Chapter 10

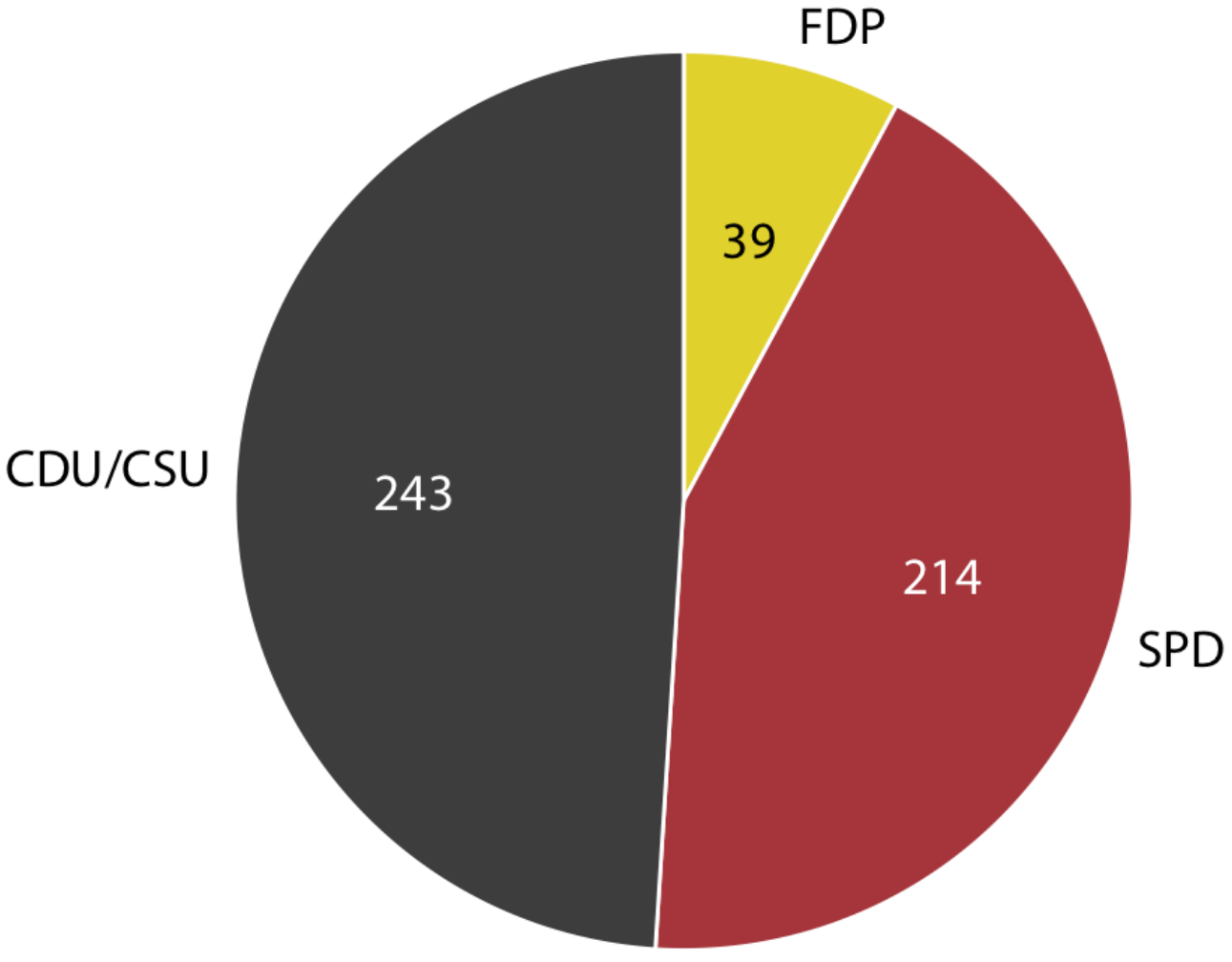
Fundamentals of Data Visualization

April 29, 2023

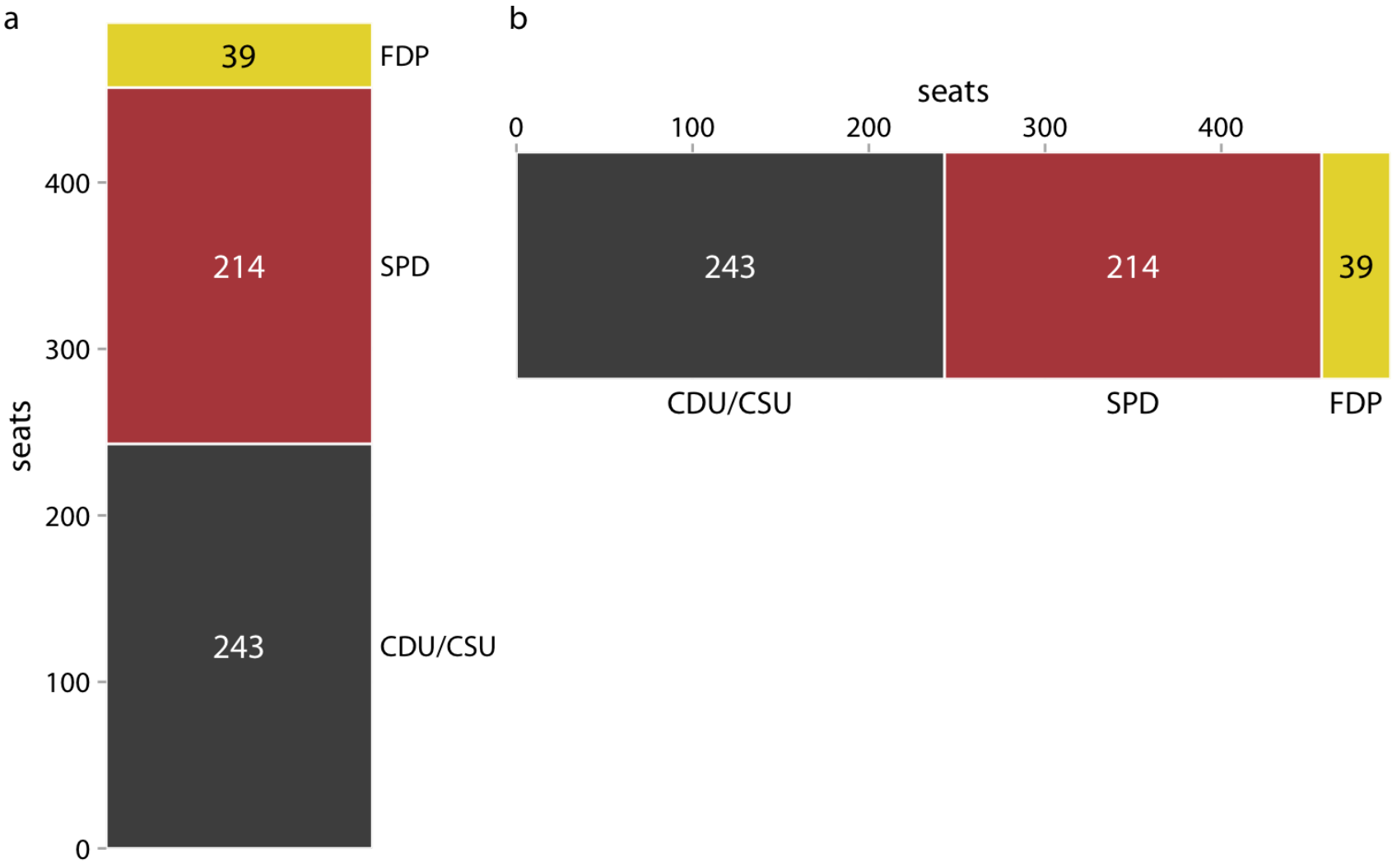
Visualizing proportions

- Examples:
  - the proportions of men and women in a group of people
  - the percentages of people voting for different political parties in an election
  - the market shares of companies.
- The archetypal such visualization is the pie chart, omnipresent in any business presentation and much maligned among data scientists.
- 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.**

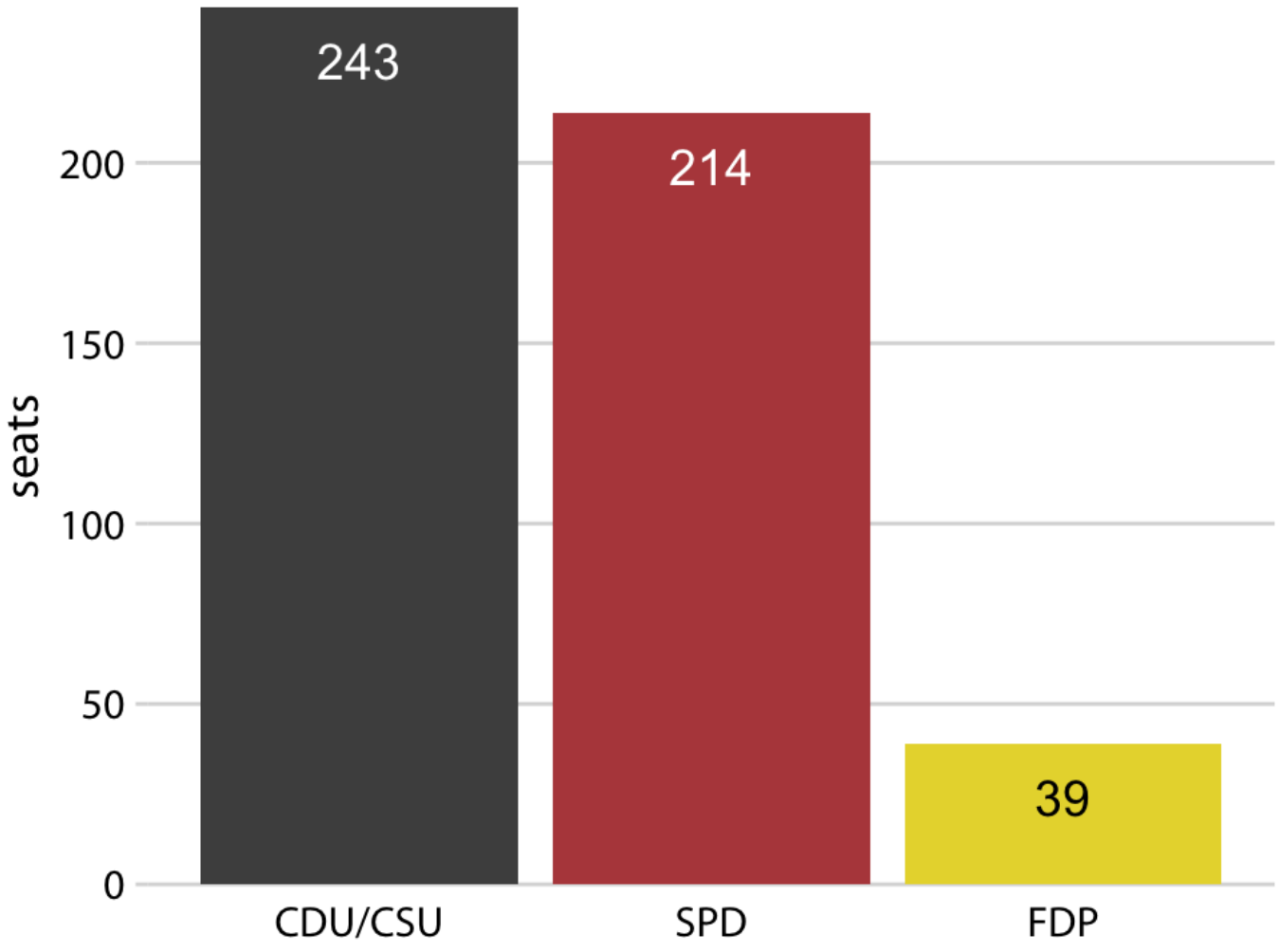
German parliament, 1976-1980



German parliament, 1976-1980



German parliament, 1976-1980



What's best?

- Many authors categorically reject pie charts.
- Others defend the use of pie charts in some applications.
- Perhaps none of these visualizations is consistently superior to any other?
- Depending on the story you want to tell, you may want to favor one or the other approach.
- In the case of the 8th German Bundestag, a pie chart shows clearly that the ruling coalition of SPD and FDP jointly had a small majority over the CDU/CSU.
- This fact is not visually obvious in any of the other plots.

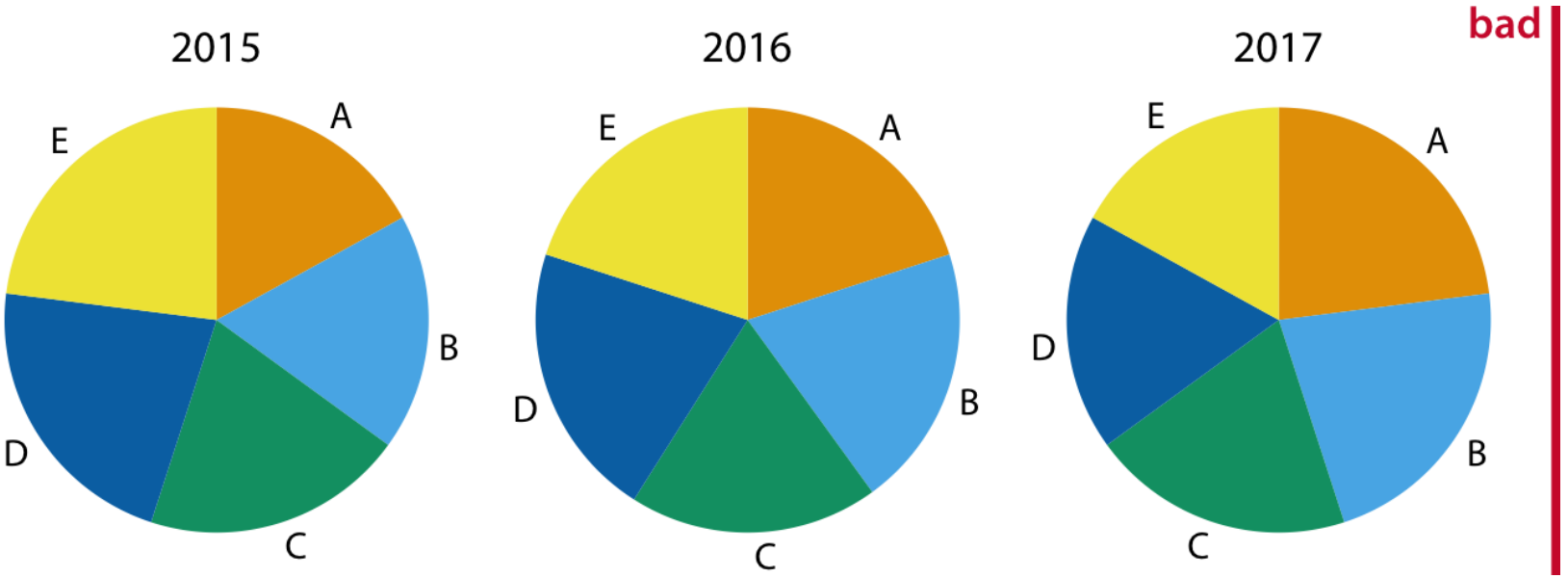
Tradeoffs

- Pie charts work well when the goal is to emphasize simple fractions, such as one-half, one-third, or one-quarter.
- 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 work when we want to directly compare the individual fractions to each other.

Tradeoffs

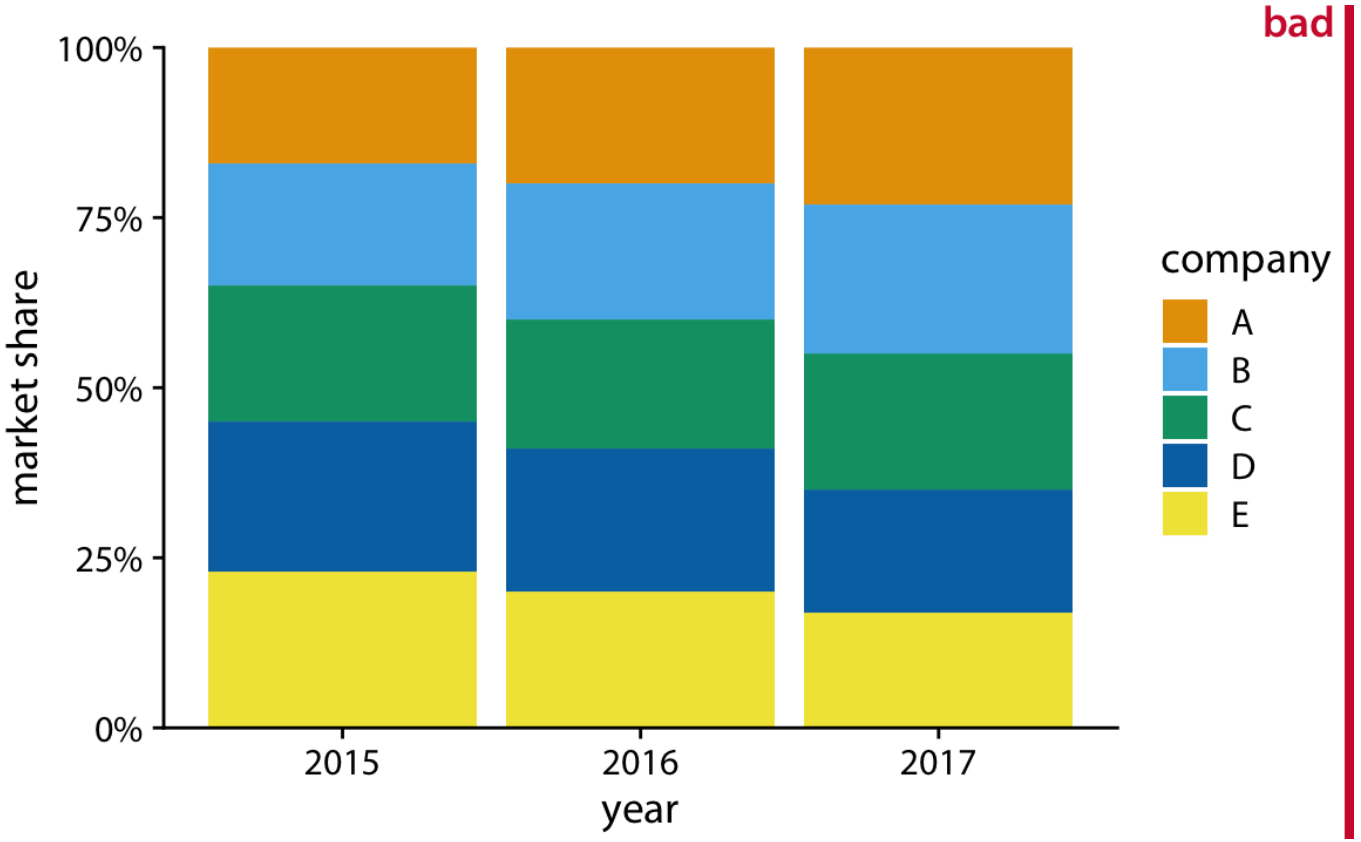
	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	×
Allows easy visual comparison of the relative proportions	×	×	✓
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✓	×	×
Looks visually appealing even for very small datasets	✓	×	✓
Works well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

Pie chart failure



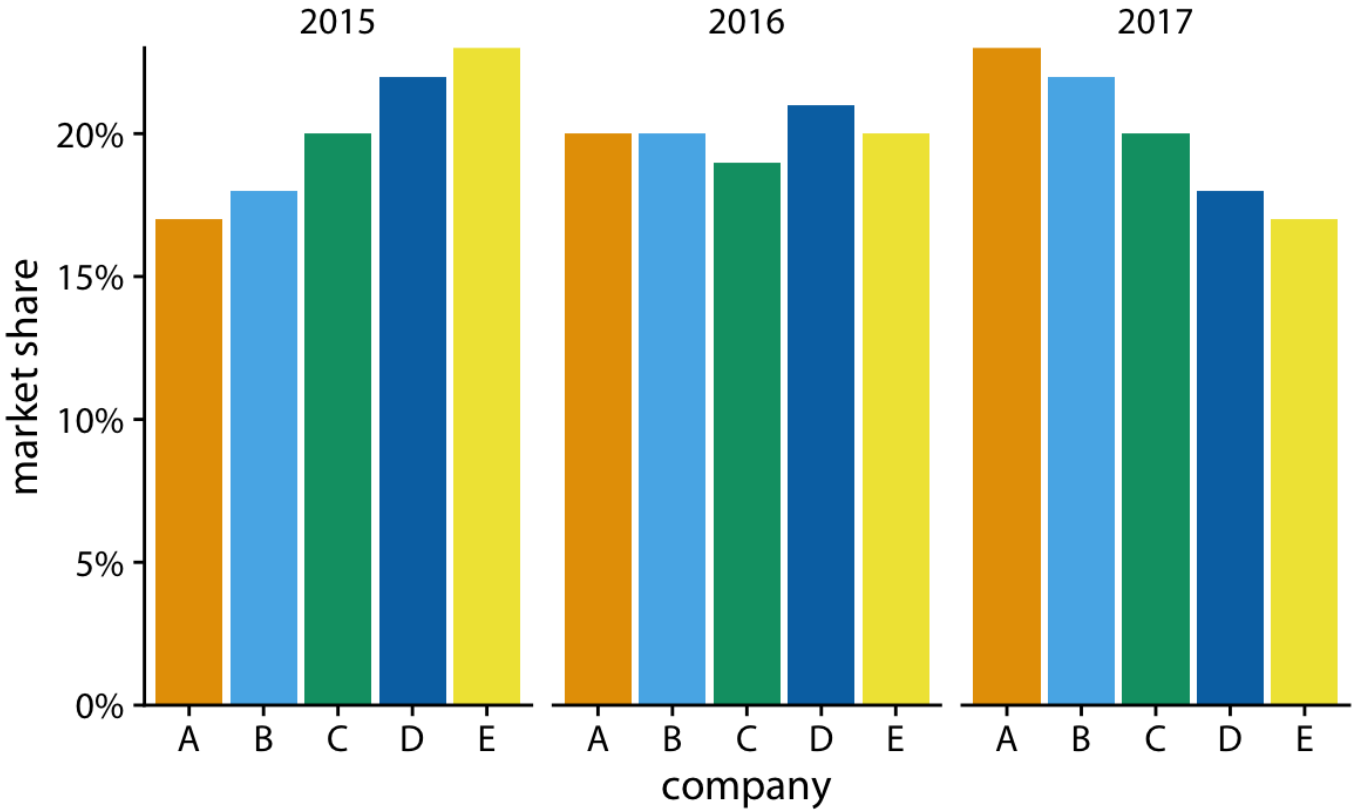
- Market share of A is growing and E is shrinking.
- What else?

Stacked bar failure



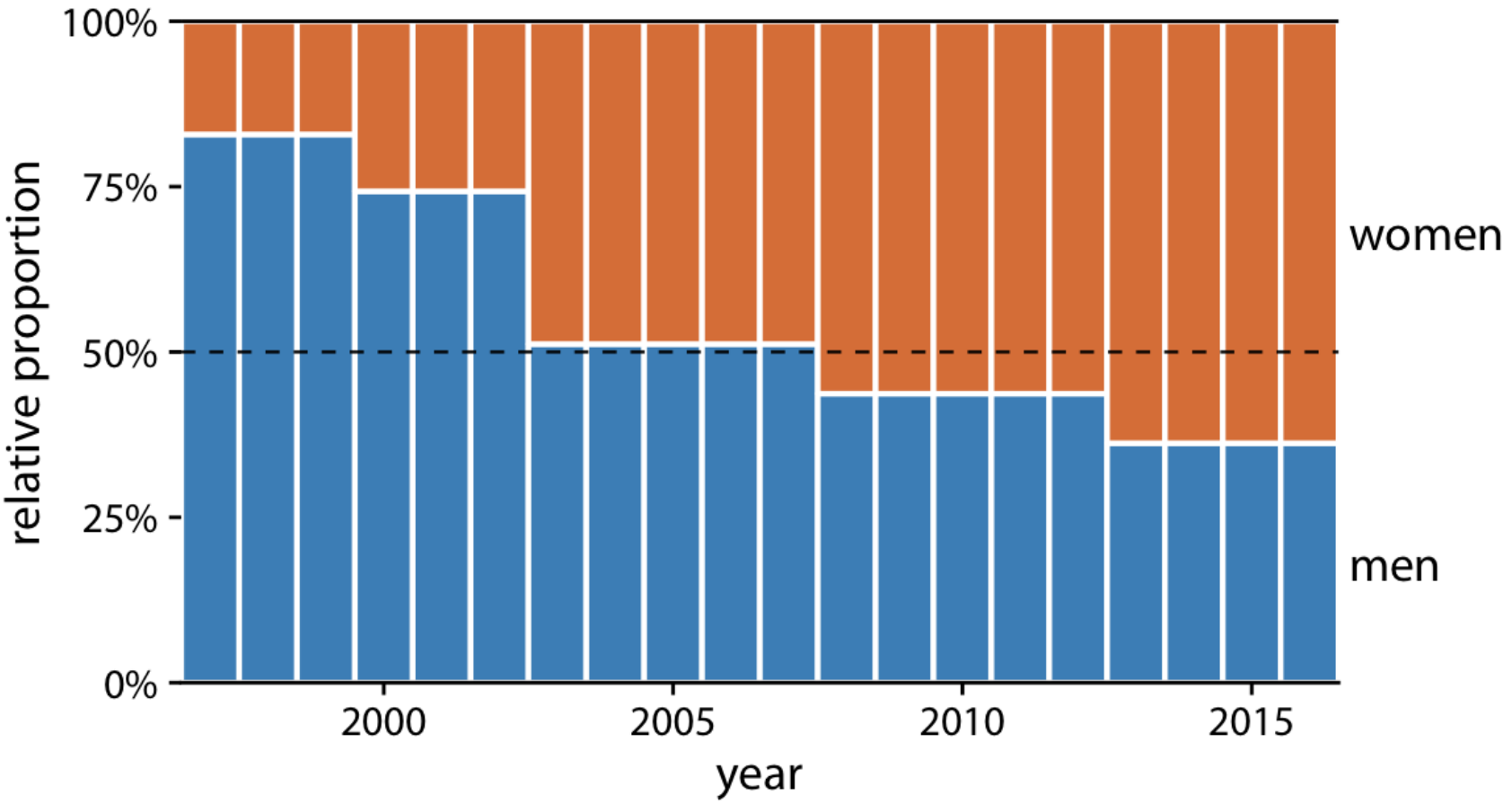
- Market share of A is growing and E is shrinking.
- What else?

Side by side bars



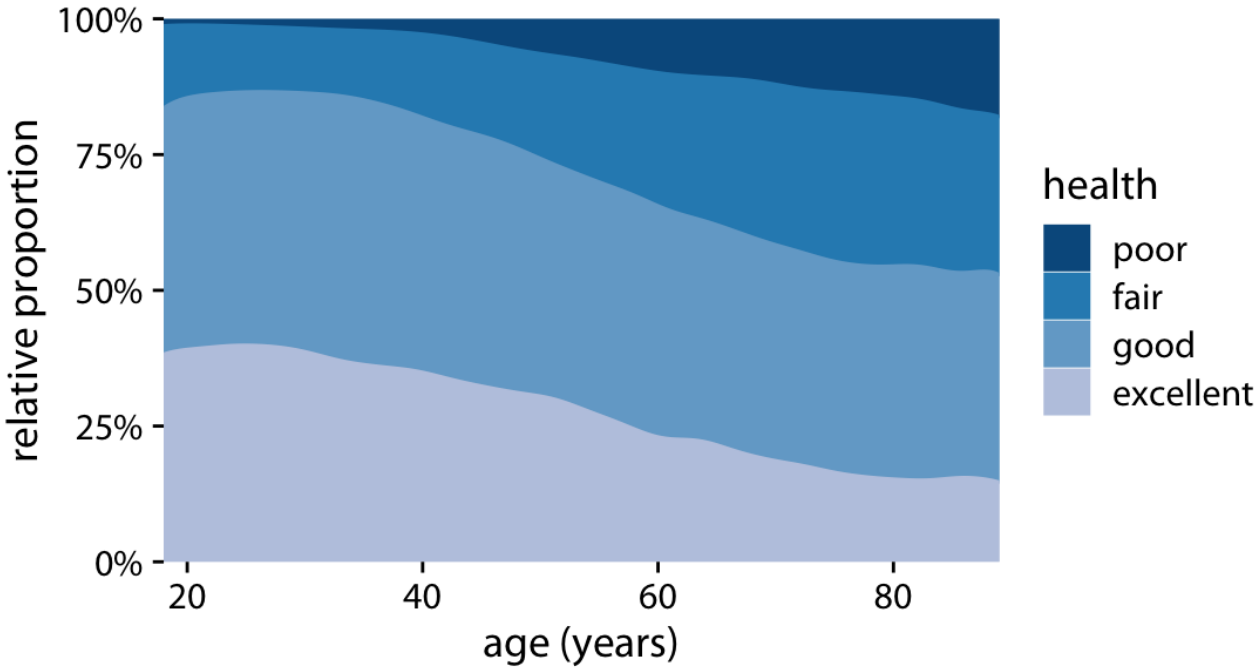
- The story is clear.

Stacked bars work when there are only two groups



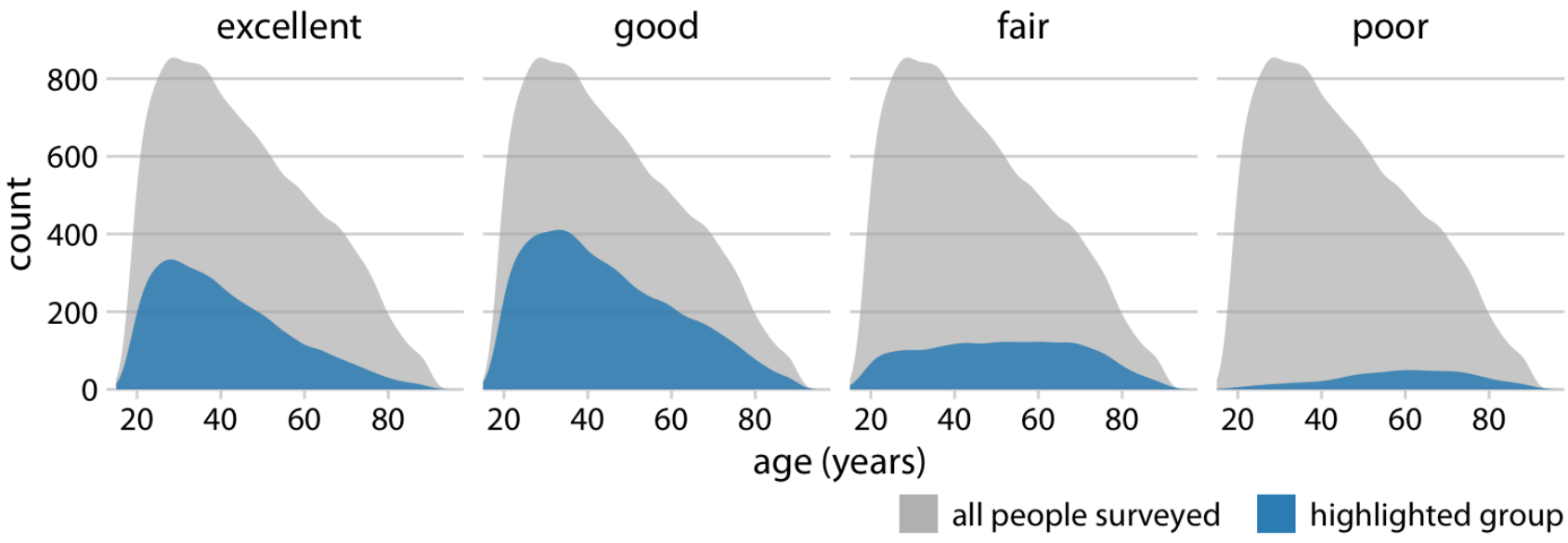
Change in the gender composition of the Rwandan parliament

Stacked densities



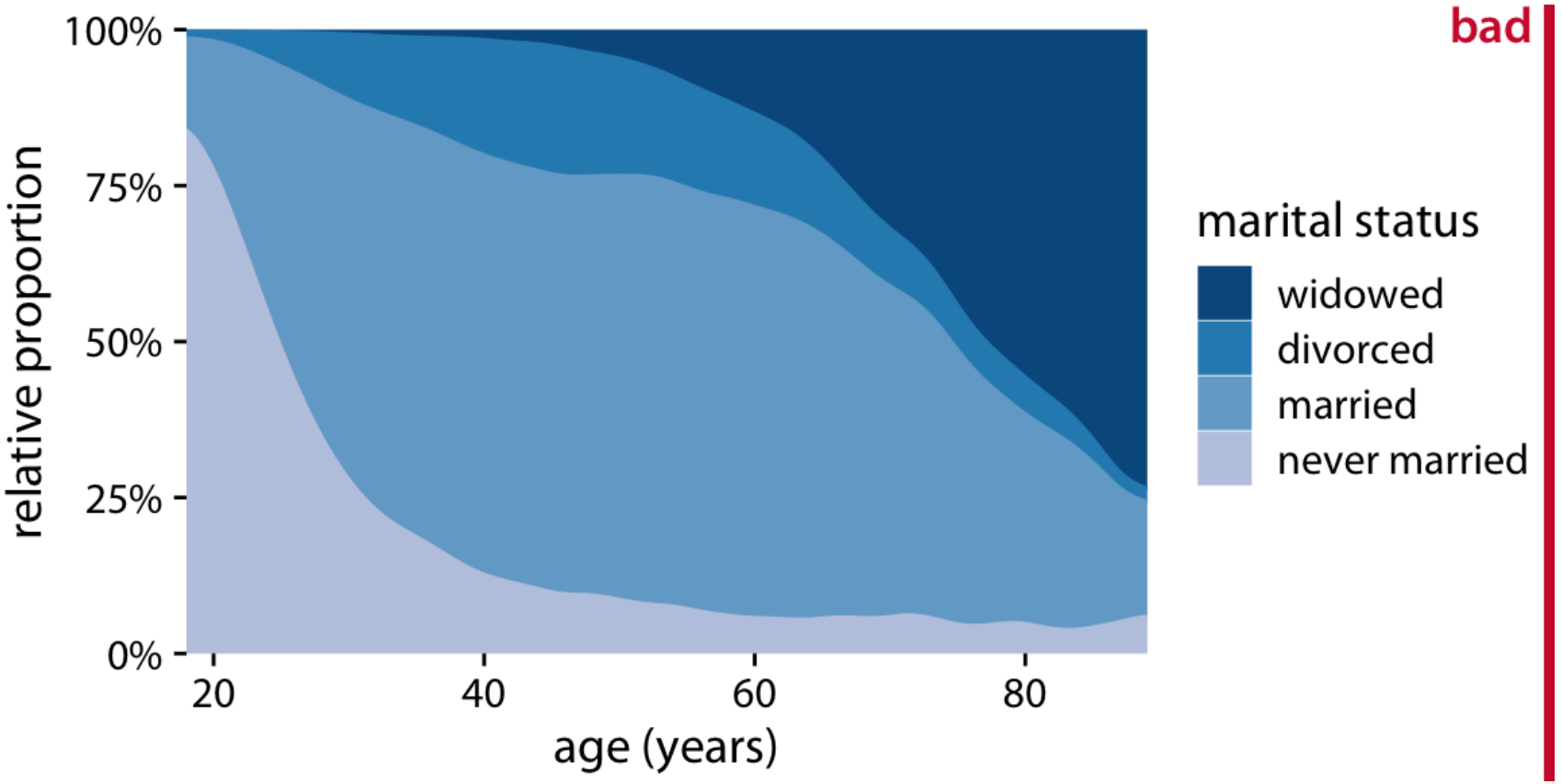
- Overall health declines with age.
- Over half remain in good or excellent health until very old age.
- Obscures the **amount** of people in favor of percentages.

Side-by-side densities

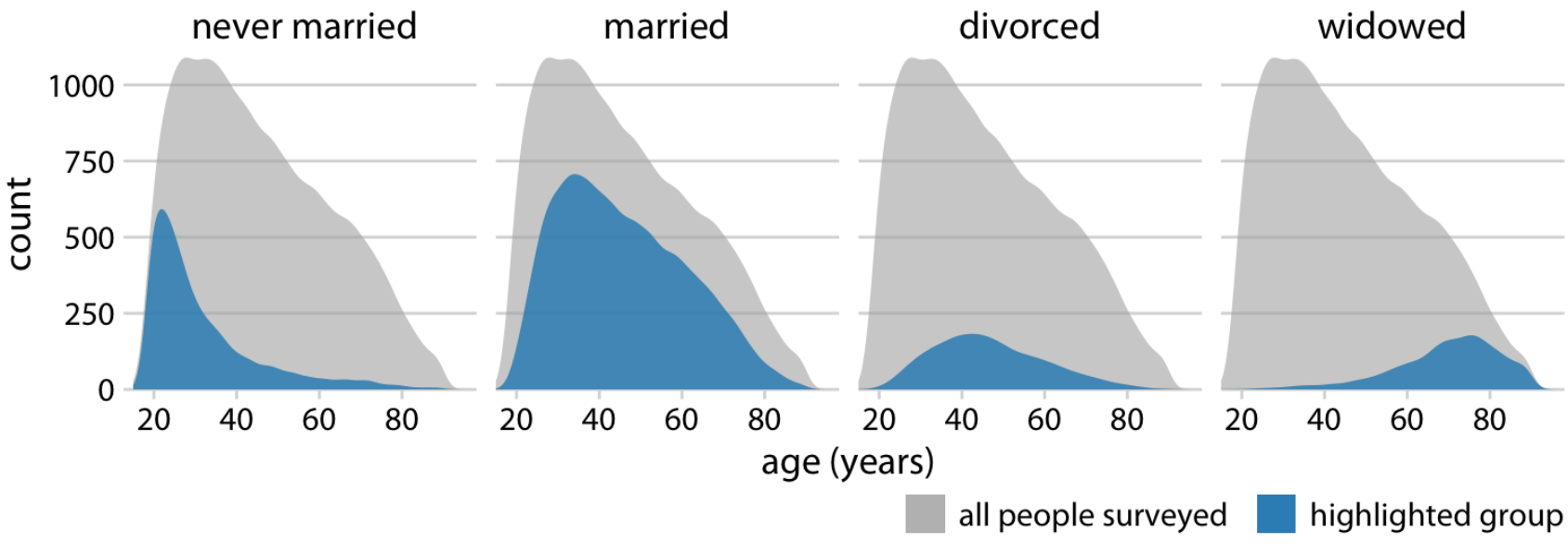


- Number in good and excellent declines with age.
- Number in fair remains constant.

Marital status does not work so well



Side-by-side reveals much more



Side-by-side proportions answer different questions