

GENE7033 – Tópicos Especiais em Genética I:

Visualização de dados para publicações científicas

Profª Drª Chirlei Glienke

Drª Desirrê Petters-Vandresen

Erros frequentes em visualização de dados e estratégias de resolução de problemas

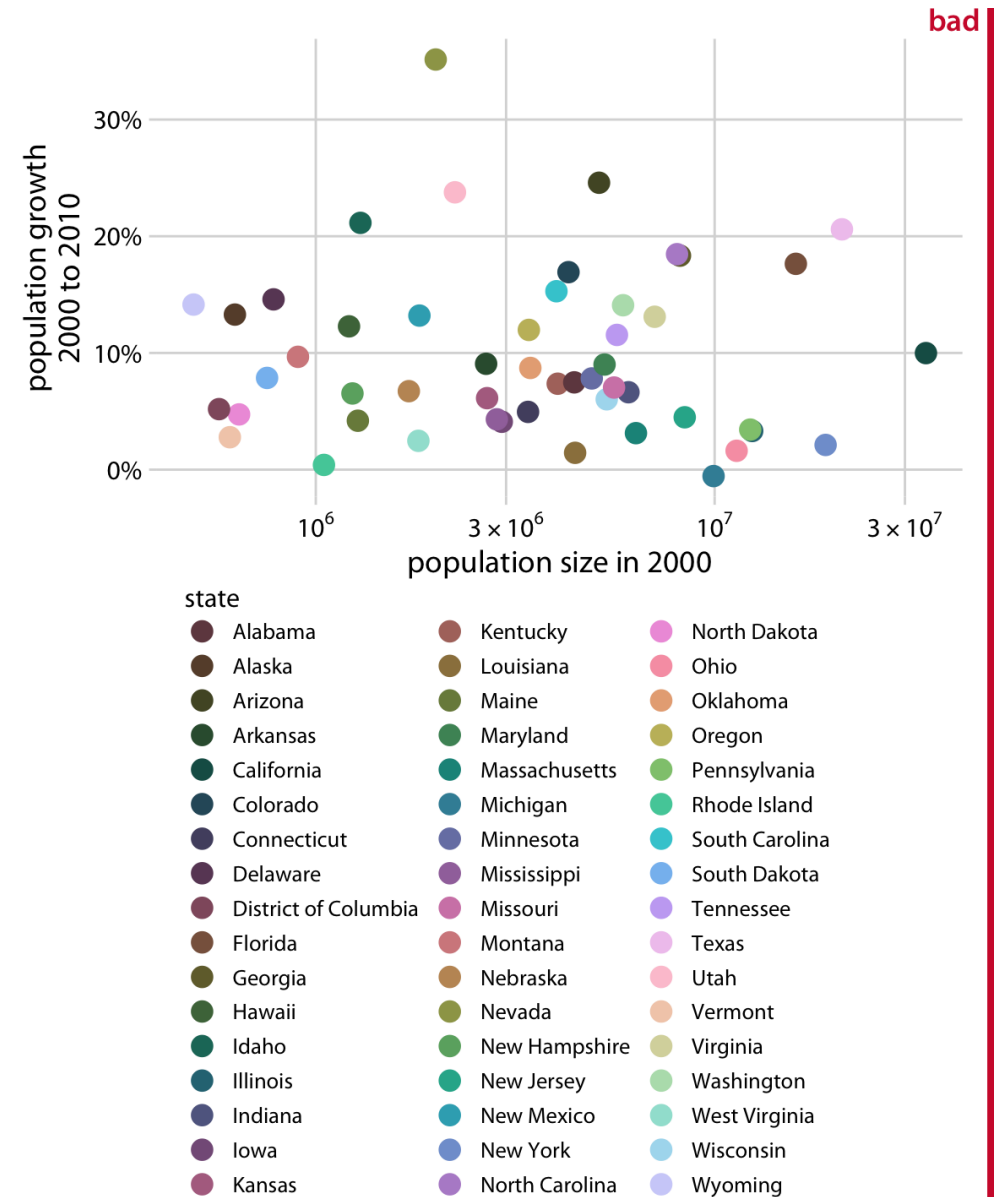
Dr^a Desirrê Petters-Vandresen

15/12/2022

Problemas no uso de cores

Excesso de informação

- Uso de cores não funciona bem quando há muitas informações diferentes a serem mapeadas
- Muitas cores similares, dificuldade de interpretação e identificação da informação

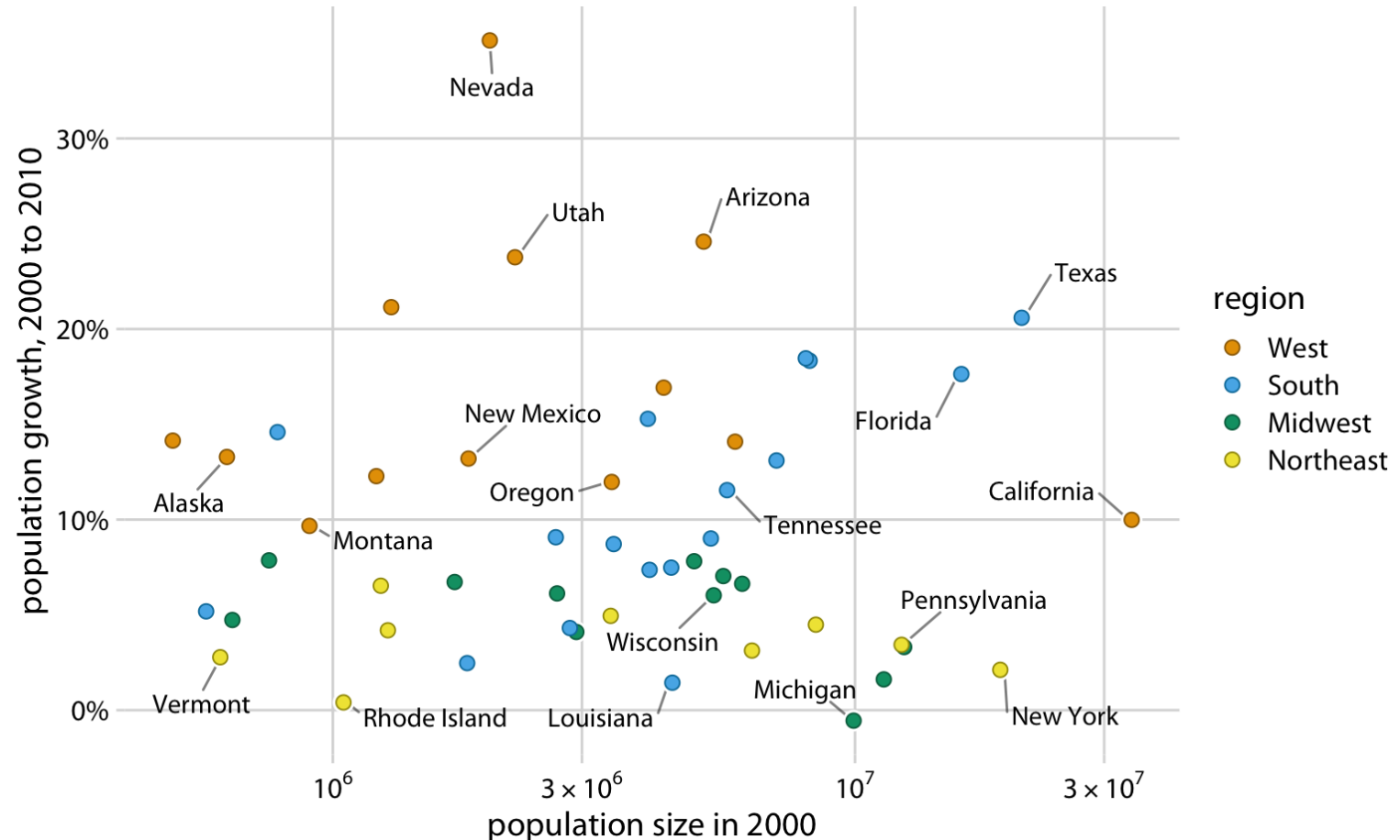


Adaptado de:

WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.

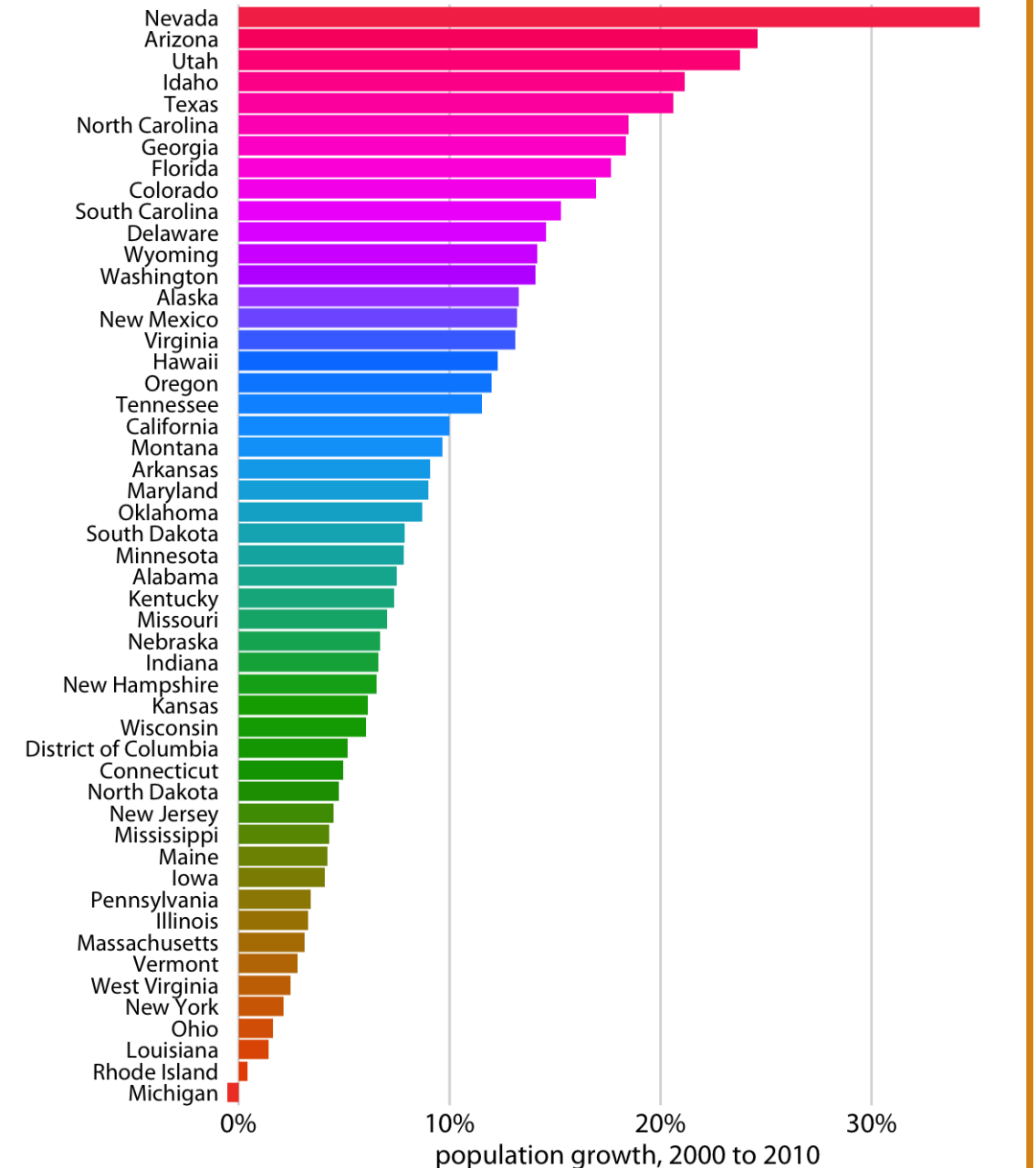
Excesso de informação

- Escalas qualitativas de cor funcionam melhor quando há entre três e cinco categorias diferentes
- Nesse exemplo, o mais adequado é:
 - Usar a cor para indicar a região geográfica
 - Adicionar legenda com o nome do estado nas situações em que essa informação for relevante
 - Se for essencial identificar todos os estados: utilizar uma tabela



Uso “gratuito” de cor

- Cores devem ser utilizadas com um propósito claro, e não para distrair o leitor
- Estados ordenados de acordo com o crescimento populacional, e cores utilizadas apenas para o efeito visual de “arco-íris”
- Esteticamente interessante, porém nada informativo
- Cores muito saturadas e intensas: dificuldade de visualização e interpretação da figura



Uso de escalas de cor inadequadas

- Premissa básica para escalas contínuas: indicação de valores maiores/menores, distância entre valores (uniformidade de percepção em toda a amplitude de cor)
- Diversas escalas populares desrespeitam essa premissa básica
- Arco-íris: passa por todas as cores do espectro visível, praticamente circular. Cores dos extremos são altamente similares, não geram a percepção de se referem a valores completamente opostos

rainbow scale



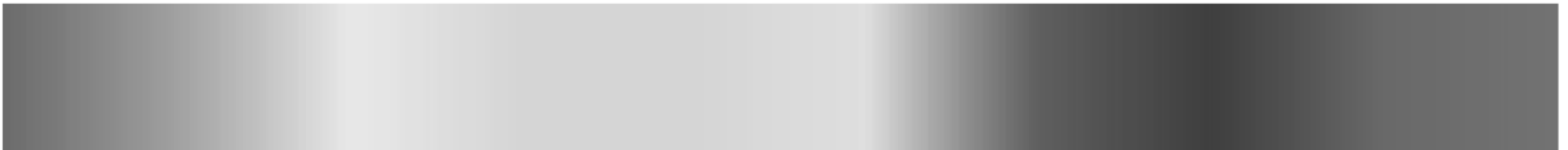
Uso de escalas de cor inadequadas

- Arco-íris: ausência de monotonicidade e uniformidade de percepção
- Mais perceptível ao converter para escala de cinza: regiões em que as cores mudam drasticamente muito rapidamente, e regiões em que quase não há mudança

rainbow scale

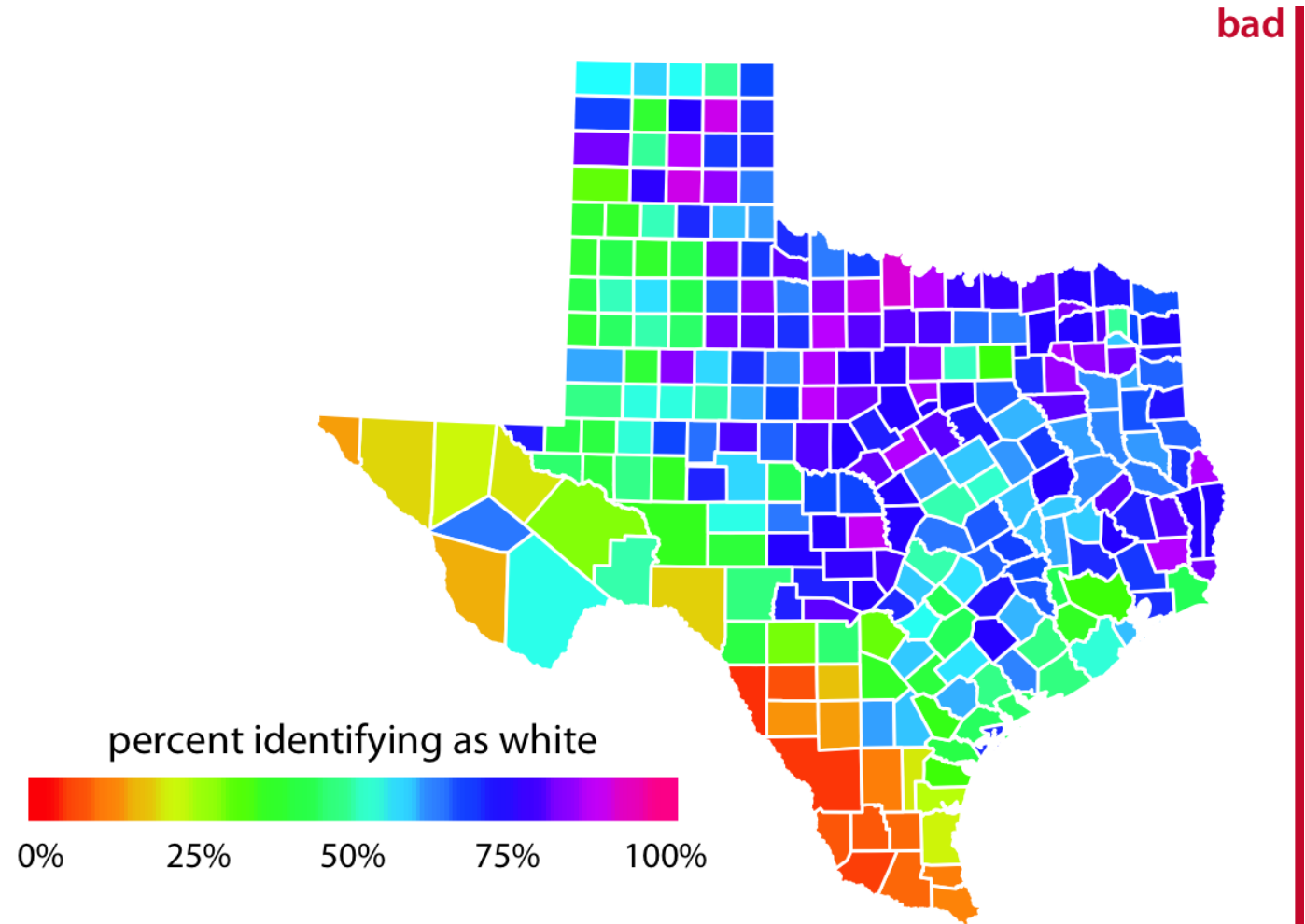


rainbow converted to grayscale



Uso de escalas de cor inadequadas

- Arco-íris: mascara percepções importantes e destaca características inexistentes ou irrelevantes no conjunto de dados
- Nesse exemplo, ênfase nos condados em que mais de 75% da população se identifica como branca
- Cores muito saturadas, visualmente desconfortável



Ignorar deficiências na visão de cores

- Diversas pessoas apresentam alguma forma de deficiência na visão de cores, e não distinguem cores que claramente são diferentes para a maior parte das pessoas
- Em geral, dificuldade de distinguir certos conjuntos de cores:
 - Deuteranomia/Deuteranopia: dificuldade de perceber/enxergar o verde, mais comum
 - Protanomia/Protanopia: dificuldade de perceber/enxergar o vermelho
 - Tritanomia/Tritanopia: dificuldade de perceber/enxergar o azul, mais rara

Ignorar deficiências na visão de cores

- Em geral, escalas sequenciais de cores não causam problemas de percepção
- Mesmo com deficiência na visão de cores, ainda há a percepção de um gradiente contínuo de cores mais escuras até cores mais claras

original



deuteranomaly



protanomaly



tritanomaly



Ignorar deficiências na visão de cores

- Em escalas divergentes, alguns contrastes popularmente utilizados são indistinguíveis

- Contraste entre verde e vermelho

original



deuteranomaly



protanomaly



tritanomaly



Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

Ignorar deficiências na visão de cores

- Em escalas divergentes, alguns contrastes popularmente utilizados são indistinguíveis

- Contraste entre verde e azul

original



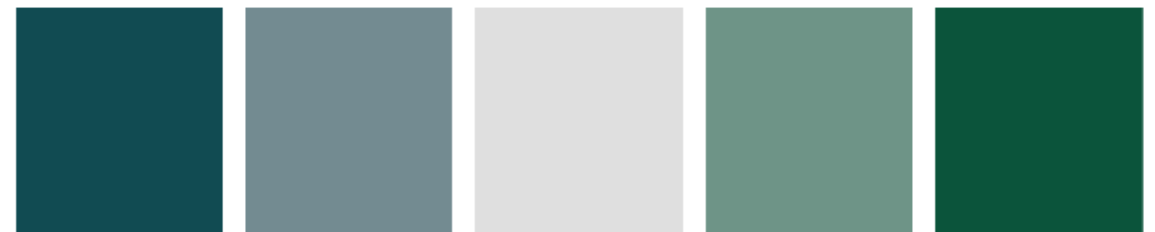
deuteranomaly



protanomaly



tritanomaly



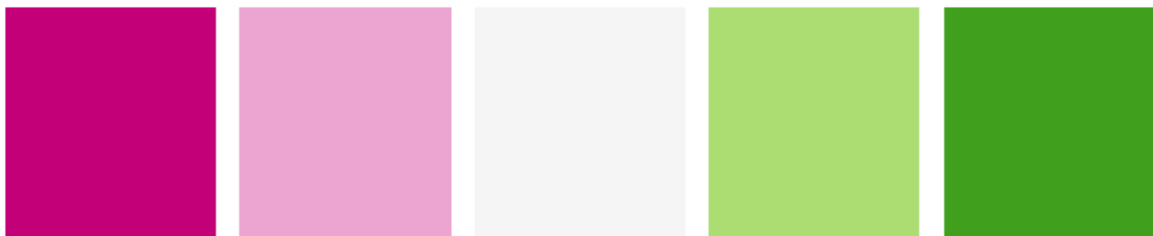
Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

Ignorar deficiências na visão de cores

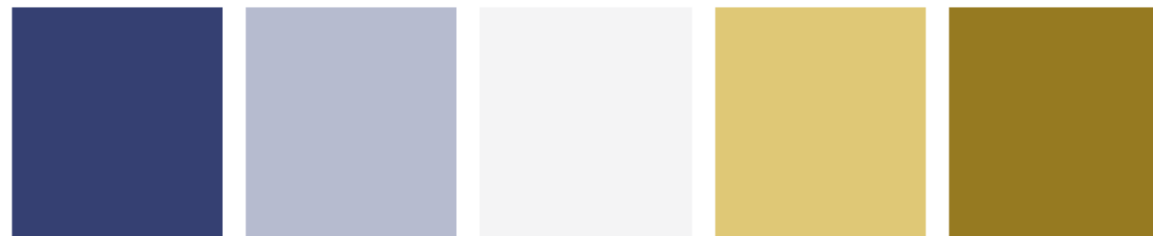
- Há contrastes levemente modificados que funcionam para pessoas com deficiências na visão de cores

- Contraste entre rosa e verde claro

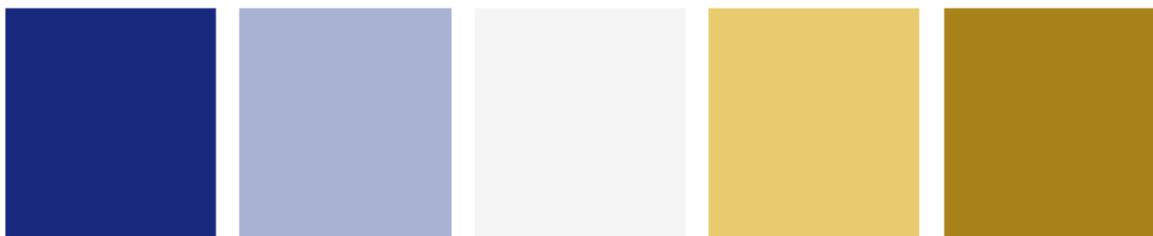
original



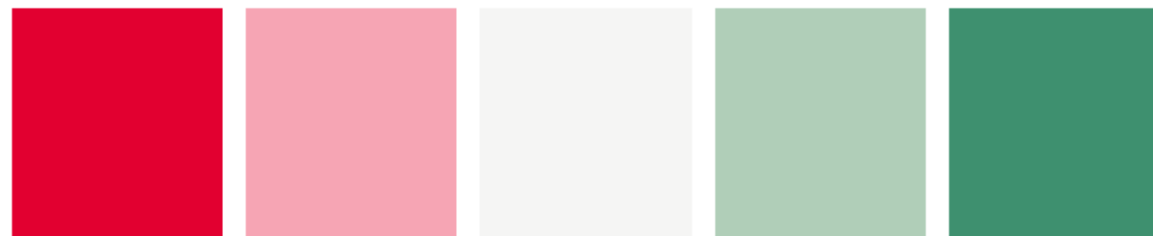
deuteranomaly



protanomaly



tritanomaly



Adaptado de:
WILKE, C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media, 2019.

Ignorar deficiências na visão de cores

- Para escalas quantitativas pode ser mais complicado encontrar cores adequadas, visto que muitas cores diferentes são necessárias
- Soluções:
 - Evitar excesso de informações
 - Testar as escalas de cor em softwares que simulam deficiências na visão de cores
- Paleta qualitativa (Okabe & Ito, 2008)



Ignorar deficiências na visão de cores

- Softwares que simulam deficiências na visão de cores:
- <https://www.color-blindness.com/coblis-color-blindness-simulator/>
- <https://pilestone.com/pages/color-blindness-simulator-1>
- <https://daltonlens.org/colorblindness-simulator>

Ignorar deficiências na visão de cores

- O tamanho dos elementos também influencia na distinção de cores
- Mais fácil distinguir cores em elementos e áreas maiores ao invés de elementos pequenos e linhas finas
- Recomendação: além de testar a escala de cores, testar também a figura final para perceber se é possível distinguir todos os elementos com base na cor

Adaptado de:
WILKE, C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media, 2019.

original



deuteranomaly



protanomaly



tritanomaly



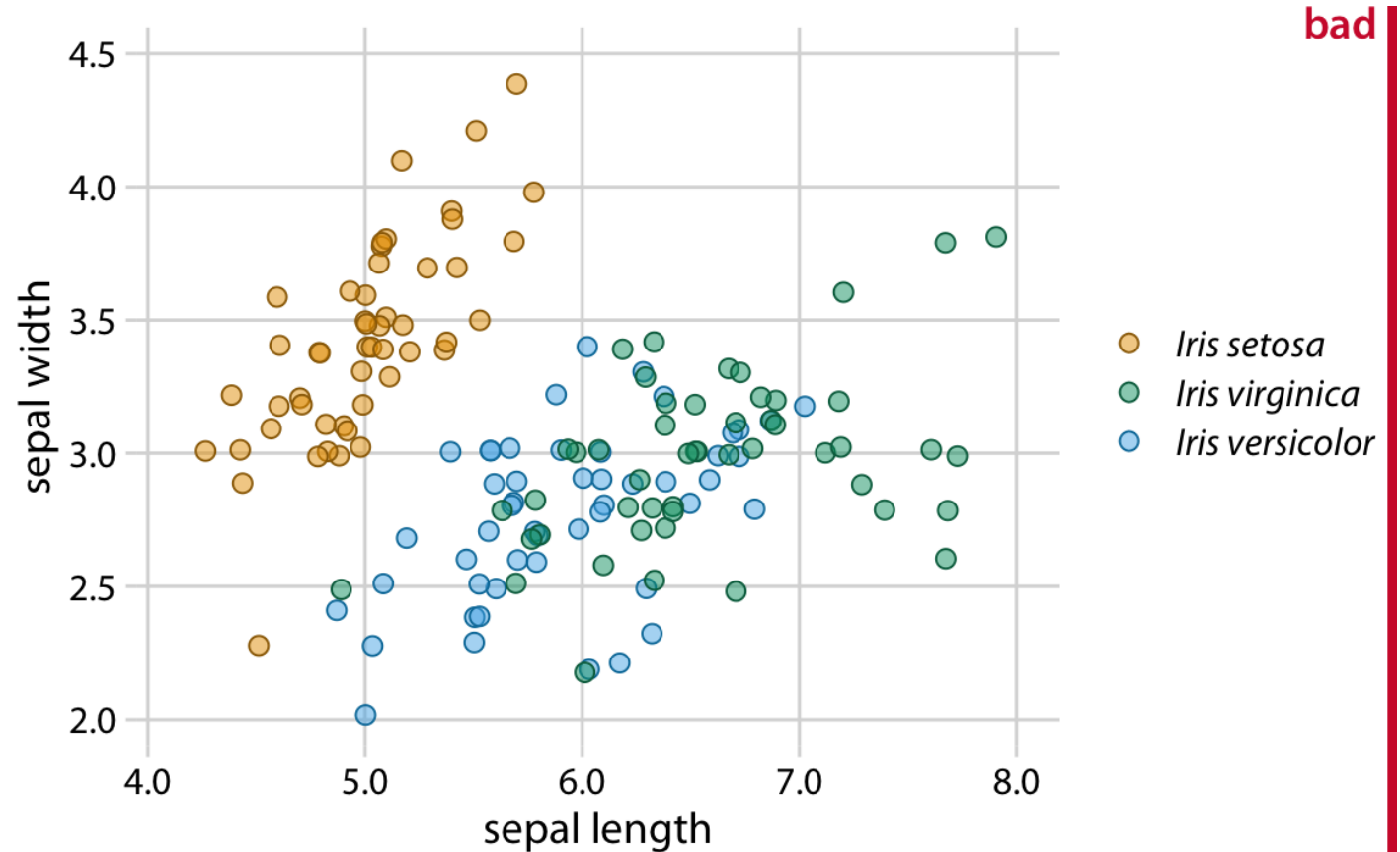
**Estratégia: codificação
redundante**

Finalidade

- Nem sempre somente o uso de cor é suficiente para transmitir a mensagem, seja por excesso de informação, tamanho reduzido dos elementos representados, ou por deficiências na visão de cores
- Utilizar a cor para reforçar a mensagem, mas não depender somente da cor para transmitir a mensagem
- Codificar a informação de forma redundante, utilizando mais de um tipo de canal visual e elementos estéticos

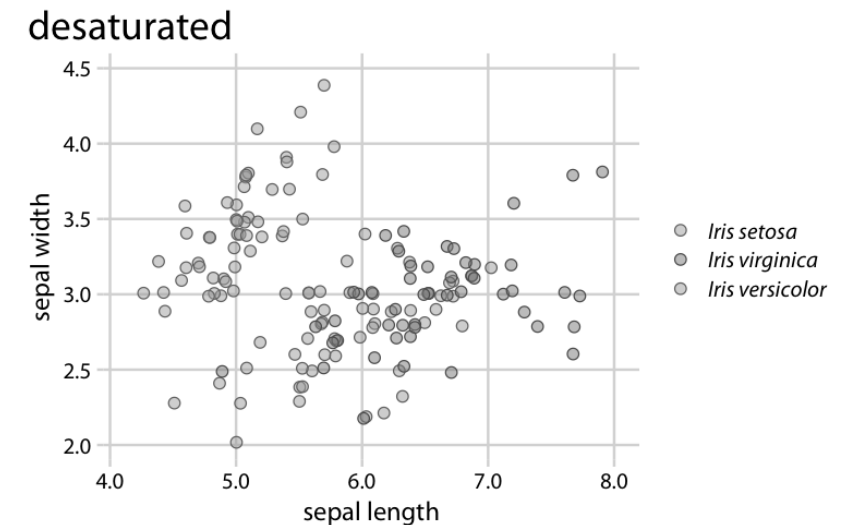
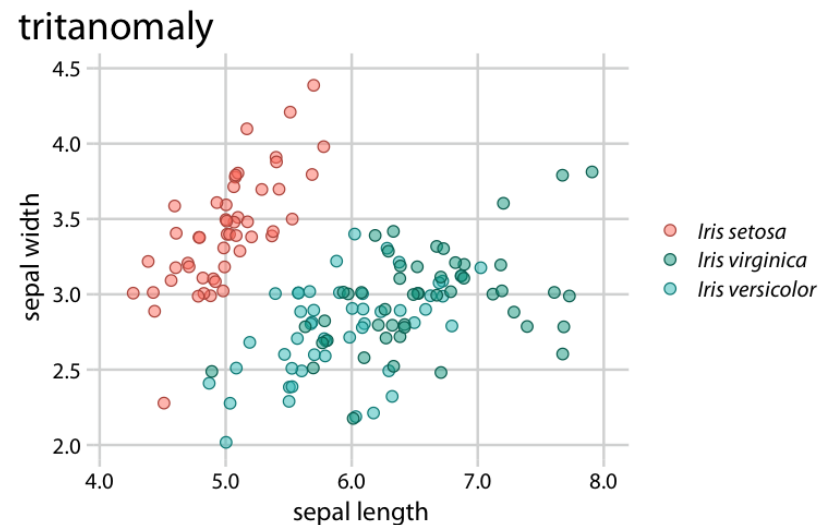
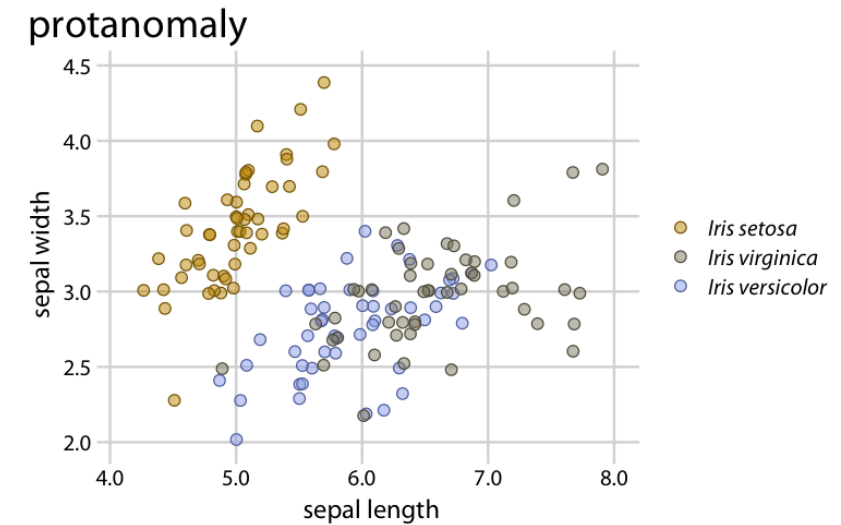
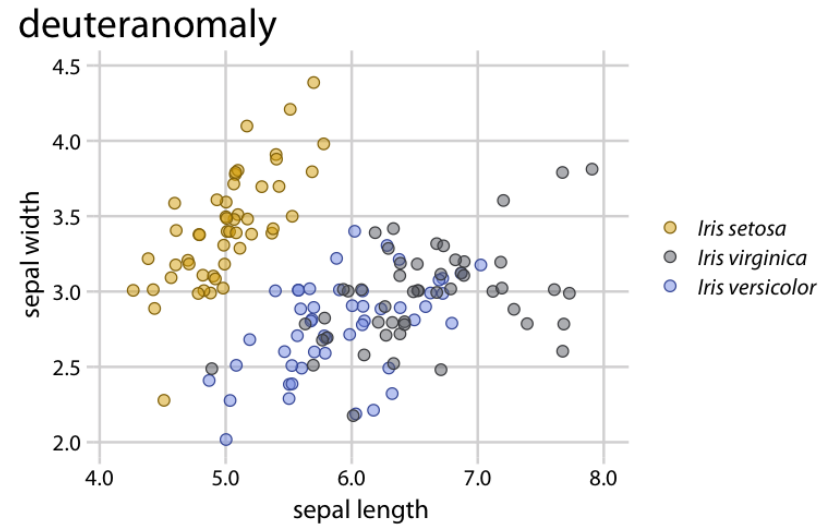
Criando legendas com codificação redundante

- Gráficos de dispersão: pontos que representam diferentes grupos comumente diferem somente na cor
- Dificuldade de separar *I. virginica* de *I. versicolor*: pontos se sobrepõem e cores são parecidas até para quem não tem deficiência na visão de cores



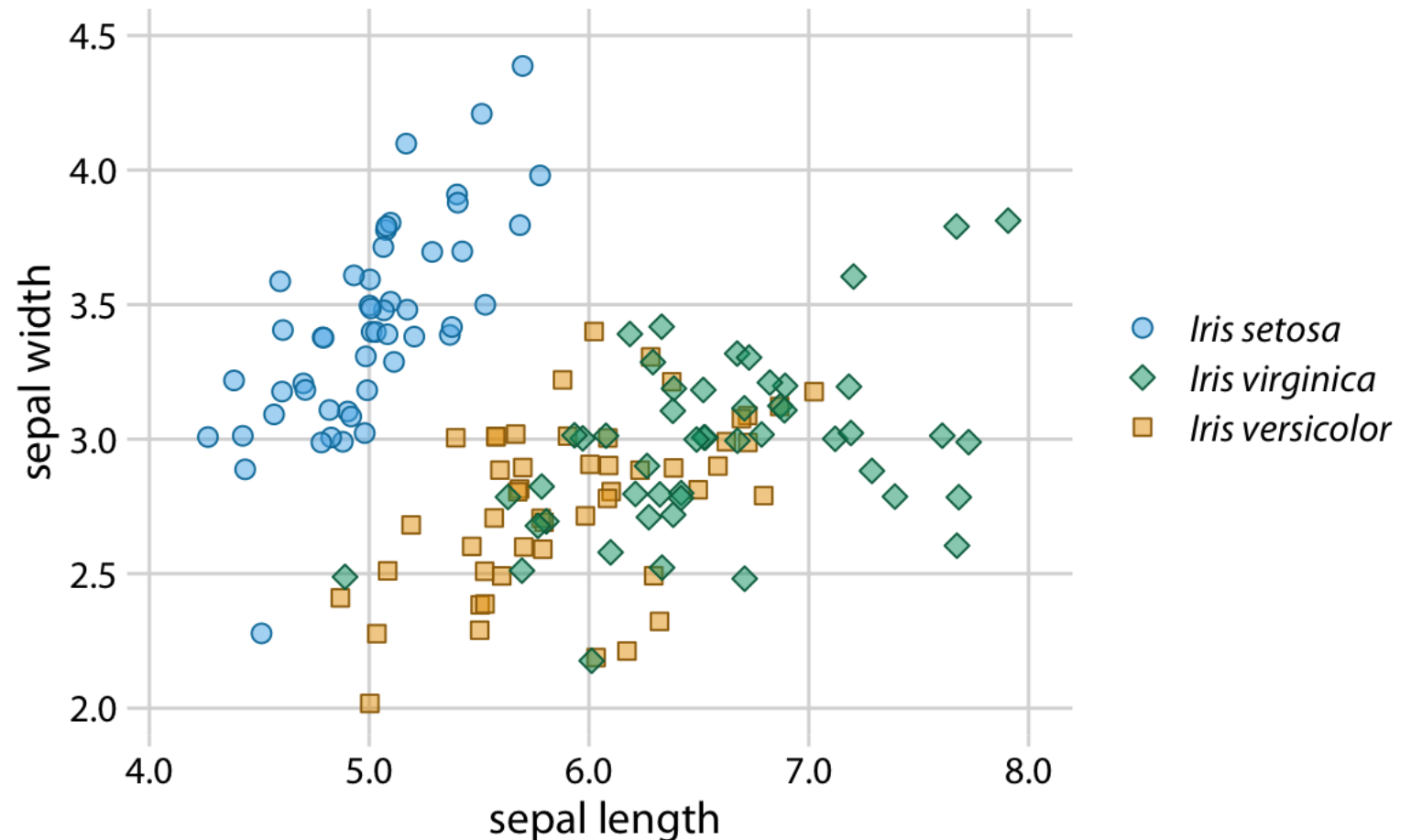
Criando legendas com codificação redundante

- Dificuldade de distinguir os pontos de cada espécie para pessoas com tritanomalia ou na imagem com cores sem saturação



Criando legendas com codificação redundante

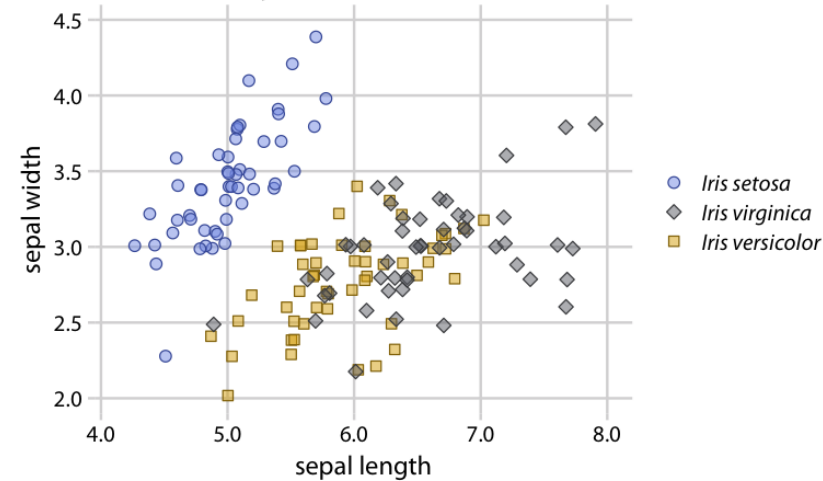
- Possíveis ajustes:
 - Mudar a ordem das cores utilizadas, para que não haja sobreposição de verde e azul
 - Utilizar diferentes formas para os pontos, para que pareçam diferentes até para quem tem deficiência na visão de cores ou na imagem em escala de cinza



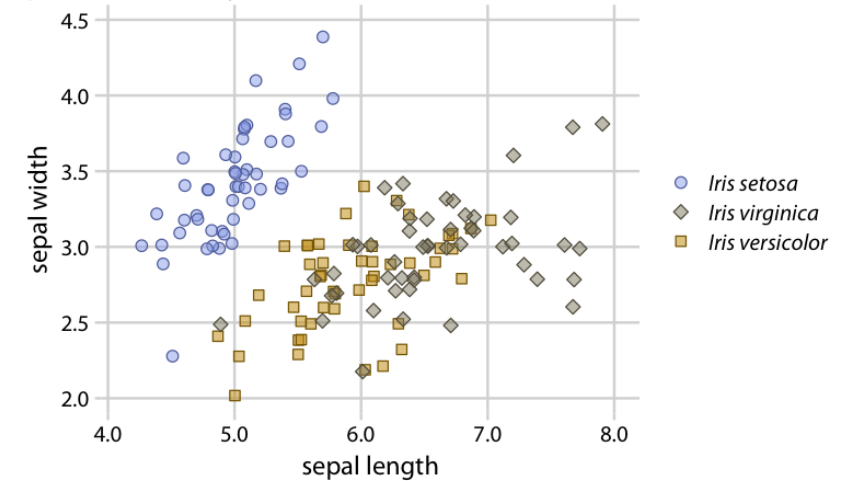
Criando legendas com codificação redundante

- Utilizar diferentes formas para os pontos, para que pareçam diferentes até para quem tem deficiência na visão de cores ou na imagem em escala de cinza

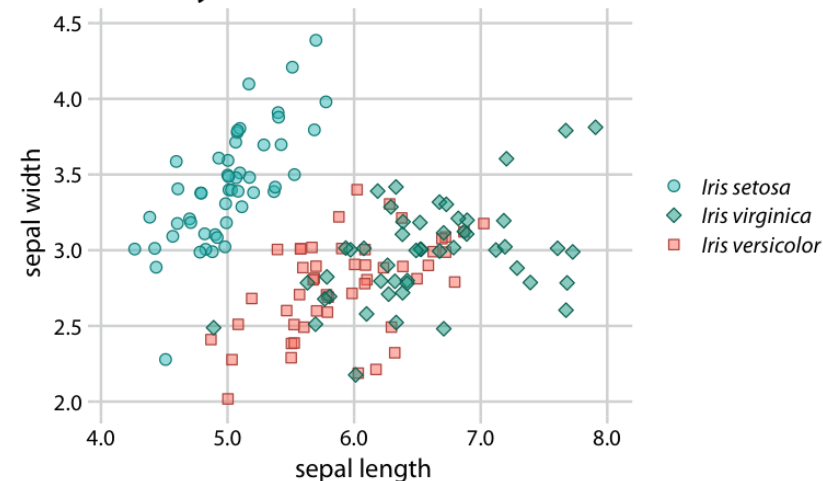
deuteranomaly



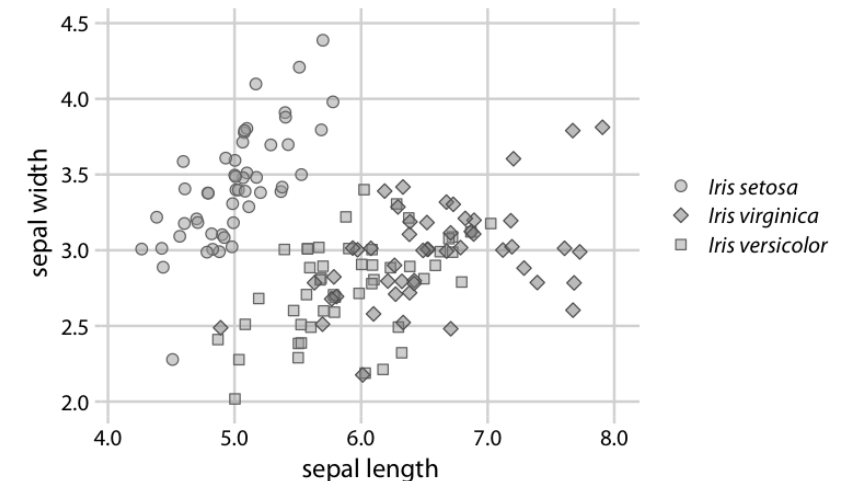
protanomaly



tritanomaly

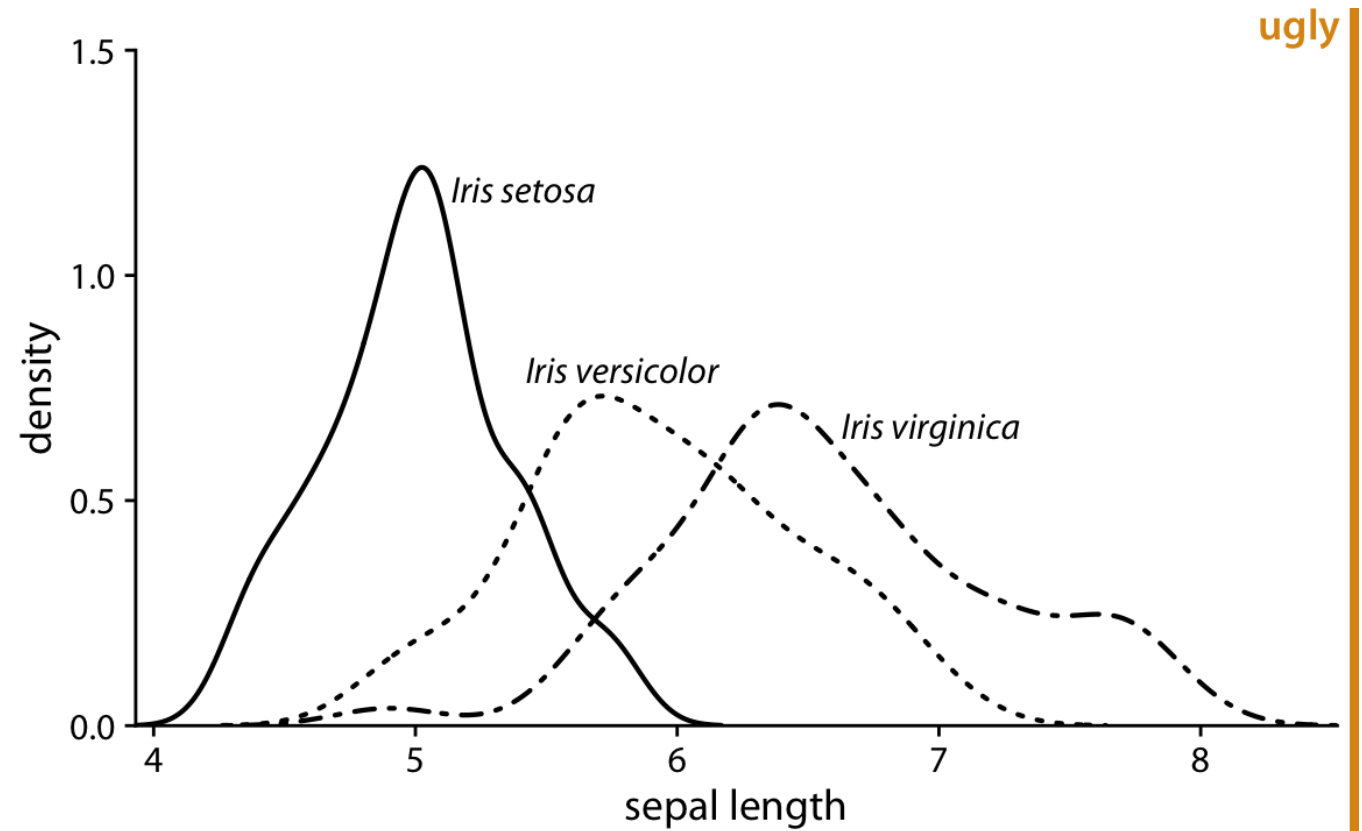


desaturated



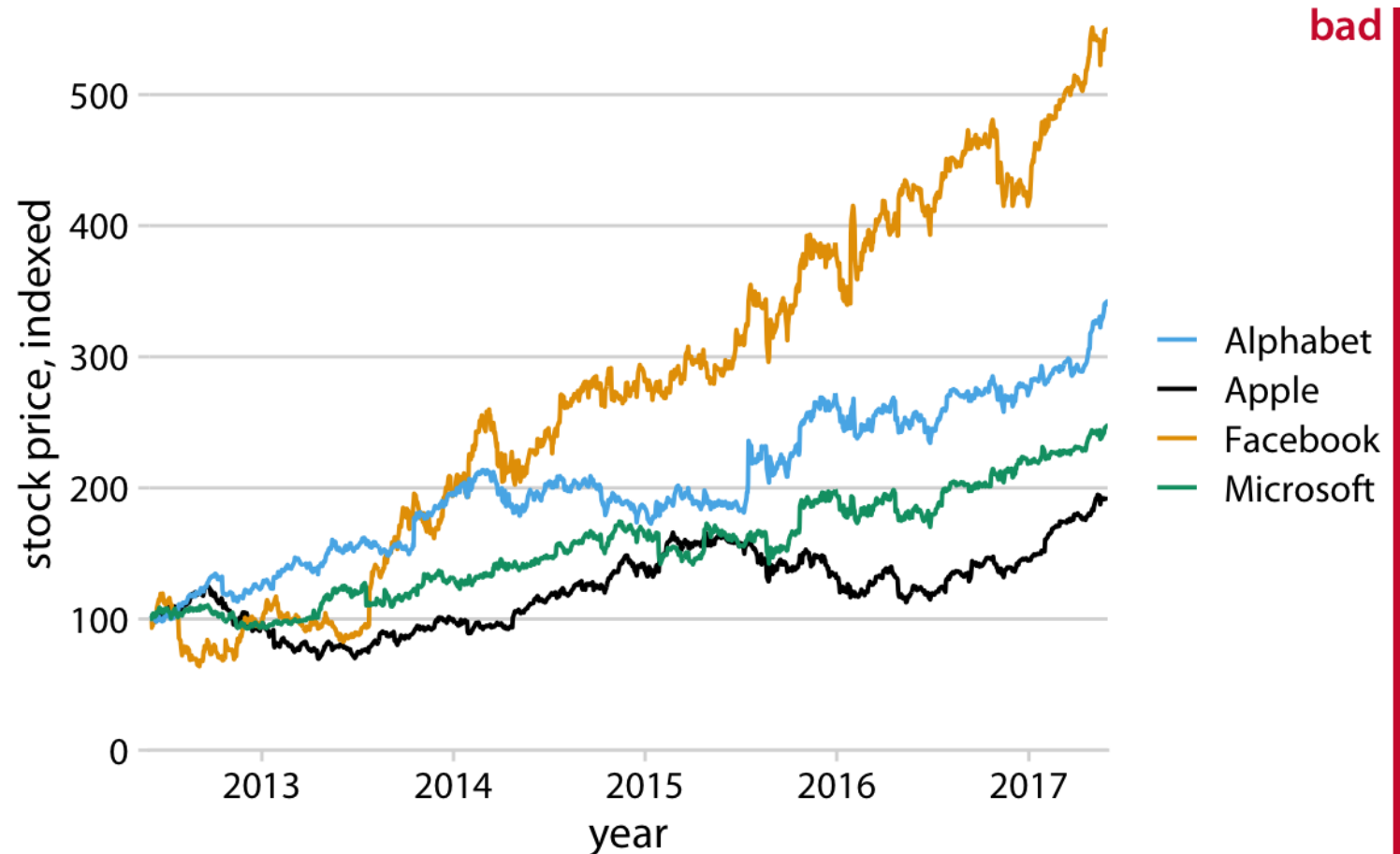
Criando legendas com codificação redundante

- Mudar o formato dos “pontos” nem sempre é viável
- Em gráficos de linhas, mudar o tipo de traçado (ex: contínuo, pontilhado) geralmente não ajuda, não é esteticamente agradável e cria ruído visual
- Esforço visual aumentar ao obrigar o leitor a comparar o tipo de traçado à legenda
- Como proceder nesses casos?



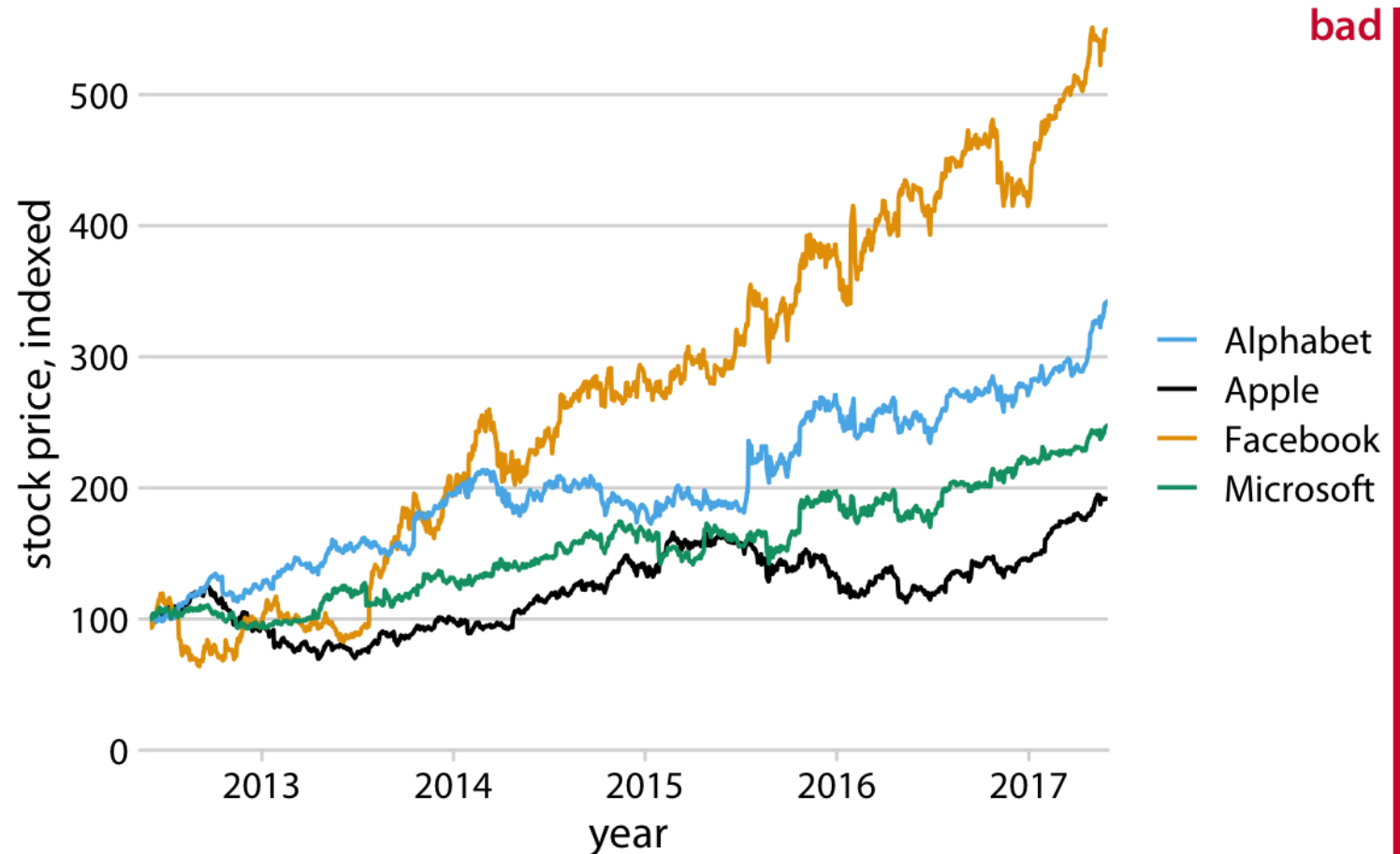
Criando legendas com codificação redundante

- Uso de cores distinguíveis mesmo quando há deficiência na visão de cores
- Problema: legenda em ordem alfabética, e não de acordo com a ordem das linhas no gráfico



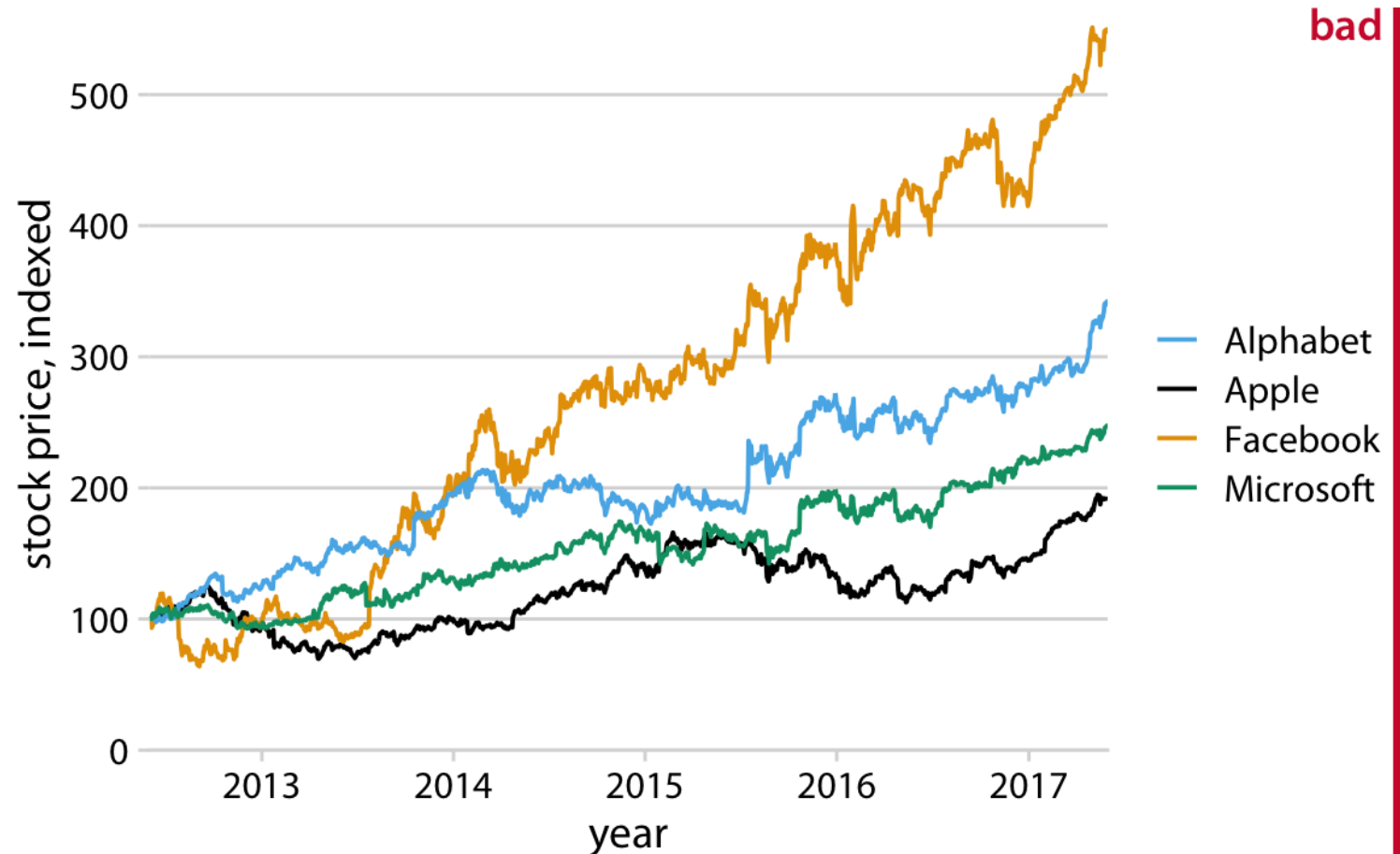
Criando legendas com codificação redundante

- Esforço mental aumentado para comparar as linhas com a legenda
- Lembrete: ao gerar as figuras e legendas automaticamente, os softwares não tem a percepção visual que temos ao olhar a figura



Criando legendas com codificação redundante

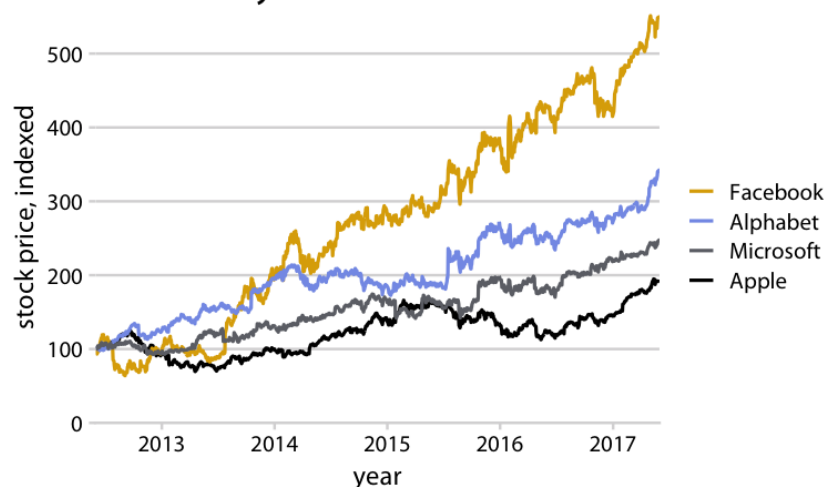
- Uso de ordem convencional e padronizada (alfabética) tende a ser o padrão, mas nem sempre é a visualmente mais adequada
- Sempre reordenar a legenda para corresponder à ordem observada no gráfico, quando ela existe



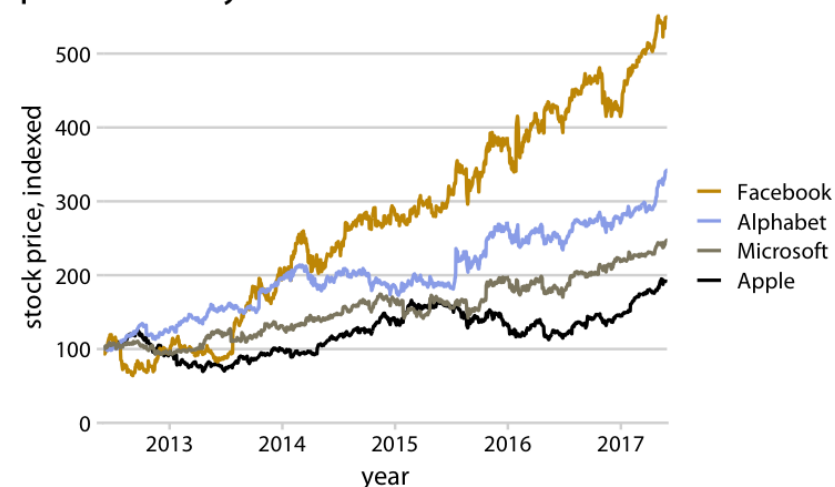
Criando legendas com codificação redundante

- Repetir a ordem na legenda auxilia principalmente nos casos de deficiência na visão de cores e em figuras em escala de cinza

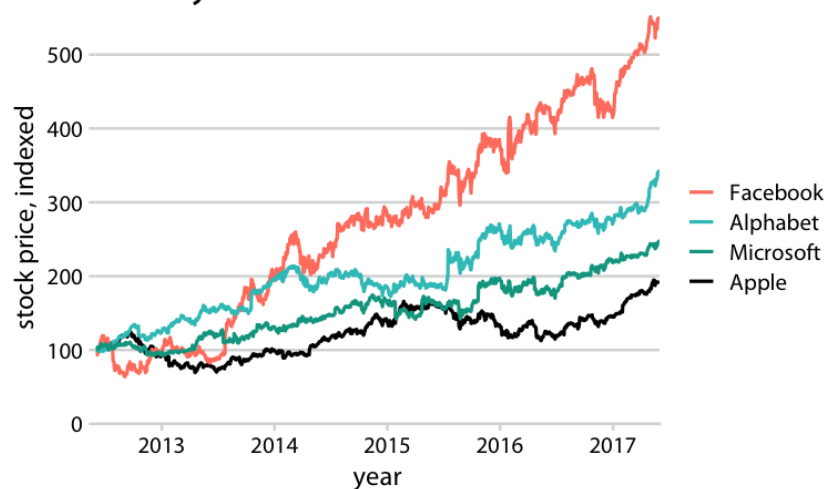
deuteranomaly



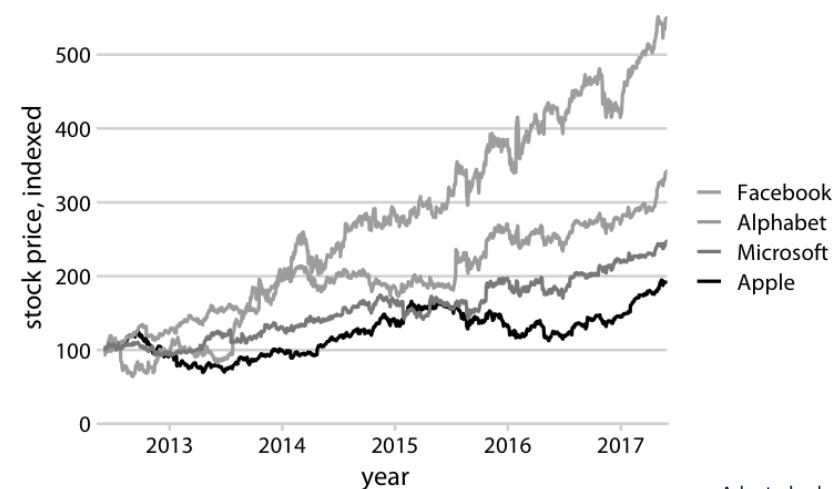
protanomaly



tritanomaly



desaturated

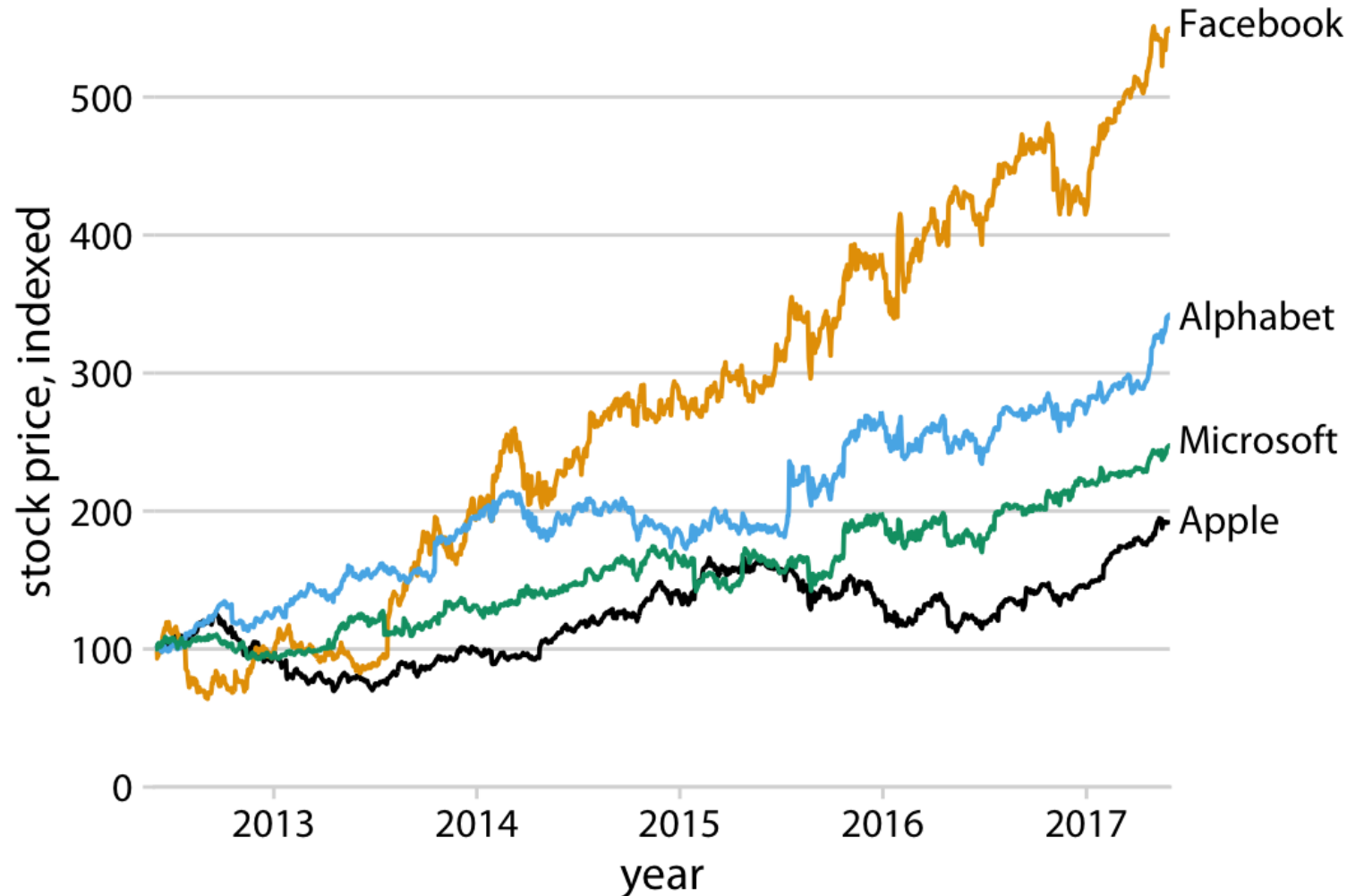


Criando figuras sem legenda

- Mesmo quando há codificação redundante, legendas sempre exigem maior esforço mental do leitor
- Uma alternativa interessante é, sempre que possível, eliminar a legenda totalmente
- Eliminar a legenda NÃO significa simplesmente escrever “*Os pontos amarelos representam X*” no texto
- Eliminar a legenda significa que o design do gráfico é feito de forma que se torne óbvio o significado dos elementos visuais, mesmo sem uma legenda explícita

Criando figuras sem legenda

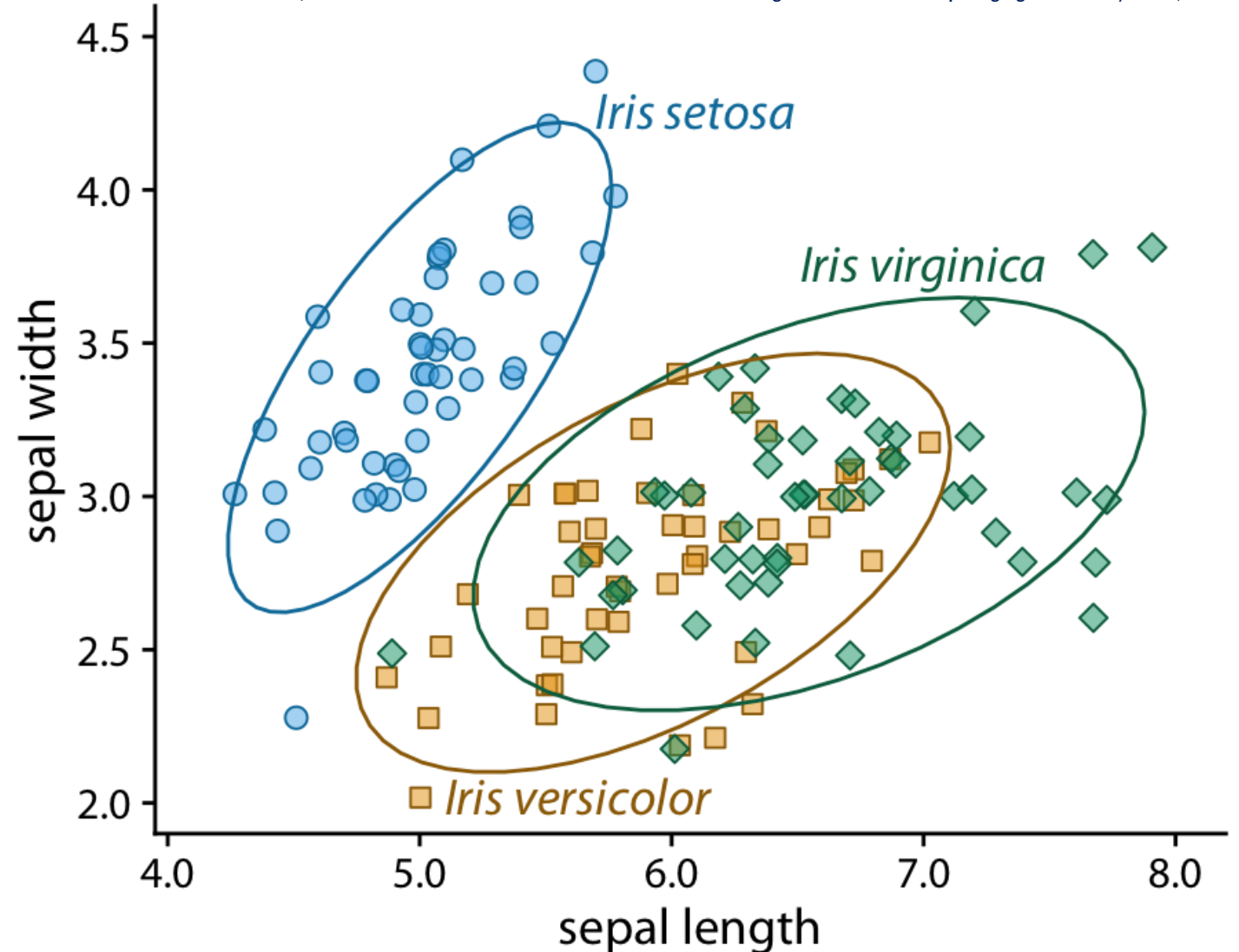
- Identificação direta: marcações em texto ou outros elementos visuais servem de guias ao resto da figura
- Em um gráfico de linhas, colocar o nome correspondente à cada linha logo ao lado



Criando figuras sem legenda

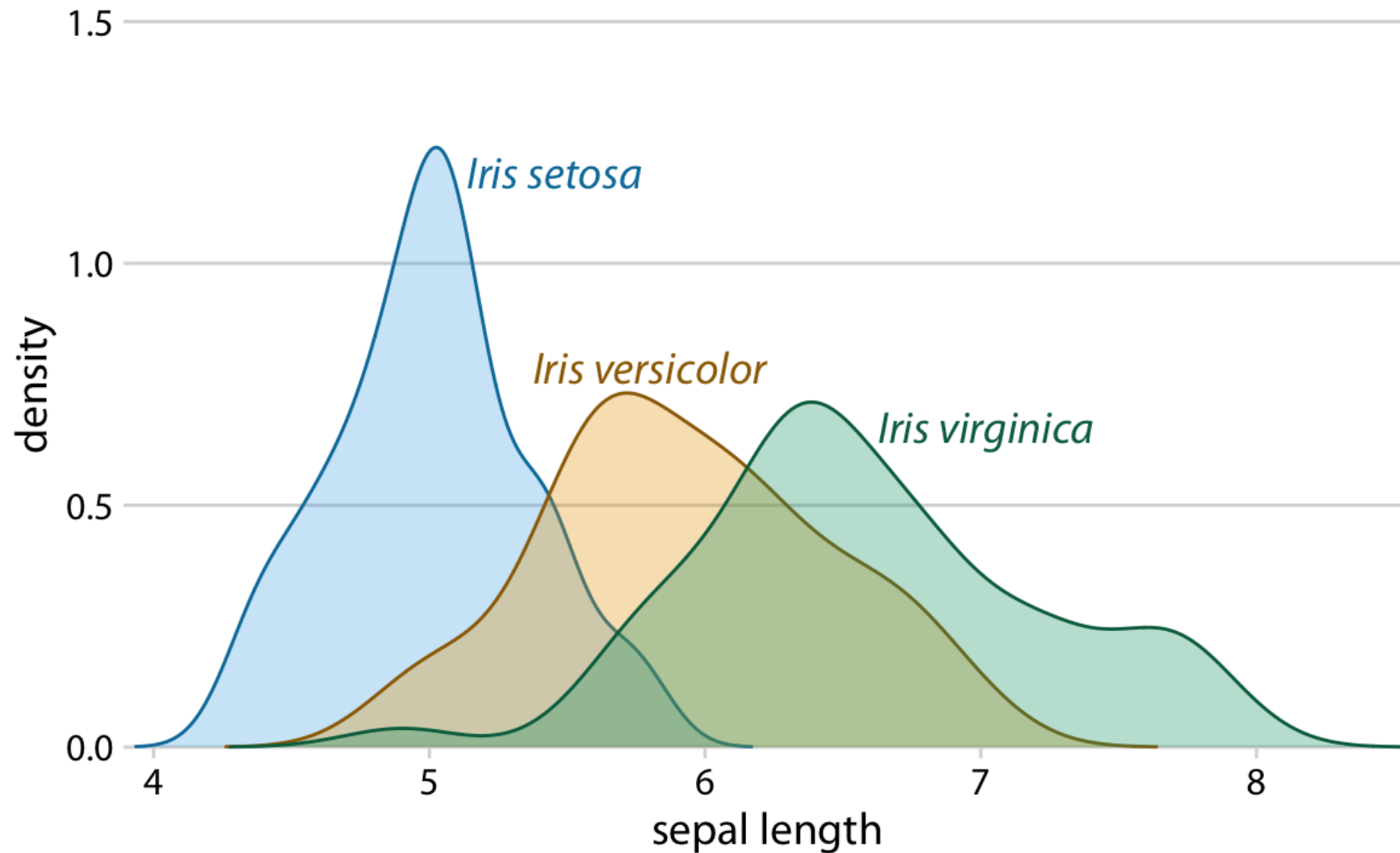
- Identificação direta: marcações em texto ou outros elementos visuais servem de guias ao resto da figura
- Em um gráfico de dispersão, utilizar formas como elipses para delimitar os grupos existentes dentro do conjunto de dados

Adaptado de:
WILKE, C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media, 2019.



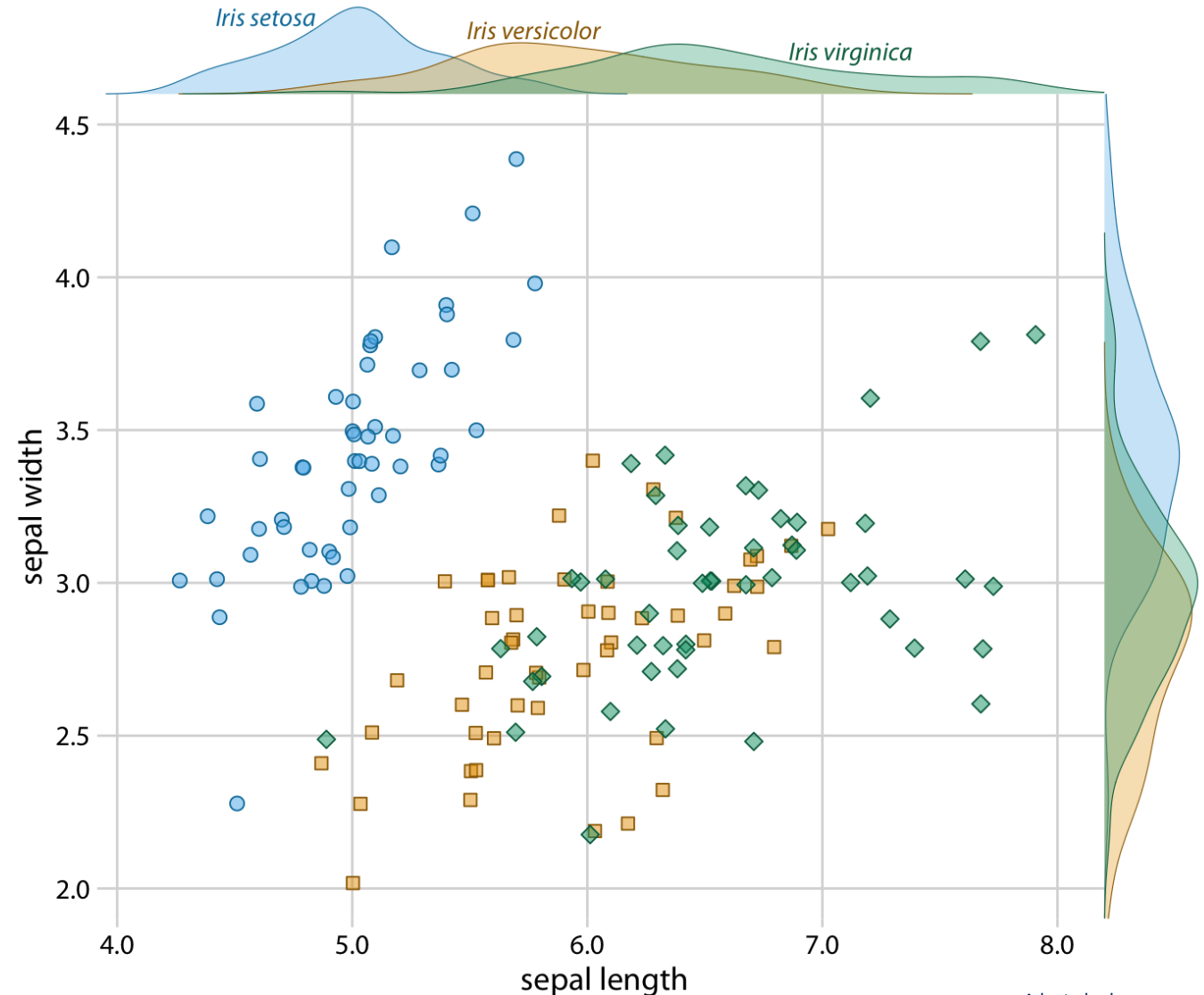
Criando figuras sem legenda

- Identificação direta: marcações em texto ou outros elementos visuais servem de guias ao resto da figura
- Em um gráfico de densidades, colorir o texto com a cor mais escura na mesma matiz da cor da área da curva



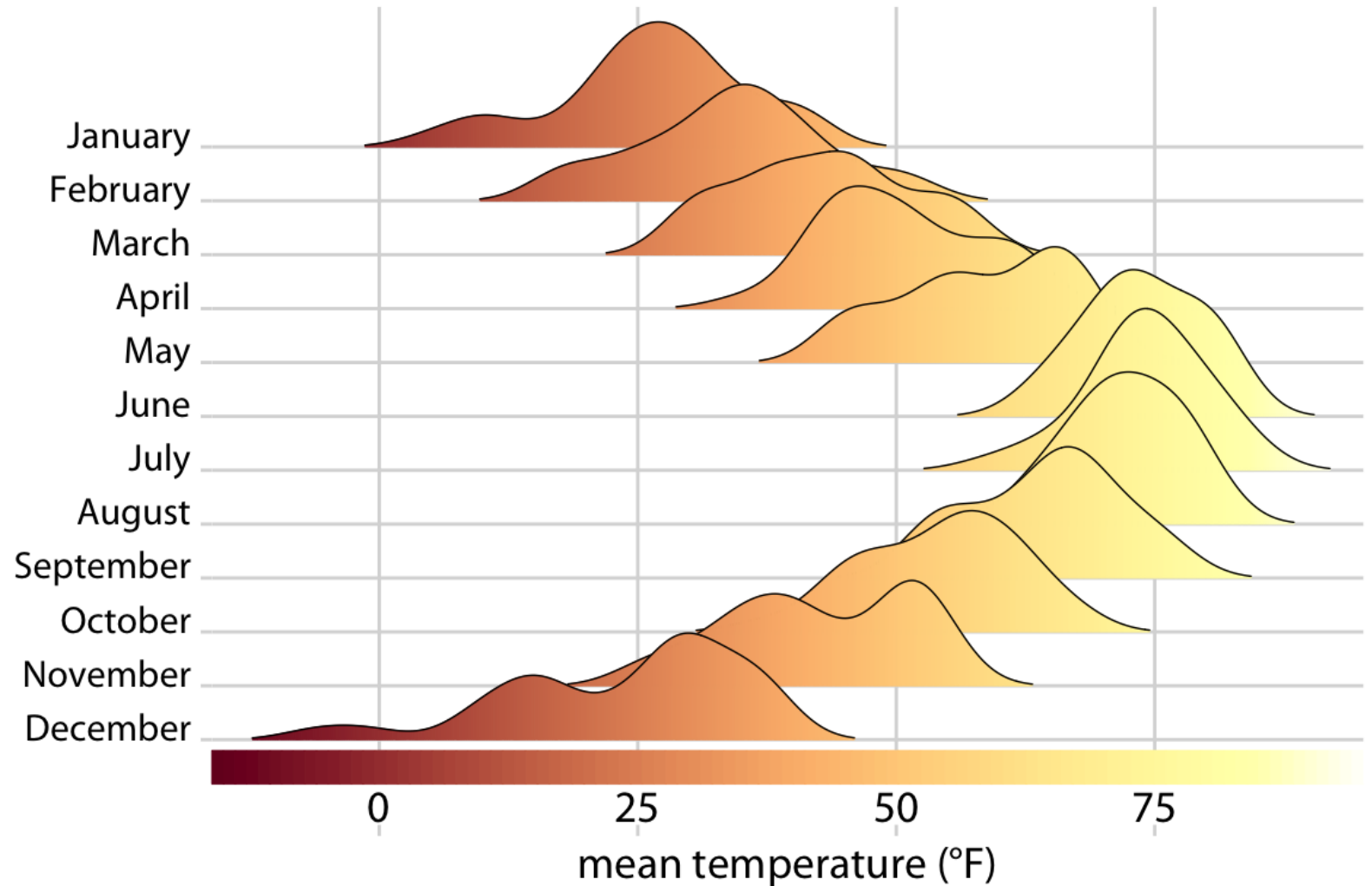
Criando figuras sem legenda

- Identificação direta: marcações em texto ou outros elementos visuais servem de guias ao resto da figura
- Combinar gráficos de densidade à gráficos de dispersão: inserir a identificação nos gráficos de densidade e manter o gráfico de dispersão livre



Criando figuras sem legenda

- Se uma mesma variável está mapeada à diferentes elementos estéticos, combinar as legendas desses elementos
- Temperatura mapeada à posição do eixo x e ao gradiente de cor: integrar a barra do gradiente de cor ao eixo

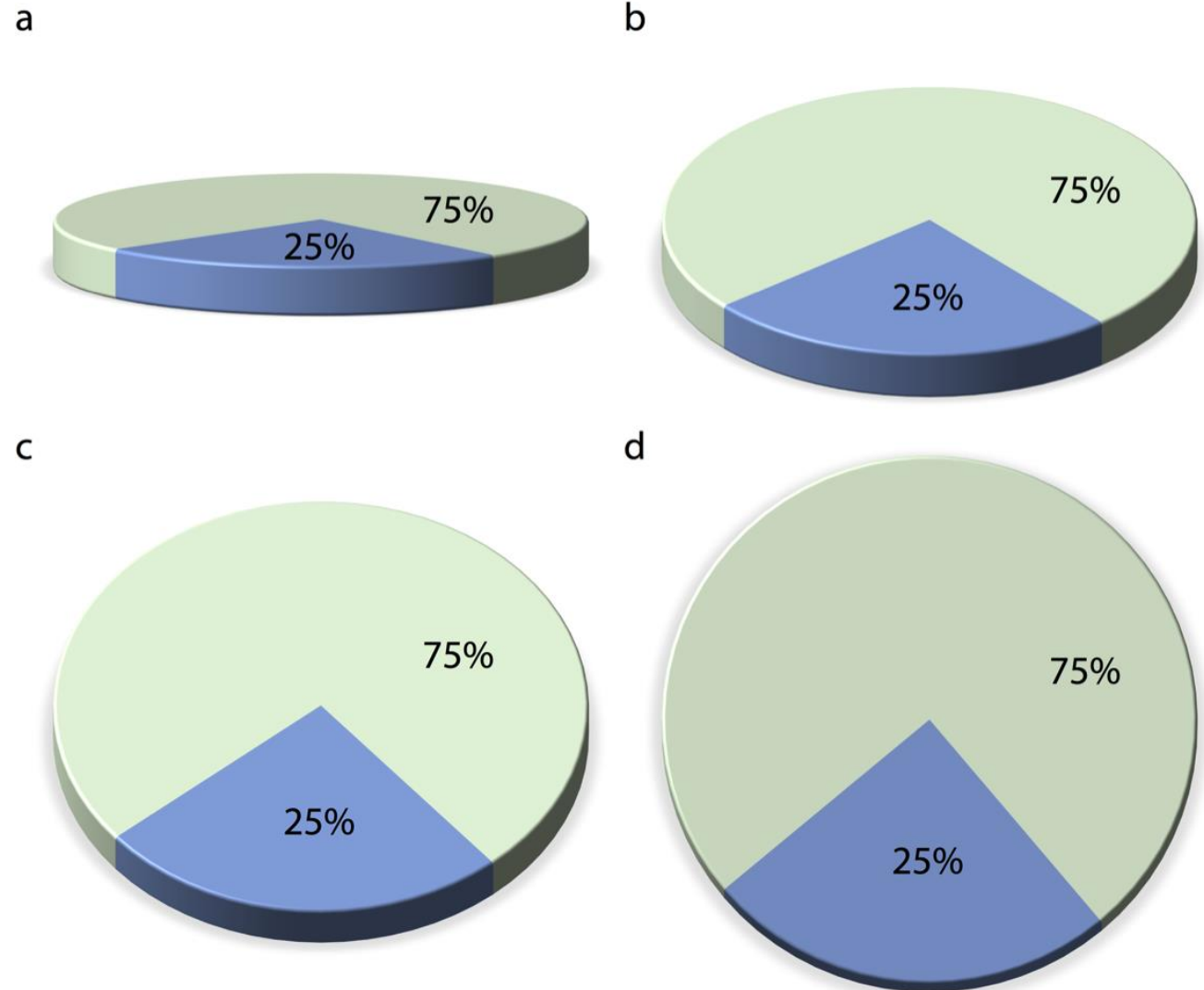


Problemas em imagens 3D

Uso “gratuito” de 3D

Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

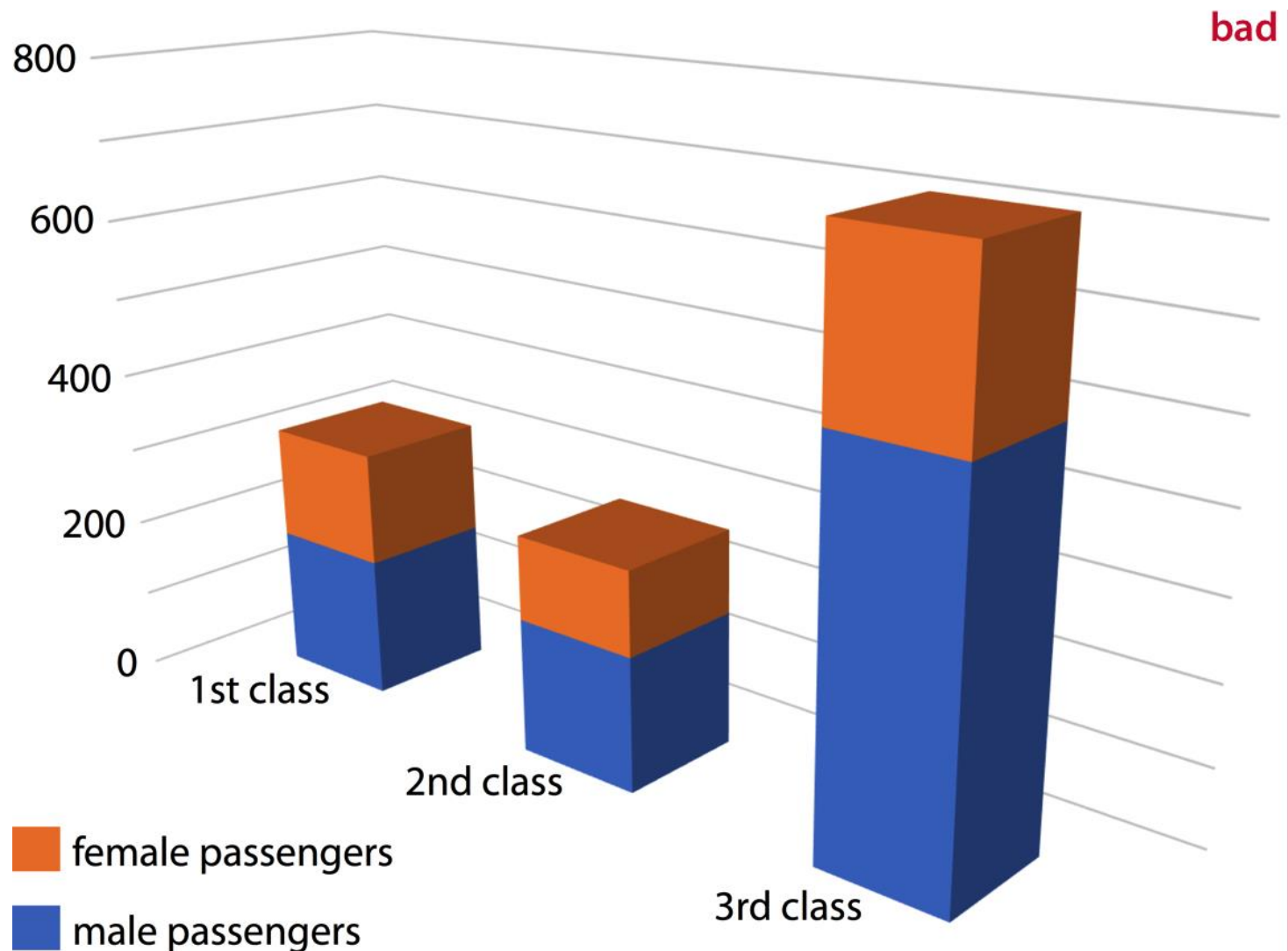
- Lembrete: toda vez que adicionamos dimensões que não existem nos dados, dificultamos a interpretação
- Problemas de percepção visual ao tentar mapear uma projeção 2D de uma imagem 3D novamente em um espaço tridimensional
- Gráfico de pizza: ao mudar o ângulo, a percepção de tamanho da fatia de 25% é modificada



Uso “gratuito” de 3D

Adaptado de:
WILKE, C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media, 2019.

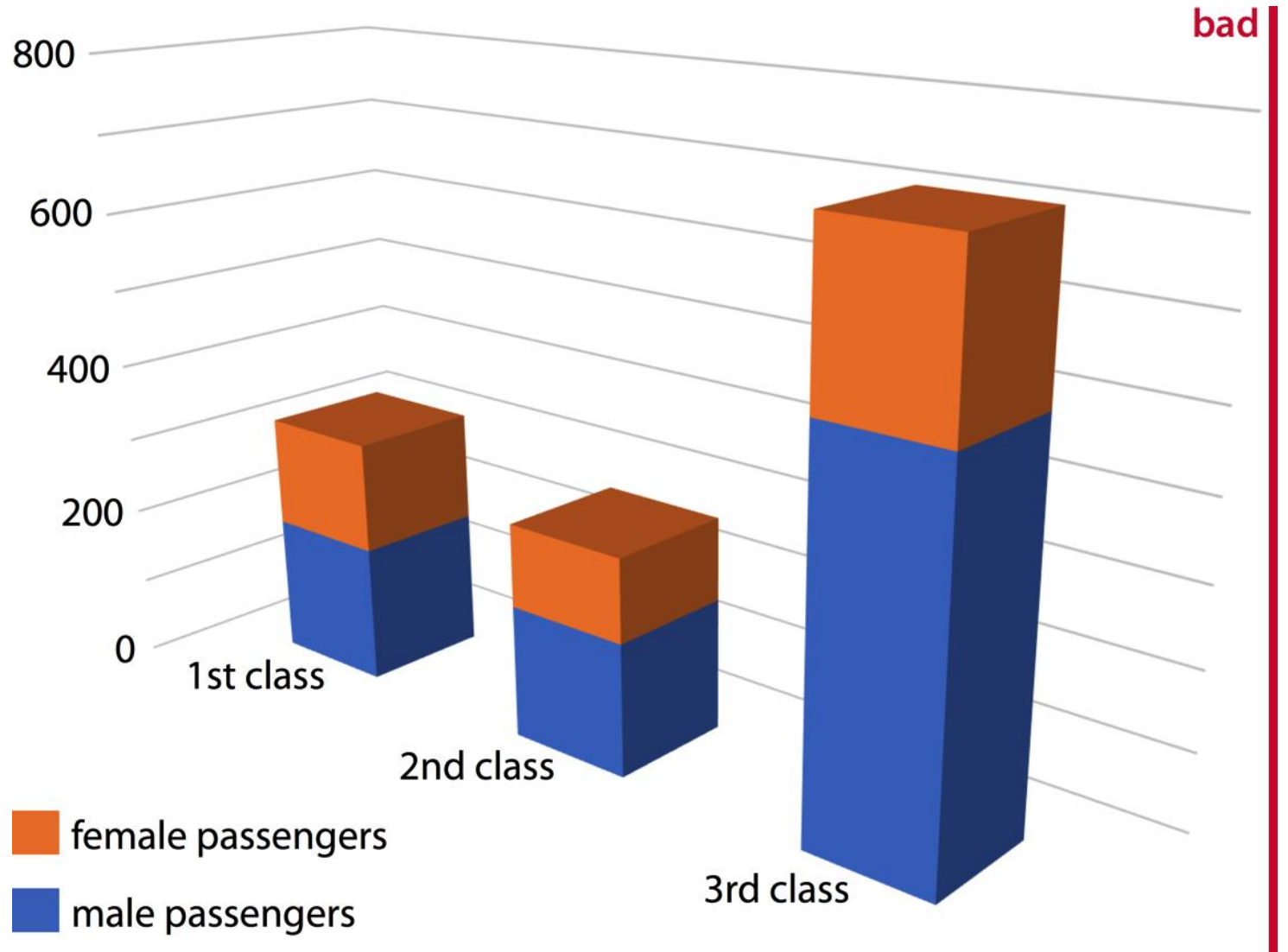
- Gráfico de barras: barras parecem menores do que realmente são
- Por exemplo, a figura sugere que menos de 300 passageiros viajaram na primeira classe do Titanic (valor real: 322)
- Ilusão surge pela distância “aparente” das colunas em relação às superfícies em que as linhas cinza do grid foram desenhadas



Uso “gratuito” de 3D

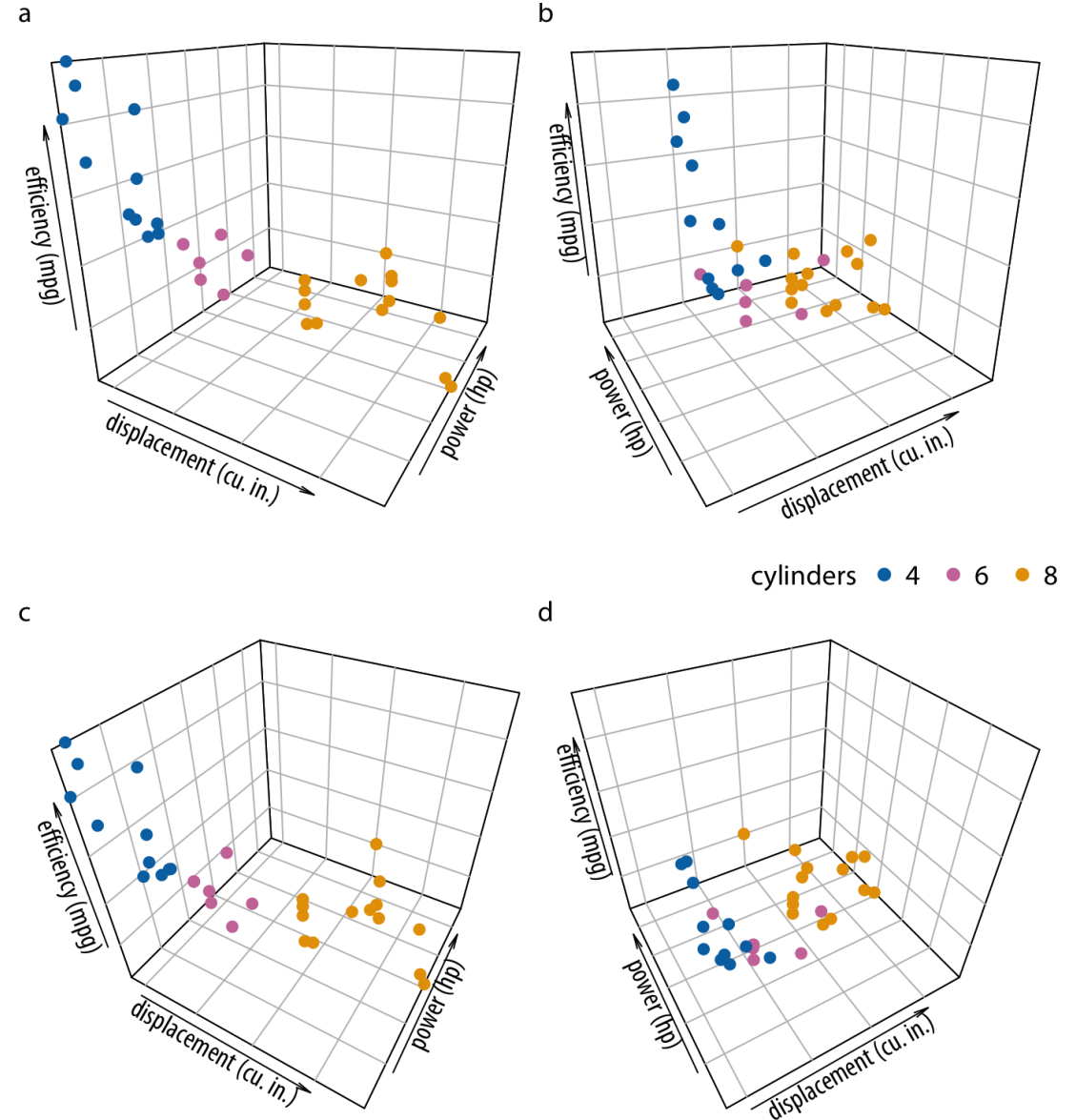
- Terceira barra domina a figura visualmente e faz o número de passageiros na terceira classe parecer maior do que realmente é
- Parece muito mais que o dobro da primeira classe, mas é 2,2 vezes maior (711 vs. 322)

Adaptado de:
WILKE, C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media, 2019.



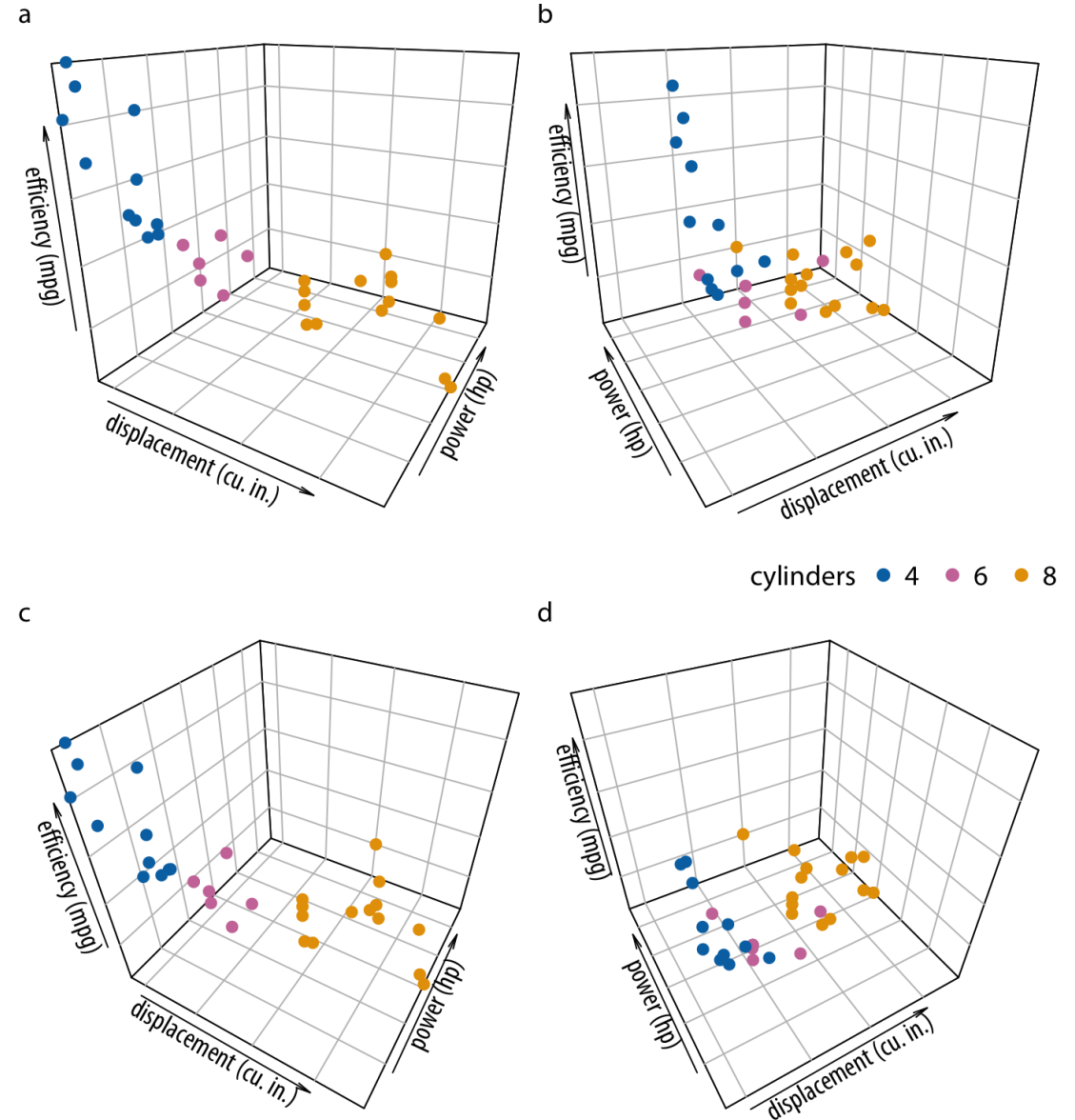
Uso de escalas de posição em 3D

- Em geral, gráficos incluindo um terceiro eixo (z) tendem a ser difíceis de interpretar
- Conjunto de dados “*mtcars*”:
adição da medida de eficiência
(milhas por galão) no eixo z
- Dependendo do ângulo de perspectiva, o aspecto do conjunto de dados parece ser diferente



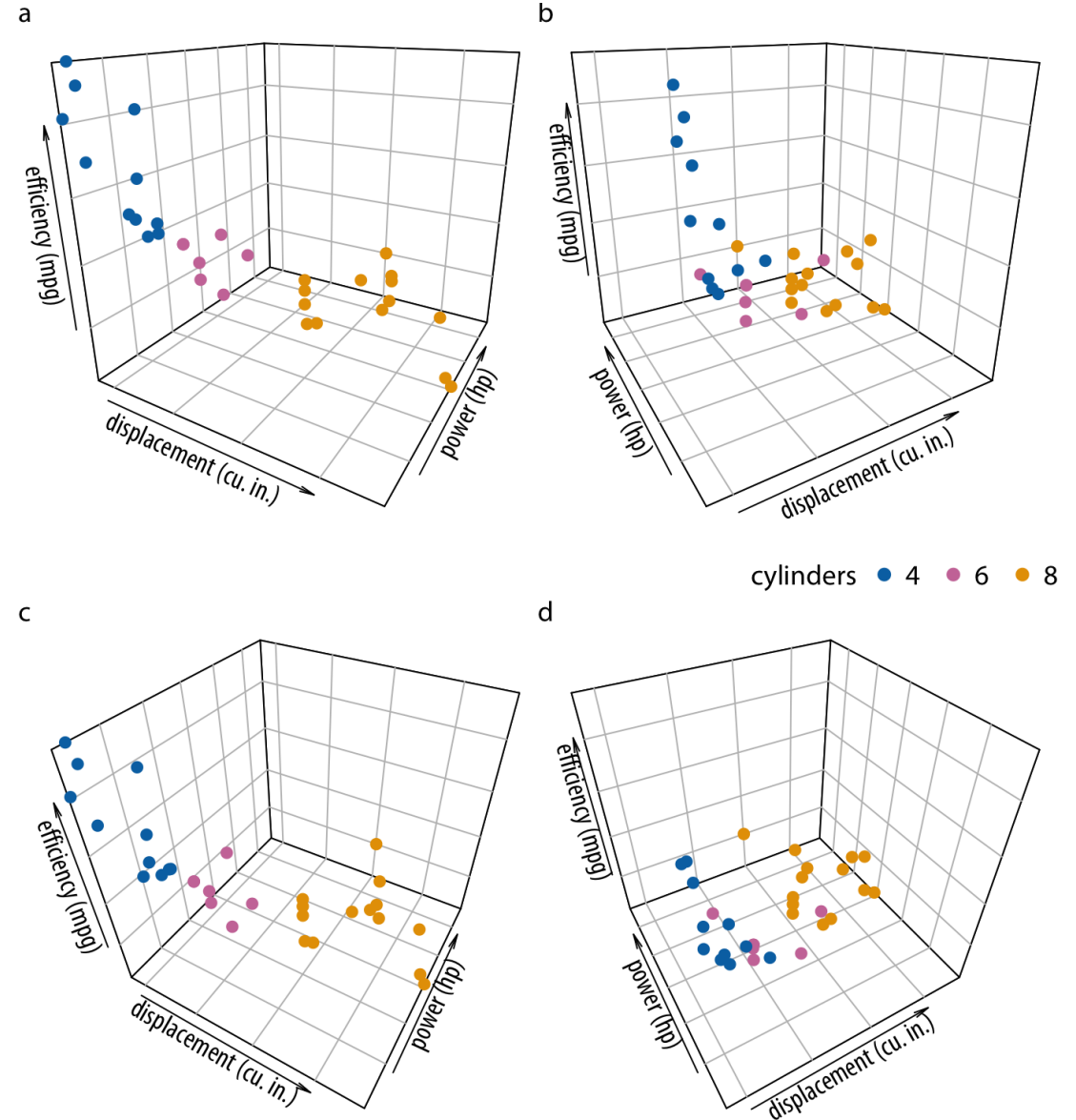
Uso de escalas de posição em 3D

- Principal problema uso de escalas de posição em 3D: ocorrência de duas transformações de dados sucessivamente
- Primeira transformação: mapeamento dos dados em um espaço 3D
- Segunda transformação: projeção dos dados posicionados no espaço 3D em uma figura plana 2D



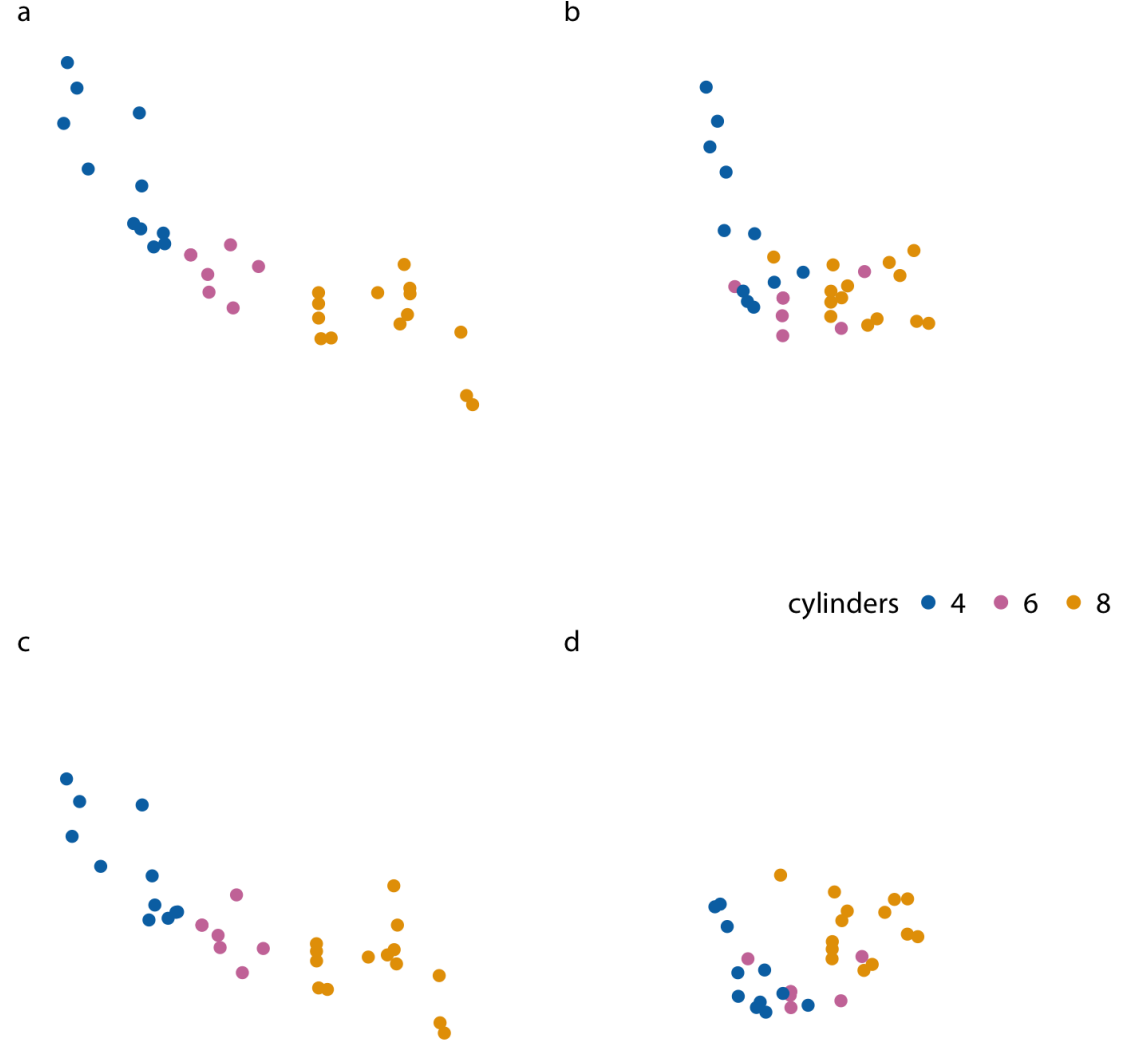
Uso de escalas de posição em 3D

- Segunda transformação não é mentalmente reversível: cada ponto na visualização 2D corresponde à uma linha de pontos no espaço 3D
- Impossível determinar a localização do ponto no espaço 3D real, apenas observando a visualização 2D



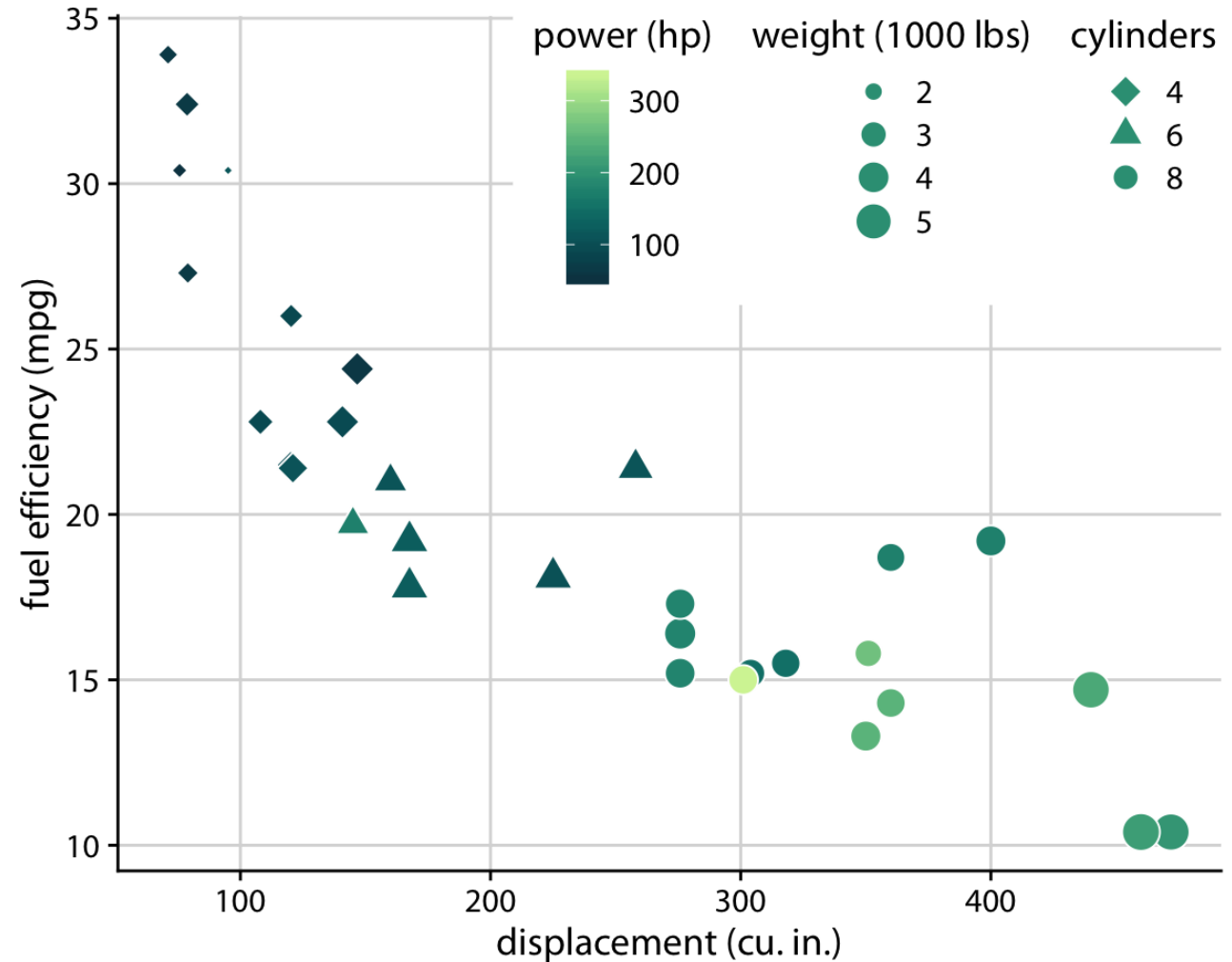
Uso de escalas de posição em 3D

- Ainda assim, nosso cérebro tenta fazer essa transformação
- Uso de pistas visuais presentes na imagem que dão um senso de tridimensionalidade e profundidade (eixos e ângulos)
- O que acontece se removermos as pistas de profundidade?



Uso de escalas de posição em 3D

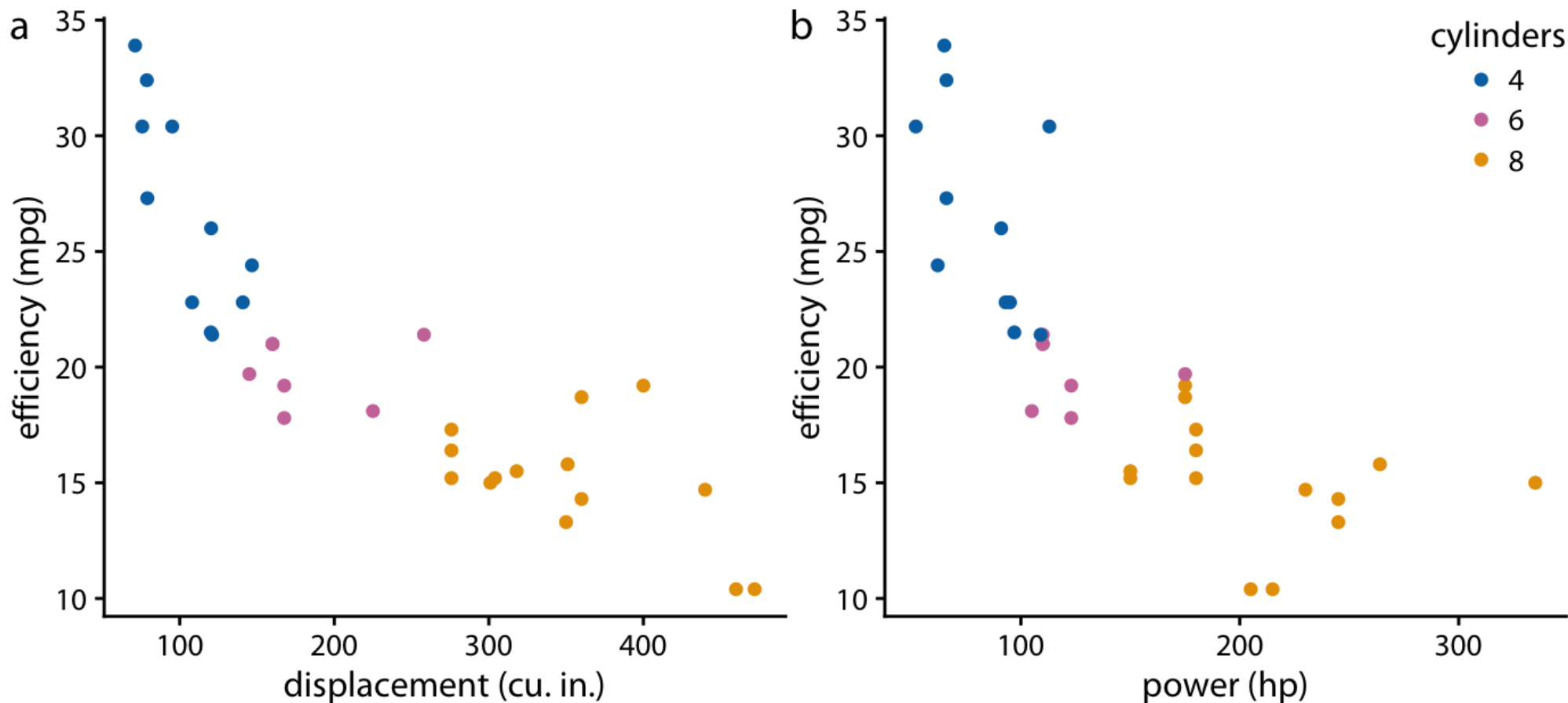
- Solução: evitar a ocorrência das duas transformações
- Mapear os dados ao espaço 2D, e evitar adicionar uma terceira dimensão como uma escala de posição
- Fazer uso de cores, formas e tamanhos para mapear variáveis



Uso de escalas de posição em 3D

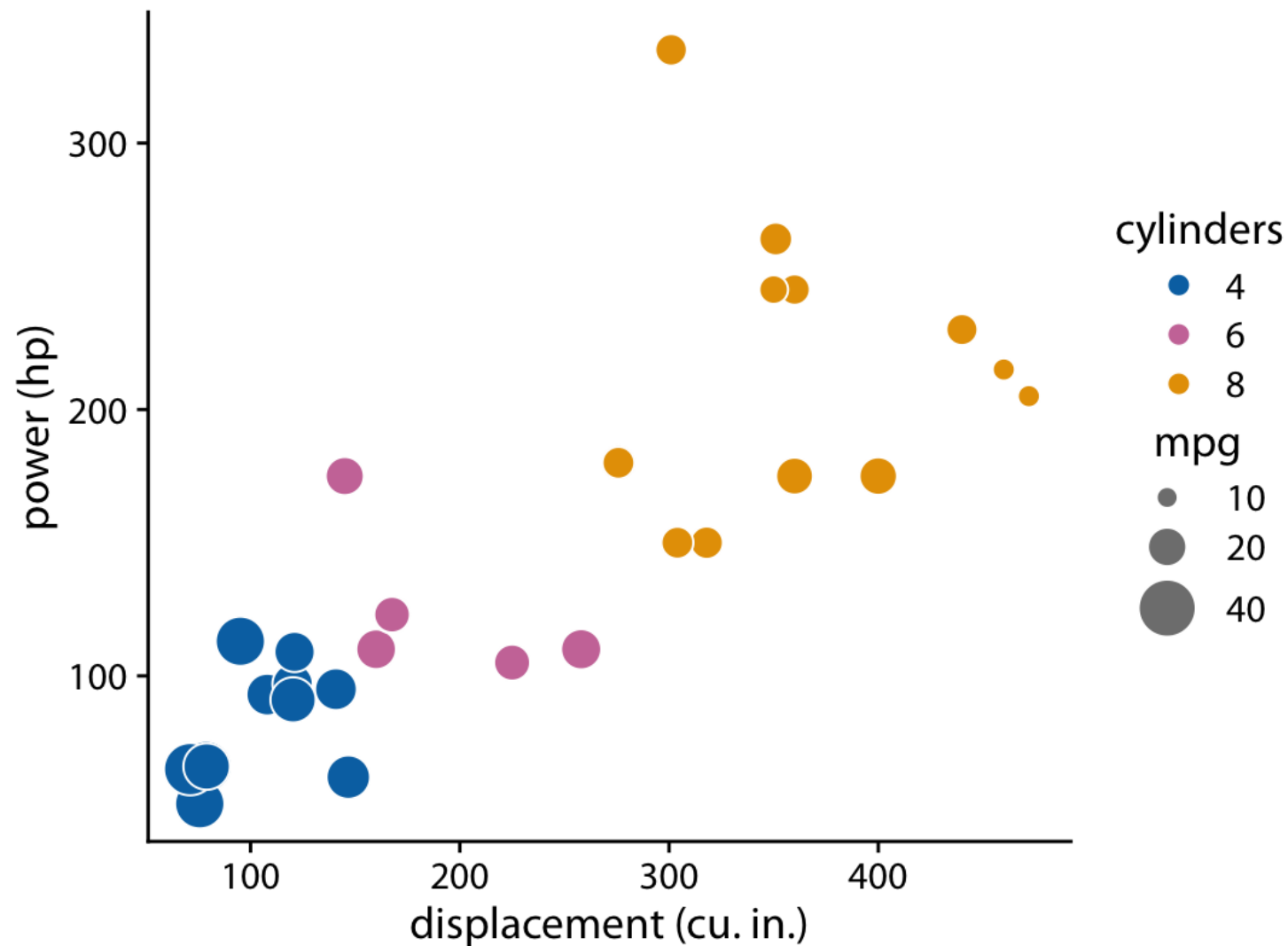
- Se o foco é na relação entre variáveis, com foco na eficiência como variável de resposta: fazer dois gráficos separados

Adaptado de:
WILKE, C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media, 2019.



Uso de escalas de posição em 3D

- Se o foco é na relação entre potência e cilindrada, com a eficiência sendo uma variável secundária:
- Plotar potência e cilindradas nos eixos x e y
- Plotar a eficiência como tamanho



Uso apropriado de visualizações em 3D

- Representação de objetos reais em 3D ou informações mapeadas a esses objetos
- Representação do relevo de uma paisagem ou local



Uso apropriado de visualizações em 3D

- Representação de objetos reais em 3D ou informações mapeadas a esses objetos
- Representação da estrutura tridimensional de uma proteína, utilizando cores para indicar regiões mais e menos conservadas



sequence conservation



Uso apropriado de visualizações em 3D

- Em geral, visualizações 3D funcionam melhor em ambientes interativos, com possibilidade de rotação e inspeção do objeto representado em diferentes ângulos
- Em visualizações não-interativas: rotação lenta, permitindo com que o cérebro construa o cenário 3D a partir das imagens obtidas a partir de diferentes ângulos
- Embora não seja possível em material impresso, perfeitamente viável em material disponibilizado online ou em apresentações de slides

