CALIFORNIA'S SCIENCE BASE: SIZE, QUALITY AND PRODUCTIVITY

A REPORT PREPARED FOR

THE CALIFORNIA COUNCIL ON SCIENCE AND TECHNOLOGY

MICHAEL R. DARBY, PH.D.

THE WARREN C. CORDNER PROFESSOR OF MONEY AND FINANCIAL MARKETS

THE JOHN E. ANDERSON GRADUATE SCHOOL OF MANAGEMENT

UNIVERSITY OF CALIFORNIA, LOS ANGELES

RESEARCH ASSOCIATE, NATIONAL BUREAU OF ECONOMIC RESEARCH

LYNNE G. ZUCKER, PH.D.

PROFESSOR OF SOCIOLOGY AND POLICY STUDIES

UNIVERSITY OF CALIFORNIA, LOS ANGELES

RESEARCH ASSOCIATE, NATIONAL BUREAU OF ECONOMIC RESEARCH

FEBRUARY 3, 1999

ABOUT THE CCST CALIFORNIA REPORT ON THE ENVIRONMENT FOR SCIENCE AND TECHNOLOGY

CCST's California Report on the Environment for Science and Technology (CREST) has analyzed the state's science and technology infrastructure to determine if California has the people, capital investment and necessary state governmental policies to maintain California's leadership in the face of increasing worldwide competition. Through eight individual research projects, CREST analyzes the state's ability to create and use new technology. By facilitating a dialog with policy makers, industry leaders, and academic communities, CCST hopes to enhance economic growth and quality of life for Californians.

ACKNOWLEDGMENTS

The California Council on Science and Technology gratefully acknowledges support from the W. M. Keck Foundation of Los Angeles, CCST's Sustaining Members, and the State of California.

The authors are indebted to two talented research assistants Xiaogang Wu, who developed the NRC and ISI university data sets, and Gihong Yi, who developed the ISI state data set.

COPYRIGHT

 Library of Congress Cataloging Number in Publications Data Main Entry Under Title:

California's Science Base: Size, Quality and Productivity

ISBN 1-930117-08-6

The California Council on Science and Technology is a nonprofit orgainization established in 1988 at the request of the California State Government and sponsored by the major post secondary institutions of California, in conjunction with leading private-sector firms. CCST's mission is to improve science and technology policy and application in California by proposing programs, conducting analyses, and helping government implement policies and initiatives for a better economy and quality of life.

The writers of this publication can be reached for questions or comments on content at the

California Council on Science and Technology
1130 K Street, Suite 280
Sacramento, California 95814

by voice at
(916) 492-0996

by fax at
(916) 492-0999

or e-mail at
ccst@ccst.ucr.edu

To order additional copies of this report or other CCST publications contact 909.787.2913 or email ccst@ccst.ucr.edu.

Table of Contents

List of Figures, Map and Tables

1. Introduction

Since the national policy decisions consequent upon Vannevar Bush's 1945 report, American economic growth has been propelled by federal support for basic, largely academic research, to create the flow of scientific and engineering knowledge which underlies technological progress and productivity growth. California has supplemented this national policy by supporting a system of world-class public universities, which, in combination with the state's outstanding private universities, have provided it with one of the finest science bases for economic development in the world.

Science and engineering have both a universal and local dimension. Discovery of basic laws of nature result in knowledge which can be shared by other scientists and engineers everywhere. However, new – especially breakthrough – knowledge is frequently tacit in nature and difficult to explain to others, inevitably embodied in the discoverers and those who actually do bench science with them. During the period that new knowledge is systemized and diffused, it is likely to be applied commercially near where it is discovered. Social scientists have documented the advantages of this "geographically localized knowledge" to the research productivity and market success of firms located near the great research centers.[1] Indeed, the very best academic scientists are often directly involved in commercialization of their discoveries through participation as principals or consultants to firms newly established for that purpose.[2]

2. Overview of California's Science Base and Strategy

To be a leader in cutting edge technologies where the greatest growth is occurring, a state must have an excellent science base which provides the source for continuous technological advances. This report investigates quantitative indicators of California's science base both in comparison to the nation as a whole and to other high-tech states. Our first set of indicators is based on data collected in preparing the 1993 National Research Council (NRC) study of U.S. research-doctorate programs (Goldberger, Maher, and Flattau 1995, see Data Appendix A for details). These data basically cover the situation as of the late 1980s and early 1990s and are the most recent comprehensive data allowing comparison of the quantity and quality of the academic science base for technological progress.

The adequacy of the science base as a determinant of growth in technological employment must be judged relative to the number of people depending on that base for geographically localized economic growth effects. Ph.D.s produced, doctorate program faculty, federal grants, publications, and citations per publication for all (science and engineering) doctorate programs in California are summarized in Figure 1 as a percentage of the national average per capita values.[3] The upper bar for each variable represents the California value while the lower bar of the pair is the average value for all 13 high-tech states.[4] California – like the high-tech states generally – ranges 15 to 20 percent above the national norm on most indicators except that faculty numbers at 85% of the national average are unusually low while citations per publication (the standard measure of scientific impact or quality) at 133% is high not only relative to the national norm but also relative to other high-tech states: With fewer faculty than the national average, California produces as many Ph.D.s, federal grants, and publications per capita as other high-tech states and those publications are of substantially higher quality.

This apparent strategy of concentrating on fewer and more stellar academics is confirmed dramatically in Figure 2 which examines these same indicators when only doctorate programs ranked by the NRC among the top ten are considered. As noted above, faculty and doctoral students in the best programs are the most likely source of the scientific breakthroughs which sometimes lead to formation of whole new industries and transformation of existing industries. Thus, they play a special role in conferring localized economic development benefits on the areas which possess them. Not surprisingly given this prior literature, top-ranked programs are much more prominent among high-tech states than in the nation generally. However, for top-ten-program Ph.D.s, faculty, federal grants, and publications, California produces between 150 and 200% more than the national average per capita while high-tech states generally produce only a bit over 50% above the national average. Even citations per publication are 14% above the national average (compared to 3% for high-tech states as a whole). Thus California's science base is built upon relatively few university faculty members of extraordinary quality.

1 See particularly, Jaffe (1989), Jaffe, Trajtenberg, and Henderson (1993), Edwin Mansfield (1995) and Zucker, Darby, and Brewer (1998).

2 Zucker and Darby (1996) and Zucker, Darby, and Armstrong (1998). These "star" scientists' involvement appears to play a major role in determining which firms utilizing breakthrough discoveries will be most successful. Interestingly, these scientists often publish more and better science during the period of their academic involvement, apparently due to the greater resources which result from their commercial activities.

3 Citations per publication are not on a per capita basis because they are normalized per publication.

4 For purposes of this study, high-tech states are defined as California, Connecticut, the DC-plex (DC, Maryland, Virginia), Illinois, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, Texas, Washington, and Wisconsin.

3. California Compared to Other High-tech States

California as the most populous state has a far larger science and engineering base than any of the other major-science states whether we consider all research doctorate programs or only top-ten programs as illustrated in Figures 3 and 4. However, California's superiority disappears on a per capita basis: On this basis, California is squarely in the middle among the other high-tech states in terms of science indicators based on all programs (Figure 5). For the population-adjusted indicators and top-ten ranked programs, California generally can claim second place to Massachusetts (Figure 6), although Connecticut, Washington, and Wisconsin are much closer to California than California is to Massachusetts.[5] The two leaders are compared on the population-adjusted, top-ten basis in Figure 7. Massachusetts's dominance derives primarily from the large number of faculty both overall (197% of the national average) and particularly on the top-ten basis (531% versus 252% for California and 152% for all high-tech states).

4. California's Science Base by Major Scientific Fields

Scientific and engineering disciplines are more relevant to particular industries than others, and we have accordingly divided the major disciplines into five distinct fields or clusters as detailed in Table 1: Biology/Medical/Chemical ("Bio/chem" for short), Computer Software/Information Processing/Interactive or Multi-Media--technical ("Information Technology"), Integrated Circuits/Semiconductors/High-Temperature Superconductors ("Semi/superconductors"), Other Engineering, and Other Science. Figures 8 and 9 report California's science base indicators by each of these major fields for all doctorate programs and for top-ten ranked programs only.

Map 1 plots the total number of top-ten science and engineering doctorate programs in both the 1982 and 1993 NRC studies for each of the high-tech states.[6] This single statistic appears to be the most comparable measure across time using the NRC data base. California clearly dominates on this total number, but Massachusetts is second in 1993 even without correction for population. Figure 10 plots the population-adjusted shares of top-ten programs overall and for all five major science and

Table 1. High-Technology Science Base Categories

High-Technology Science Base Categories
(A) "Bio/chem" (Biology/Medical/Chemical)
51 Biochemistry & Molecular Biology 52 Cell & Developmental Biology 53 Molecular & General Genetics 56 Ecology, Evolution & Behavioral 55 Pharmacology 23 Chemistry 62 Biomedical Engineering 32 Chemical Engineering 54 Neurosciences 57 Physiology
(B) "Information Technology" (Computer Software/Information Processing/ Interactive or Multi-Media--technical)
26 Computer Sciences 25 Mathematics
(C) "Semi/superconductors" (Integrated Circuits/Semiconductors/ High-Temperature Superconductors)
24 Physics 30 Electrical Engineering 34 Materials Science 31 Mechanical Engineering
(D) Other Engineering
61 Aerospace Engineering 33 Civil Engineering 63 Industrial Engineering
(E) Other Science
65 Oceanography 29 Astrophysics/Astronomy 28 Statistics/Biostatistics 27 Geoscience

California and the high-tech state average. Generally high-tech states had about 150% higher numbers of top-ten programs than would be predicted by their populations compared to 250 to 300% for California. We see that overall California declined only slightly in its share of top-ten programs (from 278 to 275% of the national average). California noticeably increased its share in the Biology/Medical/Chemical cluster and decreased its share in both the Computer Software/Information Processing/Interactive and Multi-Media--technical and Integrated Circuits/Semiconductors/ High-Temperature Superconductors clusters.

In Figure 11, we indicate the strength of each of the high-tech states overall and for three largest of the five major fields in terms of population-adjusted shares of top-ten programs in both the 1982 and 1993 studies. Connecticut fell sharply overall (from 412% in 1982 to 223% in 1993) primarily due to loss of a dominant position in the information technology cluster. The DC-

[5] Note that Ohio has no top-ten ranked doctoral programs in either the 1993 or 1982 study.

[6] The number of programs for Other Engineering and Other Sciences was too small for the numbers to be meaningful for most states.

plex made notable progress led by the Biology/Medical/Chemical cluster.

5. Using Publications Data to Track Trends over Time

As we have seen, the 1993 NRC study provides rich detail on the academic science base around 1988-1992 but limited comparability with the 1982 NRC study. Since the data on publications and citations are highly correlated with the other indicators in the 1993 NRC study, we now will use extracts from the Science Citation Index (SCI) data base of the Institute for Scientific Information (ISI) to track those variables over time by university and by state.[7]

It is tricky to follow trends in quantity and quality of a science field over time by tracking publications and citations per publication because the more time that has elapsed, the greater the possible citations to a given publication. Researchers at the ISI have developed a reasonable way to solve this problem partially by using a five-year moving window. Publications are counted as the total publications within the five years and citations are only citations to those publications which occur within the same five years. While this is only a fraction of the total citations that will ultimately be received, it should serve as a reasonable measure of changes in quality.

We examine first the ISI data based on the top 110 universities in terms of federal grants. These are reasonably consistent with the NRC data set in terms of universities covered, but it is impossible to get breakdowns by individual research doctorate programs. Experimentation revealed that we could not get consistency between university program totals and totals by journal subject for the ISI data at the discipline level (universities frequently have faculty from many programs writing in the same discipline). However, for their overlap period, totals from the two sources aligned reasonably well at the level of the five broad science areas aggregated to state totals.[8]

Figure 12 shows the trend from 1985 through 1997 for California and high-tech states generally in terms of publications per 1000 population and citations per publication, where the year refers to the final year of the 5-year moving window for which publications and the corresponding citations are measured. We see that major universities in the average high-tech state have reduced California's lead in both quantity and quality (citations/publication) of publications in science and engineering over the 13-year period.

We have attempted to control for quality of publications in another way, by counting publications and their citations at a major university only if that university has at least one NRC top-ten rated doctorate program in the same broad filed of science and engineering. Figure 13 reports publications per 1000 population and citations per publication for the aggregate of such "top-ten university articles" with the caveat that the quality screen is considerably less stringent than we could and did apply for the NRC data. On this basis, California has been more successful at maintaining a quantitative lead over and qualitative parity with the average high-tech state.

Figure 14 illustrates that when we consider all science and engineering publications -- including those from private industry, research institutes, and non-major universities – the high-tech state average per capita publication rate went above California's after 1984-88 but California did manage to maintain or widen its lead in citations per publication.

Figure 15 compares California with Massachusetts, the leading high-tech state on a per capita basis, in terms of publications per 1000 population and citations per publication for major universities with one or more top-ten NRC-ranked doctorate programs in the same broad area of science and engineering. In Figure 13 California was holding its own on these measures versus the average high-tech state, but we see that Massachusetts is increasing its lead both in publications per capita and in citations per publication. This trend has negative implications for California's competitive position versus Massachusetts for new cutting-edge technologies and industries.

6. Comparison among High-tech States over Time

As suggested by the 1993 NRC study, when we consider only publications at major universities in Figure 16, Massachusetts dominates on a per capita basis, followed in a distant second by Connecticut. We note, however, that California drifts over time from third into a pack with the DC-plex and Washington state.

[7] The ISI has two machine-readable data bases (see Data Appendixes B and C for details) available which approximately serve the purpose. One tracks the top 110 universities (selected on the basis of federal grants received) which can be aggregated to state totals for 37 states and the DC-plex and the other gives the variables by state. The difference between the two is an indication of the publishing activity by authors affiliated with other universities, firms, and research institutes. However, this difference can be misleading since an article is included in the major university total if it has any authors affiliated with them regardless of the participation of other authors and because where co-authors are from two different universities (of the top 110) in the same state the article will appear twice in the university aggregates.

[8] Comparing the NRC and ISI university data sets at the major science field and state level of aggregation, we obtained correlation coefficients of 0.968 and 0.979 for total publications and total citations, respectively.

Figure 17 reports publications per 1000 population for major universities with one or more top-ten NRC-ranked doctorate programs in the same broad area of science and engineering. On this measure, Connecticut has pulled ahead of California in the 1990s to take second place while Washington state has nearly pulled even with California to tie for third by 1993-97.

When we consider all state publications in Figure 18, the DC-plex ranks a much closer second to Massachusetts, probably because of the prodigious output of government scientists and engineers.[9] We cannot break out these different groups definitively using the ISI data set. California slips from 5th to 7th ranking among the 12 states plus DC-plex.

7. California's Science Base by Major Scientific Fields

Finally we consider how California is doing by major scientific field over time versus the average high-tech state and Massachusetts in Figures 19-23. For this purpose we focus on publications per 1000 population and citations per publication for major universities with one or more top-ten NRC-ranked doctorate programs in the same broad area of science and engineering. In every case except for Information Technology (Figure 20) California leads the average high-tech state but trails Massachusetts in publications per capita; in the exceptional case California has failed to keep up with the rising high-tech state average. In terms of the quality measure (citations per publication) equals or exceeds the high-tech state average in every case and significantly trails Massachusetts only for the Biology/Medical/Chemical area. Thus California has largely been holding its relative position except relative to Massachusetts in the Biology/Medical/Chemical area and relative to the average high-tech state in Information Technology.

8. Conclusions

California has built a science base that is the largest in the nation and, after adjusting for total state population, among the top two or three in terms of richness and productivity of top scientists. It is these top scientists who are most likely to make and exploit commercially those breakthrough discoveries which transform existing industries and create new ones. California has accomplished this despite limited resources by concentrating resources in comparatively few but excellent doctorate programs in the state's great public and private research universities.

However, over the last decade of budgetary stringency, California has not managed to maintain its relative position. In particular, California has lost important ground to Massachusetts in nearly every area of science and engineering, including the commercially crucial biological/medical/chemical, semi/superconductor, and information technology fields. Indeed in the latter programs, California's top scientists have fallen behind the national per-capita average for their peers.

Thus the glass is both half full and half empty. California's research universities and other scientific base is providing the state a great competitive advantage in this increasingly technological age. But that advantage is slowly eroding not in absolute terms, but relative to other states which are devoting the resource to build the science base for the Route 128s and Silicon Valleys of the next technological revolution.

[9] The only other state with substantial non-major university publication is New Jersey, where pharmaceutical and other science-based firms attract many excellent scientists and engineers. Moving from the university to state based measures, New Jersey advances 3 to 5 places from dead last.

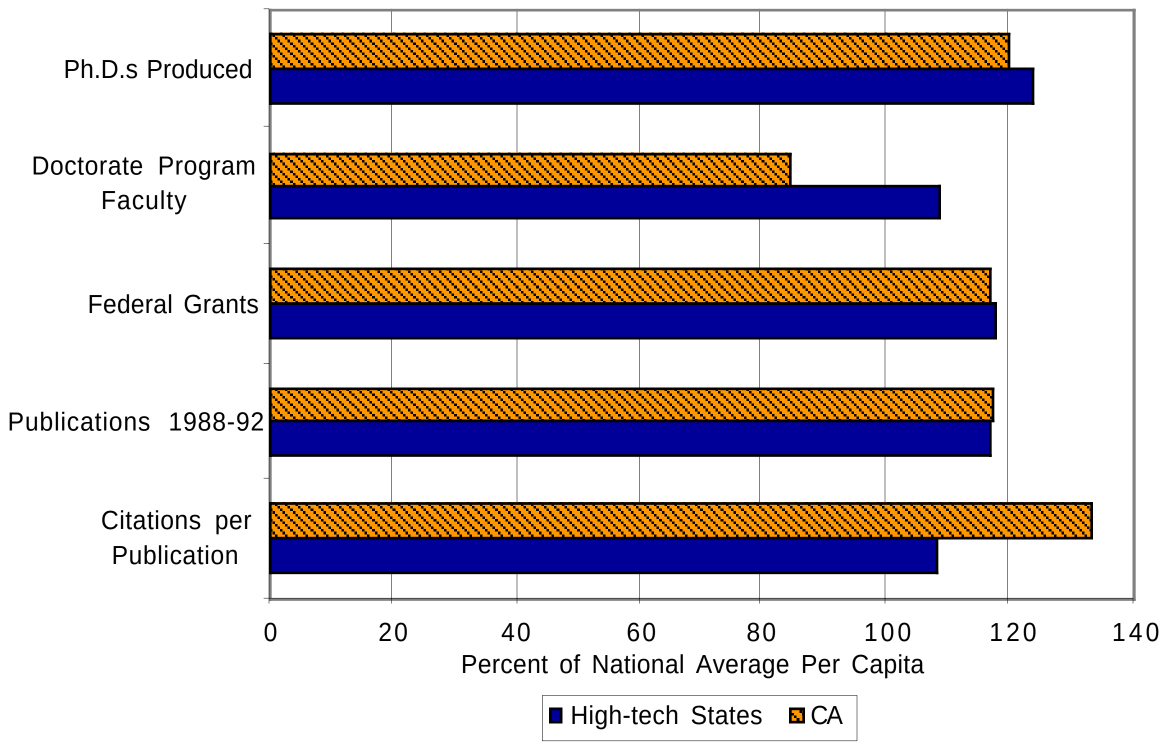


Figure 1. California's Science and Engineering Base: All Research Doctorate Programs, 1993 Study

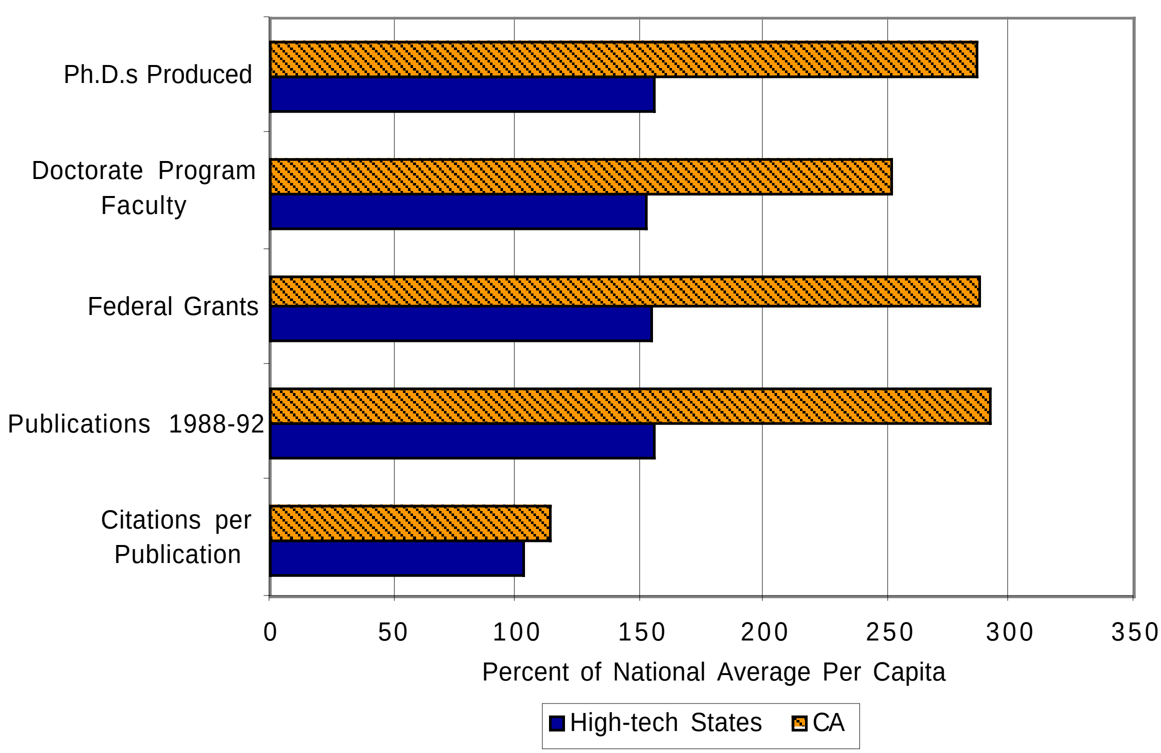


Figure 2. California's Science and Engineering Base: Top-ten NRC Ranked Research Doctorate Programs, 1993 Study

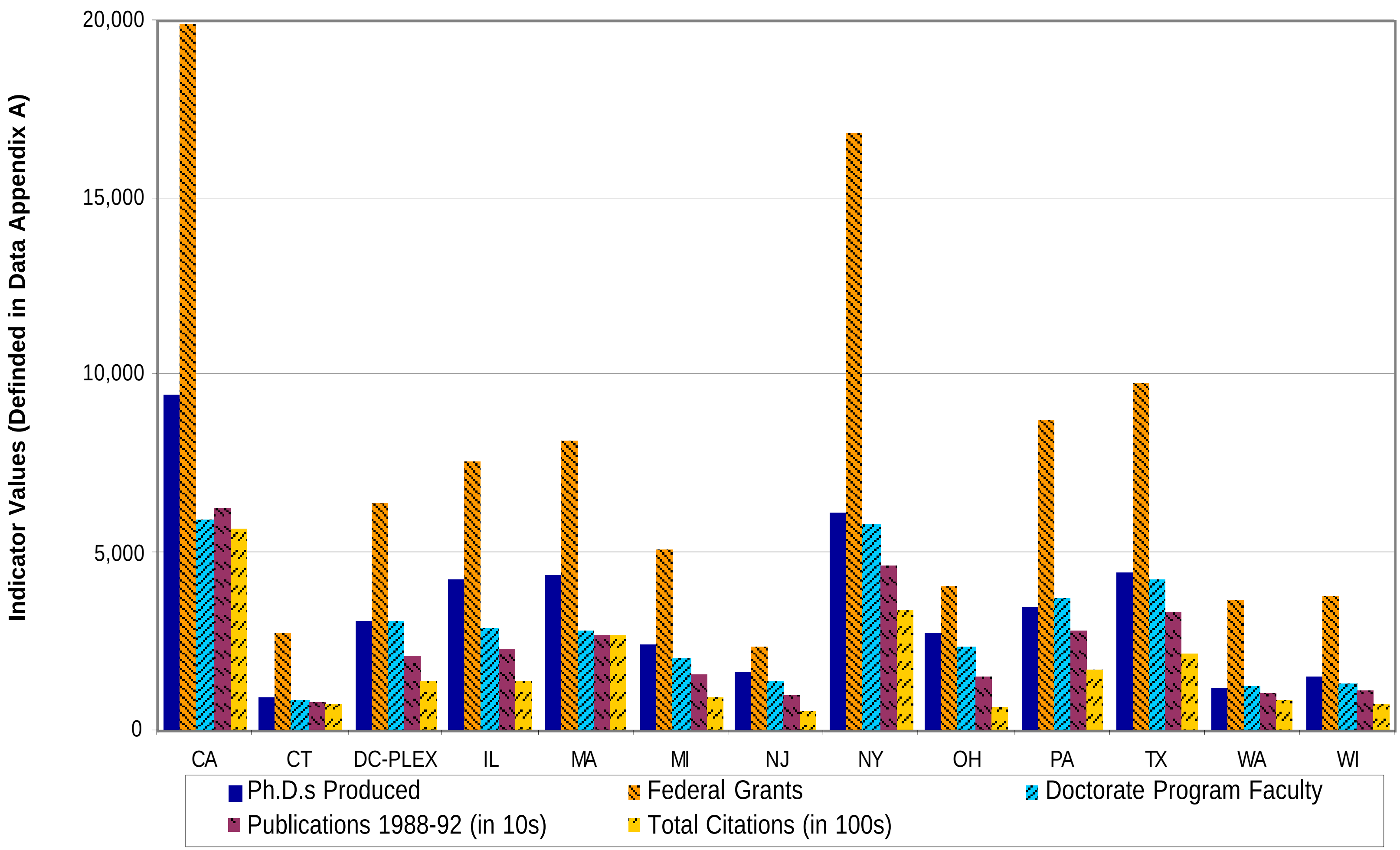


Figure 3. Science Base Indicators for High-tech States: All Research Doctorate Programs, 1993 Study (Totals Not Adjusted for Population)

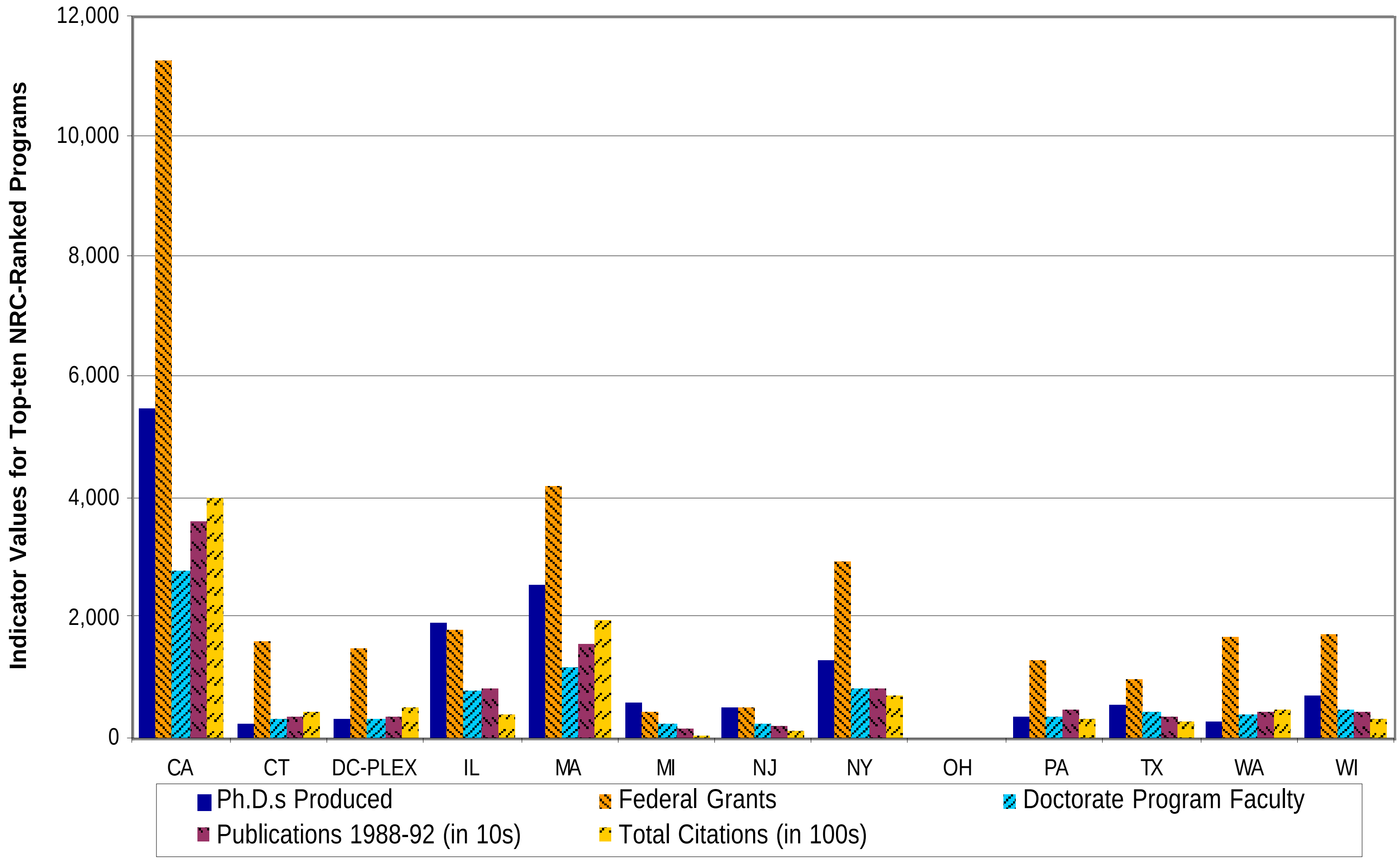


Figure 4. Science Base Indicators for High-tech States: Top-ten NRC Ranked Research Doctorate Programs, 1993 Study (Totals Not Adjusted for Population)

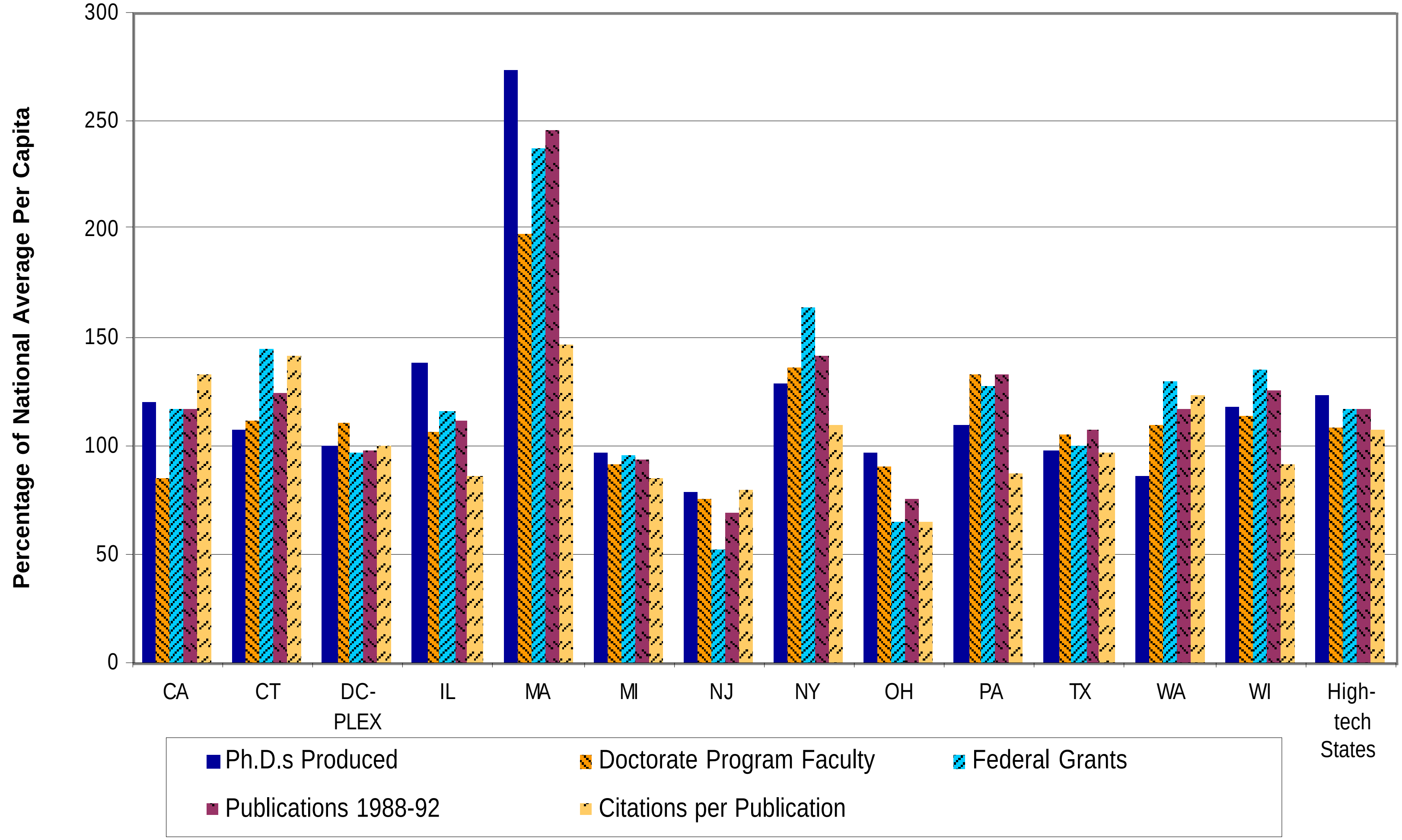


Figure 5. Science Base Indicators for High-tech States: All Research Doctorate Programs, 1993 Study

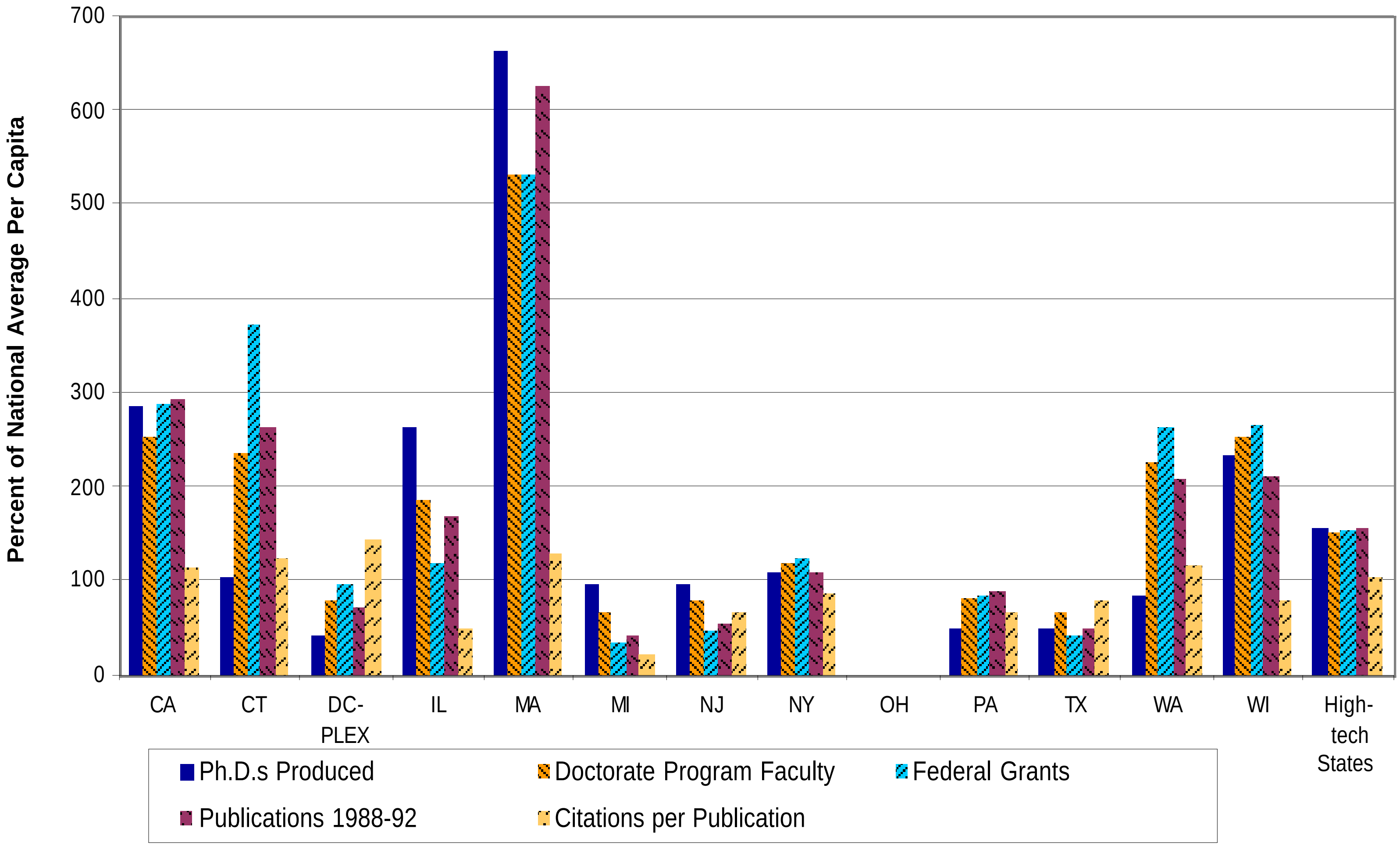


Figure 6. Science Base Indicators for High-tech States: Top-ten NRC Ranked Research Doctorate Programs, 1993 Study

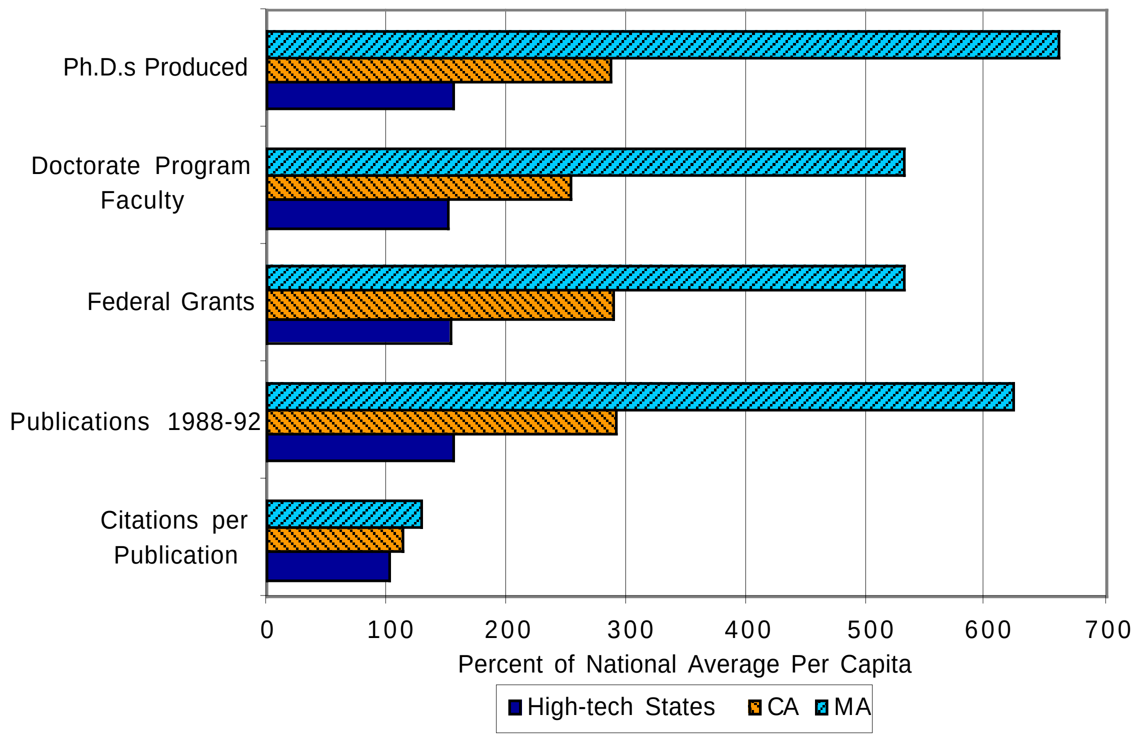


Figure 7. California's versus Massachusett's Science and Engineering Base: Top-ten NRC Ranked Research Doctorate Programs, 1993 Study

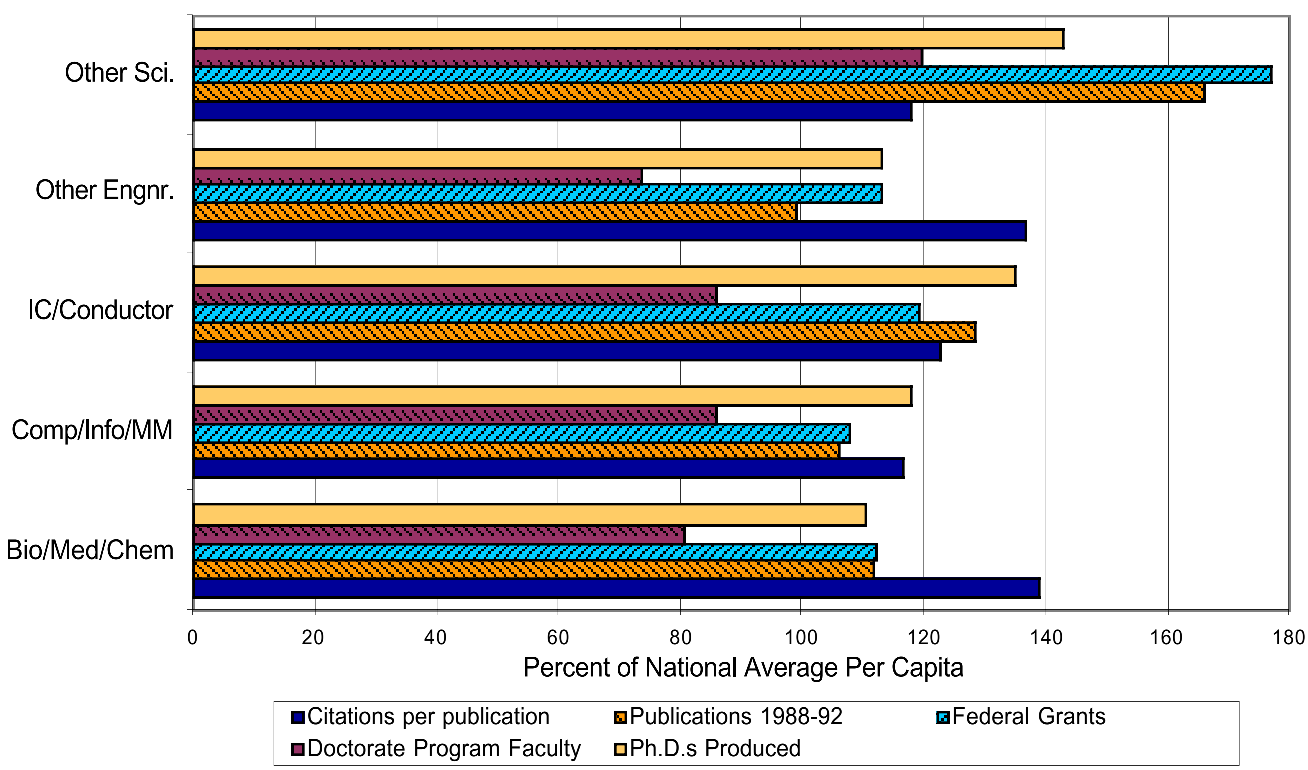


Figure 8. California's Science and Engineering Base by Major Field: All Research Doctorate Programs, 1993 Study

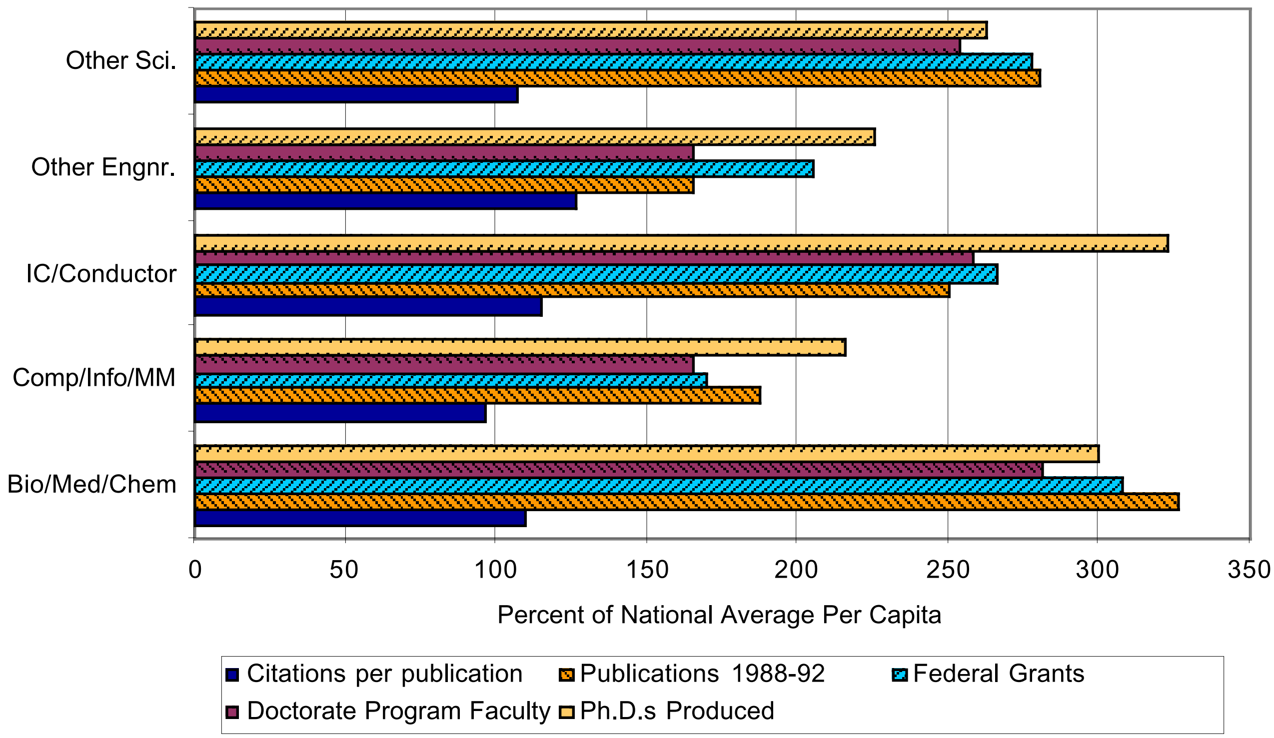


Figure 9. California's Science and Engineering Base by Major Field: Top-ten NRC Ranked Research Doctorate Programs, 1993 Study

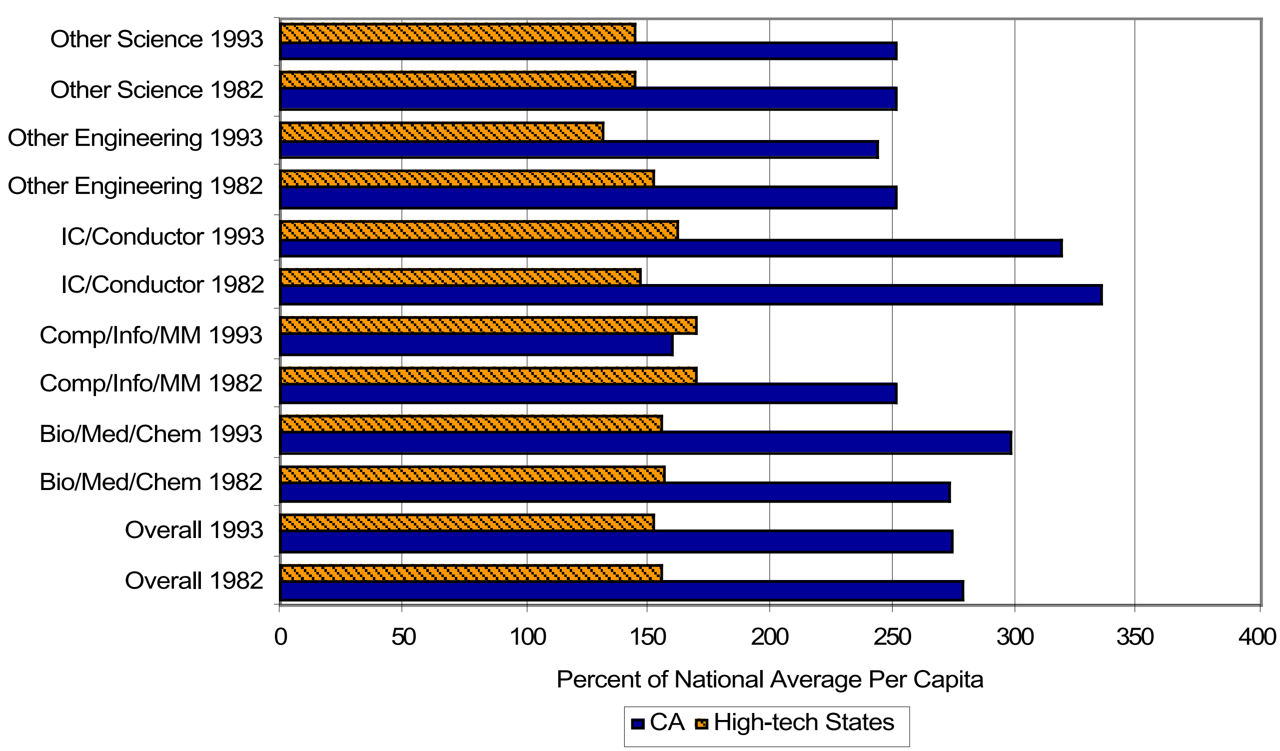


Figure 10. Population-Adjusted Shares of Top-ten NRC Ranked Doctorate Programs: Overall and by Major Science & Engineering Field California, versus High-tech State Average, 1982 and 1993 Studies

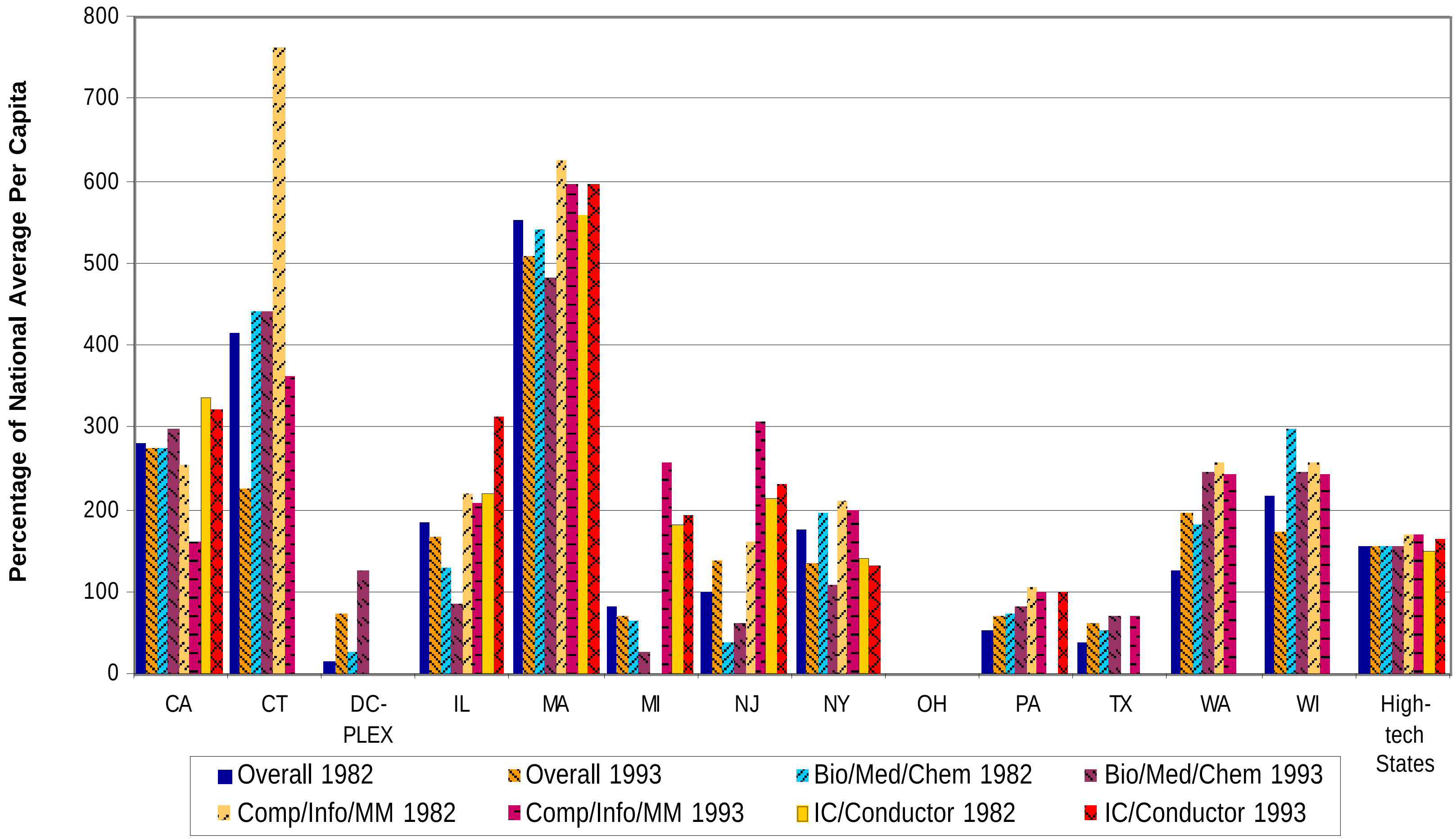


Figure 11. Share of Top-ten NRC Ranked Doctorate Programs in 1982 and 1993 Studies as Percentage Share of 1988-92 National Population: Overall and by Three Major Science & Engineering Fields

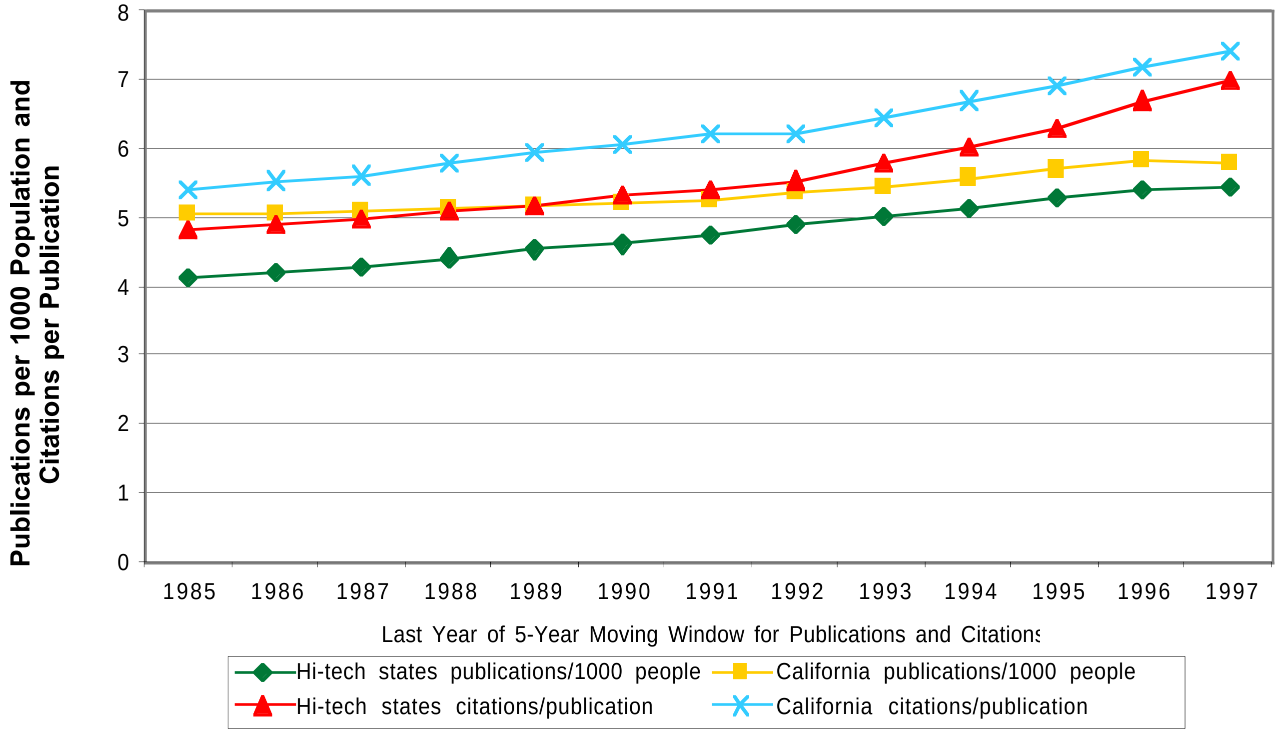


Figure 12. Trends in Major-University Publications and Citations per Publication: California versus High-tech States Average, 1981-1985 through 1993-1997

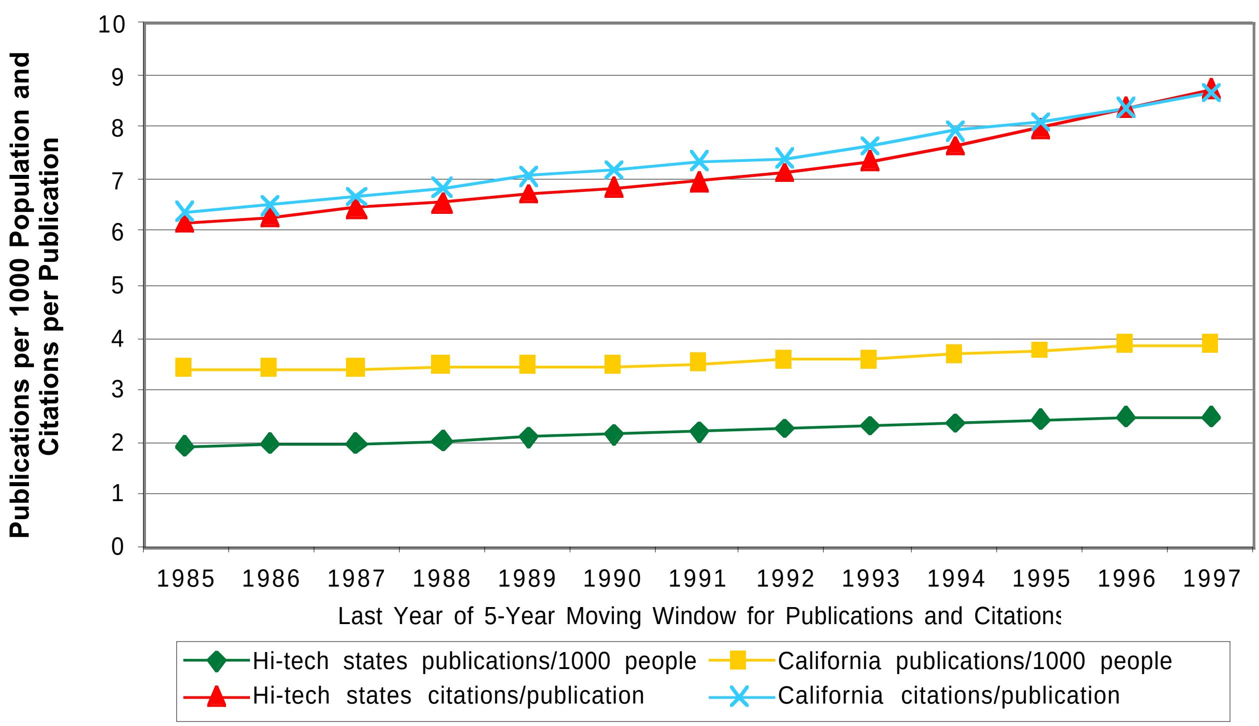


Figure 13. Trends in Publications and Citations per Publication at Major Universities with Top-ten NRC Ranked Programs in Science & Engineering Area: California versus High-tech States Average, 1981-1985 through 1993-1997

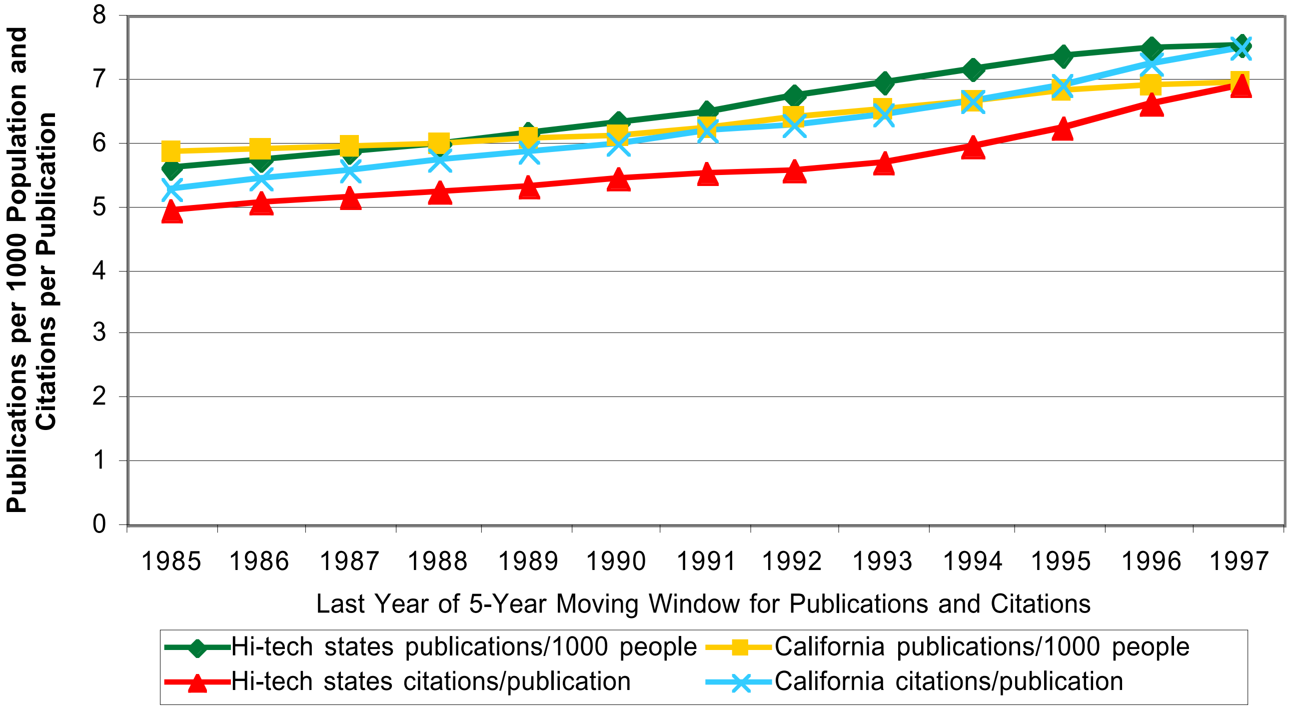


Figure 14. Trends in State Total Publications and Citations per Publication: California versus High-tech States Average, 1981-1985 through 1993-1997

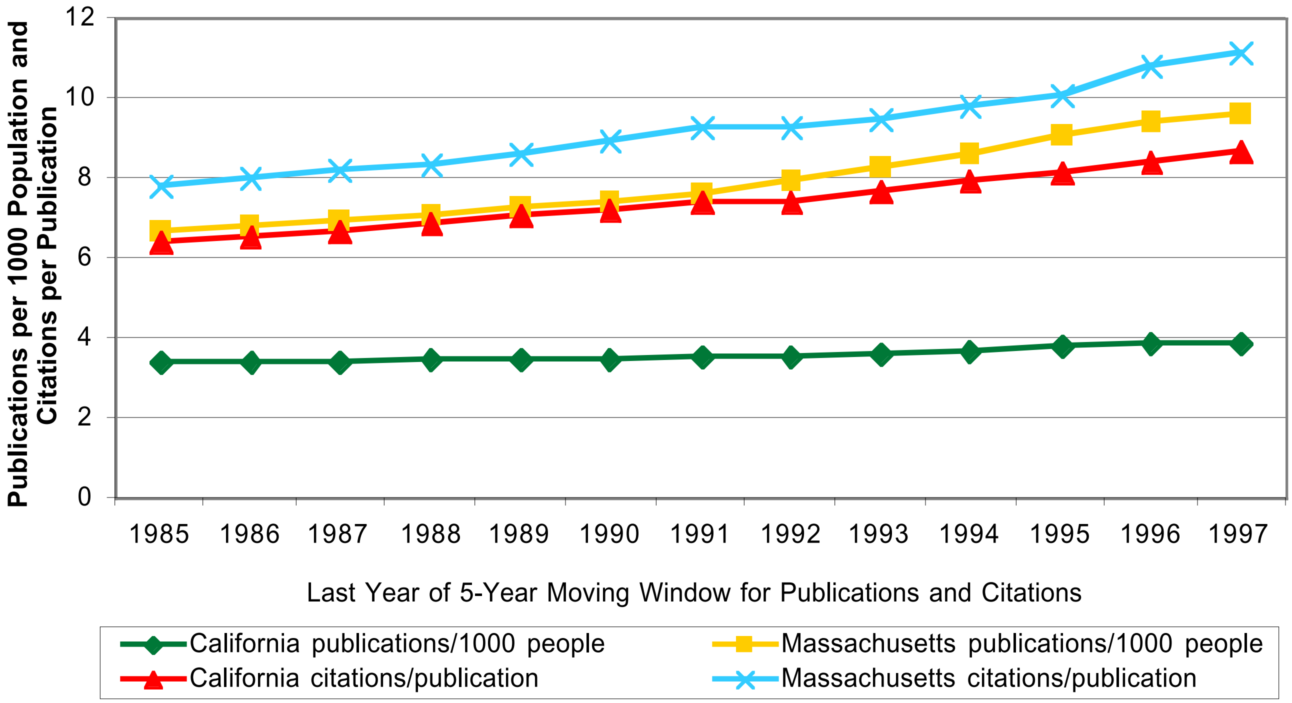


Figure 15. Trends in Publications and Citations per Publication at Major Universities with Top-ten NRC Ranked Programs in Science & Engineering Area: California versus Massachusetts, 1981-1985 through 1993-1997

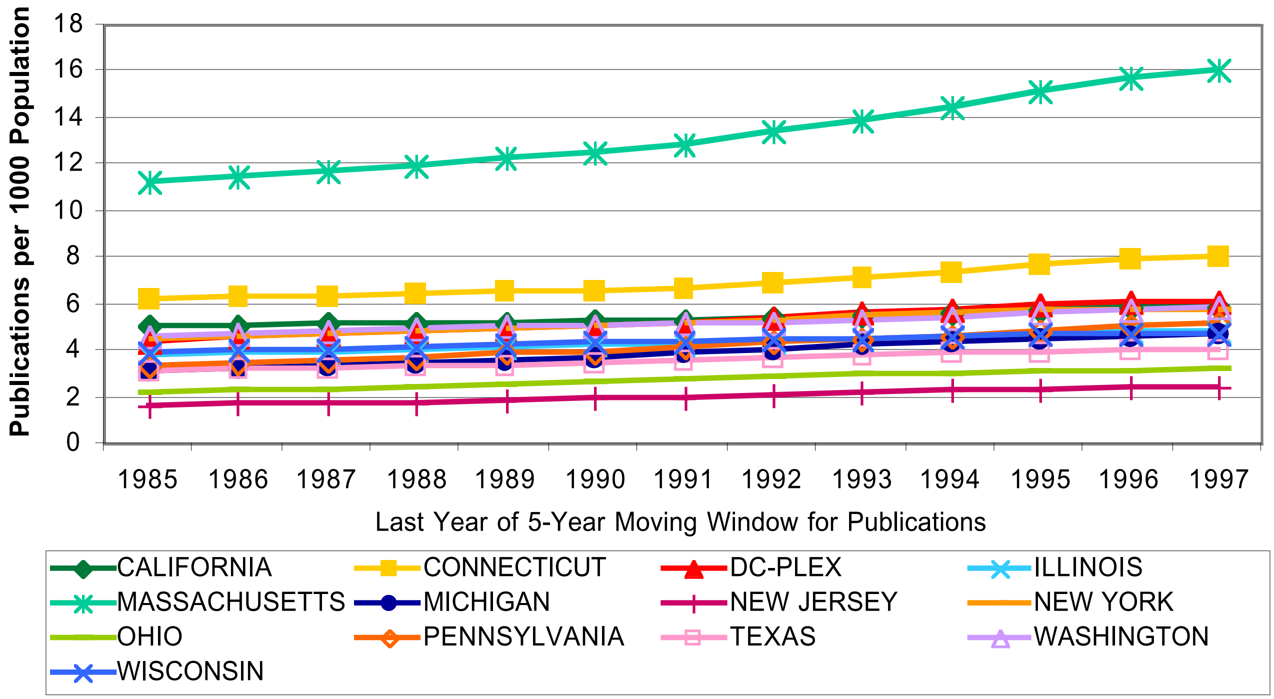


Figure 16. Major-University Publications per 1,000 People: All High-tech States, 1981-1985 through 1993-1997

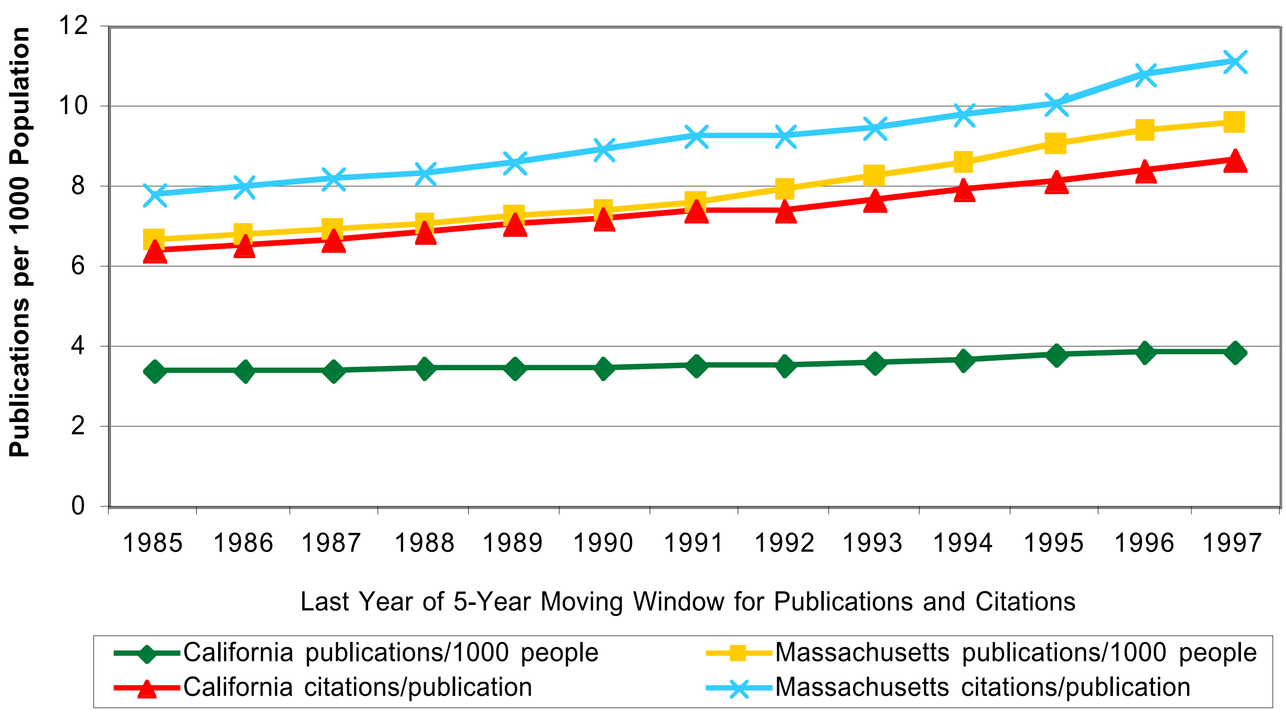


Figure 17. Publications at Major Universities with Top-ten NRC Ranked Programs in Science & Engineering Area: All High-tech States, 1981-1985 through 1993-1997

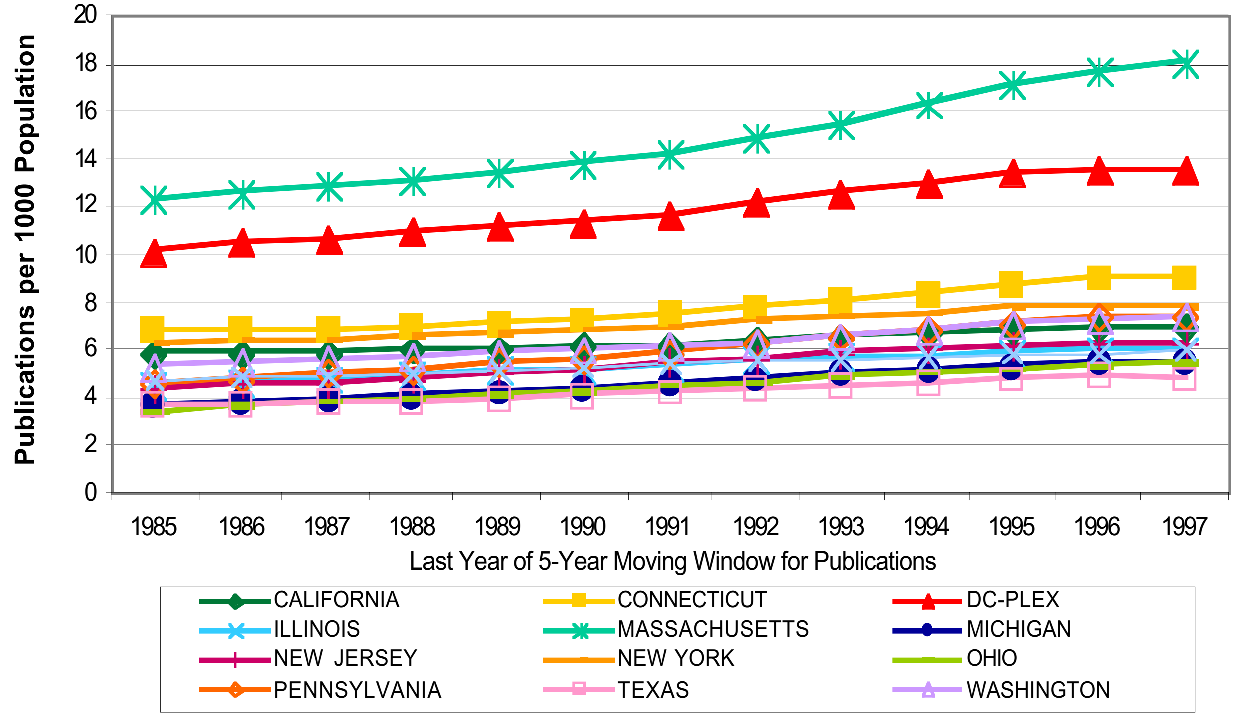


Figure 18. State Total Publications per 1,000 People: All High-tech States, 1981-1985 through 1993-1997

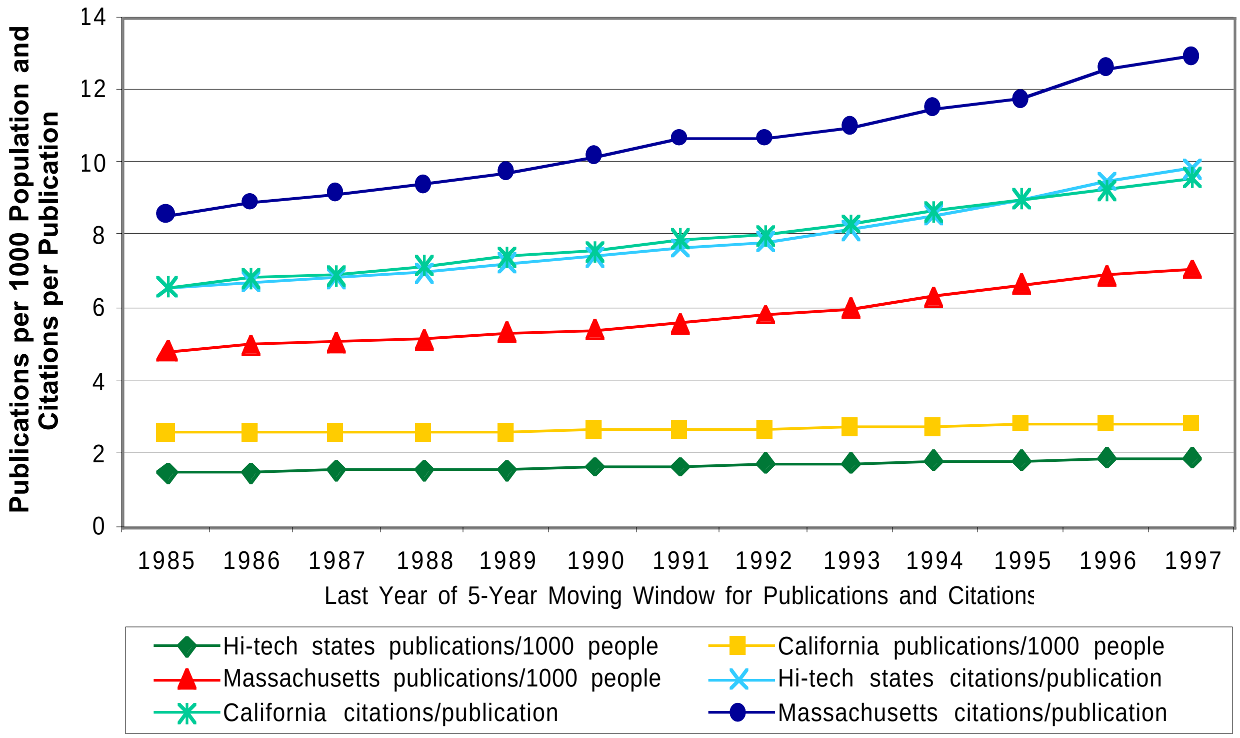


Figure 19. Trends in Publications and Citations per Publication at Major Universities with Top-ten Ranked Programs in Biology/Medical/Chemical: California, Massachusetts, and High-tech States Average, 1981-1985 through 1993-1997

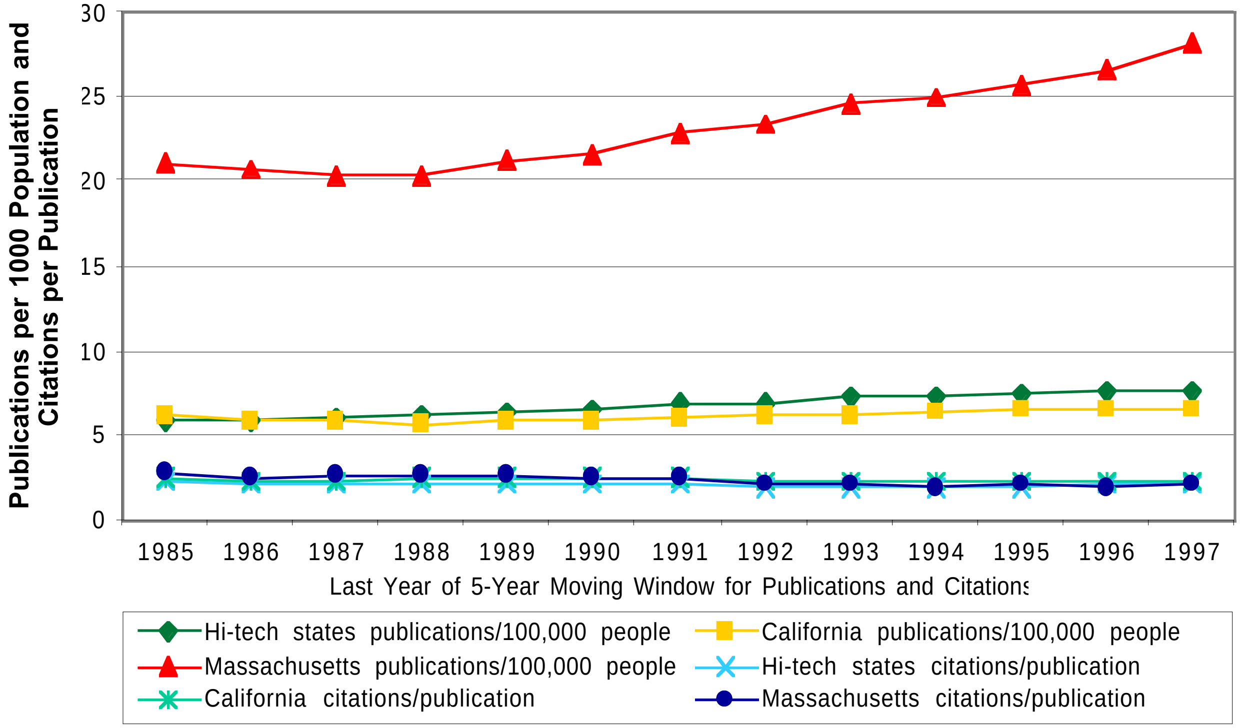


Figure 20. Trends in Publications and Citations per Publication at Major Unversities with Top-ten Ranked Programs in Information Technology: California, Massachusetts, and High-tech States Average, 1981-1985 through 1993-1997

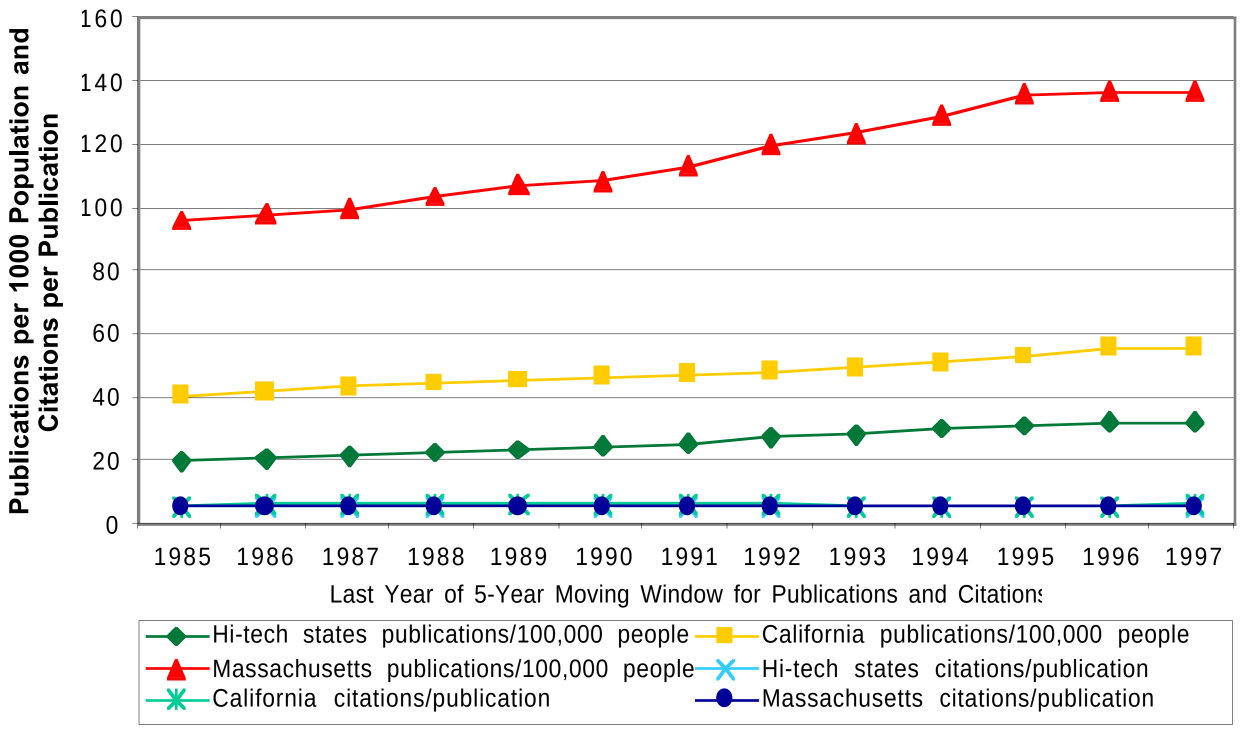


Figure 21. Trends in Publications and Citations per Publication at Major Universities with Top-ten NRC Ranked Programs in Semi- & Super-Conductors: California, Massachusetts, and High-tech States Average, 1981-1985 through 1993-1997

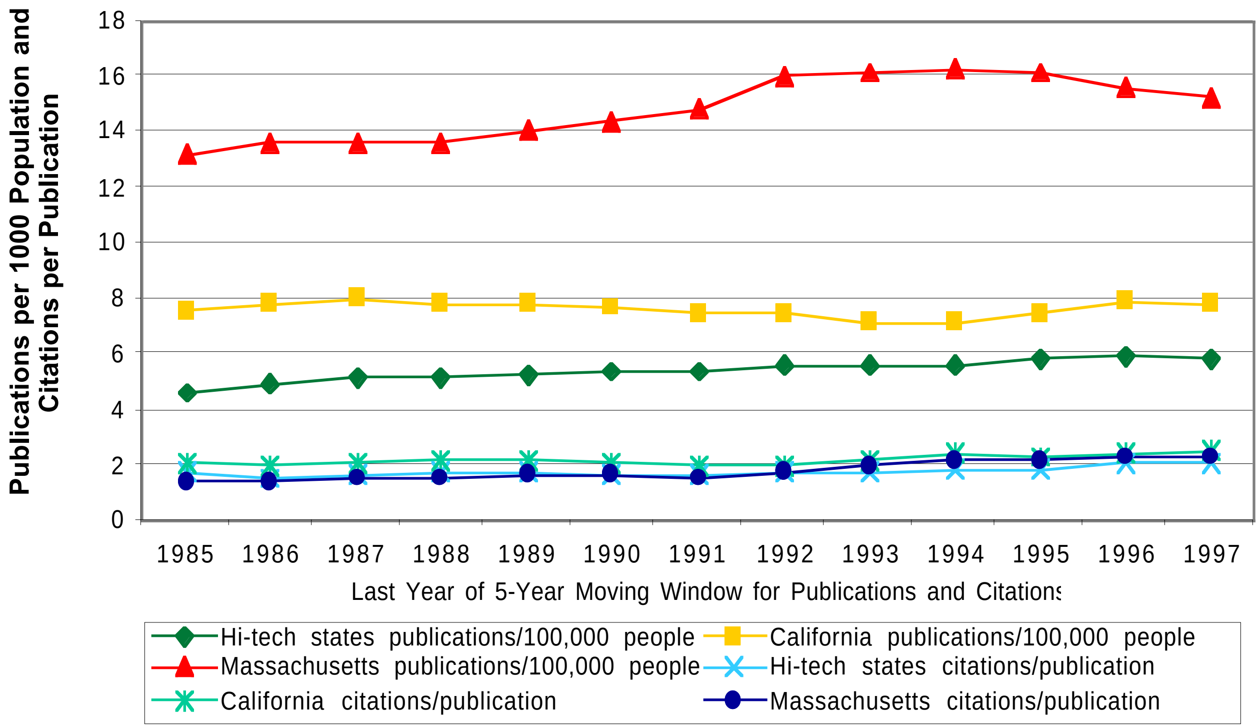


Figure 22. Trends in Publications and Citations per Publication at Major Universities with Top-ten NRC Ranked Programs in Other Engineering: California, Massachusetts, and High-tech States Average, 1981-1985 through 1993-1997

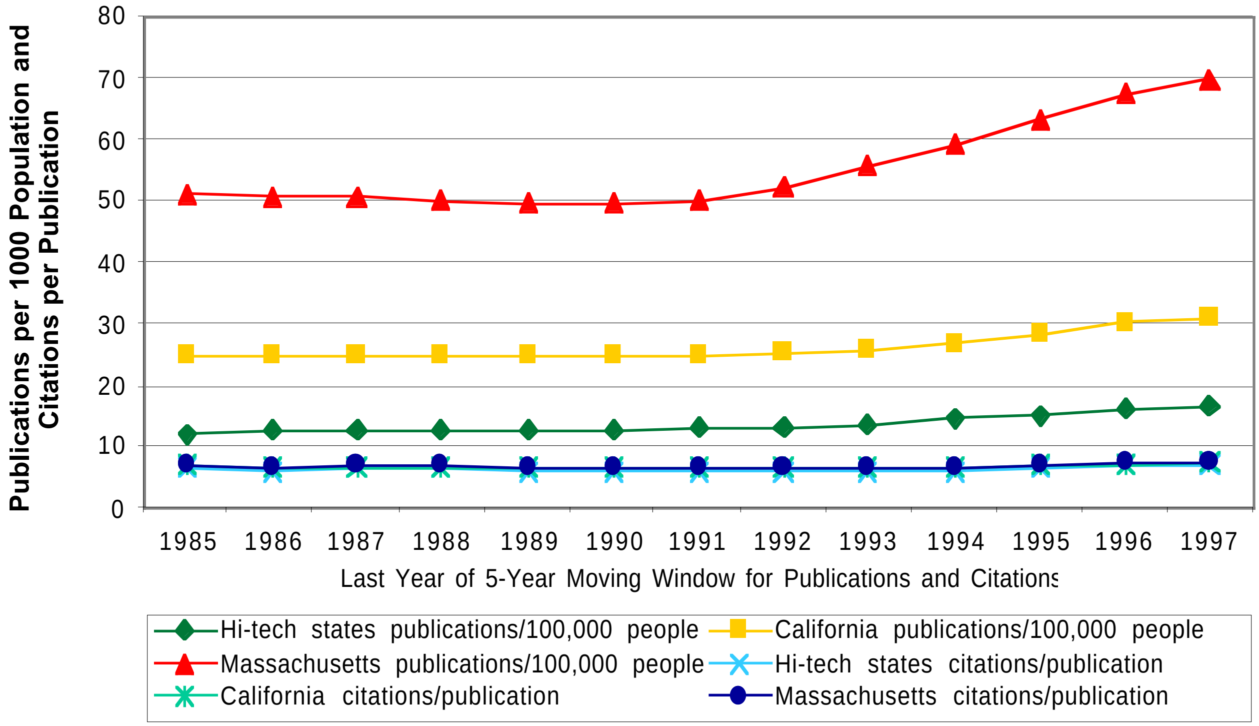
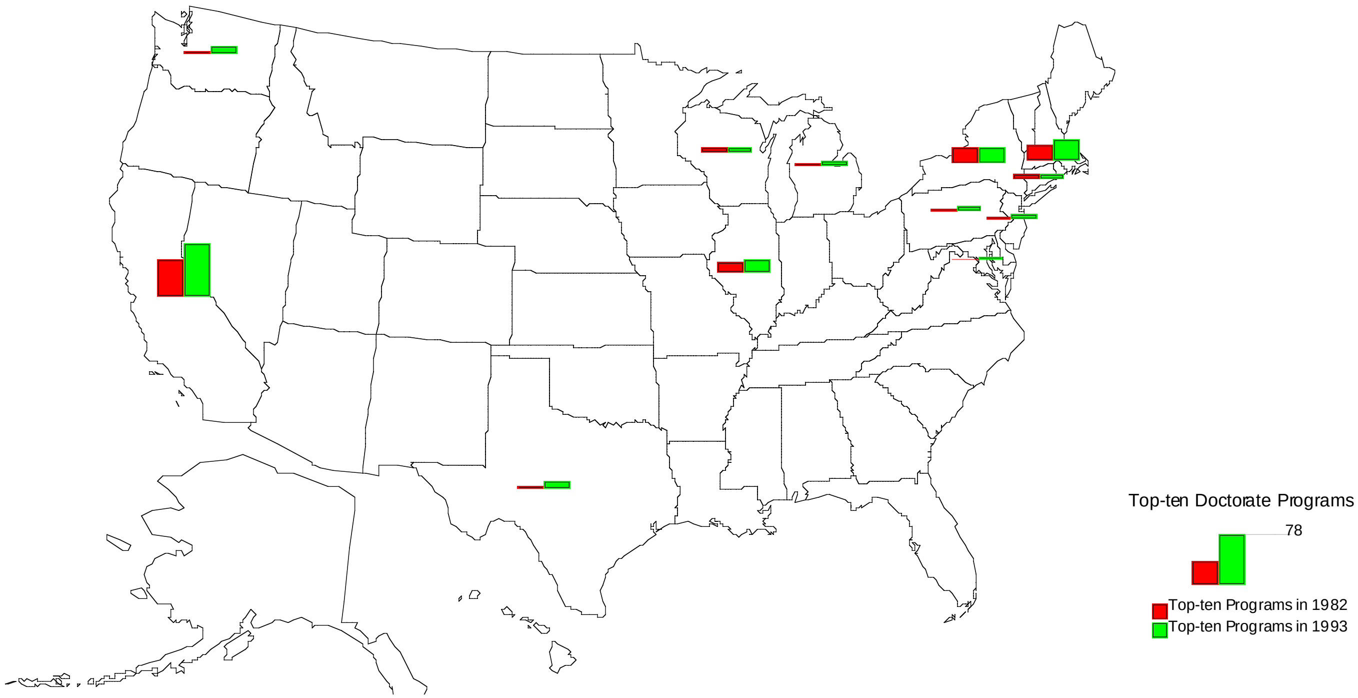


Figure 23. Trends in Publications and Citations per Publication at Major Universities with Top-ten NRC Ranked Programs in Other Sciences: California, Massachusetts, and High-tech States Average, 1981-1985 through 1993-1997

Map 1. Number of Top-ten Doctorate Programs in High-tech States



Notes

Only top-ten NRC ranked doctorate programs in the 12 high tech states + DC-plex (DC, Maryland and Virginia) are plotted. Ohio had none in 1982 and in 1993.

The number of top-ten programs in high-tech states increased from 153 to 212 (38.6 percent increase) from 1982 to 1993. All top-ten programs in the nation increased from 166 to 235 (41.6 percent increase).

9. Data Appendix

The data analyzed in this report are derived from three principal sources: The National Research Council 1993 study of U.S. research-doctorate programs (Goldberger, Maher, and Flattau 1995), two Science Citation Index data bases (Institute for Scientific Information 1981-1997) providing information on publications and their citations from 1981-1997 alternatively by state and university (where a publication is attributed to a state or university if one or more of its authors are from that state or university).

The raw data have generally been aggregated over all science and engineering disciplines and also divided into five distinct fields or clusters as detailed in Table 1: Biology/Medical/Chemical ("Bio/chem" for short), Computer Software/Information Processing/Interactive or Multi-Media--technical ("Information Technology"), Integrated Circuits/Semiconductors/High-Temperature Superconductors ("Semi/superconductors"), Other Engineering, and Other Science. We generally focus on a group of thirteen high-technology states: California, Connecticut, the DC-plex (DC, Maryland, Virginia treated as one state), Illinois, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, Texas, Washington, and Wisconsin, but also have developed aggregate data for these high-tech states as a group and the entire nation.

Detailed documentation for each data set follows.

9.1 Appendix A: 1993 NRC Study of U.S. Research-Doctorate Programs

The data files described here are contained in a Microsoft Excel data file entitled "NRCdata4HightechStates.xls." We are making this file available through the California Council on Science and Technology. The data are primarily based on the cross-section developed for the National Research Council 1993 study of U.S. research-doctorate programs (Goldberger, Maher, and Flattau 1995). Limited data are included on the number of top-ten NRC-ranked programs and program scholarly quality in the prior (1982) survey. We have supplemented that data by hand coding these variables for those 1982 programs not included in the 1993 data set. The data files used were drawn from *Research-Doctorate Programs in the United States CD-ROM* released by the National Research Council (1995). Data were compiled for each of the 2339 research doctoral programs in science and engineering in the study and then aggregated at the university and/or state level as detailed below.

We focus on six indicators of the university science base for each state: number of Ph.D.s produced in the five academic years 1987-1988 through 1991-1992, number of federal grants FY1986-1992, number of research-doctorate faculty 1992, the numbers of publications by those faculty 1988-1992 and citations to those publications 1988-1992, and the number of research doctoral programs with top-ten NRC reputational rankings in the 1993 study. The first five indicators are reported for both all the research doctoral programs and only those with top-ten NRC rankings.

9.1.1 Geographical Categorization

The District of Columbia, Maryland and Virginia together are aggregated into a single region called DC-plex, which this presentation considers a state for the sake of convenience. As a result, DC-plex and each state of the rest of the United States are the unit of analysis of this presentation. Tables 1-4 also show sums or statistics at the two aggregate levels: national and high-tech states, which, in this presentation, are California, Connecticut, DC-plex, Illinois, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, Texas, Washington, and Wisconsin.

Table A1. Zucker-Darby Categorization and NRC Standard Programs

Zucker-Darby Categorization	NRC Standard Programs
Biology/Medicine/Chemistry	Biochemistry & Molecular Biology, Cell & Developmental Biology, Molecular & General Genetics, Ecology, Evolution & Behavioral, Pharmacology, Chemistry, Biomedical Engineering, Chemical Engineering, Neurosciences, Physiology
Computer/ Information Processing/Multimedia	Computer Sciences, Mathematics
Integrated Circuit/Semi- & Super-conductor	Physics, Electrical Engineering, Materials Science, Mechanical Engineering
Other Engineering	Aerospace Engineering, Civil Engineering, Industrial Engineering
Other Sciences	Oceanography, Astrophysics/Astronomy, Statistics/Biostatistics, Geosciences

9.1.2 Academic Field Categorization

We first eliminate those 1,295 research doctorate programs in the social sciences and humanities even though they may represent a relevant intellectual base for certain technologies. The remaining 2,339 research doctorate programs are categorized into five predefined areas, also known as the Zucker-Darby categorization: **B**iology/Medicine/Chemistry, **C**omputer/Infor-mation Science/Multimedia, **I**ntegrated Circuit/ Semi- and Superconductor, Other **E**ngineering, and Other **S**ciences. Table A1 shows the correspondence between the Zucker-Darby categorization and the NRC standard programs.

9.1.3 Science-Base Indicators

All programs are coded as top-ten or non-top-ten based on the NRC reputational ranking. If the ranking of a program is less than 11, the program is identified as a top-ten Program.[10] The statistics of top ten programs are presented separately to show the contributions of very high-quality programs to scientific research.

The tables in the Microsoft Excel data file entitled "NRCdata4HightechStates.xls" contain data on the following indicators (and series derived from them as described) for the 13 high-tech states, their total, and the national total:

1. PHD: Number of Ph.D.s graduated by program for the period, academic year 1987-88 to 1991-92.
2. Faculty: The number of faculty members in the program in 1992 as reported by the institution.
3. Grants: Total number of federal grants FY 1986-92.
4. PUB5: Number of publications during the period 1988 to 1992.
5. CIT5: Number of program citations on publications during the period 1988 to 1992.
6. CPP: Citation per publication (CIT5/PUB5).
7. Mean scholarly quality: Average scholarly quality ratings of faculty from reputational surveys in the 1993 and 1982 NRC studies. [Scale: 5 (distinguished) to 0 (not sufficient for doctoral education)]
8. Top-ten PHD: Number of Ph.D.s graduated by programs which rank among national top ten for the period, academic year 1987-88 to 1991-92.
9. Top-ten faculty: The number of faculty members in programs which rank among national top ten.
10. Top-ten grants: Total number of federal grants attributed to programs which rank among national top ten.
11. Top-ten PUB5: Number of publications during the period 1988 to 1992 for programs which rank among national top ten.
12. Top-ten CIT5: Number of program citations on publications during the period 1988 to 1992 for programs which rank among national top ten.
13. Top-ten CPP: Citation per publication (CIT5/PUB5) for programs which rank among national top ten.
14. Top-ten Programs: Count of number of research doctorate programs with reputational rankings less than 11 in the 1993 and 1982 NRC studies.
15. Top-ten mean scholarly quality: Average scholarly quality ratings of faculty) for programs which rank among national top ten from reputational surveys in the 1993 and 1982 NRC studies. [Scale: 5 (distinguished) to 0 (not sufficient for doctoral education)]
16. POP5: 1988-1992 average resident population where the annual numbers are estimates of resident population taken from the web site of the United States Bureau of Census the web site of the United States Bureau of Census (www.census.gov) in August 1998.

Sheet 1 reports the PHD, Faculty, and Grants data for all and top-ten doctorate programs. Sheet 2 repeats the Faculty and reports the PUB5 and CIT5 data for all and top-ten doctorate programs. Sheet 3 reports these ten variables normalized by population (POP5) and CPP and top-ten CPP. Sheets 4 and 5 report top-ten programs for all fields and for the five broad science and engineering areas in the 1982 and 1993 studies, as well as changes and other transformations of the basic data. Sheets 6-8 report mean scholarly quality ratings (overall and top-ten, 1982 and 1993) for all doctorate programs and for computer sciences and statistics/biostatistics programs separately. Computer sciences represents a field in which programs have remained comparable over time despite about doubling in number while statistics/biostatistics represents a field comparable over time and not changing significantly in number. These reputational ratings may not be comparable across fields or time and have not been analyzed in detail by the authors. Sheets 9-11 basically replicate Sheets 6-8 by subregions within California where Northern California contains all universities 100 or more miles north of Los Angeles, and Southern California contains the remaining universities. Sheet 12 reports PHD, Faculty, Grants, PUB5, and CPP (on both overall and top-ten measures) for each and all of the five major science and engineering areas for California and the nation. Sheet 13 contains POP5.

9.2 Appendix B: State Publication Indicators of State Science Base

The data files described here are contained in a Microsoft Excel data file entitled "HightechStatePubs.xls." We are making this file available to the California Council on Science and Technology. Each table in this file contains panel data (across states and across years) for a particular science-base indicator, primarily the quantity of publication and

[10] Based on this criterion, if two programs are in the tenth tier (thus their rankings are both 10.5), then both are included in the top-ten group.

citation overall and for the five predefined high-tech areas based on *U. S. State Indicators on Diskette, 1981-1997*© from the Institute for Scientific Information.[11] In this file, a publication and citations to it are counted once if one or more of its authors list an affiliation in the state. Our presentation assumes that publication and citation are reasonable indicators of research performance in each field, by each region as defined below.

9.2.1 Geographical Categorization

The District of Columbia, Maryland and Virginia together are aggregated into a single region called DC-plex, which this presentation considers a state for the sake of convenience. As a result, DC-plex and each state of the rest of the United States are the unit of analysis of this presentation. All tables also show sums or statistics at the two aggregate levels: national and high-tech states, which, in this presentation, are California, Connecticut, DC-plex, Illinois, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, Texas, Washington, and Wisconsin.

9.2.2 Academic Field Categorization

The five predefined high-tech areas, also known as the Zucker-Darby categorization, are **B**iology/Medicine/Chemistry (B), **C**omputer/ Information Science/ Multimedia (C), **I**ntegrated Circuit/Semi- and Superconductor (I), Other **E**ngineering (E), and Other **S**ciences (S), which sequentially and respectively are separate sets of tables in each MS Excel workbook. Before the separate area tables, however, **a**ll field-combined statistics (A) are shown as the first set. (Capital letters in the following parentheses are the prefixes indicating the worksheet title of each field.)

Table B1 below shows how sub-areas match between the Zucker-Darby categorization and the ISI category description. (NOTE: The ISI labels below are from the original data set.)

9.2.3 Time Variable

All tables are in time-series format where four-digit numbers indicate the end of a five-year moving window,[12] ranging from 1985 through 1997. Thus, for instance, in the column of "1994," you will actually see the combined status of 1990, 1991, 1992, 1993, and 1994 for whichever indicator the table deals with.

9.2.4 Science-Base Indicators

The data set includes data on the first seven of the following science-base indicators overall and for each broad science and engineering area as well as state population and employment data (Table B1).

1. PUB5: sum of publications over each five-year moving window.
2. CIT5: sum of citations to the publications counted for PUB5 within each five year moving window.
3. CPP=CIT5/PUB5, citations per publication, which shows how many times each publication was cited on average during each five year moving window.
4. PUB5R=PUB5/(POP5)*1000, publications per 1000 residents over the 5 year window, which shows how many publications per 1000 residents were produced on average.
5. CIT5R=CIT5/(POP5)*1000, which shows how many citations per 1000 residents were received on average.
6. PUB5E=PUB5/(EMP5)*1000, which shows how many publications per 1000 employed residents were produced on average.
7. CIT5E=CIT5/(EMP5)*1000, which shows how many citations per 1000 employed residents were received on average.
8. POP5: five year average resident population where annual numbers are estimates of resident population taken from the web site of the United States Bureau of Census the web site of the United States Bureau of Census (www.census.gov) in August 1998.
9. EMP5: five-year average employed population where annual numbers are estimates of employment population taken from the web site of the Bureau of Labor Statistics (stats.bls.gov) within the United States Department of Labor in August 1998.

9.2.5 Tables

All tables in HightechStatePubs.xls are named as follows except for the last two: [one letter prefix] + [name of the variable shown]. For example, "Apub5" means "the table "PUB5" for "all fields combined." The last two tables show the POP5 and EMP5 population statistics used to generate quantity of publication and citation divided by five-year average resident and employment population.

[11] We used the deluxe version of the data set where ISI divided articles into 105 fields as categorized in *Current Contents*© from approximately 7,700 academic journals. In doing so, ISI attributed a paper to a state if the paper carried at least one author address of the state. For more information on this data set, contact the Institute at 215-386-0100, ext. 1411 by phone or at 215 387 1266 by facsimile. Its mailing address is 3501 Market Street, Philadelphia, PA 19104.

[12] This five-year moving window concept is adopted and sustained in this presentation mainly because it is the way the original source from ISI is compiled, not because the particular window selected by ISI is necessarily the ideal presentation of changes in publication and citation over time.

Table B1. Zucker-Darby Categorization and ISI Description

Zucker-Darby Categorization	ISI Description
Biology/ Medicine/ Chemistry	Agriculture/Agronomy, Anesthesia & Intensive Care, Animal & Plant Sciences, Animal Sciences, Neurosciences & Behavior, Biochemistry & Biophysics, Biology, Biotechnol & Appl Microbiol, Cardiovasc & Respirat Syst, Cell & Developmental Biol, Oncogenesis & Cancer Res, Agricultural Chemistry, Chemical Engineering, Chemistry & Analysis, Chemistry, Cardiovasc & Hematology Res, Dentistry/Oral Surgery & Med, Dermatology, Medical Res, Diag & Treatmt, Endocrinol, Nutrit & Metab, Entomology/Pest Control, Environment/Ecology, Experimental Biology, Food Science/Nutrition, Gastroenterol and Hepatology, General & Internal Medicine, Hematology, Immunology, Inorganic & Nucl Chemistry, Clin Immunol & Infect Dis, Molecular Biology & Genetics, Microbiology, Resrch/Lab Med & Med Techn, Medical Res, General Topics, Neurology, Endocrinol, Metab & Nutrit, Medical Res, Organs & Syst, Oncology, Ophthalmology, Organic Chem/Polymer Sci, Orthopedics & Sports Med, Otolaryngology, Pediatrics, Physical Chem/Chemical Phys, Pharmacology & Toxicology, Plant Sciences, Pharmacology/Toxicology, Psychiatry, Physiology, Clin Psychology & Psychiatry, Radiol, Nucl Med & Imaging, Reproductive Medicine, Rheumatology, Environmt Med & Public Hlth, Surgery, Urology, and Veterinary Med/Animal Health
Computer/ Information Processing/ Multimedia	AI, Robotics & Auto Control, Computer Sci & Engineering, Engineering Mathematics, Info Technol & Commun Syst, and Mathematics
Integrated Circuit/Semi- & Super-conductor	Appl Phys/Cond Matt/Mat Sci, Elect & Electronic Engn, Mechanical Engineering, Metallurgy, Materials Sci and Engn, Optics & Acoustics, Physics, and Spectrosc/Instrum/Analyt Sci
Other Engineering	Aerospace Engineering, Civil Engineering, Environmt Engineering/Energy, Engineering Mgmt/General, Geol/Petrol/Mining Engn, Instrumentation/Measurement, Nuclear Engineering, and Space Science
Other Sciences	Aquatic Sciences, Earth Sciences, and Multidisciplinary

9.3 Appendix C: University Publication Indicators of State Science Base

The data files described here are contained in a Microsoft Excel data file entitled "UniversityIndicators4HightechStates.xls." We are making this file available to the California Council on Science and Technology. Each table in this file contains panel data (across states and across years) for a particular science-base indicator, primarily the quantity of publication and citation overall and for the five predefined high-tech areas based on U. S. University Indicators on Diskette, 1981-1997© from the Institute for Scientific Information.[13] In this file, a publication and citations to it are counted once if one or more of its authors list the given university as an affiliation. Thus to the extent professors from two or more major universities in a state co-author, there will be double counting relative to the state totals when totals for the universities are aggregated to state totals. This presentation basically assumes that publication and citation reflect research performance in each field, by each region as defined below.

The data set includes information on the publication records of the 110 largest university recipients of federal funding, generally counting each campus of multiunit systems separately.[14] State data is obtained by aggregating over those of these 110 universities (if any) located in the given state. As noted in the text of this report, aggregation of university totals within a state results in double, triple, or more counting of the number of publications and citations in those cases where the authors are affiliated with two, three, or more universities within the state which appear among the 110 universities.

[13] We used the deluxe version of the data set where ISI divided articles into 105 fields as categorized in Current Contents© from approximately 7,700 academic journals. In doing so, ISI attributed a paper to a university if the paper carried at least one author address of the university. For more information on this data set, contact the institute at 215-386-0100, ext. 1411 by phone or at 215 387 1266 by facsimile. Its mailing address is 3501 Market Street, Philadelphia, PA 19104.

[14] Only the University of Massachusetts system was treated as a single entity at their request. It is important to not include in aggregations the separate reports for that system's Amherst and Worcester campuses.

9.3.1 Geographical Categorization

The District of Columbia, Maryland and Virginia together are aggregated into a single region called DC-plex, which this presentation considers a state for the sake of convenience. As a result, DC-plex and each state of the rest of the United States are the unit of analysis of this presentation. All tables also show sums or statistics at the two aggregate levels: national and high-tech states, which, in this presentation, are California, Connecticut, DC plex, Illinois, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, Texas, Washington, and Wisconsin.

Eleven (non-high-tech) states are missing since university indicators for those states are not contained in the source data base because they have no university among the top 110 in federal research funding. They are: Arkansas, Idaho, Maine, Mississippi, Montana, Nevada, North Dakota, Oklahoma, South Carolina, South Dakota, and Wyoming. Therefore, the aggregate U.S. data are only for the remaining 37 states and DC-plex.

9.3.2 Academic Field Categorization

The five predefined high-tech areas, also known as the Zucker-Darby categorization, are Biology/Medicine/Chemistry (B), Computer/Information Science/Multimedia (C), Integrated Circuit/Semi- and Superconductor (I), Other Engineering (E), and Other Sciences (S), which sequentially and respectively are separate sets of tables in each MS Excel workbook. Before the separate area tables in the Excel file, however, all field-combined statistics (A) are shown as the first set. (Capital letters in the preceding parentheses are the prefixes indicating the worksheet title of each field.)

Table C1 shows how sub-areas match between the Zucker-Darby categorization and the ISI category description. (NOTE: The ISI labels below are from the original data set.)

Table C1. Zucker-Darby Categorization and ISI Description

Zucker-Darby Categorization	ISI Descriptions
Biology/ Medicine/ Chemistry	Agriculture/Agronomy, Anesthesia & Intensive Care, Animal & Plant Sciences, Animal Sciences, Neurosciences & Behavior, Biochemistry & Biophysics, Biology, Biotechnol & Appl Microbiol, Cardiovasc & Respirat Syst, Cell & Developmental Biol, Oncogenesis & Cancer Res, Agricultural Chemistry, Chemical Engineering, Chemistry & Analysis, Chemistry, Cardiovasc & Hematology Res, Dentistry/Oral Surgery & Med, Dermatology, Medical Res, Diag & Treatmt, Endocrinol, Nutrit & Metab, Entomology/Pest Control, Environment/Ecology, Experimental Biology, Food Science/Nutrition, Gastroenterol and Hepatology, General & Internal Medicine, Hematology, Immunology, Inorganic & Nucl Chemistry, Clin Immunol & Infect Dis, Molecular Biology & Genetics, Microbiology, Resrch/Lab Med & Med Techn, Medical Res, General Topics, Neurology, Endocrinol, Metab & Nutrit, Medical Res, Organs & Syst, Oncology, Ophthalmology, Organic Chem/Polymer Sci, Orthopedics & Sports Med, Otolaryngology, Pediatrics, Physical Chem/Chemical Phys, Pharmacology & Toxicology, Plant Sciences, Pharmacology/Toxicology, Psychiatry, Physiology, Clin Psychology & Psychiatry, Radiol, Nucl Med & Imaging, Reproductive Medicine, Rheumatology, Environmt Med & Public Hlth, Surgery, Urology, and Veterinary Med/Animal Health
Computer/ Information Processing/ Multimedia	AI, Robotics & Auto Control, Computer Sci & Engineering, Engineering Mathematics, Info Technol & Commun Syst, and Mathematics
Integrated Circuit/Semi- & Super-conductor	Appl Phys/Cond Matt/Mat Sci, Elect & Electronic Engn, Mechanical Engineering, Metallurgy, Materials Sci and Engn, Optics & Acoustics, Physics, and Spectrosc/Instrum/Analyt Sci
Other Engineering	Aerospace Engineering, Civil Engineering, Environmt Engineering/Energy, Engineering Mgmt/General, Geol/Petrol/Mining Engn, Instrumentation/Measurement, Nuclear Engineering, and Space Science
Other Sciences	Aquatic Sciences, Earth Sciences, and Multidisciplinary

9.3.3 Time Variable

All tables are in time-series format where four-digit numbers indicate the end of a five-year moving window,[15] ranging from 1985 through 1997. Thus, for instance, in the column of "1994," you will actually see the combined status of 1990, 1991, 1992, 1993, and 1994 for whichever indicator the table deals with.

9.3.4 Science-Base Indicators

The data set includes data on the first nine of the following science-base indicators overall and for each broad science and engineering area as well as state population and employment data:

1. UPUB5: sum of publications over each five-year moving window.
2. UTPUB5: sum of publications over each five-year moving window counting only universities with at least one top-ten program (based on the 1993 National Research Council study) in a given broad field.
3. UCIT5: sum of citations to the publications counted for PUB5 within each five year moving window.
4. UTCIT5: sum of citations to the publications counted for UTPUB5 within each five-year moving window.
5. UCPP=UCIT5/UPUB5, citations per publication, which shows how many times each publication was cited on average during each five year moving window.
6. UPUB5R=UPUB5/(POP5)*1000, publications per 1000 residents over the 5 year window, which shows how many publications per 1000 residents produced on average.
7. UCIT5R=UCIT5/(POP5)*1000, which shows how many citations per 1000 residents were received on average.
8. UPUB5E=UPUB5/(EMP5)*1000, which shows how many publications per 1000 employed residents were produced on average.
9. UCIT5E=UCIT5/(EMP5)*1000, which shows how many citations per 1000 employed residents were received on average.
10. POP5: five year average resident population where annual numbers are estimates of resident population taken from the web site of the United States Bureau of Census the web site of the United States Bureau of Census (www.census.gov) in August 1998.
11. EMP5: five-year average employed population where annual numbers are estimates of employment population taken from the web site of the Bureau of Labor Statistics (stats.bls.gov) within the United States Department of Labor in August 1998.

[15] This five-year moving window concept is adopted and sustained in this presentation mainly because it is the way the original source from ISI is compiled, not because the particular window selected by ISI is necessarily the ideal presentation of changes in publication and citation over time.

Except for UTPUB5 and UTCIT5 (items 2 and 4 above), all other indicators are defined in the same way as those defined in state indicator tables.

9.3.5 Tables

All tables in University Indicators4 Hightech States.xls are named as follows except for the last two: [one letter prefix] + [name of the variable shown]. For example, "Apub5" means "the table UPUB5" for "all fields combined." The initial "U" means that state-level data are aggregated from university indicators. The last two tables show the POP5 and EMP5 population statistics used to generate quantity of publication and citation divided by five-year average resident and employment population.

10. References

Bush, Vannevar, *Science -- The Endless Frontier: A Report to the President on a Program for Postwar Scientific Research*, Washington, DC: U.S. Government Printing Office, 1945. [Reprinted with Foreword by Erich Bloch and Preface by Daniel J. Kevles, Washington, DC: U.S. Government Printing Office, 1990.]

Goldberger, Marvin L., Brendan A. Maher, and Pamela Ebert Flattau, eds., *Research-Doctorate Programs in the United States: Continuity and Change*, Washington, DC: National Academy Press, 1995.

Institute for Scientific Information, *Science Citation Index*, machine-readable data bases, Philadelphia: Institute for Scientific Information, 1981-1997.

Jaffe, Adam B., "Real Effects of Academic Research," *American Economic Review*, December 1989, 79(5): 957-970.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, August 1993, 63: 577-598.

Mansfield, Edwin, "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing," *Review of Economics and Statistics*, February 1995, 77(1): 55-65.

National Research Council., *Research-Doctorate Programs in the United States: Data Set*, machine-readable data base, Washington, DC: National Academy Press, 1995.

Zucker, Lynne G., and Michael R. Darby, "Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry," *Proceedings of the National Academy of Sciences*, November 12, 1996, 93(23): 12,709-12,716.

Zucker, Lynne G., Michael R. Darby, and Jeff Armstrong, "Geographically Localized Knowledge: Spillovers or Markets?", *Economic Inquiry*, January 1998. 36(1): 65-86.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer, "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, March 1998, 88(1): 290-306.