A01411671 M1. Portafolio | Entrega Final

Miguel Ángel Bermea Rodríguez | A01411671

2023-09-12

Reporte final de "El precio de los autos"

Portafolio

Resumen

Problemática La empresa automovilística china busca ingresar al mercado estadounidense y competir con sus contrapartes locales. Para lograrlo, necesita comprender los factores que afectan el precio de los automóviles en los Estados Unidos. En este informe, se aborda esta problemática mediante el análisis de datos y técnicas estadísticas.

Metodología Se realizó un análisis exploratorio de datos para comprender la distribución y relaciones entre las variables. Se seleccionaron las variables más relevantes, como enginesize, horsepower y carwidth, para construir un modelo de regresión lineal múltiple. Se realizaron pruebas de hipótesis para validar el modelo acompañado de pruebas de normalidad de los residuos, interpretación del coeficiente de determinación, verificación de media cero y prueba de homocedasticidad.

Resultados Principales El modelo de regresión lineal múltiple explica aproximadamente el 82.1% de la variación en el precio de los automóviles. Las variables enginesize, horsepower y carwidth son significativas para predecir el precio. Sin embargo, se encontró heterocedasticidad en el modelo, lo que sugiere la necesidad de ajustes adicionales.

Introducción

La empresa automovilística china enfrenta un desafío importante al ingresar al mercado automotriz de los Estados Unidos. Para tener éxito en este mercado altamente competitivo, es crucial comprender los factores que influyen en el precio de los automóviles. ¿Qué características de un automóvil impactan más en su precio? ¿Cuál es la importancia de estas características en el contexto estadounidense?

Este problema es de gran relevancia ya que una comprensión precisa de los factores que afectan el precio permitirá a la empresa china tomar decisiones estratégicas informadas, como la fijación de precios competitivos y el desarrollo de automóviles que se adapten mejor a las preferencias de los consumidores estadounidenses.

En este informe, se abordará la problemática mediante un análisis de datos y técnicas estadísticas. Se explorarán las relaciones entre las características de los automóviles y sus precios en el mercado estadounidense, y se construirá un modelo de regresión lineal múltiple para predecir el precio en función de las características seleccionadas. Además, se realizarán pruebas de hipótesis para validar la adecuación del modelo.

Este análisis proporcionará información valiosa a la empresa china, ayudándoles a tomar decisiones fundamentadas mientras ingresan al mercado automotriz de los Estados Unidos.

Contenido de Primera entrega

Finalidad de la entrega

La primera la entregarás al finalizar la semana 3. La finalidad es que entregues un borrador de la versión total de la exploración y preparación de los datos para que te sea retroalimentada. Esta entrega sólo es requisito.

Descripción

Una empresa automovilística china aspira a entrar en el mercado estadounidense. Desea establecer allí una unidad de fabricación y producir automóviles localmente para competir con sus contrapartes estadounidenses y europeas. Contrataron una empresa de consultoría de automóviles para identificar los principales factores de los que depende el precio de los automóviles, específicamente, en el mercado estadounidense, ya que pueden ser muy diferentes del mercado chino. Esencialmente, la empresa quiere saber:

- Qué variables son significativas para predecir el precio de un automóvil
- Qué tan bien describen esas variables el precio de un automóvil

Lectura de datos

```
##   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
```

Exploración y preparación de la base de datos (Portafolio de Análisis)

1. Exploración de la base de datos En esta parte, se realizan medidas estadísticas apropiadas para las variables cuantitativas y cualitativas. Se exploran las variables cuantitativas con medidas de posición, gráficos de distribución y análisis de colinealidad. También se analizan las variables categóricas mediante gráficos de barras y diagramas de pastel.

1. Medidas estadísticas apropiadas para las variables cuantitativas y cualitativas

```
##    wheelbase        carlength        carwidth       carheight
##  Min.   : 86.60   Min.   :141.1   Min.   :60.30   Min.   :47.80
##  1st Qu.: 94.50   1st Qu.:166.3   1st Qu.:64.10   1st Qu.:52.00
##  Median : 97.00   Median :173.2   Median :65.50   Median :54.10
##  Mean   : 98.76   Mean   :174.0   Mean   :65.91   Mean   :53.72
##  3rd Qu.:102.40   3rd Qu.:183.1   3rd Qu.:66.90   3rd Qu.:55.50
##  Max.   :120.90   Max.   :208.1   Max.   :72.30   Max.   :59.80
##    curbweight     enginesize        stroke      compressionratio
##  Min.   :1488   Min.   : 61.0   Min.   :2.070   Min.   : 7.00
##  1st Qu.:2145   1st Qu.: 97.0   1st Qu.:3.110   1st Qu.: 8.60
##  Median :2414   Median :120.0   Median :3.290   Median : 9.00
##  Mean   :2556   Mean   :126.9   Mean   :3.255   Mean   :10.14
##  3rd Qu.:2935   3rd Qu.:141.0   3rd Qu.:3.410   3rd Qu.: 9.40
##  Max.   :4066   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
```

```
##  symboling           CarName      fueltype          carbody   drivewheel
##  -2: 3     peugeot 504   :  6   diesel: 20   convertible: 6   4wd:  9
##  -1:22     toyota corolla:  6   gas   :185   hardtop    : 8   fwd:120
##  0 :67     toyota corona :  6                hatchback  :70   rwd: 76
##  1 :54     subaru dl     :  4                sedan      :96
##  2 :32     honda civic   :  3                wagon      :25
##  3 :27     mazda 626     :  3
##            (Other)       :177
##  enginelocation enginetype  cylindernumber
##  front:202      dohc : 12   eight :  5
##  rear :  3      dohcv:  1   five  : 11
##                 l    : 12   four  :159
##                 ohc  :148   six   : 24
##                 ohcf : 15   three :  1
##                 ohcv : 13   twelve:  1
##                 rotor:  4   two   :  4
```

2. Exploración usando herramientas de visualización Variables cuantitativas

A continuación se presentan ciertas métricas de las variables (cuartiles y valores atípicos) y también se presenta una matriz de correlación entre las variables.

Por otro lado, si se desea observar a mayor detalle la exploración de estas variables puedes ir al apartado de anexos que se encuentra al final de este reporte. En dicha sección podrás encontrar los boxplots, histogramas y diagramas de dispersión que fueron generados en esta sección o fase del proyecto.

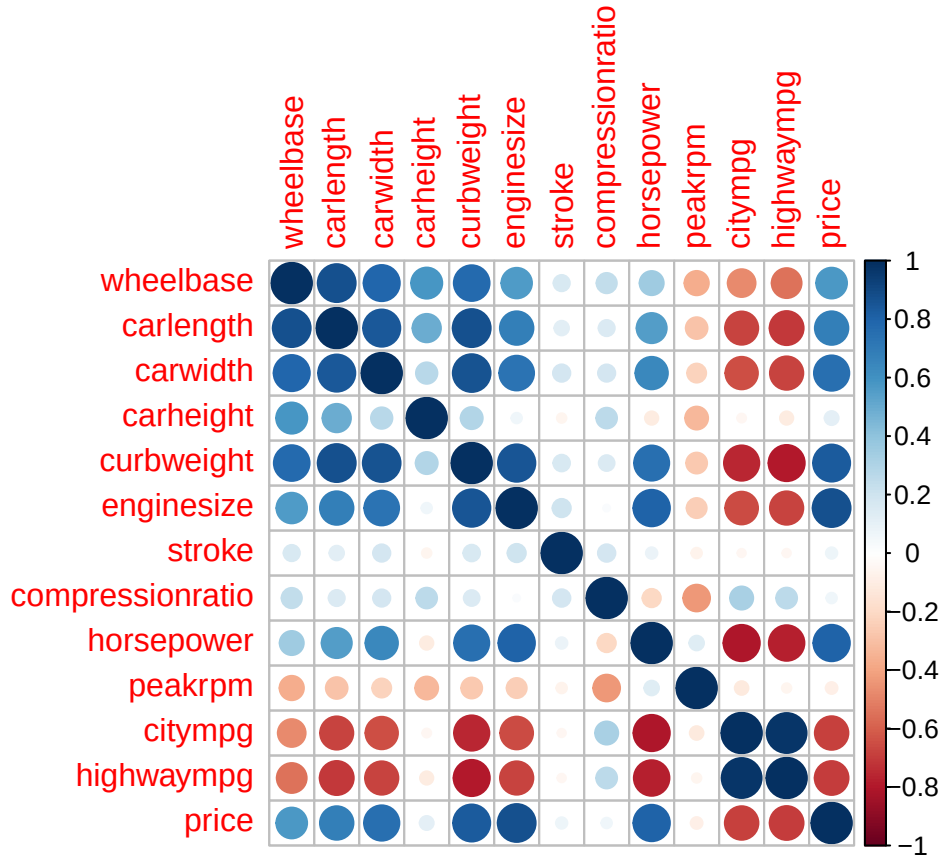
Gracias a la exploración utilizando herramientas de visualización fue que se logró observar la distribución de los datos y analizar colinealidad.

```
## [1] "Cuartiles para wheelbase"
##   25%   50%   75%
##  94.5  97.0 102.4
## [1] "Valores atípicos para wheelbase"
```

```
## [1] 115.6 115.6 120.9
## [1] "Cuartiles para carlength"
##   25%   50%   75%
## 166.3 173.2 183.1
## [1] "Valores atípicos para carlength"
## [1] 141.1
## [1] "Cuartiles para carwidth"
##  25%  50%  75%
## 64.1 65.5 66.9
## [1] "Valores atípicos para carwidth"
## [1] 71.4 71.4 71.4 71.7 71.7 71.7 72.0 72.3
## [1] "Cuartiles para carheight"
##  25%  50%  75%
## 52.0 54.1 55.5
## [1] "Valores atípicos para carheight"
## numeric(0)
## [1] "Cuartiles para curbweight"
##  25%  50%  75%
## 2145 2414 2935
## [1] "Valores atípicos para curbweight"
## integer(0)
## [1] "Cuartiles para enginesize"
## 25% 50% 75%
##  97 120 141
## [1] "Valores atípicos para enginesize"
##  [1] 209 209 209 258 258 326 234 234 308 304
## [1] "Cuartiles para stroke"
##  25%  50%  75%
## 3.11 3.29 3.41
## [1] "Valores atípicos para stroke"
##  [1] 3.90 4.17 4.17 2.19 2.19 3.90 3.90 2.07 2.36 2.64 2.64 2.64 2.64 2.64 2.64
## [16] 2.64 2.64 2.64 2.64 2.64
## [1] "Cuartiles para compressionratio"
## 25% 50% 75%
## 8.6 9.0 9.4
## [1] "Valores atípicos para compressionratio"
##  [1]  7.0  7.0 11.5 22.7 22.0 21.5 21.5 21.5 21.5  7.0  7.0  7.0 21.9 21.0 21.0
## [16] 21.0 21.0 21.0  7.0  7.0 22.5 22.5 22.5 23.0 23.0 23.0 23.0 23.0
## [1] "Cuartiles para horsepower"
## 25% 50% 75%
##  70  95 116
## [1] "Valores atípicos para horsepower"
## [1] 262 200 207 207 207 288
## [1] "Cuartiles para peakrpm"
##  25%  50%  75%
## 4800 5200 5500
## [1] "Valores atípicos para peakrpm"
## [1] 6600 6600
## [1] "Cuartiles para citympg"
## 25% 50% 75%
##  19  24  30
## [1] "Valores atípicos para citympg"
## [1] 47 49
## [1] "Cuartiles para highwaympg"
```

```
## 25% 50% 75%
##  25  30  34
## [1] "Valores atípicos para highwaympg"
## [1] 53 54 50
## [1] "Cuartiles para price"
##   25%   50%   75%
##  7788 10295 16503
## [1] "Valores atípicos para price"
##  [1] 30760.0 41315.0 36880.0 32250.0 35550.0 36000.0 31600.0 34184.0 35056.0
## [10] 40960.0 45400.0 32528.0 34028.0 37028.0 31400.5
```

```
## corrplot 0.92 loaded
```



Variables categóricas

En este apartado se trabajó con la exploración de las variables categóricas del conjunto de datos (se enlistan a continuación).

Si se desea observar la exploración de estas variables puedes ir al apartado de anexos que se encuentra al final de este reporte. En dicha sección podrás encontrar los diagramas de barras y diagramas de pastel que fueron generados en esta sección o fase del proyecto, asi como los boxplots y diagramas de barras de precio por cada categoria de cada variable.

Gracias a la exploración utilizando herramientas de visualización fue que se logró observar la distribución de los datos y analizar asociación o colinealidad.

```
## [1] "symboling"      "CarName"        "fueltype"       "carbody"
## [5] "drivewheel"     "enginelocation" "enginetype"     "cylindernumber"
```

3. Problemas de calidad de datos (valores faltantes y outliers) En esta sección, se identifican los valores faltantes por columna y se analizan los outliers en las variables cuantitativas y cualitativas. Se mencionan posibles acciones a tomar en relación a los outliers.

Valores faltantes por columna

```
##        symboling          CarName         fueltype          carbody
##                0                0                0                0
##       drivewheel   enginelocation        wheelbase        carlength
##                0                0                0                0
##         carwidth        carheight       curbweight       enginetype
##                0                0                0                0
##   cylindernumber       enginesize           stroke compressionratio
##                0                0                0                0
##       horsepower          peakrpm          citympg       highwaympg
##                0                0                0                0
##            price
##                0
```

Outliers

En la sección 2. Exploración usando herramientas de visualización:

- Para las variables cuantitvas se hizo uso de IQR para obtener los outliers y se visualizó con boxplot
- Para las variables cualitativas también se visualizó con boxplot

4. Selección de variables para el análisis de las características de los automóviles que determinan su precio Se mencionan las variables seleccionadas para el análisis, como enginesize, horsepower, y carwidth, y se plantea la posibilidad de manejar datos categóricos mediante variables dummy.

** = Tal vez la utilizo

- enginesize
- horsepower
- curbweight
- carlength
- carwidth
- wheelbase
- cylindernumber**
- carbody**

2. Preparación de la base de datos

1. Selección de datos a utilizar

- Maneja datos categóricos: transforma a variables dummy si es necesario.

En caso de que utilice los datos categóricos preseleccionados (cylindernumber y carbody), debo de tener en cuenta esto para manejarlos de la forma adecuada.

```
##   cylindernumbereight cylindernumberfive cylindernumberfour cylindernumbersix
## 1                   0                  0                  1                 0
## 2                   0                  0                  1                 0
## 3                   0                  0                  0                 1
```

```
## 4                   0                 0                 1                 0
## 5                   0                 1                 0                 0
## 6                   0                 1                 0                 0
##   cylindernumberthree cylindernumbertwelve cylindernumbertwo
## 1                   0                    0                 0
## 2                   0                    0                 0
## 3                   0                    0                 0
## 4                   0                    0                 0
## 5                   0                    0                 0
## 6                   0                    0                 0
##   carbodyconvertible carbodyhardtop carbodyhatchback carbodysedan carbodywagon
## 1                  1              0                 0           0            0
## 2                  1              0                 0           0            0
## 3                  0              0                 1           0            0
## 4                  0              0                 0           1            0
## 5                  0              0                 0           1            0
## 6                  0              0                 0           1            0
```

En caso de que solo utilice datos cuantitativos, no debo de preocuparme.

- Maneja apropiadamente datos atípicos.
  - Identificación: Ya identifique los datos atípicos mediante IQR y se pueden visualizar con los boxplots.
  - Contexto: La existencia de estos datos se debe a ser datos válidos pero poco comunes
  - Decisión: Si bien aún no tengo la certeza, si que he identificado posibles alternativas para manejar estos datos

    ```
    1) Eliminarlos: En caso de que los considere no representativos,
    puedo optar por eliminarlos. Cabe recalcar que debo estar seguro
    ya que podría afectar la integridad de mis resultados.

    2) Transformar los datos (va de la mano con lo último de esta entrega):
    En algunos casos podría aplicar transformaciones matemáticas a mis datos
    para reducir el impacto de los valores atípicos.

    3) Tratarlos por separado: Puedo analizarlos por separado para entender
    mejor su naturaleza y posible impactos en mis resultados.
    ```

    Actualmente me decanto más por tratarlos por separado.

2. Transformación de datos a utilizar (en caso de ser necesario)

- Revisa si es necesario discretizar los datos

En caso de ser necesario, tengo pensado hacer uso de la función *cut* en R para dividir una variable numérica en intervalos y asignar cada observación a uno de esos intervalos.

- Revisa si es necesario escalar y normalizar los datos

En caso de ser necesario, tengo pensado hacer uso de la función *scale* en R para estandarizar una variable numérica restando su media y dividiendo por su desviación estándar.

Destacar que la transformación de datos dependerá del modelo a utilizar (lo cual entra en Modelación y verificación del modelo (Portafolio de implementación), ya que es aqui donde buscaré el mejor modelo predictivo para la variable precio)

Contenido de Segunda entrega

Finalidad de la entrega

Se entregará al finalizar la semana 4. La finalidad es que entregues un borrador de la versión total de la modelación y verificación del modelo (solución final) para que te sea retroalimentada. Esta entrega también es sólo requisito.

Regresión lineal múltiple y Pruebas de hipótesis

La regresión lineal múltiple es una herramienta estadística útil para analizar la relación entre una variable dependiente y varias variables independientes.

En el contexto de la situación problema, la regresión lineal múltiple puede utilizarse para predecir el precio de un automóvil en función de sus características (variables seleccionadas). Esta herramienta me permite identificar qué variables son significativas para predecir el precio del automóvil y cuantificar su efecto sobre el precio.

Por otro lado, las pruebas de hipótesis pueden utilizarse para verificar si se cumplen los supuestos de un modelo de regresión lineal múltiple, como la normalidad y la homocedasticidad de los residuos. Esto me permite validar el modelo y asegurarme de que las predicciones sean confiables.

```
##
## Call:
## lm(formula = price ~ enginesize + horsepower + curbweight + carlength +
##     carwidth + wheelbase, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8209.0 -1637.8   -64.6  1448.2 14600.0
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -45156.539  12856.573  -3.512  0.00055 ***
## enginesize      82.909     12.709   6.523 5.61e-10 ***
## horsepower      53.480     12.477   4.286 2.84e-05 ***
## curbweight       2.493      1.556   1.602  0.11070
## carlength      -58.175     53.685  -1.084  0.27985
## carwidth       556.627    253.009   2.200  0.02896 *
## wheelbase       95.314     98.318   0.969  0.33350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3430 on 198 degrees of freedom
## Multiple R-squared:  0.8211, Adjusted R-squared:  0.8156
## F-statistic: 151.4 on 6 and 198 DF,  p-value: < 2.2e-16
```

Validación del modelo Coeficiente de determinación

```
## [1] 0.8210592
```

El coeficiente de determinación o R-cuadrado, es una medida que representa que tan bien la línea de regresión se ajusta a los datos, es decir, la bondad de ajuste del modelo.

En este caso, el valor de R-cuadrado es 0.8210592, lo que significa que aproximadamente el 82.1% de la variación en la variable dependiente (precio del automóvil) puede ser explicada por las variables independientes seleccionadas para el modelo.

Cabe recalcar que las variables *enginesize, horsepower* y *carwidth* son las más significativas para predecir el precio del automóvil, ya que tienen p-values bajos (menores que 0.05).

El gráfico muestra una barra que representa el valor del R-cuadrado y tiene un eje y que va desde 0 hasta 1. Esto permite ver qué tan cerca está el modelo de explicar completamente la variabilidad en la variable dependiente.

Normalidad de los residuos

(1) Hipótesis de normalidad

- H0: Los residuos siguen una distribución normal.
- Ha: Los residuos no siguen una distribución normal.

(2) Regla de decisión

Se utilizará la prueba de Anderson-Darling para evaluar la normalidad de los residuos.

Si el p-value es menor que el nivel de significancia (por ejemplo, 0.05), rechazamos la hipótesis nula y concluimos que los residuos no siguen una distribución normal.

(3) Análisis del resultado

```
##
##  Anderson-Darling normality test
##
## data:  model$residuals
## A = 2.4285, p-value = 3.684e-06
```

(4) Conclusión

El p-value es 3.684e-06, lo que es muy bajo. Esto indica que los residuos del modelo no siguen una distribución normal y, por lo tanto, se rechaza la hipótesis nula de normalidad. Esto sugiere que podría ser necesario revisar el modelo y considerar la posibilidad de transformar las variables o utilizar un modelo diferente para ajustar los datos.

Verificación de media cero

(1) Hipótesis de media cero

- H0: La media de los residuos es igual a 0.
- H1: La media de los residuos no es igual a 0.

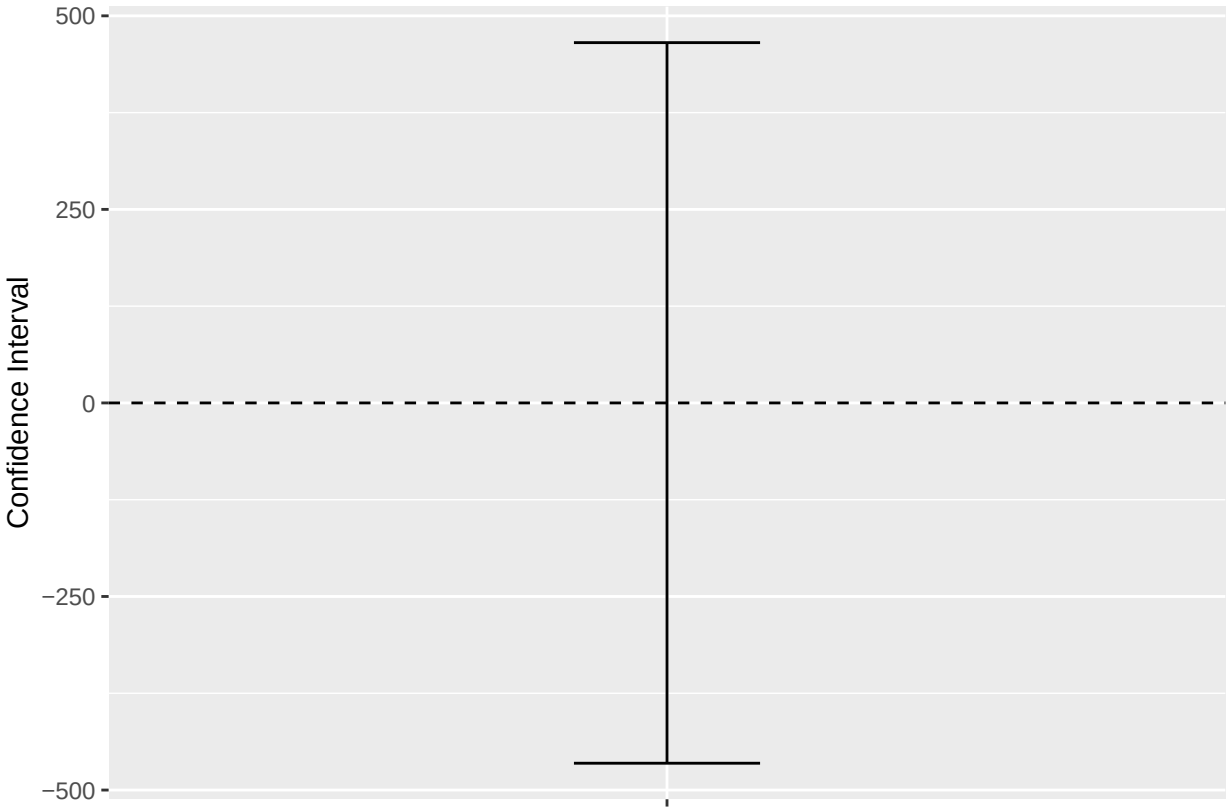
(2) Regla de decisión

Se utilizará la prueba t de una muestra aplicada a los residuos del modelo para determinar si la media es igual a un valor específico.

El p-value indica si la media de los residuos es igual a 0. Un p-value bajo (menor que 0.05) indica que la media de los residuos no es igual a 0 y, por lo tanto, se rechaza la hipótesis nula.

(3) Análisis del resultado

```
##
##  One Sample t-test
##
## data:  model$residuals
## t = -4.2715e-16, df = 204, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -465.3657  465.3657
## sample estimates:
##     mean of x
## -1.008191e-13
```



La línea punteada que representa la hipótesis nula (media igual a 0) se encuentra dentro del intervalo de confianza del 95% para la media de los residuos, esto indica que no hay evidencia suficiente para rechazar la hipótesis nula. En otras palabras, no hay evidencia suficiente para afirmar que la media de los residuos no es igual a 0.

(4) Conclusión

En este caso, el p-value es 1, lo que es muy alto. Esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de que la media de los residuos es igual a 0. Esto sugiere que el modelo está bien especificado y que no hay sesgo en las predicciones.

Homocedasticidad

(1) Hipótesis de prueba de Breusch-Pagan

- H0: No hay heterocedasticidad en el modelo, la varianza de los residuos es constante.
- Ha: Hay heterocedasticidad en el modelo, la varianza de los residuos no es constante.

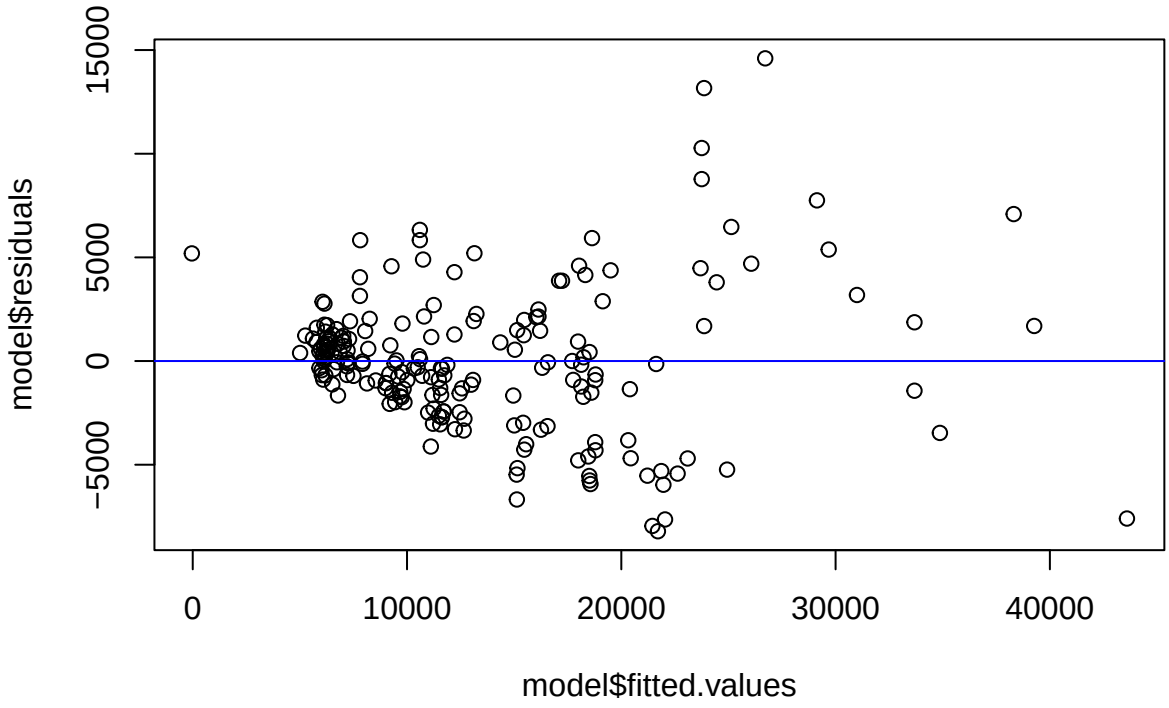
(2) Regla de decisión

Para verificar si se cumple el supuesto de homocedasticidad en el modelo, se utilizará la prueba de Breusch-Pagan. Esta prueba se utiliza para determinar si hay evidencia de heterocedasticidad (es decir, si la varianza de los residuos no es constante) en un modelo de regresión lineal múltiple.

Si el p-value es menor que el nivel de significancia (por ejemplo, 0.05), se rechaza la hipótesis nula y se concluye que hay evidencia suficiente para afirmar que hay heterocedasticidad en el modelo.

(3) Análisis del resultado

Se grafican los residuos para observar tendencia:



Se utiliza prueba de Breusch-Pagan:

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

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

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

```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 74.806, df = 6, p-value = 4.209e-14
```

(4) Conclusión

En este caso, el p-value es 4.209e-14, lo que es muy bajo. Esto indica que hay evidencia suficiente para rechazar la hipótesis nula y concluir que hay heterocedasticidad en el modelo. Esto sugiere que podría ser necesario revisar el modelo y considerar la posibilidad de transformar las variables o utilizar técnicas como la regresión ponderada para corregir este problema.

Conclusión del análisis

En base a los resultados de mi análisis, se puede concluir que el modelo de regresión lineal múltiple que ajusté para predecir el precio de un automóvil en función de la variables seleccionadas tiene un buen ajuste. El valor del coeficiente de determinación (R-cuadrado) es 0.8210592, lo que significa que aproximadamente el 82.1%

de la variación en el precio del automóvil puede ser explicada por las variables independientes seleccionadas para el modelo. Además, las variables *enginesize, horsepower* y *carwidth* son significativas para predecir el precio del automóvil, ya que tienen p-values bajos.

Sin embargo, también se encontró evidencia de heterocedasticidad en el modelo, lo que indica que la varianza de los residuos no es constante. Esto sugiere que podría ser necesario revisar el modelo y considerar la posibilidad de transformar las variables o utilizar técnicas como la regresión ponderada para corregir este problema.

En resumen, mi análisis muestra que el modelo de regresión lineal múltiple que ajusté tiene un buen ajuste y que las variables *enginesize, horsepower* y *carwidth* son significativas para predecir el precio del automóvil. Sin embargo, también se encontró evidencia de heterocedasticidad en el modelo, lo que sugiere que podría ser necesario revisarlo y la posibilidad de considerar otras variables que reemplacen a las que no fueron tan significativas, transformar las variables o utilizar técnicas para corregir este problema.

Conclusión general

En este informe, abordé la problemática planteada por la empresa automovilística china que busca ingresar al mercado estadounidense y competir con sus contrapartes locales. La empresa deseaba comprender los factores que influyen en el precio de los automóviles en los Estados Unidos, específicamente. El análisis se centró en identificar qué variables son significativas para predecir el precio de un automóvil y qué tan bien describen esas variables el precio de un automóvil en este mercado.

Los resultados indican que el modelo de regresión lineal múltiple que construí tiene un buen ajuste y puede explicar aproximadamente el 82.1% de la variación en el precio de los automóviles. Además, identifiqué que las variables enginesize, horsepower y carwidth son las más significativas para predecir el precio de un automóvil en el mercado estadounidense.

Sin embargo, también encontré evidencia de heterocedasticidad en el modelo, lo que sugiere que la varianza de los residuos no es constante. Esto puede requerir ajustes adicionales en el modelo o la consideración de otras variables que puedan mejorar su capacidad predictiva.

En resumen, este análisis proporciona a la empresa automovilística china información valiosa que les ayudará a tomar decisiones estratégicas informadas a medida que ingresan al mercado automotriz de los Estados Unidos. Les permite comprender mejor qué características de los automóviles impactan significativamente en el precio y cómo pueden adaptar sus productos y estrategias de mercado en consecuencia.

Anexos

Se proporcionan anexos con gráficos adicionales que respaldan el análisis y las conclusiones presentadas en el reporte, asi como enlaces de interés para las Evidencias de la unidad de formación.

Boxplots e histogramas de variables cuantitativas

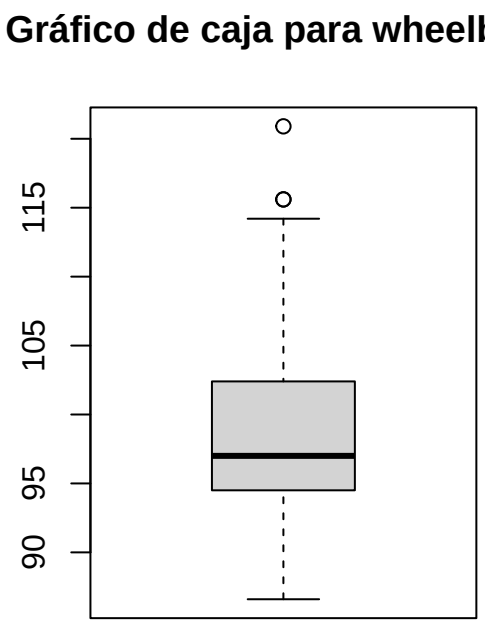
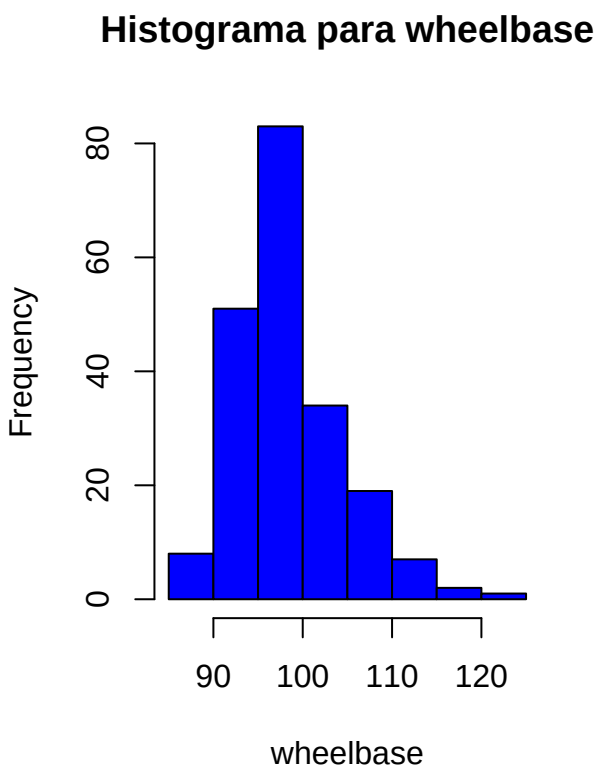


Gráfico de caja para wheelbase
90
95
105
115


Histograma para wheelbase
Frequency
0
20
40
60
80
90
100
110
120
wheelbase

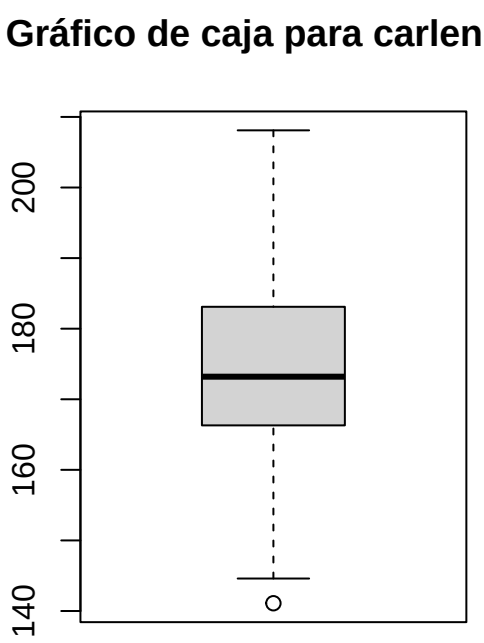
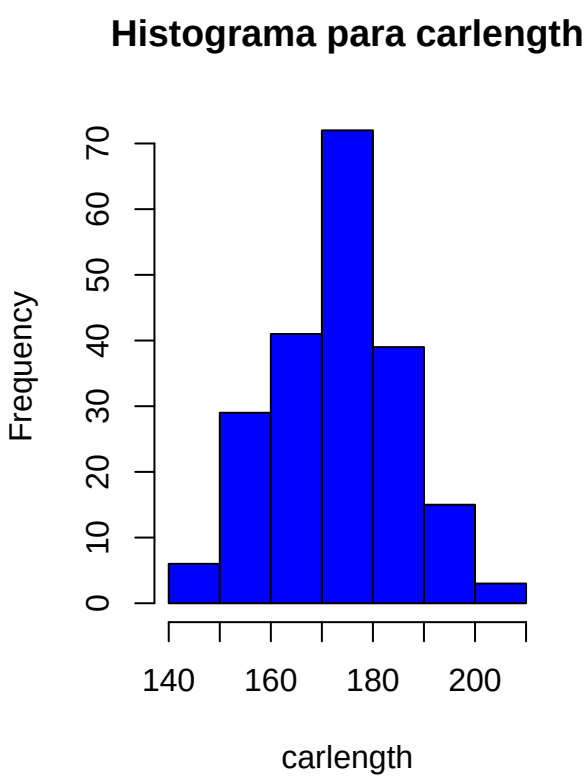


Gráfico de caja para carlength
140
160
180
200


Histograma para carlength
Frequency
0
10
20
30
40
50
60
70
140
160
180
200
carlength

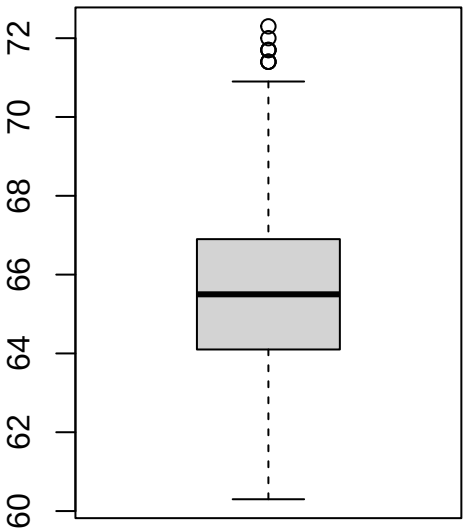
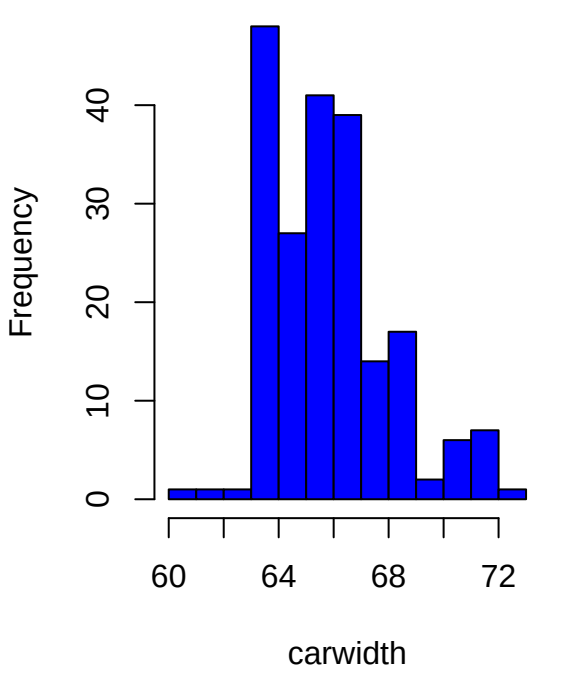


Gráfico de caja para carwidth
60
62
64
66
68
70
72


Histograma para carwidth
Frequency
0
10
20
30
40
60
64
68
72
carwidth

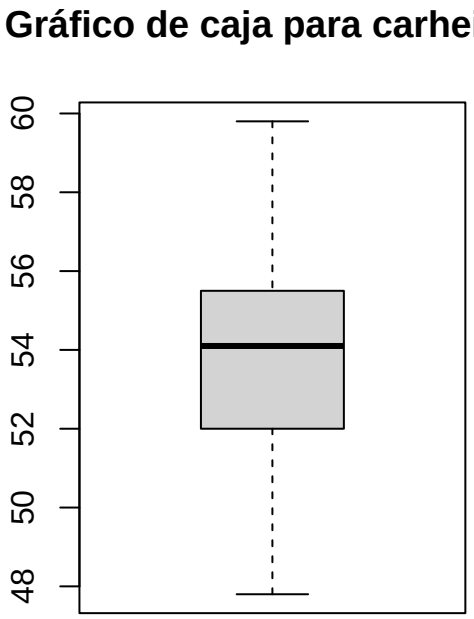
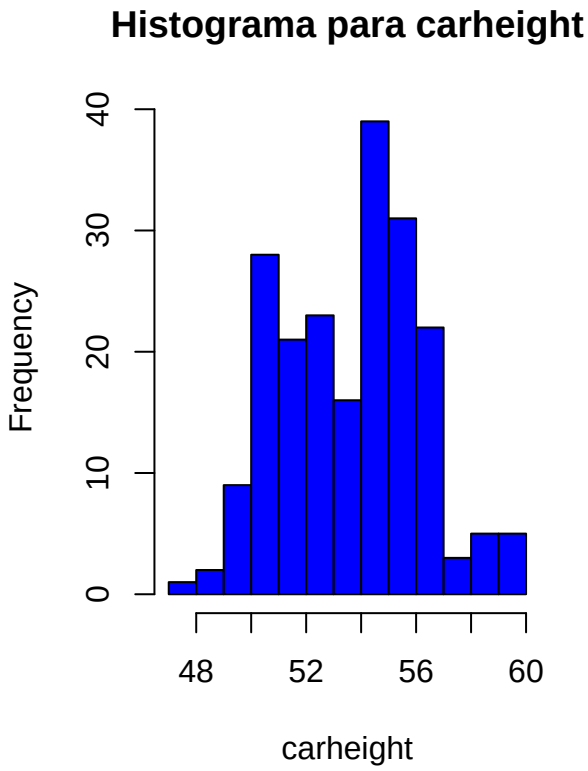


Gráfico de caja para carheight
48
50
52
54
56
58
60


Histograma para carheight
Frequency
0
10
20
30
40
48
52
56
60
carheight

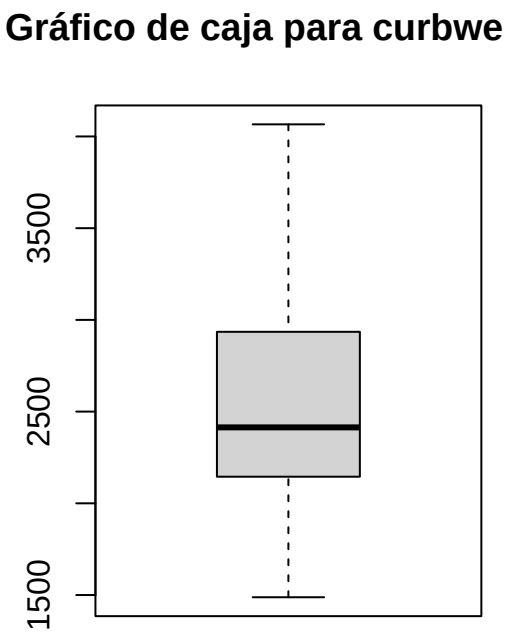
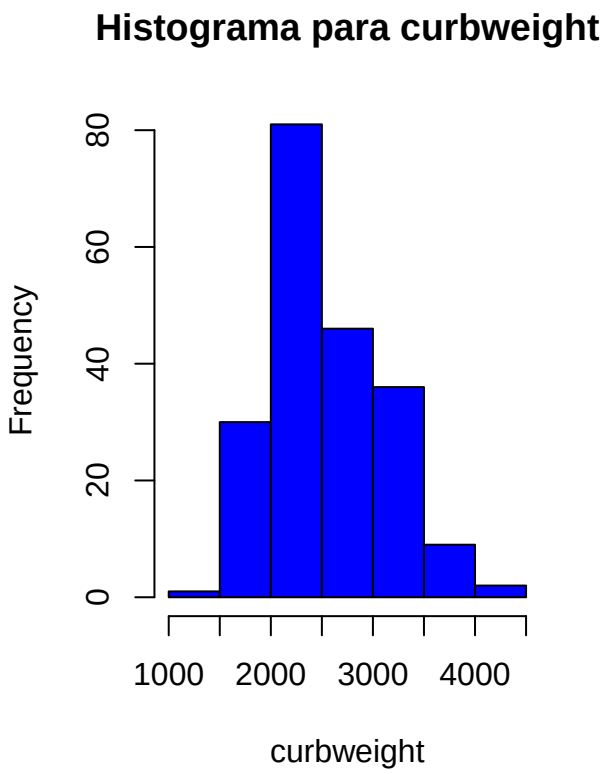


Gráfico de caja para curbweigh
1500
2500
3500


Histograma para curbweight
Frequency
0
20
40
60
80
1000
2000
3000
4000
curbweight

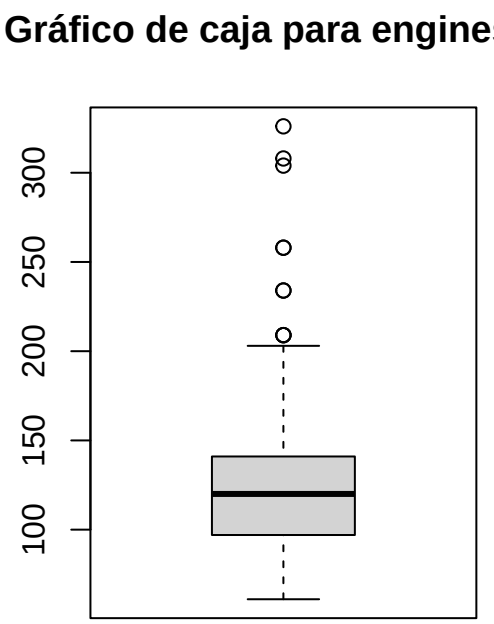
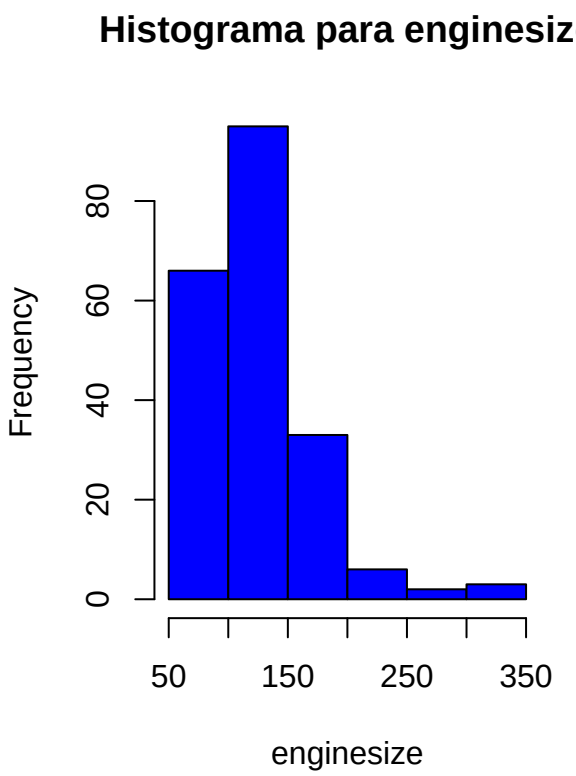


Gráfico de caja para enginesize
100
150
200
250
300


Histograma para enginesize
Frequency
0
20
40
60
80
50
150
250
350
enginesize

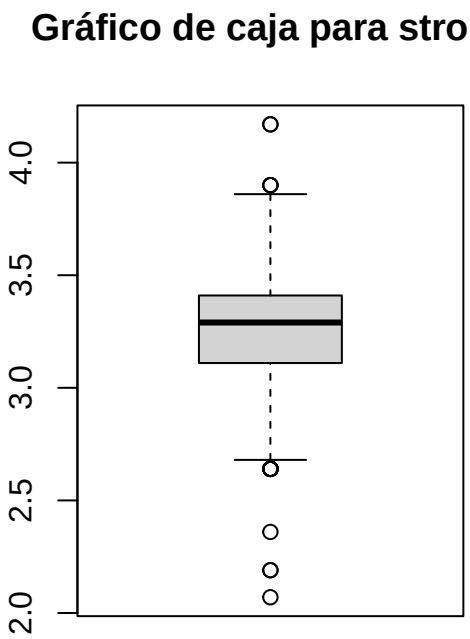


Gráfico de caja para stroke
2.0
2.5
3.0
3.5
4.0

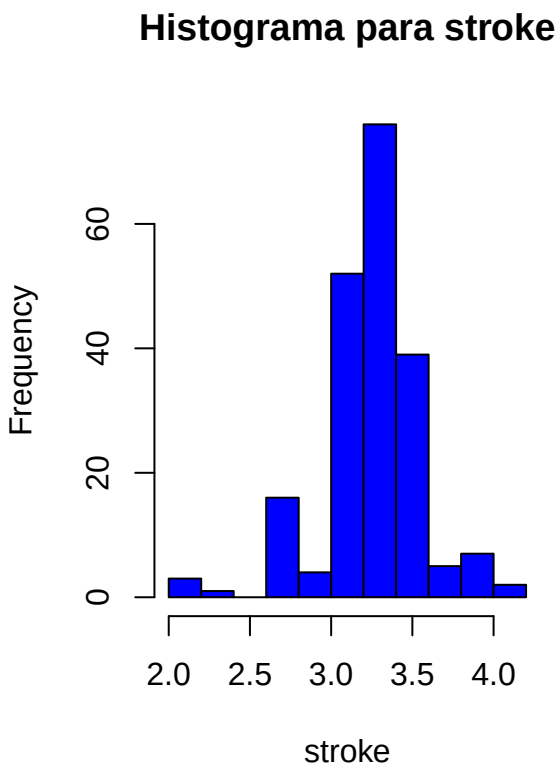
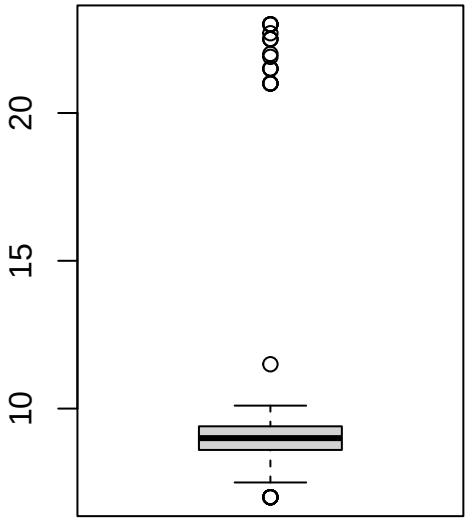
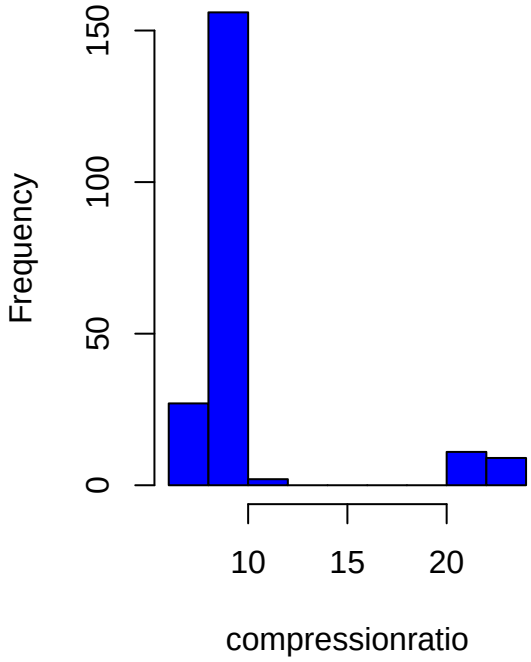

Histograma para stroke
Frequency
0
20
40
60
2.0
2.5
3.0
3.5
4.0
stroke

Gráfico de caja para compressionr



Histograma para compressionrat



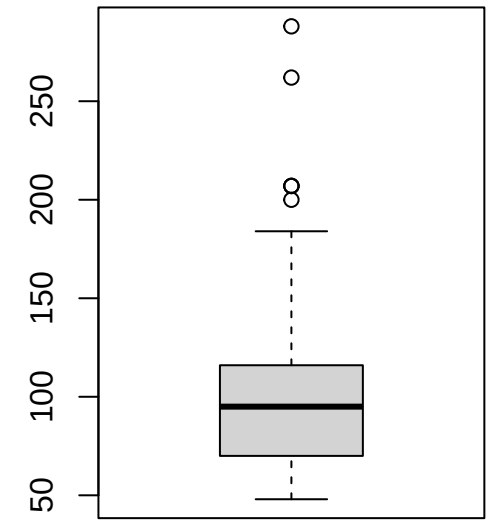
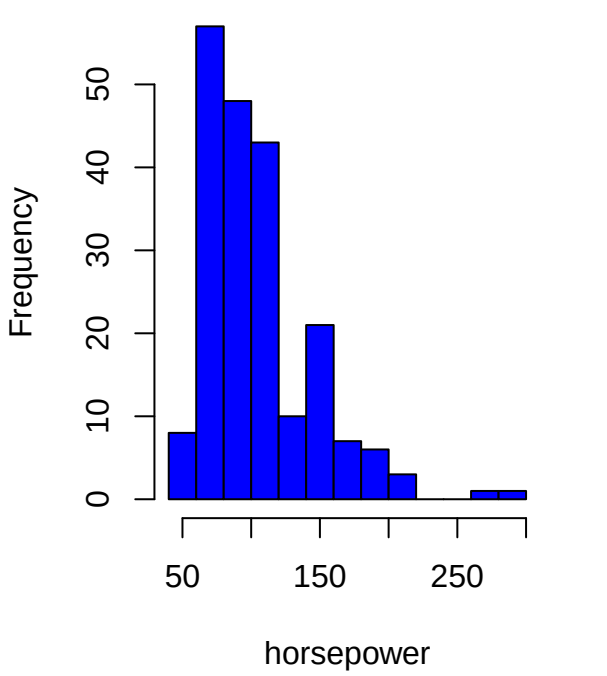


Gráfico de caja para horsepowe
50
100
150
200
250


Histograma para horsepower
Frequency
0
10
20
30
40
50
50
150
250
horsepower

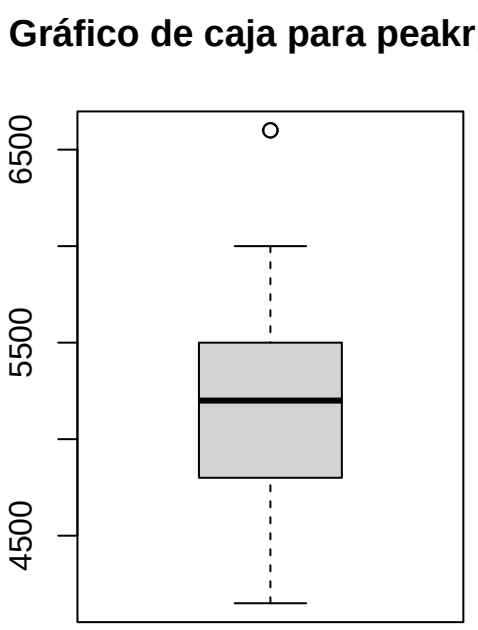
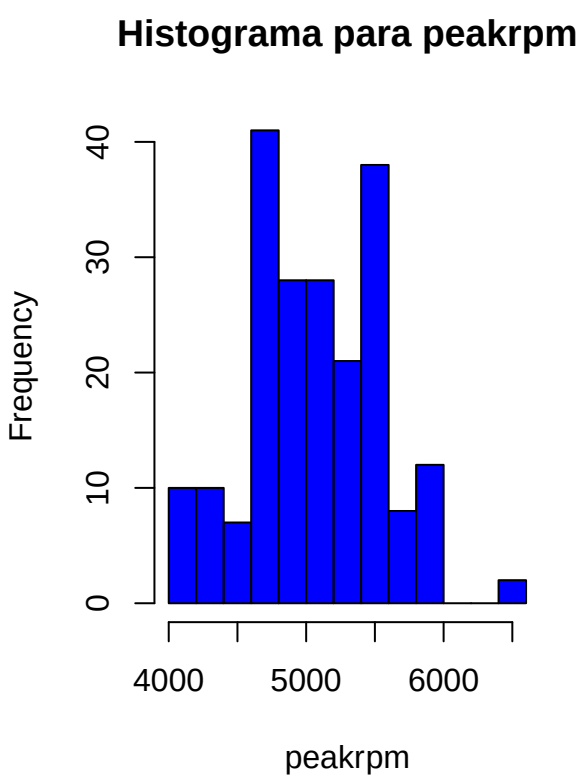


Gráfico de caja para peakrpm
6500
5500
4500


Histograma para peakrpm
Frequency
0
10
20
30
40
4000
5000
6000
peakrpm

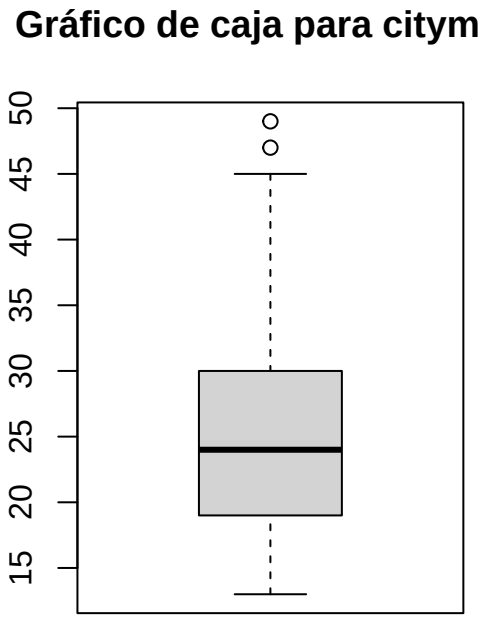


Gráfico de caja para citympg
15
20
25
30
35
40
45
50

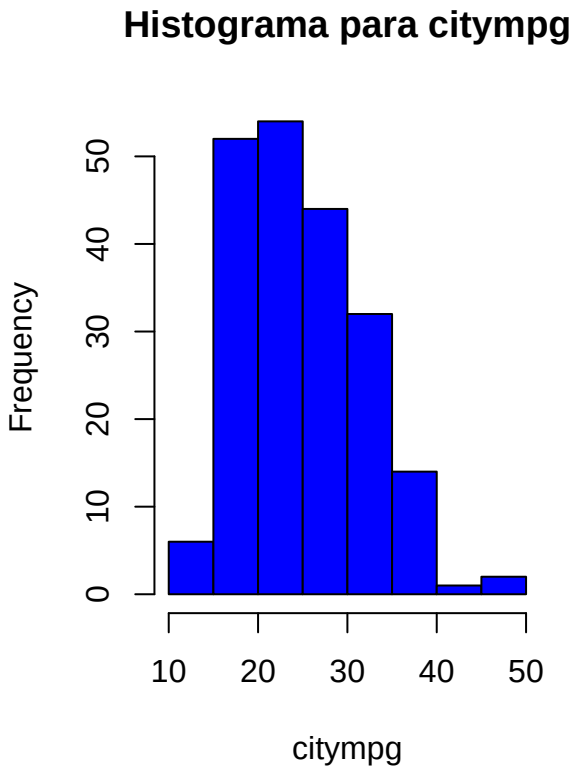
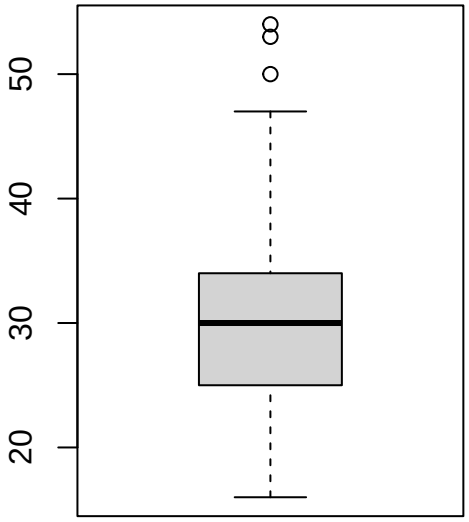
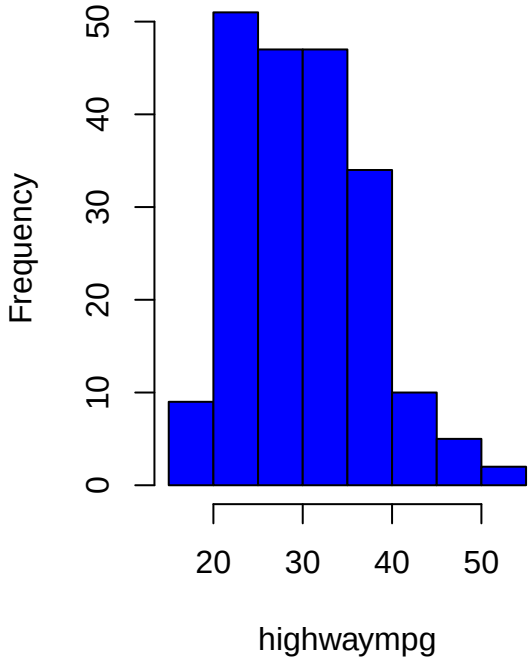
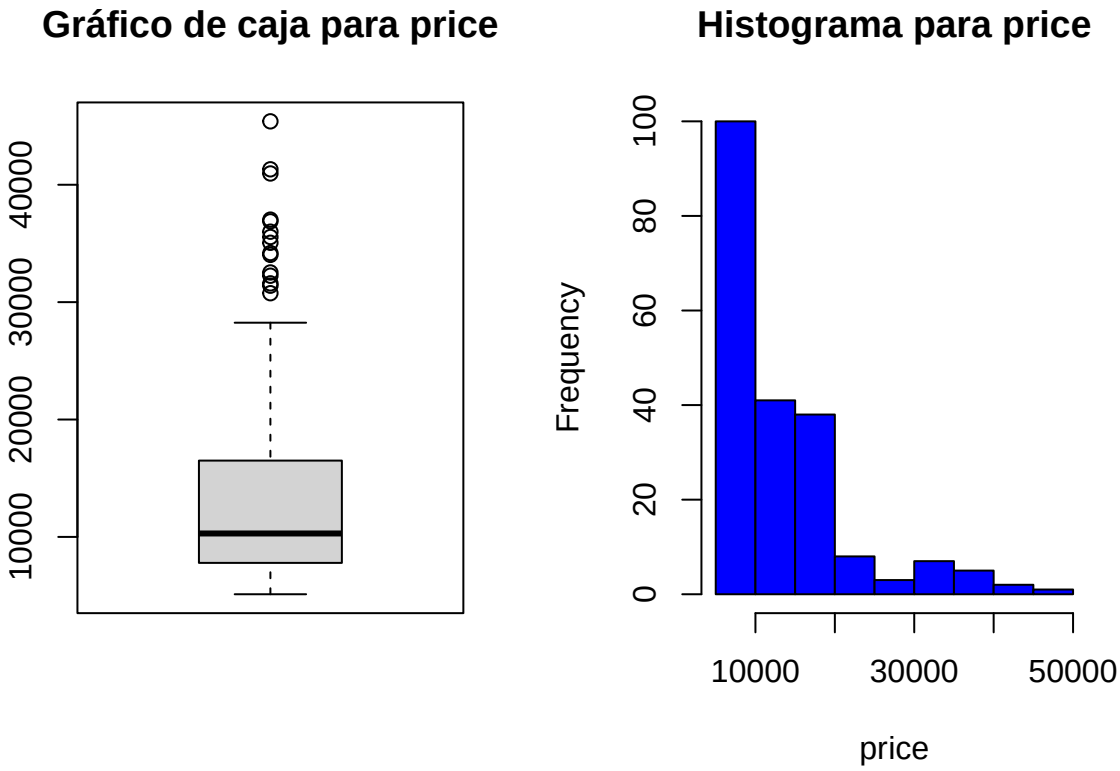

Histograma para citympg
Frequency
0
10
20
30
40
50
10
20
30
40
50
citympg

Gráfico de caja para highwaymp



Histograma para highwaympg





Diagramas de dispersión de variables cuantitativas

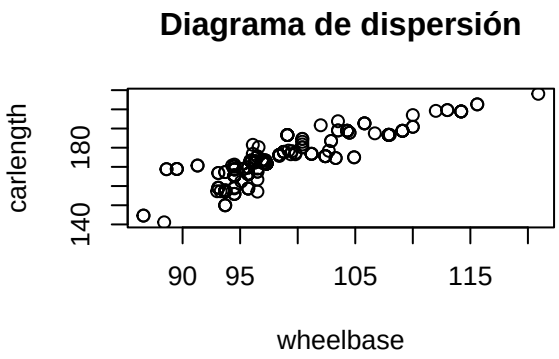


Diagrama de dispersión
carlength
140
180
90
95
105
115
wheelbase

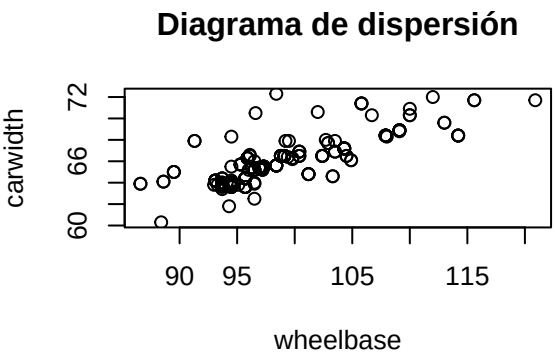


Diagrama de dispersión
carwidth
60
66
72
90
95
105
115
wheelbase

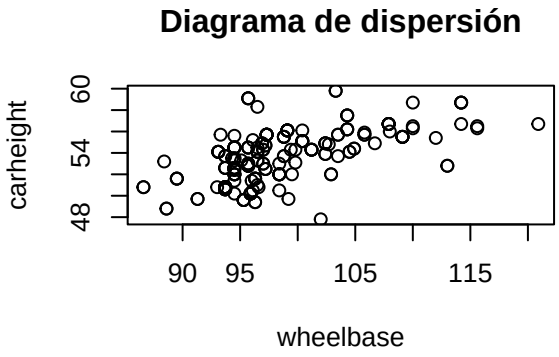


Diagrama de dispersión
carheight
48
54
60
90
95
105
115
wheelbase

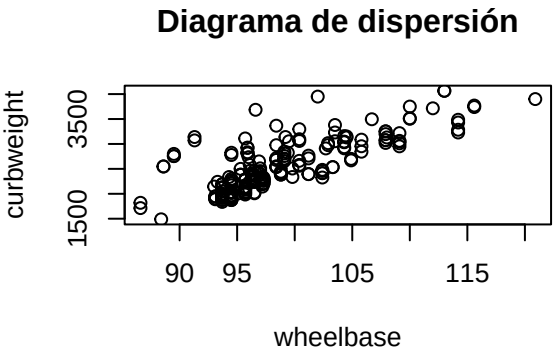


Diagrama de dispersión
curbweight
1500
3500
90
95
105
115
wheelbase

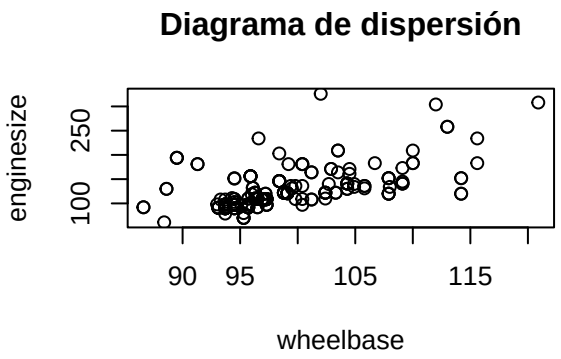


Diagrama de dispersión
enginesize
100
250
90
95
105
115
wheelbase

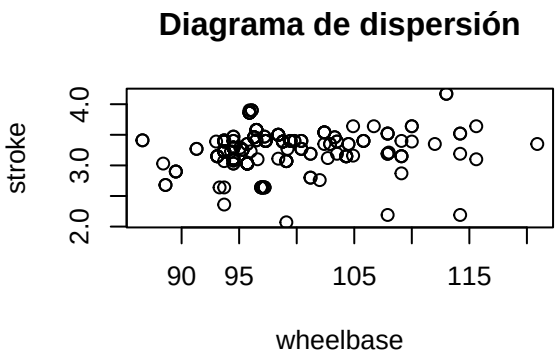


Diagrama de dispersión
stroke
2.0
3.0
4.0
90
95
105
115
wheelbase

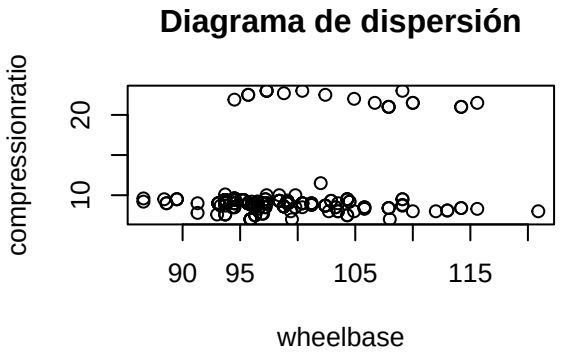


Diagrama de dispersión
compressionratio
10
20
90
95
105
115
wheelbase

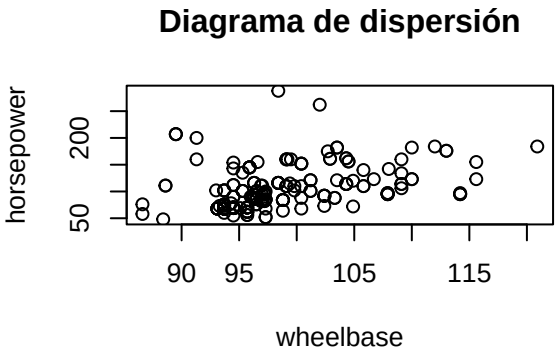


Diagrama de dispersión
horsepower
50
200
90
95
105
115
wheelbase

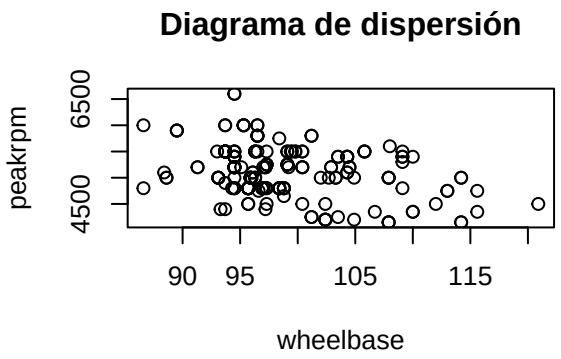


Diagrama de dispersión
peakrpm
4500
6500
90
95
105
115
wheelbase

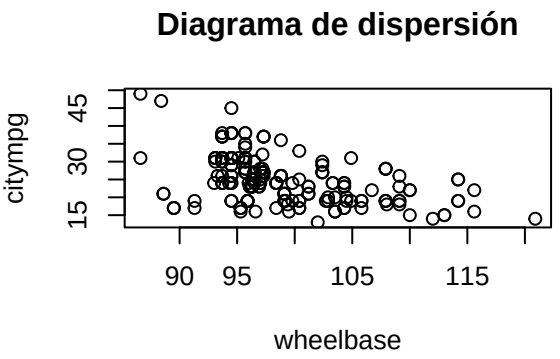


Diagrama de dispersión
citympg
15
30
45
90
95
105
115
wheelbase

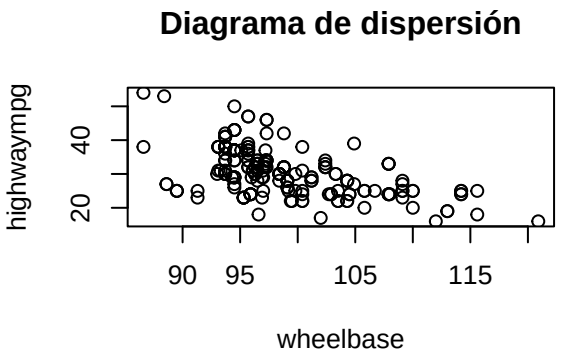


Diagrama de dispersión
highwaympg
20
40
90
95
105
115
wheelbase

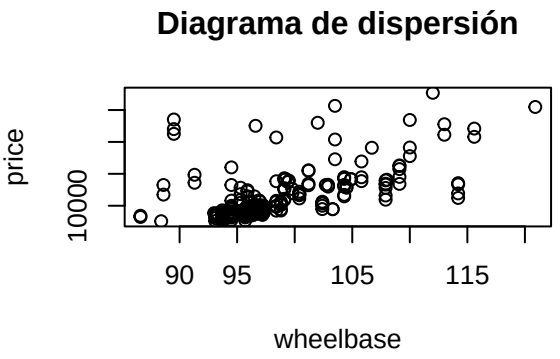


Diagrama de dispersión
price
10000
90
95
105
115
wheelbase

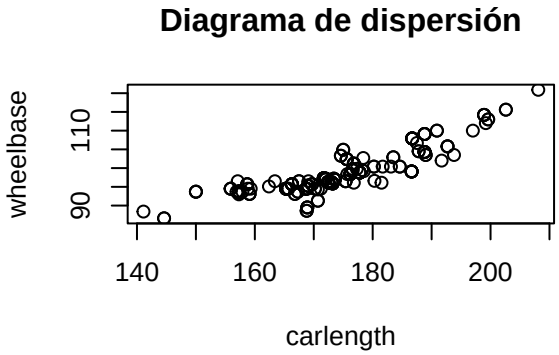


Diagrama de dispersión
wheelbase
90
110
140
160
180
200
carlength

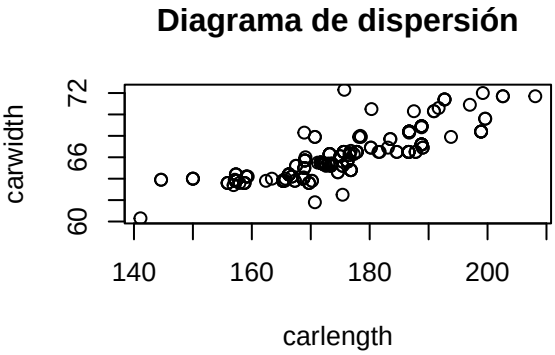


Diagrama de dispersión
carwidth
60
66
72
140
160
180
200
carlength

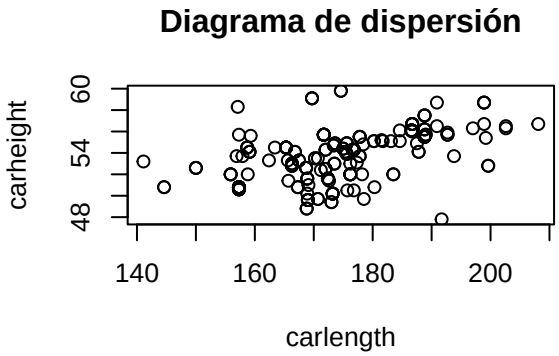


Diagrama de dispersión
carheight
48
54
60
140
160
180
200
carlength

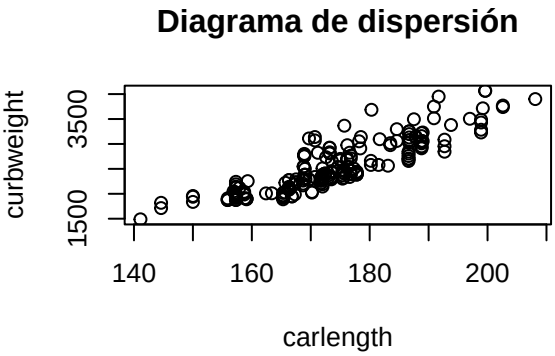


Diagrama de dispersión
curbweight
1500
3500
140
160
180
200
carlength

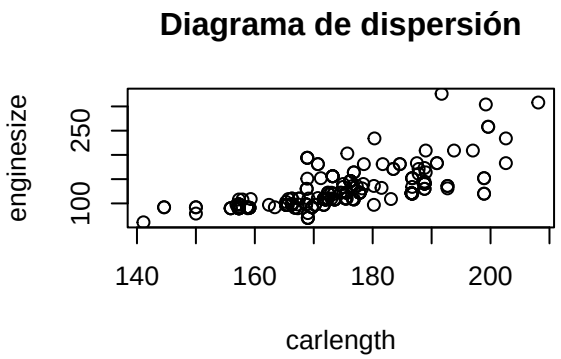


Diagrama de dispersión
enginesize
100
250
140
160
180
200
carlength

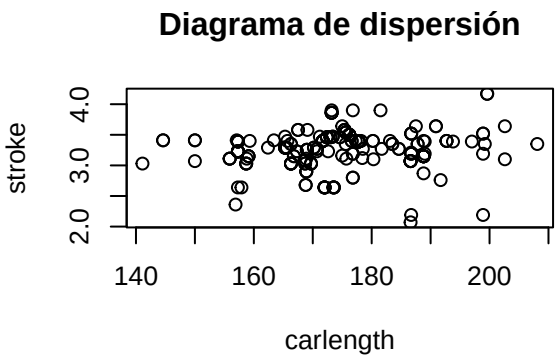


Diagrama de dispersión
stroke
2.0
3.0
4.0
140
160
180
200
carlength

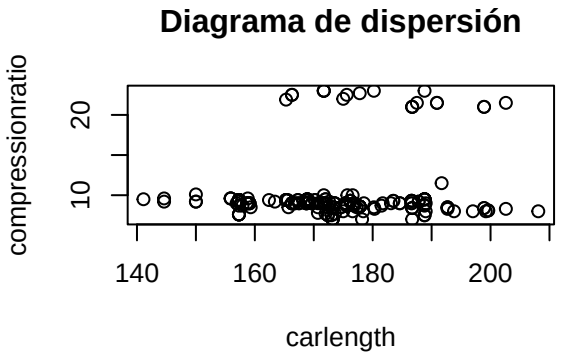


Diagrama de dispersión
compressionratio
10
20
140
160
180
200
carlength

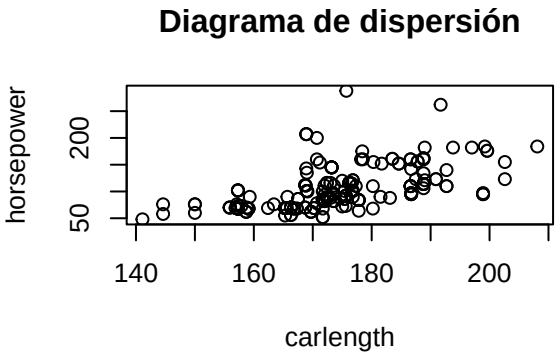


Diagrama de dispersión
horsepower
50
200
140
160
180
200
carlength

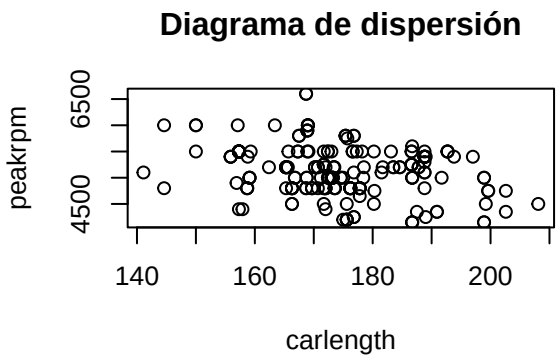


Diagrama de dispersión
peakrpm
4500
6500
140
160
180
200
carlength

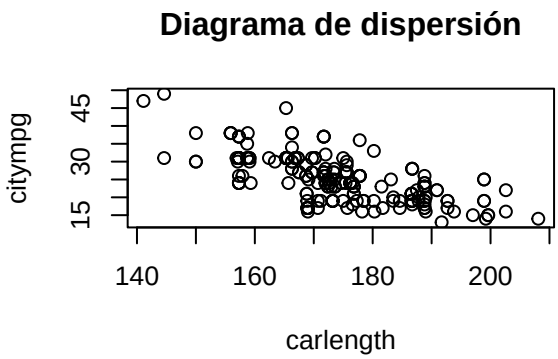


Diagrama de dispersión
citympg
15
30
45
140
160
180
200
carlength

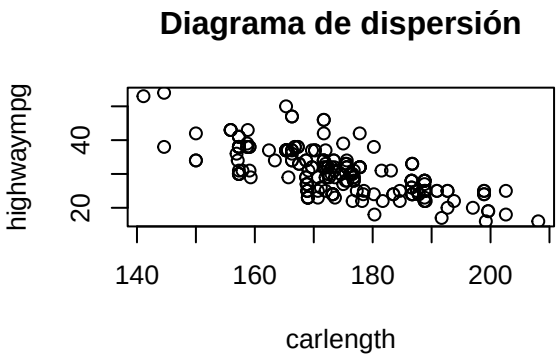


Diagrama de dispersión
highwaympg
20
40
140
160
180
200
carlength

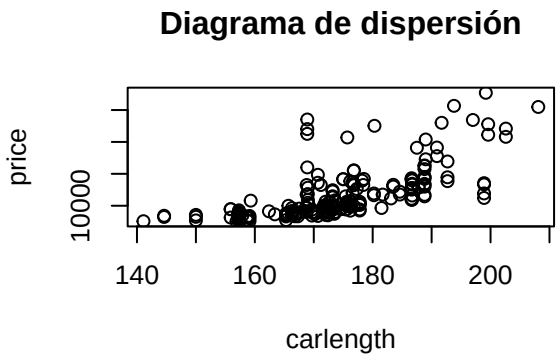


Diagrama de dispersión
price
10000
140
160
180
200
carlength

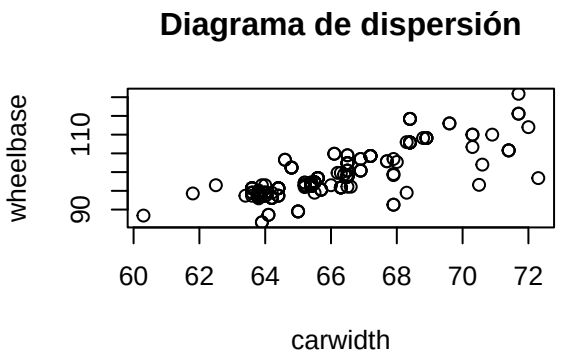


Diagrama de dispersión
wheelbase
90
110
60
62
64
66
68
70
72
carwidth

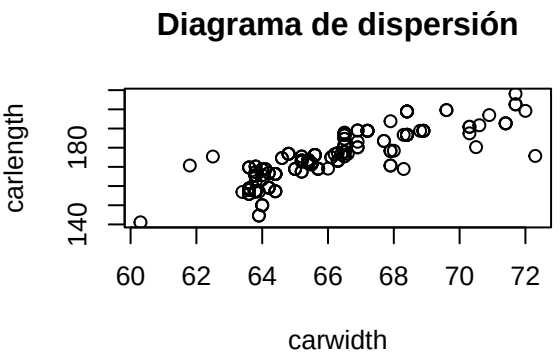


Diagrama de dispersión
carlength
140
180
60
62
64
66
68
70
72
carwidth

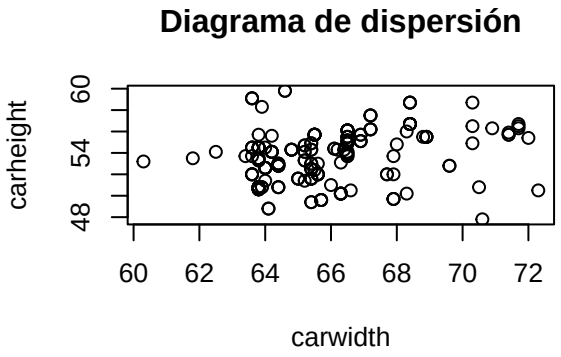


Diagrama de dispersión
carheight
48
54
60
60
62
64
66
68
70
72
carwidth

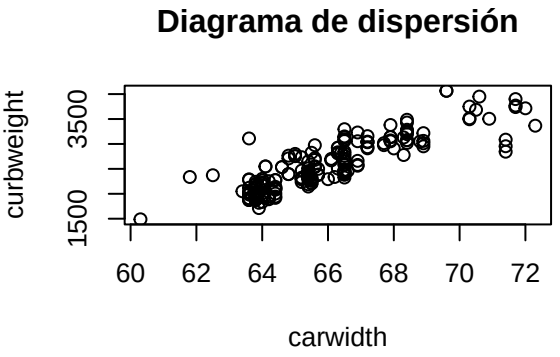


Diagrama de dispersión
curbweight
1500
3500
60
62
64
66
68
70
72
carwidth

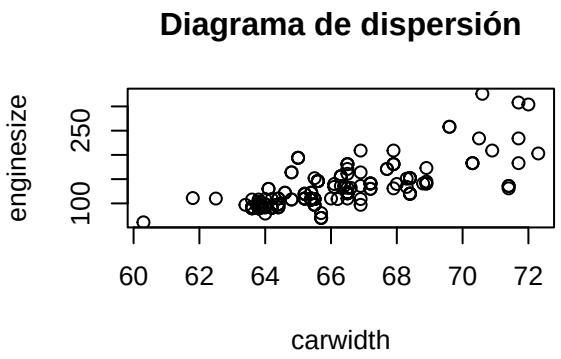


Diagrama de dispersión
enginesize
100
250
60
62
64
66
68
70
72
carwidth

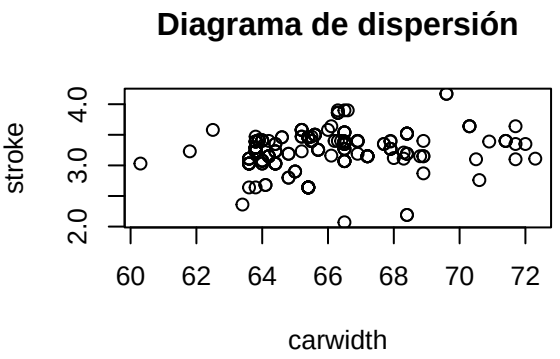


Diagrama de dispersión
stroke
2.0
3.0
4.0
60
62
64
66
68
70
72
carwidth

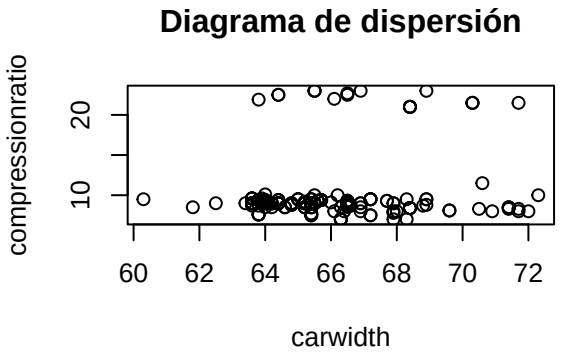


Diagrama de dispersión
compressionratio
10
20
60
62
64
66
68
70
72
carwidth

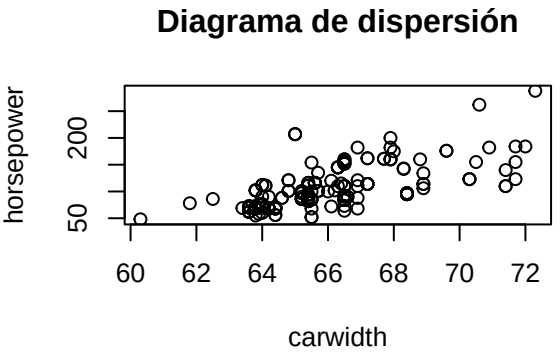


Diagrama de dispersión
horsepower
50
200
60
62
64
66
68
70
72
carwidth

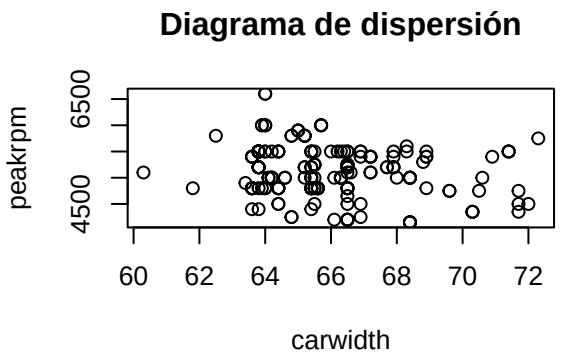


Diagrama de dispersión
peakrpm
4500
6500
60
62
64
66
68
70
72
carwidth

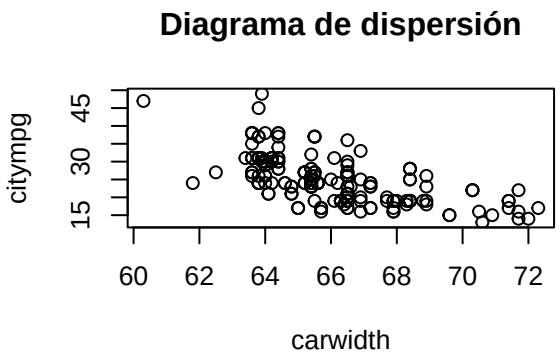


Diagrama de dispersión
citympg
15
30
45
60
62
64
66
68
70
72
carwidth

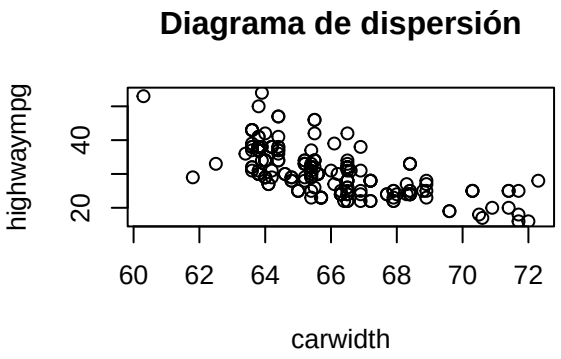


Diagrama de dispersión
highwaympg
20
40
60
62
64
66
68
70
72
carwidth

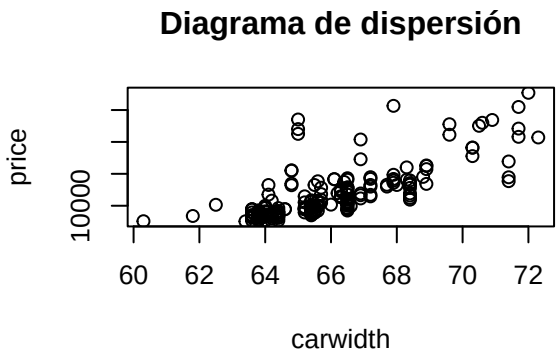


Diagrama de dispersión
price
10000
60
62
64
66
68
70
72
carwidth

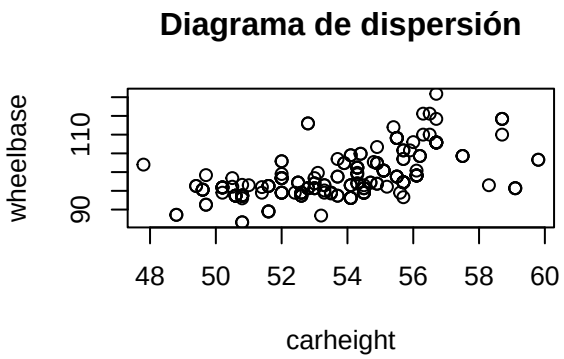


Diagrama de dispersión
wheelbase
90
110
48
50
52
54
56
58
60
carheight

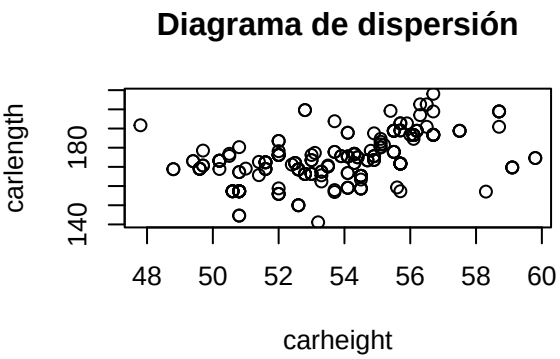


Diagrama de dispersión
carlength
140
180
48
50
52
54
56
58
60
carheight

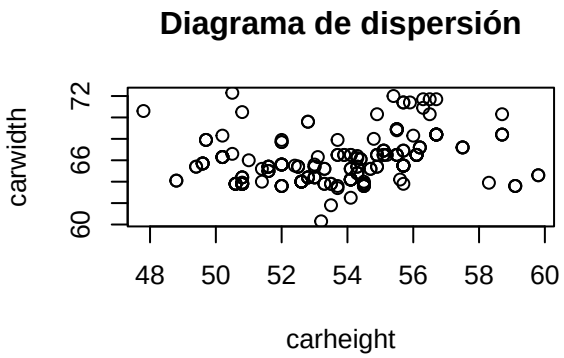


Diagrama de dispersión
carwidth
60
66
72
48
50
52
54
56
58
60
carheight

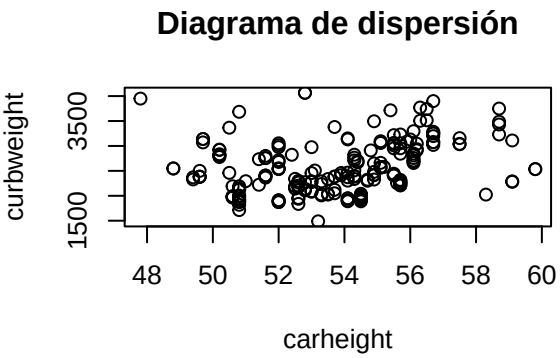


Diagrama de dispersión
curbweight
1500
3500
48
50
52
54
56
58
60
carheight

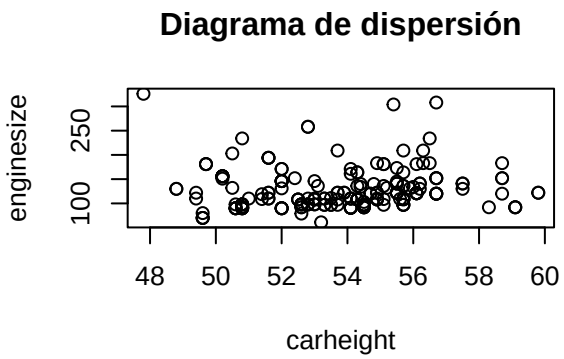


Diagrama de dispersión
enginesize
100
250
48
50
52
54
56
58
60
carheight

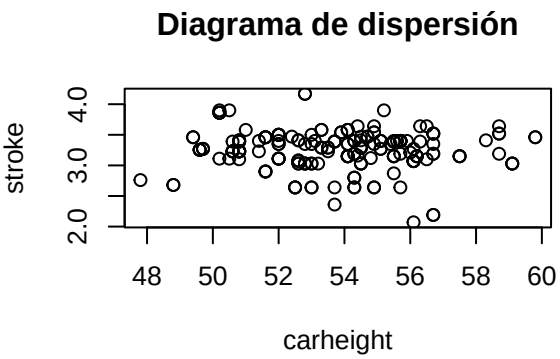


Diagrama de dispersión
stroke
2.0
3.0
4.0
48
50
52
54
56
58
60
carheight

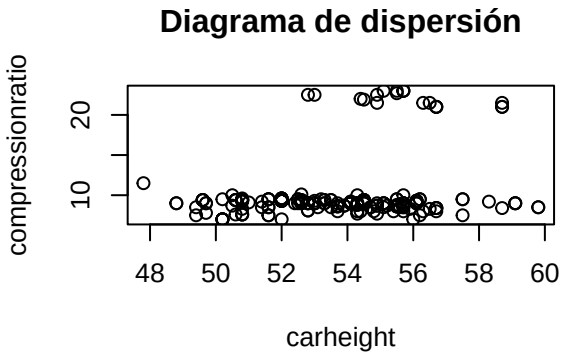


Diagrama de dispersión
compressionratio
10
20
48
50
52
54
56
58
60
carheight

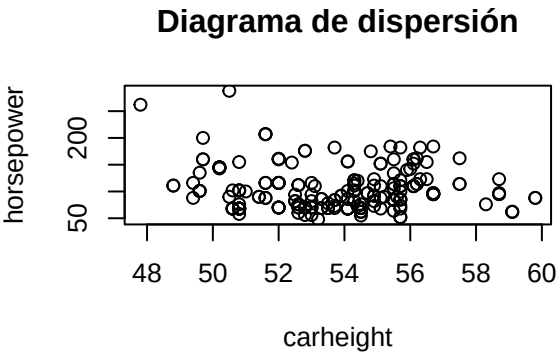


Diagrama de dispersión
horsepower
50
200
48
50
52
54
56
58
60
carheight

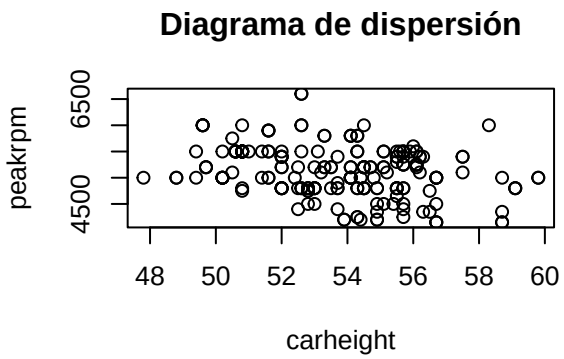


Diagrama de dispersión
peakrpm
4500
6500
48
50
52
54
56
58
60
carheight

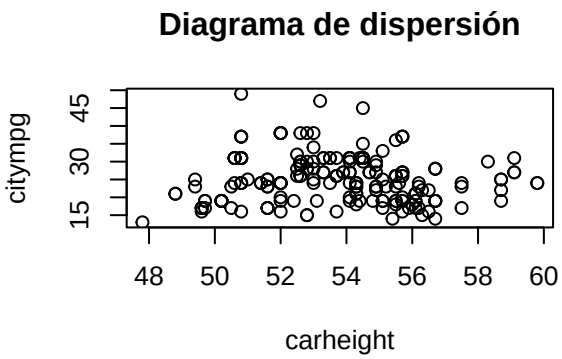


Diagrama de dispersión
citympg
15
30
45
48
50
52
54
56
58
60
carheight

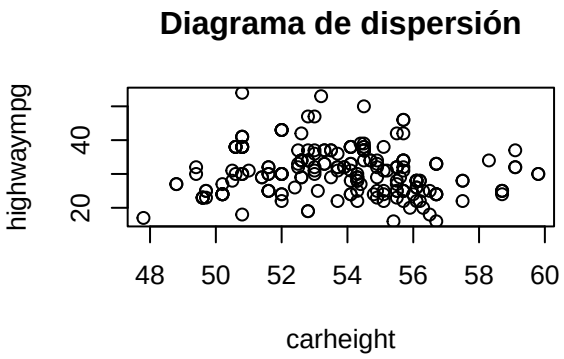


Diagrama de dispersión
highwaympg
20
40
48
50
52
54
56
58
60
carheight

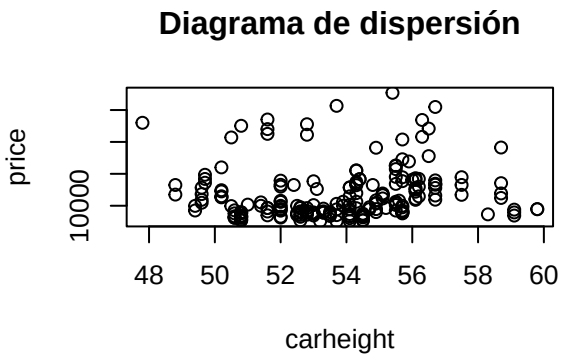


Diagrama de dispersión
price
10000
48
50
52
54
56
58
60
carheight

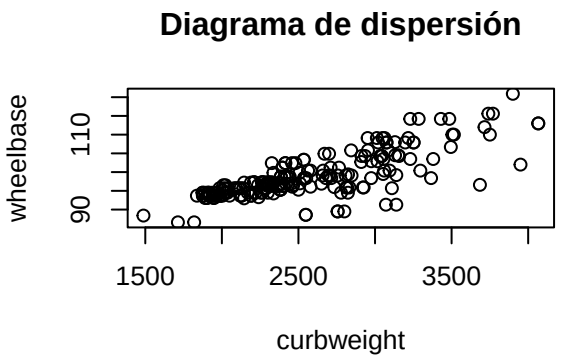


Diagrama de dispersión
wheelbase
90
110
1500
2500
3500
curbweight

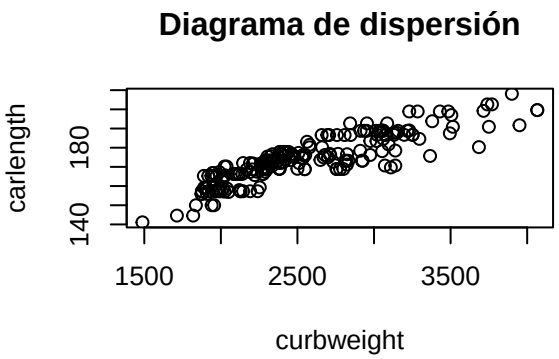


Diagrama de dispersión
carlength
140
180
1500
2500
3500
curbweight

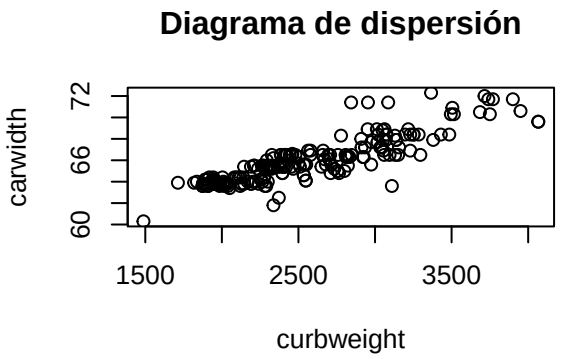


Diagrama de dispersión
carwidth
60
66
72
1500
2500
3500
curbweight

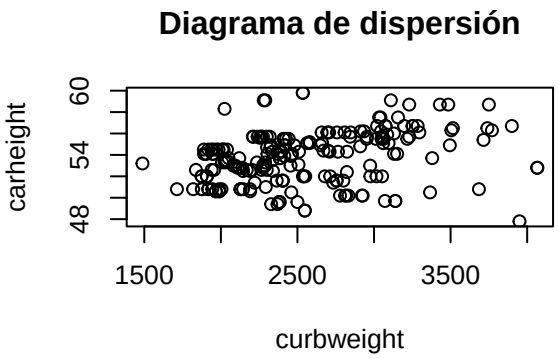


Diagrama de dispersión
carheight
48
54
60
1500
2500
3500
curbweight

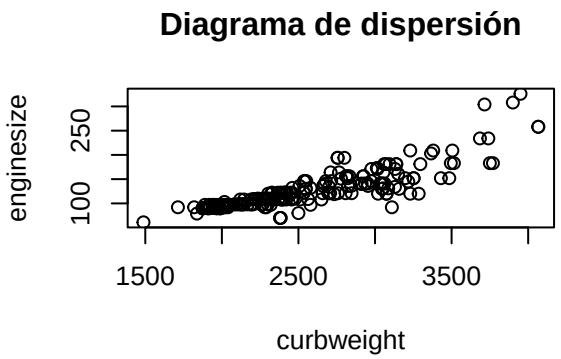


Diagrama de dispersión
enginesize
100
250
1500
2500
3500
curbweight

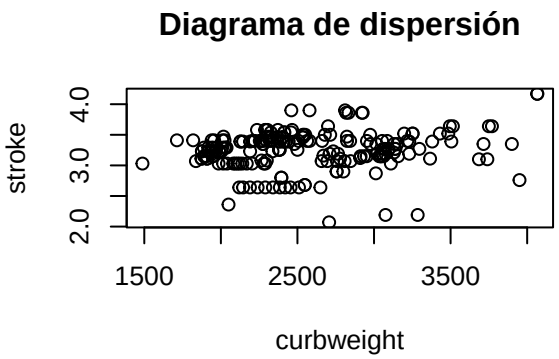


Diagrama de dispersión
stroke
2.0
3.0
4.0
1500
2500
3500
curbweight

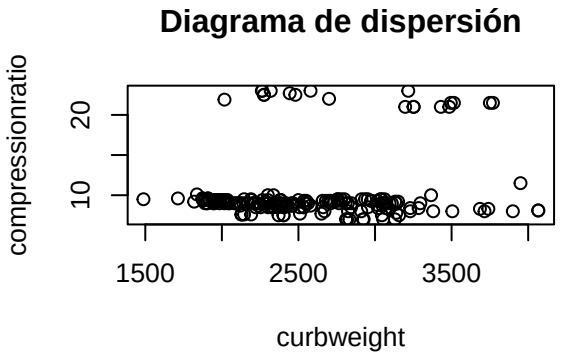


Diagrama de dispersión
compressionratio
10
20
1500
2500
3500
curbweight

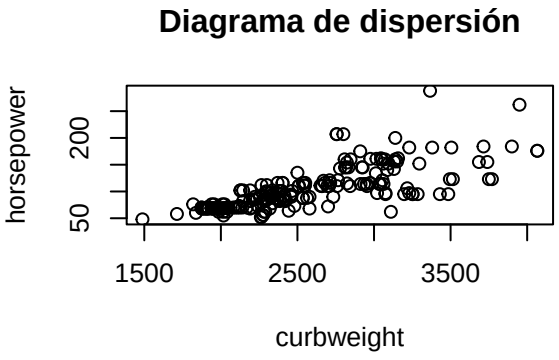


Diagrama de dispersión
horsepower
50
200
1500
2500
3500
curbweight

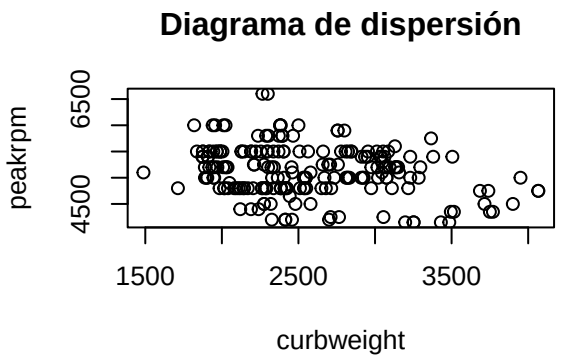


Diagrama de dispersión
peakrpm
4500
6500
1500
2500
3500
curbweight

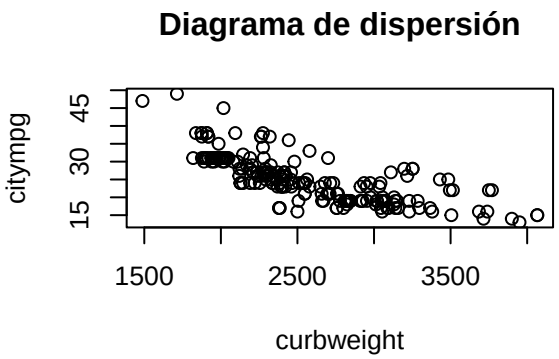


Diagrama de dispersión
citympg
15
30
45
1500
2500
3500
curbweight

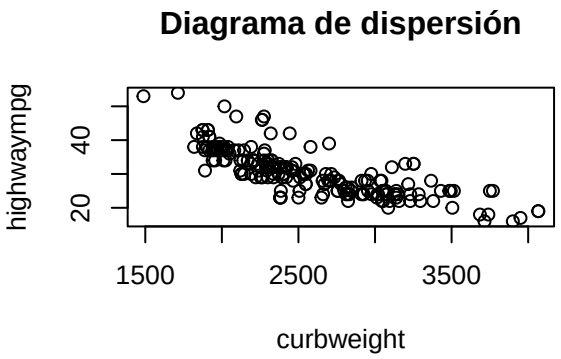


Diagrama de dispersión
highwaympg
20
40
1500
2500
3500
curbweight

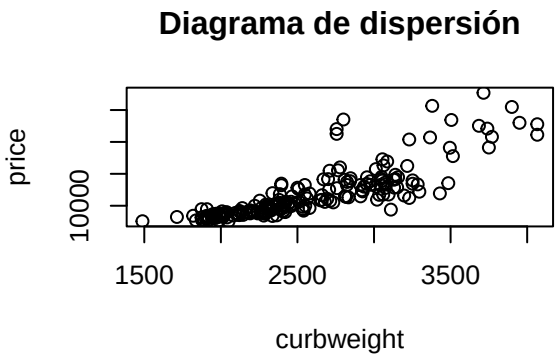


Diagrama de dispersión
price
10000
1500
2500
3500
curbweight

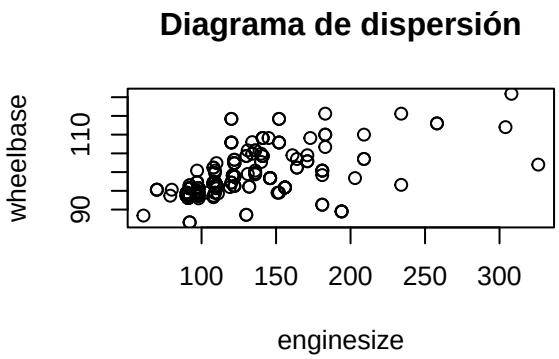


Diagrama de dispersión
wheelbase
90
110
100
150
200
250
300
enginesize

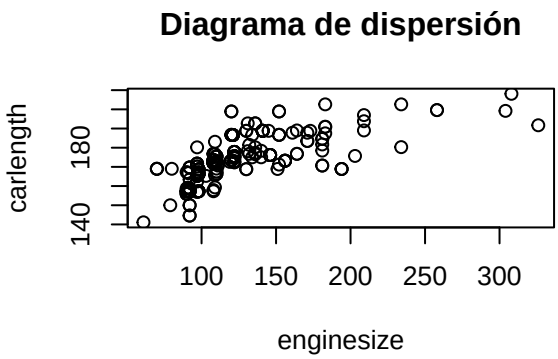


Diagrama de dispersión
carlength
140
180
100
150
200
250
300
enginesize

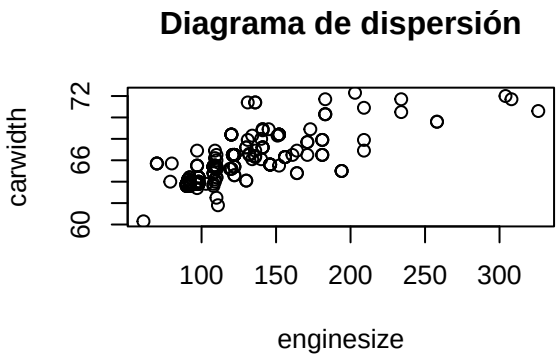


Diagrama de dispersión
carwidth
60
66
72
100
150
200
250
300
enginesize

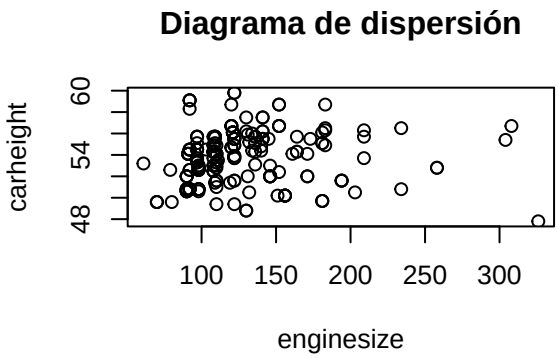


Diagrama de dispersión
carheight
48
54
60
100
150
200
250
300
enginesize

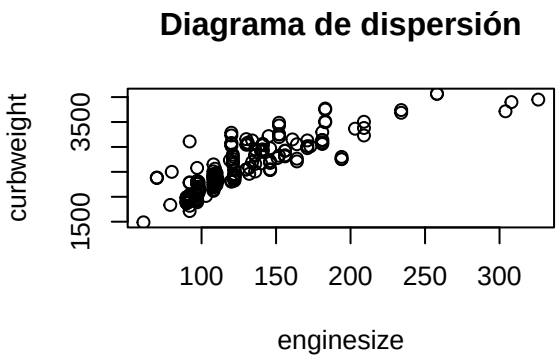


Diagrama de dispersión
curbweight
1500
3500
100
150
200
250
300
enginesize

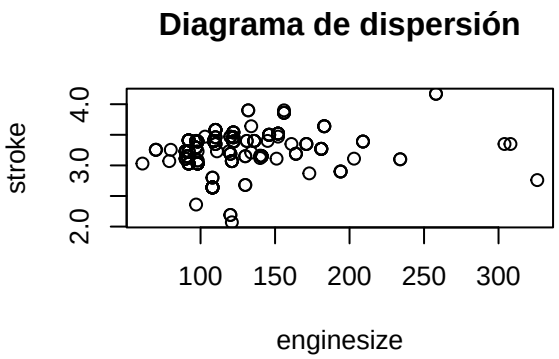


Diagrama de dispersión
stroke
2.0
3.0
4.0
100
150
200
250
300
enginesize

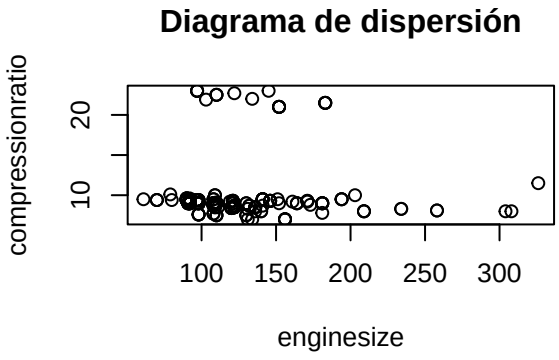


Diagrama de dispersión
compressionratio
10
20
100
150
200
250
300
enginesize

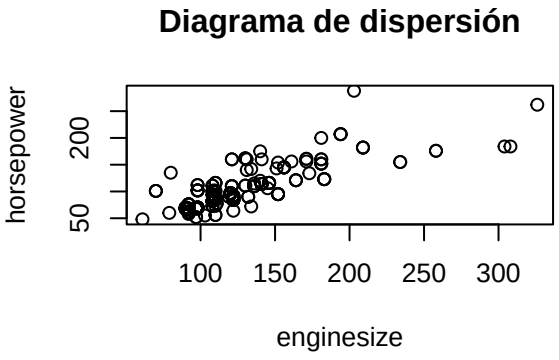


Diagrama de dispersión
horsepower
50
200
100
150
200
250
300
enginesize

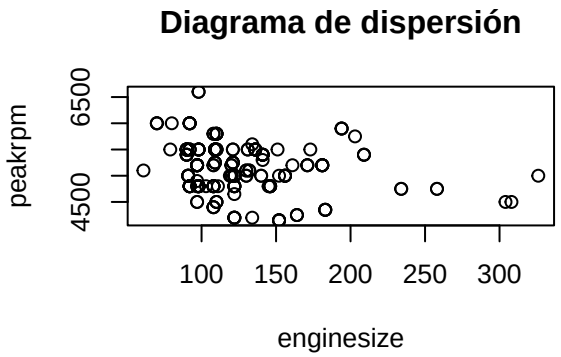


Diagrama de dispersión
peakrpm
4500
6500
100
150
200
250
300
enginesize

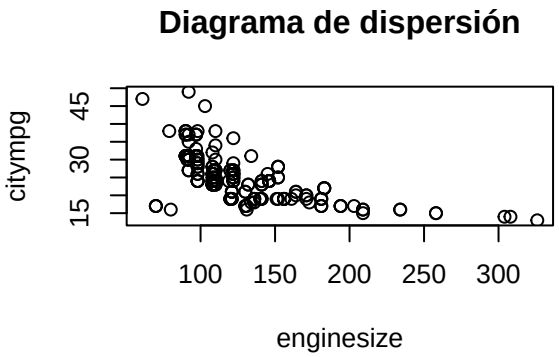


Diagrama de dispersión
citympg
15
30
45
100
150
200
250
300
enginesize

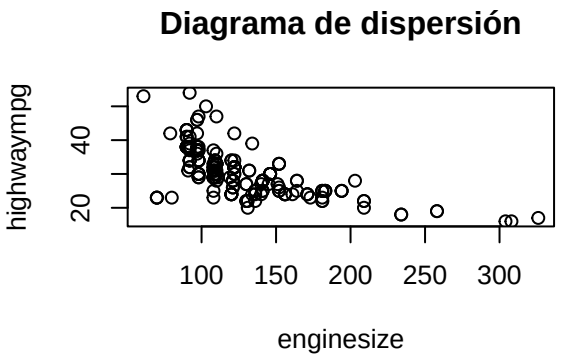


Diagrama de dispersión
highwaympg
20
40
100
150
200
250
300
enginesize

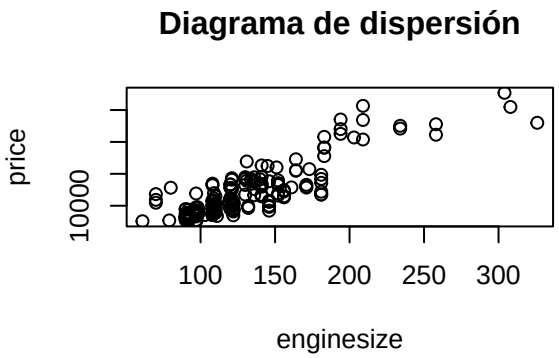


Diagrama de dispersión
price
10000
100
150
200
250
300
enginesize

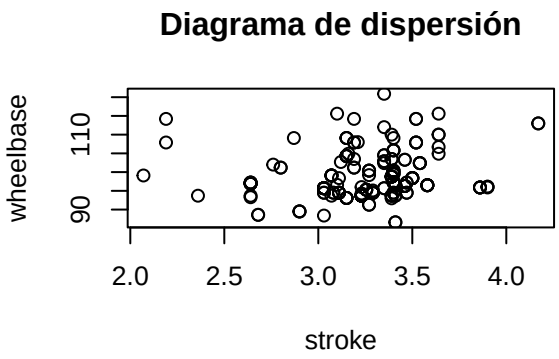


Diagrama de dispersión
wheelbase
90
110
2.0
2.5
3.0
3.5
4.0
stroke

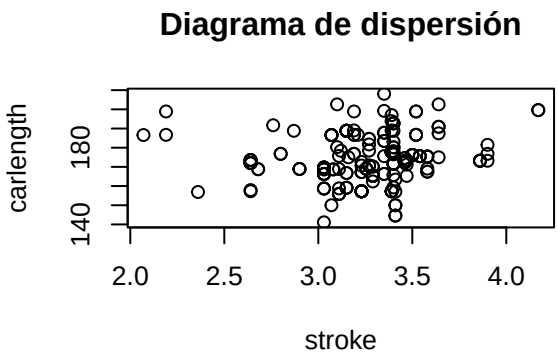


Diagrama de dispersión
carlength
140
180
2.0
2.5
3.0
3.5
4.0
stroke

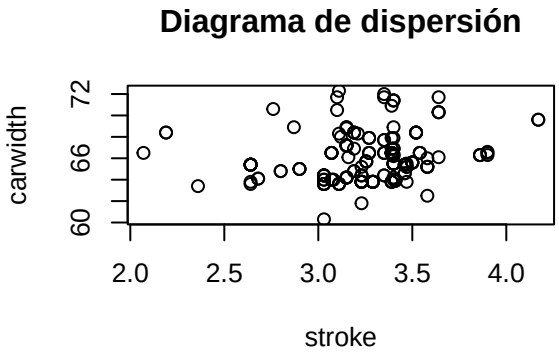


Diagrama de dispersión
carwidth
60
66
72
2.0
2.5
3.0
3.5
4.0
stroke

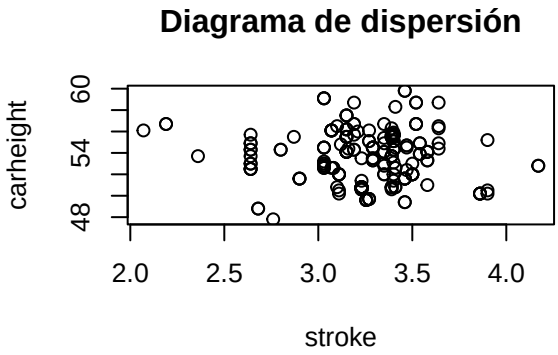


Diagrama de dispersión
carheight
48
54
60
2.0
2.5
3.0
3.5
4.0
stroke

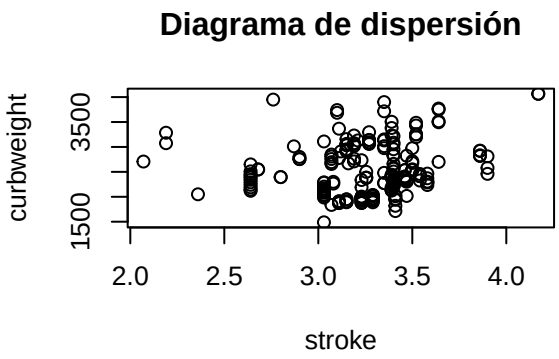


Diagrama de dispersión
curbweight
1500
3500
2.0
2.5
3.0
3.5
4.0
stroke

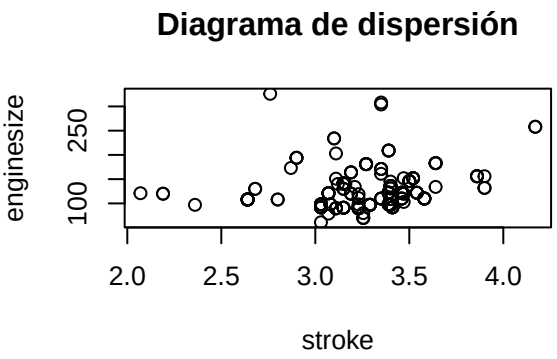


Diagrama de dispersión
enginesize
100
250
2.0
2.5
3.0
3.5
4.0
stroke

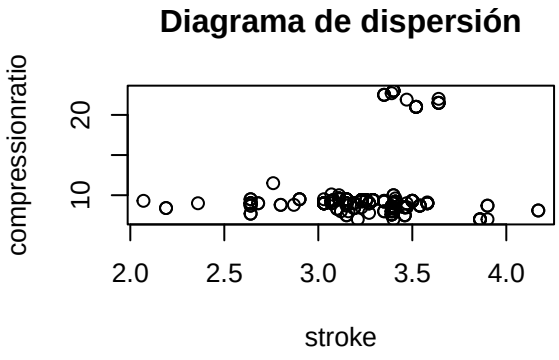


Diagrama de dispersión
compressionratio
10
20
2.0
2.5
3.0
3.5
4.0
stroke

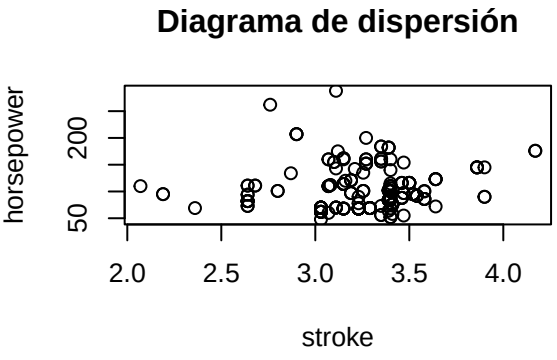


Diagrama de dispersión
horsepower
50
200
2.0
2.5
3.0
3.5
4.0
stroke

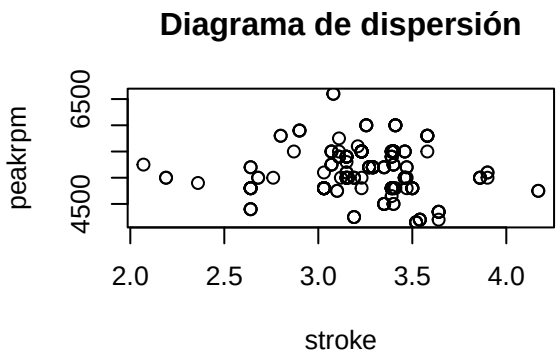


Diagrama de dispersión
peakrpm
4500
6500
2.0
2.5
3.0
3.5
4.0
stroke

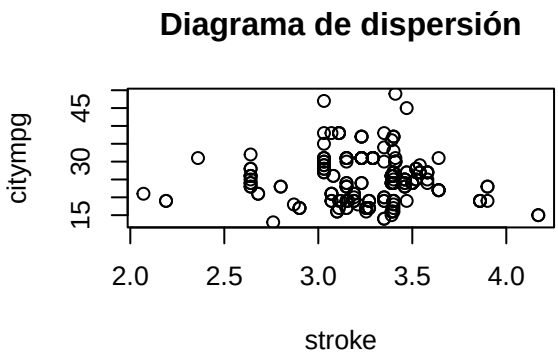


Diagrama de dispersión
citympg
15
30
45
2.0
2.5
3.0
3.5
4.0
stroke

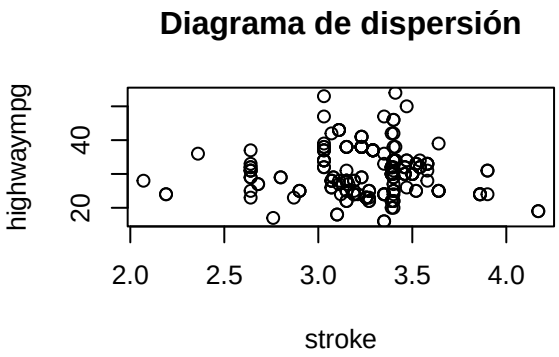


Diagrama de dispersión
highwaympg
20
40
2.0
2.5
3.0
3.5
4.0
stroke

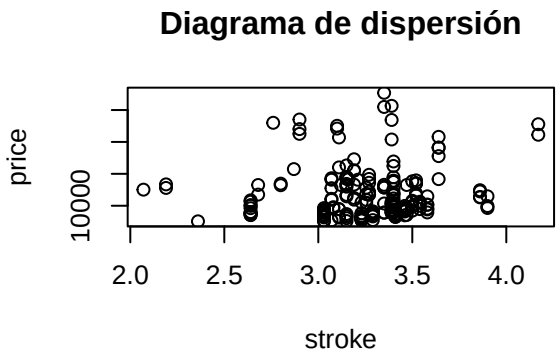


Diagrama de dispersión
price
10000
2.0
2.5
3.0
3.5
4.0
stroke

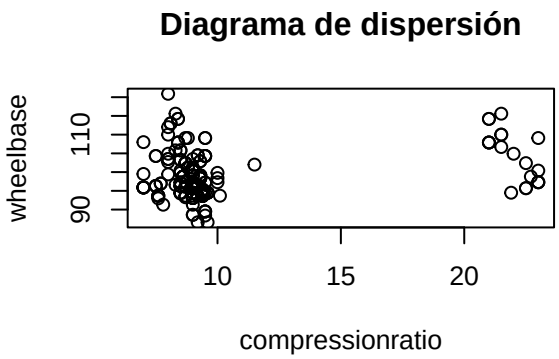


Diagrama de dispersión
wheelbase
90
110
10
15
20
compressionratio

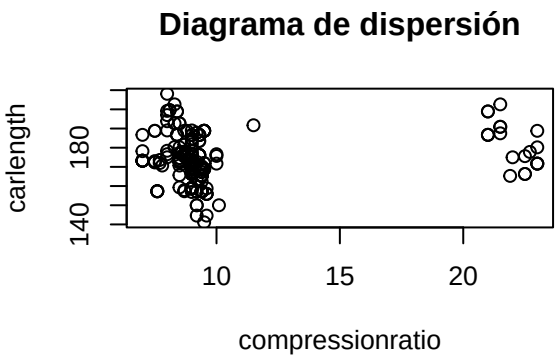


Diagrama de dispersión
carlength
140
180
10
15
20
compressionratio

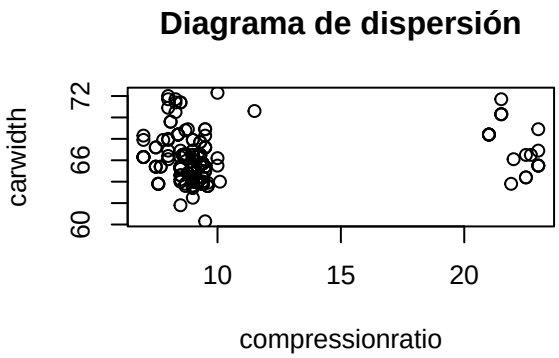


Diagrama de dispersión
carwidth
60
66
72
10
15
20
compressionratio

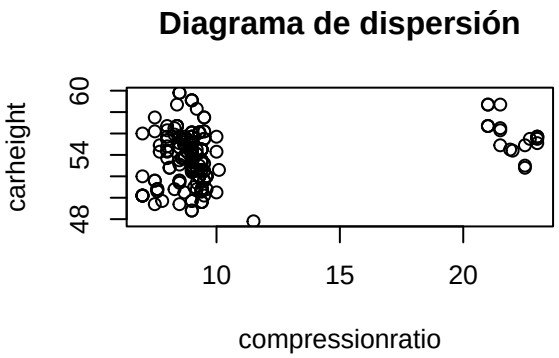


Diagrama de dispersión
carheight
48
54
60
10
15
20
compressionratio

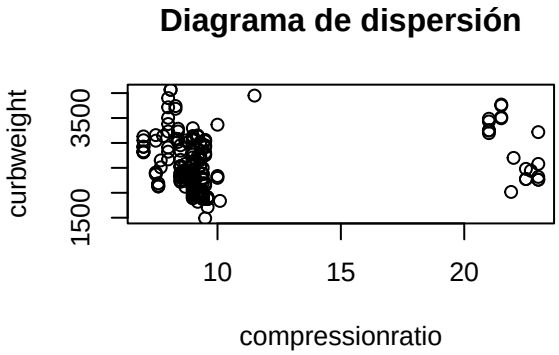


Diagrama de dispersión
curbweight
1500
3500
10
15
20
compressionratio

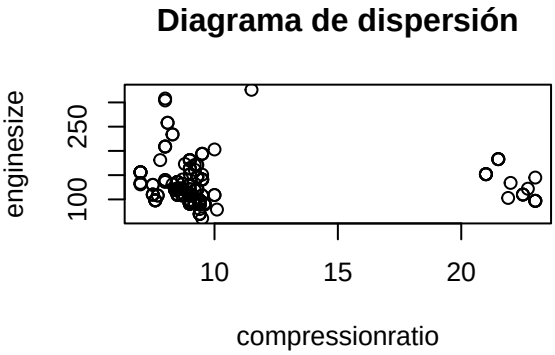


Diagrama de dispersión
enginesize
100
250
10
15
20
compressionratio

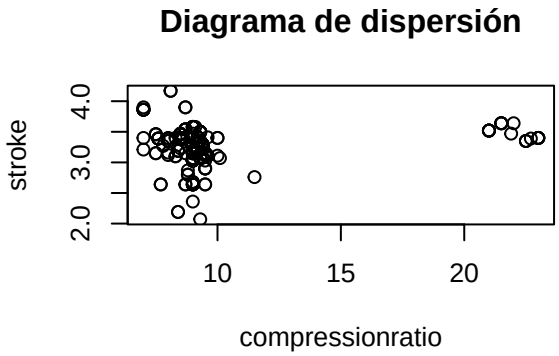


Diagrama de dispersión
stroke
2.0
3.0
4.0
10
15
20
compressionratio

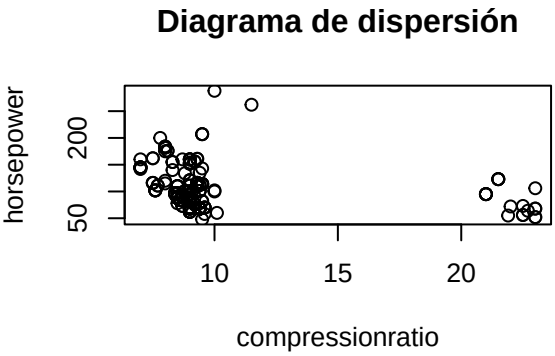


Diagrama de dispersión
horsepower
50
200
10
15
20
compressionratio

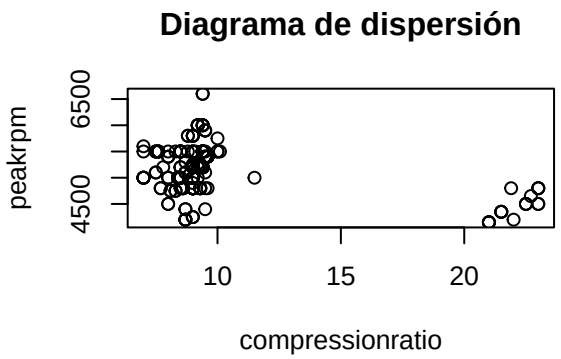


Diagrama de dispersión
peakrpm
4500
6500
10
15
20
compressionratio

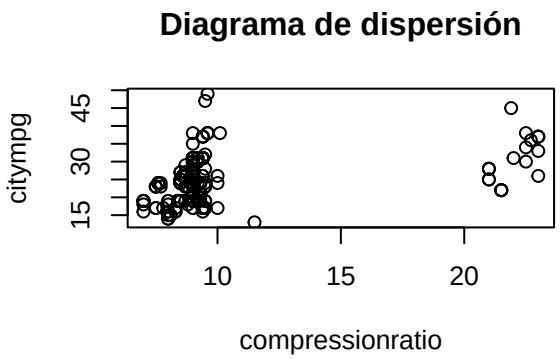


Diagrama de dispersión
citympg
15
30
45
10
15
20
compressionratio

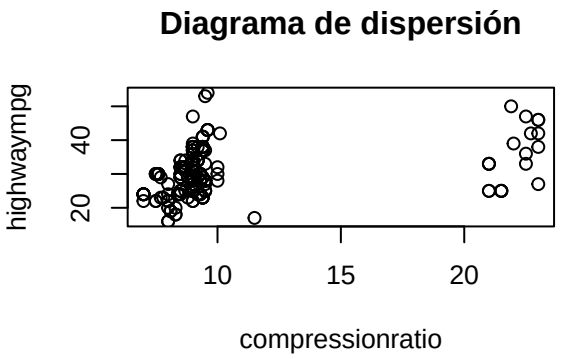


Diagrama de dispersión
highwaympg
20
40
10
15
20
compressionratio

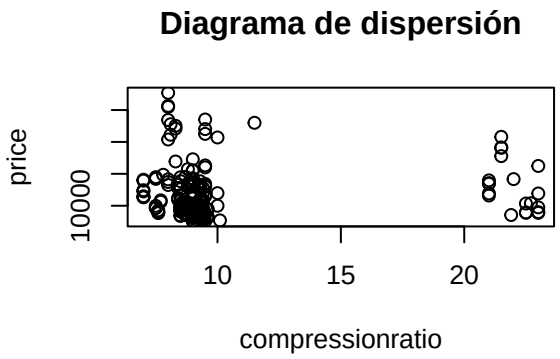


Diagrama de dispersión
price
10000
10
15
20
compressionratio

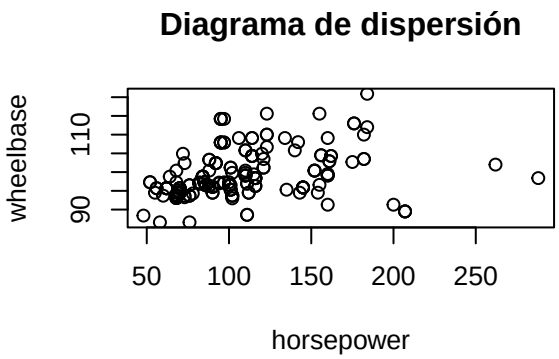


Diagrama de dispersión
wheelbase
90
110
50
100
150
200
250
horsepower

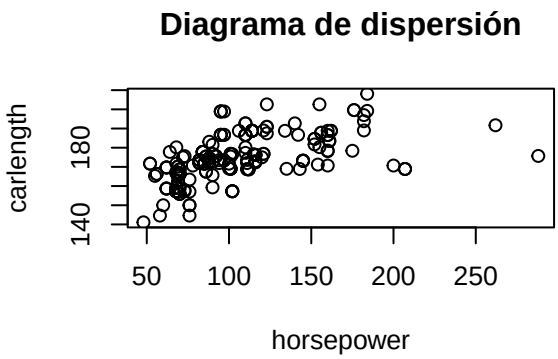


Diagrama de dispersión
carlength
140
180
50
100
150
200
250
horsepower

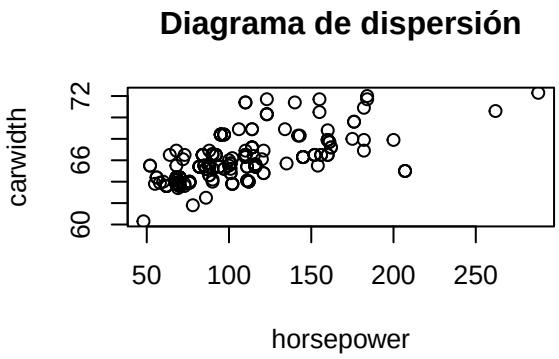


Diagrama de dispersión
carwidth
60
66
72
50
100
150
200
250
horsepower

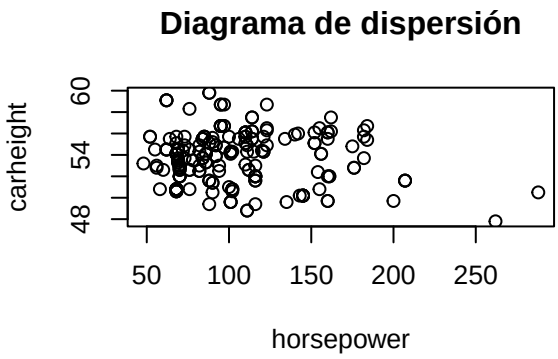


Diagrama de dispersión
carheight
48
54
60
50
100
150
200
250
horsepower

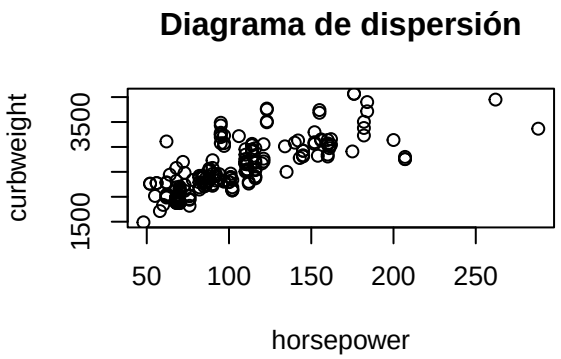


Diagrama de dispersión
curbweight
1500
3500
50
100
150
200
250
horsepower

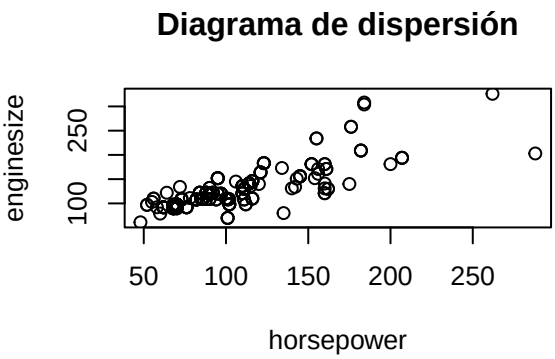


Diagrama de dispersión
enginesize
100
250
50
100
150
200
250
horsepower

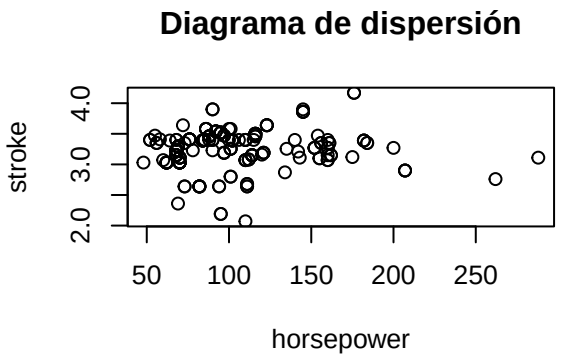


Diagrama de dispersión
stroke
2.0
3.0
4.0
50
100
150
200
250
horsepower

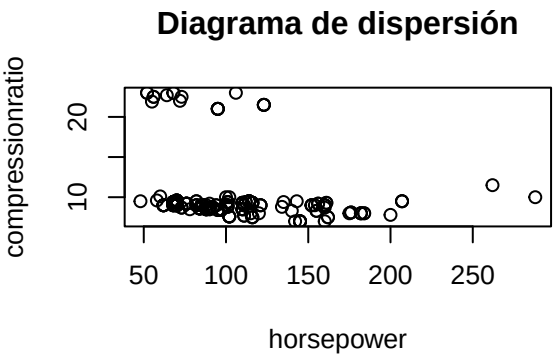


Diagrama de dispersión
compressionratio
10
20
50
100
150
200
250
horsepower

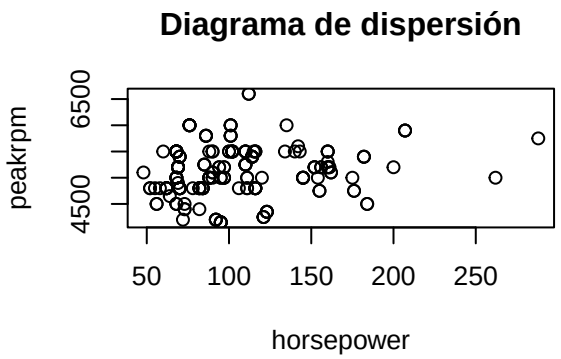


Diagrama de dispersión
peakrpm
4500
6500
50
100
150
200
250
horsepower

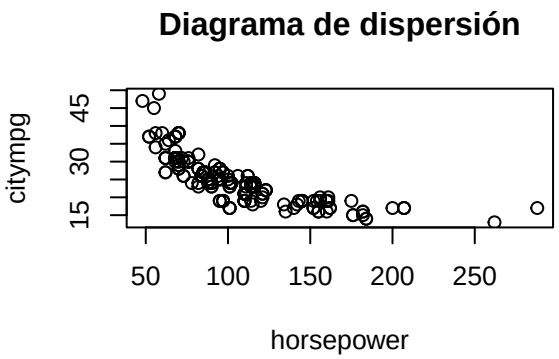


Diagrama de dispersión
citympg
15
30
45
50
100
150
200
250
horsepower

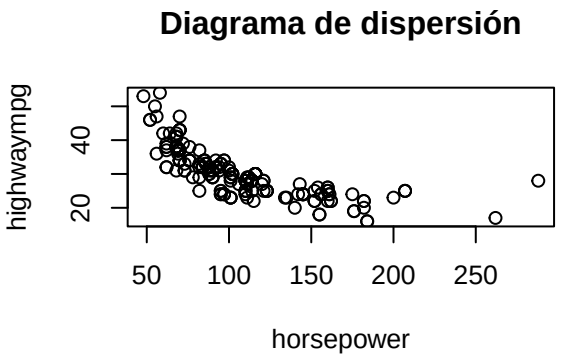


Diagrama de dispersión
highwaympg
20
40
50
100
150
200
250
horsepower

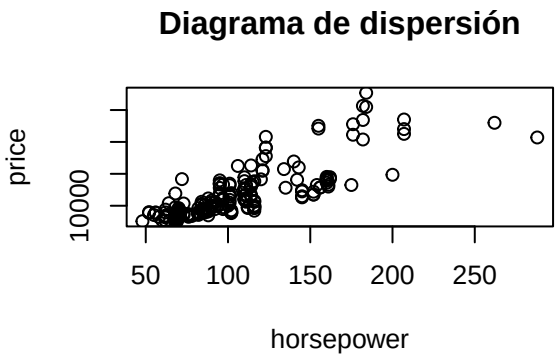


Diagrama de dispersión
price
10000
50
100
150
200
250
horsepower

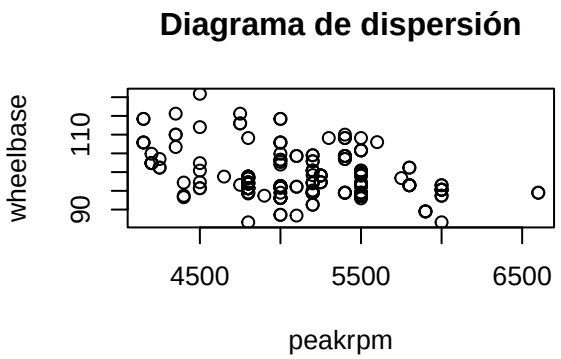


Diagrama de dispersión
wheelbase
90
110
4500
5500
6500
peakrpm

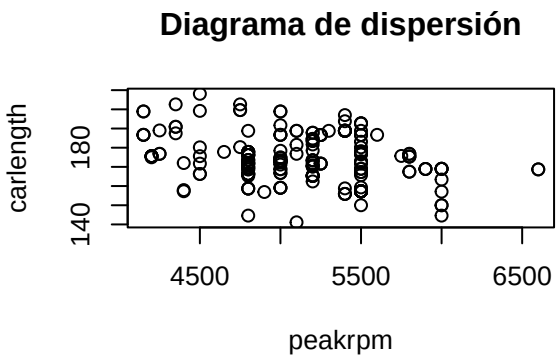


Diagrama de dispersión
carlength
140
180
4500
5500
6500
peakrpm

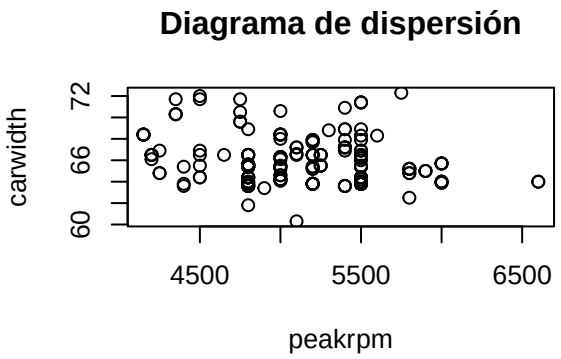


Diagrama de dispersión
carwidth
60
66
72
4500
5500
6500
peakrpm

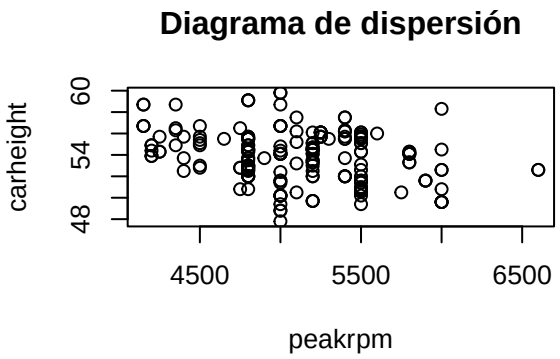


Diagrama de dispersión
carheight
48
54
60
4500
5500
6500
peakrpm

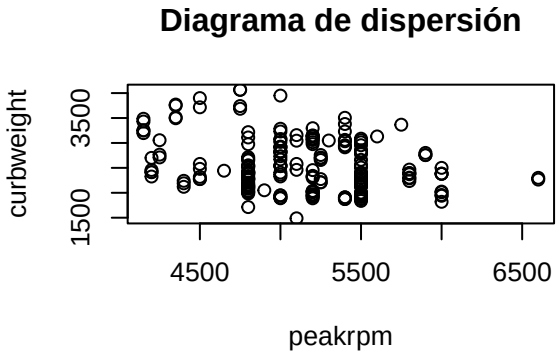


Diagrama de dispersión
curbweight
1500
3500
4500
5500
6500
peakrpm

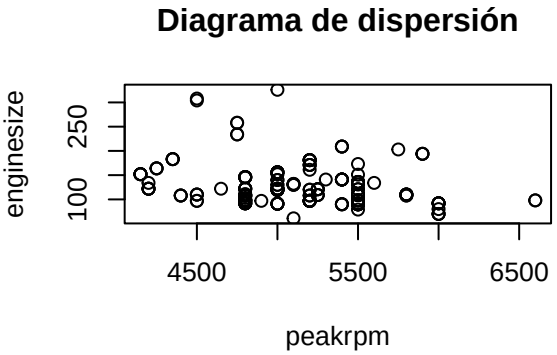


Diagrama de dispersión
enginesize
100
250
4500
5500
6500
peakrpm

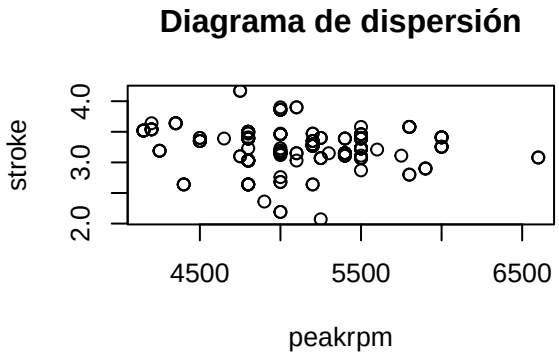


Diagrama de dispersión
stroke
2.0
3.0
4.0
4500
5500
6500
peakrpm

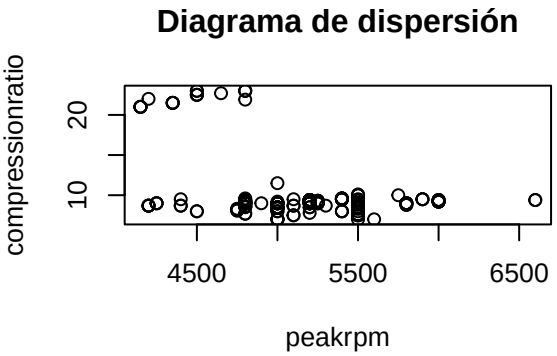


Diagrama de dispersión
compressionratio
10
20
4500
5500
6500
peakrpm

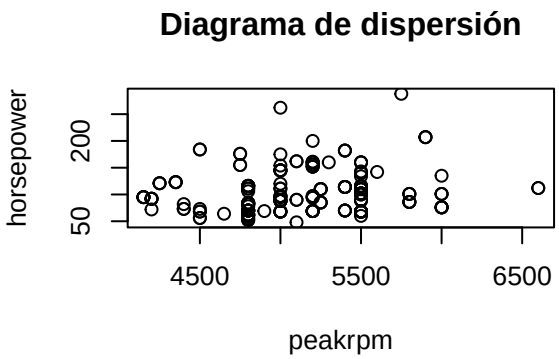


Diagrama de dispersión
horsepower
50
200
4500
5500
6500
peakrpm

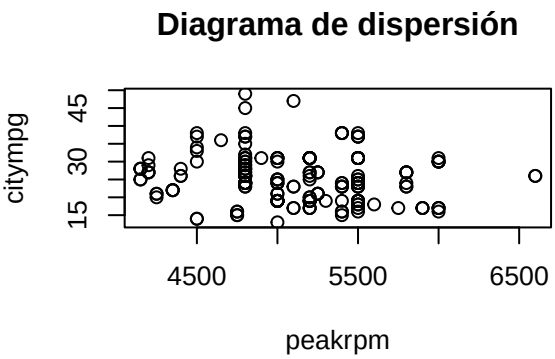


Diagrama de dispersión
citympg
15
30
45
4500
5500
6500
peakrpm

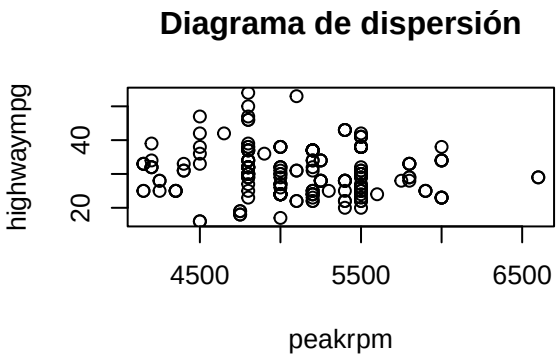


Diagrama de dispersión
highwaympg
20
40
4500
5500
6500
peakrpm

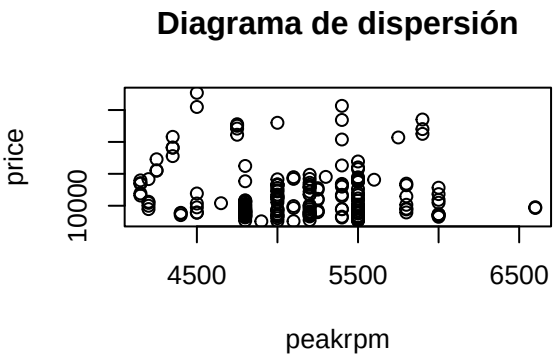


Diagrama de dispersión
price
10000
4500
5500
6500
peakrpm

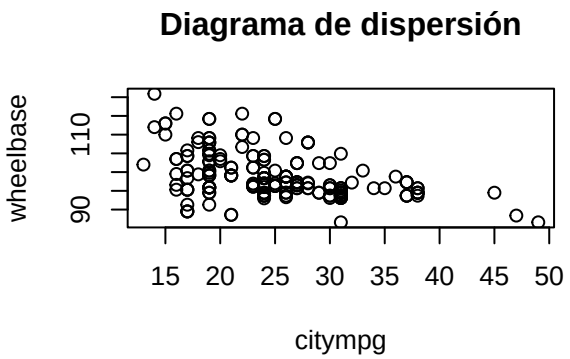


Diagrama de dispersión
wheelbase
90
110
15
20
25
30
35
40
45
50
citympg

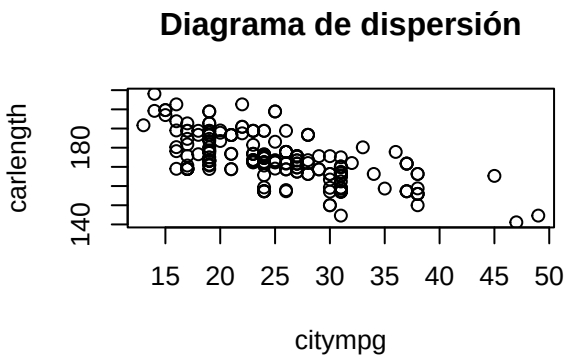


Diagrama de dispersión
carlength
140
180
15
20
25
30
35
40
45
50
citympg

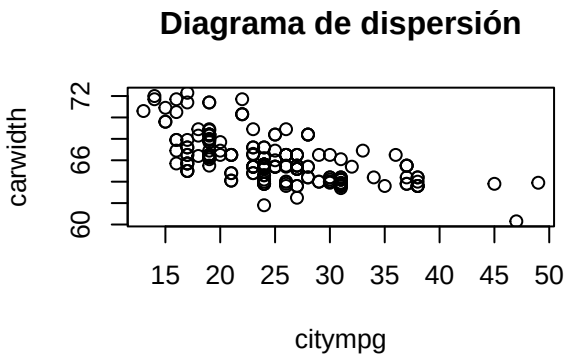


Diagrama de dispersión
carwidth
60
66
72
15
20
25
30
35
40
45
50
citympg

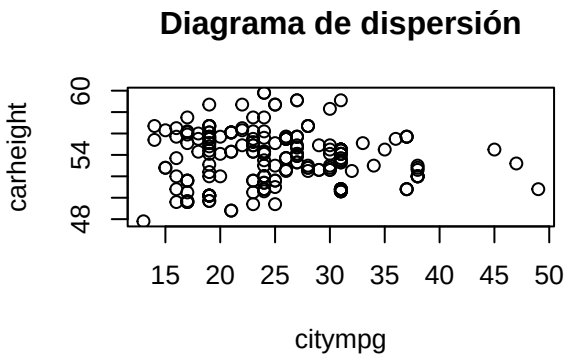


Diagrama de dispersión
carheight
48
54
60
15
20
25
30
35
40
45
50
citympg

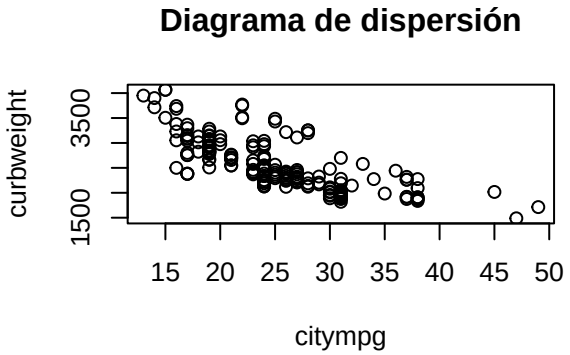


Diagrama de dispersión
curbweight
1500
3500
15 20 25 30 35 40 45 50
citympg

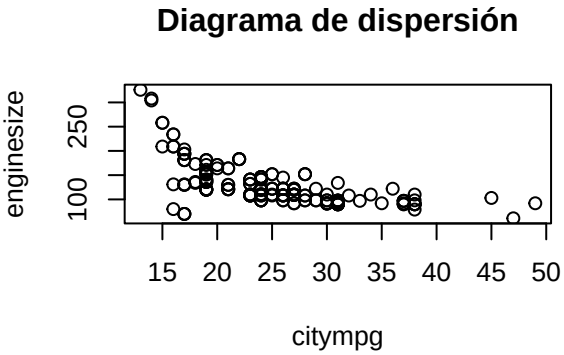


Diagrama de dispersión
enginesize
100
250
15 20 25 30 35 40 45 50
citympg

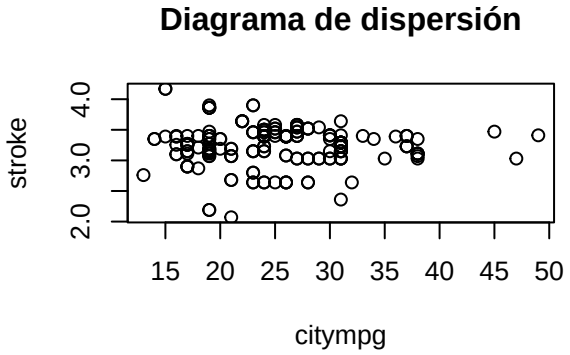


Diagrama de dispersión
stroke
2.0
3.0
4.0
15 20 25 30 35 40 45 50
citympg

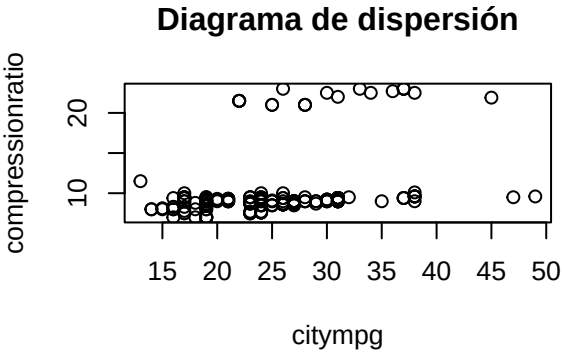


Diagrama de dispersión
compressionratio
10
20
15 20 25 30 35 40 45 50
citympg

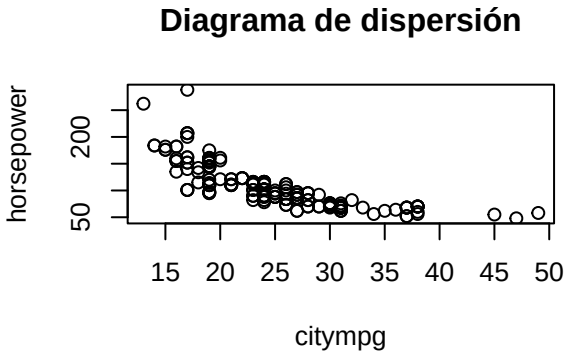


Diagrama de dispersión
horsepower
50
200
15 20 25 30 35 40 45 50
citympg

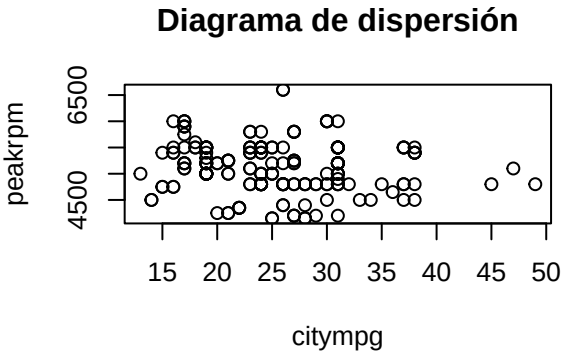


Diagrama de dispersión
peakrpm
4500
6500
15 20 25 30 35 40 45 50
citympg

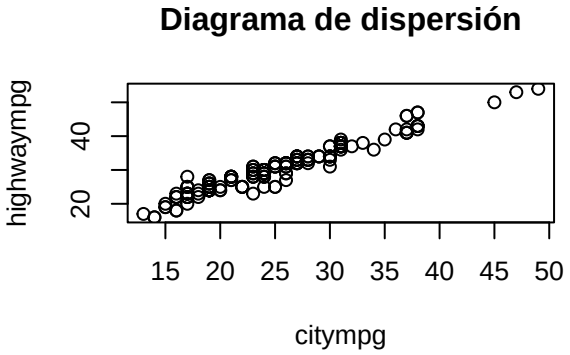


Diagrama de dispersión
highwaympg
20
40
15 20 25 30 35 40 45 50
citympg

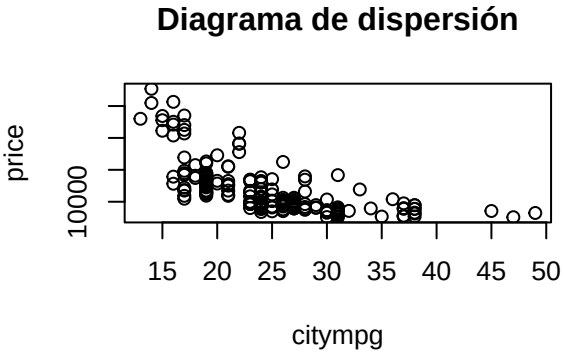


Diagrama de dispersión
price
10000
15 20 25 30 35 40 45 50
citympg

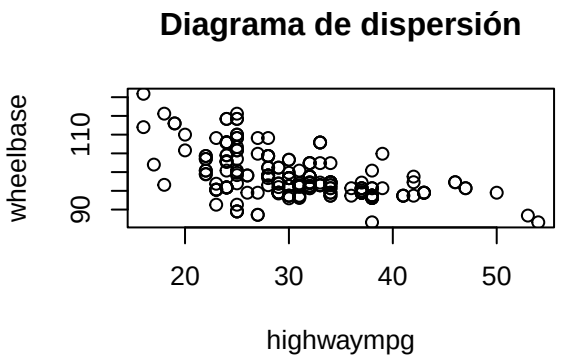


Diagrama de dispersión
wheelbase
90
110
20
30
40
50
highwaympg

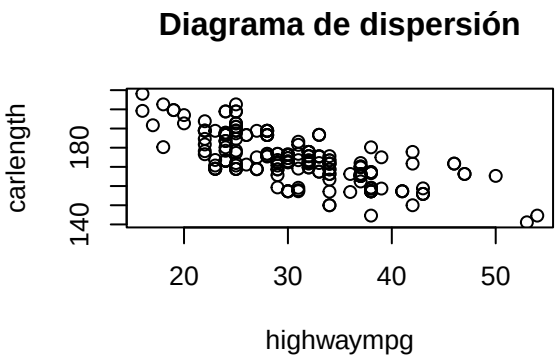


Diagrama de dispersión
carlength
140
180
20
30
40
50
highwaympg

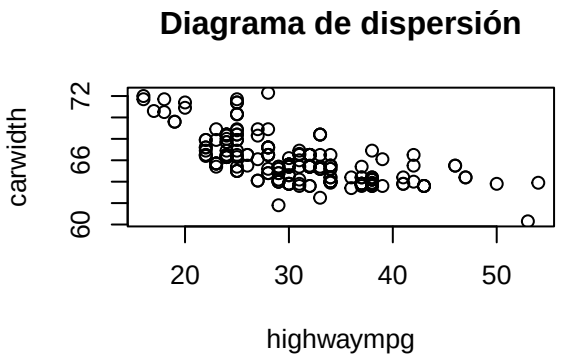


Diagrama de dispersión
carwidth
60
66
72
20
30
40
50
highwaympg

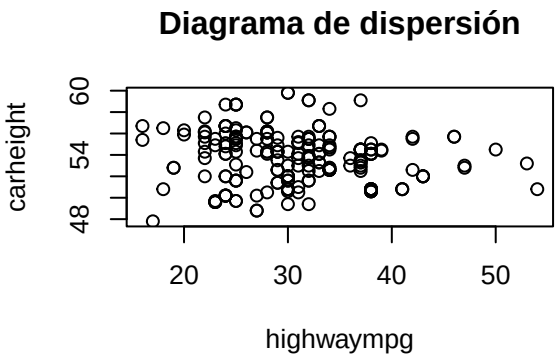


Diagrama de dispersión
carheight
48
54
60
20
30
40
50
highwaympg

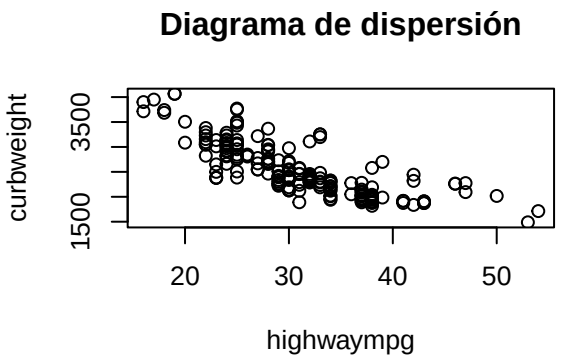


Diagrama de dispersión
curbweight
1500
3500
20
30
40
50
highwaympg

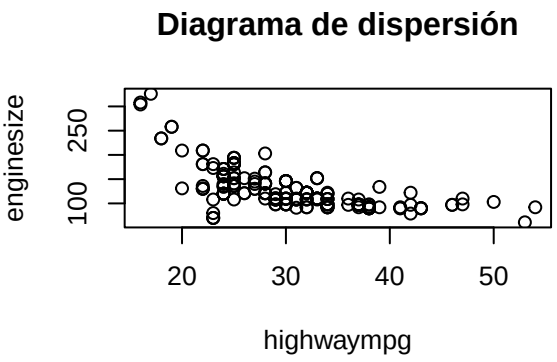


Diagrama de dispersión
enginesize
100
250
20
30
40
50
highwaympg

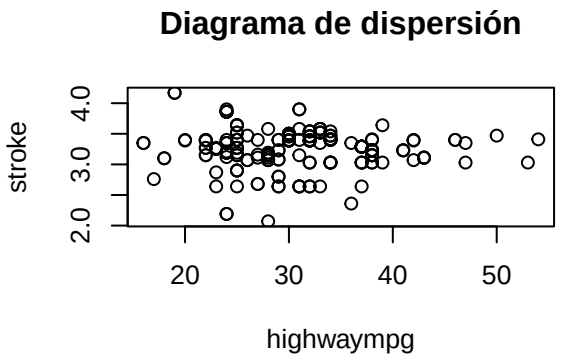


Diagrama de dispersión
stroke
2.0
3.0
4.0
20
30
40
50
highwaympg

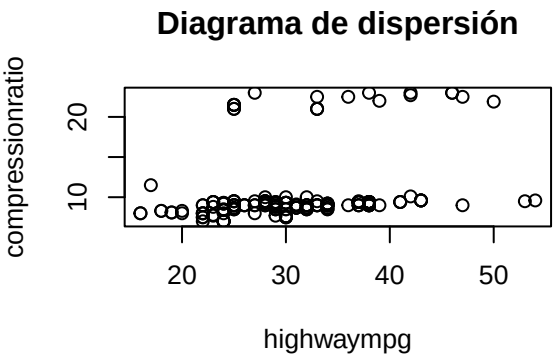


Diagrama de dispersión
compressionratio
10
20
20
30
40
50
highwaympg

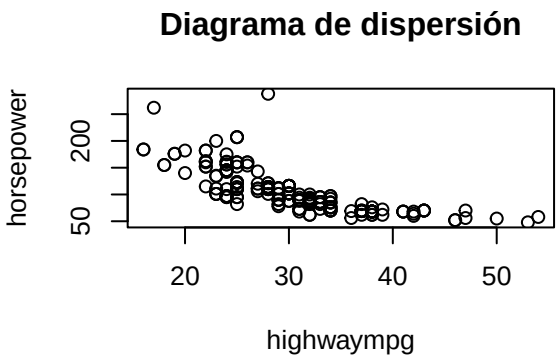


Diagrama de dispersión
horsepower
50
200
20
30
40
50
highwaympg

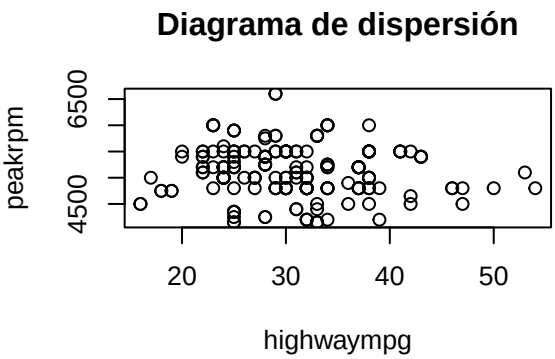


Diagrama de dispersión
peakrpm
4500
6500
20
30
40
50
highwaympg

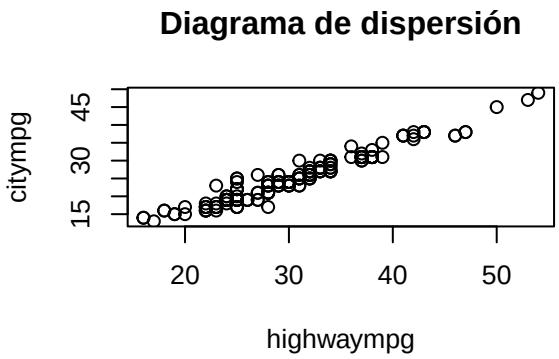


Diagrama de dispersión
citympg
15
30
45
20
30
40
50
highwaympg

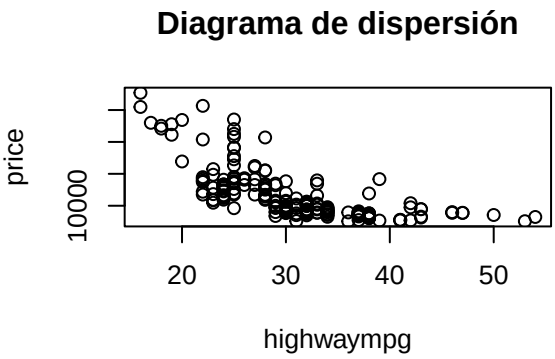


Diagrama de dispersión
price
10000
20
30
40
50
highwaympg

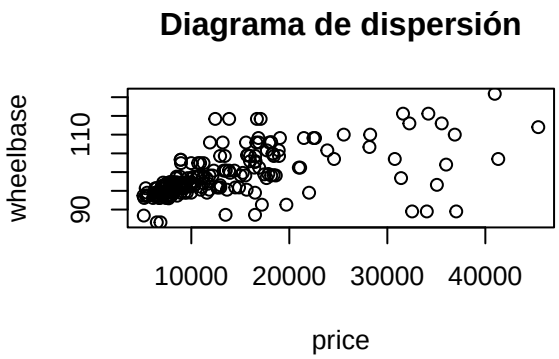


Diagrama de dispersión
wheelbase
90
110
10000
20000
30000
40000
price

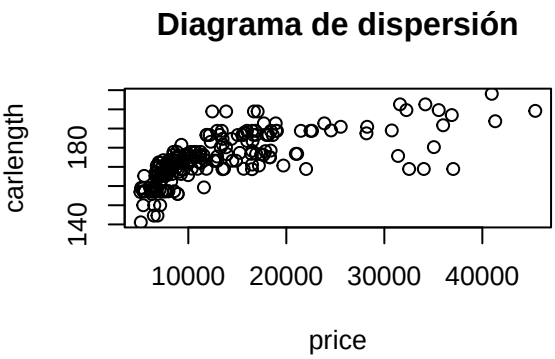


Diagrama de dispersión
carlength
140
180
10000
20000
30000
40000
price

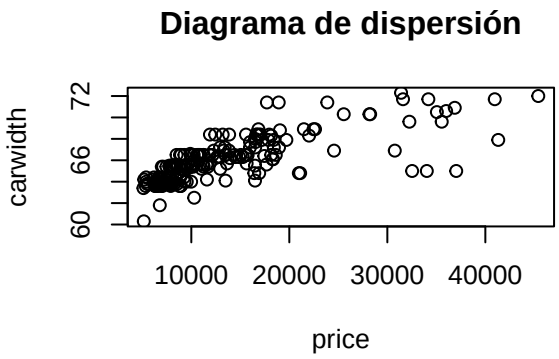


Diagrama de dispersión
carwidth
60
66
72
10000
20000
30000
40000
price

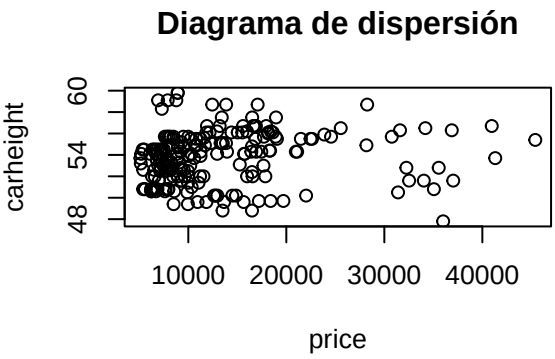


Diagrama de dispersión
carheight
48
54
60
10000
20000
30000
40000
price

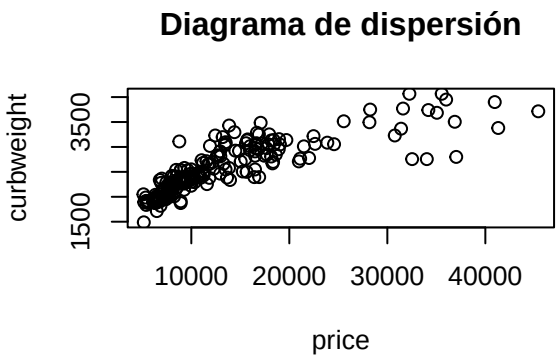


Diagrama de dispersión
curbweight
1500
3500
10000
20000
30000
40000
price

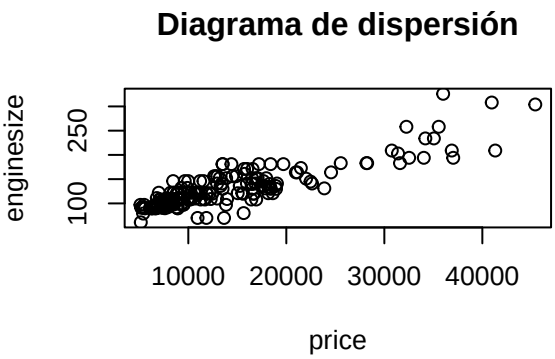


Diagrama de dispersión
enginesize
100
250
10000
20000
30000
40000
price

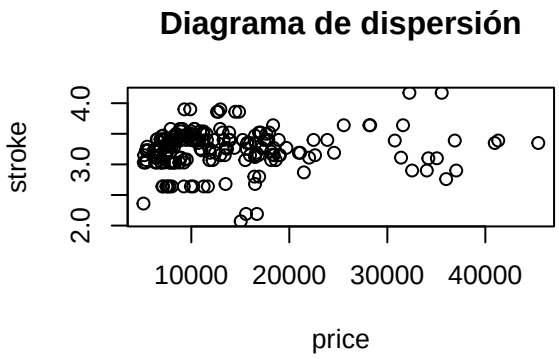


Diagrama de dispersión
stroke
2.0
3.0
4.0
10000
20000
30000
40000
price

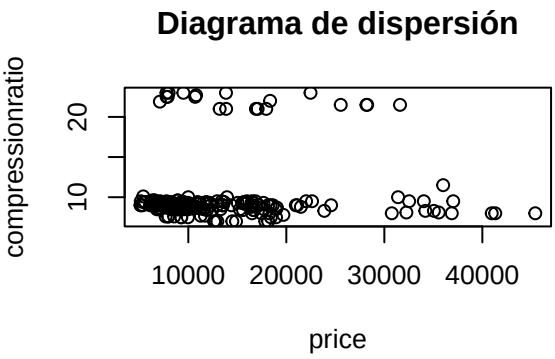
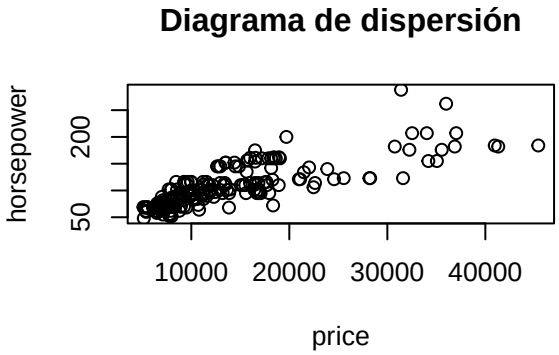
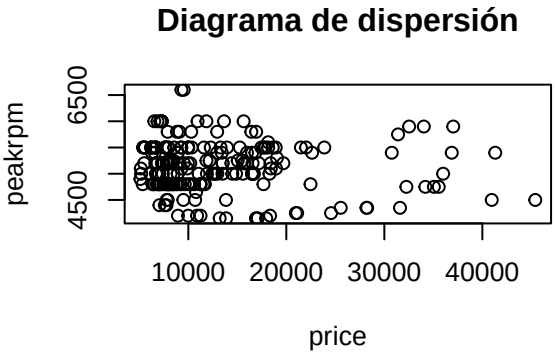
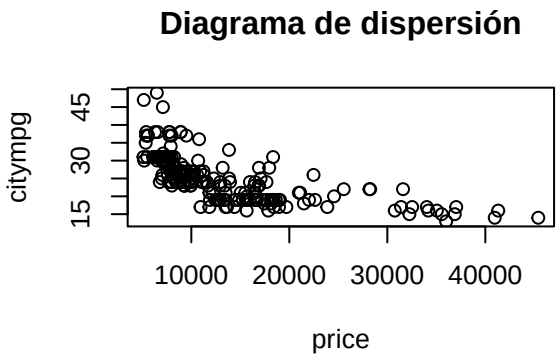
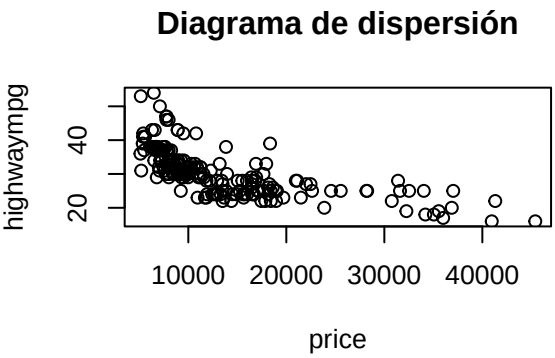


Diagrama de dispersión
compressionratio
10
20
10000
20000
30000
40000
price









Diagramas de barras y de pastel de variables categóricas

Diagrama de barras de symboling

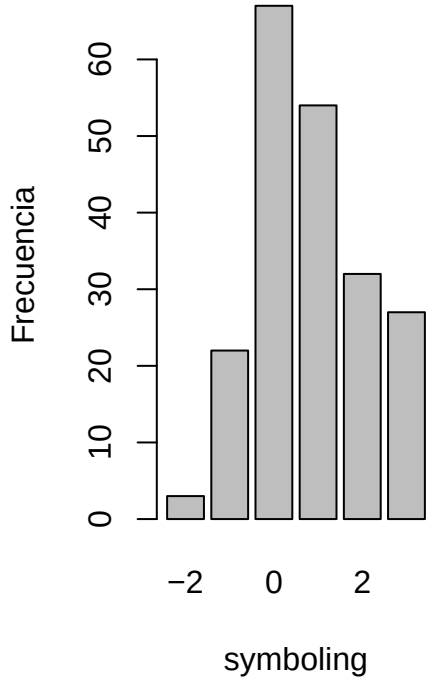


Diagrama de pastel de symboling

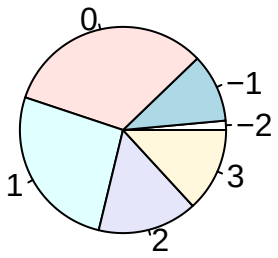


Diagrama de barras de CarName

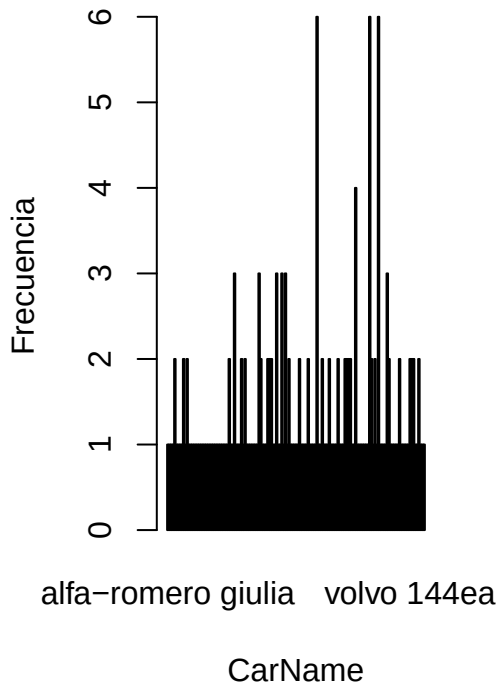


Diagrama de pastel de CarName

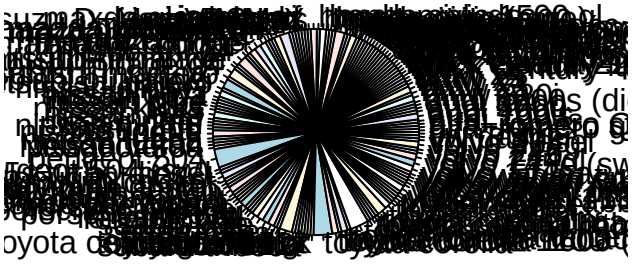


Diagrama de barras de fueltype

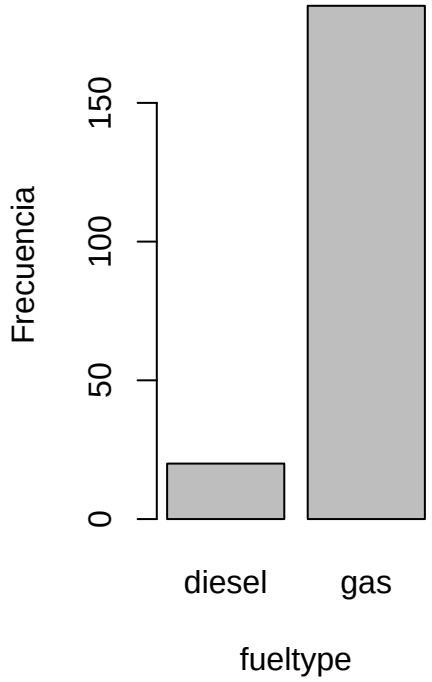


Diagrama de pastel de fueltype

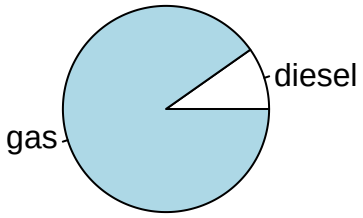


Diagrama de barras de carbody

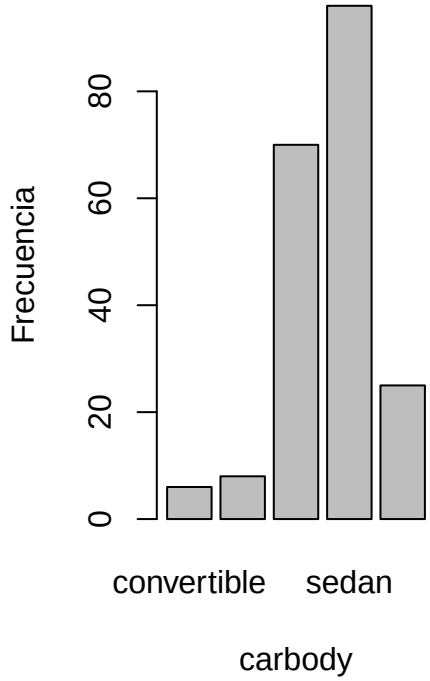


Diagrama de pastel de carbody

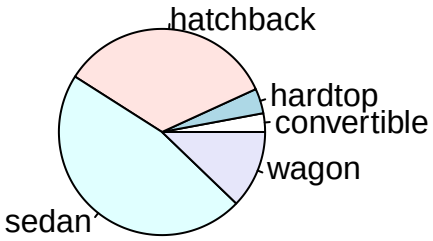


Diagrama de barras de drivewheel

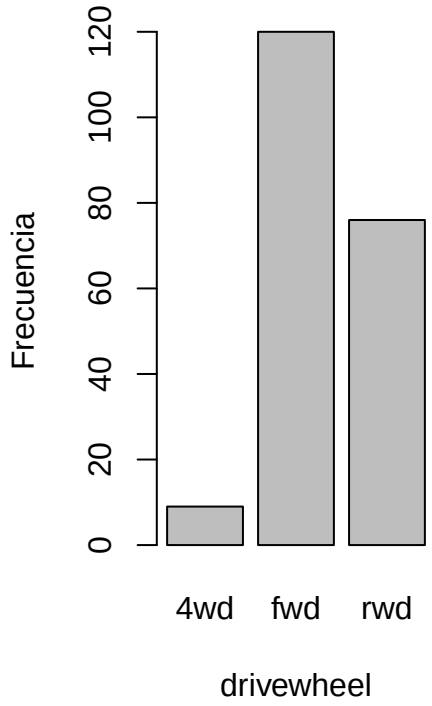
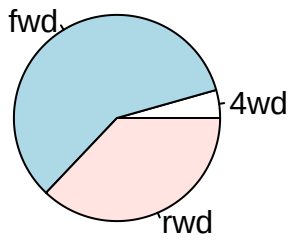
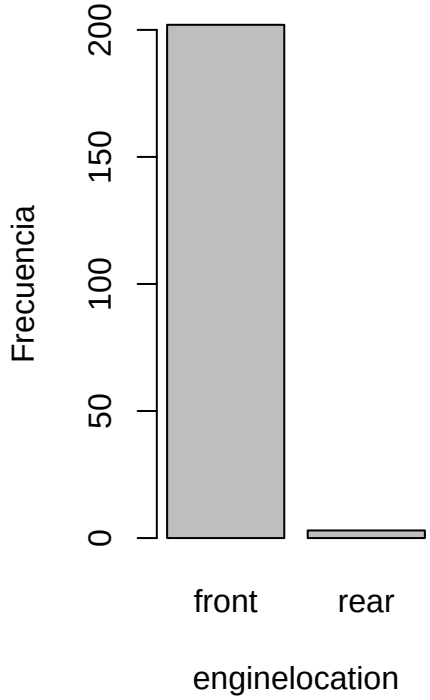


Diagrama de pastel de drivewheel



iagrama de barras de enginelocatioiagrama de pastel de enginelocatio



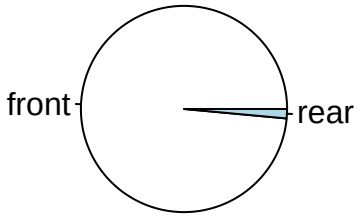


Diagrama de barras de enginetype

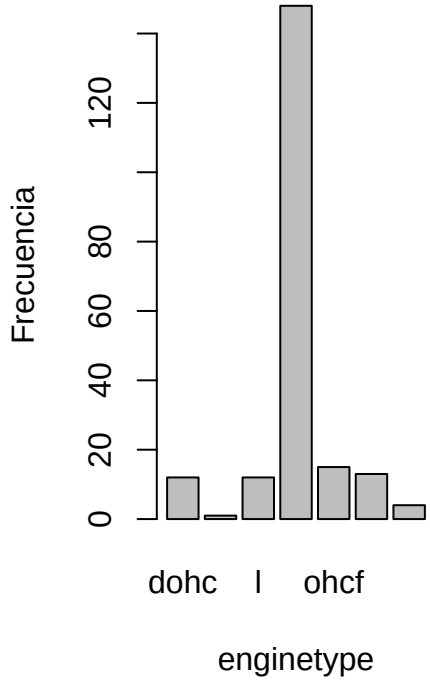
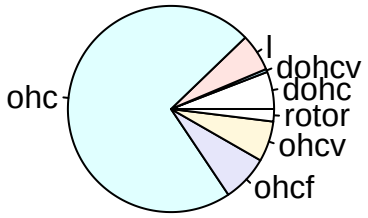
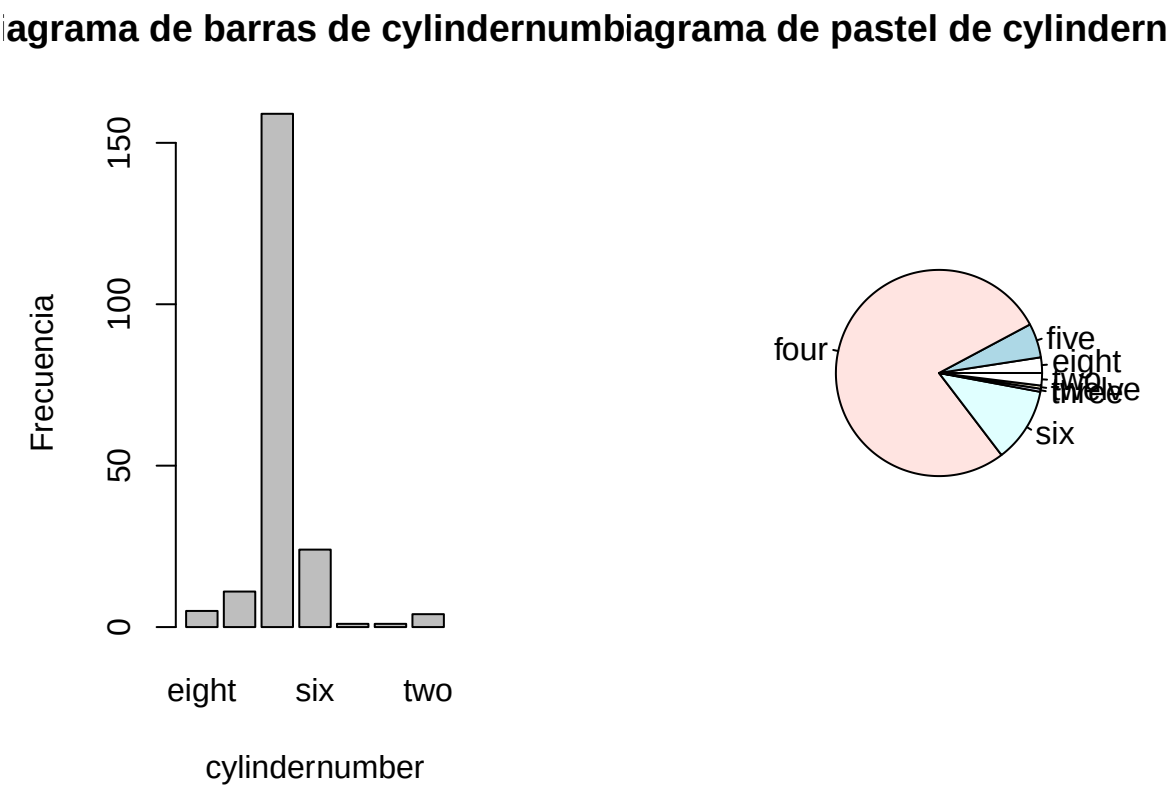
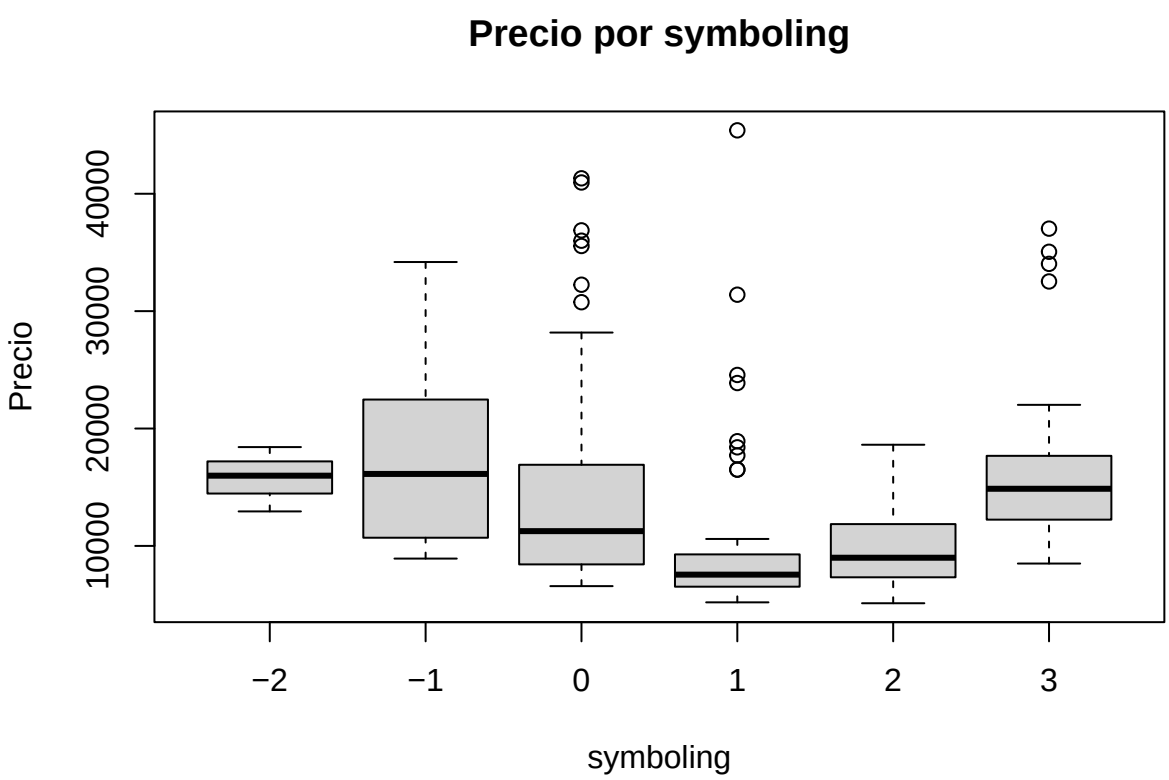


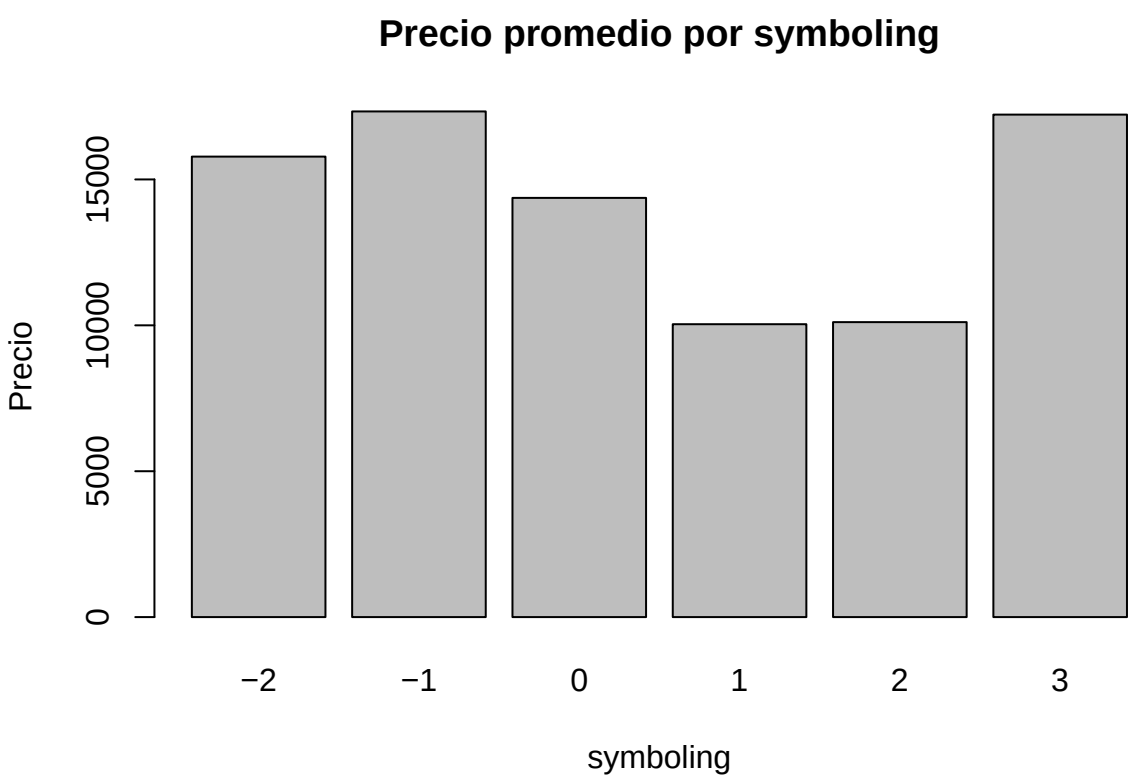
Diagrama de pastel de enginetype

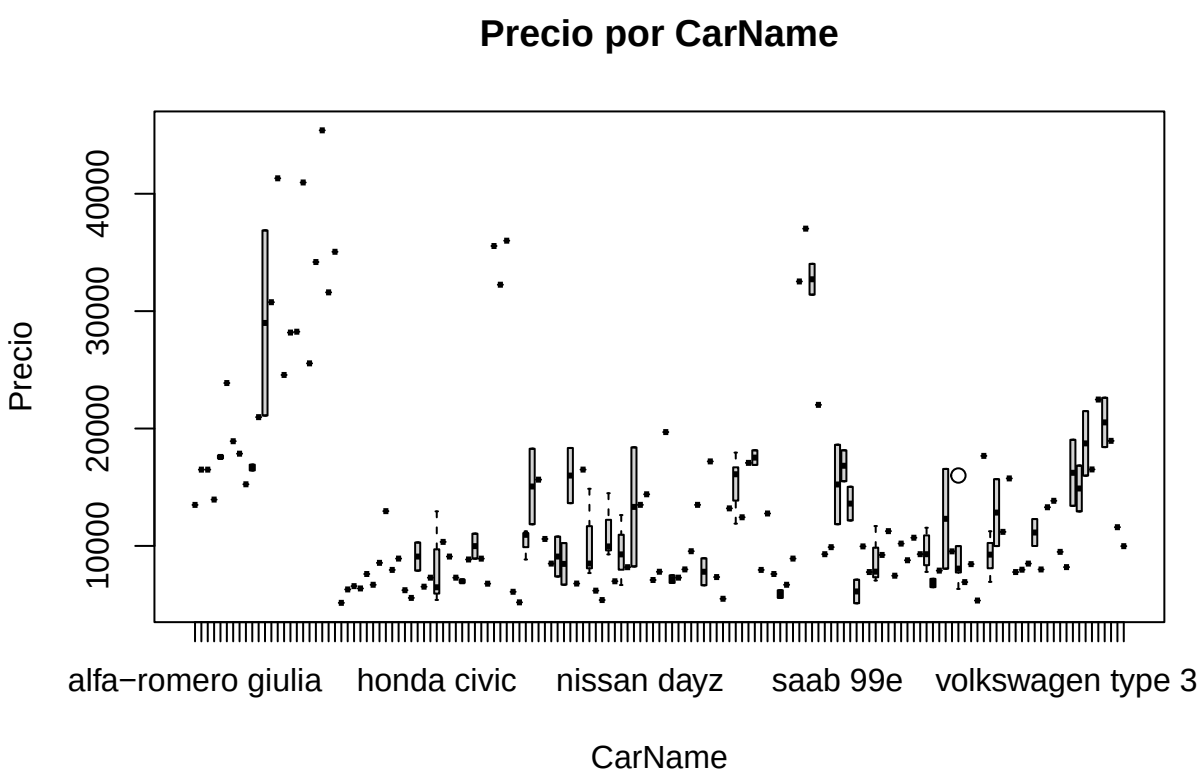


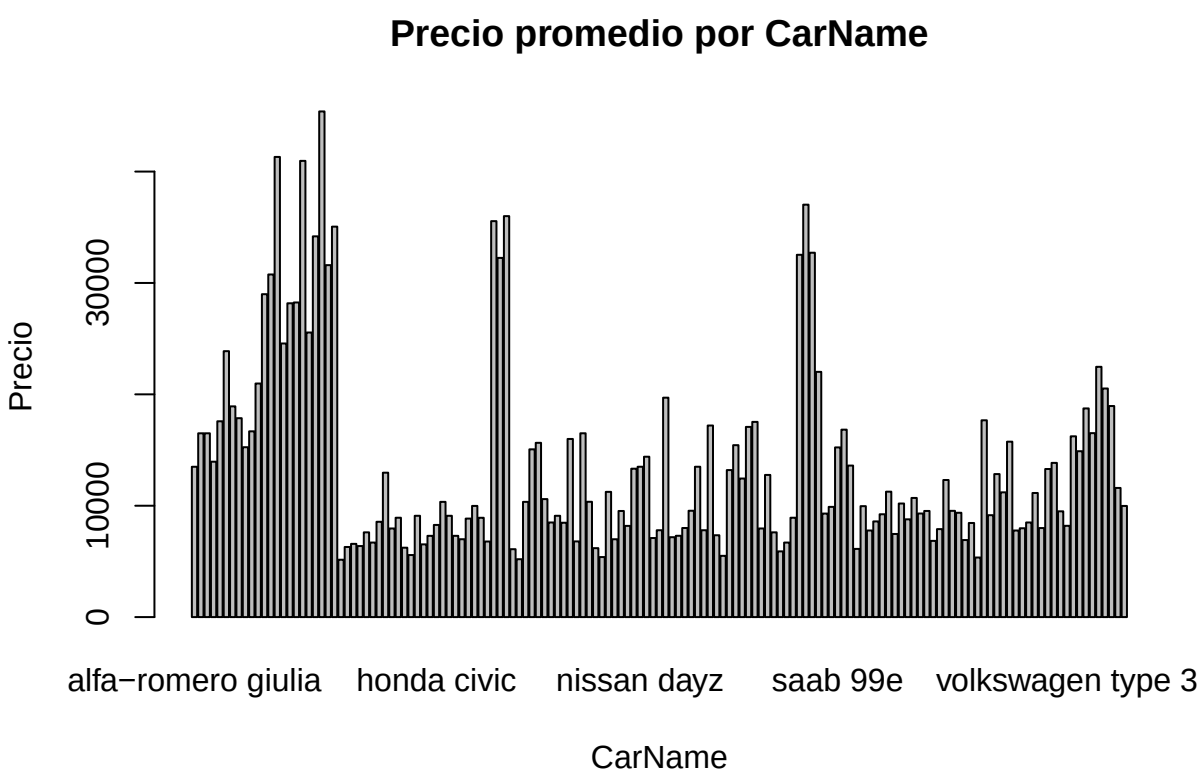


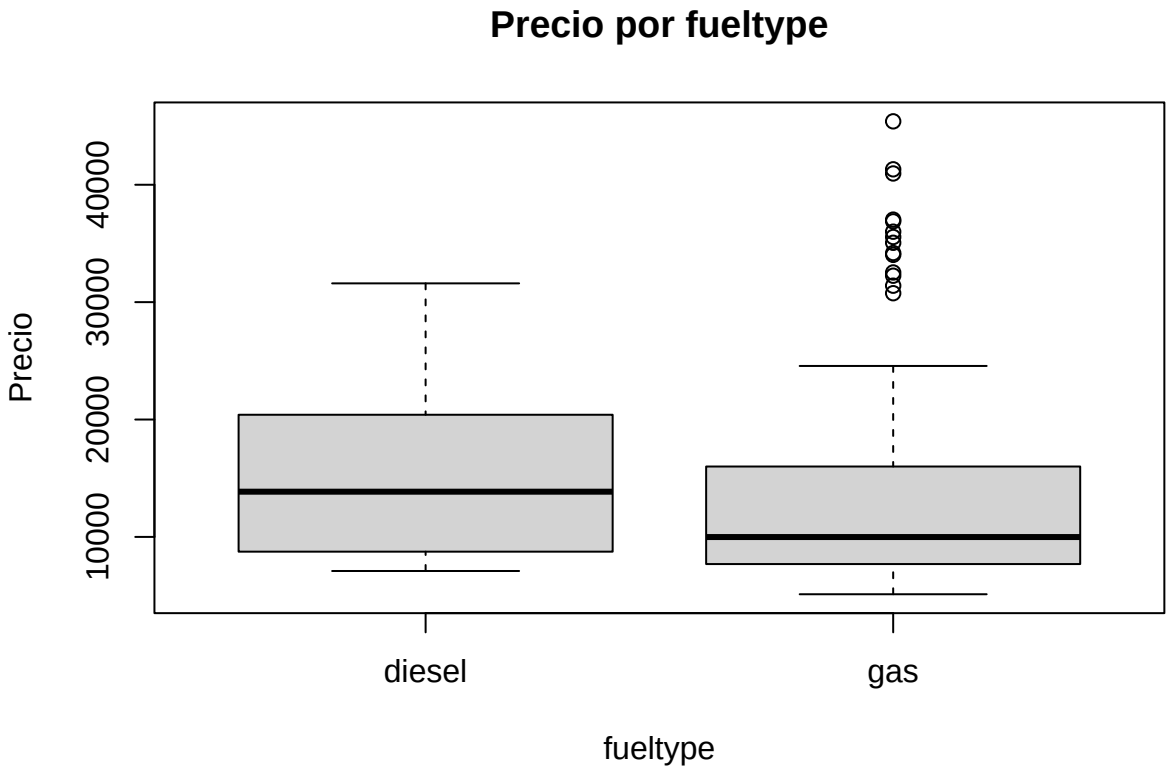
Boxplots y diagramas de barras de precio por cada categoria de cada variable categórica

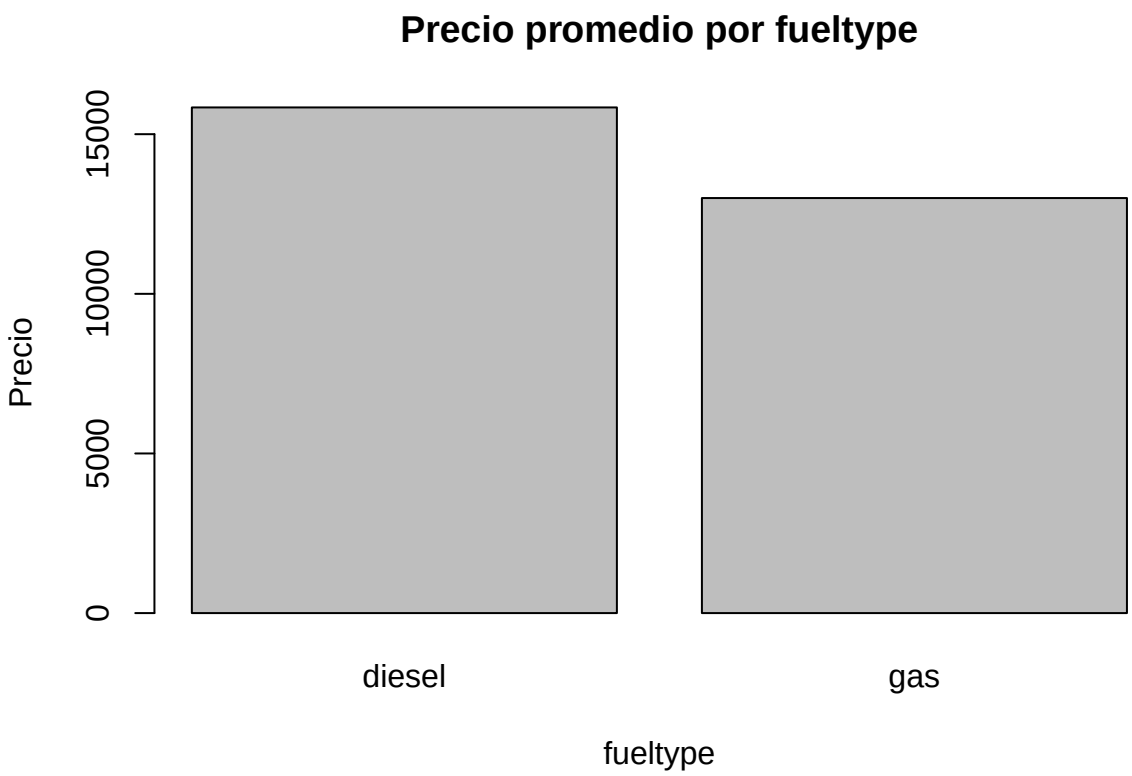

Precio por symboling
Precio
10000
20000
30000
40000
−2
−1
0
1
2
3
symboling

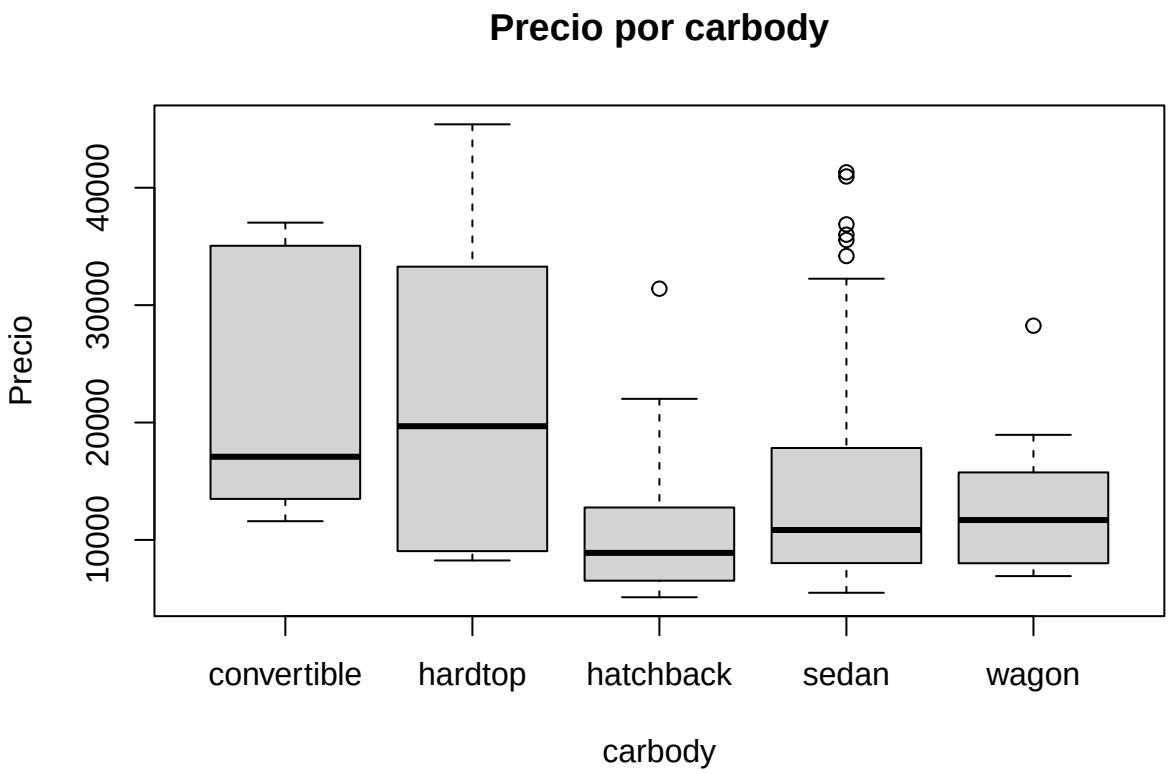

Precio promedio por symboling
Precio
0
5000
10000
15000
−2
−1
0
1
2
3
symboling

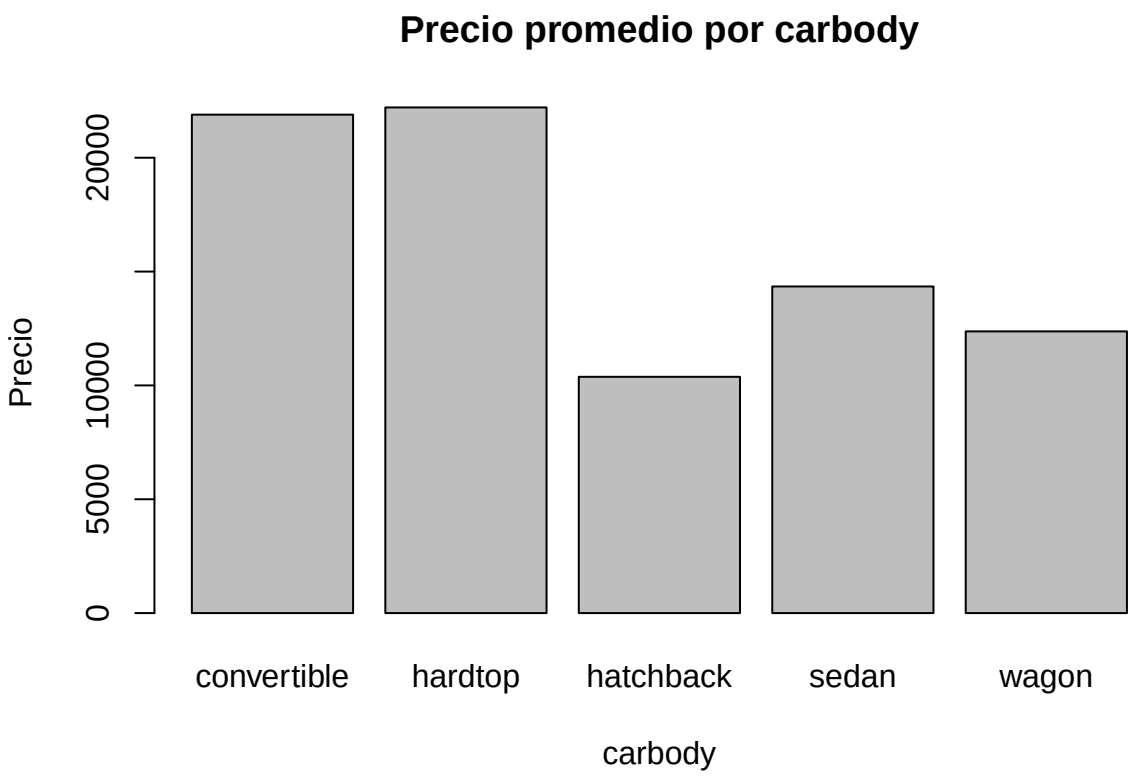

Precio por CarName
Precio
10000
20000
30000
40000
alfa−romero giulia
honda civic
nissan dayz
saab 99e
volkswagen type 3
CarName

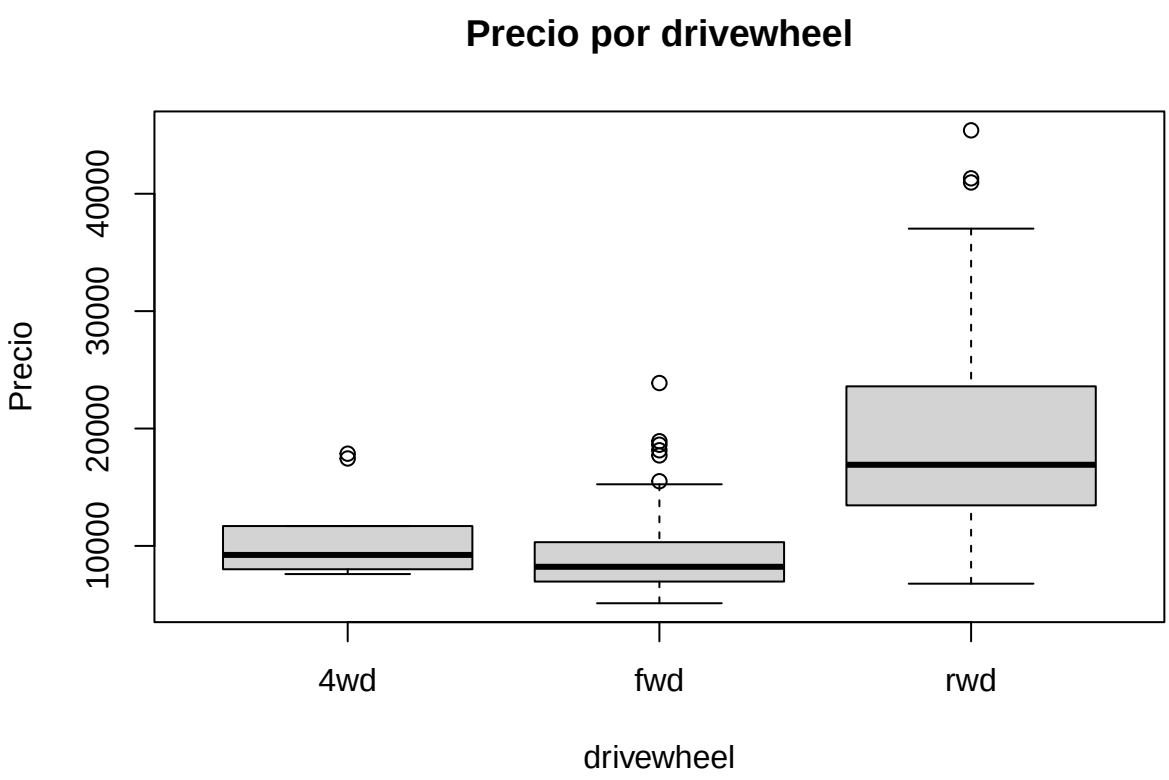

Precio promedio por CarName
Precio
0
10000
30000
alfa−romero giulia
honda civic
nissan dayz
saab 99e
volkswagen type 3
CarName

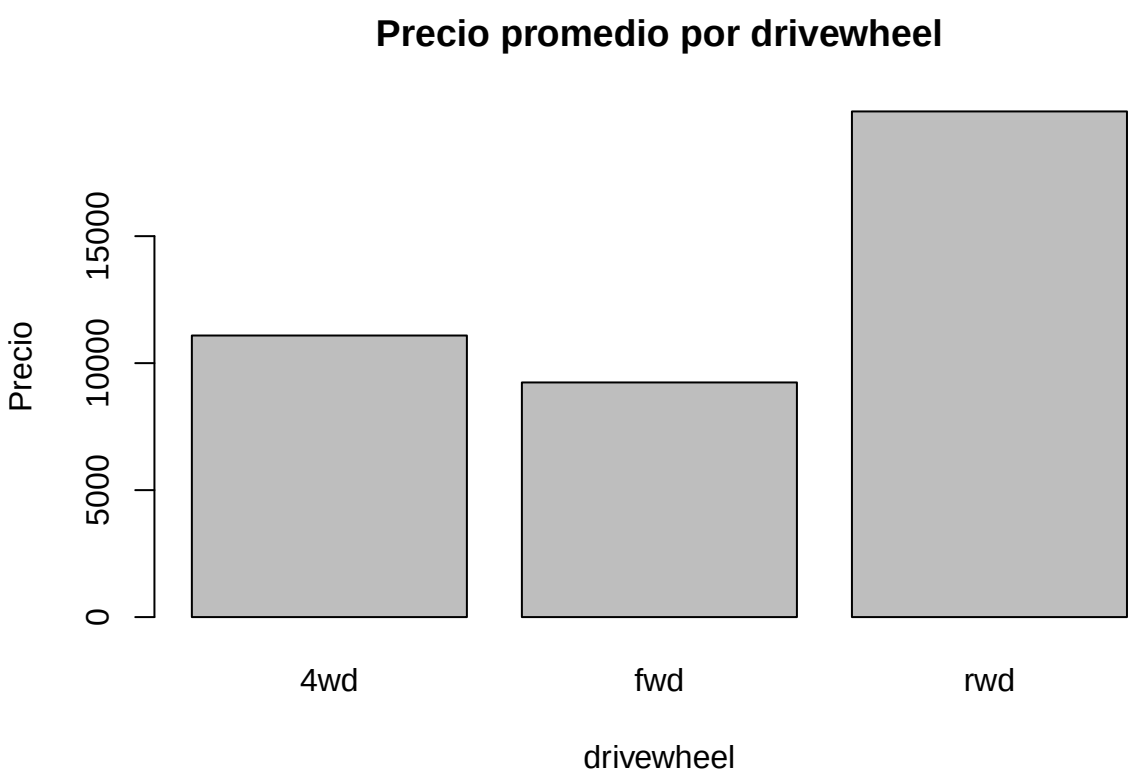

Precio por fueltype
Precio
10000
20000
30000
40000
diesel
gas
fueltype

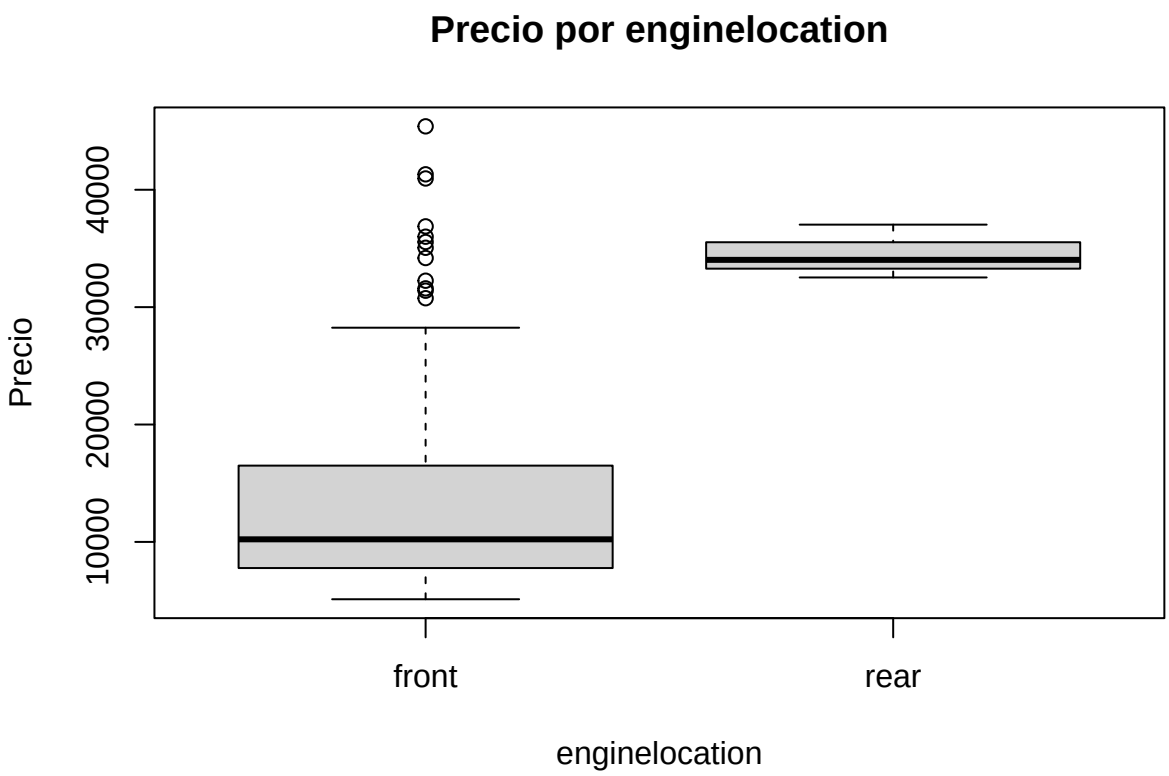

Precio promedio por fueltype
Precio
0
5000
10000
15000
diesel
gas
fueltype

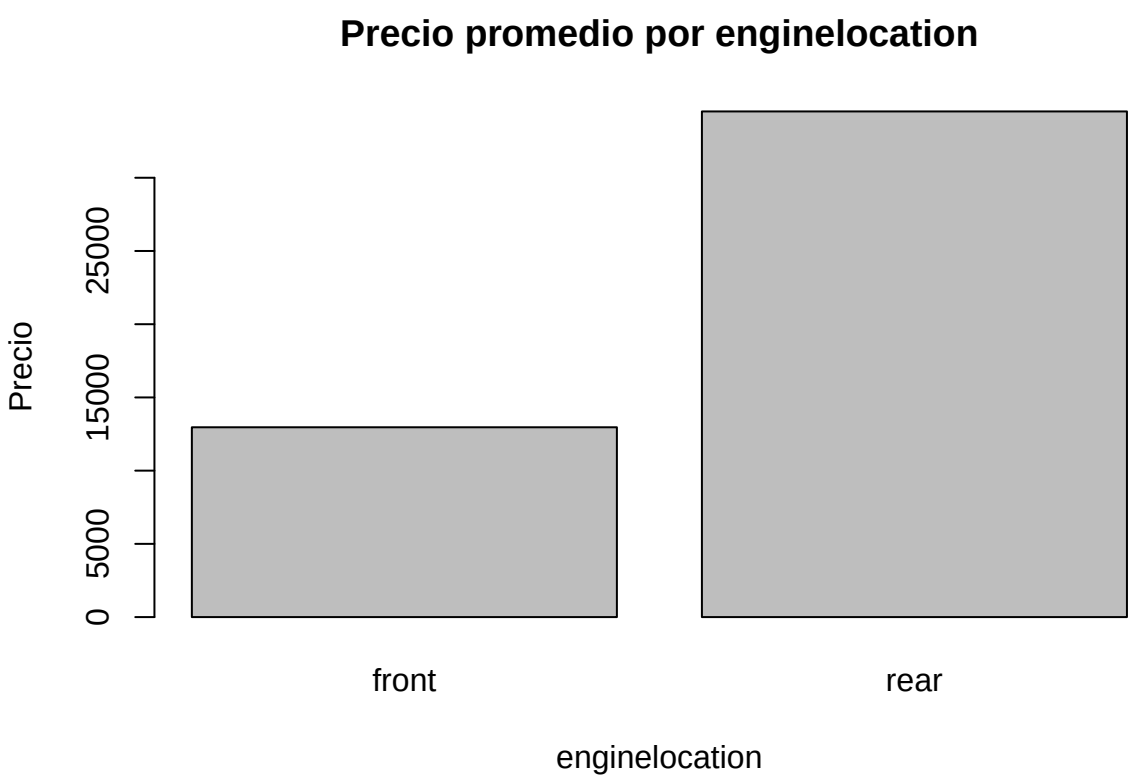

Precio por carbody
Precio
10000
20000
30000
40000
convertible
hardtop
hatchback
sedan
wagon
carbody

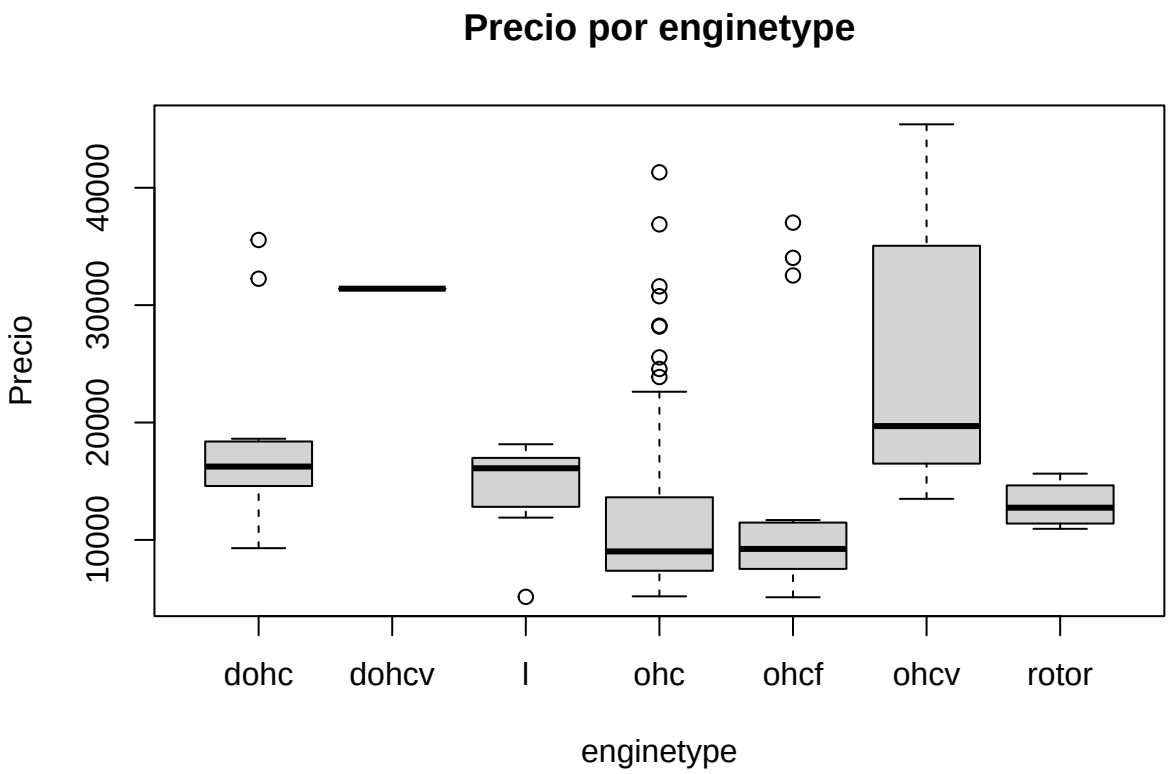

Precio promedio por carbody
Precio
0
5000
10000
20000
convertible
hardtop
hatchback
sedan
wagon
carbody

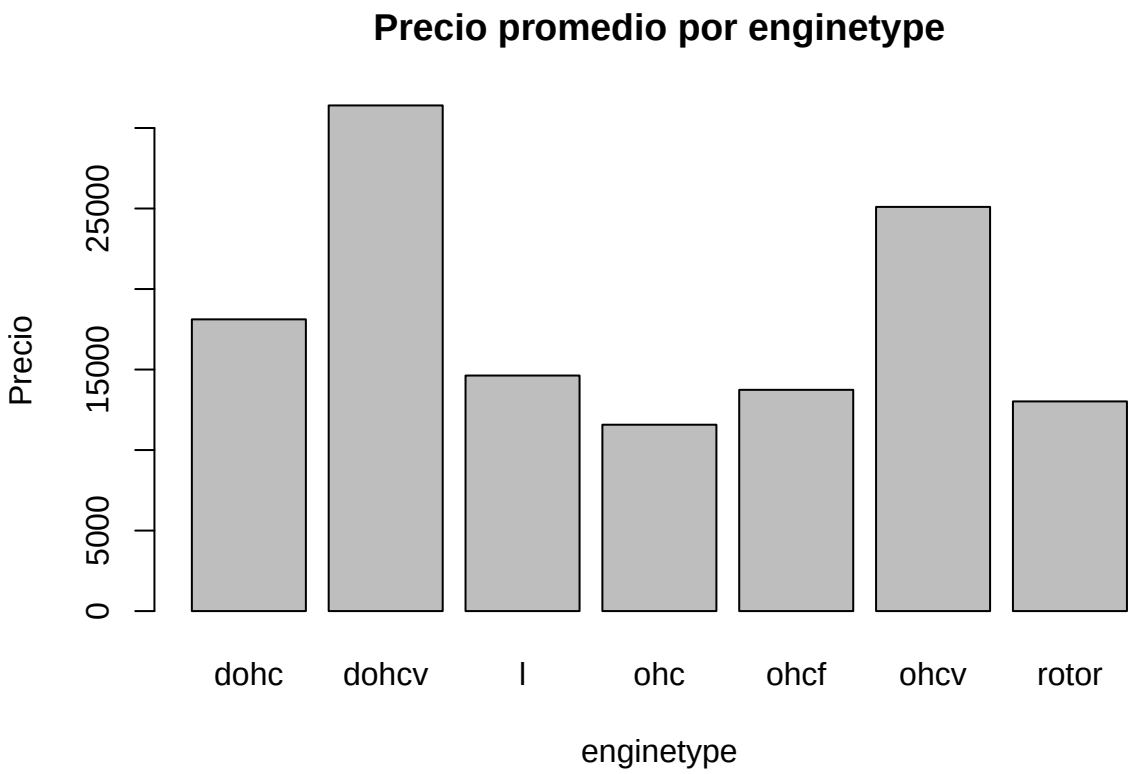

Precio por drivewheel
Precio
10000
20000
30000
40000
4wd
fwd
rwd
drivewheel

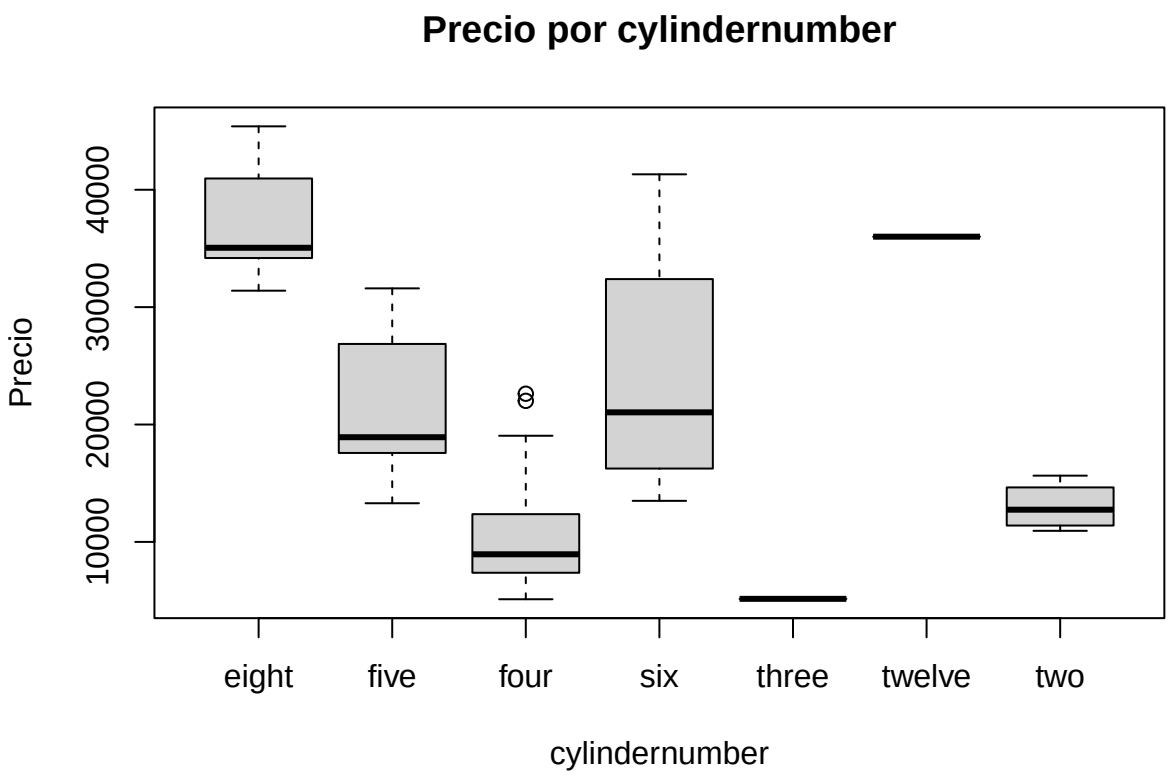

Precio promedio por drivewheel
Precio
0
5000
10000
15000
4wd
fwd
rwd
drivewheel

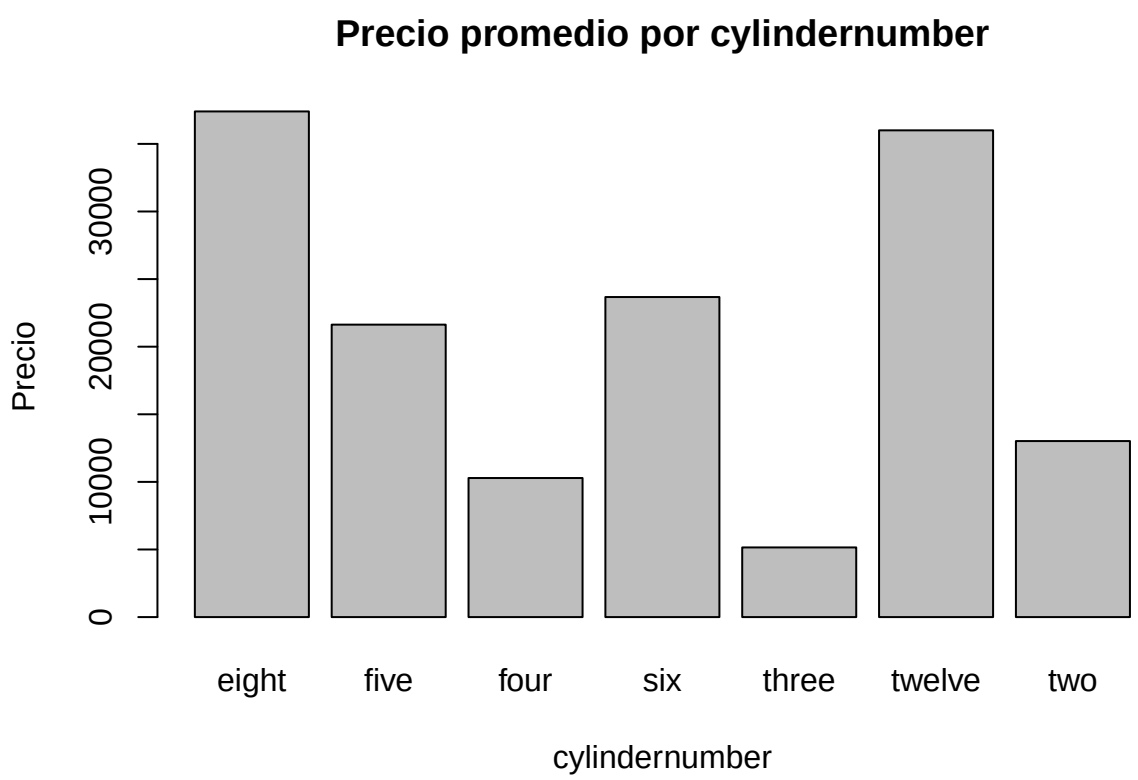

Precio por enginelocation
Precio
10000
20000
30000
40000
front
rear
enginelocation


Precio promedio por enginelocation
Precio
0
5000
15000
25000
front
rear
enginelocation


Precio por enginetype
Precio
10000
20000
30000
40000
dohc
dohcv
l
ohc
ohcf
ohcv
rotor
enginetype


Precio promedio por enginetype
Precio
0
5000
15000
25000
dohc
dohcv
l
ohc
ohcf
ohcv
rotor
enginetype


Precio por cylindernumber
Precio
10000
20000
30000
40000
eight
five
four
six
three
twelve
two
cylindernumber



- Repositorio Portafolio M1:

https://github.com/MaikiBR/m1-portfolio

- Repositorio Portafolio M2:

https://github.com/MaikiBR/m2-portfolio

- Drive de Evidencia 1. Portafolio de análisis:

https://drive.google.com/drive/folders/1JgDDc_ol1RJguCyqgyFWukCq63C6a2Z5?usp=sharing

- Drive de Evidencia 2. Portafolio de implementación:

https://drive.google.com/drive/folders/1qF6_e0_FKQUAO8aSdvJ-nbT6nyMXDTeA?usp=sharing