CP321 Data Visualization

- Visualize Proportions

Jiashu (Jessie) Zhao

- Visualizing proportions
  - *Pie Chart, Bars, Stacked Bars*
  - Compare multiple sets of proportions: *Multiple Pie Charts, Grouped Bars. Stacked Bars, Stacked Densities*
  - Nested proportions : *nested pies, mosaic plots, treemaps, or parallel sets*.
- Design: The principle of proportional ink

Visualizing proportions

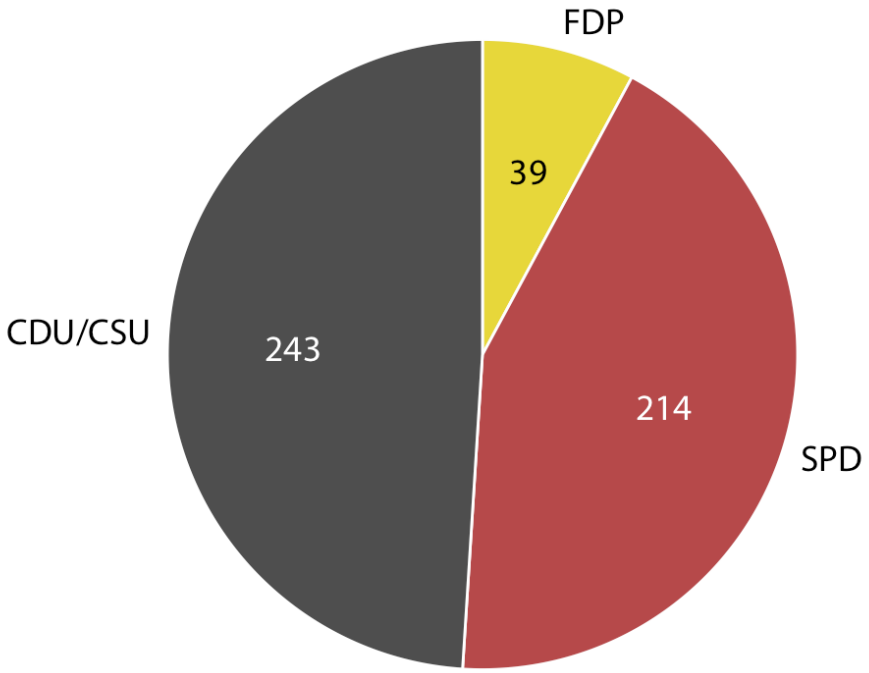
- Show how some group, entity, or amount breaks down into individual pieces that each represent a *proportion* of the whole

- There is no single ideal visualization that always works.

- You always need to pick the visualization that best fits your specific dataset and that highlights the key data features you want to show.

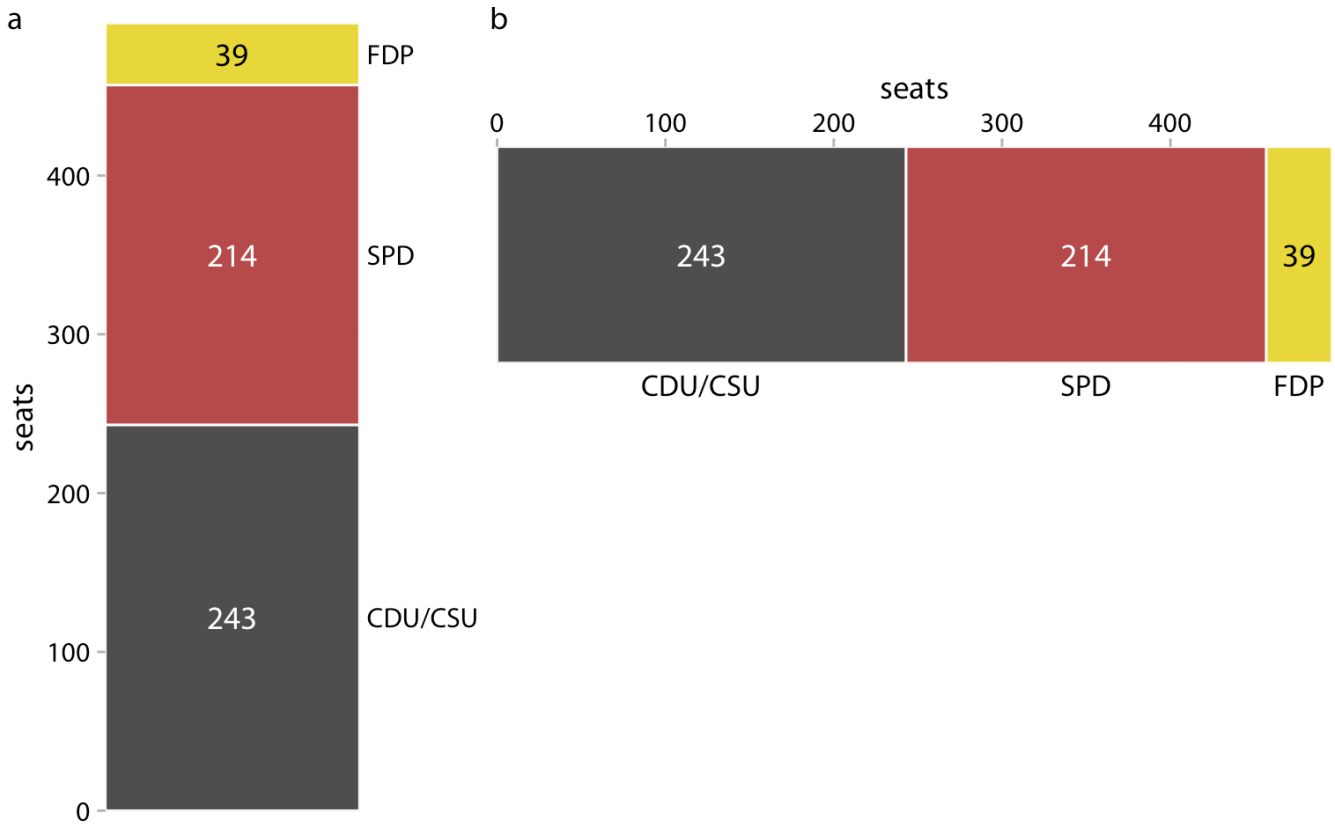
Pie chart

- A pie chart breaks a circle into slices such that the area of each slice is proportional to the fraction of the total it represents.

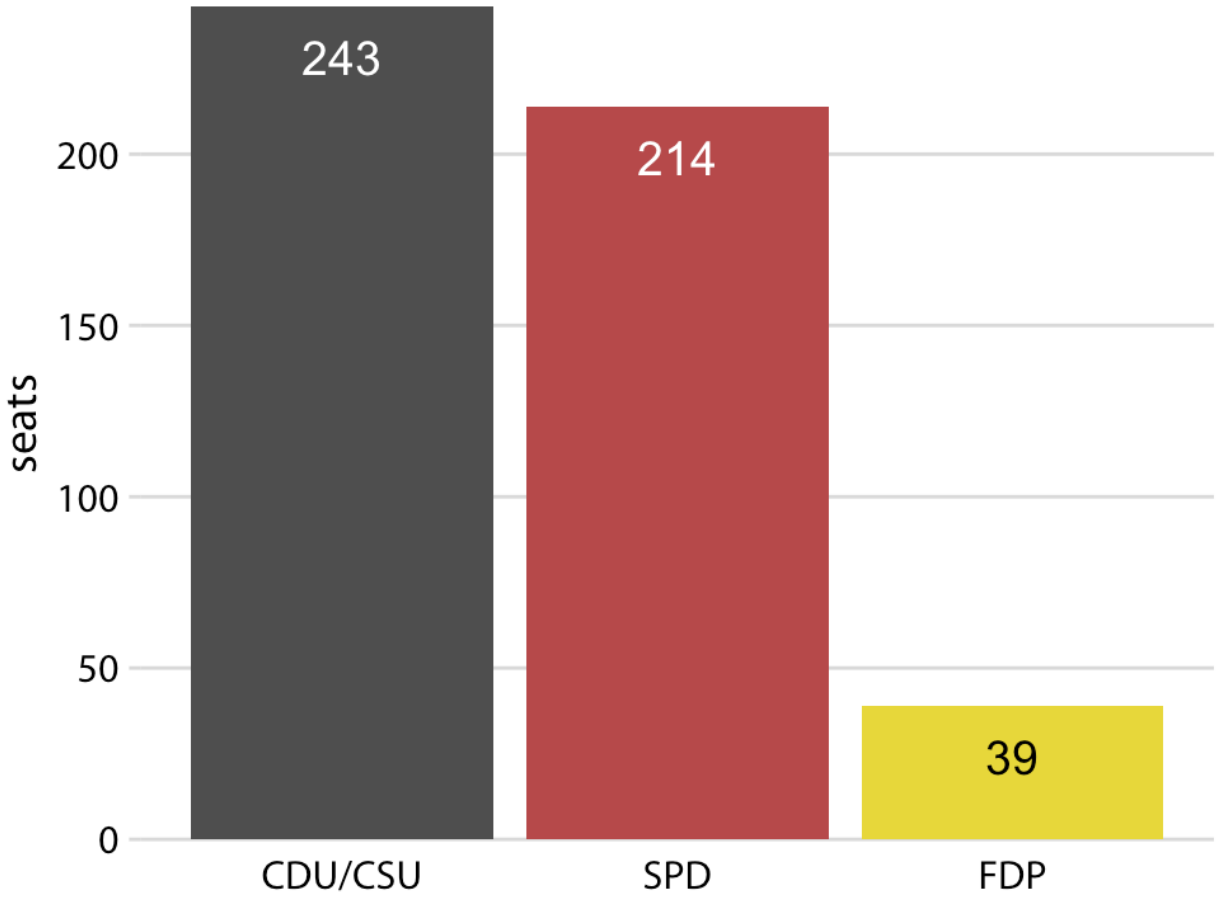


Party composition of the 8th German Bundestag, 1976–1980

Stacked bar chart for proportion



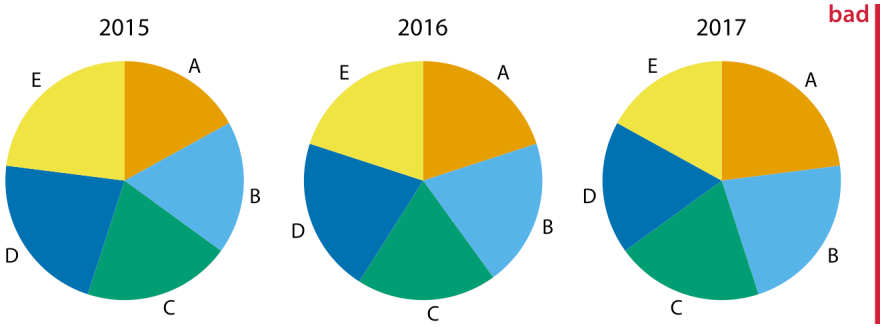
Side by side bars for proportion



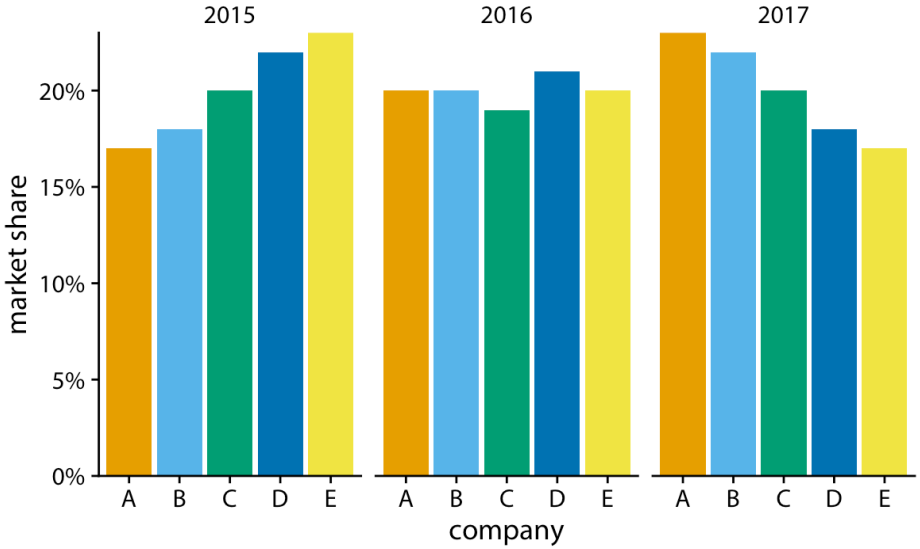
	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✔	✔	✖
Allows easy visual comparison of the relative proportions	✖	✖	✔
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✔	✖	✖
Looks visually appealing even for very small datasets	✔	✖	✔
Works well when the whole is broken into many pieces	✖	✖	✔
Works well for the visualization of many sets of proportions or time series of proportions	✖	✔	✖

Compare multiple sets of proportions

- Multiple Pie Chart

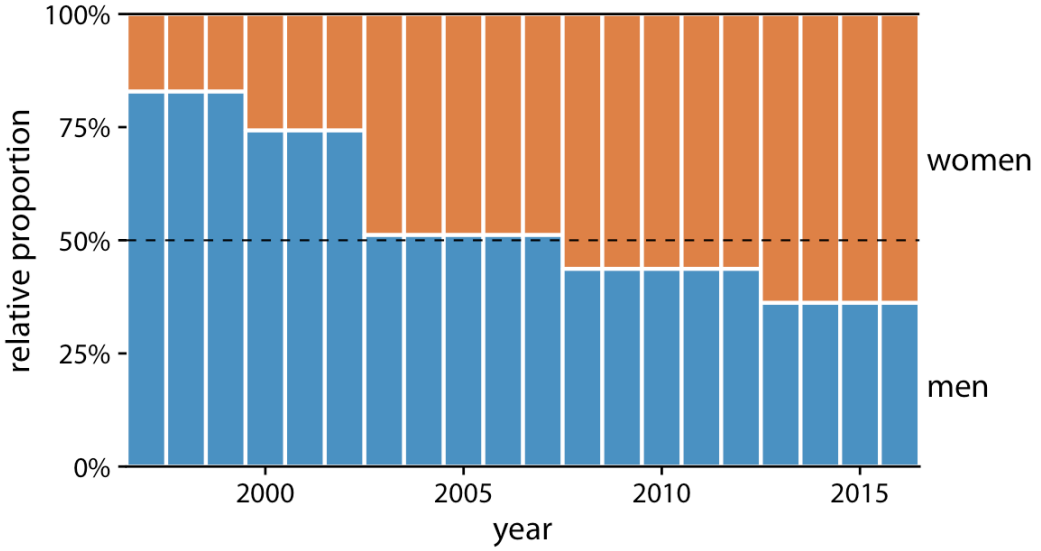


- Grouped Bar

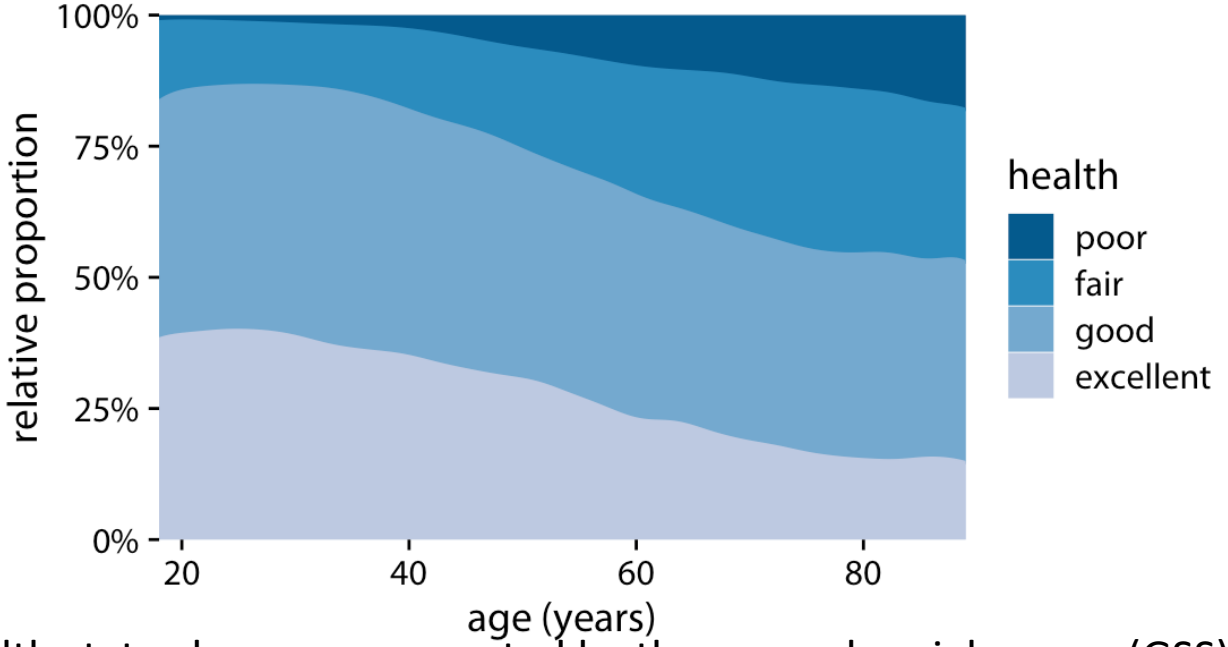


Market share of five companies, A–E

- Stacked bar
- Stacked density

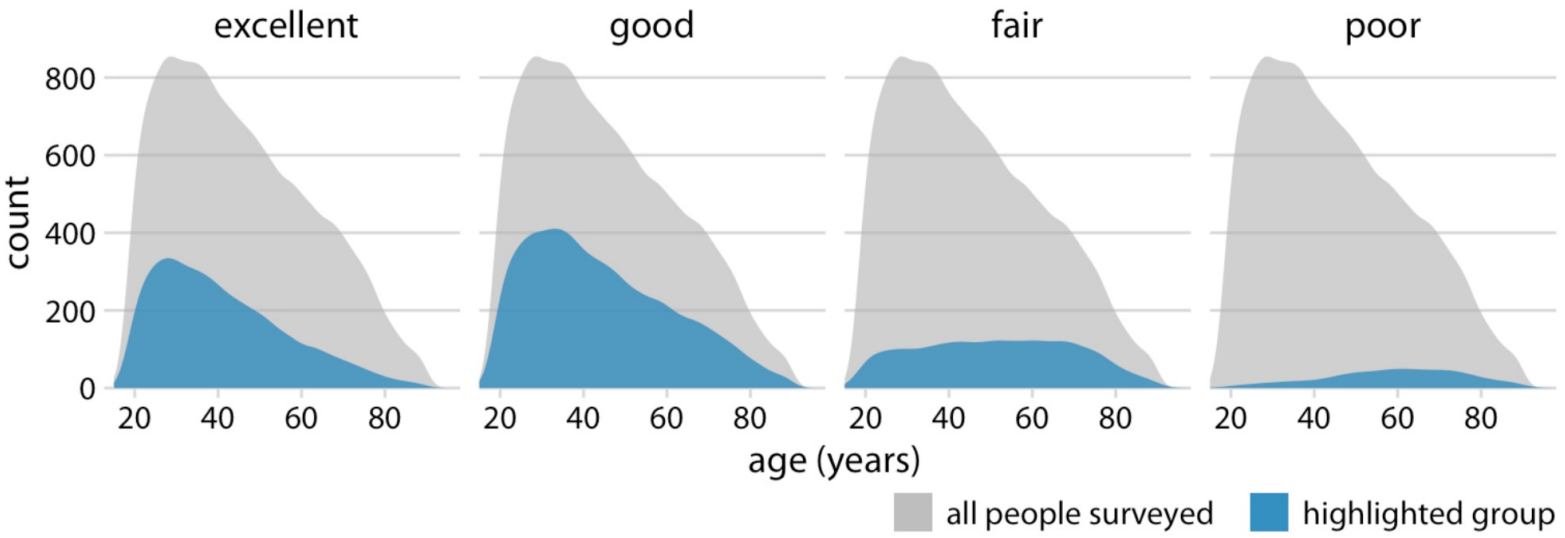


Change in the gender composition of the Rwandan parliament over time, 1997 to 2016.



Health status by age, as reported by the general social survey (GSS).

Example: Visualizing proportions separately as parts of the total

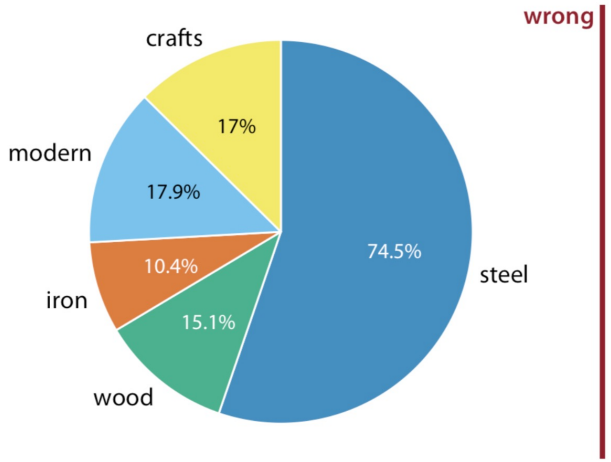


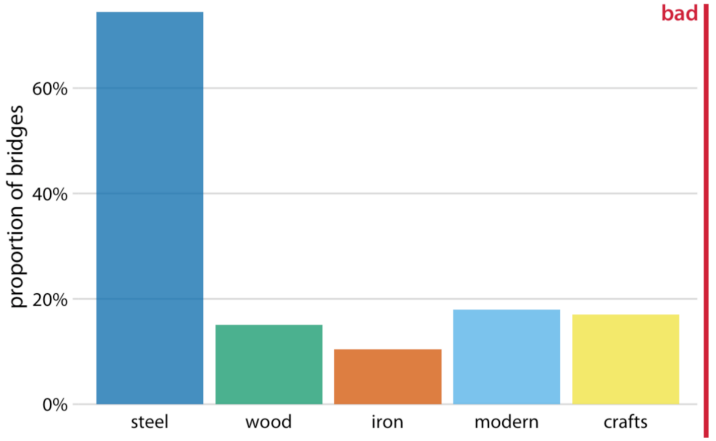
Visualizing nested proportions

- Multiple grouping variables
- Break down a dataset by multiple categorical variables at once

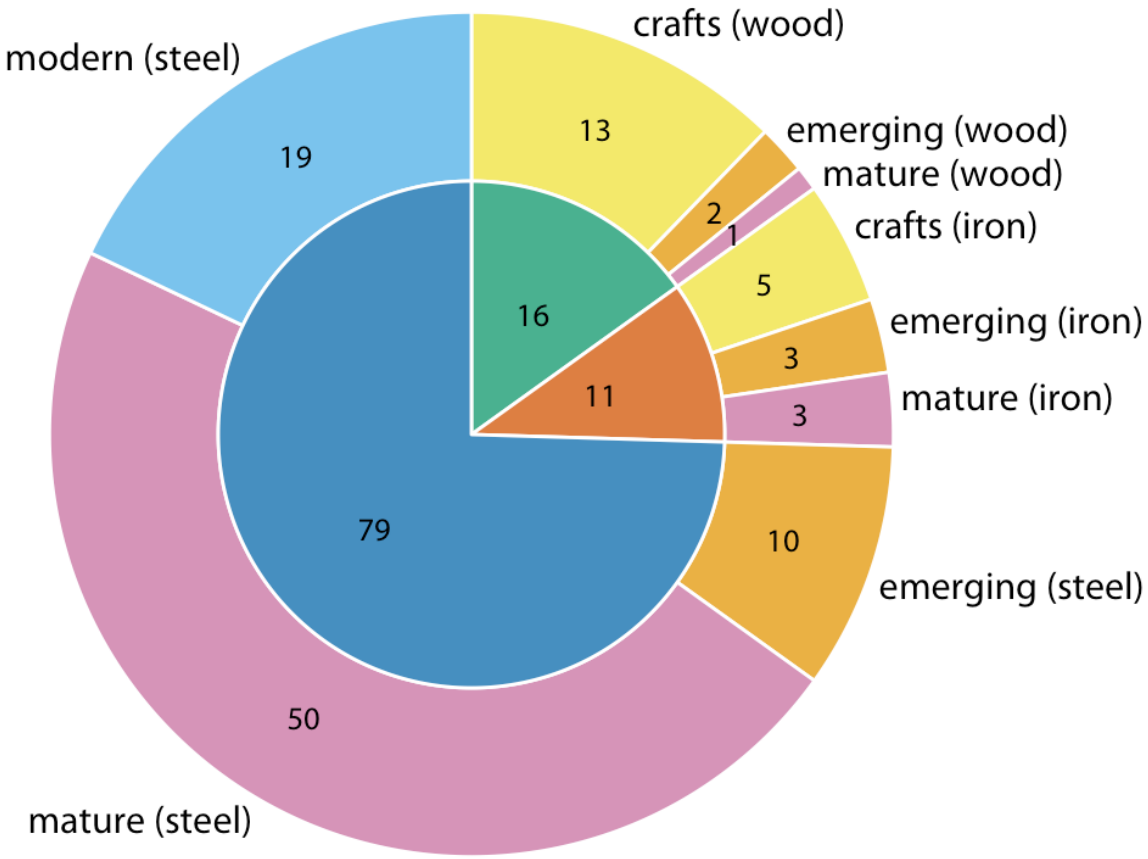
- Example: There are 106 bridges in Pittsburgh. This dataset contains various pieces of information about the bridges, such as the material from which they are constructed (steel, iron, or wood) and the year when they were erected (crafts, emerging, modern).

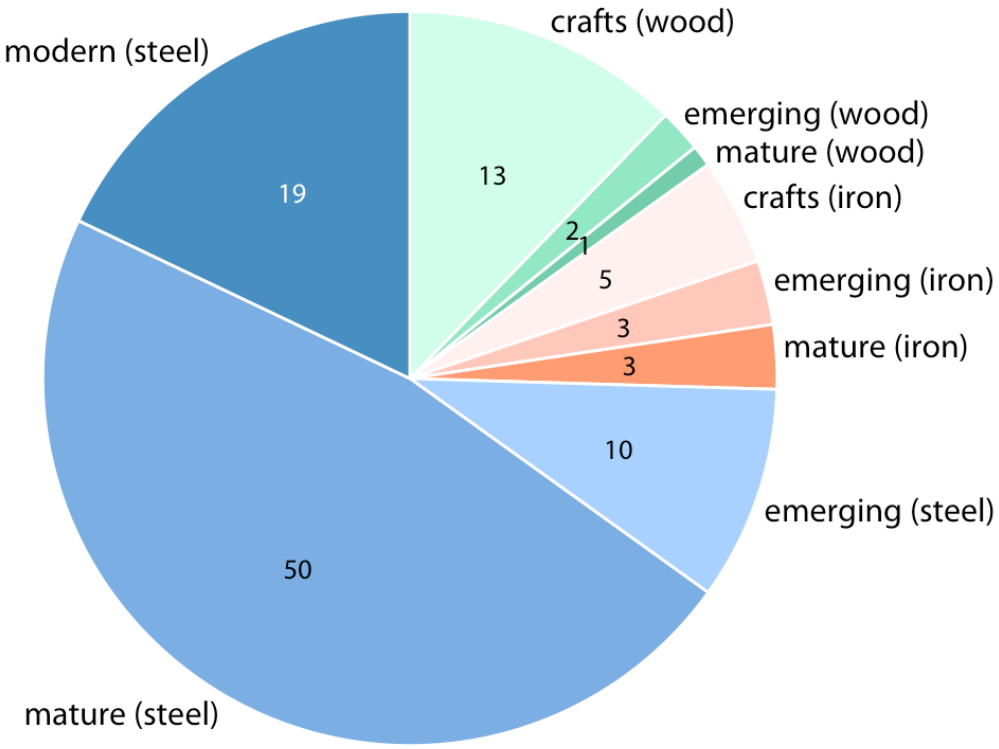
- Can we directly visualize proportions of all these categories with a pie chart or bar chart?





Nested Pies

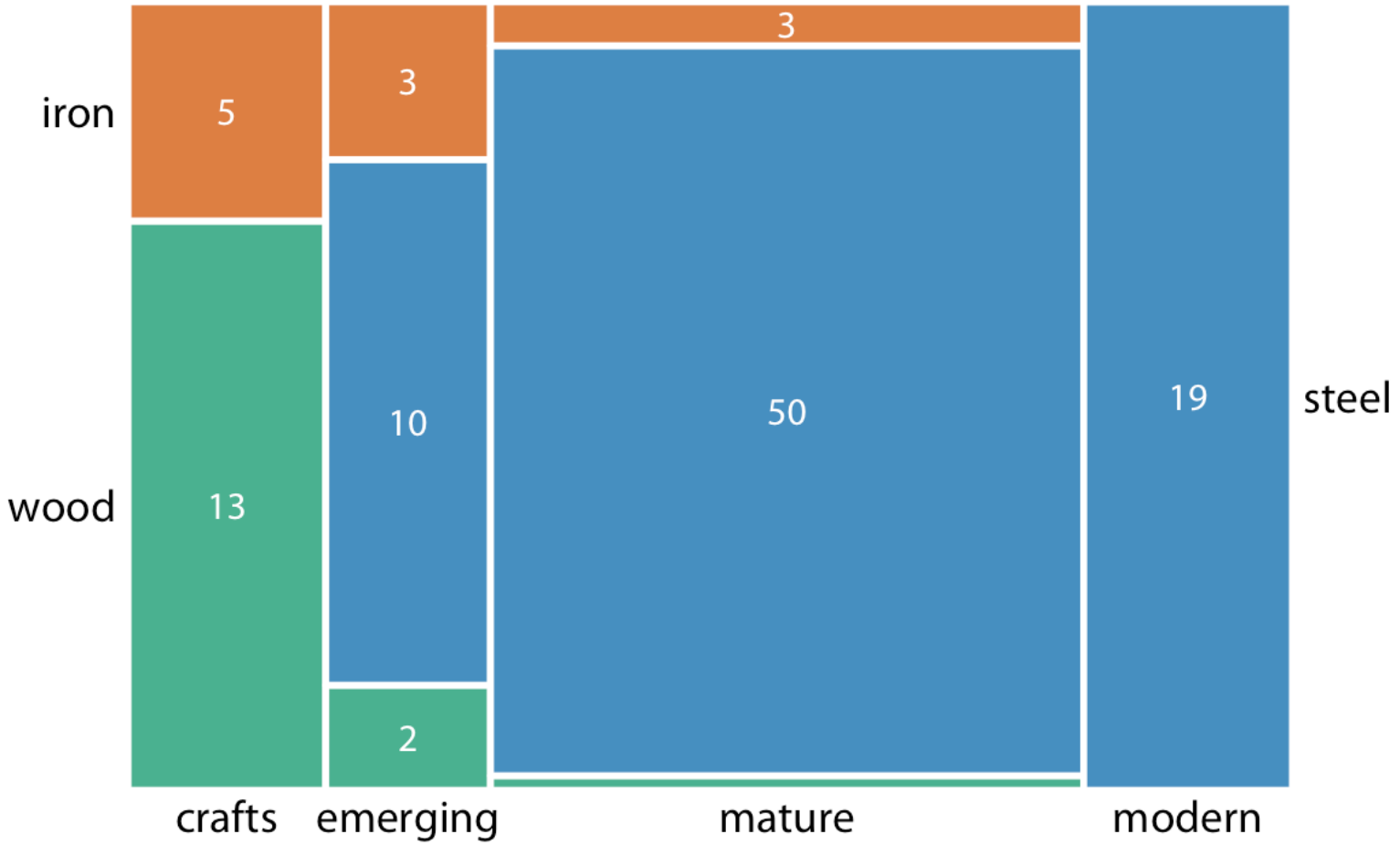

crafts (wood)
13
emerging (wood)
2
mature (wood)
1
crafts (iron)
5
emerging (iron)
3
mature (iron)
3
emerging (steel)
10
mature (steel)
50
modern (steel)
19
16
11
79


crafts (wood)
13
emerging (wood)
2
mature (wood)
1
crafts (iron)
5
emerging (iron)
3
mature (iron)
3
emerging (steel)
10
mature (steel)
50
modern (steel)
19

Mosaic plot

- Whenever we have categories that overlap, it is best to show clearly how they relate to each other.

- We begin by placing one categorical variable along the *x* axis and subdivide the *x* axis by the relative proportions that make up the categories of the y variable



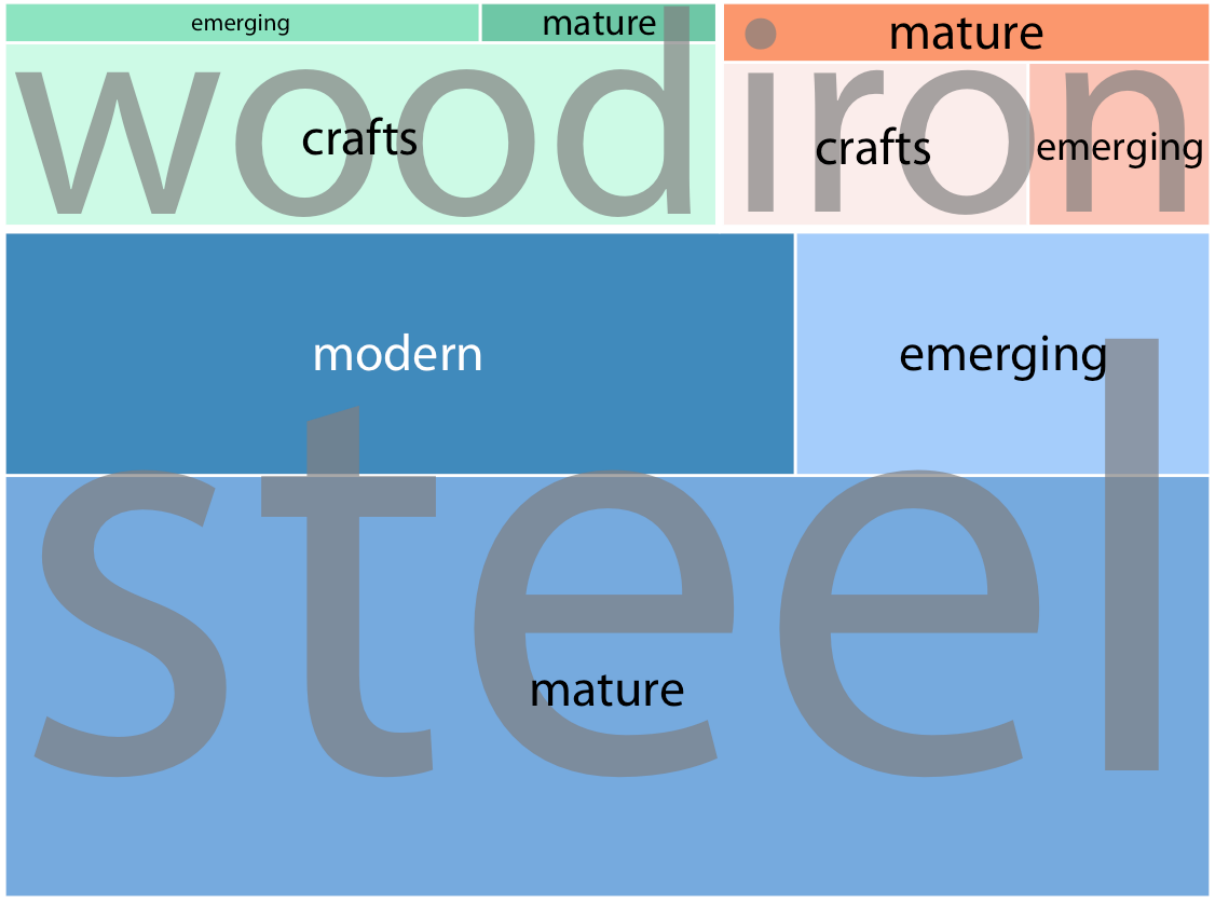
Looks similar to a stacked bar plot. What is the major difference?

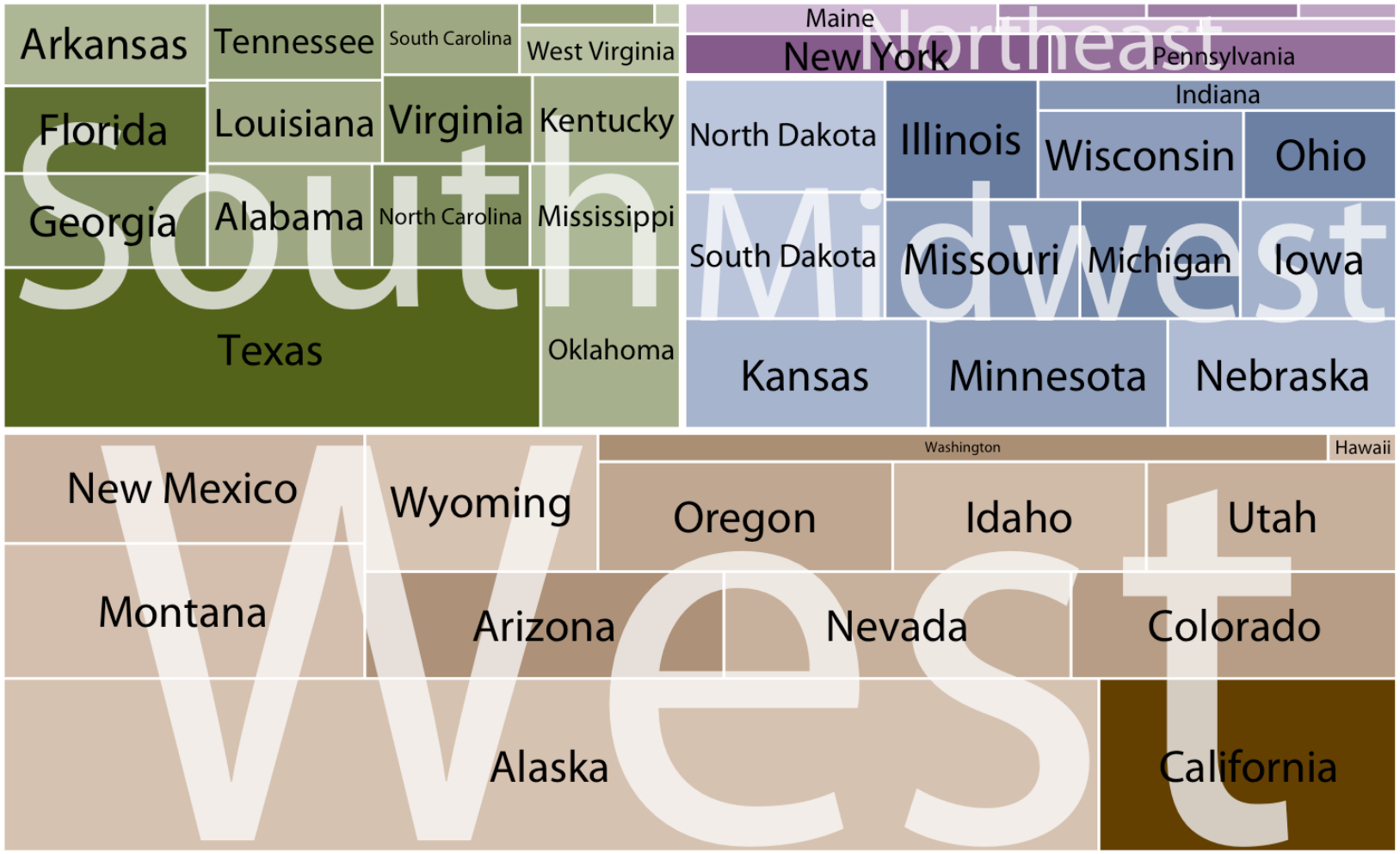
Treemap

- A series of nested rectangles of sizes proportional to the corresponding data value.

- A large rectangle represents a branch of a data tree, and it is subdivided into smaller rectangles that represent the size of each node within that branch.

- We recursively nest rectangles inside each other


emerging
mature
wood
crafts
mature
iron
crafts
emerging
modern
emerging
steel
mature

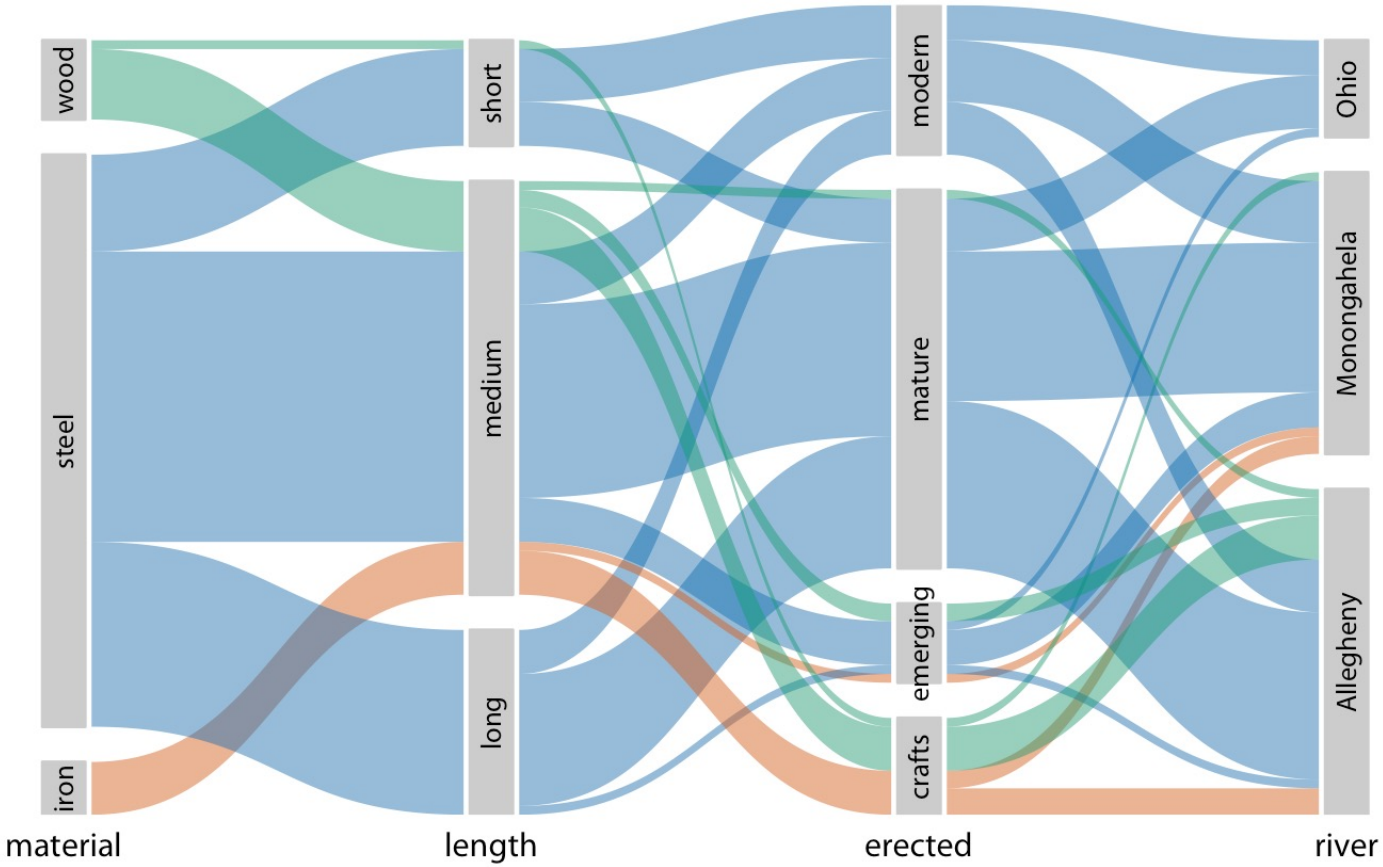

Arkansas
Tennessee
South Carolina
West Virginia
Florida
Louisiana
Virginia
Kentucky
Georgia
Alabama
North Carolina
Mississippi
South
Texas
Oklahoma
Maine
New York
Northeast
Pennsylvania
Indiana
North Dakota
Illinois
Wisconsin
Ohio
South Dakota
Missouri
Michigan
Iowa
Midwest
Kansas
Minnesota
Nebraska
Washington
Hawaii
New Mexico
Wyoming
Oregon
Idaho
Utah
Montana
Arizona
Nevada
Colorado
Alaska
California
West

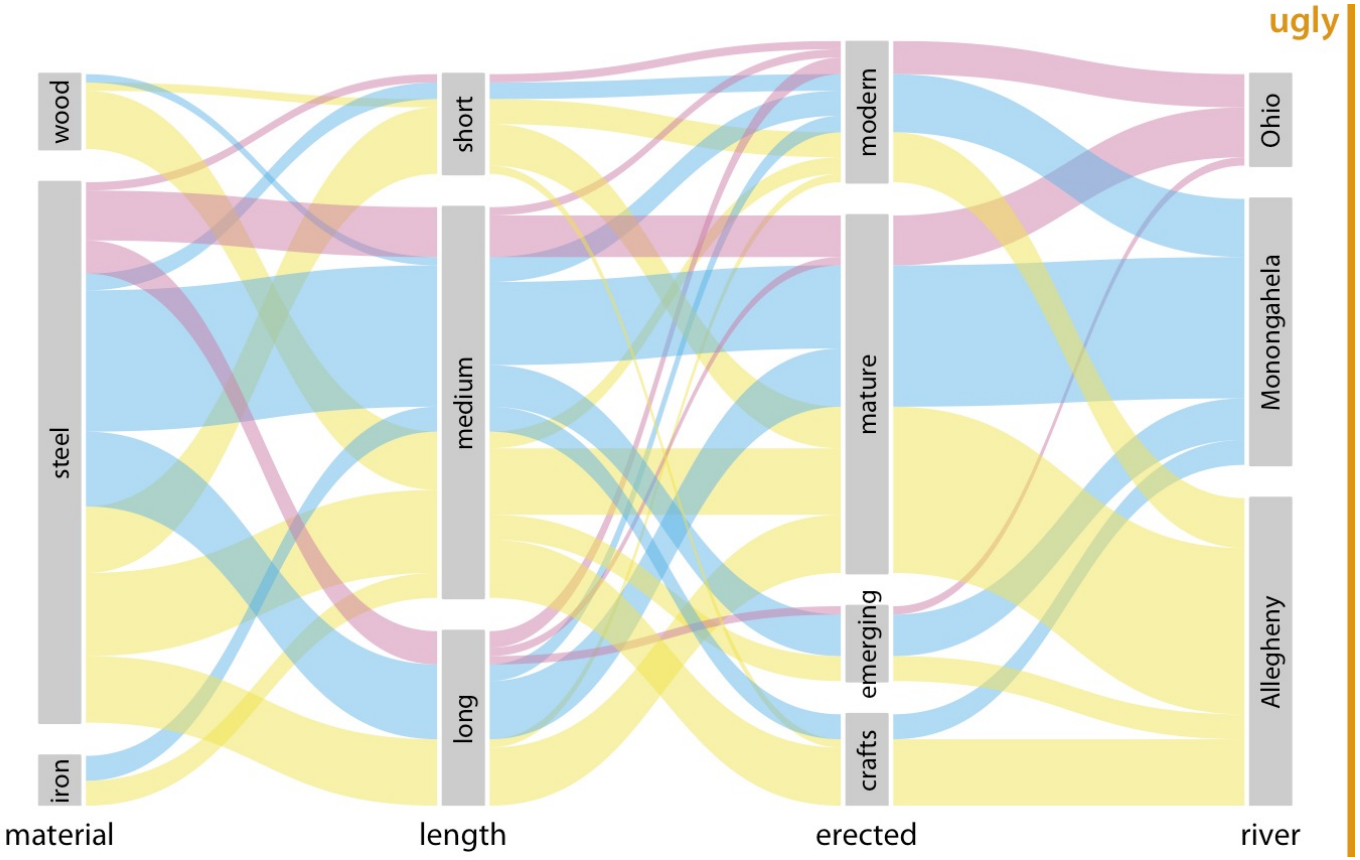
- Here, the mosaic plot emphasizes the temporal evolution in building-material use from the crafts era to the modern era, whereas the treemap emphasizes the total number of steel, iron, and wood bridges.

- More generally, in Mosaic plot, every combination of the categories are possible. There is no such requirement for treemap.

Parallel sets

- Proportions described by more than two categorical variables
- We show how the total dataset breaks down by each individual categorical variable
- Then we draw shaded bands that show how the subgroups relate to each other

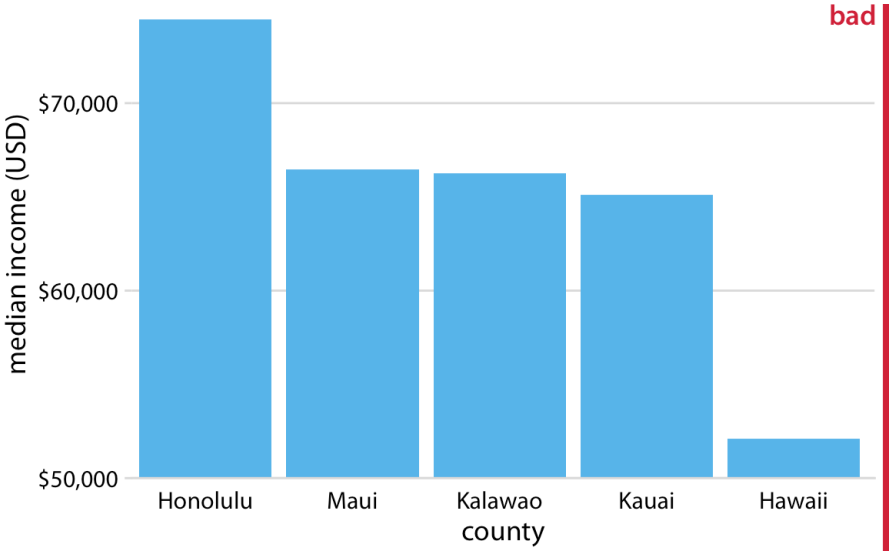

wood
steel
iron
short
medium
long
modern
mature
emerging
crafts
Ohio
Monongahela
Allegheny
material
length
erected
river

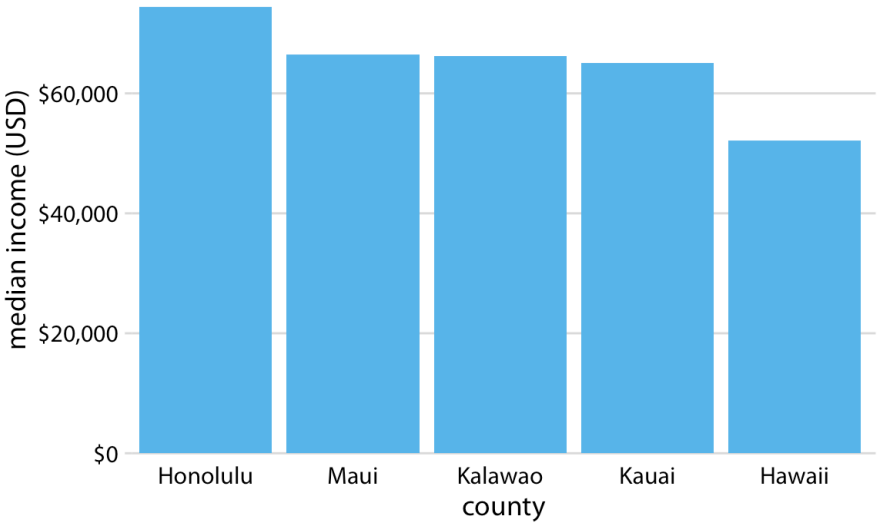

ugly
wood
steel
iron
short
medium
long
modern
mature
emerging
crafts
Ohio
Monongahela
Allegheny
material
length
erected
river

Design: The principle of proportional ink

- The *principle of proportional ink*(Bergstrom and West 2016): The sizes of shaded areas in a visualization need to be proportional to the data values they represent.

- When we use graphical elements such as bars, rectangles, shaded areas of arbitrary shape, or any other elements that have a clear visual extent





Bars should always start at 0.

- Visualize negative value