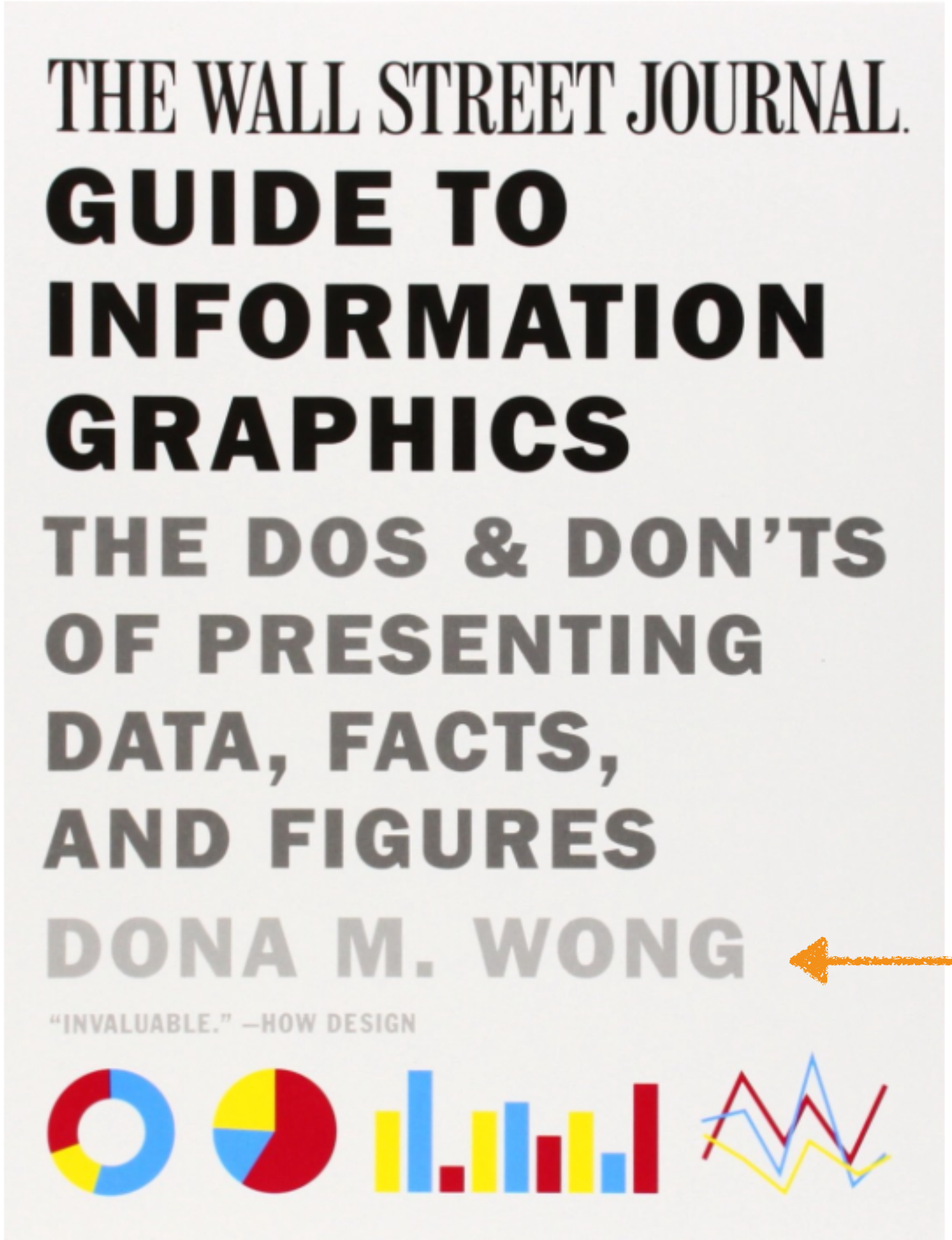
http://poloclub.gatech.edu/cse6242

CSE6242/CX4242: Data & Visual Analytics

Common visualization Issues & how to fix them

Duen Horng (Polo) Chau
Professor, College of Computing
Associate Director, MS Analytics
Georgia Tech

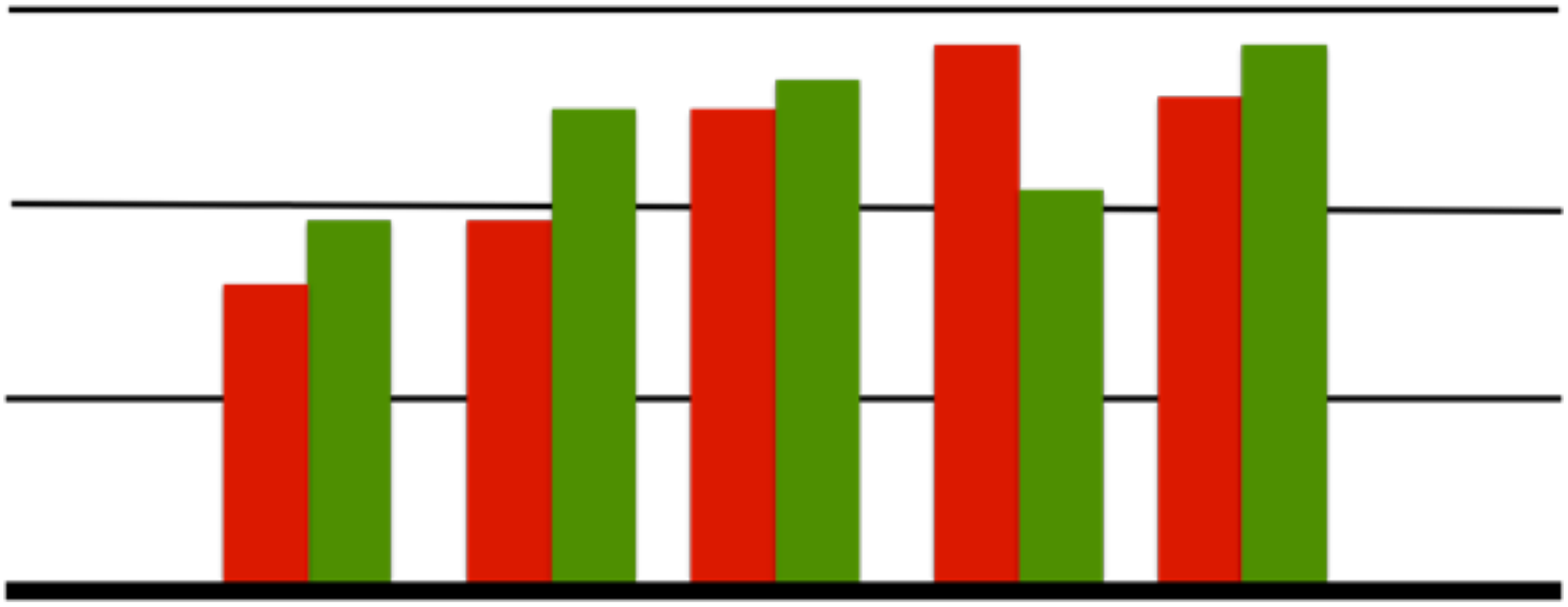
Partly based on materials by Guy Lebanon, Jeffrey Heer, John Stasko, Christos Faloutsos



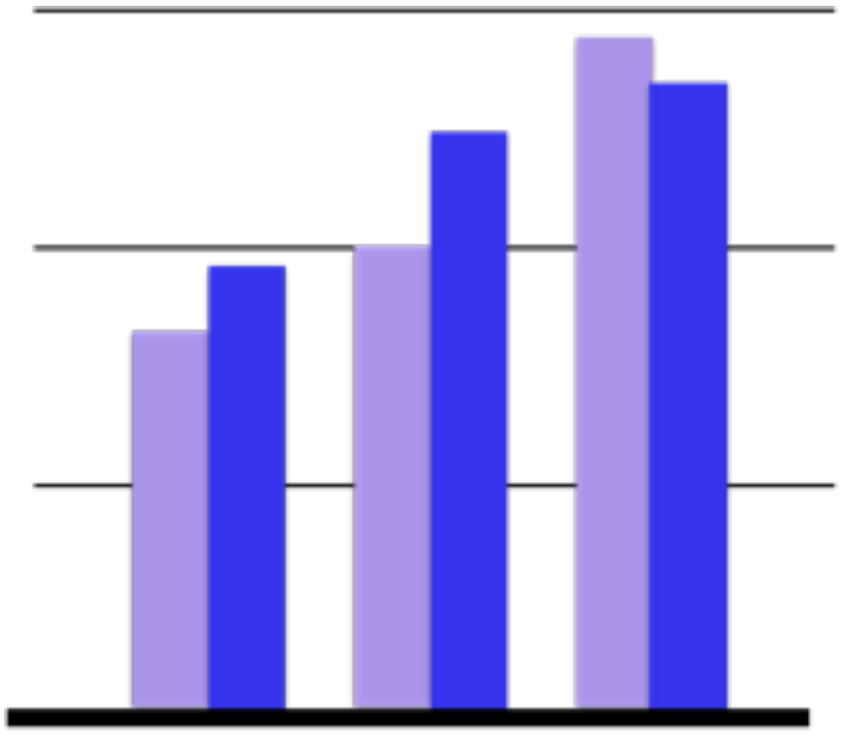
Student of
Edward Tufte

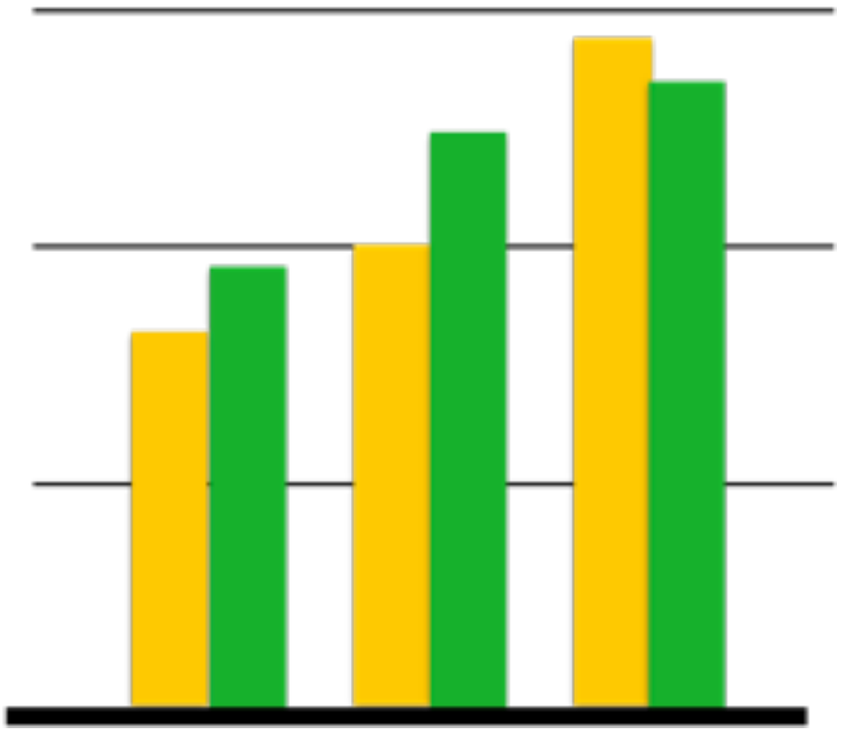
http://a.co/6BhlPfZ

Bar Charts



The color scheme reminds you of what?

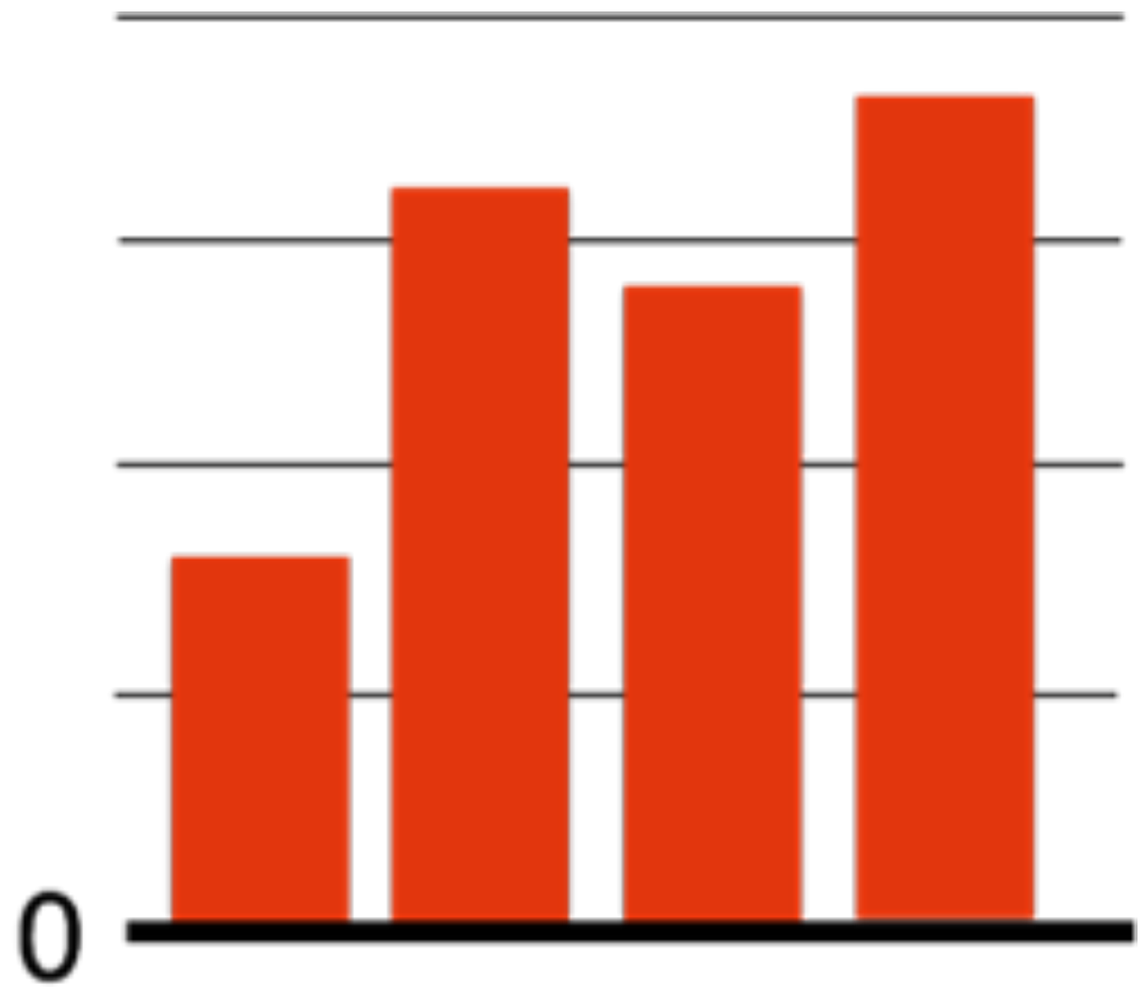




Better than Christmas

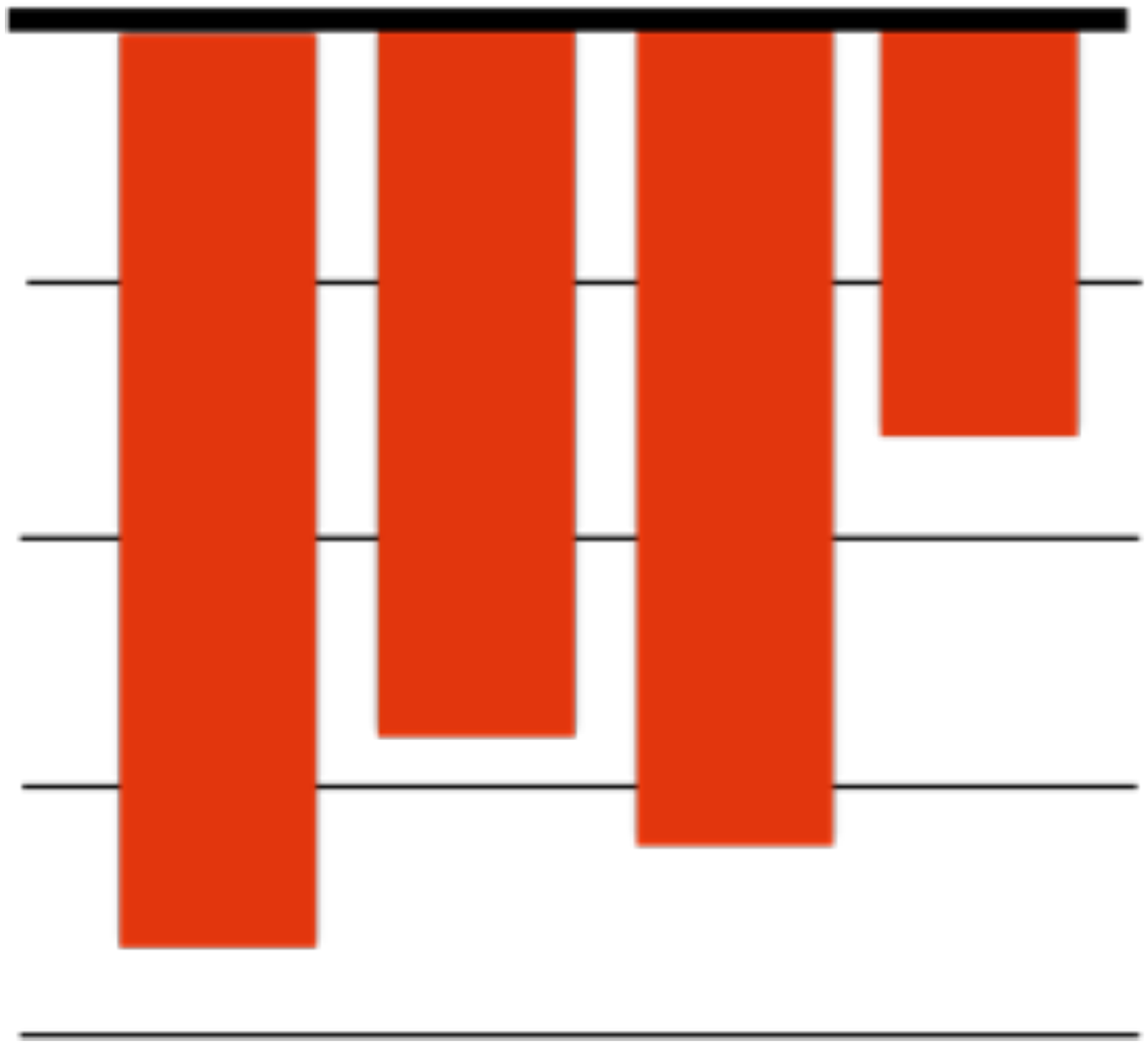
(Use color brewer to find good color schemes)

Company Profits

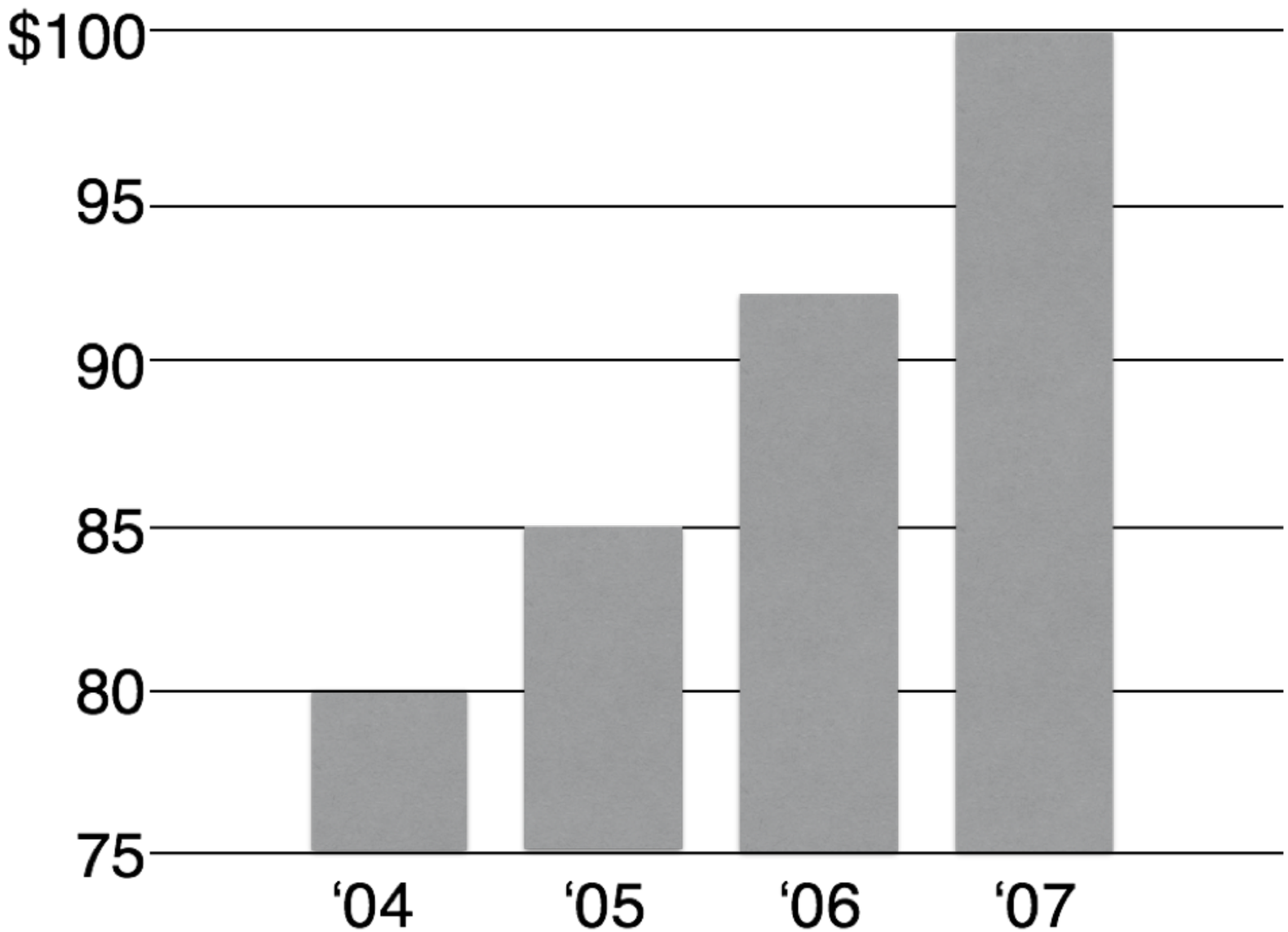


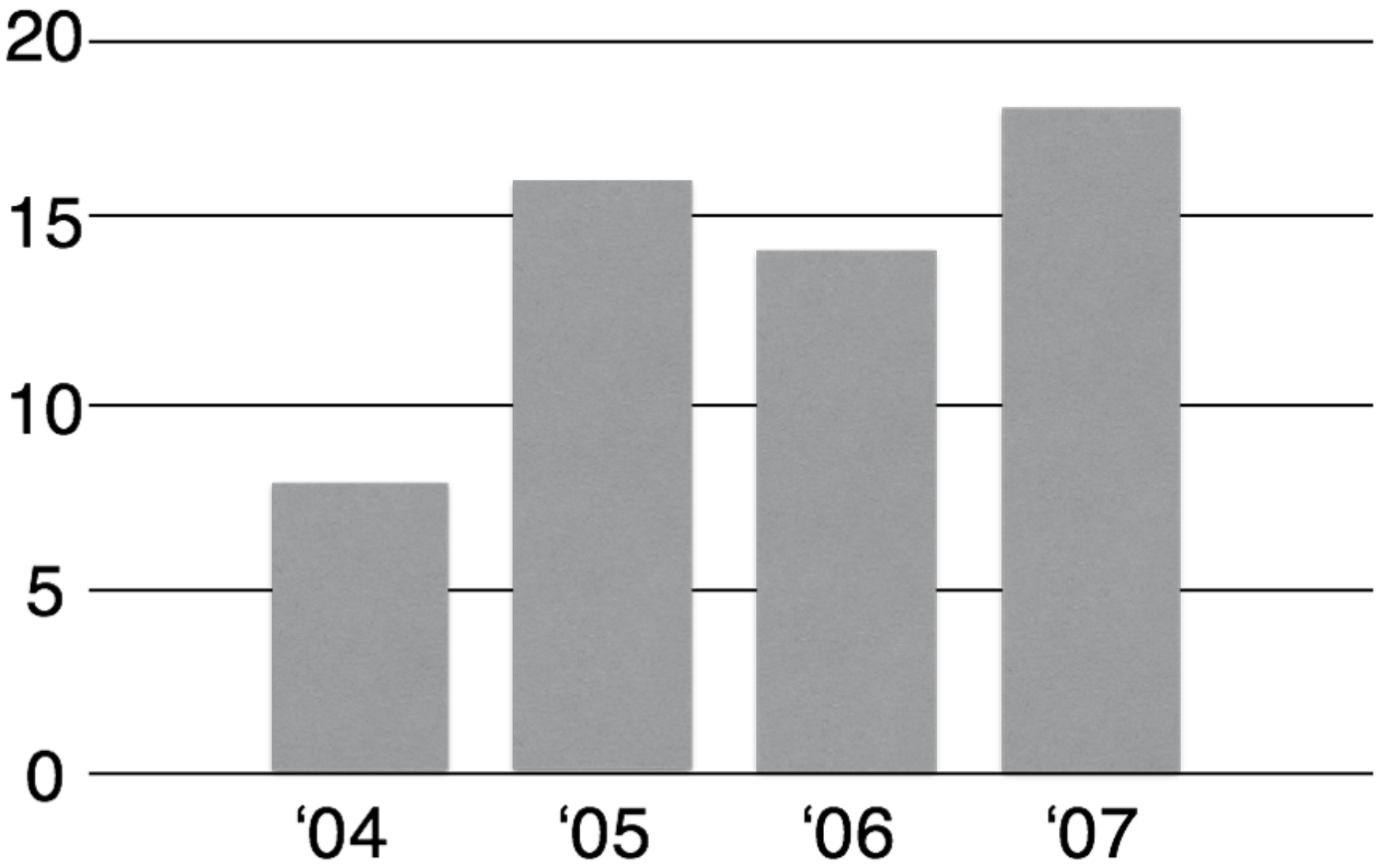
Don’t show profits in red!!

Think carefully about your color choices.



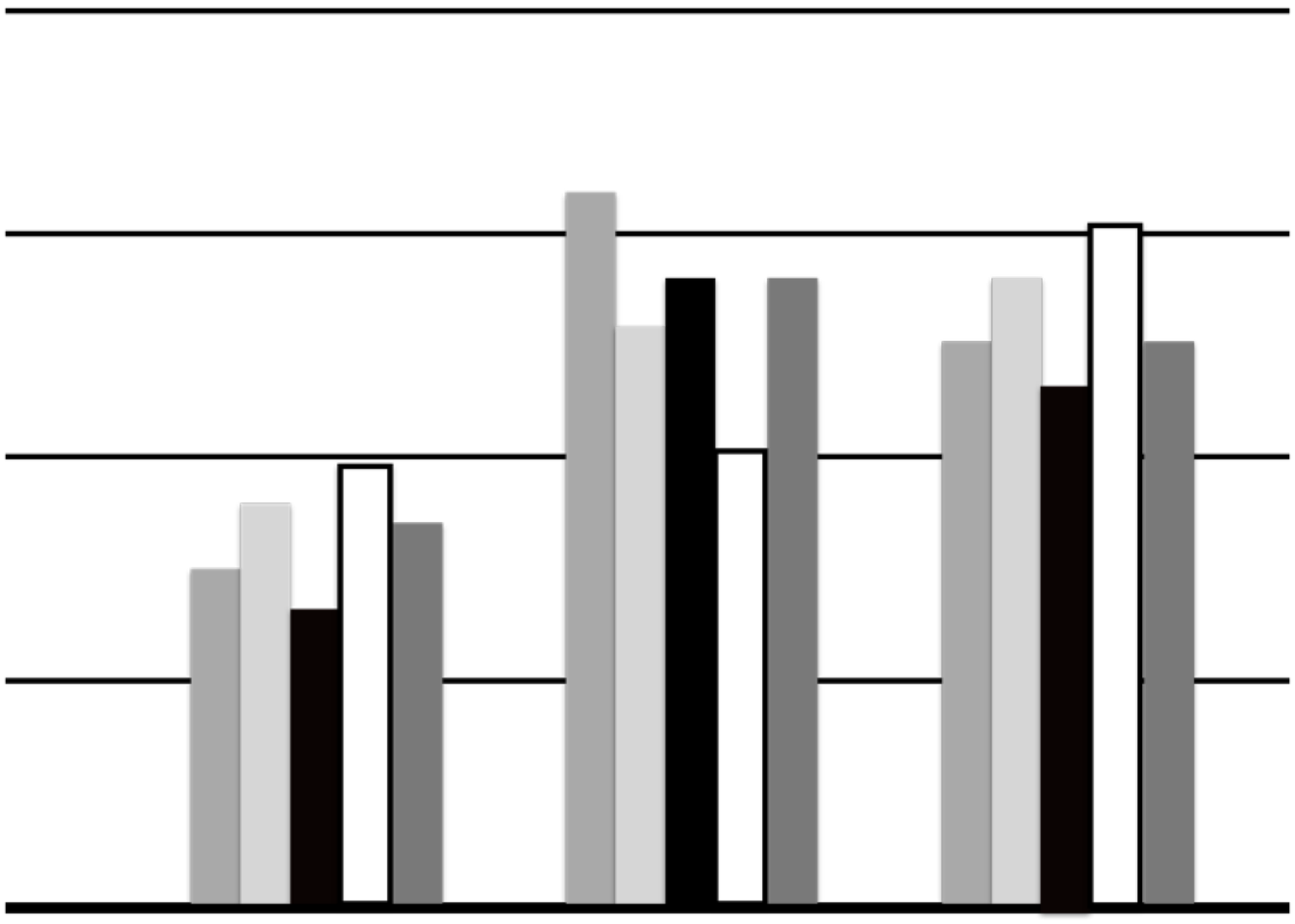
Misleading Bar Charts

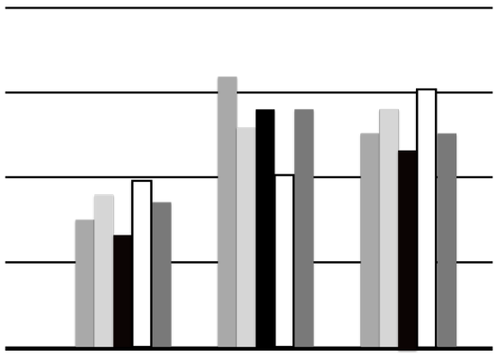




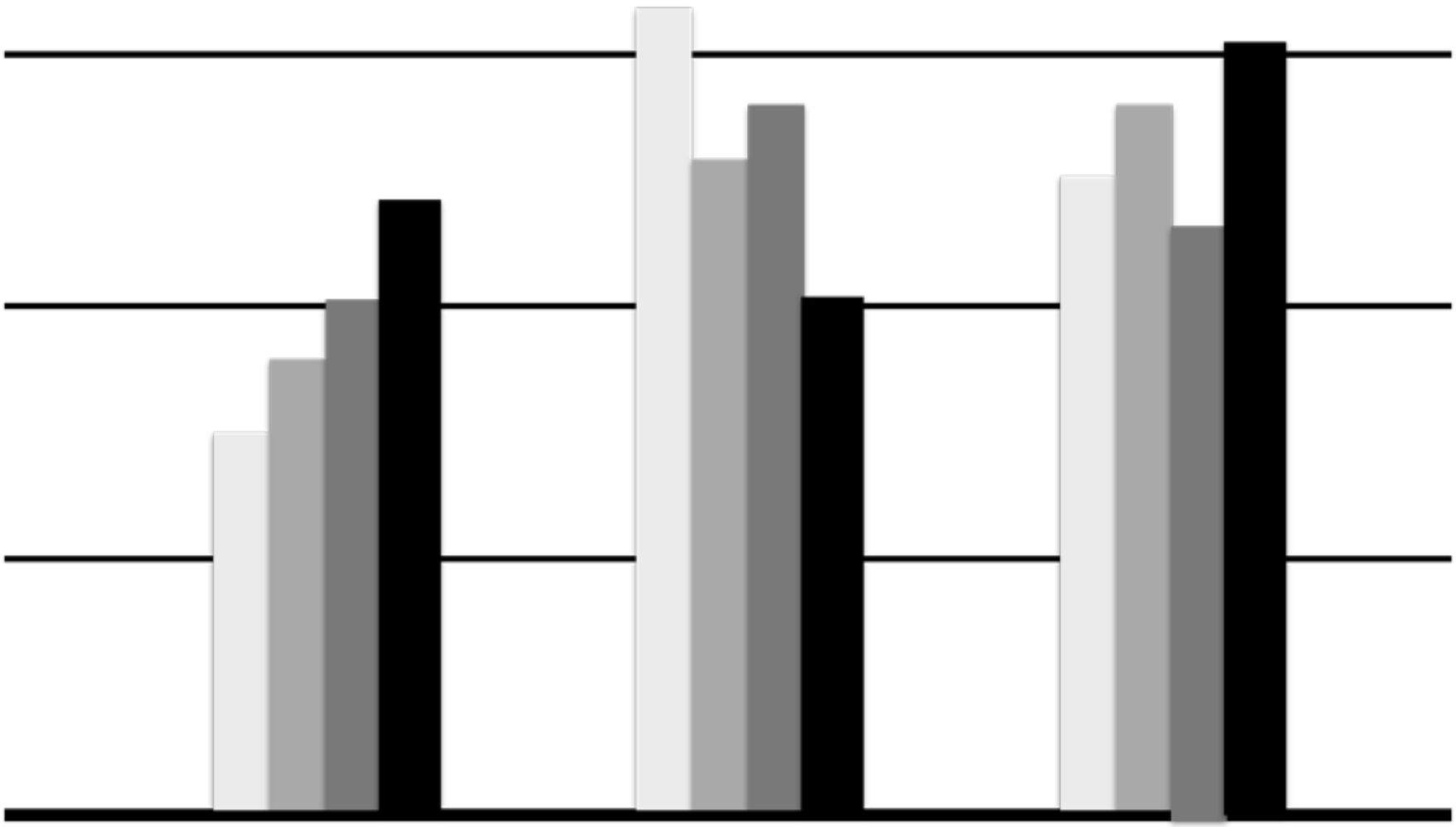
Vertical axis of bar charts should start at 0, almost always

Disorienting color bars

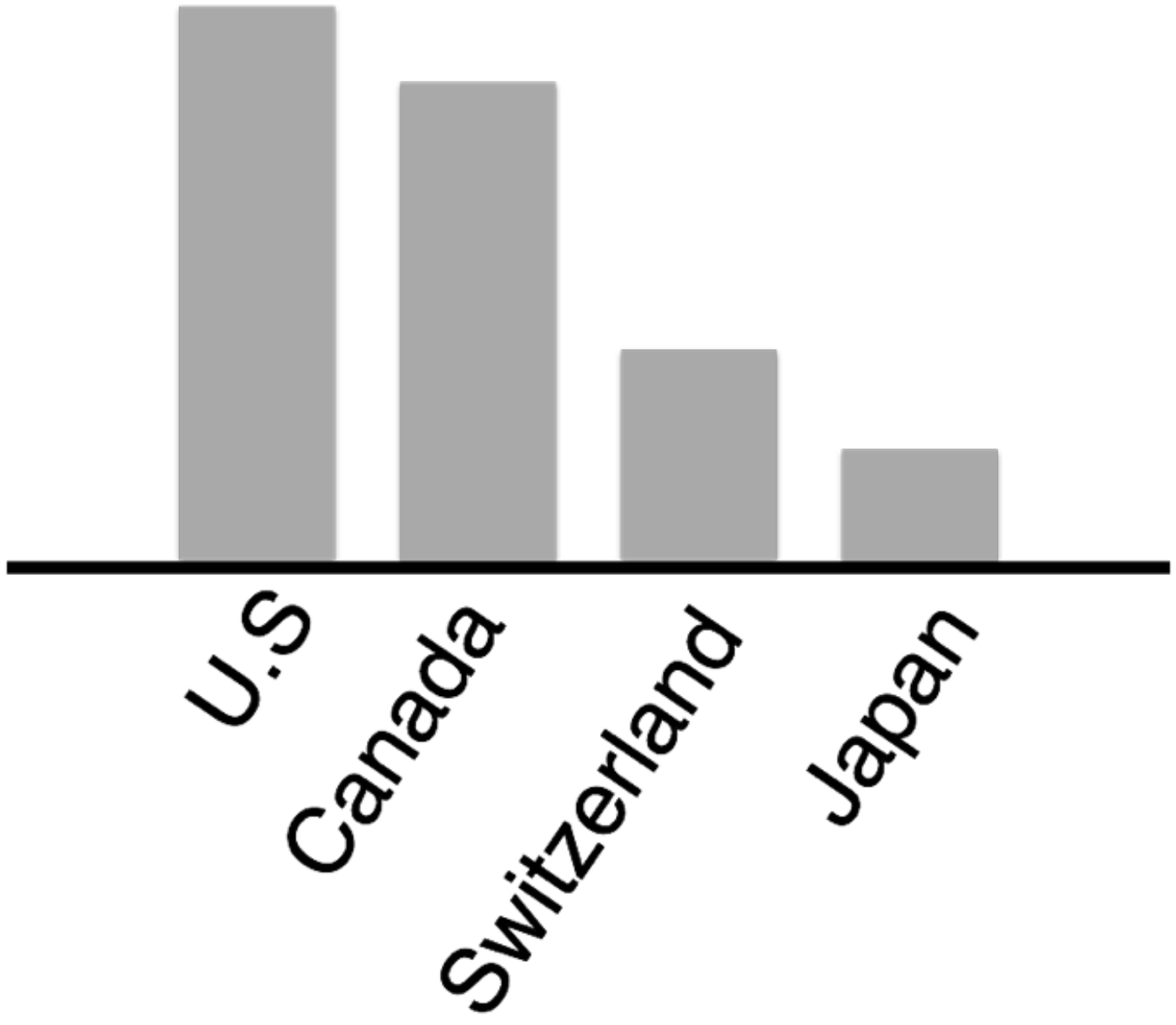




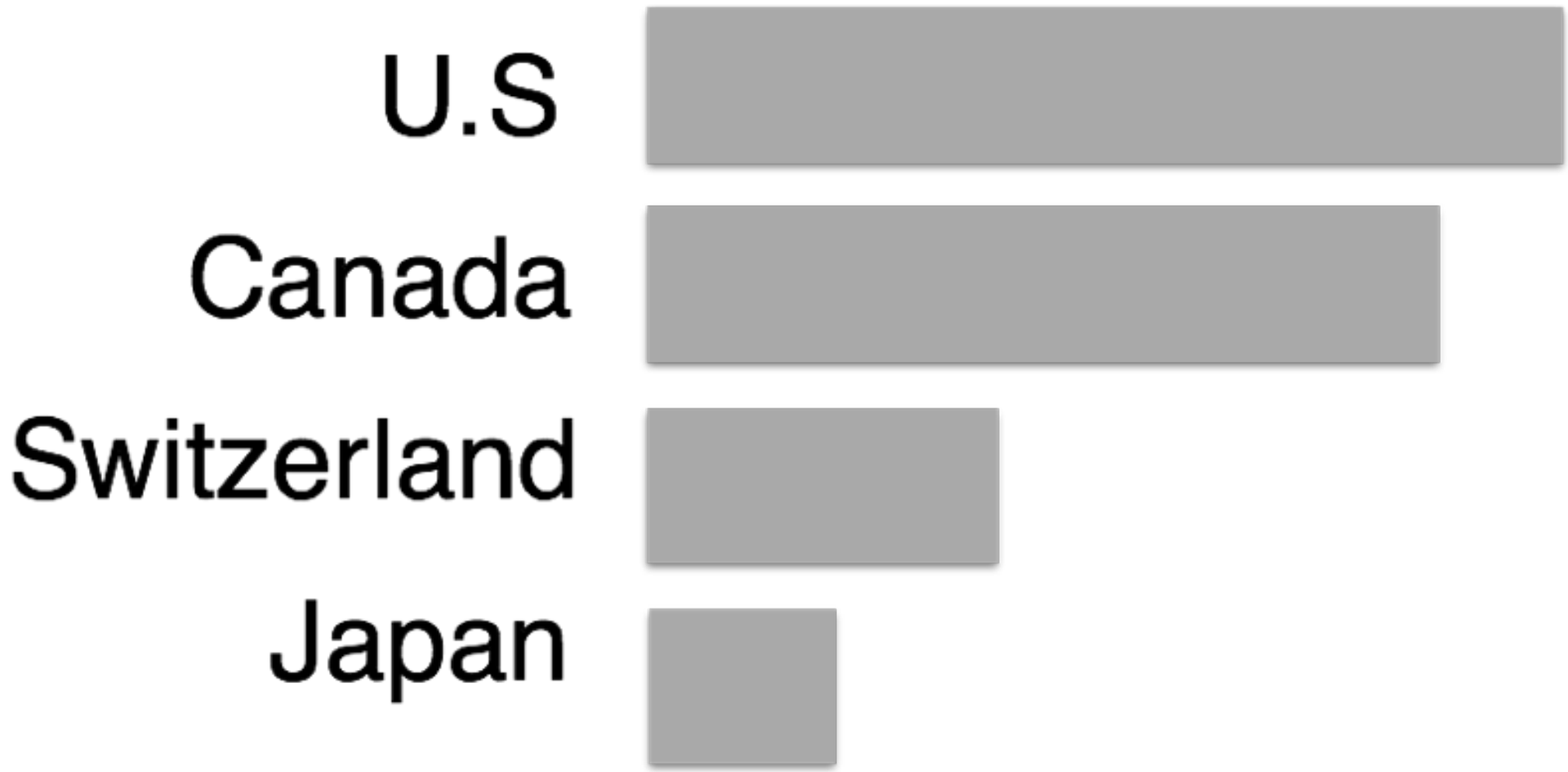
Use gradation



Avoid Tilted or Rotated Labels



Bars Can be **Horizontal**



When labels are hard to read, try horizontal layout.

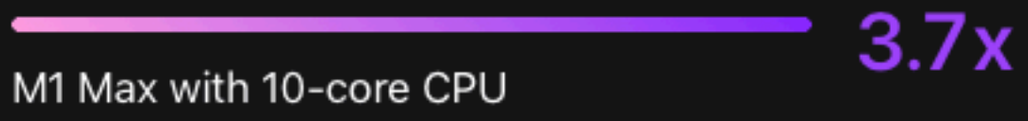
Don’t settle for the default.

CPU Performance

Xcode / NASA TetrUSS / Logic Pro / Vectorworks / Affinity Photo

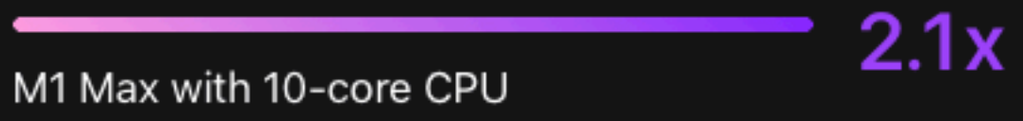
Faster project build[7]

14″ model



Quad-core Intel Core i7
13-inch MacBook Pro

16″ model



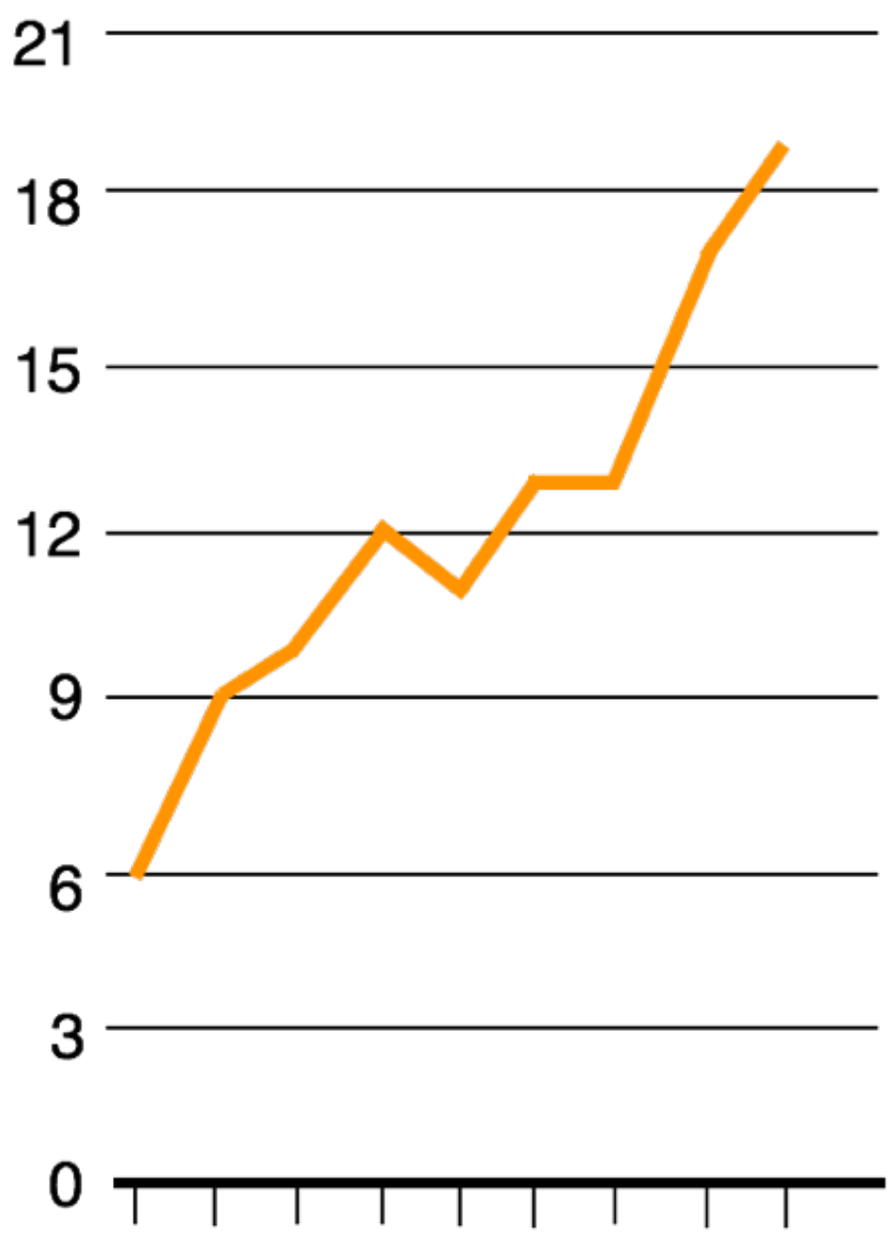
8-core Intel Core i9
16-inch MacBook Pro

https://www.apple.com/macbook-pro-14-and-16/

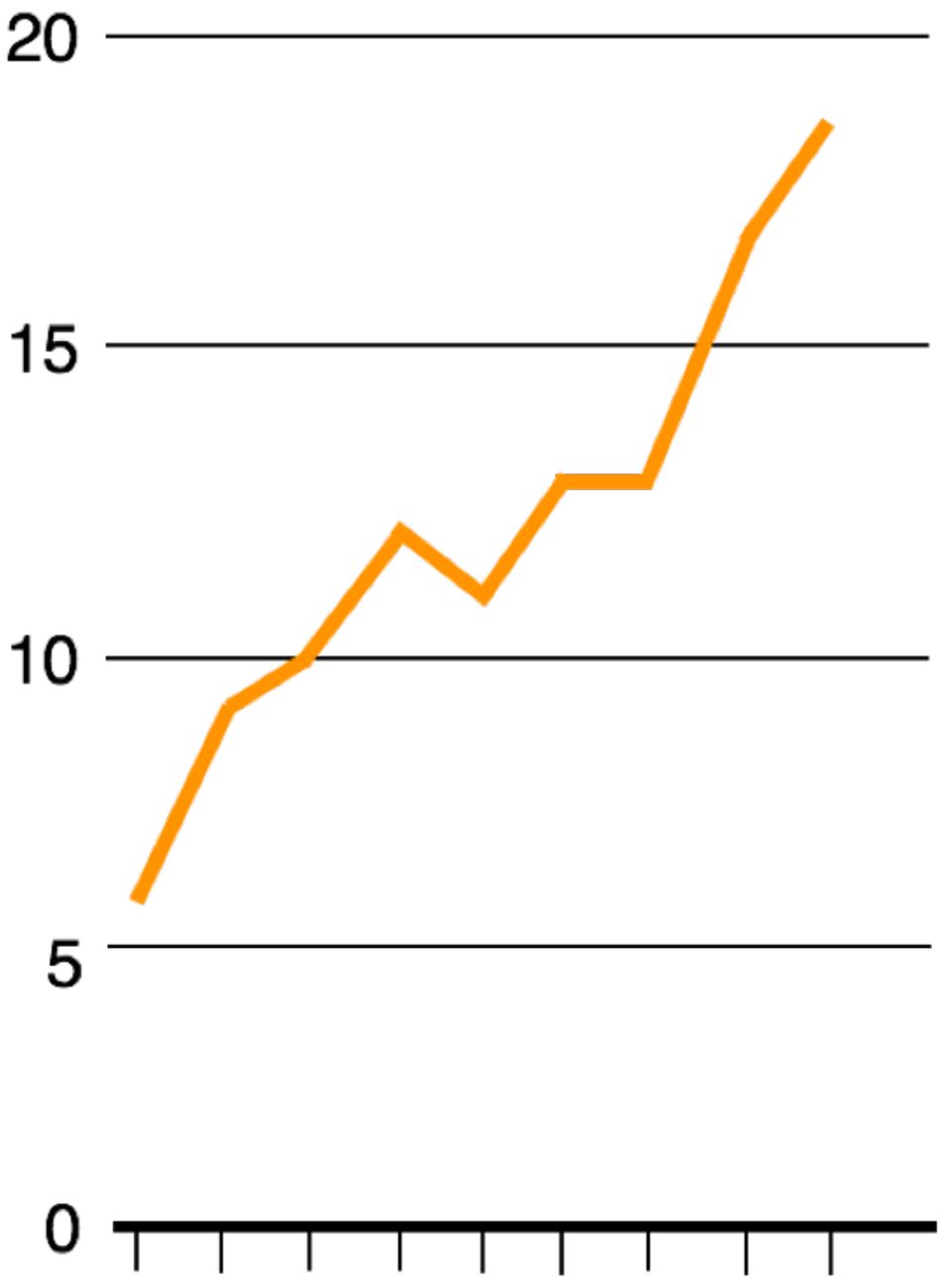
Line Charts (a.k.a. fever lines)

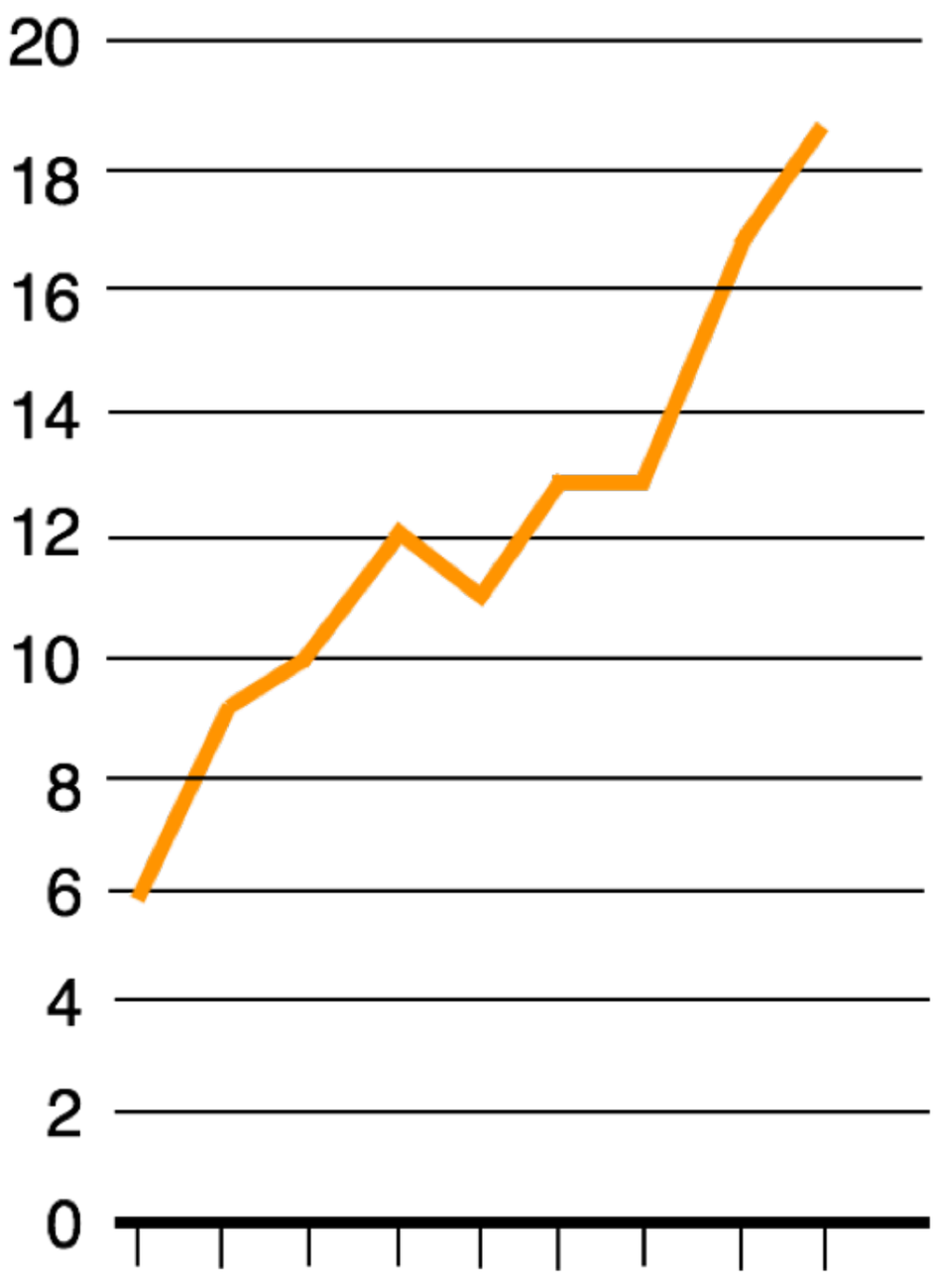
Can you improve the tick labels (e.g., 3, 6, 9, etc)?

(Ignore the missing chart title and axis labels for now)



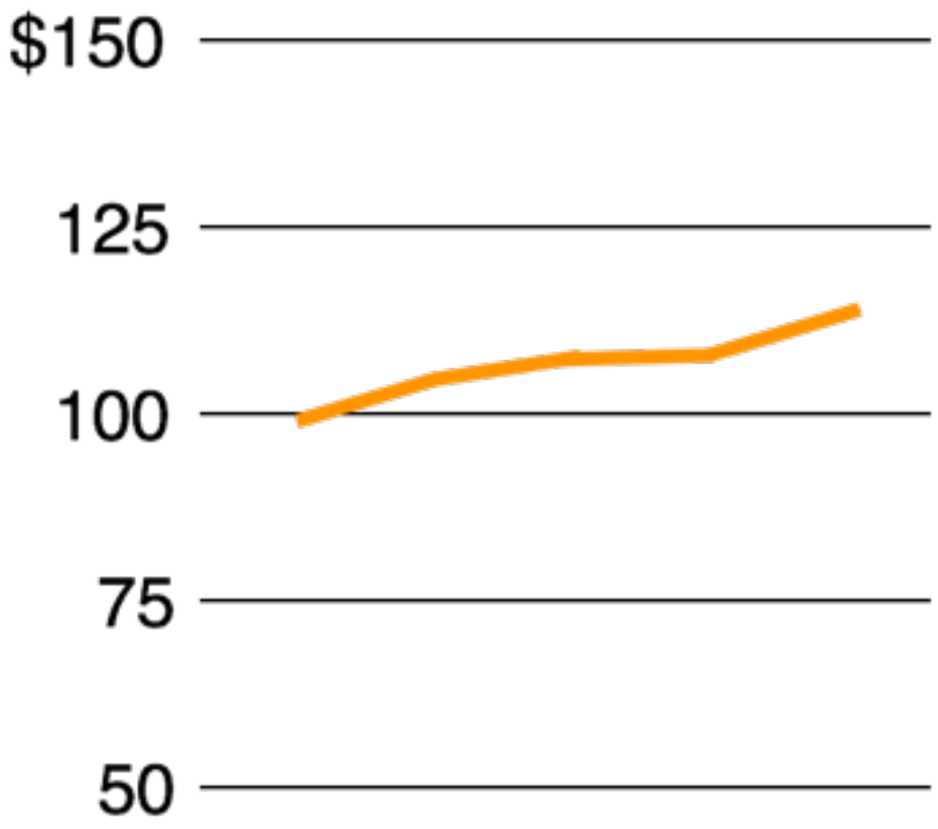
Use ticks at **common** intervals (e.g., 2, 5, 10, etc.)



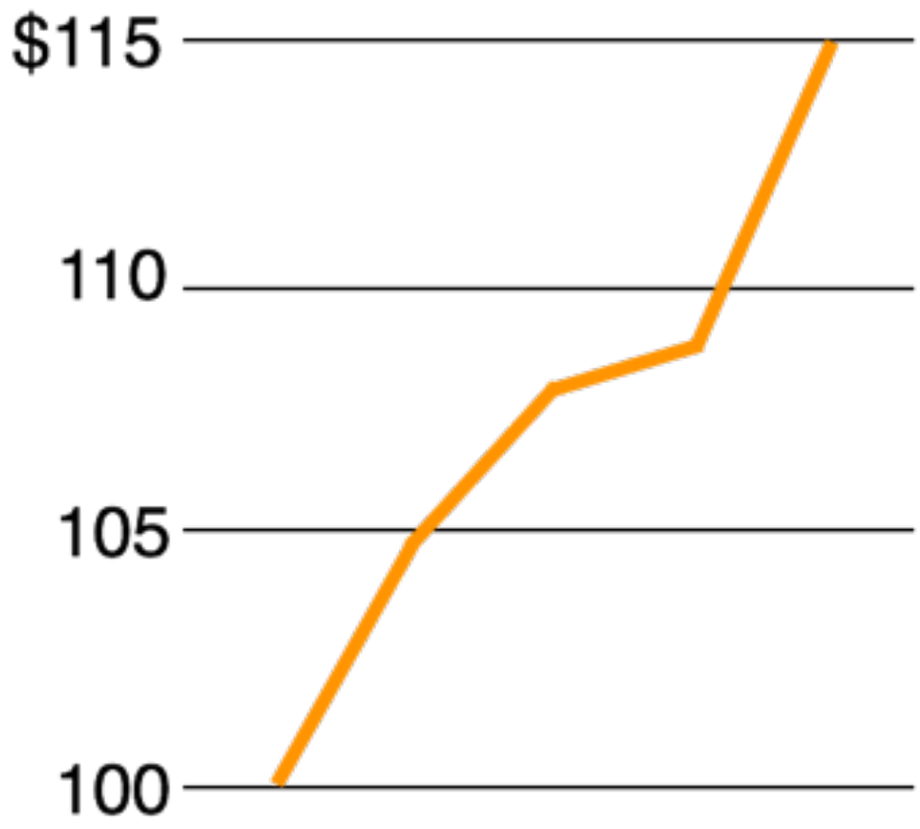


Too flat or too steep?

Too flat obscures the message

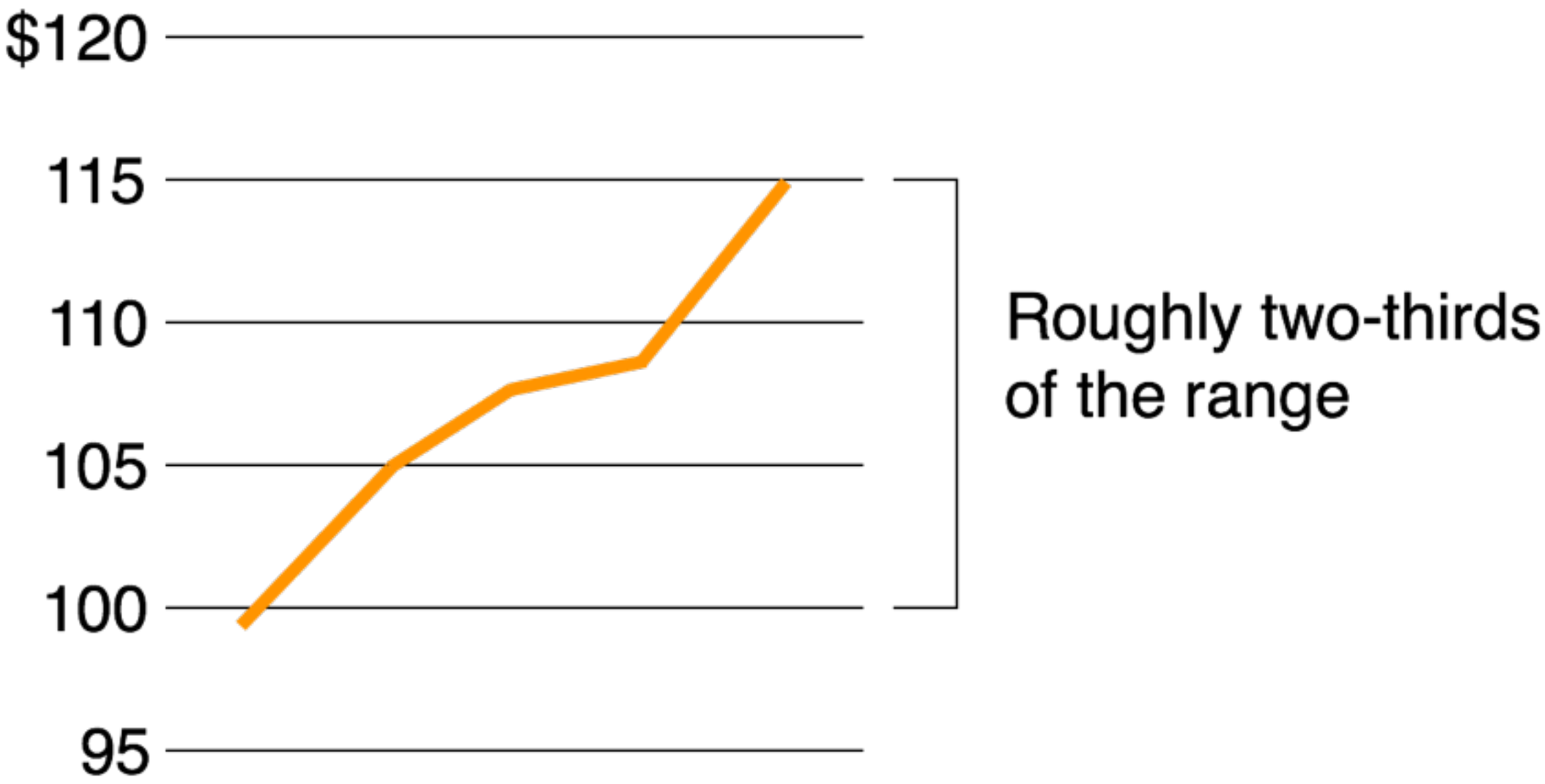


Too exaggerated overstates the trend

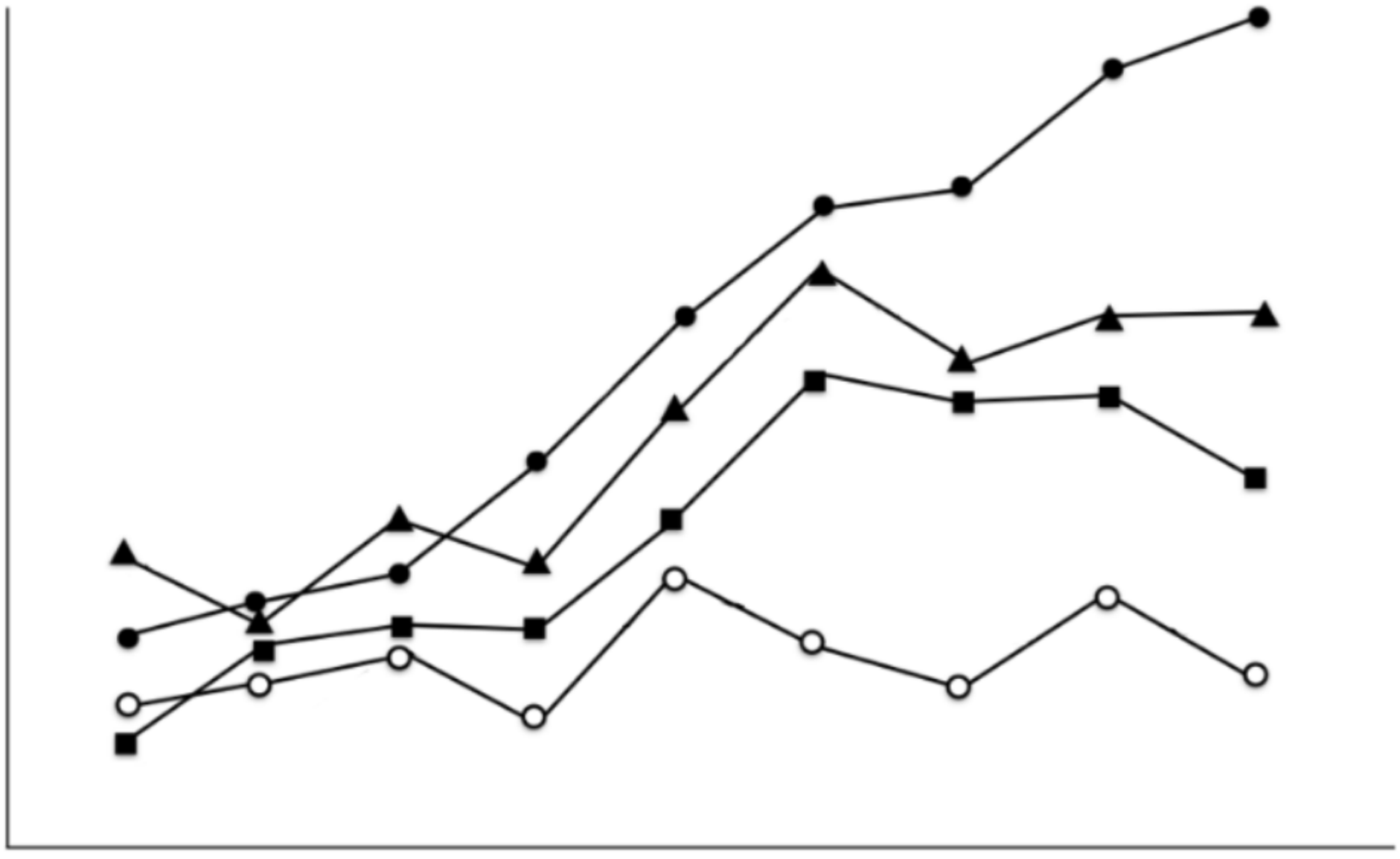


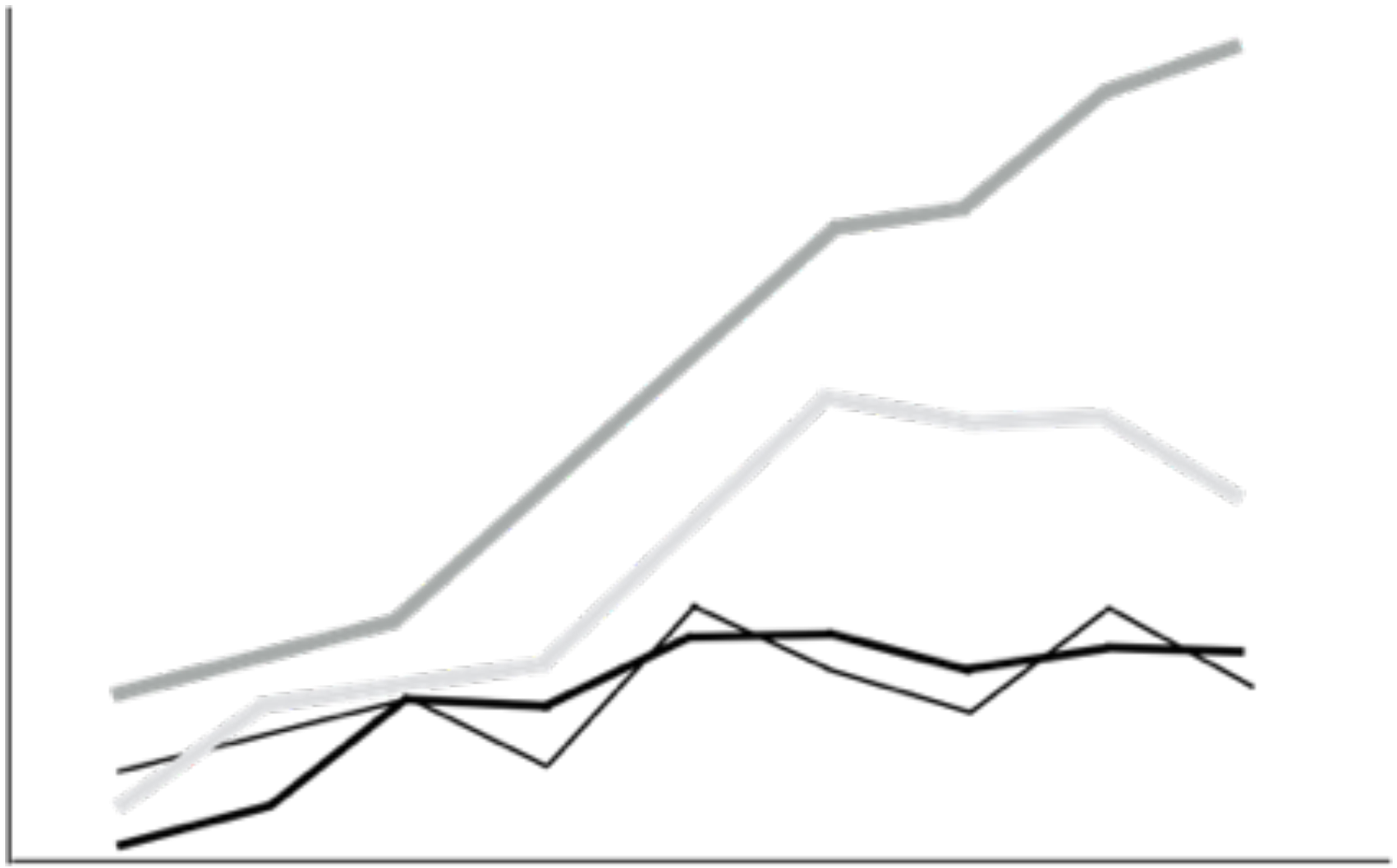
Note y-axis does not need to start at 0.
Why not as bad as in the case of bar chart?

Rule of Thumb



Multiple Patterned Lines in one chart

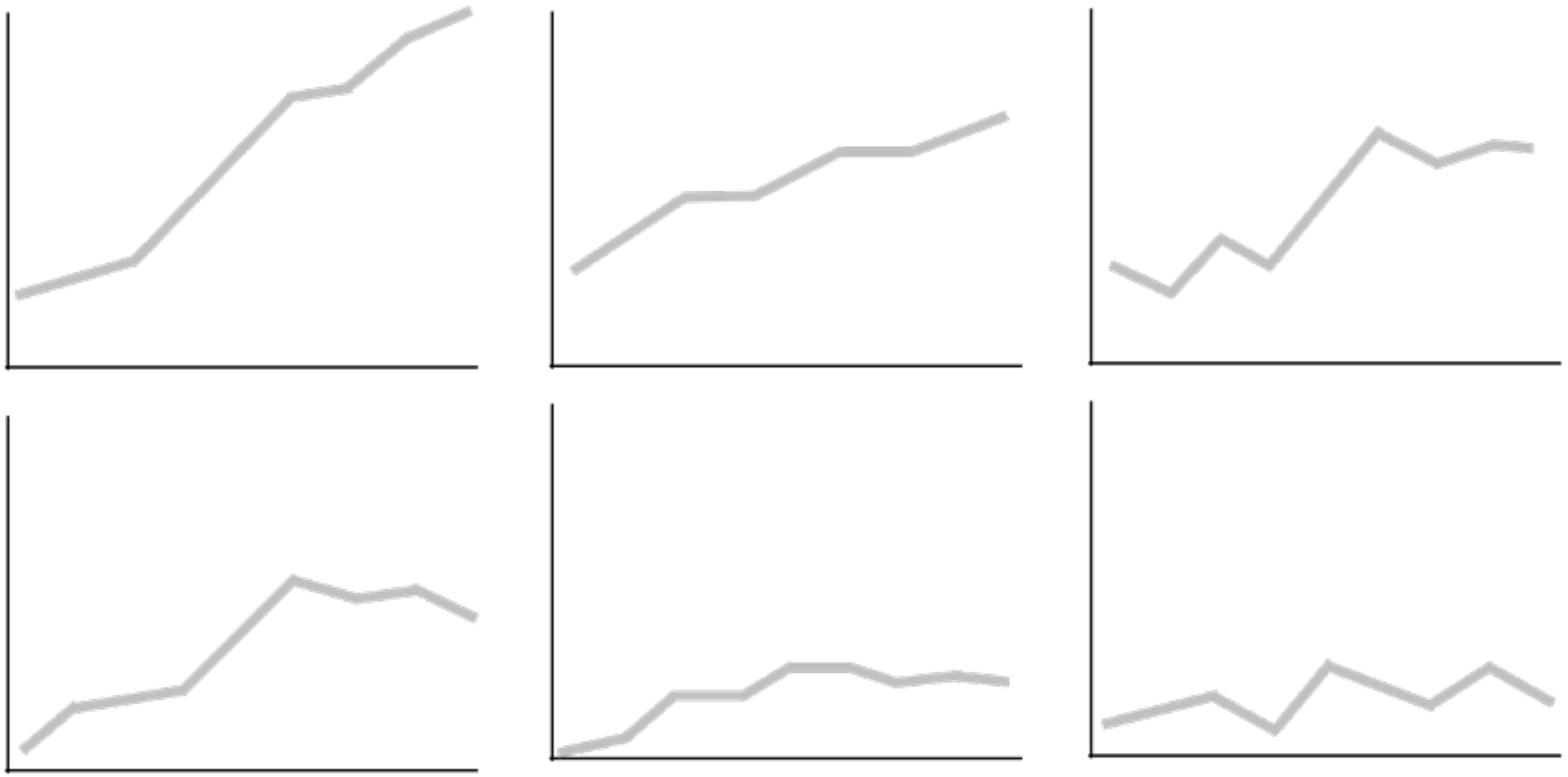




Better?

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”

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 visual aspects to improve?

(Ignore the meaning of the values for now)

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

“Everyone is special” → “No one is”

https://youtu.be/1E9pKU_N15A

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

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

Beautiful Publication-quality LaTeX Tables

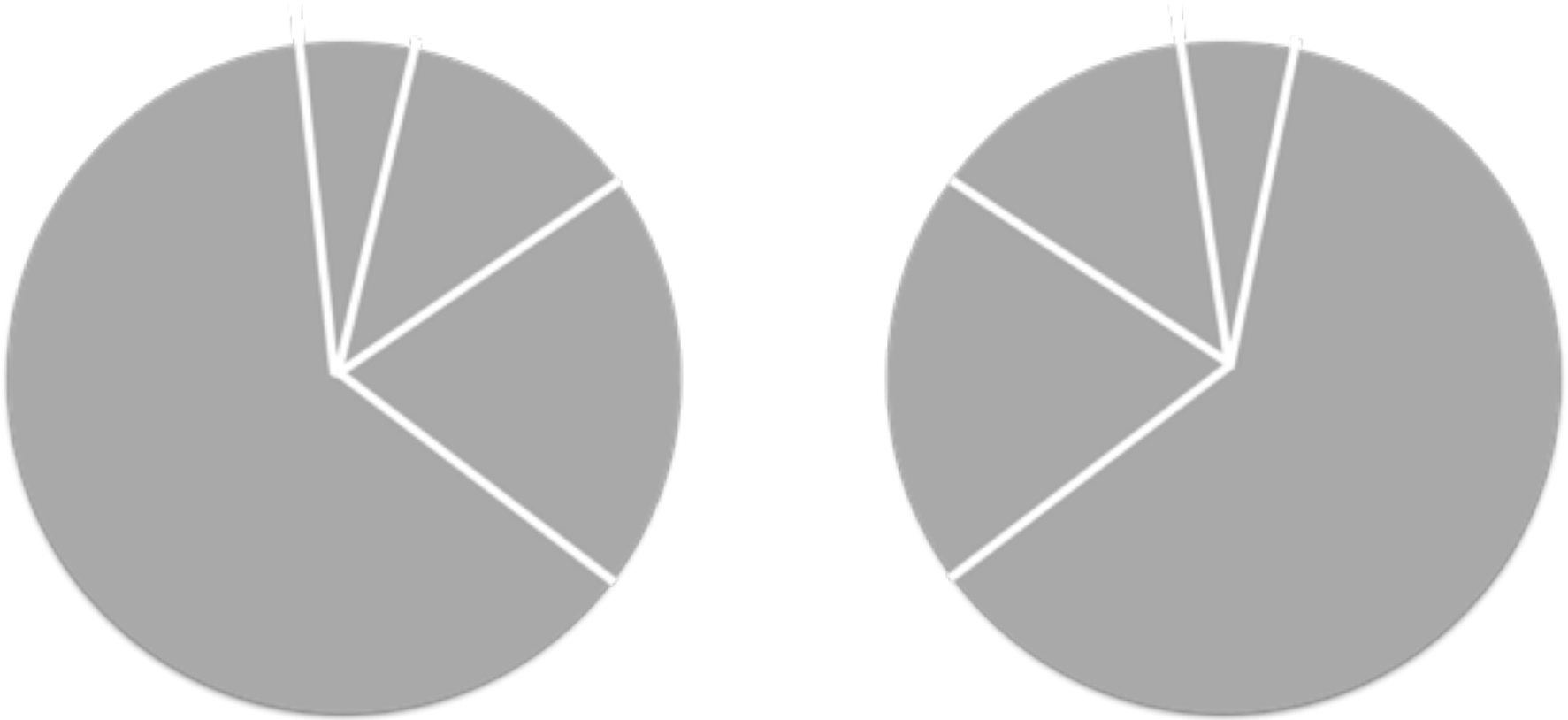
slices	abs. error (%)		abs. error (slices)	
	avg.	max.	avg.	max
< 5000	7.4	73.5	116	625
5000–10000	3.1	27.2	209	1807
10000–15000	2.4	15.6	297	2133
> 15000	1.8	9.0	317	1609

https://tex.stackexchange.com/questions/112343/beautiful-table-samples

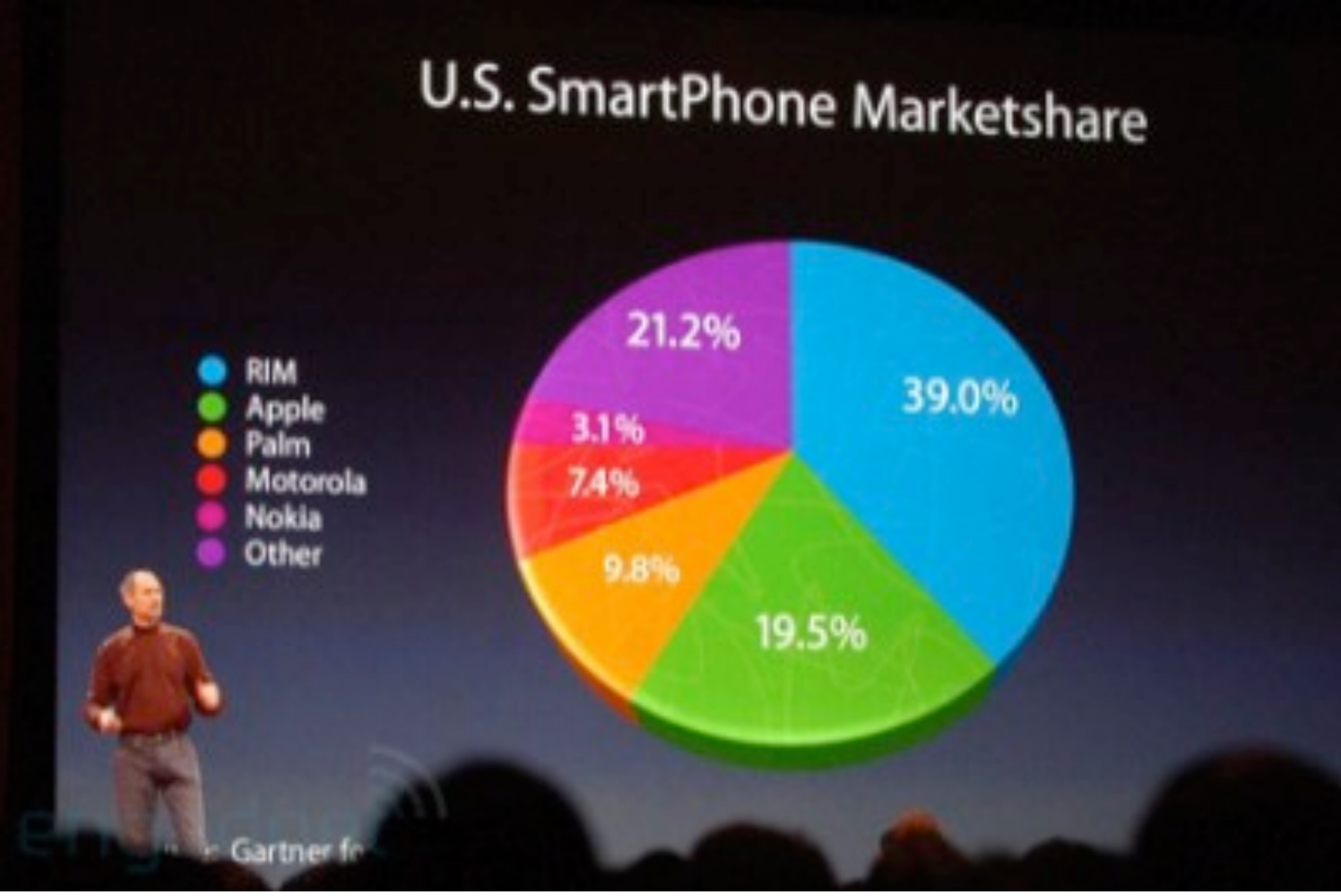
Short guide: https://www.inf.ethz.ch/personal/markusp/teaching/guides/guide-tables.pdf

Long guide: https://metacpan.org/release/LIMAONE/LaTeX-Table-v1.0.6/source/examples/examples.pdf

The Dreaded Pie Charts

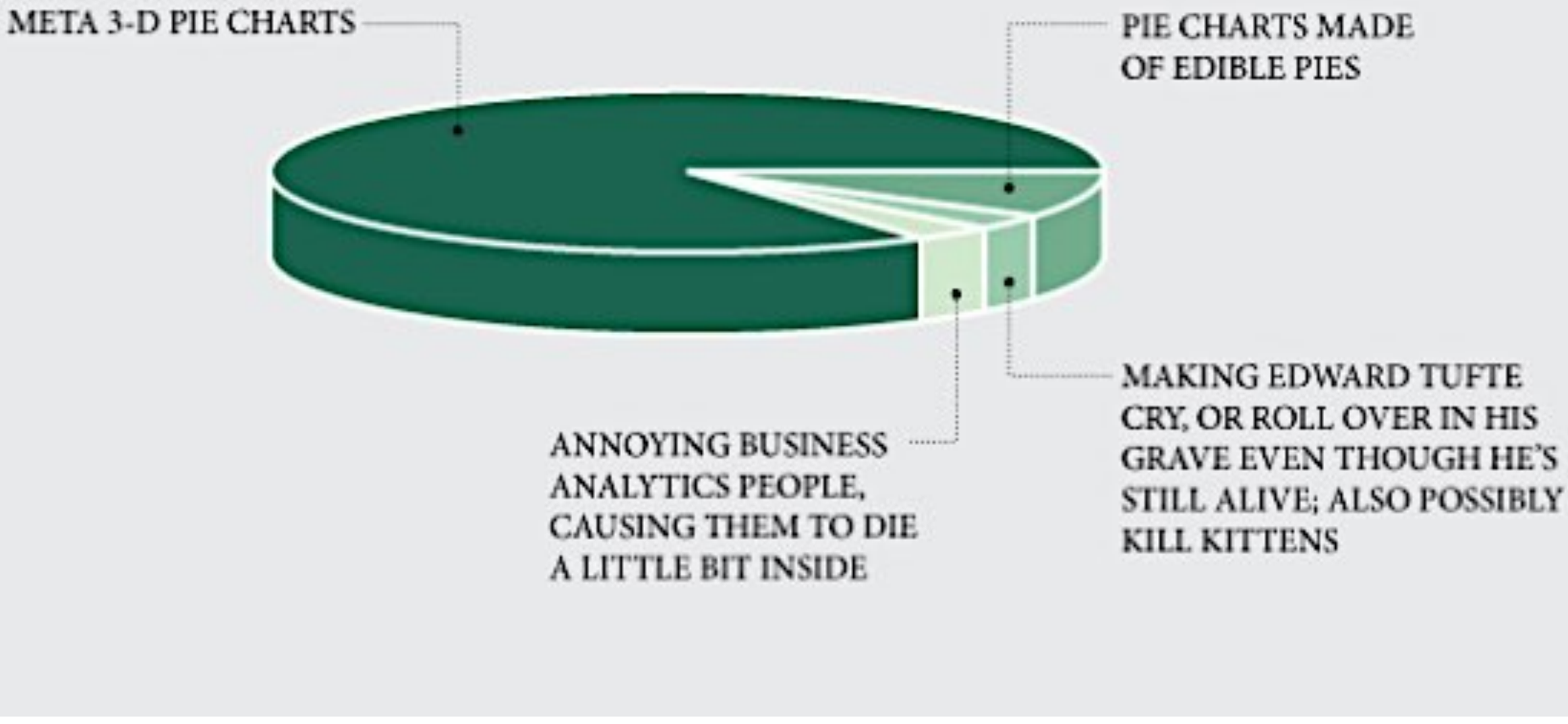


Why people like to use pie charts?

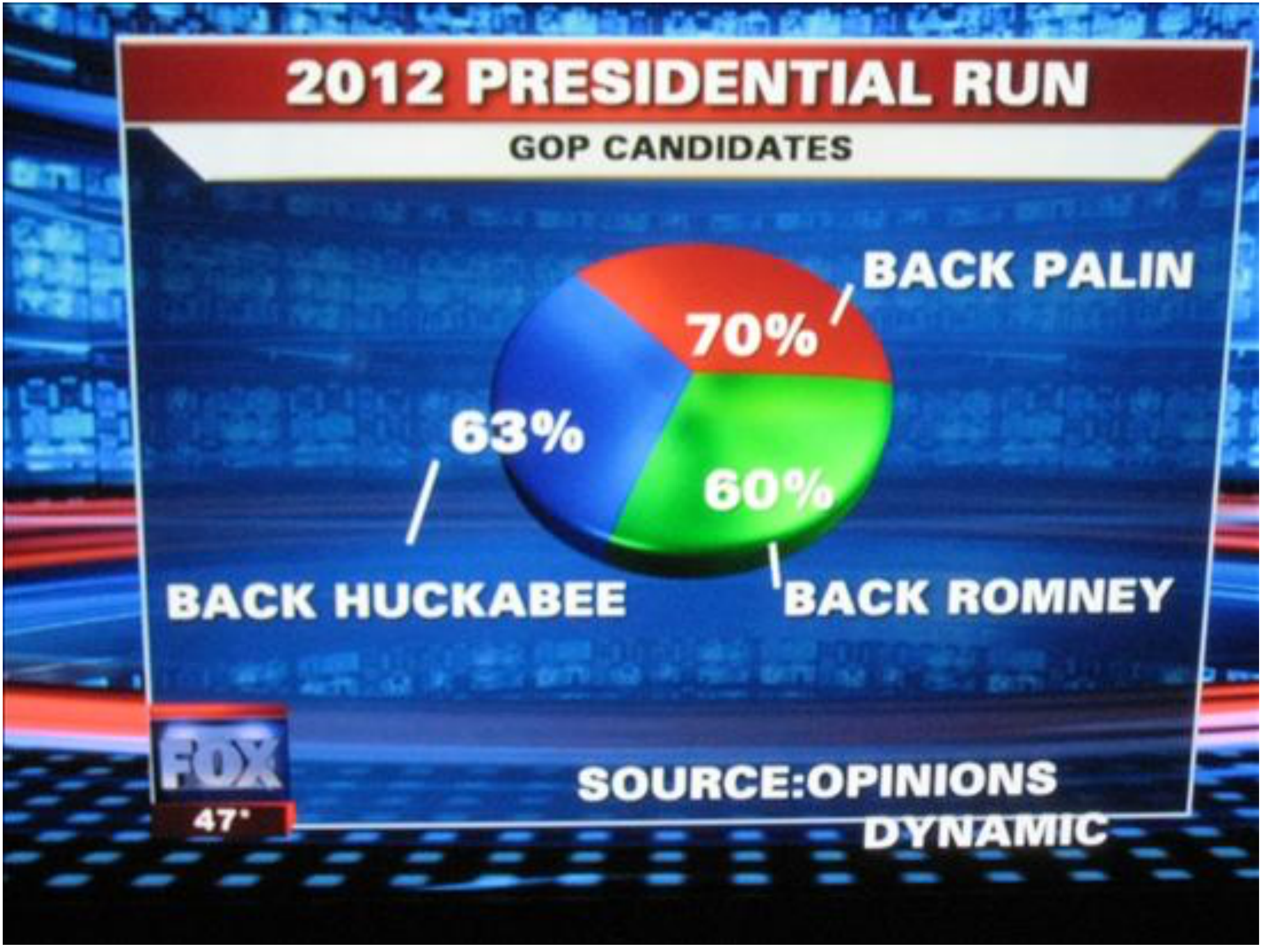


http://www.wired.com/2008/02/macworlds-iphon/

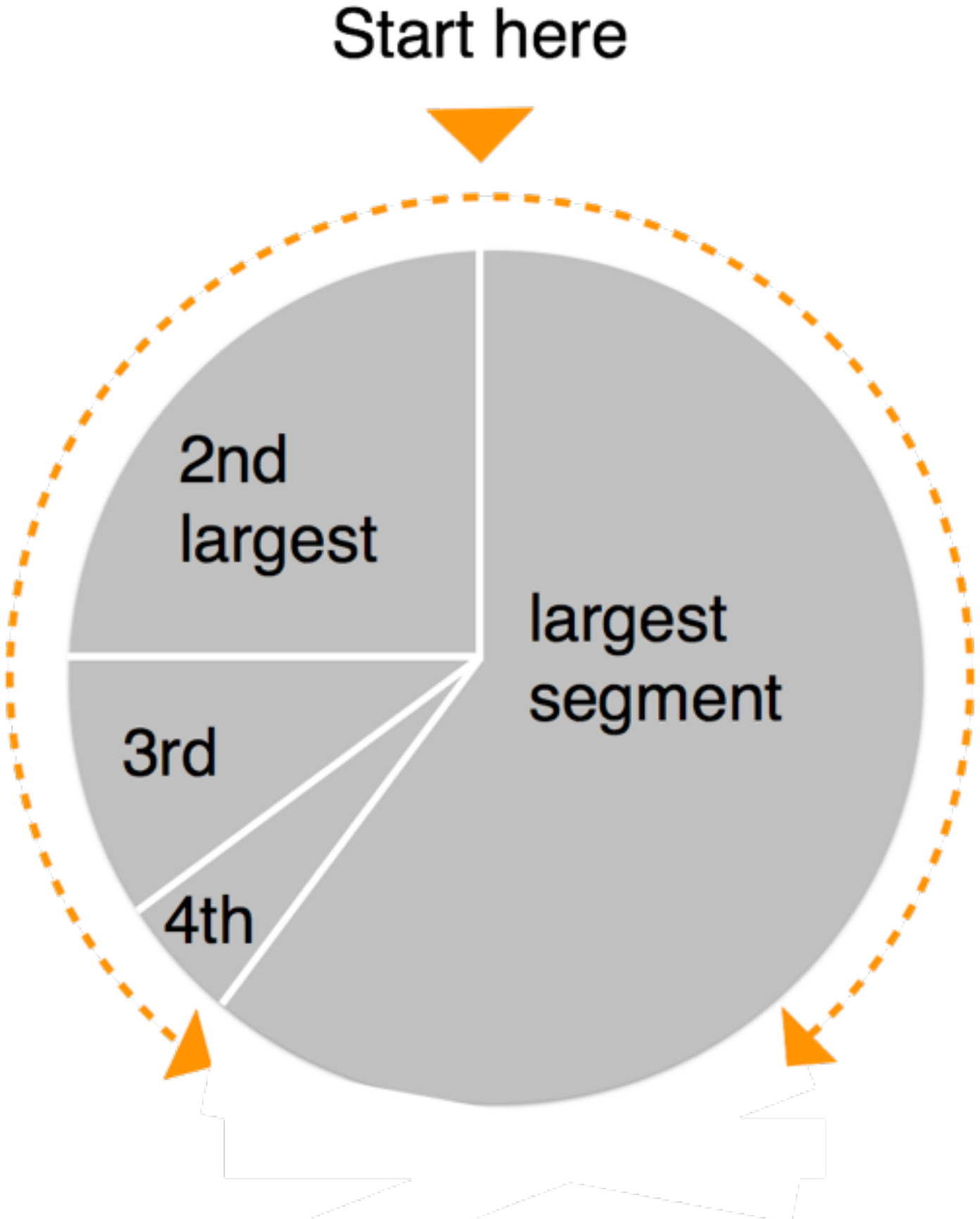
WHAT 3-D PIE CHARTS ARE GOOD FOR.



http://flowingdata.com/2012/06/15/what-3-d-pie-charts-are-good-for/



http://wonkette.com/412361/all-193-of-republicans-support-palin-romney-and-huckabee


Start here
2nd largest
largest segment
3rd
4th

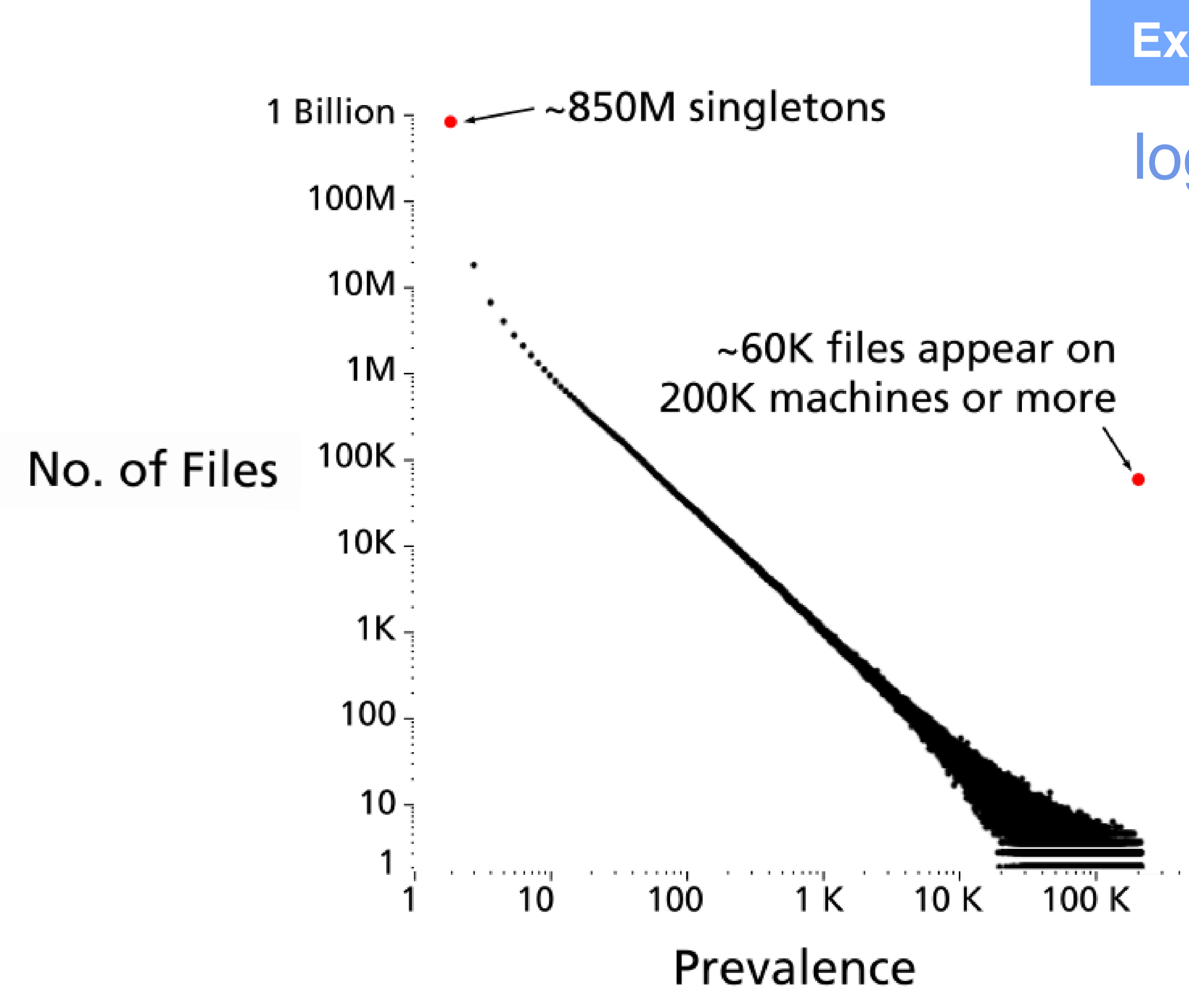
http://infosthetics.com/archives/2008/09/funniest_pie_chart_ever.html

Log scale instead of linear scale

Include numbers from different orders of magnitude

Example

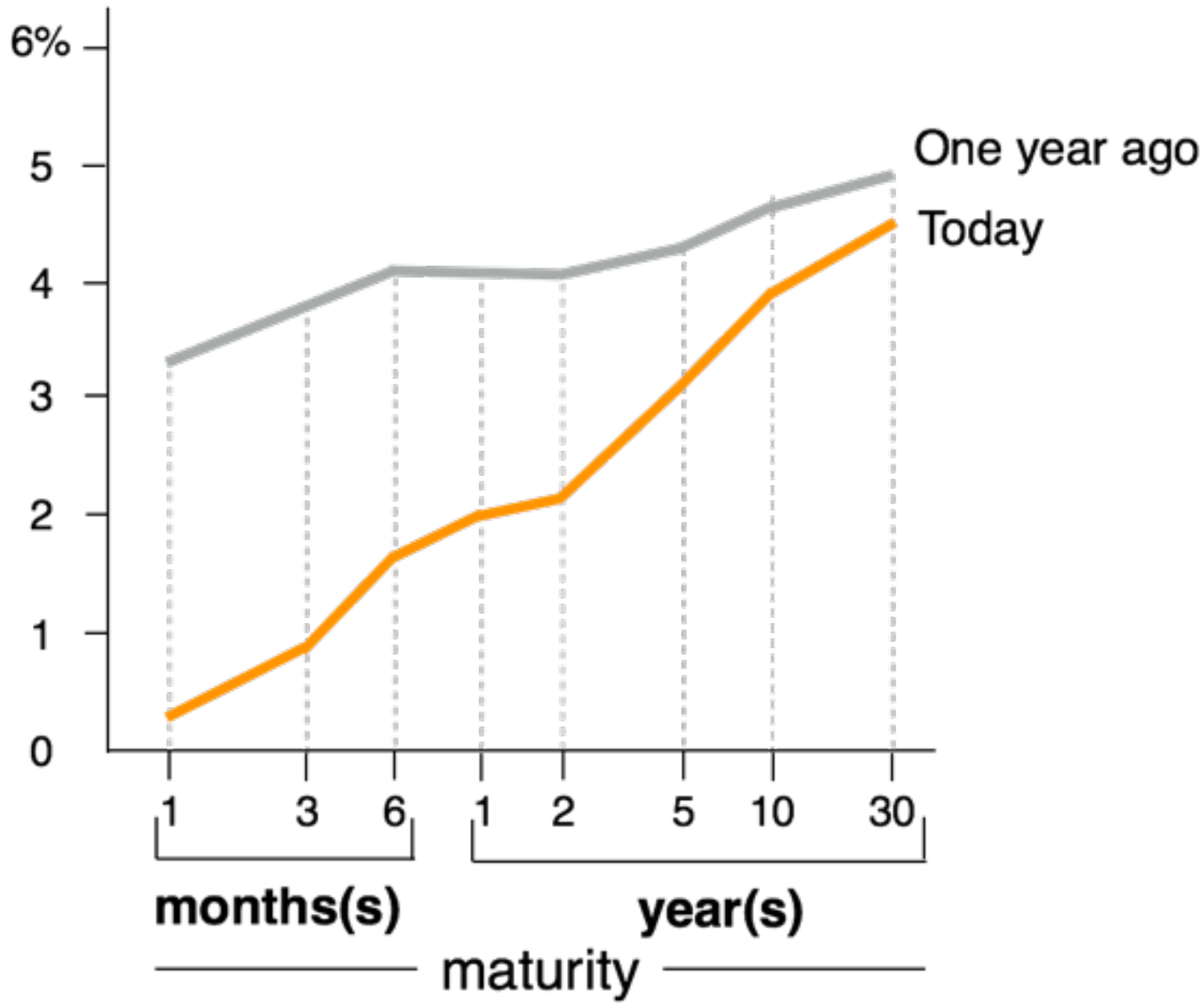
log-log



“log” also works well for time

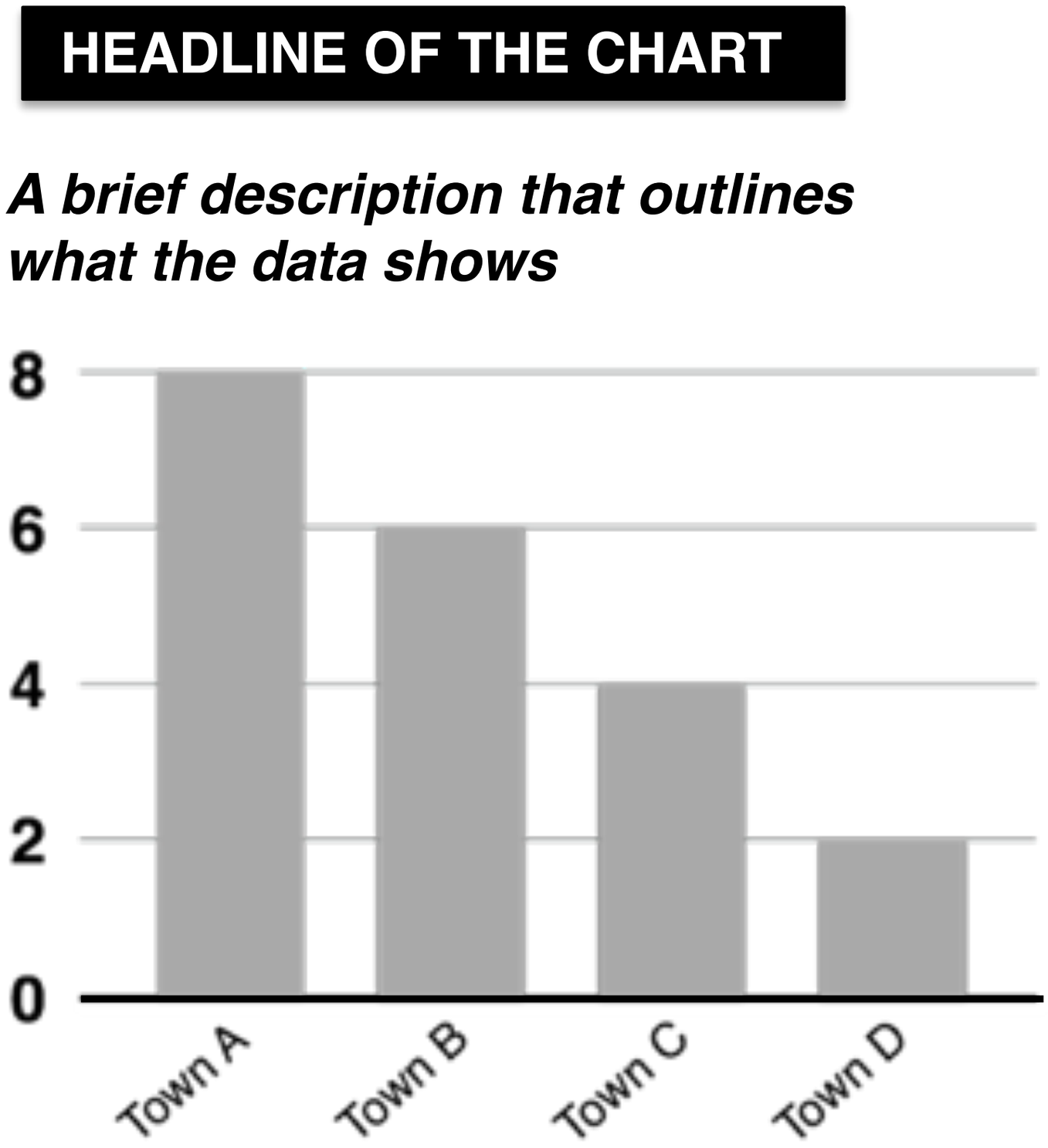
Example

The yield curve of Treasury bills, notes and bonds



Quick Exercises

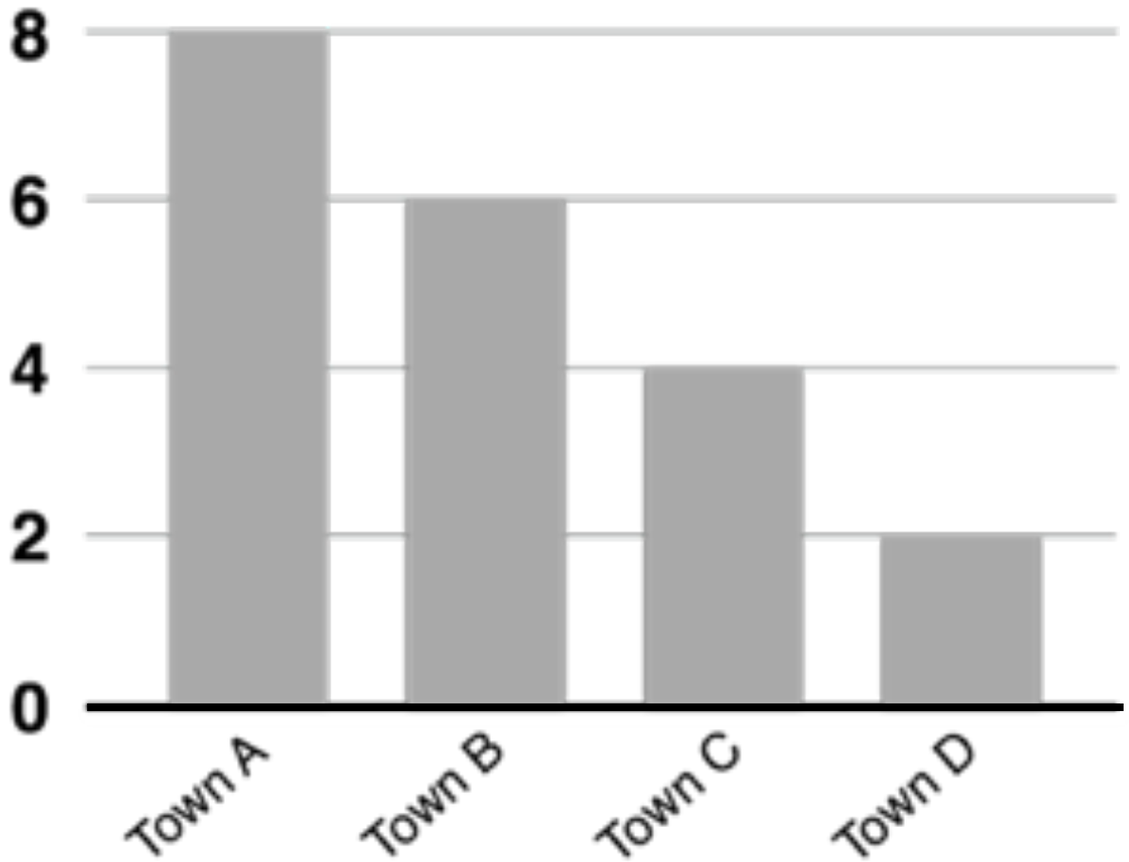
Applying what you have just learned.



Can you improve its visual design?

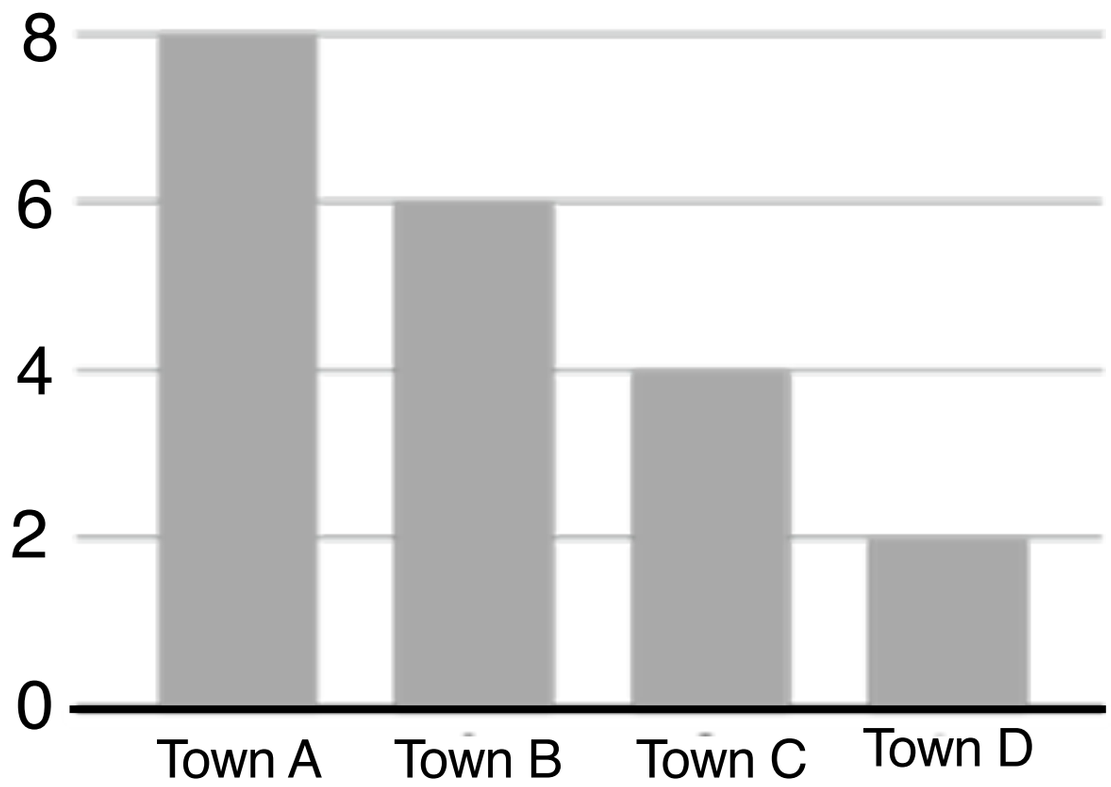
HEADLINE OF THE CHART

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?

How to fix the defaults

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

Role	Name	Year of the...	Debut	Number of Fans	Takedown Rate
Face (The Hero)	The Ultimate Warrior	Tiger	May-2011	97320.00	86.2
Face (The Hero)	Hulk Hogan	Oxen	Jan-2008	988551.00	61.978
Face (The Hero)	Macho Man Randy Savage	Monkey	Feb-2008	157618.00	59.29
Face (The Hero)	Hacksaw Jim Duggan	Pig	Mar-2008	30300.00	53.4332
Face (The Hero)	Superfly Jimmy Snuka	Dragon	Mar-2008	12341.00	52.7
Heel (The Bad Guy)	Rowdy Roddy Piper	Rooster	Jun-1968	71645.00	45.4
Heel (The Bad Guy)	The Million Dollar Man Ted DiBiase	Rat	Apr-1975	449342.00	43.7689
Heel (The Bad Guy)	Mr. Perfect Curt Henning	Rat	May-1980	13773.00	38
Heel (The Bad Guy)	Jake the Snake Roberts	Snake	Jul-1975	5609.00	37.99
Jobber (The Unknown)	Brad Smith	Sheep	Aug-2008	1103.00	36.316
Jobber (The Unknown)	Ted Duncan	Sheep	Aug-2008	200.00	33.61
Jobber (The Unknown)	Joey the Uber Nerd Cherdarchuk	Snake	Aug-2008	5.00	21.0196

How to fix the defaults

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

Role	Name	Year of the...	Debut	Thousands of Fans	Takedown Rate
Face (The Hero)	The Ultimate Warrior	Tiger	May-2011	97.3	86.2
	Hulk Hogan	Oxen	Jan-2008	988.6	62.0
	Macho Man Randy Savage	**Monkey**	**Feb-2008**	**157.6**	**59.3**
	Hacksaw Jim Duggan	Pig	Mar-2008	30.3	53.4
	Superfly Jimmy Snuka	Dragon	Mar-2008	12.3	52.7
Heel (The Bad Guy)	Rowdy Roddy Piper	Rooster	Jun-1968	71.6	45.4
	The Million Dollar Man Ted DiBiase	Rat	Apr-1975	449.3	43.8
	Mr. Perfect Curt Henning	Rat	May-1980	13.8	38.0
	Jake the Snake Roberts	Snake	Jul-1975	5.6	38.0
Jobber (The Unknown)	Brad Smith	Sheep	Aug-2008	1.1	36.3
	Ted Duncan	Sheep	Aug-2008	0.2	33.6
	Joey the Uber Nerd Cherdarchuk	Snake	Aug-2008	0.0	21.0

Practitioners' Guide

Colors: start with black & white, then carefully add colors — forces you to focus on content and layout

Fonts: **sans-serif** generally easier to read

(On Mac: Helvetica is great start)

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!

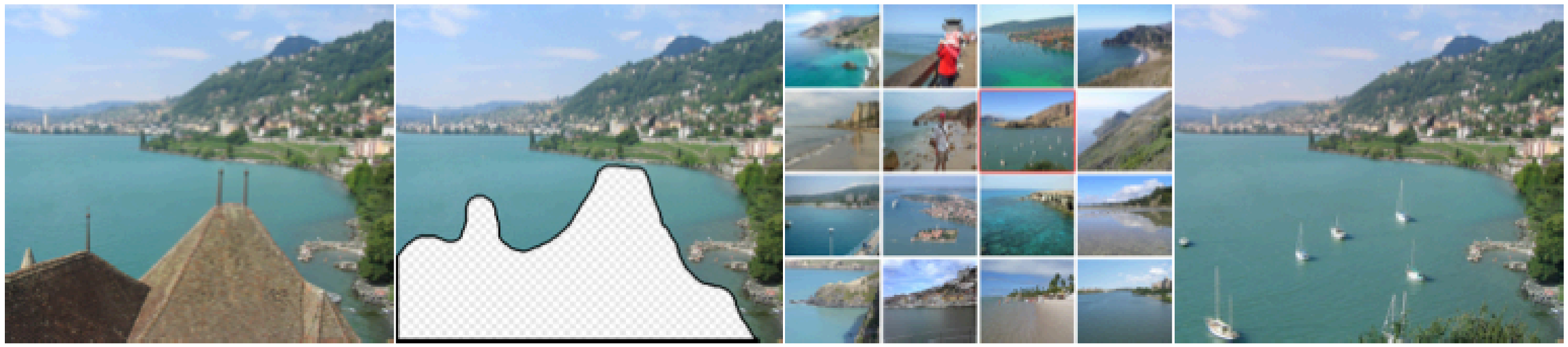
Once people fall asleep, it's hard to wake them up!
If you have good stuff, show them now.

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

Example

TRANSFORMER EXPLAINER: Interactive Learning of Text-Generative Models

Aeree Cho[*1], Grace C. Kim[*1], Alexander Karpekov[*1],
Alec Helbling[1], Zijie J. Wang[1], Seongmin Lee[1], Benjamin Hoover[1,2], Duen Horng (Polo) Chau[1]

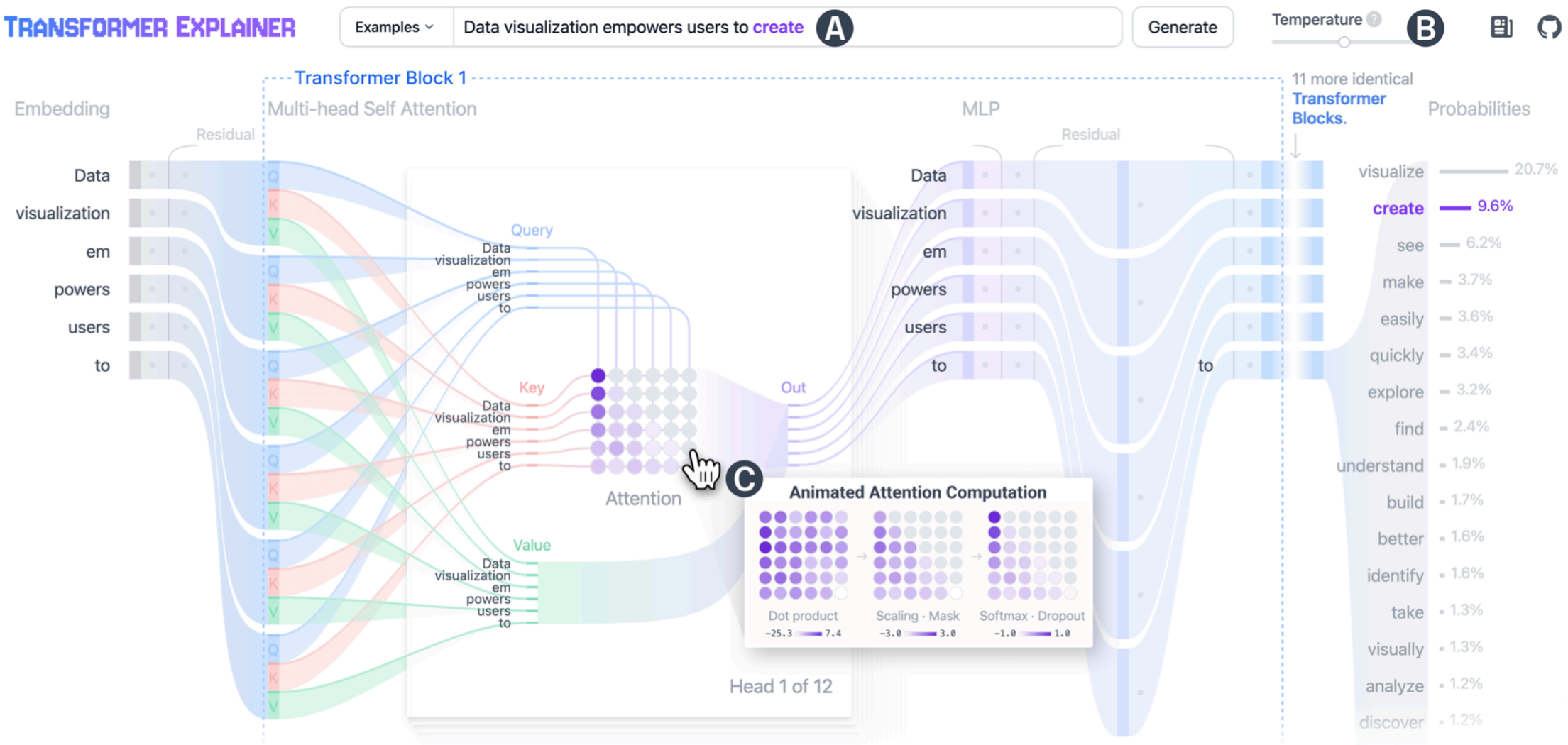


Figure 1: TRANSFORMER EXPLAINER helps users (A) visually examine how a text-generative Transformer model (GPT-2) transforms input text to predict next tokens, (B) interactively experiment in real time with key model parameters like *temperature* to understand prediction determinism, and (C) transition seamlessly between abstraction levels to visualize the interplay between low-level mathematical operations and high-level model structures.

ABSTRACT

understanding and engagement [12]. Demystifying this architecture

Practitioners' Guide:
Additional Tips for 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 figure shown later

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

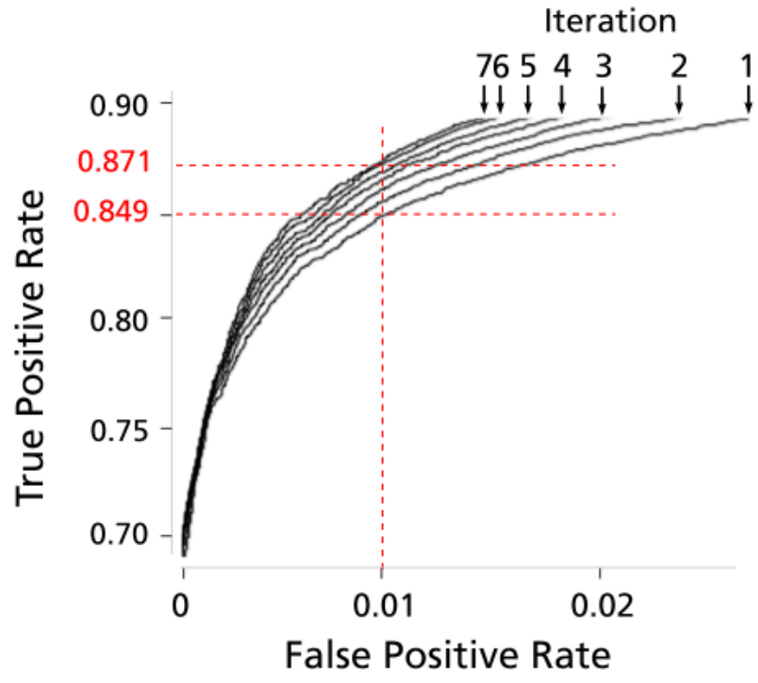


Figure 8: ROC curves of 7 iterations; true positive rate incrementally improves.

Example

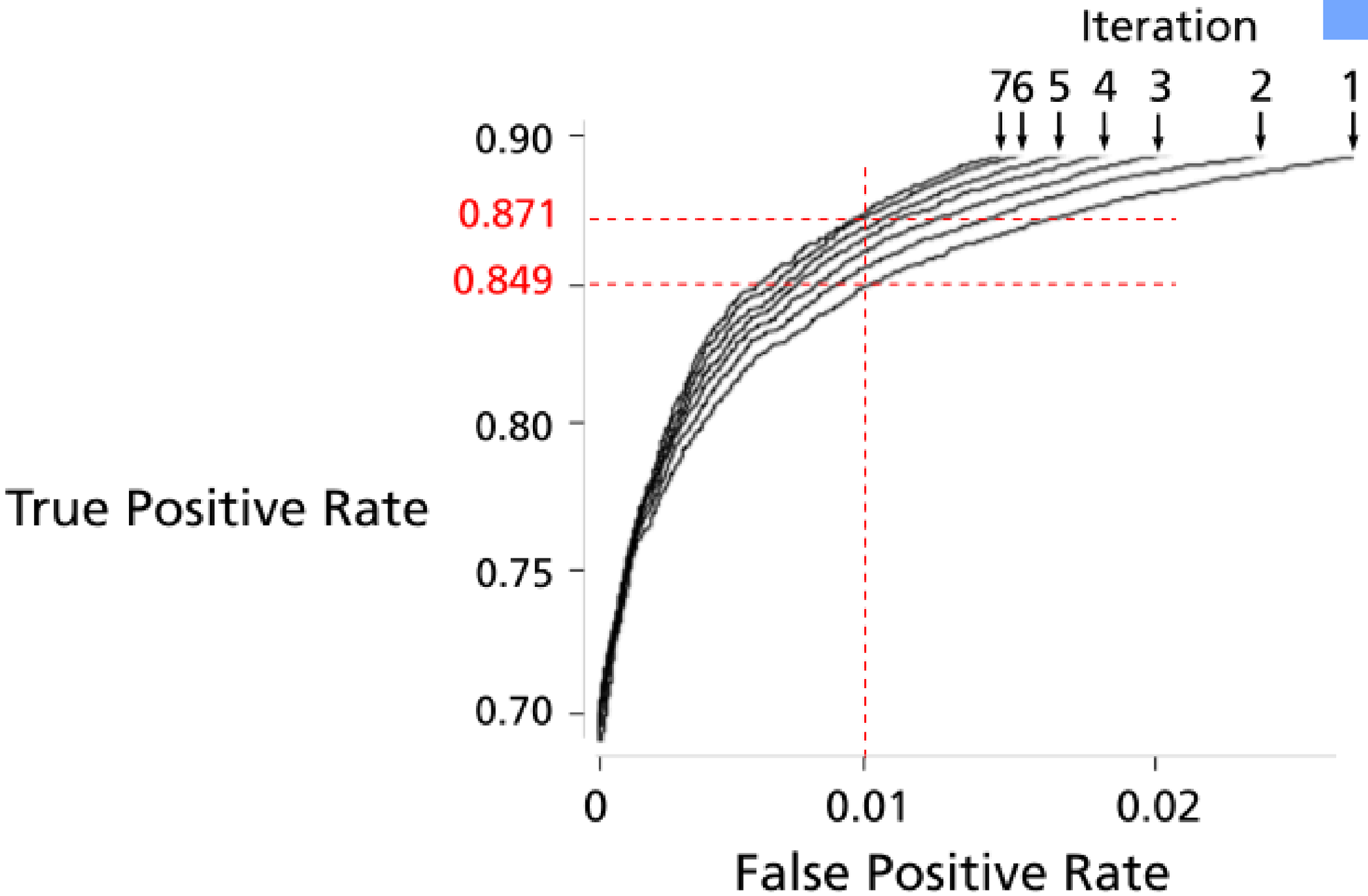


Figure 8: ROC curves of 7 iterations; true positive rate incrementally improves.

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

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.

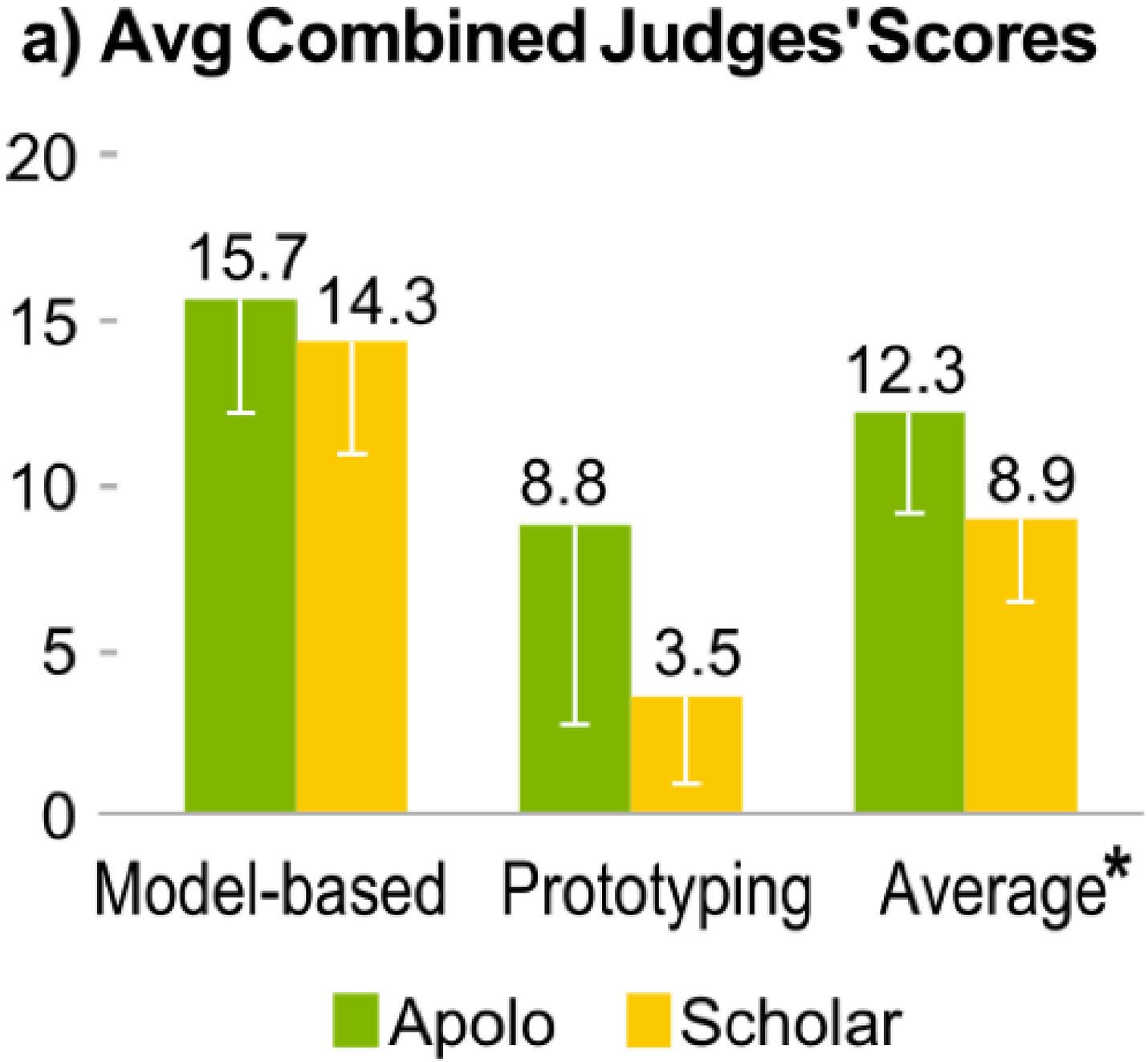
Redesign figures for presentation

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

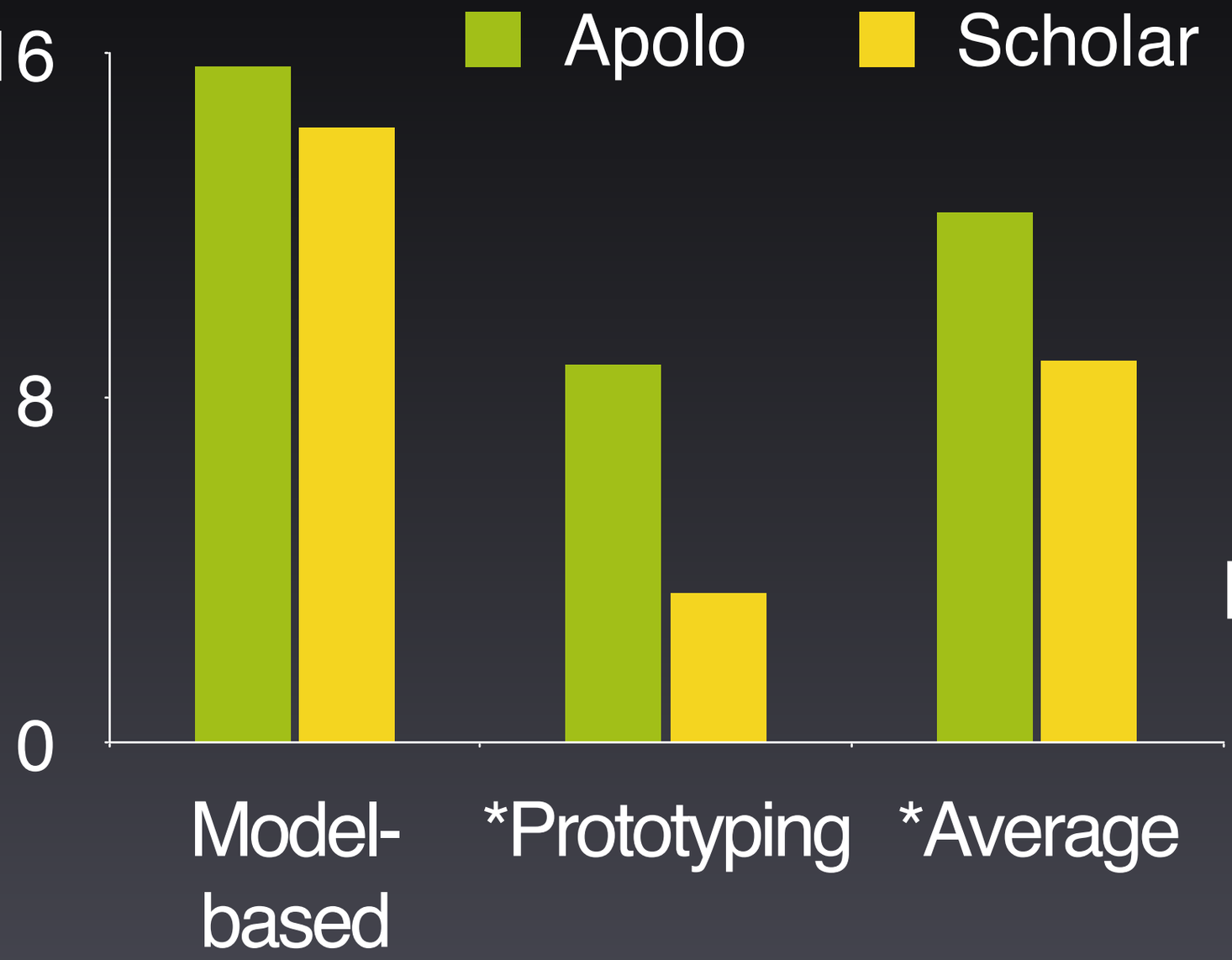
In Paper



Example

In Talk

Judges' Scores



Higher is better.
Apolo wins.

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

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