Visualization

Aleksandr Fisher

(Really) Bad Visualizations

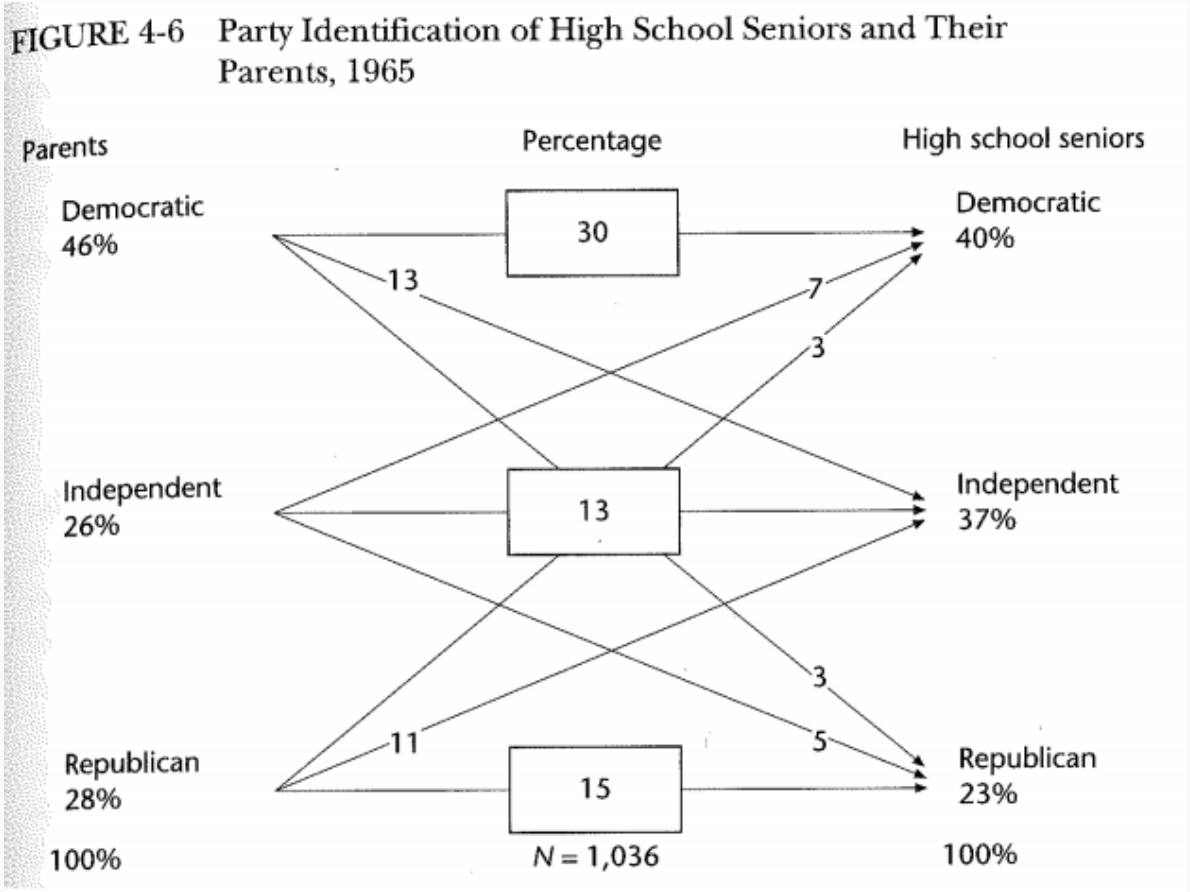


Figure 1: Panel study that examined how much high school seniors retained the party affiliations of their parents

Improved Visalization

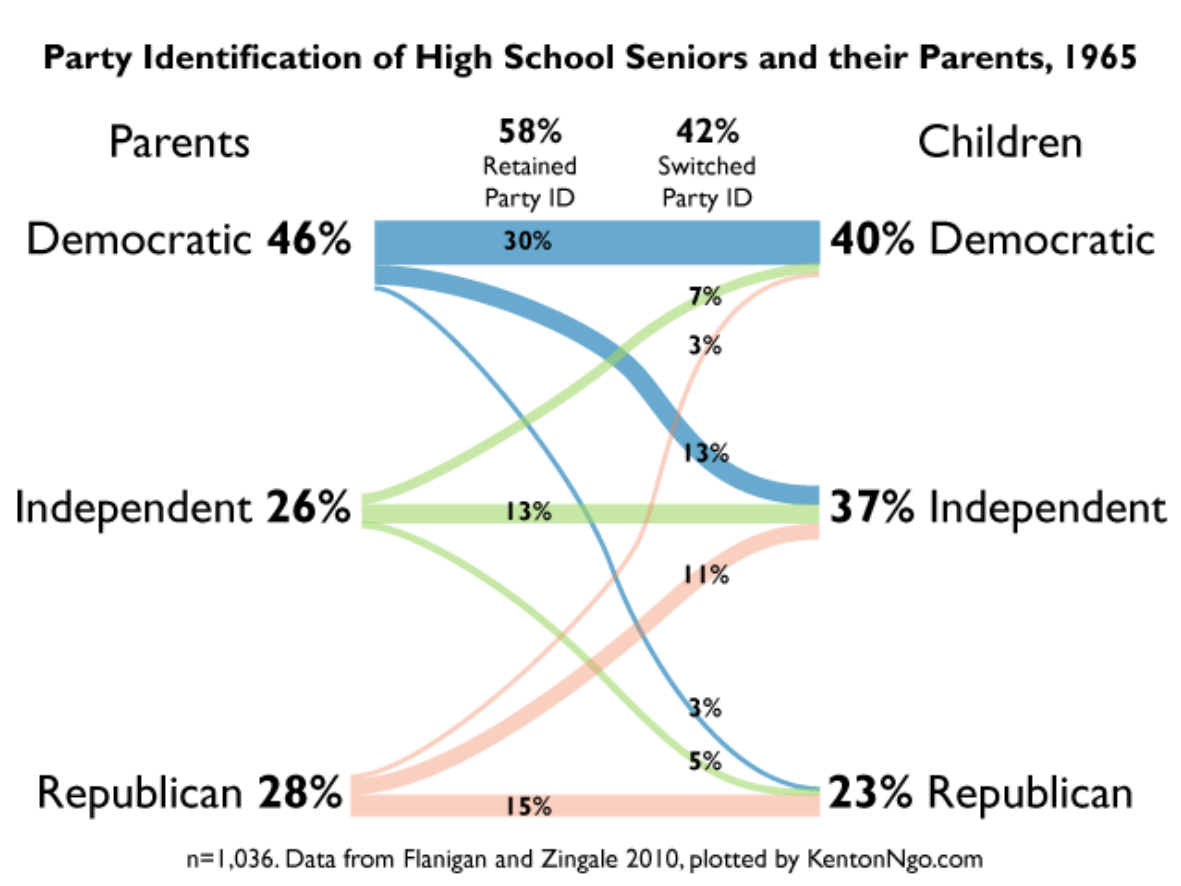


Figure 2: Panel study that examined how much high school seniors retained the party affiliations of their parents

Basic Principles

- Be honest
- Data-ink ratio
- Tell a story
- Steer reader attention
- Use balanced color palettes

In R...

- R has 5+ graphics "systems"
  - Base graphics
  - The ggplot2 package
  - The lattice package
  - The plotrix package
  - The htmlwidgets package + JavaScript's d3 library

ggplot2

- Most coherent graphics system
- Based on a “grammar” of graphics
- Easily customized using various “themes”

A bit about the grammar

- `ggplot()` creates a plot object
- `aes` describes a mapping of data to a visual element (e.g., color, shape, etc.)
- `geom_*()` displays a particular graphical representation
- `scale_*()` modifies the axes
- `coord_*()` modifies the coordinate system
- `theme_*()` modifies the overall look
- `facet_*()` creates small multiples
  - Some built-in to ggplot2
  - Some in an add-on package (ggthemes)

Ways to display a variable

- In a scatterplot, `geom point()` allows us to display a variable as:
  - X/Y Axis variable (via `aes(x=, y=)`)
  - Colour (via `aes(color=)`)
  - Alpha (via `aes(alpha=)`)
  - Size (via `aes(size=)`)
  - Shape (via `aes(shape=)`)
  - Facets (via `facet wrap()`)

Misleading charts (truncating the scale)

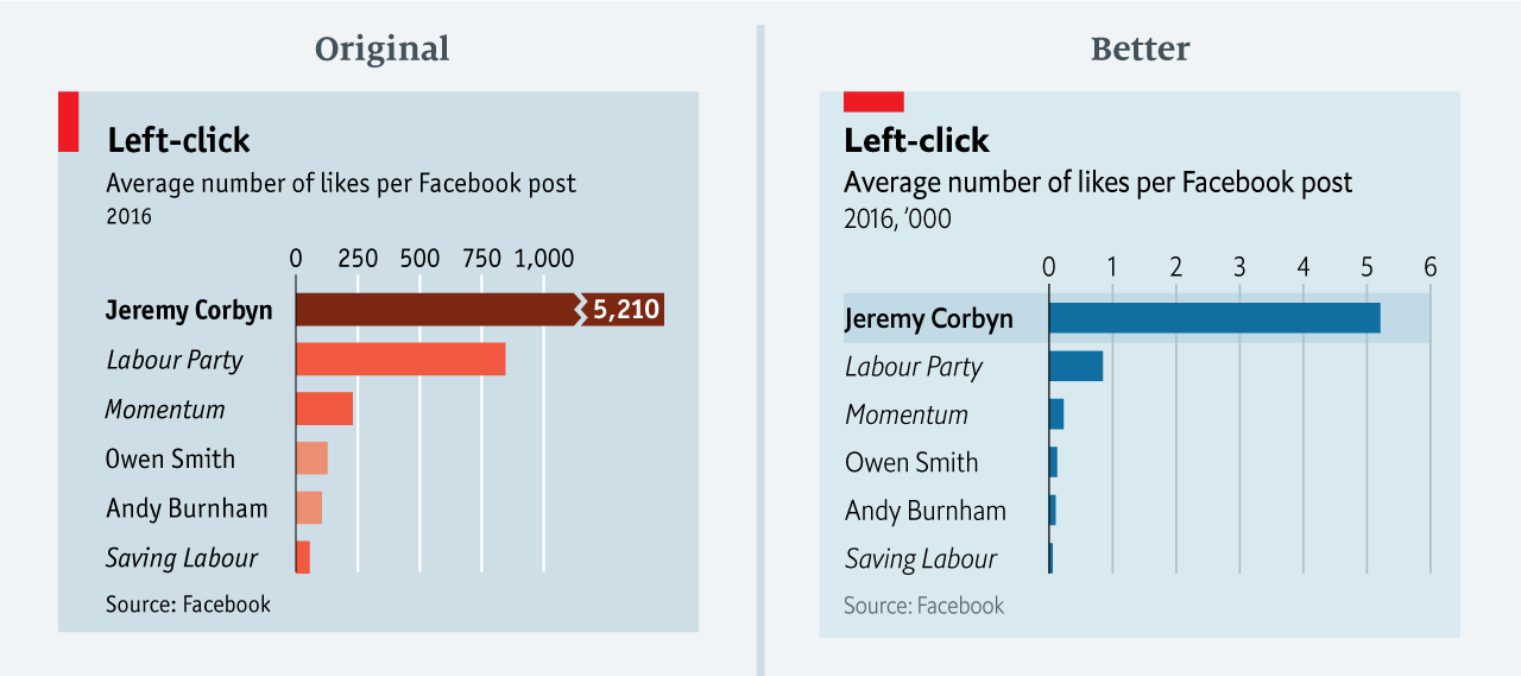


Figure 3: Mistake:Truncating the scale

Misleading charts (forced relationship)

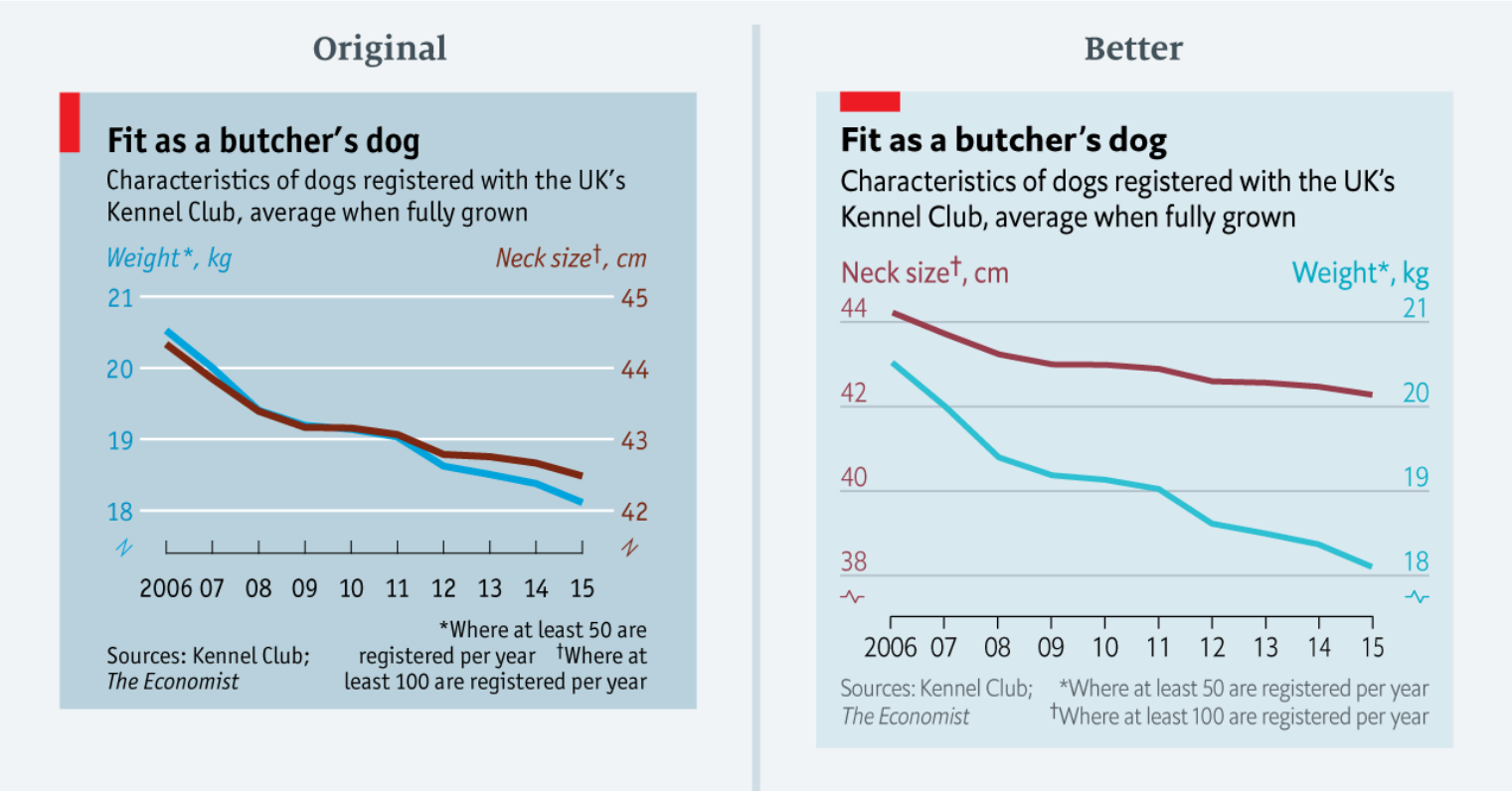


Figure 4: Mistake:Forcing a relationship by cherry-picking scales

Misleading charts (visualization method)

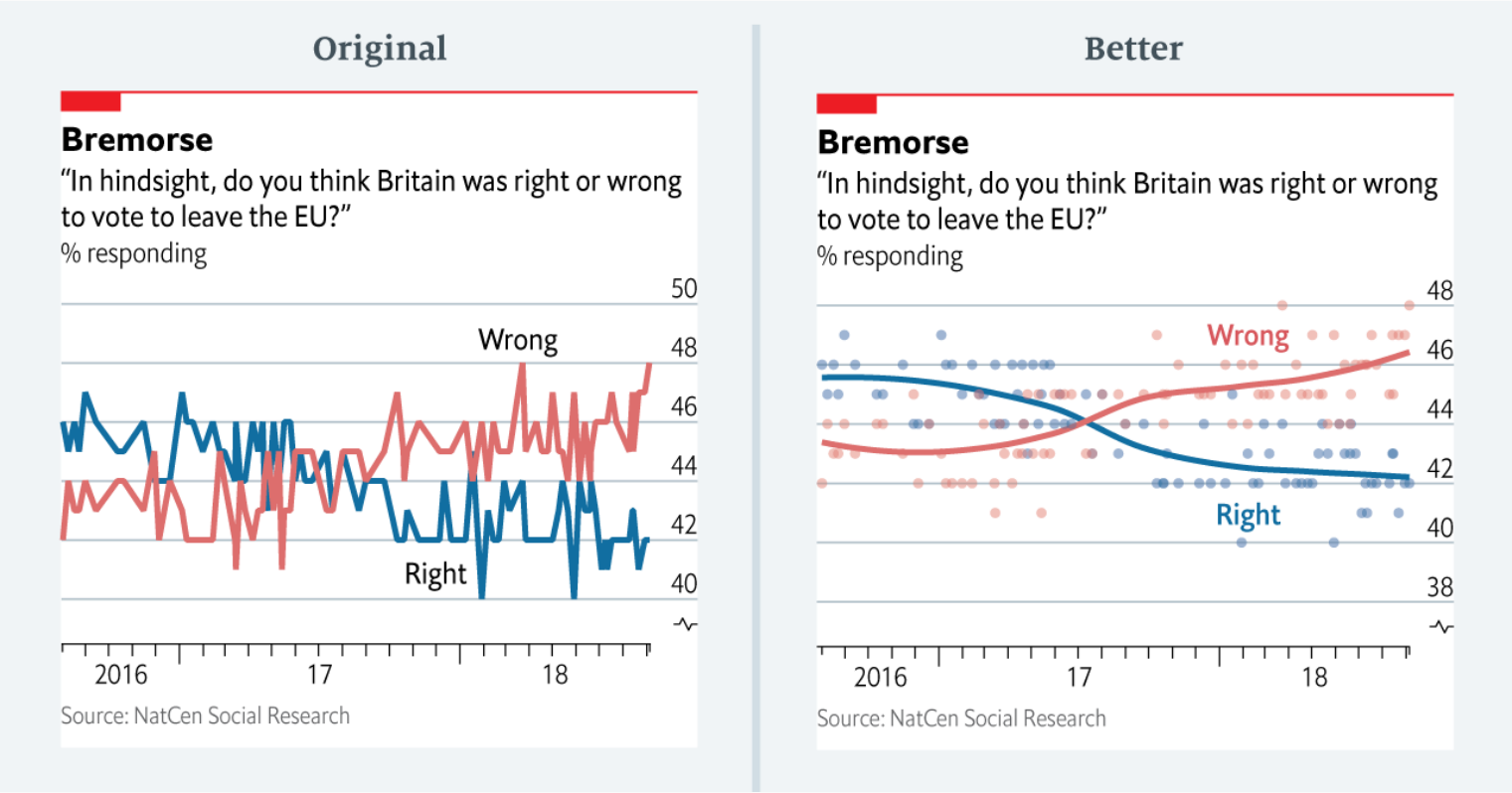
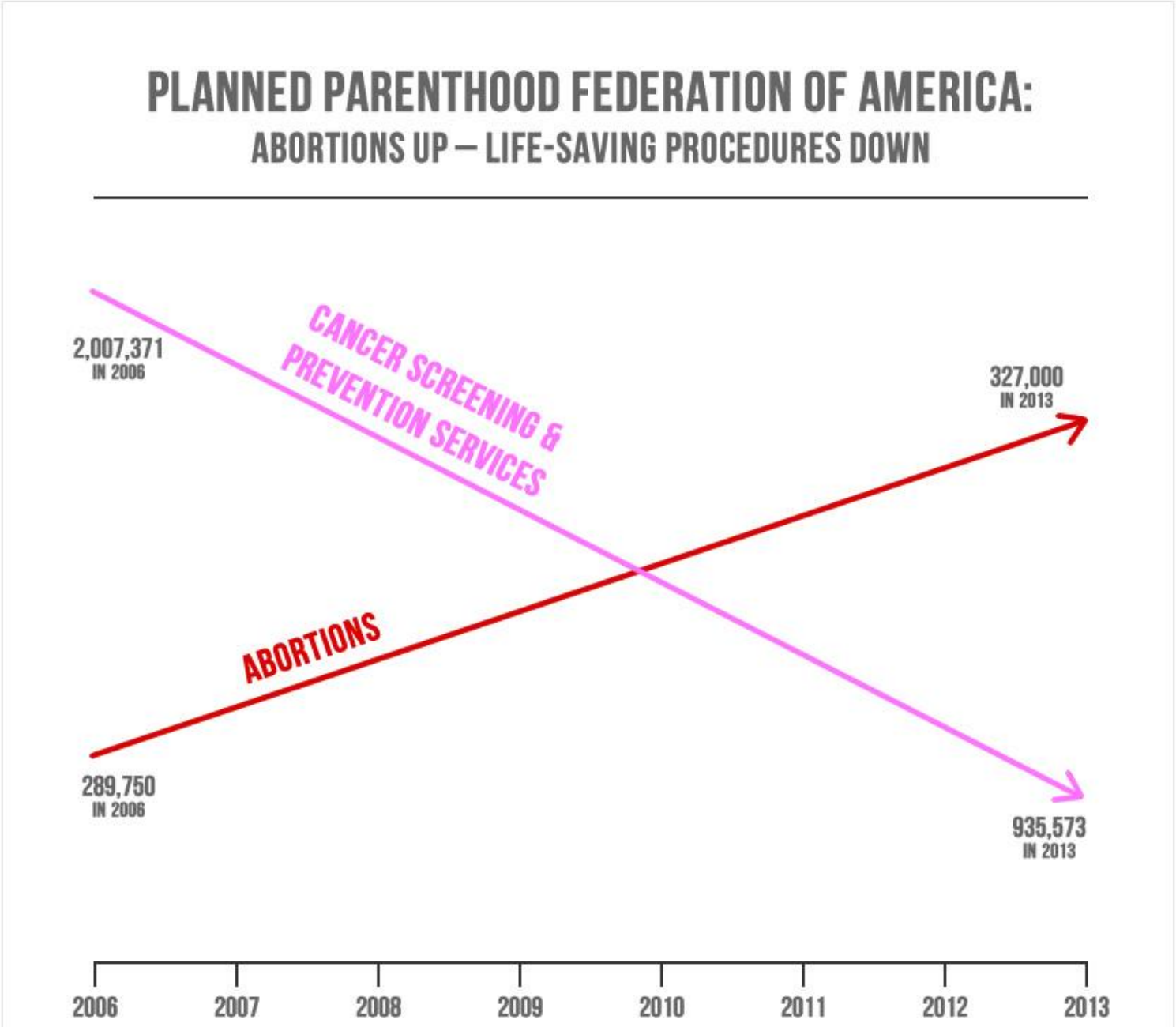


Figure 5: Mistake: Choosing the wrong visualisation method

Misleading charts



Misleading

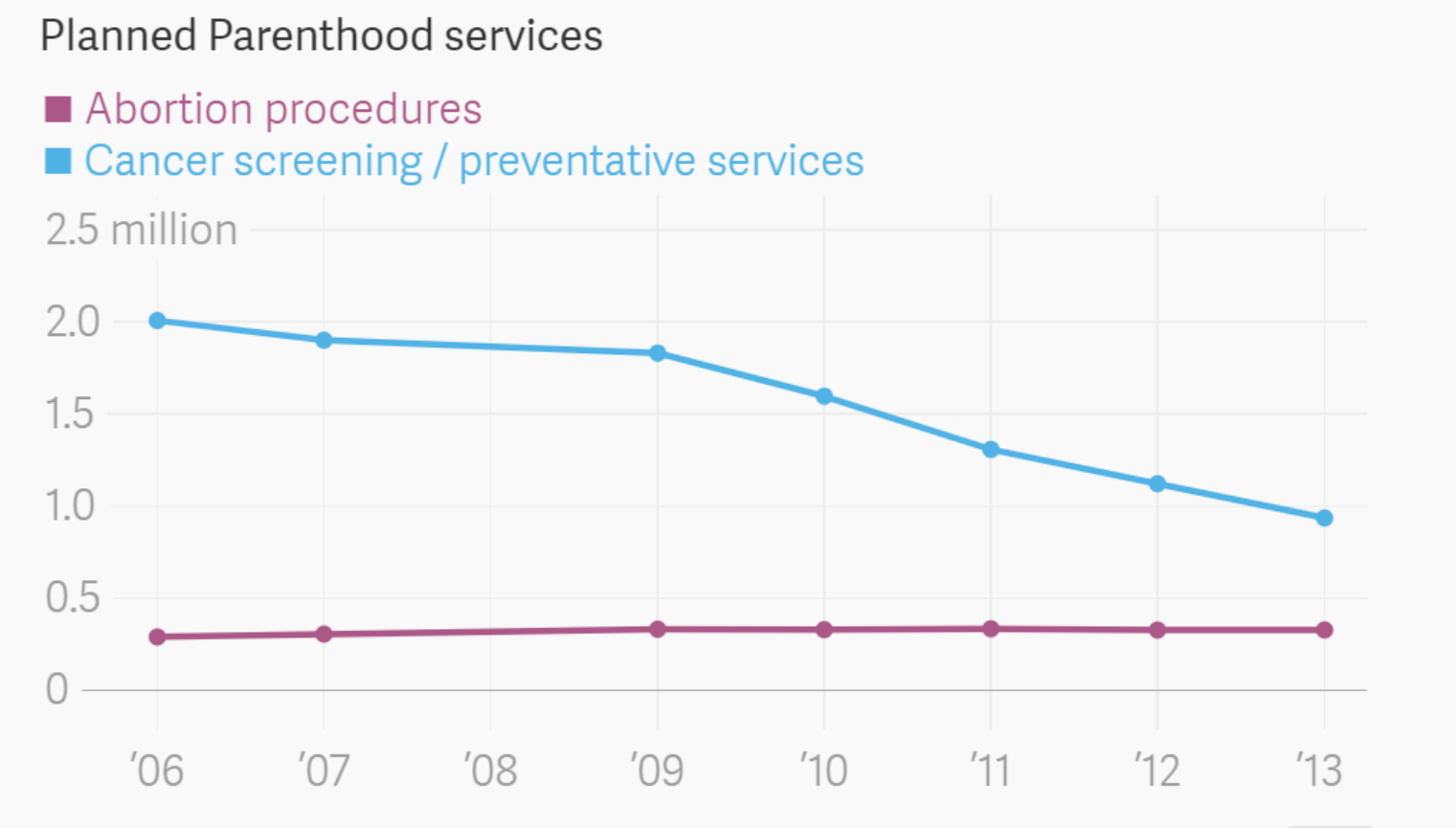


Figure 7: Mistake: Choosing the wrong visualisation method