

6

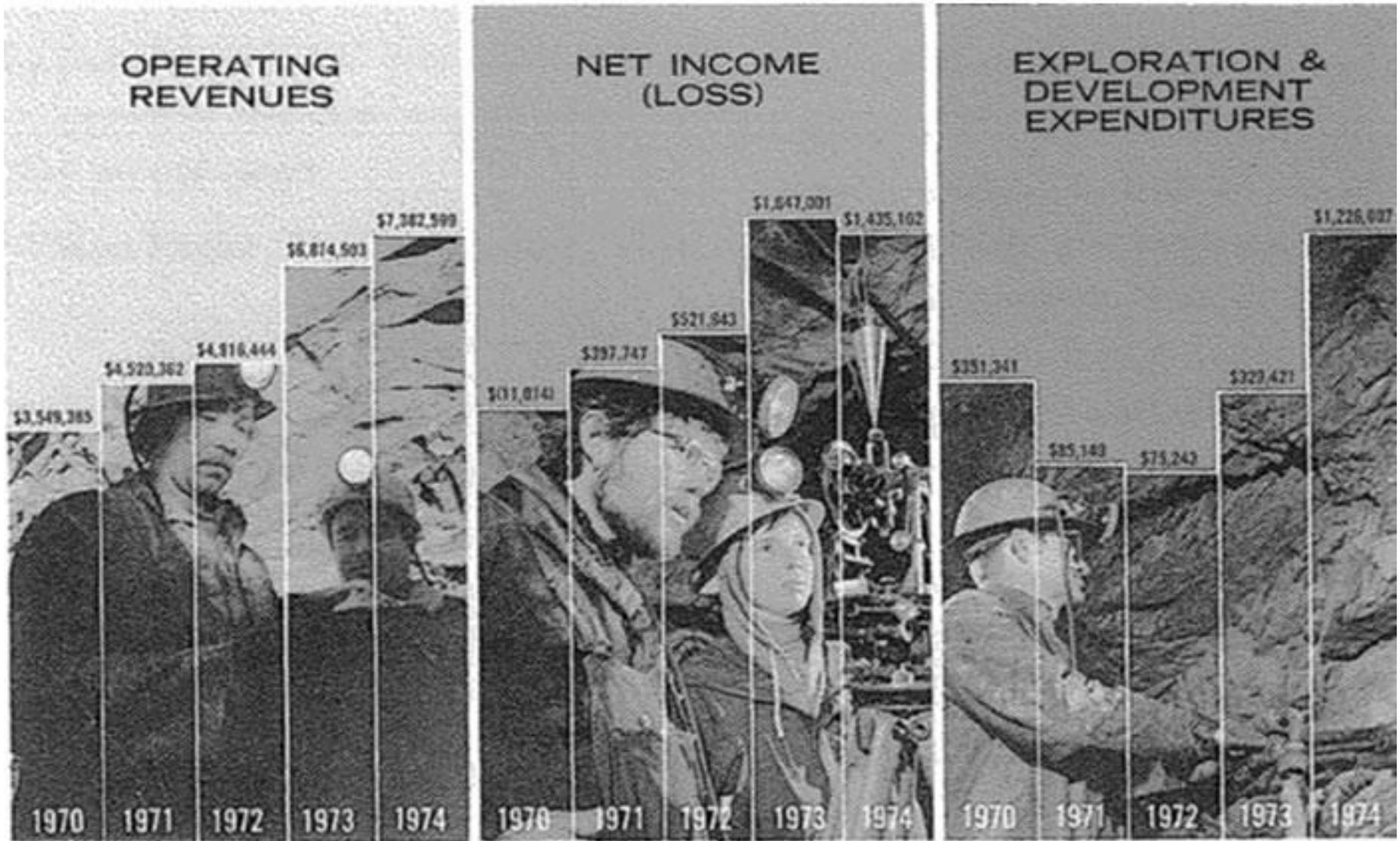
Evaluation and Integrity

Graphics, lies, and interpretation

Graphical Integrity

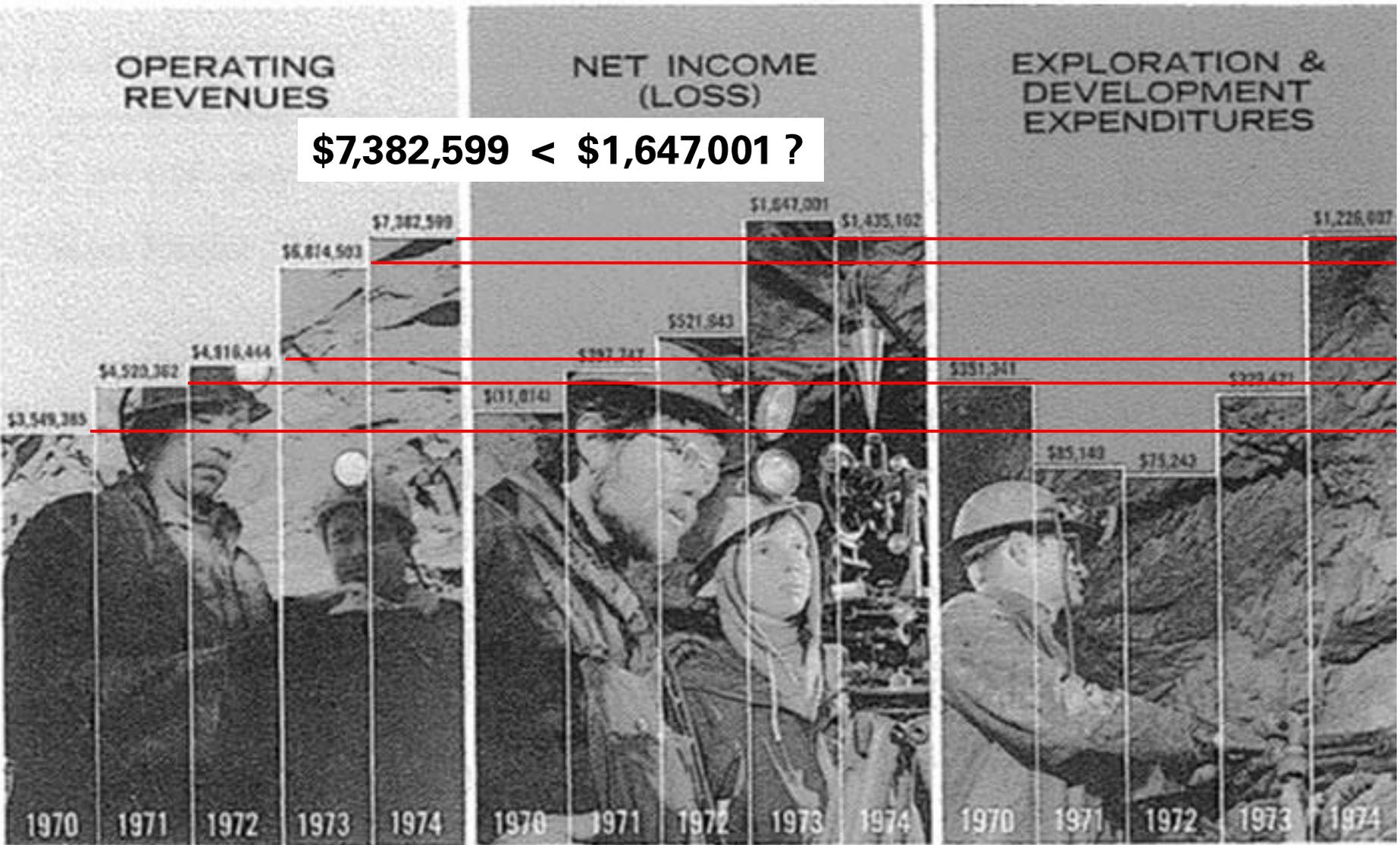
Graphical Integrity

The quality of honesty in a data graphic and the avoidance of misleading design.



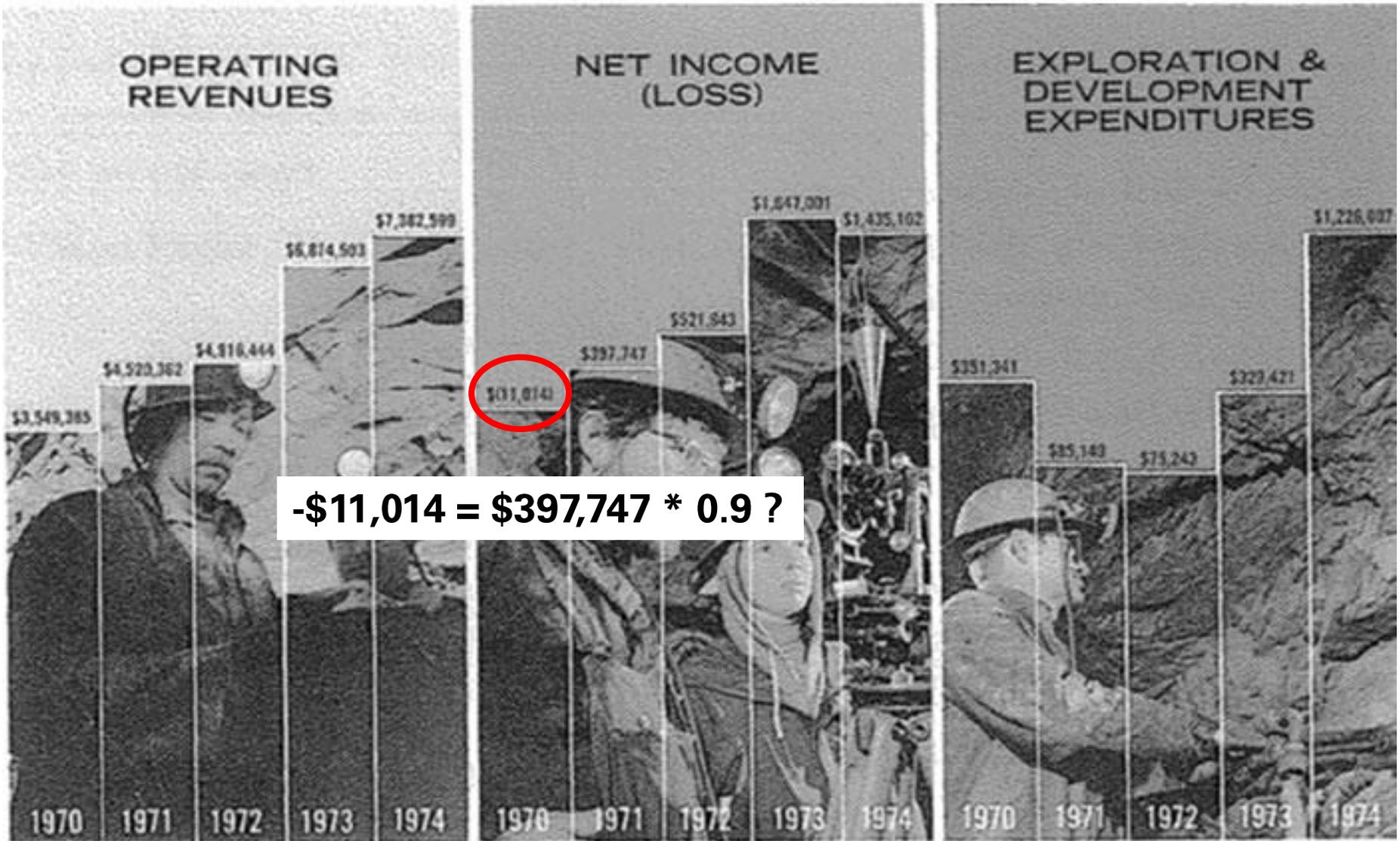
Dry Mines Inc., 1974

Let's revisit an old favorite. What is hurting this chart?



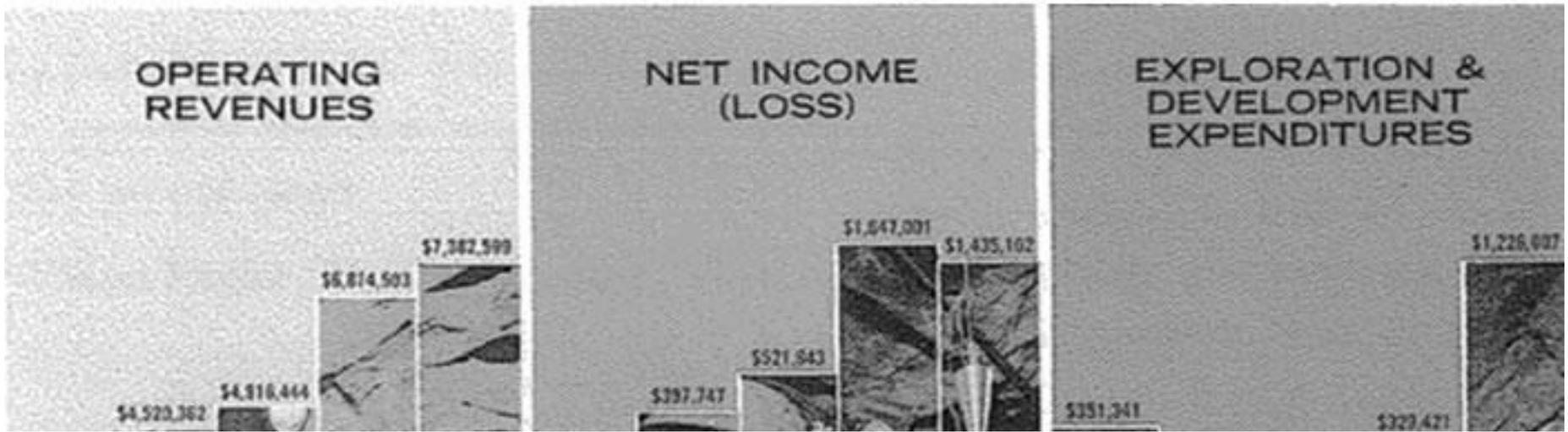
Dry Mines Inc., 1974

Lack of consistency defeats ability to make comparisons



Dry Mines Inc., 1974

And of course, the smoking gun.



A classic case of missing graphical integrity.



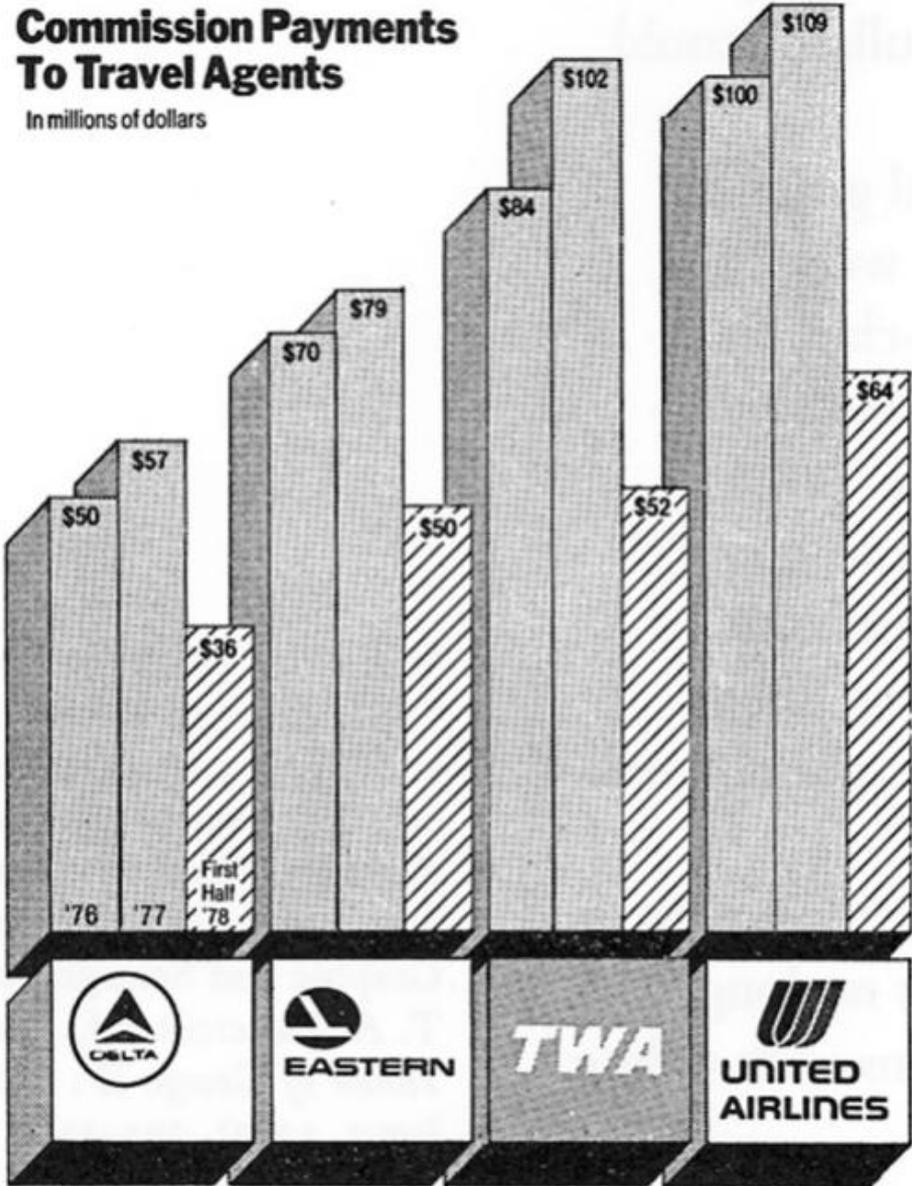
Dry Mines Inc., 1974

Design principles with **integrity**

- 1) Proportions in graphics should accurately represent raw data

Commission Payments To Travel Agents

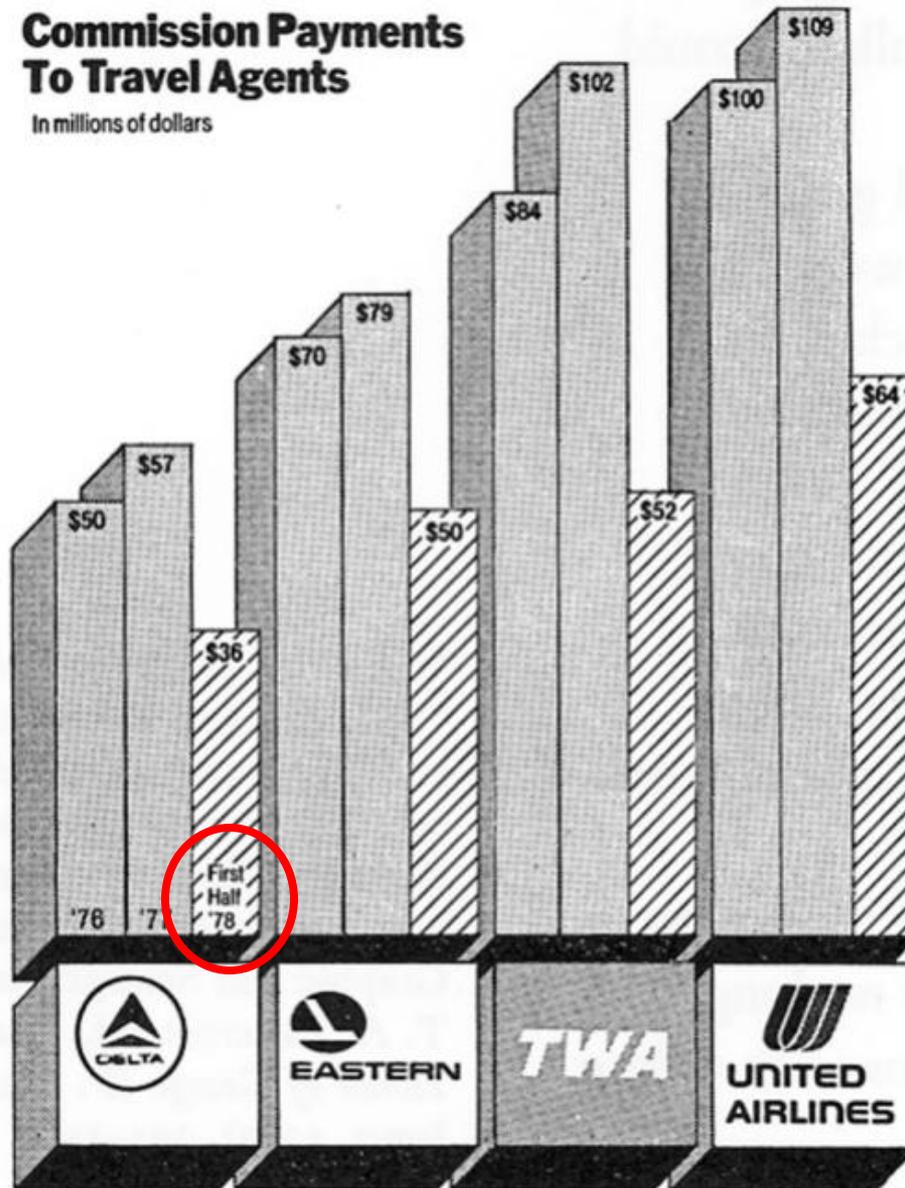
In millions of dollars



What's wrong with this picture?

Commission Payments To Travel Agents

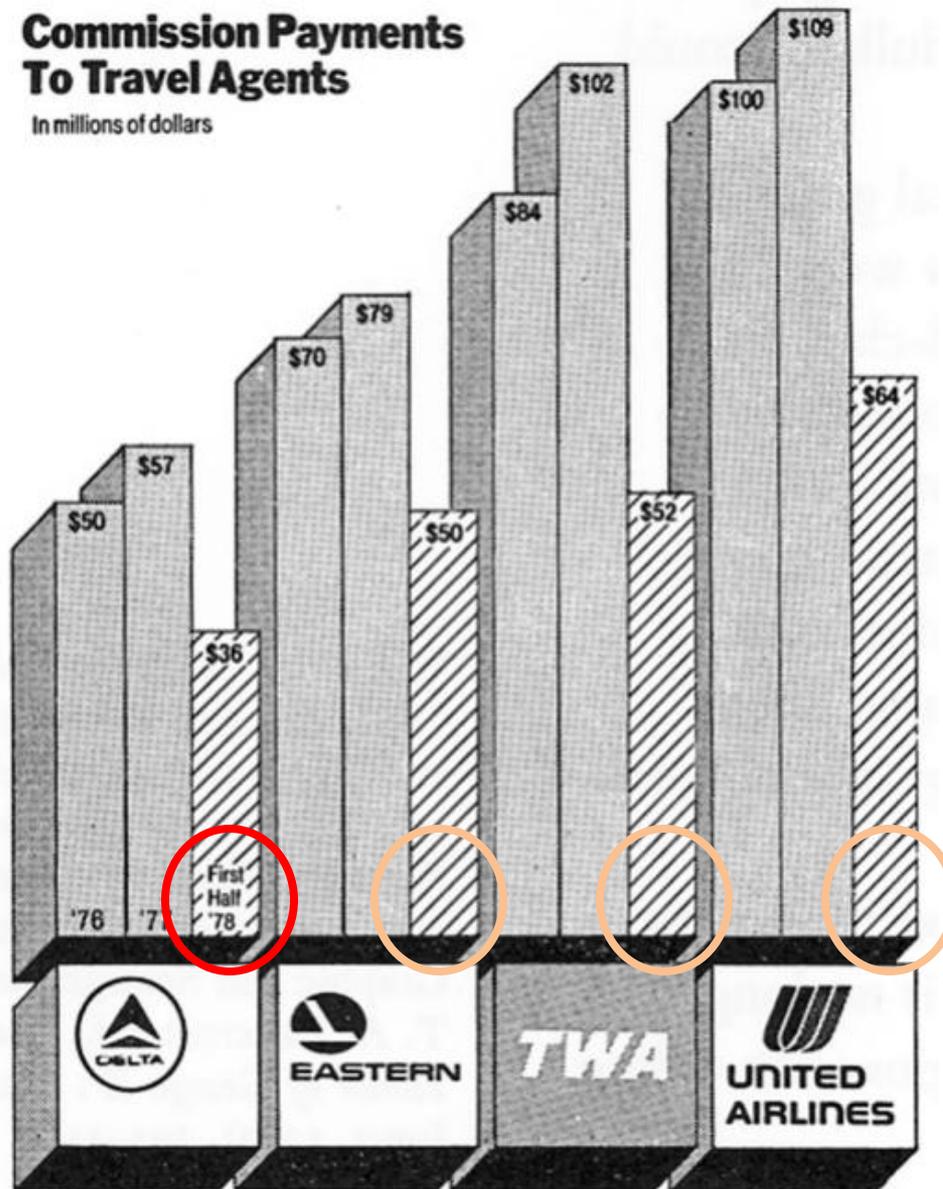
In millions of dollars



Inconsistency in labeling makes the graph misleading

Commission Payments To Travel Agents

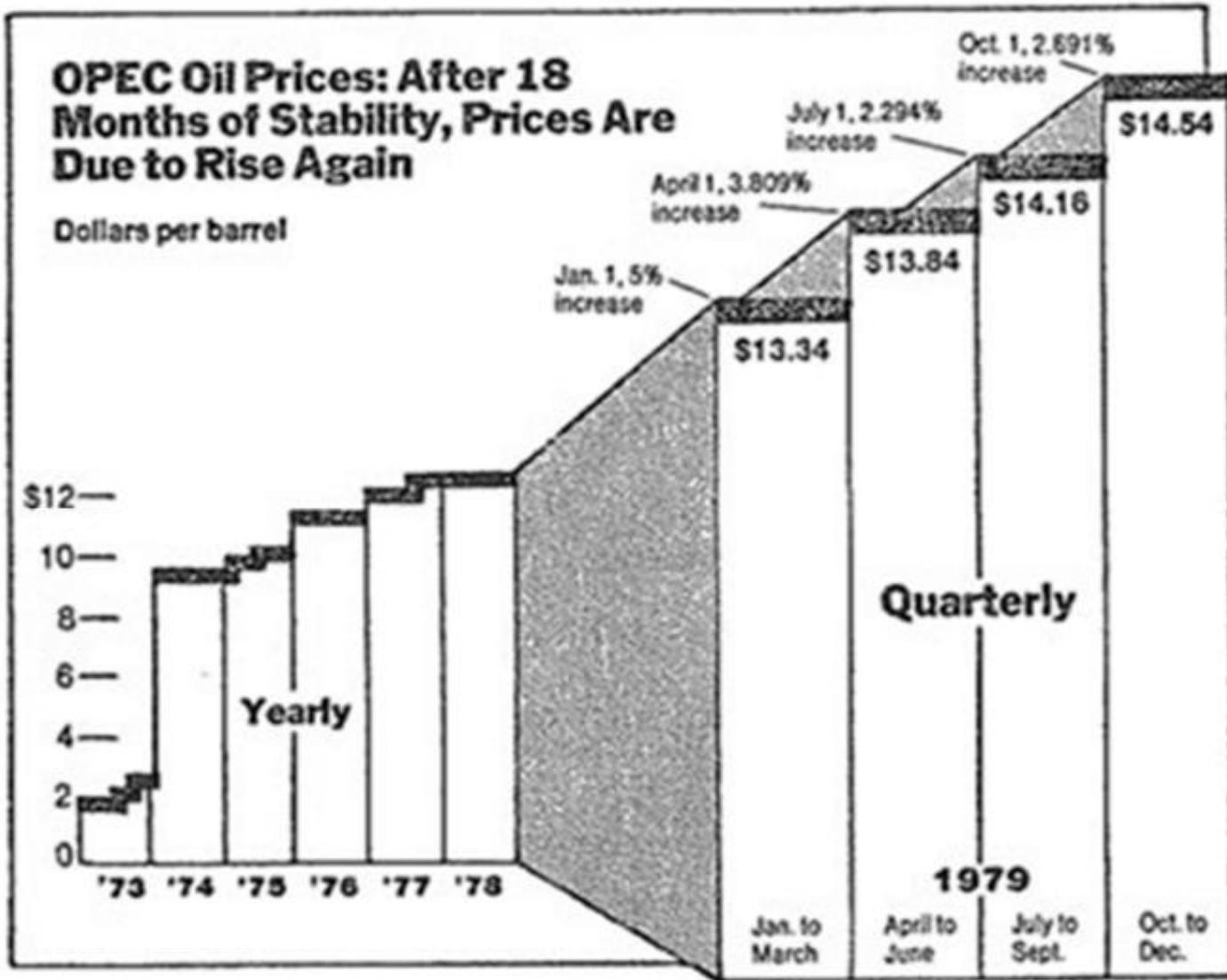
In millions of dollars



Not to mention the important text isn't even repeated.

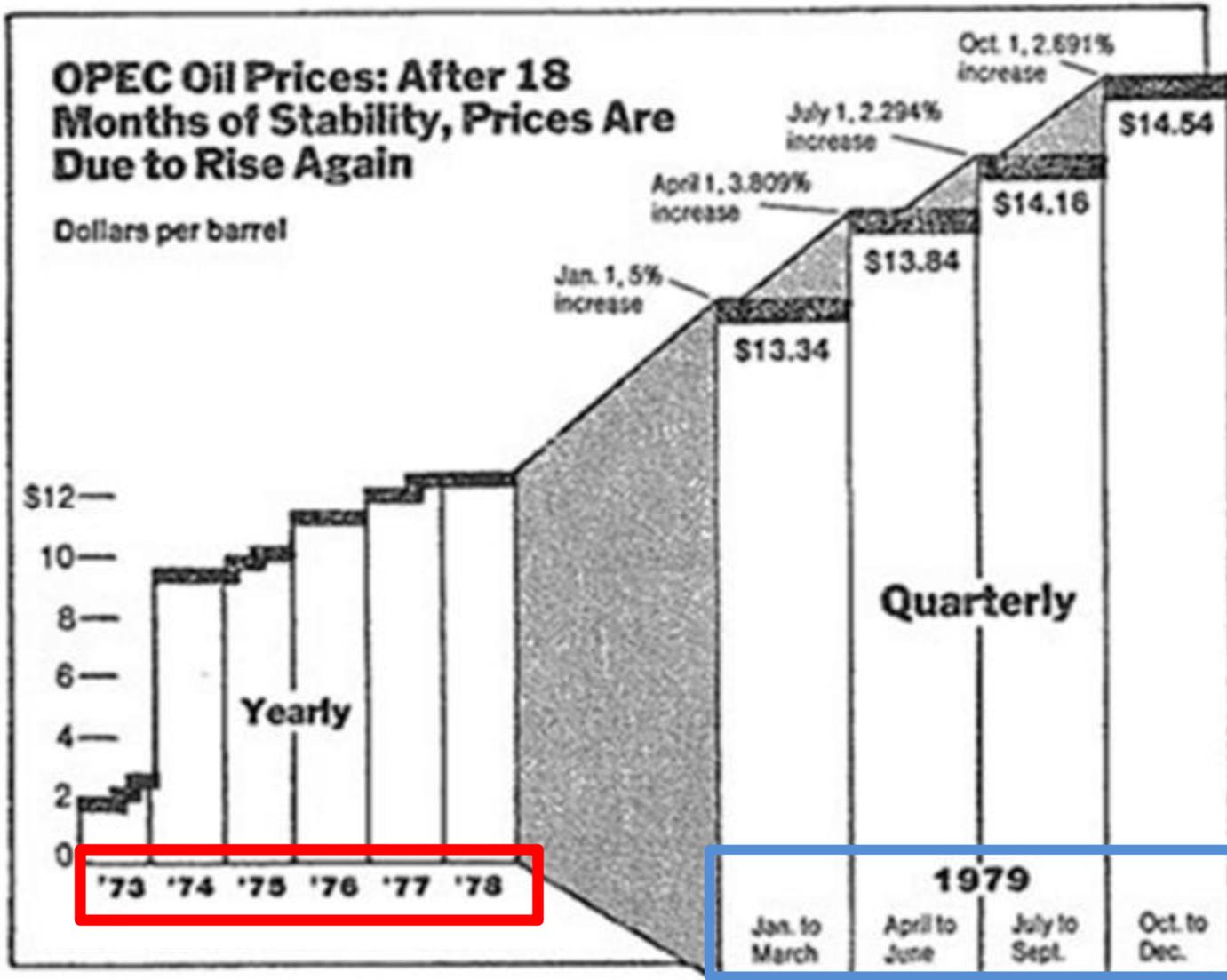
Design principles with **integrity**

- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive



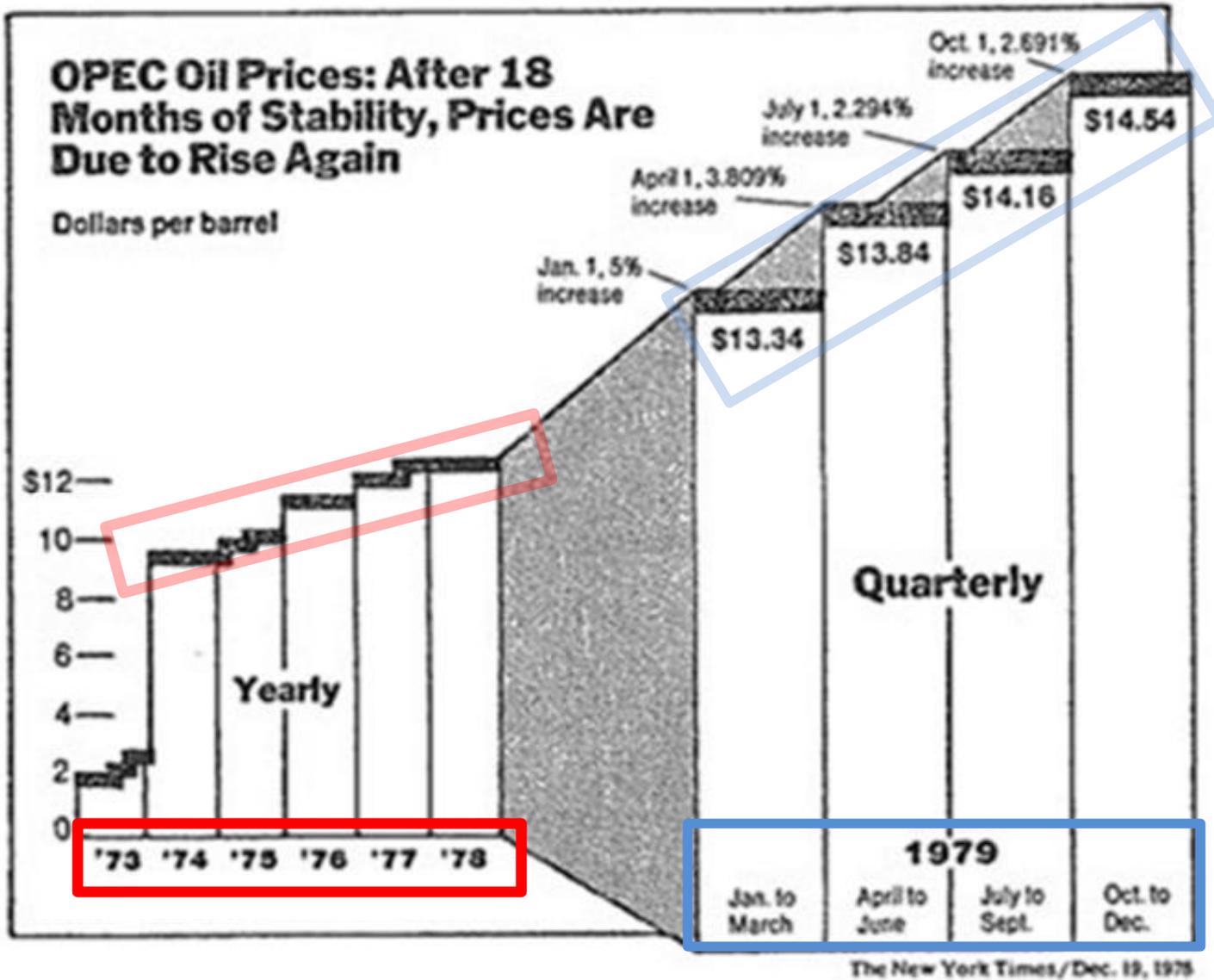
The New York Times/Dec. 19, 1978

Once again, where is the issue here?



The New York Times/Dec. 19, 1978

More inconsistency in labeling...



... and design. Are we looking at one chart or two?
 The means of comparison is different between the halves.

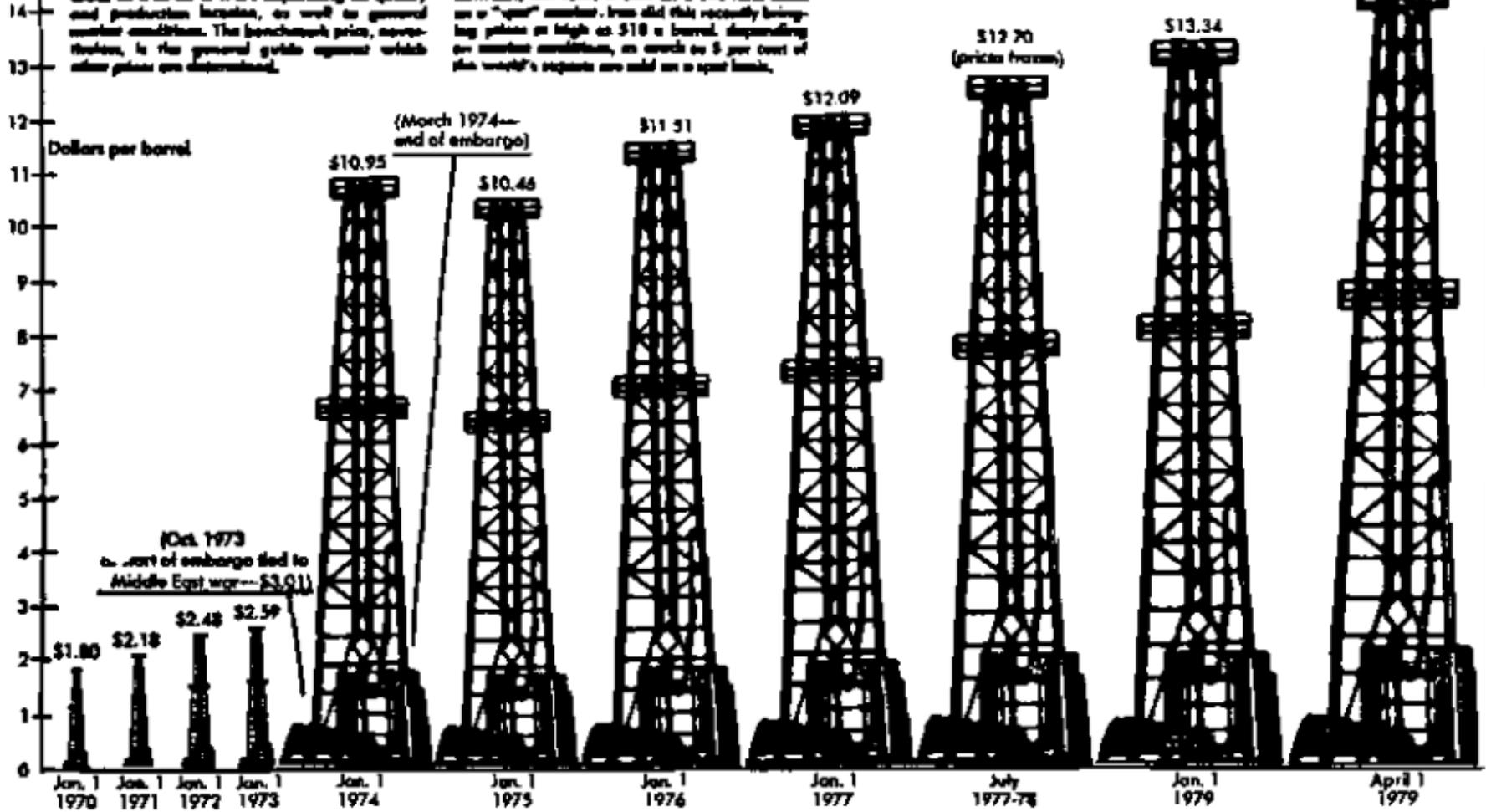
Design principles with integrity

- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive
- 3) Data should vary, not design

OPEC Benchmark Prices 1970-1979

OPEC prices are set for the so-called benchmark or "marker" light oil sold by Saudi Arabia, a high quality oil that is more widely refined. Actual market prices, however, vary widely by as much as a dollar or more depending on quality and production location, as well as general market conditions. The benchmark price, nevertheless, is the general guide against which other prices are determined.

Nearly all of the non-communist world, including non-OPEC production from countries such as Alaska or Norway, is sold at the OPEC set price. Most of this oil is sold under long-term contracts, but some is sold on a one-time basis as a "spot" market. Even oil that recently brought prices as high as \$18 a barrel, depending on market conditions, as much as \$ per cent of the world's supplies are sold on a spot basis.

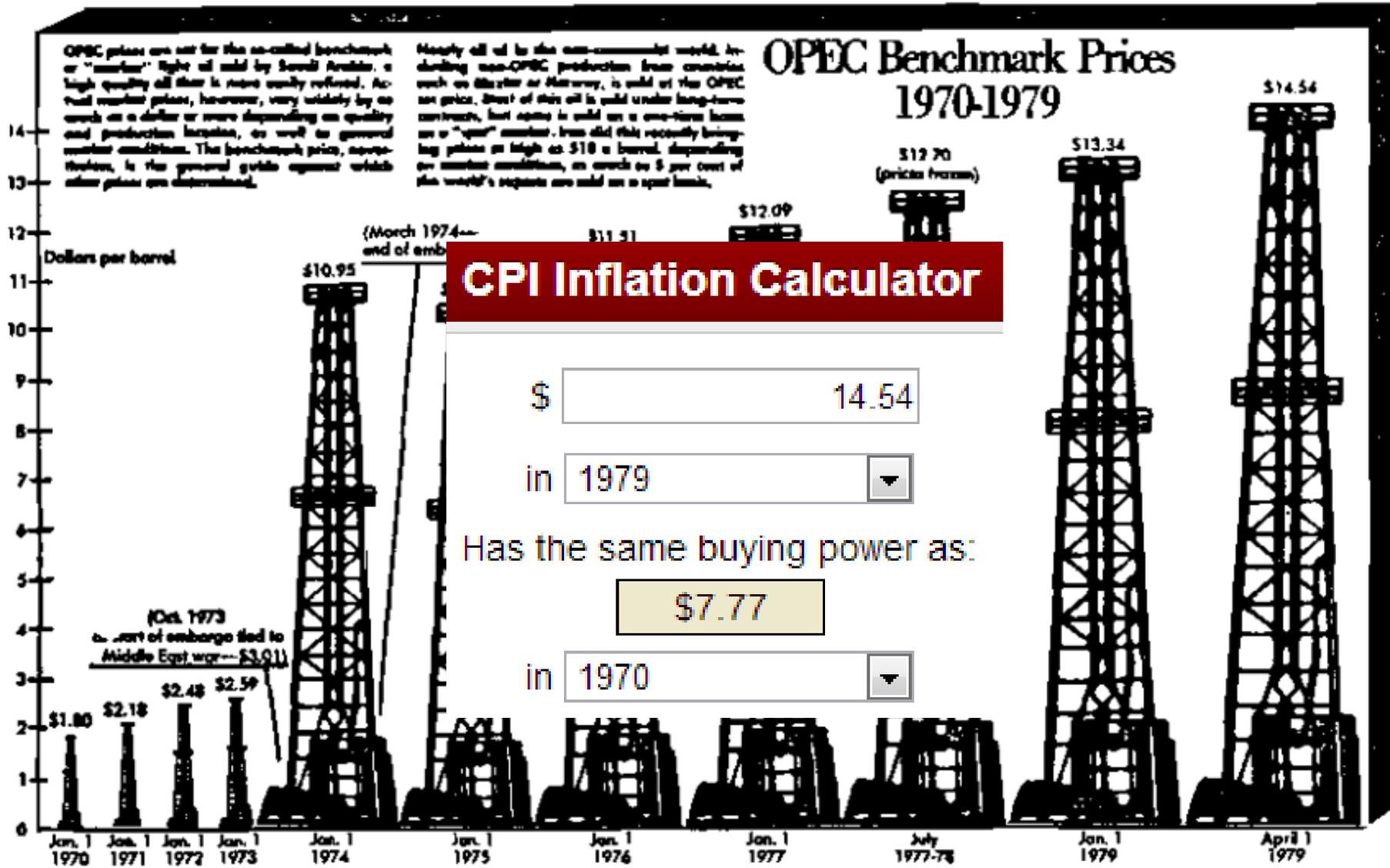


The message is pretty clear, no?

OPEC Benchmark Prices 1970-1979

OPEC prices are set for the so-called benchmark or "marker" light oil sold by Saudi Arabia, a high quality oil that is more widely refined. Actual market prices, however, vary widely by as much as a dollar or more depending on quality and production location, as well as general market conditions. The benchmark price, nevertheless, is the general guide against which other prices are determined.

Nearly all of the non-communist world, including non-OPEC production from countries such as Alaska or Norway, is sold at the OPEC set price. Most of this oil is sold under long-term contracts, but some is sold on a one-time basis as a "spot" market. Iran did this recently bringing prices as high as \$18 a barrel, depending on market conditions, as much as \$ per cent of the world's supplies are sold on a spot basis.



CPI Inflation Calculator

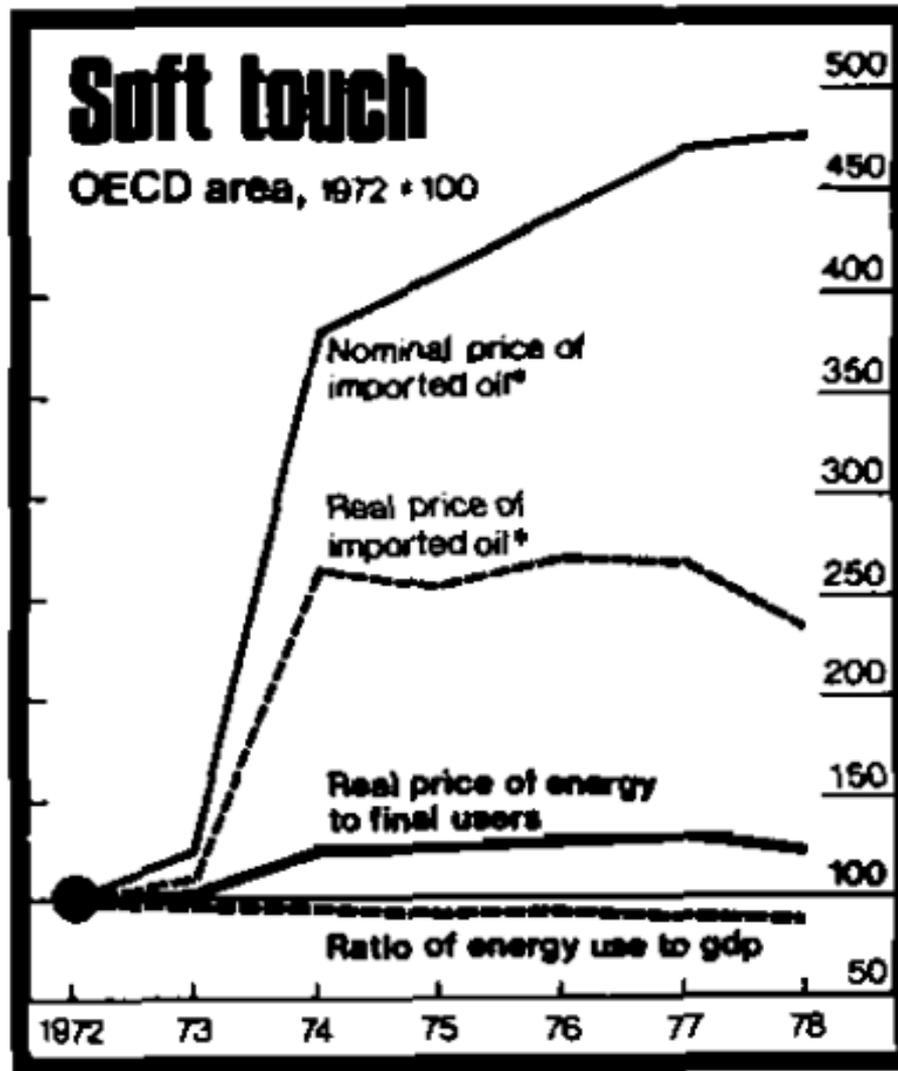
\$

in 1979

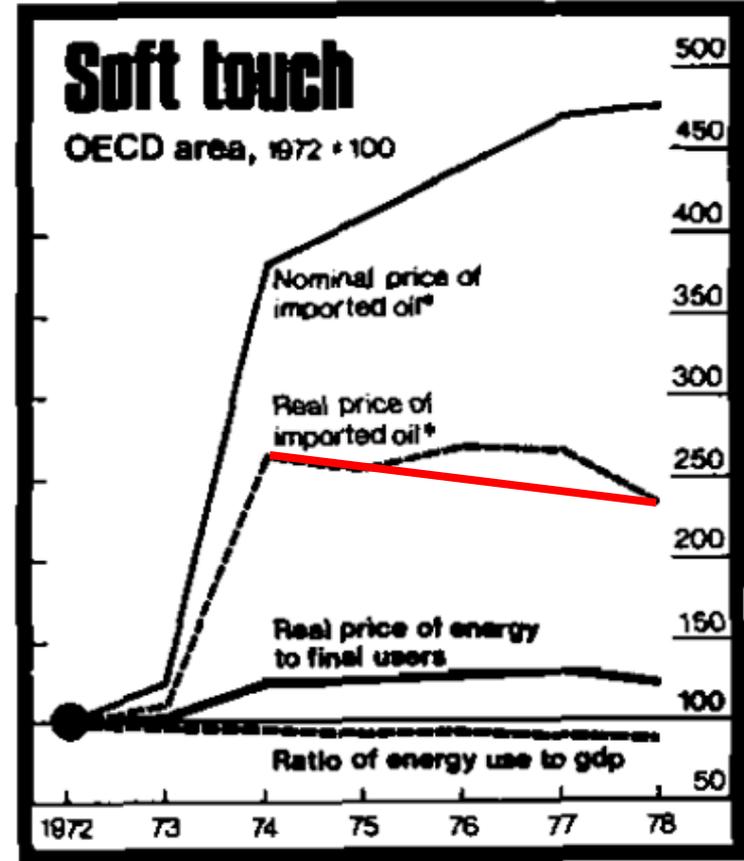
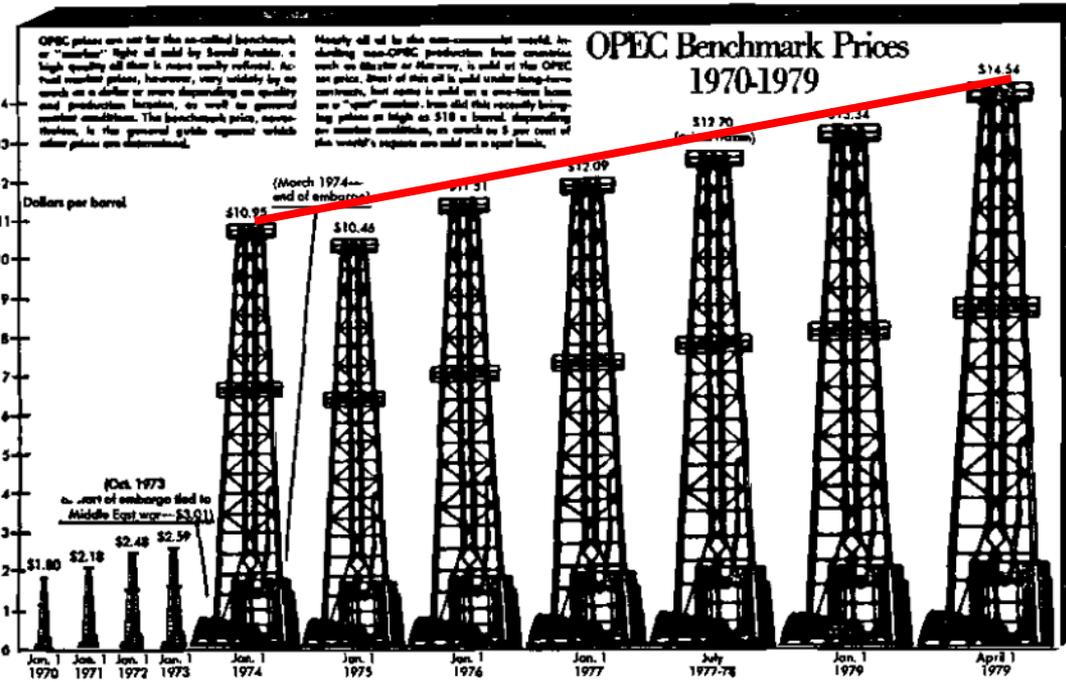
Has the same buying power as:

in 1970

Until you account for inflation.



Here's a more accurate redesigned version.



Design principles with integrity

- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive
- 3) Data should vary, not design
- 4) Displays of money and similar data must be deflated and standardized



Purchasing Power of the Diminishing Dollar
 Source: Labor Department

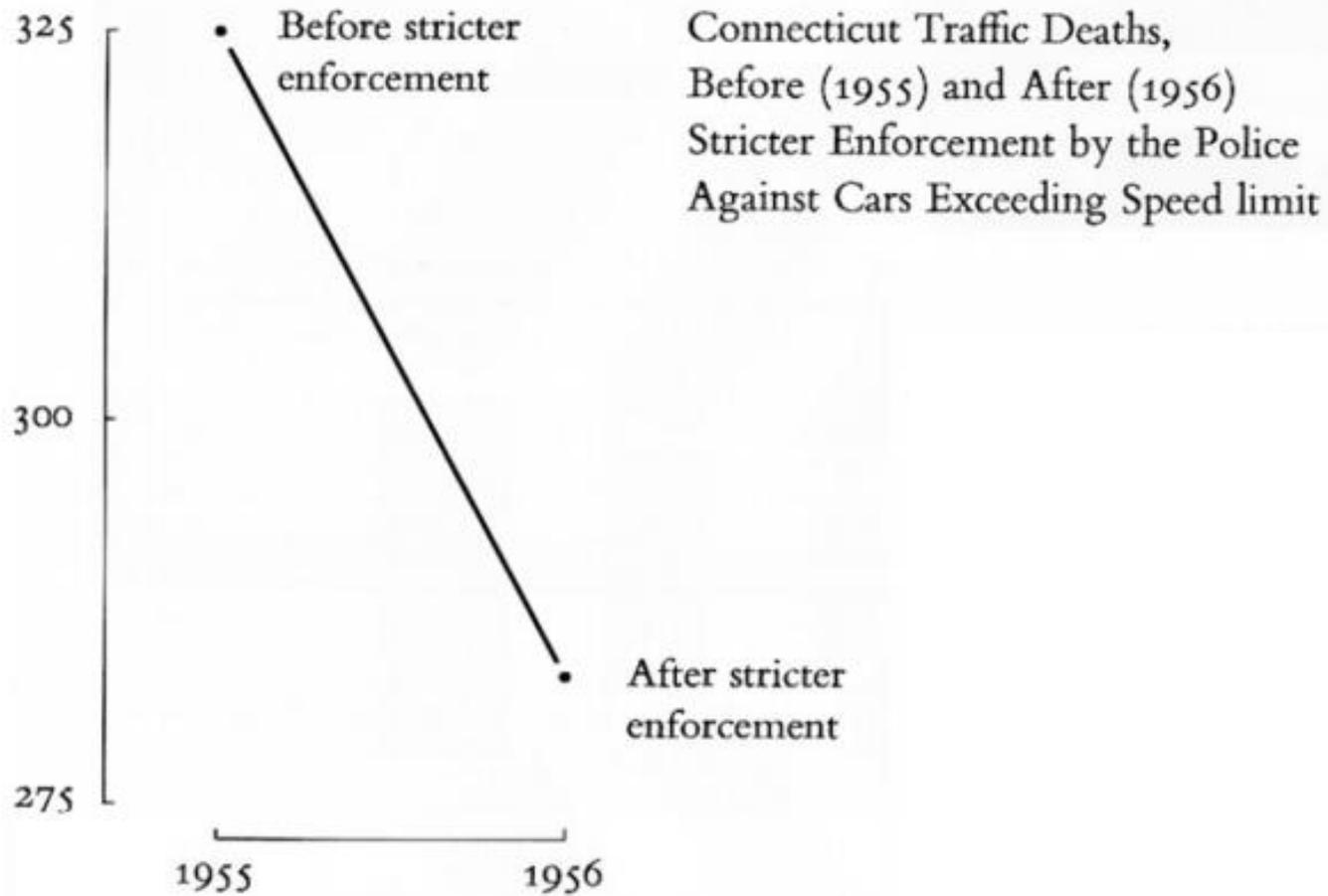
What is this trying to show?



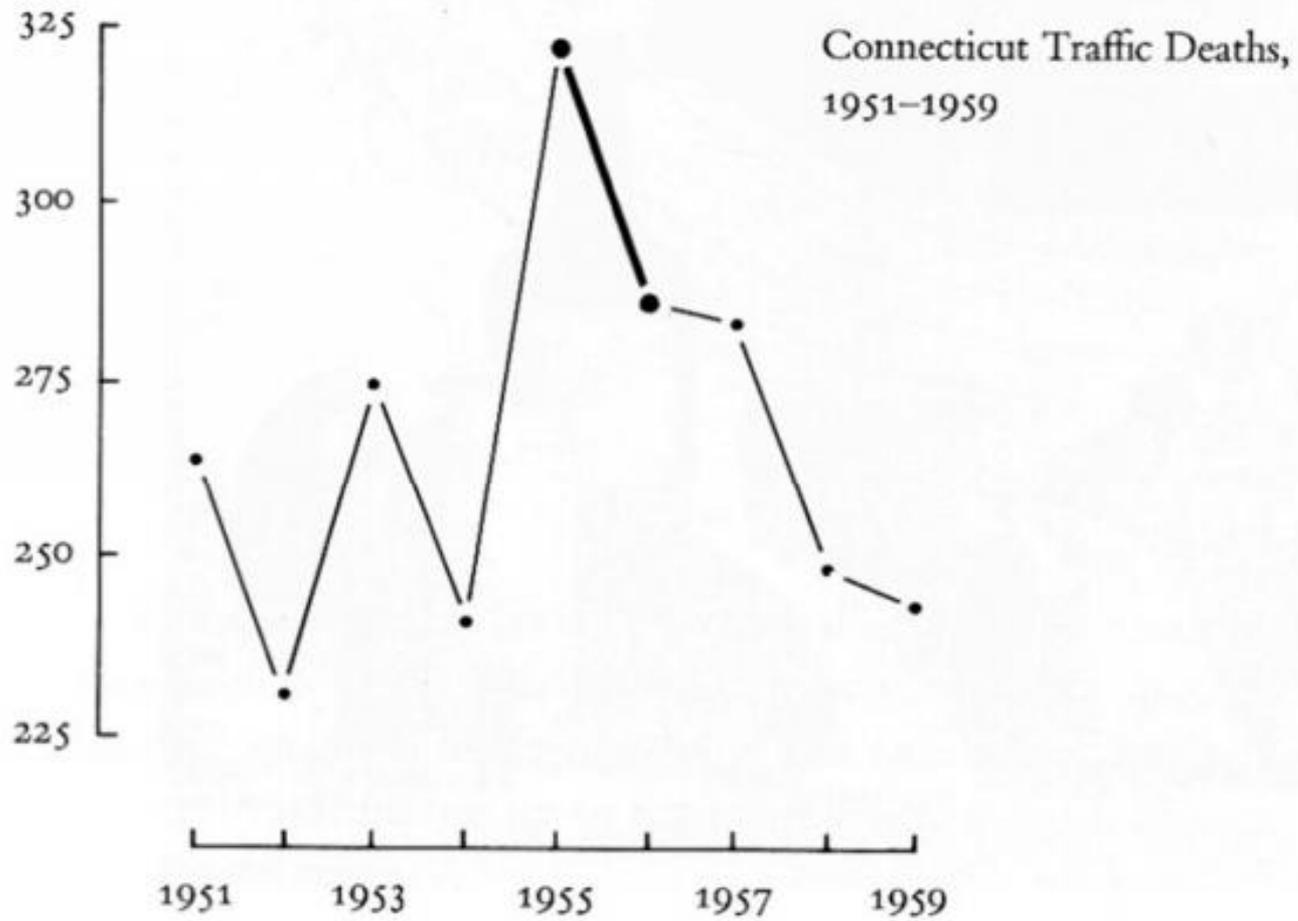
Are we modulating length? Width? Area? All of the above?

Design principles with integrity

- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive
- 3) Data should vary, not design
- 4) Displays of money and similar data must be deflated and standardized
- 5) There should be no more dimensions shown than dimensions in the data



Once again, something is missing...



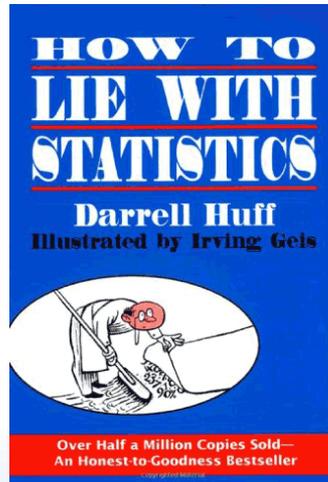
CONTEXT is king.

Design principles with **integrity**

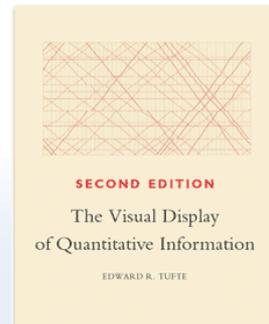
- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive
- 3) Data should vary, not design
- 4) Displays of money and similar data must be deflated and standardized
- 5) There should be no more dimensions shown than dimensions in the data
- 6) Graphics must not quote data out of context

Design principles with integrity

- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive
- 3) Data should vary, not design
- 4) Displays of money and similar data must be deflated and standardized
- 5) There should be no more dimensions shown than dimensions in the data
- 6) Graphics must not quote data out of context

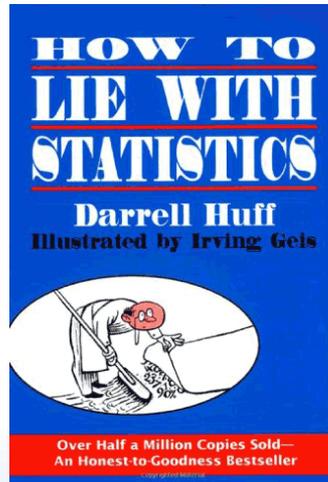


Don't be this guy.



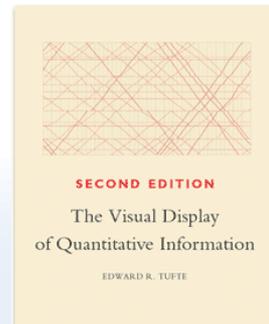
Design principles with integrity

- 1) Proportions in graphics should accurately represent raw data
- 2) Labels should be clear, accurate, and descriptive
- 3) Data should vary, not design
- 4) Displays of money and similar data must be deflated and standardized
- 5) There should be no more dimensions shown than dimensions in the data
- 6) Graphics must not quote data out of context



Don't be this guy.

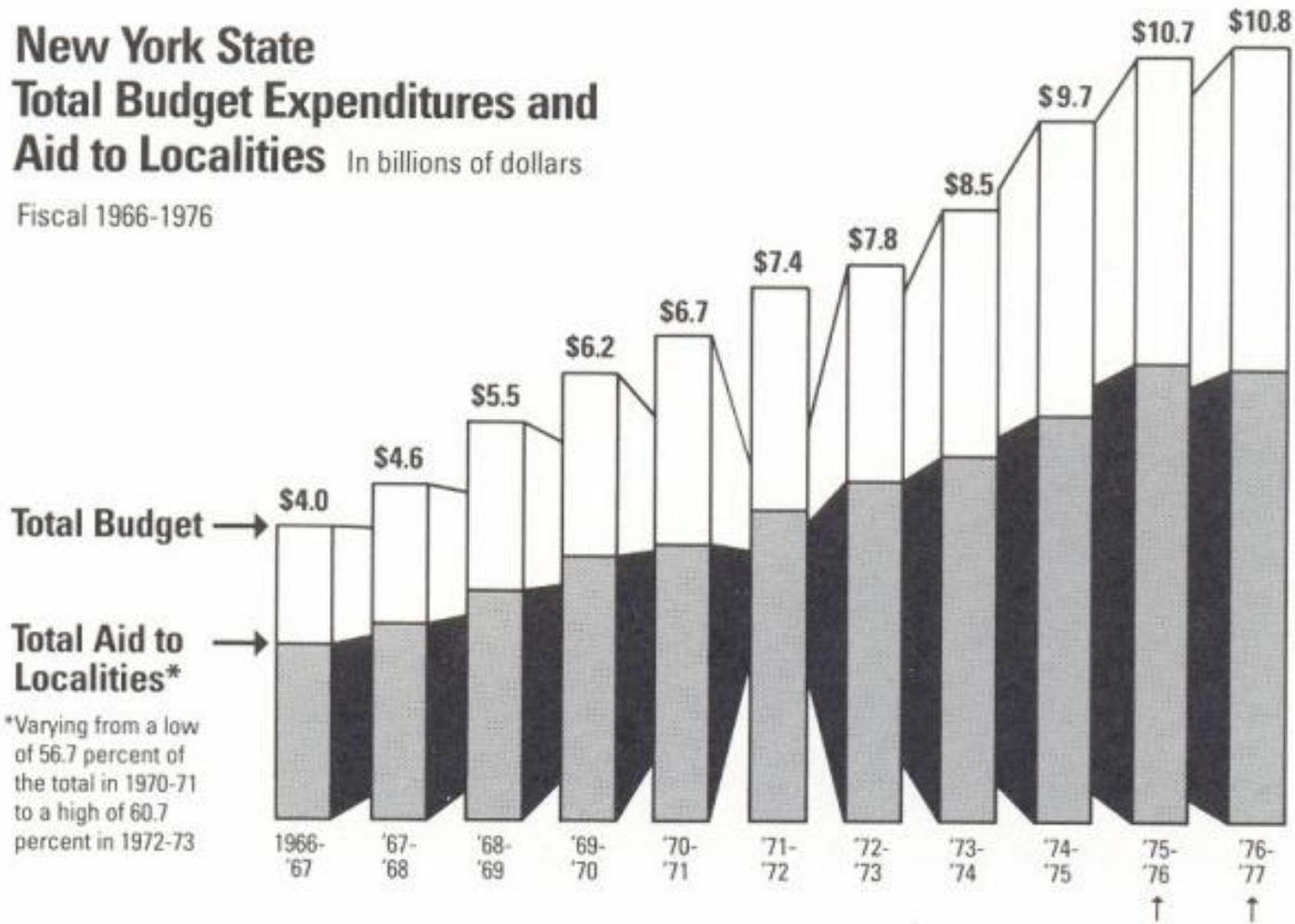
... one more example



New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976

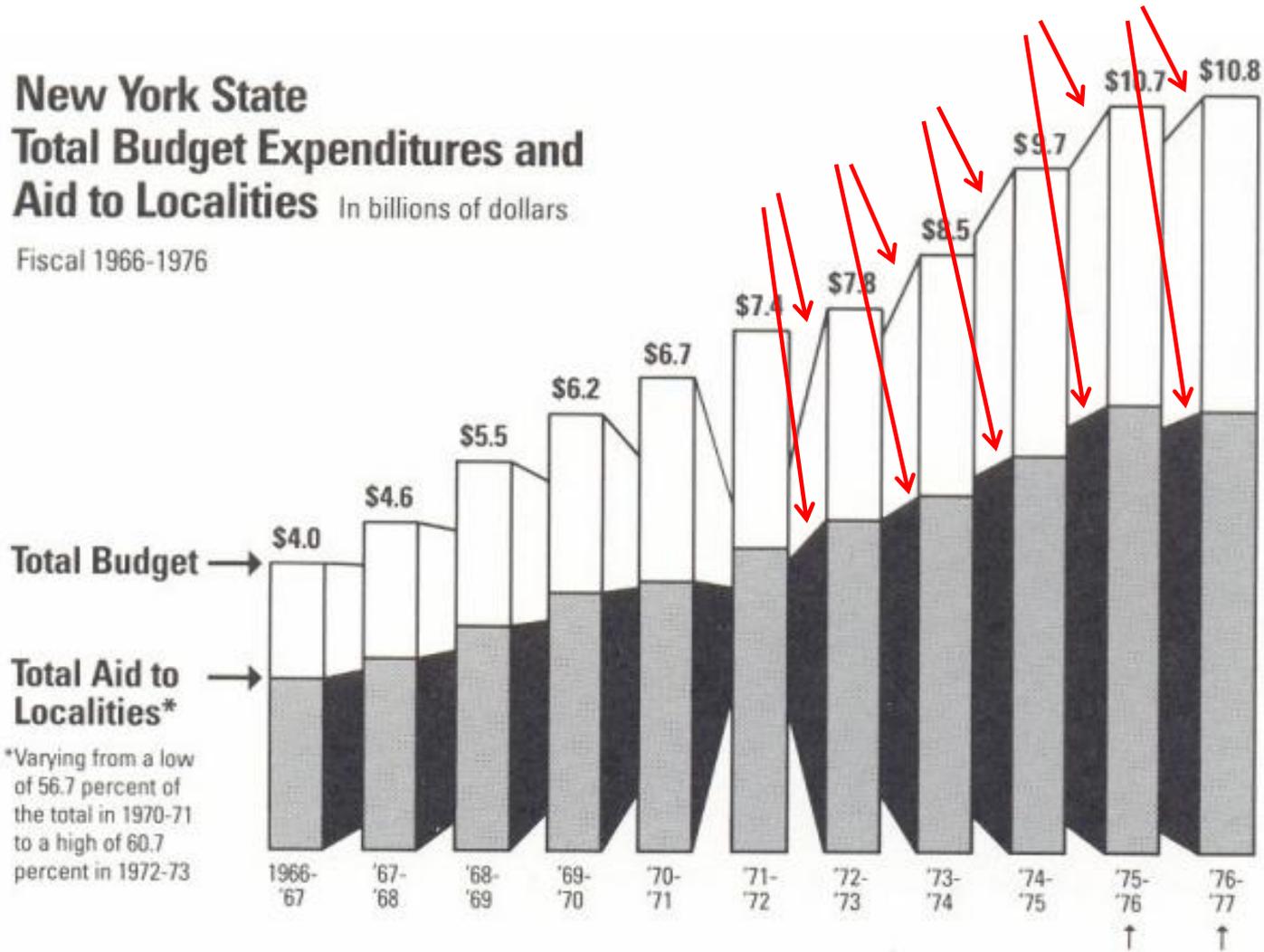


So, where does this graphic turn evil?

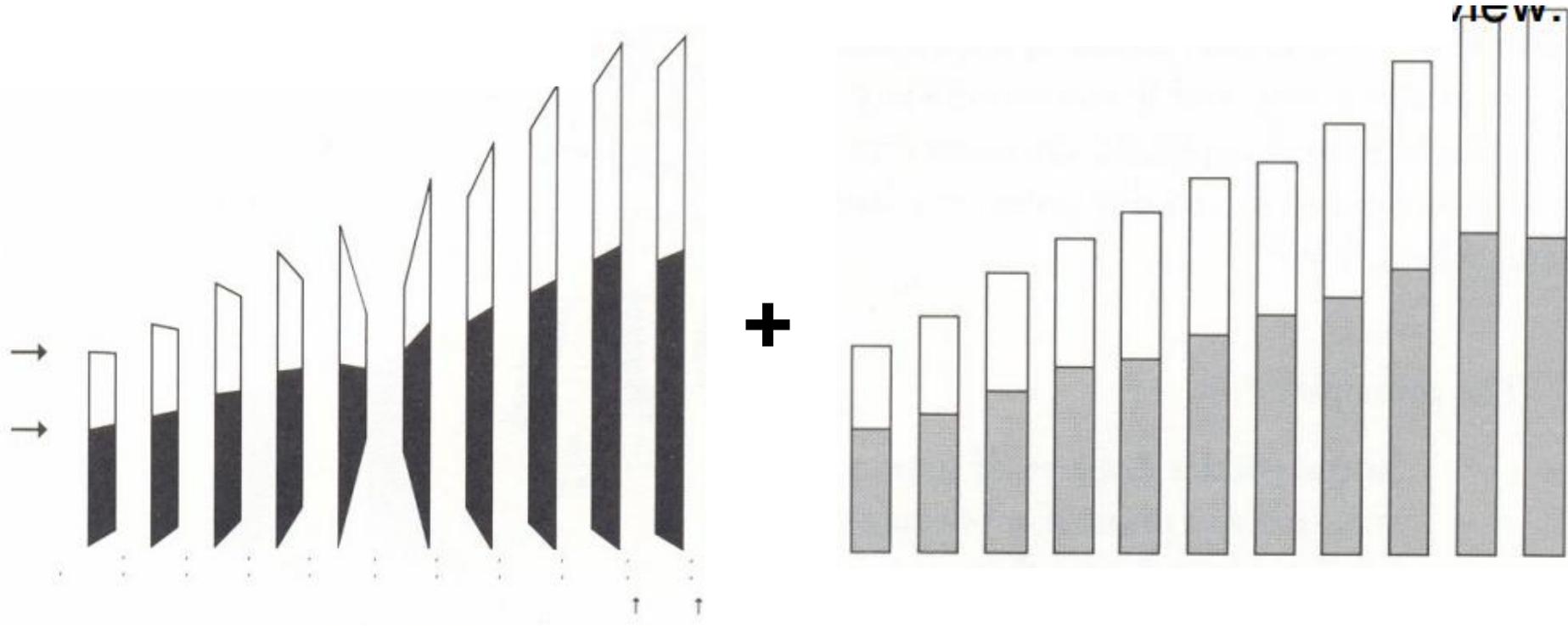
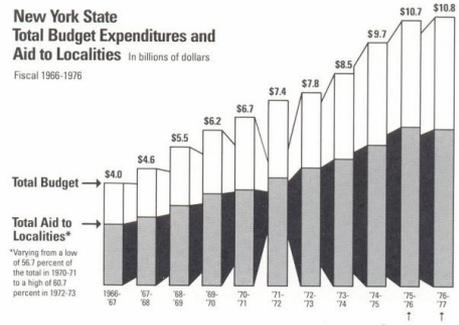
New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

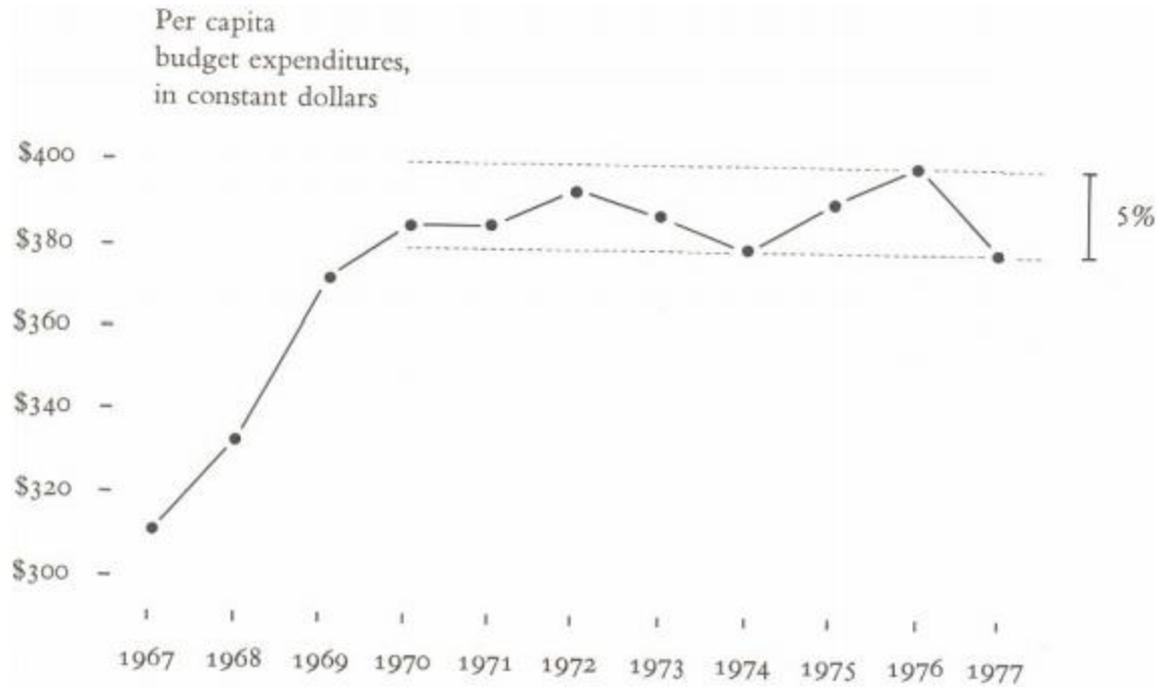
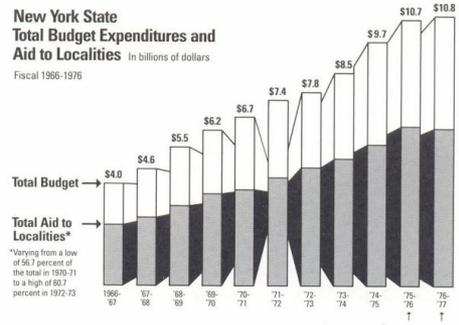
Fiscal 1966-1976



The addition of the third dimension not only makes up data but skews the data that is already there.



Removal of the extra dimensions helps remove the visual distortion. But there's more to the story.



Showing the *real* data, accounting for inflation and population changes makes the information more truthful.

“If statistics are boring then you have the wrong numbers.”

- Edward Tufte



A few pseudo-concrete means of graph evaluation

- 1) Data-Ink Ratio
- 2) Lie Factor
- 3) Graphical Sophistication
- 4) Data Density

Data-Ink Ratio

The ratio of nonremovable
To removable parts of a chart

=

data-ink

total ink used to print the graphic

Data-Ink Ratio

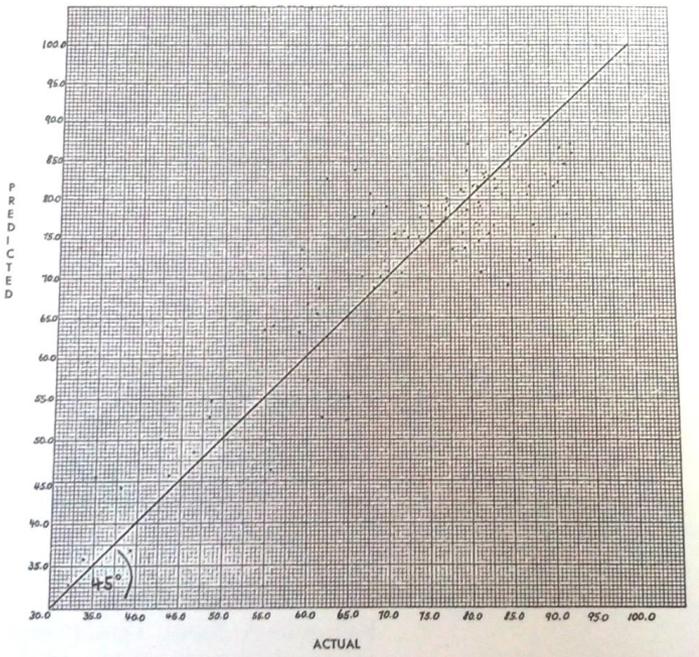
The ratio of nonremovable
To removable parts of a chart

=

data-ink

total ink used to print the graphic

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



< 0.1



1.0

Lie Factor

The discrepancy between
data and representation

=

Size of effect shown in graphic

Size of effect in data

Lie Factor = $\frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$

The discrepancy between data and representation

Also: logarithm of Lie Factor can help determine skew
 $\log(\text{LF}) < 0$: understating errors $\log(\text{LF}) > 0$: overstating errors

Lie Factor

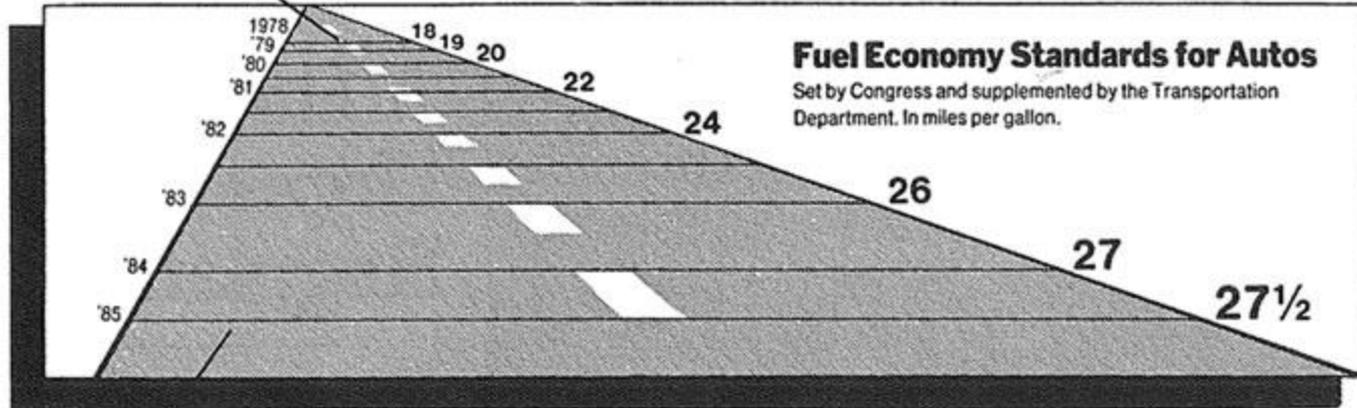
The discrepancy between data and representation

=

Size of effect shown in graphic

Size of effect in data

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.

Lie Factor

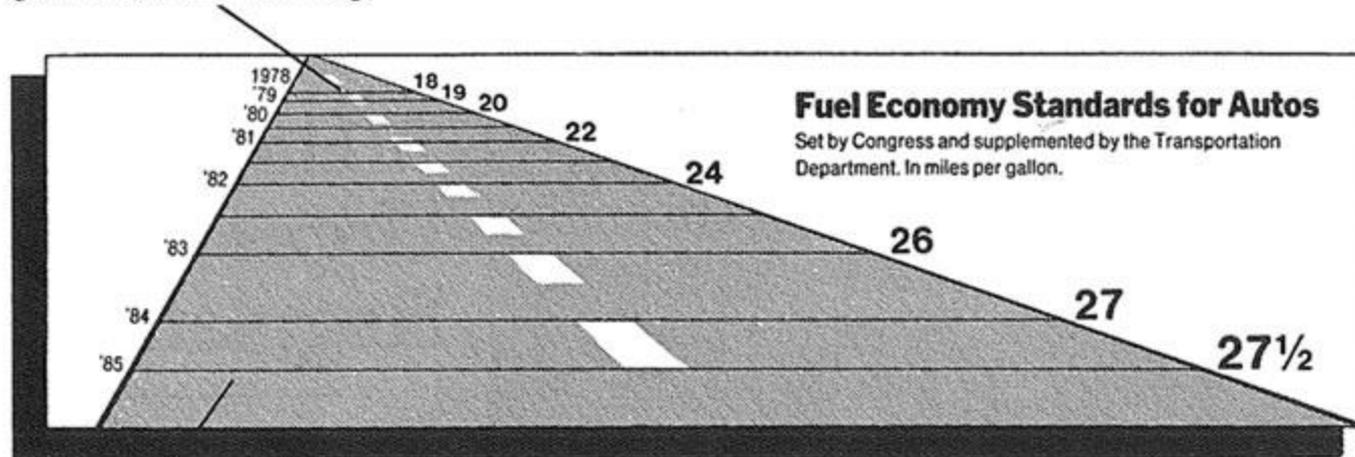
The discrepancy between data and representation

=

Size of effect shown in graphic

Size of effect in data

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.

Change in graphic: $(5.3 - 0.6) / 0.6$ 783.3%

Change in data: $(27.5 - 18) / 18$ 52.7%

Lie Factor

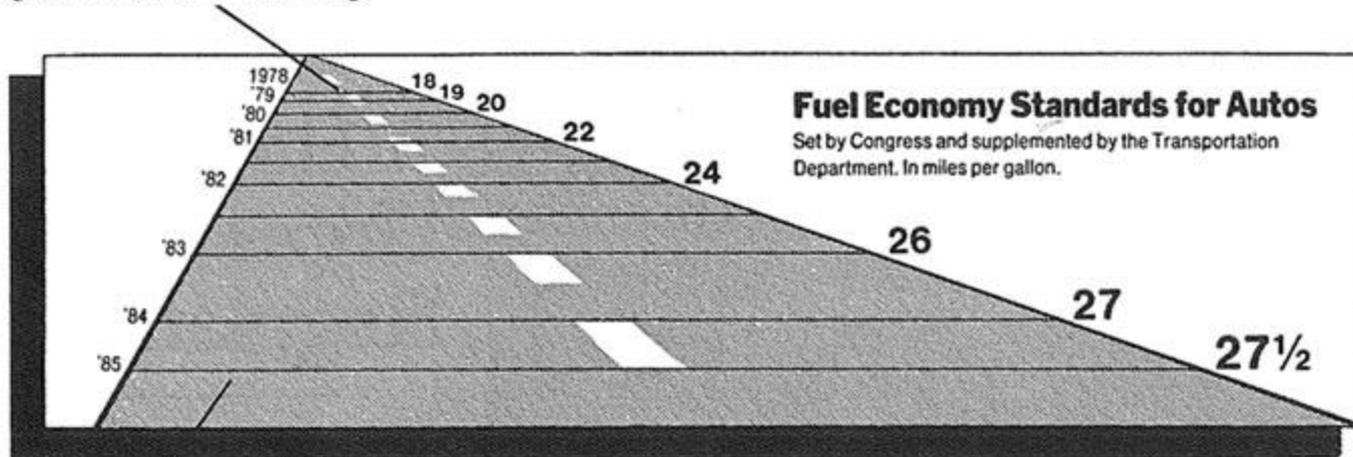
The discrepancy between data and representation

=

Size of effect shown in graphic

Size of effect in data

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.

Change in graphic:	$(5.3 - 0.6) / 0.6$	783.3%
Change in data:	$(27.5 - 18) / 18$	52.7%
Lie Factor:	$(7.83 / 0.52)$	14.8
Skew:	$\log(14.8)$	1.17

Lies via overstatement

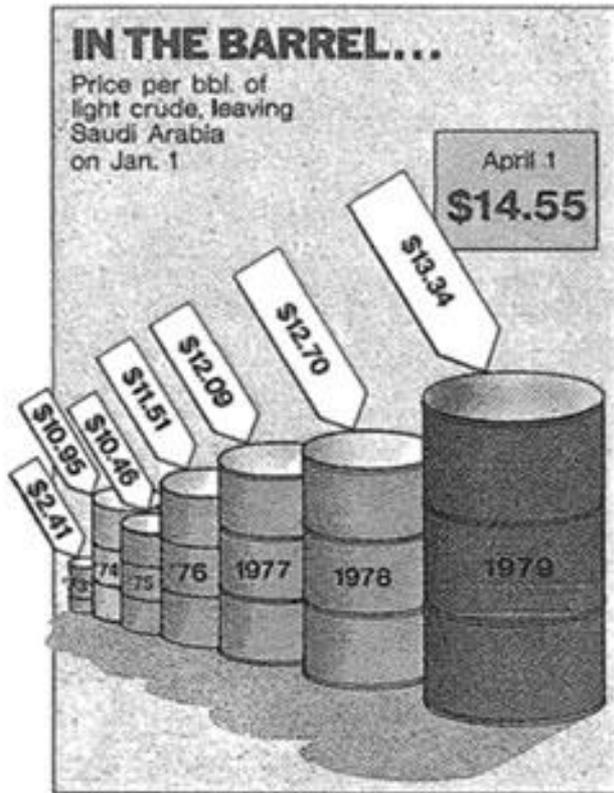
Lie Factor

The discrepancy between data and representation

=

Size of effect shown in graphic

Size of effect in data



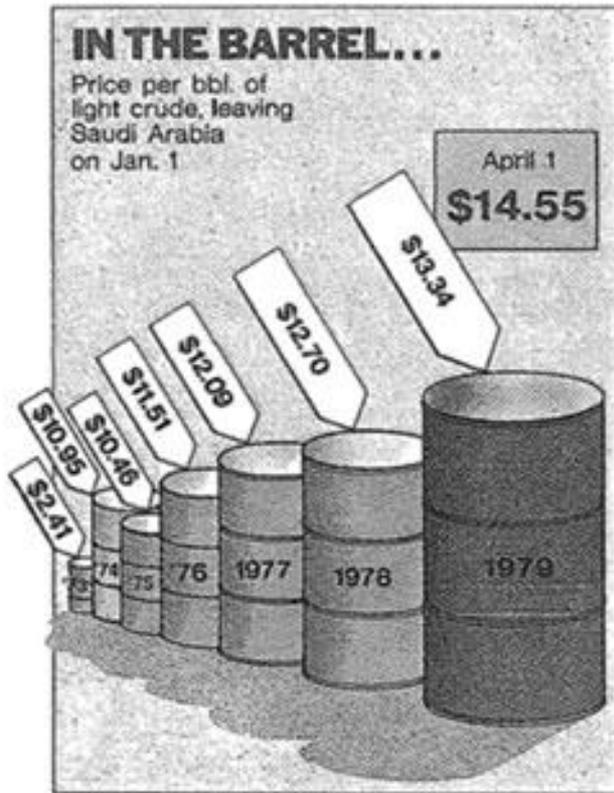
Lie Factor

The discrepancy between data and representation

=

Size of effect shown in graphic

Size of effect in data



Data increase from '73 to '79: 454%
Size (one dimension) increase: 4267.6%
Size (volume) increase: 27,000%

Overall lie factor: 59.4

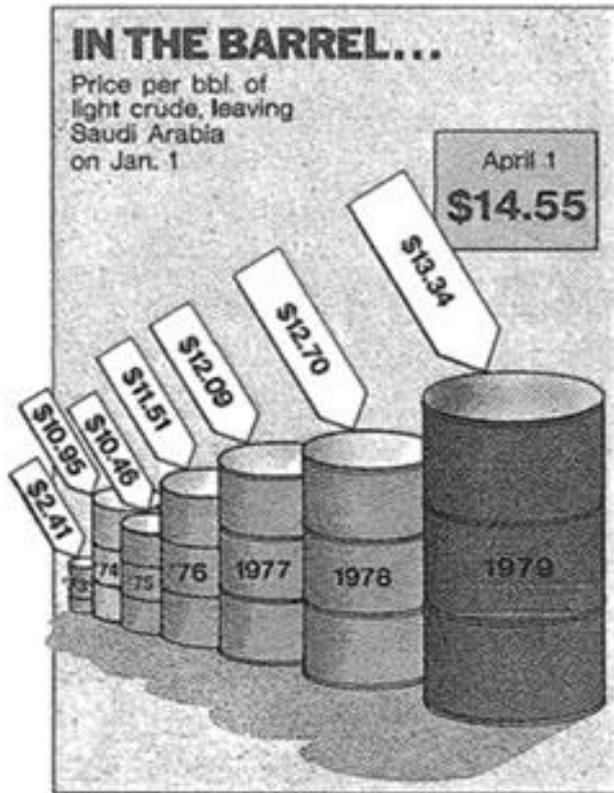
Lie Factor

The discrepancy between data and representation

=

Size of effect shown in graphic

Size of effect in data



Sophistication

Proportion of relational graphs to non-relational and time-series

=

Number of statistical graphics based upon more than one variable and not a time-series

Number of graphics in sample

Sophistication

Proportion of relational graphs to non-relational and time-series

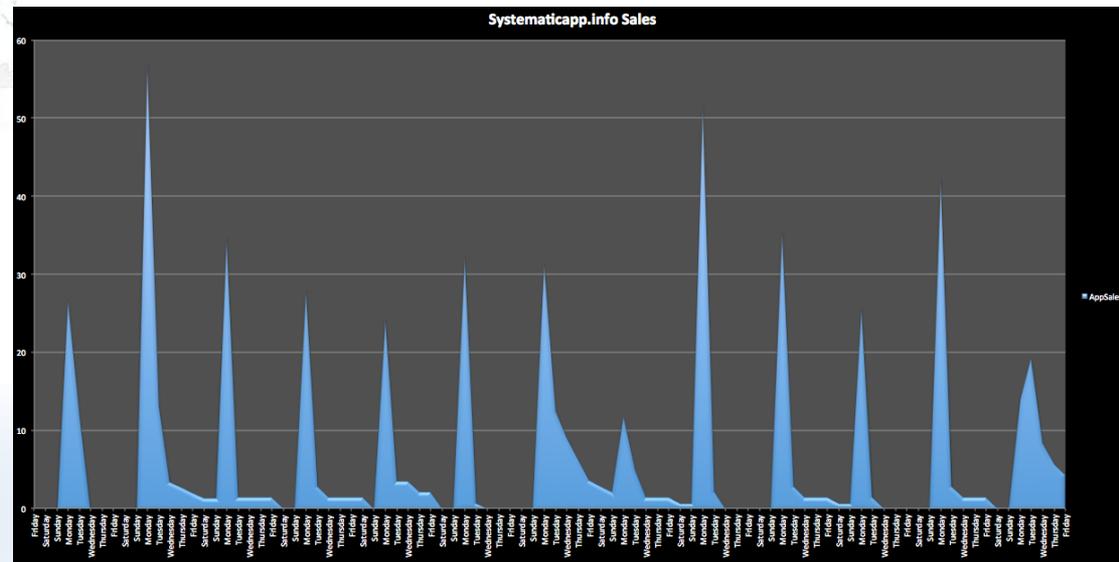
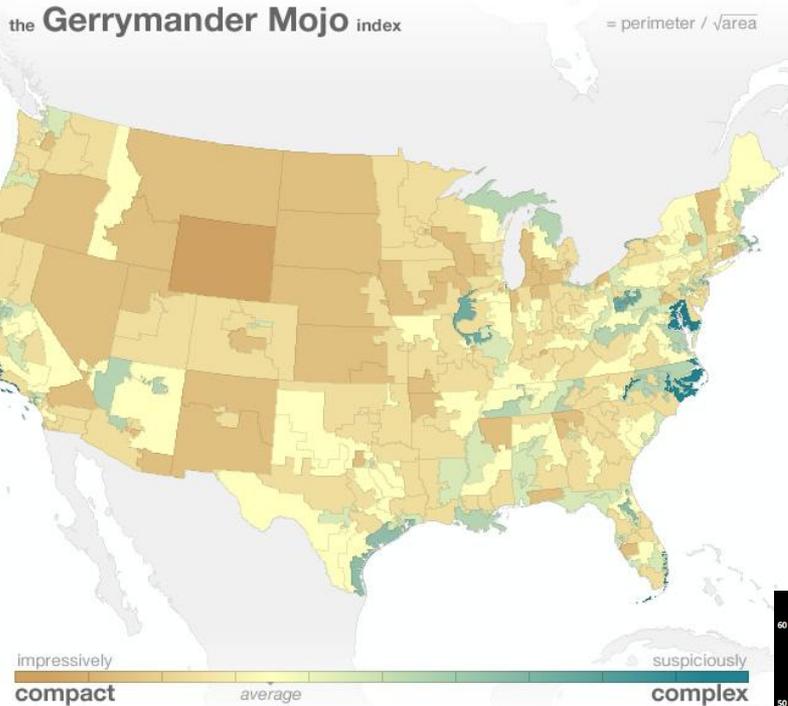
=

Number of statistical graphics based upon more than one variable and not a time-series

Number of graphics in sample

Which of these communicates interesting information?

VS



Sophistication

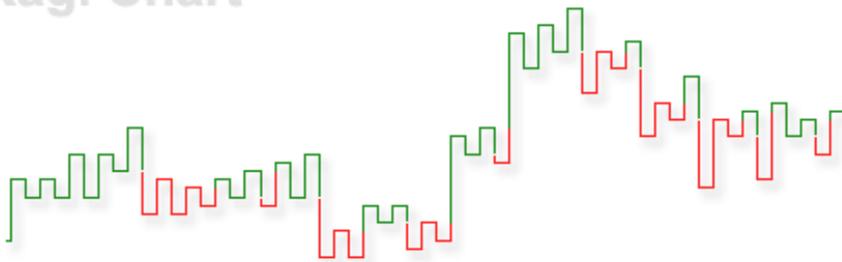
Proportion of relational graphs to non-relational and time-series

=

Number of statistical graphics based upon more than one variable and not a time-series

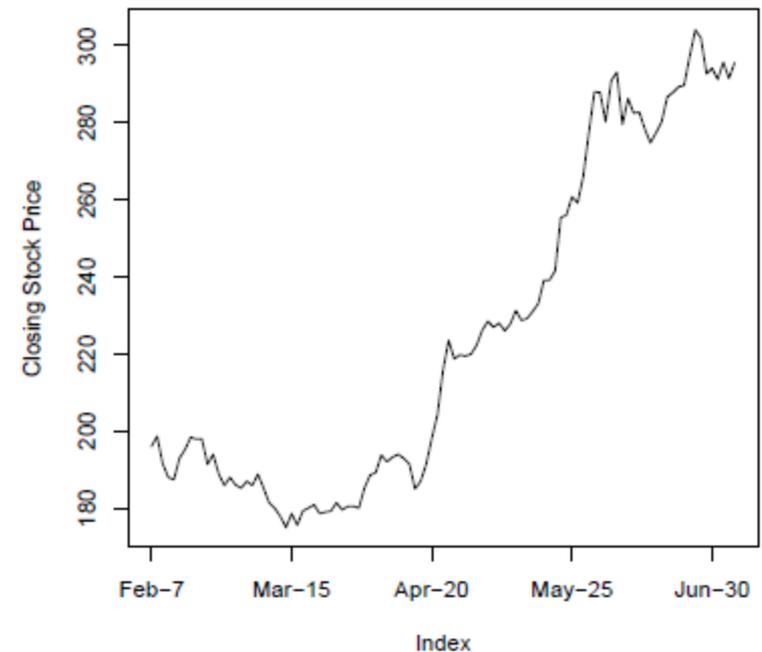
Number of graphics in sample

Kagi Chart



VS

Google Stock Price



Which of these serves a clearer purpose?

Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display

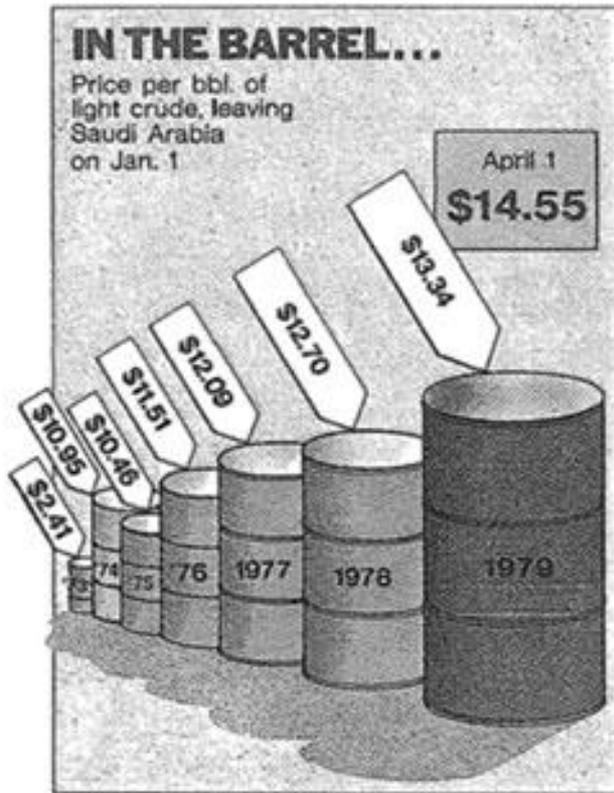
Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display



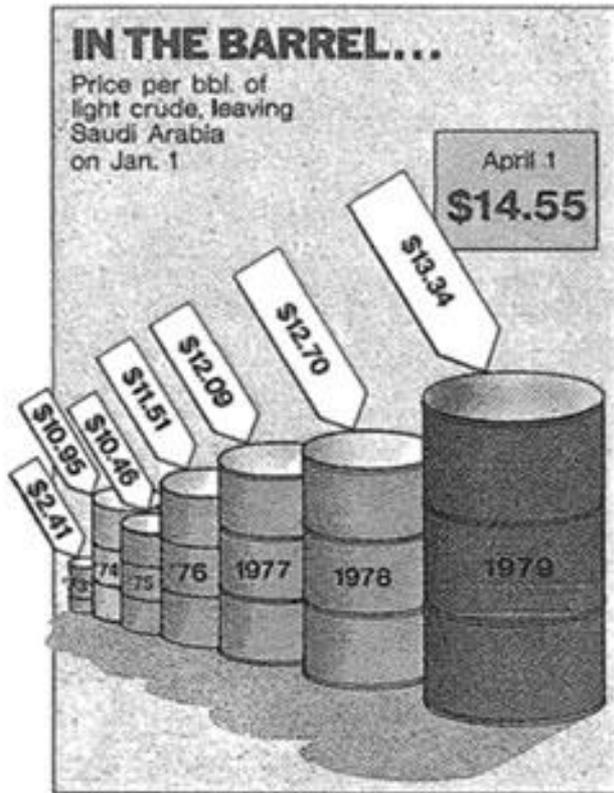
Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display



Number of numbers: 15

Approximate size of graphic: 7.5 sq in.
48.4 sq cm.

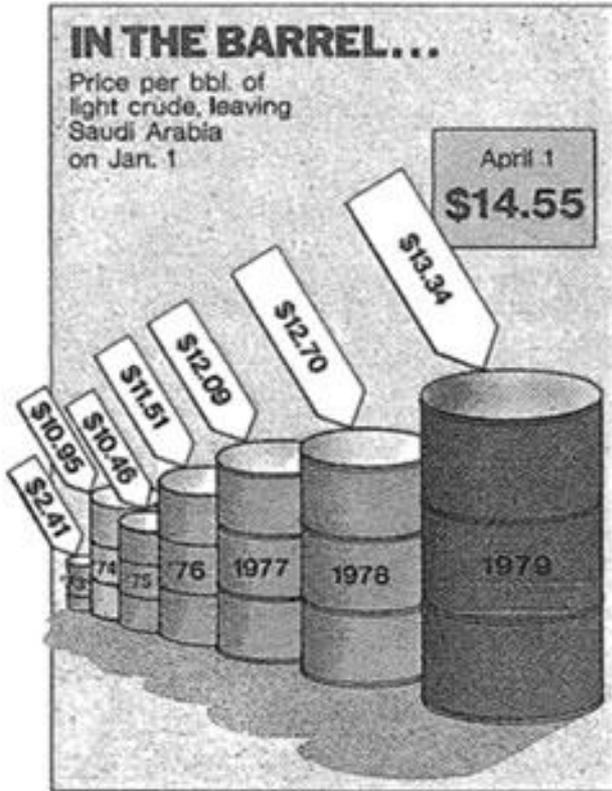
Data Density

The proportion of data to useful chart area

=

Number of entries in data matrix

Area of data display



Number of numbers: 15
Approximate size of graphic: 7.5 sq in.
48.4 sq cm.

Data Density (in): 2 val / sq in
0.31 val / sq cm

WEAK.

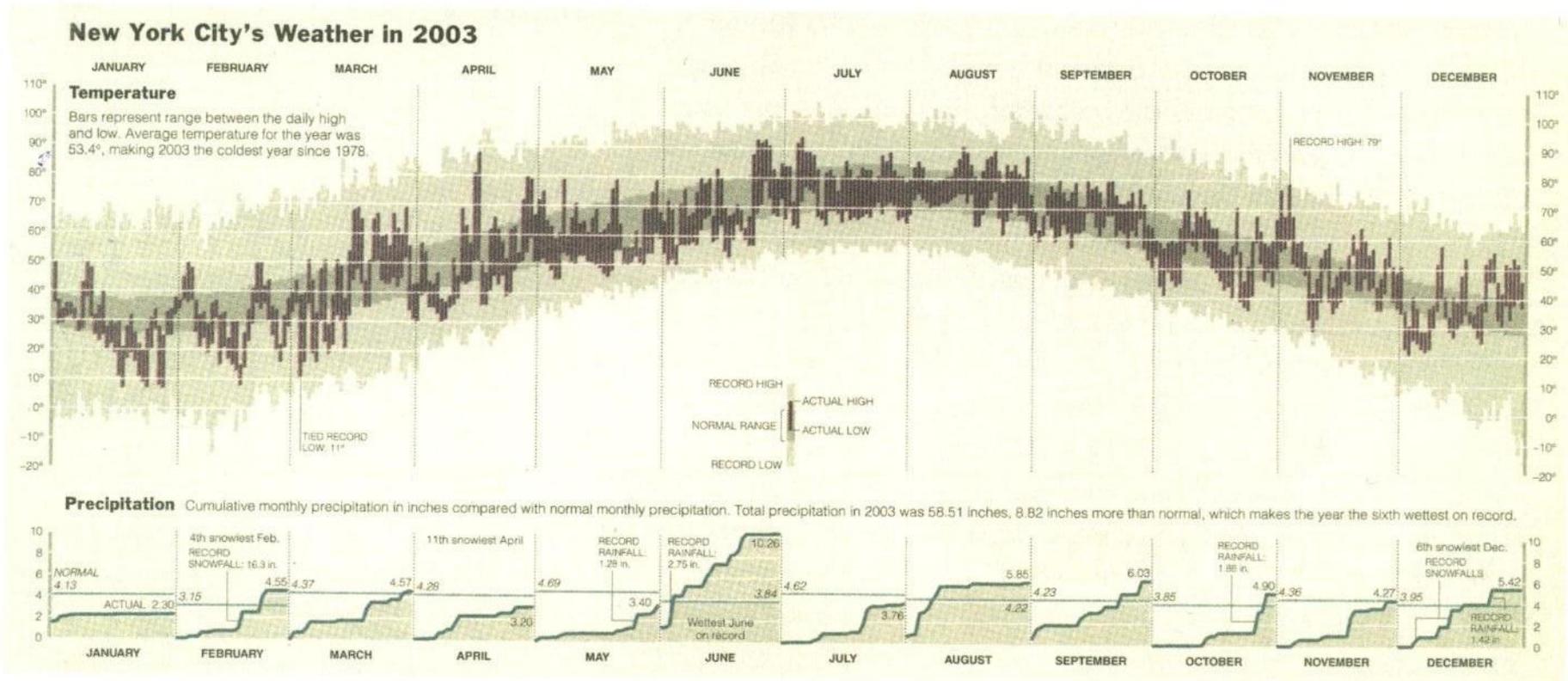
Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display



New York Times, January 4, 2004, A15.

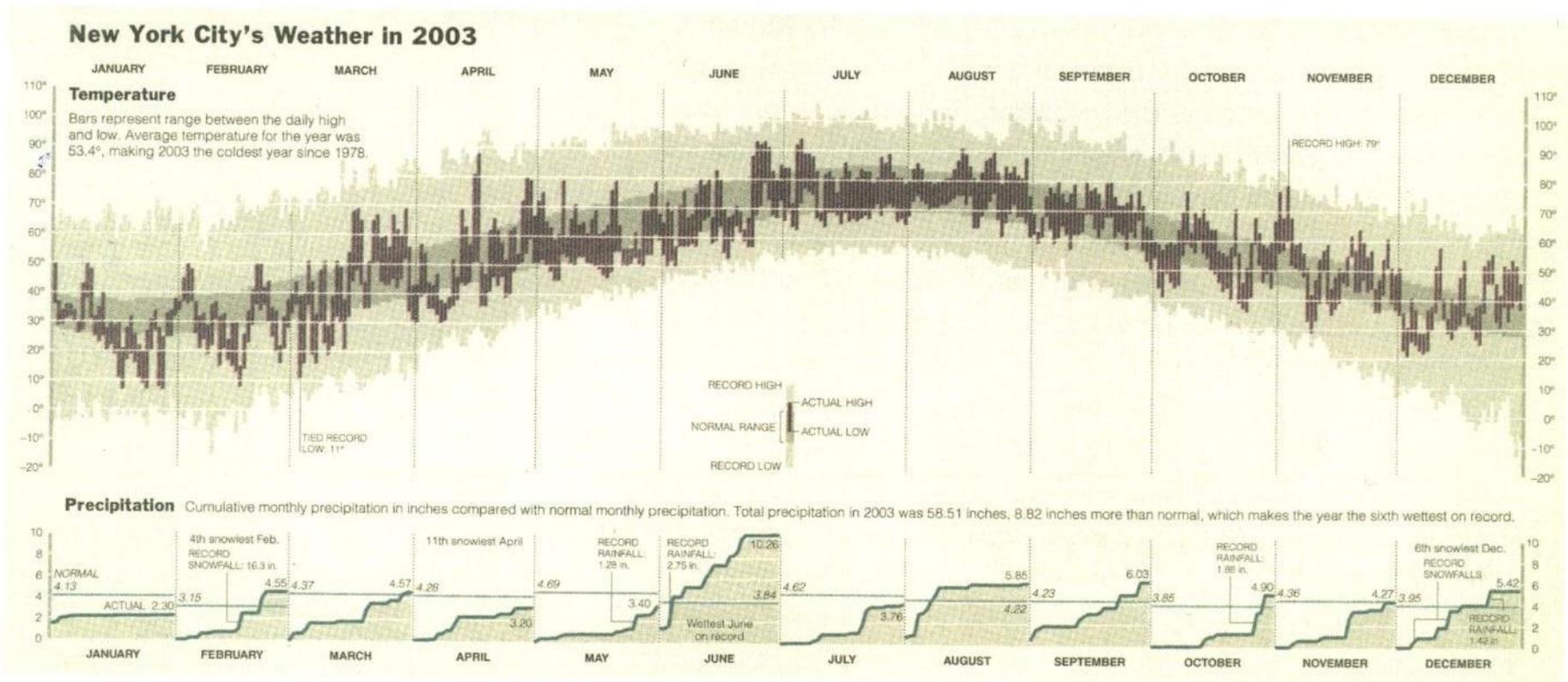
Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display



New York Times, January 4, 2004, A15.

300 values / sq in
45 values / sq cm

Graphically excellent.

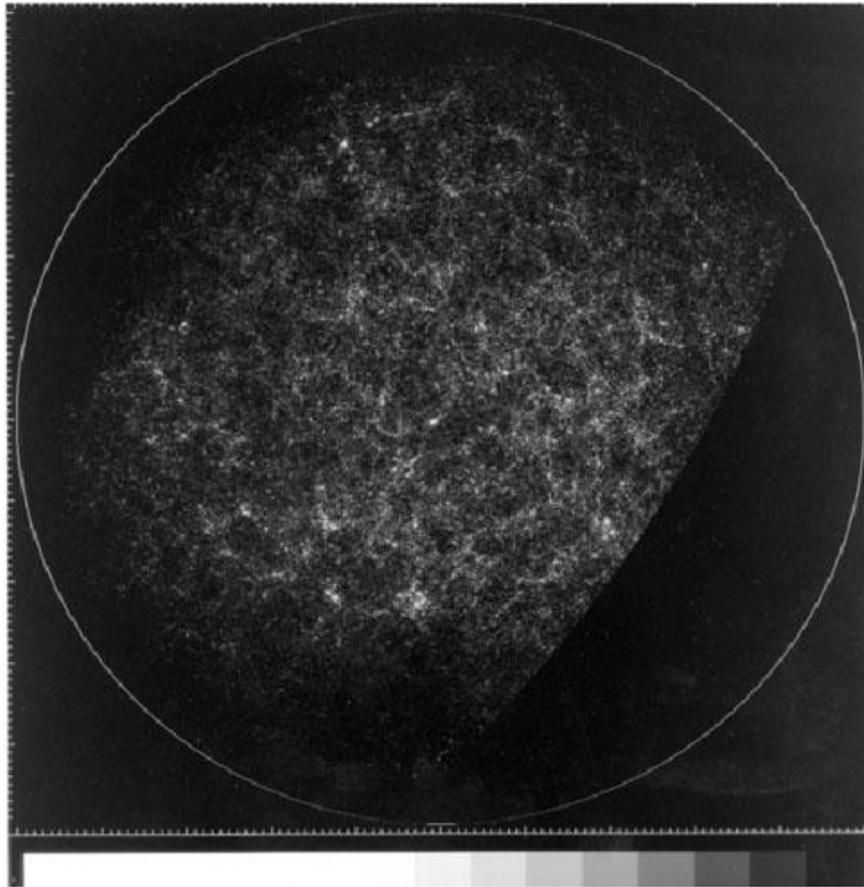
Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display



Map of the galaxies
(Harvard Astrophysics Data System)

2,275,328 encoded rectangles
(X, Y, and Color)

61 square inch surface

110,000 values / sq in
17,000 values / sq cm

Data Density

The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display

Consider the [Small Multiple](#) to help get your ideas across

Data Density

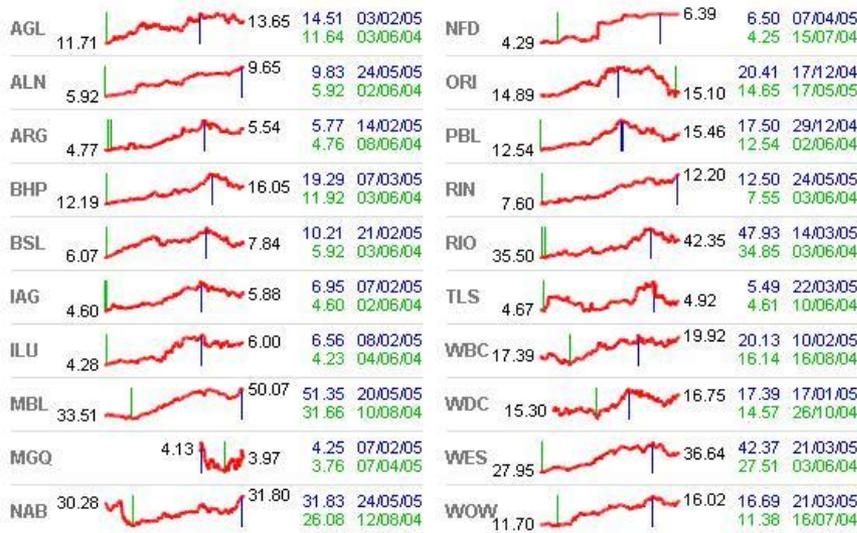
The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display

Consider the **Small Multiple** to help get your ideas across



Data Density

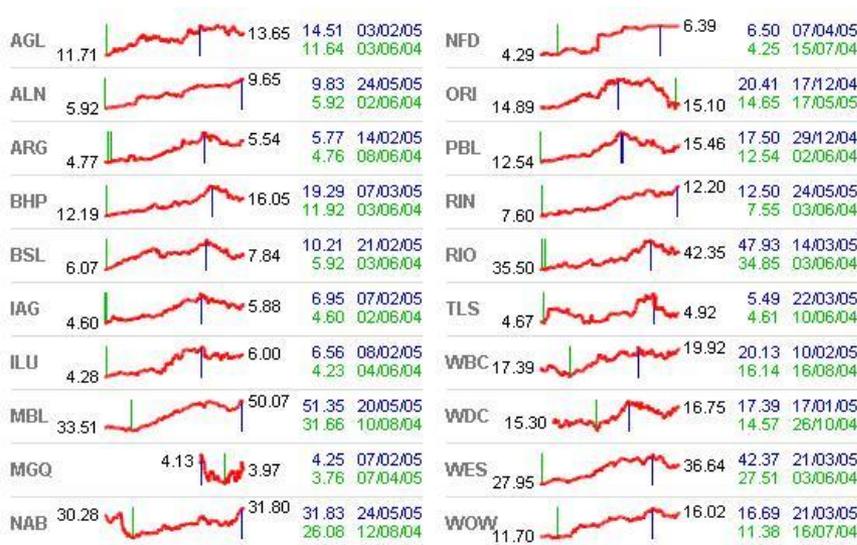
The proportion of data to
useful chart area

=

Number of entries in data matrix

Area of data display

Consider the **Small Multiple** to help get your ideas across



2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support



Orange and green colors correspond to states where support for vouchers was greater or less than the national average. The seven ethnic/religious categories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants. Where a category represents less than 1% of the voters of a state, the state is left blank.

A few pseudo-concrete means of graph evaluation

- 1) Data-Ink Ratio
- 2) Lie Factor
- 3) Graphical Sophistication
- 4) Data Density

Like all things in design, there is no formula for success.
Knowing good from bad is a great place to start.

The way you modulate objects in a visualization also has a profound impact on readability and integrity.

The way you modulate objects in a visualization also has a profound impact on readability and integrity.

Let's see some examples.

What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?



A change in position on the y-axis!

What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?



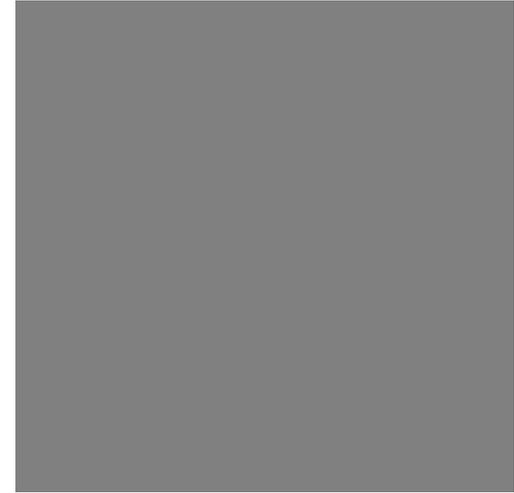
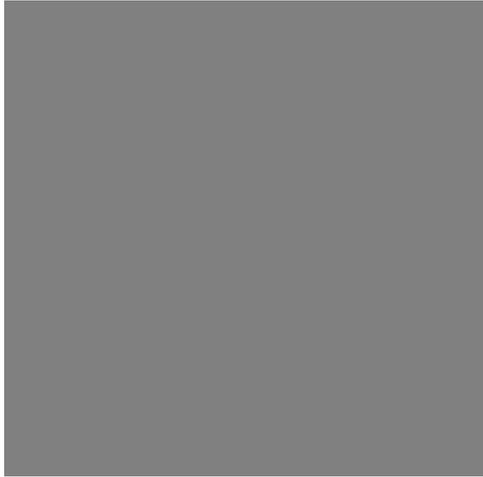
What, if anything, is the difference between the following objects?



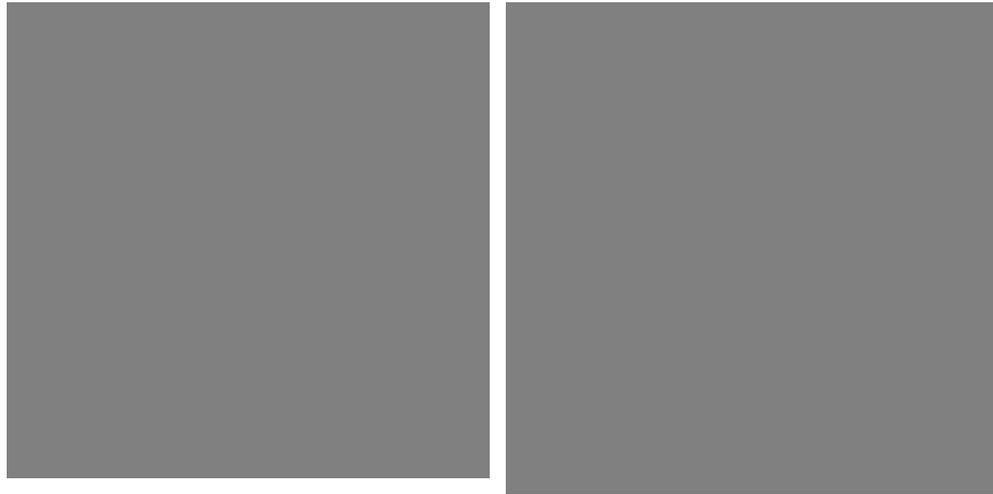
What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?



What, if anything, is the difference between the following objects?

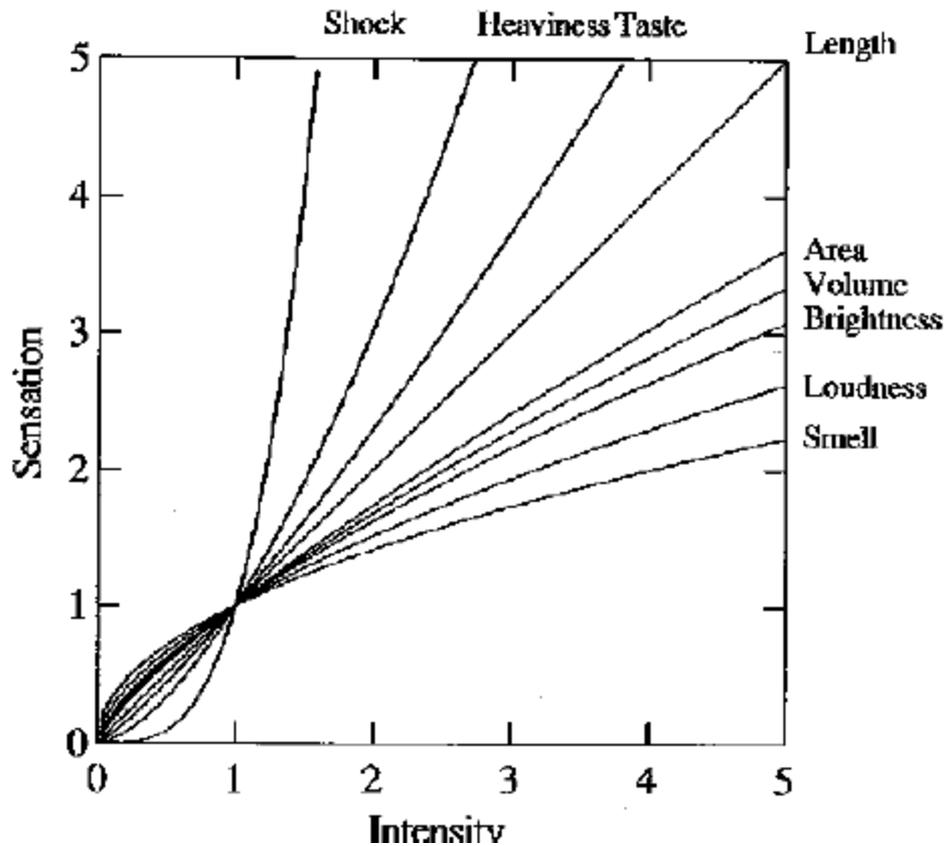


What, if anything, is the difference between the following objects?



Perception of a sensation, while not universal, can be ranked.

Perception of a sensation, while not universal, can be ranked.



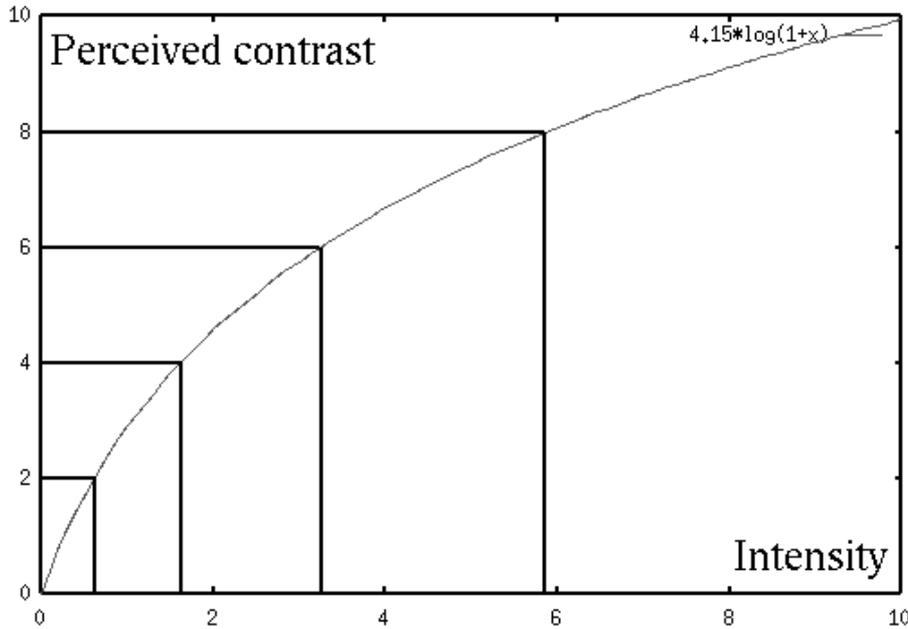
Stephen's Power Law

$$S = I^p$$

$p < 0$: underestimate

$p > 0$: overestimate

Perception of a sensation, while not universal, can be ranked.



Weber-Fechner Law

$$dS = k(dI/I)$$

Or, in English:
Magnitude is more poorly
perceived than ratios

The way you modulate objects in a visualization also has a profound impact on readability and integrity.

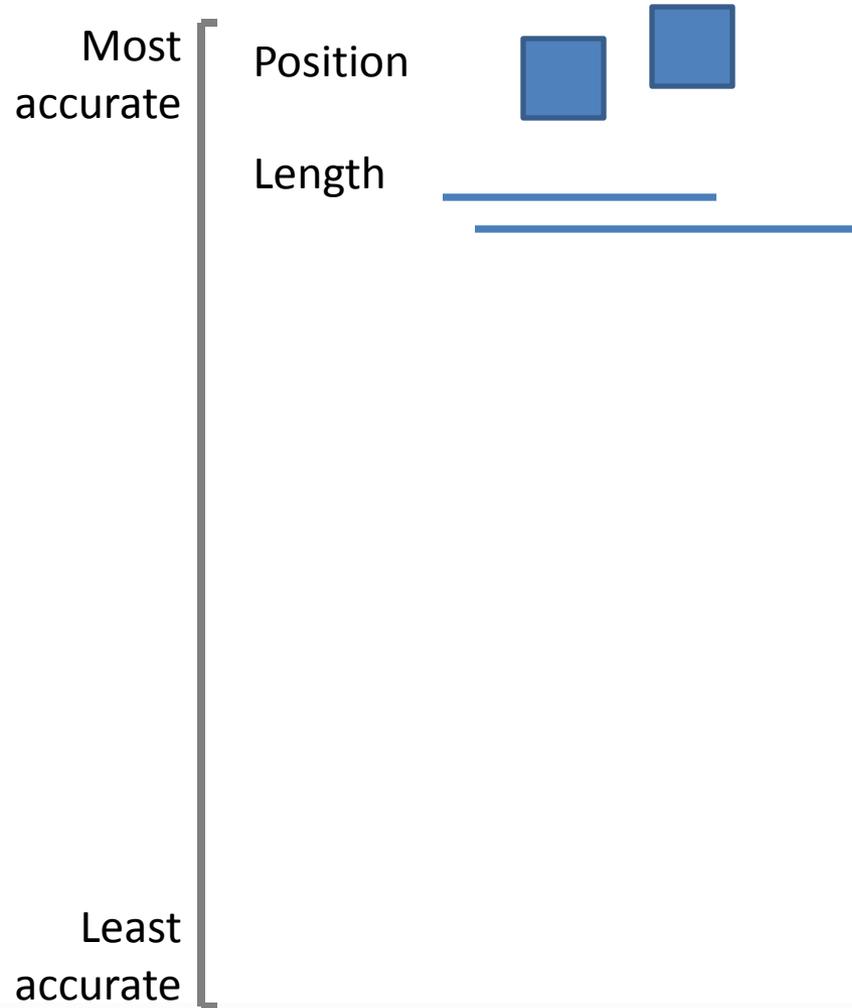
Most
accurate

Least
accurate

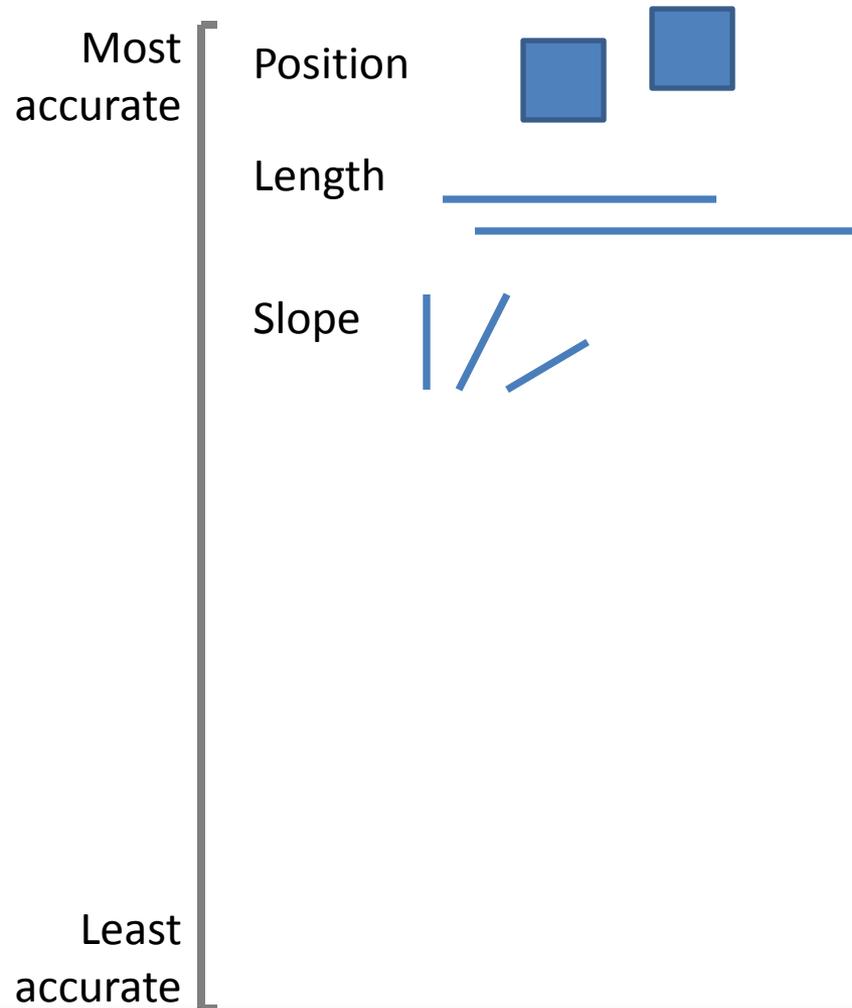
The way you modulate objects in a visualization also has a profound impact on readability and integrity.



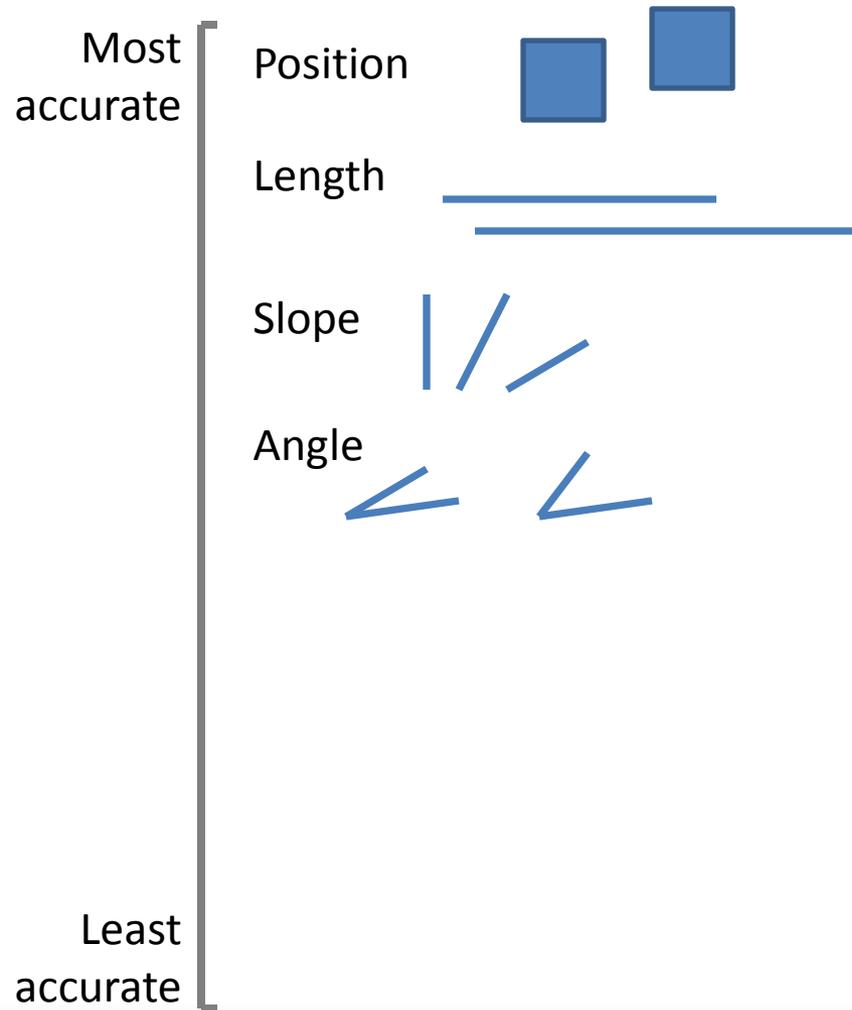
The way you modulate objects in a visualization also has a profound impact on readability and integrity.



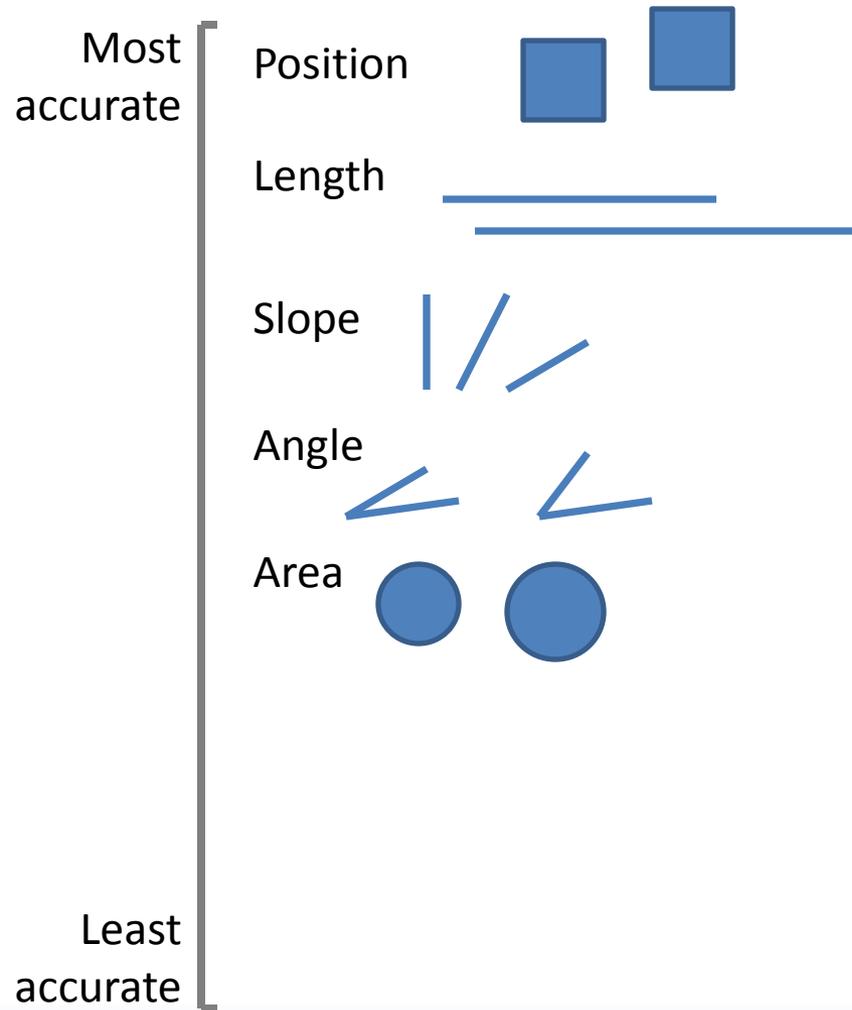
The way you modulate objects in a visualization also has a profound impact on readability and integrity.



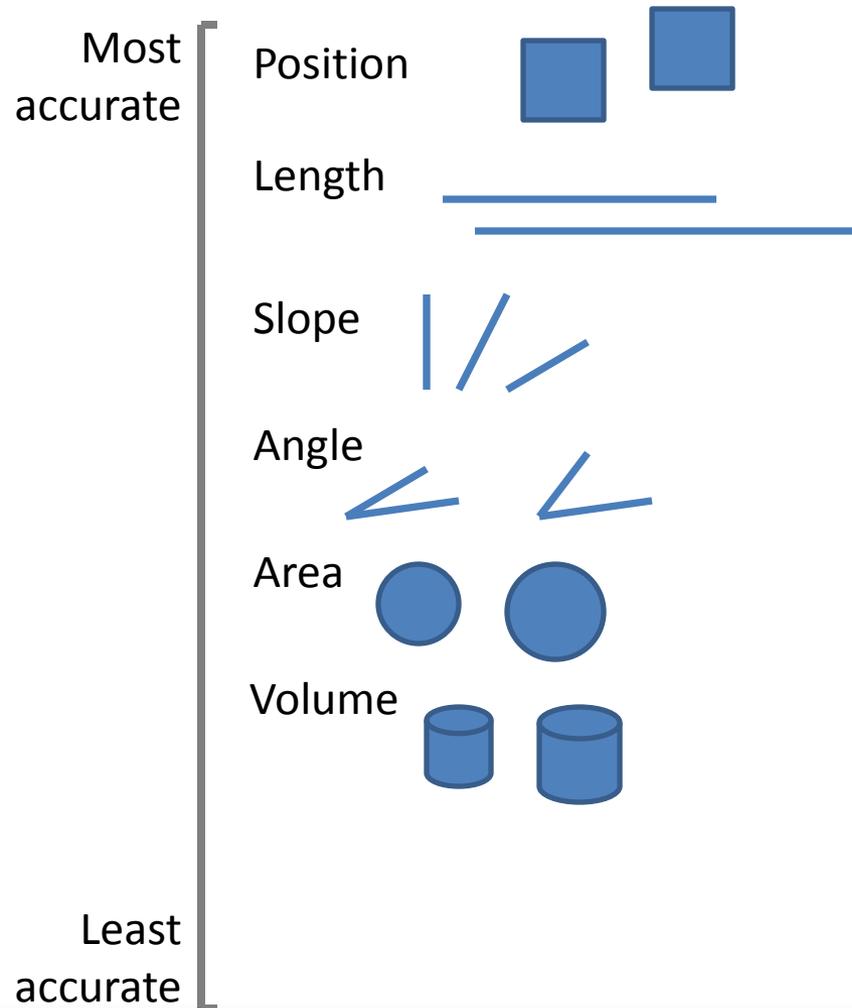
The way you modulate objects in a visualization also has a profound impact on readability and integrity.



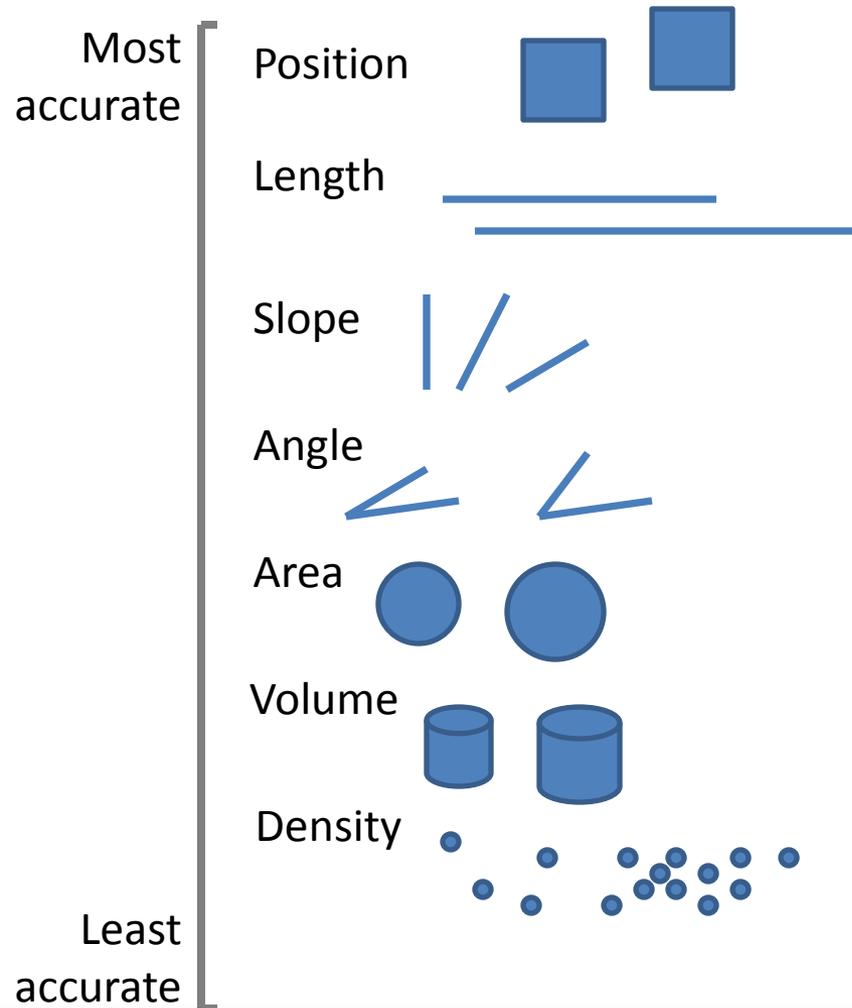
The way you modulate objects in a visualization also has a profound impact on readability and integrity.



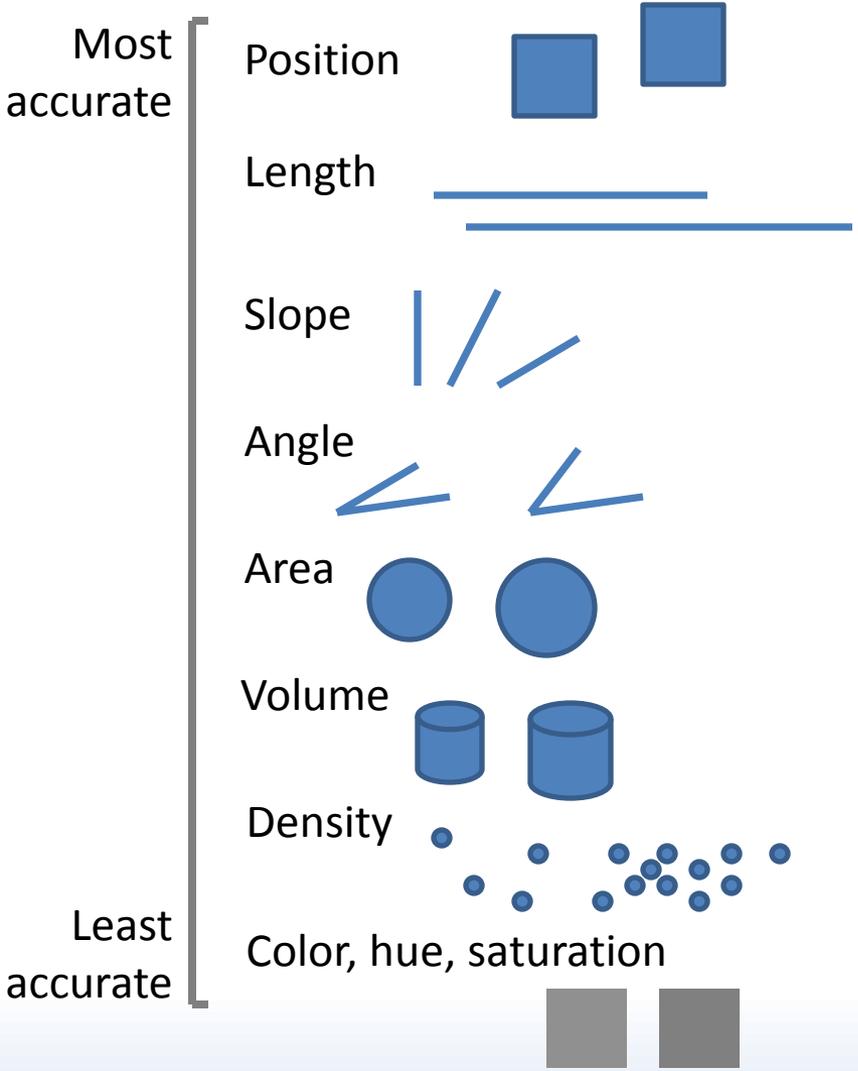
The way you modulate objects in a visualization also has a profound impact on readability and integrity.



The way you modulate objects in a visualization also has a profound impact on readability and integrity.



The way you modulate objects in a visualization also has a profound impact on readability and integrity.



Understanding the type of data you have should also affect the form of modulation you select

The three types of data

Quantitative

Ordinal

Nominal

The three types of data

Quantitative

- Data with a precise numeric value; for instance 2, 159, 25, 0x10, 0110101001.

Ordinal

Nominal

The three types of data

Quantitative

- Data with a precise numeric value; for instance 2, 159, 25, 0x10, 0110101001.

Ordinal

- Data related by order; for instance “low, medium, and high” or “small and large.”
- “Orderable categories”

Nominal

The three types of data

Quantitative

- Data with a precise numeric value; for instance 2, 159, 25, 0x10, 0110101001.

Ordinal

- Data related by order; for instance “low, medium, and high” or “small and large.”
- “Orderable categories”

Nominal

- Members of a class of things; for instance “American and European.”
- “Unorderable categories”

The three types of data

Mackinlay's Ranking of Graphical Encodings

	Quantitative	Ordinal	Nominal
Most accurate	Position	Position	Position
	Length	Density	Color Hue
	Angle	Color Saturation	Texture
	Slope	Color Hue	Connection
	Area	Texture	Containment
	Volume	Connection	Density
	Density	Containment	Color Saturation
	Color Saturation	Length	Shape
	Color Hue	Angle	Length
	Texture	Slope	Angle
	Connection	Area	Slope
	Containment	Volume	Area
Least accurate	Shape	Shape	Volume

You have the tools.
Now let's start doing.

HOMework

(Not on Collab)

Next class we will begin working with JavaScript and SVG. In the meantime, please download and install Google Chrome. Also, if you like, check out the JS examples at Raphaeljs.com

**This slide only added so I can
say my presentation had 100 slides.**