



# Tecnológico de Monterrey

Tecnológico de Monterrey – Campus Monterrey

Escuela de Ingeniería y Ciencias – Departamento Regional de Ciencias Básicas

**Unidad de formación IN1002B - Desarrollo de Proyectos de Análisis de Datos**

## **Entregable 4.2**

### **– Preparación de la Base de Datos–**

Profesores: Dr. Alexander Garrido Rios

#### **Equipo 1**

Emmanuel Naranjo Blanco – A00835704

Fernanda Martínez Valles – A01722279

Valeria Rojas Minor – A011298657

Emiliano Vazquez Rodriguez – A00344781

Santiago Alonso Mendoza Franco - A00836973

**26 de noviembre del 2023**

Monterrey, Nuevo León

## INTRODUCCIÓN

En el siguiente documento se llevará a cabo un Análisis de Exploración de Datos sobre los datos relevantes al precio de autos usados de la marca Toyota. El objetivo de este análisis es explorar la estructura y características de los datos previo a cualquier suposición, identificar errores, entender patrones, detectar valores atípicos y determinar relaciones entre las variables cuantitativas (numéricas) y cualitativas (categóricas) con el fin de determinar los principales factores que definen el precio de un auto Toyota usado. En este caso se trabaja con un dataset que contiene datos importantes sobre autos usados de la marca Toyota. Se trata de una base de datos que contiene 37 atributos y 1,436 instancias.

Partiendo de esto, se elaborará el análisis exploratorio de datos y posteriormente el procedimiento de imputación de datos donde en caso de ser requerido.

## METODOLOGÍA

Como metodología aplicada a este caso se tomará en cuenta la metodología CRISP-DM (Cross Industry Data Standard Process for Data Mining), el cual es un método probado para orientar trabajo dirigido hacia la minería de datos. Es un proceso de 6 etapas, de las cuales en el presente reporte se hará énfasis en el segundo y tercer paso que hacen referencia a la exploración de los datos por medio del análisis EDA (Análisis Exploratorio de Datos). EDA tiene el propósito de explorar y comprender la estructura y características de los datos antes de aplicar técnicas más avanzadas por medio de la búsqueda de patrones interesantes y relevantes a la pregunta de investigación. Además, en caso de ser necesario, se implementarán técnicas de imputación de datos, con el fin de corregir datos faltantes (NA's, zeros, missing values y white spaces) y obtener una base consistente para su análisis; y de este modo, en el siguiente reporte obtener el mejor modelo de regresión lineal múltiple posible.

Tomando en cuenta CRISP-DM, el primer paso se basa en entender las necesidades del negocio. La información a analizar proviene de una base suministrada por los profesores de la unidad de formación, compuesta de 37 atributos y 1,436 instancias (esto se puede observar por medio de la función DIM, que será explicada más adelante dentro del código de R. A partir del análisis de estos datos se busca identificar los factores que determinan el precio de venta de vehículos de la marca Toyota usados en México.

## ANÁLISIS EDA

A continuación, se realizará el análisis EDA, el cual estará dividido en una serie de pasos. Este análisis hace referencia al segundo y tercer paso del método CRISP-DM: Data Understanding and Data Preparation.

Inicialmente se instalan los siguientes paquetes y se procede con cargar la base de datos.

### PASO 0: Implementación de librerías

```
library("dplyr")  
  
library("ggplot2")  
  
library("tidyverse")  
  
library("pillar")  
  
library("psych")  
  
library("skimr")  
  
library("readr")  
library("GGally")  
  
library("corrplot")  
  
library("reshape2")  
  
library("gmodels")  
  
library("mice")
```

## PASO 1: Carga de la base de datos

```
ds=read.csv("corolla.csv")
View(ds)
skim(ds)
```

### Data summary

Name ds  
Number of rows 1436  
Number of columns 37

### Column type frequency:

character 2  
numeric 35

Group variables None

### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Model	0	1	15	75	0	372	0
Fuel_Type	0	1	3	6	0	3	0

### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Id	0	1	721.56	416.48	1	361.75	721.5	1081.25	1442	
Price	0	1	10730.82	3626.96	4350	8450.00	9900.0	11950.00	32500	
Age_08_04	0	1	55.95	18.60	1	44.00	61.0	70.00	80	
Mfg_Month	0	1	5.55	3.35	1	3.00	5.0	8.00	12	
Mfg_Year	0	1	1999.63	1.54	199898	1998.00	1999.0	2001.00	2004	
KM	0	1	68533.26	37506.45	1	43000.00	63389.5	87020.75	243000	
HP	0	1	101.50	14.98	69	90.00	110.0	110.00	192	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Met_Color	0	1	0.67	0.47	0	0.00	1.0	1.00	1	█-- --█
Automatic	0	1	0.06	0.23	0	0.00	0.0	0.00	1	█-- --
cc	0	1	1576.86	424.39	1300	1400.00	1600.0	1600.00	16000	█-- --
Doors	0	1	4.03	0.95	2	3.00	4.0	5.00	5	--█ █
Cylinders	0	1	4.00	0.00	4	4.00	4.0	4.00	4	--█ --
Gears	0	1	5.03	0.19	3	5.00	5.0	5.00	6	-- --█
Quarterly_Tax	0	1	87.12	41.13	19	69.00	85.0	85.00	283	█-- --
Weight	0	1	1072.46	52.64	1000	1040.00	1070.0	1085.00	1615	█-- --
Mfr_Guarantee	0	1	0.41	0.49	0	0.00	0.0	1.00	1	█-- --█
BOVAG_Guarantee	0	1	0.90	0.31	0	1.00	1.0	1.00	1	-- --█
Guarantee_Period	0	1	3.82	3.01	3	3.00	3.0	3.00	36	█-- --
ABS	0	1	0.81	0.39	0	1.00	1.0	1.00	1	-- --█
Airbag_1	0	1	0.97	0.17	0	1.00	1.0	1.00	1	-- --█
Airbag_2	0	1	0.72	0.45	0	0.00	1.0	1.00	1	█-- --█
Airco	0	1	0.51	0.50	0	0.00	1.0	1.00	1	█-- --█
Automatic_airco	0	1	0.06	0.23	0	0.00	0.0	0.00	1	█-- --
Boardcomputer	0	1	0.29	0.46	0	0.00	0.0	1.00	1	█-- --█
CD_Player	0	1	0.22	0.41	0	0.00	0.0	0.00	1	█-- --

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Central_Lock	0	1	0.58	0.49	0	0.00	1.0	1.00	1	█_ _ _ █
Powered_Windows	0	1	0.56	0.50	0	0.00	1.0	1.00	1	█_ _ _ █
Power_Steering	0	1	0.98	0.15	0	1.00	1.0	1.00	1	_ _ _ _ █
Radio	0	1	0.15	0.35	0	0.00	0.0	0.00	1	█_ _ _ _
Mistlamps	0	1	0.26	0.44	0	0.00	0.0	1.00	1	█_ _ _ █
Sport_Model	0	1	0.30	0.46	0	0.00	0.0	1.00	1	█_ _ _ █
Backseat_Divider	0	1	0.77	0.42	0	1.00	1.0	1.00	1	_ _ _ _ █
Metallic_Rim	0	1	0.20	0.40	0	0.00	0.0	0.00	1	█_ _ _ _
Radio_cassette	0	1	0.15	0.35	0	0.00	0.0	0.00	1	█_ _ _ _
Tow_Bar	0	1	0.28	0.45	0	0.00	0.0	1.00	1	█_ _ _ █

```
sapply(ds, function(x) sum(is.na(x)))
```

```
##           Id           Model           Price           Age_08_04
##           0             0             0             0
##      Mfg_Month      Mfg_Year           KM           Fuel_Type
##           0             0             0             0
##           HP           Met_Color      Automatic           cc
##           0             0             0             0
##           Doors      Cylinders           Gears      Quarterly_Tax
##           0             0             0             0
##           Weight      Mfr_Guarantee      BOVAG_Guarantee      Guarantee_Period
##           0             0             0             0
##           ABS           Airbag_1           Airbag_2           Airco
##           0             0             0             0
##      Automatic_airco      Boardcomputer      CD_Player      Central_Lock
##           0             0             0             0
##      Powered_Windows      Power_Steering           Radio           Mistlamps
##           0             0             0             0
##           Sport_Model      Backseat_Divider      Metallic_Rim      Radio_cassette
##           0             0             0             0
##           Tow_Bar
##           0
```

En primera instancia, con el uso de la función 'skim', se puede observar que hay 2 atributos de caracteres y 35 atributos numéricos. Sin embargo, de los 35 atributos numéricos, 21 son considerados lógicos ya que definen numéricamente 2 clases de respuesta. Además, con el uso de la función 'sapply', se puede observar que los atributos no cuentan con missing values.

## PASO 2: Creación de un 'subset'.

Una vez cargada la base de datos, se creará un subset de datos con uso del 'PIPE OPERATOR'. Se han seleccionado columnas específicas que corresponden a los atributos numéricos que permitirán hacer un contraste cuantitativo mediante representaciones gráficas. Por el momento no se trabajarán los datos de carácter lógico, no obstante cabe destacar que serán analizados en instancias posteriores del proyecto en cuestión ya que permiten representar y trabajar con información que tiene naturaleza binaria.

```
corolla.df = ds %>%  
  select("Id", "Model", "Price", "Age_08_04", "Mfg_Month",  
         "Mfg_Year", "KM", "HP", "cc", "Doors", "Cylinders", "Gears", "Quarterly_Tax",  
         "Weight", "Guarantee_Period")
```

```
View(corolla.df)
```

El nuevo dataframe creado, llamado 'corolla.df' se compone de los atributos numéricos contenidos en la base de datos original; es decir,  
"Id", "Model", "Price", "Age\_08\_04", "Mfg\_Month",  
"Mfg\_Year", "KM", "HP", "cc", "Doors", "Cylinders", "Gears", "Quarterly\_Tax", "Weight", y  
"Guarantee\_Period".

## PASO 3: Exploración inicial de los datos

Ahora se procederá con la exploración de datos inicial.

La función 'skim' proporciona rápidamente una visión general del marco de los datos. En este caso, nos muestra el número de valores que faltan (missing values), así como las características estadísticas básicas. La función 'dim' proporciona las dimensiones del dataframe, es decir, el número de filas y columnas.

Este es un paso básico pero crucial para entender el tamaño del conjunto de datos que estamos trabajando.

```
skim(corolla.df)
```

*Data summary*

Name	corolla.df
Number of rows	1436
Number of columns	15

---

Column type frequency:

character	1
numeric	14



Group variables          None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Model	0	1	15	75	0	372	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Id	0	1	721.56	416.48	1	361.75	721.5	1081.25	1442	
Price	0	1	1073.082	3626.96	4350	8450.00	9900.0	11950.00	32500	
Age_08_04	0	1	55.95	18.60	1	44.00	61.0	70.00	80	
Mfg_Month	0	1	5.55	3.35	1	3.00	5.0	8.00	12	
Mfg_Year	0	1	1999.63	1.54	199898	1998.00	1999.0	2001.00	2004	
KM	0	1	6853.326	37506.45	1	43000.00	63389.5	87020.75	243000	
HP	0	1	101.50	14.98	69	90.00	110.0	110.00	192	
cc	0	1	1576.86	424.39	1300	1400.00	1600.0	1600.00	16000	
Doors	0	1	4.03	0.95	2	3.00	4.0	5.00	5	
Cylinders	0	1	4.00	0.00	4	4.00	4.0	4.00	4	
Gears	0	1	5.03	0.19	3	5.00	5.0	5.00	6	
Quarterly_Tax	0	1	87.12	41.13	19	69.00	85.0	85.00	283	
Weight	0	1	1072.46	52.64	1000	1040.00	1070.0	1085.00	1615	
Guarantee_Period	0	1	3.82	3.01	3	3.00	3.0	3.00	36	

```
dim(corolla.df)
## [1] 1436 15
```

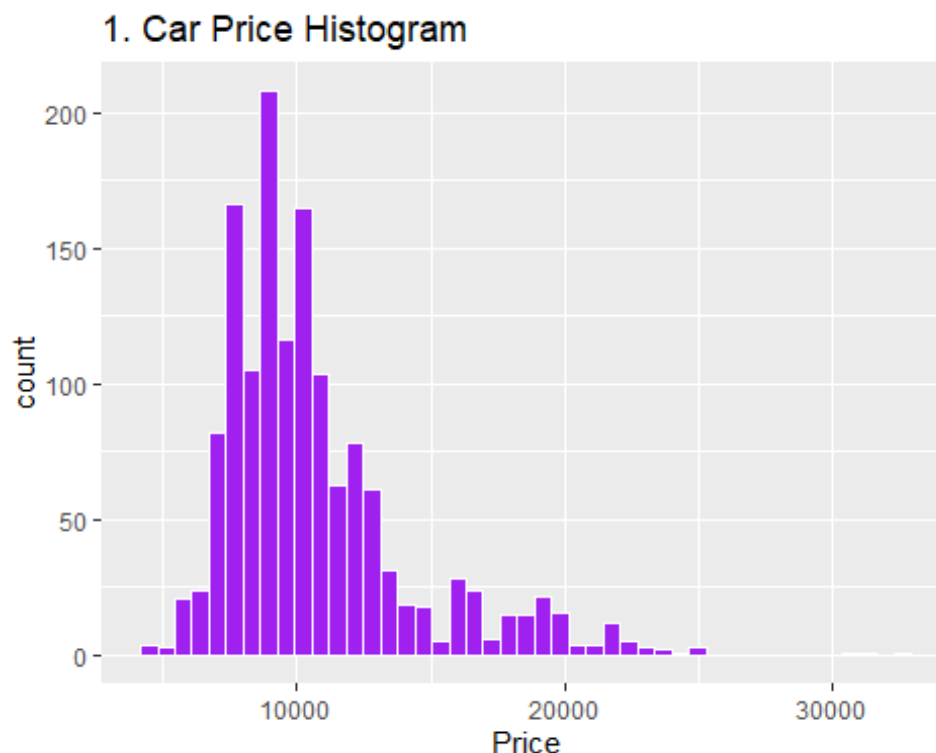
De esta manera, con el uso de la función 'skim' se nos proporcionó un resumen de los contenidos de cada atributo. En este caso se puede observar que el dataframe tiene una dimensión de 15 atributos y 1,436 instancias (observado con la función 'dim'). De los 15 atributos, es evidente que 14 tienen contenido numérico, y solo 1 tiene contenido de caracteres (perteneciente al nombre del modelo del carro). Además, de entrada se puede observar que los atributos contenidos en este nuevo dataframe no tienen datos faltantes.

Al tener este primer análisis, se puede continuar con la graficación de los histogramas en el siguiente paso.

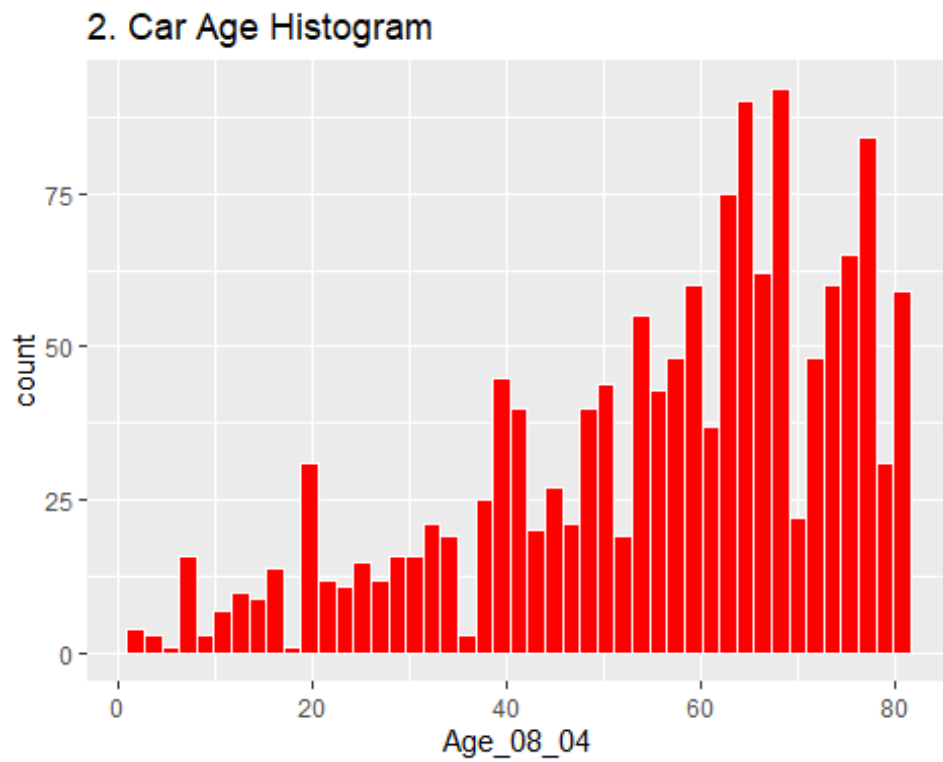
#### PASO 4: Graficación de histogramas

La función 'ggplot' permite crear gráficas de datos complejos. A continuación se muestran los histogramas hechos para las variables o atributos numéricos seleccionados específicamente en el subset de datos "corolla.df".

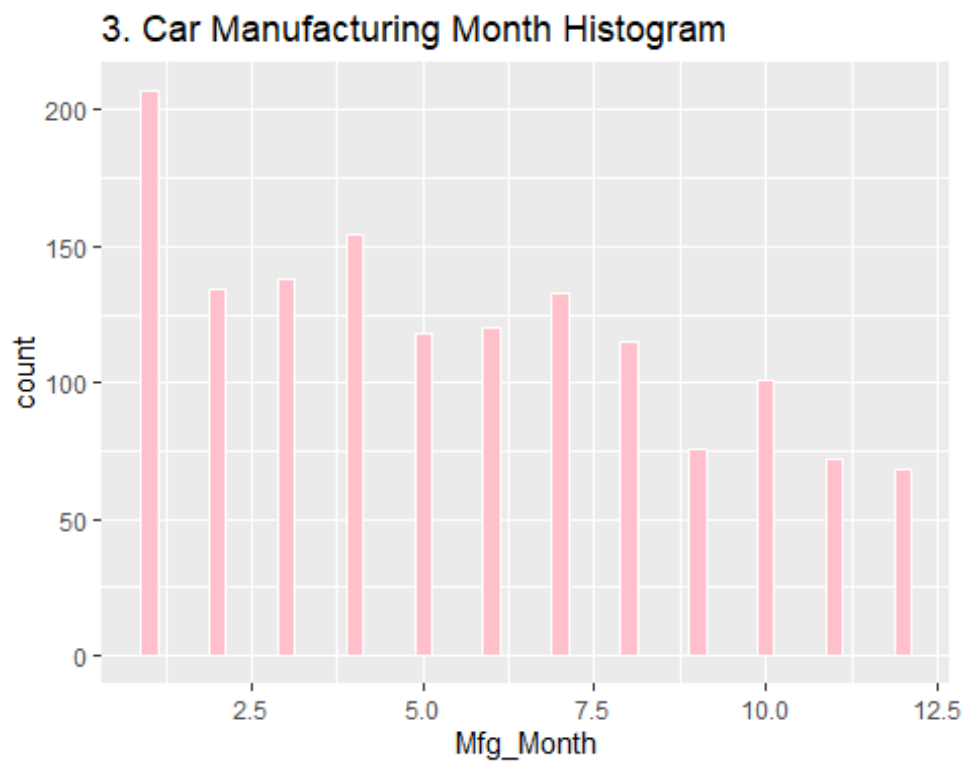
```
ggplot(corolla.df, mapping = aes(x=Price))+
  geom_histogram(bins=45, fill="purple", col="white")+
  labs(title = "1. Car Price Histogram")
```



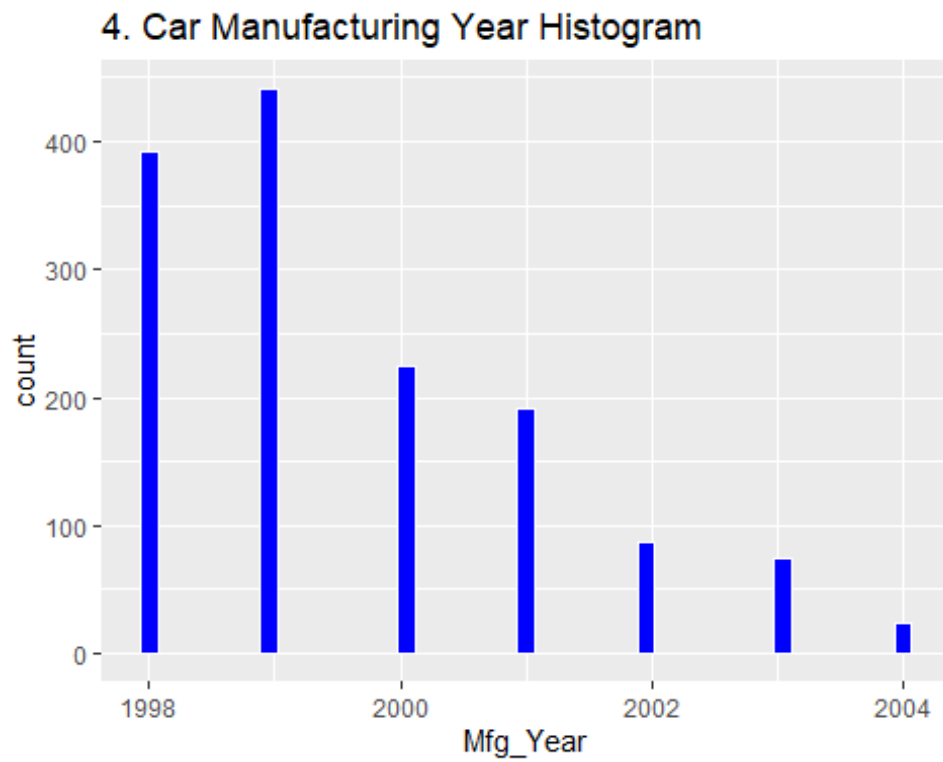
```
ggplot(corolla.df, mapping = aes(x=Age_08_04))+  
  geom_histogram(bins=45, fill="red", col="white")+  
  labs(title = "2. Car Age Histogram")
```



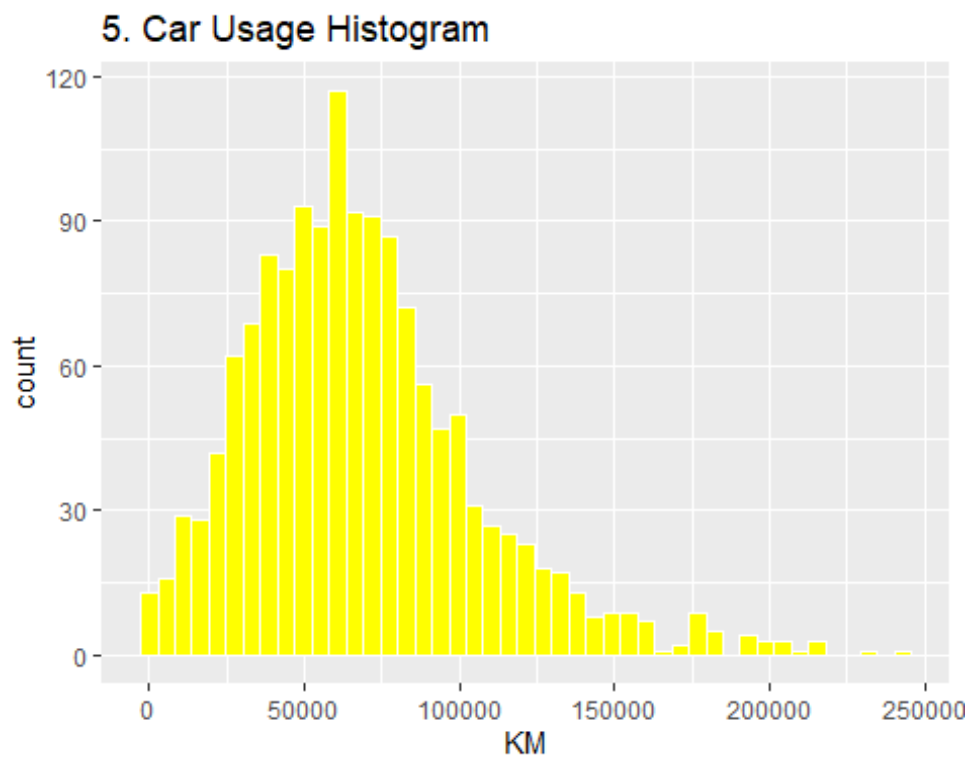
```
ggplot(corolla.df, mapping = aes(x=Mfg_Month))+  
  geom_histogram(bins=45, fill="pink", col="white")+  
  labs(title = "3. Car Manufacturing Month Histogram")
```



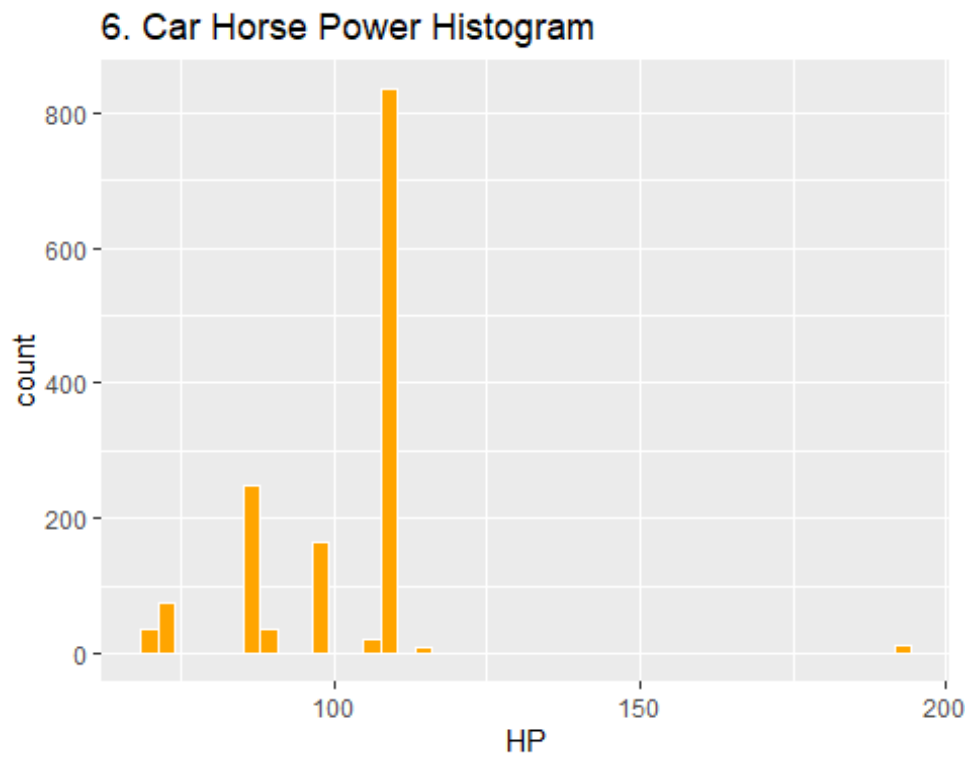
```
ggplot(corolla.df, mapping = aes(x=Mfg_Year)) +  
  geom_histogram(bins=45, fill="blue", col="white") +  
  labs(title = "4. Car Manufacturing Year Histogram")
```



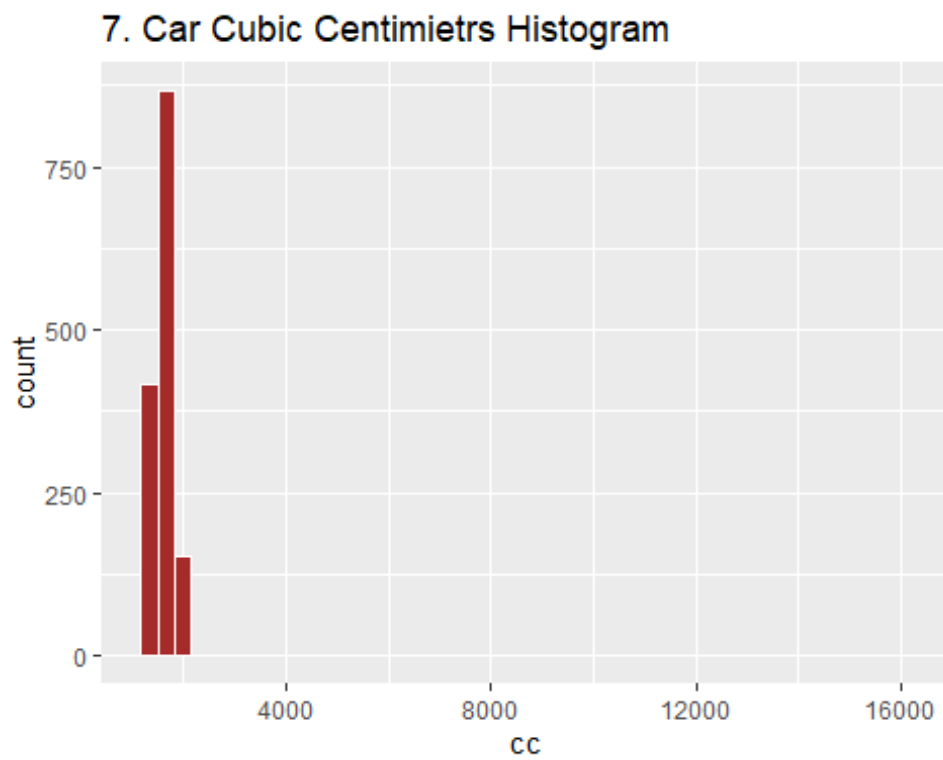
```
ggplot(corolla.df, mapping = aes(x=KM))+  
  geom_histogram(bins=45, fill="yellow", col="white")+  
  labs(title = "5. Car Usage Histogram")
```



```
ggplot(corolla.df, mapping = aes(x=HP)) +  
  geom_histogram(bins=45, fill="orange", col="white") +  
  labs(title = "6. Car Horse Power Histogram")
```

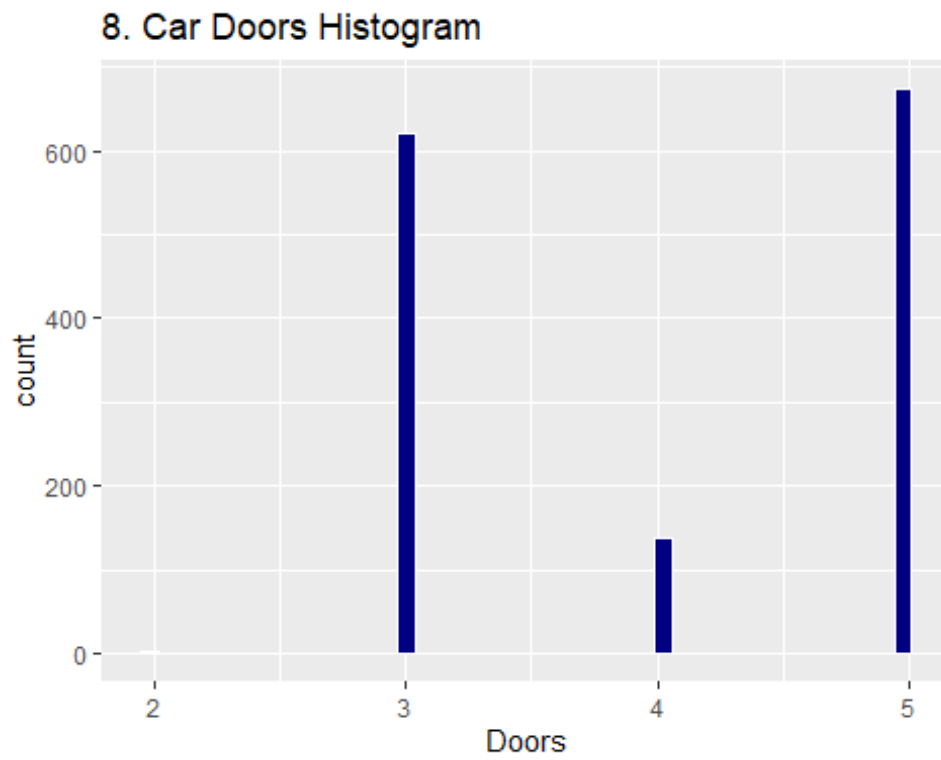


```
ggplot(corolla.df, mapping = aes(x=cc))+  
  geom_histogram(bins=45, fill="brown", col="white")+  
  labs(title = "7. Car Cubic Centimeters Histogram")
```

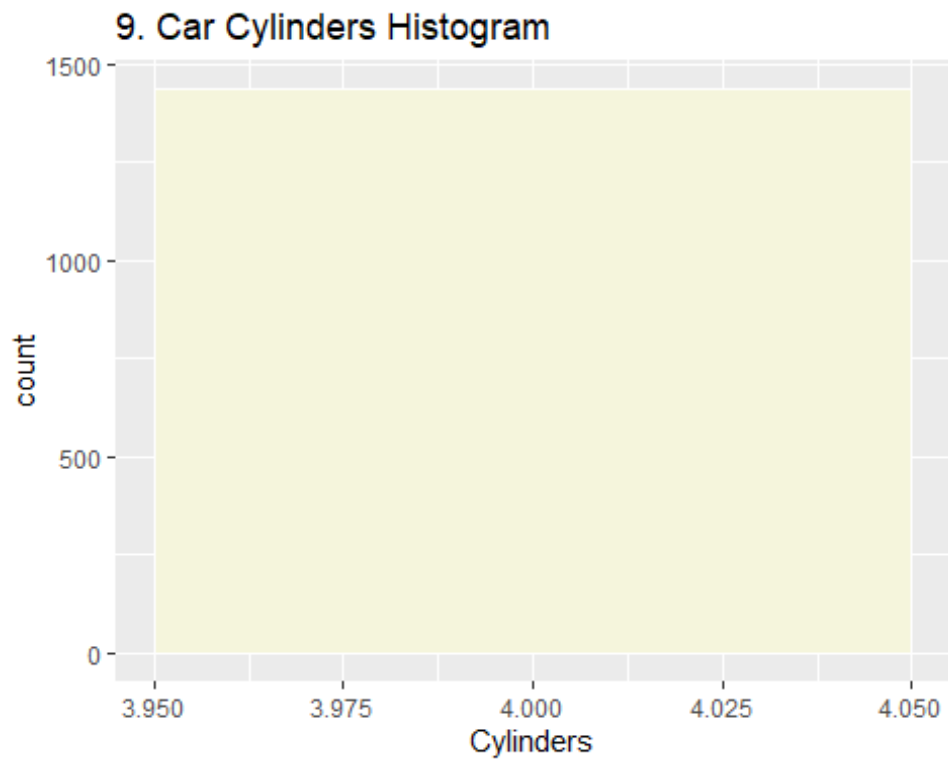




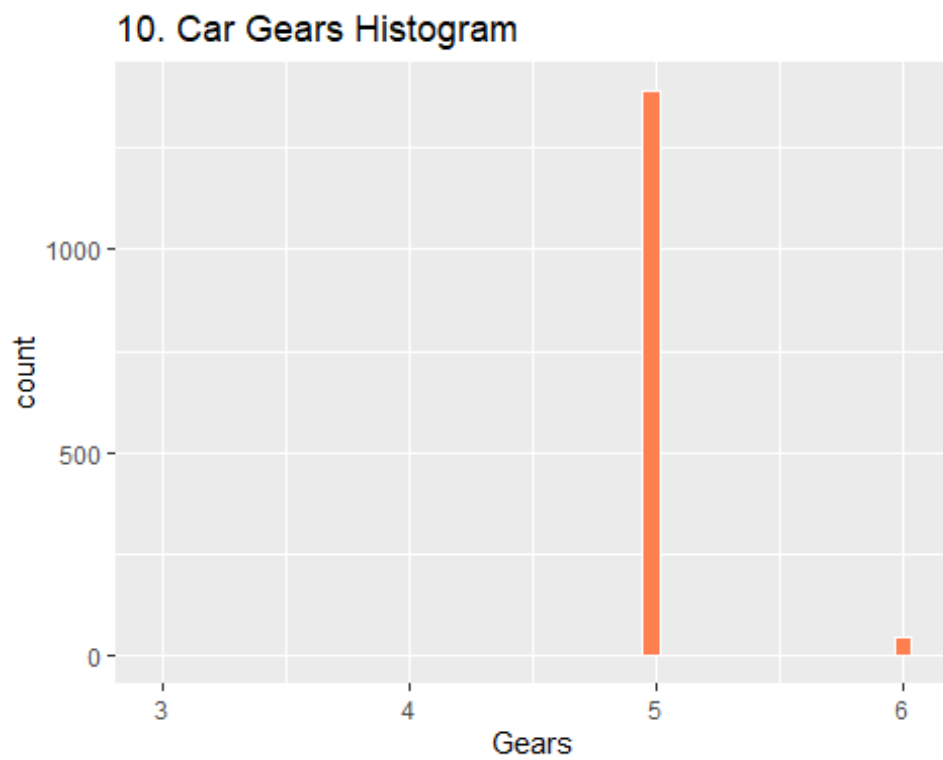
```
ggplot(corolla.df, mapping = aes(x=Doors)) +  
  geom_histogram(bins=45, fill="navy", col="white") +  
  labs(title = "8. Car Doors Histogram")
```



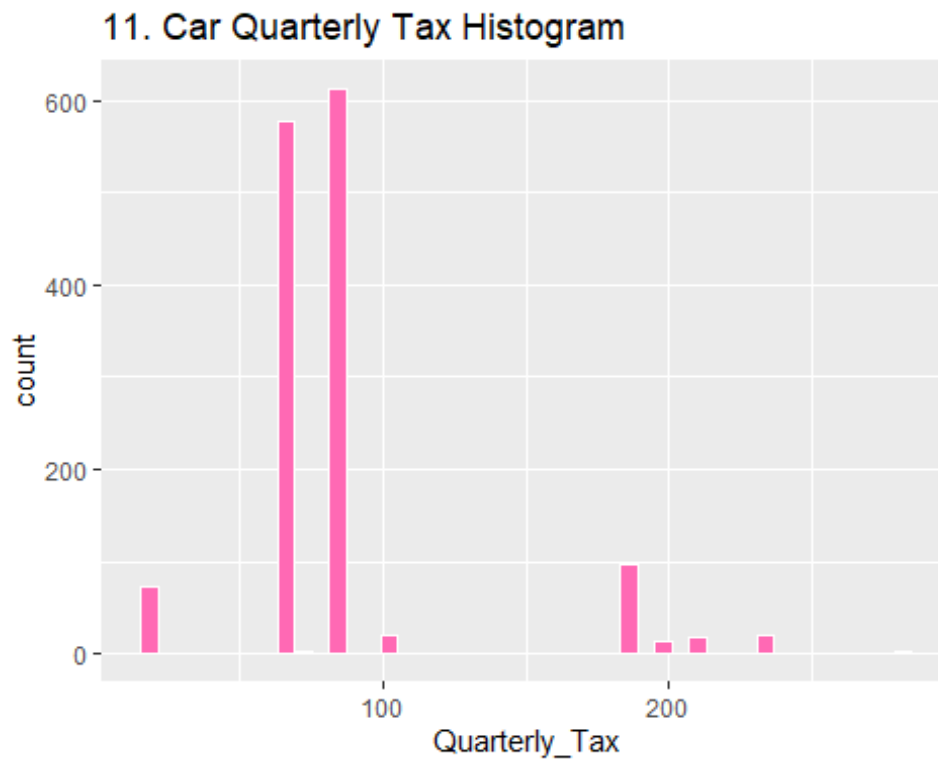
```
ggplot(corolla.df, mapping = aes(x=Cylinders))+  
  geom_histogram(bins=45, fill="beige", col="white")+  
  labs(title = "9. Car Cylinders Histogram")
```



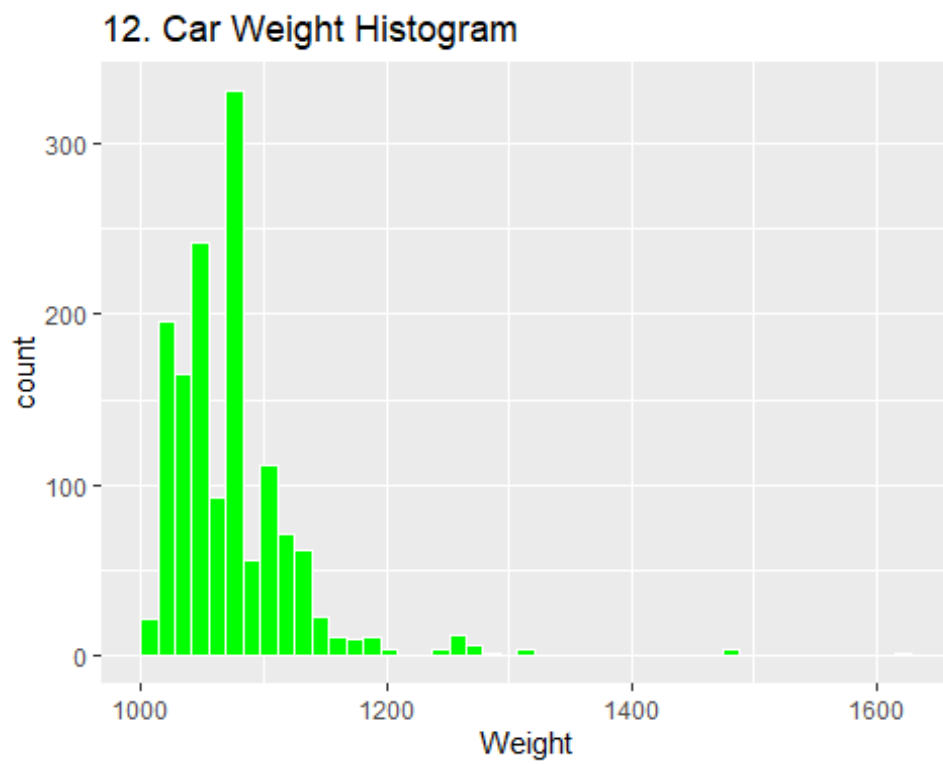
```
ggplot(corolla.df, mapping = aes(x=Gears))+  
  geom_histogram(bins=45, fill="coral", col="white")+  
  labs(title = "10. Car Gears Histogram")
```



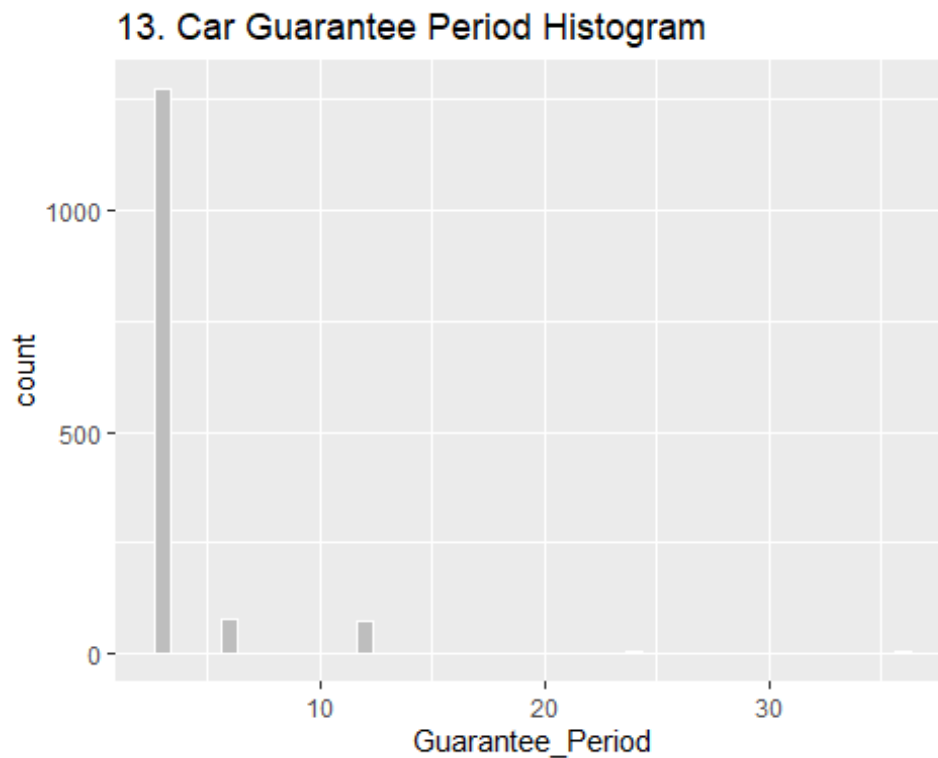
```
ggplot(corolla.df, mapping = aes(x=Quarterly_Tax))+  
  geom_histogram(bins=45, fill="hotpink", col="white")+  
  labs(title = "11. Car Quarterly Tax Histogram")
```



```
ggplot(corolla.df, mapping = aes(x=Weight))+  
  geom_histogram(bins=45, fill="green", col="white")+  
  labs(title = "12. Car Weight Histogram")
```



```
ggplot(corolla.df, mapping = aes(x=Guarantee_Period)) +  
  geom_histogram(bins=45, fill="gray", col="white") +  
  labs(title = "13. Car Guarantee Period Histogram")
```



**Evaluación Gráfica 1 (Precio):** Esta gráfica demuestra un rango de precio sesgado hacia la izquierda, probando que la mayoría de los autos manejan un precio al rededor de \$10,000.

**Evaluación Gráfica 2 (Edad del auto):** Esta gráfica demuestra que en la base de datos existe un sesgo hacia la derecha que muestra que la mayoría de los autos tienen más de 60 meses de existencia.

**Evaluación Gráfica 3 (Mes de manufactura):** En la gráfica podemos observar como en los primeros meses del año, llega a su punto máximo de manufactura. El ultimo mes del año es el más bajo en comparación de los demás. **Evaluación Gráfica 4 (Año de manufactura):** Esta gráfica representa que la mayoría de los autos en la base de datos salieron de fábrica antes del año 2000.

**Evaluación Gráfica 5 (Kilometraje):** Esta gráfica muestra un sesgo hacía la izquierda, demostrando que la mayoría de los autos en la base de datos tienen un kilometraje por debajo de los 15,000 km.

**Evaluación Gráfica 6 (Caballos de fuerza):** Se puede apreciar en la grafica que la distribución de los datos ronda por los 100 caballos de fuerza y hay otros que llegan a rondar por los 200.

**Evaluación Gráfica 7 (Centímetros cúbicos):** En esta gráfica podemos observar que la mayoría de los carros están por debajo de los 4000 centímetros cúbicos. La mayoría se sitúa entre los 1500 y los 2000 centímetros cúbicos **Evaluación Gráfica 8 (Puertas):** En esta gráfica se puede observar que únicamente existen autos con 3, 4 y 5 puertas, la mayoría cuentan con 3 o 5.

**Evaluación Gráfica 9 (Cylindros):** Con esta gráfica se puede observar que la totalidad de los carros en el dataset cuentan con 4 cilindros.

**Evaluación Gráfica 10 (Granaje):** Podemos observar que casi en su totalidad de los carros cuentas con 5 engranajes y una pequeña cantidad cuenta con 6

**Evaluación Gráfica 11 (Impuestos):** Se puede observar que la mayoría de los autos en el dataset cuentan con impuestos menores a 100.

**Evaluación Gráfica 12 (Peso):** Esta gráfica muestra un sesgo hacía la izquierda, probando que la mayoría de los autos mantienen un peso menor a 1,200 kg.

**Evaluación Gráfica 13 (Garantía):** Esta gráfica (que se puede considerar irrelevante en este caso), demuestra que la mayoría de los autos en la base de datos tienen garantía mínima.

Ahora se obtendrá la matriz de correlación para analizar la multicolinealidad, ya que las variables de ingreso deben de ser independientes entre sí, que permitirá definir en los futuros pasos del modelo CRISP DM los modelos de predicción. Esto es, entre mayor el diámetro del círculo mostrado en la gráfica, mayor es la correlación entre las variables. Para este caso en la matriz de correlación se decidió eliminar el atributo 'Cylinders' ya que todos los datos de la variable representaban un mismo valor, por lo cual no sería de relevancia.

Price

Age\_08\_04

Mfg\_Month

Mfg\_Year

KM

HP

cc

Doors

Gears

Quarterly\_Tax

Weight

Guarantee\_Period

Price

Age\_08\_04

Mfg\_Month

Mfg\_Year

KM

HP

cc

Doors

Gears

Quarterly\_Tax

Weight

Guarantee\_Period

1

0.8

0.6

0.4

0.2

0

-0.2

-0.4

-0.6

-0.8

-1



A partir de este análisis, se puede identificar claramente que las variables que más se relacionan con el precio son el kilometraje, año de manufactura del auto, edad del auto. Se puede observar que el precio del auto y la edad del mismo tienen una relación inversamente proporcional. Esto es, que conforme la edad del auto incrementa, el precio disminuye, y viceversa.

El kilometraje y el precio muestran también una relación inversamente proporcional. Entre más bajo sea el kilometraje de un auto, más alto es el precio de venta.

Por el contrario, el año de manufactura del auto y el precio muestran una relación proporcional. Esto es, entre el año aumente (entre más nuevo sea el carro), el precio aumenta de manera proporcional.

Por último, se encontró una relación proporcional entre el peso y el precio del carro. Entre más pesado sea el carro, mayor su precio de venta.

Las otras variables restantes no muestran significancia con el precio.

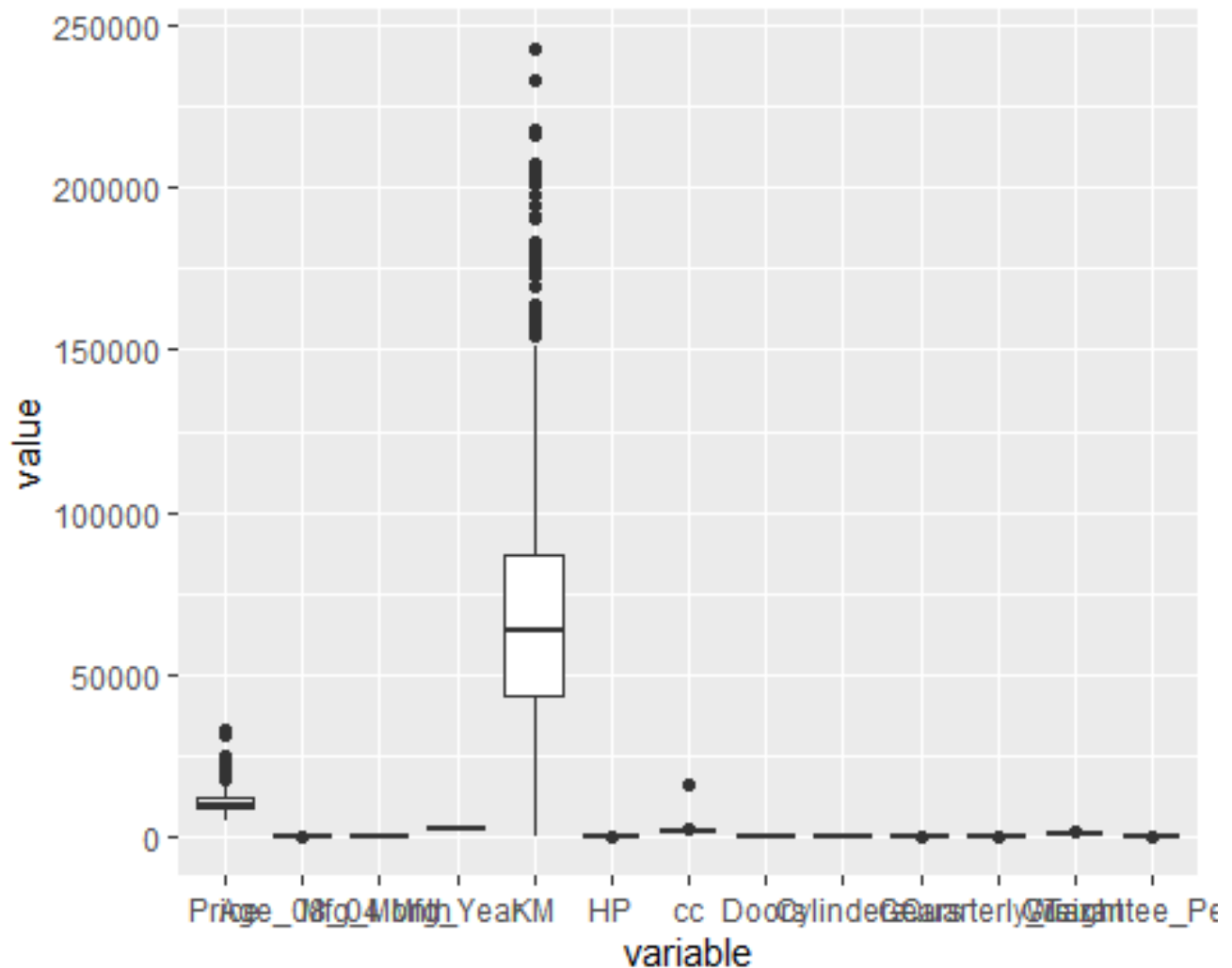
## PASO 6: Boxplot de las variables

Ahora se obtendrán los boxplots. En este tipo de gráfico se obtendrán los cuartiles, la media y los valores atípicos (outliers).

Nuevamente, con el uso de la función 'ggplot', se muestran las distribuciones de valores de datos numéricos, especialmente cuando se comparan entre varios grupos. Están diseñados para proporcionar información de alto nivel de un vistazo, ofreciendo información general sobre la simetría, la inclinación, la varianza y los valores atípicos de un grupo de datos.

```
ggplot(data=melt(corolla.df[, -c(1,2)]), mapping=aes(x=variable, y=value)) +  
geom_boxplot()
```

```
## No id variables; using all as measure variables
```



De este resultado se puede observar que la variable 'kilometraje' tiene varios outliers y es la variable con valores más dispersos. En el atributo 'precio' también se identifican varios outliers y una meda de precio diferente a los demás atributos.

Para analizar este fenómeno detenidamente en el siguiente paso se realiza el análisis de valores atípicos sobre todas las variables cuantitativas para detectar estos valores y analizarlos.

## PASO 7: Detección de outliers

Este siguiente paso consiste en la detección de outliers, el cual es esencial para comprender y validar la calidad de nuestro conjunto de datos, y proporcionar información valiosa sobre patrones y comportamientos inusuales en el conjunto de datos.

Solamente se tomarán en cuenta atributos de valores que se consideran que los outliers representarían un impacto considerable en el resultado final; y que además podrían representar errores.

```
## Price outliers
Out1 = boxplot.stats(corolla.df$Price)$out
Price_Out=which(corolla.df$Price%in%c(Out1))
Id_Price=corolla.df$Id[(Price_Out)]
Model_price_o=corolla.df$Model[(Price_Out)]
Price_Out_List=data.frame(
  ID=Id_Price,
  Model=Model_price_o,
  Price=Out1
)
Price_Out_List
```

```
##      ID
## 1      8
## 2      9
## 3     11
## 4     12
## 5     13
## 6     14
## 7     15
## 8     16
## 9     17
## 10     18
## 11     27
## 12     30
## 13     46
## 14     47
## 15     49
## 16     50
## 17     51
## 18     53
## 19     54
## 20     59
## 21     63
## 22     64
## 23     65
## 24     67
## 25     69
## 26     73
## 27     75
## 28     77
## 29     78
## 30     81
## 31     82
```

##	32	84
##	33	86
##	34	88
##	35	90
##	36	92
##	37	93
##	38	95
##	39	96
##	40	99
##	41	100
##	42	101
##	43	103
##	44	104
##	45	105
##	46	107
##	47	108
##	48	109
##	49	111
##	50	112
##	51	113
##	52	114
##	53	115
##	54	116
##	55	117
##	56	118
##	57	119
##	58	120
##	59	121
##	60	122
##	61	123
##	62	125
##	63	127
##	64	130
##	65	138
##	66	140
##	67	141
##	68	143
##	69	144
##	70	145
##	71	146
##	72	148
##	73	149
##	74	150
##	75	151
##	76	153
##	77	154
##	78	155
##	79	156
##	80	158
##	81	159

## 82 160  
## 83 161  
## 84 162  
## 85 163  
## 86 164  
## 87 165  
## 88 166  
## 89 167  
## 90 168  
## 91 169  
## 92 170  
## 93 171  
## 94 172  
## 95 173  
## 96 174  
## 97 175  
## 98 176  
## 99 177  
## 100 178  
## 101 179  
## 102 180  
## 103 181  
## 104 182  
## 105 183  
## 106 184  
## 107 185  
## 108 186  
## 109 187  
## 110 526  
##

Model

## 1 TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-  
Doors  
## 2 ?TOYOTA Corolla 1800 T SPORT VVT I 2/3-  
Doors  
## 3 TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 2/3-  
Doors  
## 4 TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT BNS 2/3-  
Doors  
## 5 TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-  
Doors  
## 6 TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-  
Doors  
## 7 TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-  
Doors  
## 8 TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-  
Doors  
## 9 ?TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-  
Doors  
## 10 ?TOYOTA Corolla 1.6 VVTI Linea Terra Comfort 2/3-

Doors	
## 11	?TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-
Doors	
## 12	TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-
Doors	
## 13	?TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-
Doors	
## 14	TOYOTA Corolla 2.0 D4D 90 5DR TERRA 4/5-
Doors	
## 15	?TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-
Doors	
## 16	TOYOTA Corolla 1.8 16V VVTI 5DR T SPORT 4/5-
Doors	
## 17	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-
Doors	
## 18	TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-
Doors	
## 19	TOYOTA Corolla 1.8 VVTI-i T-Sport 4/5-
Doors	
## 20	?TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-
Doors	
## 21	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 22	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 23	TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-
Doors	
## 24	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 25	TOYOTA Corolla 2.0 D4D 110 5DR SOL 4/5-
Doors	
## 26	?TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 27	?TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 28	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 29	TOYOTA Corolla 1.6 16V VVT I 5DR TERRA 4/5-
Doors	
## 30	TOYOTA Corolla 1.6 5drs 1 4/5-
Doors	
## 31	TOYOTA Corolla 1.6 16V VVT I 5DR TERRA 4/5-
Doors	
## 32	TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-
Doors	
## 33	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 34	TOYOTA Corolla 2.0 D4D 90 3DR TERRA COMF BNS 2/3-
Doors	
## 35	TOYOTA Corolla 2.0 D4D 110 3DR SOL 2/3-

Doors	
## 36	?TOYOTA Corolla 2.0 D4D 90 3DR SOL BNS 2/3-
Doors	
## 37	TOYOTA Corolla 2.0 D 4D Linea Sol 3 Doors 2/3-
Doors	
## 38	TOYOTA Corolla 1.6 16v VVT-i Linea Sol 2/3-
Doors	
## 39	TOYOTA Corolla 2.0 D4D Linea Sol 3 Doors 2/3-
Doors	
## 40	TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-
Doors	
## 41	TOYOTA Corolla 1.4 16V VVT I 3DR TERRA COMFORT 2/3-
Doors	
## 42	?TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-
Doors	
## 43	TOYOTA Corolla 1.4 VVT-i Sol 2/3-
Doors	
## 44	TOYOTA Corolla 1.6 16V VVT I 3DR TERRA COMFORT 2/3-
Doors	
## 45	TOYOTA Corolla 1.6 16V VVT I 3DR TERRA COMF AUT 2/3-
Doors	
## 46	?TOYOTA Corolla 1.6 VVT-i Linea Terra Comfort 2/3-
Doors	
## 47	TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-
Doors	
## 48	?TOYOTA Corolla 1.6 16V VVT I 3DR TERRA COMFORT 2/3-
Doors	
## 49	?TOYOTA Corolla VERSO 2.0 D4D SOL (7)
BNS MPV	
## 50	TOYOTA Corolla VERSO 2.0 D4D SOL
(7) MPV	
## 51	TOYOTA Corolla VERSO 2.0 D4D SOL
(7) MPV	
## 52	?TOYOTA Corolla VERSO 2.0 D4D LINEA
SOL MPV	
## 53	TOYOTA Corolla VERSO 2.0 D4D LINEA
SOL MPV	
## 54	TOYOTA Corolla 2.0 D4D 116 5DR LUNA 4/5-
Doors	
## 55	TOYOTA Corolla 2.0 D4D 90 WAGON SOL
Stationwagen	
## 56	?TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-
Doors	
## 57	TOYOTA Corolla 1.6 16V VVT I 3DR TERRA COMFORT 2/3-
Doors	
## 58	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-
Doors	
## 59	TOYOTA Corolla 2.0 D4D 110 5DR EXEC BNS 4/5-
Doors	
## 60	TOYOTA Corolla 1.6 16V VVT I 5DR TERRA COMFORT 4/5-

Doors	
## 61	?TOYOTA Corolla 2.0 D4D Linea Terra Comfort 4/5-
Doors	
## 62	TOYOTA Corolla 1.6 16V VWT I 5DR SOL 4/5-
Doors	
## 63	TOYOTA Corolla 1.6 16V VWT I 5DR EXEC 4/5-
Doors	
## 64	TOYOTA Corolla 1.4 16V VWT I 5DR SOL 4/5-
Doors	
## 65	TOYOTA Corolla 1.4 16V VWT I 5DR SOL 4/5-
Doors	
## 66	TOYOTA Corolla Luna 5drs D4D116 PK 4/5-
Doors	
## 67	TOYOTA Corolla 1.6i 16V VWT I SOL + Navigatie 4/5-
Doors	
## 68	?TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 4/5-
Doors	
## 69	?TOYOTA Corolla 1.6 16V VWT I 5DR SOL 4/5-
Doors	
## 70	TOYOTA Corolla 1.6 16V VWT I SEDAN TERRA COMF 4/5-
Doors	
## 71	TOYOTA Corolla 1.6 16V VWT I 5DR SOL 4/5-
Doors	
## 72	?TOYOTA Corolla 1.6 16V VWT I 5DR TERRA COMF AUT 4/5-
Doors	
## 73	TOYOTA Corolla 1.6 16V VWT I 5DR EXEC BNS 4/5-
Doors	
## 74	TOYOTA Corolla 1.6 16V VWT I 5DR SOL 4/5-
Doors	
## 75	TOYOTA Corolla 1.6 16V VWT I SEDAN SOL AUT4 4/5-
Doors	
## 76	TOYOTA Corolla 1.6 16V VWT I 5DR SOL BNS 4/5-
Doors	
## 77	TOYOTA Corolla 1.4 16V VWT I 5DR LUNA BNS 4/5-
Doors	
## 78	?TOYOTA Corolla 1.6 16V VWT I 5DR TERRA COMFORT 4/5-
Doors	
## 79	TOYOTA Corolla 1.6 16V VWT I 5DR EXEC 4/5-
Doors	
## 80	TOYOTA Corolla 1.6 16V VWT I SEDAN SOL 4/5-
Doors	
## 81	TOYOTA Corolla 1.6 16V VWT I 5DR TERRA COMFORT 4/5-
Doors	
## 82	TOYOTA Corolla 1.6 16V VWT I 5DR SOL 4/5-
Doors	
## 83	TOYOTA Corolla 1.6 16V VWT I 5DR SOL AUT4 4/5-
Doors	
## 84	TOYOTA Corolla 1.6 16V VWT I SEDAN SOL 4/5-
Doors	
## 85	?TOYOTA Corolla 1.6 16v VWTi Linea Sol 4/5-



Doors	
## 86	TOYOTA Corolla 1.6 16V VVT I 5DR TERRA COMFORT 4/5-
Doors	
## 87	TOYOTA Corolla 1.6 16v VVTi Terra Comfort 4/5-
Doors	
## 88	TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-
Doors	
## 89	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 90	TOYOTA Corolla 1.6 16V VVT I 5DR EXEC 4/5-
Doors	
## 91	TOYOTA Corolla 1.6 16V VVT I SEDAN SOL BNS AUT4 4/5-
Doors	
## 92	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 93	TOYOTA Corolla 1.4-16v VVT-i Linea Terra Comfort NIEUW AIRCO 4/5-
Doors	
## 94	TOYOTA Corolla 1.6-16v VVT-i Linea Terra Comfort AIRCO NIEUW 5DRS 4/5-
Doors	
## 95	TOYOTA Corolla 1.6-16v VVT-i Executive B.edition 4/5-
Doors	
## 96	TOYOTA Corolla 1.4 16V VVT I 5DR LUNA 4/5-
Doors	
## 97	?TOYOTA Corolla 1.4 16V VVT I 5DR LUNA 4/5-
Doors	
## 98	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 99	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 100	TOYOTA Corolla 1.6 16V VVT I 5DR TERRA COMFORT 4/5-
Doors	
## 101	TOYOTA Corolla 1.6 16V VVT I 5DR TERRA 4/5-
Doors	
## 102	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 103	TOYOTA Corolla 1.6 16V VVT I 5DR EXEC 4/5-
Doors	
## 104	TOYOTA Corolla 1.6 VVT-I 5DR LINEA TERRA 4/5-
Doors	
## 105	TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-
Doors	
## 106	TOYOTA Corolla 1.4 16V 5DR TERRA 4/5-
Doors	
## 107	TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors	
## 108	?TOYOTA Corolla 1.4-16v VVT-i Linea Terra Comfort NIEUW AIRCO 4/5-
Doors	
## 109	TOYOTA Corolla 1.6-16v VVT-i Linea Terra Comfort NIEUW AIRCO 5drs 4/5-
Doors	
## 110	?TOYOTA Corolla 1.6 HB LINEA TERRA 2/3-

## Doors

##	Price
## 1	18600
## 2	21500
## 3	20950
## 4	19950
## 5	19600
## 6	21500
## 7	22500
## 8	22000
## 9	22750
## 10	17950
## 11	17495
## 12	17950
## 13	19000
## 14	17950
## 15	17950
## 16	21950
## 17	17950
## 18	20500
## 19	21950
## 20	18950
## 21	18750
## 22	17950
## 23	17950
## 24	18950
## 25	22250
## 26	18950
## 27	19950
## 28	18750
## 29	18450
## 30	18950
## 31	17250
## 32	17950
## 33	17450
## 34	17950
## 35	21950
## 36	22250
## 37	19950
## 38	18900
## 39	19950
## 40	18750
## 41	17450
## 42	18990
## 43	18500
## 44	18500
## 45	19450
## 46	18800
## 47	17450
## 48	17950

##	49	32500
##	50	31000
##	51	31275
##	52	24950
##	53	24950
##	54	22950
##	55	24990
##	56	21950
##	57	17900
##	58	19250
##	59	22250
##	60	18950
##	61	19950
##	62	18950
##	63	21750
##	64	17950
##	65	18450
##	66	23000
##	67	19900
##	68	23950
##	69	19950
##	70	18500
##	71	18950
##	72	20500
##	73	24500
##	74	19450
##	75	20950
##	76	19950
##	77	18450
##	78	19500
##	79	21750
##	80	19500
##	81	18900
##	82	19750
##	83	19750
##	84	18950
##	85	20750
##	86	19600
##	87	19500
##	88	17650
##	89	19950
##	90	19950
##	91	20950
##	92	20500
##	93	17795
##	94	18245
##	95	23750
##	96	19500
##	97	18950
##	98	21950

```
## 99 19950
## 100 18950
## 101 19950
## 102 21950
## 103 22500
## 104 18500
## 105 18700
## 106 21125
## 107 21500
## 108 17795
## 109 18245
## 110 18950
```

Tal como se determinó en el análisis 'Boxplot', existen muchos outliers dentro del atributo 'Price'. Esto se debe a la gran variedad de precios que hay en el mercado. De acuerdo con la función 'skim' empleada anteriormente, la media de los precios es alrededor de \$10,730. Mientras tanto, en este análisis se observan valores que se diferencian mucho de la media.

### ## Age Outliers

```
Out2 = boxplot.stats(corolla.df$Age_08_04)$out
Age_Out=which(corolla.df$Age_08_04%in%c(Out2))
Id_Age=corolla.df$Id[(Age_Out)]
Model_age_o=corolla.df$Model[(Age_Out)]
Age_Out_List=data.frame(
  ID=Id_Age,
  Model=Model_age_o,
  Age=Out2
)
Age_Out_List
```

```
##      ID
## 1 111
## 2 112
## 3 113
## 4 184
## 5 185
## 6 186
## 7 187
##
Model
## 1      ?TOYOTA Corolla VERSO 2.0 D4D SOL (7) BNS
MPV
## 2      TOYOTA Corolla VERSO 2.0 D4D SOL (7)
MPV
## 3      TOYOTA Corolla VERSO 2.0 D4D SOL (7)
MPV
## 4      TOYOTA Corolla 1.4 16V 5DR TERRA 4/5-
```

```

Doors
## 5          TOYOTA Corolla 1.6 16V VVT I 5DR SOL 4/5-
Doors
## 6    ?TOYOTA Corolla 1.4-16v VVT-i Linea Terra Comfort NIEUW AIRCO 4/5-
Doors
## 7 TOYOTA Corolla 1.6-16v VVT-i Linea Terra Comfort NIEUW AIRCO 5drs 4/5-
Doors
##   Age
## 1   4
## 2   4
## 3   4
## 4   2
## 5   2
## 6   1
## 7   1

```

Los valores atípicos para el atributo de 'edad' son pocos. Esto puede ocurrir debido a que la mayoría de los autos en la base de datos tienen un promedio de 55 meses de existencia. En la vida real, es raro ver carros nuevos en venta en un concesionario de autos.

#### ## CC Outliers

```

Out2 = boxplot.stats(corolla.df$cc)$out
cc_Out=which(corolla.df$cc%in%c(Out2))
Id_cc=corolla.df$Id[(cc_Out)]
Model_cc_o=corolla.df$Model[(cc_Out)]
cc_Out_List=data.frame(
  ID=Id_cc,
  Model=Model_cc_o,
  cc=Out2
)
cc_Out_List

```

##	ID	Model	cc
## 1	1	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	2000
## 2	2	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	2000
## 3	3	?TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	2000
## 4	4	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	2000
## 5	5	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	2000
## 6	6	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	2000
## 7	7	?TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-Doors	2000
## 8	8	TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-Doors	2000
## 9	44	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	2000
## 10	45	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	2000
## 11	46	?TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors	2000
## 12	47	TOYOTA Corolla 2.0 D4D 90 5DR TERRA 4/5-Doors	2000
## 13	49	?TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	2000
## 14	51	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	2000
## 15	69	TOYOTA Corolla 2.0 D4D 110 5DR SOL 4/5-Doors	2000
## 16	81	TOYOTA Corolla 1.6 5drs 1 4/5-Doors	16000
## 17	88	TOYOTA Corolla 2.0 D4D 90 3DR TERRA COMF BNS 2/3-Doors	2000

## 18	90	TOYOTA Corolla 2.0 D4D 110 3DR SOL 2/3-Doors	2000
## 19	92	?TOYOTA Corolla 2.0 D4D 90 3DR SOL BNS 2/3-Doors	2000
## 20	93	TOYOTA Corolla 2.0 D 4D Linea Sol 3 Doors 2/3-Doors	1995
## 21	96	TOYOTA Corolla 2.0 D4D Linea Sol 3 Doors 2/3-Doors	1995
## 22	111	?TOYOTA Corolla VERSO 2.0 D4D SOL (7) BNS MPV	2000
## 23	112	TOYOTA Corolla VERSO 2.0 D4D SOL (7) MPV	2000
## 24	113	TOYOTA Corolla VERSO 2.0 D4D SOL (7) MPV	2000
## 25	114	?TOYOTA Corolla VERSO 2.0 D4D LINEA SOL MPV	2000
## 26	115	TOYOTA Corolla VERSO 2.0 D4D LINEA SOL MPV	2000
## 27	116	TOYOTA Corolla 2.0 D4D 116 5DR LUNA 4/5-Doors	2000
## 28	117	TOYOTA Corolla 2.0 D4D 90 WAGON SOL Stationwagen	2000
## 29	118	?TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors	2000
## 30	120	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	2000
## 31	121	TOYOTA Corolla 2.0 D4D 110 5DR EXEC BNS 4/5-Doors	2000
## 32	123	?TOYOTA Corolla 2.0 D4D Linea Terra Comfort 4/5-Doors	2000
## 33	140	TOYOTA Corolla Luna 5drs D4D116 PK 4/5-Doors	2000
## 34	189	TOYOTA Corolla 2.0 D4D SEDAN SOL 4/5-Doors	2000
## 35	190	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	2000
## 36	199	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	2000
## 37	213	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	2000
## 38	216	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	2000
## 39	245	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	2000
## 40	270	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	2000
## 41	384	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 42	387	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 43	393	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 44	400	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	2000
## 45	405	TOYOTA Corolla 2.0D 4/5-Doors	2000
## 46	419	?TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 47	421	?TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 48	468	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 49	490	TOYOTA Corolla LIFTBACK 1.9 D Linea Terra 4/5-Doors	2000
## 50	605	TOYOTA Corolla 2.0 d HB Diesel 2/3-Doors	2000
## 51	607	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 52	608	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 53	610	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 54	611	TOYOTA Corolla 2.0 DSL LIFTB LINEA LUNA 4/5-Doors	2000
## 55	612	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 56	613	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 57	616	TOYOTA Corolla 2.0 DSL HATCHB LINEA LUNA 2/3-Doors	2000
## 58	617	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 59	619	TOYOTA Corolla 2.0diesel Stationwagen	2000
## 60	620	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 61	621	TOYOTA Corolla 2.0D TERRA LIFTBACK 4/5-Doors	2000
## 62	623	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 63	625	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 64	629	TOYOTA Corolla 2.0 DSL HATCHB LINEA LUNA 2/3-Doors	2000
## 65	630	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 66	631	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 67	632	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000

## 68	634	TOYOTA Corolla 2.0D XL Sedan 4/5-Doors	2000
## 69	635	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 70	636	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 71	638	TOYOTA Corolla Sw 20D Linea Terra Stationwagen	2000
## 72	639	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 73	640	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 74	643	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 75	648	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA Stationwagen	2000
## 76	651	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 77	658	TOYOTA Corolla 2.0 D Linea Terra 3d 2/3-Doors	2000
## 78	664	TOYOTA Corolla 2.0 SDN LINEA TERRA DSL 4/5-Doors	2000
## 79	683	TOYOTA Corolla 2.0 LB LINEA TERRA D 4/5-Doors	2000
## 80	686	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 81	700	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 82	711	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 83	714	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 84	772	TOYOTA Corolla 2.0 D Linea luna Wagon Stationwagen	1975
## 85	789	TOYOTA Corolla 2.0 DSL HATCHB LINEA LUNA 2/3-Doors	2000
## 86	823	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 87	870	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 88	1049	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA COMM Anders	2000
## 89	1050	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA Stationwagen	2000
## 90	1051	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 91	1053	TOYOTA Corolla 2.0D LINEA TERRA+airco 2/3-Doors	2000
## 92	1054	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 93	1055	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 94	1056	TOYOTA Corolla 2.0 DSL HATCHB LINEA LUNA 2/3-Doors	2000
## 95	1057	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA Stationwagen	2000
## 96	1058	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 97	1059	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 98	1060	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 99	1063	TOYOTA Corolla 2.0 ST WGN LINEA TERRA DSL Stationwagen	2000
## 100	1064	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	2000
## 101	1068	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 102	1071	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 103	1074	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 104	1075	TOYOTA Corolla 2.0 DSL LIFTB LINEA LUNA 4/5-Doors	2000
## 105	1076	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA Stationwagen	2000
## 106	1078	TOYOTA Corolla 2.0D Linea Terra 2/3-Doors	2000
## 107	1079	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 108	1083	TOYOTA Corolla 2.0 DSL	2000
## 109	1085	TOYOTA Corolla 2.0 D Linea Terra Liftback 4/5-Doors	2000
## 110	1086	TOYOTA Corolla 2.0 D Linea Terra Liftback 4/5-Doors	2000
## 111	1087	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 112	1088	TOYOTA Corolla 2.0 ST WGN LINEA TERRA DSL Stationwagen	2000
## 113	1091	TOYOTA Corolla 2.0 DSL LIFTB LINEA LUNA 4/5-Doors	2000
## 114	1098	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 115	1103	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	2000
## 116	1114	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	2000
## 117	1115	TOYOTA Corolla 2.0 LB LINEA TERRA DSL 4/5-Doors	2000

```
## 118 1124      TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors 2000
## 119 1202      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 2000
## 120 1209      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 2000
## 121 1223      TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors 2000
## 122 1327      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 2000
## 123 1337      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 2000
```

Dentro de esta lista de outliers, se puede observar que el atributo 'cc' muestra outliers equivalentes a 2000. Estos datos son considerados outliers porque estan fuera del cuarto cuartil.

### ## KM Outliers

```
Out3 = boxplot.stats(corolla.df$KM)$out
KM_Out=which(corolla.df$KM%in%c(Out3))
Id_KM=corolla.df$Id[(KM_Out)]
Model_KM_o=corolla.df$Model[(KM_Out)]
KM_Out_List=data.frame(
  ID=Id_KM,
  Model=Model_KM_o,
  KM=Out3
)
KM_Out_List
```

##	ID	Model	KM
## 1	188	TOYOTA Corolla 1.9 D HATCHB SOL 2/3-Doors	243000
## 2	189	TOYOTA Corolla 2.0 D4D SEDAN SOL 4/5-Doors	180638
## 3	190	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	179860
## 4	191	TOYOTA Corolla 1.6 16V VVT I HATCHB TERRA 2/3-Doors	178858
## 5	192	TOYOTA Corolla 1.9 D HATCHB TERRA 2/3-Doors	161000
## 6	193	TOYOTA Corolla 1.8D Stationwagen	158320
## 7	381	?TOYOTA Corolla 1.9 D HATCHB SOL 2/3-Doors	216000
## 8	382	TOYOTA Corolla 1.9 D SEDAN TERRA 4/5-Doors	198167
## 9	383	TOYOTA Corolla 1.9 D HATCHB TERRA 2/3-Doors	176000
## 10	384	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	174139
## 11	385	TOYOTA Corolla 1.9 D Sedan 4/5-Doors	174000
## 12	607	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	205000
## 13	608	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	204250
## 14	609	TOYOTA Corolla 1.3 16V SEDAN LINEA TERRA 4/5-Doors	194545
## 15	610	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	190900
## 16	611	TOYOTA Corolla 2.0 DSL LIFTB LINEA LUNA 4/5-Doors	183500
## 17	612	TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors	176184
## 18	613	TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors	174833
## 19	614	TOYOTA Corolla 1.3 16V HATCHB LINEA TERRA 2/3-Doors	155720
## 20	615	TOYOTA Corolla 1.6 Linea Terra Comfort 4/5-Doors	155500
## 21	616	TOYOTA Corolla 2.0 DSL HATCHB LINEA LUNA 2/3-Doors	154783
## 22	617	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	154462
## 23	1049	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA COMM Anders	232940
## 24	1050	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA Stationwagen	218118
## 25	1051	TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors	217764
## 26	1052	TOYOTA Corolla 1.6 LB *G3* AIRCO 4/5-Doors	207114



```

## 27 1053      TOYOTA Corolla 2.0D LINEA TERRA+airco 2/3-Doors 203254
## 28 1054      TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors 200732
## 29 1055      TOYOTA Corolla 2.0 DSL LIFTB LINEA TERRA 4/5-Doors 197501
## 30 1056      TOYOTA Corolla 2.0 DSL HATCHB LINEA LUNA 2/3-Doors 194765
## 31 1057      TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA Stationwagen 191620
## 32 1058      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 183277
## 33 1059      TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors 180378
## 34 1060      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 178800
## 35 1061      TOYOTA Corolla 1.6 16V LIFTB LINEA LUNA 4/5-Doors 176177
## 36 1062      TOYOTA Corolla 1.3 16V LIFTB LINEA TERRA 4/5-Doors 176000
## 37 1063      TOYOTA Corolla 2.0 ST WGN LINEA TERRA DSL Stationwagen 172980
## 38 1064      TOYOTA Corolla 2.0 DSL SEDAN LINEA TERRA 4/5-Doors 170000
## 39 1065      TOYOTA Corolla 1.6 HB LINEA TERRA 2/3-Doors 164000
## 40 1066      TOYOTA Corolla 1.3 16V HATCHB LINEA TERRA 2/3-Doors 161775
## 41 1067      TOYOTA Corolla 1.3 16V HATCHB LINEA TERRA 2/3-Doors 160000
## 42 1068      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 159968
## 43 1069      TOYOTA Corolla 1.6 16V HATCHB LINEA TERRA 2/3-Doors 159908
## 44 1070      TOYOTA Corolla 1.3 16V HATCHB LINEA TERRA 2/3-Doors 158492
## 45 1071      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 156204
## 46 1072      TOYOTA Corolla 1.3 16V LIFTB LINEA TERRA 4/5-Doors 156169
## 47 1073      TOYOTA Corolla 1.6 3D Terra/Comf. 2/3-Doors 155210
## 48 1074      TOYOTA Corolla 2.0 DSL HATCHB LINEA TERRA 2/3-Doors 155000
## 49 1075      TOYOTA Corolla 2.0 DSL LIFTB LINEA LUNA 4/5-Doors 154900

```

Los valores atípicos para el atributo kilometraje, tal como se predijo en el análisis 'boxplot', son varios. Estos valores se ven alejados de la media que de acuerdo a la función 'skim' empleada anteriormente es de 68,533 km, y en esta lista se encuentran valores tan bajos como 15,000 km.

### ## Weight Outliers

```

Out4 = boxplot.stats(corolla.df$Weight)$out
Weight_Out=which(corolla.df$Weight%in%c(Out4))
Id_Weight=corolla.df$Id[(Weight_Out)]
Model_Weight_o=corolla.df$Model[(Weight_Out)]
Weight_Out_List=data.frame(
  ID=Id_Weight,
  Model=Model_Weight_o,
  Weight=Out4
)
Weight_Out_List

```

##	ID	Model	Weight
## 1	1	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	1165
## 2	2	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	1165
## 3	3	?TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	1165
## 4	4	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	1165
## 5	5	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	1170
## 6	6	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	1170
## 7	7	?TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-Doors	1245
## 8	8	TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-Doors	1245

## 9	9	?TOYOTA Corolla 1800 T SPORT VVT I 2/3-Doors	1185
## 10	11	TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 2/3-Doors	1185
## 11	12	TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT BNS 2/3-Doors	1185
## 12	13	TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-Doors	1185
## 13	14	TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-Doors	1185
## 14	15	TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-Doors	1185
## 15	16	TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-Doors	1185
## 16	17	?TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT 2/3-Doors	1185
## 17	22	TOYOTA Corolla 1.6 16V VVT I 3DR SOL AUT4 2/3-Doors	1170
## 18	44	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	1255
## 19	45	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	1255
## 20	46	?TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors	1270
## 21	47	TOYOTA Corolla 2.0 D4D 90 5DR TERRA 4/5-Doors	1255
## 22	49	?TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	1255
## 23	50	TOYOTA Corolla 1.8 16V VVTLI 5DR T SPORT 4/5-Doors	1195
## 24	51	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	1255
## 25	53	TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-Doors	1180
## 26	54	TOYOTA Corolla 1.8 VVTL-i T-Sport 4/5-Doors	1195
## 27	55	?TOYOTA Corolla 1.6 16V VVT I 5DR TERRA AUT4 4/5-Doors	1165
## 28	59	?TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-Doors	1180
## 29	69	TOYOTA Corolla 2.0 D4D 110 5DR SOL 4/5-Doors	1275
## 30	81	TOYOTA Corolla 1.6 5drs 1 4/5-Doors	1180
## 31	84	TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-Doors	1180
## 32	88	TOYOTA Corolla 2.0 D4D 90 3DR TERRA COMF BNS 2/3-Doors	1245
## 33	90	TOYOTA Corolla 2.0 D4D 110 3DR SOL 2/3-Doors	1265
## 34	92	?TOYOTA Corolla 2.0 D4D 90 3DR SOL BNS 2/3-Doors	1260
## 35	93	TOYOTA Corolla 2.0 D 4D Linea Sol 3 Doors 2/3-Doors	1260
## 36	96	TOYOTA Corolla 2.0 D4D Linea Sol 3 Doors 2/3-Doors	1260
## 37	105	TOYOTA Corolla 1.6 16V VVT I 3DR TERRA COMF AUT 2/3-Doors	1155
## 38	111	?TOYOTA Corolla VERSO 2.0 D4D SOL (7) BNS MPV	1480
## 39	112	TOYOTA Corolla VERSO 2.0 D4D SOL (7) MPV	1480
## 40	113	TOYOTA Corolla VERSO 2.0 D4D SOL (7) MPV	1480
## 41	114	?TOYOTA Corolla VERSO 2.0 D4D LINEA SOL MPV	1320
## 42	115	TOYOTA Corolla VERSO 2.0 D4D LINEA SOL MPV	1320
## 43	116	TOYOTA Corolla 2.0 D4D 116 5DR LUNA 4/5-Doors	1270
## 44	117	TOYOTA Corolla 2.0 D4D 90 WAGON SOL Stationwagen	1280
## 45	118	?TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors	1270
## 46	120	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	1255
## 47	121	TOYOTA Corolla 2.0 D4D 110 5DR EXEC BNS 4/5-Doors	1275
## 48	123	?TOYOTA Corolla 2.0 D4D Linea Terra Comfort 4/5-Doors	1255
## 49	140	TOYOTA Corolla Luna 5drs D4D116 PK 4/5-Doors	1320
## 50	143	?TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 4/5-Doors	1185
## 51	148	?TOYOTA Corolla 1.6 16V VVT I 5DR TERRA COMF AUT 4/5-Doors	1165
## 52	161	TOYOTA Corolla 1.6 16V VVT I 5DR SOL AUT4 4/5-Doors	1180
## 53	189	TOYOTA Corolla 2.0 D4D SEDAN SOL 4/5-Doors	1160
## 54	190	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	1205
## 55	199	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	1205
## 56	213	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	1205
## 57	216	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	1170
## 58	223	TOYOTA Corolla 1.6 HB LINEA SOL 4/5-Doors	1615

##	59	245	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	1165
##	60	270	TOYOTA Corolla 2.0 D4D LIFTB SOL 4/5-Doors	1205
##	61	400	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	1165
##	62	605	TOYOTA Corolla 2.0 d HB Diesel 2/3-Doors	1260
##	63	658	TOYOTA Corolla 2.0 D Linea Terra 3d 2/3-Doors	1260
##	64	772	TOYOTA Corolla 2.0 D Linea luna Wagon Stationwagen	1155
##	65	964	TOYOTA Corolla	1480
##	66	1049	TOYOTA Corolla 2.0 DSL WAGON LINEA TERRA COMM Anders	1172

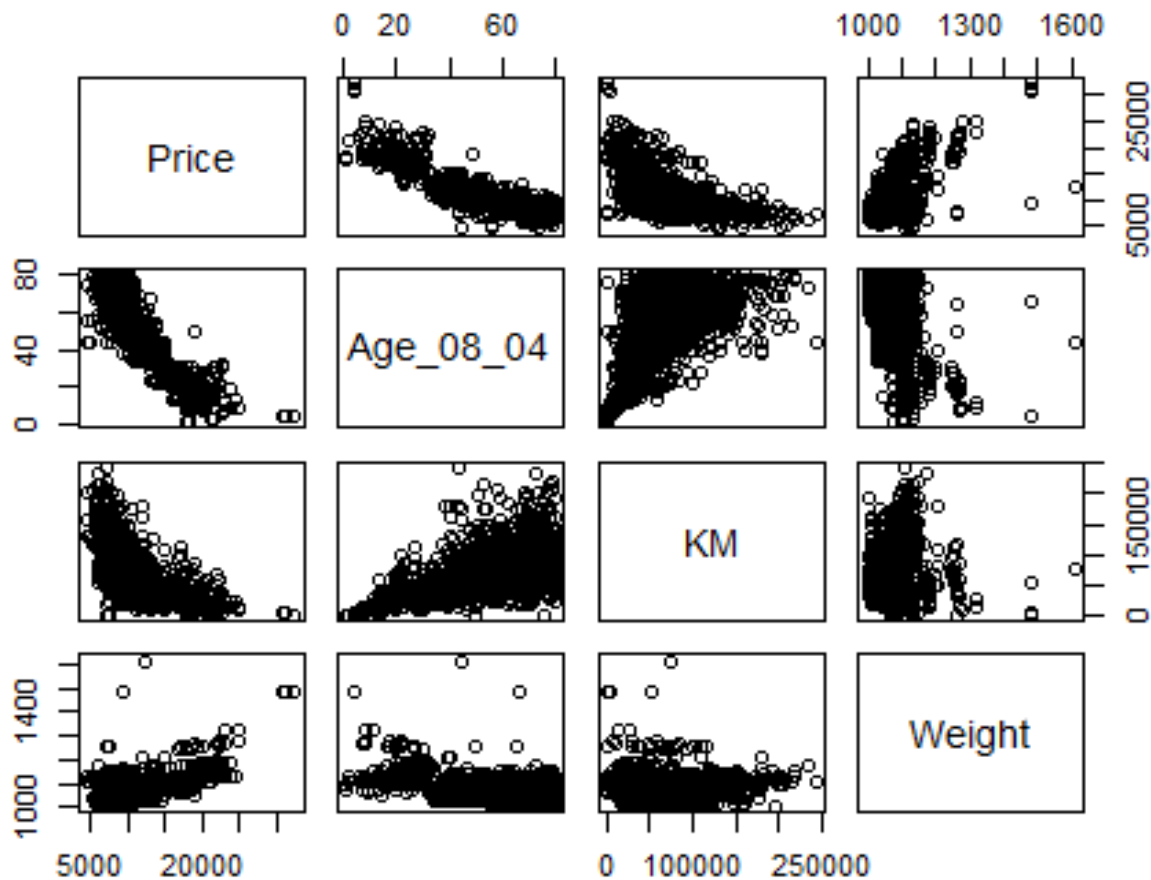
Aunque la variable 'weight' no es muy relevante a la hora de determinar el precio de un auto, existe un promedio de 1,072 kg. Los valores presentes en esta lista son considerados valores atípicos ya que sobre pasan este valor y puede llegar a afectar a la hora de tomar una decisión de compra.

## PASO 8: Elaboración de scatter plots

Las scatter plots tienen la función de observar de manera gráfica la posible relación lineal entre las variables. En esta ocasión solamente implementamos las variables que, de acuerdo a la matriz de correlación, tienen una relación significativa con el precio.

Aunque la variable 'Mfg\_Year' muestra una relación significativa con el precio, no se tomó en cuenta ya que esta esta directamente relacionada con la variable 'Age\_08\_04', de tal forma que al incluirla en este gráfico, se podría mostrar sesgo en el modelo generado futuramente.

```
pairs(corolla.df[,c(3,4,7,14)])
```



Visualmente se puede observar que hay una relación lineal entre 'Price' y 'Age\_08\_04'. Las variables son inversamente proporcionales de tal forma que entre más alto sea el valor de 'Age\_08\_04', menor será el valor de 'Price'.

La relación entre 'KM' y 'Price' no es tan fuerte al visualizarse en el gráfico, sin embargo muestran una relación inversamente proporcional. De tal manera que un auto con menor kilometraje tiene un precio más elevado.

La relación entre 'Weight' y 'Price' es directamente proporcional, de tal manera que los autos más pesados son más caros, sin embargo, visualmente la relación no es tan fuerte como la relación de 'Price' y 'Age\_08\_04'.

## PASO 9: Configuración de datos faltantes o errores

En este paso, aunque pertenece mayormente a la imputación de datos, decidimos implementar una búsqueda de datos faltantes en la base de datos. Excluimos los valores de 0 ya que hay atributos que tienen valores equivalentes a 0, específicamente aquellos relacionados con los atributos booleanos y de los cuales sí representan significado y no deben de ser reemplazados por NAs.

```
# convert whitespaces, empty or zeros into NAs (IF ANY)
corolla.df[corolla.df == " "] <- NA
corolla.df[corolla.df == ""] <- NA
# Check for changes
sapply(corolla.df, function(x) sum(is.na(x)))
```

##	Id	Model	Price	Age_08_04
##	0	0	0	0
##	Mfg_Month	Mfg_Year	KM	HP
##	0	0	0	0
##	cc	Doors	Cylinders	Gears
##	0	0	0	0
##	Quarterly_Tax	Weight	Guarantee_Period	
##	0	0	0	

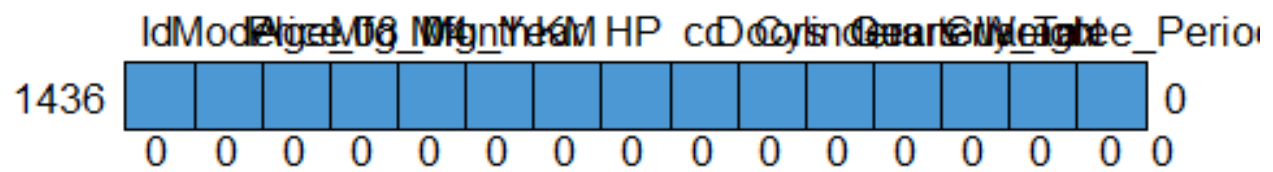
Nuevamente, se observa que en este subset no hay datos faltantes, por lo que no sería necesario realizar data imputation.

## PASO 10: Detección de datos faltantes o errors

En este paso aplicamos el uso de 'mice' para ubicar datos faltantes (NA's) en caso de haberlos.

```
md.pattern(corolla.df)
```

```
## /\      /\
## {  `---'  }
## {  0   0  }
## ==>  V <== No need for mice. This data set is completely observed.
## \  \|/  /
## `-----'
```



```
##      Id Model Price Age_08_04 Mfg_Month Mfg_Year KM HP cc Doors Cylinders
Gears
## 1436  1     1     1         1         1         1  1  1  1     1         1
1
##      0     0     0         0         0         0  0  0  0     0         0
0
##      Quarterly_Tax Weight Guarantee_Period
## 1436              1     1              1 0
##              0     0              0 0
```

```
md.pattern(ds)
```

```
## /\      /\
## {  `---'  }
## {  0  0  }
## ==>  V <== No need for mice. This data set is completely observed.
## \  \|/  /
## `-----'
```



## PASO 11: Resumen estadístico de las variables

Con uso de la función summary se pueden visualizar los datos estadísticos básicos para cada atributo.

```
summary(corolla.df)
```

```
##      Id      Model      Price      Age_08_04
## Min.   : 1.0   Length:1436   Min.    : 4350   Min.    : 1.00
## 1st Qu.: 361.8   Class :character   1st Qu.: 8450   1st Qu.:44.00
## Median : 721.5   Mode  :character   Median : 9900   Median :61.00
## Mean   : 721.6                Mean  :10731   Mean   :55.95
## 3rd Qu.:1081.2                3rd Qu.:11950   3rd Qu.:70.00
## Max.    :1442.0                Max.    :32500   Max.    :80.00
##   Mfg_Month   Mfg_Year      KM      HP
## Min.    : 1.000   Min.    :1998   Min.    : 1   Min.    : 69.0
## 1st Qu.: 3.000   1st Qu.:1998   1st Qu.: 43000 1st Qu.: 90.0
## Median : 5.000   Median :1999   Median : 63390 Median :110.0
## Mean    : 5.549   Mean    :2000   Mean    : 68533 Mean    :101.5
## 3rd Qu.: 8.000   3rd Qu.:2001   3rd Qu.: 87021 3rd Qu.:110.0
## Max.    :12.000   Max.    :2004   Max.    :243000 Max.    :192.0
##      cc      Doors      Cylinders      Gears      Quarterly_Tax
## Min.    : 1300   Min.    :2.000   Min.    :4   Min.    :3.000   Min.    :
19.00
## 1st Qu.: 1400   1st Qu.:3.000   1st Qu.:4   1st Qu.:5.000   1st Qu.:
69.00
## Median : 1600   Median :4.000   Median :4   Median :5.000   Median :
85.00
## Mean    : 1577   Mean    :4.033   Mean    :4   Mean    :5.026   Mean    :
87.12
## 3rd Qu.: 1600   3rd Qu.:5.000   3rd Qu.:4   3rd Qu.:5.000   3rd Qu.:
85.00
## Max.    :16000   Max.    :5.000   Max.    :4   Max.    :6.000   Max.
:283.00
##      Weight      Guarantee_Period
## Min.    :1000   Min.    : 3.000
## 1st Qu.:1040   1st Qu.: 3.000
## Median :1070   Median : 3.000
## Mean    :1072   Mean    : 3.815
## 3rd Qu.:1085   3rd Qu.: 3.000
## Max.    :1615   Max.    :36.000
```



## CONCLUSIÓN

Para dar cierre a la exploración de los datos, se recapitulará brevemente lo aprendido del caso. De la base de datos inicialmente planteada, compuesta por 37 atributos y 1,436 instancias, se generó un subset únicamente contemplando los atributos numéricos. Dentro de este subset se realizó un análisis profundo en el que se determinaron las variables con una relación significativa al precio de reventa de los autos, así como un análisis de datos atípicos, e incluso de datos faltantes. De todo esto, se puede concluir que no habrá necesidad de realizar imputación de datos adicionales ya que no se observan ningún dato faltante. Sin embargo, habrá que emplear una estrategia de modelación para regresión logística, además de regresión lineal ya que muchas variables son lógicas y no se pudieron analizar a profundidad en este caso.