Introduction to Data Visualization

Author: Nicholas G Reich

Key principles of effective data graphics

- ▶ Know your audience
- ▶ "**Show** the data"
- ▶ "Encourage the eye to **compare** different pieces of data"
- ▶ **Simplify** by maximizing the "data-ink ratio."
- ▶ Leverage color, shapes, facets to highlight multivariate data.
- ▶ Annotate your figures with context.

Visual cues

Graphical elements that draw attention

The choice of which visual cues you use inform the story that you are able to convey and the points you can highlight.

Table 2.1: Visual cues and what they signify.

Visual Cue	Variable Type	Question
Position	numerical	where in relation to other things?
Length	numerical	how big (in one dimension)?
Angle	numerical	how wide? parallel to something else?
Direction	numerical	at what slope? in a time series, going up or down?
Shape	categorical	belonging to which group?
Area	numerical	how big (in two dimensions)?
Volume	numerical	how big (in three dimensions)?
Shade	either	to what extent? how severely?
Color	either	to what extent? how severely? *what group?

Visual cues: position (numerical)

Where are the data in relation to each other?
e.g. points and axis alignment.

Visual cues: Length (numerical)

How big (in one dimension)?
e.g. bars, lines (aligned), lines(unaligned) ...

Visual cues: Angle (numerical)

How wide? Parallel to something else?

e.g. lines, pie wedges, ...

Visual cues: Direction/slope (numerical)

Up or down? At what slope?

e.g. lines, time-series, ...

Visual cues: Shape (categorical)

Belonging to which group?
e.g. points

Visual cues: Area/volume (numerical)

How big (in 2/3 dimensions)?

e.g. circles, squares

Visual cues: Shade/intensity (categorical or numerical)

To what extent? How severely?
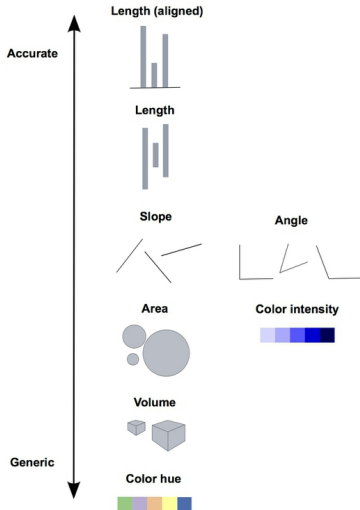
e.g. points, lines, ...

Visual cues: Color (categorical or numerical)

Belonging to which group? To what extent? How severely?
e.g. points, lines, tiles …

Research on perception of cues

In 1980s, Cleveland and McGill ran experiments to measure accuracy of human perception based on different visual cues.
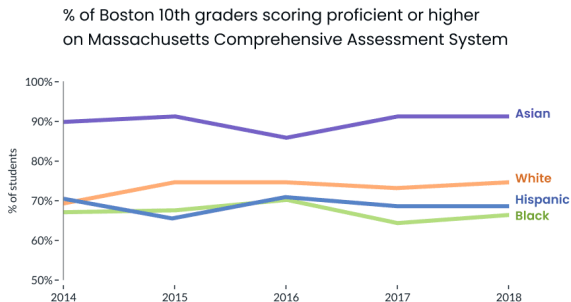

(Figure credit: Peter Aldhous)

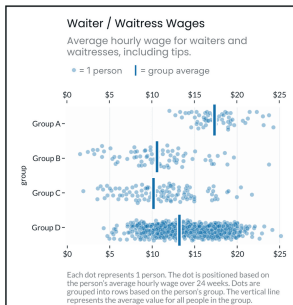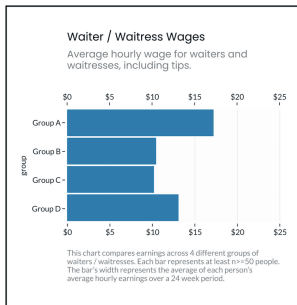# Research on reducing stereotyping from data viz

Deficit thinking

"by emphasizing direct comparisons between groups, [data visualizations] create the impression that the groups with the worst outcomes (often marginalized groups) are personally deficient relative to the groups with the best outcomes (often majority groups)."

% of Boston 10th graders scoring proficient or higher on Massachusetts Comprehensive Assessment System

(Figure credit: Eli Holder and Pieta Blakely)

Research on reducing stereotyping from data viz

Show the variability!

The insight by this recent work is that showing the individual-level variability behind the individual summary numbers might reduce the degree to which consumers of data viz engage in deficit thinking.

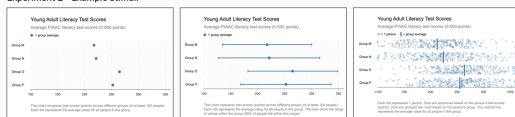(Figure credit: Eli Holder and Pieta Blakely)

Research on reducing stereotyping from data viz

Results from experiment show modest decreases in deficit thinking

"Personal attribution with Prediction Intervals was significantly lower than with Dot Plots."
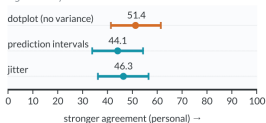
However, I add that the differences are not that large in real terms (44% vs   51%), and the population studied was data-literate volunteers from Mechanical Turk (not terribly representative sample).

(Figure credit: Eli Holder and Pieta Blakely)

Breakout rooms

For each of the following graphics, work in your breakout rooms to complete the note-catcher assignment.

Front-runners in The 'Bachelor' Or 'Bachelorette'[1]

A rose for every season

The path of every winner on every season
of the "Bachelor" and "Bachelorette"

① Each petal is a week

Week 2
3
4
6 5

② Each section of the petal is a contestant

8 contestants

2 contestants

③ Size indicates a contestant's weighted number of dates through that point in the season ...

Fewer weighted dates

More

④ ... as does order.

Underdog and winner

More weighted dates

Favorite

The Bachelorette 💔 Broke up ❤️ Still together

❤️ Trista + Ryan
2003

💔 Meredith + Ian
2004

❤️ Jen + Jerry
2005

💔 DeAnna + Jesse
2008

💔 Jillian + Ed
2009

💔 Ali + Roberto
2010

❤️ Ashley + J.P.
2011

💔 Emily + Jef
2012

❤️ Desiree + Chris
2013

💔 Andi + Josh
2014

❤️ Kaitlyn + Shawn
2015

💔 JoJo + Jordan
2016

Contestant Nick was rejected by both
Andi and Kaitlyn in their final ceremonies.
He went on to be the bachelor in 2017.

For nearly all seasons, dates did not begin until week two.
Dates are weighted so that a one-on-one counts as one date,
a two-on-one counts as 0.5 dates, etc.

FiveThirtyEight

BASED ON DATA FROM ABC, WIKIPEDIA AND BACHELOR NATION WIKIA

[1] from FiveThirtyEight

"Ageing on Facebook"

Graphic detail Ageing on Facebook

Teenagers are avoiding Facebook, as older users flock to it

Share of Americans using platforms at least once per month, estimate, by age group