

<http://poloclub.gatech.edu/cse6242>

CSE6242 / CX4242: Data & Visual Analytics

Common visualization Issues & how to fix them

Duen Horng (Polo) Chau

Assistant Professor

Associate Director, MS Analytics

Georgia Tech

Partly based on materials by

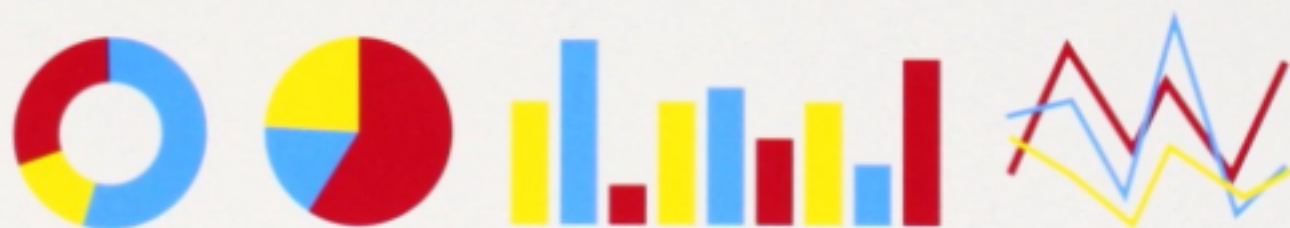
Professors Guy Lebanon, Jeffrey Heer, John Stasko, Christos Faloutsos

THE WALL STREET JOURNAL. **GUIDE TO INFORMATION GRAPHICS**

**THE DOS & DON'TS
OF PRESENTING
DATA, FACTS,
AND FIGURES**

DONA M. WONG

"INVALUABLE." —HOW DESIGN



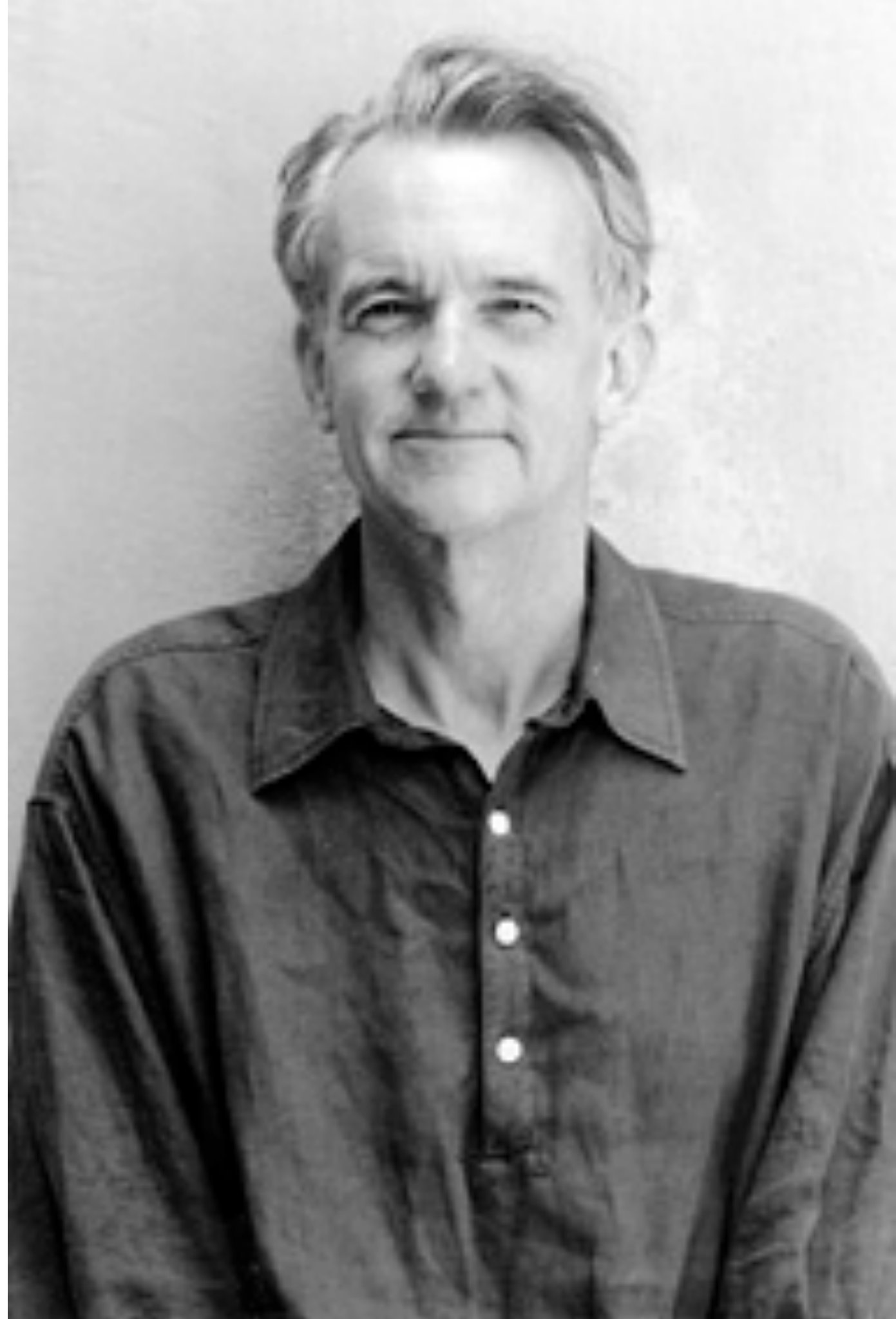
Student of
Edward Tufte

Edward Tufte

An American statistician and professor emeritus of political science, statistics, and computer science at Yale University.

He is noted for his writings on information design and as a pioneer in the field of data visualization.

-Wikipedia



Also Highly Recommended:



Copyrighted Material

Second Edition

Information Dashboard Design

Displaying data for
at-a-glance monitoring

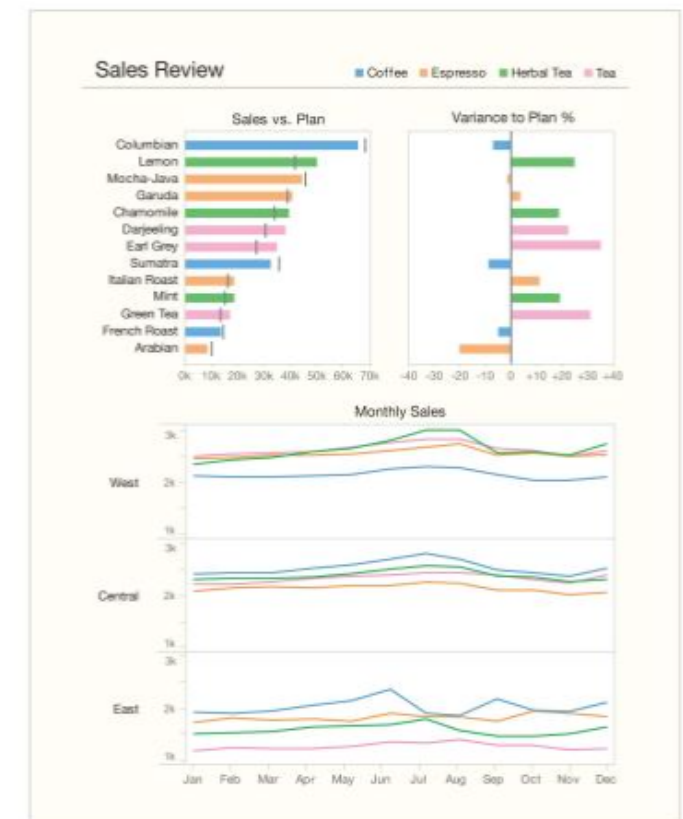
Stephen Few

Copyrighted Material

Second Edition

Show Me the Numbers

Designing Tables and Graphs to Enlighten

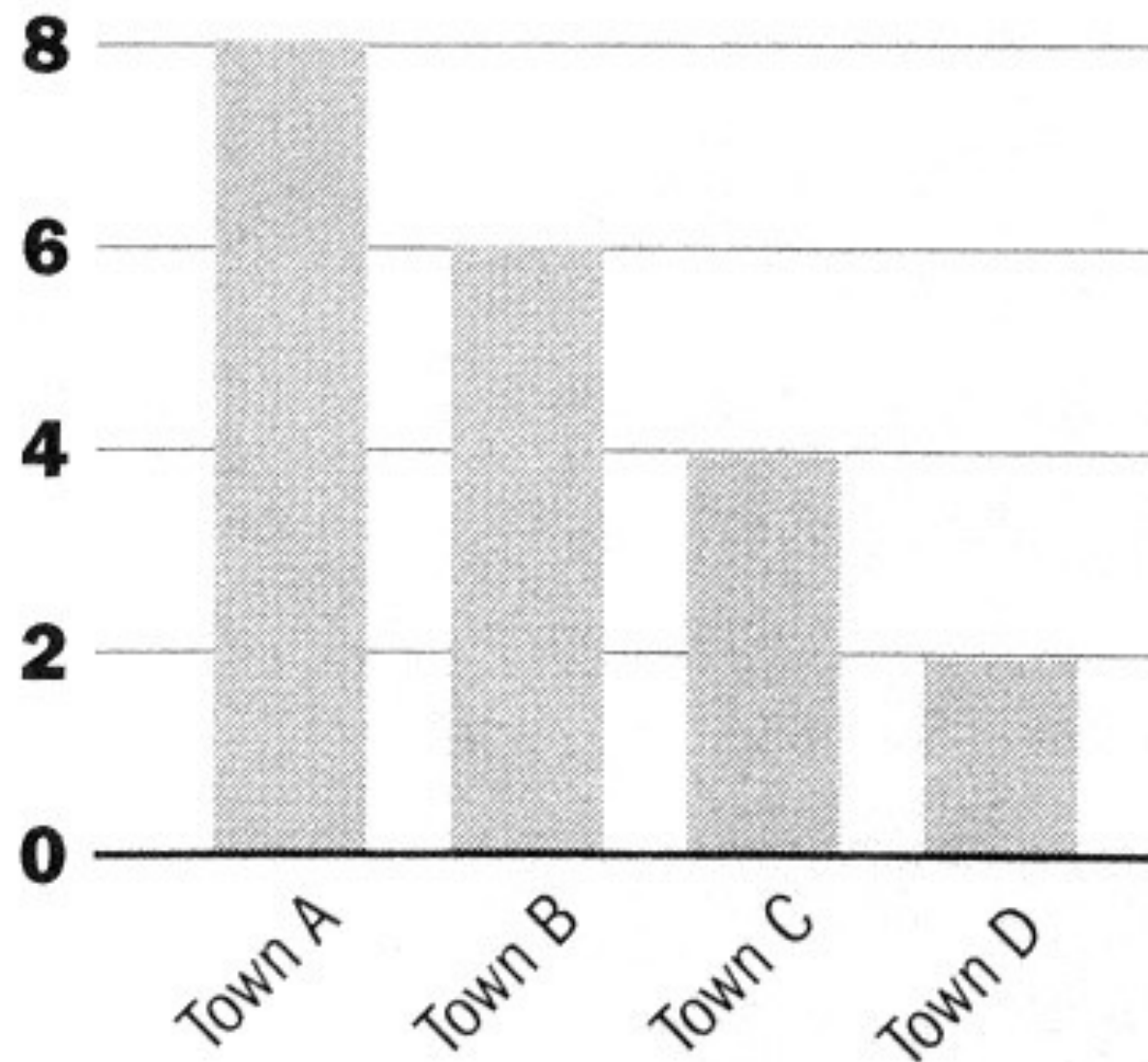


Stephen Few

Copyrighted Material

HEADLINE OF THE CHART

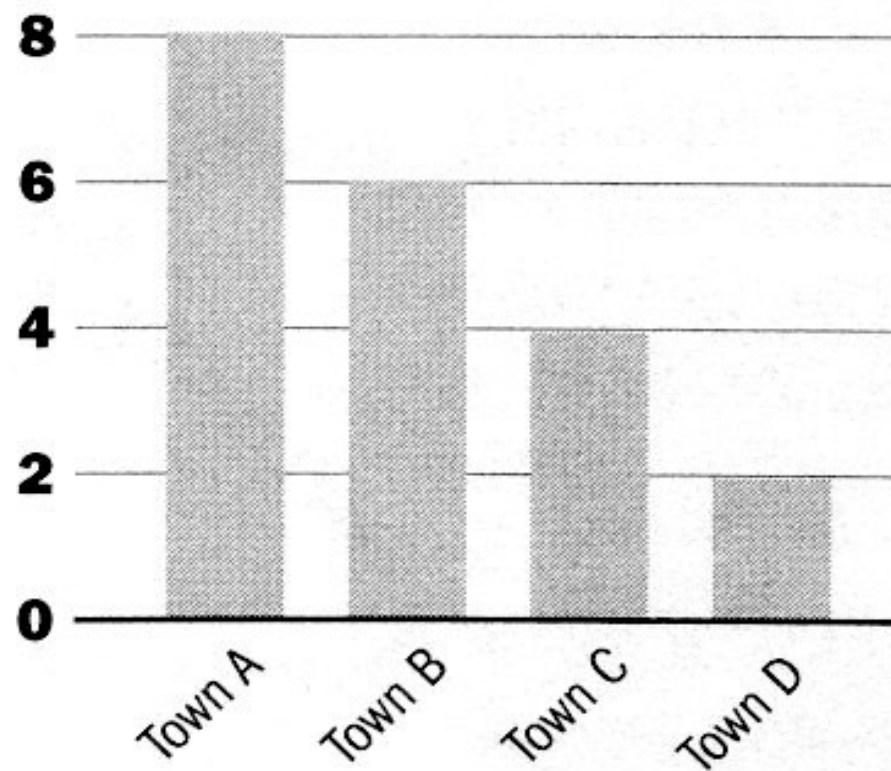
A brief description that outlines what the data shows



Can you improve its visual design?

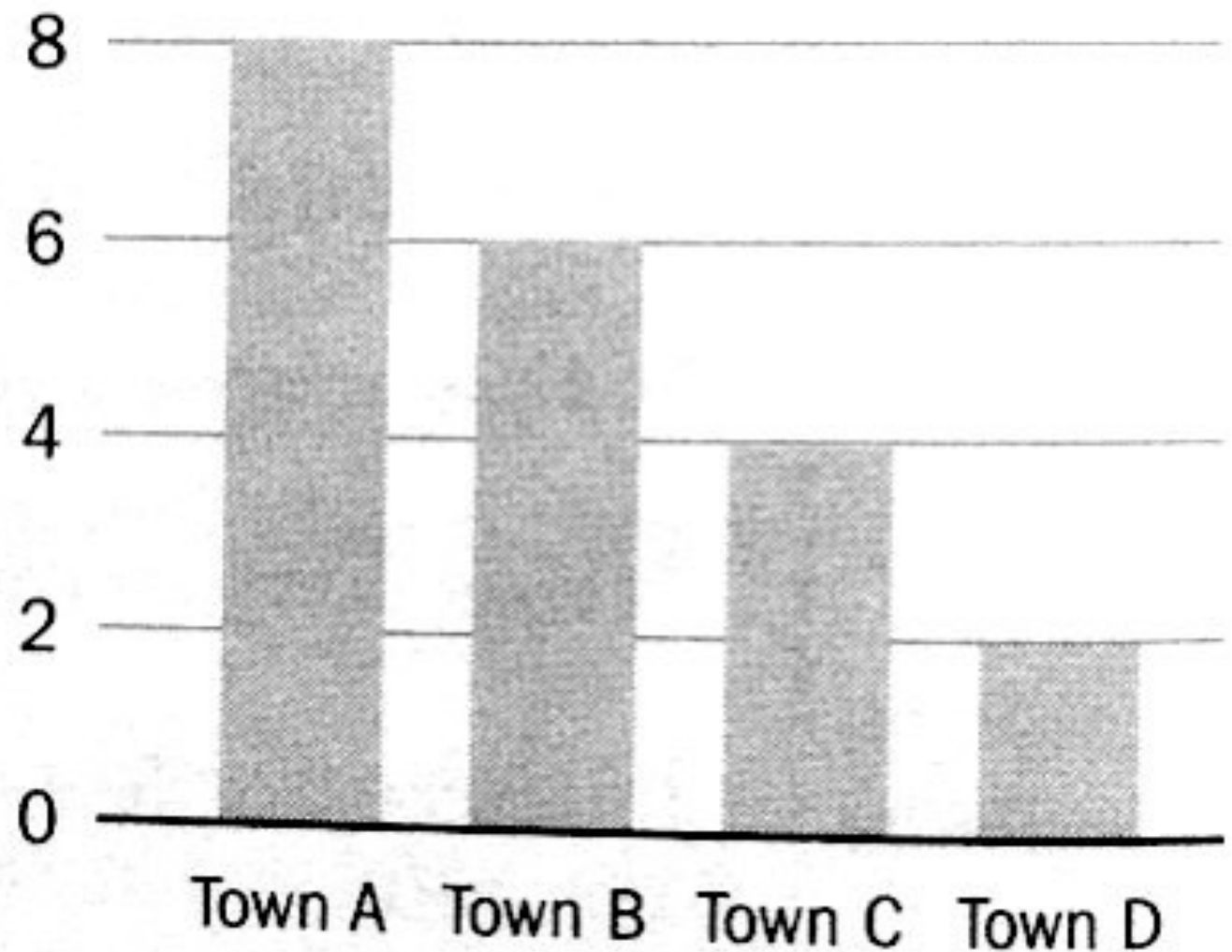
HEADLINE OF THE CHART

A brief description that outlines what the data shows



Headline of the chart

A brief description that outlines what the data shows



Which is better?

Tables

Name	Data	Data	Data
Company A	0.0	0.0	0.0
Company B	0.0	0.0	0.0
Company C	0.0	0.0	0.0
Company D	0.0	0.0	0.0

What can you improve?

What's the problem with
making everything
bold or *italic*?

Disney PRESENTS A PIXAR FILM



THE INCREDIBLES

26/11/04



https://youtu.be/1E9pKU_N15A?t=45s

www.theincredibles.co.uk

“When everyone is super,
no one is super”

Disney PRESENTS A PIXAR FILM



THE INCREDIBLES

26/11/04



https://youtu.be/1E9pKU_N15A?t=45s

www.theincredibles.co.uk

When everyone is special, no one is!

Name	Data	Data	Data
Company A	0.0	0.0	0.0
Company B	0.0	0.0	0.0
Company C	0.0	0.0	0.0
Company D	0.0	0.0	0.0

Name	Data	Data	Data	Data	Data	Data
Company A	0.0	0.0	0.0	0.0	0.0	0.0
Company B	0.0	0.0	0.0	0.0	0.0	0.0
Company C	0.0	0.0	0.0	0.0	0.0	0.0
Company D	0.0	0.0	0.0	0.0	0.0	0.0
Company E	0.0	0.0	0.0	0.0	0.0	0.0
Company F	0.0	0.0	0.0	0.0	0.0	0.0
Company G	0.0	0.0	0.0	0.0	0.0	0.0
Company H	0.0	0.0	0.0	0.0	0.0	0.0

Name	Data	Data	Data	Data	Data	Data
Company A	0.0	0.0	0.0	0.0	0.0	0.0
Company B	0.0	0.0	0.0	0.0	0.0	0.0
Company C	0.0	0.0	0.0	0.0	0.0	0.0
Company D	0.0	0.0	0.0	0.0	0.0	0.0
Company E	0.0	0.0	0.0	0.0	0.0	0.0
Company F	0.0	0.0	0.0	0.0	0.0	0.0
Company G	0.0	0.0	0.0	0.0	0.0	0.0
Company H	0.0	0.0	0.0	0.0	0.0	0.0

A lot of “chart junk”.

Low “**data to ink**” ratio (Edward Tufte)

Name	Data	Data	Data	Data	Data	Data
Company A	0.0	0.0	0.0	12.0	0.0	0.0
Company B	0.0	0.0	0.0	11.0	0.0	0.0
Company C	0.0	0.0	0.0	10.0	0.0	0.0
Company D	0.0	0.0	0.0	9.0	0.0	0.0
Company E	0.0	0.0	0.0	8.0	0.0	0.0
Company F	0.0	0.0	0.0	7.0	0.0	0.0
Company G	0.0	0.0	0.0	6.0	0.0	0.0
Company H	0.0	0.0	0.0	5.0	0.0	0.0
Company I	0.0	0.0	0.0	4.0	0.0	0.0
Company J	0.0	0.0	0.0	3.0	0.0	0.0
Company K	0.0	0.0	0.0	2.0	0.0	0.0
Company L	0.0	0.0	0.0	1.0	0.0	0.0

Higher “data to ink” ratio

Problems?

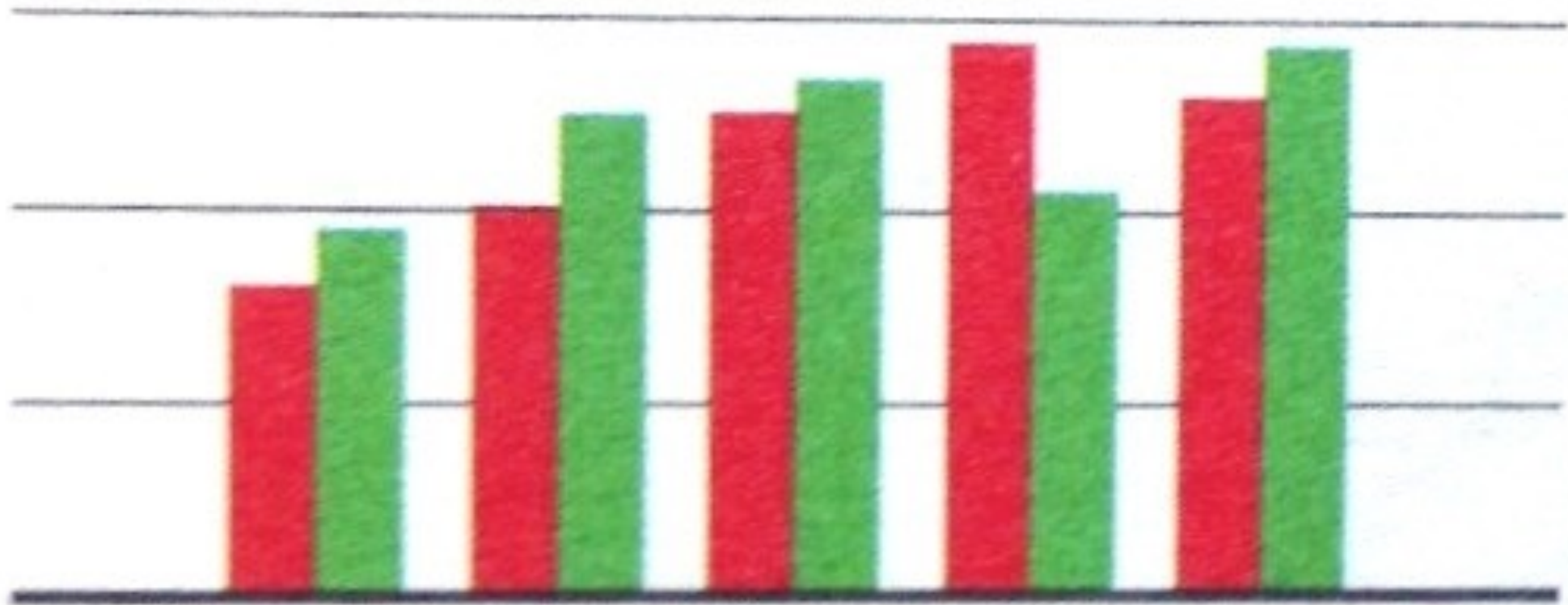
Name	Data
Company A	1000
Company B	900
Company C	80
Company D	7

Name	Data
Company A	10.82
Company B	9.49
Company C	8
Company D	7.4

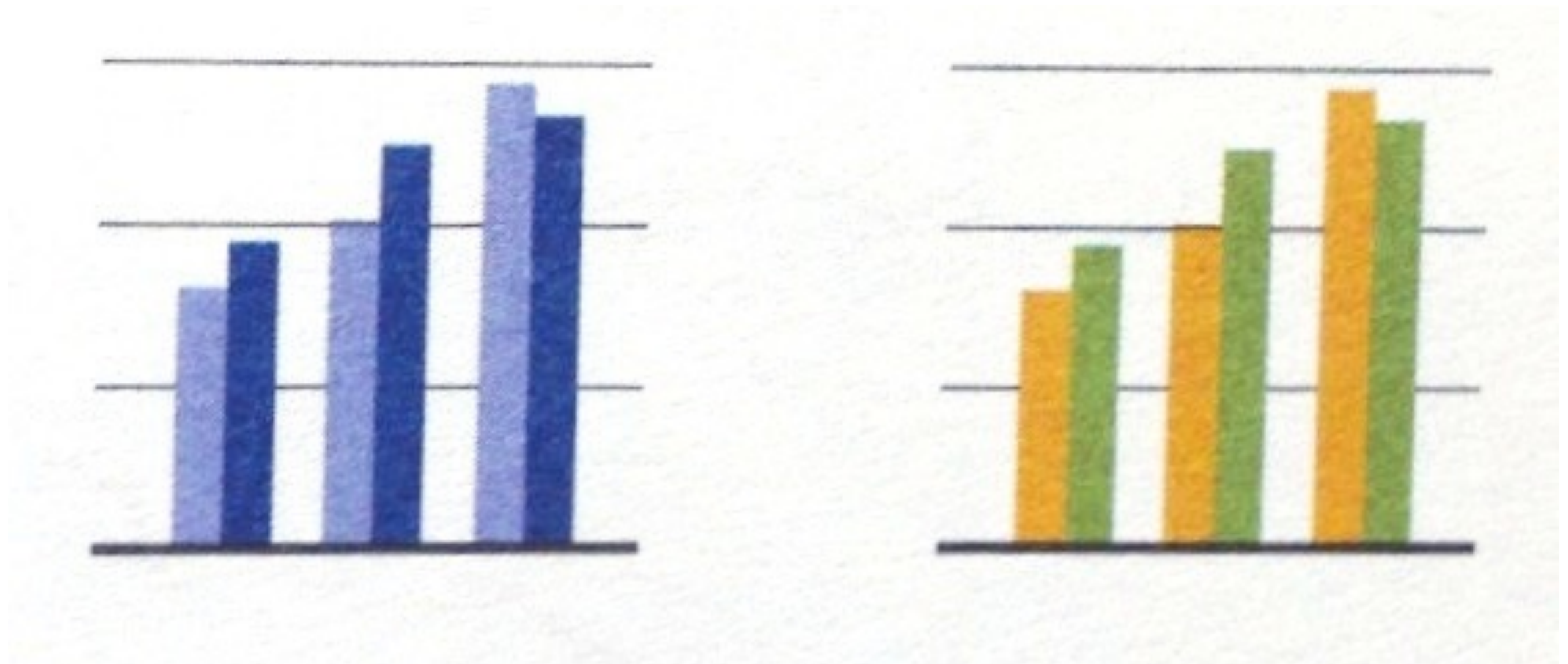
Name	Data
Company A	10.82
Company B	9.49
Company C	8
Company D	7.4

Name	Data
Company A	10.8
Company B	9.5
Company C	8.0
Company D	7.4

Bar Charts

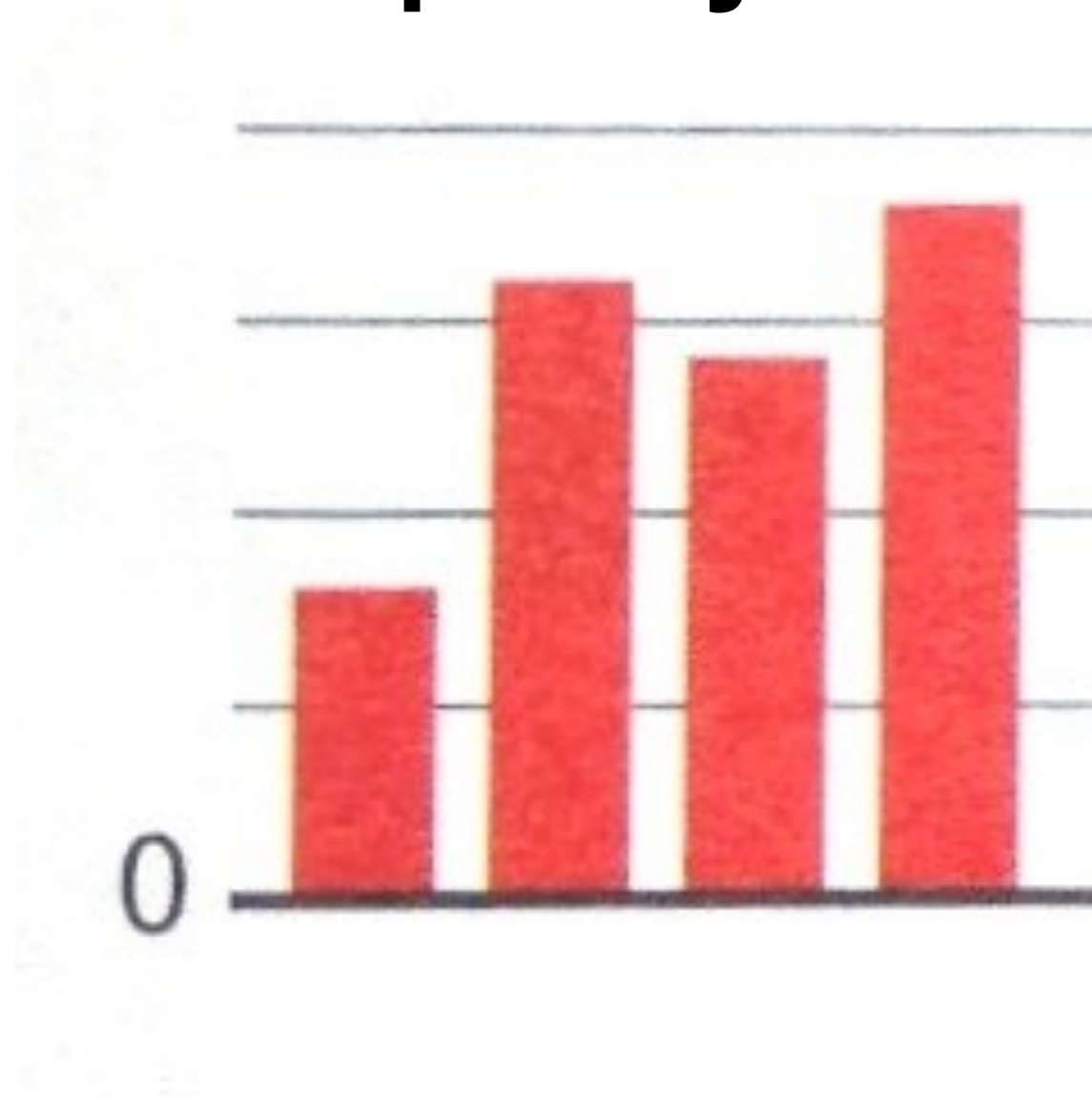


The color scheme reminds you of what?

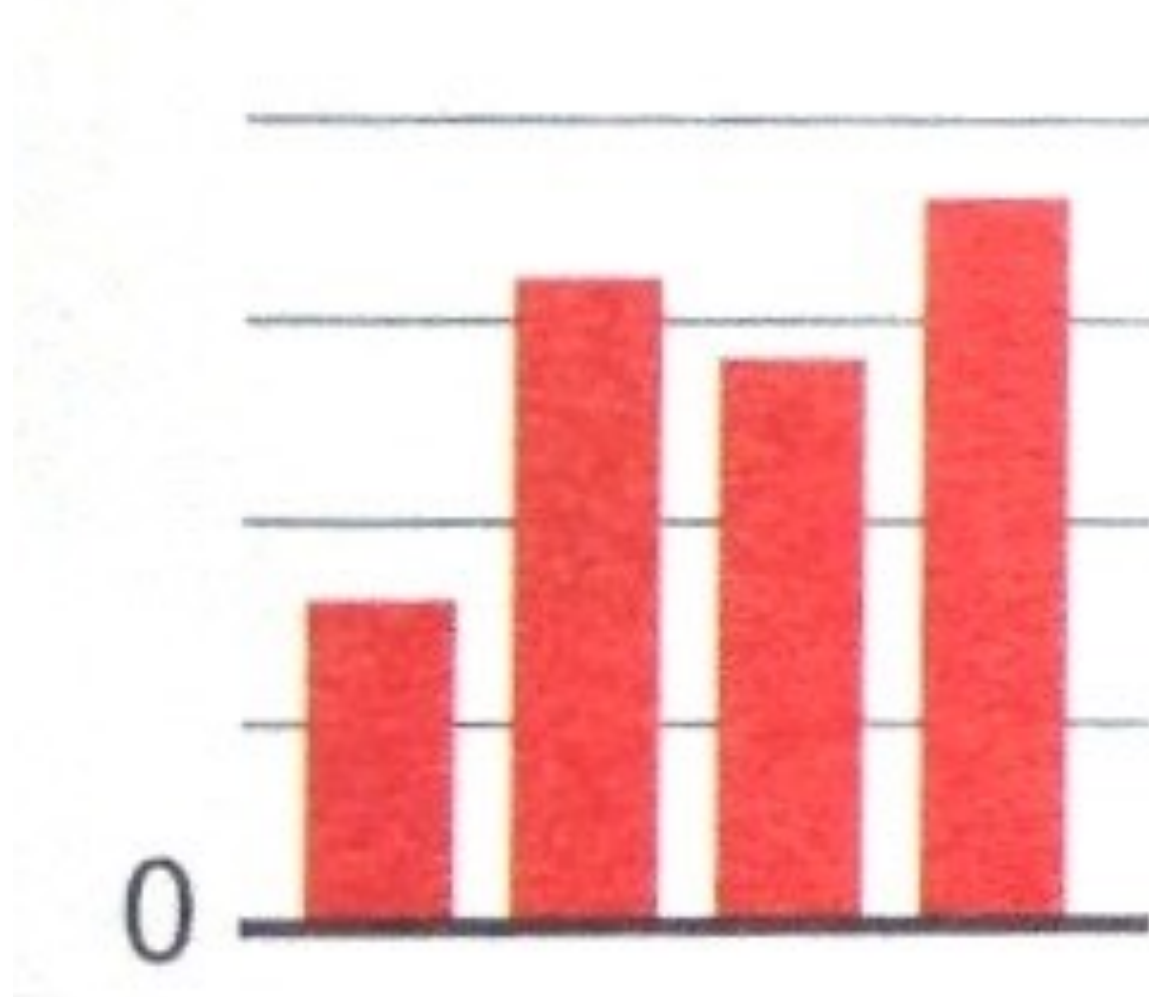


Better than Christmas
(Use color brewer to find good color schemes)

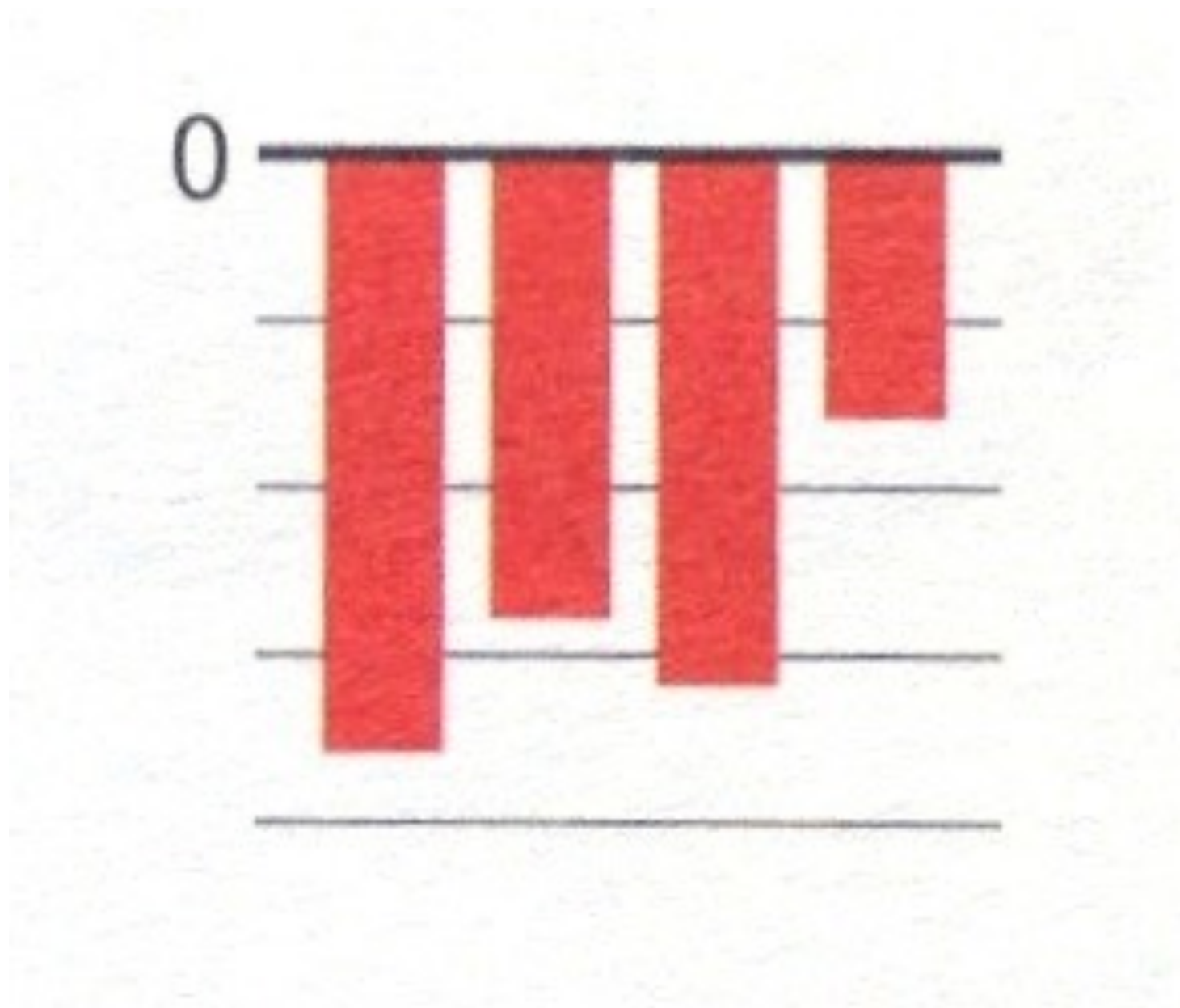
Company Profits

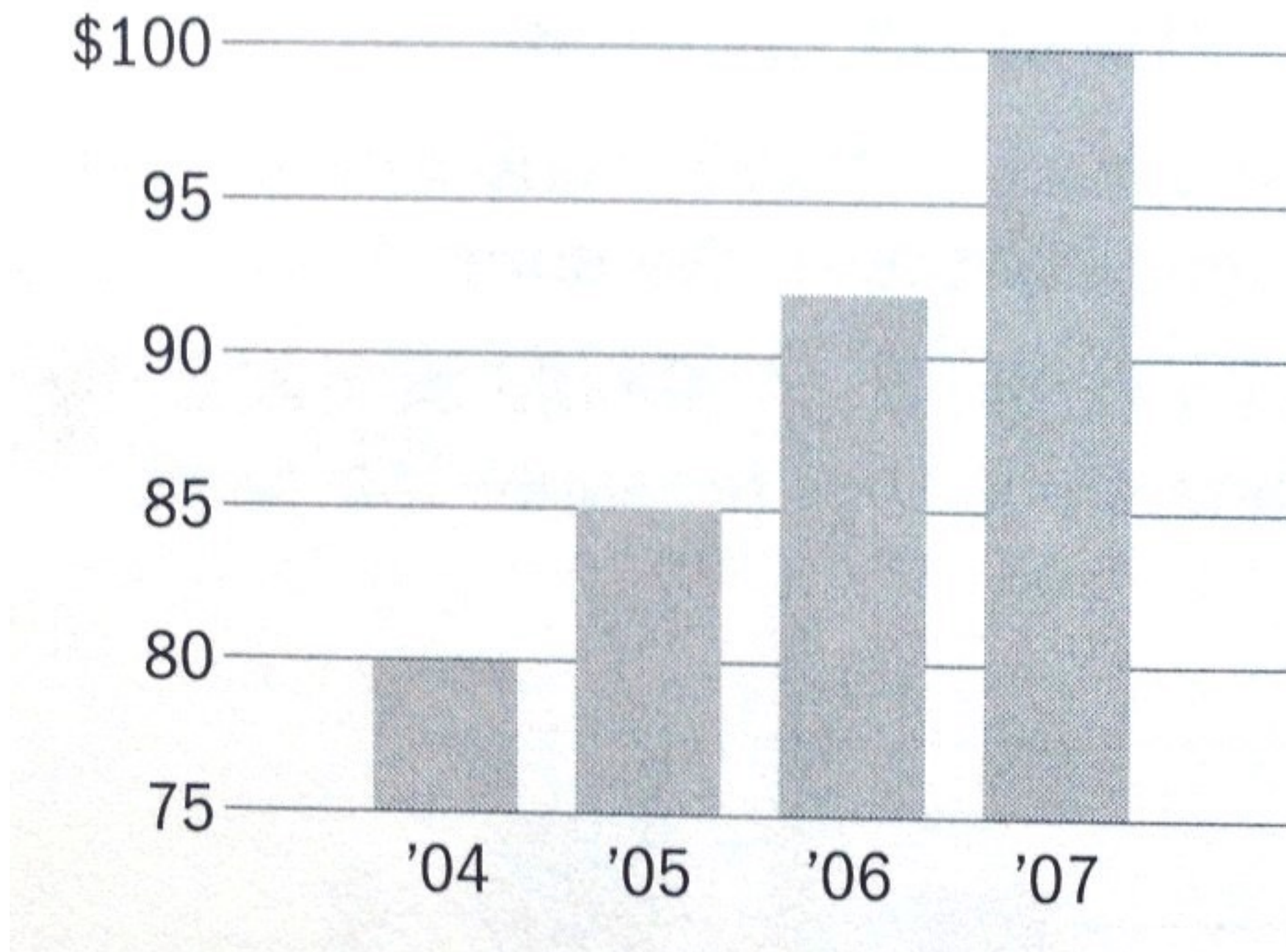


Company Profits

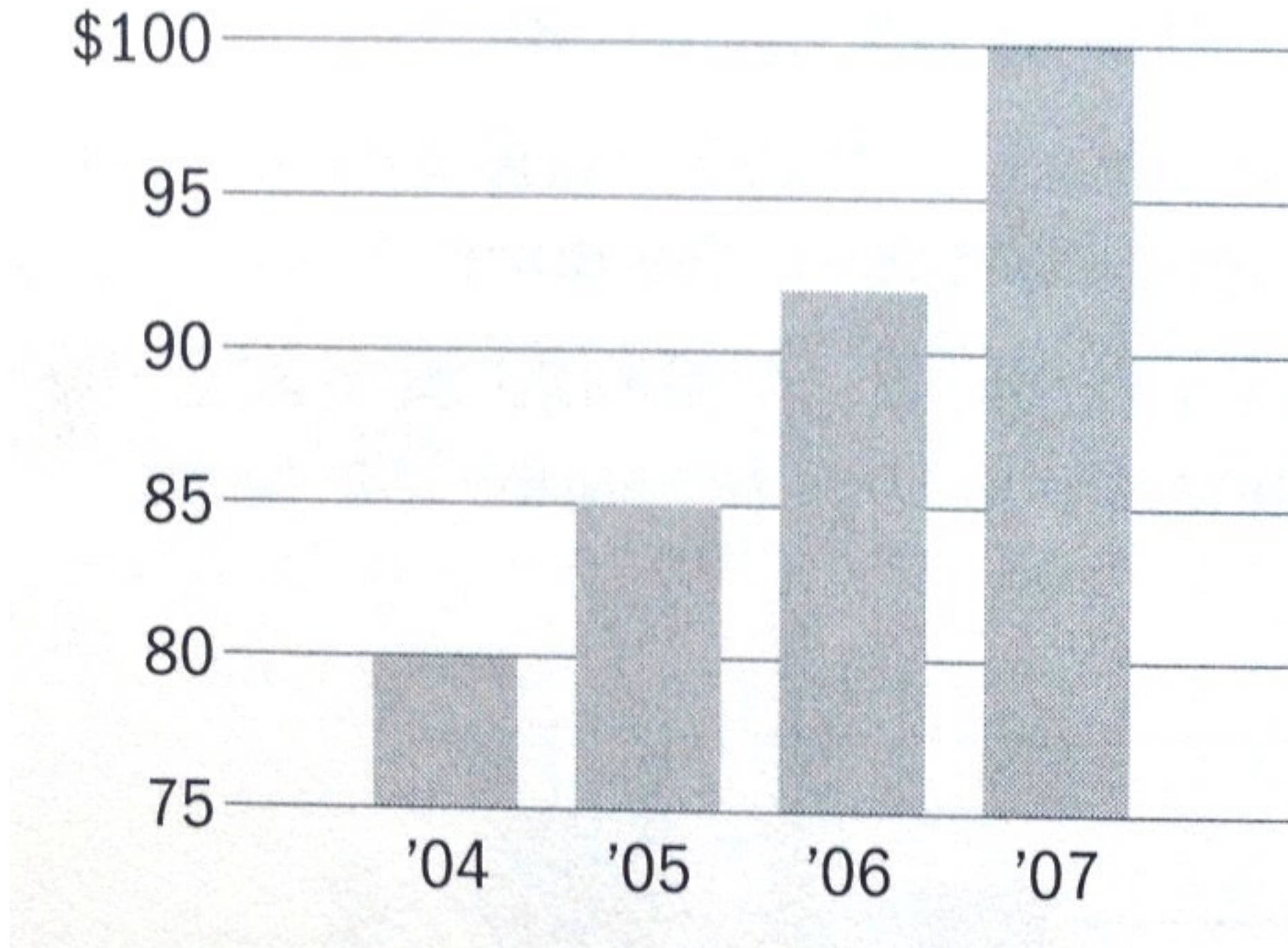


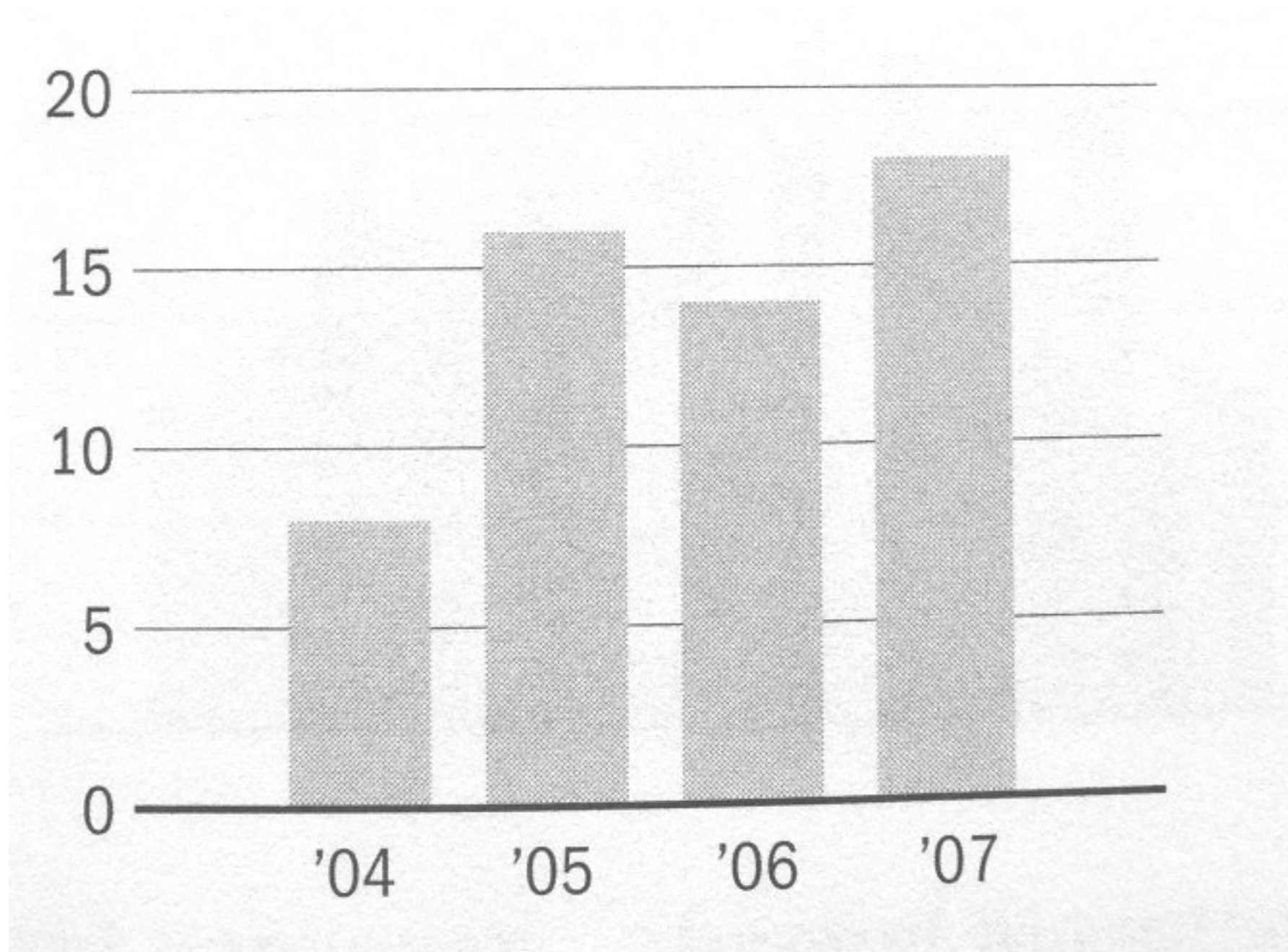
Don't show profits in **red**!!
Think carefully about your color choices.





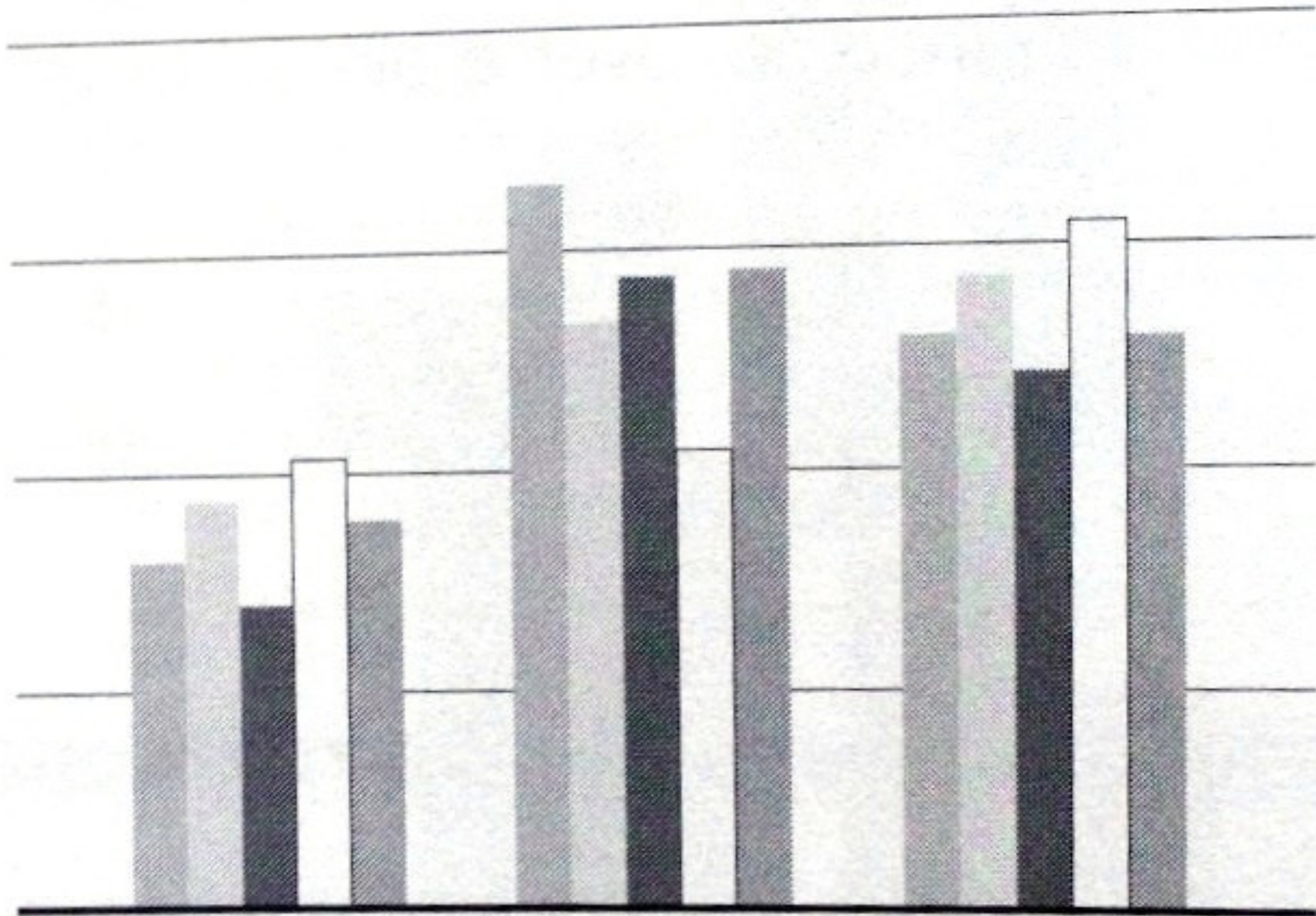
Misleading Bar Charts



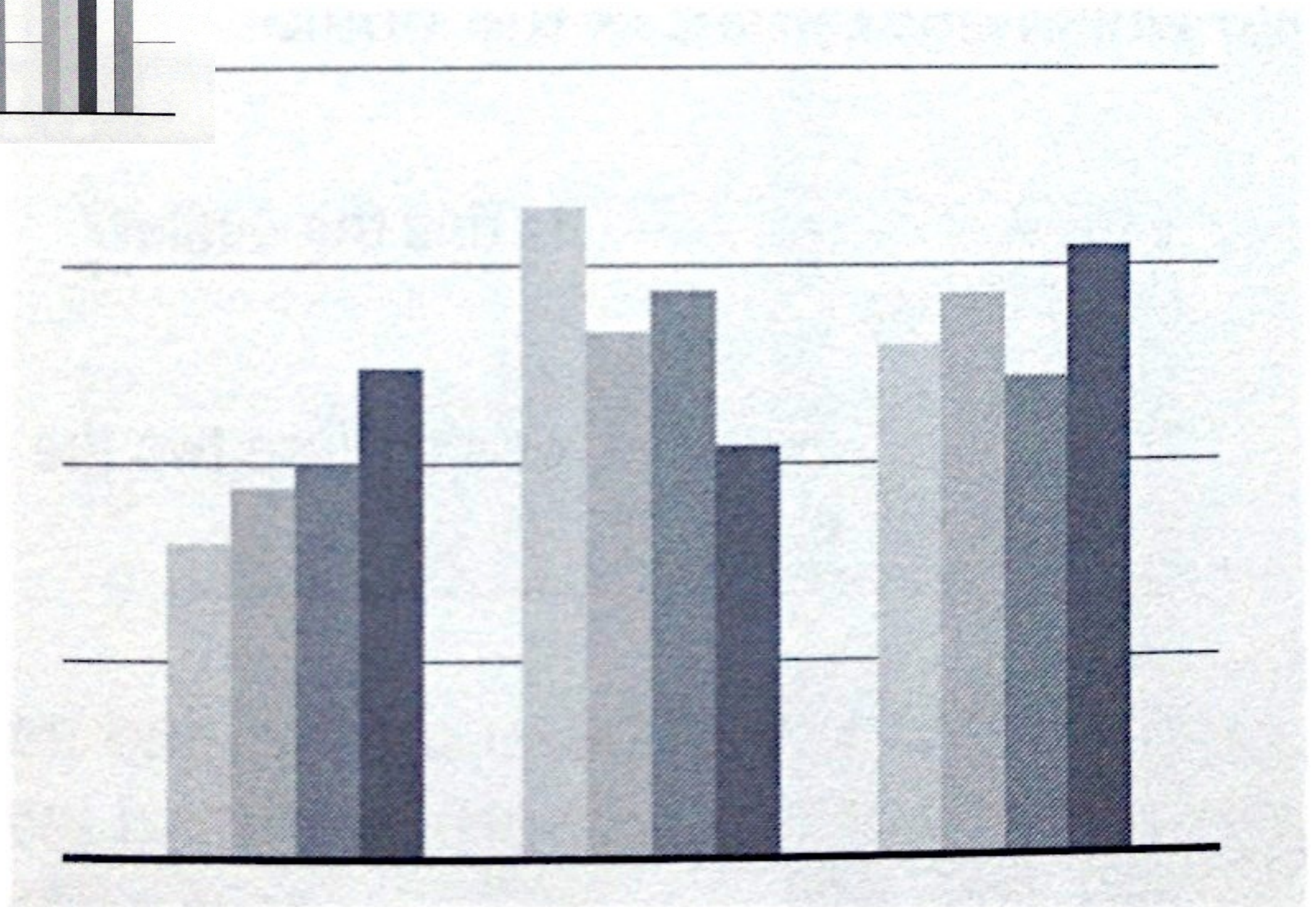
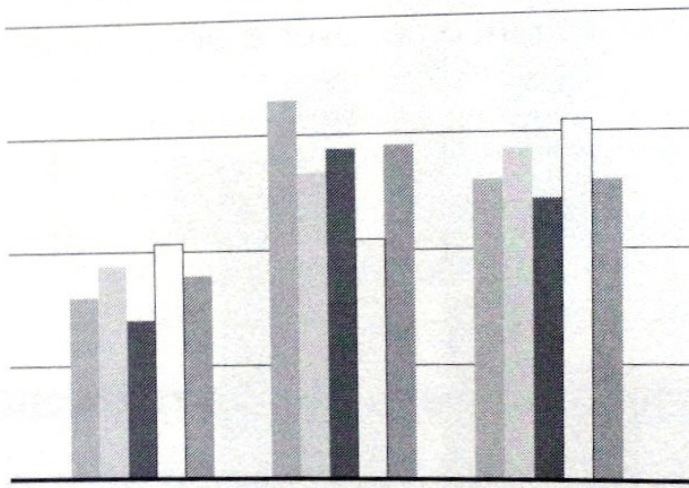


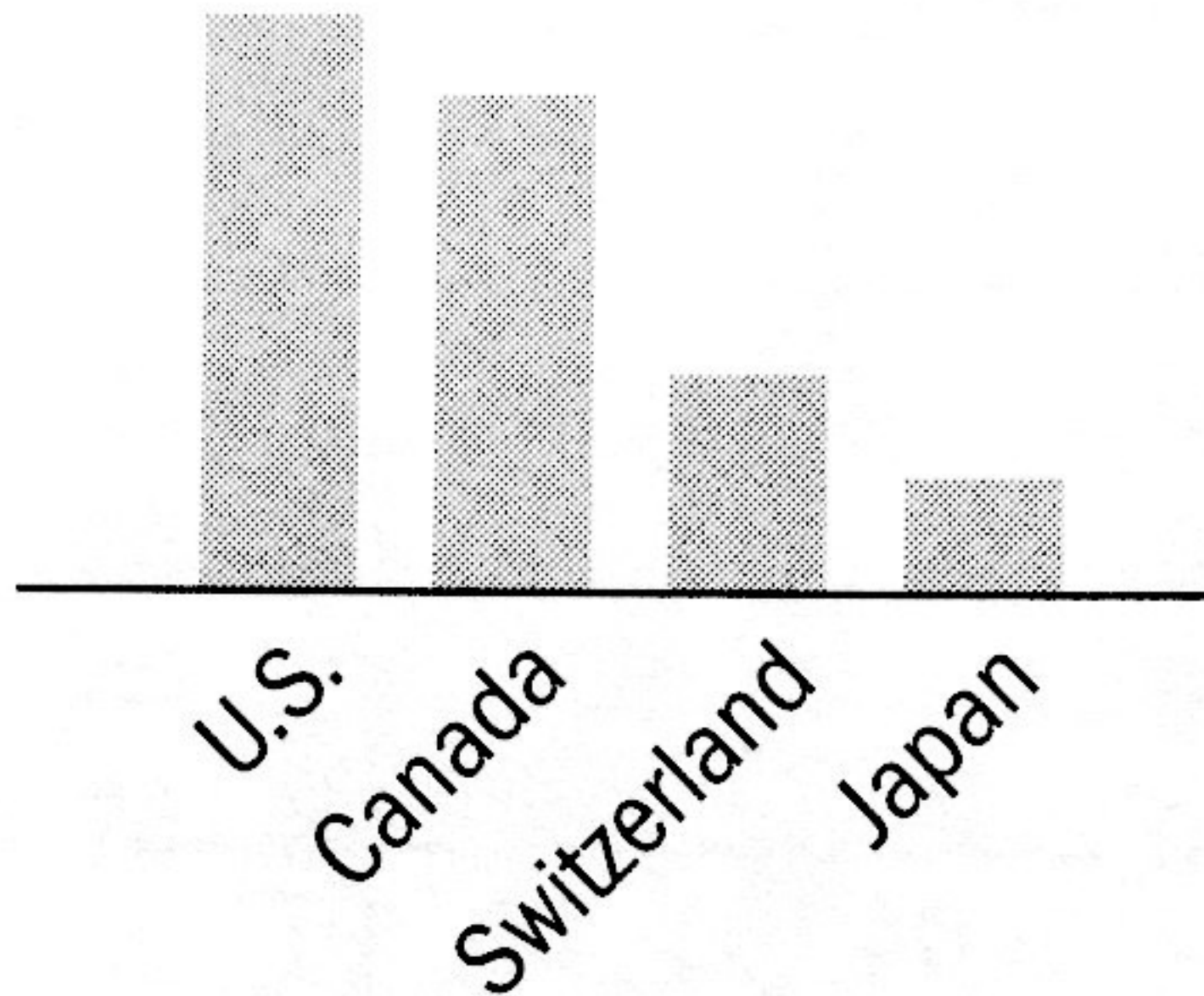
**Vertical axis of bar charts
should start at 0, almost always**

Disorienting color bars

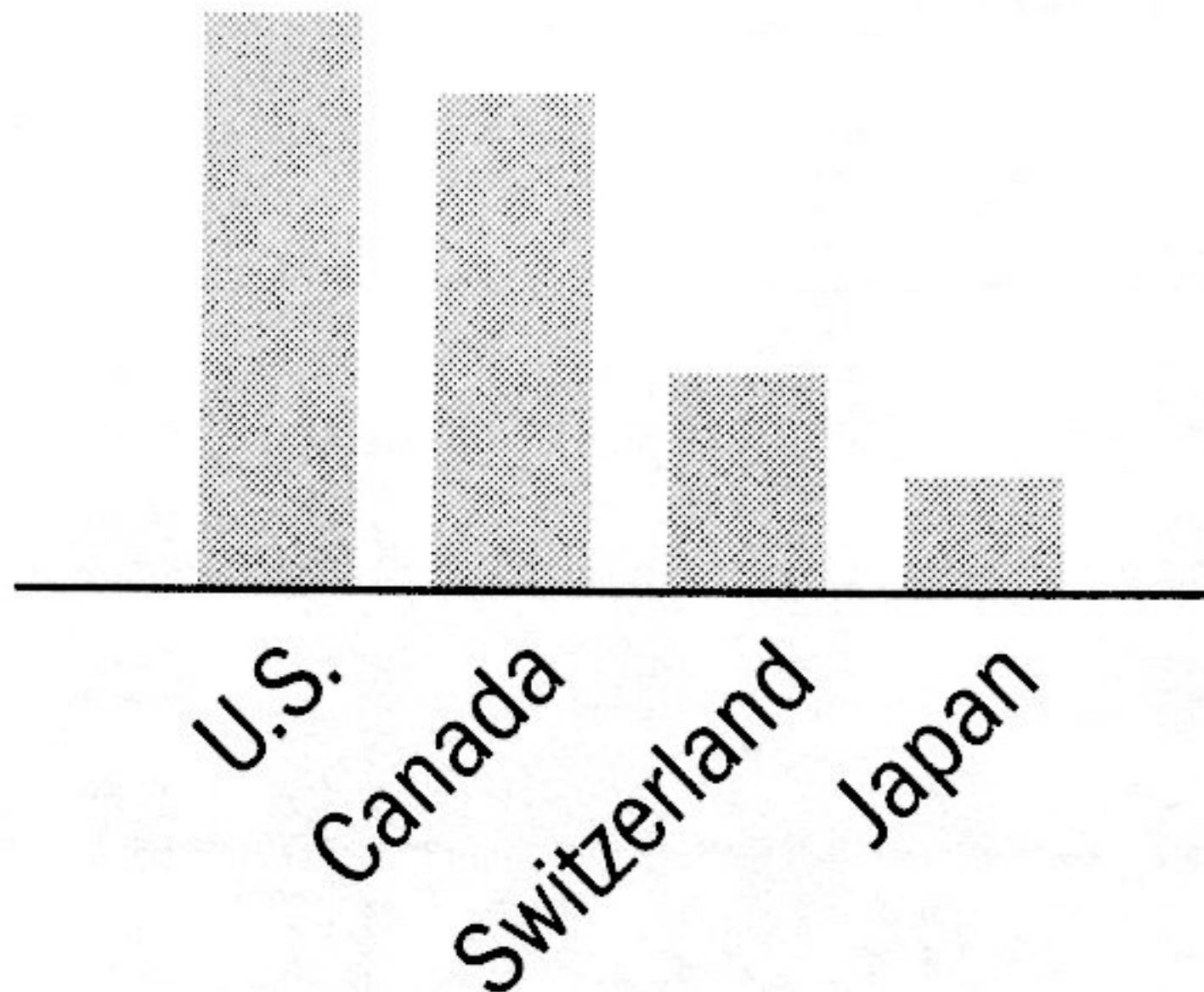


Use gradation

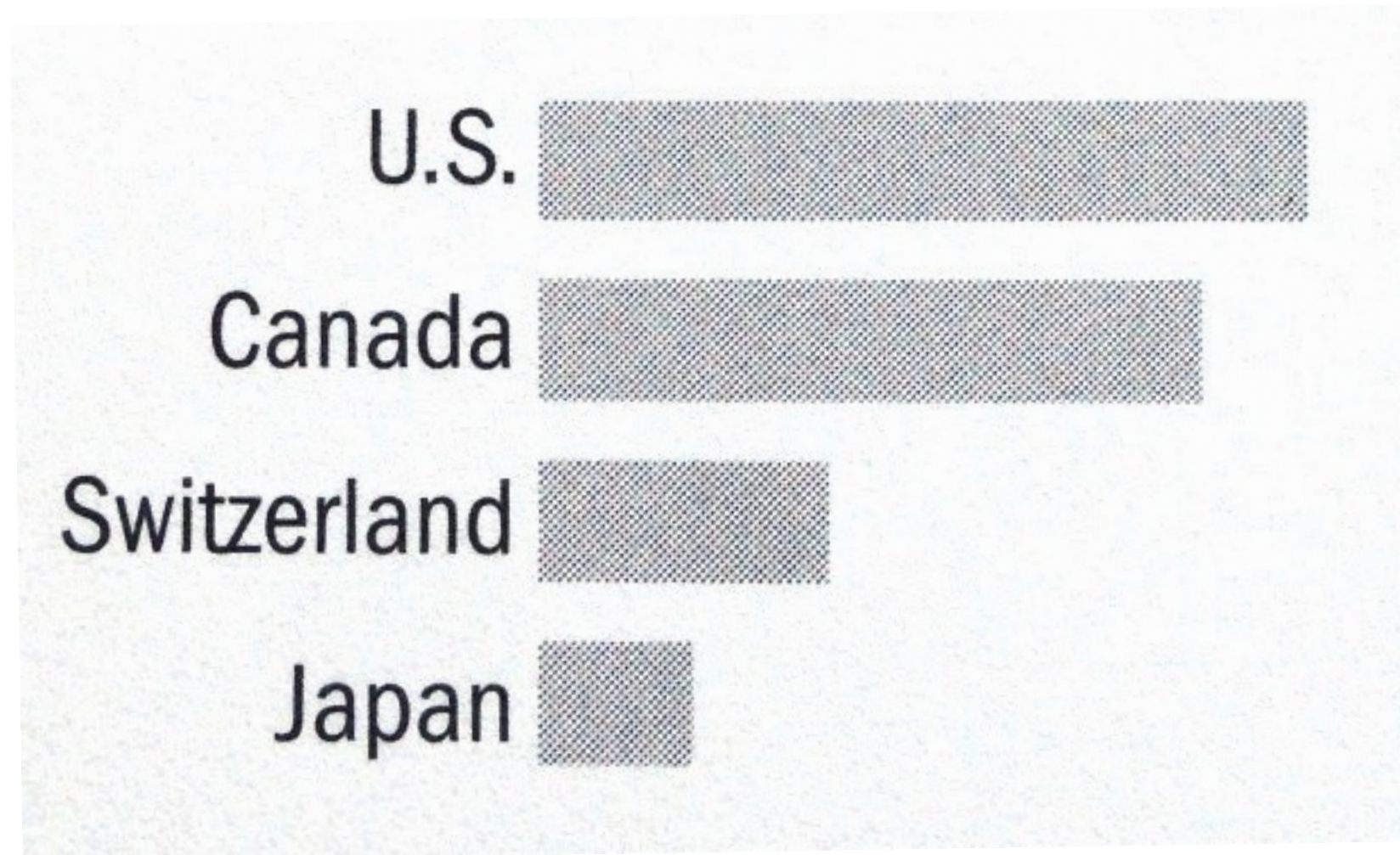




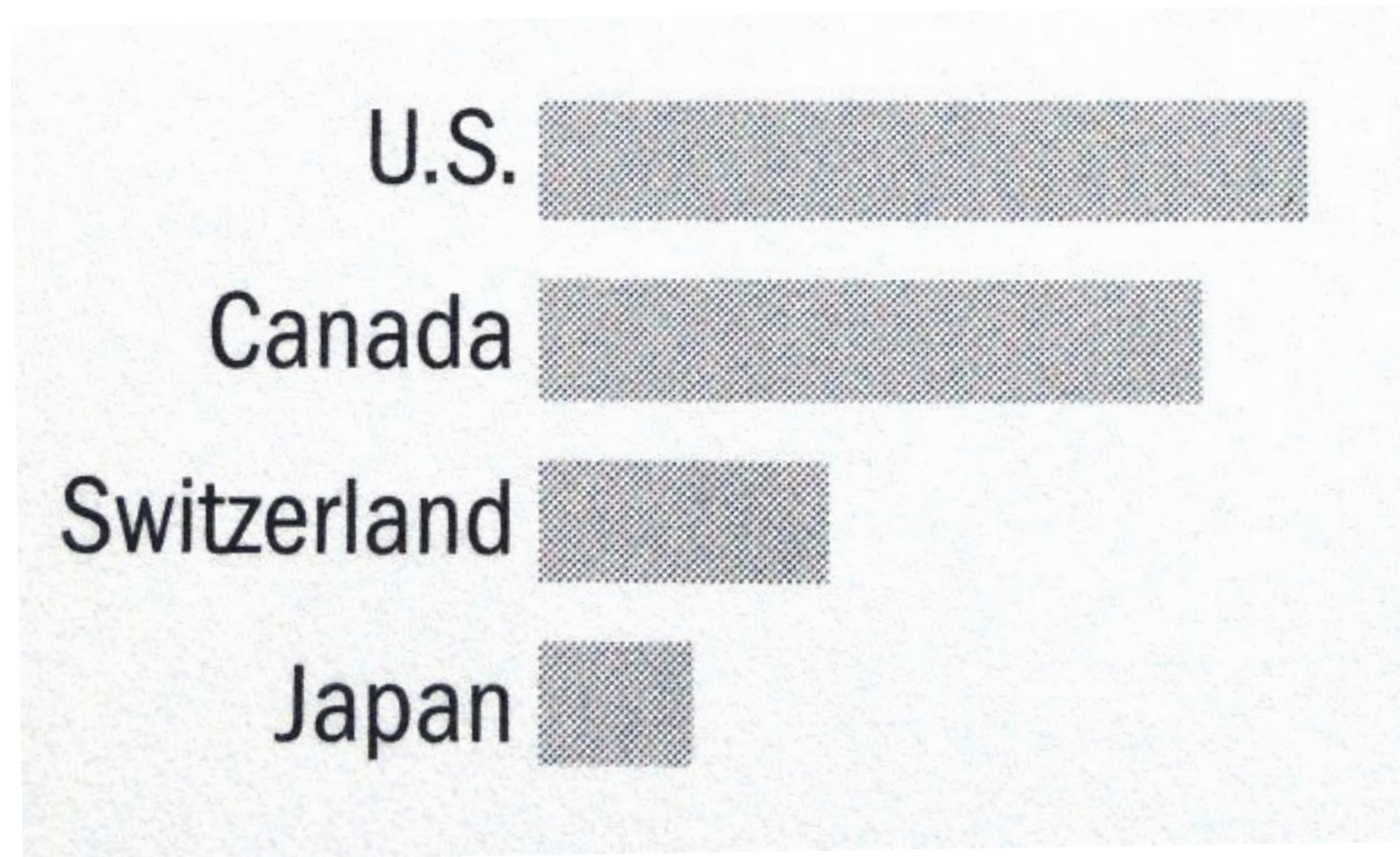
Avoid Tilted or Rotated Labels



Bars Can be **Horizontal**



Bars Can be **Horizontal**

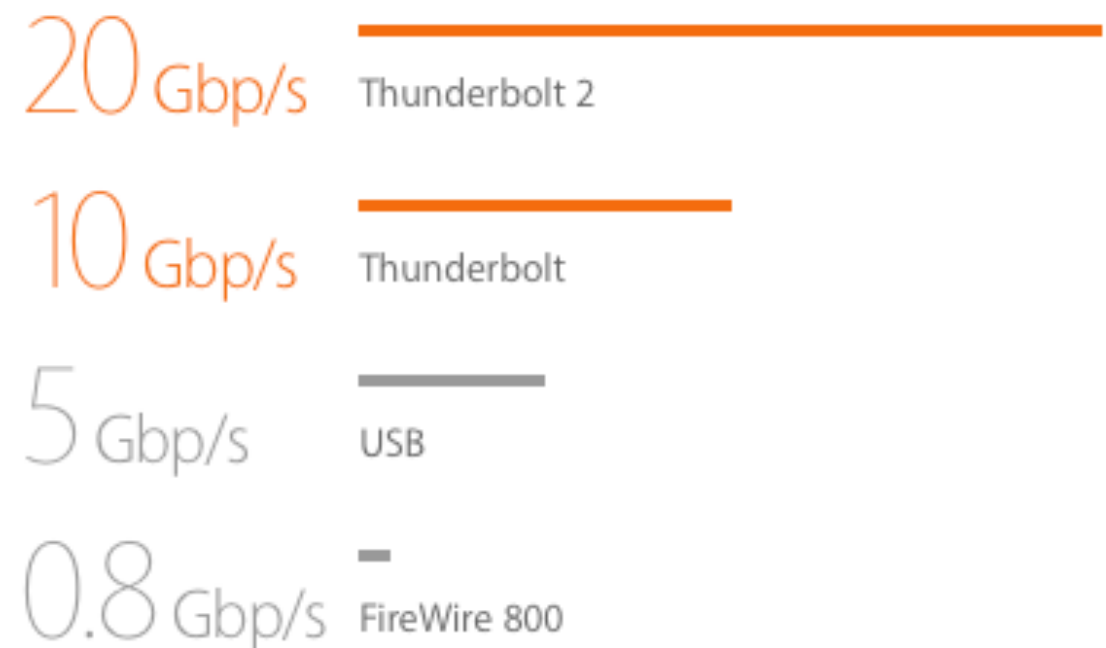


When labels are hard to read, try horizontal layout.
Don't settle for the default.

Thunderbolt 2

Blazing-fast data transfer.

Each of the two ultrafast, ultraflexible Thunderbolt 2 ports significantly expand the capabilities of iMac. For example, you can connect high-performance peripherals and move data up to 40 times faster than with USB 2, and up to 25 times faster than with FireWire 800. You also have more than enough bandwidth — up to 20 Gbps — to daisy-chain multiple high-speed devices and still maintain maximum throughput.



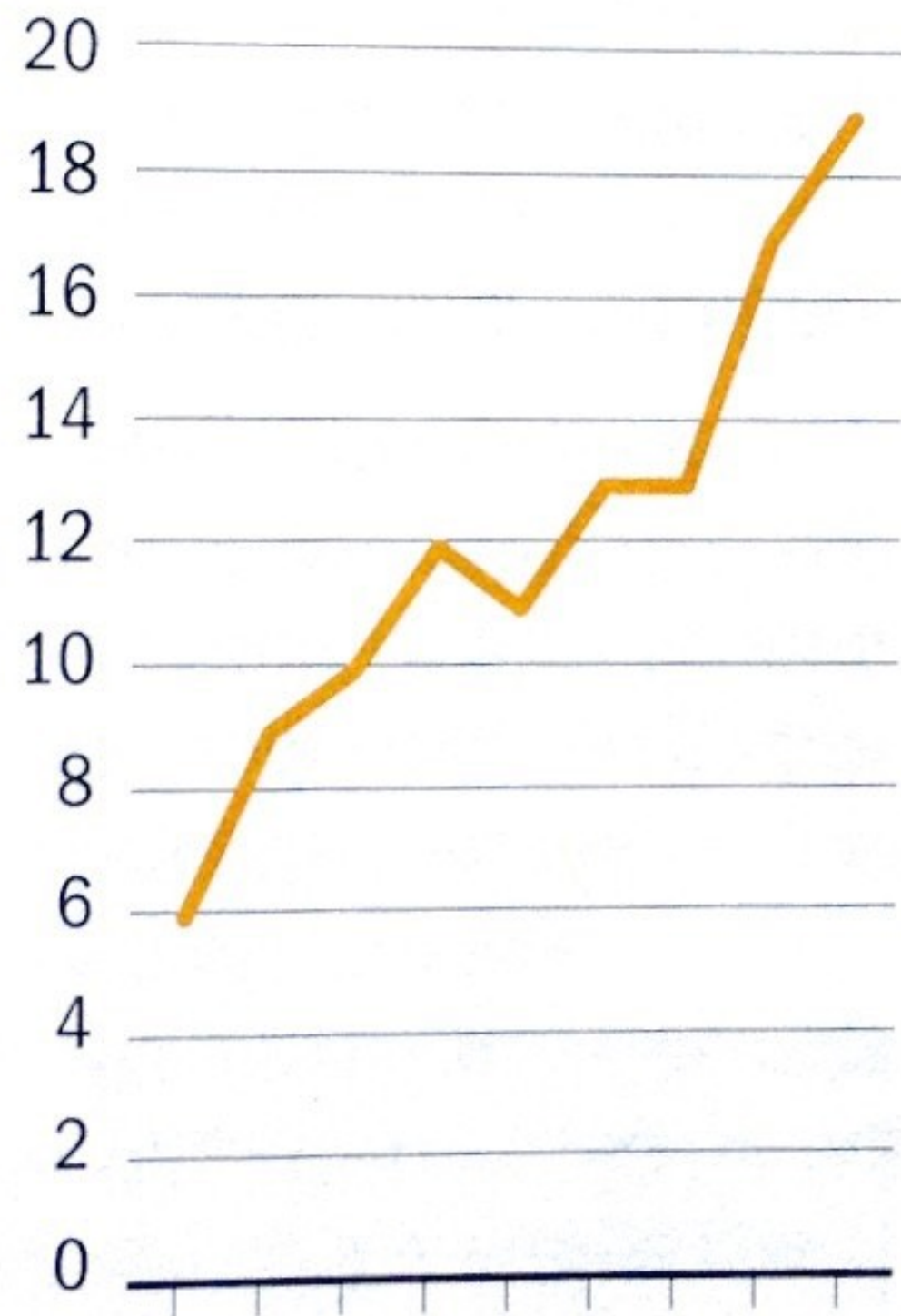
<http://www.apple.com/imac/performance/>

Line Charts (a.k.a. fever lines)



Can you improve the tick labels?

Use ticks at **common** intervals (e.g., 2, 5, 10, etc.)

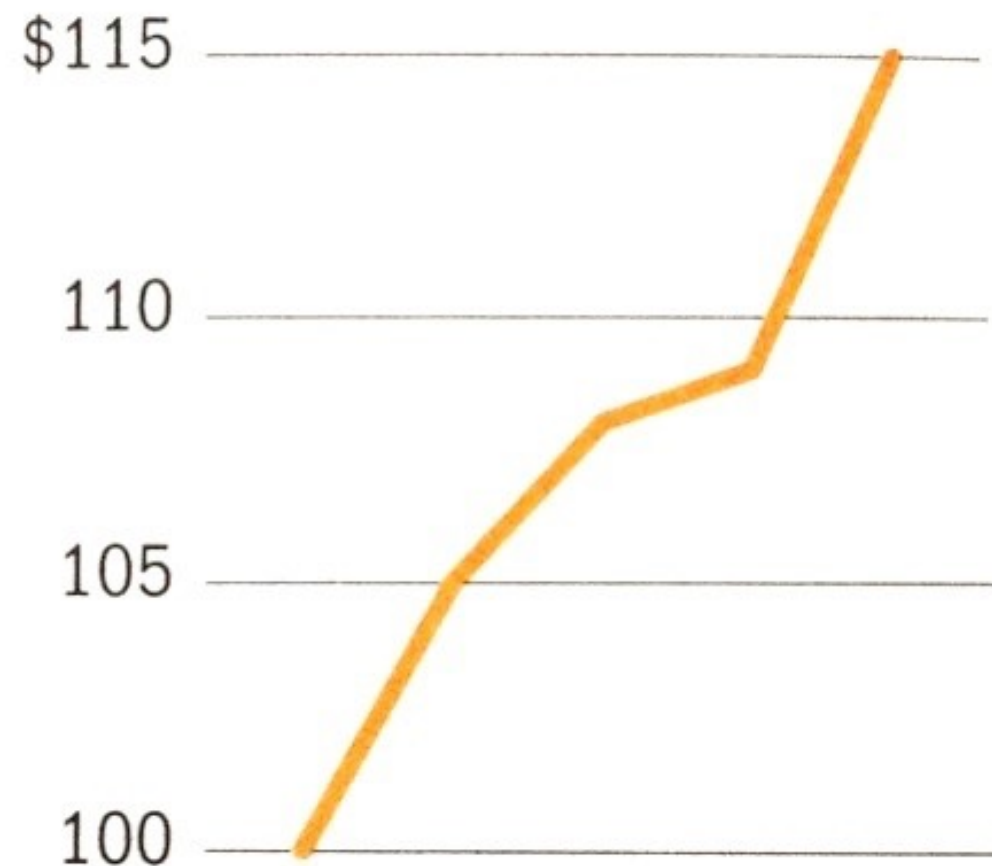


Too flat or too steep?

Too flat obscures the message

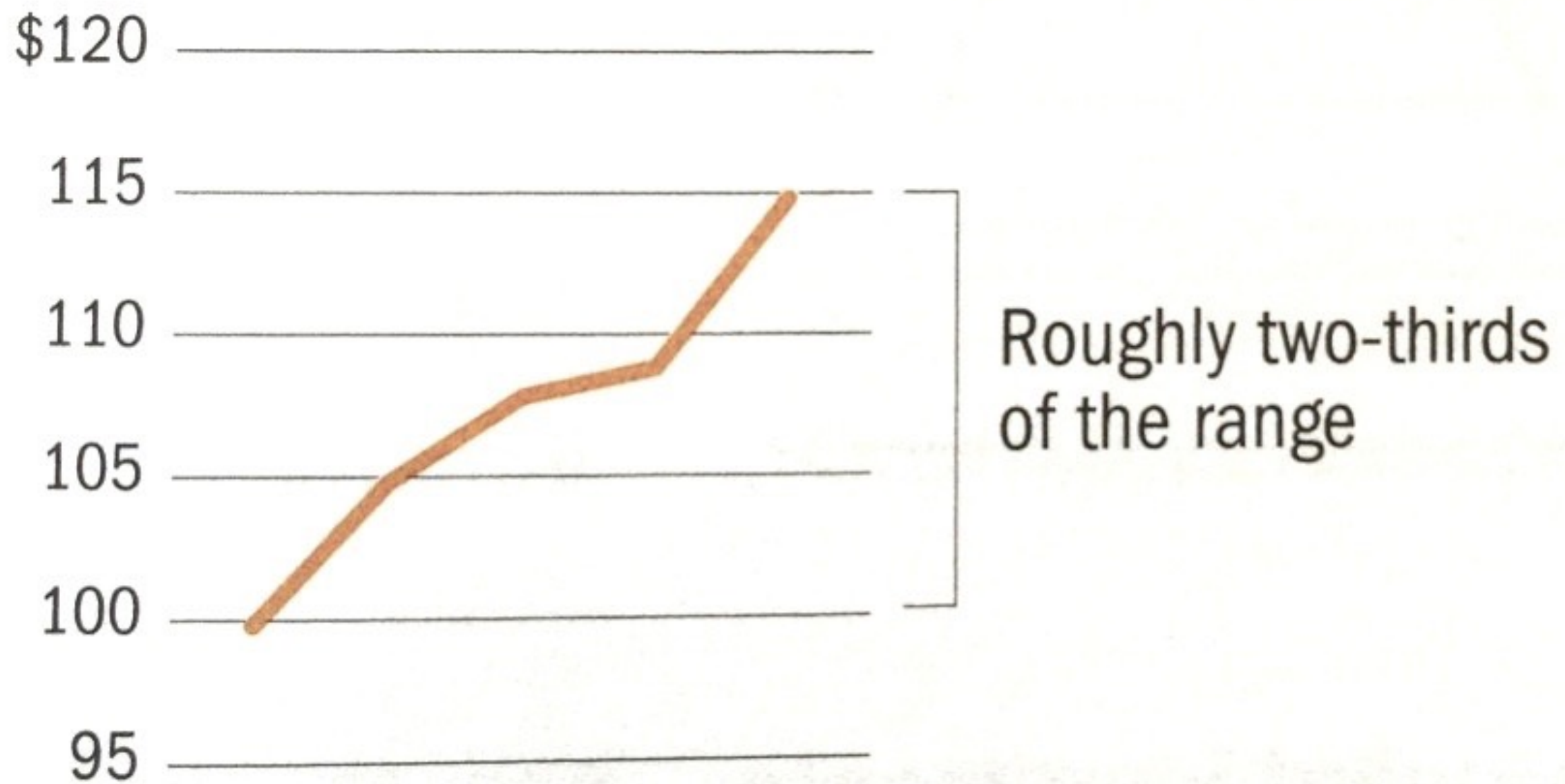


Too exaggerated overstates the trend

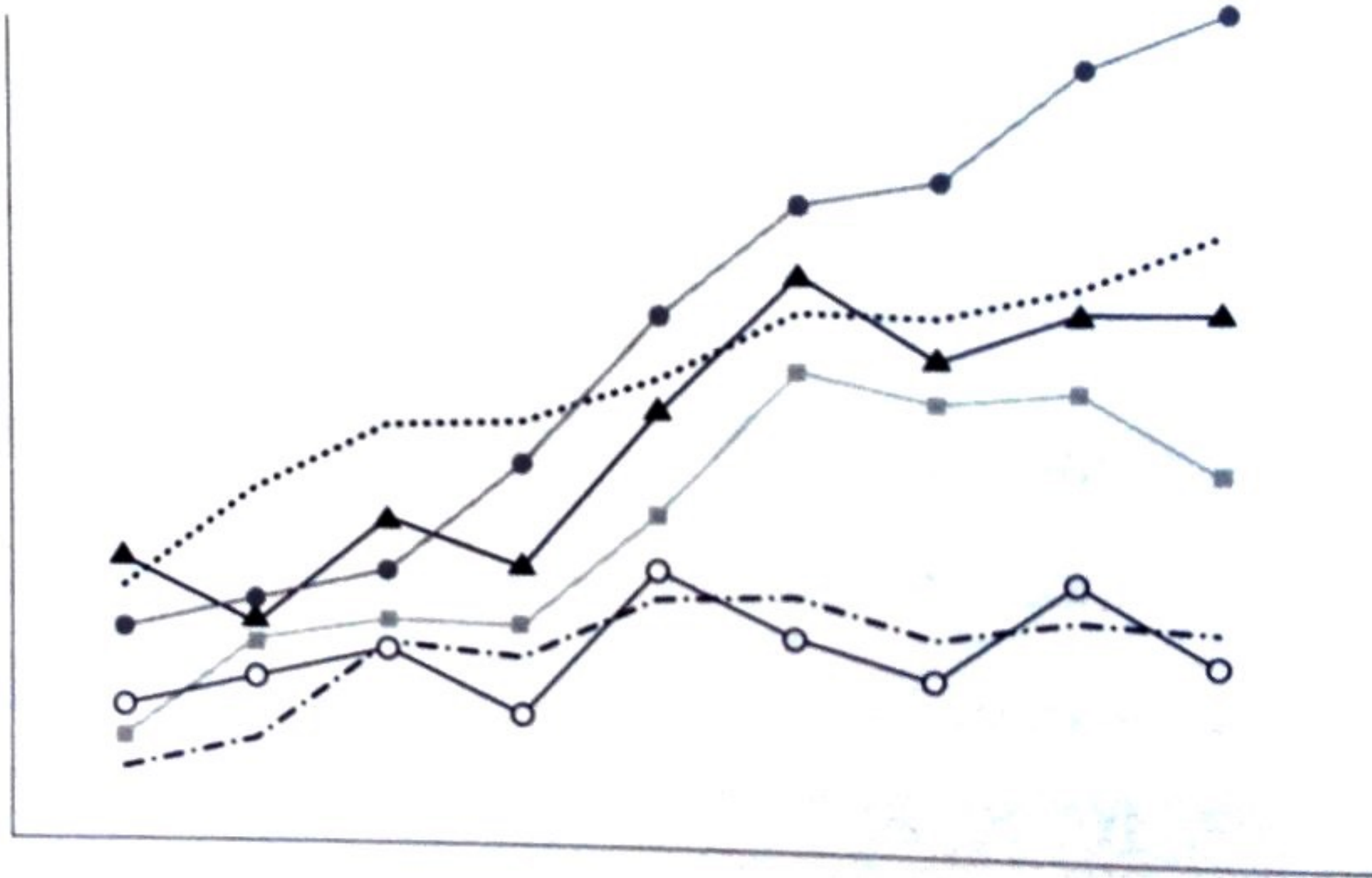


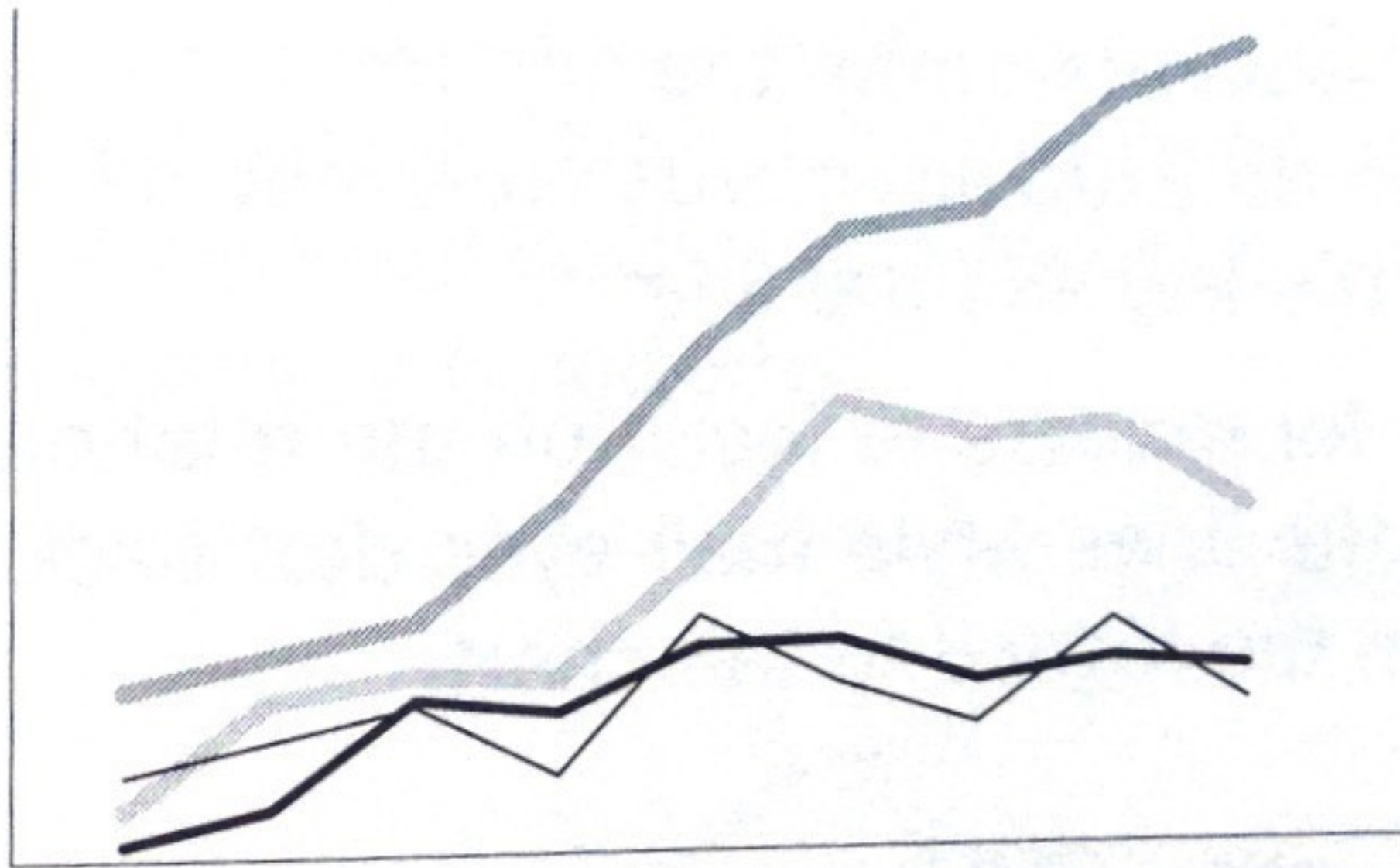
Note y-axis does not need to start at 0.
Why not as bad as in the case of bar chart?

Rule of Thumb



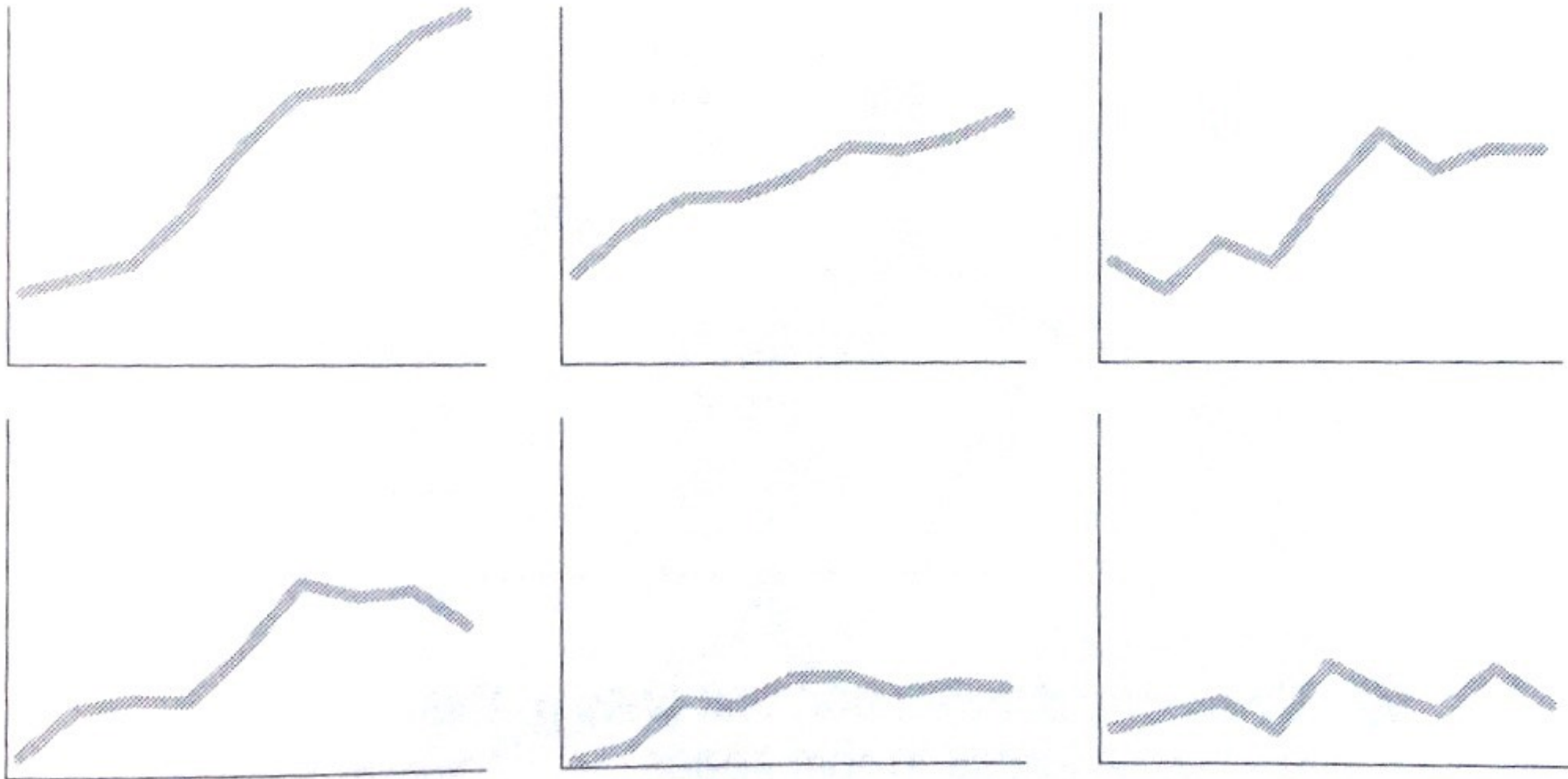
Multiple Patterned Lines in one chart





Better?

Note the “double encoding” of *line width* and *brightness*.
What if you have many lines you want to show?

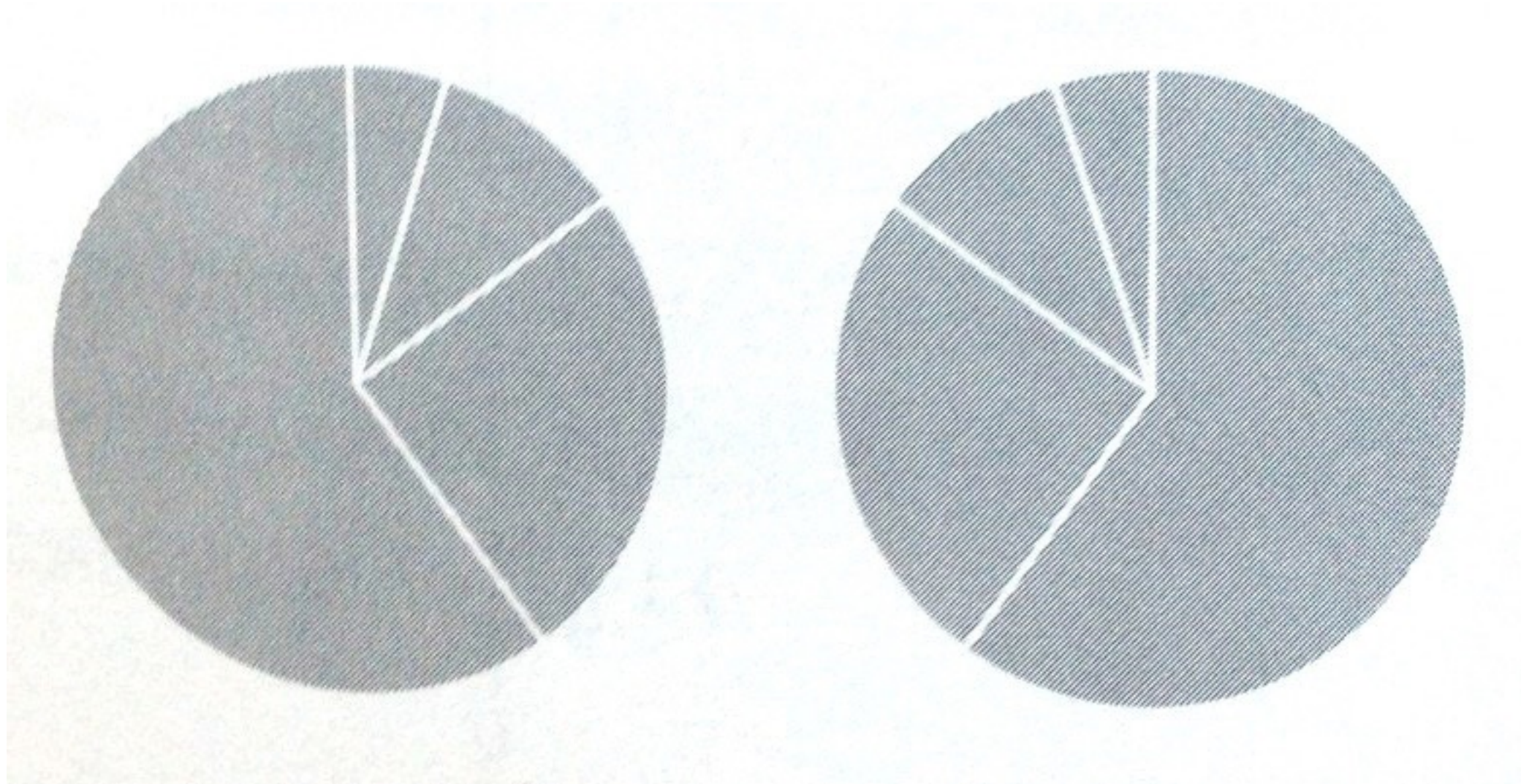


“Small Multiple” - Edward Tufte

Better than overlapping (sometimes)

“a series or grid of small similar graphics or charts, allowing them to be easily compared”

The Dreaded Pie Charts



Why people like to use pie charts?

U.S. SmartPhone Marketshare



WHAT 3-D PIE CHARTS ARE GOOD FOR.

META 3-D PIE CHARTS

PIE CHARTS MADE
OF EDIBLE PIES



ANNOYING BUSINESS
ANALYTICS PEOPLE,
CAUSING THEM TO DIE
A LITTLE BIT INSIDE

MAKING EDWARD TUFTE
CRY, OR ROLL OVER IN HIS
GRAVE EVEN THOUGH HE'S
STILL ALIVE; ALSO POSSIBLY
KILL KITTENS

2012 PRESIDENTIAL RUN

GOP CANDIDATES

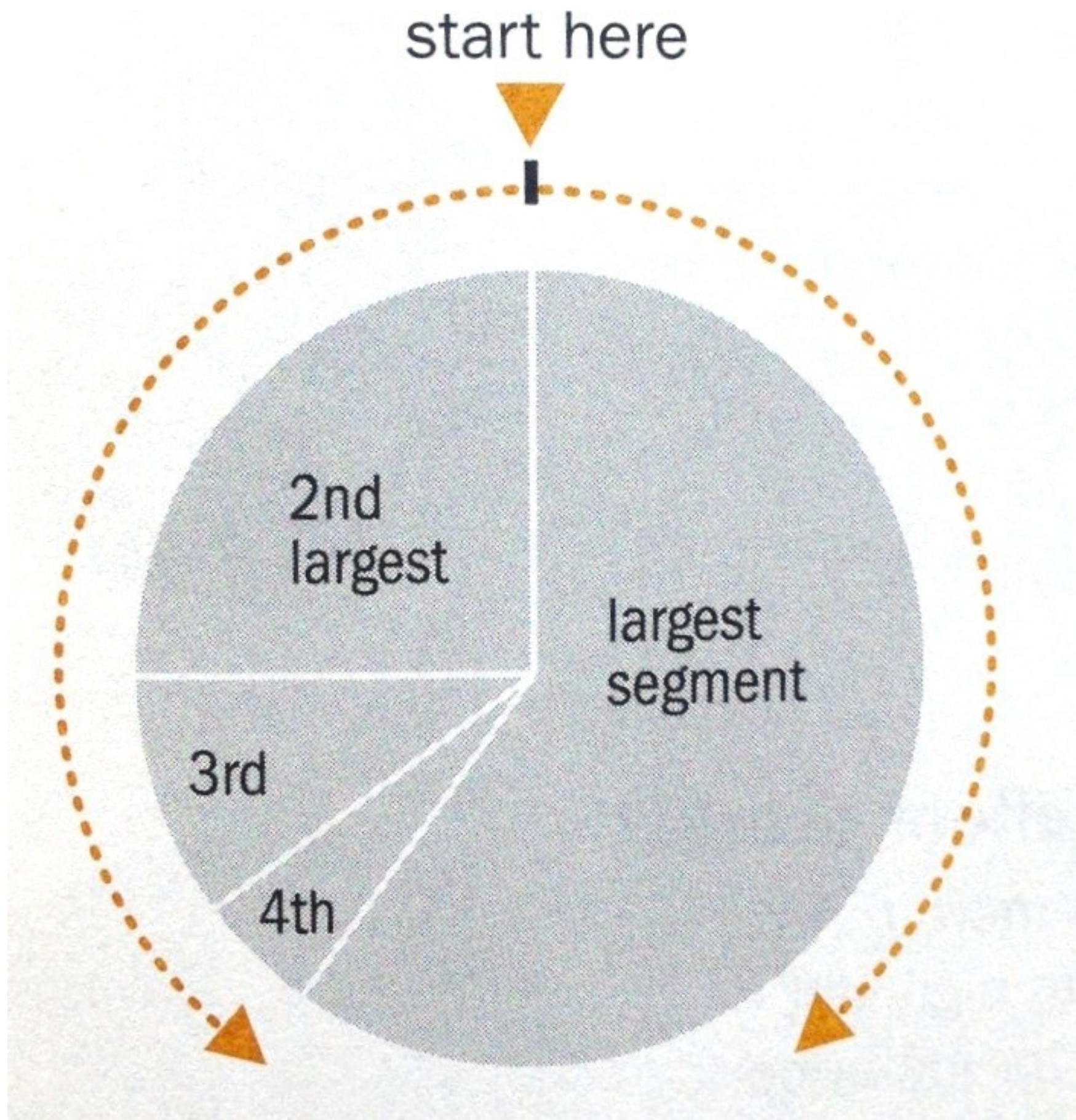


FOX

47°

SOURCE: OPINIONS

DYNAMIC

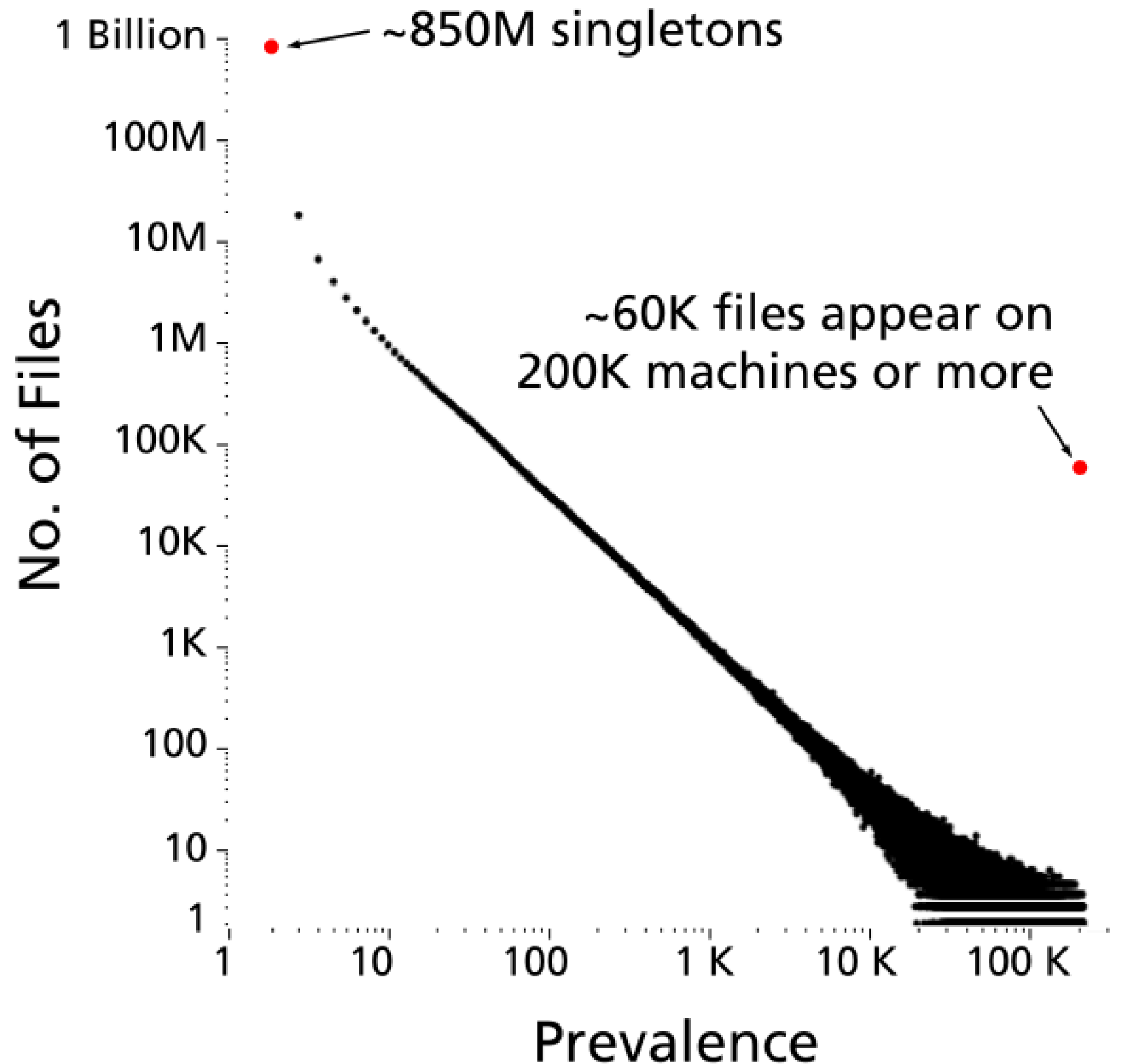




Log scale instead of linear scale

Include numbers from different orders of magnitude

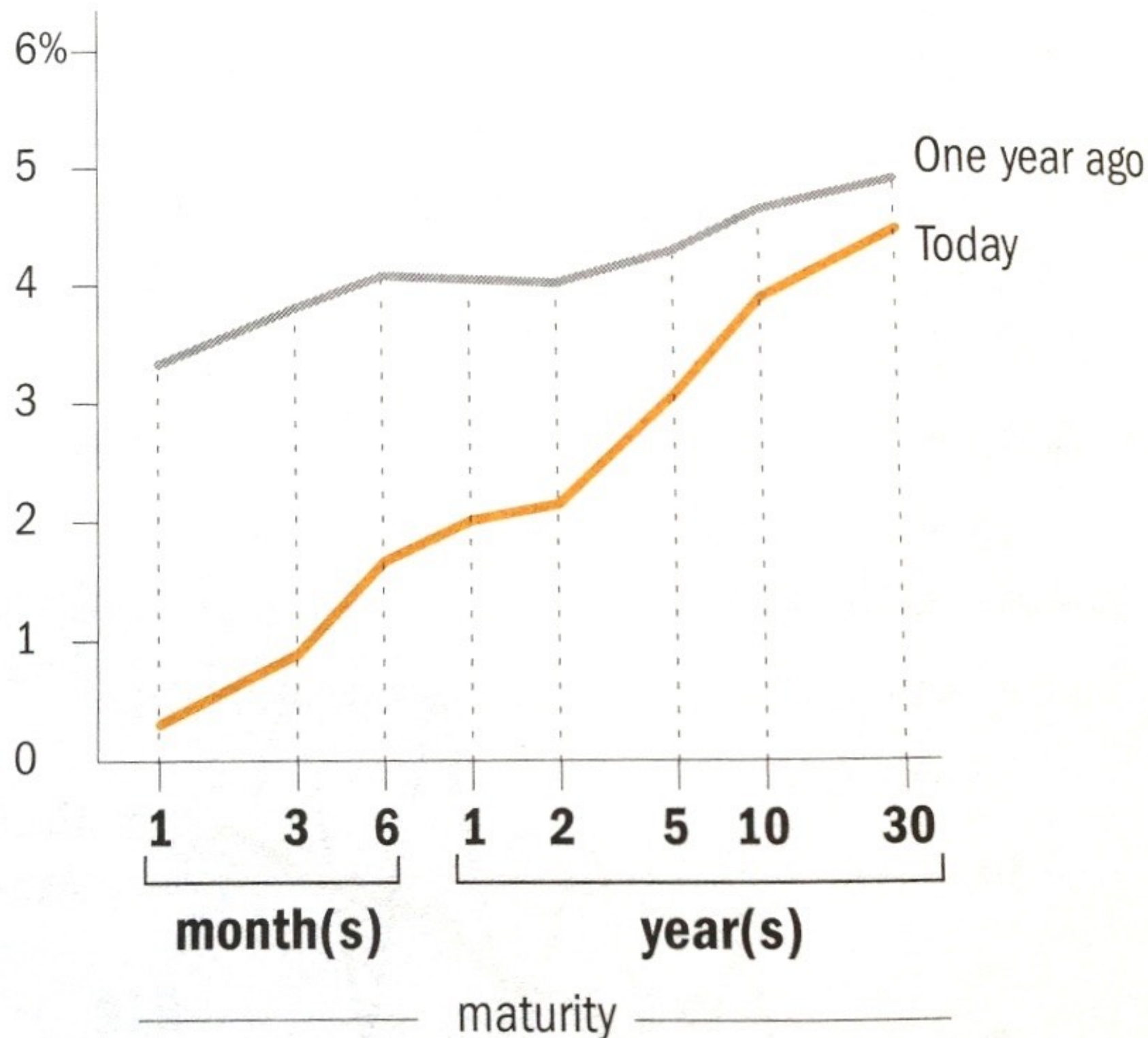
log-log



“log” also works well for time

Example

The yield curve of Treasury bills, notes and bonds



In-class Exercise.

Applying what you have just learned.

How to fix the defaults

Remove
to improve
the **data tables** edition

Created by Darkhorse Analytics

www.darkhorseanalytics.com

<http://www.darkhorseanalytics.com/blog/clear-off-the-table>

How to fix the defaults

Remove
to improve
the **data tables** edition

Created by Darkhorse Analytics

www.darkhorseanalytics.com

<http://www.darkhorseanalytics.com/blog/clear-off-the-table>

Great Work destroyed by **Poor Presentation**

Bad color schemes

Bad, tiny fonts

Too much animation

Too much data

can you read this?

100 times faster!

Don McMillan: Life After Death by PowerPoint

http://www.youtube.com/watch?v=lpvgfmEU2Ck&feature=player_embedded

Practitioners' Guide

Colors: start with black & white

Forces you to focus on content and layout

Then add **colors**, carefully

Fonts: sans-serif generally easier to read

On Mac: Helvetica is great start

On Windows: Arial?

Animation: start with **no** animation, then add meaningful ones

Practitioners' Guide: Use Pictures and Videos

“Pictures” include tables, diagrams, charts, etc.

- Pictures often more succinct & memorable
- People like pictures and love movies

And show them ASAP!

Once people fall asleep, it's hard to wake them up!

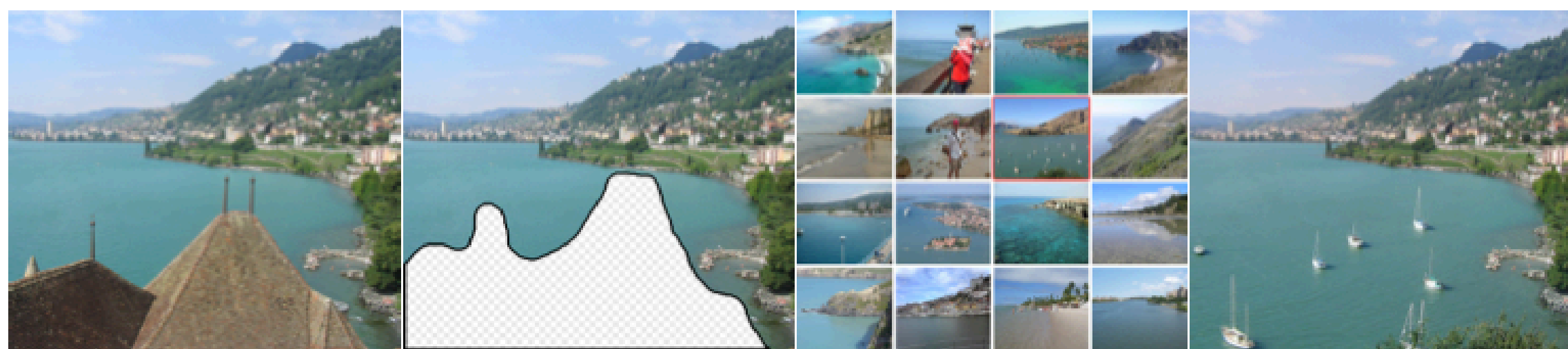
If you have good stuff, show them now.

Scene Completion Using Millions of Photographs

James Hays

Alexei A. Efros

Carnegie Mellon University



Original Image

Input

Scene Matches

Output

Figure 1: Given an input image with a missing region, we use matching scenes from a large collection of photographs to complete the image.

Abstract

What can you do with a million images? In this paper we present a new image completion algorithm powered by a huge database of photographs gathered from the Web. The algorithm patches up holes in images by finding similar image regions in the database that are not only seamless but also semantically valid. Our chief insight is that while the space of images is effectively infinite, the space of semantically differentiable scenes is actually not that large. For many image completion tasks we are able to find similar scenes which contain image fragments that will convincingly complete the image. Our algorithm is entirely data-driven, requiring no annotations or labelling by the user. Unlike existing image completion methods, our algorithm can generate a diverse set of results for each input image and we allow users to select among them. We demon-

There are two fundamentally different strategies for image completion. The first aims to reconstruct, as accurately as possible, the data that *should have been* there, but somehow got occluded or corrupted. Methods attempting an accurate reconstruction have to use some other source of data in addition to the input image, such as video (using various background stabilization techniques, e.g. [Irani et al. 1995]) or multiple photographs of the same physical scene [Agarwala et al. 2004; Snavely et al. 2006].

The alternative is to try finding a plausible way to fill in the missing pixels, hallucinating data that *could have been* there. This is a much less easily quantifiable endeavor, relying instead on the studies of human visual perception. The most successful existing methods [Criminisi et al. 2003; Drori et al. 2003; Wexler et al. 2004; Wilczkowiak et al. 2005; Komodakis 2006] operate by extending

Practitioners' Guide: Additional Tips for PhD students / Researchers

Crown-jewel pictures are important

- Overview of what readers is going to get — **cut to the chase**
- People skim and look at “interesting” things first
- Reviewers are busy and sleepy 😴 (read 5-10 papers per conference) — it's refreshing to read an interesting paper

How to do it?

- Use your **most impressive** figure
- Can be similar to another shown later

Figures should be self-contained.

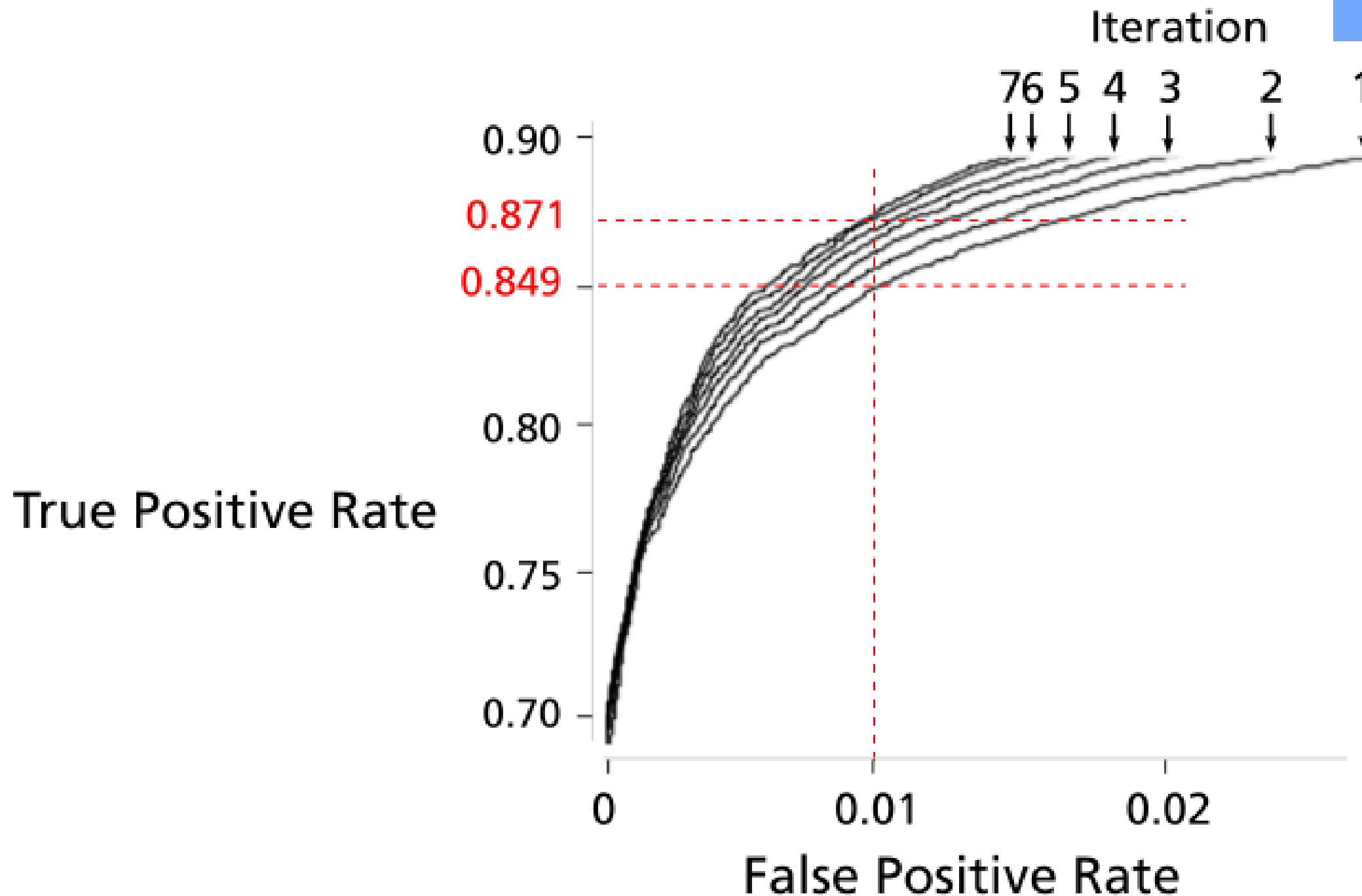


Figure 8: ROC curves of 7 iterations; true positive rate incrementally improves.

Figures should be self-contained

Why?

- Don't make people go back and forth between text & figure
- Bad figures means **bad first impression** (reject!)

How to fix?

- Succinctly describe your main (take-away) messages

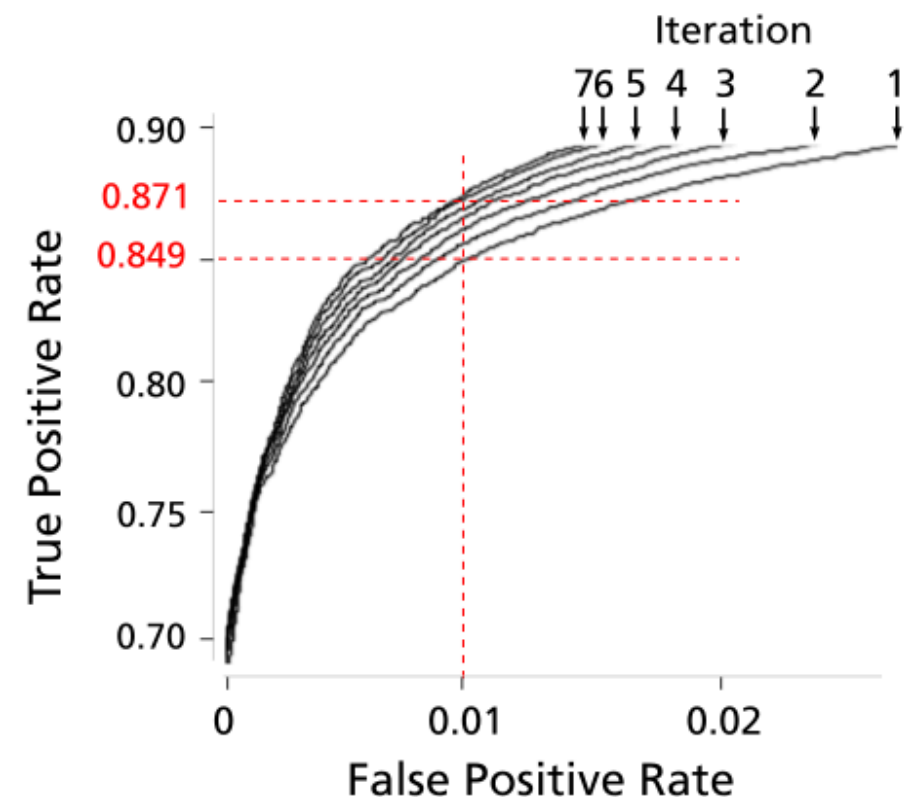


Figure 8: ROC curves of 7 iterations; true positive rate incrementally improves.

Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning

Duen Horng “Polo” Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos

School of Computer Science

Carnegie Mellon University

Pittsburgh, PA 15213, USA

{dchau, nkittur, jasonh, christos}@cs.cmu.edu

ABSTRACT

Extracting useful knowledge from large network datasets has become a fundamental challenge in many domains, from scientific literature to social networks and the web. We introduce Apolo, a system that uses a mixed-initiative approach—combining visualization, rich user interaction and machine learning—to guide the user to incrementally and interactively explore large network data and make sense of it. Apolo engages the user in bottom-up sensemaking to gradually build up an understanding over time by starting small, rather than starting big and drilling down. Apolo also helps users find relevant information by specifying exemplars, and then using a machine learning method called Belief Propagation to infer which other nodes may be of interest. We evaluated Apolo with twelve participants in a between-subjects study, with the task being to find relevant new papers to update an existing survey paper. Using expert judges, participants using Apolo found significantly more relevant papers. Subjective feedback of Apolo was also very positive.

Author Keywords

Sensemaking, large network, Belief Propagation

ACM Classification Keywords

H.3.3 Information Storage and Retrieval: Relevance feedback; H.5.2 Information Interfaces and Presentation: User Interfaces

General Terms

Algorithms, Design, Human Factors

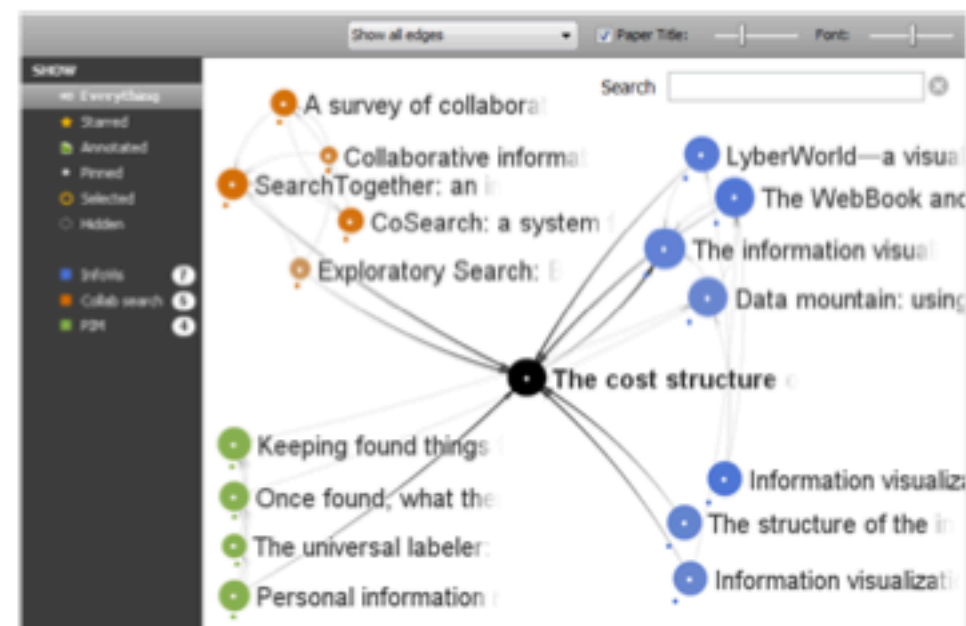


Figure 1. Apolo displaying citation network data around the article *The Cost Structure of Sensemaking*. The user gradually builds up a mental model of the research areas around the article by manually inspecting some neighboring articles in the visualization and specifying them as exemplar articles (with colored dots underneath) for some ad hoc groups, and instructs Apolo to find more articles relevant to them.

representation or schema of an information space that is useful for achieving the user’s goal [31]. For example, a scientist interested in connecting her work to a new domain must build up a mental representation of the existing literature in the new domain to understand and contribute to it.

For the above scientist, she may forage to find papers that she thinks are relevant, and build up a representation of how these papers relate to each other. As she continues to read

Use legible fonts.

If people can't see it, they won't appreciate it.

For printed materials, print them out and check!

Rule of thumb: about **7 lines** of text on a slide.

Apple UI Design Guidelines

UI Design Do's and Don'ts

iOS raises the bar for excellence in user interface design and offers great opportunities for you to deliver engaging and unique user experiences. Consider these common design concepts before you start coding to enhance the usability and appeal of your apps. To see more recommendations for great interface design, read the [iOS Human Interface Guidelines](#).

Interactivity Readability Graphics Clarity

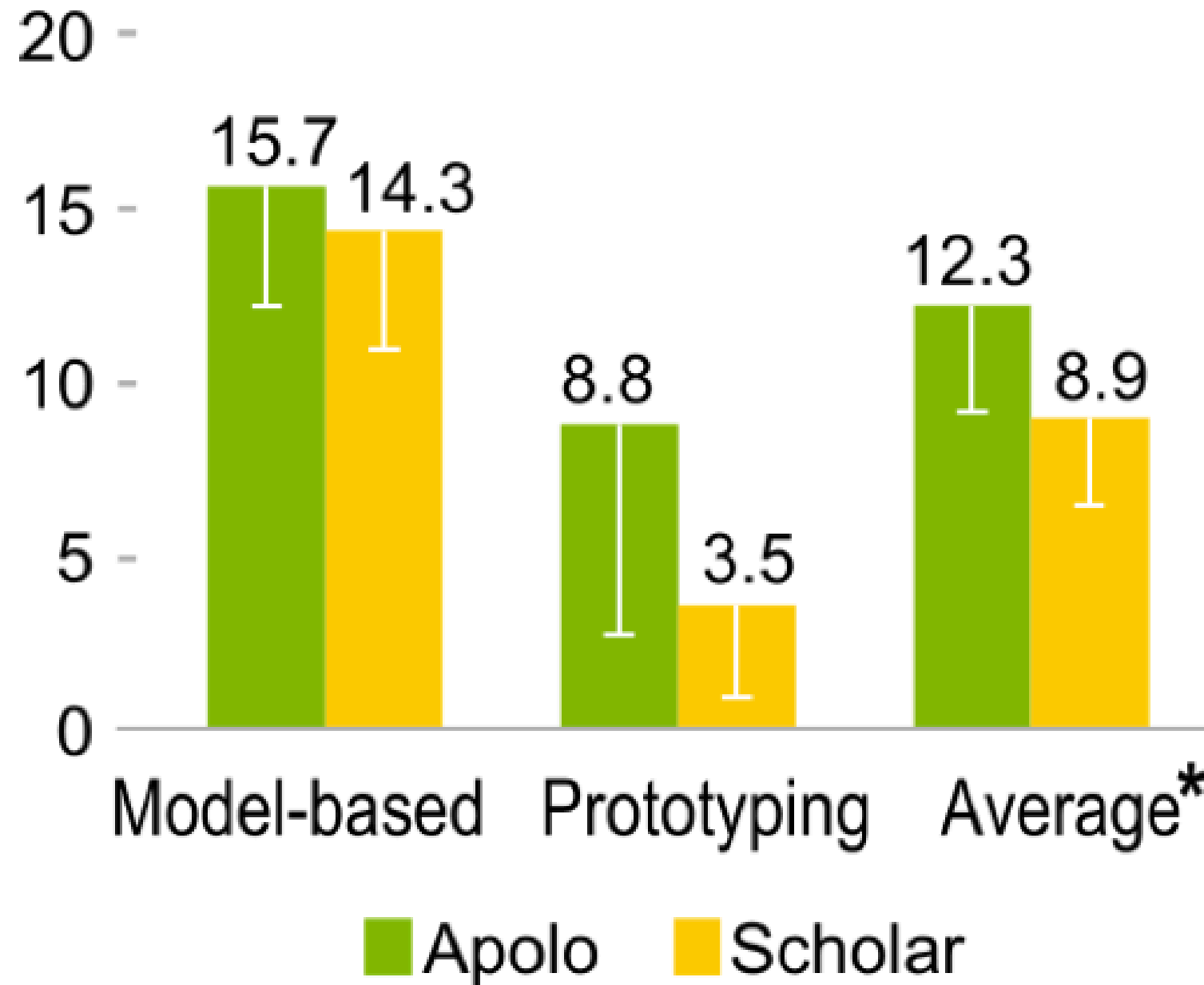
Formatting Content

Create a layout that fits the screen of an iOS device. Users should see primary content without zooming or scrolling horizontally.

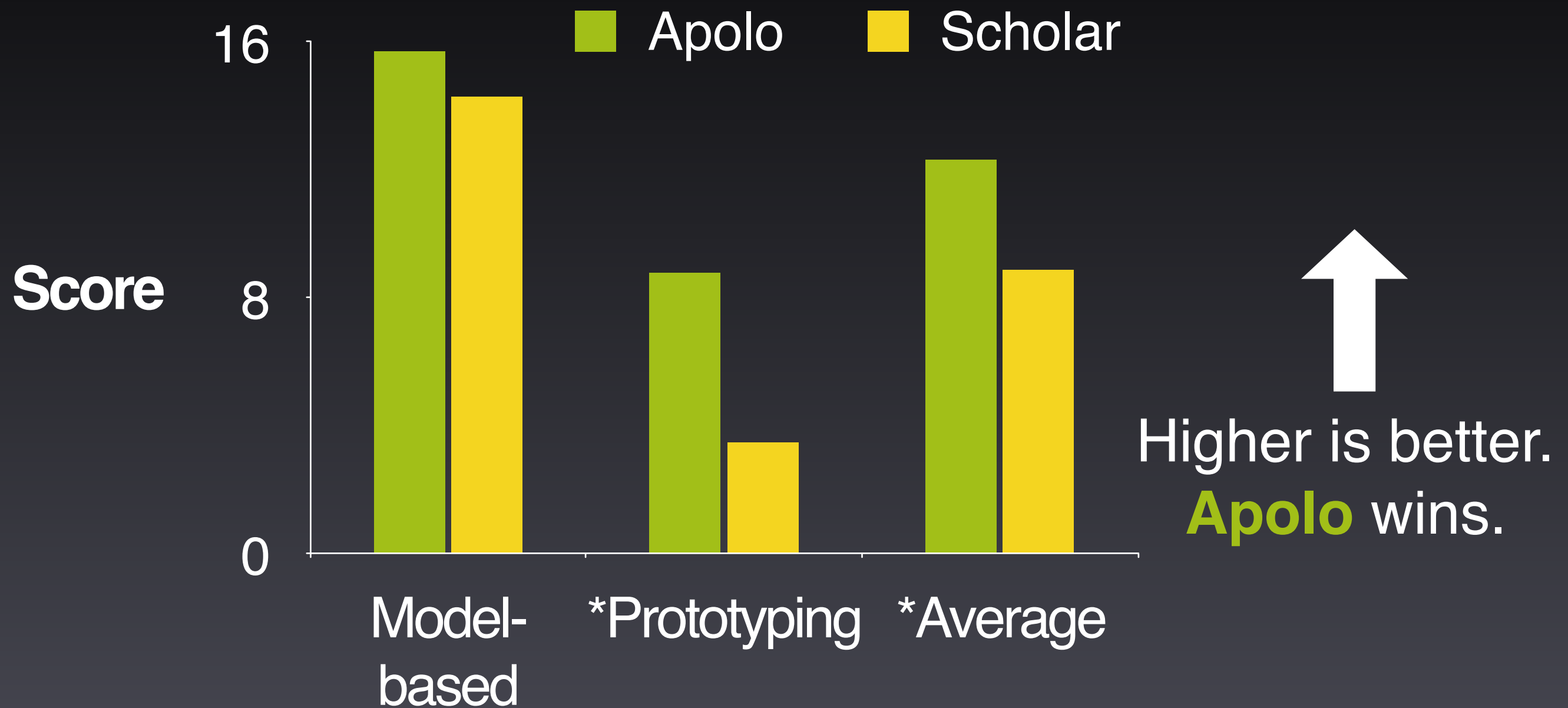
[Learn more >](#)



a) Avg Combined Judges' Scores



Judges' Scores



* Statistically significant, by *two-tailed t test*, $p < 0.05$