

Mais gráficos e representações

Representando em gráficos conjuntos de dados quantitativos →

diagrama de ramo-e-folhas



exemplos da análise exploratória de dados

Exemplo 1

Construindo um diagrama ramo-e-folhas

A seguir, temos os números de mensagens de texto enviadas no mês passado por usuários de telefonia celular em um andar de um dormitório universitário.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

Solução

Em razão de as entradas de dados irem de um número baixo (78) para um número alto (159), você deve usar valores de ramo de 7 a 15. Para construir o diagrama, liste esses ramos à esquerda de uma linha vertical. Para cada entrada de dados, liste uma folha à direita de seu ramo. Por exemplo, a entrada 155 tem um ramo de 15 e uma folha de 5. O diagrama ramo-e-folhas será desordenado. Para obter o diagrama ramo-e-folhas, reescreva o diagrama com folhas em ordem crescente da esquerda para a direita. É importante incluir uma chave para o gráfico para identificar os valores dos dados.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
 118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
 139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
 129 112 126 148 147

Número de mensagens de texto enviadas

7 | 8 Chave: 15|5=155
 8
 9
 10 | 5 8 9 9 9
 11 | 6 4 2 2 8 8 9 3 7 8 9 9 2
 12 | 9 6 2 6 2 1 6 2 6 3 1 4 4 9 6
 13 | 0 9 9 3 4 2 3
 14 | 4 5 2 0 5 8 7
 15 | 5 9

Diagrama ramo-e-folhas desordenado

Número de mensagens de texto enviadas

7 | 8 Chave: 15|5=155
 8
 9
 10 | 5 8 9 9 9
 11 | 2 2 2 3 4 6 7 8 8 8 9 9 9
 12 | 1 1 2 2 2 3 4 4 6 6 6 6 6 9 9
 13 | 0 2 3 3 4 9 9
 14 | 0 2 4 5 5 7 8
 15 | 5 9

Diagrama ramo-e-folhas ordenado

Interpretação

Com o **diagrama ramo-e-folhas** ordenado, você pode concluir que mais de 50% dos usuários de telefone celular enviaram entre 110 e 130 mensagens de texto.

Exemplo 2

Construindo variações do diagrama ramo-e-folhas

Organize os dados fornecidos no Exemplo 1 usando um diagrama de ramo-e-folhas que tenha **duas fileiras** para cada ramo. O que podemos concluir?

Solução

Construa um **diagrama ramo-e-folhas** conforme descrito no Exemplo 1, mas agora liste cada ramo duas vezes. **Use as folhas 0, 1, 2, 3 e 4 na primeira fileira do ramo** e **as folhas 5, 6, 7, 8 e 9 na segunda fileira do ramo.** A seguir, temos o gráfico ramo-e-folhas revisado.

Número de mensagens de texto enviadas

Chave: 15|5=155

7	
7	8
8	
8	
9	
9	
10	
10	5 8 9 9 9
11	4 2 2 3 2
11	6 8 8 9 7 8 9 9
12	2 2 1 2 3 1 4 4
12	9 6 6 6 6 9 6
13	0 3 4 2 3
13	9 9
14	4 2 0
14	5 5 8 7
15	
15	5 9

Diagrama ramo-e-folhas desordenado

Compare os exemplos 1 e 2. Note que, ao usar duas fileiras por ramo, você obtém um quadro mais detalhado dos dados.

Número de mensagens de texto enviadas

Chave: 15|5=155

7	
7	8
8	
8	
9	
9	
10	
10	5 8 9 9 9
11	2 2 2 3 4
11	6 7 8 8 8 9 9 9
12	1 1 2 2 2 3 4 4
12	6 6 6 6 6 9 9
13	0 2 3 3 4
13	9 9
14	0 2 4
14	5 5 7 8
15	
15	5 9

Diagrama ramo-e-folhas ordenado

Você também pode usar um diagrama de pontos para representar dados quantitativos. **Em um diagrama de pontos, cada entrada de dados é representada usando um ponto acima do eixo horizontal.** Como no diagrama ramo-e-folhas, um diagrama de pontos permite que se veja como os dados são distribuídos, determina entradas de dados específicas e identifica valores discrepantes dos dados.

Exemplo 3

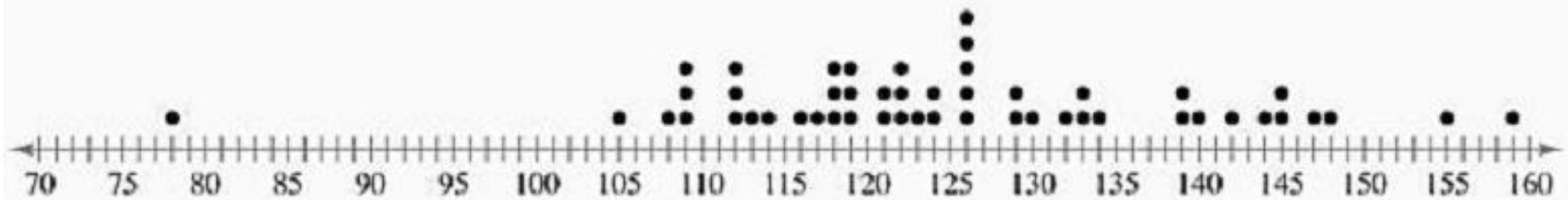
Construindo um diagrama de pontos

Use um diagrama de pontos para organizar os dados das mensagens de texto dados no Exemplo 1.

Solução

De modo que cada entrada de dados esteja incluída no diagrama de pontos, o eixo horizontal deve incluir os números entre 70 e 160. Para representar a entrada de dados, represente um ponto acima da posição da entrada no eixo. Se uma entrada for repetida, represente um outro ponto acima do ponto anterior.

Número de mensagens de texto enviadas



Interpretação

No diagrama de pontos, podemos ver que a maioria dos valores agrupa-se entre 105 e 148 e que o valor que ocorre na maioria das vezes é 126. Você também pode ver que 78 é um valor de dados incomum.

Representando em gráficos conjuntos de dados qualitativos

Gráficos de pizza (setoriais) fornecem uma maneira conveniente de se apresentar graficamente dados qualitativos como porcentagens de um inteiro. **Um gráfico de pizza (setorial) é um círculo dividido em setores que representam categorias. A área de cada setor é proporcional à frequência de cada categoria.** Na maioria dos casos, você interpretará um gráfico de pizza ou irá construí-lo usando tecnologia. O Exemplo 4 mostra como construir um gráfico de pizza à mão.

Exemplo 4

Construindo um gráfico de pizza

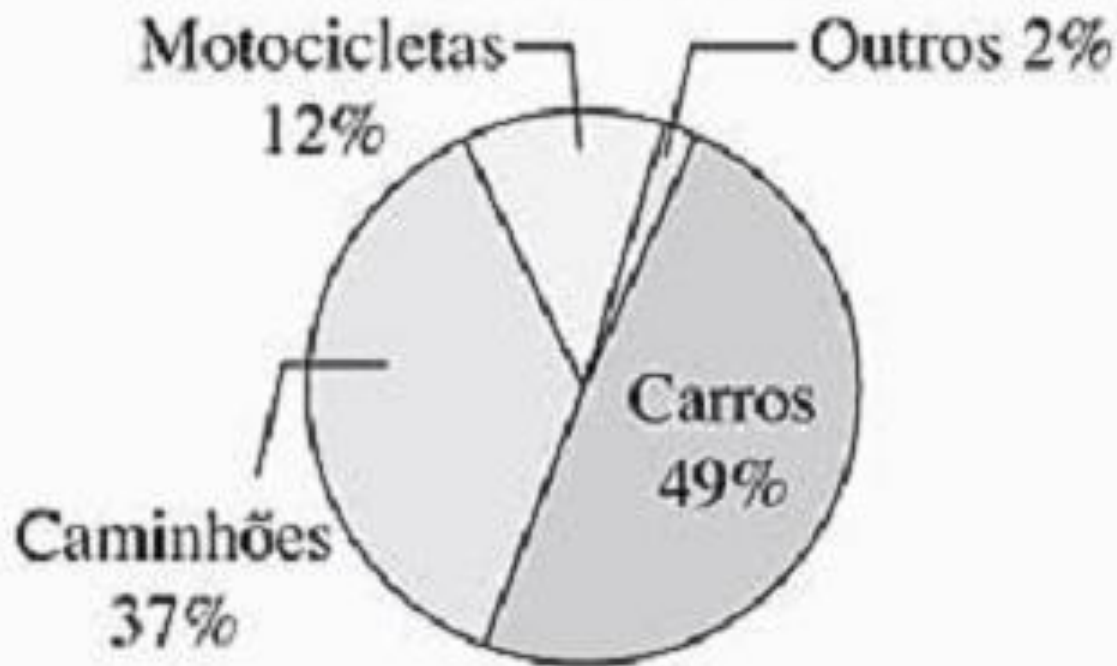
O número de ocupantes de veículos motorizados mortos em acidentes em 2005 é mostrado na tabela ao lado. Use um gráfico de pizza para organizar os dados. O que podemos concluir? (Fonte: U. S. Department of Transportation, National Highway Traffic Safety Administration.)

Tipo de veículo	Mortos
Carros	18.440
Caminhões	13.778
Motocicletas	4.553
Outros	823

Solução

	f	Frequência relativa	Ângulo
Carros	18.440	0,49	176°
Caminhões	13.778	0,37	133°
Motocicletas	4.553	0,12	43°
Outros	823	0,02	7°

Ocupantes de veículos motorizados mortos em 2005



Interpretação

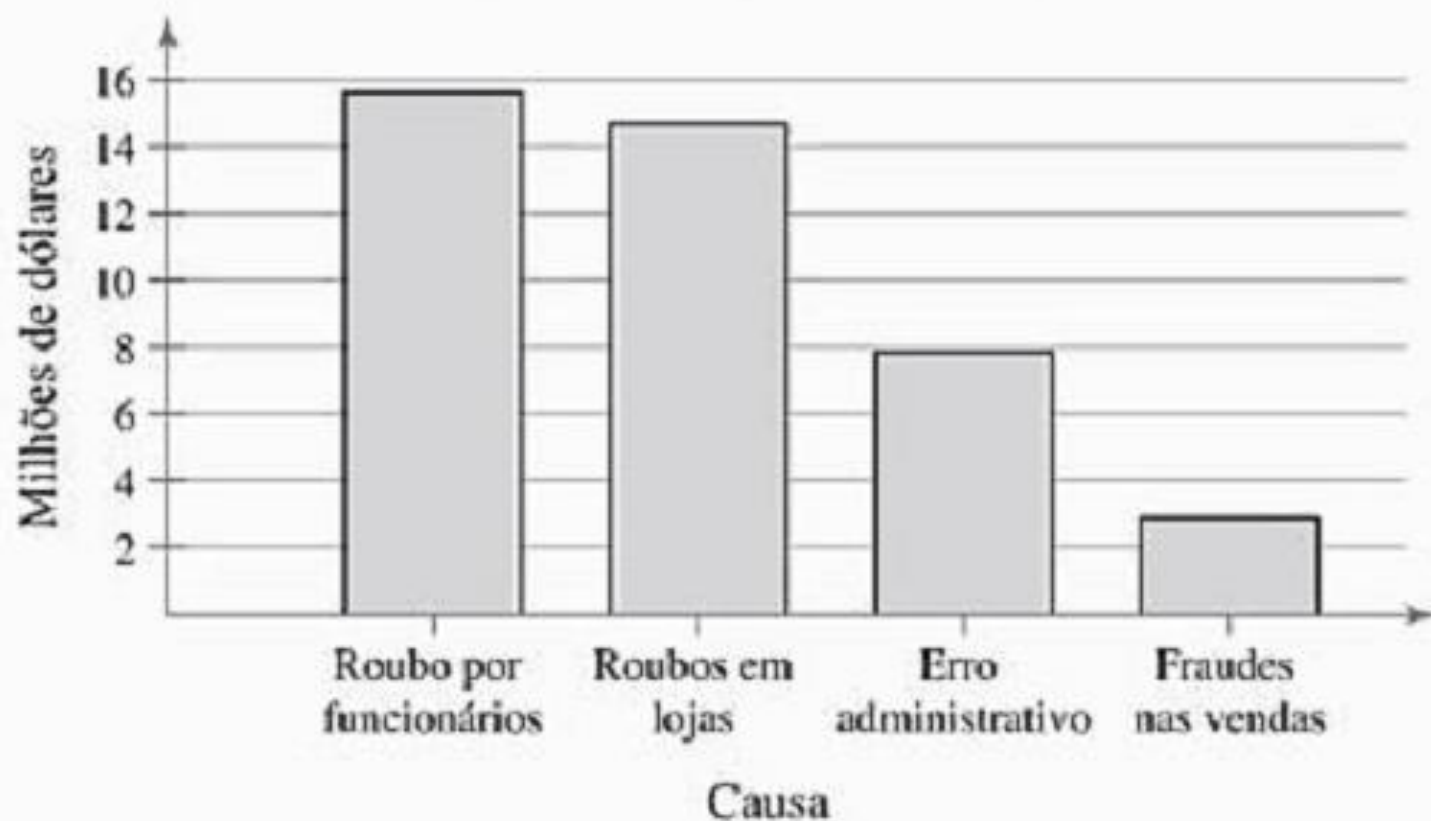
No gráfico de pizza, podemos ver que a maioria das fatalidades em acidentes com veículos motorizados ocorreu com ocupantes de carros.

Outra maneira de representar graficamente dados qualitativos é usando um gráfico de Pareto. Um gráfico de Pareto é um gráfico de barras verticais no qual a altura de cada barra representa a frequência ou a frequência relativa. As barras são posicionadas em ordem de altura decrescente, com a barra mais alta posicionada à esquerda. Tal posicionamento ajuda a enfatizar dados importantes e é frequentemente usado nos negócios.

Construindo um gráfico de Pareto

Recentemente, a indústria de varejo perdeu 41 milhões com redução nos estoques. A redução de estoque é uma perda de estoque por meio de quebra, roubo de carga, roubo em lojas e assim por diante. As causas da redução de estoque são erro administrativo (7,8 milhões), roubos por funcionários (15,6 milhões), roubo em lojas (14,7 milhões) e fraudes nas vendas (2,9 milhões). Se você fosse um varejista, para qual causa de redução de estoques você olharia primeiro? *(Fonte: National Retail Federation and Center for Retailing Education, University of Florida.)*

Causas para redução de estoques



Interpretação

No gráfico, é fácil ver que as causas da **redução de estoque** que **devem ser analisadas primeiramente** são roubo por funcionários e roubo em lojas.

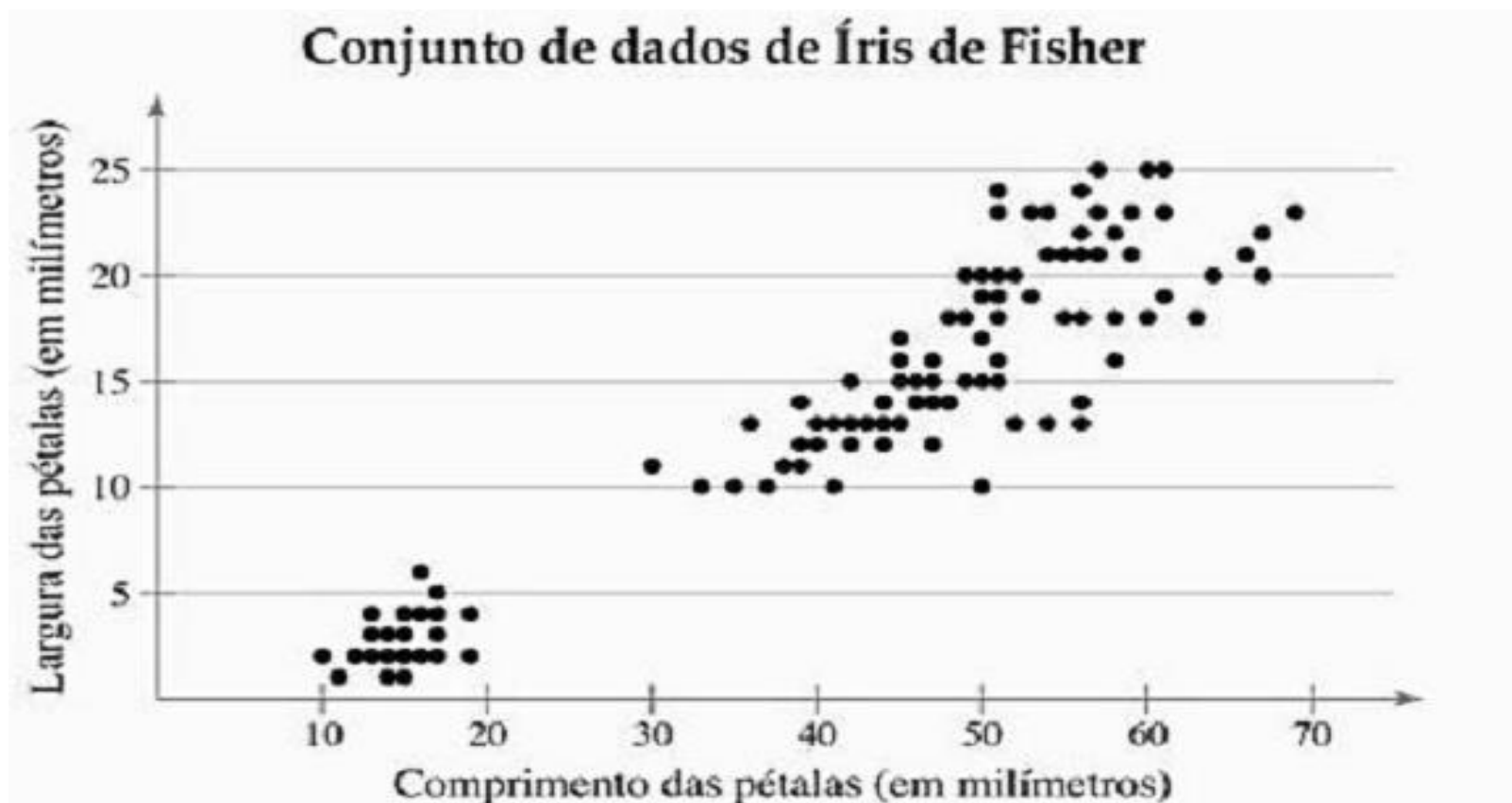
Representando em gráficos conjuntos de dados emparelhados

Se dois conjuntos de dados têm o mesmo número de entradas e cada entrada do primeiro corresponde a uma entrada do segundo, eles são chamados de **conjunto de dados emparelhados**. Por exemplo, suponha que um conjunto de dados contenha os custos de um item e um segundo conjunto de dados contém a quantidade de vendas para o item para cada custo. Já que cada custo corresponde a uma quantidade de vendas, os conjuntos de dados são emparelhados. Uma maneira de representar conjuntos de dados emparelhados é usando um gráfico de dispersão, no qual os pares ordenados são representados como pontos em um plano coordenado. Um gráfico de dispersão é usado para mostrar a relação entre duas variáveis quantitativas.

Exemplo 6

Interpretando um gráfico de dispersão

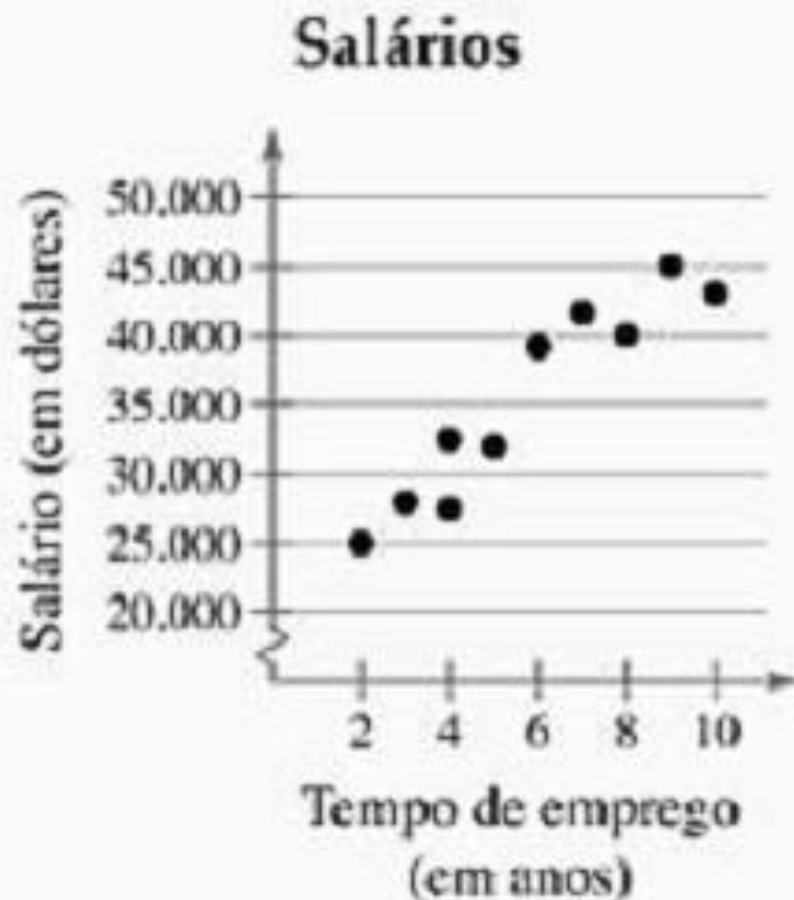
O estatístico britânico Ronald Fisher (ver p. 28) apresentou um famoso conjunto de dados chamado de conjunto de dados de Íris de Fisher. Esse conjunto de dados descreve várias características físicas tais como o comprimento de pétalas e a sua largura (em milímetros), para três espécies de íris (flor). No gráfico de dispersão mostrado, os



O período de emprego e os salários de 10 funcionários estão listados na tabela ao lado. Faça o gráfico para os dados usando um gráfico de dispersão. O que podemos concluir?

- Classifique os eixos *horizontal e vertical*.
- Represente graficamente os dados emparelhados.
- Descreva quaisquer tendências.

Tempo de emprego (em anos)	Salário (em dólares)
5	32.000
4	32.500
8	40.000
4	27.350
2	25.000
10	43.000
7	41.650
6	39.225
9	45.100
3	28.000



Você aprenderá mais sobre **gráficos de dispersão** e como analisá-los no **Capítulo 9**.

Um conjunto de dados constituído de entradas de dados tomados a intervalos regulares durante um período de tempo é chamado de **série temporal**. Por exemplo, a quantidade de precipitação medida a cada dia em um mês é um exemplo de uma série temporal. Você pode usar um **gráfico de série temporal** para representar uma série desse tipo.

Exemplo 7

Construindo um gráfico de série temporal

A tabela lista o número de usuários de telefonia celular (em milhões) e conta média de um usuário local por serviço (em dólares) para o ano de 1995 até o ano de 2005. Construa um gráfico de série temporal para o número de usuários de celulares. O que você pode concluir? (Fonte: Cellular Telecommunication & Internet Association.)

<i>Ano</i>	Usuários (em milhões)	Conta média (em dólares)
1995	33,8	51,00
1996	44,0	47,70
1997	55,3	42,78
1998	69,2	39,43
1999	86,0	41,24
2000	109,5	45,27
2001	128,4	47,37
2002	140,8	48,40
2003	158,7	49,91
2004	182,1	50,64
2005	207,9	49,98

Solução

Deixe o eixo horizontal representar os anos e o vertical representar o número de usuários (em milhões). Então, represente os dados emparelhados e conecte-os com segmentos de linha.



Interpretação

Os gráficos mostram que o número de usuários tem aumentado desde 1995, com os maiores aumentos tendo ocorrido recentemente.