01-visualizar-datos-presentacion

Renan Escalante-Chong

# Visualización de datos

## Introducción

Este tutorial te mostrará como visualizar datos usando la librería de R ggplot2. Este paquete fue creado for Hadley Wickman. La mayoría de los ejemplos de este tutorial están basados en el libro [R for data science](http://r4ds.had.co.nz)

## Data frames

Un data frame es una estructura de datos en R Un data frame es una tabla donde las columnas (columns) son variables y las filas (rows) son instancias, casos.

Nota: Este tipo de datos son datos limpios **tidy data** donde las columnas son siempre variables y las filas son instancias.

|  |  |  |  |
| --- | --- | --- | --- |
| estudiantes | edad | fac | estado |
| Juan | 22 | UNAM | Tabasco |
| Mildred | 18 | UV | Campeche |
| Sofía | 21 | preparatoria | Puebla |

## Prerequisites

To access the datasets, help pages, and functions that we will use in this chapter, load ggplot2 using the library() function. We'll also load tibble, which you'll learn about later. It improves the default printing of datasets.

library(ggplot2)

If you run this code and get the error message "there is no package called ‘ggplot2’", you'll need to first install it, then run library() once again.

install.packages("ggplot2")  
library(ggplot2)

## A graphing template

Let's use our first graph to answer a question: Do cars with big engines use more fuel than cars with small engines?

You can test your answer with the mpg dataset in ggplot2, or ggplot2::mpg:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| manufacturer | model | displ | year | cyl | trans | drv | cty | hwy | fl | class |
| audi | a4 | 1.8 | 1999 | 4 | auto(l5) | f | 18 | 29 | p | compact |
| audi | a4 | 1.8 | 1999 | 4 | manual(m5) | f | 21 | 29 | p | compact |
| audi | a4 | 2.0 | 2008 | 4 | manual(m6) | f | 20 | 31 | p | compact |
| audi | a4 | 2.0 | 2008 | 4 | auto(av) | f | 21 | 30 | p | compact |
| audi | a4 | 2.8 | 1999 | 6 | auto(l5) | f | 16 | 26 | p | compact |

## Gráficos

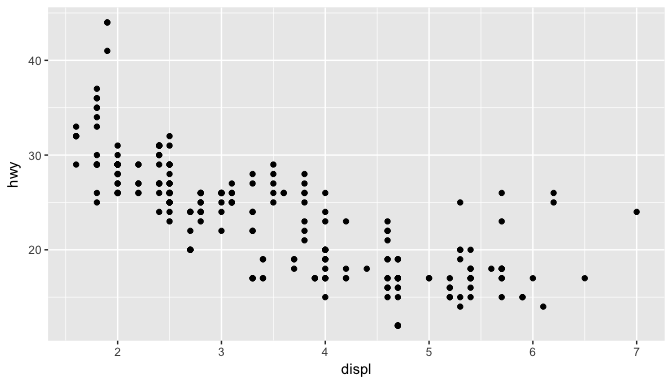
Muchas veces podemos usar números para tratar de entender un set de datos usando representaciones gráficas

Los autos con motores grandes consumen más gasolina?

## A graphing template (2)

Esta es una forma de describir un gráfico

ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy))



## Gramática de gráficos

Aquí vamos a aprender la gramática de como construir un gráfico

1. Set de datos (data)
2. Especificar que variables y que instancias del set de datos queremos (mapping)
3. Luego especificamos que tipo de plot queremos (geom\_point)
4. Por último agregamos propiedades del gráfico

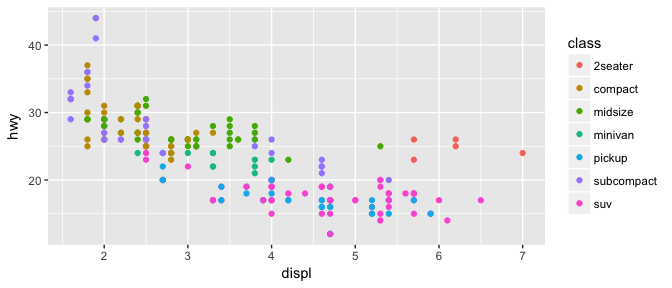
ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy))

## Exercises

1. Run ggplot(data = mpg) what do you see?
2. What does the drv variable describe? Read the help for ?mpg to find out.
3. Make a scatterplot of hwy vs cyl.
4. What happens if you make a scatterplot of class vs drv. Why is the plot not useful?

## Agregar "aesthetics" podemos agregar propiedades al gráfico

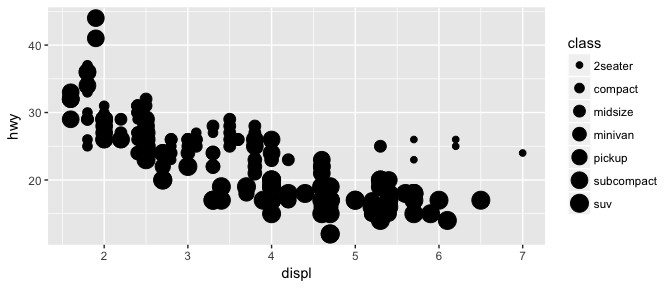
ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy,   
 color = class))



## Gráfico usando tamaño (mal ejemplo)

ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy,   
 size = class))

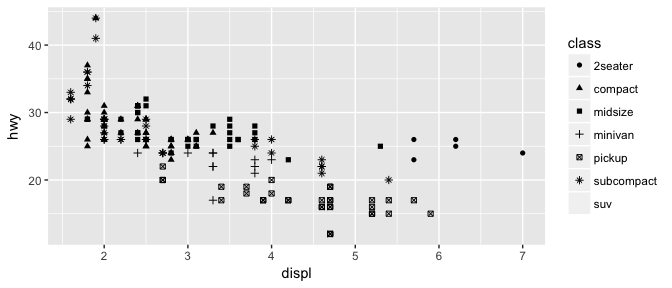
## Warning: Using size for a discrete variable is not advised.



## Gráfico usando formas

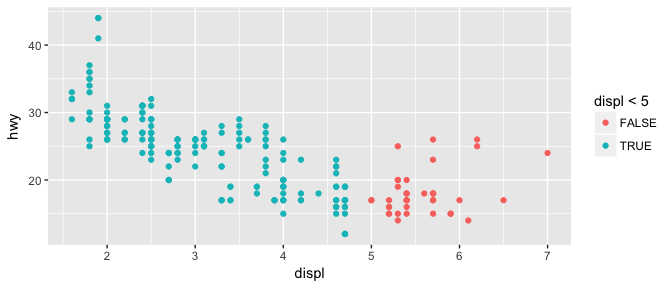
ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy, shape = class))

## Gráfico usando formas



## Selección condicional (color)

ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy, color = displ<5))



## Obtener ayuda

Escriban ?geom\_point

datos<-read.table("efectos\_medicina\_ALS.tsv", header = FALSE)  
datos\_ALS<-read.table("datos\_ALS.tsv", header = FALSE)