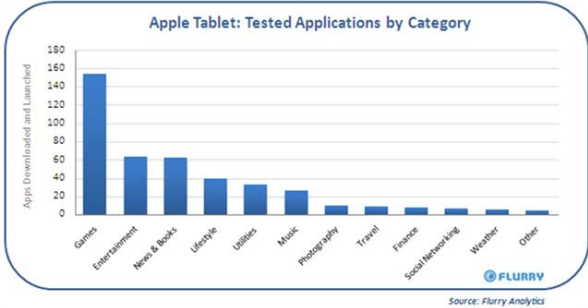
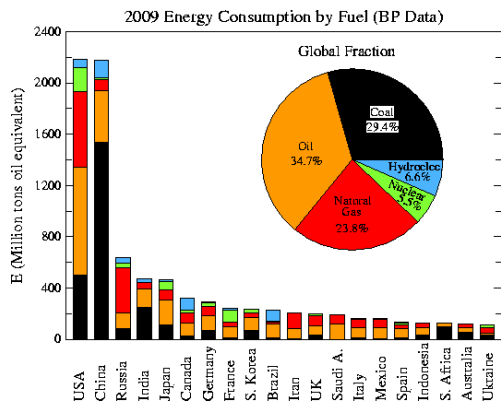
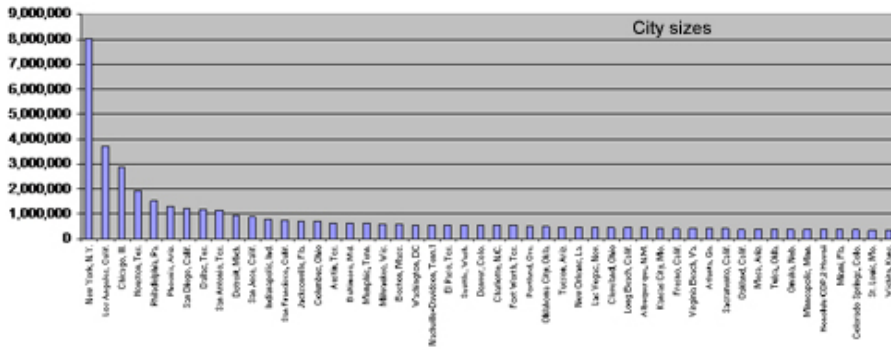
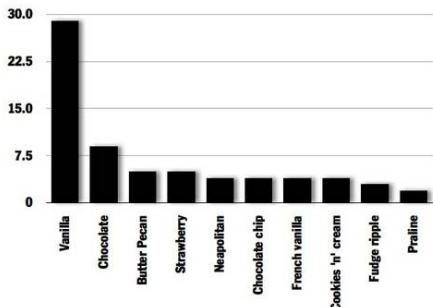


Pareto Distributions



It is a common observation that virtually everything measurable on the Internet — site popularity, site traffic, ad revenues, tag frequencies on del.icio.us, open source downloads by title, Web sites chosen to be digg'ed, Google search terms — follows such power laws or curves.

However, the real argument that Anderson made first in *Wired* magazine and then in his 2006 book, *The Long Tail: Why the Future of Business is Selling Less of More*, is that the Internet with either electronic or distributed fulfillment means that the cumulative provision of items in the long tail is now enabling the economics of some companies to move from "mass" commodities to "specialized" desires. Or, more simply put: There is money to be made in catering to individualized tastes.



The shape of Anderson's curve is a [power law](#). Power laws pop up in natural science, chaos theory, computer networks, and languages. In linguistics, [Zipf's "Law"](#) states that the distribution of word frequencies in a text follows a power law.

Let's check out Zipf's Law for ourselves. If you take the [text of Moby Dick](#), and count each word, you get this graph:

