Analysis of Science and Technology Development Across the World from 1995 to 2010

Fan Yang and Sheng Zha
{fyang,szha}@cs.umd.edu
Department of Computer Science, University of Maryland, College Park, MD 20742
October 5, 2012

1 Introduction
Across the world, the science and technology development of nations and regions significantly varies. Usually it is difficult to infer useful information by only looking at the tremendous amount of raw data collected from several decades and hundreds of nations. Therefore, a good information visualization is desirable for people to effectively explore the data to understand the progress of science and technology development in various nations and regions.

In our project, we mainly focus on the exploration of science and technology development of nations and regions across the world from 1995 to 2010. Are there any interesting facts available if we dig deep into the data? Data itself will tell us everything.

2 Data
2.1 Raw Data Source
The original raw data is publicly available at the World Bank website, which can be found from http://data.worldbank.org. The data set is divided into 18 categories, each of which contains more than 200 nations and 15 to 50 attributes and ranges from 1960 to 2011. Unfortunately, there is a huge number of missing data in early years and from underdeveloped nations, which makes freely choosing attributes to analyze inapplicable. Therefore, we need to preprocess the raw data and choose the attributes with most information before starting the exploration.

2.2 Preprocessing
First, we only extract the data from Science and Technology Development category ranging from 1995 to 2010 for our analysis as we care more about the recent development. Then we choose relevant attributes with sufficient values and rearrange years, which are originally separated attributes, as a single column. Finally, we have a new worksheet containing 4182 items including 246 nations and regions, and 66 attributes which are country name, year and topic-related attributes, ranging from 1995 to 2010.

Some important attributes are

• Scientific and technical journal articles
3 Tools

In our analysis, we use Spotfire as major software to explore the data and Tableau as an complementary tool to extend our analysis.

4 Insights

4.1 Number of researchers relies on both tertiary education rate and expenditure on research

If nationals receive better tertiary education, do they contribute more to the number of researchers in R&D field after graduation? On the other hand, does a nation’s investment on research attract more people to enter research field? With these questions in mind, we first analyze the correlations of the research and development expenditure, tertiary enrollment rate and researchers in R&D per million people of various nations in 2006, as shown in Fig 1.

We easily observe a strong correlation between tertiary enrollment rate and number of researchers: the increase in the tertiary enrollment rate indeed increases the number of researchers per million people.
Figure 2: World map illustrating the impact of two factors on researcher number. Top: the correlation between average tertiary enrollment rate and average researcher number through 1995 to 2010. Larger spot means more researchers per million people. Red means high tertiary enrollment rate while green means low tertiary rate. Bottom: the correlation between average research and development expenditure and average researcher number through 1995 to 2010. Larger spot means more researchers per million people. Red means high expenditure while green means low expenditure.
Accordingly, if the percentage of nationals enrolled into universities and colleges is low, researchers inevitably becomes insufficient, which is clearly shown in the scatter plot where nations on the bottom left have both low tertiary enrollment rates and researcher counts.

However, there are also special cases available which are Venezuela and Ukraine. Although they have very high tertiary enrollment rate, the researchers per million people are significantly low. When we vertically inspect the nations with more than 70% tertiary enrollment rate, the reason pops up that lack of research investment leads to insufficient researchers despite of high tertiary enrollment rate. Therefore, a nation should provide enough financial support to research and development even nationals have received good higher education; otherwise they may be reluctant to enter research field, which leads to slow progress in science and technology.

When digging deeper into the visualization, we find another fact that European nations have higher tertiary enrollment rate and research expenditure, therefore have more researchers per million people, for example, Finland, Sweden and Denmark, which again supports our finding that researcher number relies on both tertiary enrollment rate and research expenditure. We then use Tableau to extend out findings. The top map in Fig 2 shows the correlation between average researcher number and tertiary enrollment rate throughout 1995 to 2010 all over the world while the bottom map in Fig 2 illustrates the relationship between average research number and research expenditure. Clearly, European nations are likely to invest more money on research and improve tertiary education therefore have more researchers per million people. Other nations aiming to improve science and technology competence should learn from those European nations.

### 4.2 A few nations control most technology

Next, we investigate multiple attributes related to science and technology development across different nations to find possible remarkable patterns. We inspect four attributes which are high-tech exports (current $US), patent applications, royalty and license fee receipts and scientific and technical journal articles. These factors are important indicators of the level of science and technology development in a nation.

Fig 3 shows the treemaps of the four factors. The size of rectangles indicates the amount of each factor while the color indicates the number of researchers per million people. Instantly we can draw a conclusion that a small number of nations seize more than half of the total scientific and technical products in the world. In particular, United States, Japan, United Kingdom, China, Germany and France are the most powerful nations in science and technology, which are just the top 6 nations with the largest GDP in 2010, according to the United Nations’ report. It means that richer nations have more capacity in developing science and technology; however, poor nations such as nations in Africa involved in fighting with disease and hanger are not able to develop technology at all.

This is a clear-cut example of the so-called “digital divide” phenomena which aggravates the imbalance in science and technology between rich and poor nations and leads to scientific and technical monopoly as the result of globalization. The global community should pay attention to this as the “digital divide” may bring in instability to the world.

---


Figure 3: Illustration of amounts of four types of scientific and technical products shared by nations through 1995 to 2010. Darker color represents more researchers per million people. From top to bottom: total number of scientific and technical journal articles, total number of patent applications, total amount of high-tech exports (current $US) and total number of royalty and license fee receipts.
4.3 The dominance of US in science and technology is being challenged

Finally, we inspect the trend of science and technology development of the United States which is the most powerful nation in science and technology in the world. We choose four attributes and show that the United States will maintain the overlordship in the foreseeable future but its dominance is being challenged in some aspects.

Fig 4 shows the trends of the top 10 nations with the largest GDP according to the United Nations in 2010 \(^3\) from 1995 to 2010 using Tableau. The thickness of lines indicates the research and development expenditure, which remains stable for each nation.

Clearly, the United States is far beyond other nations regarding the number of scientific and technical journal articles and royalty and licence fee receipts. The two factors reflect a nation’s current theoretical contributions to and previous achievement in science and technology. Considering the huge superiority, we believe that the United States will still dominate the two fields in a few decades.

On the other hand, China becomes active in high-tech exports and patent applications from 2000, resulting in a rapid ascent curve, while the two attributes in the United States are stable and the number of patent applications in Japan even has dropped. The reason for that may be the national policy on stimulating science and technology development in recent decades in China. Therefore, the United States needs to find solutions to boost the high-tech exports and patent applications and to find out whether the current situation is a signal of insufficient innovation; otherwise, the dominance of the United States in science and technology will be shaken.

5 Critiques

5.1 Spotfire

Spotfire supports various kinds of visualizations from which we are most impressed by the scatter plot. Using scatter plot, information is more condensed where at most 4 attributes are shown within a single figure, which is very helpful when we need to analyze the correlations of multiple attributes at the same time. It is also good at dealing with empty values and loads data fast. Some bugs and issues are listed as follows.

- From our experience, if years are listed as separated columns, Spotfire could not effectively use them as horizontal axis to show the trend. A relaxation on the data format would be better.
- It cannot handle noisy data well in scatter plot visualization, especially when there is a single value much larger than all the other ones. Spotfire tries to enclose all values in the plot instead of wisely throwing away the noisy value.
- It is difficult for us to accurately pick up the attribute we want to show since the drop-down menu is small. We would favor a highly-customized drop-down menu.
- When drawing lines, the thickness of lines cannot adapt to any measure, which is not flexible.

5.2 Tableau

The interface of Tableau is more attractive and friendly. We are excited that we can freely drag and drop attributes into filters and marks, which greatly improves working efficiency. Another impressive functionality is that Tableau can read nation names and map them on the real world map. In our

\(^3\)http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)
Figure 4: Illustration of trend of science and technology development of top 10 nations with the largest GDP from 1995 to 2010. Top: trends of number of scientific and technical journal articles and royalty and license fee receipts. Bottom: trends of high-tech exports and number of patent applications. Nations are represented by different colors. The thickness of lines indicates the amount of research and development expenditure. Nations listed in the right are ordered according to the GDP amount.
analysis, it only made one mistake in recognizing nation names that can be easily corrected by hand. It is also very helpful by using additional attribute to control the thickness of lines.

Some negative experiences are listed as follows.

- We feel that the data loading speed of Tableau is slower than that of Spotfire. Moreover, after loading data, users cannot filter out irrelevant attributes before doing data exploration.
- Sometimes Tableau cannot successfully determine the type of attributes. It continuously treats numerical type values as string type values.
- It is hard to find how to directly convert the current visualization to another type of visualization rather than creating a complete new one.