

# **Visualización de datos con R**

Felipe Ortega

María Jesús Algar

2024-11-14

# Tabla de contenidos

<b>Prefacio</b>	<b>1</b>
Requisitos previos . . . . .	1
 <b>I Fundamentos</b>	 <b>2</b>
<b>1 Elementos de la visualización de datos</b>	<b>3</b>
1.1 Canales . . . . .	3
1.2 Selección de gráficos . . . . .	3
1.3 Paletas de colores . . . . .	3
1.4 Gráficos dinámicos . . . . .	3
1.5 Interactividad . . . . .	3
<b>2 Principios de visualización de datos</b>	<b>4</b>
2.1 Principios de E. Tufte . . . . .	4
2.2 Buenas prácticas y recomendaciones . . . . .	4
<b>3 Galería de gráficos</b>	<b>5</b>
3.1 Taxonomía de gráficos . . . . .	5
<b>4 Gramática de gráficos</b>	<b>6</b>
4.1 Origen y propósito . . . . .	6
4.2 Librerías y paquetes . . . . .	6
 <b>II Implementación</b>	 <b>7</b>
<b>5 El paquete ggplot2</b>	<b>8</b>
5.1 Anatomía de un gráfico con ggplot2 . . . . .	8
5.2 Elementos estéticos . . . . .	8
5.3 Geometrías . . . . .	8
5.4 Escalas . . . . .	8
5.5 Etiquetas y título . . . . .	8
5.6 Anotaciones . . . . .	8
5.7 Temas . . . . .	8
5.8 Extensiones de ggplot2 . . . . .	8
5.9 Taller práctico 1: construcción de gráficos paso a paso . . . . .	8
<b>6 Gráficos para evaluación de modelos</b>	<b>9</b>
6.1 Evaluación de modelos . . . . .	9
6.2 Ejemplo: modelos de regresión . . . . .	9
6.3 Ejemplo: explicabilidad de modelos . . . . .	9

<b>7</b>	<b>Visualización de series temporales</b>	<b>10</b>
7.1	Datos de series temporales . . . . .	10
7.2	Taller práctico 2: representación de series temporales . . . . .	10
7.3	Taller práctico 3: visualización de modelos de predicción . . . . .	10
<b>8</b>	<b>Visualización de datos espaciales</b>	<b>11</b>
8.1	Datos espaciales . . . . .	11
8.2	Representación de datos espaciales . . . . .	11
8.3	Taller práctico 4: el paquete <code>leaflet</code> . . . . .	11
<b>9</b>	<b>Recursos adicionales</b>	<b>12</b>
	<b>Referencias</b>	<b>13</b>
	<b>Apéndices</b>	<b>14</b>
<b>A</b>	<b>Referencia de comandos</b>	<b>14</b>
A.1	Paquete <code>ggplot2</code> . . . . .	14
A.2	Series temporales . . . . .	14
A.3	Datos espaciales . . . . .	14
<b>B</b>	<b>Paquetes R y atribuciones</b>	<b>15</b>
B.1	Requisitos previos . . . . .	15
B.2	Atribución de imágenes e iconos . . . . .	15
	<b>Referencias</b>	<b>16</b>

# Prefacio

En este taller exploramos los fundamentos prácticos para la creación de gráficos para visualización de datos utilizando el lenguaje R. La representación gráfica de la información es un apartado fundamental en todo proyecto de Ciencia de Datos, puesto que permite descubrir patrones y características no evidentes, identificar valores atípicos, así como resumir la información de forma más evidente y directa para el espectador. Aunque este apartado es en sí mismo muy amplio, en este taller vamos a centrarnos en los tipos de gráficos básicos, así como en visualizaciones que por su especial relevancia para aplicaciones en Ciencias Agrarias y Ambientales puedan resultar interesantes para los/as participantes

Este es un **taller práctico** que presenta ejemplos reales y comandos para crear paso a paso visualizaciones de datos efectivas con R. Además, junto a la explicación de los conceptos clave para entender este proceso también se ofrecen recomendaciones sobre buenas prácticas para crear gráficos más informativos y claros, evitando errores comunes y potenciando su capacidad de condensar gran cantidad de información sin que conlleve una excesiva complejidad para su correcta interpretación.

Los apuntes para este taller práctico se han realizado con Quarto, una herramienta para creación de documentación científica y programación literaria compatible con R y otros lenguajes de programación científica.

## Requisitos previos

Para poder realizar los ejemplos incluidos en este taller necesitas tener instalado R y una IDE de desarrollo para este lenguaje. Se recomienda instalar RStudio o MS Visual Code como entorno de programación.

- Instalación de R.
- Instalación de RStudio.

Adicionalmente, es necesario instalar una serie de paquetes R antes de ejecutar los ejemplos, para que todas las dependencias estén disponibles en nuestro sistema. Consulta el Apéndice Sección B.1 para comprobar el listado de paquetes R necesarios.

**Parte I**

**Fundamentos**

# **1 Elementos de la visualización de datos**

## **1.1 Canales**

- T. Munzner

## **1.2 Selección de gráficos**

- Según el tipo de variables.

## **1.3 Paletas de colores**

## **1.4 Gráficos dinámicos**

## **1.5 Interactividad**

- Interfaces interactivas
- Componentes reactivos

## **2 Principios de visualización de datos**

### **2.1 Principios de E. Tufte**

### **2.2 Buenas prácticas y recomendaciones**

- K. Healy
- C. Wilke

## 3 Galería de gráficos

En este capítulo, presentamos los principales tipos gráficos para visualización de datos que se suelen encontrar en proyectos de análisis de datos. Es esencial saber con qué opciones contamos para después elegir el tipo de gráfico y su modalidad que mejor se adapte a nuestros objetivos.

### 3.1 Taxonomía de gráficos

- Storytelling with Data.



## 4 Gramática de gráficos

Breve presentación del paradigma de *grammar of graphics*, adoptado por muchas de las principales bibliotecas de visualización de datos.

### 4.1 Origen y propósito

L. Wilkinson

### 4.2 Librerías y paquetes

Ejemplos de paquetes y librerías en R y otros lenguajes que adoptan el paradigma de la *grammar of graphics*.

# **Parte II**

## **Implementación**

## 5 El paquete ggplot2

En este capítulo se introduce el paquete `ggplot2` de R, una de las herramientas de visualización de datos más populares en la actualidad y que ha sido incluso exportado a otros lenguajes (como Python).

### 5.1 Anatomía de un gráfico con ggplot2

### 5.2 Elementos estéticos

### 5.3 Geometrías

### 5.4 Escalas

### 5.5 Etiquetas y título

### 5.6 Anotaciones

### 5.7 Temas

### 5.8 Extensiones de ggplot2

### 5.9 Taller práctico 1: construcción de gráficos paso a paso

## **6 Gráficos para evaluación de modelos**

### **6.1 Evaluación de modelos**

### **6.2 Ejemplo: modelos de regresión**

- Gráficos clásicos de resumen de resultados.
- Gráficos clásicos para evaluación de modelos.

### **6.3 Ejemplo: explicabilidad de modelos**

Casos sencillos sobre importancia de variables.

## **7 Visualización de series temporales**

### **7.1 Datos de series temporales**

Breve introducción a la estructura y representación de datos de series temporales, centrándonos en el paquete `tsibble`.

### **7.2 Taller práctico 2: representación de series temporales**

### **7.3 Taller práctico 3: visualización de modelos de predicción**

## **8 Visualización de datos espaciales**

### **8.1 Datos espaciales**

### **8.2 Representación de datos espaciales**

### **8.3 Taller práctico 4: el paquete leaflet**

## 9 Recursos adicionales

Listado de recursos adicionales de interés.

See Knuth (1984) for additional discussion of literate programming.

## Referencias

Knuth, D. E. (1984). Literate Programming. *Comput. J.*, 27(2), 97-111. <https://doi.org/10.1093/comjnl/27.2.97>



# **A Referencia de comandos**

**A.1 Paquete ggplot2**

**A.2 Series temporales**

**A.3 Datos espaciales**

## **B Paquetes R y atribuciones**

### **B.1 Requisitos previos**

Para ejecutar los ejemplos incluidos en este taller, se necesita tener instalado R y una IDE de desarrollo para este lenguaje, como por ejemplo RStudio o Microsoft VS Code.

### **B.2 Atribución de imágenes e iconos**

## Referencias

Knuth, D. E. (1984). Literate Programming. *Comput. J.*, 27(2), 97-111. <https://doi.org/10.1093/comjnl/27.2.97>