Purdue University
Purdue e-Pubs

Department of Computer Science Technical Reports | Department of Computer Science

1994

Metrics to Evaluate Academic Departments

John R. Rice
Purdue University, jrr@cs.purdue.edu

Report Number:
94-048

Rice, John R., "Metrics to Evaluate Academic Departments" (1994). *Department of Computer Science Technical Reports.* Paper 1148.
https://docs.lib.purdue.edu/cstech/1148

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

METRICS TO EVALUATE ACADEMIC DEPARTMENTS

John R. Rice
Computer Sciences Department
Purdue University
West Lafayette, IN 47907

CSD TR-94-048
July 1994

METRICS TO EVALUATE ACADEMIC DEPARTMENTS

John R. Rice

Abstract

This report discusses the types of quantitative metrics used to evaluate academic departments. It is pointed out that highly aggregated measures, the ones most commonly used, usually provide little insight into the actual "performance" of departments. More precise metrics are needed, ones that nevertheless can be understood quickly and obtained with reasonable effort. Metrics are discussed for the areas of Research/Scholarship, Service, Teaching, Space Ultilization, Courses, and Support Activities. To demonstrate the possibilities for better metrics, 15 visual metrics are given in the areas of Research/Scholarship (3), Teaching (4), Courses (2) and Space (6). These examples are actual data for the Department of Computer Sciences at Purdue University.

1. INTRODUCTION

The metrics Purdue uses to evaluate departments are all highly aggregated, e.g., *total contact hours*, *total research expenditures*, *total faculty FTE*. These are often combined to produce averages, e.g., *credit hours per faculty FTE*, *research expenditures per square foot of research lab space*. There are three interrelated shortcomings in using such aggregated measures. To create examples for this discussion I hypothesize a department of 10 people and an area of interest called X where 1000 is a satisfactory departmental value.

1. *Lack of Precision.* A department's average in area X may be satisfactory and yet there are serious shortcomings in area X. The average of 100 can be achieved by

   - 10 people produce 100 each,
   - 6 people produce 150 each and 4 people produce 25 each,
   - 1 person produces 1000 and the other 9 produce nothing.

2. *Overreliance on Averages.* Averages are well known to hide information and produce unrepresentative measures. Medians, quartiles, deciles, etc. are much more representative measures.

3. *Insufficient Information about the Nature of a Department's Operations.* The goal is to produce a handfull of metrics that characterize what a department does. It is unlikely that this can be done well with 12 or 15 numbers. The numbers used at Purdue include groups that are highly correlated (e.g., credit hours, contact hours, class sizes) so there is less real information present than one might expect.

Finally, there is a fourth shortcoming:

4. *Lack of Calibration.* There is frequently no intuitive a priori way to determine satisfactory values of a metric for a department. Thus "satisfactory" becomes defined by comparisons with other departments or similar departments at other schools. In a particular school of 8-20 departments even a small number of under-performing departments substantially affect the averages and thus make the "satisfactory" value represent lower performance than it should. Similarly, if the bulk of the departments each have a fraction of under-performing people, then this approach does not detect their existence.

The conclusion is that metrics are needed which have more precision, which provide more insight about the nature of a department's operations. The question is: **Are there such metrics that can be obtained with reasonable effort and which are easily understood?** By "obtaining with reasonable effort" we mean that the data is already available to a department (or easily derived from existing data) and metrics can be derived from it with the same levels of clerical and administrative effort used now. By "easily understood" we mean that a few minutes of study by an experienced evaluator should provide considerable insight into some aspect of a department's operations.

Note that some important metrics for a department are somewhat atomic and thus well represented by a single number. Examples of these include: Number of undergraduate majors, number of graduate students, graduation rates, and admission rates.

2. METRICS AND THEIR CHARACTERISTICS

Before discussing what to measure we present two general guidelines and a caveat:

- *The units of a metric should be easily understood.*
- *Each metric should be understandable with no more than 30 seconds to a minute's study.*
- *Metrics are unlikely to measure quality directly.*

The first guideline implies that metrics involving people should be in terms of what a person does. Units like $/person/year, courses/person/semester, Ph.D. students/person/decade are easily understood. The measures should not use simple units to hide complex formulas. For example, courses/person/year in some departments might include credits for Ph.D. students, committees chaired, editorial responsibilities and similar activities. Such activities are important and people should be given proper credit for them, but they should be accounted for separately and not counted

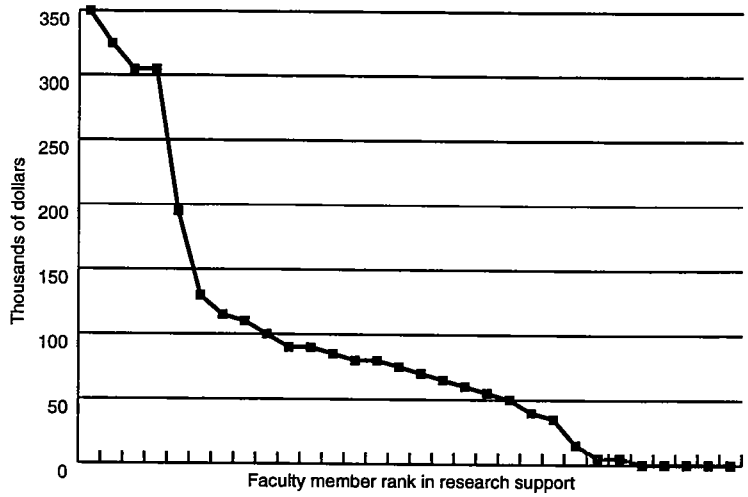


Figure 1: An example of a visual metric. The research expenditures for each faculty member are plotted in descending order. There are four people with much higher levels, perhaps they have a big equipment grant?(They do.) Then there are a few with no or very low research support, perhaps they are new Assistant Professors?(Two are), or have research support through other institutions?(Two do), or are retired?(One has been Professor Emeritus for 11 years.)

as courses taught. On the other hand, adjustments in teaching metrics must be made for things like sabbatical leaves, buy-outs from research funds, or course administration duties. Some metrics do not involve people as the basic element but the same guidelines are to be used.

The smallest metric is a single number and the desire for conciseness drives evaluation systems to try to reduce metrics for each area of interest to a single number. A table of numbers, even a short one, is hard to digest quickly. However, people are very well equipped to absorb visual information and a graph of 30–50 related numbers can be understood very quickly compared to understanding even 5 or 10 disparate numbers. Figure 1 shows a plot of research expenditures in the 1992/93 year for each faculty member in the Department of Computer Sciences. A glance shows the bulk of the faculty having substantial funding. This figure illustrates how much relevant, detailed information can be quickly grasped from a plot.

It is imperative to recognize that many of these metrics are unlikely to measure quality of activities. Numbers of publications, Ph.D. students or research dollars are likely to be correlated with research quality, but exceptions do occur in both directions. Similarly, teaching popularity probably reflects teaching quality but it can also reflect other things.

3. METRICS AND THEIR USES

It is important to decide what one wants to measure. The use of resources (people, space, money) is best suited to be studied by metrics. Even here, generic metrics primarily serve to focus attention on unusual situations. Final conclusions normally require judgements involving more information

than caputured in metrics. The quality of operations is much harder to study using metrics. A faculty can do a great job in teaching an out-of-date curriculum to hundreds of students.

The most important activities in an academic department are people related, we restrict the discussion here to faculty members. Metrics about faculty that are interesting and feasible to obtain readily include:

RESEARCH/SCHOLARSHIP.

- Publications
- Funding

SERVICE.

- Committee assignments
- Contacts (with students, contributors, etc.)
- Money raised

TEACHING.

- Number of courses
- Number of students (credits or contact hours)
- Courses developed
- Theses directed

There are other metrics that are not easily associated directly with a person. This is intrinsicly so for items like use of space for teaching, research, or support operations. It is appropriate to relate them to the "size" of something, e.g., number of faculty, number of majors, number of teaching assistants, amount of research funding. There are also situations where a clear picture of a department's operations is not obtained by data about the people in the department. For example, substantial teaching might be done by "transients" such as adjunct faculty, short term visitors, or graduate students. Metrics of this type that are interesting and feasible to obtain readily include:

SPACE.

- Office sizes (by category of occupants)
- Classroom utilization
- Research labs
  - research expenditures
  - research participants (perhaps by category)
- Support space
  - storage, conference rooms, administrative areas, lounges

COURSES.

- Size
- Student characteristics (majors or not, undergrad or grad)
- Instructor characteristics (faculty level, non-faculty, etc.)

SUPPORT.

- Administration
  - clerical, professional/administrative, faculty
- Teaching and student services
  - clerical, professional/administrative, faculty
- Technical and research
  - clerical, professional/administrative, faculty
- Outreach
  - clerical, professional/administrative, faculty
  - funds raised and spent

The purpose of this report is collect and display some visual metrics, these are given in the appendix. Each is discussed briefly to indicate some of the information that can be derived from them. These are all from a single department, and thus inadequate as a base for a judgement about their general utility. As a minimum one needs to examine the metrics for a number of departments in a single university and for a number of similar departments in different universities.

4. APPENDIX: METRICS FOR THE DEPARTMENT OF COMPUTER SCIENCES

We present metrics for the Department of Computer Sciences. Brief discussions are given along with the metric which illustrate types of information easily identified from the metrics. Not all the metrics discussed above have been generated. The metrics given here are listed.

RESEARCH/SCHOLARSHIP

- Articles published by tenured faculty, 3 years (1990–92).
- Books published by tenured faculty, 3 years (1990–92).
- Research expenditures of tenured and tenure/track faculty, 1992/93.

TEACHING

- Teaching loads of tenured and tenure-track faculty, 3 years (1990/91–1992/93).
- Teaching loads of all instructors, 3 years (1990/91–1992/93).
- Ph.D. graduates, tenured faculty, 5 years (1989–93).
- Student evaluations of all regular instructors, 3 years (1990/91–/1992/93).

Here "instructors" refers to everyone teaching lecture courses and "regular instructors" excludes temporary people who teach one year or less.

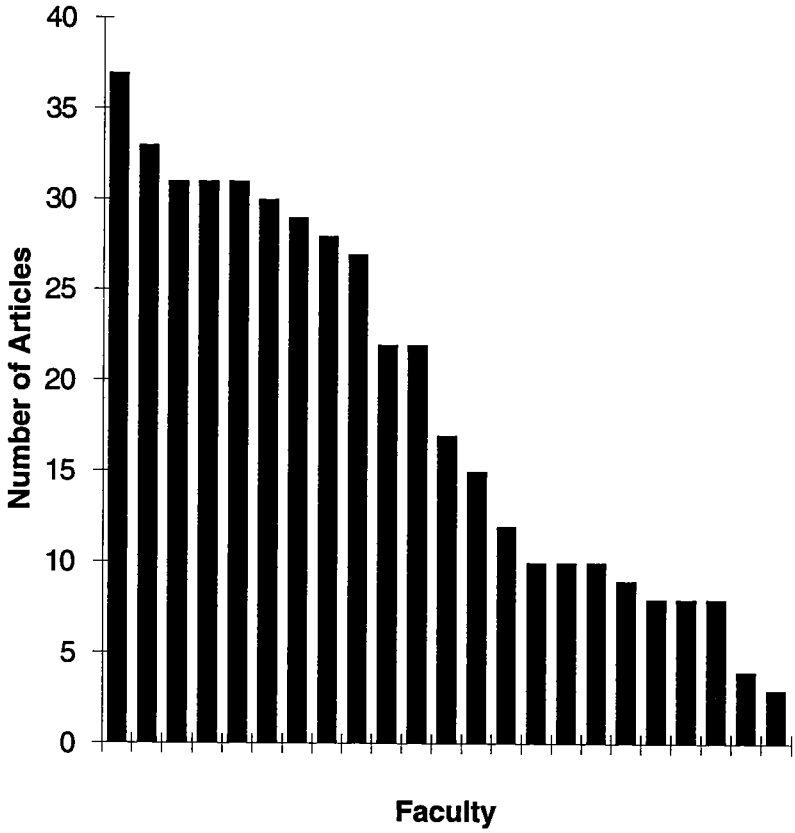
COURSES

- Size of lectures for all courses, 3 years (1990/91–1992/93).
- Size of lectures for graduate courses, 3 years (1990/91–1992/93).

SPACE

- Office sizes by categories: faculty, A/P staff, clerical.
- Office sizes for graduate students.
- Research lab expenditures per square foot, 1992/93.
- Research lab space per research-supported participant.
- Teaching lab contact hours/week per square foot, 1992/93.
- Support space per person: 4 categories (administrative, conference, lounge, storage) per (employee, faculty, grad major).

Articles Published by Tenured Faculty 1990-92

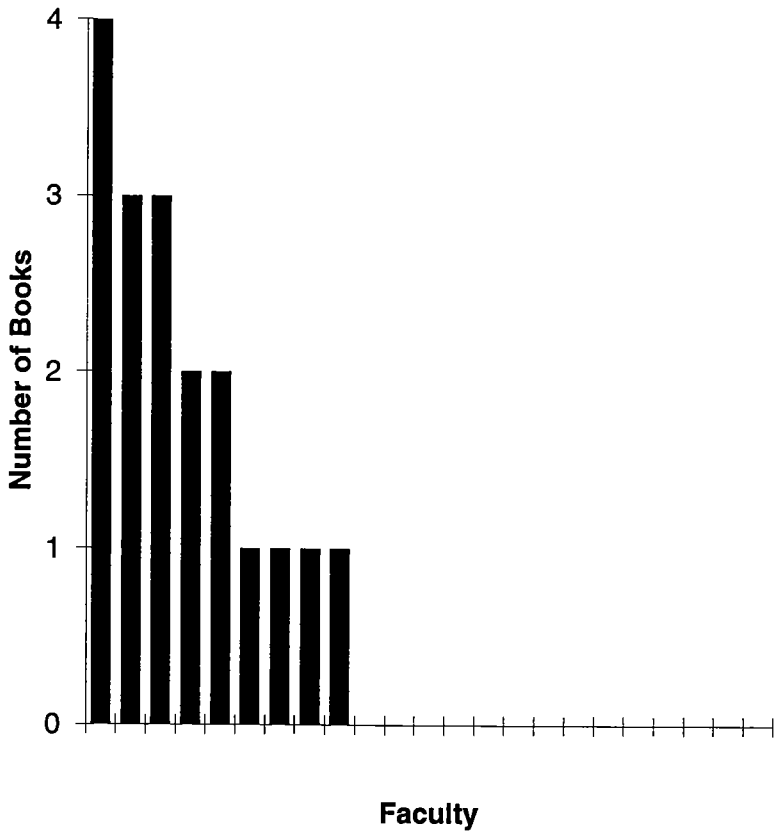


Data Definition: The data includes journal articles, conference proceedings, book chapters, and survey articles with publication dates of 1990, 1991, or 1992.

Purpose: Measure the productivity in research/scholarly publications of the tenured faculty in the department.

Comments: Only tenured faculty are included so as to exclude skewing by data from faculty just starting their research careers. This profile of publications allows one to view the typical and extreme rates of publications in the department. Recall that publications rates vary considerably from field to field so that absolute values here are not as significant as relative values.

Books Published by Tenured Faculty 1990-92

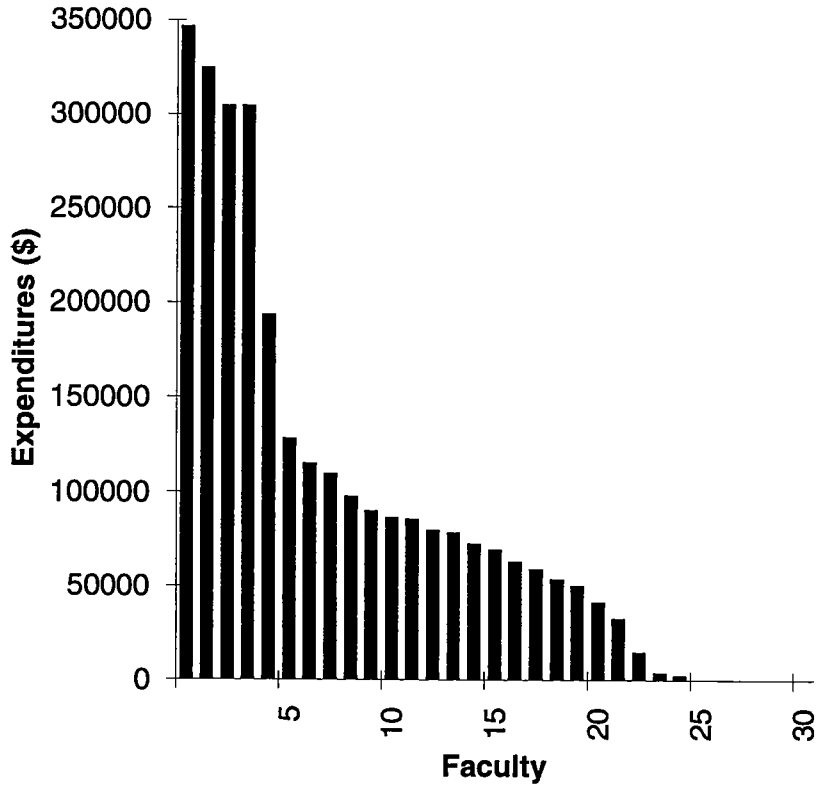


Data Definition: The data includes all books authored or edited with publication dates of 1990, 1991, or 1992.

Purpose: Measure the productivity in writing or editing books by the tenured faculty of the department.

Comments: The important and frequency of book publications vary substantially from field to field. The three year time span used here is probably too short for the best evaluation.

Fair Share Research Expenditures of 30 Tenured and Tenure-Track Faculty 1992/93



Data Definition: The total expenditures, including indirect costs, of external research grants is given for the 1992/93 academic year. If a grant has more than one principal investigator then the amount is allocated proportionally in a "fair" way.

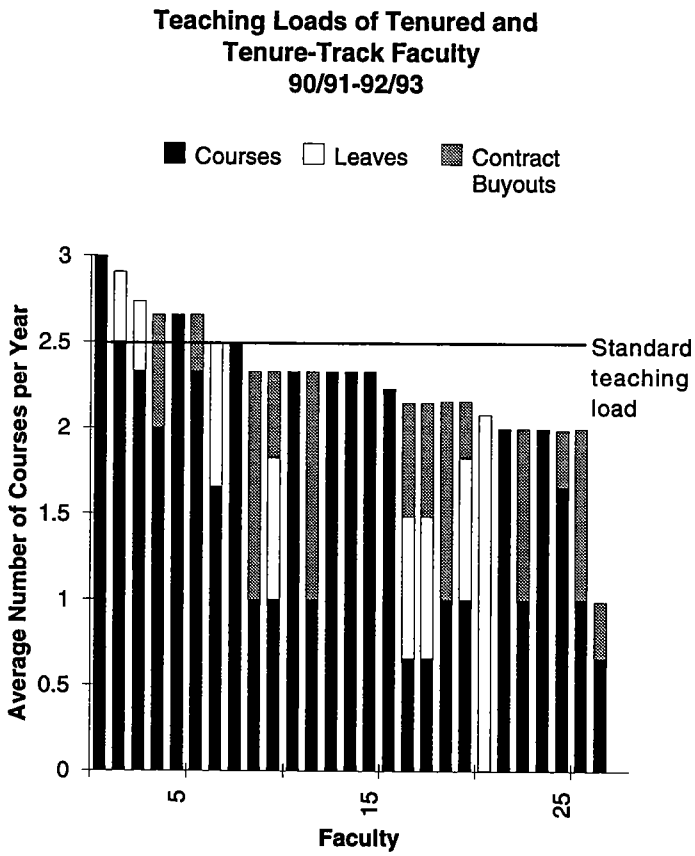
Purpose: Measure the external support for the researchof the faculty.

Comments: Perhaps the most interesting parts of these profiles are the extremes. In evaluating the research activity of a department one should examine both those who receive a lot of research funding and those who receive little or none. There are various reasons for low research funding so one should expect some, e.g.,

- New Assistant Professors have not yet established research programs.
- Research support might come in ways not visible to the accounting system, e.g., support through "gifts" or support through other institutions.

The specific circumstances of each faculty member with low research support should be determined before making judgements about the level of research/scholarly activity.

There are very substantial differences among disciplines in the normal levels of research funding. Even within a single department there can be substantial differences, e.g., between theoreticians and experimentalists.



Data Definition: The data for tenured and tenure-tack faculty include: courses taught (black), adjustments for leaves (white) (sabbatical or unpaid), and credit for course buyouts (shaded) using external research funds. The unit is one semester course meeting three times a week. Some courses may count more than one unit due to multiple sections with many assistants to supervise or to meeting more than three times a week. A faculty member with teaching duties in two departments has the load adjusted accordingly. Only those faculty in the department for 1992/93 are included.

Purpose: Measure the teaching contribution of the tenured or tenure-track faculty.

Comments: The standard faculty teaching load (2.5 courses per year in this case) is indicated by the horizontal line. There are several sources of variations in the actual loads e.g.:

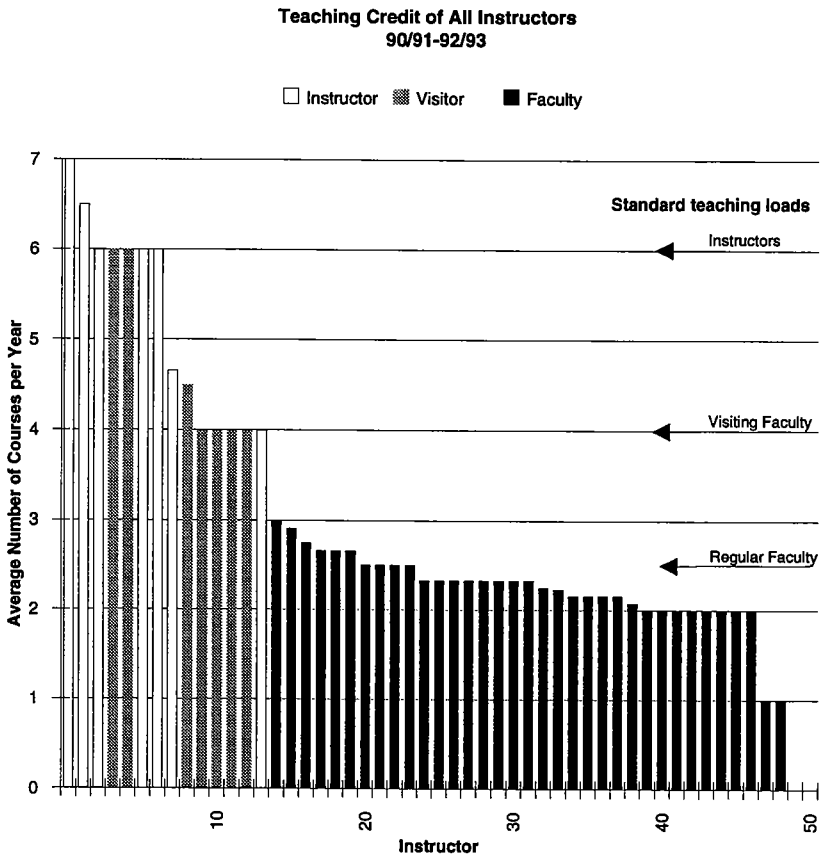
1. *Random variation.* Fractions enter from the standard load, from leaves of a fraction of a year, from fractional course buyouts, or fractional reductions in teaching from administrative work,

etc. These fractions mean that even over a period some will teach a little more, some a little less.

2. *New hire reduction.* New faculty are often given a small, temporary reduction in teaching for a first year; this happened for a short time.

3. *Administrative release.* Most larger departments have administrative positions that provide a reduced teaching load. There can be as high as 100% for some people; examples of such positions are:

   Department Head,
   Associate/Assistant Department Head,
   Directors of substantial research or educational centers,
   Chairs of Graduate, Personnel, Undergraduate, and similar committees.

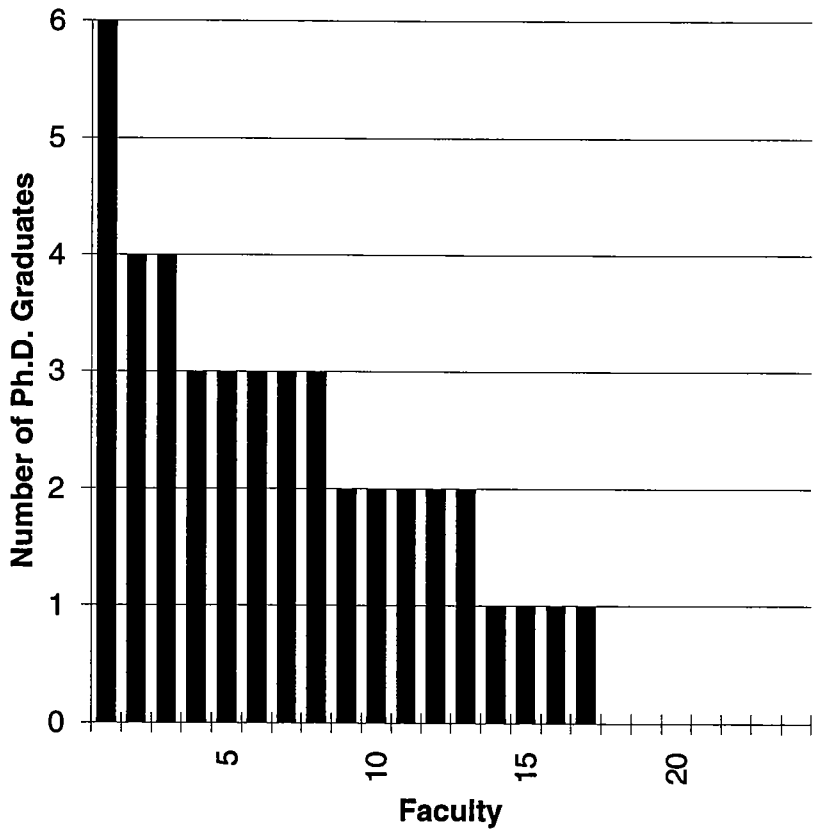


Data Definition: The data is the same as the previous graph except that faculty (black), visitors (shaded), and instructor or A/P staff (white) are included, everyone except graduate students. No recitation sections are counted. Instructors on campus for only part of the period or with part time teaching loads are included and their loads normalized to reflect their part time status.

Purpose: Measure the teaching contribution of all the instructional staff.

Comments: The standard teaching load often varies among different groups of instructors. Here different loads are indicated for visiting professors (shown as four courses per year) and the Instructor rank (six courses per year). It is common for non-standard arrangements to be made with temporary instructional staff. Note that many of these people are part time or short term, so the total number of courses taught during this three year period is much less than suggested by adding up all the credits shown above.

Ph.D. Graduates of Tenured Faculty 1989-93

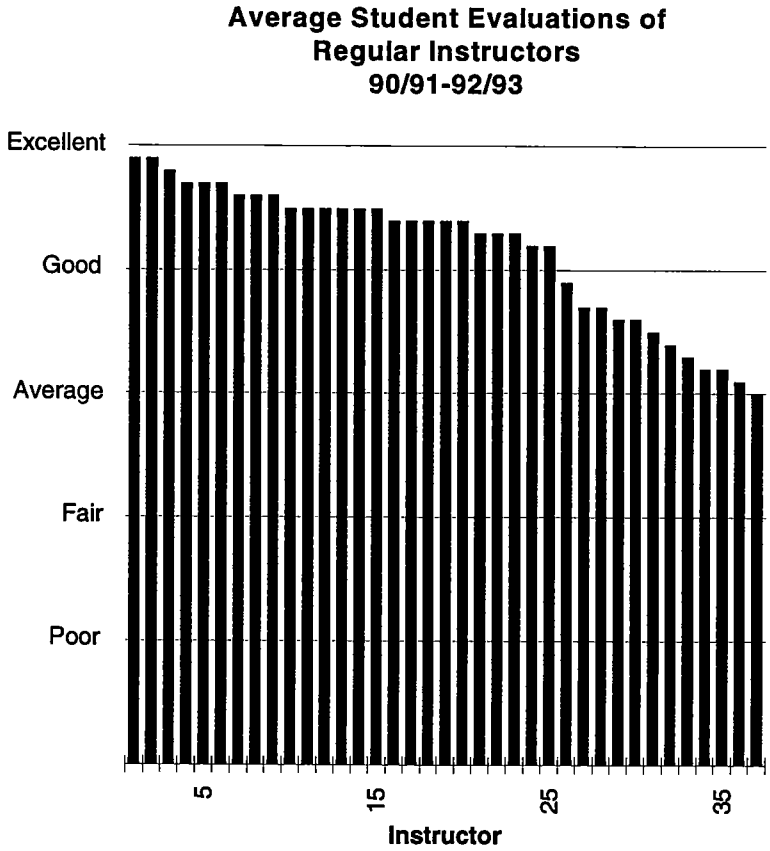


Data Definition: The number of Ph.D.s supervised and awarded during the five year period 1988–1993 is given for the tenured faculty.

Purpose: Measure the involvement of the tenured faculty with training professionals in their fields.

Comments: The normal rate of producing Ph.D. students varies greatly among fields; even within a field there may be substantial differences among subfields. Other normal sources of variation are more personal, e.g.,

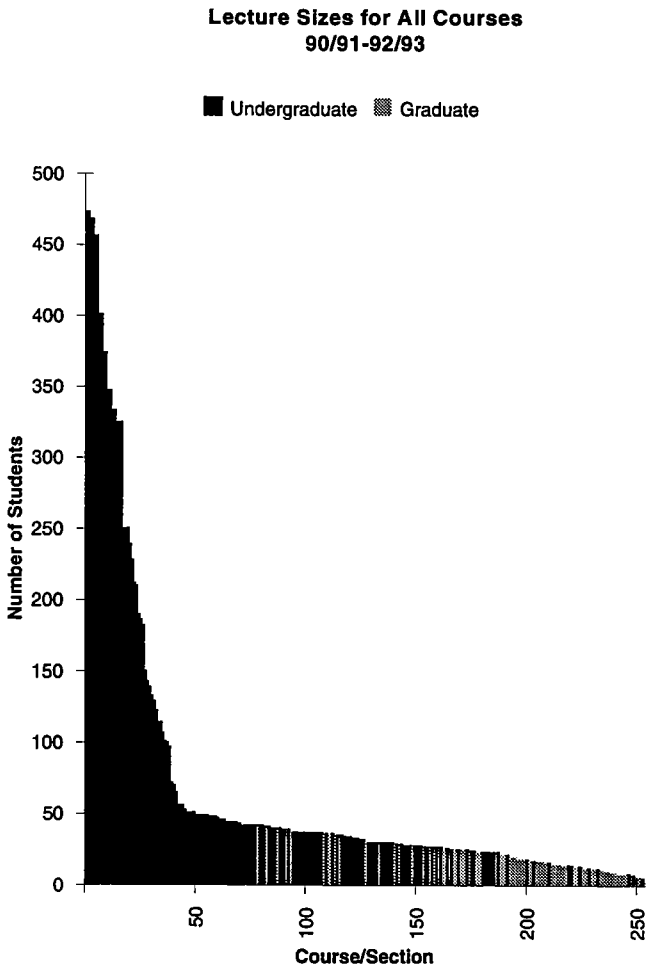
- Those just tenured have had little time for Ph.D. students to get through the pipeline,
- Some areas of research are not popular with students.



Data Definition: These data are averages for each regular instructor for the three year period 1990/91–1992/93. The student responds to the questions "Overall, I rate this instructor as ..." or something similar. Short term visitors and students are not counted as regular instructors.

Purpose: Measure the quality of instruction as perceived by the students taking courses.

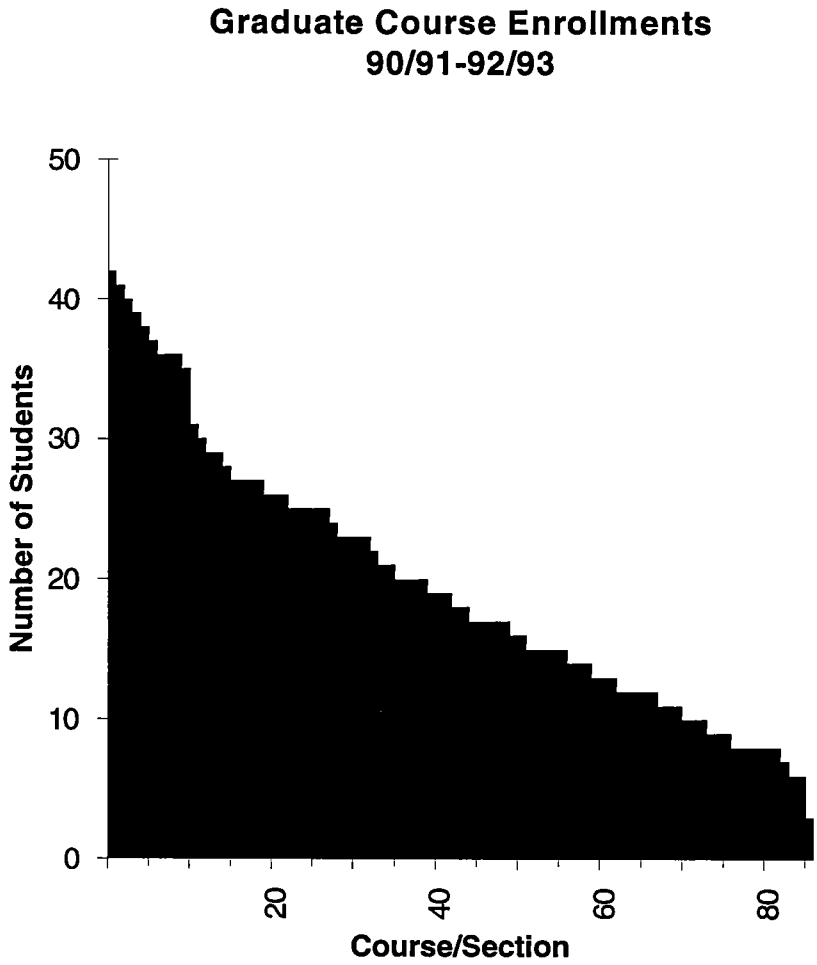
Comments: It is usually required that all instructors in a department participate in student evaluations for essentially all courses; a few evaluations are not reported when the number of student respondents is too small. These data are usually much better than "expected" based on the general discontent one hears about college teaching. Almost everyone receives at least a satisfactory evaluation (although one wonders how all instructors can be rated above average). A more detailed study of student evaluations shows that the lower evaluations primarily occur in lower division courses.



Data Definition: The sizes of lectures are given for all courses taught from 1990/91 to 1992/93. Undergraduate courses are shown in solid and graduate courses are lightly shaded. Excluded are jointly listed courses where the students actually attend a course taught in another department.

Purpose: Measure the sizes of the individual classes taught in the department.

Comments: The division between large service courses and smaller courses is easily seen and is typical for many departments in larger universities. The distinction can be even more pronounced in those departments with very large service teaching loads and a small number of majors. Many departments have a minimum course size (perhaps 7 or 10) but smaller sizes can occur as size is usually measured well after the decision is made about offering a course.

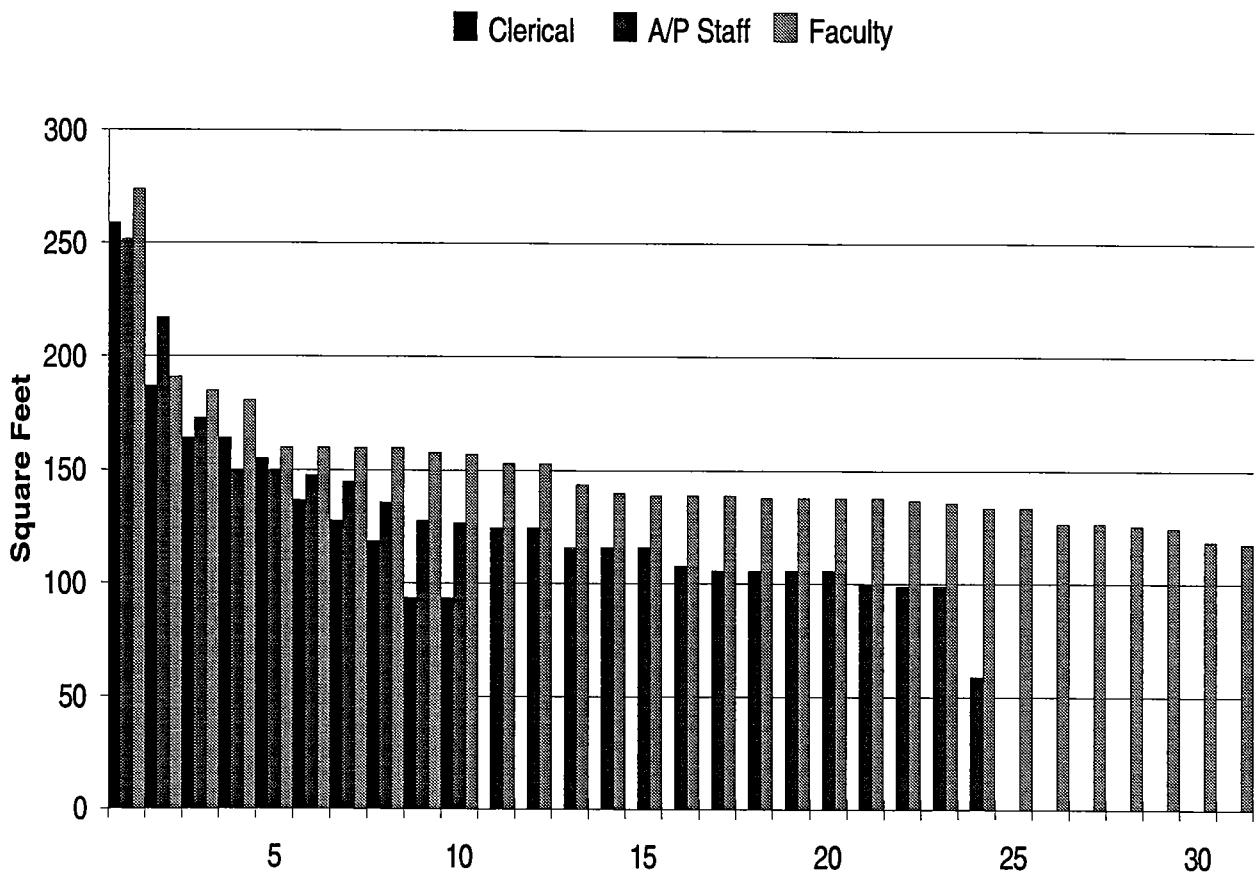


Data Definition: These data are a subset of the previous data.

Purpose: Measure the size of courses in the department's graduate program.

Comments: Graduate enrollments (and hence class sizes) are usually small enough that only single sections are given of a course. Some departments conduct specialized seminars as courses which can generate a large number of very small course sizes.

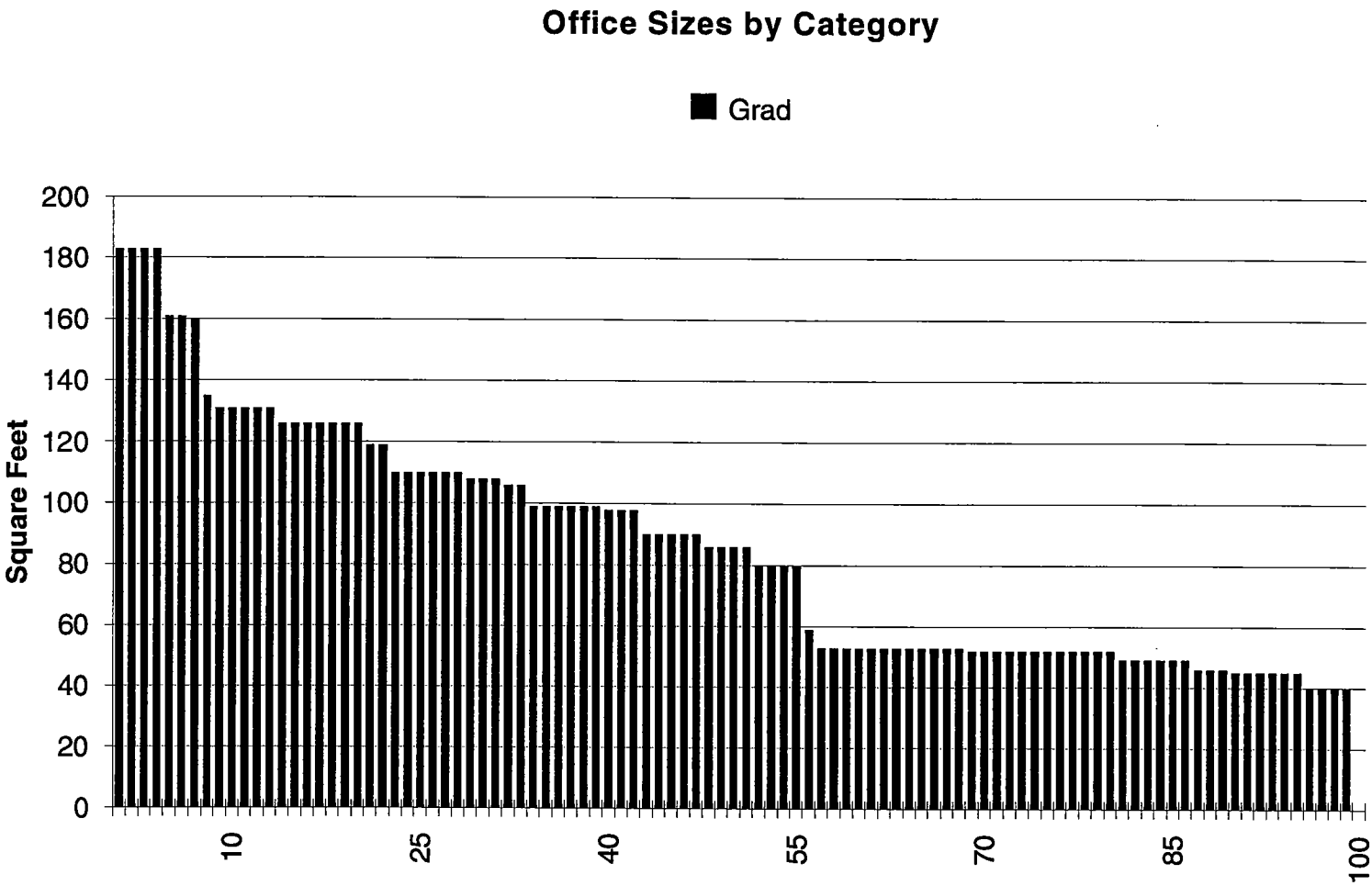
Faculty and Staff Office Sizes by Category



Data Definition: The number of square feet is given for each office. The data are given for three categories of people: Clerical, A/P Staff, and Faculty. If there are multiple people in an office, the space is divided evenly among the occupants.

Purpose: Measure the space used by offices for the department's faculty and staff.

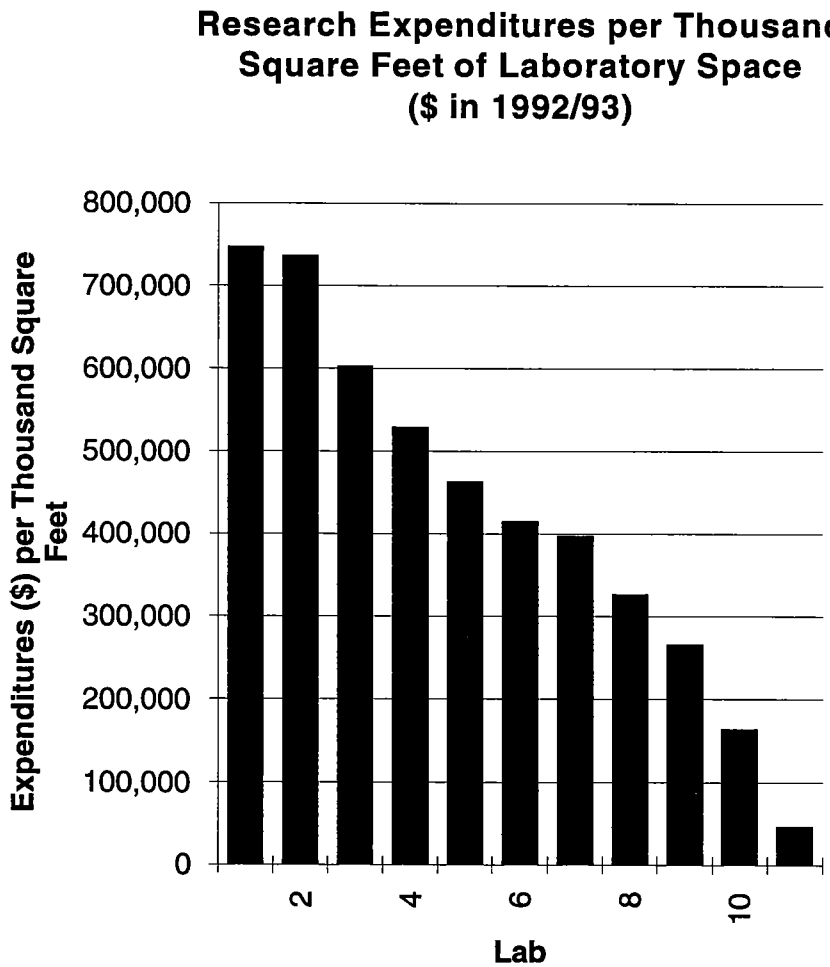
Comments: These data do not show the effect of having office machines or other equipment in the offices. Most really large offices result from this phenomenon. Occasionally there are offices designed for two people that contain only one person — or three persons.



Data Definition: The number of square feelt of office space is given for each graduate student with assigned office space. Since many graduate student offices are places (desks, cubicles, etc.) in large rooms and laboratories, the office size is computed by dividing the total area by the number of occupants.

Purpose: Measure the office space used by the department's graduate students.

Comments: Students in rooms and labs with substantial equipment or public areas appear to have large offices. Few students actually have "large" offices and most universities consider 50–80 square feet as the normal office space for a student employee. Most students with assigned office space are employees of the department.

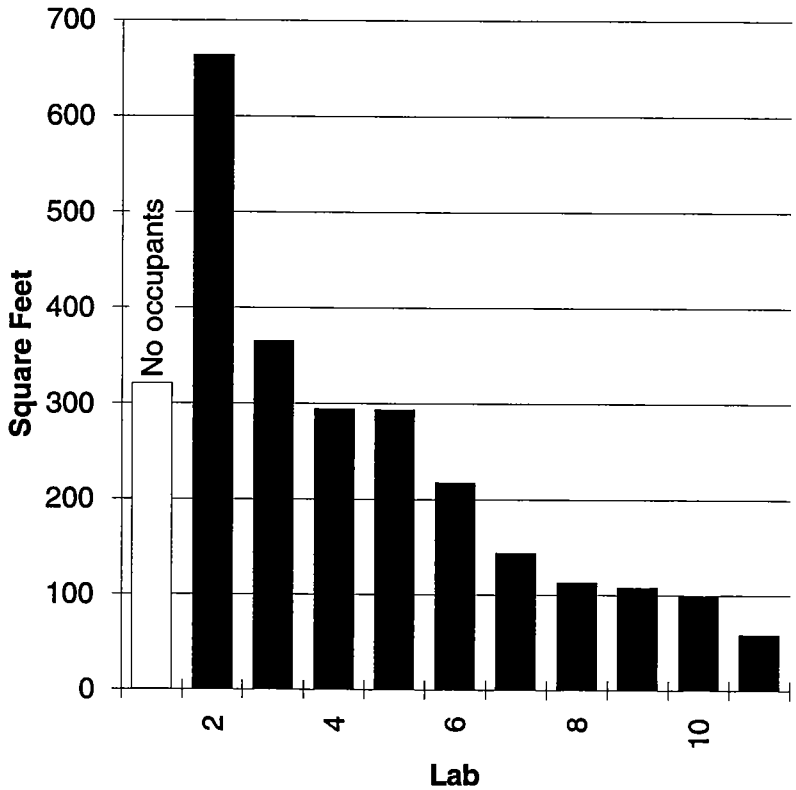


Data Definition: The amount of research expenditures per thousand square feet is given for each of the department research labs. The quantities for labs with multiple projects or projects with multiple labs are apportioned in a direct way.

Purpose: Measure the external research income associated with space in the department's research laboratories.

Comments: Faculty and other people (e.g., visitors, secretaries) associated with a laboratory might not have offices in the labs. The necessary size of a laboratory depends on the equipment used more than the dollars spent. Thus these expenditures should be expected to vary widely. Nevertheless, if the research expenditures of a laboratory are very low, then the specific circumstances of the situation are worthy of further study.

Research Lab Space per Research-Supported Occupant 92/93

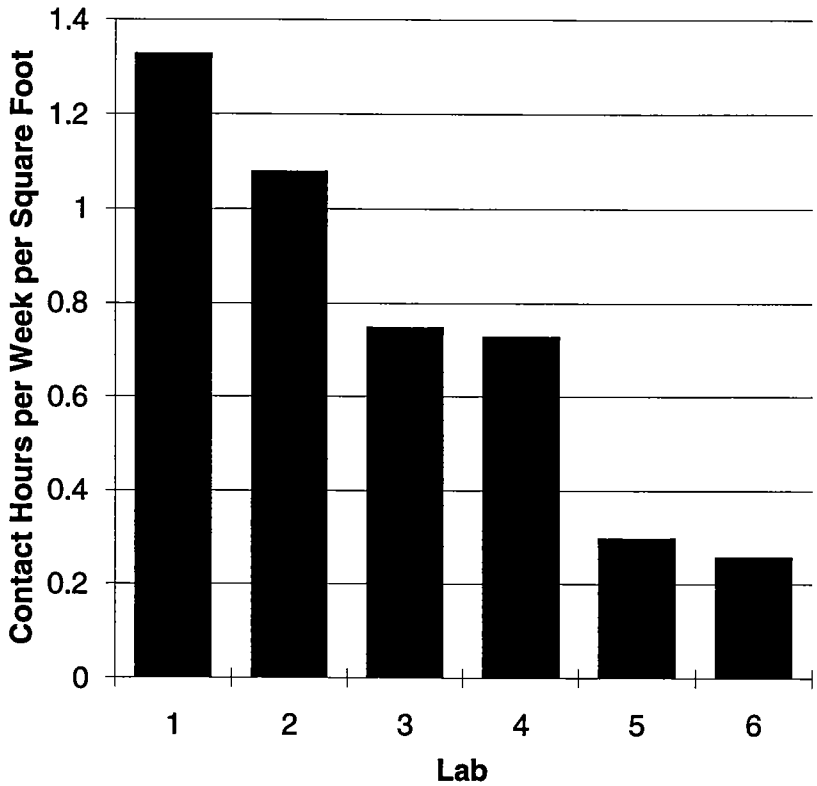


Data Definition: The number of research supported occupants for each lab is divided into the lab size to provide an office size "equivalent utilization" for the research labs. The data is for 1992/93.

Purpose: Measure the utilization of research laboratory space as office space for research-supported occupants.

Comments: Laboratories also include space for equipment and most have a small area with table for group meetings. Most labs also have non research-supported occupants such as TA's working on their research. Some laboratories have no occupants.

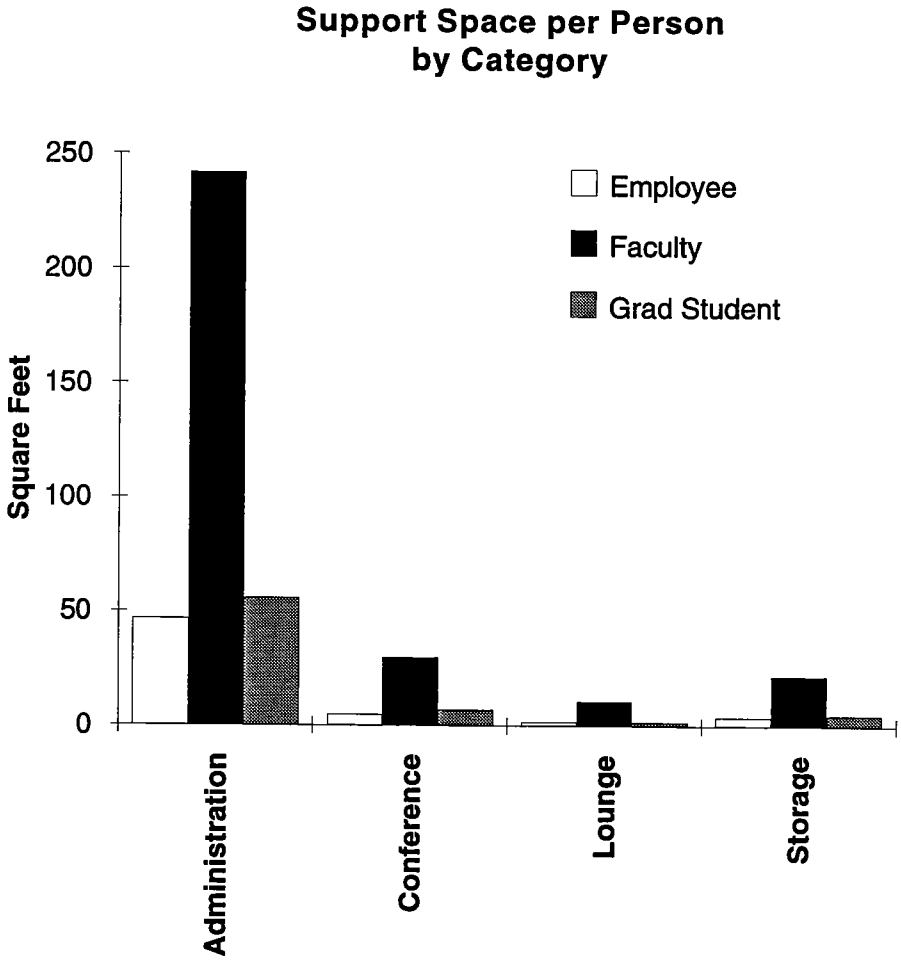
Weekly Contact Hours in Teaching Labs per Square Foot 92/93



Data Definition: The number of scheduled student contact hours per week in a lab is divided by the lab size (in square feet). The data are for 1992/93.

Purpose: Measure the utilization of departmental teaching laboratories.

Comments: Some type of laboratories also have open hours where students are free to work as they please; this use is not measured here. For comparison, a 600 ft^2 classroom scheduled 45 hours per week with 20 students/class has 1.5 contact hours per week per square foot. The space utilization also depends on the size of the laboratory equipment; in a few cases the equipment can be very large. Nevertheless, since space is perhaps the scarsest resource in a university, if the utilization is very low then the specific circumstances of the situation are worthy of further study.



Data Definition: The general purpose support space of the department is categorized as follows:

Administration, Conference rooms, Lounges, Storage

The administration category includes offices. The size of these spaces are then shown per "user" from three categories:

Employee, Faculty, Graduate Majors

Purpose: Measure the department use of space relative to the size of the department.

Comments: Undergraduate majors should also be used as a fourth "user" group. They were not included here because the bars for them are vanishingly small. The space for storage can vary greatly from field to field, the other three categories should vary much less.