

# Visualização de Dados e o ggplot2

## A Gramática por Trás dos Gráficos

Revisão de Análise Exploratória e Introdução ao ggplot2 em R

Lineu Alberto Cavazani de Freitas

Pró-Reitoria de Planejamento e Dados (PROPLAD)  
Coordenadoria de Estatística e Ciência de Dados (CECD)





# **Análise Exploratória de Dados (AED)**

# Análise Exploratória de Dados (AED)

- ▶ É a primeira e mais importante etapa em qualquer projeto de dados.
- ▶ É um processo de investigação que usa tabelas, gráficos e resumos numéricos.
- ▶ O instrumento usado está diretamente associado ao tipo de variável sob análise.

## Análise Univariada:

- ▶ Uso de tabelas de frequência.
- ▶ Visualização (Barras, Histogramas, Densidade, Boxplots).
- ▶ Medidas Resumo (Posição Central, Posição Relativa, Dispersão).

## Análise Bivariada (Associação):

- ▶ Variáveis Qualitativas: Tabelas de Contingência e Gráficos para representar frequências.
- ▶ Variáveis Quantitativas: Diagrama de Dispersão e Correlação.
- ▶ Uma Qualitativa e uma Quantitativa: Medidas resumo para os possíveis níveis e gráficos.

# Considerações Essenciais da AED

- ▶ A análise exploratória permite sumarizar o comportamento de variáveis, compreender padrões e avaliar associações.
- ▶ Defesa Contra Baixa Qualidade: A AED identifica falhas e inconsistências que podem comprometer análises futuras.
- ▶ Diferença de Propósito:
  - ▶ Análise exploratória para si: Investigação, desordenada, foco em descobrir padrões.
  - ▶ Análise exploratória para os outros: Focada no público, clareza e objetividade. Apresente apenas os insights mais relevantes. O público não está interessado no quão duro você trabalhou para obter aqueles resultados.



# Visualização de dados

# O Conceito de Visualização de Dados

- ▶ Definição: Comunicação eficaz e visual de dados.
- ▶ Excelência Gráfica (Edward Tufte): Oferecer o maior número de ideias, no menor tempo, com menos tinta, no menor espaço.
- ▶ Baseada em Ciência: Boas visualizações respeitam a percepção visual e a cognição humana.

# Princípios de Design Visual

## Princípios Gestalt

- ▶ Determinantes de como o espectador agrupa e percebe elementos.
- ▶ Focam a atenção.
- ▶ Proximidade, Similaridade, Continuidade, Fechamento.

## Elementos Visuais

- ▶ Marcas: As formas básicas (pontos, linhas, áreas).
- ▶ Canais: As propriedades que codificam variáveis (cor, tamanho, posição, forma).

# Visualização de Dados no R

Quatro Frentes Principais:

- ▶ `graphics` (pacote primitivo/básico).
- ▶ `lattice` (para gráficos em painel/mútiplos).
- ▶ `ggplot2` (Nosso Foco Principal).
- ▶ Pacotes Interativos como `plotly`, `leaflet`, dentre outros.

Destaque do `ggplot2`: Baseado na **Gramática dos Gráficos**, garantindo um modelo mental claro e coerente.





# A Gramática dos Gráficos (GoG)

# A Gramática dos Gráficos (GoG)

- ▶ A Gramática dos Gráficos, proposta por Leland Wilkinson, propõe uma abordagem teórica para a criação de gráficos estatísticos.
- ▶ Estrutura fundamental para a visualização de dados e influenciou o desenvolvimento de diversas ferramentas e bibliotecas, como o pacote ggplot2 no R, popularizado por Hadley Wickham.
- ▶ Conceito Central: Qualquer gráfico pode ser decomposto em componentes independentes e ortogonais.
- ▶ Filosofia: A visualização é construída em camadas (layers).
- ▶ Benefício: Criação de gráficos de forma lógica, modular e sequencial.

# Componentes Chave da GoG (e do ggplot2)

- ▶ **Dados:** A fonte de informação (o data frame).
- ▶ **Mapeamentos Estéticos:** Mapear variáveis a atributos visuais (aes).
- ▶ **Geometrias:** As formas visuais (geom\_).
- ▶ **Transformações Estatísticas:** Cálculos que transformam dados antes de plotar (stat\_).
- ▶ **Facetas:** Dividir o gráfico em subpainéis (facet\_).
- ▶ **Sistemas de Coordenadas:** O espaço de exibição (cartesiano, polar, mapa).

# O ggplot2

- ▶ O Pacote: Criado por Hadley Wickham, principal implementação da GoG em R.
- ▶ Sintaxe Aditiva: Gráficos construídos camada por camada usando o operador +.

**`ggplot(data)+geom_(aes(variáveis)) + camadas adicionais...`**

# Componentes Essenciais do ggplot2

1. **Dados:** O conjunto de dados inicial.
2. **Mapeamentos Estéticos** (`aes()`): Tornam os dados visíveis.
  - ▶ Posição: `x`, `y`.
  - ▶ Cor/Preenchimento: `color` (linha/contorno), `fill` (interior).
  - ▶ Aparência: `size`, `shape`, `linetype`, `alpha` (transparência).
3. **Geometrias** (`geom_()`): As formas desenhadas.
  - ▶ Exs.: `geom_point()`, `geom_line()`, `geom_bar()`, `geom_boxplot()`, `geom_histogram()`, `geom_smooth()`.

# Componentes Essenciais do ggplot2

4. **Facetas**: Dividem a figura em múltiplos painéis para comparação.
  - ▶ `facet_wrap()` (por uma variável), `facet_grid()` (por duas variáveis).
5. **Coordenadas** (`coord_`): Define o sistema de projeção.
  - ▶ Exs.: `coord_cartesian()`, `coord_polar()`.
6. **Temas** (`theme_`): Ajustam o visual não-dado do gráfico.
  - ▶ Fontes, cores de fundo, tamanhos de rótulos, etc.



# Alguns gráficos feitos com ggplot2

# Próximos Passos (Prática)

- ▶ Objetivo: Construir nossos primeiros gráficos seguindo a estrutura modular.
- ▶ Começar com a estrutura mínima: `ggplot(data) + ...`
- ▶ Adicionar camadas gradualmente (cores, temas, rótulos) para demonstrar a flexibilidade do GoG.