CSCI 491: Data Visualization
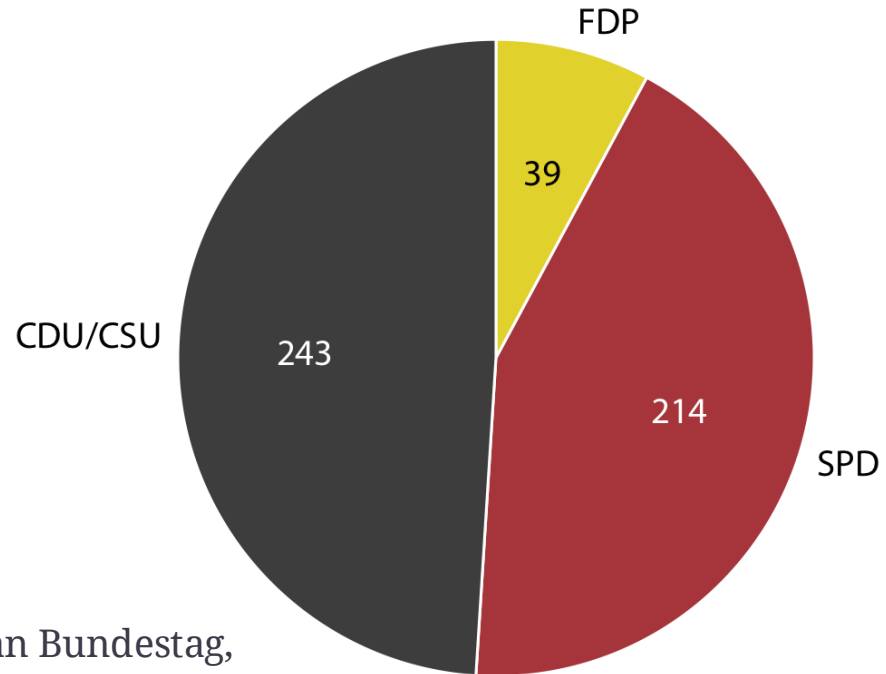
9- Visualizing Proportions

Proportions

- We often want to show how some group, entity, or amount breaks down into pieces representing a proportion of the whole.

- Common examples include the proportions of men and women in a group of people, the percentages of people voting for different political parties in an election, or companies' market shares.

- Visualizing proportions can be challenging, particularly when the whole is broken into many different pieces or when we want to see changes in proportions over time or across conditions.

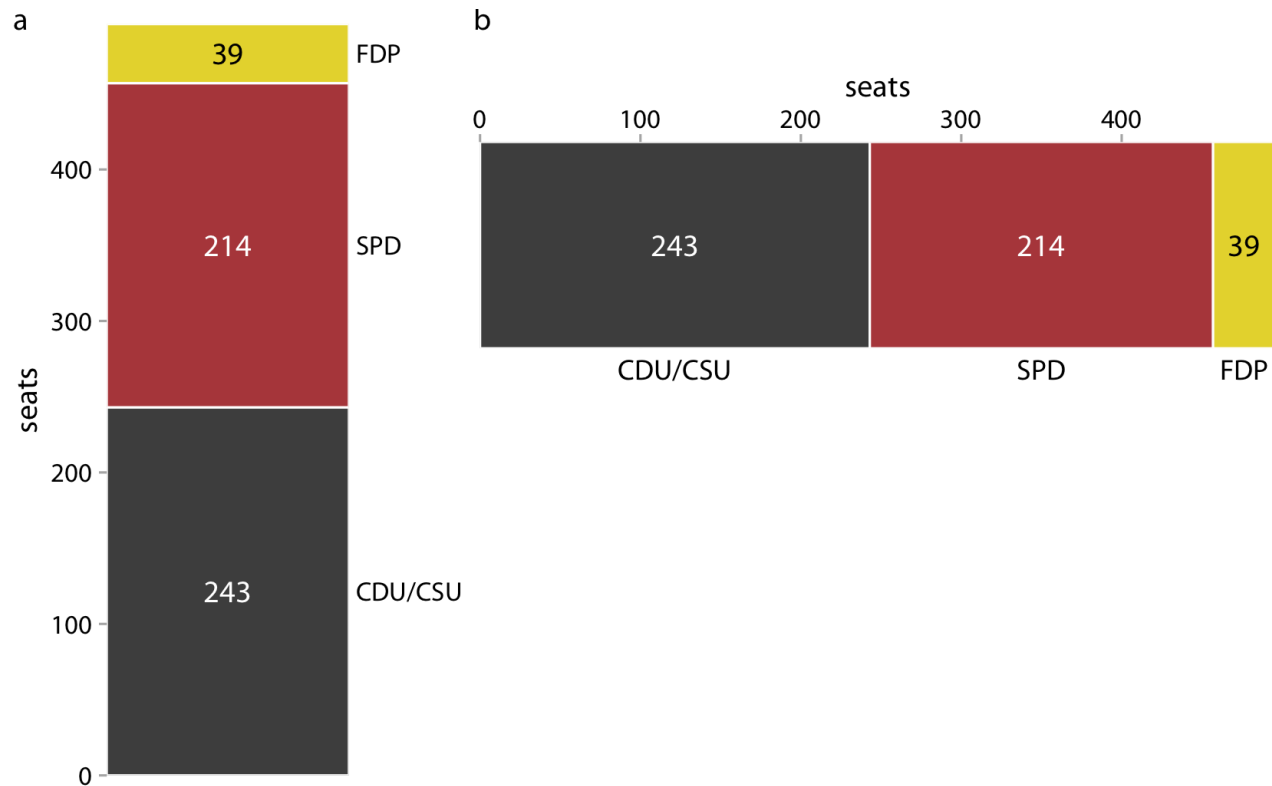- There is no single ideal visualization that always works.

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 eighth German Bundestag, 1976–1980, visualized as a pie chart.
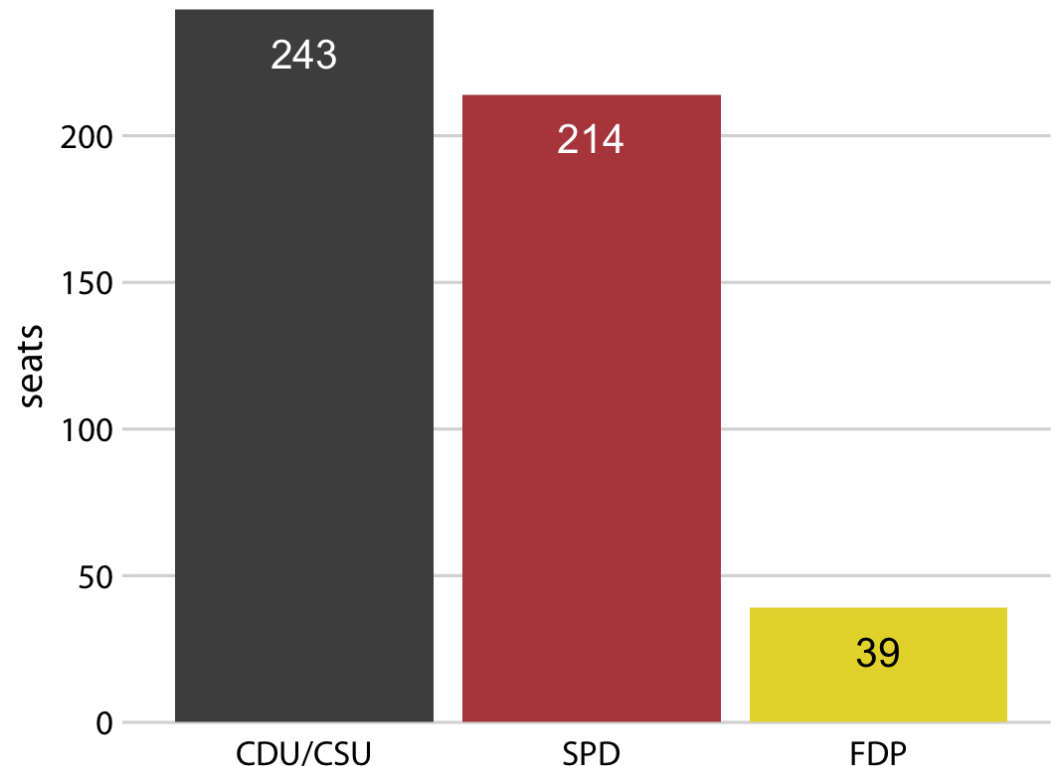
stacked bars

- The same procedure can be performed on a rectangle, and the result is a stacked bar chart.

Side-by-side Bars

- This visualization makes it easier to perform a direct comparison of the three groups
- In a side-by-side bar plot the relationship of each bar to the total is not visually obvious
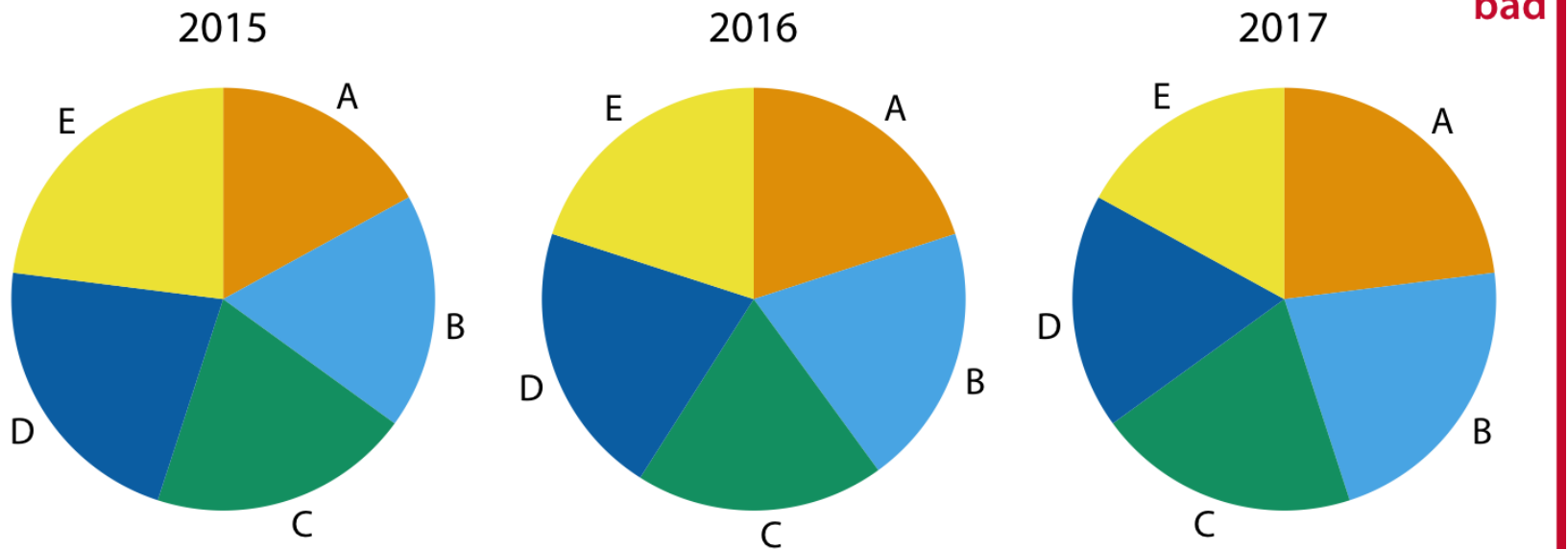
Which one is preferred?

- Many authors categorically reject pie charts and argue in favor of side-by-side or stacked bars.

- However, sometimes pie charts might be the preferred choice.

- For example, It work well when the goal is to emphasize simple fractions.

- They also work well when we have very small datasets.

- A single pie chart looks just fine, but a single column of stacked bars looks awkward.

- Stacked bars can work for side-by-side comparisons of multiple conditions or in a time series

- Side-by-side bars are preferred when we want to directly compare the individual fractions to each other.
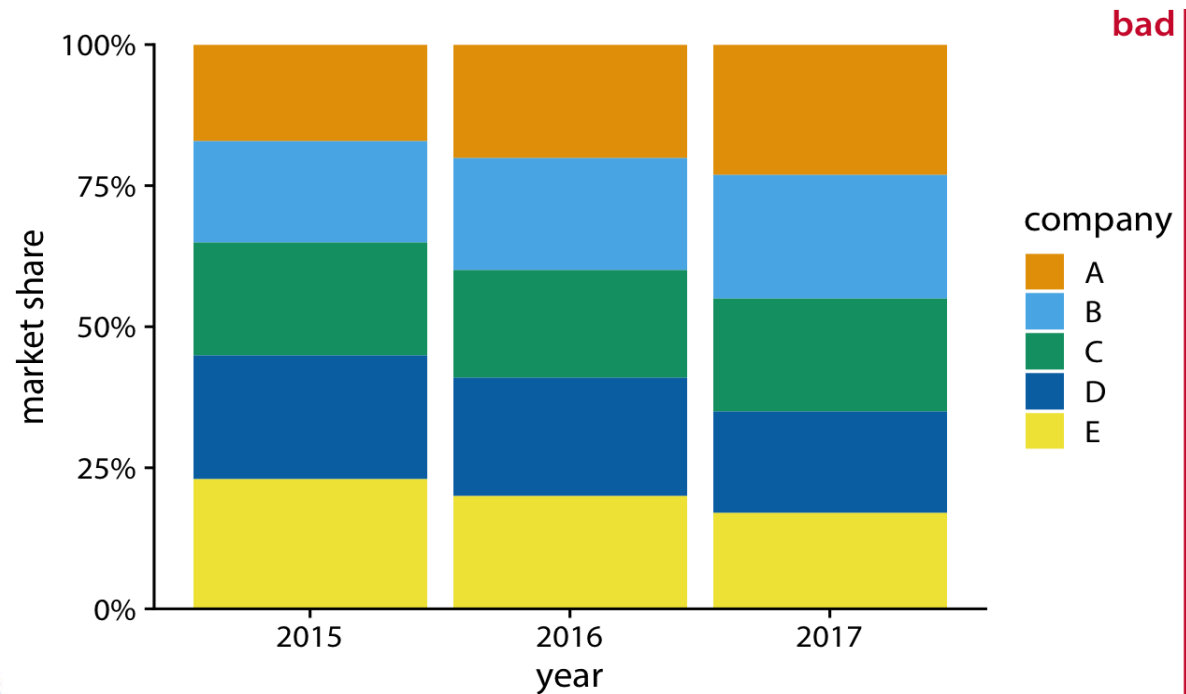
A Case for Side-by-Side Bars

- Consider the hypothetical scenario of five companies, A, B, C, D, and E, who all have roughly comparable market share of approximately 20%. Our hypothetical dataset lists each company's market share for three consecutive years.

- When we visualize this dataset with pie charts, it is difficult to see specific trends
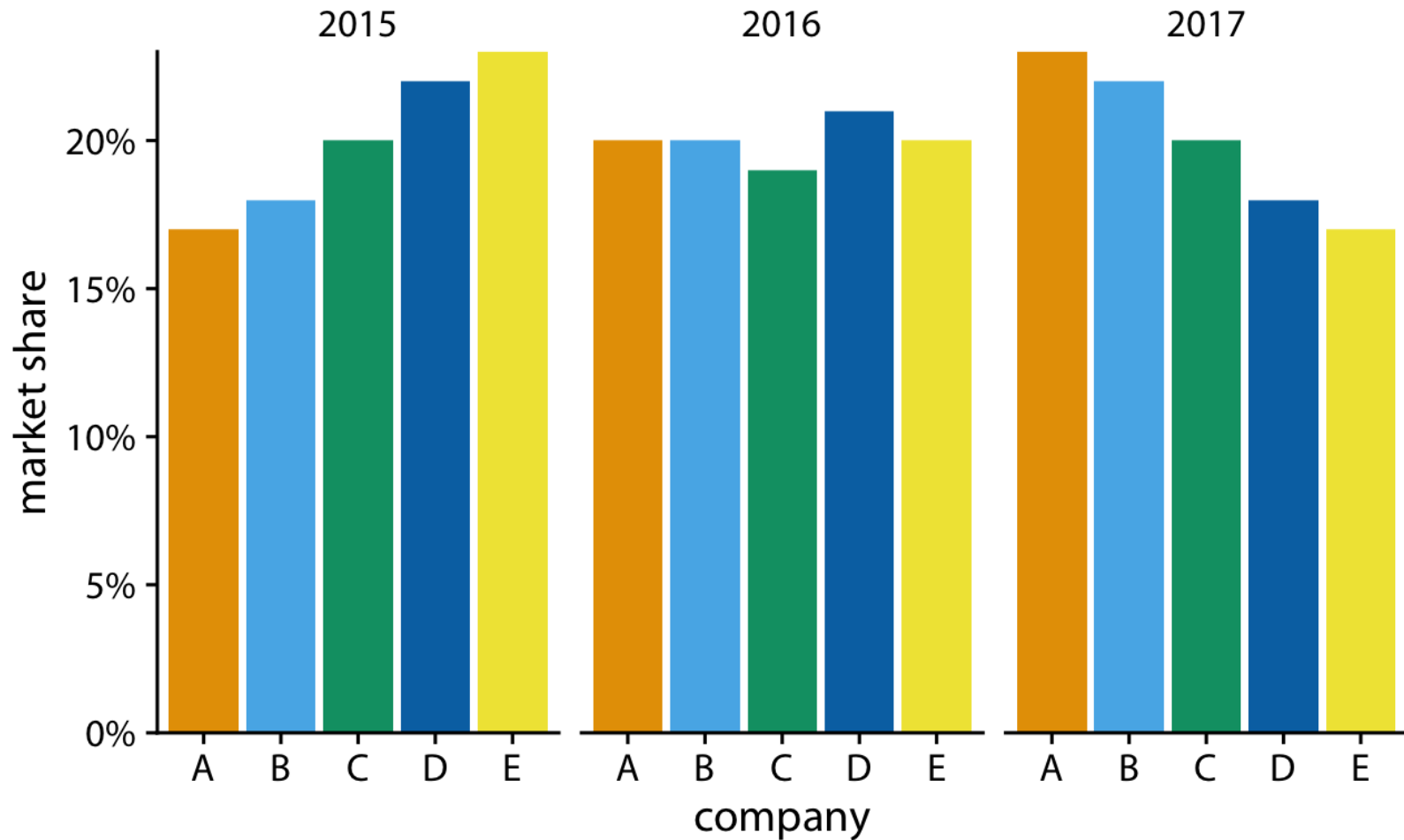
Stacked Bars

- Now the trends of a growing market share for company A and a shrinking market share for company E are clearly visible.
- However, the relative market shares of the five companies within each year are still hard to compare.
- And it is difficult to compare the market shares of companies B, C, and D across years, because the bars are shifted relative to each other across years.
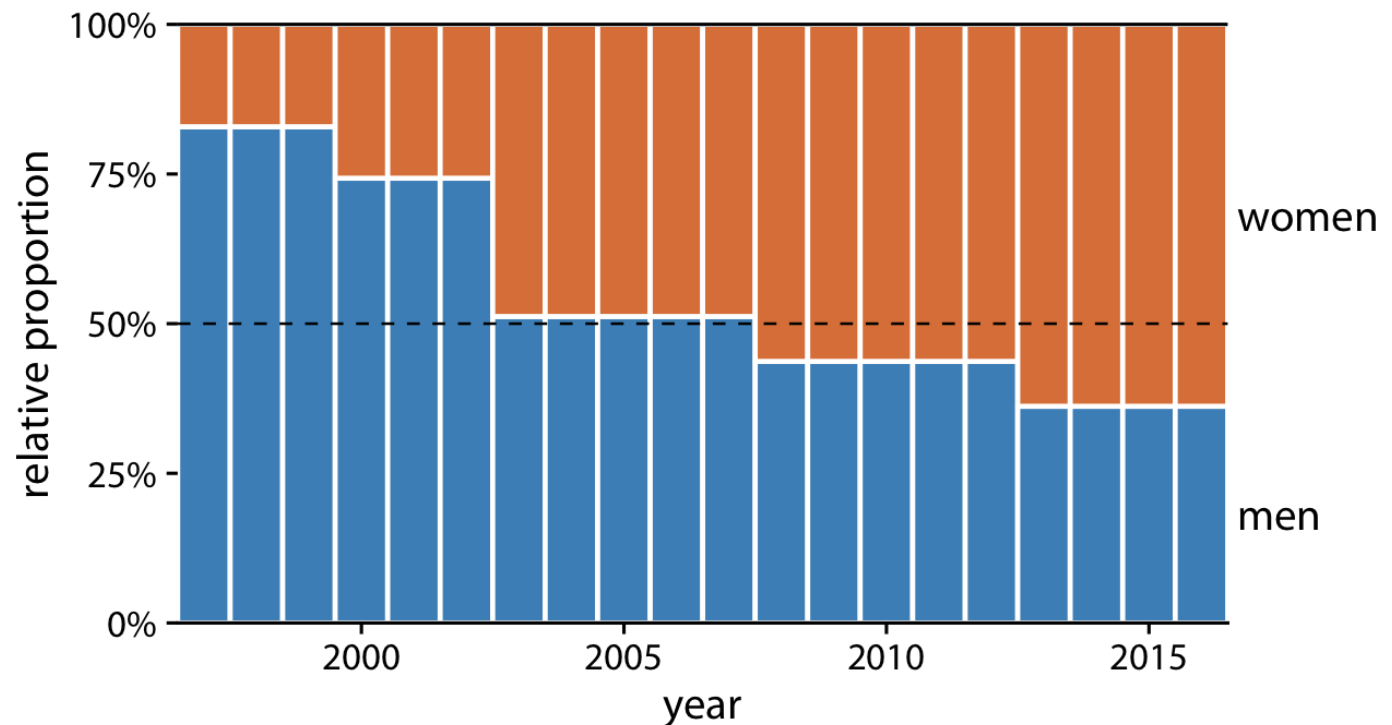
Side-by-side bars
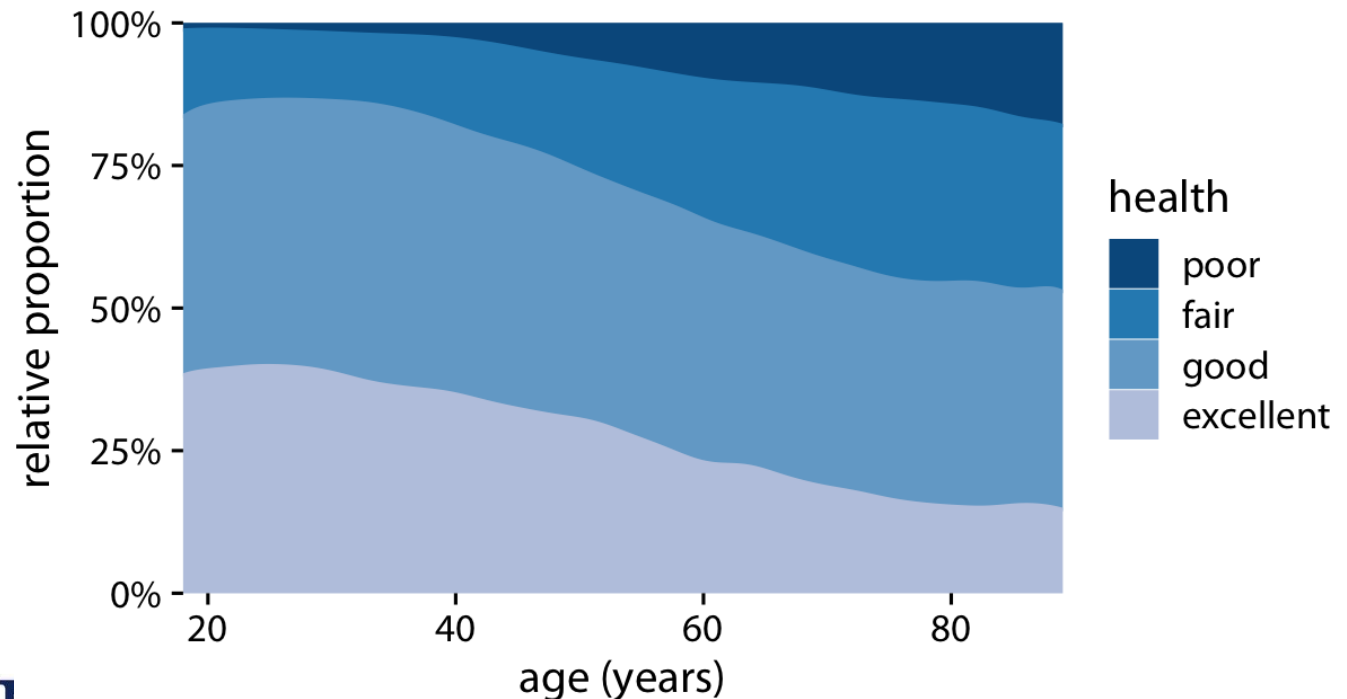
A Case for Stacked Bars and Stacked Densities

- the problem of shifting internal bars disappears if there are only two bars in each stack, and in those cases the resulting visualization can be quite clear.

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

Stacked Densities

- If we want to visualize how proportions change in response to a continuous variable, we can switch from stacked bars to stacked densities.
- By visualizing the proportions of the four health conditions as percentages of the total, the figure obscures that there are many more young people than old people in the dataset.
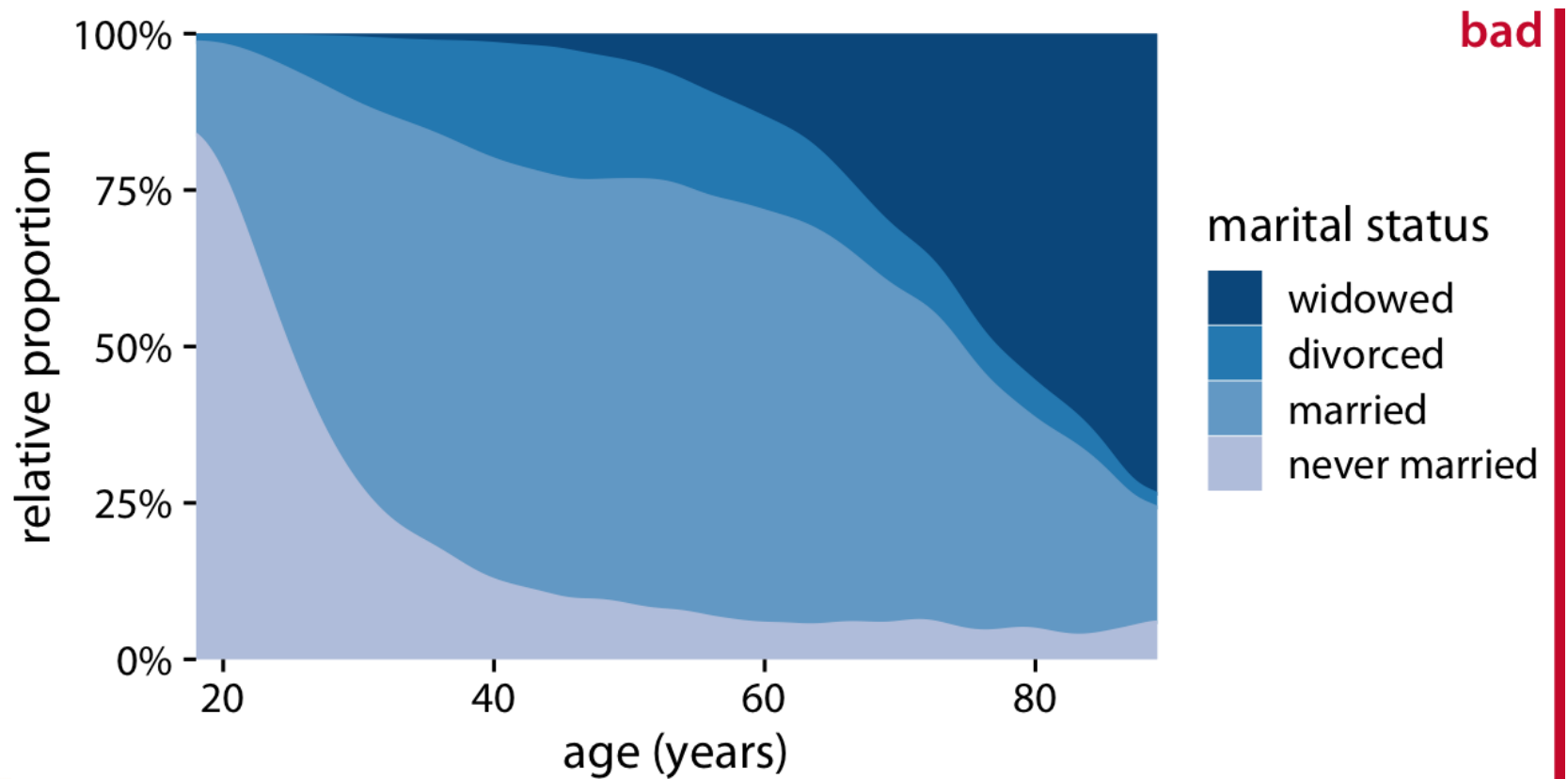
Visualizing Proportions Separately as Parts of the Total

- Side-by-side bars have the problem that they don't visualize the size of the individual parts relative to the whole, and stacked bars have the problem that the different bars cannot be compared easily because they have different baselines.

- We can resolve these two issues by making a separate plot for each part and, in each plot showing the respective part relative to the whole.
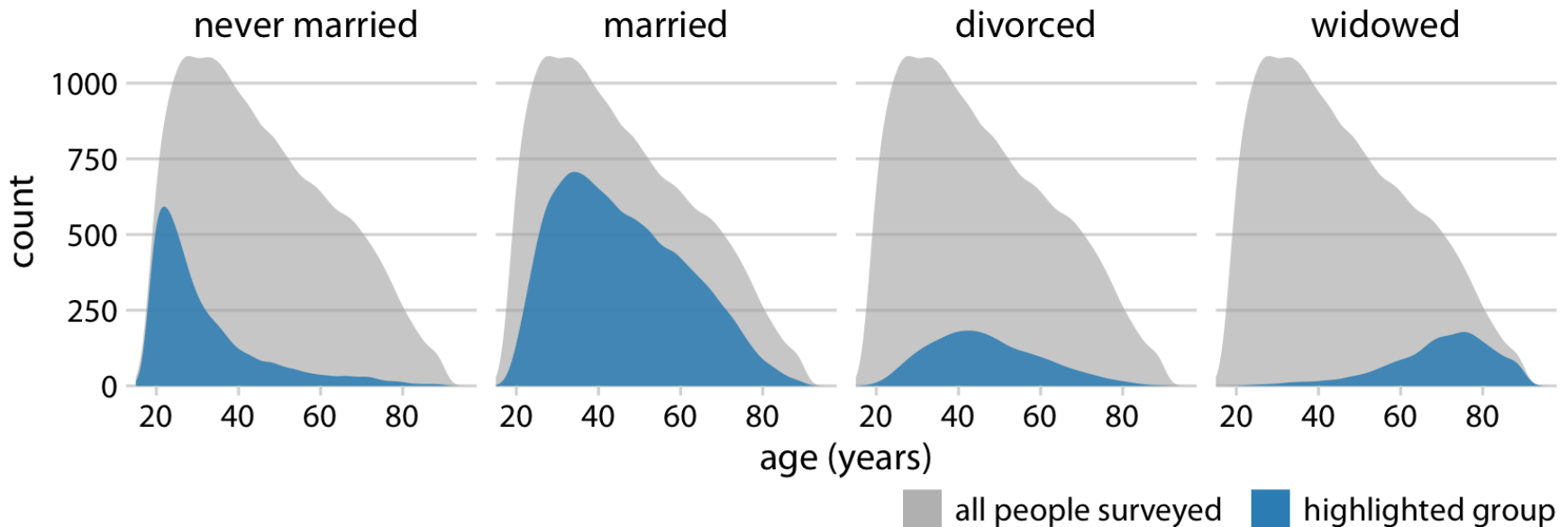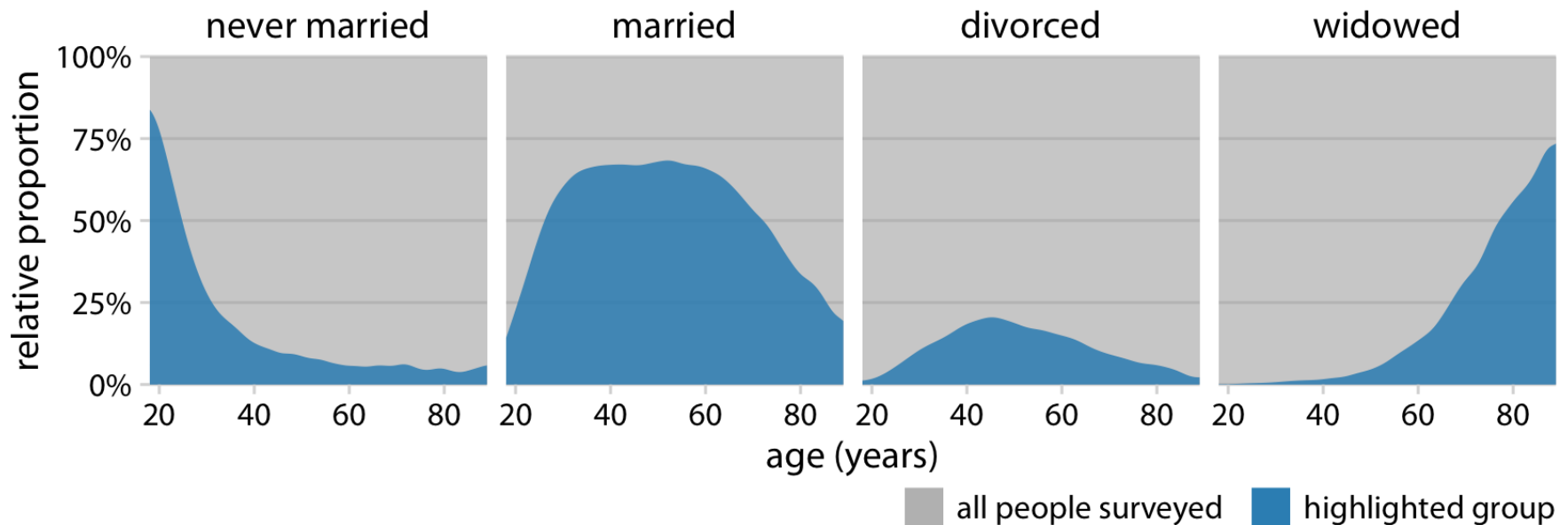
Marital status by age

Marital status by age

- Previous representation doesn't make it easy to determine relative proportions at any given point in time

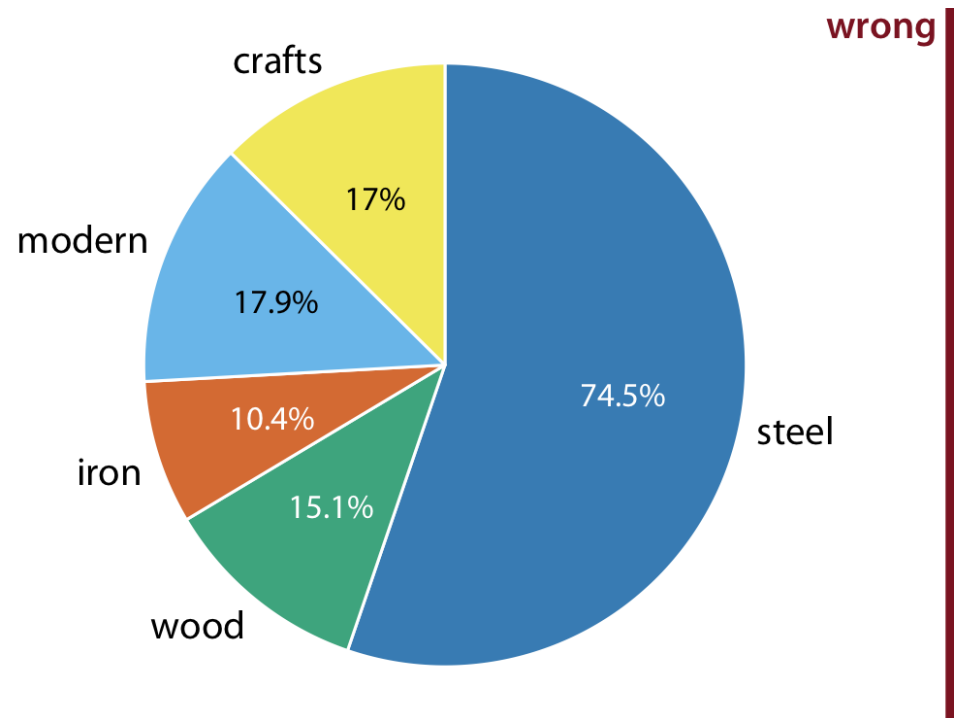# Marital status by age, shown as proportion of the total number of people in the survey

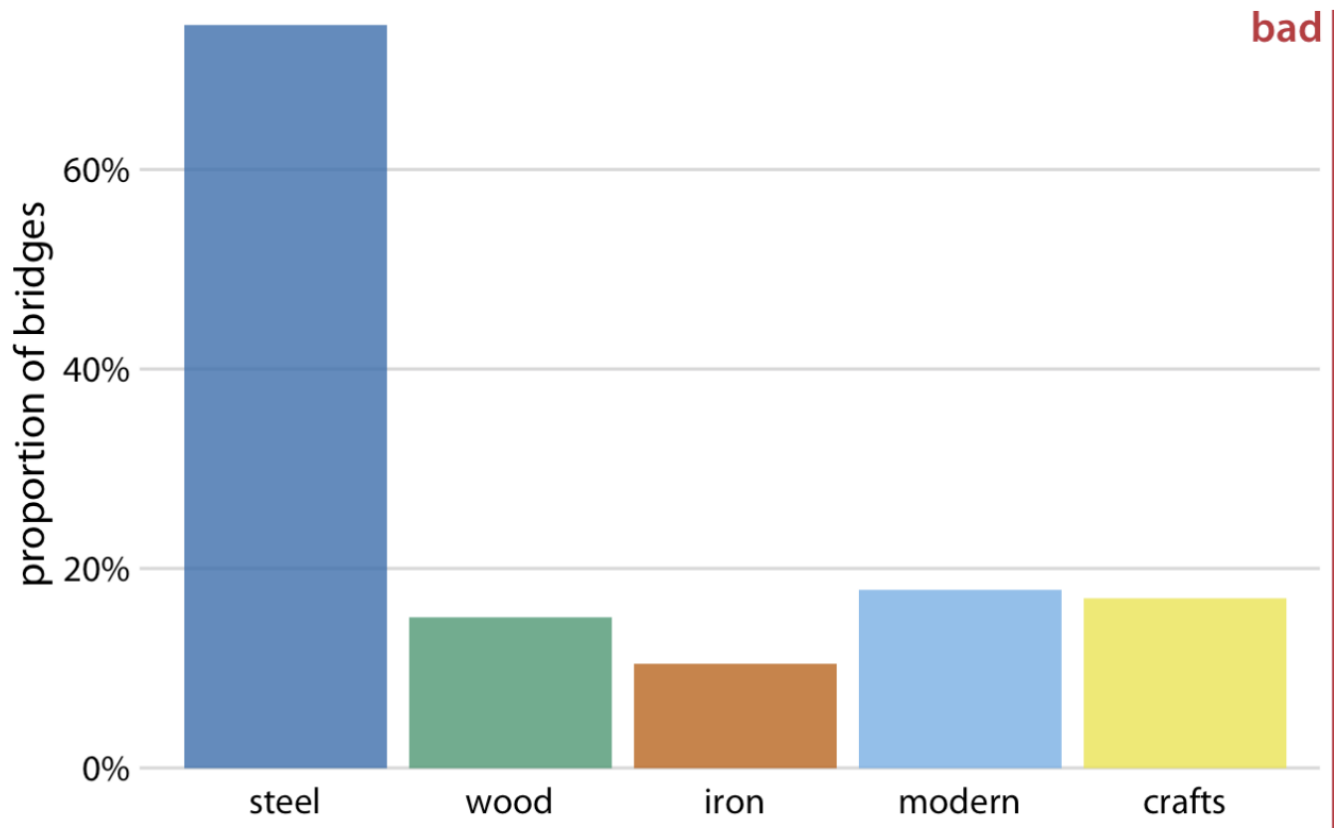# No one visualization fits all scenarios!

	Pie chart	Stacked bars	Side-by-side bars
Allows easy comparison of relative proportions	✖	✖	✔
Shows data as proportions of a whole	✔	✔	✖
Emphasizes simple fractions (1/2, 1/3, …)	✔	✖	✖
Visually appealing for small datasets	✔	✖	✔
Works well for a large number of subsets	✖	✖	✔
Works well for time series and similar	✖	✔	✖

Visualizing Nested Proportions

- A dataset of 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. Based on the year of erection, bridges are grouped into distinct categories, such as crafts bridges that were erected before 1870 and modern bridges that were erected after 1940.
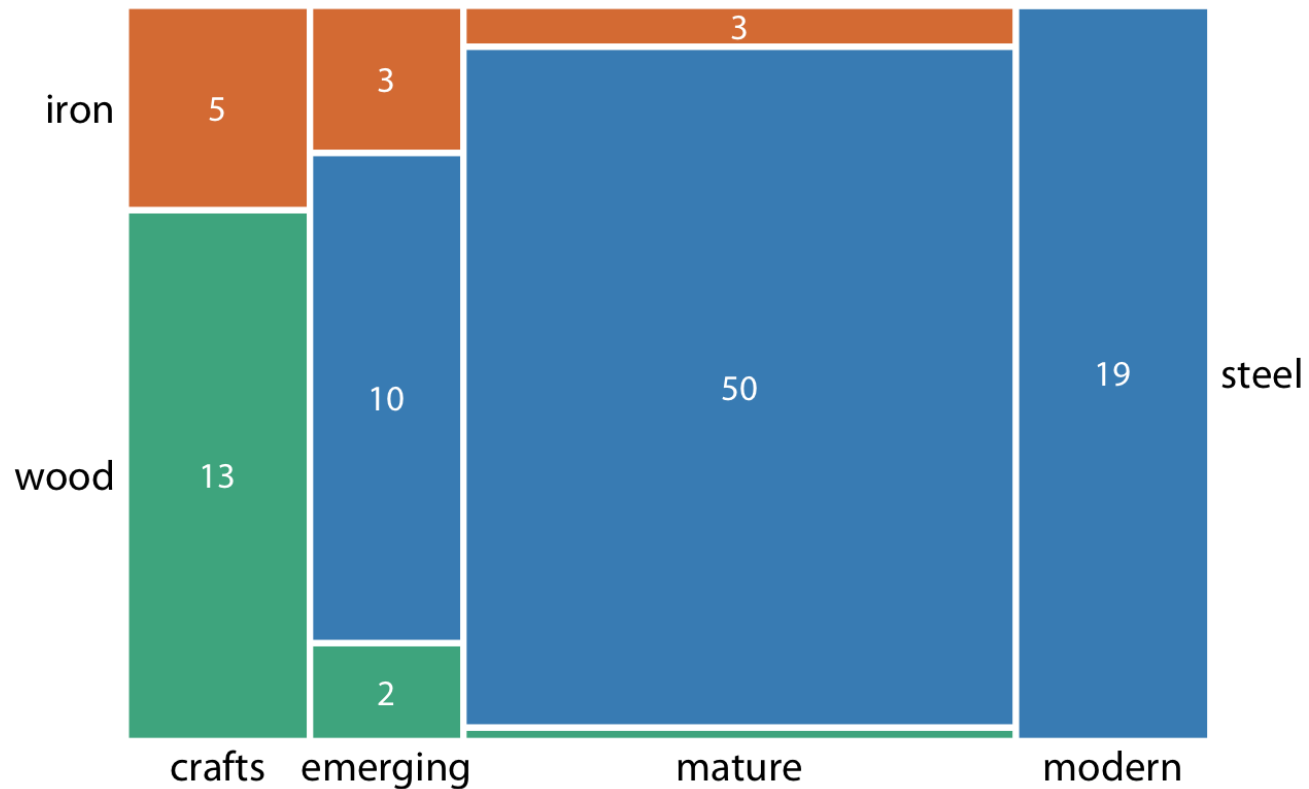
Visualizing Nested Proportions

Mosaic Plots

- What is it different from a stacked bar?

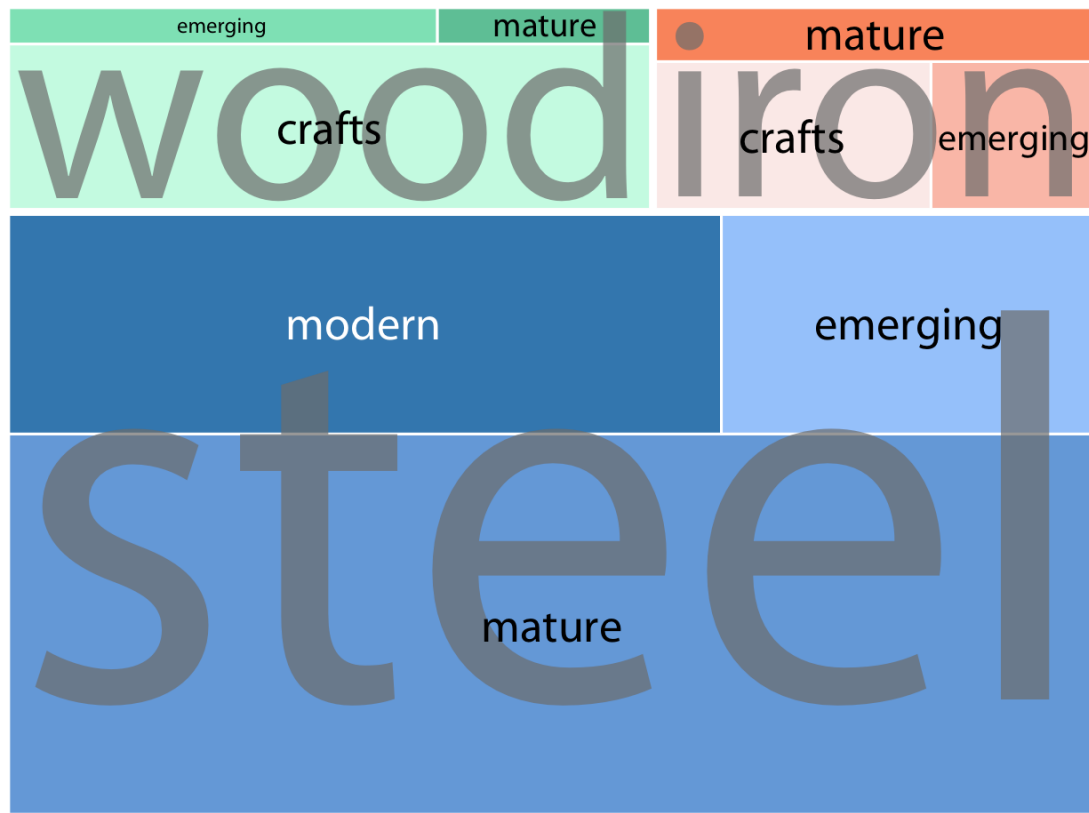Dataset: Bridges in Pittsburgh by construction material and era of construction

Mosaic Plots

- Unlike in a stacked bar plot, in a mosaic plot both the heights and the widths of individual shaded areas vary.
- Every categorical variable shown must cover all the observations in the dataset
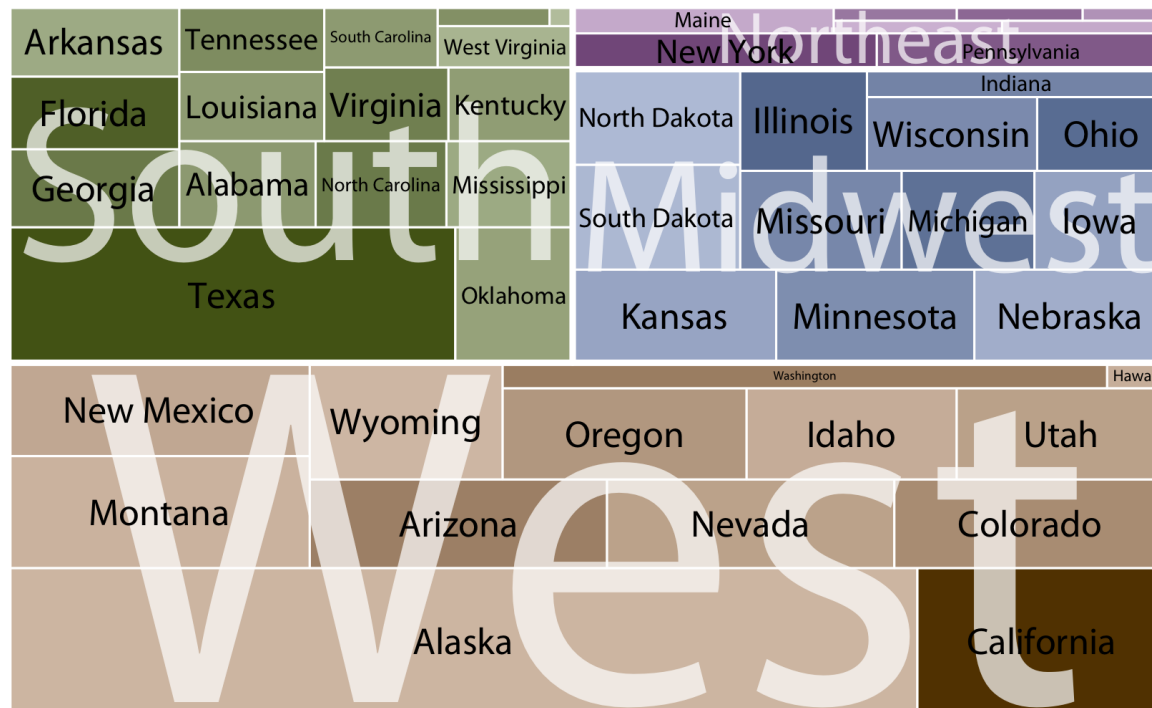
Treemaps

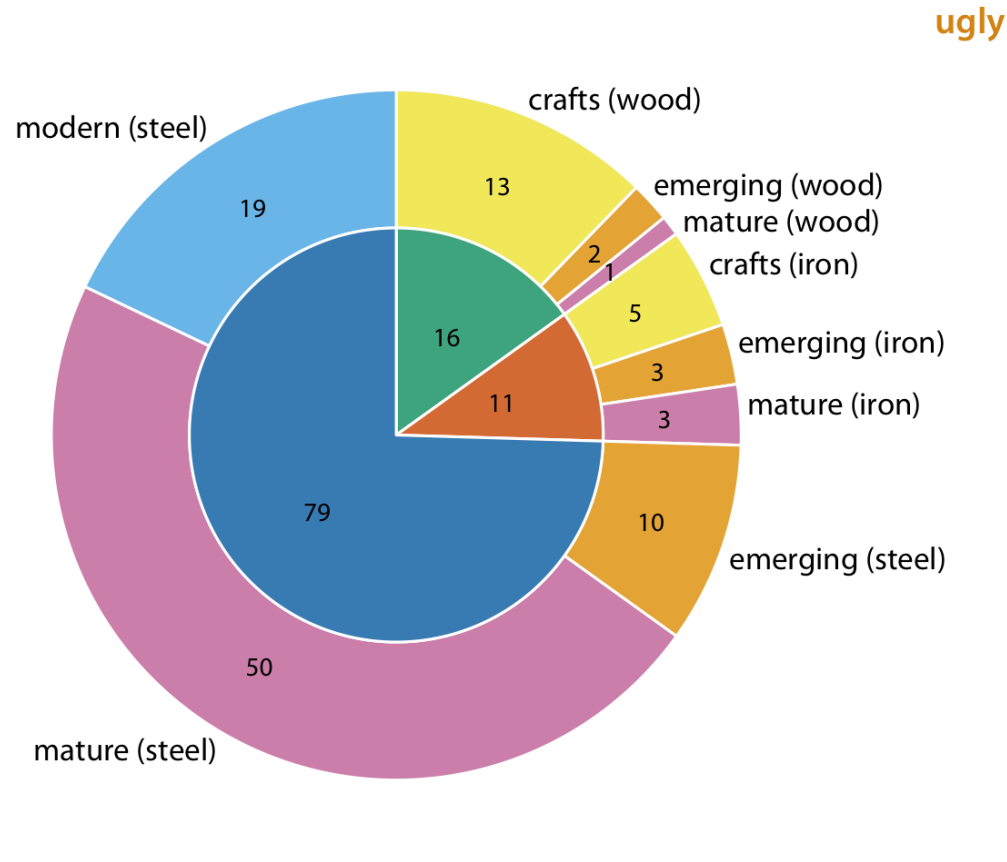- What is the main message of this plot?

Treemaps

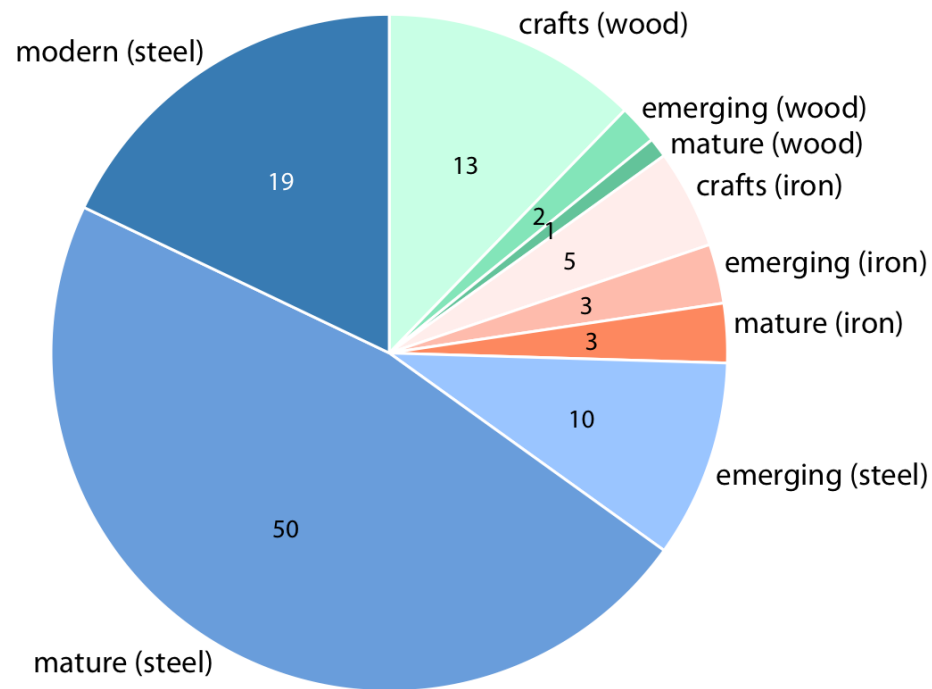- Can we use mosaic plot in this case?

Caveats of Mosaic and Treemaps

- Like stacked bars, a direct comparison among conditions can be difficult

- In mosaic plots or treemaps, this problem is exacerbated by the fact that the shapes of the different rectangles can vary
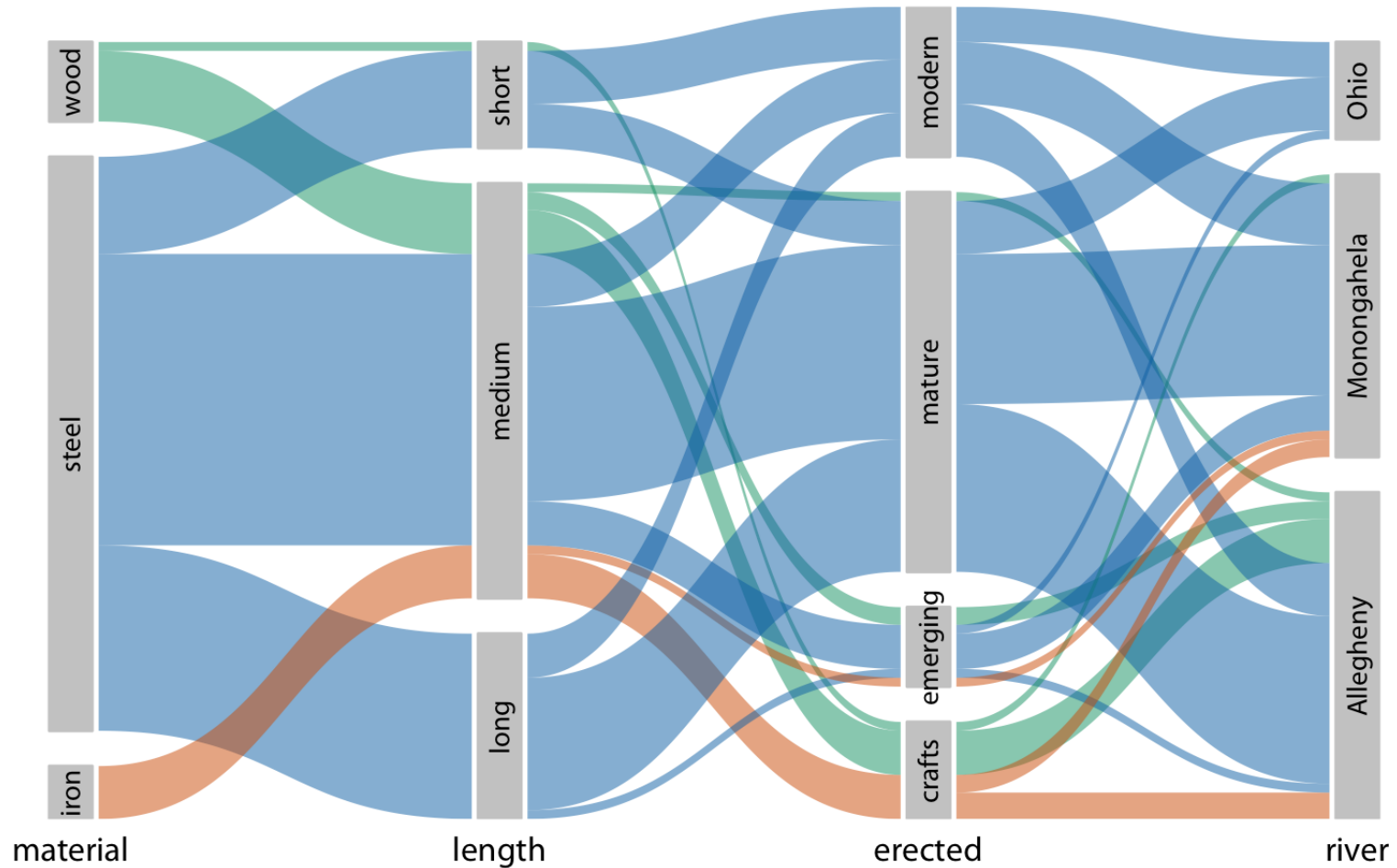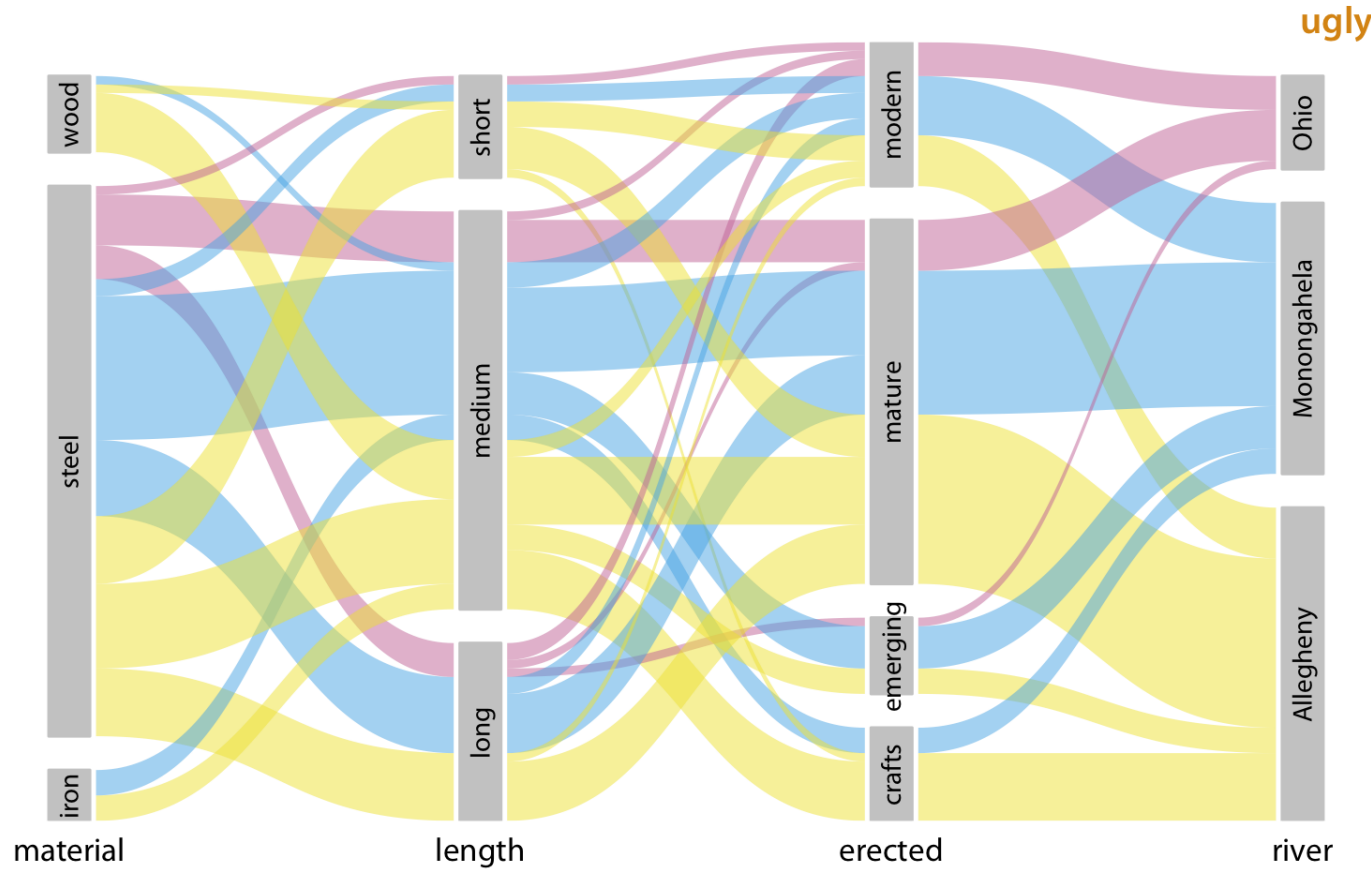
Nested Pies

Nested Pies

Parallel Sets

Parallel Sets

Parallel Sets