



McCOMBS SCHOOL OF BUSINESS

Salem Center for Policy

Introduction to data visualization

David Puelz

Outline



The data viz hall of shame

How to do better



The data viz hall of shame

Plotting pitfalls



- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Plotting pitfalls



- **Axis trickery**
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Truncating the y-axis is sometimes ok



- When you're trying to emphasize change, rather than relative magnitude.
- When you're plotting data over time.
- When zero is not a sensible baseline for comparison.

Bottom line: use your judgment; don't mislead people; watch out for "little y lies."

Plotting pitfalls



- Axis trickery
- **Violations of basic math**
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Plotting pitfalls



- Axis trickery
- Violations of basic math
- **Nearly content-free figures**
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Plotting pitfalls



- Axis trickery
- Violations of basic math
- Nearly content-free figures
- **Gratuitous chartjunk**
- Poorly chosen 3D graphics
- Bad design choices

Plotting pitfalls



- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Plotting pitfalls



- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- **Bad design choices**



How to do better

Good plots ...



- Make relevant comparisons and avoid irrelevant ones. Are truthful about magnitude.
- Minimize distractions.
- Introduce relevant conditioning variables.
- Stay firmly 2D (unless visualizing an actual 3D object, like a mountain or a building).
- Avoid meaningless variation (e.g. color differences that don't encode information).
- Have clear axis labels and informative titles/annotations.
- Highlight both overall trends and individual variation.
- Sometimes aren't plots at all – they're just tables.
- Are sensitive to the knowledge and limitations of the viewer.

Start with what's simple and effective



Simple and effective plots:

- Density plots (histograms, boxplots, violin plots)
- Barplots: to show summary statistics (means/counts/proportions) by group
- Scatter plots: to show how two numerical variables are related
- Line graphs: to show how one numerical variable changes continuously versus another (e.g. time)

Simple and effective strategies for enriching these plots:

- Variation in size, shape, and/or color (but be colorblind-friendly and don't go crazy)
- Faceting: showing the same basic plot across multiple conditions
- Labels