

Unless otherwise noted, the content of this course material is licensed under a Creative Commons Attribution – Non-commercial – Share Alike 3.0 License.
<http://creativecommons.org/licenses/by-nc-sa/3.0/>.

Copyright © 2005-2007, Jeffrey K. MacKie-Mason.

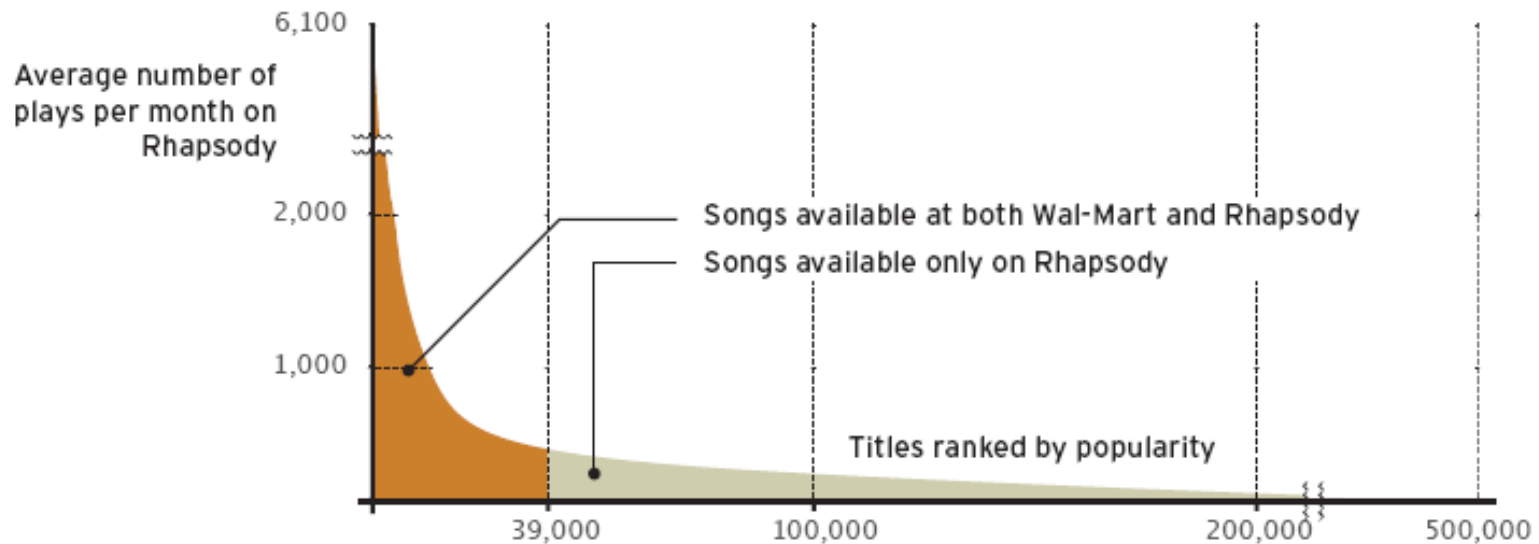
You assume all responsibility for use and potential liability associated with any use of the material. Material contains copyrighted content, used in accordance with U.S. law. Copyright holders of content included in this material should contact open.michigan@umich.edu with any questions, corrections, or clarifications regarding the use of content. The Regents of the University of Michigan do not license the use of third party content posted to this site unless such a license is specifically granted in connection with particular content. Users of content are responsible for their compliance with applicable law. Mention of specific products in this material solely represents the opinion of the speaker and does not represent an endorsement by the University of Michigan. For more information about how to cite these materials visit <http://michigan.educommons.net/about/terms-of-use>.

Any medical information in this material is intended to inform and educate and is not a tool for self-diagnosis or a replacement for medical evaluation, advice, diagnosis or treatment by a healthcare professional. You should speak to your physician or make an appointment to be seen if you have questions or concerns about this information or your medical condition. Viewer discretion is advised: Material may contain medical images that may be disturbing to some viewers.

The Long Tail

Jeff MacKie-Mason

SI 646

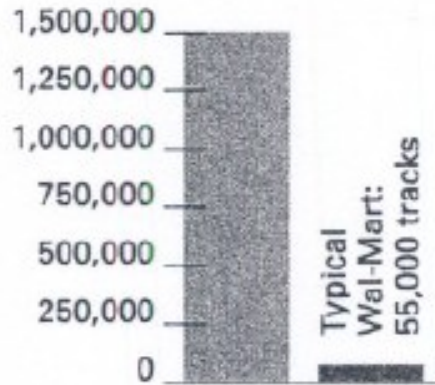


Rhapsody: More than 2 million tracks
 40% of revenue from tracks not at Wal-Mart

The new growth market: Products you can't find anywhere but online

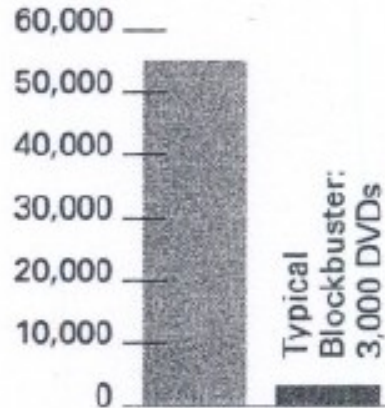
Rhapsody

Total inventory:
1.5 million tracks



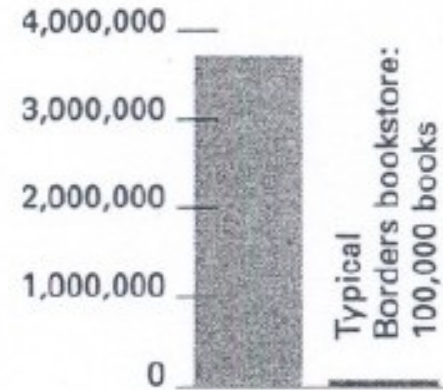
Netflix

Total inventory:
55,000 DVD titles

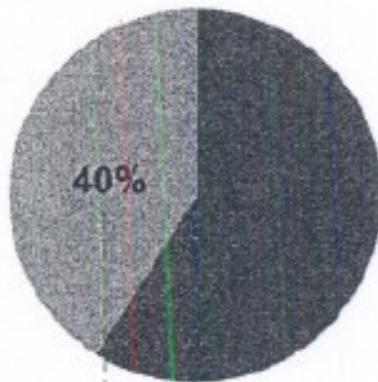


Amazon

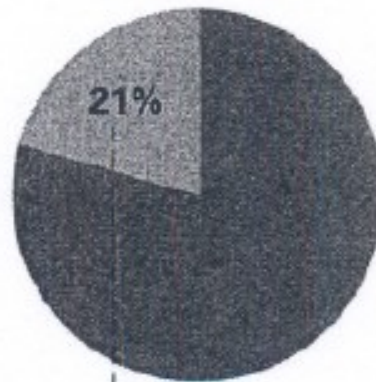
Total inventory:
3.7 million book titles



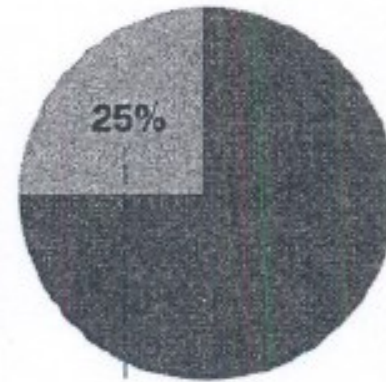
Total sales



Total sales

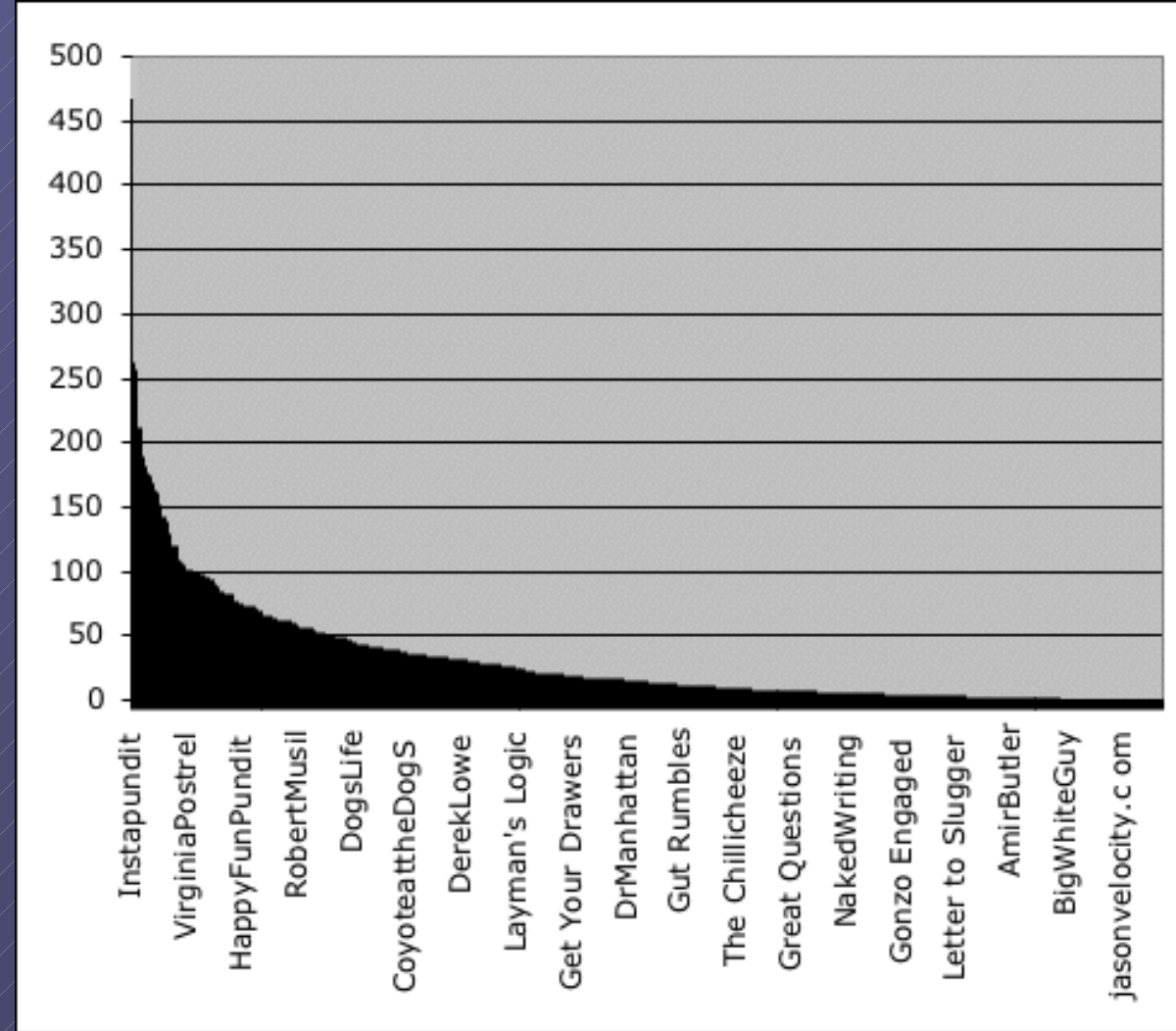


Total sales



Products not available in offline retail stores

433 blogs arranged
in rank order by
number of inbound
links



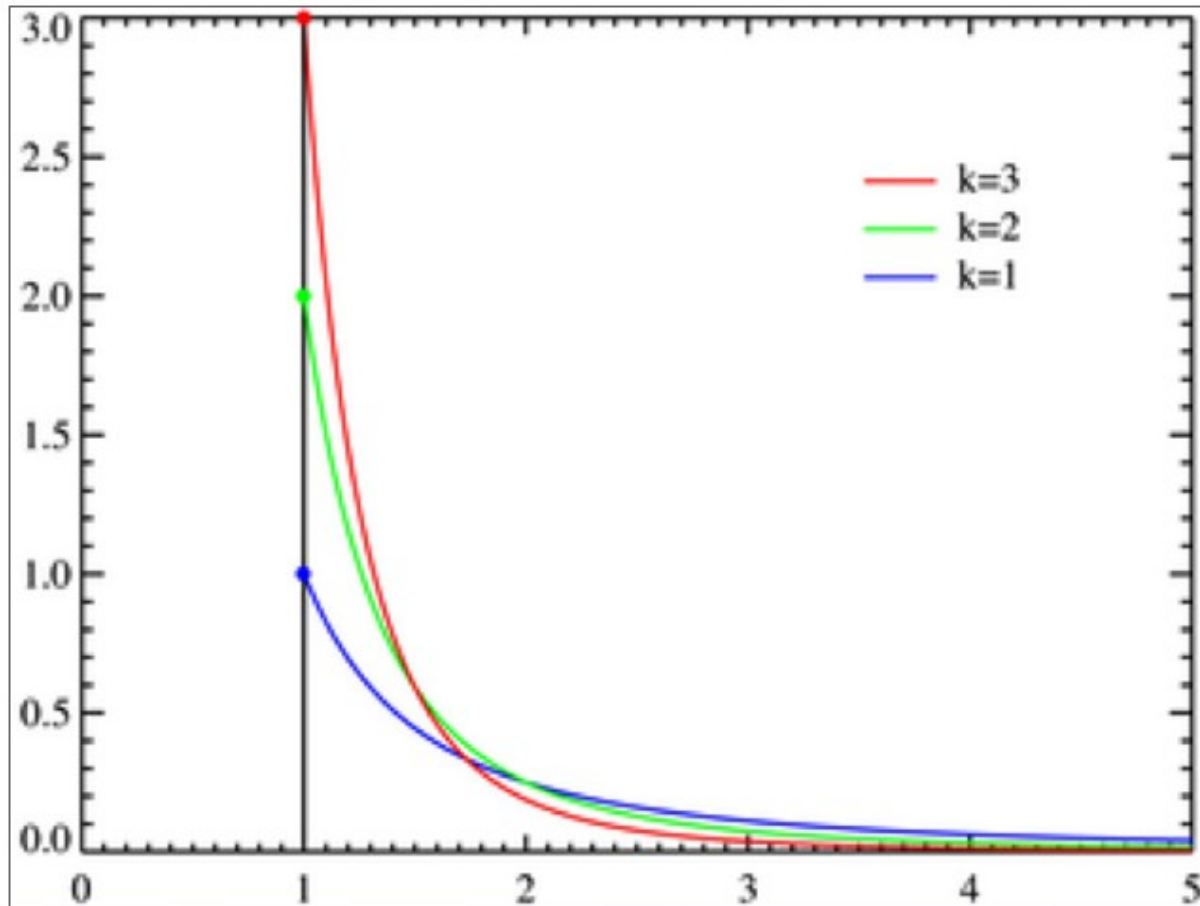
The data are drawn from N.Z Bear's 2002 work on the [blogosphere ecosystem](http://www.myelin.co.nz/ecosystem/).
The current version of this project can now be found at <http://www.myelin.co.nz/ecosystem/>.

Pareto distribution:

What is the probability that a person has wealth x ?

$$f(x; k, x_m) = k \frac{x_m^k}{x^{k+1}} \text{ for } x \geq x_m.$$

f(x)

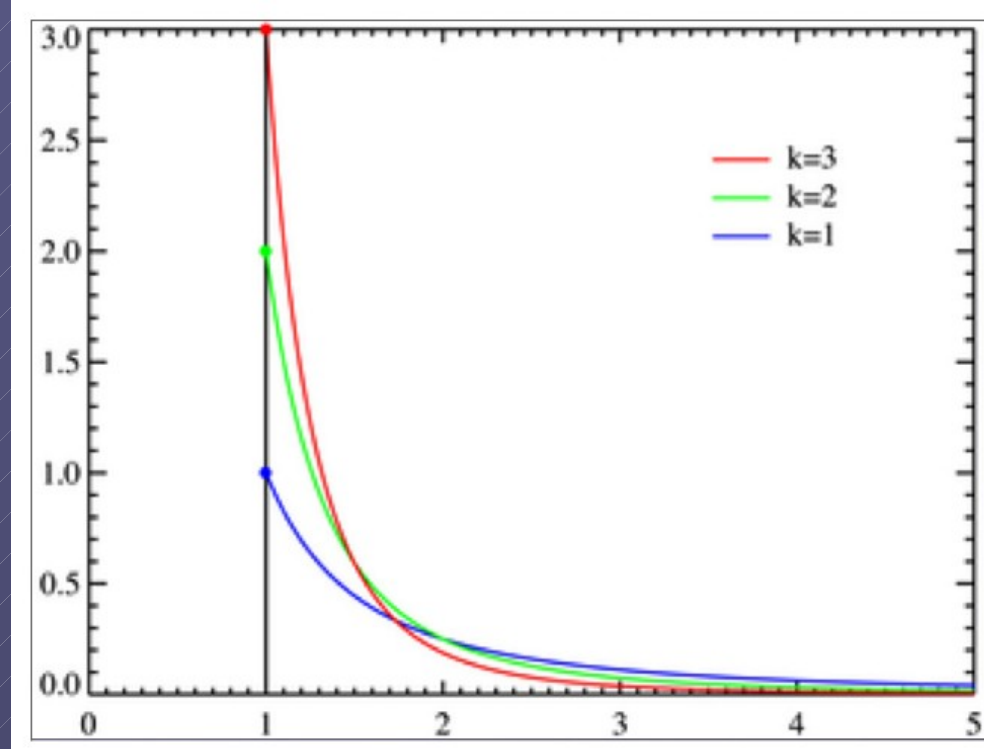


k constant shape
parameter > 0

x_m constant scale
parameter

x

Pareto distribution:
Most people have low
wealth, few have
very high



Source: Undetermined

Examples:

- Human settlement sizes

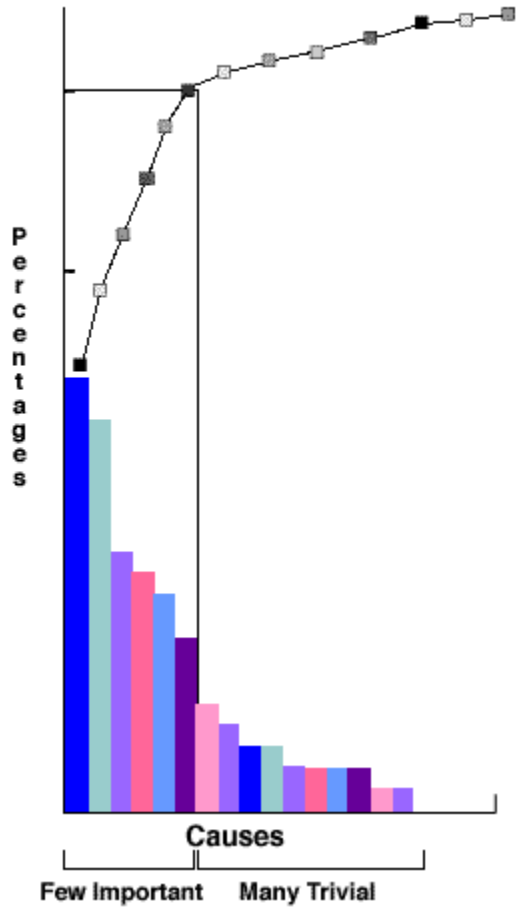
- File sizes transferred over Internet

- Sizes of oil fields

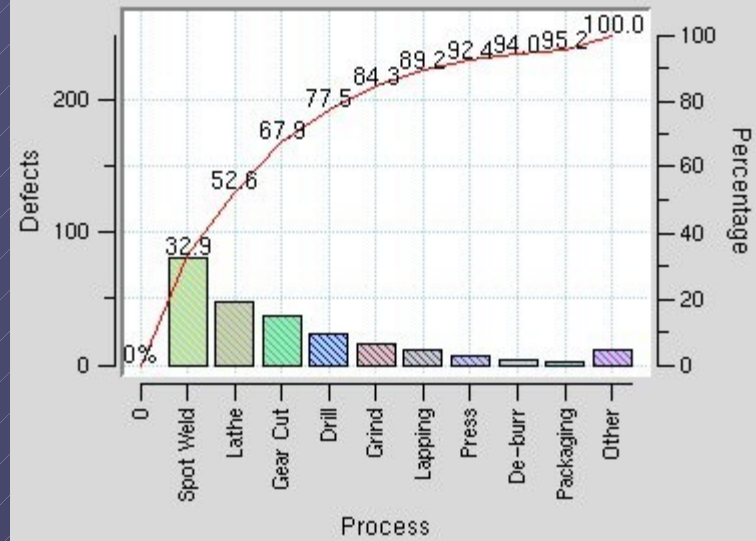
- Rates of return on corporate equities (stocks)

- Areas burnt in forest

**Causes for Delay of Care
Azoques Health Center, 1994**



Defects Found During Inspection

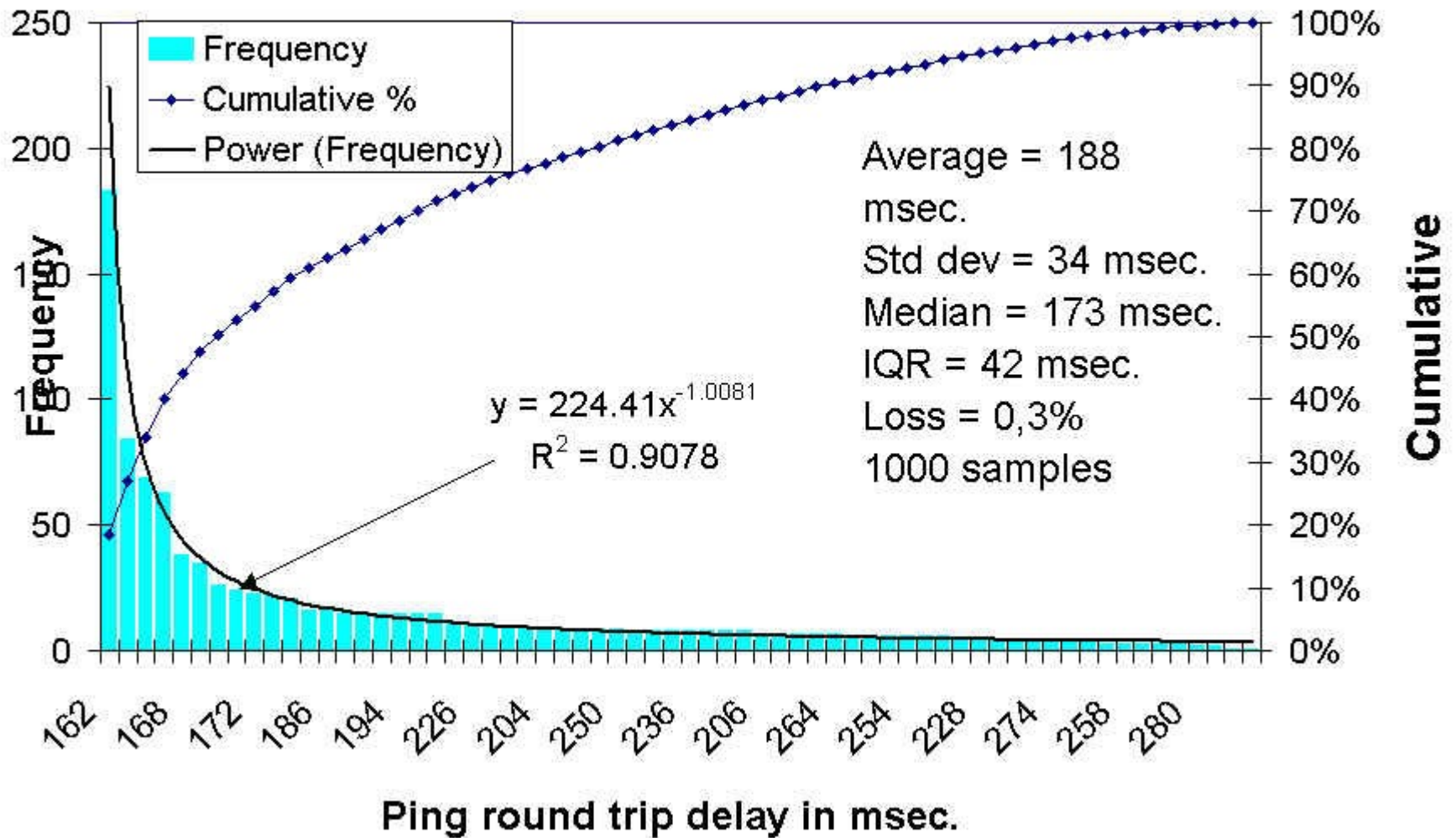


Sources:

Graph on left: <http://erc.msh.org/quality/pstools/pspareto.cfm>

Graph on right: <http://tcltk.free.fr/blt/>

SLAC<=>CERN round trip delay Pareto histogram



Zipf's law: "In a corpus of natural language utterances, the frequency of any word is roughly inversely proportional to its rank in the frequency table"

More generally, the size of the r 'th largest occurrence of the event is inversely proportional to its rank:

$$y = a r^b$$

with b close to unity

In the "Brown" corpus,

- *the* accounts for 7% = .07/1
- *of* for 3.5% = .07/2
- *and* for 2.8% = .07/2.5

The first 135 words account
for 50%

Pareto & Zipf are both
examples of a power law:

$$y = a x^k$$

Take logs of both sides: $\log y = \log a + k \log x$
which is linear: $z = a + b v$

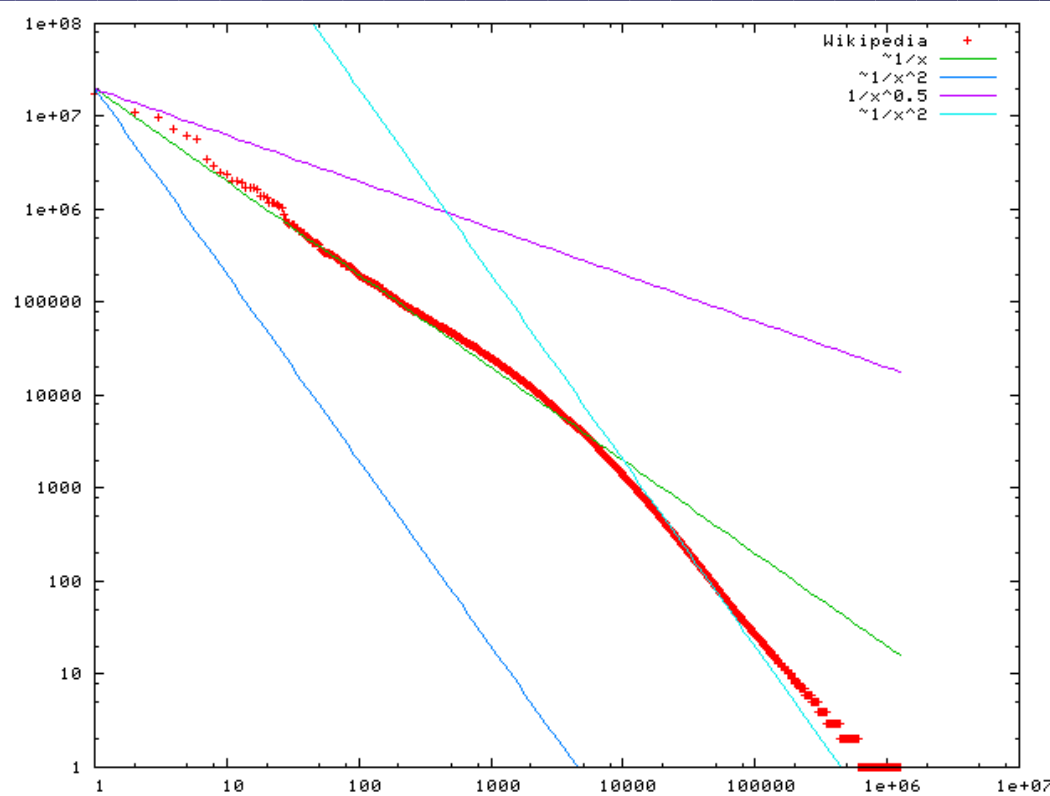
A plot of word frequency in Wikipedia (November 27, 2006).

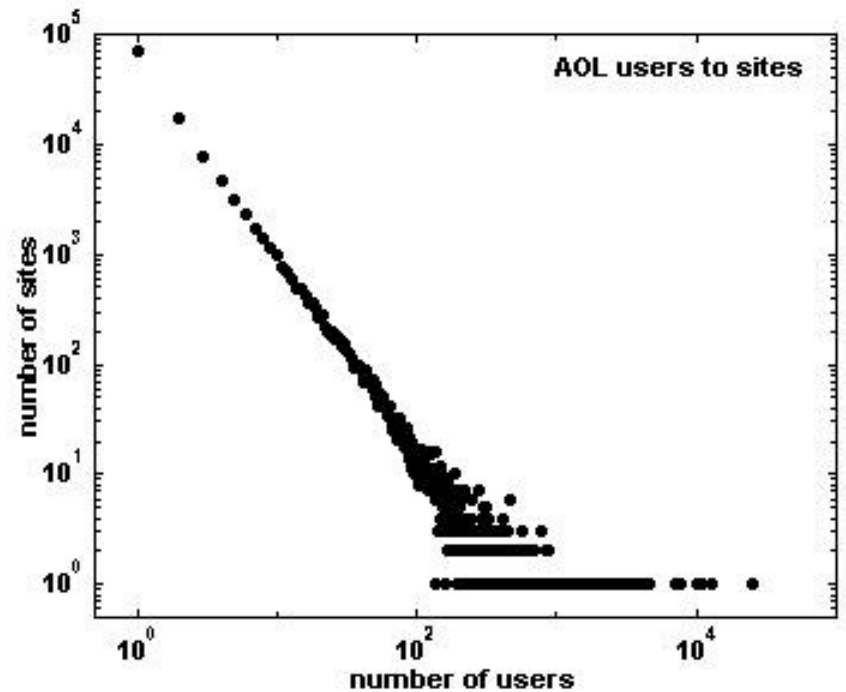
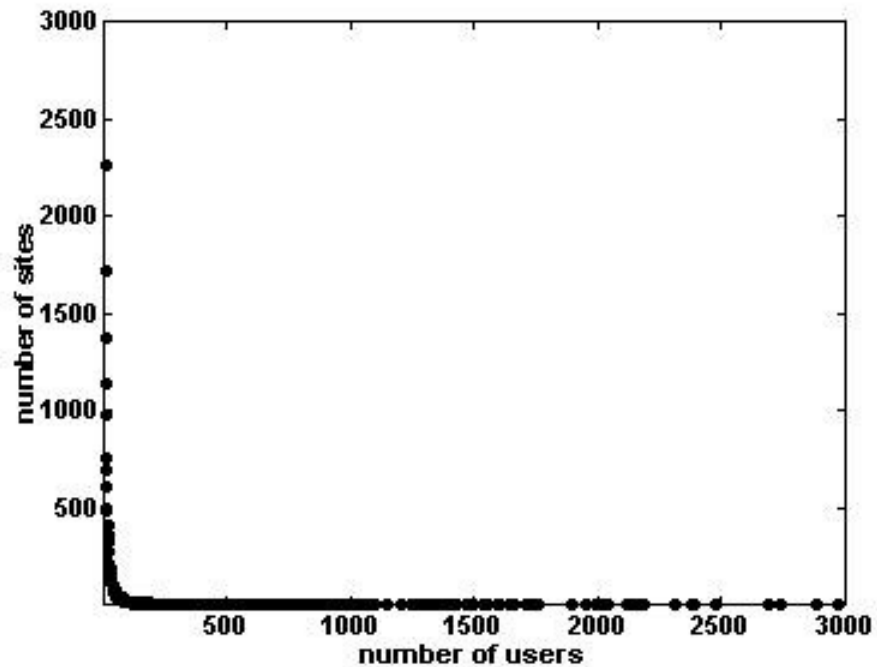
The plot is in log-log coordinates.

x is rank of a word in the frequency table; y is the total number of the word's occurrences.

Most popular words are “the”, “of” and “and”, as expected.

Zipf's law corresponds to the upper linear portion of the curve, roughly following the green $\log y = 1.3 \times 10^7 - \log x$ line.

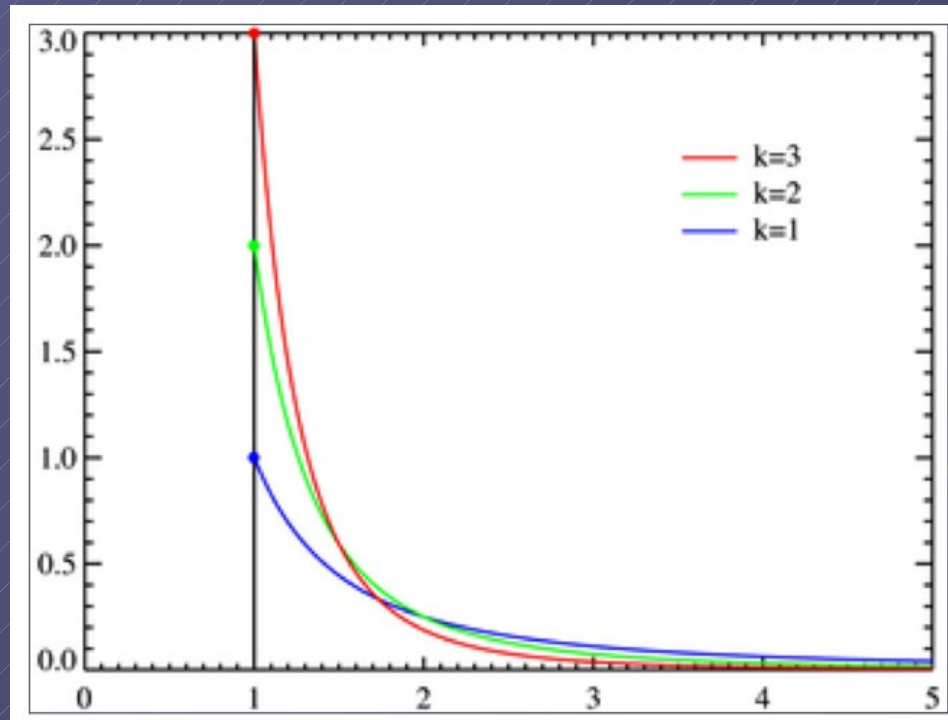




Source: Sites visited by AOL users, December day 1997.
L. Adamic, "Zipf, Power-laws, and Pareto - a ranking tutorial",
<http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html>

Pareto = Zipf = Power law

probability

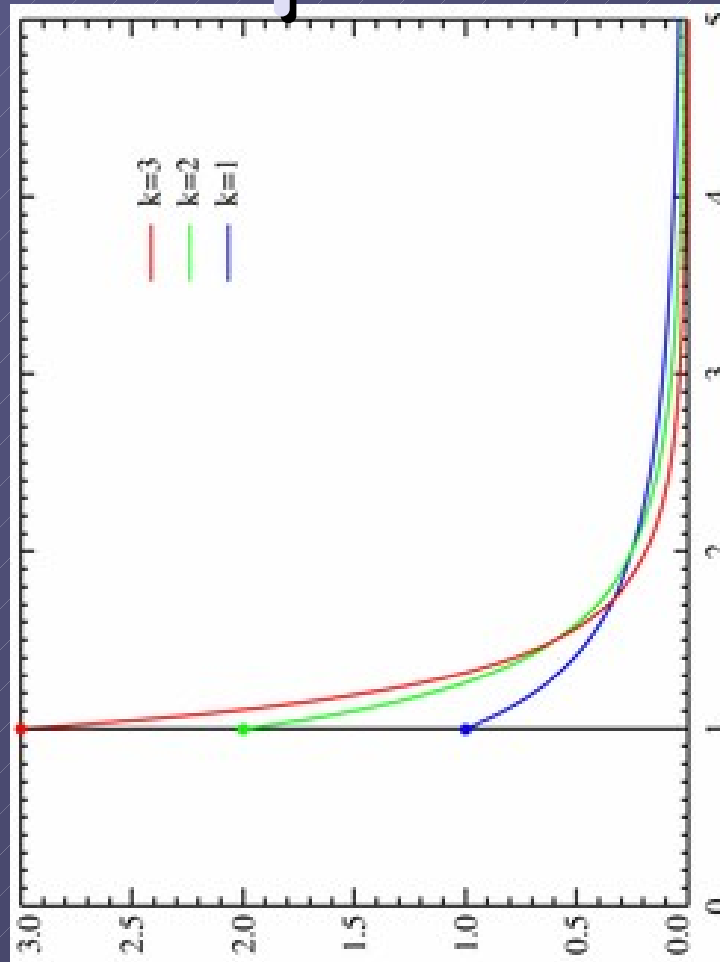


size of x

Source: Undetermined

Rotate...

Pareto = Zipf = Power law

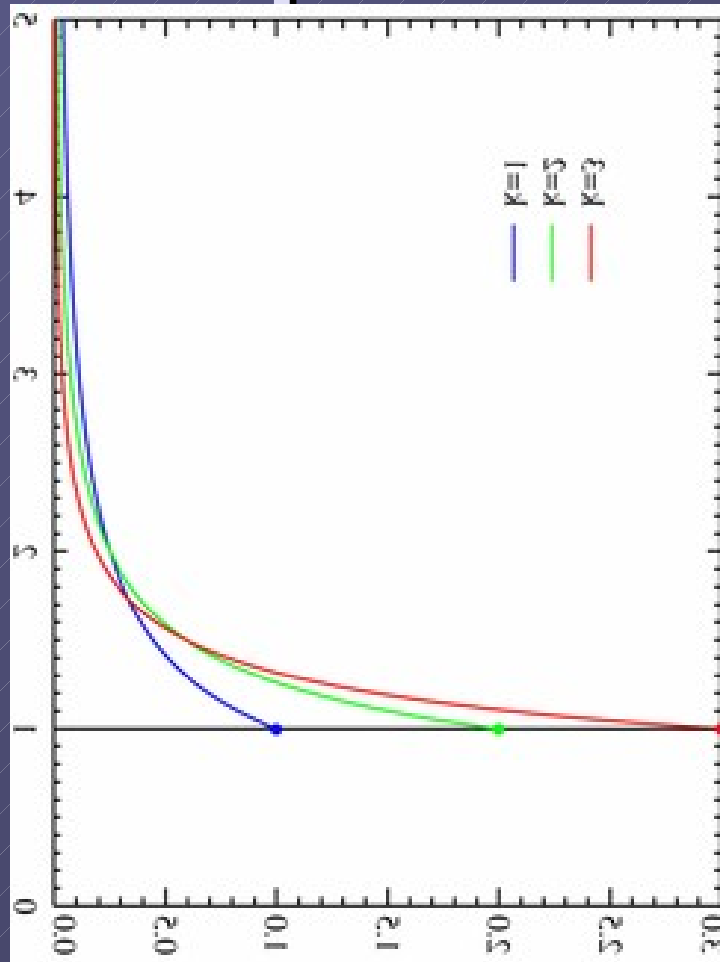


Source: Undetermined

Then flip...

Pareto = Zipf = Power law

size of x



rank of x

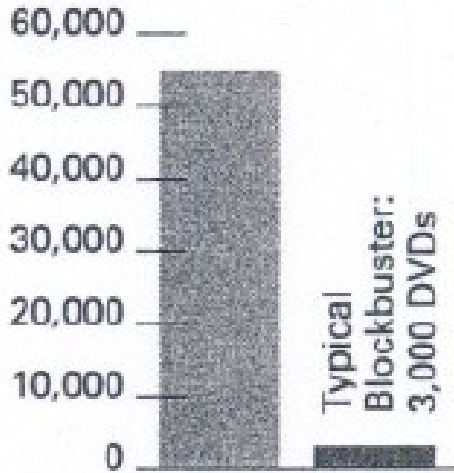
Source: Undetermined

What's this about an 80-20 rule?

For many phenomena, 80% of the consequences stem from 20% of the causes (*Pareto principle*)

Netflix

Total inventory:
55,000 DVD titles

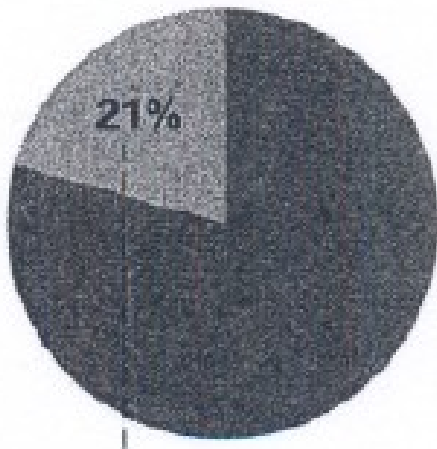


3000 out of 55,000 titles: 5.5%

79% of sales

For Netflix, “80 – 6 rule”

Total sales



**What causes
power laws for
goods variety?**

variety

quality variation

network effects

scarcity

What are we measuring? Rank
orders or significance of *what*?

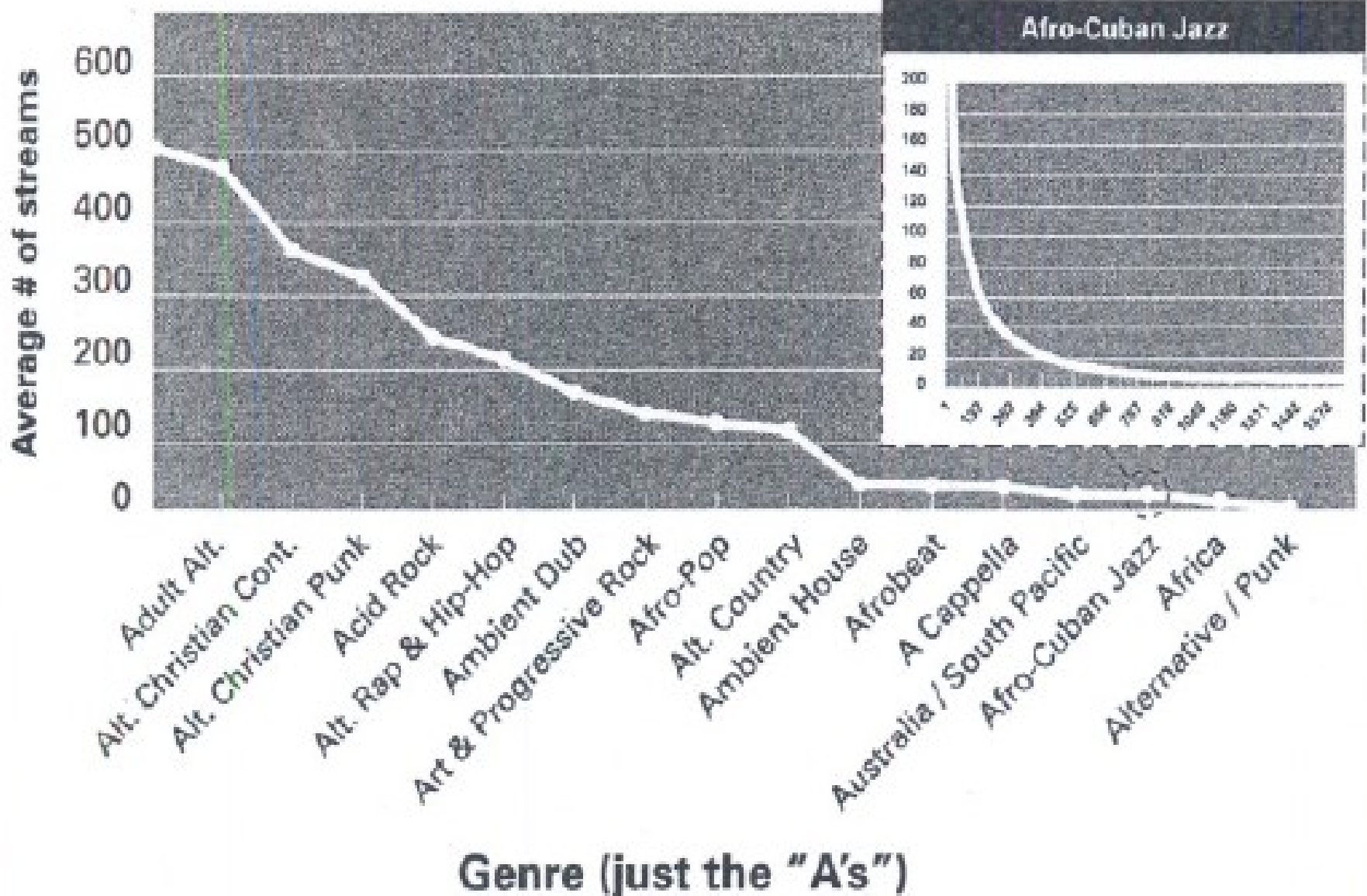
Maybe, if we're Amazon.

E.g., do we want to know about
books?

What if we're O'Reilly (technical publisher)?

Generally, goods power laws apply to subcategories (e.g., genres) as well...

Tails within Tails



Network effects and recommender services most effective at genre or subcategory level

Let's do the basic economics of
variety: demand for, supply of

Why the long tail now?

Didn't consumers demand variety before?


Need to know about and find variety.


YAHOO! SHOPPING Shop

HOME CLOTHING ELECTRONICS COMPUTERS HOME & GARDEN MORE

SHOP FOR: in


Clothing & Accessories Women's, Men's, Shoes, more...	Home & Garden Bed & Bath, Kitchen, Furniture, more...
Computers Laptops, Desktops, Software, more...	Jewelry & Watches Diamonds, Gold, Watches, more...
Electronics iPods, Cameras, TVs, Phones, more...	Sales & Bargains Center Free Shipping, Coupons, Sales, more...
DVDs, Music & Books Top Movies, Billboard 100, New Books	Sports & Outdoors Athletic Shoes, Fitness, Sports, more...
Flowers & Gifts Gifts, Roses, Gourmet Foods, more...	Toys & Baby Dolls, Car Seats, Board Games, more...
Fragrances & Beauty Perfumes, Makeup, Skin Care, more...	More Categories Autos, Health, Tools, Video Games

 Let us do the legwork. Free shipping on select furniture.



YAHOO! SHOPPING BUYING GUIDES

Need help Shopping? Visit our Buying Guides for expert advice and help choosing the product that is right for you.



- ▶ Digital Cameras
- ▶ MP3 Players
- ▶ Laptop Computers
- ▶ Television & Video
- ▶ Cell Phones
- ▶ Desktop Computers

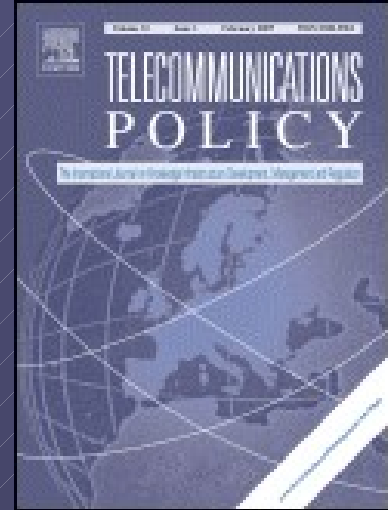
Other big changes are in cost of *providing* variety, which are...?

Suppose there are distribution
fixed costs.

$$\pi = (p - mc)Q - F$$

Offer if expect $\pi > 0 \rightarrow Q > F / (p - mc)$

If F gets smaller, more goods offered



So, fixed distribution costs favor mass market hits. Lower fixed costs favor niche goods.

See MacKie-Mason, Shenker and Varian, "Service Architecture and Content Provision: The Network Provider as Editor," in Telecommunications Policy, vol. 20, no. 3, April 1996: 203-17.

Does the marketing the Long
Tail imply a shorter head?

Does marketing to the Long tail
increase demand or just shift
it?

Anderson claims:

"Some forms of entertainment, such as music, are 'non-rivalrous' for attention, which is to say you can consume them while you're doing something else."

Agree?

Should prices be higher or lower
for products down the tail?

1. Make everything available
2. Help me find it

Anderson: Long Tail spawns two imperatives.

Do we need an economics of
abundance?

Anderson's nine rules

1. Move inventory way in...or
way out

2. Let customers do the work

3. One distribution method
doesn't fit all

4. *One product doesn't fit all*

5. One price doesn't fit all

6. Share information (lose control)

7. Think "and", not "or"

8. Trust the market to do your
job

9. Understand the power of
free