Appendix A: Description of the Application Data Tab Release

A.1 Data Files Included in this Release

This data release consists of three data files, which together provide all of the information that a user would be able to glean from the "Application Data" tab on the PTO's Public PAIR website. The first data file is called **application_data** and it includes bibliographic information on each patent application included in PatEx as of December 31, 2014. There are 9,231,170 observations in this data file, with each observation representing a unique patent application. The variables included in this file are described in more detail below. They provide information on such fields as application number, filing date, application type, identity of the examiner, group art unit of the examiner, U.S. classification and sub-classification of the underlying invention, current application status, and patent number if the application had been issued as a patent, among others. See Table A-1 for a list and brief description of all variables included in the **application_data** file.

The second data file is called **all_inventors** and includes the names and locations of all inventors listed on the front page of each application in PatEx, if such information exists for an application. There are 21,617,363 observations in this data file. Each observation represents an application/inventor pair and there are 7,842,637 unique patent applications represented in these data. There can be multiple observations for a given application. For each application/inventor pair, the file includes information on the name and location of the inventor, as well as a variable, *inventor_rank*, which provides information on whether the inventor is the first-named inventor. The rank for the first-named inventor is 1, the rank for the second-named inventor is 2, and so on. See Table A-2 for a list and brief description of all variables included in the **all_inventors** file.

The third, and final data file is called **status_codes**. This data file includes descriptions of the status codes used to populate the *appl_status_code* variable in the **application_data** file. The descriptions can be linked to **application_data** using this variable. There are 225 unique application status codes represented in this file. See Table A-3 for a list and brief description of the two variables included in this file.

A.2 Variables Included in application_data

Application Number

Each application received by the PTO is given a unique application number. The number is used to keep track of the application while it is being processed and examined. For the user, the most important use of the application number is as a key variable. For instance, one can link data from the **all_inventors** data set to the applications provided in the **application_data** data set by using this variable (*application_number*).

The application number is comprised of two parts. For all applications that were not filed under the patent cooperation treaty (PCT), the first two digits indicate the application's series number. For the most part, the series number gives a rough indication of the order in which applications were received by the PTO. For example, Series 6 applications generally were received before Series 7 applications, which were generally received before Series 8 applications, etc. Roughly 77 percent of the applications in the 2014 PAIR data release are from series 6 through 14. Only 1 percent of the applications present in the data are from prior series (see Table A-4).

The remaining 22 percent of applications are from special series (i.e. series that do not include regular nonprovisional utility or plant patent applications). Since the early 1990s, all applications for design patents have been identified using series 29 (4 percent of the total sample of applications). Provisional applications have been identified using series 60, 61, and 62 (9 percent of all applications). Reexaminations of patent applications have been given series numbers 90, 95, and 96 (less than half of 1 percent of all applications). Finally, PCT applications can be identified as those with applications that start with the three character string "PCT." These PCT applications account for nearly 9 percent of all applications included in the PAIR data.

Filing Date

For most applications, the filing date is the date on which PTO received the application.[63] For PCT applications, the filing date is the date of PTO's receipt of 35 U.S.C. 371c requirements. More than 99 percent of the applications in **application_data** have a filing date of 1910 or later. However, as we see in Sections 2.1 and 3.2 of the main report, there is very poor coverage for applications that were received prior to 1981 and limited coverage for applications received between the late 1970s and the year 2000 due to the fact that there was no pre-grant publication of applications filed prior to November 29, 2000.

The filing date variable (*filing_date*) is formatted as a numeric variable which is equal to the difference between the filing date and the first day of January 1960. For instance, if an application was received on 10 January 1960, then the date variable would be equal to 9. For dates prior to 1 January 1960, the date variable takes on negative values. In the Stata version of the data set, the %td display format is embedded, so that the dates display with the following format: ddmmmyyyy. For example, when *filing_date* is equal to 12,500, it displays in Stata as "23mar1994."

Application Type

We provide two different variables to identify application type. The first, *invention_subject_matter*, identifies the subject matter of the invention, whether it is a utility, design, or plant patent application. Roughly 85 percent of all the records in the PAIR file have a known value for this variable.[64] The second variable, *application_type*, identifies the type of application as either a regular nonprovisional, provisional, re-issue, reexamination, or PCT application. All but three of the more than 9 million records have a non-missing value for this variable.

In Table A-5 we provide a cross-tabulation of these two variables. The results indicate that almost all of the regular nonprovisional applications (99.95 percent) have a non-missing value for the subject matter variable. The same is true of applications for reexamination and re-issue. At the other end of the spectrum, the subject matter variable is never populated for PCT applications. For provisional applications, the subject matter variable was populated for less than 3 percent of such applications prior to 2006. Since 2006, the subject matter variable has been populated more often and, in recent years, it has been populated more than 90 percent of the time (see Table A-6).

Examiner Identifiers

We provide researchers with two ways to identify the examiner of record. First we provide the first, middle, and last names of the examiner as three separate variables. This is the information that is provided

[63] There are some exceptions to this. See http://www.uspto.gov/patents-getting-started/patent-basics/types-patent-applications/nonprovisional-utility-patent for more details.
[64] 99.7 percent of the records without a valid value for this variable are either PCT or provisional applications.

on the Public PAIR website. Second, we provide a numeric examiner identifier, *examiner_id*, so that analysts will be better able to group applications and issued patents by examiner. This can be particularly useful in cases where analysts want to control for the effect that examiners have on several prosecutorial measures such as time to disposal, type of disposal, likelihood of appeal, and so on.

It should be noted that, for applications pending on January 24, 2015, the examiner of record was the examiner assigned to the application as of that date. For disposed applications, the examiner is the examiner who was assigned to the application at the time of disposal.

Examiner Art Unit

The variable *examiner_art_unit* is a string variable indicating the group art unit to which the examiner of record was assigned as of the last office action recorded for the application in question. Group art units are designated as four digit numbers. The first two digits indicate the technology center (TC) to which the group art unit is assigned.[65] The designations for the technology centers have changed over the years, but currently there are eight such technology centers for examining regular utility applications.

- 1600 – Biotechnology
- 1700 – Chemical and Materials Engineering
- 2100 – Computer Architecture, Software, and Information Security
- 2400 – Computer Networks, Multiplex Communication, Video Distribution and Security
- 2600 – Communications
- 2800 – Semiconductors, Electrical and Optical Systems and Components
- 3600 – Transportation, Construction, Electronic Commerce, Agriculture, National Security and License & Review
- 3700 – Mechanical Engineering, Manufacturing, Products

There are a few instances where the group art unit variable is populated with identifiers for USPTO business units, in which patent examination does not take place. This is due to the fact that examiners sometimes switch to other business units (or other art units) between the time that an application is allowed and the time that it is issued as a patent. Because the *examiner_art_unit* variable for issued patents is based on the business unit to which the examiner of record was assigned at time of issue, it can occasionally reflect an art unit which was not the one to which the application was assigned for examination.

Classification Codes

When the PTO processes new patent applications, they assign the application into one general technology class and into one or more subclasses. Classification of new applications assists in (1) the assignment of the applications to the most relevant group art units and (2) the searches for relevant prior art during patent examination. Each class and subclass is identified by a code. The class and subclass codes for each application are provided in *uspc_class* and *uspc_subclass*.[66] As an example, consider the patented case

[65] A full listing of current group art units, along with contact information is available on the public PTO website at http://www.uspto.gov/about/contacts/phone_directory/pat_tech/. In order to find the current group art units within each technology center, click on the technology center number.

[66] For information regarding what the various class and subclass codes mean, we direct the reader to the following web page: http://www.uspto.gov/web/patents/classification/selectnumwithtitle.htm.

(patent number 8,000,000) illustrated in Exhibit 1. Here, the class code is 607 and the subclass code is 54. This class/subclass pair is defined as follows.

- Class 607 - Surgery: light, thermal, and electrical application
    - Subclass 54 - Producing visual effects by stimulation

Note that the codes included in this PatX data release are US Patent Classification codes rather the newer Cooperative Patent Classification codes which are being adopted by PTO.

Confirmation, Customer, and Attorney Docket Numbers

The confirmation number (*confirm_number*) is a four-digit number that the PTO uses to ensure that any papers filed by the applicant (or applicant's attorney) are assigned to the right file. This is not a unique identifier and should not be used as such. It is included in the PatEx data because it is available on Public PAIR.

The customer number (*customer_number*) can be used with the file COORESPONDENCE_ADDRESS to identify the entity that is listed as the correspondent for all application-related matters.[67] It is usually the law firm representing the inventor or the legal department of the firm to which the application is assigned. Roughly 60 percent of the applications available in **application_data** have a legitimate value for this variable.

When an application is filed by a patent attorney, there is usually an internal tracking number assigned by the law firm for ease of reference. That is the docket number (*atty_docket_number*) that appears in Public PAIR. The PTO takes this information from the transmittal form or Application Data Sheet (ADS) filed with the application. More than 95 percent of the applications in PAIR filed since 1998 have a value for this variable.

Application Status

The application status variable (*appl_status_code*) is coded as a one- to three-digit integer value which can be deciphered using the **status_codes** file. The variable indicates what the status of the application was as of December 31, 2014. Table A-7 lists the 10 most common application status codes. By far, the most common application status code (47 percent of all cases) is "150" which indicates that the application's current status is that of a patented case. The other two most common codes are "161" (Abandoned – Failure to Respond to an Office Action) and "250" (Patent Expired Due to Nonpayment of Maintenance Fees Under 37 CFR 1.362). These codes are found for 12 and 10 percent of all cases, respectively.

The *appl_status_date* variable indicates the date that the application entered its most recent status (or status as of the end of 2014). The formatting is the same as for the filing date variable in that it is a numeric variable which is equal to the difference between the status date and the first day of January 1960. Because several older patented cases were added to the underlying data system in September 2001, many of these applications have a most recent status date of sometime during the week of September 19, 2001, even though they had been issued as patents far earlier. The problem is that, using the status date variable, it appears that PTO issued over 750,000 patents in September of 2001. Therefore, for these

[67] The **correspondence_address** file is described in Appendix F.

cases, we recommend using the patent issue date in lieu of the status date variable. The *application_status* variable is set equal to "150" for these cases.

File Location

The *file_location* variable in PatEx provides the current site of the official file. There are several locations possible, but 97 percent of all applications are either stored electronically or at the file repository (Franconia), which is in Springfield, Virginia.[68] In this public release file, the location variable takes one of the following four distinct values, reflecting the fact that most files are located at one of these two locations.

- ELECTRONIC
- FRANCONIA
- MISSING
- OTHER

The *file_location_date* variable indicates the date on which the file first arrived at its present location. The formatting is the same as for the other date variables in that it is a numeric variable which is equal to the difference (in days) between the status date and the first day of January 1960.

Pre-Grant Publication Information

Since 2001, most applications to the PTO have been published prior to grant within 18 months of filing. Exceptions to this rule are cases in which (1) applicants have requested that an application not be published prior to grant, (2) an application is deemed un-publishable for national security reasons, or (3) an application has been abandoned prior to the end of the 18-month period after filing.[69] For applications that have been made public by PTO, the following two variables are provided.

- *earliest_pgpub_number* – This variable provides the earliest pre-grant publication number assigned by the PTO for the application.
- *earliest_pgpub_date* – The variable provides the earliest pre-grant publication date for the application. The formatting is the same as for the other date variables in that it is a numeric variable which is equal to the difference (in days) between the status date and the first day of January 1960.

PCT applications become publicly viewable because they are published internationally by the World Intellectual Property Office (WIPO). For these applications a WIPO publication number is available. In this data release we provide a variable (*wipo_pub_number*) pertaining to WIPO publication.

Disposal variables

An application can be "disposed" in one of two ways. First, all or some of the claims made in the application are allowed and ultimately issued as a patent. Second, at some point during the examination

[68] When the term "Electronic" appears as the "location" of an application or patent, the official file is an electronic image file as described in the Official Gazette Notice 1271 OG 100, published June 17, 2003.

[69] Applicants can usually only request non-publication of an application if they are not pursuing patent rights for the same invention in other national jurisdictions. See Section 2.3 on selection of applications into Public PAIR for more details.

process the applicant may abandon the application. We provide the following variables to describe each application's disposal status as of December 31, 2014.

- *patent_number* – This variable is populated for applications that resulted in an issued patent. Patent numbers are assigned sequentially based on date of issue.
- *patent_issue_date* – This variable is generally populated for applications that resulted in an issued patent. It indicates the day on which the patent was issued, and should not be confused with the date on which the underlying claims were allowed. There can be a lag of several weeks between allowance of claims and patent issue. The formatting is the same as for the other date variables in that it is a numeric variable which is equal to the difference (in days) between the status date and the first day of January 1960.
- *abandon_date* – This variable is populated for applications that have been abandoned. The formatting is the same as for the other date variables.
- *disposal_type* – This variable is constructed using the other three variables listed above. It is set equal to "ISS" (issued) if either *patent_number* or *patent_issue_date* is populated with a valid value. It is set equal to "ABN" (abandoned) if neither *patent_number* nor *patent_issue_date* is populated with a valid value AND *abandon_date* is populated with a valid value. It is set equal to "PEND" (pending) if none of the three variables (*patent_number*, *patent_issue_date*, and *abandon_date*) is populated with a valid value. Finally, for all PCT and provisional applications, the variable is set equal to "N/A" (not applicable), as these applications are truly more place-holders than anything else, and are never issued as patents nor abandoned in the traditional sense. See Table A-8 for a breakout of disposal status by year of application (starting in 1980) for regular nonprovisional utility applications.

Other variables

There are a few other variables included in **application_data**. First, *invention_title* is a string variable which provides the title of the invention, as would be found on the Public PAIR website. The *small_entity_indicator* variable is an indicator variable set equal to 1 if the applicant qualifies as a small entity. A small entity is typically either an individual inventor, a collaboration of individual inventors, a nonprofit organization, or a company with fewer than 500 employees. Small entity status typically entitles the applicant to a 50 percent discount on most fee payments to PTO.

Finally, the variable *aia_first_to_file* indicates an application that is to be judged under the first-inventor-to-file rules as laid out in the America Invents Act (AIA). This provision did not come into effect until March, 2013.

A.3 Variables Included in all_inventors

The **all_inventors** file includes information on the names and locations of the inventors for most of the applications included in the **application_data** file. The variable, *application_number*, can be used to link information regarding the inventors to specific applications. The data regarding the inventor names are fairly straight-forward. The *inventor_rank* variable indicates the order of the inventors as listed on the original application and can be used to determine who the first-named inventor is.

Determining the location (country or U.S. state) of each inventor can be done using the *inventor_country_code* and *inventor_region_code* variables. The *inventor_country_code* variable is coded

using the ISO 3166 format.[70] We also include a variable, *inventor_country_name*, which can also be used to decipher the country codes. For domestic (US) applications, the *inventor_region_code* variable can be used to determine state of residence. States are coded using standard US Postal Service 2-digit state abbreviations. The data here are not perfect as there are apparent coding errors, but we have chosen not to clean these data so as to let researchers use their own chosen algorithms for cleaning. We provide a couple of examples of how the inventor location variables can be used.

1. Suppose we want to create a subset of all applications where the first-named inventor is from Japan. We would use the **all_inventors** file and keep all records where *inventor_rank* equals 1 and where *inventor_country_code* equals "JP", which is the ISO 3166 code for Japan. We could then link the resulting file with the **application_data** file for further analysis. In Table A-9 we list the countries with the most mentions of first-named inventor in PatEx.
2. Suppose we want to create a subset of all applications where any inventor is from California. Here we would use the **all_inventors** file and keep all records where *inventor_country_code* equals "US" and where *inventor_region_code* equals "CA". We would probably also want to look at all of the *inventor_region_code* values to be certain that there were not others that could indicate California as state of residence, but using "CA" would capture almost all of the cases. Please note that in this case there might well be multiple records for a single application. Again, using *application_number*, the resulting file could be linked with the **application_data** file for further analysis. In Table A-10 we present the number of first-named inventor mentions by state.

A.4 Variables included in status_codes

The **status_codes** data set includes descriptions of what the various values of the *appl_status_code* variable mean. They can be linked to the **application_data** data set through this variable.

[70] For more information on ISO 3166 country codes see http://www.iso.org/iso/country_codes.htm.

Table A-1: List of variables included in application_data

Variable Name	Description	Type	Formatting
application_number	Application Number	str14	%14s
filing_date	Filing or 371 (c) Date	float	%td
invention_subject_matter	Invention Subject Matter	str3	%-3s
application_type	Application Type	str7	%-7s
examiner_name_last	Examiner's Family Name	str17	%-20s
examiner_name_first	Examiner's Given Name	str12	%-20s
examiner_name_middle	Examiner's Middle Name	str12	%-20s
examiner_id	Unique Examiner Identifier	str5	%9s
examiner_art_unit	Group Art Unit	str6	%-6s
uspc_class	Invention U.S. Classification	str3	%-3s
uspc_subclass	Invention U.S. Subclassification	str6	%-6s
confirm_number	Confirmation Number	int	%12.0f
customer_number	Customer number	str6	%-6s
atty_docket_number	Attorney Docket Number	str25	%-20s
appl_status_code	Application Status Code	int	%8.0f
appl_status_date	Status Date	float	%td
file_location	Location (where the file currently is)	str5	%-5s
file_location_date	Location Date	int	%td
earliest_pgpub_number	Earliest Publication No.	str15	%-15s
earliest_pgpub_date	Earliest Publication Date	int	%td
wipo_pub_number	WIPO Publication Number	long	%12.0f
patent_number	Patent Number	str7	%-10s
patent_issue_date	Issue Date of Patent	float	%td
abandon_date	Date of Abandonment	float	%td
disposal_type	Disposal Type	str4	%9s
invention_title	Title of Invention	str600	%-20s
small_entity_indicator	Entity Status	byte	%8.0f
aia_first_to_file	AIA (First Inventor to File)	byte	%8.0f

Table A-2: List of variables included in all_inventors

Variable Name	Description	Type	Formatting
application_number	Application Number	str14	%-14s
inventor_name_first	Inventor's Given Name	strL	%-20s
inventor_name_middle	Inventor's Middle Name	strL	%-20s
inventor_name_last	Inventor's Family Name	strL	%-20s
inventor_rank	Inventor Rank within Application	int	%8.0f
inventor_region_code	Region (State) of Residence	str3	%-3s
inventor_country_code	Country of Residence Code (ISO 3166)	str2	%-2s
inventor_country_name	Country of Residence Name	strL	%-20s
inventor_address_type	Residence or postal address	str9	%-20s

Table A-3: List of variables included in status_codes

Variable Name	Description	Type	Formatting
appl_status_code	Application Status Code	int	%8.0g
status_description	Application Status Description	str97	%-97s

Table A-4: Counts of PatEx applications by series umber

Series	Frequency	Percent	Cumulative Percent
02	55	0.0	0.0
03	176	0.0	0.0
04	4,209	0.1	0.1
05	86,459	0.9	1.0
06	661,447	7.2	8.2
07	742,780	8.1	16.2
08	787,490	8.5	24.7
09	833,489	9.0	33.8
10	953,149	10.3	44.1
11	943,277	10.2	54.3
12	950,821	10.3	64.6
13	926,083	10.0	74.6
14	285,226	3.1	77.7
29	376,214	4.1	81.8
35	2	0.0	81.8
60	493,040	5.3	87.1
61	357,502	3.9	91.0
62	111	0.0	91.0
90	13,408	0.2	91.2
95	2,087	0.0	91.2
96	72	0.0	91.2
PCT	814,073	8.8	100.0
Total	9,231,170		

Table A-5: Cross-tabulation of invention subject matter and application type variables
(Row percentages in *italics*)

Application Type	Invention Subject Matter: Unknown (?)	Design (DES)	Plant (PLT)	Utility (UTL)	Total
Unknown (?)	0 *0.0%*	0 *0.0%*	0 *0.0%*	3 *100.0%*	3 *100.0%*
PCT (PCT)	814,073 *100.0%*	0 *0.0%*	0 *0.0%*	0 *0.0%*	814,073 *100.0%*
Provisional (PROVSNL)	478,473 *56.3%*	11 *0.0%*	6 *0.0%*	372,123 *43.7%*	850,613 *100.0%*
Re-Exam (REEXAM)	357 *2.3%*	282 *1.8%*	7 *0.0%*	14,914 *95.8%*	15,560 *100.0%*
Regular Nonprovisional (REGULAR)	3,696 *0.0%*	470,612 *6.3%*	21,852 *0.3%*	7,030,973 *93.4%*	7,527,133 *100.0%*
Re-Issue (REISSUE)	1 *0.0%*	349 *1.5%*	2 *0.0%*	23,380 *98.5%*	23,732 *100.0%*
Total	1,296,600 *14.0%*	471,254 *5.1%*	21,867 *0.2%*	7,441,393 *80.6%*	9,231,114 *100.0%*

Table A-6: Invention subject matter by filing year, provisional applications
(Row percentages in *italics)*

	Invention Subject Matter				
Filing Year	Unknown (?)	Design (DES)	Plant (PLT)	Utility (UTL)	Total
1997-2005	349,675	0	2	10,079	359,756
	97%	*0%*	*0%*	*3%*	*100%*
2006	44,899	1	1	16,195	61,096
	74%	*0%*	*0%*	*26%*	*100%*
2007	28,761	4	3	34,647	63,415
	45%	*0%*	*0%*	*55%*	*100%*
2008	14,478	1	0	46,992	61,471
	24%	*0%*	*0%*	*76%*	*100%*
2009	9,284	1	0	50,271	59,556
	16%	*0%*	*0%*	*84%*	*100%*
2010	6,731	1	0	56,746	63,478
	11%	*0%*	*0%*	*90%*	*100%*
2011	4,916	2	0	58,196	63,114
	8%	*0%*	*0%*	*92%*	*100%*
2012	2,873	1	0	58,893	61,767
	5%	*0%*	*0%*	*95%*	*100%*
2013	1,685	0	0	39,190	40,875
	4%	*0%*	*0%*	*96%*	*100%*
2014	*30*	*0*	*0*	*874*	*904*
	3%	*0%*	*0%*	*97%*	*100%*
Total	463,332	11	6	372,083	835,432
	55%	*0%*	*0%*	*39%*	*100%*

Table A-7: Common status codes

Code	Status	Frequency	Percent
150	Patented Case	4,288,187	46.5%
161	Abandoned -- Failure to Respond to an Office Action	1,107,772	12.0%
250	Patent Expired Due to NonPayment of Maintenance Fees Under 37 CFR 1.362	919,203	10.0%
159	Provisional Application Expired	848,536	9.2%
30	Docketed New Case - Ready for Examination	350,347	3.8%
218	RO PROCESSING COMPLETED-PLACED IN STORAGE	284,700	3.1%
566	PCT - International Search Report Mailed to IB	234,837	2.5%
41	Non Final Action Mailed	168,103	1.8%
166	Abandoned -- File-Wrapper-Continuation Parent Application	165,589	1.8%
19	Application Undergoing Preexam Processing	126,741	1.4%
N/A	Other Codes	709,623	7.7%
	Missing	27,532	0.3%
Total		9,231,170	100.0%

Table A-8: Distribution of disposal type by filing year, regular nonprovisional utility applications

	disposal_type			
Filing Year	ABN	ISS	PEND	Total
1980	4,324	66,341	43	70,708
1981	8,127	63,933	43	72,103
1982	8,865	65,081	38	73,984
1983	9,706	61,647	28	71,381
1984	11,144	67,201	22	78,367
1985	12,964	71,640	16	84,620
1986	14,176	75,452	17	89,645
1987	14,849	81,757	21	96,627
1988	15,850	90,327	20	106,197
1989	18,363	96,307	18	114,688
1990	20,648	99,536	21	120,205
1991	22,552	100,441	23	123,016
1992	24,593	104,088	20	128,701
1993	25,469	108,479	19	133,967
1994	23,419	123,443	21	146,883
1995	20,038	144,786	34	164,858
1996	12,033	144,785	24	156,842
1997	9,917	169,322	32	179,271
1998	11,924	167,836	32	179,792
1999	15,744	178,557	57	194,358
2000	23,404	191,278	105	214,787
2001	70,064	198,620	316	269,001
2002	74,780	199,355	502	274,637
2003	83,479	194,526	1,005	279,013
2004	98,851	194,835	1,850	295,536
2005	112,667	195,382	3,531	311,580
2006	121,245	197,320	6,991	325,556
2007	118,814	203,536	12,585	334,936
2008	105,716	203,548	19,327	328,591
2009	85,663	191,177	26,364	303,204
2010	82,285	195,781	44,301	322,367
2011	72,188	179,118	87,338	338,644
2012	41,830	133,786	180,290	355,906
2013	10,662	53,460	270,841	334,963
2014	650	5,486	180,318	186,454
Total	1,407,003	4,618,167	836,213	6,861,383

Table A-9: Countries with most mentions of first-named inventors

Code	Country	Frequency	Percent
US	United States	4,526,501	49.9
JP	Japan	1,645,605	18.1
DE	Germany	571,548	6.3
KR	South Korea	328,274	3.6
TW	Taiwan	322,111	3.6
CA	Canada	234,133	2.6
GB	United Kingdom	231,299	2.6
FR	France	217,719	2.4
CN	China	108,570	1.2
IT	Italy	95,827	1.1
NL	Netherlands	88,156	1.0
CH	Switzerland	86,133	1.0
IL	Israel	83,703	0.9
SE	Sweden	81,760	0.9
AU	Australia	65,715	0.7
FI	Finland	41,966	0.5
IN	India	40,879	0.5
BE	Belgium	39,753	0.4
DK	Denmark	33,674	0.4
AT	Austria	33,503	0.4

Table A-10: First-named inventor mentions by US state

Code	State	Frequency	Code	State	Frequency
AL	Alabama	19,296	MT	Montana	5,725
AK	Alaska	2,111	NE	Nebraska	10,653
AZ	Arizona	73,351	NV	Nevada	23,442
AR	Arkansas	8,561	NH	New Hampshire	28,805
CA	California	1,021,662	NJ	New Jersey	201,118
CO	Colorado	90,707	NM	New Mexico	15,764
CT	Connecticut	95,690	NY	New York	306,783
DE	Delaware	23,678	NC	North Carolina	98,337
DC	District of Columbia	5,019	ND	North Dakota	3,665
FL	Florida	146,652	OH	Ohio	169,784
GA	Georgia	77,780	OK	Oklahoma	27,712
HI	Hawaii	5,274	OR	Oregon	69,503
ID	Idaho	30,964	PA	Pennsylvania	177,518
IL	Illinois	193,778	PR	Puerto Rico	1,300
IN	Indiana	75,635	RI	Rhode Island	14,551
IA	Iowa	30,350	SC	South Carolina	31,248
KS	Kansas	24,263	SD	South Dakota	3,530
KY	Kentucky	22,615	TN	Tennessee	40,665
LA	Louisiana	21,364	TX	Texas	291,672
ME	Maine	7,846	UT	Utah	42,326
MD	Maryland	76,561	VT	Vermont	15,359
MA	Massachusetts	214,221	VA	Virginia	64,577
MI	Michigan	184,662	WA	Washington	143,542
MN	Minnesota	137,342	WV	West Virginia	7,372
MS	Mississippi	7,171	WI	Wisconsin	83,068
MO	Missouri	47,163	WY	Wyoming	3,502