

# USING SPATIAL ENCODINGS EFFECTIVELY



UNIVERSITY OF  
**CALGARY**

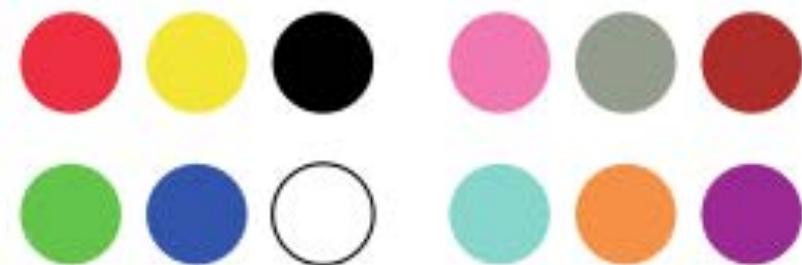
WITH SLIDES FROM JEFF HEER, CHARLES PERIN, & OTHERS

(previously)

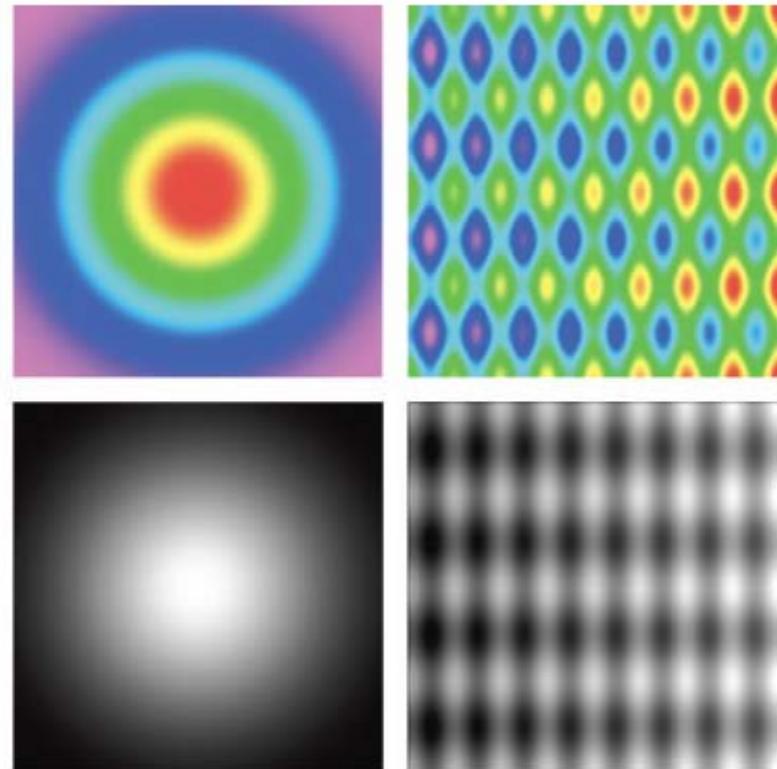
# COLOR AND VISUAL ENCODING

# COLOR

CATEGORIES - GOOD!



QUANTITIES - BAD!



# VISUAL VARIABLES

## ④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Same

Color saturation



Same

Curvature



Same

Volume (3D size)



Same

## ④ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



Shape



Most

Effectiveness

Least

**SPATIAL ENCODINGS ARE THE MOST VALUABLE.  
USE THEM TO SUPPORT SPATIAL REASONING.**

# TOPICS

Common and effective spatial encodings

Axes and scales

Space and visual embellishments

Aspect ratios and legibility

Focus + context and spatial distortion

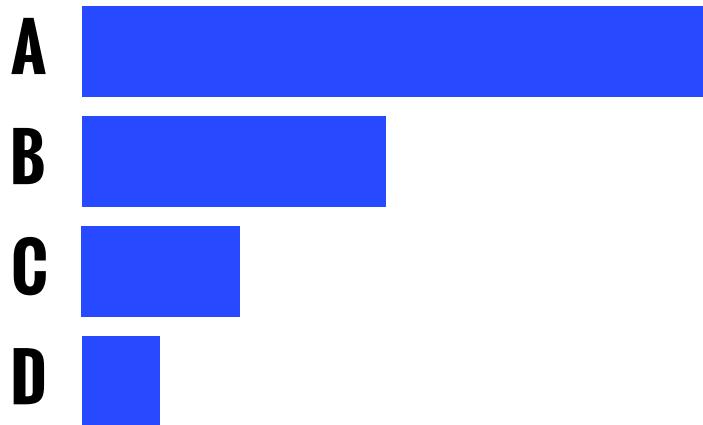
# DISPLAYING DATA USING SPACE

## Visual cues

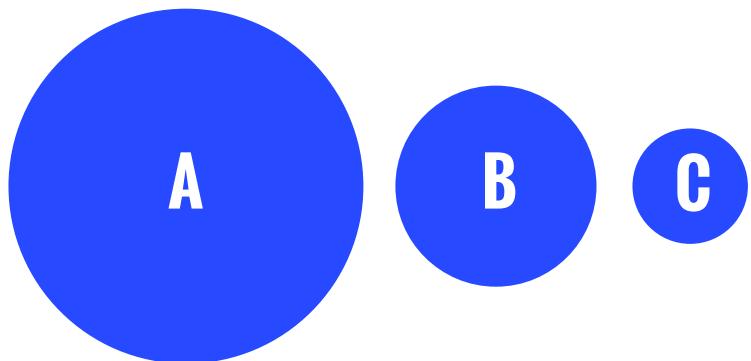
	Position	Length	Direction	Area or Volume	Color
Coordinate systems					

# VISUALIZING CATEGORIES

## BAR CHARTS



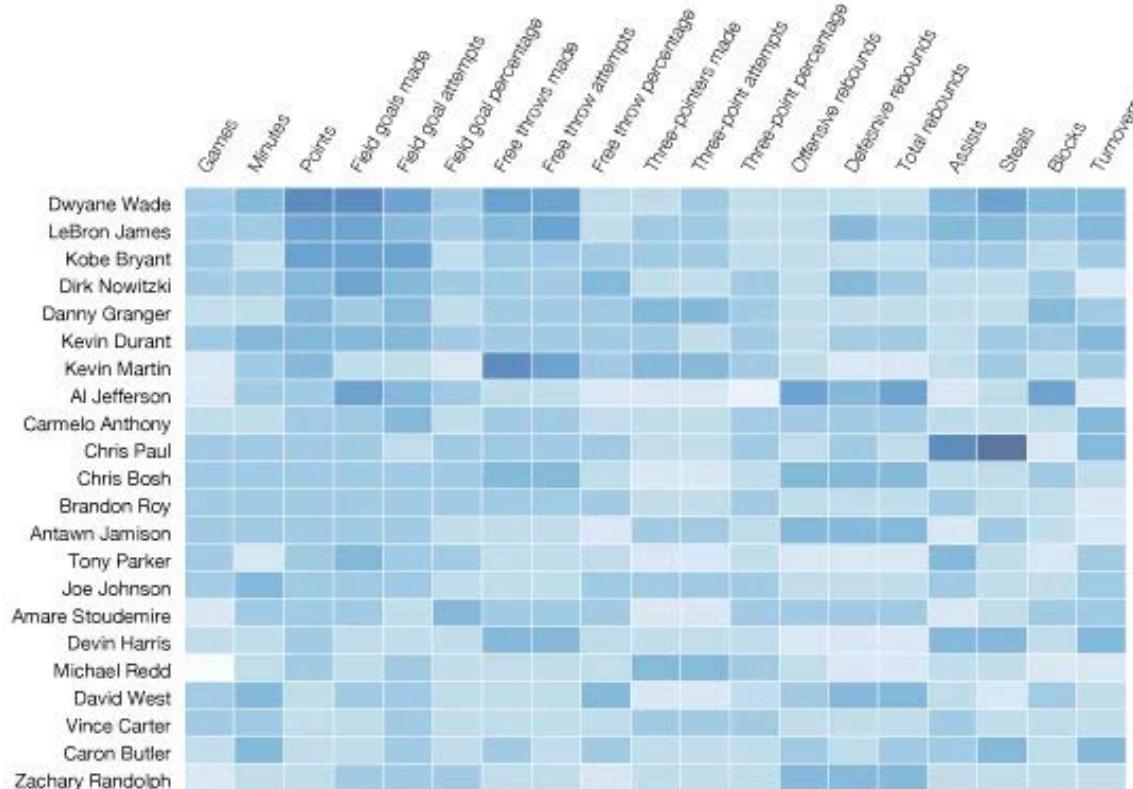
## SYMBOL PLOTS



# RELATIONSHIPS BETWEEN CATEGORIES

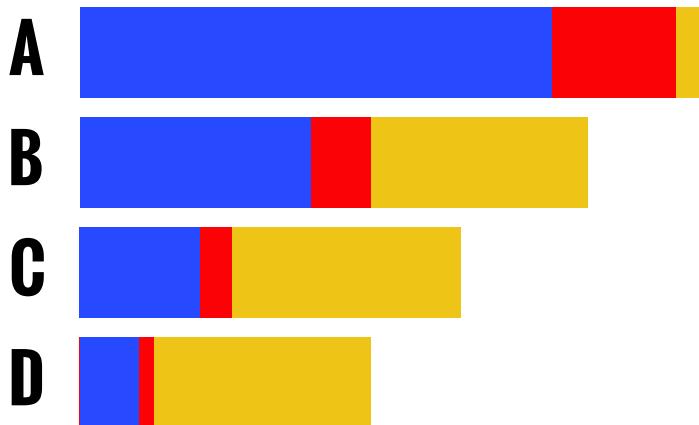
NBA per game performance of top 50 scorers

2008-2009 season

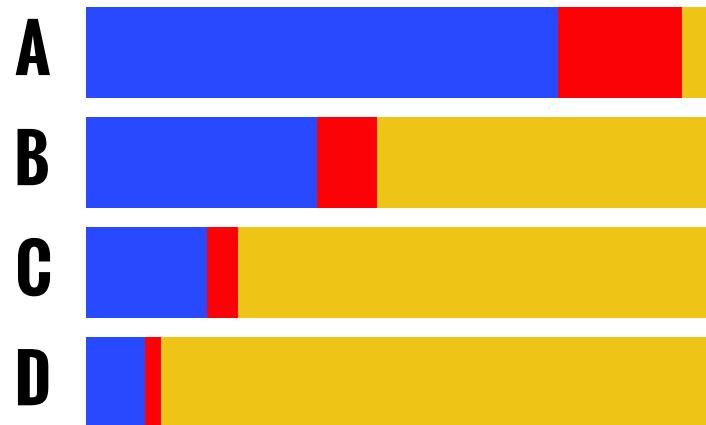


# PARTS OF A WHOLE

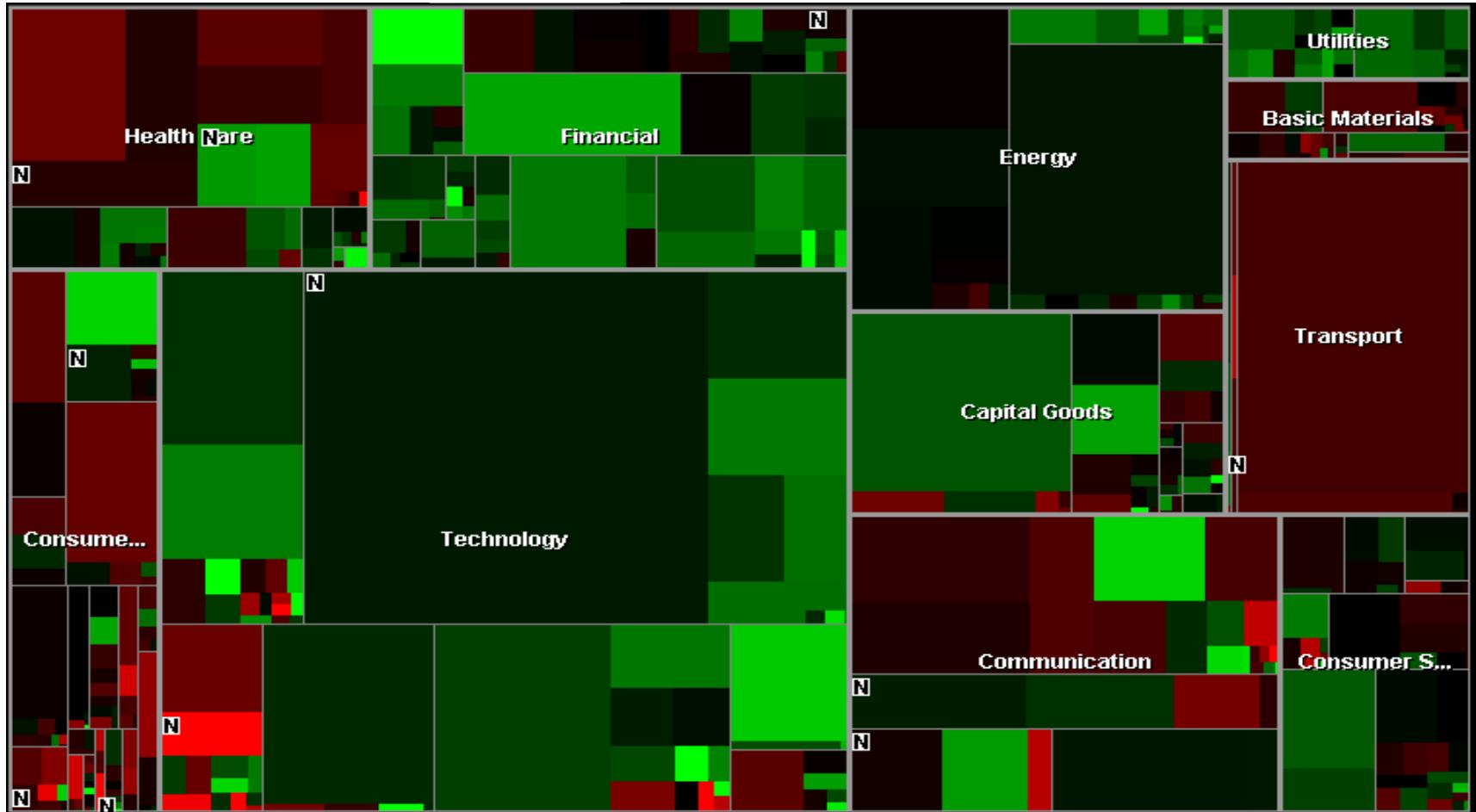
**STACKED BARS**



**(NORMALIZED)**



# NESTED CATEGORIES

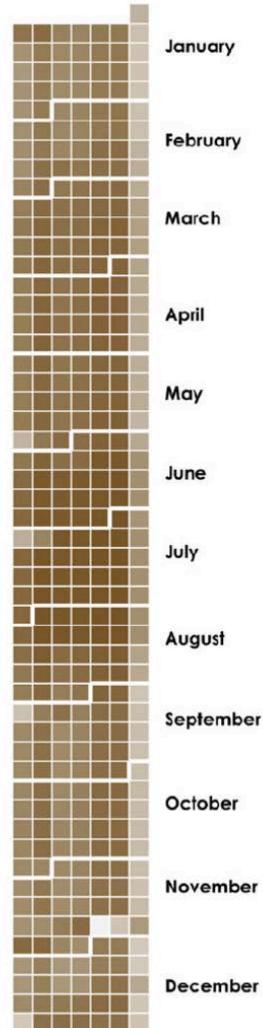


# TIME

Unemployment rate

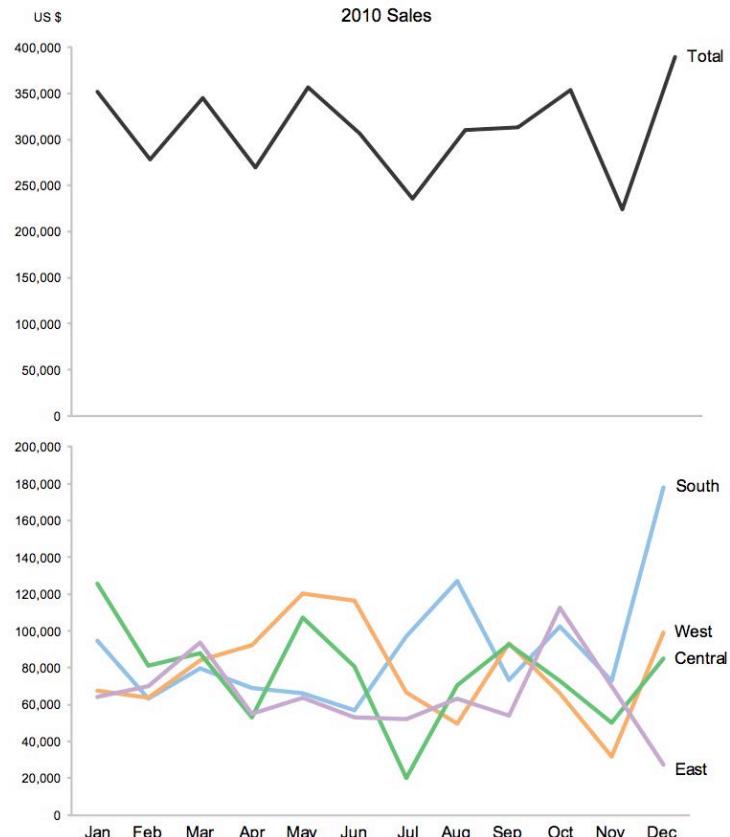
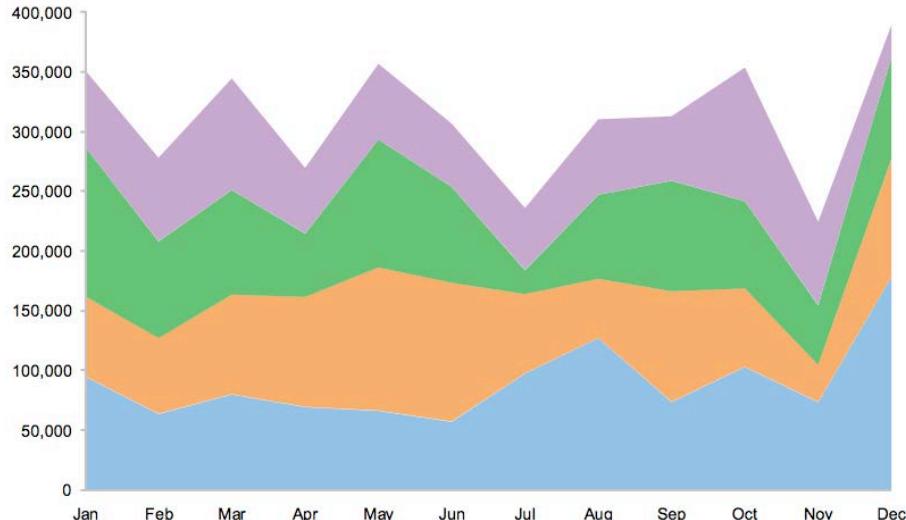


Daily Flights



NATHAN YAU

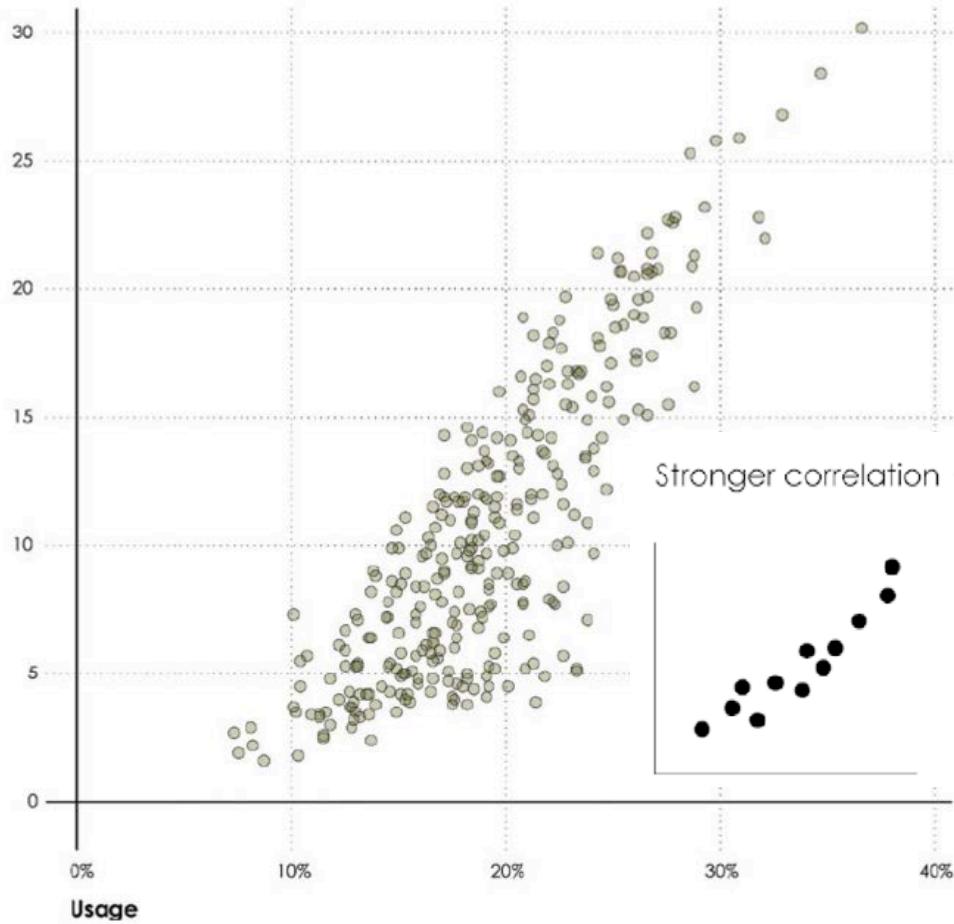
# CATEGORIES OVER TIME



STEPHEN FEW

# TWO QUANTITATIVE

Points per game



Stronger correlation

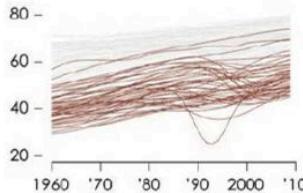
Weak correlation

# TWO QUANTITATIVE + CATEGORIES

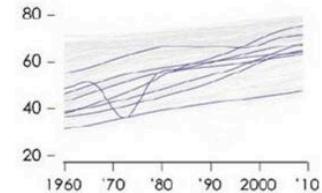
## Increasing Life Expectancy

According to data from World Bank, the number of years a person lives on average has been steadily increasing over the decades. However, as seen in some regions, war and economic turmoil can lead to sudden dips.

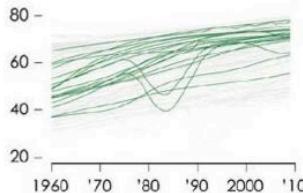
### Sub-Saharan Africa



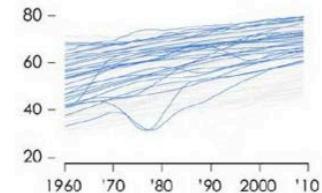
### South Asia



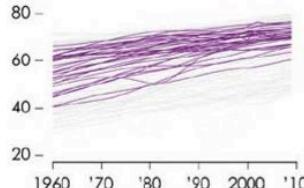
### Middle East and North Africa



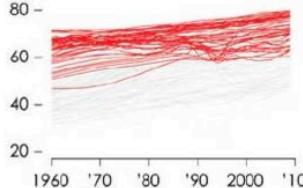
### East Asia and Pacific



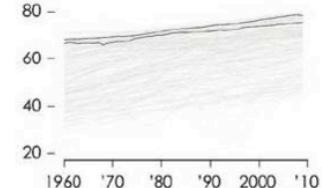
### Latin America and Caribbean



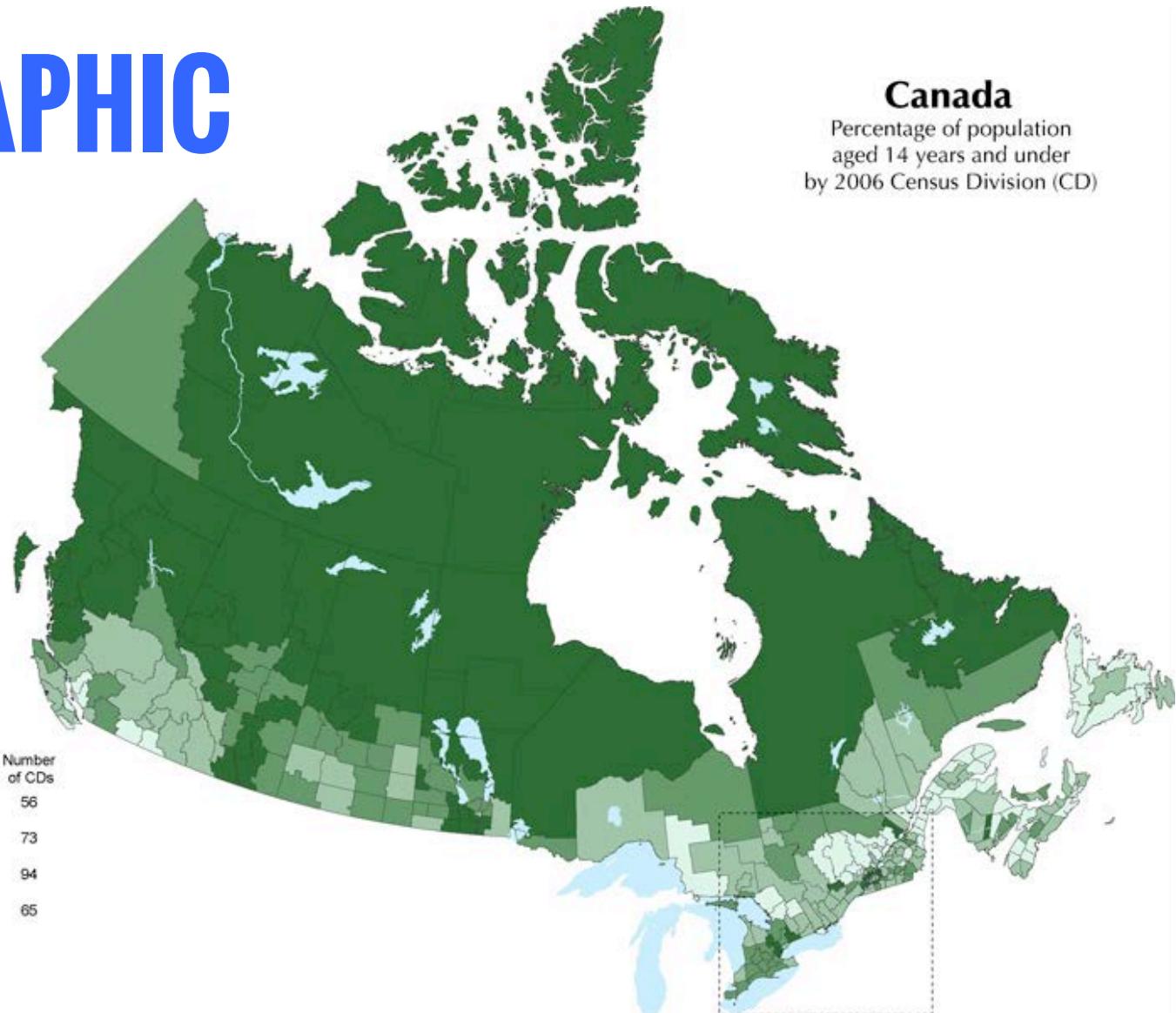
### Europe and Central Asia



### North America

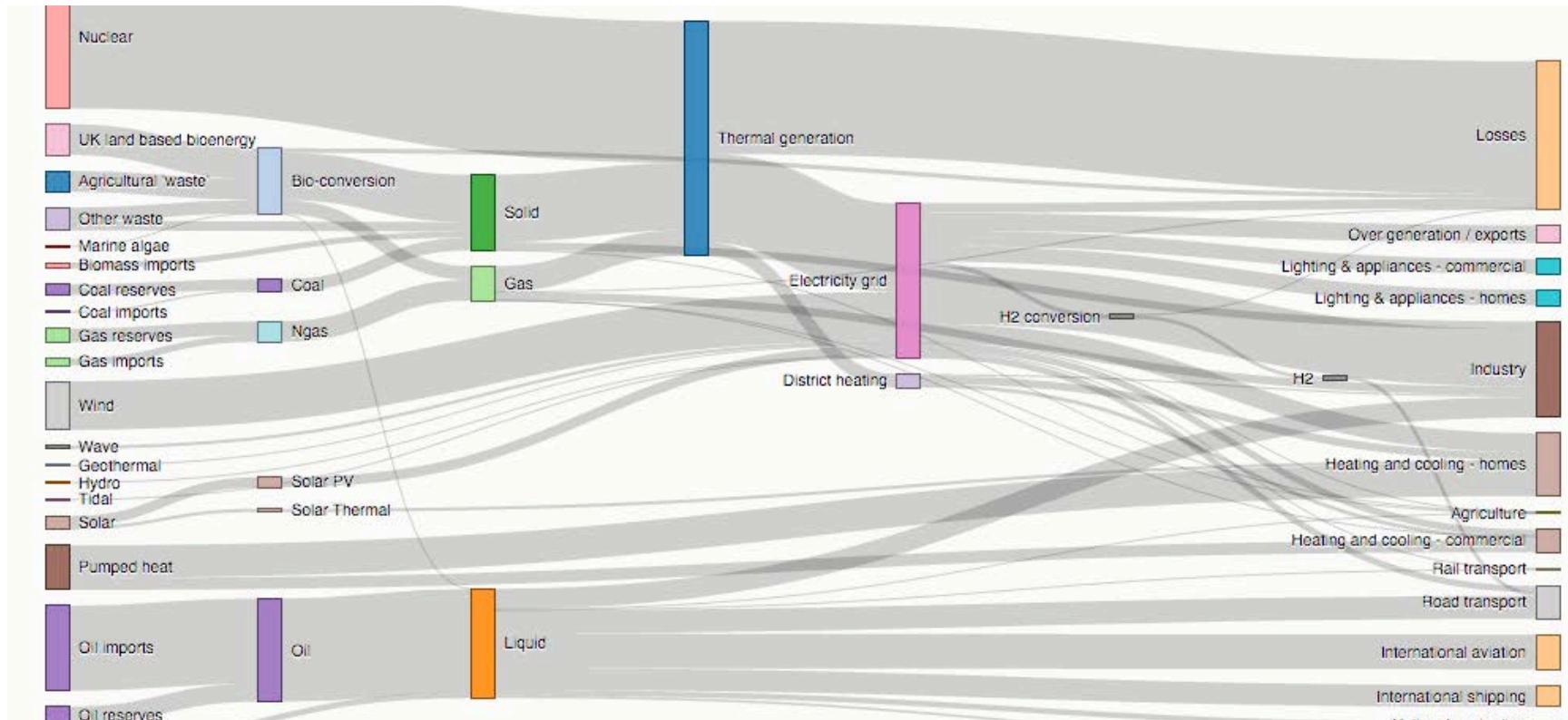


# GEOGRAPHIC



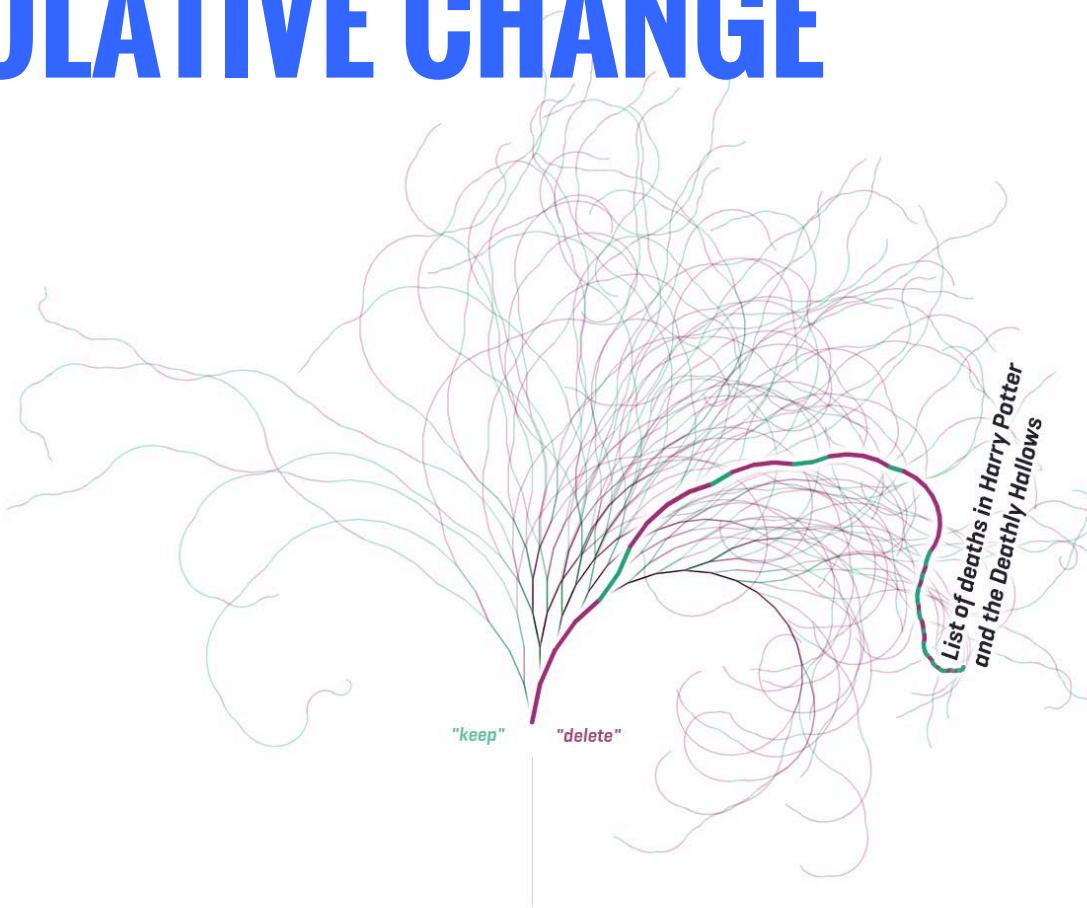
STATISTICS CANADA

# FLOW



MIKE BOSTOCK

# CUMULATIVE CHANGE

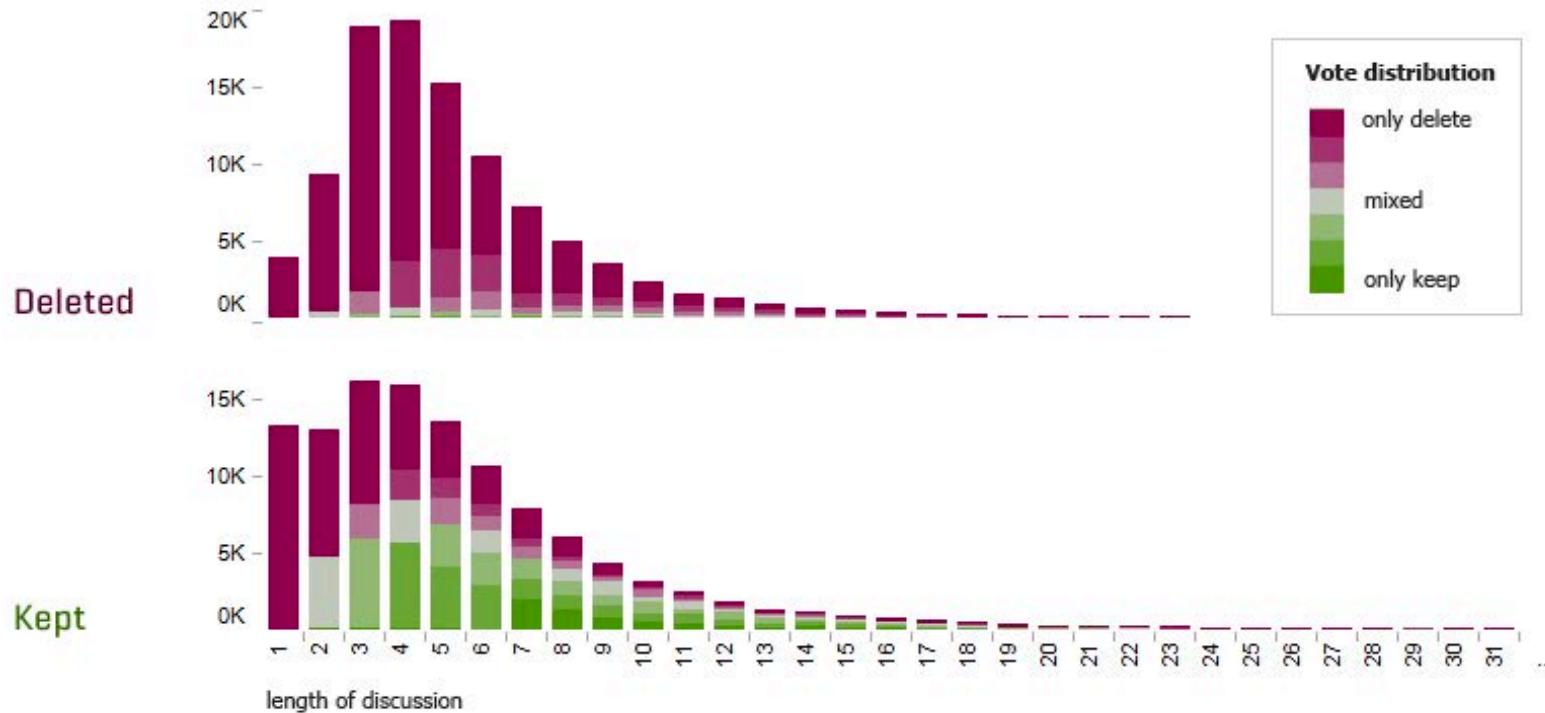


THE DELETED

The 100 longest Article for Deletion (AfD) discussions on Wikipedia,  
which resulted in deletion of the article.

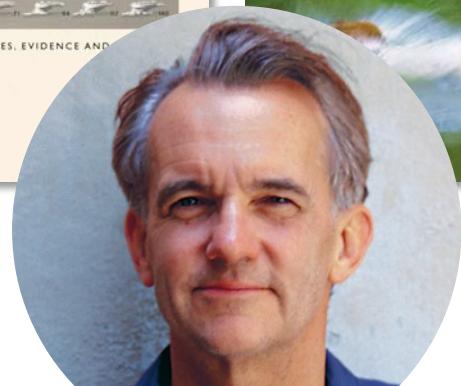
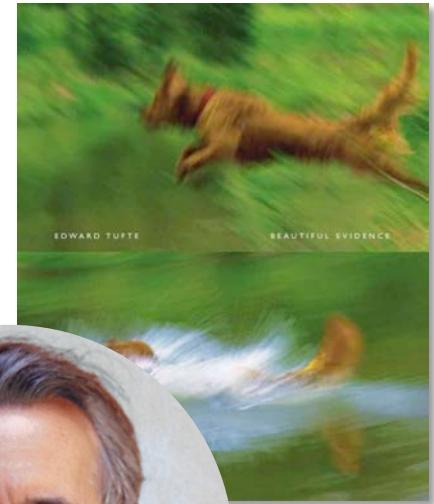
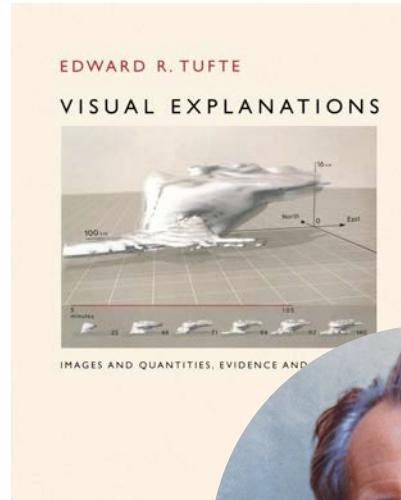
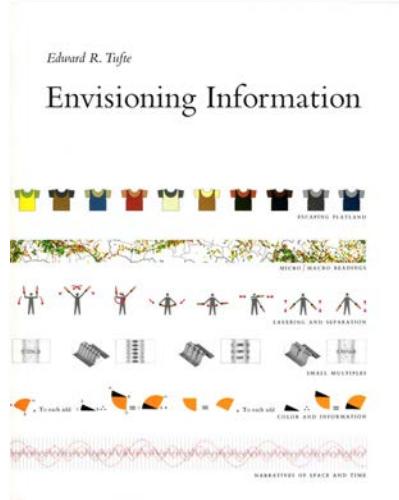
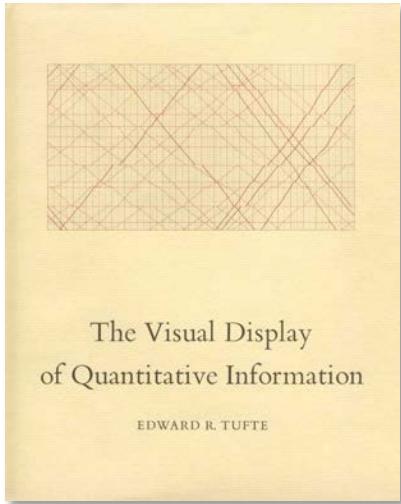
NOTABILA.NET

MORITZ STEFANER • DARIO TARABORELLI •  
GIOVANNI LUCA CIAMPAGLIA



AND MANY MORE OPTIONS  
HOW TO USE SPACE  
EFFICIENTLY?

# TUFTÉ'S INTEGRITY PRINCIPLES



**EDWARD TUFTÉ**

# TUFTÉ'S INTEGRITY PRINCIPLES

MAXIMIZE THE DATA-INK RATIO

AVOID CHART JUNK

LAYER INFORMATION

MAXIMIZE THE DATA DENSITY

- *SHRINK THE GRAPHICS*
- *MAXIMIZE THE AMOUNT OF DATA SHOWN*

# TUFTÉ'S INTEGRITY PRINCIPLES

**"CLEAR, DETAILED, AND THOROUGH  
LABELING SHOULD BE USED TO DEFEAT  
GRAPHICAL DISTORTION AND AMBIGUITY."**

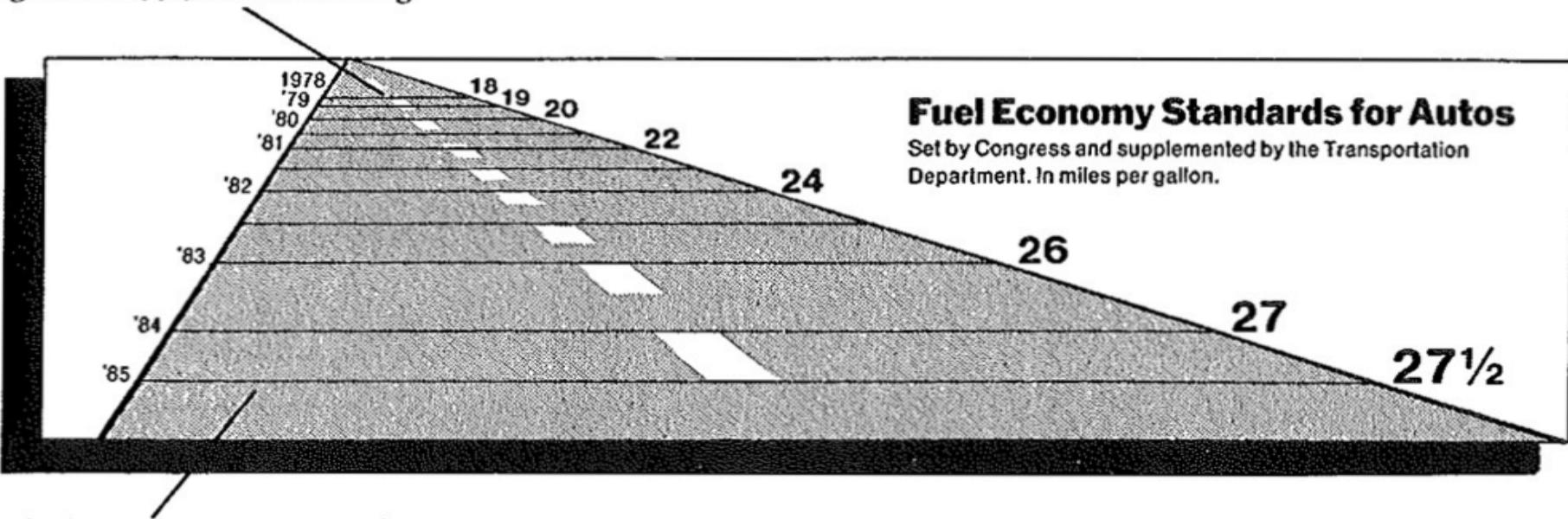
THE REPRESENTATION OF NUMBERS, AS  
PHYSICALLY MEASURED ON THE SURFACE  
OF THE GRAPHIC ITSELF, SHOULD BE  
**DIRECTLY PROPORTIONAL** TO THE  
NUMERICAL QUANTITIES REPRESENTED.



**EDWARD TUFTE**

# DISTORTION

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

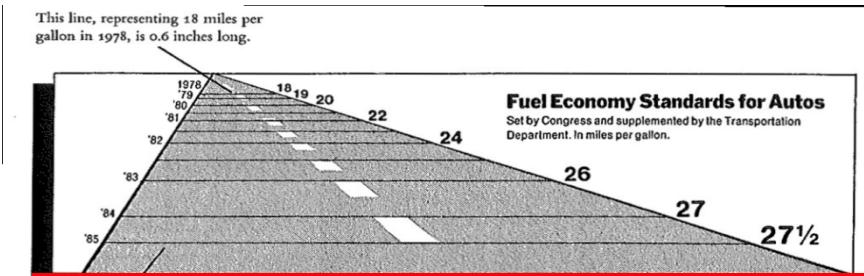
**THE LIE**

**GRAPHIC.**  $(5.5-0.6)/0.6 \times 100\% = 783\%$

**DATA:**  $(27.5-18.0)/18 \times 100\% = 53\%$

**LIE FACTOR:**  $783/53 = 14.8$

**PHIC)**



# THE LIE FACTOR MUST BE ZERO

# **DATA-INK RATIO**

**MAXIMIZE THE RATIO OF**

$$\frac{\text{(INK USED TO SHOW DATA)}}{\text{(TOTAL INK USED)}}$$

# DATA-INK RATIO

**Remove**  
to improve  
(the **data-ink** ratio)

# DATA-INK RATIO

**Remove**  
to improve  
the **data tables** edition

# DATA-INK RATIO

**Remove**  
to improve  
the **pie chart** edition

# DATA-INK RATIO

**Remove**  
to improve  
the **map** edition

# MINIMIZE CHART JUNK

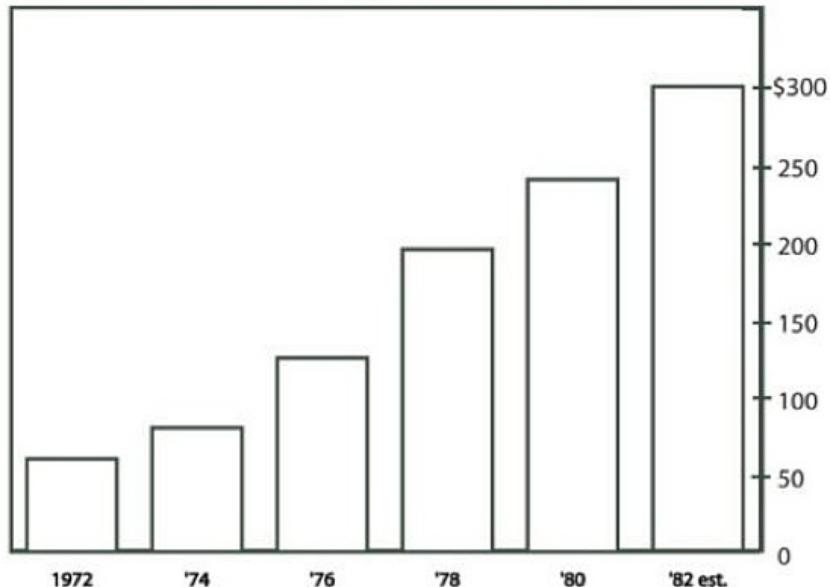
## MONSTROUS COSTS

Total House and Senate campaign expenditures, in millions



## MONSTROUS COSTS

Total House and Senate campaign expenditures, in millions



Wayne Lytle

---

The Dangers of  
*GLITZINESS*  
and other  
Visualization Faux Pas

---

or... "What's Wrong with this Visualization?"

# **DATA DENSITY**

**MAXIMIZE THE RATIO OF:**

**(NUMBER OF ENTRIES IN DATA)**

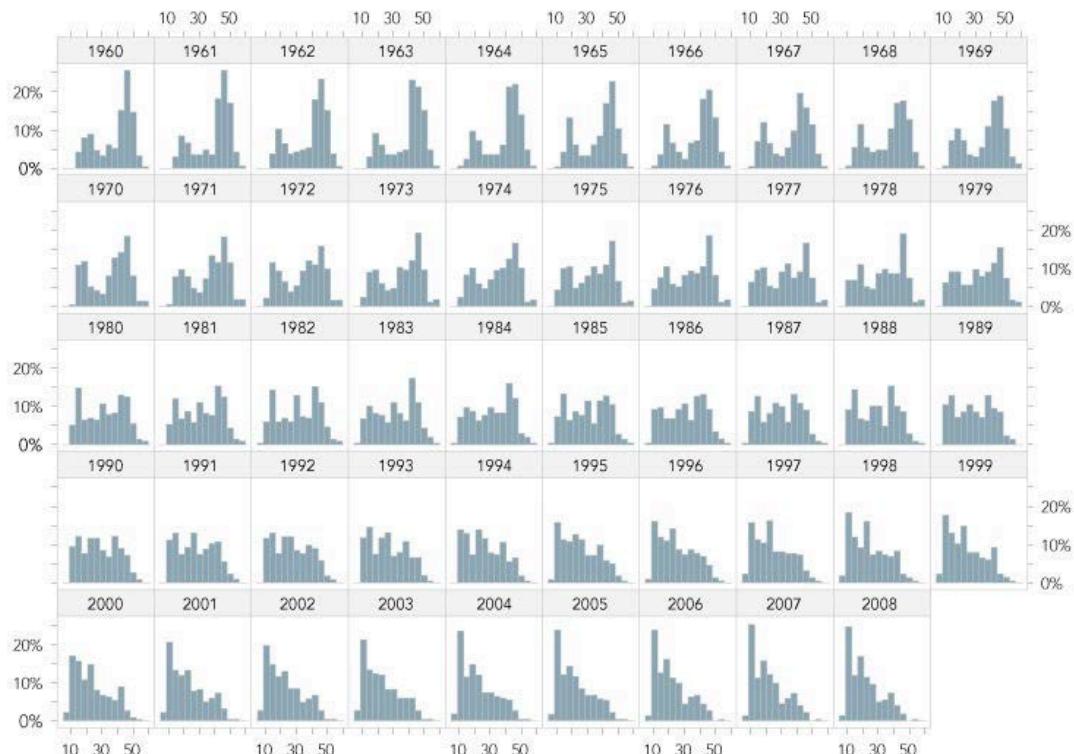
---

**(AREA OF THE GRAPHIC)**

# DATA DENSITY - SHRINK THE GRAPHICS

Annual Worldwide Distributions of Live Births

*Live births per 1,000 population*



**“SMALL MULTIPLES”**

# DATA DENSITY - SHRINK THE GRAPHICS

## GRAPHIC PROBLEMS POSED BY TIME SERIES

### Scale in years

With a scale in years, a two-year total (figure 1) should be divided by 2 (figure 2). A total for six months should be multiplied by 2.

### Pointed curves

For overly pointed curves (figure 3), the scale of the Q should be reduced; optimum angular perceptibility occurs at around 70 degrees (figure 4).

If the curve is not reducible (large and small variations), filled columns can be used (figure 5).

### Flat curves

For overly flat curves (figure 6), the scale of the Q should be increased (figure 7).

### Small variations

For small variations in relation to the total (figure 8), the total loses its importance, and the zero point can be eliminated, provided the reader is made aware of this elimination (figure 9). The graphic can be interpreted as an acceleration if a precise study of the variations is necessary; here, we use a logarithmic scale (figure 10). (See also page 240.)

### Large range

For a very large range between the extreme numbers (figure 11), we must either:

- (1) leave out the smallest variations;
- (2) be concerned only with relative differences (logarithmic scale), without knowing the absolute quantities;
- (3) select different parts (periods) within the ordered component and treat them on different scales above the common scale (figure 12).

### Obvious periodicity

If there is obvious periodicity (figure 13), and the study involves a comparison of the phases of each cycle, it is preferable to break up the cycles in order to superimpose them (figure 14). A polar construction can be used, preferably in a spiral shape (figure 15), but we should not begin with too small a circle. As striking as it seems, it is less efficient than an orthogonal construction.

### Annual curves

For annual curves of rainfall or temperature, if a cycle has two phases (figure 17), why depict only one (figure 16)?

### A contrast

Unlike what we see in figure 18, the pertinent or "new" information must be separated from the background or "redundant" information. The background involves: (a) the invariant, highlighted by a heading (Port St. Michel); (b) the highly visible identification of each component (tonnage and date). The new information (the curve) must stand out from the background (figure 19).

### Reference points

It is impossible to utilize a graphic such as figure 20, except in a general manner. There is confusion concerning the position of the points, and no potential comparison is possible, as it is in figure 21.

### Precision reading

A precision reading (utilization on the elementary level, as in figure 24) is difficult in figure 22, which results in a poor reading of the order of the points, and in figure 23, where there is ambiguity concerning the position of the points. On the other hand, figure 22 does favor overall vision (correlation).

### Null boxes

Curves accommodate null boxes poorly (figure 25). Columns (figure 26) are preferable.

### Unknown boxes

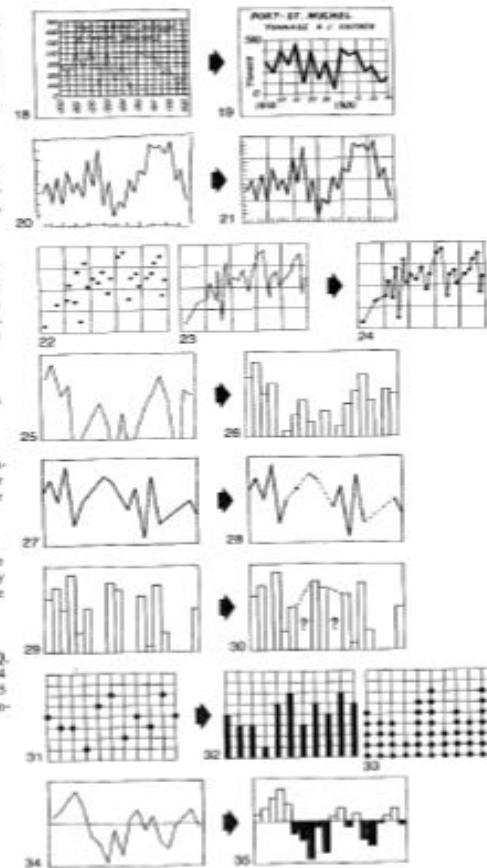
The drawing must indicate the unknowns of the information in an unambiguous way (figures 28 and 30). The reader might interpret figure 27 as a change in the structure of the curve and figure 29 as involving null values.

### Very small quantities

Except in seeking a correlation (quite improbable here) the number of ships entering into a port is represented better by figure 33 than by figures 31 or 32. The reader can perceive the numerical values at first glance.

### Positive-negative variation

This is in fact a problem involving three components O, Q,  $\neq (+-)$ , and it must be visually treated as such. Figure 34 can be improved by utilizing a retinal variable (in figure 35 a value difference: black-white) to differentiate the  $\neq$  component and thus highlight positive-negative variation.



# DATA DENSITY - SHRINK THE GRAPHICS

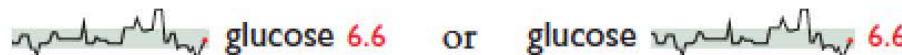
Placed in the relevant context, a single number gains meaning. Thus the most recent measurement of glucose should be compared with earlier measurements for the patient. This data-line shows the path of the last 80 readings of glucose:



Lacking a scale of measurement, this free-floating line is dequantified. At least we do know the value of the line's right-most data point, which corresponds to the most recent value of glucose, the number recorded at far right. Both representations of the most recent reading are tied together with a color accent:



Some useful context is provided by showing the *normal range* of glucose, here as a gray band. Compared to normal limits, readings above the band horizon are elevated, those below reduced:



# SPARKLINES & WORD-SCALE VIS

## Science fiction

---

From Wikipedia, the free encyclopedia

*For other uses, see [Science fiction \(disambiguation\)](#).*

33k visits in last 30 days

Science fiction is a genre of **fiction** dealing with imagination and creativity. It typically features futuristic settings, futuristic science and technology, space travel, time travel, parallel universes, and extraterrestrial life. It often explores the potential consequences of scientific and technological advancements on society and humanity.

# SPARKLINES & WORD-SCALE VIS

EASTERN EUROPE

Soviet cult and pragmatism in Transnistria

Experts worry that the next "Crimea" could be the breakaway region of Transnistria. Many locals there don't share that fear, and if the last referendum holds, a large majority would welcome a Russian annexation.



# SPARKLINES & WORD-SCALE VIS

Gonzalo Higuain slides  
a cross in from the right



and Ronaldo,  
at the front post, shoots  
off target.

# TUFTÉ'S INTEGRITY PRINCIPLES

MAXIMIZE THE DATA-INK RATIO

AVOID CHART JUNK (*SOMETIMES*)

LAYER INFORMATION

MAXIMIZE THE DATA DENSITY

- *SHRINK THE GRAPHICS*
- *MAXIMIZE THE AMOUNT OF DATA SHOWN (SOMETIMES)*

HOW DO YOUR CHARTS CHECK OUT?

## Visual cues

	Position	Length	Direction	Area or Volume	Color
Coordinate systems					

**AXES DEFINE SPACE**

### **Linear**

Values are evenly spaced



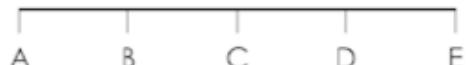
### **Logarithmic**

Focus on percent change



### **Categorical**

Discrete placement in bins



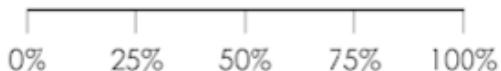
### **Ordinal**

Categories where order matters



### **Percent**

Representing parts of a whole



### **Time**

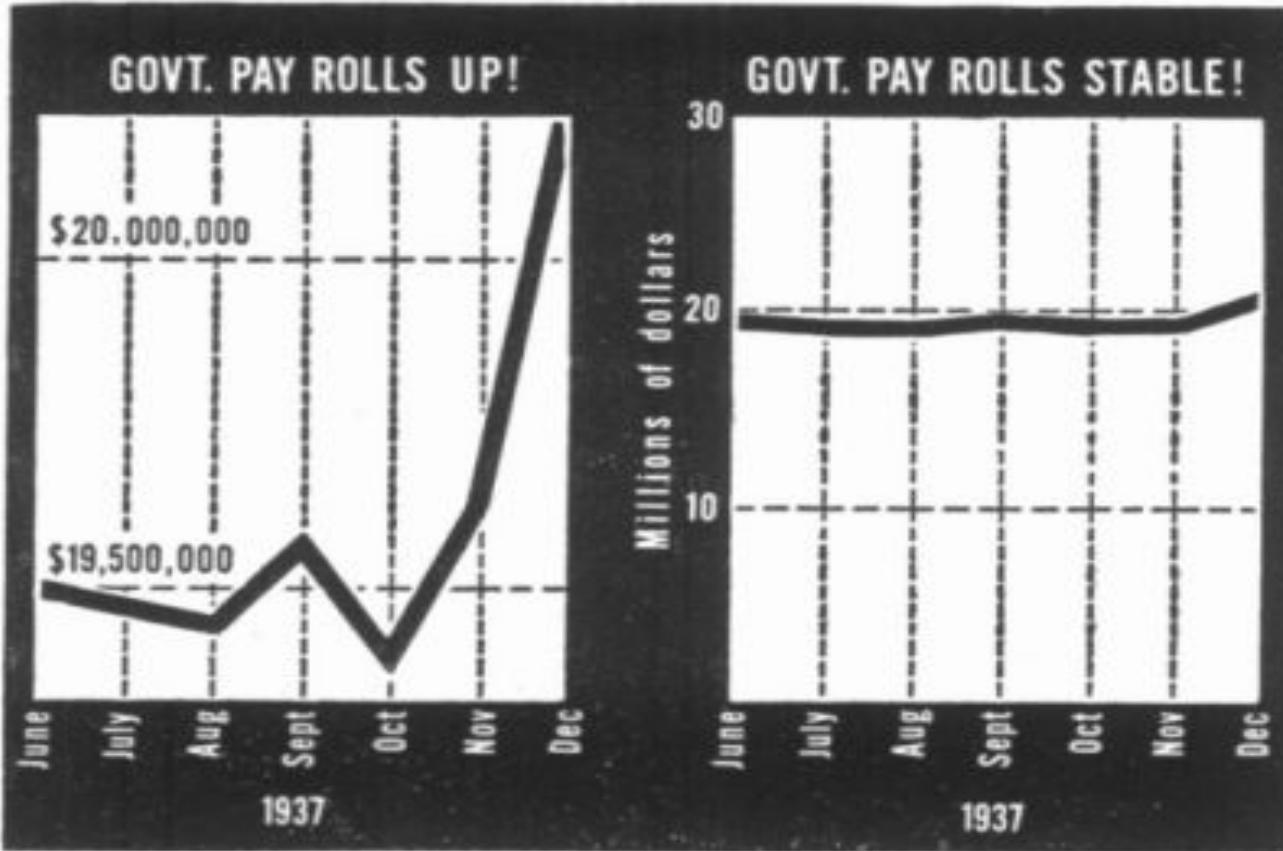
Units of months, days, or hours



**YOU HAVE TO MAKE CHOICES  
ABOUT AXES ALL THE TIME...**

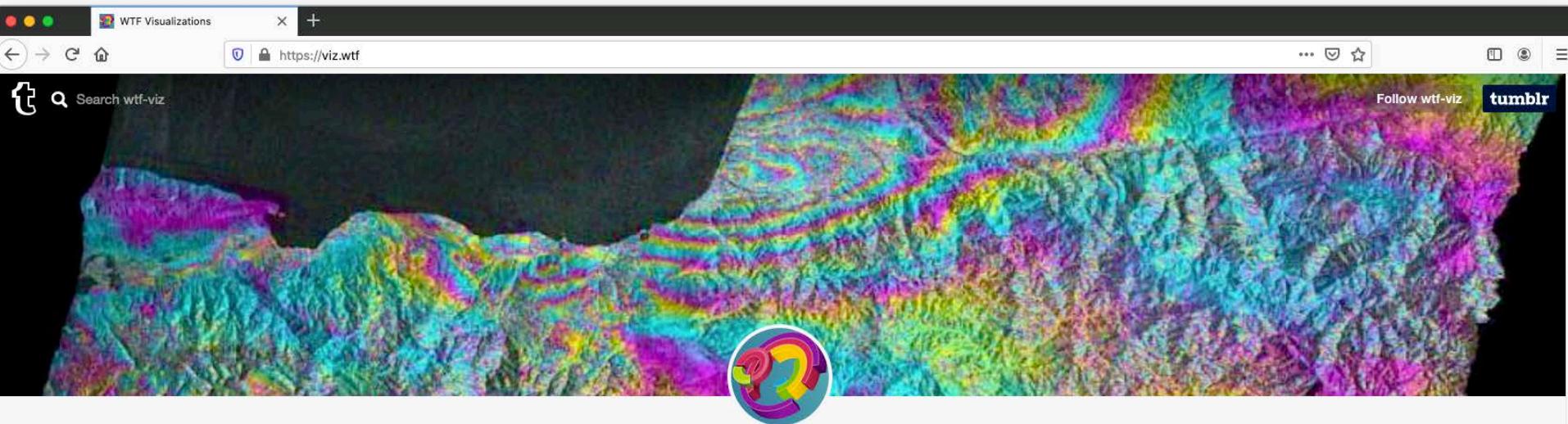
(and they can be the source of much confusion)

# WHICH GRAPH IS BETTER?



Government payrolls in 1937 [How To Lie With Statistics. Huff]

# MISLEADING AXES ARE PROBABLY THE SIGNATURE VISUALIZATION FAUX PAS...



## WTF Visualizations

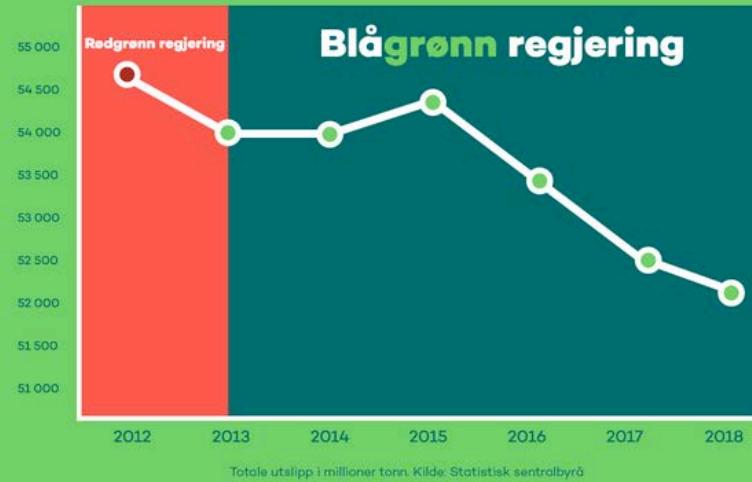
Visualizations that make no sense.

For a discussion of what is wrong with a particular visualization, tweet at us [@WTFViz](#).

[Submit a WTFViz you found.](#)

[SUBMIT A POST](#)   [ARCHIVE](#)

# Klimautslippene går ned med Venstre i regjering.



VENSTRE

How to make a 0.9% reduction in emissions look good.

#WTFViz #submission

6 notes

...





Terrible politicians make for terrible bar graphs

#BarChart #Percentages #submission

4 notes

...



## GPU performance

*iPhone 11* A13 Bionic

*iPhone XR* A12 Bionic

*Galaxy S10+* Snapdragon 855

*Huawei P30 Pro* Kirin 980

*Google Pixel 3* Snapdragon 845



iPhone 11 Presentation

#WTFViz #BarChart #submission

7 notes

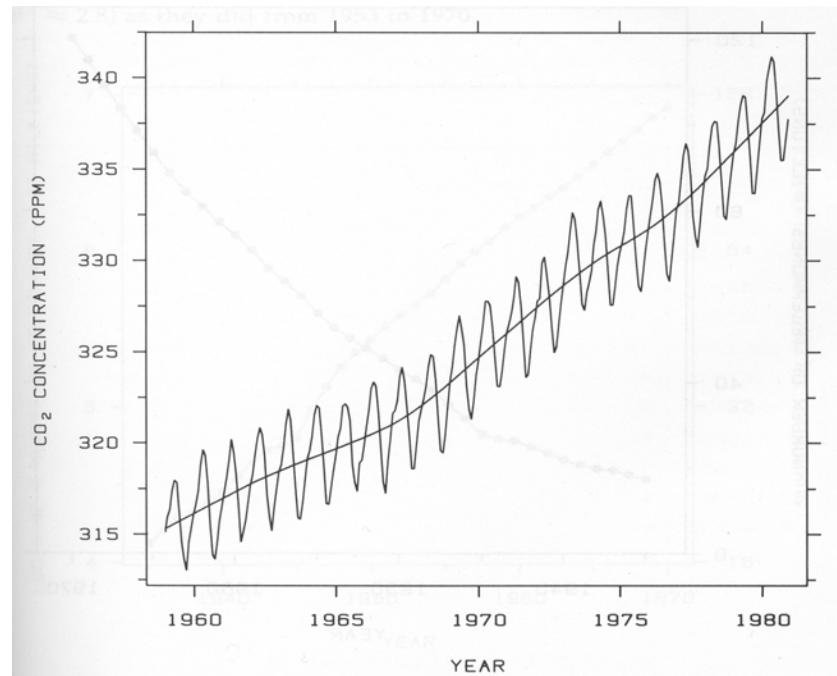
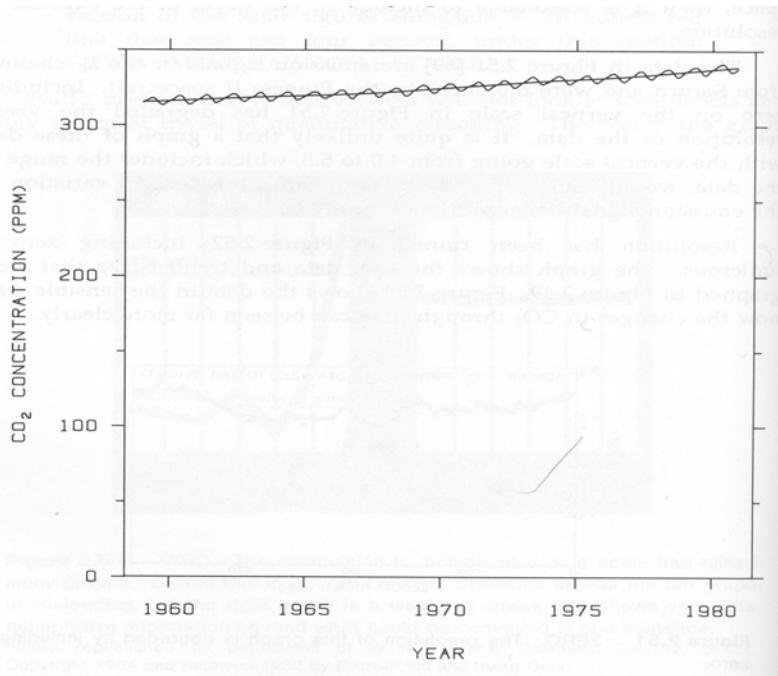
...



**WHAT CHOICES  
SHOULD YOU MAKE?**



# CLEARLY INDICATE YOUR AXIS BOUNDS



**YEARLY CO<sub>2</sub> CONCENTRATIONS [CLEVELAND 85]**

# RESCALE OR USE TWO CHARTS

**Australia's capital experienced unprecedented levels of air pollution**

Air quality index measured in Monash, Canberra

Hazardous

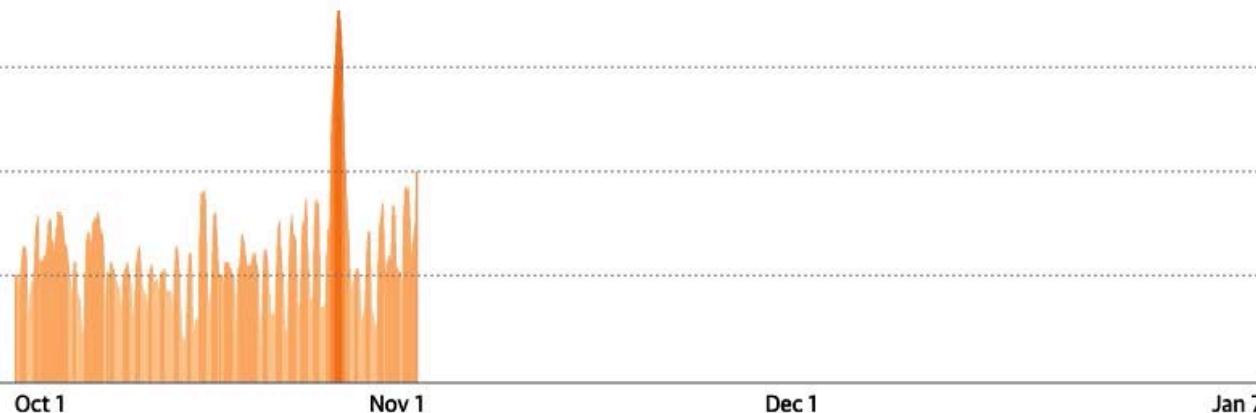
Very poor

Poor

Fair

Good

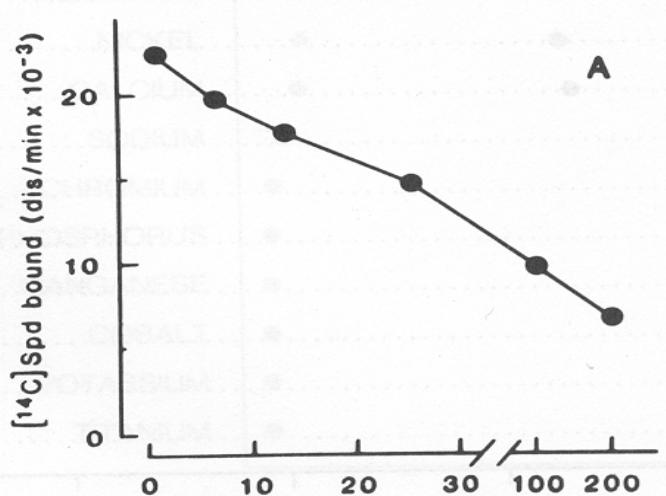
Very good



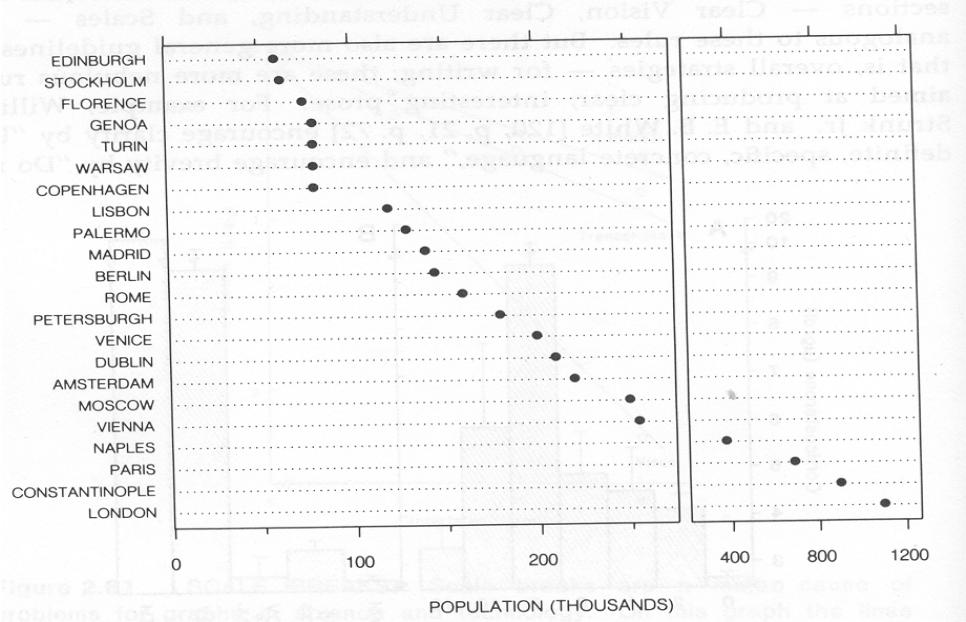
Source: data.act.gov.au

<https://www.theguardian.com/australia-news/2020/jan/07/visual-guide-see-how-australias-bushfires-are-raging-across-the-country>

# CLEARLY MARK SCALE BREAKS

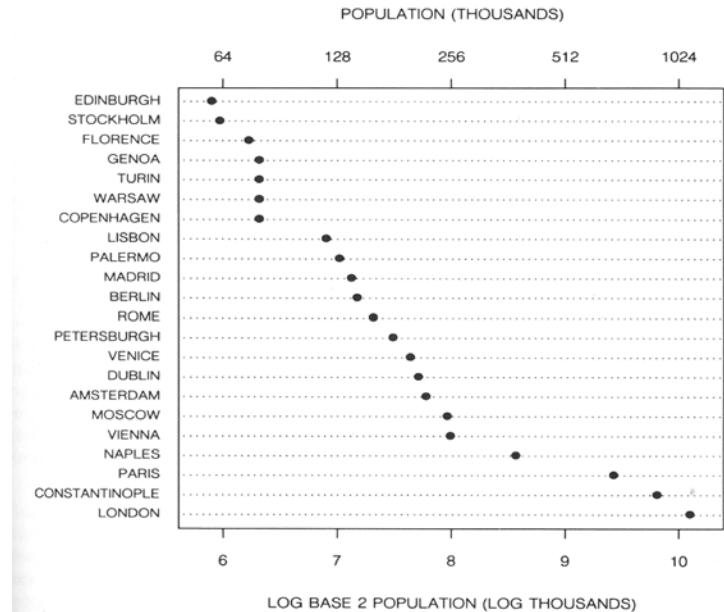
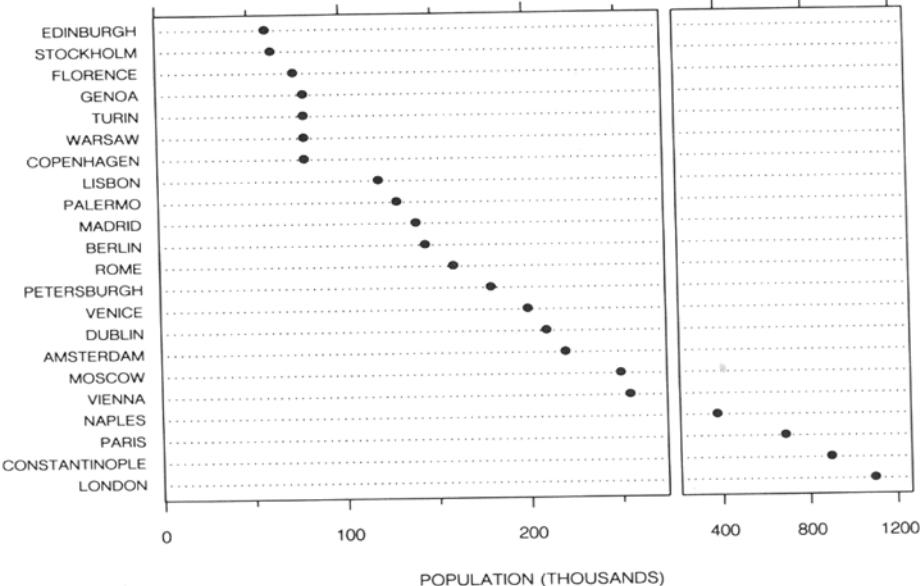


Poor scale break [Cleveland 85]



Well marked scale break [Cleveland 85]

# SCALE BREAK VS. LOG SCALE



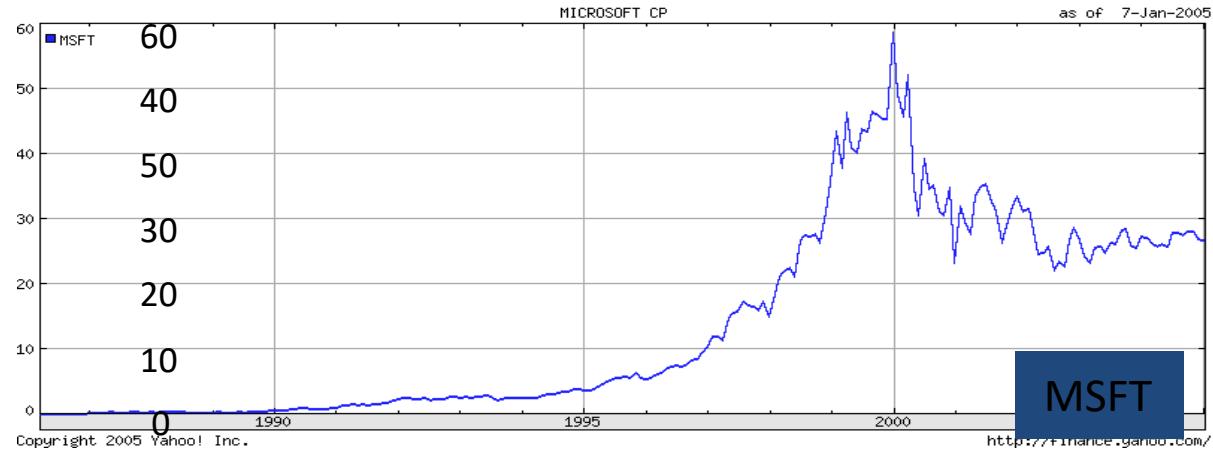
**BOTH INCREASE VISUAL RESOLUTION**

LOG SCALE - EASY COMPARISONS OF ALL DATA

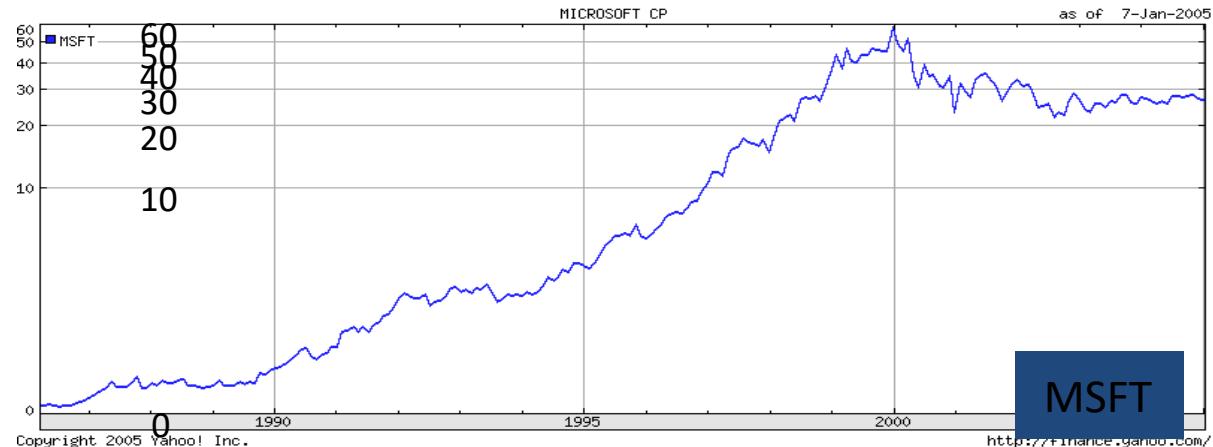
SCALE BREAK - MORE DIFFICULT TO COMPARE ACROSS BREAK

# LINEAR SCALE VS. LOG SCALE

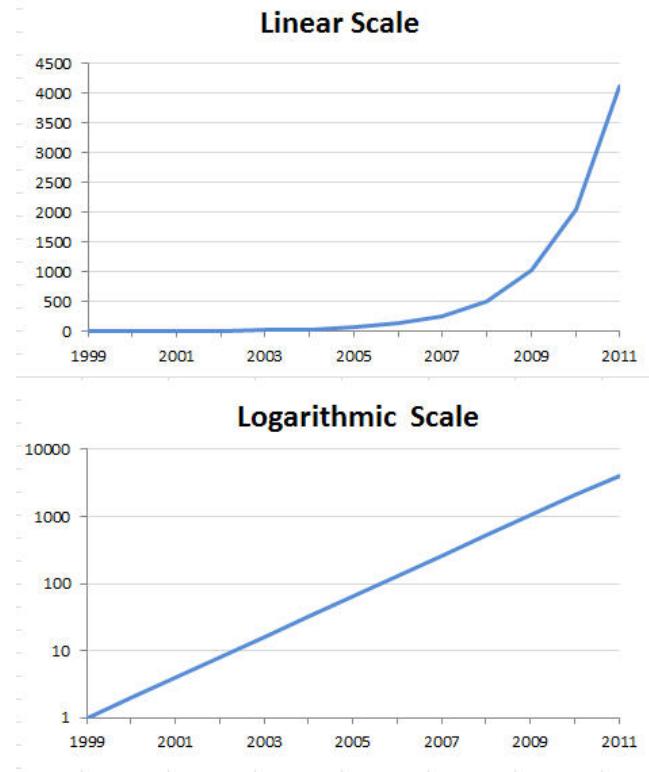
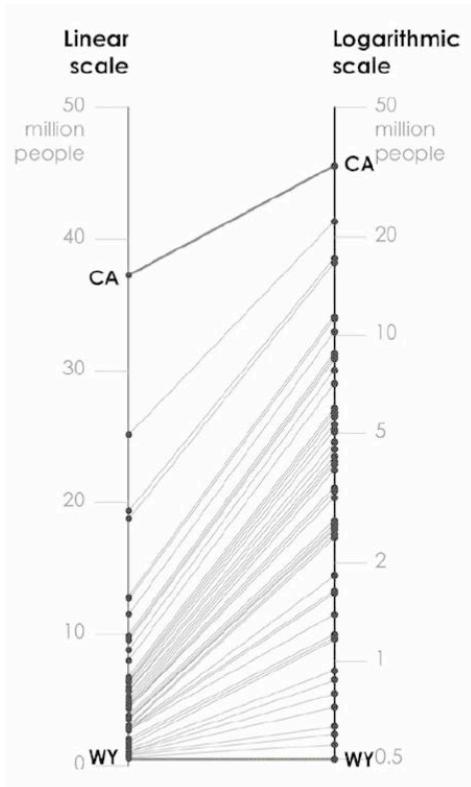
**LINEAR SCALE**  
ABSOLUTE CHANGE



**LOG SCALE**  
SMALL FLUCTUATIONS  
PERCENT CHANGE  
 $D(10,20) = D(30,60)$



# LINEAR SCALE VS. LOG SCALE

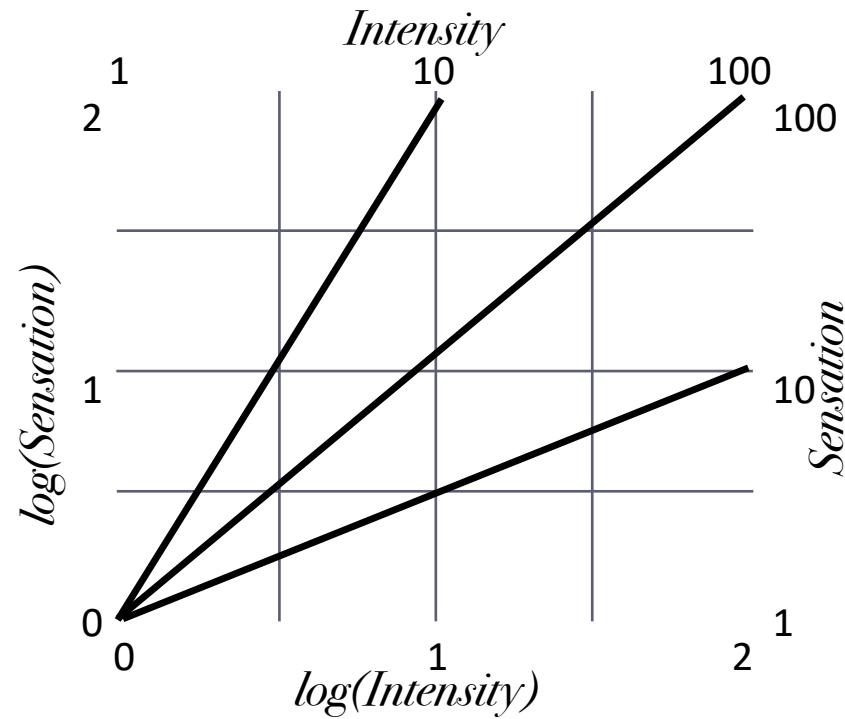
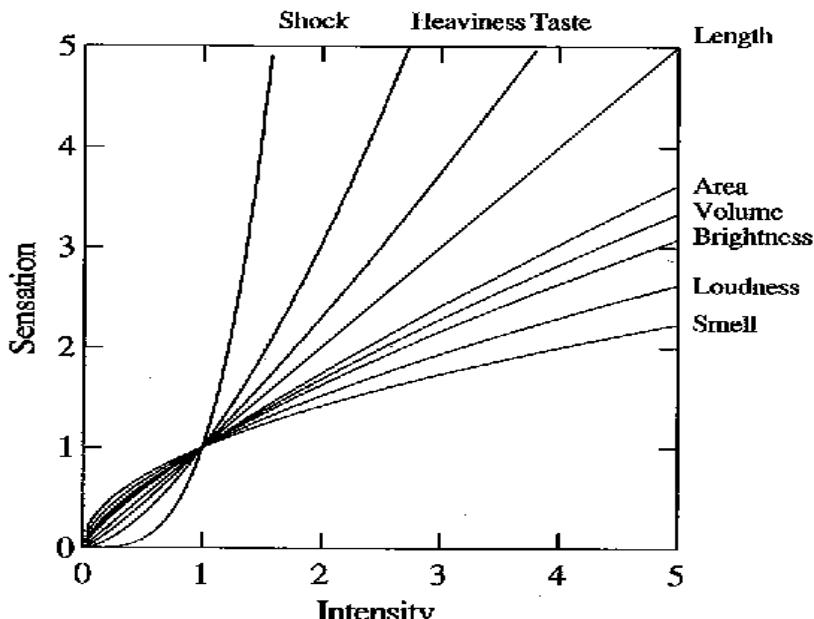


# LOG-LOG GRAPH

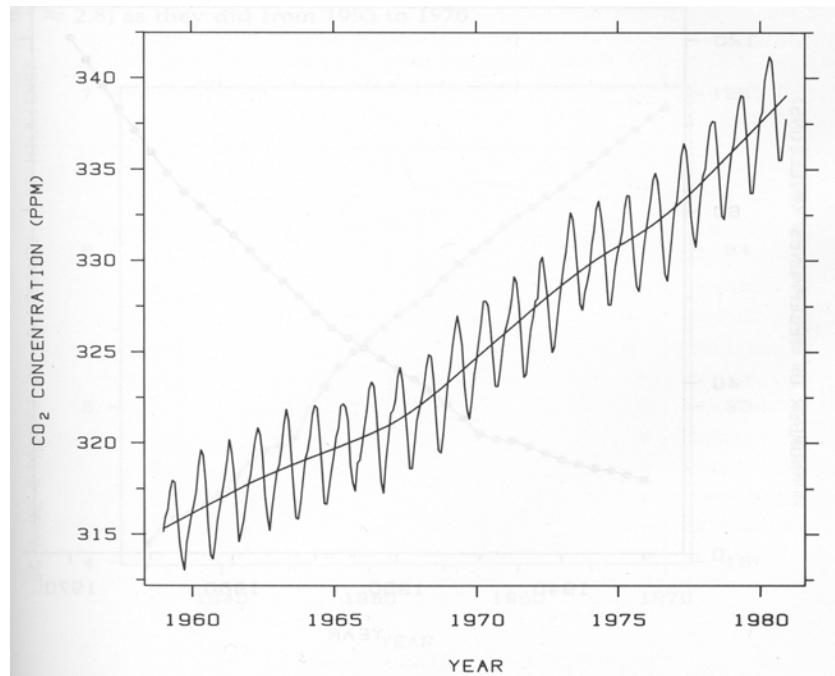
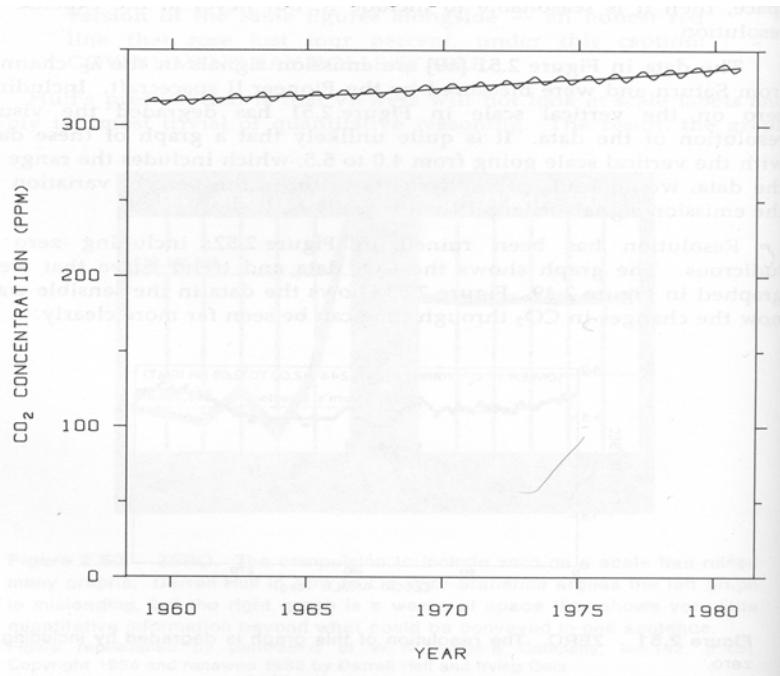
Power functions ( $y = kx^a$ ) transform into lines

Example - Steven's power laws:

$$S = kI^p \rightarrow \log S = \log k + p \log I$$



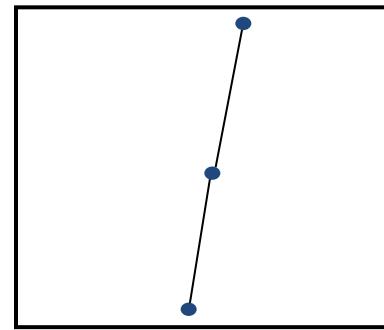
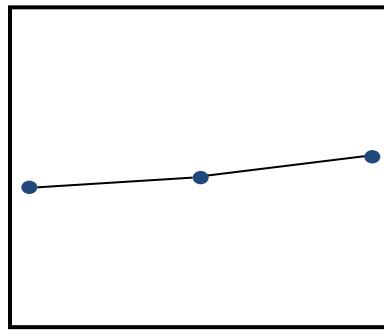
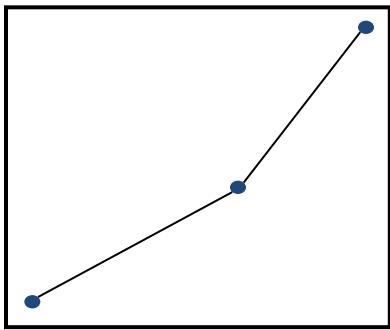
# ASPECT RATIO SELECTION



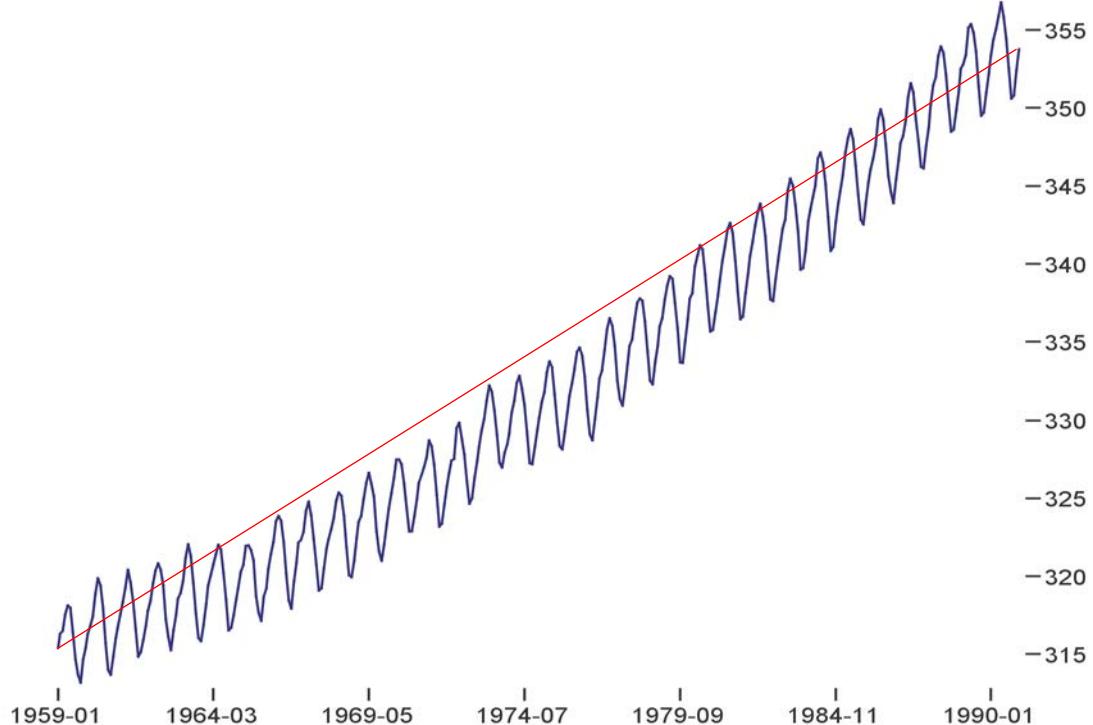
**YEARLY CO<sub>2</sub> CONCENTRATIONS [CLEVELAND 85]**

# BANKING TO $45^\circ$ [CLEVELAND]

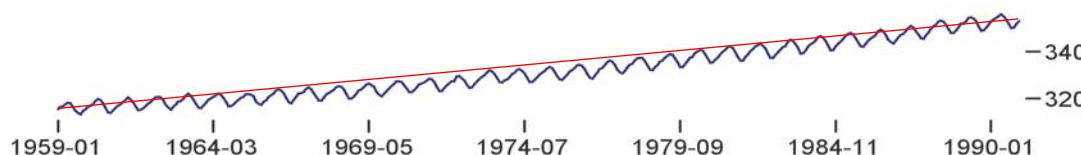
TO FACILITATE PERCEPTION OF TRENDS, MAXIMIZE THE DISCRIMINABILITY OF LINE SEGMENT ORIENTATIONS



TWO SEGMENTS ARE MAXIMALLY DISCRIMINABLE WHEN THEIR AVG ABSOLUTE ANGLE IS  $45^\circ$   
OPTIMIZE THE *ASPECT RATIO* TO BANK TO  $45^\circ$

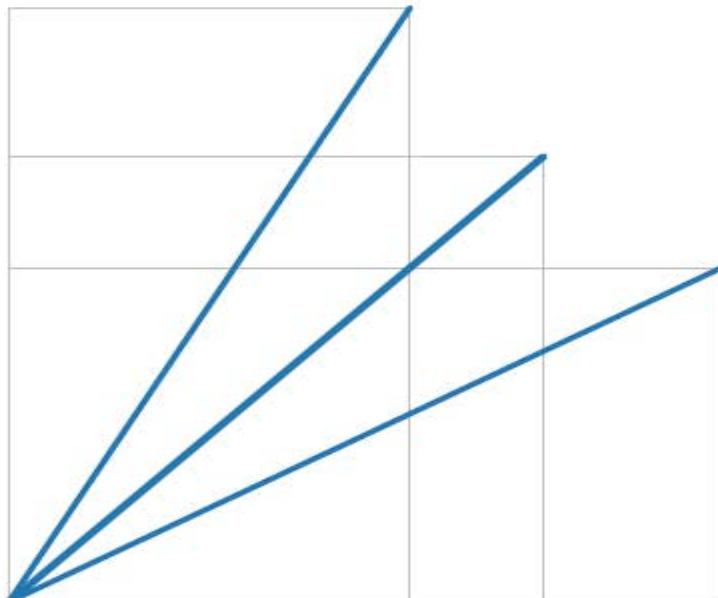


ASPECT RATIO = 1.17

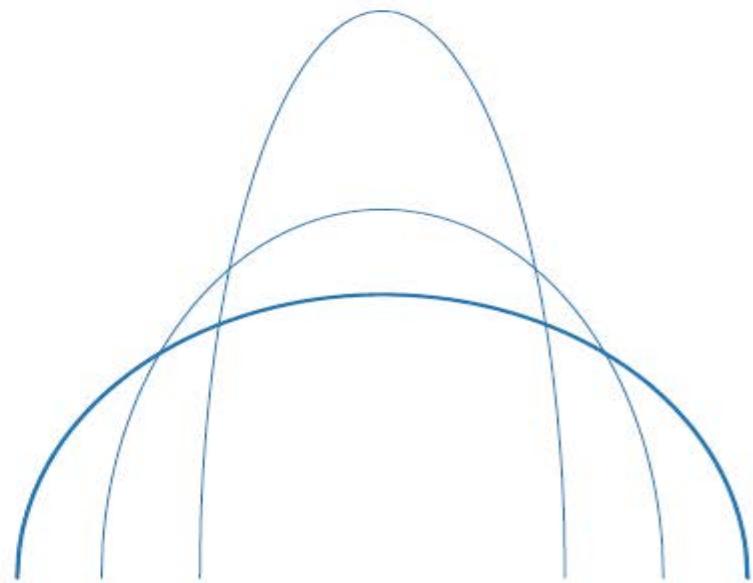


ASPECT RATIO = 7.87

# **AN ALTERNATIVE APPROACH: MINIMIZE ARC LENGTH (HOLD AREA CONSTANT)**



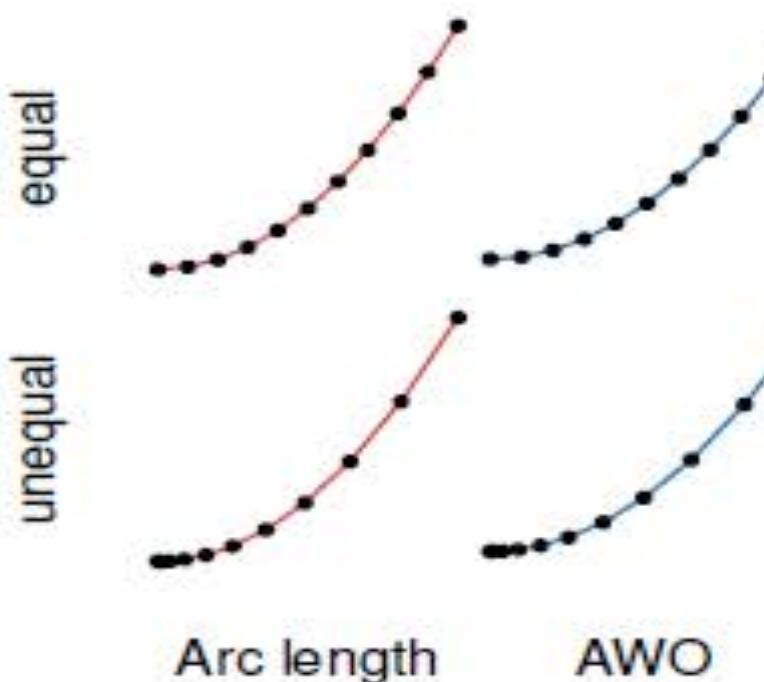
**STRAIGHT LINE -> 45 DEG**



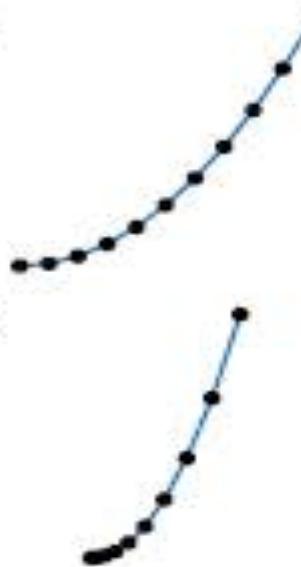
**ELLIPSE -> CIRCLE**

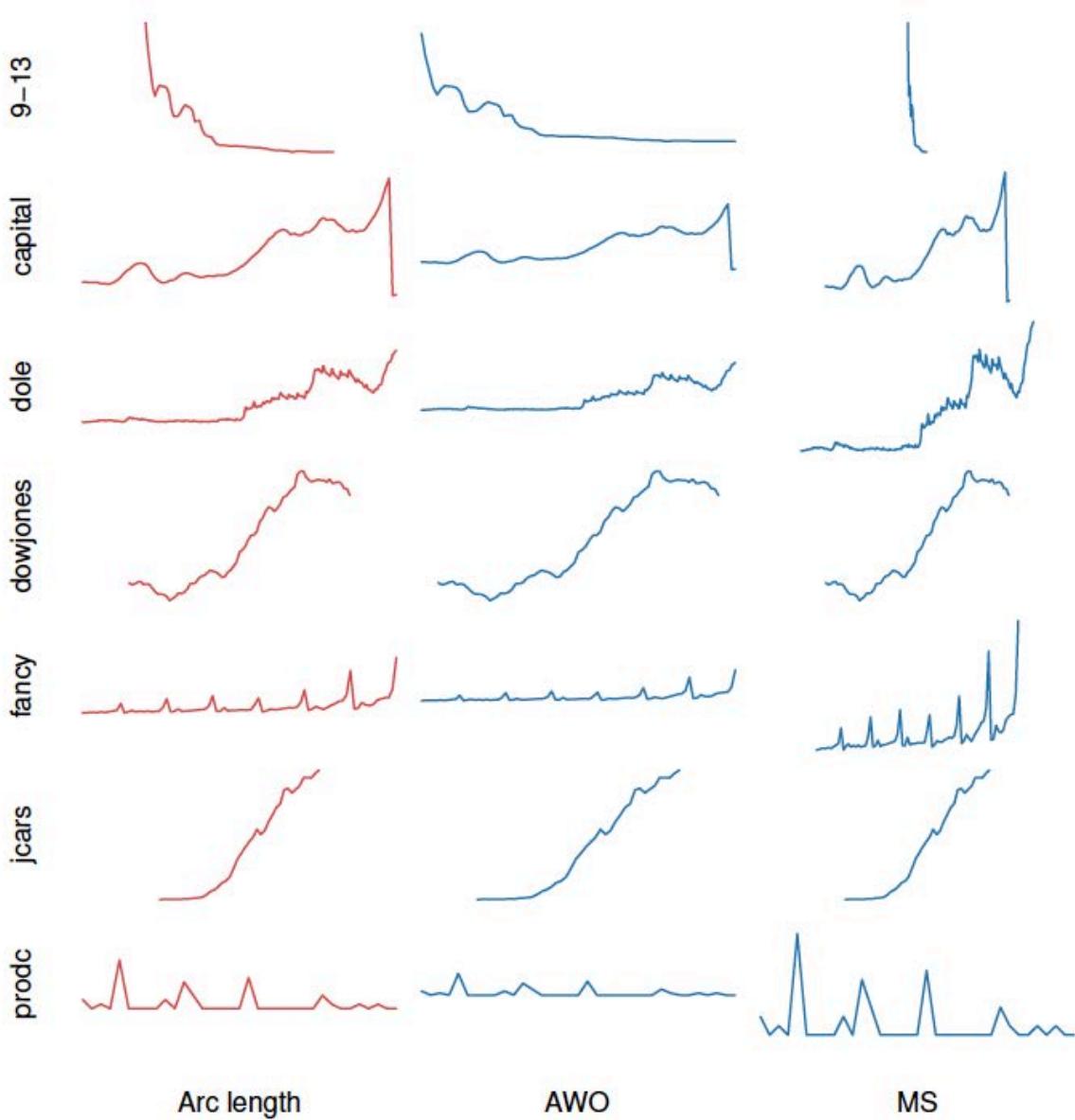
# PARAMETERIZATION INVARIANCE

WEIGHTED BY LENGTH



NOT WEIGHTED BY LENGTH





ARC-LENGTH BANKING  
PRODUCES ASPECT RATIOS IN-  
BETWEEN THOSE PRODUCED BY  
THE OTHER METHODS.

# TAKEAWAY

CHOOSE A **SCALE**, **ORIGIN**, & **ASPECT RATIO** THAT EMPHASIZES THE IMPORTANT DETAILS FOR YOUR TASK.

DESIGN THE VISUALIZATION TO MAKE THOSE CHOICES **CLEAR AND UNAMBIGUOUS** TO A VIEWER.

**HOW DO YOUR CHARTS CHECK OUT?**

# CAVEATS

*A collection of dataviz caveats by [data-to-viz.com](https://data-to-viz.com)*

Show all

Top 10

Improvement

Misleading

Map

Bar



Order your data

When displaying the value of several entities, ordering them makes the graph much more insightful.



To cut or not to cut?

Cutting the Y-axis is one of the most controversial practice in data viz. See why.



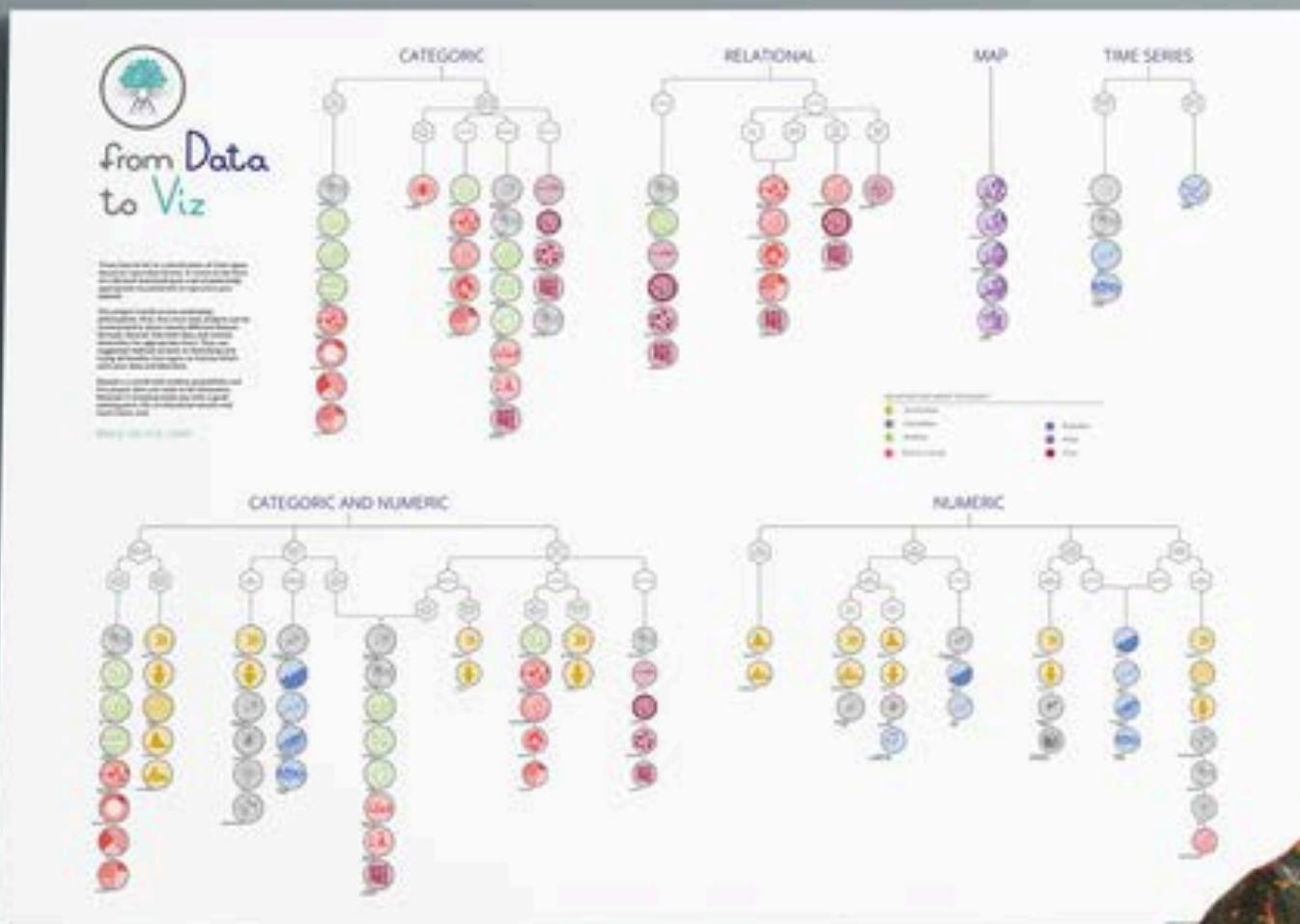
The spaghetti chart

A line graph with too many lines becomes unreadable: it is called a spaghetti graph.



Pie chart

The human eye is bad at reading angles. See how to replace the most criticized chart ever.

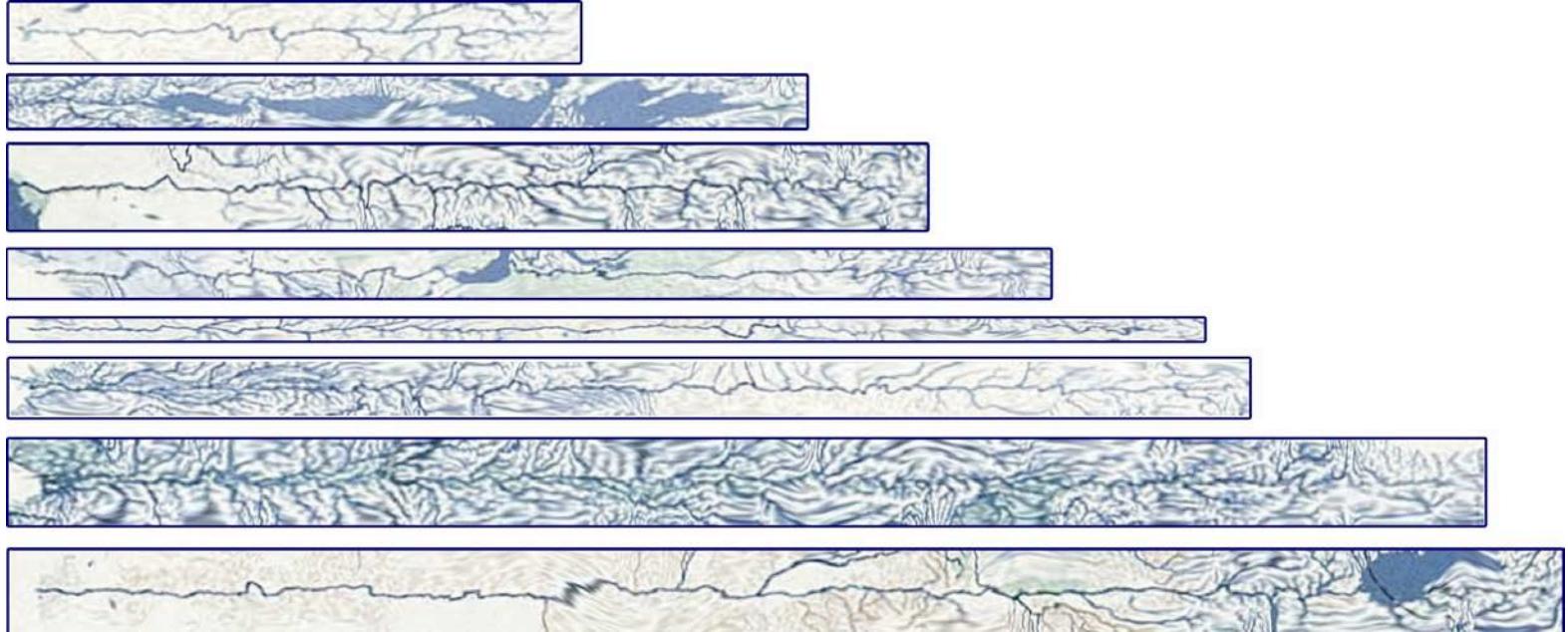


# DISTORTION

## PHYSICAL GEOGRAPHY.



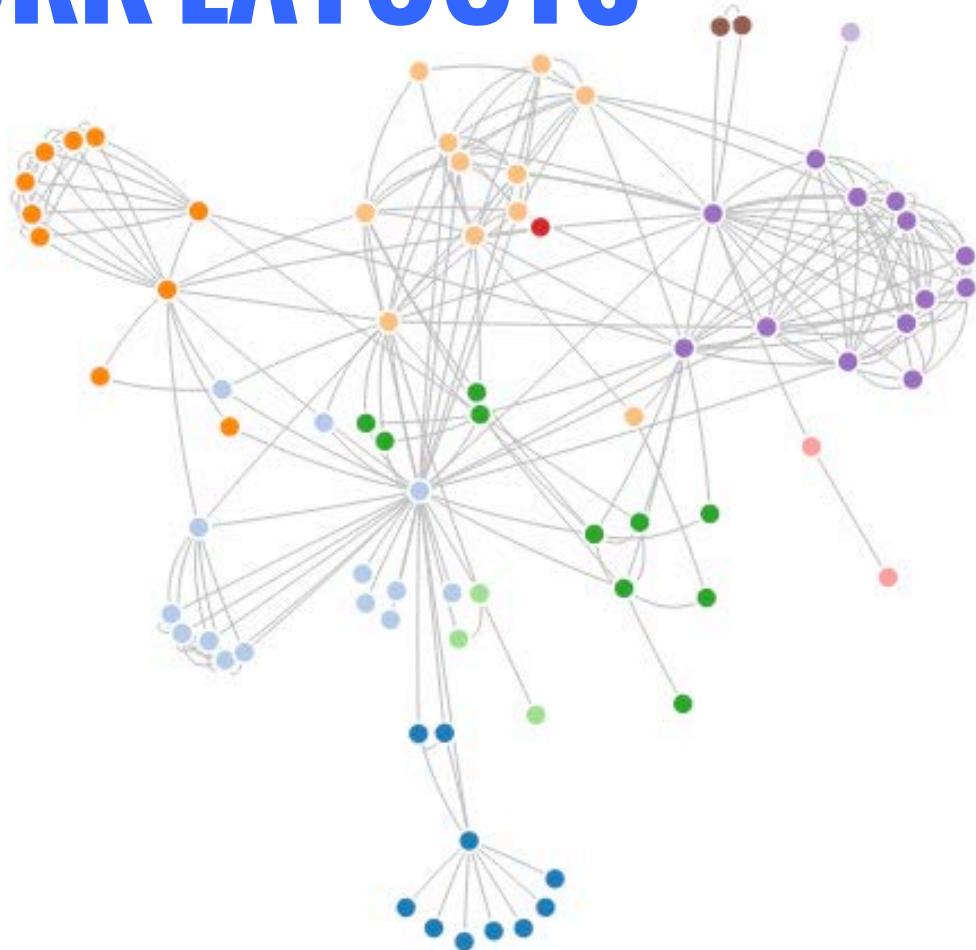




[TRANSMOGRIFICATION] BROSZ ET AL. 2013]

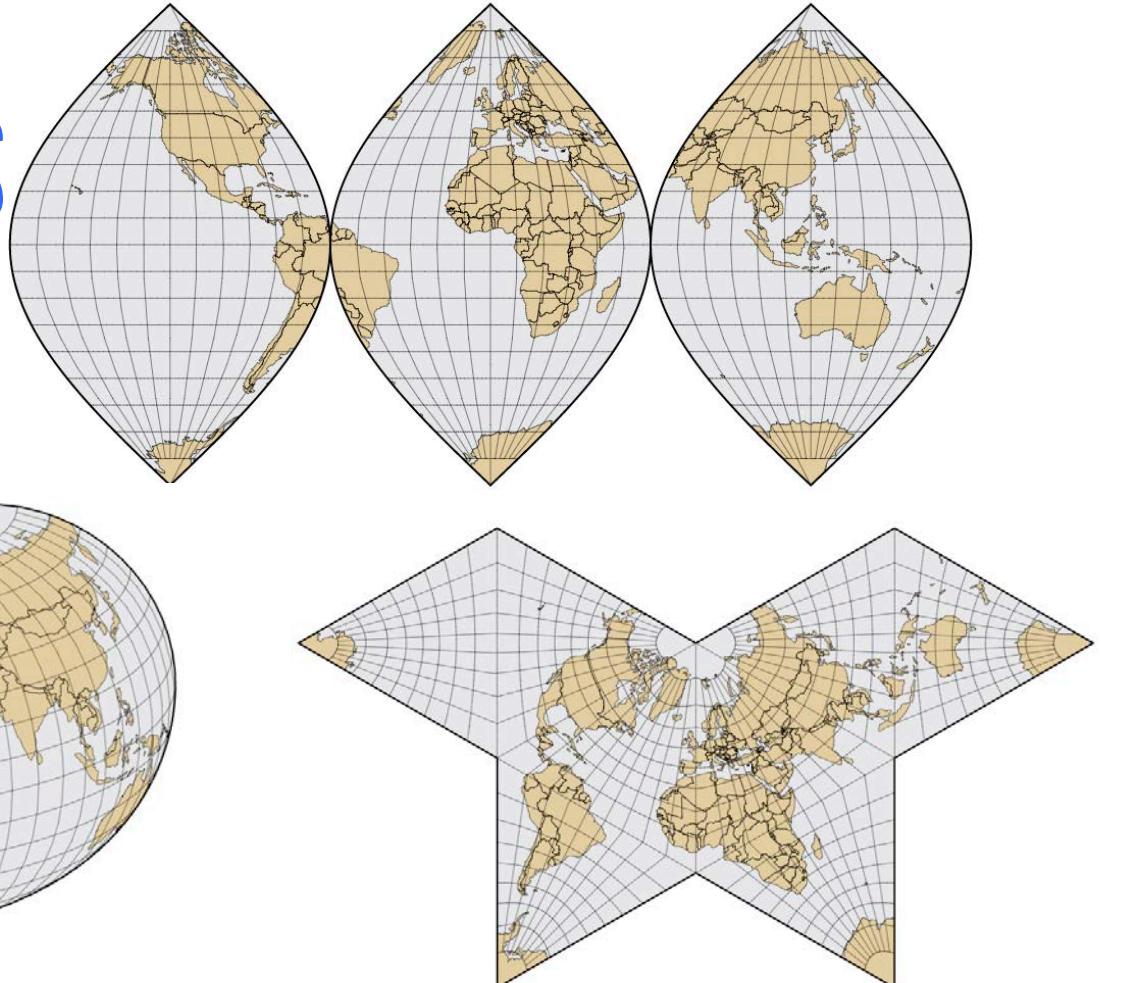
[WWW.TRANSMOGRIFIERS.ORG](http://WWW.TRANSMOGRIFIERS.ORG)

# NETWORK LAYOUTS



MIKE BOSTOCK

# MAP PROJECTIONS



JASON DAVIES

# CARTOGRAMS

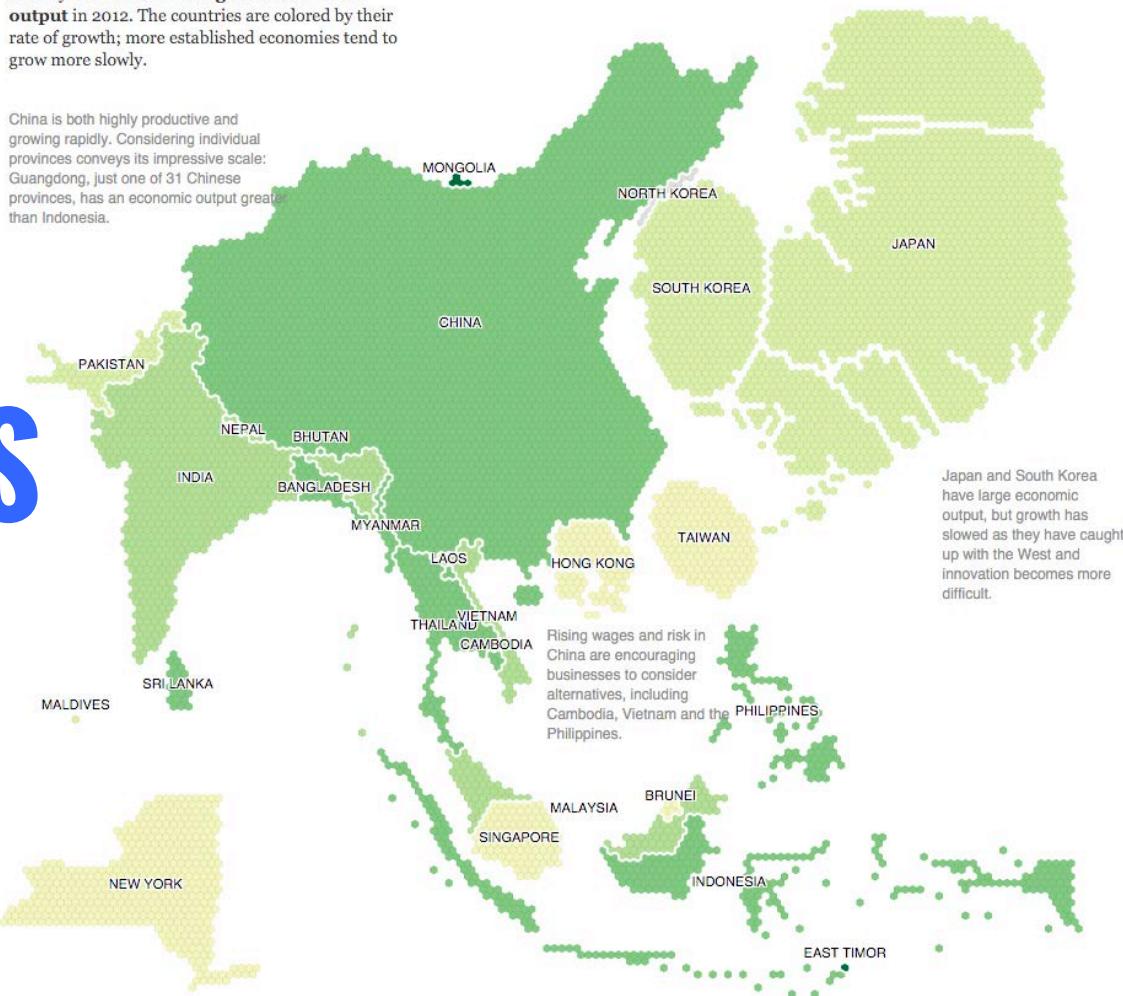
## Economic Output

In this map, geography is distorted so that each country is sized according to its economic output in 2012. The countries are colored by their rate of growth; more established economies tend to grow more slowly.

China is both highly productive and growing rapidly. Considering individual provinces conveys its impressive scale: Guangdong, just one of 31 Chinese provinces, has an economic output greater than Indonesia.

Each hexagon represents \$2.7 billion in G.D.P.

G.D.P. growth, 2011 to 2012  
2 4 6 8 +10%



MIKE BOSTOCK

**FOCUS + CONTEXT**

Mar. 29, 1976

# THE NEW YORKER

Price 75 cents

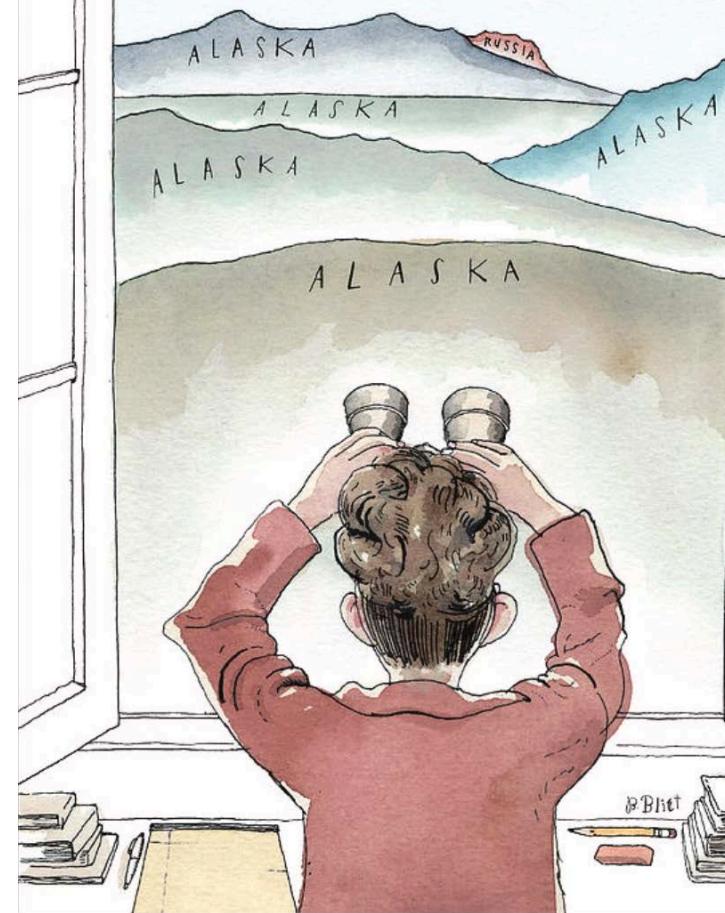


PRICE \$4.50

# THE

OCT. 6, 2008

# THE NEW YORKER



# TABLELENS

[RAO & CARD 94]

Baseball.txt - TLDemo

File Edit View Options Help

inxight

League ...	Players	At Bats	Hits	Home Runs	Runs	Rbi
N	Andres ...	321	87	10	39	42
	Jose Cruz	479	133	10	48	72
	Bo Diaz	474	129	10	50	56
	Tony Pena	510	147	10	56	52
A	Reggie J...	419	101	18	65	58

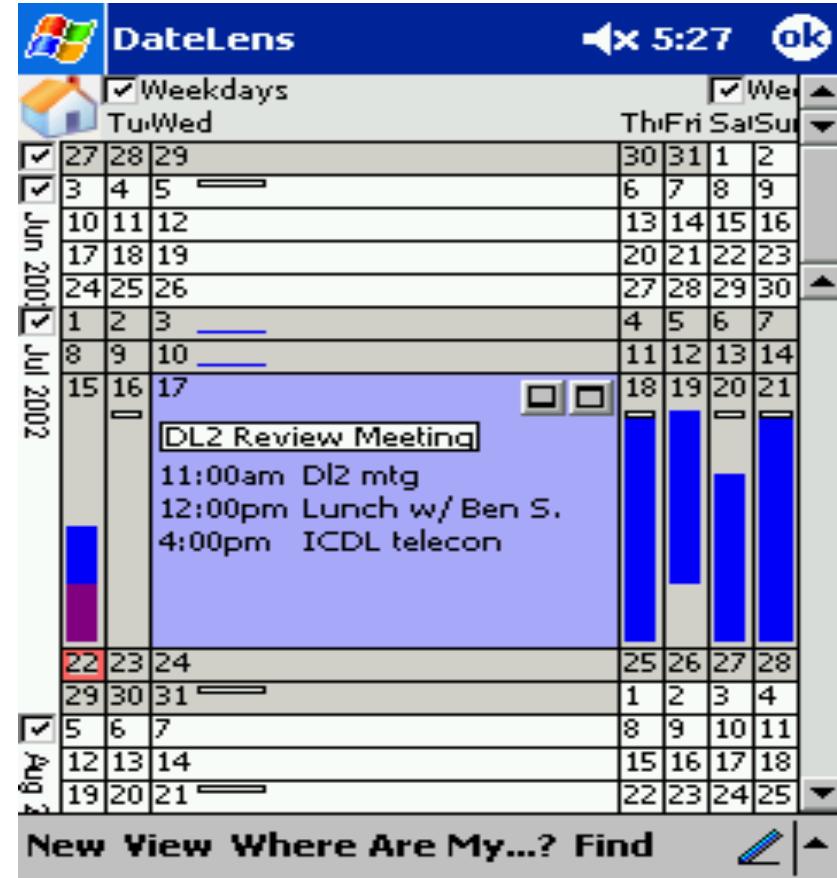
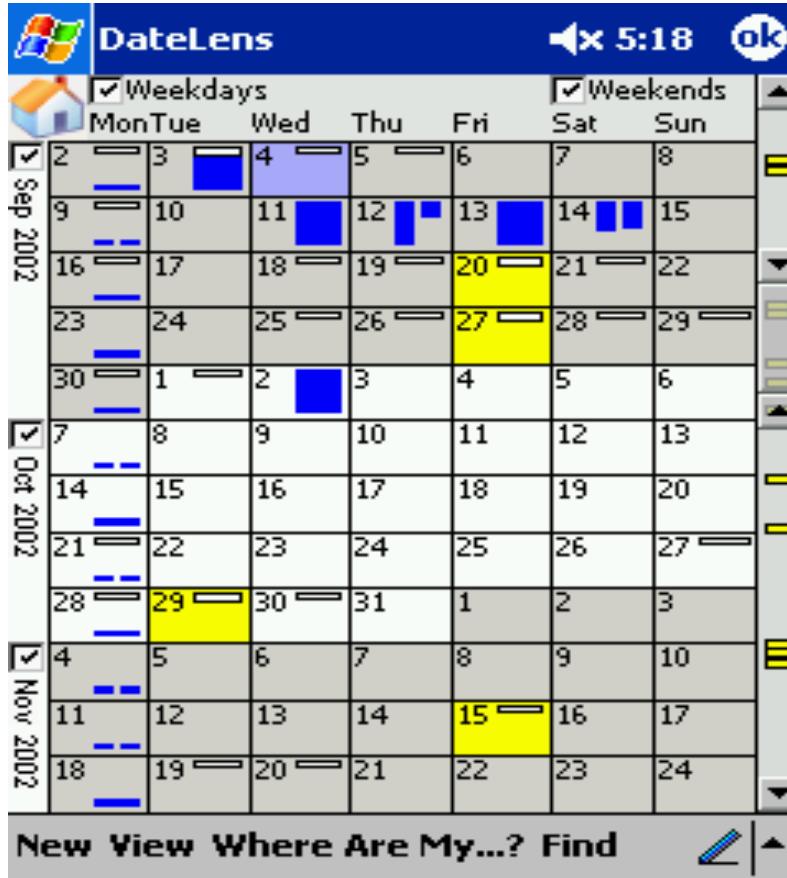
Row 79: 35

Col: Assists

Entry: 35

The screenshot shows a vintage-style Windows application window titled "Baseball.txt - TLDemo". The menu bar includes "File", "Edit", "View", "Options", and "Help". The toolbar features various icons for file operations like Open, Save, Print, and search. The "inxight" logo is in the top right corner. The main area displays a table of baseball statistics. The columns are labeled "League ...", "Players", "At Bats", "Hits", "Home Runs", "Runs", and "Rbi". The rows are categorized by league: "N" (National League) and "A" (American League). Under the "N" category, there are four entries: Andres (321 AB, 87 H, 10 HR, 39 R, 42 Rbi), Jose Cruz (479 AB, 133 H, 10 HR, 48 R, 72 Rbi), Bo Diaz (474 AB, 129 H, 10 HR, 50 R, 56 Rbi), and Tony Pena (510 AB, 147 H, 10 HR, 56 R, 52 Rbi). Under the "A" category, there is one entry: Reggie J... (419 AB, 101 H, 18 HR, 65 R, 58 Rbi). A red box highlights the data for Reggie Jackson in the "A" row. The bottom status bar shows "Row 79: 35", "Col: Assists", and "Entry: 35".

# DATELENS



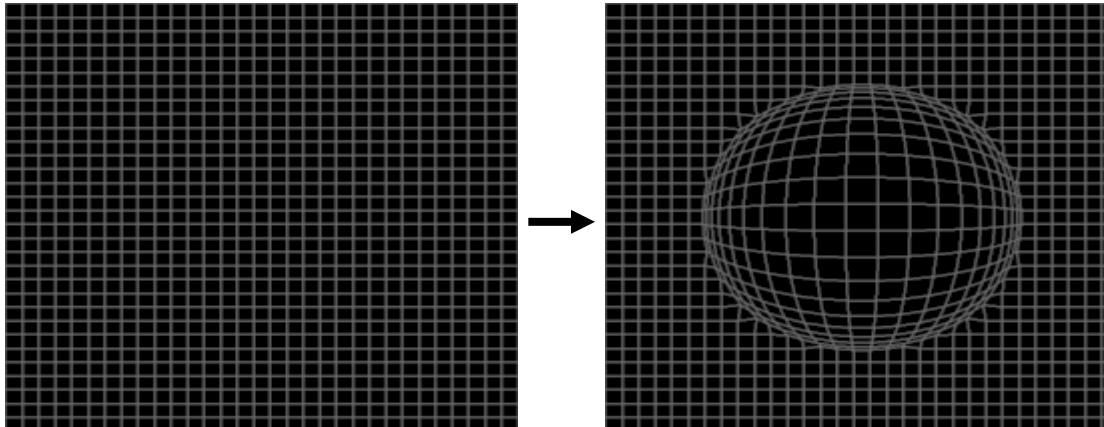
[BEDERSON ET AL. 04]

# SINGLE VIEW FOCUS + CONTEXT

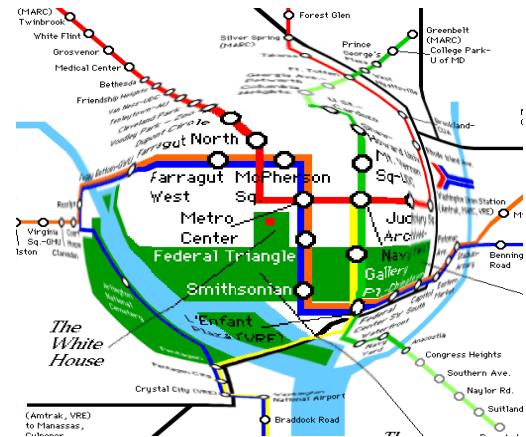
FOCUS AREA - LOCAL DETAILS

DE-MAGNIFIED AREA - SURROUNDING CONTEXT

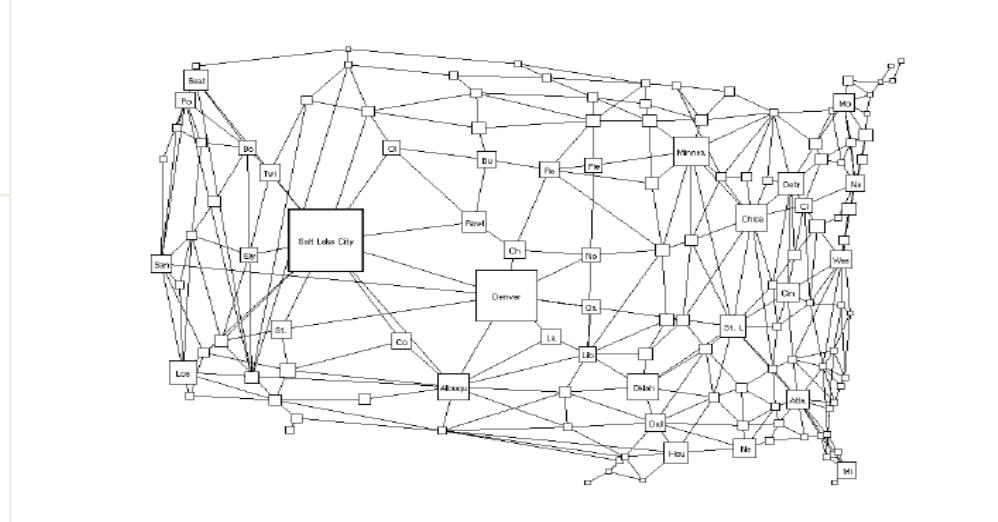
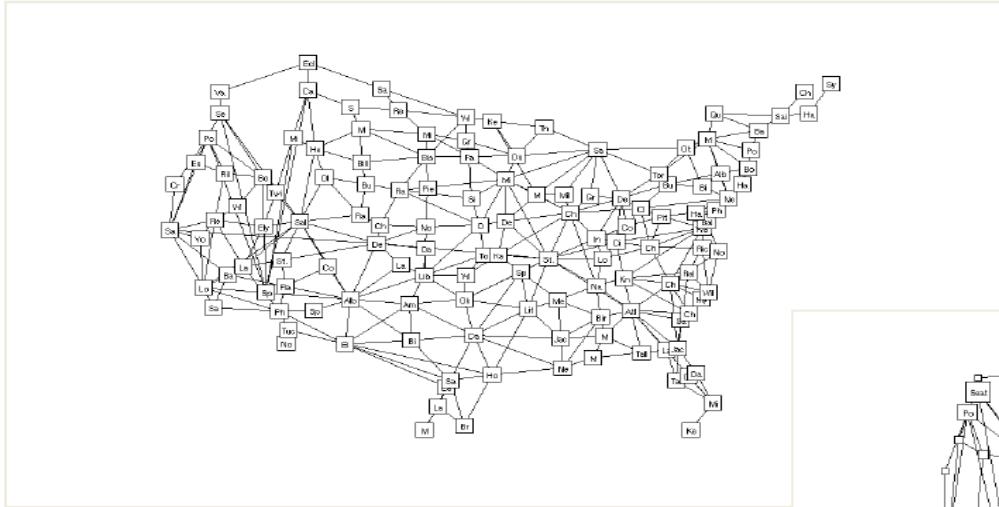
LIKE A RUBBER SHEET WITH BORDERS TACKED DOWN



Nonlinear Magnification Infocenter [<http://www.cs.indiana.edu/~tkeahley/research/nlm/nlm.html>]



# STATES AS NODES IN A GRAPH



GRAPHICAL FISHEYE VIEWS OF GRAPHS [SARKAR & BROWN 92]

# **USES (AND ABUSES) OF DISTORTION**

**OFTEN MORE HARMFUL THAN HELPFUL** – ESPECIALLY WHEN SPATIAL DIMENSIONS ALREADY HAVE FIXED SCALE AND MEANING.

**CAN ADD VALUE IF IT:**

- BUILDS ON EXPERIENCE AND ENABLES A SPECIFIC TASK
- INTENDED TO ELICIT RESPONSE, CAPTURE ATTENTION

**CONSIDER FOCUS+CONTEXT OF DATA RATHER THAN VIEW**

# THOUGHTS?