

Project Data Preparation

Graciela Rivera/Luis E. Vargas

2023-03-22

Apertura de archivo

```
datosEcommerce <- read.csv('data-2.csv', sep=",")
```

Inspección del archivo

```
names(datosEcommerce)
```

```
## [1] "InvoiceNo" "StockCode" "Description" "Quantity" "InvoiceDate"  
## [6] "UnitPrice" "CustomerID" "Country"
```

```
str(datosEcommerce)
```

```
## 'data.frame': 541909 obs. of 8 variables:  
## $ InvoiceNo : chr "536365" "536365" "536365" "536365" ...  
## $ StockCode : chr "85123A" "71053" "84406B" "84029G" ...  
## $ Description: chr "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS ...  
## $ Quantity : int 6 6 8 6 6 2 6 6 6 32 ...  
## $ InvoiceDate: chr "12/1/2010 8:26" "12/1/2010 8:26" "12/1/2010 8:26" "12/1/2010 8:26" ...  
## $ UnitPrice : num 2.55 3.39 2.75 3.39 3.39 7.65 4.25 1.85 1.85 1.69 ...  
## $ CustomerID: int 17850 17850 17850 17850 17850 17850 17850 17850 17850 13047 ...  
## $ Country : chr "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...
```

```
nrow(datosEcommerce)
```

```
## [1] 541909
```

```
summary(datosEcommerce)
```

```
## InvoiceNo StockCode Description Quantity  
## Length:541909 Length:541909 Length:541909 Min. :-80995.00  
## Class :character Class :character Class :character 1st Qu.: 1.00  
## Mode :character Mode :character Mode :character Median : 3.00  
## Mean : 9.55  
## 3rd Qu.: 10.00  
## Max. : 80995.00  
##  
## InvoiceDate UnitPrice CustomerID Country  
## Length:541909 Min. :-11062.06 Min. :12346 Length:541909  
## Class :character 1st Qu.: 1.25 1st Qu.:13953 Class :character  
## Mode :character Median : 2.08 Median :15152 Mode :character  
## Mean : 4.61 Mean :15288  
## 3rd Qu.: 4.13 3rd Qu.:16791  
## Max. : 38970.00 Max. :18287
```

```
## NA's :135080
```

Asociar el tipo de dato a Date

```
#datosEcommerce$InvoiceDate <- as.POSIXct(datosEcommerce$InvoiceDate, format="%m/%d/%Y %H:%M")
datosEcommerce$InvoiceDate <- as.Date(datosEcommerce$InvoiceDate, format="%m/%d/%Y")
str(datosEcommerce)
```

```
## 'data.frame': 541909 obs. of 8 variables:
## $ InvoiceNo : chr "536365" "536365" "536365" "536365" ...
## $ StockCode : chr "85123A" "71053" "84406B" "84029G" ...
## $ Description: chr "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS COAT HANGER" ...
## $ Quantity : int 6 6 8 6 6 2 6 6 6 32 ...
## $ InvoiceDate: Date, format: "2010-12-01" "2010-12-01" ...
## $ UnitPrice : num 2.55 3.39 2.75 3.39 3.39 7.65 4.25 1.85 1.85 1.69 ...
## $ CustomerID : int 17850 17850 17850 17850 17850 17850 17850 17850 17850 13047 ...
## $ Country : chr "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...
```

Crea nuevo atributo para ventas, apartir de los atributos de cantidad del item y el precio unitario

```
datosEcommerce$Sales <- datosEcommerce$Quantity * datosEcommerce$UnitPrice
str(datosEcommerce)
```

```
## 'data.frame': 541909 obs. of 9 variables:
## $ InvoiceNo : chr "536365" "536365" "536365" "536365" ...
## $ StockCode : chr "85123A" "71053" "84406B" "84029G" ...
## $ Description: chr "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS COAT HANGER" ...
## $ Quantity : int 6 6 8 6 6 2 6 6 6 32 ...
## $ InvoiceDate: Date, format: "2010-12-01" "2010-12-01" ...
## $ UnitPrice : num 2.55 3.39 2.75 3.39 3.39 7.65 4.25 1.85 1.85 1.69 ...
## $ CustomerID : int 17850 17850 17850 17850 17850 17850 17850 17850 17850 13047 ...
## $ Country : chr "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...
## $ Sales : num 15.3 20.3 22 20.3 20.3 ...
```

Exploración de Datos

```
head(datosEcommerce, n=10)
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate
## 1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01
## 2	536365	71053	WHITE METAL LANTERN	6	2010-12-01
## 3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01
## 4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01
## 5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01
## 6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01
## 7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01
## 8	536366	22633	HAND WARMER UNION JACK	6	2010-12-01
## 9	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01
## 10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01

	UnitPrice	CustomerID	Country	Sales
## 1	2.55	17850	United Kingdom	15.30
## 2	3.39	17850	United Kingdom	20.34
## 3	2.75	17850	United Kingdom	22.00

```
## 4      3.39      17850 United Kingdom 20.34
## 5      3.39      17850 United Kingdom 20.34
## 6      7.65      17850 United Kingdom 15.30
## 7      4.25      17850 United Kingdom 25.50
## 8      1.85      17850 United Kingdom 11.10
## 9      1.85      17850 United Kingdom 11.10
## 10     1.69      13047 United Kingdom 54.08
```

```
#kable(head(datosEcommerce,20)) %>% kable_styling()
```

Generación de Histograma

```
hist(x = datosEcommerce$Sales,
     main="Histograma",
     xlab = "Ventas",
     ylab = "Frecuencia")
```

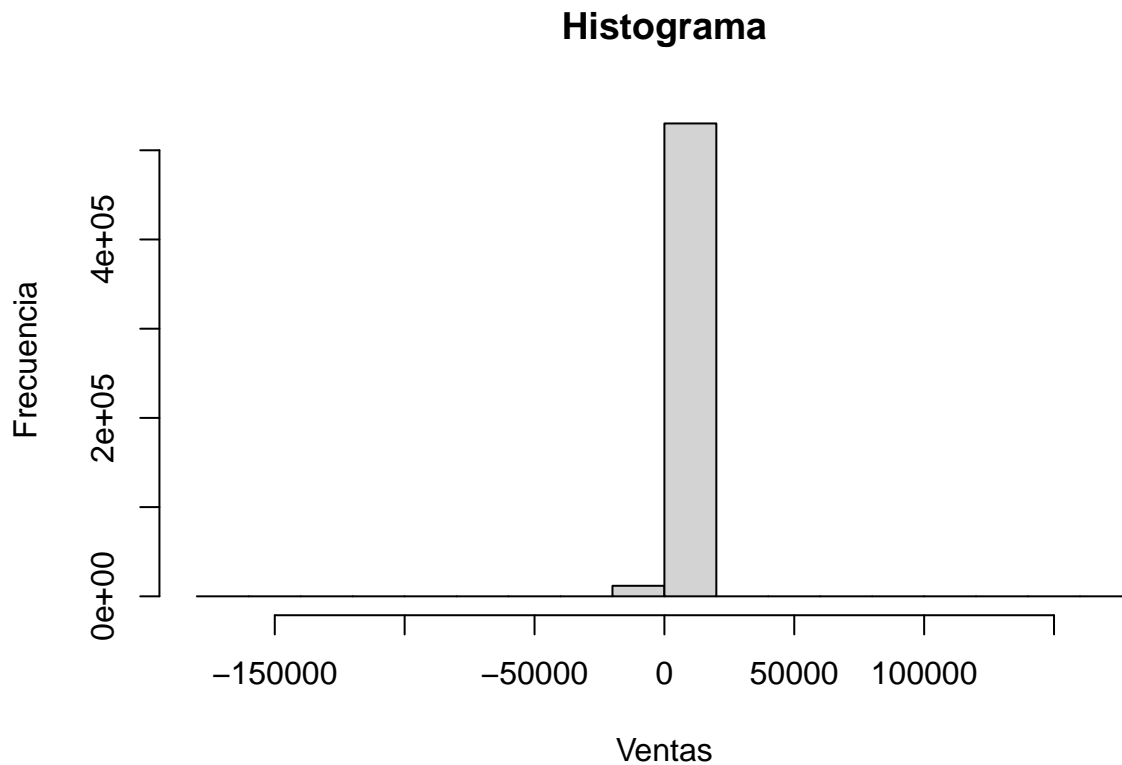


Gráfico Barras

```
barplot(prop.table(table(datosEcommerce$Country)),
       main = "Gráfico de barras")
```

Gráfico de barras

