



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Tufte's Graphics Principles

CISC7204: DATA SCIENCE & VISUALIZATION

Derek F. Wong

NLP²CT – Natural Language Processing &
Portuguese-Chinese Machine Translation Research Group

derekfw@um.edu.mo

E11-4010 (Ext: 4478)

Office Hours: Thu – 16:00~17:30, Fri 11:00~12:30



Natural Language Processing & Portuguese –
Chinese Machine Translation Laboratory
自然語言處理與中葡機器翻譯實驗室

Content

Tufte's Principles of Graphics

- Graphical *Excellence*
 - What is an excellent graphic?
- Graphical *Integrity*
 - Graphics *lie* if the visual representation distorts the numerical relationships
- *Data Graphics Theory*
 - Applying the principles of excellence and integrity

Edward Tufte



A *statistician* and Yale professor

- Evangelist for *good visual design*
- Developed *several theories* about *information design*
- Most designs are static, but *many principles* apply to interactive (*computer-based*) visualization designs
- Considered as the expert and a *pioneer* in *data visualization*

Why Does Visualization Work?

Recap

COGNITION is *limited*, **MEMORY** is *limited*

Uses **PERCEPTION** to point out **INTERESTING THINGS**

Why Do We Create Visualizations?

Find patterns

See data in context

Answer questions

Support computational analysis

Generate hypotheses

Tell a story

Inspire

Expand memory

Make decisions

What is an Excellent Graphic?

Tufte's Principles of Graphical Excellence

What makes for *graphical excellence*?

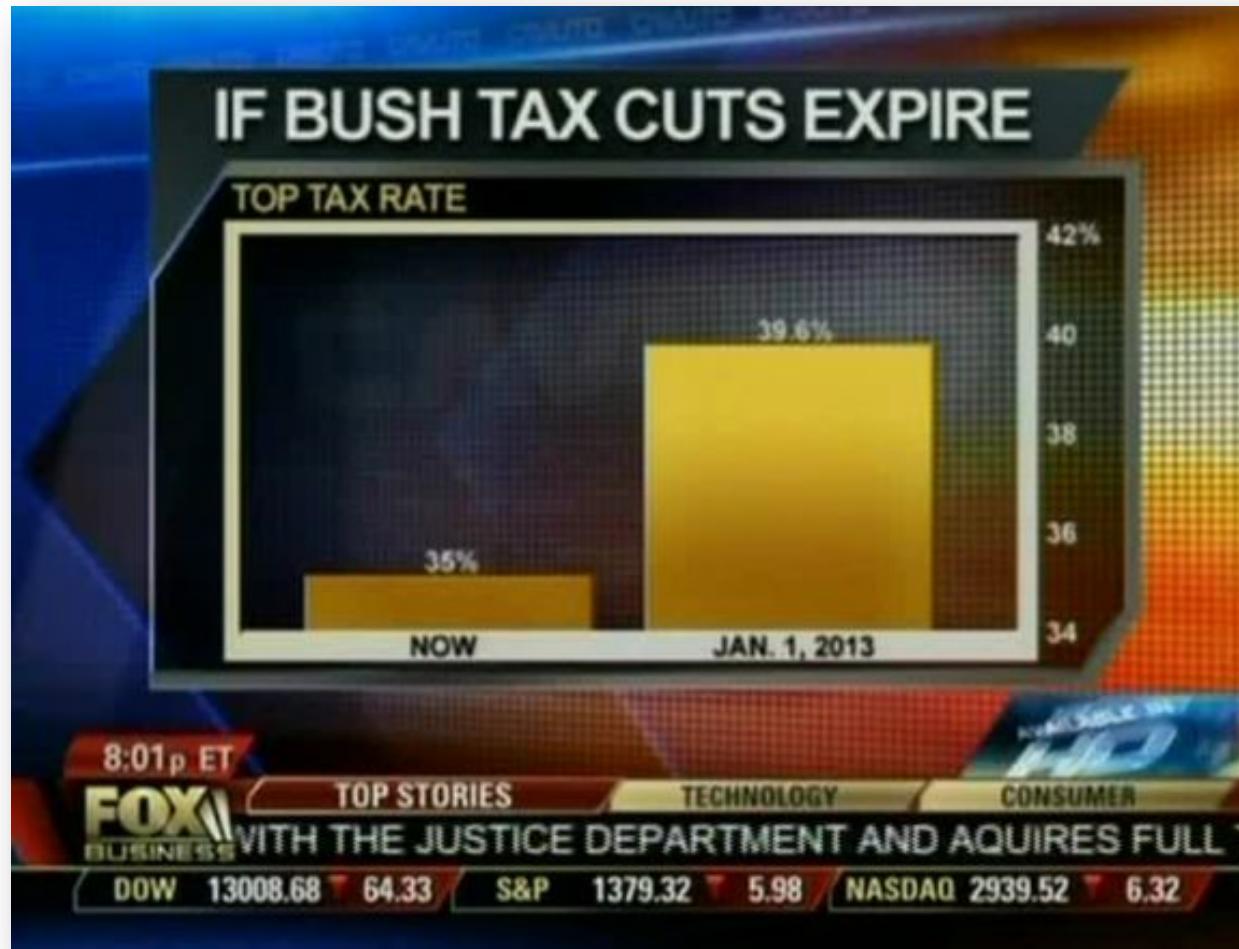
- *Well-designed* presentation of *interesting data*, a combination of *substance*, *statistics*, and *design*
- *Complex ideas* communicated with *clarity*, *precision*, and *efficiency*
- *Greatest number of ideas* in the *shortest time* with the *least ink* in the *smallest space*, “Data Richness”
- Nearly always *multivariate* in designing an excellent graphic
- Graphical excellence *requires telling the truth* about data

Graphical Integrity

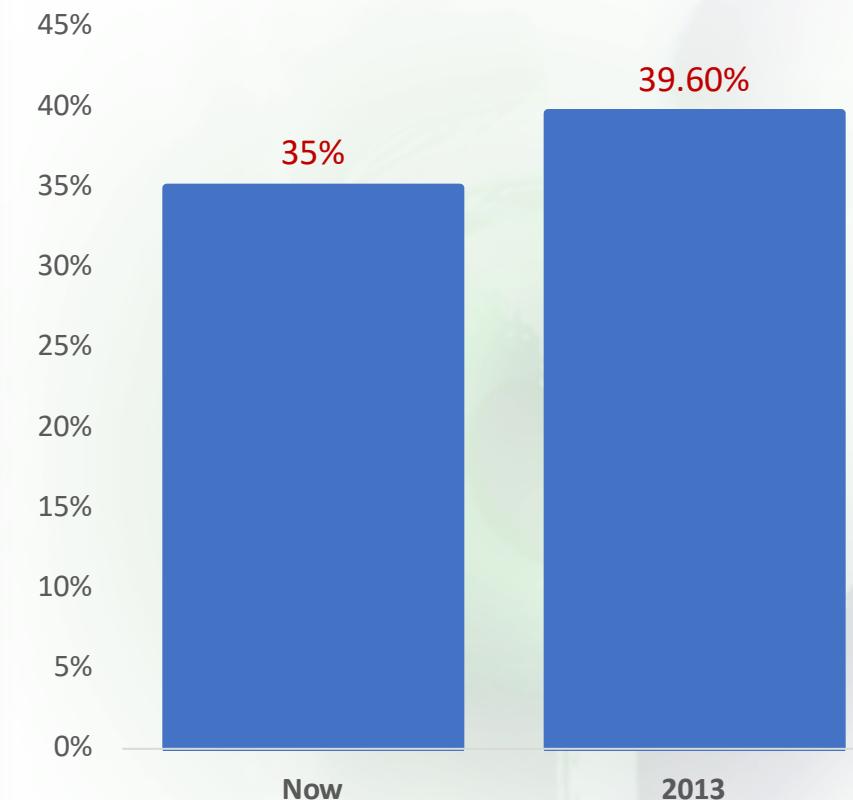
*Clear, detailed, and thorough labeling should be used to **defeat** graphical **distortion** and **ambiguity***

Graphical excellence ***requires telling the truth*** about data

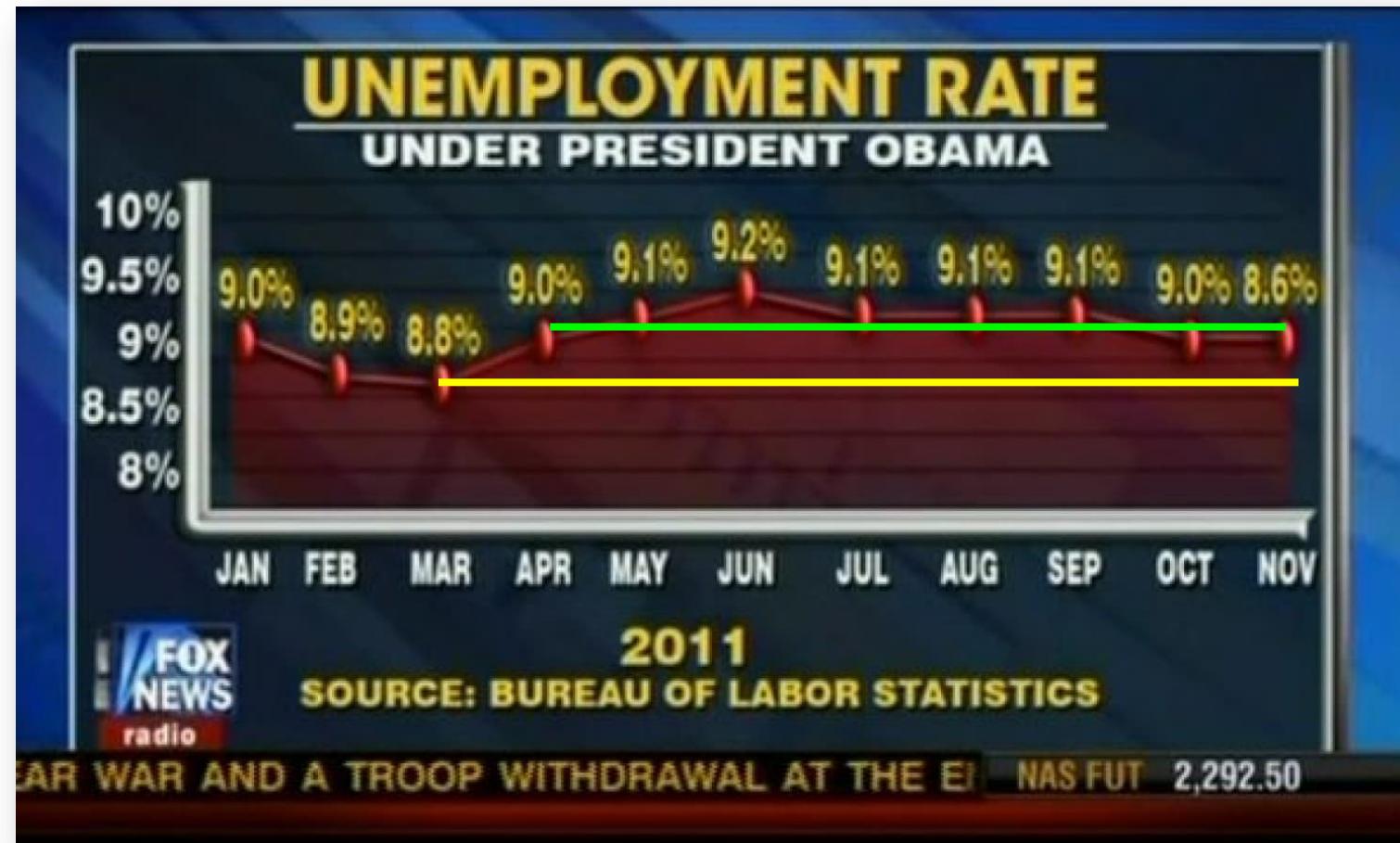
Misleading Graph



If Bush Tax Cuts Expire

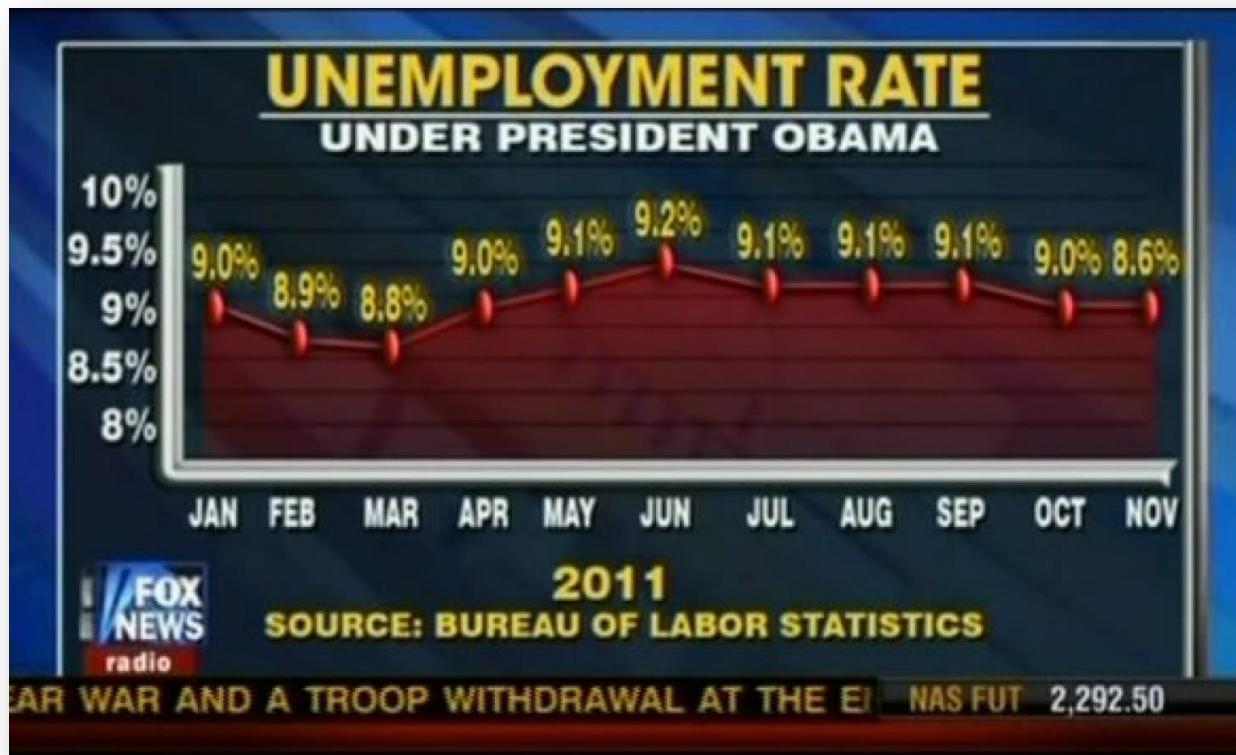


Misleading Graph



[Fox News's unemployment chart@2011](#)

Misleading Graph

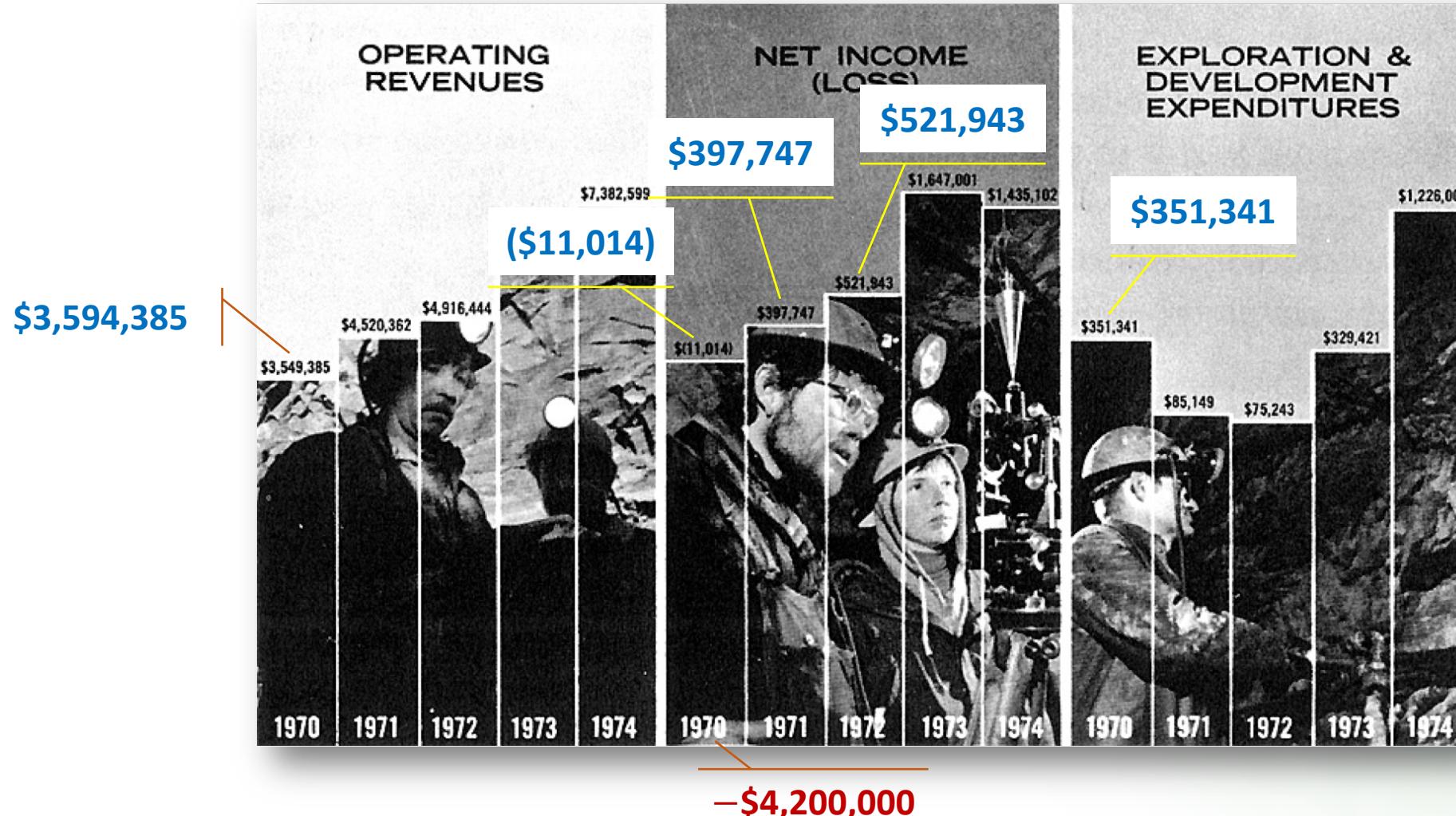


Fox News's unemployment chart@2011



Missing Scales

What is the baseline?

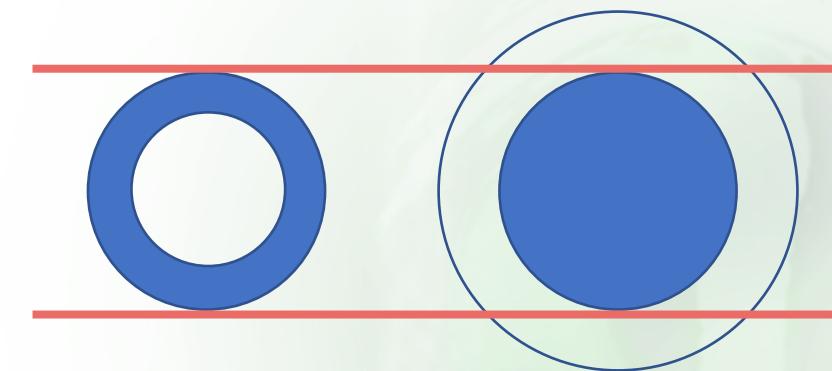
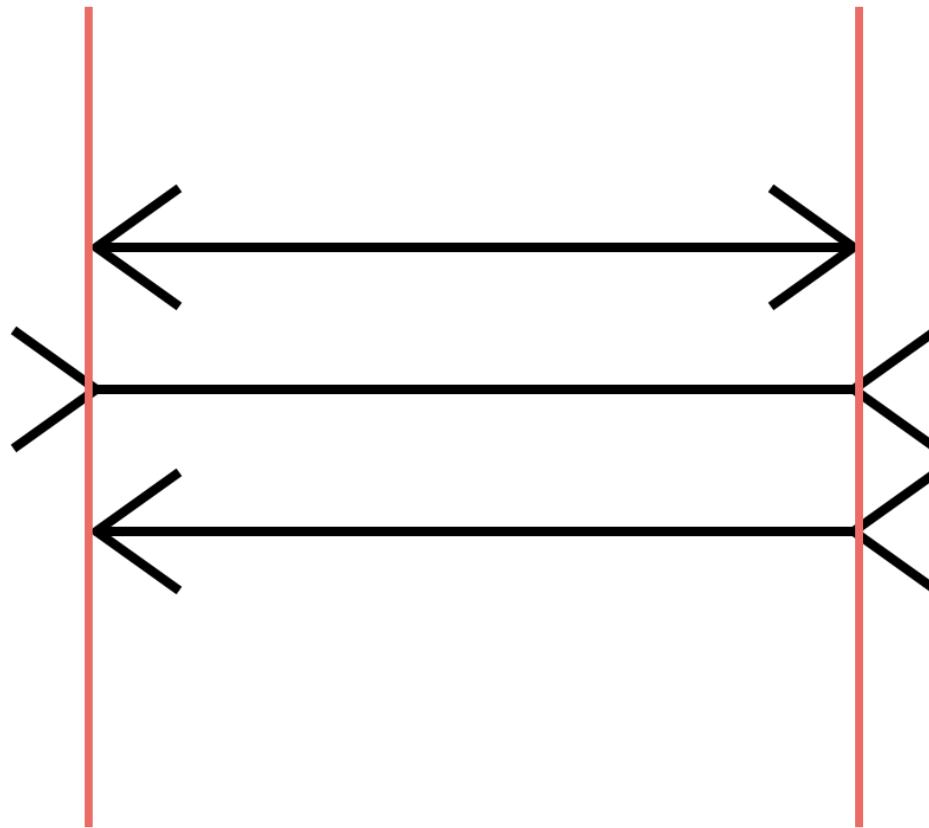


What is Distortion in a Data Graphic?

A graphic *does not distort* if the *visual representation* of data is *consistent* with the *numerical representation*

Visual representation of data ≠ numerical representation

What is Distortion in a Data Graphic?



Graphical Integrity

Clear, detailed, and thorough labeling should be used to
defeat graphical *distortion* and *ambiguity*

“Above all else – Show the data”

Two Principles

Representation of numbers

*The numbers should be **directly proportional***

Clear, detailed and thorough labeling

*Write out **explanations** and **label** important events*

First Principle: *The Lie Factor*

- Tufte define the term “*The Lie Factor*” as:

$$LF = \frac{\text{size of graphic}}{\text{size of data}}$$

- “*High*” lie factor (LF) leads to:
 - *Exaggeration* of differences or similarities
 - *Deception*
 - *Misinterpretation*

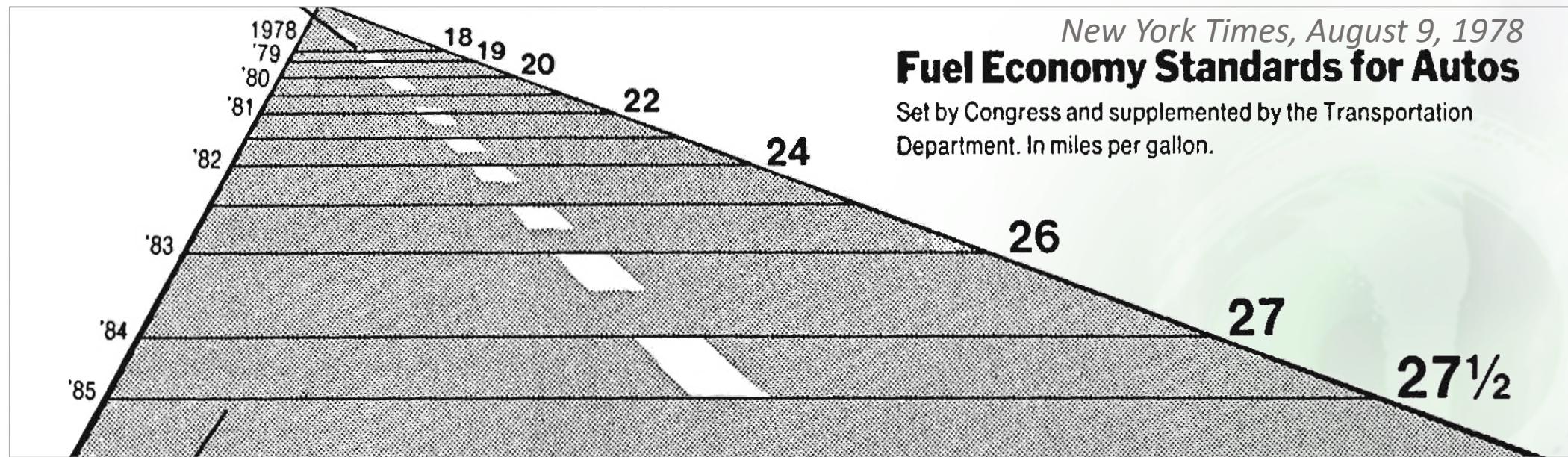
First Principle: *The Lie Factor*

- Tufte define the term “*The Lie Factor*” as:

$$LF = \frac{\text{size of graphic}}{\text{size of data}}$$

- The Lie Factor can be $LF > 1$ or $LF < 1$
- If $LF > 1$, then *size of graphic* is greater than the *size of data*
 - This leads to *exaggeration* of the data (*overstating the data*)
- If $LF < 1$, then the *size of data* is greater than the *graphic*
 - Leading to *hiding* of the data (*understating the data*)

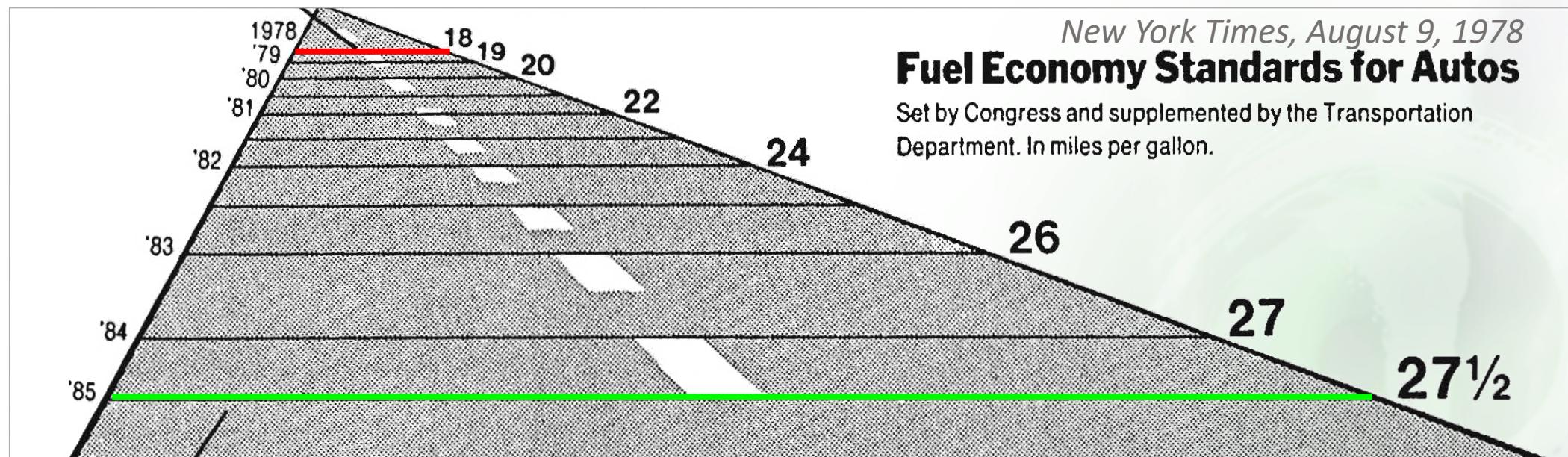
What is Wrong With This?



The US Department of Transportation had set a series of fuel economy standards to be met by automobile manufacturers, beginning with 18 miles per gallon in 1978 and moving in steps up to 27.5 by 1985

What is Wrong With This?

The line representing *18 miles per gallon* in 1978, is *0.6 inches long*



The line representing *27.5 miles per gallon* in 1985, is *5.3 inches long*

What is Wrong With This?

The *increase in real data* between 1978 to 1985 (from *18 MPG* to *27.5 MPG*) is:

$$\frac{(27.5 - 18.0)}{18.0} \times 100 = 53\%$$

The *difference in length* between 1978 to 1985 (from *0.6 inches* to *5.3 inches*) is:

$$\frac{(5.3 - 0.6)}{0.6} \times 100 = 783\%$$

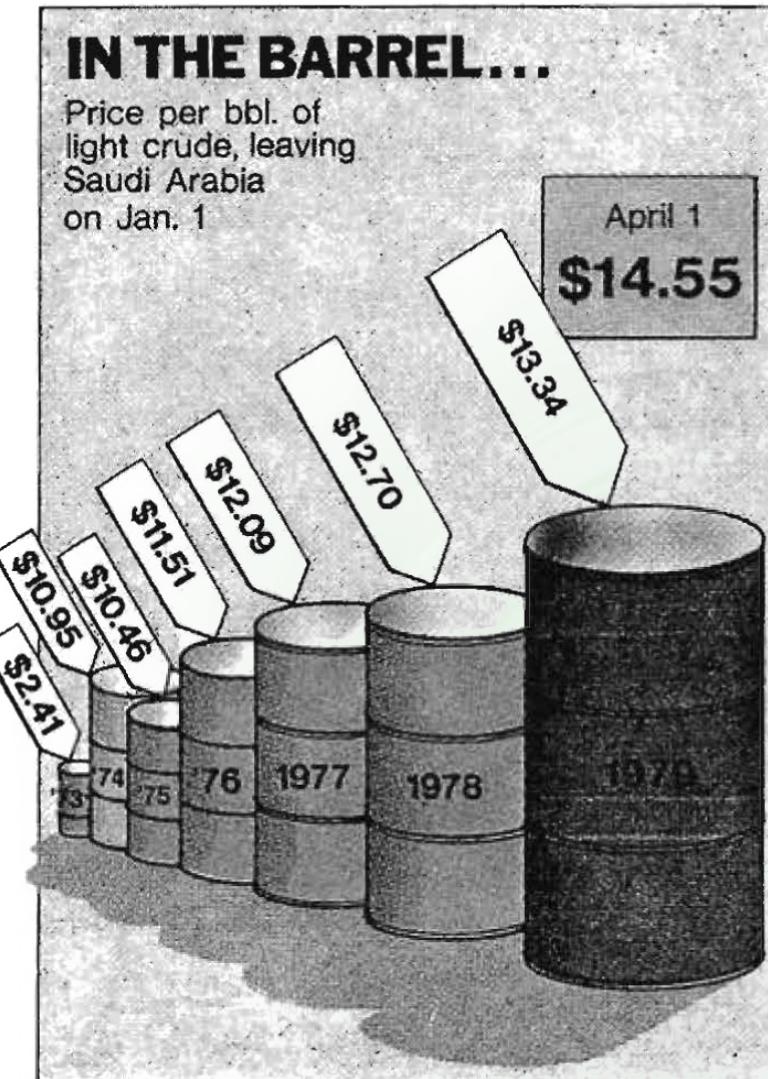
Lie Factor is:

$$\frac{783}{53} = 14.8$$

Lie Factor Example

\$2.41 to \$13.34 is *approx.* 5 *times* increase

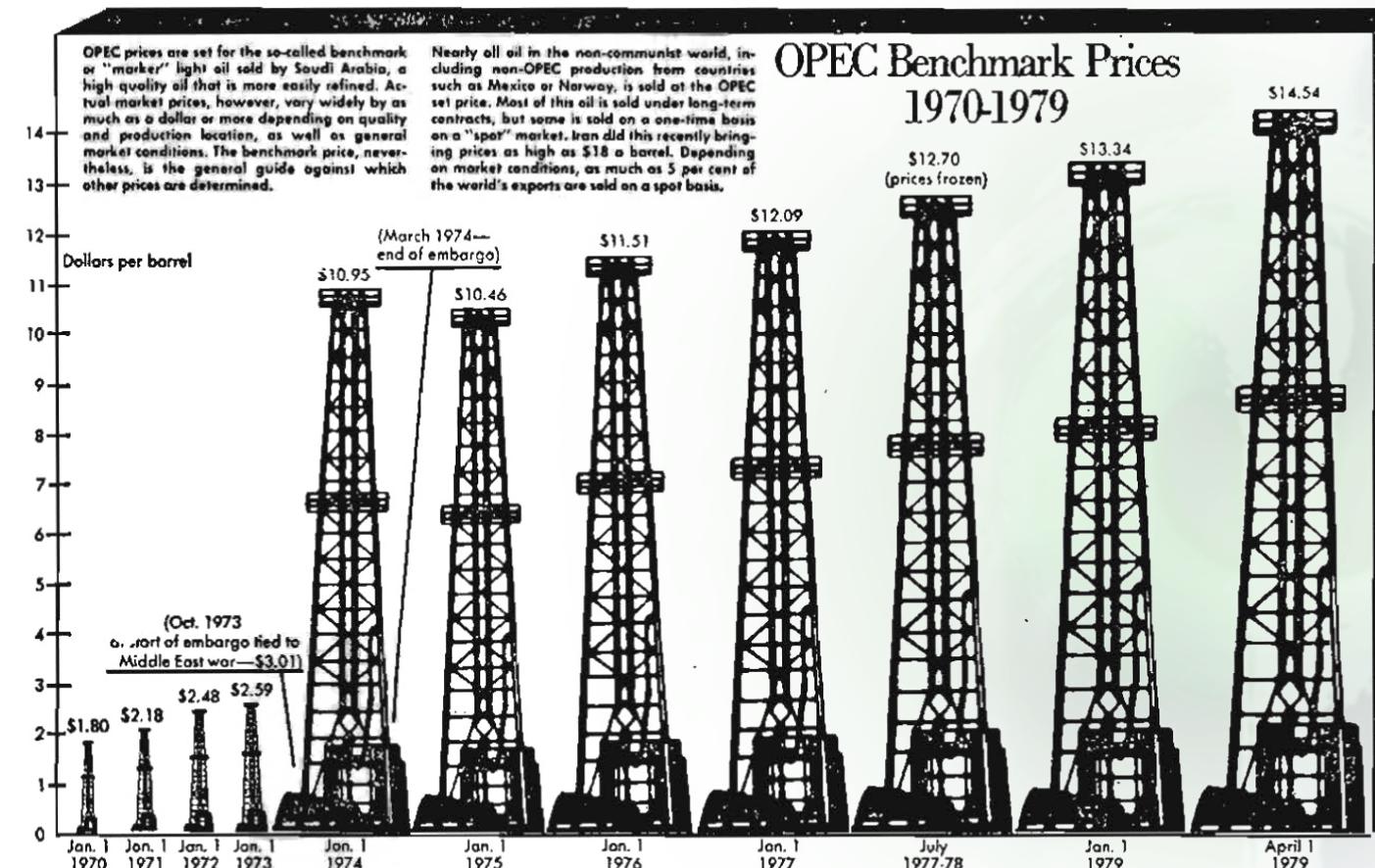
But taking into account the *barrel size*, the design has a *lie factor* of **9.4!**



Lie Factor Example

This design contains a *lie factor*
of 9.5

Washington Post, March 28, 1979

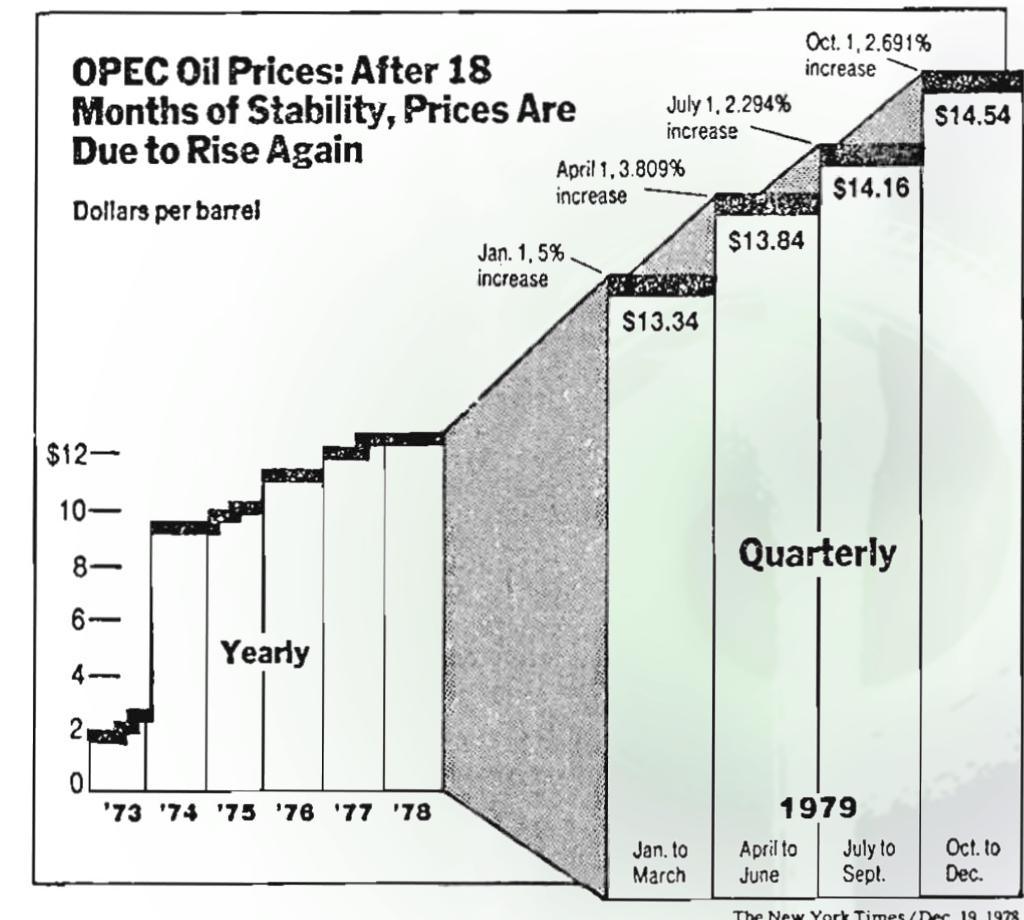


Other Ways to Lie

Design Variation

Beware of the “3D” effect. It *distorts* the *telling of the data*

- There are *five vertical scales* here:
 - 1973 ~ 1978: 1 inch = \$8.00
 - Jan ~ Mar: 1 inch = \$4.73
 - Apr ~ Jun: 1 inch = \$4.37
 - Jul ~ Sep: 1 inch = \$4.16
 - Oct ~ Dec: 1 inch = \$3.92
- And *two horizontal scales*:
 - 1973-1978: 1 inch = 3.8 years
 - 1979: 1 inch = 0.57 years



The *lie factor* of **15.1**

New York Times, December 19, 1978

Data vs. Design Variation

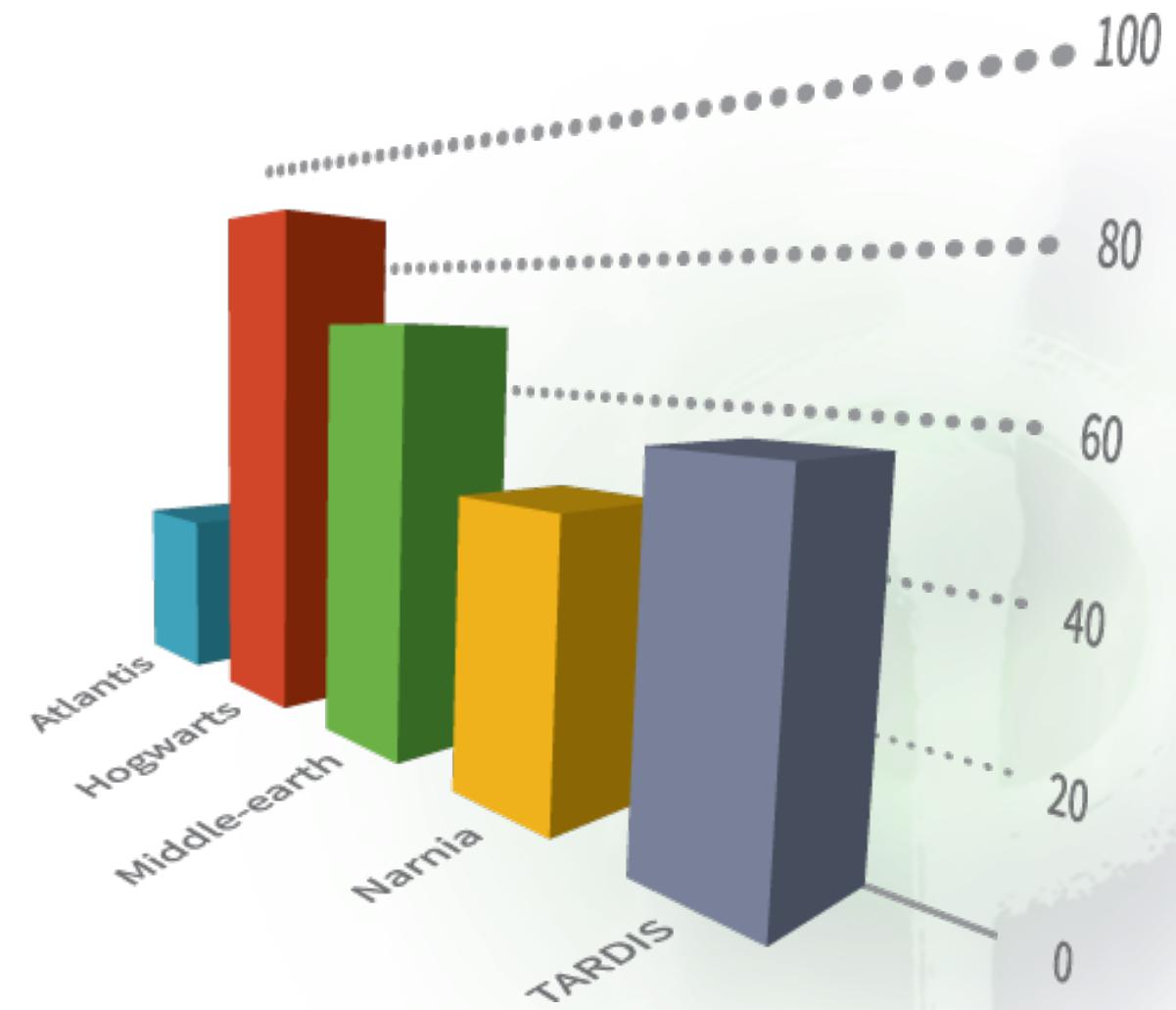
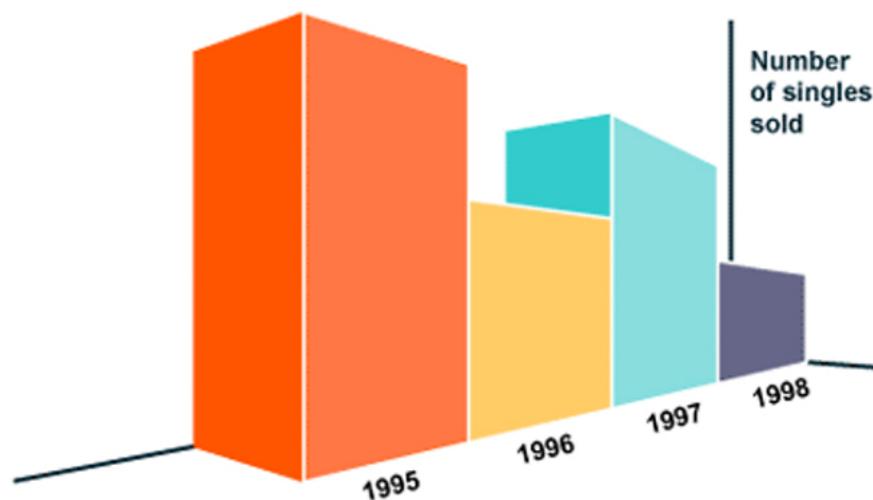
Show *Data Variation*, not *Design Variation*

- *Lie Factor = 9.4 for In the Barrel*
- *Lie Factor = 9.5 for the OPEC Benchmark Prices*
- *Lie Factor = 15.1 for the OPEC Oil Prices*
- *Stay consistent* with your design through the data representation

Let the *data variation present itself*

Other Ways to Lie

3D Effect



Other Ways to Lie

Encoding

Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.



Numbers have magnitude and close values should be reflected

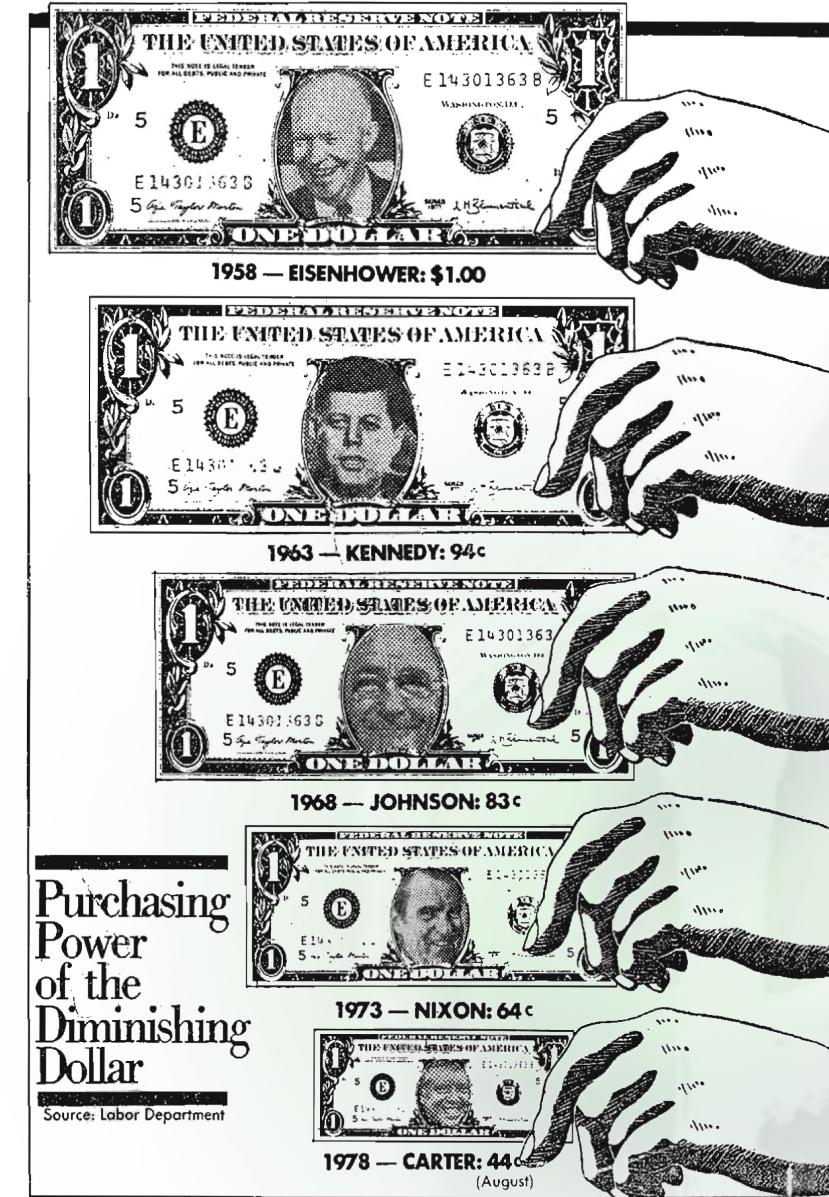
Other Ways to Lie

Double Encoding

- Here, both width and height encode the same information. The effect is multiplicative

$$0.44(\text{width}) \times 0.44(\text{height}) = 0.19$$

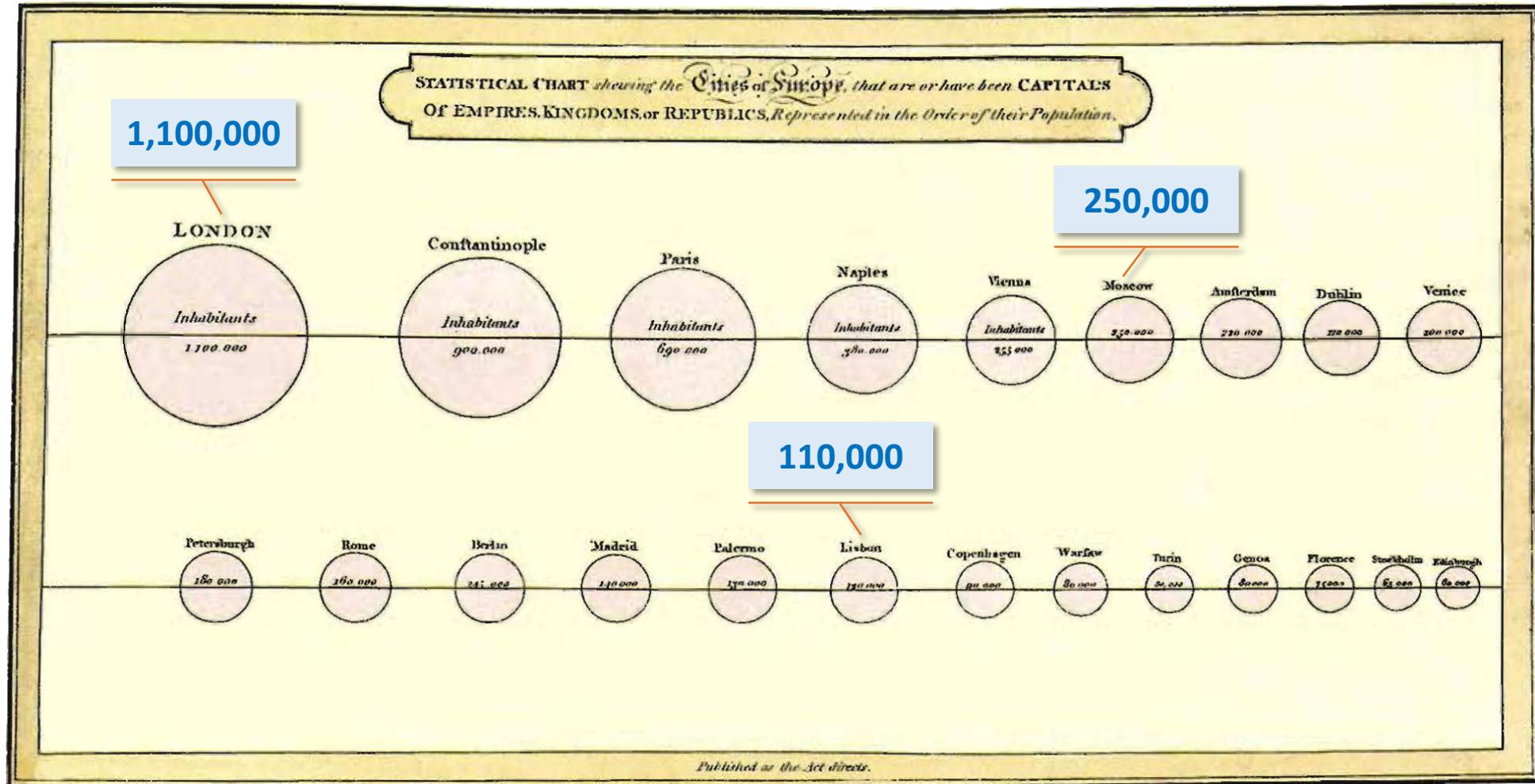
Principle: *The number of graphical dimensions should not exceed the number of data dimensions*



Washington Post, October 25, 1978

Other Ways to Lie

Unintended Encoding



Other Ways to Lie

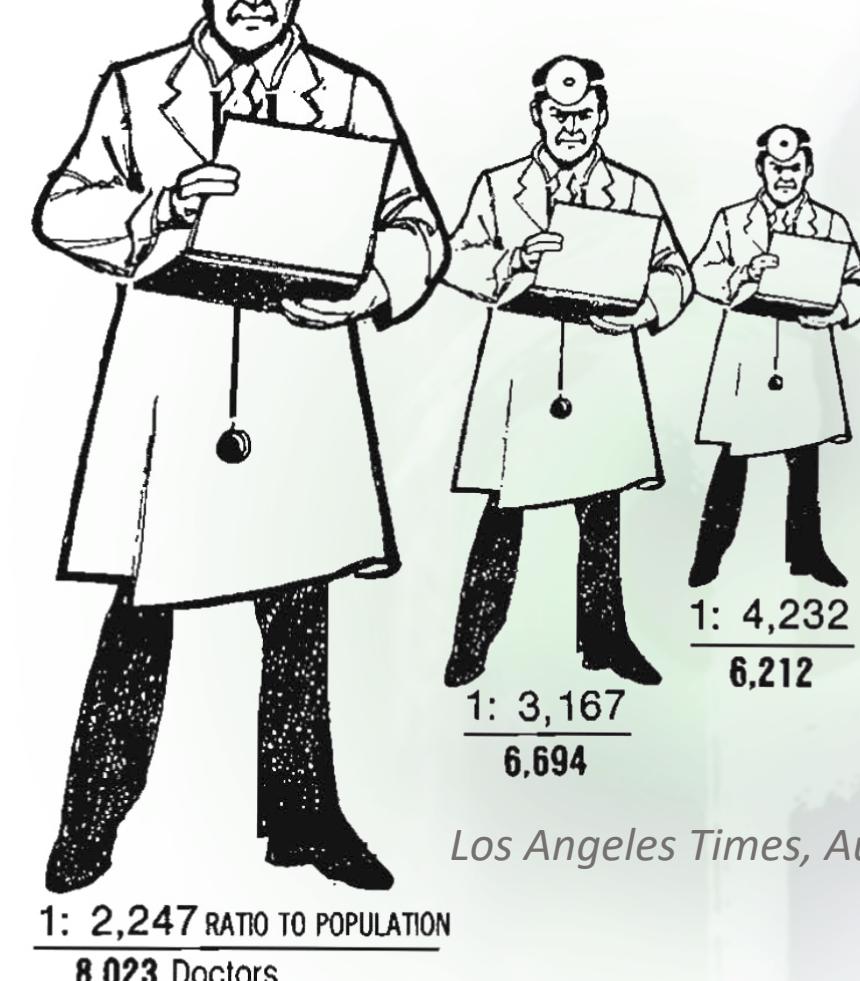
Alignment

This design contains a *lie factor* of 2.8

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

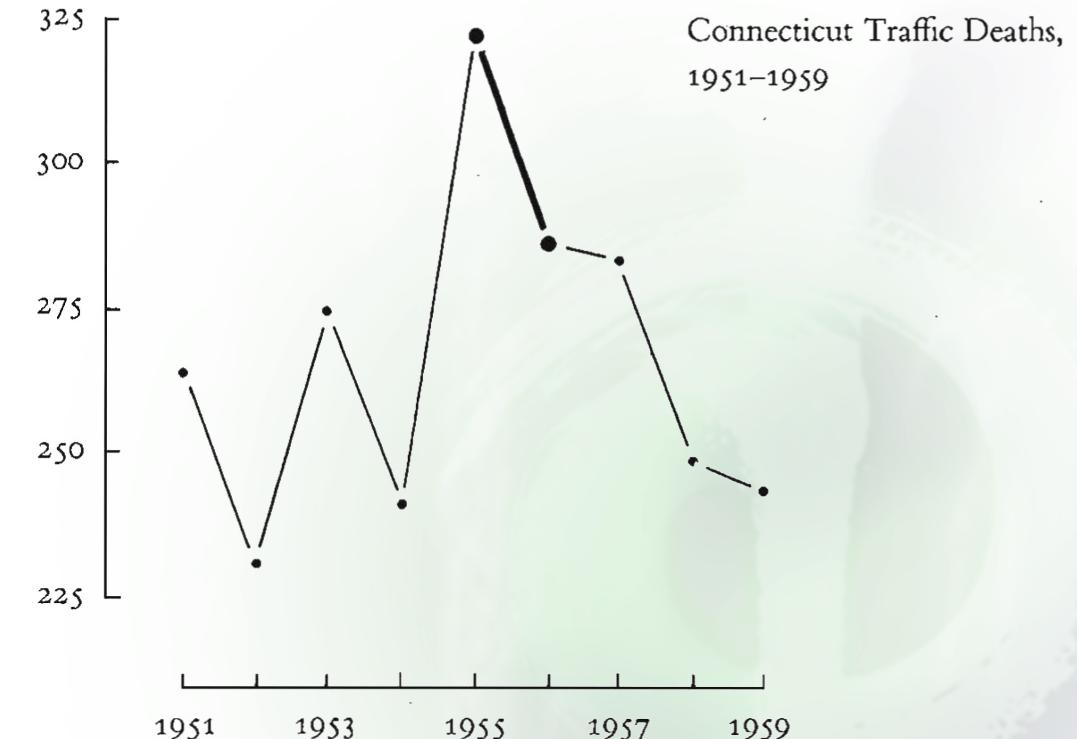
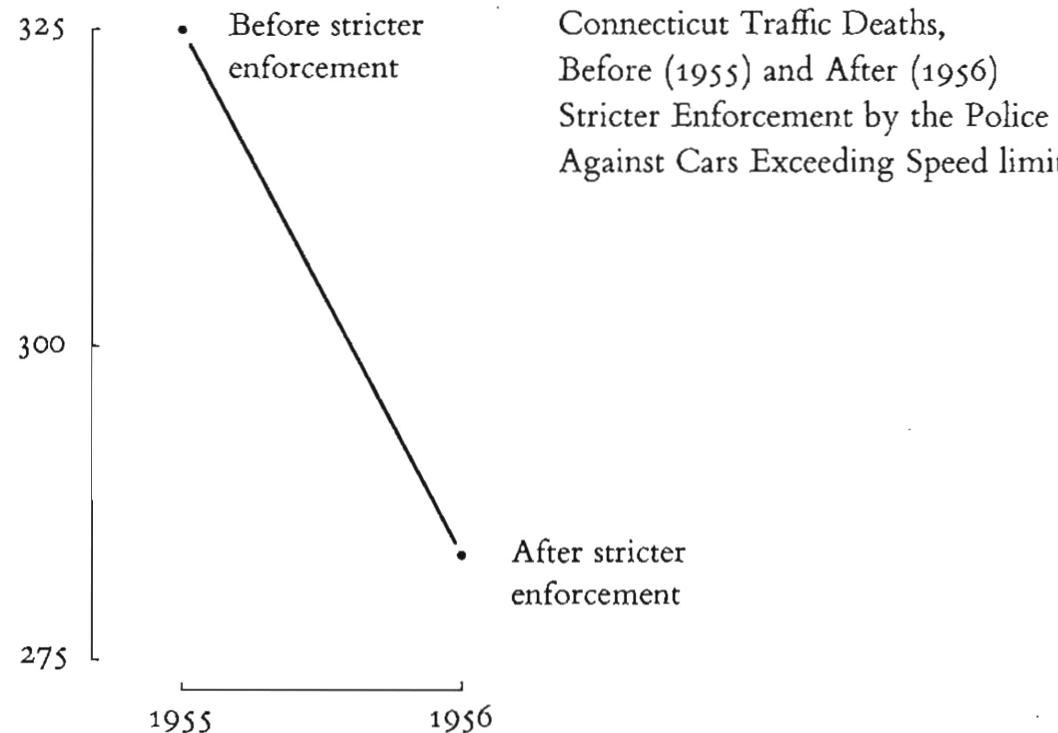
1964	1975	1990
27 %	16.0 %	12.0 %



Los Angeles Times, August 5, 1979

Other Ways to Lie

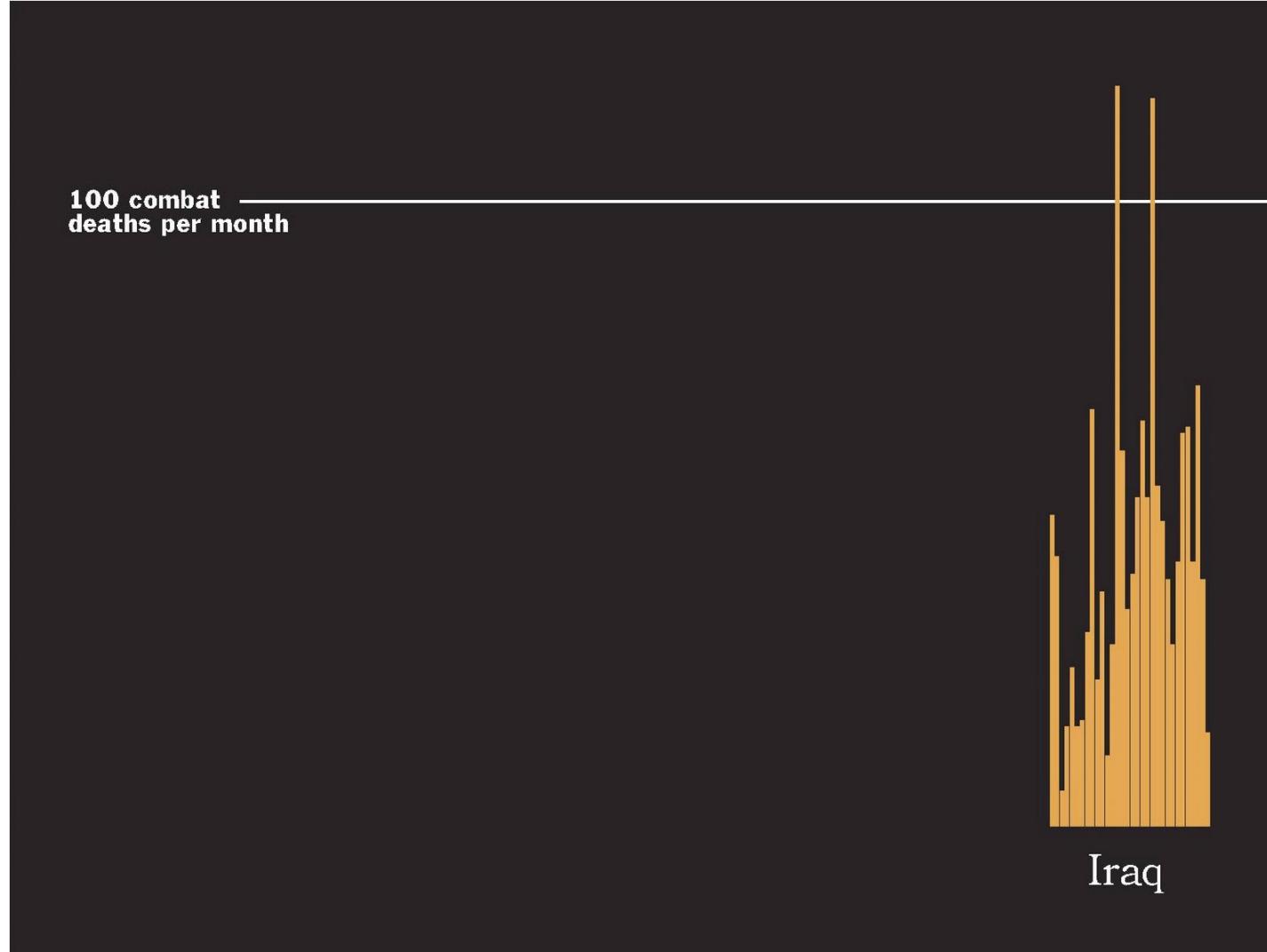
Limiting Context



Principle: *Graphics must not quote data out of context*

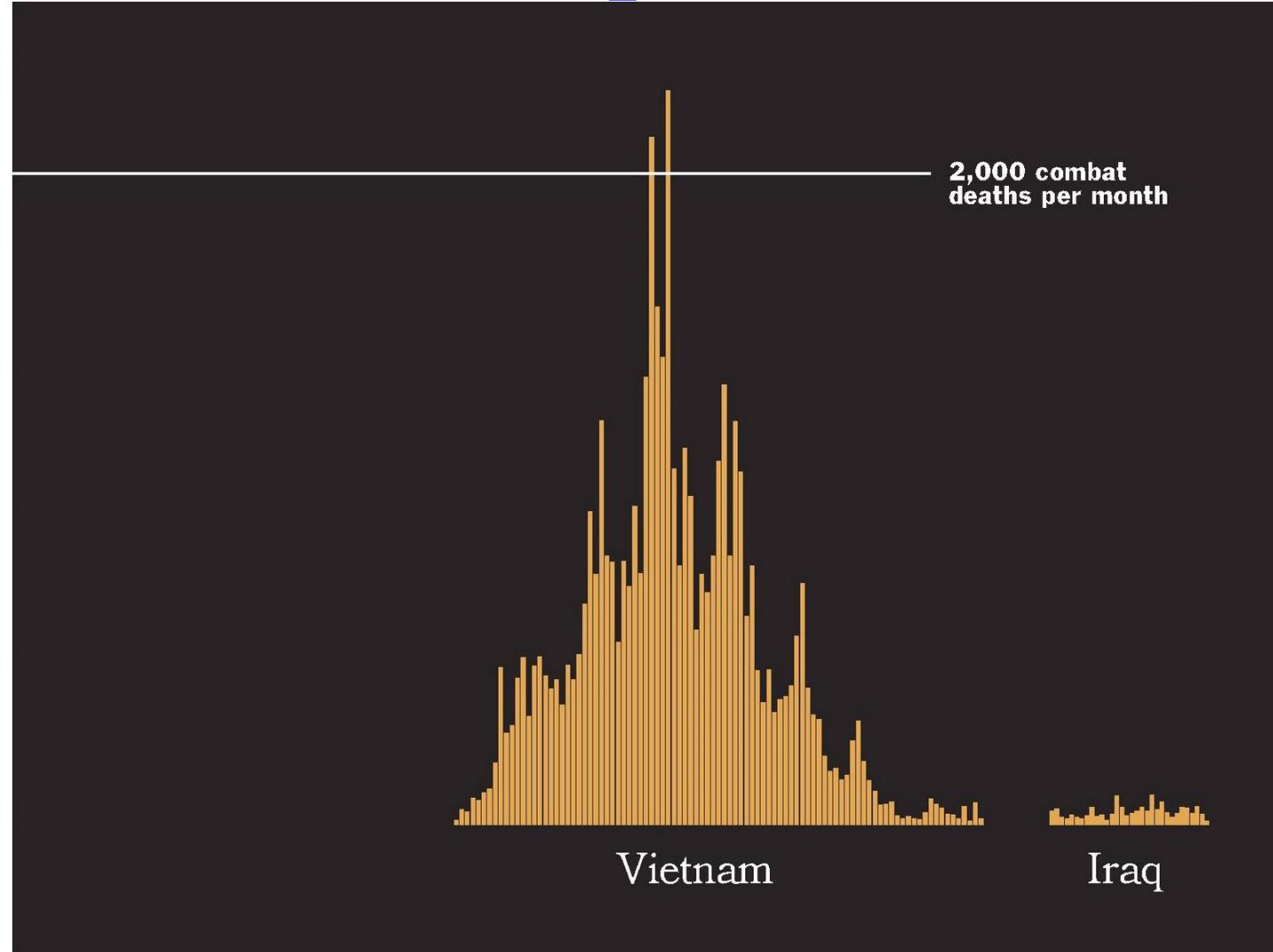
Other Ways to Lie

Limiting Context



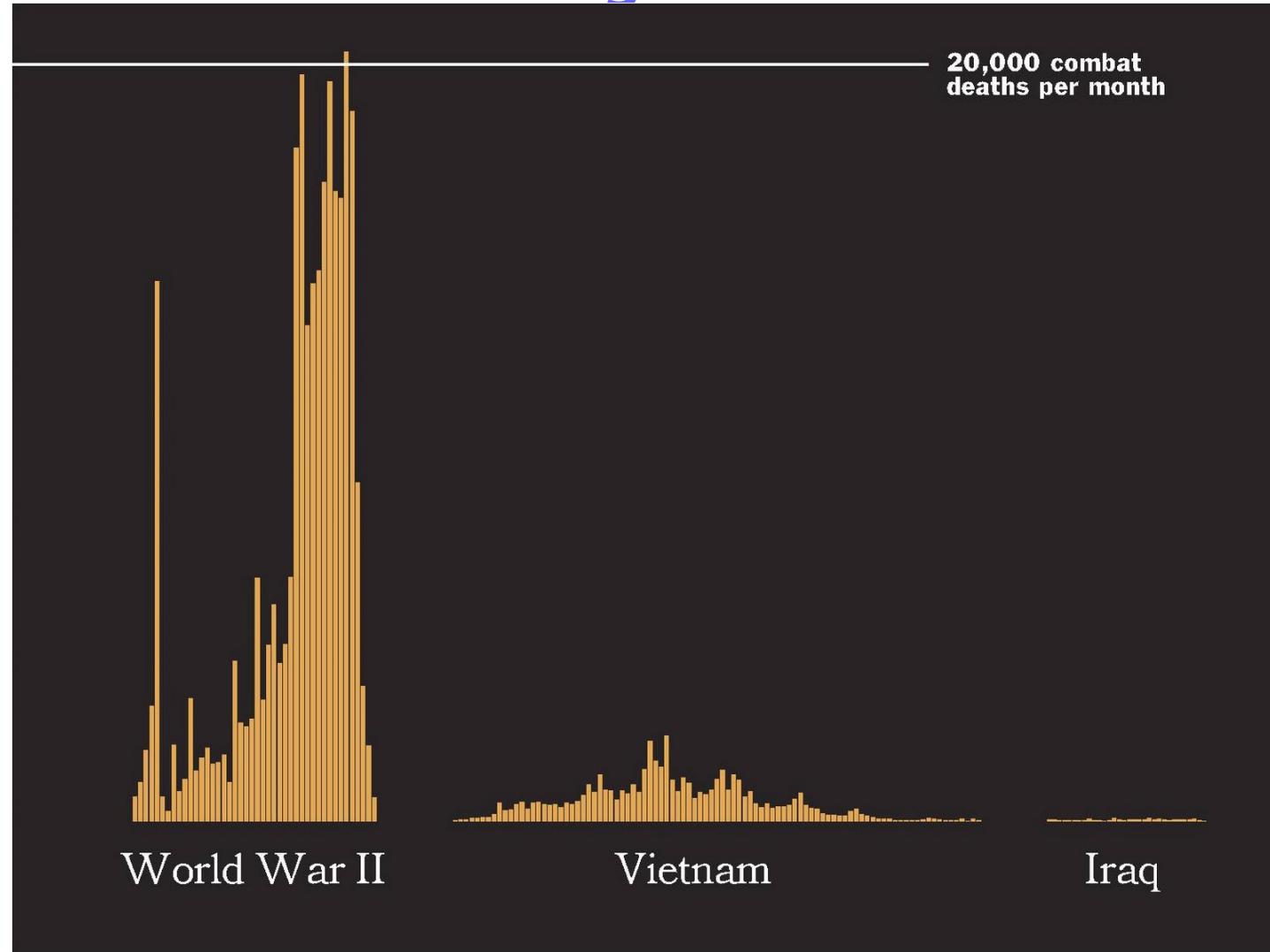
Other Ways to Lie

Limiting Context



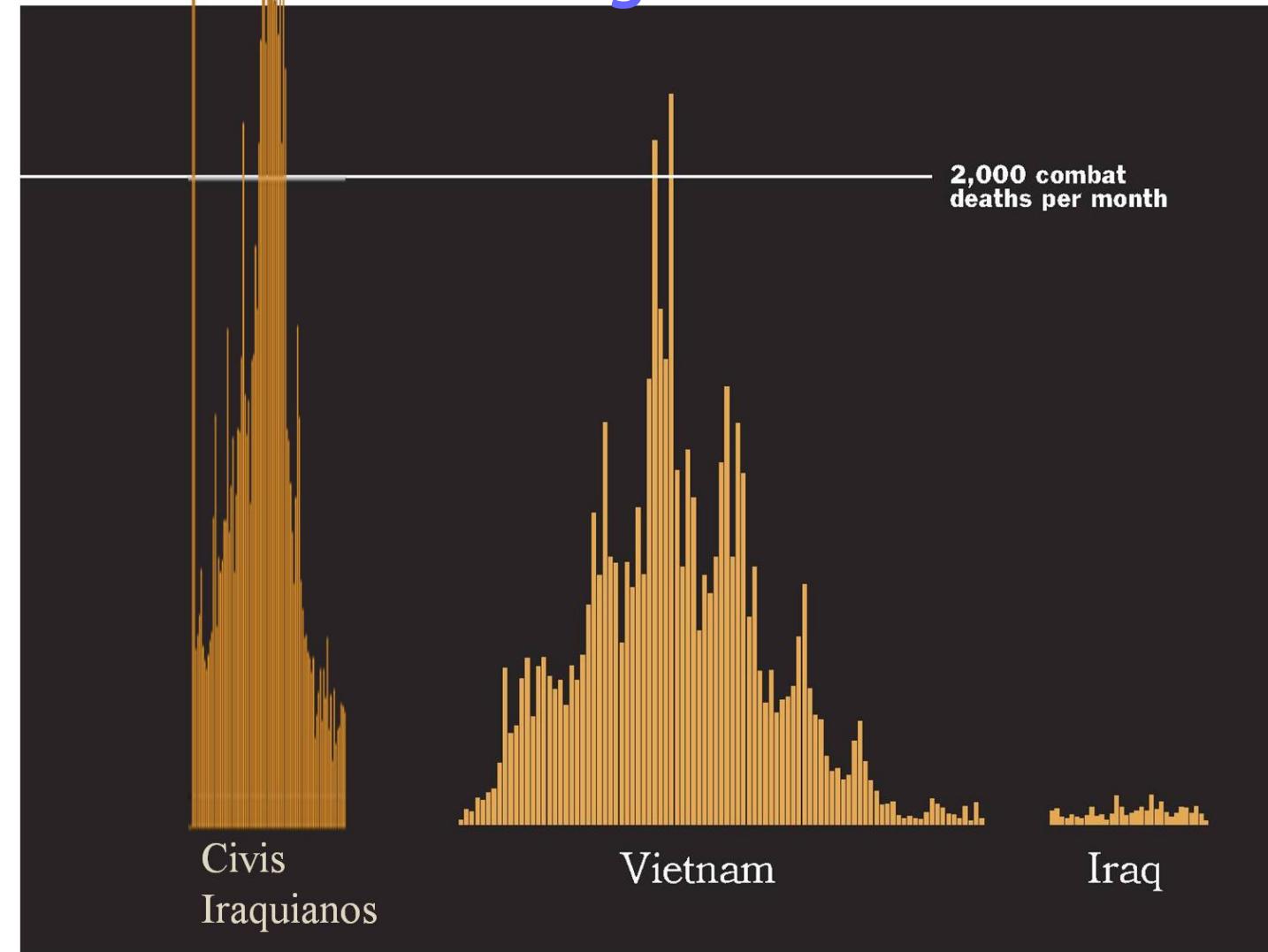
Other Ways to Lie

Limiting Context



Other Ways to Lie

Limiting Context



Tufte's Principles of Graphical Integrity

- *Physical area* on the graphic should be *directly proportional* to the *data represented*
- *Data and important events* should be *labeled* and *explained* on the graphic
- Show *data variation*, not *design variation*
- Time-series displays of money should use deflated and standardized units
- The *number of chart dimensions* should *not exceed* the *number of data dimensions*
- *Do not quote* data *out of context*

References

- Edward Tufte (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire CT.
- Erik W. Anderson, Kristin C. Potter, Laura E. Matzen, Jason F. Shepherd, Gilbert Preston, Cláudio T. Silva (2011). [A User Study of Visualization Effectiveness Using EEG and Cognitive Load](#). Comput. Graph. Forum 30(3): 791-800.

Acknowledgements

Some of the materials are adapted from:

- Alvitta Ottley
- Doug Smith