

# Apresentando Dados

Tabelas e Gráficos

Reinaldo Madarazo - 2015

# Tipos de Dados

- As variáveis em estatística são classificadas em **categóricas** ou **numéricas** e as numéricas são ainda classificadas em **discretas** ou **contínuas**.
- ***Variáveis categóricas (variáveis qualitativas)*** apresentam valores que somente podem ser posicionados em categorias como “sim” e “não”. Um exemplos desse tipo de variável é solicitar aos clientes que indiquem o dia da semana em que realizaram suas compras.

# Tipos de Dados

- ***Variáveis numéricas (variáveis quantitativas)*** apresentam valores que representam quantidades.
- ***Variáveis discretas*** apresentam valores numéricos que surgem de um processo de contagem, como por exemplo solicitar “o número de revistas que uma pessoa assina”.
- ***Variáveis contínuas*** produzem respostas numéricas que surgem de um processo de mensuração, como por exemplo o “tempo de espera para ser atendido”, cujo valor só depende da precisão do instrumento de medida.

# Escalas

- **Escala Nominal:** classifica os dados em categorias distintas sem nenhum tipo de classificação. Ex.:

Variável Categórica	Categorias
Possui computador	Sim <input type="checkbox"/> Não <input type="checkbox"/>
Tipos de ações que possui	Crescimento <input type="checkbox"/> Valorização <input type="checkbox"/> Outro <input type="checkbox"/> Nenhum <input type="checkbox"/>
Provedor de Internet	Microsoft Network <input type="checkbox"/> AOL <input type="checkbox"/> Outro <input type="checkbox"/>

- **Escala Ordinal:** classifica os dados em categorias distintas nas quais está implícita uma classificação.

Variável categórica	Categorias Ordenadas
Satisfação com o produto	Muito insatisfeito, insatisfeito, neutro, satisfeito, muito satisfeito
Classificação de Títulos na Standard & Poor	AAA, AA, A, BBB, BB, B, CCC, CC, C, DDD, DD, D
Notas do aluno	A, B, C, D, F

# Escalas

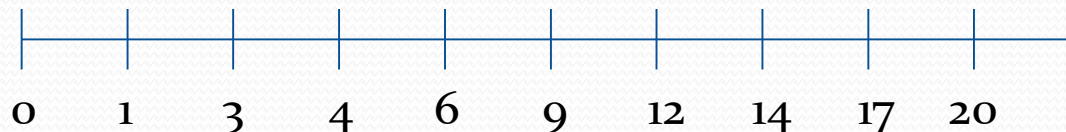
- **Escala intervalar:** é uma escala ordenada na qual a diferença entre medidas é uma quantidade significativa, embora não envolva um verdadeiro ponto zero. Ex.: Temperatura de um dado lugar ao meio dia, que corresponde a  $28^{\circ}\text{C}$  é 2 graus mais quente que uma leitura de  $26^{\circ}\text{C}$ . A diferença de 2 graus é a mesma caso as temperaturas sejam  $32^{\circ}\text{C}$  e  $30^{\circ}\text{C}$ .
- **Escala de razão:** é uma escala ordenada na qual a diferença entre medidas envolve um ponto zero verdadeiro, como por exemplo, altura, peso, idade. Uma pessoa que pesa 108 kg pesa o dobro que uma pessoa de 54kg.
- Obs: uma medida de temperatura de  $4^{\circ}\text{C}$  não é o duas vezes mais quente que uma de  $2^{\circ}\text{C}$  feita ao meio dia, uma vez que o zero é arbitrário. Isso pode ser visto mudando a escala para graus Fahrenheit.

# Escalas lineares

- Quando representadas em um eixo, as escalas devem ser obrigatoriamente lineares:



- As escalas a seguir estão erradas:



# Tabela resumida

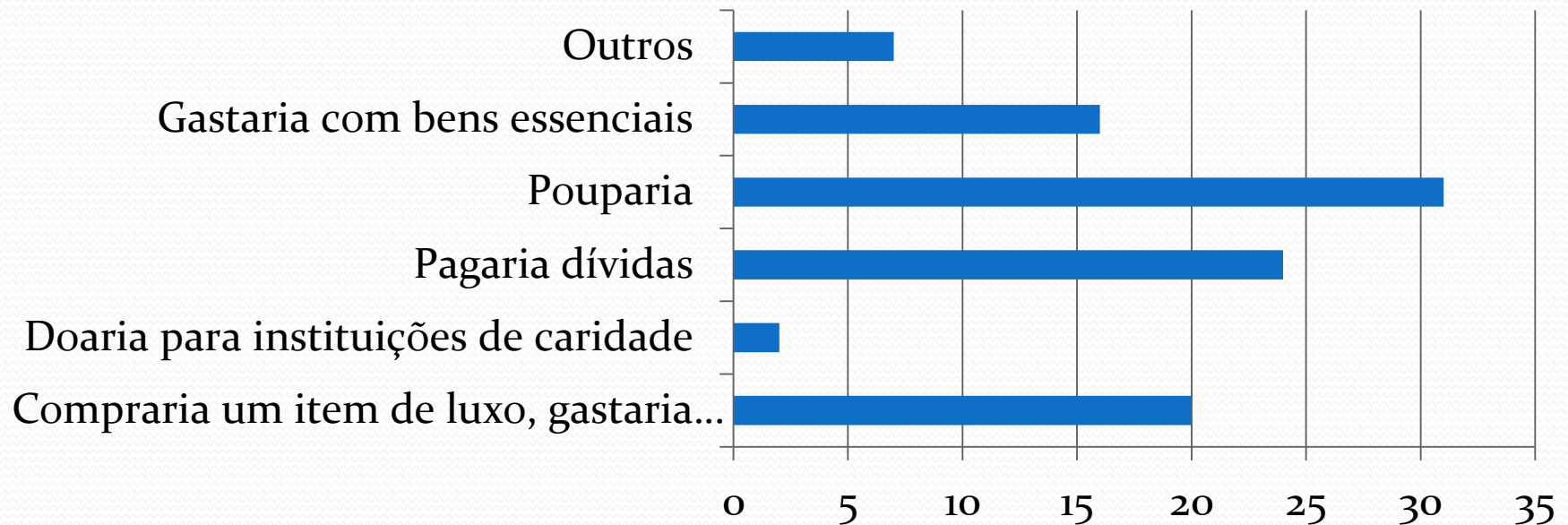
- Uma tabela resumida indica a *frequência*, a *quantidade* ou a *percentagem* de itens em um conjunto de categorias, de tal modo que se possa verificar diferenças entre categorias.
- A tabela abaixo ilustra uma tabela resumida, baseada em uma pesquisa recente, que perguntou às pessoas o que elas fariam caso tivessem \$ 1000 excedentes para gastar (“caso tivesse \$ 1000 excedentes para gastar, o que faria com o dinheiro?”).

O que você faria com o dinheiro?	%
Compraria um item de luxo, gastaria com férias ou presentes	20
Doaria para instituições de caridade	2
Pagaria dívidas	24
Pouparia	31
Gastaria com bens essenciais	16
Outros	7

# O Gráfico de Barras

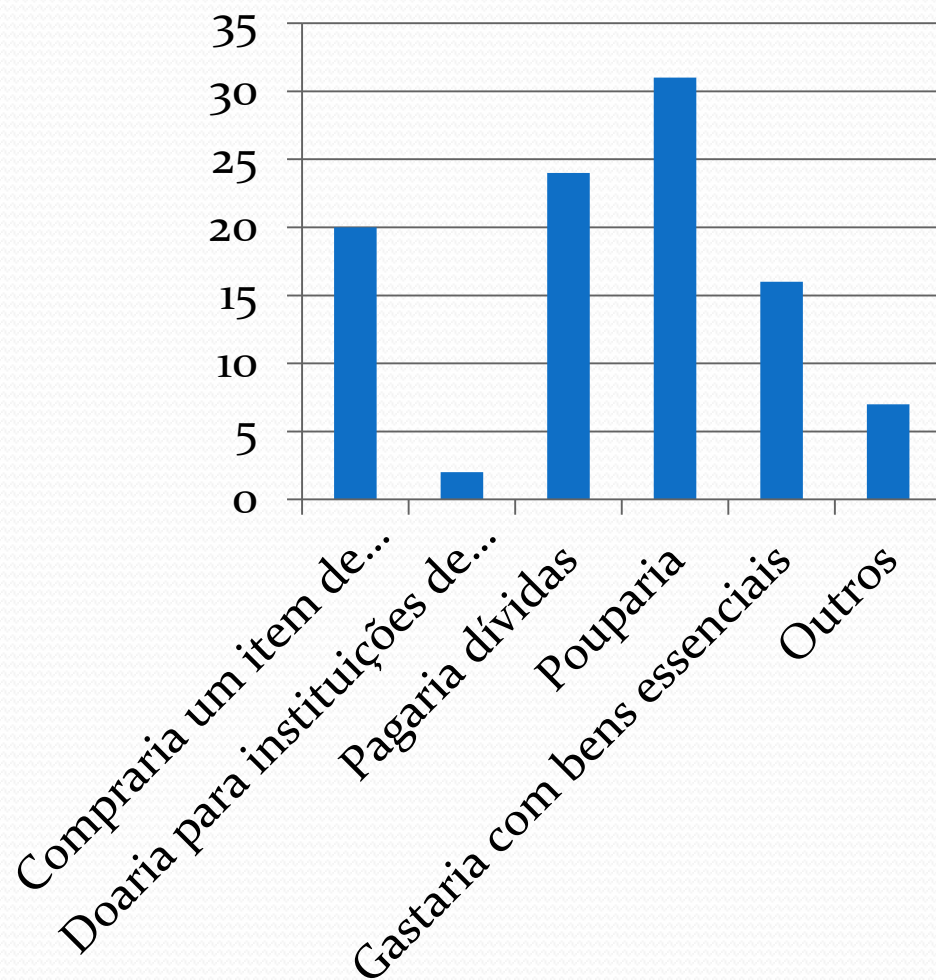
Em um **gráfico de barras**, uma barra ilustra cada uma das **categorias**, cujo comprimento representa a quantidade, a frequência ou a percentagem de valores que se posicionam em uma determinada categoria.

Abaixo o gráfico de barras para as hipóteses de gasto de \$1000 excedentes.



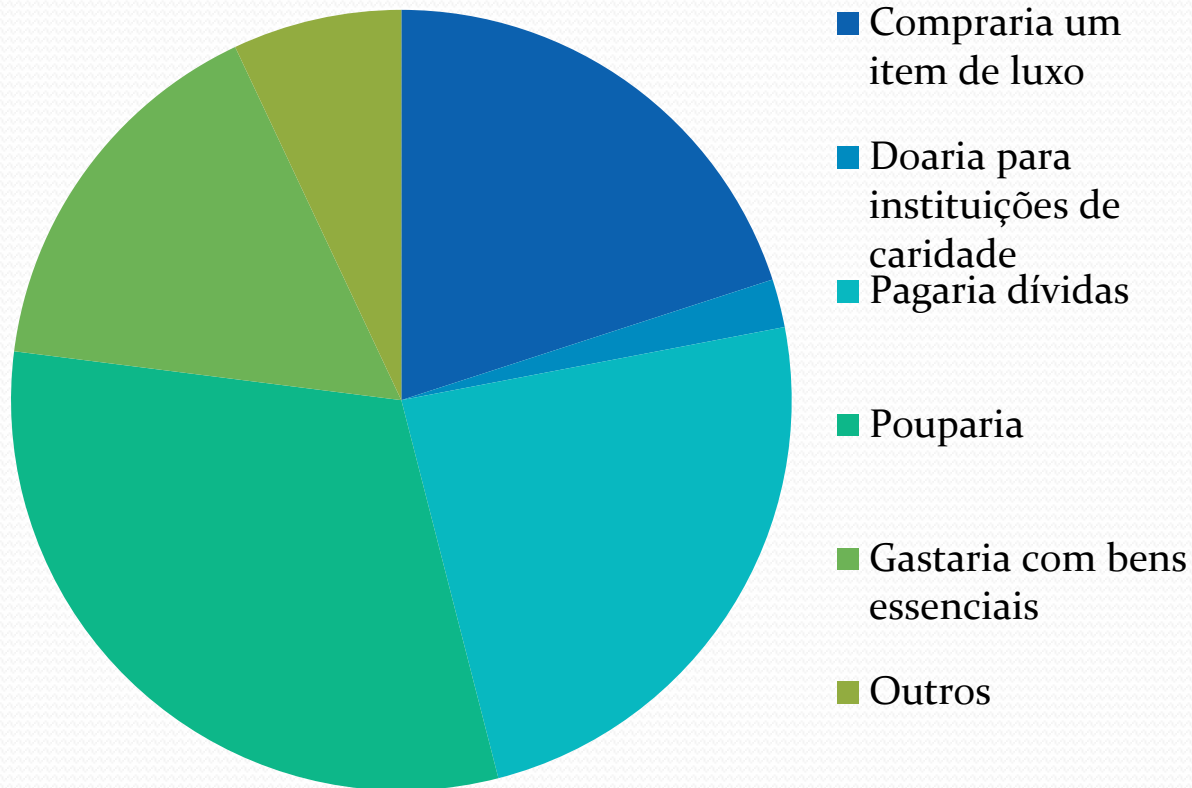


# O Gráfico de Colunas



O gráfico de colunas é semelhante ao gráfico de barras, onde cada coluna representa uma **categoria** e sua altura representa a quantidade, a frequência ou a percentagem de valores. Ao lado o gráfico de colunas para as hipóteses de gasto de \$1000 excedentes.

# O Gráfico de Pizza

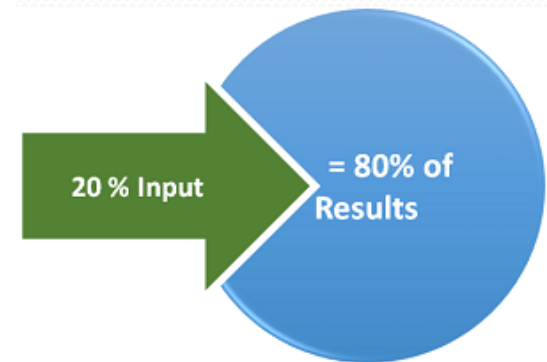
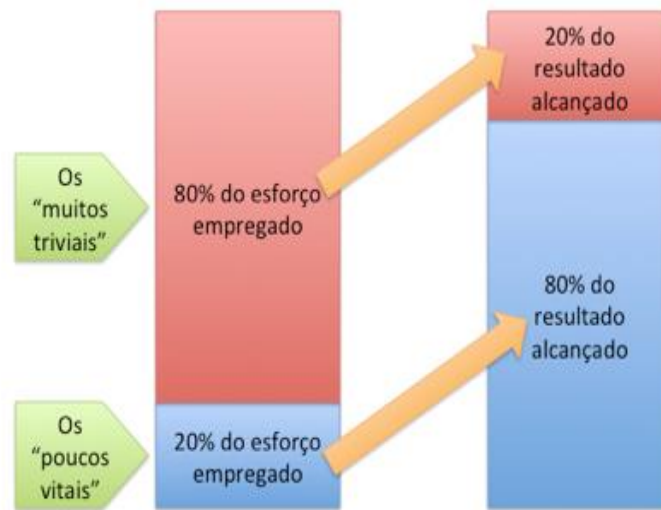


O *Gráfico de Pizza* é um círculo desmembrado em fatias que representam as **categorias** e o tamanho (abertura em graus) de cada fatia varia de acordo com a percentagem em cada uma das categorias.

No exemplo ao lado, 31% dos respondentes afirmaram que poupariam os \$ 1000. Assim, na construção do gráfico, os 360 graus que compõem o círculo são multiplicados por 0,31, resultando em uma fatia que ocupa 111,6 graus dos 360 graus correspondentes ao círculo.

# O Princípio de Pareto

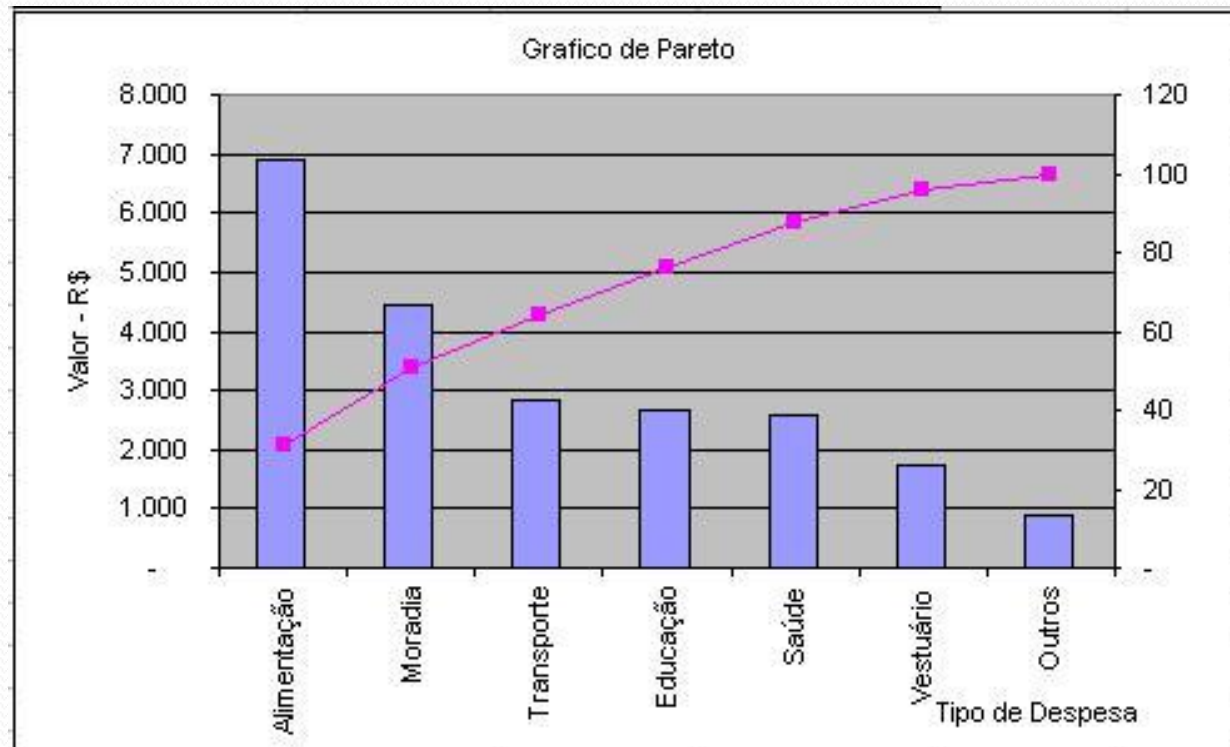
- O **princípio de Pareto** existe quando a maioria dos itens em um conjunto de dados ocorre em uma pequena quantidade de categorias, e os poucos itens remanescentes estão espalhados por uma grande quantidade de categorias. Esses dois grupos são geralmente chamados de “*poucos vitais*” e “*muitos triviais*”.



Vilfredo Pareto (1848-1923)

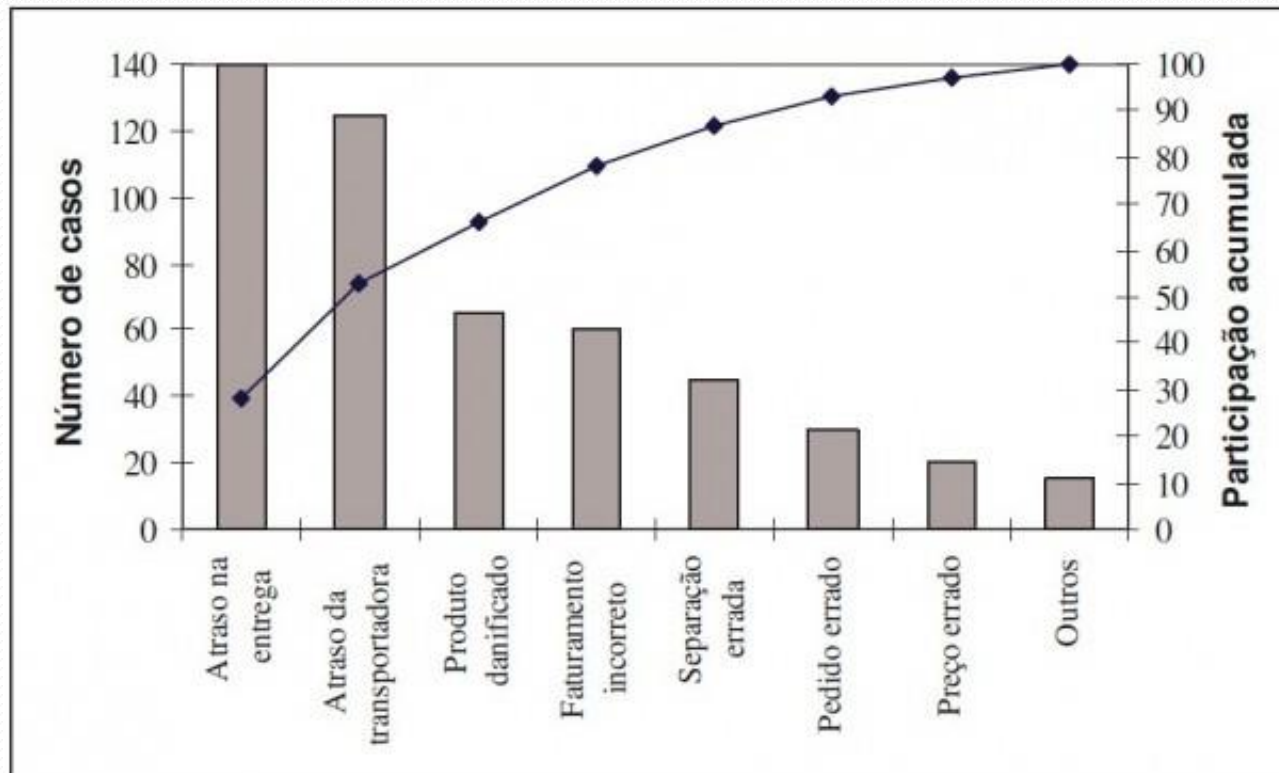
# O Diagrama de Pareto

- No *Diagrama de Pareto*, as respostas categorizadas são apresentadas sob a forma de um **gráfico de colunas** em ordem *descendente*, de acordo com suas respectivas *frequências*, e são combinados com uma linha correspondente à *percentagem acumulada* no mesmo gráfico. O diagrama de Pareto consegue identificar os “**poucos vitais**” e os “**muitos triviais**”.



# Diagrama de Pareto

- Em situações nas quais os dados envolvidos consistem em itens com defeito ou não-conformes, o diagrama de Pareto representa uma ferramenta poderosa para a priorização de esforços voltados para melhorias.



# Organização de Dados Numéricos

- Quando a quantidade de valores é muito grande, é conveniente organizar os dados em uma disposição ordenada.
- **Disposição Ordenada:** consiste em uma sequência de dados em uma ordem de classificação do menor valor para o maior valor. Por exemplo, os valores abaixo correspondem às notas oriundas de uma amostra de  $n=7$  resultados de provas de contabilidade:

6,3 - 6,4 - 6,8 - 7,1 - 7,5 - 8,8 - 9,4

Com base nesses dados, podemos verificar que as notas estão compreendidas entre 6,3 e 9,4.

# Disposição Ramo-e-Folha

- A disposição **Ramo-e-Folha** organiza os dados em grupos (chamados de ramos) de tal modo que os valores dentro de cada grupo (as folhas) se ramifiquem para a direita de cada linha.
- A disposição resultante permite verificar o modo como os dados estão distribuídos e os locais onde existem concentrações.

# Disposição Ramo-e-Folha

- Exemplo: os dados a seguir correspondem aos gastos com almoço:

5,40	4,30	4,80	5,50	7,30	8,50	6,10	4,80
4,90	4,90	5,50	3,50	5,90	6,30	6,60	

- Para formar a disposição ramo-e-folha, utiliza-se as unidades como os ramos e os decimais (folhas) são arredondados uma casa para cima. Por exemplo, o primeiro valor (5,40) tem como ramo o número 5 e folha o 40 arredondada para 4. Os valores serão dispostos da seguinte maneira:

```
3 | 5
4 | 3 8 8 9 9
5 | 4 5 5 9
6 | 1 3 6
7 | 3
8 | 5
```



# Exercício:

- Os dados a seguir representam o tempo em segundos para a carga de um aplicativo em um sistema compartilhado, com 50 observações:

5,2	6,4	5,7	8,3	7,0	5,4	4,8	9,1	5,5	6,2
4,9	5,7	6,3	5,1	8,4	6,2	8,9	7,3	5,4	4,8
5,6	6,8	5,0	6,7	8,2	7,1	4,9	5,0	8,2	9,9
5,4	5,6	5,7	6,2	4,9	5,1	6,0	4,7	14,1	5,3
4,9	5,0	5,7	6,3	6,0	6,8	7,3	6,9	6,5	5,9

- Escreva esses dados na disposição ramo-e-folha.

# Solução

4	7 8 8 9 9 9 9	(7)
5	0 0 0 1 1 2 3 4 4 4 5 6 6 7 7 7 7 9	(18)
6	0 0 2 2 2 3 3 4 5 7 8 8 9	(13)
7	0 1 3 3	(4)
8	2 2 3 4 9	(5)
9	1 9	(2)
10		(0)
11		(0)
12		(0)
13		(0)
14	1	(1)

# Tabelas e Gráficos de dados numéricos

- ***Distribuição de frequências***: é uma tabela resumida, na qual os dados são dispostos em *grupos de classe* ordenados numericamente.
- Na construção da tabela, deve-se ter atenção à seleção da *quantidade* apropriada de grupos de classe, à determinação da *amplitude* adequada para um grupo de classe e ao estabelecimento de *limites* para cada grupo de classe, visando evitar sobreposições.
- A quantidade de grupos de classe utilizadas depende da quantidade de dados disponíveis. Uma maior quantidade de valores permite uma maior quantidade de grupos de classe.

# Número de grupos de classe

- Existem 3 metodologias para a determinação do número de classes que serão utilizadas:
  - **Escolha arbitrária:** em geral, a distribuição de frequências de ter entre 5 a 15 grupos. Quantidades demasiadamente pequena ou demasiadamente grande faz com que haja uma pequena quantidade de novas informações.
  - **Regra da raiz quadrada:** o número de classes  $C$  pode ser obtido pela expressão  $C = \sqrt{n}$  onde  $n$  é a quantidade de dados.
  - **Regra de Sturges:** o número de classes  $C$  pode ser obtido pela seguinte expressão, onde  $n$  é a quantidade de dados:

$$C = \frac{\log(n)}{\log 2} + 1 = 1 + 3,3 \log(n)$$

# Amplitudes

- Para se determinar a *amplitude de um intervalo de classe (grupo)*, divide-se a *amplitude* (maior valor – menor valor) do conjunto de dados pela quantidade desejada de classes:

$$\textit{Amplitude do intervalo} = \frac{\textit{Amplitude}}{C}$$

- Por exemplo, no caso do tempo de carga de um sistema, a amplitude vale  $14,1 - 4,7 = 9,4$ .
- Admitindo um valor arbitrário de  $C=10$ , obtemos como amplitude do intervalo  $9,4/10 = 0,94$ .

# Amplitudes

- No entanto, deve-se escolher uma ***amplitude de intervalo*** que simplifique a leitura e a interpretação. Portanto, em vez de se utilizar uma amplitude de **0,94**, adotaremos como amplitude de intervalo o **valor 1**.
- Assim, a primeira classe deve ter como *limite inferior* o valor **4** e o *superior menor que* **5**. A próxima classe teria como *limite inferior* o valor **5** e o *superior menor que* **6** e assim por diante.
- O centro de cada classe é denominado ***Valor Médio da Classe*** e está posicionado a meio caminho entre os limites inferior e superior.

# Exercício:

- Utilizando os dados referentes ao tempo de carga de um sistema, determine a quantidade, as classes e o valor médio de cada classe.

(a) *Usando a regra da raiz quadrada:*

$$n = 50 \Rightarrow C = \sqrt{50} = 7,07 = 7 \text{ classes}$$

(b) *Usando a regra de Sturges:*

$$C = \frac{\log(50)}{\log(2)} + 1 = \frac{1,69897}{0,30103} + 1 = 6,64 = 7$$

(c) *Usando a regra arbitrária:* com base nos valores obtidos acima e para simplificar a exibição dos dados na classe, escolheremos o valor **C=10**. Este valor, embora arbitrário, foi escolhido com base nos resultados anteriores.

# Solução

- Indicaremos com o símbolo  $\mid$  para indicar que o limite inferior engloba o valor (é igual) e o limite superior é menor que o valor. Dessa forma, as classes são as seguintes:

Classe	Valor Médio	Classe	Valor Médio
4 $\mid$ 5	4,5	10 $\mid$ 11	10,5
5 $\mid$ 6	5,5	11 $\mid$ 12	11,5
6 $\mid$ 7	6,5	12 $\mid$ 13	12,5
7 $\mid$ 8	7,5	13 $\mid$ 14	13,5
8 $\mid$ 9	8,5	14 $\mid$ 15	14,5
9 $\mid$ 10	9,5		



# Distribuição de Frequências

- Agora é possível montar uma tabela relacionando as classes obtidas, o valor médio da classe e o número de ocorrências de valores em cada classe (frequência).
- Através da distribuição ramo-e-folha é fácil obter a frequência dos dados, isto é, o número de valores que se encontram dentro de cada classe, que é denominada de *Frequência Absoluta*.
- *Exercício:* Monte a tabela de Distribuição de Frequências dos valores de carga de um sistema.

# Solução

Classes	Valor Médio	Frequência Absoluta
4 — 5	4,5	7
5 — 6	5,5	18
6 — 7	6,5	13
7 — 8	7,5	4
8 — 9	8,5	5
9 — 10	9,5	2
10 — 11	10,5	0
11  — 12	11,5	0
12  — 13	12,5	0
13  — 14	13,5	0
14  — 15	14,5	1

# Distribuição de Frequências Relativas e Percentagens

- De um modo geral, é interessante conhecer a proporção ou percentagem do total que está em cada um dos grupos.
- Assim, a distribuição de Frequências Relativas ou a Distribuição de Percentagens é preferida em relação à distribuição de frequências.
- A frequência relativa pode ser obtida dividindo o valor da frequência absoluta pela quantidade total de valores.
- Escrevendo esses resultados em %, temos a distribuição de percentagens.

# Distribuição Acumulada

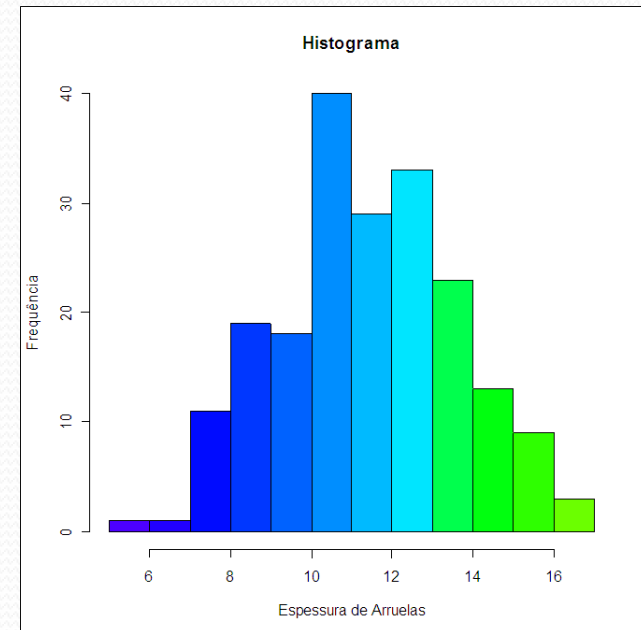
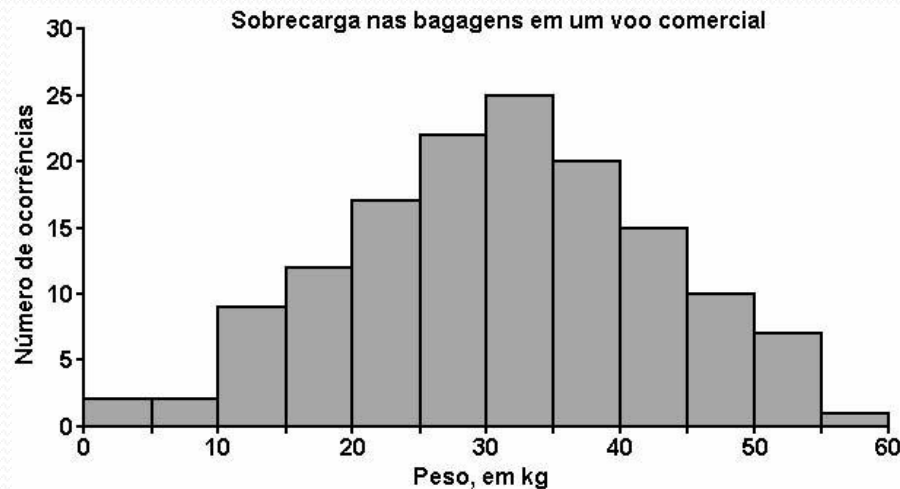
- A distribuição de percentagens acumuladas proporciona um modo de apresentar informações sobre a percentagem de itens que estão abaixo de um determinado valor.
- É construída somando-se os valores das percentagens de cada classe.
- Exercício: monte a tabela de frequências absolutas, relativas, percentagem e acumulada dos 50 valores dos tempos de carga de um sistema.

# Solução

Classes	Valor Médio	Frequência Absoluta	Frequência Relativa	Percentagem %	Percentagem Acumulada %
4 — 5	4,5	7	0,14	14	14
5 — 6	5,5	18	0,36	36	50
6 — 7	6,5	13	0,26	26	76
7 — 8	7,5	4	0,08	8	84
8 — 9	8,5	5	0,10	10	94
9 — 10	9,5	2	0,04	4	98
10 — 11	10,5	0	0	0	98
11 — 12	11,5	0	0	0	98
12 — 13	12,5	0	0	0	98
13 — 14	13,5	0	0	0	98
14 — 15	14,5	1	0,02	2	100

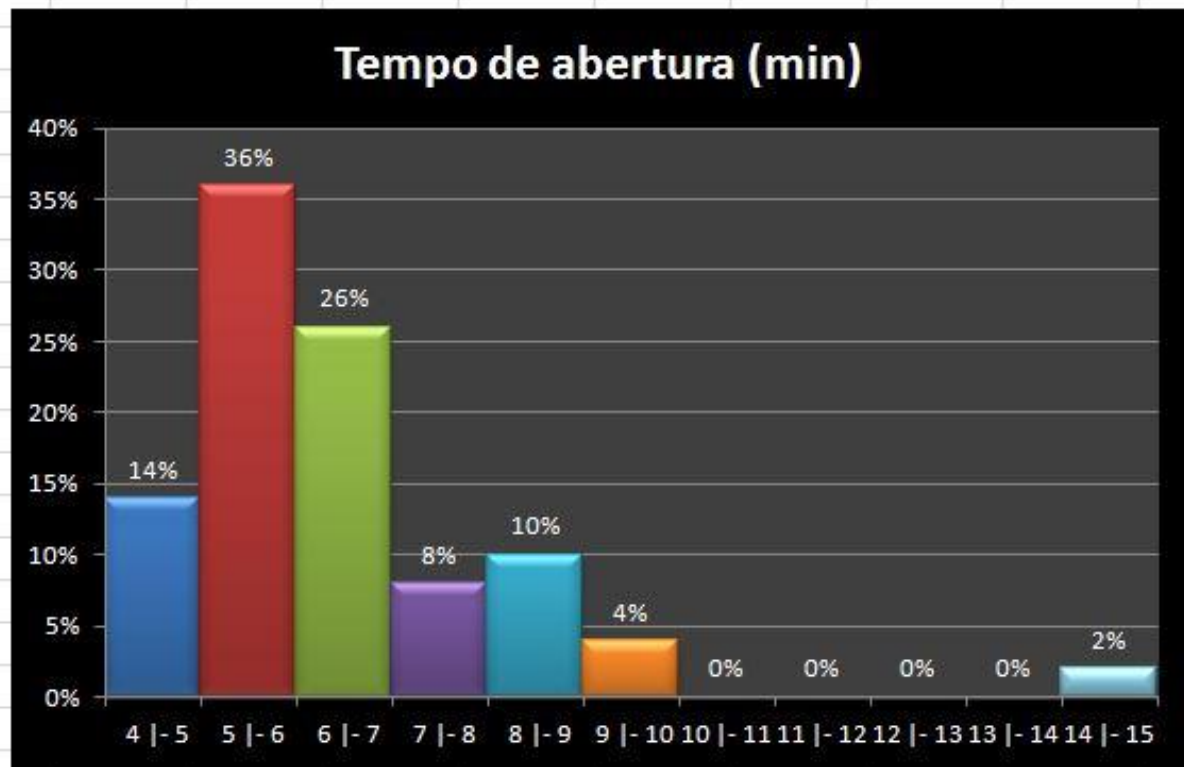
# O Histograma

- Um histograma é um gráfico de colunas para dados numéricos em grupo no qual as frequências ou percentagens de cada grupo de dados numéricos são representados por barras verticais individuais.



# Exercício:

- Construa o histograma para os dados do tempo de carga de um sistema.



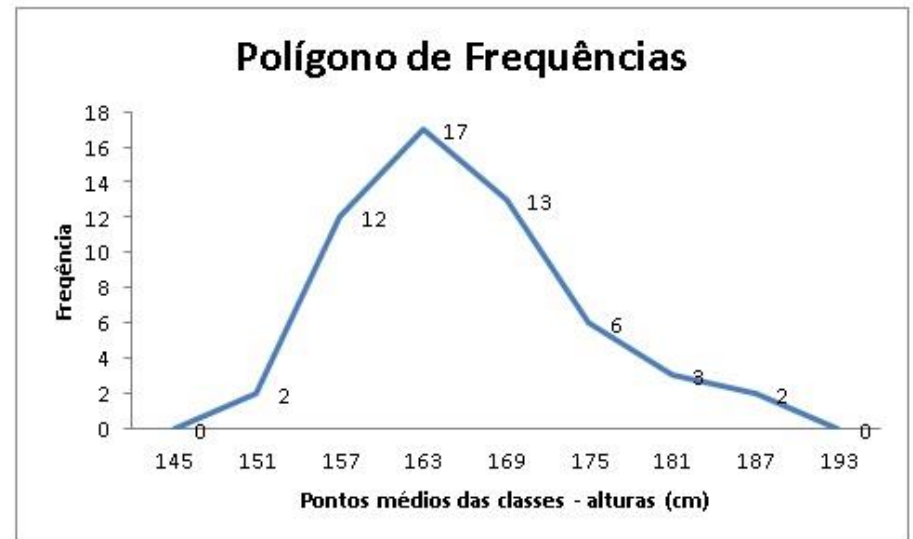
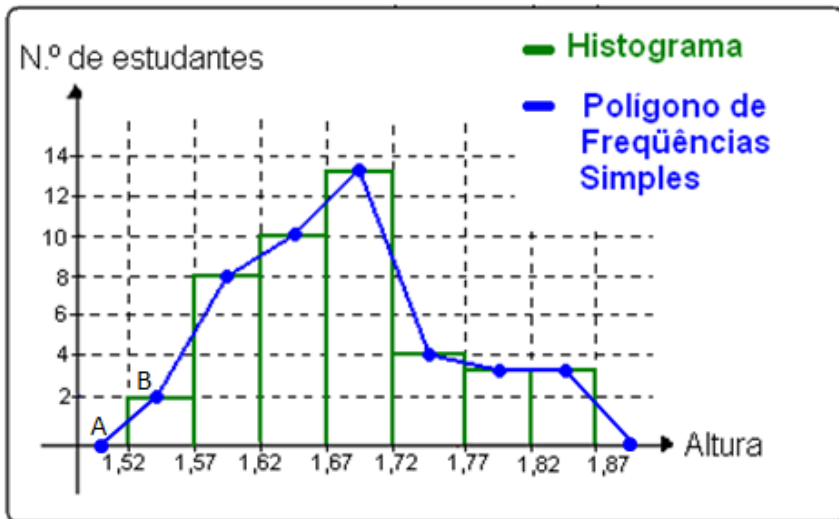
# Polígono

- A construção de vários histogramas no mesmo gráfico torna-se confusa. A superposição de barras verticais de um histograma em um outro histograma dificulta a interpretação. Quando se deseja representar dois ou mais grupos de dados, devem ser usados os *polígonos de percentagem*.
- ***Um polígono de percentagens*** é formado fazendo-se com que o ponto médio de cada classe represente os dados naquela classe e, depois, interligando-se a sequência de pontos médios em suas respectivas percentagens de classe.



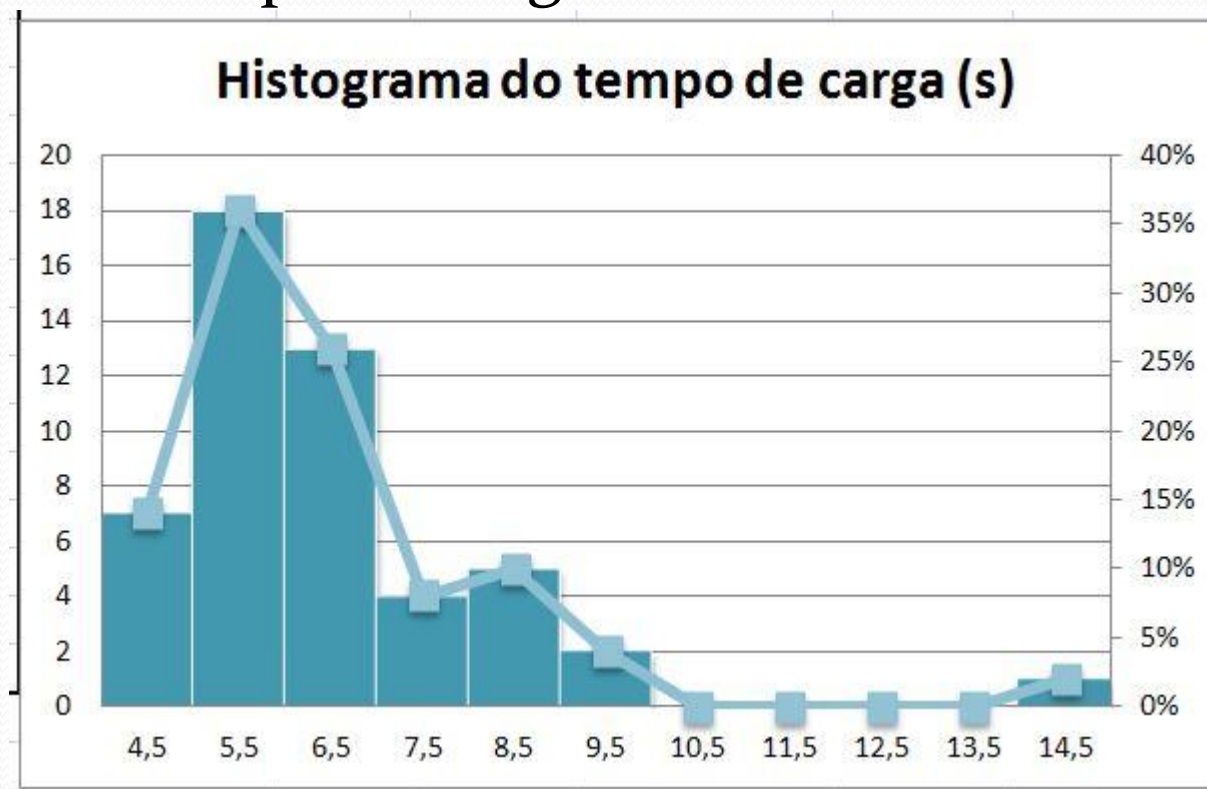
# Polígonos

- Exemplos:



# Exercício:

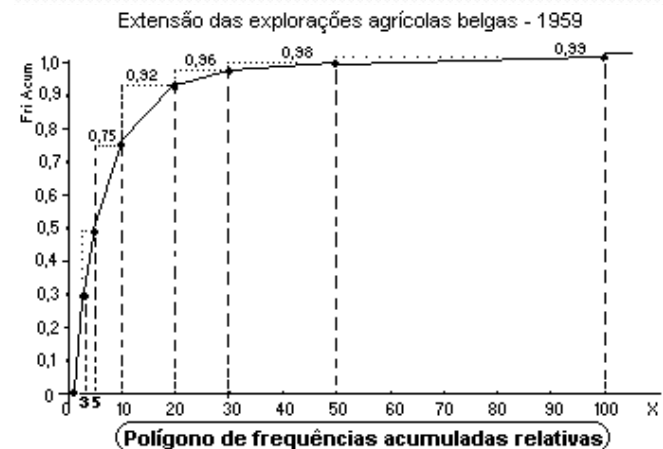
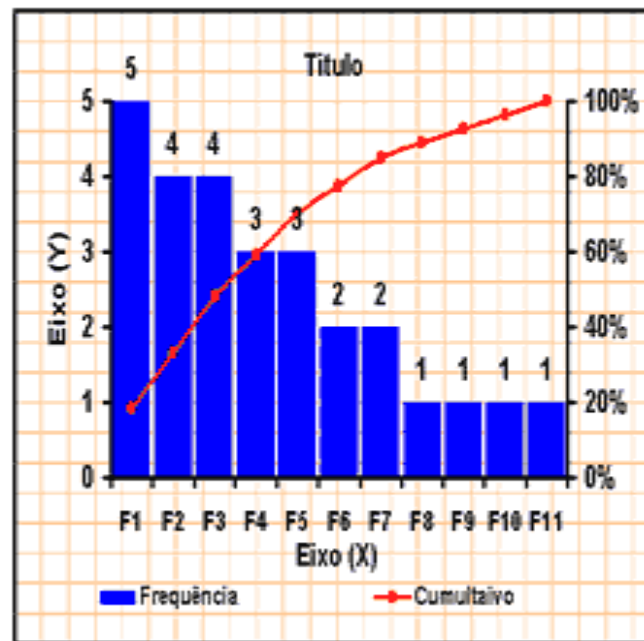
- Construa o polígono de frequências do histograma dos valores do tempo de carga de um sistema:



# Polígono de Percentagens Acumuladas (Ogiva)

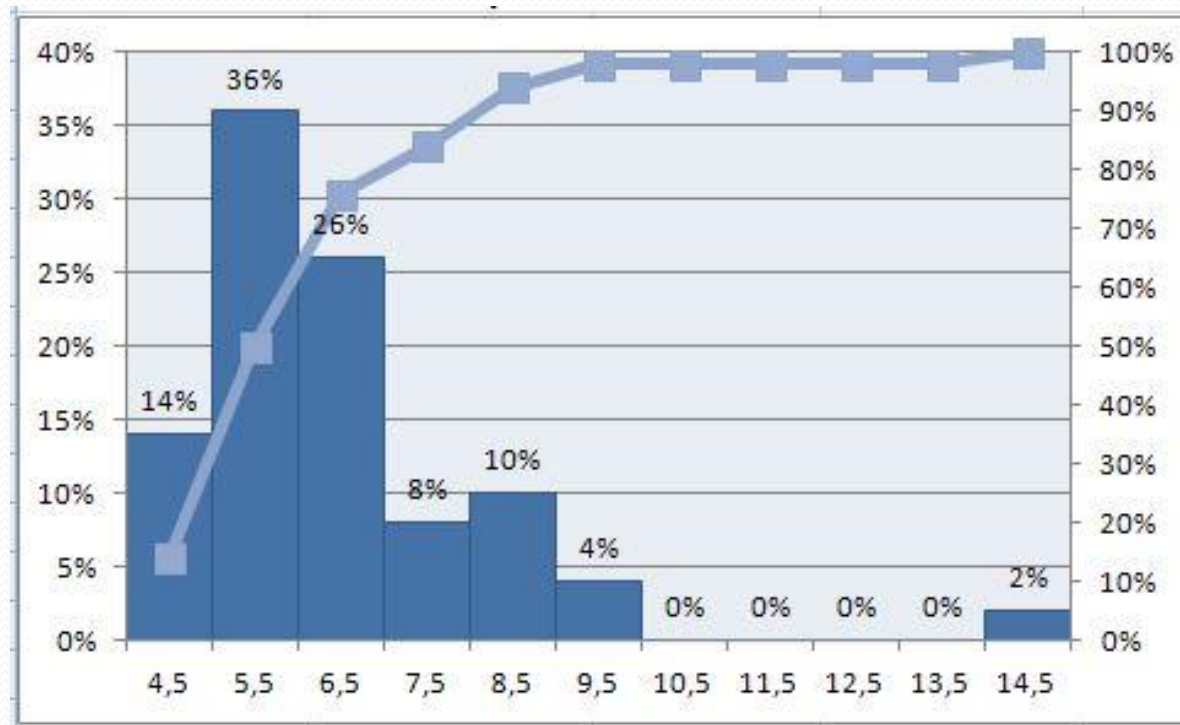
- O **Polígono de Percentagens Acumuladas** ou **Ogiva**, exibe a variável de interesse ao longo do eixo X e a percentagem acumulada ao longo do eixo Y.

Frequência Cumulativa			
F1	5	19%	19%
F2	4	33%	15%
F3	4	48%	15%
F4	3	59%	11%
F5	3	70%	11%
F6	2	78%	7%
F7	2	85%	7%
F8	1	89%	4%
F9	1	93%	4%
F10	1	96%	4%
F11	1	100%	4%
	27		



# Exercício:

- Construa o polígono de frequências acumuladas dos 50 valores dos tempos de carga de um sistema.

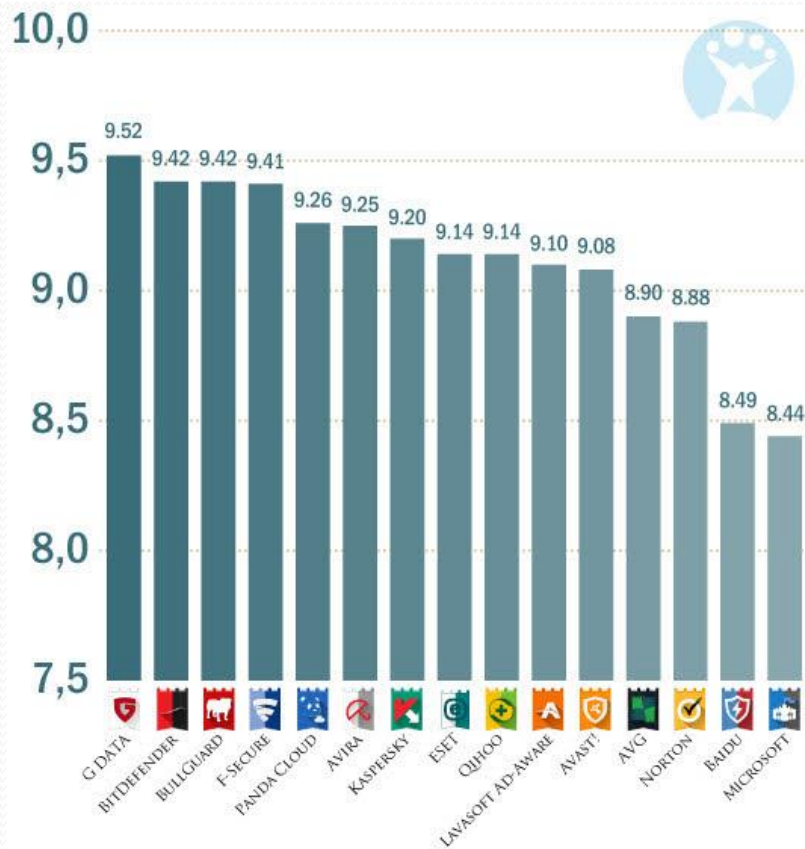


# Uso indevido de gráficos

- Gráficos Pictóricos: São gráficos através de figuras que simbolizam fatos estatísticos, ao mesmo tempo que indicam proporcionalidades. São gráficos que carecem de precisão, mas, como são muito atrativos, são largamente usados em publicidade.
- Regras fundamentais para a construção de um pictograma:
  - Os símbolos devem explicar-se por si próprios.
  - As quantidades maiores são indicadas por meio de um número maior de símbolos, mas não por um símbolo maior.
  - Os símbolos comparam quantidades aproximadas, não detalhes minuciosos.
  - Os gráficos pictóricos só devem ser usados para comparações, nunca para afirmações detalhadas ou isoladas.
- Uso indevido de Gráficos: Podem trazer uma ideia falsa dos dados que estão sendo analisados, chegando mesmo a confundir o leitor. Trata-se, na realidade, de um problema de construção de escalas.

# Uso indevido de gráficos

- Uso de escalas inapropriadas:



## Awarded Best Security For 2011

**Bitdefender Internet Security 2011 & 2012**

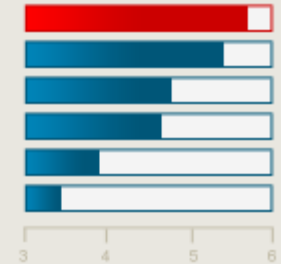
Kaspersky Internet Security 2011 & 2012

Panda Internet Security 2011 & 2012

Norton Internet Security 2011 & 2012

ESET Smart Security 4.2 & 5.0

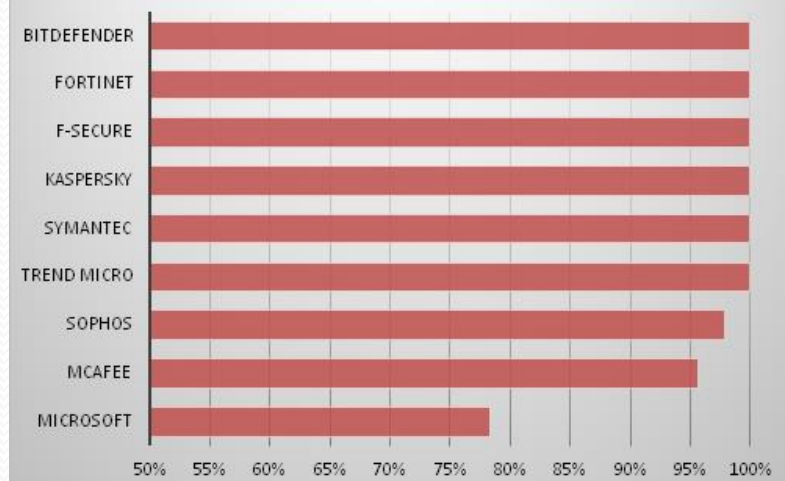
McAfee Total Protection 2011 & 2012



Highest overall score in protection, repair and usability  
Home User category  
<http://www.av-test.org/certifications>

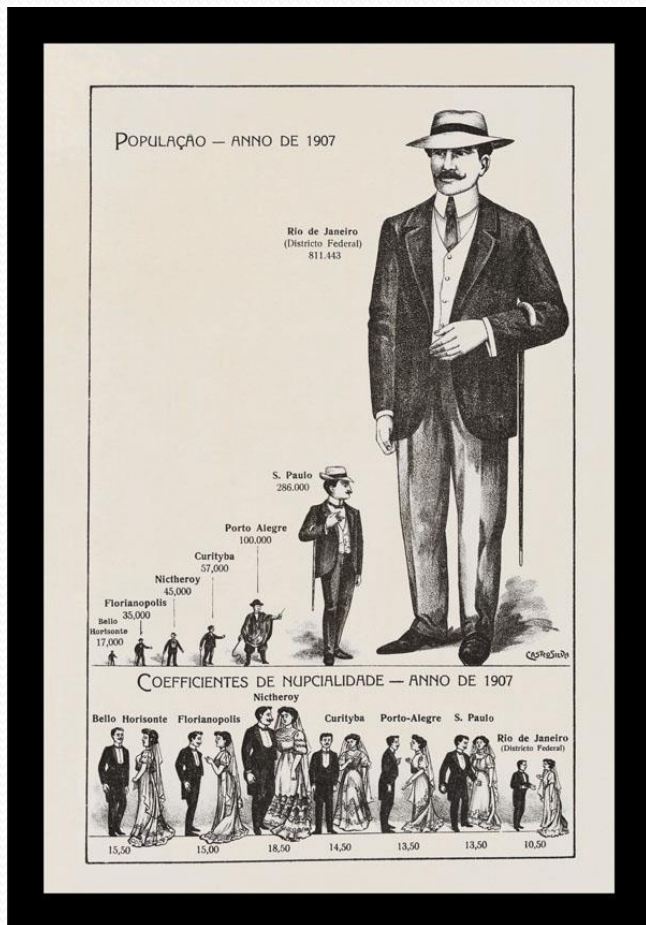
**AVTEST**

## Real World Protection on Windows 8 Corporate Products - December 2013 - av-test.org

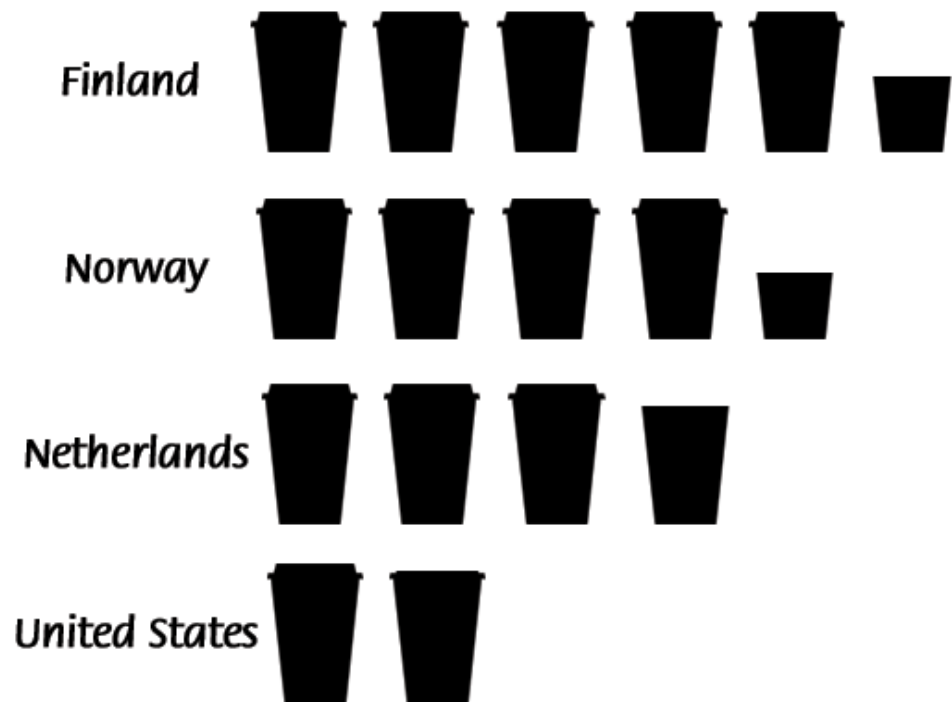




# Gráficos Pictóricos




Average Daily Coffee Consumption  
Around the World Per Capita



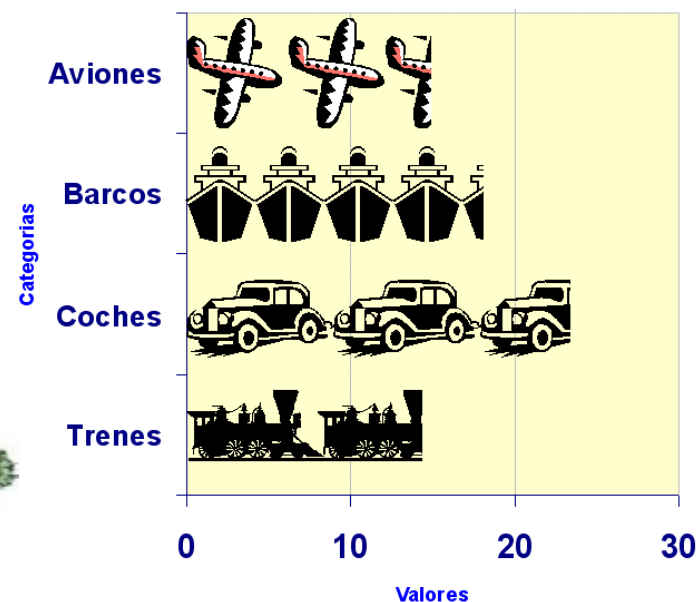
# Gráficos Pictóricos



 = 32 mil hectares de floresta ardida



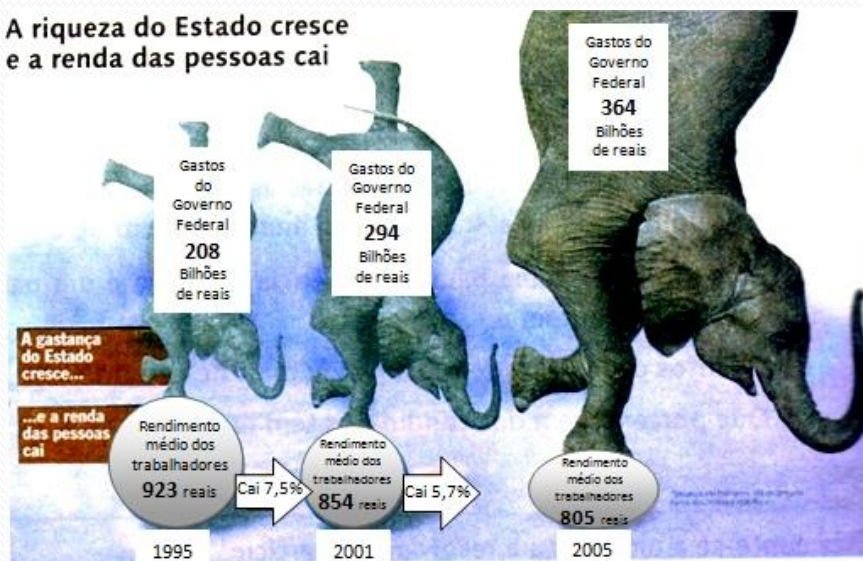
**Pictórico**





# Gráficos Pictóricos

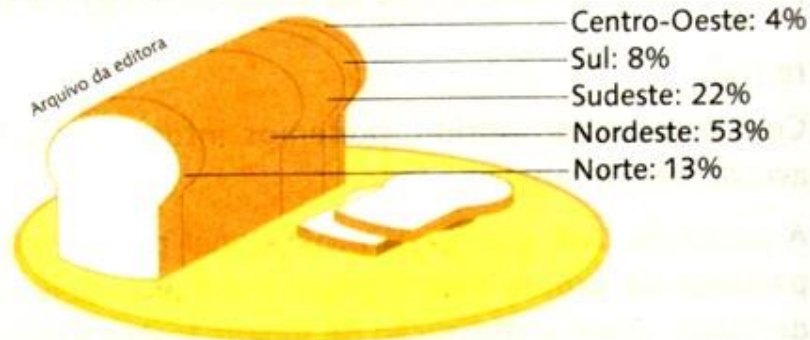
A riqueza do Estado cresce  
e a renda das pessoas cai



VEJA. São Paulo: Abril, ano 39, n. 41, 18 out. 2006. p. 54-5.

Onde o sapato aperta

Em relativos e absolutos, o Nordeste concentra o maior número de pobres do Brasil.



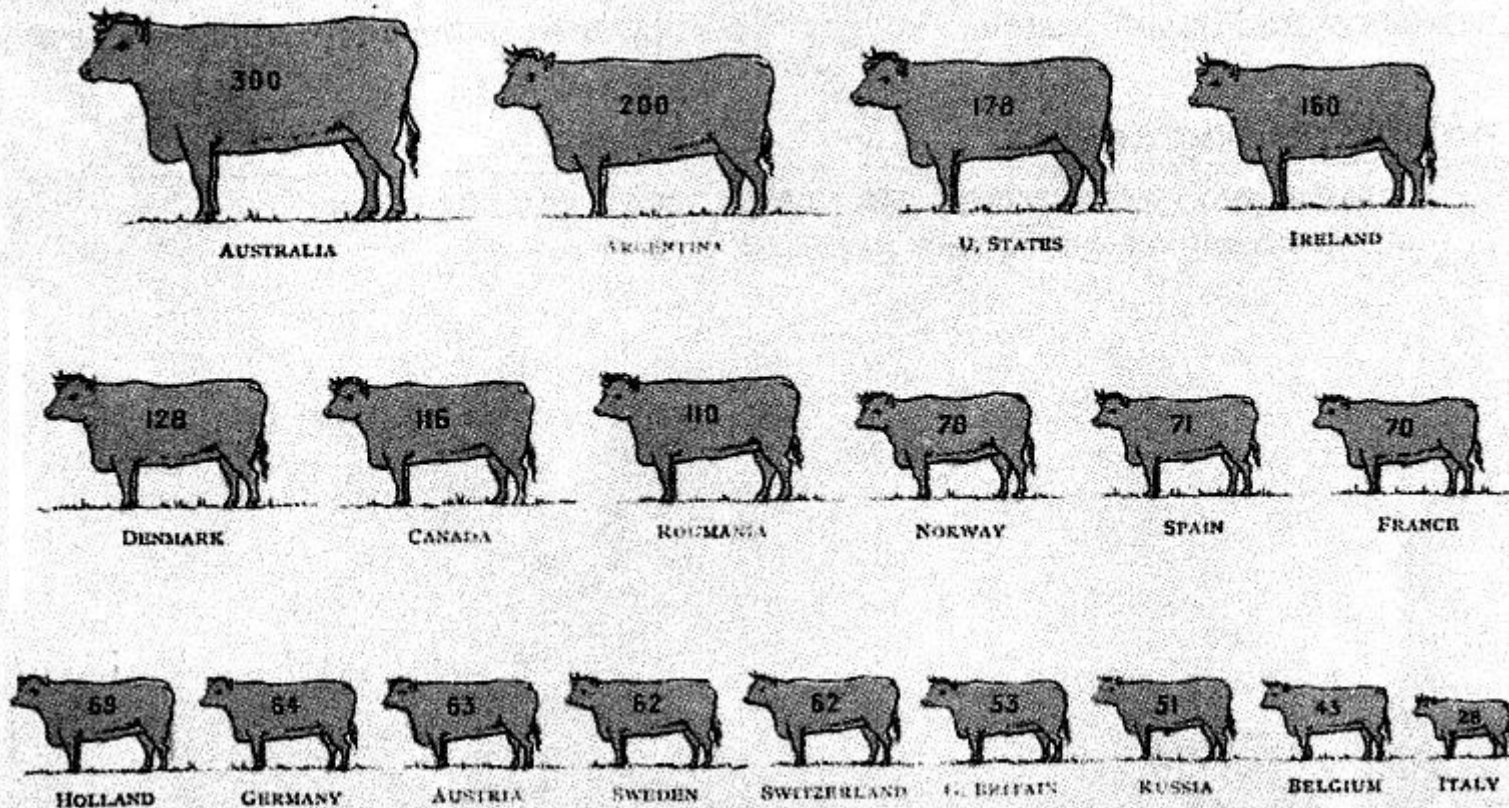
URBIM, Emiliano. Onde estão os pobres do Brasil? *Superinteressante*. São Paulo: Abril, ano 23, n. 1, jan. 2009. p. 36-7.

# Gráficos Pictóricos



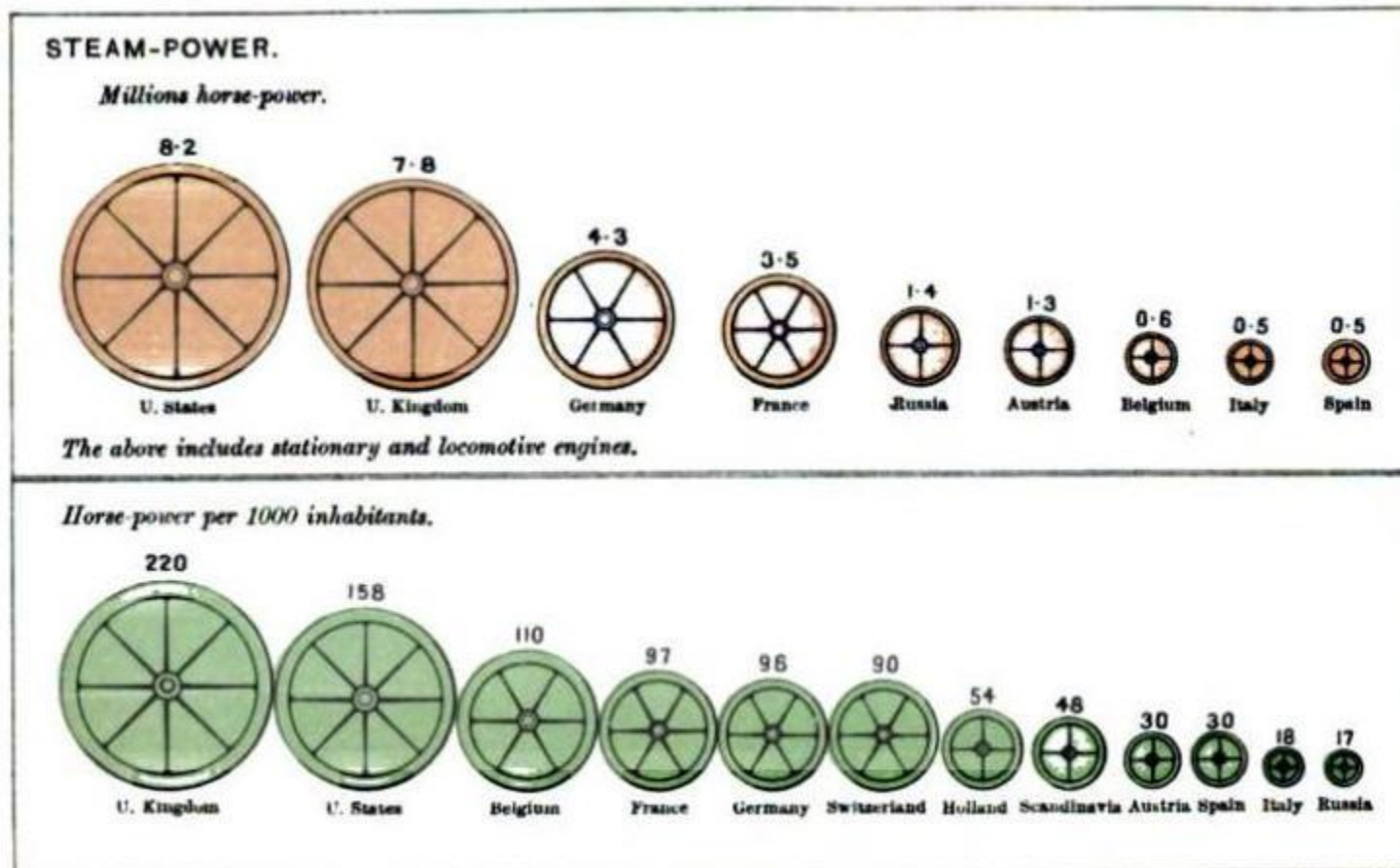
# Gráficos Pictóricos

*Production of Meat, lbs. yearly per Inhabitant.*

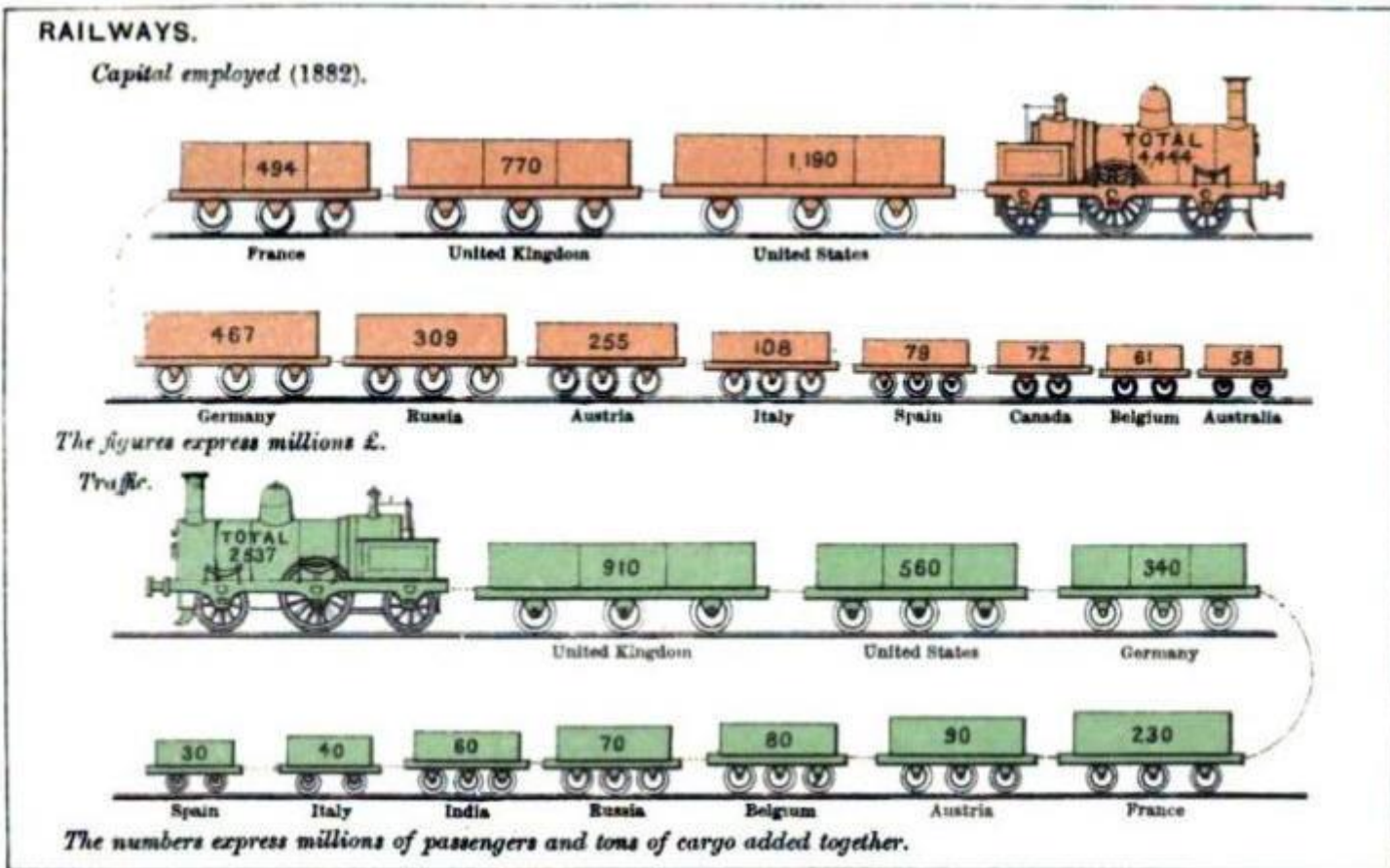




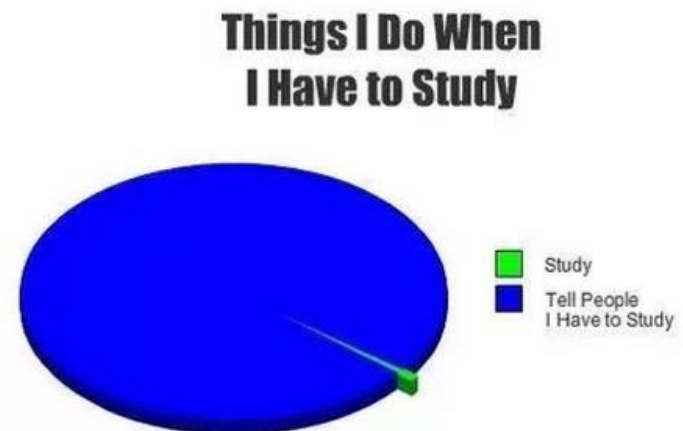
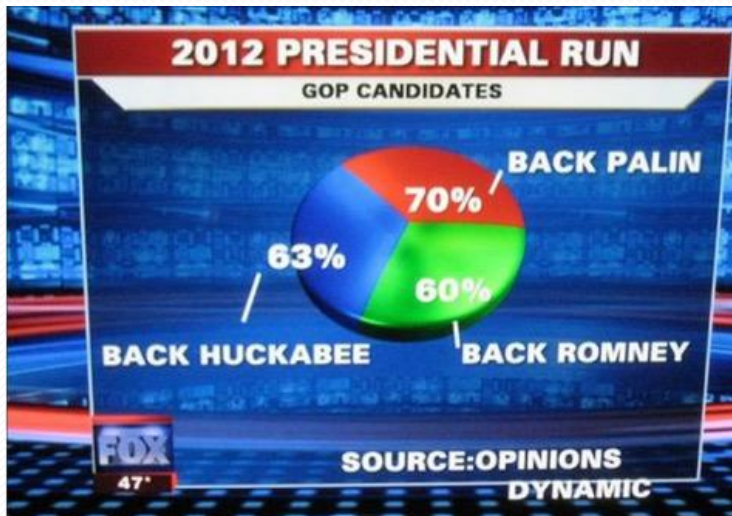
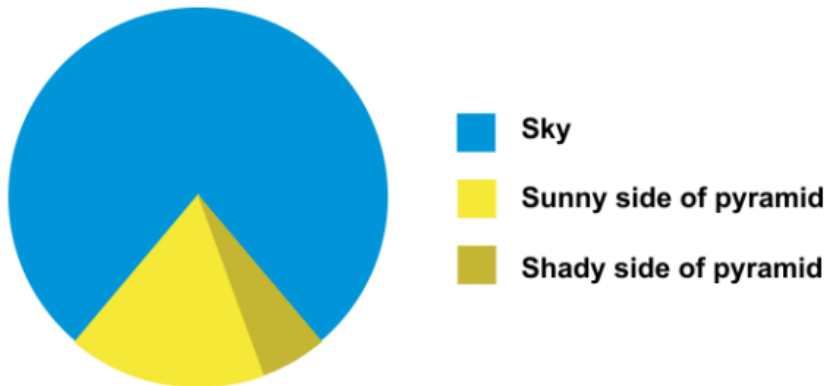
# Gráficos Pictóricos



# Gráficos Pictóricos

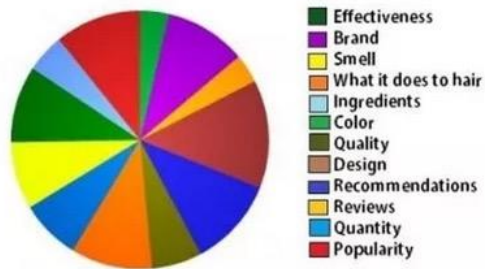


# Gráficos, Gráficos, Gráficos...

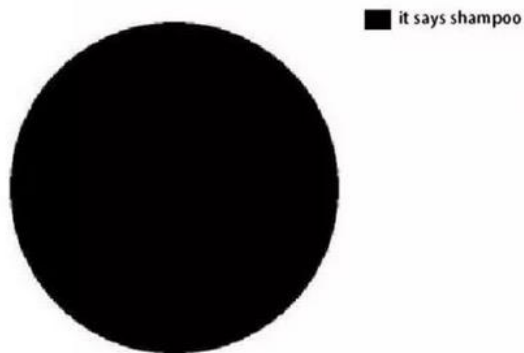


# Gráficos, Gráficos, Gráficos...

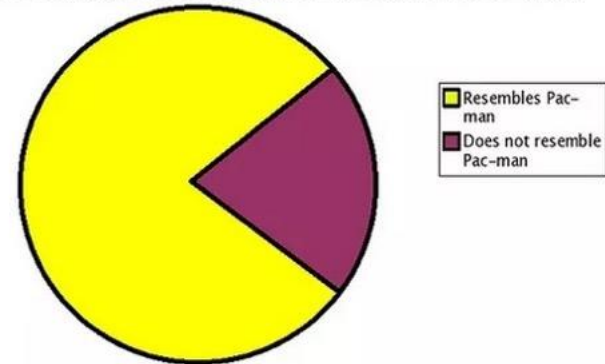
## HOW WOMEN CHOOSE SHAMPOO:



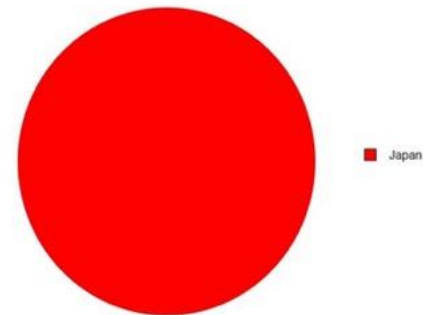
## HOW MEN CHOOSE SHAMPOO:



## Percentage of Chart Which Resembles Pac-man



## JAPAN



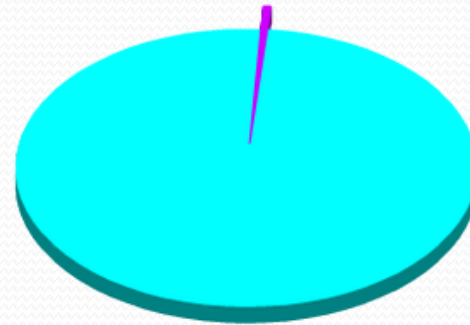
# Gráficos, Gráficos, Gráficos...

**What are you doing right now**



 reading this

**Parts of a Pie-Chart:**



 Big Part

 Tiny Part



# Bibliografia

- Levine, Stephen, Krehbiel e Berensom. Estatística, Teoria e Aplicações usando o Microsoft Excel em Português. Quinta Edição. Ed. LTC.
- Estatística Descritiva Básica – prof. Ilydio Pereira de Sá (<http://www.magiadamatematica.com/uss/pedagogia/estatsticaapostila2.pdf>)
- Imagens da Internet.