Problem Set #[1]

MACS 30200, Dr. Evans
Zhuo Leng

Context and Data Source

The U.S. Patent data file is created by the National Bureau of Economic Research and has three major components: information about U.S. patents granted between 1963 and 1999; citations made to the patents listed between 1975 and 1999 and matched Compustat information. The NBER U.S. Patent data file can be downloaded at http://nber.org/patents/, and there are two compressed formats available: SAS transport (.tpt) files and ASCII comma-separated variable (.csv) files.

These data are described in detail in the follow publication: Bronwyn H. Hall, Adam B. Jaffe and Manuel Trajtenberg (2001). The NBER Patent Citation Data File:Lessons, Insights and Methodological Tools. NBER Working Paper 8498.

Many academic papers have used this data, including:

Youtie, J., Iacopetta, M. Graham, S. J Technol Transfer. Assessing the nature of nanotechnology: can we uncover an emerging general purpose technology? Technol Transfer (2008) 33: 315. doi:10.1007/s10961-007-9030-6

rdi, P., Makovi, K., Somogyvri, Z. et al. Prediction of emerging technologies based on analysis of the US patent citation network Scientometrics (2013) 95: 225. doi:10.1007/s11192-012-0796-4

Bronwyn H. Hall. A Note on the Bias in Herfindahl-type Measures Based on Count Data University of California at Berkeley and NBER September 2000 (revised July 2001, January 2005)

The database is builded on the original U.S. patent record data provided by USPTO. The original patent data has 10 variables, including basic patent information such as Patent number Grant date State of first inventor, Assignee type, Main U.S. patent class and Number of claims etc.

Based on the USPTO patent data, the authors of this database constructed the following 10 new variables:

1. Technological category 2. Technological sub-category 3. Number of citations made 4. Number of citations received 5. Percent of citations made by this patent to patents granted since 1963 6. Measure of generality 7. Measure of originality 8. Mean forward citation lag 9. Mean backwards citations lag 10. Percentage of self-citations made upper and lower bounds

Then, the database is also linked to outside data (Compustat) by match the assignee as the key variable, in order to link relate patent records to the corporation that owns it and bring in data about the corporations.

Summary Statistics

descriptive statistics table

Table 1: descriptive statistics

variables	**count**	**mean**	**std**	**min**	**max**
number of Claims	1984055	12.083	10.268	1	868
Number of Citations Made	2139314	7.720	9.000	0	7.700
Number of Citations Received	2923922	4.779	7.346	0	7.790
Percent of Citations Made to Patents Granted Since 1963	2088795	0.843	0.249	0	1
Measure of Generality	2240348	0.320	0.285	0	0.939
Measure of Originality	2042151	0.349	0.281	0	0.950
Mean Forward Citation Lag	2074641	8.31	5.80	0	96
Mean Backward Citation Lag	2088785	14.010	11.769	0	154
Share of Self-Citations Made - Upper Bound	1703004	0.136	0.256	0	1
Share of Self-Citations Made - Lower Bound	1703004	0.136	0.256	0	1
Share of Self-Citations Received - Upper Bound	1599160	0.132	0.260	0	1
Share of Self-Citations Received - Lower Bound	1599160	0.125	0.250	0	1

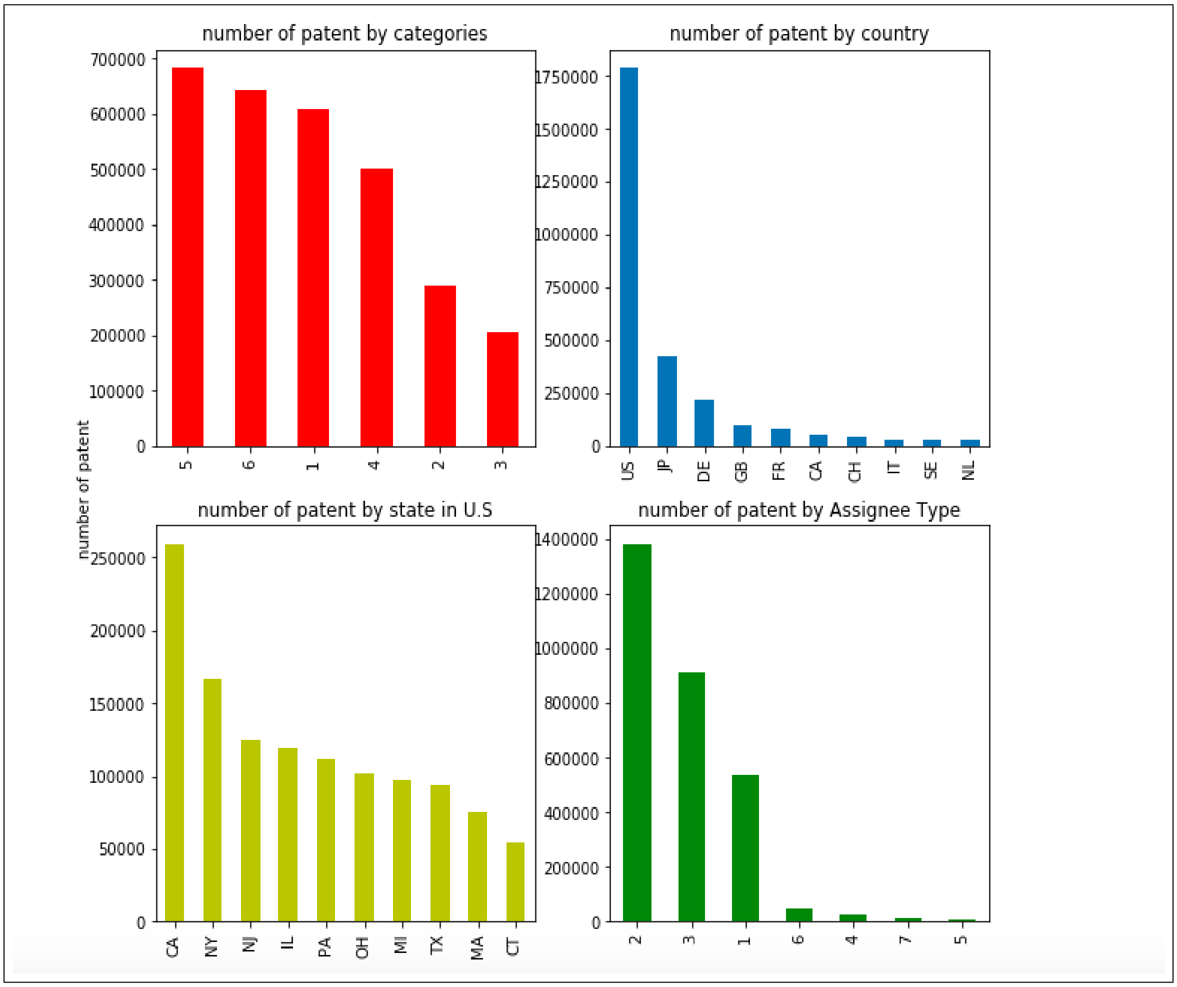
1.The upper and lower bound for self-citations, together with the number of citations received,allow to compute the number of citations received having an assignee identifier (same for citations made).

2.The application year is available for patents granted since 1967.

Source: USPTO, and Jaffe and Trajtenberg computations

Visualization of Data

General bar plot: number of patent by categories, country, state in U.S and assignees type



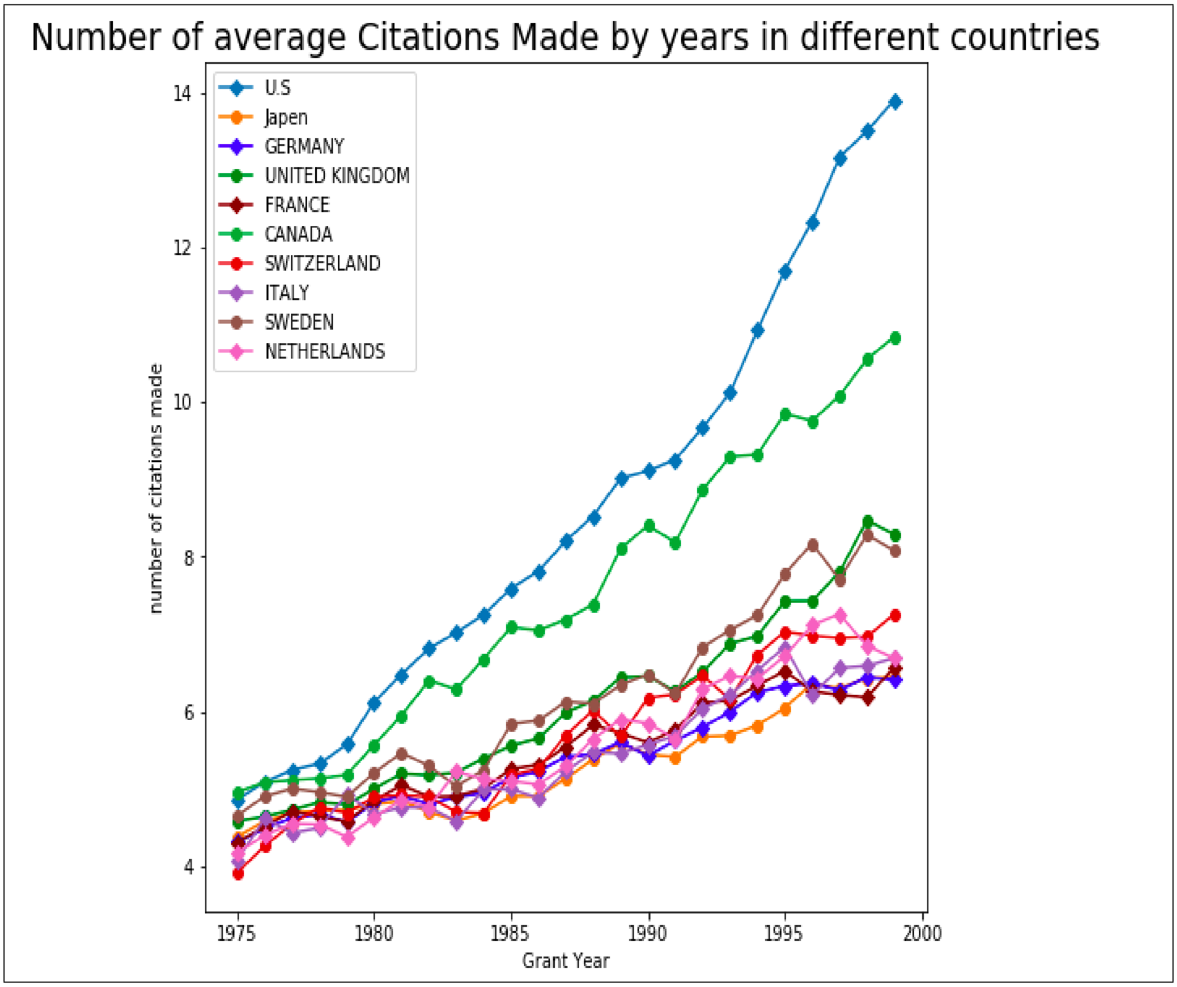
From the bar plot, we could know that the categories of patent which has the most patents are type5: Mechanical. Type 6,1,4, which represent Others, Chemical and Electrical Electronic types, although have a little bit less patent than Mechanical type, occupy majority in overall counts. Type 2:Computers Communications and type3: Drugs Medical patent are relatively less, with counts lower than 300000.

Take a look at number of patent by countries, US has the most patents, with number of counts more than 1750000, and it's number of patent occupy more then 2/3 of that of overall patents all over the world.

Then look at the bar chart of number of patent by state in U.S, only consider the patent in U.S country, CA gets most number of patent. IL ranks fourth in all state in U.S.

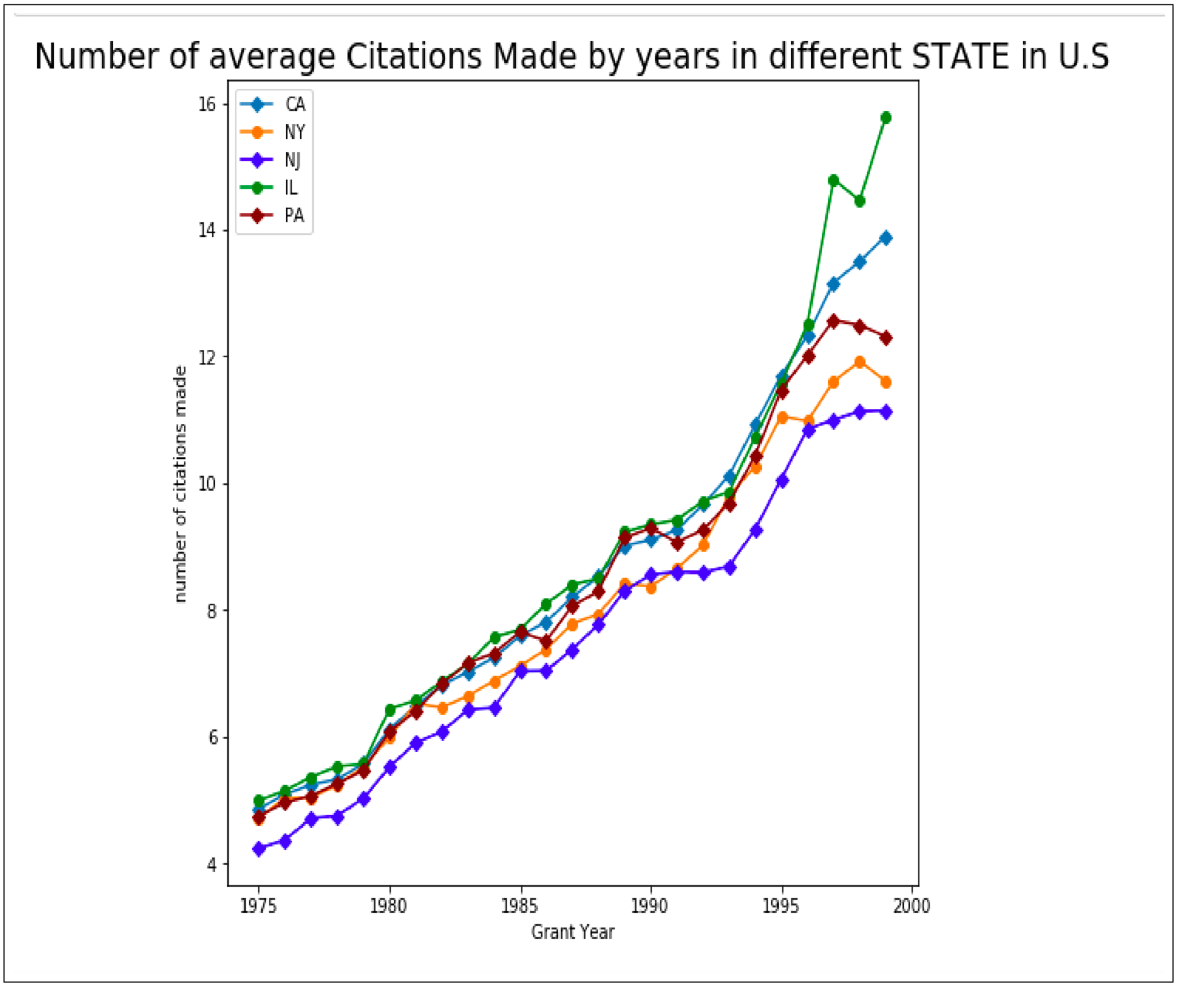
From bar plot of number of patent by assignees type, we could make conclusion that type 2,3,1, which represent assigned to a U.S. nongovernment organization, assigned to a non-U.S., nongovernment organization and unassigned respectively are the most counts type over all patent.

Number of average Citations Made by years in different countries)



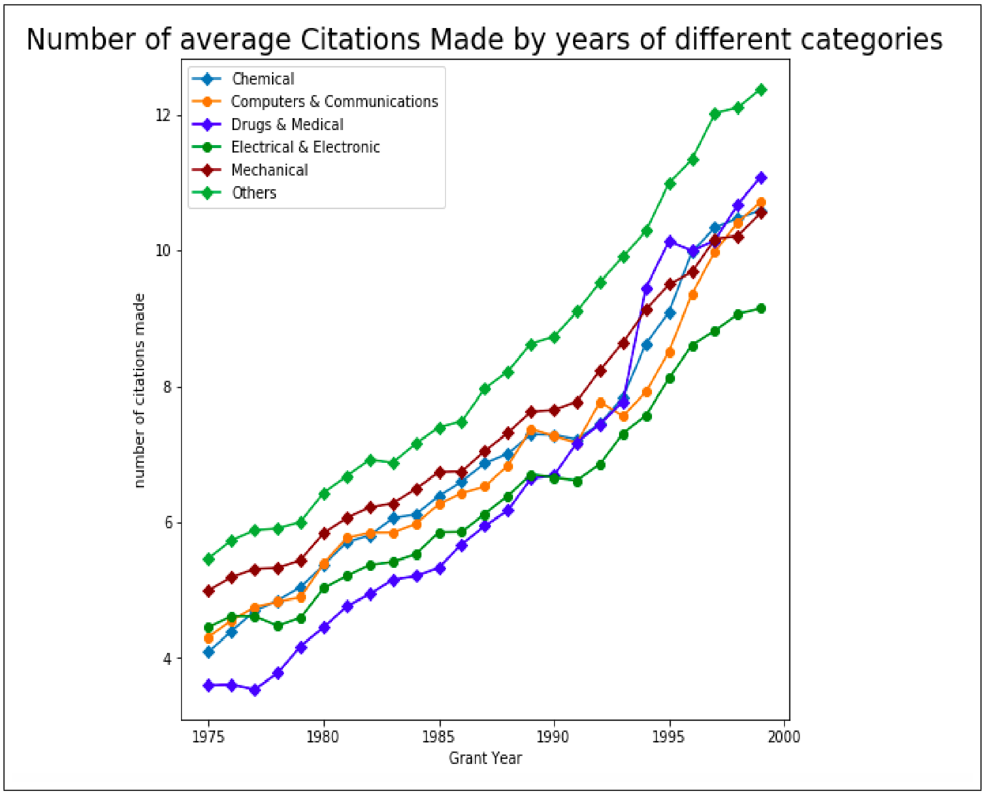
Above plot is number of average citations made by year of countries who rank top 10 of the number of patent. From the plot, we could notice, the overall number of average citations increase with the increase of Grant Years. U.S, which rank first of number of patent also ranks first of number of average citations in each year. From the first bar plot, we know that Japen and Germany rank 2 and 3 respectively in number of patent. However, these two countries seems have least number of average citations made by years.

Number of average Citations Made by years in different state(top 5 number of patent states)



Above plot is number of average citations made by year of states who rank top 5 of the number of patent. From the plot, we could notice, the overall number of average citations increase with the increase of Grant Years in each states. Although IL rank 4 of number of patents, it has the most number of average citations made by each year. Nj has least number of average citations made by years. In all, the numbers of average citations made by year do not have much difference in various states.

Number of average Citations Made by years of different categories



Above plot is number of average citations made by year of categories. From the plot, we could notice, before 1990, drugs and medical has the least number of average citations. However, by 2000, it increase to top 2 most number of citations categories. electrical electronic type after 1990 has least number of citation among all 6 types.