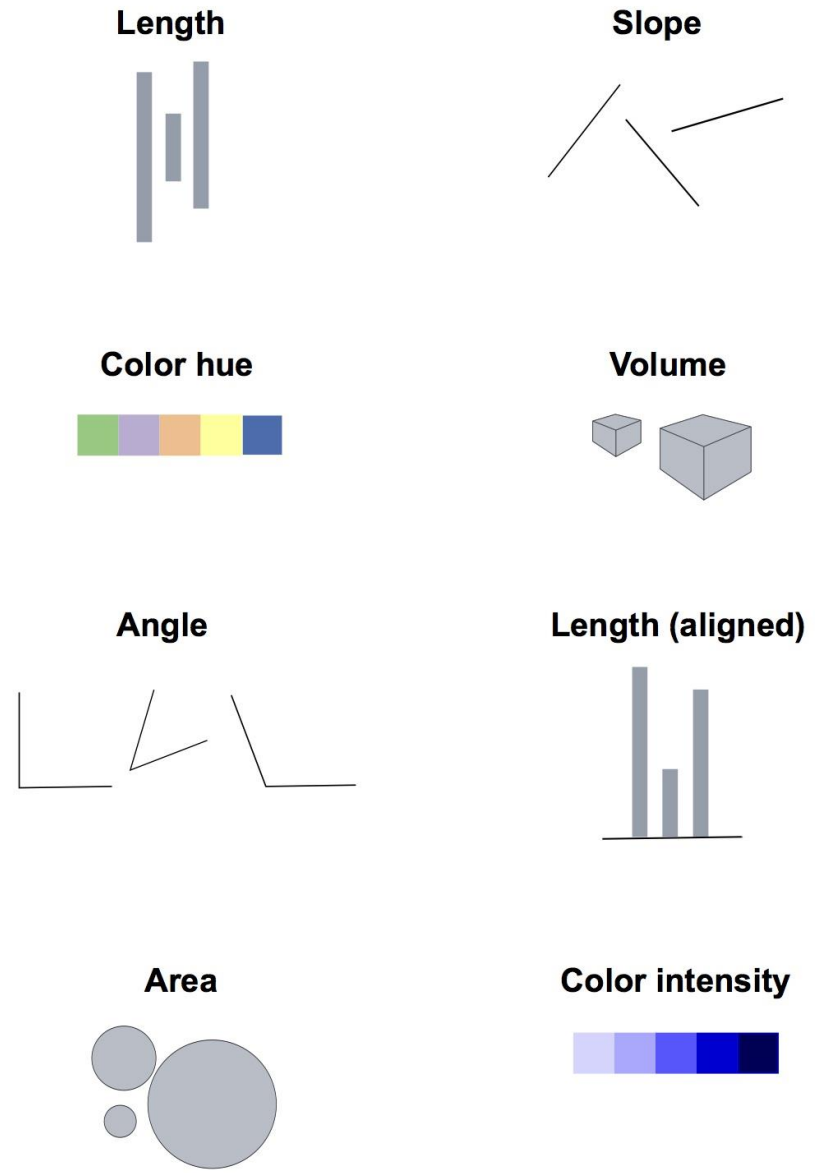
Data visualization: basic principles

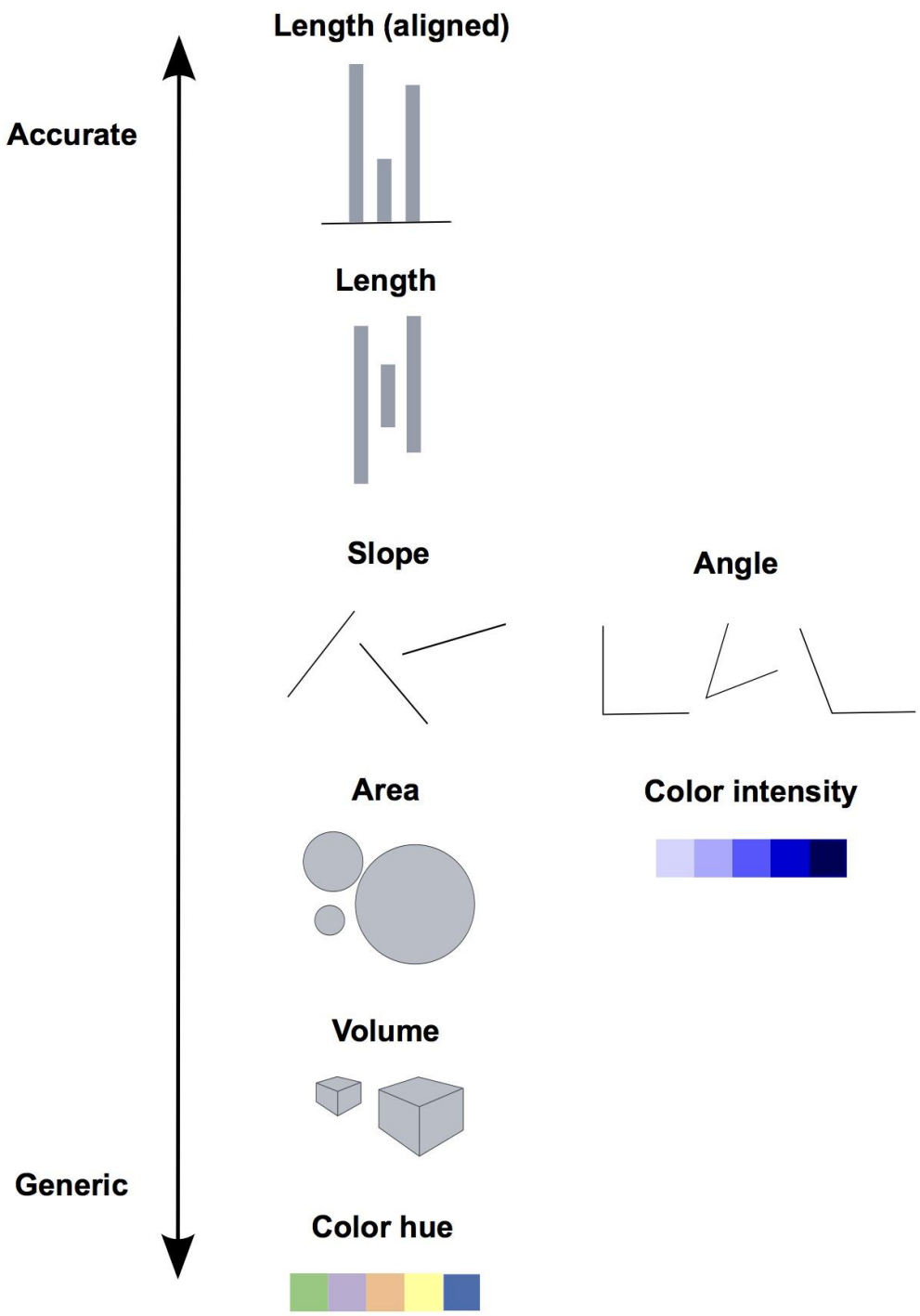
Peter Aldhous

@paldhous

Visualization: encoding data by visual cues



Our brains do not treat those cues equally!



Design for the human brain!

What type of chart should I use?

Distribution

Relationship

Comparison

What do you want to show?

Connection

Composition
(parts of the whole)

Location

Distribution

Relationship

Comparison

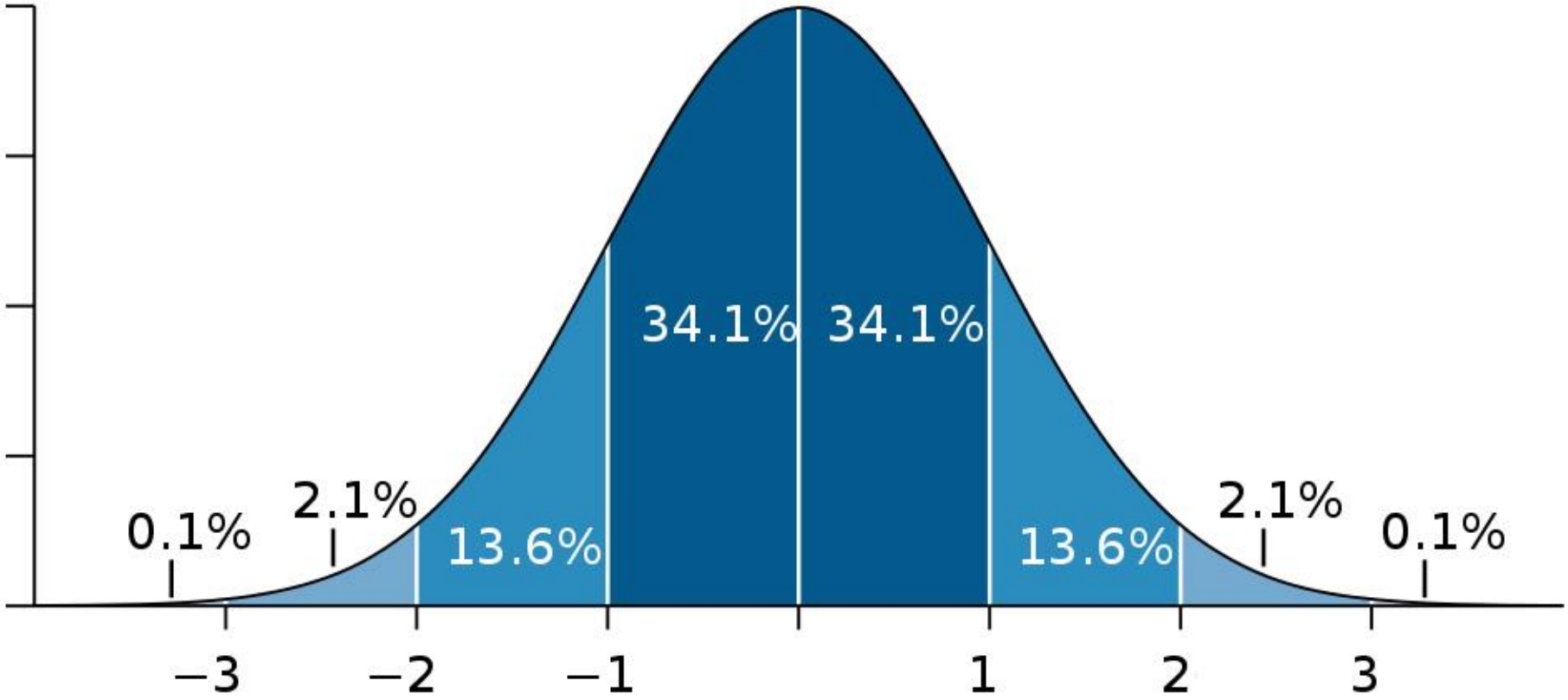
What do you want to show?

Connection

Composition
(parts of the whole)

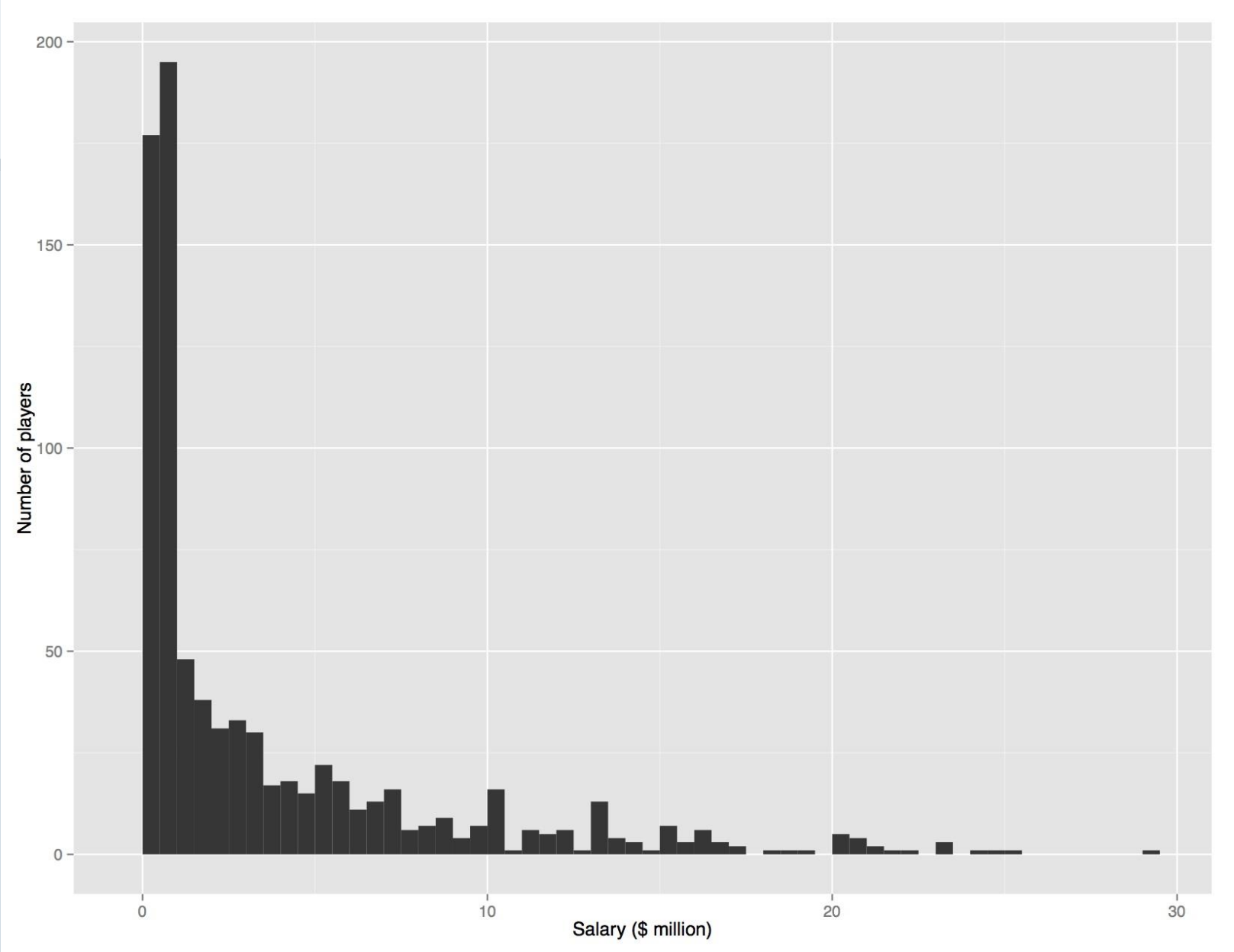
Location

Consider the distribution



Wikimedia Commons

Consider the distribution



Distribution

Relationship

Comparison

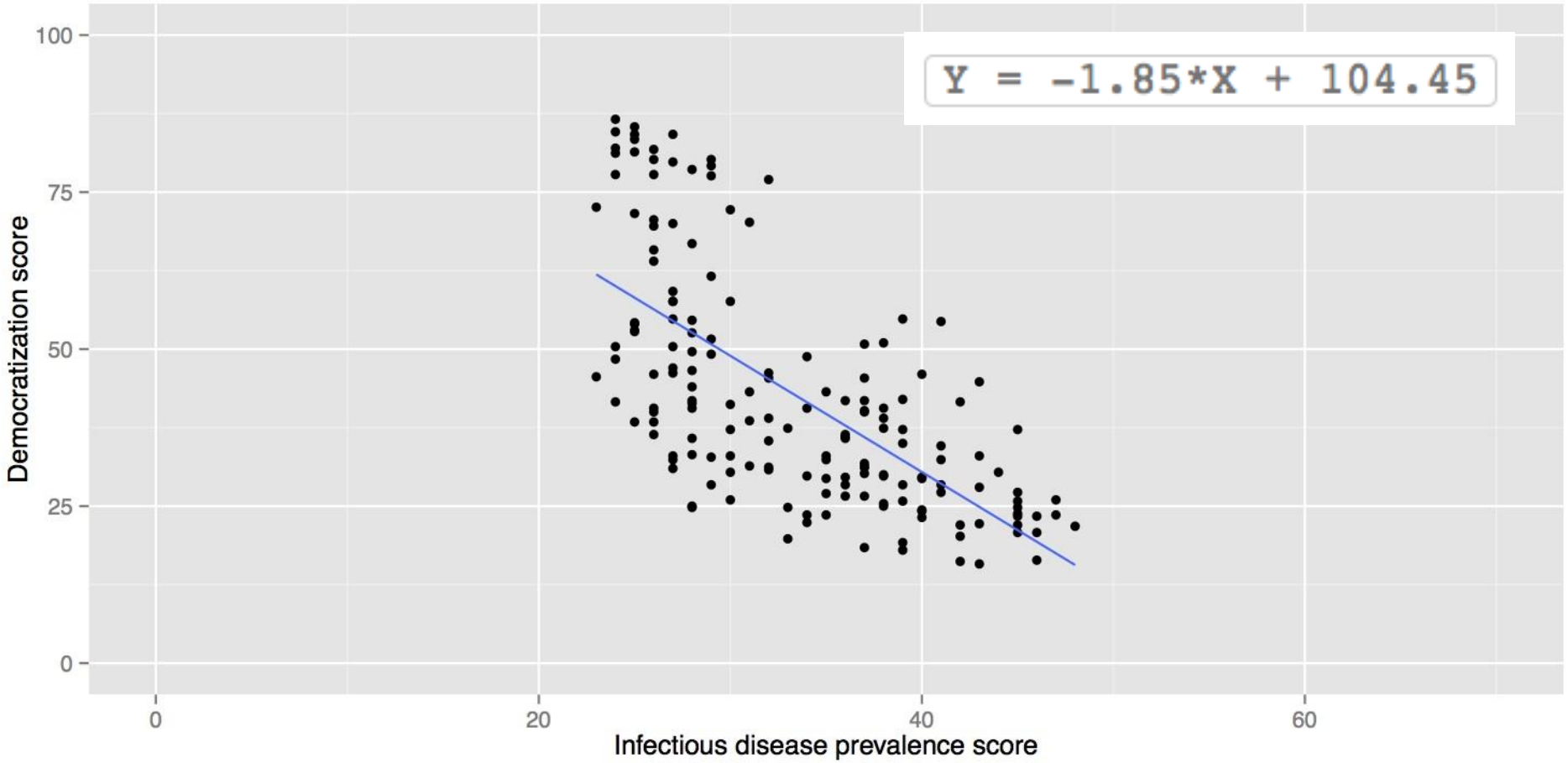
What do you want to show?

Connection

Composition
(parts of the whole)

Location

Relationships between variables: scatter plots and trend lines



Relationships between variables: correlation and its pitfalls

I can’t stress enough:

- Correlation does not prove causation
- Beware of “lurking” variables

Distribution

Relationship

Comparison

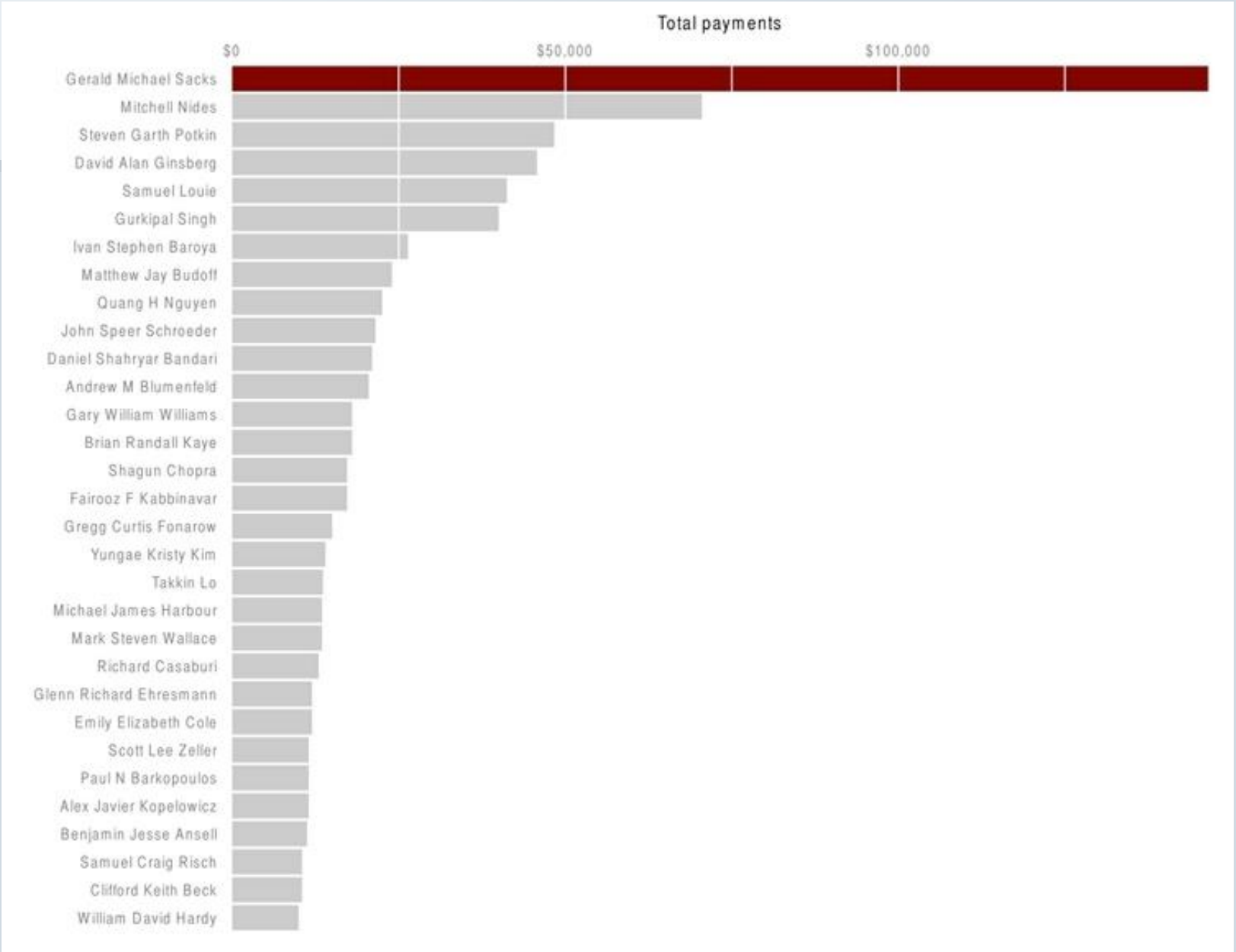
What do you want to show?

Connection

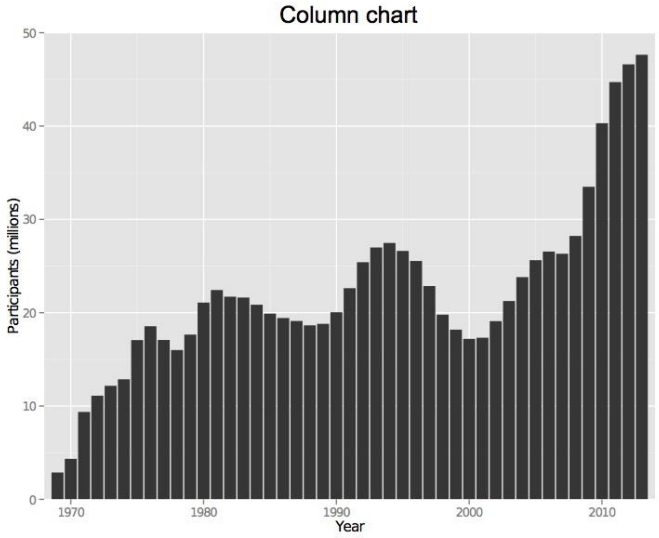
Composition
(parts of the whole)

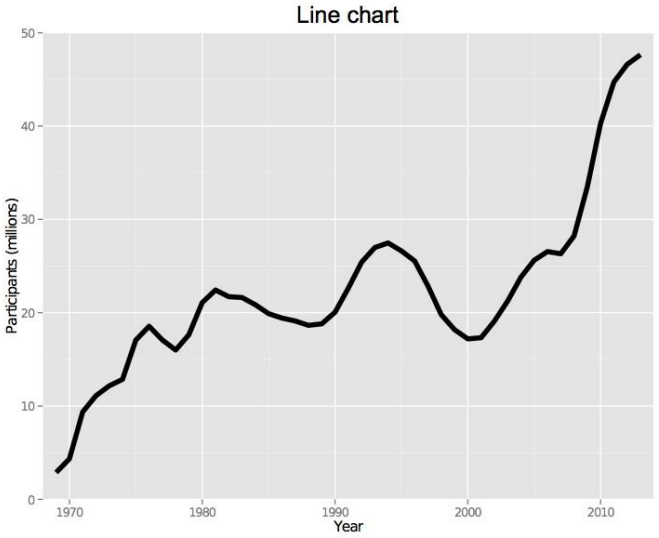
Location

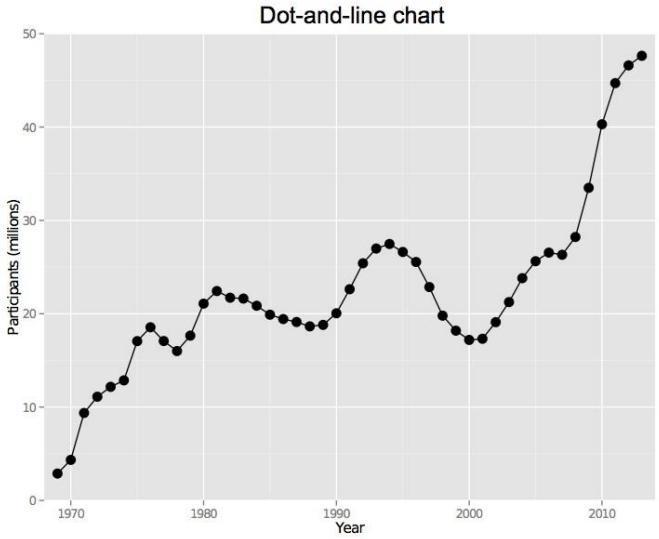
Simple comparisons: bars and columns

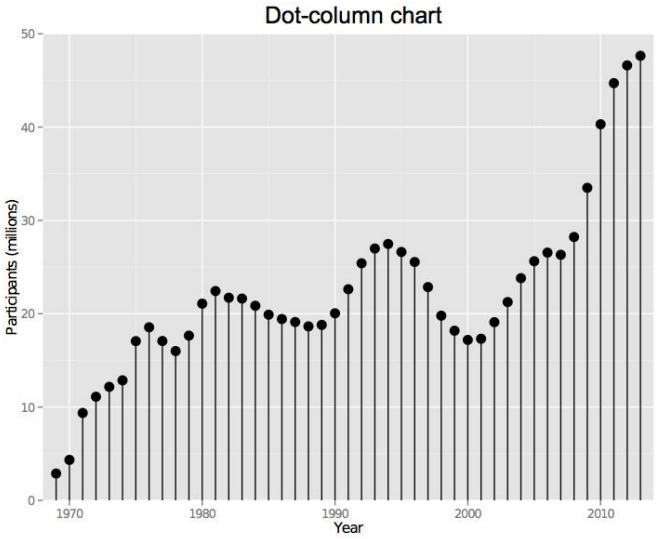


Comparisons: change over time

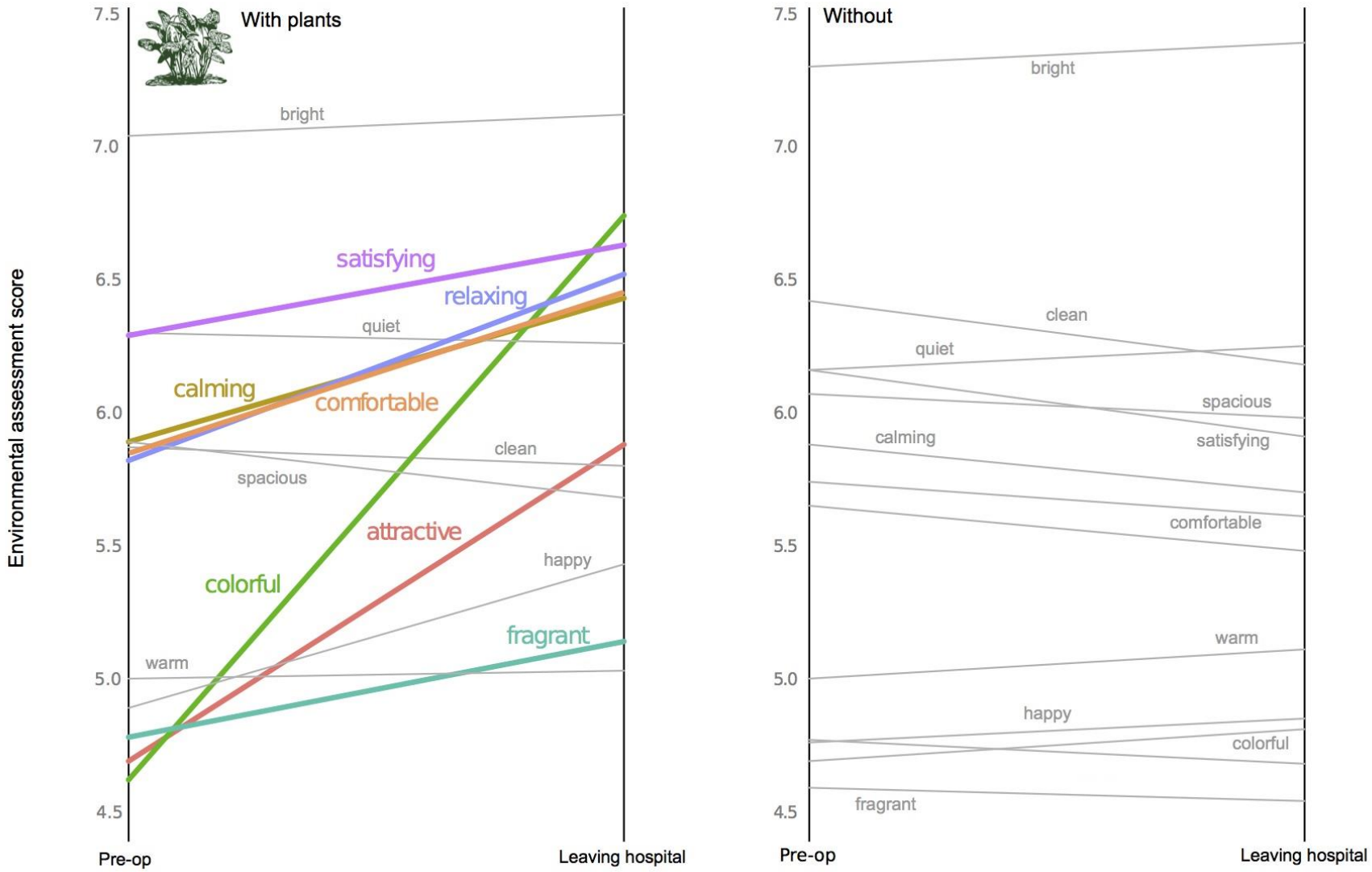








Two points in time, many categories



Distribution

Relationship

Comparison

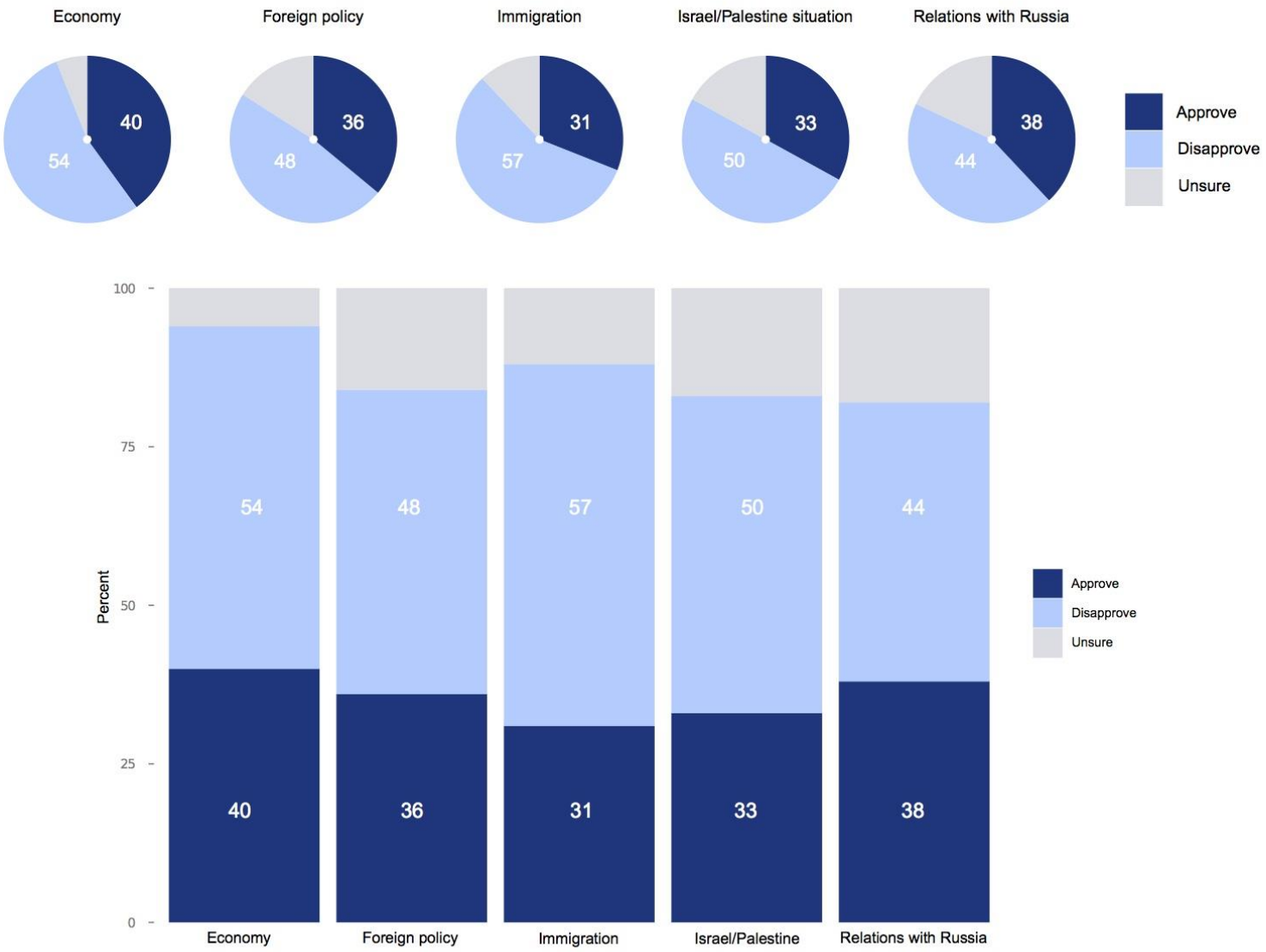
What do you want to show?

Connection

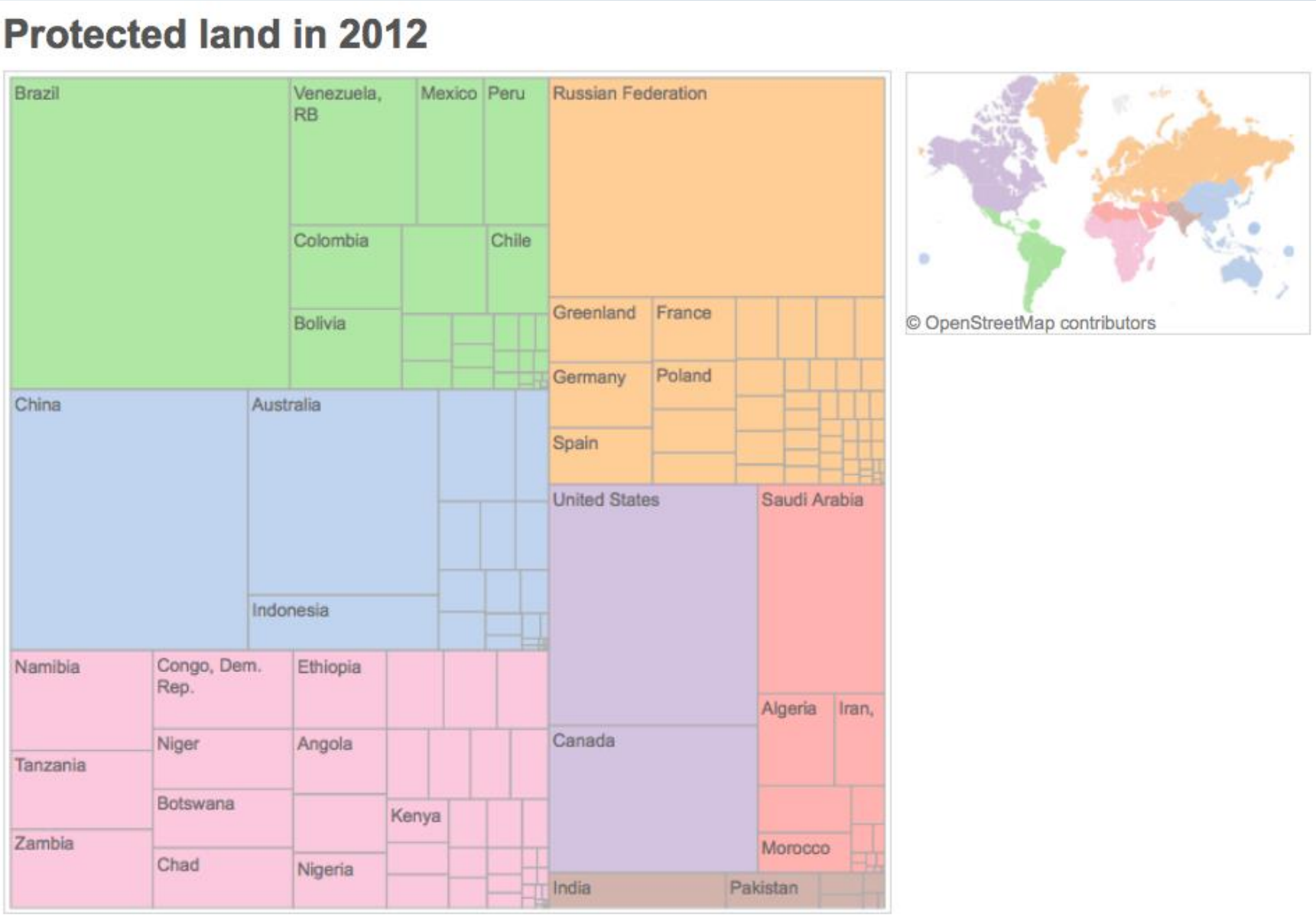
Composition
(parts of the whole)

Location

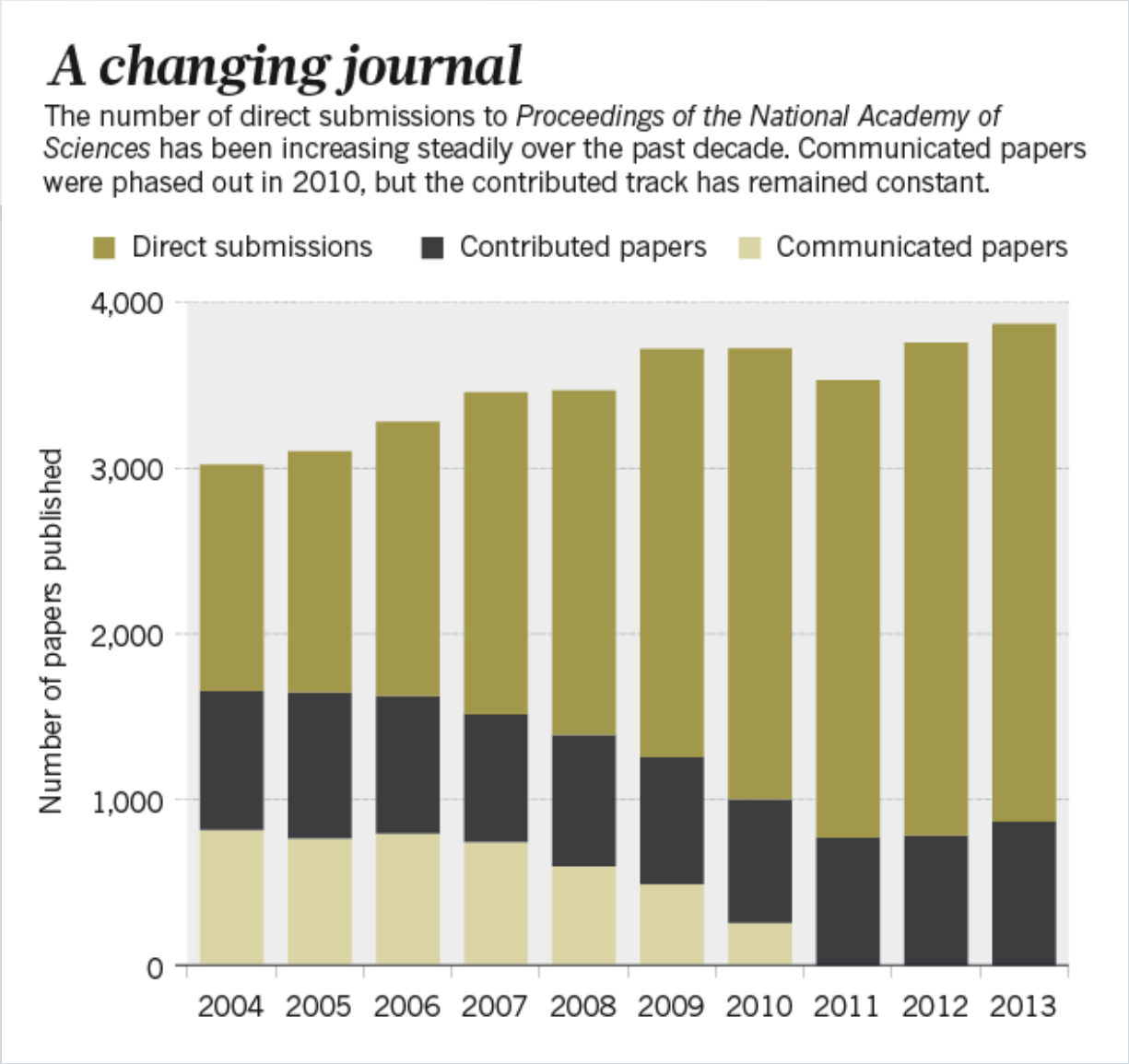
Composition: parts of the whole



Composition: parts of the whole

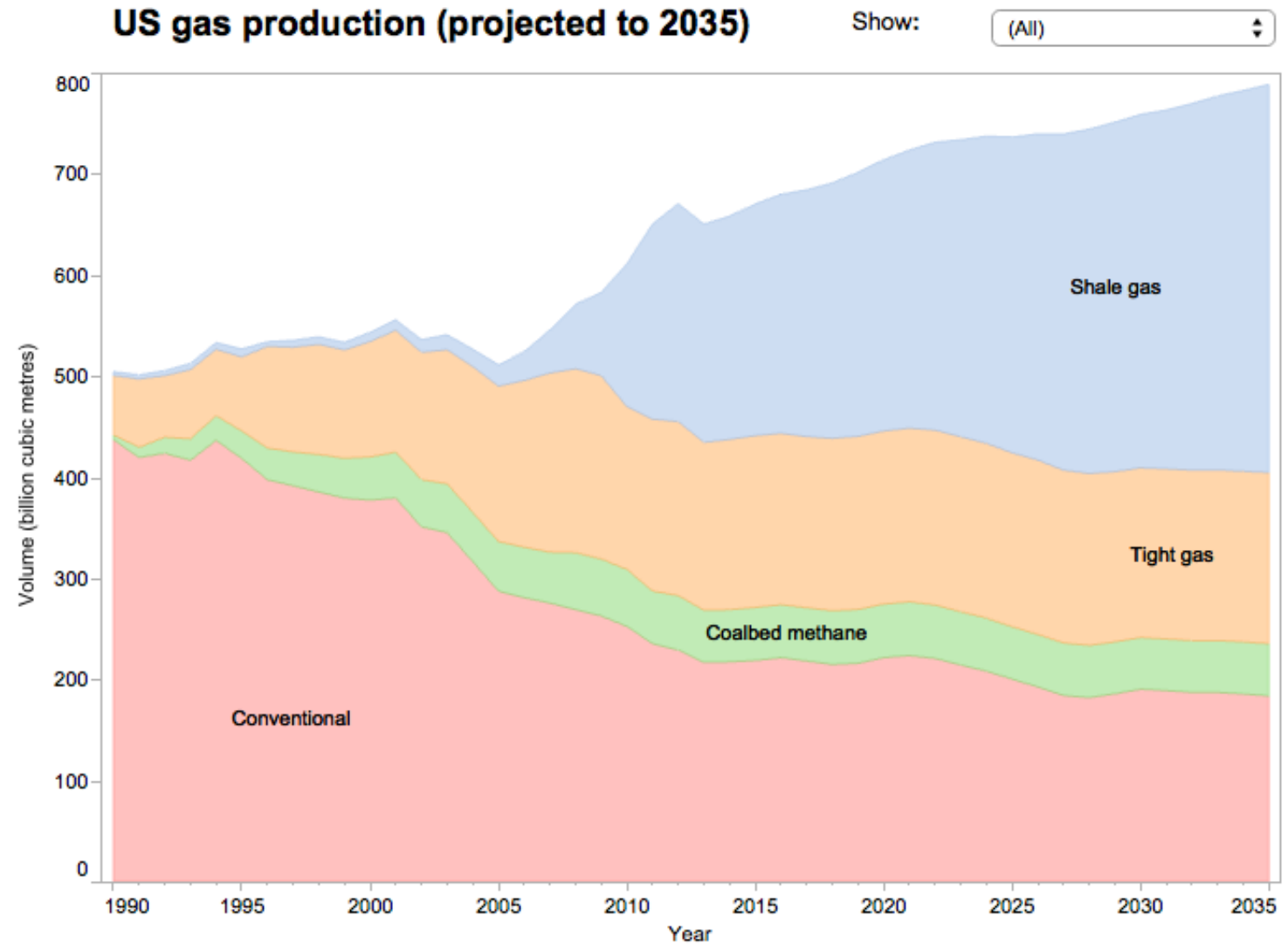


Composition: change over time



Nature

Composition: change over time



Source: U.S. Energy Information Administration

Distribution

Relationship

Comparison

What do you want to show?

Connection

Composition
(parts of the whole)

Location

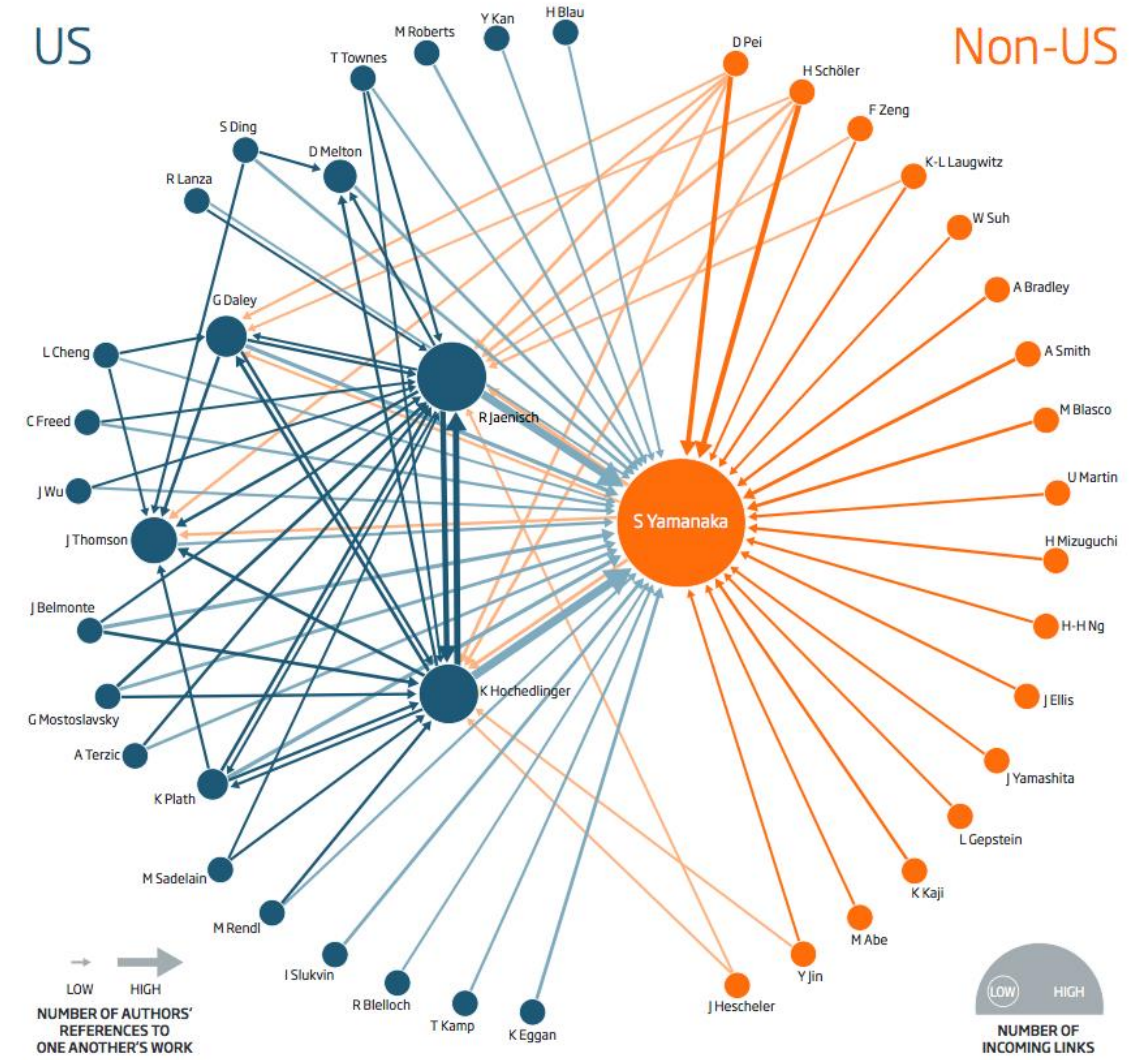
Connection: network graphs

New Scientist

THE STEM CELL WARS

When a Nobel prize is up for grabs, do scientists across the globe compete on a level playing field? **Peter Aldhous** investigates

The most influential players in cellular reprogramming are revealed by recording how many times the scientists have referred to each other's work. Each link shows where one researcher cited another four or more times in papers in leading journals (for analysis, see "The strongest link", below right)



Distribution

Relationship

Comparison

What do you want to show?

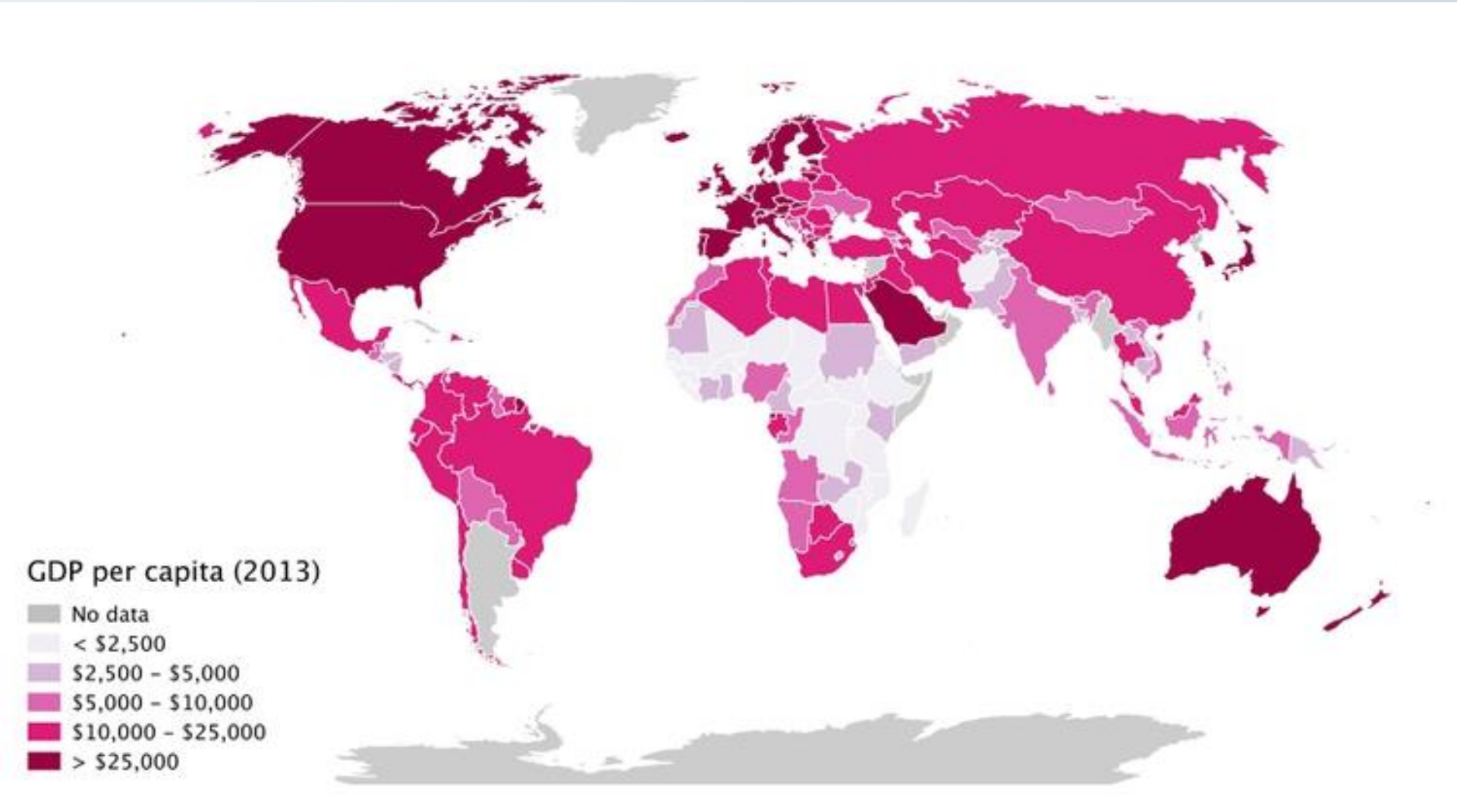
Connection

Composition
(parts of the whole)

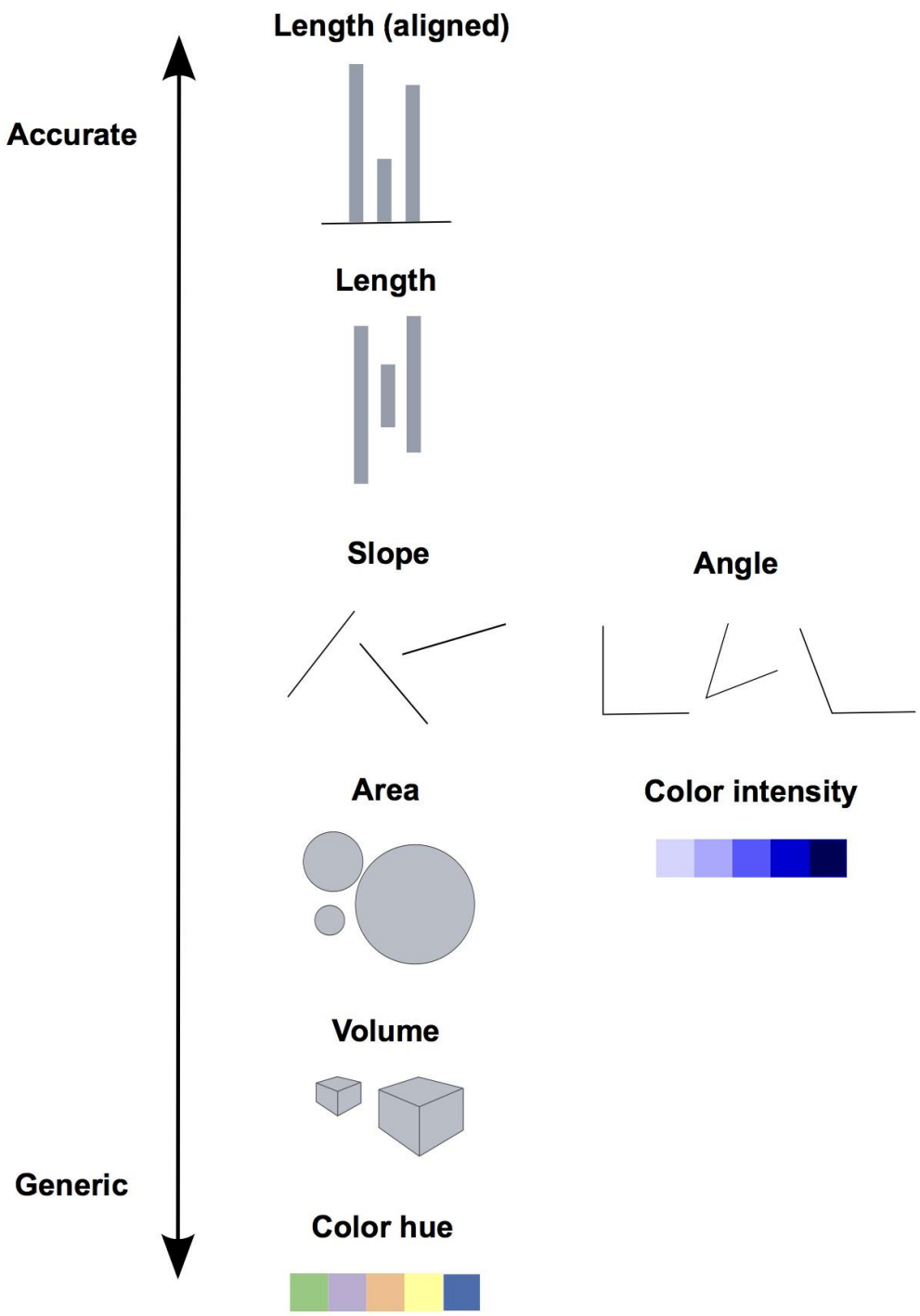
Location

Location plus data: scaled circles

Location plus data: choropleth maps



Remember the perceptual hierarchy of visual cues!



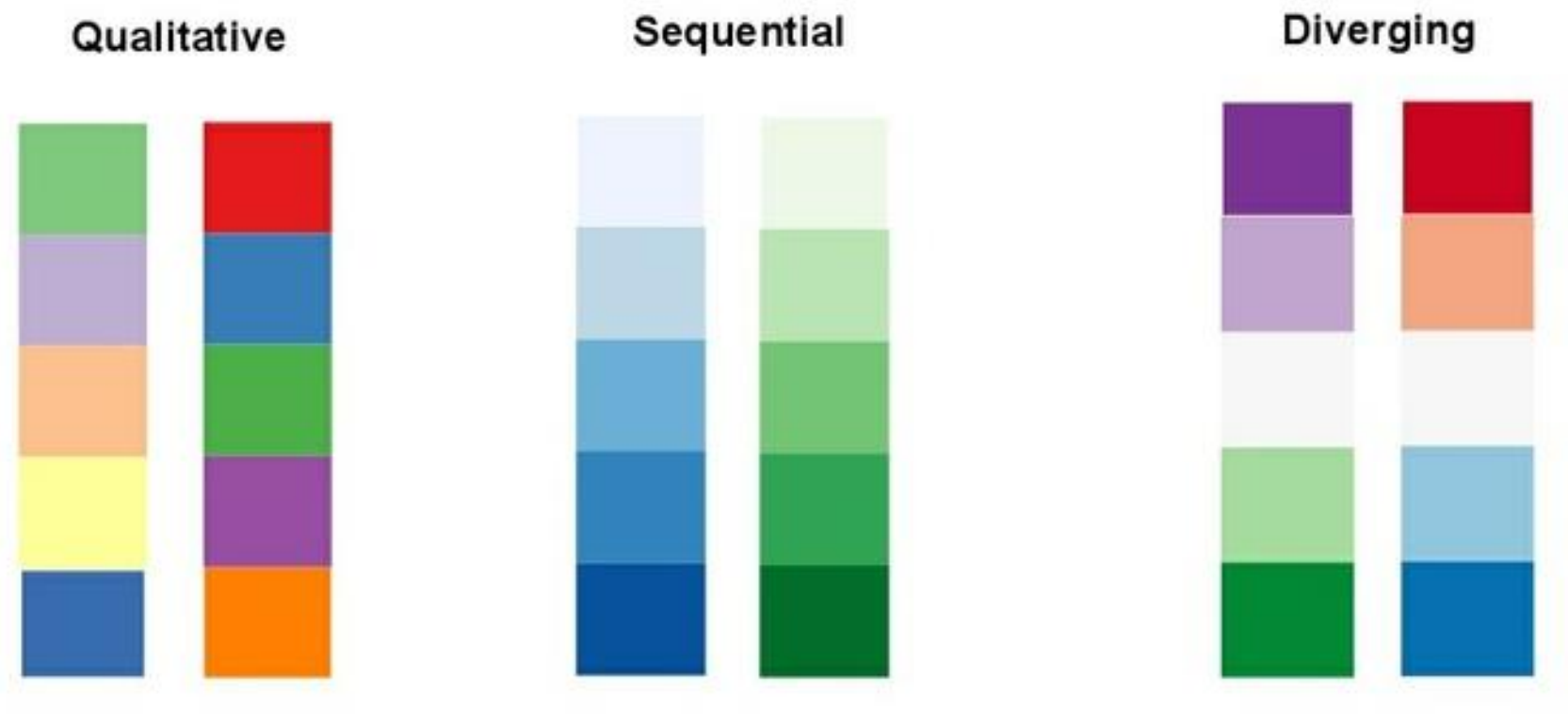
So ask yourself: Is a map the best way to tell the story?

Using color effectively

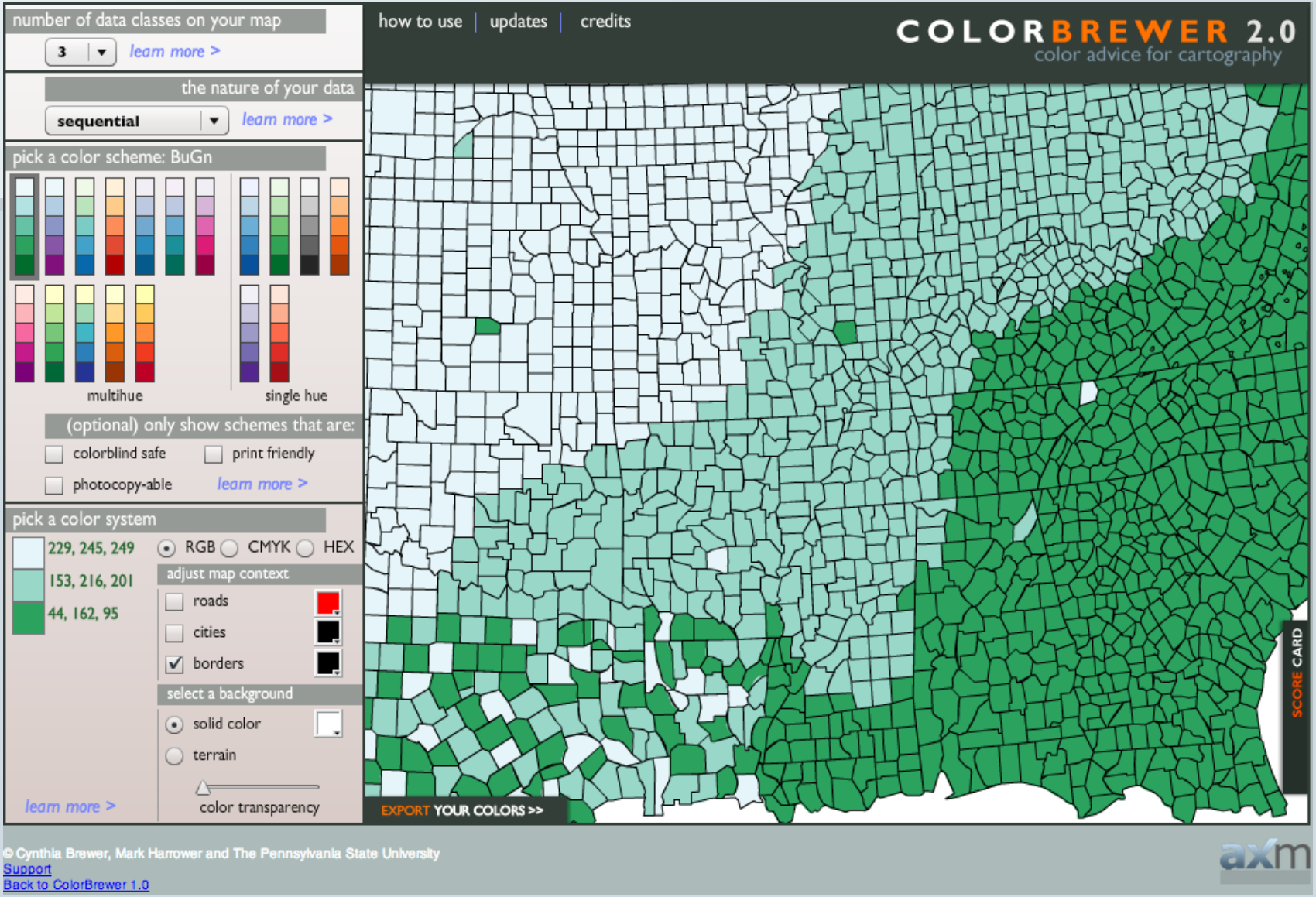
The color wheel

Wikimedia Commons

Using color: fit to your data

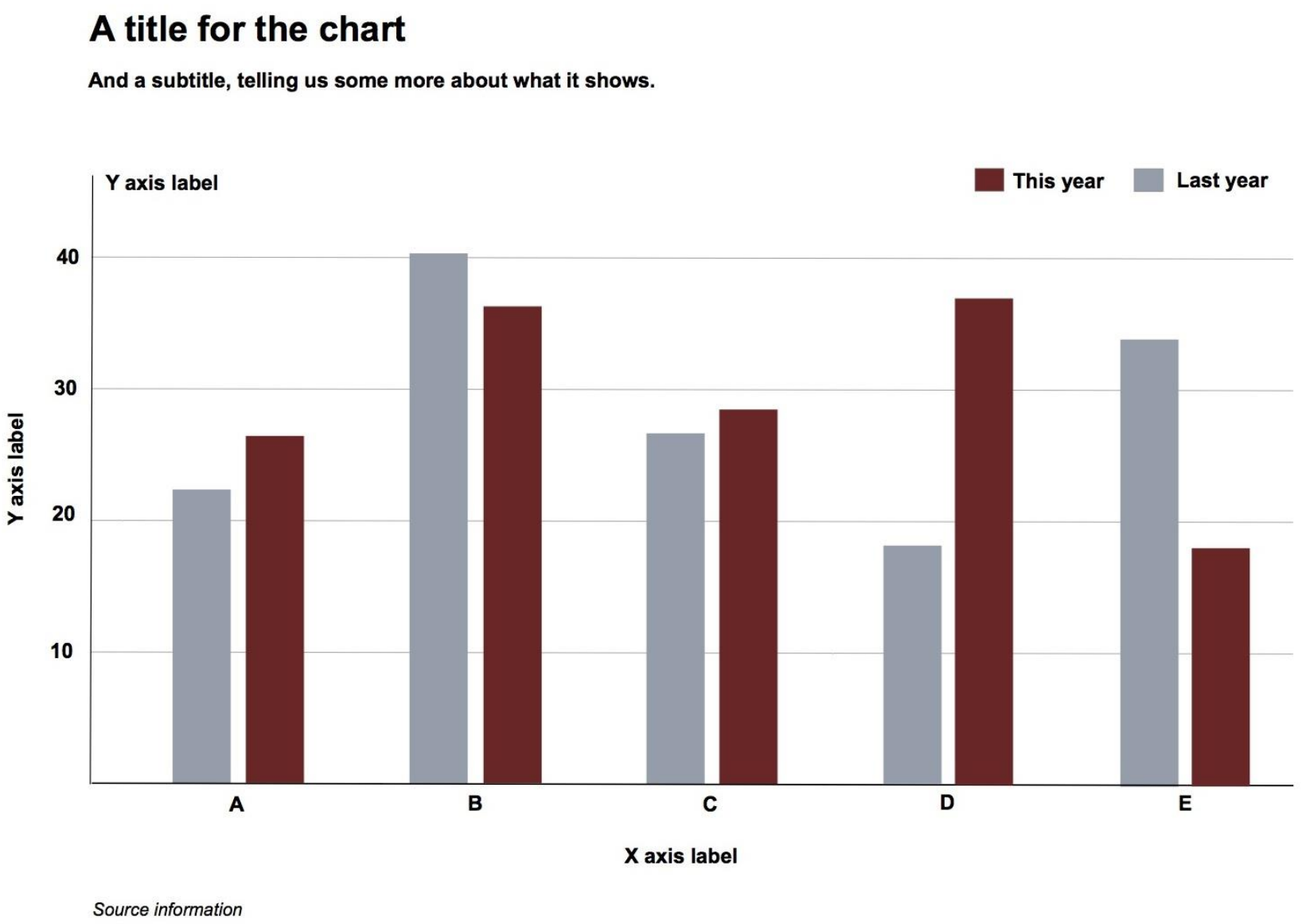


ColorBrewer is your friend!

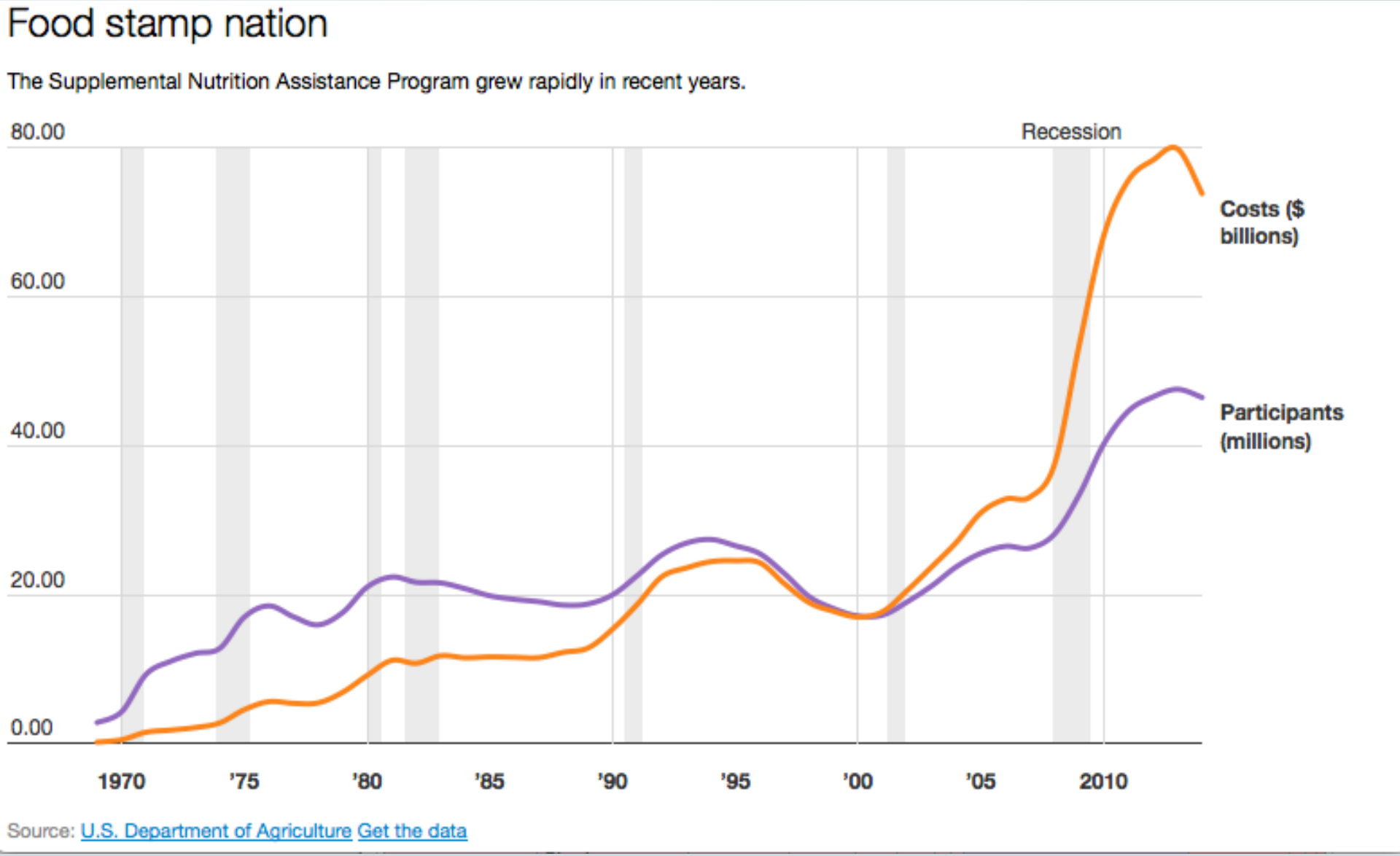


http://www.colorbrewer2.org/

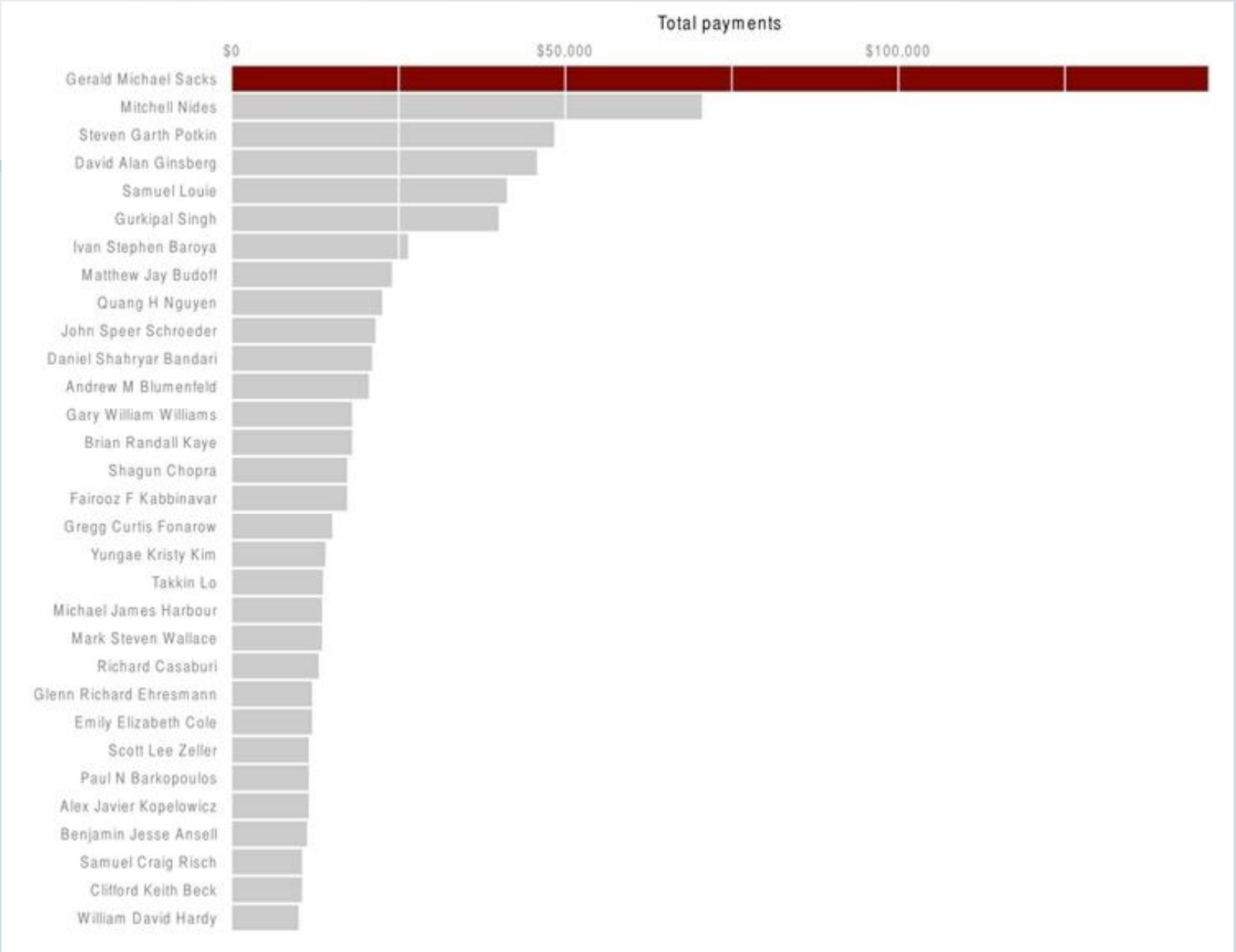
Chart furniture



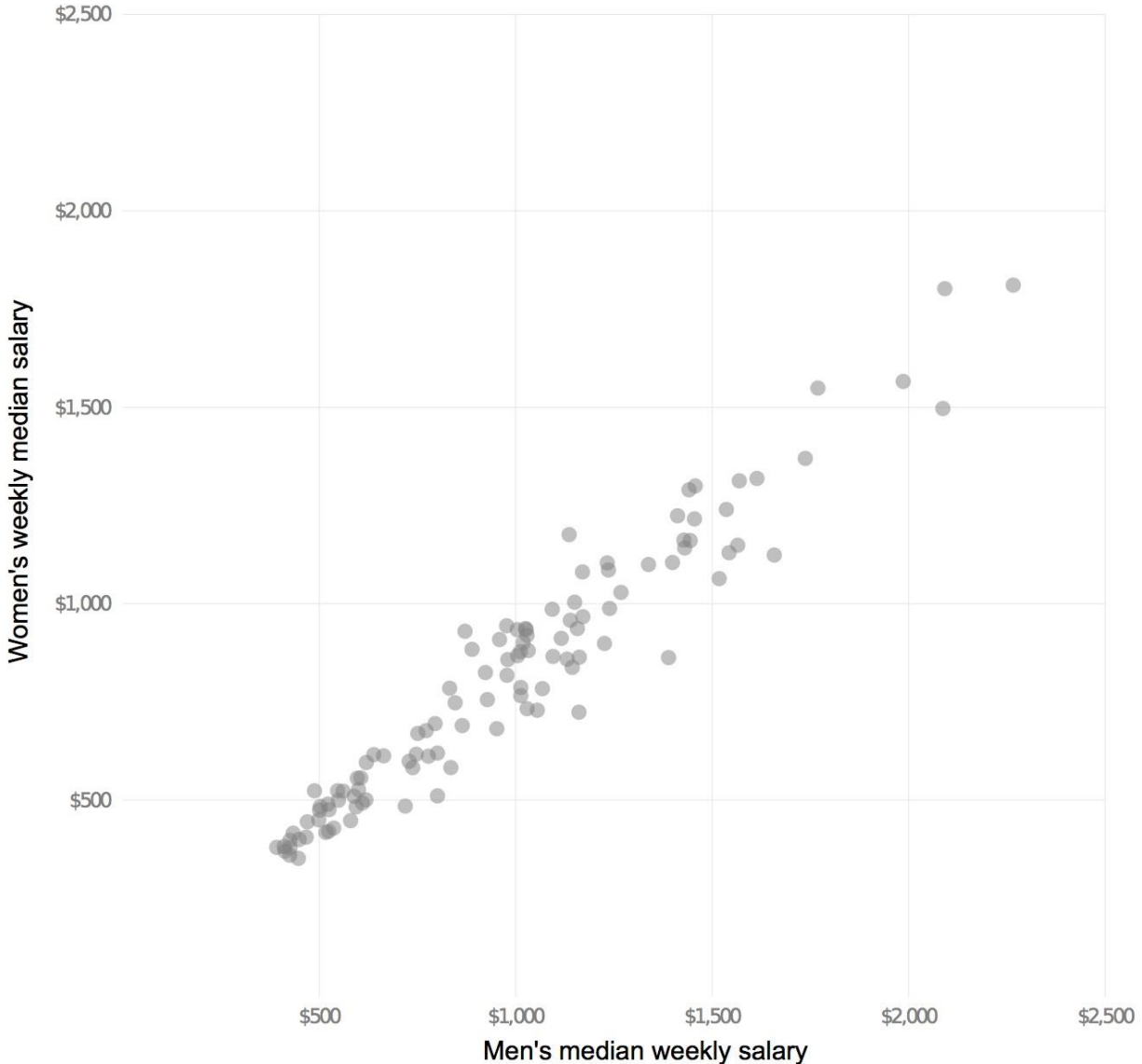
Using chart furniture to encode data



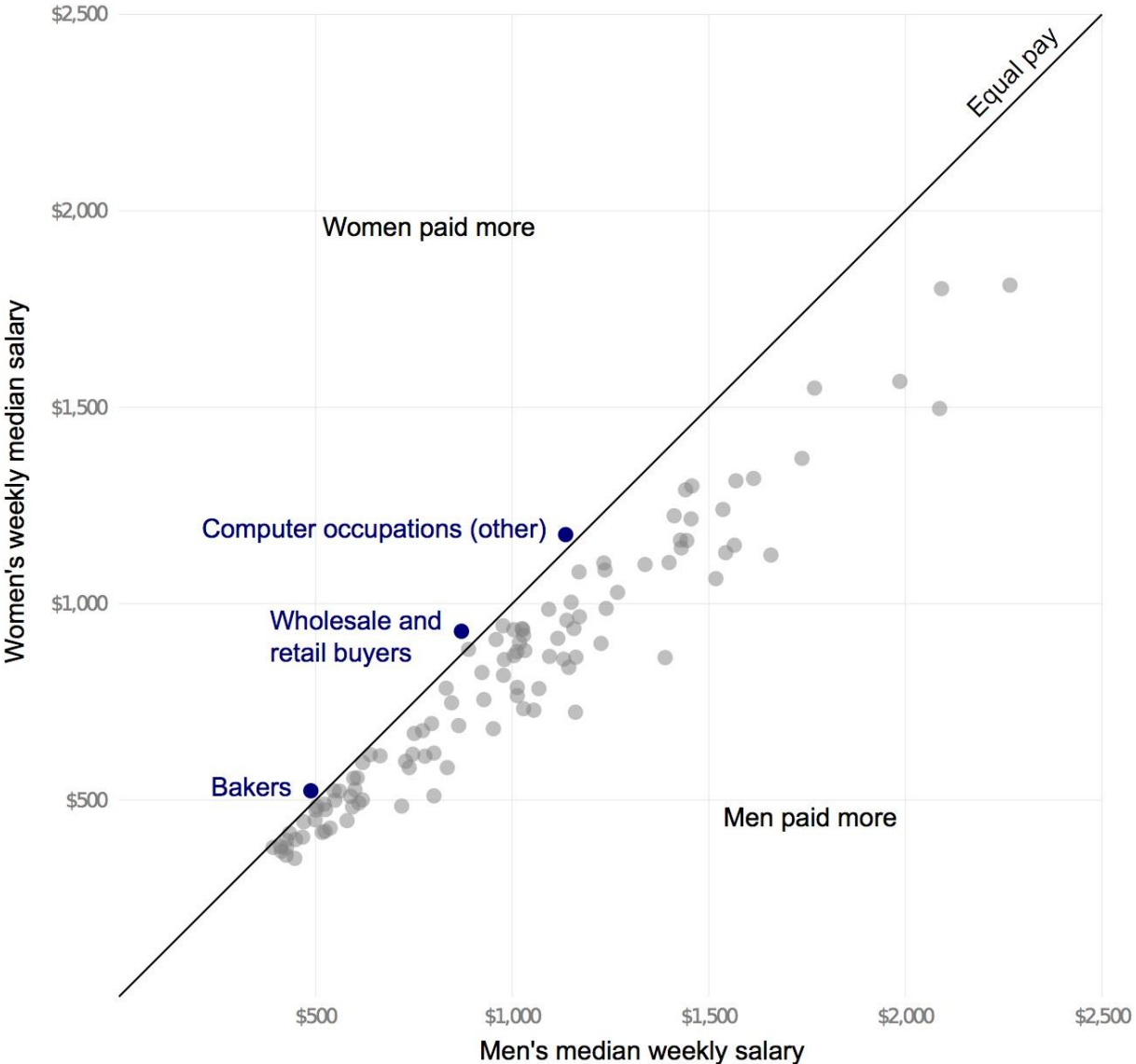
Highlight the story: minimize chart junk



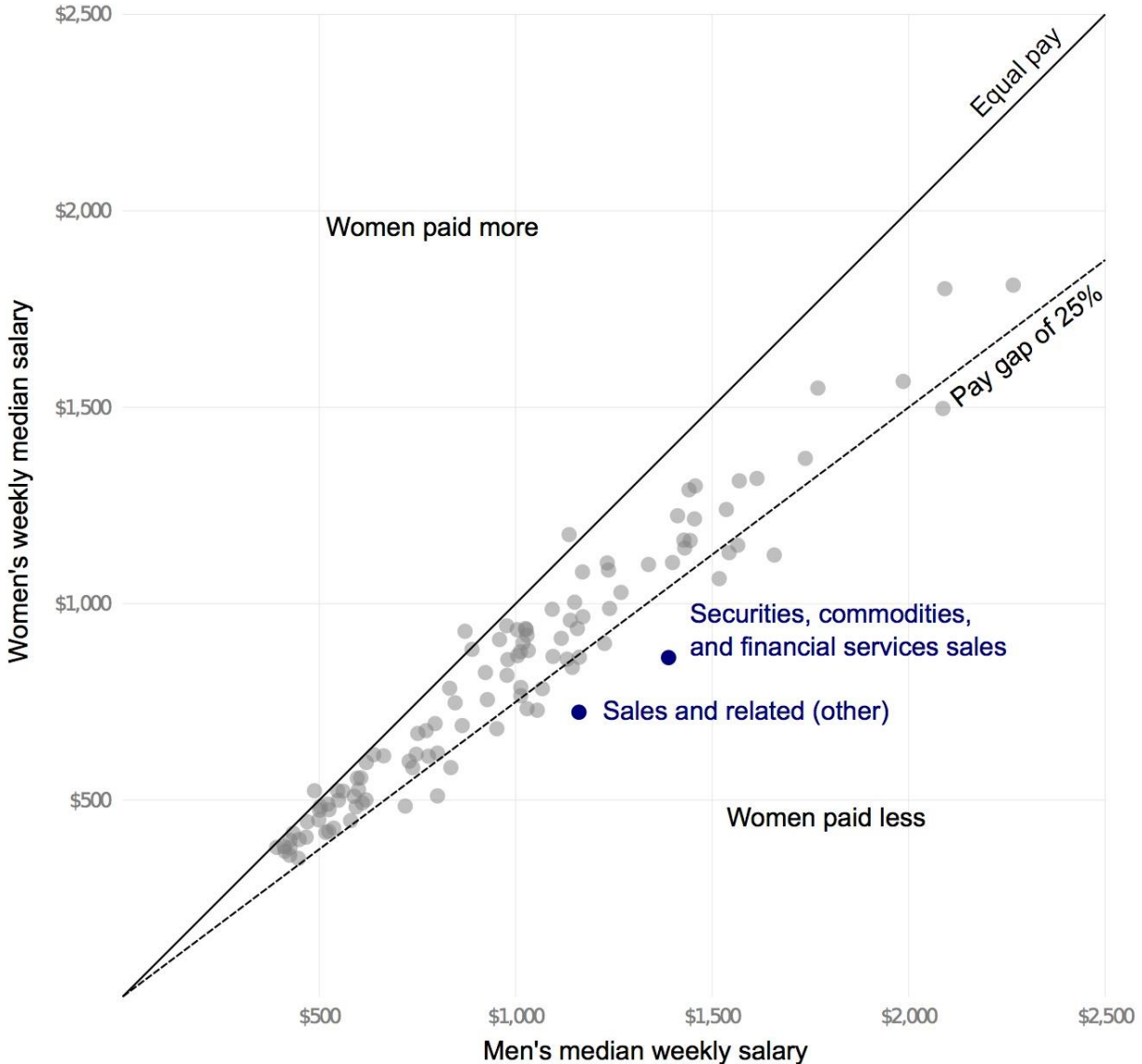
Highlight the story: labels and annotation



Highlight the story: labels and annotation



Highlight the story: labels and annotation



When in doubt:

keep it clean, clear and simple!

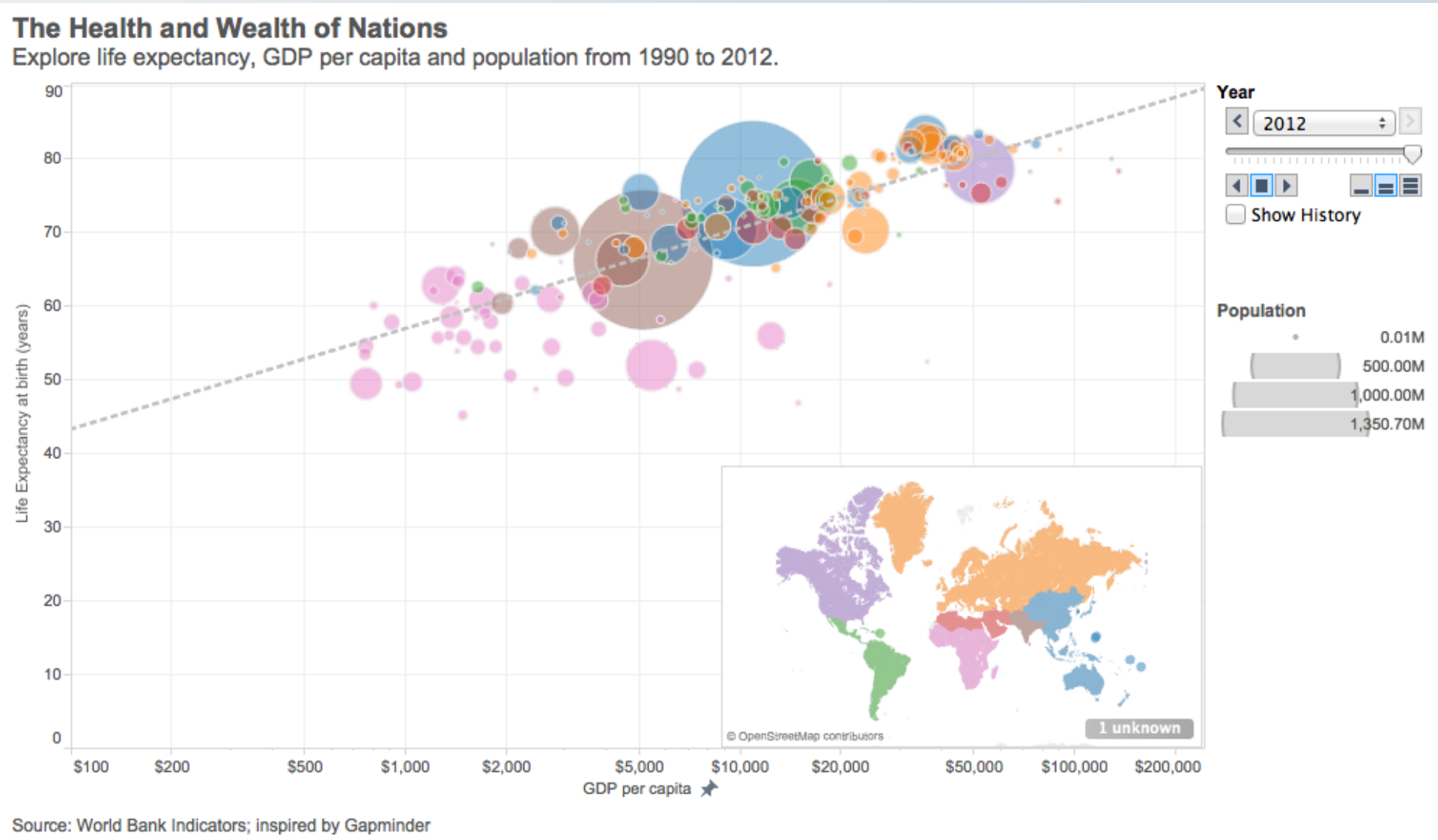
(But aim for clarity over simplicity)

Experiment! Sketch!

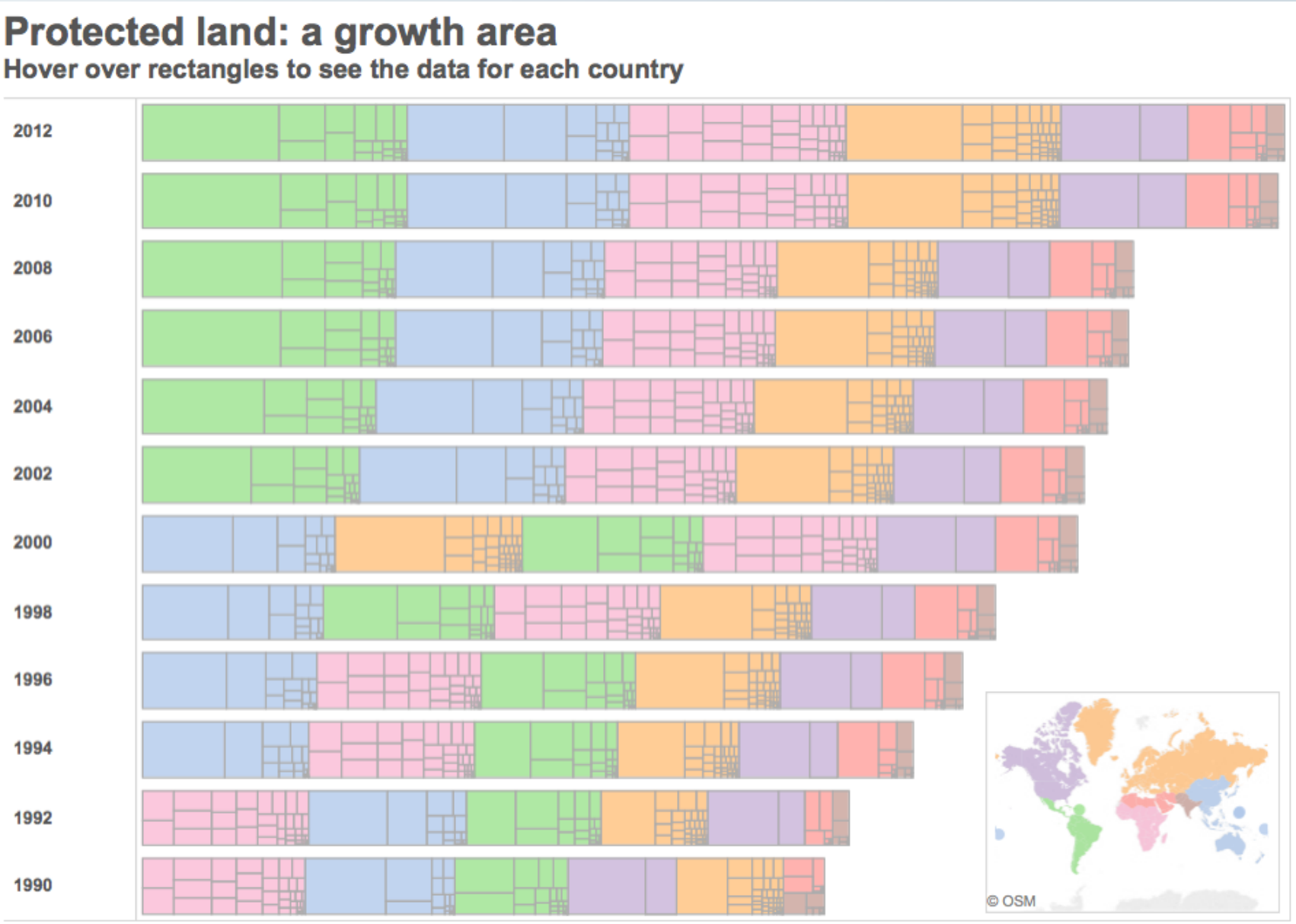
That may be how you find the story

Show people. If they're confused, try another approach

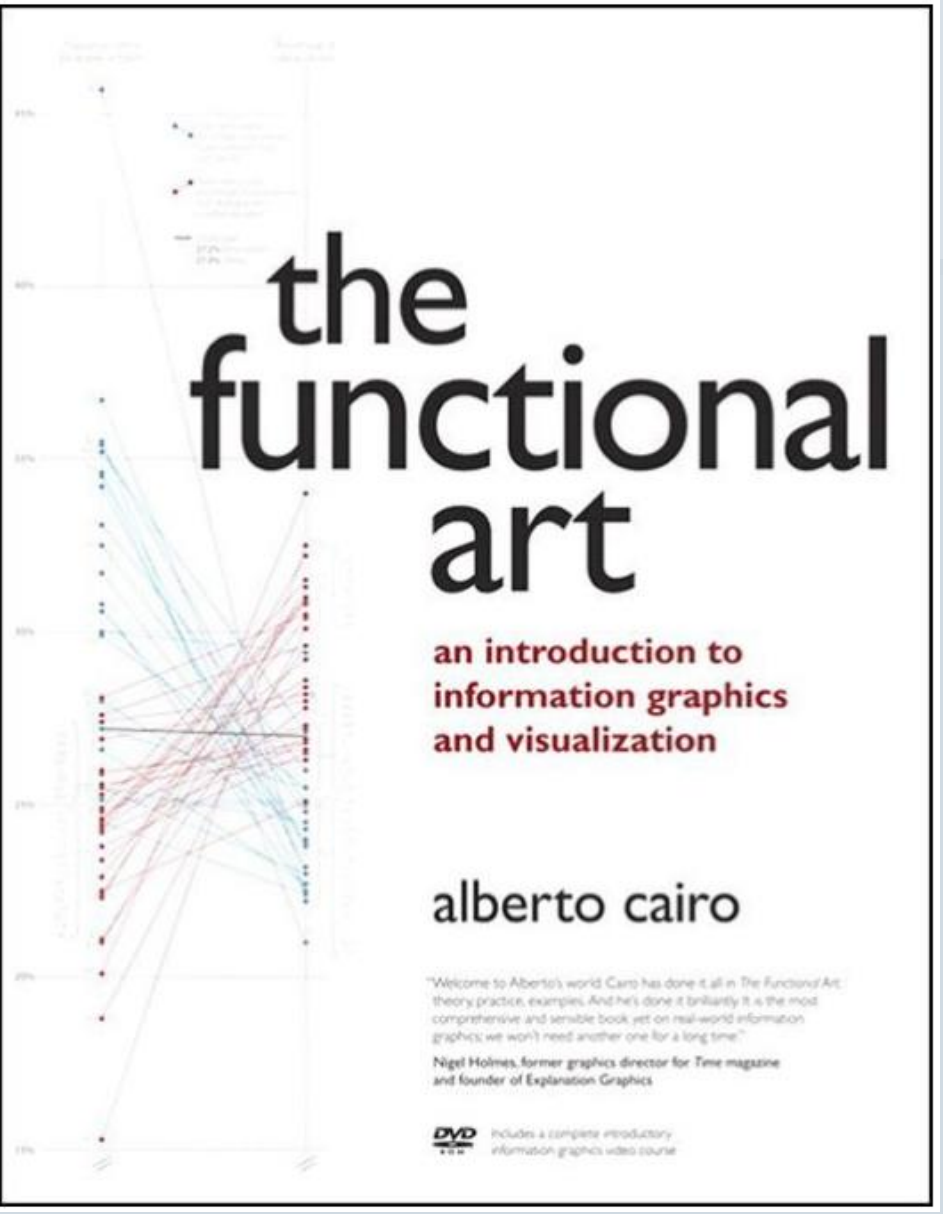
Putting it all together

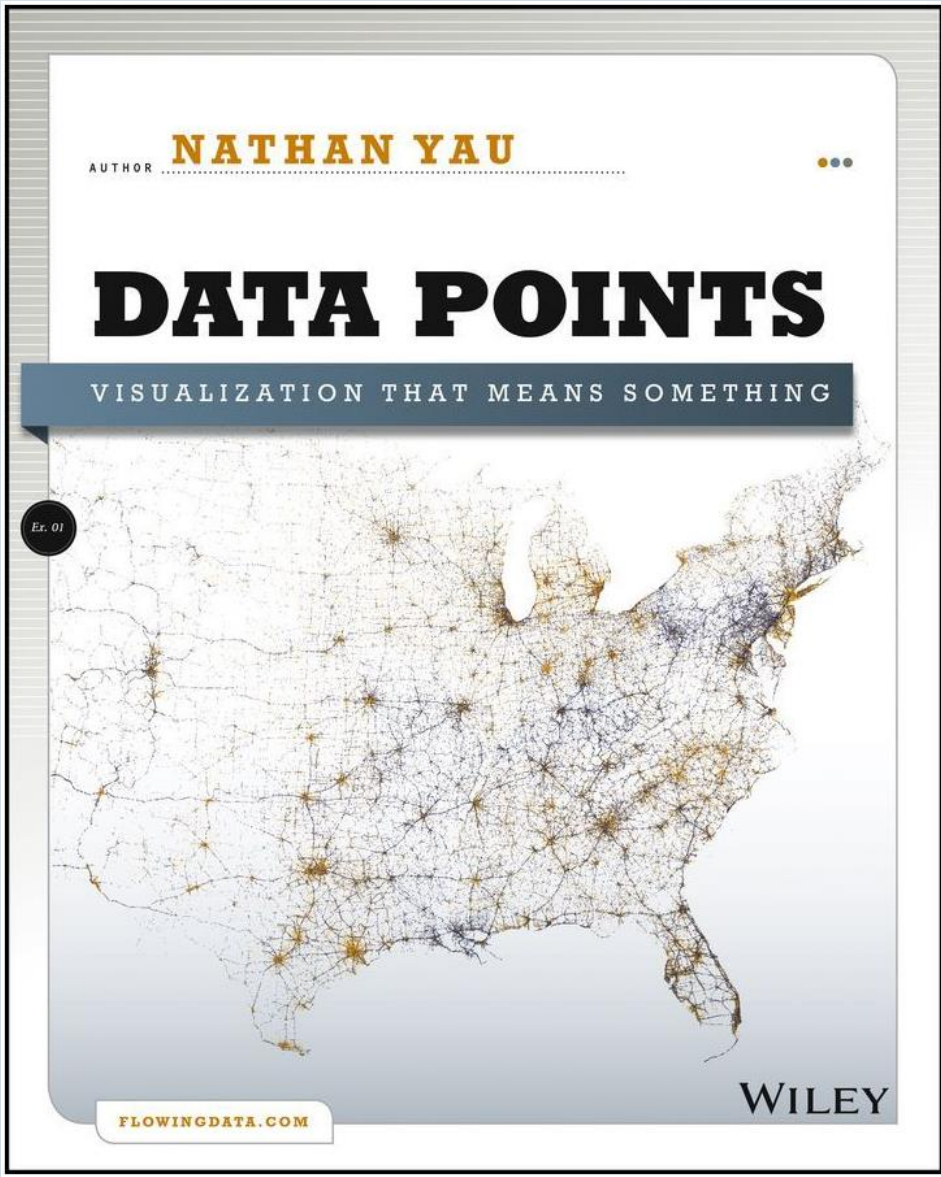


Putting it all together



Recommended reading





Data visualization: basic principles

Peter Aldhous

@paldhous