl.out.3$aic
```

```
## [1] 6154.5
```

```r
spl.out.4$aic
```

```
## [1] 6143.428
```

```r
spl.out.5$aic
```

```
## [1] 6138.731
```

```r
spl.out.6$aic
```

```
## [1] 6135.987
```

```r
spl.out.7$aic
```

```
## [1] 6134.897
```

```r
spl.out.8$aic
```

```
## [1] 6134.975
```

```r
poly.out.2$aic
```

```
## [1] 6144.381
```

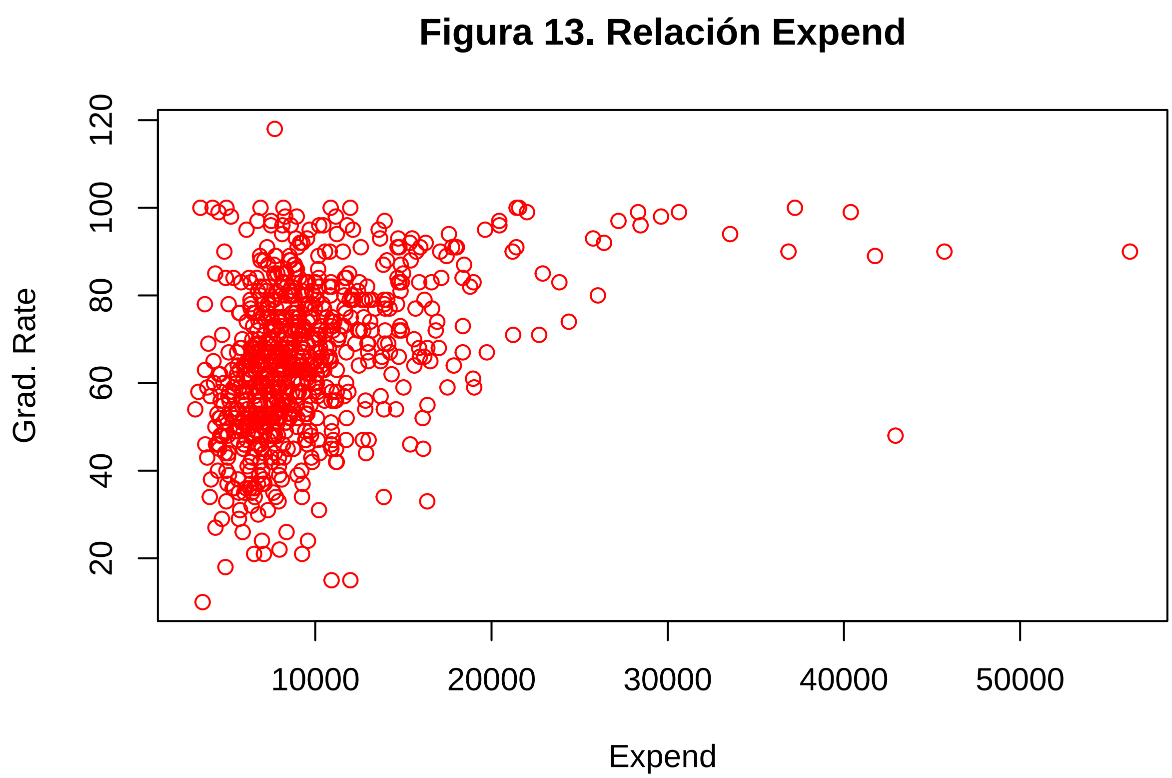
```r
poly.out.4$aic
```

```
## [1] 6143.845
```

Para la variable OutState el mejor ajuste se obtiene usando splines suavizados de grado de libertad igual a 5

3.10 Estudio de la variable *Expend*

```r
plot(Expend, Grad.Rate,
     xlab="Expend",  ylab="Grad. Rate",
     main="Figura 13. Relación Expend",
     col="red")
```



Ajuste con splines suavizados

```
fit_spl.expend=smooth.spline(Expend, Grad.Rate ,cv=TRUE)
fit_spl.expend$df
```

```
## [1] 4.385853
```

```
spl.expend.1=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5),
                 data=college)

spl.expend.2=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 Expend ,data=college)

spl.expend.3=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 s(Expend, 1), data=college)

spl.expend.4=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 s(Expend, 2), data=college)

spl.expend.5=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 s(Expend, 3), data=college)

spl.expend.6=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
```

```r
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 s(Expend, 4), data=college)

spl.expend.7=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 s(Expend, 5), data=college)

spl.expend.8=gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps,5) + Top25perc +
                 s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                 s(Expend, round(fit_spl.expend$df,3)), data=college)

anova(spl.expend.1, spl.expend.2, spl.expend.3, spl.expend.4, spl.expend.5,
      spl.expend.6, spl.expend.7, spl.expend.8, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5)
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + Expend
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + s(Expend, 1)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + s(Expend, 2)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + s(Expend, 3)
## Model 6: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + s(Expend, 4)
## Model 7: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + s(Expend, 5)
## Model 8: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + s(Expend, round(fit_spl.expend$df, 3))
##   Resid. Df Resid. Dev          Df Deviance       F    Pr(>F)
## 1    755.00     115149
## 2    754.00     114558  1.00000000   591.22  3.9431 0.0474272 *
## 3    754.00     114557  0.00025998     0.46 11.7177 0.0007683 ***
## 4    753.00     113063  0.99977590  1494.41  9.9692 0.0016569 **
## 5    752.00     112615  1.00005559   447.97  2.9876 0.0843154 .
## 6    751.00     112502  0.99986533   113.05  0.7541 0.3854363
## 7    750.00     112452  1.00029987    49.48  0.3299 0.5659698
## 8    750.61     112480 -0.61403742   -27.50  0.2987 0.4756803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ajuste con funciones polinómicas

```r
poly.expend.1= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) +
                    Top25perc + s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) +
                    s(Outstate, 5) + Expend, data=college)

poly.expend.2= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                    s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                    poly(Expend ,2) ,data=college)

poly.expend.3= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                    s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                    poly(Expend ,3) ,data=college)

poly.expend.4= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                    s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                    poly(Expend ,4) ,data=college)


poly.expend.5= gam(Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps ,5) + Top25perc +
                    s(Room.Board, 4) + Personal + poly(P.Undergrad ,2) + s(Outstate, 5) +
                    poly(Expend ,5) ,data=college)

anova(poly.expend.1, poly.expend.2, poly.expend.3, poly.expend.4, poly.expend.5)
```

```
## Analysis of Deviance Table
##
## Model 1: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + Expend
## Model 2: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + poly(Expend, 2)
## Model 3: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + poly(Expend, 3)
## Model 4: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + poly(Expend, 4)
## Model 5: Grad.Rate ~ Private + s(perc.alumni, 2) + poly(Apps, 5) + Top25perc +
##     s(Room.Board, 4) + Personal + poly(P.Undergrad, 2) + s(Outstate,
##     5) + poly(Expend, 5)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       754     114558
## 2       753     113048  1  1509.46 0.001512 **
## 3       752     112602  1   445.71 0.084743 .
## 4       751     112588  1    14.27 0.757726
## 5       750     112495  1    92.98 0.431083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
spl.expend.2$aic
```

```
## [1] 6132.897
```

```r
spl.expend.3$aic
```

```
## [1] 6132.895
```

```r
spl.expend.4$aic
```

```
## [1] 6124.692
```

```r
spl.expend.5$aic
```

```
## [1] 6123.607
```

```r
poly.expend.2$aic
```

```
## [1] 6124.591
```

```r
poly.expend.3$aic
```

```
## [1] 6123.522
```

Para la variable *Expend* la función que mejor ajusta es usando polinomios de orden 3.

```r
detach(college)
```

4. Modelo resultante

El modelo propuesto es el que hemos denominado como *poly.expend.3*, se trata de un modelo aditivo generalizado que incluye las siguientes variables:

gam(Grad.Rate ~ Private + Top25perc + Personal + s(perc.alumni, 2) + poly(Apps ,5) + s(Room.Board, 4) + poly(P.Undergrad ,2) + s(Outstate, 5) + poly(Expend ,3) , data=college)

- **Private**: la funcion que mejor ajusta esta variable es la lineal, polinomio de grado 1.
- **Top25perc**: la funcion que mejor ajusta esta variable es la lineal, polinomio de grado 1.
- **Personal**: la funcion que mejor ajusta esta variable es la lineal, polinomio de grado 1.
- **perc.alumni**: la funcion que mejor ajusta esta variable con splines suavizados con 2 grados de libertad
- **Room.Board**: la funcion que mejor ajusta esta variable con splines suavizados con 4 grados de libertad
- **Outstate**: la funcion que mejor ajusta esta variable con splines suavizados con 5 grados de libertad
- **P.Undergrad**: la funcion que mejor ajusta esta variable con splines suavizados con 2 grados de libertad
- **Expend**: la funcion que mejor ajusta esta variable con splines suavizados con 3 grados de libertad
- **Apps**: la funcion que mejor ajusta esta variable con splines suavizados con 5 grados de libertad

El AIC de este modelo es igual a **6123.5** , es un AIC alto, por tanto vemos que el modelo aditivo generalizado propuesto no es muy satisfactorio para este conjunto de datos.