COMP7507 Visualization and Visual Analytics

Project Report

Visualization on Financial Data of HK Listed Stocks

Group 24

Ng Kwong Hei, Eddie (2006401749)

Tong Jin Tao Virgil (2009246580)

Fang Fei, Karen (3035347873)

Foreword **4**

The Data **4**

Tools of the Trade **5**
- Data Acquisition 5
- Data Processing and Visualization 5
  - D3.js and jQuery 5
    - Advantages 5
    - Advantages 5
    - Disadvantages 5
    - Disadvantages 5

The Past: HSI Time Machine - How Regions and Industries Changed Over Time **6**
- Introduction 6
- Data Processing 6
- Design 7
- Final Visualization 8
- Analysis and Insights 8
- Limitations and Future Work 9

Now: HSI Current Constituents **9**
- Introduction 9
- Data Processing 9
- Design 10
- Final Visualization 10
- Analysis and Insights 10
  - Relationship to “The Past” 10
  - Difference between Price Returns and Total Returns 10
  - Huge Growth of the Information Technology Industry 11
  - Subtle Differences within an Industry 11
- Limitations and Future Work 11

The Future: Analytical Tool Based on Fundamentals **11**
- Introduction 11
- Data Processing 12
- Design 12
  - Design Considerations 12
  - Competing Designs 12
- Analysis and Insights 14
  - Explanations for Past Performances 14
  - Differences among Industries 14
  - Guidance for future investment decisions 15
- Limitations and Future Work 15
  - Knowledge Limitations 16
  - Data Limitations 16
  - Technical Limitations 16

The Extras: Hidden Network of Directors **16**
Introduction 16
Data Processing 17
Design 17
Final Visualization 18
Analysis and Insights 18
Limitations and Future Work 19

Contributions **19**

References **20**

Foreword

In our project proposal we suggested a number of intuitive visualizations for companies listed in Hang Seng Index (HSI), such that even users with limited financial knowledge would fully comprehend the data and grasp the insights. Our group's focus can be concretely described by a number of user stories.
In this report, we provide a brief recap of the steps our group has taken to answer these user requirements. The product is a suite of data visualization templates, featuring a number of techniques covered in the course materials, with a particular emphasis on user interactions.

The Data

Our data set covers the companies listed in the Hang Seng Index (HSI) from 2006 to 2016, for a total of 60 stocks[#].
For each stock, the data collected can be divided into the following categories:

1. Static data
   This comprises of the relatively static characteristics of a stock, for instance, the business characteristics, or the corporate governance aspects of a firm.

   - Company name
   - Domicile
   - Industry
   - List of directors (include all stocks listed in Hong Kong in order to better dissect the relationship between stocks in Hong Kong)

2. Price data
   Price data and other derived values fall into this category. To facilitate rendering of visual elements within web pages, we chose to compress the data into a monthly frequency (month-end values are being used).

   - Closing prices
   - Closing prices, adjusted for dividends and splits
   - Market capitalizations
   - Monthly returns
   - Yearly returns

3. Fundamental data
   Fundamentals are exclusively related to the operational characteristics of a company, for instance, its total assets and total liabilities, annual sales, and so on. For the purposes of this study the company fundamentals are transformed into six financial ratios:

   - Return on assets (ROA)
   - Return on equity (ROE)
   - Free cash flow to assets
   - Free cash flow to equity
   - Operating profit margin
   - Net profit margin

In order to give a holistic view of any individual stock, or the index as a whole, we would also need a way to efficiently query the data and use the resulting subset to populate the visualizations. A key observation is that individual stocks can be uniquely identified by the exchange symbol; for the case of prices and fundamentals, the unique key would be the stock symbol and the timestamp combined. Motivated by this observation, we have

chosen to store the data in a relational database-like manner. More specifically, the data set is broken down into a large number of small files organized into several folders, allowing our web portal to rapidly gather data across multiple data categories on user demand.

Tools of the Trade

Prototyping our visualization idea was both challenging and rewarding. Our team made good use of the mainstream web development toolkits (HTML, JavaScript, CSS) as well as back-end scripting languages (Python).

Data Acquisition

A major headwind for our project was that many types of financial data we looked for (with the possible exception of prices) were not readily bundled into downloadable datasets by third parties^. Instead the data was everywhere - including various finance websites targeting retail investors, plus a "database website" hosted by the famous investor David Webb. To complicate matters, Sina Finance - the website with the largest set of historical data available - only supports Chinese web pages. Combining these data sources manually is simply too time-consuming to even consider.

Fortunately, after digging deep into the structure of these websites, we found ways to automate the retrieval of numerical and text data from these websites, by writing a bundle of web-scraping scripts to intelligently make queries to the back-end servers of these websites, and sifting through the responses for raw data to capture and store. Python, with its excellent libraries for web page parsing and data manipulation, was ideal for this task, and with this valuable tool at hand we managed to create a comprehensive set of data in a short amount of time.

Data Processing and Visualization

After collecting an initial set of data we proceeded to build the visualization templates using the front-end technologies of choice, including HTML, JavaScript (D3.js and jQuery), and CSS. Due to the limited time we had, we skipped the development efforts for one of our visualizations, and instead used an open-source network analysis and visualization tool NodeXL to generate a network chart.

D3.js and jQuery

Comparing D3 with Tableau, the de facto standard for data visualization, we initially made the choice to move into D3 partly because Tableau is proprietary, and partly because we had little knowledge of both tools and would like to try out a tool that supports our preferred programming language, JavaScript.

After using D3 to create the charts for this project, we would argue that both tools have their merits and limitations, with D3 coming out slightly ahead.

Our Model (D3 and jQuery)	**Tableau**
Advantages ● D3 is open-source and has a large user community. ● Visualizations are fully customizable. ● Possible to create highly interactive displays of data.	Advantages ● Tableau has lots of built-in data visualization templates for users to choose from. ● Tableau supports numerous data sources such as Excel files, or even SQL databases.
Disadvantages ● Plain D3 has limited data manipulation capabilities. With other libraries (jQuery being the most prominent) this can be overcome. ● D3 only supports data retrieval from plain text files, reducing its usefulness for large data sets.	Disadvantages ● Tableau is proprietary. ● Data pre-processing is still needed. ● Some visual features are not customizable, especially user interactions.

^ A recurring theme is the lack of quality financial data from retail finance websites. Missing or mistyped values are frequent; for websites less prone to data impairments, one could easily identify questionable treatments of the fundamentals, such as incorrect reporting dates and extrapolation of data.

The Past: HSI Time Machine - How Regions and Industries Changed Over Time

Introduction

'The Past' aims to provide a holistic view of historical changes in HSI, helping users to learn how it has grown to the current state, as well as gain insight in the markets through event studies. We believe, learning about the past is beneficial for users to gain deeper understanding of the current index and further enable them to predict the future.

In particular, this visualization is designed to answer the following question:

Which group of stocks (grouped by region / industry) account for the performance of an index within a time period?

In answering this particular question, professional investors are given the chance to look into the historical performance of industry as well as region, which allow them to gain insight in what industry will be dominating in the future. Industry analysis is of crucial importance for professional investors in the sense that financial performance of companies are highly linked with the industries they belong to. One would not expect to beat the market by picking the strongest stock in a sunset industry. Apart from industry analysis, regional analysis on stocks pave way for a detailed analysis on the rising influence of Chinese market in Hong Kong.

Data Processing

Our visualization on the historical changes in HSI component grouped by region and industry are based on the market capitalization of the stock. Our take on picking market capitalization as our visualized data is that market capitalization provides a simplistic but meaningful measure on the company influence in the financial market. Solely picking market price or market transaction quantity would not be able to comprehensively reflect the company influence since market price is subject to manipulation through stock split and stock consolidation and market quantity merely portray the liquidity of the stock instead of the company value.

Since most of the financial website merely provide transactional data of stocks without historical changes in issued share, we have to resort to David Webb's webb-site.com for querying the issued share of the stocks within our scope.

The scope of stocks are selected based on the historical changes in HSI component in the past 10 years. The reason of limiting the period to be 10 years is that the market capitalization data in David Webb's webb-site.com are only complete in the past 10 years though it should be reminded that a few of the monthly data are not available for several relatively small stock that we have to conduct linear interpolation to make sure a smooth monthly transition.

The industry label is not a difficult decision for us given that they are clearly defined by the HSI into four main categories (Finance, Utilities, Properties and Commerce & Industry). Nevertheless, region label requires our judgement since there are no official definition from HSI. Our classification are based on the location of headquarter as well as business coverage of the stock since the location of headquarter is not sufficient to classify the stock into region when many companies might just move their headquarter to Hong Kong to enjoy tax benefit.

Design

There are three four dimensions of data involved in this visualization:

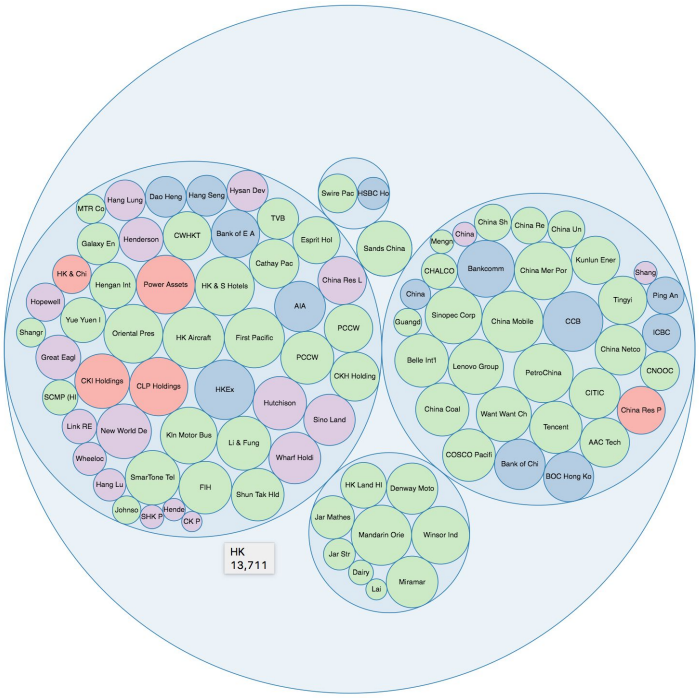
- Stock Region (Category)
- Stock Industry (Category)
- Time (Quantity)
- Stock Market Capitalization(Quantity)

We came up with the design based on the data types and how they are usually visualized:

- Hierarchical: Treemap. We decided to use Circular Treemap rather than normal Treemap because it's easier to compare sizes and the hierarchical structure is clearer.
- Temporal: Time Series or Animation. We chose animation to show more dimensions of data.
- Multivariate: Since one of the dimensions (time) has been removed by using animation, it become three-dimensional data and could be shown using colors, size and position.

Initially, we proposed a time-based animation of Circular Treemap (inspired by Circle Packing and Bubble Chart), with the outer bubbles showing regions, and inner bubbles (size based on market capitalization) inside showing individual stocks, color-grouped by industries. Such presentation formats allow us to observe how bubbles grow or shrink, and identify the trend on region or industry by the bubbles' groups or colors, throughout the time periods.

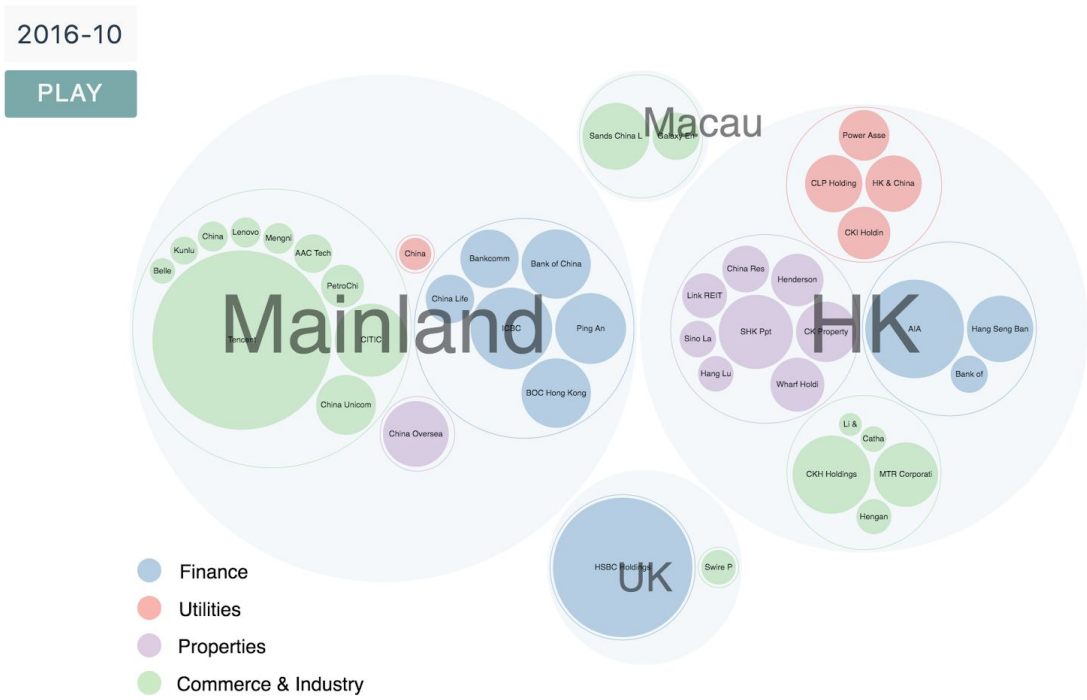
After implementing the initial visualization(as shown below), we found that since the position of bubbles change every time we update the market capitalization of stocks, it's hard to observe proportion of different industries (shown by color) and how they change overtime.



In the lecture, we learnt that group location with color is one of the most effective use of conjunctions of preattentive features. With that in mind, we decided to group the stocks of different industries together in each region. That way we can easily observe the proportion of industries by the groups and colors.

Also, we changed a couple of style details, including adding labels for regions and removing the outest circle, to improve the clarity of this visualization.

Final Visualization



This is the visualization of the most recent timestamp (October 2016). The full version of visualization (animated) can be viewed live here. It has been implemented using the Circle-packing layout probvided by D3.js, which can automatically calculate the position and size of the bubbles based on the input values. After careful design and multiple adjustments, the final visualization has the following great features:

1. Clear hierarchy structure of stocks using Circular Treemap.
2. Rich Interaction: Users can click on the 'Play' button to play and pause the animation. By selecting time using the drop down list, users can view the visualization of different timestamp and also change the start time of animation.
3. Easy to create insight through animation: Since the layout is calculated by value (market capitalization) and the largest bubble is always placed at certain position (on the left), every change of the layout indicates the change of proportion ranking in regions or industries and the group with highest value can always be easily observed at certain position.

Analysis and Insights

Group by Region

Hong Kong had undergone a transitional period in the nineties but curiously the "handover" from the British to the Chinese in terms of financial influence in Hong Kong market has not yet been completed.

In the beginning of the animation, it is shown that the total market capitalization of the British companies remained slightly higher than that of Chinese companies even though prior to the handover in 1997, several British companies have already been dropped off from the Hang Seng Index. The main support for British companies is HSBC, the pseudo-central bank of Hong Kong in the colonial period.

Nevertheless, it doesn't take long for Chinese companies to surpass British companies and even Hong Kong companies in terms of weighting in HSI when more Chinese financial institutions are accepted into HSI. From the trend as shown by the visualization, it is reasonable to predict that Chinese companies will continue to dominate Hong Kong financial market.

Group by Industry

Property companies has always dominated Hong Kong market in the past decades which fits the general depiction of Hong Kong economy. The industry distribution are fairly balanced in Hong Kong yet not in Mainland or other regions. This might be partly due to the fact that only the most reputable companies in other region will consider to list in Hong Kong instead of their own region, which is evidence by the UK region where the industry distribution are not heavily focus in finance yet our visualization only show financial company. Nonetheless, it might also reflect that particular region only focus in a particular industry, such as Macau where casino are thriving.

An interesting point should be made upon the observation that companies in the industry of Commerce and Industry took over companies in Financial industry - the main driver of this changes is the high-flyer stock - Tencent, whose growth rate has surprised many equity analyst. This might worth take note by professional investor as in bank and property companies no longer deem as a lucrative investment when their stocks have been stagnant for the past few years while stock in information technology sector provide a much higher return.

Limitations and Future Work

The main limitation we had on our visualization is insufficient of historical data for us to visualize the data in a much longer timeframe including the handover of Hong Kong. With this part of data, it would be much obvious to see how British companies left the Hong Kong market since prior to the handover, the British companies knew that they won't be able to survive in a market under the rule of Chinese market after taking advantage of the British colonial period. This might also allow us to visualize how the Hong Kong companies take control of the financial market during the period of power vacuum when British influence is dwindling and Chinese influence has not yet arrived.

In the future, our team would suggest to expand the scope from HSI component stock to all companies listed in the main market and classify stock into a more narrowly defined industry for a better illustration of which industry are performing better. A zoomable circle hierarchy visualization would be beneficial to achieve this goal.

Now: HSI Current Constituents

Introduction

Recall that HSI consists of stocks from various countries and industries. Oftentimes, industries as a whole may significantly outperform or underperform the index as a whole. (The same effects may also apply to the country classification, though the effects are typically less pronounced.)

For investors new to the Hong Kong stock market, they may be interested in knowing which sectors have outperformed in the past, and vice versa. This visualization is designed to answer exactly this question.

Data Processing

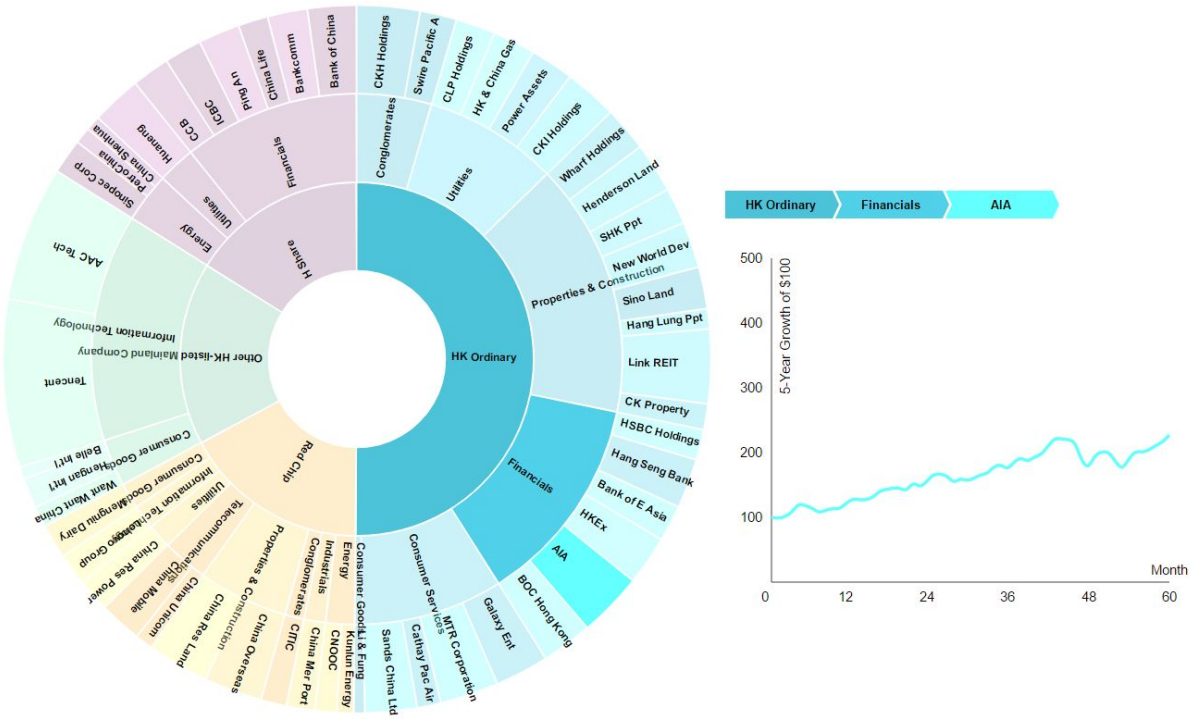
The adjusted prices of stocks are rebased to 100 on-the-fly, followed by real-time computation of the 5-year returns. The reason we require the return figures to be computed in the web page is because a single source of truth (the same data files) should be used to derive all the data visualizations, a principle Tableau also closely adhered to.

Design

The resulting visualization is a sunburst chart showing the portfolio hierarchy. When the user hovers above a category, a line chart next to the sunburst would be updated to display the selected category, together with historical price performance of the group.

An example of the final product is shown below. Breadcrumbs, a visual element typically used to display short text, is used to show categorical information. Another interesting feature is when the user displays a category other than individual stocks, the average total returns would be shown. This essentially allows the user to get a feeling for what would happen if the initial investment was spread evenly among the stocks within a group.

Final Visualization



Analysis and Insights

Relationship to "The Past"

The sunburst chart is similar to the circle packing chart in the previous section, weighing the individual stocks by 5-year growth of $100.

Difference between Price Returns and Total Returns

One surprise from the sunburst chart is that most of the companies and sectors have positive returns over the past 5 years. This is in stark contrast with the perceived returns of the individual stocks.

To see why this is the case, we pick *HK & China Gas* as an example. In October 2011, its closing price is $17.64, while in October 2016, its closing price is $15.20, representing a loss of $2.44 or 13.8% of the initial price. Yet the chart is displaying a rather steady upward trend. Is the use of adjusted prices introducing a flaw? In fact this can be immediately answered by the fact that *HK & China Gas* had paid 5 stock dividends during the past 5 years, each time amounting to exactly 10% of the outstanding shares, effectively scaling the stock price by a factor of $1.1^{-5} = 0.6244$. If investors are not careful enough and used raw stock prices in calculating returns, they would be

missing out on the 50-plus percent returns *HK & China Gas* had generated over 5 years, and deem it as a stock not worth investing in.

In fact this is a very common mistake even among seasoned investors - the anchoring effect of the day-to-day price fluctuations is so strong that the positive effects dividends and stock splits are ignored. This is also the reason why the vertical axis of the price chart would have a title of "Growth of $100", to lower the chance of misinterpretations by users.

Huge Growth of the Information Technology Industry

Two stocks that clearly stand out in the sunburst chart are *Tencent* and *AAC Tech*, each recording a mouth-watering gain by growing an initial investment of $100 to around $500 over the past 5 years. While past performance does not equal future success, we shall attempt to provide an explanation in the next visualization.

Subtle Differences within an Industry

Next we turn our attention to the H-share Energy industry. The industry as a whole is one of the laggards - recording zero growth in 5 years - but one stock in the group, *Sinopec Corp*, is significantly outperforming the other two. The underlying reason is because the main business of *Sinopec Corp* is oil refining, which benefited significantly from the oil price crash during the period, while the other two are energy producers and were badly hurt. This shows that the information provided by the visualization is insufficient, but at least it provides a subtle clue, prompting users to investigate further.

Limitations and Future Work

For this chart the sole emphasis is on user interactions. We hope the user would rapidly discover insights by navigating the whole HSI universe using the sunburst chart, getting a feel for the different industries and stocks in the process.
During the initial planning stage, we spent a significant amount of time discussing an important interaction element, of which we did not have sufficient time to implement - the ability for the user to freeze the display of the line chart, such that the user can select multiple stocks and industries and compare their total returns. This interesting feature is certainly high up in the list of future work.

The Future: Analytical Tool Based on Fundamentals

Introduction

Having created visual plots summarizing the past and the current status of the market, the next milestone to achieve is to demonstrate ways to facilitate investment decisions. Specifically, we aim to show, in a single view, the fundamental strength of a company plus its recent stock price performance.

In theory, it is possible to go as far as creating specialized visualizations for trading strategies. There are many trading strategies that make use of prices and fundamentals, leading to numerous possibilities of creating visualization charts for them. The problem with this approach is that, since the sole purpose of the visualization is to speed up the decision-making process, the choice of visual elements would be highly subjective.

Alternatively, visualizations could be structured to present general information about the stocks that users are considering an investment in. The problem with this approach is that users would need to find their own ways of interpreting the visualizations, since the charts would present information instead of subjective opinions.

Confronted with this trade-off, our team had taken the middle road - we carefully picked a handful of financial metrics that are applicable to all companies in the universe, and injected a subjective element by allowing the user to configure the visualization to only display their stocks and time periods of choice.

Data Processing

Initially the fundamentals are collected from various websites. However, having only these figures would make comparison among companies extremely biased - for instance, the fundamentals may be reported in different currencies; or the business characteristics in different industries can be diverse. To facilitate data visualization and analysis, the data is first preprocessed into last-twelve-month (LTM) figures; afterwards, financial ratios are calculated from the raw figures, such that currency and industry impacts are minimized.

Design

Design Considerations

As one would expect, the many financial ratios must be presented in the same chart space; how price performance would integrate into the visualization becomes the key. Besides, the message conveyed by the visualization should be forward-looking: with the information presented in the chart, users should be able to form an opinion about the prospects of stock prices and formulate their trading plans. An important consideration is whether the visualization accurately reflects proven or widely accepted investment principles. If conflicting information is present, the user would easily be misled rather than helped by the visualization.

Competing Designs

The following lists some of the designs our team had considered implementing, and the pros and cons of implementing them.

- Line chart
  - Advantages
    - *Efficient use of screen space*: for each time series, since every time point is displayed in the chart, a single line chart can represent a large number of data points.
    - *Intuitive*: users are well-trained to use line charts.
    - *Useful for showing trends*: Line graphs are ideal for showing trends over time.
    - *Flexible*: can be extended with text annotations. A common scenario is to annotate the time points and add explanations for significant events.
  - Disadvantages
    - *Might cause confusion to user*: users have a tendency to compare the lines shown in the same chart. However, the different financial ratios for the same stock are obviously not comparable.
    - *Visually unappealing*: as the number of financial ratios increase, the chart would be overloaded with information. Plus the lines may cross each other (though the intersections have little meaning).
    - *Difficult to incorporate price performance into the chart*: if price is shown as a separate time series, the graph would become very confusing to users.

- Radar chart
  - Advantages
    - *Intuitive*: data values can be easily read since every data point lines on one axis. Moreover, the area enclosed by the data points in the radar chart would be closely corresponding to the stock's fundamental strength. The assumption here is that the financial ratios (the axes) are appropriately selected.
    - *Interactive*: by allowing the user to select the stocks and periods to display, the single radar chart can be extended to perform comparisons among stocks, or for a single stock across multiple time periods.
    - *Flexible*: it is possible to add animations to the radar chart, to demonstrate the evolution of fundamentals for a single stock.

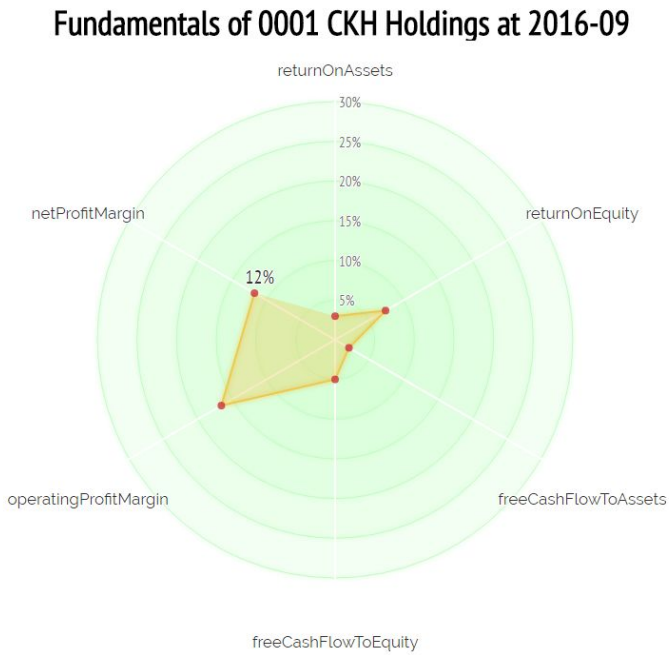
- Disadvantages
  - *Inefficient use of screen space*: a radar chart cannot display too many records, otherwise the chart would become cluttered. Because both the central and peripheral parts of the chart would usually be empty, the use of screen space is rather inefficient.
  - *Less popular among users*: users are less familiar with radar charts as a means of visualization.
  - *Cannot visualize trends*: since the time axis is not present in a radar chart, trends of individual financial ratios over time cannot be shown.

- Scatter chart

  A third design worth mentioning is the scatter chart, where the financial ratios become the axes and individual stocks are represented as data points. Unfortunately it failed to meet the requirements of displaying high-dimensional data. We still list out the advantages and disadvantages we have identified for this type of visualization.
  - Advantages
    - *Intuitive*: the fundamental metrics are represented by the location of the data points, achieving pre-attentive salience.
    - *Useful for comparison across stocks/time*: a scatter plot can show a large number of records at once.
    - *Flexible*: the scatter chart can be greatly enhanced by adding user interactions, such as selecting stocks by highlighting a region in the plot, or showing trends by charting the historical records alongside the most recent entries for a single stock.
  - Disadvantages
    - *Not scalable*: for instance, a 2-D scatter plot can only represent at most two financial ratios.

Our team eventually geared towards implementing the static radar chart, for its advantages significantly outweigh the drawbacks.

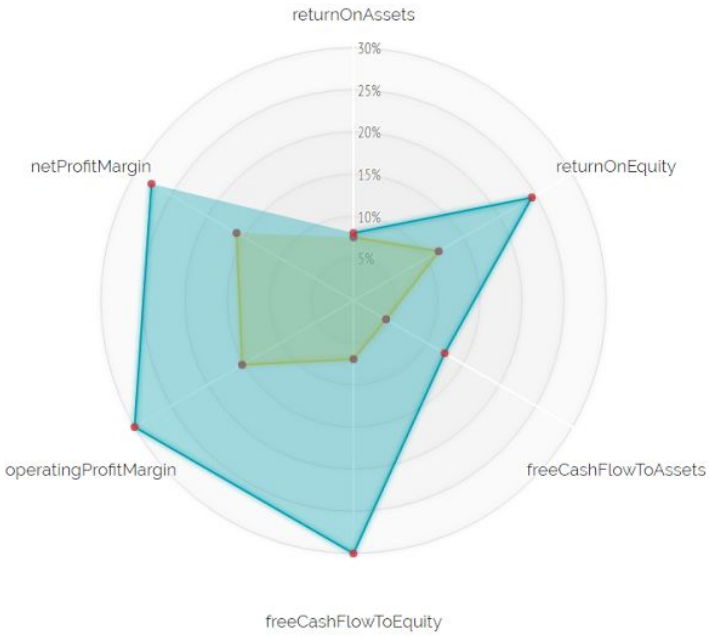
<u>Final Visualization</u>



Analysis and Insights

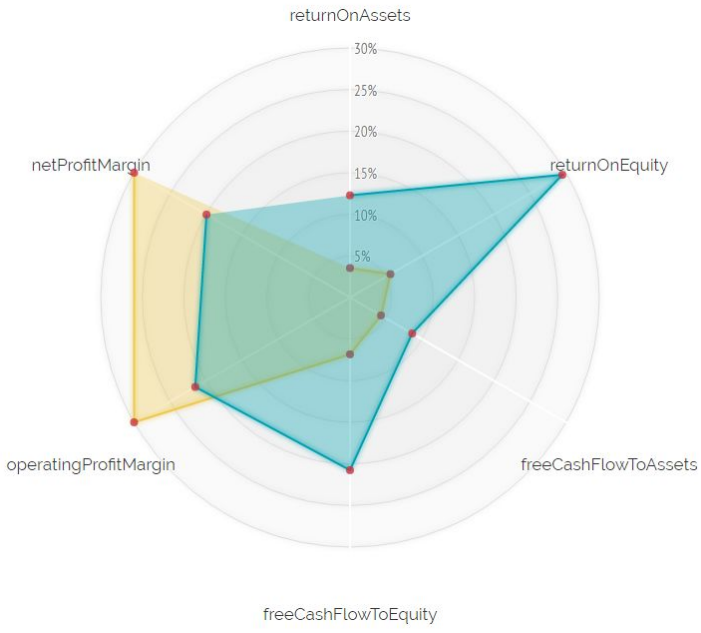
Explanations for Past Performances

Tencent stands out in terms of fundamental strength. Compare this to a laggard stock, *China Mobile* - the area enclosed by the financial ratios is much larger. This may account for the strong outperformance in stock prices.



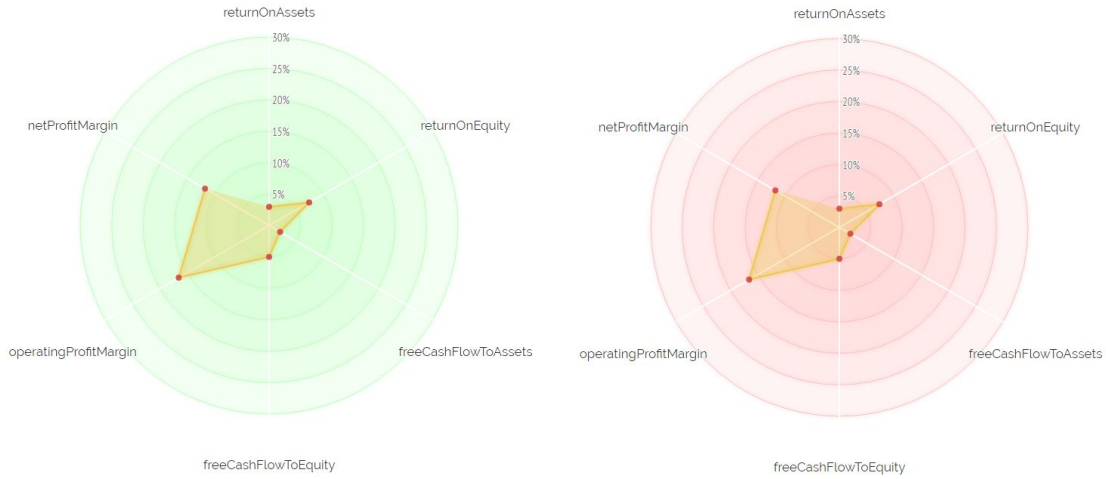
Differences among Industries

An example is to compare *China Mer Port*, a bellwether in the shipping ports industry, against *Sands China*, a blue chip casino operator. Apparently the ports enjoy very high profit margins (compared to the casinos), but the return on assets is low. Here one can easily reason that the ports industry is asset-intensive (again, compared to the casinos).



Guidance for future investment decisions

One might wonder what role price returns would play in fundamental analysis. In fact, its importance cannot be overstated - the price momentum effect in stocks was well-documented in literature and studied by investment professionals. While our data set is probably too small to verify the momentum effect in the HSI, we have developed the visualization in alignment with the philosophy documented in those papers, using the 12-month lagged return to determine the background color of the chart. For users who are aware of the momentum effects, this visual hint could be helpful in their decision-making process.



Limitations and Future Work

In contrast to the ambitious title of "The Future", this visualization is probably a baby step - more of a proof of concept than a fully-fledged consumer product. As investors gain more sophistication, however, they would be less impressed by what little insights the retail investment websites have to offer. We believe most users would benefit from having visual aids when studying their investment targets, creating demand for visualization tools like ours. This is a huge motivation for us to keep improving. The following aspects have been identified.

Knowledge Limitations

Historically, many trading strategies became fashionable only to go out of favour as time passed. In the same manner, no single visualization would appeal to all of the diverse user preferences. For instance, technical analysis is highly recommended by some market practitioners, while at the same time criticized by others who would favour fundamental analysis. In our opinion, retail investment website owners have avoided making this trade-off - the websites are structured to contain large amounts of price and fundamental information - to attract the most users. In contrast, both academia and investment management companies have produced lots of quality research, which the general public is largely unaware of. Our tool has demonstrated that certain investment ideas can be visualized in intuitive ways; there is a long road ahead to translate other little-known investment philosophies into useful visual tools.

Data Limitations

High quality price and fundamental data are both hard to obtain. Most retail investment websites do not provide data beyond five years, effectively forbidding the calculation of many useful financial metrics. For instance it is impossible to compute the 5-year earnings growth rate (%) that is widely used in academic research.
Moreover, given the HSI data set, some of the visualization experiments simply cannot be carried out. One example is the ranking approach popularized by Greenblatt[4], who clearly stated that the algorithm does not apply to financial companies and utilities. As a third of the constituents in HSI are financials and utilities, any attempts to replicate Greenblatt's strategy for the HSI would be severely biased.

Technical Limitations

Due to tight time constraints, we only managed to enhance the radar chart to display multiple records, without little time left to add a legend. Plans for adding more user interactions were also scrapped. Priorities would be on developing a fix for these issues, in addition to fine-tuning the visual appearances of the chart.

The Extras: Hidden Network of Directors

Introduction

In this section, we would like to unearth the relationship between stocks in another aspect - the interpersonal relationship between directors. Our assumption on such analysis is that stocks are essentially the business of people and it would be an useful angle for professional investor to dissect the relationship between stock in terms of their director's social circle. In essence, we would like to seek the answer for the following particular question:

Whether directorship of board member in listed companies highly concentrated?

In answer the above question, we are interested to know the answer to a broader question every Hong Kong citizen would like to know - if the listed companies are rigged by the hands of a few billionaires. It is often rumored that there are a group of so-called "zhuangjia" manipulating the financial stock market into their own financial gain. Although the visualization of the network of directors in listed companies are unable to uncover their under-table dealing, it gives us an opportunities to visualize the sign of their intricate network of directorship in listed stock.

To further clarify, the data presented in this section do not suggest that the existence of social circle indicate an intimate relationship between these directors, it just merely demonstrate these directors should have been at least know each other and our team would like to see whether there are a high concentration of individuals in a significantly large quantity of stock.

Data Processing

The directorship of board member in listed companies are scrapped from David Webb's webb-site.com since it contained the full record of the directors in one centralized website with a uniform format that allows us to obtain the information without prasing the information one by one.

It is reminded that the stock range is not only comprised of those in HSI but also all companies that are listed in the main board in order to provide a more comprehensive view on the listed companies instead of merely focus on the few symbolic companies in Hong Kong. By including the non-HSI companies, the contrast between social group of these elite groups in Hong Kong will be more significant in our visualization.

To visualize data in the format of a network graph, our team has decided to put all the companies and directors into nodes with a label to indicate whether they are companies or director. Weighting on edges are defined to be the position they held in the companies. Obviously the edge weighting for the position director held are proportional to the seniority of that particular position, for example, the chairman in the companies would be the highest with independent director rank relatively lower.

Furthermore, we have filtered the number of nodes by restricting the frequency of nodes in edges to be more than 6 in order to remove the noise in the data. Those with only one or two position in listed companies might not provide sufficient insight for us to dissect the social circle of these listed companies' director.

Design

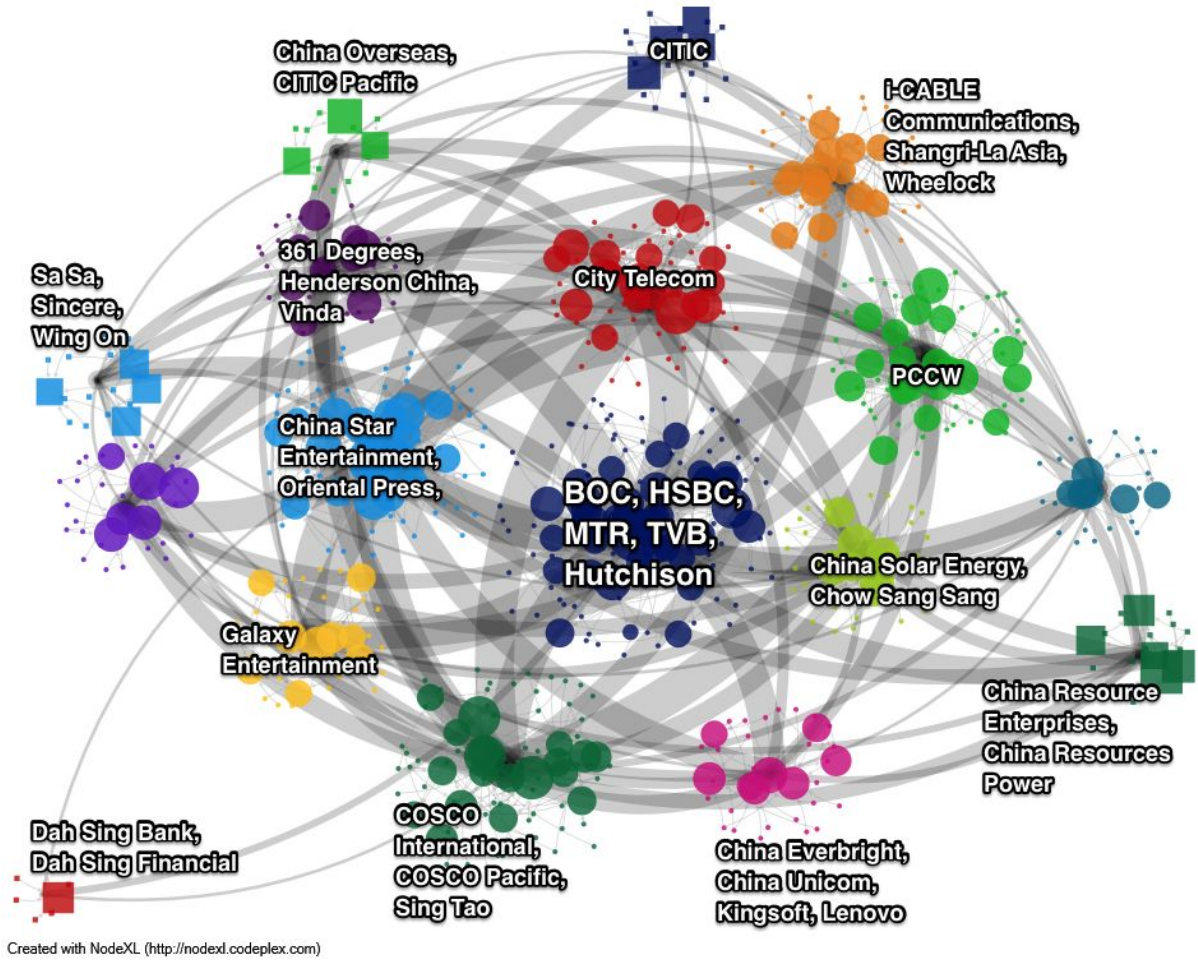
In the beginning, Heatmap was the visualization technique of choice rather than Node-link Diagram, because there are two types of nodes in this data set (directors and companies), which we considered ambiguous to be put in one Node-Link Diagram. However, soon we realized Heatmap is lack of capability of effectively visualizing large amount of data (over 10,000 items involved).

After investigating various network visualizations, we came up with a original design of visualizing this huge network of stocks and directors: A size-by-in-degree Directed Node-Link Diagram, using both directors and companies as vertices, and companies -> directors relationship as edges. The key thing about this design is that as there's no edge pointing from directors to companies, the in-degree of companies would always be 0. Then we calculate the vertex size (from 1 to 100) based on the in-degree of each node, thus the size of the nodes of companies is fixed to 1, which can be easily differentiated from the nodes representing directors.

Different from the other visualizations in our project, this graph has been conducted by NodeXL, as not only does it provide different kinds of network visualization layouts, but also it has features like computing graph metrics and cluster detection, that allows us to perform in depth analysis on the network.

In order to visualize the inner relationships among large amounts of companies and their directors, we applied the following steps to make use of different visual elements: 1. Add tooltips to vertices with company or person name; 2. Adjust vertice size using its in-degree; 3. Group and color the vertices by cluster using Clauset-Newman-Moore algorithm, which is usually used for finding community structure in very large networks; 4. Use Fruchterman-Reingold algorithm, a force-directed layout algorithm to draw the graph in an aesthetically pleasing way, with as few crossing edges as possible; 5. In layout options, we chose to lay out each of the graph's groups in its own box and combine the edges between clusters together, so that the clusters would be clearer and the graph would be more readable.

Final Visualization



Analysis and Insights

The above network graph provide an alternative way for us to look into the question we mention in our introduction. Although we do not see one large cluster of directors, it is worthwhile to note that the directors in listed companies are more or less connected in the sense that there are no cluster being isolated at all. These companies are often connected by an independent directors who are either accountants or lawyers. It is suspected that these clusters of directors are often associated with the lawyer firm and accountant firms.

The dark blue colored cluster of directors in the middle are often directors for the more reputable companies such as HSBC, BOC, Cathay Pacific and TVB. This might be the core elite social circle of Hong Kong, which includes Li Ka Shing, the billionaire often satirized as the man who controls Hong Kong, and his two beloved sons.

The clustering on the above visualization from the data of directorship also coincide with the general understanding of social circle. Oriental Daily and China Star Entertainment are grouped into the same cluster, which fit the public impression on the business owners of these two stock. The visualization also successfully categorized the group of Dah Sing type of stocks and China Resources type of stocks.

Limitations and Future Work

Data Limitations

The limitation on the above network graph lies with the data available to us. It is relatively challenge to visualize the true hidden layer of directors social circle given that they do not have to appoint people who are in their own circle.

As mentioned in the above insight, we have pointed out that the cluster groups are linked with mostly independent director who are lawyer and accountant. Nevertheless, we have not visualized this part of feature since the information required is significant and we do not have sufficient time to label each of the independent directors.

Visualization Tools Limitations
As stated in the above part, our team has resorted to NodeXL instead of D3 due to the huge amount of data available to us. Furthermore, NodeXL provide an embedded function for us to automatically generate the modality and in-degree for our network graph, which ease our pressure in visualizing the graph in a meaningful approach.

Future work
As mentioned in the above, our team would add their respective lawyer firm and accountant firm as well as the identity of these independent directors to further strengthen our hypothesis that these social circles are often connected by these professional entities.

In the future, we would seek to implement the visualization in D3.js (with the computed data of metrics and clusters by NodeXL) so that user can interact with the data and clearly show each of the node with detailed information for their own analysis. This way, our project can be completely interactive, easy-to-share and open source to the public.

Moreover, we would like to link up the stock performance of the cluster group to see whether there exist a correlation of the stock performance in the cluster group. This provide an extra perspective in analysing the stock other than industry analysis. For a small market like Hong Kong, it is relatively easier for board of directors to manipulate their stock performance by various financial strategies.

Contributions

		Eddie	Karen	Virgil
Proposal		✓	✓	✓
Data Collection and Processing	The Past			✓
	Now	✓		✓
	The Future	✓		
	The Extras			✓
Visualization and Analysis	The Past		✓	✓
	Now	✓	✓	
	The Future	✓	✓	
	The Extra		✓	✓

Web Page		✓	
Report	✓	✓	✓
Presentation	✓	✓	✓

References

1. N. Jegadeesh and S. Titman (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. http://www.bauer.uh.edu/rsusmel/phd/jegadeesh-titman93.pdf. Accessed: 18 November 2016.
2. N. Jegadeesh and S. Titman (1999). Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. http://www.nber.org/papers/w7159.pdf. Accessed: 18 November 2016.
3. C. S. Asness, T. J. Moskowitz and L. E. Pedersen (2013). Value and Momentum Everywhere. http://pages.stern.nyu.edu/~lpederse/papers/ValMomEverywhere.pdf. Accessed: 18 November 2016.
4. Joel Greenblatt (2005). The Little Book that Beats the Market, 1st Ed. Wiley Press.