

MÓDULO 2 - AULA 2

Formas de Visualização de Dados

Gabarito - Resultados Esperados

ATIVIDADE 1: Quarteto de Anscombe

1.2 Estatísticas por conjunto:

Conjunto	Média X	Média Y	DP X	DP Y	Correlação
Conjunto 1	9.0	7.50	3.32	2.03	0.82
Conjunto 2	9.0	7.50	3.32	2.03	0.82
Conjunto 3	9.0	7.50	3.32	2.03	0.82
Conjunto 4	9.0	7.50	3.32	2.03	0.82

Observe: Todos os conjuntos têm estatísticas praticamente idênticas!

1.3 Interpretação dos gráficos:

- **Conjunto 1:** Relação linear 'normal' - os pontos seguem a linha de regressão
- **Conjunto 2:** Relação não-linear (curva parabólica) - a reta não captura o padrão
- **Conjunto 3:** Relação linear com um outlier que puxa a reta
- **Conjunto 4:** Sem relação real - um único outlier define toda a correlação

ATIVIDADE 2: Datasaurus Dozen

2.1 Datasets disponíveis (13 conjuntos):

away, bullseye, circle, dino, dots, h_lines, high_lines, slant_down, slant_up, star, v_lines, wide_lines, x_shape

2.2 Estatísticas (todos os datasets):

Média X ≈ 54.26 | Média Y ≈ 47.83 | DP X ≈ 16.77 | DP Y ≈ 26.94 | Correlação ≈ -0.06

Todos os 13 conjuntos têm as mesmas estatísticas, mas formam figuras completamente diferentes!

2.4 O dinossauro:

O dataset 'dino' forma a silhueta de um T-Rex quando visualizado como gráfico de dispersão, demonstrando que dados podem 'esconder' padrões surpreendentes.

ATIVIDADE 3: Boas Práticas na Visualização

Dados do IPCA:

Ano	2009	2010	2011	2012	2013

IPCA (%)	4.31	5.92	6.50	5.84	5.91
----------	------	------	------	------	------

Regra: Em gráficos de barras, o eixo Y SEMPRE deve começar em zero para permitir comparação proporcional correta.

ATIVIDADE 4: Uso Adequado de Cores

Dados de eventos cardiovasculares após dengue (IRR):

Desfecho	Dias 1-7	Dias 8-14
Insuficiência cardíaca	27.24	2.45
AVC isquêmico	15.56	3.17
Infarto agudo	13.53	1.16
AVC hemorrágico	10.90	4.33

Interpretação: IRR > 1 indica aumento de risco. O risco é muito maior nos primeiros 7 dias após a infecção.

ATIVIDADE 5: Tipos de Gráficos Adequados

5.1 Pizza vs Barras:

- Grupo A: 45% | Grupo B: 35% | Grupo C: 20%
- **Recomendação:** Barras horizontais facilitam a comparação - humanos são ruins em comparar ângulos!

5.2 Série temporal:

O gráfico de linha mostra casos de Influenza por semana epidemiológica (2023), evidenciando a sazonalidade da doença com pico aproximadamente na semana 14.

ATIVIDADE 6: Contextualização Completa

Elementos de um bom gráfico:

- ✓ **Título:** 'Distribuição do peso ao nascer'
- ✓ **Subtítulo:** 'Amostra de 20 recém-nascidos de uma maternidade de São Paulo'
- ✓ **Eixo X:** 'Peso ao nascer (gramas)' - com unidade!
- ✓ **Eixo Y:** 'Frequência (n)'
- ✓ **Anotação:** 'Média: 3167 g' indicando a linha de referência
- ✓ **Caption:** 'Fonte: Dados simulados para fins didáticos'