

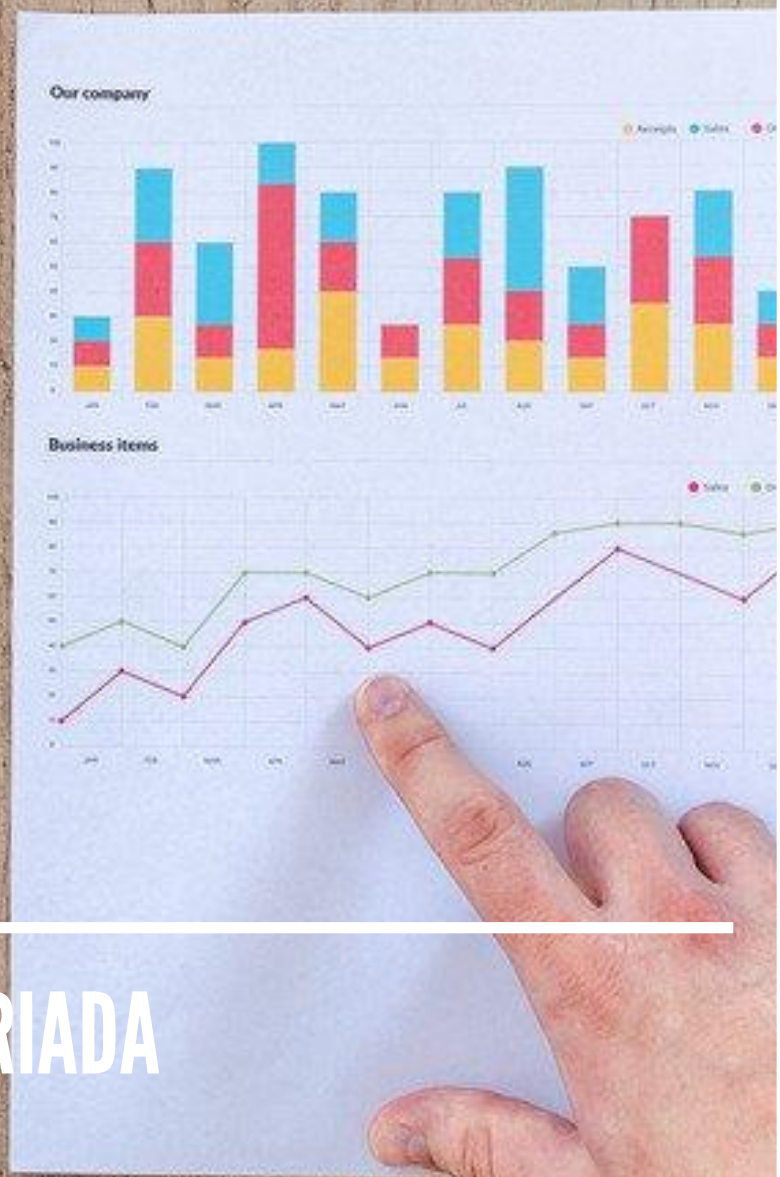
BIOLOGIA/BIOMEDICINA

---

# BIOESTATÍSTICA

Prof<sup>a</sup>. Letícia Raposo  
profleticiaraposo@gmail.com

# ESTADÍSTICA DESCRIPTIVA BIVARIADA



## OBJETIVOS DA AULA

- Compreender os principais conceitos de estatística descritiva bivariada;
  - Escolher o(s) método(s) adequado(s), incluindo tabelas, gráficos e/ou medidas-resumo, para descrever o comportamento das variáveis;
  - Estudar as associações entre duas variáveis qualitativas por meio de tabelas de contingência e medidas de associação;
- 



## OBJETIVOS DA AULA

- Estudar as correlações entre duas variáveis quantitativas por meio de tabelas de distribuição conjunta de frequências, gráficos e medidas de correlação;
- Gerar tabelas, gráficos e medidas-resumo por meio do R.



# ESTATÍSTICA DESCRITIVA

---

**Univariada**

**Estuda uma única variável**

**Bivariada**

**Estuda duas variáveis**

**Multivariada**

**Estuda mais de duas variáveis**

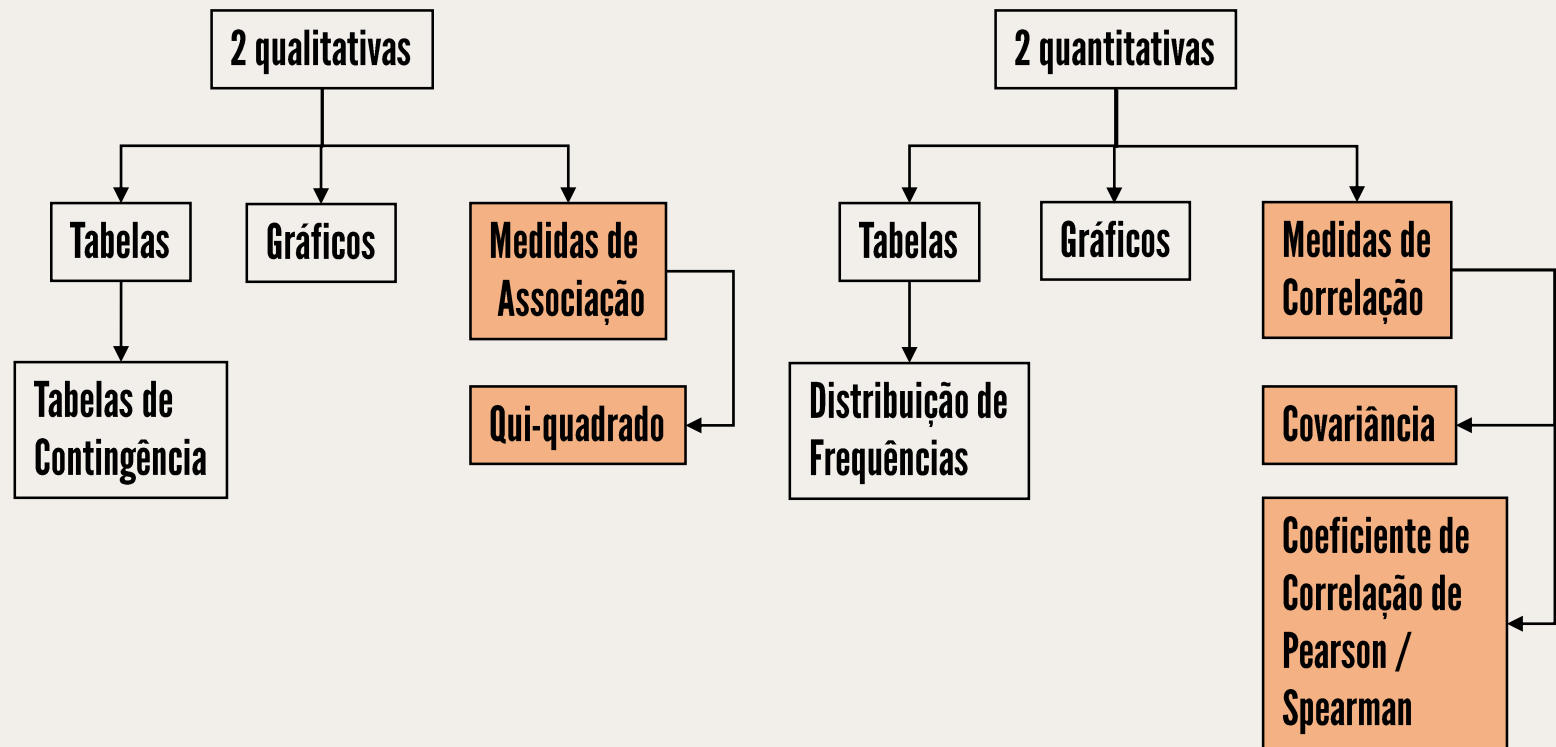
A análise bivariada tem como objetivo estudar as relações (associações para variáveis qualitativas e correlações para variáveis quantitativas) entre duas variáveis.

**Distribuição conjunto de frequências  
(tabelas de contingência ou de  
classificação cruzada)**

**Representações gráficas**

**Medidas-resumo**

# ANÁLISE BIVARIADA



\*Serão vistas (com mais detalhes) mais a frente no curso.



# ASSOCIAÇÃO ENTRE DUAS VARIÁVEIS QUALITATIVAS



# TABELAS DE DISTRIBUIÇÃO CONJUNTA DE FREQUÊNCIAS

*Tabelas de contingência ou tabelas de dupla entrada*

Dados		
Família	Nível de instrução	Uso de programas*
1	Nenhum	Não
2	Segundo grau	Não
3	Primeiro grau	Sim
4	Primeiro grau	Sim
5	Segundo grau	Sim
⋮	⋮	⋮

Uso de programas	Nível de instrução		
	Nenhum	Primeiro grau	Segundo grau
Sim			
Não			

\* Programas alimentares



# TABELAS DE DISTRIBUIÇÃO CONJUNTA DE FREQUÊNCIAS

---

Uso de programas	Nível de instrução do chefe da casa			Total
	Nenhum	Fundamental	Médio	
Sim	31	22	25	78
Não	7	16	19	42
Total	38	38	44	120

# TABELAS DE DISTRIBUIÇÃO CONJUNTA DE FREQUÊNCIAS

Uso de programas	Nível de instrução do chefe da casa			Total
	nenhum	fundamental	médio	
Sim	31 (39,7)	22 (28,2)	25 (32,1)	78 (100,0)
Não	7 (16,7)	16 (38,1)	19 (45,2)	42 (100,0)
Total	38 (31,7)	38 (31,7)	44 (36,7)	120 (100,0)

Nota: os números entre parênteses são porcentagens em relação aos totais das linhas.

Uso de programas	Nível de instrução do chefe da casa			Total
	nenhum	fundamental	médio	
Sim	31 (81,6)	22 (57,9)	25 (56,8)	78 (65,0)
Não	7 (18,4)	16 (42,1)	19 (43,2)	42 (35,0)
Total	38 (100,0)	38 (100,0)	44 (100,0)	120 (100)

Nota: os números entre parênteses são porcentagens em relação aos totais das colunas.

PERFIL LINHA

PERFIL COLUMNA

# REPRESENTAÇÕES GRÁFICAS

---

GRÁFICO DE BARRAS  
MÚLTIPLAS



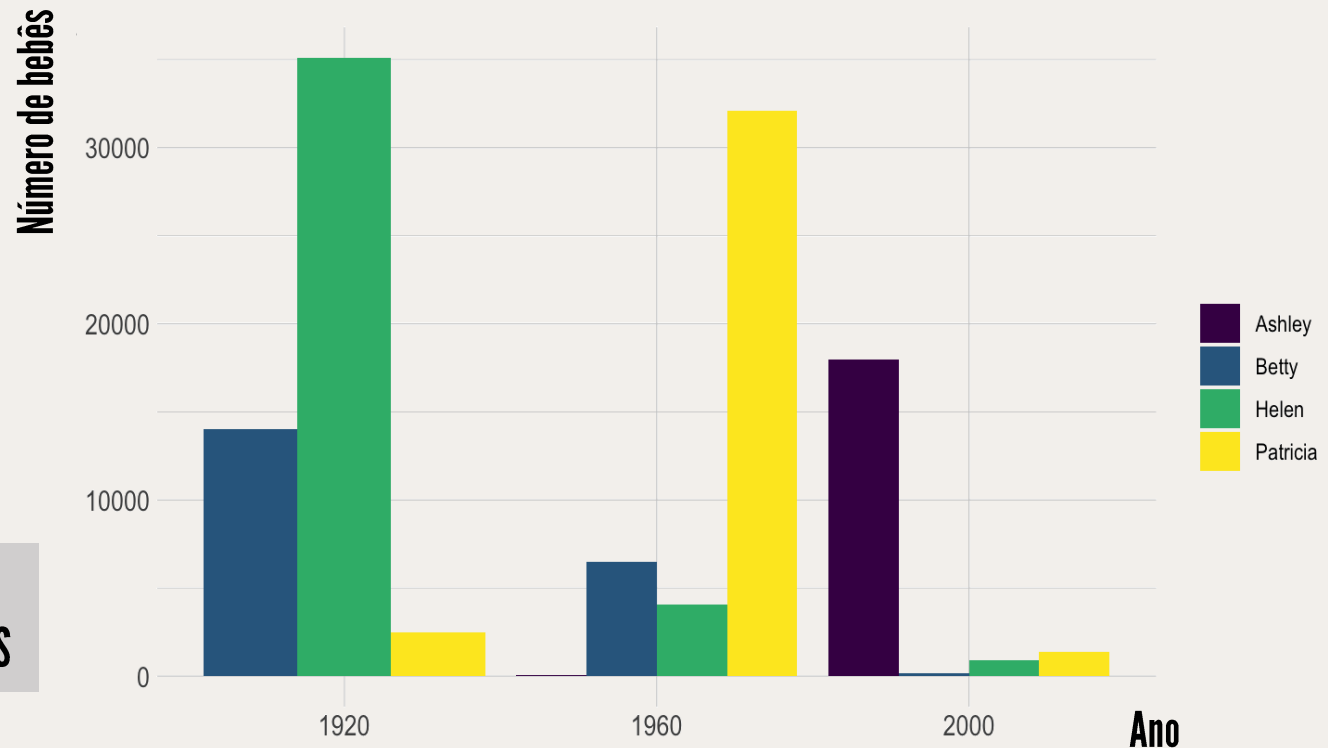
GRÁFICO DE  
BARRAS EMPILHADAS



# GRÁFICO DE BARRAS MÚLTIPLAS



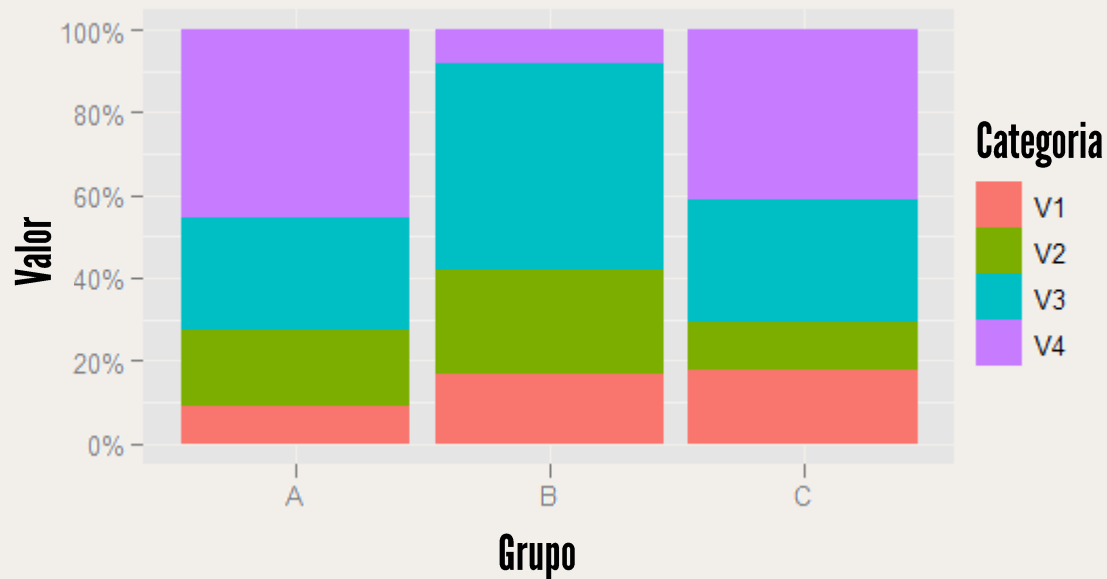
Representam mais de uma distribuição de frequências, ou distribuições de frequências conjuntas de duas variáveis qualitativas.



**ÚTIL PARA COMPARAR  
SUBGRUPOS DE GRUPOS**

<https://www.data-to-viz.com/graph/barplot.html>

# GRÁFICO DE BARRAS EMPILHADAS



ÚTIL PARA ESTUDAR A EVOLUÇÃO DOS SUBGRUPOS



# ESTATÍSTICA QUI-QUADRADO

---

- Mede a discrepância entre uma tabela de contingência observada e uma tabela de contingência esperada, partindo da hipótese de que não há associação entre as variáveis estudadas.
- Se a distribuição de frequências observadas for exatamente igual à distribuição de frequências esperadas, o resultado da estatística qui-quadrado é zero.
- Um valor baixo de  $\chi^2$  indica independência entre as variáveis.

# ESTATÍSTICA QUI-QUADRADO

---

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- $O_{ij}$ : quantidade de observações na  $i$ -ésima categoria da variável X e na  $j$ -ésima categoria da variável Y;
- $E_{ij}$ : frequência esperada de observações na  $i$ -ésima categoria da variável X e na  $j$ -ésima categoria da variável Y;
- $I$ : quantidade de categorias (linhas) da variável X;
- $J$ : quantidade de categorias (colunas) da variável Y.

$$O_{ij} = 51$$

$$E_{ij} = \frac{56 \times 80}{100} = 44,8$$

Gênero X Acidente de carro

	Sem acidente	Acidente	Total
Mulheres	51	5	56
Homens	29	15	44
Total	80	20	100

# ASSOCIAÇÃO ENTRE DUAS VARIÁVEIS QUANTITATIVAS



# ASSOCIAÇÃO ENTRE DUAS VARIÁVEIS QUANTITATIVAS

---

Avaliar se existe relação entre as variáveis quantitativas estudadas, além do grau de correlação entre elas.

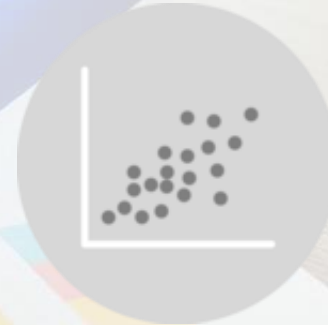
Tabela de distribuição conjunta de frequências: mesmo procedimento das variáveis qualitativas.

- Variáveis discretas;
- Variáveis contínuas agrupadas em intervalos de classe.

# REPRESENTAÇÕES GRÁFICAS

---

GRÁFICO DE DISPERSÃO





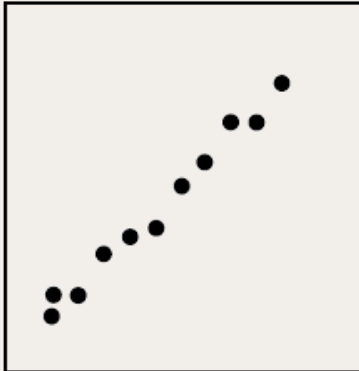
# GRÁFICO DE PONTOS OU DISPERSÃO

---



- Representa os valores das variáveis X e Y em um plano cartesiano.
- Permite avaliar:
  - Se existe ou não alguma relação entre as variáveis em estudo;
  - O tipo de relação entre as duas variáveis, isto é, a direção em que a variável Y aumenta ou diminui em função da variável de X;
  - O grau de relação entre as variáveis;
  - A natureza da relação (linear, exponencial, etc).

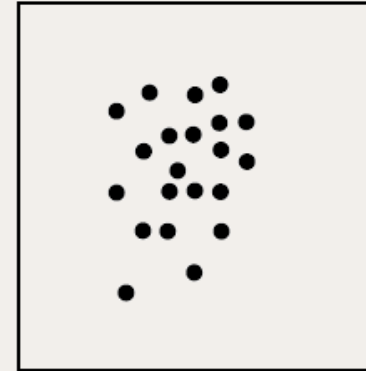
# GRÁFICO DE PONTOS OU DISPERSÃO



**Forte correlação positiva**



**Moderada correlação positiva**



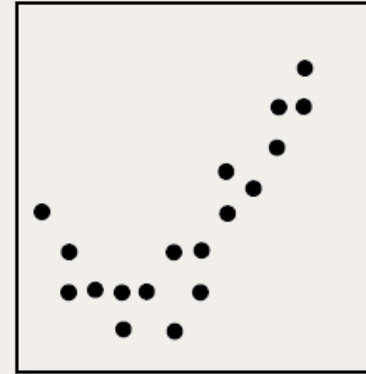
**Nenhuma correlação**



**Moderada correlação negativa**



**Forte correlação negativa**

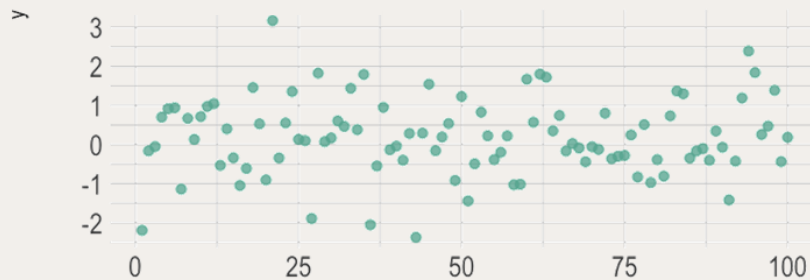


**Correlação curvilínea**

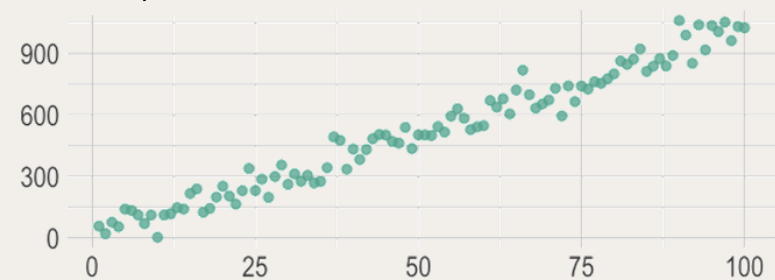
# GRÁFICO DE PONTOS OU DISPERSÃO



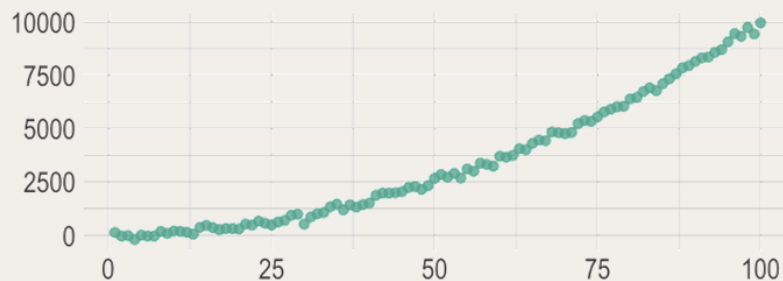
Nenhuma tendência



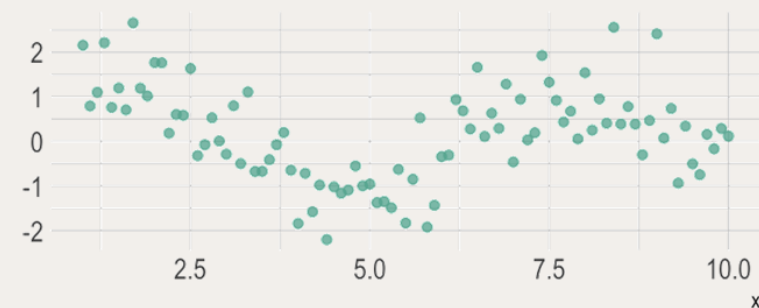
Relação linear



Quadrática



Seno



<https://www.data-to-viz.com/graph/scatter.html>

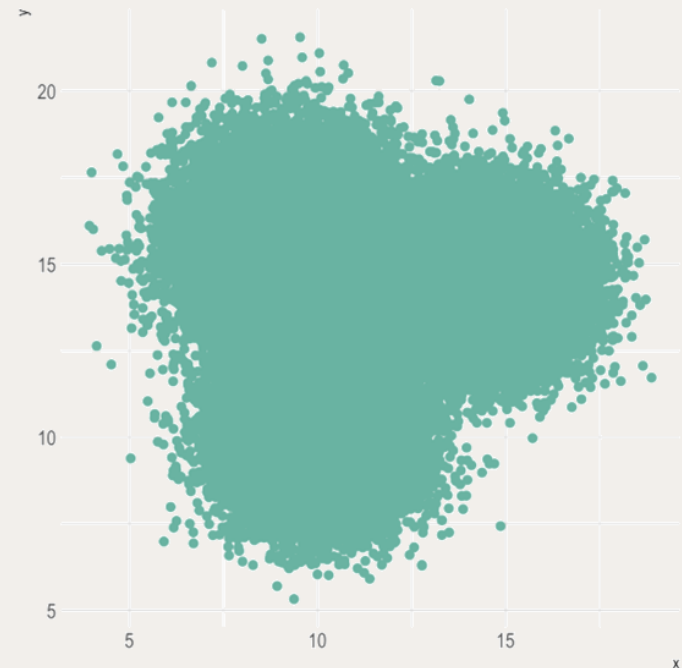
# GRÁFICO DE PONTOS OU DISPERSÃO



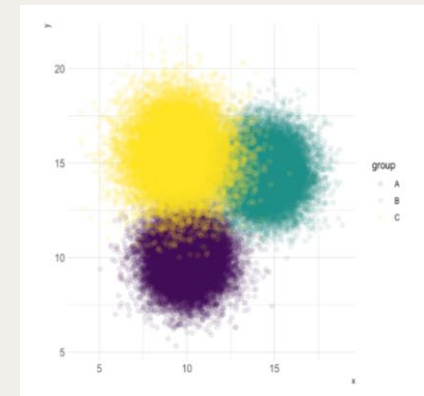
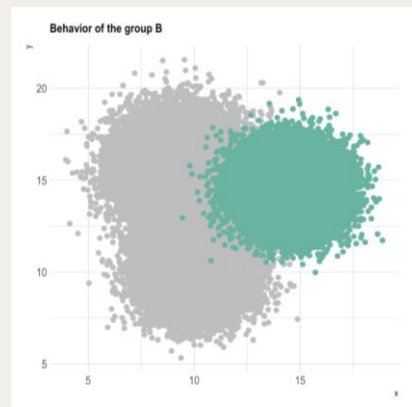
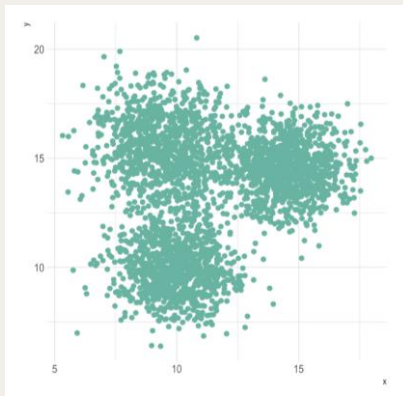
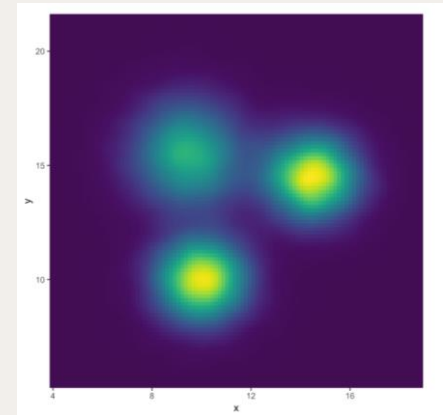
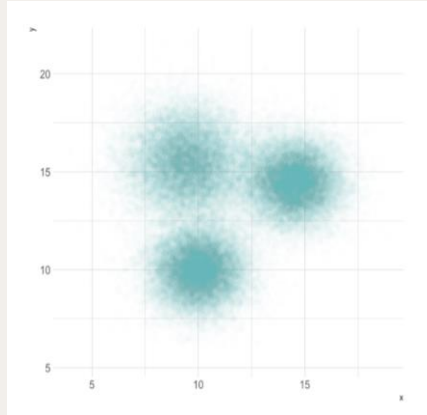
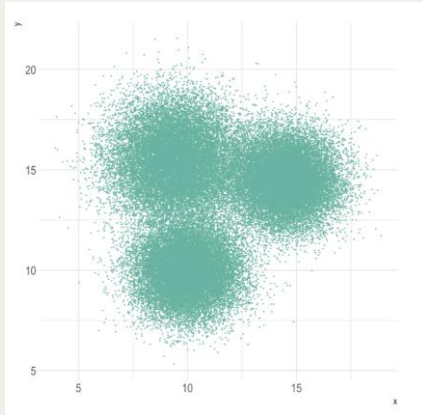
## Dicas:

Evite o *overplotting*;

- Reduza o tamanho dos pontos;
- Usar transparência;
- Densidade 2D;
- Amostrar apenas 5% dos dados;
- Destacar algum grupo;
- Colorir os grupos.



# GRÁFICO DE PONTOS OU DISPERSÃO



<https://www.data-to-viz.com/caveat/overplotting.html>

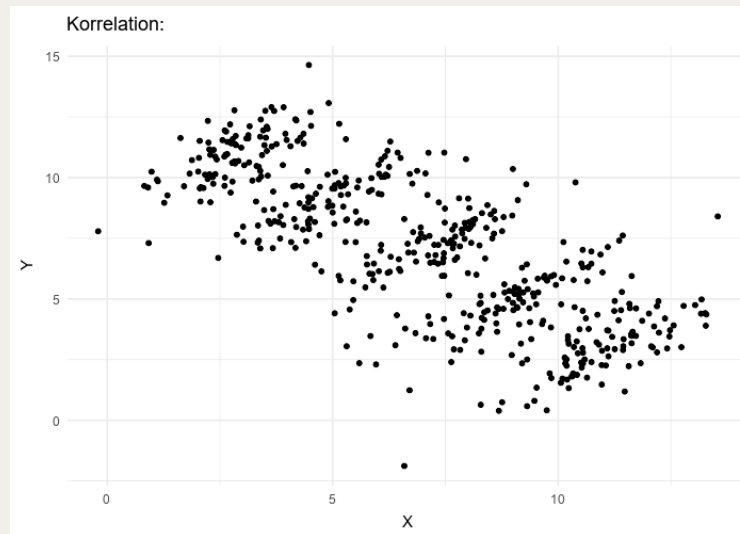


# GRÁFICO DE PONTOS OU DISPERSÃO



## Dicas:

Não se esqueça de mostrar subgrupos se existirem. De fato, eles podem revelar padrões ocultos importantes em seus dados, como no caso do paradoxo de Simpson.



# MEDIDAS DE CORRELAÇÃO

---

Covariância: mede a variação conjunta entre duas variáveis quantitativas X e Y.

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

$X_i$ : i-ésimo valor de X;

$Y_i$ : i-ésimo valor de Y;

$\bar{X}$ : média dos valores de  $X_i$ ;

$\bar{Y}$ : média dos valores de  $Y_i$ ;

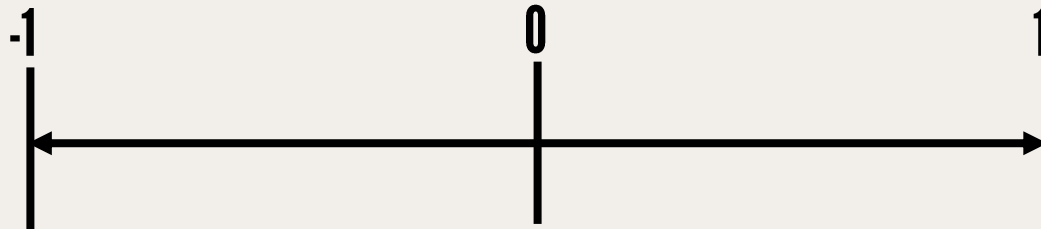
$n$  : tamanho da amostra.

**A DEFICIÊNCIA DA COVARIÂNCIA É QUE SEU VALOR CALCULADO DEPENDE DIRETAMENTE DAS UNIDADES DE MEDIDA.**

# COEFICIENTE DE CORRELAÇÃO DE PEARSON

$$\rho = \frac{cov(X, Y)}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1 S_X S_Y}$$

SEU VALOR É INDEPENDENTE DA  
UNIDADE MEDIDA.



Correlação linear  
negativa perfeita entre  
as variáveis X e Y

Não existe correlação  
linear entre as  
variáveis X e Y

Correlação linear  
positiva perfeita entre  
as variáveis X e Y

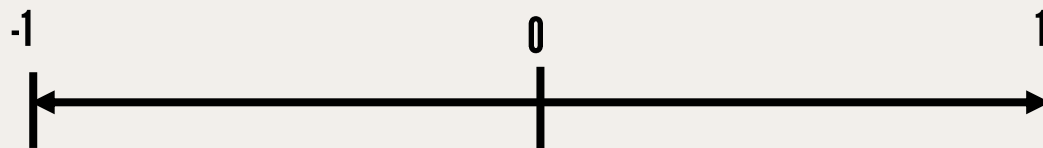
# COEFICIENTE DE CORRELAÇÃO DE POSTOS DE SPEARMAN

Indicado quando:

- Os dados não formam uma nuvem comportada, com alguns pontos bem distantes dos demais;
- Parece existir uma relação crescente ou decrescente num formato de curva;
- Existe uma ordenação clara, por exemplo, escores numa escala de 1 a 20.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{(n^3 - n)}$$

$d_i = (\text{posto de } x_i \text{ dentre os valores de } x) - (\text{posto de } y_i \text{ nos valores de } y)$



**Correlação negativa perfeita  
entre as variáveis X e Y**

**Não existe correlação entre  
as variáveis X e Y**

**Correlação positiva perfeita  
entre as variáveis X e Y**

# ASSOCIAÇÃO ENTRE VARIÁVEIS QUALITATIVAS E QUANTITATIVAS





# ASSOCIAÇÃO ENTRE VARIÁVEIS QUALITATIVAS E QUANTITATIVAS

---

- É comum analisar o que acontece com a variável quantitativa dentro de cada categoria da variável qualitativa.
- As medidas-resumo podem ser calculadas para a variável quantitativa em cada categoria da variável qualitativa.

# REPRESENTAÇÕES GRÁFICAS

---

HISTOGRAMA



GRÁFICO DE DENSIDADES



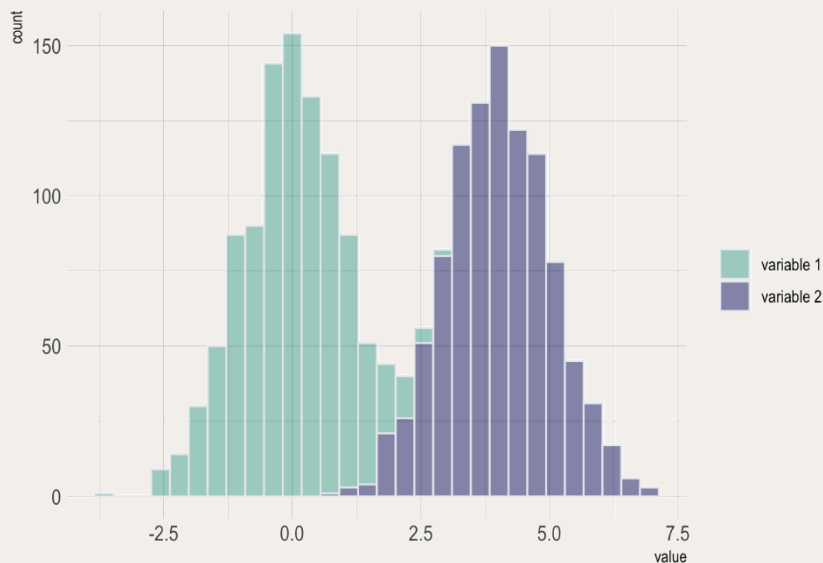
BOXPLOT



# HISTOGRAMA

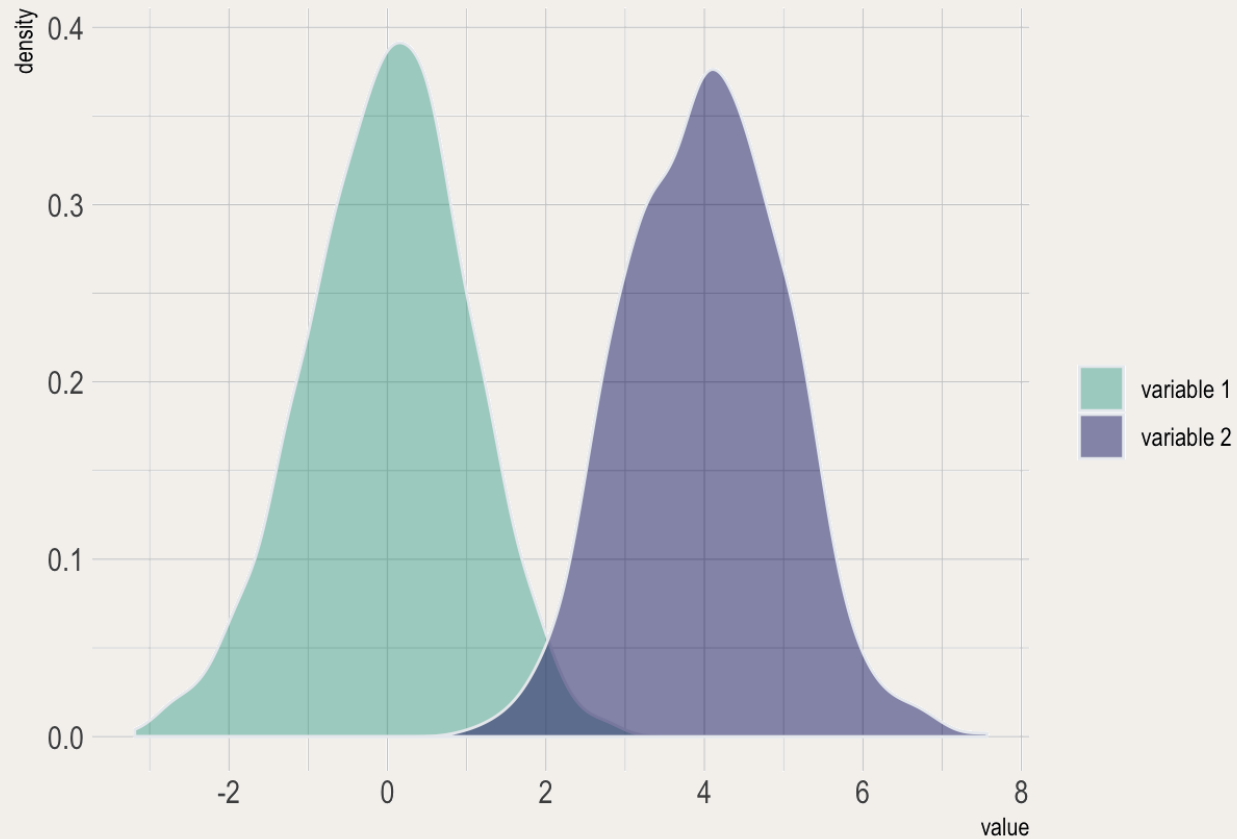


- O histograma permite comparar a distribuição de algumas variáveis.
- Não compare mais de 3 ou 4, isso tornaria a figura desordenada e ilegível. Essa comparação pode ser feita mostrando as duas variáveis no mesmo gráfico.



<https://www.data-to-viz.com/graph/histogram.html>

# GRÁFICO DE DENSIDADES



<https://www.data-to-viz.com/graph/density.html>

# BOXPLOT



Pode resumir a distribuição de uma variável numérica para vários grupos. O problema é que resumir também significa perder informação, e isso pode ser uma armadilha.

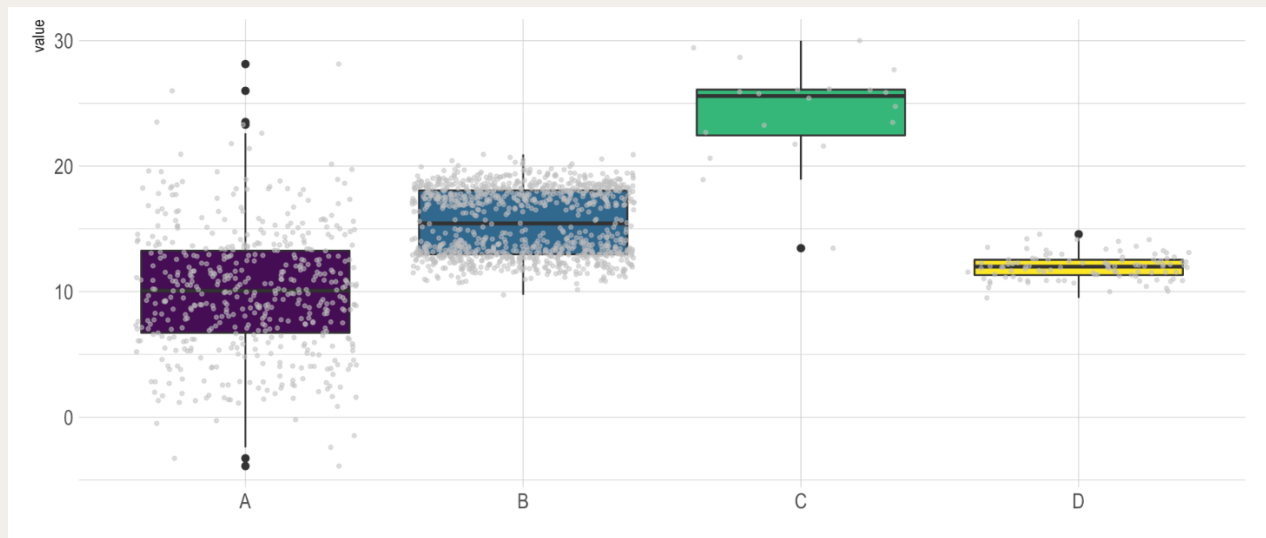


É fácil concluir que o grupo C tem um valor maior que os outros. No entanto, não podemos ver a distribuição de pontos em cada grupo ou seu número de observações.

# BOXPLOT



Se a quantidade de dados com a qual você está trabalhando não for muito grande, a adição de tremulação (“*jitter*”) em cima do *boxplot* poderá tornar o gráfico mais interessante.



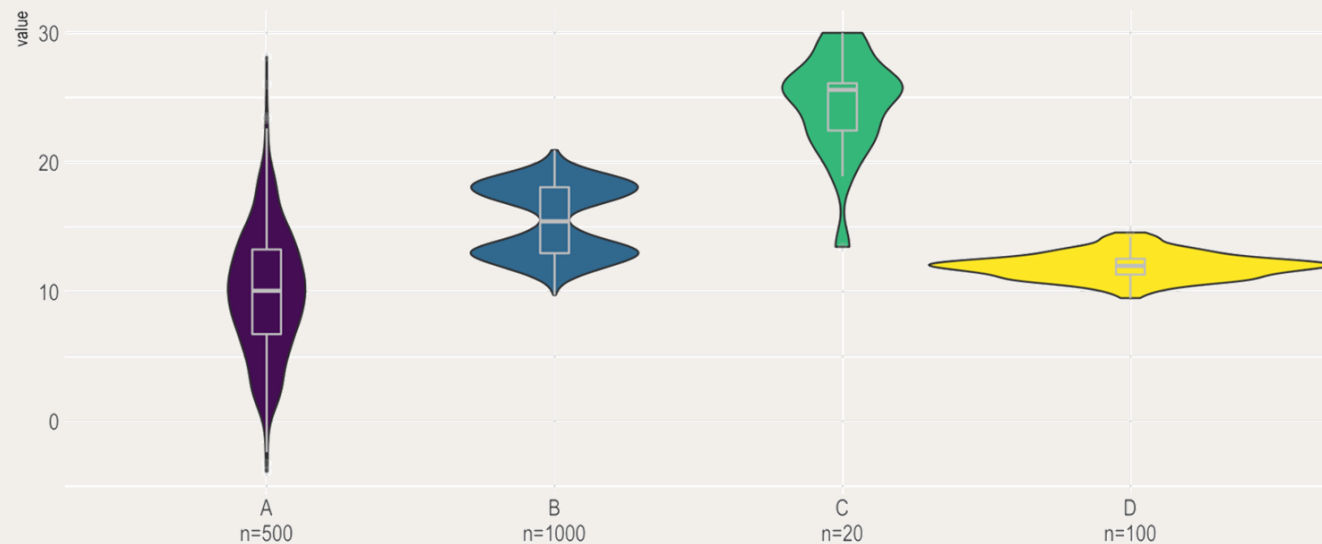
**O grupo C tem uma amostra pequena comparada aos outros grupos. Além disso, parece que o grupo B tem uma distribuição bimodal: os pontos são distribuídos em dois grupos: em torno de  $y = 18$  e  $y = 13$ .**

# GRÁFICO DE VIOLINO



Quando a amostra é grande, usar o “jitter” não é mais uma opção, pois os pontos se sobrepõem, tornando a figura não-interpretável. Uma alternativa é o gráfico do violino, que descreve a distribuição dos dados para cada grupo.

**A distribuição bimodal do grupo B torna-se óbvia.**



**Maneira poderosa de exibir informações, porém são subutilizadas em comparação com os *boxplots*.**

# SITES LEGAIS PARA GRÁFICOS

---



THE R GRAPH  
GALLERY

<https://www.r-graph-gallery.com/>

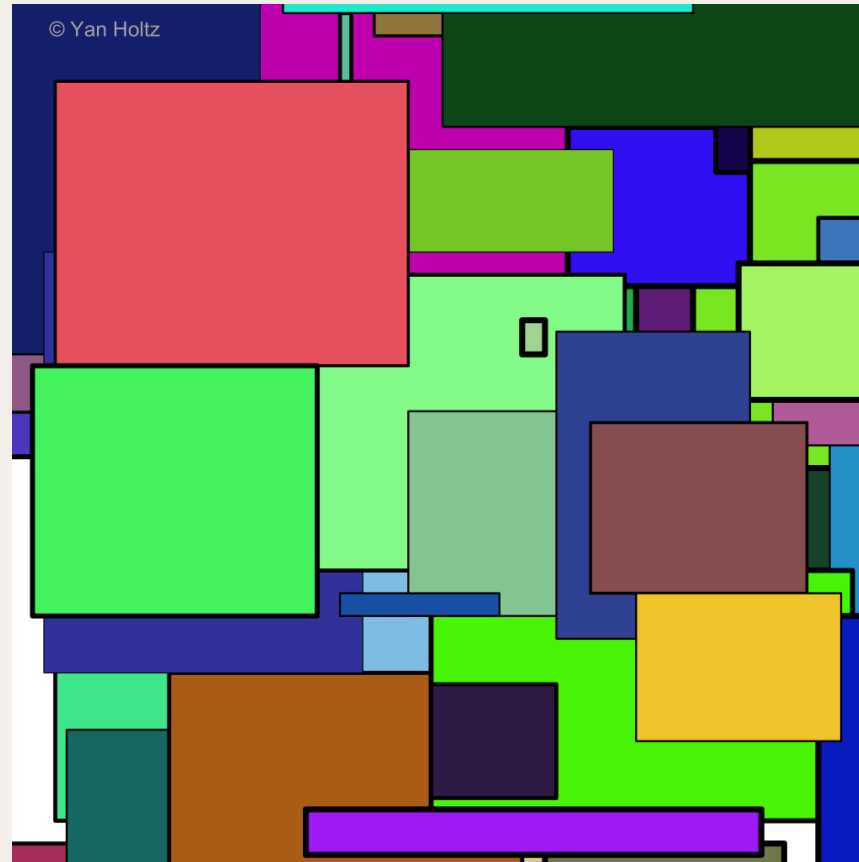


from Data to Viz

<https://www.data-to-viz.com/index.html>



# ARTE DO DIA FEITA EM R



<https://www.r-graph-gallery.com/portfolio/data-art/>

# REFERÊNCIAS BIBLIOGRÁFICAS

---

- BARBETTA, Pedro Alberto. Estatística aplicada às ciências sociais. Ed. UFSC, 2008.
- DANCEY, Christine P.; REIDY, John G.; ROWE, Richard. Estatística Sem Matemática para as Ciências da Saúde. Penso Editora, 2017.
- MAGNUSSON, Willian E. Estatística [sem] matemática: a ligação entre as questões e a análise. Planta, 2003.

