

Principios de Buena Visualización

Visualización de Datos con R

Prof. Nicolás Sidicaro - FCE-UBA

2025-10-28

Encuentro 3

Principios de Buena Visualización

Desarrollar criterio para evaluar y crear visualizaciones efectivas

Objetivos del encuentro

Al finalizar esta clase serán capaces de:

- **Evaluar** la calidad de una visualización aplicando principios de diseño
- **Identificar** errores comunes y problemas en gráficos existentes
- **Decidir** cuándo es apropiado usar un gráfico versus una tabla
- **Aplicar** fundamentos de diseño para mejorar sus visualizaciones

Agenda

1. Fundamentos del diseño

- Principios visuales fundamentales
- Uso del color, tipografía y escalas
- Percepción visual y cognición

2. Análisis de malas visualizaciones

- Gráficos engañosos y sobrecarga de información

3. Cuándo NO hacer un gráfico

- Criterios para tablas vs. gráficos

4. Actividades prácticas

- Identificación de problemas y rediseño

Fundamentos del Diseño

El Principio Fundamental

Claridad antes que belleza

"La excelencia en el diseño estadístico consiste en ideas complejas comunicadas con claridad, precisión y eficiencia" - Edward Tufte

Los tres pilares:

1. **Claridad:** El mensaje debe ser evidente e inmediato
2. **Precisión:** Los elementos visuales reflejan fielmente los datos
3. **Eficiencia:** Máxima información con mínimo ruido visual

Principios de Percepción Visual

Jerarquía de precisión perceptual (Cleveland & McGill, 1984):

1. Posición en escala común (más preciso)
2. Posición en escalas no alineadas
3. Longitud, dirección, ángulo
4. Área
5. Volumen, curvatura
6. Sombreado, saturación de color (menos preciso)

Implicación práctica:

Usar **barras** en lugar de **áreas** para comparaciones precisas

Principios de Gestalt en Visualización

¿Qué son los principios de Gestalt?

Leyes de percepción visual que explican cómo el cerebro organiza y procesa la información visual de manera automática.

Relevancia para visualización:

- Nos ayudan a **predecir** cómo los usuarios interpretarán nuestros gráficos
- Permiten **diseñar** visualizaciones más intuitivas
- Explican por qué algunas visualizaciones "funcionan" y otras no

Principio fundamental:

"El todo es mayor que la suma de sus partes" - percibimos patrones y estructuras globales

Gestalt 1: Proximidad

Elementos cercanos espacialmente se perciben como relacionados o pertenecientes al mismo grupo

Aplicaciones en visualización:

- **Agrupación de categorías:** Barras agrupadas vs. separadas
- **Paneles múltiples:** Facetas relacionadas juntas
- **Elementos de UI:** Leyendas cerca de sus gráficos correspondientes
- **Etiquetas:** Cerca de los elementos que describen

Error común:

Leyenda muy alejada del gráfico - el usuario no asocia fácilmente los colores

Regla práctica:

La distancia visual debe reflejar la relación conceptual de los datos

Gestalt 2: Similitud

Elementos con características visuales similares (color, forma, tamaño) se perciben como grupo

Aplicaciones en visualización:

- **Codificación por color:** Todas las series de datos de una categoría en el mismo color
- **Formas consistentes:** Mismos símbolos para mismos tipos de datos
- **Tamaños coherentes:** Elementos de igual importancia en igual tamaño
- **Estilos de línea:** Sólida para datos reales, punteada para proyecciones

Error común:

Usar colores/formas aleatorios sin significado semántico

Regla práctica:

La similitud visual debe indicar similitud conceptual

Gestalt 3: Cerramiento (Cierre)

Tendemos a completar mentalmente figuras incompletas o cerrar espacios abiertos

Aplicaciones en visualización:

- **Líneas de tendencia:** No necesitan conectar todos los puntos para ser percibidas
- **Áreas bajo curvas:** El cerebro "cierra" el área automáticamente
- **Formas implícitas:** Scatter plots forman "nubes" o patrones sin líneas explícitas
- **Bordes de gráficos:** No siempre necesarios si hay otros elementos delimitadores

Ejemplo práctico:

En un scatter plot, podemos ver clusters sin dibujar círculos alrededor

Regla práctica:

Usar elementos mínimos necesarios - el cerebro completa el resto

Gestalt 4: Continuidad

El ojo sigue naturalmente líneas, curvas y patrones continuos

Aplicaciones en visualización:

- **Series temporales:** Líneas conectadas facilitan seguimiento de tendencias
- **Flujo de lectura:** De izquierda a derecha, arriba hacia abajo
- **Conexiones:** Líneas suaves mejor que ángulos abruptos
- **Jerarquía visual:** Crear "caminos" visuales para guiar la atención

Error común:

Interrumpir líneas de tendencia sin justificación (datos faltantes)

Regla práctica:

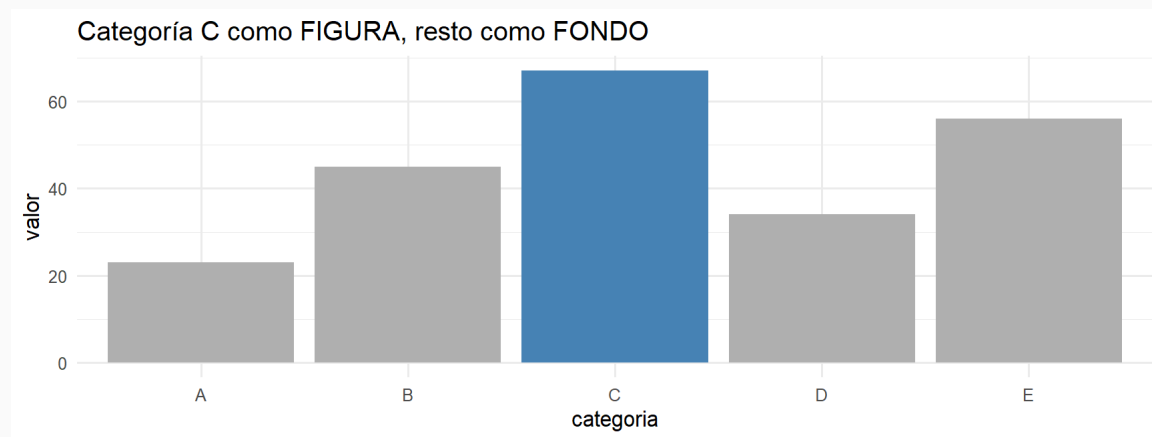
Facilitar el "recorrido visual" del usuario por los datos

Gestalt 5: Figura y Fondo

Distinguimos automáticamente entre elemento principal (figura) y contexto (fondo)

Aplicaciones en visualización:

- **Contraste:** Datos principales destacados, elementos secundarios atenuados
- **Capas:** Datos en primer plano, grillas/ejes como fondo
- **Énfasis:** Color saturado para figura, grises para fondo
- **Jerarquía:** Lo importante "emerge" visualmente



Gestalt 6: Simplicidad (Prägnanz)

Preferimos la interpretación más simple y estable de los estímulos visuales

Aplicaciones en visualización:

- **Diseño minimalista:** Eliminar elementos decorativos innecesarios
- **Patrones simples:** Preferir estructuras regulares y predecibles
- **Interpretación directa:** Evitar ambigüedad en codificación visual
- **Carga cognitiva mínima:** Una idea principal por visualización

Error común:

Gráficos sobrecargados que requieren interpretación compleja

Regla práctica:

"Simplicidad elegante" - máximo insight con mínima complejidad visual

Gestalt 7: Destino Común

Elementos que se mueven o cambian juntos se perciben como grupo relacionado

Aplicaciones en visualización:

- **Animaciones:** Series que cambian juntas en el tiempo
- **Interactividad:** Elementos que responden conjuntamente al hover/click
- **Transiciones:** Cambios coordinados entre estados del gráfico
- **Patrones temporales:** Variables que siguen tendencias similares

Ejemplo:

En un gráfico animado, líneas que suben/bajan juntas se perciben como relacionadas

Regla práctica:

Usar movimiento/cambio para mostrar relaciones entre datos

Aplicación Práctica de Gestalt

Checklist de principios de Gestalt:

1. ☐ **Proximidad:** ¿Elementos relacionados están cerca visualmente?
2. ☐ **Similitud:** ¿Uso color/forma de manera consistente y semántica?
3. ☐ **Cerramiento:** ¿Permito que el cerebro complete patrones naturalmente?
4. ☐ **Continuidad:** ¿Facilito el recorrido visual por los datos?
5. ☐ **Figura-fondo:** ¿Los datos importantes emergen del contexto?
6. ☐ **Simplicidad:** ¿Es la interpretación más simple la correcta?
7. ☐ **Destino común:** ¿Uso cambio/movimiento para mostrar relaciones?

Resultado esperado:

Visualizaciones intuitivas que aprovechan la percepción natural del cerebro

Uso Estratégico del Color

Funciones del color en visualización:

1. Distinguir categorías

- Máximo 6-8 colores categóricos
- Usar paletas accesibles (ColorBrewer, Viridis)
- Evitar rojo-verde juntos (daltonismo)

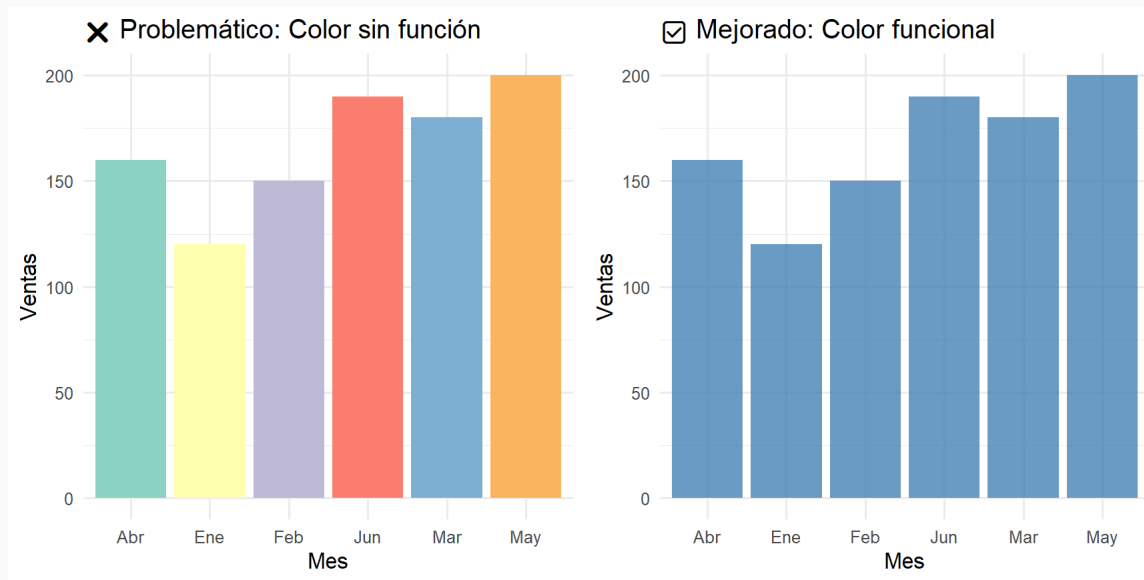
2. Mostrar magnitud/intensidad

- Escalas secuenciales: claro → oscuro
- Escalas divergentes: dos colores desde neutro
- Consistent with data direction

3. Enfatizar elementos clave

- Color para destacar, gris para contexto
- Regla 80-20: 80% neutro, 20% destacado

Uso del Color: Errores Comunes



Principio: El color debe **comunicar**, no solo **decorar**

Tamaño y Proporciones

Principios de escalamiento:

1. Longitud vs. Área

- **Barras:** escalamiento por longitud (lineal)
- **Círculos:** cuidado con área vs. diámetro
- **Pictogramas:** usar altura, no área bidimensional

2. Proporciones áureas

- Gráficos: ratio ~1.6:1 (ancho:alto)
- Evitar gráficos demasiado altos o anchos
- Mantener legibilidad en diferentes medios

3. Jerarquía de tamaños

- Título: mayor énfasis visual
- Subtítulos: información complementaria
- Ejes: legibles pero discretos

Espaciado y Alineación

White space (espacio en blanco):

- No es "espacio perdido" - mejora legibilidad
- Agrupa elementos relacionados
- Separa elementos diferentes
- Guía el ojo del lector

Alineación:

- Elementos relacionados: misma alineación
- Etiquetas de ejes: perpendiculares y legibles
- Leyendas: posición que no obstruya datos
- Márgenes: suficientes para no apretar

Principio de proximidad:

Elementos relacionados deben estar visualmente próximos

Escalas y Ejes

Reglas fundamentales:

1. Eje Y en gráficos de barras:

- Siempre comenzar en cero
- Excepción: datos que no pueden ser cero (temperaturas, fechas)
- Justificar explícitamente cualquier excepción

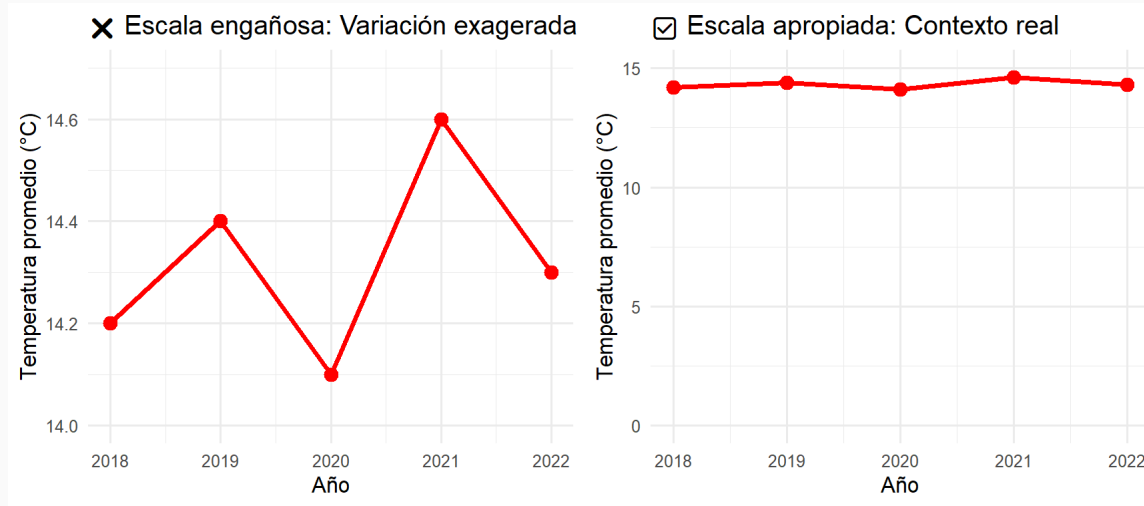
2. Proporciones honestas:

- 1 unidad visual = 1 unidad de datos
- Evitar transformaciones no lineales sin etiquetar
- Intervalos consistentes en escalas

3. Límites apropiados:

- Mostrar rango completo relevante
- No truncar para exagerar diferencias
- Contexto suficiente para interpretación

Escalas: Ejemplo de Distorsión



Mismo dato, diferente interpretación según la escala

Tipografía y Legibilidad

Jerarquía tipográfica:

Títulos (14-18pt):

- **Función:** Comunicar el mensaje principal
- **Ubicación:** Centrado o alineado izquierda
- **Estilo:** Bold, sin ser excesivo

Subtítulos (10-12pt):

- **Función:** Contexto adicional, fuente, fechas
- **Estilo:** Regular, color gris

Ejes y etiquetas (8-10pt):

- **Función:** Información técnica
- **Legibilidad:** Contraste suficiente
- **Rotación:** Evitar más de 45°

Tipografía: Reglas Prácticas

Límites recomendados:

- Máximo 2-3 tipografías diferentes
- Máximo 3-4 tamaños de fuente
- Consistent font family en todo el proyecto

Legibilidad:

- Contraste mínimo: 4.5:1 (WCAG guidelines)
- Longitud de línea: 45-75 caracteres
- Interlineado: 120-150% del tamaño de fuente

Evitar:

- Comic Sans, Papyrus en contextos profesionales
- TEXTO EN MAYÚSCULAS EXTENSO
- Colores de bajo contraste (gris claro sobre blanco)

Longitud y Proporción de Elementos

Barras y columnas:

- Ancho: 0.6-0.8 del espacio disponible
- Separación: 20-40% del ancho de barra
- Agrupación: separación menor dentro de grupos

Líneas:

- Grosor: 1-3pt según importancia
- Estilo: sólido para datos principales
- Puntos: 2-4pt, visibles pero no dominantes

Elementos de referencia:

- Líneas de grilla: 0.25-0.5pt, color gris claro
- Ejes: 0.5-1pt, color gris medio
- Líneas de referencia: 1pt, color contrastante

Análisis de Malas Visualizaciones

Categorías de Problemas Frecuentes

1. Distorsiones de escala

- Ejes truncados sin justificación
- Escalas logarítmicas no etiquetadas
- Proporciones visuales incorrectas

2. Uso problemático del color

- Colores sin función semántica
- Paletas inaccesibles (problemas daltonismo)
- Demasiados colores simultáneos

3. Sobrecarga informativa

- Demasiados elementos en una visualización
- Múltiples mensajes confusos
- Decoración que distrae del mensaje

Categorías de Problemas (cont.)

4. Elección incorrecta de gráfico

- Gráficos de torta con >7 categorías
- Barras 3D que distorsionan percepción
- Dobles ejes Y que crean correlaciones falsas

5. Manipulación temporal

- Cherry picking de fechas
- Intervalos temporales inconsistentes
- Omisión de contexto histórico relevante

6. Problemas de proporción

- Pictogramas con escalamiento bidimensional
- Áreas que no representan proporciones reales
- Elementos decorativos desproporcionados

El Problema de los Gráficos 3D

¿Por qué evitar 3D en visualización de datos?

1. Distorsión perceptual:

- La perspectiva altera las proporciones reales
- Elementos atrás parecen más pequeños
- Dificulta comparación precisa

2. Oclusión:

- Elementos se superponen y ocultan
- Pérdida de información visual

3. Complejidad innecesaria:

- No aporta información adicional
- Aumenta carga cognitiva
- Dificulta lectura rápida

Alternativas efectivas: Usar color, tamaño o facetas para tercera dimensión

El Problema del Doble Eje Y

¿Cuándo es problemático?

- Variables con escalas muy diferentes: Crea correlaciones visuales falsas
- Sin justificación técnica: Solo para que "se vea bien"
- Interpretación ambigua: ¿Cuál variable sigue cuál eje?

¿Cuándo puede usarse?

- Misma unidad, diferente magnitud: Ej. ventas en miles vs millones
- Relación causal conocida: Variables conceptualmente relacionadas
- Etiquetado explícito: Colores y etiquetas claras por eje

Alternativa recomendada: Gráficos separados o estandarización de variables

Cuándo NO Hacer un Gráfico

La Pregunta Fundamental

¿Cuál es el objetivo de comunicación?

- Mostrar valor exacto → Tabla
- Comparar magnitudes → Gráfico de barras
- Ver tendencia temporal → Gráfico de líneas
- Entender distribución → Histograma
- Buscar correlación → Scatter plot
- Consultar referencia → Tabla

Regla general:

Si necesitan el **número exacto**, probablemente necesiten una tabla

Criterios Específicos para Tablas

Usar TABLA cuando:

1. Precisión numérica crítica

- Decisiones financieras, médicas, técnicas
- Cálculos posteriores requieren valores exactos
- Diferencias pequeñas pero significativas

2. Pocas observaciones (< 5-7 puntos)

- El patrón visual no es evidente
- Cada valor individual es importante
- Espacio limitado para visualización

3. Múltiples métricas por observación

- Cada fila tiene 4+ variables relevantes
- Comparación multidimensional necesaria
- Dashboard con indicadores específicos

Criterios Específicos para Gráficos

Usar GRÁFICO cuando:

1. Patrones y tendencias

- Series temporales con >10 puntos
- Comparaciones de magnitud relativa
- Relaciones entre variables

2. Audiencia general

- Presentaciones ejecutivas
- Comunicación pública
- Medios de comunicación

3. Impacto visual necesario

- Llamar atención sobre problema/oportunidad
- Presentaciones orales
- Infografías y reportes visuales

Matriz de Decisión: Tabla vs Gráfico

Criterio	Favorece Tabla	Favorece Gráfico
Precisión de valores	Crítica	Relativa suficiente
Número de observaciones	< 7	> 10
Número de variables por obs.	> 3	1-2
Tipo de audiencia	Técnica especializada	General/ejecutiva
Medio de presentación	Documento escrito	Presentación oral
Objetivo principal	Consulta/referencia	Mostrar patrón
Frecuencia de consulta	Regular/diaria	Ocasional

Casos Límite: Cuándo es Difícil Decidir

Situaciones ambiguas:

1. Datos longitudinales cortos

- 3-6 puntos temporales
- **Solución:** Línea + tabla de valores

2. Comparaciones múltiples

- Varias categorías, varias métricas
- **Solución:** Gráfico principal + tabla detalle

3. Audiencia mixta

- Técnicos y ejecutivos juntos
- **Solución:** Gráfico con tabla en anexo

4. Espacio limitado

- Dashboards, informes compactos
- **Solución:** Sparklines o tablas condensadas

Checklist de Decisión Práctica

Antes de crear cualquier visualización:

1. ☐ ¿Quién va a usar esto y para qué?
2. ☐ ¿Necesitan el valor exacto o la tendencia general?
3. ☐ ¿Cuántos datos tengo? (cantidad y dimensiones)
4. ☐ ¿Dónde se va a mostrar? (papel, pantalla, presentación)
5. ☐ ¿Con qué frecuencia lo van a consultar?
6. ☐ ¿Qué decisión deben tomar con esta información?

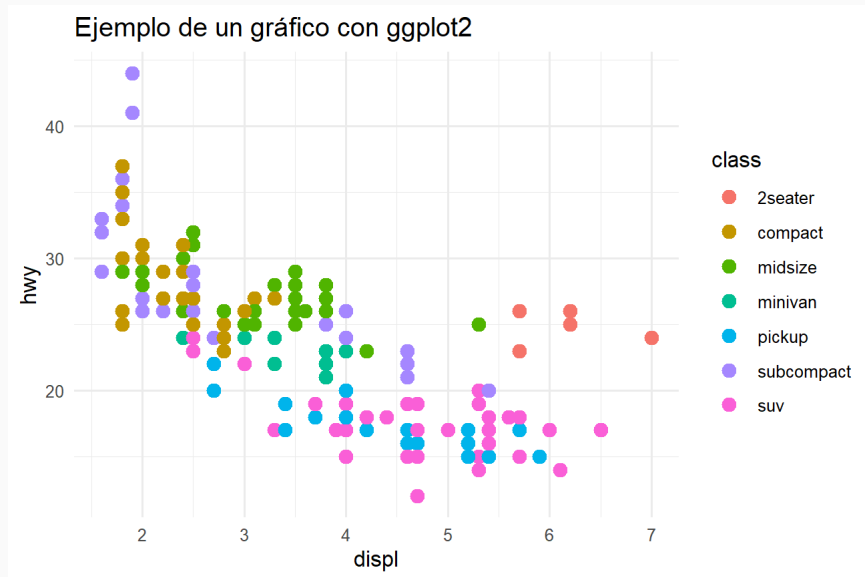
Regla práctica:

Si dudas, haz ambos - gráfico simple + tabla de respaldo

Introducción a ggplot2

¿Qué es ggplot2?

- **ggplot2** es un paquete de R para crear visualizaciones de datos
- Basado en la "Gramática de Gráficos" (*Grammar of Graphics*)
- Creado por Hadley Wickham
- Parte del ecosistema **tidyverse**



La Gramática de Gráficos

Un gráfico es una combinación de capas independientes:

1. **Datos** (data)
2. **Estéticas** (aesthetics - aes)
3. **Geometrías** (geoms)
4. **Estadísticas** (stats)
5. **Escalas** (scales)
6. **Coordenadas** (coordinates)
7. **Facetas** (facets)
8. **Tema** (theme)

Componentes Esenciales de un Gráfico

Los 3 componentes básicos:

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

1. **Datos:** El dataset que queremos visualizar
2. **Estéticas (aes):** Cómo se mapean las variables a propiedades visuales
3. **Geometrías (geom):** El tipo de gráfico (puntos, líneas, barras, etc.)

Componentes Esenciales: Datos

Los datos deben estar en formato **tidy** (ordenado):

- Cada variable es una columna
- Cada observación es una fila
- Cada valor es una celda

```
# Usando el dataset incorporado 'mpg'  
head(mpg, 3)
```

```
## # A tibble: 3 × 11  
##   manufacturer model displ  year   cyl trans      drv    cty   hwy fl      class  
##   <chr>         <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>  
## 1 audi         a4      1.8  1999     4 auto(l5) f       18    29 p      compact  
## 2 audi         a4      1.8  1999     4 manual(m5) f       21    29 p      compact  
## 3 audi         a4      2    2008     4 manual(m6) f       20    31 p      compact
```

Componentes Esenciales: Estéticas (aes)

Las estéticas mapean variables a propiedades visuales:

- Posición: `x`, `y`
- Color: `color`, `fill`
- Forma: `shape`
- Tamaño: `size`
- Transparencia: `alpha`
- Tipo de línea: `linetype`

```
aes(x = displ, y = hwy, color = class, size = cyl)
```

Componentes Esenciales: Geometrías

Las geometrías definen el tipo de gráfico:

Geometría	Tipo de gráfico
<code>geom_point()</code>	Gráfico de dispersión
<code>geom_line()</code>	Gráfico de líneas
<code>geom_bar()</code>	Gráfico de barras
<code>geom_histogram()</code>	Histograma
<code>geom_boxplot()</code>	Diagrama de caja
<code>geom_smooth()</code>	Línea de tendencia

