

actividadintegradora2

Facundo Colasurdo Caldironi

2024-09-06

##Exploración de la base de datos

ME TOCA GRUPO 2 Segundo grupo. Altura del auto, ancho del auto y si es convertible o no. carheight
carwidth carbody

```
M=read.csv("file:///Users/facundocolasurdocaldironi/Downloads/precios_autos.csv") #leer la base de datos
head(M)
```

```
##      symboling          CarName fueltype      carbody drivewheel
## 1           3      alfa-romero giulia      gas convertible      rwd
## 2           3      alfa-romero stelvio      gas convertible      rwd
## 3           1 alfa-romero Quadrifoglio      gas  hatchback      rwd
## 4           2          audi 100 ls      gas      sedan      fwd
## 5           2          audi 100ls      gas      sedan      4wd
## 6           2          audi fox      gas      sedan      fwd
##      enginelocation wheelbase carlength carwidth carheight curbweight enginetype
## 1           front      88.6      168.8      64.1      48.8      2548      dohc
## 2           front      88.6      168.8      64.1      48.8      2548      dohc
## 3           front      94.5      171.2      65.5      52.4      2823      ohcv
## 4           front      99.8      176.6      66.2      54.3      2337      ohc
## 5           front      99.4      176.6      66.4      54.3      2824      ohc
## 6           front      99.8      177.3      66.3      53.1      2507      ohc
##      cylindernumber enginesize stroke compressionratio horsepower peakrpm citympg
## 1           four      130      2.68              9.0      111      5000      21
## 2           four      130      2.68              9.0      111      5000      21
## 3           six      152      3.47              9.0      154      5000      19
## 4           four      109      3.40             10.0      102      5500      24
## 5           five      136      3.40              8.0      115      5500      18
## 6           five      136      3.40              8.5      110      5500      19
##      highwaympg price
## 1           27 13495
## 2           27 16500
## 3           26 16500
## 4           30 13950
## 5           22 17450
## 6           25 15250
```

```
carbody_counts <- table(M$carbody)

# Mostrar la frecuencia de cada categoría
print(carbody_counts)
```

```
##
## convertible      hardtop    hatchback      sedan      wagon
##              6              8              70              96              25
```

```
# Verificar cuántos autos son convertibles
num_convertibles <- carbody_counts["convertible"]

# Mostrar el número de autos convertibles
print(paste("Número de autos convertibles:", num_convertibles))
```

```
## [1] "Número de autos convertibles: 6"
```

Calcula medidas estadísticas apropiadas para las variables: cuantitativas (media, desviación estándar, cuantiles, etc) cualitativas: cuantiles, frecuencias (puedes usar el comando table o prop.table) Summary general

```
summary(M)
```

```
##      symboling      CarName      fueltype      carbody
## Min.      :-2.0000 Length:205      Length:205      Length:205
## 1st Qu.: 0.0000   Class :character Class :character Class :character
## Median : 1.0000   Mode  :character Mode  :character Mode  :character
## Mean      : 0.8341
## 3rd Qu.: 2.0000
## Max.      : 3.0000
##      drivewheel      enginelocation      wheelbase      carlength
## Length:205      Length:205      Min.      : 86.60 Min.      :141.1
## Class :character Class :character 1st Qu.: 94.50 1st Qu.:166.3
## Mode  :character Mode  :character Median : 97.00 Median :173.2
##                                     Mean      : 98.76 Mean      :174.0
##                                     3rd Qu.:102.40 3rd Qu.:183.1
##                                     Max.      :120.90 Max.      :208.1
##      carwidth      carheight      curbweight      enginetype
## Min.      :60.30 Min.      :47.80 Min.      :1488 Length:205
## 1st Qu.:64.10 1st Qu.:52.00 1st Qu.:2145 Class :character
## Median :65.50 Median :54.10 Median :2414 Mode  :character
## Mean      :65.91 Mean      :53.72 Mean      :2556
## 3rd Qu.:66.90 3rd Qu.:55.50 3rd Qu.:2935
## Max.      :72.30 Max.      :59.80 Max.      :4066
##      cylindernumber      enginesize      stroke      compressionratio
## Length:205      Min.      : 61.0 Min.      :2.070 Min.      : 7.00
## Class :character 1st Qu.: 97.0 1st Qu.:3.110 1st Qu.: 8.60
## Mode  :character Median :120.0 Median :3.290 Median : 9.00
##                                     Mean      :126.9 Mean      :3.255 Mean      :10.14
##                                     3rd Qu.:141.0 3rd Qu.:3.410 3rd Qu.: 9.40
##                                     Max.      :326.0 Max.      :4.170 Max.      :23.00
##      horsepower      peakrpm      citympg      highwaympg      price
## Min.      : 48.0 Min.      :4150 Min.      :13.00 Min.      :16.00 Min.      : 5118
## 1st Qu.: 70.0 1st Qu.:4800 1st Qu.:19.00 1st Qu.:25.00 1st Qu.: 7788
## Median : 95.0 Median :5200 Median :24.00 Median :30.00 Median :10295
## Mean      :104.1 Mean      :5125 Mean      :25.22 Mean      :30.75 Mean      :13277
## 3rd Qu.:116.0 3rd Qu.:5500 3rd Qu.:30.00 3rd Qu.:34.00 3rd Qu.:16503
## Max.      :288.0 Max.      :6600 Max.      :49.00 Max.      :54.00 Max.      :45400
```

carheight carwidth carbody (convertible)

```
summary(M$carheight)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    47.80  52.00   54.10   53.72  55.50   59.80
```

```
summary(M$carwidth)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    60.30  64.10   65.50   65.91  66.90   72.30
```

```
summary(M$carbody)
```

```
##      Length      Class    Mode
##         205 character character
```

```
table_carbody= table(M$carbody)
prop_table_carbody = prop.table(table_carbody)
table_carbody
```

```
##
## convertible      hardtop  hatchback      sedan      wagon
##           6           8           70           96           25
```

```
prop_table_carbody
```

```
##
## convertible      hardtop  hatchback      sedan      wagon
## 0.02926829 0.03902439 0.34146341 0.46829268 0.12195122
```

Analiza la correlación entre las variables (analiza posible colinealidad entre las variables)

```
cor(M[, c("carheight", "carwidth", "price")], use = "complete.obs")
```

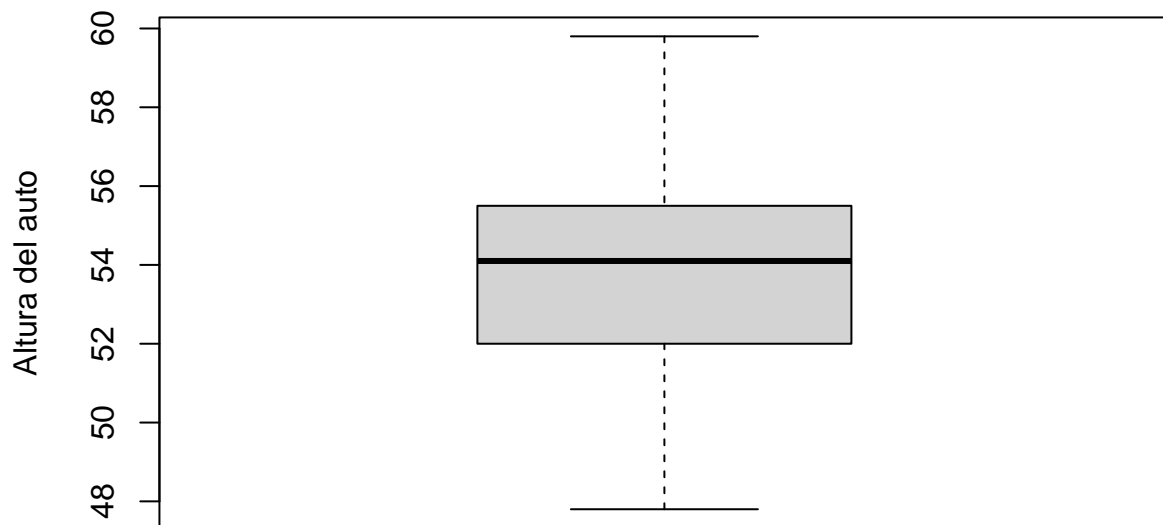
```
##           carheight carwidth    price
## carheight 1.0000000 0.2792103 0.1193362
## carwidth  0.2792103 1.0000000 0.7593253
## price     0.1193362 0.7593253 1.0000000
```

Nos indica una relacion positiva baja entre la altura y el ancho del vehiculo, es decir, usualmente si aumenta mientras el otro incrementa, mas no es fuerte.

Explora los datos usando herramientas de visualización (si lo consideras necesario): Variables cuantitativas: Boxplot (visualización de datos atípicos) Histogramas Diagramas de dispersión y correlación por pares

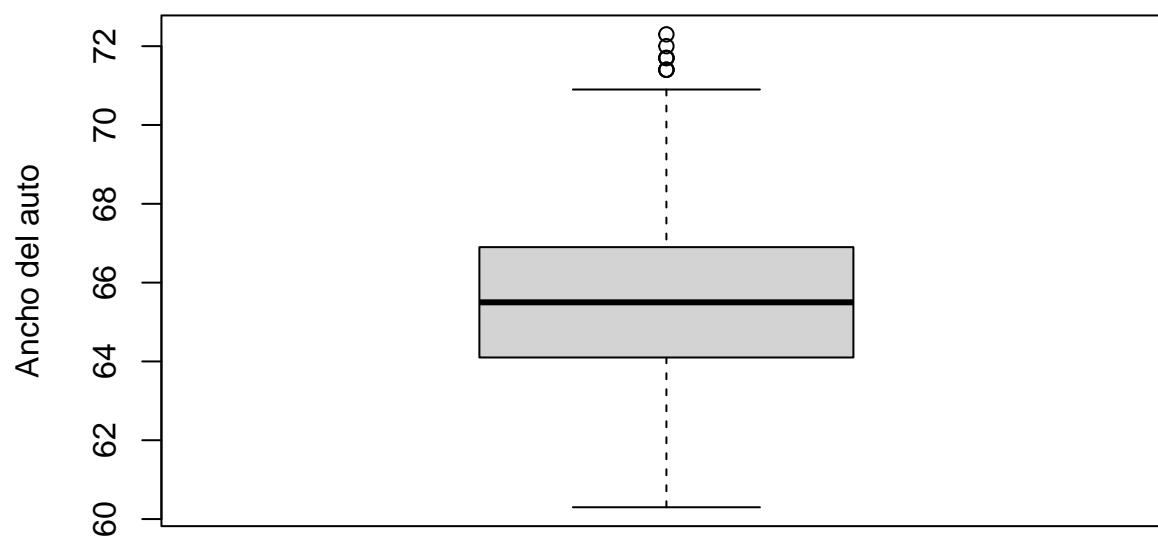
```
boxplot(M$carheight, main="Boxplot de carheight", ylab="Altura del auto")
```

Boxplot de carheight

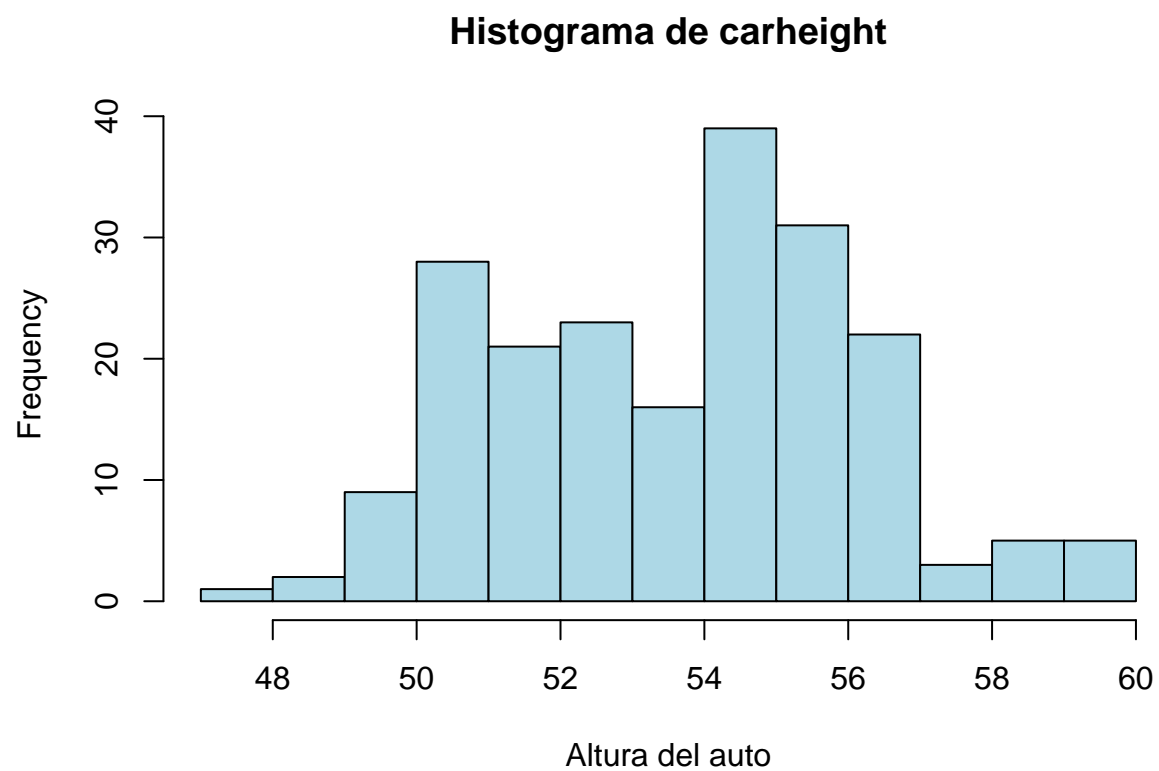


```
boxplot(M$carwidth, main="Boxplot de carwidth", ylab="Ancho del auto")
```

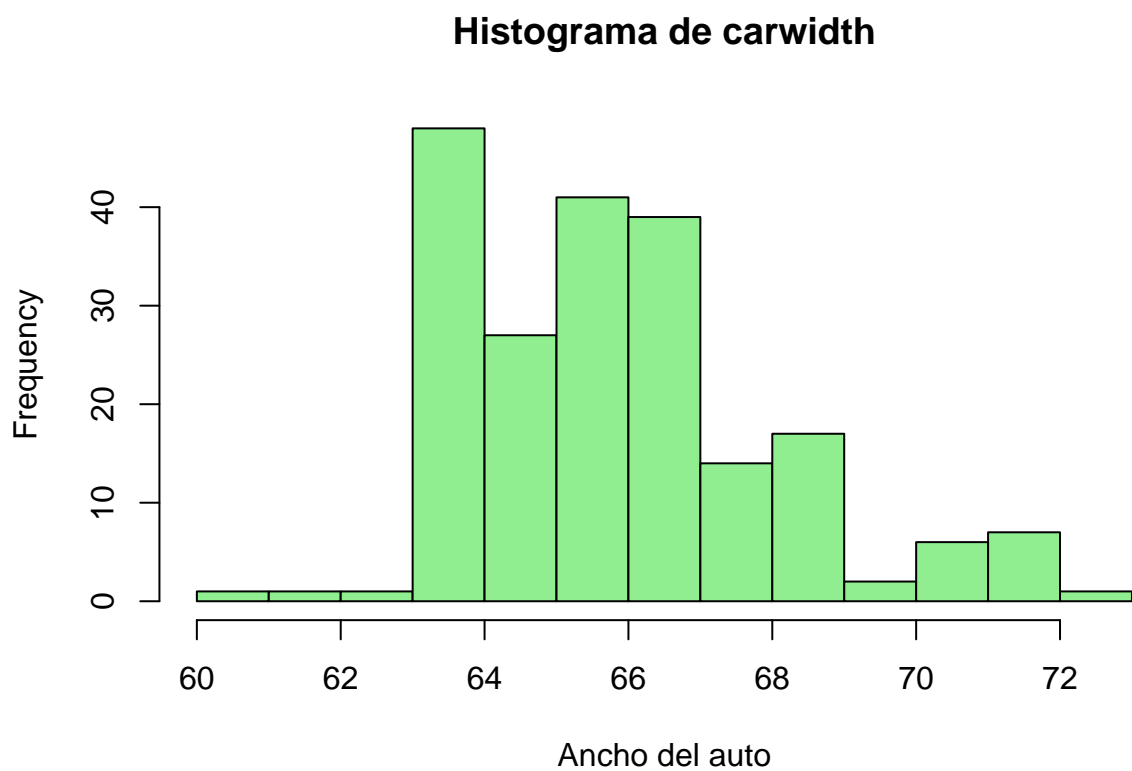
Boxplot de carwidth



```
hist(M$carheight, main="Histograma de carheight", xlab="Altura del auto", col="lightblue", breaks=10)
```

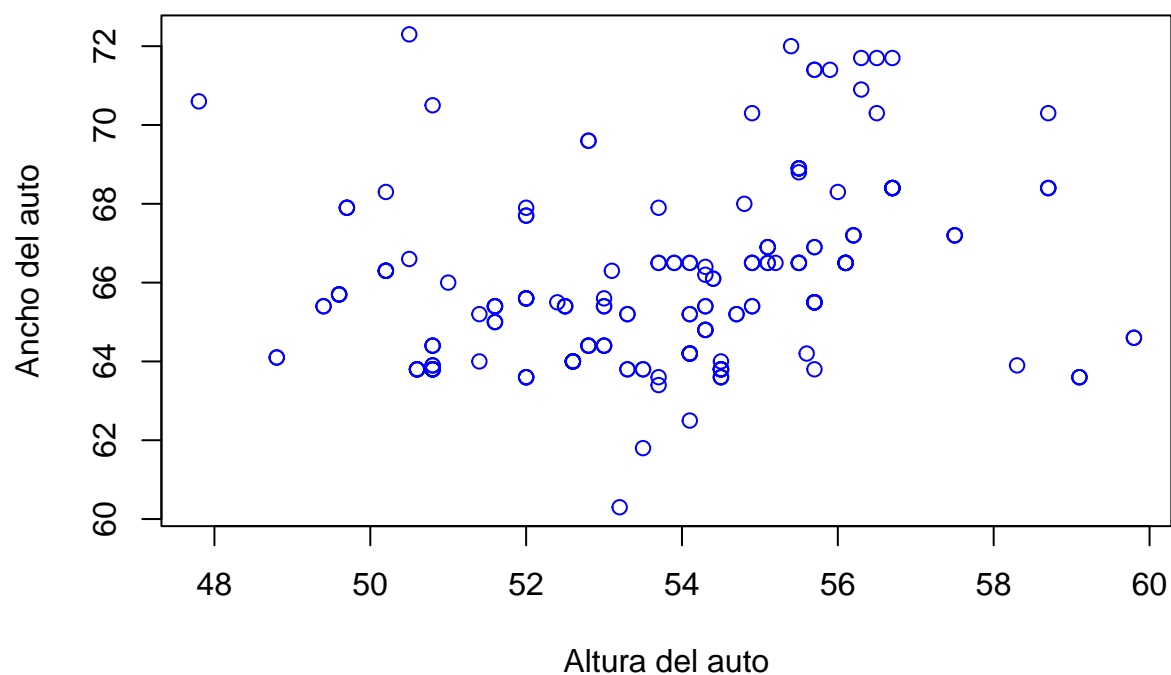


```
hist(M$carwidth, main="Histograma de carwidth", xlab="Ancho del auto", col="lightgreen", breaks=10)
```



```
plot(M$carheight, M$carwidth, main="Diagrama de dispersión: Altura vs Ancho", xlab="Altura del auto", ylab="Frecuencia",
```

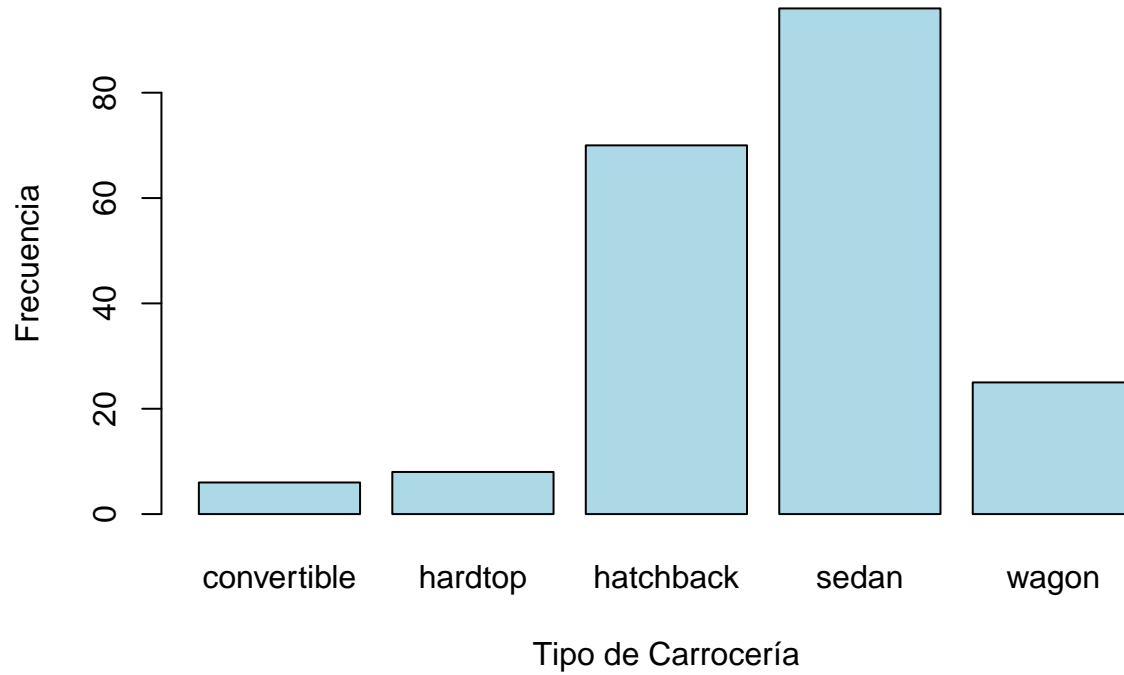
Diagrama de dispersión: Altura vs Ancho



Variables categóricas Distribución de los datos (diagramas de barras, diagramas de pastel) Boxplot por categoría de las variables cuantitativas

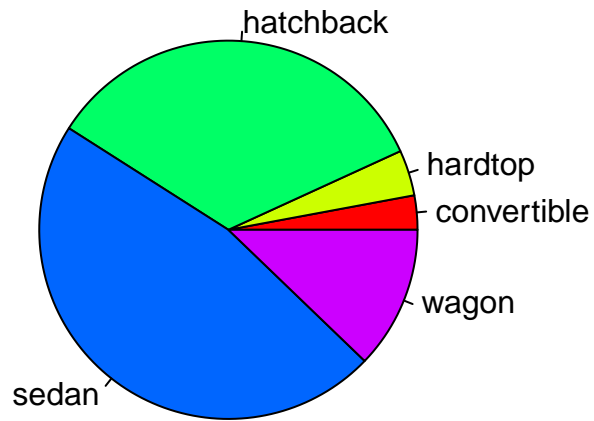
```
barplot(table(M$carbody), main = "Distribución de Carrocerías", col = "lightblue", ylab = "Frecuencia",
```


Distribución de Carrocerías

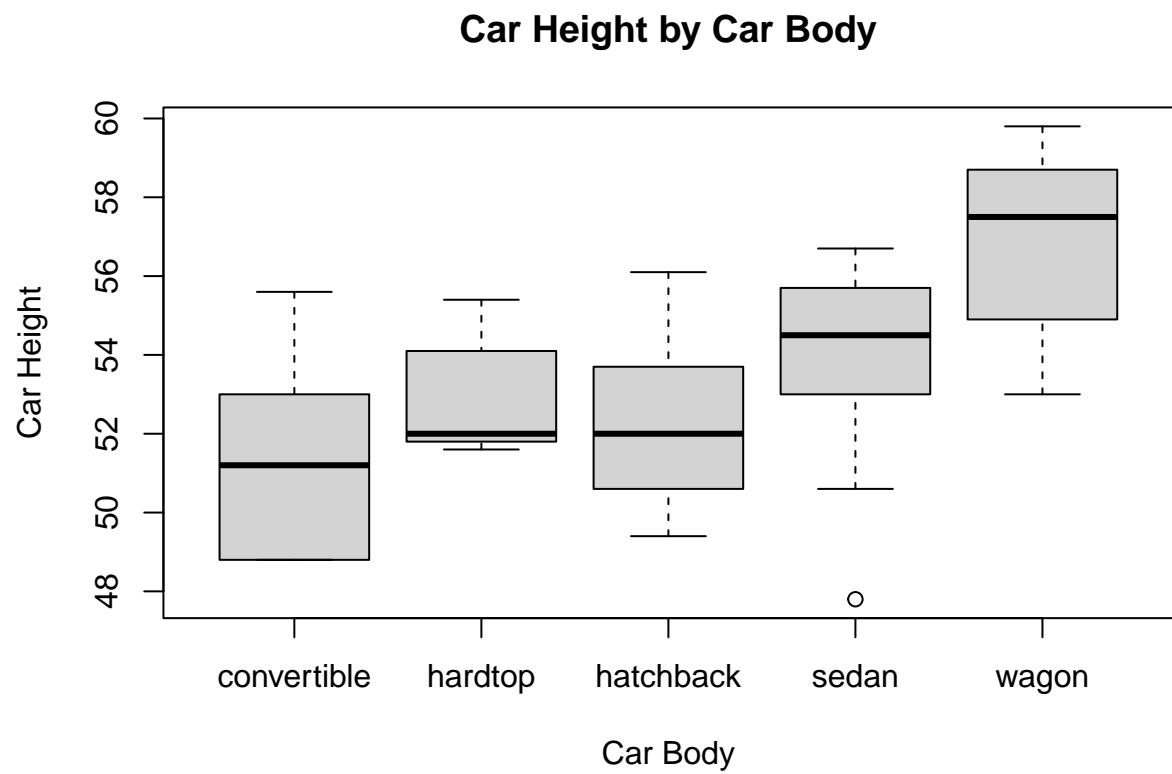


```
pie(table(M$carbody), main = "Distribución de Carrocerías", col = rainbow(length(table(M$carbody))))
```

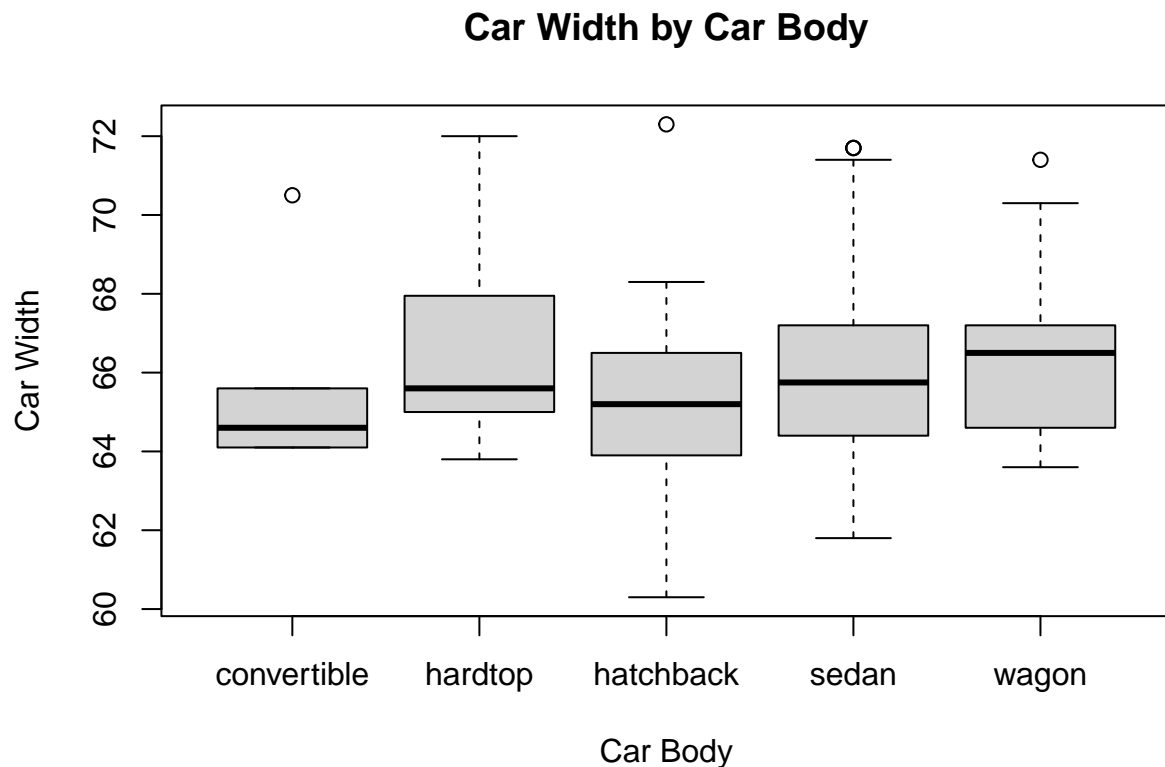
Distribución de Carrocerías



```
boxplot(M$carheight ~ M$carbody, data = M, main="Car Height by Car Body", xlab="Car Body", ylab="Car Height")
```



```
boxplot(M$carwidth ~ M$carbody, data = M, main="Car Width by Car Body", xlab="Car Body", ylab="Car Width")
```



Modelación y verificación del modelo Encuentra la ecuación de regresión de mejor ajuste. Propón al menos 2 modelos de ajuste para encontrar la mejor forma de ajustar la variable precio.

Modelo simple modelo de regresion multiple

```
model_simple = lm(price ~ carheight + carwidth + carbody, data = M)
summary(model_simple)
```

```
##
## Call:
## lm(formula = price ~ carheight + carwidth + carbody, data = M)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-11103.9	-2404.6	-657.1	1430.6	22217.3

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-150934.9	12535.9	-12.040	< 2e-16 ***
carheight	-225.0	177.7	-1.266	0.207
carwidth	2811.7	161.7	17.388	< 2e-16 ***
carbodyhardtop	-2256.9	2554.2	-0.884	0.378
carbodyhatchback	-10416.6	2005.7	-5.194	5.10e-07 ***
carbodysedan	-8796.5	2042.1	-4.307	2.60e-05 ***
carbodywagon	-10218.4	2328.8	-4.388	1.86e-05 ***

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

```
##
## Residual standard error: 4703 on 198 degrees of freedom
## Multiple R-squared:  0.6636, Adjusted R-squared:  0.6534
## F-statistic: 65.1 on 6 and 198 DF,  p-value: < 2.2e-16
```

```
model_multiple = lm(price ~ carheight * carwidth + carbody, data = M)
summary(model_multiple)
```

```
##
## Call:
## lm(formula = price ~ carheight * carwidth + carbody, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11220.7  -2458.1   -563.4   1382.6  22008.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -436072.67   221003.36  -1.973   0.0499 *
## carheight       5043.33    4080.68    1.236   0.2180
## carwidth       7117.82    3336.14    2.134   0.0341 *
## carbodyhardtop  -2157.81    2551.05   -0.846   0.3987
## carbodyhatchback -10424.67    2002.28  -5.206 4.82e-07 ***
## carbodysedan    -8774.32    2038.77  -4.304 2.64e-05 ***
## carbodywagon   -10319.42    2326.15  -4.436 1.52e-05 ***
## carheight:carwidth   -79.53      61.54  -1.292   0.1978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4695 on 197 degrees of freedom
## Multiple R-squared:  0.6664, Adjusted R-squared:  0.6546
## F-statistic: 56.23 on 7 and 197 DF,  p-value: < 2.2e-16
```

Para cada uno de los modelos propuestos: Realiza la regresión entre las variables involucradas

Analiza la significancia del modelo: Valida la significancia del modelo con un alfa de 0.04 (incluye las hipótesis que pruebas y el valor frontera)

Hipotesis Modelo simple $H_0 : u = 0$ $H_1 : u \neq 0$

Modelo multiiple $H_0 : u = 0$ $H_1 : u \neq 0$ alpha = 0.04

```
summary(model_simple)
```

```
##
## Call:
## lm(formula = price ~ carheight + carwidth + carbody, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11103.9  -2404.6   -657.1   1430.6  22217.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      -150934.9    12535.9 -12.040 < 2e-16 ***
## carheight        -225.0      177.7  -1.266   0.207
## carwidth         2811.7      161.7  17.388 < 2e-16 ***
## carbodyhardtop   -2256.9     2554.2  -0.884   0.378
## carbodyhatchback -10416.6     2005.7  -5.194 5.10e-07 ***
## carbodysedan     -8796.5     2042.1  -4.307 2.60e-05 ***
## carbodywagon     -10218.4     2328.8  -4.388 1.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4703 on 198 degrees of freedom
## Multiple R-squared:  0.6636, Adjusted R-squared:  0.6534
## F-statistic: 65.1 on 6 and 198 DF, p-value: < 2.2e-16
```

```
anova(model_simple)
```

```
## Analysis of Variance Table
##
## Response: price
##           Df      Sum Sq    Mean Sq  F value    Pr(>F)
## carheight   1 185414441 185414441   8.3827 0.004214 **
## carwidth    1 7442659231 7442659231 336.4856 < 2.2e-16 ***
## carbody     4 1012041757 253010439 11.4387 2.256e-08 ***
## Residuals 198 4379523933 22118808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Indica cuál es el porcentaje de variación explicada por el modelo. Porcentaje de variacion explicada: 66.36%

Valida la significancia de Bi con un alfa de 0.04 (incluye las hipótesis que pruebas y el valor frontera de cada una de ellas)

```
summary(model_multiple)
```

```
##
## Call:
## lm(formula = price ~ carheight * carwidth + carbody, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11220.7  -2458.1   -563.4   1382.6  22008.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -436072.67  221003.36  -1.973   0.0499 *
## carheight      5043.33   4080.68    1.236   0.2180
## carwidth       7117.82   3336.14    2.134   0.0341 *
## carbodyhardtop -2157.81   2551.05   -0.846   0.3987
## carbodyhatchback -10424.67  2002.28  -5.206 4.82e-07 ***
## carbodysedan   -8774.32   2038.77  -4.304 2.64e-05 ***
## carbodywagon  -10319.42   2326.15  -4.436 1.52e-05 ***
## carheight:carwidth    -79.53     61.54   -1.292   0.1978
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4695 on 197 degrees of freedom
## Multiple R-squared:  0.6664, Adjusted R-squared:  0.6546
## F-statistic: 56.23 on 7 and 197 DF,  p-value: < 2.2e-16
```

```
anova(model_multiple)
```

```
## Analysis of Variance Table
##
## Response: price
##
##           Df      Sum Sq   Mean Sq F value    Pr(>F)
## carheight    1 185414441 185414441   8.411 0.004153 **
## carwidth     1 7442659231 7442659231 337.624 < 2.2e-16 ***
## carbody      4 1012041757 253010439  11.477 2.144e-08 ***
## carheight:carwidth 1   36813013   36813013   1.670 0.197777
## Residuals   197 4342710920   22044218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Indica cuál es el porcentaje de variación explicada por el modelo. Porcentaje de variación explicada: 66.64%

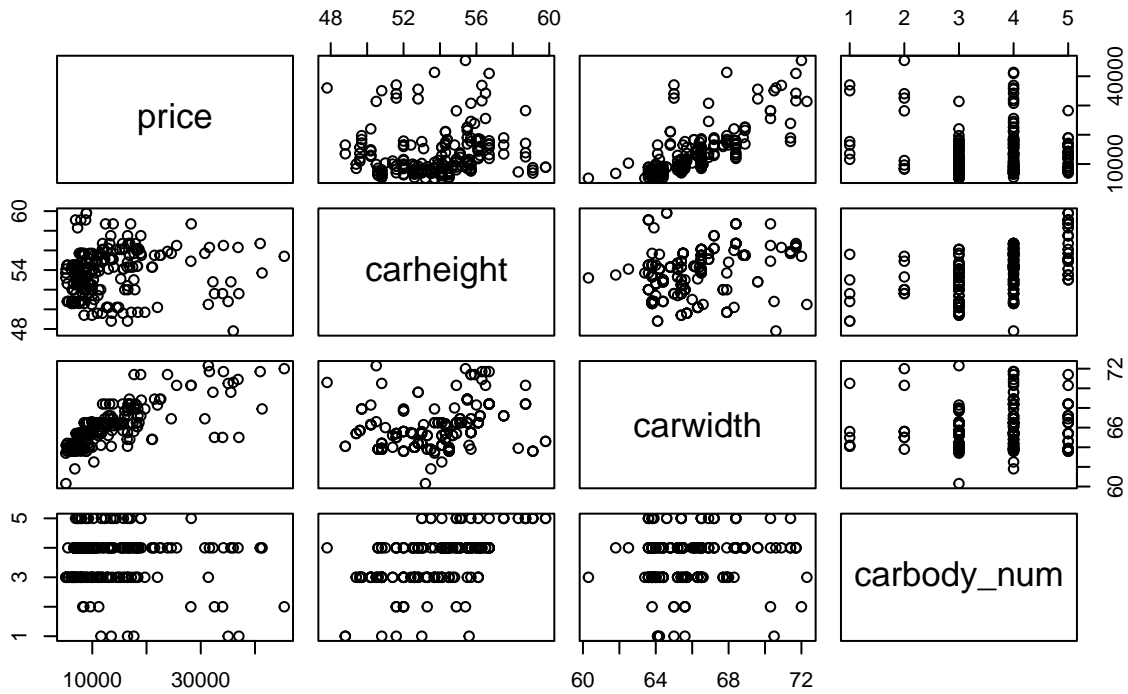
Dibuja el diagrama de dispersión de los datos por pares y la recta de mejor ajuste.

```
ModeloNuevo <- lm(price ~ carheight * carwidth + carbody, data = M)

M$carbody_num <- as.numeric(factor(M$carbody)) # Convertir 'carbody' a numérico para graficar

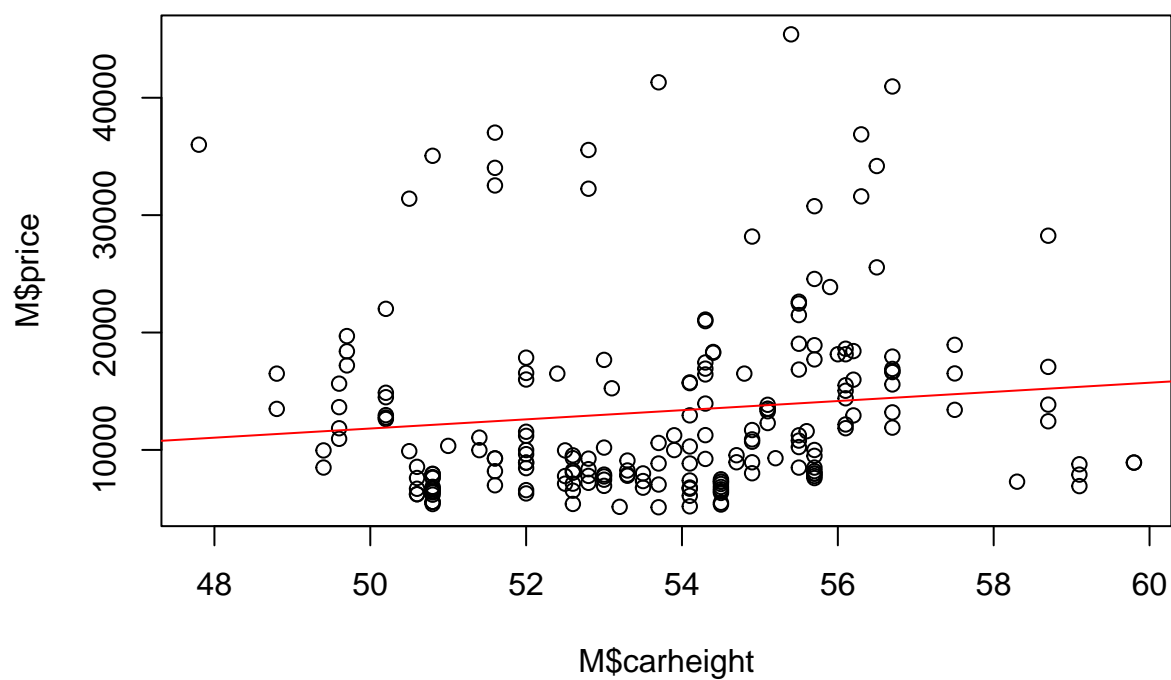
pairs(M[,c("price", "carheight", "carwidth", "carbody_num")],
      main = "Matriz de dispersión de Price, CarHeight, CarWidth y CarBody")
```

Matriz de dispersión de Price, CarHeight, CarWidth y CarBody



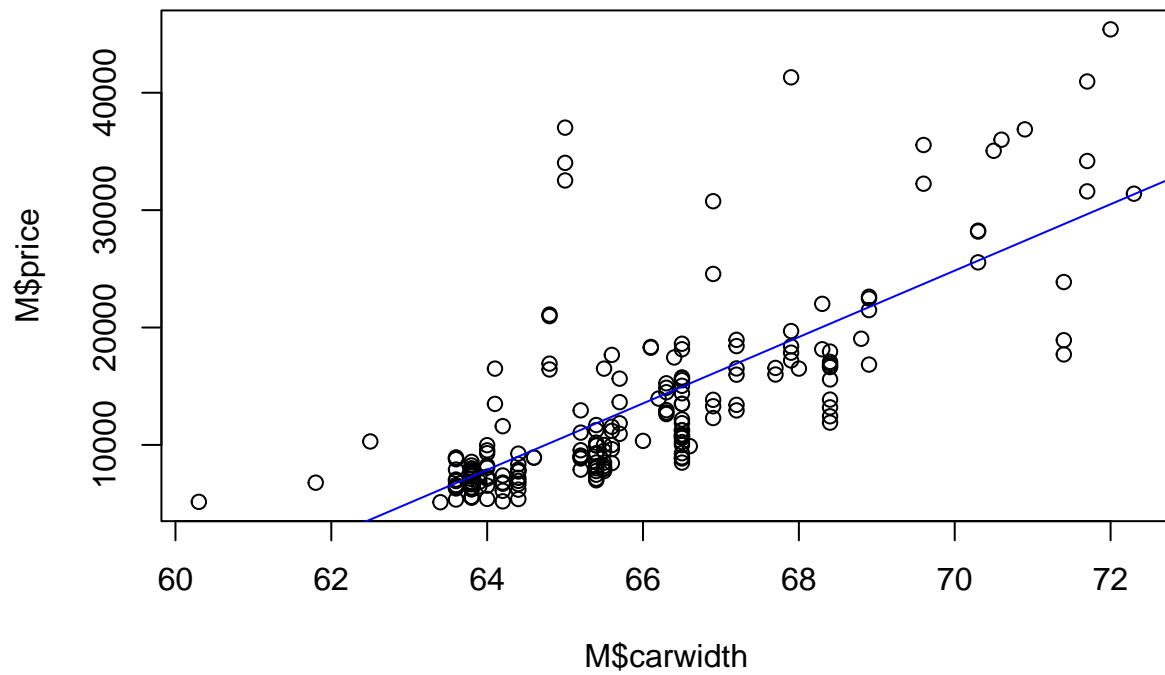
```
plot(M$carheight, M$price, main="Dispersión de Price vs CarHeight")
abline(lm(price ~ carheight, data = M), col="red") # Ajuste de línea para carheight
```


Dispersión de Price vs CarHeight



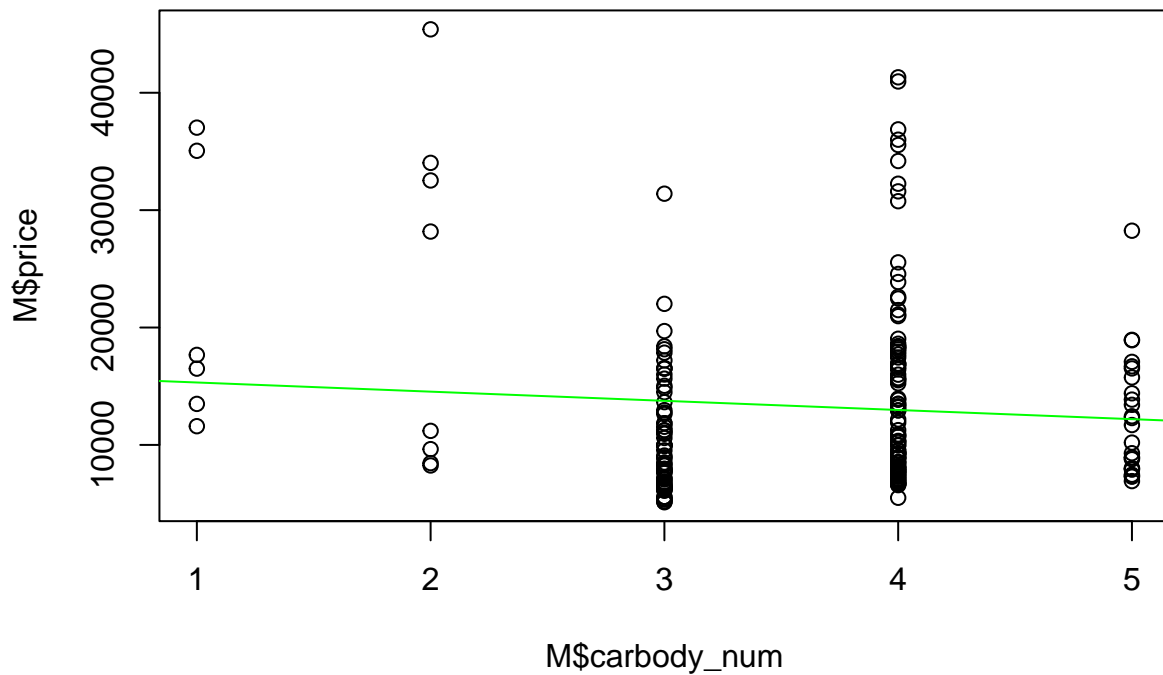
```
plot(M$carwidth, M$price, main="Dispersión de Price vs CarWidth")  
abline(lm(price ~ carwidth, data = M), col="blue") # Ajuste de línea para carwidth
```

Dispersión de Price vs CarWidth



```
plot(M$carbody_num, M$price, main="Dispersión de Price vs CarBody")  
abline(lm(price ~ carbody_num, data = M), col="green") # Ajuste de línea para carbody_num
```

Dispersión de Price vs CarBody



Interpreta en el contexto del problema cada uno de los análisis que hiciste. Observando los dos modelos, se tomo la decision de usar el modelo multiple ya que tiene un mejor ajuste entre la interacciones de carwidht y precio, a su vez, que este tiene muchos mas datos pertinentes a comparacion del modelo multiple, lo que mejora su validacion a comparacion del modelo

Analiza la validez de los modelos propuestos: Normalidad de los residuos

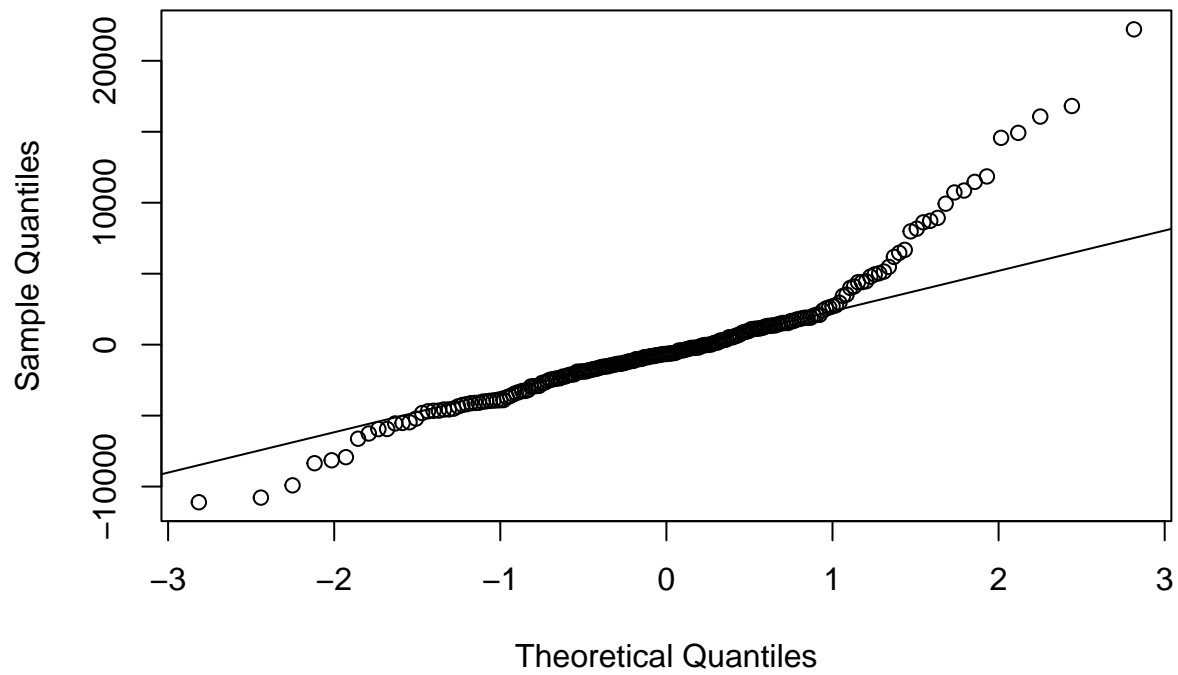
H0: Los datos provienen de una poblacion normal H1: los datos no provienen de una poblacion noormal

```
library(nortest)
ad.test(model_simple$residuals)
```

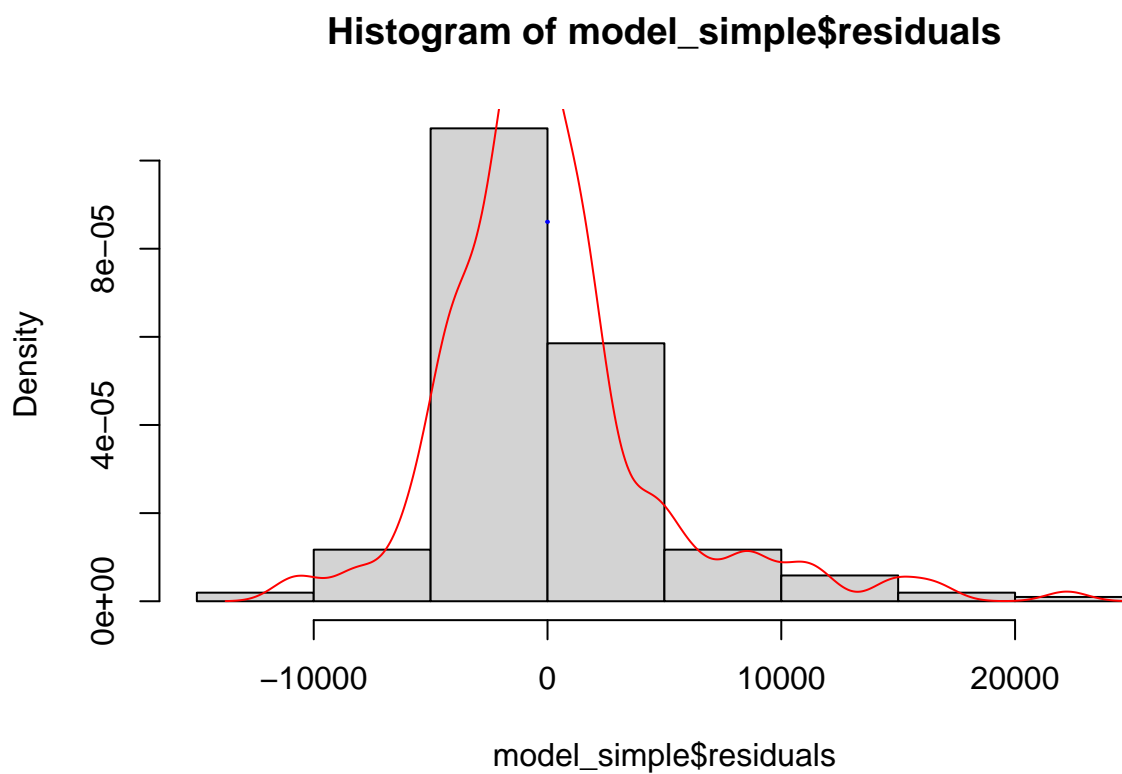
```
##
## Anderson-Darling normality test
##
## data: model_simple$residuals
## A = 6.3689, p-value = 1.103e-15
```

```
qqnorm(model_simple$residuals)
qqline(model_simple$residuals)
```

Normal Q-Q Plot



```
hist(model_simple$residuals,freq=FALSE)
lines(density(model_simple$residual),col="red")
curve(dnorm(x,mean=mean(model_simple$residuals),sd=sd(model_simple$residuals)), from=-20, to=20, add=TRUE, col="blue",lwd=2)
```

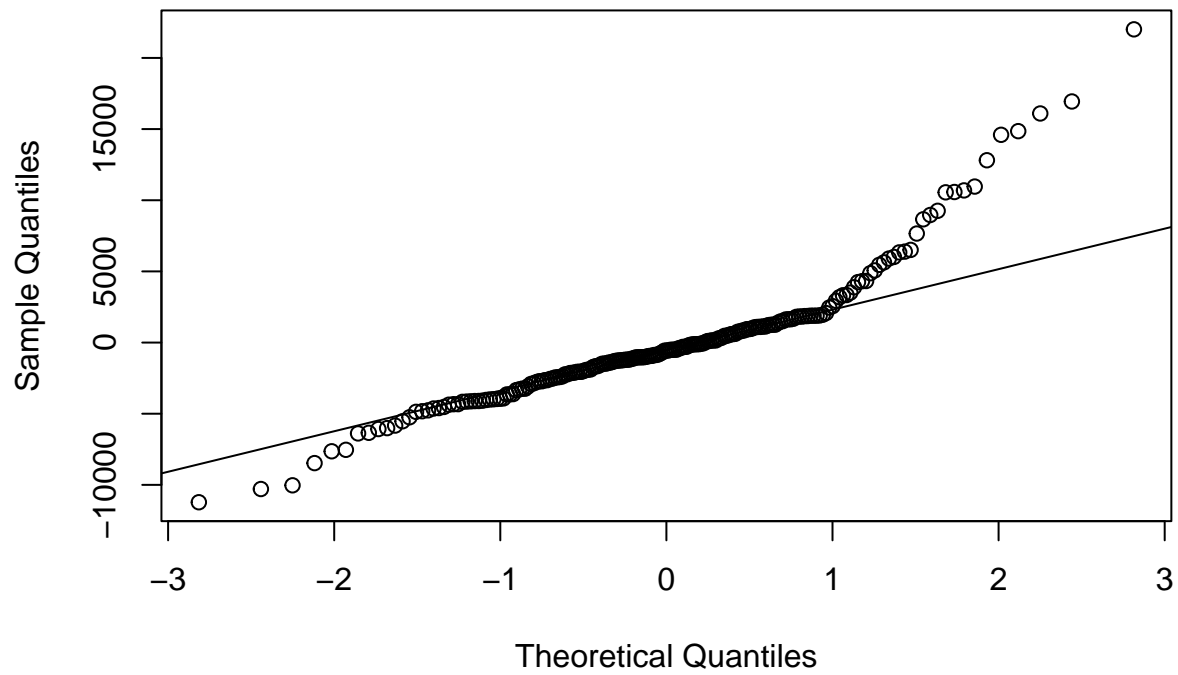


```
library(nortest)
ad.test(model_multiple$residuals)
```

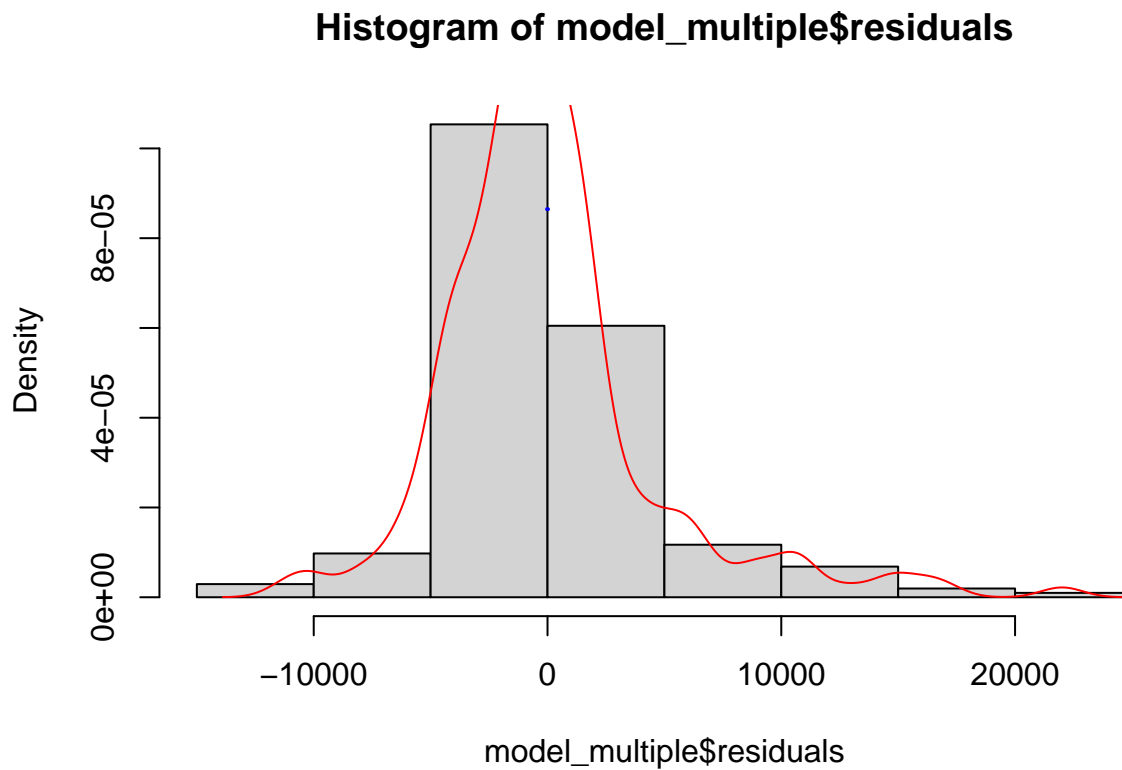
```
##
##  Anderson-Darling normality test
##
## data:  model_multiple$residuals
## A = 6.5149, p-value = 4.947e-16
```

```
qqnorm(model_multiple$residuals)
qqline(model_multiple$residuals)
```

Normal Q-Q Plot



```
hist(model_multiple$residuals,freq=FALSE)
lines(density(model_multiple$residual),col="red")
curve(dnorm(x,mean=mean(model_multiple$residuals),sd=sd(model_multiple$residuals)), from=-20, to=20, add=TRUE, col="blue",lwd=2)
```



En el primer modelo H_0 se rechaza ya que el valor p es menor que 0.04

En el segundo modelo H_0 no se rechaza ya que el valor p es mayor que 0.04

Verificación de media cero $H_0 : u = 0$ $H_1 : u \neq 0$

```
t.test(model_simple$residuals)
```

```
##
## One Sample t-test
##
## data: model_simple$residuals
## t = 1.1555e-16, df = 204, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -638.0485 638.0485
## sample estimates:
## mean of x
## 3.739394e-14
```

```
t.test(model_multiple$residuals)
```

```
##
## One Sample t-test
##
## data: model_multiple$residuals
```

```
## t = 2.2556e-15, df = 204, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -635.3612 635.3612
## sample estimates:
## mean of x
## 7.268668e-13
```

Homocedasticidad, linealidad e independencia para la homocedasticidad H0: La varianza de los errores es constante (homocedasticidad) H1: La varianza de los errores no es constante (heterocedasticidad)

para la linealidad H0: No hay términos omitidos que indicate linealidad H1: Hay una especificación errónea en el modelo que indica no linealidad

para la independencia H0: Los errores no están correlacionados H1: Los errores están correlacionados

```
library(lmtest)
```

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

```
##
```

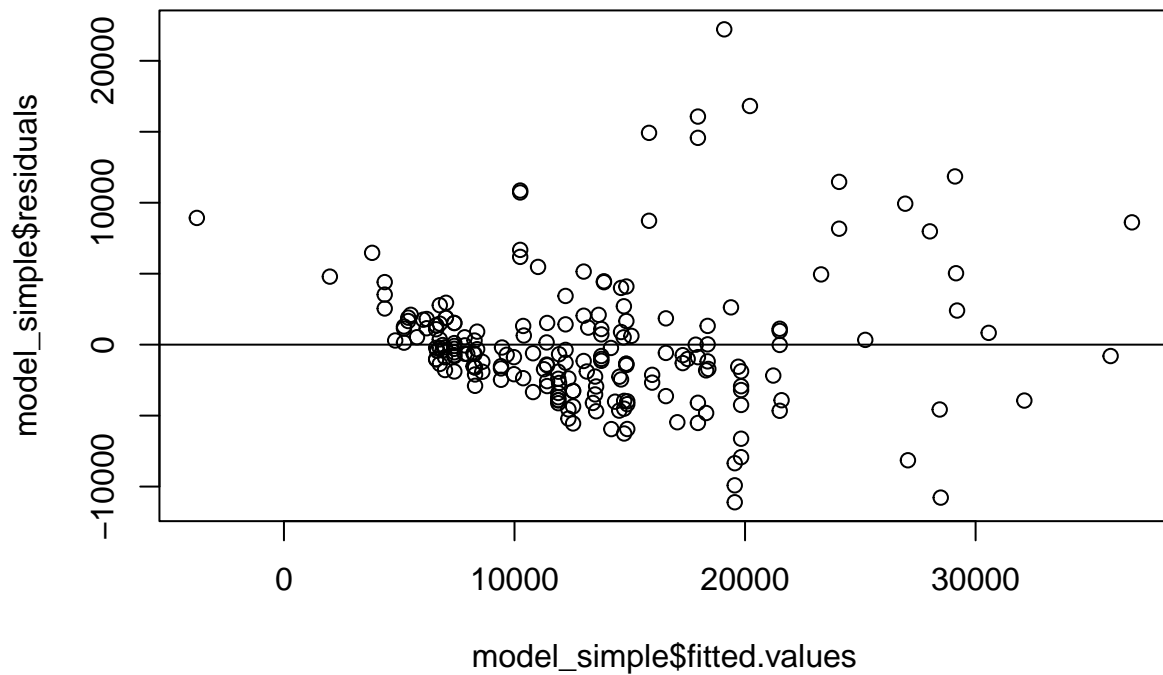
```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
plot(model_simple$fitted.values,model_simple$residuals)
abline(h=0)
```

```
resettest(model_simple)
```

```
##
## RESET test
##
## data: model_simple
## RESET = 6.8362, df1 = 2, df2 = 196, p-value = 0.001349
```

```
bptest(model_simple)
```

```
##
## studentized Breusch-Pagan test
##
## data: model_simple
## BP = 37.966, df = 6, p-value = 1.141e-06
```

```
gqtest(model_simple)
```

```
##
## Goldfeld-Quandt test
##
## data: model_simple
## GQ = 0.67139, df1 = 96, df2 = 95, p-value = 0.9736
## alternative hypothesis: variance increases from segment 1 to 2
```

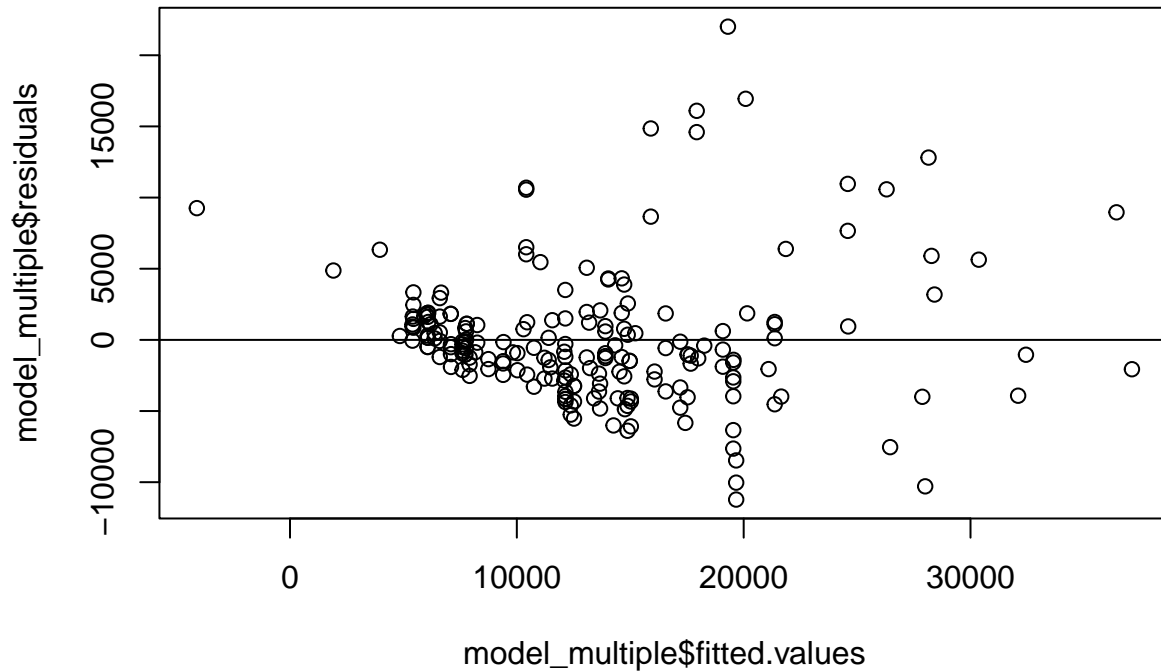
```
dwtest(model_simple)
```

```
##  
## Durbin-Watson test  
##  
## data: model_simple  
## DW = 0.76974, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(model_simple)
```

```
##  
## Breusch-Godfrey test for serial correlation of order up to 1  
##  
## data: model_simple  
## LM test = 80.703, df = 1, p-value < 2.2e-16
```

```
plot(model_multiple$fitted.values,model_multiple$residuals)  
abline(h=0)
```



```
resettest(model_multiple)
```

```
##
```

```
## RESET test
##
## data: model_multiple
## RESET = 6.9666, df1 = 2, df2 = 195, p-value = 0.001196
```

```
bptest(model_multiple)
```

```
##
## studentized Breusch-Pagan test
##
## data: model_multiple
## BP = 39.458, df = 7, p-value = 1.598e-06
```

```
gqtest(model_multiple)
```

```
##
## Goldfeld-Quandt test
##
## data: model_multiple
## GQ = 0.67423, df1 = 95, df2 = 94, p-value = 0.9717
## alternative hypothesis: variance increases from segment 1 to 2
```

```
dwtest(model_multiple)
```

```
##
## Durbin-Watson test
##
## data: model_multiple
## DW = 0.8073, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(model_multiple)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: model_multiple
## LM test = 79.076, df = 1, p-value < 2.2e-16
```

Interpreta cada uno de los análisis que realizaste. Para el modelo simple: Homocedasticidad H_0 se rechaza porque el valor p es menor 0.04. Independencia H_0 se rechaza porque el valor p es menor 0.04. Linealidad H_0 se rechaza porque el valor p es menor 0.04.

Modelo: Homocedasticidad H_0 se rechaza porque el valor p es menor 0.04. Independencia H_0 se rechaza porque el valor p es menor 0.04. Linealidad H_0 se rechaza porque el valor p es menor 0.04.

Emite una conclusión final sobre el mejor modelo de regresión lineal y contesta la pregunta central: Concluye sobre el mejor modelo que encuentres y argumenta por qué es el mejor?

De los dos modelos anteriores, el mejor modelo disponible es el modelo de regresión múltiple, ya que ayuda a obtener una mayor cantidad de características físicas, como la anchura, la altura y el tipo de carrocería, influyen en su precio.

¿Cuáles de las variables asignadas influyen en el precio del auto? ¿de qué manera lo hacen? La variable que mas influye en el precio del auto es carwidth, ya que muestra una relacion positiva entre el carwidth y el precio del mismo, mientras mas ancho sea un auto, mas caro sera.

##Intervalos de predicción y confianza Con los datos de las variables asignadas construye la gráfica de los intervalos de confianza y predicción para la estimación y predicción del precio para el mejor modelo seleccionado: carheight carwidth carbody

```
library(ggplot2)
Ip <- predict(object = model_multiple, interval = "prediction", level = 0.97)
```

```
## Warning in predict.lm(object = model_multiple, interval = "prediction", : predictions on current data
```

```
M2 <- cbind(M, Ip)
```

```
M_convertible <- subset(M, carbody == "convertible")
M_non_convertible <- subset(M, carbody != "convertible")
```

```
summary(M_convertible[, c("carheight", "carwidth", "price")])
```

```
##      carheight      carwidth      price
## Min.   :48.80   Min.   :64.10   Min.   :11595
## 1st Qu.:49.30   1st Qu.:64.12   1st Qu.:14246
## Median :51.20   Median :64.60   Median :17084
## Mean   :51.43   Mean   :65.58   Mean   :21890
## 3rd Qu.:52.65   3rd Qu.:65.45   3rd Qu.:30709
## Max.   :55.60   Max.   :70.50   Max.   :37028
```

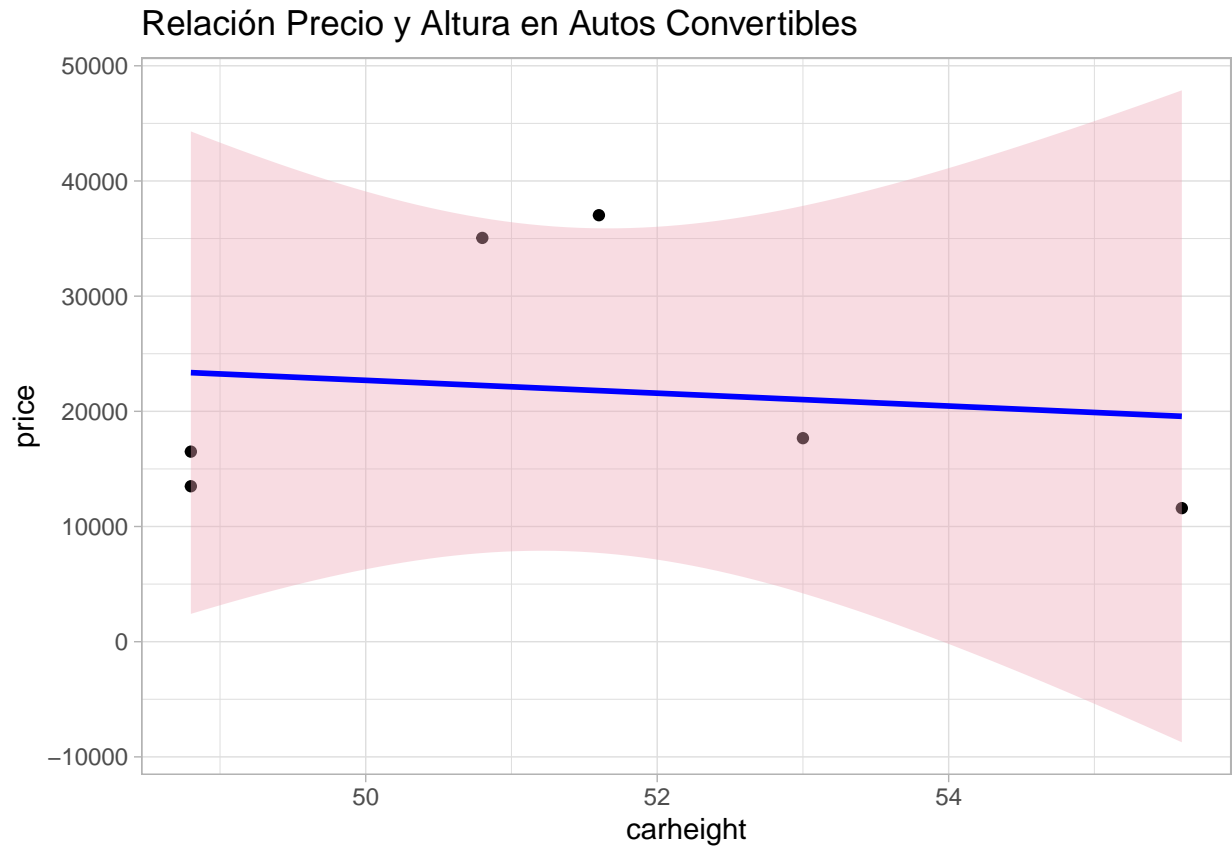
```
summary(M_non_convertible[, c("carheight", "carwidth", "price")])
```

```
##      carheight      carwidth      price
## Min.   :47.80   Min.   :60.30   Min.   : 5118
## 1st Qu.:52.00   1st Qu.:64.00   1st Qu.: 7775
## Median :54.10   Median :65.50   Median : 9995
## Mean   :53.79   Mean   :65.92   Mean   :13017
## 3rd Qu.:55.50   3rd Qu.:66.90   3rd Qu.:16465
## Max.   :59.80   Max.   :72.30   Max.   :45400
```

```
# Gráfico para Convertibles
```

```
ggplot(M_convertible, aes(x = carheight, y = price)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue", fill = "pink2") +
  ggtitle("Relación Precio y Altura en Autos Convertibles") +
  theme_light()
```

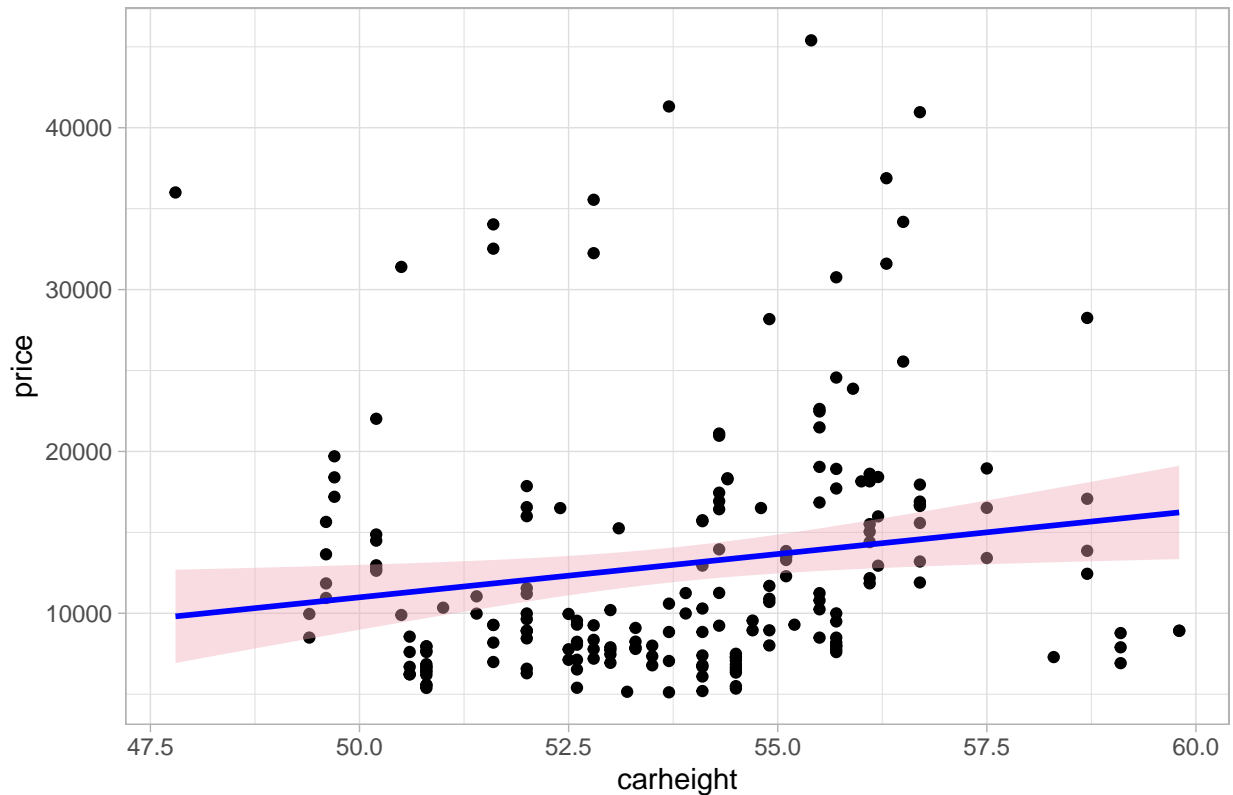
```
## 'geom_smooth()' using formula = 'y ~ x'
```



```
# Gráfico para No Convertibles
ggplot(M_non_convertible, aes(x = carheight, y = price)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue", fill = "pink2") +
  ggtitle("Relación Precio y Altura en Autos No Convertibles") +
  theme_light()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Relación Precio y Altura en Autos No Convertibles



Modelo con Interacciones

```
model_interaction = lm(price ~ carheight * carwidth * carbody, data = M)
summary(model_interaction)
```

```
##
## Call:
## lm(formula = price ~ carheight * carwidth * carbody, data = M)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-11721.4	-2229.1	-534.1	1197.2	20838.2

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-9.338e+05	5.428e+06	-0.172	0.863584
carheight	1.506e+04	1.069e+05	0.141	0.888123
carwidth	1.517e+04	8.441e+04	0.180	0.857535
carbodyhardtop	2.757e+07	7.165e+06	3.848	0.000164 ***
carbodyhatchback	1.693e+04	5.461e+06	0.003	0.997530
carbodysedan	2.439e+05	5.439e+06	0.045	0.964281
carbodywagon	1.486e+06	5.497e+06	0.270	0.787195
carheight:carwidth	-2.413e+02	1.662e+03	-0.145	0.884724
carheight:carbodyhardtop	-5.093e+05	1.371e+05	-3.715	0.000269 ***
carheight:carbodyhatchback	-3.312e+02	1.076e+05	-0.003	0.997547
carheight:carbodysedan	-6.049e+03	1.071e+05	-0.056	0.955024
carheight:carbodywagon	-2.678e+04	1.080e+05	-0.248	0.804406

```
## carwidth:carbodyhardtop          -4.133e+05  1.100e+05  -3.758 0.000229 ***
## carwidth:carbodyhatchback        -9.038e+02  8.491e+04  -0.011 0.991519
## carwidth:carbodiesedan            -4.038e+03  8.457e+04  -0.048 0.961970
## carwidth:carbodywagon            -2.367e+04  8.546e+04  -0.277 0.782075
## carheight:carwidth:carbodyhardtop  7.626e+03  2.105e+03   3.623 0.000376 ***
## carheight:carwidth:carbodyhatchback 1.444e+01  1.672e+03   0.009 0.993117
## carheight:carwidth:carbodiesedan   9.608e+01  1.665e+03   0.058 0.954046
## carheight:carwidth:carbodywagon    4.244e+02  1.679e+03   0.253 0.800675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4298 on 185 degrees of freedom
## Multiple R-squared:  0.7375, Adjusted R-squared:  0.7105
## F-statistic: 27.35 on 19 and 185 DF,  p-value: < 2.2e-16
```

Interpreta en el contexto del problema

El 73.74% de la variación en el precio de los automóviles se explica por este modelo, lo cual es considerado como un buen ajuste, aunque claramente hay espacio de mejora en el mismo. Las interacciones más relevantes para predecir el precio son aquellas que involucran la carrocería en combinación con la altura y el ancho del automóvil.

En resumen, aunque algunas variables individuales como carheight y carwidth no son significativas por sí solas, las interacciones entre estas variables y ciertos tipos de carrocería, tienen un gran efecto, esto se puede denotar mucho más sencillamente en los vehículos convertibles y no convertibles, ya que debido al número reducido de convertibles, no es muy útil la gráfica, mientras que es mucho más fácil de comprender en los vehículos no convertibles.

Más allá: Contesta la pregunta referida a la agrupación de variables que propuso la empresa para el análisis: ¿propondrías una nueva agrupación de las variables a la empresa automovilística?

Si, propondría una nueva agrupación de variables físicas (carheight, carwidth), ya que estas variables están directamente relacionadas con el tamaño físico del vehículo, ya que el tamaño afecta aspectos como el costo de fabricación y la aerodinámica, lo que influye en el precio final.

Retoma todas las variables y haz un análisis estadístico muy leve (medias y correlación) de cómo crees que se deberían agrupar para analizarlas.

```
media_carheight <- mean(M$carheight, na.rm = TRUE)
media_carwidth <- mean(M$carwidth, na.rm = TRUE)
media_price <- mean(M$price, na.rm = TRUE)

cat("Media de carheight:", media_carheight, "\n")
```

```
## Media de carheight: 53.72488
```

```
cat("Media de carwidth:", media_carwidth, "\n")
```

```
## Media de carwidth: 65.9078
```

```
cat("Media de price:", media_price, "\n")
```

```
## Media de price: 13276.71
```

```

correlacion <- cor(M[, c("carheight", "carwidth", "price")], use = "complete.obs")

cat("Matriz de correlación:\n")

```

```
## Matriz de correlación:
```

```
print(correlacion)
```

```
##           carheight carwidth   price
## carheight 1.0000000 0.2792103 0.1193362
## carwidth  0.2792103 1.0000000 0.7593253
## price     0.1193362 0.7593253 1.0000000
```

```
dimensiones <- M[, c("carheight", "carwidth")]
```

```
carroceria <- M$carbody
```

```

precio_por_carroceria <- aggregate(M$price, by = list(M$carbody), FUN = mean, na.rm = TRUE)
colnames(precio_por_carroceria) <- c("carbody", "mean_price")
print(precio_por_carroceria)

```

```
##           carbody mean_price
## 1 convertible    21890.50
## 2   hardtop     22208.50
## 3  hatchback    10376.65
## 4     sedan    14344.27
## 5     wagon    12371.96
```