Tipología y ciclo de vida de los datos: PRA2

Autor: Miquel Rived

Enero 2022

Contents

```
# Cargamos los paquetes R que vamos a usar
library(ggplot2)
library(dplyr)
library(knitr)
library(stringr)
library(lubridate)
library(purrr)
library(RcmdrMisc)
```

```
# library(ggbiplot)
library(summarytools)
library(kableExtra)
```

Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

El dataset empleado en esta práctica es el obtenido en la PRA1 del semestre pasado y es el resultado de un proceso de *scraping* de la web de *reviews* de cámaras digitales dpreview.com. Contiene 2490 registros con 125 columnas de tipo numérico y categórico, representando cada fila una cámara digital.

Lo más interesante del conjunto de datos extraído es la gran cantidad de especificaciones diferentes que se encuentran, así como el gran abanico de cámaras digitales que lo abarcan, además de que al haber sido extraídas de una web, constituyen datos reales referentes a cámaras.

En primer lugar se quiere analizar que características técnicas afectan más en el aumento de precio de una cámara digital.

Por otro lado, se tratará de determinar qué cámaras son las más valoradas por los usuarios o los expertos, por lo que se pretenderá analizar las marcas más valoradas, si el precio influye en la valoración final, o que tipo de especificaciones son las que buscan los usuarios en una cámara digital para realizar una valoración alta.

Los datos permiten responder a preguntas del tipo:

- ¿Cuál es la cámara mejor valorada por los usuarios?
- ¿Cuál es la cámara más cara y más ergónomica?
- ¿Qué cámara es capaz de disparar más fotografías en modo ráfaga?
- ¿Cuál es la cámara con GPS más ligera y mayor autonomía de batería?

Estas preguntas pueden variar a lo largo de las prácticas, ya que disponemos de una gran variedad de cámaras y campos para analizarlas que seguro que al visualizar con mayor detenimiento nos hacen hacernos nuevas preguntas.

La diferencia principal con los otros estudios encontrados de dpreview.com es la gran variedad de campos que hemos seleccionado para analizar las cámaras.

Integración y selección de los datos de interés a analizar

> Se debe abrir el archivo de datos y examinar el tipo de datos con los que R ha interpretado cada variable. Examinar también los valores resumen de cada tipo de variable.

En primer lugar cargamos el archivo utilizando la función `read.csv`. añadimos el parámetro `stringsAsFactors=False` para que no convierta en factores los atributos que detecte como `String`. Posteriormente, nosotros realizaremos esa conversión para aquellos campos que consideremos oportuno.

```
# Cargamos el fichero de datos
data <- read.csv('dpreview.csv',stringsAsFactors = FALSE)
```

Verificamos el tipo de datos asignado por R a cada atributo utilizando el comando `str` (`structure`):

```
# Verificamos la estructura del conjunto de datos
str(data)
```

```
## 'data.frame':    2490 obs. of  125 variables:
##  $ Maximum.shutter.speed..electronic.: chr  "" "" "1/16000 sec" "1/32000 sec" ...
##  $ review_score                      : num  NA NA NA NA NA NA NA NA 87 NA ...
##  $ Build.quality                     : num  NA NA NA NA NA ...
```

```
##  $ Number.of.cross.type.focus.points : int  25 NA NA NA NA NA NA NA NA NA ...
##  $ Autofocus                         : chr  "Contrast Detect (sensor)Phase DetectMulti-areaCenterSel
##  $ link                              : chr  "https://www.dpreview.com/products/pentax/slrs/pentax_k3
##  $ Format                            : chr  "MPEG-4, H.264" "MPEG-4, H.264" "MPEG-4, H.264, H.265" "
##  $ Touch.screen                      : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ brand                             : chr  "Pentax" "Sigma" "Fujifilm" "Fujifilm" ...
##  $ Screen.dots                       : chr  "1,620,000" "2,100,000" "2,360,000" "1,620,000" ...
##  $ Continuous.drive                  : chr  "12.0 fps" "10.0 fps" "5.0 fps" "20.0 fps" ...
##  $ Microphone.port                   : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Minimum.shutter.speed             : chr  "30 sec" "30 sec" "30 sec" "4 sec" ...
##  $ Shutter.priority                  : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Dimensions                        : chr  "135 x 104 x 74 mm (5.31 x 4.09 x 2.91â\200³)" "113 x 70
##  $ Number.of.lenses                  : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Performance                       : num  NA NA NA NA NA ...
##  $ review_link                       : chr  "" "" "" "" ...
##  $ Viewfinder.magnification          : chr  "1.05Ã- (0.7Ã- 35mm equiv.)" "0.83Ã-" "0.61Ã- (0.77Ã- 35
##  $ Image.stabilization.notes         : chr  "" "" "" "" ...
##  $ Screen.Quality                    : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ GPS.notes                         : chr  "" "" "" "" ...
##  $ own_gear                          : int  3 0 31 17 49 10 0 21 99 187 ...
##  $ Boosted.ISO..minimum.             : int  NA 6 50 80 50 50 100 NA 80 50 ...
##  $ Viewfinder...screen.rating        : num  NA NA NA NA NA ...
##  $ Speed.and.Responsiveness          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Optical.zoom                      : chr  "" "" "" "" ...
##  $ want_gear                         : int  72 26 152 88 286 34 3 106 136 231 ...
##  $ Flash.modes                       : chr  "Auto, Auto + Red-eye Reduction, Flash On, Flash On + Re
##  $ Custom.white.balance              : chr  "Yes (3 slots)" "Yes" "Yes (3 slots)" "Yes" ...
##  $ Uncompressed.format               : chr  "RAW" "RAW" "RAW + TIFF" "RAW" ...
##  $ announcement_date                 : chr  "Announced 2 days ago" "Announced 1 week ago" "Announced
##  $ Timelapse.recording               : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ White.balance.presets             : int  8 6 7 7 7 8 6 NA 7 12 ...
##  $ Focal.length.multiplier           : chr  "1.5Ã-" "1Ã-" "0.79Ã-" "1.5Ã-" ...
##  $ Max.resolution                    : chr  "6192 x 4128" "9520 x 6328" "11648 x 8736" "6240 x 4160"
##  $ Video.Quality                     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Live.view                         : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Battery.description               : chr  "D-LI90" "BP-51 lithium-ion battery and charger" "NP-W23
##  $ Sensor.photo.detectors            : chr  "27 megapixels" "62 megapixels" "" "" ...
##  $ Boosted.ISO..maximum.             : int  NA 102400 102400 51200 102400 NA 25600 NA 51200 204800 .
##  $ Screen.type                       : chr  "TFT LCD" "TFT LCD" "TFT LCD" "TFT LCD" ...
##  $ brand_camera_family               : chr  "Pentax Interchangeable Lens Cameras" "Sigma Interchange
##  $ Still.Image.Quality               : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Viewfinder.coverage               : chr  "100%" "100%" "100%" "100%" ...
##  $ Remote.control                    : chr  "Yes (via wireless)" "Yes (via smartphone or wired remot
##  $ Value                             : num  NA NA NA NA NA ...
##  $ Sensor.size                       : chr  "APS-C (23 x 15.5 mm)" "Full frame (36 x 24 mm)" "Medium
##  $ Maximum.shutter.speed             : chr  "1/8000 sec" "1/8000 sec" "1/4000 sec" "1/4000 sec" ...
##  $ Metering.modes                    : chr  "MultiCenter-weightedHighlight-weightedSpot" "MultiCente
##  $ Environmentally.sealed            : chr  "Yes" "Yes" "Yes" "No" ...
##  $ Wireless.notes                    : chr  "802.11b/g/n + Bluetooth" "" "802.11ac + Bluetooth" "802
##  $ Focal.length..equiv..             : chr  "" "" "" "" ...
##  $ name                              : chr  "Pentax K-3 Mark III" "Sigma fp L" "Fujifilm GFX 100S" "
##  $ Battery                           : chr  "Battery Pack" "Battery Pack" "Battery Pack" "Battery Pa
##  $ Viewfinder.resolution             : chr  "" "3,680,000" "3,690,000" "2,360,000" ...
##  $ Movie...video.mode                : num  NA NA NA NA NA ...
```

```
##  $ Articulated.LCD                  : chr  "Fixed" "Fixed" "Tilting" "Tilting" ...
##  $ Image.quality..raw.              : num  NA NA NA NA NA ...
##  $ Connectivity                     : num  NA NA NA NA NA ...
##  $ review_award                     : chr  "" "" "" "" ...
##  $ JPEG.quality.levels              : chr  "Best, better, good" "High, med, low" "Super fine, fine,
##  $ Screen.size                      : chr  "3.2â\2003" "3.2â\2003" "3.2â\2003" "3â\2003" ...
##  $ Weight..inc..batteries.          : chr  "820 g (1.81 lb / 28.92 oz)" "427 g (0.94 lb / 15.06 oz)
##  $ Effective.pixels                 : chr  "26 megapixels" "61 megapixels" "102 megapixels" "26 mega
##  $ quick_specs                      : chr  "26 megapixels | 3.2â\2003 screen | APS-C sensor" "61 meg
##  $ Lens.mount                       : chr  "Pentax KAF2" "Leica L" "Fujifilm G" "Fujifilm X" ...
##  $ Manual.exposure.mode             : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Metering...focus.accuracy        : num  NA NA NA NA NA ...
##  $ Sensor.type                      : chr  "CMOS" "BSI-CMOS" "BSI-CMOS" "BSI-CMOS" ...
##  $ Normal.focus.range               : chr  "" "" "" "" ...
##  $ Durability                       : chr  "" "" "" "" ...
##  $ Speaker                          : chr  "Mono" "Mono" "Mono" "Mono" ...
##  $ Resolutions                      : chr  "" "" "" "" ...
##  $ had_gear                         : int  4 1 8 11 31 11 2 6 15 19 ...
##  $ Image.quality..jpeg.             : num  NA NA NA NA NA ...
##  $ Optics                           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Viewfinder.type                  : chr  "Optical (pentaprism)" "Electronic (optional)" "Electroni
##  $ MSRP                             : chr  "$1999 (body only)" "$2499 (body only), $2999 (with EVF)"
##  $ Ergonomics.and.Handling          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Flash.performance                : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ GPS                              : chr  "None" "None" "None" "None" ...
##  $ Performance..speed.              : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Self.timer                       : chr  "Yes" "Yes (2 or 10 sec)" "Yes" "Yes" ...
##  $ USB                              : chr  "USB 3.2 Gen 1 (5 GBit/sec)" "USB 3.2 Gen 1 (5 GBit/sec)
##  $ Other.resolutions                : chr  "" "" "" "" ...
##  $ Features                         : num  NA NA NA NA NA ...
##  $ Wireless                         : chr  "Built-In" "Built-In" "Built-In" "Built-In" ...
##  $ Built.in.flash                   : chr  "No" "No" "No" "No" ...
##  $ Exposure.compensation            : chr  "Â±5 (at 1/3 EV, 1/2 EV steps)" "Â±5 (at 1/3 EV steps)"
##  $ Exposure.and.focus.accuracy      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Ergonomics...handling            : num  NA NA NA NA NA ...
##  $ External.flash                   : chr  "Yes (via hot shoe or flash sync port)" "Yes (via flash
##  $ AE.Bracketing                    : chr  "Â±5 (2, 3, 5 frames at 1/3 EV, 1/2 EV steps)" "Â±3 (3, 5
##  $ Battery.Life..CIPA.              : int  800 240 460 380 530 510 330 350 325 410 ...
##  $ USB.charging                     : chr  "Yes" "Yes (USB Power Delivery supported)" "Yes" "Yes" .
##  $ HDMI                             : chr  "Yes (micro HDMI)" "Yes (micro-HDMI)" "Yes" "Yes (micro-H
##  $ Modes                            : chr  "3840 x 2160 @ 30p, MOV, H.264, Linear PCM3840 x 2160 @ 1
##  $ Digital.zoom                     : chr  "Yes" "Yes (1.5x - 5x)" "" "Yes" ...
##   [list output truncated]
```

Los atributos identificados, junto a su significado se resumen en la siguiente tabla:

```
description_df <- data.frame (
                    #campo  = colnames(data),
                    tipo = sapply(data, class),
                    descripción = c(
                      "Velocidad de obturación máxima .",
                      "Puntuación de review",
                      "Puntuación calidad de construcción",
                      "Número de puntos de enfoque",
                      "Tipo de autofoco",
```

```
                    "Enlace a la página",
                    "Formatos de almacenamiento de vídeo",
                    "Pantalla táctil",
                    "Marca",
                    "Puntos de la pantalla",
                    "Disparos en modo ráfaga",
                    "Conexión para micrófono",
                    "Velocidad de obturación mínima",
                    "Prioridad de obturador",
                    "Dimensiones",
                    "Número de lentes",
                    "Puntuación de rendimiento",
                    "Enlace a la review",
                    "Amplificación del visor",
                    "Observaciones sobre la estabilización de imagen",
                    "Puntuación de calidad de pantalla",
                    "Observaciones sobre GPS",
                    "Número de usuarios que poseen la cámara",
                    "ISO mínimo",
                    "Puntuación de visor",
                    "Puntuación de velocidad y respuesta",
                    "Zoom óptico",
                    "Número de usuarios que desean tener la cámara",
                    "Modos de flash",
                    "Balances de blancos personalizados",
                    "Formato sin compresión",
                    "Fecha de publicación",
                    "Grabación de timelapse",
                    "Preajustes de balances de blancos",
                    "Multiplicador de la distancia focal",
                    "Resolución máxima",
                    "Puntuación de calidad de vídeo",
                    "Vista en directo",
                    "Descripción de la batería",
                    "Resolución del sensor",
                    "ISO máximo",
                    "Tipo de pantalla",
                    "Familia de la cámara",
                    "Puntuación de imagen fija",
                    "Cobertura del visor",
                    "Mando a distancia",
                    "Puntuación general",
                    "Tamaño del sensor",
                    "Velocidad de obturación máxima",
                    "Modos de medición",
                    "Estanqueidad",
                    "Observaciones sobre conexión inalámbrica",
                    "Distancia focal equivalente",
                    "Modelo de cámara",
                    "Formato de bateria",
                    "Resolución del visor",
                    "Puntuación de modo vídeo",
                    "Modo de pantalla articulada",
```

```
                    "Puntuación de calidad de imagen en RAW.",
                    "Puntuación de conectividad",
                    "Premio (Gold, Silver)",
                    "Niveles de calidad JPEG",
                    "Tamaño de la pantalla",
                    "Peso con batería", # Extraer gramos por expresión regular
                    "Píxeles efectivos",
                    "Especificaciones rápidas",
                    "Montura de lente",
                    "Modo de exposición manual",
                    "Puntuación de medición y enfoque",
                    "Tipo de sensor",
                    "Distancia mínima de enfoque",
                    "Durabilidad",
                    "Altavoz",
                    "Resoluciones",
                    "Número de usuarios que han poseido la cámara",
                    "Puntuación de calidad de imagen JPG",
                    "Puntuación de la óptica",
                    "Tipo de visor",
                    "Precio recomendado", # extraer mediante expresión regular el número
                    "Puntuación de ergonomía y manejo",
                    "Puntuación de rendimiento de flash",
                    "GPS",
                    "Puntuación de rendimiento en velocidad",
                    "Cuenta atrás",
                    "USB",
                    "Otras resoluciones",
                    "Puntuación de características",
                    "Wireless",
                    "Flash incluido",
                    "Compensación de la exposición",
                    "Puntuación de exactitud de exposición y enfoque",
                    "Puntuación de ergonomía y manejo",
                    "Flash externo",
                    "AE Bracketing",
                    "Ciclos de vida de la batería",
                    "Carga por USB",
                    "HDMI",
                    "Modos",
                    "Zoom digital",
                    "Apertura máxima",
                    "Rango del flash",
                    "ISO",
                    "Puntuación de review (porcentaje)",
                    "Modos de escena",
                    "Enfoque manual",
                    "Estabilización de imagen",
                    "Prioridad de apertura",
                    "Orientación del sensor" ,
                    "Puntuación CIPA de estabilización de imagen",
                    "Observaciones sobre grabación en vídeo",
                    "URL de imagen de la cámara",
```

```
                      "Tipo de micrófono",
                      "Puntuación con rendimeinto con luz baja e ISO alto",
                      "Campo de visión",
                      "Bracketing de balance de blancos",
                      "Puntuación de las características de la camera y fotografía",
                      "Procesador",
                      "Ratio de imagen ancho x alto",
                      "Distancia mínima modo macro",
                      "Tipo de cuerpo", # IMPORTANTE
                      "Conector de cascos",
                      "Número de reviews", # Extraer número con expresión regular
                      "Tipos de almacenamientos",
                      "Número de puntos de enfoque", # IMPORTANTE
                      "Almacenamiento incluido")
                  )
kable(description_df,caption="**Propiedades de las cámaras**") %>% kable_styling()
```

En primer lugar, a continuación mostraremos las columnas que vamos a seleccionar del dataset, además de renombrarlas a los nombres indicados.

Variables cuantitativas

- Los relativos a la **puntuación del equipo de expertos** de dpreview:
  - `review_score`: Puntuación de la review. Renombramos a `puntuacion.review`
  - `Build.quality`: Puntuación calidad de construcción. Renombramos a `puntuacion.calidad_construccion`
  - `Performance`: Puntuación del rendimiento. Renombramos a `puntuación.rendimiento`
  - `Screen.Quality`: Puntuación de la calidad de la pantalla. Renombramos a `puntuacion.calidad.pantalla`
  - `Viewfinder...screen.rating`: Puntuación del visor. Renombramos a `puntuacion.visor`
  - `Speed.and.Responsiveness`: Puntuación velocidad y respuesta. Renombramos a `puntuacion.velocidad_respues`
  - `Video.Quality`: Puntuación de calidad de vídeo. Renombramos a `puntuacion.calidad_video`
  - `Still.Image.Quality`: Puntuación de imagen fija. Renombramos a `puntuacion.calidad_imagen_fija`
  - `Value`: Puntuación general. Renombramos a `puntuacion.general`
  - `Movie...video.mode`: Puntuación del modo vídeo. Renombramos a `puntuacion.modo_video`
  - `Image.quality..raw.`: Puntuación de la calidad de imagen RAW. Renombramos a `puntuacion.calidad_raw`
  - `Connectivity`: Puntuación de conectividad. Renombramos a `puntuacion.conectividad`
  - `Metering...focus.accuracy`: Puntuación de medición y enfoque. Renombramos a `puntuacion.precision`
  - `Image.quality..jpeg.`: Puntuación de imagen JPG. Renombramos a `puntuacion.calidad_jpg`
  - `Optics`: Puntuación de óptica. Renombramos a `puntuacion.optica`
  - `Ergonomics...handling`: Puntuación de ergonomía y manejo. Renombramos a `puntuacion.ergonomia_manejo`
  - `Flash.performance`: Puntuación de rendimiento del flash. Renombramos a `puntuacion.rendimento_flash`
  - `Performance..speed.`: Puntuación de rendimiento en velocidad. Renombramos a `puntuacion.rendimiento_velocidad`
  - `Features`: Puntuación de características. Renombramos a `puntuacion.caracteristicas`
  - `Exposure.and.focus.accuracy`: Puntuación de precisión de exposición y enfoque. Renombramos a `puntuacion.precision_exposicion_enfoque`
  - `review_value`: Puntuación de review (porcentaje). Renombramos a `puntuacion.review`
  - `Low.light...high.ISO.performance`: Puntuación con rendimeinto con luz baja e ISO alto. Renombramos a `puntuacion.luz_baja_alto_ISO`
  - `Camera.and.Photo.Features`: Puntuación de las características de la camera y fotografía. Renombramos a `puntuacion.caracteristicas_camara_foto`

Table 1: **Propiedades de las cámaras**

	tipo	descripción
Maximum.shutter.speed..electronic.	character	Velocidad de obturación máxima .
review_score	numeric	Puntuación de review
Build.quality	numeric	Puntuación calidad de construcción
Number.of.cross.type.focus.points	integer	Número de puntos de enfoque
Autofocus	character	Tipo de autofoco
link	character	Enlace a la página
Format	character	Formatos de almacenamiento de vídeo
Touch.screen	character	Pantalla táctil
brand	character	Marca
Screen.dots	character	Puntos de la pantalla
Continuous.drive	character	Disparos en modo ráfaga
Microphone.port	character	Conexión para micrófono
Minimum.shutter.speed	character	Velocidad de obturación mínima
Shutter.priority	character	Prioridad de obturador
Dimensions	character	Dimensiones
Number.of.lenses	integer	Número de lentes
Performance	numeric	Puntuación de rendimiento
review_link	character	Enlace a la review
Viewfinder.magnification	character	Amplificación del visor
Image.stabilization.notes	character	Observaciones sobre la estabilización de imagen
Screen.Quality	numeric	Puntuación de calidad de pantalla
GPS.notes	character	Observaciones sobre GPS
own_gear	integer	Número de usuarios que poseen la cámara
Boosted.ISO..minimum.	integer	ISO mínimo
Viewfinder...screen.rating	numeric	Puntuación de visor
Speed.and.Responsiveness	numeric	Puntuación de velocidad y respuesta
Optical.zoom	character	Zoom óptico
want_gear	integer	Número de usuarios que desean tener la cámara
Flash.modes	character	Modos de flash
Custom.white.balance	character	Balances de blancos personalizados
Uncompressed.format	character	Formato sin compresión
announcement_date	character	Fecha de publicación
Timelapse.recording	character	Grabación de timelapse
White.balance.presets	integer	Preajustes de balances de blancos
Focal.length.multiplier	character	Multiplicador de la distancia focal
Max.resolution	character	Resolución máxima
Video.Quality	numeric	Puntuación de calidad de vídeo
Live.view	character	Vista en directo
Battery.description	character	Descripción de la batería
Sensor.photo.detectors	character	Resolución del sensor
Boosted.ISO..maximum.	integer	ISO máximo
Screen.type	character	Tipo de pantalla
brand_camera_family	character	Familia de la cámara
Still.Image.Quality	numeric	Puntuación de imagen fija
Viewfinder.coverage	character	Cobertura del visor
Remote.control	character	Mando a distancia
Value	numeric	Puntuación general
Sensor.size	character	Tamaño del sensor
Maximum.shutter.speed	character	Velocidad de obturación máxima
Metering.modes	character	Modos de medición
Environmentally.sealed	character	Estanqueidad
Wireless.notes	character	Observaciones sobre conexión inalámbrica
Focal.length..equiv..	character	Distancia focal equivalente
name	character	Modelo de cámara
Battery	character	Formato de bateria

```
data<- rename(data,
              puntuacion.calidad_construccion=Build.quality,
              puntuación.rendimiento=Performance,
              puntuacion.calidad.pantalla=Screen.Quality,
              puntuacion.visor=Viewfinder...screen.rating,
              puntuacion.velocidad_respuesta=Speed.and.Responsiveness,
              puntuacion.calidad_video=Video.Quality,
              puntuacion.calidad_imagen_fija=Still.Image.Quality,
              puntuacion.general=Value,
              puntuacion.modo_video=Movie...video.mode,
              puntuacion.puntuacion_calidad_raw=Image.quality..raw.,
              puntuacion.conectividad=Connectivity,
              puntuacion.precision=Metering...focus.accuracy,
              puntuacion.calidad_jpg=Image.quality..jpeg.,
              puntuacion.optica=Optics,
              puntuacion.ergonomia_manejo=Ergonomics...handling,
              puntuacion.rendimento_flash=Flash.performance,
              puntuacion.rendimiento_velocidad=Performance..speed.,
              puntuacion.caracteristicas=Features,
              puntuacion.precision_exposicion_enfoque=Exposure.and.focus.accuracy,
              puntuacion.pro_review=review_value,
              puntuacion.luz_baja_alto_ISO=Low.light...high.ISO.performance,
              puntuacion.caracteristicas_camara_foto=Camera.and.Photo.Features)
```

- Las relativos a las **características físicas de la cámara**:
  - `Weight..inc..batteries`: Peso. Renombramos a `caracteristicas.peso`
  - `Dimensions`: Dimensiones (ancho x alto x fondo). Renombramos a `caracteristicas.dimensiones`

```
data<- rename(data,
              caracteristicas.peso=Weight..inc..batteries.,
              caracteristicas.dimensiones=Dimensions)
```

- Las relativas al **interés por parte de los usuarios de la comunidad** de dpreview:
  - own_gear. Renombramos a `usuario.tiene`
  - had_gear. Renombramos a `usuario.ha_tenido`
  - want_gear. Renombramos a `usuario.desea`
  - review_score: Puntuación media por parte de los usuarios de la comunidad. Renombramos a `usuario.puntuacion`

```
data<- rename(data,
              usuario.tiene=own_gear,
              usuario.ha_tenido=had_gear,
              usuario.desea=want_gear,
              usuario.puntuacion=review_score)
```

- El **precio de la cámara**: `MSRP`. Renombramos a `precio`

```
data<- rename(data,
              precio=MSRP)
```

- Las relativas a **características técnicas** de la cámara:
  - `Max.resolution`: resolución, renombramos a `caracteristicas.resolucion` -`Maximum.shutter.speed`: velocidad máxima de obturador, renombramos a `caracteristicas.velocidad_obturador`

```
data<- rename(data,
              caracteristicas.resolucion=Max.resolution,
```

```
                caracteristicas.velocidad_obturador=Maximum.shutter.speed
               )
```

Variables cualitativas

- Marca y modelo de la cámara:
  - `Brand`: Marca, renombramos a `marca`
  - `name`: Modelo, renombramos a `modelo`
- Otras características:
  - `Body.type`: Tipo de cuerpo, renombramos a `tipo_cuerpo`
  - `Sensor.type`: Tipo de sensor, renombramos a `tipo_sensor`
  - `announcement_date`: Fecha de lanzamiento, renombramos a `fecha_lanzamiento`

```
data<- rename(data,
               marca=brand,
               modelo=name,
               tipo_cuerpo=Body.type,
               tipo_sensor=Sensor.type,
               fecha_lanzamiento=announcement_date
              )
```

Convertimos a factor los atributos:

```
data$marca<-as.factor(data$marca)
data$tipo_cuerpo<-as.factor(data$tipo_cuerpo)
data$tipo_sensor<-as.factor(data$tipo_sensor)
data$caracteristicas.velocidad_obturador<-as.factor(data$caracteristicas.velocidad_obturador)
data$GPS <- as.factor(data$GPS)
```

Finalmente aplicamos la reducción de la dimensionalidad, filtrando únicamente las columnas que nos interesan

```
data <- data[ , grepl( "puntuacion|caracteristicas|tipo_|fecha_lanzamiento|marca|modelo|precio|GPS" , na
```

Extraccción de valores numéricos

Atributo `precio`

Los valores de la columna `precio` tiene diversos tipos de valores. En la mayoría de las filas contiene un único precio en dólares, precedido del símbolo de dolar (p.ej. `$899`), pero en algunos casos incluye varios precios en función del kit a comprar (con objetivo, sin objetivo, con varios ojetivos,...) y en otros casos se muestra el precio tanto en dólares como en otras monedas (euro y libras).

Para simplificar la extracción, nos quedaremos con el primer precio que aparezca precedido del símbolo del dólar utilizando una expresión regular.

Otras posibles alternativas en el tratamiento de este atributo podría haber sido obtener todos los posibles valores precedidos del símbolo dólar y quedarnos con el máximo, o hacer una conversión entre divisas según el cambio euro->dólar y libra->dolar.

```
data$precio<-as.integer(str_extract(str_extract(data$precio, "(\\$\\d+)"),"(\\d+)"))
```

Atributo `puntuacion.pro_review`

Este atributo contiene un porcentaje que incluye el símbolo `%`. Interesa convertirlo a entero y eliminar ese porcentaje. Utilizaremos para ello una expresión regular y la función `str_extract` del paquete `stringr`

```
data$puntuacion.pro_review<-as.integer(str_extract(data$puntuacion.pro_review, "(\\d+)"))
```

Atributo `caracteristicas.peso`

Este atributo expresa el peso tanto en gramos como en libras y onzas. Utilizaremos una expresión regular para quedarnos con el peso en gramos

```
data$caracteristicas.peso<-as.integer(str_extract(data$caracteristicas.peso, "(\\d+)"))
```

Atributo `caracteristicas.dimensiones`

Este atributo expresa el tamaño (ancho x alto x fondo) en milímetros. A partir del mismo es posible establecer una nueva variable volumen

```
data$dim.ancho<-as.integer(str_match(data$caracteristicas.dimensiones, "(\\d+) x (\\d+) x (\\d+)")[,2])
data$dim.alto<-as.integer(str_match(data$caracteristicas.dimensiones, "(\\d+) x (\\d+) x (\\d+)")[,3])
data$dim.fondo<-as.integer(str_match(data$caracteristicas.dimensiones, "(\\d+) x (\\d+) x (\\d+)")[,4])

# Calculamos el volumen
data$caracteristicas.volumen <- data$dim.ancho  * data$dim.alto * data$dim.fondo

# data %>% select(caracteristicas.dimensiones,dim.ancho,dim.alto,dim.fondo, caracteristicas.volumen)

# Eliminamos las columnas innecesarias
data <- data %>% select (-c(caracteristicas.dimensiones,dim.ancho, dim.alto,dim.fondo))
```

Ajuste de variables discretas

Tipo de sensor

La variable `tipo_sensor` contiene diversos valores que se pueden agrupar en dos grandes grupos: CCD y CMOS. Realizamos esta simplificación:

```
# https://stackoverflow.com/questions/25372082/create-column-based-on-presence-of-string-pattern-and-if
data$tipo_sensor<-ifelse(grepl("CMOS",data$tipo_sensor),'CMOS','CCD')
```

Extracción de fecha

La fecha de lanzamiento de la cámara viene especificada en el atributo `fecha_lanzamiento`.

El formato del campo varía en función de si el lanzamiento ha sido hace menos de un año (p.ej. "`Announced 7 months ago`"), o hace más de un año (p.ej. "`Announced Jan 15, 2016`"). En cualquier caso, de cara al análisis sólo nos interesa quedarnos con el año de lanzamiento. Utilizaremos una expresión regular para obtener el año, y en caso de que no tenga éxito, asignaremos el valor actual.

```
data$fecha_lanzamiento <- as.integer(str_extract(data$fecha_lanzamiento, "(\\d\\d\\d\\d)"))
# Asignamos por defecto al año 2020
data[is.na(data$fecha_lanzamiento),"fecha_lanzamiento"] = 2020
data$fecha_lanzamiento <-as.factor(data$fecha_lanzamiento)
```

Resumen de los datos

A continuación se muestra un resumen de las variables del dataset

```
summary(data)
```

```
##  usuario.puntuacion puntuacion.calidad_construccion       marca
##  Min.   : 10.00     Min.   :28.57                   Canon    :303
##  1st Qu.: 77.47     1st Qu.:64.29                   Sony     :299
##  Median : 84.24     Median :75.71                   Nikon    :264
```

```
##  Mean   : 81.69     Mean   :74.25                  Fujifilm :253
##  3rd Qu.: 89.16     3rd Qu.:82.86                  Olympus  :253
##  Max.   :100.00     Max.   :98.57                  Panasonic:211
##  NA's   :446        NA's   :2131                   (Other)  :907
##  puntuacion.calidad.pantalla  GPS.notes         puntuacion.visor
##  Min.   :84.29              Length:2490        Min.   :21.43
##  1st Qu.:84.29              Class :character   1st Qu.:68.57
##  Median :84.29              Mode  :character   Median :78.57
##  Mean   :84.29                                 Mean   :75.51
##  3rd Qu.:84.29                                 3rd Qu.:85.71
##  Max.   :84.29                                 Max.   :97.14
##  NA's   :2487                                  NA's   :2220
##  puntuacion.velocidad_respuesta fecha_lanzamiento caracteristicas.resolucion
##  Min.   :71.43                  2007   : 178      Length:2490
##  1st Qu.:71.43                  2008   : 172      Class :character
##  Median :71.43                  2010   : 171      Mode  :character
##  Mean   :71.43                  2009   : 170
##  3rd Qu.:71.43                  2012   : 165
##  Max.   :71.43                  2011   : 162
##  NA's   :2487                   (Other):1472
##  puntuacion.calidad_video puntuacion.calidad_imagen_fija puntuacion.general
##  Min.   :78.57            Min.   :47.14                  Min.   :14.29
##  1st Qu.:78.57            1st Qu.:47.14                  1st Qu.:57.14
##  Median :78.57            Median :47.14                  Median :64.29
##  Mean   :78.57            Mean   :47.14                  Mean   :62.88
##  3rd Qu.:78.57            3rd Qu.:47.14                  3rd Qu.:71.43
##  Max.   :78.57            Max.   :47.14                  Max.   :95.71
##  NA's   :2487             NA's   :2487                   NA's   :2131
##  caracteristicas.velocidad_obturador    modelo          puntuacion.modo_video
##  1/2000 sec:892                      Length:2490        Min.   : 14.29
##  1/4000 sec:402                      Class :character   1st Qu.: 50.00
##  1/1000 sec:254                      Mode  :character   Median : 64.29
##            :190                                         Mean   : 63.78
##  1/8000 sec:146                                         3rd Qu.: 78.57
##  1/1500 sec:130                                         Max.   :100.00
##  (Other)   :476                                         NA's   :2145
##  puntuacion.puntuacion_calidad_raw puntuacion.conectividad caracteristicas.peso
##  Min.   : 26.79                    Min.   :42.86           Min.   :  46.0
##  1st Qu.: 69.14                    1st Qu.:64.29           1st Qu.: 170.0
##  Median : 76.43                    Median :71.43           Median : 228.0
##  Mean   : 74.69                    Mean   :69.50           Mean   : 327.9
##  3rd Qu.: 82.89                    3rd Qu.:76.43           3rd Qu.: 406.0
##  Max.   :100.00                    Max.   :85.71           Max.   :1860.0
##  NA's   :2190                      NA's   :2308            NA's   :95
##  puntuacion.precision tipo_sensor        puntuacion.calidad_jpg
##  Min.   :32.14        Length:2490        Min.   :21.43
##  1st Qu.:67.86        Class :character   1st Qu.:58.13
##  Median :71.43        Mode  :character   Median :65.51
##  Mean   :70.88                           Mean   :64.60
##  3rd Qu.:75.00                           3rd Qu.:71.60
##  Max.   :90.00                           Max.   :87.14
##  NA's   :2220                            NA's   :2131
##  puntuacion.optica     precio     puntuacion.rendimento_flash       GPS
##  Min.   :36.73     Min.   :   1   Min.   :21.43                       :1358
```

```
##  1st Qu.:57.14     1st Qu.: 349   1st Qu.:28.57              Built-in: 109
##  Median :64.29     Median : 599   Median :42.86              None    : 918
##  Mean   :65.06     Mean   :1222   Mean   :42.35              Optional: 105
##  3rd Qu.:74.49     3rd Qu.:1199   3rd Qu.:50.00
##  Max.   :89.59     Max.   :9999   Max.   :71.43
##  NA's   :2357      NA's   :2009   NA's   :2402
##  puntuacion.rendimiento_velocidad puntuacion.caracteristicas
##  Min.   :14.29                    Min.   :33.33
##  1st Qu.:46.43                    1st Qu.:61.90
##  Median :57.14                    Median :71.43
##  Mean   :55.65                    Mean   :70.02
##  3rd Qu.:64.29                    3rd Qu.:80.95
##  Max.   :78.57                    Max.   :95.71
##  NA's   :2401                     NA's   :2131
##  puntuacion.precision_exposicion_enfoque puntuacion.ergonomia_manejo
##  Min.   :28.57                           Min.   :14.29
##  1st Qu.:57.14                           1st Qu.:64.29
##  Median :64.29                           Median :71.43
##  Mean   :62.83                           Mean   :69.38
##  3rd Qu.:71.43                           3rd Qu.:78.57
##  Max.   :82.86                           Max.   :94.29
##  NA's   :2401                            NA's   :2131
##  puntuacion.pro_review puntuacion.luz_baja_alto_ISO
##  Min.   : 0.00         Min.   :14.29
##  1st Qu.:73.00         1st Qu.:56.82
##  Median :78.00         Median :66.88
##  Mean   :77.47         Mean   :63.80
##  3rd Qu.:83.00         3rd Qu.:75.32
##  Max.   :92.00         Max.   :93.64
##  NA's   :2127          NA's   :2131
##  puntuacion.caracteristicas_camara_foto                      tipo_cuerpo
##  Min.   :87.14                          Compact                     :1336
##  1st Qu.:87.14                          Ultracompact                : 499
##  Median :87.14                          Rangefinder-style mirrorless: 138
##  Mean   :87.14                          SLR-like (bridge)           : 116
##  3rd Qu.:87.14                          Mid-size SLR                :  99
##  Max.   :87.14                          SLR-style mirrorless        :  90
##  NA's   :2487                           (Other)                     : 212
##  caracteristicas.volumen
##  Min.   :  48336
##  1st Qu.: 134059
##  Median : 222222
##  Mean   : 434752
##  3rd Qu.: 558260
##  Max.   :4831200
##  NA's   :54
```

Limpieza de datos

¿Los datos contienen ceros o elementos vacíos? ¿Como gestionarías cada uno de estos casos?

Comprobar atributos con valores vacíos

Vamos a verificar si el dataset incluye elementos vacíos:

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

```
##                       usuario.puntuacion         puntuacion.calidad_construccion
##                                      446                                    2131
##                                   marca             puntuacion.calidad.pantalla
##                                       0                                    2487
##                                GPS.notes                       puntuacion.visor
##                                       0                                    2220
##          puntuacion.velocidad_respuesta                      fecha_lanzamiento
##                                    2487                                       0
##              caracteristicas.resolucion                puntuacion.calidad_video
##                                       0                                    2487
##          puntuacion.calidad_imagen_fija                     puntuacion.general
##                                    2487                                    2131
##     caracteristicas.velocidad_obturador                                 modelo
##                                       0                                       0
##                   puntuacion.modo_video       puntuacion.puntuacion_calidad_raw
##                                    2145                                    2190
##                puntuacion.conectividad                   caracteristicas.peso
##                                    2308                                      95
##                   puntuacion.precision                           tipo_sensor
##                                    2220                                       0
##                puntuacion.calidad_jpg                      puntuacion.optica
##                                    2131                                    2357
##                                 precio             puntuacion.rendimento_flash
##                                    2009                                    2402
##                                    GPS        puntuacion.rendimiento_velocidad
##                                       0                                    2401
##             puntuacion.caracteristicas puntuacion.precision_exposicion_enfoque
##                                    2131                                    2401
##             puntuacion.ergonomia_manejo                  puntuacion.pro_review
##                                    2131                                    2127
##           puntuacion.luz_baja_alto_ISO  puntuacion.caracteristicas_camara_foto
##                                    2131                                    2487
##                             tipo_cuerpo                caracteristicas.volumen
##                                       0                                      54
```

Se comprueba que efectivamente el `dataset` contiene un gran número de elementos vacíos, especialmente en las columnas de puntuación. El primer tratamiento que realizaremos es asignar `NA` a los atributos con valor cadena vacía.

```
data <- data %>% mutate_all(list(~na_if(.,"")))
```

Eliminamos todos aquellos atributos tipo "puntuación" que tengan más del 90% de atributos con valor `NA`

```
#https://stackoverflow.com/questions/31848156/delete-columns-rows-with-more-than-x-missing
data.puntuacion <- data%>%select(starts_with("puntuacion"))
atributes_to_remove<-colnames(data.puntuacion[, which(colMeans(is.na(data.puntuacion)) > 0.9)])
```

```
data<-data %>% select(-atributes_to_remove)
```

Además, dado que nuestra variable objetivo es el precio, eliminamos todas aquellas filas para las que el precio no se encuentra definido. Esto supondrá una reducción significativa en el `dataset`.

```
data <- data %>% filter (!is.na(precio))
```

Verificar duplicación de registros

No existe un campo ID. Se puede verificar la existencia de duplicados mediante la combinación de campos Brand+Name.

Para comprobar si hay algún registro repetido utilizamos la función `unique` que proporciona los elementos únicos de una lista, en combinación con la función `nrow`, que obtiene el número de filas de un dataframe.

```
if (nrow(unique(data[,c("marca","modelo")]))!=nrow(data[,c("marca","modelo")])){
  print("Hay alguna cámara repetida")
}else{
  print("No hay ninguna cámara repetida")
}
```

```
## [1] "No hay ninguna cámara repetida"
```

Imputación de valores vacios

Puntuaciones Para todos los atributos de tipo ´puntuacion.xxxx´, imputamos inicialmente con la **media de la misma marca y tipo de cuerpo**:

```
data<-data %>%
    group_by(marca,tipo_cuerpo) %>%
    mutate_each(funs(replace(., which(is.na(.)), mean(., na.rm=TRUE))),
       starts_with('puntuacion'))
```

En caso de los valores que sigan siendo `NA` utilizamos la **media de la marca**:

```
data<-data %>%
    group_by(marca) %>%
    mutate_each(funs(replace(., which(is.na(.)), mean(., na.rm=TRUE))),
       starts_with('puntuacion'))
```

Y si continua siendo `NA` imputamos por la **media general de la puntuación**:

```
data<-data %>%
    mutate_each(funs(replace(., which(is.nan(.)), mean(., na.rm=TRUE))),
       starts_with('puntuacion'))
```

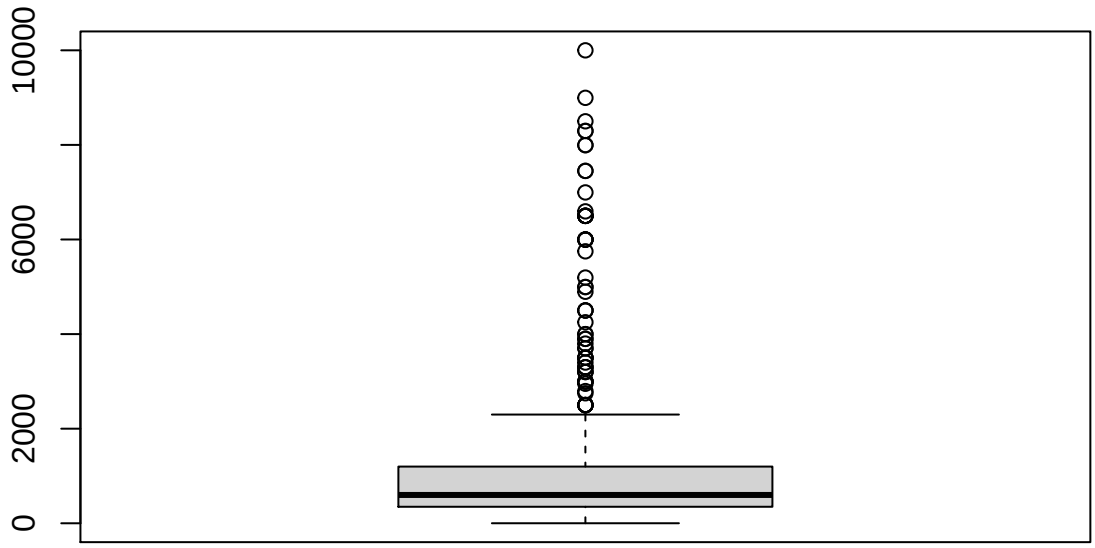
Identificación y tratamiento de valores extremos

Identificamos a continuación los valores extremos. Nosotros lo realizaremos para los valores de precio, peso y la puntuación de proreview.

Precio

Veamos los valores extremos del precio de las cámaras

```
boxplot(data$precio)
```




```
out = boxplot.stats(data$precio)$out
out
```

```
##  [1] 2499 5999 6500 4895 5995 2999 3499 8295 3899 2499 8295 6499 5999 4500 6499
## [16] 3500 3999 5750 3999 9999 4995 2499 3699 2999 4995 2499 7995 4499 3399 7995
## [31] 2499 2795 3199 3299 4500 6595 6499 3199 3499 8995 5999 2950 6499 5195 7450
## [46] 3299 3000 3199 4250 7450 2499 3799 3699 3899 6995 3299 8499 2499 6499 2749
## [61] 3499 2999 5999
```

Para el precio tenemos muchos valores extremos. Como podemos comprobar los datos con la misma página de dpreview.com para ver si se ha realizado bien la extracción, antes de tomar cualquier decisión podemos comprobar si los datos se han extraído bien de la página.

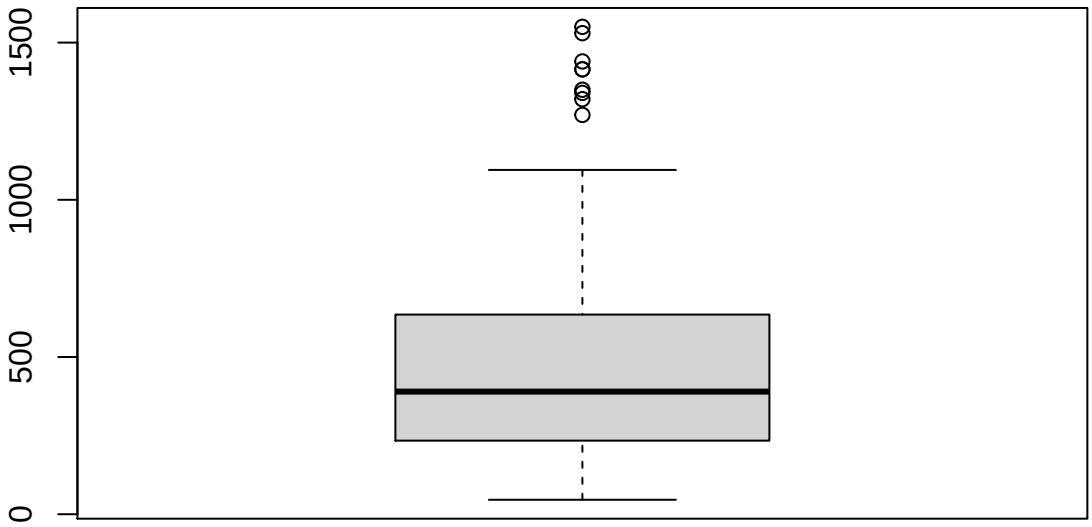
Tras una breve comprobación, vemos como cámaras como LEICA M10-R vale 8295 dólares, LEICA M10 Monochrom vale también 8295 dólares, o Fujifilm GFX 100 Specs cuesta 9999 dólares.

Por lo tanto, tras comprobar los valores extremos en la web, vemos que no son errores sino valores reales. Y decidimos quedárnoslos y trabajar con ellos.

Peso

Veamos los valores extremos del peso de las cámaras

```
boxplot(data$caracteristicas.peso)
```




```
out = boxplot.stats(data$caracteristicas.peso)$out
out
```

```
## [1] 1440 1270 1320 1415 1530 1415 1550 1350 1340
```

```
filter(data, caracteristicas.peso > 1700)$modelo
```

```
## character(0)
```

Este caso es similar al anterior, por lo que también comprobaremos un par de cámaras para ver si nos encontramos ante casos reales o errores.

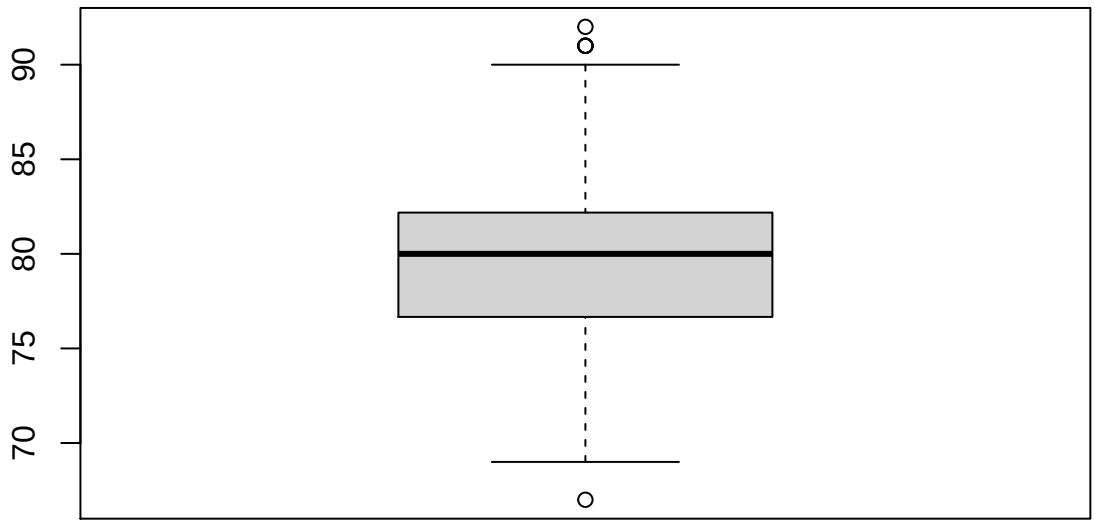
En la misma página web, podemos observar como Kodak DCS315 pesa 1.800 g y Kodak DCS760 pesa 1860 g.

Por lo tanto, nos encontramos ante valores reales y trabajaremos con ellos.

Puntuacion.proreview

Veamos los valores extremos de la puntuación proreview de las cámaras

```
boxplot(data$puntuacion.pro_review)
```




```
out = boxplot.stats(data$puntuacion.pro_review)$out
out
```

```
## [1] 91 91 92 91 91 67
```

En este caso también nos encontramos varios valores extremos, pero todos se comprenden entre el 0 y el 100, por lo que los damos como buenos ya que son los valores entre los que puede estar la cámara.

Archivo limpio

Para guardar el `dataframe` resultado del procesamiento utilizamos la función `write.csv`:

```
write.csv(data, 'dpreview_clean.csv', row.names = FALSE)
```

Análisis de los datos

Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).

En primer lugar se seleccionan los datos que se quieren analizar/comparar. En este ejercicio, el objetivo es realizar los análisis exploratorios que se han determinado en el primer apartado.

Para ello se realizará la comprobación de normalidad y homogeneidad de la varianza.

Análisis estadístico para comparar los grupos de datos: - Análisis del precio en función del tipo de cuerpo - Análisis del tipo de sensor por fecha de lanzamiento - Análisis de la evolución del tipo de cuerpo. -

Comparación de valoración de los usuarios entre cámaras Canon y Nicon - Creación de modelo para predecir la puntuación general de una cámara - Creación de modelo para predecir el precio de una cámara

Comprobación de la normalidad y la homogeneidad de la varianza

La comprobación de la normalidad es necesaria para poder realizar análisis posteriores como por ejemplo el contraste de hipótesis. Para la comprobación de la normalidad nos basaremos en el teorema del límite central, según el cual una muestra mayor de 30 se podrá considerar que sigue una distribución normal.

Precio en función del tipo de cuerpo

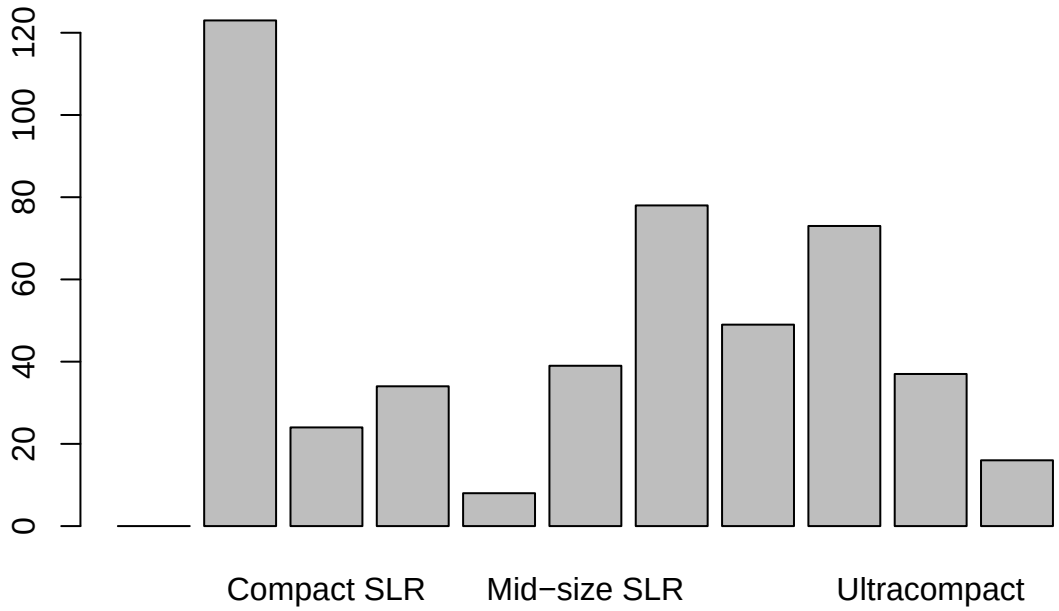
```
cuerpos = na.omit(unique(data$tipo_cuerpo))
for (cuerpo in cuerpos){
  print(cuerpo)
  print(length(filter(data, tipo_cuerpo == cuerpo)$precio))

}
```

```
## [1] "Mid-size SLR"
## [1] 39
## [1] "Rangefinder-style mirrorless"
## [1] 78
## [1] "SLR-style mirrorless"
## [1] 73
## [1] "Large sensor compact"
## [1] 34
## [1] "Compact"
## [1] 123
## [1] "VR/Action camera"
## [1] 16
## [1] "Compact SLR"
## [1] 24
## [1] "SLR-like (bridge)"
## [1] 49
## [1] "Large SLR"
## [1] 8
## [1] "Ultracompact"
## [1] 37
```

Como vemos, todas las muestras, a excepción de "VR/Action camera" cuentan con más de 30 elementos, por lo que se puede aplicar el Teorema del Límite Central y podemos considerar que la media de cada muestra sigue una distribución normal.

```
plot(x = data$tipo_cuerpo)
```



Puntuación general para Nikon y Canon

```
length( filter(data, marca=="Nikon")$puntuacion.general)
```

```
## [1] 75
```

```
length( filter(data, marca=="Canon")$puntuacion.general)
```

```
## [1] 89
```

En este caso, ambos conjuntos superan las 30 muestras y podemos asumir normalidad por el TLC.

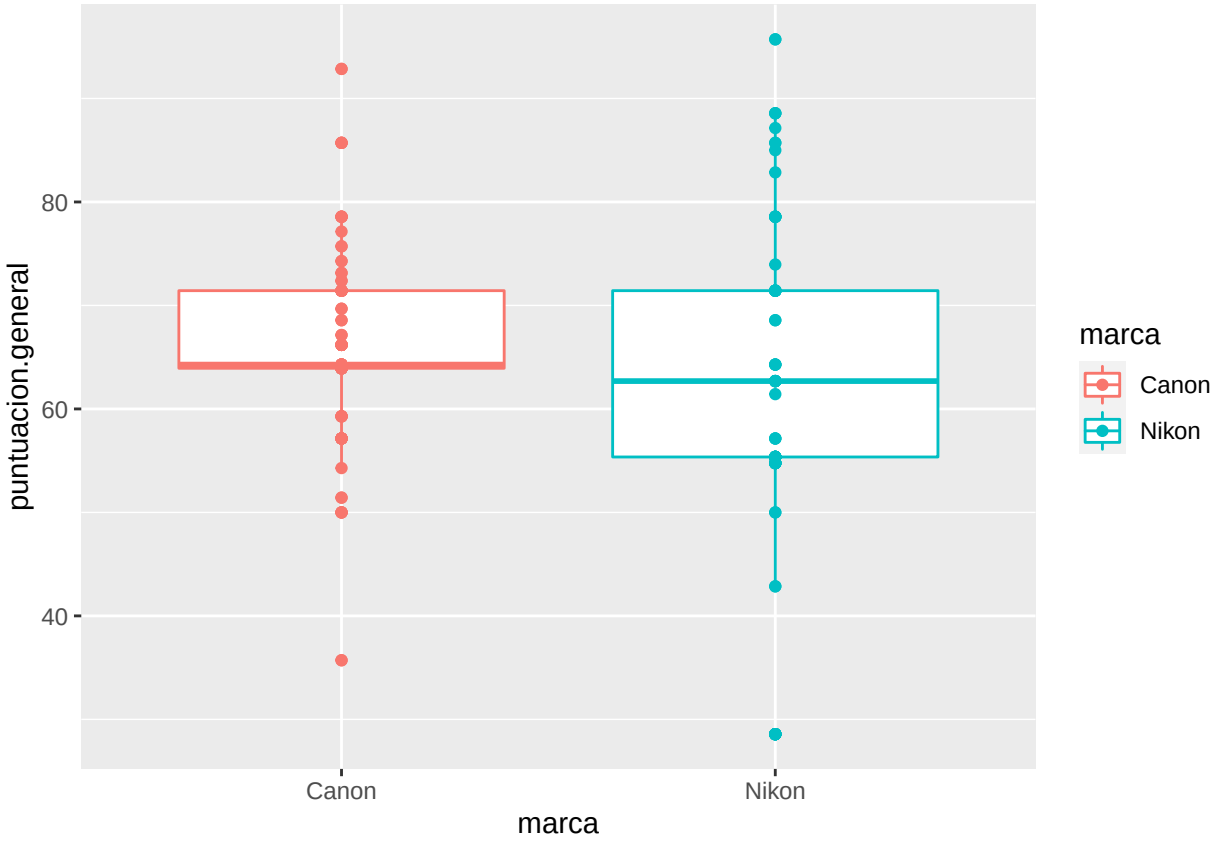
Procedemos a comprobar la homogeneidad de la varianza para los dos grupos mediante el test de Barlett.

```
bartlett.test(list(data$puntuacion.general[data$marca == "Nikon"],data$puntuacion.general[data$marca ==
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  list(data$puntuacion.general[data$marca == "Nikon"], data$puntuacion.general[data$marca == "Ca
## Bartlett's K-squared = 26.87, df = 1, p-value = 2.176e-07
```

Nos encontramos ante un p-value muy inferior a 0.05, por lo que podemos afirmar que ambas varianzas son similares.

```
data_canon_nikon = filter(data, marca %in% c("Nikon","Canon"))
ggplot(data=data_canon_nikon, aes(x=marca, y= puntuacion.general, colour = marca)) + geom_boxplot() + g
```

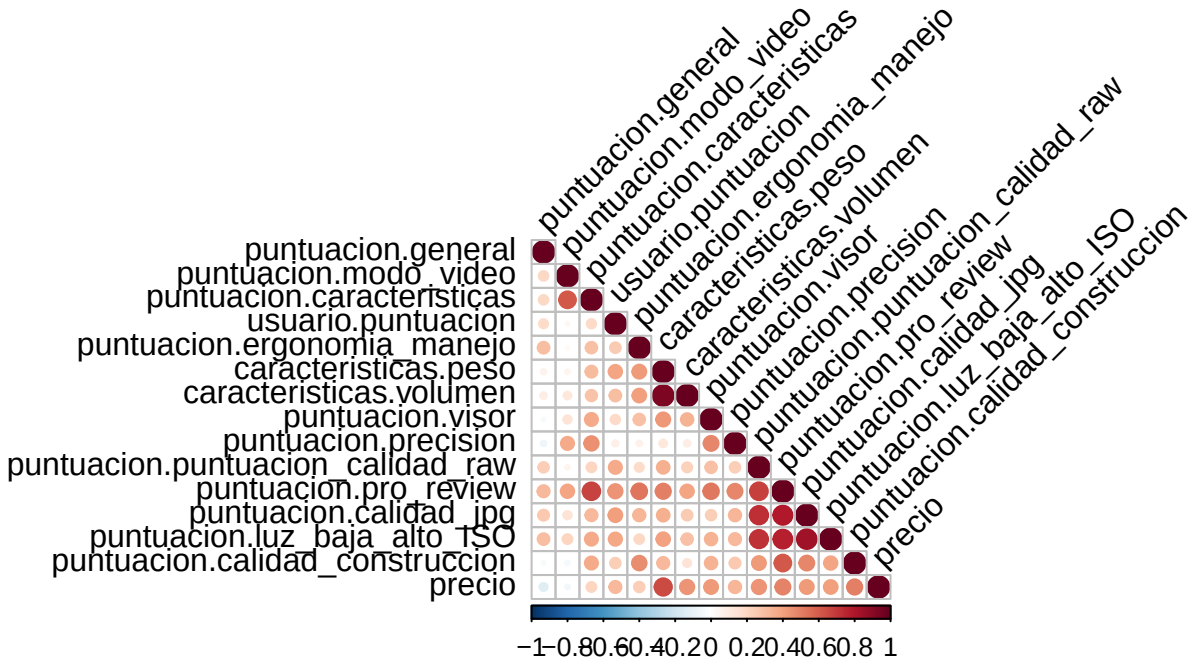


Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.

Correlación entre variables

En primer lugar vamos a realizar un análisis visual de la correlación entre las variables numéricas.

```
# Fuente: http://www.sthda.com/english/wiki/correlation-matrix-an-r-function-to-do-all-you-need
#install.packages("corrplot")
library(corrplot)
source("http://www.sthda.com/upload/rquery_cormat.r")
nums <- unlist(lapply(data, is.numeric))
rquery.cormat(data[ , nums])
```




```
## $r
##                                    puntuacion.general puntuacion.modo_video
## puntuacion.general                                  1                      
## puntuacion.modo_video                             0.2                      1
## puntuacion.caracteristicas                        0.2                   0.61
## usuario.puntuacion                               0.18                  0.033
## puntuacion.ergonomia_manejo                       0.3                  0.027
## caracteristicas.peso                            0.066                  0.052
## caracteristicas.volumen                         0.097                   0.13
## puntuacion.visor                               -0.032                   0.15
## puntuacion.precision                            -0.07                   0.37
## puntuacion.puntuacion_calidad_raw                0.24                   0.05
## puntuacion.pro_review                            0.32                   0.39
## puntuacion.calidad_jpg                           0.26                   0.15
## puntuacion.luz_baja_alto_ISO                      0.3                   0.21
## puntuacion.calidad_construccion                -0.034                 -0.049
## precio                                         -0.14                  -0.053
##                                    puntuacion.caracteristicas usuario.puntuacion
## puntuacion.general                                                               
## puntuacion.modo_video                                                            
## puntuacion.caracteristicas                                  1                    
## usuario.puntuacion                                       0.19                   1
## puntuacion.ergonomia_manejo                              0.29                0.25
## caracteristicas.peso                                     0.31                0.39
## caracteristicas.volumen                                  0.28                 0.3
## puntuacion.visor                                         0.37                0.19
```

```
## puntuacion.precision                                0.45              0.085
## puntuacion.puntuacion_calidad_raw                   0.21               0.37
## puntuacion.pro_review                               0.67               0.44
## puntuacion.calidad_jpg                              0.32               0.41
## puntuacion.luz_baja_alto_ISO                        0.37               0.38
## puntuacion.calidad_construccion                     0.37               0.24
## precio                                              0.21               0.32
##                                  puntuacion.ergonomia_manejo
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo                             1
## caracteristicas.peso                                 0.42
## caracteristicas.volumen                              0.41
## puntuacion.visor                                     0.29
## puntuacion.precision                                0.072
## puntuacion.puntuacion_calidad_raw                    0.18
## puntuacion.pro_review                                0.53
## puntuacion.calidad_jpg                               0.33
## puntuacion.luz_baja_alto_ISO                          0.2
## puntuacion.calidad_construccion                      0.46
## precio                                               0.24
##                                  caracteristicas.peso caracteristicas.volumen
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso                                1
## caracteristicas.volumen                          0.93                      1
## puntuacion.visor                                 0.43                   0.34
## puntuacion.precision                             0.13                  0.096
## puntuacion.puntuacion_calidad_raw                0.35                   0.22
## puntuacion.pro_review                             0.5                   0.39
## puntuacion.calidad_jpg                           0.35                   0.25
## puntuacion.luz_baja_alto_ISO                      0.4                   0.29
## puntuacion.calidad_construccion                  0.32                   0.14
## precio                                           0.65                   0.44
##                                  puntuacion.visor puntuacion.precision
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor                                1
## puntuacion.precision                         0.48                   1
## puntuacion.puntuacion_calidad_raw            0.29                0.24
## puntuacion.pro_review                        0.52                0.48
## puntuacion.calidad_jpg                       0.24                0.33
## puntuacion.luz_baja_alto_ISO                 0.34                0.32
## puntuacion.calidad_construccion              0.34                0.26
```

```
## precio                                        0.42                0.33
##                                   puntuacion.puntuacion_calidad_raw
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw                                  1
## puntuacion.pro_review                                           0.68
## puntuacion.calidad_jpg                                          0.73
## puntuacion.luz_baja_alto_ISO                                    0.72
## puntuacion.calidad_construccion                                 0.42
## precio                                                          0.44
##                                   puntuacion.pro_review puntuacion.calidad_jpg
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review                                     1
## puntuacion.calidad_jpg                                 0.79                    1
## puntuacion.luz_baja_alto_ISO                           0.77                 0.84
## puntuacion.calidad_construccion                        0.61                 0.48
## precio                                                 0.49                 0.42
##                                   puntuacion.luz_baja_alto_ISO
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO                                  1
## puntuacion.calidad_construccion                            0.39
## precio                                                     0.41
##                                   puntuacion.calidad_construccion precio
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
```

```
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO
## puntuacion.calidad_construccion                                1
## precio                                                       0.5      1
##
## $p
##                                 puntuacion.general puntuacion.modo_video
## puntuacion.general                               0
## puntuacion.modo_video                      2.6e-07                    0
## puntuacion.caracteristicas                 5.4e-07                2.2e-51
## usuario.puntuacion                          0.0015                   0.71
## puntuacion.ergonomia_manejo                3.3e-12                   0.46
## caracteristicas.peso                         0.089                    0.3
## caracteristicas.volumen                      0.031                  0.012
## puntuacion.visor                              0.13                0.00086
## puntuacion.precision                         0.042                9.4e-16
## puntuacion.puntuacion_calidad_raw          5.2e-05                   0.65
## puntuacion.pro_review                      6.8e-11                2.7e-16
## puntuacion.calidad_jpg                     2.5e-07                 0.0016
## puntuacion.luz_baja_alto_ISO               1.3e-09                1.8e-06
## puntuacion.calidad_construccion               0.17                  0.094
## precio                                       0.0018                    0.2
##                                 puntuacion.caracteristicas usuario.puntuacion
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas                               0
## usuario.puntuacion                                   5e-04                  0
## puntuacion.ergonomia_manejo                          2.3e-12              1e-06
## caracteristicas.peso                                 4.8e-12            5.5e-14
## caracteristicas.volumen                                2e-10            4.1e-09
## puntuacion.visor                                     9.5e-16            0.00014
## puntuacion.precision                                 1.7e-20              0.089
## puntuacion.puntuacion_calidad_raw                    0.00019              2e-12
## puntuacion.pro_review                                1.3e-59            8.6e-18
## puntuacion.calidad_jpg                               1.1e-11            2.8e-14
## puntuacion.luz_baja_alto_ISO                         9.3e-17            9.5e-13
## puntuacion.calidad_construccion                      4.3e-14            1.9e-06
## precio                                               1.8e-05            7.2e-10
##                                 puntuacion.ergonomia_manejo
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo                                0
## caracteristicas.peso                                 4.7e-21
## caracteristicas.volumen                              3.8e-19
## puntuacion.visor                                     7.3e-10
```

```
## puntuacion.precision                                  0.37
## puntuacion.puntuacion_calidad_raw                  0.00093
## puntuacion.pro_review                              9.6e-39
## puntuacion.calidad_jpg                             8.6e-14
## puntuacion.luz_baja_alto_ISO                       1.5e-06
## puntuacion.calidad_construccion                    1.2e-27
## precio                                             8.8e-08
##                                 caracteristicas.peso caracteristicas.volumen
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso                               0
## caracteristicas.volumen                        3.8e-204                    0
## puntuacion.visor                               2.4e-20              3.9e-14
## puntuacion.precision                            0.0067                0.036
## puntuacion.puntuacion_calidad_raw              4.7e-15              2.6e-06
## puntuacion.pro_review                          1.2e-30              1.8e-18
## puntuacion.calidad_jpg                         1.1e-14              5.9e-08
## puntuacion.luz_baja_alto_ISO                   1.8e-16              7.4e-09
## puntuacion.calidad_construccion                1.8e-14                1e-04
## precio                                           2e-60                4e-28
##                                 puntuacion.visor puntuacion.precision
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor                               0
## puntuacion.precision                      1.1e-32                    0
## puntuacion.puntuacion_calidad_raw         1.6e-10              5.2e-09
## puntuacion.pro_review                     1.9e-32                4e-26
## puntuacion.calidad_jpg                      1e-07              2.1e-13
## puntuacion.luz_baja_alto_ISO              2.1e-13              1.4e-13
## puntuacion.calidad_construccion           4.8e-15              1.1e-06
## precio                                    4.4e-21              7.9e-13
##                                 puntuacion.puntuacion_calidad_raw
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw                                 0
## puntuacion.pro_review                                      1.2e-56
## puntuacion.calidad_jpg                                     2.6e-79
## puntuacion.luz_baja_alto_ISO                               9.7e-74
## puntuacion.calidad_construccion                              3e-22
```

```
## precio                                                    1.9e-24
##                                  puntuacion.pro_review puntuacion.calidad_jpg
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review                                 0
## puntuacion.calidad_jpg                          4.4e-99                     0
## puntuacion.luz_baja_alto_ISO                    1.2e-88                2.6e-133
## puntuacion.calidad_construccion                 5.2e-47                 1.2e-25
## precio                                          3.7e-30                 1.2e-20
##                                  puntuacion.luz_baja_alto_ISO
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO                               0
## puntuacion.calidad_construccion                       6.2e-17
## precio                                                  2e-19
##                                  puntuacion.calidad_construccion precio
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO
## puntuacion.calidad_construccion                                0
## precio                                                  5.6e-32      0
##
## $sym
##                                  puntuacion.general puntuacion.modo_video
## puntuacion.general               1
## puntuacion.modo_video                               1
```

```
## puntuacion.caracteristicas                          ,
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision                                 .
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review                .               .
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO
## puntuacion.calidad_construccion
## precio
##                                   puntuacion.caracteristicas usuario.puntuacion
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas        1
## usuario.puntuacion                                           1
## puntuacion.ergonomia_manejo
## caracteristicas.peso              .                          .
## caracteristicas.volumen
## puntuacion.visor                  .
## puntuacion.precision              .
## puntuacion.puntuacion_calidad_raw                            .
## puntuacion.pro_review             ,                          .
## puntuacion.calidad_jpg            .                          .
## puntuacion.luz_baja_alto_ISO      .                          .
## puntuacion.calidad_construccion   .
## precio                                                       .
##                                   puntuacion.ergonomia_manejo
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo       1
## caracteristicas.peso              .
## caracteristicas.volumen           .
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review             .
## puntuacion.calidad_jpg            .
## puntuacion.luz_baja_alto_ISO
## puntuacion.calidad_construccion   .
## precio
##                                   caracteristicas.peso caracteristicas.volumen
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso              1
## caracteristicas.volumen           *                    1
## puntuacion.visor                  .                     .
```

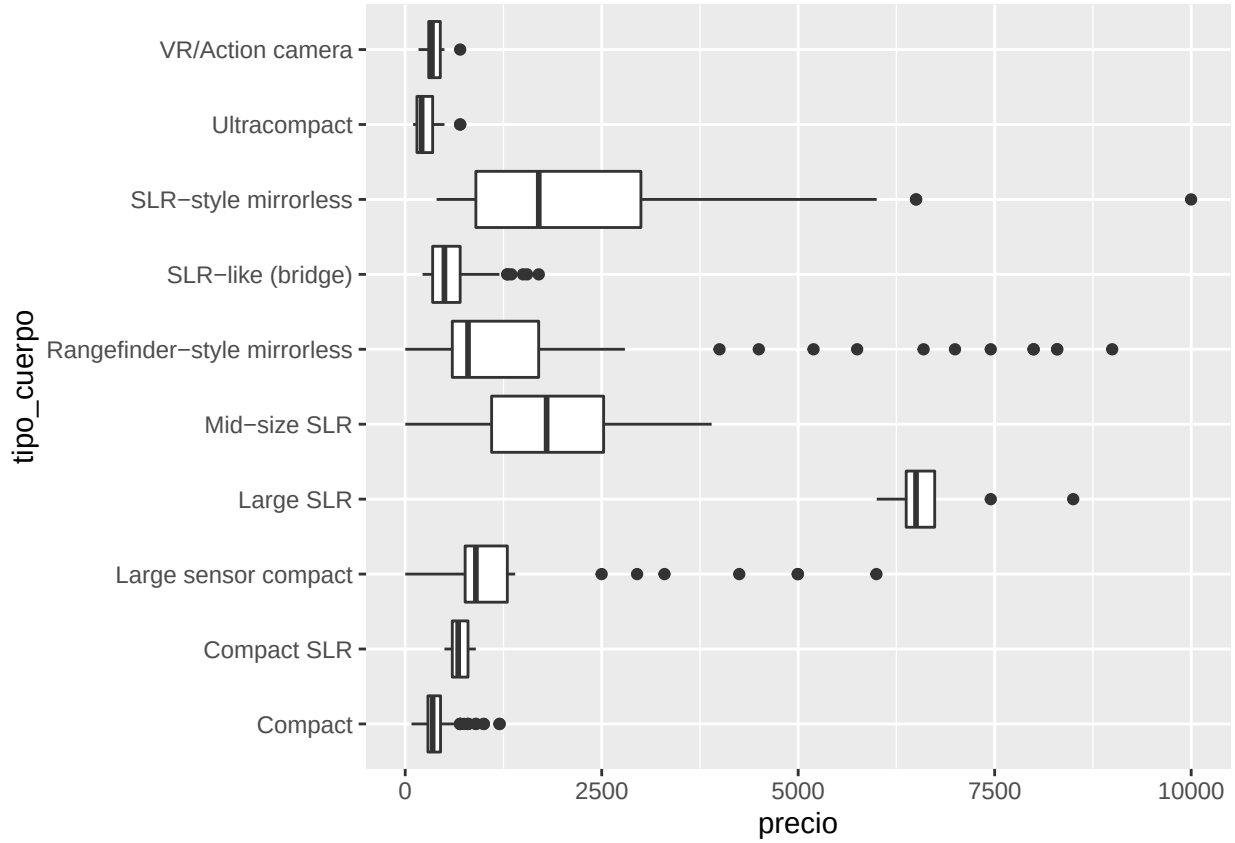
```
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw .
## puntuacion.pro_review             .                     .
## puntuacion.calidad_jpg            .
## puntuacion.luz_baja_alto_ISO      .
## puntuacion.calidad_construccion   .
## precio                            ,                     .
##                                   puntuacion.visor puntuacion.precision
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor                  1
## puntuacion.precision              .                 1
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review             .                 .
## puntuacion.calidad_jpg                              .
## puntuacion.luz_baja_alto_ISO      .                 .
## puntuacion.calidad_construccion   .
## precio                            .                 .
##                                   puntuacion.puntuacion_calidad_raw
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw 1
## puntuacion.pro_review             ,
## puntuacion.calidad_jpg            ,
## puntuacion.luz_baja_alto_ISO      ,
## puntuacion.calidad_construccion   .
## precio                            .
##                                   puntuacion.pro_review puntuacion.calidad_jpg
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review             1
## puntuacion.calidad_jpg            ,                     1
## puntuacion.luz_baja_alto_ISO      ,                     +
## puntuacion.calidad_construccion   ,                     .
```

```
## precio                              .                          .
##                                     puntuacion.luz_baja_alto_ISO
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO        1
## puntuacion.calidad_construccion     .
## precio                              .
##                                     puntuacion.calidad_construccion precio
## puntuacion.general
## puntuacion.modo_video
## puntuacion.caracteristicas
## usuario.puntuacion
## puntuacion.ergonomia_manejo
## caracteristicas.peso
## caracteristicas.volumen
## puntuacion.visor
## puntuacion.precision
## puntuacion.puntuacion_calidad_raw
## puntuacion.pro_review
## puntuacion.calidad_jpg
## puntuacion.luz_baja_alto_ISO
## puntuacion.calidad_construccion     1
## precio                              .                                  1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1
```

```
#corrplot(cor(data[, nums]), type="upper")
```

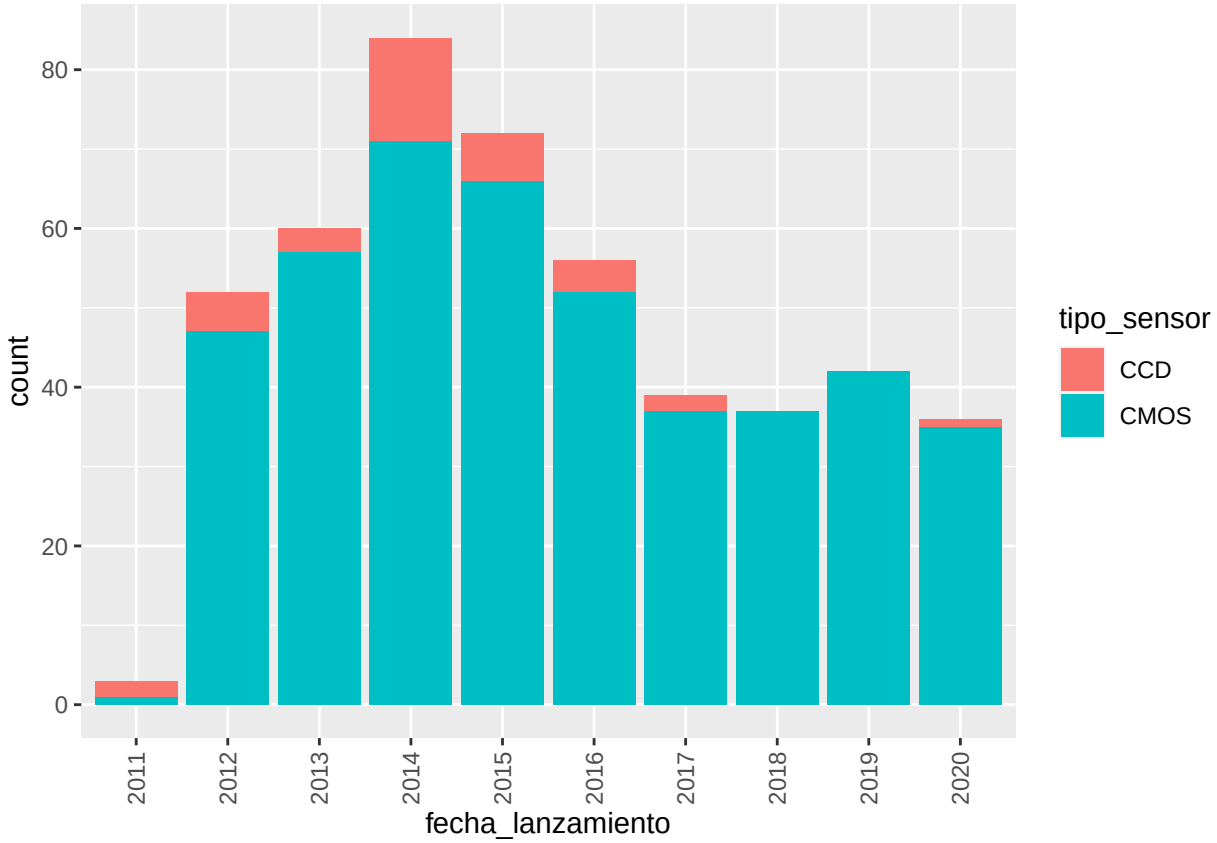
Análisis exploratorio

```
ggplot(data, aes(x=precio, y=tipo_cuerpo)) +
    geom_boxplot()
```



Lógicamente las cámaras más caras son las SLR "grandes", que son las que se emplean profesionalmente. Las compactas, ultracompactas y cámaras de acción son las más asequibles.

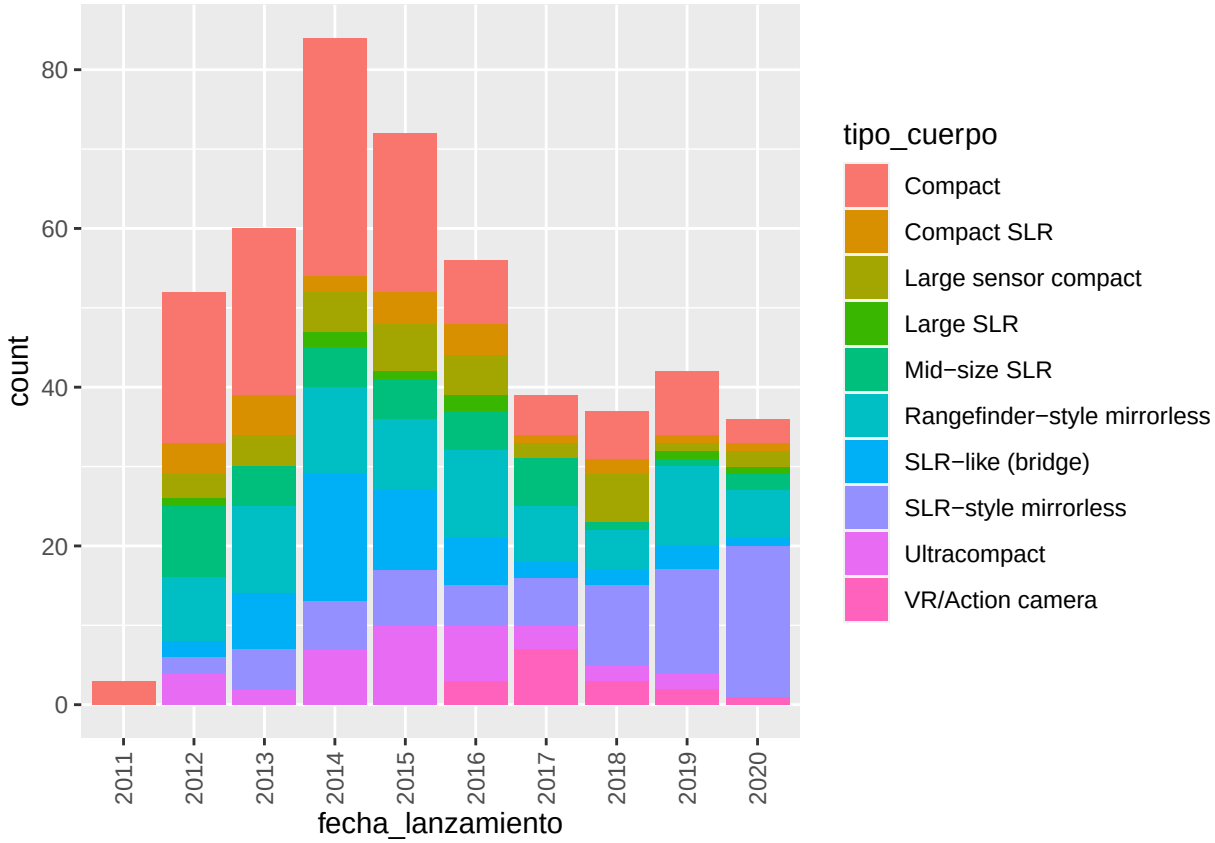
```
ggplot(data, aes(x=fecha_lanzamiento, fill=tipo_sensor)) +
    geom_bar()+theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Se puede observar que el *boom* de las cámaras digitales tuvo lugar entre el año 2000 y el año 2015. Posiblemente el declive se debió a mejora de las cámaras de los dispositivos móviles. Asimismo se puede observar, a través del código de colores, la evolución en el tipo de sensores que equipan las cámaras, pasando de sensores CCD en las fases iniciales a sensores CMOS, que han copado el mercado desde el 2013.

Veamos la evolución a continuación en función del tipo de cámara:

```
ggplot(data, aes(x=fecha_lanzamiento, fill=tipo_cuerpo)) +
    geom_bar()+theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Como habíamos indicado anteriormente, la aparición de dispositivos móviles con cámaras de calidad ha hecho desaparecer prácticamente por completo el segmento de las cámaras compactas y ultracompactas, manteniendose únicamente actualizadas el sector de las cámaras profesionales.

Pruebas de hipótesis

Siempre ha habido un gran debate entre los profesionales de la fotografía acerca de qué marca es mejor, si Canon o Nikon. Vamos a verificar mediante un contraste de hipótesis sobre la variable `puntuacion.general` si existe una diferencia significativa entre ambas marcas.

Las hipótesis nula y alternativa serían:

- H_0: No existe una diferencia significativa entre Canon y Nikon
- H_1: Sí existe una diferencia significativa entre Canon y Nikon

Suponemos normalidad en las muestras por el Teorema del Límite Central dado que contamos con un conjunto de muestras grande.

Comprobamos la homogeneidad de varianzas

```
var.test( data %>%filter(marca=="Nikon")%>%pull(puntuacion.general), data %>%filter(marca=="Canon")%>%pu
```

```
##
##  F test to compare two variances
##
## data:  data %>% filter(marca == "Nikon") %>% pull(puntuacion.general) and data %>% filter(marca == "C
## F = 3.2306, num df = 74, denom df = 88, p-value = 2.097e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

```
##  2.088823 5.039121
## sample estimates:
## ratio of variances
##           3.230636
```

El p-value da un valor por debajo de 0.05, por lo que podemos suponer que las muestras tienen la misma varianza con un 95% de confianza.

Aplicamos a continuación el contraste de hipótesis

```
t.test( data %>%filter(marca=="Nikon")%>%pull(puntuacion.general),
        data %>%filter(marca=="Canon")%>%pull(puntuacion.general), # dos muestras
        alternative = "two.sided", # contraste bilateral
        paired = FALSE, # muestras independientes
        var.equal = TRUE ) # se supone homocedasticidad
```

```
##
##  Two Sample t-test
##
## data:  data %>% filter(marca == "Nikon") %>% pull(puntuacion.general) and data %>% filter(marca == "
## t = -1.8959, df = 162, p-value = 0.05975
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.1405108  0.1453615
## sample estimates:
## mean of x mean of y
##  62.69053  66.18811
```

En este caso el resultado del t-test ofrece un `p-value` de 4.251e-08, inferior a 0.05, con lo que rechazamos la hipótesis nula y confirmamos con un 95% de confianza que nohay diferencia significativa entre la puntuación general de Canon y Nikon.

Regresión lineal

Puntuacion

A continuación construiremos un modelo de regresión lineal para tratar de predecir la puntuación de la cámara en función de diferentes valores.

Los valores que se analizarán son el precio, el peso, la puntuación en ergonomía, la puntuación de las características y la puntuación de la precisión de la cámara.

```
lineal_regression = lm(puntuacion.general ~ precio + caracteristicas.peso + puntuacion.ergonomia_manejo
summary(lineal_regression)
```

```
##
## Call:
## lm(formula = puntuacion.general ~ precio + caracteristicas.peso +
##     puntuacion.ergonomia_manejo + puntuacion.caracteristicas +
##     puntuacion.precision, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.272  -7.803   1.886   7.197  36.087
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 29.3791477  8.6155551   3.410 0.000709 ***
```

```
## precio                       -0.0021229  0.0005005  -4.241 2.71e-05 ***
## caracteristicas.peso          0.0037700  0.0031463   1.198 0.231478
## puntuacion.ergonomia_manejo   0.4139530  0.0709077   5.838 1.03e-08 ***
## puntuacion.caracteristicas    0.3151020  0.0732127   4.304 2.07e-05 ***
## puntuacion.precision         -0.2556355  0.1116966  -2.289 0.022571 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.14 on 441 degrees of freedom
##   (34 observations deleted due to missingness)
## Multiple R-squared:  0.1903, Adjusted R-squared:  0.1811
## F-statistic: 20.73 on 5 and 441 DF,  p-value: < 2.2e-16
```

Podemos ver como el precio o el peso prácticamente no tienen relación con la puntuación general que recibe una cámara.

La puntuación en la ergonomia o las características si que afectan de forma positiva en la puntuación general que recibirá la cámara. En cambio, la puntuación en la precisión de la cámara afecta de forma negativa en la puntuación final que esta tendrá.

Aún y así, la R-squared de la regresión lineal hecha es de 0.25 por lo que las previsiones que se pueden hacer con este modelo no se ajustarán mucho a la realidad.

Precio A continuación, veremos si podemos crear un modelo que se ajuste más a la realidad que el anterior, para predecir el precio de una cámara.

Los valores que se analizarán son la puntuación general, si tiene GPS o no, la puntuación en ergonomía, la puntuación de las características , la puntuación del visor y el tipo de cuerpo de la cámara

```
lineal_regression = lm(precio ~ puntuacion.general + GPS + puntuacion.ergonomia_manejo + puntuacion.cara
summary(lineal_regression)
```

```
##
## Call:
## lm(formula = precio ~ puntuacion.general + GPS + puntuacion.ergonomia_manejo +
##     puntuacion.caracteristicas + puntuacion.visor + tipo_cuerpo,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2947.9  -567.9   -81.8   340.5  7276.9
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             -1205.397    816.263  -1.477  0.14047
## puntuacion.general                        -35.200      4.871  -7.227 2.21e-12
## GPSNone                                   -91.622    182.872  -0.501  0.61661
## GPSOptional                               380.253    225.124   1.689  0.09191
## puntuacion.ergonomia_manejo                 9.793      6.985   1.402  0.16163
## puntuacion.caracteristicas                  6.696      7.234   0.926  0.35517
## puntuacion.visor                           34.930      7.638   4.573 6.26e-06
## tipo_cuerpoCompact SLR                    771.875    278.797   2.769  0.00587
## tipo_cuerpoLarge sensor compact           777.021    244.812   3.174  0.00161
## tipo_cuerpoLarge SLR                     5368.026    461.171  11.640  < 2e-16
## tipo_cuerpoMid-size SLR                   994.215    252.588   3.936 9.63e-05
## tipo_cuerpoRangefinder-style mirrorless  1349.019    175.222   7.699 9.18e-14
```

```
## tipo_cuerpoSLR-like (bridge)               -8.073    202.084  -0.040  0.96815
## tipo_cuerpoSLR-style mirrorless          1782.200    203.546   8.756  < 2e-16
## tipo_cuerpoUltracompact                  -252.130    223.284  -1.129  0.25943
## tipo_cuerpoVR/Action camera              -209.247    676.454  -0.309  0.75722
##
## (Intercept)
## puntuacion.general                       ***
## GPSNone
## GPSOptional                              .
## puntuacion.ergonomia_manejo
## puntuacion.caracteristicas
## puntuacion.visor                         ***
## tipo_cuerpoCompact SLR                   **
## tipo_cuerpoLarge sensor compact          **
## tipo_cuerpoLarge SLR                     ***
## tipo_cuerpoMid-size SLR                  ***
## tipo_cuerpoRangefinder-style mirrorless ***
## tipo_cuerpoSLR-like (bridge)
## tipo_cuerpoSLR-style mirrorless          ***
## tipo_cuerpoUltracompact
## tipo_cuerpoVR/Action camera
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1143 on 439 degrees of freedom
##   (26 observations deleted due to missingness)
## Multiple R-squared:  0.5263, Adjusted R-squared:  0.5101
## F-statistic: 32.52 on 15 and 439 DF,  p-value: < 2.2e-16
```

En este caso vemos como la mayoría de factores tienen una gran repercusión en el precio final.

Por ejemplo, el hecho de no tener GPS disminuye en 130 dólares el precio final.

También podemos ver como el tipo de cuerpo "Large SLR" aumenta el precio final en 5.320 dólares.

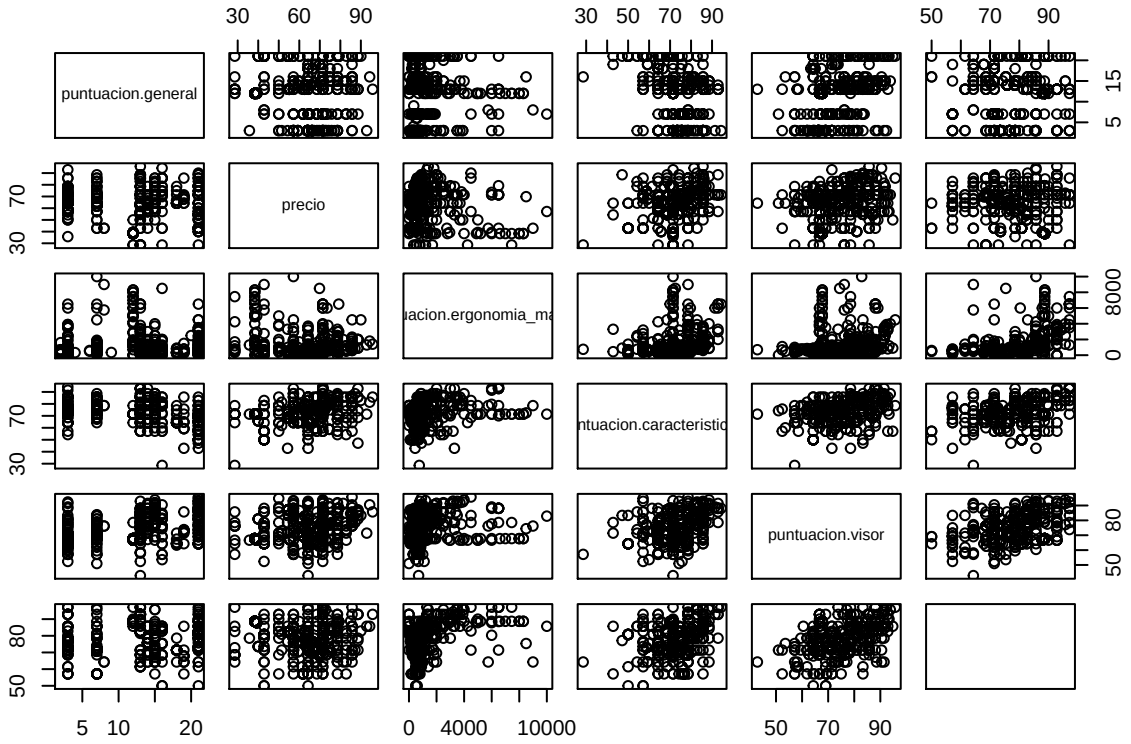
Además, un hecho sorprendete es que la puntuación general es inversamente proporcional al precio que tendrá la cámara.

En este caso, la R-squared nos da un valor de 0.5272, que aunque no es el valor deseado, si que es un valor con el que se puede trabajar y hacer predicciones.

Correlación

En este apartado buscaremos si existe correlación entre las variables numéricas que hemos usado con anterioridad, que son: puntuacion.general, precio, puntuacion.ergonomia_manejo, puntuacion.caracteristicas y
puntuacion.visor.

```
data_correlation = select(data, x=c(puntuacion.general, precio, puntuacion.ergonomia_manejo, puntuacion
colnames(data_correlation) = c("puntuacion.general", "precio", "puntuacion.ergonomia_manejo", "puntuaci
plot(data_correlation)
```



En los gráficos anteriores podemos ver las diferentes características numéricas pintadas una frente a otra. Aunque no se aprecia ningún caso de relación lineal, haremos el análisis de correlaciones en profundidad.

```
cor.test(data$puntuacion.general, data$precio, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.general and data$precio
## S = 13978072, p-value = 0.004397
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.1327136
```

```
cor.test(data$puntuacion.general, data$puntuacion.ergonomia_manejo, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.general and data$puntuacion.ergonomia_manejo
## S = 10993626, p-value = 3.088e-12
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.3178872
```

```r
cor.test(data$puntuacion.general, data$puntuacion.caracteristicas, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.general and data$puntuacion.caracteristicas
## S = 12455838, p-value = 8.745e-07
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2271624
```

```r
cor.test(data$puntuacion.general, data$puntuacion.visor, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.general and data$puntuacion.visor
## S = 17611468, p-value = 0.0471
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.09272483
```

```r
cor.test(data$precio, data$puntuacion.ergonomia_manejo, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$precio and data$puntuacion.ergonomia_manejo
## S = 10795394, p-value = 3.891e-13
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.3301867
```

```r
cor.test(data$precio, data$puntuacion.caracteristicas, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$precio and data$puntuacion.caracteristicas
## S = 11073911, p-value = 6.96e-12
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.3129058
```

```r
cor.test(data$precio, data$puntuacion.visor, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$precio and data$puntuacion.visor
## S = 10294338, p-value = 1.345e-15
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```
##       rho
## 0.3612753
```

```
cor.test(data$puntuacion.ergonomia_manejo, data$puntuacion.caracteristicas, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.ergonomia_manejo and data$puntuacion.caracteristicas
## S = 11404854, p-value = 1.693e-10
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2923721
```

```
cor.test(data$puntuacion.ergonomia_manejo, data$puntuacion.visor, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.ergonomia_manejo and data$puntuacion.visor
## S = 12261179, p-value = 2.136e-07
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2392403
```

```
cor.test(data$puntuacion.caracteristicas, data$puntuacion.visor, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  data$puntuacion.caracteristicas and data$puntuacion.visor
## S = 10953457, p-value = 2.045e-12
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.3203795
```

El test de correlación devuelve un resultado entre [-1,1] donde los extremos indican una alta correlación y el 0 una correlación nula.

En este cruce entre las variables anteriores, la mayoría de ellas muestran valores muy próximos al 0, por lo que en la mayoría de ellas no existe correlación.

Aún y así podemos encontrar una correlación positiva entre la puntuación de las características y la puntuación del visor, entre el precio y la punutuación de ergonomía y manejo, y entre el precio y la puntuación de las características.

Representación de los resultados a partir de tablas y gráficas.

Todos los resultados obtenidos se han ido presentando en cada uno de los apartados en tablas y gráficas.

Resolución del problema. A partir de los resultados obtenidos, ¿Cuáles son las conclusiones? ¿Los resultados permiten responder al problema?

Creemos que el dataset con el que se ha trabajado en esta práctica resulta ideal para los objetivos de la misma. Es un *dataset* que requiere mucho tratamiento de limpieza, extracción de valores y conversión de tipos. El hecho de que no sea un *dataset* académico clásico imposibilita encontrar referencias de trabajos previos de limpieza.

Los datos, una vez tratados, nos han servido para elaborar modelos, que aunque con baja precisión, se podrían emplear para realizar predicciones.

En cuanto a las conclusiones que hemos obtenido a partir de los análisis realizados sobre los datos, a continuación las listamos:

- Las cámaras más caras son las SLR grandes, mientras que las más baratas son compactas, ultracompactas y cámaras de acción asequibles.
- El boom de las cámaras dígitales tuvo lugar entre 2014 y 2015, mientras que a partir de ahí los precios han ido disminuyendo como consecuencia de las mejoras en las cámaras de los móviles. *Las cámaras compactas y ultracompactas prácticamente han desaparecido debido a la explosión de las cámaras en los teléfonos móviles, mientras que las cámaras profesionales se han mantenido.
- Se puede afirmar con un p-value de 4,2e-8 que no hay diferencias en las puntuaciones de las cámaras Canon y Nikon.
- Valores como el peso o el precio prácticamente no tienen relación con la puntuación general de la cámara, en cambio la ergonomía o las características de esta si que afectan de forma positiva, mientras que la precisión afecta de forma negativa.
- Mientras que la puntuación general de la cámara es inversamente proporcional al precio final de esta, el hecho de tener GPS aumenta 130 dólares el precio final, o el hecho de ser de cuerpo tipo “Large SLR” lo aumenta en 5.320 dólares. *Aunque entre la mayoría de variables no existen correlaciones, podemos ver como si que existen entre la puntuación de las características y la puntuación del visor, entre el precio y la puntuación de ergonomía, y entre el precio y la puntuación de las características.

Referencias

- Introducción a la limpieza y análisis de datos. Calvo, M., Pérez, D., Subirats, L. (2019). Editorial UOC.
- Identificar y eliminar duplicados: [https://www.datanovia.com/en/lessons/identify-and-remove-duplicate-data-in-r/]
- Tabla con medidas de tendencia central y dispersión: [https://cran.r-project.org/web/packages/qwraps2/vignettes/summary-statistics.html]
- My favourite R package for: summarising data [https://dabblingwithdata.wordpress.com/2018/01/02/my-favourite-r-package-for-summarising-data/]