Data Description

Xiuyuan (Flora) Zhang

4/14/2018

Part I. Your assignment is to write a data description of a paper for your assigned dataset from the list below.

1. Describe how to access data, where it is stored, who curates it. Make sure to use the original source and curator in addition to the NBER site to which I have linked.

The data set is from The NBER Patent Citations Data Files. It combines two data sources: patent data from Patent Oce and rm level dataset from Compustat. The patent data is stored in NBER US Patent Citation Database. NBER is working on a funded update and extension of this data so now a new release of files contained existing data up to date through 2006 are now available in Patent Data Project. Original data from 1963 to 1999 are still available on NBER patent website. All the data can be accessed free of charge. This data set is curated by Bronwyn H. Hall, Adam B. Jaffe and Manuel Trajtenberg and those curators also wrote the paper *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools.*

2. Cite other key papers that have used this data.

1. Hall, Bronwyn H., Adam Jae, and Manuel Trajtenberg. *Market value and patent citations.* RAND Journal of economics (2005): 16-38.

2. Jung Won Sonn and Michael Storper, 2008, *The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975 -1997,* Environment and Planning A 2008, volume 40, 1020-1039.

3. Natarajan Balasubramanian and Jagadeesh Sivadasan, 2011, *What Happens When Firms Patent? New Evidence from U.S. Economic Census Data,* The Review of Economics and Statis- tics, February 2011, Vol. 93, No. 1, 126-146.

In the three paper listed above, Hall, et al.(2005) used this dataset to study the relationship between a rms patents and the market value of the rm. Sonn and Storper(2008) used the NBER Patent Citation Data File to examine the importance of geographical proximity in knowledge production. Also Balasubramanian and Sivadasan(2011) used this dataset along with U.S Census microdata to examine what happens when firms patent.

3. Describe how the data were collected.

The rst data source is patents granted from January 1, 1963 to December 30, 1999 from United States Patent and Trademark Oce (USPTO). The original data set consists of 10 variables and contains information on patent grant and applica- tion date, country of the inventor, and patent class, etc. The curators of the data set also added other new variables on technological category, number of citations made and received, and two measures of generality and originality. 'Citations made' refers to the citation one specific patent made while 'citations received' refers to all of the citations one patent receive after being granted. 'Measure of

generality' represents how wide of the field one specific patent inuence, while 'measure of originality' is computed using citations made by one patent.

## 4. Include a table that gives descriptive statistics for at least 8 key variables.

Statistic	Count	Mean	Std. Dev.	Min	Max
Grant Year	2923922.0	1983.55	10.98	1963.0	1999.00
Grant Date	2923922.0	8785.65	4013.28	1096.0	14606.00
Application Year	2699606.0	1983.11	10.13	1901.0	1999.00
Number of Claims	1984055.0	12.08	10.27	1.0	868.00
Number of Citations Made	2139314.0	7.72	9.00	0.0	770.00
Number of Citations Received	2923922.0	4.78	7.35	0.0	779.00
Percent of Citations Made	2088795.0	0.84	0.25	0.0	1.00
Measure of Generality	2240348.0	0.32	0.28	0.0	0.94
Measure of Originality	2042151.0	0.35	0.28	0.0	0.95
Mean Forward Citation Lag	2074641.0	8.31	5.80	0.0	96.00
Mean Backward Citation Lag	2088785.0	14.10	11.77	0.0	154.00

5. Include at least one key visualization of the data that exhibits an interesting characteristic.

The Number of Granted Patents by Grant Year
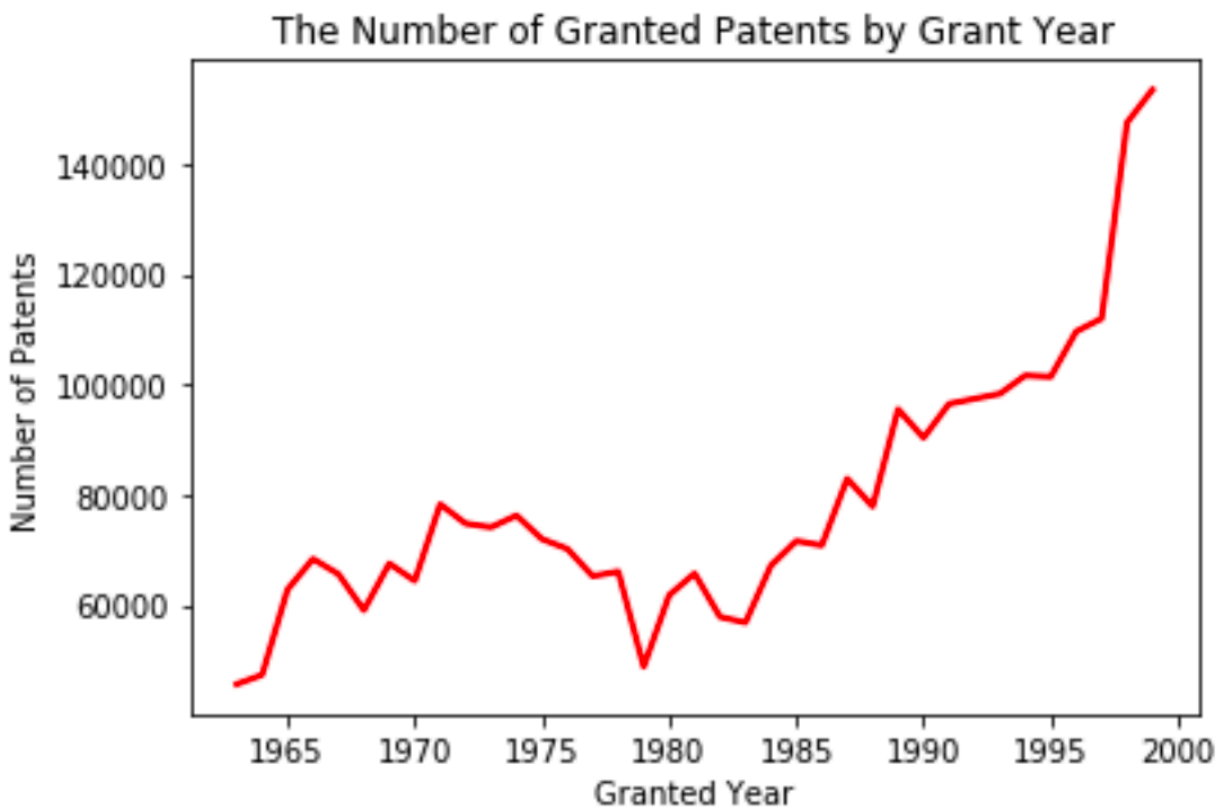
This graph shows a general trend for number of granted patents by year. One can see from the graph that, while there are up and downs for each year comparing to previous years, there is a general trend of increase in number of granted patents. The growth is rather rapid starting from the year 1983.

6. Show at least one conditional (slice) description of the data (e.g., all variable descriptive statistics by nationality of survey respondent). This can be a table or visualization.
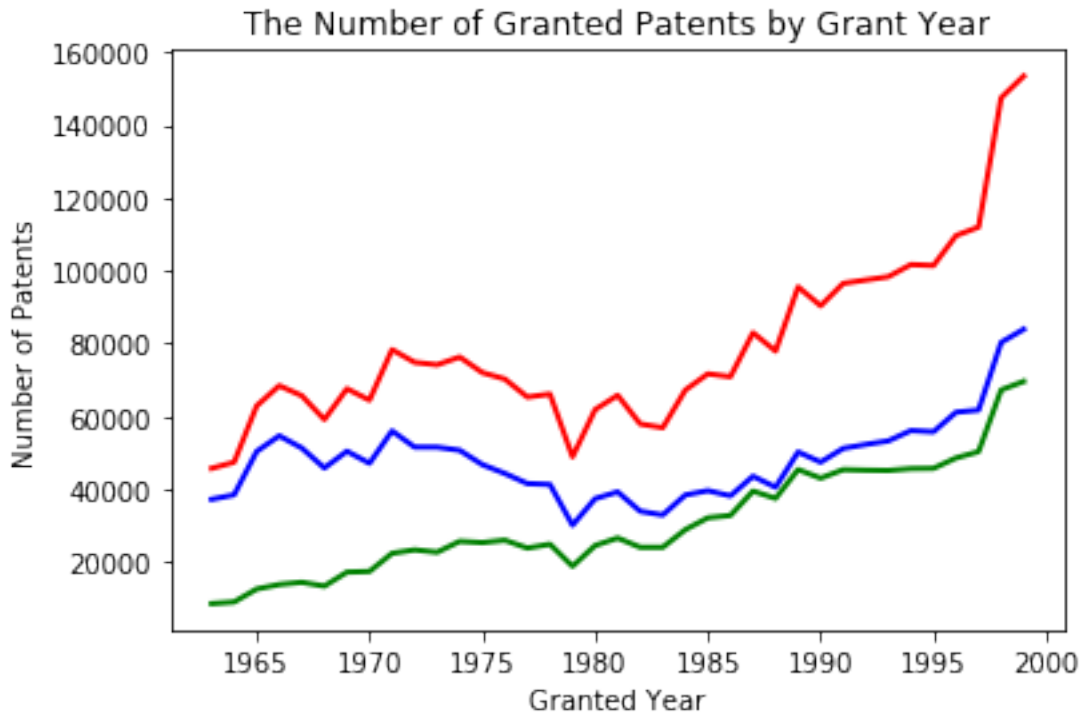
This graph conditions on the country (U.S. or Non U.S.) in which patents are granted in by application year. One can see from the graph that there are more patents granted in the U.S. than other countries. However, the gap between started to becomes smaller and smaller starting from 1980. Thus, the rapid increase we saw from previous general trend plot has a major contributor: the increase of patents by other non U.S. countries.