

# Chapter 10

## Fundamentals of Data Visualization

April 29, 2023

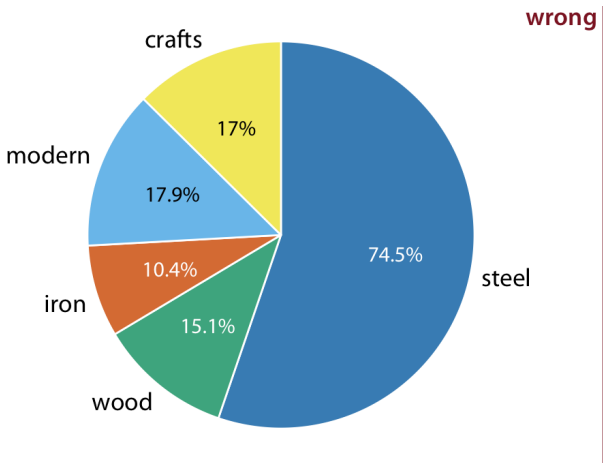
# Nested proportions

- German parliament by party and gender.
- Health status by age and marital status.
- Solutions:
  - Mosaic plots
  - Treemaps
  - Parallel sets

# Pittsburgh bridge dataset

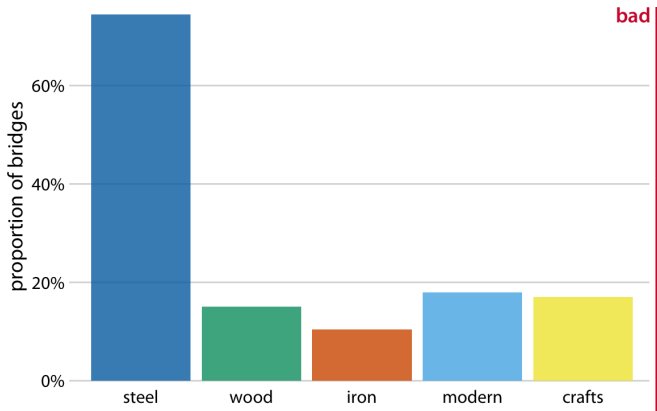
- 106 bridges in Pittsburgh.
- The material from which they are constructed:
  - steel
  - iron
  - wood
- The year when they were erected.
- Based on the year of erection, bridges are grouped into distinct categories:
  - crafts bridges that were erected before 1870
  - modern bridges that were erected after 1940

## Wrong approach



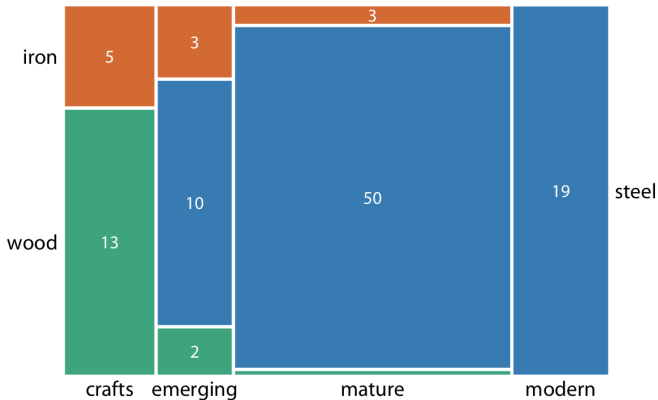
- Pie charts must add to 100%
- Double counting

## Side-by-side misleading



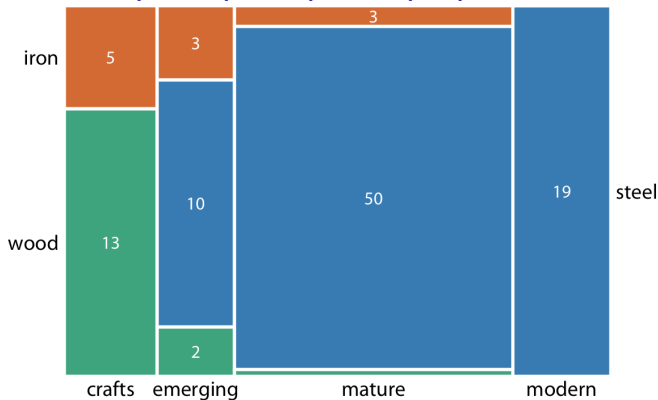
- Does not have to add to 100%
- Suggests there are five types of bridge

## Mosaic plot



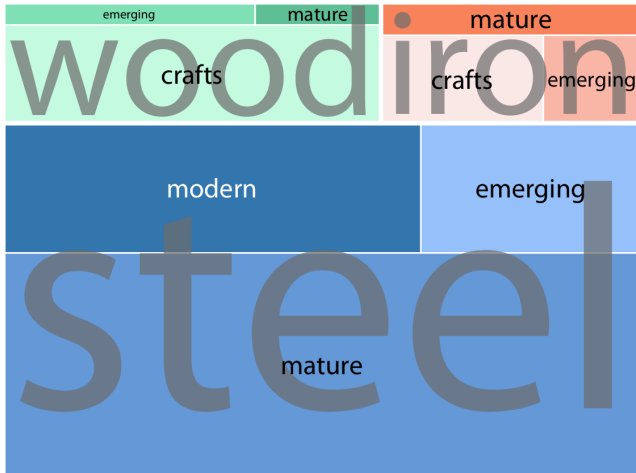
- Similar to stacked bar
- But both heights and widths matter
- Added emerging and mature categories
- Every categorical variable shown must cover all the observations in the dataset.

## Mosaic plot: principle of proportional ink



- Start by subdividing the x axis proportionally to one categorical variable.
- Then subdivide the y axis for each category of the x axis.
- Each rectangle is proportional to the number of cases for that combination.

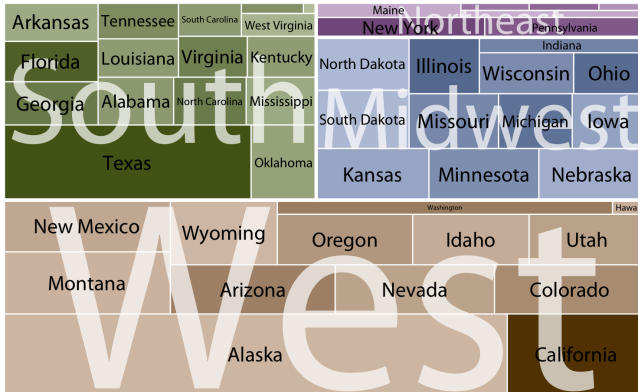
## Tree map: nested rectangles



- Mosaic plot emphasizes the time evolution of categories
- Tree map emphasizes the relative proportions of categories.

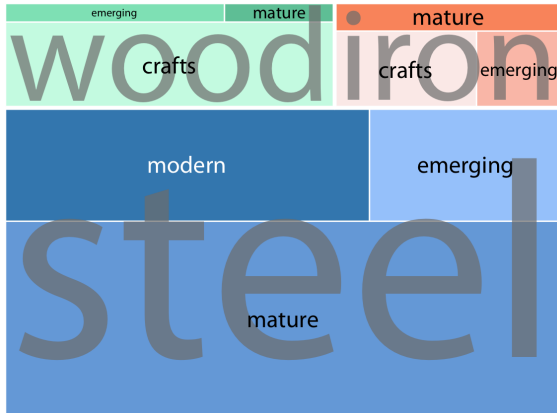


# Mosaics vs. Tree maps



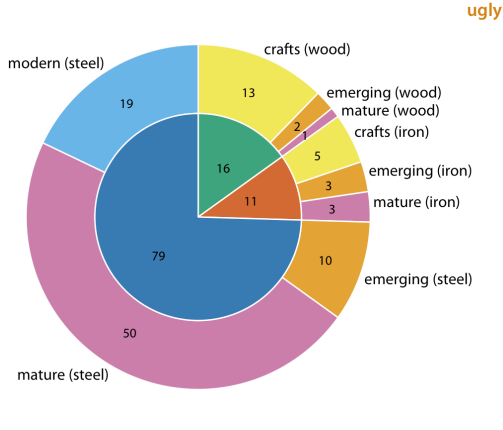
- Mosaics assume all proportions shown can be identified as combinations of orthogonal categorical variables.
- Tree maps tend to work well when the proportions cannot be described by combining multiple categorical variables.

## Tree map



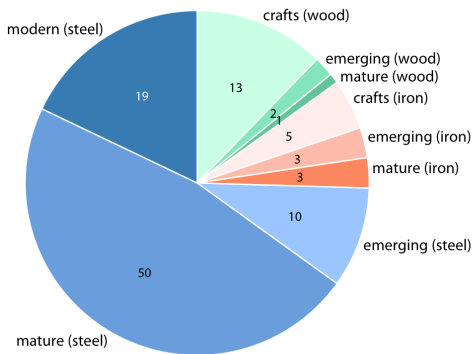
- Direct comparisons of areas is difficult.
- No common baseline.
- No common shape.
- For example, emerging and mature iron bridges are the same size (3).
- Adding numbers to the areas can help.

# Nested pies



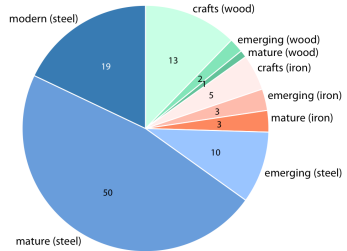
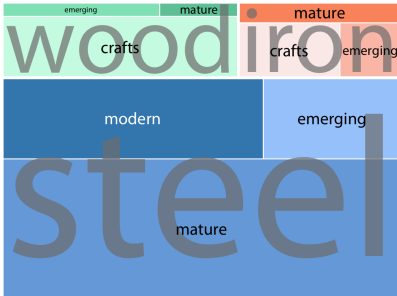
- The inner circle shows the breakdown of the data by one variable (here, building material) and the outer circle shows the breakdown of each slice of the inner circle by the second variable (here, era of bridge construction).

## Nested pies



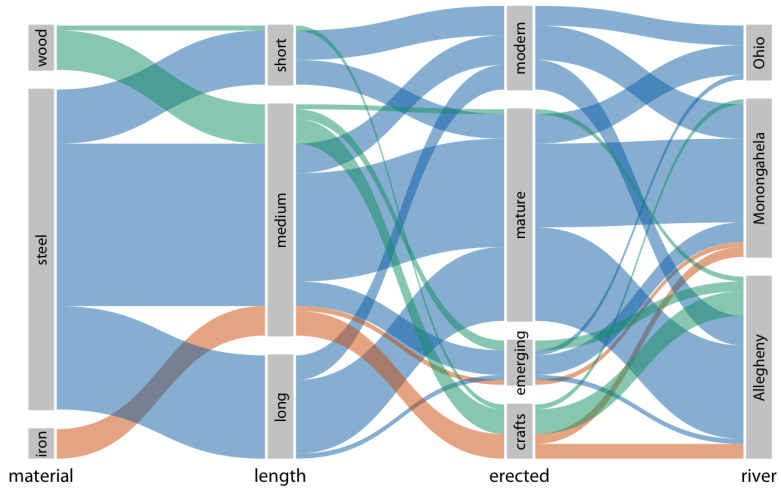
- Alternatively, we can first slice the pie into pieces representing the proportions according to one variable (e.g. material) and then subdivide these slices further according to the other variable (construction era)

## Treemap vs. Nested pie

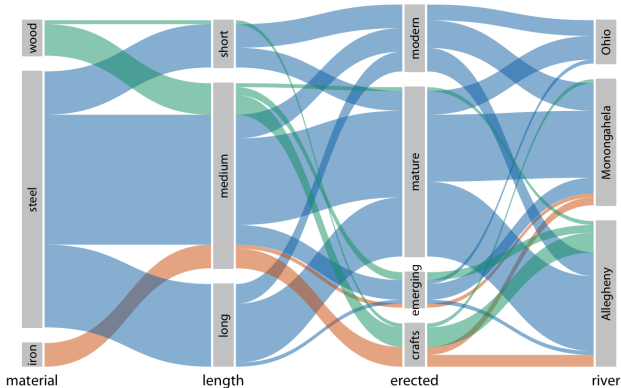


- Treemap uses space better
- Permits inside labels
- Some pie slices are too thin
- But every rectangle is reasonable size

# Parallel sets plot



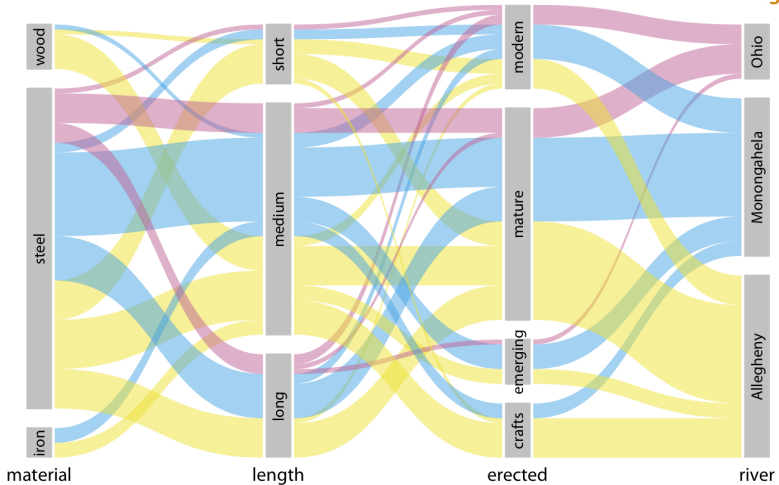
# Parallel sets plot



- wood bridges are mostly of medium length
- wood bridges primarily erected during the crafts period
- wood bridges span primarily the Allegheny river
- iron bridges are all of medium length
- iron bridges primarily erected during the crafts period
- iron bridges span the Allegheny and Monongahela rivers in approximately equal proportions.

## Parallel sets colored by river

ugly



- Colors should start at the left
- Rearrange categories to minimize crossing



## Parallel sets colored by river

