

R taller - Visualización de datos con ggplot2 - Ejercicios

Anabelle Laurent

10/26/2021

```
library(dplyr)
library(ggplot2)
library(datos)
```

Descripción de los datos “vehiculos”

`vehiculos` es una base de datos que contiene datos de economía de combustible de la Agencia de Protección Medioambiental (EPA) de EE.UU., 1985-2015.

Variable	Descripción
id	Identificador único EPA
fabricante	Fabricante
modelo	Nombre del modelo
anio	año de fabricación.
clase	tipo de auto
transmision	tipo de transmisión
traccion	tipo de tracción
cilindros	número de cilindros
motor	tamaño del motor del automóvil, en litros
combustible	tipo de combustible
autopista	millas por galón de combustible en autopista
ciudad	millas por galón de combustible en ciudad

```
dim(vehiculos) # dimensión
```

```
## [1] 33442 12
```

```
str(vehiculos) # estructura de los datos
```

```
## tibble [33,442 x 12] (S3: tbl_df/tbl/data.frame)
## $ id : num [1:33442] 13309 13310 13311 14038 14039 ...
## $ fabricante : chr [1:33442] "Acura" "Acura" "Acura" "Acura" ...
## $ modelo : chr [1:33442] "2.2CL/3.0CL" "2.2CL/3.0CL" "2.2CL/3.0CL" "2.3CL/3.0CL" ...
## $ anio : num [1:33442] 1997 1997 1997 1998 1998 ...
## $ clase : chr [1:33442] "Automóviles subcompactos" "Automóviles subcompactos" "Automóviles subcompactos" ...
## $ transmision: chr [1:33442] "Automática 4-velocidades" "Manual 5-velocidades" "Automática 4-velocidades" ...
## $ traccion : chr [1:33442] "Delantera" "Delantera" "Delantera" "Delantera" ...
## $ cilindros : num [1:33442] 4 4 6 4 4 6 4 4 6 5 ...
```

```
## $ motor      : num [1:33442] 2.2 2.2 3 2.3 2.3 3 2.3 2.3 3 2.5 ...
## $ combustible: chr [1:33442] "Regular" "Regular" "Regular" "Regular" ...
## $ autopista  : num [1:33442] 26 28 26 27 29 26 27 29 26 23 ...
## $ ciudad     : num [1:33442] 20 22 18 19 21 17 20 21 17 18 ...
```

```
head(vehiculos) # devuelve la primera parte del conjunto de datos
```

```
## # A tibble: 6 x 12
##   id fabricante modelo      anio clase  transmision traccion cilindros motor
##   <dbl> <chr>      <chr>      <dbl> <chr>  <chr>      <chr>      <dbl> <dbl>
## 1 13309 Acura      2.2CL/3.0CL 1997 Autom~ Automática~ Delante~     4  2.2
## 2 13310 Acura      2.2CL/3.0CL 1997 Autom~ Manual 5-v~ Delante~     4  2.2
## 3 13311 Acura      2.2CL/3.0CL 1997 Autom~ Automática~ Delante~     6  3
## 4 14038 Acura      2.3CL/3.0CL 1998 Autom~ Automática~ Delante~     4  2.3
## 5 14039 Acura      2.3CL/3.0CL 1998 Autom~ Manual 5-v~ Delante~     4  2.3
## 6 14040 Acura      2.3CL/3.0CL 1998 Autom~ Automática~ Delante~     6  3
## # ... with 3 more variables: combustible <chr>, autopista <dbl>, ciudad <dbl>
```

```
veh<-vehiculos %>% filter(anio>2010)
```

Ejercicio 1: Cantidad de autos por fabricante Consejo:

usa ‘veh<-vehiculos %>% filter(anio>2010)’ para tener un base de datos más pequeña

Ejercicio 2: Cantidad de autos por fabricante con las barras en orden ascendente Consejo: - usa

veh<-vehiculos %>% filter(anio>2010) - usa dplyr

- usa n() que asigna el número de filas (piense en el número de observaciones)

- geom_bar(stat="identity") La altura de las barras representan valores en los datos (deja los datos como están)

Ejercicio 3: Relacion entre las millas por galón de combustible en autopista y los cilindros

Ejercicio 4: Distribucion des millas por galón de combustible en autopista y los cilindros por el fabricante BMW

Ejercicio 5: distribucion de millas por galón de combustible en ciudad por clase en 1990 y 2010 Consejo:

anio tiene que ser un factor

Descripción de los datos “diamantes”

diamantes es un conjunto de datos que contiene los precios de casi 54.000 diamantes

Variable	Descripción
precio	Precio en dólares americanos (\$326-\$18,823)
quilate	Peso del diamante (0.2-5.01)
corte	Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)
color	Color del diamante, de J (peor) a D (mejor)

Variable	Descripción
claridad	Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))
profundidad	Porcentaje de la profundidad total = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43-79)
tabla	Ancho de la parte superior del diamante con relación a su punto más ancho (43-95)
x	Largo en milímetros
y	Ancho en milímetros
z	Profundidad en milímetros

```
dim(diamantes) # dimensión
```

```
## [1] 53940 10
```

```
str(diamantes) # estructura de los datos
```

```
## tibble [53,940 x 10] (S3: tbl_df/tbl/data.frame)
## $ precio      : int [1:53940] 326 326 327 334 335 336 336 337 337 338 ...
## $ quilate     : num [1:53940] 0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...
## $ corte       : Ord.factor w/ 5 levels "Regular"<"Bueno"<...: 5 4 2 4 2 3 3 3 1 3 ...
## $ color       : Ord.factor w/ 7 levels "D"<"E"<"F"<"G"<...: 2 2 2 6 7 7 6 5 2 5 ...
## $ claridad    : Ord.factor w/ 8 levels "I1"<"SI2"<"SI1"<...: 2 3 5 4 2 6 7 3 4 5 ...
## $ profundidad: num [1:53940] 61.5 59.8 56.9 62.4 63.3 62.8 62.3 61.9 65.1 59.4 ...
## $ tabla      : num [1:53940] 55 61 65 58 58 57 57 55 61 61 ...
## $ x          : num [1:53940] 3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...
## $ y          : num [1:53940] 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...
## $ z          : num [1:53940] 2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...
```

```
head(diamantes) # devuelve la primera parte del conjunto de datos
```

```
## # A tibble: 6 x 10
##   precio quilate corte      color claridad profundidad tabla      x      y      z
##   <int>   <dbl> <ord>    <ord> <ord>          <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    326    0.23 Ideal     E     SI2            61.5    55  3.95  3.98  2.43
## 2    326    0.21 Premium  E     SI1            59.8    61  3.89  3.84  2.31
## 3    327    0.23 Bueno    E     VS1            56.9    65  4.05  4.07  2.31
## 4    334    0.29 Premium  I     VS2            62.4    58  4.2   4.23  2.63
## 5    335    0.31 Bueno    J     SI2            63.3    58  4.34  4.35  2.75
## 6    336    0.24 Muy bueno J     VVS2           62.8    57  3.94  3.96  2.48
```

Ejercicio 6: relacion entre quilate y precio + pone una tercera dimension de tu preferencia

Ejercicio 7: agregar una regresión

Ejercicio 8: Haz un gráfico por diferentes grupos de corte