The Sorenson Workstation FIVE Data

by

Olav Sorenson

Jeffrey S. Skoll Chair in Technical Innovation and Entrepreneurship
Professor of Strategic Management
Joseph L. Rotman School of Management
University of Toronto

Contents

1. Conditions of Use

The data sets described in this document are part of the FIVE Project on Firm and Industry Evolution and Entrepreneurship, organized by Constance Helfat (Tuck School of Business at Dartmouth) and Steven Klepper (Carnegie-Mellon University). If you download or otherwise obtain any of the data sets described in this document, or obtain modified versions of these data sets, you agree to abide by the conditions of use set forth here.

The organizers of the FIVE Project, Constance Helfat and Steven Klepper, are not responsible for the content of any of the FIVE Project data sets. The FIVE Project organizers and the FIVE Project data set authors make no representations or warranties of any kind concerning any of the FIVE Project data, including with regard to the absence or presence of errors.

All FIVE Project data sets are protected by copyright. The copyright holders for each data set are indicated on the first page of the documentation for each data set. Permission is granted to reproduce FIVE Project data for non-profit educational and research use only. Requests to reproduce FIVE Project data for other uses should be addressed to: Professor Constance Helfat, Tuck School of Business at Dartmouth, Hanover, NH 03755, constance.helfat@dartmouth.edu.

This document describes the Sorenson Workstation FIVE Data, which should be referred to by this name in any derivative works. In any written and published work, users of these data should cite this document, one of the data set author's original articles (either Sorenson, 2000, or Sorenson, 2003), and the FIVE Project: Data Overview (Helfat & Klepper, 2007).

You may create a new data set by modifying a FIVE Project data set, including but not limited to adding or removing data or merging data from different FIVE Project data sets. If you utilize any FIVE Project data to create new data sets, within six years from the date that you obtain the FIVE Project data, you must provide the FIVE Project with a copy of your modified data set by contacting: Constance Helfat, constance.helfat@dartmouth.edu.

Any written documents or statistical estimates that use FIVE Project data, including but not limited to working papers and publications, must be reported to: Constance Helfat, constance.helfat@dartmouth.edu.

Additional conditions of use (specified by the author of this document):
If you supplement the Sorenson Workstation FIVE Data, you must rename that variant of the original data with Sorenson appearing first in the list of data set authors, along with the word FIVE. For example, if Joe Campus added product

performance characteristics, the data set would be called the "Sorenson-Campus Workstation FIVE Data." If you merge two or more FIVE data sets, the new name of the data set must include the names of all of the original data set authors and the word FIVE.

2. Data Set Author

Olav Sorenson
Jeffrey S. Skoll Chair in Technical Innovation and Entrepreneurship
Professor of Strategic Management
Joseph L. Rotman School of Management
University of Toronto

3. Data Set Summary

The data describe all firms that I could identify selling a "computer workstation" at any point in time between 1980, the genesis of the industry (for a brief description of the history of the industry, see Sorenson, 2003), and 1996. In total, it covers 175 firms for 690 firm-years (Data File 2) and 1,276 products for 2,735 product-years (Data File 1). See Data Description for a discussion of the definition of a "computer workstation" and an explanation of the process used to identify computer workstations and the firms that produced them.

4. Author's Research Using This Data Set

Sorenson O. 2000. Letting the market work for you: an evolutionary perspective on product strategy. *Strategic Management Journal* 21: 577-592.

Sorenson O. 2003. Interdependence and adaptability: organizational learning and the long-term effect of integration. *Management Science* 49: 446-463.

Sorenson O. 2005. "Social networks and the persistence of clusters: Evidence from the computer workstation industry,' pp. 297-316 in *Clusters, networks and innovation*, Breschi S, Malerba F (eds). Oxford University Press: New York.

Sorenson O, McEvily S, Ren CR & Roy R. 2006. Niche width revisited: organizational scope, behavior and performance. *Strategic Management Journal* 27: 915-936.

5. Additional References

Barron DN, West E & Hannan MT. 1994. A time to grow and a time to die: Growth and mortality of credit unions in New York City, 1914-1990. *American Journal of Sociology* 100: 381-421.

Berstein Research. 1992. *The Workstation Market*. Sanford C. Berstein & Co.: New York.

Helfat, CE & Klepper, S. 2007. FIVE Project: Data Overview. http://papers.ssrn.com/paper=1028022.

Sun Microsystems. 1986. *Sun news* (Tech. Rep.). Sun Microsystems: Palo Alto, CA.

6. Data Set Sources

The primary source for identifying products is Data Sources, a catalog published by Information Access Company of Foster, CA at varying frequencies – either semi-annually or quarterly – over the period covered by the data. I identified additional products within this category by reading product reviews and scanning advertisements in Byte and IEEE Graphical Computing and Applications for the entire period (Byte did not yield information on any products not also found in IEEE). I should note that although these sources yielded additional products from 1980-1982, they provided little more than supplemental information from 1983 onward. I collected additional company level data from corporate reports and from articles available on Lexis-Nexis (for each firm, I performed a search in Lexis-Nexis using the company name listed in Data Sources).

7. Data Description

I originally collected these data to understand better how vertical integration influences the ability of organizations to learn from experience (Sorenson, 2003). With respect to this question, I chose the computer workstation industry because product directories allowed me to determine to some extent whether the firms in the industry produced the components used in their machines or whether they purchased them from other companies. Though originally designed to investigate integration and learning, I have also used these data to explore the effects of product line breadth and product innovation on firm performance (Sorenson, 2000, Sorenson et al, 2006) and the evolution of the industry geographically (Sorenson, 2005). Much of the data nonetheless remains unexploited. In particular, very little has been done with the product level data

set, and with the exception of vertical integration, most of the firm-level variables have been used only as controls.

7.1. Product/Industry Definition

Although computer workstations formed an established product category by 1985, even the term "computer workstation" did not exist at the time of introduction of what experts generally consider the first product, the Apollo DOMAIN. Some furniture manufacturers did, however, use the term in the 1970s and early 1980s to refer to desks designed to house terminals. Moreover, despite the common usage of the term computer workstation in the press, Data Sources never adopted it as a product category (the products found here appeared under either "Desktop Systems" or "Graphics Systems"). Given the absence of a standard catalog, and to capture the early period of the market/industry before a consistent category name had been adopted, I decided to collect information on computer systems with a consistent set of attributes.

Definitions of computer workstations abound. Most of these definitions have the disadvantage of being relevant at some particular cross-section in time and therefore not particularly useful in collecting longitudinal data. For example, Berstein, 1992, defined the workstation product class by the following characteristics: a 32-bit CPU; stand-alone single-user capability; multi-tasking 32-bit operating system; bit-mapped graphics; utilization of virtual memory; and built-in networking. This definition poses multiple problems. First, Bernstein requires that workstations be 32-bit machines, but some early offerings had 16-bit processors and architectures. For example, the Orcatech Design Engineering Workstation, introduced in 1982 and one of the first products with "workstation" in its name, had a proprietary 16-bit processor. This firm eventually exited strategically from the workstation market, but for two years, the press cited it as one of the up-and-comers that could potentially rival Apollo. At the other end, by 1996 most of the major RISC chip manufacturers had already started producing 64-bit chips for the workstation market. Second, this definition requires machines to use bit-mapped graphics. Although all workstations on the market today use this technology, some of the early rivals to Apollo chose to design their systems around raster graphics. The ubiquity of bit-mapped graphics therefore appears an artifact of hindsight on the evolution of the product.

I therefore decided to define the product according to some abstract features of its functionality, rather than according to specific components. In particular, I defined a "computer workstation" as a distributed computing machine for primary use by a single user. This definition differentiated workstations importantly from three other classes of computing machines: terminals, servers, and personal computers. Terminals do not have significant local processing capabilities. Although in the 1980's terminals evolved to include hardware for

processing some graphics rendering locally, these machines still relied on a server for all general processing. Servers meanwhile are used in distributed computing, but computer manufacturers do not design them for use by a single user; multiple users access their resources through either terminals or workstations. Personal computers, where all processing, memory, and most data storage occur on the user's machine, sit at the opposite end of the spectrum. Though intended for a single user, they involve no distributed processing.

Meeting this definition requires both specific hardware and system software. On the hardware side, workstations must have a CPU and memory for general processing and a monitor for display. These hardware requirements help to distinguish workstations from terminals and servers. Terminals do not have CPUs for general processing and servers often have no built-in display capabilities. Workstations also need an operating system that allows for distributed computing. Experts considered the ability to share virtual memory and file resources across machines the key attribute of the Apollo system, widely considered the first workstation. UNIX did not even provide a distributed computing environment until Sun Microsystems introduced the NFS protocol for file-sharing in November of 1984, which other manufacturers offering UNIX systems quickly adopted (Sun, 1986). Although personal computers often look very similar to workstations in their physical configurations, MS-DOS, MacOS and other personal computer operating systems did not allow for this seamless sharing of resources during the 1980s and early 1990s.

Because of the definition that I adopted, 1996 serves as a natural endpoint for the data set. After the arrival of the Windows NT operating system (first released in 1993, but not widely adopted until version 4.0 became available in 1996), distinctions between personal computers and computer workstations blur—or at least they shift from being qualitative, in terms of functionality, to being quantitative, in terms of speed and price. I believe that this blurring accurately reflects the demise of the product as a distinct category/industry.

I used these hardware and software features to identify computer workstations at the product level. Though the classification process often required additional research, I adopted the following baseline approach: (1) From the “Desktop Systems” category, I included all systems with operating systems capable of distributed computing (except MacOS and Windows NT), but excluded systems with “Server” in the name or without graphics capabilities or designated as a multiple user system. (2) From the “Graphics Systems” category, I included all systems with operating systems capable of distributed computing.

7.2. Excluded Fields

In the papers that I have written from these data, users may note that I mention and sometimes use product-level unit sales. I obtained that information from IDC

and the terms of the agreement do not allow me to grant access to others. Even if the intellectual property issues could be resolved, I have unfortunately been unable to locate an uncorrupted file containing this information, so I am afraid that it may have become unrecoverable.

8. Tables

8.1. Descriptive Statistics, File 1: SorensonworkprodFIVEdata

	Variable	N	Mean	SD	Min	Max
Clock speed	mhz	956	40.8	42.4	1	300
Memory	ram	2355	131,178	843,323	64	1.60E+07
Storage	hd	1521	2,980	13,998	20	138,600
Display	monitor	2354	0.26	0.44	0	1
Listprice	price	2613	27,389	31,963	400	600,000

8.2. Descriptive Statistics, File 2: SorensonworkfirmFIVEdata

Variable	N	Mean	SD	Min	Max
year	690	1990	3.9	1980	1996
found	672	1972	20.2	1880	1996
employ	580	15,800	54,241	1	405,535
sales	630	2.70E+09	.9.9e+09	0	7.50E+10
RandD	105	1.90E+08	9.90E+08	294,877	6.40E+09
products	690	22.5	75	0	598
ptypes	690	3.8	5.3	0	33
system	690	0.41	0.49	0	1
graphic	690	0.85	0.36	0	1
cpu	690	0.02	0.13	0	1
os	690	0.25	0.43	0	1
apps	690	0.59	0.49	0	1
comm	690	0.19	0.39	0	1
monitor	690	0.11	0.32	0	1
disk	690	0.16	0.36	0	1
memory	690	0.08	0.27	0	1
board	690	0.11	0.32	0	1

9. Variable List and Definitions: File 1: SorensonworkprodFIVEdata

Descriptive statistics for the continuous and indicator variables below appear in the Tables section.

9.1. Year of Observation:

Year

[Variable name: *year*; 4 digit numeric variable]: Year of observation, ranging from 1980 to 1996.

9.2. Firm Identifiers:

Firm Name

[Variable name: *firmname*; 40 character string variable]: Name of company.

FIVE Firm ID (unique firm identifier for the FIVE Project)

[Variable name; *fivefirmID*; 6 digit numeric variable]: Unique identifier for each Firm Name that appears in the FIVE data sets. Firm Names and FIVE Firm IDs may appear in more than one FIVE data set. If a firm's name changed over time, the FIVE Firm ID for the firm sometimes may change as well. In addition, different FIVE data set authors may have coded different firm names for the same company, which could result in different FIVE Firm IDs for the same firm in different data sets.

Data Set Firm ID (firm identifier assigned in the original data set)

[Variable name: *firmID*; 3 digit numeric variable]: Unique identifier that I created for each firm in the original data set.

CUSIP Header ID (firm identifier assigned by Compustat)

[Variable name: *CUSIPheaderID*; 8 digit numeric variable]: Unique identifier for each firm assigned by Compustat for 2007. For a given Firm Name, this ID is constant across all years in the data set. These IDs come from matching each Firm Name in the FIVE data sets to an identical or similar name in Compustat. There is no guarantee that these matches are accurate. There also are many Firm Names for which no match was found.

CUSIP Header Firm Name (company name assigned by Compustat)

[Variable name: *CUSIPheadername*; character variable]: Name of company given in Compustat for the CUSIP Header ID.

CUSIP Historical ID (firm identifier assigned by Compustat)

[Variable name: CUSIP*historyID*; 8 digit numeric variable]: Unique identifier for each firm assigned by Compustat in their historical databases. For each Firm Name, the CUSIP Historical ID is that shown in Compustat for the year of observation in the FIVE data set. These IDs come from matching each Firm Name in the FIVE data sets to an identical or similar name in Compustat. There is no guarantee that this match is accurate. There also are many Firm Names for which no match was found.

CUSIP Historical Firm Name (company name assigned by Computstat)

[Variable name: *CUSIPhistoryname*; character variable]: Name of company given in Compustat for the CUSIP Historical ID.

9.3. Industry Submarket or Product Identifiers:

Product Name:

[Variable name: *productname*; 40 character string variable]: Either the name listed in Data Sources or the name appearing in the advertisement or product review from which I identified the computer workstation.

Product ID (product identifier assigned in the original data set)

[Variable name: *productID*; 4 digit numeric variable]: Unique identifier that I created for each product. Note that although these IDs uniquely identify products, they may not nest within companies; several companies continue to sell existing systems under the same names after acquiring or merging with another company.

9.4. Other Variables:

Operating System

[Variable name: *os*; 2 digit numeric variable]: Operating system installed on system by the firm selling the workstation. The table below provides the codes, names, year of first appearance, number of product-years with that operating system, and whether the operating system had UNIX as its underlying engine.

Code	OS name	First year	Number	UNIX-based
1	SunOS	1985	477	Yes
2	Solaris	1990	147	Yes
4	DEC OSF	1994	10	No
5	MAX/OS	1990	13	Yes?
7	DEC UNIX	1986	209	Yes
8	ULTRIX	1990	50	Yes
9	HP/UX	1984	225	Yes
10	AIX	1988	120	Yes
11	Other UNIX	1982	708	Yes
12	DG Aviion	1986	41	No
13	DEC 3000	1994	27	No
14	MicroVMS	1983	317	No
15	SGI	1984	23	Yes
16	EP/IX	1991	17	Yes
18	Classix	1991	13	Yes
19	DOMAIN	1983	116	No
20	Other	1980	68	No
21	Aegis	1981	33	No
22	Rainbow TOS	1990	7	No
23	CTOS	1983	4	No
24	NextStep	1990	5	Yes
25	OrcaOS	1983	9	No
26	CP/M	1982	3	No
	Missing		79	

CPU

[Variable name: *cpu*; 2 digit numeric variable]: Central processing unit used in machine. The table below provides the codes, names, number of products with that CPU, whether the chip has a RISC architecture (i.e. "Reduced Instruction Set"), and the designer (and usually manufacturer) of the chip.

Early workstations incorporated processors that had been developed for other applications, but the industry quickly became a driving force in microprocessor development. The first Apollo machine as well as most of the other early computer workstations used the Motorola 68000 processor. This chip, released in 1979, offered the first mass production 32-bit processor (Intel did not release a 32-bit processor until the 80386 in 1984). RISC technology did not appear in the industry until 1988, when Sun Microsystems began to push RISC technology. It nevertheless became the dominant architecture by 1993.

Code	CPU name	First year	Number	Architecture	Designer
1	SPARC	1987	322	RISC	Sun
2	SuperSPARC	1993	125	RISC	Sun
3	ALPHA	1994	74	RISC	DEC
4	IBM Power	1994	37	CISC	IBM
5	microSPARC II	1993	54	RISC	Sun
6	HyperSPARC	1995	12	RISC	Sun
8	MIPS R4400	1993	49	RISC	MIPS
9	MIPS R3000	1990	175	RISC	MIPS
10	MIPS R4000	1992	42	RISC	MIPS
11	PA-RISC	1993	53	RISC	HP
12	PowerPC 604	1995	15	RISC	IBM
13	SuperSPARC II	1996	2	RISC	Sun
14	Pentium	1995	28	CISC	Intel
15	Intel 80960	1992	143	CISC	Intel
16	MC68020	1984	328	CISC	Motorola
17	MC68030	1986	163	CISC	Motorola
18	MC68040	1991	34	CISC	Motorola
19	MC88100	1992	67	CISC	Motorola
20	Alpha AXP	1994	24	CISC	DEC
21	CVAX	1986	72	CISC	DEC
22	MC68000	1980	85	CISC	Motorola
23	SPARC64	1996	2	RISC	Sun
24	RS6000	1994	10	CISC	IBM
27	TMS320C65	1991	6	CISC	TI
28	MIPS R4600	1995	11	RISC	MIPS
29	MIPS R8000	1994	15	RISC	MIPS
30	Unknown	1980	402	Unknown	Unnamed
31	TMS34020	1989	29	CISC	TI
32	LSI R33020	1994	6	CISC	LSI
33	80486	1991	21	CISC	Intel
34	AMD 29000	1995	14	CISC	AMD
35	MIPS R6000	1991	3	RISC	MIPS
36	80860	1992	8	CISC	Intel
37	MIPS R2000	1990	14	RISC	MIPS
38	MC88000	1990	34	CISC	Motorola
39	MC68010	1983	71	CISC	Motorola
40	80286	1987	32	CISC	Intel
41	8085	1991	6	CISC	Intel
42	Clipper C300	1990	11	CISC	Intergraph
43	Clipper C400	1993	3	CISC	Intergraph
44	Sun-3/4	1990	3	CISC	Sun
45	NEC V30	1988	5	CISC	NEC
46	AMD 2901	1990	14	CISC	AMD
47	Clipper C100	1990	4	CISC	Intergraph
48	8086	1983	10	CISC	Intel
49	Whizzard	1981	47	CISC	Megatek
50	MicroVax	1983	14	CISC	DEC
51	Apollo 32	1982	10	CISC	Apollo
52	Apollo S	1986	12	CISC	Apollo

Architecture

[Variable name: *bus*; 2 digit numeric variable]: Architecture for integrating the various components of the computer system. The most common buses in personal computers were ISA, then EISA (and IBM's proprietary MCA) and then PCI, but computer workstation manufacturers tended to design their own buses to improve system performance. The table below provides the codes, names, year of first appearance, and the number of product-years with that bus.

Code	Bus name	First year	Number
1	VME bus	1982	504
2	Sbus	1990	376
3	ISA w/ PCI	1985	86
4	Mbus	1993	8
5	Proprietary VME-type	1994	5
6	EISA	1992	67
7	Unnamed proprietary	1980	1,318
8	PCI local bus	1994	28
9	MCA	1992	97
10	NuBus	1990	5
11	Multibus	1983	4
	Missing		237

Clock Speed

[Variable name: *mhz*; 3 digit numeric variable]: Clock speed of machine in cycles (megahertz). Note that one cannot readily compare the speed of CISC and RISC machines because of their architectural differences.

Memory

[Variable name: *ram*; 7 digit numeric variable]: Amount of random access memory installed in the standard configuration of the machine (in kilobytes).

Storage

[Variable name: *hd*; 6 digit numeric variable]: Size of hard drive installed on the machine in megabytes.

Display

[Variable name: *monitor*; 1 digit numeric variable]: A code indicating whether the system has an external monitor (=0), or an internal monitor (=1).

List Price

[Variable name: *price*; 6 digit numeric variable]: List price of standard configuration in U.S. dollars. Please note that I have not deflated any of these amounts to adjust for inflation. Though not consistent across manufacturers, these prices generally do not include the cost of the monitor.

10. Variable List and Definitions: File 2: SorensonworkfirmFIVEdata

Descriptive statistics for the continuous and indicator variables below appear in the Tables section.

10.1. Year of Observation:

Year

[Variable name: *year*; 4 digit numeric variable]: Year of observation, ranging from 1980 to 1996.

10.2. Firm Identifiers:

Firm Name

[Variable name: *firmname*; 40 character string variable]: Name of company.

FIVE Firm ID (unique firm identifier for the FIVE project)

[Variable name; *fivefirmID*; 7 digit numeric variable]: Unique identifier for each *Firm Name* that appears in the FIVE data sets. Firm names and FIVE firm identifiers may appear in more than one FIVE data set. If a firm's name changed over time, the FIVE Firm ID for the firm sometimes changes as well. In addition, different FIVE data set authors may have coded different Firm Names for the same company, which could result in different FIVE Firm IDs for the same firm in different data sets.

Data Set Firm ID (firm identifier assigned in the original data set)

[Variable name: *firmID*; 3 digit numeric variable]: Unique identifier that I created for each firm in the original data set.

CUSIP Header ID (firm identifier assigned by Compustat)

[Variable name: CUSIP*headerID*; 8 digit numeric variable]: Unique identifier for a firm assigned by Compustat for 2007. For a given firm name, this ID is constant across all years in the data set. These IDs come from matching each Firm Name in the FIVE data sets to an identical or similar firm name in Compustat. ***There is no guarantee that these matches and IDs are accurate.*** If no match was found, the CUSIP Header ID is blank.

CUSIP Header Firm Name (company name assigned by Compustat)
[Variable name: CUSIPheadername; character variable]: Name of company given in Compustat for the CUSIP Header ID.

CUSIP Historical ID (firm identifier assigned by Compustat)
[Variable name: CUSIP*historyID*; 8 digit numeric variable]: Unique identifier for a firm assigned by Compustat in their historical databases. For each Firm Name, the CUSIP Historical ID is that shown in Compustat for the year of each observation in the FIVE data set. These IDs come from matching each Firm Name in the FIVE data sets to an identical or similar firm name in Compustat. ***There is no guarantee that these matches and IDs are accurate.*** If no match was found for a given Firm Name in a given year, the CUSIP Historical ID is blank. This means that a firm may have CUSIP Historical IDs assigned for some but not all years in which the firm appears in the data set.

CUSIP Historical Firm Name (company name assigned by Computstat)
[Variable name: CUSIPhistoryname; character variable]: Name of company given in Compustat for the CUSIP Historical ID.

10.3. Entry/Exit Variables:

Year of Founding
[Variable name: *found*; 4 digit numeric variable]: Year in which the company began operations (not the year that it entered the computer workstation market). I assessed market entry from the product-level data as the year in which it first offered a computer workstation product.

For some analyses (Sorenson, 2000), I created either a continuous or dummy variable for de alio status. The continuous version simply calculated the number of years the firm had been in operation prior to entering the computer workstation market. The dichotomous version collapsed this measure to 0 versus 1 (for any number of years greater than zero).

Notes on Exit Event
[Variable name: *fate*; 50 character string variable]: Notes on the nature of exit events, acquisitions, name changes, etc.

Nature of Exit Event
[Variable name: *estate*; 1 digit numeric variable]: A code indicating whether the observation is left-censored (=0), has exited the market (=1), has been acquired and continued operations (=2), has spun off and changed its name (=3), or has changed its name (=4). In exit rate analyses, I grouped acquisitions with failures.

In 13 of the 14 cases, the press discusses the firms' financial troubles prior to being acquired.

10.4. Firm Size Variables:

Revenue

[Variable name: *sales*; 10 digit numeric variable]: Company sales in U.S. dollars; these figures also include sales outside of the computer workstation market. Please note that I have not deflated any of these amounts to adjust for inflation. For public companies, these figures came from annual reports. For private companies, they came from Data Sources or Lexis-Nexis searches.

Firms without information tend to be quite small. To deal with the missing cases in analyses that use this variable as a control in my dissertation, I assigned them a random value from the bottom decile of firms with information available for a particular time window: 351,000 to 2,475,000 for missing fields from 1980 to 1984; 867,000 to 3,400,000 for 1985 to 1988; 180,000 to 4,900,000 for 1989 to 1992; and 180,000 to 5,400,000 for 1989 to 1992. The low end of the size range drops beginning in 1989 due to the entry of "clone" manufacturers. See Barron et al, 1994, for a discussion of this technique for dealing with missing fields. Tests of the validity of this method revealed no significant difference between the exit rates of these firms and those of other firms in the bottom decile (based on a t-test of a dummy variable for those cases with randomly assigned sizes).

Number of Employees

[Variable name: *employ*; 6 digit numeric variable]: Number of people employed at the firm in a given year; these figures also include employees outside of the computer workstation division for companies with other products.

10.5. Location Variables:

Zip Code

[Variable name: *zip*; 5 digit numeric variable]: Zip code of corporate headquarters. With the exception of a few of the largest firms, almost none of the companies in the data have more than one facility. In general, those interested in the geography of the industry will find the latitude and longitude measures more useful.

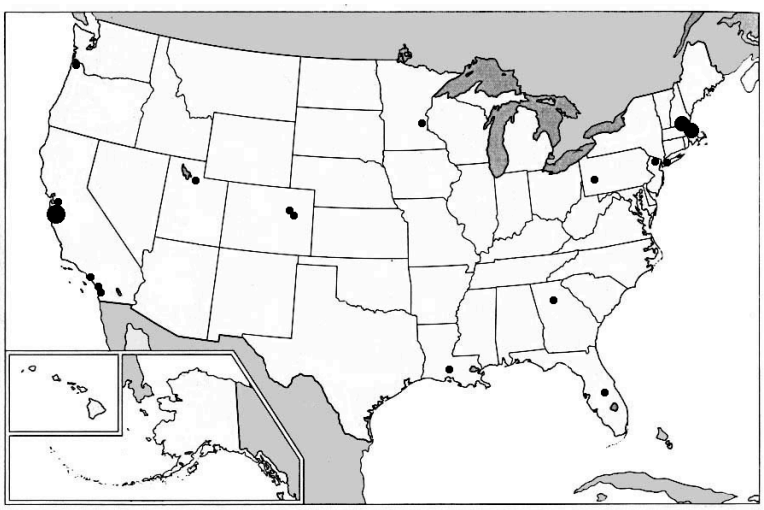

Figure 1: *Distribution of firms in 1984*

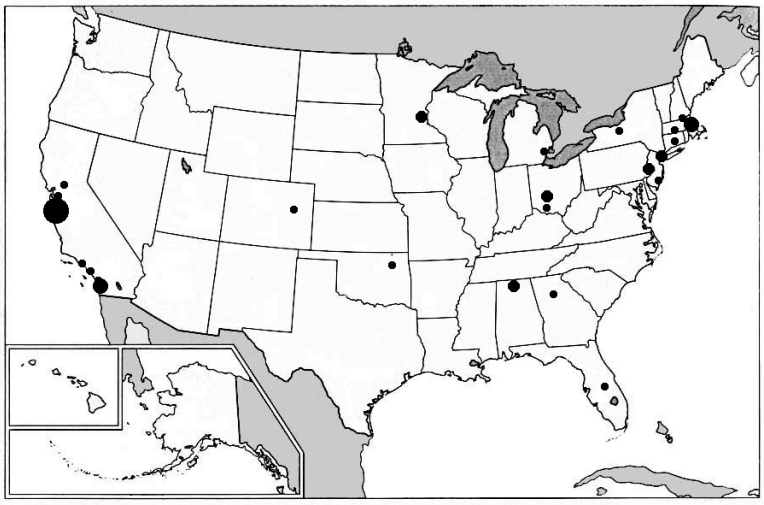

Figure 2: *Distribution of firms in 1996*

Latitude

[Variable name: *lat*; 7 digit numeric variable]: Latitude of the headquarters in radians.

Longitude

[Variable name: *lon*; 7 digit numeric variable]: Longitude of the headquarters in radians.

10.6. Research and Development Variables:

R&D Expenditure

[Variable name: *RandD*; 9 digit numeric variable]: Amount of money spent on R&D at the firm in a given year; these figures also include spending outside of computer workstations.

10.7. Products and Markets Variables:

Number of Computer Products

[Variable name: *products*; 3 digit numeric variable]: A count of the number of products that the firm offers in all categories covered by Data Sources.

Number of Computer Markets

[Variable name: *ptypes*; 2 digit numeric variable]: A count of the number of product categories of Data Sources in which the firm offers at least one product.

Product Section 1

[Variable name: *system*; 1 digit numeric variable]: A dummy variable indicating whether the firm has products listed in the "Desktop Systems" category.

Product Section 2

[Variable name: *graphic*; 1 digit numeric variable]: A dummy variable indicating whether the firm has products listed in the "Graphics Systems" category.

10.8. Other Variables:

CEO

[Variable name: *ceo*; 30 character string variable]: Name of CEO.

Ownership

[Variable name: *own*; 1 digit numeric variable]: A code indicating whether the firm is private (=0), public (=1) or a subsidiary of another firm (=2).

Vertical Integration

A series of indicator variables capture whether the firm produces products that could serve as subsystems for their computer workstation products. In general, I do not know if they incorporated these components or whether they produced all of them (e.g., firms could engage in "tapered" integration---sourcing a portion of their needs for an item internally and a portion externally). The variables are: *cpu* - central processing unit; *os* - operating system; *apps* - software applications; *comm* - communications hardware; *monitor* - monitor; *disk* - hard disk drives; *memory* - random access memory; and *board* - motherboard.

