STAT679 Milestone2

Qizheng Xia, Yixuan Wang, Ziyue Zheng

November 18, 2022

1 Choice of Design

1.1 Background

As we discussed in milestone 1, we focus on visualizing financial data which has grown significantly in size and complexity. A single stock may have hundreds of thousands trade per day and thousands of quotes per second. It is extremely high-volume and long-period time-series data for a market. Thus, we would like to use some plots to visualize, such as line chart, candlestick chart in our project.

From the characteristics above, there are two main points we want to solve in our visualization platform. First is to show as much information as possible in a plot, and second is to show the relationship between stocks.

1.2 Data source and pre-processing

Our data is from Yahoo finance API. In tidyQuant package in R, there is a function named getSymbols. It is a wrapper to load data from various sources including Yahoo finance. In practice we let user select their interested stock and time period, then we call this function to require data. The data contains detail information of the selected stock/index in the selected period.

We added a new row call return calculating from (close price-open price)/open price. It will be used in the correlation computing.

1.3 Design

We use shinyapp in R to build our platform, and upload the code in github (Github link: https://github.com/ZiyueZHENG/STAT679_DataVisual.git) and also publish it in the web-ste(https://zzheng232.shinyapps.io/stat679_datavisual/).

Our platform is combined with two part. The first part is a line chart with close prices and a histogram with volume, selected by stock name and date range. We can select any interval data by brush in the line chart, and form a candlestick chart within the relevant interval under the original line chart, including the open price, close price, and the highest and lowest transaction prices at that date. There also form a histogram for volume within the interval under candlestick chart.

For example, in the input panel, we can request data with stock name and date range, like we request the price data of stock AACG in 11-09-2015 and 11-28-2017.

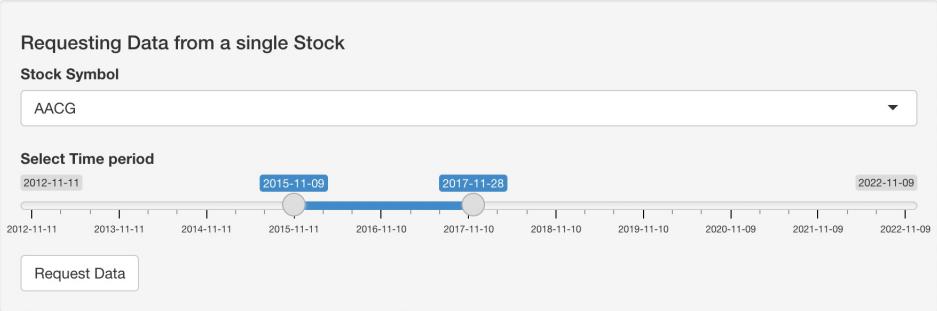

Figure 1: Request data

Then, using this requested data, we build a line chart with closing price and a histogram with volume. We can brush an area in line plot, with selection to finish the candlestick plot.

Figure 2: Line chart and histogram of requested data

After brush, the data in our brush will been selected to finish our candlestick plot. It includes opening price, closing price, maximum and minimum price in a day, which shows much more information than line chart. And the histogram changes its range from the requested date to brushed date. Also, while hovering on the candlestick plot, the right panel shows the exact value of information including prices, volumes and returning percent at this date.

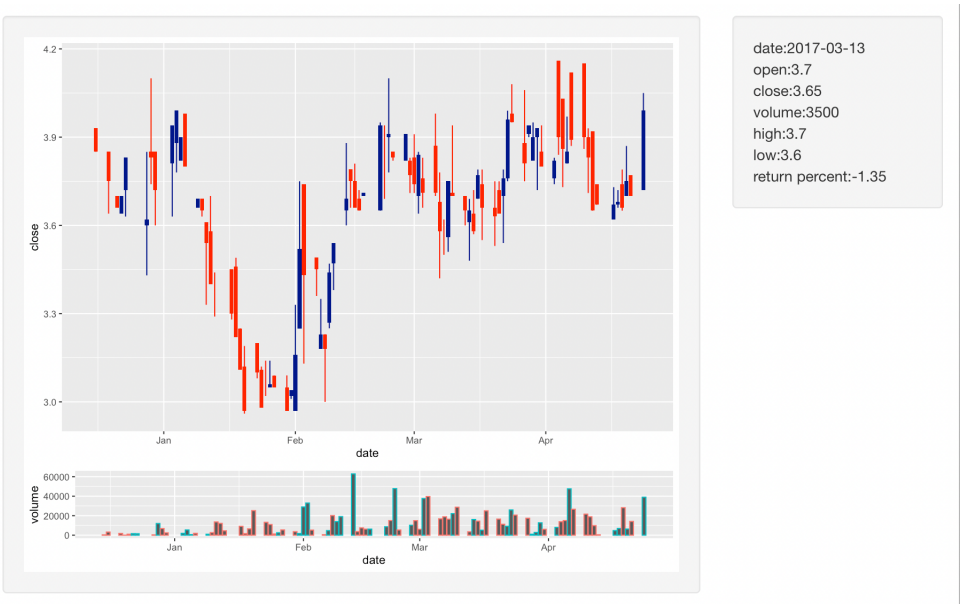

Figure 3: Candlestick chart and histogram of brushed data

In the candlestick plot, we can see the highest price and lowest in one day, and the color represents whether the return percent is positive or negative.

In the second part, we want to know the relationship among multiple stocks, so we use a correlation chart showing the correlation of multiple selected stocks over the selected time frame. We use two inputs, SelectInput to choose stocks and SilderInput to select time. For example, we select AACG, AACI, AADI, AAOI, AAPL and AAWW, 6 stocks for total from 03-01-2022 to 06-20-2022 to calculate their correlation matrix.

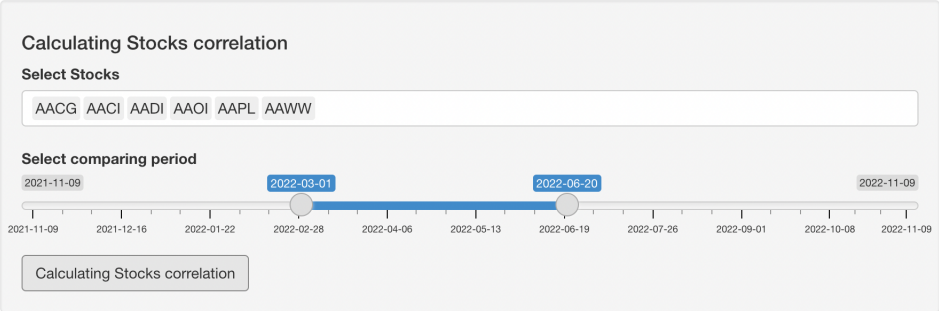

Figure 4: Request data for correlation

In the matrix, the larger the circle is, the stronger correlation is between two stocks. Moreover, if the circle is blue, it means two stocks have positive relationship, while if circle is red, they have negative relationship. For example, in the figure 5, AACG and AAPL have strong positive relationship, nearly 0.8, while AACG has little negative relationship with AACI. Thus, from the matrix, it helps us analysis whether the price of one stock will be effected or effect other stocks, and how the influnce will be like.

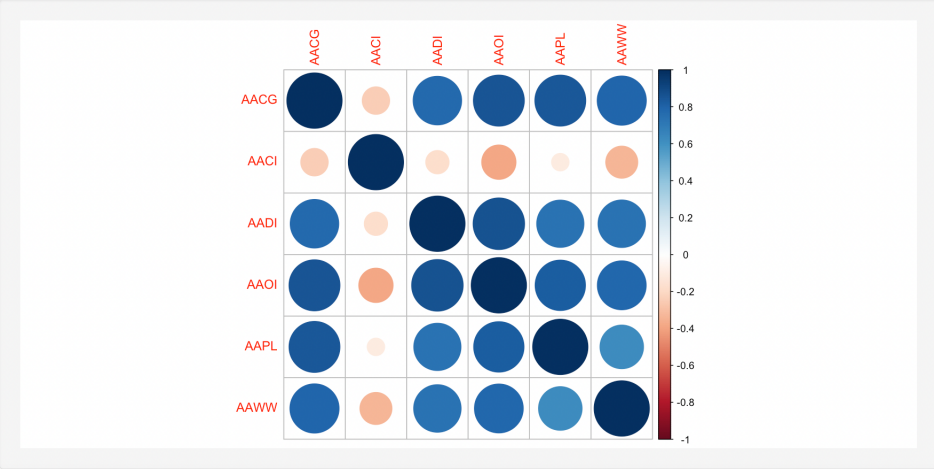

Figure 5: Correlation matrix

2 Trade-off

1. We decided to use candlestick plot rather than line plot because even line chart can show the trend of stock price in a long period, candlestick plot can show different kind of stock prices. So, we use both two plots here, line chart for trend and candlestick chart for detailed information in a small period.

2. We decided to use hover in candlestick plot rather than click because hover is more flexible. When clicking on an exact date, related information is pinned to the right panel, it allows for showing more information. But when we need to see the exact information for multiple days, it will require many repeated clicks which is not that convenient. So we decided to use hover, which also pinned the information to the right when hovering over the exact date, while reducing a lot of repeated clicks.

3 Improvement

3.1 Data

For requesting data for line chart, there are some stocks that might cause errors, like requesting AACIW. This is because the data for this stock is no longer in the Yahoo Finance dataset, possibly due to the company going out of business or no longer providing data, etc.

Since the data of companies is updated day by day, it is impossible to get a certain data of

company's names. For our later improvement, we would like to use some update function to check the validation of company with real-time detection, such as using 'try' function.

3.2 Panel

For our whole visualization, we divided them into two parts, but we put them into the same panel. This may cause the entire panel to appear very cluttered.

For our later improvement, we are going to divide them into two panel by adding a sidebar. We set two options in the sidebar, named as stock price volatility and stocks correlation. While clicking each bottom, it turns to corresponding plots.

3.3 Chart

For the multiple stocks correlation part, it is better to visualize the stocks trend or value at the same time. For example, we could add a line chart or a horizon chart next to the correlation plot.

3.4 Table

For our shinyapp. we show several plots for our visualization, like line chart, histogram, candlestick chart and correlation matrix. They are good at showing time series data and relationship among data sets. But, if people want to see some details of stocks in a range of time, the plots may not work.

For our later improvement, we would like to consider adding some tables in the shinyapp, so people can see more details of information of stocks.