

Visualización de datos con ggplot2

BSc. Fabrizio Vasquez
Universidad Nacional de Ingeniería
XXI Congreso Nacional de Estudiantes de Estadística – Lima 2019



Marcha de Napoleón

por Charles Minard



Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869

Les nombres d'hommes présents sont représentés par les longueurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Ségur, de Fozzard, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps de Lénie Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mielno et qui rejoignent Orel et Witebsk, avaient toujours marché avec l'armée.

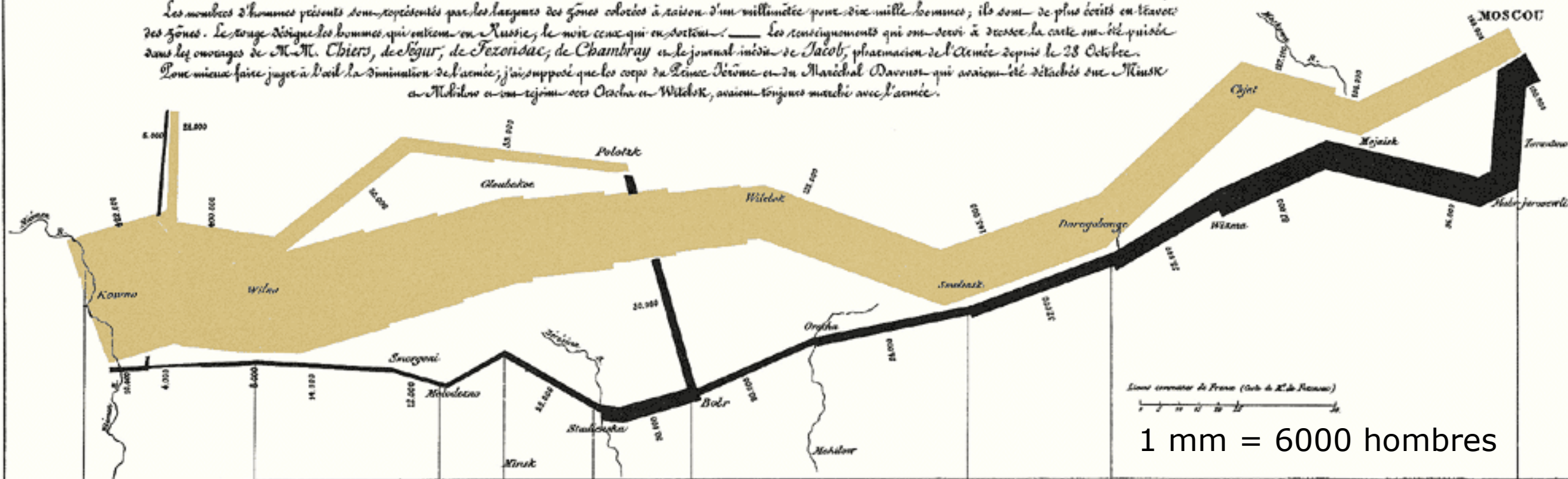
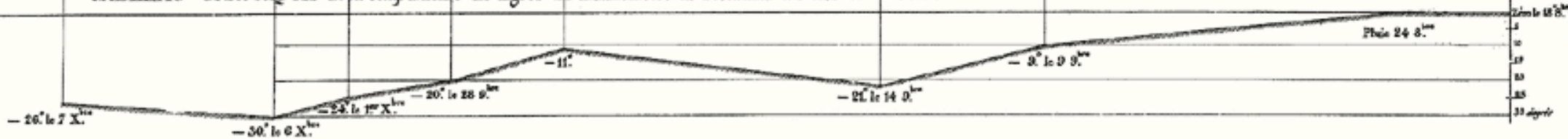
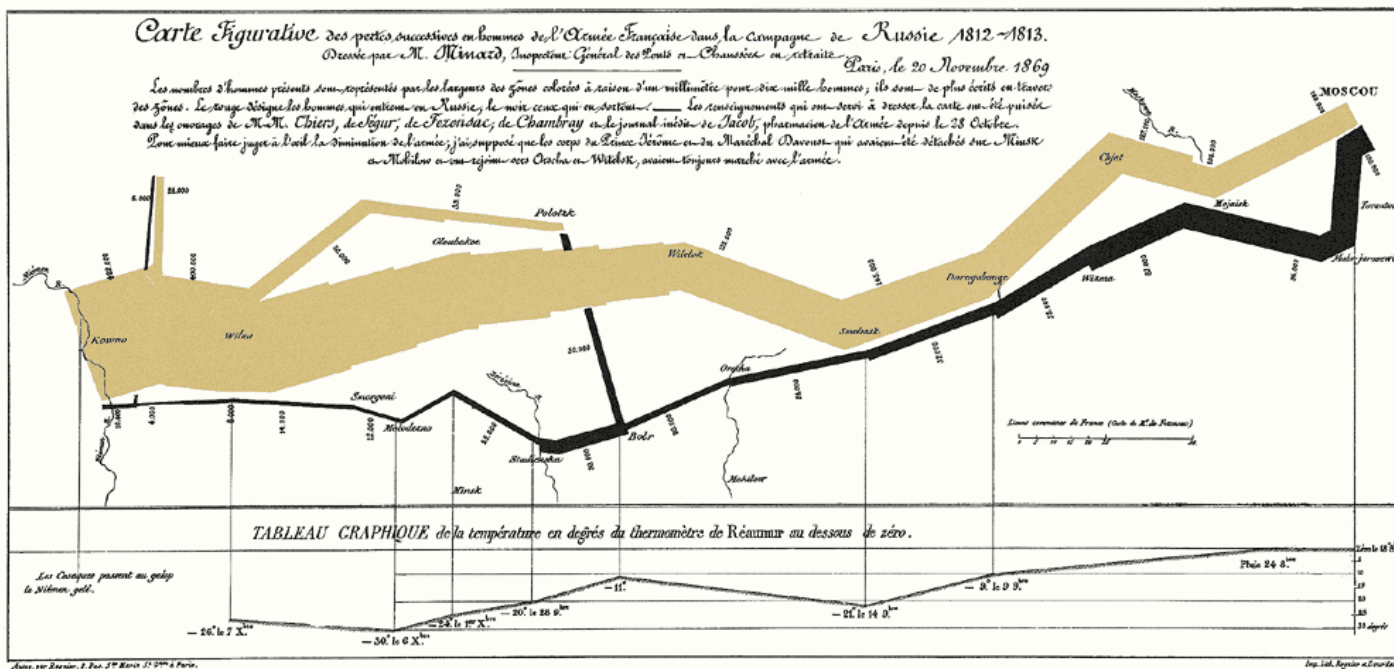


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Marcha de Napoleón

por Charles Minard



1 Espacio Geográfico

2 Número de soldados

3 Sentido de marcha del ejército

4 Tiempo

5 Temperatura a la vuelta

1 mm = 6000 hombres

SPOILER
ALERT

Spoilers...

- Necesitamos saber “algo” de R.

```
algo <- c(“tipos y estructuras de datos”,  
          “manipulación básica de objetos”,  
          “cositas adicionales que Fabrizio se olvidó”  
        )  
  
if(nivel_de_r==“Nada mano :v”){  
  print(“Atento a las explicaciones”)} else {  
  print(“¡Disfrute del minicurso!”)  
}
```

Referencias



- [1] ggplot2, Elegant Graphics for Data Analysis, Hadley Wickham
- [2] R Graphics Cookbook, Winston Chang

Asegurarnos

```
#Si no tenemos instalado la librería ggplot2  
install.packages("ggplot2")
```

```
#Cargar la librería a R  
library(ggplot2)
```

```
#Cargar los datos a utilizar  
load("data_curso_ggplot2.Rdata")
```


¿Qué es ggplot2?

- Paquete de visualización de datos en R.
- Creado por Hadley Wickham en 2005, como una implementación de la gramática de los gráficos de Leland Wilkinson.
- Es uno de los paquetes más usados dentro del software R.



Hadley Wickham



Leland Wilkinson

mapping = trazado

Actividad o proceso de crear una imagen o diagrama que representa algo.

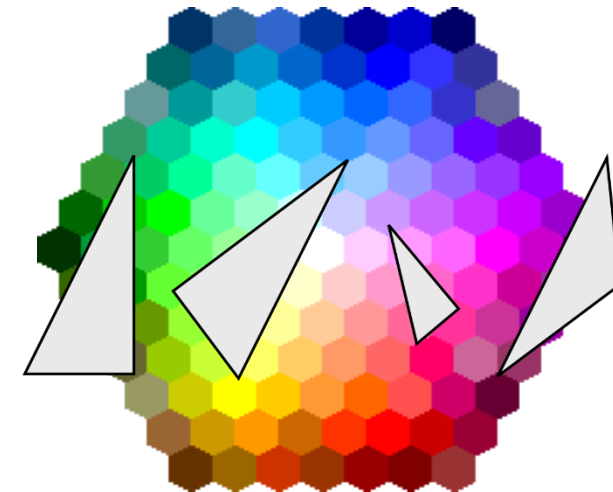
¿Y qué es la gramática de los gráficos? (gg)

- Creado por Wilkinson en 1999.
- Describe las características internas que recaen detrás de un gráfico.

Invoice #	Created	Status	Customer	Due	Service	Worked
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs
INV-0001	Aug 12, 2018	Outstanding	AC Accounting	Sept 12, 2018	UI Design	10 hrs

Datos

Mapeo

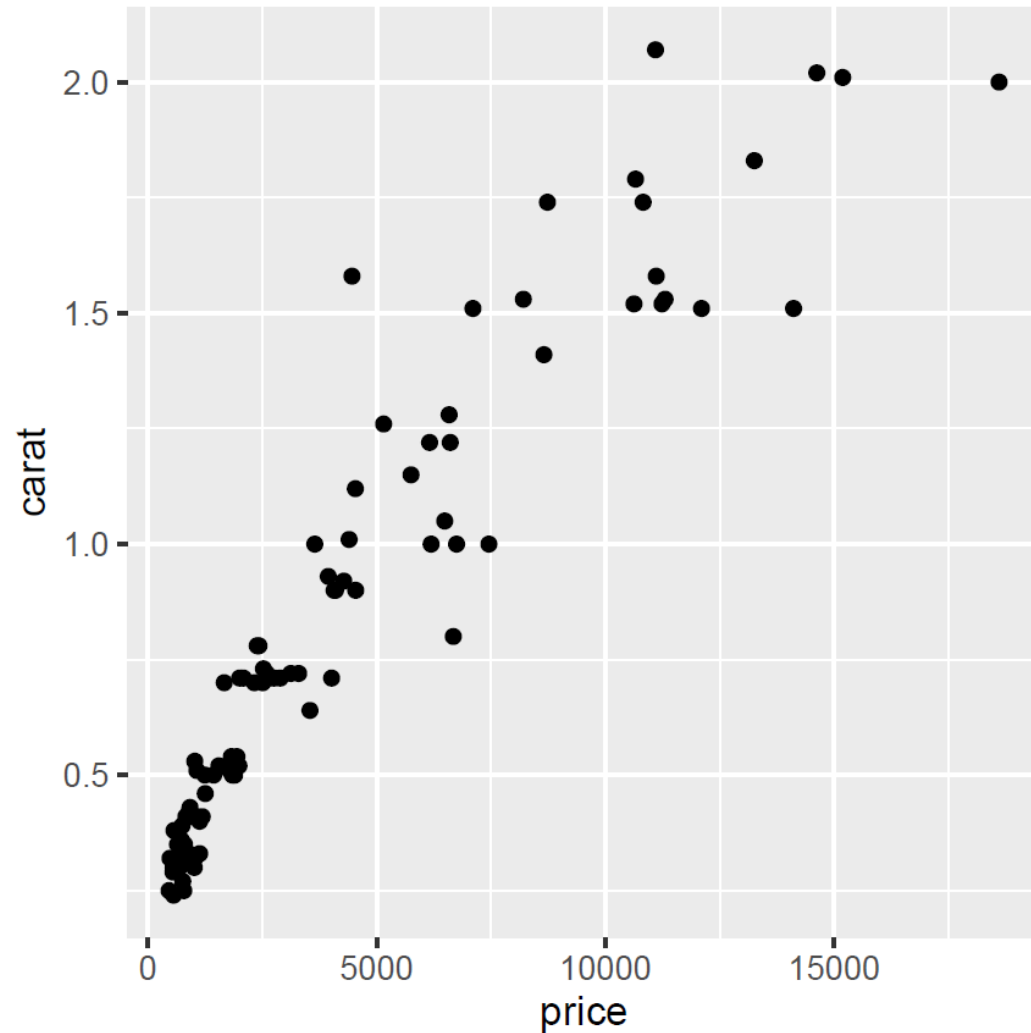


Atributos estéticos (color, forma y tamaño) de objetos geométricos (línea, punto, círculo)

Componentes de un gráfico

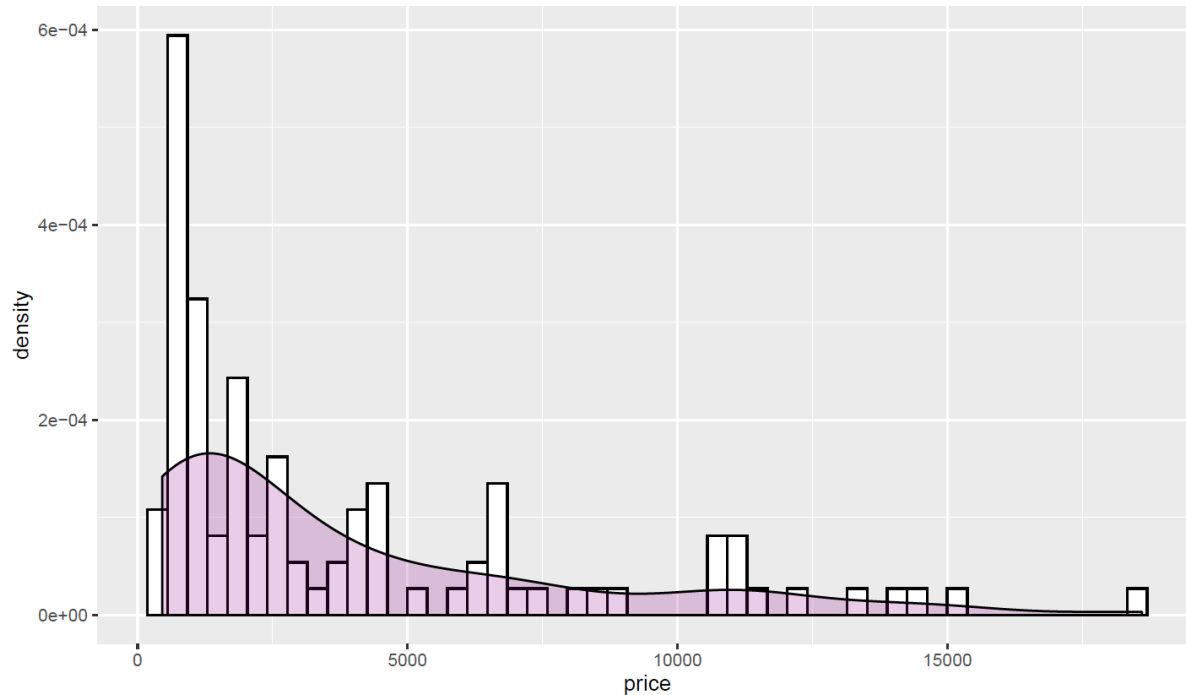
- **Conjunto de datos** (en ggplot2 nos restringimos un poco) y un conjunto de **mapeos estéticos (aesthetic mappings)**.
- Múltiples **capas (layers)**.
- **Escalas (scales)** por cada estética.
- Una especificación de **facetas (facets)**.
- Un sistema de **coordenadas (coordinate system)**.

Componentes de un gráfico



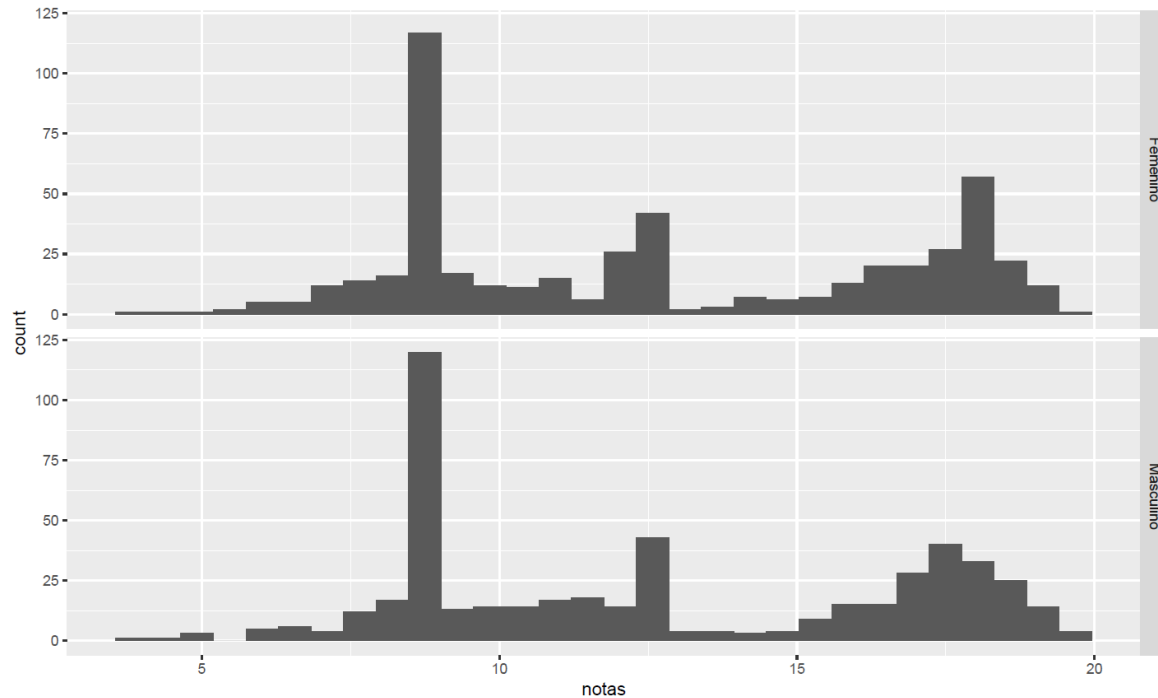
- **Conjunto de datos:** diamonds.
- **Mapeos:** carat y price.
- **Capas:** puntos.
- **Escalas:** posición x e y.
- **Facetas:** ninguna.

Componentes de un gráfico



- **Conjunto de datos:** diamonds.
- **Mapeos:** Price y density.
- **Capas:** barras y densidad.
- **Escalas:** posición x e y.
- **Facetas:** ninguna.

Componentes de un gráfico



- **Conjunto de datos:**
- **Mapeos:**
- **Capas:**
- **Escalas:**
- **Facetas:**

Gráfico de dispersión

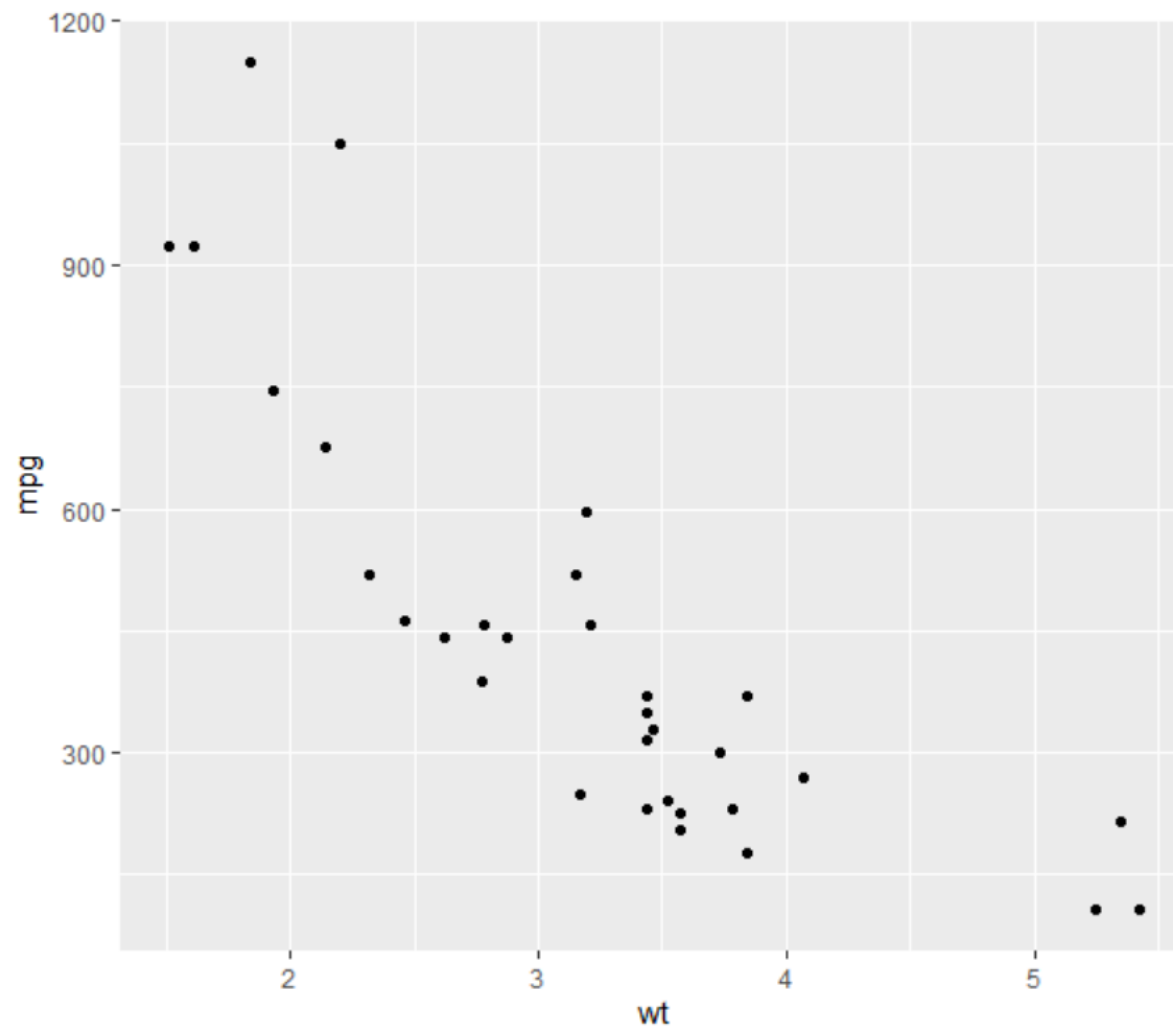
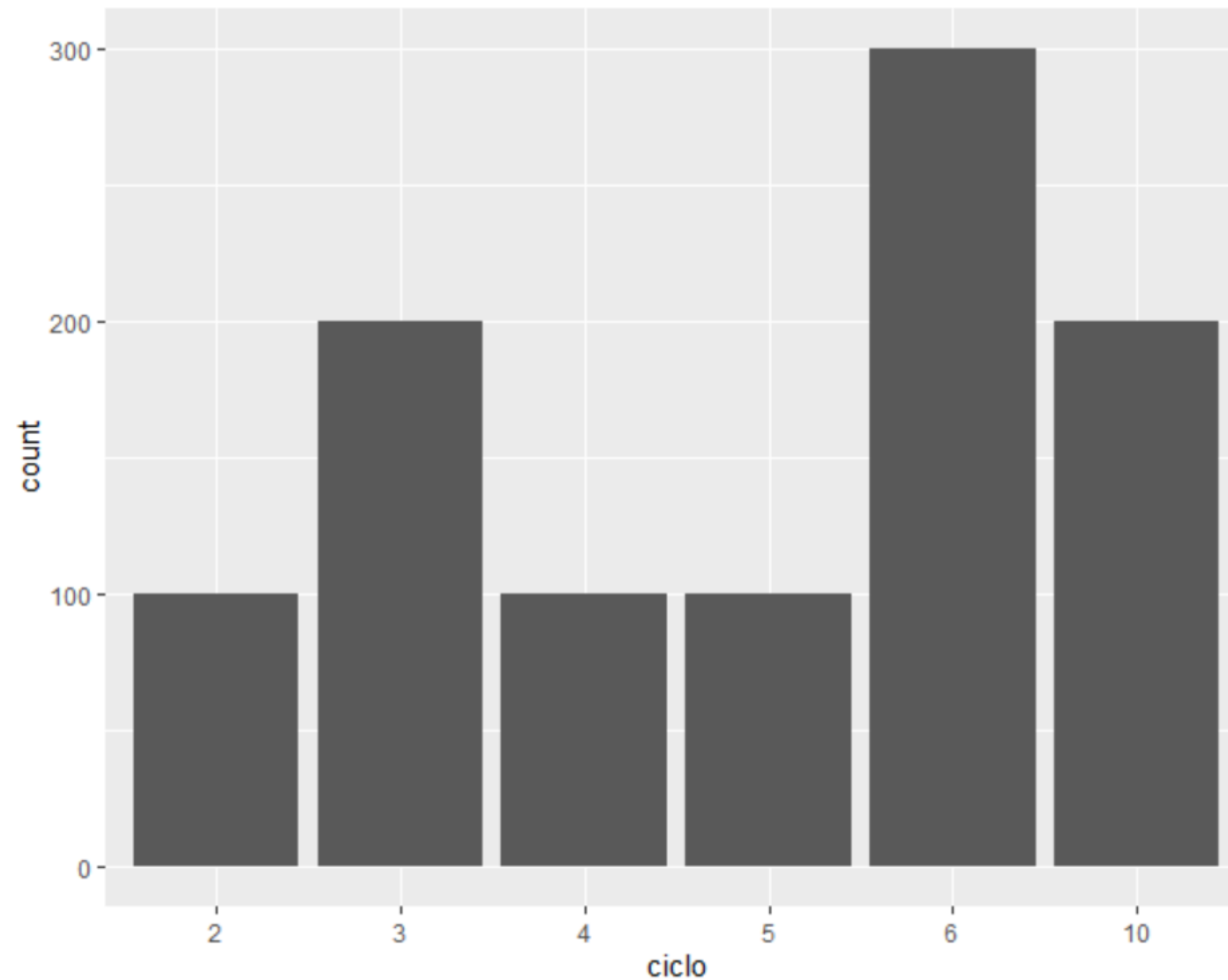
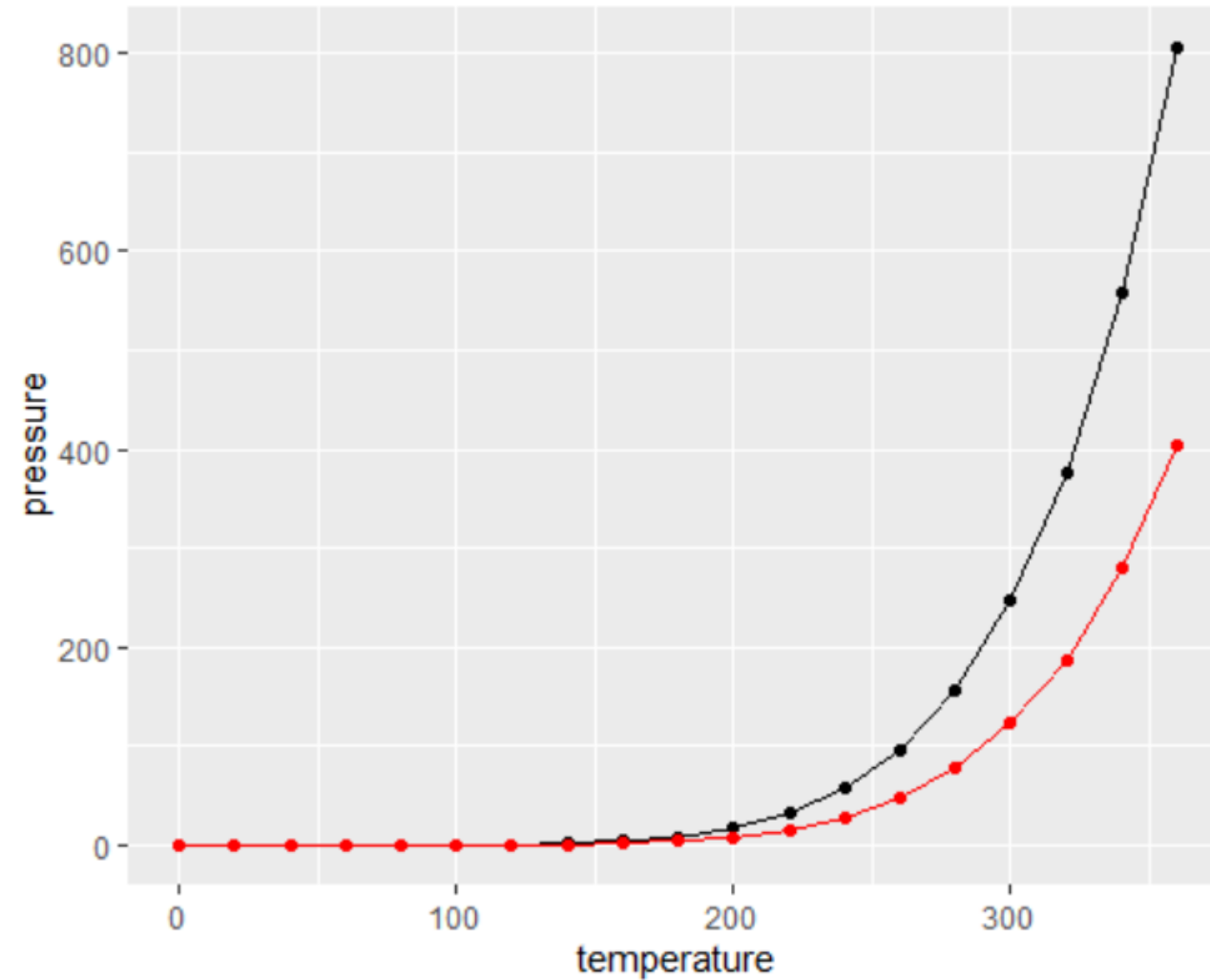


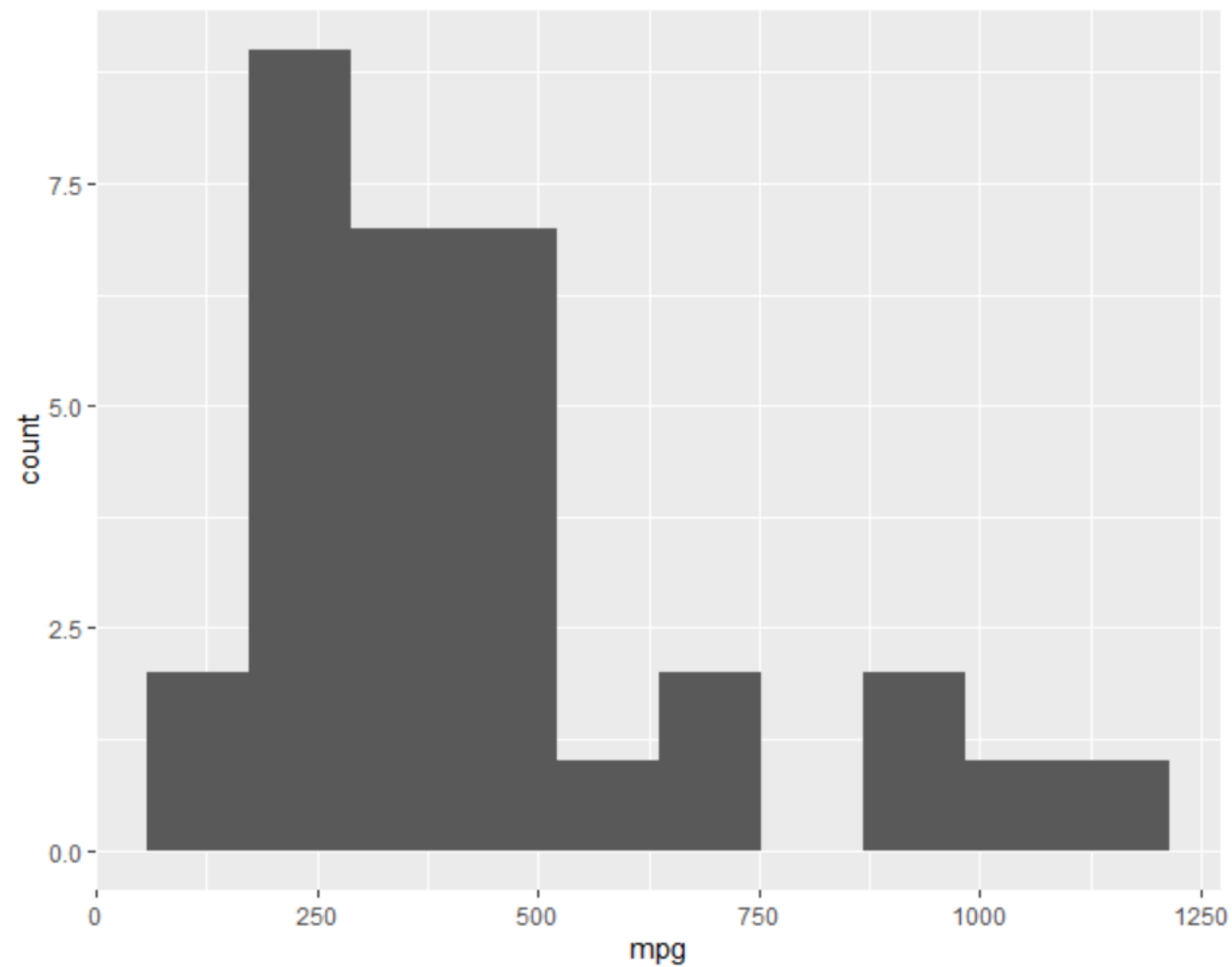
Gráfico de barras



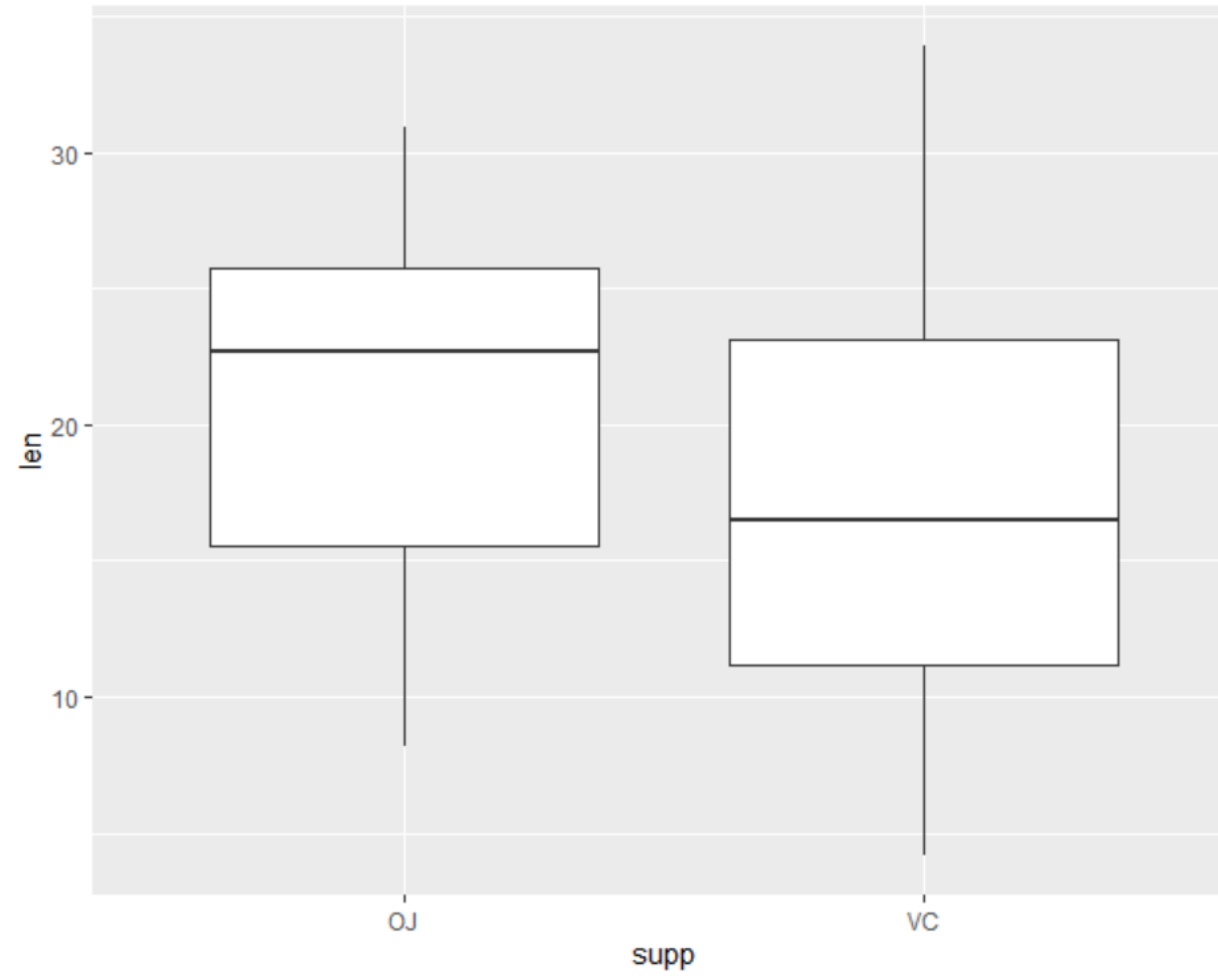
Gráficos de líneas



Histograma



Boxplot



Curvas

