

$$S = \sum_{i=1}^n (y_i - \bar{y})^2$$

# *Entendendo Estatística Divertidamente*

*Profa. Adriana Silva*

Seja bem vindX!!!

Câmera ligada e

Microfone mutado sempre  
que não estiver falando

# Estatística Descritiva

## Tipos de Medidas

Estatisticamente existem inúmeras medidas que nos norteiam sobre nossas análises. Estas medidas são divididas em classes:

- Medidas de Posição ✓
- Medidas de Dispersão
- Análise Gráfica
- Medidas de Assimetria
- Medidas de Associação

# Estatística Descritiva

## Medidas de Dispersão

- Amplitude

A amplitude é definida como sendo a diferença entre o maior e o menor valor do conjunto de dados. A Amplitude também é chamada de Range.

Mede a dispersão total no conjunto de dados.

É uma medida simples que não leva em consideração como os dados são efetivamente distribuídos entre os valores extremos

$$A = X_{maior} - X_{menor}$$

# Estatística Descritiva

## Medidas de Dispersão

- Amplitude

Salário
R\$600.00
R\$1,300.00
R\$1,300.00
R\$3,000.00
R\$4,500.00
R\$6,000.00
R\$15,000.00
R\$30,000.00
R\$50,000.00
R\$100,000.00



$$A = X_{maior} - X_{menor}$$

$$A = 100.000,00 - 600,00 = R\$ 99.400,00$$



# Estatística Descritiva

## Medidas de Dispersão

- Variância

A variância da amostra é a média aproximada das diferenças ao quadrado entre cada uma das observações e a média aritmética da amostra.

É a diferença entre o valor de cada observação e a média dividido pelo tamanho amostral menos 1.

$$s^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

# Estatística Descritiva

## Medidas de Dispersão

- Variância

Salário
R\$600.00
R\$1,300.00
R\$1,300.00
R\$3,000.00
R\$4,500.00
R\$6,000.00
R\$15,000.00
R\$30,000.00
R\$50,000.00
R\$100,000.00

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n - 1}$$

$$S^2 = \frac{(600 - 21170)^2 + \dots + (100000 - 21170)^2}{9}$$

$$S^2 = 1023589000$$

# Estatística Descritiva

## Medidas de Dispersão

- Desvio Padrão

Desvio padrão é a raiz quadrada da variância.

O desvio padrão indica o afastamento dos valores observados em relação a média aritmética. É um conceito imprescindível para análises gráficas, determinação de confiabilidade e estudos de distribuições.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

# Estatística Descritiva

## Medidas de Dispersão

- Desvio Padrão

Salário
R\$600.00
R\$1,300.00
R\$1,300.00
R\$3,000.00
R\$4,500.00
R\$6,000.00
R\$15,000.00
R\$30,000.00
R\$50,000.00
R\$100,000.00

$$S^2 = 1023589000$$

$$S = \sqrt{1023589000} = R\$31.993,57$$



# Estatística Descritiva

## Medidas de Dispersão

- Coeficiente de Variação

Indica qual é o tamanho do Desvio Padrão em relação à média. Geralmente é expressa como uma porcentagem.

$$CV = \left( \frac{\text{Desvio Padrão}}{\text{Média}} \times 100 \right)$$

- No exemplo dos salários:

$$S = \sqrt{1023589000} = R\$ 31.993,57$$

$$\bar{x} = \frac{600 + 1300 + \dots + 100000}{10} = R\$ 21.170,00$$

$$cv = \frac{R\$ 31.993,57}{R\$ 21.170,00} \times 100 = 151,12\%$$

# Estatística Descritiva

## Medidas de Dispersão

- O Coeficiente de Variação nos ajuda entender se o Desvio Padrão calculado é pequeno ou grande.
- É interpretado como a variabilidade dos dados em relação à média.
- Quanto menor o CV mais homogêneo é a variável.
- Adimensional, isto é, um número puro, que será positivo se a média for positiva; será zero quando não houver variabilidade entre os dados

# Estatística Descritiva

## Medidas de Dispersão

- Os dados se dispersam em relação a média.
- Quanto esses dados estão espalhados?



- Peso médio 75kg



- Peso médio 75kg

# Estatística Descritiva

## Tipos de Medidas

Estatisticamente existem inúmeras medidas que nos norteiam sobre nossas análises. Estas medidas são divididas em classes:

- Medidas de Posição
- Medidas de Dispersão
- Análise Gráfica
- Medidas de Assimetria
- Medidas de Associação



Para que  
serve um  
gráfico?

# Estatística Descritiva

## Análise Gráfica

O gráfico é uma forma de apresentação dos dados, cujo objetivo é o de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão do que o conjunto de dados em si. O gráfico é um instrumento que possibilita transmitir muitas vezes o significado de planilhas ou tabelas complexas de uma forma mais eficiente e mais simples.

Uma representação gráfica tem por objetivo mostrar as relações que existem entre elementos dentro de um banco de dados.

Lembram da Florence Nightingale?

Florence dizia que o diagrama (“Diagrama da Rosa”) deveria atingir os olhos, coisa que um monte de palavras e números jamais conseguiriam.

# Estatística Descritiva

## Análise Gráfica

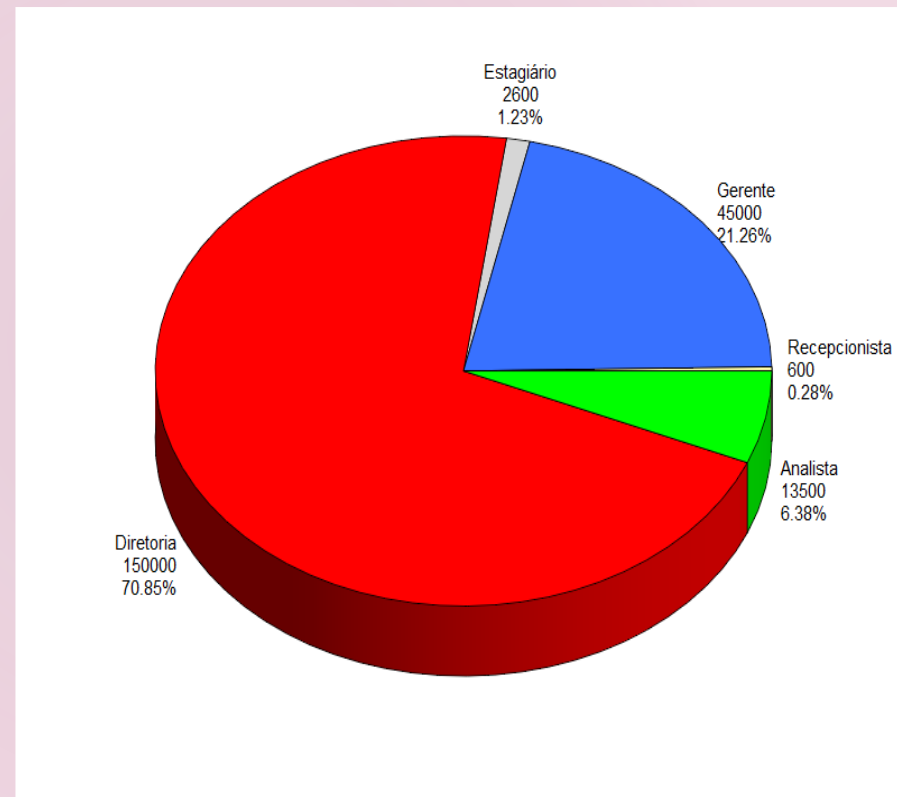
- Gráfico de Pizza
  - É um diagrama circular
  - Cada fatia representa as frequências (ou outra medida de interesse) de cada categoria
  - É usado para dados qualitativos nominais
  - Também conhecido como gráfico de setores ou gráfico circular.

# Estatística Descritiva

## Análise Gráfica

- Gráfico de Pizza

Cargo	Salário
Recepcionista	R\$600
Estagiário	R\$1.300
Estagiário	R\$1.300
Analista	R\$3.000
Analista	R\$4.500
Analista	R\$6.000
Gerente	R\$15.000
Gerente	R\$30.000
Diretoria	R\$50.000
Diretoria	R\$100.000





# Estatística Descritiva

## Análise Gráfica

cole nussbaumer knaflic

storytelling  
com  
dados

um guia sobre  
visualização de dados  
para profissionais  
de negócios



56 a escolha de um visual eficaz

### A ser evitado

Discutimos os visuais que mais uso para comunicar dados em um ambiente comercial. Existem ainda alguns tipos e elementos de gráfico específicos que devem ser evitados: gráficos de pizza, gráficos de rosca, 3D e eixos y secundários. Vamos discuti-los.

### Gráficos de pizza são ruins

Tenho desprezo bem documentado a gráficos de pizza. Resumindo, eles são ruins. Para entender como cheguei a essa conclusão, vamos ver um exemplo.

O gráfico de pizza da Figura 2.21 (baseado em um exemplo real) mostra a participação no mercado de quatro fornecedores: A, B, C e D. Se eu pedisse para você fazer uma observação simples — qual é o maior fornecedor, de acordo com este visual —, o que diria?

Participação no mercado de fornecedores

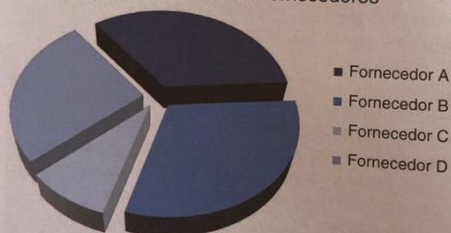


FIGURA 2.21 Gráfico de pizza

A maioria das pessoas concordará que o "Fornecedor B," representado em azul médio na parte inferior direita, parece ser o maior. Se você precisasse avaliar a proporção que o Fornecedor B ocupa no mercado global,

35%?

40%?

A partir de meu questionamento anterior, talvez você possa dizer que está acontecendo algo suspeito aqui. Dê uma olhada no que acontece quando acrescentamos os números às fatias da pizza, como mostrado na Figura 2.22.

Participação no mercado de fornecedores

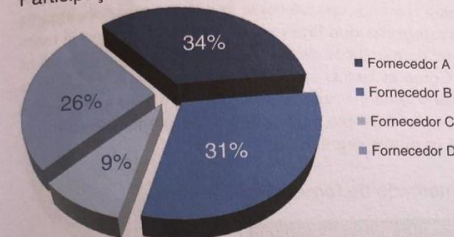


FIGURA 2.22 Gráfico de pizza com segmentos legendados

Na verdade, o "Fornecedor B" — que parece maior, com 31% — é menor que o "Fornecedor A" acima dele, que parece menor.

Vamos discutir dois problemas que apresentam um desafio para a interpretação correta desses dados. A primeira coisa que chama sua atenção (e causa suspeita, se você for um leitor de gráficos perspicaz) é a estranha perspectiva em 3D aplicada ao gráfico, inclinando a fatia e fazendo as partes superiores parecerem mais afastadas e, portanto, menores do que realmente são, enquanto as partes inferiores parecem mais perto e, assim, maiores do que são. Vamos falar mais sobre 3D em breve, mas por enquanto vou enunciar uma regra relevante sobre visualização de dados: não use 3D! Não serve para nada e pode prejudicar muito, como vemos aqui, pois distorce a percepção visual dos números.

Mesmo quando eliminamos o 3D e achatamos a pizza, os desafios à interpretação permanecem. O olho humano não consegue atribuir corretamente valores quantitativos no espaço bidimensional. Dito mais simplesmente: é difícil ler os gráficos de pizza. Quando os segmentos têm tamanhos parecidos, é difícil (se não impossível) dizer qual é o maior. Quando seus tamanhos não são parecidos, o melhor que você pode fazer

# Estatística Descritiva

## Análise Gráfica

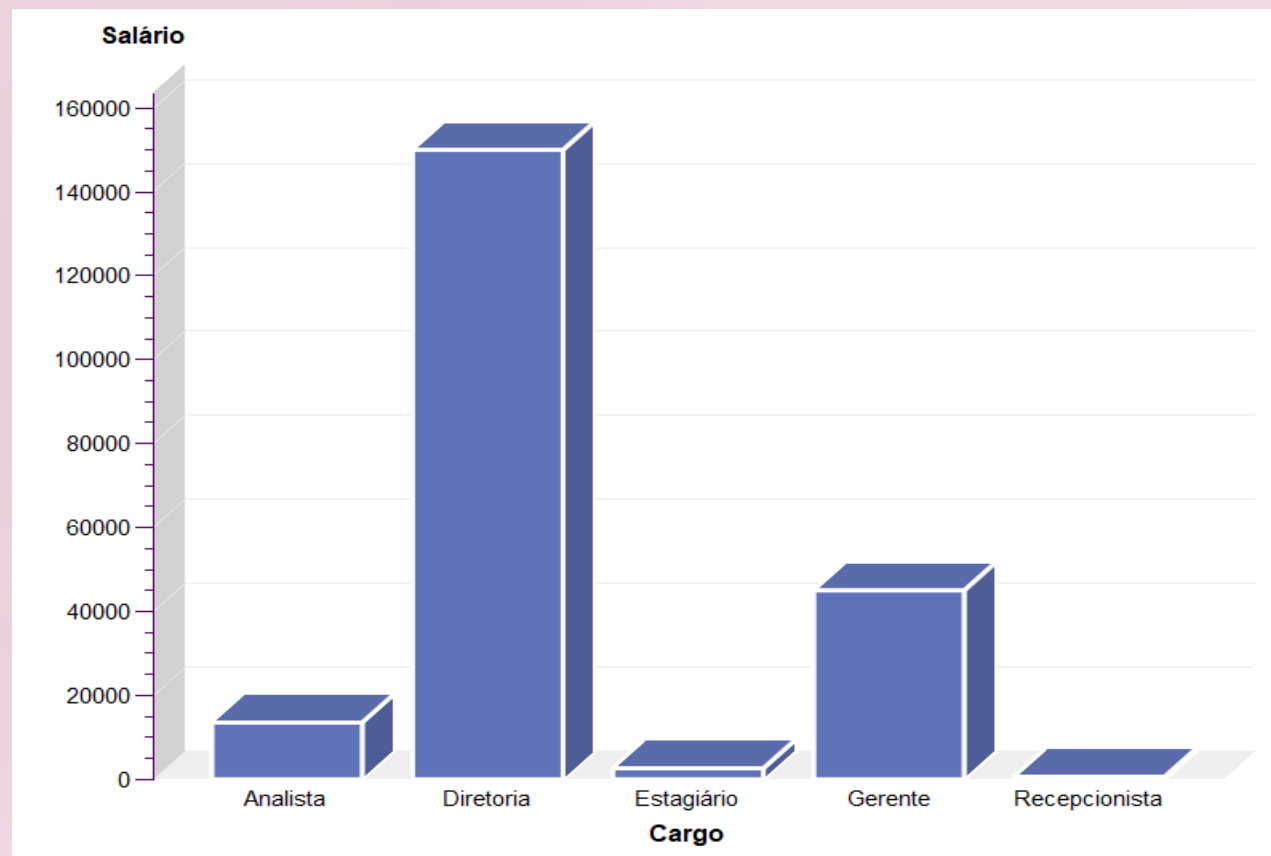
- Gráfico de Barras
  - É um gráfico representado por barras
  - Cada barra representa as frequências (ou outra medida de interesse) de cada categoria
  - É usado para dados qualitativos nominais

# Estatística Descritiva

## Análise Gráfica

- Gráfico de Barras

Cargo	Salário
Recepcionista	R\$600
Estagiário	R\$1.300
Estagiário	R\$1.300
Analista	R\$3.000
Analista	R\$4.500
Analista	R\$6.000
Gerente	R\$15.000
Gerente	R\$30.000
Diretoria	R\$50.000
Diretoria	R\$100.000







# Estatística Descritiva

## Análise Gráfica

gráfico a ser empregado, dadas algumas das dificuldades na interpretação visual discutidas aqui.

Enquanto estamos no assunto dos gráficos de pizza, vamos ver rapidamente outra "iguarria visual" a evitar: o gráfico de rosca.

O gráfico de rosca

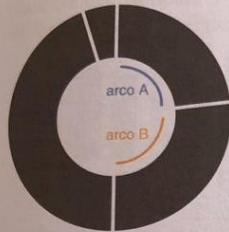


FIGURA 2.24 Gráfico de rosca

Com pizzas, estamos pedindo ao nosso público para comparar ângulos e áreas. Com um gráfico de rosca, estamos pedindo para comparar o comprimento de um arco com o comprimento de outro (por exemplo, na Figura 2.24, o comprimento do arco A comparado ao arco B). Quão confiante você se sente na capacidade de seus olhos de atribuir valores quantitativos a um comprimento de arco?

Não muito? Foi o que pensei. Não use gráficos de rosca.

### Nunca use 3D

Uma das regras de ouro da visualização de dados é esta: nunca use 3D. Repita comigo: nunca use 3D. A única exceção é se você estiver representando uma terceira dimensão (e, mesmo assim, as coisas ficam complicadas rapidamente; portanto, cuidado ao fazer isso) — e você nunca deve usar 3D para representar uma única dimensão. Como vimos no exemplo de gráfico de pizza anterior, 3D distorce nossos números, tornando difícil ou impossível interpretá-los ou compará-los.

A ser evitado 59

60 a escolha de um visual eficaz

Adicionar 3D aos gráficos introduz elementos desnecessários a ele, como painéis laterais e chão. Ainda pior do que essas distrações, os aplicativos de geração de gráficos fazem algumas coisas muito estranhas ao representar valores em 3D. Por exemplo, em um gráfico de barras em 3D, você poderia pensar que seu aplicativo de geração de gráficos representaria a frente da barra ou talvez a parte traseira. Infelizmente, muitas vezes as coisas são mais complicadas. No Excel, por exemplo, a altura da barra é determinada por um plano tangente invisível, cruzando a altura correspondente no eixo y. Isso gera gráficos como o que aparece na Figura 2.25.

Numero de publicações

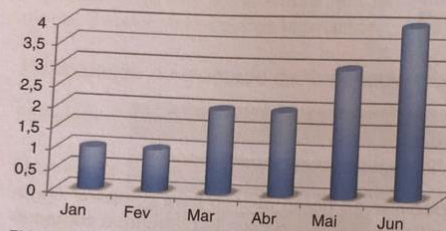


FIGURA 2.25 Gráfico de colunas em 3D

A julgar pela Figura 2.25, quantas publicações havia em janeiro e fevereiro? Eu representei uma publicação para cada um desses meses. No entanto, do modo como leio o gráfico, se comparasse a altura da barra com as linhas de grade e seguisse para a esquerda do eixo y, estimaria visualmente um valor de 0,8, talvez. Essa visualização de dados é simplesmente ruim. Não use 3D.

### Eixo y secundário: geralmente não é uma boa ideia

Às vezes é útil representar dados que estão em unidades totalmente diferentes no mesmo eixo x. Frequentemente, isso gera o eixo y secundário: outro eixo vertical no lado direito do gráfico.

Veja o exemplo mostrado na Figura 2.26.

Eixo y secundário



FIGURA 2.26 Eixo y secundário

Interpretar a Figura 2.26, requer tempo e leitura para entendermos quais dados devem ser lidos em relação a qual eixo. Por isso, você deve evitar o uso de um eixo y secundário ou à direita. Em vez disso, considere se umas das seguintes estratégias atenderá suas necessidades:

1. Não mostrar o segundo eixo y. Em vez disso, legende diretamente os pontos de dados pertencentes a esse eixo.
2. Separe os gráficos verticalmente e tenha um eixo y separado para cada um (ambos ao longo da esquerda), mas aproveite o mesmo eixo x para ambos.

A Figura 2.27 ilustra estas opções.

# Estatística Descritiva

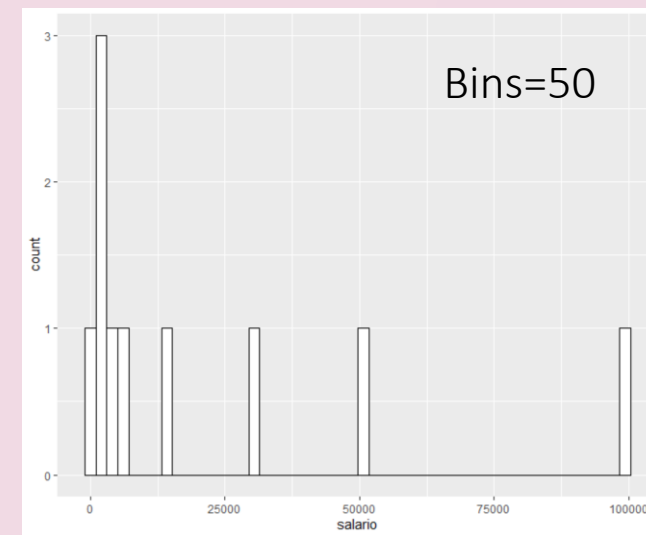
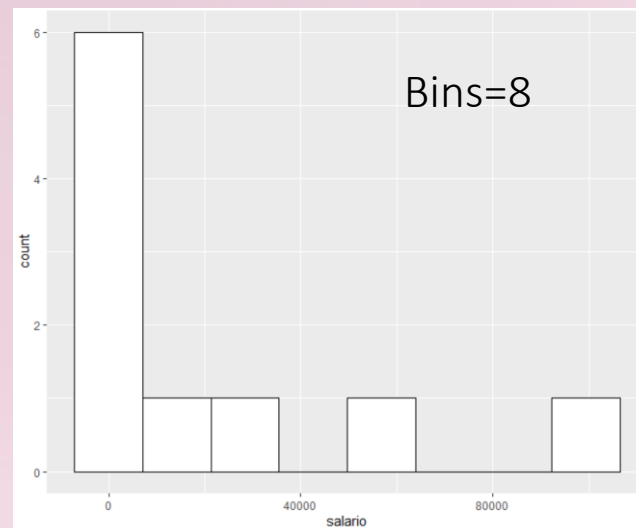
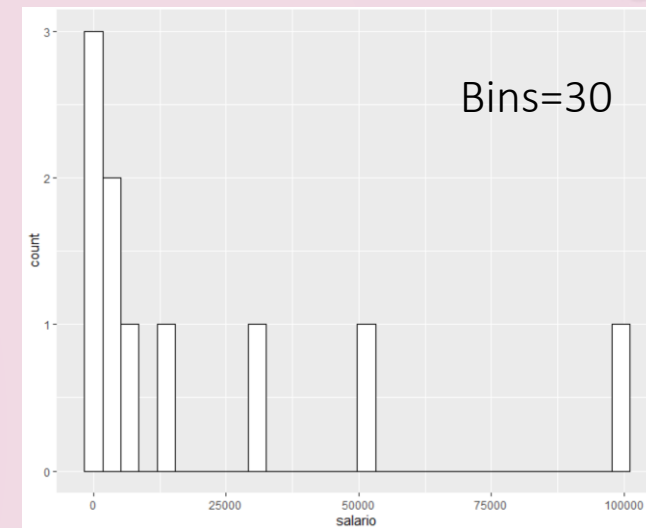
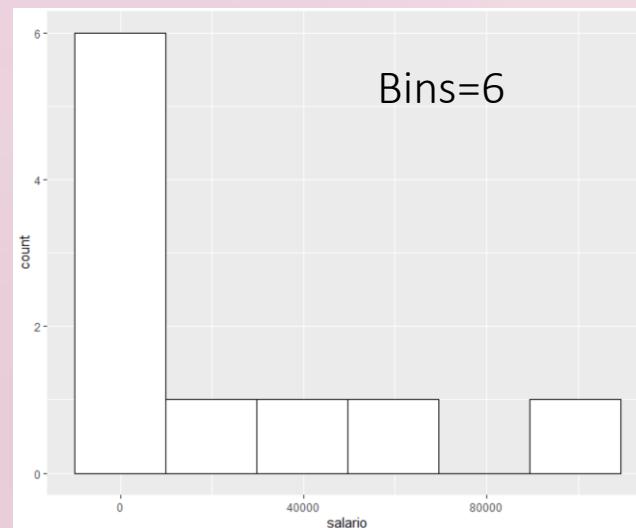
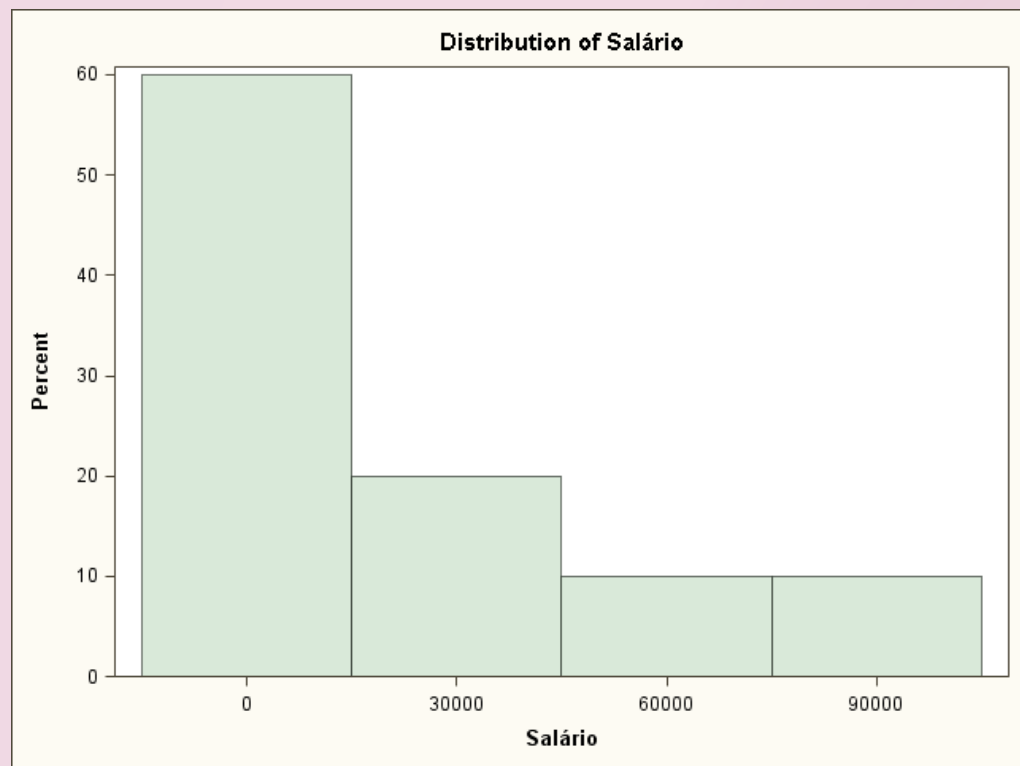
## Análise Gráfica

- Histograma
  - Representação gráfica da distribuição de frequências de dados quantitativos contínuos.
  - Pode ser um gráfico por valores absolutos, frequência relativa ou densidade.

# Estatística Descritiva

## Análise Gráfica

- Histograma



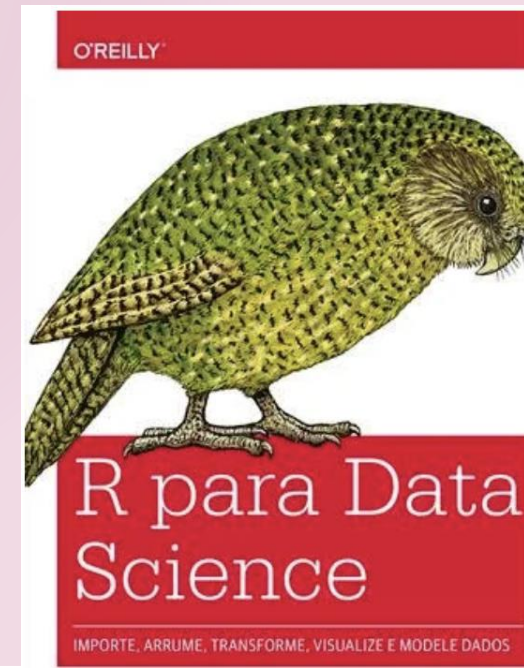
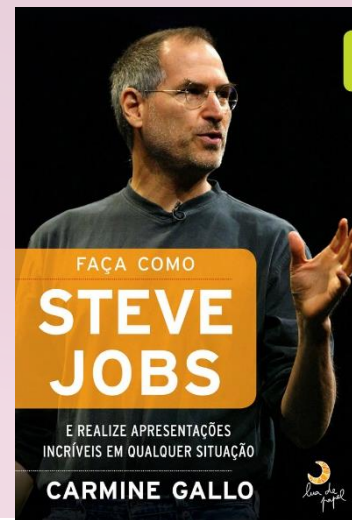
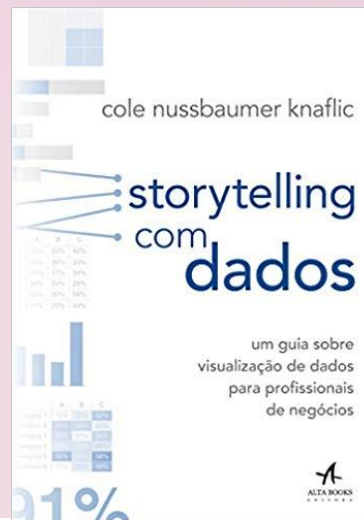
Não esqueça de deixar seu  
feedback!

=]



# Referência

- Moore, D., McCabe, G., Duckworth, W., Sclove, S. *A prática da Estatística Empresarial*. LTC, Rio de Janeiro, 2006.
- Anderson, D., Sweeney, D., Williams, T. *Estatística Aplicada à Administração e Economia*. Segunda Edição. Cengage Learning, São Paulo, 2011.
- [www.asn.rocks](http://www.asn.rocks)
- [www.curso-r.com](http://www.curso-r.com)





# It's kind of fun to do the IMPOSSIBLE



dri@asn.rocks



/in/adrianamms  
/in/asn.rocks



asn.rocks



**www.asn.rocks**

