

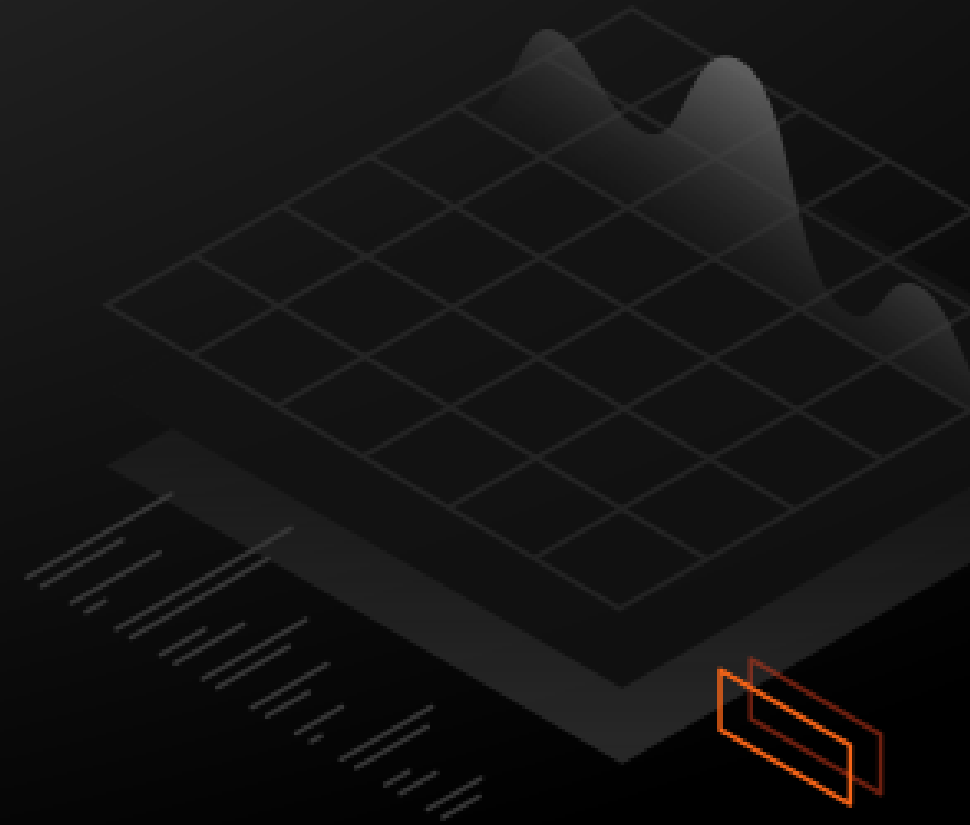
COMPUTACIÓN ESTADÍSTICA

EPG3308

02 AED Viz **ggplot2**

JOSHUA KUNST
@JBKUNST

2023-03-23

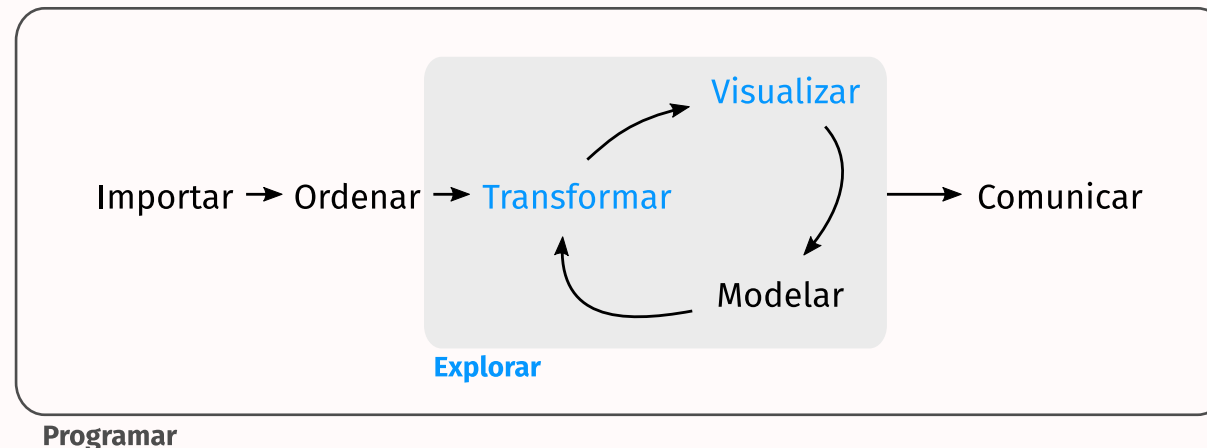


ANÁLISIS EXPLORATORIO DE DATOS

El **Análisis Exploratorio de Datos** se refiere a un ciclo iterativo en el que:

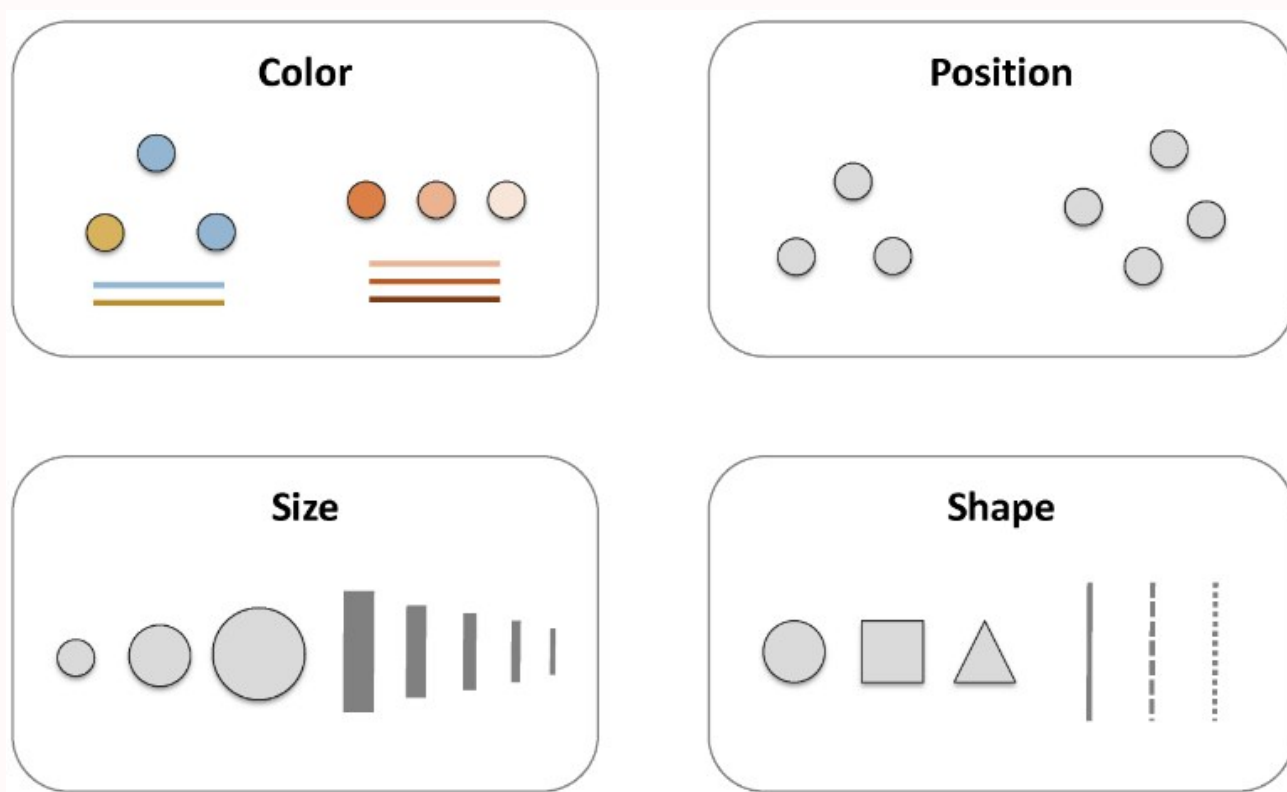
- Generas preguntas acerca de tus datos.
- Buscas respuestas visualizando, transformando y modelando tus datos.
- Usas lo que has aprendido para refinar tus preguntas y/o generar nuevas interrogantes.

Lo anterior es un *copy & paste* desde R4DS.



VISUALIZACIÓN

Definiremos la visualización, en el contexto de análisis de datos, como una manera de representar/codificar información, y por tanto una herramienta para analizar datos.



ANTES DE SEGUIR

- Una **variable** es una cantidad, que puede ser medida: estatura, magnitud de un sismo, velocidad.
- Un **valor** es un estado de la variable cuando se mide: 1.20 metros, 8° Richter.
- Una **observación** o caso es un conjunto de mediciones -no de la misma variable necesariamente- pero si en un mismo instante y a un mismo objeto.

country	year	cases	population
Afghanistan	1999	2666	19995360
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	210258	1272915272
China	2000	210766	1280425583

variables

country	year	cases	population
Afghanistan	1999	2666	19995360
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	210258	1272915272
China	2000	210766	1280425583

observations

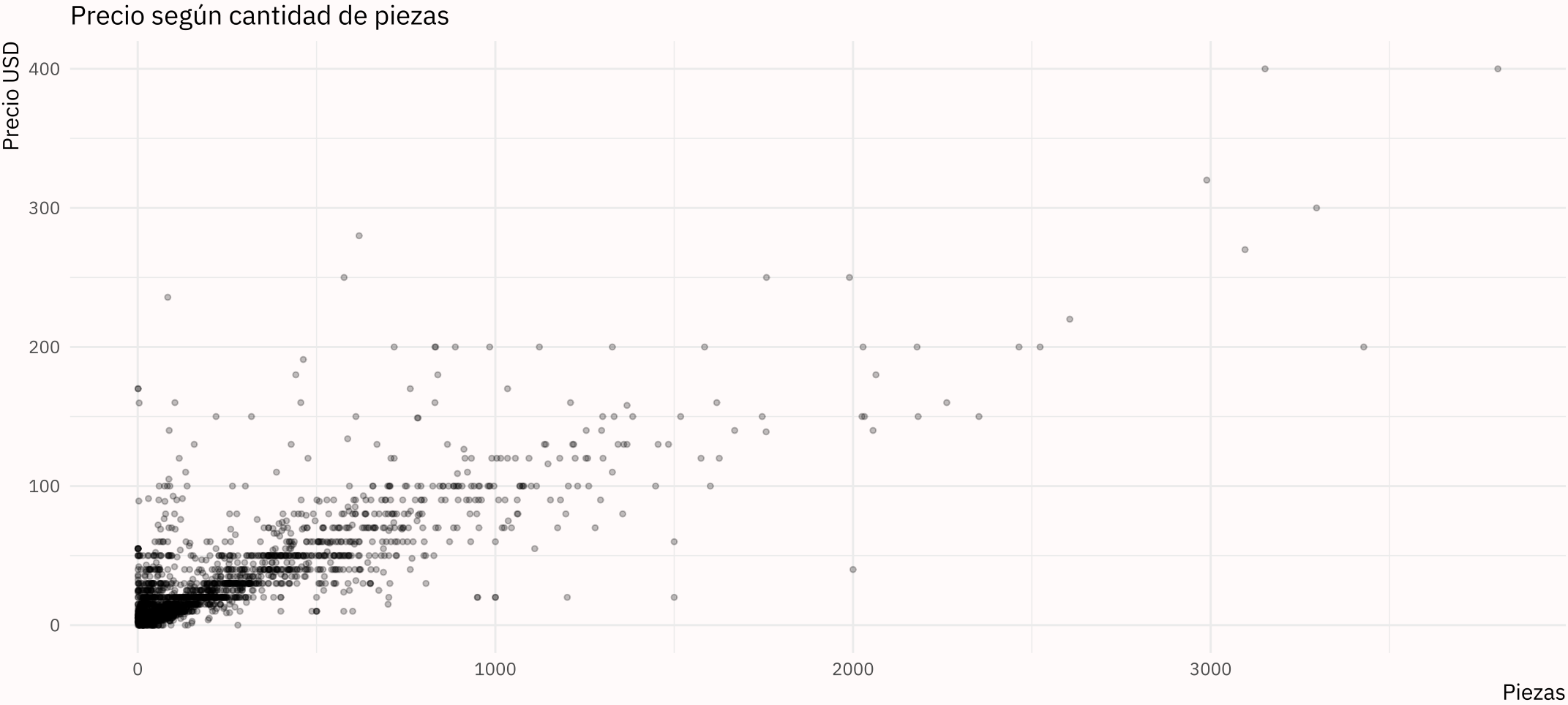
country	year	cases	population
Afghanistan	99	2666	19995360
Afghanistan	00	2666	20595360
Brazil	99	37737	172006362
Brazil	00	80488	174504898
China	99	210258	1272915272
China	00	210766	1280425583

values

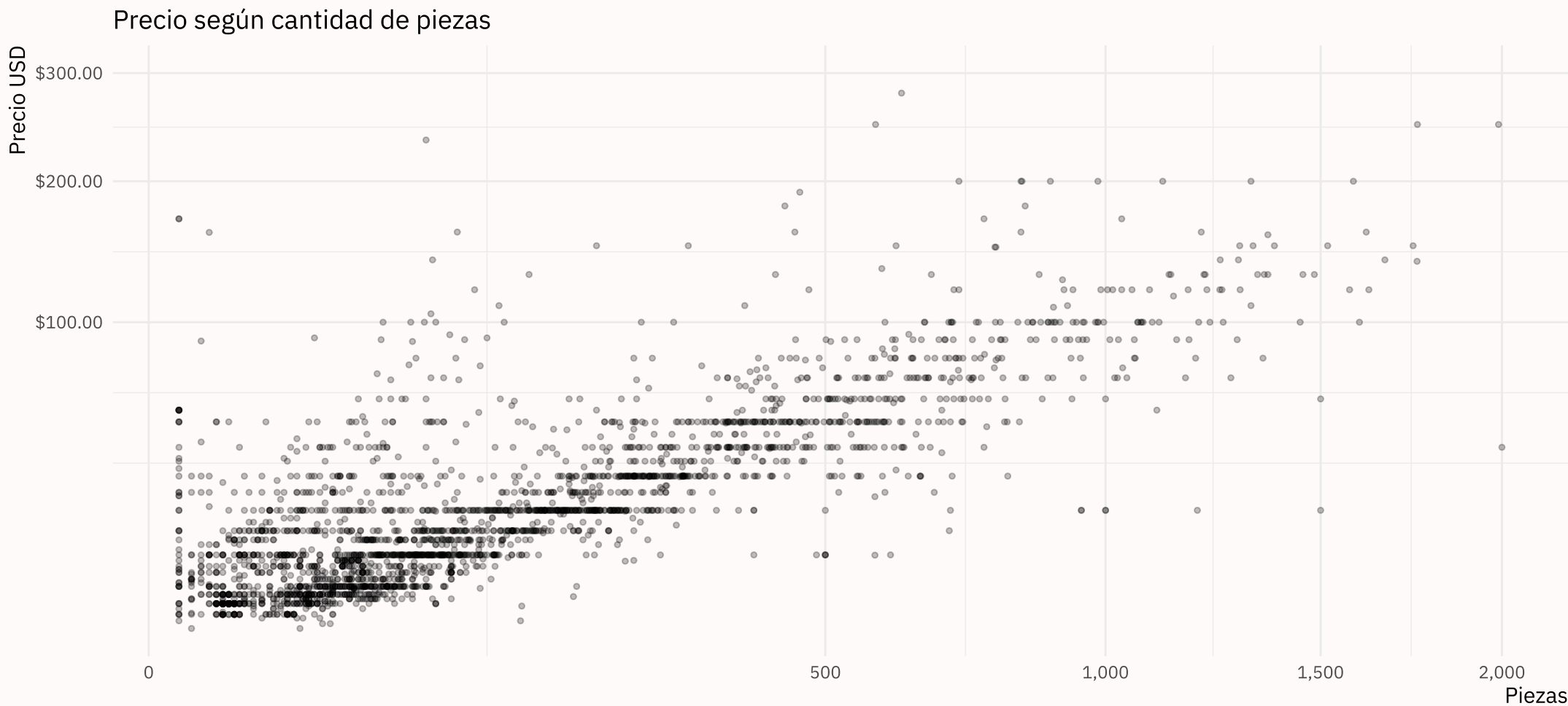
UN PRIMER EJEMPLO: SET DE LEGOS!

AutoSave Off legosets.csv Search Joshua Benjamin Kun													
File Home Insert Page Layout Formulas Data Review View Help													
16176													
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Item_Number	Name	Year	Theme	Subtheme	Pieces	Minifigures	GBP_MSRP	USD_MSRP	CAD_MSRP	EUR_MSRP	Packaging	Availability
6157	376	Low loader with Excava	1971	LEGOLAND	Large Vehicle	92	NA	1.75	NA	NA	NA	Box	Retail
6158	377	Crane and Float Truck	1971	LEGOLAND	Large Vehicle	91	NA	1.85	NA	NA	NA	Box	Retail
6159	380	Village Set	1971	LEGOLAND	Building	485	NA	5.9	NA	NA	NA	Box	Retail
6160	604	Excavator	1971	LEGOLAND	Vehicle	22	NA	0.33	NA	NA	NA	Box	Retail
6161	605	Taxi	1971	LEGOLAND	Vehicle	19	NA	0.33	NA	NA	NA	Box	Retail
6162	608	Kiosk	1971	LEGOLAND		33	NA	0.33	NA	NA	NA	Box	Retail
6163	640	Fire Truck	1971	LEGOLAND	Vehicle	45	NA	0.55	NA	NA	NA	Box	Retail
6164	642	Double Excavator	1971	LEGOLAND	Large Vehicle	33	NA	0.55	NA	NA	NA	Box	Retail
6165	643	Mobile Crane	1971	LEGOLAND	Vehicle	35	NA	0.55	NA	NA	NA	Box	Retail
6166	644	Double Tanker	1971	LEGOLAND		61	NA	0.55	NA	NA	NA	Box	Retail
6167	645	Milk Truck with Trailer	1971	LEGOLAND	Vehicle	57	NA	0.55	NA	NA	NA	Box	Retail
6168	646	Mobile Site Office	1971	LEGOLAND	Vehicle	51	NA	0.55	NA	NA	NA	Box	Retail
6169	647	Lorry With Girders	1971	LEGOLAND	Vehicle	52	NA	0.55	NA	NA	NA	Box	Retail
6170	648	Service Station	1971	LEGOLAND	Building	55	NA	0.55	NA	NA	NA	Box	Retail
6171	795	2 Large Baseplates, Red	1971	Basic	Supplementarie	2	NA	NA	2.1	NA	NA	Not specifiec	Not specified
6172	796	2 Large Baseplates, Gre	1971	Basic	Supplementarie	2	NA	NA	2.1	NA	NA	Not specifiec	Not specified
6173	798	2 Medium Baseplates, C	1971	Basic	Supplementarie	2	NA	NA	3.4	NA	NA	Not specifiec	Not specified
6174													
6175													
6176													
6177													

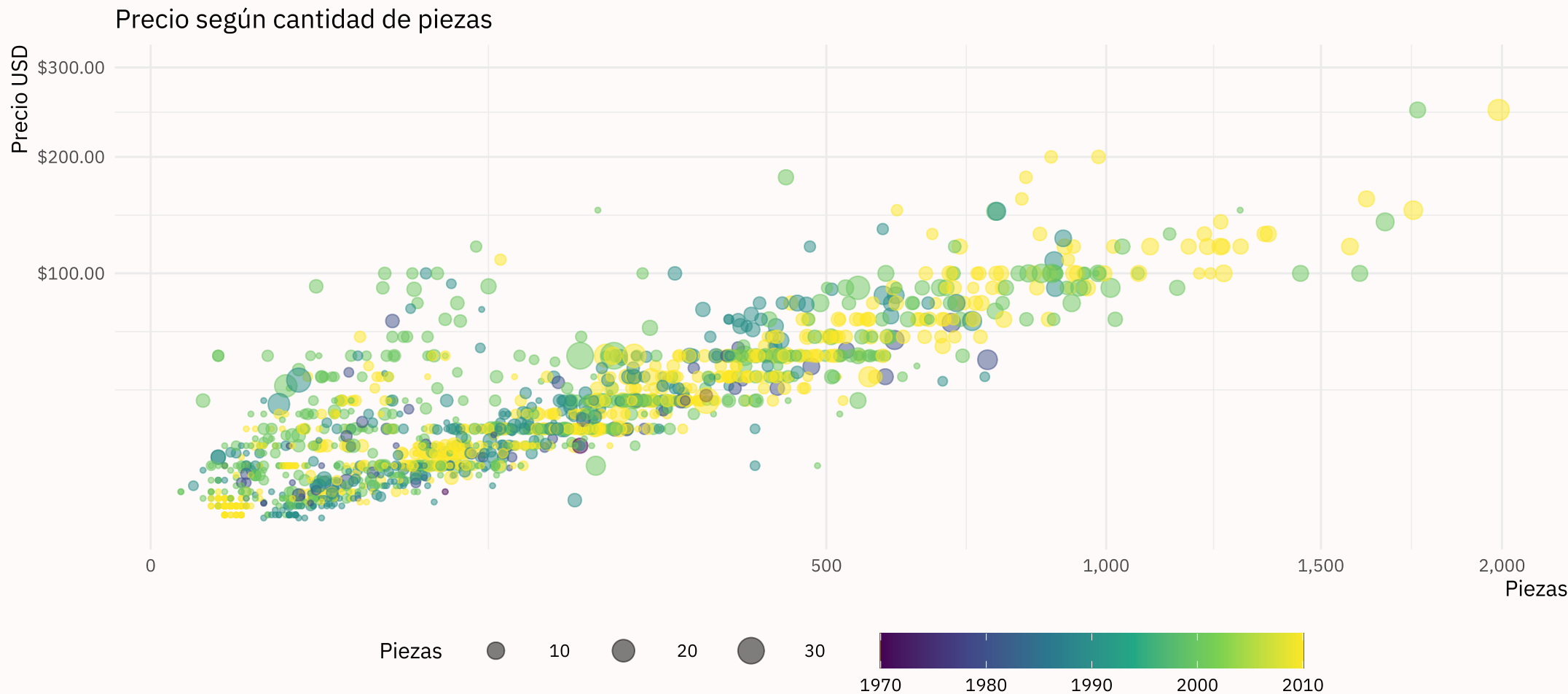
EXPLORAR (V1)



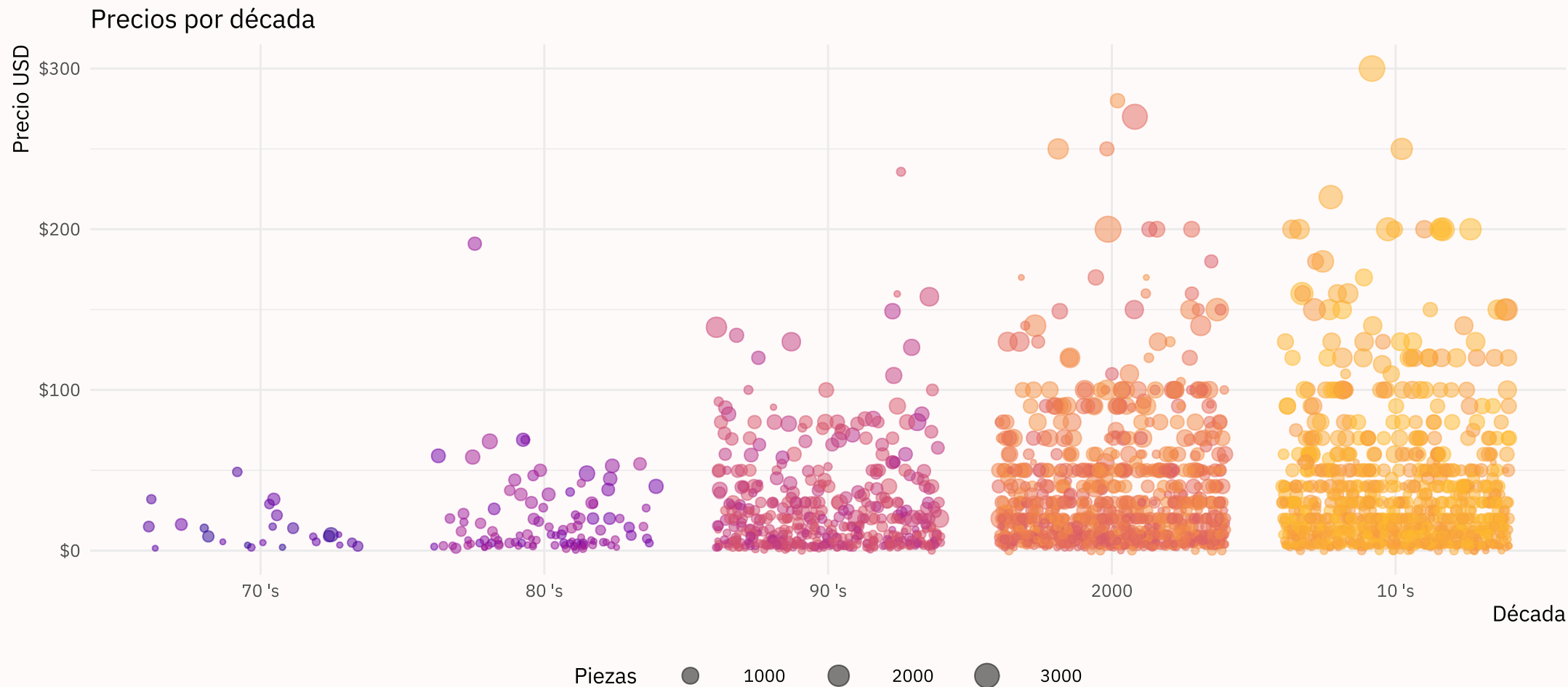
EXPLORAR (V2)



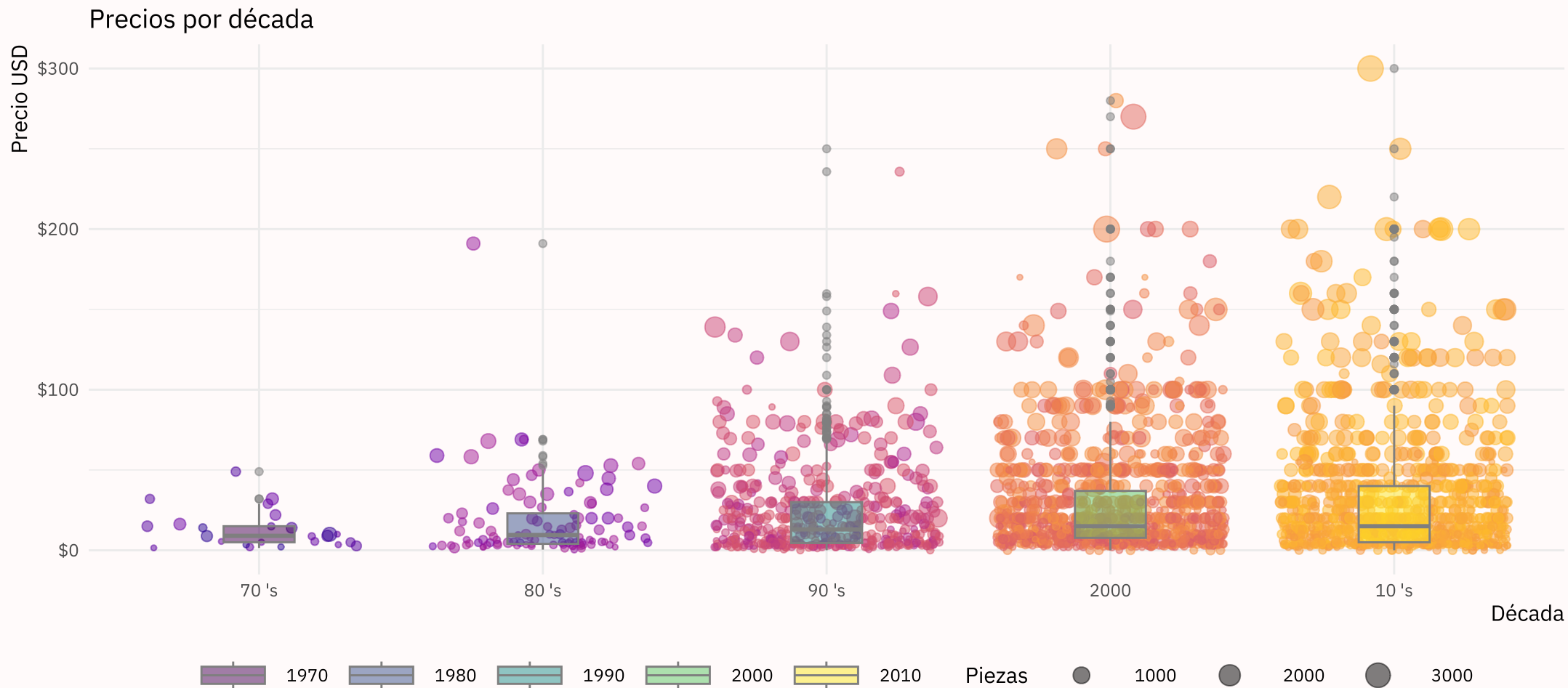
CODIFICAMOS MAYOR INFORMACIÓN. BUENO/MALO?



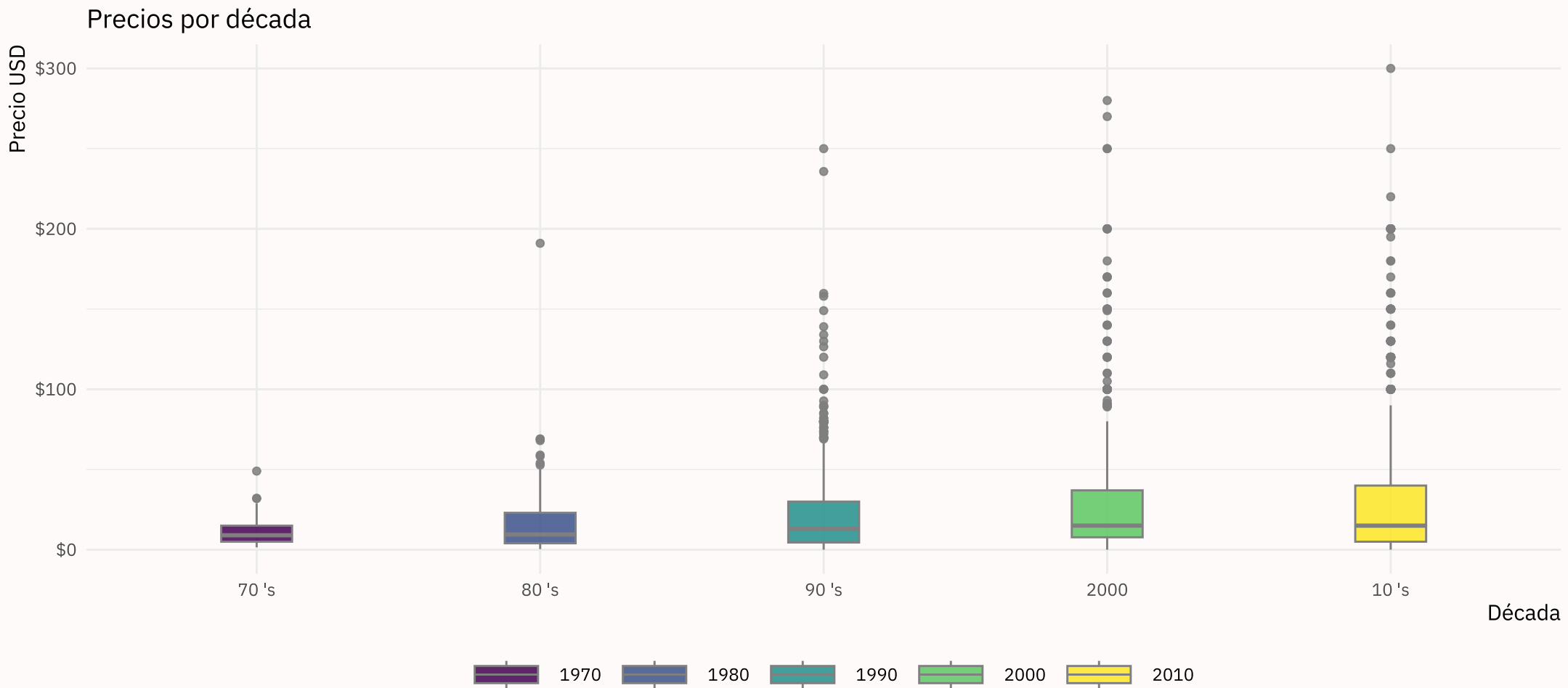
CADA VISUALIZACIÓN CUENTA UNA HISTORIA



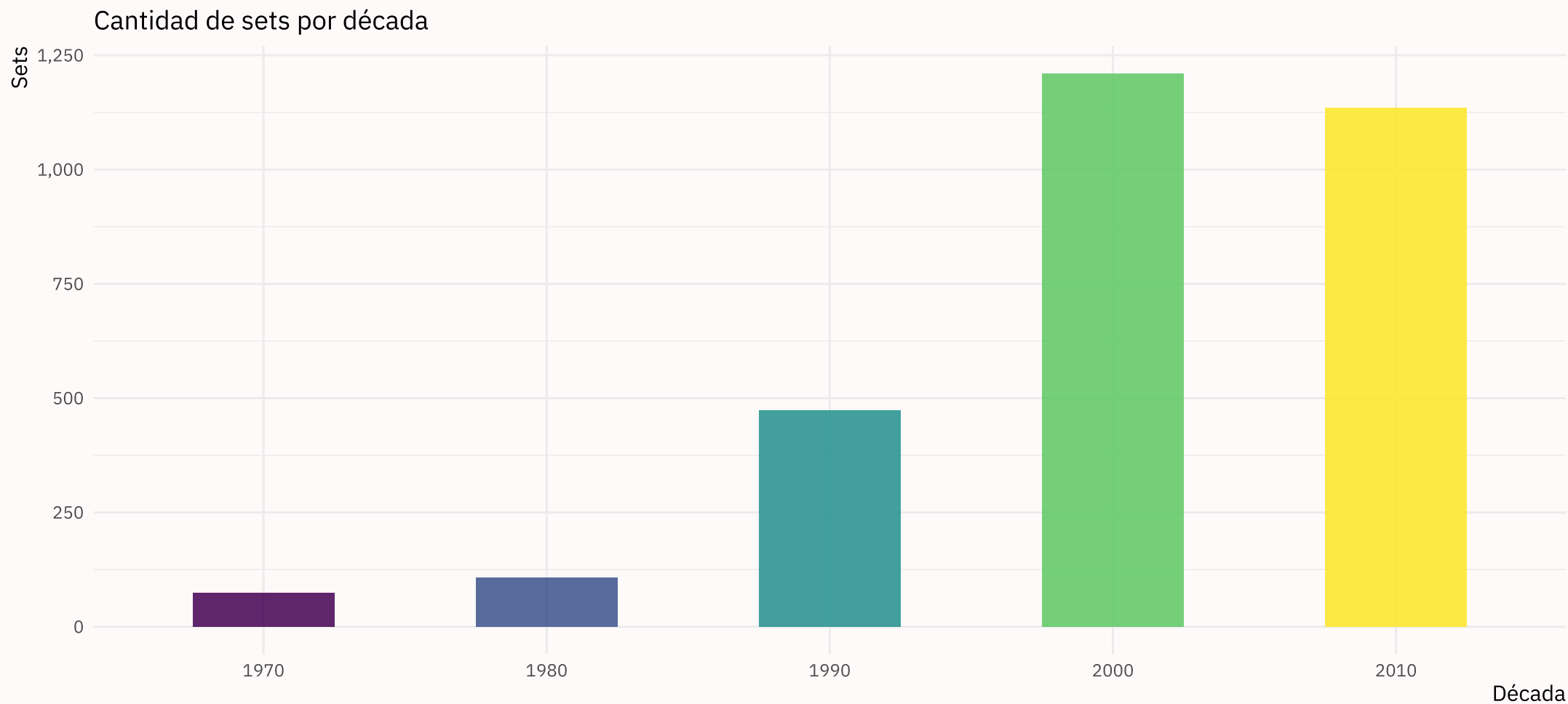
EXPLORAR TENDENCIA ANTES QUE DETALLES...



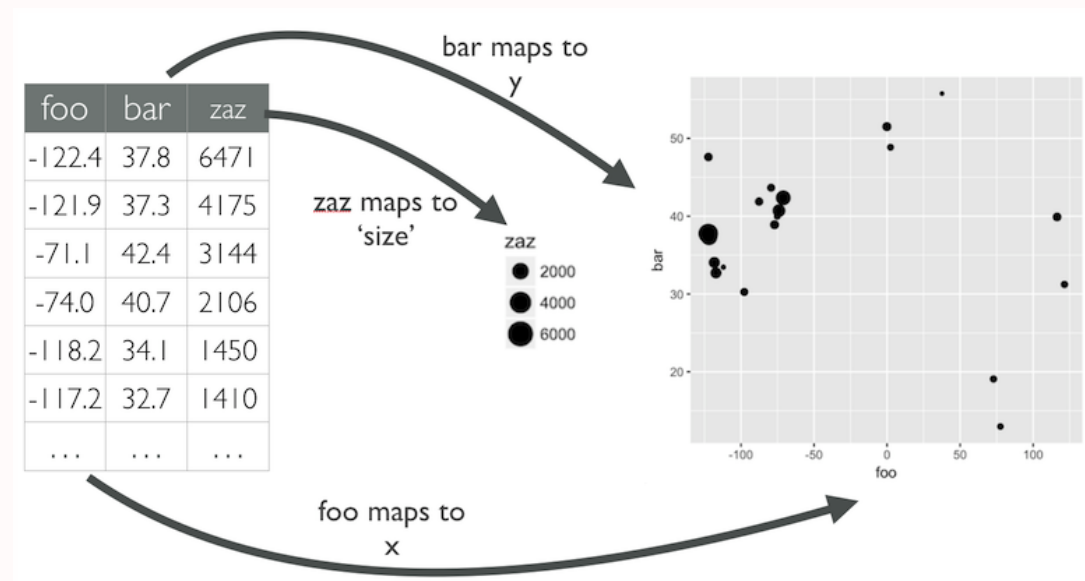
EXPLORAR TENDENCIA ANTES QUE DETALLES... (V2)



O SIMPLEMENTE CONTAR UNA HISTORIA MÁS SENCILLA



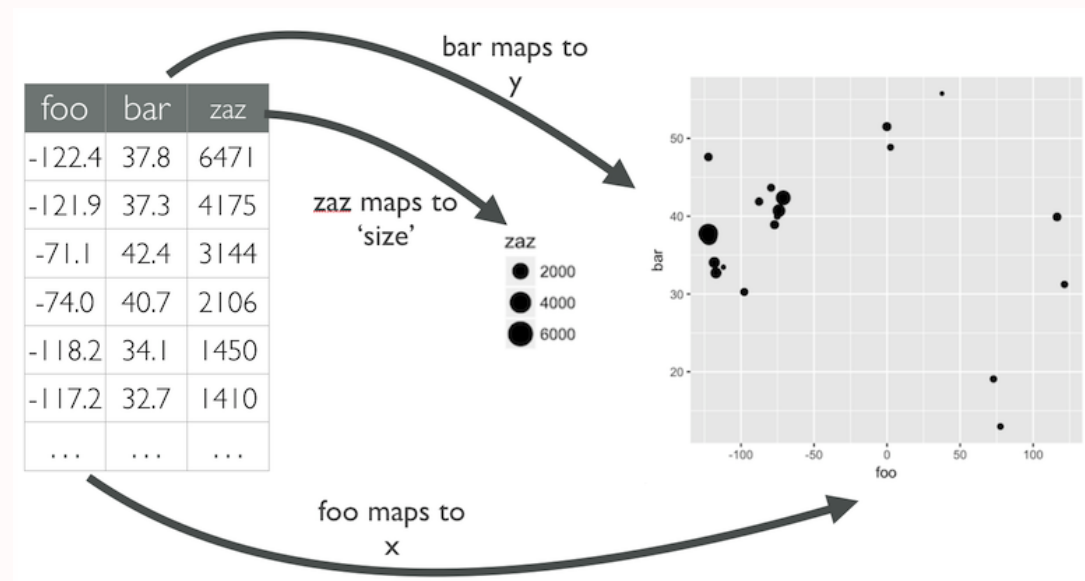
ggplot2



De forma **general**:

```
ggplot(data = <DATOS>) +  
  <GEOM_FUNCIÓN>(mapping = aes(<MAPEOS>))
```

ggplot2



De forma **particular**:

```
ggplot(data = datos) +  
  geom_point(mapping = aes(x = foo, y = bar, size = zaz))
```

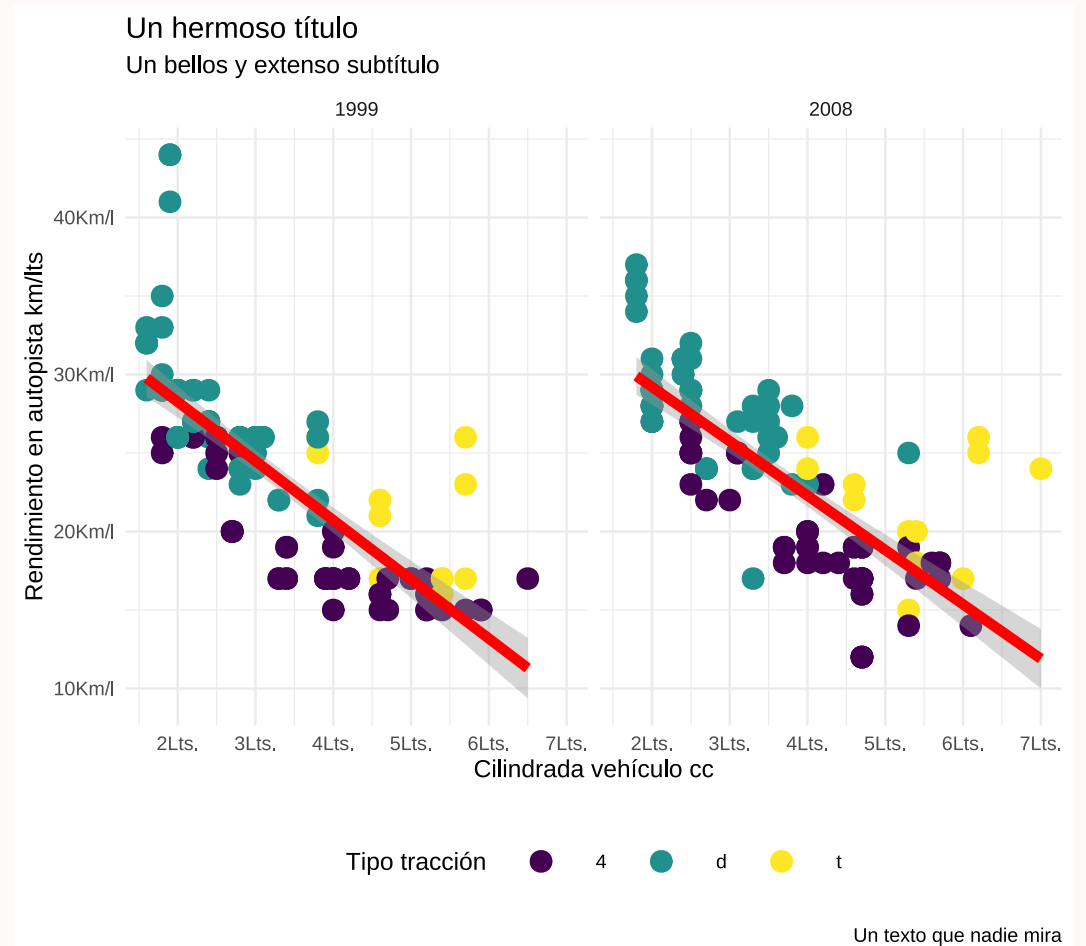
ggplot2

`ggplot2` funciona bajo un marco de layers o capas. Estas las iremos agregando de abajo hacia arriba.



EJEMPLO LAYERS

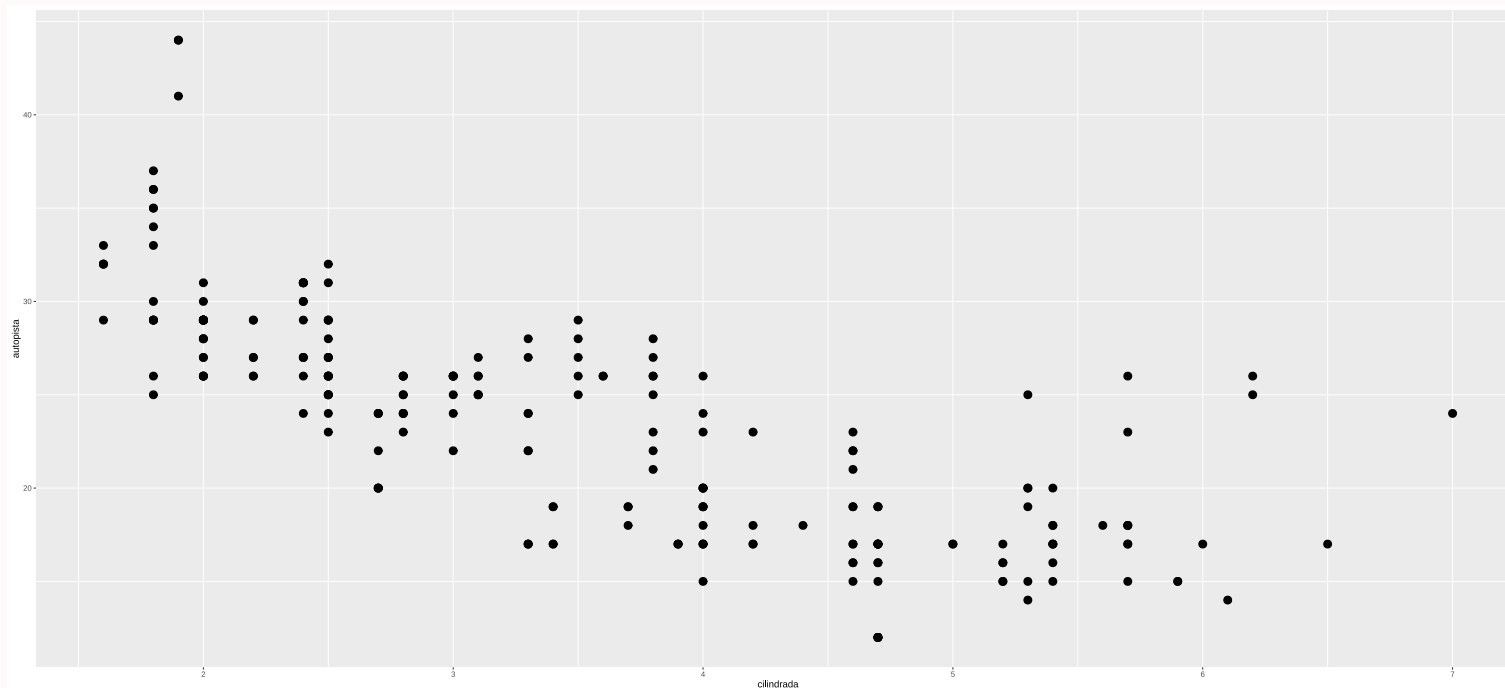
```
ggplot(  
  millas, # data  
  aes(cilindrada, autopista) # aesthetics  
) +  
  geom_point(aes(color = traccion)) + # geometr.  
  facet_wrap(vars(anio)) + # facets  
  stat_smooth(color = "red", method = "lm") + #  
  scale_x_continuous(labels = comma_format(suffi  
  scale_y_continuous(labels = comma_format(suffi  
  theme_minimal(base_size = 5) + # theme  
  theme(legend.position = "bottom") + # theme  
# otros  
  scale_color_viridis_d() +  
  labs(  
    title = "Un hermoso título",  
    subtitle = "Un bellos y extenso subtítulo",  
    caption = "Un texto que nadie mira",  
    x = "Cilindrada vehículo cc",  
    y = "Rendimiento en autopista km/lts",  
    color = "Tipo tracción"  
  )
```



HAGAMOS UN *COPY & PASTE*

```
library(datos)
library(ggplot2)

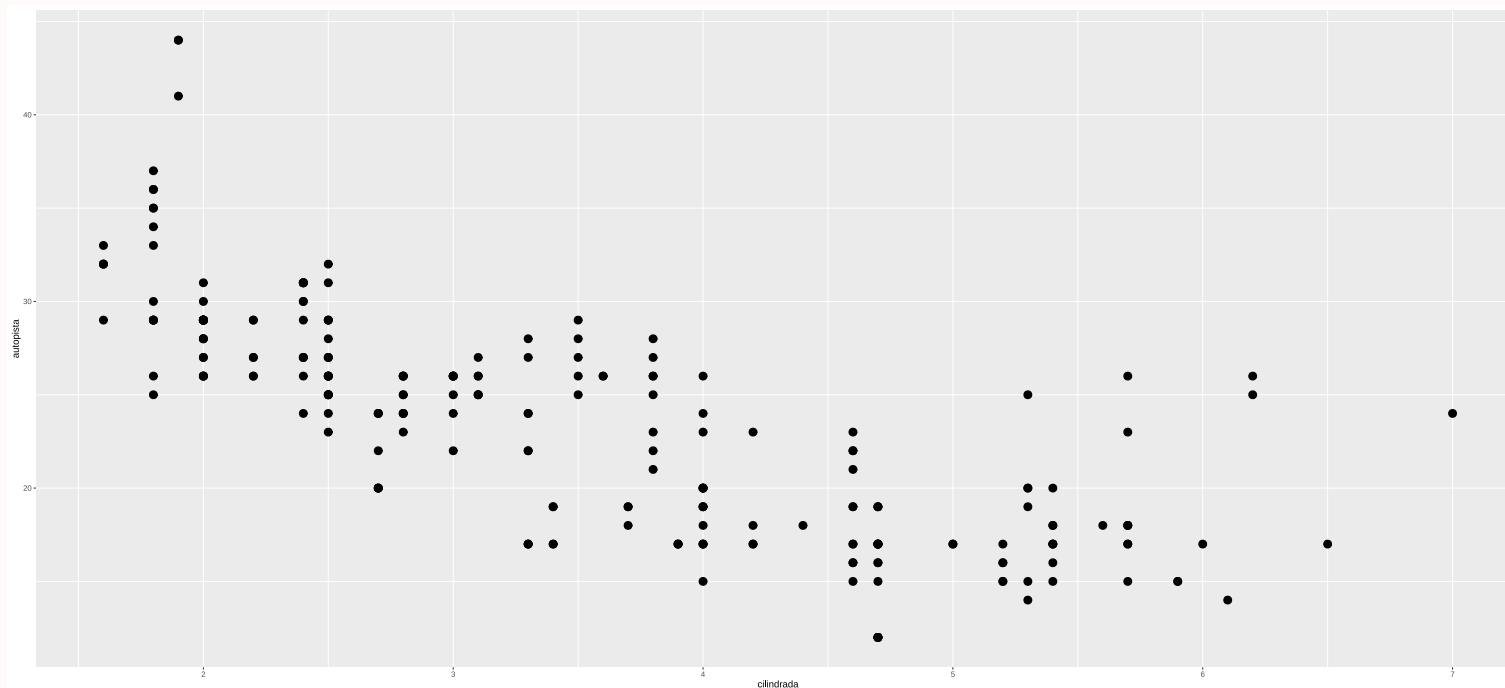
ggplot(data = millas) +
  geom_point(mapping = aes(x = cilindrada, y = autopista))
```



HAGAMOS UN *COPY & PASTE* (V2)

```
library(datos)
library(ggplot2)

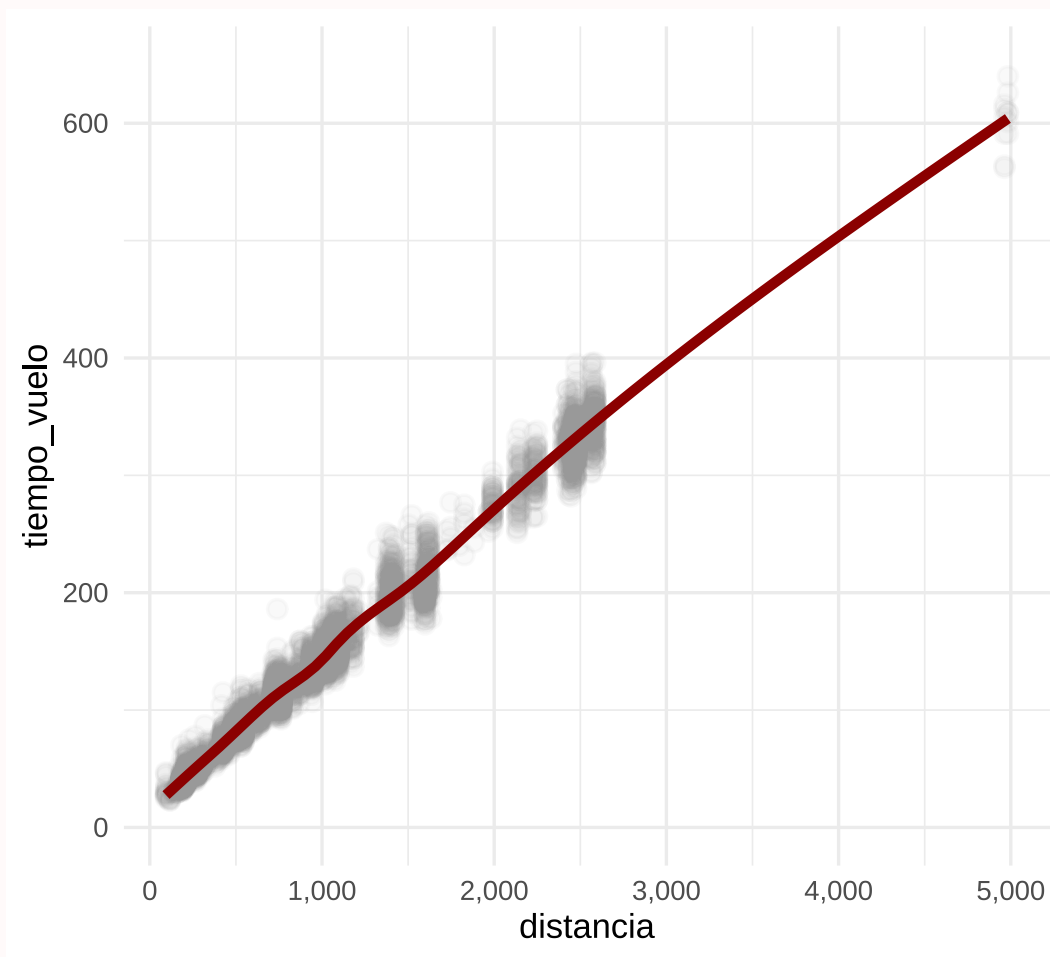
ggplot(millas) +
  geom_point(aes(cilindrada, autopista))
```



SIRVE OBTENER UNA MUESTRA?

```
vuelos2 <- sample_n(vuelos, 5000)

ggplot(vuelos2, aes(distancia, tiempo_vuelo)) +
  geom_point(alpha = 0.05, color = "gray60") +
  geom_smooth(se = FALSE, color = "darkred") +
  scale_x_continuous(labels = scales::comma) +
  scale_y_continuous(limits = c(0, 650))
```



ggplot2 theme elements reference

Set minimal as the baseline theme:

```
theme_minimal() +  
theme(theme.element = element_type())
```

Use `element_blank()` to remove an element

Axis titles, text, ticks, and lines can be specified per axis using theme inheritance by putting `.x/.y` at the end of the theme element.

```
axis.line.y = element_line()
```

```
axis.title.y = element_text()
```

```
panel.grid.major = element_line()
```

```
panel.grid.minor = element_line()
```

```
axis.text.y
```

```
axis.text = element_text()
```

```
axis.text.x
```

`text = element_text()` ← modifications will be applied to all text elements

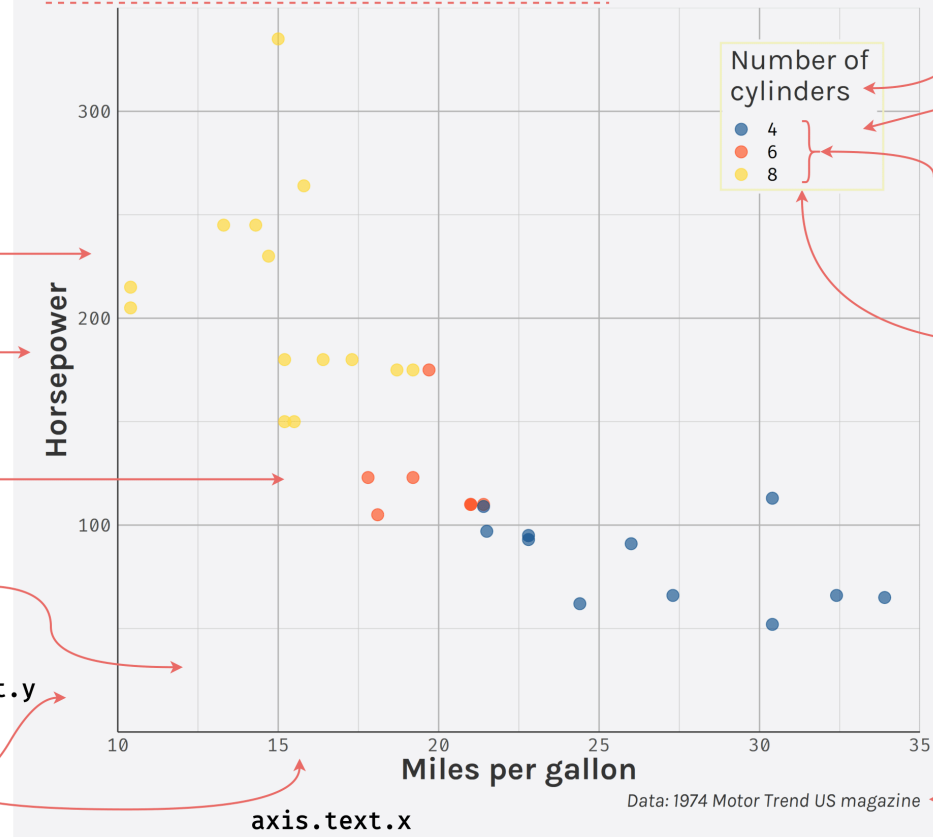
```
plot.title.position = "plot"  
plot.caption.position = "plot"
```

"plot" means that they will be aligned to the entire plot (instead of the panel)

```
plot.title = element_text()  
plot.subtitle = element_text()
```

```
plot.margin = margin(25, 25, 25, 25)
```

**Miles per Gallon & Horsepower
of 32 Automobiles (1973-74 models)**



```
legend.title = element_text()
```

```
legend.background = element_rect()
```

```
legend.text = element_text()
```

```
legend.position = c(.85, .85) / "none" /  
"left" / "right" /  
"bottom" / "top"
```

```
plot.background = element_rect()
```

```
plot.caption = element_text()
```

isabella-b

Full list of elements at ggplot2.tidyverse.org/reference/theme

EJERCICIO 2: REPLIQUE LOS GRÁFICOS DE LEGOS

Para cargar los datos utilice:

```
library(tidyverse)

set.seed(123)

url_datos <- "https://raw.githubusercontent.com/seankross/lego/master/data-tidy/legosets.csv"

legos <- read_csv(url_datos) |>
  mutate(year2 = floor(Year/10)*10) |>
  sample_n(3000)
```

Antes de replicar los gráficos de legos -si bien utiliza funciones que no hemos visto-, interprete el código que lee los datos.

MÁS INFORMACIÓN SOBRE VISUALIZACIÓN

- [Visualización de datos](#) en R4DS.
- A *simple* intro to `ggplot2`, [link](#).
- Para motivarse: *ggplot Wizardry Hands-On*, [link](#). [Cedric Sherer](#)
- Visualización desde el análisis de datos. [Slides](#).
- Más *flipbooks* con el paquete [flipbookr](#).
- Documentación `ggplot2` <https://ggplot2.tidyverse.org/index.html>.
- Explorando histogramas, [link](#).