

# TYPE GRAPHIC

DANILO MACIEL



"Os gráficos de uma forma que  
ninguem nunca te contou"

## ENTENDENDO OS GRÁFICOS

Sem sombras de dúvidas você já se perguntou:

- "para este Insight, qual gráfico usar?"

aqui você entenderá que para cada insight existe uma aplicação recomendada.

Por isso usar o gráfico adequado para transmitir a ideia apresentada é mais que essencial.

## TIPOS DE GRÁFICOS

Existem mais de 500 tipos de gráficos no mundo, alguns exclusivos para regras de negócio específica, nem sempre o "GRÁFICO BONITINHO" lhe trará a ideia que deseja passar.

Linhas, colunas, tabelas, dispersão, inclinação, bolhas, aranhas, boxplot, são apenas uma parte dos gráficos que iremos estudar.

## QUANDO E COMO USA-LOS

Iremos entender os principais gráficos usados no mercado, seus conceitos, sua aplicação correta, quando e como usá-los, e como não usa-los.

Em qual cenário o gráfico X ou Y é o mais indicado? Sempre se atentando aos conceitos, e nunca esquecendo que o seu gráfico deve ser claro e objetivo na apresentação do insight.

# RESUMO

Este pequeno resumo lhe dará embasamento TÉCNICO, CONCEITUAL, EXPLANATÓRIO e principalmente em sua usabilidade pratica, em nossos relatórios é normal nos depararmos com as dúvidas do tipo – “Que tipo de gráfico devo usar agora” não conhecer certos tipos de gráficos é absolutamente normal, é como se eu te perguntasse; - “Para que ser o gráfico de BOXPLOT”? alguns tipos de gráficos são exclusivos para certas visualizações e por não fazer parte do nosso dia a dia acaba batendo as dúvidas.

Iremos entender os principais gráficos usados atualmente, segunda a referência do livro *storytellingwithdata* existem cerca de mais de 600 tipos de gráficos e alguns deles usados exclusivamente para regras de negócios específicas.

Este material tem como referência o livro *storytellingwithdata* da INCRIVEL Cole Nussbaumer.

Cole, é autora de "contar histórias com dados: um guia de visualização de dados para profissionais de negócios" (Wiley, novembro de 2015) e escreve o popular blog [www.storytellingwithdata.com](http://www.storytellingwithdata.com). Suas oficinas e apresentações são altamente procuradas por indivíduos, empresas e organizações filantrópicas dedicadas a dados em todo o mundo. Seu talento único foi aprimorado ao longo da última década através de papéis analíticos em bancos, private equity e, mais recentemente, como gerente da equipe Google People Analytics. No Google, ela usou uma abordagem baseada em dados para informar programas e práticas de gerenciamento de pessoas inovadoras, garantindo que o Google atraiu, desenvolveu e manteve grandes talentos e que a organização estava melhor alinhada para atender às necessidades do negócio. Cole viajou para os escritórios do Google em todo os EUA e na Europa para ensinar o curso que ela desenvolveu na visualização de dados. Ela também atuou como membro adjunto do corpo docente da MICA, onde ensinou Introdução à Visualização de Informações.

" O sucesso da sua apresentação de relatório está 100% ligada a boa organização do storytelling"



"Este material que iremos estudar é designado para profissionais que não possuem conhecimento prévio em análises de dados, porem iremos nos aprofundar no conhecimento técnico e conceitual, mesmo que você já seja um profissional da área de dados vale a pena explorar alguns pontos, temos certeza que muitos deles você desconhecia".



## Sumário

Conhecendo os tipos de gráficos.....	5
o que é um gráfico de linha? .....	6
CENÁRIOS IDEAL PARA UTILIZAÇÃO DO GRAFICO DE LINHA .....	6
CENÁRIOS NÃO IDEAL PARA UTILIZAÇÃO DO GRAFICO DE LINHA.....	6
Como funciona um gráfico de linhas? .....	6
Que tipo de dados podem ser usados em um gráfico de linha? .....	8
Quantas linhas diferentes posso colocar em um gráfico? .....	8
Outras formas de visualização (Gráfico de Linhas) .....	9
O que é um gráfico de Barras ou Colunas? .....	9
Quais são os tipos comuns de gráficos de barras? .....	9
Quais são as dicas para criar gráficos de barras?.....	10
Outras formas de visualização (Gráfico de Barras) .....	11
O que é um gráfico de área? .....	12
Quando os gráficos de área são úteis? .....	13
Outras formas de visualização (Gráfico de Área) .....	14
O que é um gráfico de pizza? .....	15
Existem dois casos de uso principais para um gráfico de pizza: .....	15
Quero usar uma torta: como faço uma boa? .....	17
Outras formas de visualização (Gráfico de pizza) .....	18
O que é um gráfico de Dispersão? .....	19
Quais são as dicas para projetar gráficos de dispersão? .....	23
Outras formas de visualização (Gráfico de Dispersão) .....	25
O que é um gráfico de BOLHAS?.....	25
Como você interpreta um gráfico de bolhas? .....	26
Como saber quando usar um gráfico de bolhas?.....	27
Quais são as dicas de design para criar um gráfico de bolhas eficaz?.....	27
Outras formas de visualização (Gráfico de Bolhas) .....	27
O que é um gráfico de inclinação?.....	28
Dados categóricos: destacando diferenças entre duas categorias .....	28
Quais são algumas razões para não usar um gráfico de inclinação? .....	29
Outras formas de visualização (Gráfico de Inclinação).....	29
Como usar gráficos de TABELA? .....	30
Quando você não deve usar uma tabela para comunicar dados?.....	30
Como você pode projetar uma mesa melhor? .....	31
Outras formas de visualização (Tabelas).....	32
O que é um gráfico de PONTOS? .....	33
Como faço um gráfico de pontos? .....	37
Outras formas de visualização (Gráfico de Pontos) .....	37
O que é um gráfico de ARANHA?.....	38



Coordenadas cartesianas .....	38
Coordenadas polares .....	38
Qual é a diferença entre um gráfico de radar e um gráfico de aranha? .....	39
Quando (e como) devo usar gráficos de aranha de forma eficaz? .....	39
Coisas a ter em mente ao usar um gráfico de aranha Reduza o desafio de comparar coisas por área .....	42
Em um ambiente ao vivo, apresente seu público a este gráfico gradualmente .....	42
Outras formas de visualização (Gráfico Aranha).....	43



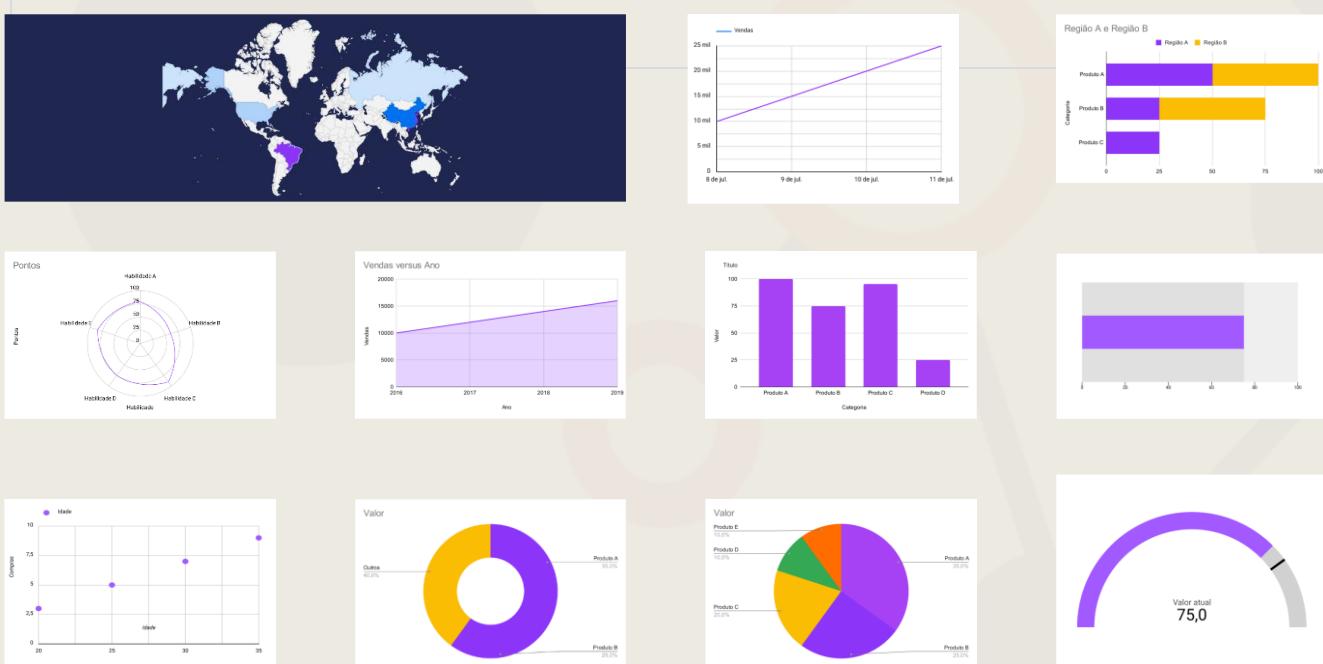
## Leia com atenção

A compreensão dos diferentes tipos de gráficos **desempenha um papel fundamental na interpretação e comunicação eficaz de dados**. Os gráficos representam visualmente informações complexas, permitindo que as pessoas assimilem dados de maneira rápida e compreensível. A variedade de gráficos disponíveis, como barras, linhas, pizza e dispersão, oferece uma gama diversificada de ferramentas para ilustrar padrões, tendências e relações nos dados.

Ao conhecer os tipos de gráficos, os profissionais em diversas áreas podem escolher a representação visual mais apropriada para os dados específicos que desejam transmitir. Por exemplo, um gráfico de barras pode ser ideal para comparar quantidades entre diferentes categorias, enquanto um gráfico de linhas é eficaz para mostrar tendências ao longo do tempo. Já um gráfico de pizza pode destacar a proporção de partes em relação a um todo.

Além disso, a familiaridade com diferentes tipos de gráficos facilita a interpretação precisa dos dados, evitando interpretações equivocadas. Cada tipo de gráfico destaca aspectos específicos dos dados, e a escolha correta contribui para uma comunicação mais clara e eficiente.

A habilidade de criar e interpretar gráficos é valiosa em vários contextos, desde apresentações empresariais até relatórios acadêmicos. Ela não apenas simplifica a complexidade dos dados, mas também capacita as pessoas a tomar decisões informadas com base em informações visualmente acessíveis. Em resumo, o conhecimento dos tipos de gráficos é uma habilidade essencial para aqueles que buscam comunicar efetivamente informações quantitativas e qualitativas.



# Conhecendo os tipos de gráficos

*“O sucesso de muitas visualizações depende de uma sólida compreensão dos conceitos básicos.”*

## o que é um gráfico de linha?

Os gráficos de linhas (ou gráficos de linhas) são melhores quando você deseja mostrar **como o valor de algo muda ao longo do tempo** ou comparar como várias coisas mudam ao longo do tempo em relação umas às outras. Sempre que você ouvir a **frase-chave “ao longo do tempo”**, essa é a sua pista para considerar o uso de um gráfico de linhas para seus dados.

## CENÁRIOS IDEAL PARA UTILIZAÇÃO DO GRAFICO DE LINHA

- Comparando **muitos dados** de uma só vez
- Mostrando **mudanças e tendências** ao longo do tempo
- Incluindo **contexto e anotação importantes**
- Exibindo **dados de previsão** e incerteza
- Destacando **anomalias** dentro e entre as séries de dados

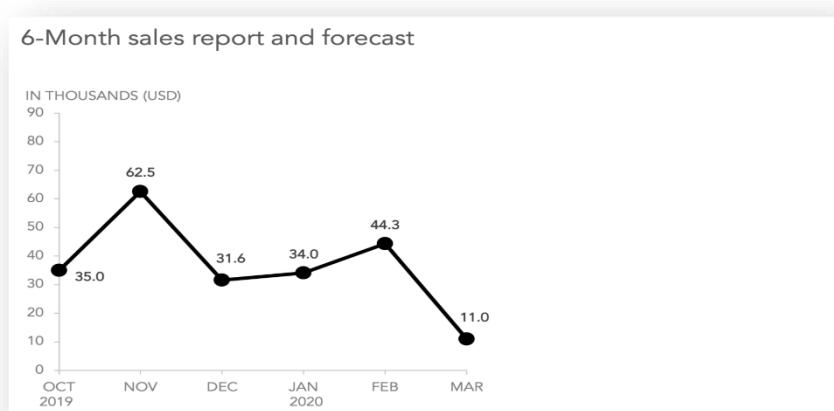
## CENÁRIOS NÃO IDEAL PARA UTILIZAÇÃO DO GRAFICO DE LINHA

- Exibindo **quantidades** de coisas
- Trabalhando com dados **categóricos**
- Fazendo comparações **de parte para todo**
- Mostrando conjuntos de dados **esparsos**

## Como funciona um gráfico de linhas?

Pense nas linhas em um gráfico de linhas como pequenas histórias.

- Primeiro, **medimos o valor de algo que queremos acompanhar em diferentes pontos no tempo**. Por exemplo, no final de cada mês, nossa loja deseja registrar quanto dinheiro arrecadamos com as vendas.
- Assim que tivermos muitas dessas medições, **traçamos esses pontos em nosso gráfico e, em seguida, desenhamos uma linha que conecta esses pontos**, para facilitar a visualização de como esses valores aumentaram, diminuíram ou permaneceram os mesmos ao longo do tempo.





Mas isso não é tudo que podemos fazer com gráficos de linha!

Além de traçar essa linha, podemos incluir outras informações contextuais importantes em nosso gráfico. Podemos optar por **marcar, rotular e/ou anotar** pontos individuais em nossa linha; podemos adicionar **linhas de referência e regiões**; podemos até incluir pontos de dados de previsão e **distinguir entre dados reais e projetados** alterando a aparência de nossa linha e nossos marcadores de dados .

Aqui está a aparência de um gráfico como esse.



Nossos dados de vendas coletados ainda são mostrados no gráfico, como a linha preta sólida, mas agora conseguimos colocar esses dados em contexto.

- Uma faixa de referência cinza mostra qual foi o **intervalo dos totais de vendas mensais** e também **rotulamos esses valores específicos** .
- 
- Uma linha verde nos mostra qual é nossa **meta de vendas** , para facilitar a visualização se vale a pena comemorar nossos números de vendas em um determinado mês.
- Para reduzir alguma confusão visual, **os rótulos de dados e a maioria dos marcadores de pontos individuais em nossa linha foram removidos**. Ficam apenas os marcadores dos meses em que atingimos nossos valores máximo e mínimo de vendas, pois são importantes para a história que pretendemos contar com este gráfico.
- Uma linha pontilhada mostra as **vendas projetadas** para os próximos três meses.
- **Anotações** em momentos-chave no tempo explicam o raciocínio dos números, ações que tomamos em resposta a esses números e nossas expectativas.

**Lembre-se:** os gráficos de linha enfatizam como nossos valores medidos mudam ao longo do tempo ou conforme uma variável independente especificada muda. Podemos aproveitar o espaço em branco em nossos gráficos de linhas para incluir um contexto informativo que ajude o público a entender por que essas mudanças específicas são importantes.



## Que tipos de dados podem ser usados em um gráfico de linha?

Um gráfico de linha típico terá **dados contínuos** ao longo das dimensões vertical (eixo y) e horizontal (eixo x). O eixo y geralmente mostra o **valor de qualquer variável que estamos medindo**; o eixo x é usado com mais frequência para mostrar **quando o medimos, cronologicamente** ou com base em alguma **variável independente** (por exemplo, conforme aceleramos o motor de nosso carro antigo, medimos o nível de decibéis em diferentes RPM).

Embora alguns gráficos de linha não usem dados contínuos no eixo x (particularmente gráficos de inclinação e diagramas de coordenadas paralelas, que são variantes especializadas de gráficos de linha), o que absolutamente **não podemos** usar em nosso eixo x são dados que não têm qualquer relação significativa entre as categorias mostradas.

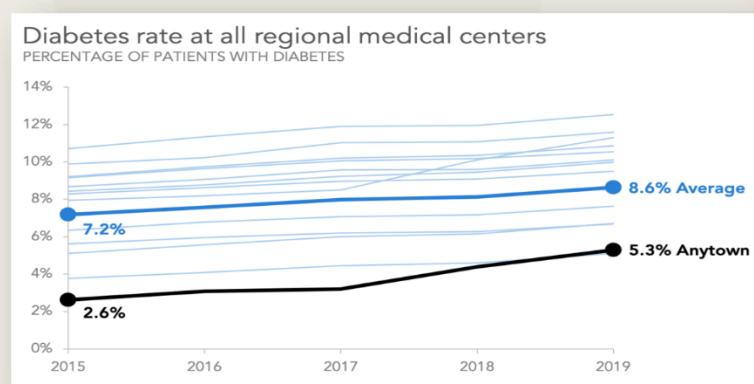


**Gráficos de linha só fazem sentido quando há uma relação significativa entre pontos sucessivos na linha.**

Portanto, uma mercearia poderia razoavelmente usar um gráfico de linhas para mostrar seus dados de vendas de produtos por mês, mas usar um para mostrar as diferenças de vendas entre vários tipos de produtos **não é recomendado**.

## Quantas linhas diferentes posso colocar em um gráfico?

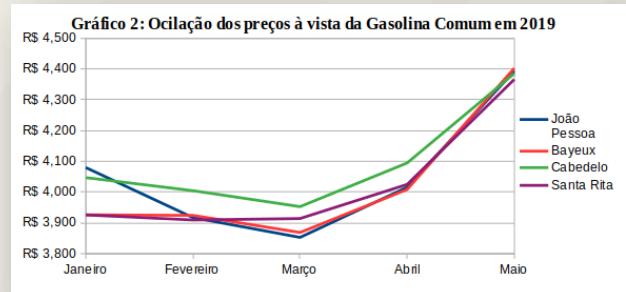
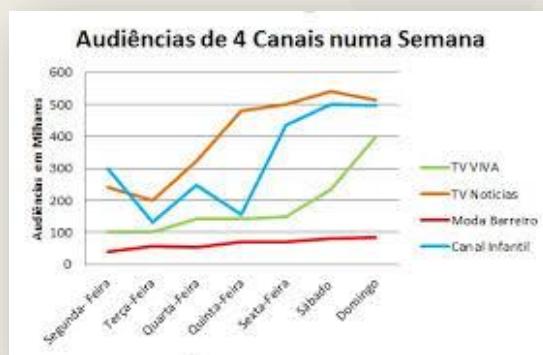
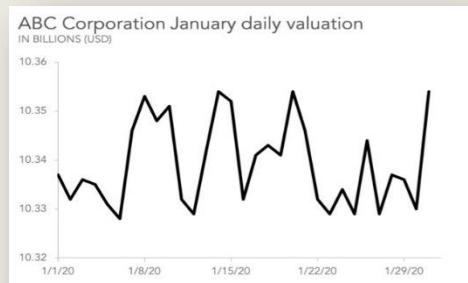
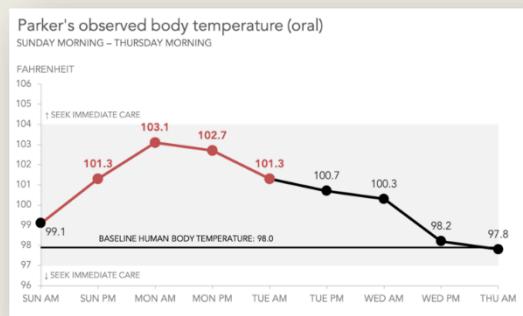
As pessoas têm dificuldade em acompanhar mais de **quatro ou cinco coisas** simultaneamente, então considere usar isso como sua diretriz. Se você realmente precisar mostrar mais linhas do que isso, tente usar cor, espessura de linha e rotulagem para focar nas linhas ou pontos mais importantes.



Este gráfico acima mostra cinco anos de dados para taxas anuais de diabetes de pacientes em nosso centro médico local, Anytown Medical, bem como as taxas de outros 12 centros em nossa região. Como provavelmente o que mais nos preocupa é nosso próprio centro (principalmente porque nossas taxas são baixas), estamos usando cor, peso e rotulagem para trazer o foco para Anytown e para a média regional, ao mesmo tempo em que incluímos os valores para todos outros centros.



## Outras formas de visualização (Gráfico de Linhas)



## O que é um gráfico de Barras ou Colunas?

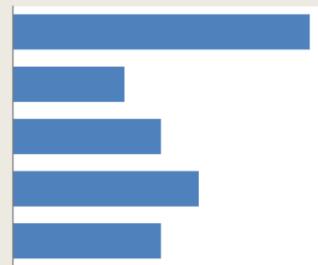
Um gráfico de barras é um tipo de gráfico comum para representar graficamente dados categóricos ou dados classificados em grupos. Consiste em vários retângulos alinhados a uma linha de base comum. O comprimento de cada um é proporcional ao valor que representa — em outras palavras, em um gráfico de barras, os dados são codificados por comprimento. Nossos olhos são muito bons em comparar comprimentos quando os objetos estão alinhados, tornando este gráfico fácil de interpretar - apenas um dos motivos pelos quais os gráficos de barras são comuns.

## Quais são os tipos comuns de gráficos de barras?



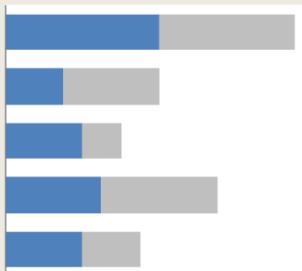
### GRÁFICOS DE BARRAS VERTICAIS

Barras verticais padrão (também conhecidas como gráficos de colunas) são uma das variedades mais comuns, e praticamente todas as ferramentas possuem funcionalidades integradas para criá-las. Só porque são comuns não significa que sejam chatos. Leia mais sobre como as barras verticais básicas podem ser usadas para comunicar várias histórias.



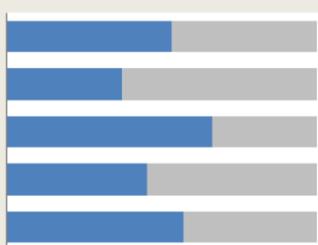
### GRÁFICOS DE BARRAS HORIZONTAIS

As barras horizontais são semelhantes à versão vertical virada de lado. Essa é uma ótima opção para nomes de categoria longos porque o espaço no lado esquerdo permite texto posicionado horizontalmente (uma alternativa ao texto diagonal no eixo x de um gráfico de barras vertical). As barras horizontais são confortáveis de ler, pois a estrutura imita como processamos as informações (nas culturas ocidentais), onde lemos da esquerda para a direita, começando no topo. Confira este artigo para uma reforma simples de barra horizontal.



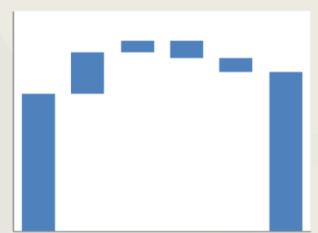
### BARRAS EMPILHADAS

barras empilhadas vêm em formas verticais e horizontais. Eles permitem que você compare os totais digitalizando as extremidades, mas também mostram as peças do subcomponente dentro de cada categoria. O desafio com as barras empilhadas é que é difícil comparar os subcomponentes em categorias que não estão alinhadas com a mesma linha de base. Aqui está um post com mais exemplos de quando as barras empilhadas funcionam.



### BARRAS 100% EMPILHADAS

Uma alternativa às barras empilhadas padrão é um gráfico de barras 100% empilhadas (variação horizontal mostrada à esquerda), onde as porcentagens relativas são exibidas em vez dos valores absolutos. O benefício desta versão é que ela oferece duas linhas de base consistentes para comparação: uma à esquerda e outra à direita. Leia mais para ver por que essa pode ser uma ótima maneira de visualizar os dados da pesquisa.



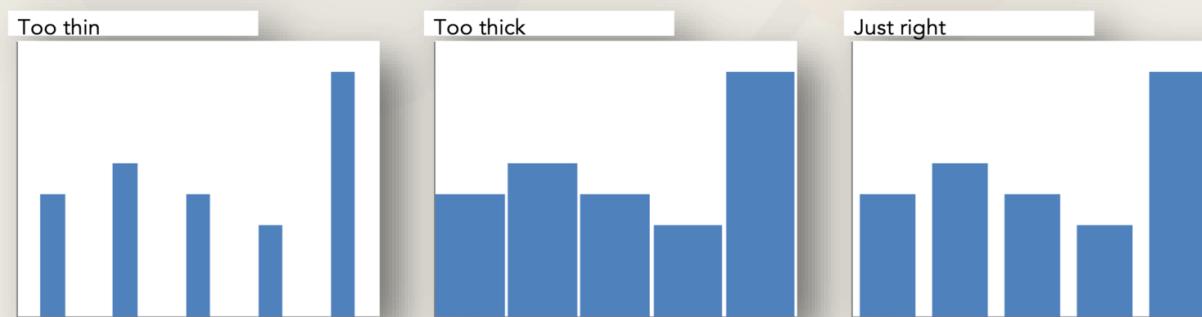
### GRÁFICO DE CACHOEIRA

O gráfico de cascata é uma variação mais avançada de um gráfico de barras que não é tão amplamente utilizado. Este gráfico às vezes pode ser útil para visualizar a mudança, onde a primeira barra representa um valor inicial e a última barra mostra a quantidade final. As barras menores ilustram a mudança incremental do início ao fim. Aprenda a criar um gráfico em cascata no Excel.

Quais são as dicas para criar **gráficos de barras**?

Aqui estão algumas etapas de formatação a serem consideradas ao criar gráficos de barras.

- Ajuste o espaço entre as barras.** Não existe uma regra específica aqui, mas de um modo geral, o gap deve ser mais fino que a largura das barras. No Excel, normalmente escolho uma porcentagem de largura de lacuna entre 30% e 40%. O objetivo é encontrar um equilíbrio saudável: você não deseja ter muito espaço em branco para que os dados não fiquem proeminentes, mas também não deseja que as barras fiquem tão próximas umas das outras que apareçam como uma única forma, em vez de do que elementos distintos. Use o bom senso aqui! (Uma exceção a isso seria se você estivesse criando um histograma, que é um tipo de gráfico semelhante, mas diferente para mostrar a distribuição subjacente de um conjunto de dados. Saiba mais sobre as diferenças entre histogramas e gráficos de barras.)

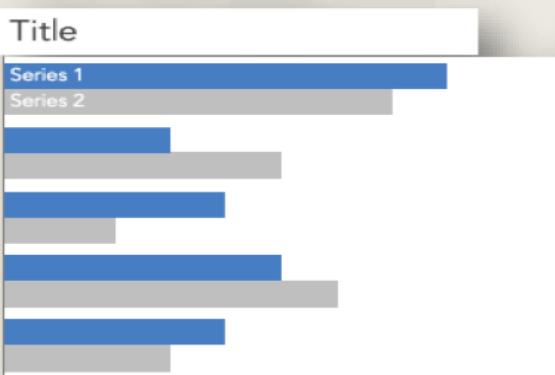


- Peça barras intencionalmente.** Seja cuidadoso com a ordem dos compassos e como o leitor irá interpretá-los. Se houver uma ordem inerente em seus dados, por exemplo, categorias ordinais, como faixas etárias ou uma escala Likert, considere fazer uso dessa ordem natural. Se não for esse o caso, considere organizar as barras com base nos dados, por exemplo, do maior para o menor ou do menor para o maior, dependendo do que você está tentando transmitir.

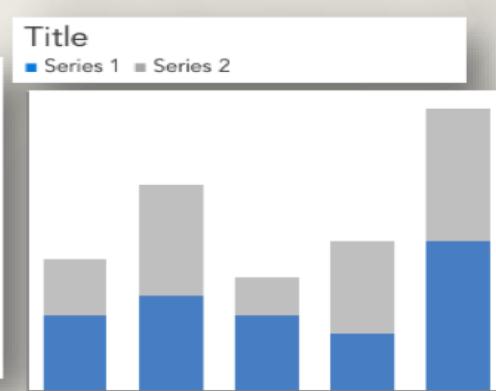


- Considere uma lenda.** Muitas vezes me pego brincando com rótulos de gráfico de barras mais do que com outros tipos de gráfico, especialmente quando não estou usando um eixo para rotular categorias. Os gráficos de barras normalmente não têm muitos espaços em branco, portanto, para evitar confusão, prefiro rotular o nome da categoria dentro de uma barra. Isso às vezes é difícil de fazer com barras verticais ou empilhadas, onde o espaço horizontal é limitado. Por esse motivo, não é incomum adicionar uma legenda a um gráfico de barras se não conseguir ajustar o rótulo dentro da barra ou diretamente ao lado de uma forma visualmente atraente. Quando uso uma legenda, gosto de posicioná-la abaixo do título, para que o leitor saiba como ler o gráfico antes de se deparar com os dados. No visual abaixo, você pode ver como eu normalmente rotularia as séries em um gráfico de barras horizontal versus vertical.

Easy to label directly

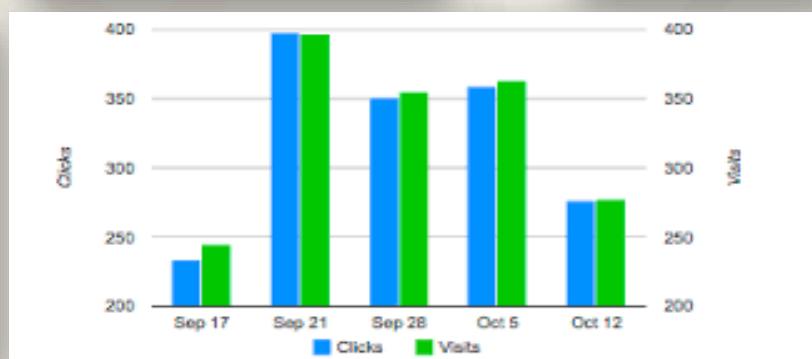
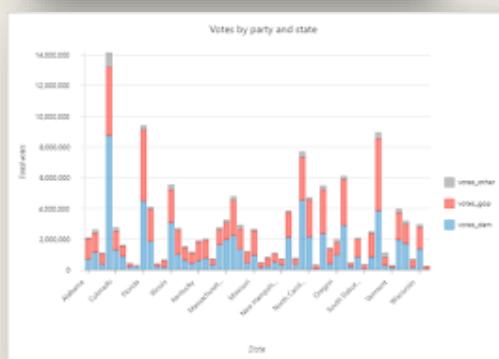
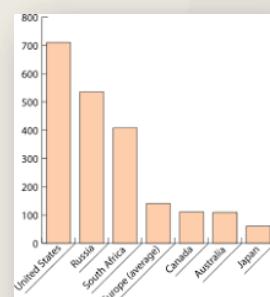
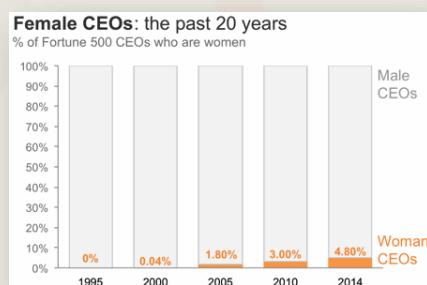
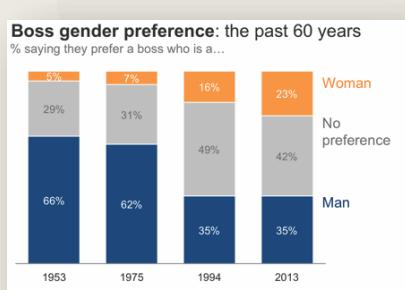


Hard to label directly (use a legend)



- Mova os rótulos de dados diretos dentro das barras. As barras podem parecer muito pesadas e consumir muita tinta. Se rotular os valores de cada barra diretamente, considere mover os rótulos de dados para dentro das extremidades das barras para evitar desordem. Essa pequena mudança de posição pode levar nossos olhos a pensar que há menos, preservando a mesma quantidade de detalhes.
- Use uma linha de base zero. Esta é uma regra importante a seguir. Como os dados são codificados por comprimento, todo o intervalo de barras precisa estar presente; caso contrário, nossa medição relativa será imprecisa. Se você precisar alterar a linha de base para comunicar diferenças menores, é provável que haja um tipo de gráfico alternativo mais adequado às suas necessidades. Ainda não está convencido? Leia mais para ver um exemplo e um pouco de matemática sobre porque isso é fundamental.

## Outras formas de visualização (Gráfico de Barras)





# O que é um gráfico de área?

Um gráfico de área é uma forma especializada do gráfico de linha , onde, em vez de simplesmente conectar nossos pontos de dados com uma linha contínua, também preenchemos a região abaixo dessa linha com uma cor sólida. Isso pode parecer uma pequena mudança cosmética, mas tem um efeito significativo em como percebemos os dados no gráfico.

Os gráficos de área podem ser eficazes para:

- Mostrando a ascensão e queda de várias séries de dados ao longo do tempo
- Transmitindo valores totais ao longo do tempo, bem como algumas divisões subcategorias (mas apenas até certo ponto)
- Enfatizando uma relação parte-todo ao longo do tempo quando uma parte é muito grande, ou muda de muito grande para muito pequena
- Mostrando a mudança ao longo do tempo em painéis individuais de um pequeno gráfico múltiplo

## CENÁRIOS NÃO IDEAL PARA UTILIZAÇÃO DO GRAFICO DE ÁREA

- Conjuntos de dados em escalas que não têm uma relação significativa com zero
- Mostrando vários conjuntos de dados voláteis ao longo do tempo
- Mostrando pequenas diferenças nos valores

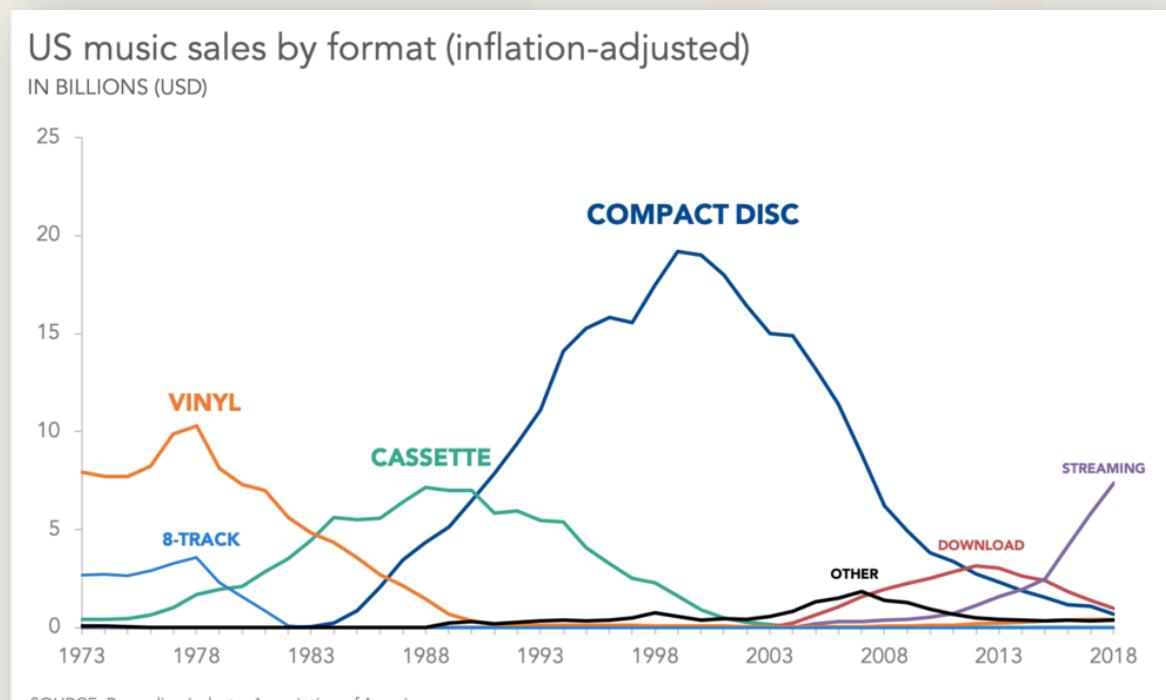
# Como funciona um gráfico de área?

Os gráficos de área são iguais a um gráfico de linha regular, com duas exceções:

O espaço entre o eixo x e cada linha no gráfico é preenchido com alguma cor (possivelmente matiz variável ou opacidade, dependendo da preferência do designer).

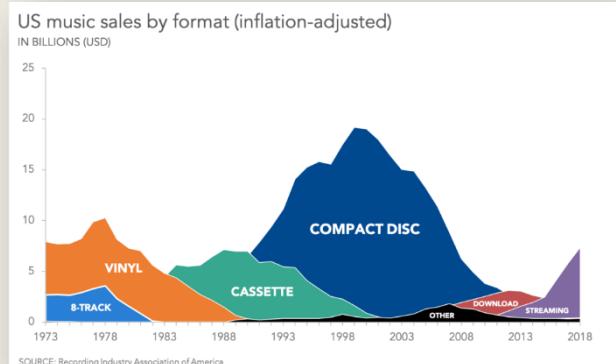
O eixo x DEVE SER ZERO.

Aqui está um gráfico de linha das vendas de música ao longo do tempo nos EUA, divididas por formato:





Lembre-se de que quando olhamos para gráficos de linha, nos concentramos em como os valores no eixo vertical mudam, para cima ou para baixo, à medida que avançamos da esquerda para a direita ao longo de nosso eixo horizontal, e comparamos as linhas entre si mais do que sua altura a partir do eixo x. Pela sua natureza, os gráficos de linha são melhores para mostrar a mudança de valores ao longo do tempo do que uma quantidade exata em cada ponto medido.



Nossa percepção muda, porém, quando um criador de gráfico adiciona um preenchimento de cor entre a linha e a linha de base, transformando nosso gráfico de linha em um gráfico de área.

Em vez de ver linhas, vemos cada série de dados como um polígono, como uma forma; e julgamos cada série de dados menos pela forma como ela muda ao longo do tempo e mais pelo que percebemos como sendo a área geral da região colorida.

## Quando os gráficos de área são úteis?

Gráficos de área são tentadores de usar no lugar de gráficos de linha. Eles são coloridos, atraentes e são tão incomuns que o fator novidade tende a chamar a atenção. Como discutimos anteriormente, no entanto, gráficos de linha e gráficos de área não são intercambiáveis, devido à forma como os visualizadores percebem as diferentes apresentações de dados. Raramente (mas não nunca!) encontramos casos de uso ideais para o gráfico de área. Ainda assim, aqui estão alguns casos para os quais o gráfico de área pode ser uma opção razoável.

Mostrando várias séries, em vários períodos diferentes, se houver variação significativa entre as séries e mudanças ao longo do tempo

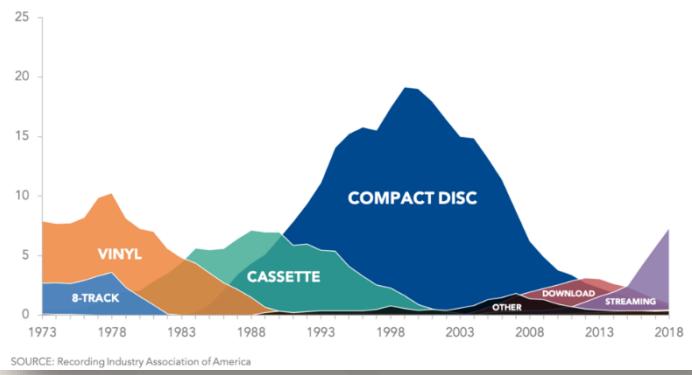
Às vezes, é importante mostrar como as quantidades mudaram ao longo do tempo em várias séries de dados relacionadas. Os gráficos de barras são bons para mostrar uma quantidade em um ponto no tempo, mas várias séries em vários pontos no tempo são difíceis de mostrar como barras. O gráfico fica lotado e desordenado muito rapidamente.

O gráfico de área tenta dar uma noção de quantidade em todos os pontos do tempo, criando polígonos irregulares de cores preenchidas. No entanto, onde as séries de dados se sobreponem, pode ser um desafio para o público ler que dois (ou mais) preenchimentos estão ocupando o mesmo espaço.

Continuando com o mesmo exemplo que estamos usando, que mostra as vendas de música ajustadas pela inflação por formato ao longo do tempo, tentamos mitigar esse desafio com uma pequena alteração na opacidade de nossas cores de “preenchimento”, conforme mostrado abaixo.

Com preenchimentos puramente opacos, não saberíamos, por exemplo, quantos cassetes foram vendidos antes de 1983, ou quantos CDs foram vendidos antes de 1990. (Ainda é difícil ver quantos cassetes foram vendidos antes de 1979, porque três séries se sobrepõem, e é a série de dados "Cassete" que fica mais para trás. Esse tipo de gráfico também exige que um designer faça escolhas ponderadas sobre a ordem da série de dados.)

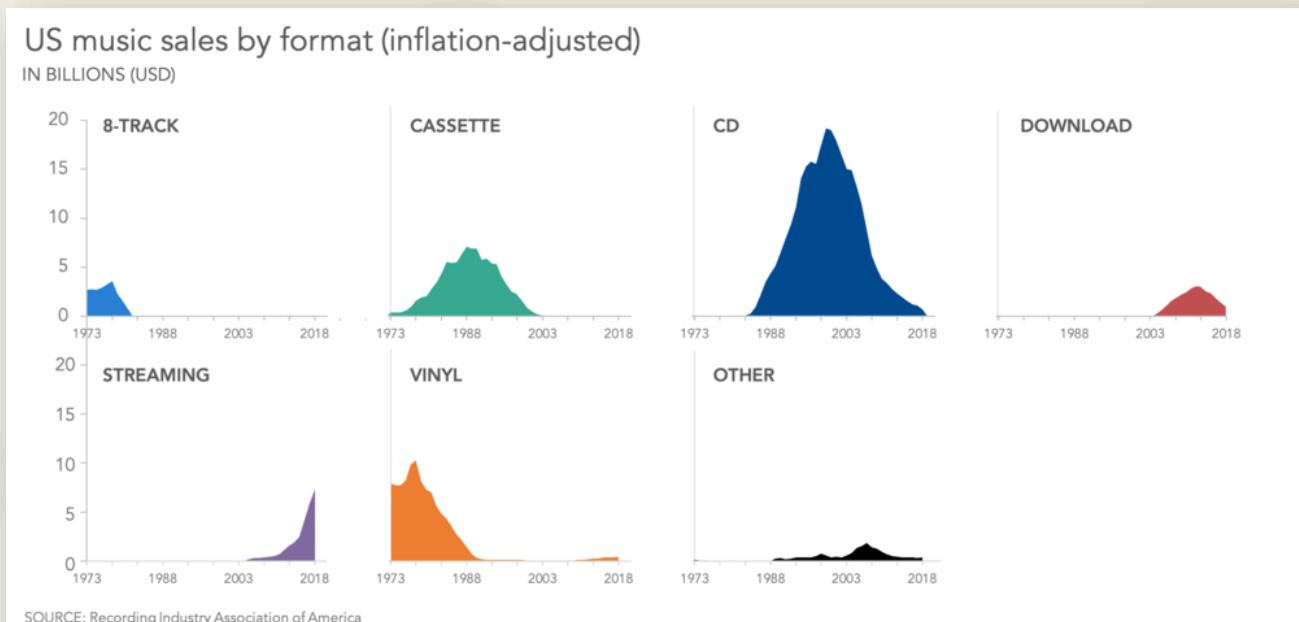
US music sales by format (inflation-adjusted)  
IN BILLIONS (USD)



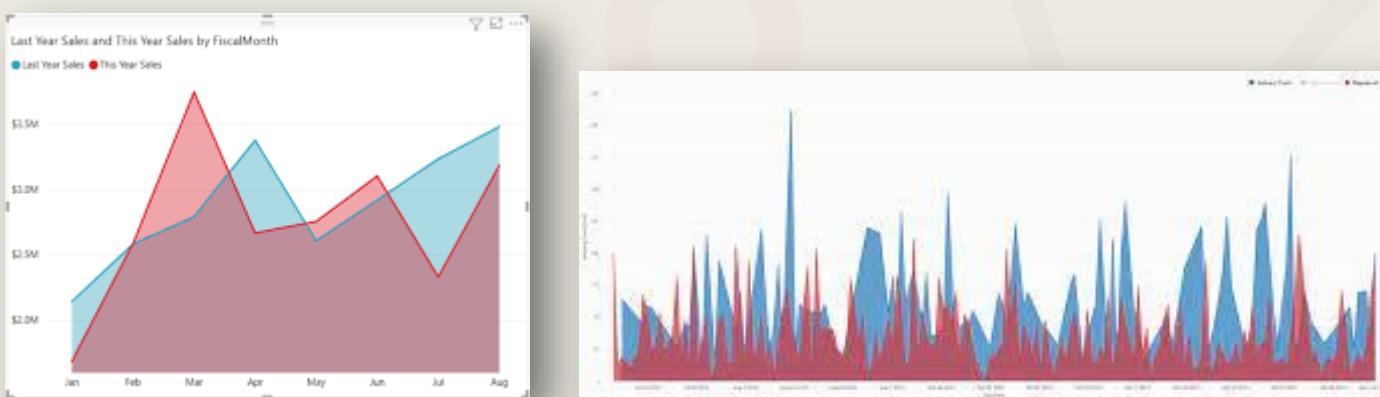
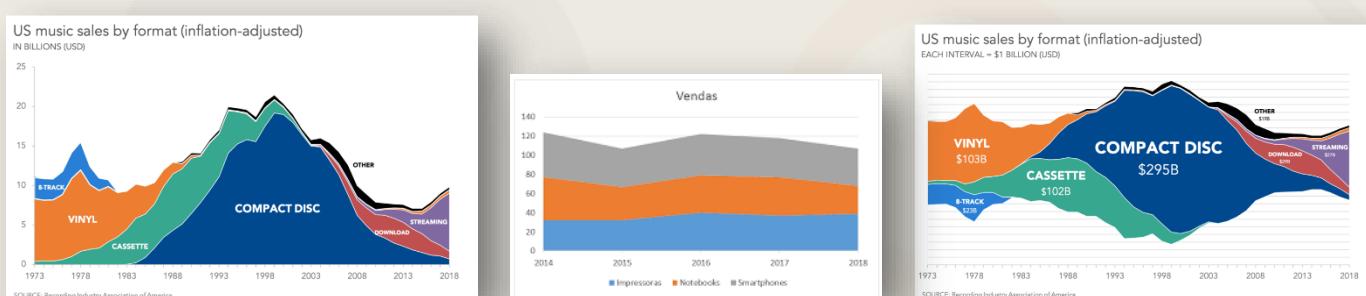


Mostrando várias (ou dezenas) séries que mudam ao longo do tempo em um pequeno gráfico múltiplo

Os gráficos de área podem ser úteis em um pequeno gráfico múltiplo (também chamado de gráfico de painel ou gráfico de treliça). Quando você tem várias séries de dados, em vez de sobrepor-las todas em um gráfico, você pode exibir cada uma delas como um pequeno gráfico único em uma série regularmente ordenada de gráficos semelhantes. Um gráfico de área pode ajudar os visualizadores a comparar a quantidade e a mudança ao longo do tempo nesses painéis.



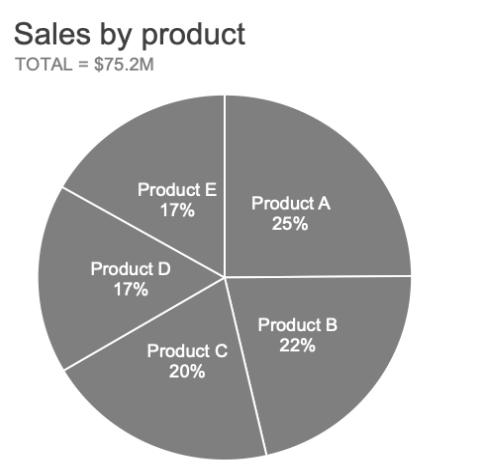
## Outras formas de visualização (Gráfico de Área)





# O que é um gráfico de pizza?

Um gráfico de pizza expressa uma relação parte-a-todo em seus dados. Imagine uma torta de verdade (vou deixar você escolher sua variedade favorita!). Cada fatia representa um componente e todas as fatias somadas são iguais ao todo. As tortas são talvez tão comumente usadas porque, como Manuel Lima escreve em *O Livro dos Círculos*, os humanos gravitam em círculos como o símbolo universal de unidade, perfeição e infinito. Gostamos que as coisas sejam inteiras - e entendemos como as respectivas peças se relacionam com o todo. Um gráfico de pizza nos dá o conforto de um relacionamento parte-a-todo.



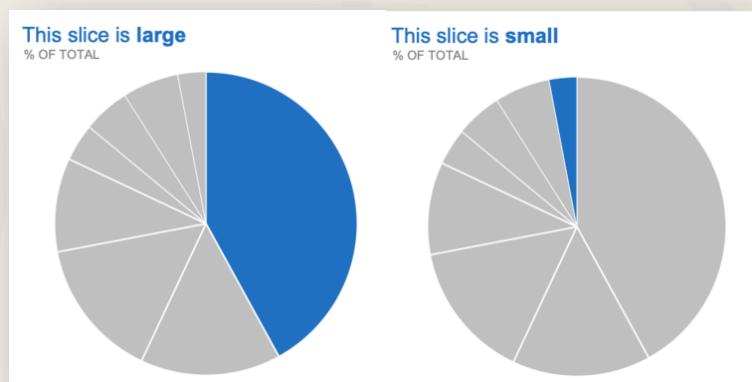
Aqui está uma breve visão geral de como consumimos uma torta. Cada fatia de pizza tem três componentes visuais: seu ângulo central, área e comprimento do arco. Uma pesquisa recente de Robert Kosara e Drew Skau demonstrou que lemos gráficos de pizza por área - ou possivelmente em combinação com o comprimento do arco - o que desmentiu algumas crenças anteriores de que as pizzas são lidas por ângulo. Você pode ler mais sobre a pesquisa deles em um tour ilustrado pelos resultados do estudo do gráfico de pizza.

## Existem dois casos de uso principais para um gráfico de pizza:

Se você deseja que seu público tenha uma noção geral do relacionamento parte-a-todo em seus dados e comparar os tamanhos precisos das fatias é menos importante.

Para transmitir que um segmento do total é relativamente pequeno ou grande.

Nosso exemplo de vendas suporta o primeiro caso de uso. Suponha que queremos nos concentrar em quanta receita um determinado produto – ou conjunto de produtos – contribui para o total. Dada a nossa construção natural de como lemos um relógio, é bastante fácil dizer que o Produto A representa aproximadamente um quarto da receita total e que os Produtos A, B e C combinados representam aproximadamente dois terços da receita total. Para essas relações gerais de parte para todo, um gráfico de pizza pode transmitir essas informações de maneira eficaz.



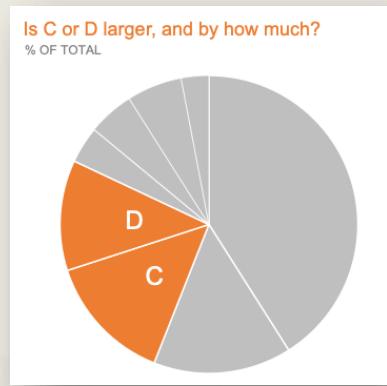
Os dois gráficos de pizza abaixo são exemplos do segundo caso de uso. Com cada um deles, posso facilmente chegar a uma conclusão ampla - uma fatia da torta é relativamente maior ou menor que as outras.



## CENÁRIOS NÃO IDEAL PARA UTILIZAÇÃO DO GRAFICO DE PIZZA

não devem ser usadas para avaliar os tamanhos relativos das categorias, comparar dados entre as pizzas e visualizar porcentagens que não somam 100%. Vejamos um exemplo de cada.

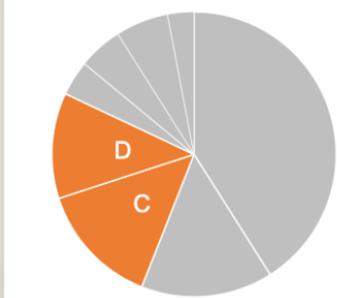
Comparando o tamanho das categorias: os gráficos de pizza se decompongão rapidamente se precisarmos que nosso público tenha uma compreensão mais sutil de nossos dados do que simplesmente grande, pequeno ou quase o mesmo. Para ilustrar, considere o seguinte exemplo. Se eu pedisse para você comparar o tamanho das fatias C e D, qual você concluiria ser maior? Indo um passo adiante, com que facilidade você pode estimar quanto maior?



Este exemplo demonstra uma limitação dos dados codificados como fatias de pizza: os olhos humanos não estão bem equipados para comparar áreas. O desafio torna-se ainda mais pronunciado sem uma linha de base consistente – como um eixo y em um gráfico de barras horizontal – contra a qual fazer a comparação visual. Conforme ilustrado a seguir, comparar o tamanho de C x D é muito mais fácil no gráfico à direita. Também estamos comparando uma dimensão (comprimento) em duas categorias, em vez de áreas de formato irregular.

This is hard!

Is C or D larger, and by how much?  
% OF TOTAL



This is easier!

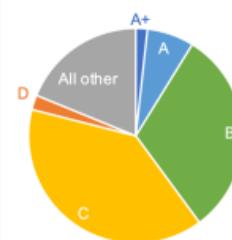
Is C or D larger, and by how much?  
% OF TOTAL



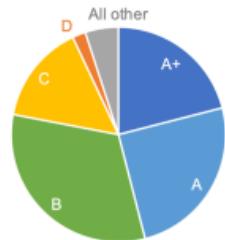
Comparar dados entre pizzas: esta é uma tarefa difícil, porque as coisas que você deseja comparar são separadas visualmente e estão localizadas em lugares diferentes em cada pizza. Para ilustrar, confira o par de tortas abaixo, extraído do Exercício 2.1 do Storytelling with Data: Let's Practice!

New client tier share

% of Total Accounts

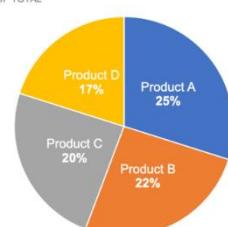


% of Total Revenue



This does not make sense!

Sales by product  
% OF TOTAL



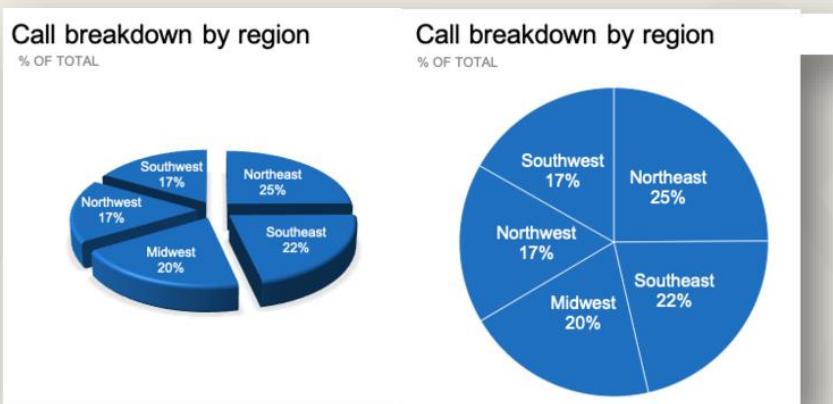
Mostrando dados que não somam 100%: se as fatias somam algo diferente de 100%, está errado. Por definição, a torta deve retratar um todo significativo.



## Quero usar uma torta: como faço uma boa?

Ao usar gráficos de pizza, aqui estão algumas dicas rápidas de design para torná-los mais eficazes.

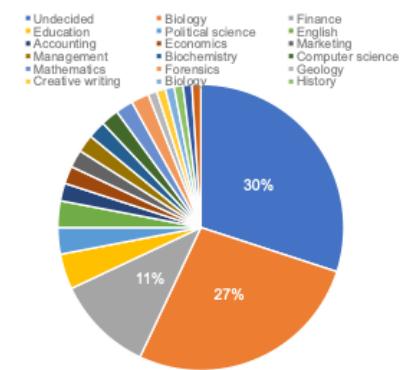
Evite efeitos 3D e explosivos. Esses elementos introduzem desordem desnecessária e podem distorcer os dados. Essa incapacidade de fazer uma comparação visual precisa foi validada no estudo de Kosara e Skau mencionado acima, em que os participantes do estudo julgaram erroneamente os valores em gráficos de pizza explosivas em um grau mais alto do que na pizza plana.



Não tenha uma tonelada de fatias. Não há uma regra exata aqui, portanto, considere cuidadosamente seu conjunto de dados e as necessidades do público. Você pode melhorar a legibilidade desenfatizando pequenas categorias com semelhança de cor (como mostrado na seguinte torta do meio com cinza) ou agregando em uma categoria “Todas as outras”, conforme ilustrado à direita.

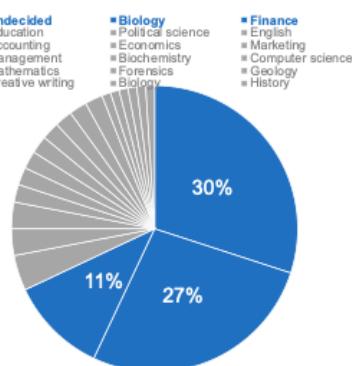
### Too many slices

**Declared majors**  
% OF ANDERSON HALL RESIDENTS



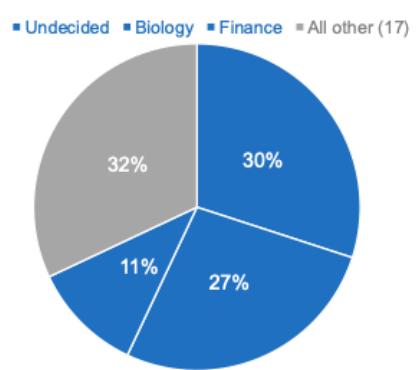
### De-emphasize

**Declared majors**  
% OF ANDERSON HALL RESIDENTS



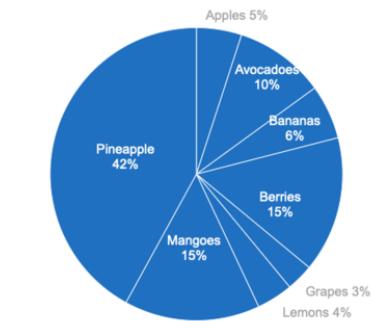
### Aggregate

**Declared majors**  
% OF ANDERSON HALL RESIDENTS

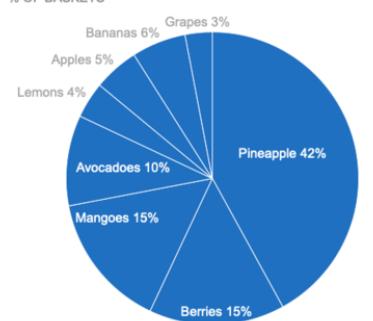


**Classifique seus dados de forma significativa.** Confira as duas tortas abaixo. À esquerda, os dados são classificados em ordem alfabética (começando com maçãs em zero grau). A torta certa é classificada em ordem decrescente começando com a maior categoria (abacaxi). Isso leva em consideração nossa construção natural de leitura em torno de um círculo.

**Fruit purchases**  
% OF BASKETS

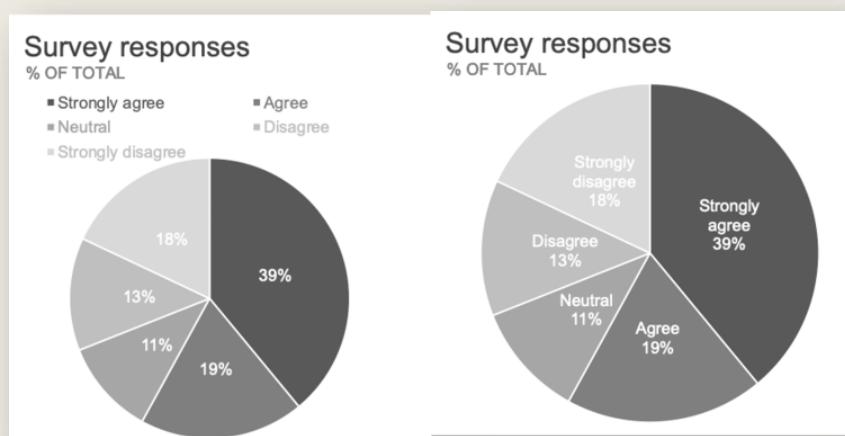


**Fruit purchases**  
% OF BASKETS

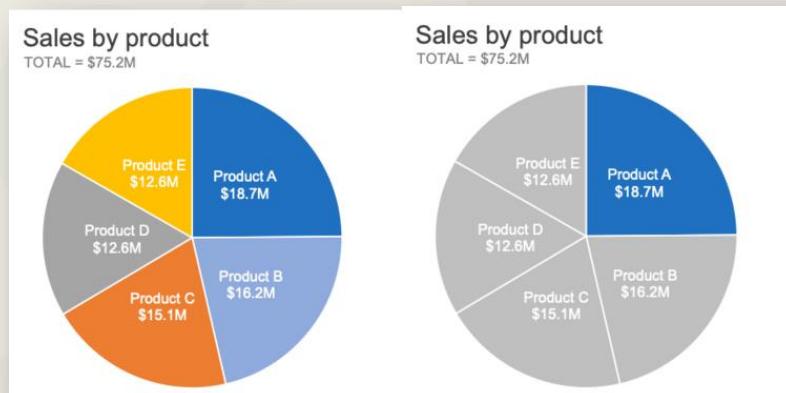




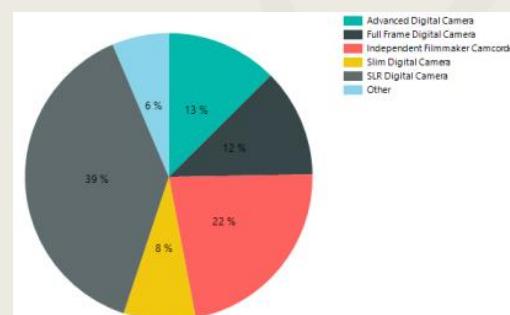
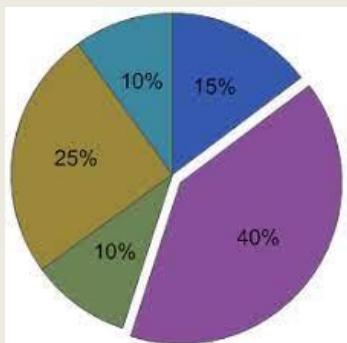
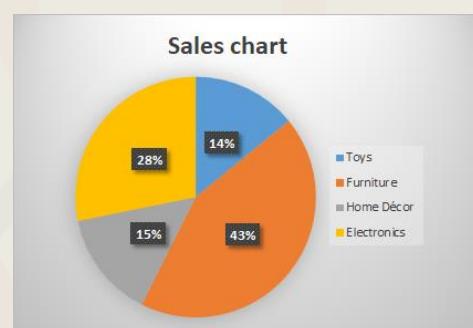
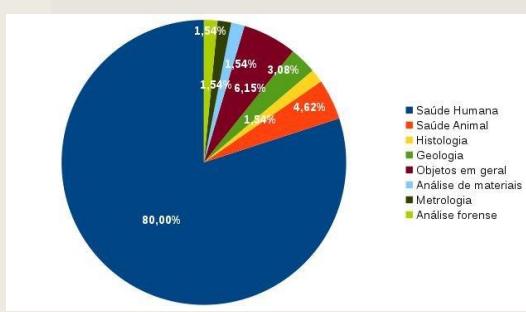
Elimine a legenda e rotule os dados diretamente. Confira a diferença nos gráficos a seguir. Rotular diretamente, conforme mostrado à direita, reduz a tarefa de ir e vir entre a legenda e os dados. Sempre que possível, tente rotular suas fatias diretamente com o valor, desde que haja espaço suficiente para tornar os rótulos legíveis.



**Use cores estratégicamente.** Quando usada com moderação, a cor é indiscutivelmente a decisão mais importante que tomamos para focar a atenção do público. Vamos supor que gostaríamos de destacar a contribuição do Produto A para nosso portfólio de vendas. Qual das tortas a seguir chama mais efetivamente sua atenção para o produto A?



## Outras formas de visualização (Gráfico de pizza)





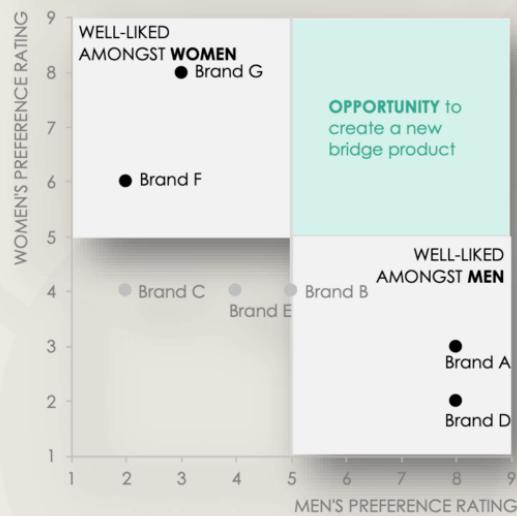
# O que é um gráfico de Dispersão?

Um gráfico de dispersão mostra a relação entre duas variáveis numéricas plotadas simultaneamente ao longo dos eixos horizontal e vertical. Eles são comuns em campos científicos e frequentemente usados para entender os dados, em vez de se comunicar com eles. Isso não quer dizer que você nunca deva se comunicar com um, mas você deve tomar medidas explícitas para garantir que seu gráfico seja claro para um público desconhecido (algo que você deve fazer com todos os gráficos!). Vejamos um cenário em que um gráfico de dispersão funciona bem para comunicar uma descoberta.

Imagine que você é um analista do setor de beleza e sua empresa deseja formular um novo produto para os lábios. Antes de embarcar nessa empreitada, você é solicitado a fazer uma pesquisa para ver se há um buraco no mercado. Você descobre que os produtos para os lábios estão polarizando entre compradores masculinos e femininos, então há uma oportunidade de criar um novo produto que preencha a lacuna.

Você pode considerar mostrar a relação entre as pontuações de classificação masculina e feminina usando um gráfico de dispersão, como o abaixo.

Lip care is polarizing between men and women



É de natureza exploratória. Um gráfico de dispersão provavelmente foi usado para descobrir essa descoberta. Muitos pacotes de software estatístico geram gráficos de dispersão para testar a correlação entre duas variáveis. Embora o gráfico usado na fase de descoberta nem sempre seja ideal para comunicar os insights finais, ele funciona nesse caso. Posso ver os dois conjuntos de classificações simultaneamente e também enfatizar o buraco no mercado. Eu categorizei e rotulei os pontos para tornar este gráfico legível, mas um gráfico alternativo - um que seja mais familiar - também pode funcionar (confira esta postagem para ver três alternativas para comparar métricas). Tudo depende do seu público e cenário específico.

Esteja atento ao posicionamento variável nos eixos. Neste exemplo, não importa qual variável está ao longo do eixo horizontal ou vertical, mas nem sempre será o caso. Às vezes, você terá uma variável independente e uma dependente. Uma variável independente é exatamente o que seu nome indica: ela não é afetada pela outra variável. Uma variável dependente é provavelmente o que você está tentando medir, o que significa que ela é afetada por sua variável independente. É prática comum colocar a métrica

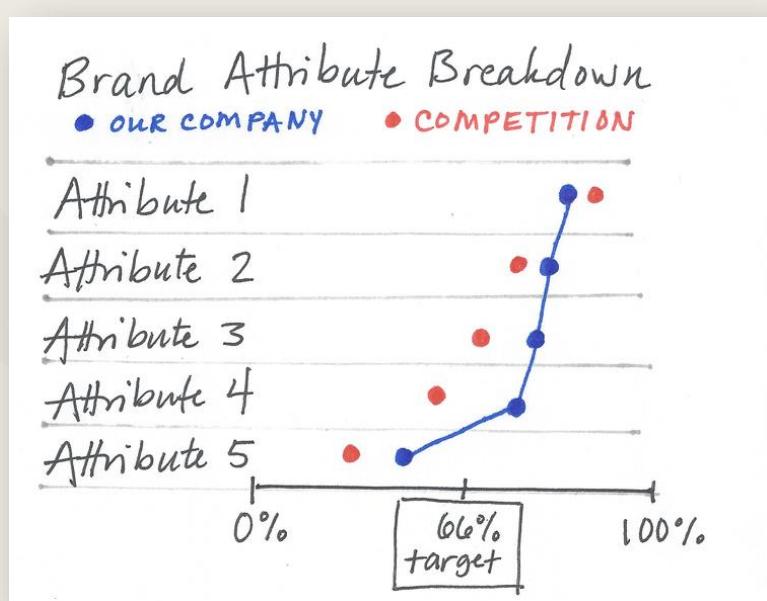
independente ao longo do eixo horizontal ou x e a variável dependente ao longo do eixo y. Não é errado invertê-los, mas pode ser inesperado, causando um pouco de confusão inicial.

Os gráficos de dispersão mostram pontos, não linhas. Os gráficos de dispersão são muito semelhantes aos gráficos de linhas, pois ambos exibem dois valores numéricos; no entanto, os gráficos de dispersão tendem a se concentrar em pontos de dados individuais (representados com um ponto) em vez de agregar vários pontos em uma linha distinta.



## O exemplo: dados de pesquisa de mercado

A equipe estava analisando alguns dados de pesquisa de mercado, querendo comparar sua empresa com seu principal concorrente em algumas dimensões. Eles originalmente visualizaram isso com um gráfico de dispersão conectado. Não funcionou bem. As pessoas descobriram que estavam conversando sobre como ler o gráfico corretamente por muito tempo antes de chegarem ao ponto em que pudessem realmente olhar para os dados e ver o que poderiam dizer com eles. Não vou me esforçar para recriá-lo completamente aqui, mas para dar uma noção, farei um esboço rápido (todos os dados foram alterados e o cenário generalizado para preservar a confidencialidade):

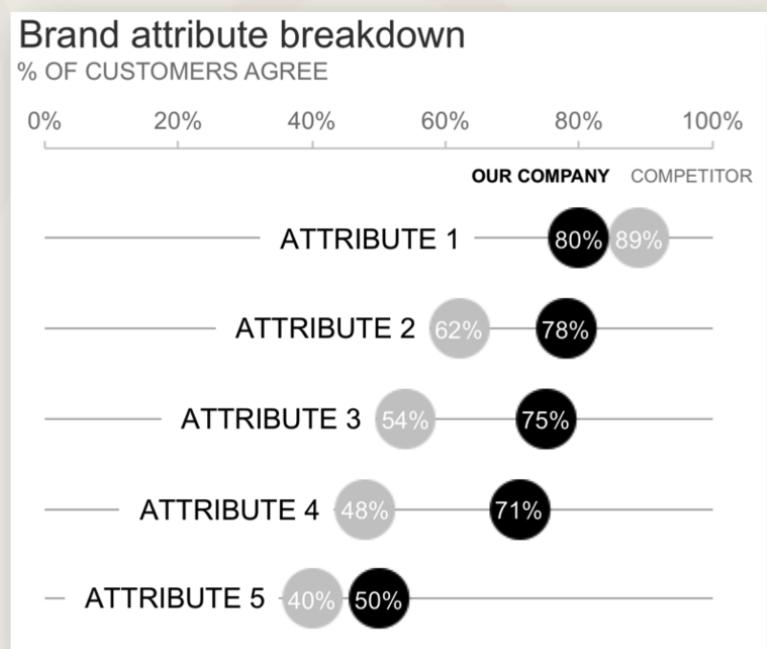


Havia dois pontos principais a fazer com esses dados:

A empresa estava se saindo melhor do que o concorrente em todas as áreas, exceto ATRIBUTO 2. A empresa estava atingindo sua meta em todas as áreas, exceto no ATRIBUTO 5. Isso parece bem direto, certo? Você pode obter essas informações por meio do visual acima, mas há melhorias que podemos fazer nele e em outras exibições em potencial que também podem funcionar. Originalmente pensei que deveríamos olhar para duas alternativas: (1) um gráfico de pontos e (2) um gráfico de inclinação. Na verdade, foi Elizabeth da minha equipe que adicionou uma terceira opção — o “velho gráfico de barras chato” — à mistura (que bom que ela o fez!). Vamos dar uma olhada em cada um deles.

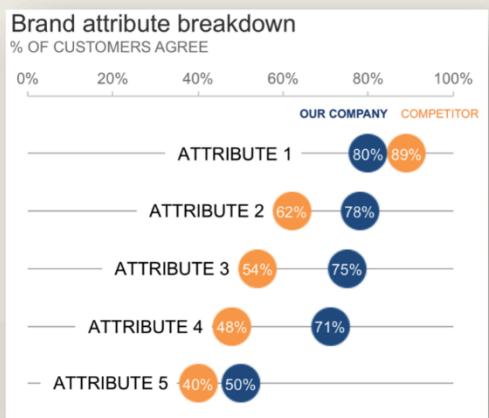
### Opção 1: o gráfico de pontos

Primeiro, o gráfico de pontos. Esta foi principalmente uma tentativa de melhorar seu visual original com uma visão semelhante e algumas pequenas modificações. Não utilizo com muita frequência, mas descobri que existem alguns bons casos de uso. Eu pensei que este seria um cenário apropriado para isso. O visual base ficou assim:

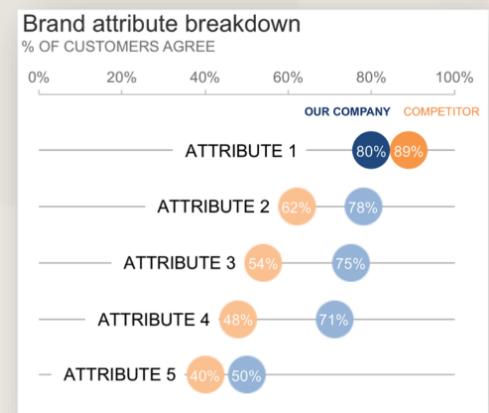




Em vez de conectar os pontos para baixo, como no gráfico de dispersão conectado original, que faz a comparação primária de como NOSSA EMPRESA está se saindo nos vários atributos, as linhas horizontais aqui desenham os olhos da esquerda para a direita. Isso torna a comparação primária NOSSA EMPRESA x CONCORRENTE, que parecia ser o ponto principal aqui.

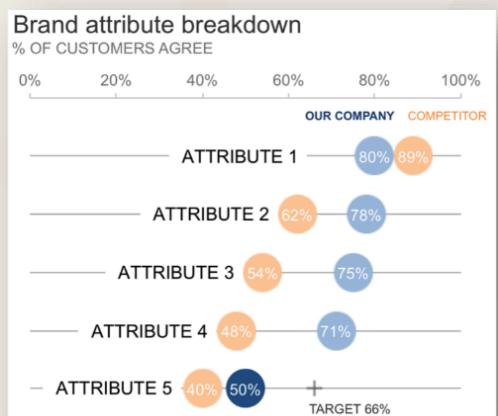


Agora que incorporei a cor, posso variar a intensidade para enfatizar certos pontos. Por exemplo, poderíamos primeiro chamar a atenção para o ATRIBUTO 1, onde NOSSA EMPRESA pontua abaixo do CONCORRENTE:



Como você lê um gráfico de dispersão?

A maior parte deste post é dedicada a como você pode usar gráficos de dispersão para fins explicativos, apesar de serem um tipo de gráfico mais técnico. Conforme mencionado anteriormente, se você usar um gráfico de dispersão, esteja ciente de que pode ser necessário dividi-lo para o seu público e, se for a primeira vez que eles olham para um desses gráficos, talvez seja necessário explicar como lê-lo. Vou orientá-lo no processo que utilizei ao examinar gráficos de dispersão, bem como links para recursos adicionais. Verifique os eixos. Uma das primeiras coisas que faço ao ler qualquer gráfico é escanear cada eixo. É especialmente importante ao ler gráficos de dispersão, pois eles contêm várias variáveis, portanto, é necessário determinar qual variável está ao longo de qual eixo. Também devo mencionar que nem todos os gráficos de dispersão exibirão variáveis semelhantes; alguns podem comparar métricas diferentes com unidades de medida e escala variáveis. Entender isso desde o início facilitará a interpretação da relação assim que você começar a examinar os dados.



Visualizar seções. No exemplo acima, criei seções agrupando os pontos em quadrantes. Isso é útil não apenas para comunicar a descoberta no cenário de cuidados com os lábios, mas também para revelá-la. Na próxima vez que você olhar para um gráfico de dispersão, pergunte-se quais linhas você pode desenhar ou quais quebras e agrupamentos naturais existem que o ajudarão a entender a comparação.



Identifique a forma. Costumo dar um passo para trás no gráfico, para tentar resumir os pontos individuais em uma forma unificada. Eles estão todos se movendo em uma única direção linear? Ou talvez seja mais uma curva exponencial? Os pontos aumentam conforme meus olhos se movem ao longo do eixo? Ter uma descrição em mente me ajuda a descobrir e explicar o relacionamento. Por exemplo, neste exemplo de custo por milha , podemos ver que os pontos de dados formam uma forma de u. Isso significa que o custo é relativamente alto para usos mais curtos e mais longos, mas como dirigimos uma quantidade média, o custo é mais gerenciável. Tenha em mente que pode não haver uma forma discernível, o que é um achado perfeitamente válido (e sugere uma relação fraca ou inexistente entre as variáveis).

Eu seria negligente em não compartilhar nenhum aviso neste momento. Ainda posso ouvir meus professores de estatística cravando este mantra em minha cabeça: "correlação não implica causalidade". Só porque você descobriu um relacionamento não significa que identificou a causa subjacente. Por exemplo, no exemplo de cuidados com os lábios, descobrimos que as marcas preferidas entre os homens são menos preferidas entre as mulheres. Isso significa que se um homem gosta de uma marca, uma mulher não? Absolutamente não, ou não haveria oportunidade de criar um produto ponte. Podemos usar as descobertas para informar e tomar decisões fundamentadas,

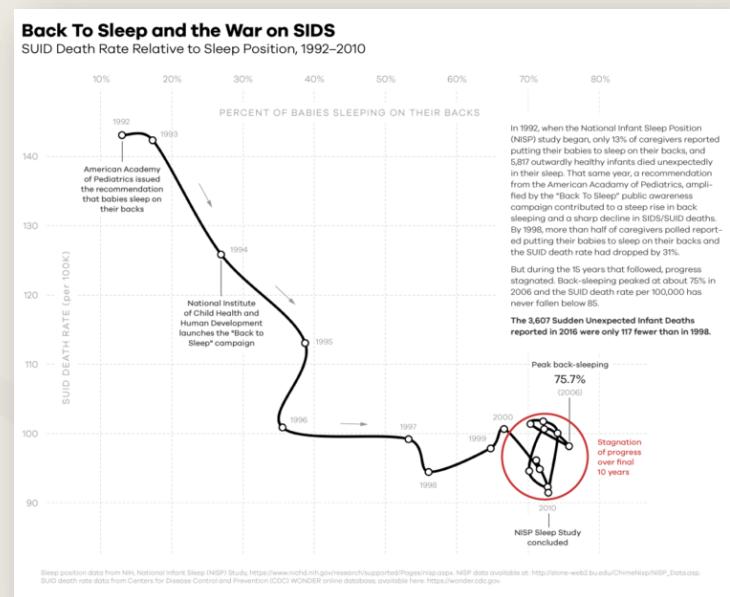
mas generalizar o insight seria um uso indevido dos dados. Isso também se aplica a outros tipos de gráficos, no entanto, acho que os gráficos de dispersão são mais suscetíveis a erros de leitura. Acho que é porque eles focam no relacionamento, tornando mais fácil especular e extrapolar.

### Quais são as variações comuns de gráficos de dispersão?

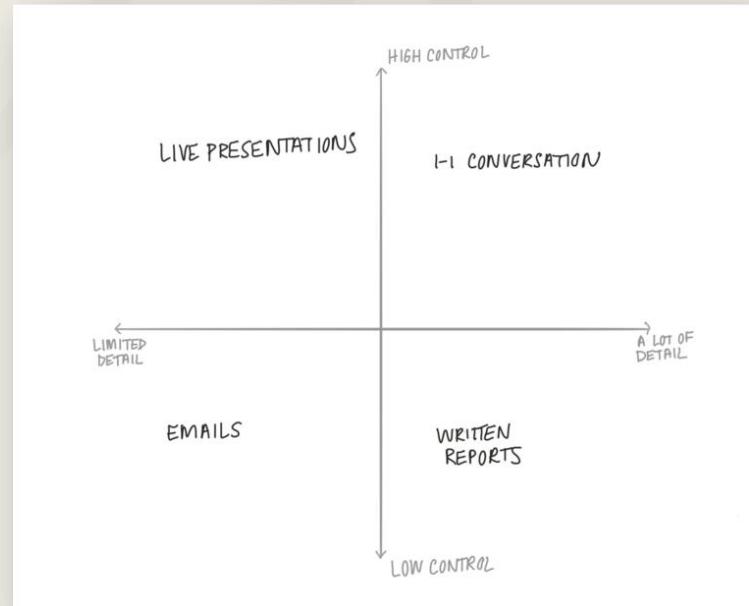
Gráficos de bolhas: Vamos imaginar que o exemplo de cuidado labial inclua outra dimensão (como tipo de embalagem, preço, ingrediente ativo etc.). Podemos codificar essa terceira variável alterando o tamanho, a cor ou até mesmo a forma dos pontos de dados. Este é um gráfico de bolhas. Os gráficos de bolhas são úteis para mostrar relacionamentos multidimensionais, mas isso tem um custo, pois são difíceis de ler. Depois de alterar o tamanho dos círculos, começamos a codificar as informações por área. Nossos olhos não são muito bons em medir áreas, então comparações específicas são mais difíceis de fazer. Também devo mencionar que muitos aplicativos gráficos não oferecem modelos de gráficos de bolhas enlatados. Isso significa que alguém precisa calcular a área correta de cada círculo, o que pode introduzir erro humano. Além disso, enquanto brincamos com cores e formas, pedimos ao público que faça muito trabalho mental para processar as informações. As pessoas normalmente só podem reter cerca de quatro informações visuais em sua memória de curto prazo a qualquer momento, portanto, esteja atento ao esforço cognitivo colocado no leitor! Dito isso, existem exemplos fantásticos de gráficos de bolhas explicativos, como o famoso Vídeo da BBC . O que torna a explicação de Hans tão eficaz é sua disposição e entusiasmo para conduzir seu público através de seu gráfico de bolhas animado. Se você estiver considerando um gráfico de bolhas - ou qualquer gráfico multidimensional - considere seu público e quanto esforço deseja gastar explicando como ler seu gráfico.



Gráficos de dispersão conectados: penso em um gráfico de dispersão conectado como um híbrido entre um gráfico de dispersão tradicional e um gráfico de linhas. Ainda existem duas dimensões ao longo dos eixos x e y, mas uma terceira dimensão, geralmente o tempo, é sobreposta com linhas. O desafio é que estamos acostumados a ler o tempo da esquerda para a direita, então vê-lo se movendo em todas as direções de ponto a ponto pode ser perturbador. Como acontece com todas as variações, cada gráfico tem um caso de uso ideal e, com anotações criteriosas, rotulagem e atenção focada, podem ser esclarecidos para qualquer público. Para saber mais, confira a postagem do convidado de Dan Zvinca sobre as várias funções das linhas e veja o exemplo do envio do #SWDchallenge de Bill Rapp em janeiro de 2018 (abaixo).



Grade 2x2 (gráfico de dispersão conceitual):  
Esta não é uma visão multidimensional per se, mas uma que pode ser prática em um ambiente de negócios. Muitas vezes queremos entender as relações entre duas coisas ou explicar como uma decisão pode impactar outra. Um gráfico de dispersão conceitual, também conhecido como grade 2x2, pode ajudar seu público a entender essas comparações, porque nosso sistema visual é muito mais rápido no processamento de informações do que nosso sistema verbal. Por exemplo, podemos usar um para explicar as diferenças entre apresentar seus dados ao vivo e comunicar por meio de um documento escrito. As duas variáveis são a quantidade de controle que você deseja ter e a quantidade de detalhes necessários para transmitir seu ponto de vista. Observe como é rápido entender quando um meio é ideal sobre o outro.



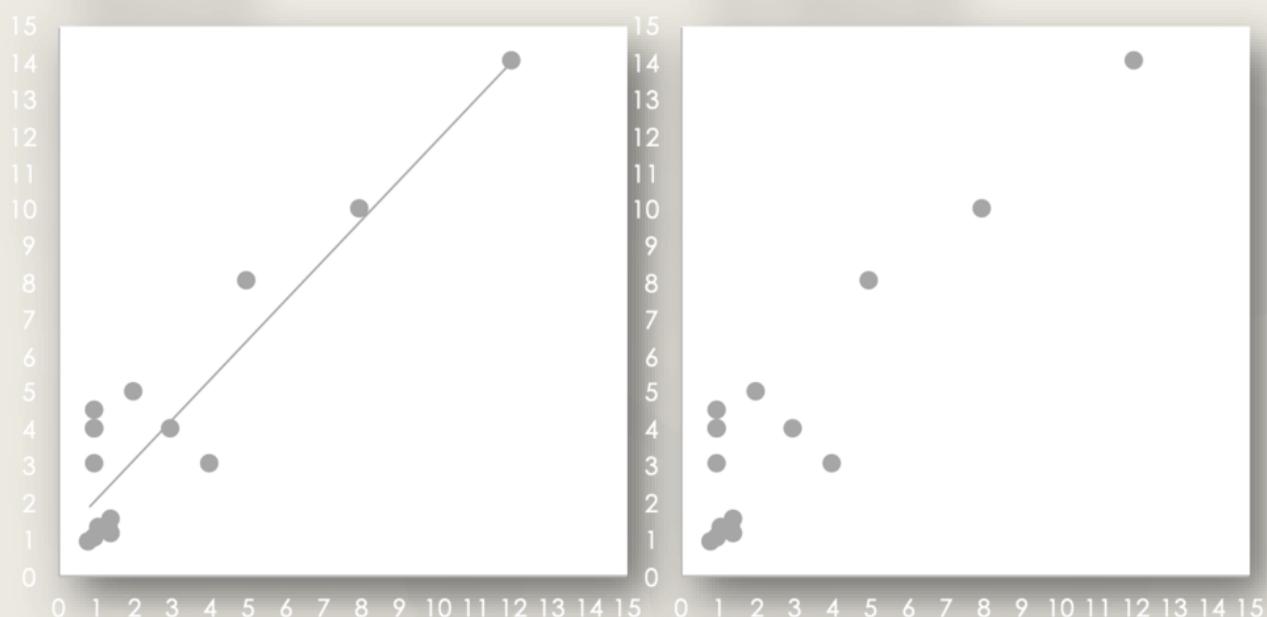
## Quais são as dicas para projetar gráficos de dispersão?

Aqui estão algumas etapas de formatação a serem consideradas ao projetar gráficos de dispersão.

Declutter removendo linhas de tendência. Durante o processo analítico, você pode ajustar um modelo para descrever a relação mostrada em seu gráfico de dispersão e exibir esse modelo como uma linha que passa pelos pontos de dados. Minha preferência - e percebo que isso pode ser controverso - é remover a linha ajustada ao comunicar os dados. Isso adiciona confusão e, se a tendência subjacente não for óbvia, a implantação de uma linha pode levar a contenção ou confusão.



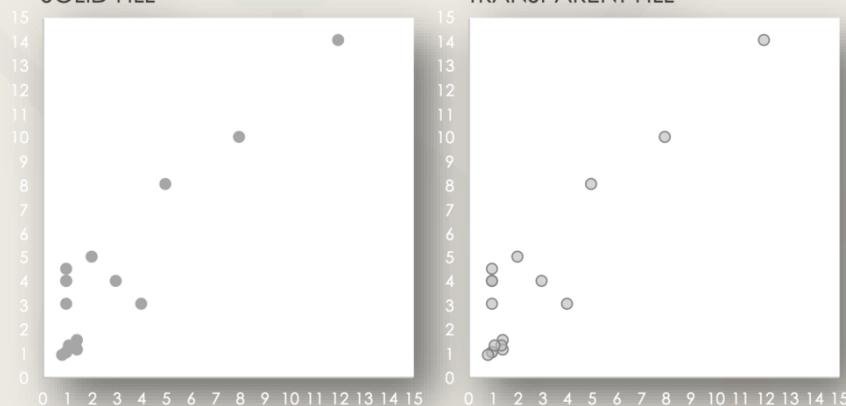
TRENDLINE



NO TRENDLINE

Torne os pontos de dados sobrepostos transparentes. Se muitos pontos de dados se sobrepõem, pode ser difícil ver cada valor ou o volume de pontos em uma seção específica. Um truque pode ser brincar com a opacidade dos marcadores de dados para que os pontos de dados individuais fiquem visíveis.

SOLID FILL

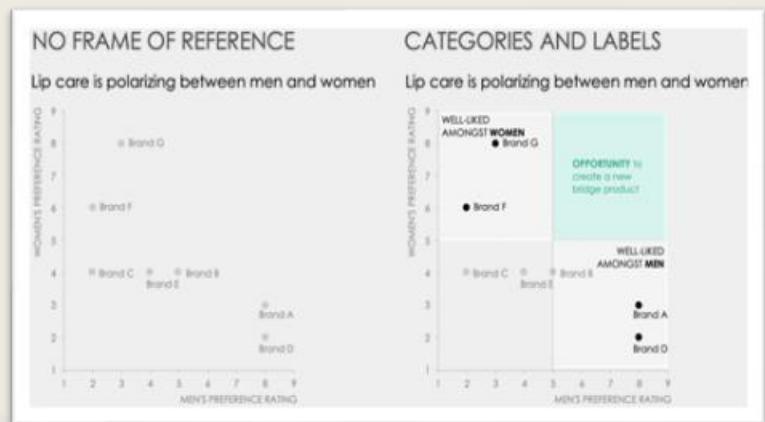


TRANSPARENT FILL

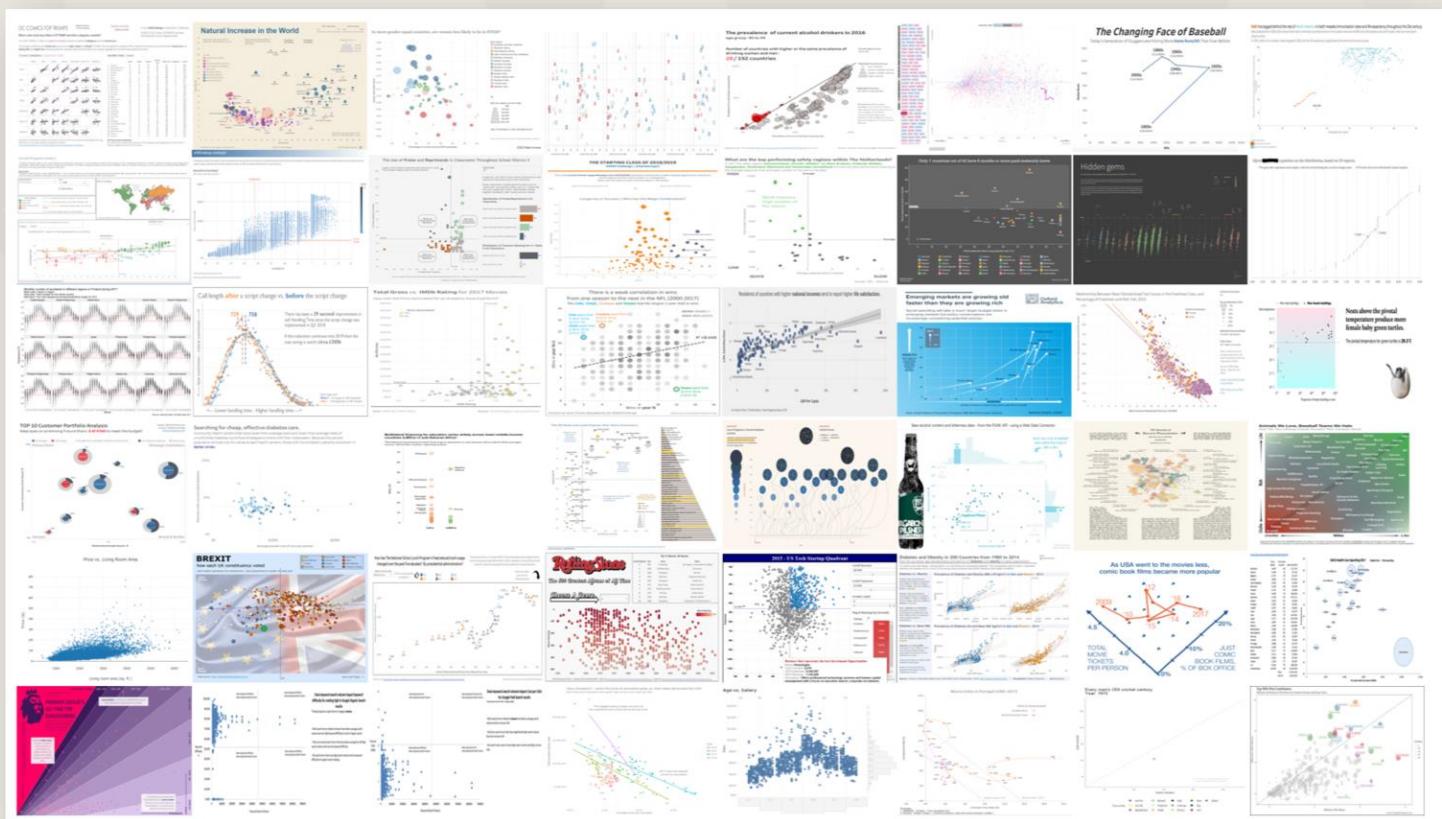
Você não precisa de uma linha de base zero. Semelhante aos gráficos de linha, os gráficos de dispersão codificam dados por posição ao longo do eixo. Isso significa que não é necessário que sua linha de base comece em zero da mesma forma que para as barras . Com isso dito, você ainda deve estar atento sempre que se desviar da linha de base zero, pois isso pode causar confusão.



Crie seções e adicione rótulos para maior clareza. Adicionar anotações e categorizar pontos de dados pode facilitar o consumo de gráficos de dispersão. No exemplo de cuidados com os lábios, observe como as descrições concisas não apenas funcionam bem em espaços restritos, mas também facilitam a discussão desse gráfico. Posso me referir ao quadrante da oportunidade e você provavelmente sabe o que quero dizer! Sem implementar uma construção clara para ler o gráfico de dispersão, o gráfico é altamente exploratório e exige que o leitor repita a análise para descobrir a relação. Confira a diferença na imagem abaixo ou leia estas duas postagens para obter exemplos adicionais de como rotular e categorizar gráficos de dispersão:



## Outras formas de visualização (Gráfico de Dispersão)

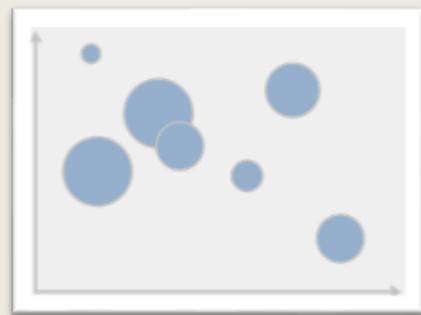


## O que é um gráfico de BOLHAS?

Uma extensão de um gráfico de dispersão, um gráfico de bolhas é comumente usado para visualizar relacionamentos entre três ou mais variáveis numéricas. Cada bolha em um gráfico representa um único ponto de dados. Os valores para cada bolha são codificados por 1) sua posição horizontal no eixo x, 2) sua posição vertical no eixo y e 3) o tamanho da bolha. Às vezes, a cor da bolha ou seu movimento na animação podem representar mais dimensões.



- Como interpretar um gráfico de bolhas
- Como identificar o caso de uso apropriado
- Como projetar um impactante (e formas relacionadas de praticar)



## Como você interpreta um gráfico de bolhas?

Ilustrarei como um gráfico de bolhas funciona construindo-o peça por peça. Imagine que você trabalha para uma organização global e está reunindo alguns dados para uma análise competitiva. Você reuniu a tabela a seguir, que mostra a participação de mercado de cada concorrente, volume de vendas, crescimento de vendas ano a ano e sua região principal (América do Norte, Europa/Oriente Médio/Africa e Ásia-Pacífico).

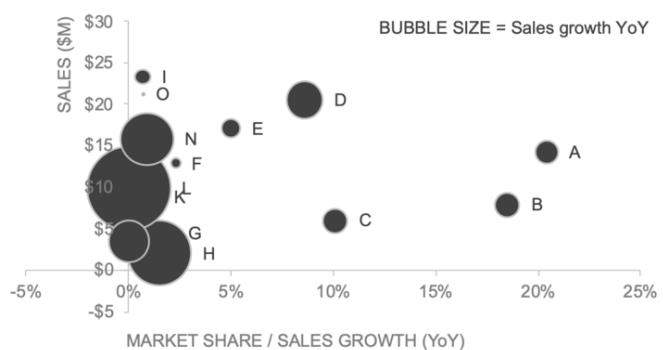
As duas primeiras colunas de dados (participação de mercado e volume de vendas de cada concorrente) são exibidas no gráfico abaixo. Isso é chamado de gráfico de dispersão, que visualiza a relação entre duas séries: o eixo x (participação de mercado) e o eixo y (volume de vendas).

**Competitive landscape**

COMPETITOR	Market share	Sales (\$M)	Sales growth	Region
A	20.4%	\$14.2	9.0%	NA
B	18.5%	\$7.9	10.0%	NA
C	10.1%	\$6.0	9.8%	NA
D	8.6%	\$20.5	15.0%	EMEA
E	5.0%	\$17.1	7.6%	EMEA
F	2.3%	\$12.9	3.3%	NA
G	2.0%	\$4.4	5.6%	EMEA
H	1.5%	\$2.1	26.8%	APAC
I	0.7%	\$23.3	5.9%	EMEA
J	0.7%	\$16.7	-4.2%	EMEA
K	0.2%	\$8.8	24.8%	APAC
L	0.0%	\$9.9	35.1%	APAC
M	0.0%	\$3.5	17.0%	NA
N	0.9%	\$15.8	21.8%	NA
O	0.7%	\$21.2	0.8%	NA

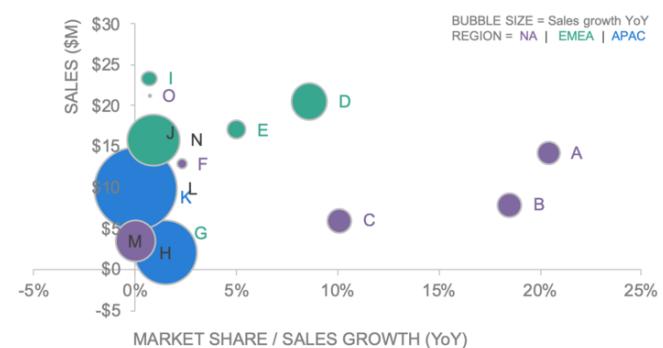
Um gráfico de bolhas se expande no gráfico de dispersão adicionando dimensões adicionais. Podemos adicionar o crescimento das vendas YoY como a terceira dimensão, codificada pelo tamanho das bolhas:

**Competitive landscape**



Essa terceira dimensão dá uma noção visual de quanto os concorrentes diferem uns dos outros com relação à variação de suas vendas: quanto maior o crescimento, maior a bolha. Poderíamos até dar um passo adiante e codificar a quarta variável (Região) por cor:

**Competitive landscape**





## Como saber quando usar um gráfico de bolhas?

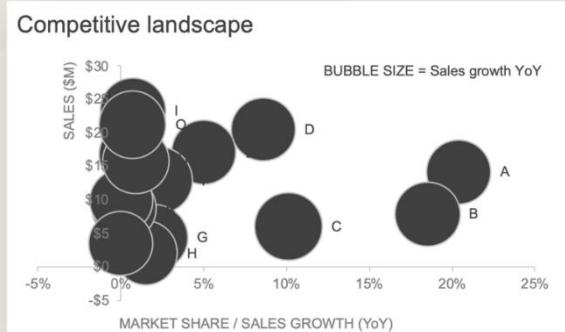
Estamos pedindo muito ao público quando mostramos a eles um gráfico de bolhas. Como são muitas informações para os espectadores acompanharem, precisamos pensar quando os benefícios do gráfico de bolhas superam os custos. Você pode considerar o uso de um quando seus dados tiverem pelo menos três dimensões - use um gráfico de dispersão para apenas duas dimensões - e haja variação suficiente nas métricas para uma comparação visual de três vias atraente. É importante, mesmo neste caso de uso, ajudar seu público a tomar decisões de design inteligentes; descreverei algumas práticas recomendadas para criar gráficos de bolha eficazes momentaneamente.

Existem outros cenários em que um gráfico de bolhas pode não ser a escolha mais eficaz. Por exemplo, você não deve usar um gráfico de bolhas.

Para fins de ilustração, vamos modificar ligeiramente o crescimento das vendas ano a ano, tornando-o mais consistente entre os concorrentes:

Competitive landscape				
COMPETITOR	Market share	Sales (\$M)	Sales growth	Region
A	20.4%	\$14.2	15.0%	NA
B	18.5%	\$7.9	15.1%	NA
C	10.1%	\$6.0	15.8%	NA
D	8.6%	\$20.5	14.9%	EMEA
E	5.0%	\$17.1	15.2%	EMEA
F	2.3%	\$12.9	15.6%	NA
G	2.0%	\$4.4	15.7%	EMEA
H	1.5%	\$2.1	14.9%	APAC
I	0.7%	\$23.3	15.3%	EMEA
J	0.7%	\$16.7	15.5%	EMEA
K	0.2%	\$8.8	15.0%	APAC
L	0.0%	\$9.9	14.8%	APAC
M	0.0%	\$3.5	15.1%	NA
N	0.9%	\$15.8	15.4%	NA
O	0.7%	\$21.2	15.7%	NA

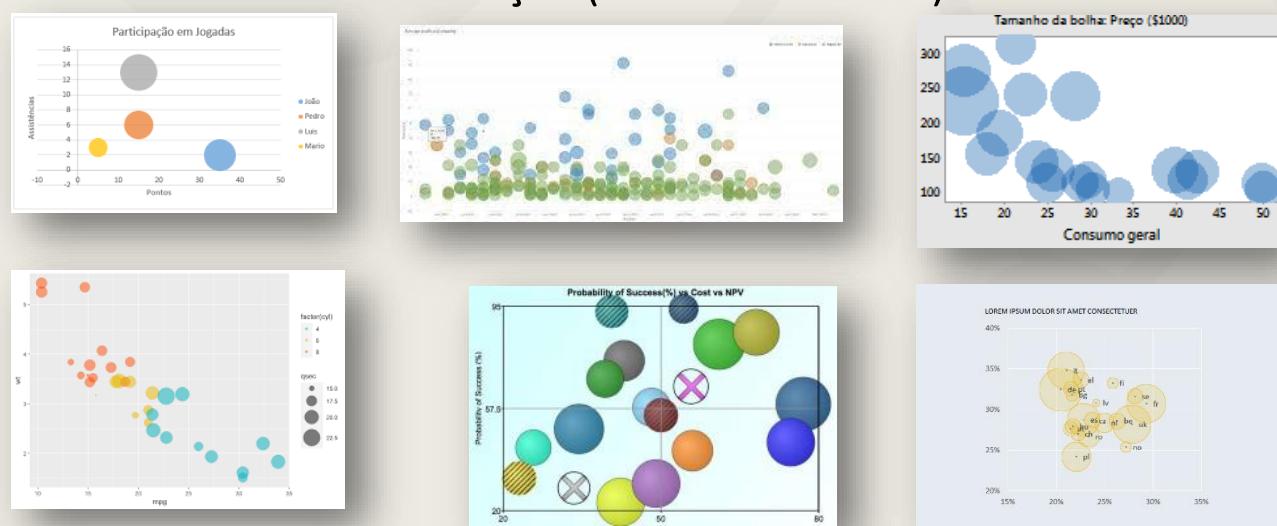
O gráfico de bolhas resultante é mostrado abaixo. Devido à variação mínima, as bolhas são tão semelhantes em tamanho que é difícil extrair qualquer percepção significativa dessa dimensão adicional. Os olhos humanos não estão bem equipados para fazer comparações visuais precisas por área, que abordarei com mais detalhes abaixo.



## Quais são as dicas de design para criar um gráfico de bolhas eficaz?

Como mencionei anteriormente, mesmo o melhor gráfico de bolhas pode ser um desafio para o visualizador entender à primeira vista. Aqui estão algumas coisas que você pode fazer pelo seu público para facilitar a interpretação dos gráficos de bolhas.

## Outras formas de visualização (Gráfico de Bolhas)





# O que é um gráfico de Inclinação?

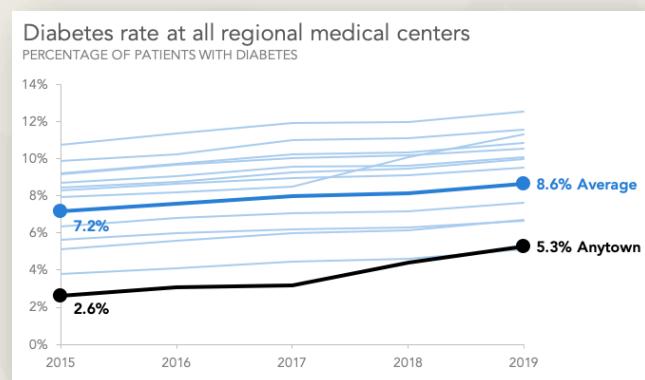
Slopegraphs podem ser usados com dados contínuos ou dados categóricos.

Para gráficos de inclinação que mostram um ponto inicial e final no tempo, ele mostra como os valores mudaram ao longo do período intermediário.

Para gráficos de inclinação que mostram comparações categóricas, ele mostra quais séries de dados são marcadamente diferentes nas duas categorias e em qual direção.

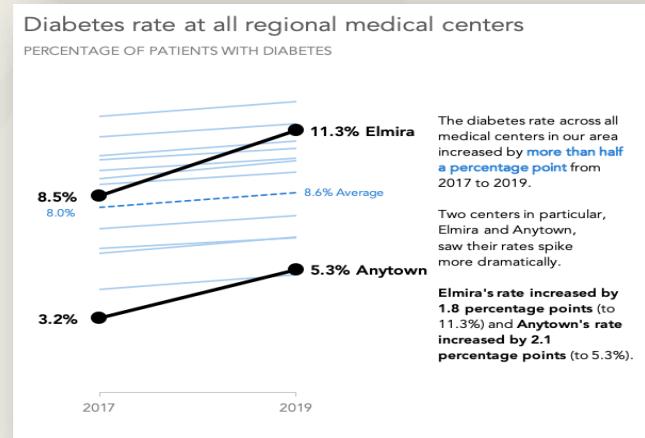
Dados contínuos: destacando a mudança ao longo do tempo

Um gráfico de linha com várias séries, cada uma medida em vários pontos no tempo, pode ficar visualmente confuso rapidamente. Se os valores em cada série tendem a flutuar ao longo do tempo, o ruído no gráfico pode obscurecer os padrões gerais de crescimento ou declínio. Quando você está mais interessado em comunicar "mudança ao longo do tempo" em vez das variações de cada período individual - ou se deseja enfatizar uma série específica que está subindo ou descendo notavelmente em comparação com outras - um gráfico de inclinação é uma opção a ser considerada. Por exemplo, pegue este gráfico de linha das taxas de diabetes em centros médicos em nossa área.



Cada linha representa um dos treze centros médicos da nossa região. Nossa região, Anytown, teve uma das taxas mais baixas de pacientes com diabetes na área nos últimos cinco anos, mas parecia aumentar notavelmente entre 2017 e 2019.

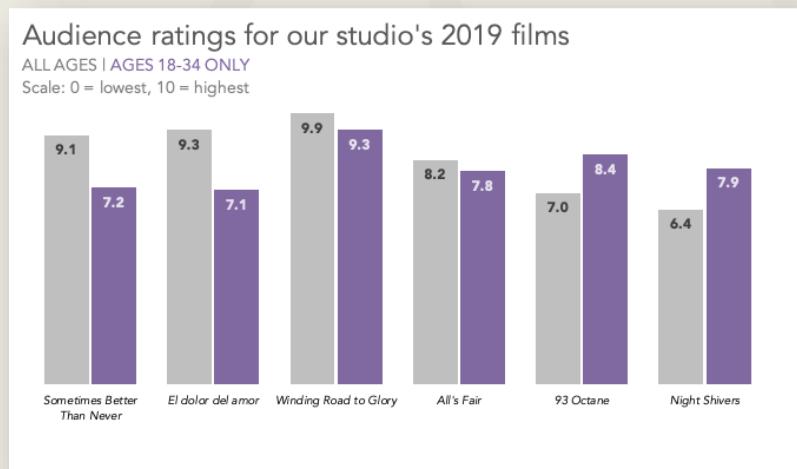
Um gráfico de inclinação pode ajudar a mostrar isso de forma mais óbvia e pode apontar quais outros centros tiveram um aumento semelhante.



## Dados categóricos: destacando diferenças entre duas categorias

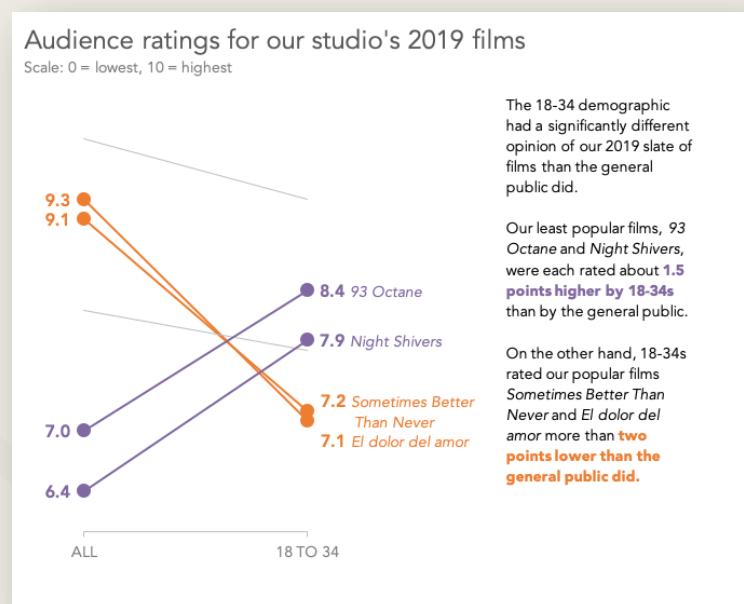
Quando comparamos duas categorias que têm uma escala comum, um gráfico de inclinação pode ser uma excelente maneira de mostrar como as coisas se comparam nessas categorias.

Imagine, por exemplo, que trabalhamos para um estúdio de cinema. Lançamos seis filmes em 2019 e queremos comparar as avaliações gerais do público desses filmes com as avaliações de pessoas no grupo demográfico principal de 18 a 34 anos. Poderíamos mostrar esses resultados como gráficos de barras emparelhados:





Essa visão certamente torna mais fácil comparar as avaliações dos principais dados demográficos com as avaliações gerais de um único filme, mas um gráfico de inclinação tornaria as diferenças em toda a lista de filmes mais óbvias e proeminentes. (Também podemos “aumentar o zoom” para ver as diferenças detalhadas entre cada categoria, porque os gráficos de inclinação, como os gráficos de linhas em geral, não precisam necessariamente começar com uma linha de base zero; os gráficos de barras não nos permitem essa flexibilidade.)



## Quais são algumas razões para não usar um gráfico de inclinação?

Ao trabalhar com dados contínuos de séries temporais, é importante lembrar que os slopegraphs, por design, mostram APENAS a diferença entre um ponto inicial e um ponto final. Estamos, portanto, excluindo todos e quaisquer dados da parte intermediária de nosso período de tempo. Se essas informações fornecerem um contexto essencial, usar um gráfico de linha padrão seria uma escolha melhor.

Slopegraphs não são a melhor escolha para dados categóricos quando não há nenhuma conexão real entre as categorias selecionadas. (É quando os gráficos de inclinação acabam parecendo pilhas de espaguete seco jogados no chão.) Um gráfico de inclinação categórico ideal usa categorias relacionadas (mas não necessariamente relacionadas causalmente).

Uma última ressalva é que os gráficos de inclinação, como todos os gráficos de linha, destinam-se principalmente a enfatizar a mudança entre dois valores ou as taxas comparativas de mudança em várias séries - não as quantidades absolutas desses valores. Use um gráfico de barras se for importante transmitir que suas vendas cresceram precisamente 20% ou US\$ 300.000; use um gráfico de inclinação para mostrar que suas vendas subiram mais rapidamente do que as de seus concorrentes ou que aumentaram enquanto as vendas de outros escritórios diminuíram.

## Outras formas de visualização (Gráfico de Inclinação)

**NÃO DISPONIVEL**



# Como usar gráficos de Tabela?

O que é uma mesa?

Uma tabela organiza os dados em uma estrutura tabular, composta por linhas e colunas. Eles são intuitivos de ler, razoavelmente fáceis de construir e incrivelmente flexíveis. Ao contrário de alguns dos gráficos anteriores que discutimos - barras , pizzas , linhas , etc . - As tabelas não são restritas a um tipo específico de dados ou relacionamento subjacente. Eles podem conter valores numéricos, texto, símbolos ou até mesmo uma combinação de tipos de dados. Por esses motivos, as tabelas são comuns e preferidas por muitos.

Embora aparentemente simples na superfície, as tabelas são mais complexas do que parecem. É fácil pensar que eles são intercambiáveis com outros visuais básicos, como barras e linhas, mas há uma distinção importante a ser entendida: processamos tabelas de maneira diferente de gráficos. As tabelas interagem com nosso sistema de comunicação verbal, o que significa que as lemos. Nossos olhos examinam as linhas e as colunas. Em comparação, os gráficos interagem com nosso sistema visual. Nossa sistema visual é muito mais rápido no processamento de grandes quantidades de informação , especialmente quando há um padrão ou forma que queremos transmitir. Entender isso antecipadamente nos ajuda a perceber quando é apropriado usar uma tabela ou, dito de outra forma, quando nosso público deve ler os dados em vez de decodificar um visual.

Você tem dados variados. Às vezes, precisamos comunicar métricas relacionadas com várias unidades de medida ou intervalos variados. Resumos executivos, scorecards, detalhamentos demográficos e metadados são coisas que vêm à mente. Nesse caso, o usuário final provavelmente fará referência a esses itens para procurar valores específicos, portanto, uma tabela é uma escolha eficaz. Além disso, as tabelas não estão sujeitas às mesmas restrições de espaço bidimensionais que os gráficos, portanto, pode ser mais fácil exibir diferentes unidades e intervalos em uma única visualização.

Seu público tem necessidades diferentes. As tabelas podem ser ótimas quando você tem um público diversificado. Já mencionei que as tabelas são intuitivas, mas também permitem que cada usuário final procure valores específicos. Se você antecipar que seu público desejará verificar dados diferentes, as tabelas podem ser uma boa solução.

Você está lendo linha por linha. Discutir números e itens orçamentários dentro de uma tabela é comum entre o público financeiro ou contábil. Um ex-diretor meu tinha formação em contabilidade e, como resultado, desenvolveu uma afinidade por tabelas. Ele aprendeu a identificar padrões dentro de tabelas tão bem que solicitou tabelas de dados em vez de gráficos. Trago isso por um bom motivo - se o seu público tem um forte desejo por mesas, reconheça que esse é outro excelente motivo para usá-las.

Você está complementando a história principal. Ao comunicar dados para fins explicativos, raramente mostramos todos os dados, optando por destacar os principais tópicos. Pode ser útil fornecer ao público o conjunto de dados completo ou detalhamentos adicionais, especialmente se seu público quiser consultar os detalhes posteriormente ou explorar os dados de forma independente. Para atender a essa necessidade, considere vincular a uma tabela subjacente ou adicionar uma tabela detalhada a um apêndice.

## Quando você não deve usar uma tabela para comunicar dados?

Sempre haverá exceções, mas duas instâncias principais vêm à mente quando pode não ser o ideal usar uma mesa. Ambos os cenários estão relacionados à forma como processamos as tabelas.

Você está apresentando seus dados ao vivo - pessoalmente ou virtualmente. Se você compartilhar uma tabela de dados em uma apresentação ao vivo, seu público provavelmente começará a lê-la. Infelizmente, não podemos ouvir e ler ativamente simultaneamente: se o seu público está lendo a tabela, eles o desligaram. Nessa situação, eu recomendaria um gráfico (geralmente, menos trabalhoso e demorado

para processar) sobre uma tabela. Se você for obrigado a usar uma tabela durante uma apresentação, considere como seu público irá interpretá-la. Você pode fazer uma pausa, dando-lhes tempo para ler e depois continuar falando? Existe uma maneira de limitar linhas e colunas desnecessárias ou direcionar a atenção para um subconjunto dos dados, tornando a tabela mais rápida de digerir?



Você deseja enfatizar um padrão ou forma nos dados. Imagine que você deseja compartilhar um padrão sazonal que identificou em alguns dados mensais contínuos. Para localizar isso em uma tabela, você teria que comparar e manter vários valores mensais em sua cabeça, o que é um processo mentalmente desgastante. Se houver um padrão ou algo significativo sobre a forma de seus dados, um gráfico será uma escolha melhor do que uma tabela. Dito isso, é possível adicionar componentes visuais a uma tabela — exploraremos isso mais adiante.

## Como você pode projetar uma mesa melhor?

Quase todas as dicas compartilhadas abaixo têm a ver com tornar as tabelas mais fáceis de escanear. As tabelas geralmente contêm grandes quantidades de detalhes, portanto, devemos tomar medidas intencionais para torná-las fáceis de consumir.

Esclarecer horizontal ou vertical. As tabelas são consumidas nas linhas ou nas colunas. Evite que seu público pense em qual caminho digitalizar adicionando dicas visuais. Existem várias maneiras de fazer isso; dois dos meus métodos preferidos são adicionar uma série resumida e aproveitar o espaço em branco. Se eu prevejo que meu público desejará resumir os dados em um total ou média, adicionar uma série de resumos - e diferenciá-los visualmente - pode economizar uma etapa e orientá-los durante a leitura.

Summarize series

•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
TOTAL	•	•	•	•	•	•

Use white space to create columns or rows

•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•

Remova bordas e sombreamentos desnecessários. Muitos padrões de tabela incluem bordas pesadas e sombreamento. Eles podem ajudar a separar os cabeçalhos dos dados ou estabelecer categorias e valores de resumo, mas sua presença deve ser mínima. Minha preferência é usar linhas de borda em vez de sombreamento, pois isso me permite usar a cor de forma mais estratégica, como em um mapa de calor (mais sobre isso a seguir).

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

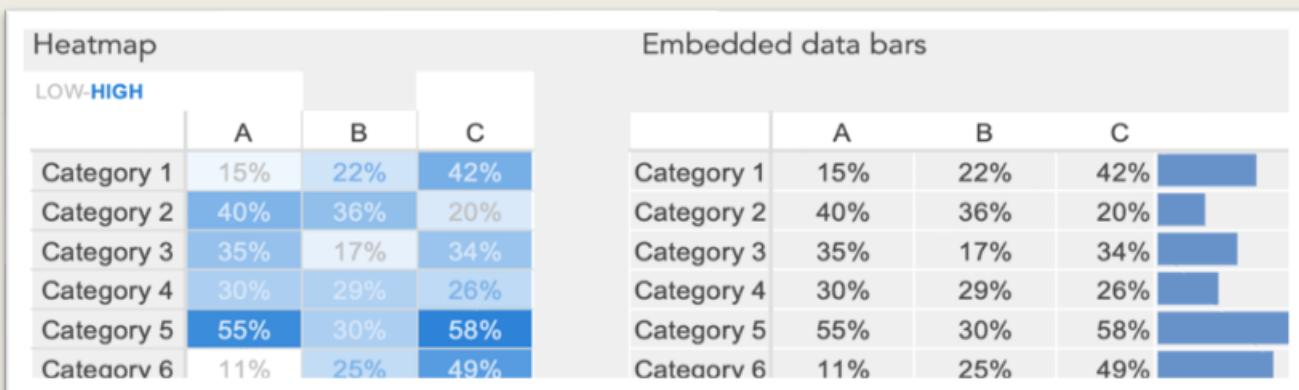
Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Incorpore um elemento visual. Considere adicionar elementos visuais para ajudar a chamar a atenção para os dados mais importantes ou para comunicar um padrão. Você pode adicionar cores, mini gráficos ou até mesmo barras de dados. Um caso especializado é conhecido como mapa de calor, em que a saturação da cor é usada para representar visualmente os valores relativos. Consulte as postagens relacionadas sobre como melhorar uma tabela, incorporar barras horizontais e uma reforma de mapa de calor para obter exemplos específicos.



Ordem intencionalmente. Quando processamos uma página, normalmente começamos no topo e nos movemos da esquerda para a direita, em um padrão de ziguezague na página. Isso também é verdade para uma tabela e pode nos ajudar a decidir como ordenar linhas e colunas. Se houver uma métrica importante ou um valor de resumo, pode fazer sentido movê-lo para o canto superior esquerdo da tabela, portanto, é uma das primeiras coisas que seu público vê. Tenha em mente que isso não será apropriado em todos os cenários. Se sua tabela incluir um componente de tempo, por exemplo, considere aproveitar a construção natural do tempo movendo-se da esquerda para a direita, mesmo que os dados recentes sejam os mais importantes.

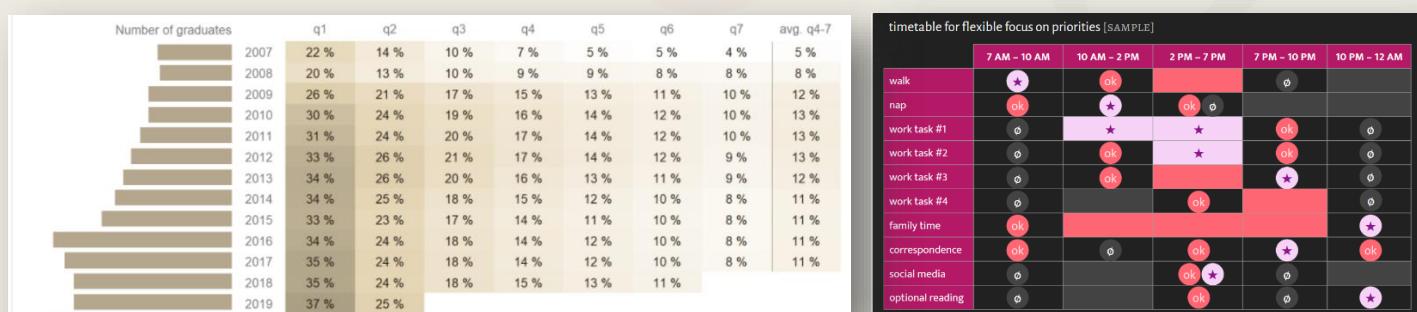
Use rótulos de supercategoria. A criação de rótulos de nível superior adiciona estrutura a uma tabela, facilitando a navegação. Os rótulos de supercategoria também têm o benefício adicional de limitar a redundância (confira esta postagem relacionada sobre o poder da categorização). Considere o caso de uma tabela de resumo executivo contendo KPIs para algumas unidades de negócios. Em vez de listar cada KPI separadamente, agrupe os KPIs relacionados, para que alguém possa identificar rapidamente todas as medidas relacionadas a um departamento específico.

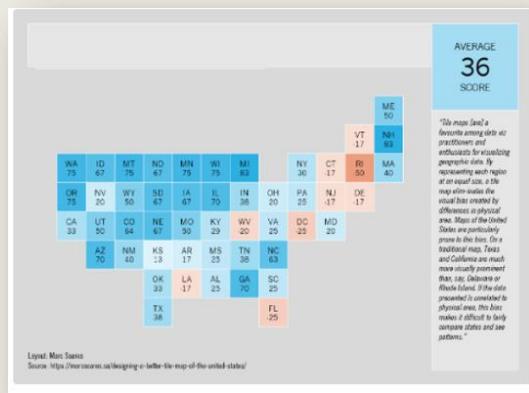
Alinhe cuidadosamente. Existem várias abordagens para justificar os dados em uma tabela. Uma preferência popular é alinhar o texto à esquerda e justificar os números à direita. Essa é uma ótima diretriz, mas não uma regra. Em vez disso, procure ser cuidadoso sobre o alinhamento que você usa. O alinhamento central permite espaço em branco para distinguir visualmente as colunas? Se estiver usando decimais, é mais fácil comparar valores relativos se você alinhar os pontos decimais? Jogue com diferentes opções e veja qual funciona melhor para seus dados. (Em uma nota relacionada, certos tipos de letra permitem alinhamento tabular criando colunas de números; para saber mais, leia o artigo de Matthew Ström sobre como projetar tabelas melhores.)

Mixed alignment			Center alignment			Align decimals (for numbers)		
Group	Metric A	Metric B	Group	Metric A	Metric B	Group	Metric A	Metric B
Group 1	1.11	AAA	Group 1	1.11	AAA	Group 1	1.11	AAA
Group 2	2.2	BBBB	Group 2	2.2	BBBB	Group 2	2.2	BBBB
Group 3	3.33	CC	Group 3	3.33	CC	Group 3	3.33	CC
Group 4	44.4	D	Group 4	44.4	D	Group 4	44.4	D
Group 5	555.5	EEEEE	Group 5	555.5	EEEEE	Group 5	555.5	EEEEE

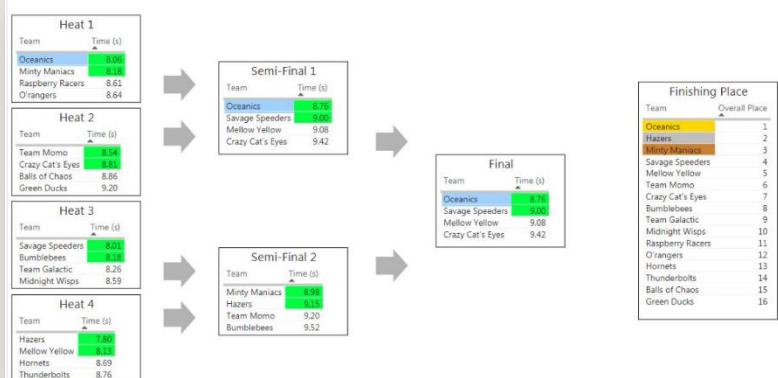
## Outras formas de visualização (Tabelas)

Referência: community.storytellingwithdata





The Oceanics cruise to victory, winning every matchup.



	Calories	Carbs (g)	less of these : (			more of these : )				
			Fat (g)	Sugar (g)	Sodium (mg)	Protein (g)	Fiber (g)	Calcium (%)	Potassium (%)	Iron (%)
<b>Breakfast - Oatmeal</b>										
1 large banana	121	31	0	17	1	2	4	0	12	1
1 cup vanilla almond milk	30	1	3	0	115	1	0	30	2	2
1 cup rolled oats	300	54	6	2	0	10	8	8	19	41
<b>Lunch - Homemade "Lunchables"</b>										
1 oz. cheddar cheese	113	0	9	0	174	7	0	12	0	1
15 wheat thins	140	22	5	5	200	2	3	2	0	4
1tbsp hummus	70	5	5	0	130	2	1	0	2	4
10 baby carrots	23	5	0	3	43	1	1	1	0	2
8 grapes	20	7	0	15	0	0	0	1	0	0
15 seasoned almonds	170	5	16	1	150	6	3	8	5	6
1 hardboiled egg	78	0	5	0	62	6	0	2	1	3
<b>Snack - Halloween Cookie</b>										
1 pillsbury sugar cookie	85	11	4	7	2	1	0	0	0	0
<b>Dinner - Crockpot Chicken</b>										
4 oz. crockpot chicken	620	48	22	0	1,111	52	1	0	0	11
1 cup white rice	410	3	21	0	772	48	0	0	0	0
<b>Dessert - Fudge</b>										
1 piece of Fudge	70	13	2	12	8	0	0	0	0	0
Total	1,840	202	76	62	1,996	90	21	64	41	75

## O que é um gráfico de PONTOS?

O termo “dot plot” pode ser usado para qualquer gráfico que esteja codificando dados em um ponto ou um pequeno círculo. Existem alguns tipos comuns nos quais focarei aqui. Se você já se perguntou: o que é um gráfico de pontos? Como interpreto um gráfico de pontos? Quando devo usar um gráfico de pontos? ou quais são os prós e os contras dos gráficos de pontos?

### Quais são os tipos comuns de gráficos de pontos?

Existem três variações comuns do gráfico de pontos:

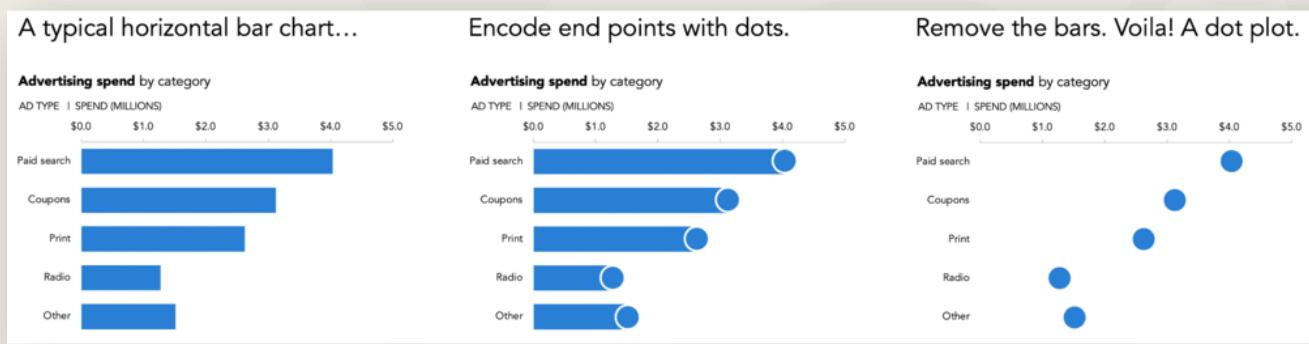


O primeiro, eu me referia como um gráfico de pontos tradicional (rotulado simplesmente como “gráfico de pontos” acima). Este gráfico foi usado indo muito para trás no tempo. Eles foram originalmente desenhados à mão e são usados para mostrar a distribuição dos dados. Isso pode ser útil se você quiser ver seus dados para ter uma noção da forma ou identificar valores discrepantes. Parece semelhante a um histograma (que codifica uma



distribuição de dados em barras em vez de pontos). No entanto, neste tipo de gráfico de pontos, como cada ponto representa um ponto de dados individual, os pontos não precisam necessariamente ser espaçados uniformemente ao longo do eixo horizontal. Eles tendem a funcionar melhor para pequenos conjuntos de dados (quando você tem mais dados, algo que agregue os pontos individuais, como um histograma ou gráfico de caixa, pode ser mais adequado).

Hoje, o tipo mais comum de gráfico de pontos que encontro (pelo menos em um ambiente de negócios) é o desenvolvido por William S. Cleveland, às vezes chamado de gráfico de pontos de Cleveland. Este gráfico codifica dados quantitativos entre categorias. Uma maneira de pensar nisso é como uma alternativa a um gráfico de barras. Confira o exemplo a seguir para ver o que quero dizer.

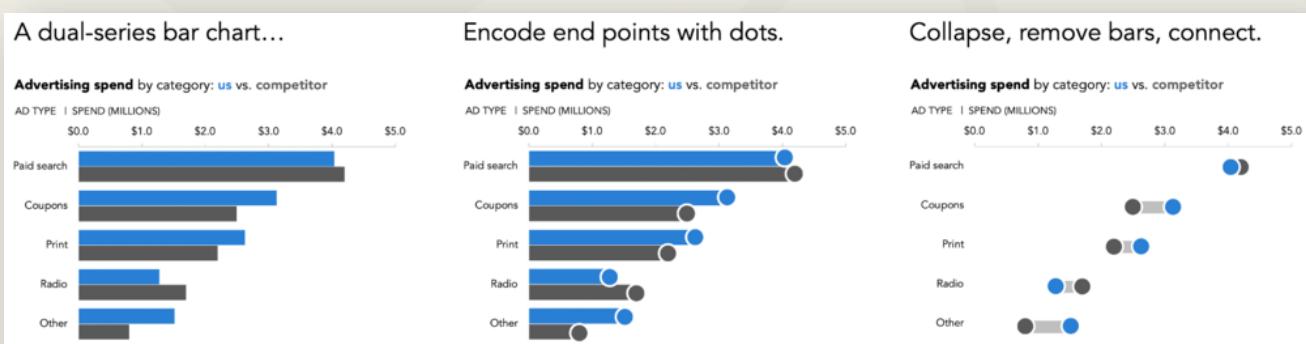


Você não precisa realmente passar por essa progressão para criar um gráfico de pontos. Em vez disso, este exemplo destina-se simplesmente a ilustrar como esse tipo de gráfico de pontos funciona, fazendo a transição para ele de algo familiar.

A terceira visualização acima não precisa ser a final. Talvez você possa imaginar alterações de formatação adicionais que poderíamos aplicar. Por exemplo, poderíamos rotular os valores individuais dentro de cada ponto, eliminando a necessidade do eixo x. Podemos adicionar linhas ou linhas de grade claras de cada categoria à esquerda em seu respectivo ponto para dar maior estrutura ao gráfico (esta última variação com linhas de eixo a ponto é geralmente chamada de gráfico de pirulito).

Ilustrei com uma versão horizontal neste caso, mas você também pode girar (imagine os pontos substituindo um gráfico de barras verticais de maneira semelhante à que ilustrei acima). A orientação escolhida dependerá dos próprios dados, do comprimento dos nomes das categorias e de outros elementos. Assim como os gráficos de barras, pode valer a pena considerar os dois layouts (vertical e horizontal) para determinar qual faz mais sentido ou parece mais intuitivo na situação e nos dados fornecidos (em caso de dúvida, peça feedback a outras pessoas).

Uma variação comum adicional é o gráfico de pontos conectados. Ele funciona da mesma forma que o gráfico de pontos de Cleveland, mas representa graficamente duas (ou às vezes mais) séries de dados. Vamos usar o mesmo exemplo anterior, só que agora adicionarei uma segunda série de dados.



O gráfico de pontos conectados codifica os valores absolutos (através das posições dos pontos), mas enfatiza a *diferença* ou *alteração* entre as séries de dados devido à conexão visual entre os pontos.



## Quais são alguns prós e contras dos gráficos de pontos?

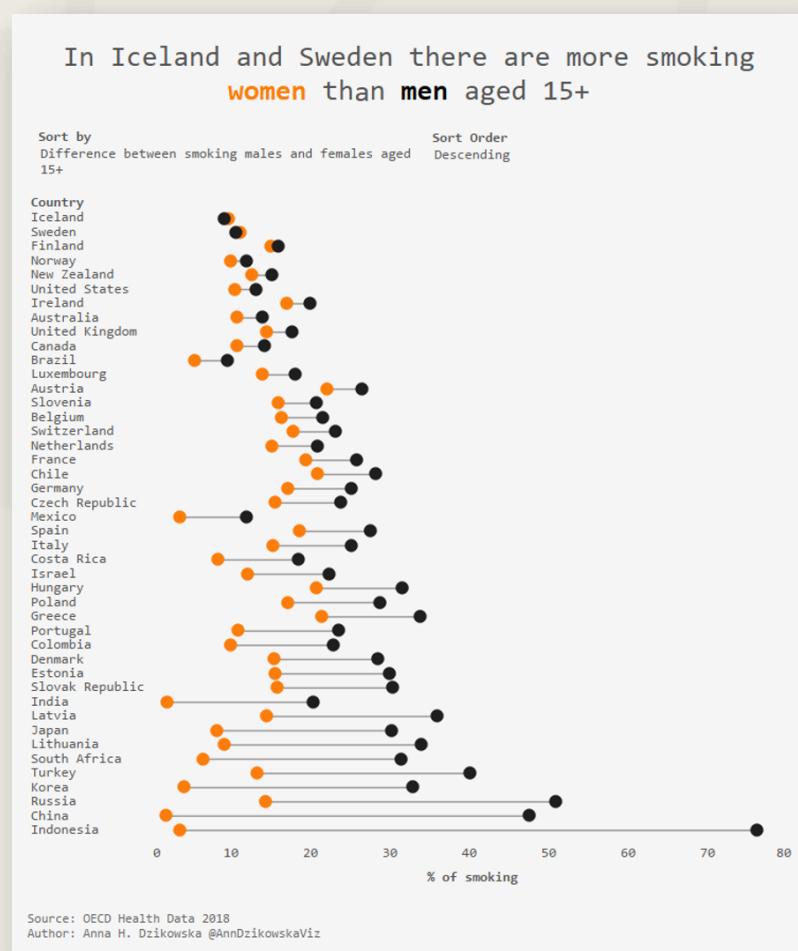
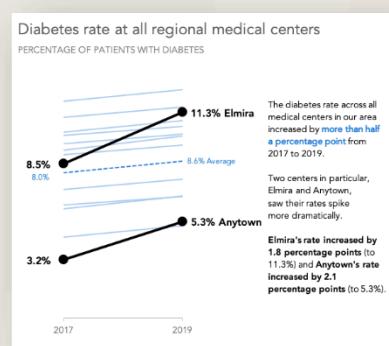
Em comparação com o gráfico de barras mais comum, os pontos em um gráfico usam menos tinta, o que significa que há mais espaço para adicionar anotações ou simplesmente ter um gráfico mais leve. Para Cleveland e gráficos de pontos conectados, uma liberdade adicional que às vezes é útil é que o eixo não precisa começar em zero (como acontece com um gráfico de barras ou um gráfico de pontos tradicional, que codifica dados por comprimento ou altura; com o outro variações do gráfico de pontos, focamos mais na posição relativa no espaço do que na distância do eixo).

Dito isso, os gráficos de pontos não são tão comuns quanto as barras. Em alguns casos, isso pode ser uma vantagem (entediado com barras? Tente um gráfico de pontos!). No entanto, eles podem ser recebidos por audiências desconhecidas com resistência ou confusão, então avalie se faz sentido usar um dado a situação específica e tome medidas cuidadosas para rotular e explicar.

Onde posso ver gráficos de pontos de exemplo?

Vamos ver mais alguns aqui! O gráfico de pontos conectados é o mais matizado e tem alguns casos de uso interessantes e variados, então vamos dar uma olhada em alguns exemplos deles. Por meio deles, você também verá diversas opções de design ilustradas.

Às vezes, o gráfico de pontos conectados pode ser útil para mostrar a comparação de grupos (um gráfico de inclinação é outra alternativa). Por exemplo, o visual a seguir (criado por Anna Dzikowska) mostra a porcentagem de mulheres em comparação com a porcentagem de homens que fumam por país.





Como você classifica os dados é uma consideração ao projetar um gráfico de pontos (como é para muitos tipos de gráficos). No exemplo anterior, os países são organizados pela diferença – com casos em que uma proporção maior de mulheres fuma do que de homens no topo, até a maior diferença entre a proporção de mulheres e homens na base.

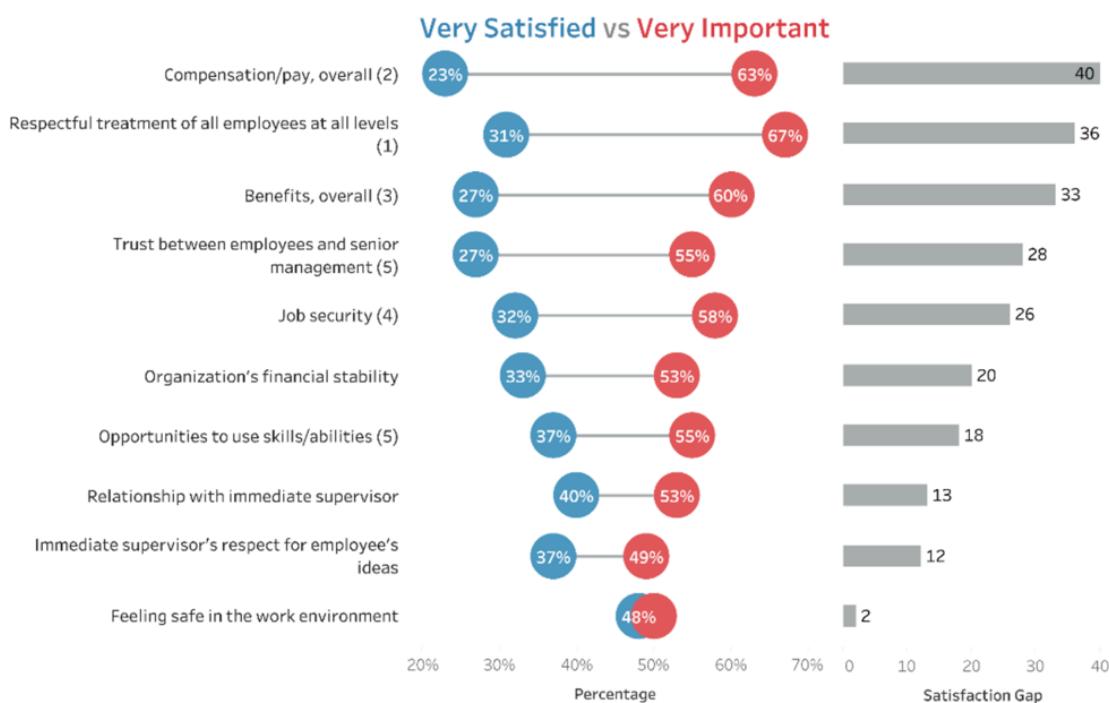
Isso destaca um desafio que às vezes surge, que é quando as séries flip-flop, mudando de posição uma em relação à outra, não necessariamente se destacam. Por exemplo, a Islândia e a Suécia no exemplo anterior são os dois únicos casos em que os pontos laranja estão acima dos pontos pretos; em todos os outros casos, os pontos pretos são maiores. Mas é preciso pensar e processar para reconhecer isso (ou pelo menos foi para mim, mesmo que tenha sido mencionado no título!). Por causa disso, às vezes pode ser útil aplicar formatação adicional para deixar essa diferença clara.

Por exemplo, no diagrama de pontos conectado a seguir (criado por Louise Shorten), a cor é usada para ajudar a diferenciar visualmente a direcionalidade no gráfico de pontos. Este exemplo também ilustra outro caso de uso potencial para um gráfico de pontos conectado: comparação de tempo. O seguinte mostra a mudança de preço do MacDonald's Big Mac entre dois pontos no tempo.

A comparação de tempo talvez seja um caso de uso relativamente mais raro para um gráfico de pontos (e restritivo, pois normalmente funciona apenas para dois pontos no tempo). No outro extremo do espectro, um lugar comum que vejo gráficos de pontos conectados usados em um ambiente de negócios é comunicar dados de pesquisa. Veja o exemplo a seguir, por exemplo (criado por Leigh Fonseca). Ele visualiza a “lacuna de satisfação” – a diferença entre a proporção de funcionários que dizem que um determinado fator de seu trabalho é muito importante em comparação com aqueles que expressam alta satisfação.

## THE SATISFACTION GAP

How can we expect employees to be engaged when they're not satisfied with their most important job aspects?



Source: Society for Human Resource Management, 2016 Employee Job Satisfaction and Engagement Report  
<https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/2016-Employee-Job-Satisfaction-and-Engagement-Report.pdf>

@missleigh

Os gráficos de pontos para dados de pesquisa também são usados às vezes para simplesmente relatar resultados em itens ou categorias de pesquisa (usando um gráfico de pontos de Cleveland) ou para comparar dados de pesquisa entre grupos (com um gráfico de pontos conectado).



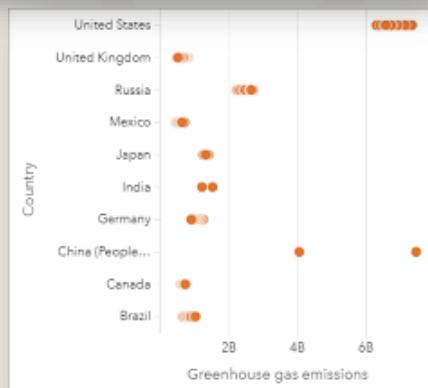
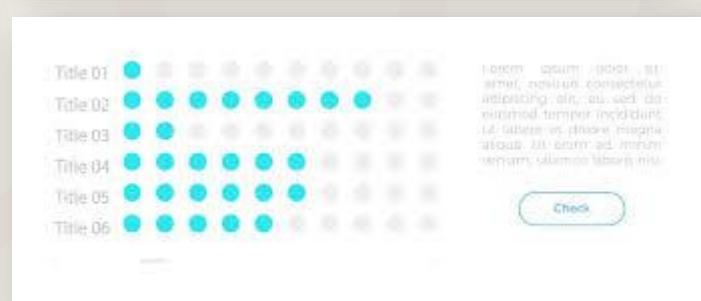
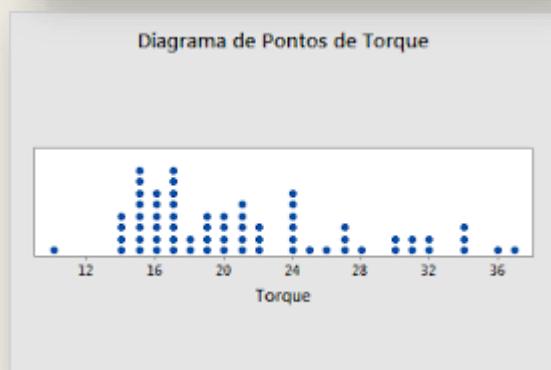
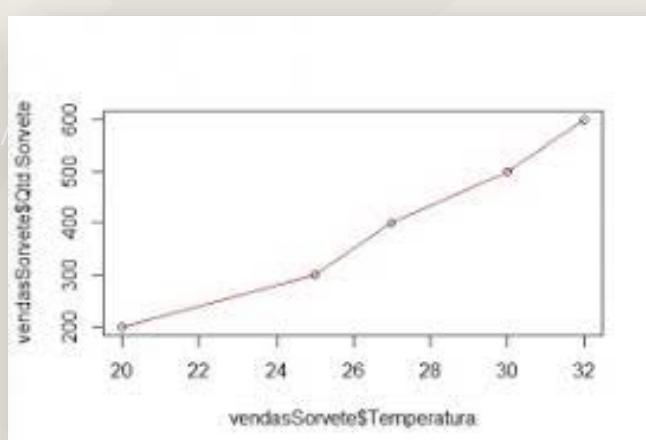
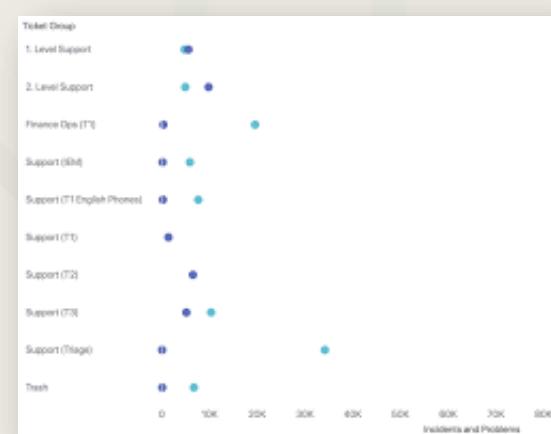
# Como faço um gráfico de pontos?

Agora que exploramos os tipos comuns de gráficos de pontos e vimos vários exemplos, você pode estar se perguntando: "Como eu realmente faço um desses gráficos?" Algumas ferramentas possuem funcionalidade integrada para criar gráficos de pontos. Mesmo naqueles que não o fazem, no entanto, muitas vezes um pensamento um pouco fora da caixa tornará possível a criação de um.

Por exemplo, no Excel (e imagino que isso funcionaria na maioria das ferramentas), você pode usar um gráfico de dispersão para criar um gráfico de pontos de Cleveland. Dependendo se você está atrás de orientação vertical ou horizontal, você coloca seus valores de dados como valores y ou x (respectivamente) e então insere valores para o outro que você usará para separar suas categorias (por exemplo, na horizontal Gráfico de pontos de Cleveland que mostrei em minha explicação no início deste post, os valores x no gráfico de dispersão são meus valores de dados e os valores y são 1, 2, 3, 4 e 5, que são simplesmente para separar as categorias). Para orientação vertical, usar um gráfico de linha também funcionaria, onde você adiciona marcadores de dados para seus pontos e formata a linha que os conecta para ficar invisível.

Criar o gráfico de pontos conectado em uma ferramenta em que a funcionalidade integrada não existe é um pouco mais complicado, mas ainda é possível. Isso normalmente será uma combinação de um gráfico de dispersão ou de linhas (com pontos enfatizados e linhas ocultas, conforme descrito acima) para os pontos e um gráfico de barras empilhadas (com a proporção à frente do primeiro ponto não preenchida) para a conexão.

## Outras formas de visualização (Gráfico de Pontos)



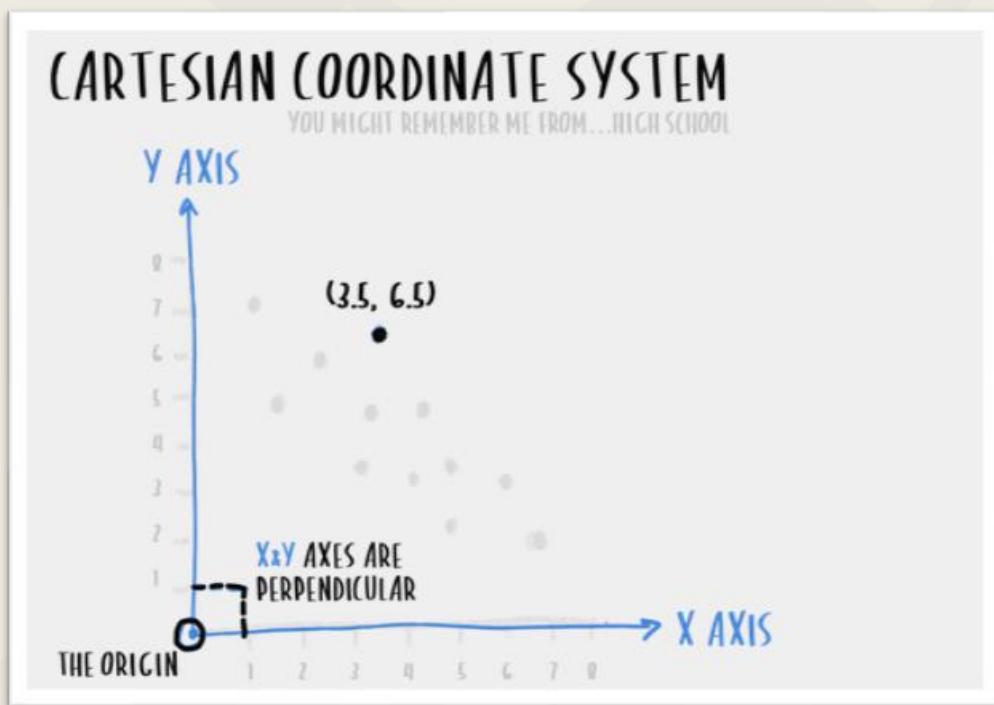


# O que é um gráfico de Aranha?

Assim como gráficos de linhas, gráficos de dispersão e muitos gráficos usados em comunicações comerciais, os gráficos de aranha mostram dados plotados em dois eixos diferentes. No entanto, o sistema de coordenadas que usamos para plotar dados em um gráfico de aranha é um pouco menos comum: a maioria dos gráficos de negócios é desenhada usando coordenadas cartesianas, mas os gráficos de aranha usam coordenadas polares.

## Coordenadas cartesianas

Os gráficos bidimensionais mais comuns, como os gráficos de dispersão, usam coordenadas cartesianas — usamos os valores X e Y para determinar as localizações dos pontos em dois eixos perpendiculares: o horizontal e o vertical, respectivamente.



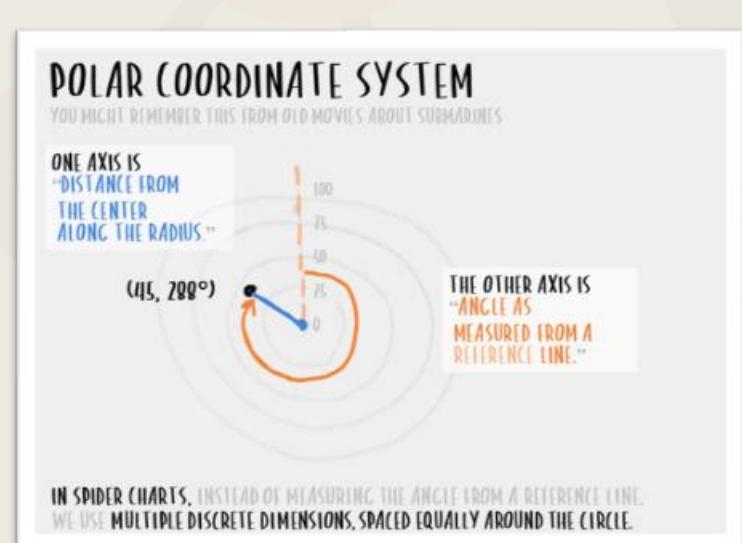
Um sistema de coordenadas cartesianas plota pontos de dados ao longo de dois eixos perpendiculares.

Cada ponto recebe um par de valores (o X e o Y), e podemos olhar para a localização X/Y de muitos pontos ao mesmo tempo. Dessa forma, podemos ter uma boa noção da relação entre a variável X e a variável Y em todos os itens de nosso conjunto de dados.

Dito de forma mais simples: esse tipo de gráfico e sistema de coordenadas destina-se a ajudar o visualizador a comparar duas medições relacionadas em vários itens diferentes.

## Coordenadas polares

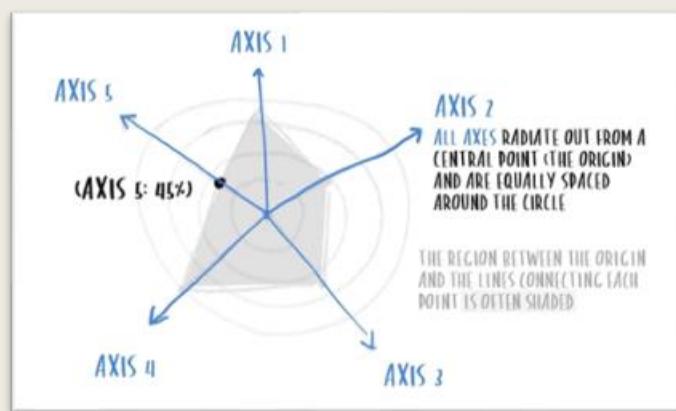
Em um sistema de coordenadas polares, embora usemos dois valores para determinar as localizações dos pontos, esses valores são "distância do centro do círculo" (ou "distância do polo") e "ângulo relativo a uma linha de referência".





Um sistema de coordenadas polares também usa dois eixos: um eixo que irradia do centro do círculo e um eixo definido pelo tamanho do ângulo medido a partir de uma linha de referência. Em outras palavras: um eixo é linear e um eixo é radial.

Para gráficos de aranha, nosso segundo eixo (o laranja no desenho acima) não é contínuo. Em vez disso, usamos um número discreto de eixos fixos, cada um representando uma única dimensão, espaçada igualmente ao redor do círculo.



Vê como eles se parecem com raios saíndo do cubo de uma roda? Em um gráfico de aranha, cada dimensão tem seu próprio raio, e os raios são distribuídos uniformemente ao redor da roda. Quanto mais próximo do final do pico, maior o valor. Mais próximo do centro significa mais próximo de zero.

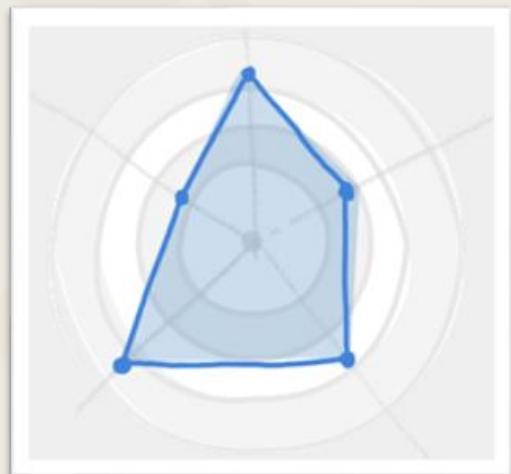
Com a liberdade de usar muitos eixos lineares, poderíamos mostrar várias dimensões diferentes para um único item simultaneamente... mas se quiséssemos comparar vários itens ao mesmo tempo, usariamos vários gráficos individuais (em um pequeno formato, talvez). Se não parecesse muito confuso, também poderíamos fazer uma única sobreposição com vários itens empilhados uns sobre os outros.

Esse tipo de gráfico, então, é mais adequado para comparar várias dimensões diferentes em um espaço compacto, mas para apenas um ou alguns itens.

## Qual é a diferença entre um gráfico de radar e um gráfico de aranha?

Os termos “gráfico de radar” e “gráfico de aranha” são usados de forma mais ou menos intercambiável. É lógico pensar qu talvez só o chamaríamos de gráfico de aranha se fôssemos conectar os pontos que traçamos, para tornar o gráfico em si uma reminiscência de uma teia de aranha..., mas não é o caso. Quer sejam chamados de radar ou gráficos de aranha, os pontos estão quase sempre conectados.

Se várias séries forem exibidas no mesmo gráfico, o gráfico começará a se parecer cada vez mais com uma teia de aranha - especialmente se o designer optar por não usar cores para preencher nenhuma área do gráfico. Frequentemente, a área entre a origem do gráfico e as linhas que conectam os pontos é preenchida com uma cor semitransparente.



Ocasionalmente, você também verá bandas alternadas de círculos concêntricos como preenchimentos. Eles servem para substituir ou aumentar as linhas de grade e ajudar a enfatizar o tamanho e a forma da área preenchida.

Além disso, dependendo da ferramenta de sua escolha, suas linhas de grade podem ser desenhadas como círculos suaves ou podem ser desenhadas como segmentos de linha conectando cada dimensão sucessiva. O efeito, então, é que ao invés de suas linhas de grade serem uma série de círculos concêntricos, elas são uma série de polígonos concêntricos... como você verá em breve no exemplo abaixo.

## Quando (e como) devo usar gráficos de aranha de forma eficaz?

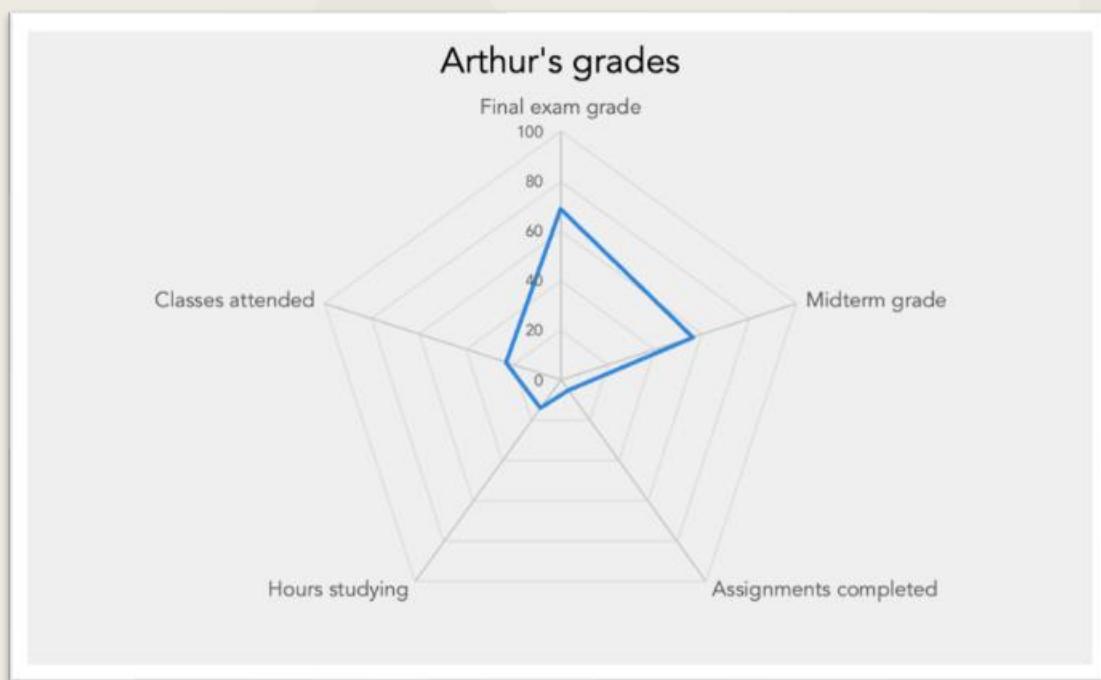
Os gráficos de aranha são melhores quando usados para comparar rapidamente várias dimensões em um espaço compacto. Eles podem chamar a atenção, devido à sua estrutura circular e à sua relativa novidade em comparação com outros gráficos de negócios, para que possam ser eficazes quando você precisa envolver visualmente seu público. O público em geral pode achá-los confusos ou intimidadores de ler sem orientação adicional (que você pode fornecer — falaremos mais sobre isso em uma seção posterior), mas o público técnico pode achá-los intrigantes.



Vamos imaginar que temos uma sala de aula com alunos que acabaram de fazer o exame final. Para cada aluno, sabemos algumas informações diferentes:

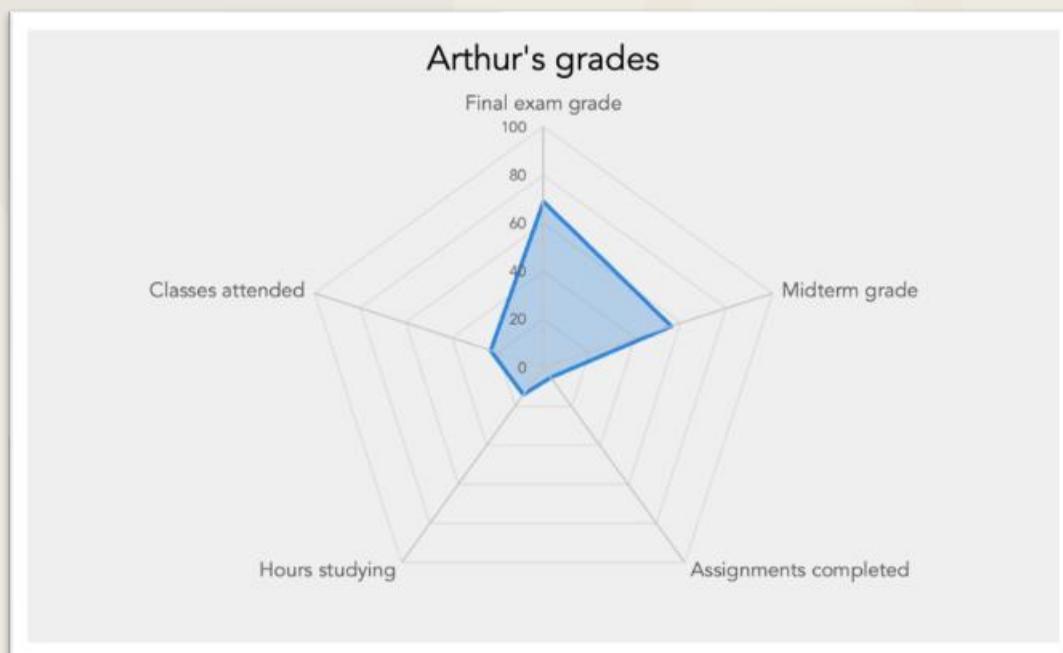
- Nota do exame final
- Nota do exame intermediário
- Tarefas concluídas durante o período
- Horas gastas estudando para a prova
- Dias de ausência (ou, para ser mais positivo: número de aulas assistidas)

Imagine então que traçamos esses valores para um aluno, o Arthur.



Um gráfico de aranha plotando os dados brutos representando cinco dimensões do desempenho de sala de aula de Arthur.

Aqui você pode ver como minha ferramenta preferida, o Excel, está desenhando essas linhas de grade como polígonos, em vez de círculos. O conjunto de dados para Arthur é plotado como uma única linha, conectando uma série de pontos. Às vezes, você verá gráficos de aranha onde a área contida por essa linha fechada é preenchida com uma cor semitransparente.



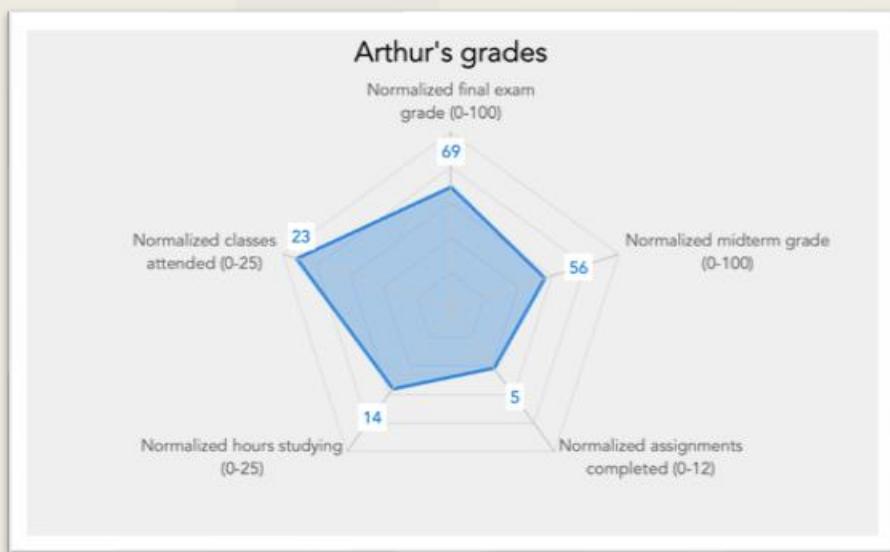
Um gráfico de aranha plotando os dados brutos representando cinco dimensões do desempenho de sala de aula de Arthur, com a área circundante preenchida com uma cor semitransparente.



Devido às limitações de nossa ferramenta, estamos usando o mesmo intervalo (0-100) para cada uma dessas cinco dimensões. No entanto, o domínio é diferente para algumas dessas medições. Por exemplo, havia apenas 25 dias de aula e 12 tarefas no total, então não faz sentido esses eixos irem até 100.

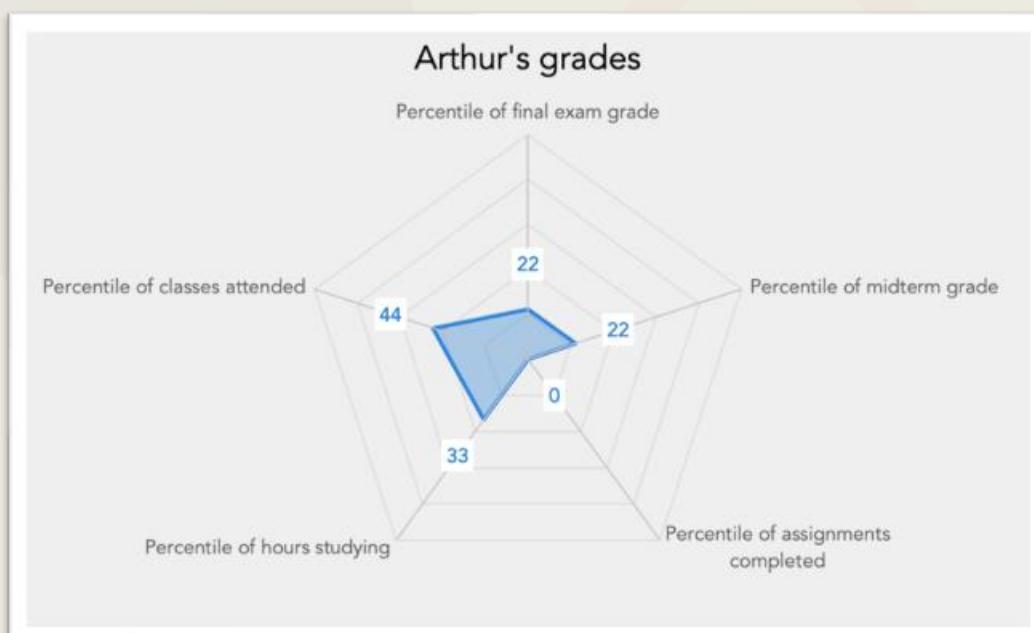
Uma maneira de contornar essa limitação é normalizar os dados. Para cada dimensão, sabemos qual valor é o valor mínimo possível (zero), e qual é o valor máximo possível... ou, se não houver limite para o valor máximo possível (tipo, quantas horas você estudou para isso teste?), então podemos usar o valor máximo observado dentro da classe como limite de nosso domínio.

Quando normalizamos os dados de Arthur - ainda usando os valores reais para nossos rótulos de dados - o gráfico de aranha agora se parece com isso:



Um gráfico de aranha sombreado traçando os dados normalizados representando cinco dimensões do desempenho de sala de aula de Arthur. Cada eixo agora é dimensionado de 0 no centro até o valor máximo possível para essa dimensão na borda.

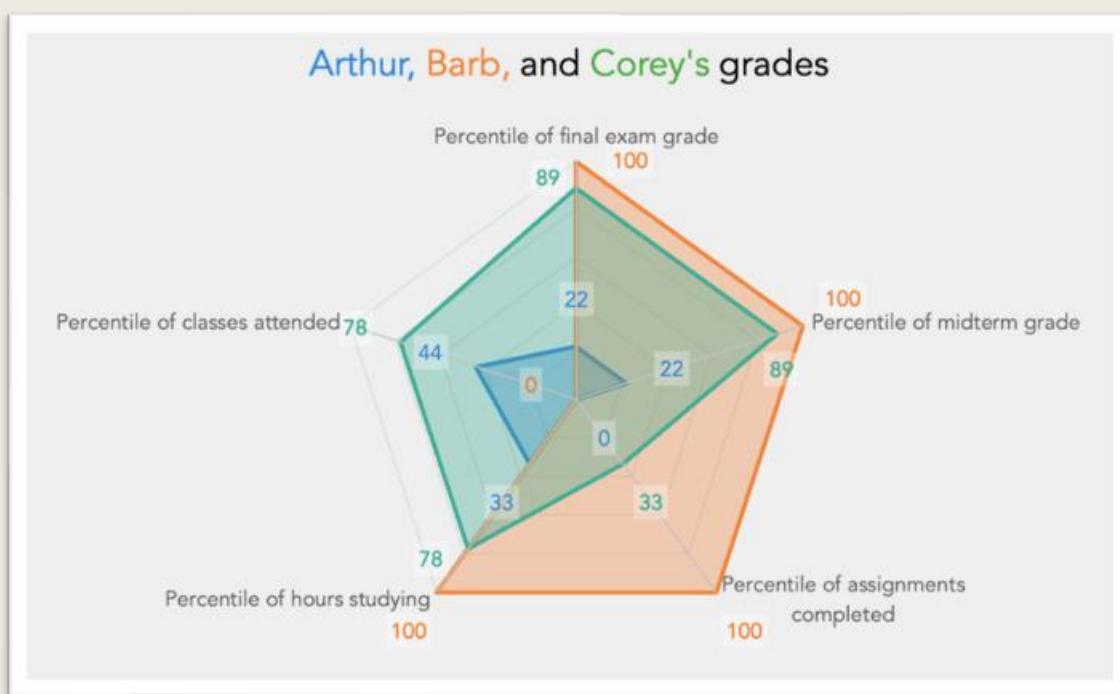
É possível que esta seja a visão que queremos usar. Também podemos dar um passo adiante e pensar sobre o desempenho de Arthur em comparação com seus colegas de classe; isto é, ao longo de cada uma dessas dimensões, em que percentual de resultados reais ele caiu? Ao calcular os percentis, podemos desenhar um gráfico de aranha onde o intervalo de valores de cada eixo vai de 0 a 1, ou do percentil mais baixo ao mais alto. Isso nos dará uma melhor visão comparativa dos dados.



Um gráfico de aranha sombreado traçando os dados normalizados representando cinco dimensões do desempenho de sala de aula de Arthur. Cada eixo agora mostra o percentil em que Arthur caiu, dentro das pontuações de seus colegas, ao longo de cada dimensão.



Quando aplicamos essa transformação aos nossos dados, o gráfico de Arthur ficou bem menor. Não estava claro antes, porque não tínhamos o contexto do desempenho do restante de seus colegas, mas Arthur parece ser um dos alunos com desempenho mais baixo. Vamos colocar dois de seus amigos, Barb e Corey, para comparação.



Um gráfico de aranha sombreado traçando os dados normalizados representando cinco dimensões do desempenho de três colegas. Cada eixo agora mostra o percentil em que Arthur, Barb e Corey caíram, dentro das pontuações de seus colegas, ao longo de cada dimensão.

Com este contexto, podemos fazer algumas avaliações visuais. Arthur não ia às aulas com tanta frequência quanto seus colegas, fazia o menor número de tarefas, mal estudava e estava perto do fundo do poço nas provas intermediárias e finais. Barb era uma superestrela: ela fez todas as tarefas, estudou muito e se saiu bem em ambos os testes. A única crítica contra ela como aluna é que ela faltava muito às aulas, embora isso não parecesse prejudicar seu desempenho. Corey, como Barb, foi bem-sucedido nos testes, mas aprendeu mais assistindo às aulas do que fazendo as tarefas.

## Coisas a ter em mente ao usar um gráfico de aranha

### Reduza o desafio de comparar coisas por área

No centro de tudo, quando representamos nossos dados em um gráfico de aranha, estamos pedindo ao público para julgar e avaliar esses dados comparando áreas. As pessoas são boas em comparar dados em uma dimensão por vez — os gráficos de barras são superfícies de ler —, mas são notoriamente ruins em comparar várias dimensões simultaneamente... mesmo que sejam formas simples como quadrados ou círculos. Com gráficos de aranha, estamos pedindo ao público para comparar as áreas relativas, ângulos e picos de polígonos irregulares (possivelmente enquanto eles estão sobrepostos uns sobre os outros).

Isso já parece desafiador, mas há ainda outra camada de complexidade incorporada: nossas métricas para cada dimensão escolhida aumentarão linearmente ao longo de nosso raio. Se sombreamos as áreas entre o raio e o centro de nosso gráfico, a área desse sombreamento aumentará geometricamente, não linearmente. Isso significa que acabaremos super-representando visualmente valores maiores versus valores menores.

## Em um ambiente ao vivo, apresente seu público a este gráfico gradualmente

Será útil, em um cenário ao vivo, se você puder narrar e animar a construção de seu gráfico de aranha, peça por peça. Dessa forma, esse tipo de gráfico denso e potencialmente intimidador poderia ser introduzido lenta e deliberadamente. Ao reservar um tempo para acompanhar seu público pela estrutura e significado do gráfico, você poderá levá-los a uma compreensão complexa e diferenciada de sua análise. Se não estiver em um ambiente ao vivo, você provavelmente terá que anotar seus gráficos liberalmente (ou iterar para um tipo de gráfico mais simples para garantir que os tópicos não sejam perdidos).

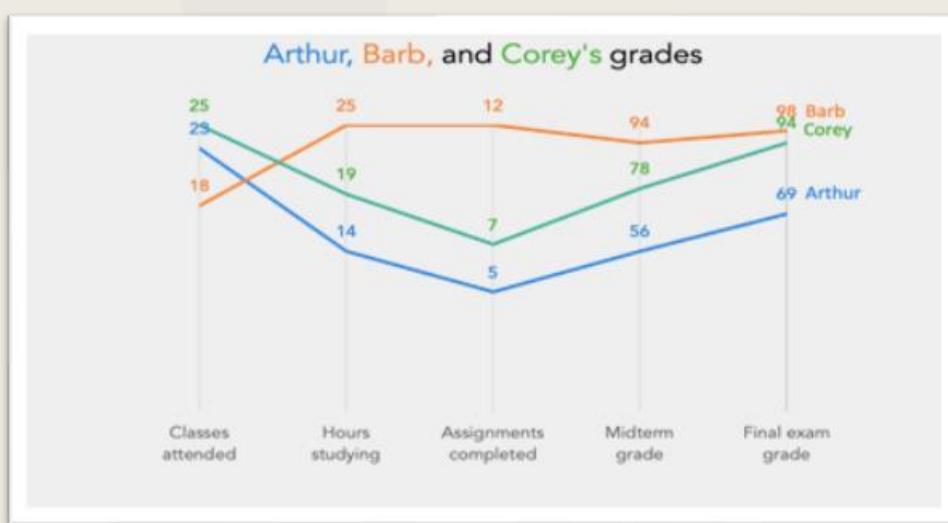
Ordene os eixos de forma cuidadosa e consistente



Como teoricamente você poderia colocar os itens em qualquer ordem, a área sombreada pode ter formas e tamanhos totalmente diferentes, dependendo de suas escolhas. Uma maneira de mitigar isso é agrupar métricas significativas o mais próximo possível e usar a mesma ordem consistentemente sempre que usar um gráfico de aranha para todos os conjuntos de dados estruturados de forma semelhante. Se seus dados se prestam a agrupamentos ou supercategorias, você pode usar cores para ajudar seu público a distingui-los visualmente.

### Considere alternativas semelhantes

Um gráfico de coordenadas paralelas é um pouco como um gráfico de aranha não agrupado. Em vez de cada eixo partir do mesmo ponto de origem e irradiar um do outro, colocamos esses eixos em paralelo, verticalmente. Então acabamos comparando as coisas no eixo Y, pois cada uma de nossas métricas é um pico ao longo do eixo X.



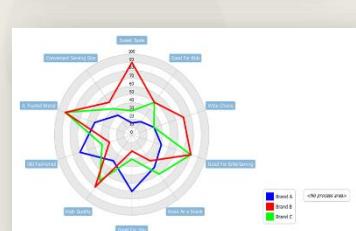
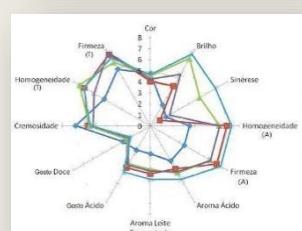
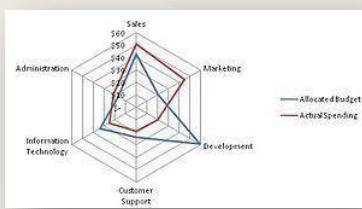
Este gráfico de coordenadas paralelas é baseado nos mesmos dados que usamos em nossos gráficos de aranha, mas ao “desenrolar” o gráfico em uma linha reta, ele pode facilitar a visualização das comparações entre os alunos.

Não é tão elegante quanto um gráfico de aranha visualmente, mas pode tornar as comparações mais fáceis de ver, pois se assemelha ao gráfico de linhas muito mais comum e, como tal, parecerá mais familiar para um público mais amplo. Uma desvantagem, no entanto, é que não é, de fato, um gráfico de linha comum (cada eixo vertical é uma escala diferente), e sua semelhança com um pode levar um visualizador casual a interpretar mal o visual totalmente.

### Dicas finais:

Os gráficos de aranha podem ser atraentemente eficientes, porque podem codificar visualmente muitas informações diferentes em uma pequena área. Tenha em mente, porém, que um público desconhecido pode precisar despender muito esforço para lê-los ou para entender uma lição importante. Quando um gráfico de aranha é essencial para comunicar algo específico, certifique-se de apoiar a compreensão do público por meio de narração ou animação passo a passo.

## Outras formas de visualização (Gráfico Aranha)





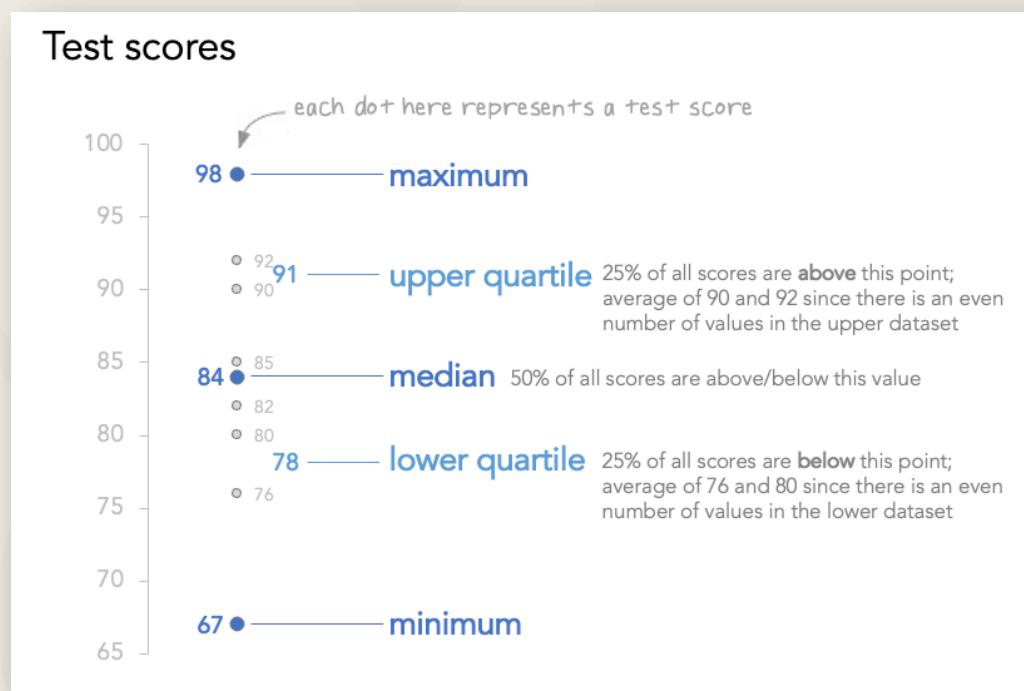
# O que é um gráfico de BOXPLOT?

## O que é um boxplot?

Um boxplot – às vezes chamado de box and whisker plot – é uma exibição densa que visa fornecer muito contexto para qualquer conjunto de dados. O gráfico representa visualmente várias métricas resumidas em uma única exibição:

- mínimo : o menor valor
- máximo: o maior valor
- mediana : o valor do meio quando os números são ordenados (ou a média dos dois valores do meio, se houver um número par de pontos de dados) - também conhecido como 50º percentil porque metade dos valores está acima (e abaixo) da mediana
- quartil inferior : a mediana da metade inferior (qualquer coisa menor que a mediana de todo o conjunto de dados) — também conhecido como 25º percentil
- quartil superior: a mediana da metade superior (algo mais que a mediana de todo o conjunto de dados) – também conhecido como 75º percentil

Vamos ilustrar essas medidas com um pequeno conjunto de dados de pontuações de teste com os seguintes valores: 67, 76, 80, 82, 84, 85, 90, 92 e 98.



## Como analisar um boxplot?

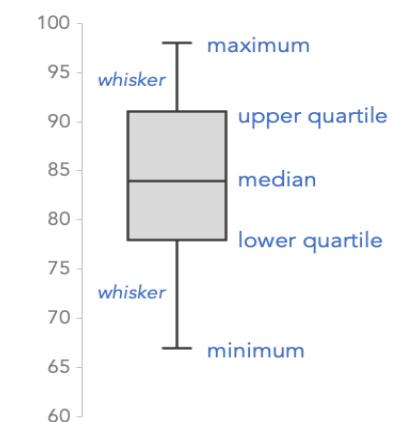
A parte “caixa” de um boxplot descreve os quartis inferior e superior. Dentro da caixa há uma linha que indica o valor mediano. Existem linhas que se estendem para fora da caixa - conhecidas como bigodes - para representar uma faixa de valores em um determinado conjunto de dados. Se houver outliers, os pontos individuais alinhados com os bigodes são plotados para denotar os valores extremos.



Abaixo está um boxplot para o conjunto de dados de pontuação do teste acima. Observe que, se houvesse outliers em nosso conjunto de dados, os bigodes representariam o "mínimo" e o "máximo", excluindo quaisquer valores extremos que seriam plotados como pontos. Nesse caso, o eixo não começa em zero porque estamos focando apenas na faixa de valores possíveis.

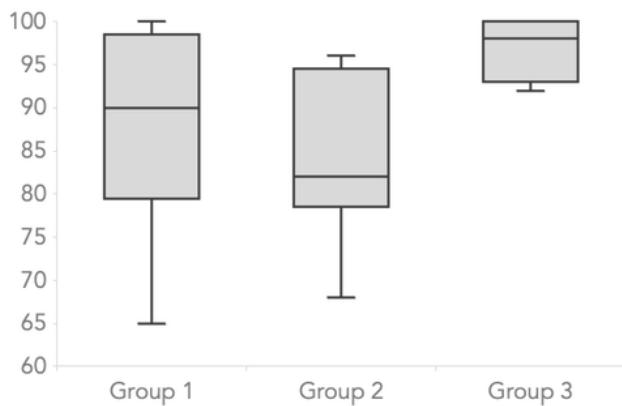
- Whisker = Bigode
- upper quartile = Quartil superior
- median = Mediana
- lower quartile = Quartil inferior

### Test scores

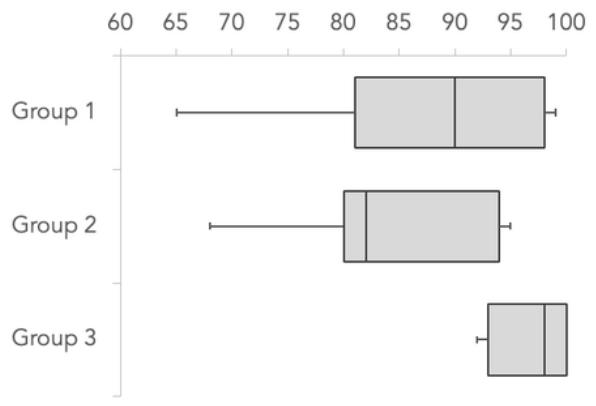


Assim como os gráficos de barras, os boxplots podem ser horizontais ou verticais. Um boxplot horizontal é uma boa opção para nomes longos de categorias, enquanto uma orientação vertical pode ser uma escolha melhor se ordenar as categorias da esquerda para a direita for mais lógico. Ser cuidadoso sobre a orientação do gráfico que você escolher pode melhorar a legibilidade e a compreensão dos dados.

Vertical boxplots



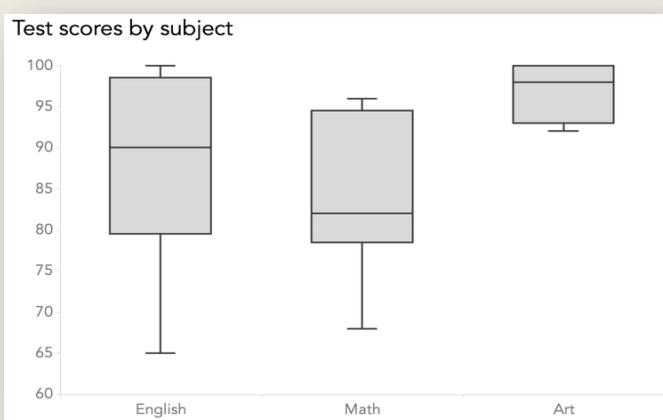
Horizontal boxplots





## Quando você **deve usar** um boxplot?

Os boxplots oferecem uma exibição condensada ao avaliar as distribuições de vários conjuntos de dados lado a lado. Por exemplo, o gráfico abaixo compara as pontuações dos testes entre as disciplinas.



Nesse cenário, podemos ver que a variação de pontuações é maior para o inglês em comparação com o art. Em geral, as pontuações de Arte são mais altas e mais agrupadas. A pontuação mediana em matemática é menor do que as outras disciplinas, e nenhuma das disciplinas tem valores discrepantes ou pontuações extremas nos testes.

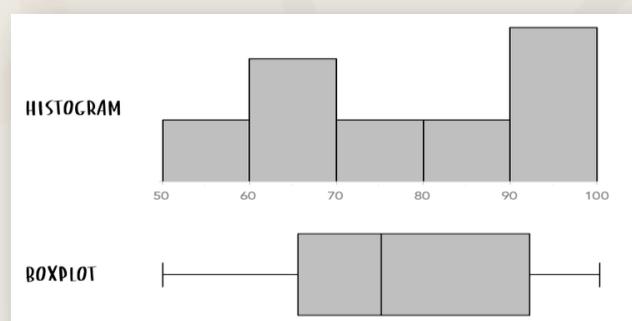
## Quando você **não deve usar** um boxplot?

Um boxplot pode ser um visual familiar se você avaliar as estatísticas do conjunto de dados regularmente, mas esse tipo de gráfico não é familiar para muitas pessoas e pode apresentar uma curva de aprendizado para elas. A escolha de uma exibição totalmente diferente pode evitar confusão para o público que não conhece esse gráfico ou como analisar as estatísticas que ele representa.

**Se você estiver apresentando um boxplot, considere o seguinte para facilitar a interpretação:**

- **Narre e construa o gráfico peça por peça.** Um tipo de gráfico complexo como esse pode ser opressor quando apresentado de uma só vez. Em vez disso, narre e anime a construção de um boxplot enquanto você apresenta, conduzindo seu público pela estrutura do gráfico e o significado de cada métrica.
- **Incluir palavras.** É sempre bom fornecer clareza e contexto por meio do uso cuidadoso de títulos e anotações de gráficos. Se você não estiver apresentando em um ambiente ao vivo, anotar liberalmente um boxplot (ou iterar para um tipo de gráfico mais simples) pode ajudar os visualizadores que, de outra forma, não entenderiam como interpretar o gráfico.

Outra coisa a ter em conta é que um **boxplot** não exibe a **distribuição** de dados com tanta precisão quanto um **histograma** ou outros gráficos de distribuição. Em certas situações, os boxplots podem obscurecer informações subjacentes importantes. Por exemplo, pode ser difícil dizer se seu conjunto de dados é multimodal – o que significa que os valores se agrupam em mais de um ponto – a partir de um boxplot. Na ilustração abaixo, a maioria dos valores está agrupada na faixa de 60 e 90, conforme mostrado no histograma, mas isso não é tão aparente no boxplot que representa o mesmo conjunto de dados.

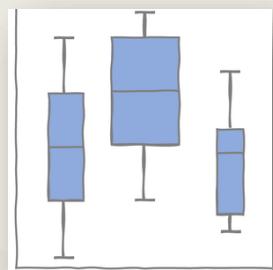




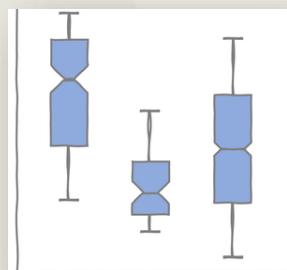
Nesse cenário, podemos ver que a variação de pontuações é maior para o inglês em comparação com o art. Em geral, as pontuações de Arte são mais altas e mais agrupadas. A pontuação mediana em matemática é menor do que as outras disciplinas, e nenhuma das disciplinas tem valores discrepantes ou pontuações extremas nos testes.

### Quais são as variações comuns de um boxplot?

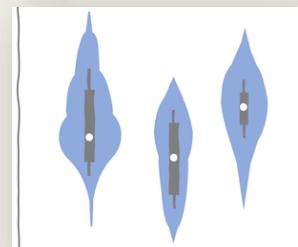
Em um **boxplot de largura variável**, a largura da caixa é ajustada para representar o tamanho dos dados dentro de cada grupo. Um conjunto de dados maior será mais largo do que um menor neste gráfico.



Um **boxplot com entalhes** se estreita no valor mediano. Os entalhes visualizam uma forma de intervalo de confiança em torno da mediana, o que ajuda a comparar diferentes conjuntos de dados. Se os entalhes das caixas de dois boxplots diferentes não se sobreponerem, haverá uma diferença estatisticamente significativa entre as medianas.



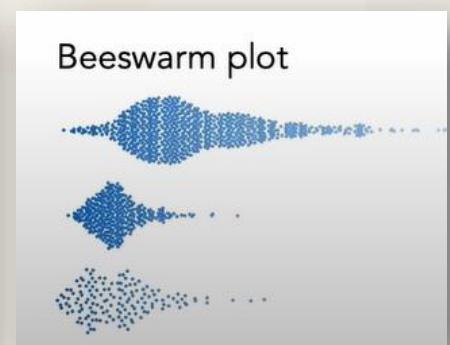
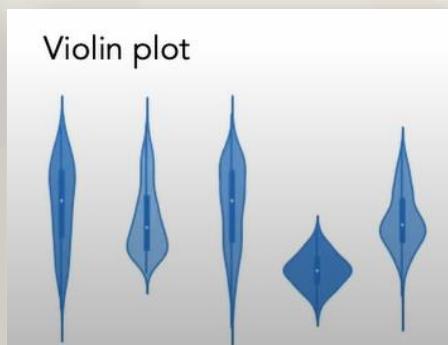
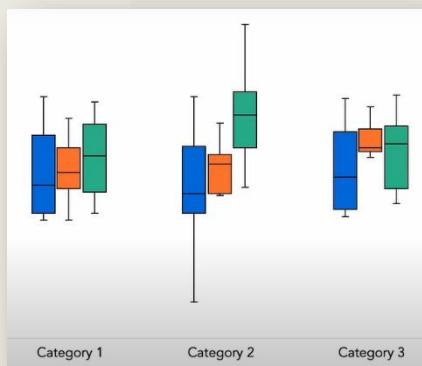
Um **gráfico de violino** combina um boxplot com uma curva de distribuição. A largura de cada curva corresponde à frequência aproximada de pontos de dados em cada grupo. Como os plots de violino exibem mais informações, eles são mais ocupados do que um boxplot tradicional.



Referência para mais vídeos de BOXPLOT

<https://www.youtube.com/watch?v=nbqfAEeAkHI>

## Outras formas de visualização (Gráfico Boxplot)





ESTA OBRA TERÁ CONTINUIDADE...