

Process Book for Stock Data Visualization

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 visulization works about stock data visulization:

- [D3 Show Reel](#)

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 donut.

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.

4 Questions

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

- 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 dataset from two candidate datasets:

- The A share¹ of the Chinese stock market.
- 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](#).

1. In finance an **A share** is a designation for a “class” of common or perferred 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.

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.

About the data collection work:

1. We have the full dataset for type 1 data.
2. We have a sample of type 2 data. This is enough for implementing our project.

We can access any data we want through Bloomberg Professional Workstation at any time. We just didn't try to get the full dataset of type 2, because some decisions can be made only when we implement part 3, and we are still working on it. We don't want to waste time to collect a huge amount of data that we won't finally use in our project.

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.

The quantities derived from the collected data that we will present:

- Stock Prices (technical analysis²). The stock price data contains two parts: (1) the stock price for the entire market and we use S&P 500 as benchmark; and (2) the stock price for individual stock price, for each individual stock that is in the basket of S&P 500. These data are all numerical values in the original dataset, so there is pretty much nothing we have to modify or pre-process.
- Data from Financial Statements (fundamental analysis³): 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 calculated 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.

These two types of data analysis are usually combined together in a presentation to provide an overview for different aspects.

6 Exploratory Data Analysis

We initially use line charts to 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, which in our case, is the 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.

2. Investors will directly 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 its 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 undervalued or overvalued by only looking at their price.

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 `googl`), 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 significant 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 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.

7 Design Evolution

7.1 Visualization Design

We are inspired by many examples demonstrated in class and examples outside of class on the Internet — stock data presentation and visulization is so poplular 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.

Design 1.

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 trival. 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 insprired 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 Design 1 is our choice. We put the other two designs in the Appendix. In Design 1 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.

7.2 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. And we actually put this idea straightly into our design, which is discussed in detail in section 5.1.
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.
4. They thought our idea is clear and concise. After we explained our design to them, they told us that they understood what we are trying to do, and they think that our visualization would achieve our goals.

7.3 Must-Have Features

Design 1 is the main implementation of our final design as described in the last paragraph of the section 6. The must-have features are the features showed in part 1 and 3. For part 2, at least we'll implement three views.

7.4 Optional Features

The optional features are those in design 1 part 4. If we have enough time, we'll implment five or more views for part 2 and more filters.

8 Implementation

8.1 Project Schedule

One of the member of this project, Jie Ma, is going to take the CFA (Chartered Financial Analyst) level I exam, who is the member that has a better understanding of the data, while the other member, Jian Lan, has much more expericen in coding, and he will play a more important role when doing the implementation.

1. **2016-10-24 ~ 2016-10-30** Submit Proposal (2016-10-24); Data Collection (Start); Code framework.

2. **2016-10-31 ~ 2016-11-06** Webpage Layout Design (in details); Implement Design Part I; Data Collection (End, hopefully).
3. **2016-11-07 ~ 2016-11-13** Implement Design Part II; Try to implement Design Part III Project Milestone (2016-11-11).
4. **2016-11-14 ~ 2016-11-20** Finish Design Part III; Refactor and Cleanup the code, adjust code structures. If there is still time left we want to use ES6 features to re-write the code (the code will be translated to ES5 with proper tools when being loaded by the webpage, thus it doesn't matter if the browsers have implemented ES6 features or not).
5. **2016-11-21 ~ 2016-11-27** Screencast; Website.
6. **2016-11-28 ~ 2016-12-02** Minor Adjustments; Redo the Screencast part if the adjustments affect the final result view; Submit the Final Project (2016-12-02).

8.2 Current Work

We have finished the basic views of part 1 and part 2. A lot of details will be added later. Honestly speaking, the view of our prototype now is not attractive, but we promise it will be (after adding enough details). We will finish the basic view of part 3 during this weekend. After then we will cleanup the code a little and try to add details to them.

One thing we need to mention is that we use the code framework from our homework 5. Recently we are focusing on implementing functionalities, and have no chance to edit the css file from homework 5. You may see some wierd info in the `styles.css` file, and please ignore them.

9 Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

This part will be finished later.

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 people 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.

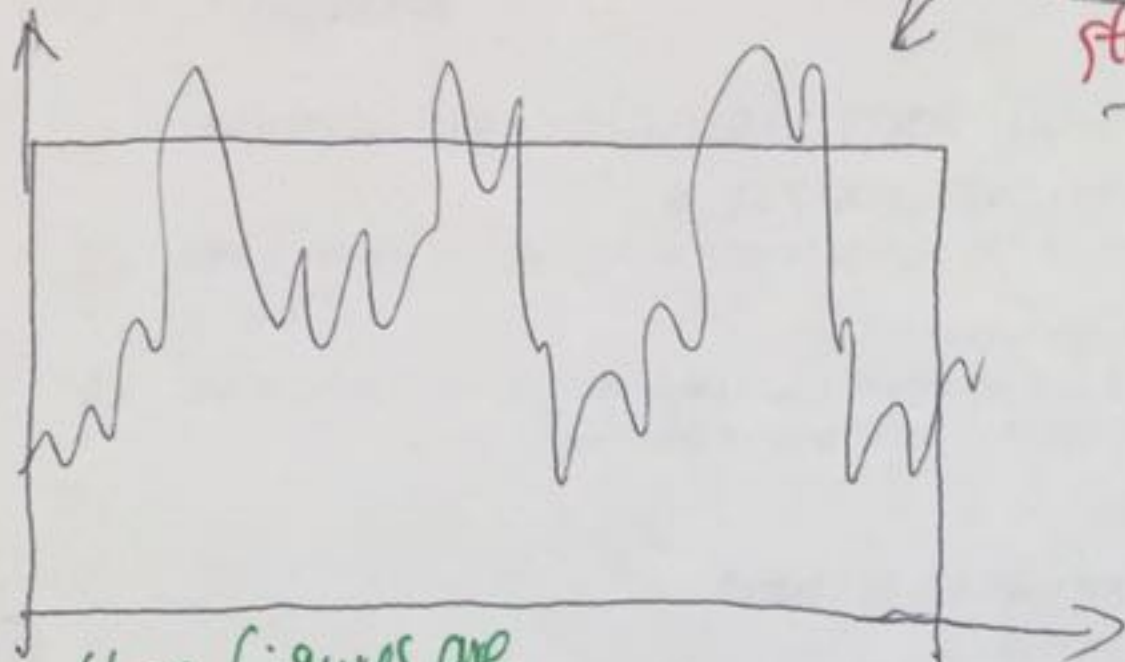
Design 3.

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.

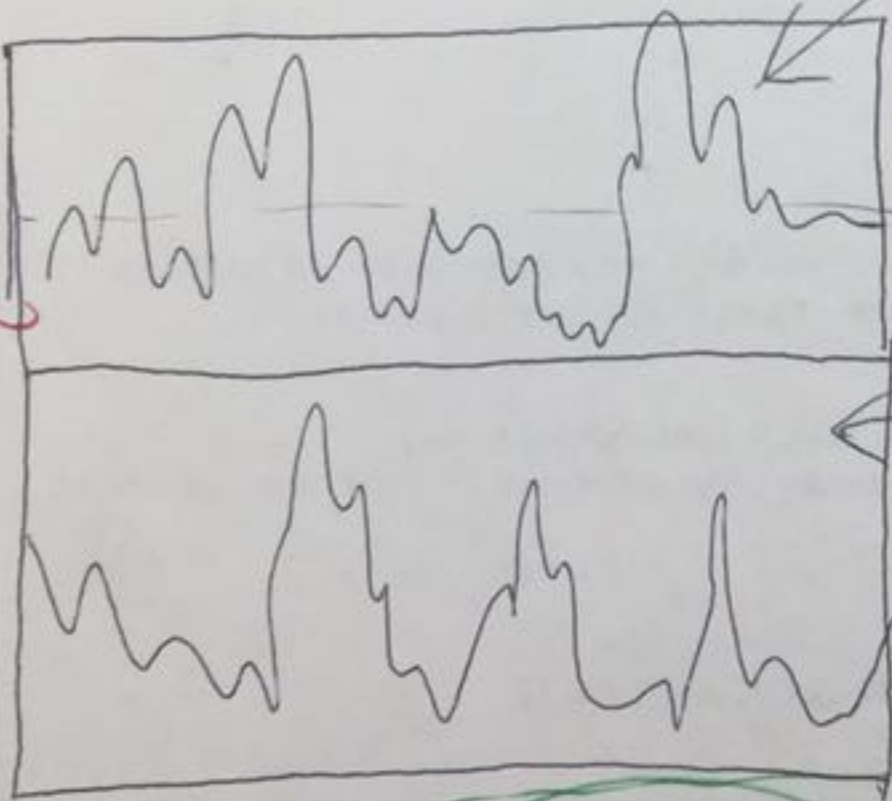
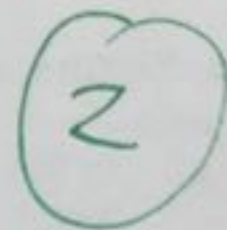
Design 1: page 1

entire market
stock price
data.



these figures are
related to
stock price

hist stock price
of selected stock
1.



hist stock price
of selected stock
2.

Stock 1.
Recommendation
Index

stock 2.
Recommendation
Index

Financial statements
Related.

select stock	▼
Apple	
GOOGLE	
Amazon	
...	

Statistics	▼
Revenue	
ROE	
EPS	
cash ratio	
...	

ROE Comparison

	0%	25%	50%	75%	100%
Apple	0%				
GOOGLE	0%				

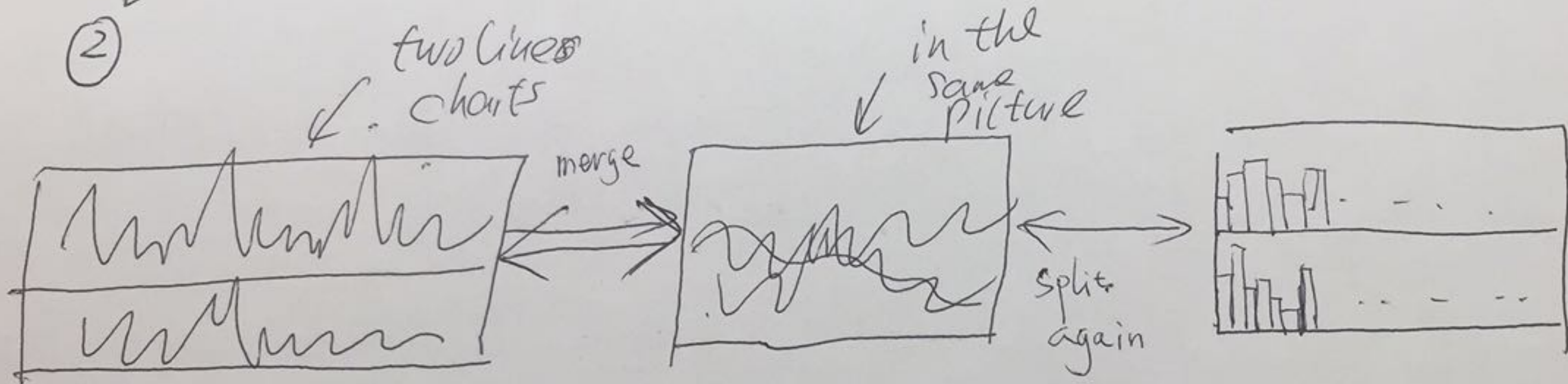
Cash ratio

	0%	25%	50%	75%	100%
Apple.	0%				
GOOGLE	0%				

Design 1: page 2.

has transitions between views.

(2)

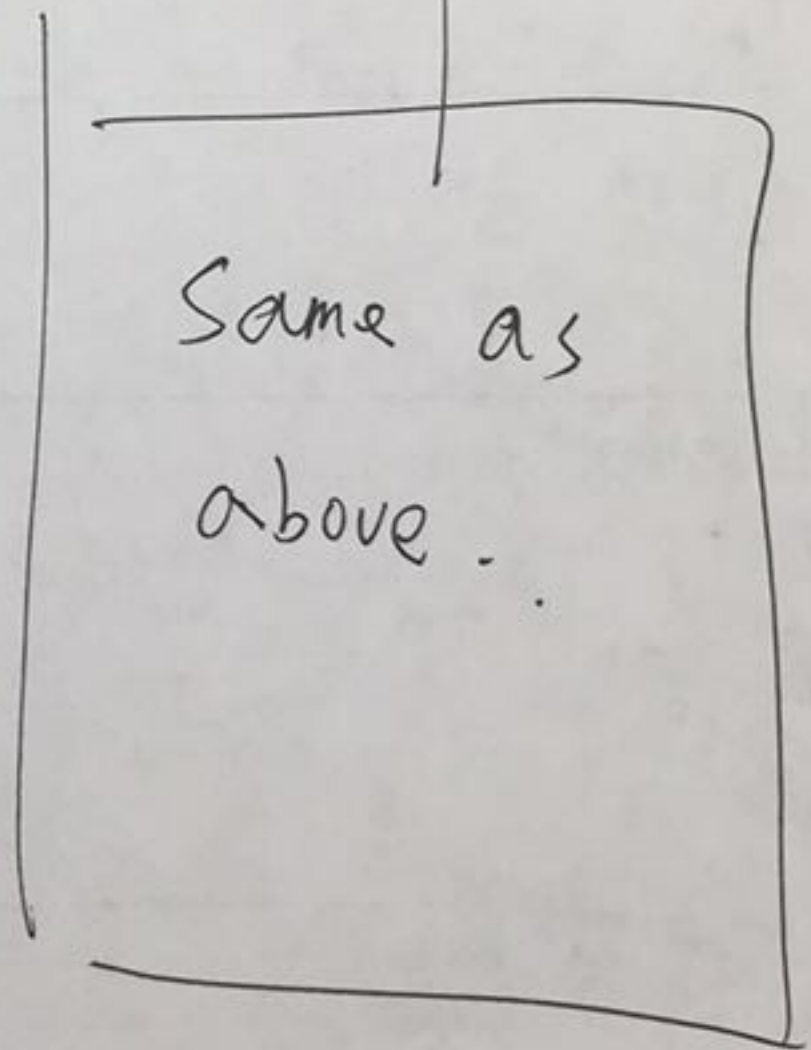
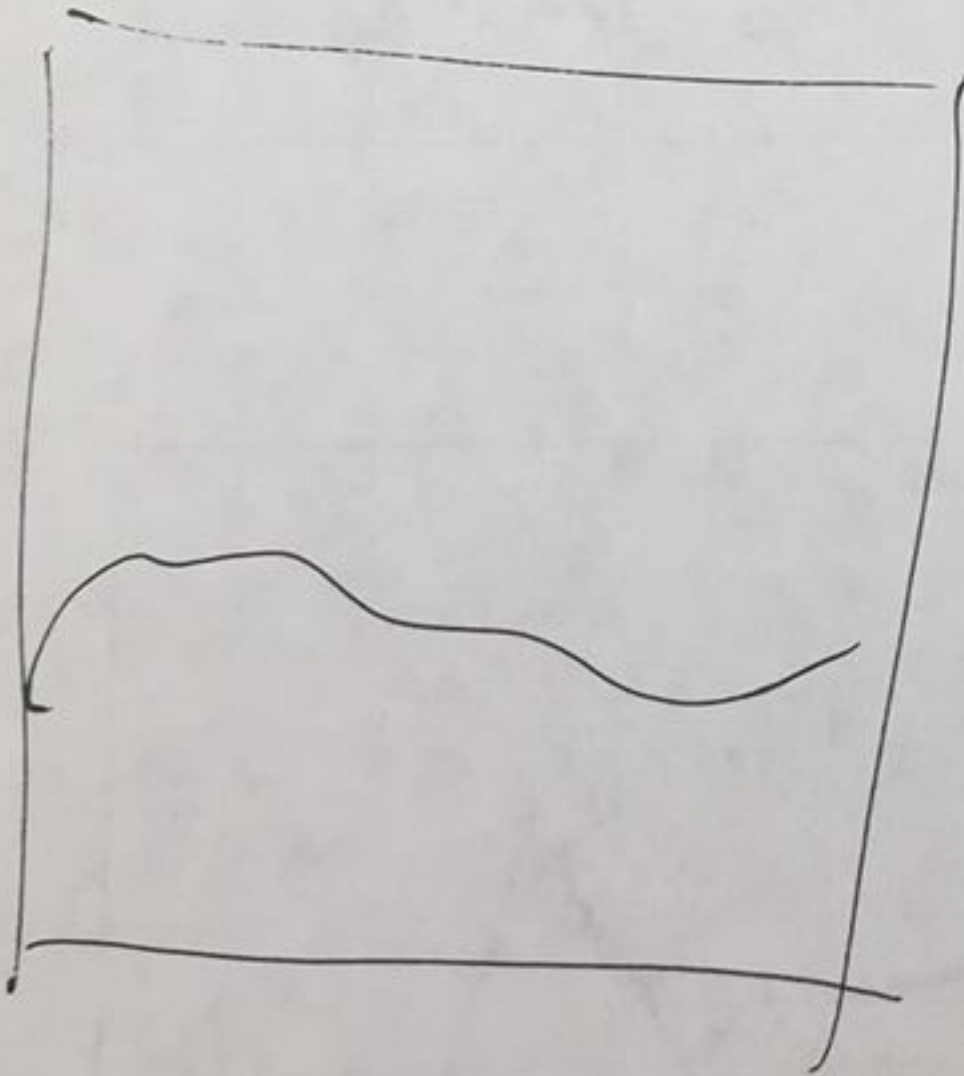
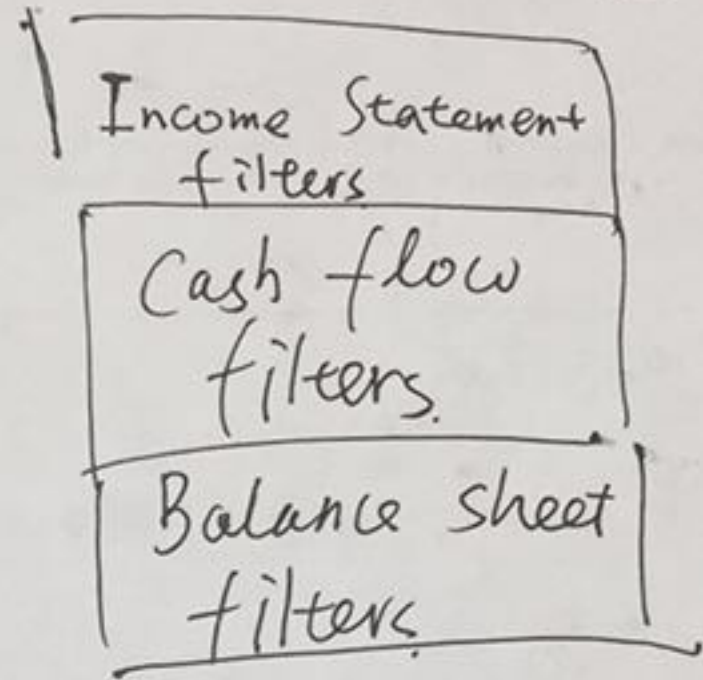
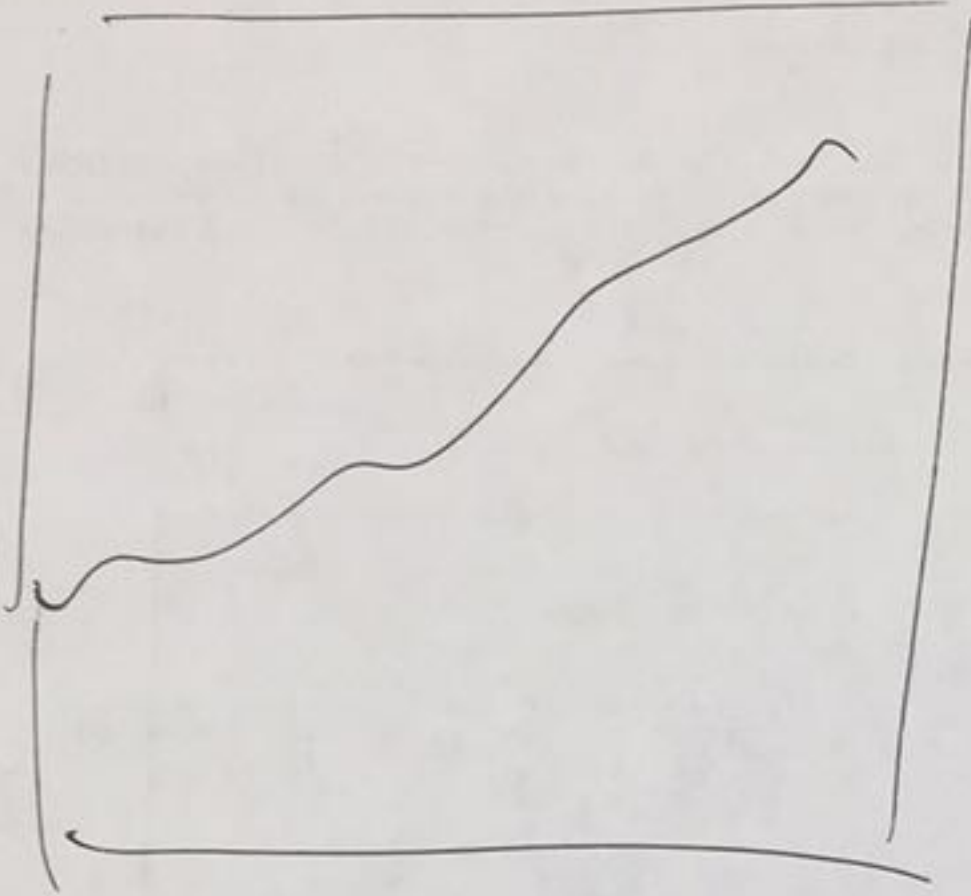


they all have tips.

and so on.

Use this to choose
1 fig view or 2 figs view

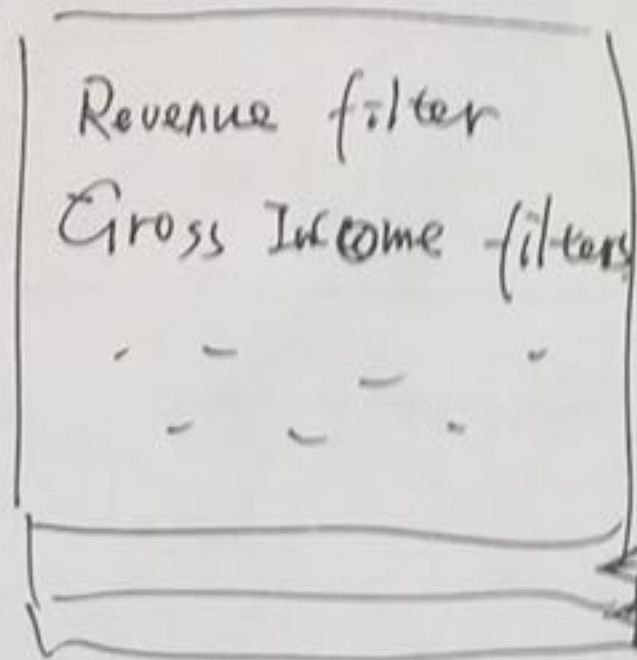
1 fig | 2 figs
bottoms



Desty 2-1.

Q

When focus on Income Statement filters Category.



Zoom in Income Statement filters to see the details of each filter.

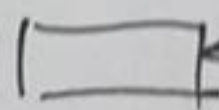
cash flow filters.
Balance sheet filters.

②

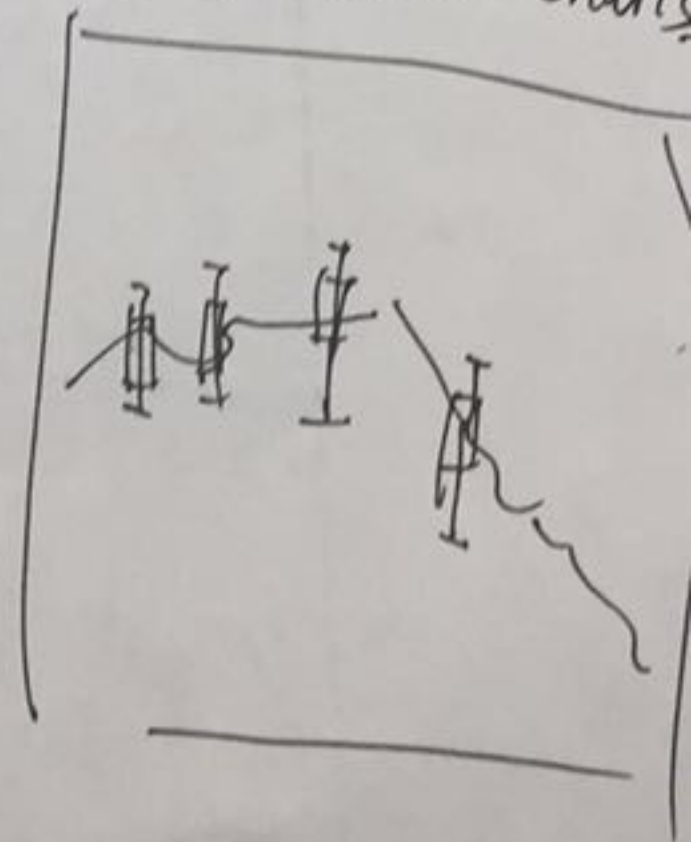
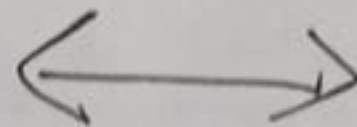
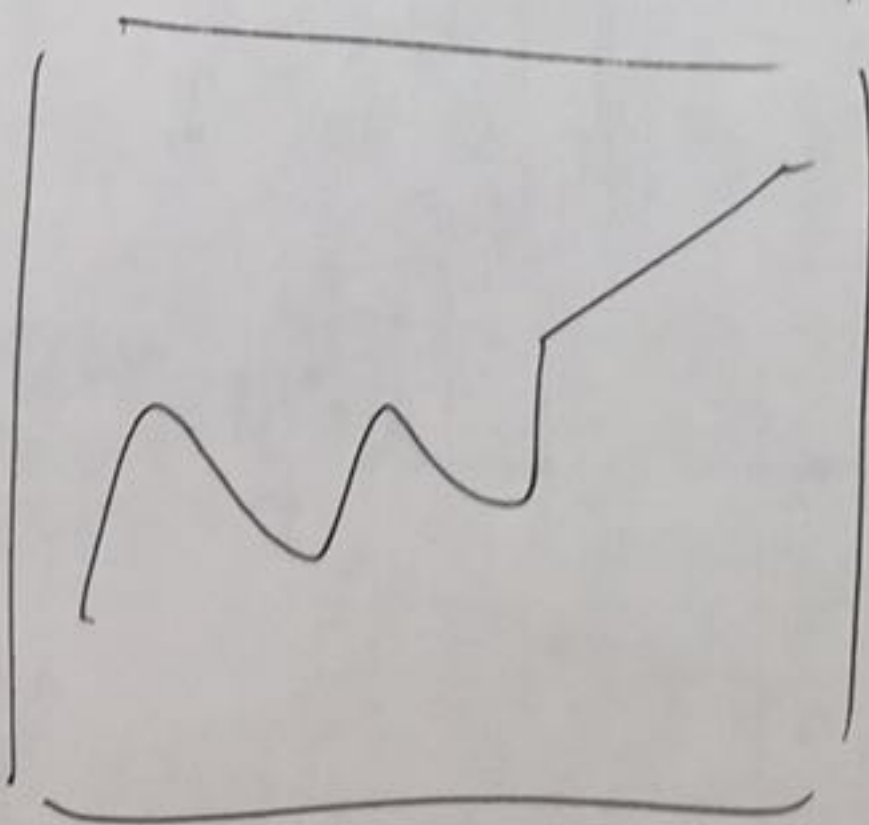
③

Figures have transitions between two different

~~view~~ views.



drop-down menu to select views between line charts and candle charts.

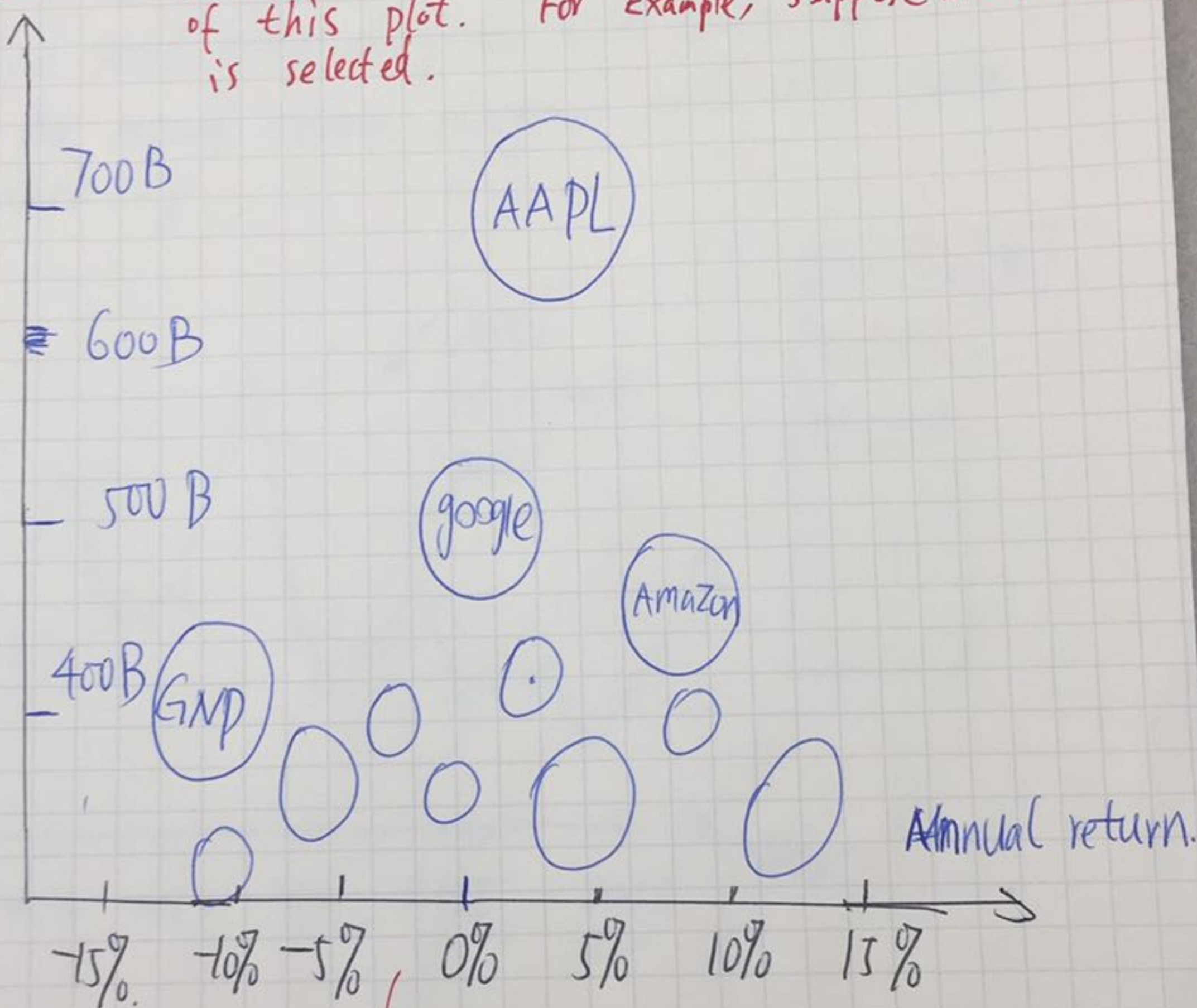


Design 2-2.

Design 3. (page 1)

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

select ▼
Revenue ✓
Net Income
Research
ROE
EPS
...



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.

