

Stock Data Visualization Proposal

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 Background 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 Project Objectives

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?

4 Data

There are two resources we can use:

- The A share¹ of the Chinese stock market. These data are retrieved from The Wind Info financial database². For simplicity, you can consider Wind Info as a Bloomberg-like company in China. It has the most complete information on Chinese stocks and stocks related information.
- The stocks of Nasdaq/NYSE stock market. The data can be accessed through the Bloomberg Professional Workstation in Marriott Library of our university. Here is a guide to using the workstation: [Bloomberg Professional Workstation Guide](#).

The data in these two database are in the same format (csv file) and structure, we can choose any one of them. However, we prefer the stocks of Nasdaq/NYSE stock market.

Now we can guarantee that it's easy for us to get the A share data³. The Nasdaq/NYSE data, on the other hand, before we get the reservation and access the workstation, we can't guarantee anything.

5 Data Processing

All the retrieved data are saved in csv format. The data are extremely clean. From the sample we have now, we may have none or only little cleanup work to do. In most of the cases, the raw data is good enough for being presented. The only data processing work we need to do is to combine some data together. Thanks for `d3` the the practice we did in our homework, this is also not hard to do.

The quantities derived from the collected data that we plan to present:

- Stock Prices (technical analysis⁴).
- Data from Financial Statements (fundamental analysis⁵): 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.

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.

2. <http://www.wind.com.cn/en/edb.html>

3. Thanks for Jian's friend DaZhong Wang, who is working for the Xiamen University as a Postdoc in Economy. He is the person who has permission to access the Wind Info financial database, and he told us we can use these data legally for non-commercial purpose.

4. 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.

5. 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 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.

6 Visualization Design

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.

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 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.

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 indicator 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.

Design 1 is our choice. In design 1 we combine different visulization views together. Each kind of view has their defects, and this combination make complementary result. Switching between different views shows different aspects. 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.

7 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.

8 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.

9 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.

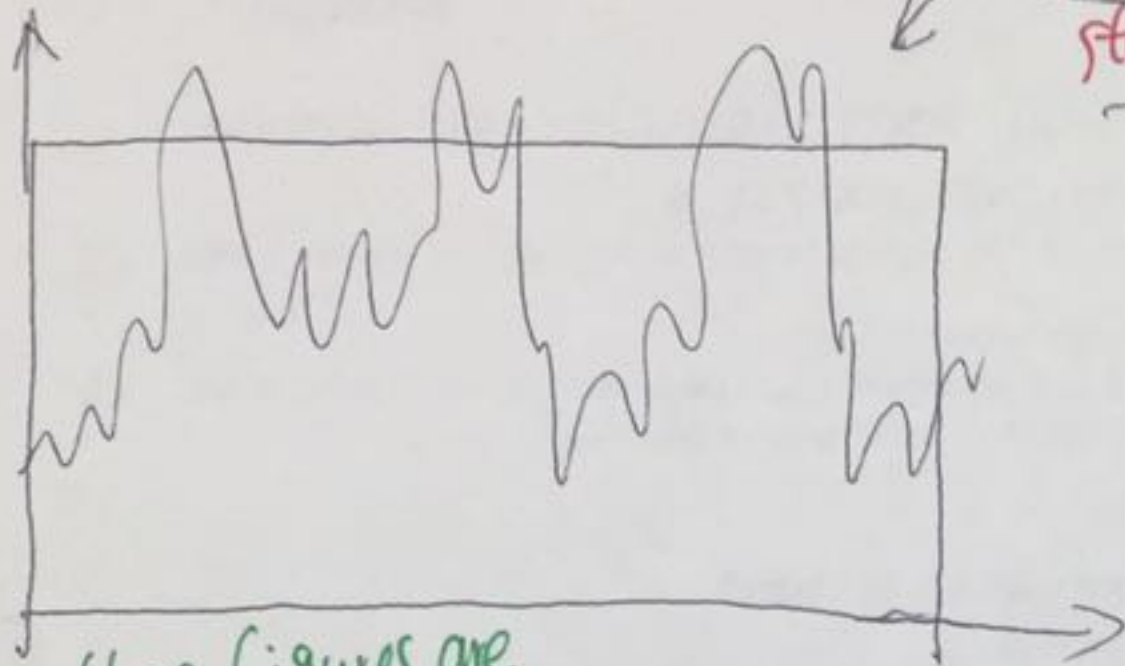
1. **2016-10-24 ~ 2016-10-30** Submit Proposal (2016-10-24); Data Collection (Start); Code framework.
2. **2016-10-31 ~ 2016-11-06** Webpage Layout Design (in details); Implement Design Part I; Data Collection (End, hopefully).
3. **2016-11-07 ~ 2016-11-13** Implement Design Part II; Try to implement 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 bowsers 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).

Reference

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

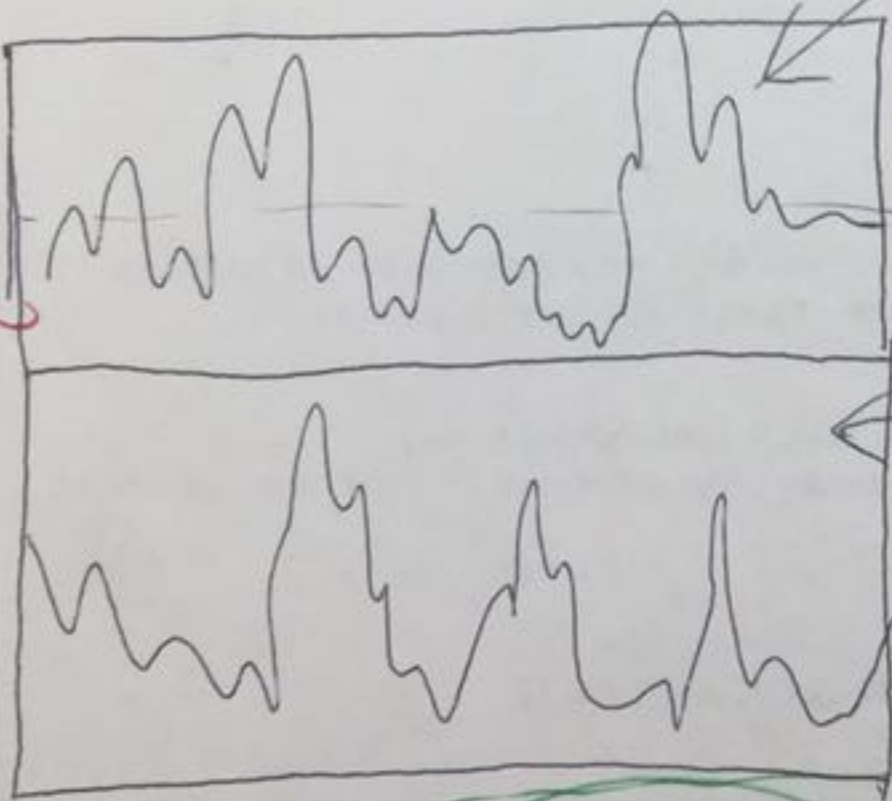
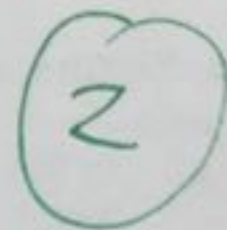
Design 1: page 1

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

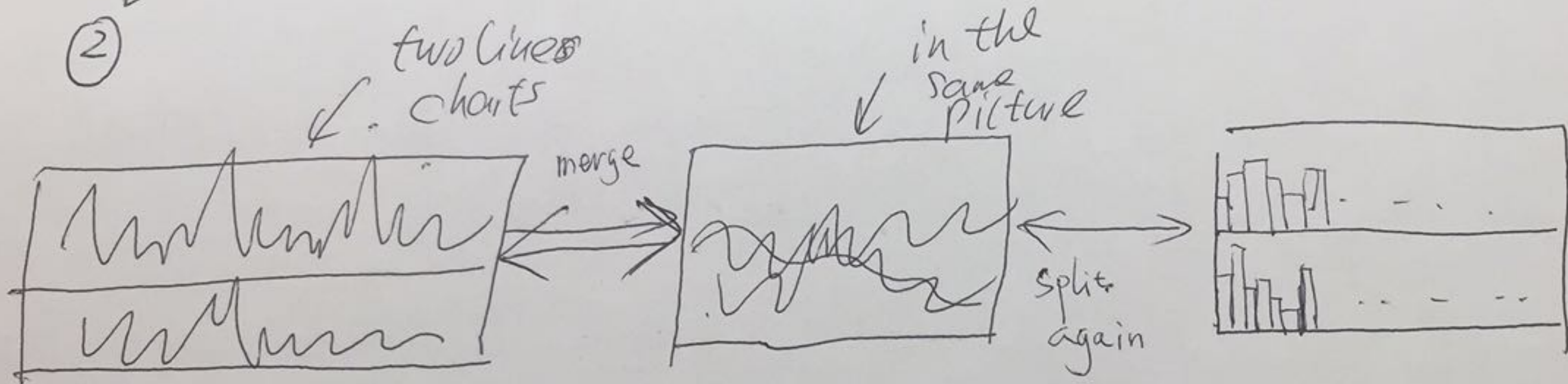
Cash ratio

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

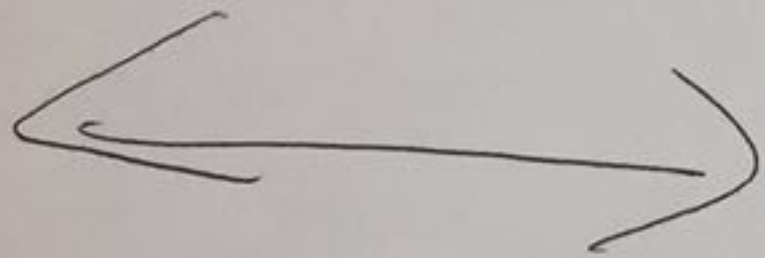
Design 1: page 2.

has transitions between views.

(2)



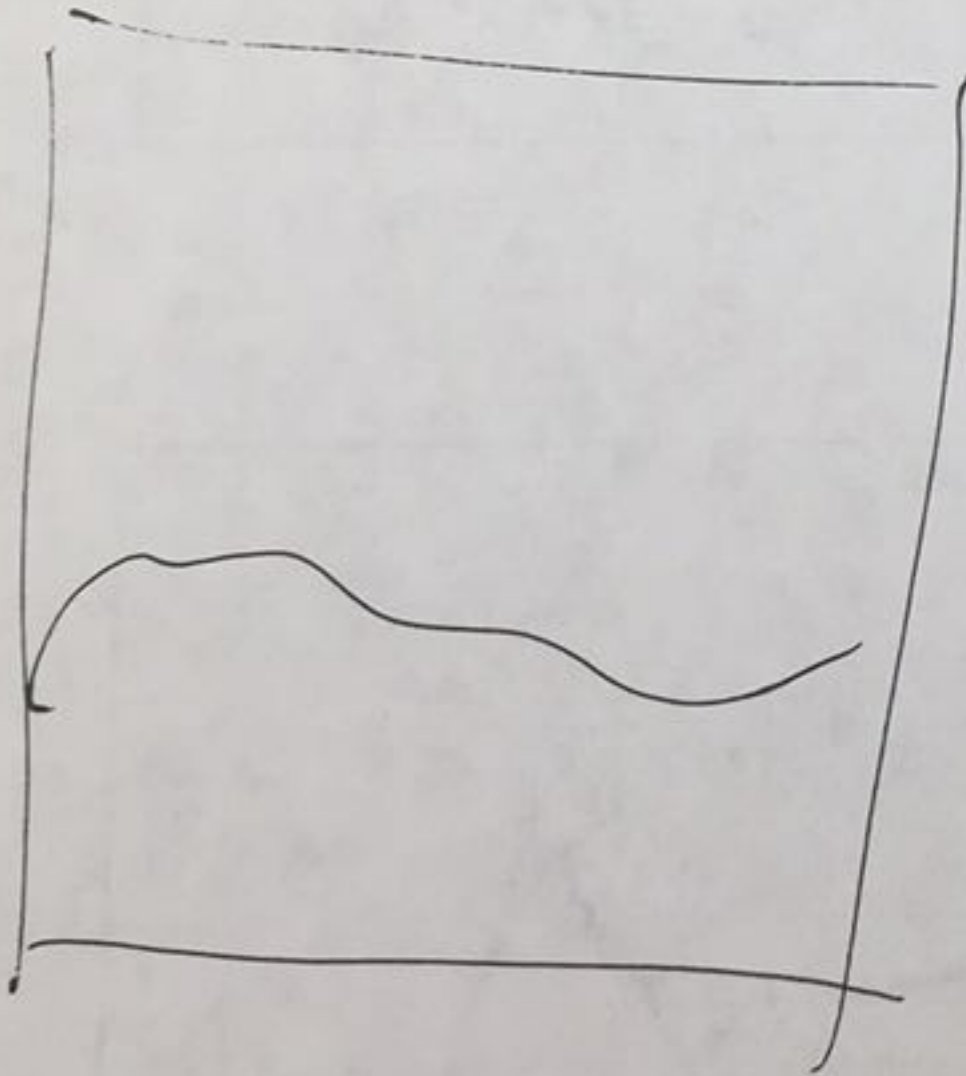
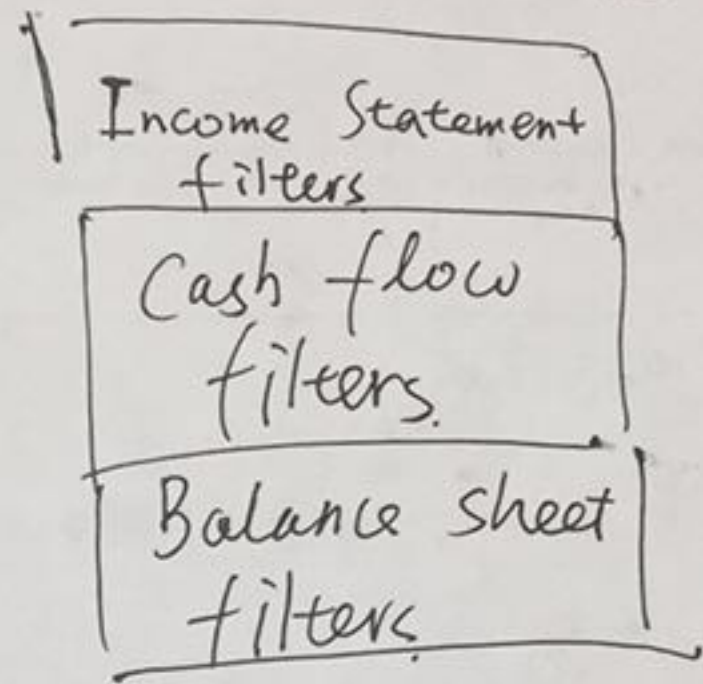
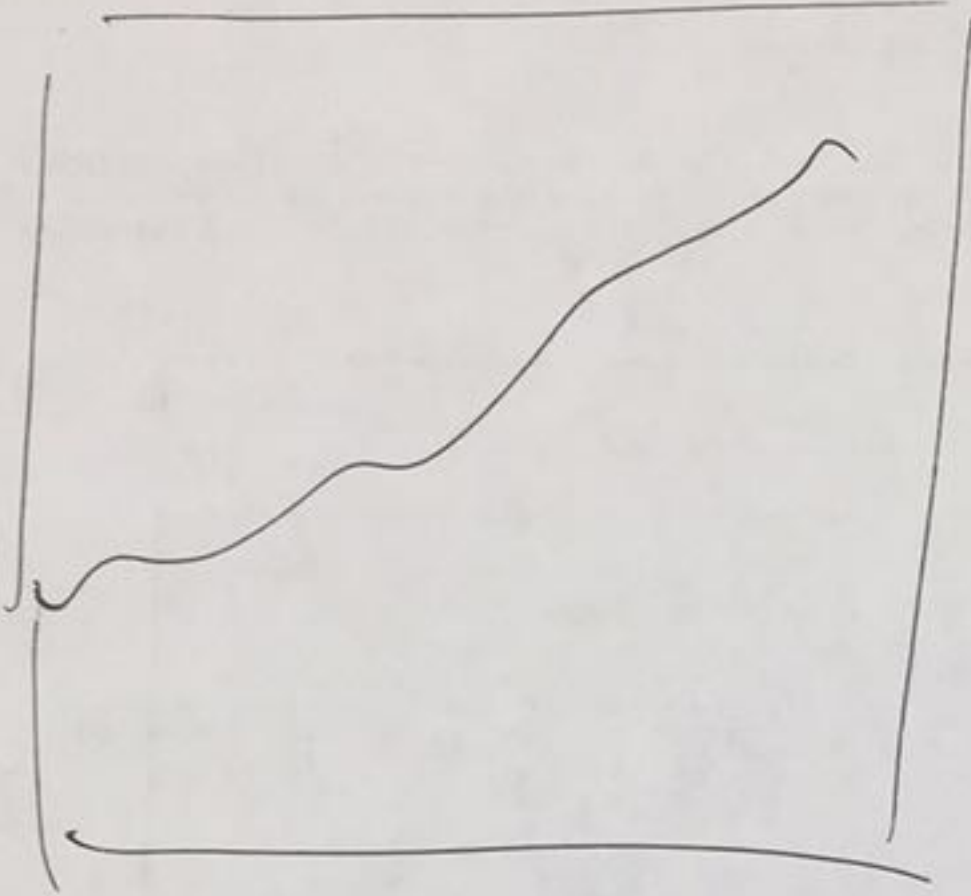
they all have tips.



and so on.

Use this to choose
1 fig view or 2 figs view

1 fig | 2 figs
bottoms

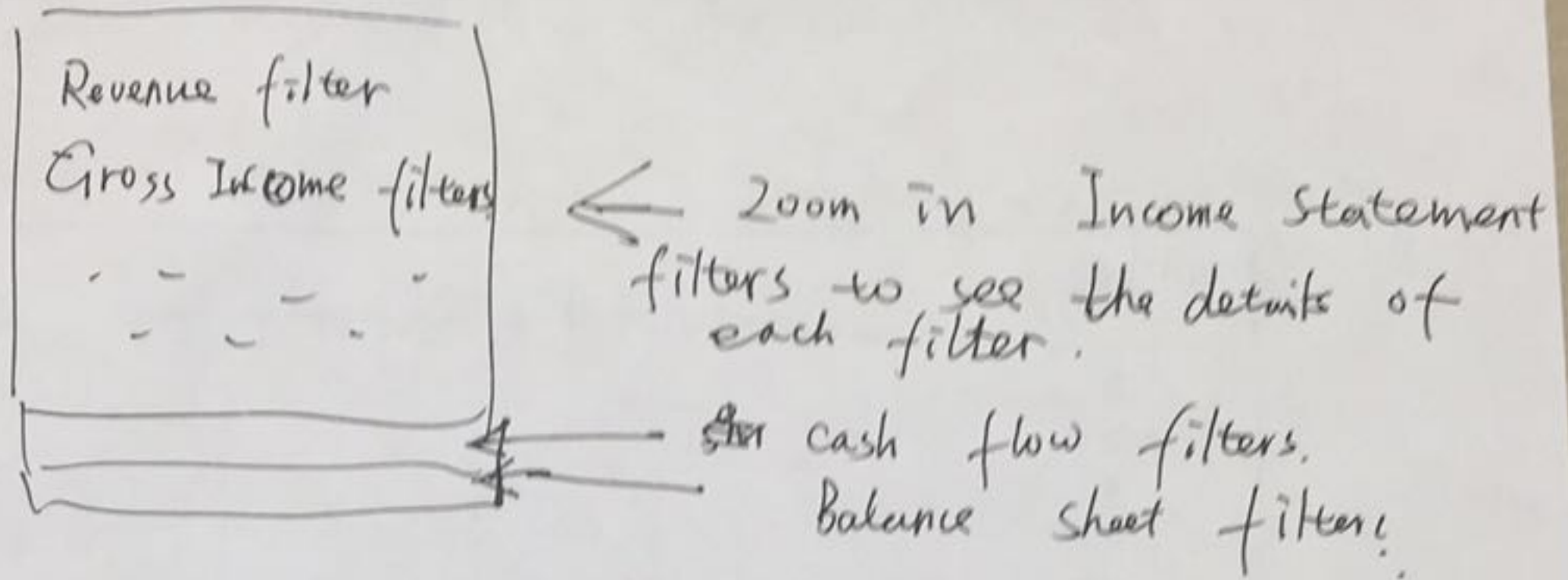


Same as
above.

Desty 2-1.

Q

When focus on Income Statement filters Category.

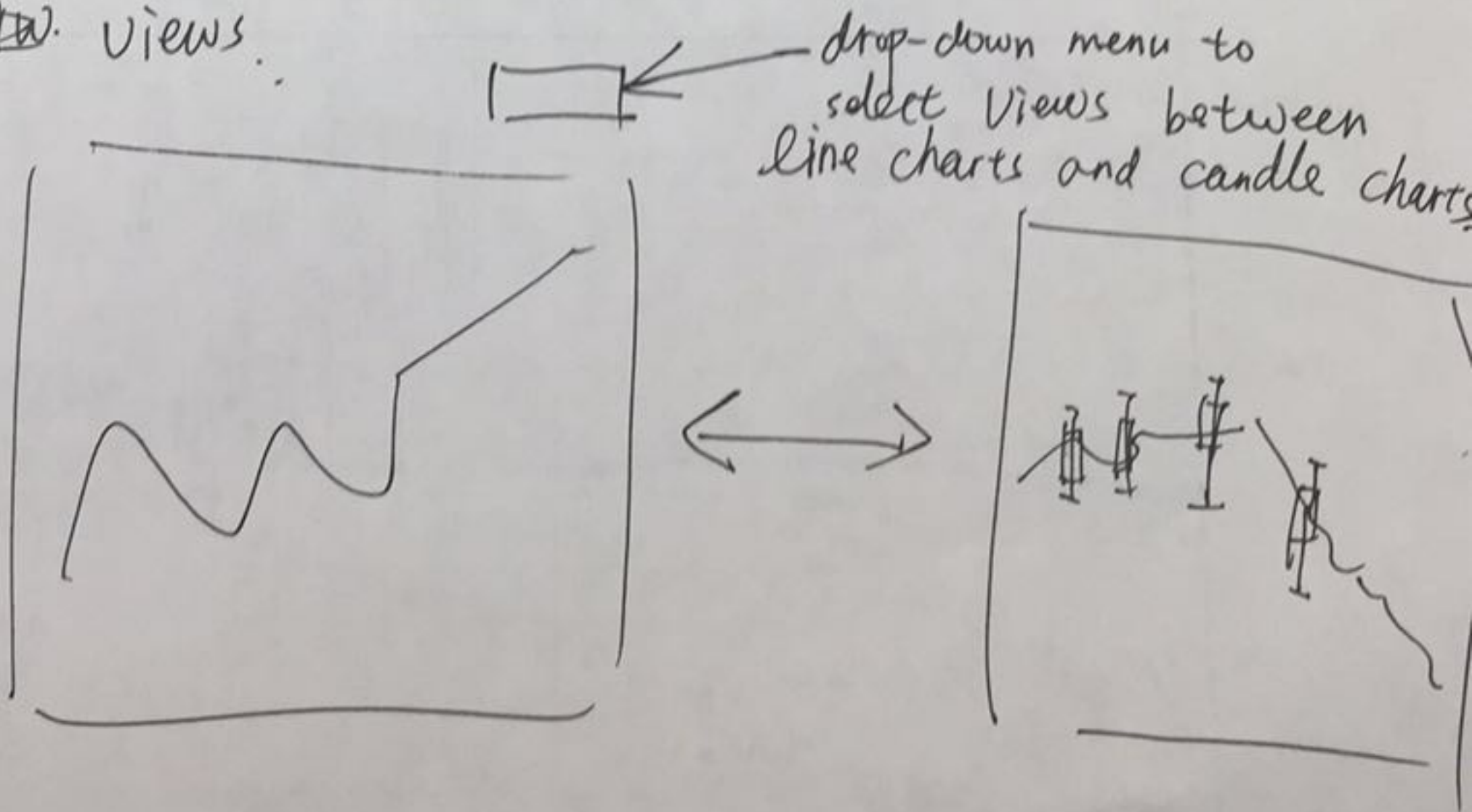


②

③

Figures have transitions between two different

~~view~~ views.

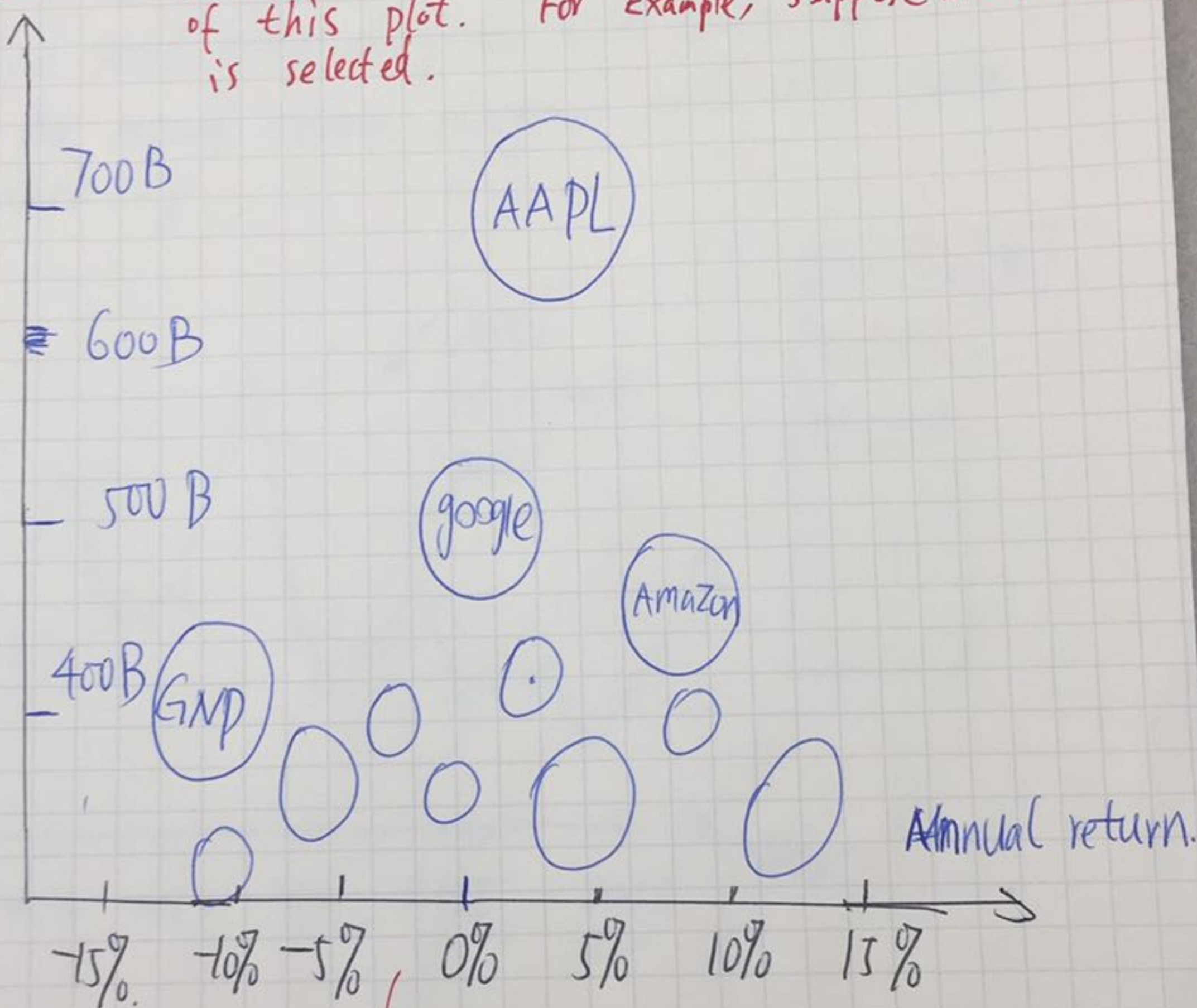


Design 2-2.

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.

