Where Are We? Feedback So far?

Collection

Cleaning

Integration

Analysis

Visualization

Presentation

Dissemination

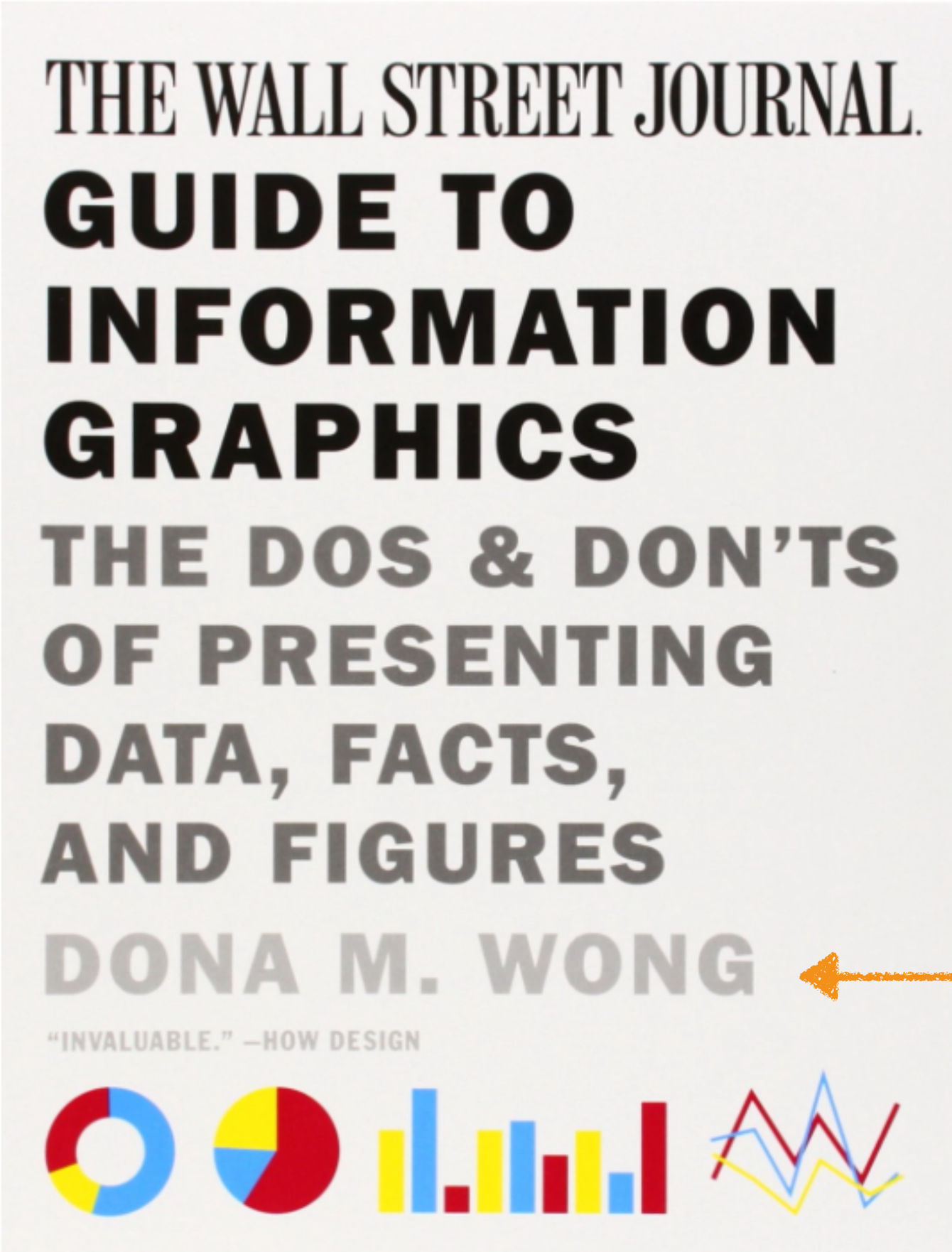
CSE 6242 / CX 4242

1. How to Identify Vis Issues?

2. Class Project

Duen Horng (Polo) Chau
Georgia Tech

Partly based on materials by
Professors Guy Lebanon, Jeffrey Heer, John Stasko, Christos Faloutsos



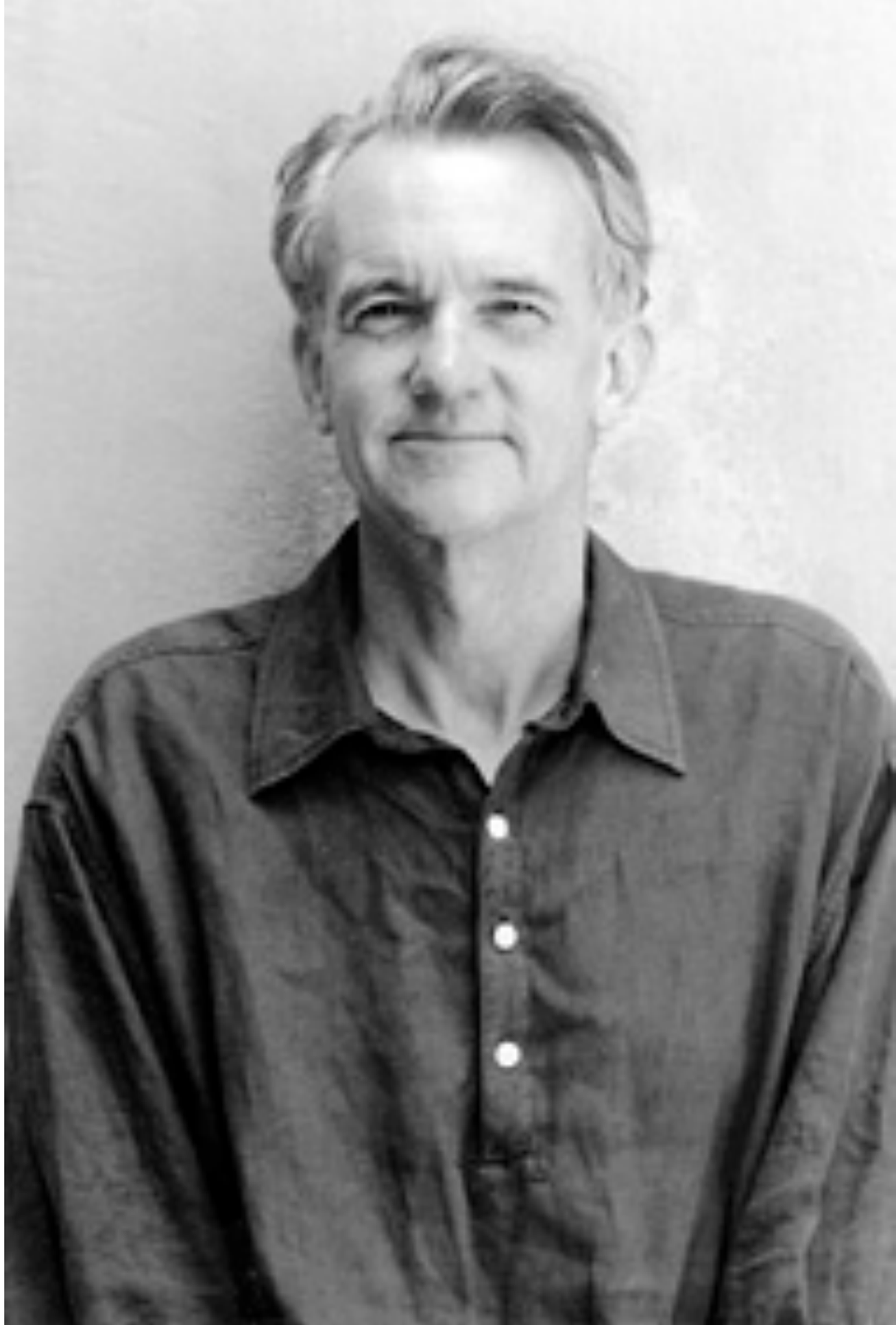
Student of Edward Tufte

Edward Tufte

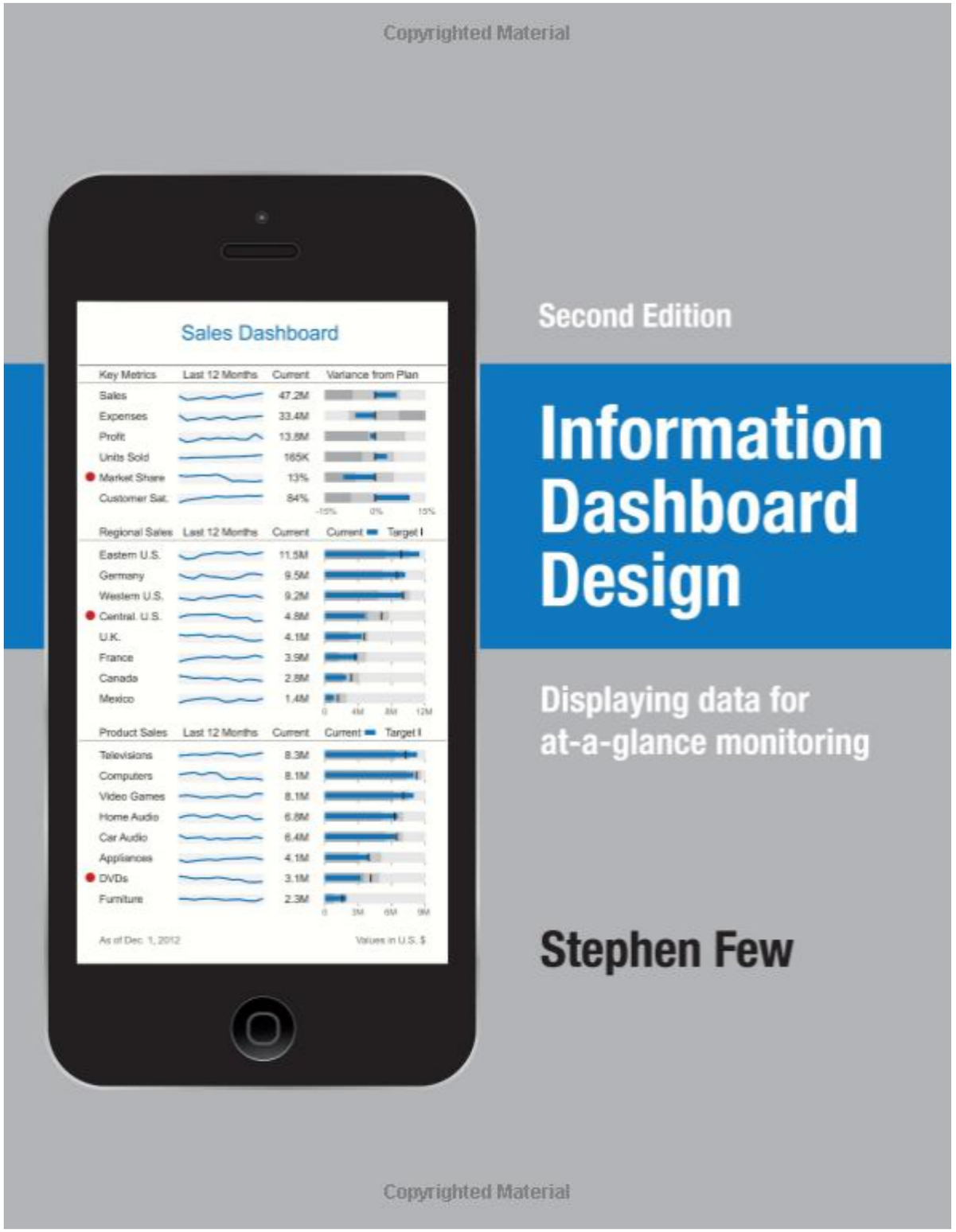
An American statistician and professor emeritus of political science, statistics, and computer science at Yale University.

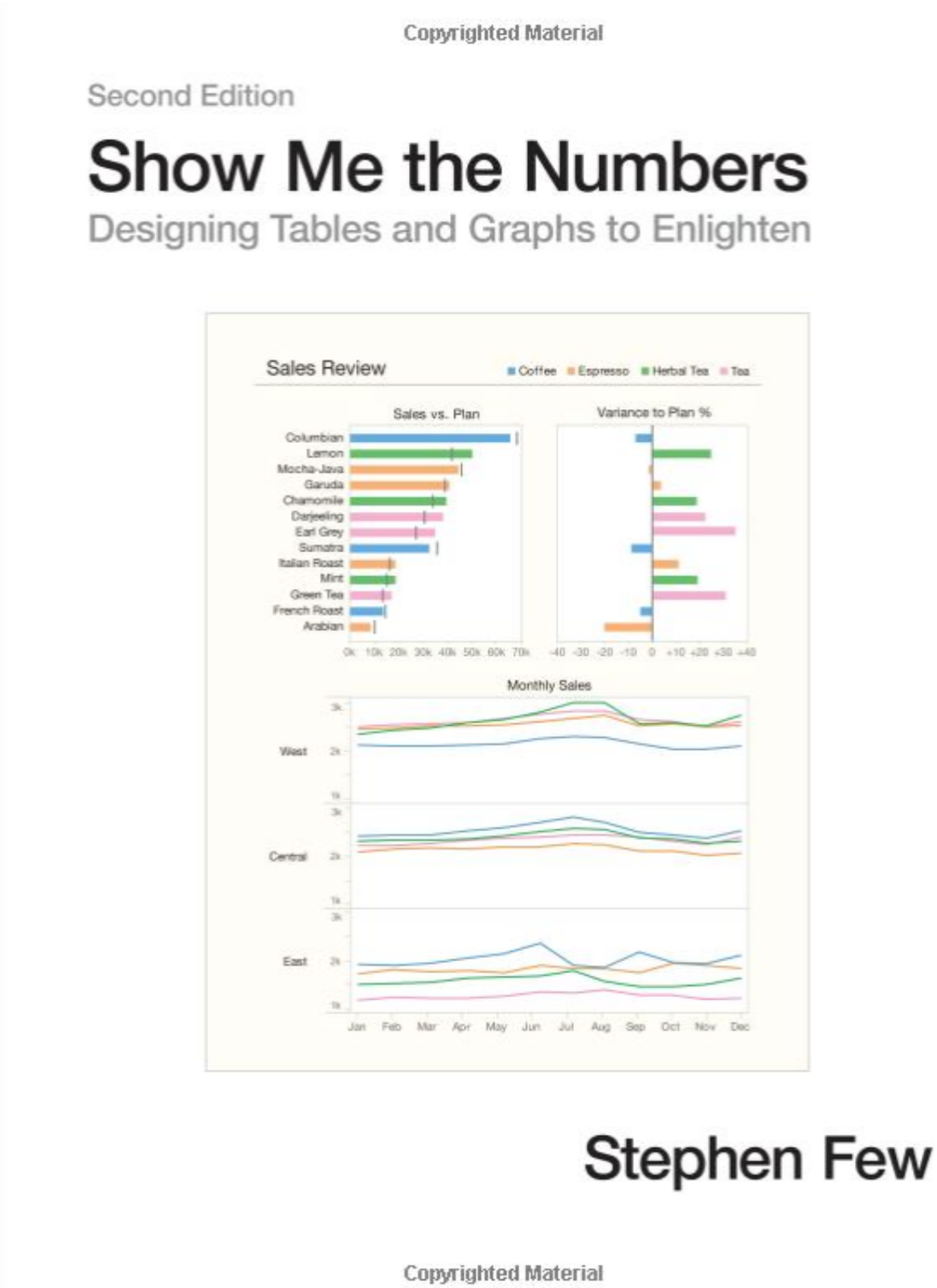
He is noted for his writings on information design and as a pioneer in the field of data visualization.

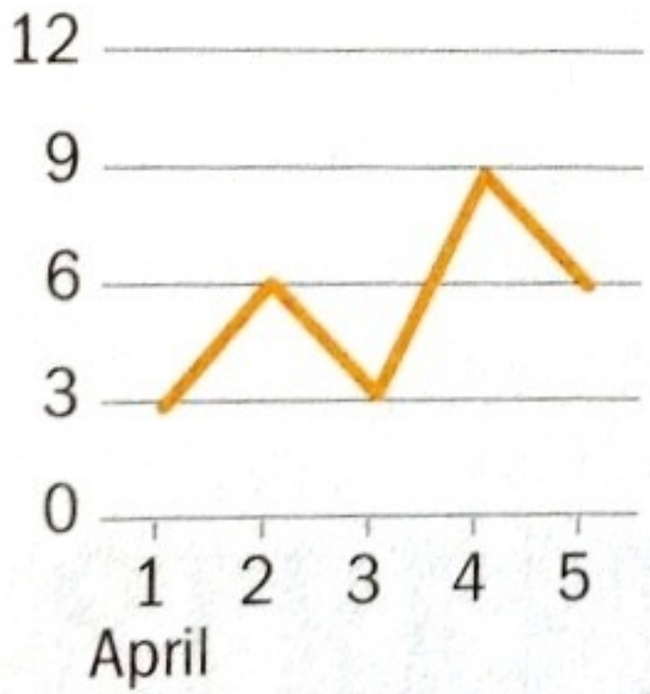
-Wikipedia

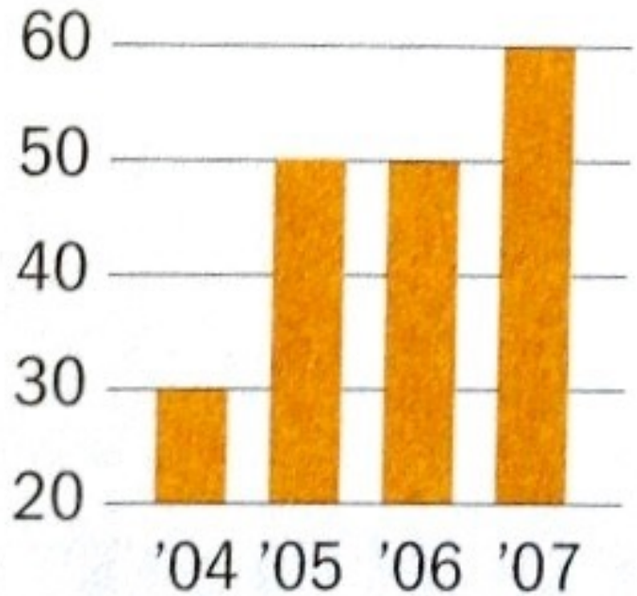


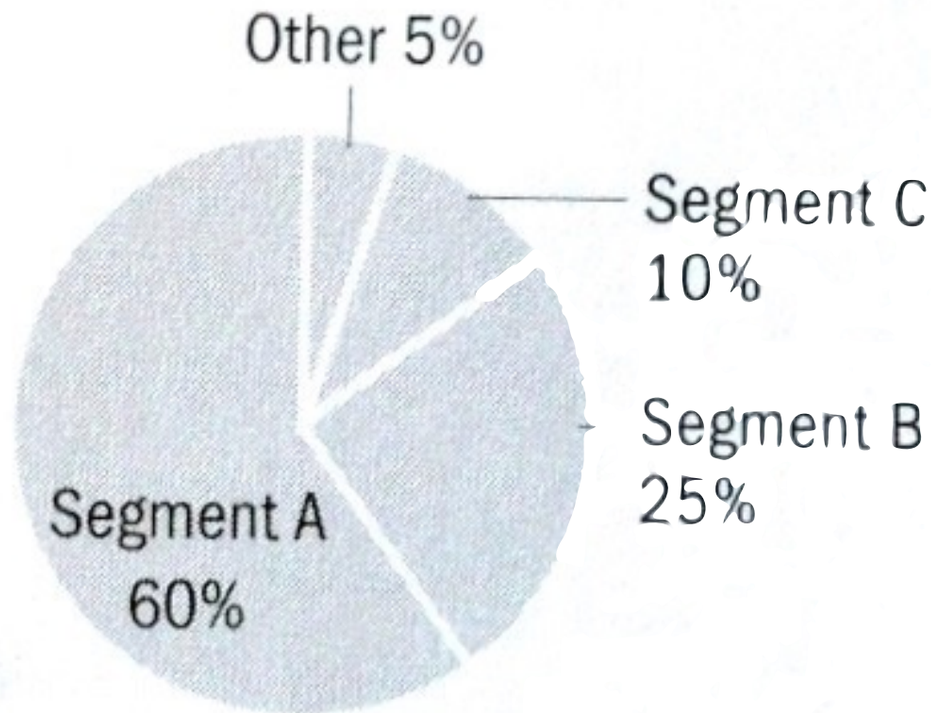
Also Highly Recommended:



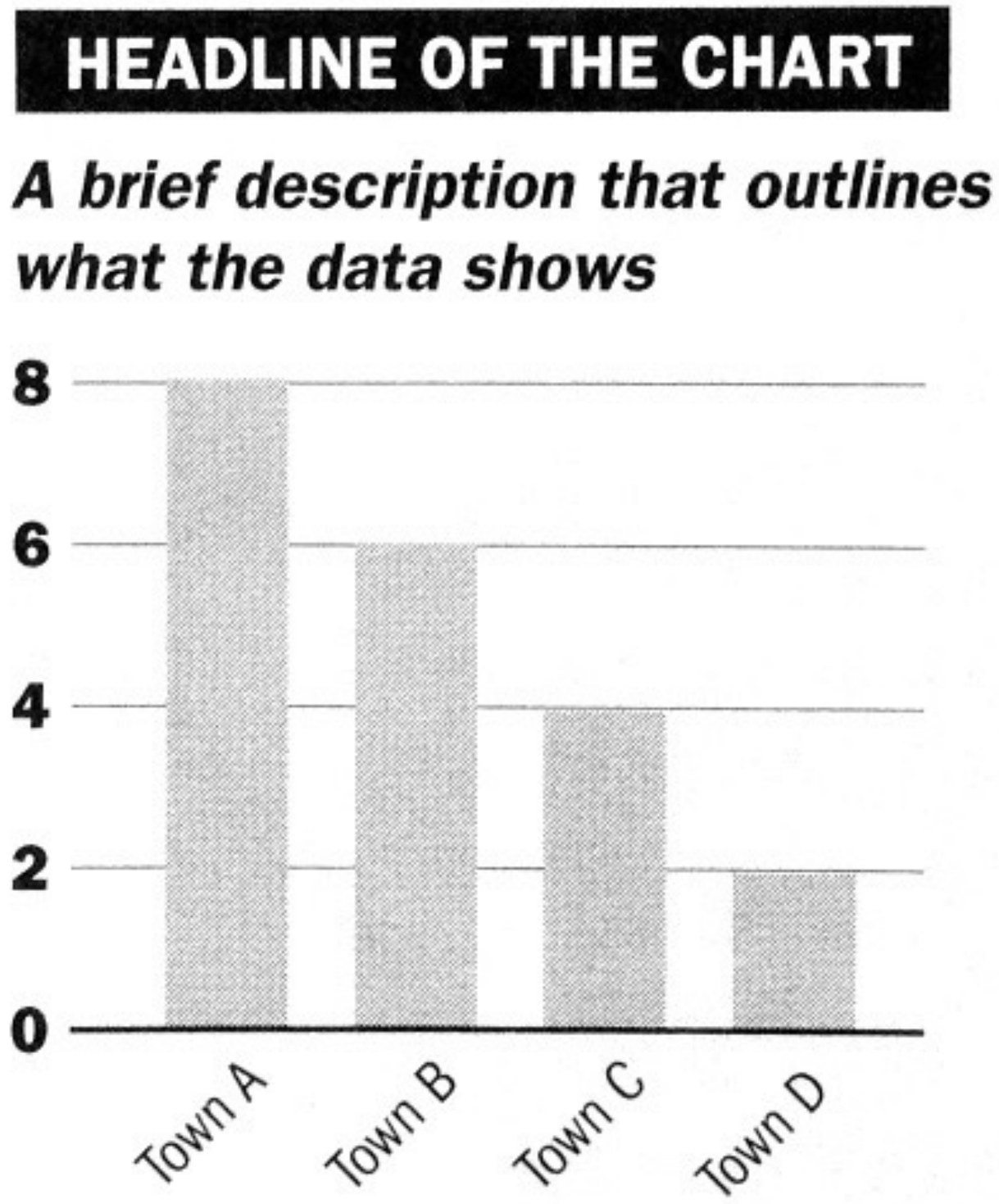




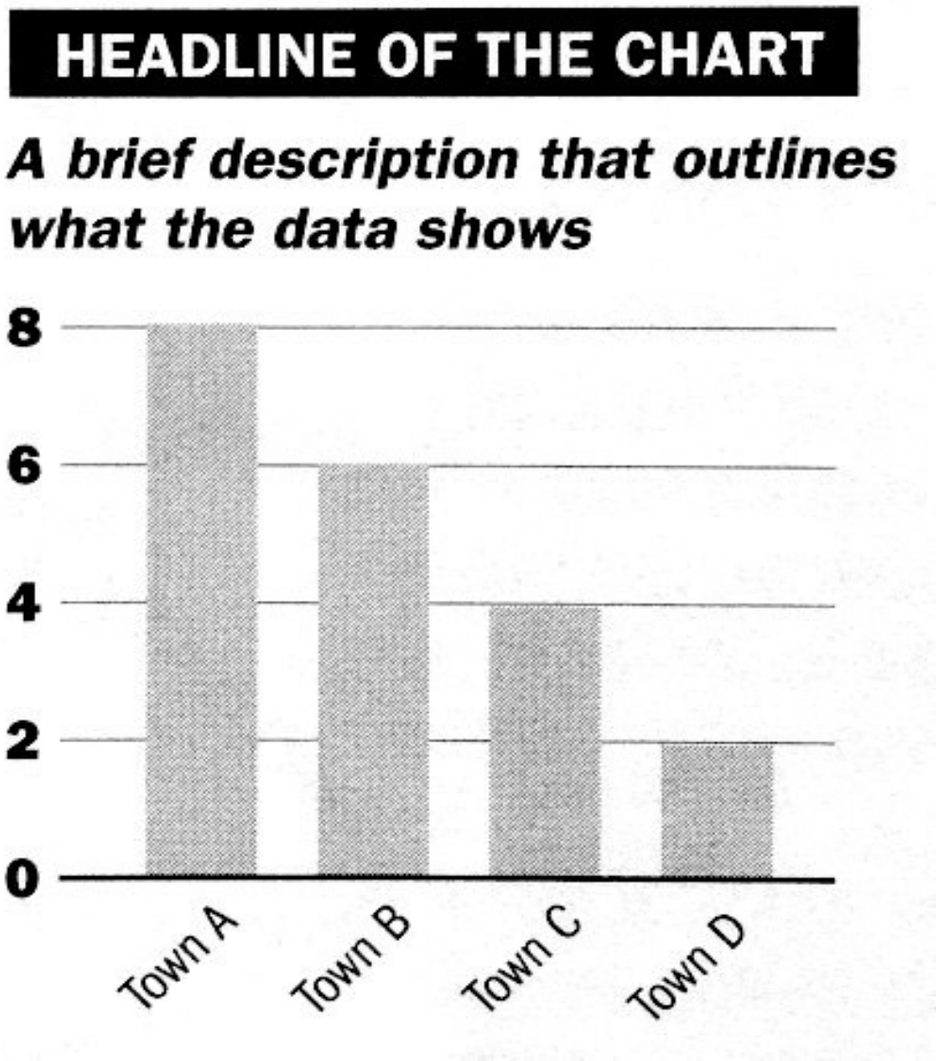


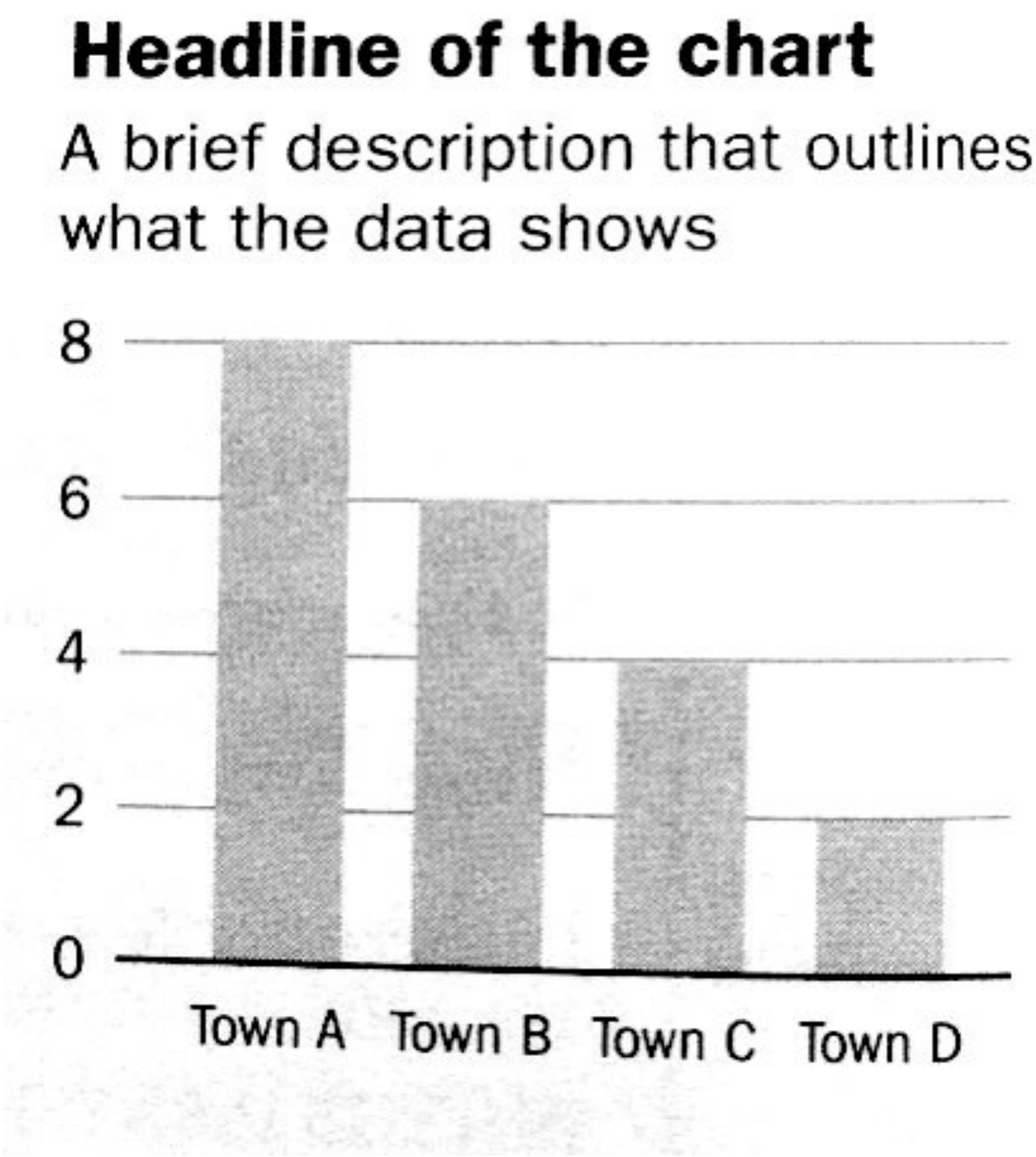


Good charts? How would you improve them?



How about this one?





Which is better?

Tables

What are they good for?

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

Can you improve this table's design?

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

https://www.youtube.com/watch?v=1E9pKU_N15A
http://www.youtube.com/watch?v=A8I9pYCl9AQ

“When everyone is super, no one is super”

https://www.youtube.com/watch?v=1E9pKU_N15A
http://www.youtube.com/watch?v=A8I9pYCl9AQ

“When everyone is super, no one is super”

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

Better? High “data to ink” ratio

Aligning Numbers

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

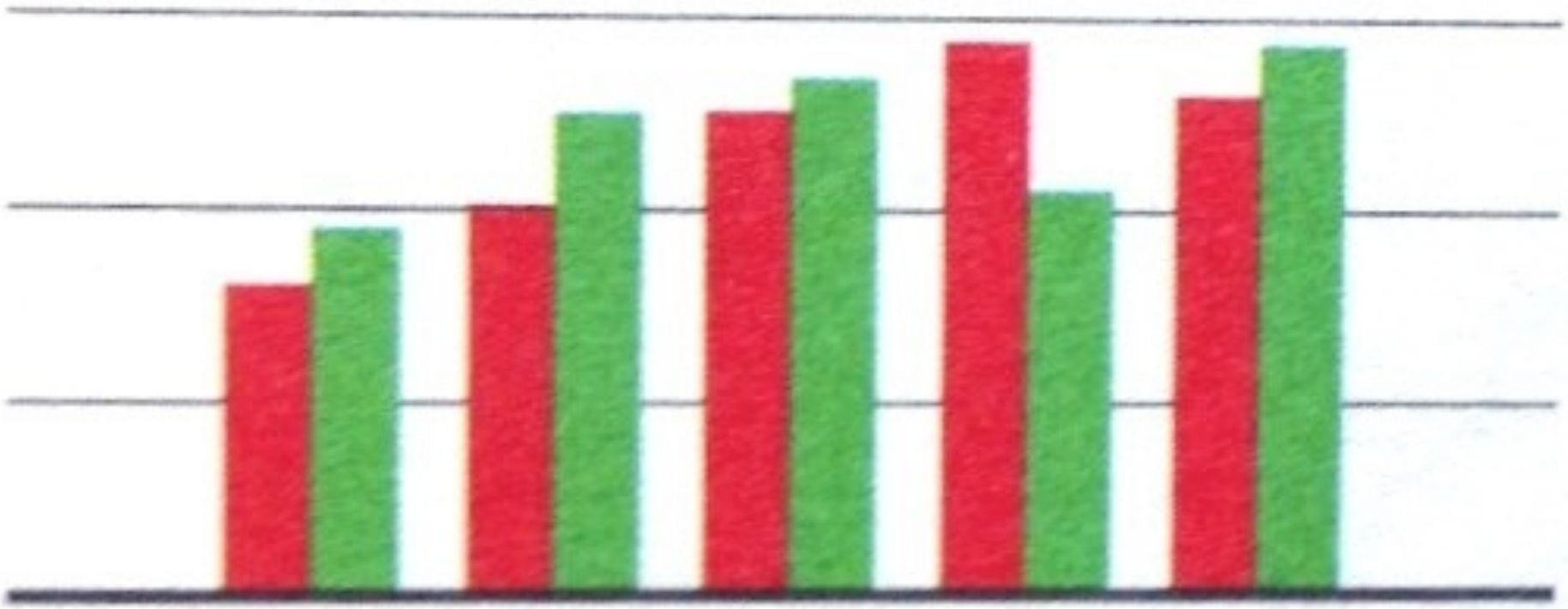
Name	Data
Company A	10.82
Company B	9.49
Company C	8
Company D	7.4

Look good? Or not?

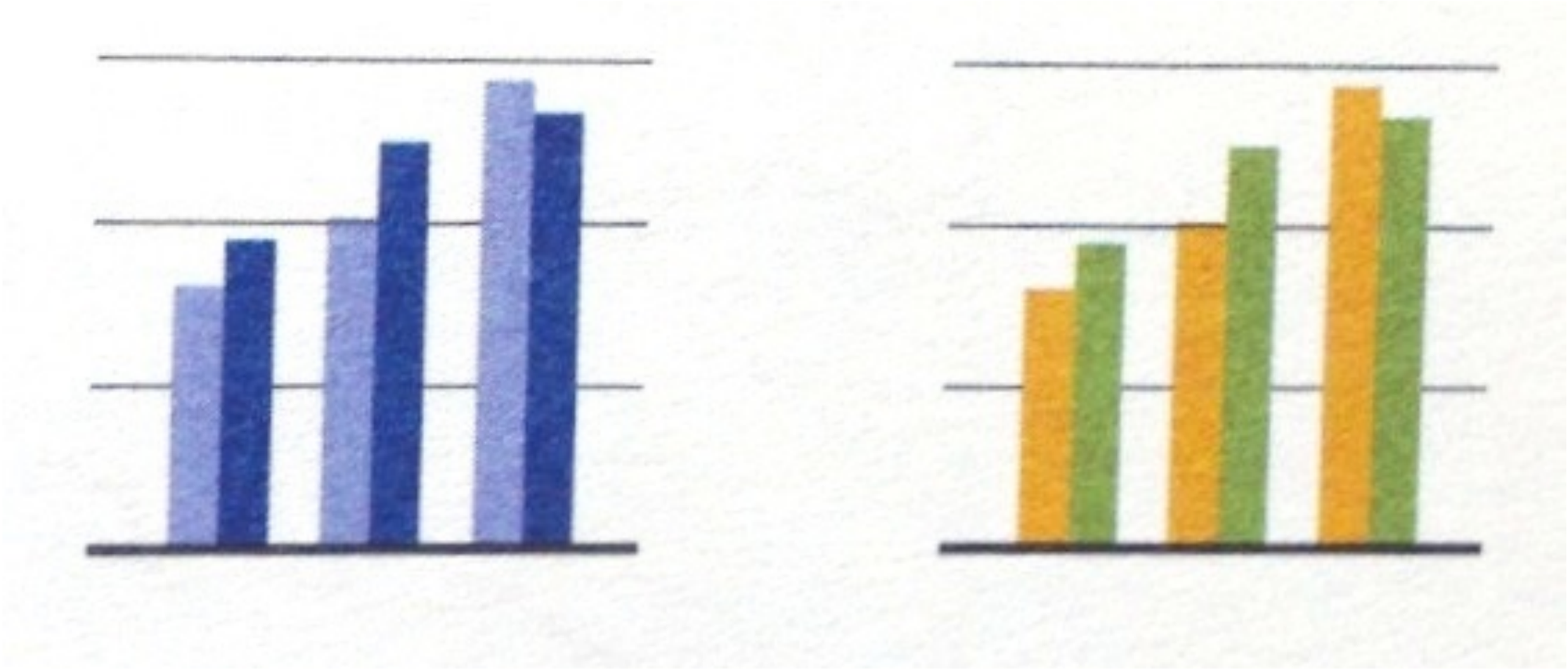
Name	Data
Company A	10.82
Company B	9.49
Company C	8
Company D	7.4

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

Bar Charts

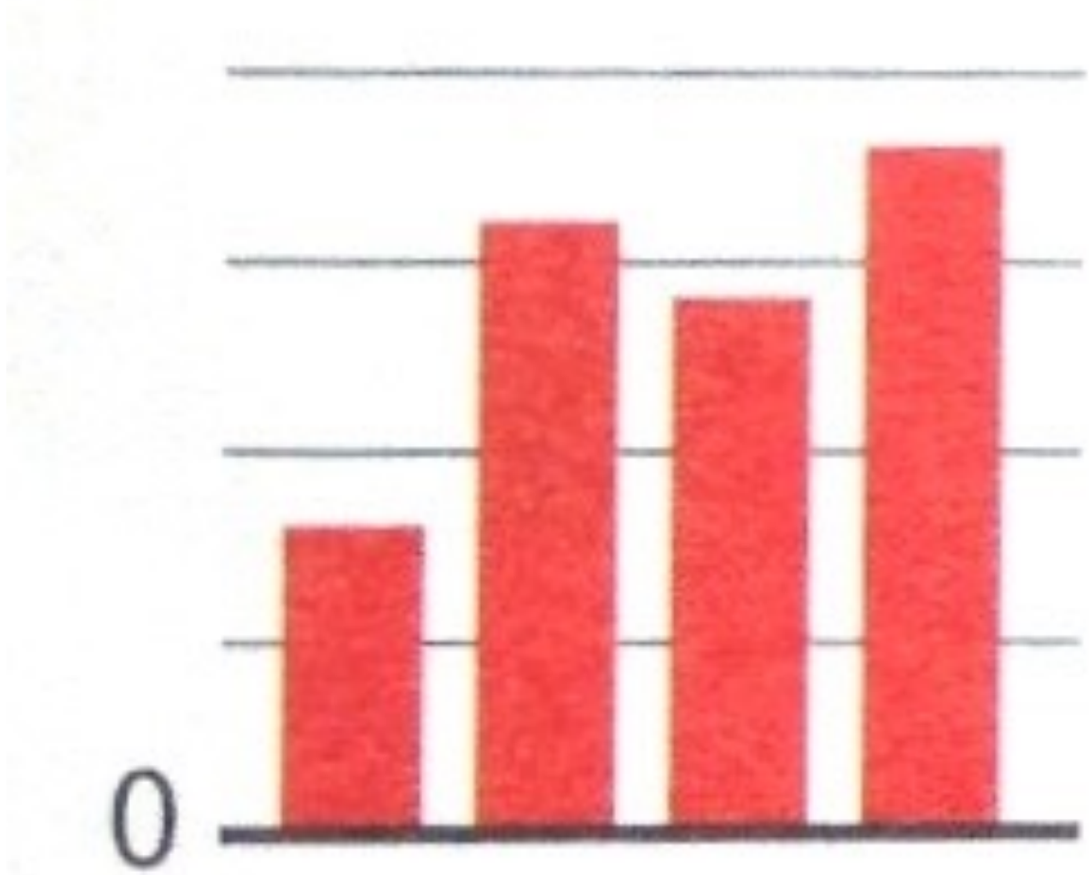


This reminds you of what?

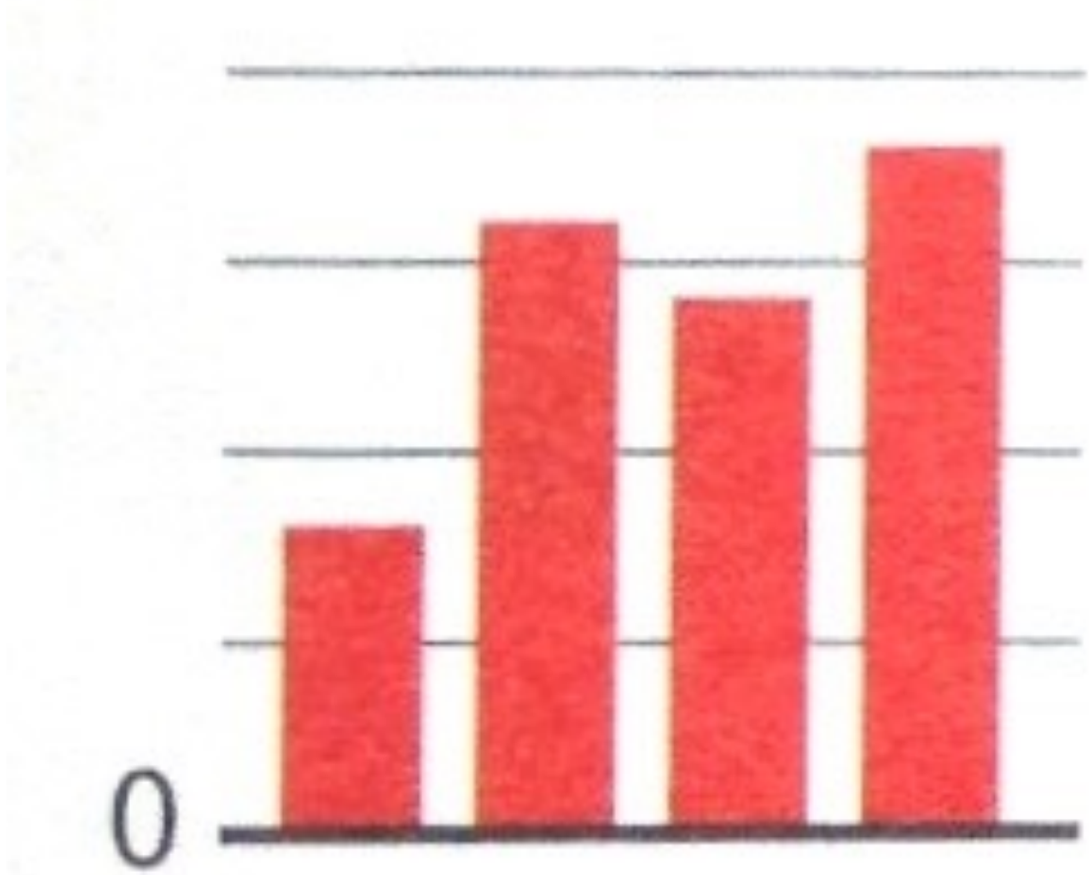


Better than Christmas.

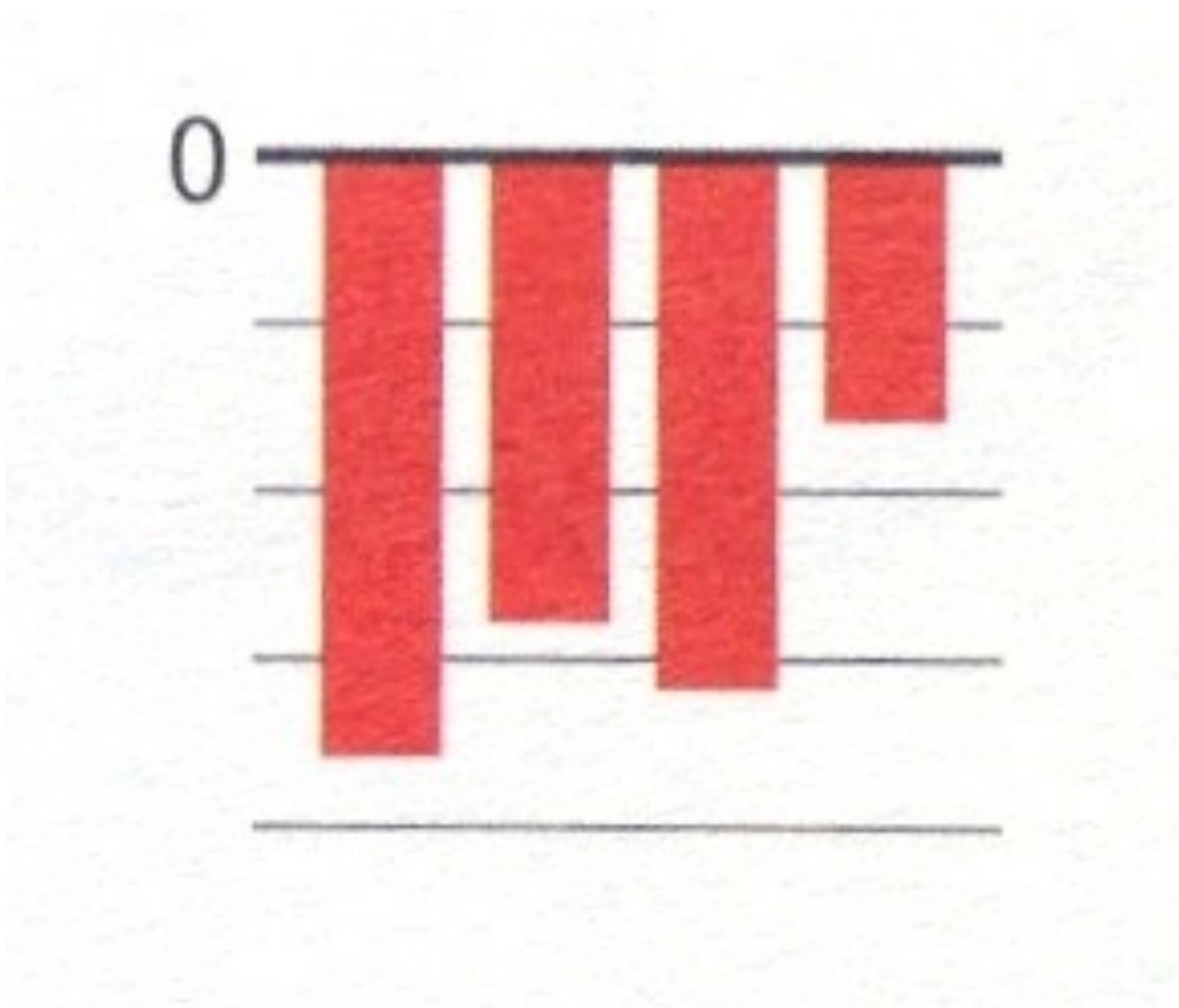
Company Profits

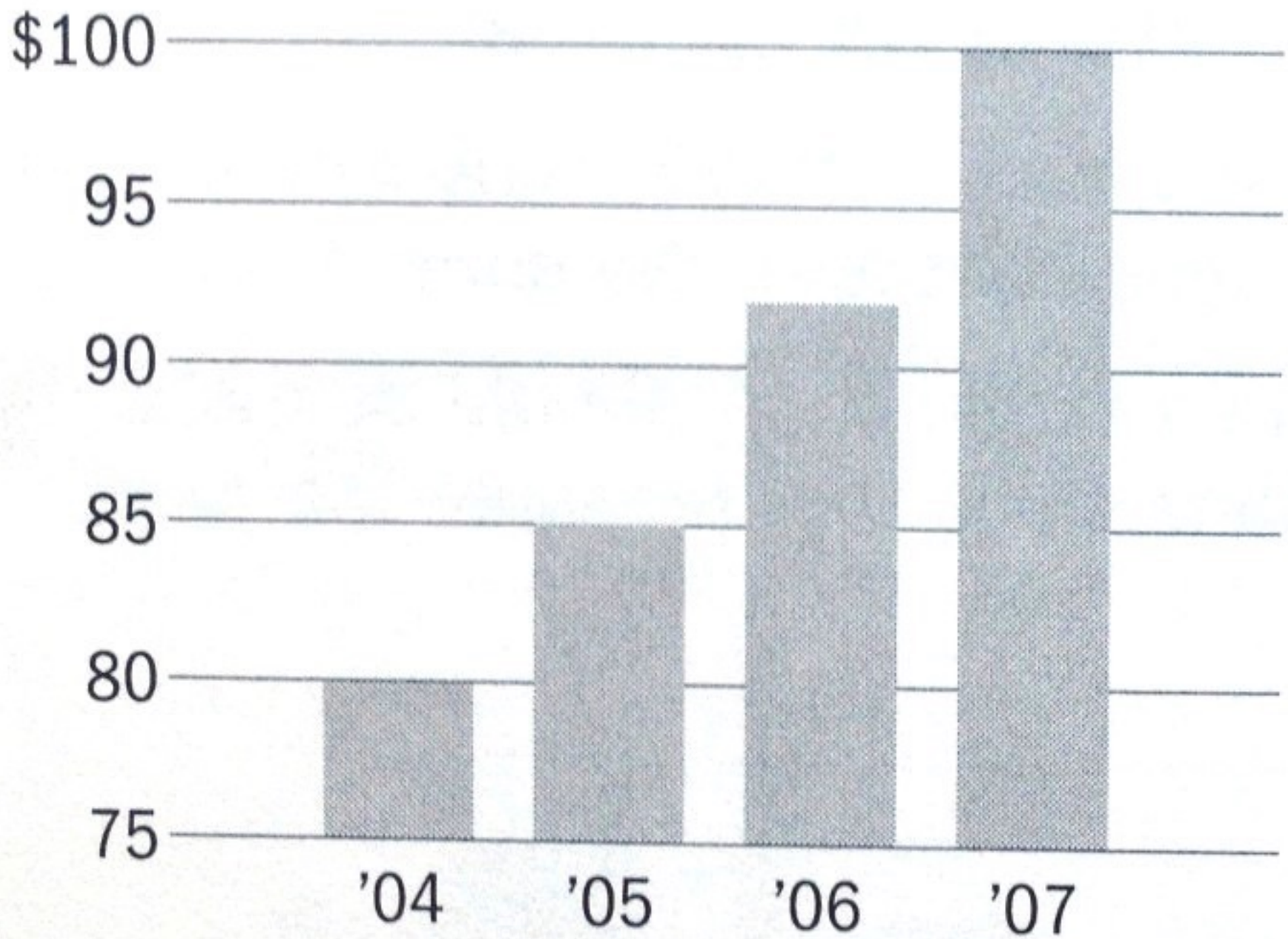


Company Profits

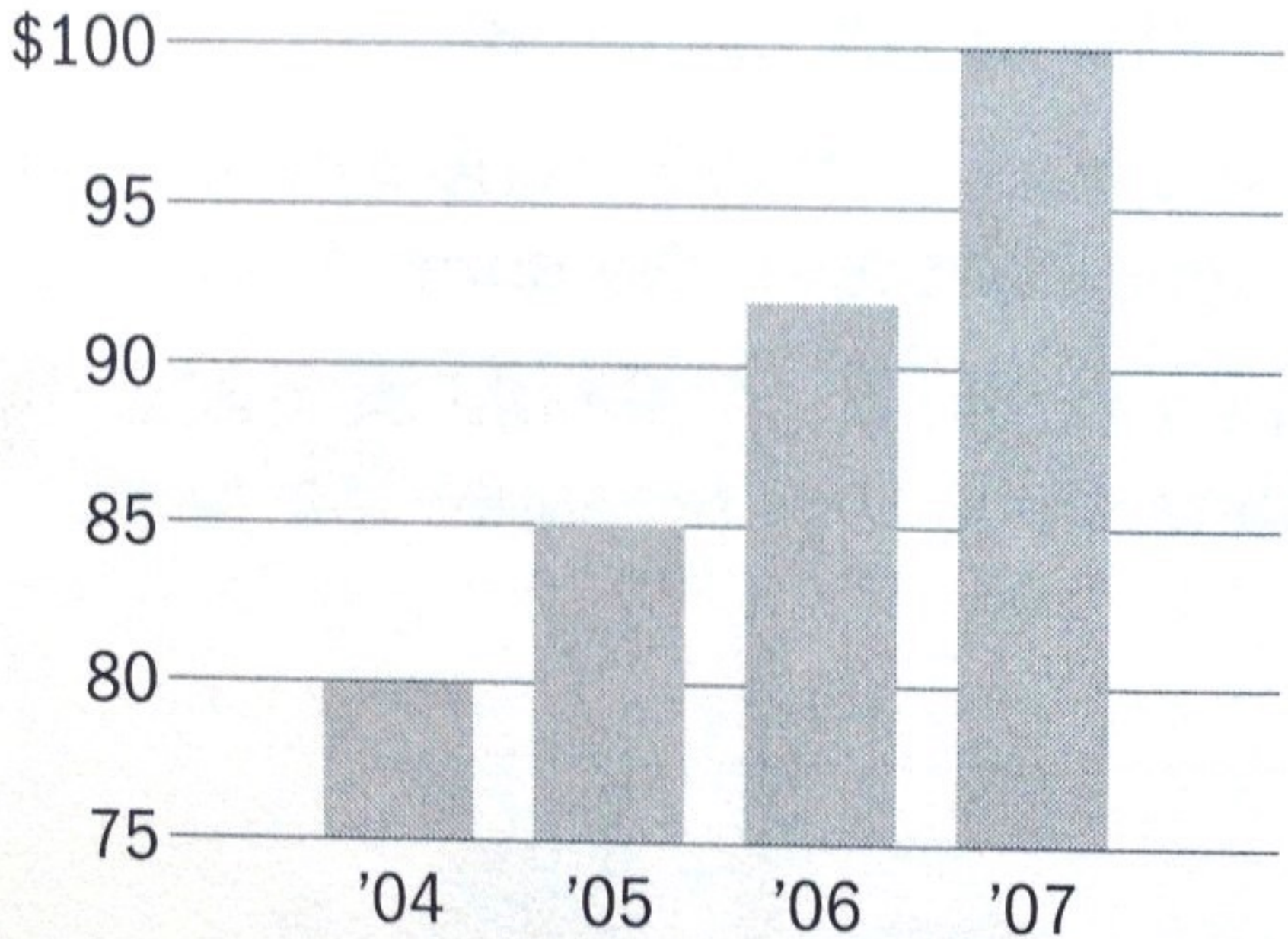


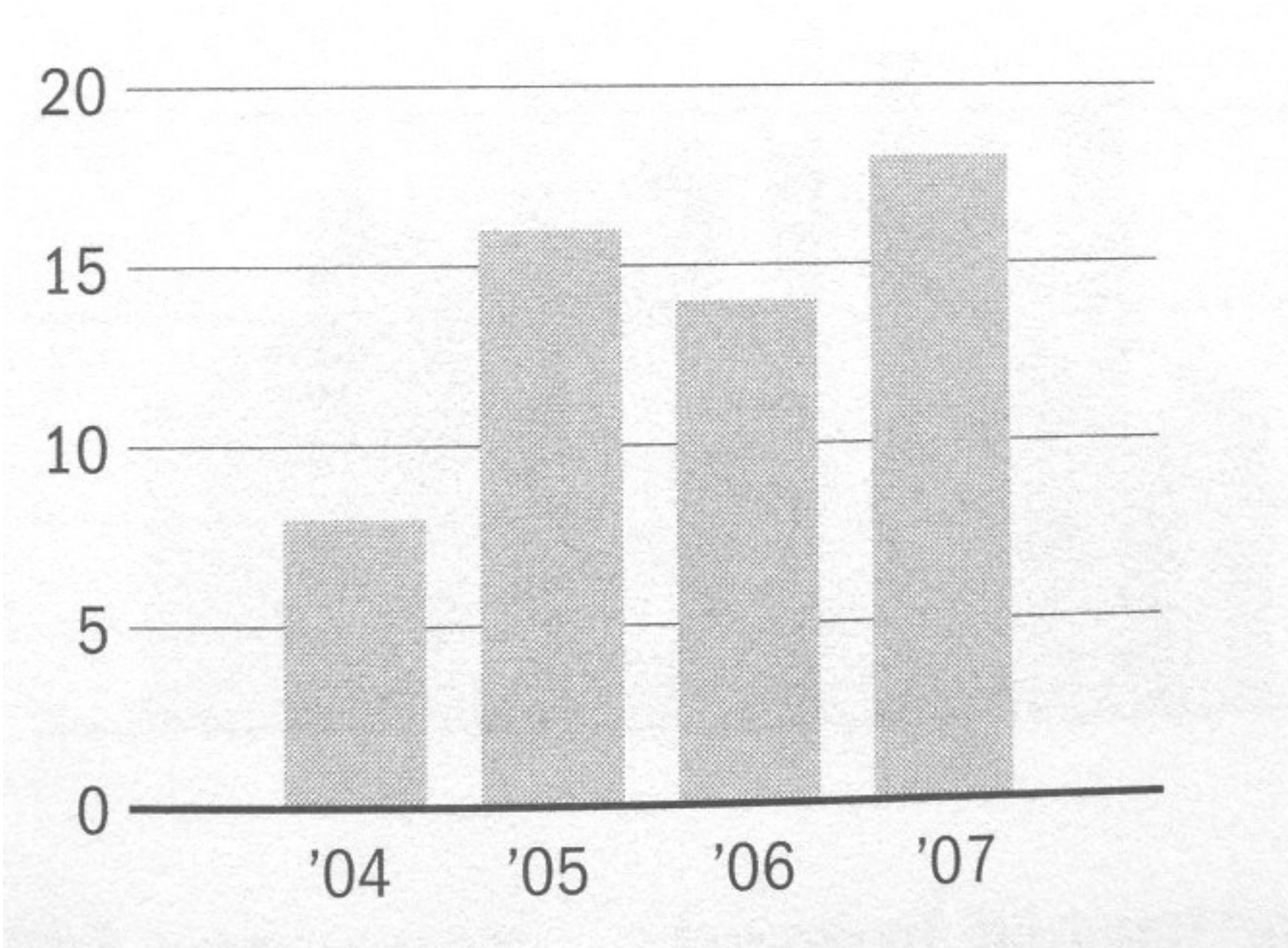
Don’t show profits in red!!
Think carefully about your color choices.


0


$100
95
90
85
80
75
'04
'05
'06
'07

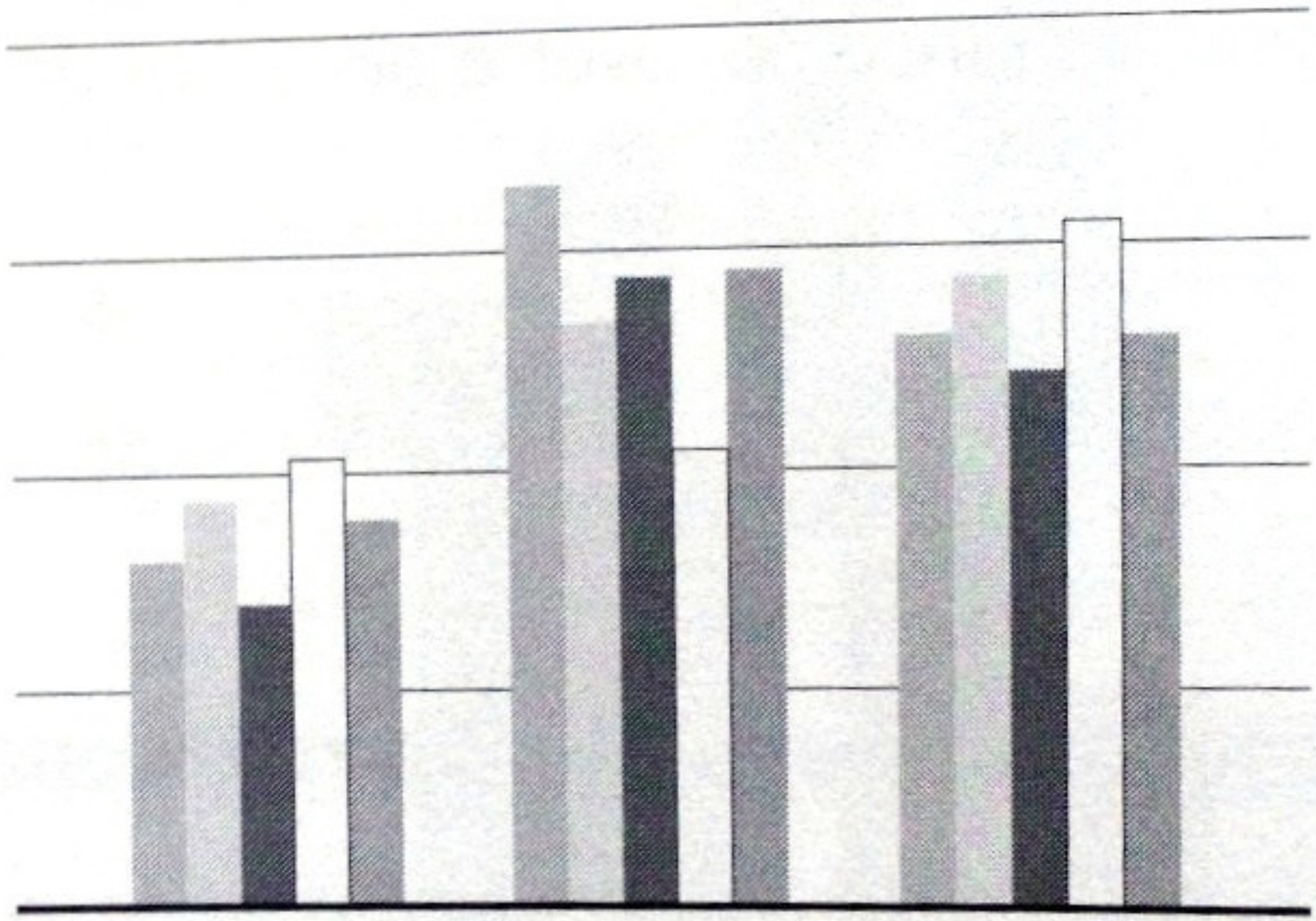
Misleading Bar Charts

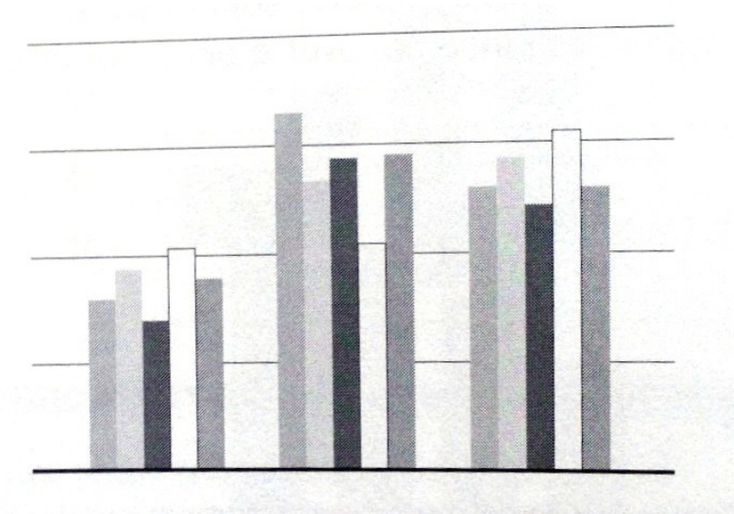




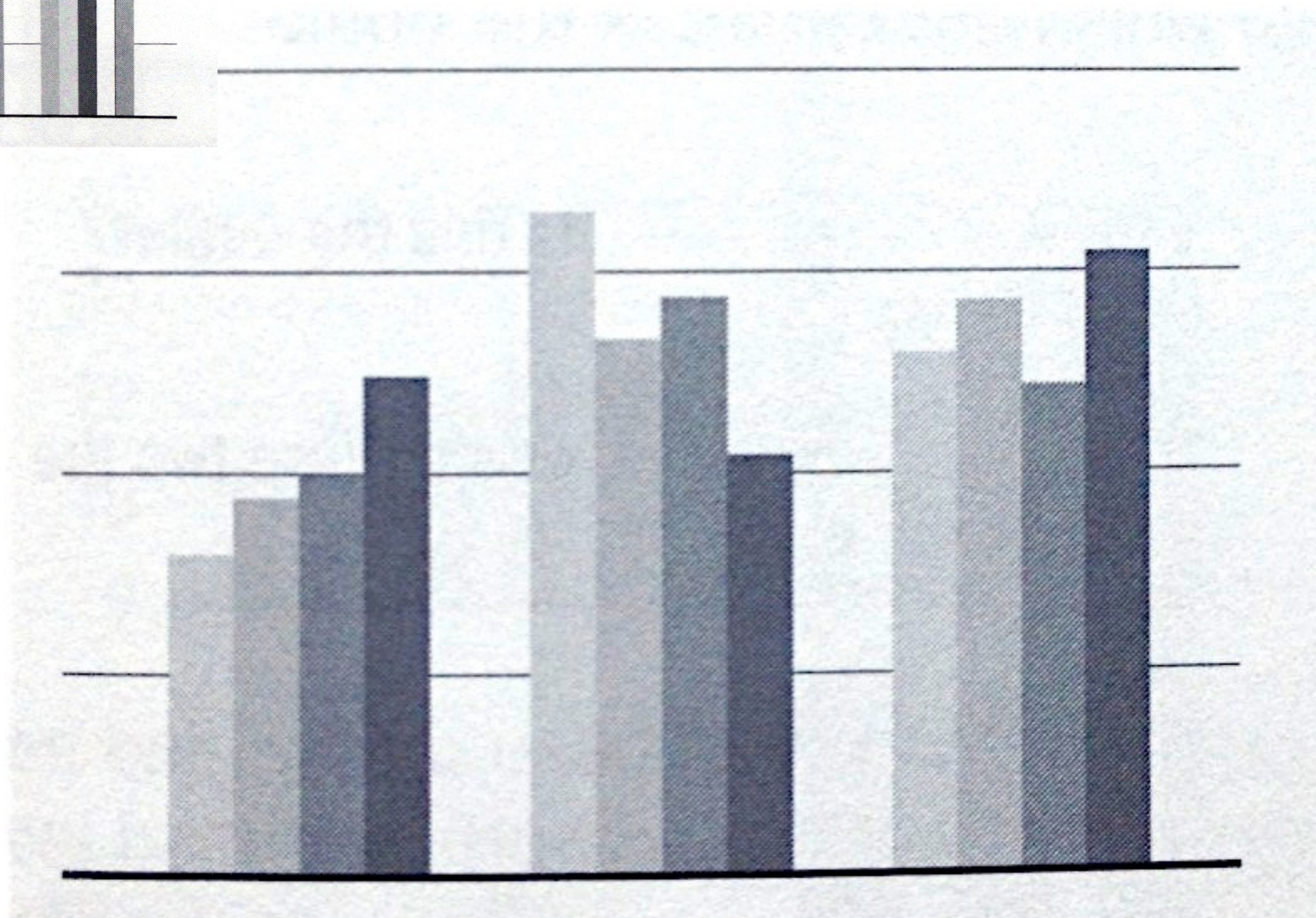
Vertical axis of bar charts should start at 0, almost always

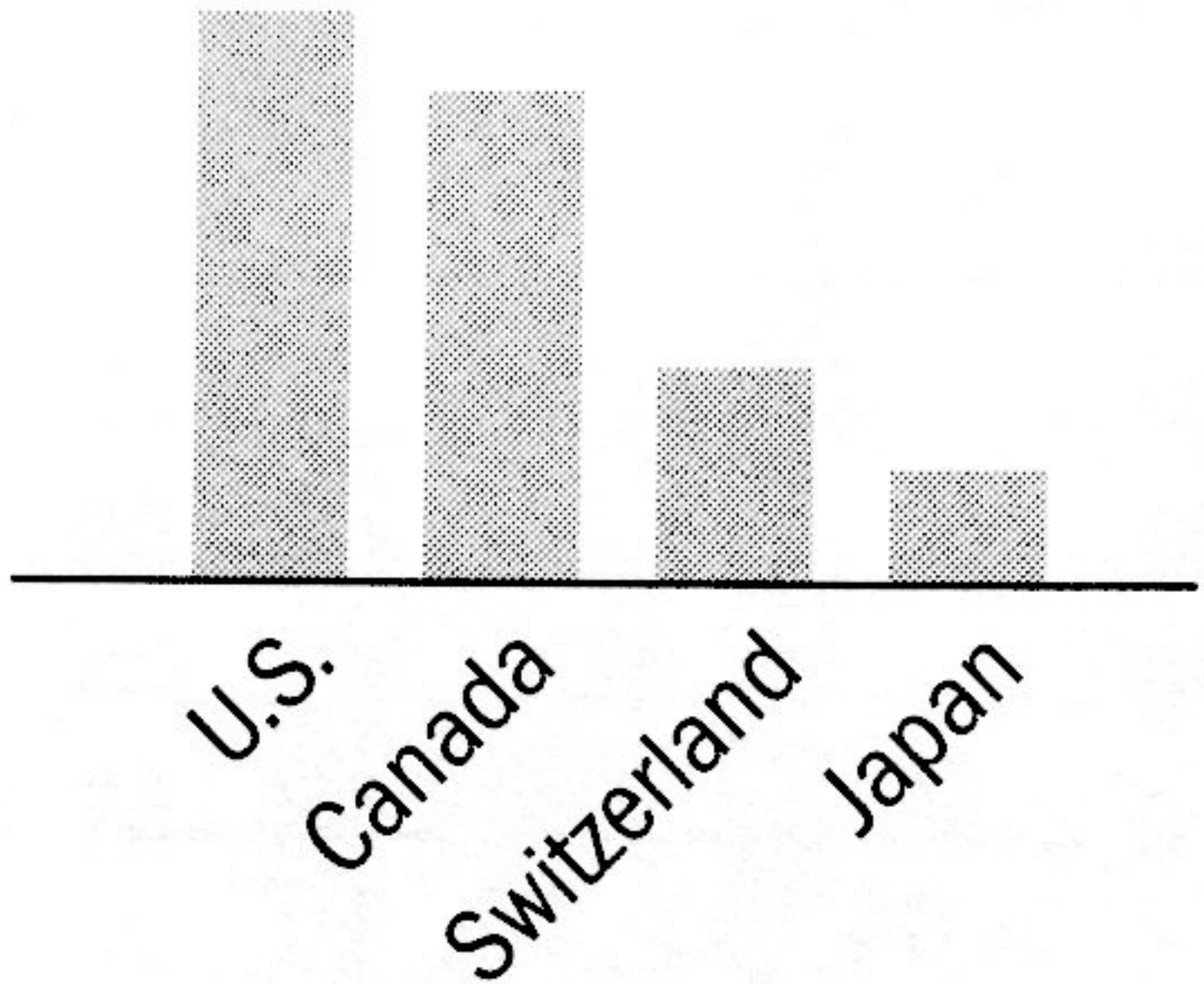
Disorienting color bars



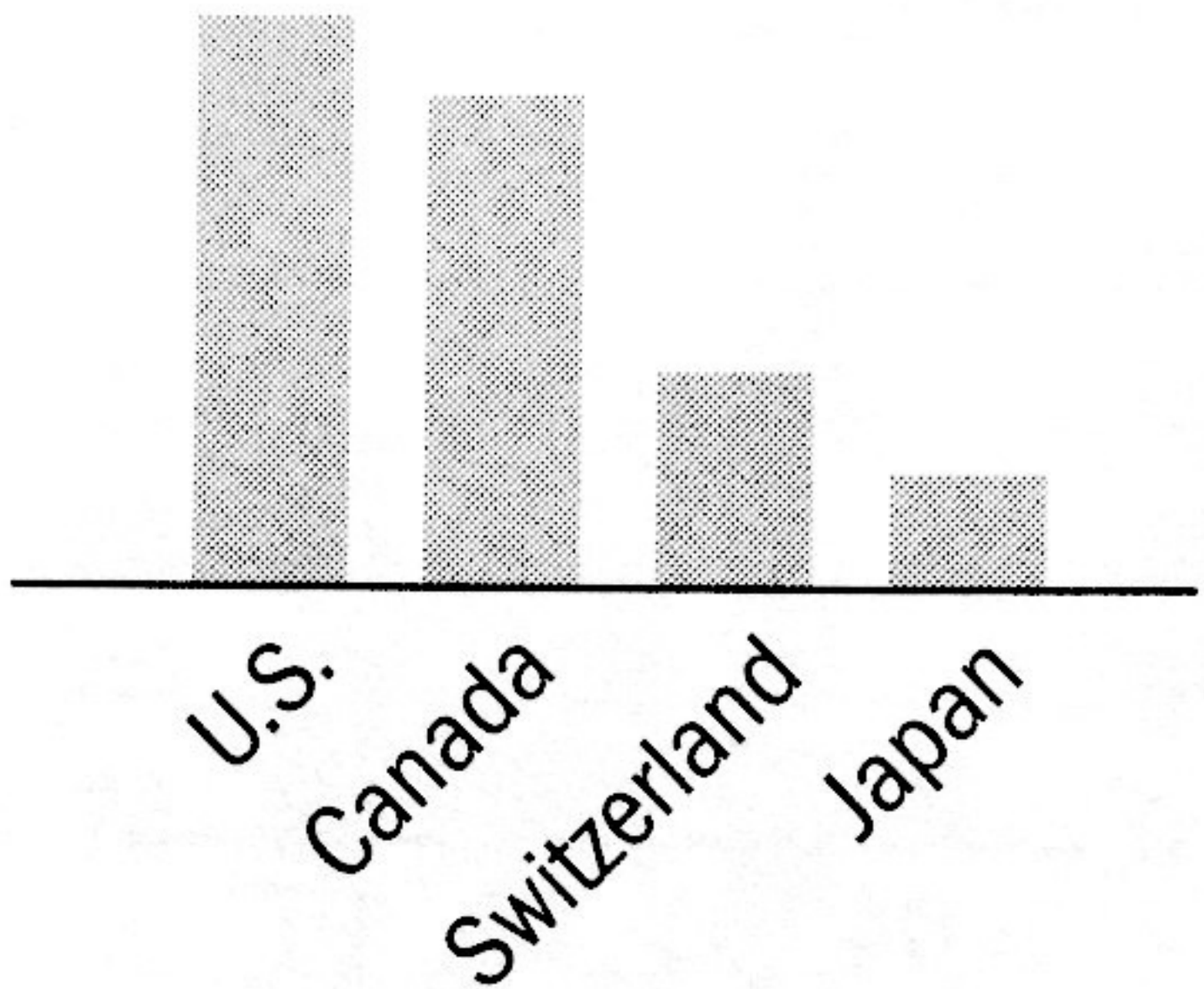


Better?




U.S.
Canada
Switzerland
Japan

Exercise For Your Necks



Bars Can be Horizontal

http://www.apple.com/macbook-pro/performance-retina/

Bars Can be Horizontal

When labels are hard to read, try horizontal layout.

http://www.apple.com/macbook-pro/performance-retina/

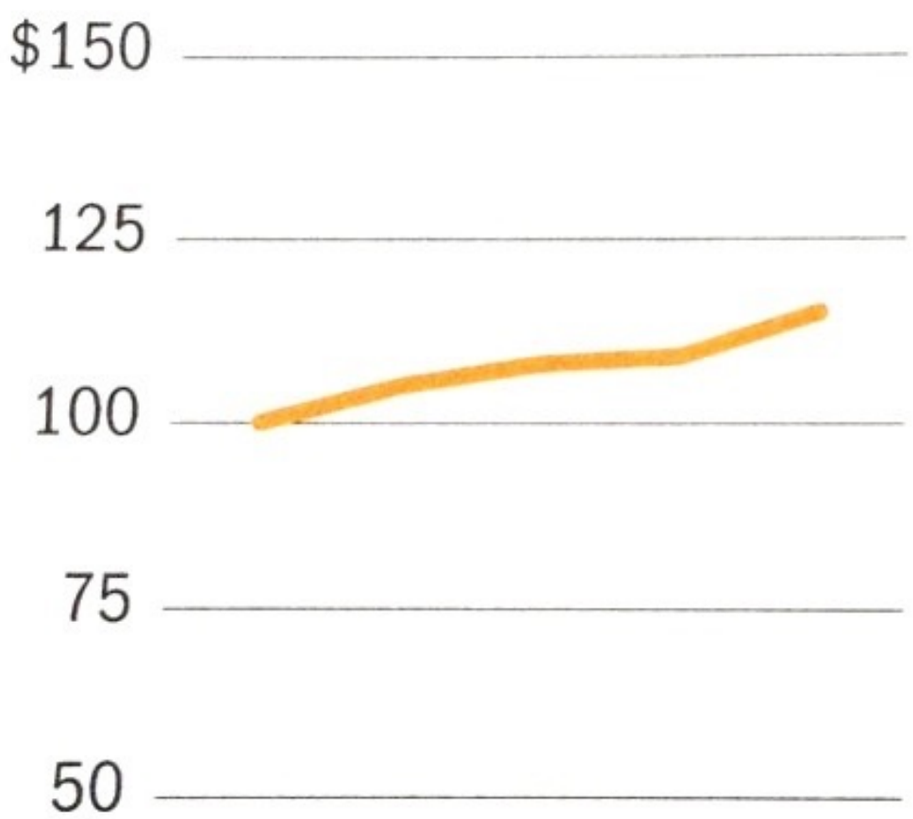
Line Charts

Does this look all right to you?

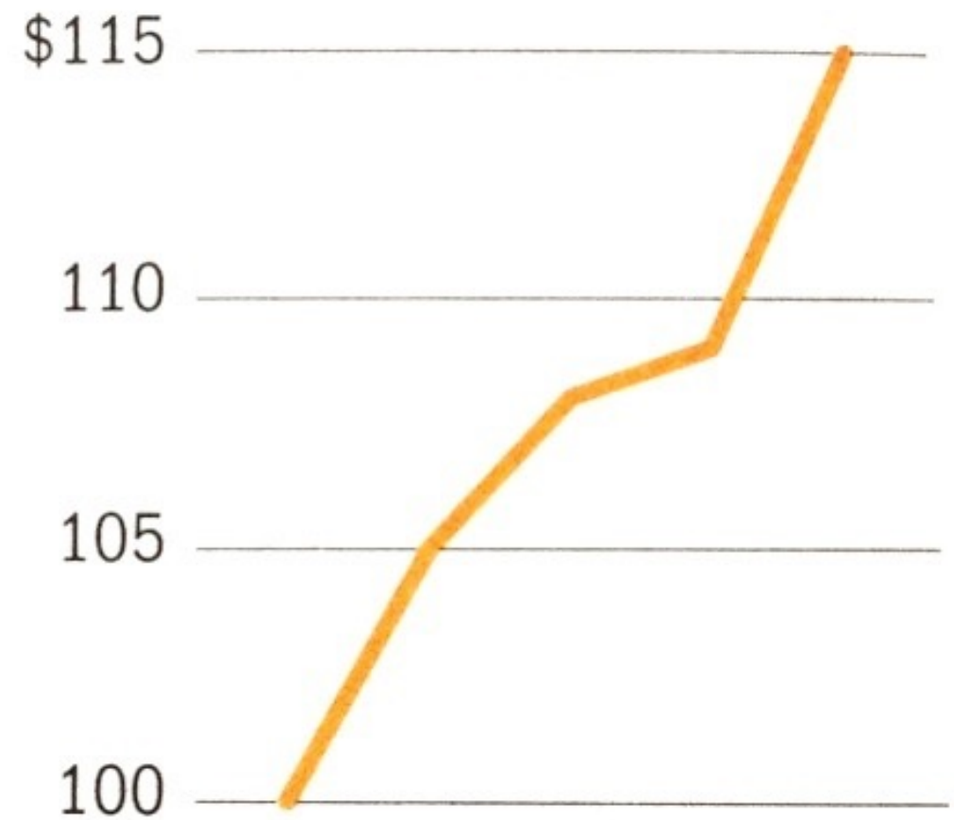
Use “ticks” at regular intervals (e.g., 2, 5, 10, etc.)

Fever Line

Too flat obscures the message

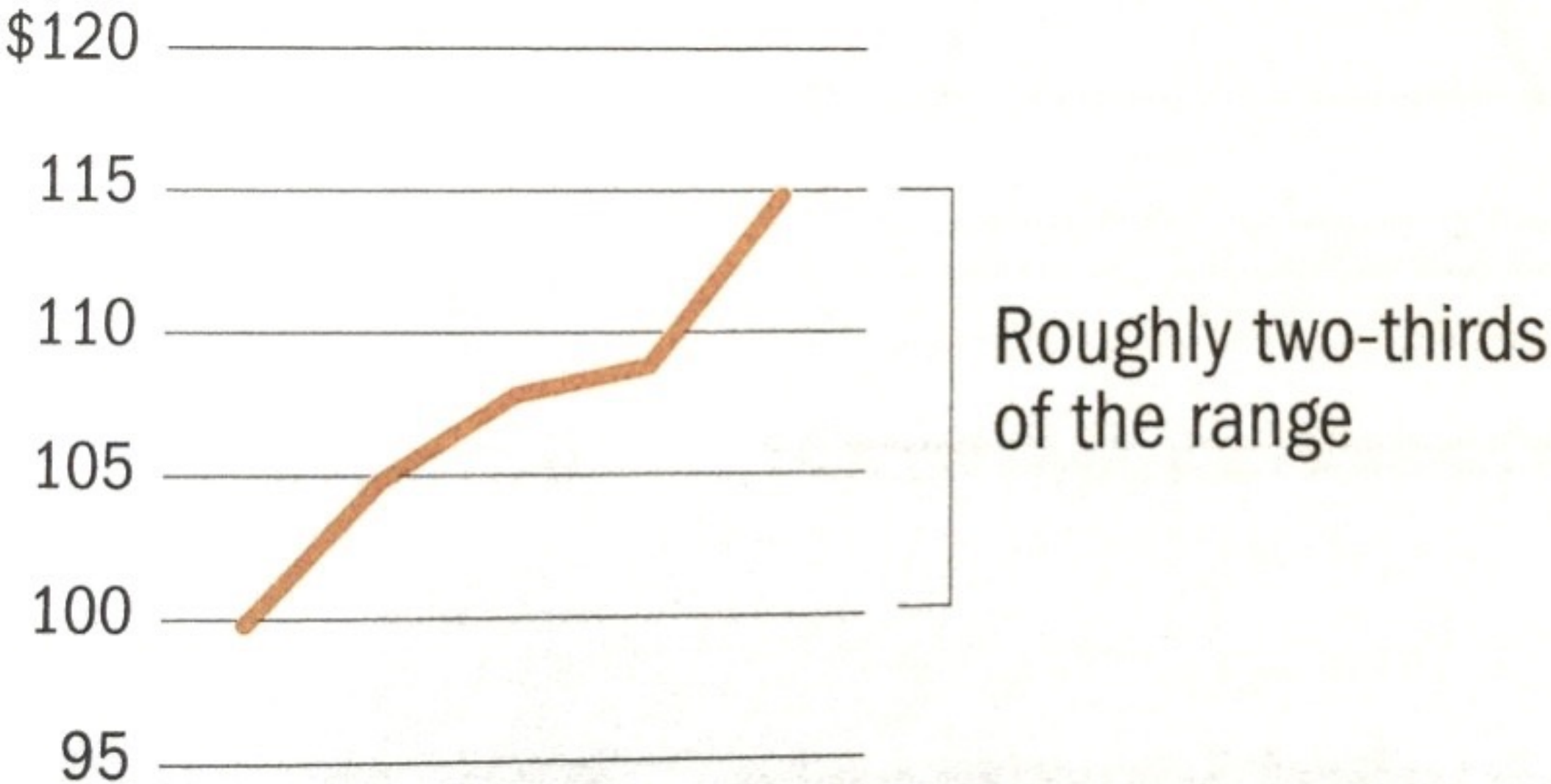


Too exaggerated overstates the trend

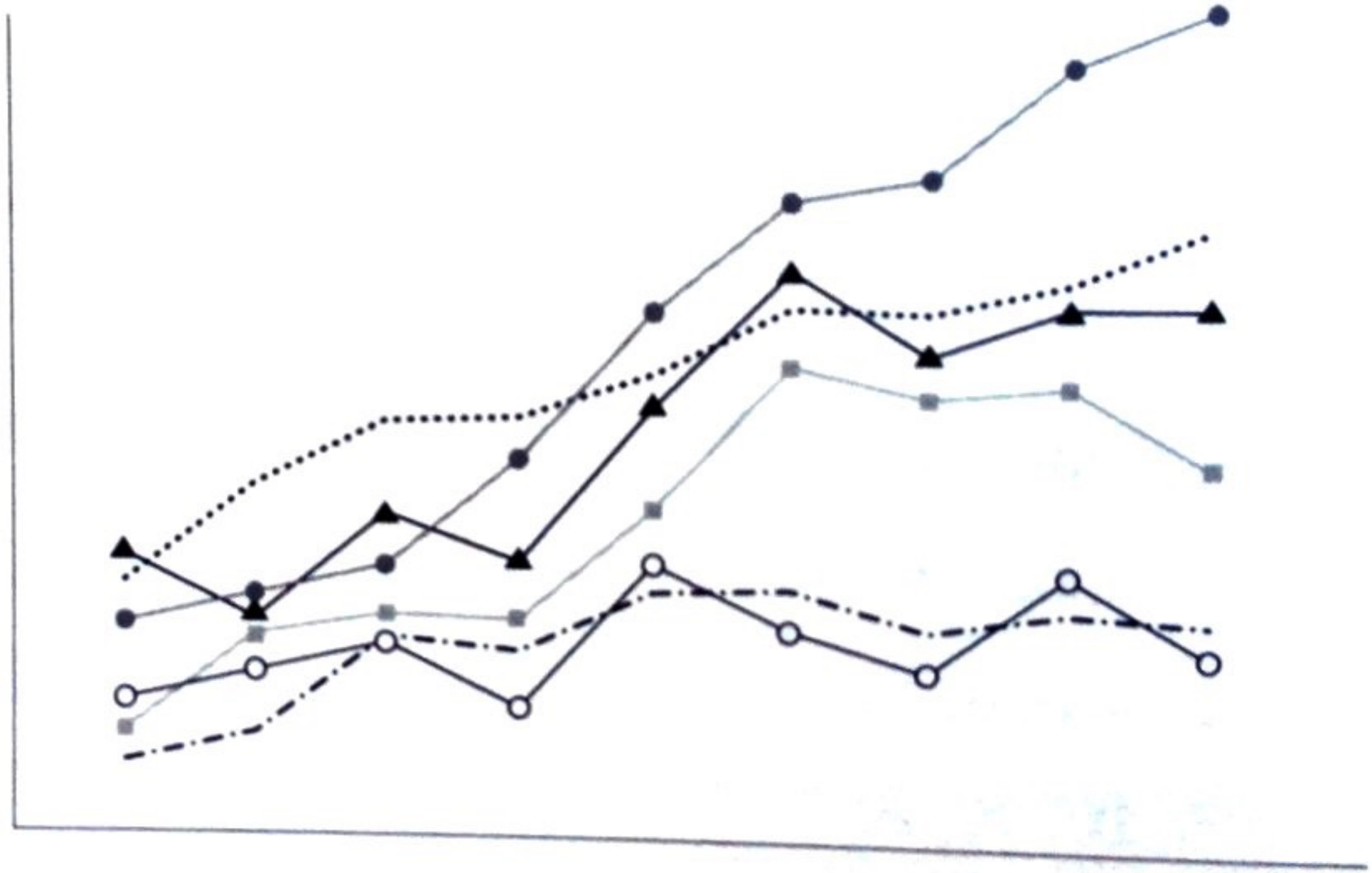


Note y-axis does not need to start at 0.
Why not as bad as in the case of bar chart?

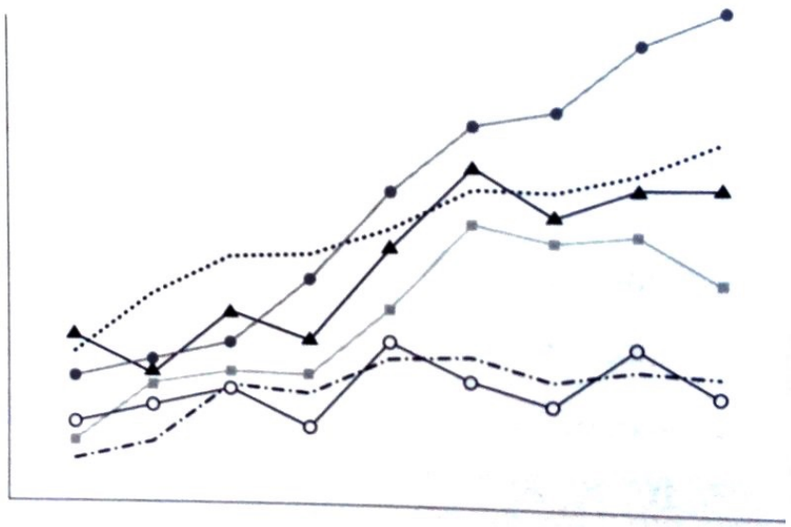
Fever Line

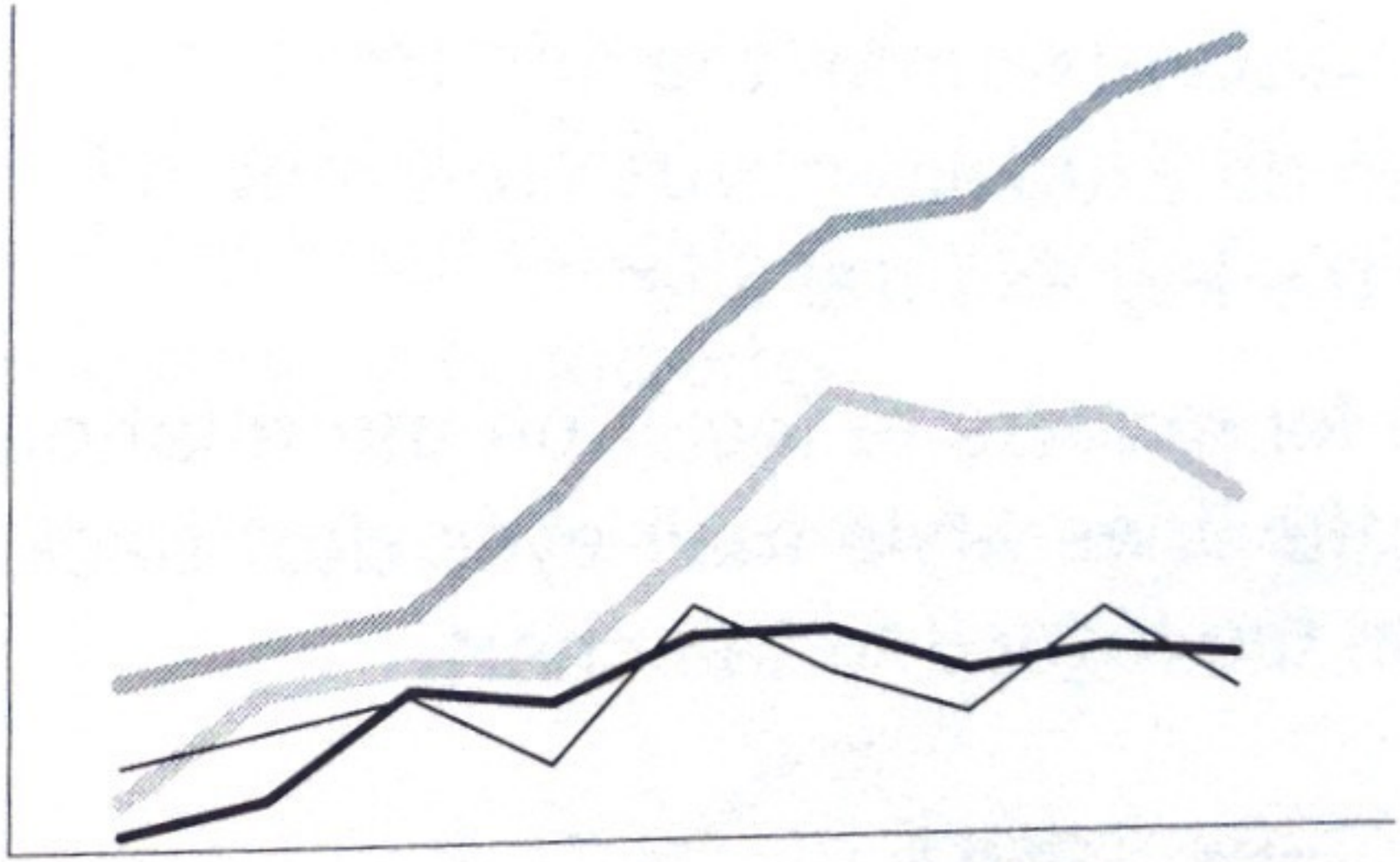


Multiple Lines in one chart

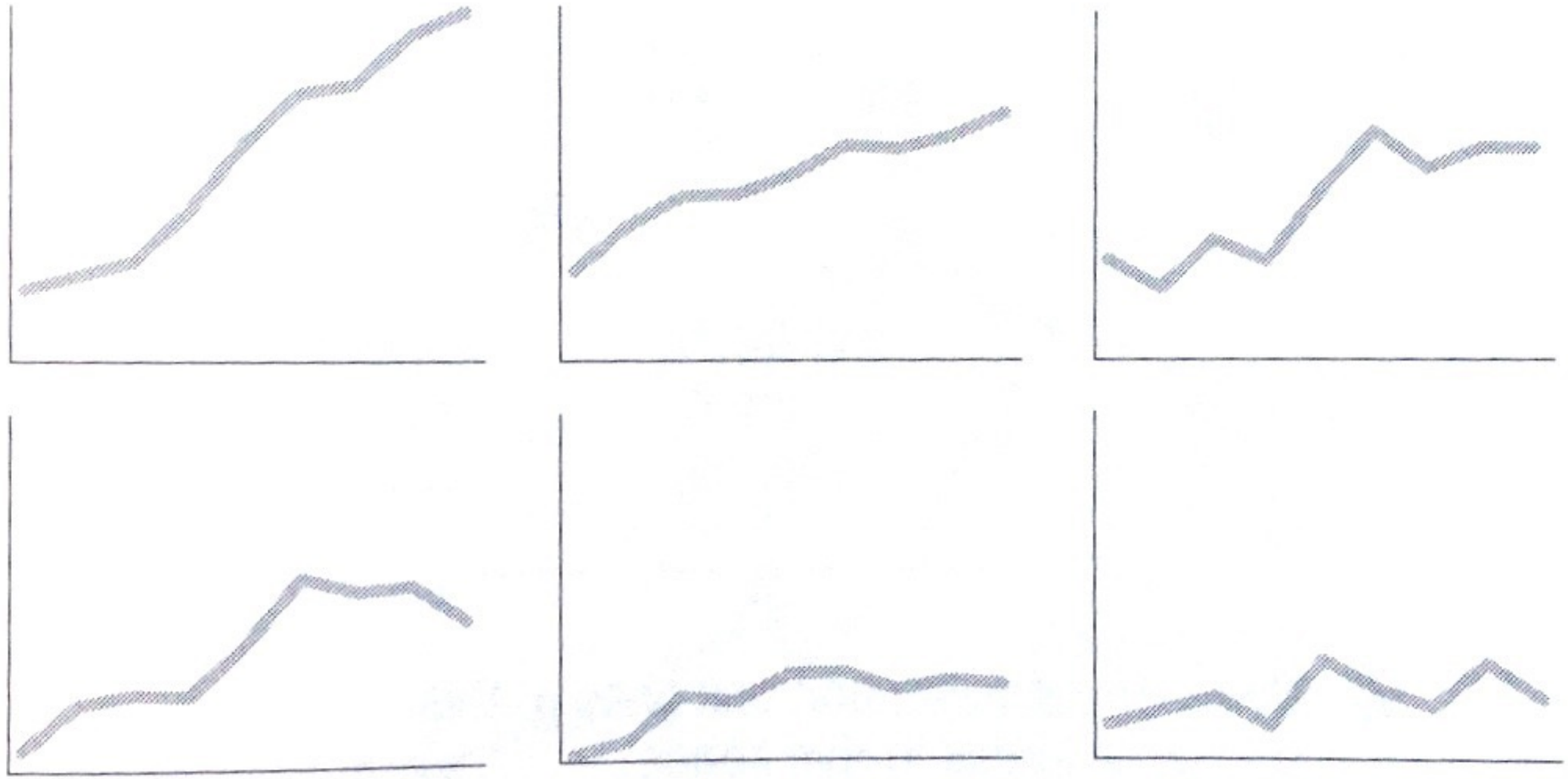


We see this often in academic papers. Better ways?





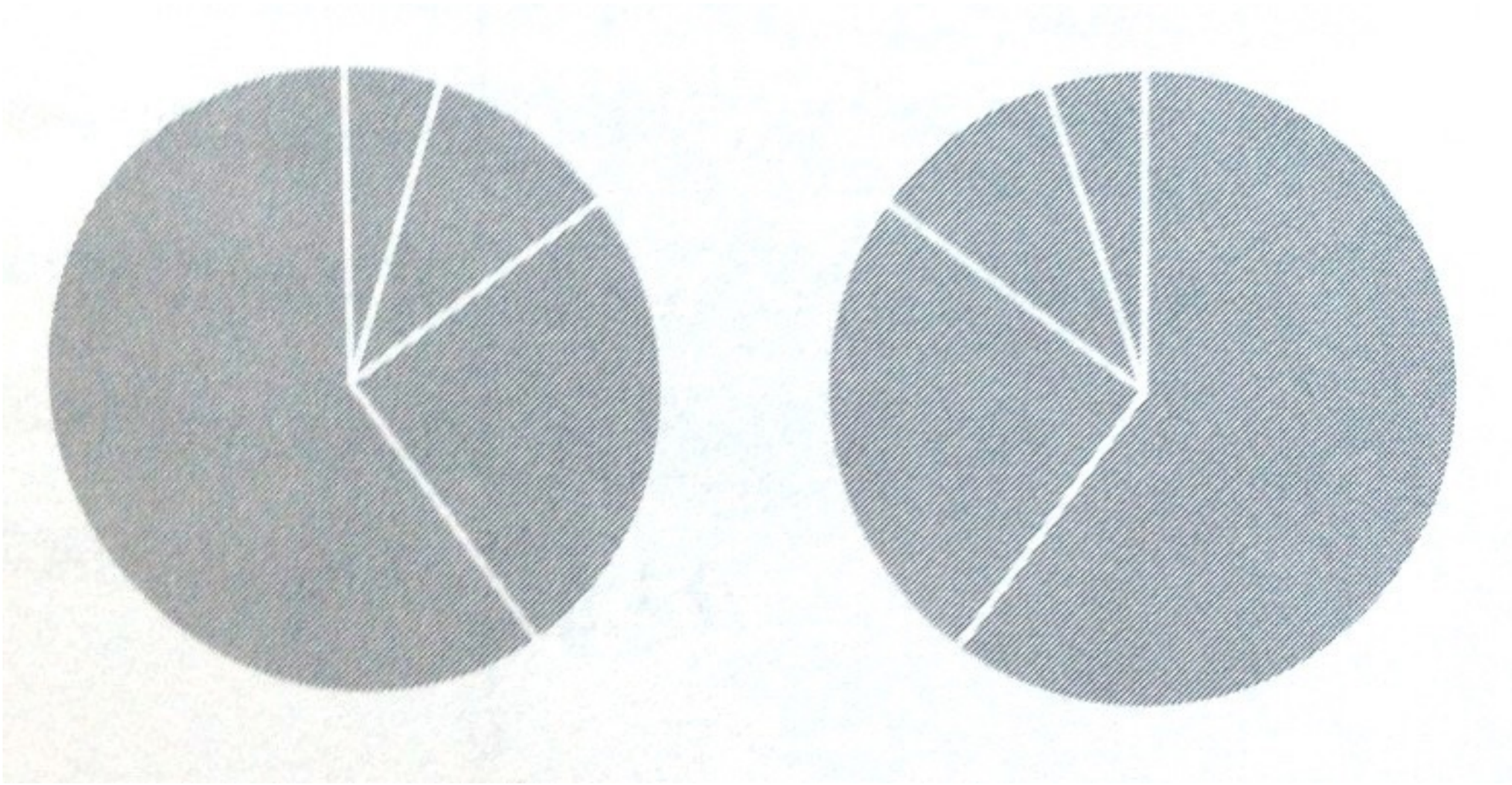
Which one is more effective? Why?
What if you have many lines you want to show?



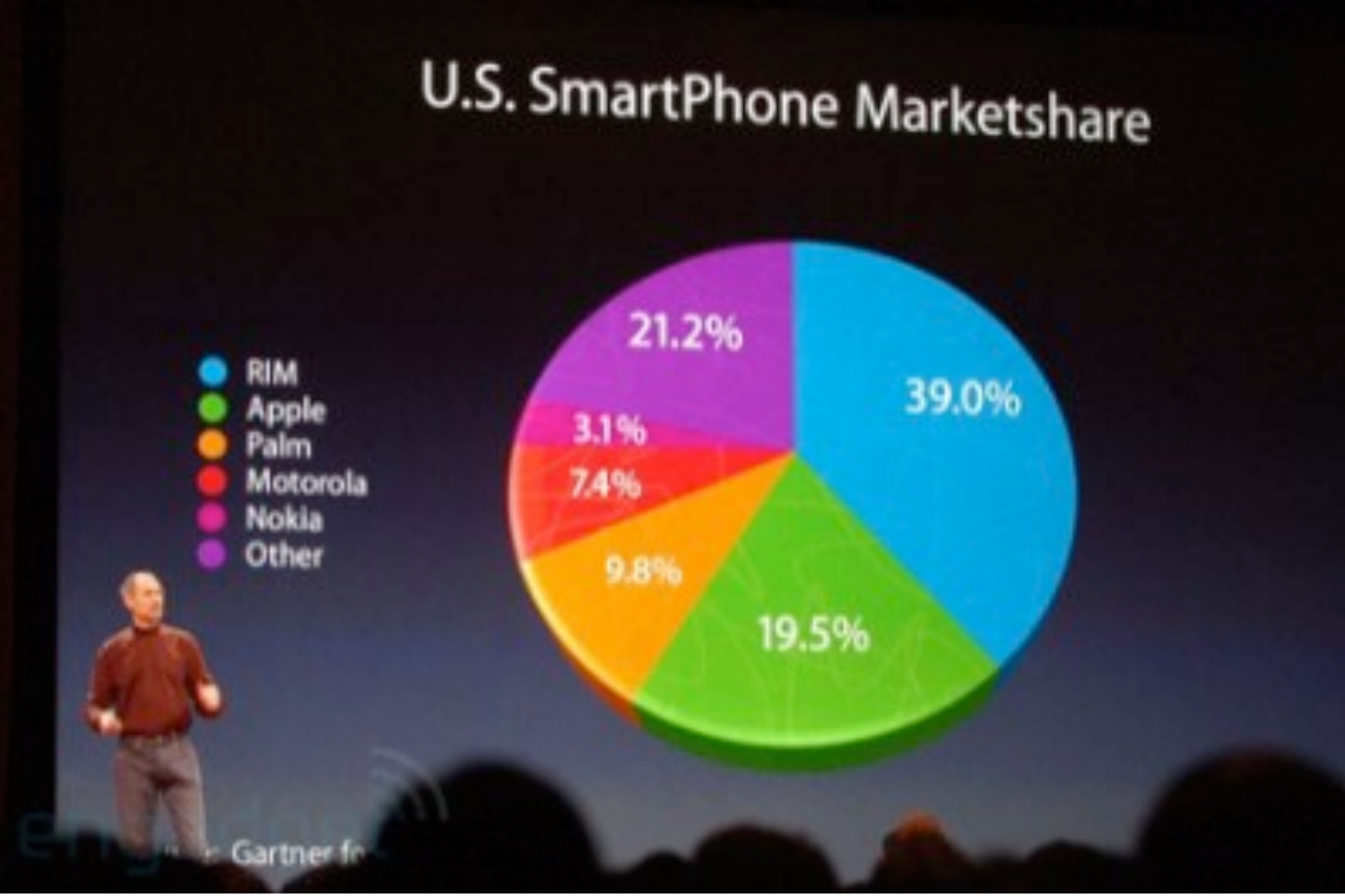
“**Small Multiple**” - Edward Tufte
Better than overlapping (sometimes)

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

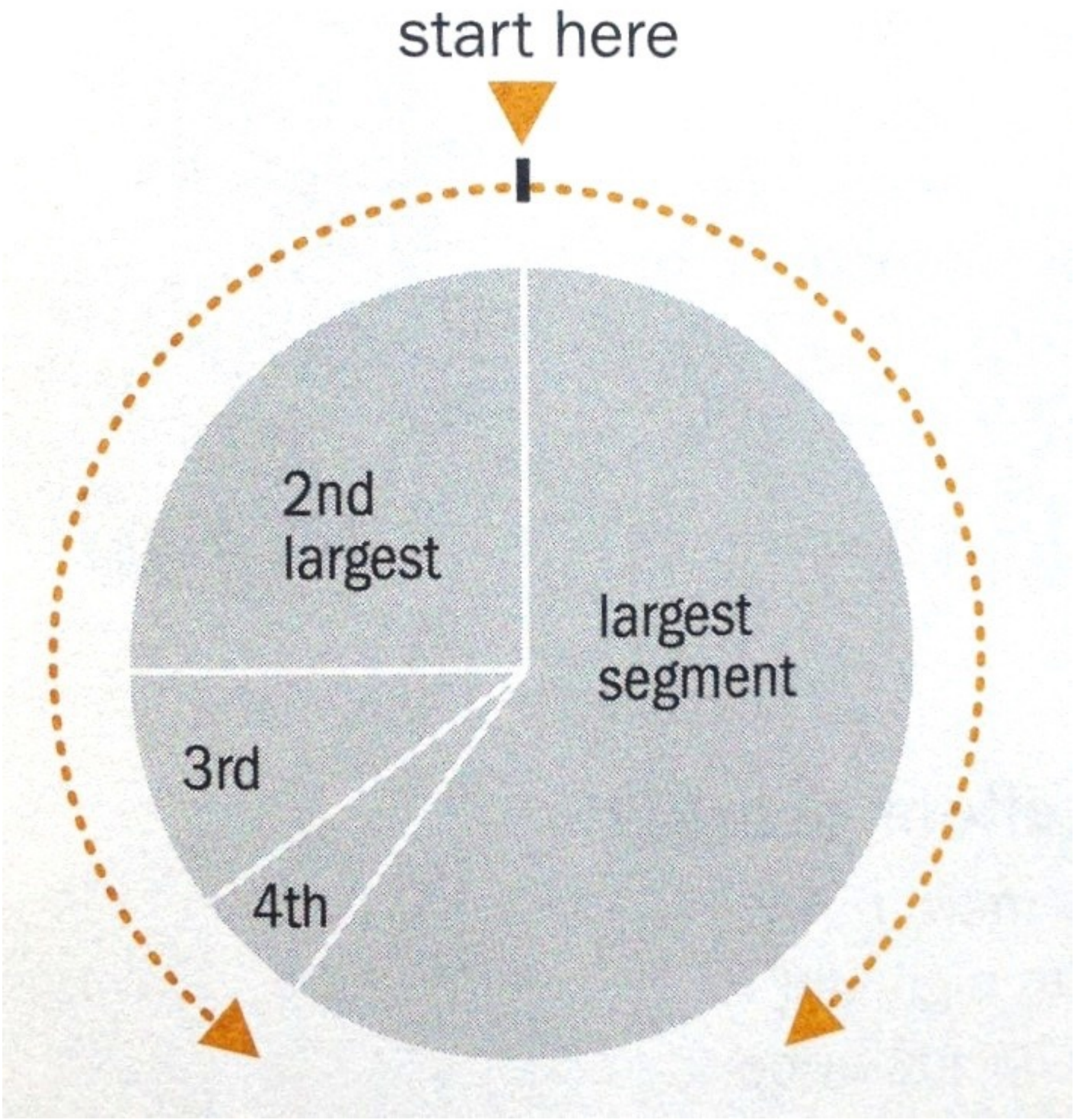
The Dreaded Pie Charts



Why people like to use pie charts?



http://www.wired.com/2008/02/macworlds-iphon/

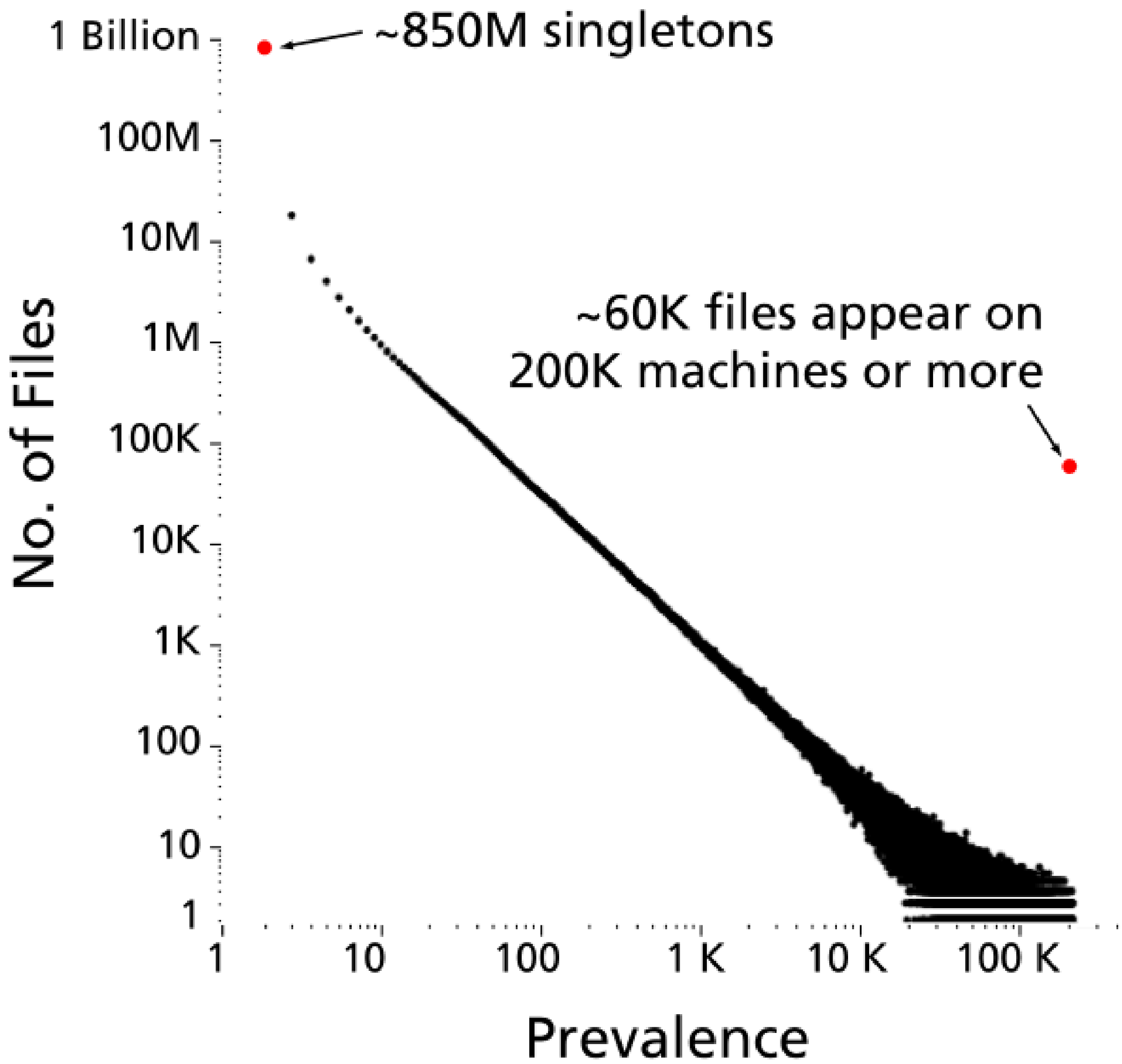

start here
2nd largest
largest segment
3rd
4th

Log scale instead of linear scale

Include numbers from different orders of magnitude

Example

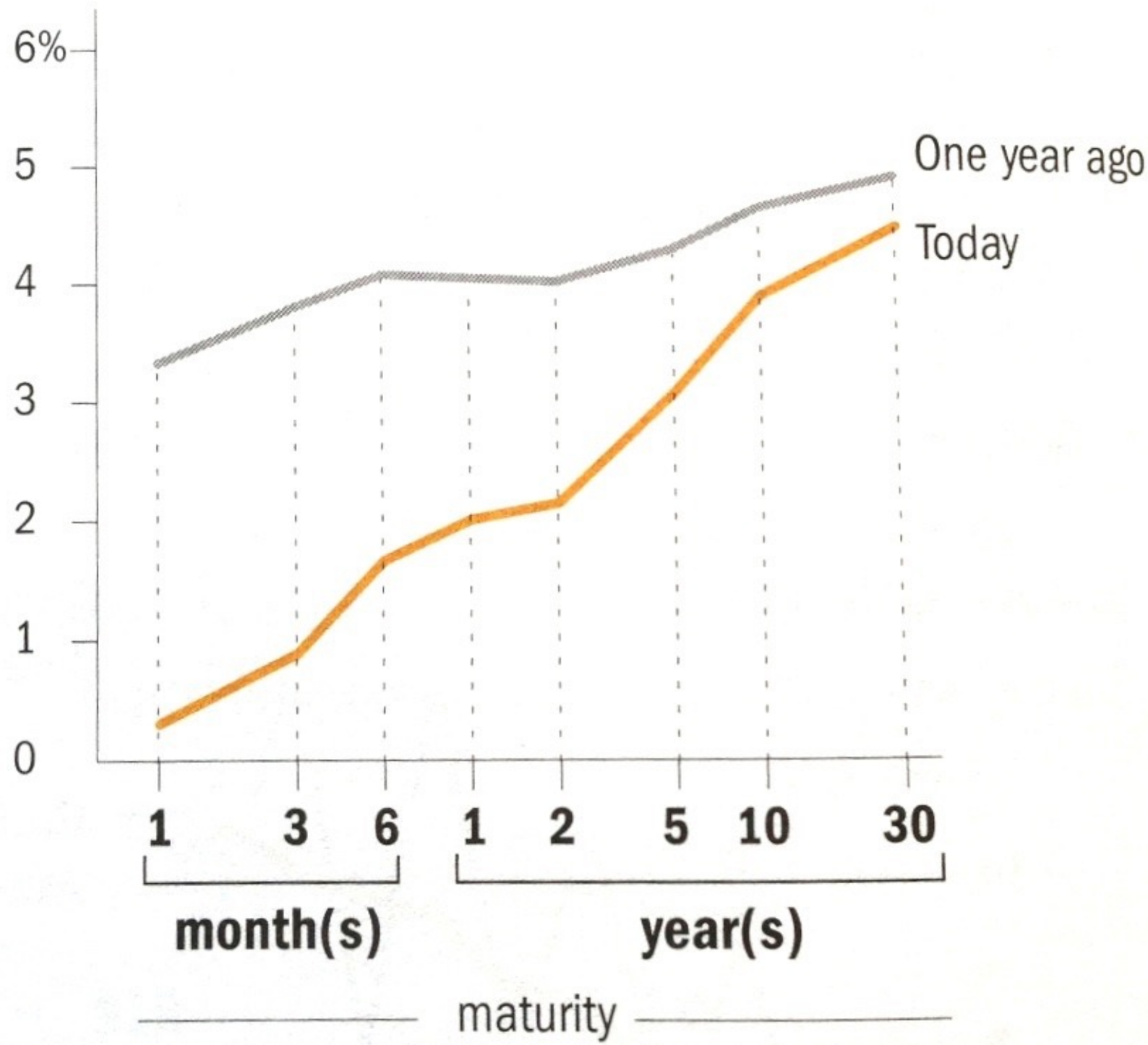
log-log



“log” also works well for time

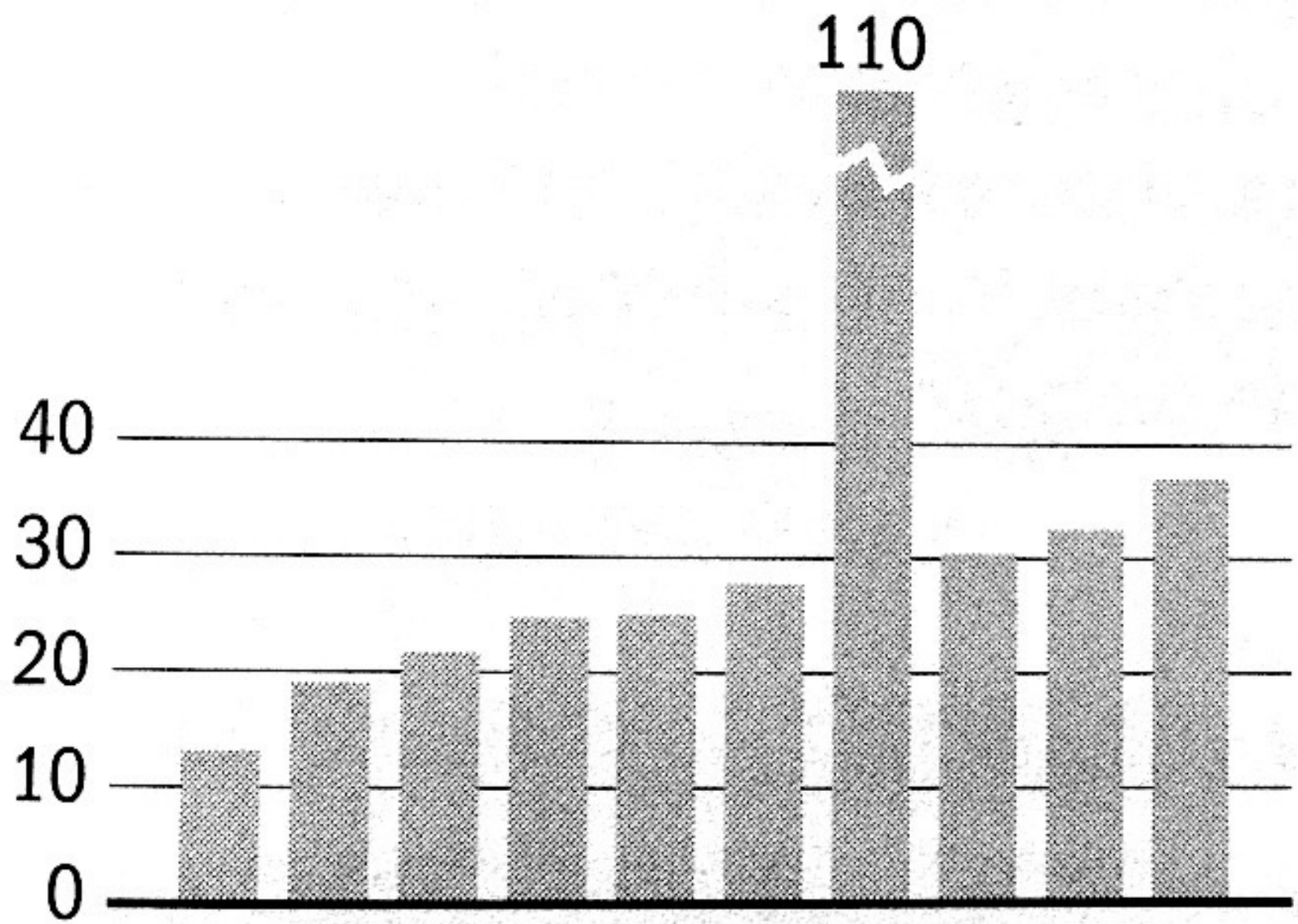
Example

The yield curve of Treasury bills, notes and bonds



OK for outliers that are *really* different

Use broken bars sparingly



Destroying your great results with poor powerpoint

Bad color schemes

can you read this?

Bad fonts

100 times faster!

Too much animation

Too much data

Don McMillan: Life After Death by PowerPoint

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

Destroying your great results with poor powerpoint

How to fix?

- **Color schemes**: start with black & white, add colors if needed
- **Fonts**: sans-serif font looks nicer
  - On Mac: Helvetica is always good
  - On Windows: Arial?
- **Too much animation**: start with **no** animation, then add if appropriate
- **Too much data:** don't just copy figures from paper and past them on the slides!

Don McMillan: Life After Death by PowerPoint

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

Suggestions: use pictures whenever appropriate

"Pictures" include most *non-text* elements: tables, diagrams, charts, etc.

Why?

- "A picture is worth a thousand words"
- People like pictures and love movies.
- Picture is often more succinct, memorable

Figures should be self-contained

Why?

- Don’t make people go back and forth between text & figure
- People **skim**; look at “interesting” things first
- Especially in academia, busy reviewers look at figures first
- Bad figures -> **bad first impression**
  (lower chance of paper acceptance)

How to fix?

- Succinctly describe your main messages
  (what you want the readers to learn)

Example

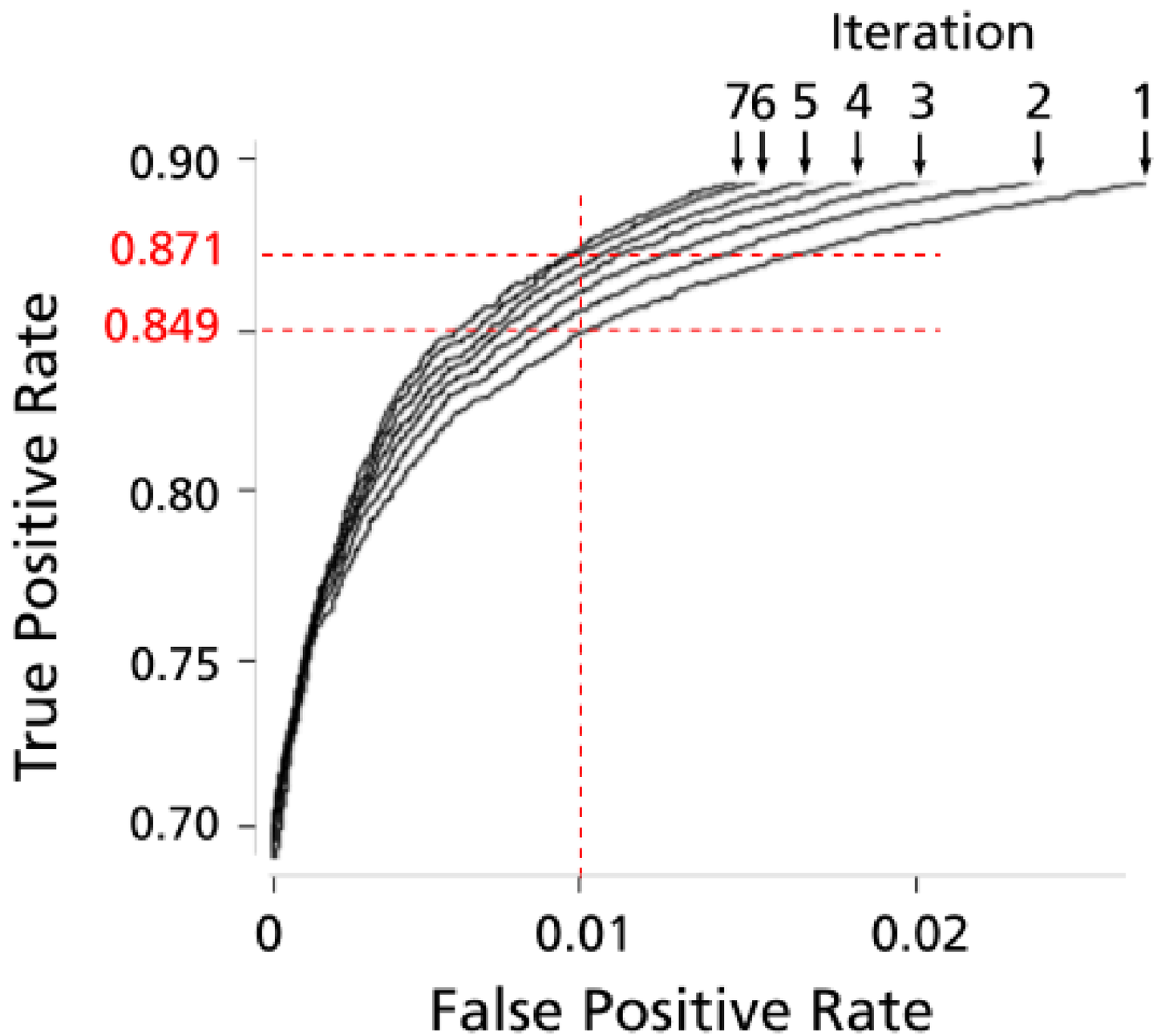


Figure 8: ROC curves of 7 iterations; true positive rate incrementally improves.

http://www.cs.cmu.edu/~dchau/polonium_sdm2011.pdf

Example

SIGGRAPH2007 Computer Graphics Proceedings, Annual Conference Series, 2007

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

Crown-jewel figure on first page

(nice to have)

Why?

- Give an overview of what readers is going to get -- cut to the chase
- Again, people like to see interesting things

How to do it?

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

Example

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

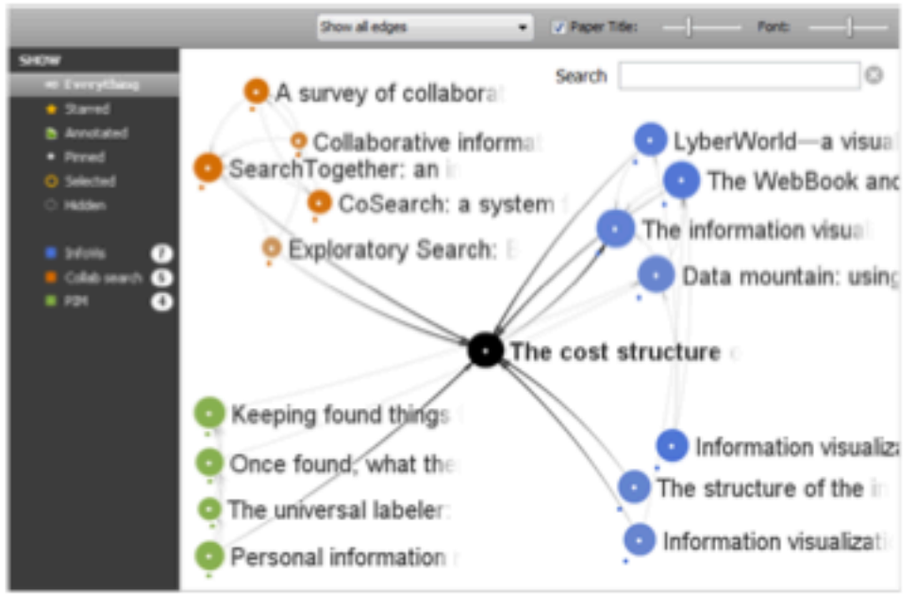


Figure 1. Apolo displaying citation network data around the article *The Cost Structure of Sensemaking*. The user gradually builds up a mental model of the research areas around the article by manually inspecting some neighboring articles in the visualization and specifying them as exemplar articles (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

Suggestion: Design in grayscale first

Then add color

If it doesn’t look good in black and white, it’s not gonna look good with color

(Why iPhone comes in black or white first?)

Suggestion: Use legible fonts

If people can’t see it, they won’t appreciate it

For printed materials, print them out and check!

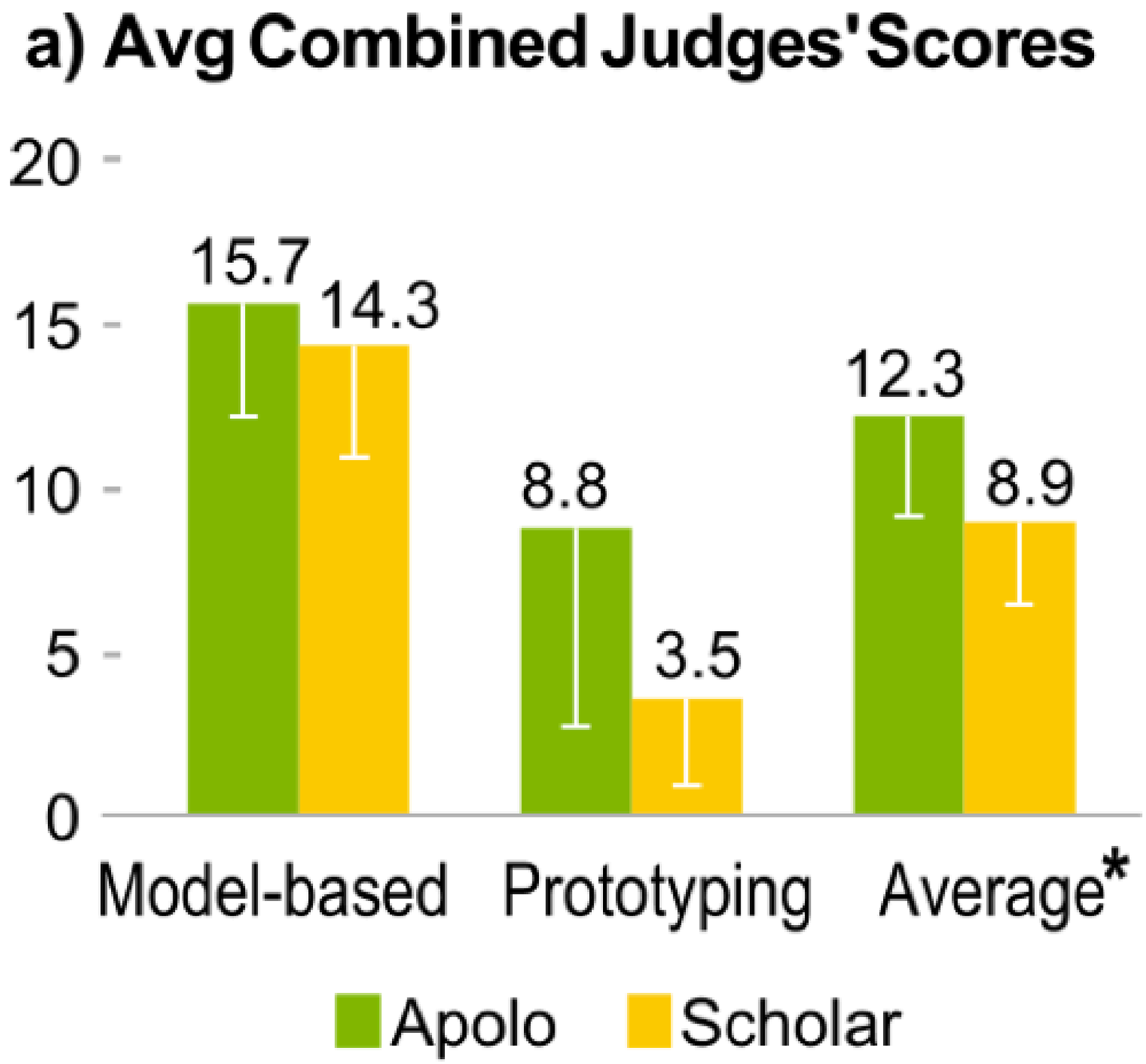
For slides, rule of thumb is about 7 lines of text per slide.

Suggestion: you probably need to redo your figure for slides

Designing for print is different from designing for the screen

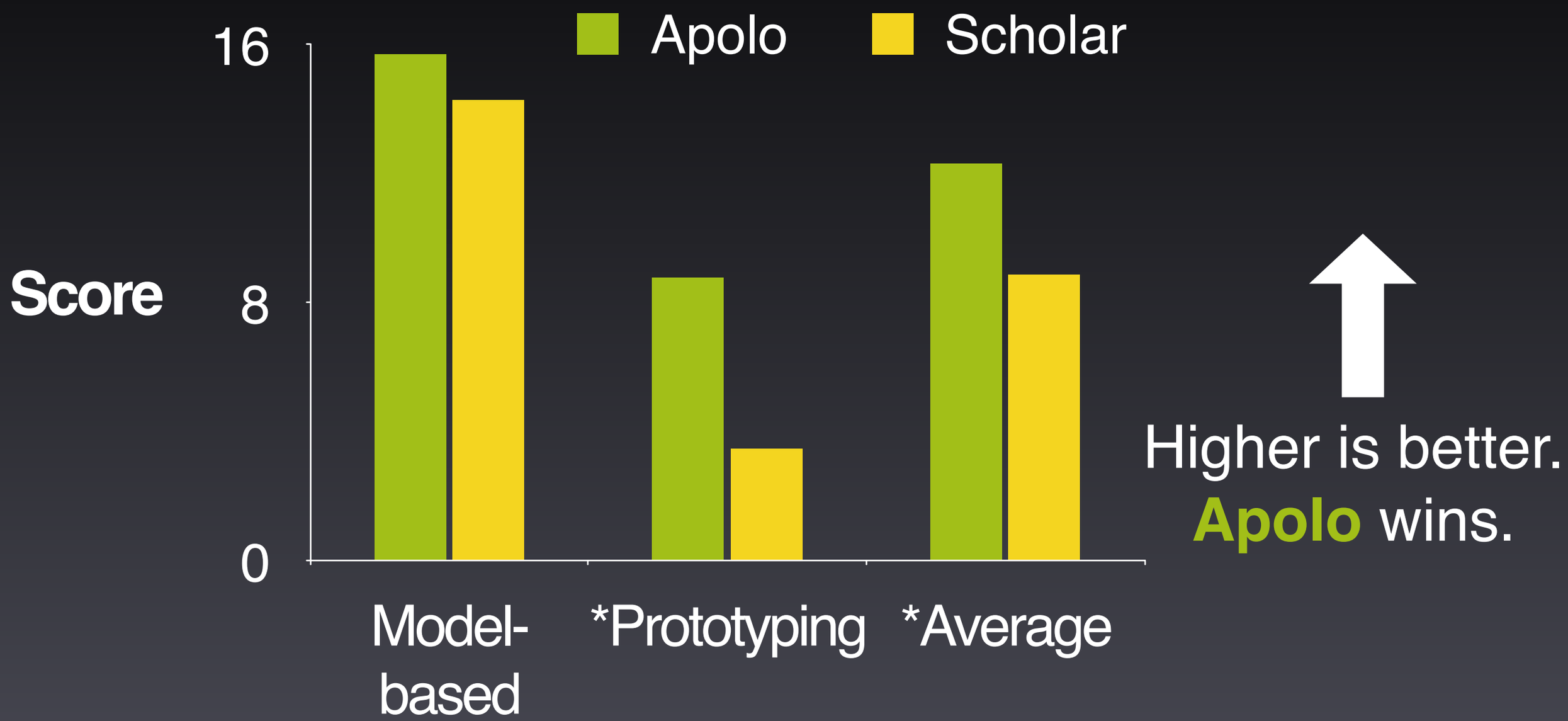
- Resolution (which is higher?)
- Levels of details (people mostly want a few “take-away” messages from your talk)

Example


a) Avg Combined Judges'Scores
20
15
10
5
0
15.7
14.3
8.8
3.5
12.3
8.9
Model-based
Prototyping
Average*
Apolo
Scholar

Example

Judges' Scores



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

Project

Description is out, on course homepage

50% of course grade

4-5 students in a group

You may mix grads, undergrads, on-campus, Q & Q3 students

You must first read the “Teaming” section for caveats and suggestions

Project

3 core requirements

1. **Large** dataset
2. **Non-trivial** analysis/algorithms/computation
3. An **interactive user interface** that interact with the algorithms

Grading & Schedule

- Proposal (writeup + in-class presentation)
- Progress report (mostly as a “checkpoint”)
- Final report (writeup + poster presentation)

George Heilmeier

Former Director of DARPA

Heilmeier Questions

Preflight checklist for (your) successful projects

1. What are you trying to do?
   Articulate your objectives using absolutely no jargon.
2. How is it done today; what are the limits of current practice?
3. What's new in your approach; why it will be successful?
4. Who cares?
5. If you're successful, what difference will it make?
6. What are the risks and payoffs?
7. How much will it cost?
8. How long will it take?
9. What are the midterm and final "exams" to check for success?

http://en.wikipedia.org/wiki/George_H._Heilmeier
http://smlv.cc.gatech.edu/2010/10/17/heilmeiers-questions/

Aurigo: An Interactive Tour Planner for Personalized Itineraries

Alexandre Yahi*, Antoine Chassang*, Louis Raynaud*, Hugo Duthil*, Duen Horng (Polo) Chau
Georgia Institute of Technology
{alexandre.yahi, antoine.chassang, l.raynaud, hduthil, polo}@gatech.edu

ABSTRACT

Planning personalized tour itineraries is a complex and challenging task for both humans and computers. Doing it manually is time-consuming; approaching it as an optimization problem is computationally NP hard. We present Aurigo, a tour planning system combining a recommendation algorithm with interactive visualization to create personalized itineraries. This hybrid approach enables Aurigo to take into account both quantitative and qualitative preferences of the user. We conducted a within-subject study with 10 participants, which demonstrated that Aurigo helped them find points of interest quickly. Most participants chose Aurigo over Google Maps as their preferred tools to create personalized itineraries. Aurigo may be integrated into review websites or social networks, to leverage their databases of reviews and ratings and provide better itinerary recommendations.

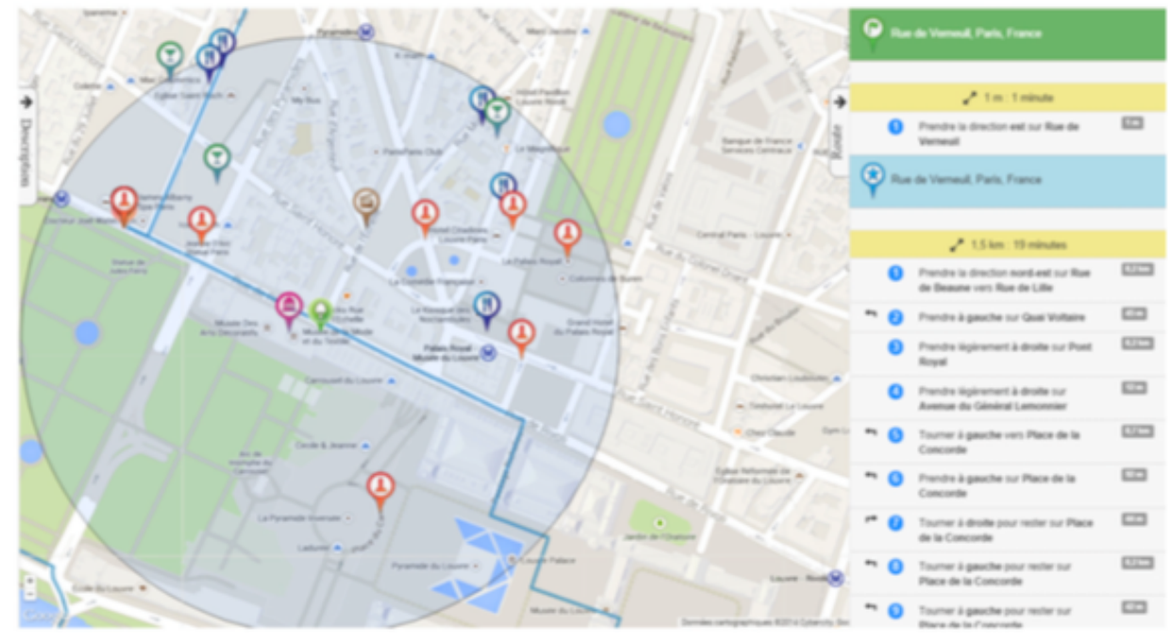
Author Keywords

User Interfaces; Visualization; Recommendation; Tour itinerary planning

ACM Classification Keywords

(e.g. HCI): User Interfaces

IUI'15 Full conference paper



ISPARK: Interactive Visual Analytics for Fire Incidents and Station Placement

Subhajit Das, Andrea McCarter, Joe Minieri, Nandita Damaraju, Sriram Padmanabhan, Duen Horng (Polo) Chau
Georgia Tech
Atlanta, GA, USA
{das, andream, jminieri, nandita, sriramp, polo}@gatech.edu

ABSTRACT

In support of helping to reduce the response time of firefighters, and thus deaths, injuries, and property loss due to fires, we introduce ISPARK. The ISPARK system determines where fire stations should be located, analyzes the primary causes of fires, the existing infrastructure, and response times, by using visualizations which show the GIS mapping of fire stations on a dashboard. Incidents and response times are shown as additional layers, with clustering of fire incidents to determine predicted fire station locations, forecasting of fire incidents using regression, causal, infrastructure, and personnel analysis, creating an interactive, multi-faceted method for locating fire stations. A comparison of urban and rural fire incident response times is another dimension of this study. We demonstrate ISPARK's usage and benefits using a publicly available dataset describing 300,000 fire incidents in the states of Massachusetts and Maine. ISPARK is generalizable to other geographic areas

Figure 1: Screenshot of ISPARK showing actual (pink) and predicted (green) fire station locations n Maine determined by our approach, using coordiates with actual driving distances from fire stations o actual fire incidents. Fire incidents are shown as small yellow dots. ISPARK reduces the average

KDD'15 Workshop paper

Categories and Subject Descriptors