

## One reason for the current importance of Statistics...

1

STOCK MARKET VALUES OF U.S. CORPORATIONS  
18th AUGUST, 2005

Microsoft	\$287 billion
Wal-Mart	\$197 billion
Time Warner	\$85.9 billion
Google	\$79.6 billion
eBay	\$54.8 billion
Disney	\$52.8 billion
Yahoo	\$48.5 billion
McDonald's	\$42.3 billion
GM	\$19.3 billion
Amazon	\$18.2 billion

2

Why has Google been so successful?

Answer:

Google's founders (Stanford University PhD students; S. Brin and L. Page) invented the

PageRank algorithm

for ranking web-pages in the mid-1990s — which vastly improved upon existing search technology.

3

## Opportunity for Statistics?

One of the fundamental structures of Statistics is the

logistic regression

A Google search of

"PageRank algorithm"+"logistic regression"

on 28th February, 2011, gave 2,880 hits.

4

## Some Recent Internet Retrieval Research

### ENHANCING RETRIEVAL WITH HYPERLINKS: A GENERAL MODEL BASED ON PROPOSITIONAL AUGMENTATION SYSTEMS

by Justin Picard and Jacques Savoy

5

Catalysts included technologies such as:

- data-base systems
- bar coding (e.g. supermarket scanners)
- the Internet
- digitisation
- microarrays
- functional magnetic resonance imaging

7

## The Data Explosion

The rise of Google is a consequence of what many are now calling the

### THE DATA EXPLOSION

This phrase had about 70,600 Google hits on 28th February, 2011.

6

## Dealing With The Data Explosion

Two growth areas in quantitative sciences  
are:

- Data Mining
- Machine Learning

8

## Two Relevant Items (find on web)

- The Ori Allon story.
- The Hal Varian speech.

9

Comparing genetic composition of a healthy individual and a patient diagnosed with cancer via **microarray technology**

scientists can pinpoint

**which genes**

may be involved in cancer development and **design best treatment for cancer patient.**

11

## Data Mining by Example

Supermarket chain **mines scanner data** and discovers that **significant profit increases** can be made by putting

**beer displays**

in the

**nappy section.**

10

## Statistical Machine Learning

**Statistical techniques** are enjoying increasing use in machine learning problems and leading to significant improvements.

12

## Relevance to **Advanced Data Analysis**

All statistical techniques for dealing with

- Internet searching,
- AIDS vaccine trials,
- hedge fund management,
- traffic flow management (cars and Internet downloads)
- many other problems.

rely upon certain **principles and methods**.

13

## Australia Needs You!

On 7th February, 2007 the findings of a report

### **Mathematics and Statistics: Critical Skills for Australia's Future**

was discussed at the Australian Academy of Science in Canberra.

See:

[www.review.ms.unimelb.edu.au/Forum.html](http://www.review.ms.unimelb.edu.au/Forum.html)

15

Many advanced

**statistical principles and methods**

will be covered in **Advanced Data Analysis**

14

## Jobs in Statistics

[www.statsci.org/jobs](http://www.statsci.org/jobs)

16