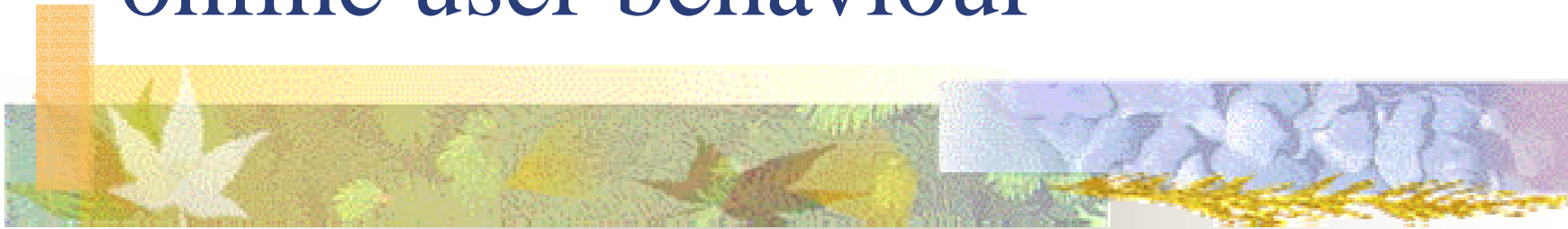# What usage statistics say about online user behaviour

Philip Davis, Librarian

Cornell University

# What do these stats mean?

| A. Monthly Statistics by Journal | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Subscribed Journal Usage** | | | | | | |
| **Journal** | **Jan 2002** | **Feb 2002** | **Mar 2002** | **Apr 2002** | **May 2002** | **Jun 2002** | **Total Use** |
| | | | | | | | |
| **A** | 100 | 131 | 136 | 193 | 80 | 56 | 696 |
| **B** | 0 | 0 | 0 | 0 | 10 | 0 | 10 |
| **C** | 8 | 6 | 39 | 5 | 88 | 368 | 514 |

# Usage statistics do not tell us…

- ***what*** is being downloaded
- ***who*** did the downloading
- ***why*** an article was downloaded
- ***how many*** individuals are responsible for the statistics

# Why we can't know everything

- Patron confidentiality
- Use IP address as a surrogate for "user"
- Some IPs represent aggregate users
    - Library proxy server
    - Public computers in libraries and labs
    - Dial-in modem users
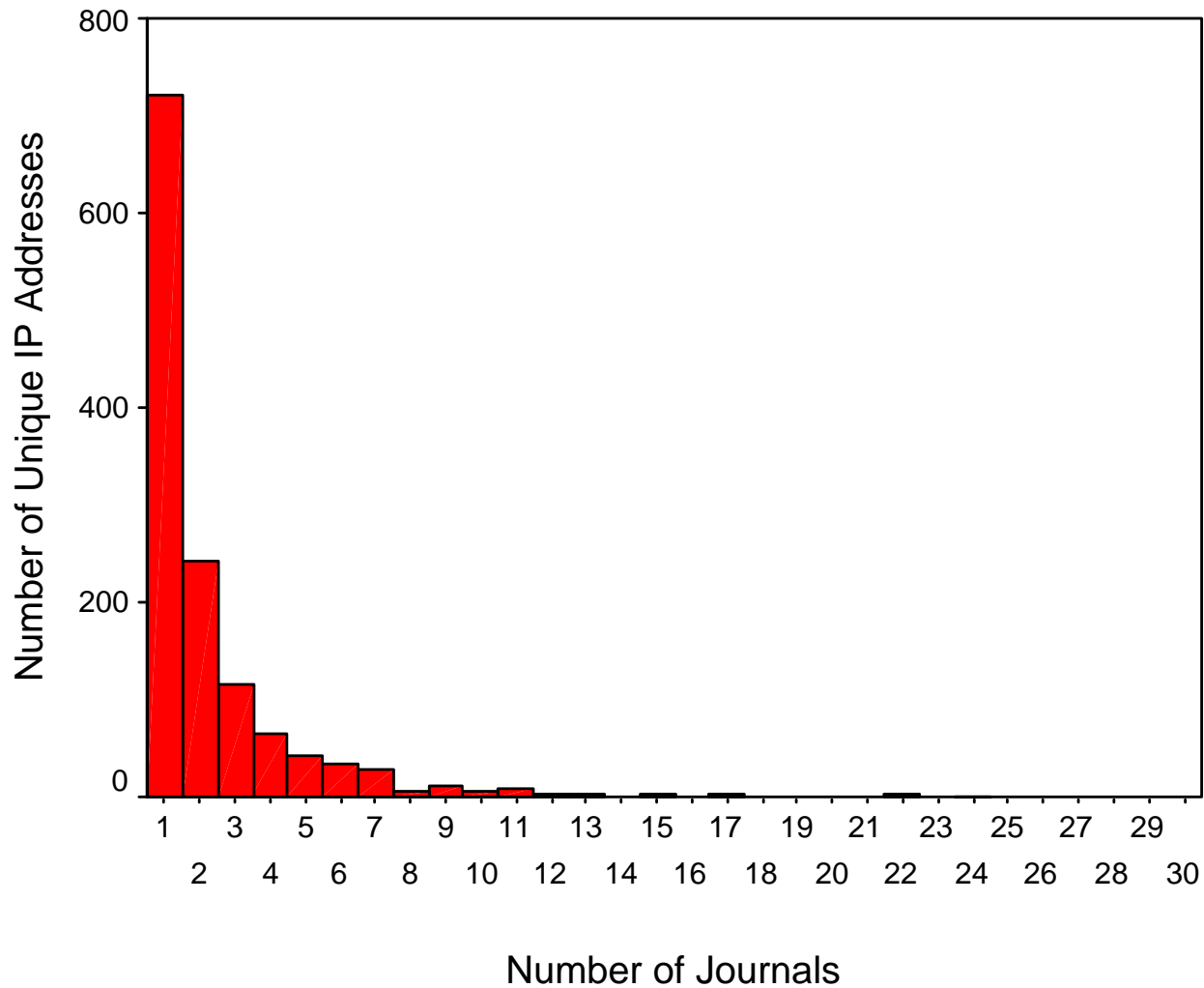- Some IPs are assigned dynamically

# Results of two studies (ACS)

- Article downloads by IP address
  - Previous studies have reported only aggregate use analysis

- How scientists find the articles they read
  - Using referral URL data. The location from which users were referred to the ACS site.
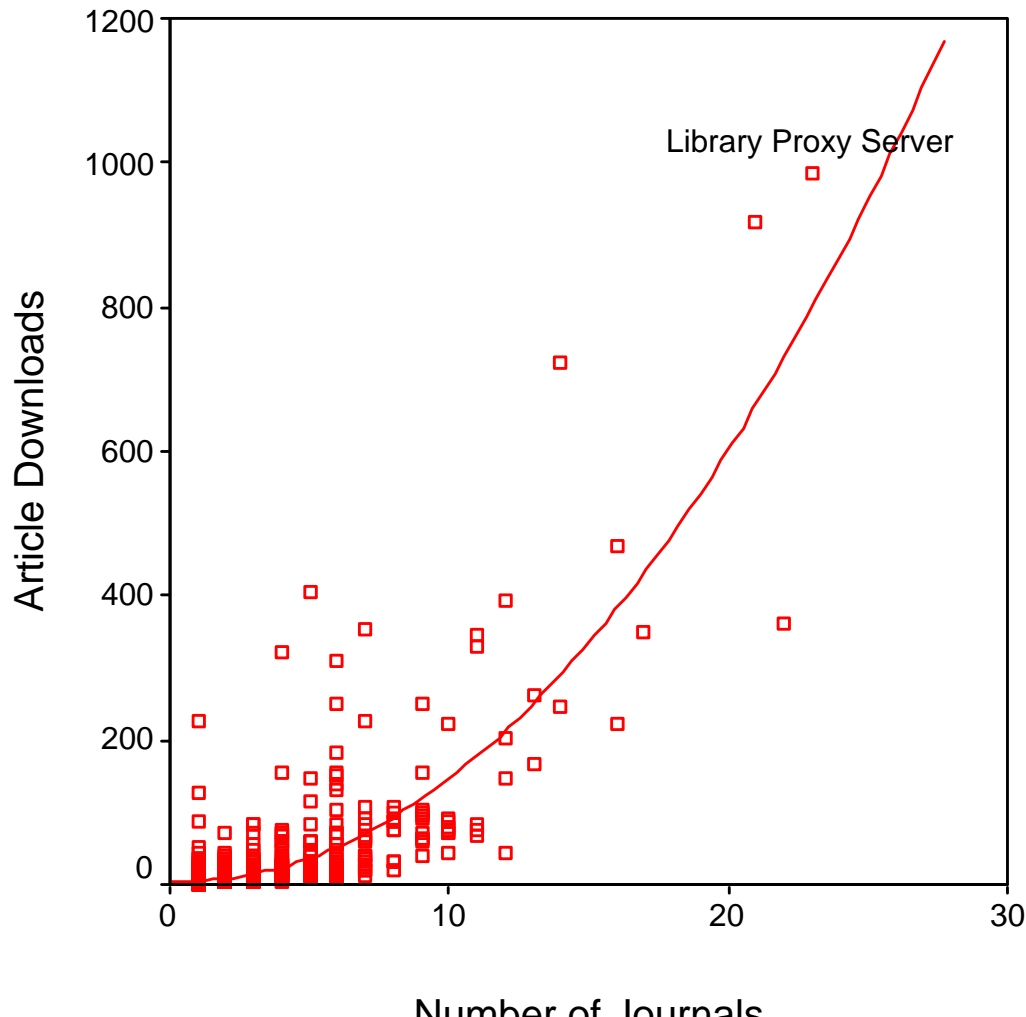
# Most users download few articles
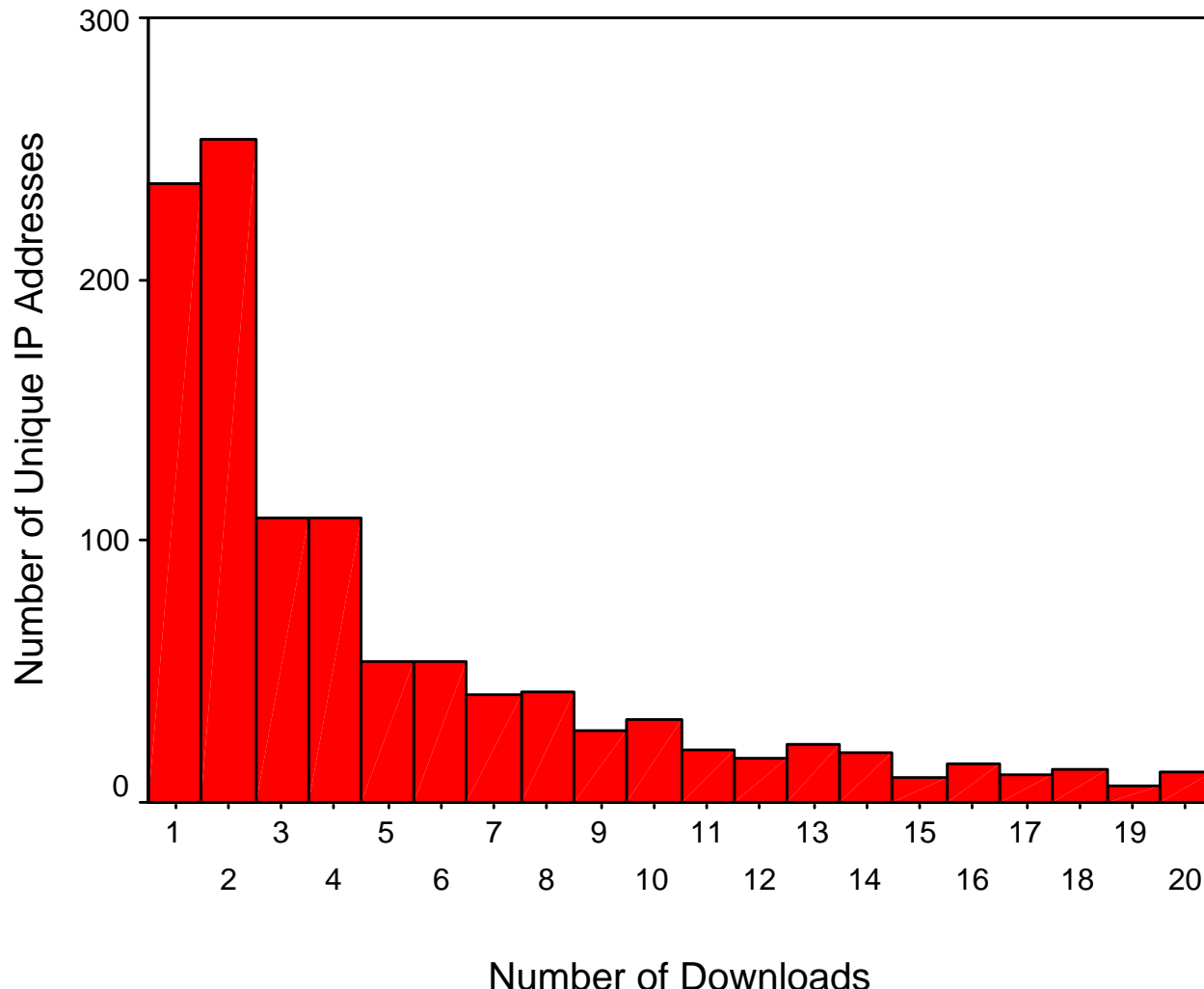
# …from few journals
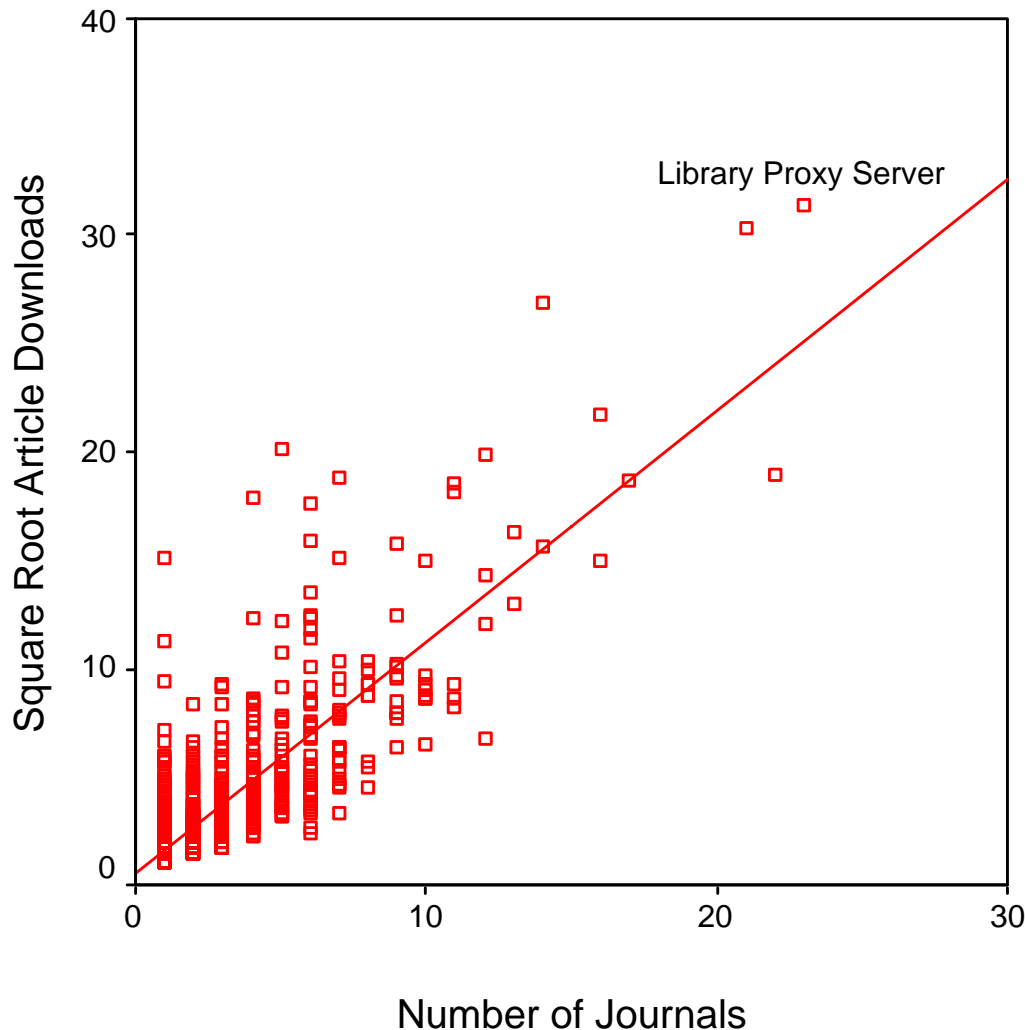
# The relationship is quadratic



Each point represents a "user"
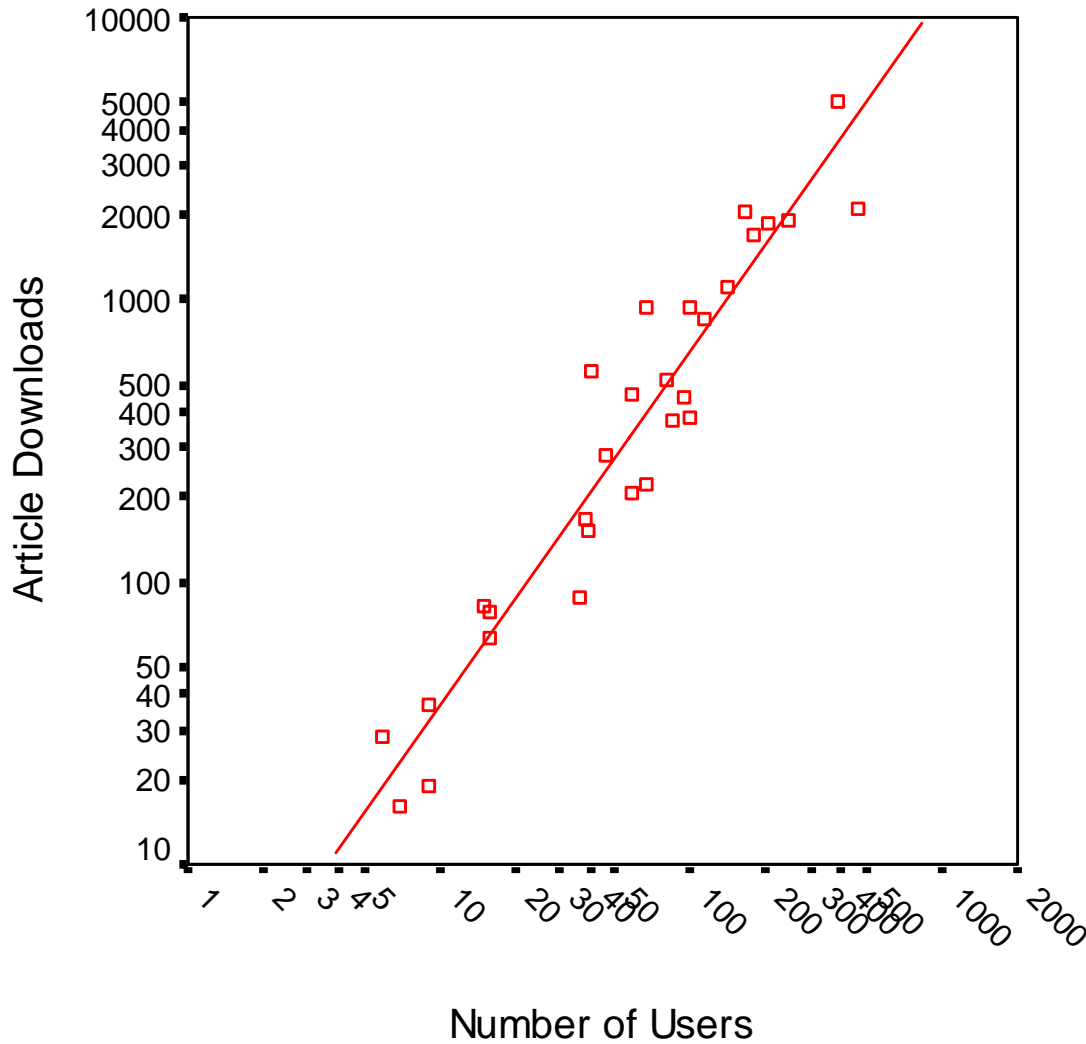
N = 1283

Rsq = 0.6798

# In fact, its an inverse square law



$x/y^2$

N = 1283

Rsq = 0.6858

# Population size may be estimated



Each point represents a journal

Rsq = 0.9169

Article Downloads

Number of Users

# Analysis of *individual* use illustrates

- Most users download few articles from few journals
  - A small number of users have a very large effect on total downloads
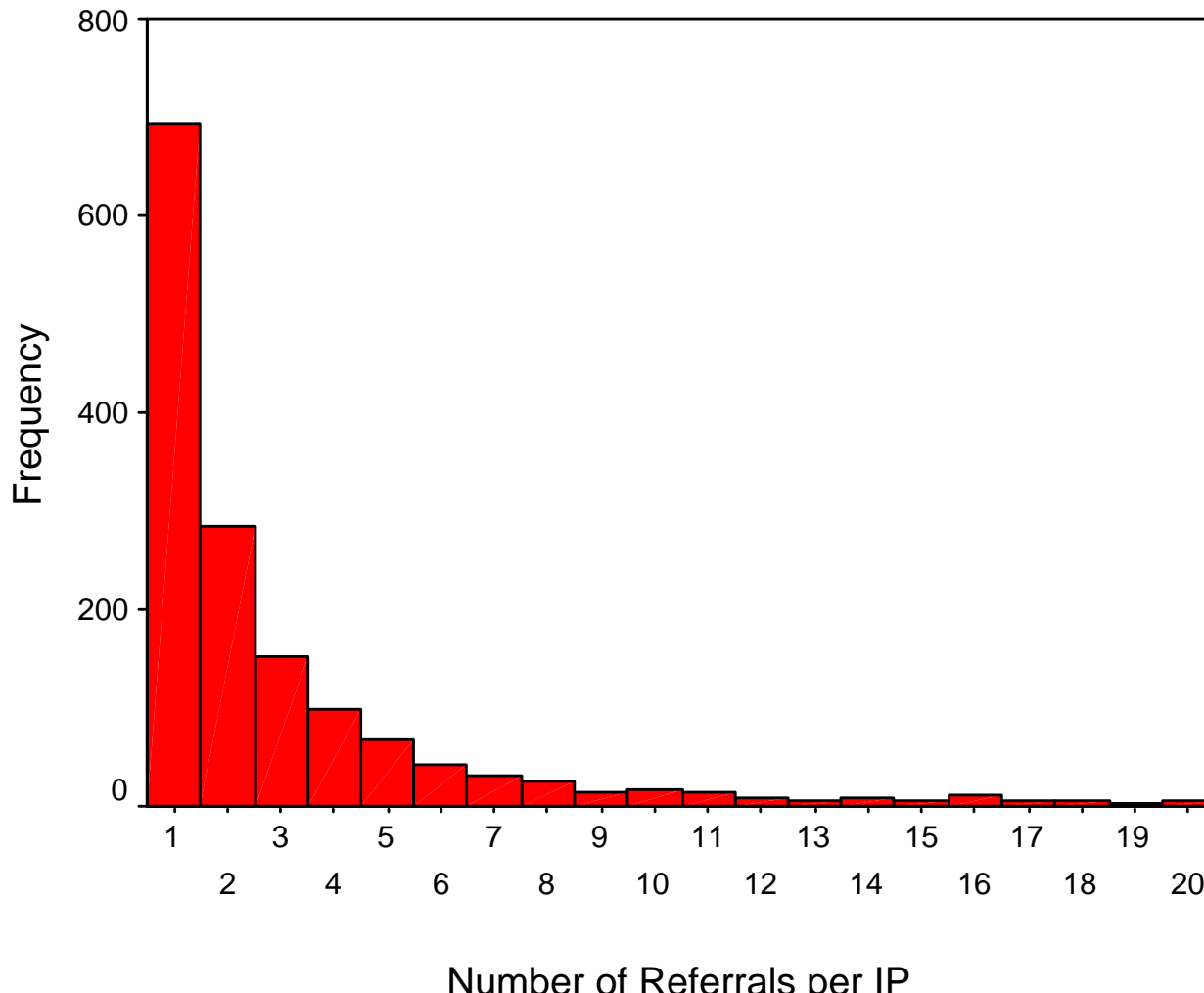- User population size may be estimated by total use

# Different paths to same destination

| Referral Type | Total Referrals | Unique IPs | Referrals per IP |
|---|---|---|---|
| library catalog | 2,482 | 552 | 4.5 |
| bib database | 2,372 | 324 | 7.3 |
| e-journal list | 1,813 | 405 | 4.5 |
| web page | 1,108 | 190 | 5.8 |
| web search | 996 | 491 | 2 |
| email (web based) | 592 | 79 | 7.5 |
| article link | 571 | 204 | 2.8 |
| other | 15 | 9 | 1.7 |
| Total Referrals | 9,949 | 1591 | 6.3 |

# Web page referrals

| Web Page Referral | Frequency | Percent |
|---|---:|---:|
| ACS Journal Web Page | 366 | 33 |
| News | 272 | 25 |
| Department/lab | 200 | 18 |
| Faculty | 75 | 7 |
| Course web page | 43 | 4 |
| Commercial | 31 | 3 |
| Organization | 21 | 2 |
| Personal | 19 | 2 |
| Other | 81 | 7 |
| Total | 1108 | 100 |

# Most users referred infrequently



Number of Referrals per IP

# …from few sources



Number of Unique Domain Referrals per IP
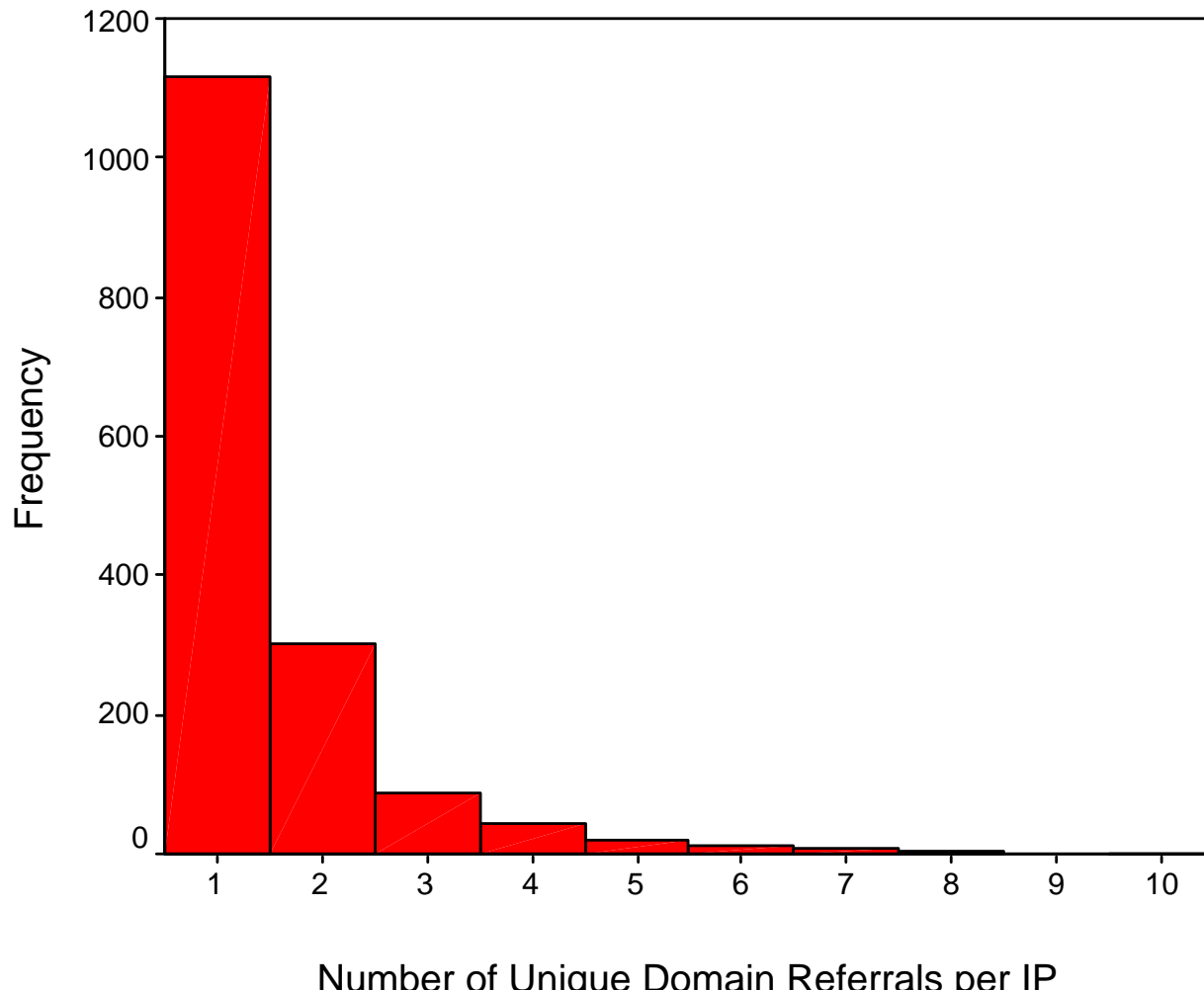
# Yielding same inverse square law
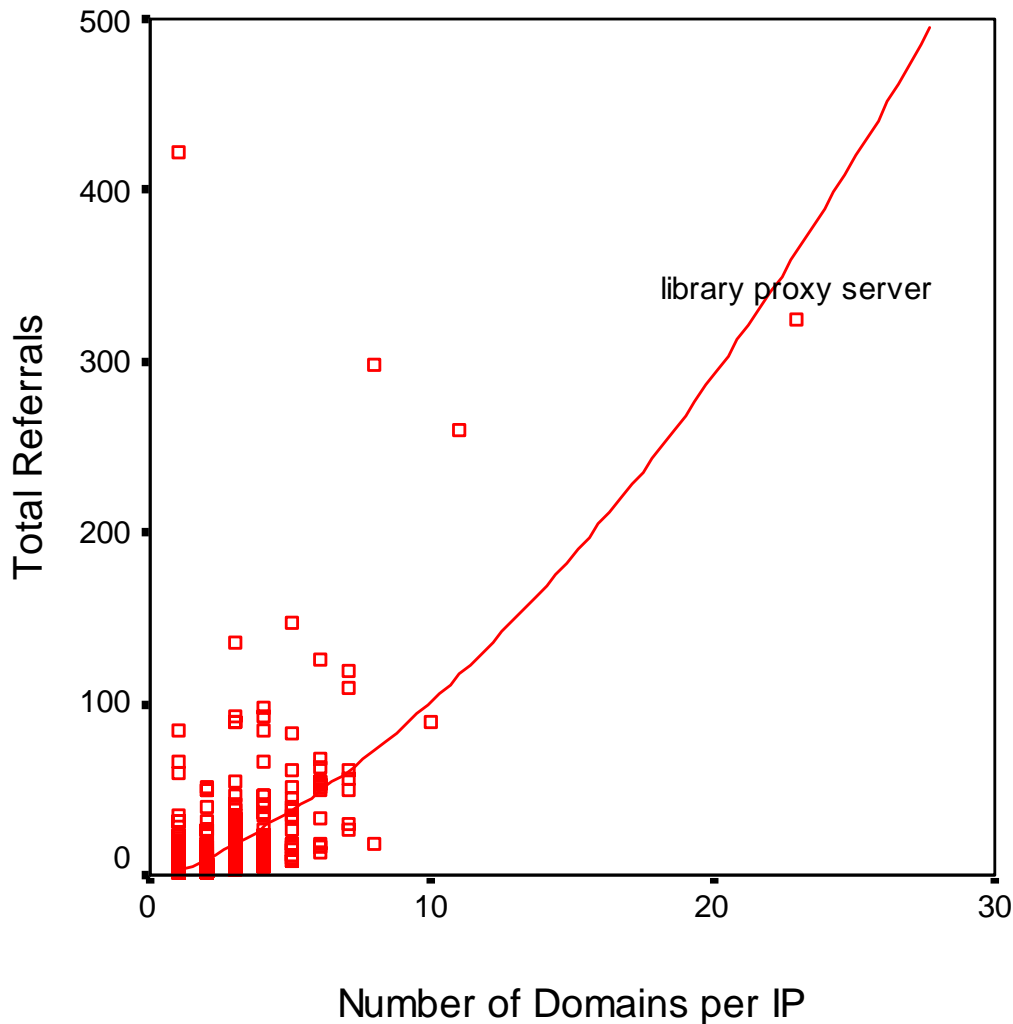


Each point represents a "user"

N = 1591
Rsq = 0.4107

# In summary

- Scientists will use many different pathways to the same literature
    - But use few and consistent methods of referral
- Underestimated the use of e-mail and bookmarking as a source of referral
- Underestimated bibliographic indexes
- Overestimated importance of library catalog

# Implications

## Libraries

- Develop **redundant tools** to facilitate access to literature

## Publishers

- Facilitate direct linking to article
- Adoptions of linking standards

**"Save the time of the reader"**

-- S.R. Ranganathan, from the Five Laws of Library Science

P. Davis and L. Solla. An IP-level analysis of usage statistics for electronic journals in chemistry: Making inferences about user-behavior.  JASIST 54(11), 2003 in press.

P. Davis.  Information seeking behavior of scientists: a transaction log analysis of referral URLs.  (in review, JASIST, June 19, 2003).

**http://people.cornell.edu/pages/pmd8/**