Final Report
CSE 6242
Team 113 (bhurt3, ralbright7, rouldnoughi3, pchen43)
11/26/2019

<div align="center">Team 113: Portfolio Analysis Tool</div>

Introduction and Problem Definition

Our project team built a long-only stock portfolio management tool for analyzing historical performance of a user-defined portfolio of stocks that is geared towards the individual investor. The tool plots the return history, performance metrics, and other important information of a portfolio entered by the user. Our approach focuses on the questions asked by the majority performing this type of analysis, instead of pursuing a more complete response to all the answers an individual investor might ask.

Our tool is not the first such tool for portfolio analysis. There is stock broker software, as well as some other free options such as Portfolio Visualizer[1]. The problem with the current set of available tools is either: monetary (many tools are not free, as with the broker software), or complexity/scope (overly complex for the individual investor who is perhaps less sophisticated, as with Portfolio Visualizer).

Method

Our tool is free, simpler, and easier to use than the currently available options – this is the primary value-add of the tool. Individually, these are not 'innovations' per se, but they are additive to the current set of options, and therefore innovative as a whole.

In our survey of other software and websites that try to perform similar functions, we have found that they are generally more complex, require more sophistication to use, and need more setup (e.g. clicks or uploads) before seeing results. Our tool updates all of its charts and tables as soon as the first position is entered – something that, for example, Portfolio Visualizer does not allow (you must enter the entire portfolio first). Our tool also allows the user to input positions as transactions. This feature results in a portfolio that is dynamic – the user can buy and sell stocks throughout the life of the portfolio, a more realistic alignment to an individual investor's behavior (as opposed to the "all up front" approach of existing software that analyzes the portfolio with all positions held at the beginning).

To accomplish our goal, we have an interface where the user adds positions, and in return gets a series of graphs and tables describing the input portfolio. Below we demonstrate and explain the different pieces of our tool.

---

[1] https://www.portfoliovisualizer.com/

After signing in, the user first must add cash to their portfolio (similar to funding an investment account). After adding cash, which is treated as another line-item in the stock database, the user then chooses a date and "buys" stocks in the portfolio via buy transactions. The user can also "sell" positions via sell transactions (although not more than they currently own – selling short is prohibited). More cash can be added later, and at any point the user may not buy more of a stock than they have cash available (all transactions are done in dollars, number of shares is calculated behind the scenes). The user input field is shown below:

Trade Date: ?	Buy/Sell: ?	Ticker: ?	Dollar Value: ?	Execute: ?
08/18/2010	BUY × ▼	AAL - AMERICAN AIRLINES GROUP INC × ▼	10000	**SUBMIT**

Once at least one stock has been added, the below charts and tables generate (example with a couple stocks added, along with explanations of the information shown):

Portfolio Value ($) and gain (%) charts:
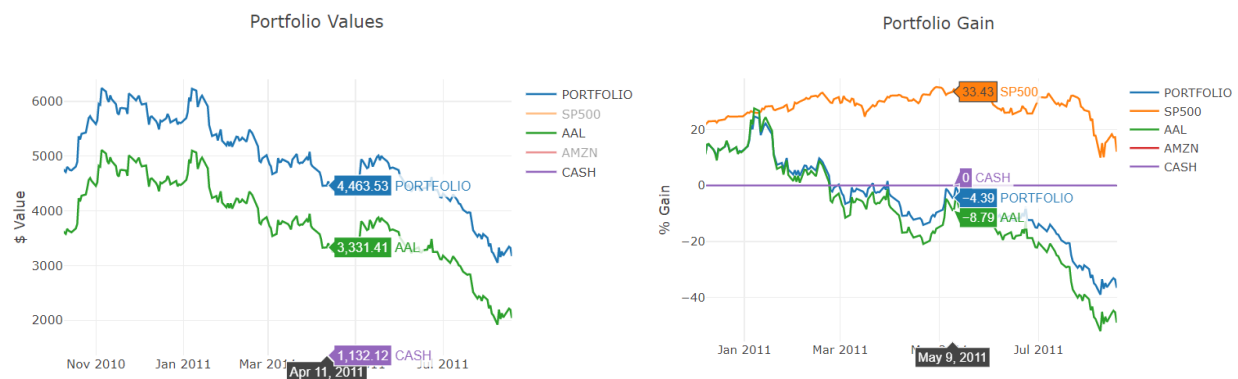
The calculation for these charts are:
- Portfolio Value:
    o Compounded $ return of the individual stocks over the time period
    o Portfolio value as a sum of the individual stocks
- Portfolio Gain:
    o Compounded percent return of the individual stocks over the time period
    o Portfolio percent return
- Both of the above account for changes in value, i.e. buy and sell transactions, along with corporate actions like stock splits, etc.)

Individual stock return bar chart and volatility vs. return scatter plot:
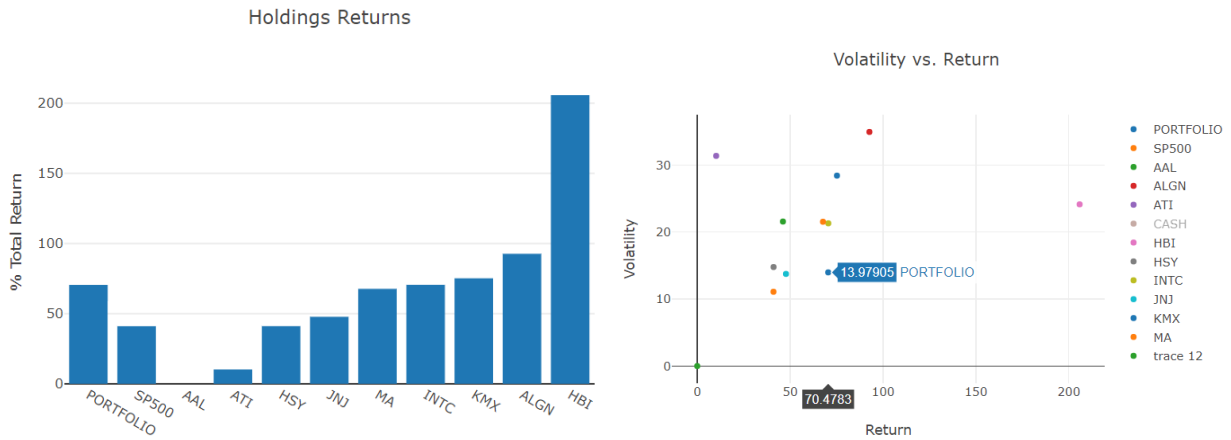
The calculation for these charts are:
- Holdings Returns: individual stock returns over the life of the portfolio from the compounded set of daily returns
- Volatility vs. Returns: Volatility as the standard deviation of the daily returns for each stock, return as calculated in the bar chart. The purpose of this chart is to get a visual feeling of the risk/return trade-off for each stock.

Portfolio-level and individual stock (holdings) metrics tables:

Portfolio Metrics: ?

Date	Value	Tot ret (%)	Sp500 tot (%)	Alpha (%)	Beta	Sharpe	Sortino	Volatility (%)
2011-08-18	$3,172.47	-36.55	12.02	-47.64	0.8	-0.79	-1.12	29.74

Holdings Metrics: ?

Ticker	Date	Value	Portfolio (%)	Tot ret (%)	Sp500 tot (%)	Alpha (%)	Beta	Sharpe	Sortino	Volatility (%)
CASH	2015-01-02	$0.00	0	0	0	0	0	0	0	0
ALGN	2015-01-02	$3,853.27	11.3	92.66	41.94	31.83	1.49	1.11	2.03	34.94
ATI	2015-01-02	$2,203.78	6.46	10.19	40.27	-63.49	1.81	0.31	0.47	31.37
GOOGL	2015-01-02	$2,922.49	8.57	46.12	41.03	3.51	1.05	0.99	1.73	21.56
HBI	2015-01-02	$6,116.12	17.94	205.81	41.76	164.8	1.01	2.44	4.38	24.14

Details on the above:
- Date: as of date of the portfolio. The user can change this using a slider to see the value of the portfolio at different times.
- Value: dollar value of the positions at the specific date
- Portfolio (%): percent of the portfolio in each position
- Tot ret (%): total (compounded and including dividends) return of the position
- Sp500 tot (%): total return of the S&P 500 index during the time period of the position held in the portfolio (will be the same for all positions added on the same date).
- Alpha (%): excess return of each position over the S&P500, after adjusting for beta
- Beta: slope coefficient of the linear regression of the position's daily return on the S&P500 daily return
- Sharpe: Sharpe ratio of the position (excess return / volatility)
- Sortino: Sortino ratio of the position (excess negative return for only days of negative performance / volatility)
- Volatility (%): standard deviation of the stock's return time series

All of the above graphics have hover-over tool tips to display specific information on each. As tooltips, they won't be obtrusive to those who are familiar with them, but educational/helpful for those who want a refresher. On the line charts, specific lines can be clicked on to be hidden from the display. All the charts also have a zoom ability to dive in on specific time periods.

As a note on the data and back-end, we sourced our data (from the Center for Research in Security Prices, "CRSP", and Compustat database) hosted on AWS redshift and stored in a PostgreSQL database. The original (downloaded) database contained daily return time series from 2007

through 2018 for 32,000+ securities. We cut this down to contain the return histories for only current and past constituents of the S&P500 (707 stocks). The original database was 43gb, cut down to ~1gb.

Our website is also hosted from a duo of AWS accounts on AWS Elastic Beanstalk. Since our group prefers python over javascript, we are using the dash libraries (built by plotly) for our visualizations.

Experiments and Evaluation

We took a two-pronged approach towards evaluating our tool: a demonstration of experiments on the tool to show its capabilities, and a user survey to assess its ease of use. Below is a list of the highest priority experiments, and what we expect to demonstrate:

Test portfolios that:	Demonstrates the tool can:
Have many positions	Handle reasonably large portfolios without much latency
Have a gap with no holdings in between two periods of holdings	Calculate sensical metrics over gaps
Enter and exit the same stock multiple times	Calculate sensical metrics over changing position size
Enter positions at different times	Present sensical results over the life of a dynamic portfolio
Try to sell more of a position than it owns	Throw an error when trying to go short a stock
Try to add a stock that is not in the S&P500	Throw an error for adding a non-S&P500 stock

The above tests can be accomplished in two experiments:
1. Create a portfolio that enters and exits a position more than once.
2. Create a portfolio with a large number of securities

Both of these are explained in depth below. For the last two tests shown above, the tool has this functionality, but cannot be easily demonstrated (outside of actual use). Trying to sell more of a position than owned will not be processed, and the only positions that can be added are contained in a dropdown (searchable), these are all members of the S&P500.

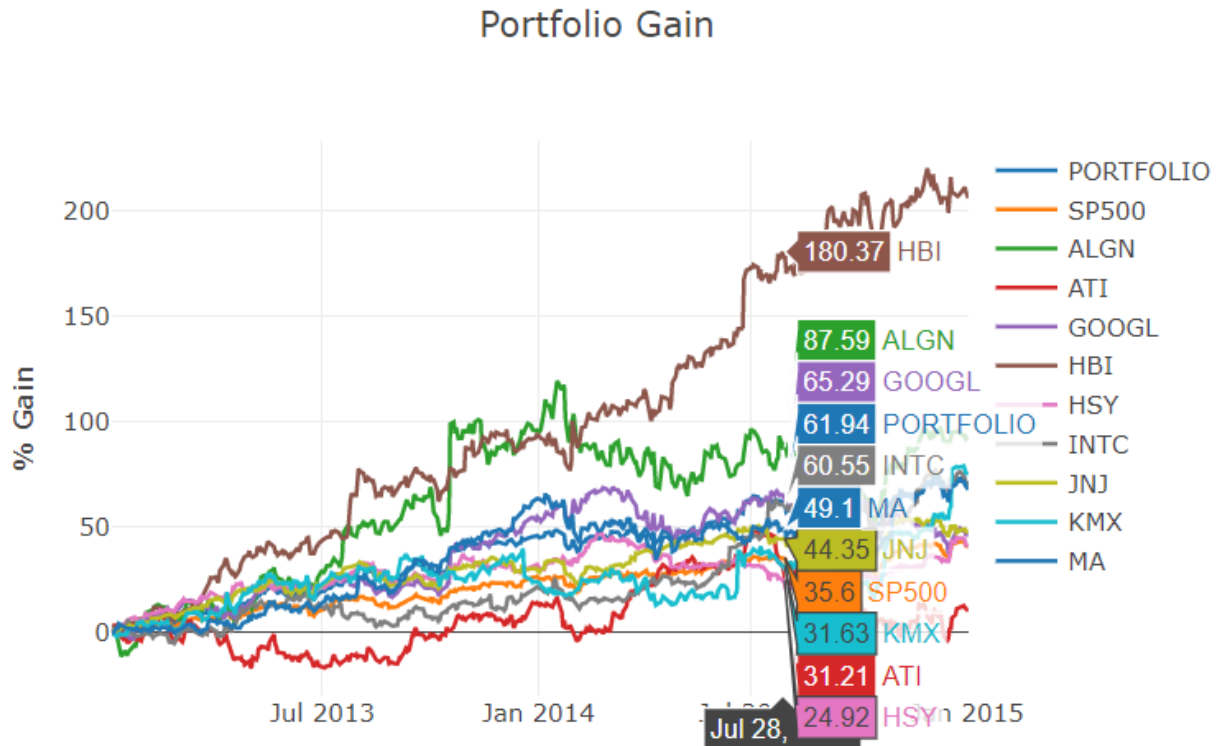**Experiment 1:** Create a portfolio that enters and exits a position more than once.

Below is the Portfolio Values chart from this experiment (other charts not required to show the success). In this experiment, the user started with $5,000 in cash, immediately bought $2,000 of Amazon stock (AMZN) and held it for 2 months. One can see the portfolio value follows the AMZN value, since it is the only holding during this time period.

Around April, the user sold all AMZN stock, cash rebounds up to near $5,000 (just under due to a small loss on the AMZN holding), and the portfolio value then matches cash since it has no non-cash holdings. Finally, the user buys AMZN again in June, this time $3,000, and holds it through August.
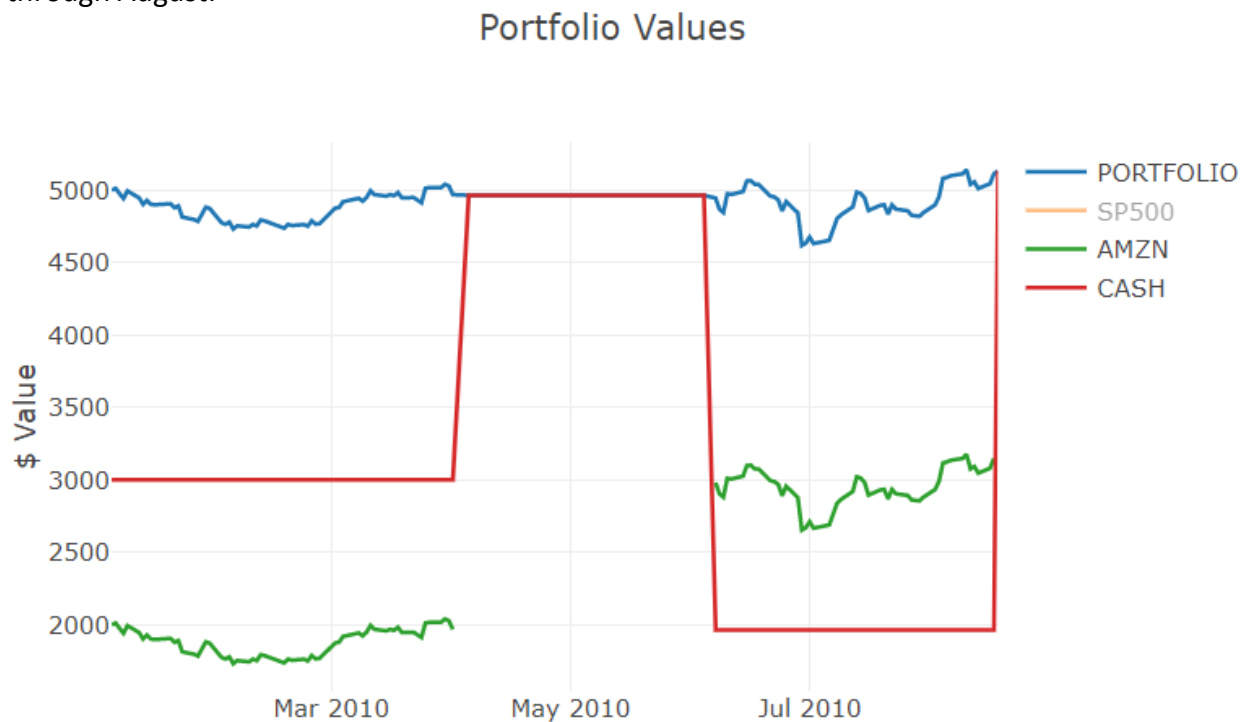
Portfolio Values

Experiment 2: Create a portfolio with a large number of securities

Below the Portfolio Gain chart and Holdings Metrics table are shown for a portfolio holding 9 stocks.

Portfolio Gain

Holdings Metrics: ?

Ticker	Date	Value	Portfolio (%)	Tot ret (%)	Sp500 tot (%)	Alpha (%)	Beta	Sharpe	Sortino	Volatility (%)
CASH	2015-01-02	$0.00	0	0	0	0	0	0	0	0
ALGN	2015-01-02	$3,853.27	11.3	92.66	41.94	31.83	1.49	1.11	2.03	34.94
ATI	2015-01-02	$2,203.78	6.46	10.19	40.27	-63.49	1.81	0.31	0.47	31.37
GOOGL	2015-01-02	$2,922.49	8.57	46.12	41.03	3.51	1.05	0.99	1.73	21.56
HBI	2015-01-02	$6,116.12	17.94	205.81	41.76	164.8	1.01	2.44	4.38	24.14
HSY	2015-01-02	$2,821.66	8.28	41.08	41.58	16.3	0.61	1.24	1.95	14.77
INTC	2015-01-02	$3,410.88	10	70.54	40.46	28.84	1.02	1.36	2.01	21.3
JNJ	2015-01-02	$5,910.09	17.33	47.75	40.64	19.77	0.69	1.49	2.11	13.75
KMX	2015-01-02	$3,503.95	10.28	75.2	40.57	22.89	1.28	1.13	1.65	28.42
MA	2015-01-02	$3,353.42	9.84	67.67	40.8	18.07	1.22	1.31	1.96	21.54

The tool had no trouble adding additional positions and no decrease in wait times was recorded. The hover-over tool tips on the line chart make deciphering the multitude of lines easy (this occurs on all of the charts). Positions added at different times simply start their performance history at different places on the line charts, and are treated in the metrics tables like every other position (showing information from their respective holding periods until the 'current' as-of date).

User Survey

Our own experiments on the webpage were successful for all proposed tests. In order to expand our test base to non-developers of the project, we also surveyed users from other project teams. We received 8 responses (out of an expected 12-18).

Below are the questions we posed to the users, and their responses. Discussion to follow (graphs in the appendix).

How familiar are you with finance/portfolio management tools?

ANSWER CHOICES	RESPONSES	
Not at all familiar	10.00%	1
Not so familiar	0.00%	0
Somewhat familiar	40.00%	4
Very familiar	50.00%	5
Extremely familiar	0.00%	0
TOTAL		10

How much experience do you have using similar tools?

ANSWER CHOICES	RESPONSES	
No experience	10.00%	1
Little experience	0.00%	0
Some experience	50.00%	5
Very experienced	40.00%	4
Extremely experienced	0.00%	0
TOTAL		10

Was the application responsive and quick to load?

ANSWER CHOICES	RESPONSES	
Not at all responsive	10.00%	1
Not so responsive	10.00%	1
Somewhat responsive	20.00%	2
Very responsive	50.00%	5
Extremely responsive	10.00%	1
TOTAL		10

Was it easy to discover all the options and interactions available within the application?

ANSWER CHOICES	RESPONSES	
Very difficult	10.00%	1
Difficult	20.00%	2
Neither easy nor difficult	40.00%	4
Easy	30.00%	3
Very easy	0.00%	0
TOTAL		10

How would you rank the difficulty of completing the task of adding positions?

ANSWER CHOICES	RESPONSES	
Very difficult	10.00%	1
Difficult	20.00%	2
Neither easy nor difficult	50.00%	5
Easy	20.00%	2
Very easy	0.00%	0
TOTAL		10

The visualizations/graphics are easy to understand and interpret.

ANSWER CHOICES	RESPONSES	
Strongly disagree	0.00%	0
Disagree	10.00%	1
Neither agree nor disagree	20.00%	2
Agree	60.00%	6
Strongly agree	10.00%	1
TOTAL		10

Did the application display enough information? Was there information that you would have liked to see that was not included?

ANSWER CHOICES	RESPONSES	
Lacking a lot of information	20.00%	2
Lacking some information	50.00%	5
Adequate information provided	20.00%	2
Some additional information was included	0.00%	0
Lot of additional information was included	10.00%	1
TOTAL		10

Summarization of additional comments:
- Of the comments, users indicated that they would have liked some more direction at the beginning of the process (that they needed to add a cash position to the portfolio, from which the value of their positions would be drawn).

As one can see from the above, generally the reviews were positive – most importantly, there was consensus (among our small sample of users) that the graphics were easy to interpret and the site was responsive. Of the negatives, users were looking for some additional usability help – in particular adding positions. To address this, further improvements include additional explanations and walk throughs.

Results and Conclusions

From our experiments, the tool is fast and delivers the right range of information: enough to answer the essential questions for portfolio analysis (i.e. "what are my returns and basic exposure to risk?"), but not too much to slow the application down, or create unnecessary complication.

The tool has a bright future: there are improvements to be made (e.g. speeding up the actions, adding ease-of-use features like an overlay or first-time-user walk through), but importantly, it provides utility in its current state. It has an advantage over broker software by being free, and it has an advantage over Portfolio Visualizer by being simple, quick, easy to understand.

Distribution of Effort:

Task	Completed by:
Get data (from the CRSP and compustat database)	ralbright7
Setup in our environment (host on AWS, setup on flask)	ralbright7
Munge data – remove non-S&P500 constituents	ralbright7
Write python/sql processes to calculate data needed for visualizations	ralbright7
Feed data into the web page	ralbright7
Create, review data feed triggers to webpage	rablright7, rouldnoughi3
Create primary line chart	bhurt3
Create individual returns bar chat	bhurt3
Create portfolio/holding metrics table	bhurt3, pchen43
Create position scatterplots	bhurt3
Coordinate/create/execute/analyze peer survey	pchen43
Proposal/presentation/progress report/poster/final report	bhurt3
Add tooltips	ralbright7
Add line chart for % return amount	ralbright7
Add user input fields	ralbright7
Deploy webpage to AWS	rouldnoughi3

Survey

Given the more applied nature of our project, our survey of existing research gravitated towards two broad categories: visualization techniques for time-series and financial data, and existing (academic-generated) tools for visualizing time-series and financial data.

Visualization Techniques:

There is a plethora of visualization techniques to consider[2]. In the paper by D. Marghescu, many of the techniques we plan to use are demonstrated. Single and multiple line charts, scatter plots, and versions of heatmaps all reside in the realm of easy-to-understand and broadly familiar visualizations that we are looking to employ.

Our goal in this research was to find potential visualizations to add to our tool while keeping in mind that simplicity of use and understanding is the primary goal. Interactive graphics like treemaps, where you can dive down limbs of data with increasing specificity[3], and factor heatmaps, that use color to indicate sensitivity to various factors[4], both have potential to be added to our tool. The shortcomings of techniques like these are that while they may offer new perspectives, we must be careful that they do not materially detract from the straightforward presentation of our data and are not onerous for the user to decipher.

There is another category of potential visualizations which are those that apply advanced techniques to depict data but require a non-insignificant amount of training and understanding to glean information from. Graphics like pixel-based box models[5], pixel graphics[6], clustering[7], and correlation distributions and graph representations[8][9], may provide results with increased resolution of the dynamics of the data, however the loss of simplicity to achieve these gains generally outweighs their usefulness.

---

[2] Marghescu, D. (2007). Multi-dimensional Data Visualization Techniques for Exploring Financial Performance Data. *AMCIS*.

[3] Jungmeister, W-A., & Turo, D. (1992) Adapting Treemaps To Stock Portfolio Visualization. Digital Repository at the University of Maryland

[4] Van Gelderen, E., Huij, J., Kyosev, G., (2019) Factor Investing from Concept to Implementation. The Journal of Portfolio Management, Quantitative Special Issue

[5] Ziegler, H., Nietzschmann, T., & Keim, D.A. (2007). Relevance Driven Visualization of Financial Performance Measures. *EuroVis*.

[6] Ziegler, H., Nietzschmann, T., & Keim, D. A. (2008). Visual Analytics on the Financial Market: Pixel-based Analysis and Comparison of Long-Term Investments. *2008 12th International Conference Information Visualisation*, 287–295.

[7] Lemieux, Victoria & Rahmdel, Payam & Walker, Rick & Wong, B.L. & Flood, Mark. (2014). Clustering Techniques And their Effect on Portfolio Formation and Risk Analysis. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 1-6.

[8] Simon, P. M. and Turkay, C. (2018), Hunting High and Low: Visualising Shifting Correlations in Financial Markets. Computer Graphics Forum, 37: 479-490.

[9] Abrams, J., Celaya-Alcula, J., Baldwin D., Gonda, R., Chen, Z. (2016) Analysis of Equity Markets: A Graph Theory Approach. ResearchGate Working Paper.

Existing Tools:

Investors with large brokerage or paid-for broker software and institutional investors have access to expensive and often proprietary tools oriented towards stock analysis. The tools available to the individual, main-street, investor are much more limited. Our survey of existing tools, though limited in the most part to academic publications, provides insight into the types of platforms being developed for visualizing data. These are typically stand-alone applications that employ novel visualizations such as a combination of line and pixel graphics using non-overlapping shapes[10], cluster-based factor mapping[11], and probabilistic forward-looking outcomes based visualizations aimed at showing likely financial performance[12] (this in particular is a worrisome endeavor given uncertainty in the market and the danger of a paradigm shift that is unobservable ahead of time). These tools are the focus of academic pursuit and do not appear to be readily available to the public market for free, if at all. They forward the progress of creating new ways to visualize performance (both forward- and backward-looking), but in exchange they forfeit simplicity. They are helpful, though, in providing high-level inspiration for avenues of visualization (e.g. focus on correlation, or volatility relative to size), as well as cautioning against attempting some of their practices.

Finally, we investigated another tool for assess time series (non-specific to finance), that endeavors to replace typical time series visuals with multi-scale drill-down graphics[13]. Analytics of pure timeseries is useful for understanding a portfolio's return progression, although more than just time series analysis is needed and applying the advancement from a tool like this would pass beyond the scope of our project – this paper serves to underline that many of these cutting edge techniques, while worth consideration are likely too complex to include. Our research into many of the current and new tools being written about does less to inform what we should include, and more to help define the problem - the actual need in the market is for a simple tool that is straightforward and easy to use.

---

[10] Schäfer, M., Kahl, R., Zhang, L., Schreck, T., & Keim, D.A. (2011). A Novel Explorative Visualization Tool for Financial Time Series Data Analysis.

[11] Yue, X., Bai, J., Liu, Q., Tang, Y., Puri, A., Li, K., & Qu, H. (2019). sPortfolio: Stratified Visual Analysis of Stock Portfolios. *IEEE Transactions on Visualization and Computer Graphics*, 1–1.

[12] Rudolph, S., Savikhin, A., & Ebert, D. S. (2009). FinVis: Applied visual analytics for personal financial planning. *2009 IEEE Symposium on Visual Analytics Science and Technology*, 195–202.

[13] Walker, James & Borgo, Rita & Jones, Mark. (2015). TimeNotes: A Study on Effective Chart Visualization and Interaction Techniques for Time-Series Data. IEEE transactions on visualization and computer graphics. 22.
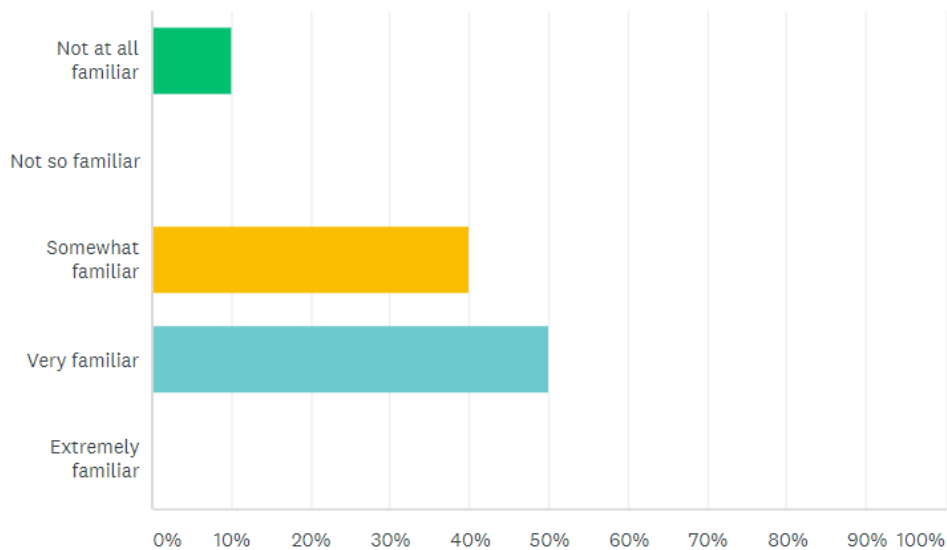
Appendix

Charts from the survey (same information, easier to consume):

How familiar are you with finance/portfolio management tools?
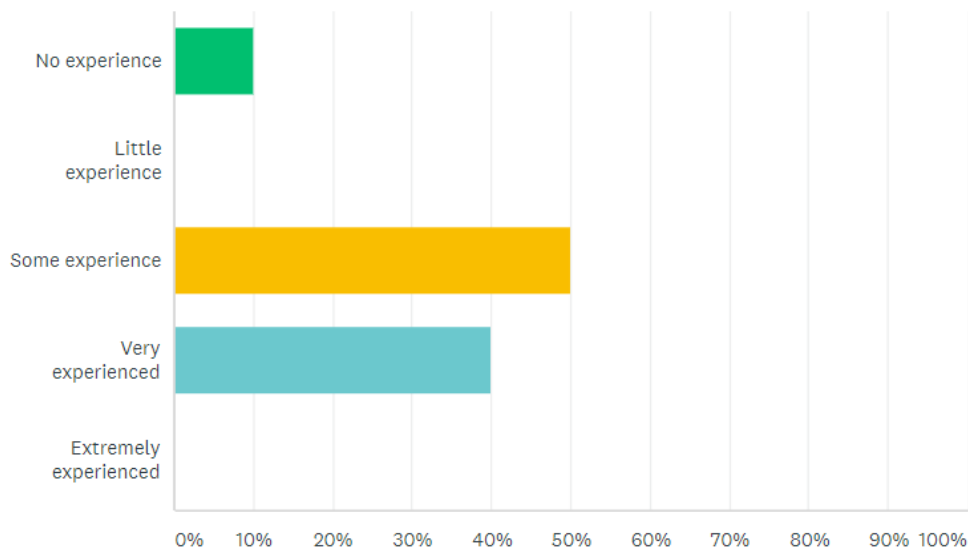
Answered: 10   Skipped: 0

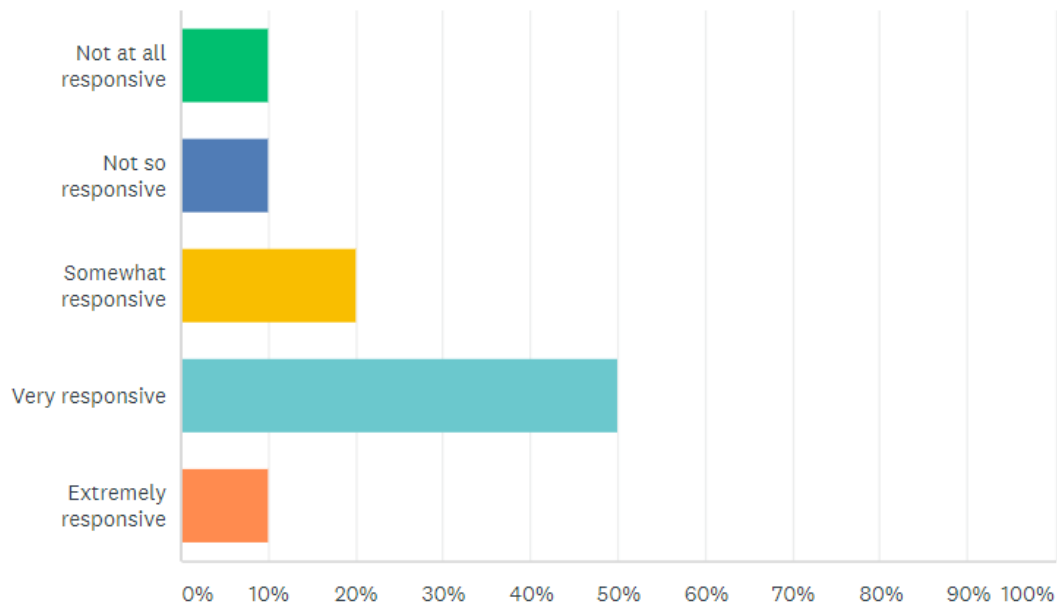# How much experience do you have using similar tools?

Answered: 10   Skipped: 0

Q3

Was the application responsive and quick to load?

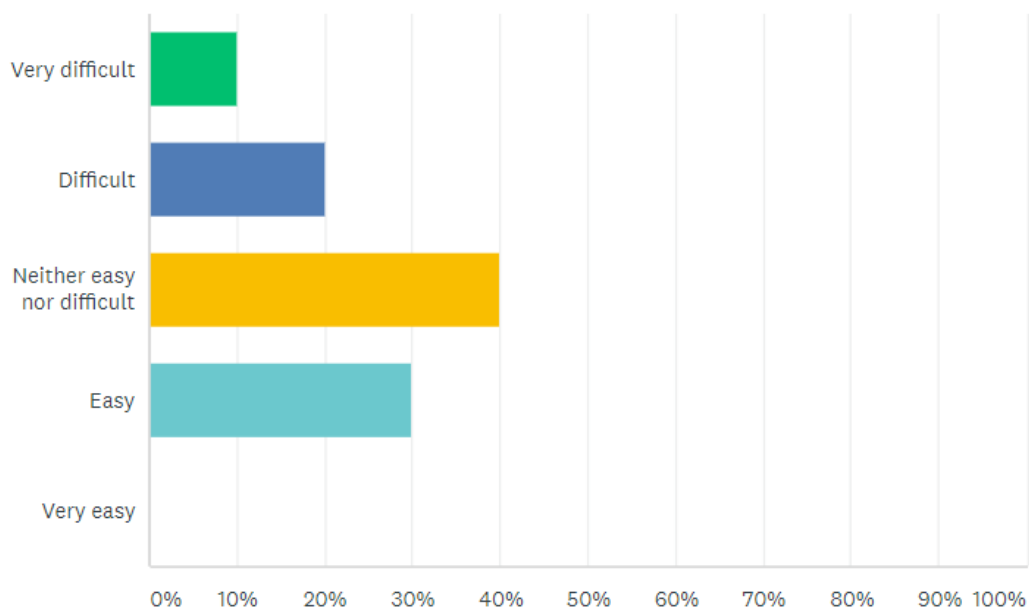Answered: 10   Skipped: 0

Q4

Was it easy to discover all the options and interactions available within the application?
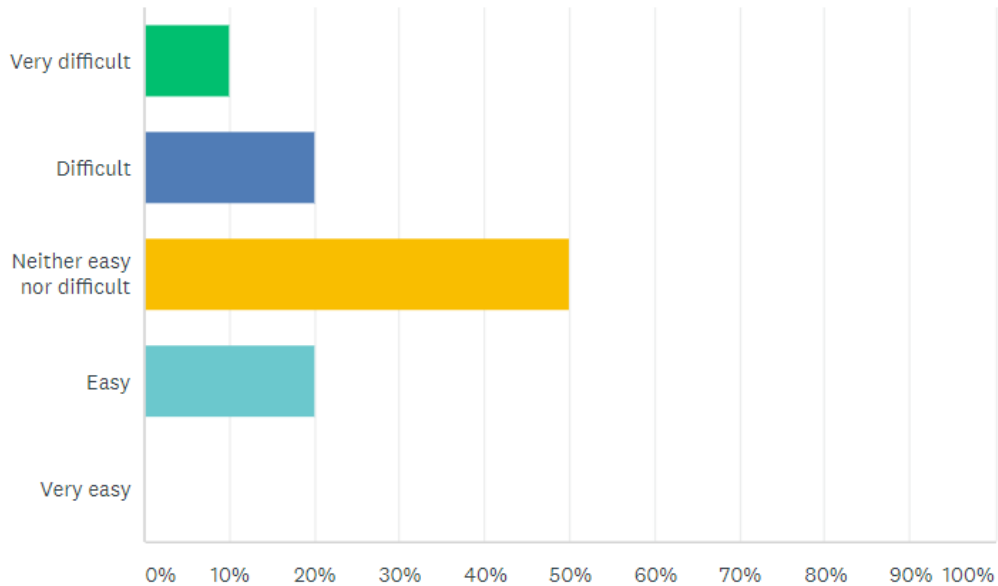
Answered: 10   Skipped: 0

Q5

How would you rank the difficulty of completing the task of adding positions?
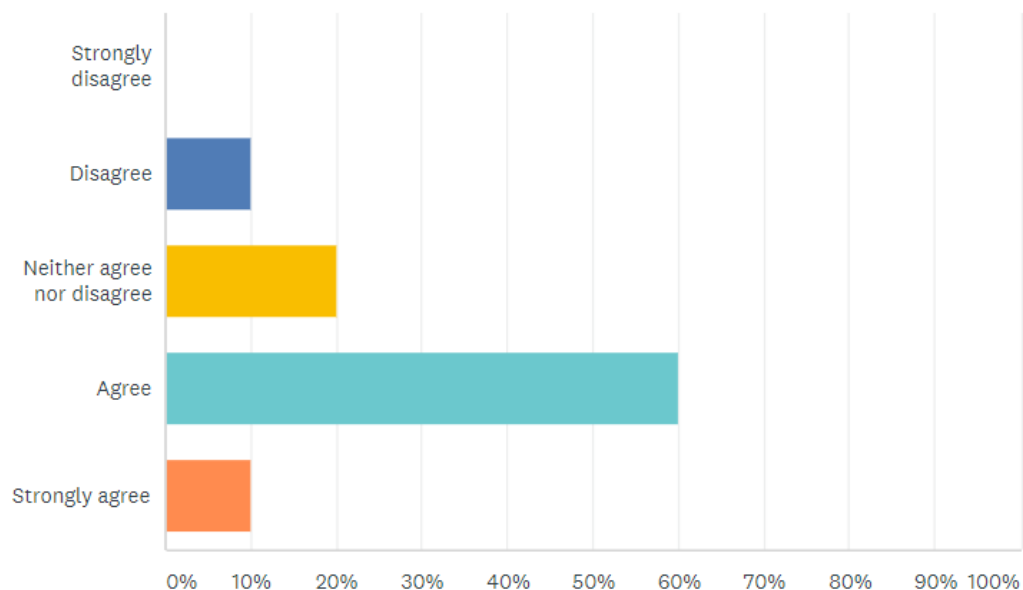
Answered: 10   Skipped: 0

Response	Percentage
Very difficult	10%
Difficult	20%
Neither easy nor difficult	50%
Easy	20%
Very easy	0%

Q6

The visualizations/graphics are easy to understand and interpret:

Answered: 10   Skipped: 0

Response	Percentage
Strongly disagree	0%
Disagree	10%
Neither agree nor disagree	20%
Agree	60%
Strongly agree	10%

Q7

Did the application display enough information? Was there information that you would have liked to see that was not included (feel free to detail in Question 8)?

Answered: 10   Skipped: 0