

# **Data visualization: basic principles**

**Peter Aldhous**

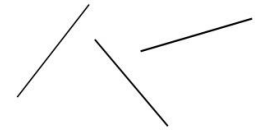
[@paldhous](#)

# Visualization: encoding data by visual cues

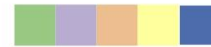
**Length**



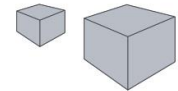
**Slope**



**Color hue**



**Volume**



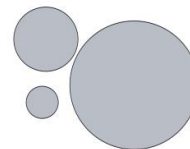
**Angle**



**Length (aligned)**



**Area**



**Color intensity**



**Our brains do not  
treat those cues  
equally!**

Accurate

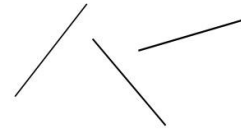
Length (aligned)



Length



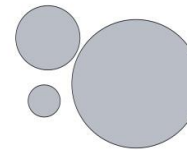
Slope



Angle



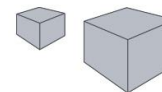
Area



Color intensity



Volume



Generic

Color hue



# Design for the human brain!



**What type of chart should I use?**

Distribution

Relationship

Comparison

**What do you want to show?**

Connection

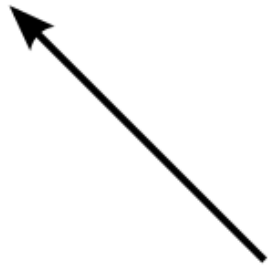
Composition  
(parts of the whole)

Location

**Distribution**

Relationship

Comparison



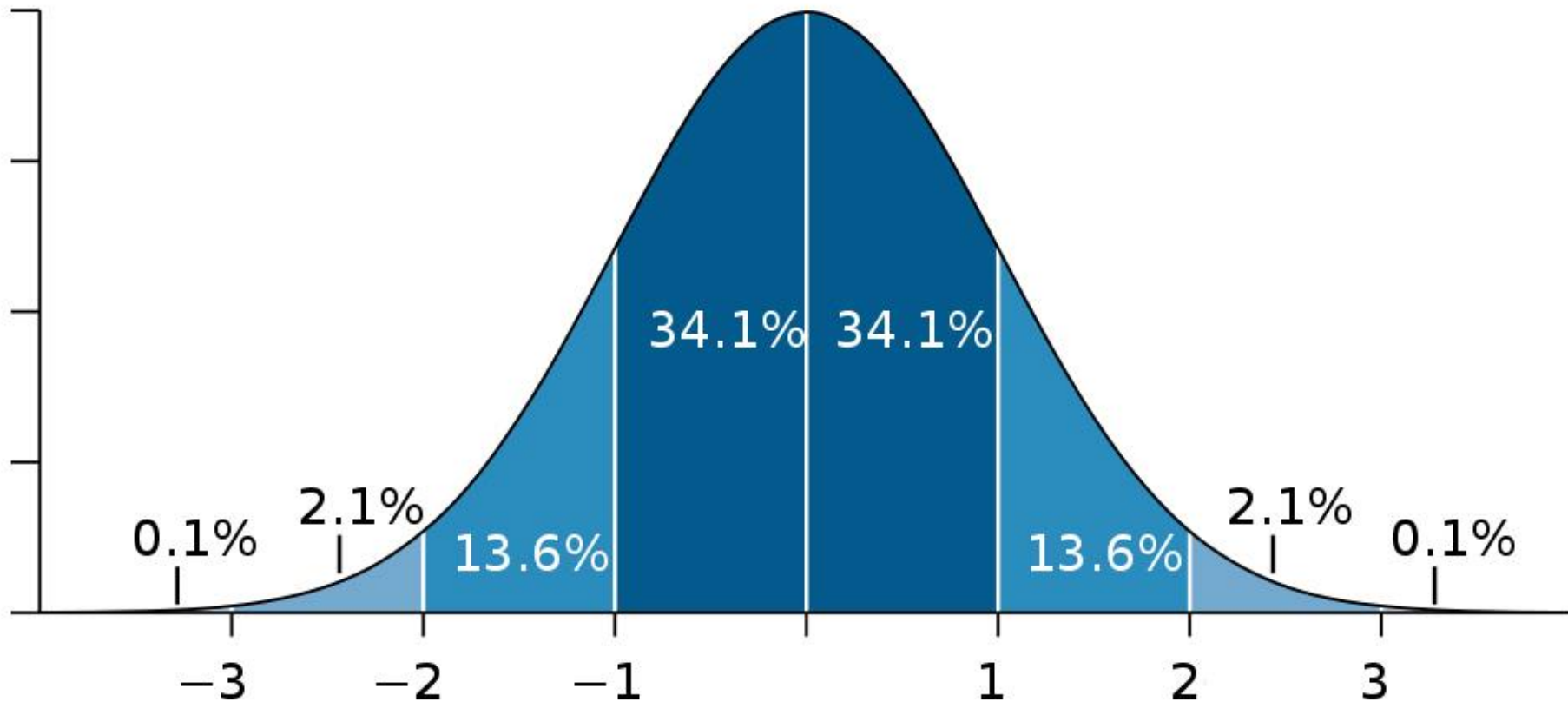
**What do you want to show?**

Connection

Composition  
(parts of the whole)

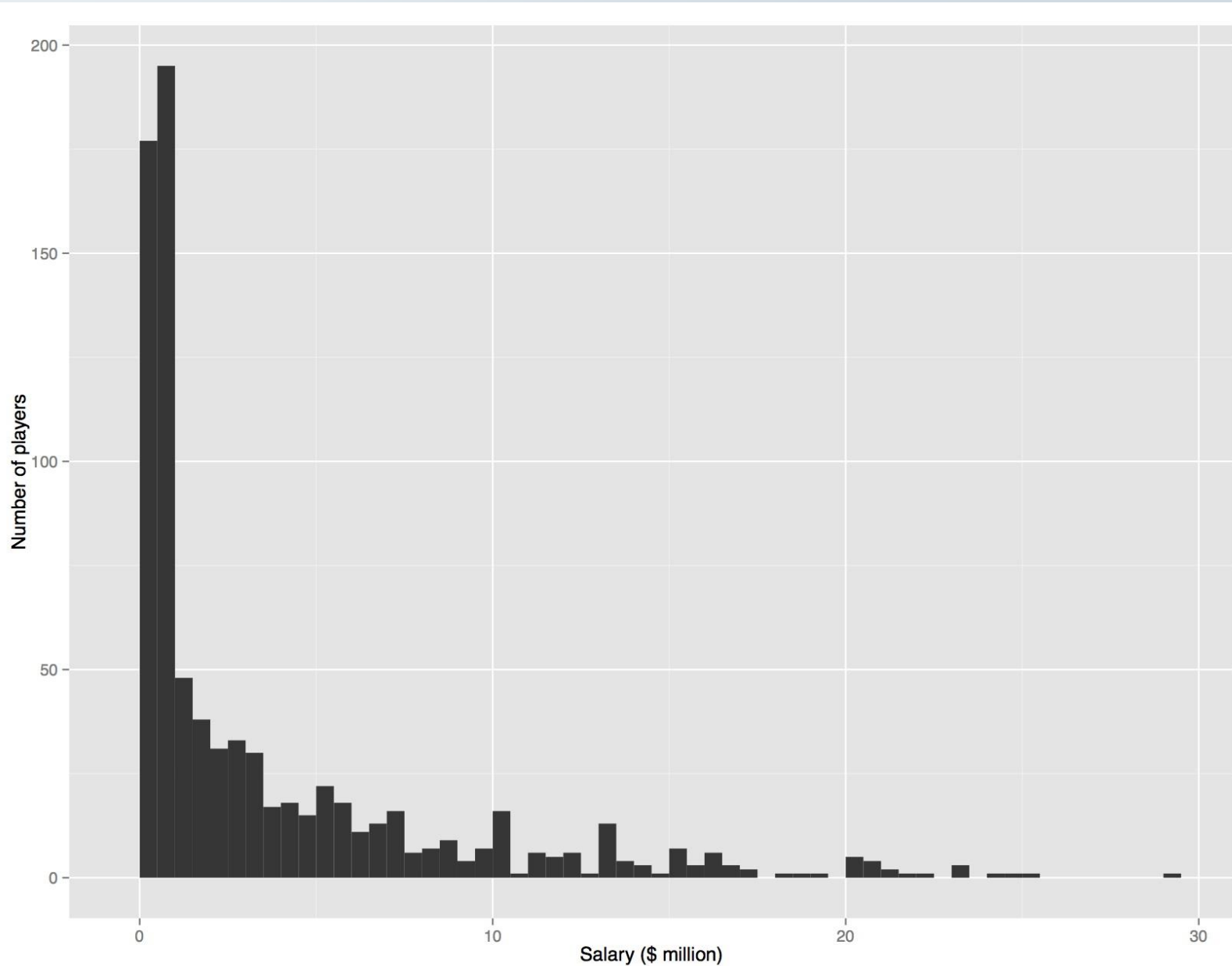
Location

# Consider the distribution





# Consider the distribution



Distribution

**Relationship**

Comparison



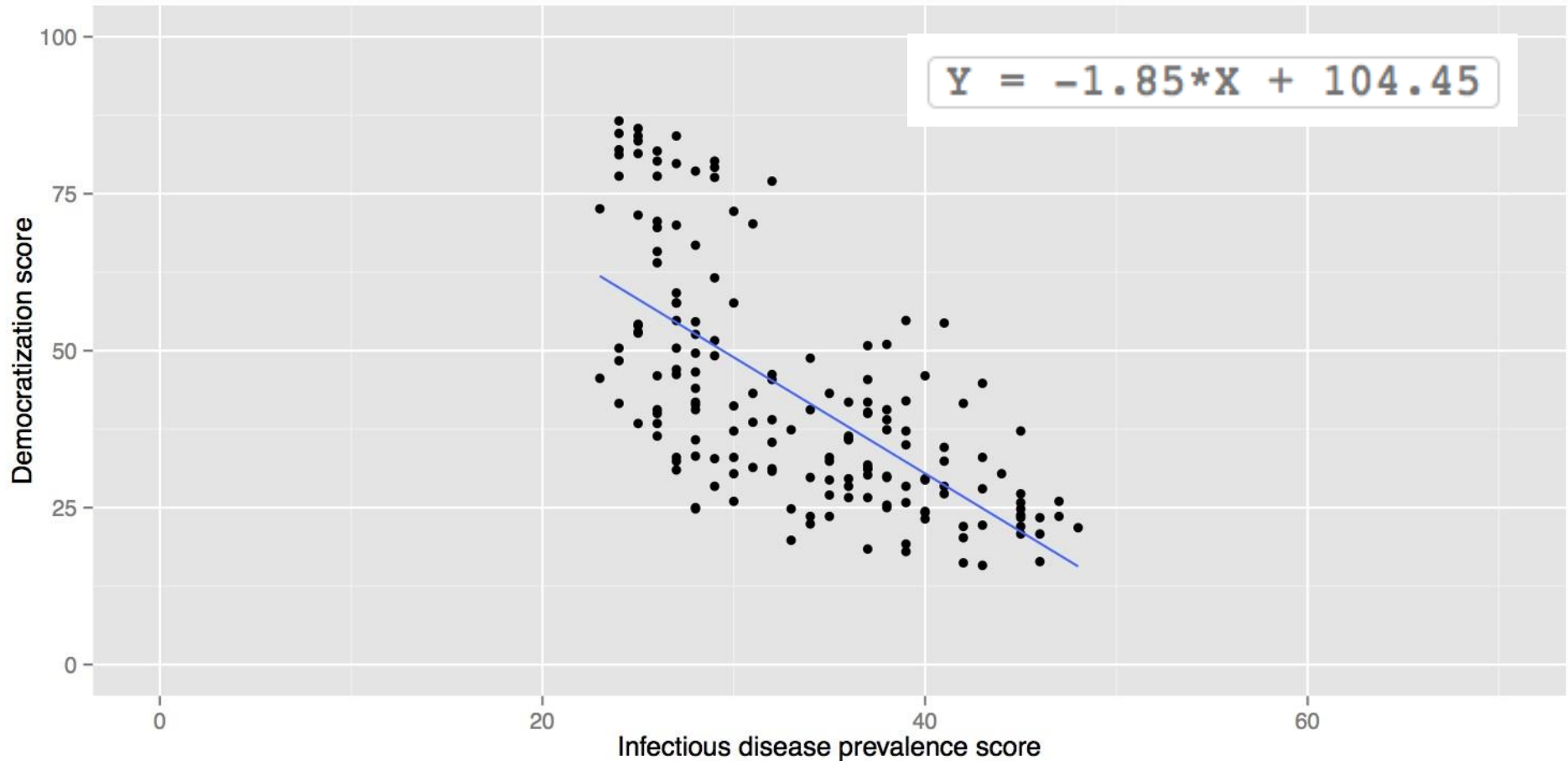
**What do you want to show?**

Connection

Composition  
(parts of the whole)

Location

# Relationships between variables: scatter plots and trend lines



# **Relationships between variables: correlation and its pitfalls**

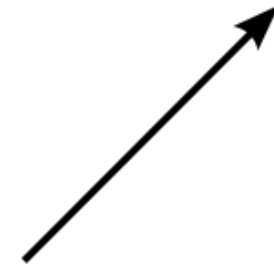
I can't stress enough:

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

Distribution

Relationship

**Comparison**



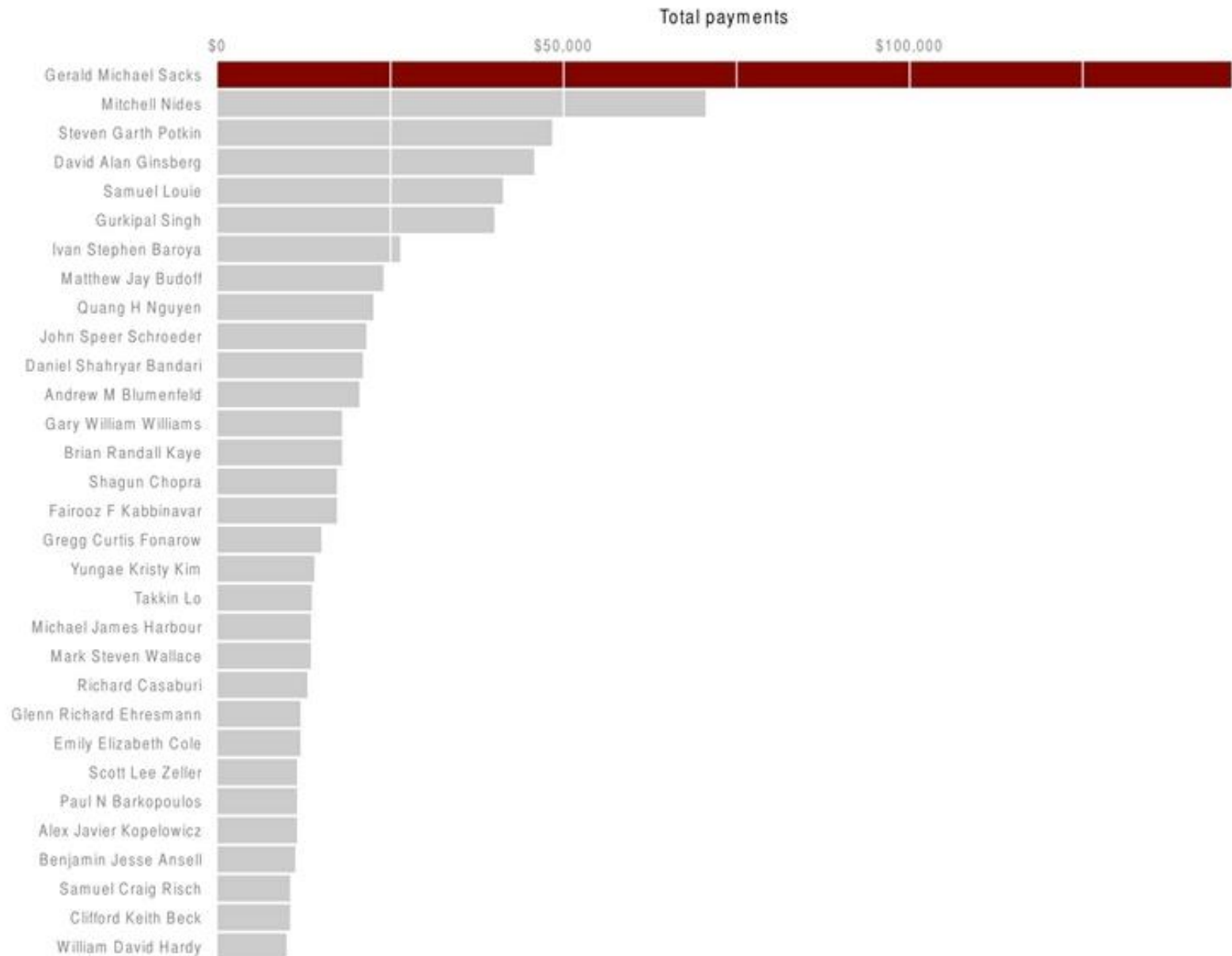
**What do you want to show?**

Connection

Composition  
(parts of the whole)

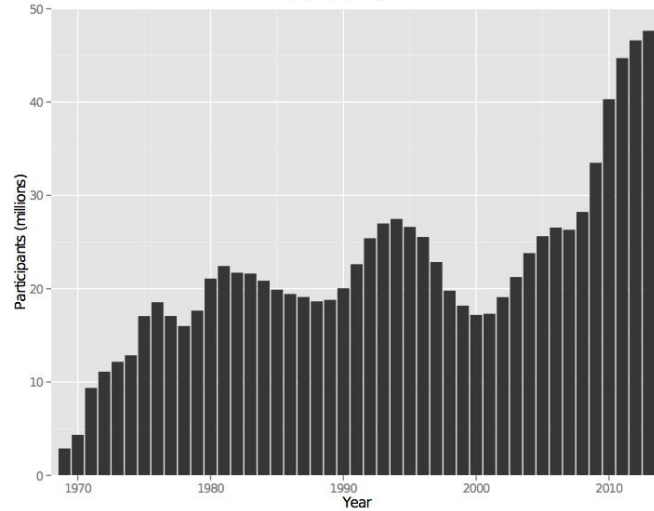
Location

# Simple comparisons: bars and columns

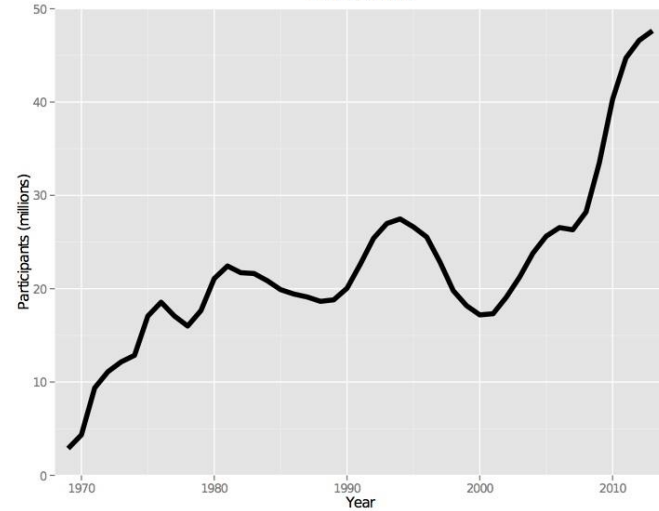


# Comparisons: change over time

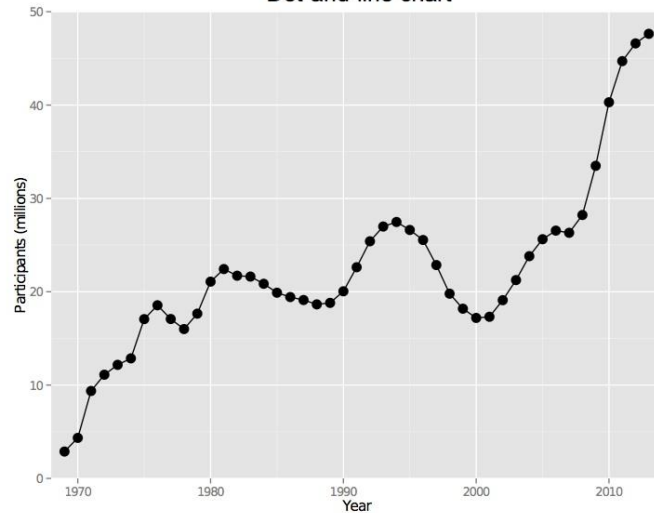
Column chart



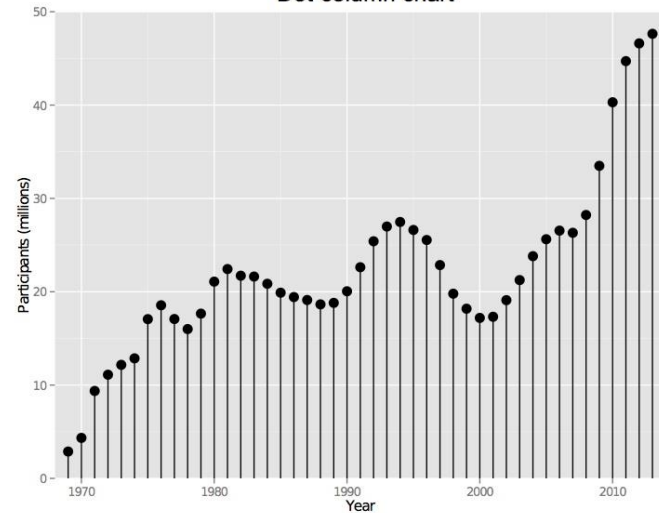
Line chart



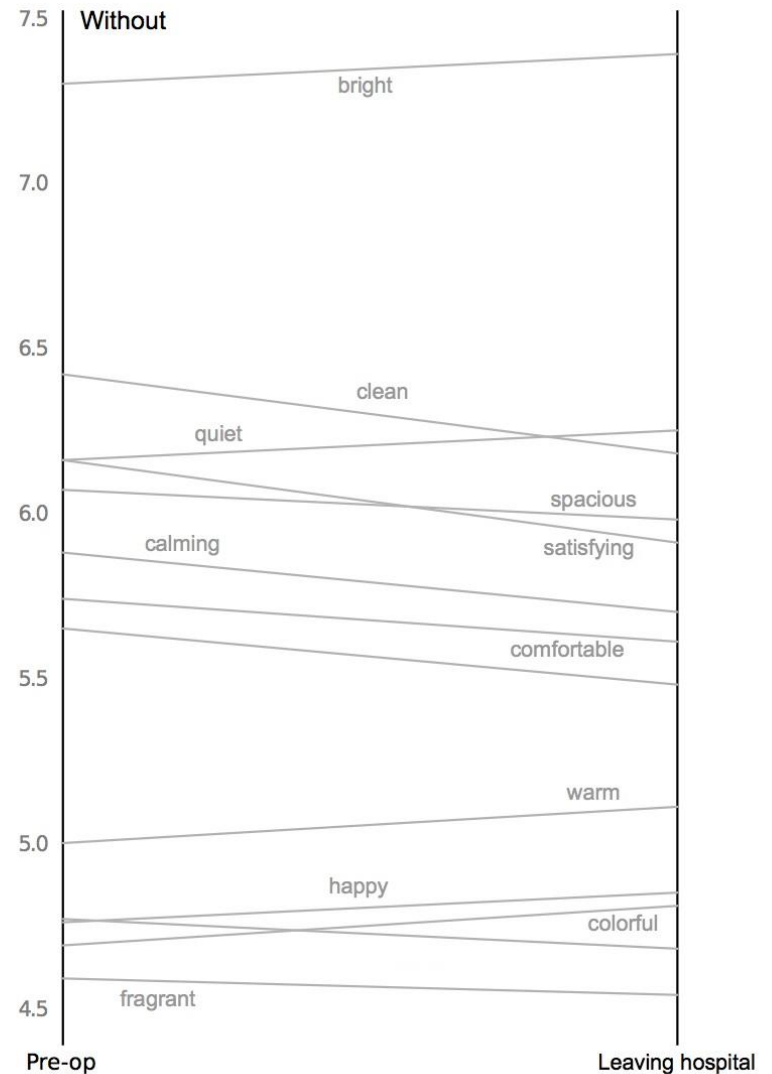
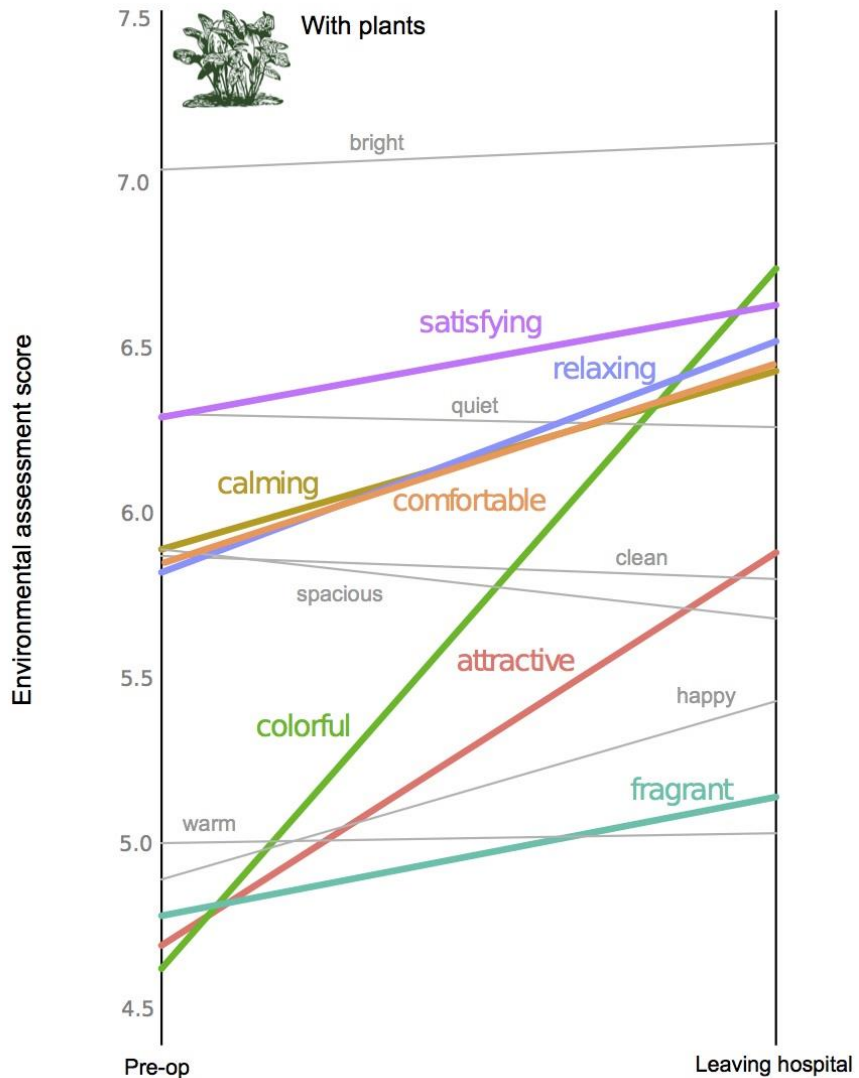
Dot-and-line chart



Dot-column chart



# Two points in time, many categories





Distribution

Relationship

Comparison

**What do you want to show?**

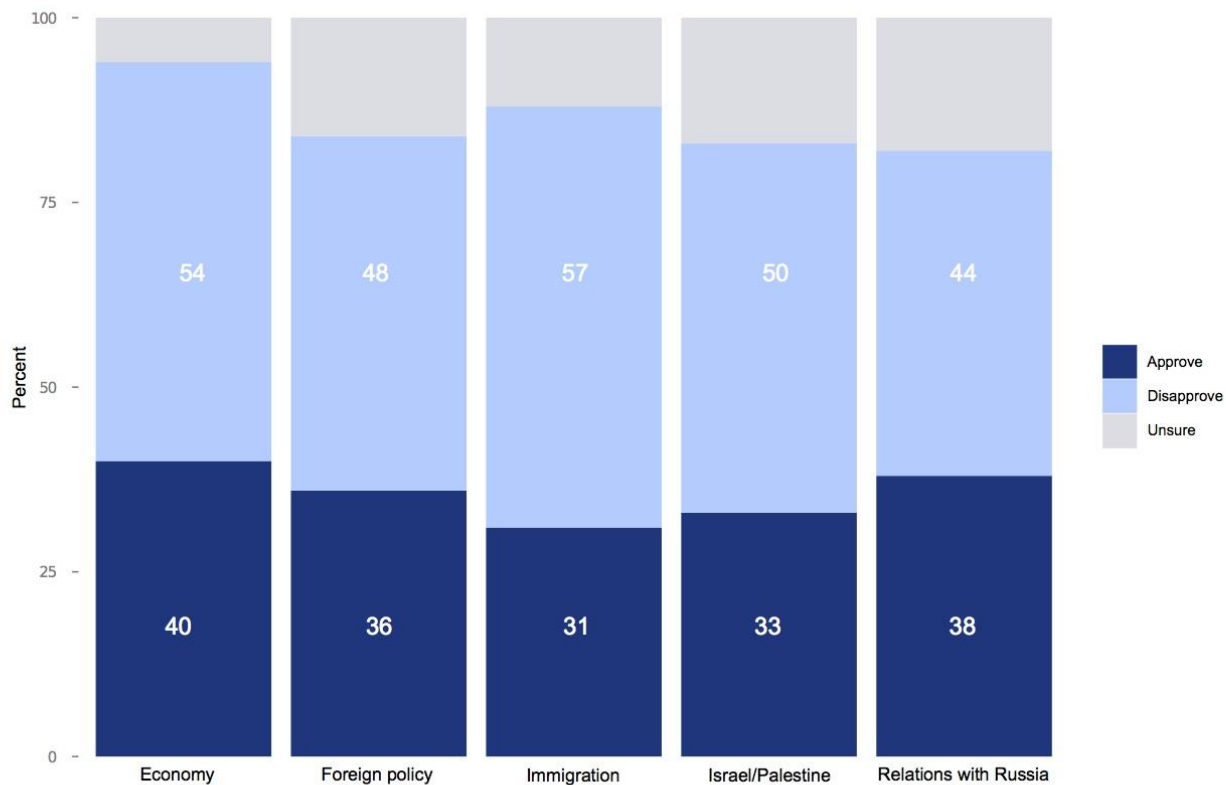
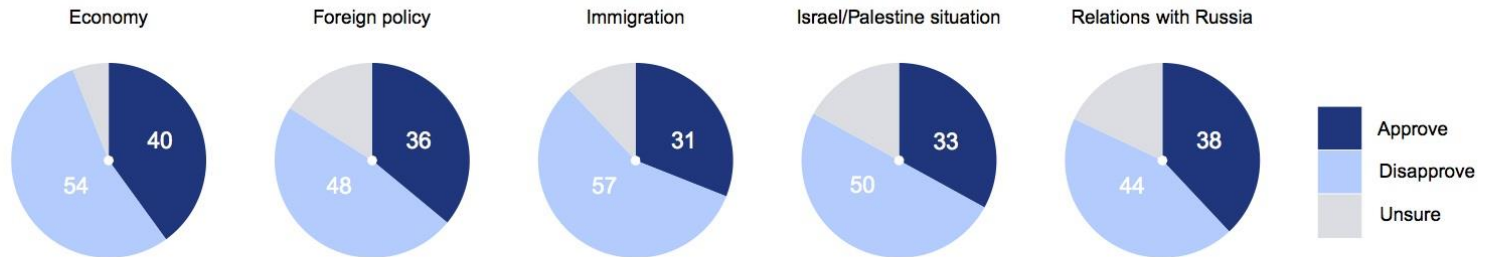


Connection

**Composition**  
**(parts of the whole)**

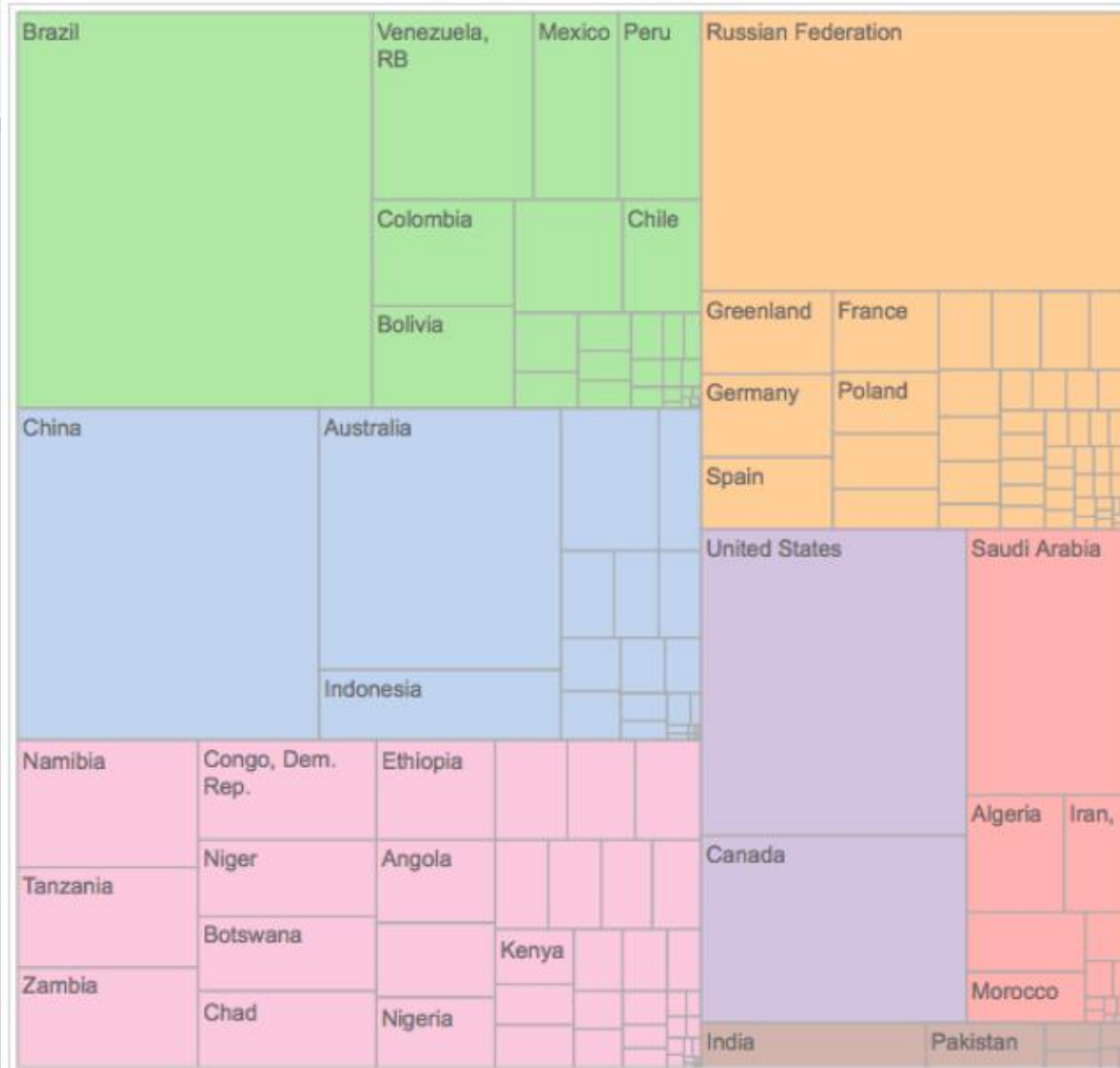
Location

# Composition: parts of the whole



# Composition: parts of the whole

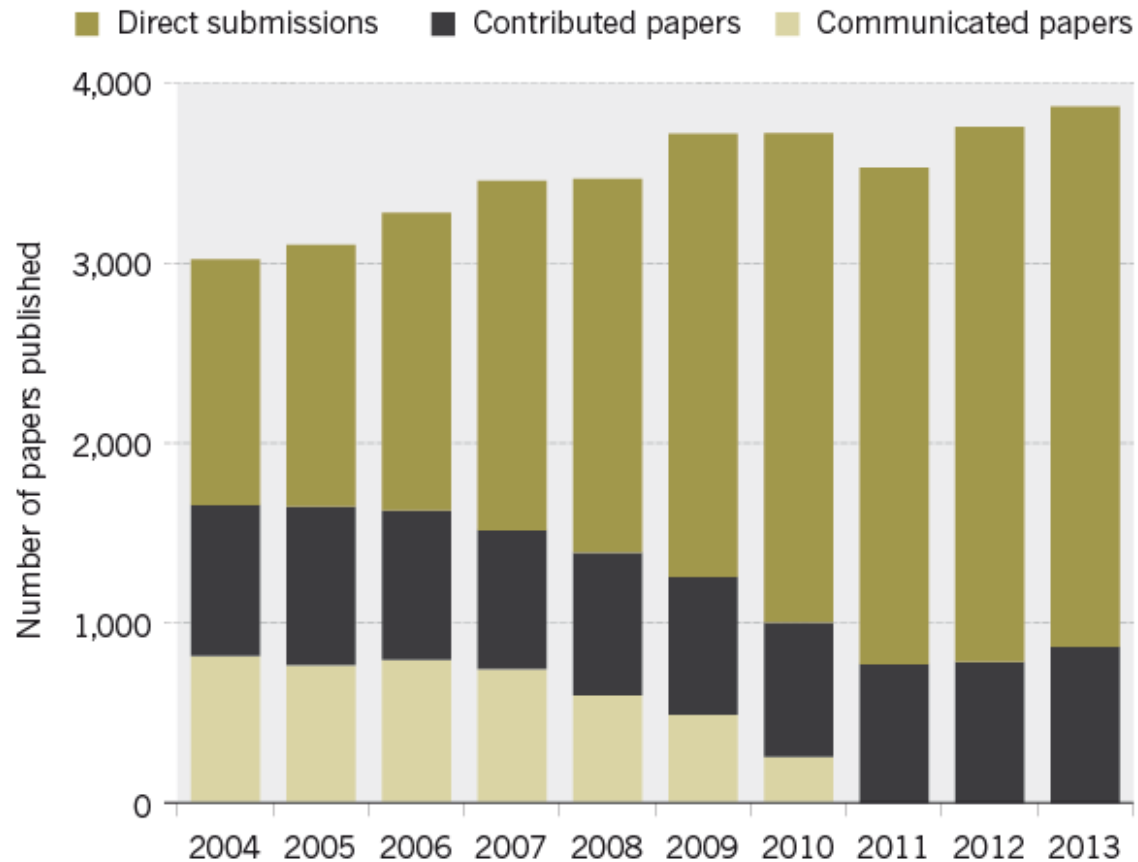
## Protected land in 2012



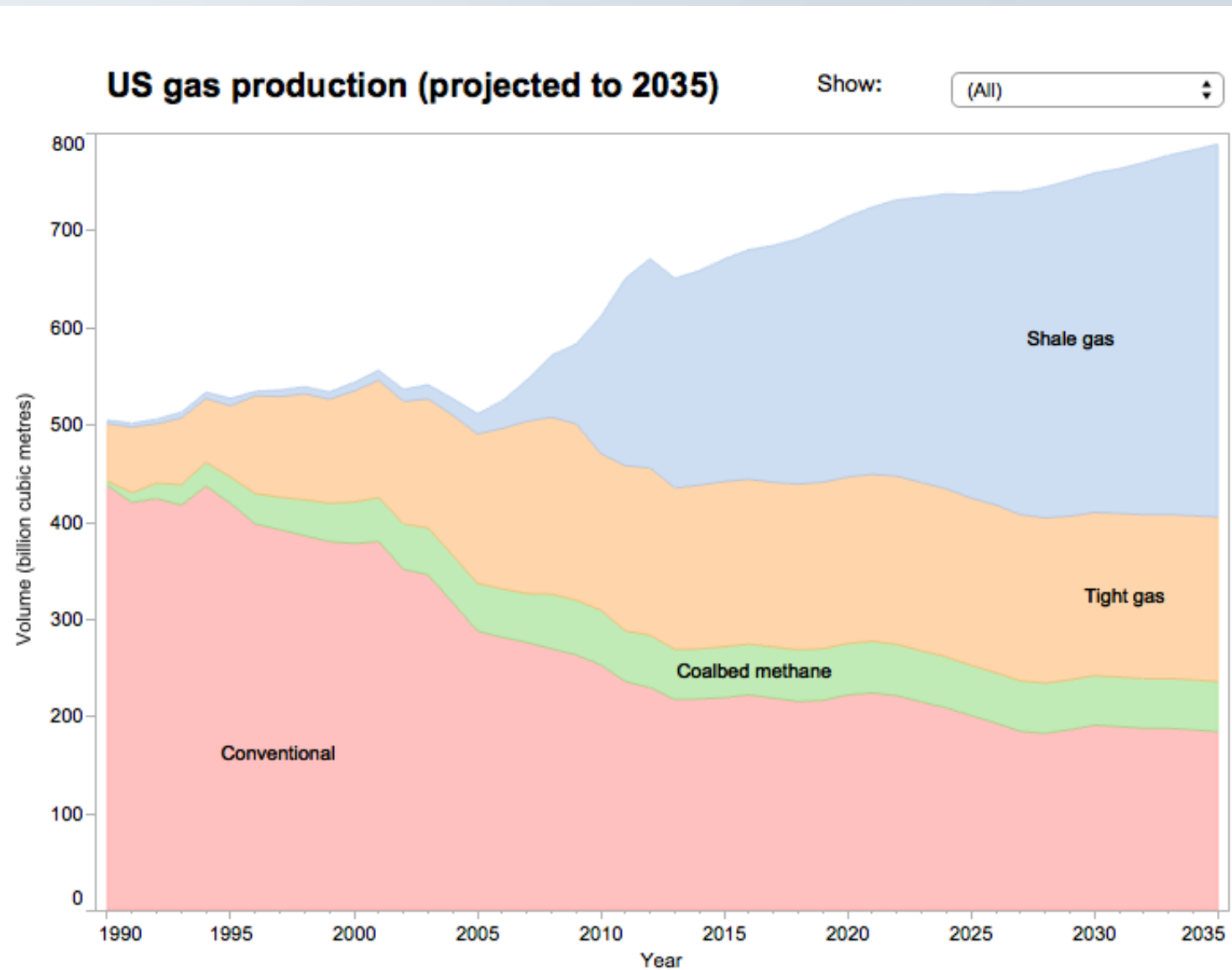
# Composition: change over time

## *A changing journal*

The number of direct submissions to *Proceedings of the National Academy of Sciences* has been increasing steadily over the past decade. Communicated papers were phased out in 2010, but the contributed track has remained constant.



# 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)

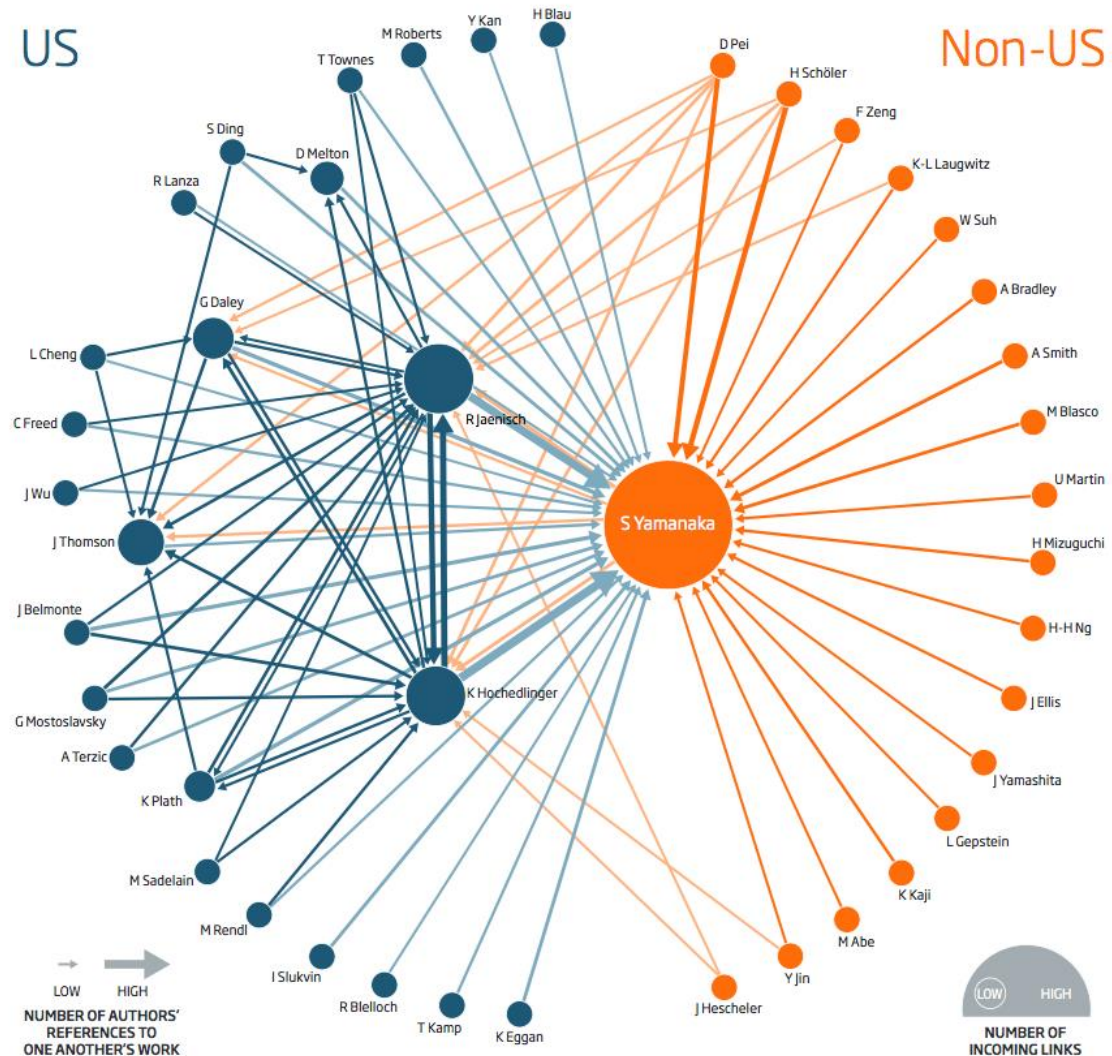
Location

# Connection: network graphs

## 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?**



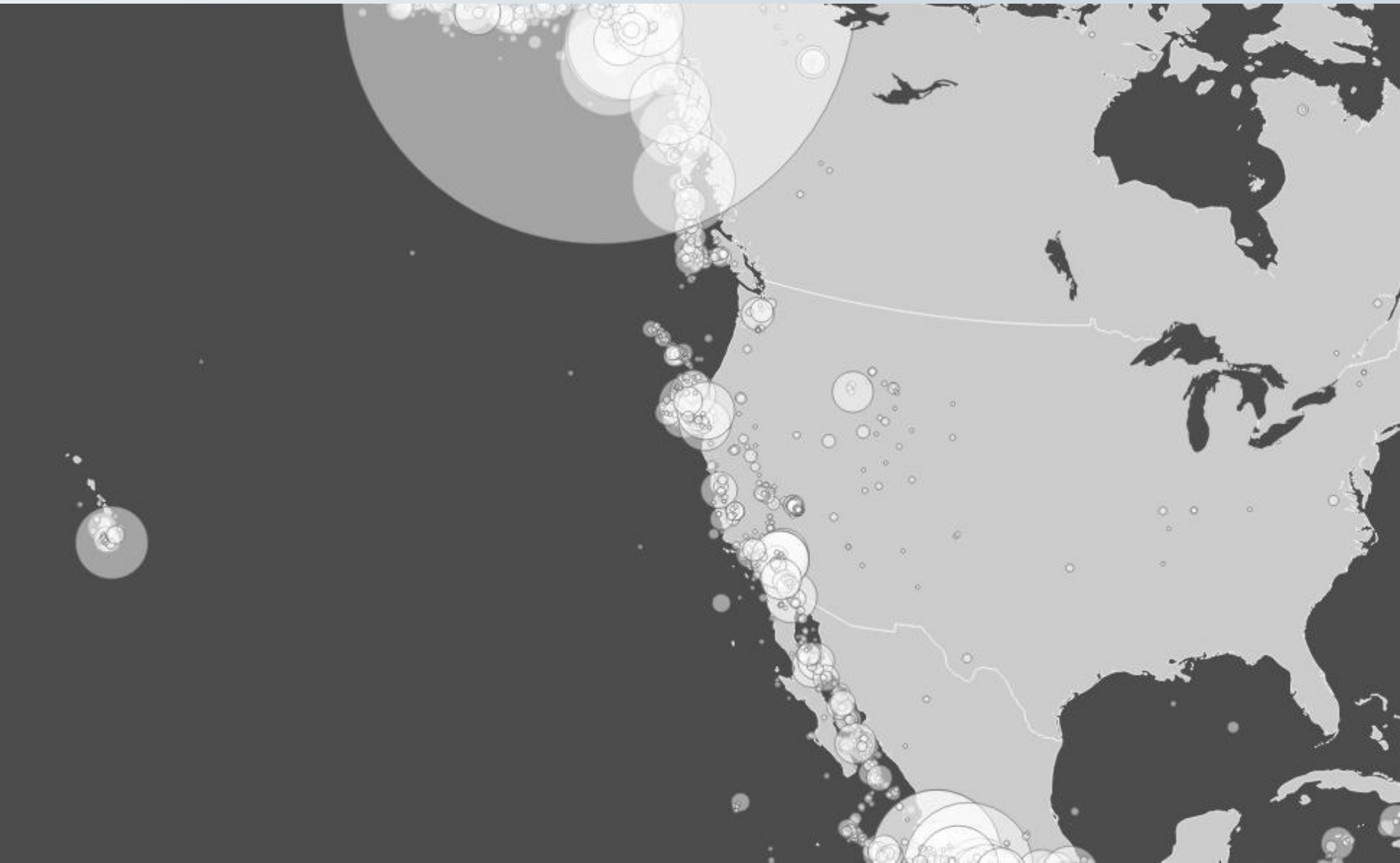
Connection

Composition  
(parts of the whole)

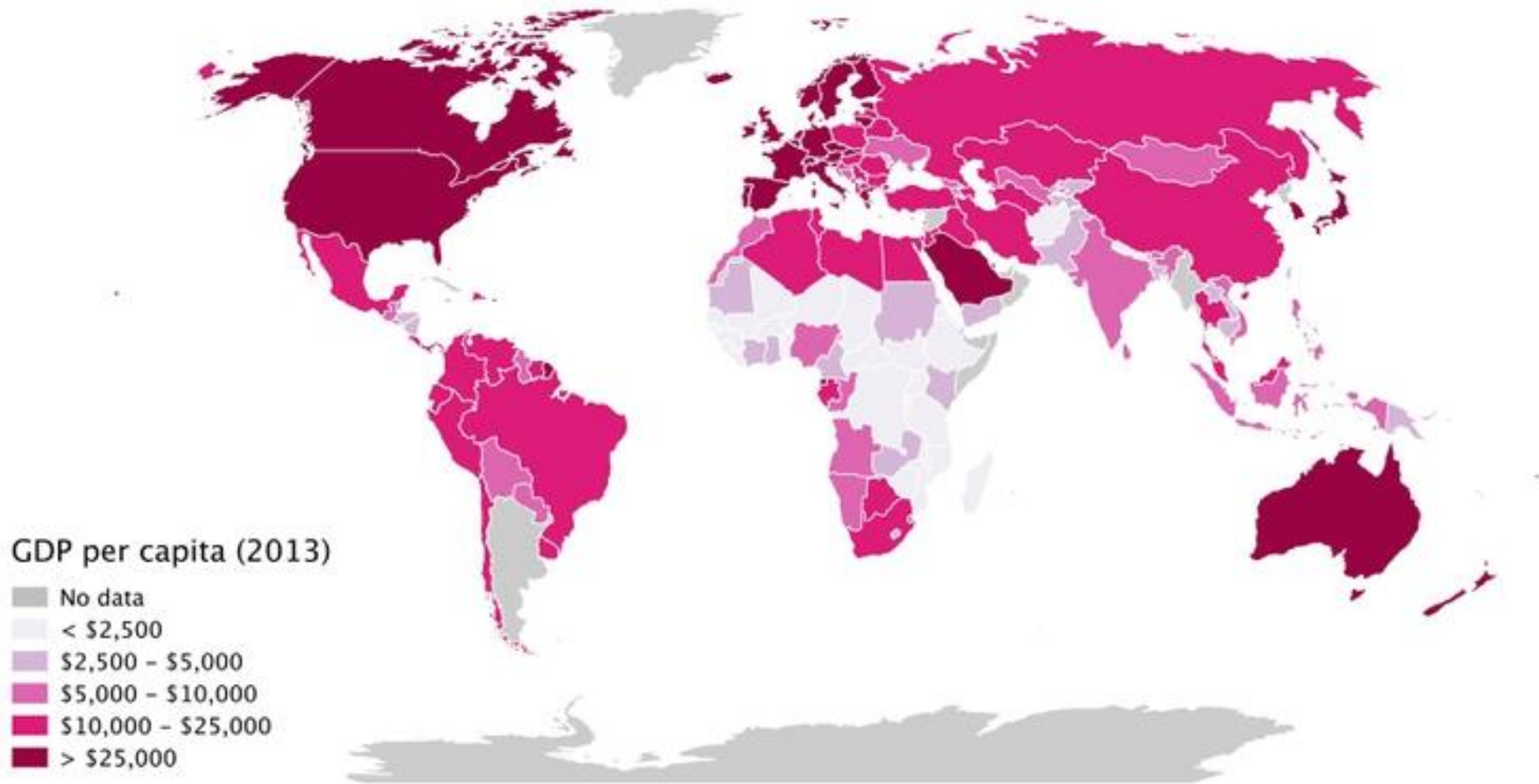
**Location**



# Location plus data: scaled circles



# Location plus data: choropleth maps



**Remember the  
perceptual  
hierarchy of  
visual cues!**

Accurate

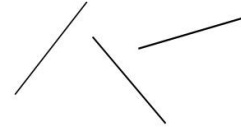
Length (aligned)



Length



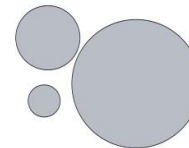
Slope



Angle



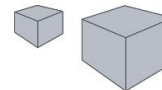
Area



Color intensity



Volume



Color hue



Generic

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

# **Using color effectively**

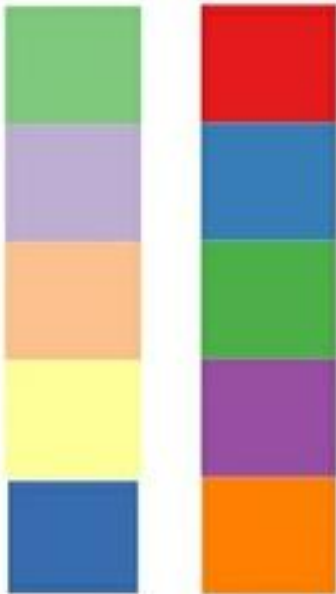
# The color wheel



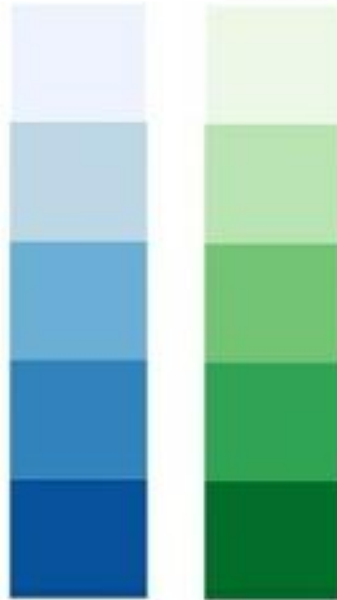
Wikimedia Commons

# Using color: fit to your data

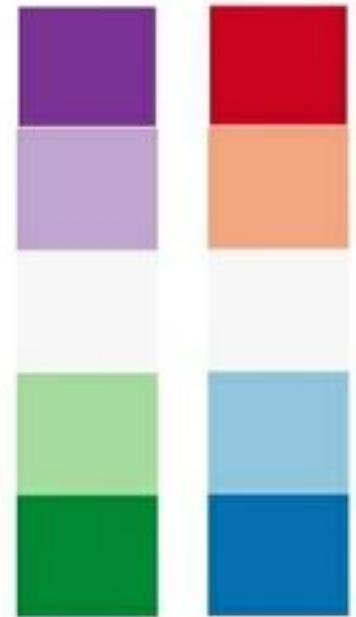
Qualitative



Sequential



Diverging



# ColorBrewer is your friend!

number of data classes on your map  
3 | ▼ [learn more >](#)

the nature of your data  
sequential ▼ [learn more >](#)

pick a color scheme: BuGn

multihue single hue

(optional) only show schemes that are:  
☐ colorblind safe ☐ print friendly  
☐ photocopy-able [learn more >](#)

pick a color system  
229, 245, 249 153, 216, 201 44, 162, 95  
● RGB ○ CMYK ○ HEX

adjust map context  
☐ roads ☐ cities ☒ borders

select a background  
● solid color ○ terrain  
color transparency

EXPORT YOUR COLORS >>

how to use | updates | credits

COLORBREWER 2.0  
color advice for cartography

SCORE CARD

© Cynthia Brewer, Mark Harrower and The Pennsylvania State University

[Support](#)

[Back to ColorBrewer 1.0](#)

axm

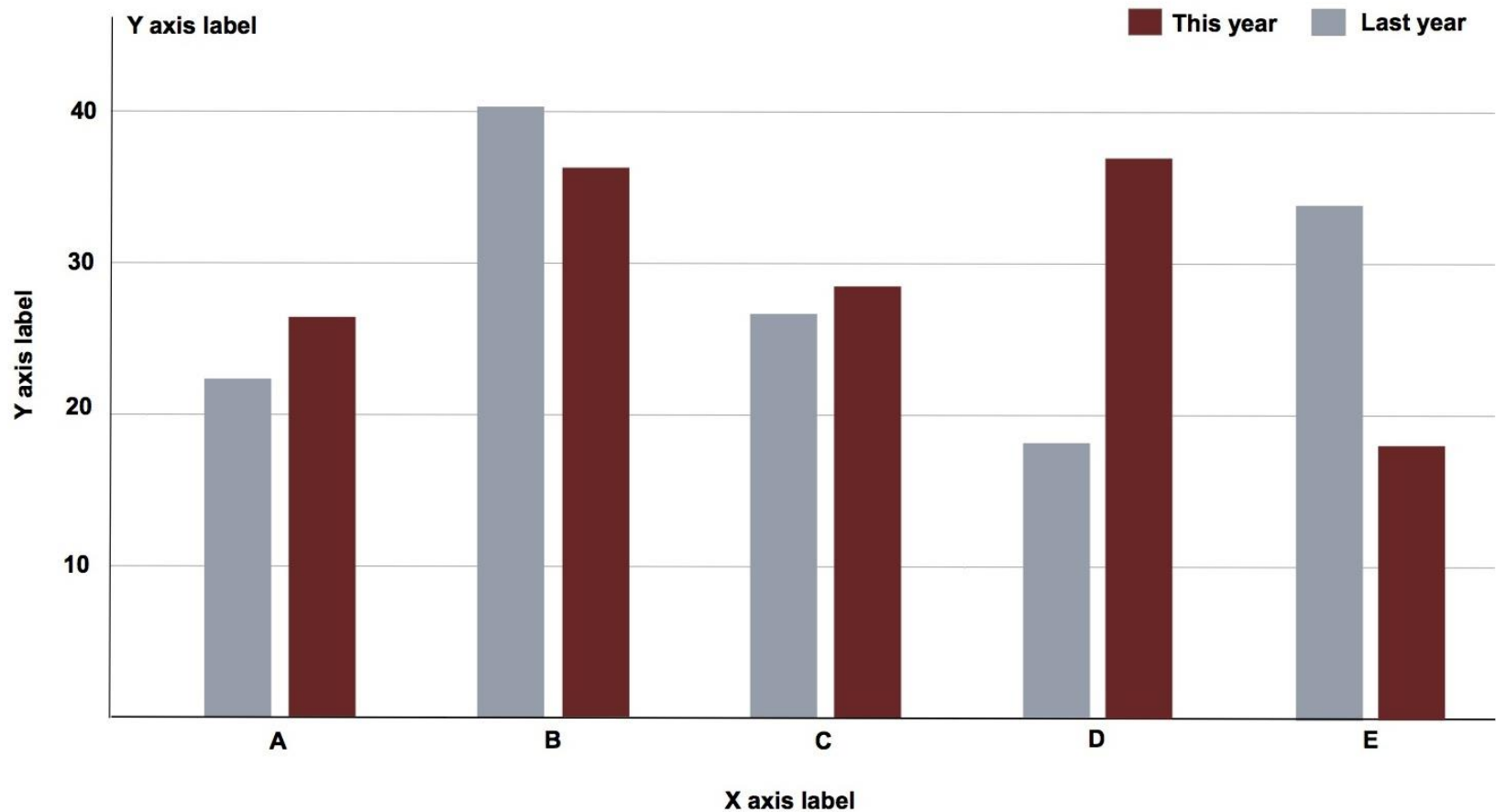
<http://www.colorbrewer2.org/>



# Chart furniture

## A title for the chart

And a subtitle, telling us some more about what it shows.

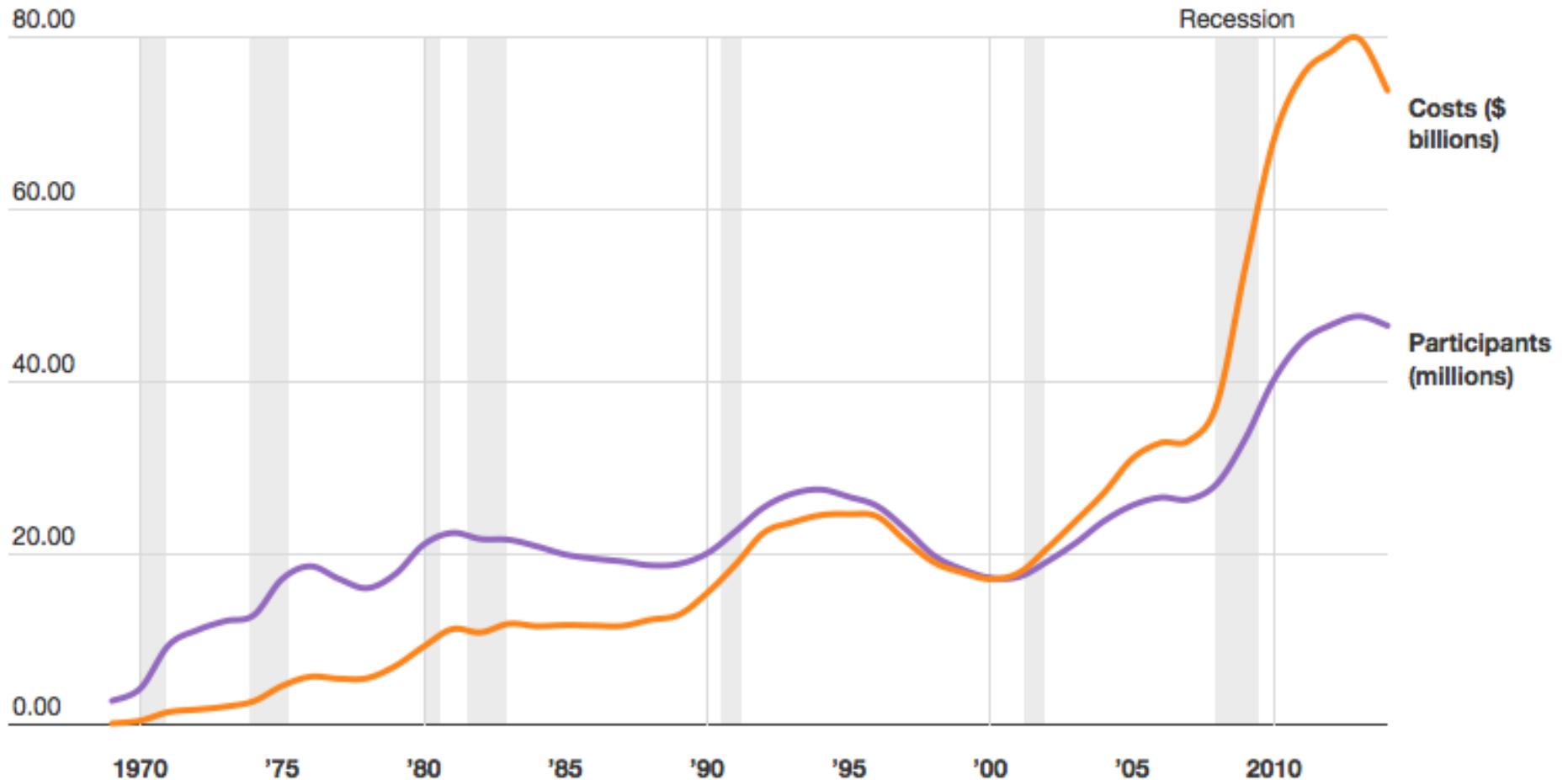


Source information

# Using chart furniture to encode data

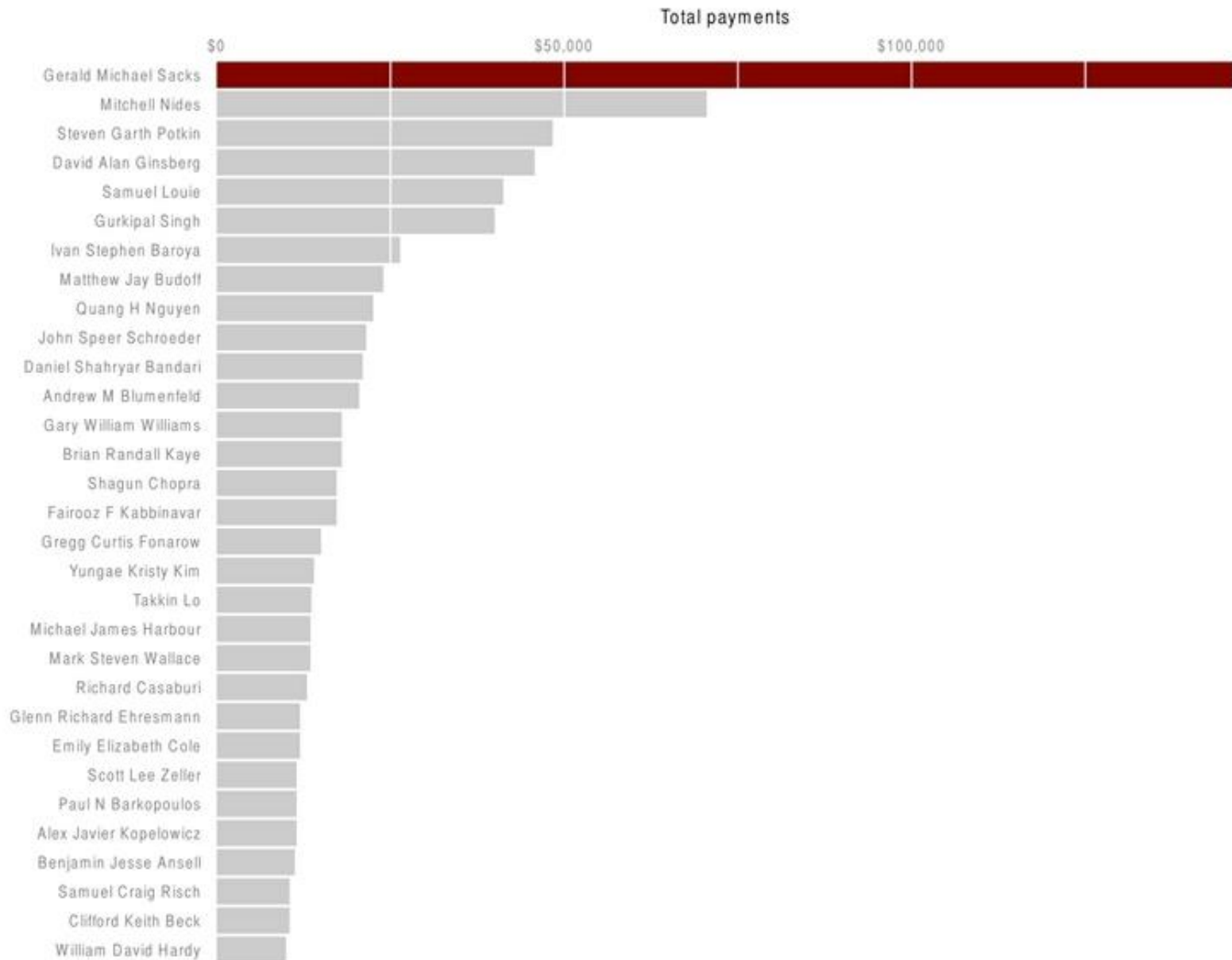
## Food stamp nation

The Supplemental Nutrition Assistance Program grew rapidly in recent years.

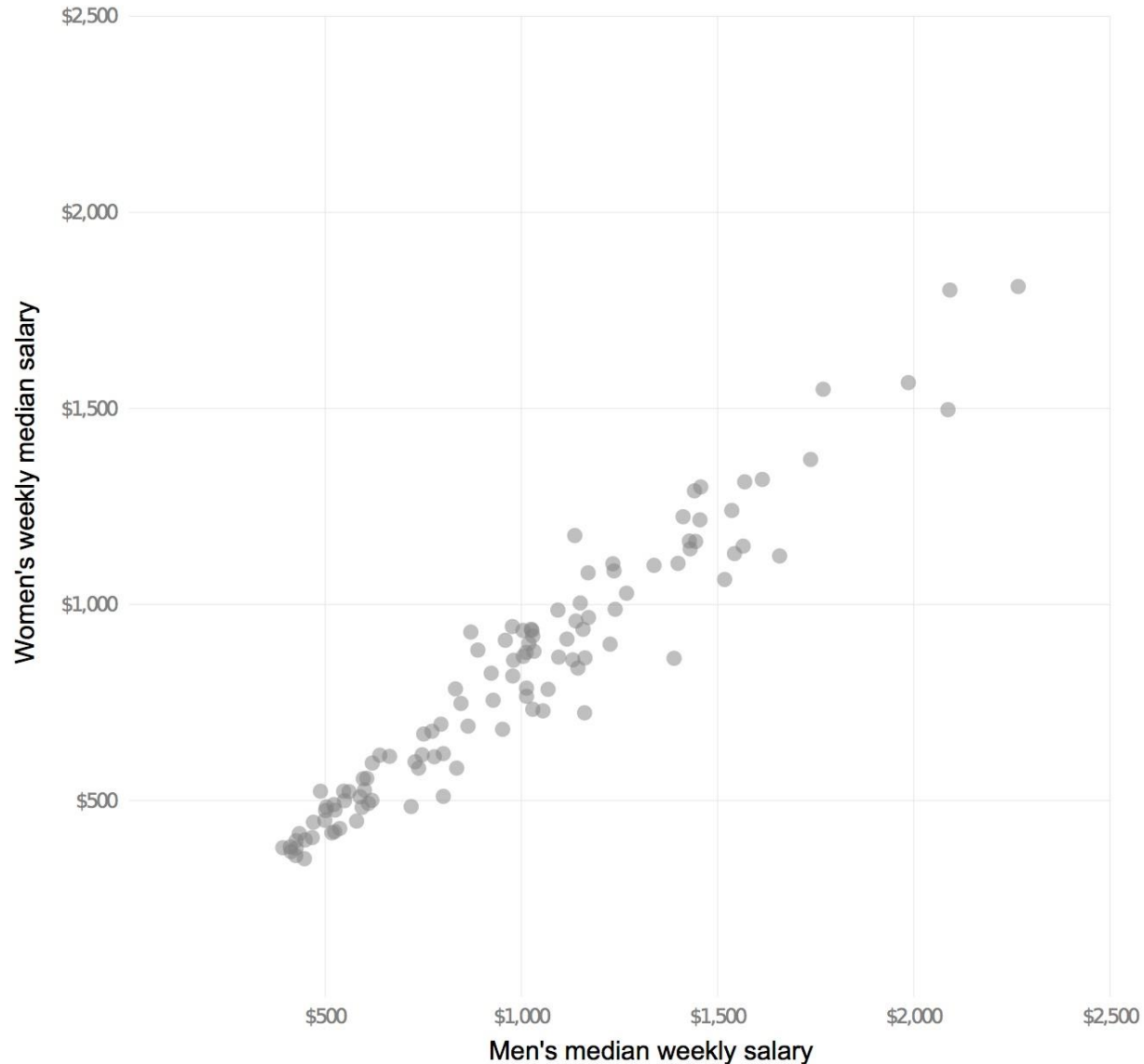


Source: [U.S. Department of Agriculture](#) [Get the data](#)

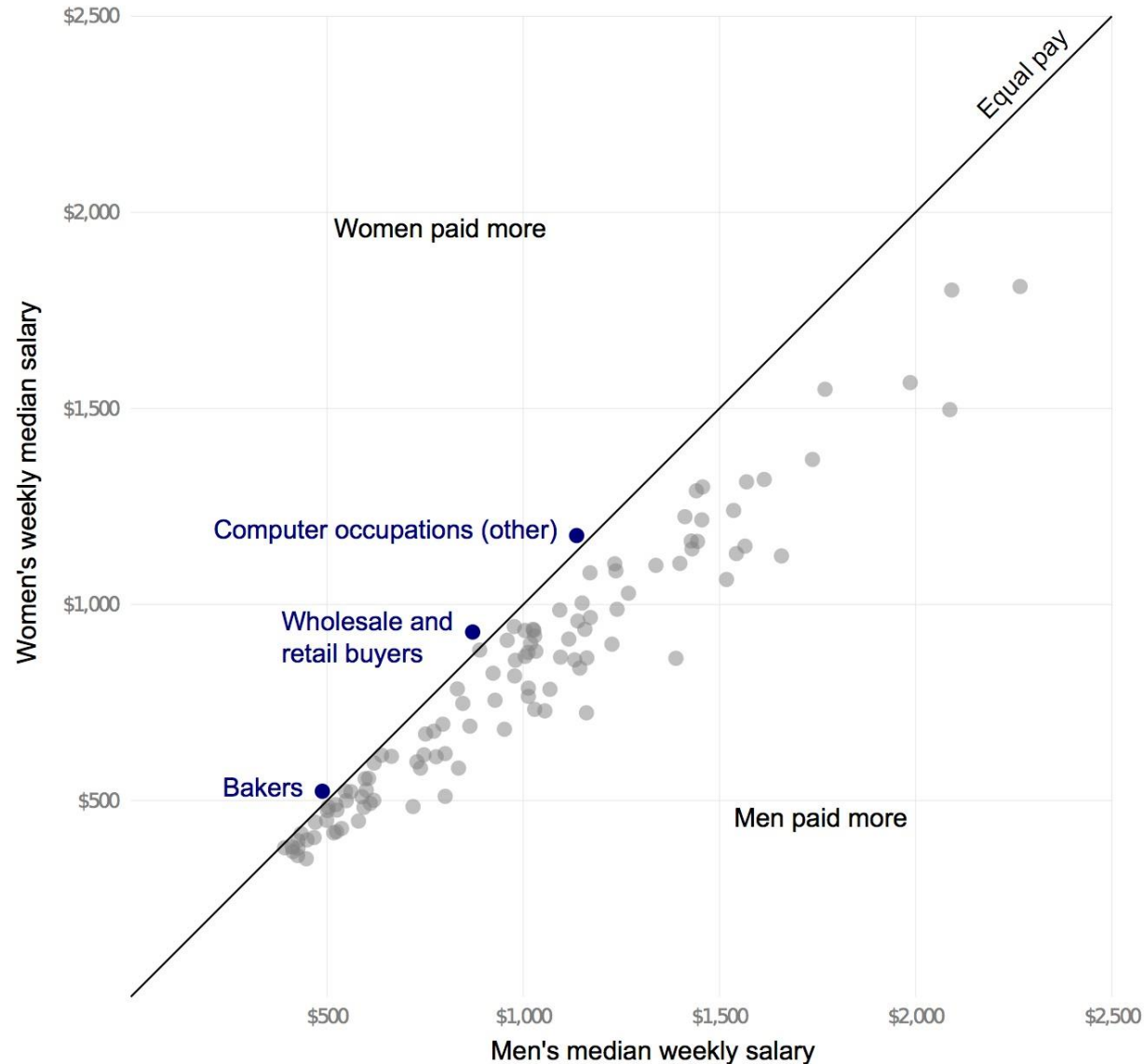
# Highlight the story: minimize chart junk



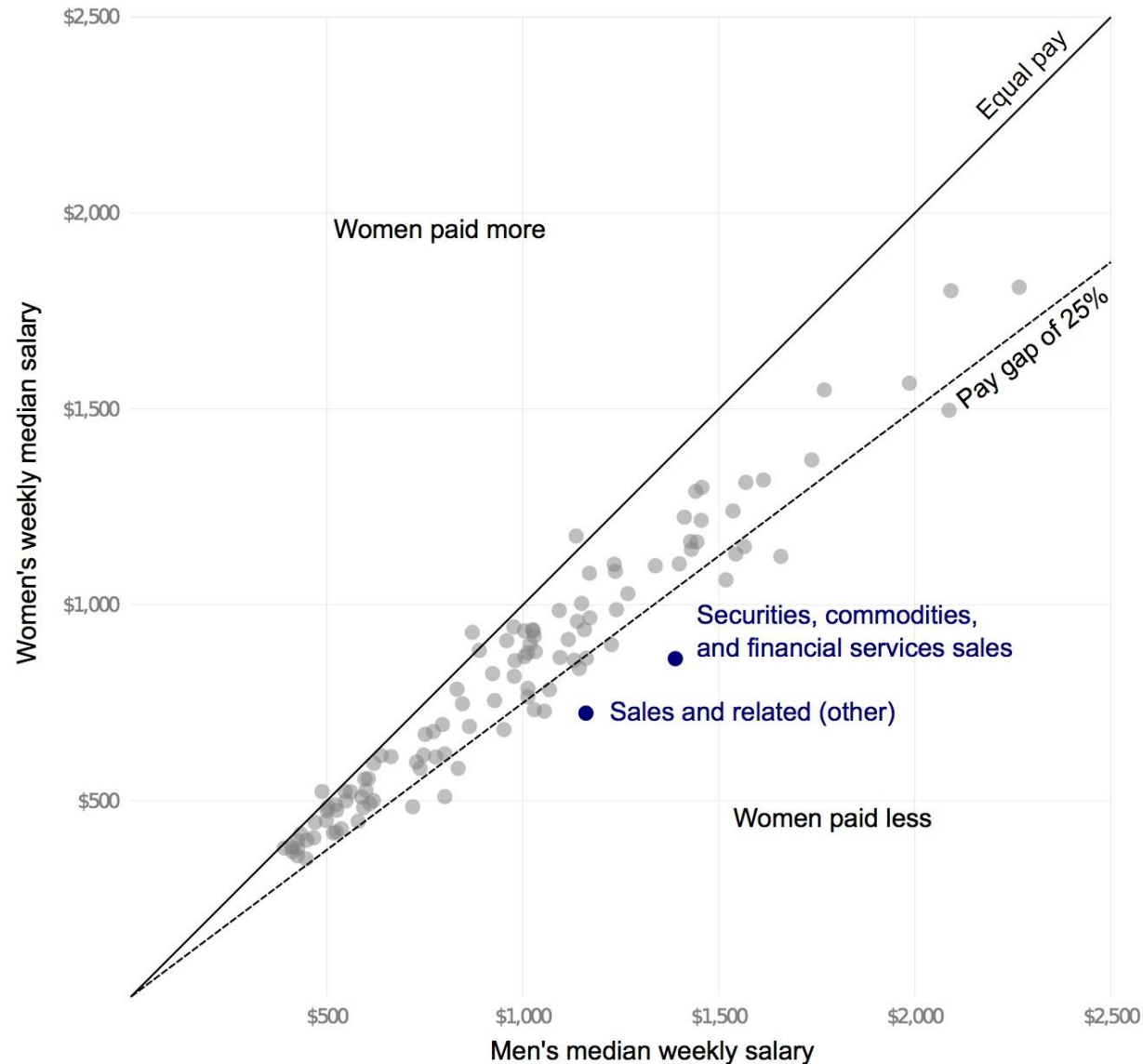
# Highlight the story: labels and annotation



# 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

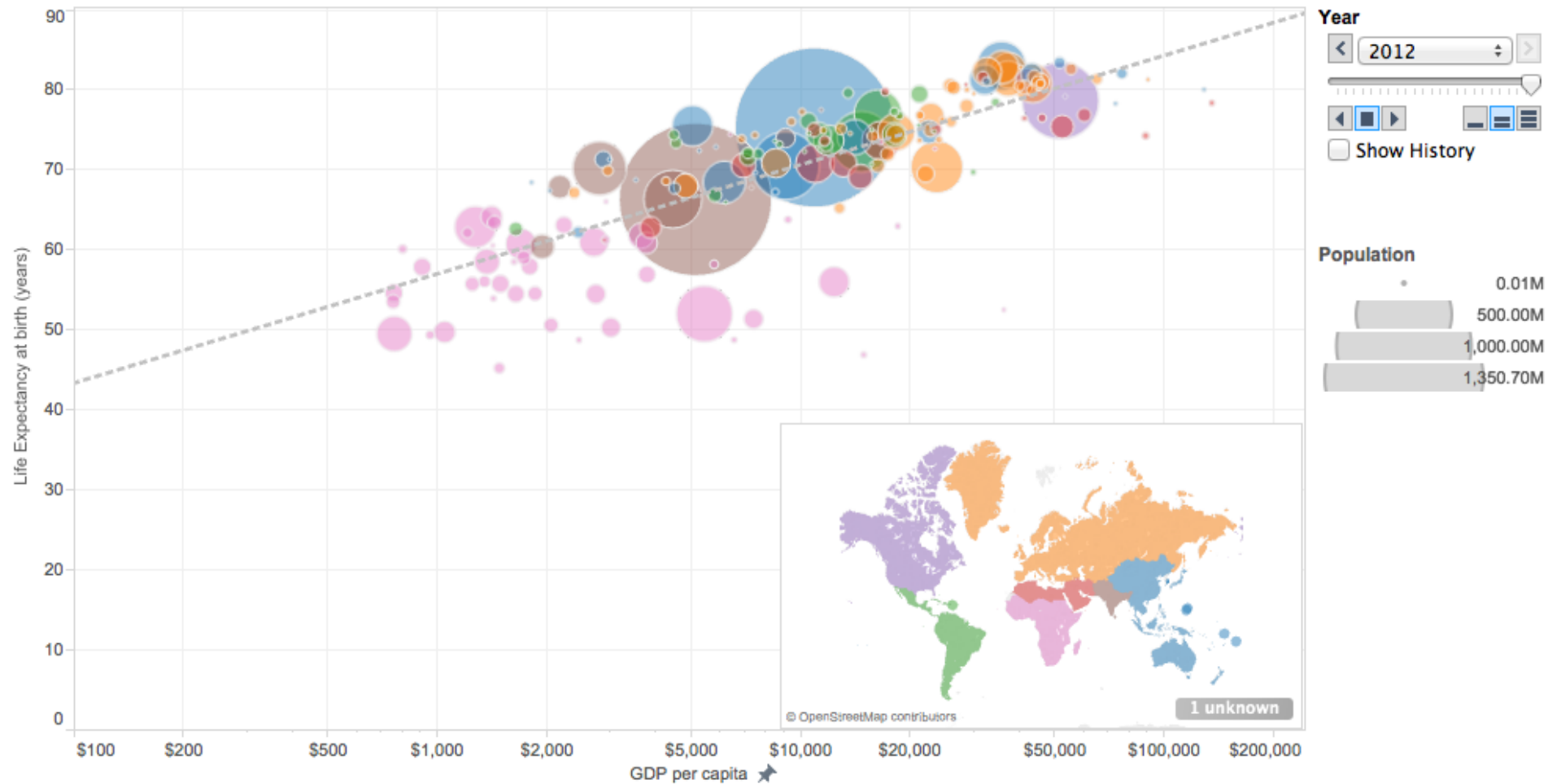
Show people. If they're confused, try another approach



# Putting it all together

## The Health and Wealth of Nations

Explore life expectancy, GDP per capita and population from 1990 to 2012.

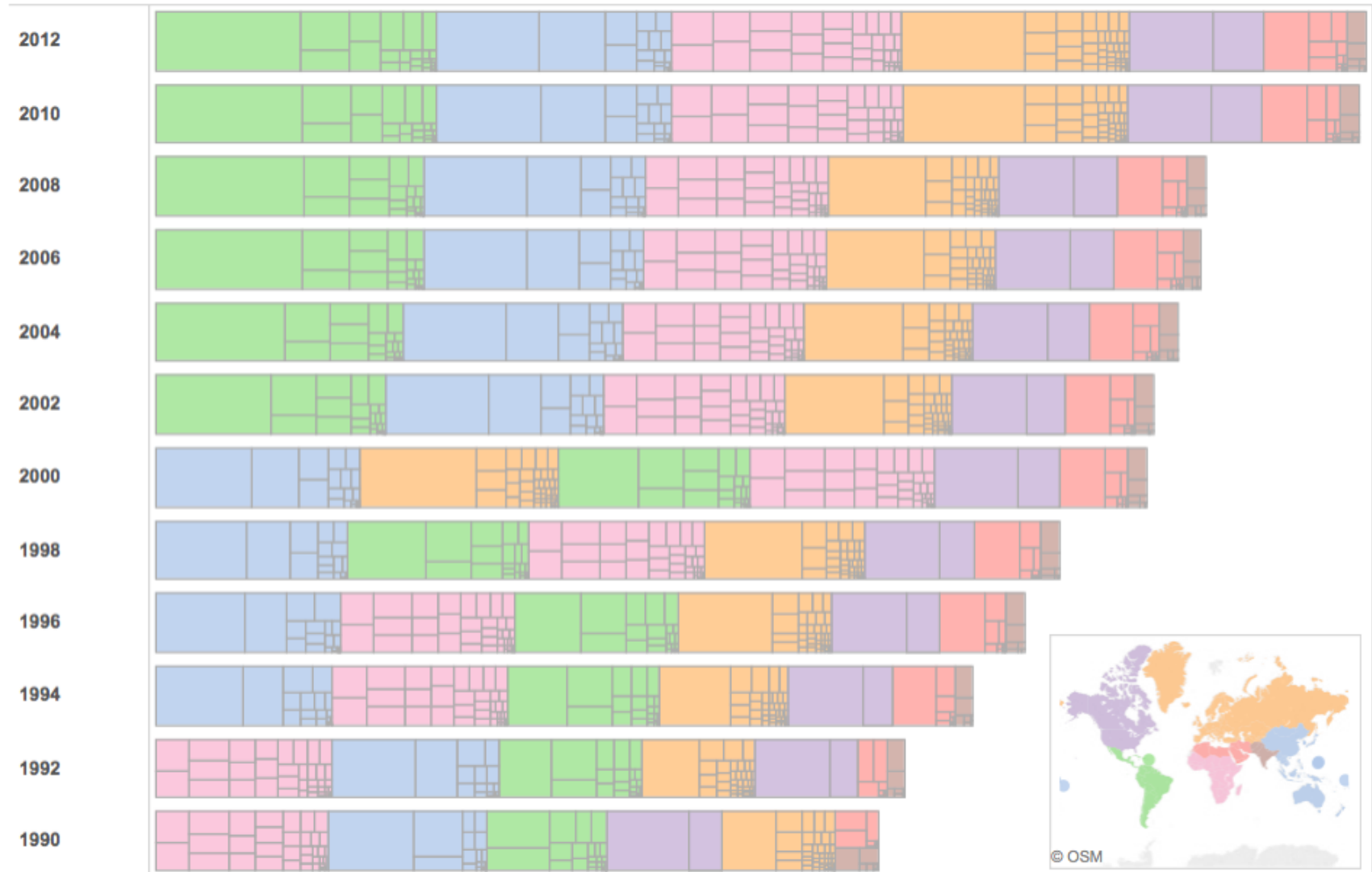


Source: World Bank Indicators; inspired by Gapminder

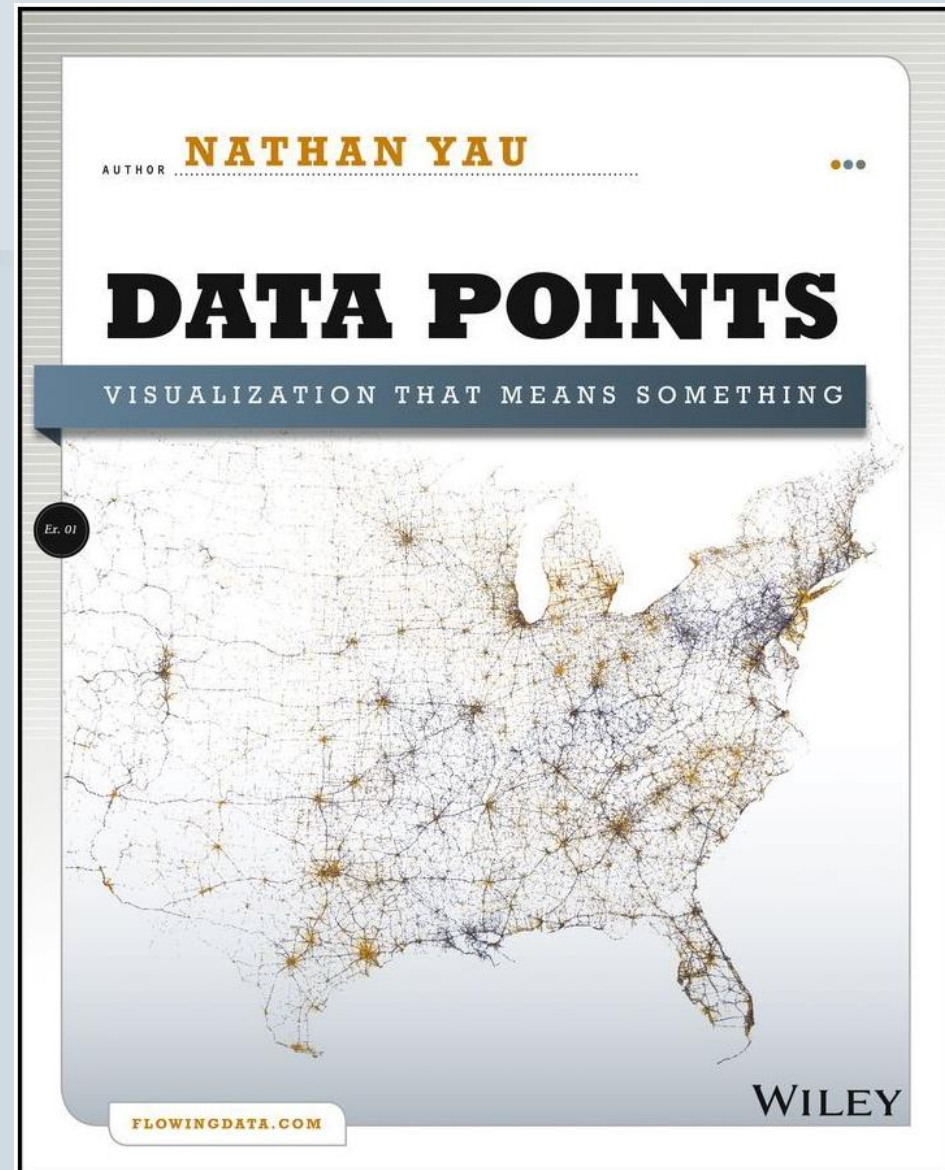
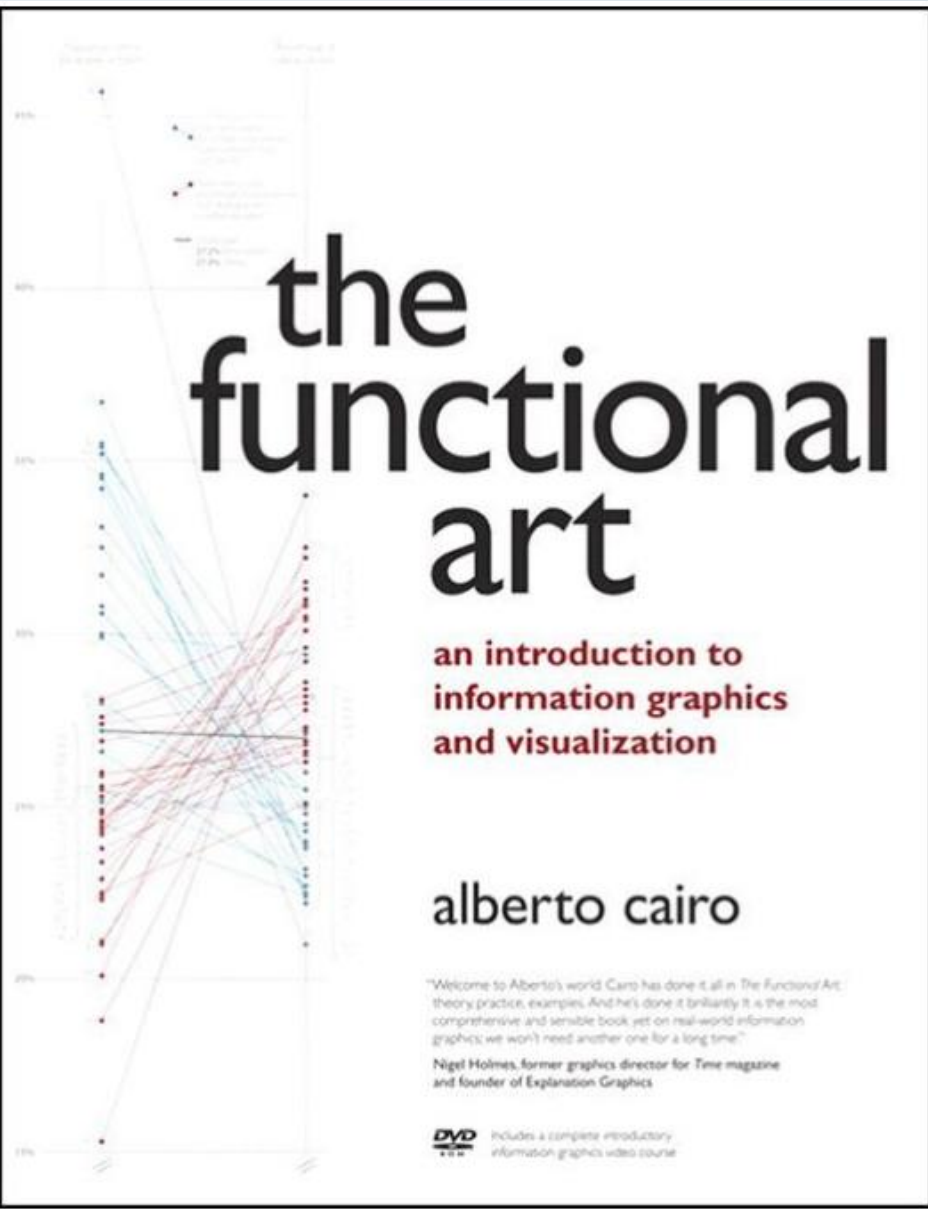
# Putting it all together

## Protected land: a growth area

Hover over rectangles to see the data for each country



# Recommended reading



# **Data visualization: basic principles**

**Peter Aldhous**

[@paldhous](#)