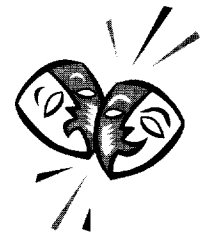


péter's picks & pans



Péter Jacsó
University of Hawaii

Critical Comparisons of American Colleges and Universities

DemoTables/DemoGraphics

World Databases

Factographic databases—ones that provide the facts in a structured, visually appealing format—are becoming increasingly important amidst today's information glut. What makes two databases—Critical Comparisons of American Colleges and Universities and DemoGraphics/DemoTable—stand out is the intuitive way they allow searching for data and the creative way they present results. What makes Bowker-Saur's World Databases the pan are screamingly wrong, nonsense data, poor searchability, and a pricing policy that adds insult to injury. Credit is due to Cheryl Toyama, my former student, for drawing my attention to the educational database.

information sources (and a great many useless ones), but also it offers a platform for talented individuals to introduce novel approaches for information retrieval and presentation. Critical Comparisons of American Colleges and Universities (<http://www.memex-press.com>) is one example. This new service and database from Memex Press, Inc. is the brainchild of William Busa, a former professor at The Johns Hopkins University. He saw the opportunity to usefully organize the large amount of information collected by government agencies and national organizations about colleges and universities. Searching is intu-

itive and Memex presents the data in context in a way that is an impressively good example of information visualization.

Entering this market is not for the faint of heart—it has been the stronghold of Peterson and, more recently, of *U.S. News & World Report*. Busa combines, synthesizes, and distills data collected by the Department of Education, the Bureau of Labor



the picks

CRITICAL COMPARISONS OF AMERICAN COLLEGES AND UNIVERSITIES

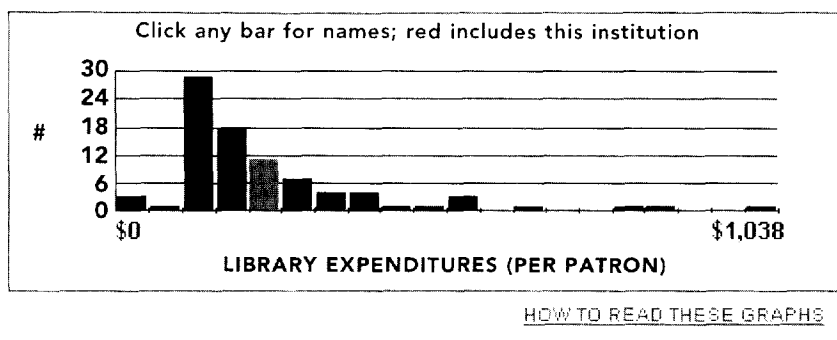
The beauty of the World Wide Web lies not only in that it makes possible the publication of many useful

Critical Comparisons of American Colleges and Universities can be searched by a number of criteria, separately or in combination, through a self-explanatory template.

COST		LIBRARY	
In State Tuition: ANY	▼	Matl Expenditures: ANY	▼
Out of State Tuition: TYPICAL	▼	Data Expenditures: ANY	▼
Annual Increases: ANY	▼	Inter-Library Loans: ANY	▼
Room & Board: ANY	▼	Volumes Patron: ANY	▼
ANY	▼	TYPE OF SCHOOL (CARNEGIE CLASSIFICATION)	
HIGH	▼	▼	
HIGH OR TYPICAL	▼	▼	
TYPICAL	▼	▼	
TYPICAL OR LOW	▼	▼	
LOW	▼	▼	
CALIFORNIA	▼	▼	
FACULTY REPUTATION		STUDENT SERVICES	
Rank: HIGH	▼	Expenditures: ANY	▼

FULL-TIME UNDERGRADUATES		CAMPUS CRIME RATES	
Number: ANY	▼	Property: ANY	▼
% of all Students: ANY	▼	Violent: TYPICAL OR LOW	▼
% Female: ANY	▼	Drug/Alcohol: TYPICAL OR LOW	▼
% African American: ANY	▼	FINANCES	
% Hispanic American: ANY	▼	Revenues: ANY	▼
% Asian American: HIGH	▼	Revenues From Tuition: ANY	▼
% Native American: ANY	▼	Endowment: ANY	▼
% Non-Resident Aliens: ANY	▼	Endowment Yield: ANY	▼
INSTITUTIONAL SCHOLARSHIPS			
% of Tuition: ANY			

EXPENDITURE (PER PATRON) FOR LIBRARY MATERIALS IS WITHIN THE NORM:



Within each category are several comparative graphs, such as this one for student library expenditures.

Statistics, and the National Academy of Science's National Research Council to paint an objective portrait of nearly 1,400 U.S. colleges and universities.

The database can be browsed by name or Carnegie classification (Research University I & II, Doctoral Universities I & II, Masters (Comprehensive) Universities I & II, and Baccalaureate Universities I & II). The database can be searched by a number of criteria, separately or in combination, through a self-explanatory template.

Six options exist for the various criteria (except for the state names and the college classifications). These are implicitly combined in a Boolean operation. This is very intuitive, allowing for a search on, say, institutions in California that have a good reputation, typical or low out-of-state tuition fees, and low crime.

The results of the search are first presented in a list format, with the names of the universities and colleges hyperlinked to their descriptions. These are not the run of the mill fact sheets that often prevent you from seeing the forest from the trees, but visually informative textual-graphical portraits.

The software presents the portraits enhanced with short interpretive descriptions of these categories: cost, faculty reputation, library services, crime, student services, scholarships, student body, finances, and majors. Within each category are several comparative graphs. These put the institution in appropriate context by

the given criterion. Within the cost category, for example, there are three charts for in-state tuition, out-of-state tuition, and room & board charges for institutions belonging to the same Carnegie classification.

The horizontal axis shows the range of the possible values of the evaluation criterion (such as in-state tuition or expenditures for books or databases per students). The values are stratified into 20 bins, each of which represents an interval. I sorely missed a "not available" bin for institutions that did not provide data for the criterion, causing them to appear in the lowest category. This yields the odd graphing of Harvard and Georgetown Universities as ones that did not spend a penny on book purchases. Luckily, this is easy to correct.

The Y-axis represents the number of institutions that belong to the appropriate bin. The bin-group of the target institution being analyzed appears in red (shaded when printed in black and white). It clearly shows the relative position of the institution within its class. Clicking on any of the bars will list the institutions belonging to the bin-group. In turn, the entries in the list are hyperlinked to the description. This concept of binning and active graphing (along with the hyperlinking) matches very well the train of thought of someone looking for further information, triggered by data being presented, while being able to step back easily to follow the previous path. For example, what is that single university, quite alone in the extreme ranges, that

spends \$1,038 for books per student? You click on the bar to learn that it's Princeton. And while you are there you want to see how it fares in terms of other criteria so you click on its name. When your enthusiasm is cooled by the tuition fee, you click twice to get back to your original target institution.

Does the name Memex ring a bell? It honors the concept developed by Vannevar Bush in the 1930s and presented in 1945 in his famous *Atlantic Monthly* article, "As We May Think." It is regarded as the pioneering article for the now popular hyperlink concept. Vannevar Bush would be happy to see this database, especially as it is free unless you log in from an ISP with the .edu domain. Educational institutions have to pay \$24.75 per year for unlimited access from their domain. It is less than the real cost of the interlibrary loan for Bush's seminal article.



DEMO TABLES/DEMOGRAPHICS

I might not have spotted this gem of a database without the Web (<http://www.magnet.at/heilig.htm>). The software developed by Gerhard K. Heilig is an outstanding tool for easy searching and spectacularly good display of a mountain of demographic data collected and projected for 1950-2050 by the United Nations Population Division. The data is available from UNPD in ASCII format, but it is Heilig's software that brings the most and best out of this file.

DemoTables looks like a spreadsheet, but it is much more than that. With the click of a button you can rotate through time periods, or change the details of geographic coverage drilling down from global data to continental, regional, subregional, and country level. You can sort the table by any of the columns with another click. This is a superbly neat, fast, and convenient way of rearranging a spreadsheet the size of Texas, but this is just for starters.

DemoTables '96

Vital Rates: 1990-95

Select Area	Select Year							
	Birth / Death Rates		Nat. Growth	Fertility	Infant Mort	Life Expectancy		Difference
Standard Table	CBR	CDR	NGR	TFR	IMR	e(0)M	e(0)F	e(0)F-M
Northern Europe	13.1	11.3	2	1.8	7	72.8	78.8	6.0
Western Europe	11.3	10.4	1	1.5	7	73.2	80.2	7.0
Australia/New Zealand	15.2	7.5	8	1.9	7	74.5	80.3	5.8
Northern America	15.4	8.8	7	2.0	9	72.8	79.5	5.7
More developed regions	12.4	10.1	2	1.7	11	70.4	78.0	7.6
Southern Europe	10.7	9.8	1	1.4	11	72.7	79.3	6.6
Europe	11.5	11.3	0	1.5	13	68.5	76.9	8.4
Eastern Europe	11.5	12.6	-1	1.5	15	63.0	73.6	10.6
Oceania	19.1	7.8	11	2.5	26	70.3	75.6	5.3
Polynesia	25.4	5.5	20	3.5	35	67.2	71.6	4.4
Central America	23.8	5.5	23	3.4	37	67.6	73.4	5.8
Latin America & Caribbean	24.9	6.7	18	2.9	40	65.3	71.8	6.5
Eastern Asia	17.5	7.2	10	1.9	4*	67.6	71.9	4.3
South America	23.6	7.0	17	2.8	4*	64.4	71.4	7.0
Micronesia	33.0	5.8	27	4.4	41	65.7	69.1	3.4
Caribbean	22.8	7.8	15	2.7	43	66.4	70.8	4.4
South-eastern Asia	26.4	8.0	18	3.2	54	61.7	65.6	3.9
Southern Africa	32.0	9.1	23	4.2	55	59.3	64.9	5.6
Melanesia	32.1	9.2	23	4.7	59	58.1	60.3	2.2
Western Asia	29.7	7.2	23	4.1	60	64.4	66.4	4.0
World total	24.1	9.3	1.5	3.0	62	62.2	66.5	4.3
Asia	24*	8.3	1.6	2.8	62	63.2	66.0	2.8

Help
Definitions
Print
Go Back

DemoTables looks like a spreadsheet, but it is much more than that.

Most of the data are also available through the DemoGraphics program as a variety of graph, chart, and thematic map formats. What makes them unusual is that these are not static, but data-driven animated charts, graphs, and maps. You choose the type of chart (line, pie, thermometer, population tree, scattergram, and—sometimes—a combination), one to three geographic entities, and the criteria by which they will be compared. You can run the animation

step-wise in 5-year intervals, or continuously through 100 years, at your choice of slow, medium, or fast pace. It is awesome to see, for example, how infant mortality rate (IMR) has been dropping around the world, as illustrated by the changing of the color codes of the countries on the map from deep red (very high IMR) to pale yellow (low IMR) as quintennials go by. This database is a godsend, not only for researchers, but also for students to learn about population

dynamics, represent statistical data across time, find interesting demographic facts, and explore trends.

Playing around with the database I was surprised to see that in two countries (Maldives and, to a lesser extent, Nepal), the life expectancy for males was higher than for females (for 1990-95). It is a unique phenomenon even in countries that don't wage wars and refrain from tribal genocide. It piqued my interest, so I clicked the button to sort the table by countries to see if perhaps the maternal mortality rate in the Maldives and Nepal was extremely high. I wish I could brag about what a clever mental leap I had, but the data did not support it. I am puzzled, but it is not the fault of the database or the software. That this product can make available data in a way that triggers creative thinking and teaches numeracy is splendid. I wish the software were made available to let users import and understand their own time series.



the pan

WORLD DATABASES (WDBS) ON DATASTAR

I often welcome a new database when it shows the promise of challenging the hegemony of another, fueling competition that benefits the consumer. There are plenty of promises for the new World Databases (WDBS on DataStar) directory from Bowker-Saur, but these are only in the publicity blurb. Reality shows an embarrassingly bad database, and makes me wonder how the producer and DataStar thought they could get away with it when the primary target users are likely to be information professionals. Bowker-Saur and DataStar don't seem to be able and/or they don't care in the least about handling factographic data properly.

The absurdities start to strike you even in DataStar's own Database Description for WDBS (in the BASE

DemoGraphics '96

A population education tool from the United Nations Population Fund (UNFPA). Developed by Gerhard K. Herle

Most of the data are also available through the DemoGraphics program as a variety of graph, chart, and thematic map formats.

database or, on DataStar Web, click the I button from within WDBS). It claims that WDBS answers such questions as "which newsletters are covered by ABI/INFORM in full-text?" It would be lovely, but there is absolutely no way to get such information from WDBS since individual journal titles in listed databases are not part of the description of the databases. Update frequency for WDBS is promised as "365 days." With the vision of daily updating in my head, I envisioned the Grinch who stole Christmas, or—more appropriately for the British origin of the database—a Dickensian sweatshop without a break, even for Boxing Day. Well, it ain't so. The last time the database was updated was December 23, 1997. Perhaps on Boxing Day, the creators rested and cashed in their bonuses. More likely, the update frequency should have read "Every 365 days." In that case, why doesn't it say "Annual reload?" Users may become incredulous as they read the sample records for Science Citation Index (SCI) in the Database Description. WDBS claims that there are 12,000,000 records in the database. Sorry, wrong number. The correct figure for the entire database is about 16,000,000. WDBS claims that SCI is updated weekly (so far so good) and 780,000 records are added per update. No one spotted that it would result in a phenomenal growth rate quite out of line with the actual size of the file (40,560,000 records per year). In reality, the number of records added weekly ranges from 13,000 to 21,000. Bowker-Saur may have had second thoughts because for some implementations of SCI the records say: "Updating: Not updated; 780,000 records per update." So is it or isn't it? And if it is, with how many records and how often?

WDBS claims that SCI covers 5,200 journals. ISI itself pegs this number at less than 4,000. WDBS claims that abstracts are available. It is true, but it does not tell the entire truth. You don't need to be a super searcher to know that abstracts are available only from 1991 forward. If you have any knowledge of this database then you know that, even from 1991, abstracts are available only if the

article had author abstracts. ISI does not create abstracts. Usually, the sample records used by the producer for documentation and promotional materials are well-chosen and carefully examined to let the producer put its best foot forward. If this is WDBS' best foot, it is a bad omen. My test searches suggest that such grossly erroneous, misleading, incomplete, and simply nonsense data characterize the entire WDBS database. The same uninformative statement about abstract availability appears for databases where only a minority of records have abstracts, such as the SPORT database (17% with abstracts).

No matter what database I looked up, its description was filled with nonsense. When WDBS claimed that Information Science Abstracts (ISA) adds 8,800 records per update, I thought that Bowker-Saur has mistaken the volume of the yearly update for that of the monthly update. But when I saw the monthly update volume for Mental Health Abstracts (MHA) reported as 18,000, this belief became untenable. It takes more than 5 years for IFI/Plenum to add that many (or is it that few?) records to this ailing database.

Bowker-Saur claims that a detailed subject index allows quick and easy access to relevant information. Unfortunately, this subject index is not browsable, and the pull-down menu on DataStar Web for field qualified searching does not offer the descriptor field. It may be consoling for some that the address of the database producer and vendor is searchable, so there is nothing to prevent one from looking up all the databases produced in East Grinstead (home of Bowker-Saur). I grinned (although you might grimace) when looking at the descriptors assigned to MHA—such as UFOs and Disaster-Relief. You can easily memorize that the coverage of MHA includes Psychology, Psychiatry, Children, and Substance-Abuse—these appear two or three times in the descriptor field. Repetition is the mother of knowledge as my mother repeatedly told me. Other interesting descriptors assigned to this database may take a little longer to memorize because of the lacks in browsing

ability. How do you look up "see-W-7" in this descriptor: "Other-Addictive-Behaviours-see-W-7?" Others, such as "Materia-Medica" that appear in just five records in the entire MHA database, can hardly be justified as a topical descriptor in WDBS.

Credit goes to the producer for advising that "in some cases price information is also included in the records." I would have taken that for granted, but maybe we are better off without price information. The records that do have price information are grossly outdated or plain wrong. I would not have expected the new Dialog DialUnit prices, but the prices that *are* given are at least pre-1997. It is easy to recognize that the US \$65 charges for both offline and online prints in ISA are excessive (a misprint of 65 cents, I presume). But it was years ago when the print charge was 65 cents. Most of the nonsense is repeated for records of each and every online, CD-ROM, and tape versions of a database (at \$4 a pop) even when the database name provides a clue.

Take ABI/INFORM as an example. The descriptions of ABI/INFORM Global, ABI/INFORM Research, ABI/INFORM Select, and most of the 34 versions keep repeating that the database is updated monthly, 40,000 records are added to it at each update, and it covers 1,100 journals. In some records, the descriptive part makes it clear that far fewer journals are included in ABI/INFORM Select, but the rest of the record just parrots the wrong information dragged from one record to the next.

The software matches the quality of the data content. You cannot search by type of medium (online, CD-ROM), or type of database (bibliographic, numeric) and other criteria that are available for the Gale Database Directory on DataStar (GDDB). You can get a lot of useful things for \$4 apiece and \$60 per hour. This appalling database is certainly not one of them.

Communications to the author should be addressed to Péter Jacsó, 1030 Aoloa Place, 207-A, Kailua, HI 96734; 808/956-5817; jacso@hawaii.edu.