CSC 544
Data Visualization

Joshua Levine
josh@arizona.edu

Lecture 05
Design Principles

Jan. 30, 2023

Today’s Agenda

- Reminders:
  - A01 posted, questions?
  - P01 posted, questions?
  - Design Critiques
- Goals for today: Introduce design principles for visualization

Reminder:
Main Themes for CSC544

- **Mechanics**: <u>how</u> do I build a visualization?
  - Javascript, CSS, HTML, d3
- **Principles**: <u>why</u> should I build it in this way?
  - mathematical and perceptual arguments
- **Techniques**: <u>what</u> do I use to turn principles and mechanics in an actual visualization?
  - algorithms, software libraries

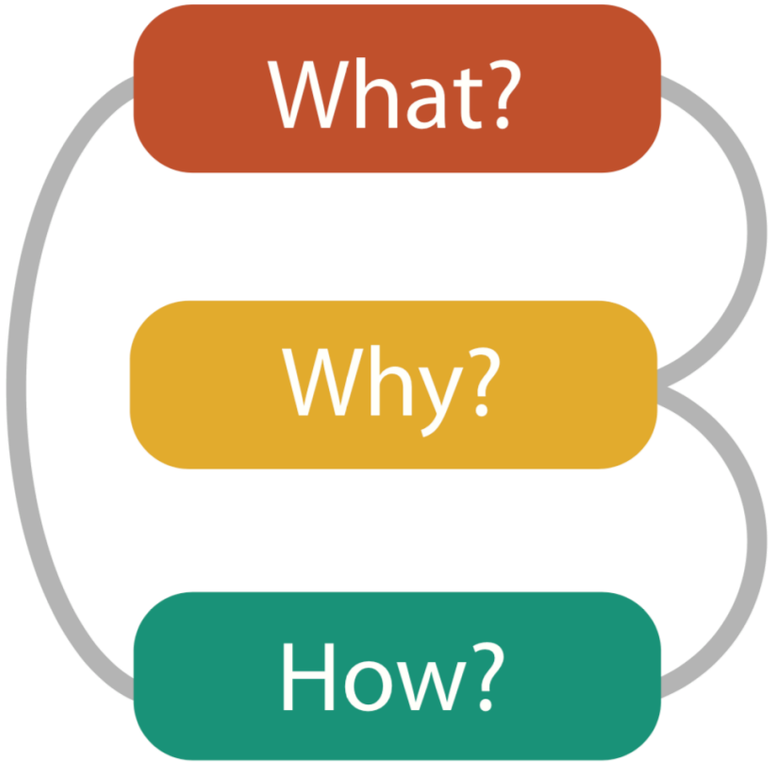
Reminder:
Main Themes for CSC544

- **Mechanics**: <u>how</u> do I build a visualization?
  - Javascript, CSS, HTML, d3
- **Principles**: <u>why</u> should I build it in this way?
  - mathematical and perceptual arguments
- **Techniques**: <u>what</u> do I use to turn principles and mechanics in an actual visualization?
  - algorithms, software libraries

An Analysis Framework for Visualization

Munzner's What-Why-How Framework

- **What** data the user sees?
- **Why** the user intends to use vis?
- **How** the visual encoding and interaction idioms are constructed in terms of design choices?

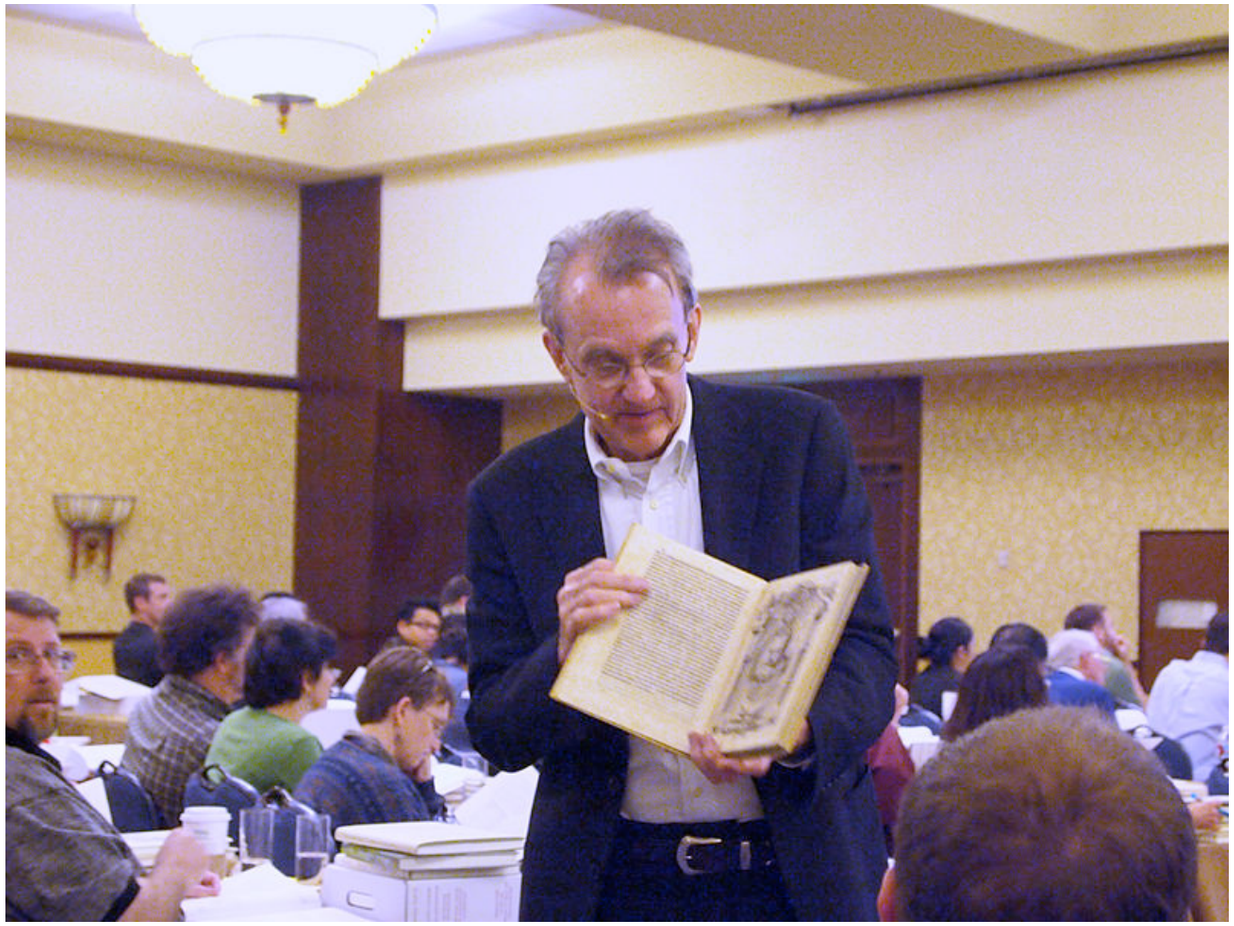


- Each of these questions will be mapped to a data-task-idiom trio to evaluate the quality of the visualization.

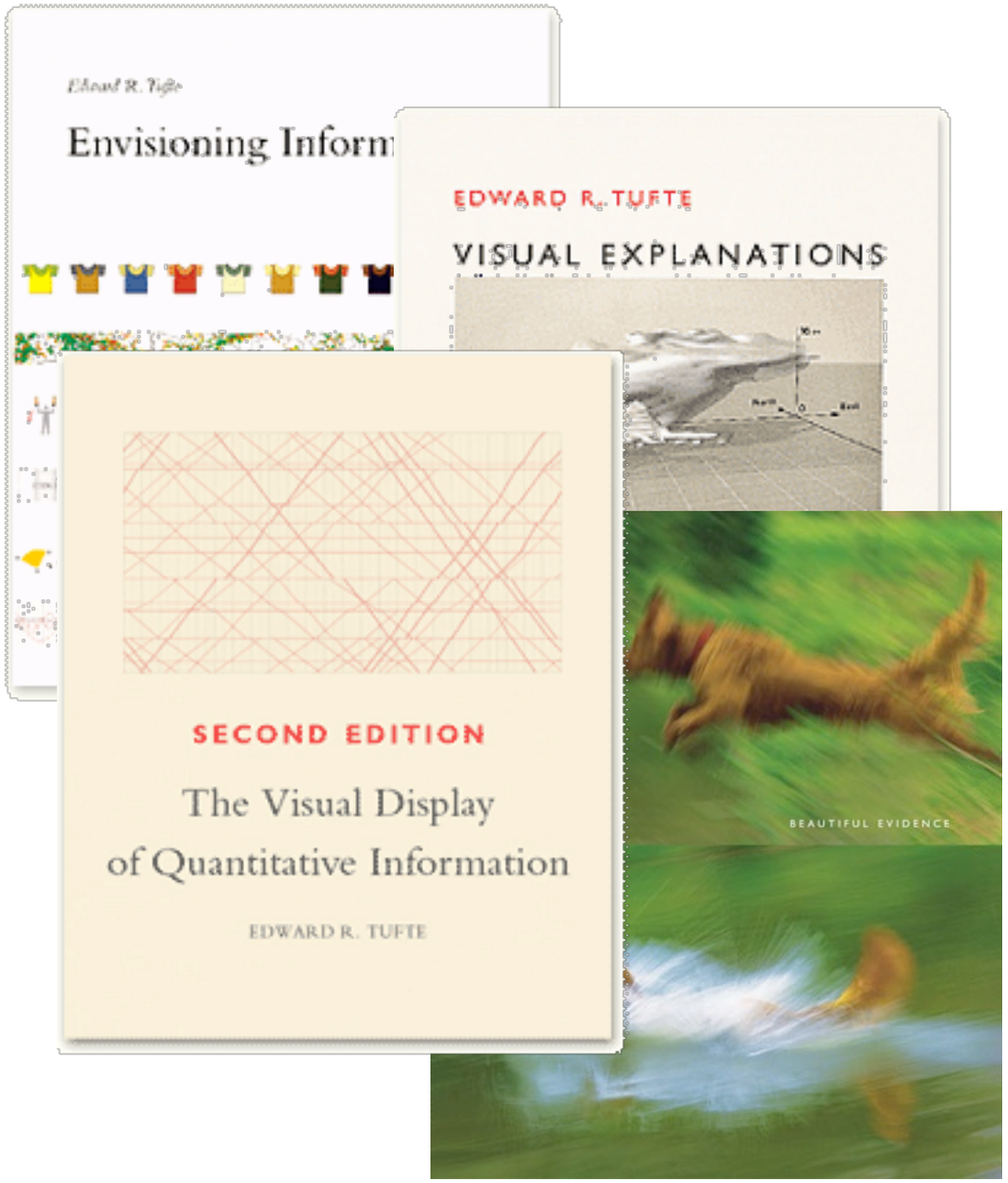
Graphical Integrity and Excellence

Design Excellence

"Well-designed presentations of interesting data are a matter of substance, of statistics, and of design."



Edward Tufte



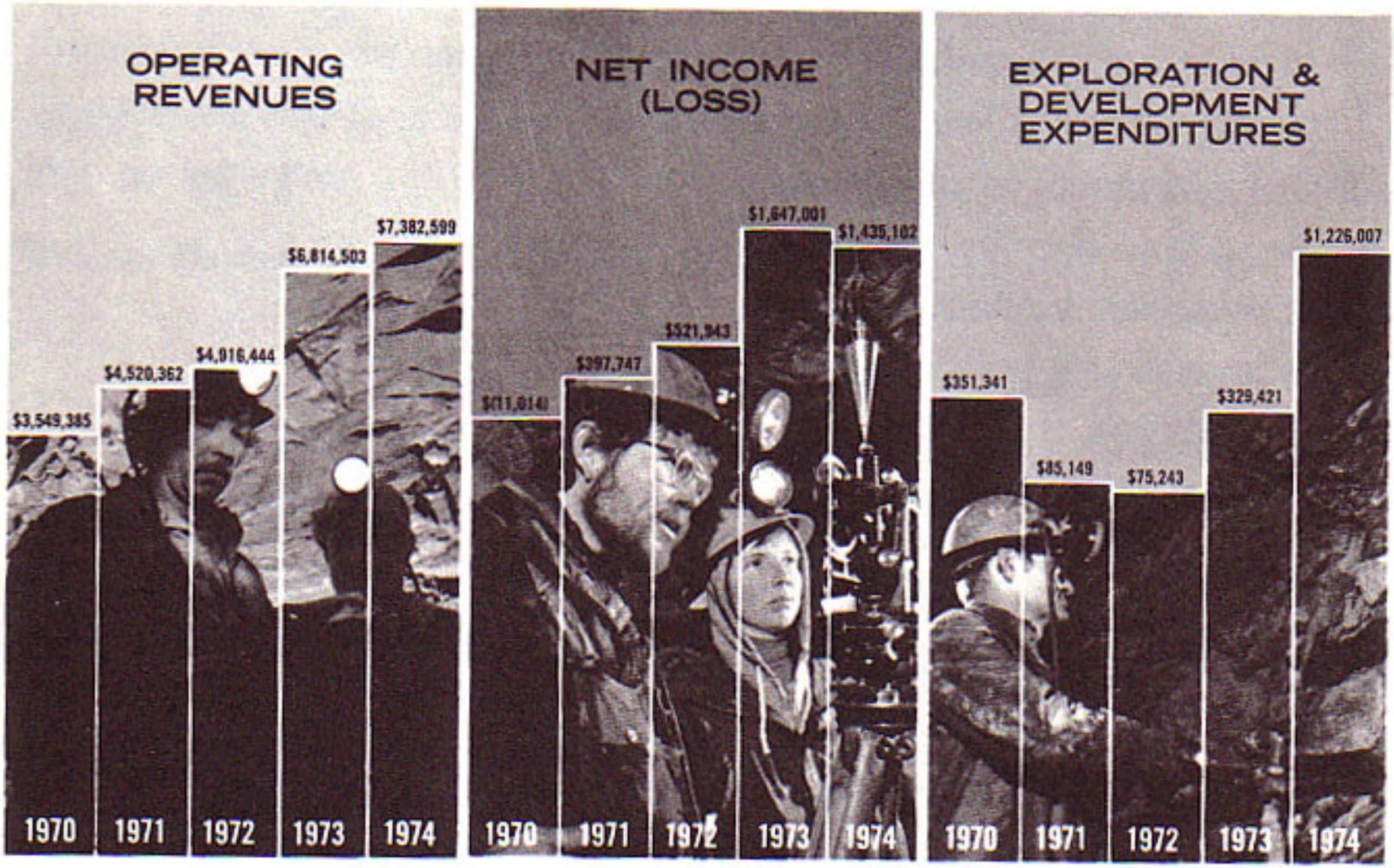
Graphical Integrity

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

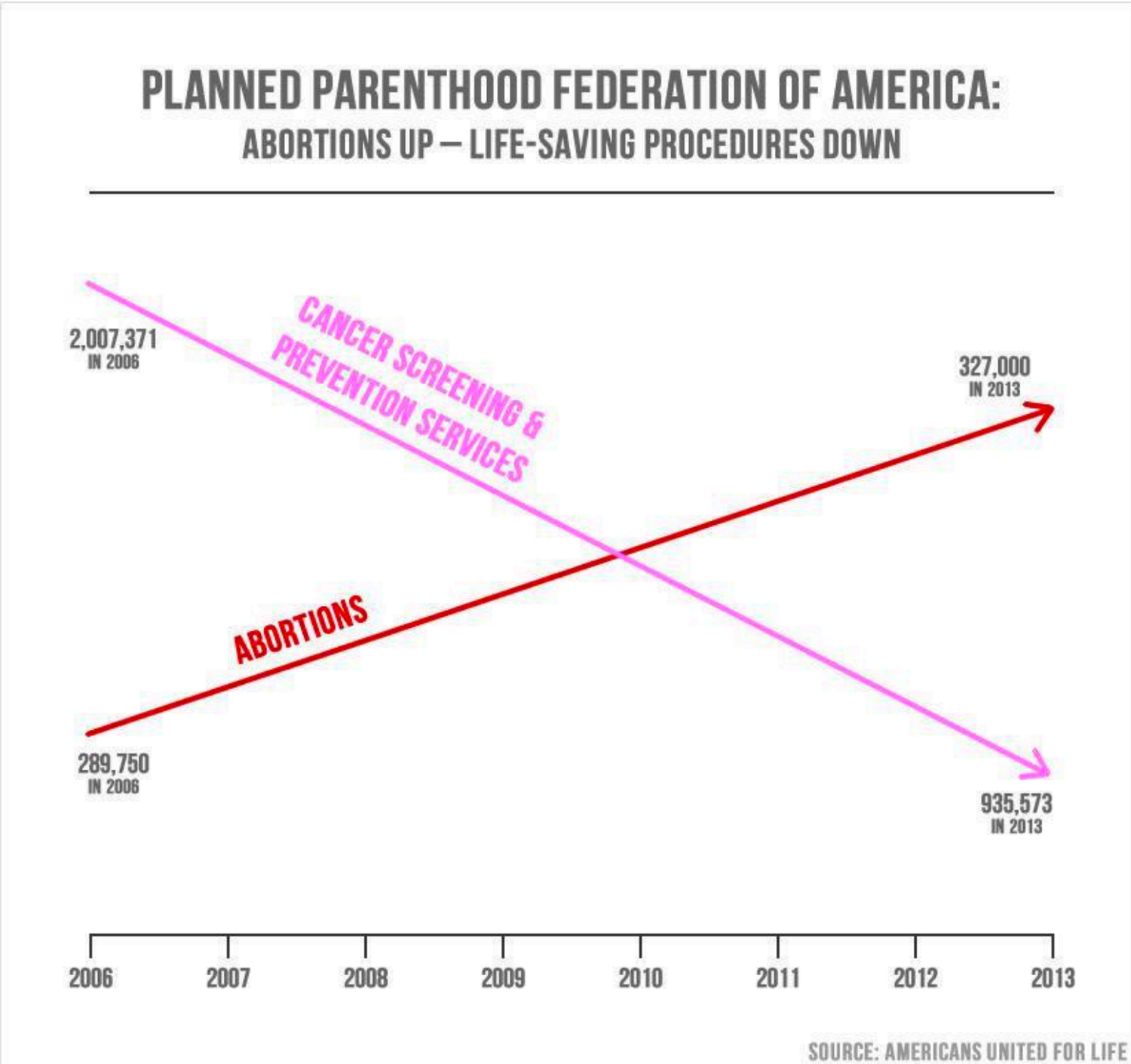
- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

- Show data variation, not design variation.

Integrity: Missing Scales



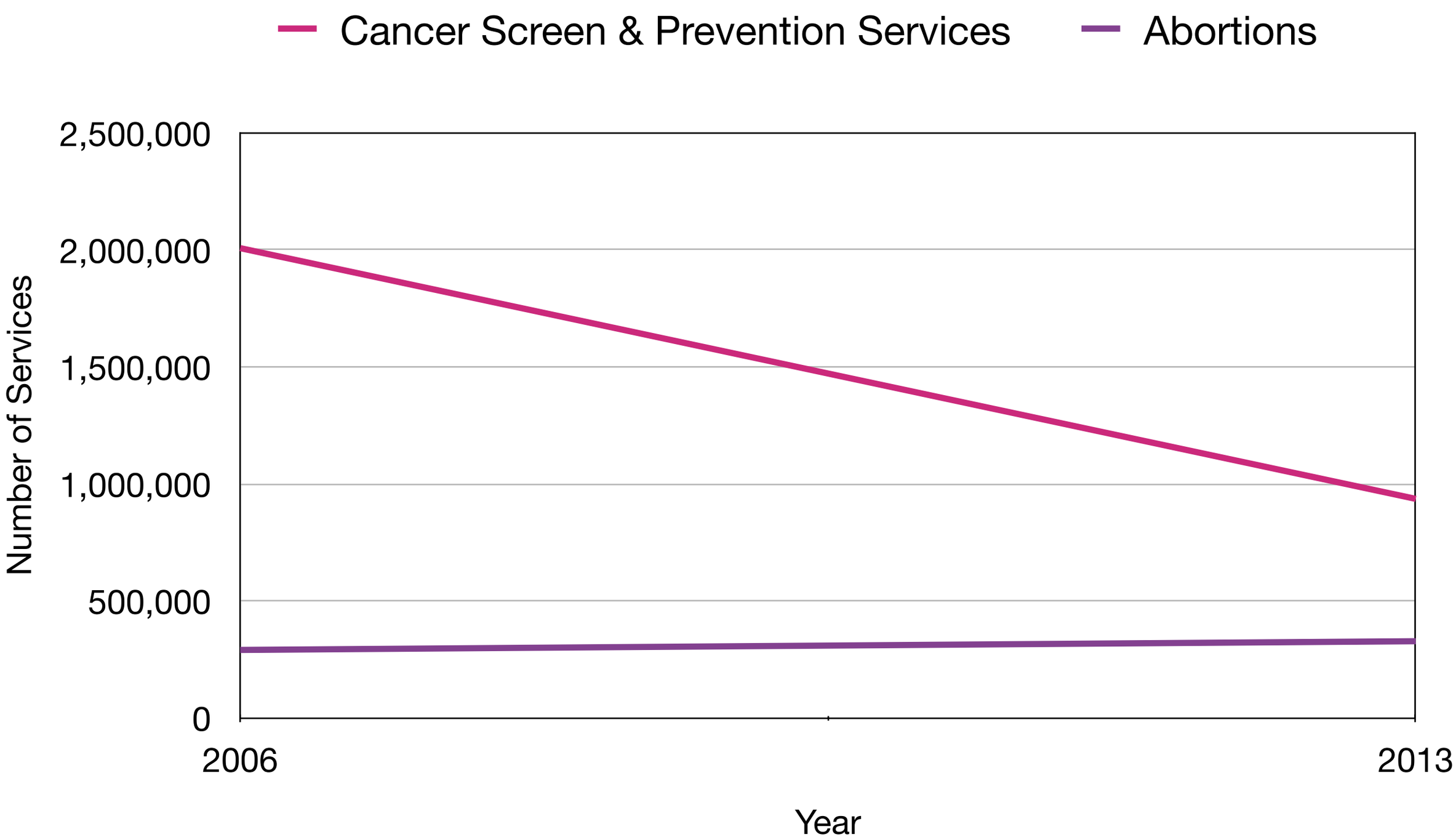
Scale Distortions



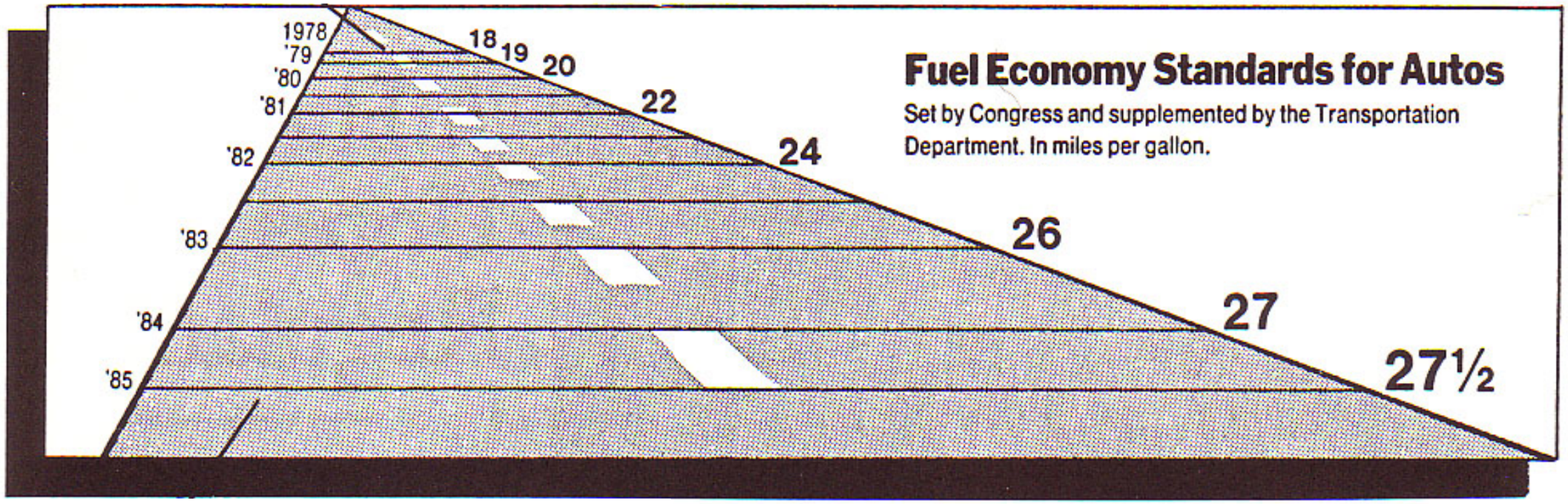
In 2013, did Planned Parenthood do more abortions than cancer screening & prevention services?

https://twitter.com/Bound4LIFE/status/648873097454579712

Scale Distortion: Plotted on the same axis



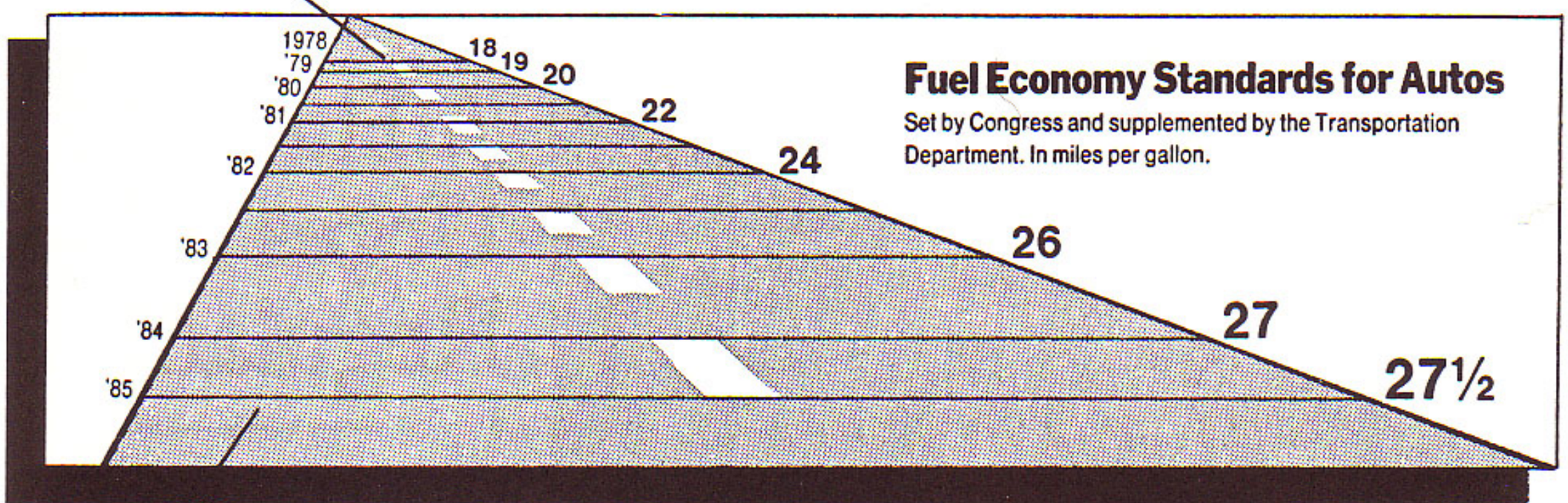
Integrity: Distortions



New York Times, August 9, 1978, p. D-2.

Integrity: Distortions

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.

New York Times, August 9, 1978, p. D-2.

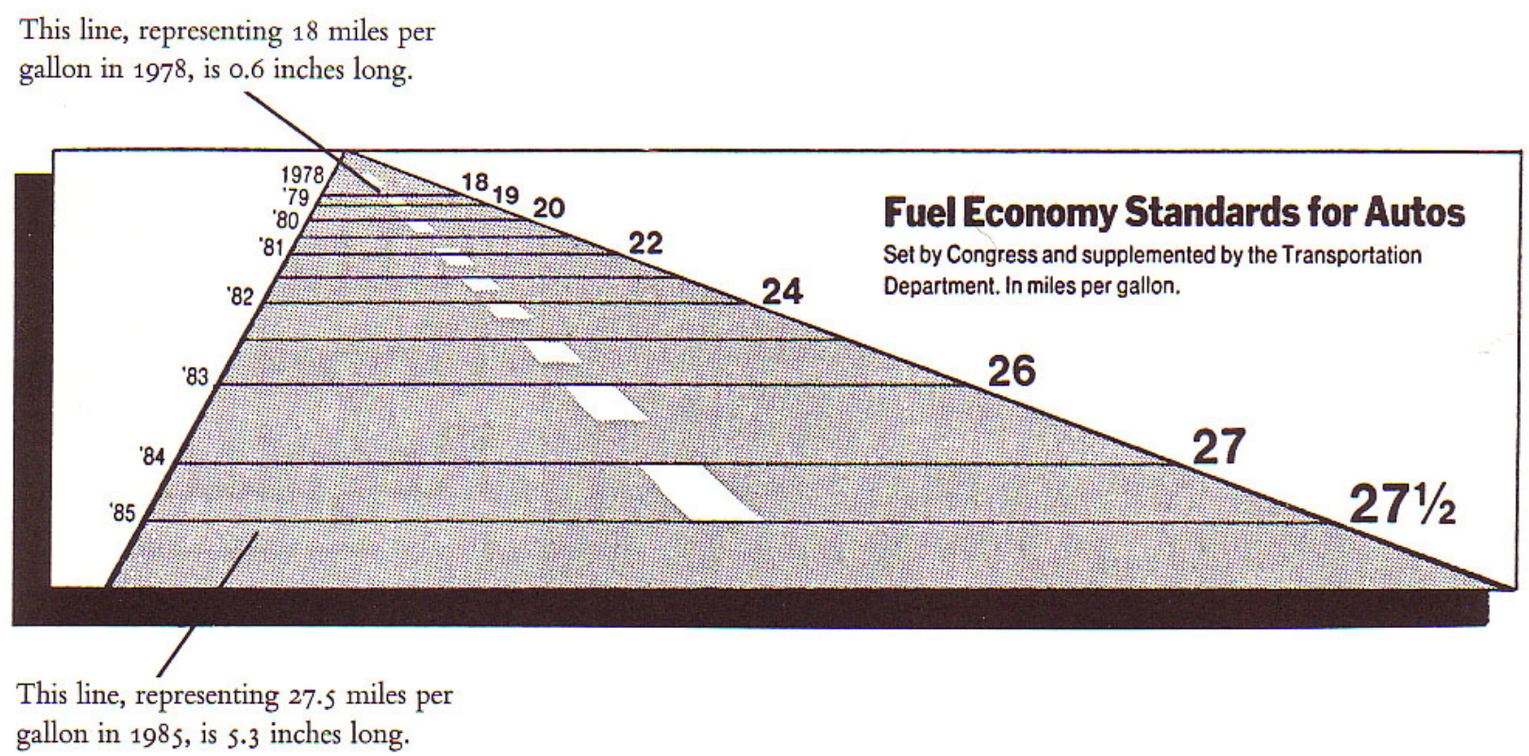
The Lie Factor

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

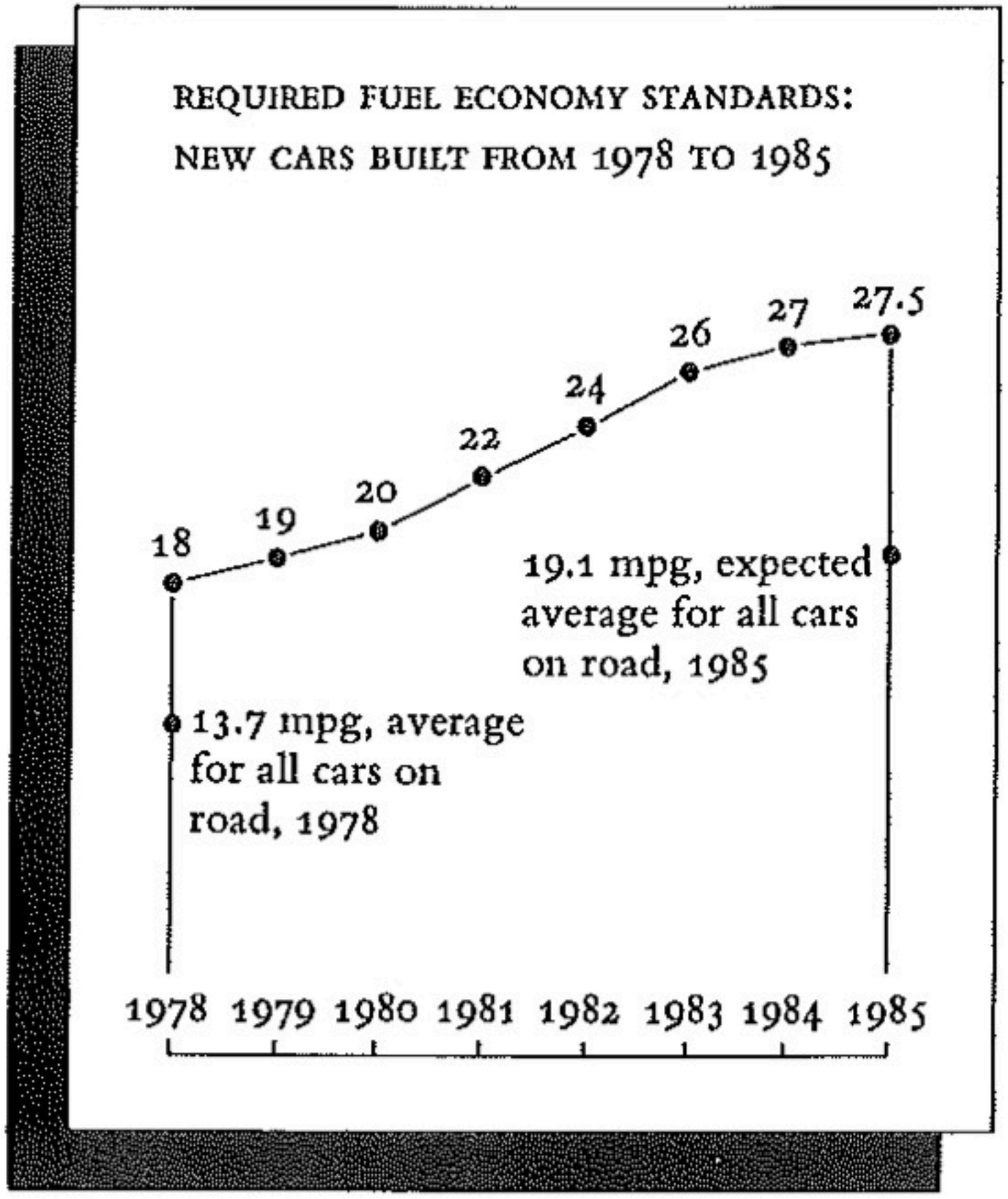
|graph| = (5.3 - 0.6) / 0.6
= 783%

|data| = (27.5 - 18) / 18
= 53%

Lie Factor = 783 / 53
= 14.8

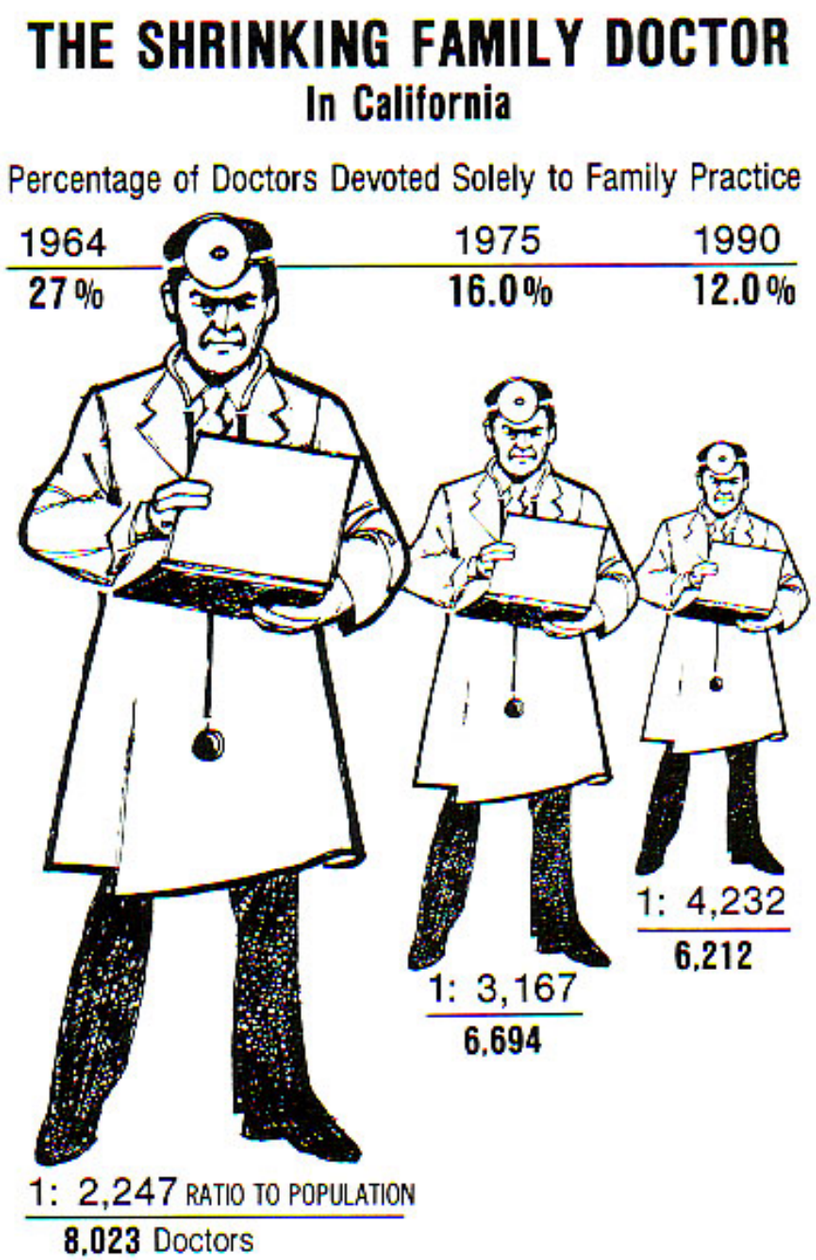


New York Times, August 9, 1978, p. D-2.

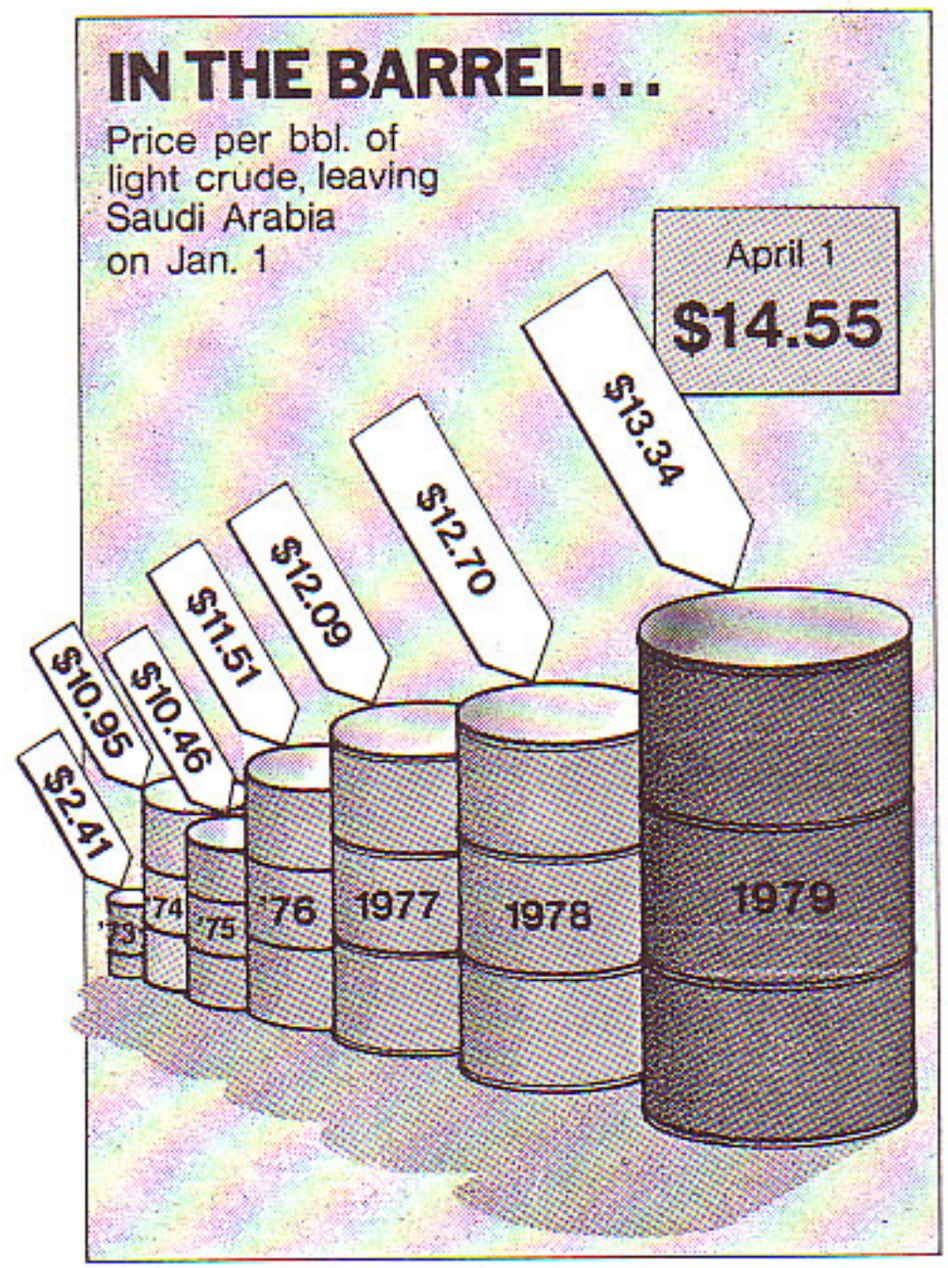

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985
18
19
20
22
24
26
27
27.5
19.1 mpg, expected average for all cars on road, 1985
13.7 mpg, average for all cars on road, 1978
1978 1979 1980 1981 1982 1983 1984 1985

Tufte 2001

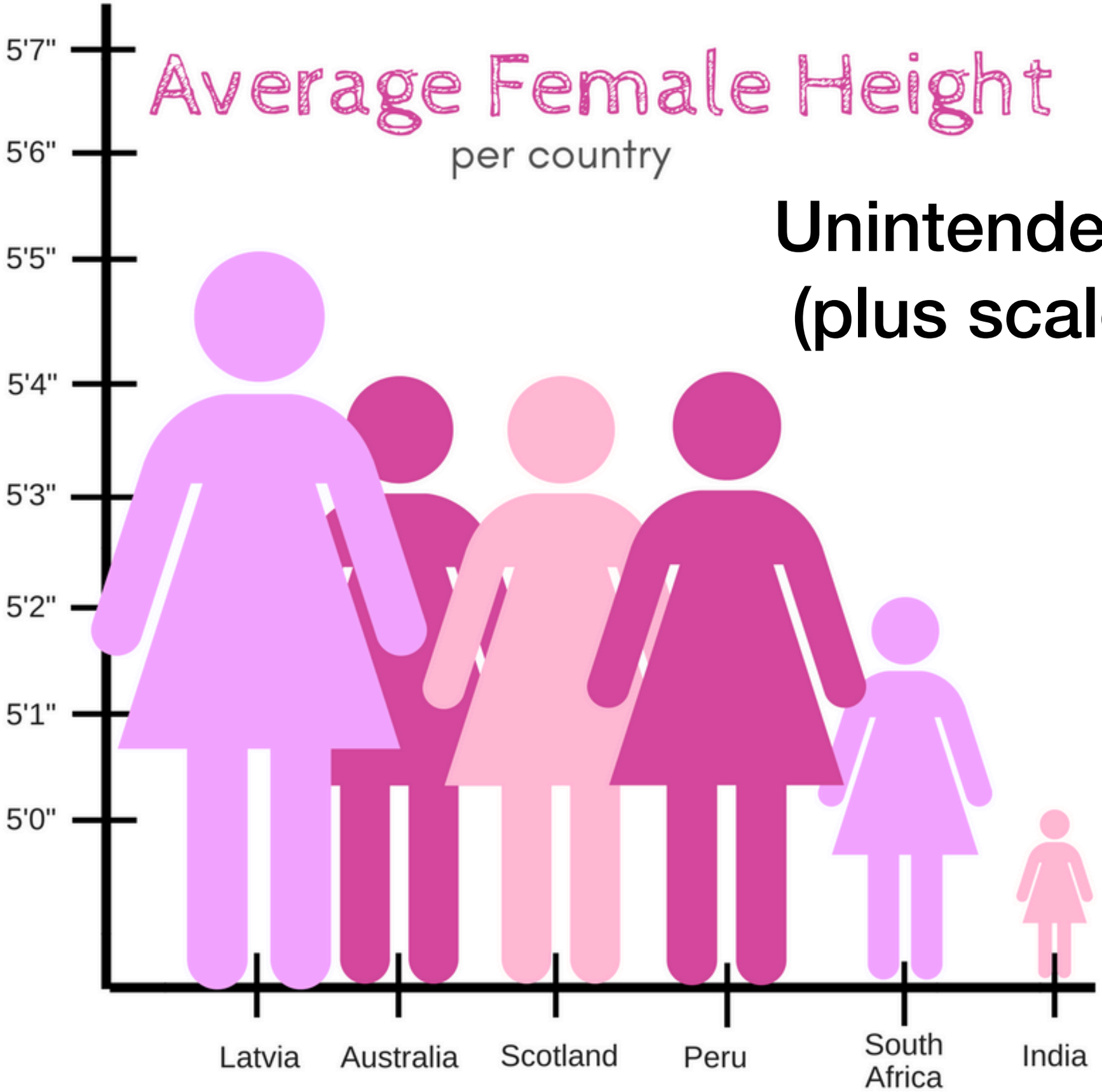
Size Coding: Height/Area/Volume



LA Times, 1979

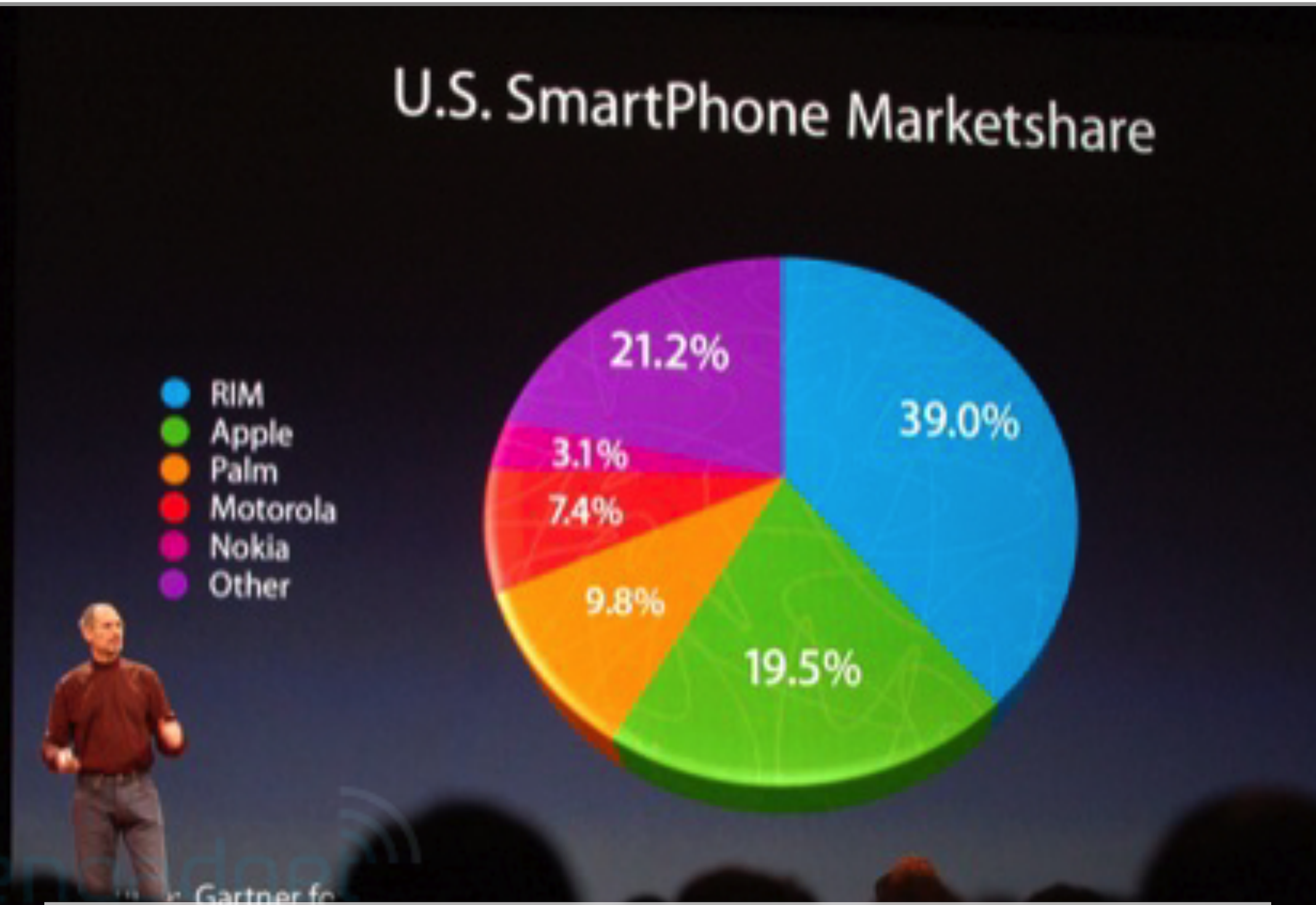


Times, 1979



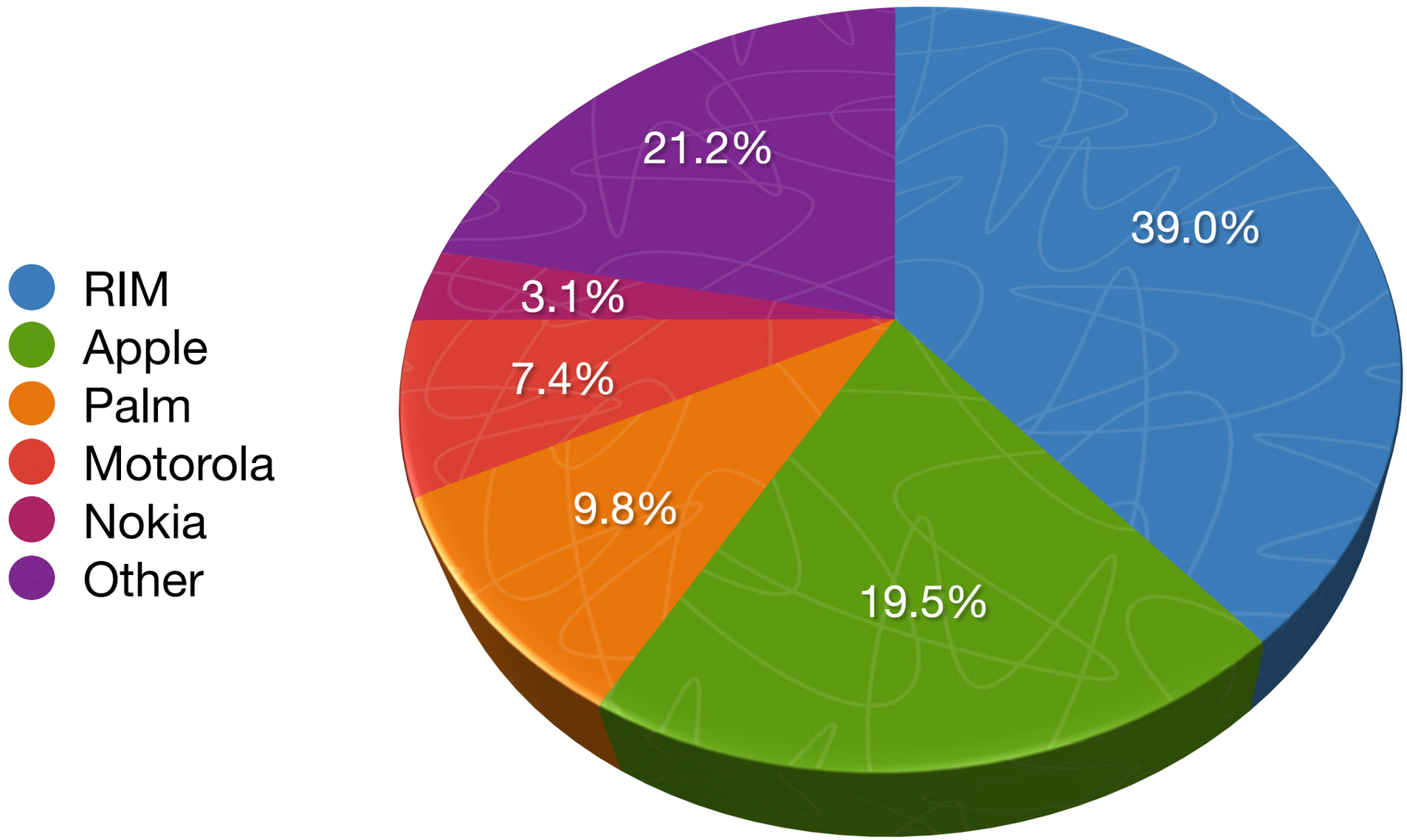
Unintended Size Coding (plus scale distortions!)

https://twitter.com/reina_sabah/status/1291509085855260672

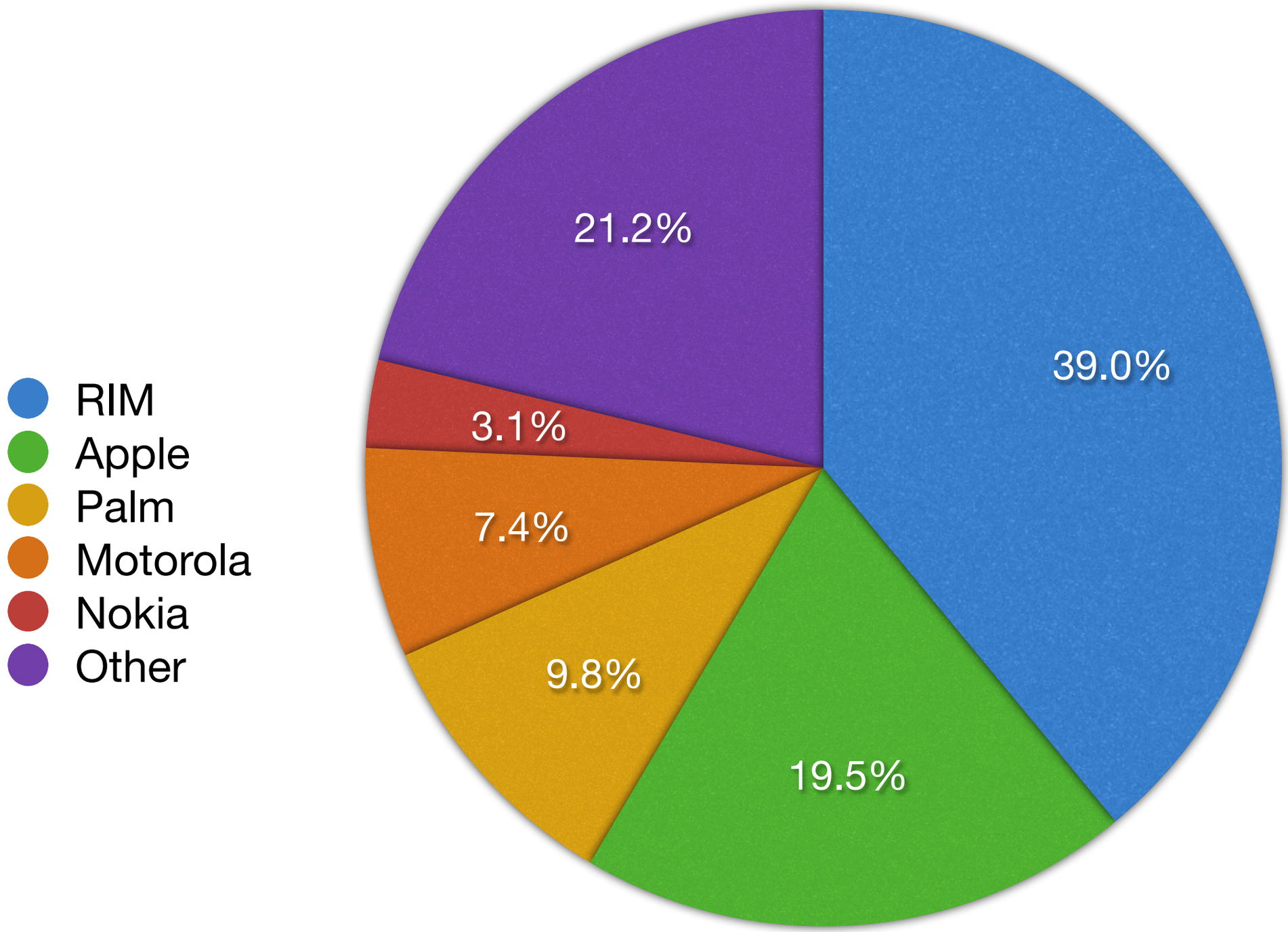


http://www.guardian.co.uk/technology/blog/2008/jan/21/liesdamnliesandstevejobs

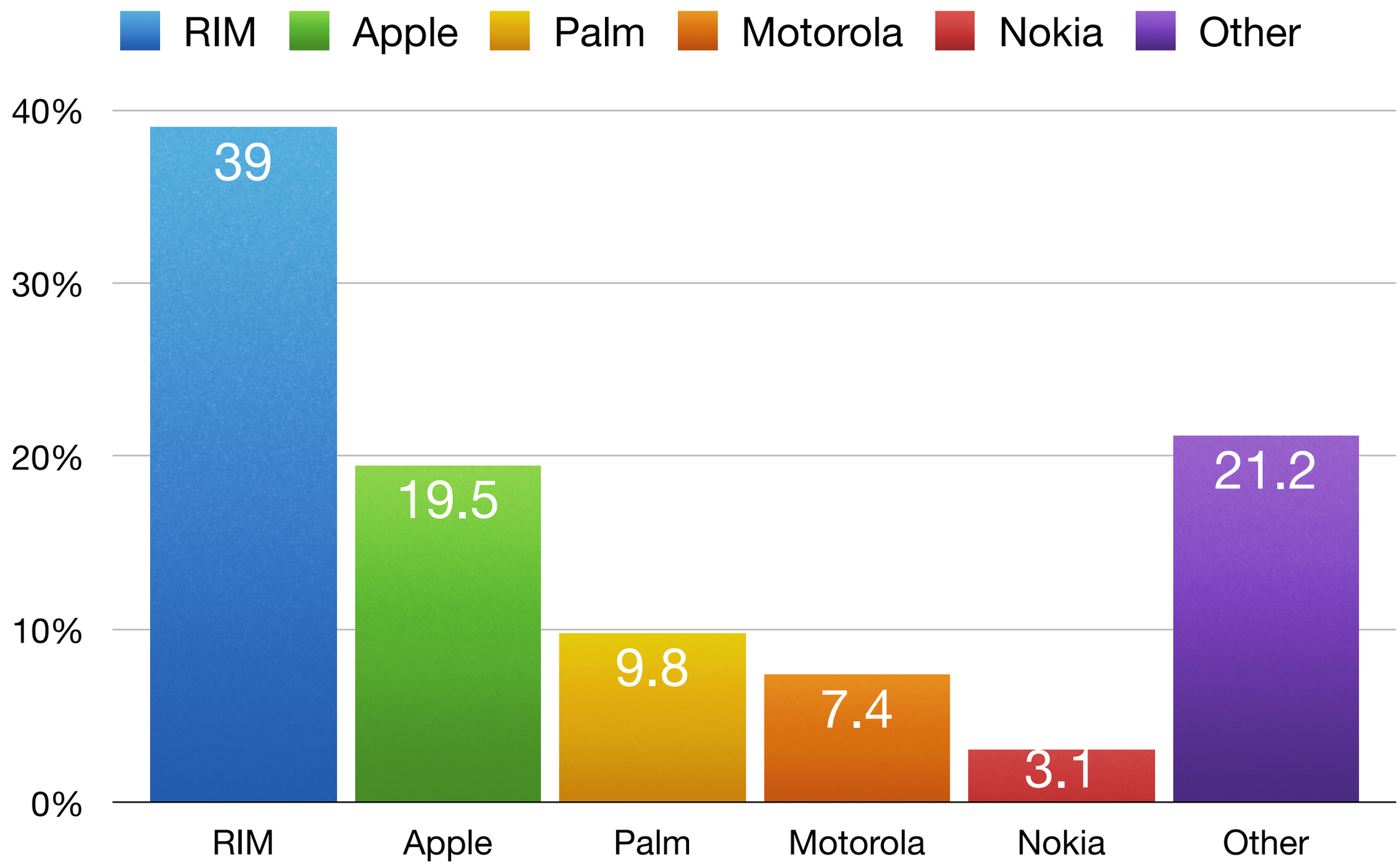
U.S. Smart Phone Marketshare



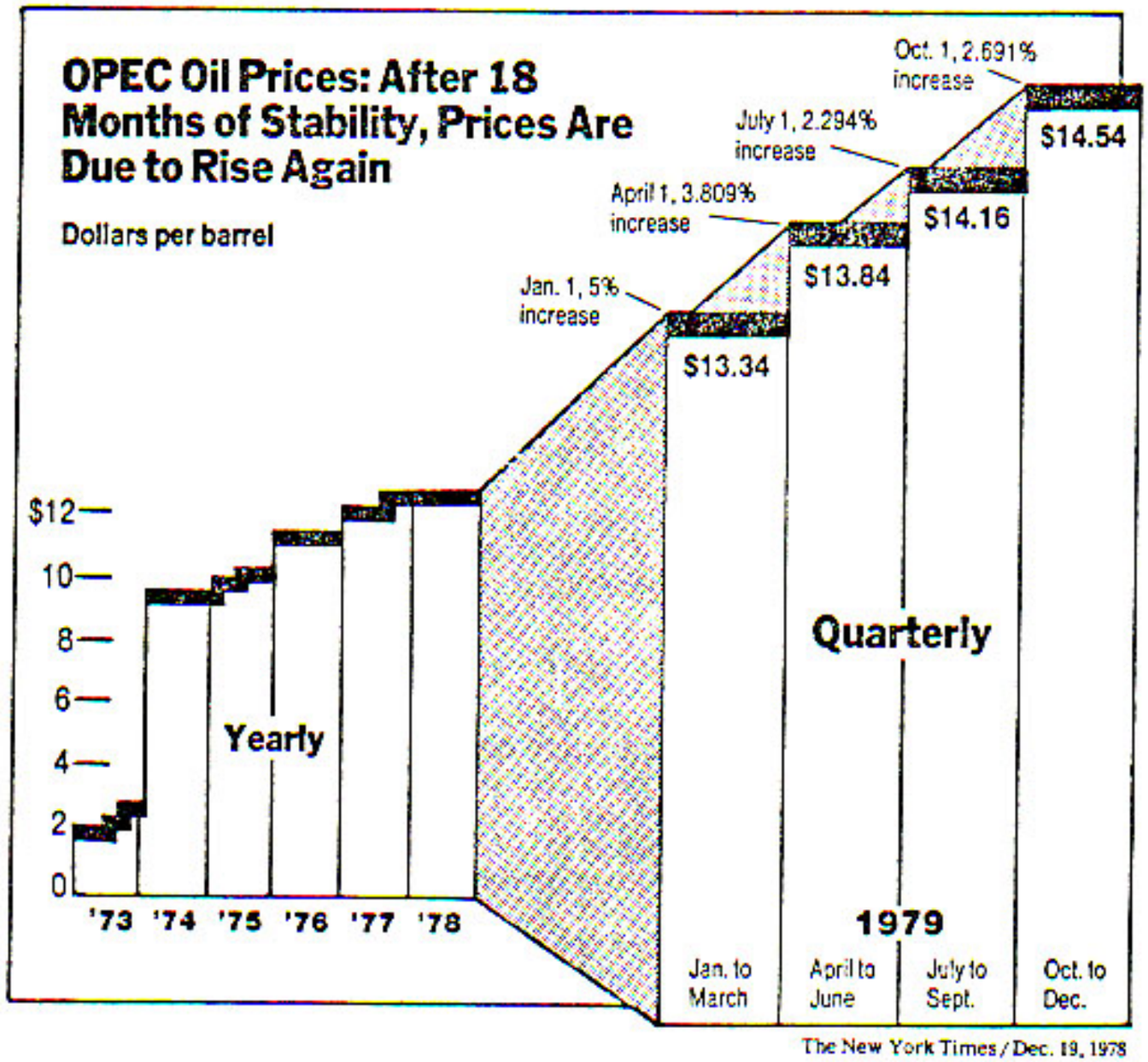
U.S. Smart Phone Marketshare



U.S. Smart Phone Marketshare



Key Idea: Show Data Variation, not Design Variation



1" Equivalents

1973-1978	$8.00
Jan.-Mar.	$4.73
Apr.-June	$4.37
July-Sept.	$4.16
Oct.-Dec.	$3.92

Design Principles

(…or approaches to achieving
Integrity and Excellence)

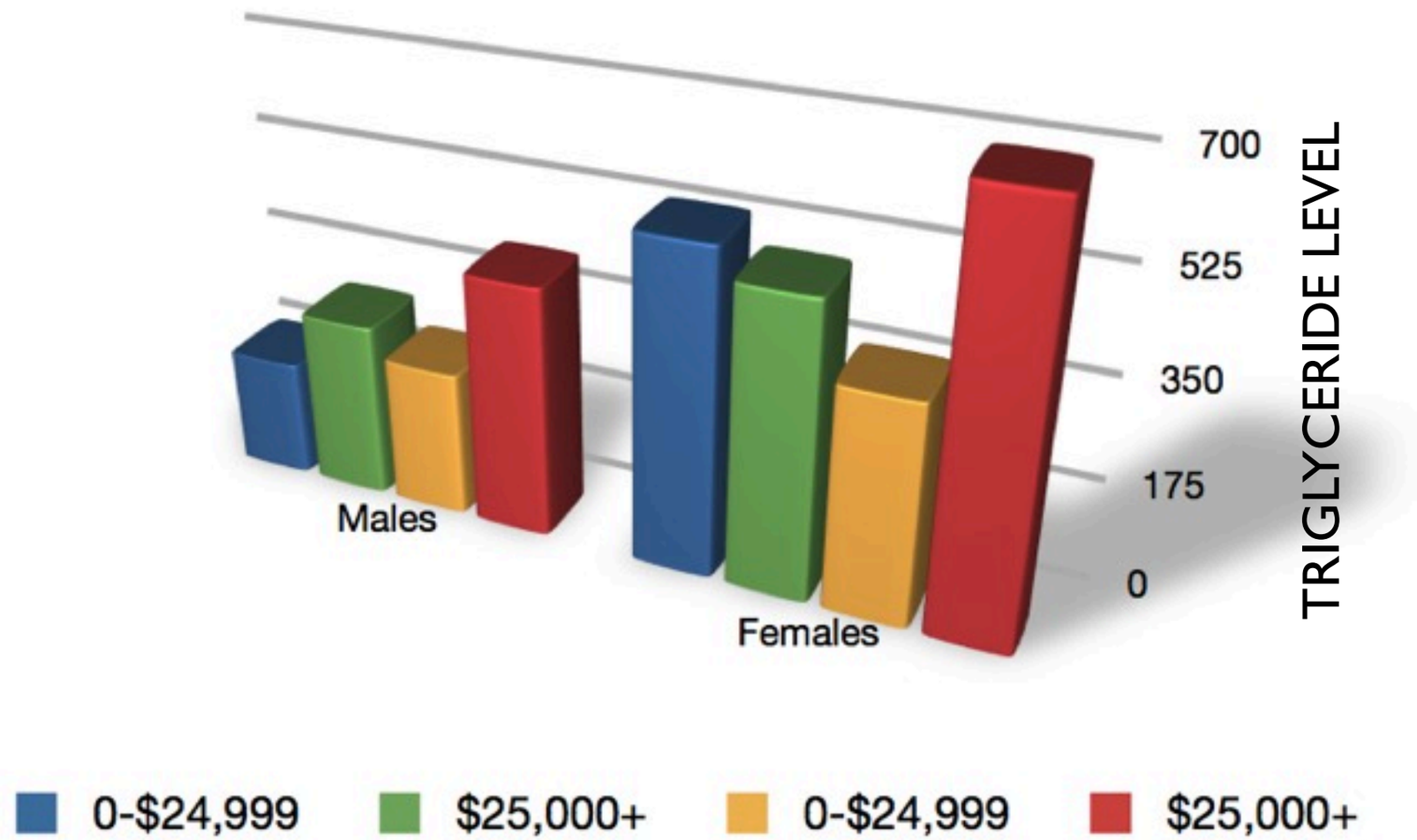
Graphical Excellence

- **Graphical excellence** is that which
  - gives the viewer the greatest number of ideas
  - in the shortest time
  - with the least ink
  - in the smallest space

A. Einstein, “An explanation should be as simple as possible, but no simpler.”

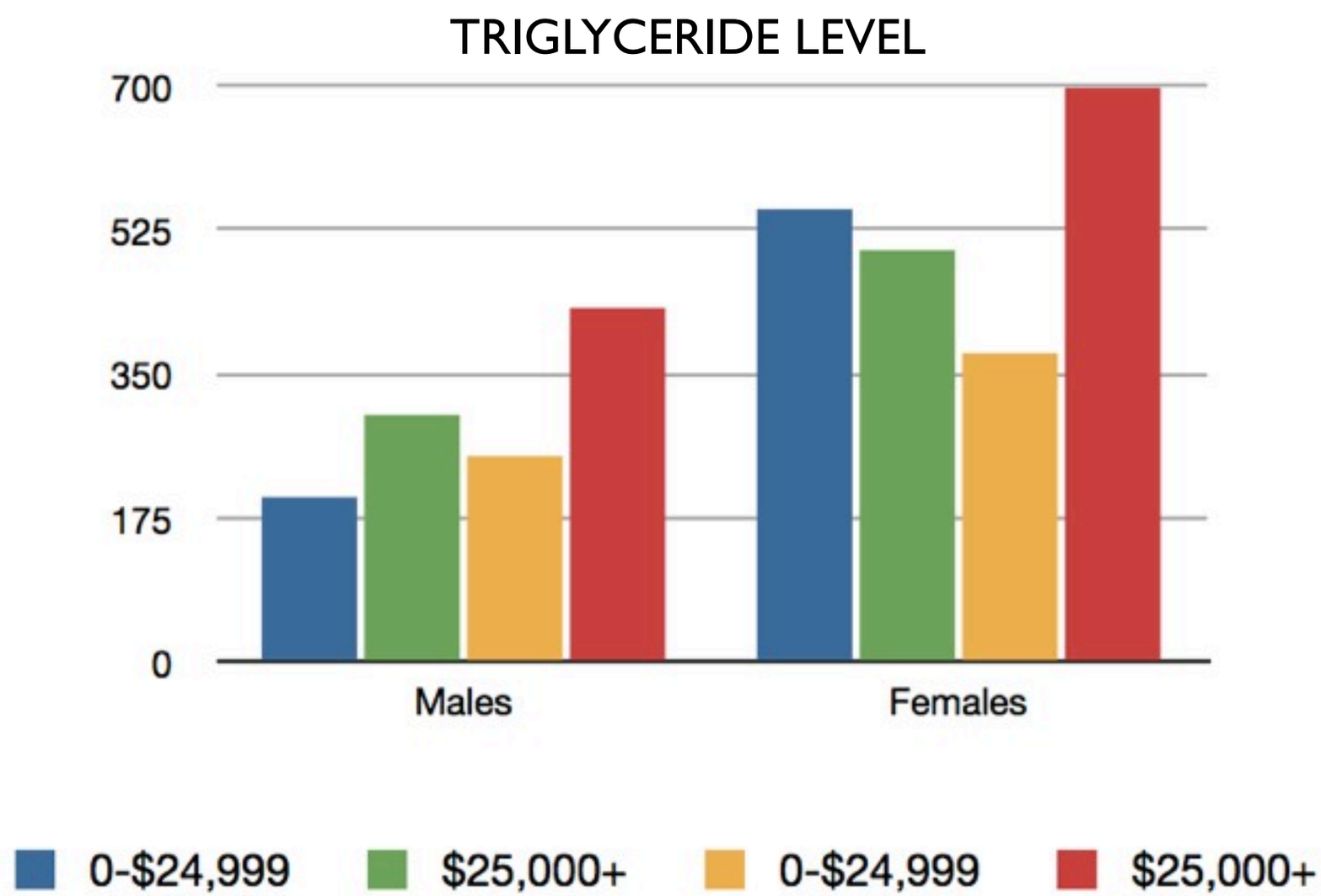
Maximize Data-Ink Ratio

$$\textbf{Data-Ink Ratio} = \frac{\textbf{Data Ink}}{\textbf{Total Ink Used in Graphic}}$$

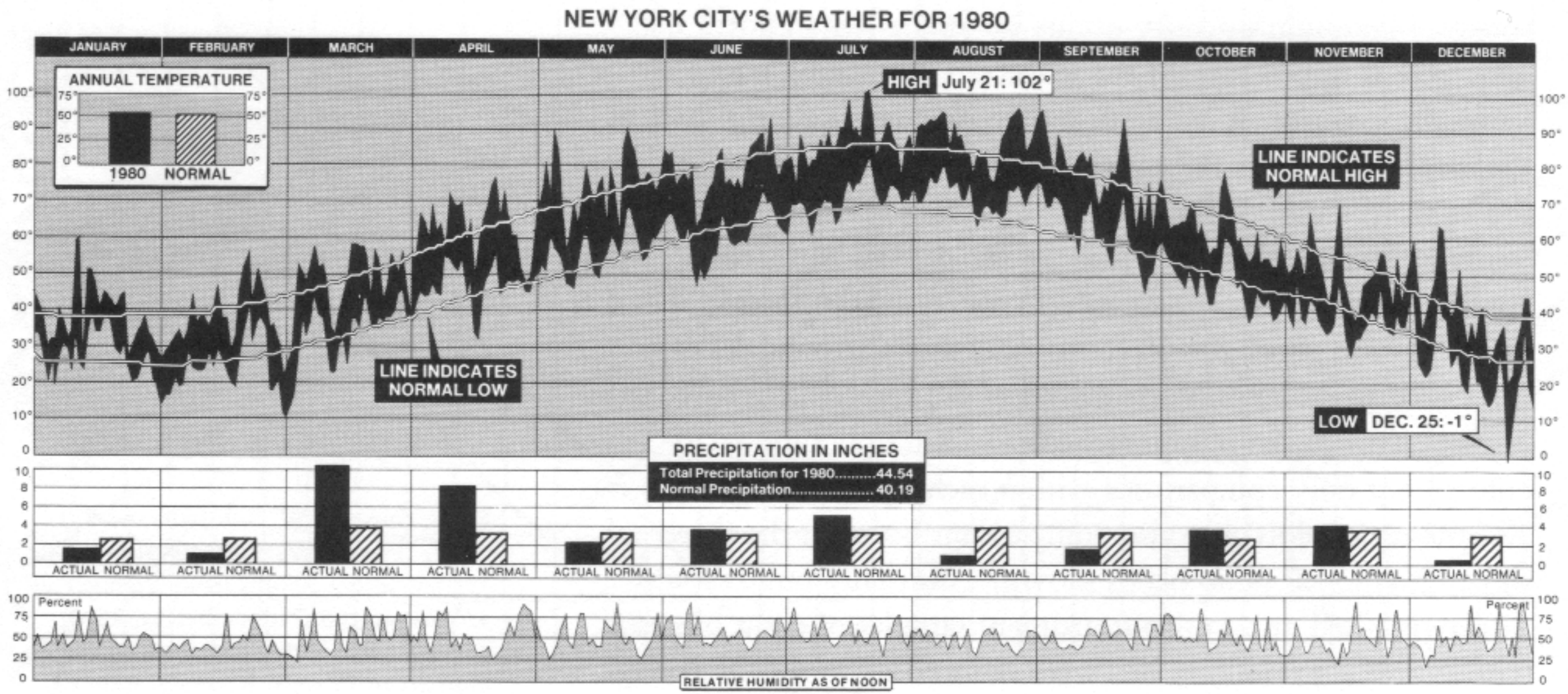


Maximize Data-Ink Ratio

$$\textbf{Data-Ink Ratio} = \frac{\textbf{Data Ink}}{\textbf{Total Ink Used in Graphic}}$$

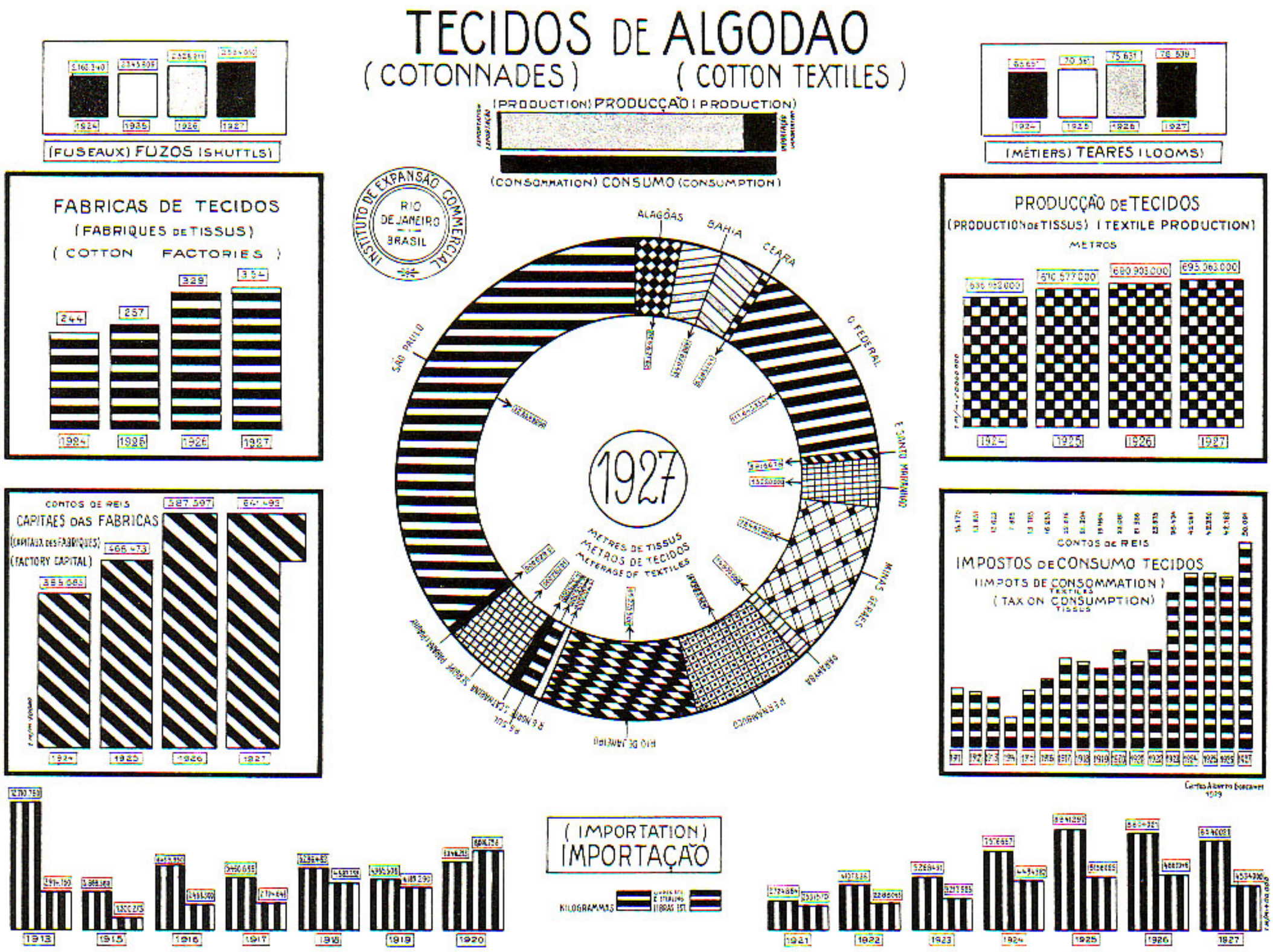


Data-Ink Ratio
Good or Bad?

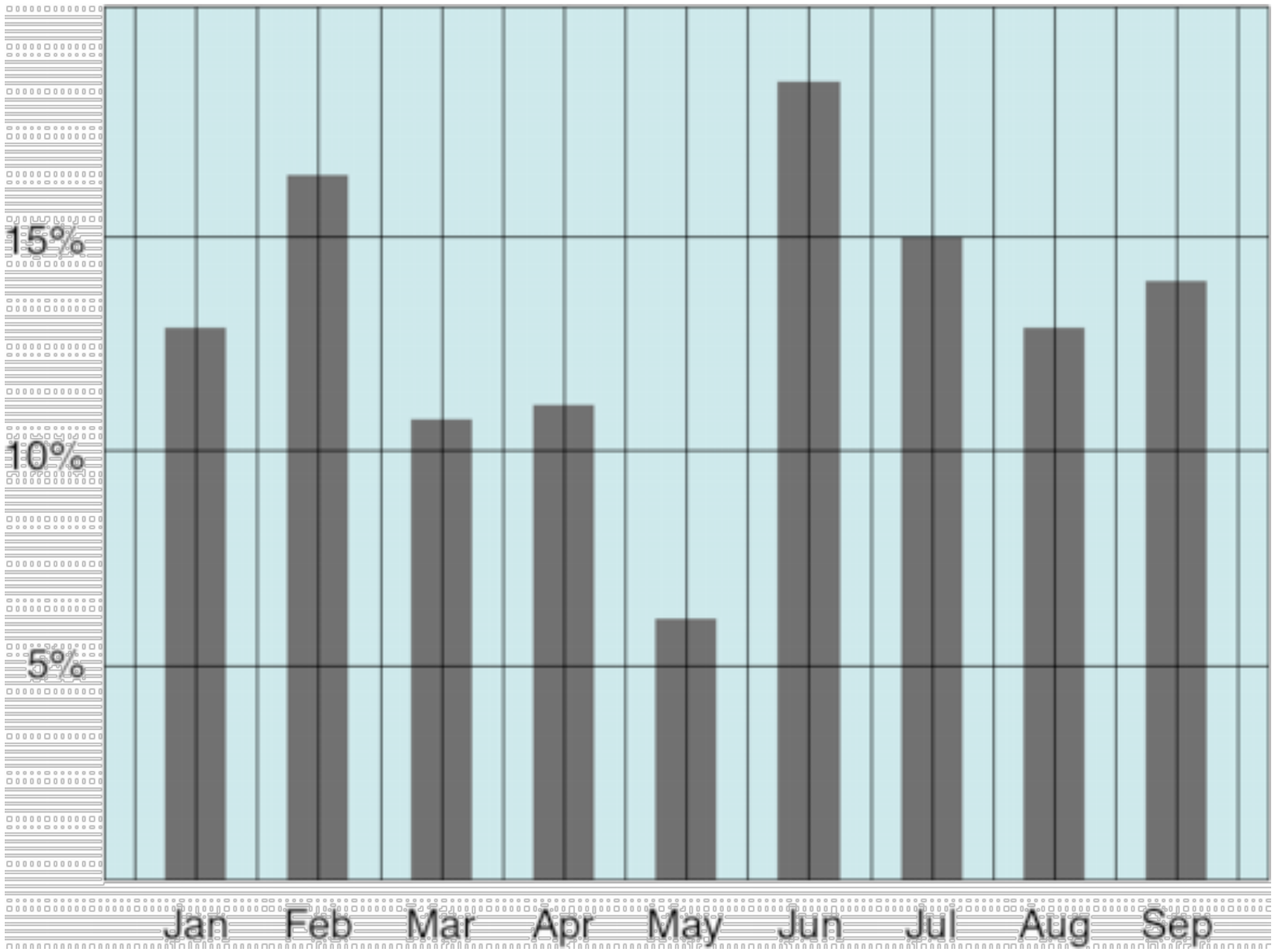


New York Times, January 11, 1981, p. 32.

Avoid Chartjunk: Remove extraneous visual elements that distract from the message

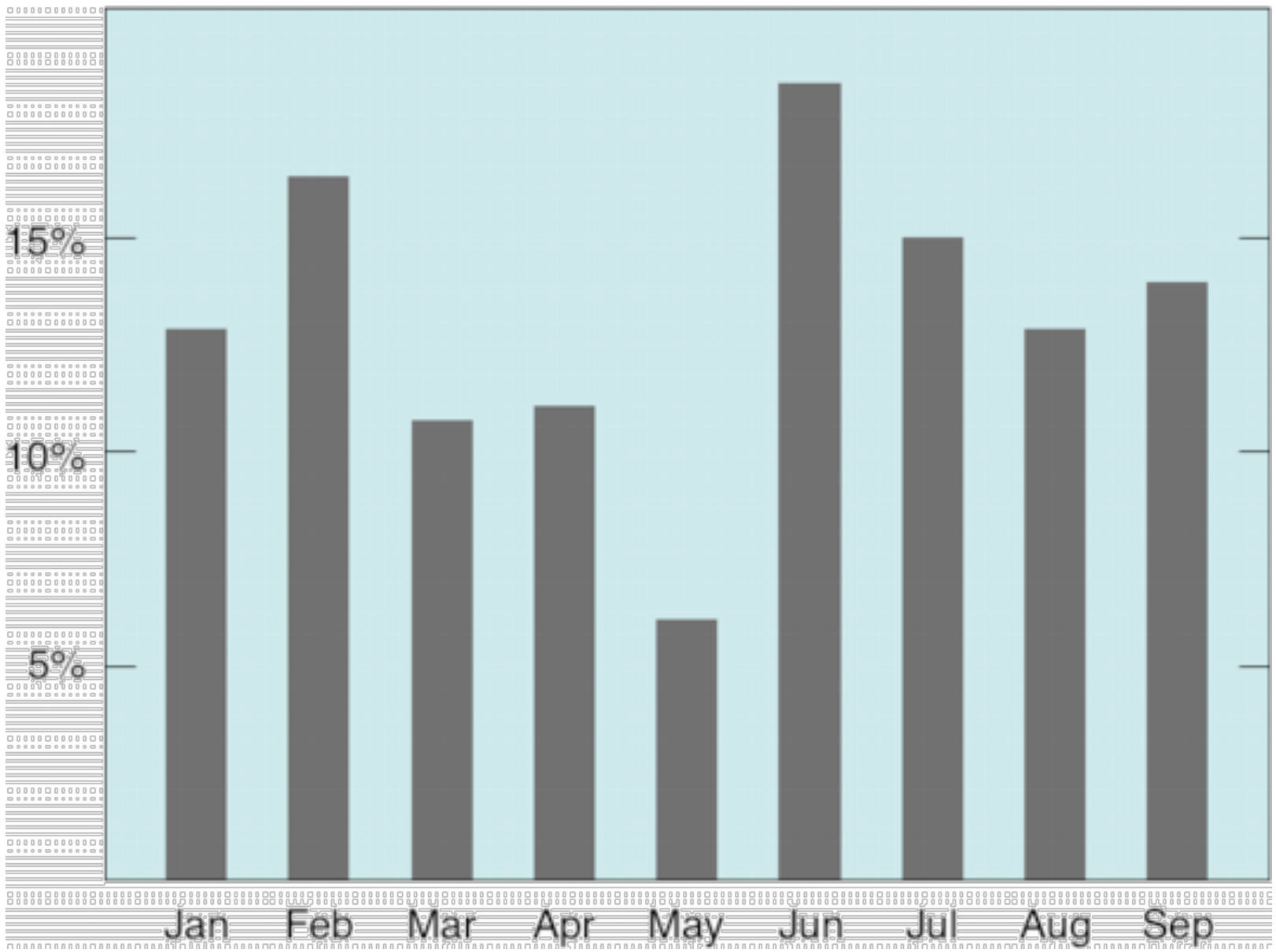


Avoiding Chartjunk



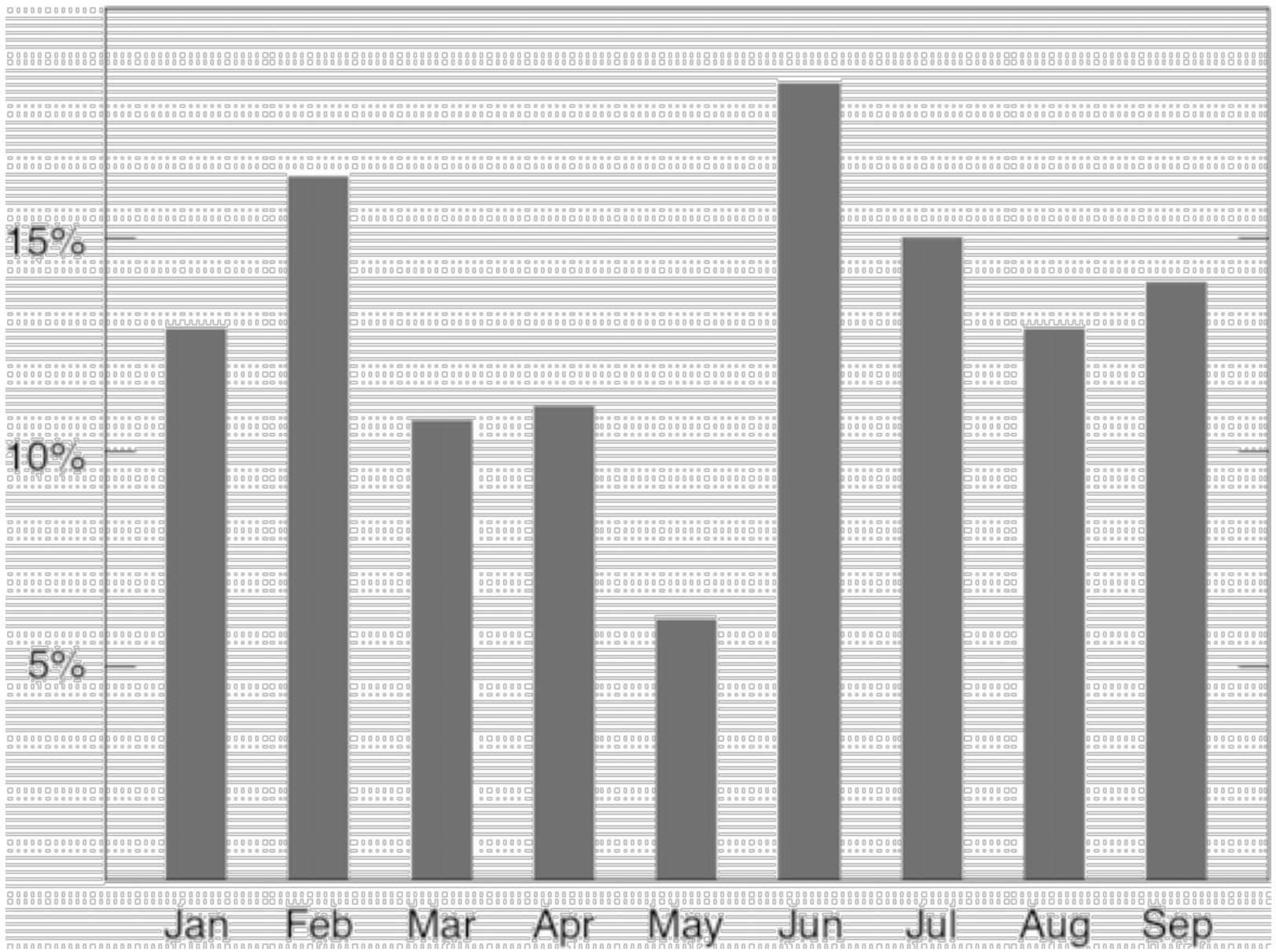
http://www.tbray.org/ongoing/data-ink/di1

Avoiding Chartjunk



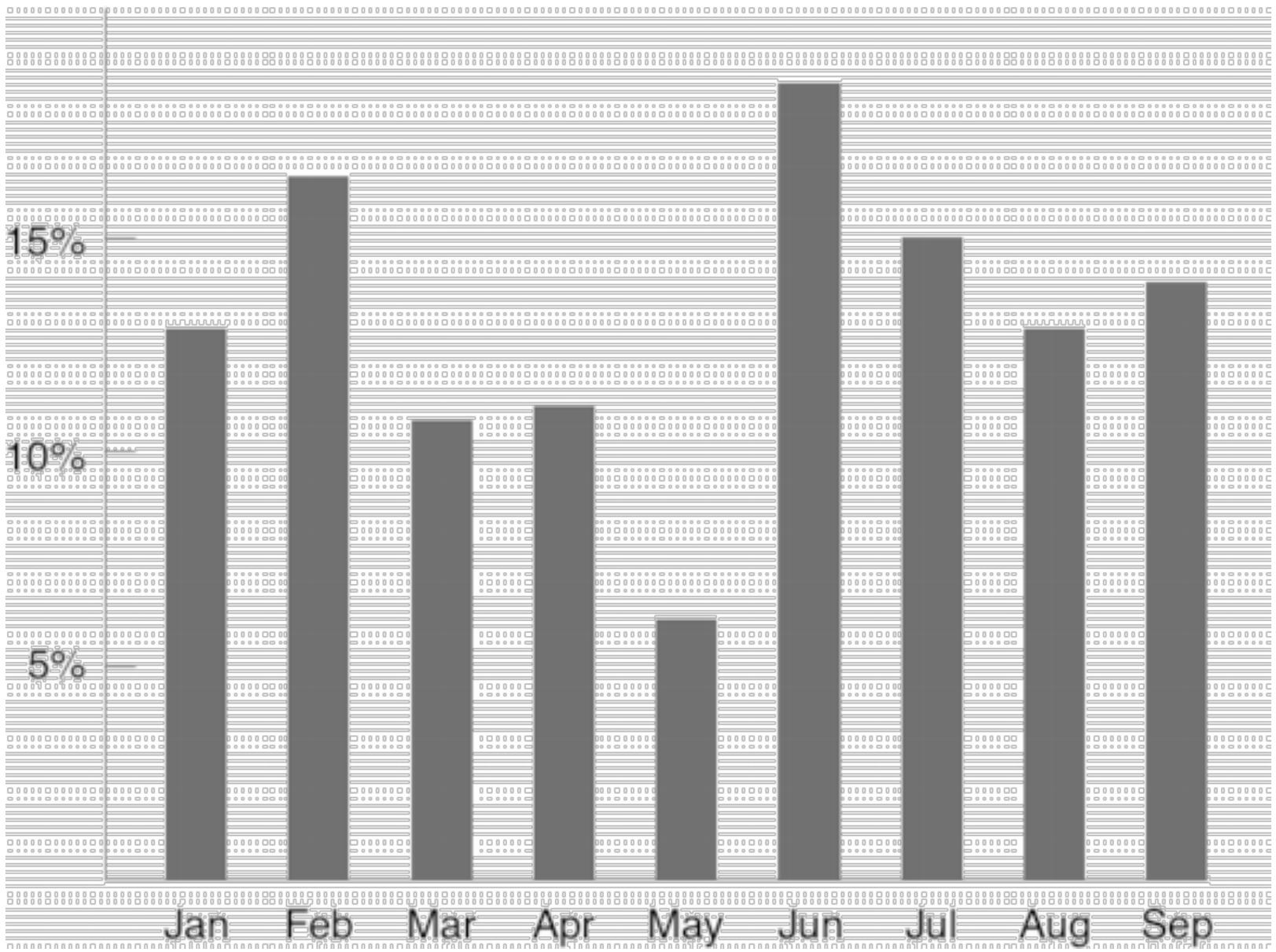
http://www.tbray.org/ongoing/data-ink/di1

Avoiding Chartjunk



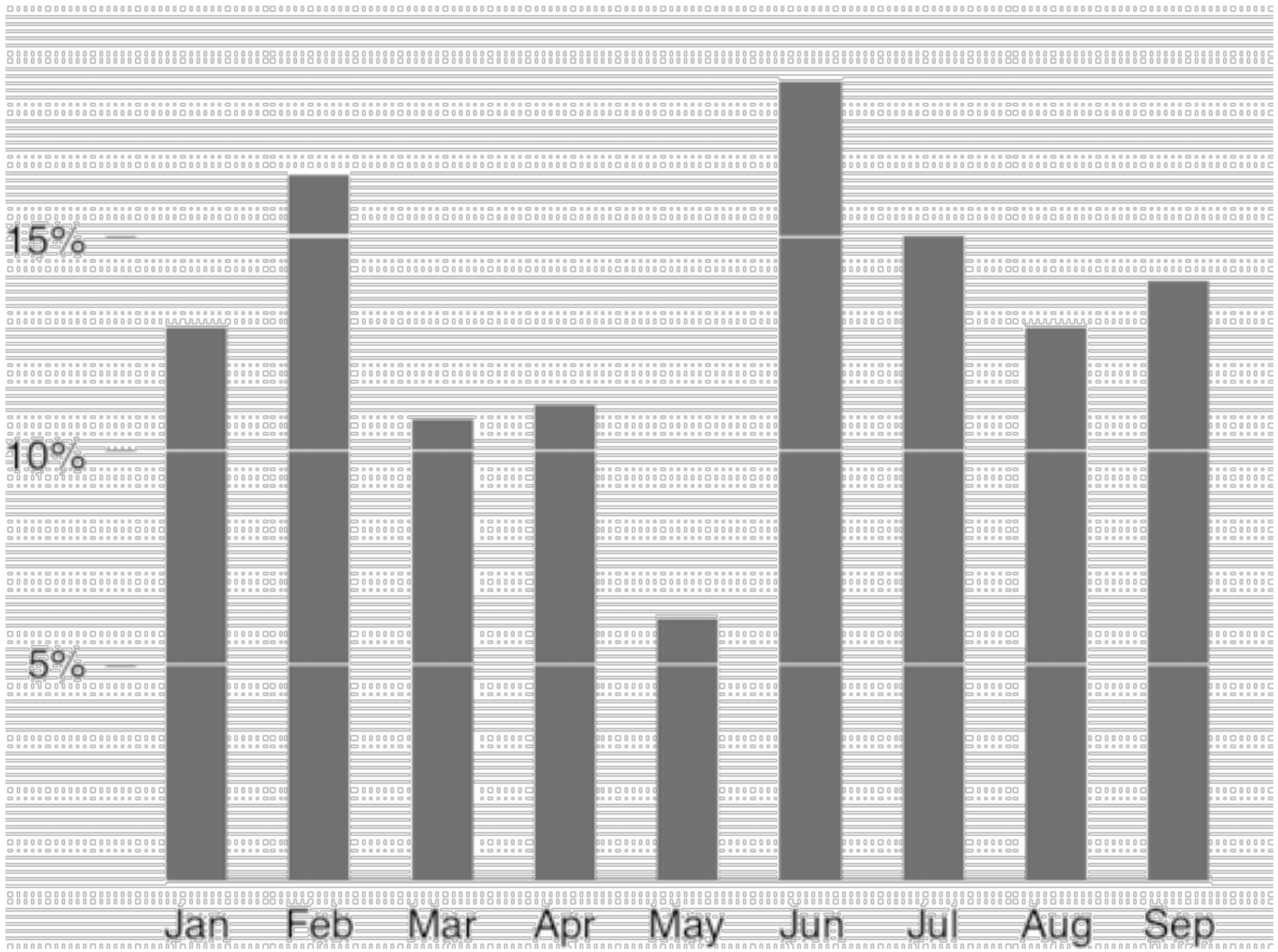
http://www.tbray.org/ongoing/data-ink/di1

Avoiding Chartjunk



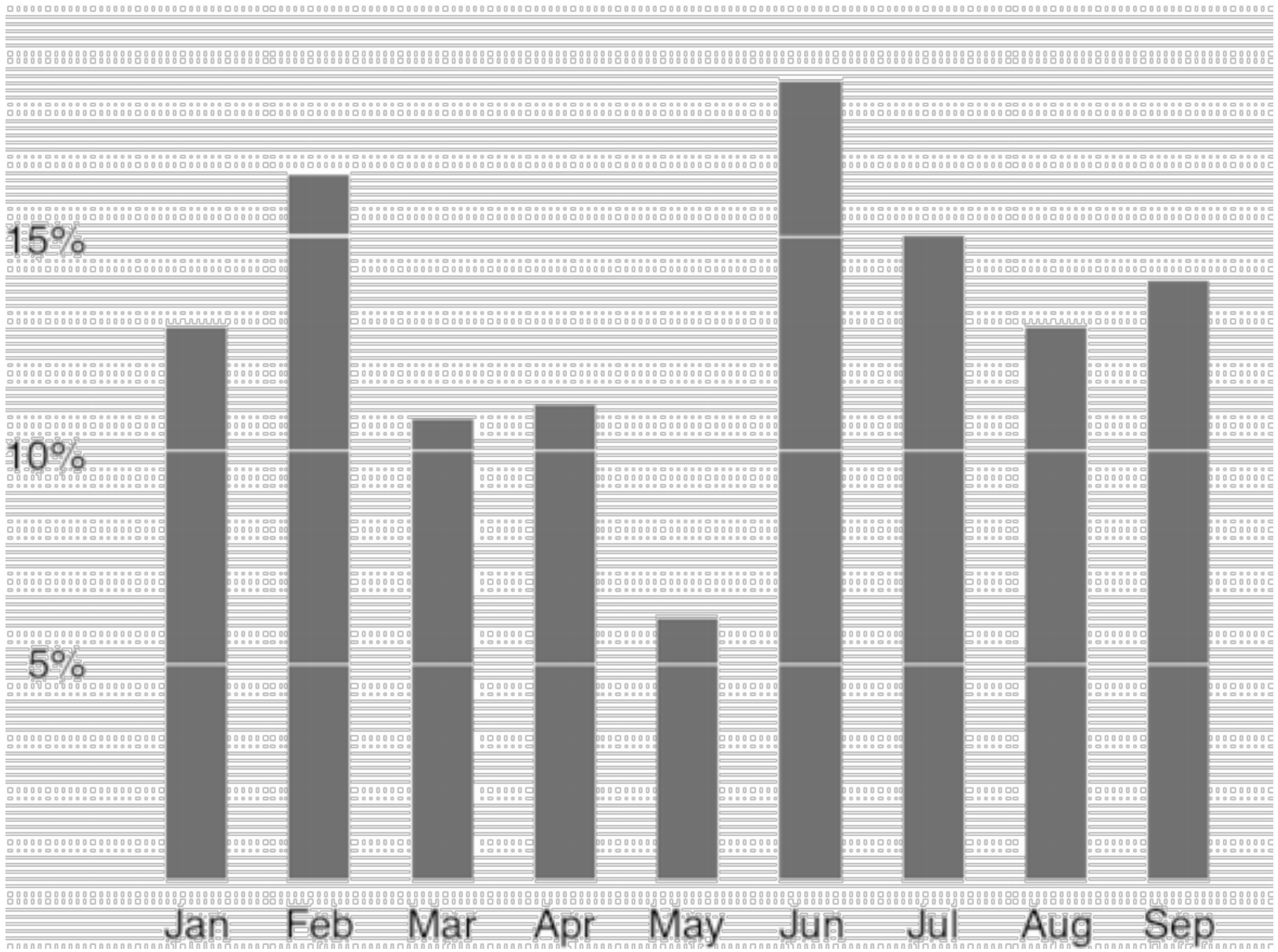
http://www.tbray.org/ongoing/data-ink/di1

Avoiding Chartjunk



http://www.tbray.org/ongoing/data-ink/di1

Avoiding Chartjunk



http://www.tbray.org/ongoing/data-ink/di1

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce 'chart junk' – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).

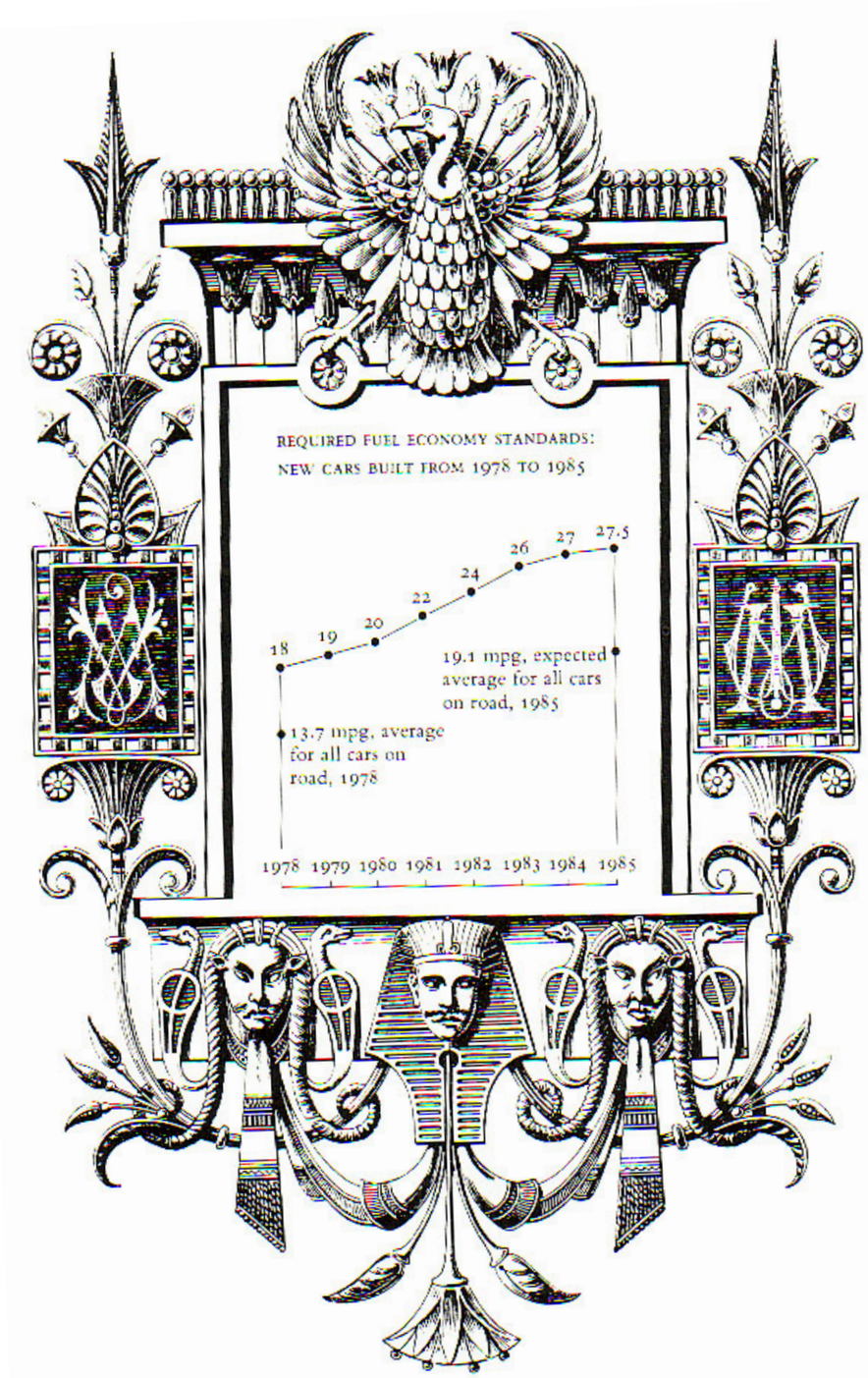
Is Chartjunk Harmful?

Pros

- Persuasion
- Memorability
- Engagement

Cons

- Unbiased Analysis
- Trustworthiness
- Interpretability
- Space Efficiency



Make Use of Multifunctioning Elements

- Graphical elements that convey data information and a design function

Building data measures out of the data increases the quantitative detail and dimensionality of a graphic. The stem-and-leaf plot constructs the distribution of a variable with numbers themselves:

0 | 9 = 900 feet

0	98766562
1	97719630
2	6998776654442221100985 0
3	876655412099551426
4	9998844331929433361107
5	976666665544222100977 31
6	898665441077761065
7	98855431100652108073
8	653322122937
9	377655421000493
10	0984433165212
11	4963201631
12	45421164
13	47830
14	00
15	676
16	52
17	92
18	5
19	39730

Stem-and-leaf displays:
heights of 218 volcanoes, unit 100 feet.

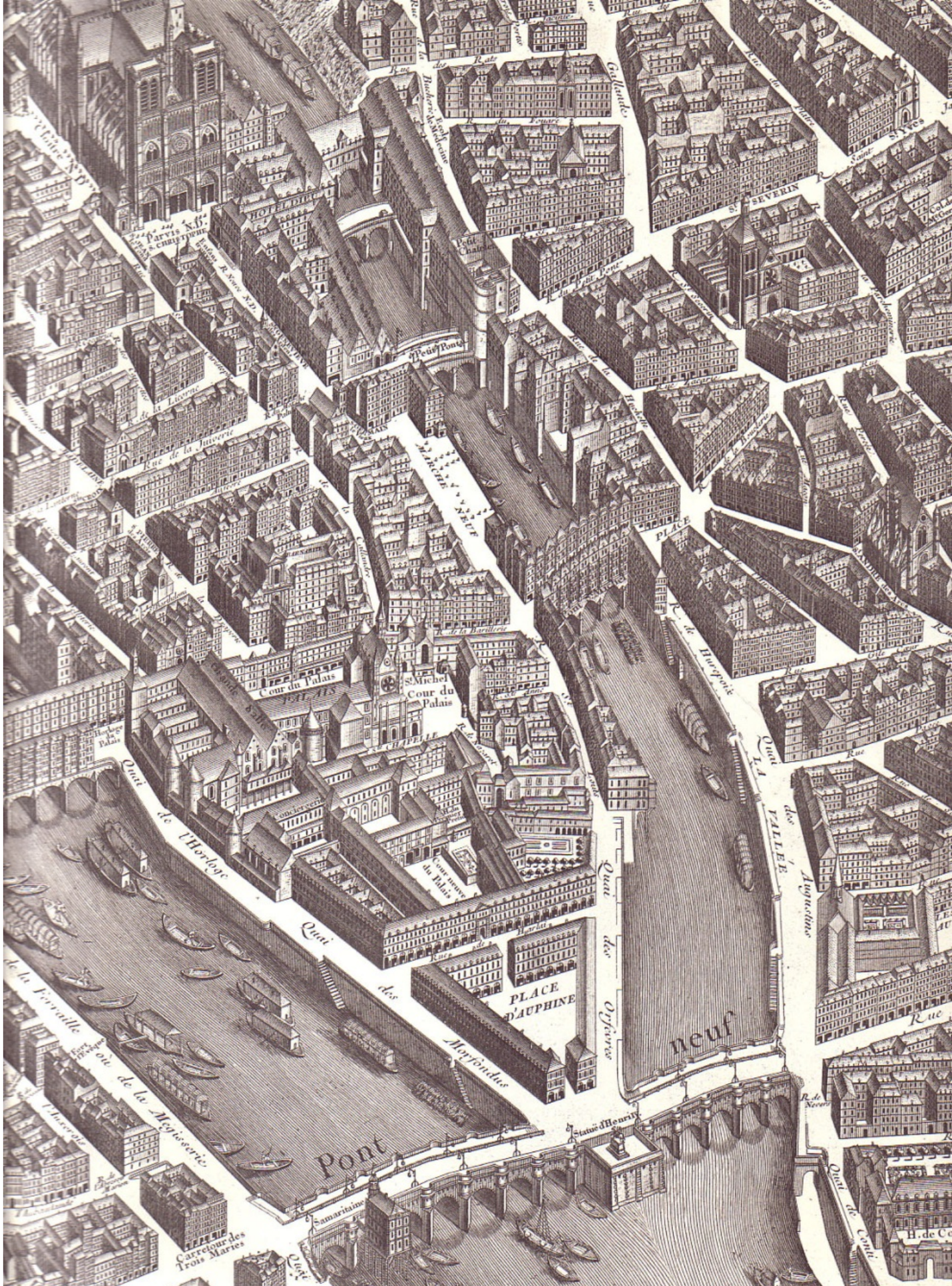
19 | 3 = 19,300 feet

Make Use of Multifunctioning Elements

- Graphical elements that convey data information and a design function

Brian L. Joiner, "Living Histograms," *International Statistical Review*, 43 (1975), 339–340. But, for further developments, see Mark F. Schilling, Ann E. Watkins, and William Watkins, "Is Human Height Bimodal?" *The American Statistician*, 56 (August 2002), 223–229.

Multifunctioning Visualizations: Macro/Micro Readings



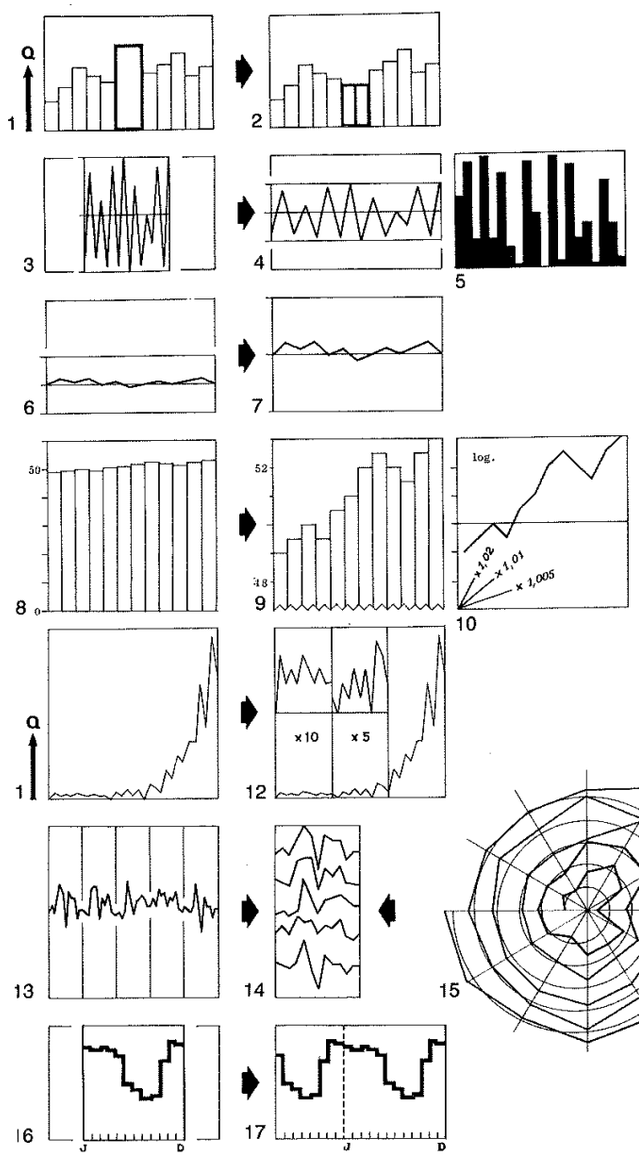
Bretez-Turgot, "Plan de Paris",1739

Maximize Data Density

$$\textbf{Data Density} = \frac{\textbf{Number of elements of data}}{\textbf{Area of the data graphic}}$$

How? Shrink the graphics

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 "reference" information. The background involves: (a) the invariant, highlighted by a heading (Port St. Michel); (b) the highly visible identification of each component (tonnage and dates). 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

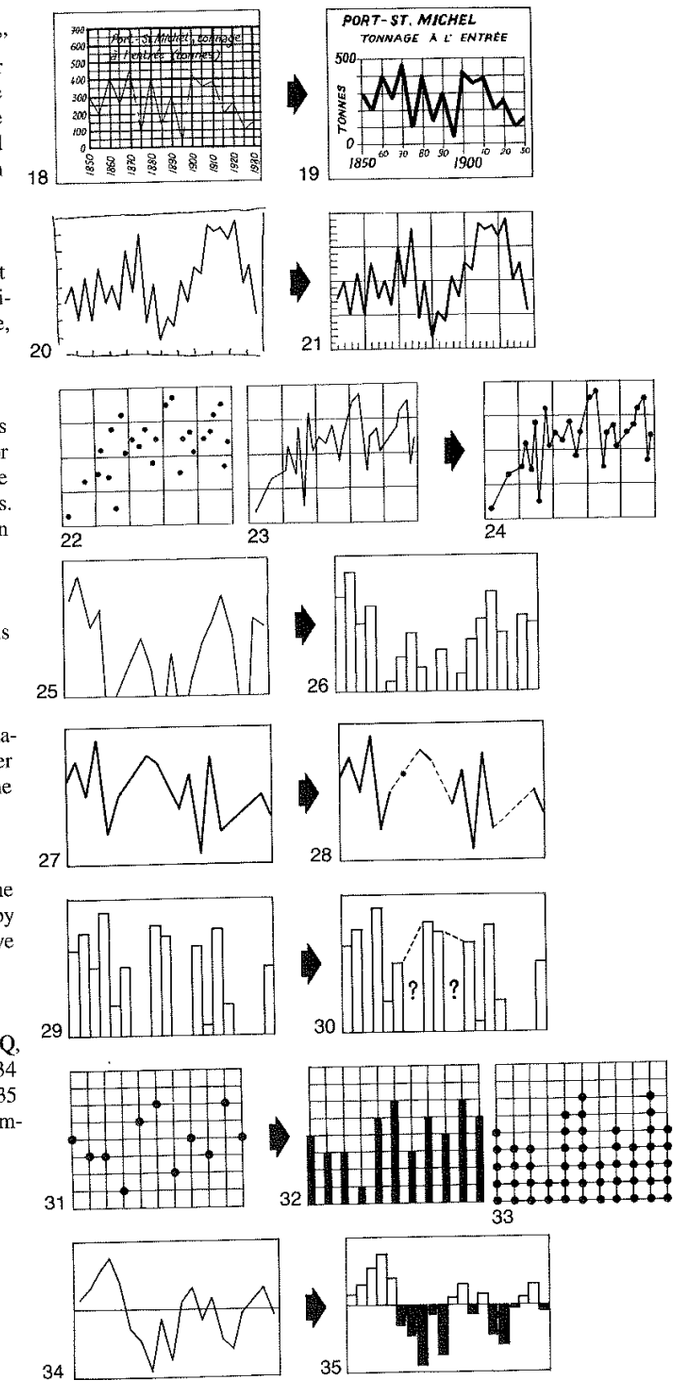
The drawing must indicate the unknowns of the information in an unambiguous way (figures 28 and 30).The reader might interpret figure 27 as a change in the structure of the curve and figure 29 as involving null values.

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**, ≠ (+ –), 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 ≠ component and thus highlight positive-negative variation.



Bertin 1967

Sparklines

	1999.1.1	65 months	2004.4.28	low	high
Euro foreign exchange $	1.1608		1.1907	.8252	1.2858
Euro foreign exchange ¥	121.32		130.17	89.30	140.31
Euro foreign exchange £	0.7111		0.6665	.5711	0.7235

	2003.4.28	12 months	2004.4.28	low	high
$	1.1025		1.1907	1.0783	1.2858
¥	132.54		130.17	124.80	140.31
£	0.6914		0.6665	0.6556	0.7235

Sparklines: Intense, Simple, Word-Sized Graphics

THE most common data display is a noun followed by some numbers. For example, a medical patient's current level of glucose is typically reported in a clinical record as a word and number:

glucose 128

When placed in a relevant context, a single number takes on additional meaning. Thus the most recent measurement of glucose might be — and often should be — compared to earlier measurements. This line shows a path of the last 80 readings of glucose:

glucose 128

Lacking a scale of measurement, this free-floating line is dequantified. But 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. These 2 representations of the most recent reading are tied together with a color accent:

glucose 128

Additional 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:

glucose 128

For clinical analysis, the task is quickly to detect and assess wayward deviations from normal limits, shown here by visual deviations outside the gray band. Multiplying this format brings in additional data from the medical record; a stack, which can show hundreds of variables and thousands of measurements, allows fast effective parallel comparisons:

glucose 128

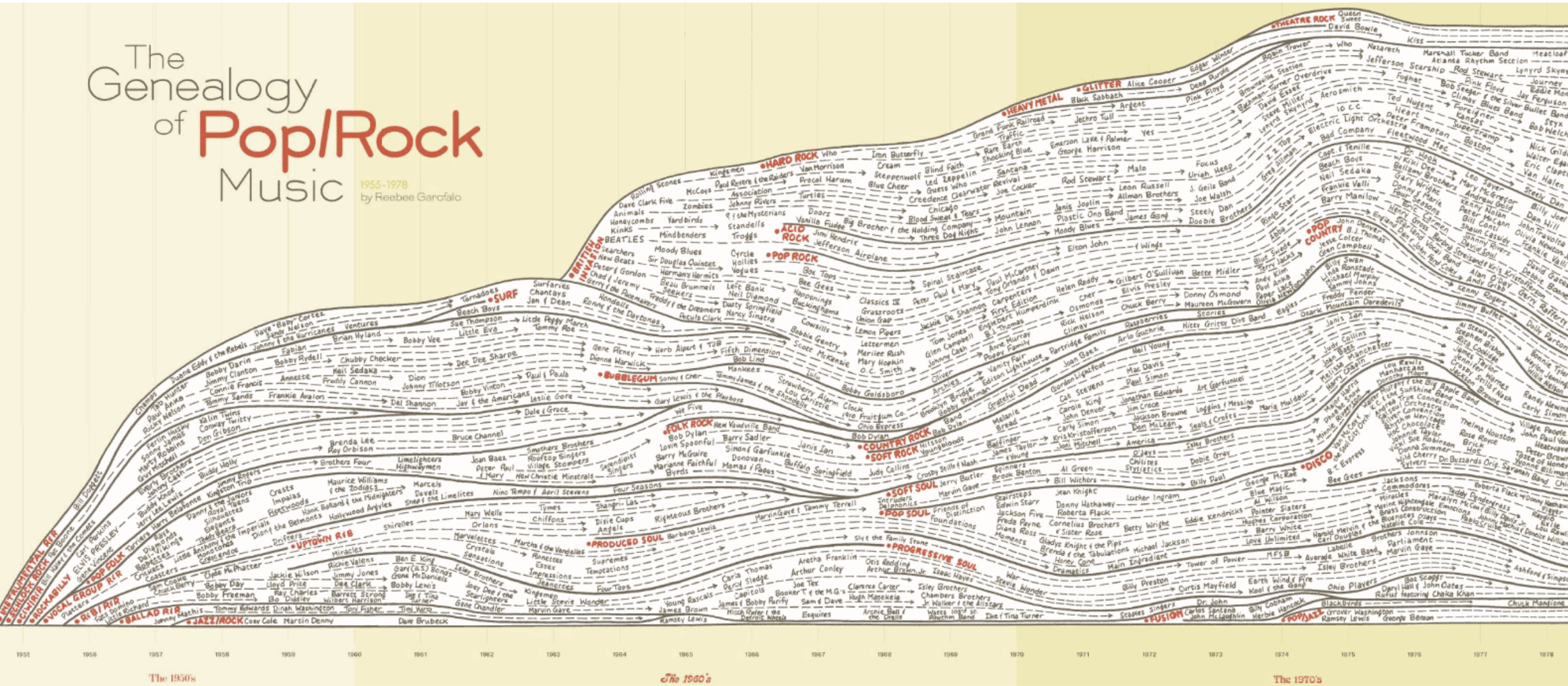
respiration 16

temperature 99.2

WBC 8,800

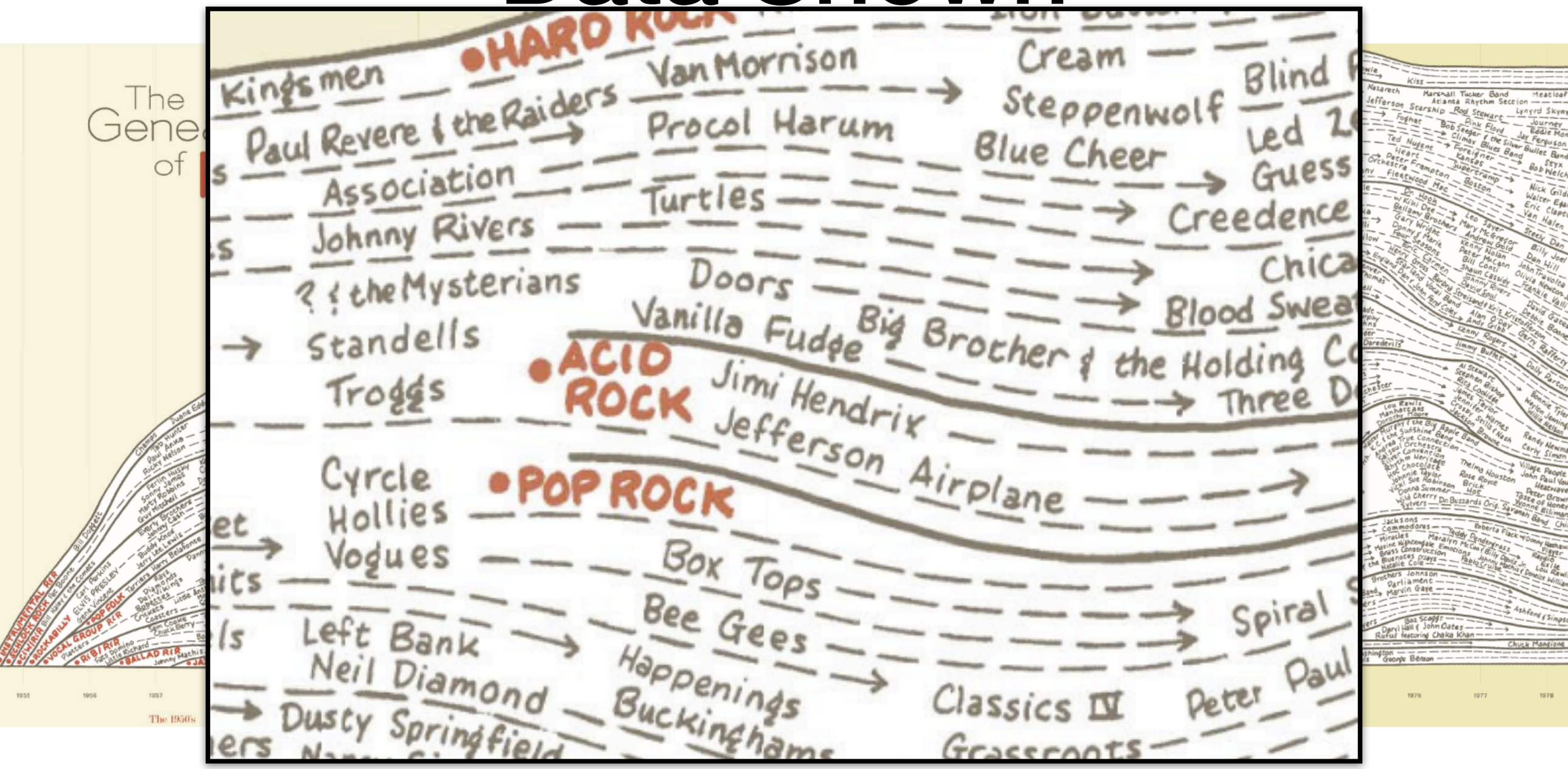
These little data lines, because of their active quality over time, are called *sparklines*: small, high-resolution graphics embedded in a context of words, numbers, images. Sparklines are data-intense, design-simple, word-sized graphics.

Maximize the Amount of Data Shown



Steve Chappel and Reebe Garofalo
in Rock 'N' Roll is Here to Pay: The History and Politics of the Music Industry, 1977

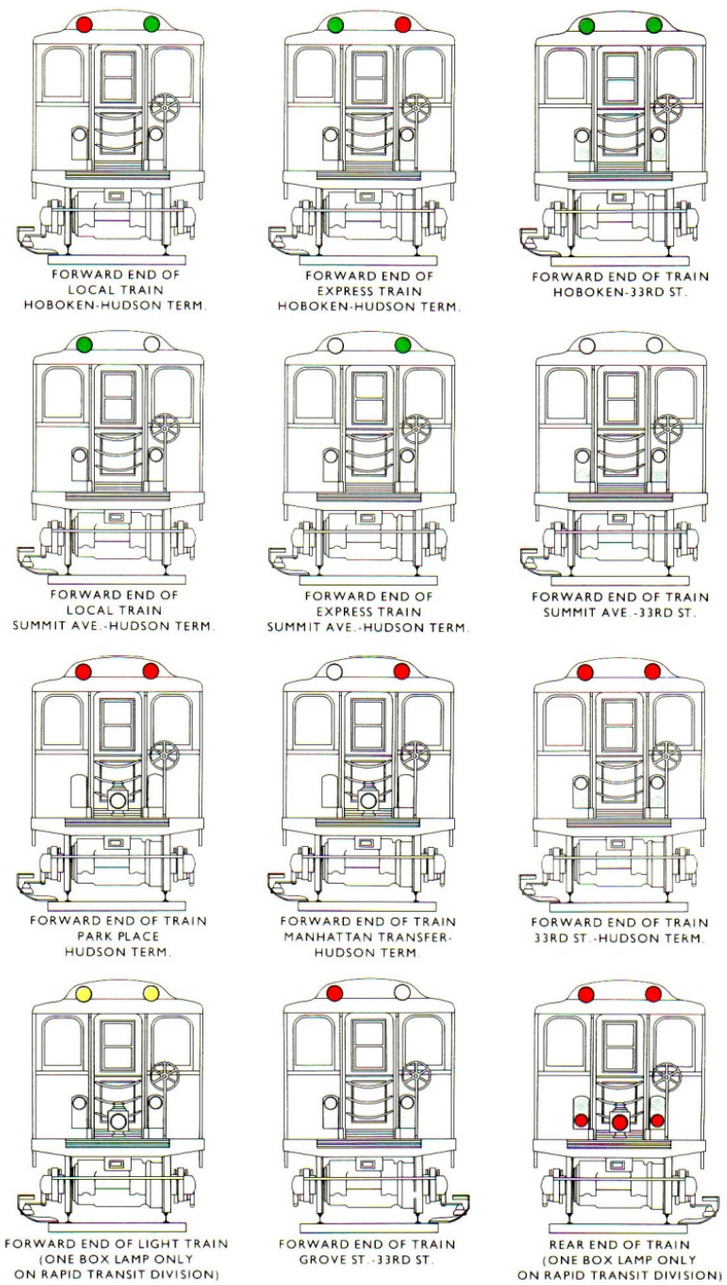
Maximize the Amount of Data Shown



Steve Chappel and Reebe Garofalo
in Rock 'N' Roll is Here to Pay: The History and Politics of the Music Industry, 1977

Small Multiples

- Repeat visually similar elements nearby rather than spreading apart



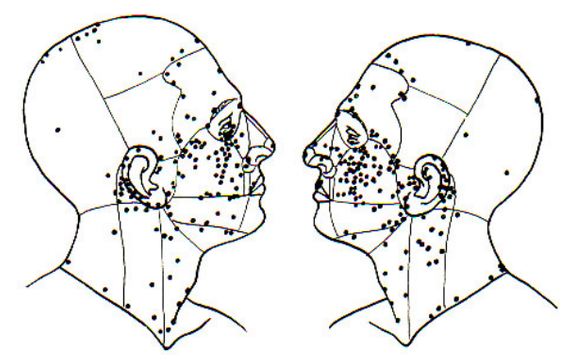


Abb. 1. Verteilung von 269 primären Melanomen auf Kopf und Hals

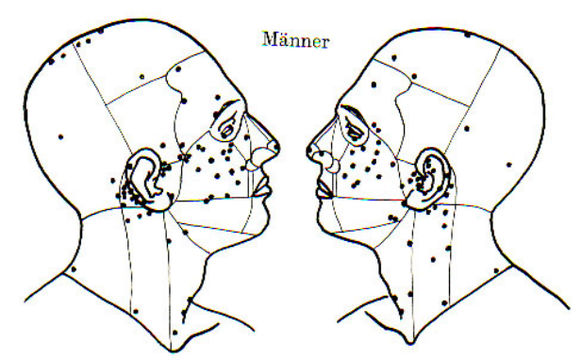


Abb. 2

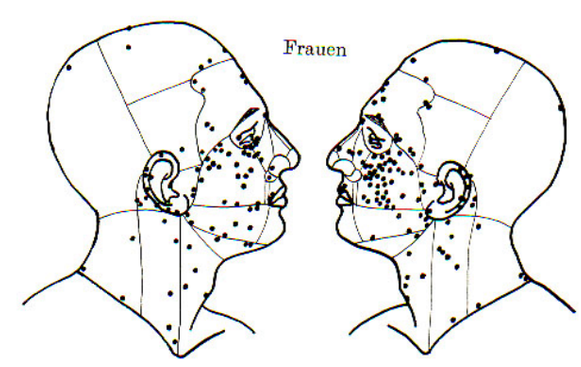
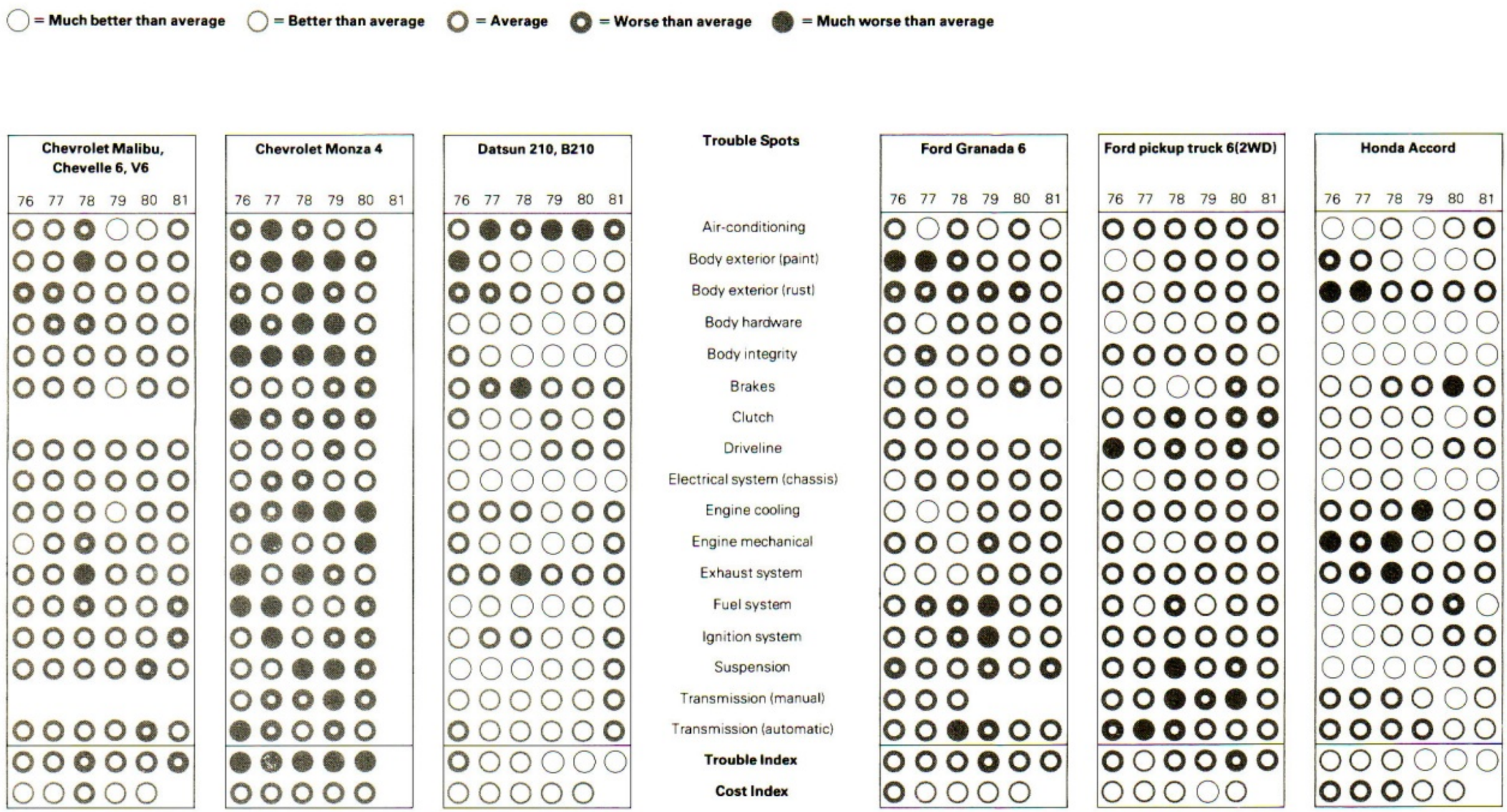


Abb. 3

Abb. 2 u. 3. Differenzierung der Melanomverteilung nach Geschlechtern

Small Multiples

- Repeat visually similar elements nearby rather than spreading apart



Summary: (Some of) Tufte's Design Principles

- Maximize the data-ink ratio
- Avoid chart junk (sometimes)
- Use multifunctioning elements (macro/micro readings)
- Maximize the data density (sometimes)
- Use small multiples

- Utilize narratives of space and time
- Escape flatland

Case Study

Visualizations by W. E. B. Du Bois

Context

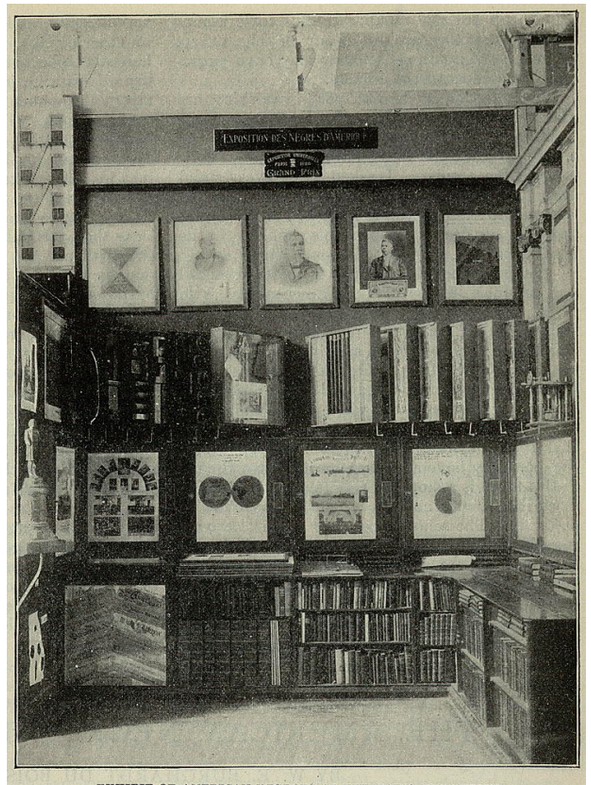


EXHIBIT OF AMERICAN NEGROES AT THE PARIS EXPOSITION.

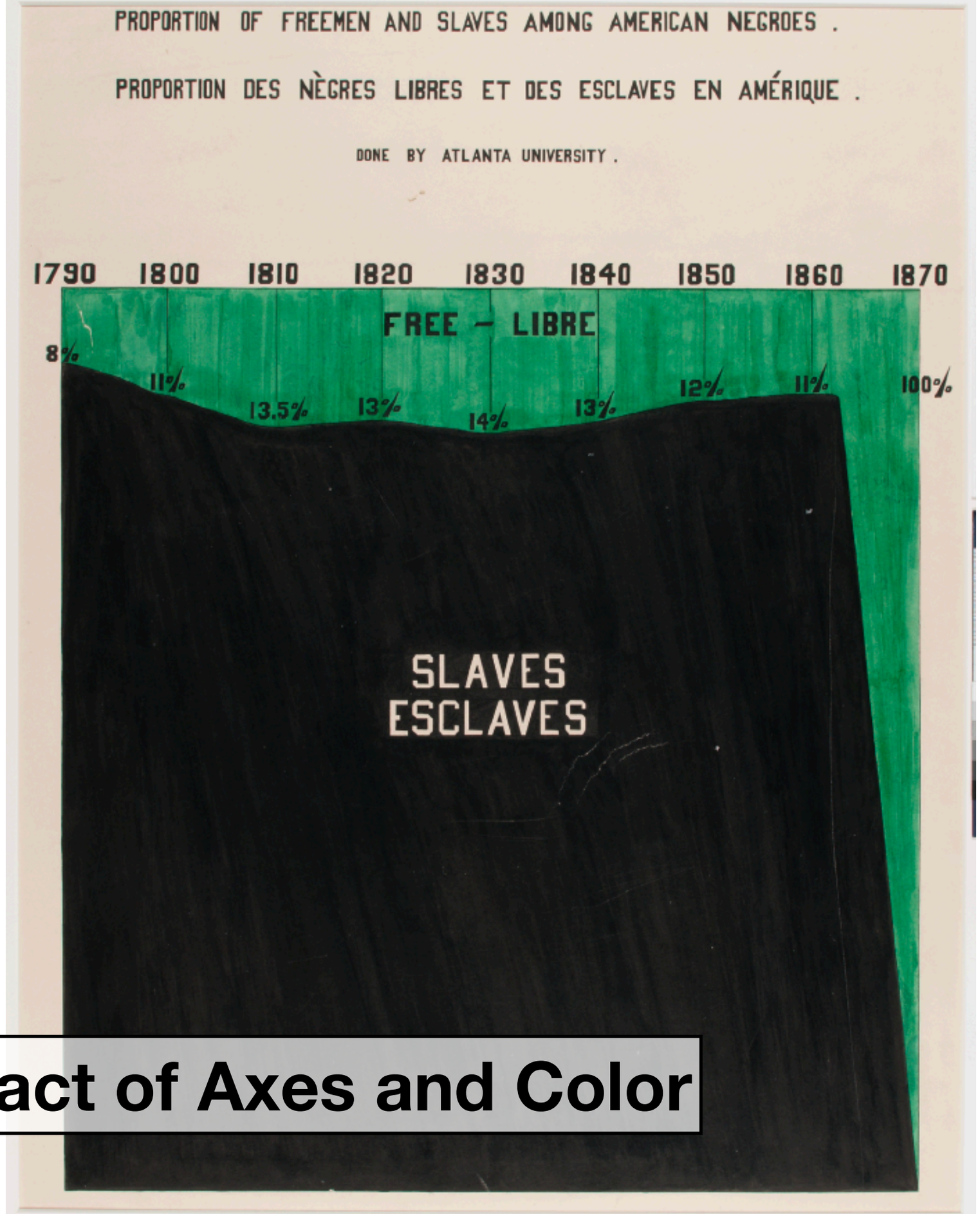
- Powerful display of sociological data collected and presented meticulously
- Designed for the 1900 Paris Exposition
- Goal was to tell the story of African Americans to an international audience
- While mainstream news ignored the exhibit, comprehensive articles have since been published

https://en.wikipedia.org/wiki/The_Exhibit_of_American_Negroes

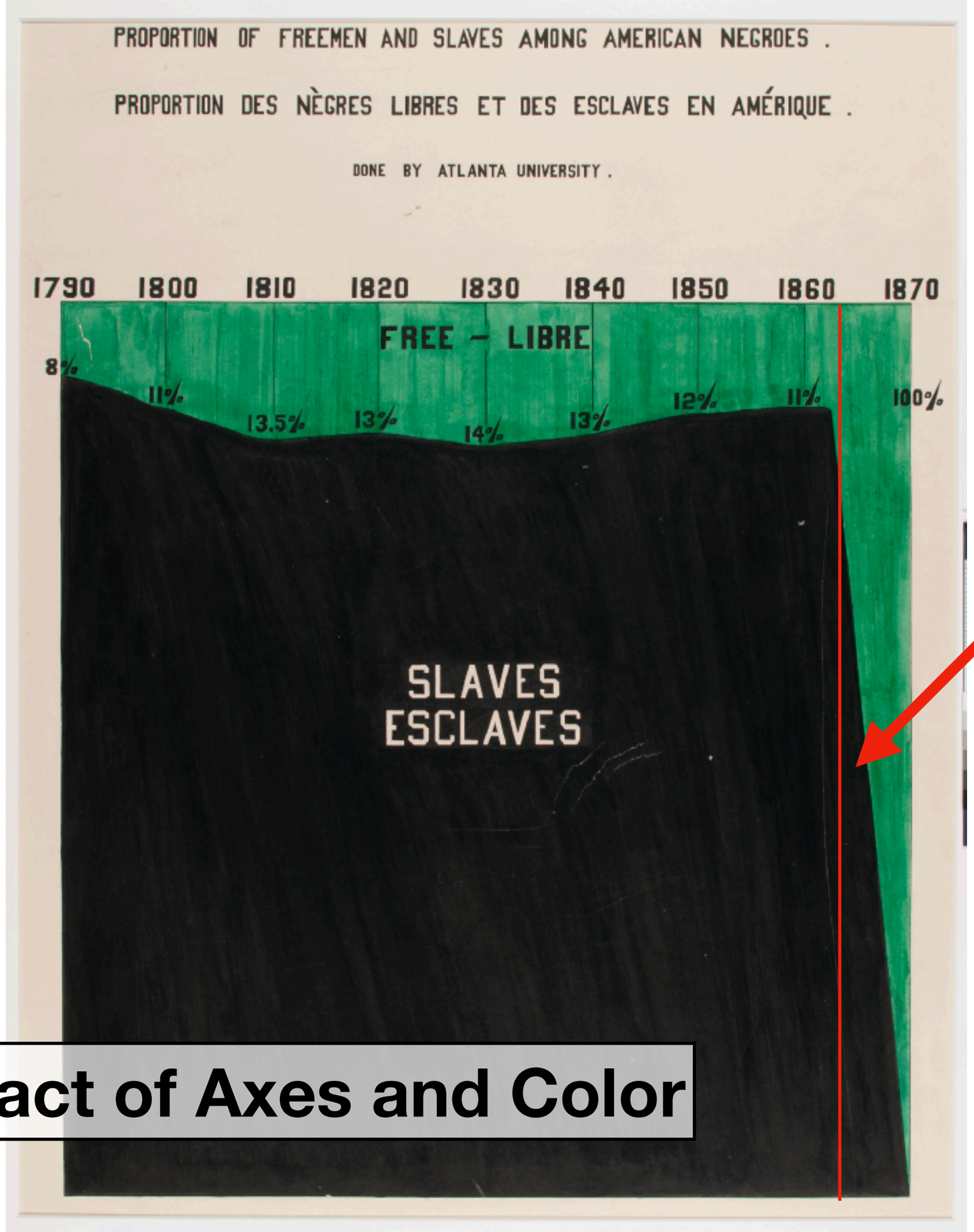
“Thirty-two charts, 500 photographs, and numerous maps and plans form the basis of this exhibit. The charts are in two sets, one illustrating conditions in the entire United States and the other conditions in the typical State of Georgia”

–W. E. B. Du Bois

https://medium.com/nightingale/w-e-b-du-bois-staggering-data-visualizations-are-as-powerful-today-as-they-were-in-1900-64752c472ae4



The Impact of Axes and Color



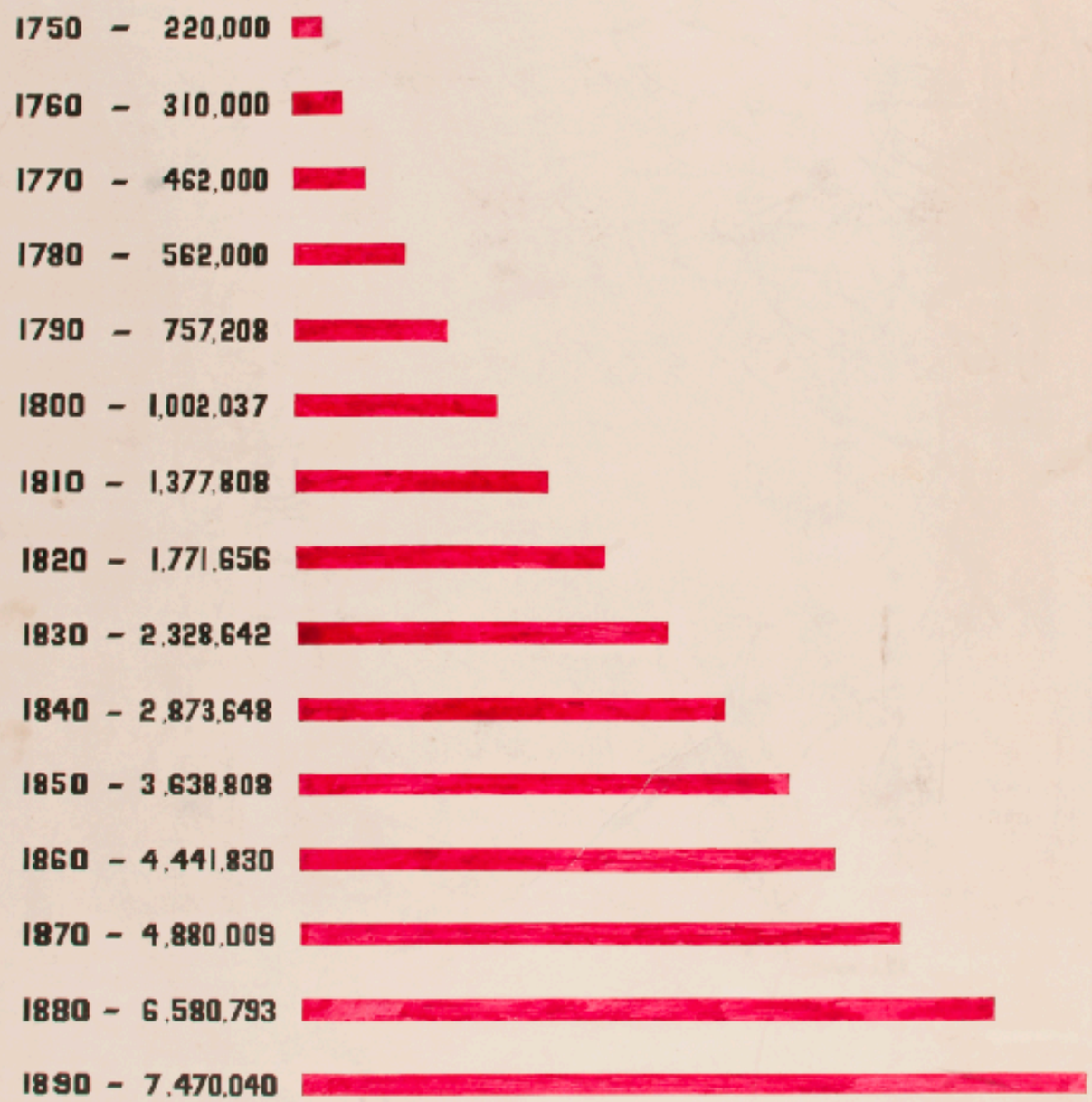
Emancipation Proclamation signed in 1863

The Impact of Axes and Color

Increase of the Negro population in the United States of America.

Accroissement de la population Negre aux Etats Unis d' Amerique.

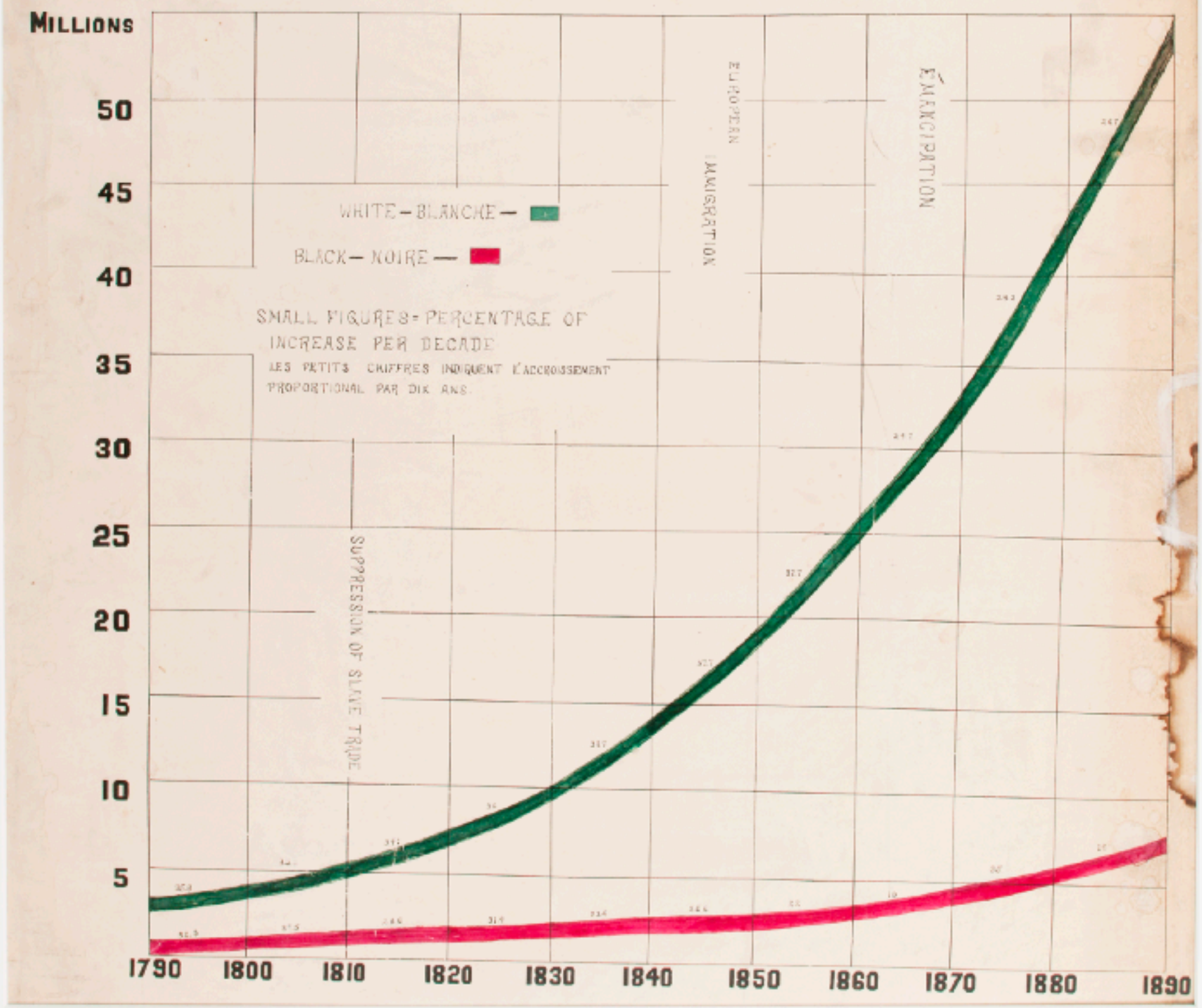
Done by Atlanta University.



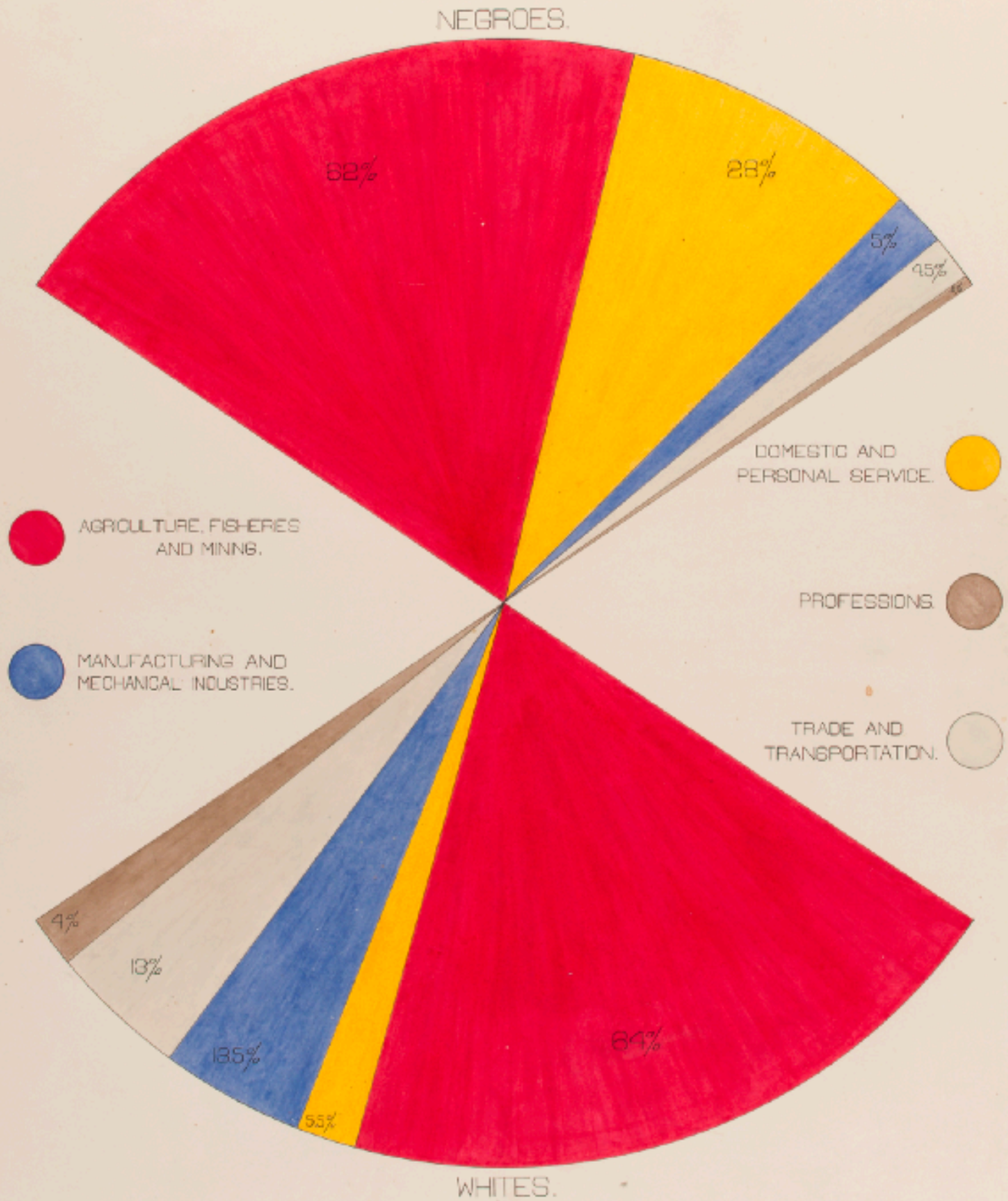
Comparative rate of increase of the White and Negro elements of the population of the United States.

Accroissement proportionnel des elements blancs et noirs aux Etats Unis.

Done by Atlanta University.

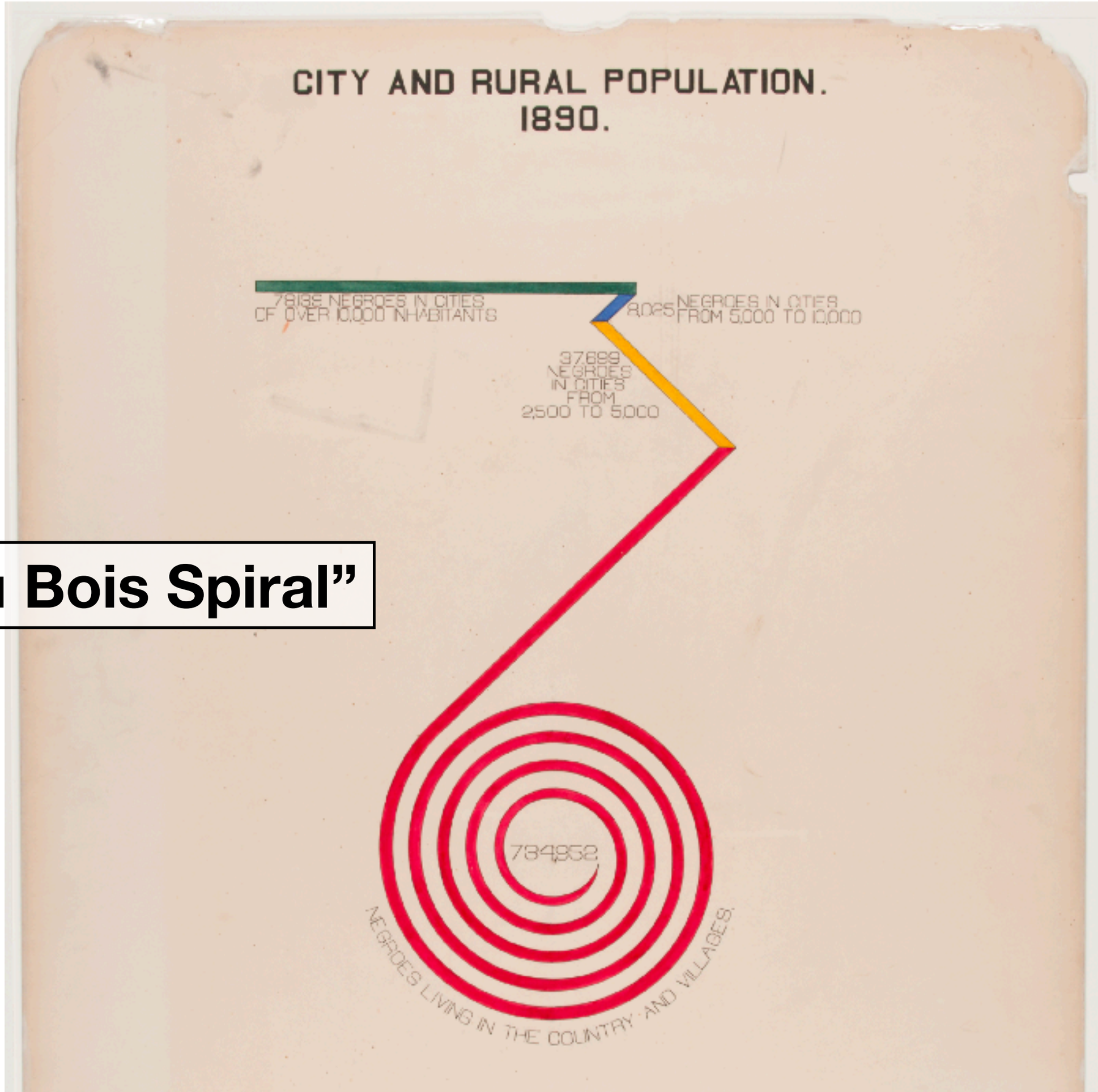


OCCUPATIONS OF NEGROES AND WHITES IN GEORGIA.

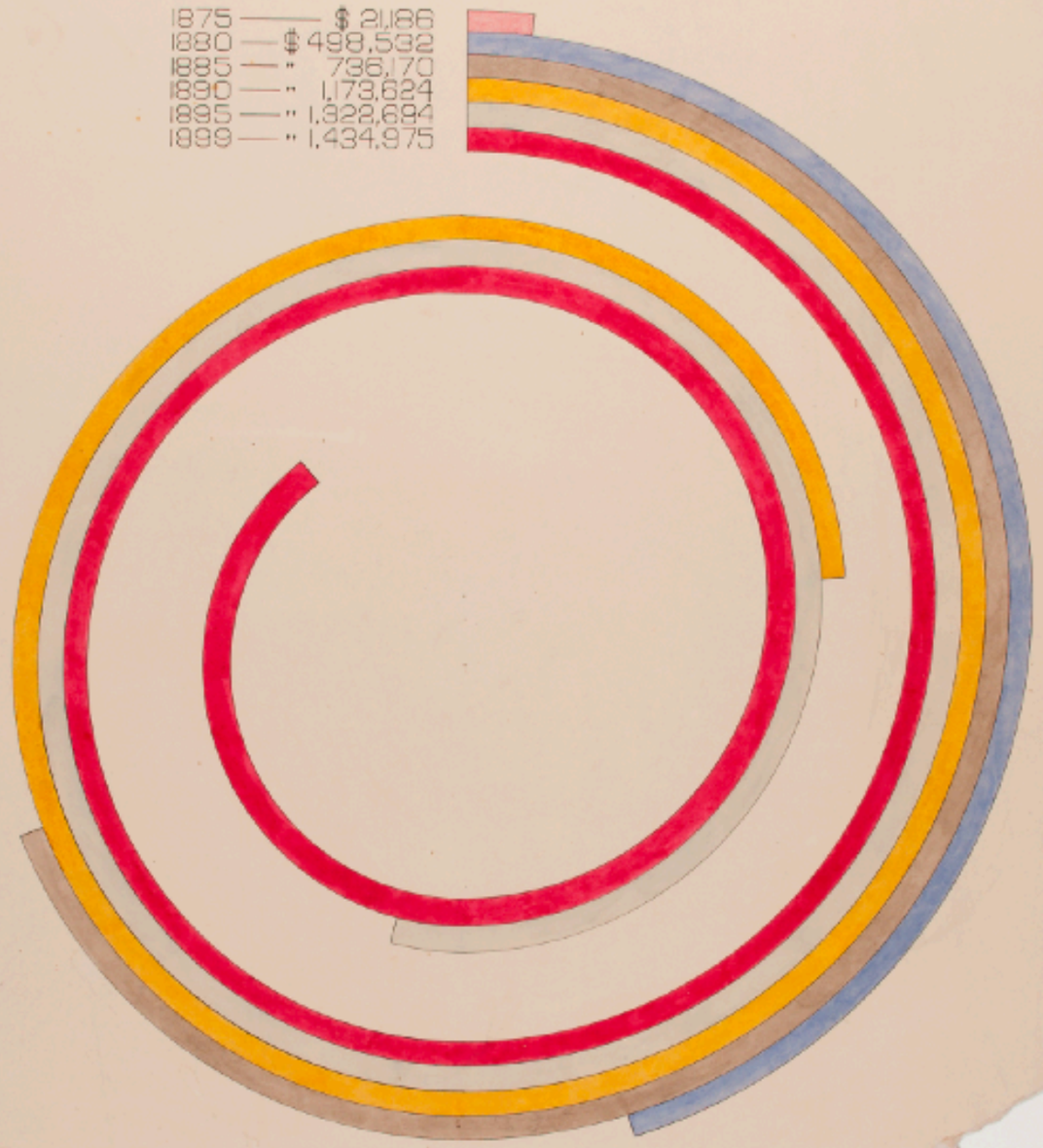


[27]

“Du Bois Spiral”

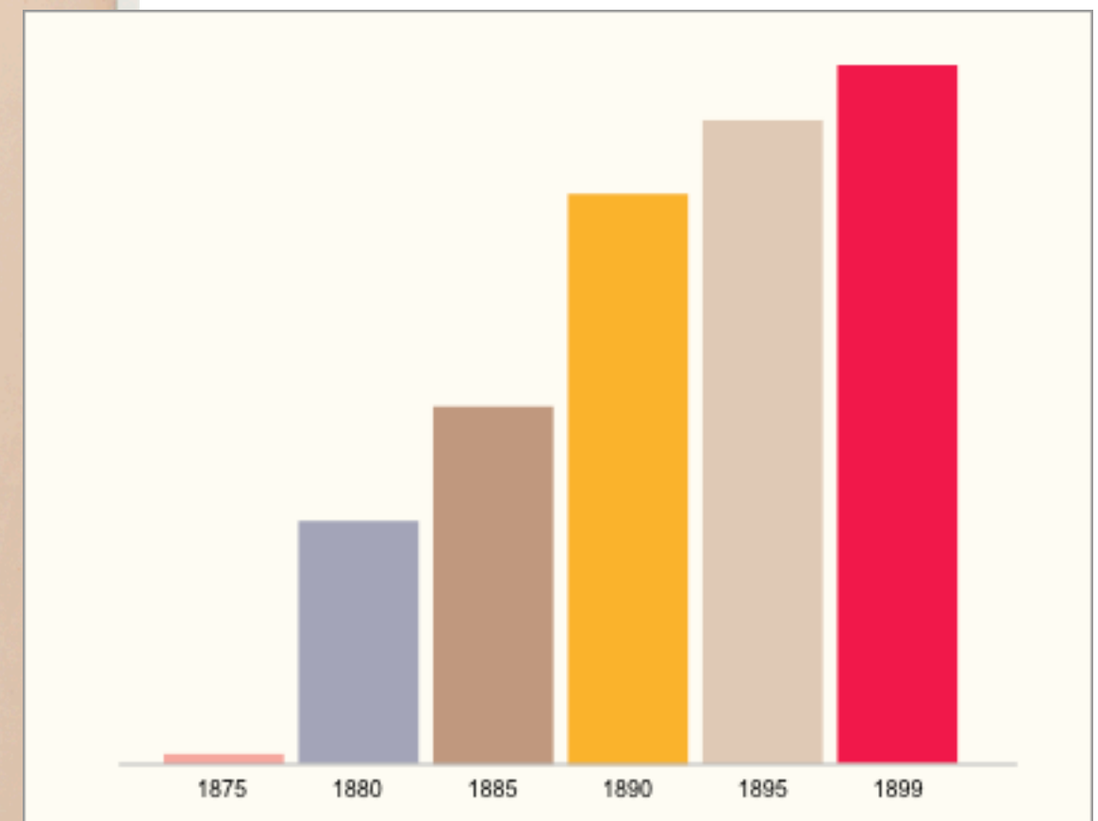


ASSESSED VALUE OF HOUSEHOLD AND KITCHEN FURNITURE
OWNED BY GEORGIA NEGROES.



[25]

ASSESSED VALUE OF HOUSEHOLD AND KITCHEN FURNITURE
OWNED BY GEORGIA NEGROES.
1875 —— $ 21,186
1880 — $ 498,532
1885 — " 736,170
1890 — " 1,173,624
1895 — " 1,322,684
1899 — " 1,434,975


1875
1880
1885
1890
1895
1899

Lec06 Reading

- Ware, Information Visualization, Chapter 3
- Healey and Enns, Attention and Visual Memory in Visualization and Computer Graphics. https://www.csc2.ncsu.edu/faculty/healey/PP/

Reminder
Assignment 01

Assigned: Monday, January 23
Due: Monday, February 6, 4:59:59 pm

Reminder
Project Milestone 01

Assigned: Monday, January 25
Due: Monday, February 22, 4:59:59 pm