

Visualización de datos con R y Rstudio

Temario.....	1
Introducción.....	1
Objetivos.....	2
1. Comunicación Efectiva.....	2
2. Selección del Tipo de Gráfico.....	2
3. Principios de Diseño Visual.....	2
4. Revelar Patrones y Tendencias.....	2
5. Contar Historias.....	2
6. Respalda la Toma de Decisiones.....	2
Pasos de diseño de una visualización de datos.....	2
Base R Graphics.....	3
Funciones.....	3
Ejemplo 1) Diagrama de dispersión.....	3
Argumentos.....	4
Ejemplo 2) Diagrama de dispersión - Primeras Variaciones.....	5
ggplot2.....	6
Componentes de Mapeo.....	7
Capa (Layer).....	7
Escala (Scales).....	7
Coordenadas (Coord).....	7
Faceta (Facet).....	8
Tema (Theme).....	8

Temario

- Introducción a la Visualización de Datos
- Gráficos Básicos en R: scatterplot, histograma, boxplot, heatmaps, densidad, pca, biplot
- ggplot2
- Visualización Exploratoria de Datos
- Generación de reportes
- Presentación e interpretación de resultados
- Otras Librerías de Visualización en R

Introducción

La visualización de datos se refiere al proceso de representar información de manera gráfica y comprensible.

En lugar de simplemente examinar conjuntos de números o texto, la visualización utiliza gráficos y diagramas para presentar datos de manera visual, facilitando la interpretación y comprensión de patrones, tendencias y relaciones.

La visualización de datos:

Proporciona la capacidad de comprender grandes cantidades de datos. La información importante de más de un millón de mediciones está inmediatamente disponible.

Posibilita la percepción de propiedades emergentes que no se anticiparon.

Permite que los problemas con los datos mismos se hagan inmediatamente evidentes. Una visualización comúnmente revela cosas no solo sobre los datos en sí, sino también sobre la forma en que se recopilan. Con una visualización adecuada, los errores y artefactos en los datos a menudo saltan a la vista. Por esta razón, las visualizaciones pueden ser invaluableles en el control de calidad.

Promueve la comprensión de características tanto a gran escala como a pequeña escala de los datos. Puede ser especialmente valiosa al permitir la percepción de patrones que vinculan características locales.

Facilita la formación de hipótesis.

Objetivos

1. Comunicación Efectiva

- Definición: La visualización de datos busca comunicar de manera efectiva información compleja a través de representaciones gráficas.
- Claridad del Mensaje: El objetivo principal es transmitir mensajes de manera clara y comprensible para el público objetivo.

2. Selección del Tipo de Gráfico

- Adaptación a los Datos: Implica elegir el tipo adecuado de gráfico que se ajuste a la naturaleza de los datos que se están presentando.
- Contextualización: Considerar el contexto y los objetivos para seleccionar el formato más apropiado.

3. Principios de Diseño Visual

- Facilitar la Interpretación: Aplicar principios de diseño visual para hacer que la información sea fácilmente interpretable.
- Eliminación de Ruido: Eliminar elementos innecesarios que puedan distraer o confundir al espectador.

4. Revelar Patrones y Tendencias

- Identificación de Patrones: La visualización efectiva revela patrones y tendencias en los datos que pueden no ser evidentes en formas tabulares.
- Análisis Exploratorio: Facilita el análisis exploratorio de datos para descubrir información relevante.

5. Contar Historias

- Narrativa Visual: Permite construir una narrativa visual alrededor de los datos, facilitando la comprensión de la audiencia.
- Impacto Emocional: Puede agregar un componente emocional que fortalezca el impacto del mensaje.

6. Respalda la Toma de Decisiones

- Información Accionable: Proporcionar información que sea útil para respaldar decisiones informadas.
- Persuasión Informada: Ayudar a persuadir y respaldar argumentos con base en la visualización de datos.

Pasos de diseño de una visualización de datos

1. Adquirir los datos
2. Organizarlos de acuerdo a alguna estructura
3. Filtrar datos inadecuados o incorrectos
4. Minar (extraer) para discernir patrones o relaciones
5. Representar dichos patrones en forma visual
6. Refinar la representación para asegurar su efectividad
7. Interactuar con la representación para ulteriores exploraciones

Base R Graphics

Funciones gráficas básicas que están integradas directamente en el lenguaje R y no requieren la instalación de ninguna librería adicional.

Funciones

- **plot()**
- **hist()**
- **barplot()**
- **boxplot()**
- **heatmap()**

Ejemplo 1) Diagrama de dispersión

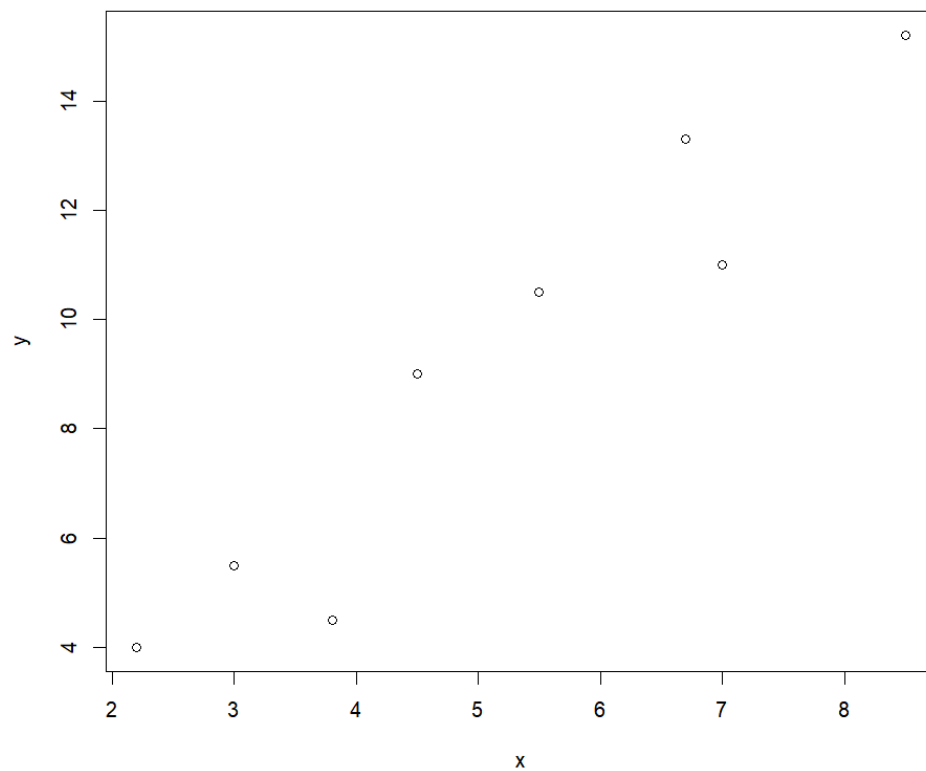
- Variables a graficar:

```
x <- c(2.2, 3, 3.8, 4.5, 7, 8.5, 6.7, 5.5)
```

```
y <- c(4, 5.5, 4.5, 9, 11, 15.2, 13.3, 10.5)
```

- Graficar los puntos

```
plot(x, y)
```



Argumentos

main: título

sub: subtítulo

xlab o ylab: nombre de los ejes

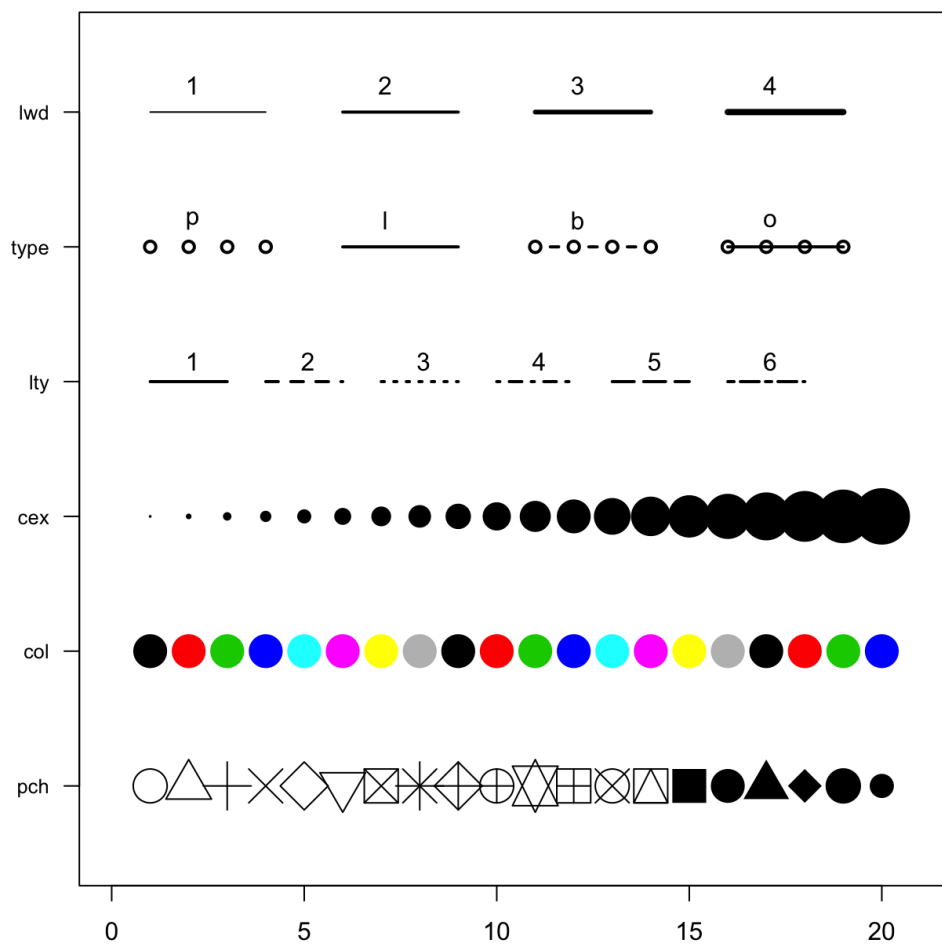
xlim o ylim: rango de los ejes

col: color de los bordes

lty: tipo de líneas

lwd: ancho de líneas

alpha: transparencia

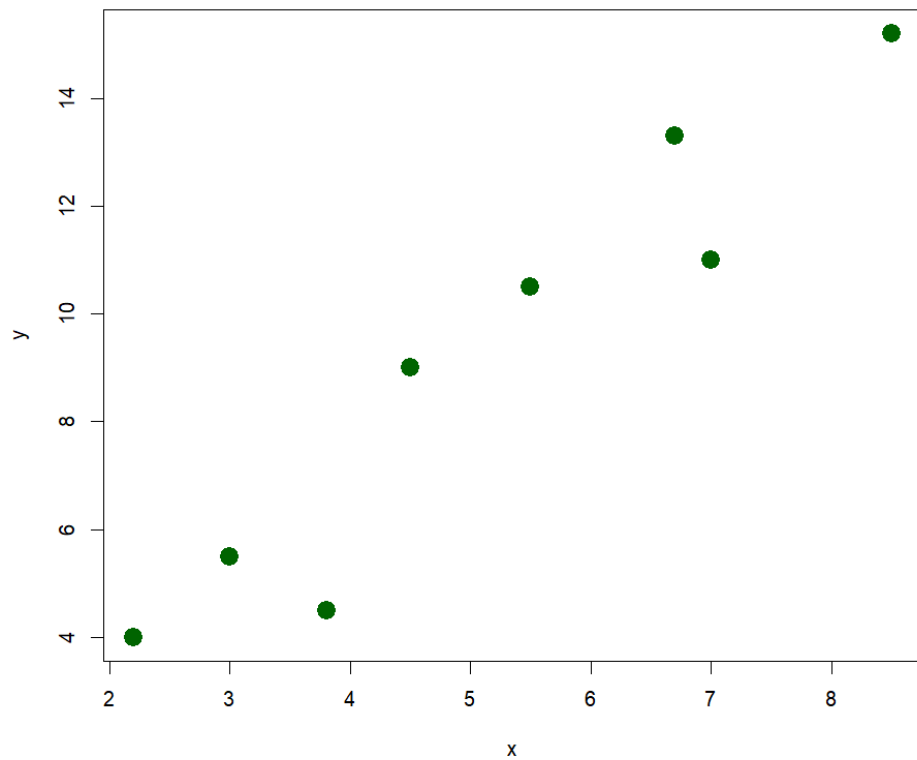


Ejemplo 2) Diagrama de dispersión - Primeras Variaciones

```

plot(x, y,          #Datos
pch=19,            #Símbolo
col="darkgreen",   #Color Símbolo
cex=2)             #Tamaño Símbolo

```

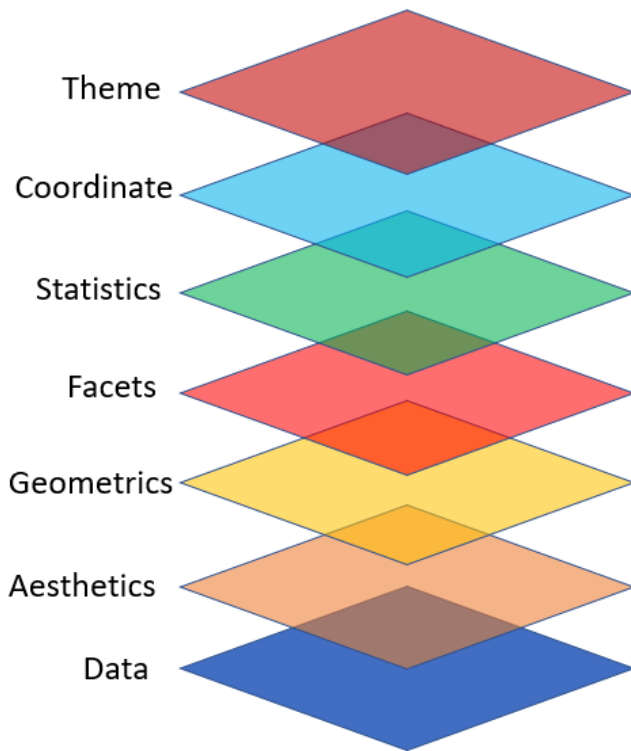


ggplot2

Los gráficos están compuestos por:

Datos: información que se desea visualizar.

Mapeo: descripción de cómo se asignan las variables de los datos a atributos estéticos.



Componentes de Mapeo

Capa (Layer)

Colección de elementos geométricos y transformaciones estadísticas.

Elementos geométricos (geoms) representan lo que se ve en el gráfico: puntos, líneas, polígonos, etc.

Transformaciones estadísticas (stats) resumen los datos, como agrupar y contar observaciones para crear un histograma o ajustar un modelo lineal.

Escalas (Scales)

Mapean valores en el espacio de datos a valores en el espacio estético (color, forma o tamaño).

Incluyen la leyenda y los ejes, facilitando la lectura de los valores originales desde el gráfico (mapeo inverso).

Coordenadas (Coord)

Describen cómo se asignan las coordenadas de los datos al plano gráfico.

Proporcionan ejes y líneas de cuadrícula para facilitar la lectura del gráfico.

Usualmente se utiliza el sistema de coordenadas cartesianas, pero hay otras opciones como coordenadas polares y proyecciones de mapas.

Faceta (Facet)

Especifica cómo dividir y mostrar subconjuntos de datos como múltiples gráficos pequeños (condicionamiento o latticing/trellising).

Tema (Theme)

Controla detalles de visualización, como tamaño de fuente y color de fondo.

ggplot2 tiene valores predeterminados cuidadosamente seleccionados.

Actividad de Cierre

- 1) Usar data() para visualizar los datos precargados en R
- 2) Importar el set de datos ToothGrowth usando data(ToothGrowth)

Este set de datos contiene información sobre el efecto de la vitamina C en el crecimiento de los dientes de Conejillos de indias

- 3) Explorar los datos de la tabla.
 - a) ¿Qué columnas tiene?
 - b) ¿Qué tipo de valor hay en cada columna?
 - c) ¿En qué condiciones se realizaron los ensayos?

Los datos corresponden a:

- len : numeric, tooth (odontoblast, actually) length
- supp : factor, supplement type, 2 levels, VC is ascorbic acid, and OJ is orange juice
- dose : numeric, dose (mg/day)

Resultados:

```
library(ggplot2)
```

```
# load data
```

```
data(ToothGrowth) #Dataset - The Effect of Vitamin C on Tooth Growth in Guinea Pigs
```

```
ToothGrowth
```

```
qplot(len,data=ToothGrowth)
```



```
qplot(len,data=ToothGrowth,bins=5)
```

```
qplot(len,data=ToothGrowth,bins=5, geom = "histogram", fill = as.factor(supp))
```

```
qplot(len,data=ToothGrowth,bins=5, geom = "density", fill = as.factor(supp))
```

```
qplot(supp,len,data=ToothGrowth)
```

```
qplot(supp,len,data=ToothGrowth) + geom_boxplot(aes(fill = supp))
```

```
qplot(supp,len,data=ToothGrowth, facets=~dose) + geom_boxplot(aes(fill = supp))
```

```
qplot(supp,len,data=ToothGrowth, facets=~dose, main="Tooth growth of guinea pigs by  
supplement type and dosage (mg)",xlab="Supplement type", ylab="Tooth length") +  
geom_boxplot(aes(fill = supp))
```

```
qplot(supp,len,data=ToothGrowth, facets=~dose, geom = "boxplot", main="Tooth  
growth of guinea pigs by supplement type and dosage (mg)",xlab="Supplement type",  
ylab="Tooth length",fill = supp)
```

```
qplot(supp,len,data=ToothGrowth, facets=~dose, geom = "violin", main="Tooth growth  
of guinea pigs by supplement type and dosage (mg)",xlab="Supplement type",  
ylab="Tooth length", fill = supp)
```