Process Book for Stock Data Visulization*

by Jian Lan && Jie Ma

1 Basic Info

Project Repository. https://github.com/lanjiann/stock_data_visualization

Group Members.

- Jian Lan
  - UID: u0752670
  - EMAIL: lanjiann1985@gmail.com / jiann.lan@utah.edu
  - Degree: Master candidate in Computer Science
- Jie Ma
  - UID: u0748120
  - EMAIL: jiema1989@gmail.com / ma@math.utah.edu
  - Degree: PHD candidate in Mathematics

2 Overview and Motivation

In the financial system funds flow from those who have surplus funds to those who have a shortage of funds, either by direct, market-based financing or by indirect, bank-based finance. The former British Prime Minister William Gladstone expressed the importance of finance for the economy in 1858 as follows: "*Finance is, as it were, the stomach of the country, from which all the other organs take their tone.*"[1]

The **stock market** is one of the most importance component of the financial system. It's an important indicator of the economy, and when people use it to make financial decision, it also has an intensive reaction to the economy. Thus, analysis to the stock market is a valuable work. Every day there is a huge amount of data generated by the stock market, and it's nearly impossible to read out the useful information from the data directly for most of the people (if you are statistician, this is another story). We'd better find out a more intuitive way to present the information hidden behind the data. Visualization is our choice.

3 Related Work

When searching for project ideas, we find a lot of fascinating visualization works about stock data visualization:

- D3 Show Reel

*. This document has been written using the GNU TEXMACS text editor (see `www.texmacs.org`).

This visualization shows the data of four stocks, and it includes seamless transitions between the visualization types of lines, horizons, areas, stacked areas, streamgraph, overlapping areas, grouped bars, stacked bars, bars, and donuts.

Admittedly, these aren't the most useful visualizations. It is automatic, and it can be controlled by the users. However, it give us an idea of showing stock data in multiple ways, which help to users to inspect the data in different aspects.

- **TenchanJS**

  A visual, stock charting and technical analysis library built on D3. We like its indicators, and plan to add similar indicators to our visualization work.

- Bloomberg Stock Visualization

  With the inspiration of this visualization, we decide to create a table in the thrid part and display the Financial Statements Infomation, which will interact with the second part.

- D3-stock

  Stock prices visualization using d3.js

  See, all in one chart, with multiple time spans (1wk, 1mth, 6mths, 1yr, 5yr):

  - actual stock prices
  - a mobile average
  - trading volumes

4 Questions

The primary questions that we want to answer through our visualization:

- How do the stock prices change against time (use the stock price data only)?
- What is the actual performance of companies (also use stock related information)?
  - Which stock(s) outperform(s) in the past?
  - Which stock(s) are undervalued in the past and should be recommended for buying to investors?
  - Which ones should investors avoid buying even though they shined in the past?

5 Data

5.1 Data Source

In the proposal we said we would choose one data set from two candidate data sets:

- The A share[1] of the Chinese stock market.

1. In finance an **A share** is a designation for a "class" of common or preferred stock that typically has weakened voting rights or other benefits compared to B share. The equity structure, or how many types of shares are offered, is determined by the corporate charter. *A share* is also a way of pricing sales charges (loads) on mutual funds in the United States.

- The American stock market.

We finally choose the American stock market. We use the data of S&P 500 companies. Our project need two types of data:

1. The whole market index (S&P 500 index) and the daily stock price of every S&P 500 companies. We download these data from Yahoo Finance.

2. The financial statement of the S&P 500 companies. We download these data through the Bloomberg Professional Workstation in Marriott Library of our university.

We downloaded our data from BloomBerg Professional Machine in our Marriott Library.

5.2 Data Processing

All the data we download are in xls/xlsx format, and which can be dumped to csv format. D3 has method to read csv file directly. No complicated data processing.

- S&P Index: We have about 26 years S&P Index data in a `.csv` file. No pre-process required.
- Stock Prices: Each company has a `.csv` file. No pre-process required.
- Financial Statements: The original dataset we download is a big file that contains the entire information for all the 505 companies of 5 years. It is a `.csv` file but not in the form we require. A script is written to process and split this big file into individual `.csv` files, which are formatted well to be used directly. This is also not complicated.

When we were trying to implement our project, we notice that the two data sets, daily stock prices data set and financial statements data set doesn't include identical stock tickers. Since most tickers are in common, we decided to eliminate non-common tickers, and 411 tickers left. Though the this final data set is perfectly complete, it is also representative.

6 Exploratory Data Analysis

Two aspects we should consider:

- S&P Index and Stock Prices (technical analysis[2]): The stock price data contains two parts: (1) the stock price for the entire market and we use S&P 500 Index as benchmark; and (2) the stock price for individial stock price, for each individual stock that is in the basket of S&P 500.
- Data from Financial Statements (fundamental analysis[3]): There numerous statistics we can use to compare between different companies. For example, we can compare ROE, EPS, etc across different companies. These values are either directly included in the documents, or will be calcluated by existing formula. Specifically, by Financial Statements, we mean the company ratios statements, which contains all the common size data and a variety of useful ratios. The reason we choose to use this set of data is that ratios can be compared between companies more easily since they are all normalized. For example, profit margin is the ratio of net income to revenue. By comparing this value between selected companies, we have a clear view about which company's operating capability is more efficient.

2. Investors will dierctly look at the price of the stock and its trend to determine whether or not, and when he should buy or sell the stock. The validity of this method is based on 'efficient market hypothesis', which says that all the information of the underlying company of the stock is 100% percent encoded in the price of its stock.

3. People who use this method will not look at the stock price directly; instead, they examine the financial statements of the company to determine it performance. By financial statements, we basically mean the following three tables: Income statement table, Balance Sheet Table and Cash Flow Table. The validity of this method is based on the hypothesis that the market is inefficient, that it is unable to fully reflect the performance of the underlying company. A direct implication of this hypothesis is that many stocks are either underevaluated or overperformed by only looking at their price.

These two types of data analysis are usually combined together in a presentation to provide an overview for different aspects. This is also what we have done in our project.

Since both these two types of data are time series. Line chars of different time scales and price scales are used to initially look at our data. That is, we constructed lines that represent the trend of stock prices over a certain period of time. Specifically, we constructed two separate charts: the first of which is the trend of the index, S&P 500 index; The S&P 500 index is a basket of stock prices, and it calculates the average stock prices for all the stocks that are in the basket. The average performance will indicate the overall performance of all these 500 stocks.

Next, in another chart, we created a similar line charts for whichever stock is selected by clicking on the selection menu, upon selection. For example, if Google is selected (ticker `goog`), then the stock price of it will show up in this panel. Once it shows up, the trend of it is clear and it can be compared to the S&P index panel above it. If the the trend looks pretty much the same as the index chart, then we would say that Google's performance is good and its stock is worthwhile of buying, since it is least not worse than average; However, if it outperformed the index for most of the time interval, that is, its stock price goes up constantly and the slope beats the index by a great amount, then we have to be careful about whether it stock is really worth of buying, since from what people have learned from the past is that if a stock outperformed the market average by so much for a period of time, then its performance will definitely drop to a low level such that there might be a signifiant decline in its price.

Th insights we gain from the index chart is that there are two main recessions that occur in the year of 2000 and 2008; it took almost 3 years and only 1 year for the full market to get out of the bull market. From the individual stock price chart, if we choose AAPL (Apple company) for example, then we see that initially, around the year 2000, it performance is very normal, since compated to the index market, its stock prices has limited increases only limitedly; However, it is able to keep up with the trend of the index for the entire period, and even when the entire market is in recession, its stock price for most of the time keeps increasing or decreasing but less rapidly than the market. Based these observation, we reserve no determination in recommending buying APPL's stock, since it is able to perform better than average even under unfavorable conditions.

Financial statement related values are different. They are not daily, which means they don't have many data points, then we think bar charts is better than lines for them.

7 Design Evolution

7.1 Visualization Design (Original Design, Design 1)

We are inspired by many examples demonstrated in class and examples outside of class on the Internet — stock data presentation and visualization is so popular that we can learn a lot from the exist examples, and then try to create better ones. The design sketches are at the end of this proposal file.

1. Show the minimap of the whole market. Use brush to select time interval.
2. According to the selected time interval, present the closing price with Line charts, Areas Charts, Stack Areas, Overlapping Areas, Stacked Bars. Switch between them by clicking drop-down menu. At most two stocks can be chosen and shown in this part.
   - If only one stock is chosen, it is trivial. If two stocks are chosen (these two stock can be two stocks in the same time, or the same stock in two different time), they can be presented in the same figure or in separate ones. You can also select to merge two figures or re-separate them.
   - In each figure we show a tiny indicator, which can be dragged by the mouse cursor, and it extends with dash lines to the time axis to show the exact time of the price. It also has tooltips.

3. To show Financial Statements related visualization, we create two drop-own menus, one to control the stock(s) (e.g: Apple, Google and Amazon) the user wants to compare, and the other drop-own menu controls the attributes the user wants to compare (e.g: Revenue, ROE, etc).

   – We will create an individual plot for the attribute the user wishes to compare between companies, and then we put all the selected companies in this figure such that the selected attribute of all the selected companies can be compared directly. We are inspired by the idea of Winner of 2015: New York City Schools.

   – Based on the last paragraph, we do not need to control the number of stocks and attributes we want to compare.

4. Compute a recommendation index for the selected stocks based on the past stock prices and financial statements, and show the data on the right margin of this page.

This is the Design 1 in our proposal. We put the other two designs in the Appendix. In this design we combine different visualization views together. Each kind of view has their defects, and this combination make complementary result. Switching between different views shows different aspects.

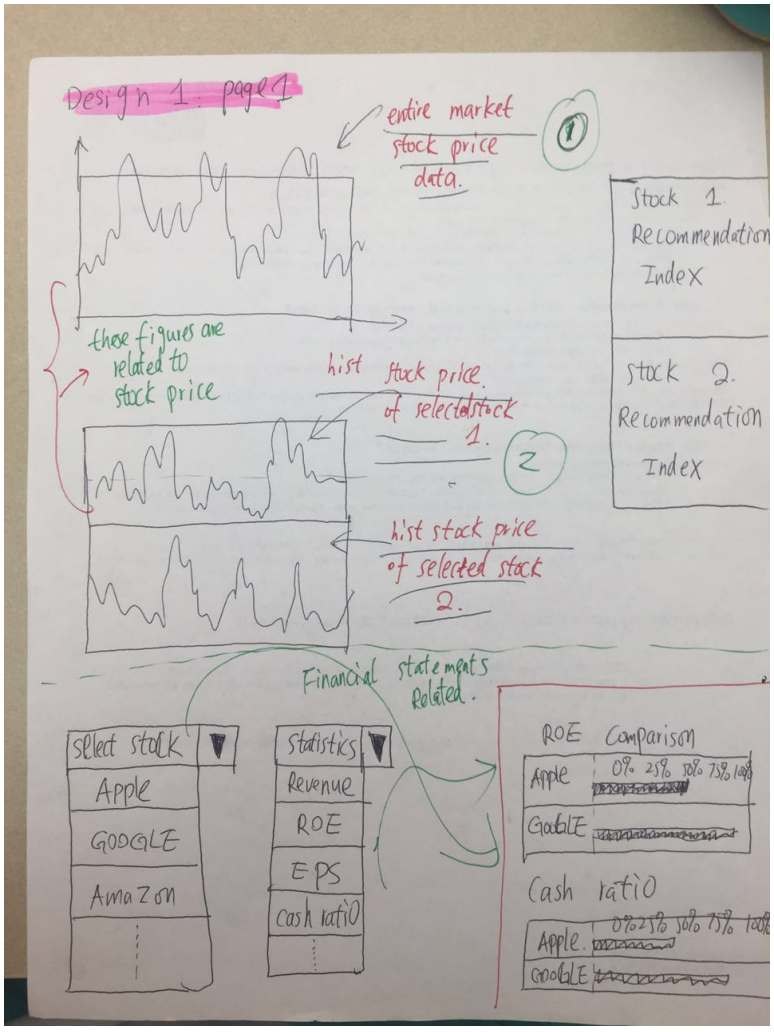


Figure 1. Original Design - Overview

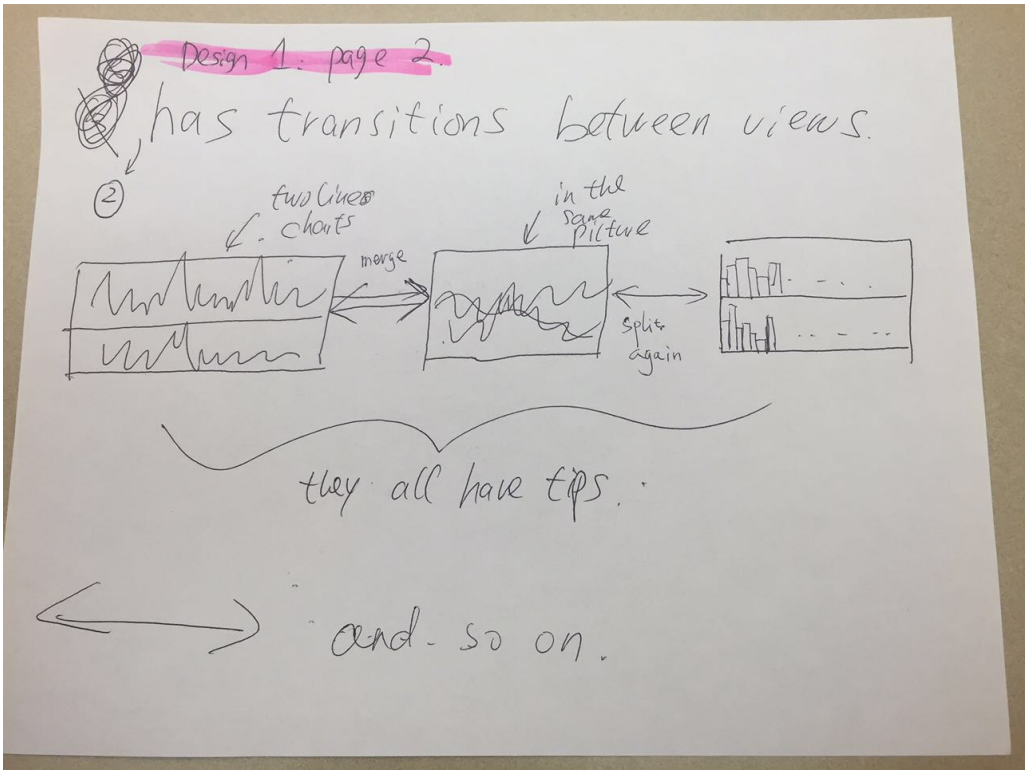


Figure 2. Original Design - Views Switch

7.2 Design Suggestions from In-class Peer Review (2016-10-25)

As we discussed with members from other teams, they suggested that:

1. The idea of part 1 (create a minimap of the whole market with axis. Use brush to select time interval) sounds good, but it's not practical in use: people can't use brush to select a very exact time interval. This is right, our stock price data is daily and the whole time interval is tens of years. Thus, we decide change our mind add to text frame on the top to let people fill in the start date and end date of the interval they want to inspect. Then the time axis of part 1 and part 2 will change according to this.

2. It is more important to compare single stock price with S&P500. It will show the contrast between two trends or consistency of the two trends, and this comparison is more meaningful than the comparison of two stocks in different categories. This is a good point.

3. They also suggest that we need to find a webpage for reference as our template such that the page doesn't look weird. That is, we should make full use of the entire webpage, and rearrange all the things we have so that the structure of the page looks attractive.

7.3 Final Design

In the final design we emphasize the transition behaviors, which make the final display more attractive, and it's also very different from the original design. Though we think we keep the main

idea of the original design.

Part I. S&P 500 index (Time Series; Line Chart)

Fixed size svg. It also has a focus circle which can move along the S&P 500 index value line following the mouse. The price of the focused data point is printed out to the up left corner.

Part II. Two Stock Prices (Time Series; Line Chart)

Two ticker selectors are given. Use to select tickers and the related stock prices will be display in lines. These two tickers

There are three modes/views:

i. Single Lines Chart: Two lines use common time scale and price scale.

ii. Separate Line Charts: Two charts, and each has its own line, time scale, and price scale.

iii. Single Lines Chart (Normalized): Two lines use common time scale, but different price scale. Each price scale is shown in one y axis. Two y axis have the same range but different domain.

Use the Chart Type selector to switch among different views. Transition effects are added.

With using these three selectors (two ticker selectors and one chart type selector), the time scale and price scale of S&P 500 index (Part I) will also change:

i. Single Lines Chart: Use the common time scale.

ii. Separate Line Charts: Use the widest time scale.

iii. Single Lines Chart (Normalized): Use the common time scale.

Justification: The time scale can be large or small. The performance of line charts are good for all the time scales we may use. This is the reason why we decide to use line charts only in our final design.

Also, with the interaction (adjust and sync scales automatically), though we didn't put the svg of Part I and the svg of Part II together, we can compare them. This is our method to follow the in-class peer review suggestion 2.

Part III. Finantial Statements Related Values (Time series; Bar Chart)

These are also affected by the three selectors of Part II.

The picked values are: Cash Ratio, Current Ratio, Quick Ratio, Growth Margin, Sales Growth, Operation Income/Current Debt, Total Debt/EBITDA, Total Debt/Asset, Total Debt/Equity, Receivable Days, Effective Tax Rate.

There are not many negative values. For preventing them to affect the layout we use a different scale for these negative values, which ensure the negative value bars are not tool

big in size.

Certain financial statements related values for a specific company or year might be not available. Not available value doesn't mean no value, thus we use fixed size grey color bars to represent these values.

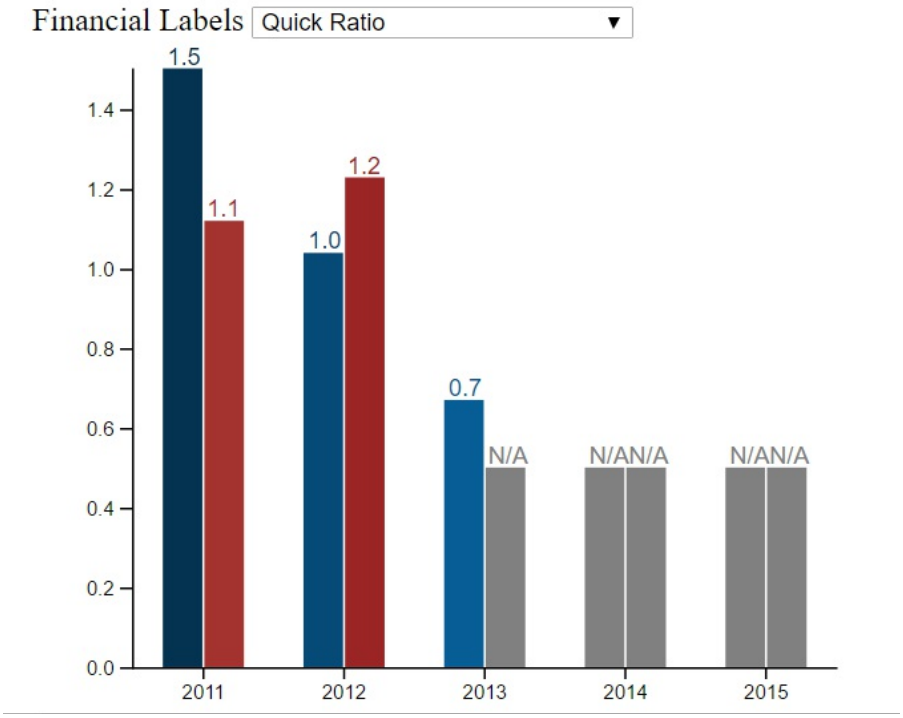


Figure 3. Not Available Values

The bar chart will change according to the selected tickers (use ticker selectors of Part II) or the financial statements value types (use financial labels selector).

Figure 4. Final Design - Single Lines Chart (ticker1: aapl, ticker2: goog, financial label: Cash Ratio)

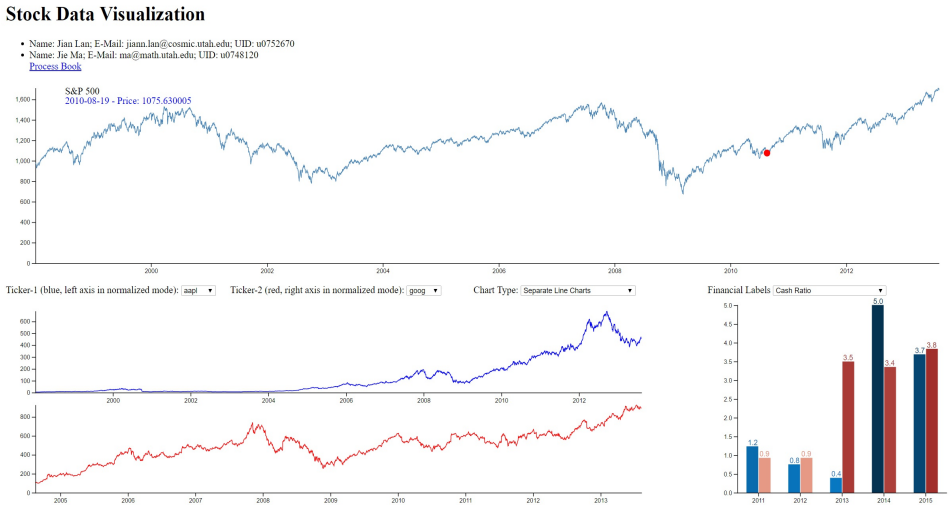


Figure 5. Final Design - Separate Line Charts (ticker1: aapl, ticker2: goog, financial label: Cash Ratio)

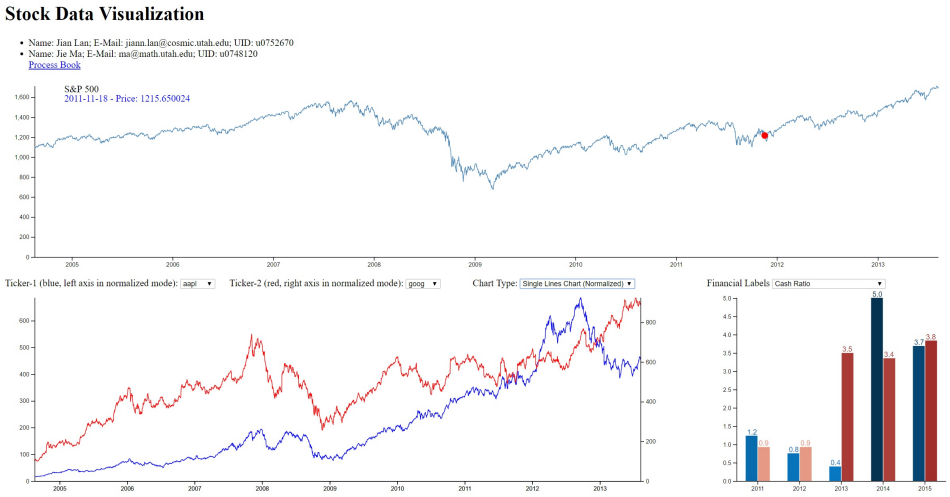


Figure 6. Final Design - Single Lines Chart (Normalized) (ticker1: aapl, ticker2: goog, finantial label: Cash Ratio)

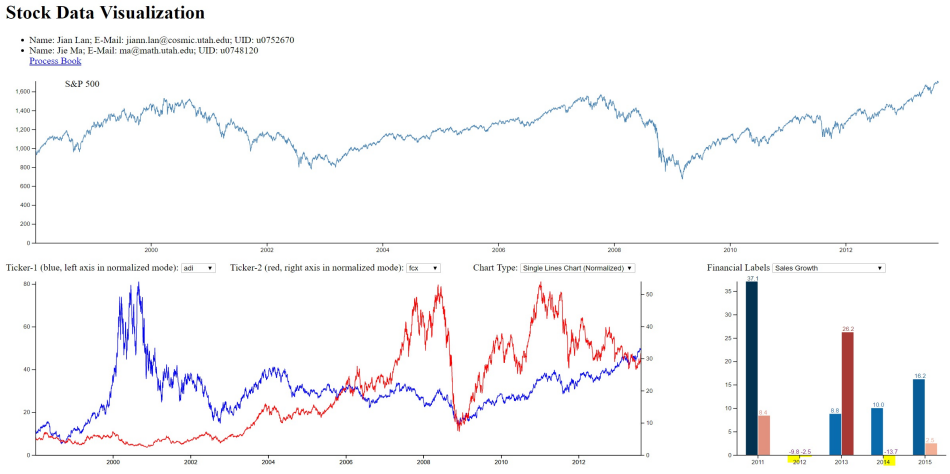


Figure 7. Final Design - Single Lines Chart (Normalized) (ticker1: adi, ticker2: fcx, financial label: Sales Growth)

8 Implementation

Part I. S&P 500 index (Time Series; Line Chart)

Load the `S&P500Index.csv` data. Draw a line the corresponding axis. The mouse position can be found, and read out its `x` coordinate, use the time scale and `d3.invert` method to find out the index value of this specific time point. Add a focus to the line, and print out the index value.

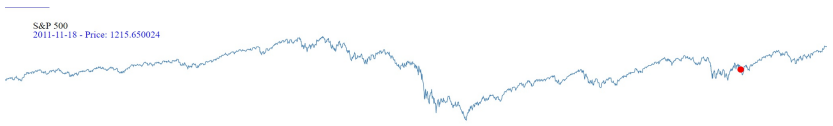
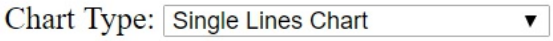


Figure 8. Focus and The Index Value

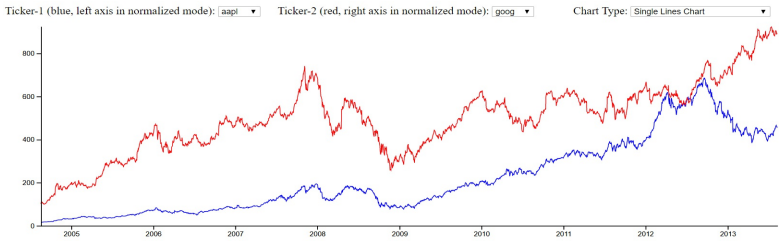
Part II. Two Stock Prices (Time Series; Line Chart)

- Ticker selectors:

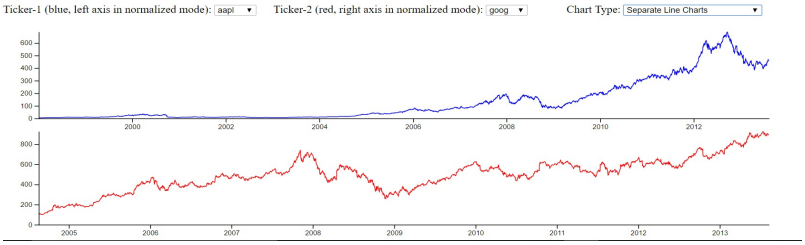
- Chart Type selector:



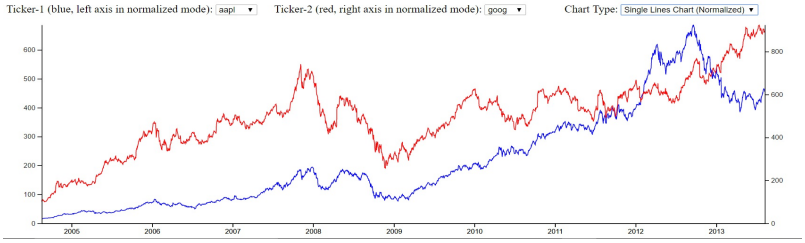
- Single Lines Chart: Choose the intersection of the two time scales of selected tickers as the common time scale and plot the x axis. Choose the union of the two price scales of selected tickers as the common price scale and plot the y axis.



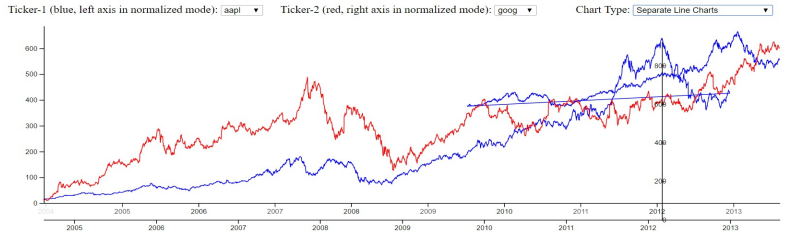
- Separate Line Charts: Two charts, and each has their own line, time scale, and price scale.



- Single Lines Chart (Normalized): Common x scale as in Single Lines Chart. Each line has their own y scale.

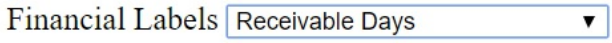


– Transition between views: Here is a example of an on going transition. Use `object.transtion().duration(3000)`.

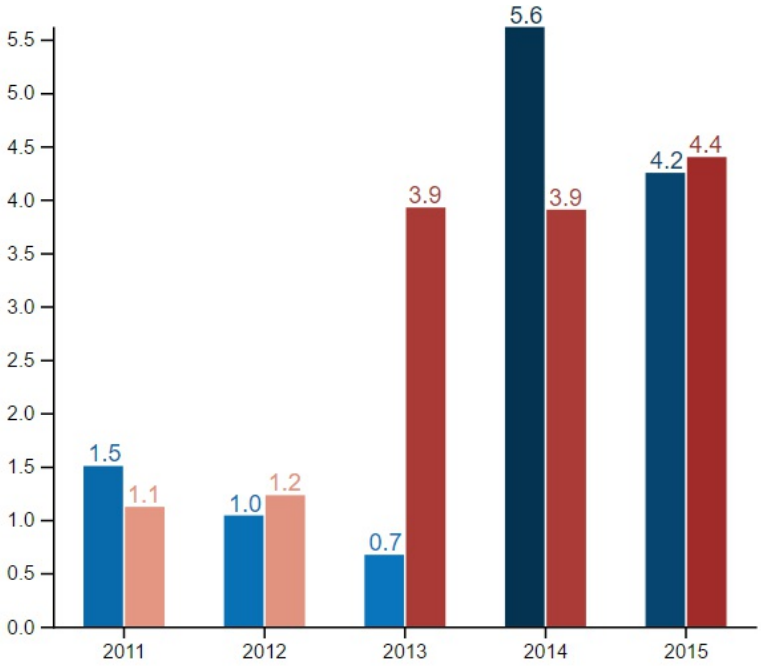


Part III. Financial Statements Related Values (Time series; Bar Chart)

– Financial Lable selector: We first create a drop-down menu include financial indictors we choose. Once selected, the bar chart will change.



– Bar chart: The bar chart also follow the change of ticker selectors in Part II. Ticker change can be detected, and new `csv` files will be loaded both for Part II (daily price of stocks) and Part III (the values of a financial labels).



We also perform colorscale techniques on the data. Specifically, we let both blue and red to range from light to dark, to indicate the relative height of each bar. Since there are also missing data, that is either due to the reason that for some certain companies, those companies do not have those data in their dataset, or due to the reason that even though the labels are in the data set, but for some specific years, the data is not available, we use grey bars to indicate that data are unavailable.

- Transition of Bars:

  We add transition effect to bars. Use `object.transition().duration(2000)`.

9 Evaluation

- What we have learned about the data by using your visualizations:

  Building interactive behaviors between different time series data and automatically adjust their scale is a good way to help us inspect data. We can concentrate on comparing them, rather than, before comparing them, spend more time to adjust their scale.

- The answers to your questions of Section 4:

  Our questions are hard to answer if there is only data. However, the answer is almost intuitive if with an proper visualization. Just inspect the data with our visualization you can answer these questions. We gave an example to illustrate this in Section 6.

- According to our test, our visualizations work fine.

- How could you further improve it?

  - If there is more time, we should add more content to `css` can make our visualizations have a better style.

  - Find a efficient way to add time interval selection to our visualizations can improve it. I tried this when we were implementing this project, but for some reason our method is very inefficient and make the page work very slow. Thus, we didn't add this feature to our find result of this project now.

  - The ticker selectors should also include the information of full company names, rather than only tickers.

Reference

[1] Duisenberg, Willem F. "The role of financial markets for economic growth." *BIS Review No* 48 (2001).

Appendix

1 Design

Design 2. (inspired by the desgin of "Sex, Drugs, and Munchies")

1. Show the stock price of the whole market at most two large figures. Each figures has their own filters.

   - Data can be presented as line chart or candle chart, and switch between them.

   - You can select the time interval from a list of 1 Day, 1 Month, 1 Year, and 5 Years.

- If there are two figures, you can adjust their filters independently.

2. Base on the Financial Statements, theoretically we can create hundreds of filters to filter our collected data, though we won't create so many. We can choose the most important ones to create filters base on Jie's financial knowledge. These filters can be classed as three major categories: income statements, cash flow, and balance sheet. We can do the similar thing as peple did in the project of "Sex, Drugs, and Munchies". Filter the data will lead to three possible cases (use a drop down menu, you can switch between these three cases):

   - The whole market figure will no longer for the whole data, but for the filtered data.

   - Draw the filter data also in the whole market figure. Put them together, and than people can compare them.

   - Draw the filter data in a separate figure.

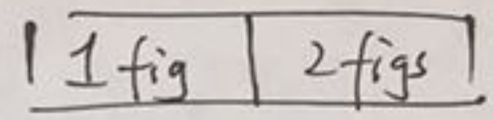
Design 3.

1. For Design 3, we create a plot, in which we have bubbles, each of which corresponds to a single company. To the left (or right) of the bubble plot, we append a drop-down menu, which contains a selection of attributes we use as the 'y' variable in the plot. The radius of the circle will also be used as an indictor of the performance in that 'y' variable.

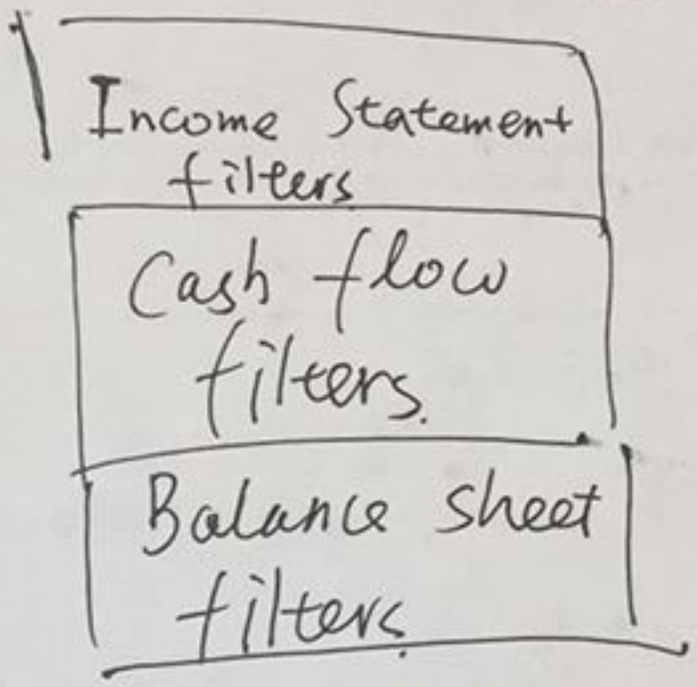
2. We append tooltips to each circle such that some of the most basic information of the company will show up. Once one of them is clicked, two new plots will show up below it. The first one will show the historical stock price as a function of year, and the second will show the annual return of that stock. The time period during which the two plots will show is controlled by a drop-down menu to the right of them.

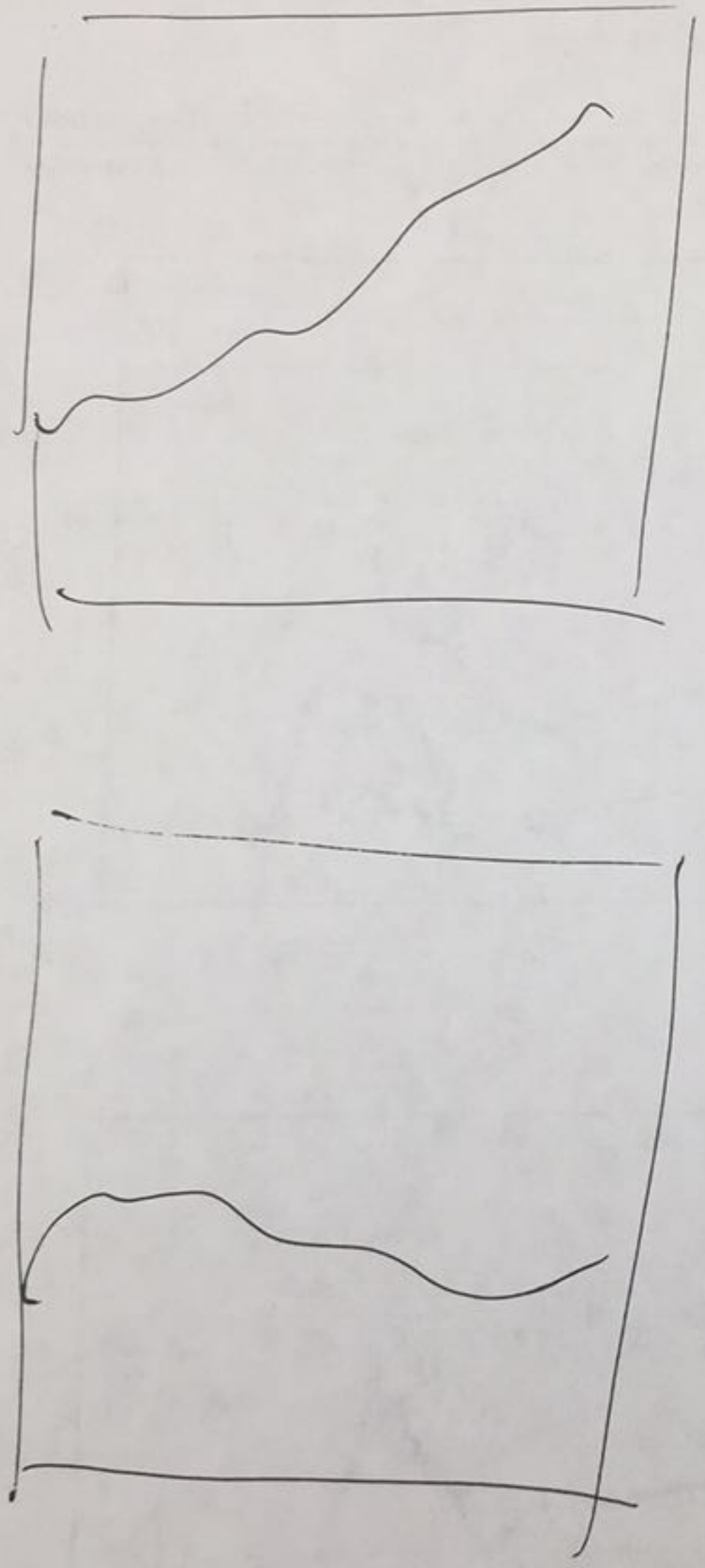
The filters in Design 2 are good, but it's boring; We only implement so many filters, and too many of them will make the project to complicated to convey useful information, and the layout is hard to design. We can pick up the most important filters in design 2, and add them to design one.The drawback of Design 3 is that since there are many circles, and lots of them might just overlap with each other, some of them companies might actually not be visually accessible, since they might be blocked by adjacent circles.

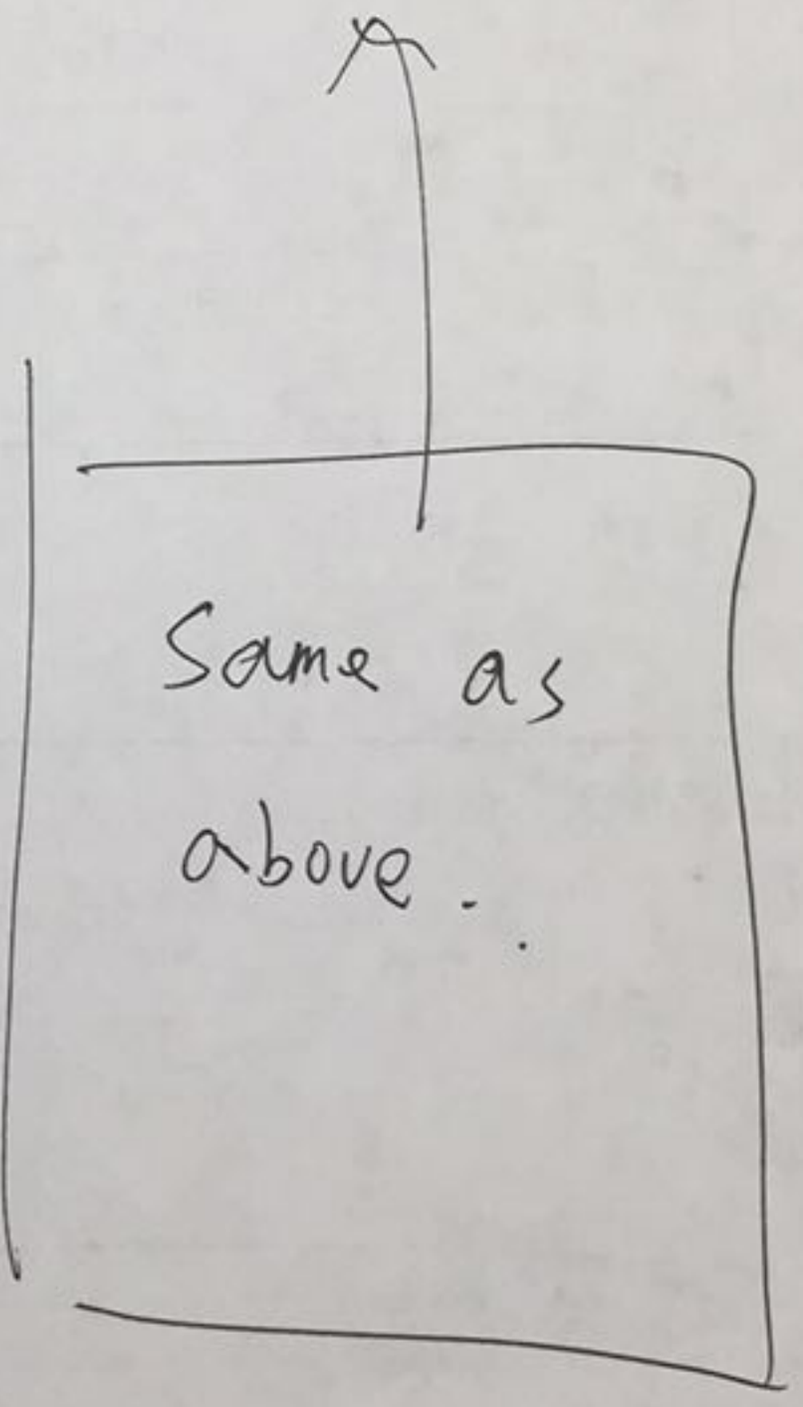
Use this to choose
1 fig view or 2 figs view



bottoms

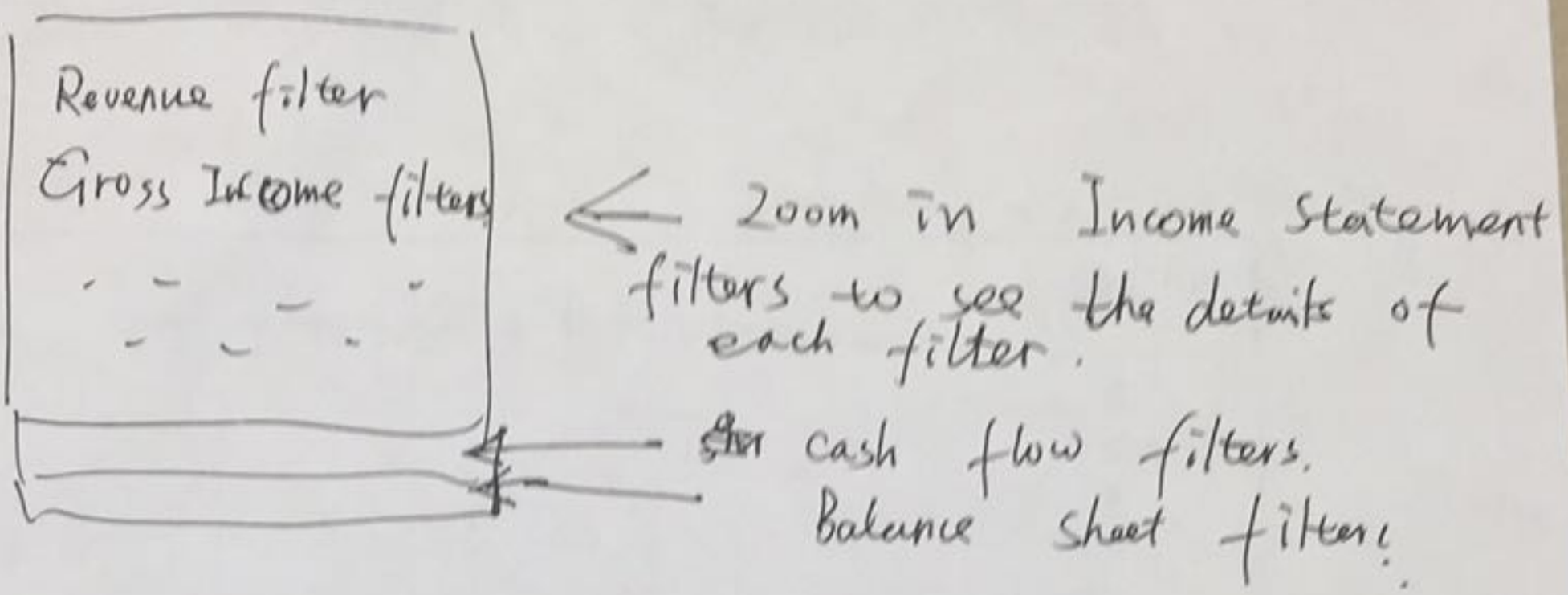






Desig 2 - 1.

When focus on Income Seatement filters Category.

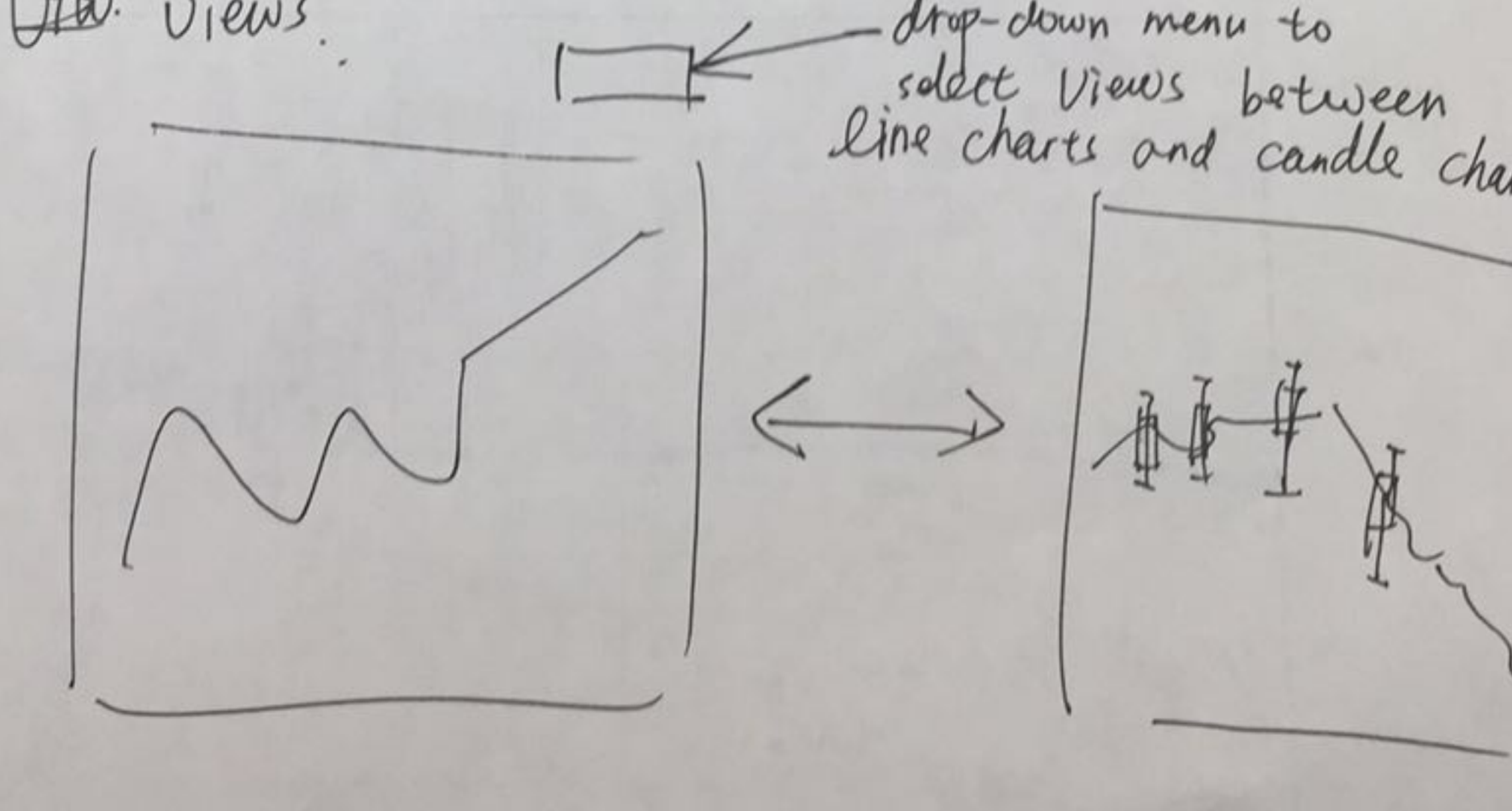


(2)

---

(3)

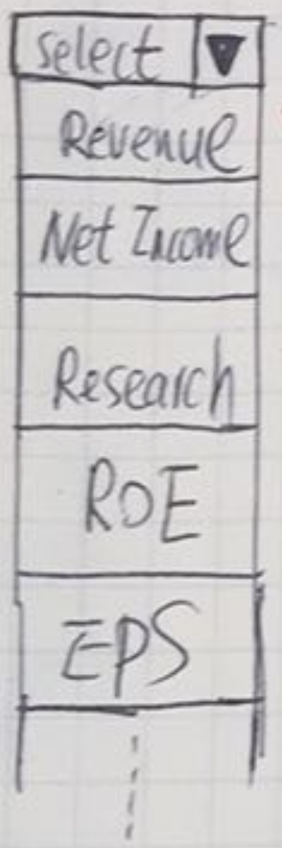
Figures have transitions between two different views.

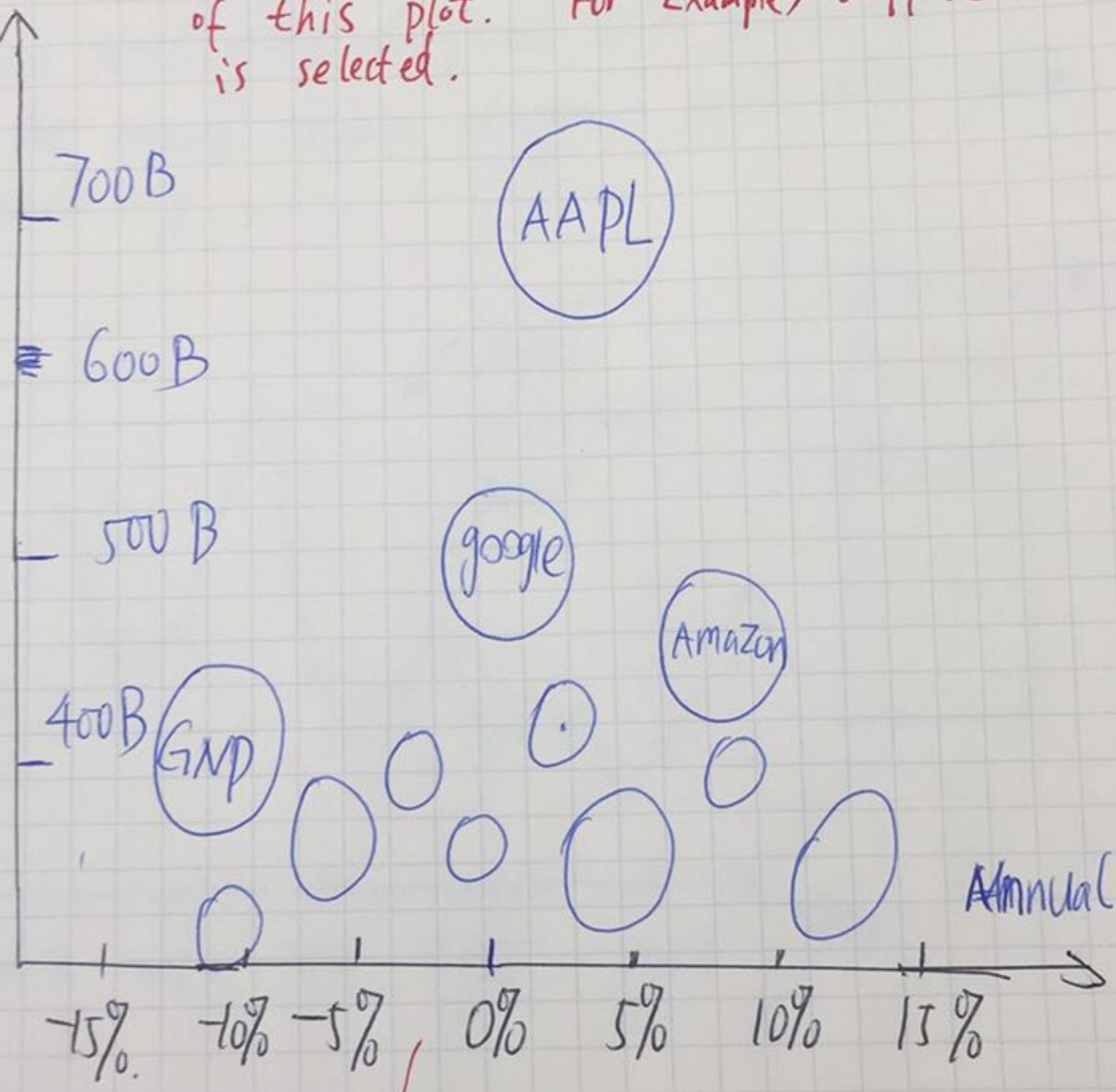


Design 2-2.

Design 3.

This is a Drop-down Menu, and the selected attributed will be used as Y-coordinate of this plot. For Example, suppose now Revenue is selected.





x-axis represents annual return, and if it is greater than 0, the stock generates positive profit.

Design 3 (page 2)
upon clicking on any circle on the last page, the historical stock price diagram will show up below. ~~The~~ Annual return diagram will also show up in another svg.