

Estatística Descritiva: Tabelas e Gráficos

Prof. André Martins

Referências

- Bussab, W. e Morettin, P., Estatística Básica, 5^a Edição
- Berry, D. A., Statistics, A Bayesian Perspective
- Freedman, D., Pisani, R. E Purves, R.,
 Statistics
- Gallery of Data Visualization (http://www.math.yorku.ca/SCS/Gallery/)

Problema

- Vamos supor que você esteja tentando determinar se a prática de exercícios pela mãe durante a gravidez é benéfica ao bebê e se quanto mais exercício a mãe fizer, se seria melhor ou não.
- Como medida da saúde do recémnascido, você está utilizando o seu peso ao nascer



 Você registrou o peso de 462 crianças e o quanto a mãe praticou exercício durante a gravidez (os primeiros 16 dados estão na tabela ao lado) e precisa interpretar e obter alguma conclusão destes dados.

Nível de	Peso (em	
exercício	gramas)	
nenhum	3242,82	
mudando	3547,59	
mudando	3929,22	
nenhum	2765,92	
baixo/moderado	3134,82	
mudando	2693,38	
mudando	3144,96	
nenhum	3508,47	
alto	3728,29	
nenhum	4012,09	
nenhum	3973,98	
mudando	3342,50	
mudando	3278,79	
mudando	3369,27	
baixo/moderado	3583,00	
nenhum	2323,93	

Perguntas

- Existem muitas perguntas que você pode querer responder a partir dos dados.
- Por exemplo, que pesos são comuns para recém-nascidos, o quanto uma grávida costuma se exercitar, se o grau de exercício influencia no peso do bebê, etc.



- Algumas perguntas são fundamentais:
 - Estamos interessados apenas no grupo observado ou o grupo deve fornecer informações sobre um conjunto maior de indivíduos (amostra)?
 - Que tipos de variáveis estão presentes e o que pode ser feito com elas?
 - Quais perguntas queremos responder? De que forma os dados podem ajudar (ou não) a responder estas perguntas?

Resumos

- A simples visualização de todos os dados, ainda que contenha toda a informação, muitas vezes, não diz nada. Existe mais ali do que nosso cérebro consegue processar.
- É imprescindível ter técnicas de resumo que passem informações de uma forma clara e fácil de entender para podermos resolver problemas de forma eficiente.

Resumos sempre escondem algo

- Hamlet é a história de um príncipe atormentado que morre no final, assim como todos os personagens principais.
- O Senhor dos Anéis é a história de um grupo de pessoas que viajam de um lado para o outro.
 - O resumo pode estar correto (ou não, se os termos não forem definidos de forma clara), mas não servir ao problema em questão.
 - Há sempre um balanço entre a concisão (e clareza) e a quantidade de informação transmitida.
 - A decisão sobre quais informações são importantes e quais não é fundamental.

Estatística Descritiva: Resumindo Dados

- A Estatística Descritiva lida com as formas de obter informações úteis a partir de um conjunto de dados, de forma a facilitar a resolução de problemas.
- Ela o faz a partir de medidas resumo, gráficos e tabelas.

Métodos de Análise de Dados

- Simplesmente olhar para os dados não fornece um quadro claro do que pode estar acontecendo, especialmente quando a quantidade de dados for muito grande.
- A Estatística Descritiva possui uma grande quantidade de instrumentos de resumo que podem ser aplicados às diversas situações.

Métodos Gráficos e Resumos Numéricos

- Existem dois tipos de métodos que podem ser utilizados, frequentemente de forma complementar:
 - Métodos Gráficos ou Tabulares
 - Métodos Numéricos

Existem Inúmeros Resumos Úteis para a Resolução de Diferentes Problemas

- Tabelas de Frequências, Gráficos de Setores, Gráficos de barras, Gráfico de Pareto, Histogramas, Ogiva (frequência cumulativa), Ramos e folhas, Gráficos de pontos, Gráfico de Quantis, Gráficos de caixa (box-plots), Diagramas de dispersão, Gráfico temporal, Tabelas de Contingência, Gráfico qxq (quantil x quantil), etc...
- Média, Mediana, Moda, Quantis, Desvio Padrão, Variância, Intervalo Interquartil, Coeficiente de Variação, Coeficiente de Assimetria, Curtose, Coeficiente de Correlação Linear, Covariância, Coeficientes de Associação, etc...

Resumindo

 Vamos aqui, explorar apenas algumas destas técnicas, de forma a entender o que elas dizem (e o que não).

Tipos de Variáveis

- Antes de resumir algo, precisamos saber sobre o que estamos falando, ou seja, qual o tipo de variável estamos interessados. Estas podem ser:
 - Qualitativas
 - Quantitativas

Qualitativas

- Medem uma qualidade, podendo ser
 - ordinais (possuem uma ordem natural), como, por exemplo, o índice de aprovação de um político: péssimo, ruim, regular, bom ou ótimo)
 - nominais (não há uma ordem natural), como, por exemplo, o sexo de uma pessoa.

Quantitativas

- Medem uma quantidade, podendo ser
 - discretas (os possíveis valores são contáveis), como o número de alunos em uma sala ou o número de partículas no universo.
 - contínuas (podem ser observados quaisquer valores dentro de um intervalo), como a altura de uma pessoa.

Univariável ou Multivariável

Da mesma forma, existem métodos específicos para lidar com problemas onde estamos interessados em apenas uma variável, ou quando estamos interessados em descrever como duas (ou mais) variáveis se relacionam entre si.

Variáveis Qualitativas

Variáveis Qualitativas possuem uma série de medidas associadas a elas, mas, por não serem quantidades de algo, existem várias operações que não fariam sentido se aplicadas (como somar), mesmo se as variáveis se apresentarem na forma de um número (o número do RG de uma pessoa, por exemplo).

Tabela de Freqüência

Indica a frequência observada (relativa ou absoluta). No exemplo, se queremos saber, dentre as mulheres estudadas, quantas se encontram em cada categoria de exercício, obtemos a tabela:

Tabela de Frequência

- Mostra a frequência com que cada observação aparece nos dados (também pode se referir a classes de observações).
 - Frequência absoluta: número de eventos observados de um tipo
 - Frequência relativa: dada em porcentagem (ou como fração). Se foram observados xi do tipo i, dentre n dados, a frequência relativa percentual será:

 $(x_i/n)*100\%$

Tabela de Freqüência

Indica a freqüência observada (relativa ou absoluta). No exemplo, se queremos saber, dentre as mulheres estudadas, quantas se encontram em cada categoria de exercício, obtemos a tabela:

frequência		
Exercício	absoluta	
nenhum	185	
mudando	213	
baixo/ moderado	49	
alto	15	

Tabela de Freqüência

Indica a frequência observada (relativa ou absoluta). No exemplo, se queremos saber, dentre as mulheres estudadas, quantas se encontram em cada categoria de exercício, obtemos a tabela:

	frequência	frequência
Exercício	absoluta	relativa
nenhum	185	40,04%
mudando	213	46,10%
baixo/ moderado	49	10,61%
alto	15	3,25%

Frequência Cumulativa

- Mede frequência absoluta ou relativa até um certo ponto e não apenas em um valor.
- Por exemplo, número de pessoas que tem escolaridade igual ou menor que ensino médio (não apenas igual a).

Frequência Cumulativa

Supondo que podemos colocar a categoria mudando antes de baixo/moderado, temos:

	frequência	frequência
Exercício	absoluta	relativa
nenhum	185	40,04%
mudando	398	86,15%
baixo/ moderado	447	96,75%
alto	462	100,00%

Gráfico de Barras

- Apresenta a frequência absoluta ou relativa (NÃO cumulativa), ou seja, quantas observações, ou a fração de observações para um dado valor da variável em estudo (ou classe de valores).
- A altura das barras representa o que foi mais observado.

Gráficos de Barras

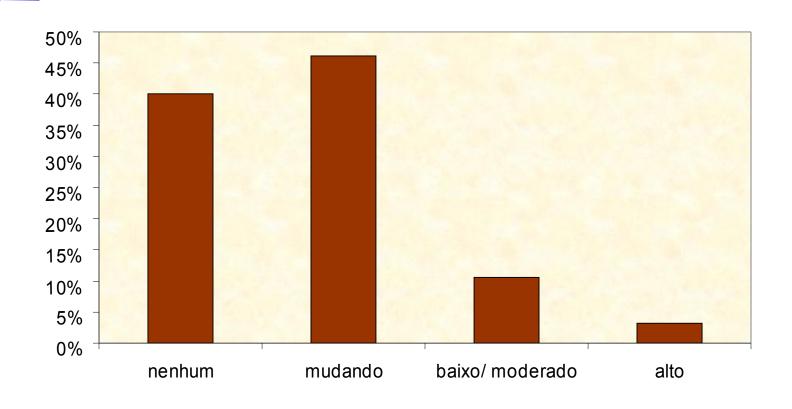
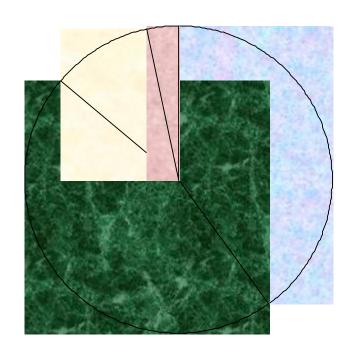


Gráfico de Setores

Dada uma frequência <u>relativa</u> observada fi da observação i, o gráfico de setores irá apresentar uma fatia de circunferência de ângulo 360*fi associada a observação i.

Gráfico de Setores (Pizza ou Torta)



□ nenhum■ mudando

□ alto

□ baixo/ moderado

Cuidado com a pizza!

- Gráficos de pizza são péssimos para visualizar comparações ou evoluções temporais. Use APENAS para fornecer a visualização de um caso, em um instante!
- Para comparações, use o gráfico apropriado para o seu caso (MAIS CLARO)!

Dados Quantitativos

 Dados quantitativos podem ser resumidos na forma de gráficos e tabelas ou a partir de medidas resumo (Média, Moda, Mediana, Desvio Padrão, etc.)

Dados Quantitativos : Tabela de Frequência

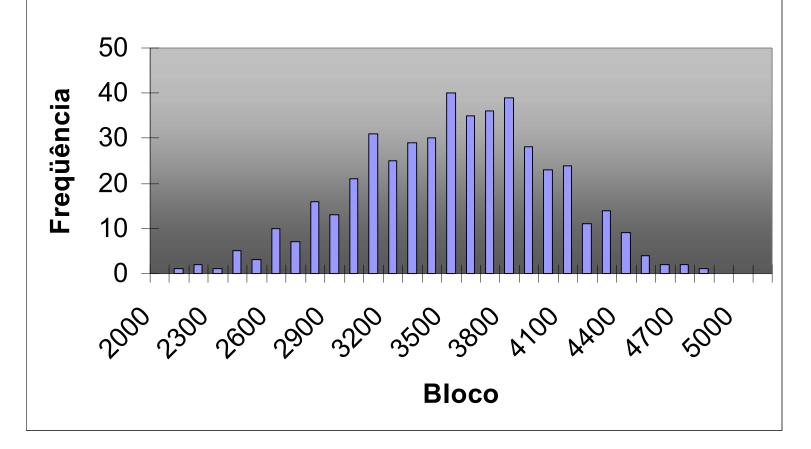
- Alguns bons conselhos de estilo:
 - Determina-se o número de classes excludentes (em geral, 5, 10 ou 20, ou um número próximo à raiz quadrada do número de dados)
 - Determina-se o tamanho de cada classe (Maior Dado-Menor Dado)/Número de Classes
 - Determina-se os Limites das classes
 - Use valores <u>aproximados</u> para o tamanho e os limites das classes!
 - Conta-se quantos dados estão em cada classe

Métodos Gráficos: Histograma

- Uma forma de apresentar dados quantitativos é o histograma, com os dados classificados por classes.
- O histograma se parece com um gráfico de barras, mas possui algumas diferenças.
- O objetivo é visualizar de que forma os dados se distribuem pelos diversos valores diferentes observados (onde é mais comum, onde é mais raro).



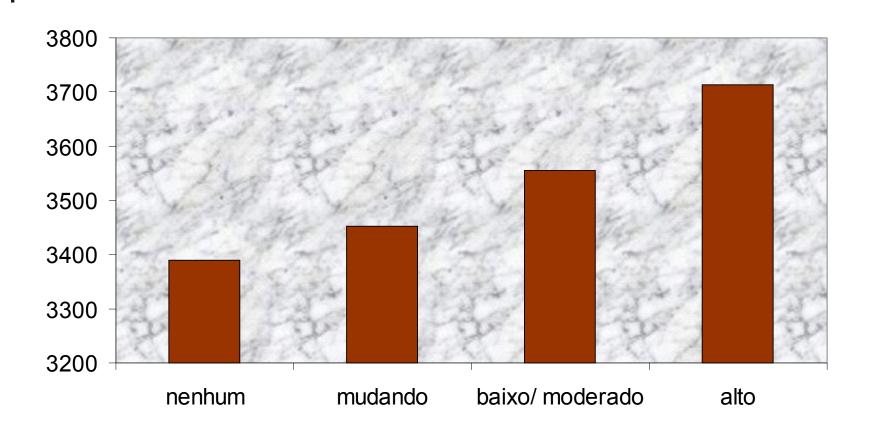




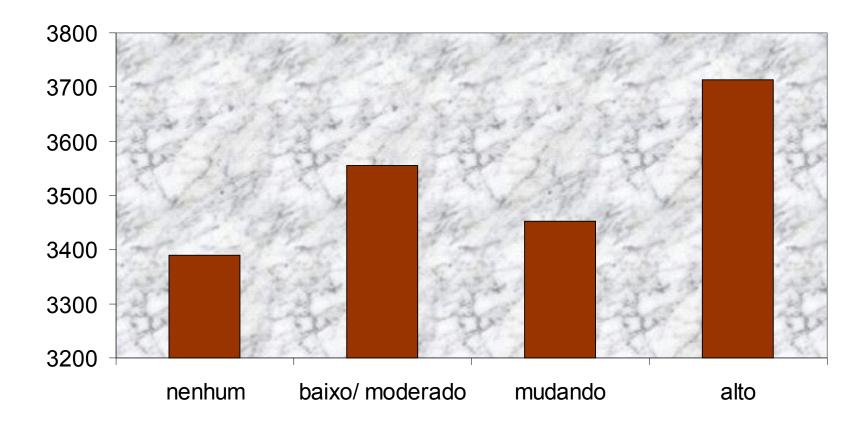
De volta ao problema inicial

- Com base no que discutimos, podemos ter, como primeira análise, um gráfico que mostre o peso médio dos bebês em cada situação diferente para as mães.
- Teríamos então:

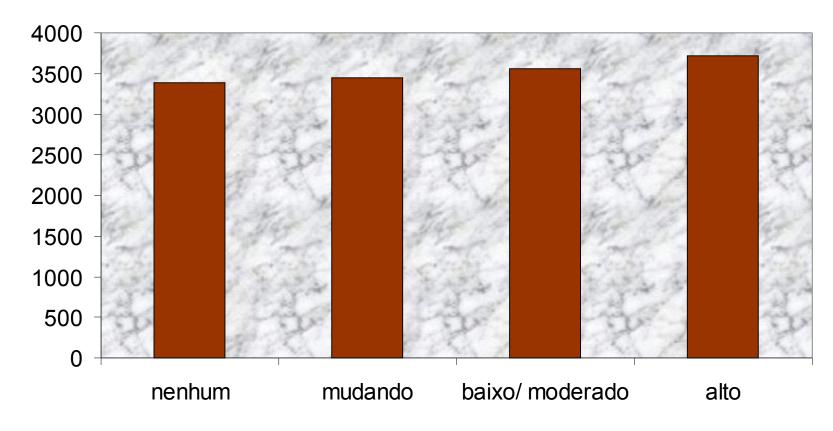
Qual a conclusão?



Será?







Verificando Relações

- Frequentemente, queremos descobrir se existe alguma relação (associação) entre duas variáveis diferentes.
- Exitem tabelas e gráficos que permitem uma visualização rápida. A melhor escolha depende do tipo das variáveis:
 - Qualitativa x Qualitativa
 - Quantitativa x Qualitativa
 - Quantitativa x Quantitativa

Relação entre Variáveis Qualitativas: Tabulação Cruzada

- A relação entre duas variáveis qualitativas pode ser representada em uma tabulação cruzada.
- Conta-se quantos valores correspondem a cada par de possíveis resultados, para as duas variáveis.
- O resultado pode ser apresentado como frequência absoluta ou relativa, em relação às colunas ou às linhas (nunca ambas).

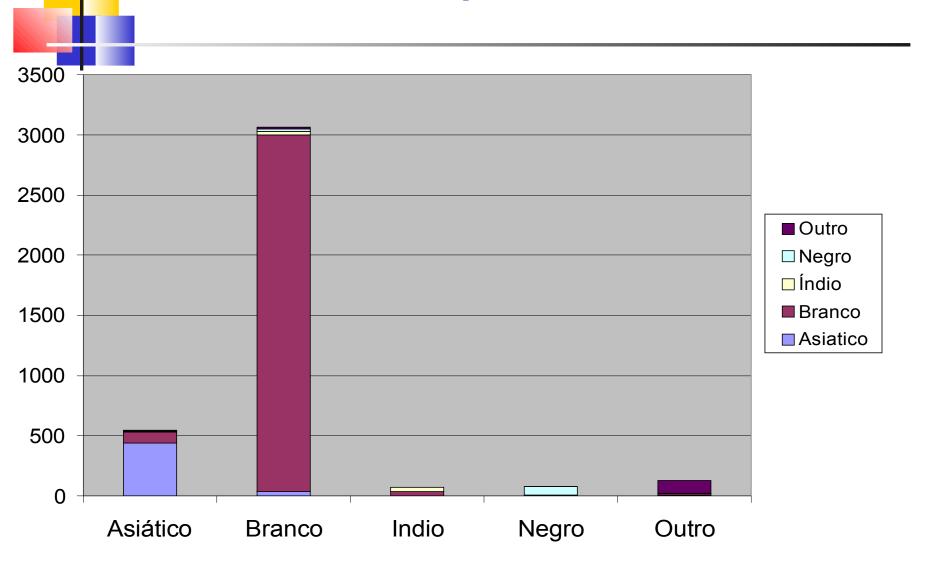


		Esposa				
		Asiático	Branco	Indio	Negro	Outro
Marido	Asiatico	441	34	1	2	4
	Branco	92	2968	38	6	19
	Índio	3	27	28	0	1
	Negro	4	18	0	72	1
	Outro	3	17	2	1	107

Relação entre Variáveis Qualitativas: Gráfico de Barras

- O Gráfico de barras, com barras segmentadas de acordo com categorias diferentes, pode ser usado para apresentar a relação entre duas variáveis qualitativas.
- Neste caso, é sempre bom apresentar valores relativos.

NÃO Use frequências absolutas



Melhor (frequências relativas):

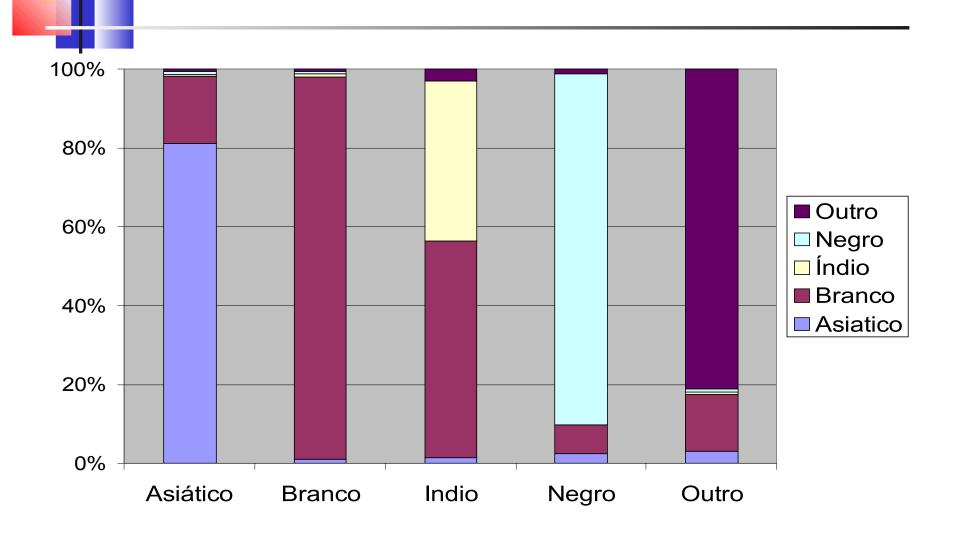
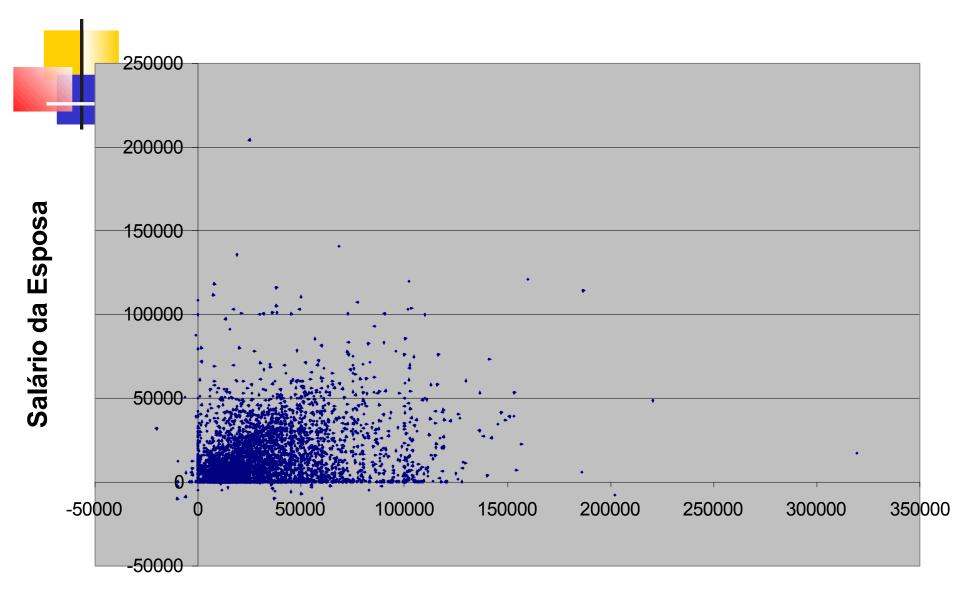


Diagrama de Dispersão (Quantitativa x Quantitativa)

- Mostra a relação entre duas variáveis quantitativas.
- Cada par observado de duas variáveis (x,y) é marcado como um ponto a partir de suas coordenadas.
- Não una os pontos!



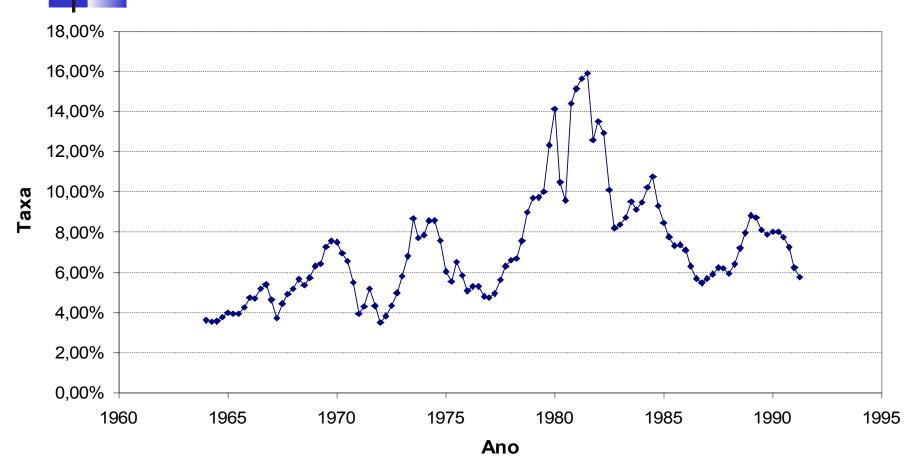
Salário do Marido

Gráfico Temporal ou Seqüencial

- Mostra a evolução de uma variável ao longo do tempo.
- É criado da mesma forma que o diagrama de dispersão, afinal é um diagrama de dispersão onde a variável x é o tempo.
- Neste caso, pode-se unir pontos consecutivos.



Juros Americanos



Comparando Dados Quantitativos (ou associação entre quantitativa e qualitativa)

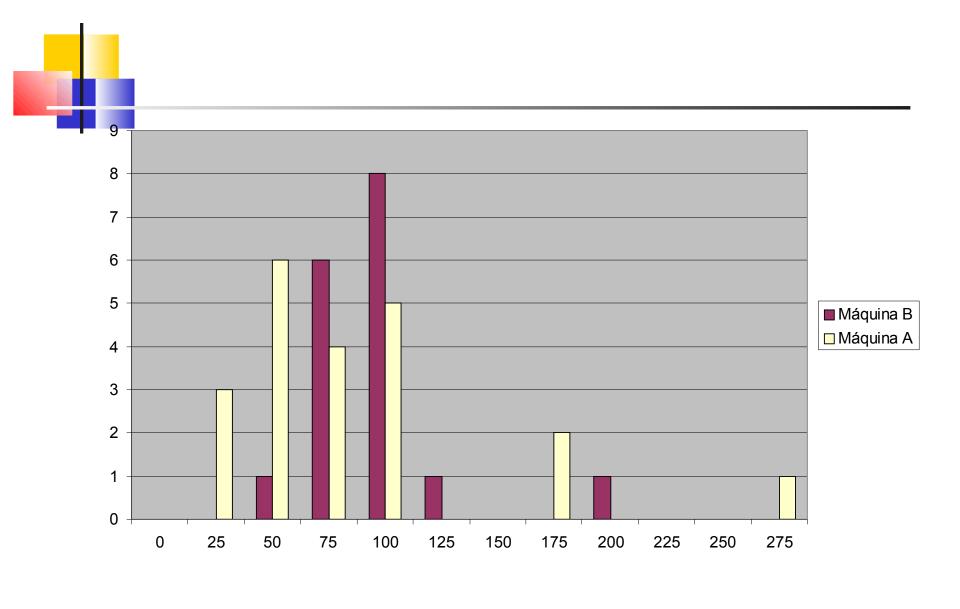
- Por vezes, queremos comparar duas distribuições de dados quantitativos, para saber se os dados se comportam da mesma forma nos dois casos.
- Por exemplo, saber se os pesos dos bebês cujas mães não fizeram exercícios tem ume distribuição igual aos pesos das mães que não fizeram.



- Uma alternativa é se preparar histogramas para cada uma das distribuições. Neste caso, para facilitar a comparação, devemos apresentar os resultados no mesmo gráfico.
- No entanto, colocar muitas barras na mesma classe pode atrapalhar a visualização.

Exemplo

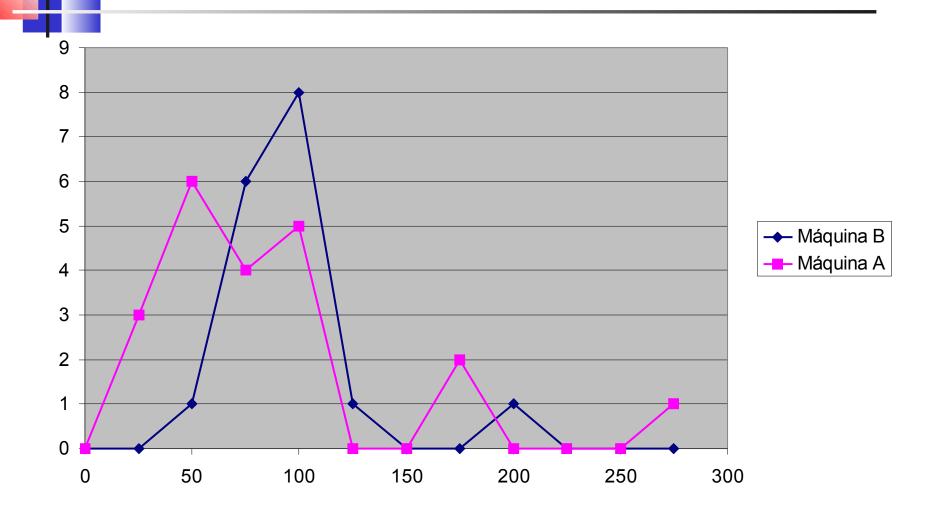
- Em uma indústria, existem duas máquinas que são fundamentais para a operação normal. No entanto, ambas as máquinas são muito delicadas e constantemente requerem reparos. Visando poder se preparar melhor para estas eventualidades, o gerente anotou o tempo transcorrido (em dias) entre os reparos mais recentes, para a máquina A e a máquina B:
 - A: 12 58 9 45 32 68 97 255 45 68 12 94 36 62 78 42 84 164 26 90 172
 - B: 122 63 180 96 49 78 95 82 63 94 88 80 62 71 60 91 65



Polígonos de Frequência

No entanto, existem outras alternativas melhores. Se, ao invés de barras, apresentarmos o histograma por pontos unidos, para cada máquina, teremos

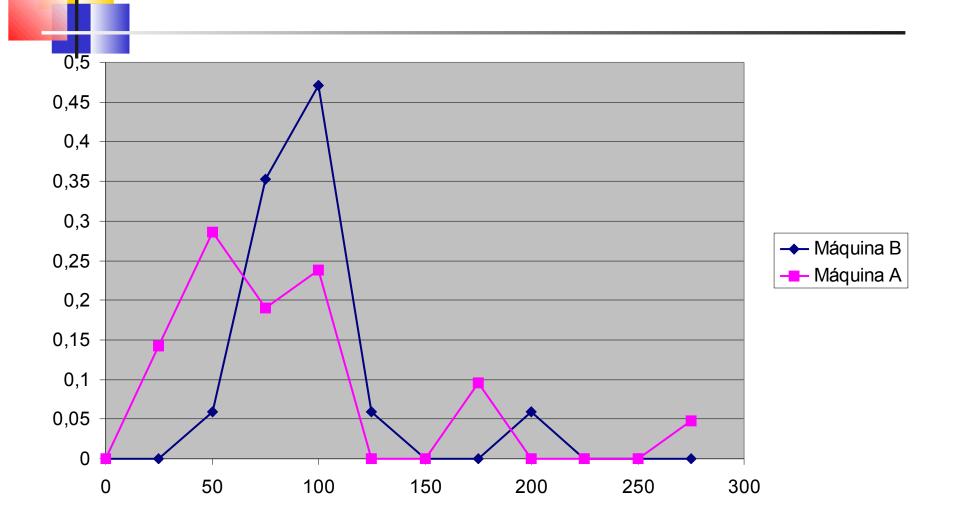
Ainda não está bom...

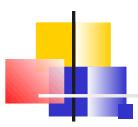




- Notemos que, no gráfico anterior, foi apresentada a frequência absoluta. Se a amostra de um caso for muito maior que a do outro, cada classe tenderá a ser mais frequente, atrapalhando a leitura.
- Corrigindo e apresentando a frequência relativa:

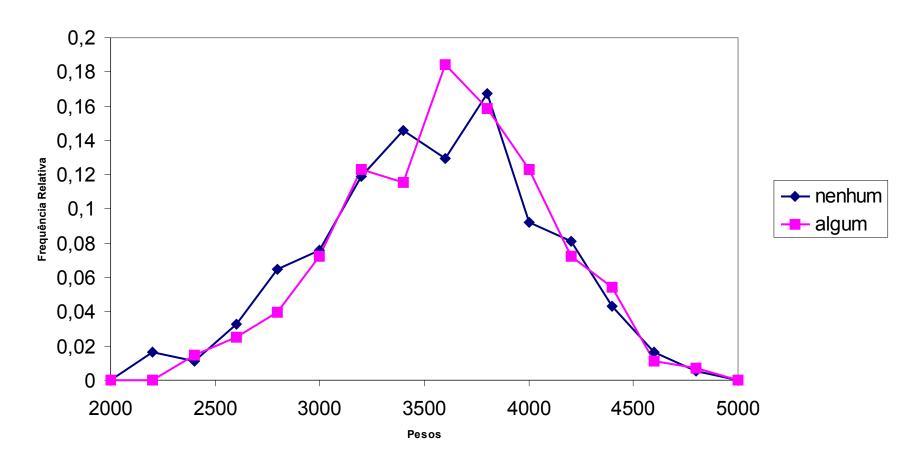
Agora sim





Para os bebês, teremos

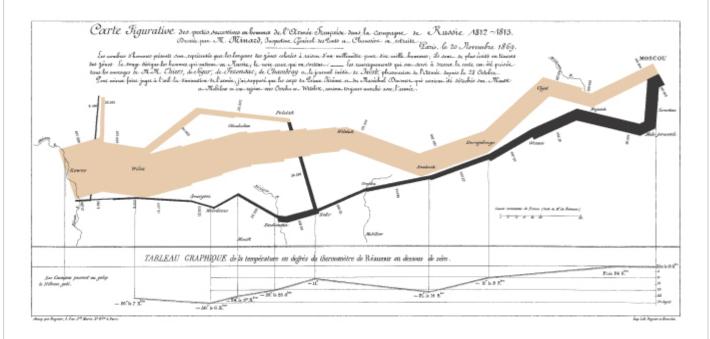
Distribuição dos Pesos dos Bebês por Exercício da Mãe



Associação entre Variáveis Qualitativas e Quantitativas

- Outra boa forma de representar esta associação é através de gráficos de caixa (box plots) que serão estudados mais adiante no curso.
- Por enquanto, você pode analisar o valor médio para cada categoria qualitativa, como no gráfico que fizemos de peso médio dos bebês por intesidade de exercício.

Um bom exemplo



Napoleon's March to Moscow The War of 1812

Cluries Joseph Minuré

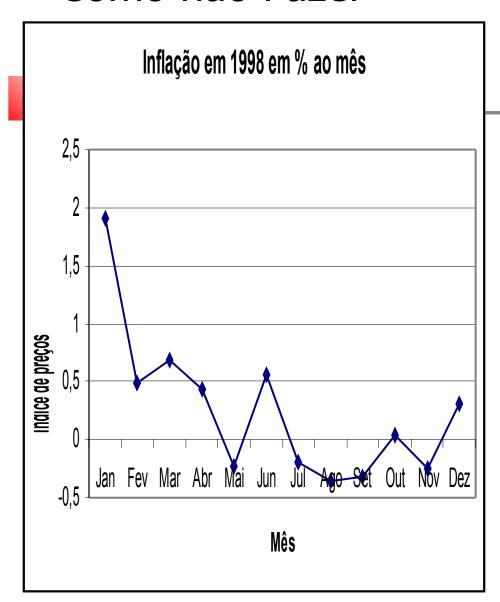
This close of Charles Jeorgh Mixard (1911-1919), the Frenck engineer, shows the terrible fits of Napoleun's may in Russia. Described in E. J. Marcy at memory to the first may not fitted by E. J. Marcy at memory to delive the per of the kinetics by in the real education, that complication of fits any past discuss, described in the person of the describing for the describing losses entitled in Napoleun's Russian campaign of situ. Registrating at the left on the Polish-Harstin bender our the Nierane River, the close the band shows the time of the same (junction campaign of situ. Registrating at the left on the Polish-Harstin bender our the Nierane River. the close is lower to the same (junction campaign of situ. Registrating at the left of the Same distincts the size of the same at each place on the map is September, the same resided Mostovo, which was by the native and described.

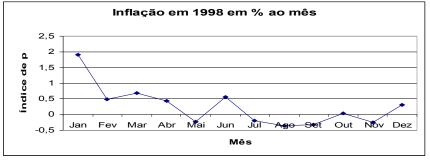
When the Charles are the same than the close of the same three close that the same three closes are the same three closes and the same three closes are three closes are the same three closes are three closes

such and dates at the between of the classe. It was a litterfly cold winter, and many force on the numb one of Branis, As the graphic thores, the consequence of the Benezian Bitter was a distinct, and the army faulty stranged both into Polician with only motion men manifeing. Also shown as the movement of madilary troops, as they single to present the rate and the flack of the advancing army. Minarchly graphic tells in othic obstract strength with in madilarization data, for more enlightening than just a single transfer beauting over time. See vanishes are platted: the size of the army, its bottless on a two-discussional surface, direction of the same's newtoness, and temperature on various dates changing the recent from Moscows. It may would be the best centralizing significe over thoses.

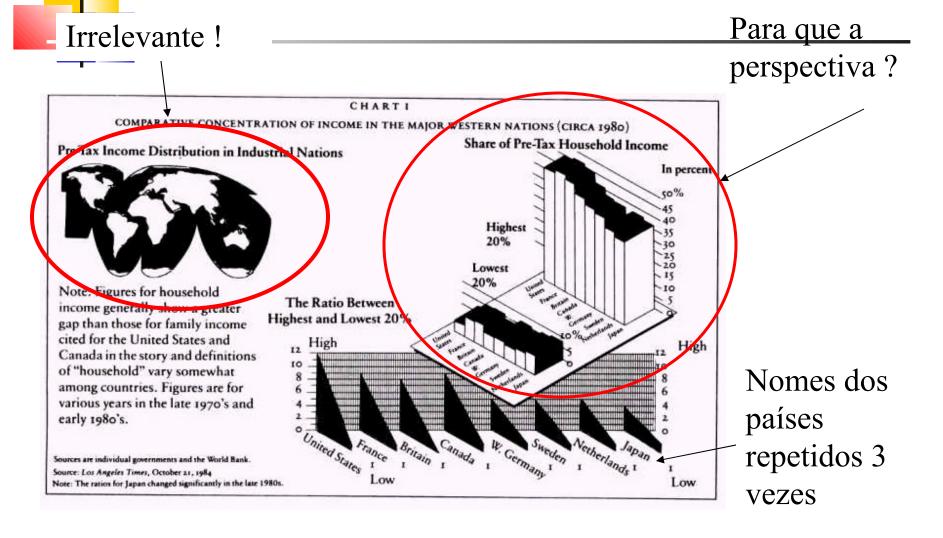
Grand B. Walle, The Pland Display of Quantitative Information — Graphics Press, Box 450, Clarabites, Communicati 19410.

Como não Fazer





Formas NÃO recomendadas



Formas NÃO recomendadas

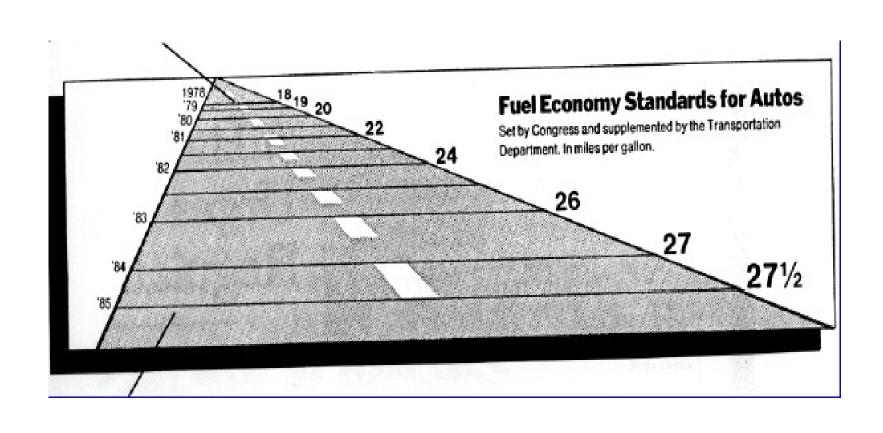
THE SHRINKING FAMILY DOCTOR

Percentage of Doctors Devoted Solely to Family Practice

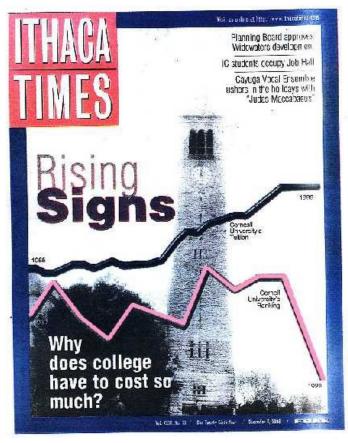


A área cresce mais rápido que a altura, o gráfico da impressão de que a diferença é muito maior do que realmente é.

Confuso!







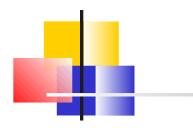
O que o gráfico ao lado parece dizer?

Mais exemplos (bons e ruins)

http://www.math.yorku.ca/SCS/Gallery/

Alguns Princípios de Desenho de Gráficos (Tufte)

- Acima de tudo, mostre os dados.
- Maximize a razão dados/tinta.
- Apague a tinta que não se refere a dados.
- Apague a tinta que se refere a dados e seja redundante.
- Revise e edite.
- Não use um gráfico só para tornar uns poucos números "bonitos"!



Interpretando medidas

Faculdade A afirma que, em média, o aluno formado ali em 95, recebe, no total, R\$21.123,45 por mês.



Um estudo recente revelou que os casais europeus escovam os dentes, em média, 0,97 vezes por dia. O mesmo estudo realizado no Brasil apresentou um resultado médio de 3,21 vezes por dia.



As estatísticas mostram claramente que os aviões estão se tornando cada vez menos seguros. Mais pessoas morreram em acidentes aéreos na última década do que na década de 20!



O consumo de uísque escocês faz bem a saúde. Foi verificado em uma pesquisa realizada no Brasil que pessoas que bebem uísque escocês regularmente vivem, em média, mais do que o restante da população.



 No ano passado, o grupo onde a doença D cresceu mais foi o de adolescentes entre 12 e 14 anos.



A propaganda de uma determinada marca de espremedor de sucos afirma que o seu espremedor espreme 26% a mais de suco, resultado comprovado por testes de laboratório.

Esta é de verdade!

A taxa de mortalidade da marinha americana durante a guerra hispanoamericana foi de 9 a cada 1.000. Durante o mesmo período da guerra, a taxa de mortalidade entre civis na cidade de Nova Iorque foi de 15 a cada 1.000. Portanto, entrar para a marinha, mesmo durante uma guerra, tornaria a vida de uma pessoa mais segura.



Um novo programa para reabilitação de prisioneiros antes de eles serem soltos está sendo testado na Califórnia. O objetivo é reduzir a taxa de reincidência – a percentagem daqueles que retornam à prisão até dois anos depois de serem postos em liberdade. O programa envolve vários meses de treinamento em um acampamento no estilo militar, com uma disciplina bastante rígida. A admissão ao programa é voluntária. De acordo com o porta-voz da prisão: "Aqueles que fazem o treinamento militar tem menor chance de retornar à prisão do que os outros detentos".



Um pai preocupado leu em uma revista que, em média, os bebês começam a andar por volta de uma certa idade, digamos, 12 meses. Como o seu filho já completou 13 meses e ainda não anda, ele conclui que a criança deve ter algum tipo de problema de desenvolvimento.