

The Global Market in Proportional View

Data Visualisation Report

Wan Yilun, Wu Zhendong - November 7, 2014

1 Introduction

When we scan the information on Stock Market Websites, what we usually obtain are the numbers and the basic line-charts of the stocks – which intuitively show us the rising or the falling of certain stock prices. However, we seldom witness other kind of charts that can be used as proportional comparison in stock market. As a new approach, we use pie-chart relevant visualisation to deal with vast number of Nasdaq companies as well as Financial Times Top500 datasets, aiming to explore and compare several important company financial figures in different perspectives. After processing and interpret the datasets, we successfully realise the planned visualisations and discover some interesting findings.

2 Visualisation Purpose

With a quick glance on the present stock market websites, what we see is always typically line charts or area charts with large number of tabs, showing the figures including price of the stocks, volumes of certain hour, which are those that investors and brokers most interested in. For example, in popular stock websites like Nasdaq¹ or Yahoo Finance², users can use interactive charts to get the price, volume and other relevant figures by interacting with the charts, but all of the figures are presented in bar charts and area charts, with time as the horizontal axis. These charts are good at retrieving information that is strongly connected with time attributes so that we can see the past changing of other certain attributes, with the hope of predicting future trend.

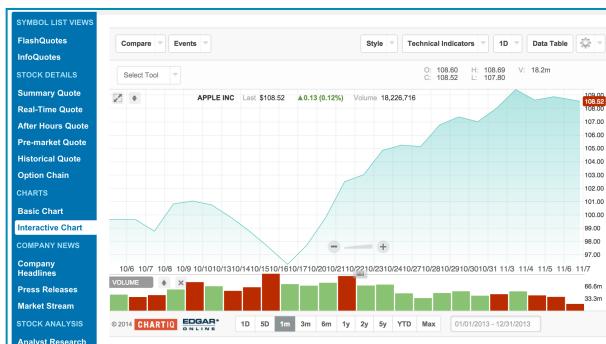


Chart 1.1 Nasdaq Interactive Chart



Chart 1.2 Yahoo Finance Interactive Chart

However, when we turn to questions like “What is the proportion of Chinese companies in the Global market?”, “Is US companies still the dominators in the global market?” or “Did the proportion of Asia’s companies in world Top 500 changes over the past 5 years?”, these simple line charts and bar charts fail to respond to us. When we try to find pie charts or relevant diagrams that

¹ Nasdaq Interactive Charts, <http://www.nasdaq.com/symbol/aapl/interactive-chart>

² Yahoo Finance Interactive Charts: <http://finance.yahoo.com/echarts?s=%5EGSPC+Interactive#>

are more suitable to answer these issues, we can hardly retrieve any correlative charts. In some cases we can find Rank lists of companies that are break in to countries, but are not visualised in pie chart for users to have a more intuitively insight on the percentage that each countries have. Like the *Top 10 countries with the most Global 500 companies rank list*³, we can have a quick look on which country has the most companies in this year and the top 5 countries that on the table. However, its not direct and clear for us to see the proportions of each country merely looking at the numbers in the table. Moreover, we can

hardly see the proportion changes over time. Most of the datas are still in simple table forms or line charts form, showing that this area is still a blank considering visualisation perspective.

In this situation, we decide to develop pie-charts relevant visualisations in order to show the global market situations in a proportional view. Moreover, time attribute is taken into consideration too, for we expect to see the proportional changing trends over time. Some specific questions have been come up with and are expected to be answered by the final result outcomes of our visualisation.

1. What is the proportion of the Total Market Capitalisation of all Chinese (or any other countries) companies` stocks in Nasdaq Market or in Global Market in certain year?
2. What about proportion of the Total Number of all Chinese companies in certain year? Can we compare it with other countries in a more direct way?
3. What is the above figures and diagrams looks like in the past years? Is the proportions of those figures getting larger or smaller over time?
4. Is the companies with Asian countries background start to play an important part in the top companies list these years? What are the proportions of it during the past 5-10 years?
5. What is Apple`s Market Proportion`s in the Top global companies? How does it change over the past 10 years? Does it play an important role in US`s proportion?
6. What`s Book Industry (or any other industry)`s total Market cap`s proportion in Nasdaq market or in the global market? Are this proportion changing over time?

| Rank | Country | Companies |
|------|----------------|-----------|
| 1 | United States | 128 |
| 2 | China | 95 |
| 3 | Japan | 57 |
| 4 | France | 31 |
| 5 | Germany | 28 |
| 6 | United Kingdom | 28 |
| 7 | South Korea | 17 |
| 8 | Switzerland | 13 |
| 9 | Netherlands | 13 |
| 10 | Canada | 10 |

Table 1.1 Top 10 countries with the most Global 500 companies

³ wikipedia, http://en.wikipedia.org/wiki/Fortune_Global_500

3.1 Data Selection

First issue before picking the data is to choose the **Reference Attribute** that we focus on. This attribute must be easy to understand by the mass, and it should make sense when we adapt it into piecharts. Available attributes on the list include Market Capitalisation, Prices of the Stock, Turnovers of the companies, revenue, number of the employees and etc. With some analysis and comparisons we decide to choose the Market Capitalisation as the reference attribute to compare between the companies and the countries, and there are several reasons for this. Firstly, hierarchy structure has been included in our target visualisation, so it makes more sense when we sum up the figures in higher level of the hierarchy when we adapt market capitalisation instead of the stock prices or the revenues. Secondly, Market Capitalisation is a relatively simpler concept to understand compare with other attributes, the result will be more intuitive to the users when they access to the visualisation for the first time. Thirdly, the total market capitalisation is an open figure in the market space compare to other more sensitive figures like the employees' numbers and so on, so the data will be more easier to acquire and the Data-missing event will occur less often.

After setting up the visualisation target and the key attribute, we start to look up the available datasets. The first option that comes into our scope is the **Companies List of Nasdaq Stock**⁴. This dataset is clean while contains complete information of each companies, not only contains the Market capitalisation that we need , but also includes other figures like Company Names, Last sale price, countries that it belongs, sectors and industries attribute of the company, and its stock symbol in the market and so on. We keep this dataset and retain the country attributes in order to establish a hierarchical structure basing on geometric location.

With this datasets we build the first visualisation basing on d3.js, but later we realise that although this data set is perfect for demonstration of the situation in the present year of 2014, we can hardly retrieve the similar forms of datas in the past years. At first we aim to acquire the data by accessing online datasets and download the datas of the past years, but it turns out that this kind of database is hard to access in Nasdaq Official websites. We turn to other third party website like yahoo finance tempting to obtain the past Market Capitalisation of the companies, but it turns out that most of the past figures have been covered, while more up-to-date datas which investors more interested in are brought to the front. Moreover, when we take a look on our first visual outcome, we realise that putting few thousands companies on the outer part of the pie

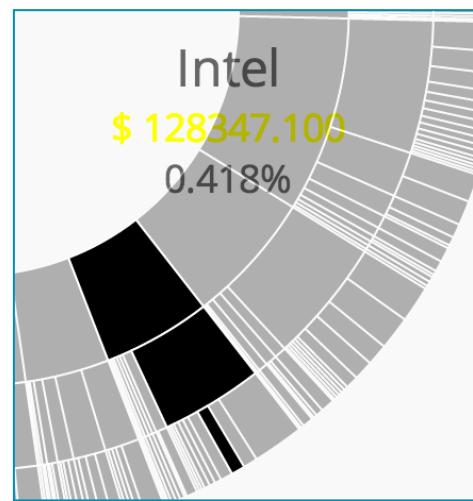


Chart 3.1 Intel in the Nasdaq Stocks
Visualisation 1st edition

⁴ Nasdaq Companies List, <http://www.nasdaq.com/screening/company-list.aspx>

chart is not realistic, most of the small companies will be hidden for the space and the proportion that they occupy in the big picture is too small. From the Chart 3.1 we can see that even for the large company like Intel, it only take a really small fraction in the whole pie-like diagram. In this perspective, user can have a more clear view of the big picture, but it will takes a while to find the companies that they are interested in this case, even though the target companies might be relatively important players in the market. With further research and discussion we figure out that we can focus more on the top companies in the global market rather than putting all the companies into the picture. In this way we can adjust the data size to certain amount so that it can be more precise in interpreting the data. Moreover, companies rank base on Market Capitalisation also reflects the economic power of the countries and continents that companies belongs to in global market in a way.

For the second data search approach we focus on the **Top Company List**, and we find that we can quote the Institutional Reports developed by authority institutions that work on this. Relative reports includes Fortune Global 500⁵, Financial Times 500⁶, Forbes Global 2000⁷ and so on. After comparison we choose Financial Times 500 as our data sets, as Fortune Global 500 use Revenue as their criterion, and Forbes 2000 `s data size does not meet our size requirements. Moreover, consider Financial Times is a UK based media, it`s datas are more independent and Objective. Last but not least, Financial Times Global 500 opens a number of its historical reports to the registered users, while other Report Institutions require an relatively expensive membership fee in order to access the past data.

3.2 Data Processing

Data Processing Tools: Adobe Acrobat/Microsoft Excel/Sublime

The raw data we collected in Financial Times Global 500 official websites are annually reports base on PDF format. They cover the year from 2006 to 2014. By the Internet Archive⁸ we can also access to the report date back in year1996. Considering the total span of the time, we choose the data start from 2006 to 2014 for every 2 years. During the data processing period, we encounter several Difficulties.

1. PDF - Excel Formatting Issue - Dealing with the PDF based tables require large amount of manual works, for the tables that convert to excel by Adobe Acrobat or any other tools will result in misplacing of certain import ant data. We have to sort and fix the false data in order to prevent the distortion of the truth.

⁵ Fortune Global 500, <http://fortune.com/fortune500/>

⁶ Financial Times Global 500, <http://www.ft.com/intl/indepth/ft500>

⁷ Forbes Global 2000, <http://www.forbes.com/global2000/>

Web Archive,Financial Times Website in year 10-31-2002, http://web.archive.org/web/20021031135258/http://specials.ft.com/ln/ftsurveys/ft5_glob.htm

2. Large number of N/A cases - As we expected, some part of the data is missing from the data set so we have to do research and fill in the N/A cases. Moreover, because the data was converted from PDF files, there may happens that part of the data were lost and need to be repaired. For instances, in Chart3.2, the names of the companies are hidden and can not be restored directly, in this case we need to do search and refill by codings in order to repair the data.
3. The raw data of PDF format reports are divided into years, and in each years' report the data has been part by pages. We have to remerge the data as well as dealing with the duplicated cases. For instances, IBM appears in all the five reports, when merging the data, they should not be treated as 5 single rows but should be merge into one line.
4. The datasets type that d3.js requires varies from .csv to .json , therefore we have to sort and convert the datasets into the format that we need. Relevant adapted tools include sublime and Microsoft Excel.
5. The hierarchical attribute of Continent that we desire is not included in the datasets. We have to write programs to form the Continent Columns basing on the countries that certain company belongs.

| Company | Country |
|------------------|---------|
| 1 Apple | US |
| 2 Exxon Mobil | US |
| 13 Berkshire Hat | US |
| 3 PetroChina | China |
| 11 Wal-Mart Sto | US |
| 9 General Elect | US |
| 4 Microsoft | US |

Chart 3.2 A Typical missing case example

4.1 Visual Tools

Data Interpretation Tools:

D3.js /Tableau Public/Adobe Kuler

While choosing the tools we have to consider the complexity of the attributes that we want to includes. Each of the companies process two quantitative data - the Market Capitalisation, and the rank of the Company. Moreover, the datas we need to present have important hierarchy structures that should not be ignored. For instance, Microsoft belongs to United States of America, while US is part of North America Continent. While counting the data in

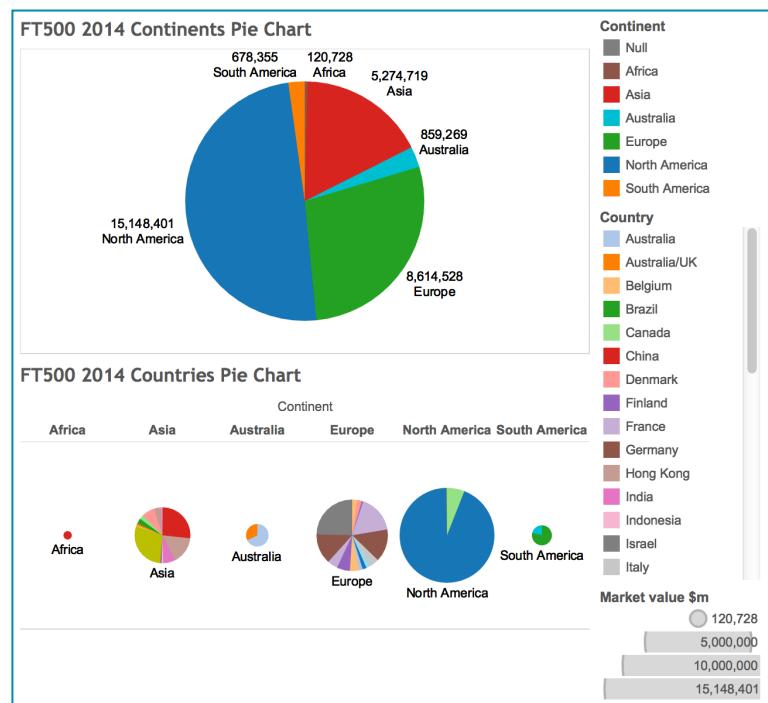


Chart 4.1.1 - 2014 Financial Times Global 500 Visualisation

each section, we not only have to add up the total capitalisations of each section, but also need to calculate the total counts of the companies in each part. After all, with the hope of exploring the proportions changing trend over time, we have to include the time axis to our visualisation.

After setting up the goal, we tried several visualisation tools to demonstrate the data we retrieved and settled. Tools including excel diagram, tableau tables are used during this period. Outcome in this step includes the Chart4.1.1, which is the Financial Times Global 500 Visualisation for the year 2014 report. However, after certain numbers of attempts we find out that demonstrating these datas with statistics pie charts will need a numerous number of them as well as large space. If we include the time axis into consideration, the outcome turns out to be the Chart 4.1.2⁹.

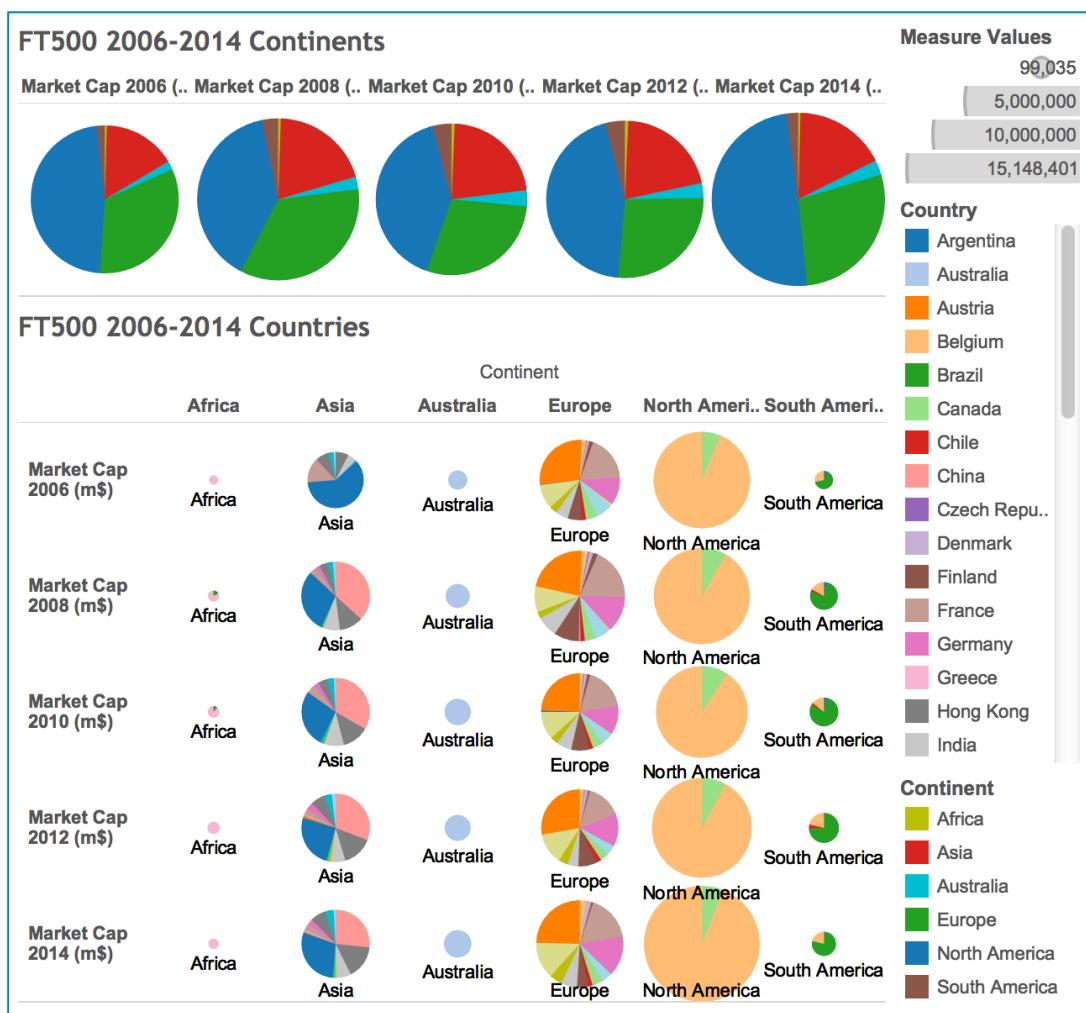


Chart 4.1.2 - 2006 to 2014 - Financial Times Global 500 Companies` Total Market Capitalisation that break into Continents and Countries

This chart is interactive and users can put the mouse on certain section to see specific information. This visualisation already provide a brief look on the proportion view of the Top 500

⁹ Financial Times Global 500 Companies` Total Market Capitalisation that break into Continents and Countries 2006 - 2014, https://public.tableausoftware.com/views/FT5002006-2014/Dashboard1?:embed=y&:display_count=no

companies that base on geographical hierarchy, and users can acquire the certain data easily by filtering through year, Continent and country. However, from the chart we can see that the data presentation occupies a huge space, while the changes over time are not easy to be analysed. Moreover, it only breaks into countries and did not show the data of certain companies. Although the chart is interactive already and showing the proportional views in a way, it fails to answer some of our questions that relevant to the changes over time. Also, we can not further explore the detailed proportional figures within a country. We conclude that Static Pie charts are naturally not good at interpreting hierarchical structure and time changing trends, therefore this visualisation is not the target visualisation and we need to come up with new tools for better demonstration.

The second form of visualisation we refer to is a hierarchical sunburst diagram. Sunburst diagram is a visualisation form that developed upon simple pie charts, it includes the hierarchical structure that each record process and interpret it in a direct way. Take Chart 4.1.3¹⁰ for example, the chart clearly shows the hierarchical structure and the proportions that each section holds. This diagram form is more suitable for our target visualisation for it includes more layers in to the pie chart and greatly reduce the waste of space comparing with using a large number of simple pie charts. By adopting the sunburst diagram, the proportions of each section can be shown together with the structure of the records. As for the tool that realises the goal, d3.js is one of the best choices despite the fact that we have to input a large portion of time diving in coding work. With this tool we can add interactive effect on visualisation and save spaces for information. Once the user hovers the mouse over certain section, the information will pop up and show the relevant data and calculated proportions result.

In this part of the visualisation we use the Dataset that acquire from Nasdaq Stock, aiming to test hierarchy structure. d3.js requires the data format of .json, which is a more strict form in order to show the hierarchy attribute. After programming to process the data, we got the interactive Sunburst Diagram that base on Sequences sunburst visualisation technique. The outcome and relevant analysis is shown in Part 4.2, the demonstration charts include Chart 4.1.2, and Chart 4.1.3.

After achieving the first sunburst interactive diagram, as we mentioned in Part 3.1, we find that this graph still undergo issues dealing with the changing of time. Technically speaking we can develop 5 different graphs of this kind representing each year of the time, but in this way we fail to meet the goal of saving space and directly showing the differences. Moreover, demonstrating 5 graphs together will consume a numerous amount of computing resources for users' system, and the delaying effect is somehow critical and will greatly harm the user experience for our demonstration.

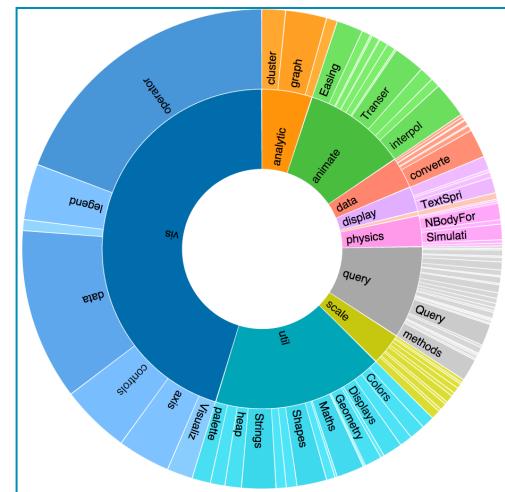


Chart 4.1.3 A Sunburst Visualisation Example

¹⁰ vgrocha, Sunburst bilevel partition with tooltips, <http://bl.ocks.org/vgrocha/1580af34e56ee6224d33>

Under this situation, we further explore the available tools until we discover Zoomable sunburst with updating data¹¹, which is also a d3.js tool but is capable of dealing with updating data. Take the Chart 4.1.4 as an example, the sunburst diagram on the left shows the origin datas when the “Count” checkbox is checked. When user click on the “Size” checkbox,



Chart 4.1.4 - The data updating effect of the Zoomable Sunburst Diagram

We come up with the idea that by modifying the checkbox numbers and effects we can shift its function into time bar effect. In other words, we can create a line of the checkbox in sequence, arrange them horizontally and form the time bar controller. By the time users click on any checkbox that represent a time, the diagram will update according to the dataset we include(as the effect shown in Chart 4.1.5) In this case we perfectly solve one of the critical questions that we ask at the beginning of our visualisation tour - “How does these figures change over time?”

Moreover, the outcome of this visualisation tool is capable to zoom in and zoom out when user click on certain section of the diagram, which allow users to explore the proportion distribution within each section. For example, user can see the proportions of China in Asia market by zoom into Asia Section, then change the time checkbox and see the changes of China's proportion.

Eventually, considering we are adopting d3.js for visualisation, platform of website becomes the first choice for us to integrate all the visual outcomes that we produce during this visualisation tour. By this means we can also share our discoveries and outcomes with others, retain comments and suggestions for further improvement.

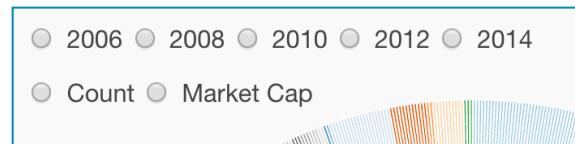


Chart 4.1.5 Virtual Time Bar effect

¹¹ kerryrodden, Zoomable sunburst with updating data, <http://bl.ocks.org/kerryrodden/477c1fb081b783f80ad>

4.2 Outcomes and Insights

There are mainly three parts of our final visualisation results.

First part is the Tableau Worksheet that we create basing on the primitive data. The diagram is relatively large and abundant, therefore we does not include it into our webpage. However, we also been tried and tested other views in tableau to cover the shortage of pie view based visualisation. For instances, in the Chart 4.2.1 shown below, we can clearly see the proportion of each continent's top 500 countries in the scope of the whole globe, together with the measuring of the total Market capitalisation for each year. From this diagram we easily can tell that the total Market capitalisation of the top 500 companies are fluctuating over time, rather than growing by time. The Chart 4.2.2 on the right clearly shows how the total Market Capitalisation of each continent changes over time. These charts cover the deficit of the sunburst charts which are not able to show the total sum for each year.

This worksheet can be viewed online anytime through the site URL that stated in the footage of page 7, Chapter 4.1.

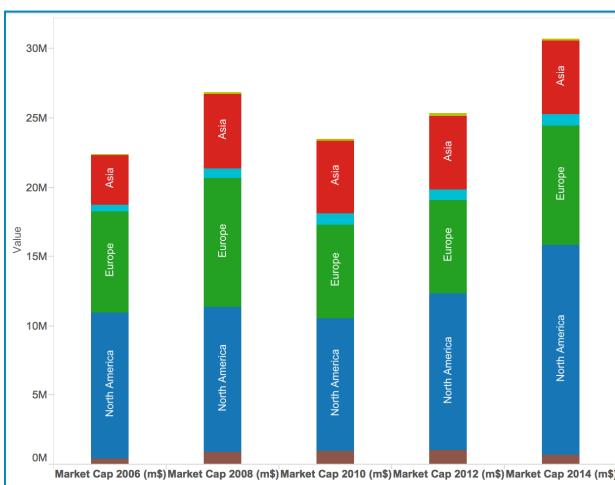


Chart 4.2.1 A Sunburst Visualisation Example

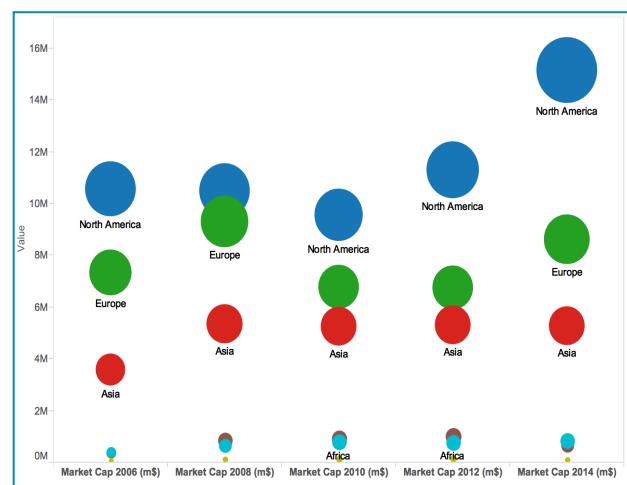
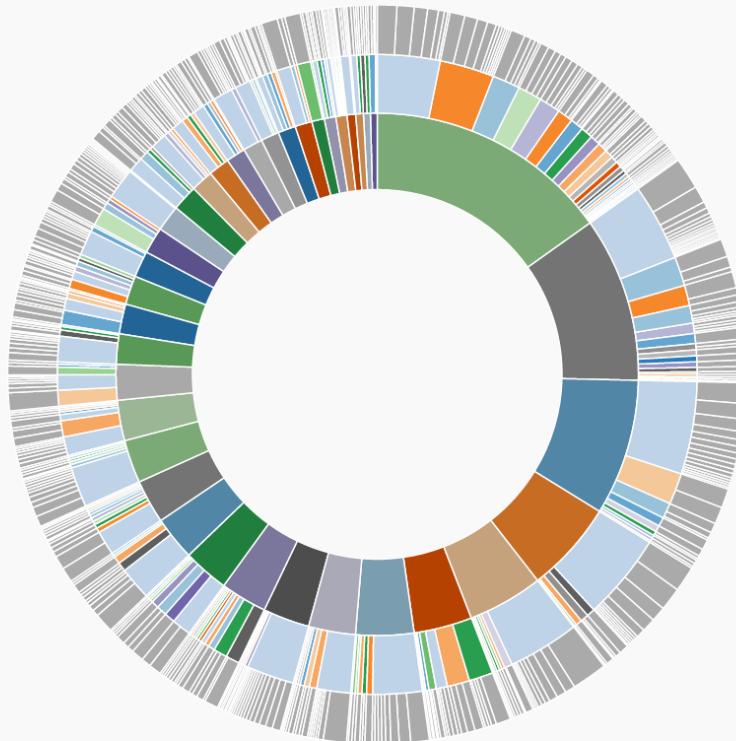


Chart 4.2.2 A Sunburst Visualisation Example

The second part is the Nasdaq Companies in a Proportional View, we put it into the third section of our web page, and the final layout effect is shown in Chart 4.2.3 and Chart 4.2.4. In this view we can see all the market capitalisations of all the Nasdaq Companies. (The data is extracted on Nov.4, 2014)



Data base on [Nasdaq Companies List](#).

Chart 4.2.3 - Nasdaq Markets in a proportional view



Chart 4.2.4 - Layer demonstration of the Nasdaq Proportion Visualisation. Users can hover the mouse on each section to see the relative data

In this view we can see the market capitalisations of all Nasdaq Companies as well as the sum up in each industry sector. (The data is extracted on Nov.4, 2014). Users can move the mouse over the chart to explore the data of each section. The sectors in inner circle represent the different industries in Nasdaq market, the sectors in middle circle represent the countries, while the smallest sectors on the outer circles represent the companies. The angle of each sector represent the total market capitalisation and relevant portion of this sector. Each Country has an distinct colour represent itself.

From this diagram we can answer the question that "What's Book Industry (or any other industry)`s total Market cap`s proportion in Nasdaq market or in the global market?" simply by hover the mouse on to the certain sector.

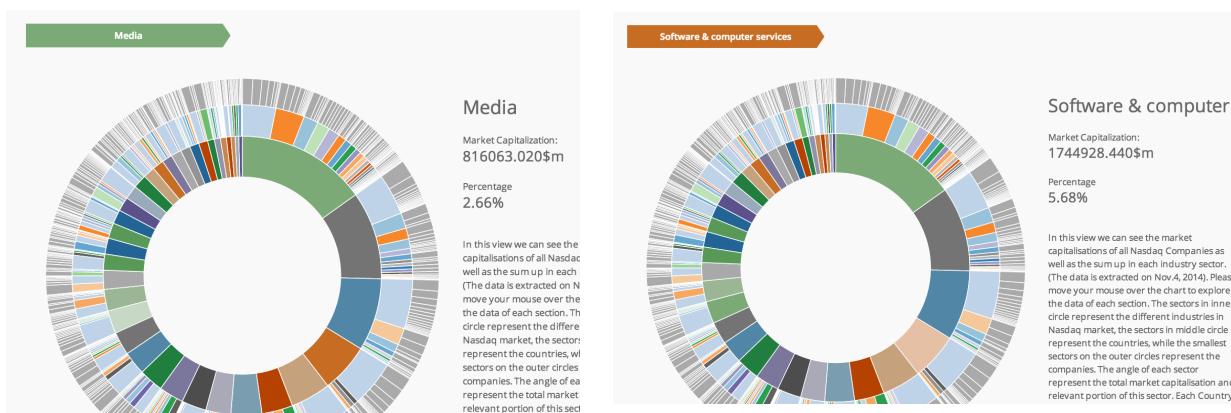


Chart 4.2.5 - Exploring Industry proportions examples

Within the industry sector ,the countries are arranged due to ranking of the total market capitalisation, therefore we can see the dominator in each section in this view. For instances, in Bank Area we figure out that China plays the second biggest role, while in Oil and Gas area, UK takes the second place(US is always the dominator as usual)

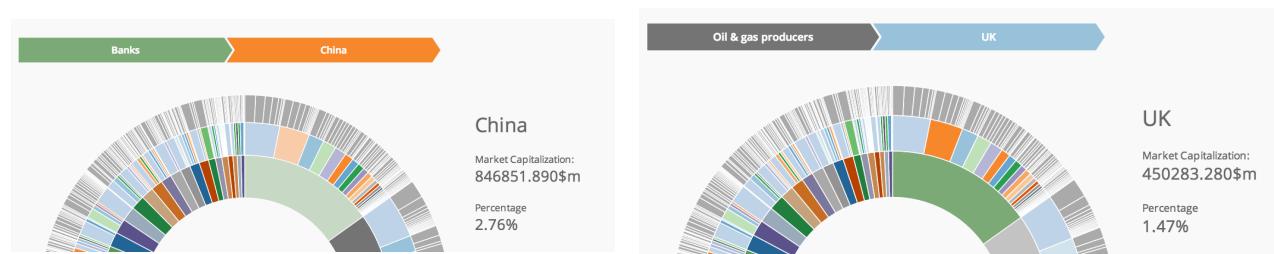


Chart 4.2.6 - Exploring Industry proportions examples

The third part is the Global Top 500 Countries in a Proportional View, we put it into the forth section of our web page, and the final layout effect is shown in Chart 4.2.7 and Chart 4.2.8. In this view we can see the market capitalisations of companies, countries as well as the continents together with a time axis scale.

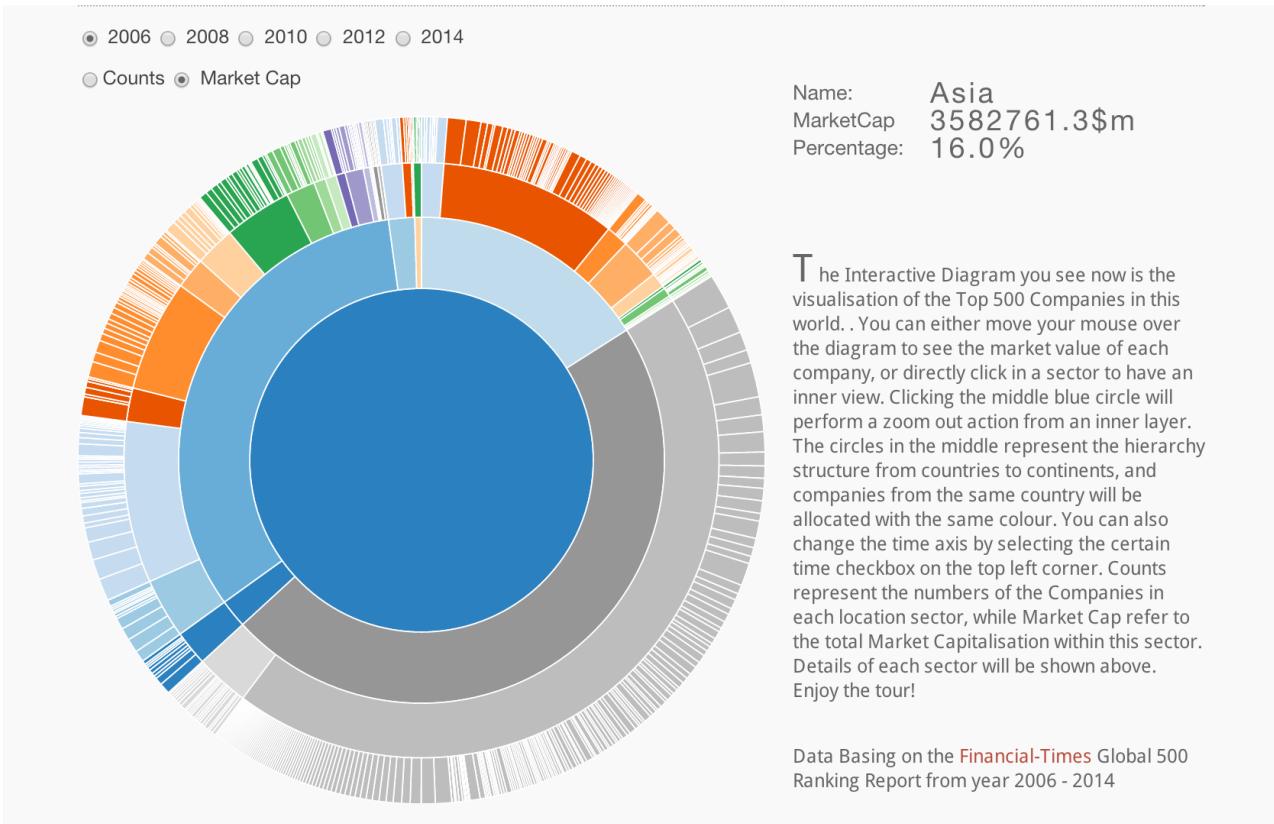


Chart 4.2.7 - Global Top 500 Companies in Proportional View

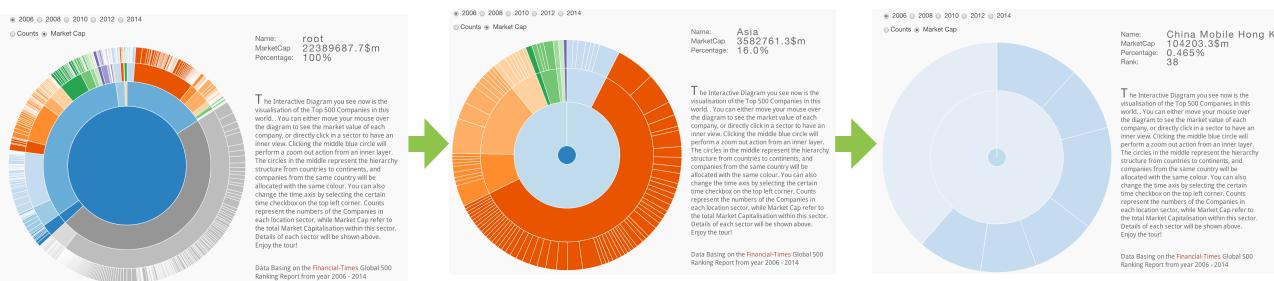


Chart 4.2.8 - Zoom in effect in Proportional View

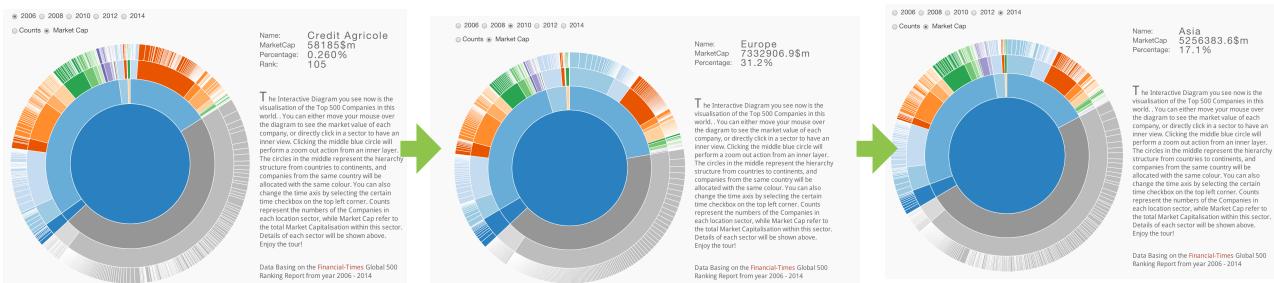


Chart 4.2.9 - Time changing in Proportional View

Note : Some companies are under the heading of more than two countries, in these case we choose the country that has the largest proportion of the company.

This Interactive Diagram is the visualisation of the Top 500 Companies in this world. To start with users should pick a year and a data type on the top left corner. Users can either move the mouse over the diagram to see the market value of each company, or directly click in a sector to have an inner view. Clicking the middle blue circle will perform a zoom out action from an inner layer. The circles in the middle represent the hierarchy structure from countries to continents, and companies from the same country will be allocated with the same colour. Users can also change the time axis by selecting the certain time checkbox on the top left corner. Counts represent the numbers of the Companies in each location sector, while Market Cap refer to the total Market Capitalisation within this sector. Details of each sector that users mouse hover on will be shown above the instruction text.

With this visualisation we can easily answer the following question.

1. What is the proportion of the Total Market Capitalisation of all Chinese (or any other countries) companies in Top 500 companies in certain year? What about proportion? Can we compare it with other countries in a more direct way?

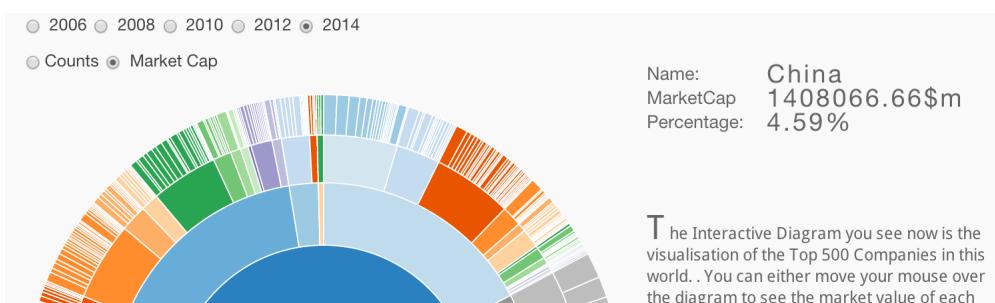


Chart 4.2.10 - Answer to Question 1

2. What is the above figures and diagrams looks like in the past years? Is the proportions of those figures getting larger or smaller over time?

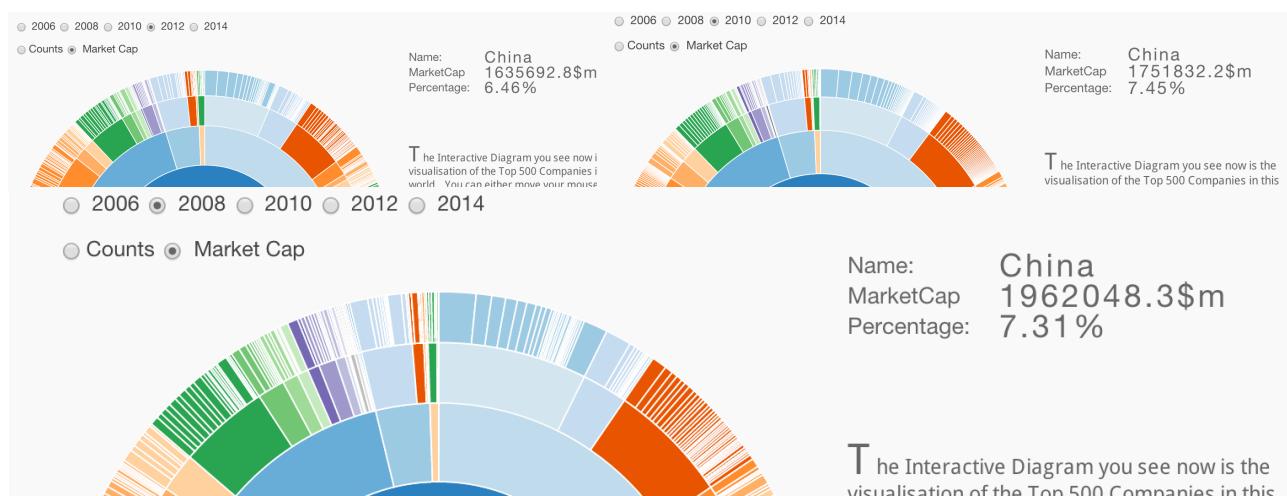


Chart 4.2.11 - Answer to Question 2

3. What is Apple's Market Proportion's in the Top global companies? Does it play an important role in US's proportion? How does it change over time?

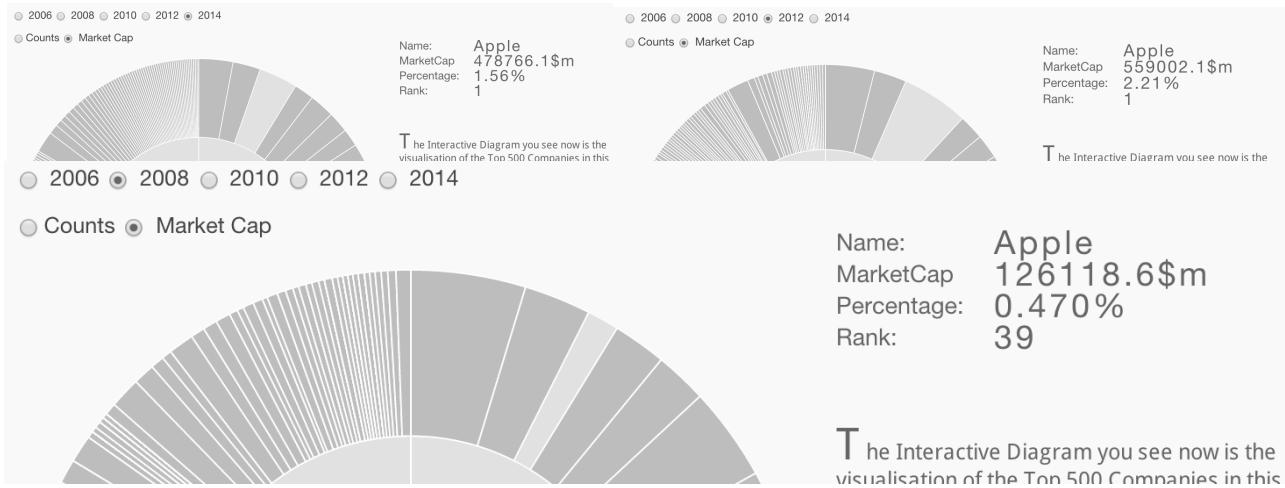
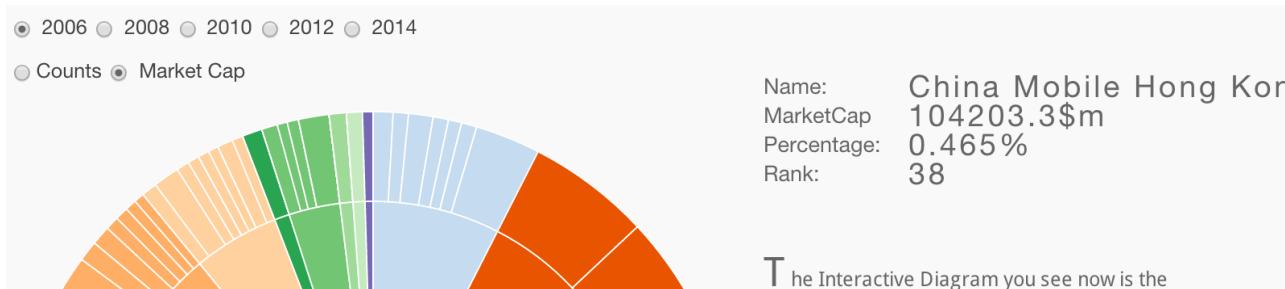
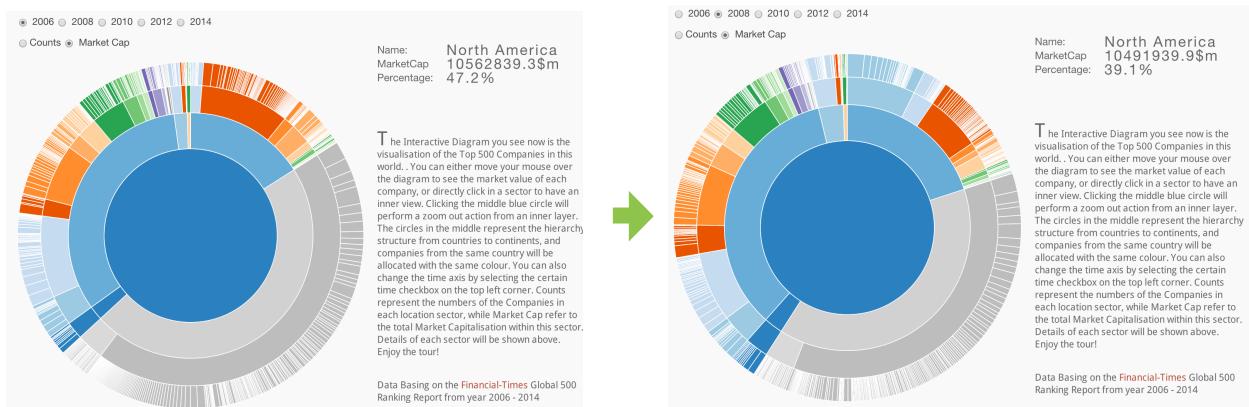


Chart 4.2.12 - Answer to Question 3

By further discovery we also find out that in 2006 there is still not any Chinese companies included in Financial Times Global 500. But there already some Hong kong companies have been put on the list including the China Mobile Hong Kong.



From 2006 to 2008 some of the US proportions have been taken away by Europe companies and Asian Companies, and China plays an important role in this. (First time been included into the ranking)



5 Remaining Tasks

Real time data retrieving

Up to now most of the data we acquire are old datas that obtain from the reports or table lists and been processed for a long time before visualisation. We hope in the future we can link to relevant database and achieve the goal that up-to-date data can be imported into our Visualisation Diagram from the Online Sources.

More clear and precise data source

After accessing the data that extracted from Nasdaq official website we find that total market capitalisation of the companies data is always up-to-date, while history figures are not available to quote and retrieve. Only the stock price and the trading volume of the website's history figures are open to the public. the datas are expensive , especially for the financially relative datas, they are are always been kept by companies or institutions and are not open to public. We plan to find datas that are more precise and clear in the future to ensure the data will not be distorted.

More data to be imported.

Temporarily we focus merely on the Financial Times Global 500, basing on the database limitation. We look forward to include more datas like the Forbes 2000, Fortune 500, or other history data of the Stock markets around the world. With these datasets, we can build a more precise and rich data visualisation, providing researchers and users an intuitional look on the global market in a proportional view together with time based axis.

Adjust the Colour Indicators

In the Zoomable visualisation of Financial Times Global 500 from year 2006 - 2014, the colour does not meet our ideal status. For example US is in colour grey while China is in blue, which is not the first colour impression of the countries. We plan to adjust the colour indicators

6 Summary

Through this Visualisation tour we explore and achieve some new visualising diagrams that are suitable for showing the hierarchy data together with a time changing axis. Moreover, we not only discover new insights through our diagrams, but also develop a new tool which is capable to find further novel and interesting facts. By visualising the Top 500 companies and the Nasdaq global markets in a proportional and an interactive view, we come up with a new way of looking the economic world.

7 Data Source

Nasdaq Official Website: <http://www.nasdaq.com/quotes/>

Financial Times Global 500 Reports 2014:

<http://www.ft.com/cms/s/0/988051be-fdee-11e3-bd0e-00144feab7de.html>

Financial Times Global 500 Reports 2012:

<http://www.ft.com/cms/s/0/988051be-fdee-11e3-bd0e-00144feab7de.html>

Financial Times Global 500 Reports 2010:

<http://www.ft.com/cms/s/0/988051be-fdee-11e3-bd0e-00144feab7de.html>

Financial Times Global 500 Reports 2008:

<http://www.ft.com/cms/s/0/988051be-fdee-11e3-bd0e-00144feab7de.html>

Financial Times Global 500 Reports 2006:

<http://www.ft.com/cms/s/0/988051be-fdee-11e3-bd0e-00144feab7de.html>

