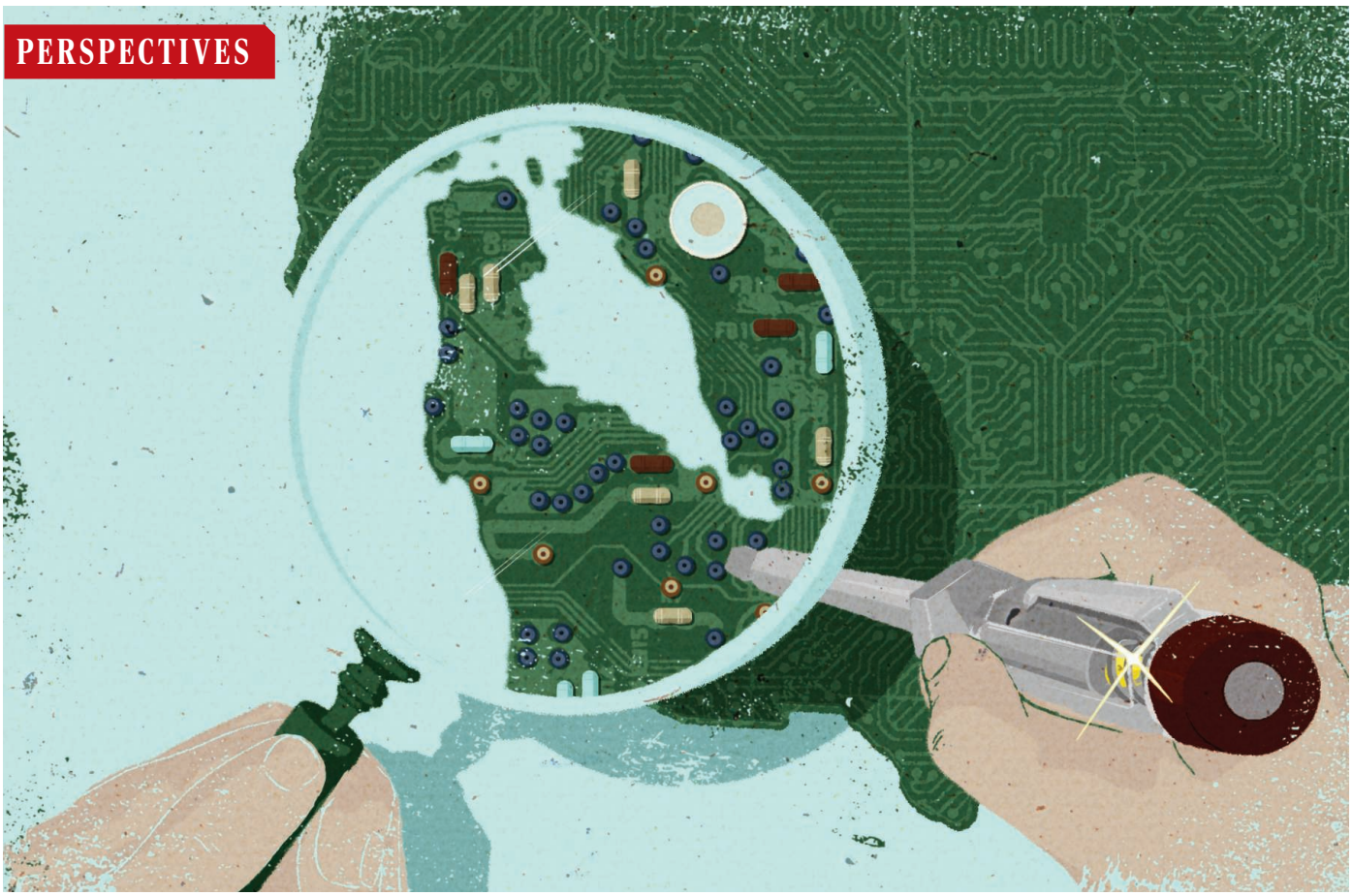
INSIGHTS

Secondary response to infection *p. 609*

Surface control of nanoparticle properties *p. 615*

PERSPECTIVES



INNOVATION ECONOMICS

Where is Silicon Valley?

Forecasting and mapping entrepreneurial quality

By **Jorge Guzman**[1] *and* **Scott Stern**[1,2*]

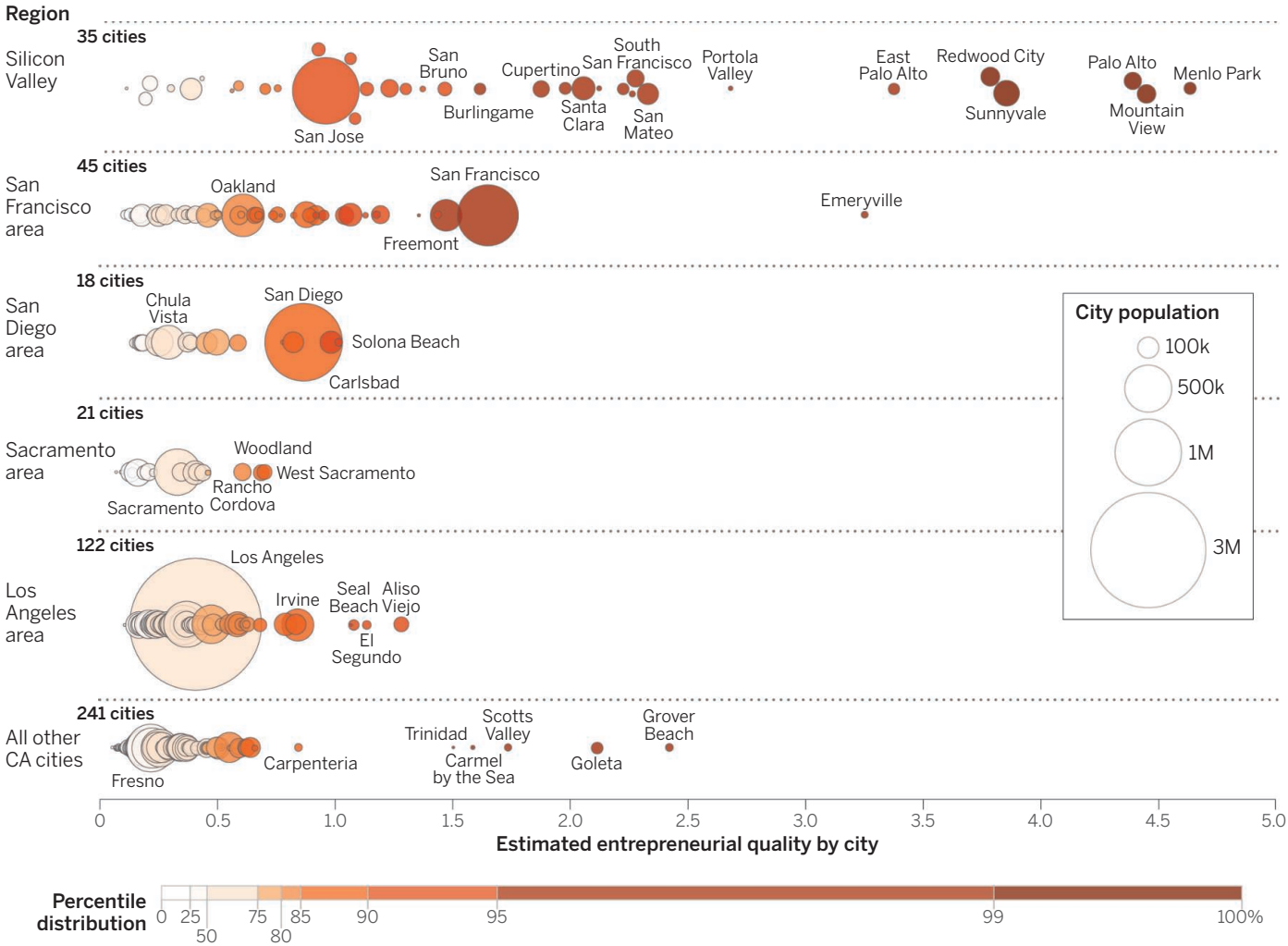
Although economists, politicians, and business leaders have long emphasized the importance of entrepreneurship (*1*, *2*), defining and characterizing entrepreneurship has been elusive (*3*, *4*). Researchers have been unable to systematically connect the type of high-impact entrepreneurship found in regions such as Silicon Valley with the overall incidence of entrepreneurship in the population (*5*–*7*). This has important implications: Researchers arrive at alternative conclusions about roles and patterns of entrepreneurship (*8*–*10*), and policy-makers are given conflicting recommendations about whether or how to promote entrepreneurship for economic and social progress (*11*, *12*).

To break this impasse, we introduce a new method for studying the founding and growth of entrepreneurial ventures. Whereas most prior studies have focused on the quantity of entrepreneurial ventures (e.g., the number of new businesses per capita in a given region), we focus on characterizing their quality. Rather than assume that all ventures have an equal ex ante probability of success, our method allows us to estimate the probability of growth based on information publicly available at or near the time of founding.

POLICY

We implement our approach using for-profit business registrations in California from 2001 to 2011 (*13*), combined with data from the U.S. Patent and Trademark Office and SDC Platinum [details on data and methods are in the supplementary materials (SM)]. We estimate outcomes on the basis of a small number of start-up characteristics: (i) firm name characteristics, including whether the firm name is eponymous [named after

California quality is all over the map



Estimating entrepreneurial quality by city. Each bubble represents a city. Bubble size reflects city population. Bubble color varies according to quality scale at bottom of figure. Each row represents distinct geographic region. See SM.

the founder (*14*)], is short or long, is associated with local business activity or regionally traded clusters (e.g., dry cleaning versus manufactured goods), or is associated with a set of high-technology industry clusters (*15*, *16*); (ii) how the firm is registered, including whether it is a corporation [rather than partnership or limited liability company (LLC)] and whether it is incorporated in Delaware (*17*); and (iii) whether the firm establishes control over formal intellectual property (IP) rights within 1 year of registration (*18*).

To ensure that our estimate reflects the quality of start-ups in a location rather than assuming that start-ups from a given location are associated with a given level of quality, we exclude location-specific measures from the set of observable start-up characteristics.

We estimate entrepreneurial quality as the probability of achieving a meaningful growth outcome—defined as an initial public offering (IPO) or an acquisition (*19*) within 6 years of founding—as a function of these start-up characteristics. This predictive, location-agnostic algorithm can then be used to independently characterize the entrepreneurial quality of firms and locations.

ESTIMATING ENTREPRENEURIAL QUALITY. We estimate entrepreneurial quality through a logit model with a randomly selected sample of 70% of all firms registered in 2001–2006 (keeping the other 30% as a test sample). Our model incorporates business registration and IP factors in a single regression, with all coefficients significant at the 5% level (*20*) (table S1). When we look at firm name characteristics, eponymous firms are more than 70% less likely to grow than noneponymous firms, whereas firms with short names are 50% more likely to grow than firms with long names, and firms that include words associated with high-technology clusters are 92% more likely to grow than others. Looking at legal form and IP, corporations are >6 times more likely to grow than noncorporations, and firms with trademarks are >5 times more likely to grow than nontrademarked firms. Patenting and Delaware jurisdiction play an outsized role: Each alone is associated with a >25 times increase in the probability of growth relative to not being present. When both are present at the same time, there is nearly a 200 times increase in the probability of growth.

As a validation test, we estimate entrepreneurial quality for the test sample withheld from the original regression and so

[1]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [2]National Bureau of Economic Research, Cambridge, MA 02138, USA. *E-mail: sstern@mit.edu

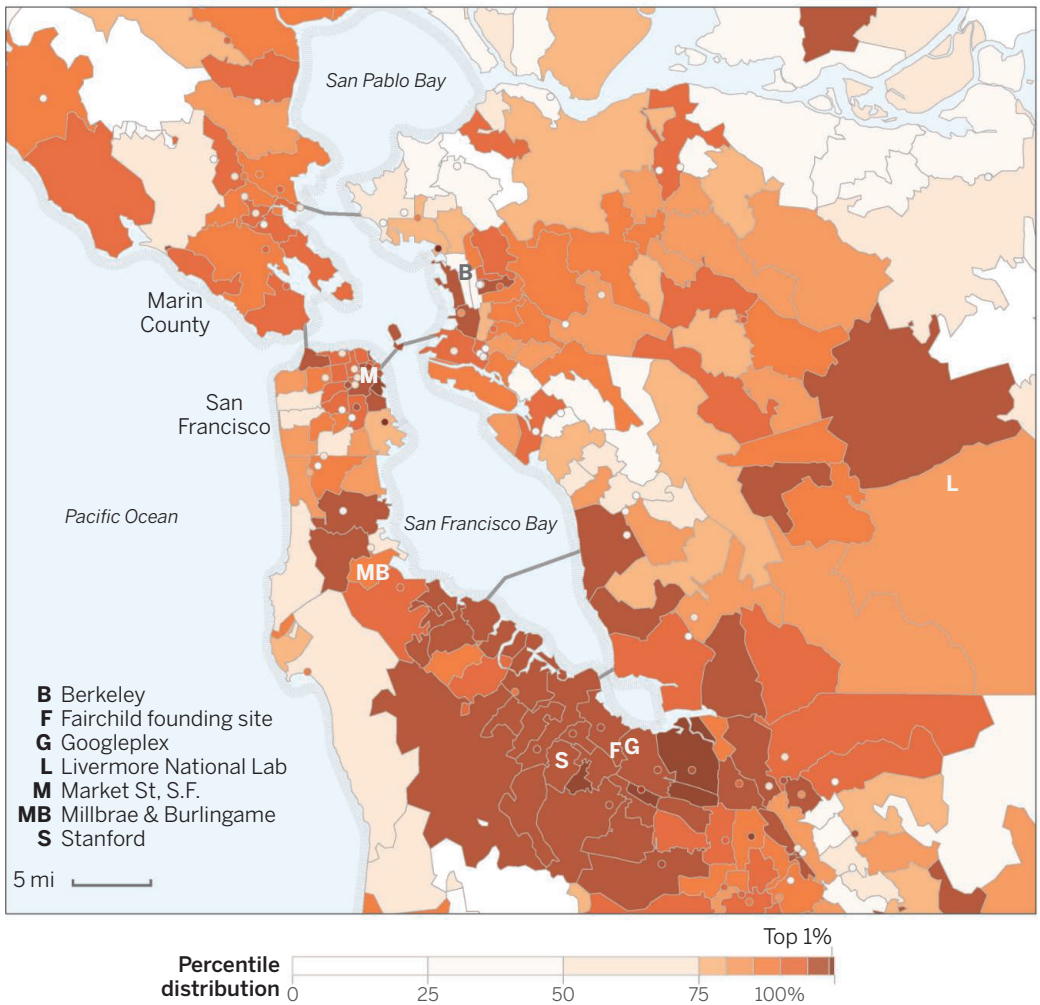
compare our predictions of entrepreneurial quality to the actual outcome distribution. Our estimate of entrepreneurial quality is strongly related to out-of-sample outcomes: 76% of all growth outcomes in the test sample are within the top 5% of the distribution of estimated entrepreneurial quality, with 56% drawn from the top 1% of that distribution (fig. S1). Highlighting the extreme uncertainty associated with entrepreneurship, growth is still rare: Even within the top 1% of estimated entrepreneurial quality, the average firm has only a 5% chance of realizing a growth outcome. This is consistent with recent findings that start-up growth is skewed relative to overall firm growth—Gibrat's law (*21*).

MAPPING ENTREPRENEURSHIP. The centerpiece of our analysis focuses on recent cohorts before a growth outcome has occurred (i.e., all start-ups from 2007 to 2011). We estimate the entrepreneurial quality for each firm and then calculate the average estimated quality of firms by city and, separately, by ZIP Code. These scores can be interpreted as the expected number of growth outcomes per 1000 start-ups in the 2007–2011 cohorts.

Average quality across municipalities is shown in the first figure. Silicon Valley stands out from other regions across California: Start-ups in Menlo Park, Mountain View, Palo Alto, and Sunnyvale have 20 times the average quality of the median city and 90 times that of the lowest-ranked cities in California. Among large cities, San Francisco registers an entrepreneurial quality level nearly 8 times that of Fresno.

Entrepreneurial quality is mapped for the San Francisco Bay area at the ZIP Code level in the second figure. The quality of entrepreneurial activity is distinctively higher in the area that ranges just north of San Jose through San Francisco, with a contiguous mass of intense entrepreneurial quality from just southeast of Google (and the founding location of Fairchild) through Milbrae and Burlingame. In contrast, the Los Angeles region has a much lower level of entrepreneurial quality (fig. S2). Large economic areas can vary significantly in their quality. We investigated the statistical relation between quality and quantity (fig. S3): At best, the relation is weak and noisy. Intriguingly, across regions, entrepreneurial quality is centered around research institutions, such as universities and national laboratories. Stanford is at the heart of Silicon Valley, and University of California (UC) Berkeley; Lawrence Livermore; Caltech; University of California, Los Angeles (UCLA); and UC Irvine each host a region of distinctive entrepreneurial quality.

Better by the Bay



Mapping estimated entrepreneurial quality by ZIP Code. San Francisco Bay area. Dots indicate single-address ZIP Codes. See SM.

IMPLICATIONS. By focusing on entrepreneurial quality, we can evaluate more clearly the role of location and institutions in firm growth. For example, our method allows us to estimate a locational entrepreneurship "premium" as the difference between realized and expected growth outcomes for a region. Between 2001 and 2006, Silicon Valley had 60% more actual growth events than predicted by our model, whereas Los Angeles registered 13% fewer than predicted.

Our method can be extended to evaluate entrepreneurial quality at arbitrary levels of geographic aggregation (e.g., a specific street in Palo Alto) (fig. S4). This facilitates fine-grained analysis of entrepreneurial dynamics (*22*), distinguishing empirically (although not causally) between locations at a high level of granularity.

Finally, beyond our characterization of Silicon Valley in the aggregate, our results highlight the role of research institutions as centers of entrepreneurial quality. Characterizing the two-way relation between entrepreneurial quality and scientific research activity is a promising agenda for future research. Although one would need to be cautious about using these estimates as a policy tool (for example, one could imagine "gaming" of various sorts), clarifying the conditions that facilitate positive growth outcomes has important implications for policy-makers and regional stakeholders.

REFERENCES AND NOTES

1. Z. Acs, D. Audretsch, *Innovation and Small Firms* (MIT Press, Cambridge, MA, 1990).
2. W. Baumol, R. Litan, C. J. Schramm, *Good Capitalism, Bad Capitalism, and the Economics of Growth and Prosperity* (Yale Univ. Press, New Haven, CT, 2007).
3. E. Hurst, B. W. Pugsley, "What do small businesses do?" (Brookings paper 73-118, Brookings Institution, Washington, DC, 2011).
4. M. Henrekson, T. Sanandaji, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1760 (2014).
5. R. W. Fairlie, *Kaufman Index of Entrepreneurial Activity: 1996–2013* (Ewing Marion Kaufman Foundation, Kansas City, MO, 2014).
6. L. Klapper, R. Amit, M. F. Guillén, in *International Differences in Entrepreneurship*, J. Lerner, A. Schoar, Eds. (Univ. of Chicago Press, Chicago, 2010), pp. 130–158.
7. J. E. Amoros, N. Bosma, *Global Entrepreneurship Monitor: 2013 Executive Report* (London Business School, London, and Babson College, Wellesley, MA, 2014).
8. R. Decker, J. Haltiwanger, R. Jarmin, J. Miranda, *J. Econ. Perspect.* **28**, 3 (2014).
9. S. Kaplan, F. Murray, *Technology and Organization*, N. Phillips, G. Sewell, D. Griffiths, Eds. (Emerald Group, Bingley, UK, 2010), pp. 107–147.
10. S. Shane, S. Venkatamaran, *Acad. Manage. Rev.* **25**, 217 (2000).
11. J. Lerner, *Boulevard of Broken Dreams* (Princeton Univ, Press, Princeton, NJ, 2009).
12. S. Samila, O. Sorenson, *Res. Policy* **39**, 1348 (2010).
13. Formal registration includes corporation, LLC, limited partnership, and general partnership.
14. S. Belenzon, A. K. Chatterji, B. Daley, *Eponymous Entrepreneurs* (2014); https://faculty.fuqua.duke.edu/~bd28/BCD_EE.pdf.
15. M. Delgado, M. E. Porter, S. Stern, "Defining clusters of related industries" (NBER Working Paper 20375, National Bureau of Economic Research, Cambridge, MA, 2014); www.nber.org/papers/w20375.
16. We define traded and local industries in line with the definition used in the economic cluster literature [e.g., (*15*)], and high-technology clusters are drawn from the U.S. Cluster Mapping Project [see (*15*)].
17. Many firms with the intention to grow register in the state of Delaware, where corporate law is beneficial owing to a large legal canon. Venture capitalists often prefer companies to incorporate in Delaware.
18. Our use of firm names builds on a basic assumption that entrepreneurs choose firm names conscientiously to serve as a signal to consumers, investors, and employees and that there are costs in impersonating a different type of firm.
19. An IPO or acquisition represents a significant and observable equity growth outcome from the perspective of the founders. Our ongoing research agenda also explores alternative growth outcomes in terms of employment, firm revenues, and so on.
20. Results are similar when we look at business registration and IP factors alone; see columns 1 and 2 in fig. S1.
21. L. M. B. Cabral, J. Mata, *Am. Econ. Rev.* **93**, 1075 (2003).
22. J. Guzman, S. Stern, *Nowcasting and Placecasting Entrepreneurial Quality and Performance* (NBER, Cambridge, MA, 2014); www.nber.org/chapters/c13493.pdf.

ACKNOWLEDGMENTS

We thank the Jean Hammond (1986) and Michael Krasner (1974) Entrepreneurship Fund and the Edward B. Roberts (1957) Entrepreneurship Fund for financial support. We thank participants in the Massachusetts Institute of Technology (MIT) Innovation and the Digital Economy Seminar, the MIT Regional Entrepreneurship Acceleration Program, the Micro@Sloan Seminar, the Asian Innovation and Entrepreneurship Association–NBER Conference, and three anonymous referees for comments. I. Cockburn, M. Delgado, J. Gans, H. Varian, and C. Fazio provided valuable advice. We thank R. J. Andrews for excellent development of figures and visualizations, A. Carracuzzo and MIT Libraries for data support, and I. DiMambro for editorial assistance.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/347/6222/606/suppl/DC1

10.1126/science.aaa0201

IMMUNOLOGY

There goes the macrophage neighborhood

Migrating dendritic cells disrupt lymph node macrophages and limit the immune response to secondary infection

By **Heather D. Hickman**

The lymph node is a highly structured organ optimized for generating adaptive immune responses. Lymph fluid carrying pathogens and their antigens from infected tissue is first distributed into a large cavity just beneath the node's surface, which is populated by a dense layer of specialized macrophages. These subcapsular sinus (SCS) macrophages filter incoming lymph, capture pathogens, and relay pathogen-derived antigen to B cells in subjacent follicles, provoking them to produce antibodies (see the figure). At the original infection site, migratory dendritic cells (DCs) are activated, acquire antigen, and deliver it to the node through the lymph, generating a secondary wave of immune cell activation. Until now, this influx of DCs has been viewed as beneficial to the host, as they activate T cells within the node's paracortex. However, on page 667 of this issue, Gaya *et al.* (*1*) demonstrate that incoming DCs can be harmful. These cells can disrupt the SCS macrophage layer and reduce the host's ability to mount a humoral (antibody) response to a secondary pathogen.

Resident antigen-presenting cells in the lymph node are commonly classified into two major subsets: DCs and macrophages. Both populations are a complex, heterogeneous mixture of cells with somewhat nebulous differences and overlapping capabilities. Even so, it is clear that different cellular subsets within each population preferentially localize to distinct regions of the lymph node where they can optimally activate discrete aspects of immune responses (*2*). For example, CD8α DCs reside in the interior of the node, are efficient exogenous antigen gatherers, and are needed for optimal T cell activation after viral infection (*3*). Several subsets of DCs are not present (in appreciable numbers) in steady-state lymph nodes, but traffic to nodes from peripheral tissue sites after infection or inflammation. Because activation, and particularly migration, take time, hours to days may elapse before immigrant DCs can influence the immune response. It is unclear how these migratory DCs precisely navigate nodal architecture to situate themselves in the node's interior; however, their arrival is essential for eliciting maximal T cell responses to many pathogens (*4*).

SCS macrophages, typically distinguished by the expression of the cell surface marker CD169 and the absence of F4/80 (found on medullary macrophages in the node), form a sessile, carpet-like layer along the floor of the SCS. After subcutaneous injection of viruses or antigen-antibody immune complexes, SCS macrophages transfer antigen on cellular processes to closely apposed B cells that lack direct access to SCS contents (*5*–*8*). This antigen-capture process both activates B cells and removes infectious material from the lymph, preventing entry into the bloodstream. Accordingly, depletion of SCS macrophages from the node before infection can result in failure to control pathogen dissemination, leading to the infection of distal organs (*6*, *9*, *10*).

> ***"The evolutionary advantage of reduced responses to ... secondary challenges is puzzling."***

Although carefully scrutinized previously with primary infection models, the behavior and function of SCS macrophages have not been systematically followed for extended periods after infection. To close this gap, Gaya *et al.* used sophisticated techniques to image skin-draining murine lymph nodes 1 week after cutaneous infection with a variety of pathogens (including *Staphylococcus*, group B *Streptococcus*, and vaccinia virus). Intriguingly, the authors observed fragmentation of the SCS macrophage layer after infection with any of the pathogens, with as much as 80% of the layer disrupted. Gaya *et al.* also assessed the ability of various additional stimuli to deplete the SCS macrophage layer. Whereas the injection of inert beads or dead

National Institutes of Health, Bethesda, MD, USA. E-mail: hhickman@mail.nih.gov

www.sciencemag.org/content/347/6222/606/suppl/DC1

Supplementary Materials for

Where is Silicon Valley?

Jorge Guzman and Scott Stern*

*Corresponding author. E-mail: sstern@mit.edu

Published 6 February 2015, *Science* **347**, 606 (2015)
DOI: 10.1126/science.aaa0201

This PDF file includes:

CONTENTS

Table S1. Logit regression on growth (IPO or acquisition within 6 years) as dependent variable. Coefficients reported are incidence rate ratios (changes in probability relative to 1). Growth is a binary variable equal to 1 if a firm achieves an IPO or significant acquisition within 6 years of founding. Robust standard errors in brackets.

	Business registration observables	**External observables**	**All**
	(1)	**(2)**	**(3)**
Eponymous	0.195*		0.239*
	[0.099]		[0.12]
Short Name	1.973**		1.504*
	[0.24]		[0.19]
Local	0.314*		0.378*
	[0.11]		[0.14]
Technology	3.499**		1.918+
	[0.80]		[0.52]
Corporation	9.223**		6.117**
	[1.67]		[1.15]
Delaware Jurisdiction	54.65**		
	[6.84]		
Trademark		8.252**	5.364**
		[1.74]	[0.90]
Patent		54.45**	
		[10.2]	
Interactions			
Patent Only			25.02**
			[6.93]
Delaware Only			35.93**
			[5.48]
Patent and Delaware			195.5**
			[36.5]
Constant	0.0000324**	0.000337**	0.0000398**
	[0.0000067]	[0.000023]	[0.0000085]
N	585,162	585,162	585,162
Pseudo-R2 (%)	25	19	32

+ p<.05, * p<0.01, ** p<.001.

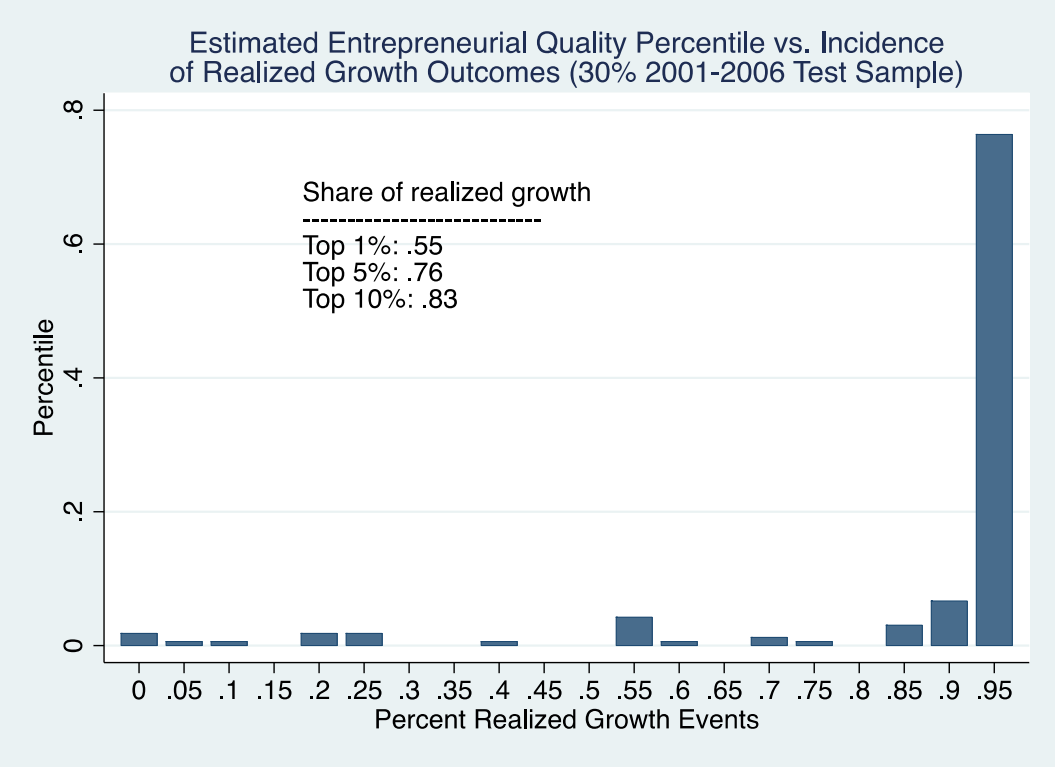


Fig. S1. Estimated entrepreneurial quality percentile vs. incidence of realized growth outcomes (30% 2001–2006 test sample).

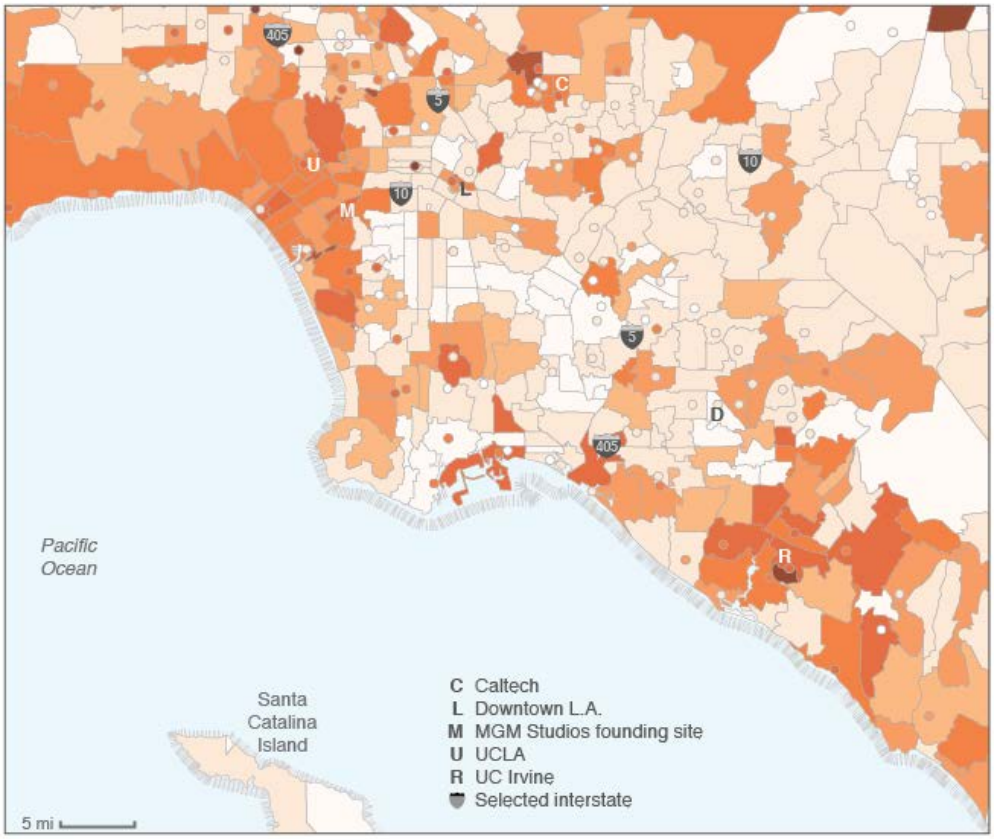


Fig S2. Los Angeles Region estimated entrepreneurial quality map by ZIP Code.

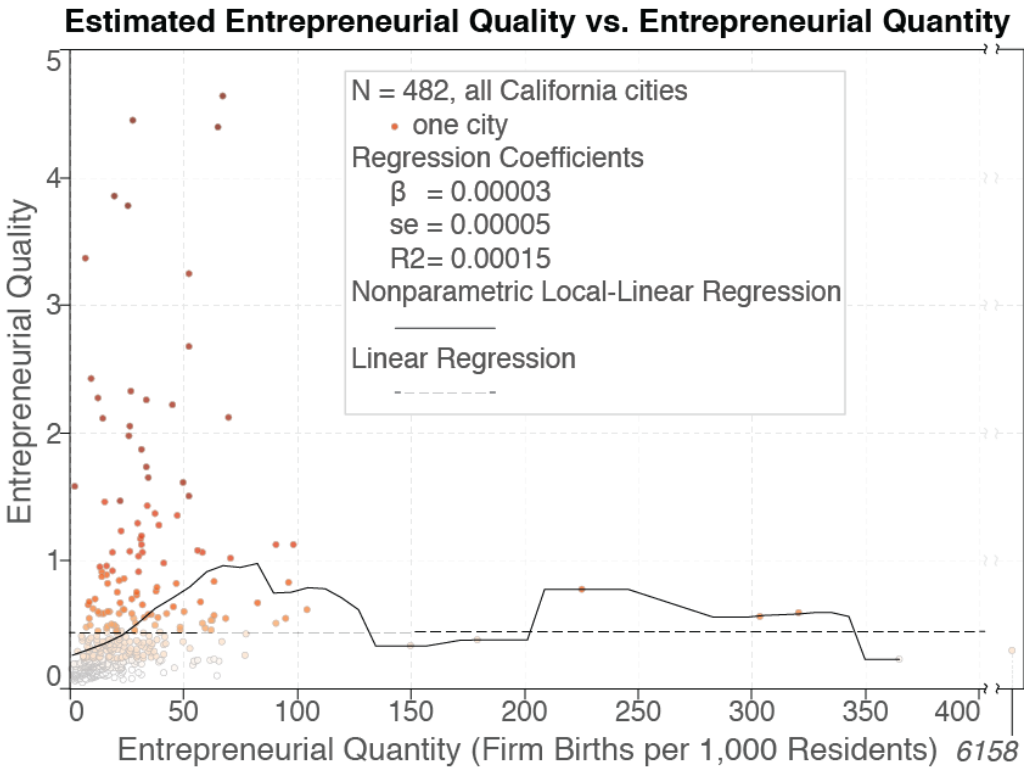
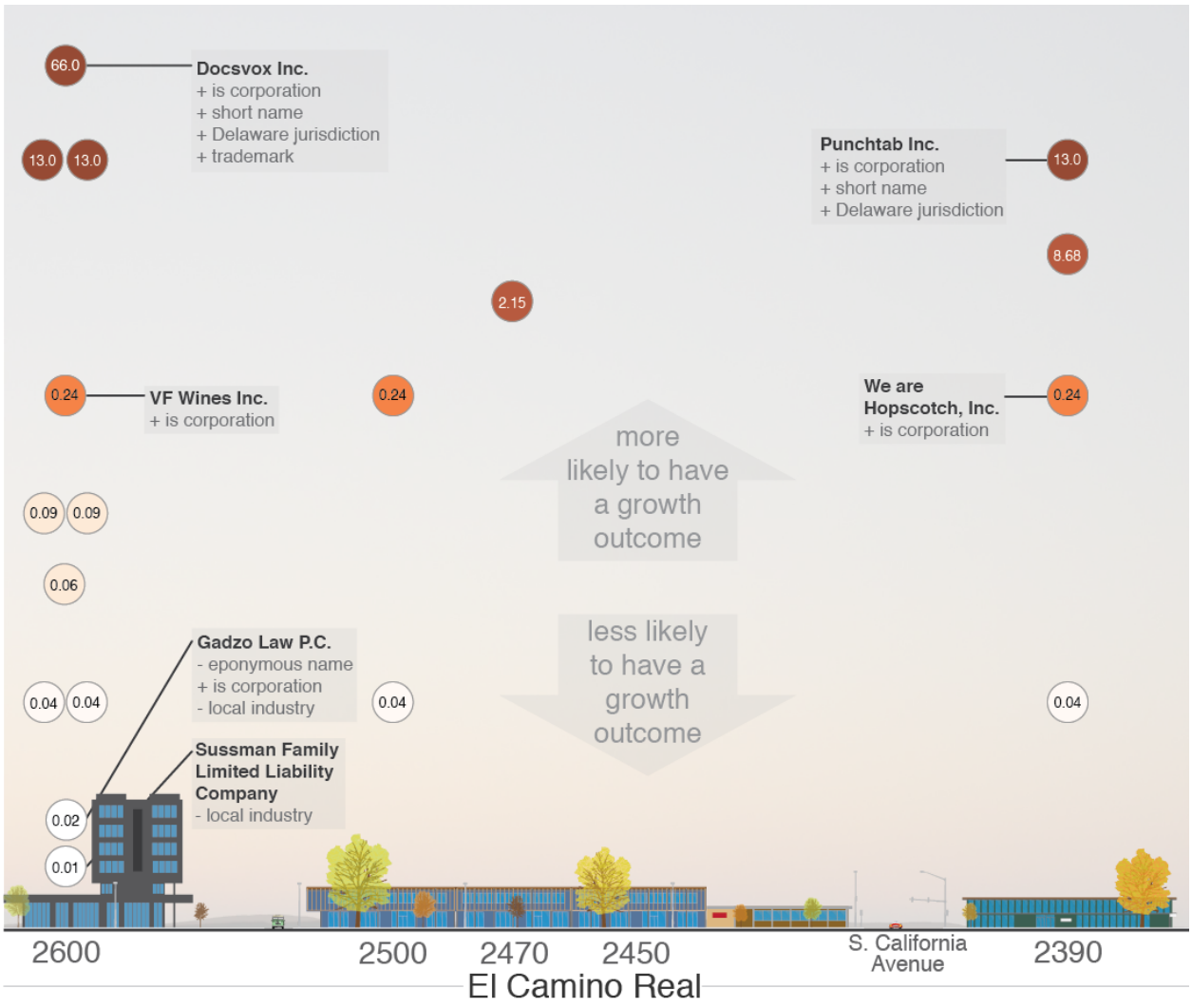


Fig. S3. Estimated entrepreneurial quality vs. entrepreneurial quantity.



The micro-geography of estimated entrepreneurial quality. This figure renders the estimated entrepreneurial quality of start-ups founded in 2010 and 2011 within a two-block length of El Camino Real in Palo Alto, CA. The blocks were selected due to the relatively high incidence of new business registrants within this location.

Fig. S4. The microgeography of entrepreneurial quality.

I. California Business Registration Records

Business registration records are a potentially rich and systematic data source for entrepreneurship and business dynamics. While it is possible to found a new business without business registration (e.g., a sole proprietorship), the benefits of registration are substantial, including limited liability, protection of the entrepreneur's personal assets, various tax benefits, the ability to issue and trade ownership shares, credibility with potential customers, and the ability to deduct expenses. Among business registrants, there are several categories, and the precise rules governing each category varies by jurisdiction and time. This study focuses on the state of California from 2001 to 2011, at which point one could register the following: corporations, limited liability companies, limited liability partnerships, limited partnerships, and general partnerships [see (*1*) for further information].

The data in this paper comes from the California Secretary of State (http://kepler.sos.ca.gov, data received on January 24, 2014) containing three files, two for corporation records and one for partnerships (which also include LLCs). The first corporation file contains a master record of all firms ever registered in California as that record exists at the moment of extraction; the second corporation file contains a record of all changes to each file on record. The corporation master file includes the following fields: corporation id; incorporation date; tax status (nonprofit or for profit); firm status (active, deceased, merged, etc.); jurisdiction (California or another U.S. state); address; firm name; name of president or manager; address of the principal office (for firms foreign to California); and California county (for California firms). The partnership master file includes the following fields: a firm id; registration date; firm status (active, deceased, merged, etc.); jurisdiction (California, or another U.S. state); address; address of the principal office; and up to two general partners or managers.

After combining these files, we generate unique firm identifiers. For this paper, we select a data set of the for-profit firms first registered in California from 1 January 2001 to 31 December 2011, satisfying one of the following two conditions: (i) for-profit firms whose jurisdiction is California and (ii) for-profit Delaware firms whose main office is in California. Table S2 lists the number of observations in our data set for each annual cohort year from 2001 to 2011. It is useful to note that, for those firms registered in Delaware, we use the year they registered in California as their founding date. Both the links to the underlying data and the program files used to construct the data set are available through links in the Materials section.

As a final note, this paper uses a subset of the business registration records that we have now gathered from several states, including Massachusetts, Texas, Florida, Washington, and New York. Although our evaluation of these additional states is at a more preliminary stage, we have found very similar qualitative findings in terms of the impact of factors observable at or near the time of registration on subsequent growth outcomes and the ability of these models to offer detailed characterization of growth entrepreneurship clusters (e.g., the identification of the role of the Route 128 corridor and the Kendall Square area near MIT for growth entrepreneurship in Massachusetts). See Guzman and Stern (*2*) for further details.

Table S2. Number of observations per year. N is the number of observations after limiting the sample to for-profit firms registered in California and for-profit firms registered in Delaware with their main office in California.

Year	N*	Share of Total	Cumulative Share
2001	102,350	6.44	6.44
2002	117,776	7.41	13.84
2003	130,718	8.22	22.06
2004	150,680	9.47	31.54
2005	167,186	10.51	42.05
2006	167,236	10.52	52.56
2007	171,675	10.79	63.36
2008	156,165	9.82	73.18
2009	141,710	8.91	82.09
2010	139,968	8.80	90.89
2011	144,906	9.11	100

II. Data and Methods

Our data set is drawn from the complete set of California business registrants from 2001 to 2011. Our analysis draws on the complete population of firms satisfying one of the following conditions: (i) a for-profit firm whose jurisdiction is in California or (ii) a for-profit firm whose jurisdiction is in Delaware but whose principal office address is in California. The resulting data set is composed of 1,590,370 observations. For each observation, we construct variables related to (i) the growth outcome for the startup, (ii) measures based on business registration observables and (iii) measures based on external observables that can be linked to the startup.

Growth outcome. The growth outcome utilized in this paper, *Growth*, is a dummy variable equal to 1 if the startup achieves an initial public offering (IPO) or is acquired at a meaningful positive valuation within 6 years of registration. Both outcomes, IPO and acquisitions, are drawn from Thomson Reuters SDC Platinum (*3*). Although the coverage of IPOs is likely to be nearly comprehensive, the SDC data set excludes some acquisitions. However, although the coverage of significant acquisitions is not universal in the SDC data set, previous studies have "audited" the SDC data to estimate its reliability, finding a nearly 95% accuracy (*4*). We observe 501 positive growth outcomes for the 2001–2006 start-up cohorts), yielding a mean for *Growth* of 0.0006. The median acquisition price is $155.8 million (ranging from a minimum of $9.7 million to a maximum of $21.6 billion). In our main results, we assign acquisitions with an unrecorded acquisitions price as a positive growth outcome, because an evaluation of those deals suggests that most reported acquisitions were likely in excess of $5 million. In unreported specifications, we drop firms where the acquisition price is not reported; neither the pattern of coefficient estimates nor our qualitative findings regarding the geography of growth entrepreneurship is affected by this choice. All of our results are also robust to including firms registered in Delaware into cohorts associated with their initial Delaware registration date (rather than their California registration date).

Start-up characteristics. The core of the empirical approach is to map growth outcomes to observable characteristics of start-ups at or near the time of business registration. We develop two types of measures: (i) measures based on business registration observables and (ii) measures based on external indicators of start-up quality that are observable at or near the time of business registration. We review each of these in turn.

Measures based on business registration observables. We construct six measures of start-up quality based on information directly observable from the business registration record. First, we create binary measures related to how the firm is registered, including *corporation*, whether the firm is a corporation (rather than partnership or LLC) and *Delaware jurisdiction*, whether the firm is incorporated in Delaware. C*orporation* is an indicator equal to 1 if the firm is registered as a corporation and 0 if it is registered either as an LLC or partnership.[1] In the period of 2001 to 2006, 0.13% of corporations achieve a growth outcome versus only 0.03% of noncorporations. *Delaware jurisdiction* is equal to 1 if the firm is registered under Delaware, but has its main office in California (all other foreign firms are dropped before analysis). Delaware jurisdiction is favorable for firms which, due to more complex operations, require more certainty in corporate law, but it is associated with extra costs and time to establish and maintain two registrations. Between 2001 and 2006, 4.5% of the sample registers in Delaware; 70% of firms achieving a growth outcome do so.

[1]Previous research highlights performance differences between incorporated and unincorporated entrepreneurs (*5*).

Second, we create four measures that are based on the name of the firm, including a measure associated with whether the firm name is eponymous (named after the founder), is short or long, is associated with local industries (rather than traded), or is associated with a set of high-technology industry clusters.

Drawing on the recent work of Belenzon, Chatterji, and Daley (*6*), we use the firm and founder name to establish whether the firm name is eponymous (i.e., named after one or more of the founders). *Eponymy* is equal to 1 if the first, middle, or last name of the top managers is part of the name of the firm itself.[2] We require names be at least four characters to reduce the likelihood of making errors from short names. Our results are robust to variations of the precise calculation of eponymy (e.g., names with a higher or lower number of minimum letters). We have also undertaken numerous checks to assess the robustness of our name matching algorithm. 10% of the firms in our training sample are eponymous [an incidence rate similar to (*6*)], though less than 3% for whom *growth* equals one. It is useful to note that, while we draw on (*6*) to develop the role of eponymy as a useful start-up characteristic, our hypothesis is somewhat different than (*6*): we hypothesize that eponymous firms are likely to be associated with lower entrepreneurial quality. Whereas (*6*) evaluates whether serial entrepreneurs are more likely to invest and grow companies which they name after themselves, we focus on the cross-sectional difference between firms with broad aspirations for growth (and so likely avoid naming the firm after the founders) versus less ambitious enterprises, such as family-owned "lifestyle" businesses.

Our second measure relates to the length of the firm name. Based on our review of naming patterns of growth-oriented start-ups versus the full business registration database, a striking feature of growth-oriented firms is that the vast majority of their names are at most two words (plus perhaps one additional word to capture organizational form (e.g., "Inc."). Companies such as Google or Spotify have sharp and distinctive names, whereas more traditional businesses often have long and descriptive names (e.g., "Green Valley Home Health Care & Hospice, Inc."). We define *short name* to be equal to one if the entire firm name has three of less words, and zero otherwise. 13.5% of firms within the 2001–2006 period have a short name, but the incidence rate among growth firms is more than 36%. We have also investigated a number of other variants (allowing more or less words, evaluating whether the name is "distinctive" (in the sense of being both noneponymous and also not an English word). While these are promising areas for future research, we found that the three-word binary variable provides a useful measure for distinguishing entrepreneurial quality.

Finally, we construct two measures based on how the firm name reflects the industry or sector that the firm within which the firm is operating. To do so, we take advantage of two features of the U.S. Cluster Mapping Project, which categorizes industries into (i) whether that industry is primarily local (demand is primarily within the region) versus traded (demand is across regions) and (ii) among traded industries, a set of 51 traded clusters of industries that share complementarities and linkages (*7*). We augment the classification scheme from the U.S. Cluster Mapping Project with the complete list of firm names and industry classifications contained in Reference USA, a business directory containing more than 10 million firm names and industry codes for companies across the United States (*8*). Using a random sample of 1.5

[2]For corporations, we consider top managers only the current president, for partnerships and LLCs, we allow for any of the two listed managers. The corporation president and two top partnership managers are listed in the business registration records themselves.

million Reference USA records, we create two indices for every word ever used in a firm name. The first of these indices measures the degree of localness, and is defined as the relative incidence of that word in firm names that are in local versus non-local industries

$$\rho_i = \frac{\sum_{j=\{local\ firms\}} \mathbf{1}[w_i \subseteq name_j]}{\sum_{j=\{non-local\ firms\}} \mathbf{1}[w_i \subseteq name_j]}).$$

We then define a list of Top Local Words, defined as those words that are (i) within the top quartile of ρ_i and (ii) have an overall rate of incidence greater than 0.01% within the population of firms in local industries (see Table S10 for the complete list). Finally, we define *local* to be equal to one for firms that have at least one of the Top Local Words in their name, and zero otherwise. Just more than 15% of firms have local names, although only 3% of firms for whom *growth* equals one. We undertake a similar exercise for the degree to which a firm name is associated with a high-technology cluster. We draw on firm names from industries include in three USCMP clusters: Aerospace Vehicles and Defense, Biopharmaceuticals, and Information Technology and Analytical Instruments. We then create a list of names, Top High-Technology Words, which includes those words within the top quartile of relative incidence within industries within these three clusters, and with an incidence greater than 0.01% within the population of firms within the high-technology clusters (see Table S11 for the complete list). *Technology* is defined as equal to one for firms that have at least one of the Top High-Technology Words in their name, and zero otherwise. Less than 1% of firms register a positive value, though 7.2% of growth firms do.

Measures based on external observables. We construct two measures related to start-up quality based on information in intellectual property data sources. Although this paper only measures external observables related to intellectual property, our approach can be utilized to measure other externally observable characteristics that may be related to entrepreneurial quality (e.g., measures related to the quality of the founding team listed in the business registration, or measures of early investments in scale (e.g., a Web presence).

Building on prior research matching business names to intellectual property (*9*, *10*), we rely on a name-matching algorithm connecting the firms in the business registration data to external data sources. Importantly, because we match only on firms located in California, and because firms names legally must be "unique" within each state's company registrar, we are able to have a reasonable level of confidence that any "exact match" by a matching procedure has indeed matched the same firm across two databases. In addition, our main results use "exact name matching" rather than "fuzzy matching"; in small-scale tests using a fuzzy matching approach [the Levenshtein edit distance (*11*)], we found that fuzzy matching yielded a high rate of false positives due to the prevalence of similarly named but distinct firms (e.g., Capital Bank v. Capitol Bank, Pacificorp Inc v. Pacificare Inc.).

Our matching algorithm works in three steps.

First, we clean the firm name by:

- expanding eight common abbreviations ("Ctr.", "Svc.", "Co.", "Inc.", "Corp.", "Univ.", "Dept.", "LLC.") in a consistent way (e.g., "Corp." to "Corporation")
- removing the word "the" from all names
- replacing "associates" for "associate"
- deleting the following special characters from the name: .|'" @_

Second, we create measures of the firm name with and without the organization type, and with and without spaces. We then match each external data source to each of these measures of the firm name. The online appendix contains all of the data and annotated code for this procedure.

This procedure yields two variables. Our first measure of intellectual property captures whether the firm is in the process of acquiring patent protection during its first year of activity. *Patent* is equal to 1 if the firm holds a patent application in the first year. All patent applications and patent application assignments are drawn from the Google U.S. Patent and Trademark Office (USPTO) Bulk Download archive. We use patent applications, rather than granted patents, because patents are granted with a lag and only applications are observable close to the data of founding. Note that we include both patent applications that were initially filed by another entity (e.g., an inventor or another firm), as well as patent applications filed by the newly founded firm. While only 0.6% of the firms in 2001–2006 have a first-year patent, 39% of growth firms do.

Our second intellectual property measure captures whether a firm registers a trademark during its first year of business activity. *Trademark* is equal to 1 if a firm applied for a trademark within the first year, and 0 otherwise. We build this measure from the Stata-ready trademark DTA file developed by the USPTO Office of Chief Economist (*12*). Between 2001 and 2006, 0.7% of all firms register a trademark, while 33% of growth firms do.

Table S3. Summary statistics for training period (2001–2006). SD, standard deviation.

	All Firms			**Growth = 0**			**Growth = 1**		
Characteristic	**N**	**Mean**	**SD**	**N**	**Mean**	**Std. Dev.**	**N**	**Mean**	**SD**
Eponymous	835946	0.101	0.301	835445	0.101	0.301	501	0.018	0.133
Local	835946	0.152	0.359	835445	0.152	0.359	501	0.032	0.176
Technology	835946	0.007	0.080	835445	0.006	0.080	501	0.072	0.259
Short Name	835946	0.135	0.342	835445	0.135	0.341	501	0.367	0.483
Corporation	835946	0.626	0.484	835445	0.626	0.484	501	0.872	0.334
Delaware jurisdiction	835946	0.046	0.209	835445	0.045	0.208	501	0.697	0.460
Patent	835946	0.006	0.077	835445	0.006	0.075	501	0.385	0.487
Trademark	835946	0.007	0.082	835445	0.007	0.081	501	0.255	0.437

Table S4. Summary statistics for training, test, and prediction samples. Summary statistics for all samples used in all procedures. We build our model with data from 2001 to 2006, and separate the data randomly into a 70% training and a 30% test sample. Test results are robust, with 76% of the realized growth events in the top 5% of our predicted distribution. We then predict on 2007–2011 data. We are unable to use 2012 and 2013 data for prediction because patent applications are only observable 18 months after submission and at the time we retrieved our data sets less than 18 months had elapsed since December 31st, 2012, the last day of 2012.

	Model building period: 2001–2006					**Prediction period: 2007–2011**	
	Training data ***Random 70% Sample*** **N=585162**		**Test data** ***Remaining 30% Sample*** **N=250784**		***t* test training vs. test samples**	**All Data** **N=754424**	
	Mean	**SD**	**Mean**	**SD**		**Mean**	**SD**
Eponymous	0.10035	0.30046	0.10103	0.30137	-0.95	0.07508	0.26353
Local	0.15252	0.35953	0.1509	0.35795	1.90	0.1523	0.35931
Technology	0.00655	0.08064	0.00644	0.08001	0.53	0.0061	0.07785
Short Name	0.13471	0.34142	0.1352	0.34193	-0.59	0.14074	0.34775
Corporation	0.62612	0.48383	0.62612	0.48383	0.00	0.52524	0.49936
Delaware jurisdiction	0.04577	0.20898	0.04579	0.20903	-0.05	0.05035	0.21867
Patent	0.00576	0.07569	0.00618	0.07837	-2.29	0.00572	0.07543
Trademark	0.00685	0.08251	0.00658	0.08085	1.41	0.00932	0.09609
Growth	0.00057	0.02396	0.00066	0.02564	-1.43		

Table S5. Probability of growth outcome as a function of start-up characteristics. Outcome variable growth is equal to 1 if a firm achieves an IPO or acquisition within 6 years of founding. Model 1 uses only information available form the business registry of California, model 2 only external information, and model 3 uses both. Robust standard errors in brackets.

	Regression Coef. (1)	**Regression Coef. (2)**	**Regression Coef. (3)**
Business Registration Observables	-1.633*		-1.432*
Eponymous	[0.51]		[0.51]
	2.222**		1.811**
Corporation	[0.18]		[0.19]
	-1.158*		-0.972*
Local	[0.36]		[0.37]
	1.252**		0.651+
Technology	[0.23]		[0.27]
	0.679**		0.408*
Short Name	[0.12]		[0.13]
	4.001**		
Delaware Jurisdiction	[0.13]		
External Observables			
Trademark		2.110**	1.680**
		[.210]	[0.17]
Patent		3.997**	
		[.187]	
Interaction			3.220**
Patent Only			[0.28]
			3.581**
Delaware Only			[0.15]
			5.276**
Patent and Delaware			[0.19]
Constant	-10.34**		-10.13**
	[0.21]		[0.21]
Observations	585162	585162	585162
Pseudo-R2	0.25	0.29	0.32

+p < .05, *p < .01, ** p < .001.

Table S6. Probability of growth outcome as a function of start-up characteristics (robustness tests). Outcome variable growth is equal to 1 if a firm achieves an IPO or acquisition within 6 years of founding. Model 1 runs our main regression only for corporations, model 2 runs our main regression only for Traded firms (firms where Local=0). Robust standard errors in brackets

	Corporations Only	**Traded Only**
	(1)	**(2)**
Eponymous	0.251*	0.189*
	[0.13]	[0.11]
Short name	1.497*	1.512*
	[0.20]	[0.20]
Local	0.409+	
	[0.16]	
Technology	1.975+	1.939+
	[0.53]	[0.53]
Corporation		6.029**
		[1.14]
Trademark	4.392**	5.490**
	[0.75]	[0.92]
Interactions		
Only Patent	28.42**	24.11**
	[8.20]	[6.68]
Only Delaware	43.78**	34.14**
	[6.98]	[5.21]
Patent and Delaware	229.0**	185.6**
	[44.3]	[34.4]
Constant	0.000225**	0.0000416**
	[0.000029]	[0.0000089]
Observations	366380	495911
Pseudo-R2	33%	32%

+ p<.05, * p<.01, ** p<.001

III. Entrepreneurial Quality Estimation

Our approach combines three interrelated insights. First, because the challenges to reach a growth outcome as a sole proprietorship are formidable, a practical requirement for any growth-oriented entrepreneur is business registration. For the purposes of this study, we focus on the state of California and observe the full population of state business registrants using publicly available records.

Second, it is possible to measure informative characteristics of each firm at or close to the time of registration and so to distinguish among business registrants in terms of their entrepreneurial quality. These characteristics include the names of the firm and its president or managers, the firm's location, the firm's organization type and local jurisdiction, and whether the firm seeks a patent or trademark.

Finally, although growth outcomes are observed with a lag, we can create a mapping between the set of meaningful growth outcomes and characteristics observable near the time of founding ("start-up characteristics"). Our primary model, presented in the third column of Table S1, is a logistic regression based on a training sample of California business registrants from 2001 to 2006. For a firm *i* in region *j* initially registered during cohort year *t*, we can measure a growth outcome $g_{i,j,t+s}$ that is realized within *s* years of founding as a function of start-up characteristics $X_{i,j,t}$:

$$\theta_{i,j,t} = 1000 \times P\left(g_{i,j,t+s} \middle| X_{i,j,t}\right)$$
$$\theta_{i,j,t} = 1000 \times f(\alpha + \beta X_{i,j,t})$$

We then calculate $\hat{\theta}_{i,j,t}$, our estimate of entrepreneurial quality equal to the growth probability of firm *i* located in location *j* and registered at time *t* (multiplied by 1000). We calculate $\hat{\theta}_{i,j,t}$ for all firms in our sample, including firms in the 30% "test" sample from 2001 to 2006 as well as the "prediction" sample from 2007 to 2011. Table S3 contains summary statistics on our training period.

Our primary findings regarding the geography of entrepreneurial quality presented in Figures 1 and 2 and Figs. S1 to S3, and Table S1 are based on the prediction sample from 2007 to 2011[3]. We calculate $\widehat{\Theta_{j,t}}$, the average entrepreneurial quality of start-ups in *j* founded at *t* (multiplied by 1000), at two different levels of geographic aggregation: ZIP Code (five-digit) and city. We use the 482 California cities listed in the 2010 Census Incorporated Places and Minor Civil Subdivisions. Our procedures accounts for 86% of all firms registered within the period; misspellings ("Los Amgeles," "Palo Alato"), use of neighborhood names rather than city names (e.g., "Sherman Oaks"), or firms in unincorporated areas are excluded from the city-level analysis.

Table S7 shows summary statistics for estimated entrepreneurial quality at three levels of aggregation: firm-level, ZIP Code–level, and city-level. The mean of entrepreneurial quality is less than one, indicating that there is less than a one in a thousand chance of achieving a growth outcome. The distribution is highly skewed, even at the level of ZIP Codes and cities: For firms

[3] The main regression used is presented in Table S1 with coefficients as incidence ratios, and Table S5 presents the actual regression coefficients. In table S6, we include robustness tests by running our model on the subsets of corporations and traded firms.

in cities within the top one percent of entrepreneurial quality, the probability of a growth outcome is nearly 10 times higher than the average firm in the population.

Finally, in Figure 1, we group the cities into six regions of interest: Silicon Valley, San Francisco, Sacramento, Los Angeles, San Diego, and Other cities. For Sacramento, Los Angeles, and San Diego, we define these regions according to the Census-defined Metropolitan Statistical Areas (MSA). To represent the difference between the Northern and Southern Bay Area, we define the San Francisco Area as the areas covered by San Francisco County, Alameda County, Contra Costa County, and Marin County. We define Silicon Valley as those cities in San Mateo and Santa Clara counties. Table S8 contains a list of regions and the counties that belong to each one in our definition.

One important concern in policy applications of this methodology, is that our measures might change incentives of firms, such that they try to "game" the result by select into high-quality measures they previously did not care about (e.g. changing its name form long to short). We note that this possibility, though real, is bounded by the incentives of the founders. For example, it is unlikely that a founder with no intention to grow would incur the significant yearly expense require to keep a registration in Delaware (which we estimate around $1000). Similarly, firms that signal in their name being a local business (e.g. "Taqueria") are unlikely to change their names in ways that affect their ability to attract customers. Finally, we also note that any effects from "gaming" would be short-lived because, as low quality firms select into a specific measure the correlation between such measure and growth—and therefore the weight our prediction model would assign to it—weakens.

Table S7.Summary statistics for average entrepreneurial quality of firms, ZIP Codes, and cities. Summary statistics for the predicted probability of growth in all firms, ZIP Codes, cities in the 2007–2011 time period. Numbers are multiplied by 1000 for readaility. Cities are those listed in the U.S. Census 2010 Incorporated Places and Minor Civil Subdivisions data set.

	Firms	**ZIP Codes**	**Cities**
All data			
N	754424	3243	482
Mean	0.0009	0.748	0.678
Median	0.0002	0.389	0.431
Standard deviation	0.0102	2.719	0.748
99^{th} percentile	0.0101	8.405	4.483
Top 1%			
Mean	0.0479	19.310	5.258
Median	0.0109	10.140	5.366
Standard deviation	0.0898	18.300	0.311

Table S8. Definition of regions.

County	Region
Los Angeles County	Los Angeles Area
Orange County	Los Angeles Area
Sacramento County	Sacramento Area
Placer County	Sacramento Area
Yolo County	Sacramento Area
Sutter County	Sacramento Area
Yuba County	Sacramento Area
Alameda County	San Francisco Area
Contra Costa County	San Francisco Area
San Francisco County	San Francisco Area
Marin County	San Francisco Area
San Diego County	San Diego Area
San Mateo County	Silicon Valley
Santa Clara County	Silicon Valley

IV. Ranking of Entrepreneurial Quality by City

Table S9. Ranking of entrepreneurial quality by city.

Rank	City	Quality	Rank	City	Quality
1	Menlo Park	4.64	51	Hayward	0.88
2	Mountain View	4.45	52	San Diego	0.86
3	Palo Alto	4.40	53	Carpinteria	0.85
4	Sunnyvale	3.86	54	Irvine	0.84
5	Redwood City	3.78	55	Santa Monica	0.83
6	East Palo Alto	3.37	56	Vista	0.82
7	Emeryville	3.25	57	Tiburon	0.82
8	Portola Valley	2.68	58	Costa Mesa	0.79
9	Grover Beach	2.42	59	Del Mar	0.78
10	San Mateo	2.33	60	Ross	0.76
11	South San Francisco	2.28	61	Hillsborough	0.76
12	Los Altos Hills	2.26	62	Novato	0.75
13	Los Altos	2.22	63	Orinda	0.73
14	Woodside	2.13	64	West Sacramento	0.70
15	Goleta	2.11	65	Belmont	0.70
16	Santa Clara	2.06	66	Woodland	0.68
17	Foster City	1.98	67	Culver City	0.68
18	Cupertino	1.87	68	Mill Valley	0.67
19	Scotts Valley	1.74	69	Dublin	0.67
20	San Francisco	1.65	70	Corning	0.66
21	Burlingame	1.61	71	San Rafael	0.66
22	Carmel-By-The-Sea	1.59	72	Santa Barbara	0.64
23	Trinidad	1.51	73	Azusa	0.63
24	Fremont	1.47	74	Camarillo	0.62
25	San Bruno	1.46	75	Santa Cruz	0.62
26	Larkspur	1.44	76	Malibu	0.62
27	Atherton	1.37	77	Oakland	0.61
28	Belvedere	1.36	78	Commerce	0.60
29	San Carlos	1.30	79	Rancho Cordova	0.60
30	Aliso Viejo	1.28	80	Clayton	0.60
31	Milpitas	1.23	81	Beverly Hills	0.60
32	Pleasanton	1.20	82	Tracy	0.59
33	San Anselmo	1.18	83	Alameda	0.59
34	El Segundo	1.13	84	Carson	0.59
35	Campbell	1.13	85	Millbrae	0.59
36	Sausalito	1.13	86	San Luis Obispo	0.59
37	Los Gatos	1.09	87	Pasadena	0.59
38	Seal Beach	1.08	88	Encinitas	0.58
39	Berkeley	1.07	89	San Clemente	0.58
40	Saratoga	1.07	90	Agoura Hills	0.57
41	Rolling Hills	1.06	91	Westlake Village	0.56
42	San Ramon	1.04	92	Villa Park	0.56
43	Solana Beach	1.02	93	Monte Sereno	0.56
44	Carlsbad	0.99	94	Sebastopol	0.55
45	San Jose	0.96	95	Oxnard	0.55
46	Hercules	0.95	96	Newport Beach	0.55
47	Morgan Hill	0.92	97	Rosemead	0.55
48	Corte Madera	0.92	98	Santa Fe Springs	0.54
49	Livermore	0.92	99	Laguna Beach	0.53
50	Newark	0.89	100	Dana Point	0.52

Rank	City	Quality	Rank	City	Quality
101	Sand City	0.52	151	Torrance	0.37
102	Sonoma	0.51	152	Thousand Oaks	0.37
103	La Palma	0.51	153	El Cerrito	0.37
104	Ventura	0.50	154	Walnut	0.37
105	Piedmont	0.50	155	Laguna Hills	0.37
106	Vacaville	0.50	156	Long Beach	0.37
107	Lafayette	0.50	157	Corona	0.36
108	Escondido	0.49	158	Union City	0.36
109	Moraga	0.48	159	Loma Linda	0.36
110	Banning	0.48	160	Laguna Niguel	0.36
111	Burbank	0.48	161	Coronado	0.36
112	Santa Ana	0.48	162	Glendale	0.36
113	Sonora	0.47	163	Anaheim	0.35
114	Manhattan Beach	0.46	164	San Juan Capistrano	0.35
115	Monterey	0.46	165	Ontario	0.35
116	Loomis	0.46	166	Folsom	0.35
117	Concord	0.46	167	Cypress	0.35
118	Norco	0.45	168	Grass Valley	0.35
119	Sierra Madre	0.45	169	Rancho Cucamonga	0.34
120	El Cajon	0.45	170	Gardena	0.34
121	La Quinta	0.45	171	Santa Rosa	0.34
122	Napa	0.45	172	Calabasas	0.34
123	Placentia	0.45	173	Duarte	0.34
124	Danville	0.44	174	Brea	0.34
125	Brisbane	0.44	175	Watsonville	0.33
126	Rocklin	0.44	176	Chico	0.33
127	Lake Forest	0.43	177	Sacramento	0.33
128	West Hollywood	0.43	178	Hawthorne	0.33
129	Lompoc	0.41	179	Madera	0.33
130	Davis	0.41	180	Martinez	0.33
131	Tustin	0.41	181	Buena Park	0.33
132	Walnut Creek	0.41	182	Healdsburg	0.32
133	Los Angeles	0.41	183	Huntington Beach	0.32
134	Roseville	0.40	184	Rohnert Park	0.32
135	Gridley	0.40	185	Hermosa Beach	0.32
136	Bell Gardens	0.40	186	Redding	0.31
137	Pleasant Hill	0.39	187	La Mirada	0.31
138	Rancho Santa Margarita	0.39	188	Rancho Mirage	0.31
139	Poway	0.39	189	Fountain Valley	0.31
140	Petaluma	0.39	190	Claremont	0.31
141	Irwindale	0.39	191	Oroville	0.31
142	Daly City	0.38	192	Tehachapi	0.31
143	Monterey Park	0.38	193	Windsor	0.30
144	Hollister	0.38	194	La Verne	0.30
145	Palos Verdes Estates	0.38	195	El Monte	0.30
146	Newman	0.38	196	Huntington Park	0.30
147	San Marcos	0.37	197	Half Moon Bay	0.30
148	Moorpark	0.37	198	Vernon	0.30
149	Capitola	0.37	199	Yorba Linda	0.30
150	Manteca	0.37	200	Covina	0.29

Rank	City	Quality	Rank	City	Quality
201	Cerritos	0.29	251	Palm Springs	0.22
202	Compton	0.29	252	Redlands	0.22
203	Coalinga	0.29	253	Grand Terrace	0.22
204	Chula Vista	0.29	254	Auburn	0.22
205	Indian Wells	0.29	255	Diamond Bar	0.22
206	Simi Valley	0.29	256	Glendora	0.22
207	Temecula	0.29	257	San Gabriel	0.22
208	South El Monte	0.29	258	Colfax	0.22
209	Orange	0.29	259	Albany	0.22
210	San Leandro	0.28	260	Paramount	0.22
211	Mission Viejo	0.27	261	Los Alamitos	0.22
212	San Juan Bautista	0.27	262	Temple City	0.22
213	Placerville	0.27	263	Bakersfield	0.22
214	Palm Desert	0.27	264	Lakewood	0.22
215	Alhambra	0.26	265	Rolling Hills Estates	0.22
216	Nevada City	0.26	266	Eureka	0.22
217	Fontana	0.26	267	Atwater	0.22
218	La Canada Flintridge	0.26	268	Gilroy	0.21
219	Rio Vista	0.26	269	Hemet	0.21
220	Port Hueneme	0.25	270	Arcadia	0.21
221	Brentwood	0.25	271	Fresno	0.21
222	San Fernando	0.25	272	Lynwood	0.21
223	Redondo Beach	0.25	273	La Puente	0.21
224	San Dimas	0.25	274	Chino Hills	0.21
225	Oceanside	0.25	275	Perris	0.21
226	Pomona	0.25	276	Upland	0.21
227	Solvang	0.25	277	Santa Maria	0.21
228	Fort Bragg	0.25	278	Fairfax	0.21
229	Riverside	0.25	279	Yuba City	0.21
230	Richmond	0.25	280	Murrieta	0.21
231	San Bernardino	0.25	281	Baldwin Park	0.21
232	Kerman	0.25	282	Signal Hill	0.21
233	Dixon	0.24	283	Santa Clarita	0.21
234	Monrovia	0.24	284	Del Rey Oaks	0.21
235	Montebello	0.24	285	Lancaster	0.21
236	Maywood	0.24	286	Stanton	0.20
237	Chino	0.24	287	Exeter	0.20
238	Fullerton	0.24	288	Waterford	0.20
239	Yountville	0.24	289	Wildomar	0.20
240	Visalia	0.24	290	South Lake Tahoe	0.20
241	Westminster	0.24	291	Brawley	0.20
242	Garden Grove	0.23	292	Apple Valley	0.20
243	Colton	0.23	293	Big Bear Lake	0.20
244	Inglewood	0.23	294	Lemoore	0.20
245	Industry	0.23	295	Whittier	0.20
246	South Pasadena	0.23	296	Anderson	0.20
247	San Marino	0.23	297	Pico Rivera	0.20
248	Canyon Lake	0.23	298	Artesia	0.20
249	Imperial Beach	0.23	299	Corcoran	0.20
250	Desert Hot Springs	0.23	300	Beaumont	0.19

Rank	City	Quality	Rank	City	Quality
301	Pacifica	0.19	351	Ripon	0.17
302	Montclair	0.19	352	California City	0.16
303	Patterson	0.19	353	Ceres	0.16
304	La Habra	0.19	354	Sutter Creek	0.16
305	San Joaquin	0.19	355	Lomita	0.16
306	West Covina	0.19	356	Palmdale	0.16
307	Livingston	0.19	357	Plymouth	0.16
308	Cudahy	0.19	358	Vallejo	0.16
309	Hanford	0.19	359	Fairfield	0.16
310	Benicia	0.19	360	Mammoth Lakes	0.16
311	Turlock	0.19	361	Truckee	0.16
312	Lincoln	0.19	362	Santa Paula	0.16
313	Stockton	0.18	363	Pacific Grove	0.16
314	La Habra Heights	0.18	364	Escalon	0.16
315	Ojai	0.18	365	El Centro	0.16
316	Ukiah	0.18	366	Lake Elsinore	0.16
317	Kingsburg	0.18	367	Dunsmuir	0.16
318	Laguna Woods	0.18	368	Mount Shasta	0.16
319	Arroyo Grande	0.18	369	Elk Grove	0.16
320	Yucca Valley	0.18	370	Jurupa Valley	0.16
321	Calexico	0.18	371	Greenfield	0.16
322	National City	0.18	372	Clovis	0.16
323	Lawndale	0.18	373	Lathrop	0.15
324	Atascadero	0.18	374	Pismo Beach	0.15
325	South Gate	0.18	375	Clearlake	0.15
326	Cathedral City	0.18	376	Wasco	0.15
327	Norwalk	0.18	377	Weed	0.15
328	Modesto	0.18	378	Selma	0.15
329	Eastvale	0.17	379	Shasta Lake	0.15
330	Hesperia	0.17	380	Adelanto	0.15
331	Farmersville	0.17	381	Paso Robles	0.15
332	Antioch	0.17	382	Porterville	0.15
333	Cotati	0.17	383	Lemon Grove	0.15
334	Santee	0.17	384	Arvin	0.15
335	Bell	0.17	385	Avenal	0.15
336	Menifee	0.17	386	San Pablo	0.15
337	St. Helena	0.17	387	Blue Lake	0.15
338	Hawaiian Gardens	0.17	388	Etna	0.15
339	Downey	0.17	389	Sanger	0.15
340	Salinas	0.17	390	Jackson	0.15
341	Bellflower	0.17	391	Lodi	0.14
342	Victorville	0.17	392	Fillmore	0.14
343	Rancho Palos Verdes	0.17	393	Indio	0.14
344	Yucaipa	0.17	394	Dinuba	0.14
345	Rialto	0.17	395	Avalon	0.14
346	Los Banos	0.17	396	Highland	0.14
347	La Mesa	0.17	397	Willits	0.14
348	San Jacinto	0.17	398	Calimesa	0.14
349	Pittsburg	0.17	399	Seaside	0.14
350	Moreno Valley	0.17	400	Fowler	0.14

Rank	City	Quality
401	Merced	0.14
402	Crescent City	0.14
403	Coachella	0.14
404	Galt	0.14
405	Marina	0.14
406	Live Oak	0.13
407	Buellton	0.13
408	Ione	0.13
409	King City	0.13
410	Tehama	0.13
411	Oakley	0.13
412	Suisun City	0.13
413	Ridgecrest	0.13
414	Citrus Heights	0.13
415	American Canyon	0.13
416	Tulare	0.13
417	Lakeport	0.13
418	Bradbury	0.13
419	Woodlake	0.13
420	Morro Bay	0.13
421	Susanville	0.12
422	Lindsay	0.12
423	Chowchilla	0.12
424	Calistoga	0.12
425	Delano	0.12
426	Paradise	0.12
427	Arcata	0.12
428	Montague	0.12
429	Alturas	0.12
430	Imperial	0.12
431	Reedley	0.12
432	Barstow	0.12
433	Orange Cove	0.12
434	Westmorland	0.12
435	Holtville	0.12
436	Fortuna	0.12
437	Gonzales	0.12
438	Marysville	0.11
439	Point Arena	0.11
440	Red Bluff	0.11
441	Riverbank	0.11
442	Guadalupe	0.11
443	Colma	0.11
444	Winters	0.11
445	Taft	0.11
446	Williams	0.11
447	Hughson	0.11
448	Pinole	0.11
449	Bishop	0.11
450	Mcfarland	0.11
451	Oakdale	0.11
452	Blythe	0.10
453	Soledad	0.10
454	Cloverdale	0.10
455	Portola	0.10
456	Parlier	0.10
457	Ferndale	0.10
458	Orland	0.10
459	Twentynine Palms	0.10
460	Hidden Hills	0.10
461	Shafter	0.10
462	Gustine	0.10
463	Rio Dell	0.10
464	Tulelake	0.09
465	Yreka	0.09
466	Biggs	0.09
467	Wheatland	0.09
468	Fort Jones	0.09
469	Huron	0.09
470	Angels Camp	0.08
471	Mendota	0.08
472	Dos Palos	0.08
473	Amador City	0.08
474	Maricopa	0.08
475	Colusa	0.08
476	Firebaugh	0.08
477	Willows	0.07
478	Isleton	0.07
479	Needles	0.07
480	Calipatria	0.07
481	Dorris	0.07
482	Loyalton	0.05

Table S10. Top local words

ABSTRACT	CHIROPRACTIC	FLORIST	MARY'S	REALTY	UNISEX
AC	CHR	FLOWER	MASONIC	REHAB	UPHOLSTERY
ACCOUNTING	CHRIST	FLOWERS	MASONRY	REHABILITATION	UROLOGY
ACUPUNCTURE	CHRISTIAN	FOOT	MASSAGE	REMODELING	USED
ADVENTIST	CHURCH	FUNERAL	MEDICINE	REPAIR	VETERINARY
ALLERGY	CLEANERS	GARAGE	MENTAL	REPAIRS	WASH
ALTERATIONS	CLEANING	GIFT	METHODIST	RESTAURANT	WELLNESS
AMBULANCE	CLINIC	GOD	MEXICAN	ROOFING	WINDOW
AME	COFFEE	GOSPEL	MIDDLE	ROOTER	WINDOWS
ANIMAL	COLLISION	GOURMET	MINISTRIES	RSTRNT	WITNESSES
ANKLE	CONCRETE	GRILL	MINISTRY	SALON	WOK
ANTIQUES	COND	GRILLE	MIRROR	SALOON	WOMEN'S
APARTMENT	CONDO	GROCERY	MISSIONARY	SCHL	WORSHIP
APARTMENTS	CONDOMINIUM	GROOMING	MONTESSORI	SCHOOL	WRECKER
APOSTOLIC	CONDOMINIUMS	GUTTER	MOTORS	SCHOOLS	ZION
APPLIANCE	CONGREGATIONAL	GUTTERS	MTHDST	SEAMLESS	
APPLIANCES	CONSIGNMENT	HAIR	MUFFLER	SENIOR	
APTS	CONTR	HANDYMAN	NAIL	SEPTIC	
ASPHALT	CONTRACTING	HEAD	NAILS	SHEAR	
ASSISTED	CONVENIENCE	HEALING	NAZARENE	SHOE	
ATTORNEY	COOLING	HEALTH	NURSING	SHOES	
AUTO	COSMETIC	HEARING	NUTRITION	SIDING	
AUTOMOTIVE	COUNSELING	HEATING	OFFICES	SKIN	
BAKERY	COURSE	HOLY	ORAL	SMOG	
BAPT	COVERING	HOSP	ORIENTAL	SPA	
BAPTIST	CPA	HOSPICE	ORTHODONTICS	SPINE	
BAR	CU	HOSPITAL	ORTHOPEDIC	SPIRITS	
BARBER	CUISINE	HTG	OUR	ST	
BBQ	CUTS	IGLESIA	OUTREACH	START	
BEER	DAY	IMPROVEMENT	OVERHEAD	STEAK	
BEGINNINGS	DAYCARE	IMPROVEMENTS	OWNERS	STORES	
BEHAVIORAL	DELI	INSPECTION	PAIN	STYLES	
BETHEL	DENTAL	INSPECTIONS	PAINTING	STYLING	
BIBLE	DENTISTRY	INSULATION	PAUL'S	SUPERMARKET	
BISTRO	DERMATOLOGY	INSURANCE	PAVING	SURGERY	
BODY	DETAIL	INTERNAL	PEDIATRIC	SUSHI	
BOOKKEEPING	DETAILING	ITALIAN	PEDIATRICS	SWIMMING	
BOUTIQUE	DINER	JAPANESE	PENTECOSTAL	TABERNACLE	
BRAKE	DONUTS	JEHOVAH'S	PEST	TACO	
BRIDAL	DOORS	JESUS	PHARMACY	TAILOR	
BUFFET	DRAIN	JEWELERS	PHOTOGRAPHY	TAN	
BUILDERS	DRUG	JOHN'S	PHYSICAL	TANNING	
CAFE	DRY	KARATE	PHYSICIANS	TAQUERIA	
CALVARY	DRYWALL	LANES	PIANO	TATTOO	
CARE	ELEMENTARY	LAUNDROMAT	PIZZA	TAVERN	
CARPENTRY	EPISCOPAL	LAUNDRY	PIZZERIA	TAX	
CARPET	ESTATE	LAW	PLASTERING	TEMPLE	
CARS	EVANGELICAL	LAWN	PLUMBING	TERMITE	
CATERING	EXCAVATING	LDS	PRACTICE	TERRACE	
CATHOLIC	EXCAVATION	LIQUOR	PRESBYTERIAN	THAI	
CEMETERY	EXTERMINATING	LIQUORS	PRESCHOOL	THERAPEUTIC	
CHAPEL	FAITH	LOCK	PROPERTIES	THERAPY	
CHEVROLET	FAMILY	LOCKSMITH	PROPERTY	THRIFT	
CHICKEN	FELLOWSHIP	LOUNGE	PSYCHOLOGICAL	TILE	
CHILD	FIRM	LUBE	PUB	TIRE	
CHILDCARE	FITNESS	LUTHERAN	PUMPING	TIRES	
CHIMNEY	FLOOR	MALL	RADIATOR	TOWING	
CHINA	FLOORING	MART	REAL	TRANSMISSION	
CHINESE	FLOORS	MARTIAL	REALTORS	TRANSMISSIONS	

Table S11. Top high-technology words

AERO
AEROFLEX
AEROJET-GENERAL
AERONAUTICS
AEROSPACE
AEROSTRUCTURES
AEROSTRUCTURES-VOUGHT
AGILENT
AIRCRAFT
AIRFOILS
ALLIANT
ALPHARMA
ALPS
ALTERA
AMETEK
AMPHENOL
ANALOG
ANSYS
APP
ASTRA
ATK
ATMEL
AVID
AVIONICS
AXCELIS
B/E
BAE
BECKMAN
BIOLOGICAL
BIOLOGICALS
BIOPHARMACEUTICALS
BIOSCIENCE
BIOSCIENCES
BIOTECHNOLOGY
BIOTHERAPEUTICS
BOEING
BOMBARDIER
BOTTOMLINE
BRISTOL-MYERS
BRK
BROADCOM
BRUKER
C&D
C4
CAPACITORS
CELGENE
CESSNA
CHROMALLOY
CIRCUIT
CIRCUITS
CIRRUS
CISCO
COIL
COILS
COMBUSTION
COMPONENT
COMPONENTS
CONDUCTOR
CONEXANT
CONNECTOR
CONNECTORS
CONTROLS
COULTER
CUBIC
CURTISS-WRIGHT
DEVICES
DIODES
DUCOMMUN
DYNETICS
ELECTRO
ELECTRON
ELECTRONICS
EXTERRAN
FLEXTRONICS
FLIR
FLUKE
FREESCALE
FUJITSU
GARMIN
GENZYME
GKN
GLAXOSMITHKLINE
GLIDER
GOODRICH
GRUMMAN
HARLAND
HELICOPTER
HEWLETT-PACKARD
HITACHI
HONEYWELL
HOSPIRA
HYNIX
IBM
IDEC
IMMUNE
INFRARED
INSTRS
INSTRUMENT
INSTRUMENTATION
INSTRUMENTS
INTEL
INTERCONNECT
INTERSIL
INTGRD
INVENSYS
ITT
JABIL
KAMAN
KYOCERA
L-3
LATTICE
LAUNCH
LINEAR
LOCKHEED
LSI
M/A-COM
MAGNETIC
MAGNETICS
MARVELL
MEASUREMENTS
MEASURING
MERCK
METROLOGY
METTLER-TOLEDO
MICRO
MICROCHIP
MICROELECTRONICS
MICRON
MICROS
MICROSEMI
MICROSOFT
MICROSYSTEMS
MICROWAVE
MKS
MOLECULAR
MOLEX
MOOG
MWARE
NANO
NAVIGATION
NDT
NIKON
NORTHROP
NOVARTIS
NOVELLUS
NUANCE
NXP
OEM
OPTO
OPTOELECTRONICS
ORACLE
ORBITAL
PACKET
PANASONIC
PARKER-HANNIFIN
PASSUR
PCB
PCC
PERIPHERALS
PERKIN
PFIZER
PHARMA
PHARMACEUTICAL
PHARMACEUTICALS
PHOTOMASKS
PHOTRONICS
PHRMCTCLS
PLEXUS
PMC-SIERRA
PRINTED
PROBE
PROPELLER
PROPULSION
PTC
PTI
QUADRA
QUARTZ
QUINT
RAYTHEON
RECTIFIER
RELIV
RESISTOR
RF
ROLLS-ROYCE
ROSEMOUNT
SANMINA-SCI
SANYO
SATCOM
SCIENTIFIC
SEAGATE
SEMI
SEMICONDUCTOR
SEMICONDUCTORS
SENSING
SENSOR
SENSORS
SENSUS
SHIELDING
SIEMENS
SIKORSKY
SILICON
SIMCO
SIMULATION
SMSC
SOARING
SPACECRAFT
SPEAKER
SQUIBB
STATIC
SUNDSTRAND
SYMANTEC
SYNOPSYS
TDK
TECHSYSTEMS
TECT
TEKTRONIX
TELEDYNE
TEVA
TEXTRON
THERAPEUTICS
THERM-O-DISC
THERMO
TIBCO
TOPPAN
TOSHIBA
TRANSFORMER
TRIUMPH
TRONICS
TRX
TTI
TTM
TURBINE
UFC
VACCINES
VARIAN
VEECO
VIBRATION
VISHAY
WAFER
ZENECA
ZODIAC

Cited Works

1. California Corporations Code (Thomson West, 2013); http://www.leginfo.ca.gov/.html/corp_table_of_contents.html.
2. J. Guzman, S. Stern, *Nowcasting and Placecasting Entrepreneurial Quality and Performance* [National Bureau of Economic Research (NBER), Cambridge, MA, 2014]; http://www.nber.org/chapters/c13493.pdf.
3. Thomson Reuters's SDC Platinum is a commonly used database of financial information. More details are available at http://thomsonreuters.com/sdc-platinum/.
4. B. Barnes, N. Harp, D. Oler, "Evaluating the SDC mergers and acquisitions database" (SSRN Working paper 2201743, SSRN, Rochester, NY, 2014); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201743.
5. R. Levine, Y. Rubinstein, "Smart and illicit: Who becomes an entrepreneur and does it pay?" (NBER Working Paper 19276, NBER, Cambridge, MA, 2013); http://www.nber.org/papers/w19276.
6. S. Belenzon, A. Chatterji, B. Daley, "Eponymous entrepreneurs" (Working paper, https://faculty.fuqua.duke.edu/~sb135/bio/BCD_EE.pdf.
7. M. Delgado, M. Porter, S. Stern, "Defining clusters in related industries" (NBER Working paper 20375, NBER, Cambridge, MA, 2014); http://www.nber.org/papers/w20375
8. ReferenceUSA, Business Historical Data [2010 yearly snapshot retrieved through MIT Libraries].
9. N. Balasubramanian, J. Sivadasan, "NBER Patent Data-BR Bridge: User guide and technical documentation" (SSRN Working paper 1695013, SSRN, Rochester, NY, 2009); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1695013.
10. W. R. Kerr, Shihe Fu, "The Survey of Industrial R&D--Patent Database Link Project." *J. Technol. Transf.* **33**, no. 2 (2008).
11. V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals." *Doklady Akad. Nauk SSSR* **163**(4): 845–848 (1965).
12. S. Graham, G. Hancock, A. Marco, A. F. Myers, "The USPTO case files data set: Descriptions, lessons and insights" (SSRN Working Paper 2188621, Rochester, NY, 2013); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2188621.