

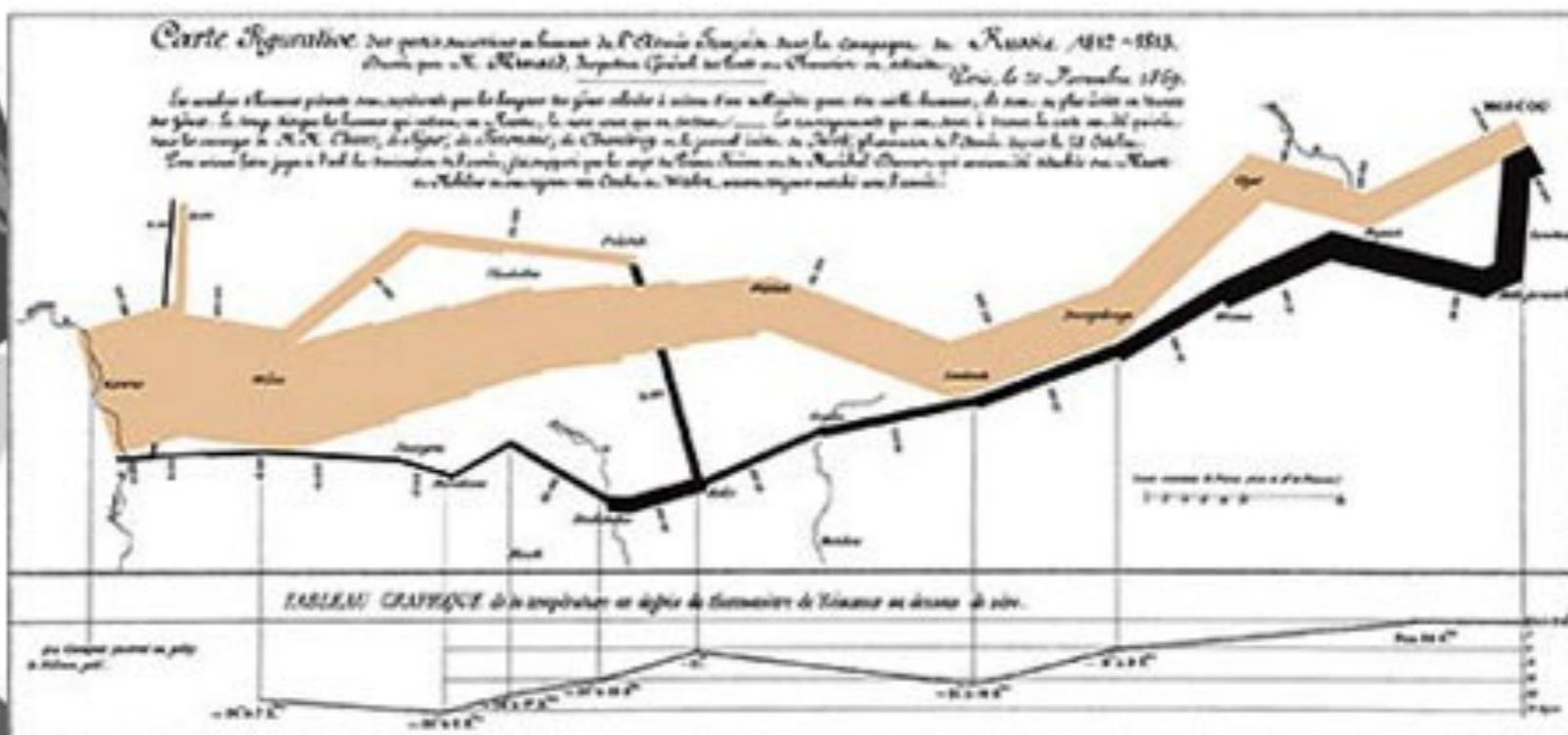
# Thinking Graphically

## Lessons From Edward Tufte

Slides adapted from presentation by Mika Aldaba, Graphic Designer

# Who Is Edward Tufte?

an analytical design theorist, educator, and landscape sculptor best known for his self-published books on analytical design



Illustrated by Merchant  
for the Brunswick Review

# Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.7	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.8	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

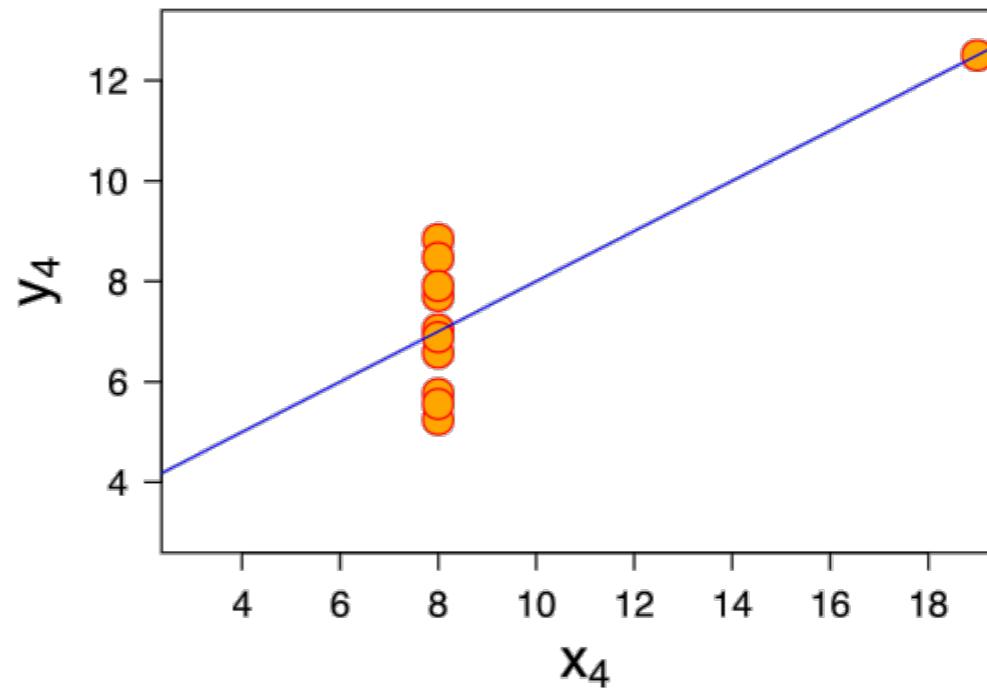
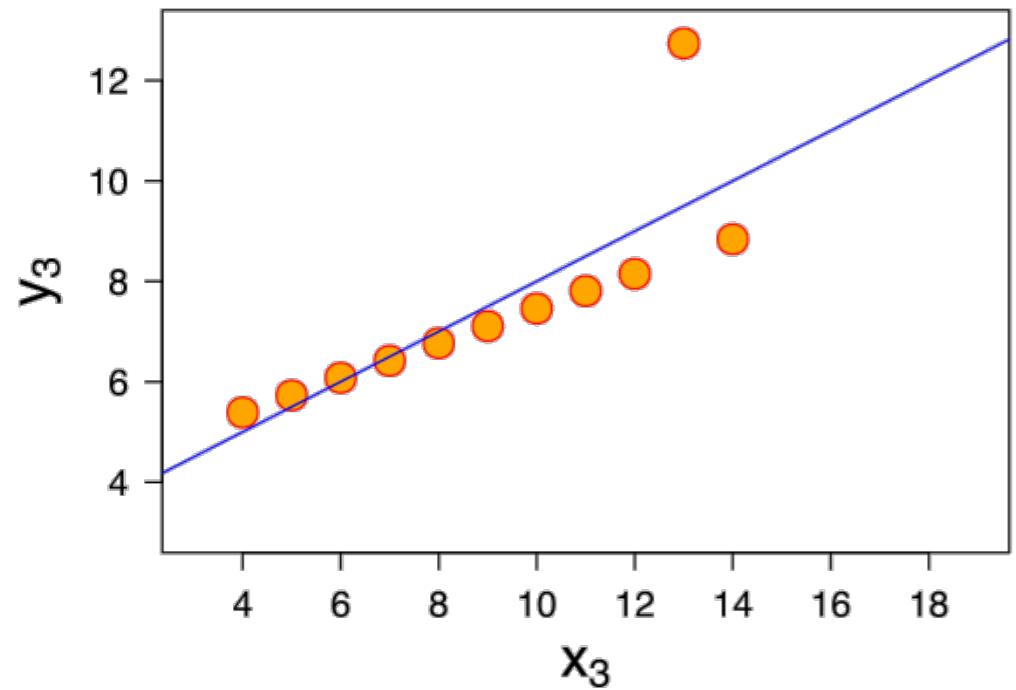
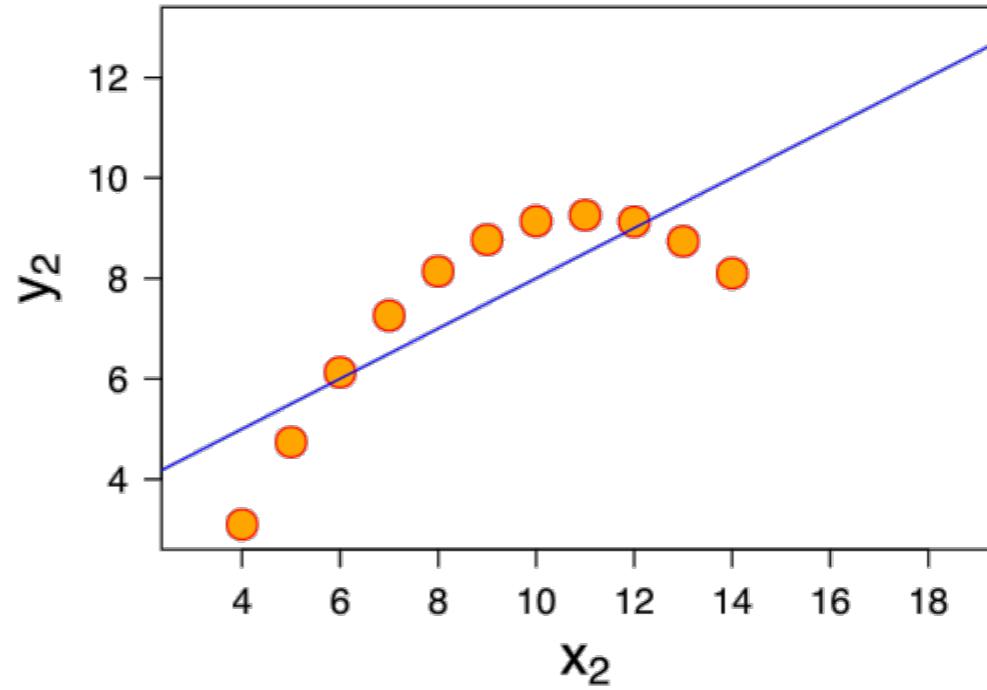
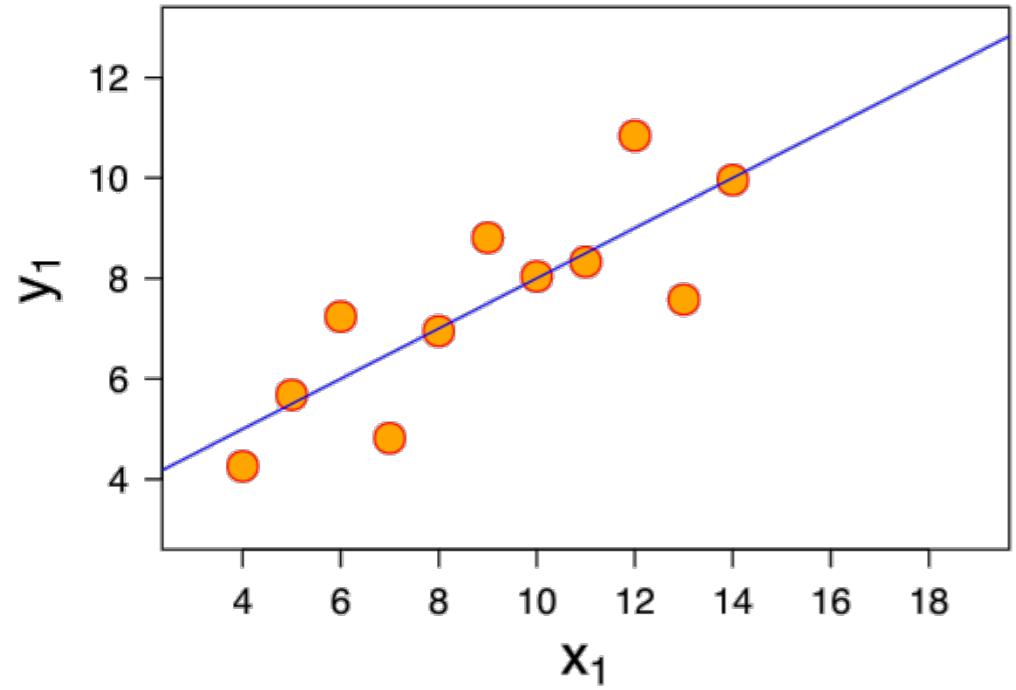


How would you describe the differences in these datasets?

# Anscombe's Quartet

Metric	Value
mean of the y values	7.5
equation of the least-squared regression line	$y = 3 + 0.5x$
sums of squared errors (about the mean)	110
regression sums of squared errors (variance accounted for by x)	27.5
residual sums of squared errors (about the regression line)	13.75
correlation coefficient	0.82
coefficient of determination	0.67

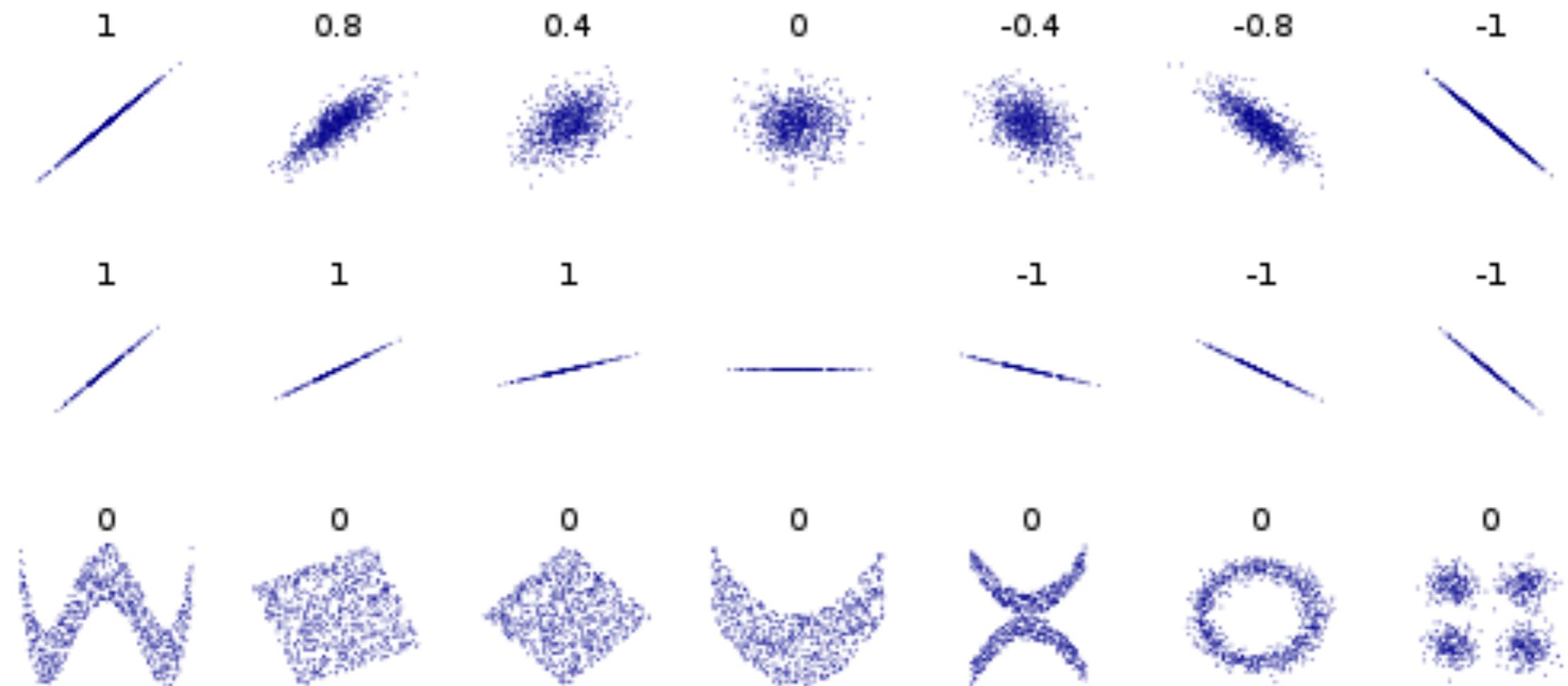
# Anscombe's Quartet



Graphics  
reveal  
data

# Another example: Pearson Correlation

---



# Other reasons?

---

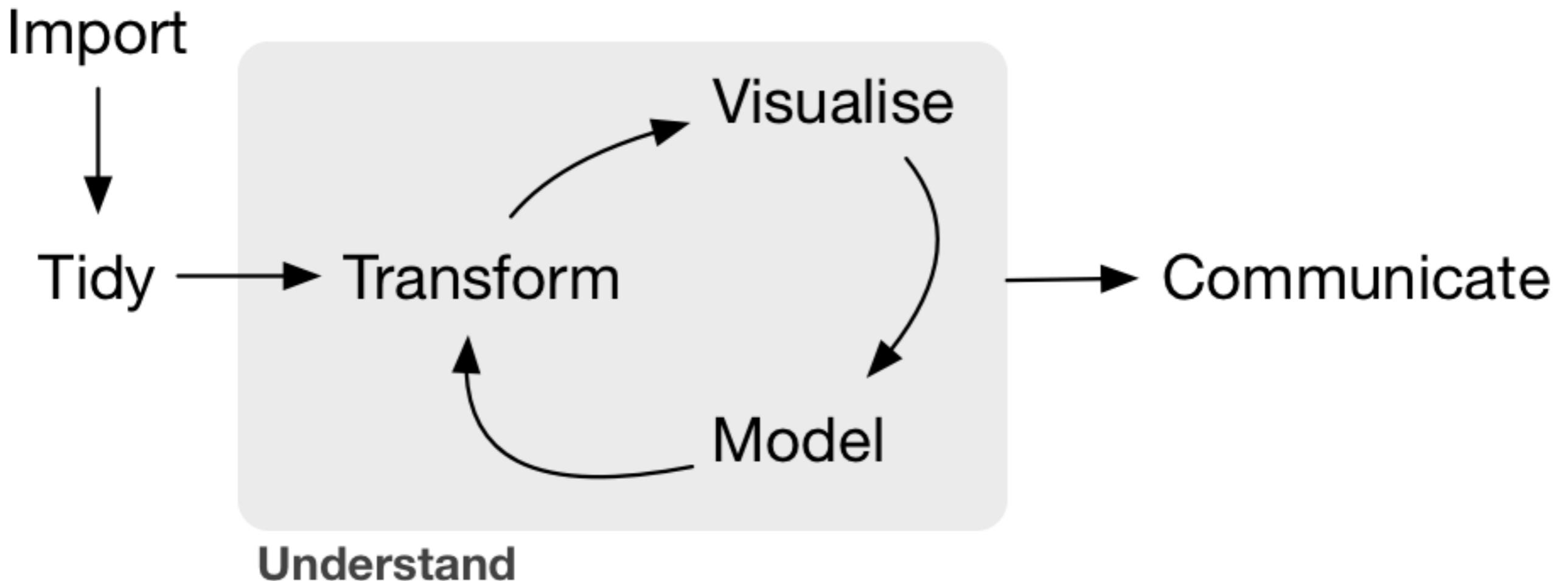
Visualization is the highest bandwidth channel into the human brain

As data volumes grow, visualization becomes a necessity rather than a luxury.

- “A picture is worth a thousand words”

# Visualization is important...

---



# Graphical Excellence

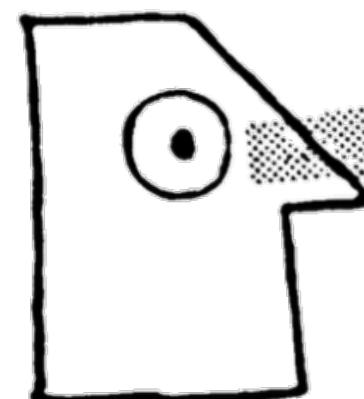
---

1. Show the data
2. Induce the viewer to think about substance rather than methodology
3. Avoid distortion
4. Present many numbers in a small space
5. Make large data sets coherent
6. Encourage comparisons
7. Reveal the data at several levels from broad overview to fine structure
8. Serve a purpose
9. Be integrated with statistical and verbal descriptions of data

-- Principles of *Graphical Excellence*, Edward Tufte

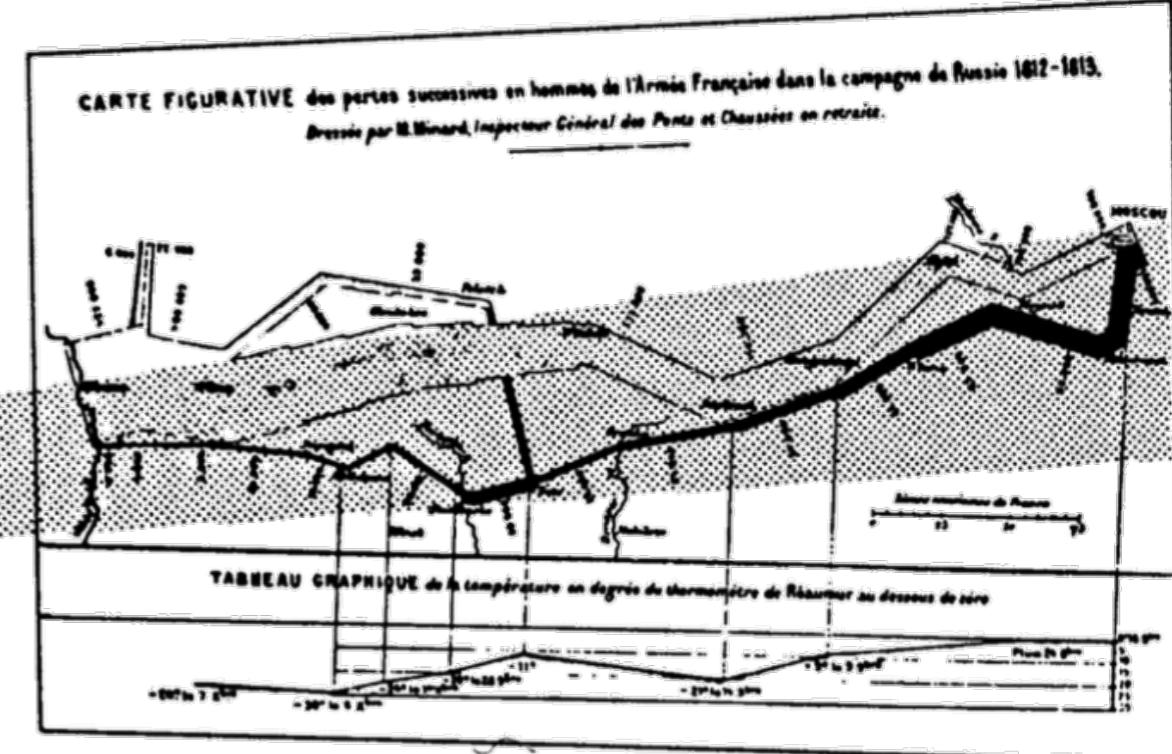
**ideas**

**time**



**ink**

**space**



Graphical excellence gives the viewer:

- the greatest # of ideas
- in the shortest time
- with the least ink
- in the smallest space.

# One of the oldest graphical displays

---



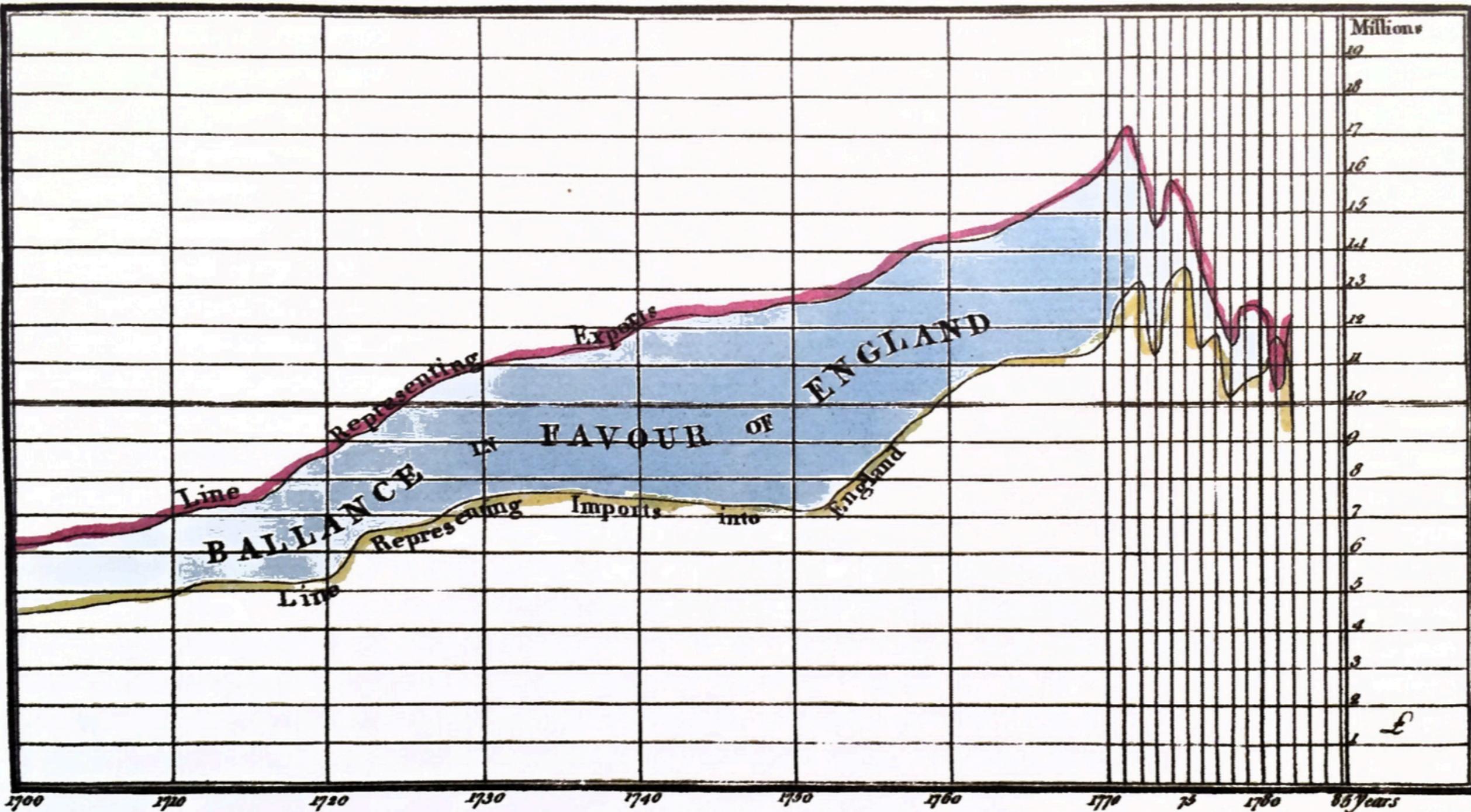
Yu Chi Thu

*Map of the tracks of Yu the Great*

Carved in stone in 1137 A.D.  
(idea probably older)

Nothing similar in Europe until  
~1550

*CHART of all the IMPORTS and EXPORTS to and from ENGLAND  
From the Year 1700 to 1782 by W. Playfair*



*The Divisions at the Bottom, express YEARS, & those on the Right hand MILLIONS of POUNDS*

*J. Andrie Sculp:*

*Published as the Act directs. 20.<sup>th</sup> Aug. 1785*

One of the earliest timeseries charts

# Life-cycle of the Japanese Beetle

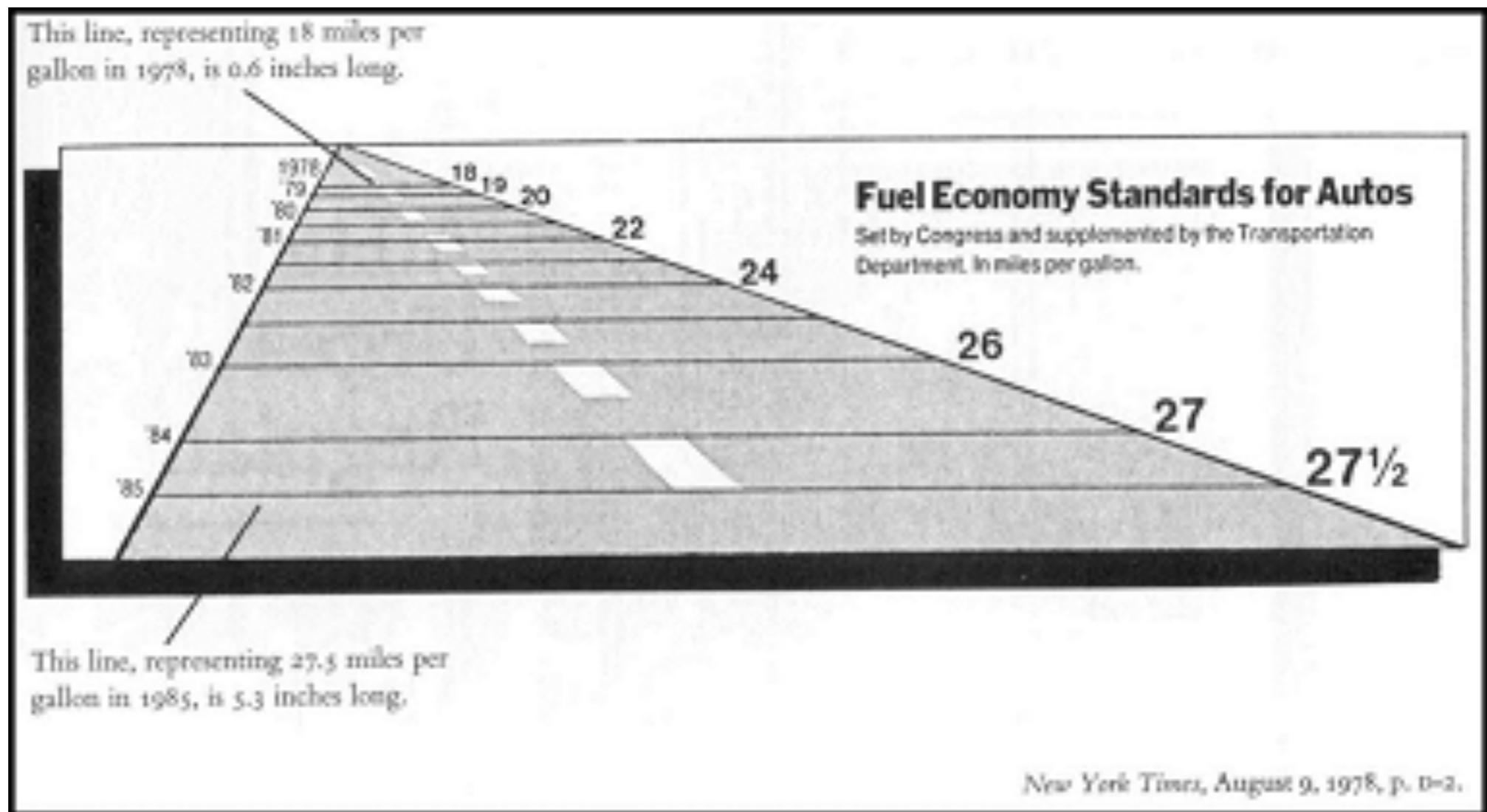


Strive For  
Graphical  
Integrity

Visual  
representations of  
data must tell the  
truth.

# The Lie Factor

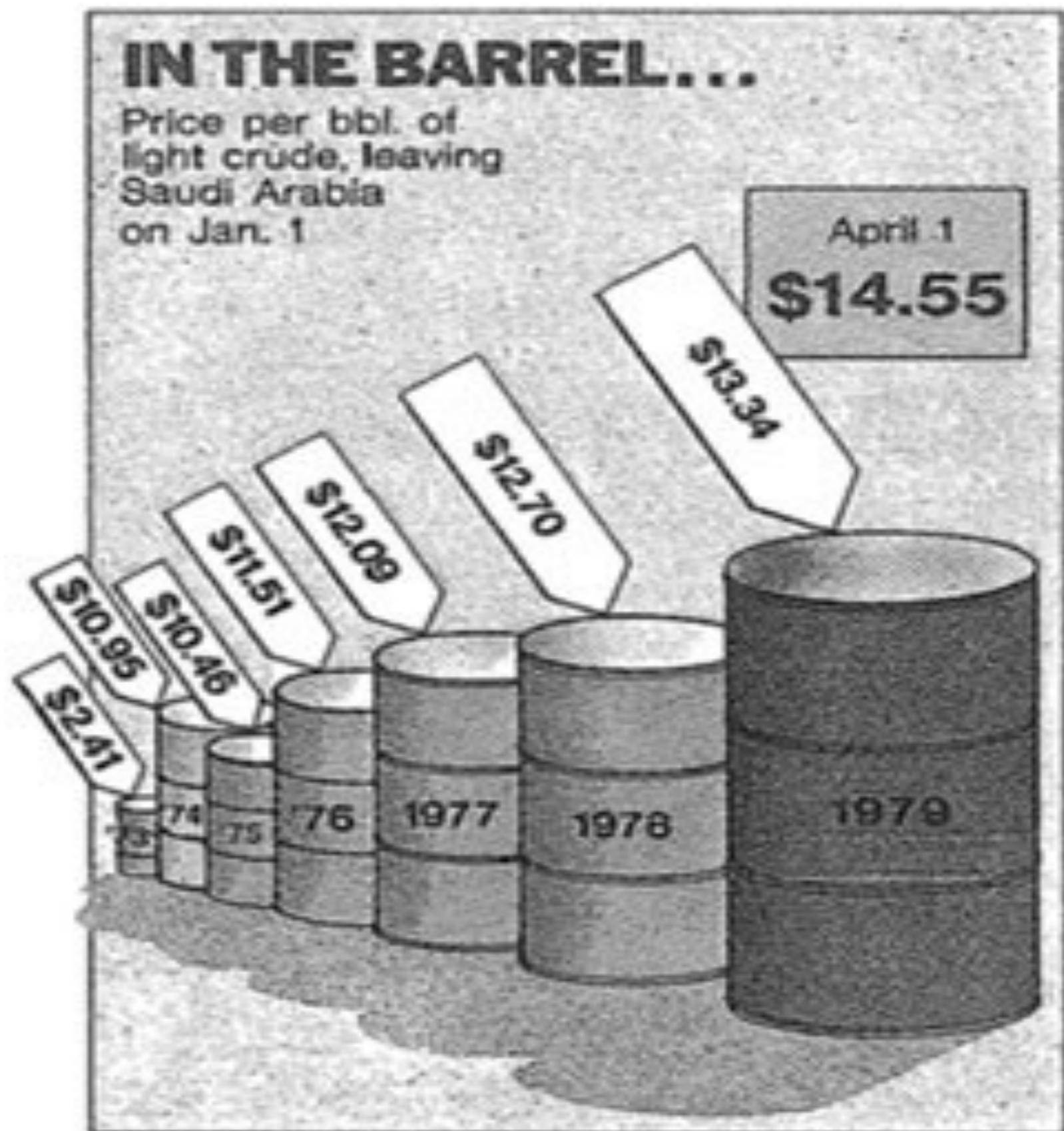
$$\text{Lie Factor} = \frac{\text{Size of effect in the visualization}}{\text{Size of effect in the data}}$$



According to Tufte the Lie Factor of this graph is 14.8. A numerical change of 53% is represented by a graphical change (size of horizontal lines) of 783%.

# Graphical Integrity

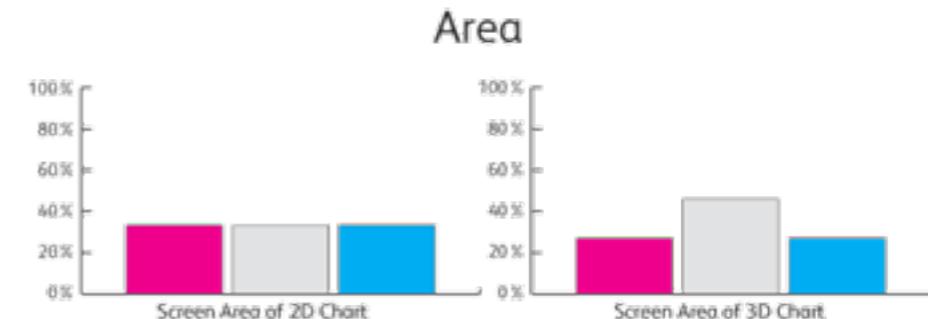
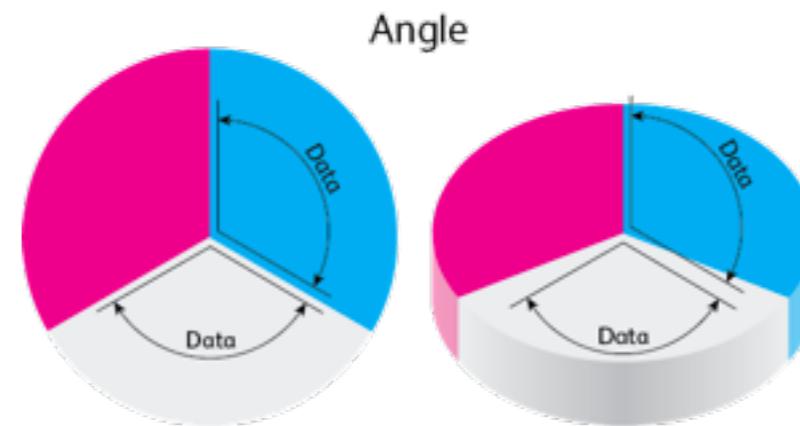
The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented



# Graphical Integrity

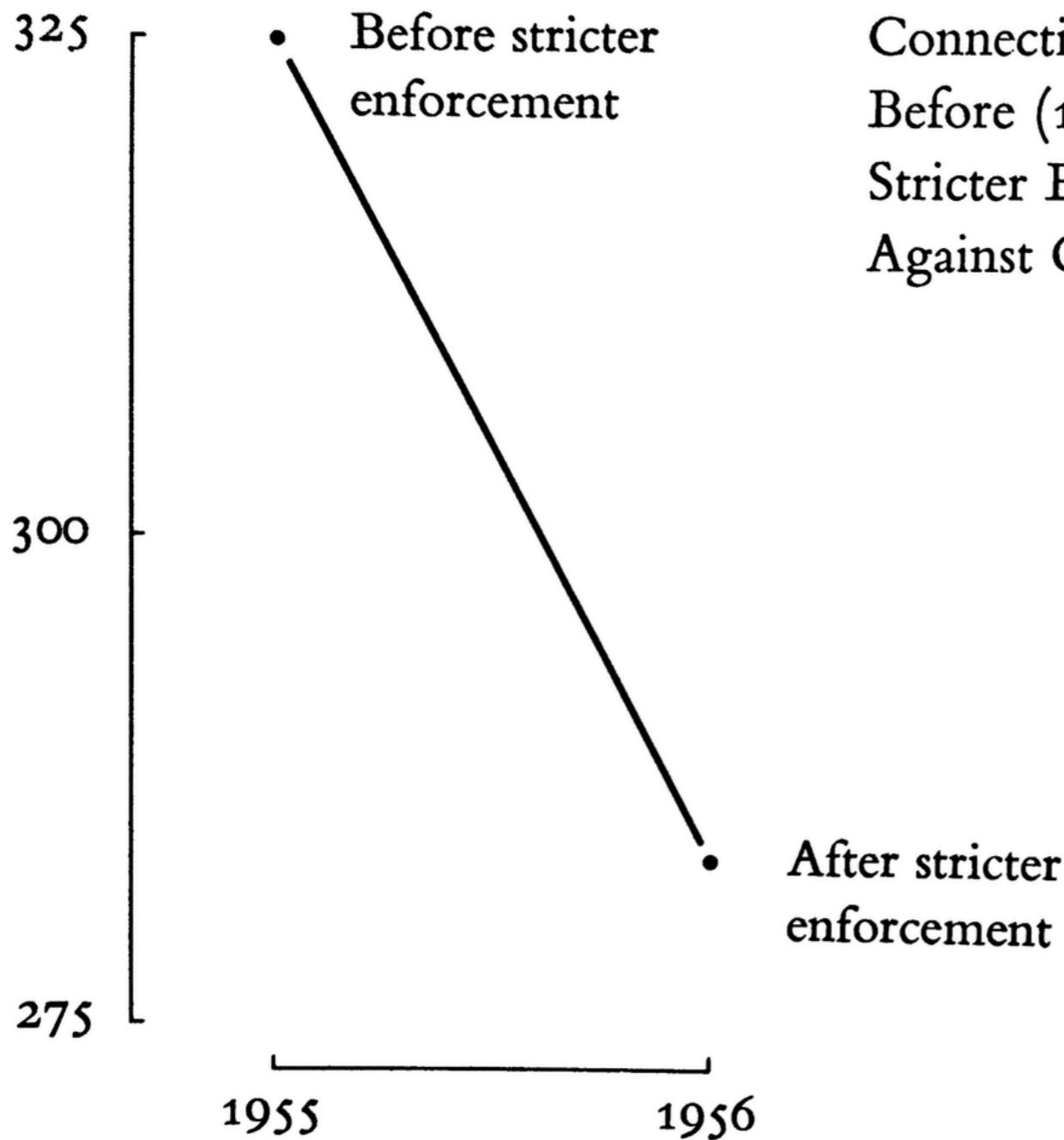
---

The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.



# Graphical Integrity

---

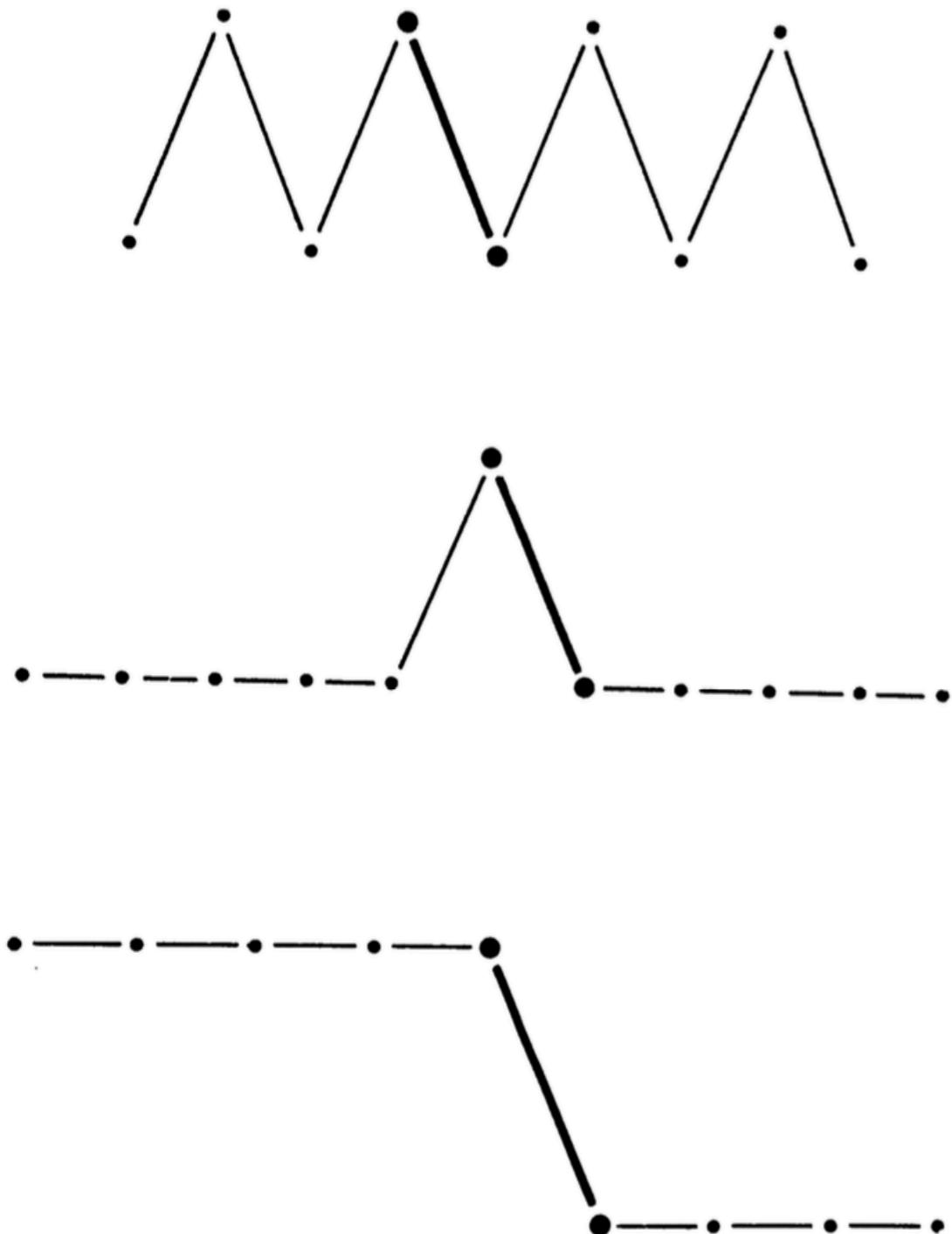


Connecticut Traffic Deaths,  
Before (1955) and After (1956)  
Stricter Enforcement by the Police  
Against Cars Exceeding Speed limit

Graphics must not  
quote data out of  
context.

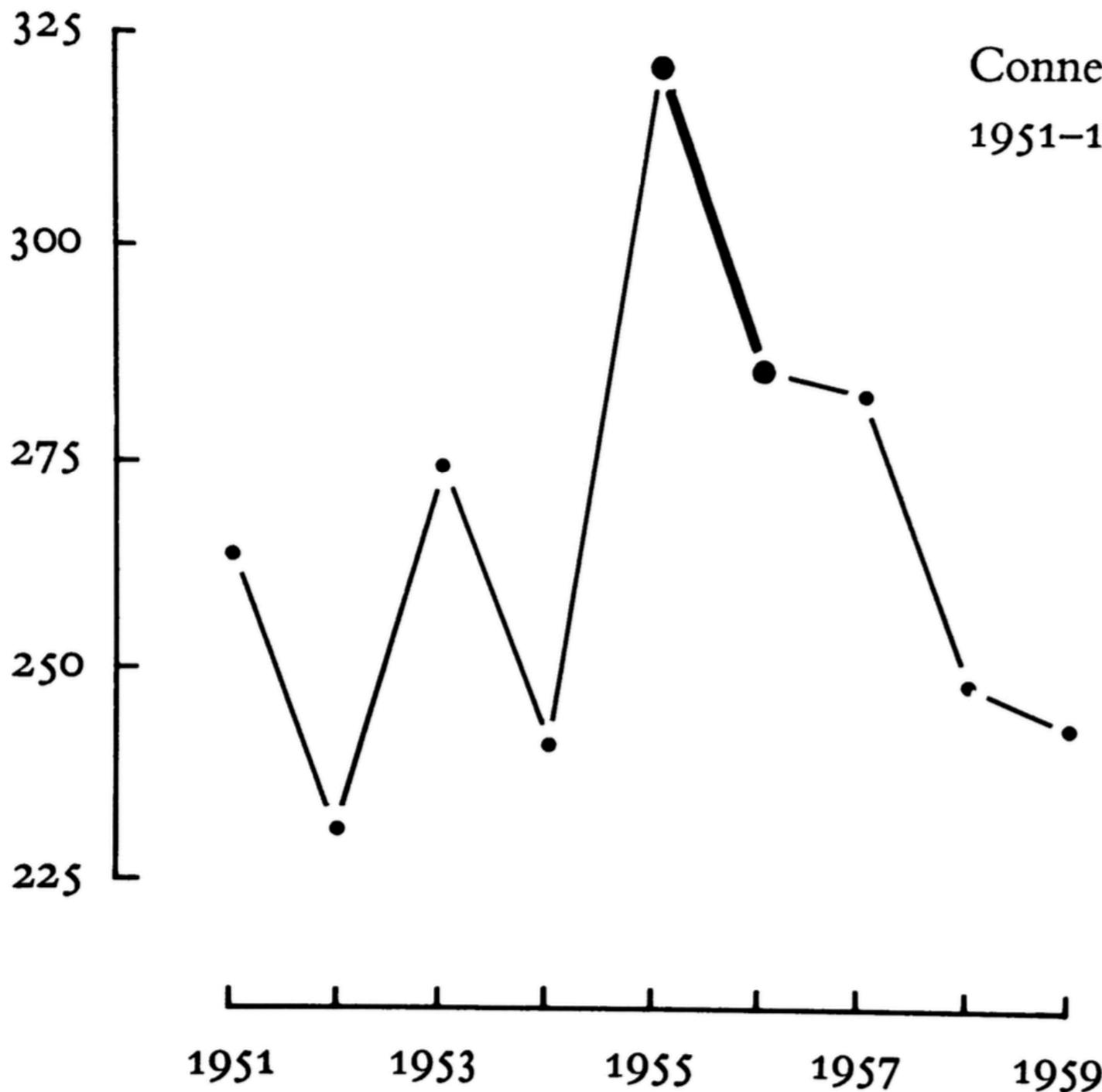
# Graphical Integrity

---



Graphics must not  
quote data out of  
context.

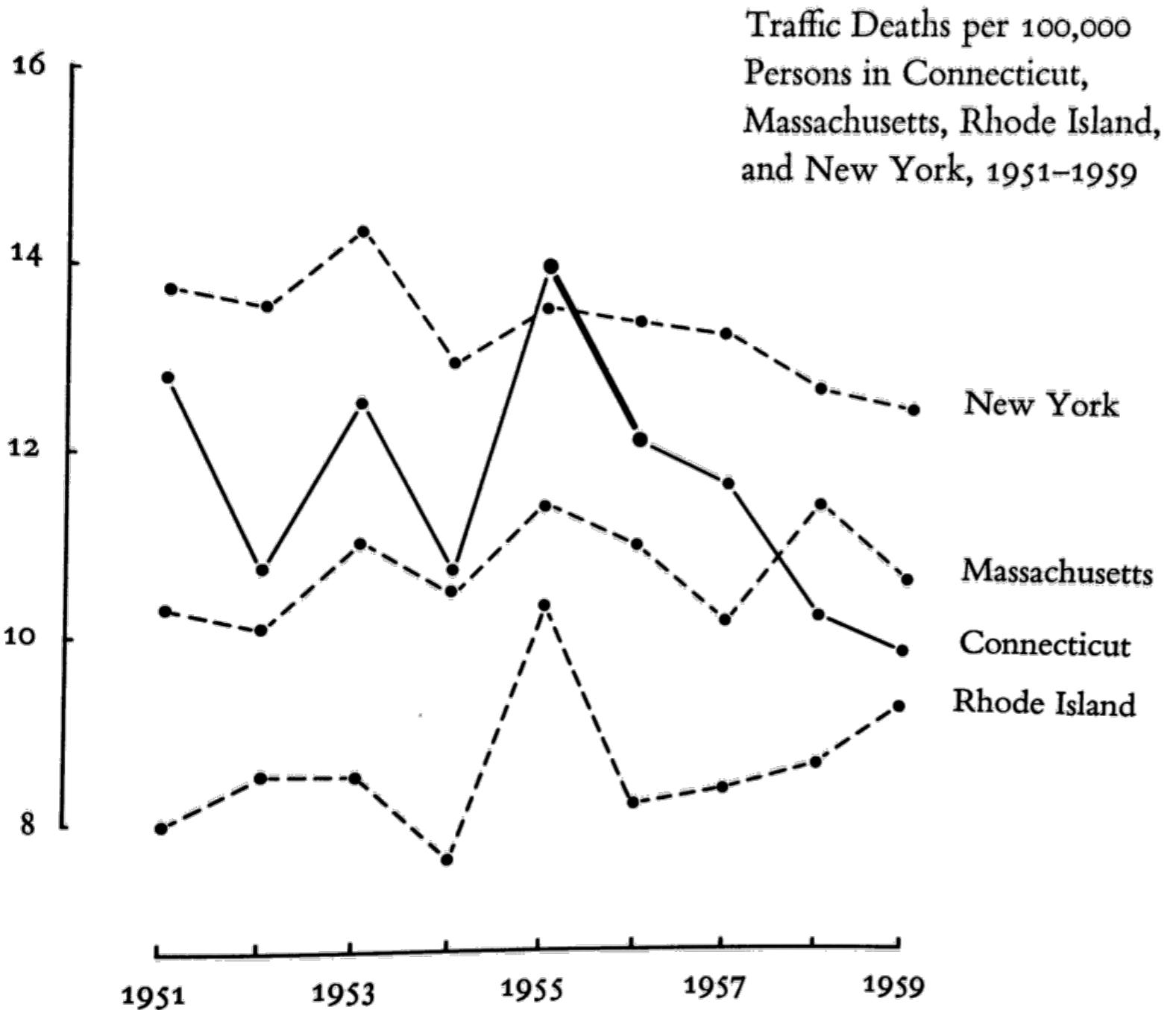
# Graphical Integrity



Connecticut Traffic Deaths,  
1951–1959

Graphics must not  
quote data out of  
context.

# Graphical Integrity



Graphics must not quote data out of context.

# Maximize Data Ink

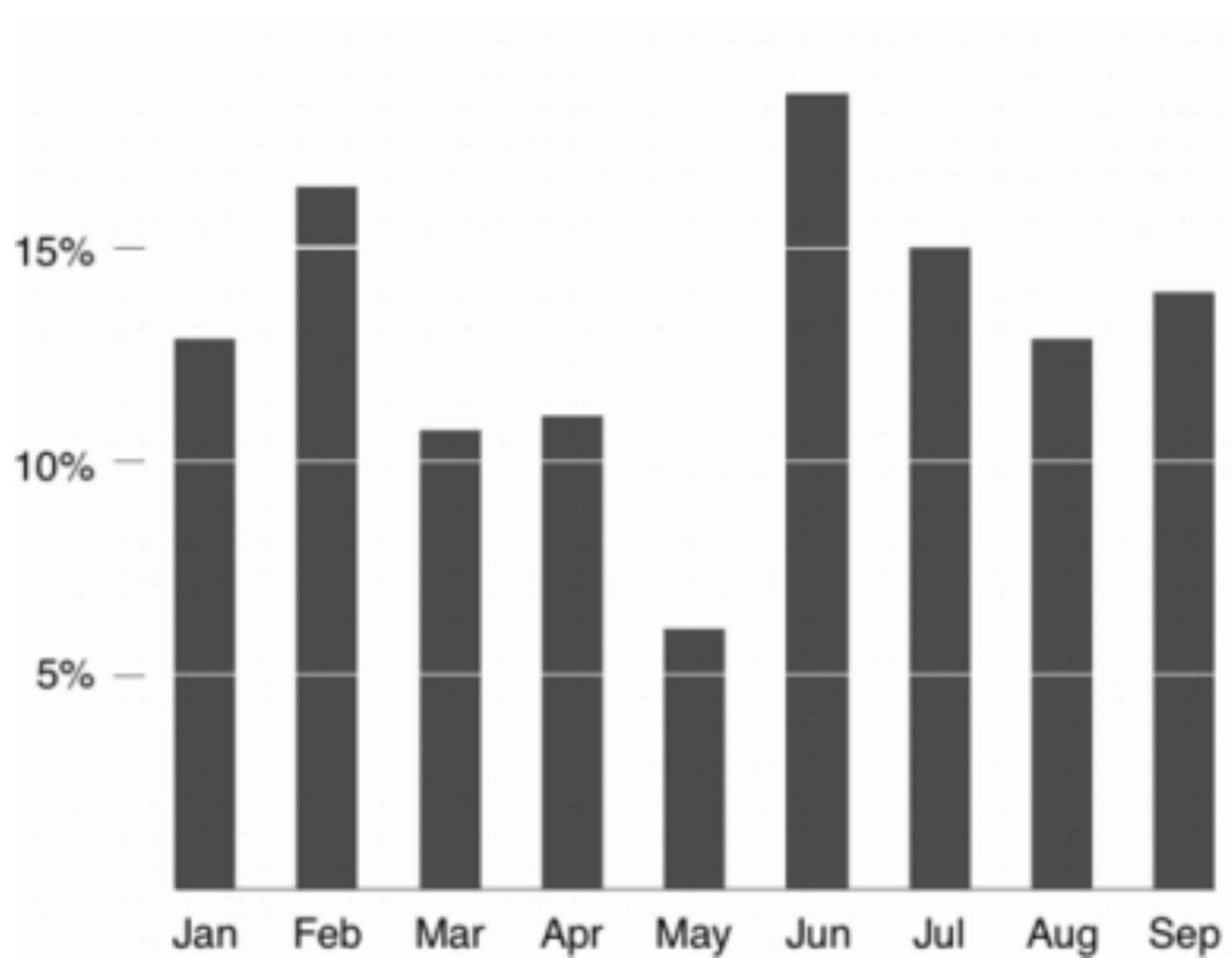
**Data-ink ratio** = 
$$\frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

= proportion of a graphic's ink devoted to the non-redundant display of data-information

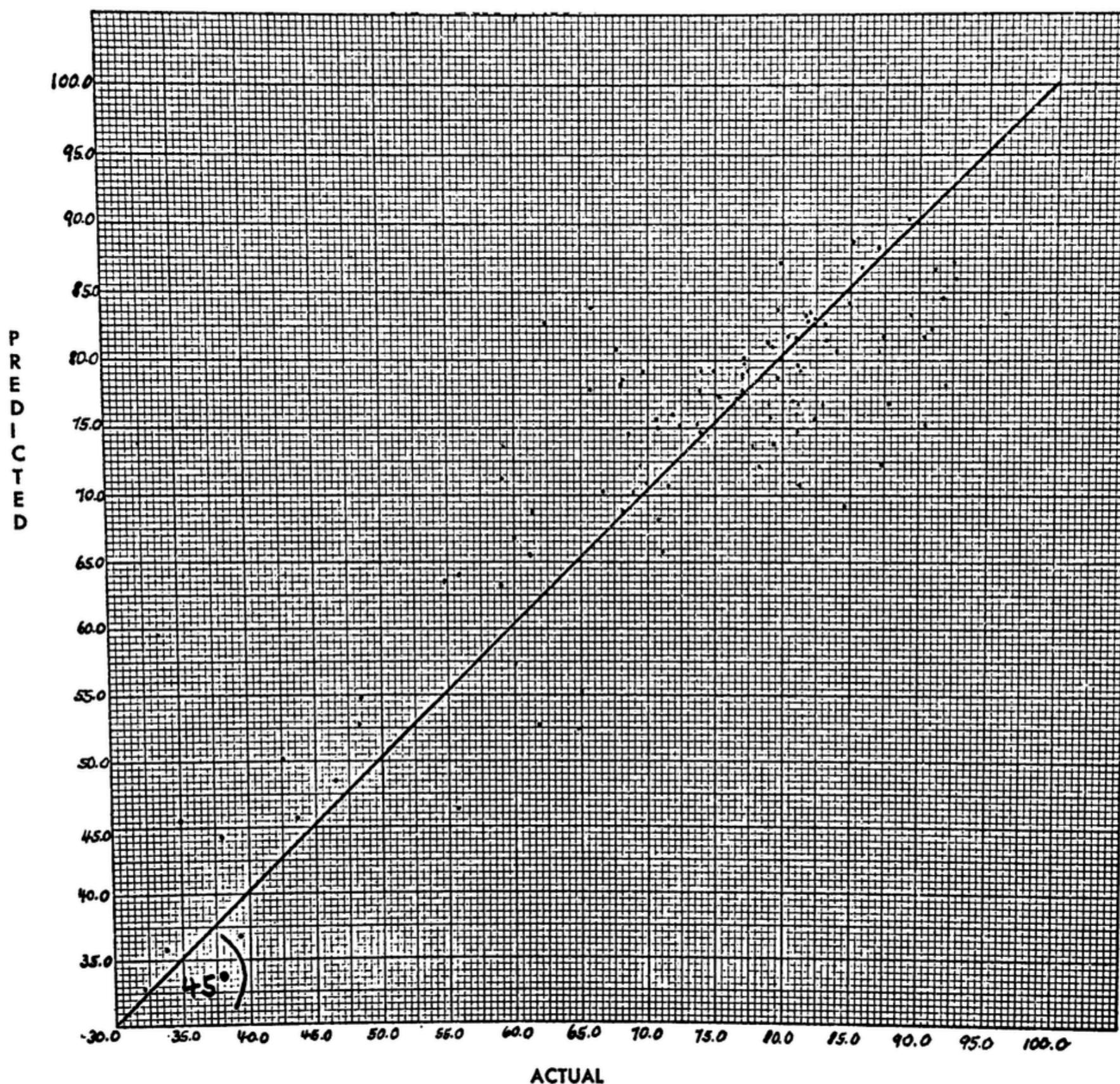
=  $1.0 - \text{proportion of a graphic that can be erased}$

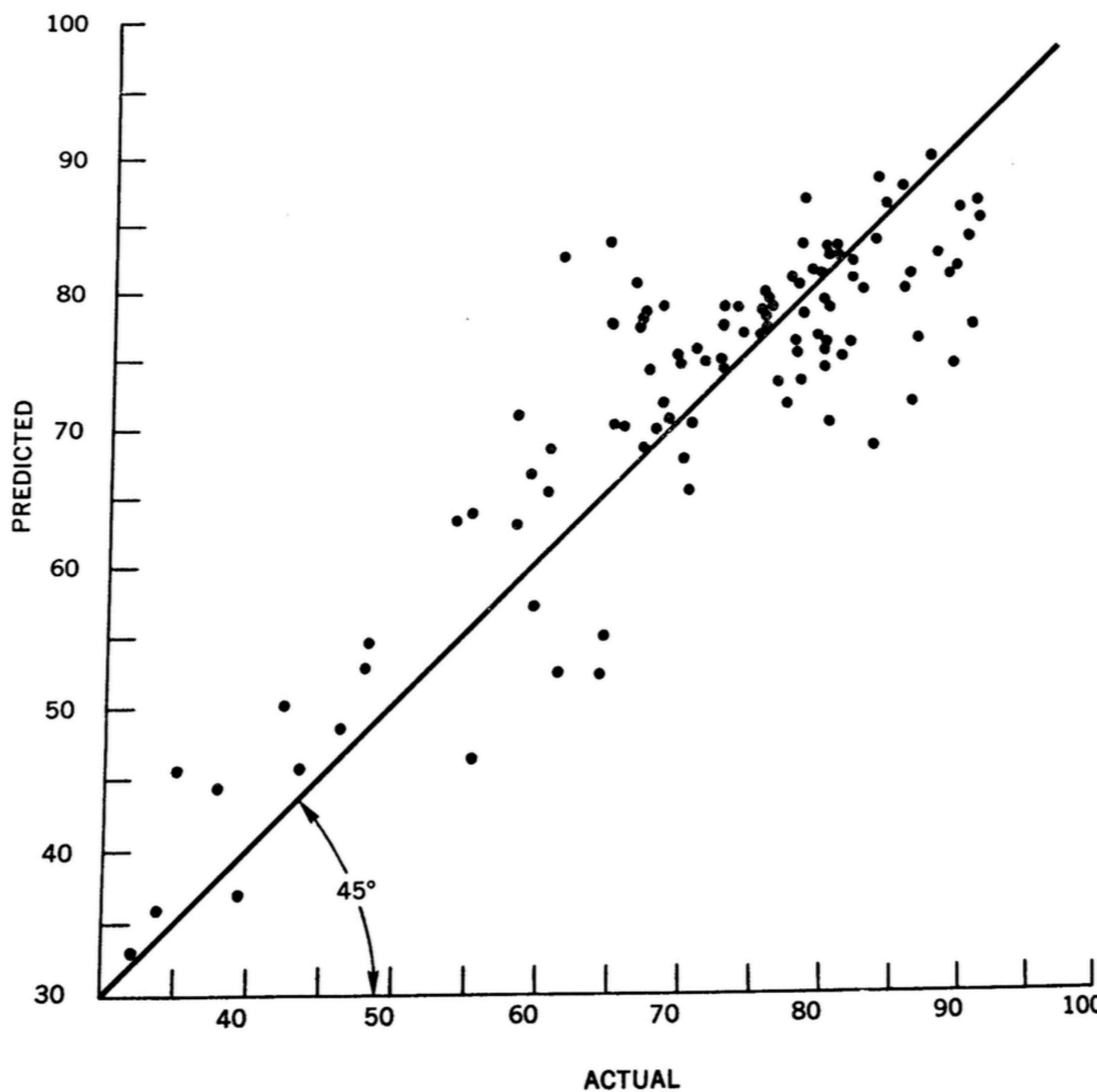
# Example: High or Low Data Ink ratio?

---

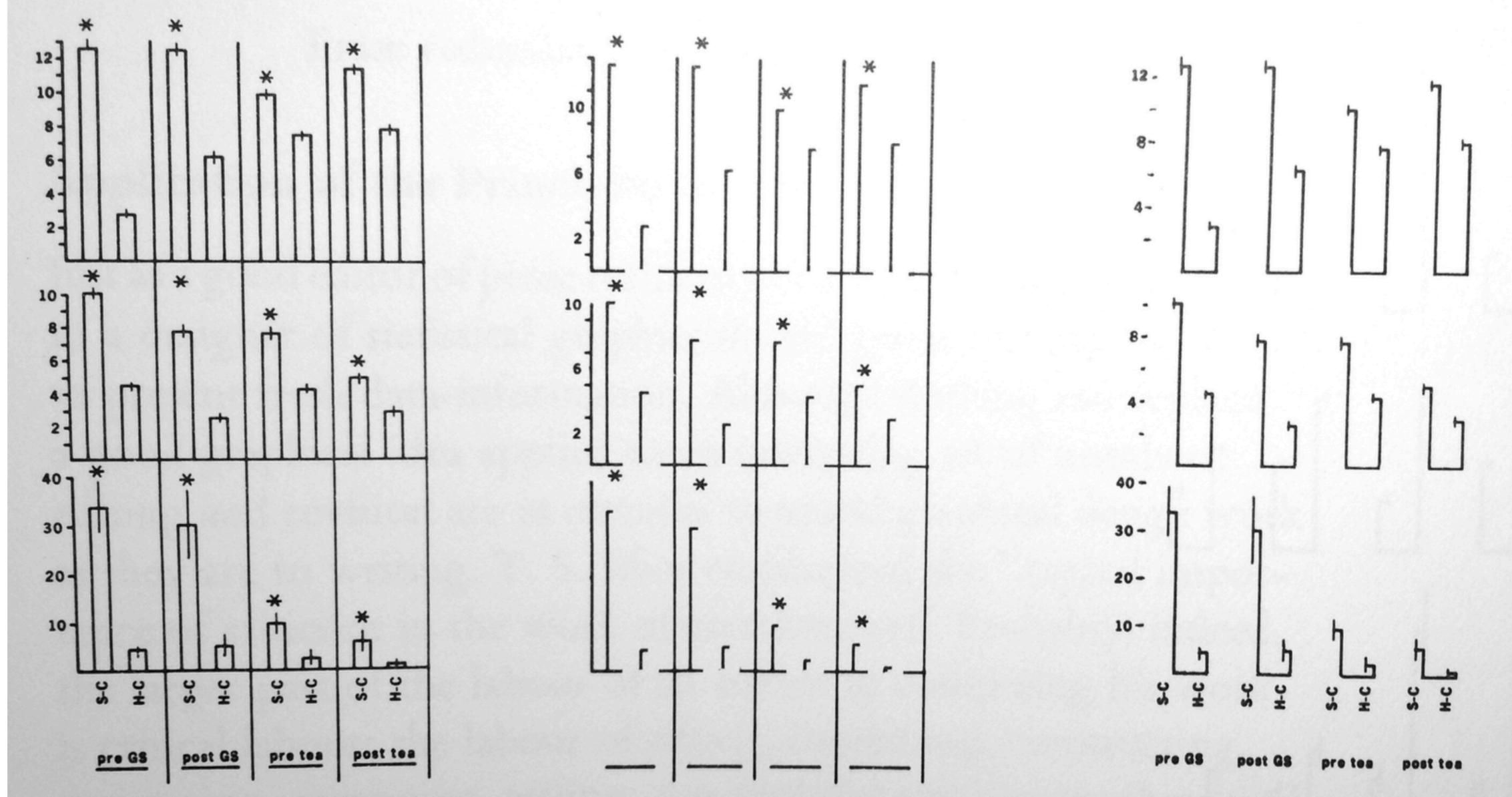


Relationship of Actual Rates of Registration to Predicted Rates  
(104 cities 1960).





# Maximize Data Ink



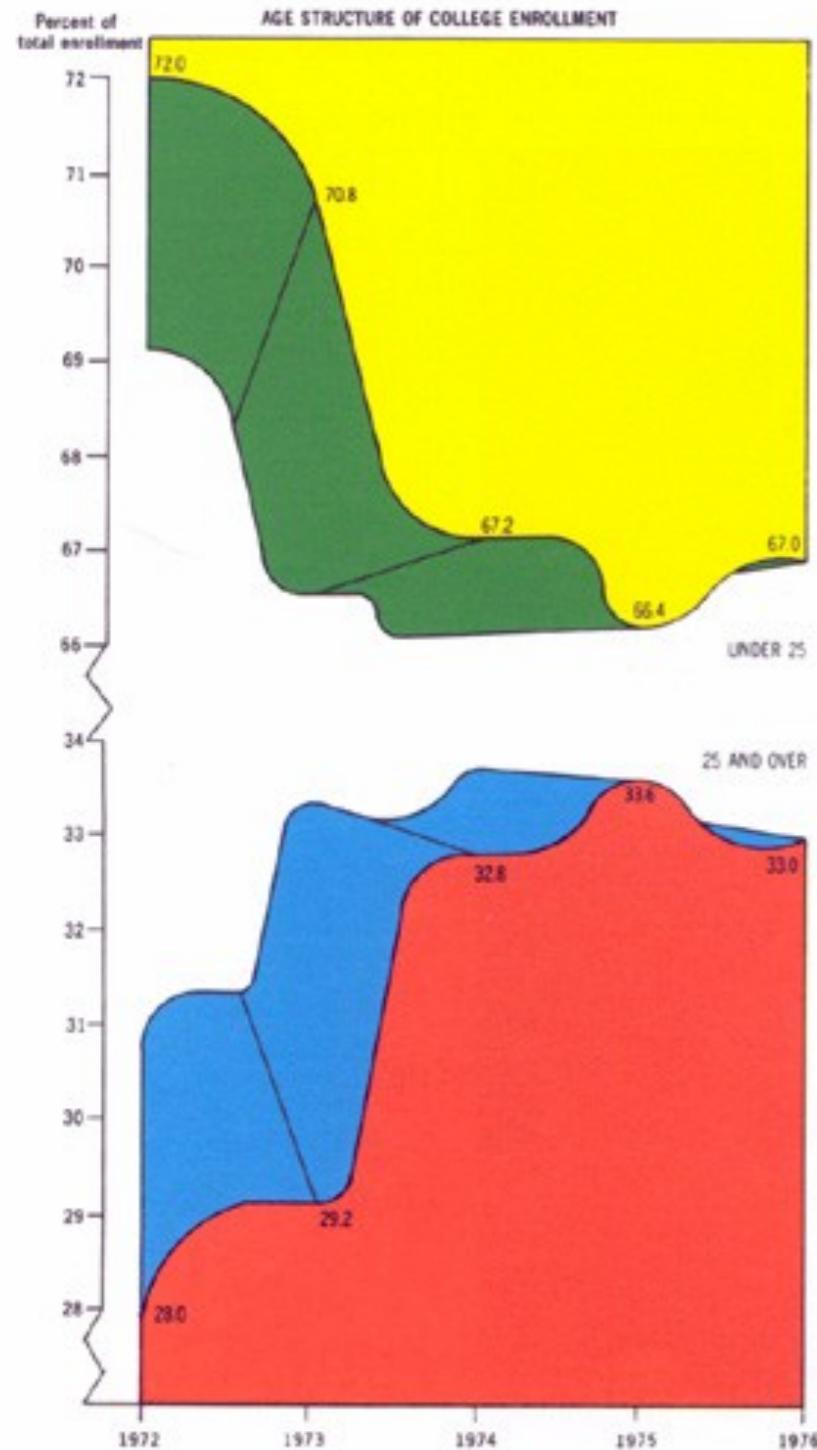
Trim the fluff from your graphics as much as possible!

## Avoid Chart Junk

The excessive and unnecessary use of graphical effects in graphs used to demonstrate the graphic ability of the designer rather than display the data.

# CHART JUNK

---



5 Colors  
2 Curvy lines  
2 Polygons  
2 Broken Axis  
3 Dimensions  
To display: 5 numbers

*Possibly the worst graph  
ever made*

# MONSTROUS COSTS

Total House and Senate campaign expenditures,  
in millions



Aim for high  
data density

The proportion of the  
total size of the graph  
that is dedicated  
displaying data.

# Shrink Principle

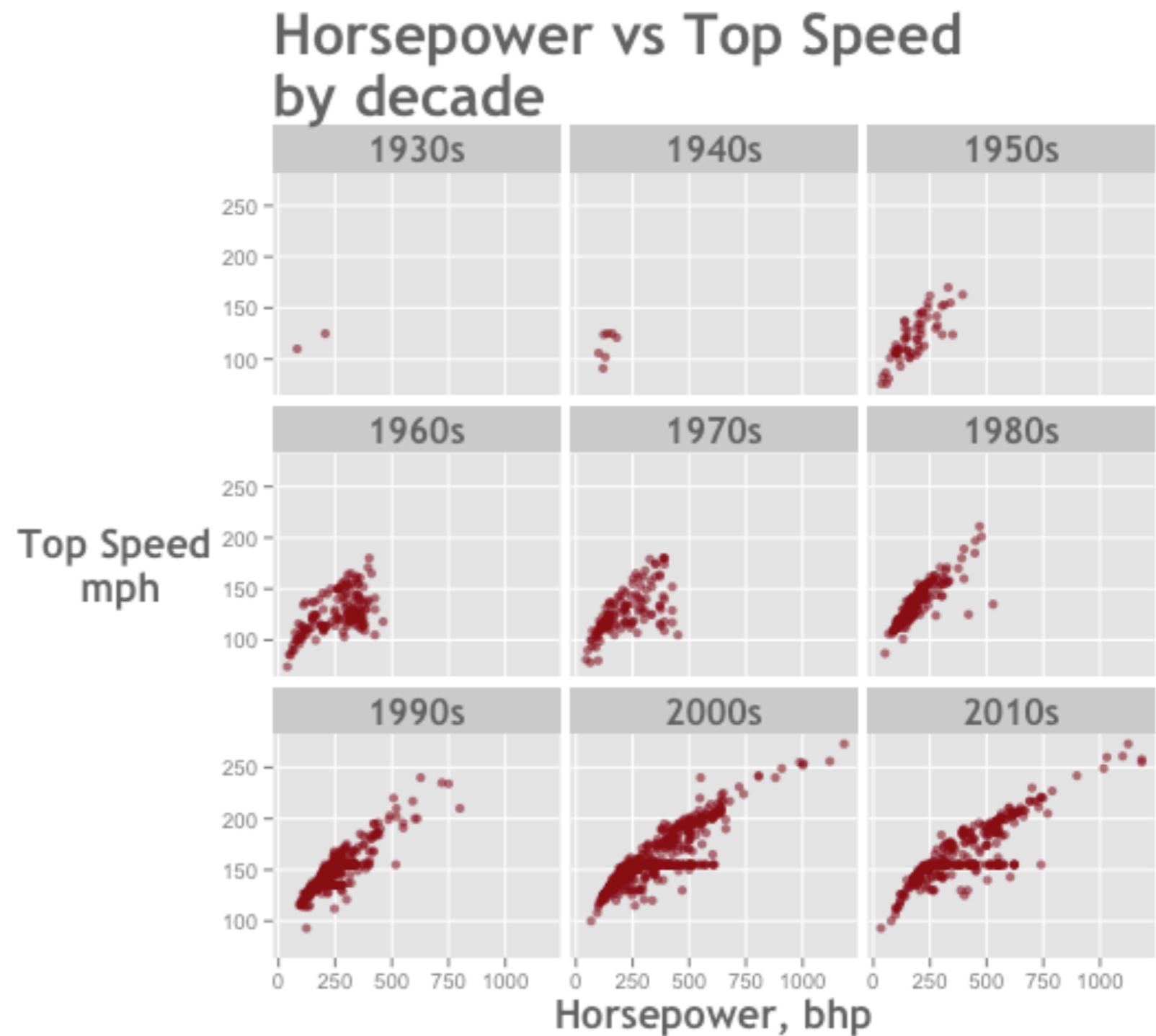
Many graphs can  
be shrunk way  
down without  
losing legibility or  
information



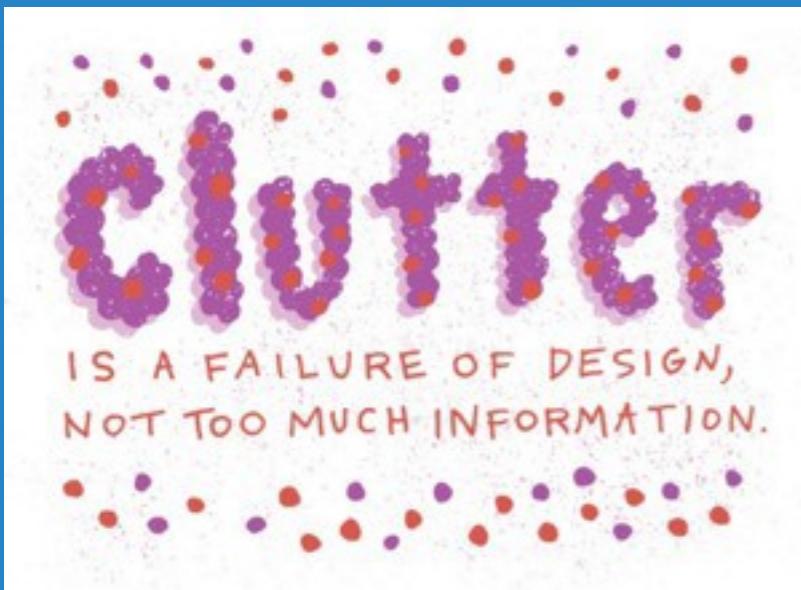
# Small Multiples

---

Series of the same  
small graph repeated  
in one visual



# Clutter



Visualization too cluttered?

Don't remove data, change the design.

Credibility comes from detail and in many cases one can clarify a design by adding detail.

## 2001 | Afghanistan NATO Invasion; Taliban Deposed

VIEWING

WORLD REFUGEES  
**12,031,996**

POPULATION  
6,195,665,261

REFUGEES / POPULATION  
1 of 515

TOP 3 OF 162 ORIGINS  
AFGHANISTAN  
3,809,767

BURUNDI  
553,999

IRAQ  
530,511



1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Hyperakt and Ekene Ijeoma visualized migrations  
over time and space in The Refugee Project  
<http://www.therefugeeproject.org>

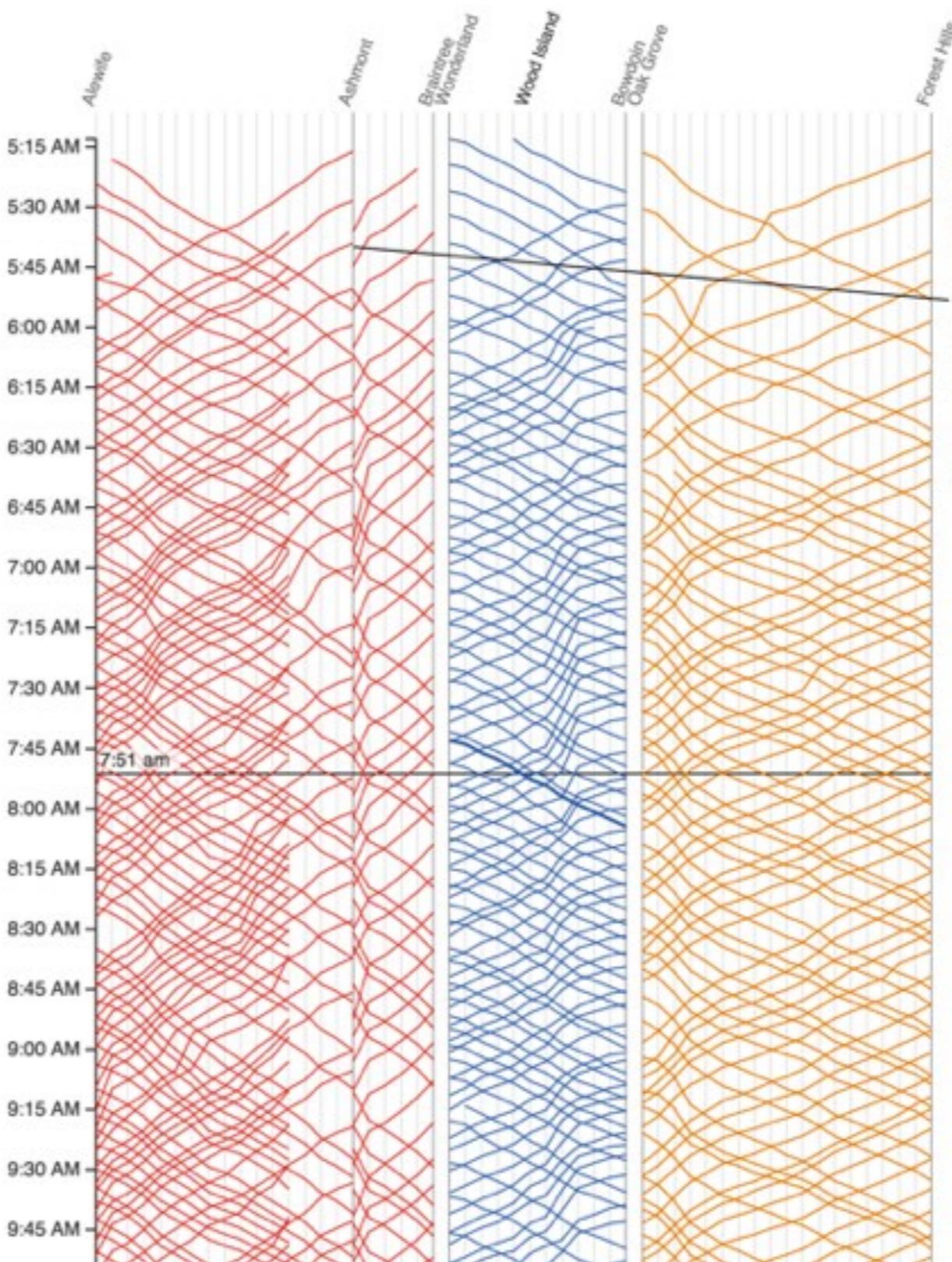


## Subway Trips on Monday February 3, 2014



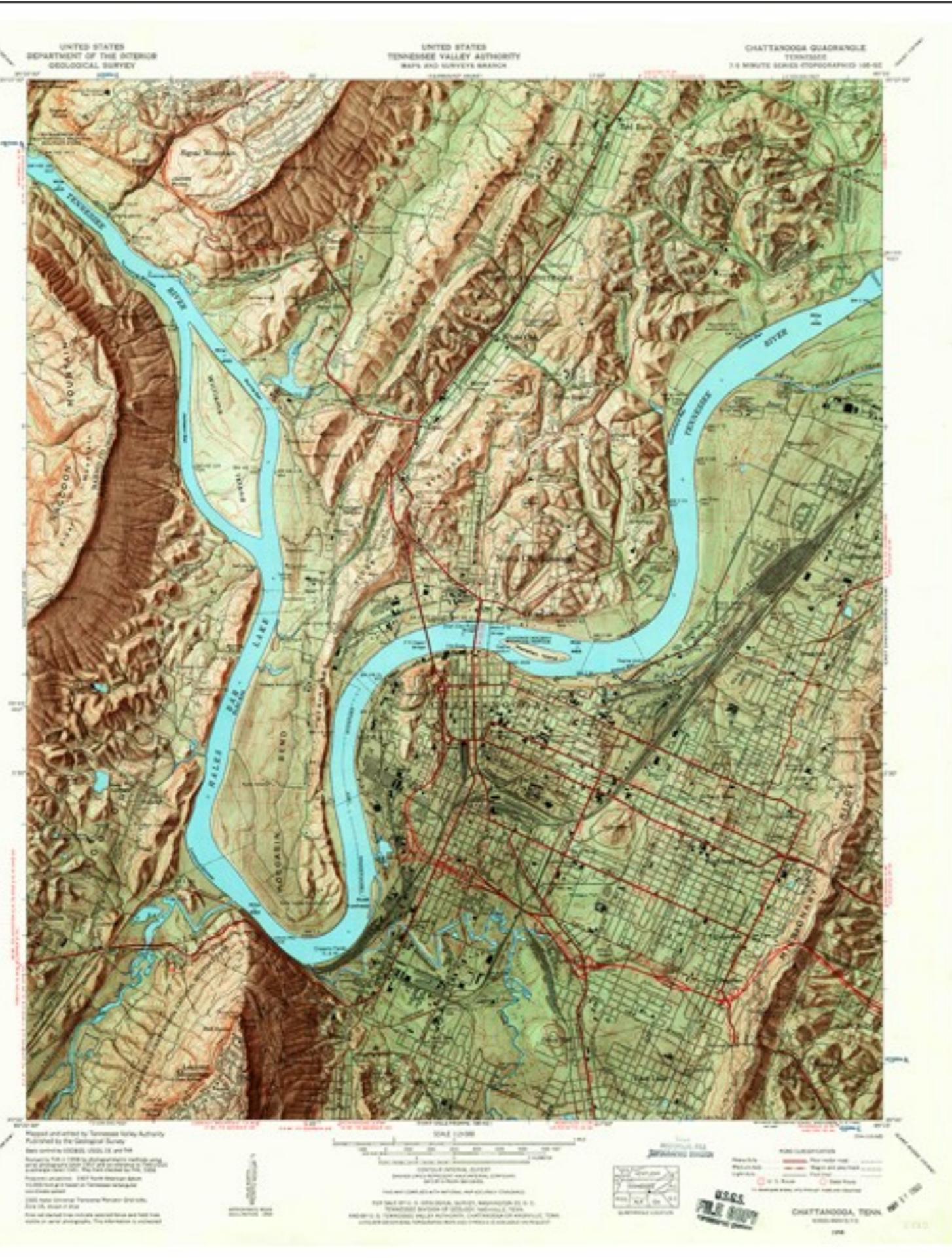
Trains are on the right side of the track relative to the direction they are moving.

See the [morning rush-hour](#), [midday lull](#), [afternoon rush-hour](#), and the [evening lull](#).



## Layering & Separation

Use color or other  
differentiation to  
separate important  
classes of information.



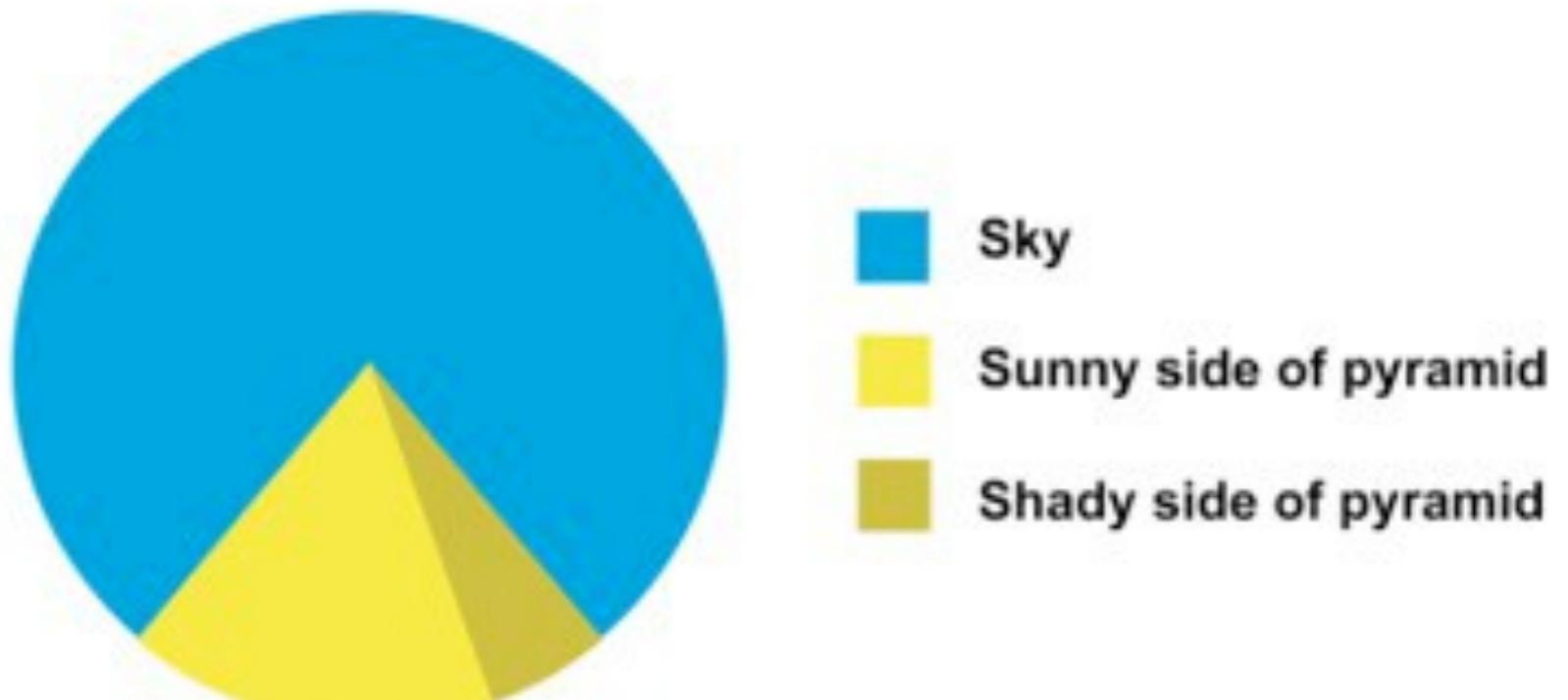
Muted colors, subtle shading and thin contour lines allow multiple types of data to be layered together in this 1958 topographic map of Chattanooga, Tennessee.

$$1 + 1 = 3 \text{ (or more)}$$

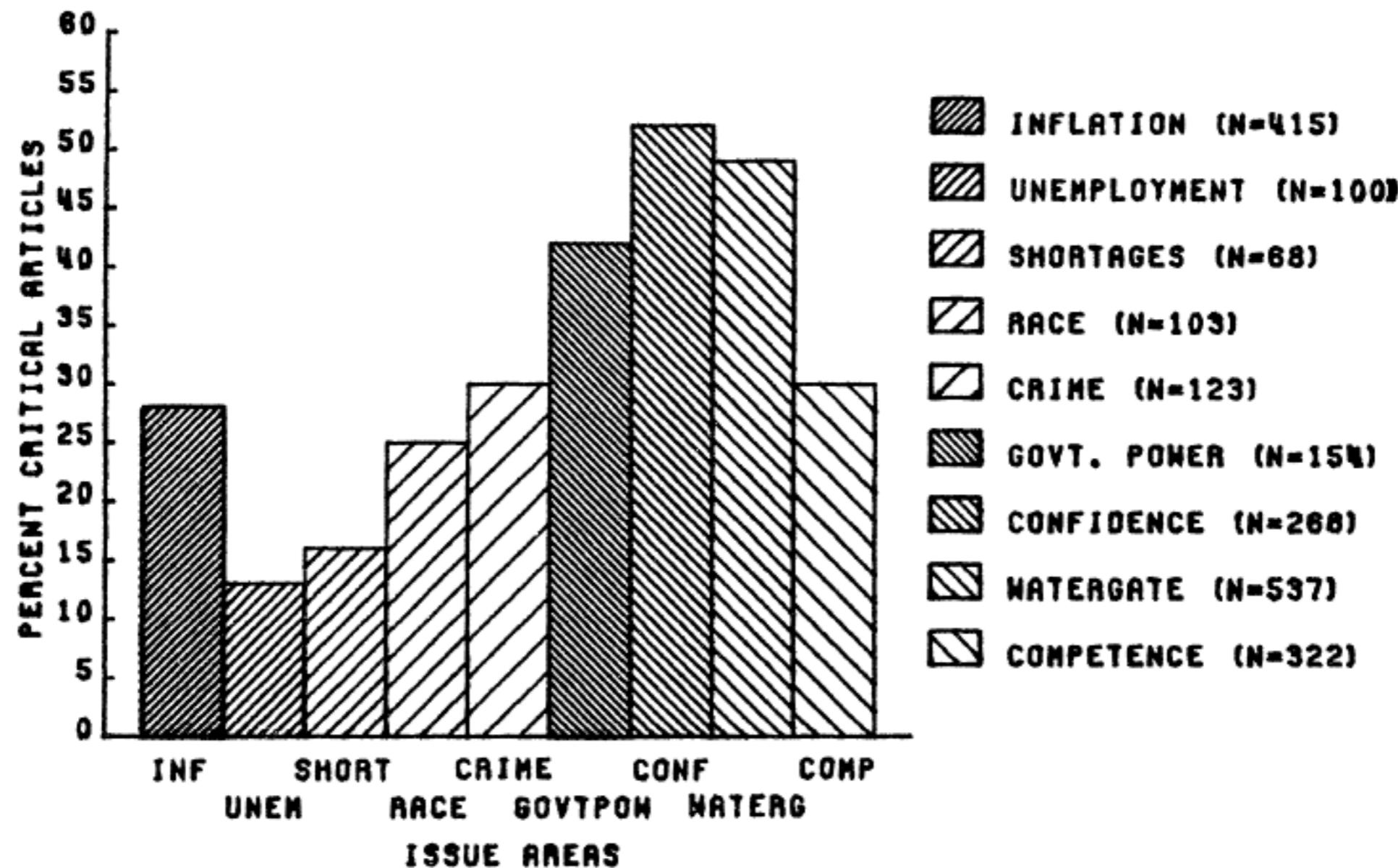
---

Effective layering  
of information is  
difficult because  
of interactions

Can create non-  
information  
patterns and  
texture

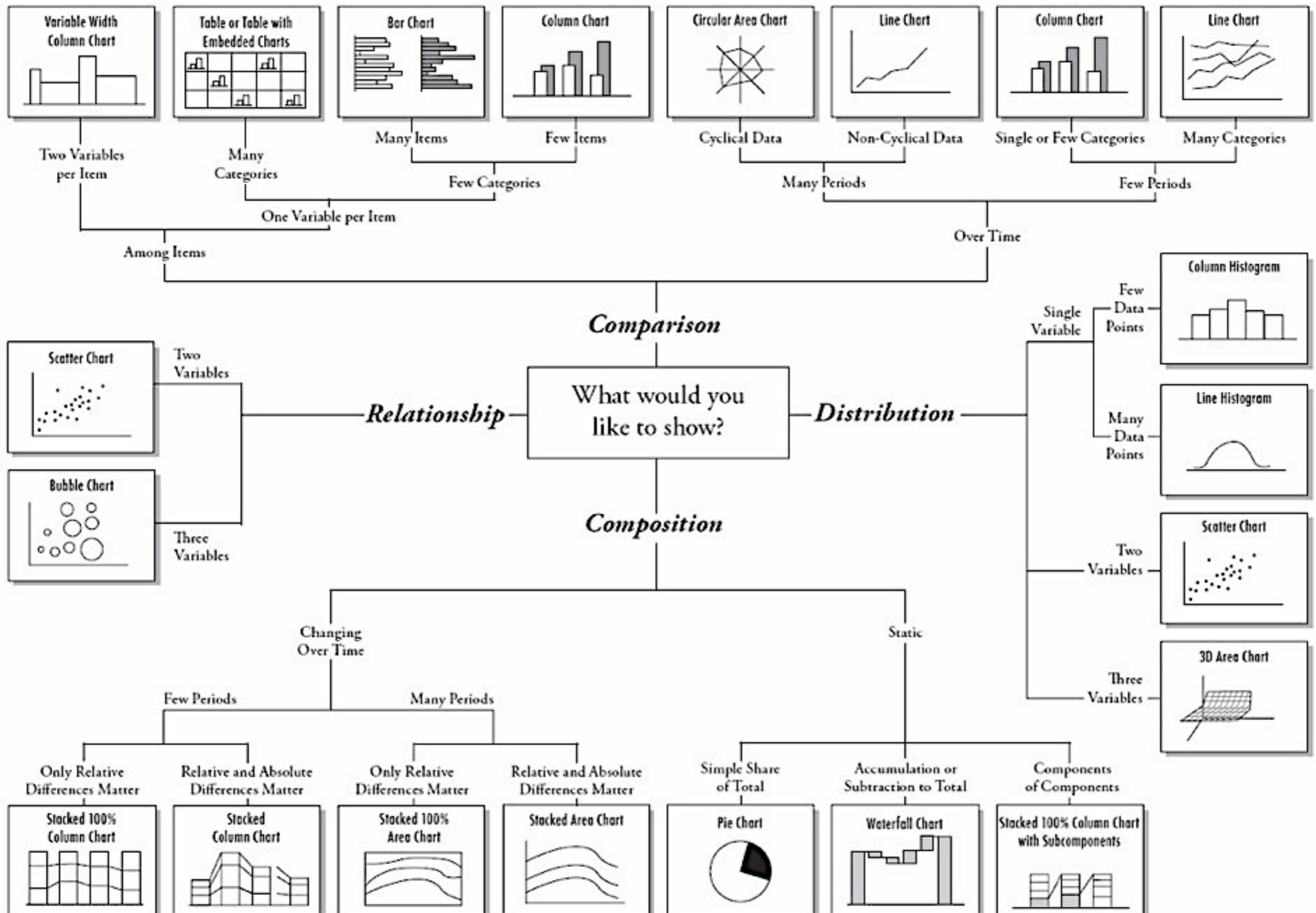


# Moire effect



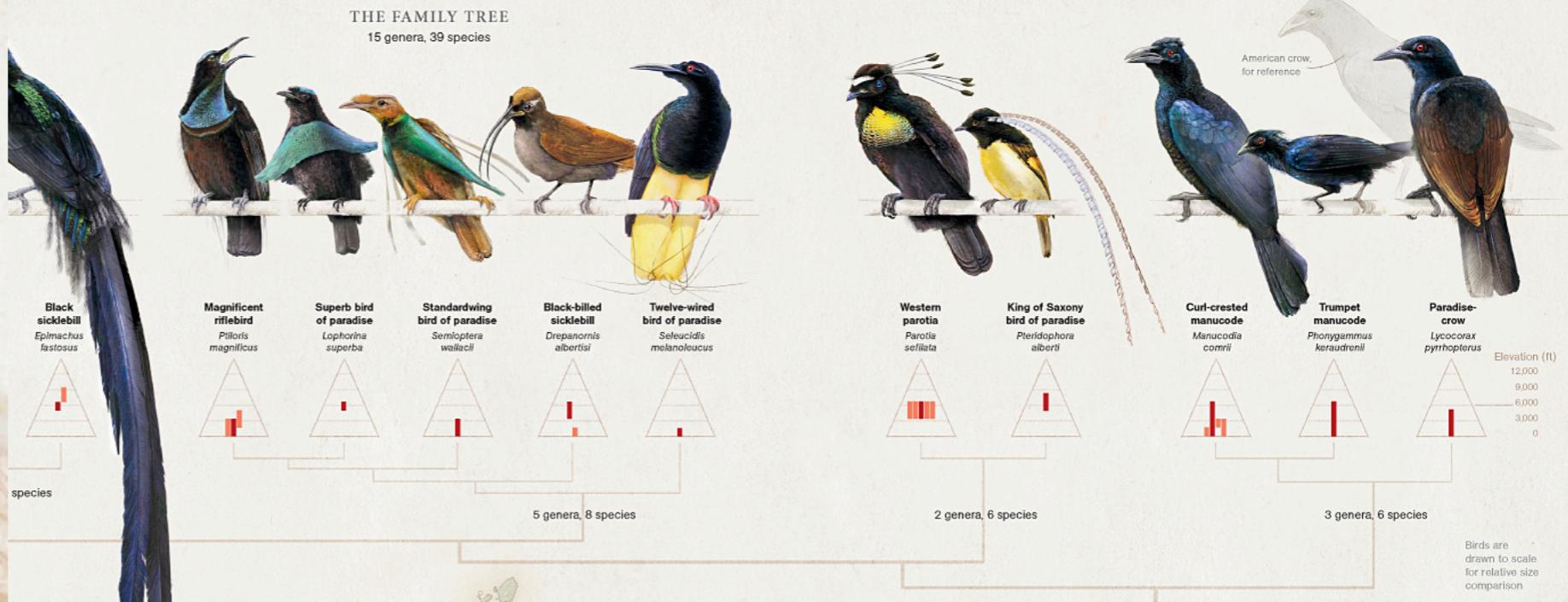
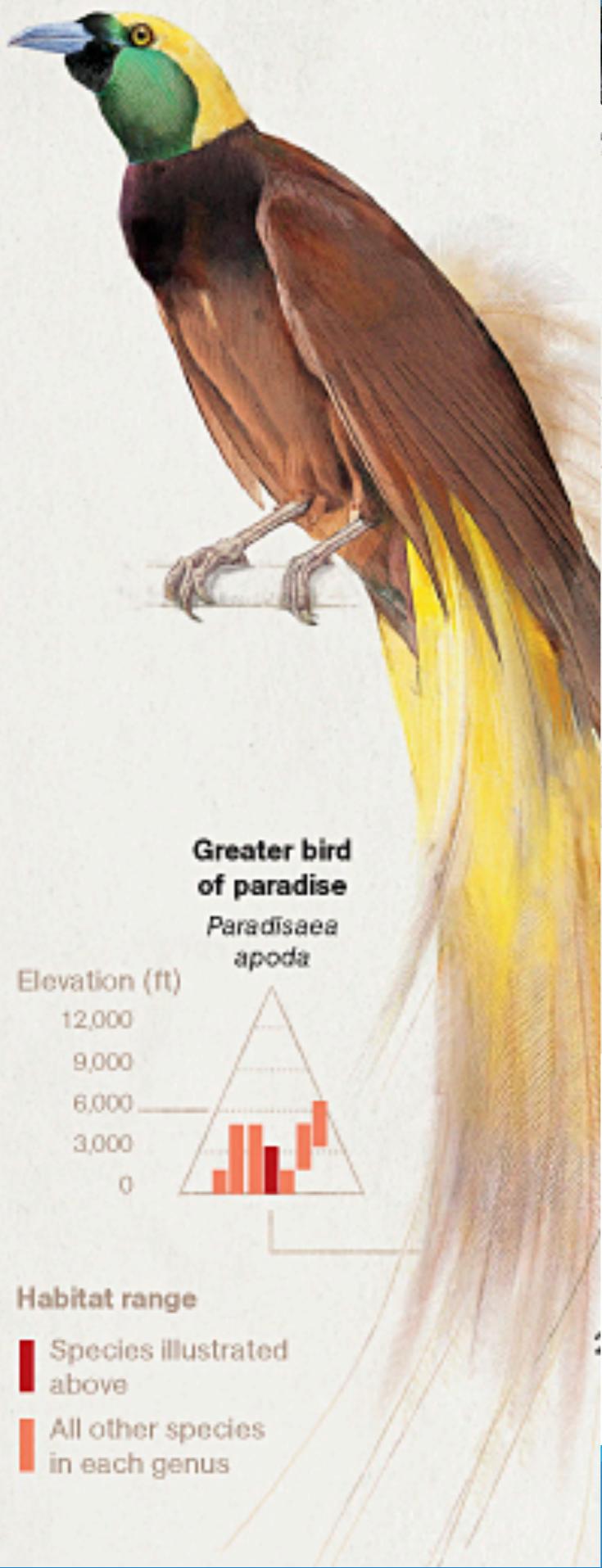
Source: Center for Political Studies Media Content Analysis Study, 1974; available through the University of Michigan, ICPSR. Not to be cited without full bibliographical reference to the present article.

Select the appropriate graphic for the data & analysis



# Aesthetics & Technique

- Carefully chosen graphical format
- Design uses words, numbers, and drawings
- Displays an accessible complexity of detail
- Has a narrative quality, a story to tell about the data
- Technical details of production done with care
- Avoid content-free decoration, including chartjunk



Juan Velasco. Cornell ornithologist Edwing Scholes and biologist/photographer Tim Laman. Senior Graphics Editor Fernando Baptista, Graphics Specialist Maggie Smith and freelance researcher Fanna Gebreyesus. National Geographic

# The final word...

The principles should not be applied rigidly or in a peevish spirit; they are not logically or mathematically certain; and *it is better to violate any principle than to place graceless or inelegant marks on paper*

--Tufte, Visual Display of Quantitative Information, page 191