

GENE7033 – Tópicos Especiais em Genética I:

Visualização de dados para publicações científicas

Profª Drª Chirlei Glienke

Drª Desirrê Petters-Vandresen

Visualizando proporções

Dr^a Desirrê Petters-Vandresen

24/11/2022

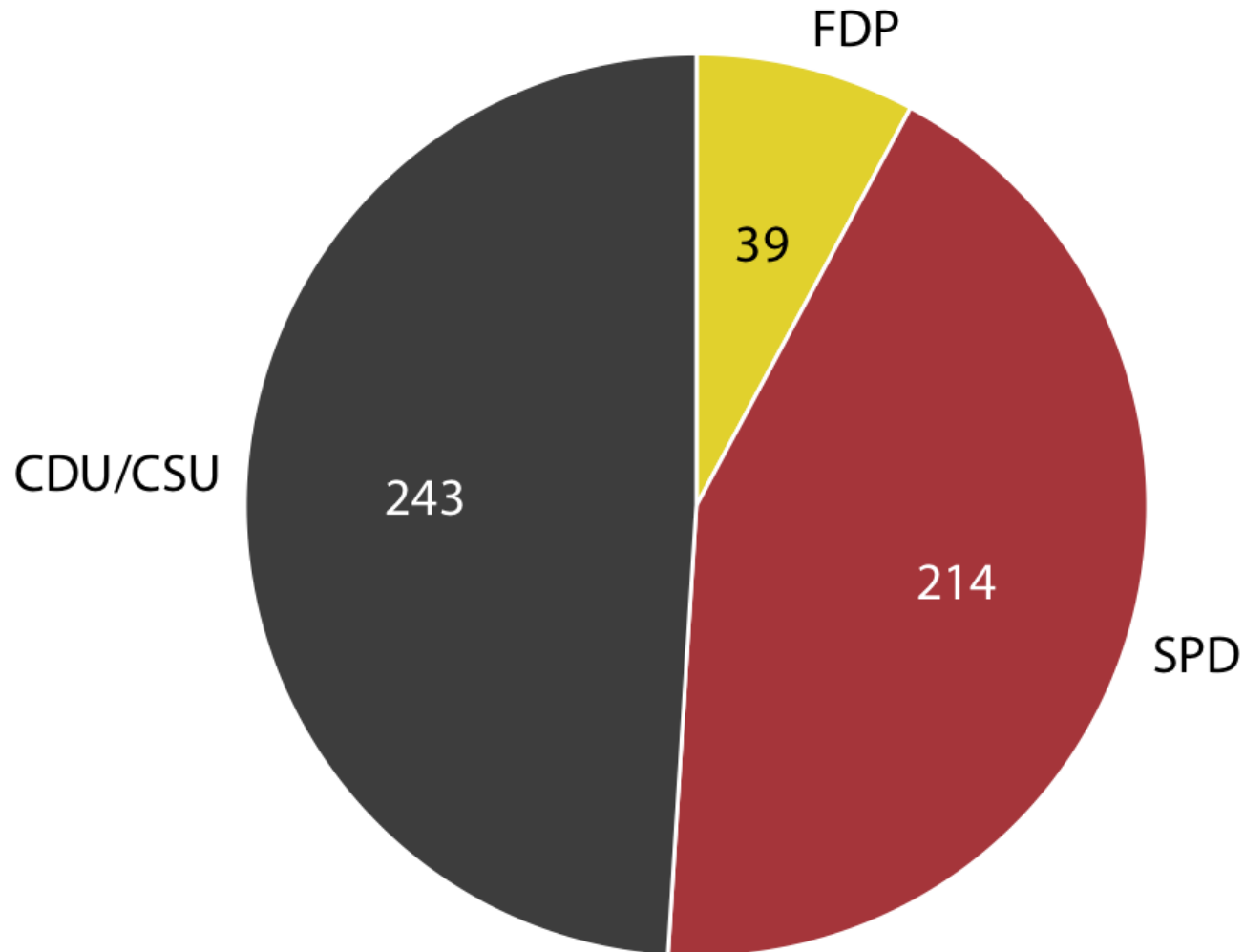
Finalidade

- Visualização de como um variável pode ser subdividida em diferentes grupos, os quais representam proporções do todo
- Exemplos
 - Proporção de homens e mulheres numa população
 - Proporção de votos em cada candidato em uma eleição
- Desafios:
 - Muitos subgrupos a serem representados
 - Visualizar mudanças em proporções ao longo do tempo ou entre condições/tratamentos
- Várias possibilidades dependendo do contexto

Gráfico de pizza

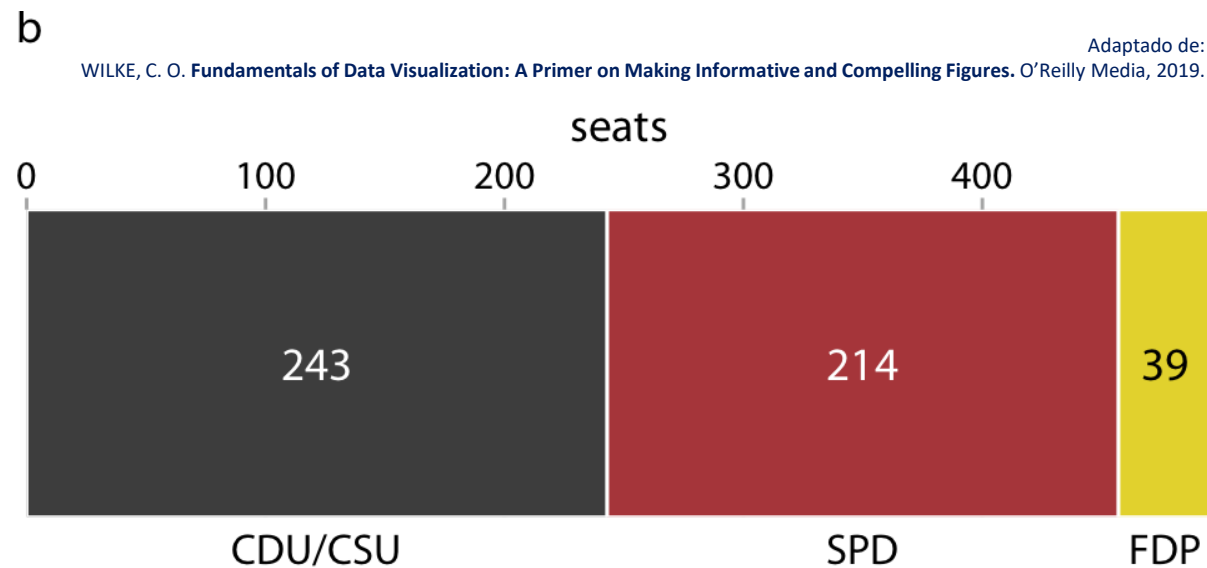
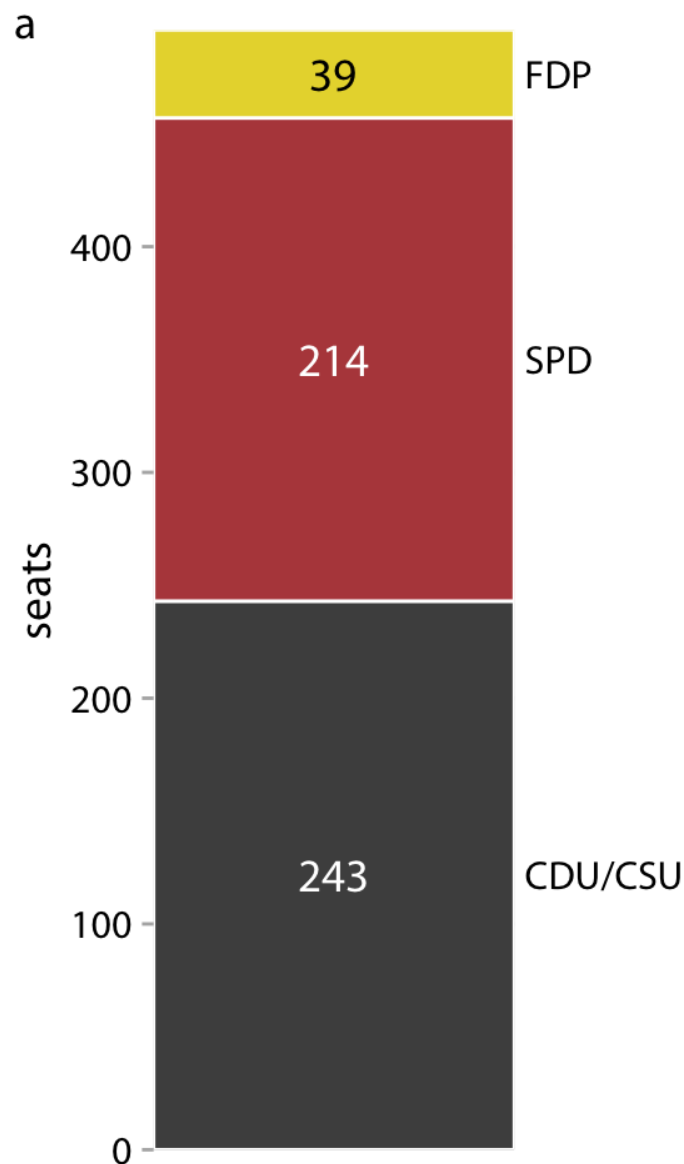
- Amplamente difundido em apresentações de negócios e igualmente “odiado” entre estatísticos e cientistas de dados
- Divisão de um círculo em fatias de modo que cada fatia seja proporcional à fração do total que ela representa
- Funcionam bem para:
 - Ênfase em frações simples ($1/2$, $1/3$, $1/4$)
 - Conjuntos de dados pequenos e/ou com poucos subgrupos

Gráfico de pizza – um estudo de caso



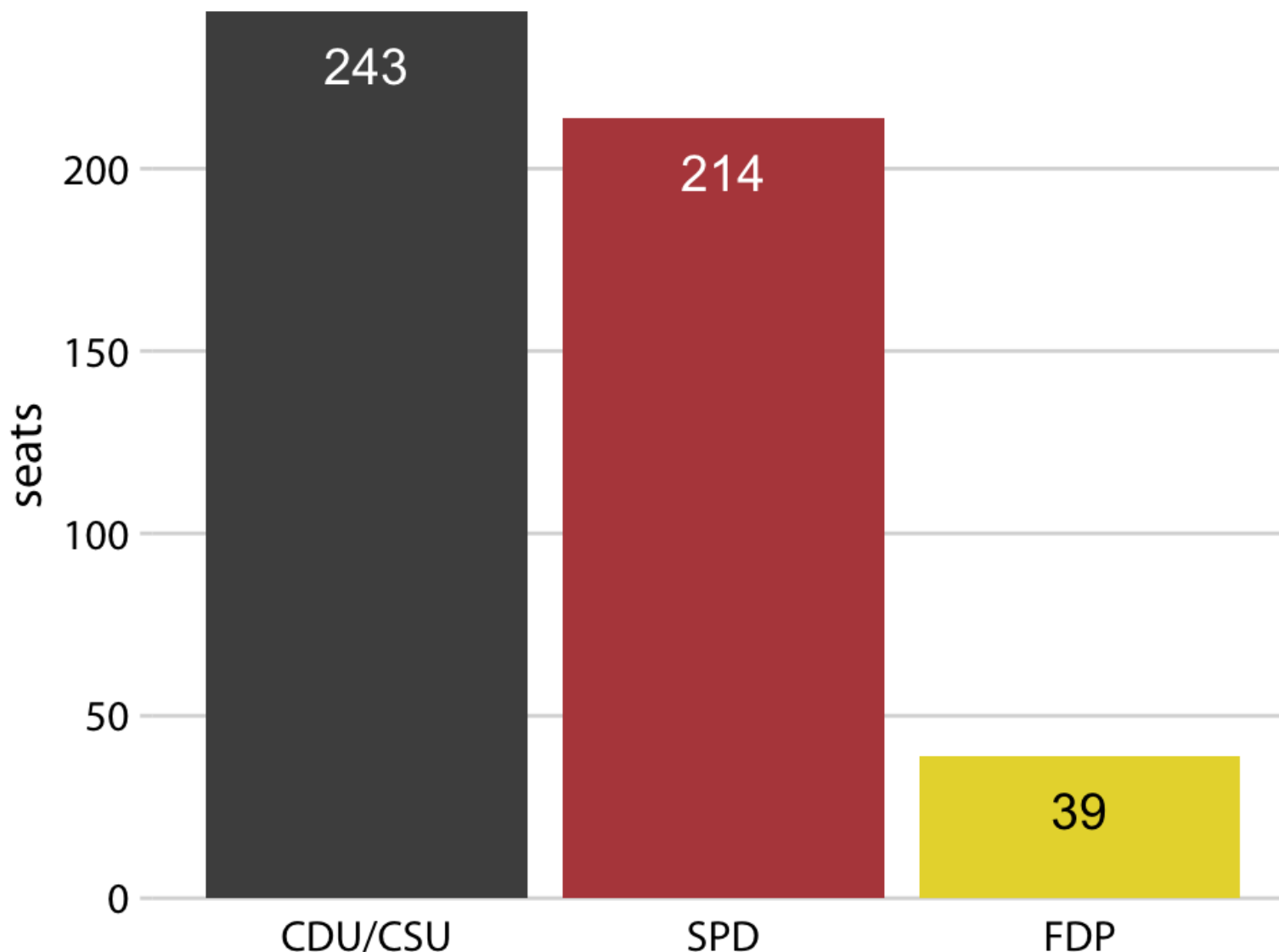
- Parlamento alemão (Bundestag) de 1961 até 1983:
- Três partidos, CDU/CSU e SPD com a maior quantidade de membros e FDP com poucos membros
- Coalisção SPD e FDP: ligeira maioria em relação à CDU/CSU

Gráfico de pizza – um estudo de caso



- Ligeira maioria da coalisção SPD e FDP sobre CDU/CSU não está representada de forma tão óbvia
- Aspecto visualmente estranho no uso de somente um gráfico de barras empilhadas, enquanto somente um gráfico de pizza não gera estranheza

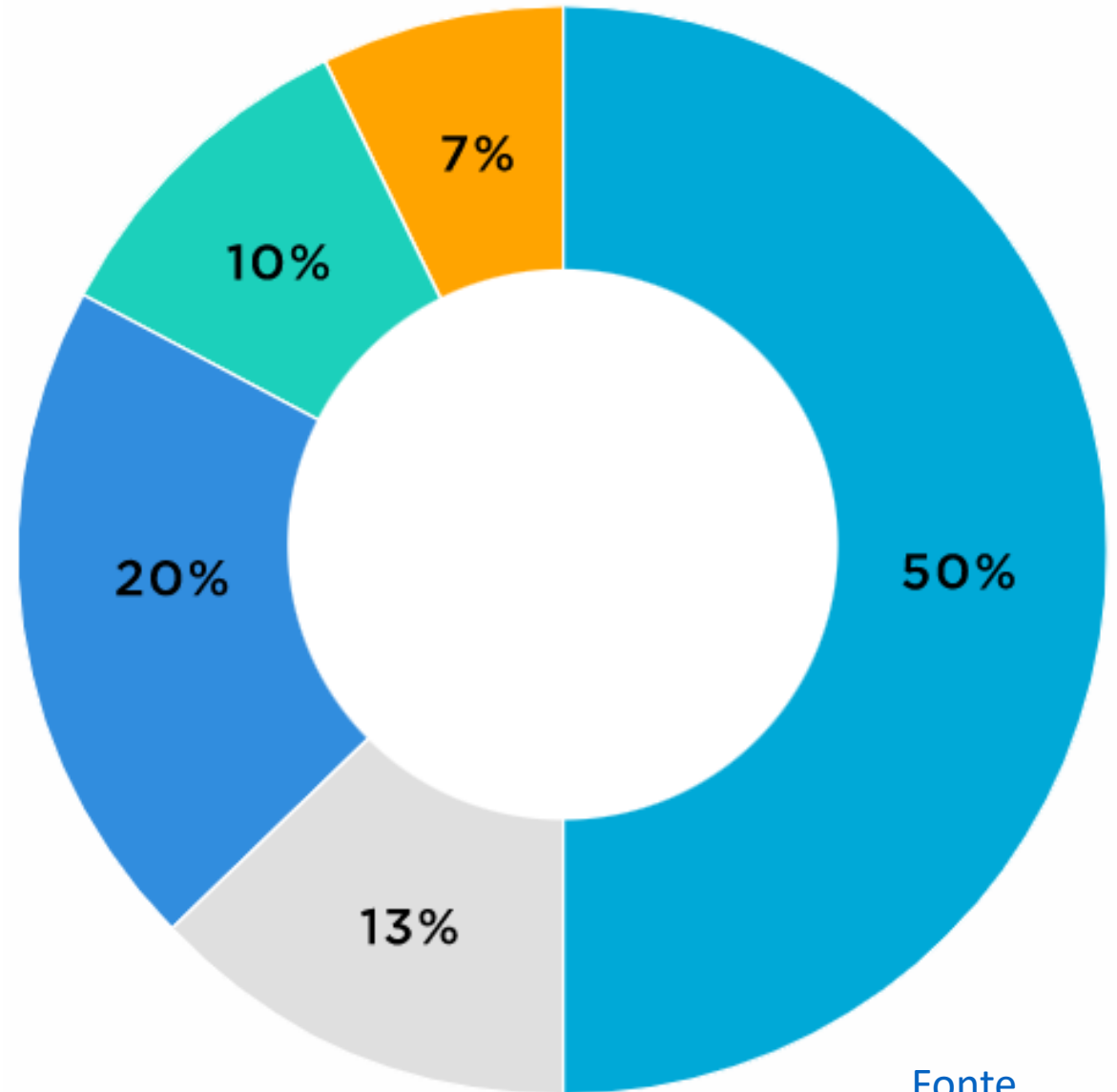
Gráfico de pizza – um estudo de caso



- Fácil comparação entre os grupos
- Dificuldade de percepção do quanto cada grupo corresponde no total
- Ligeira maioria da coalisção SPD e FDP também não está óbvia

Gráfico de rosca/donut

- Adaptação do gráfico de pizza: centro cortado para se assemelhar a uma rosca/donut
- Muito além da questão estética:
 - Remoção do centro induz o leitor a se concentrar no comprimento do arco
 - Evita comparação da área de cada fatia com a área da pizza inteira
 - Espaço dentro da rosca pode ser utilizado para representar dados e rótulos



Fonte

Gráfico de barras agrupadas – um estudo de caso

- Market share de cinco empresas hipotéticas (A, B, C, D, E), com valores próximos de 20% no início
- Visualização da evolução de market share ao longo de 3 anos: dificuldade de comparação entre empresas

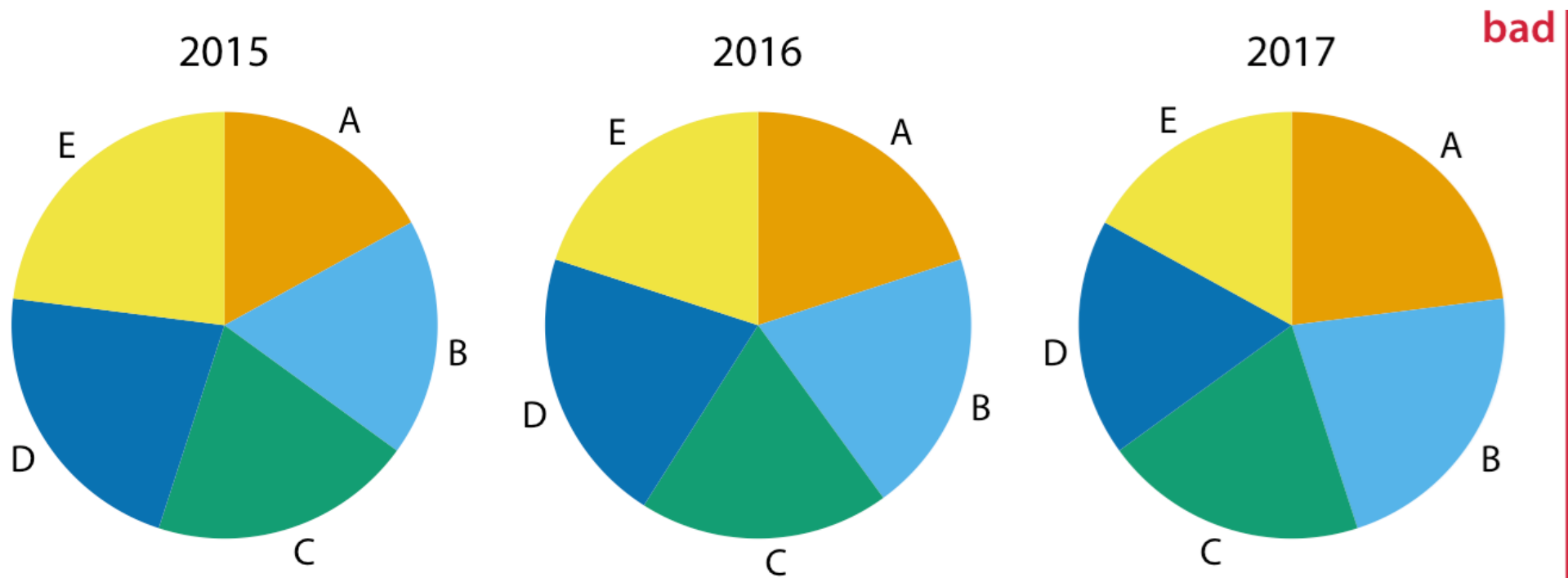
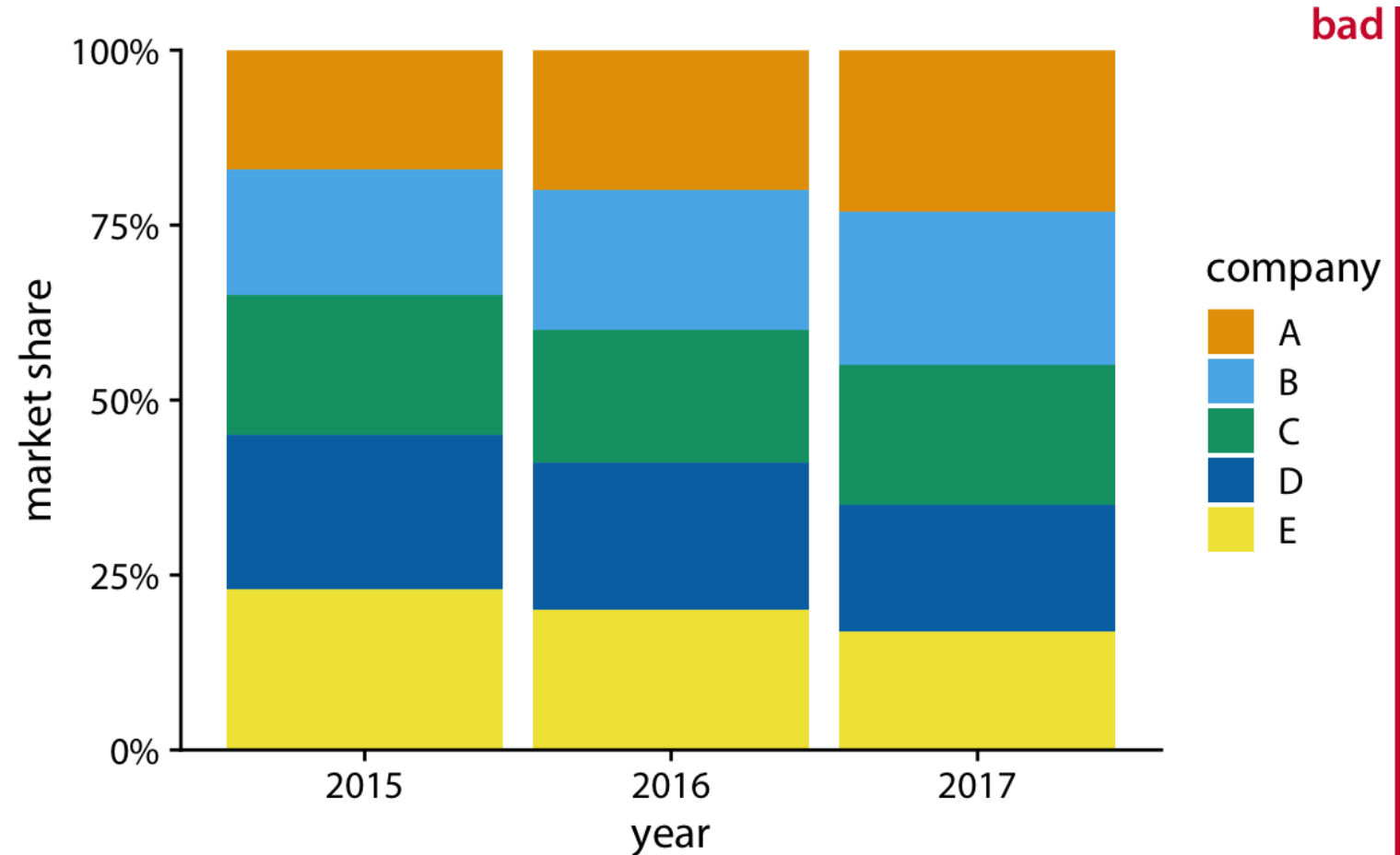


Gráfico de barras agrupadas – um estudo de caso

- Maior facilidade na percepção de crescimento da empresa A e diminuição de market share da empresa E
- Problema: linha de base instável para comparação das outras empresas



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

Gráfico de barras agrupadas – um estudo de caso

- Facilidade de percepção do aumento de market share das empresas A e B e diminuição para empresas D e E
- Mudança no panorama geral entre 2015 e 2017 (empresa A saindo de menor market share para maior entre as cinco empresas)

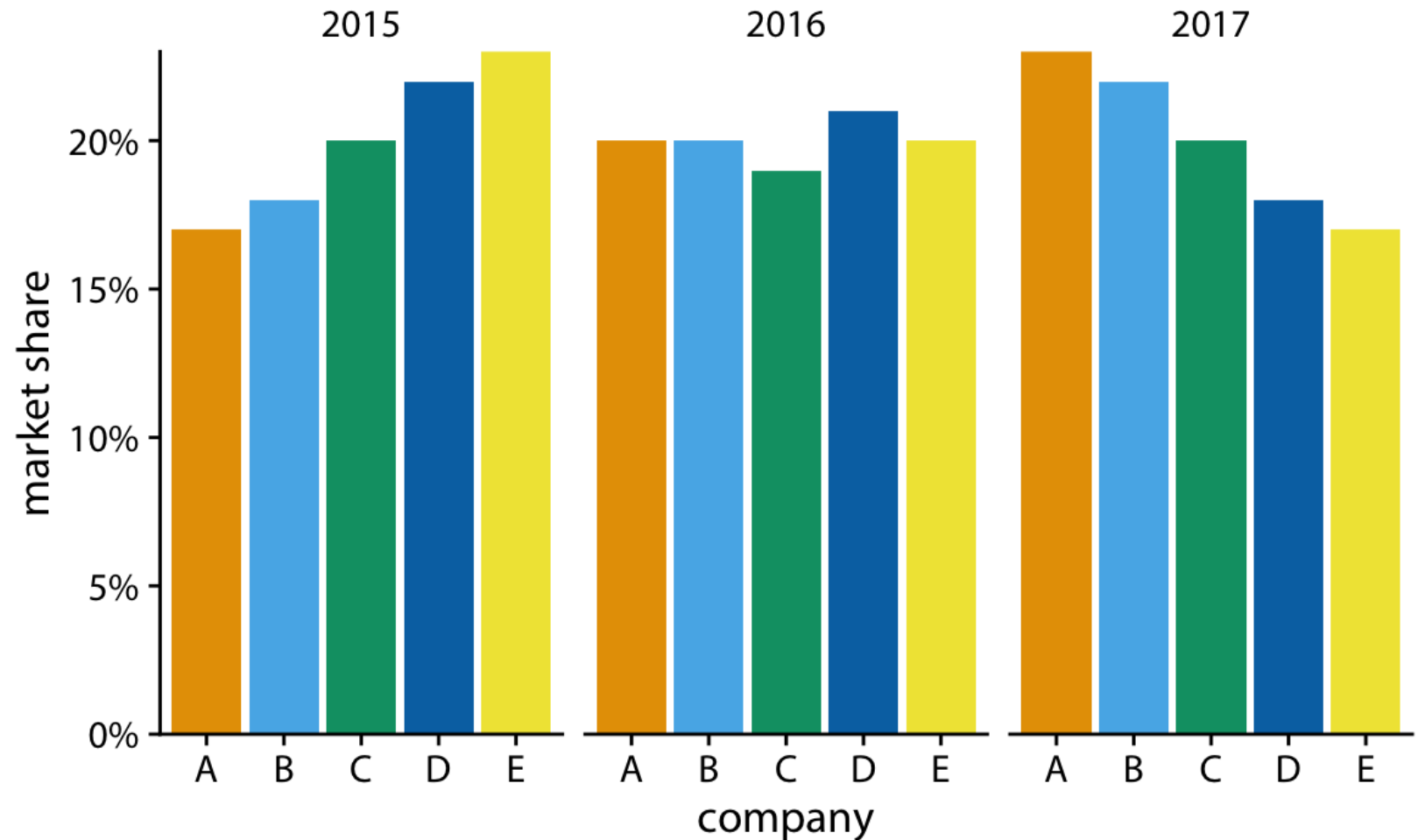
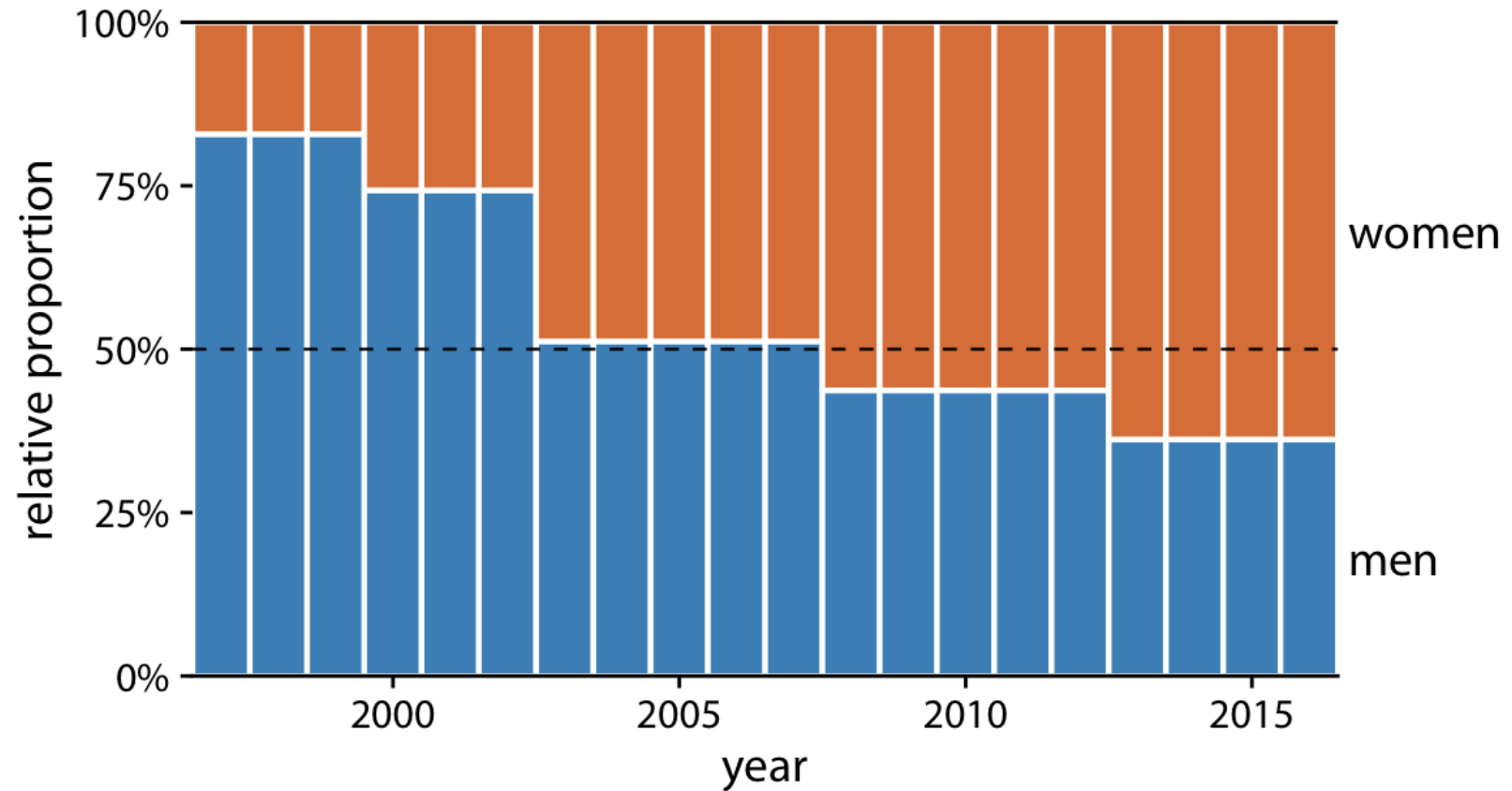


Gráfico de barras empilhadas – um estudo de caso

- Apesar de não recomendado com muitos subgrupos (linhas de base instáveis), funciona bem com dois grupos e as barras são ajustadas para totalizar 100%
- Exemplo: evolução da participação de mulheres no parlamento de Ruanda
- Adição de limiar de 50% para auxiliar na interpretação



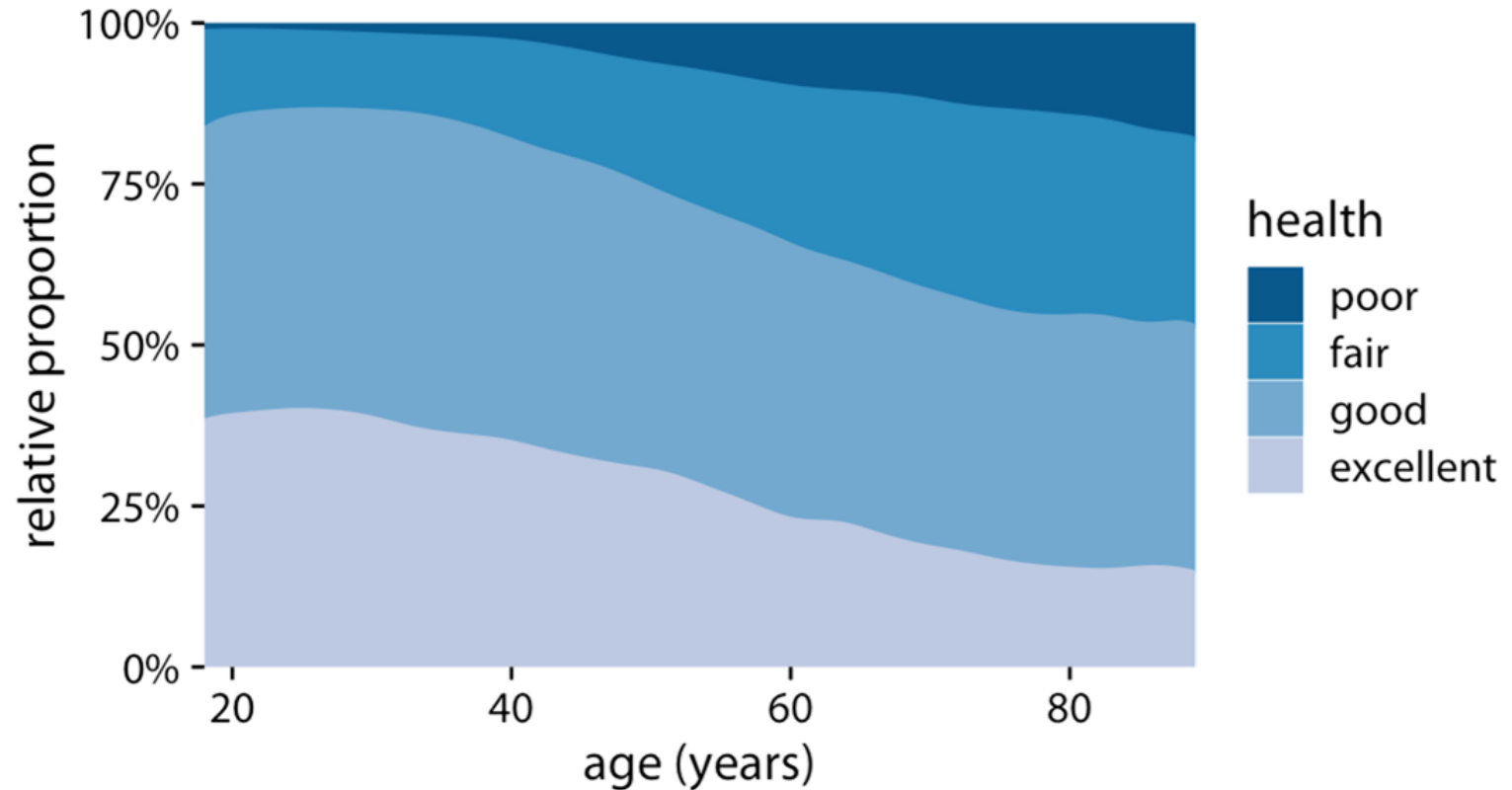
Adaptado de:
WILKE, C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

Comparativo

	Gráfico de pizza	Gráfico de barras empilhadas	Gráfico de barras agrupadas
Funcional para visualizar os dados como proporções que compõem um todo			
Fácil comparação visual entre proporções relativas			
Ênfase em frações simples (1/2, 1/3, 1/4)			
Visualmente funcional mesmo para conjuntos de dados pequenos			
Funcional quando o conjunto todo é subdividido em muitas proporções			
Funcional para visualização simultânea de proporções em muitos conjuntos, ou de proporções em séries temporais			

Gráfico de densidades empilhadas – um estudo de caso

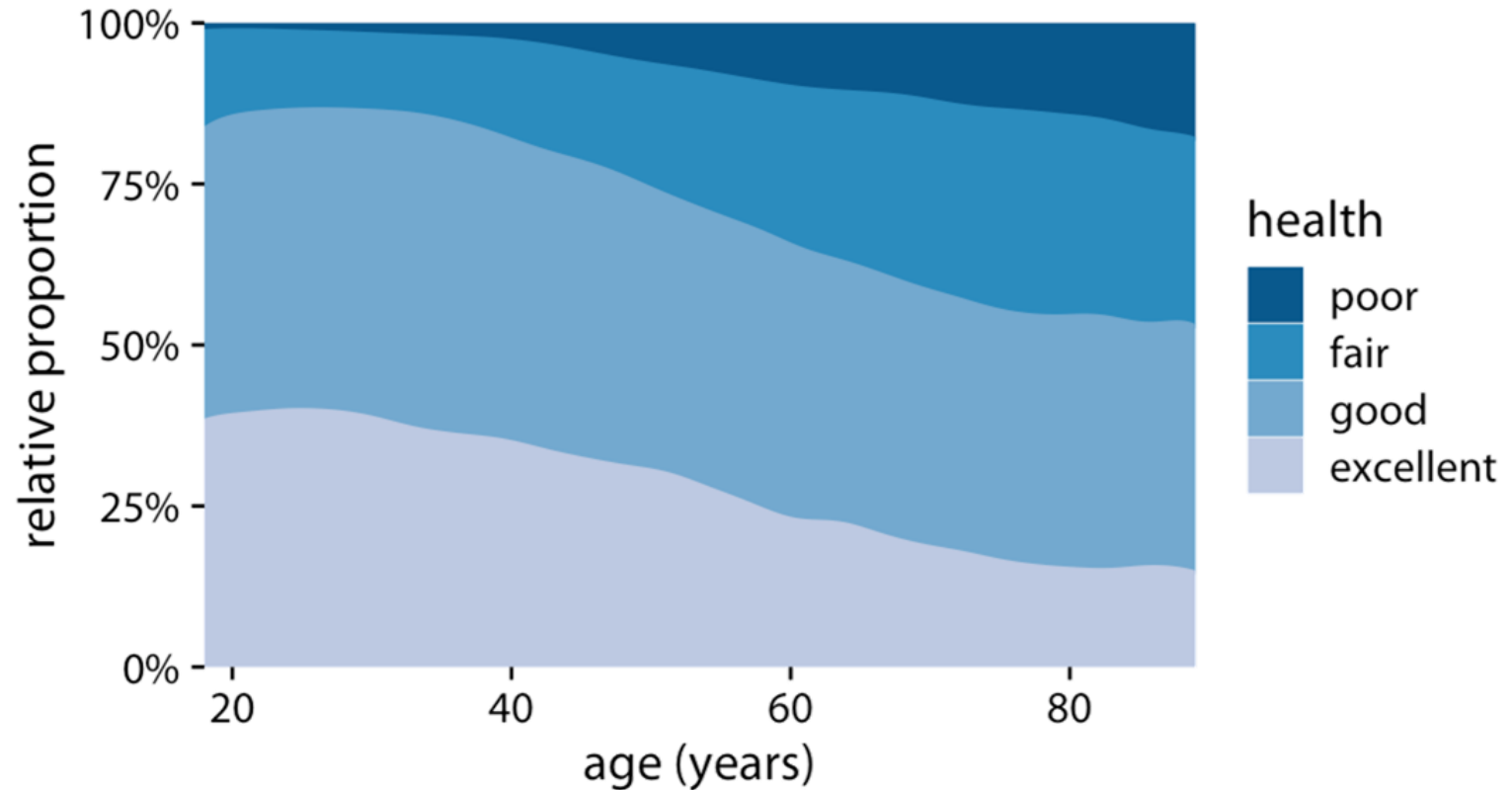
- Representação da mudança de proporções para variáveis contínuas
- Exemplo: condições de saúde ao longo de diferentes idades
- Percepção de que a saúde piora com o envelhecimento, mas que boa parte da população se mantém em boas condições de saúde até em idades avançadas



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

Gráfico de densidades empilhadas – um estudo de caso

- Limitação: sem representação adequada da proporção de cada grupo etário na amostra
- Porcentagem de pessoas em boas condições de saúde se mantém constante ao longo das idades
- Número absoluto de pessoas em boas condições de saúde diminui à medida que a quantidade de pessoas diminui (falecimento de pessoas mais velhas)



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

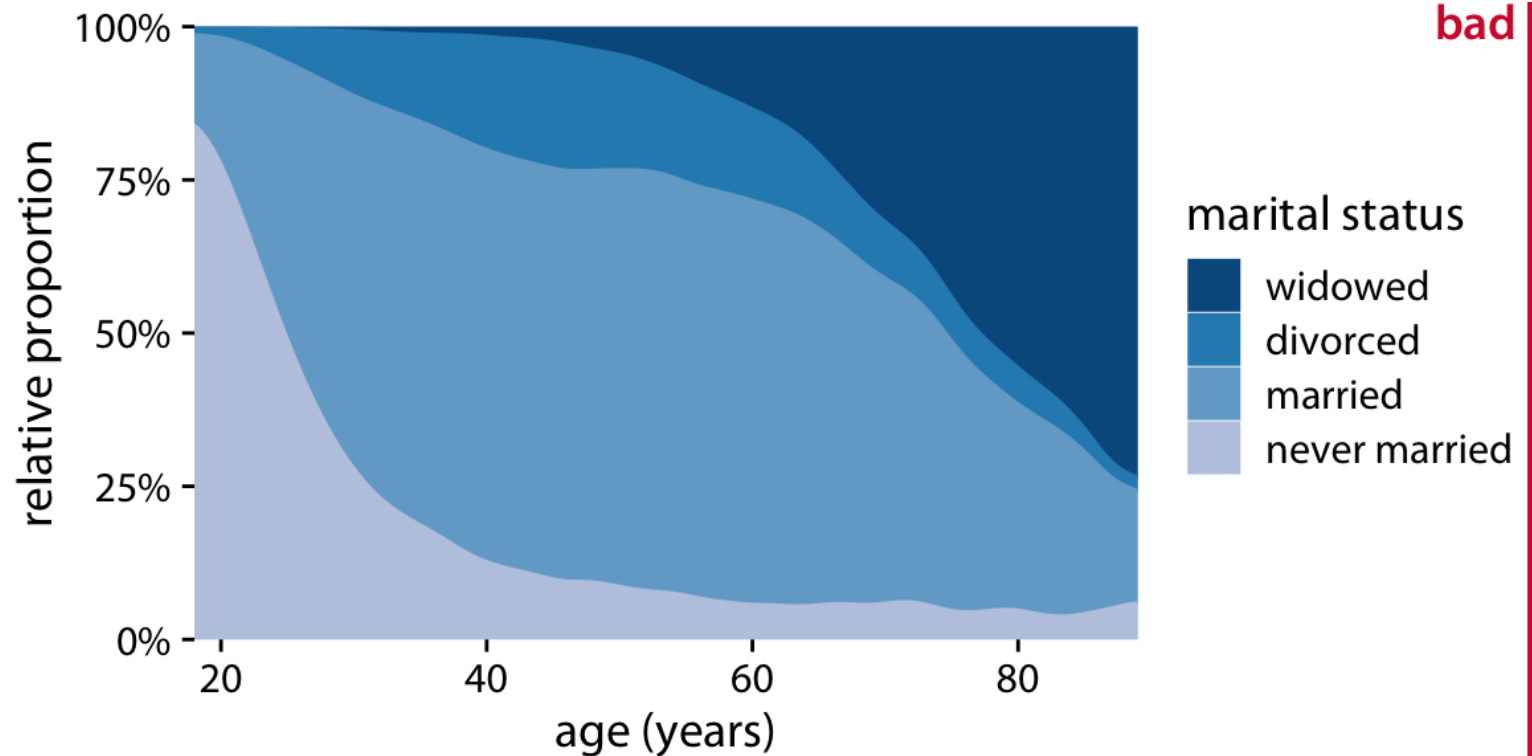
Gráfico de densidade– um estudo de caso

- Em números absolutos, a quantidade de pessoas em condições excelentes ou boas de saúde decai após os 30-40 anos, mas a quantidade de pessoas em condições razoáveis permanece constante



Gráfico de densidade– um estudo de caso

- Mudanças de estado civil de acordo com a idade são muito mais drásticas
- Quantidade de pessoas nunca casadas relativamente alta aos 20 anos
- Quantidade de pessoas viúvas relativamente alta a partir de 70 anos
- Viés que distorce a representação dos outros grupos



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

Gráfico de densidade– um estudo de caso

- Proporção de pessoas casadas tem um pico perto dos 30 anos, divorciados tem um pico perto dos 40, e viúvos com um pico perto dos 70 anos

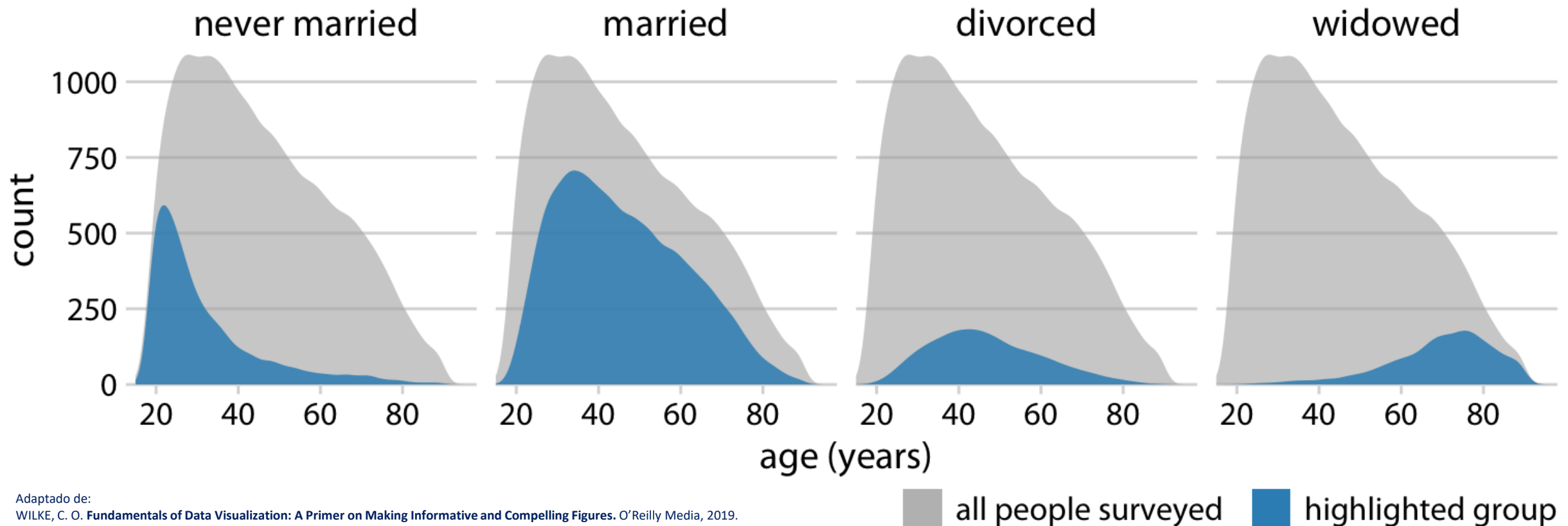
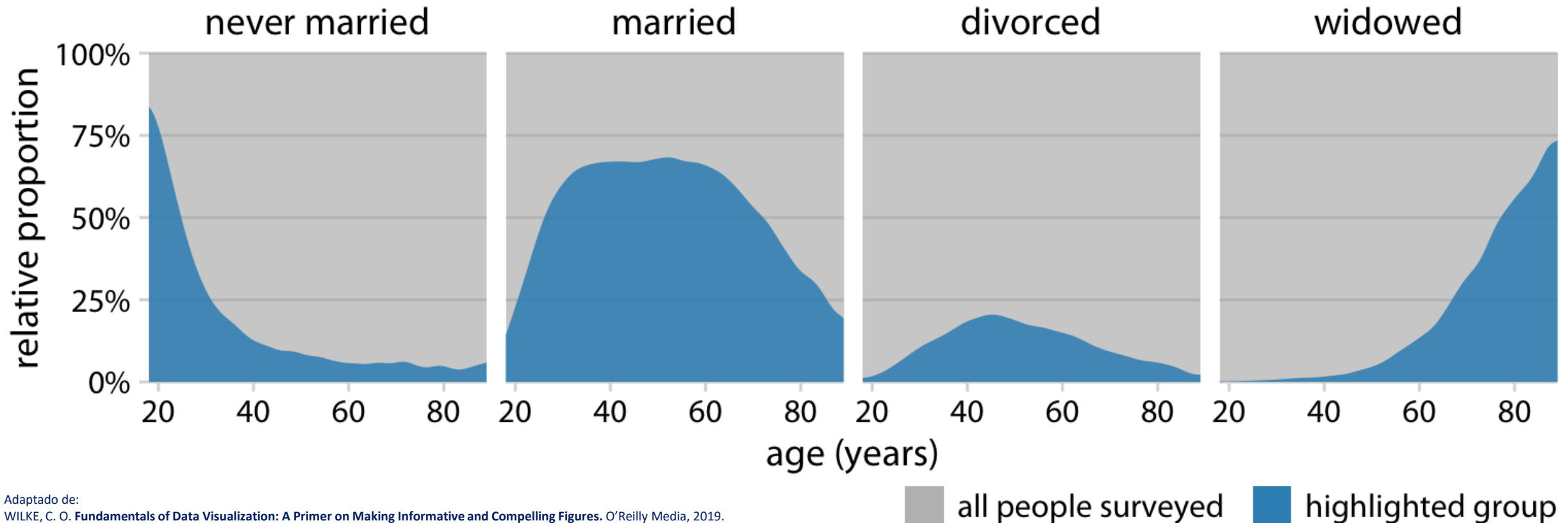


Gráfico de densidade– um estudo de caso

- Proporção relativa ao invés de números absolutos: determinar momentos em que 50% das pessoas estão casadas, 50% estão viúvas



Visualizando proporções aninhadas

Dr^a Desirrê Petters-Vandresen

24/11/2022

Finalidade

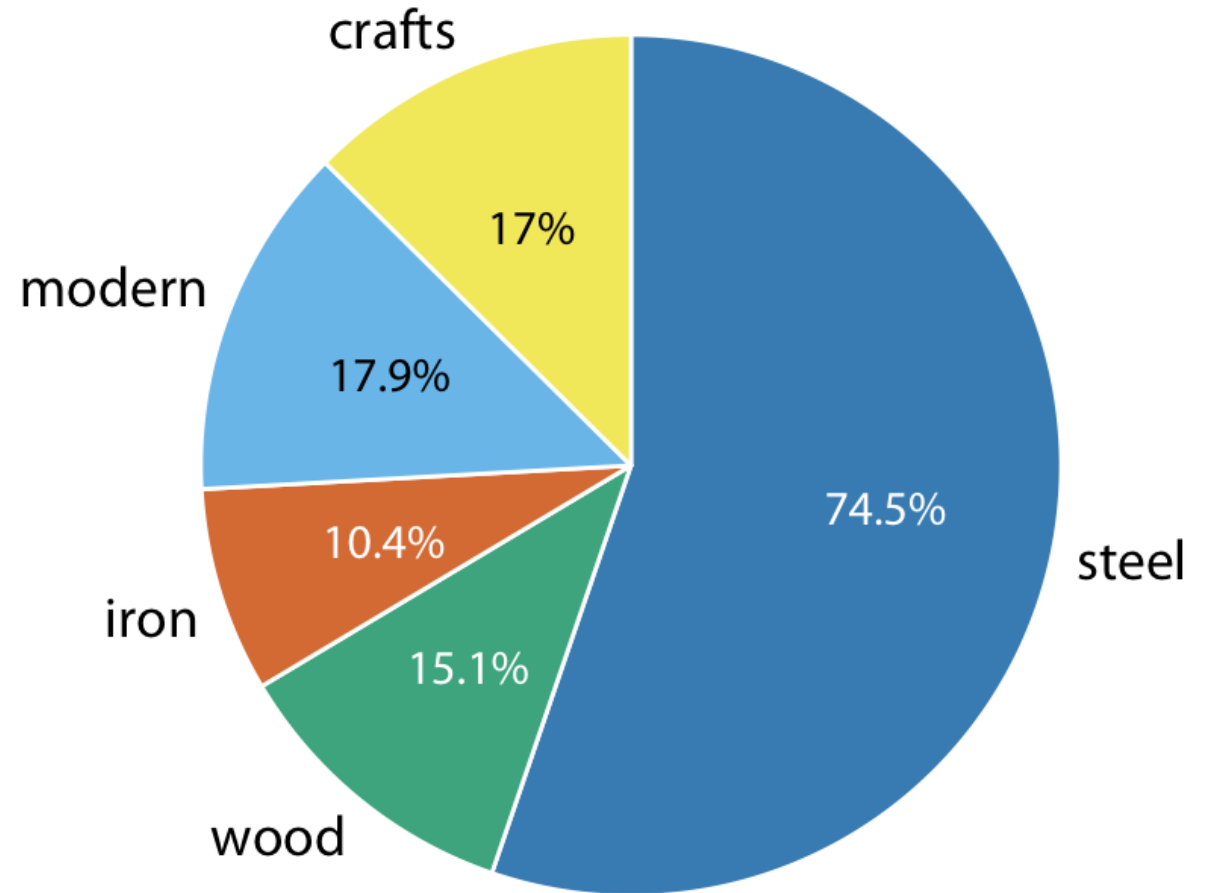
- Visualização de como um variável pode ser subdividida em diferentes grupos, os quais representam proporções do todo, e de como essas proporções podem ser novamente subdivididas
- Exemplos
 - Proporção de homens e mulheres numa população
 - Proporção de homens e mulheres com diferentes níveis de escolaridade
 - Proporção de votos em cada candidato em uma eleição
 - Proporção de homens/mulheres que votaram em cada candidato
 - Proporção de pessoas de diferentes níveis de escolaridade que votaram em cada candidato
- Cada variável categórica adiciona cria uma camada adicional de subdivisão dos dados



[Fonte](#)

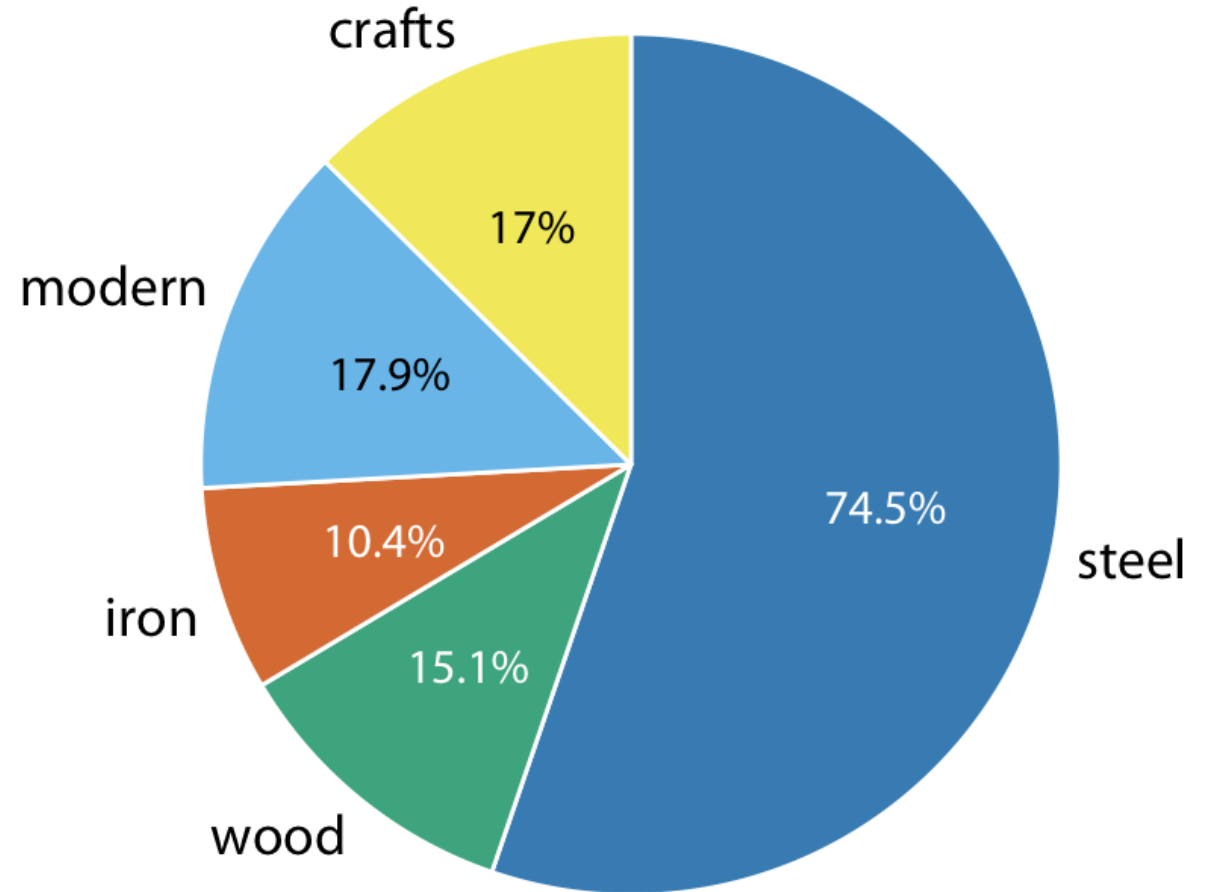
Estratégias a evitar: gráfico de pizza

- 106 pontes de Pittsburgh
 - Material utilizado: aço, ferro, madeira
 - Época de construção: artesanais (antes de 1870) e modernas (depois de 1940)
- Visualizar proporção de acordo com material e época de construção



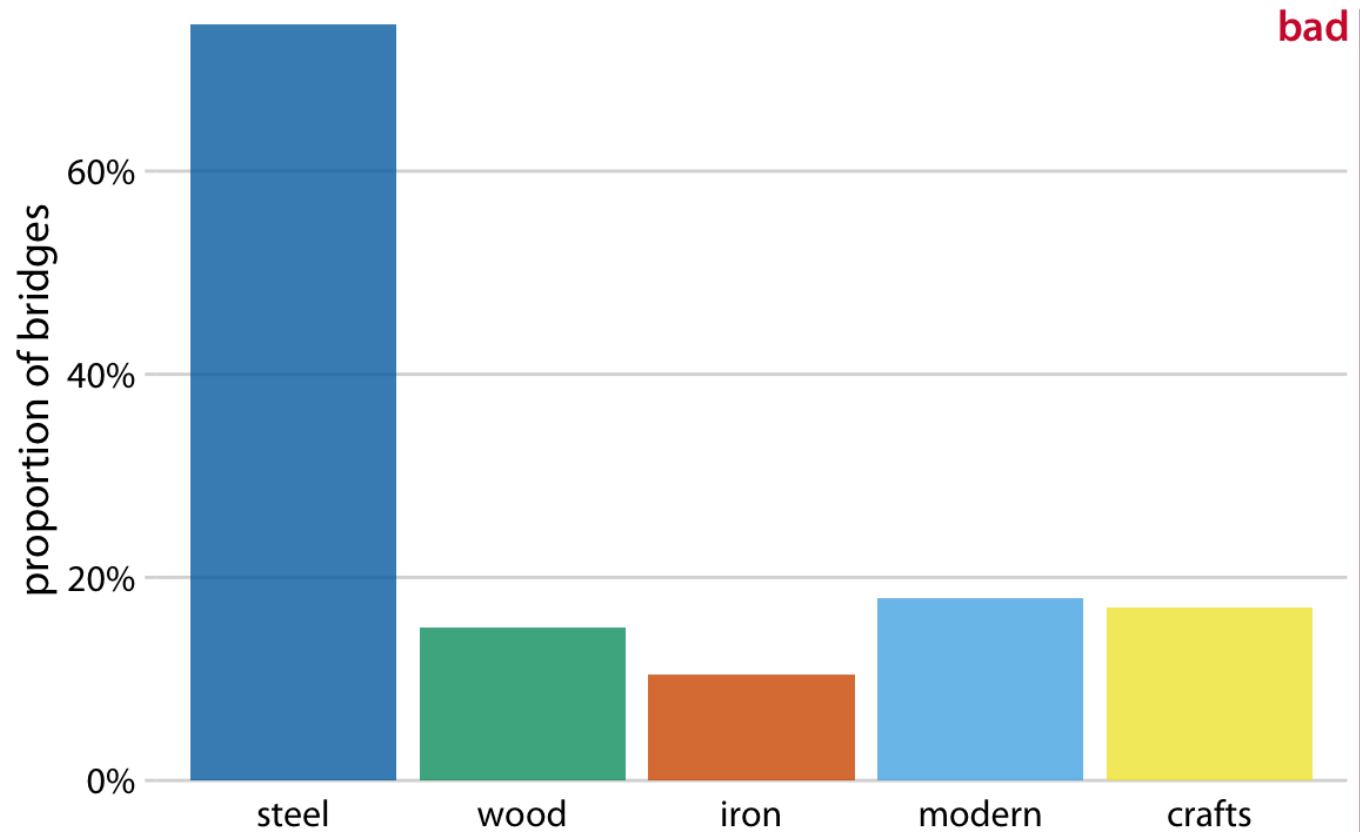
Estratégias a evitar: gráfico de pizza

- Problema: num gráfico de pizza, as fatias devem totalizar 100% (e não 135%)
- Algumas pontes estão sendo contabilizadas duas vezes
- Material utilizado ou época de construção, separadamente, já totalizam 100%: seja antiga ou moderna, toda ponte é feita ou de aço, ferro ou madeira



Estratégias a evitar: gráfico de barras

- Contabilizar o mesmo dado não é necessariamente um problema se o tipo de gráfico utilizado não exige que as proporções totalizem 100%
- Limitações: gráficos de barras não demonstram que há sobreposição entre categorias
- Qual é a proporção de cada tipo de material em relação à época de construção?
- Leitores desatentos: conclusão de que há cinco tipos de pontes, e que pontes modernas não são feitas nem de aço, ferro ou madeira

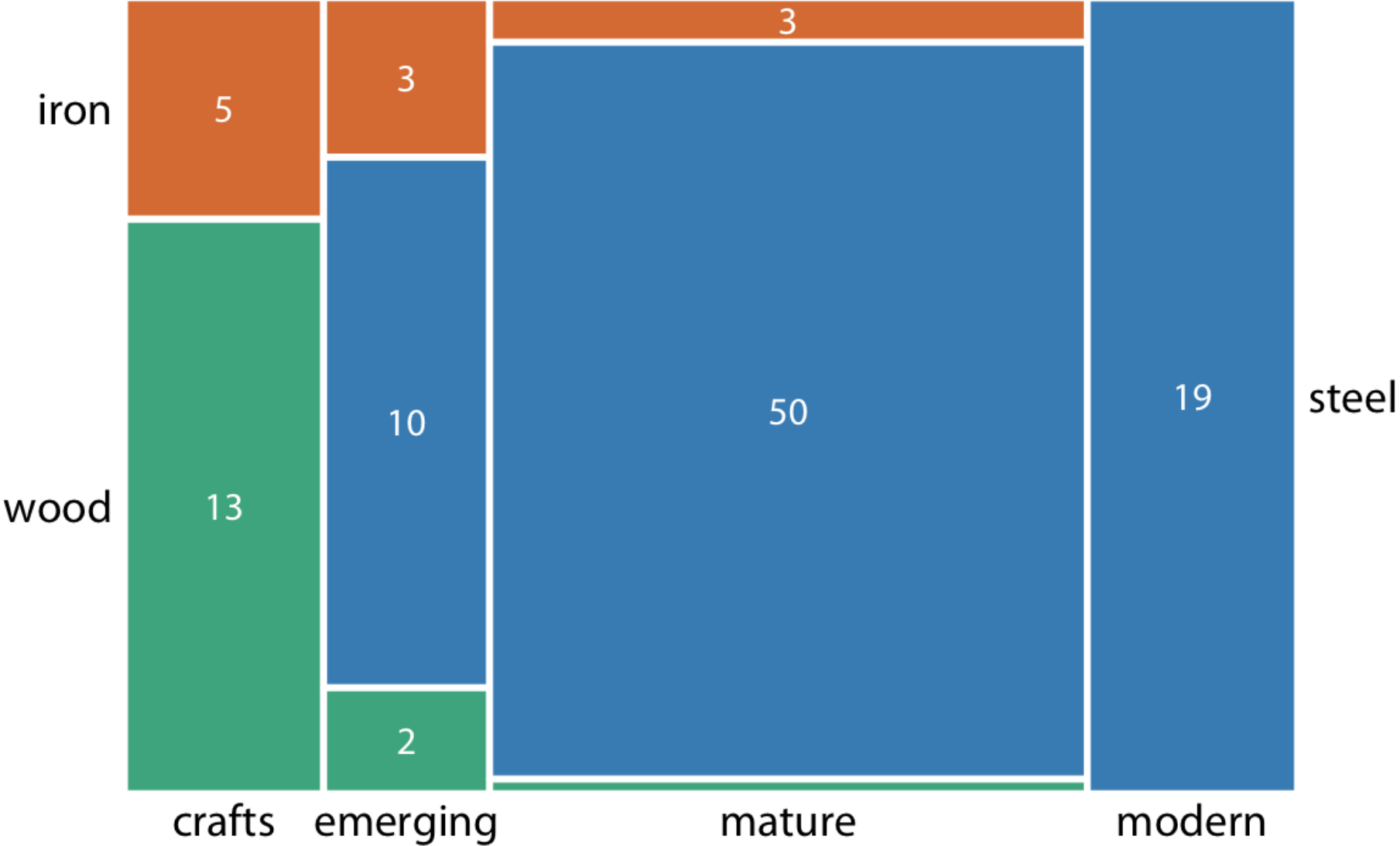


Adaptado de:
WILKE, C. O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019.

Gráfico de mosaico

- Semelhante à um gráfico de barras, mas tanto a altura quanto largura da barra podem variar
- Premissa obrigatória: todos grupos de variáveis categóricas apresentados devem cobrir todas as observações no conjunto de dados (nenhuma observação pode estar fora de alguma categoria)
- Construção:
 - Primeira variável categórica no eixo x, e subdividir o eixo proporcionalmente à quantidade de observações em cada subcategoria
 - Segunda variável categórica no eixo y, e subdividir o eixo proporcionalmente à quantidade de observações em cada subcategoria
- Áreas dos retângulos proporcionais à quantidade de observações para cada combinação das variáveis categóricas

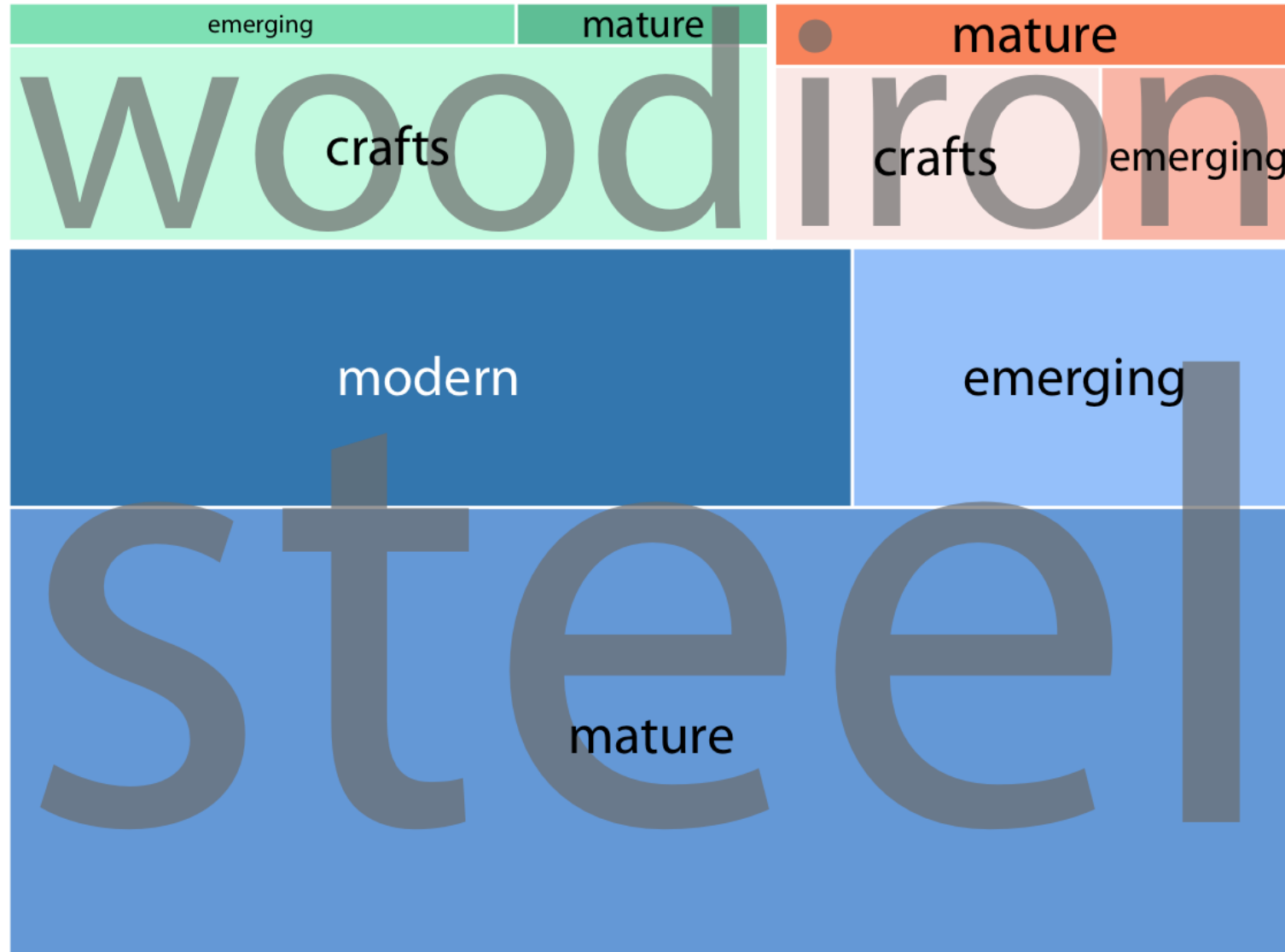
Gráfico de mosaico



Mapa de árvore

- Semelhante ao gráfico de mosaico: um retângulo maior é subdividido em retângulos menores cujas áreas representam a proporção de observações
- Diferença no método de construção:
 - Retângulos são recursivamente inseridos um dentro do outro, de maneira aninhada
- Possível realizar cada vez mais subdivisões internas, mas rapidamente o gráfico se torna poluído e confuso

Mapa de árvore



Comparativo de estratégias

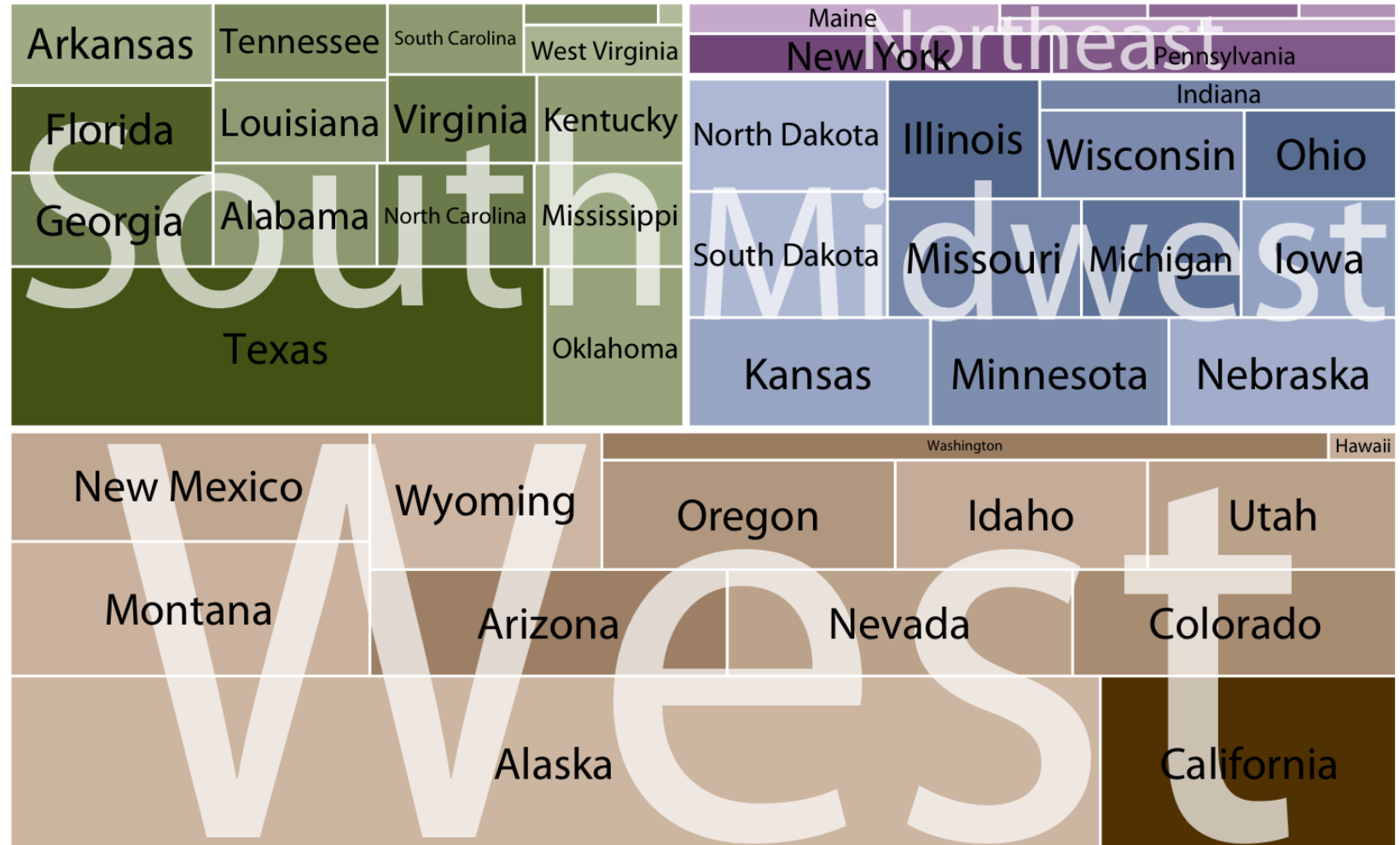
- Diferença de ênfase e aplicação
- Gráfico de mosaico:
 - Ênfase na evolução temporal
 - Premissa de que todas as proporções apresentadas podem ser identificadas pela combinação de duas ou mais variáveis categóricas dispostas de forma ortogonal
 - Premissa de que todas as combinações entre as variáveis são possíveis

Comparativo de estratégias

- Diferença de ênfase e aplicação
- Mapa de árvore:
 - Ênfase na proporção que cada categoria representa do total
 - Funcionam bem quando as combinações das variáveis categóricas nem sempre são possíveis ou informativas

Mapa de árvore

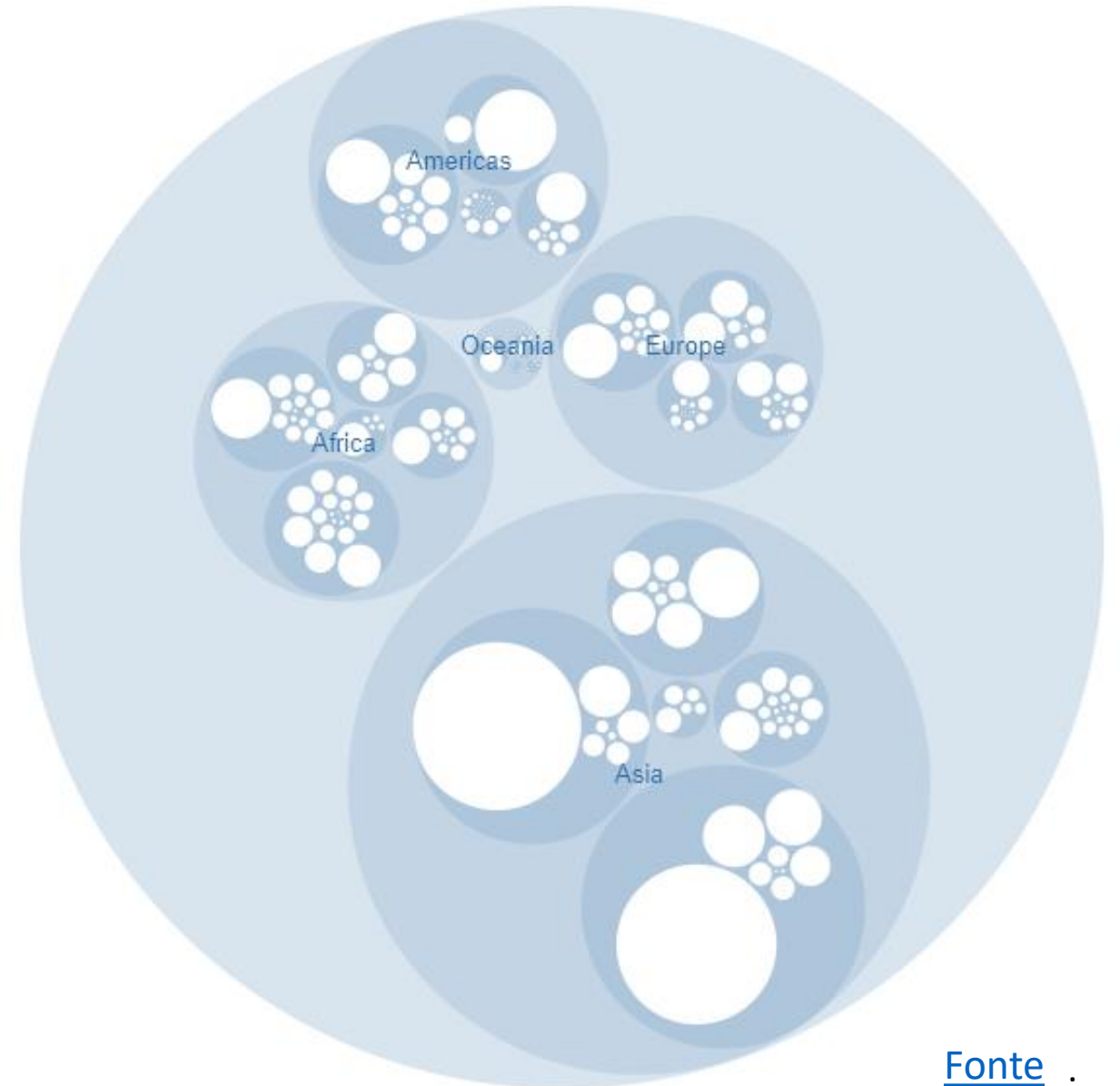
- Quatro regiões dos EUA e seus estados, representados proporcionalmente às suas áreas e cor proporcional à quantidade de habitantes
- Estados de uma região não possuem “combinações” ou “relação” com os estados de outras regiões



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

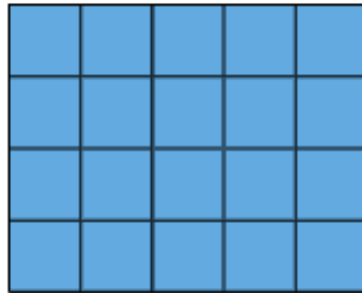
Mapa de árvore circular / empacotamento circular

- Equivalente ao mapa de árvore, mas áreas representadas de forma circular
- Não recomendado quando é necessário representar valores altamente precisos, pela dificuldade visual de interpretar uma área circular e correlacionar com um valor
- Vantagem: ressalta a organização de grupos e subgrupos e interpretação de relações hierárquicas

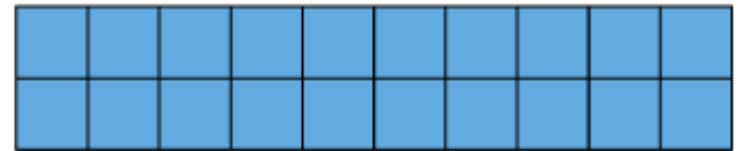


Comparativo de estratégias

- Principais limitações semelhantes às limitações dos gráficos de barras:
 - Dificuldade de comparação visual entre condições (linhas de base instáveis entre retângulos)
 - Retângulos de áreas iguais podem ter perímetros distintos, resultando em formas drasticamente diferentes
 - Retângulo A: perímetro de 18
 - Retângulo B: perímetro de 24
 - Ambos com área de 20



Rectangle A



Rectangle B

Comparativo de estratégias

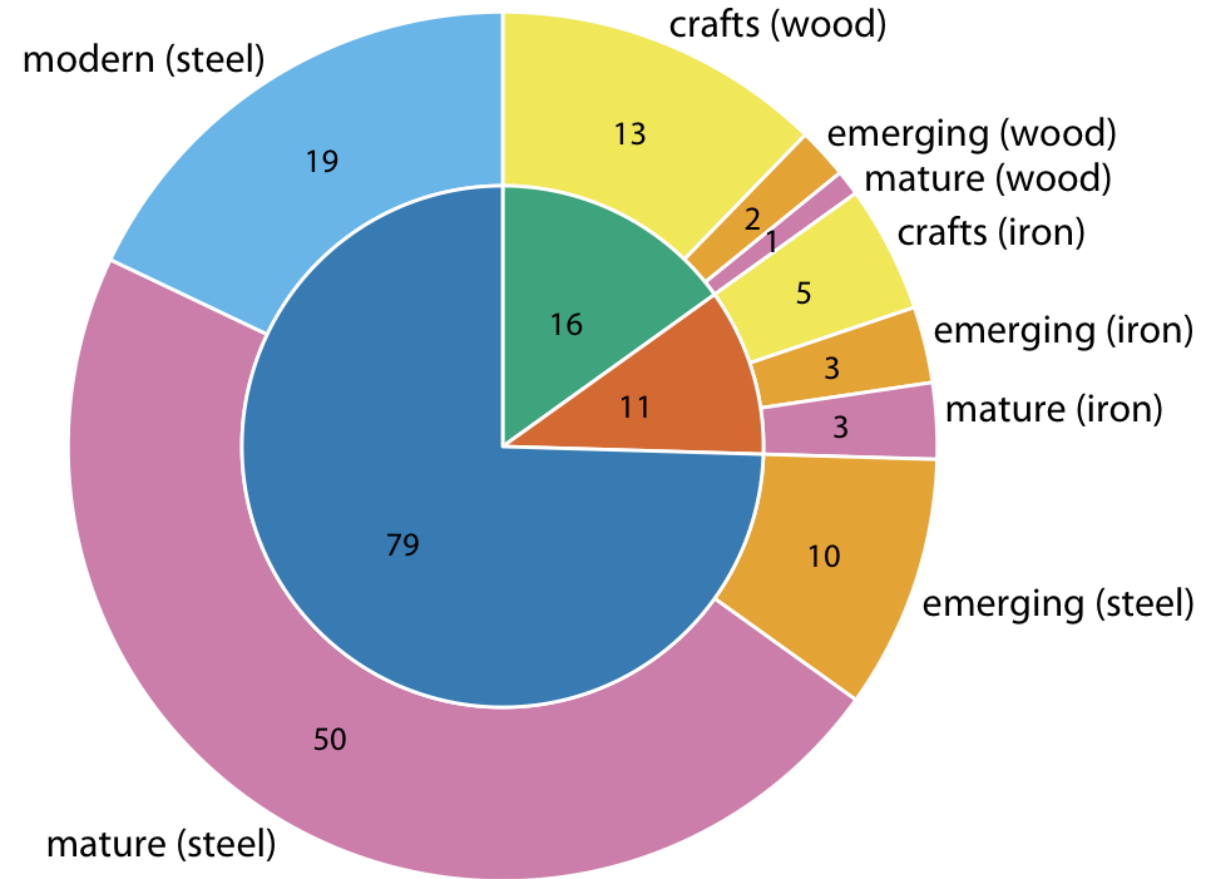
- Problema do formato dos retângulos não é tão simples de resolver pela própria natureza da representação de proporções aninhadas
- Sempre que possível, incluir as quantidades ou porcentagens dentro dos retângulos para auxiliar a interpretação

Gráficos de pizza aninhados

- Semelhante ao gráfico de mosaico e mapa de árvore: uso de área para representar valores
- Diferença no uso de sistema de coordenadas: cartesiana (mosaico e árvore) ou polar (pizza)
- Duas principais possibilidades de utilizar gráficos de pizza para representar probabilidades aninhadas

Gráficos de pizza aninhados

- Desenhar um gráfico de pizza com círculo interno e círculo externo
- Círculo interno demonstra a proporção da primeira variável categórica
- Círculo externo demonstra a proporção da segunda variável categórica em cada fatia do círculo interno

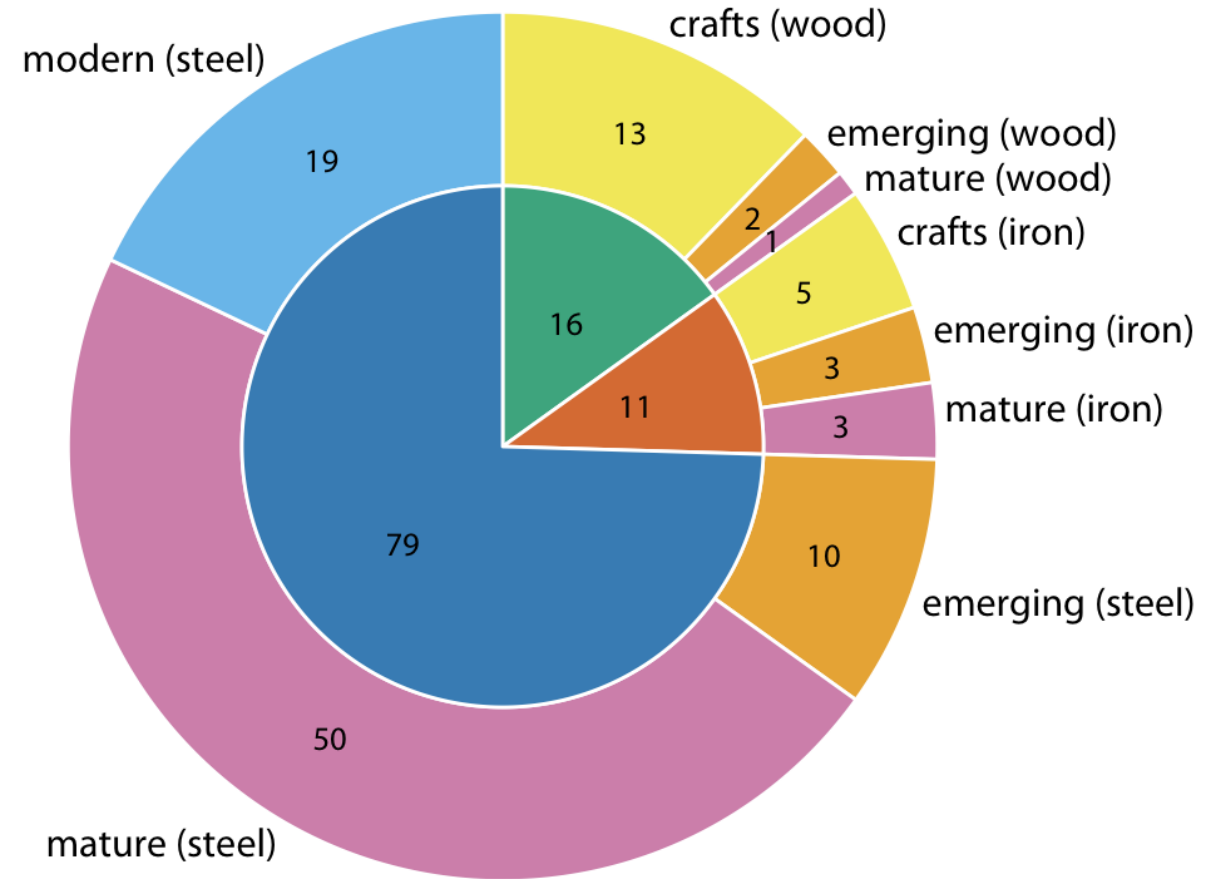


ugly

Gráficos de pizza aninhados

- Problemas:

- O uso de dois círculos ainda não deixa óbvio que há sobreposição entre as variáveis
- Pontes ainda sendo contabilizadas duas vezes: a soma dos valores resulta em 212 (dobro de 106 pontes)



ugly

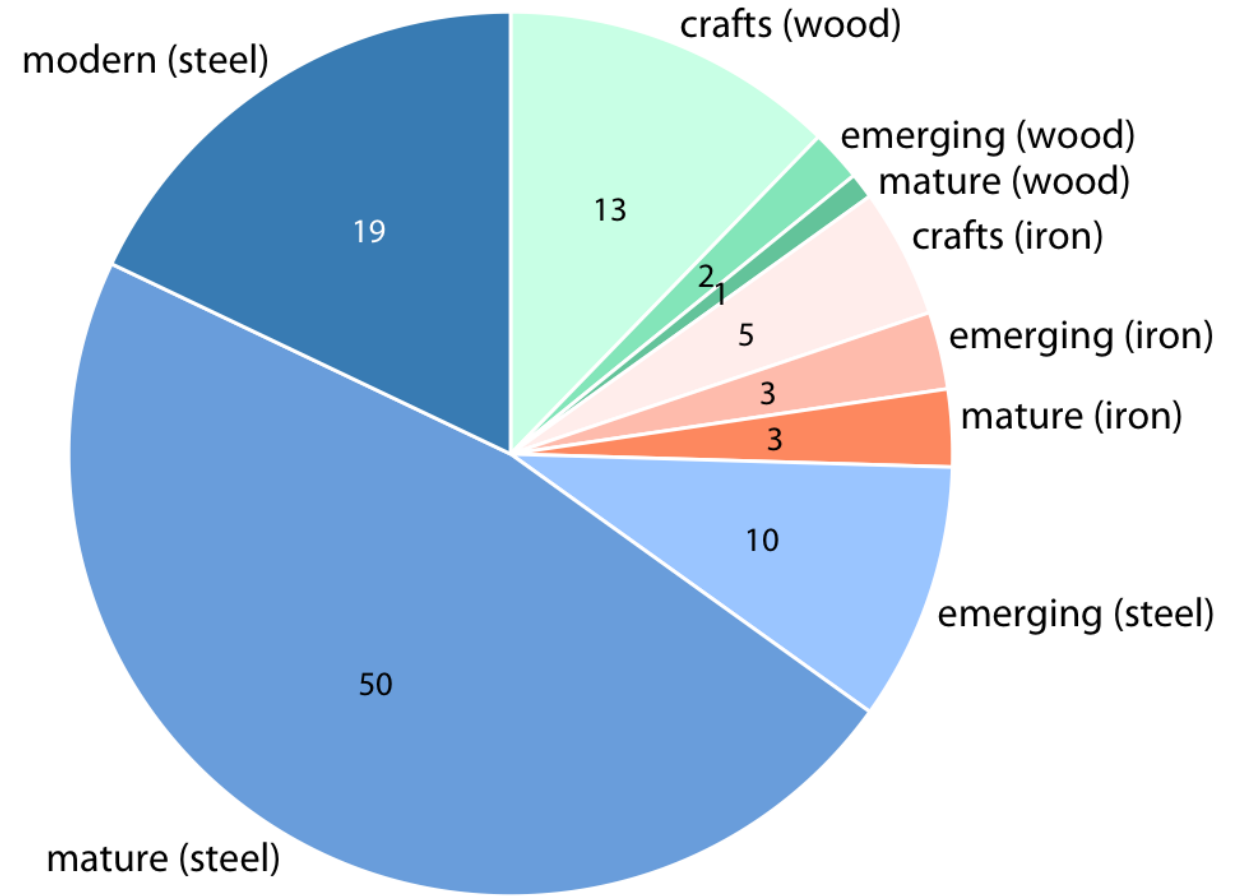
Gráfico de explosão solar (sunburst)

- Adaptação dos gráficos de pizza aninhados a partir do gráfico de rosca (donut)
- Cada nível da hierarquia é representado por um donut/anel:
 - Anel mais interno representando o topo dos dados hierárquicos
 - Anel mais externo representando o maior nível de subdivisão dentro da hierarquia



Gráficos de pizza aninhados

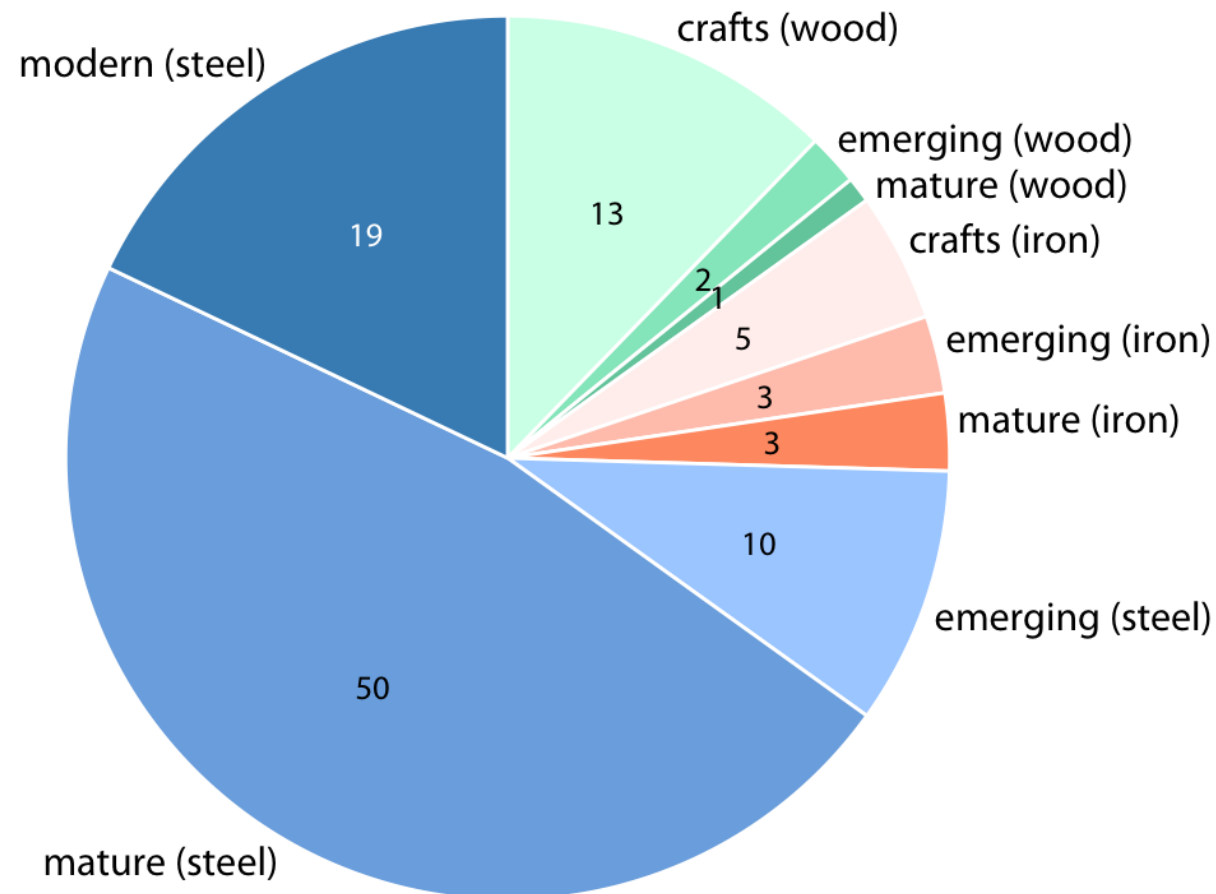
- Realizar uma divisão inicial das fatias de acordo com a primeira variável categórica (e codificar por matiz)
- Subdividir as fatias de acordo com a segunda variável categórica e utilizar diferenças de saturação de luminosidade dentro da mesma matiz:
 - Cores mais escuras: pontes mais recentes



Gráficos de pizza aninhados

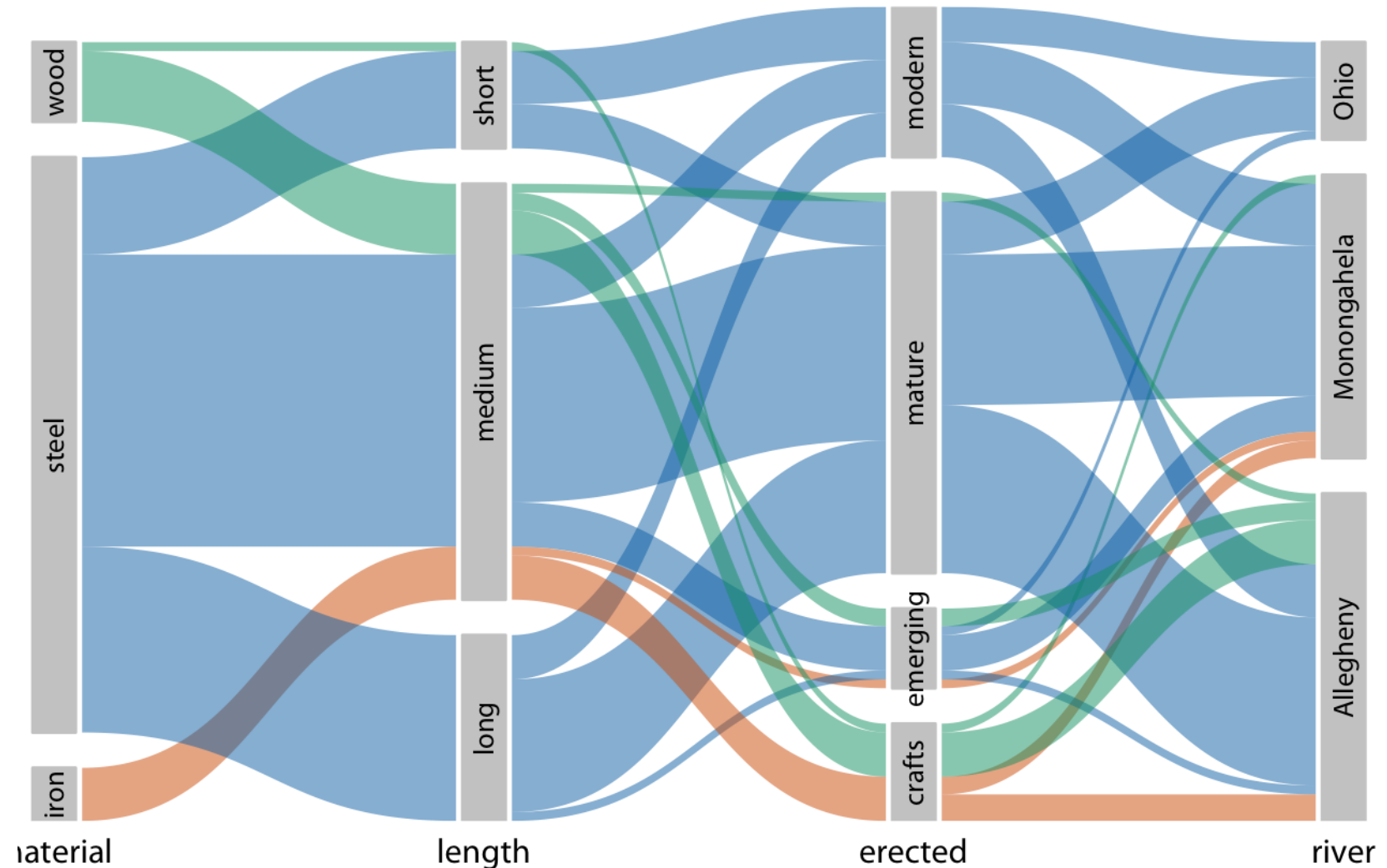
- Problemas:

- Desperdício de espaço vazio de impressão ao redor do gráfico (se comparado a um mapa de árvore)
- Impossibilidade de inserir as legendas dentro das fatias (o que auxiliaria na percepção visual de unidade)
- Algumas fatias podem se tornar muito pequenas e de difícil comparação
 - No gráfico de árvore, os retângulos correspondentes ainda tem tamanhos razoáveis



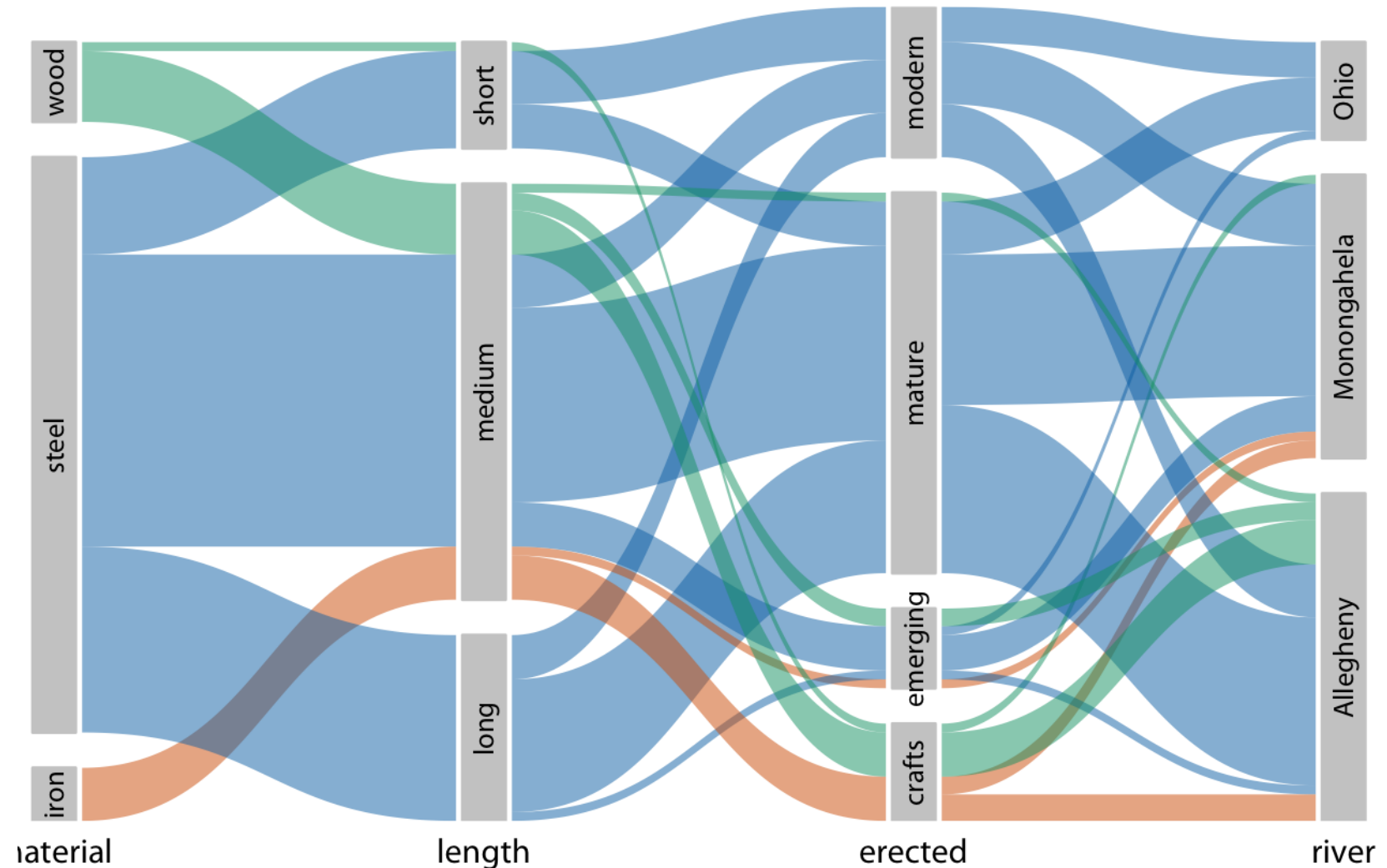
Gráficos de conjuntos paralelos (*parallel sets*)

- Aumento de variáveis categóricas dificulta a interpretação de gráficos de mosaico, mapas de árvore e gráficos de pizza aninhados
- Gráfico de conjuntos paralelos: como o conjunto de dados se subdivide de acordo com diferentes variáveis categóricas
- Bandas entre as variáveis para mostrar como os subgrupos estão relacionados



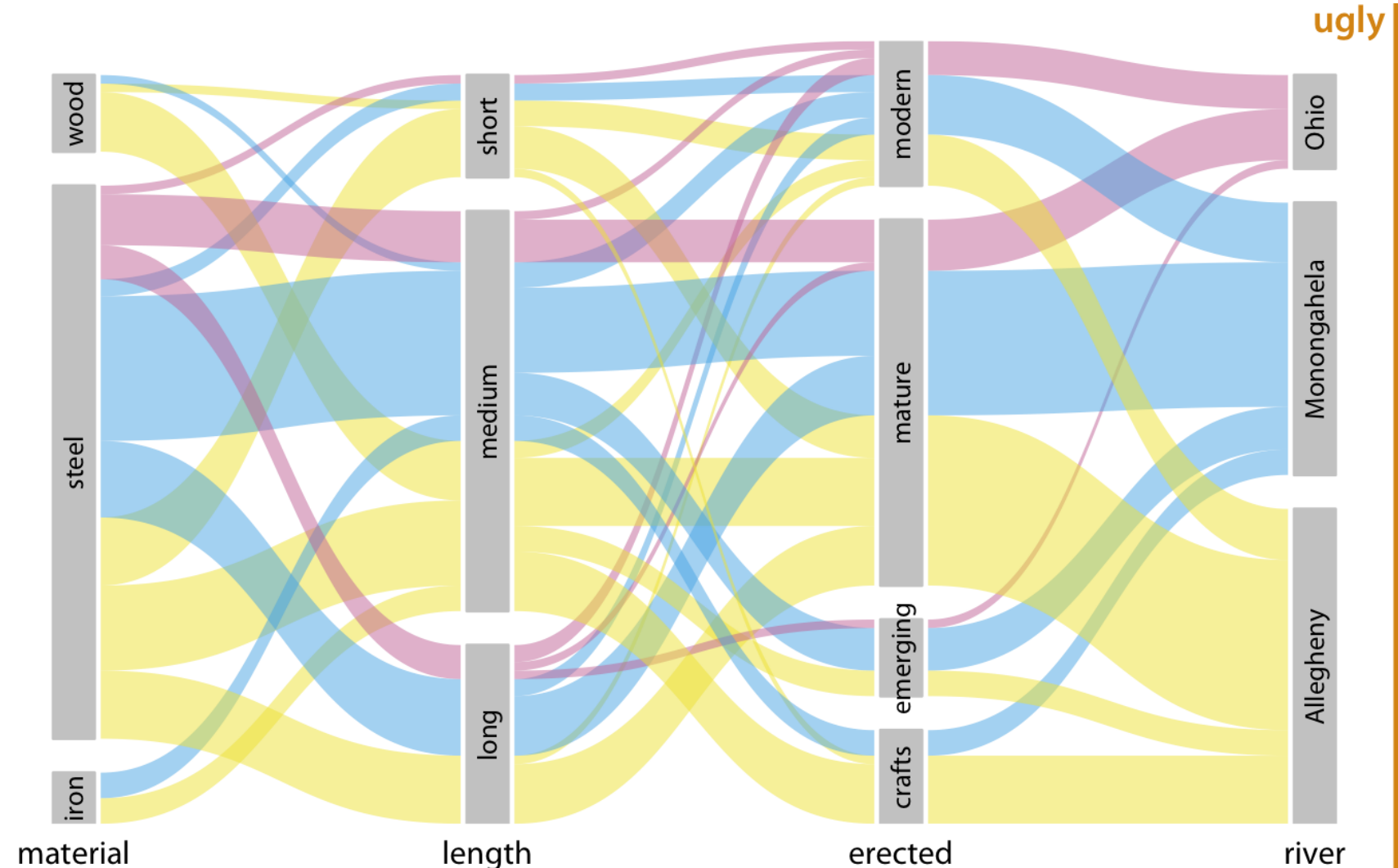
Gráficos de conjuntos paralelos (*parallel sets*)

- Pontes classificadas de acordo com:
 - Material
 - Comprimento
 - Época de construção
 - Rio em que estão localizadas
- Bandas coloridas de acordo com o material utilizado



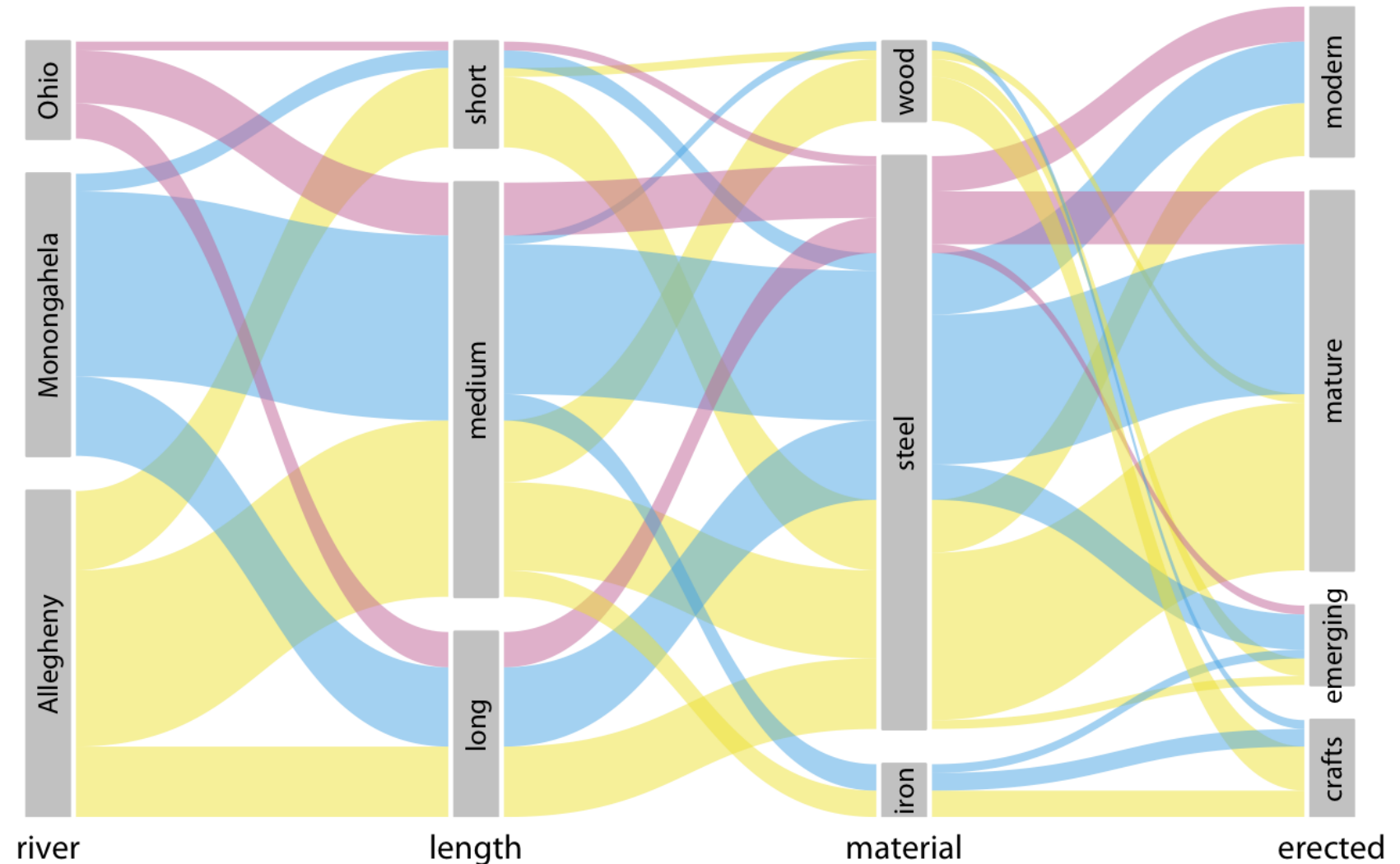
Gráficos de conjuntos paralelos (*parallel sets*)

- Visualização modificada se for colorida de acordo com outro critério
- Cor de acordo com o rio onde as pontes estão localizadas
- Praticamente qualquer tipo de ponte pode ser encontrada nos três rios



Gráficos de conjuntos paralelos (*parallel sets*)

- Recomendação: critério utilizado para colorir o gráfico deve estar à esquerda (leitura da esquerda para direita)
- Reorganizar as subdivisões para evitar muito emaranhamento das bandas



Gráficos de conjuntos paralelos (*parallel sets*)

- Distribuição dos passageiros do Titanic de acordo com diferentes categorias
- Ressalta a diferença de proporções de homens e mulheres entre os sobreviventes

