Titre: Title:	HFT Data Visualization
Auteur: Author:	Javad Yaali
Date:	2021
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Yaali, J. (2021). HFT Data Visualization [Master's thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/9728/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/9728/
Directeurs de recherche: Advisors:	Thomas Hurtut
Programme: Program:	Génie informatique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

HFT Data Visualization

JAVAD YAALI
Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Novembre 2021

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

HFT Data Visualization

présenté par **Javad YAALI**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Jinghui CHENG, président

Thomas HURTUT, membre et directeur de recherche

Manuel MORALES, membre

DEDICATION

This thesis work is dedicated to my dearest friend, Som, who has been a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having him in my life.

ACKNOWLEDGEMENTS

Throughout the writing of this thesis I have received a great deal of support and assistance. I would first like to thank my supervisor, Professor Thomas Hurtut, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

Also, I would like to appreciate our collaborator, Professor Vincent Gregoire for his guidance and support during several challenge of this work and his help for knowing better the financial industry.

In addition, I acknowledge all my committee members, Professor Jinghui Cheng and Professor Manuel Morales for their valuable time that devoted to review and evaluate my research works.

Finally, I deeply appreciate supports that I receive from my family during the challenges that I had in my life.

RÉSUMÉ

Le trading à haute fréquence (HFT, pour “High Frequency Trading”) est un type de trading basé sur des infrastructures à haut débit. Les machines de trading génèrent d’énormes quantités de messages de trading difficiles à explorer pour les chercheurs financiers et les traders. Les outils de visualisation des données financières se concentrent généralement sur la gestion de portefeuille et l’analyse des relations entre le risque et le rendement. Outre la relation risque-rendement, il existe d’autres caractéristiques qui attirent les chercheurs financiers comme la liquidité[1] et les événements de mouvements rapides des indicateurs de marché. Les chercheurs HFT peuvent extraire ces caractéristiques des données HFT car elles montrent tous les détails des mouvements du marché. Dans ce mémoire, nous proposons une méthodologie de visualisation, que nous appelons HFTViz, conçue pour aider les chercheurs en données financières à explorer l’ensemble de données HFT fournies sur le marché d’échange NASDAQ. HFTViz fournit un tableau de bord complet visant à faciliter l’exploration des données HFT. HFTViz propose un design à deux niveaux. Premièrement une vue d’ensemble du marché à une date précise. Deuxièmement, après avoir sélectionné quelques actions à étudier en détail, la visualisation fournit également une vue détaillée des messages de trading, des volumes de trading et des mesures de liquidité. Nous évaluons la méthodologie proposée en réalisant une études de cas regroupant cinq experts du domaine, dans laquelle nous illustrons l’utilisabilité de HFTViz.

[1] Dans les affaires, l’économie ou l’investissement, la liquidité du marché est une caractéristique du marché par laquelle un individu ou une entreprise peut rapidement acheter ou vendre un actif sans provoquer de changement radical dans le prix de l’actif. La liquidité implique le compromis entre le prix auquel un actif peut être vendu et la rapidité avec laquelle il peut être vendu.

ABSTRACT

High Frequency Trading (HFT), mainly based on high speed infrastructure, is one element of the trading industry. However, trading machines generate enormous quantities of trading messages that are difficult to explore for financial researchers and traders. Visualization tools of financial data usually focus on portfolio management and the analysis of the relationships between risk and return. Beside risk-return relationship, there are other features that attract financial researchers like liquidity[2] and moments of flash crashes in the market. HFT researchers can extract those features from HFT data since it has shown every detail of the market movement. In this work, we present HFTViz, a visualization tool designed to help financial researchers explore the HFT dataset provided on NASDAQ exchange. HFTViz provides a comprehensive dashboard aimed to facilitate HFT data exploration. HFTViz contains two sections. It first proposes an overview of the market on a specific date in order to enable user for selecting different stocks. After selecting some stocks to investigate in detail, HFTViz also provides a detailed view of the trading messages, the trading volumes and the liquidity measures. Also, in the details panel, HFTViz provides synchronous hovering and zooming for controlling the level of details for all stocks simultaneously. In order to evaluate our work, we did two case studies gathering five domain experts and we illustrate the usefulness of HFTViz, which eases their analysis process [1].

[2]In business, economics or investment, market liquidity is a market's feature whereby an individual or firm can quickly purchase or sell an asset without causing a drastic change in the asset's price. Liquidity involves the trade-off between the price at which an asset can be sold, and how quickly it can be sold.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

HFT	High Frequency Trading
ETF	Exchange Traded Funds
S&P	Standard and Poor
NASDAQ	National Association of Securities Dealers Automated Quotations
CME	Chicago Mercantile Exchange
LOB	Limit Order Book
HTML	HyperText Markup Language
CSS	Cascading Style Sheets
UCD	User-Centered Design
AT	Algorithmic Trading

CHAPTER 1 INTRODUCTION

1.1 Context and Motivation

Early in trading, exchanges were a place for traders to meet and trade their products. Also, this concept is the same in floor-based trading. In floor-based trading, market participants meet each other and trade their securities and assets. In the past twenty years, with the growth of internet and computing power, trading experience significantly changed. Nowadays, trading is automated and has different levels of speed for traders. By automating trading, we have two emerging areas that leverage speed and automation for trading such as algorithmic trading (AT) and high-frequency trading (HFT) [3].

Algorithmic trading changes the way that investors trade in the market. It generates orders for trading automatically. Many companies used to leverage its power on sell side[1] for years. However, new market access model enables buy side[2] of the market to use algorithms for trading decision making. They can design their own algorithms or use standard software solutions from independent software vendors (ISV). Beside that, the sell side companies still use the majority of algorithms for their clients. Using algorithms and automation reduce trading costs for investors. Because of that, AT became popular in recent years [3].

High frequency trading (HFT) has gained lots of attention in the past fifteen years. Most of the attentions is because of the flash crash in the U.S market on May 6, 2010. Basically, AT focuses on automating execution of the clients orders but HFT relates to the implementation of special strategies by advance market participants in technology. With this explanation, HFT can be considered as a subgroup of AT. However, both AT and HFT help participants to use technology capacity for speeding up different trading aspects like reception of market data, financial analysis, order submission and reception of execution confirmations [3]. Beside that, High Frequency Trading (HFT) has become one of the most profitable ways for trading. For example, the BYX exchange reduced its order processing time sevenfold, from 445 μs in 2009 to 64 μs in 2018. Likewise, the round-trip communication time between Nasdaq and the Chicago Mercantile Exchange (CME) has nearly halved, from over 14.5 ms in 2010 to 7.9 ms today [4]. We can name different usages for HFT data like trading, financial market regulation and research on different events in the market. Traders want to implement different strategies to gain more from the market. The result of this purpose can then be

[1] In this side of the market, companies are trying to create, promote and sell traded securities to public.

[2] Also, in this side of the market, participants are trying to invest large amount of money in different securities for fund management.

used to influence the improvement of their strategies and algorithms. HFT data is useful to regulate the financial markets since it contains all details of the market in fine time scales. The regulators can investigate market data and detect financial frauds. Also, academic researchers can investigate current market micro-structure and propose novel models for analysing financial market.

Beside beneficial usage of the HFT data, there are some challenges for users that are working with the data. Loading large data, cleaning and extracting different metrics from them can be overwhelming for the users. Also, there are different users that want to observe stocks liquidity to make decision about their trading strategy or research hypothesis. As a research team, we are interested in understanding the practices, challenges and needs of HFT users in order to design a tool to facilitate their tasks using HFT data. In order to fulfill this goal, we conducted different interviews with two domain experts. As a result, we divide our target users into two groups. One is the financial researchers that work in academia. The second group is financial analysts who work with HFT data in order to make optimal financial decisions to reduce risk and increase return. Using this knowledge, we designed HFTViz. To evaluate our work, we conducted two case studies to understand if our design choices are fulfilling the users needs.

1.2 Research Objectives

HFTViz provides a visual interface to HFT data. It enables users to investigate several levels of time aggregation, i.e. from hours to milliseconds. Altogether, research objectives of HFTViz are:

1. An HFT data visualization design enabling the exploration and understanding of financial information like price changes, liquidity measures and number of quoting messages in different timescales from milliseconds to hours;

2. HFTViz, a prototype for HFT data exploration;

3. A case study of HFTViz gathering five experts in financial data analysis.

To accomplish previously mentioned objectives, we conducted multiple interviews to identify users needs. Then, we started designing the visualization. After finalizing the design, we evaluate it by organizing two case studies with five domain experts. At the end, we proposed a visual design that enables users to investigate HFT data in details.

1.3 Thesis Outline

In the following chapter of this thesis, we present a literature review of previous studies on user-centered design, financial data visualization and time series visualization. After literature review, in chapter 3 (based on the paper we have submitted to Information Visualization Journal) we discuss design methods that we used for HFTViz and evaluation results. Chapter 4 present a discussion around the lessons that we learned during the process and future improvements that we can provide. Finally, chapter 5 provides a summary from our work and concludes the main points of this research.

CHAPTER 2 LITERATURE REVIEW

In this chapter, we review the related works to HFTViz. In the first section, we present User-Centered Design (UCD). After that, we review the related works in financial market visualization. Finally, we finish this chapter by mentioning related works in time series visualization.

2.1 User-Centered Design

User-centered design (UCD) is a phrase that describes a methodology for helping designers to create better tools to answer the users needs. This method can be used in different industries and products. This process concentrates on potential users to involve them in the design steps. This user involvement can be varied from an occasional user advice to intensive participation in product design. The UCD concept introduced by Donald Norman from his research laboratory at the University of California at San Diego (UCSD) and was advanced in his book "The Design of Everyday Things" [5]. In [5], the authors introduced two main points that UCD is concentrating on. The first one is identifying a right problem. For solving that problem, the second point is appeared. The second point focuses on a way of solving identified problem which has to meet the human needs and potentials. He also adds to UCD concept in his book " The Psychology Of Everyday Things (POET)" [6]. In POET, he suggests four main points to add more usability into the UCD design which we briefly mention below [6].

1. At each step, make it easy to find all possible actions
2. Produce a design that conceptual model of the system, alternative actions and the results of actions are visible.
3. Facilitate system states evaluation
4. Create a reasonable mapping between intentions and needed actions; between actions and their results; and between the information that is visible and interpretation of the system state. [6, 7]

UCD process could be explained as a four steps iterative process starting from observing users, generating idea, prototype and test. After testing the design, designers find the designs

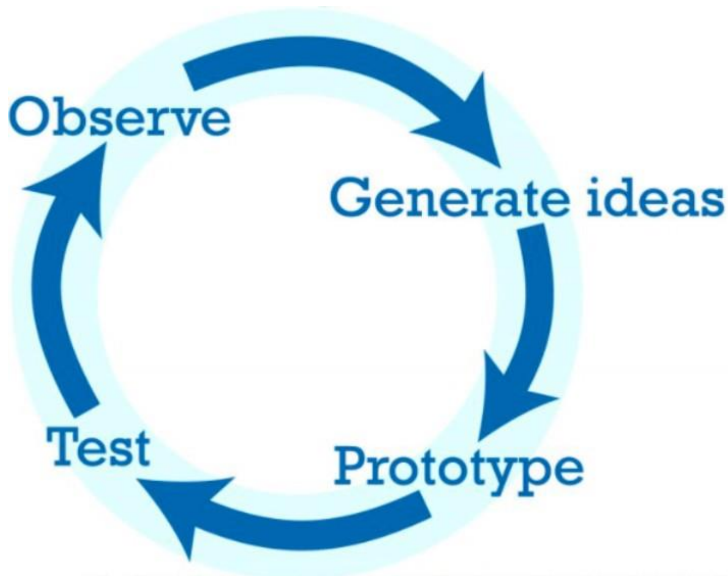

Figure 2.1 UCD process steps [2]

problem and by observing users, they can refine the problems with generating ideas and producing another prototype. (Figure 2.1) [5].

In addition to Norman's works, other researchers described more key features for user centered design. Gould and Lewis [8] work suggested three points that are vital for UCD process which we mention them below.

1. At first step, designers have to understand the potential users. This understanding consists of two major part. One is directly understanding user's behavioural. The other one is studying the nature of the work expected to be accomplished.

2. In development process, target users should work with the prototype to achieve the real work. Also, their interactions and performance should be observed, recorded and analyzed.

3. When user testing problems are found, designers have to fix them. For fixing them, there should be an iterative design process that contains design, test and analyze and redesign [8].

In [9] the authors use a large number of software projects in order to examine UCD theories and analyze them on the projects. The result of their work gathered in 12 principles which capture factors such as user focus, active user involvement, simple design representation, prototyping, evaluate use in context, explicit and conscious design activities, process customization and a user-centric attitude by the team development [9]. Since proposed principles are abstract in nature, the authors also provide an activity list for aforementioned principles to specify the main techniques that help designers to apply principles in their real world project [9]. In addition to the mentioned principles, forming the idea is the starting point for UCD process. Because of that, researchers have created some design guidelines to organize the process and facilitate idea generation. Shneiderman eight golden rules [10] can be mentioned as one of the most influential guidelines for the design of human-machine interaction. Shneiderman [10] proposed the eight golden rules of interface design based on the concept of direct manipulation, which leverages physical symbols in user interactions [11]. We briefly discuss these rules below [12].

1. **Strive for consistency**, that emphasizes on different aspects of design like layouts, colors, terminologies and sequence of actions should be consistent through the design and prototype. If the system has some exceptions, those should be limited and comprehensible for the end-users.

2. **Seek universal usability**, which points out importance of knowing different users types (based on the attributes of users like age, disabilities, etc.) and make flexible designs.

3. **Offer informative feedback**, which mentions that the product present appropriate feedback based on user actions; for example the system should give modest feedback to frequent actions and special feedback to infrequent and major actions.

4. **Design dialogues to yield closure**, which identifies that sequence of actions between user and product should be formed as a dialogue that consists of a beginning, middle, and end.

5. **Prevent errors**, mentions that the designed system should have a troubleshooting process which means end-users can not make serious mistakes. The product should be designed in a way that if a user makes a mistake and the error was raised, an easy and meaningful process helps the user to solve the error.

6. **Permit easy reversal of actions**, asserts that a design should provide easy processes to reverse different actions that were made by the users.

7. **Keep users in control**, which indicates that in the process of working with the designed product, surprises and annoyances should be minimized. Also, the product should enable users to perform their desired tasks easily.

8. **Reduce short-term memory load**, which specifies that designers should create interfaces that avoid users memorize information from one step to the other [12].

In this project, we follow a UCD approach in the design of the information tool for creating HFTViz. In particular, we used an interview method to research users and understand their needs [12].

2.2 Financial Visualization

Previous works related to financial data visualization methods can be organized in two major categories (1) fraud detection in financial market (2) trading and investments. In this section we present related works in each aforementioned category. After that in the next section, we focus on the time series visualization section in order to find the limitations that HFTViz can solve.

2.2.1 Fraud Detection in Financial Market

Detecting specific events is very important task in different areas. As an example, detecting flash crashes in the stock market can be beneficial for algorithmic traders to gain lots of money. Another example could be in healthcare area. Detecting a rare disease could be vital for patients. Also, finding a fraud in the financial transaction guarantees the security of the financial network and users can consume different financial services safely.

Analyzing time series data allows the identification of insights such as frequency, trends, and changes. Moreover, finding outliers can be beneficial to identify rare situations in data. In this section we focus on related works in financial frauds detection (FFD). Each day, there are millions of transaction in financial network around the world. Most of the transactions are legitimate but there is a small number of criminal attempts in them which may create dangerous situation for customers or financial institutions. Because of that, each transaction should be assessed by financial institutes but since transactions formed a multidimensional time series data, it's a difficult task to detect FFDs [13].

Some researchers gather surveys in fraud detection. In [14], Bolton and Hand did a review about FFD categories. They categorized the available tools for statistical fraud detection based on their technologies in four areas: credit card fraud, money laundering, telecommunication fraud, and computer intrusion. In [15], Kou et al. gathered different techniques for identifying the same types of fraud as described in [14] as a survey. We can classify the different approaches into two categories: misuse and anomaly detection [13]. Both categories show outlier detection, neural networks, expert systems, model-based reasoning, data mining, state transition analysis, and information visualization as techniques to identify FFDs. We use aforementioned surveys in order to understand diverse fraud domains and how they are normally tackled. We identified FinanceVis proposed in [16] as a survey of visual techniques for financial data. This web-based tool includes papers related to financial data visualization. Also, this tool enables us to analyze similarity of fraud transaction with the normal one. Ko et al. [17] presented a survey of ways to exploring financial data based on a motivation from lack of data. In this work, they gathered financial experts and conducted an interview in order to find their concerns about FFDs in financial data. In [18], Kirkland et al. presented one of the first works using visual techniques to support FFDs. In their work they connected AI topics like pattern recognition and data mining to information visualization in order to support analysis, fraud detection, and knowledge discovery [13].

In [19], authors present WireVis which its main idea is enabling users to investigate through large amount of transaction data using multiple coordinated views. For helping fraude detection, they use similarities between accounts an a factor for fraud detection through time. One of the limitation of the WireVis in [19] is that the tool can not analyze single account to find its possible frauds. Instead, the tool has to leverage clustering in order to find suspicious accounts. In [13], authors use the similar approach with clustering. The difference of EVA with WireVis is that EVA provides more flexible analysis through accounts rather than focusing on analysis of keyword patterns in the transactions. Another approach that can be used for fraud detection is data flow that is presented by [20]. In this approach, for creating better understanding of events and making transaction decision, data are gathered for users to enable them draw analytical conclusions. By presenting EventFlow [21], authors aimed to design a tool for facilitating analysis, queries on data and data transaction of time series datasets. This work aims tracking entities and event related to them by creating aggregated data representation. Huang et al. [22] presented a general approach for event monitoring with a visual analysis framework for stock market security. One of the problems of FDDs can be number of false alarms that are produced by the systems like AI algorithms. For reducing number of false alarms, Huang et al. [22] present their visual approach with combining 3D tree map for market performance analysis and a node-link diagram for network analysis. In [23], authors gathered current limitations and needs in FFD. They proposed a framework in order to predict time and type of actions for the investigators to apply using visual analysis techniques. They investigate through different visual techniques and gathered ones that support different cognitive processes. Beside that, the authors proposed future challenges in FFD area and made a discussion around efficiency of data visualization for FFD. Also in the health domain there are use cases for anomaly detection. In [24], Rind et al. presented a study concentrating on data visualization system that are used for investigating health records. Moreover, in [25], Wagner et al. proposed a systematic overview and categorization for virus defected systems using visual analysis. Both health and software areas have some similar studies that are related to FFD. The reason is of similarity is type of data that they use. Both of them are using multivariate and temporal data [13].

2.2.2 Trading and Investment

By introducing Capital Asset Pricing Model (CAPM) in the 1960s, quantitative investment began [26]. Since then different theories have been presented to interpret stock returns relationship with different macro and company-specific risk factors using mathematical modelling. Nowadays, because of the technology investors can create different strategies based on risk factors and test multiple portfolios to see the results. According to the IPE report in 2019 [27], the value of assets under management in top 400 ranked by worldwide companies, reached 66.4 trillion. One of the popular methods for quantitative investment is using risk factors which leverages mathematical techniques to build portfolios. Usually, the steps for creating a portfolio start from searching for potential companies to invest using different methods like searching online through companies balance sheet, making on-site visits and interacting with companies leader. However in quantitative investment gathering data plays a key role in decision making. Investors do that by gathering historical data from companies that are publicly available like book-to-price ratio. After gathering data, they apply their quantitative model in order to find significant risk factors that can affect companies returns more that others. In quantitative investing, risk factor models help investors to understand the relationship between a stock and the whole market [28], by running multivariate regression; including different risk factors like price-to-earnings and price-to-book. After the multivariate regression, risk factors returns should be independent from each other in order to construct diversified portfolio for investors. From different models in quantitative portfolio investing, Barra Risk Model is one of the most widely-adopted model. The model was created by Bar Rosenberg, the founder of Barra Inc., with collaborating with Grinold and Kahn [29]. Barra model has been updated many times by MSCI[1], who acquired Barra Inc, in order to adapt with different stocks market. The reason for its popularity in finance world is because of the combination of traditional factors with the new ones in creating the model [30]. As an example, in the Barra model for U.S market, we have two categories of factors that each of them has 13 risk factors related to style of the market and industries that are in the market [31]. This factor number is more than other famous models like CAPM

[1] MSCI. https://www.msci.com.

and Fama-French [32]. Because of the large number of factors, Barra model can be better predictor for stocks return [33].

Based on the progress of computational power and novel visualization systems, Portfolio data visualization has been a popular topic in the past two decades. Earlier research work focused on portfolio returns visualization since it has a key role for measuring the performance of a portfolio. In [34], Matthias et al. presented a visualization based on line chart. They divided the background of the line chart and colored it in order to encode the return values. To compare the return and volatility of stocks in real time, Ziegler et al. [35] used a visualization based on heatmap for using the space of the screen efficiently. Both [34] and [35] tried to take advantage of background for encoding extra information. Beside those aforementioned works there are many previous works were based on heatmap [36], [37], [38]. Those works leverage heatmap power for space utilization to encode more information regarding the returns in different time periods. One way to observe the markets returns is using clustering techniques. As an example, Lei et al. [39] clustered all stocks and grouped them by their returns. After grouping, they present groups in co-centered, sequential rings. In [40], Xiong et al. visualized returns that play a key role for constructing mutual funds[2] portfolio. Beside mentioned works on portfolio returns visualization, we can add some works that focused on helping non-expert users in portfolio selection. For example, systems like FinVis [41] or PortfolioCompare [42] were created to help general users to pick portfolio that has better performance based on expected returns. For displaying stock holdings like different exchanges together, Jungmeister et al. [43] and Csallner et al. [44] both used treemaps to accomplish this task and made improvements in user interactions. [45], [46], [47], Dwyer et al. presented a visualization that they called "2.5 Dimensional". This visualization leveraged clustering in order to group funds by their stock holdings [33].

2.3 Time Series Visualization

When we gather variables over time, we create a time series. Formally, we can define time series (X) as: $X = x_1, x_2, ..., x_n$, where n is the number of observed time steps and $x_t = x^t_1, x^t_2 ..., x^t_m \in R^m$ is the observed values for m variables at time t (from 1 to n). If we have one variable in our analysis, we can call our time series a univariate ($m = 1$). Otherwise, if we have multiple variable in our data, we can call it multivariate time series($m > 1$) [48].

After organizations felt the need of having a monitoring system in order to detect abnormal-

[2]A mutual fund is an open-end professionally managed investment fund that pools money from many investors to purchase securities. Mutual funds are "the largest proportion of equity of U.S. corporations." Mutual fund investors may be retail or institutional in nature.

ities over time, multivariate time series became popular. Because of that, there are many research done in this area. Some of the well-known applications are cybersecurity [49], fraud detection [50] and environmental issues [51]. Large internet companies or high-performance computing centers can be two major sources for generating large multivariate time series since they have lots of users that work with their services. In these companies, it's vital to monitor server's metrics such as CPU utilization, memory utilization, and power usage over time for detecting abnormal behaviours. After unusual event identification, companies can make decisions in order to improve security and performance of the services. The cluster trace datasets from Google [52] and Alibaba [53] can be examples of having such monitoring systems [48].

In our work, HFT dataset is multivariate time series. It contains different variables such as ask/bid/cancel/modify price and volume of trade. Basically, line graphs and heatmap are common approach to visualize a time series data [51]. On one hand, line graphs can present movements of a variable well but it creates visual clutter if we have a large number of variables. On the other hand, heatmap technique solves visual clutter problem and utilize the space for showing more graphs. Also it uses different color encoding such as color saturation and shades in order to visualize change in time series value. The limitation of heatmap can be requiring large displays to show all the details of multivariate time series at once. Beside heatmap, we can use animations for visualizing changes through time and avoid using large screens to show every detail. The limitation of animation is that it forces the users to memorize the details in a specific timestamp in order to use it in a future analysis or replay the animation to see that event again. Those aforementioned limitations reduce the performance for exploration and analysis of data. [54] [48].

Beside current approach to solve visual clutter problem, we can use visual elements in small multiples [55]. Those multiples show snapshots of data for a few time sequences. Two often solutions for multivariate data are radar chart [56] and parallel coordinates [57]. Radar chart projects multidimensional data into two-dimensional circular space. In this chart, we draw equiangular axes in order to show variables. After that, the distance from center on axes reflects the value of the variable [58]. After applying values of each item on the axes, we use a polygon to connect the points for that item. So we have one polygon for each item on the graph that shows its variables values. Using this polygon, each item has a shape in the visualization. Like Radar chart, the parallel coordinates [59] is another approach for presenting multivariate data. In this method, we project N-dimensional data into 2-dimensional sets. Each data item represented by a line or hyperplanes. This representation allows us to visualize multidimensional relation [48].

Another popular category of time series is financial data. Financial data categories are not all about stock market. They can be insurance or banking data. In [60], Lei et al. presented an approach for visualizing financial time series data that they called "visual signatures". Moreover, in [61], Sorenson et al. presented a time series visualization which helps users to analyze price fluctuation. Their visual design combines glyph-based events with line chart in order to fulfill their task. Besides that, there are some surveys that gathered different visualization techniques for time series data like FinanceVis.net [16] and towards visual analysis [17] [33].

Overall, different visualization approaches were presented by researchers. However, there are some limitations to aforementioned works. One of them is that they are working with financial time series data that has low frequency from years to days. But in our work, we have large multivariate time series data that has frequency from days to microseconds. In the next chapter, we present our tool, HFTViz, for HFT data visualization.

CHAPTER 3 ARTICLE 1 : HFTVIZ: VISUALIZATION FOR THE EXPLORATION OF HIGH FREQUENCY TRADING DATA

(This chapter is based on the paper we have submitted to Information Visualization Journal. The authors of the paper are Javad YaAli, Thomas Hurtut and Vincent Gregoire. This paper is submitted on October 13th 2021.)

High Frequency Trading (HFT), mainly based on high speed infrastructure, is a significant element of the trading industry. However, trading machines generate enormous quantities of trading messages that are difficult to explore for financial researchers and traders. Visualization tools of financial data usually focus on portfolio management and the analysis of the relationships between risk and return. Beside risk-return relationship, there are other aspects that attract financial researchers like liquidity and moments of flash crashes in the market. HFT researchers can extract these aspects from HFT data since it shows every detail of the market movement. In this paper, we present HFTViz, a visualization tool designed to help financial researchers explore the HFT dataset provided by NASDAQ exchange. HFTViz provides a comprehensive dashboard aimed at facilitate HFT data exploration. HFTViz contains two sections. It first proposes an overview of the market on a specific date. After selecting desired stocks from overview visualization to investigate in detail, HFTViz also provides a detailed view of the trading messages, the trading volumes and the liquidity measures. In a case study gathering five domain experts, we illustrate the usefulness of HFTViz.

3.1 Introduction

With the development of high speed electronic chips, High Frequency Trading (HFT) has become one of the most profitable ways for trading. Financial markets have been transformed by faster speeds in recent years. For example, the BYX exchange reduced its order processing time sevenfold, from 445 μs in 2009 to 64 μs in 2018. Likewise, the round-trip communication time between Nasdaq and the Chicago Mercantile Exchange (CME) has nearly halved, from over 14.5 ms in 2010 to 7.9 ms today [4]. HFT data can be used by traders, financial market regulators and academic researchers. Traders want to detect different events like abrupt changes in stock prices to see their algorithm's performance. The result of this exploration can then be used to influence the improvement of their strategies and algorithms. Regulators are another possible user of HFT data. This type of user wants HFT data to investigate the market movements and detect financial frauds. Likewise, academic researchers tackle questions about market micro structure and algorithmic trading effects on the market, using

HFT data. In this design study paper, we introduce a visual analytic system that helps users from these different categories to investigate events in HFT data.

HFTViz provides a visual interface to HFT data. It enables users to investigate several levels of time aggregation, i.e. from hours to milliseconds. Altogether, contributions of HFTViz are:

- An HFT data visualization design enabling the exploration and understanding of financial information like price changes, liquidity[1] measures and number of quoting messages in different timescales from milliseconds to hours;
- HFTViz, a prototype for HFT data exploration;
- A case study of HFTViz gathering five experts in financial data analysis.

3.2 Related Works

Previous work related to high frequency trading data visualization can be organized into two categories: financial data visualization and time series visualization. In the financial data visualization, we concentrate on approaches related to financial markets. After that, we present the time series visualization section to discuss works leveraging time series visualization techniques in different domains.

3.2.1 Financial Data Visualization

Previous works related to financial data visualization methods can be organized in two groups: high frequency data and low frequency data. Low frequency data has a range from years to days. In the contrary, high frequency data has a time range from days to milliseconds.

Previous works related to financial data visualization methods can be organized in two groups: high frequency data and low frequency data. Low frequency data has a range from years to days. In the contrary, high frequency data has a time range from days to milliseconds.

One of the significant aspects of HFT data is its dimension. As an example, there are around 3,000 companies in NASDAQ exchange, with the following temporal data: price, volume of trade, trade type (buy and sell) and message type (cancel, add and modify). Hence, HFT datasets are multivariate high dimensional. For handling high dimensionality, previous

[1] In business, economics or investment, market liquidity is a market's feature whereby an individual or firm can quickly purchase or sell an asset without causing a drastic change in the asset's price. Liquidity involves the trade-off between the price at which an asset can be sold, and how quickly it can be sold.

research uses dense fields of color to encode information. As an example, pixel oriented visualization technique in [62] represent large datasets by mapping individual data values to pixels. This technique is not useful in our case since we have 500 companies in S&P500[2] to show and millions of messages in milliseconds scale.

Several previous research target the visualization of multidimensional financial data. In [1], the authors use geographical location of companies headquarter office, sector, financial performance (e.g., cash flow, personnel expenses) to create a visualization. Their goal is to provide firm-to-firm transaction network for financial analysts. FinaVistory platform visualizes narratives that are generated by computer analysis to address all economical concerns related to the news [63]. The approach iConViz proposes a platform to help financial regulatory authorities and banks manage the risk associated with corporation loans [64]. For detecting frauds, Leite et al. propose EVA, a Visual Analytics approach for supporting fraud investigation, fine-tuning fraud detection algorithms, and thus, reducing false positive alarms [13]. Closer to trading activities, TradAO assists users in exploring the whole optimization process of a trading algorithm and evaluating its detailed performances [65]. Besides financial analysis, trading and fraud detection, there are emerging visualizations for cryptocurrencies. For example, BitVis enables users to analyze the behavior of their bitcoin accounts [66].

Besides finance, there are other areas that have high dimensional data. For example, in Genomics there are similar characteristics and challenges for handling dimension of genomes data. In [67], authors leverage summarizing approach to create an overview visualization for showing the broad view of Genomics dataset and its alignments trends. They conclude that large-scale visualizations should leverage simplicity and only provide details and excess dimensionality on demand.

3.2.2 Time Series Visualization

Time series visualization started with line charts that were used by Lambert and by Playfair in the 18th century [68]. Along the years, different visual designs were proposed by researchers in order to facilitate extracting insights from time series (see [69, 70] for an overview of time oriented visualizations). We can categorize the purpose of these visual techniques, for example some focused on seasonal effects in time series data [71,72], others aggregate different time series through clustering in order to place similar data together [73]. Another category concentrates on proposing a way to explore and compare a set of time series [74, 75]. One aspect of time series that grabs researcher's attention is the scalability of their visualizations.

[2] Stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States.

One of the oldest visualization approaches for scalability is to connect each time point with line and visualize line charts in small multiples [68] or sparklines [76]. Recent approaches are focused to propose different representations for line chart itself. For example, In [77], authors propose two-tone pseudo coloring visualization in order to represent line charts in one dimension. Also, in [78] authors propose horizon graphs that have different vertical levels that are color encoded and split the vertical range of values in a line chart into a few vertical bands, that are then overlapped. Those aforementioned representations of line chart can save vertical space while it saves all the details of the line chart. In other works, scalability is addressed using color-based representations, referred to as heatmap or color fields. In heatmap, color saturation or brightness encode the range of values over time instead of using position. This approach is seen in many systems [77, 79–81] and answers well to scalability by stacking multiple such sequences of small height [82, 83]. In [84], for representing multiple time series, the aforementioned visual design split the space vertically and try to encode each time series per heatmap. Instead, the same space can be served to represent multiple visualizations [84]. Another representation of multiple line charts can be created by overlapping the layers of area charts with different colors (e.g. stacked [85] or braided [84]). One of the problem in scalability with the majority of those space sharing techniques is visual clutter because of the large number of time series to represent. [86].

Despite current progresses in financial data and time series visualization, there are still two major limitations. One is the frequency of the data that is used in aforementioned visualizations. The other is details of price quoting in the market. In HFT data, we have different aspect of quoting like Bid, Ask, Cancel messages that help us track market movement in details. In conclusion, despite the fact that there are many methods for visualizing financial data, they do not adapt well to HFT data characteristics. The goal of this paper is to introduce a visualization that helps a financial researcher investigate HFT datasets easily.

3.3 Data and User Questions

In this section we introduce the data we use in more details, and the questions our typical target user has when analyzing such data. The dataset for this visualization is **NASDAQ-ITCH** data[3]. The structure of each message in the dataset has different features like ID, timestamp, type of the message (Bid / Ask / Cancel / Modify), volume of shares and price. The ID is the unique field for each message to identify the message in the stream. Timestamp is the point of time during the trading date. Type of message identifies whether our message is to buy a stock (Ask), sell the stock (Bid), cancel the previous order or modify the previous

[3]NASDAQ-ITCH

order's volume or price. Volume and price are amount and fee of the shares that traders want to trade in NASDAQ respectively.

Analysis of HFT dataset is challenging because of the amount of messages that are sent from NASDAQ server. According to [87] a typical file containing a single trading day consists of something like 30-50 millions of messages (BX-exchange) up to 230 millions of messages (NASDAQ), thus speed makes a crucial difference. The data contains the messages that NASDAQ server sends to the subscribed trading machines.

Beside understanding HFT data, in order to characterize the design constraints, we have to identify the target user questions which are necessary for the user to reach their higher level goals. To that end, we gathered the questions by conducting five informal interviews with three experts in Financial Engineering who use the dataset as part of their research process. Firstly, we asked them to describe the usage of the dataset in the daily workflow of a financial expert. After that, we extract the user questions from their description and conduct another interview with them to fine-tune the details of each question. We did that process five times. In addition to interviews, we asked them to participate in our evaluation process and let us know their opinions.

We have identified five principle questions (Table 3.1) that financial researchers need to answer in order to achieve scientific intuition in the process of analyzing the data. We have formed them according to their priority for our interviewers. Questions Q1 and Q2 endeavors to find out the messages price and volume patterns in the data respectively. The price of the stock in time t is mean of bid and ask price on that time [88]. Questions Q3 and Q4 investigate the different aspects of the liquidity feature which is calculated from message type, price and volume. Based on US Securities and Exchange Commission[4], stock's liquidity generally refers to how rapidly shares of a stock can be bought or sold without substantially impacting the stock price. Stocks with low liquidity may be difficult to sell and may cause you to take a bigger loss if you cannot sell the shares when you want to. Finally, Q5 tries to discover the daily dynamic of the stock correlation with the US market.

3.4 Proposed Design

Our goal in the design of HFTViz is to create a tool for financial researchers that enables them to easily investigate the data. The challenges with HFT datasets are: (1) there is lots of data point to show and investigate (2) there are different methods and mathematical metrics to consider. For handling this, we divide the effects of the data into two groups.

[4]www.investor.gov

Table 3.1 Questions for analysis of NASDAQ dataset with the feature addressed by each. The features are Price, Volume, Liquidity and Correlation.

	Question	Feature			
		Price	Volume	Liquidity	Correlation
1	How does a single stock messages (bid / ask), price and volume pattern change around news ?	✓	✓	-	-
2	What is a daily market price dynamic?	✓	-	-	-
3	What is a stock liquidity changing pattern ?	-	-	✓	-
4	What liquidity measures best capture the dynamics of the limit order book?	-	-	✓	-
5	How do stock and market movements change in different time scales?	-	-	-	✓

One is large-scale effects and the other one is detail-scale effects. For handling large-scale effects, we propose an overview visualization to answer to our user needs **(Q2)**. In addition, for answering detail-scale user needs **(Q1, Q3, Q4, Q5)** we design a dashboard that covers different aspect of the detailed user needs.

3.4.1 Overview Visualization

Our goal in designing overview visualization is to answer the large-scale effect of our user questions. This section starts with an introduction to overview visualization. Then, following subsections describe details about different aspects of the overview visualization. Firstly, we have to consider that users should be able to view data patterns in different displays. As discussed in the time series visualization section, one of the potential designs can be color field design [67]. This design makes the patterns in the data to pop out more in the visualization [67]. Since the number of messages in each day is large, we need an aggregation method to summarize the data points and visualize it in the screen. Other aspects of design include a tool for arranging the companies for effective comparison and also interaction techniques to help exploration (see Fig. 3.1). In the following subsections, **Perception** and **Visual Search**, we will describe details of the overview visualization.

Perception

Visualizing 300 millions of messages can be complex. Since human visual system can be overwhelmed by many information, our design must rely on the concepts that show patterns easily to the users and support search for details [89].

One of the visual concepts that we can use in our design is preattentive visual processing. This aspect of the human visual perception allows user to rapidly find patterns in a visually cluttered environments. Leveraging this processing, we can implement some features in

Figure 3.1 HFTViz overview and details panels.

HFTViz that simplify visual search by making some patterns visually pop out. Since HFTViz knows the users needs, so it can use preattentive processing to highlight patterns. Also, in [90] authors mention that preattentive features can be processed in parallel so using this concept in our design, we can reduce visual search task's cognitive load. For applying the impact of preattentive processing, we use color scheme. To avoid false attention, we use semantically informed color choices. We chose different shades of green and red for positive and negative price movements respectively (Fig. 3.2). This color encoding enables us to leverage preattentive processing to find patterns in price movements.

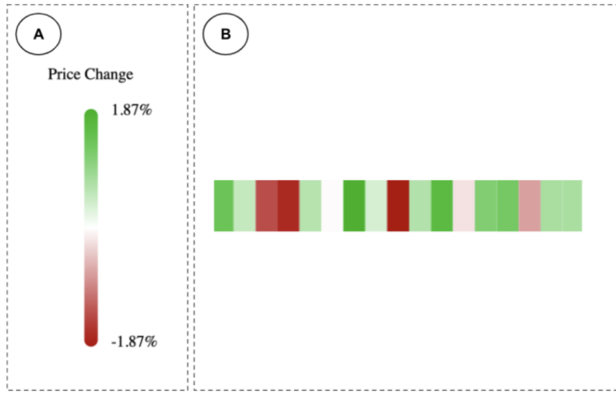

Figure 3.2 Price change visualization.

Visual Search

In case of not finding the patterns using preattentive processing, visual search can be helpful. In this process, users scan the scene with their attention to search for patterns. Visual search can be time-consuming for the users if they don't have perceptual aid [91]. Due to our visual system limitations we can not recognize everything at once even when the pattern is shown in the current scene [92]. Since we have at least 500 companies in the overview section, visual search plays a key role in the overview visualization.

In [93] authors mentioned that the direction of searching a display is the same as reading order. In our design we implement two tools for making flexible visual search. One is a search bar that aids user to search a company by name. The other one is sorting method. User can sort the companies with their market capitalization or trading volume. In addition, for helping users to have better visual search, we provide a padding space between each company block.

Aggregation

In large datasets, we usually have sequences that are larger that the horizontal screen size. Because of that even if we can put all of the data in the screen, it creates visual clutter and makes visualization hard to interpret. For handling this complexity, HFTViz implements horizontal aggregation blocks in the visualization. Price changes are grouped into blocks that can be shown in the screen (Fig.3.2B, Fig.3.3). Using blocks can be an approach for down sampling the data and creates more control on the information flow in the block [67]. The blocks work like bin size in histograms. The number of blocks are defined by user. The users can change the blocks. Since we visualize the price changes in the overview, to aggregate the blocks together, we need to sum them up and create bigger blocks.

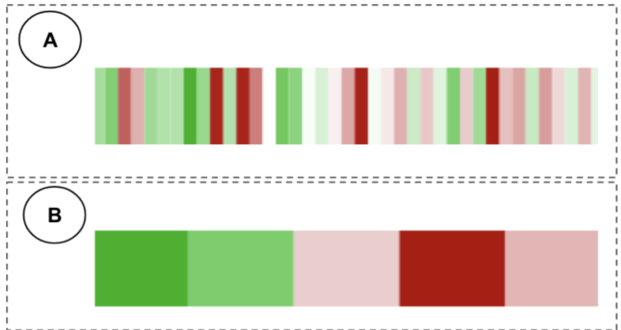

Figure 3.3 Percentage of price change movement for two intervals.

3.4.2 Detailed Visualization

The design of HFTViz follows the Shneidermans' mantra "Overview first, zoom and filter, then details on demand" [94]. In the overview visualization, users select the stocks based on their price movement. Besides selecting stocks using overview visualization panel, we add SPY ETF[5] (as a S&P500 market indicator) by default to the selected stocks in order to help users compare stocks details with a market indicator. After selecting the stocks and adding market indicator, users can load the detailed visualization by clicking on View details button. Detailed visualization panel consists of two main components: the limit order book (LOB) view and the liquidity measures view. Since HFTViz provides various interactions, to show various data granularity of data, on user demand, it is crucial to arrange the information properly, so that users could perform analysis efficiently. According to the theory proposed in [95], it is more efficient for users to compare views side-by-side, rather than commit visible items to memory. For this reason, we attempt to use the screen space fully in our system design, so that we could display different data granularity (LOB movements and its liquidity measures) side-by-side (Fig. 3.4). This avoids that users are forced to rely on a mental map to perform comparisons [33].

LOB view:

We provide LOB view for the users to discover different aspects of LOB efficiently. The market changes every microsecond, which means different time periods may have different

[5] An exchange-traded fund is a type of investment fund and exchange-traded product, i.e. they are traded on stock exchanges. ETFs are similar in many ways to mutual funds, except that ETFs are bought and sold throughout the day on stock exchanges while mutual funds are bought and sold based on their price at day's end.

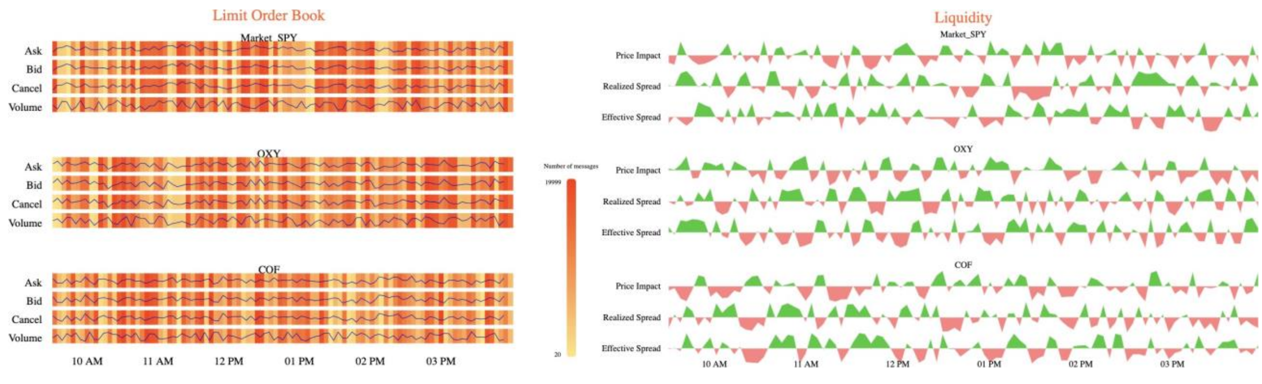

Figure 3.4 Detailed visualization panel.

patterns and insights, such as different bid price, ask price, cancel price and volume of the traded shares **(Q1)**. We also provide quoting message number as a parameter to show the focus of the traders in the market.

There are two main parts in the LOB view, price movements and the message number. Price movements define changing price through time for bid, ask and cancel messages. For showing volume of the trade through time, we use the concept of price movements. In addition, message number shows the number of quoting messages (sell, buy or cancel) in a specific time period. For each selected stock, we divide its section into bid, ask and cancel. Furthermore, we add volume section to visualize the traded volume of the stock.

The design of the price movements is built on a line chart to capture the changes in the price. We use the x-axis to encode the timeline and the y-axis to encode the exact amount. The amount for bid, ask and cancel section is their price but for the volume section, it represents the traded stocks in that period.

For the message number, we use the background of each section (bid, ask, cancel and volume) to represent the number of quoting messages. Quoting messages is calculated as the sum of bid, ask and cancel messages and it shows concentration of stock trading. For designing the message number visualization, we use heatmap for each section (i.e. bid, ask, cancel and volume). In each section, the x-axis represents timeline and the color of each block of the heatmap encodes the number of messages. Because of high number of messages in daily HFT data, we use aggregation technique as previously explained in Aggregation. Moreover, we enable zooming interaction for users to investigate the details on demand (Fig. 3.5, Fig. 3.6). In this approach, we divide the detailed heatmap with all messages into a discrete set of images, called 1-D tiles, details can be seen in these tiles by using the zoom function

(Q5) [96]. The Zoom function is also applicable to all selected stocks and liquidity measures simultaneously in order to help users to investigate different timescales efficiently. Another interaction that helps users to investigate the data by using HFTViz is hovering. When a user hovers on a block of heatmap that corresponds to a timestamp, HFTViz shows the number of messages, bid/ask/cancel price and trading volume in all selected stocks at that timestamp. In addition to stocks, hovering also enables users to see the value of the liquidity measures in hovered timestamp.

We separate the LOB view into two main sections because we want to observe the pattern in number of messages and price movements together. An alternative design could be applying multiple lines for each of the price and message number instead of using heatmap for the message number. Nonetheless, there is a weakness to this alternative design. Having multiple lines together creates clutter in the visualization since we have four sections for each stock and multiple stocks to choose. Because of that, we choose heatmap for encoding the information regarding the message number.

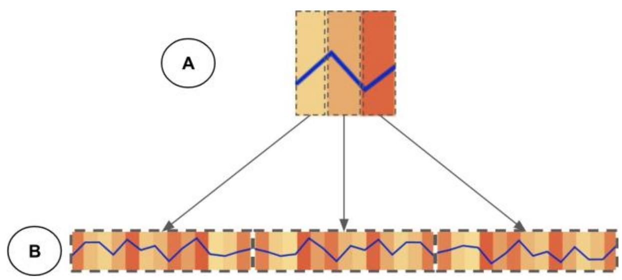

Figure 3.5 LOB view of HFTViz.

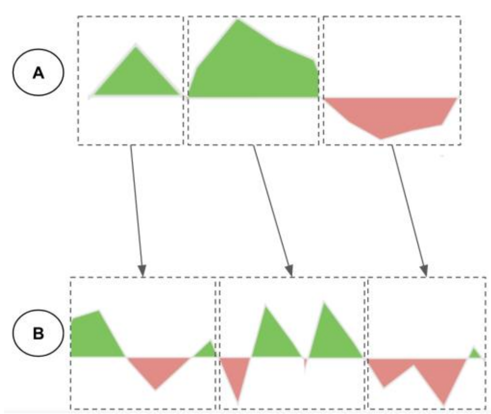

Figure 3.6 1-D map view format with zooming level.

Liquidity Measures view:

The liquidity measures view shows different liquidity metrics for the selected stocks in addition to SPY ETF. Those metrics help users to observe the trade effects on the stock prices. Furthermore, the liquidity measures visualization enables users to investigate different patterns in the market to see which measure best captures the dynamics of the LOB **(Q3, Q4)**.

For visualizing the measures, we select top three liquidity metrics in importance to our potential users. Based on our set of user questions, we need to visualize the liquidity movements while we show its trend. As we explained earlier in Overview Visualization, color fields allow patterns and trends to pop out. As a result, we choose area chart for visualizing the liquidity measures. This design choice enables users to investigate quantity movements while they are observing the trend of positive or negative liquidity. We encode positive liquidity in green color and negative liquidity in red color in order to help users discover the trends.

An alternative design could be the line chart with green and red colors encoding positive and negative values respectively. However, there is a drawback for this alternative choice. The line thickness in the visualization when we choose different stocks could be very small. As a result, users could not identify trends efficiently. Our design choice uses colors to encode the space below the line chart efficiently for popping out trending in liquidity measures.

3.4.3 Implementation

We implement a prototype of HFTViz using HTML, CSS and JavaScript. This prototype uses D3.js for implementing all visual designs and the React.js framework for managing interactions. A demo of this visualization is publicly available at: HFTViz demo page. Please note that the prototype is optimized for 1080p screens and the Google Chrome browser. Because of the NASDAQ-ITCH data license, HFTViz demo is using synthetic data and it performs all functionalities that we explained in previous sections.

3.5 Evaluation

We conducted a case study to evaluate the proposed HFTViz design. We invited five experts for a one-hour online interview using a video-conference platform. The experts were enrolled in the evaluation because of their interest in HFTViz and their compatibility with the scope of the project. In addition, we followed an official written protocol describing the sequence of questions and features to be presented. Furthermore, we recorded the interviews with each expert and their interactions on-screen during the entire demo session for post-evaluation transcription. Our goal was to understand whether the questions that drove HFTViz development were answered; moreover, we wanted to assess the validity of our assumptions and design choices.

Each interview started with a 10 minute presentation of HFTViz. During that presentation, we onboarded the user and showed all possible interactions. After that, we gave them 15 minutes in order for exploration, asking questions related to working with the tool and identifying a potential goal that they would like to investigate. Then, we asked them to identify the goal that they chose and share their specified screen for opening the visualization. Subsequently, they start working with the tool. This step took 25 minutes for the users. During that process, we record all the movements and selections on the screen. After investigation step, we started a discussion with the user for gathering their opinion regarding the tool. We asked them detailed questions about different sections in the visualization. Also, we organized some open questions to collect ideas for improving HFTViz in the future.

After gathering all the interviews recordings, we organize them into two sections with respect to the goals that users identified. The first section is "Investigating news effects on stock market" and the second section is "Exploring the effect of other markets on stock market".

3.5.1 Investigating News Effects on Stock Market

In this case study, we have three users that are working in financial academia and industry. They mainly had keen interest in observing the effect of different news in the market. The first user, U_1 wanted to observe the effect of Apple vs. Epic Game's lawsuit. The rest of the users, U_2 and U_3, needed to investigate the effect of Federal reserve's announcement on the interest rates and Merk's earnings announcement on its own stock respectively. All the users were quite familiar with the HFT dataset and concepts but they have never seen the dataset visualized. Also, they didn't have a previous experience using HFTViz. In addition, two of them participated in the process of identifying the user question but they didn't involve in the visual design process.

At first, the users started working with the search bar in order to find their desired stocks. When they picked a stock, they see its position in the "Market Cap" sorted overview panel in HFTViz. After selecting their first stock using search bar, they looked at the overview panel to select related stocks in "Market Cap" or "Volume". They picked related stocks by clicking on their sections in the panel Fig. 3.7.

The users then explored the details of the selected stocks by clicking on "View details" button.

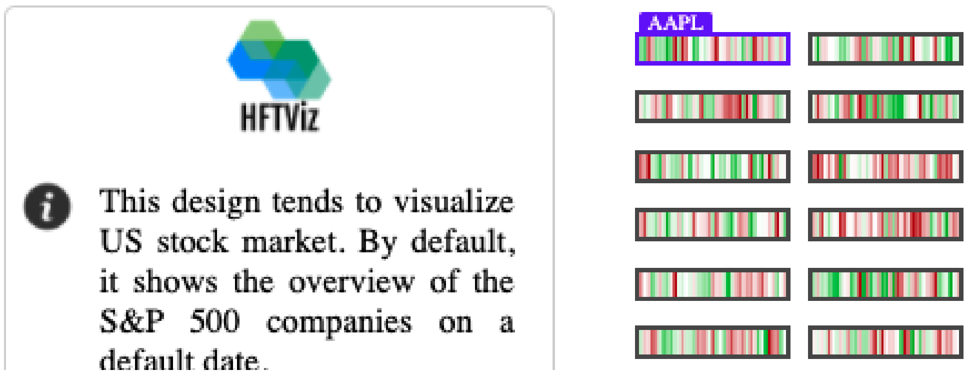

Figure 3.7 Choosing AAPL by market cap criteria.

After the details panel was loaded, the users start to interact with LOB section of the first selected stock at the top of the page and compare it with the market movements using hovering interaction. After hovering, the number of messages that shows the flow of messages in a particular time grabbed their attention to find the moments that their selected stocks had some attention from traders in stock market Fig. 3.8. This view helped U_1 to identify Apple's stock movement. Beside HFTViz, U_1 checked the Yahoo Finance in order to find the time of the news regarding the lawsuit. Also, U_1 combined the liquidity movements around news time with the number of messages in LOB to find the effect of the news on stocks price Fig. 3.9. To see the details of price movements, they zoom in interested area that shown in Fig. 3.8. Also, they mentioned that zooming feature helped a lot to see different details.

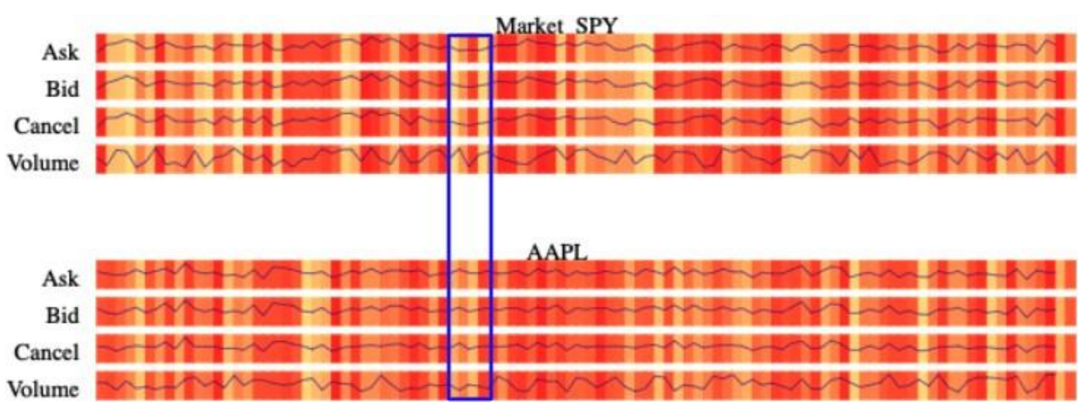

Figure 3.8 Number of messages in LOB view of details panel.

U_2 and U_3. U_2 also started with Yahoo Finance to find the Merck & Co. announcement date in 2020. After that, U_2 started using HFTViz in order to find the details of Merck's announcement on its stock price. U_2 selected Merck and UnitedHealth Group Incorporated

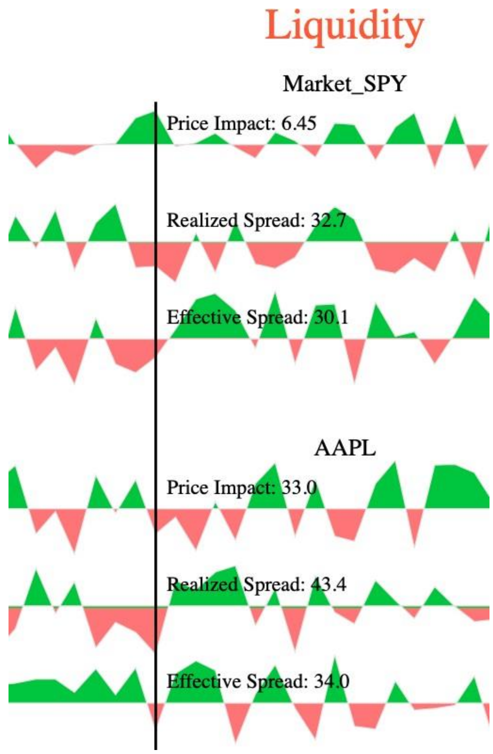

Figure 3.9 Liquidity changes around AAPL and Market SPY.

(UNH) stocks by searching in HFTViz search bar. Then, U_2 investigated the details by clicking on the "View details" button. After loading the stocks, U_2 found the high number of messages by hovering and zooming on LOB section. Similar to U_2, U_3 started investigating on the HFT with Yahoo Finance. After finding Federal Reserve's announcement date and time, U_3 used different sorting and selected Xcel Energy, Ameriprise Financial and DTE Energy by clicking on the overview panel. In addition, U_3 selected IBM, Coca cola by searching in the search bar. After finishing the selection part, U_3 loaded the details panel by clicking on the "View details" button. In the details panel, U_3 hovered on 3PM time to see the effect of the Fed's announcement on SPY ETF that was added at the top of panel. U_3 zoomed in that period and identified the movement of the market changed after that announcement.

At the end of the journey, they asserted to have better understood the effect of news on the stock market. They found related stocks with searching in the search bar or by sorting with different criteria. Also they mentioned that they could find the special periods in the market by using number of messages heatmap in the details sections. Furthermore, they commented on the liquidity section. They mentioned that the design of that section helped them to find most popular liquidity metrics easily and see its movements. Beside the benefits of the HFTViz, they also mentioned some improvements that could make HFTViz more suitable for financial users. They suggested the hovering label to become larger for being more clear to read. Another point that they noticed was changing the sequence of selected stocks at details panel in order to put different stocks near the Market SPY at the top for making comparison. In general, users are satisfied with HFTViz and they like to work with it in their workplace.

3.5.2 Exploring the Effect of Other Markets on Stock Market

In this section, we have two users that are portfolio analysts. They used different type of data in their works in order to create well performed portfolio. In addition, they have a keen interest in HFT world since they can investigate different phenomenon in the stock market in details. The first user of this case study, U_4 wanted to explore the effect of commodities market on stock market. Because of that, U_4 aimed to analyze the shortage of the oil on the stocks. In order to answer that question, U_4 started with oil demand dataset. After analyzing that, U_4 wanted to investigate related stocks in US market using HFTViz. For finding the related stocks using HFTViz, U_4 started with the search bar and select New Fortress Energy Inc., Xcel Energy Inc., CenterPoint Energy, Inc., Chesapeake Energy Corporation and NOV Inc. After that, U_4 went to the details panel to see the selected stocks in details. By looking at the number of messages and liquidity simultaneously, U_4 saw a pick in those companies stocks prices. With those finding, U_4 decided to create a statistical model to test the relationship between commodities and stock market.

Another user in this section, U_5, wanted to investigate the effect of Covid-19 on stock market. For this reason, U_5 was interested in looking at different industries. U_5 started with Feb. 20th of 2020 to find the effect of beginning of the pandemic. After that U_5 sorted the stocks by market capitalization and picked Apple, Google and Facebook. U_5 then went to the details and started looking at the liquidity measures first. After investigating the liquidity to find the start of big sell of at that date, U_5 changed the date into one year later to see the recovery of those stocks. U_5 finally noticed that the recovery of the giant tech companies is much faster than the rest of the S&P500 companies.

Furthermore, the users pointed out some limitations of HFTViz. In the current version of HFTViz, we focused on displaying all information on S&P500 companies and selected stocks; however, also more filters of companies (i.e. filtering by companies sector) would be of interest to the users. The overall feedback for HFTViz was very encouraging and indicates that our system generally answers the questions presented in "Data and User Questions" section.

3.6 Discussion

Two case studies show the effectiveness and usability of our system for the exploration of HFT data. Overall, our collaborating users are satisfied with its abilities, especially with LOB visualization. The large overview presents a whole picture of the S&P500 world based on percentage of price changes. Supplementary information is provided by the details panel. In details panel, HFT data is organized into two sections. The first section is Limit Order Book (LOB) view and the second one is Liquidity measures. LOB view helps users to investigate each stock in terms of its bid, ask, cancel messages and volume of trade. Also, Liquidity view helps users to find out time periods that a stock has low liquidity which means there is a special period in the market. When compared with standard financial data analysis software, our system has already processed data and provided a comprehensive analysis of the whole S&P500 market and individual stock with well-coordinated views. To the best of our knowledge, HFTViz is the first visualization system analyzing High Frequency Trading data, which closely follows domain tasks and requirements. Both domain experts and users with basic knowledge about HFT could step into this world via HFTViz [97].

Although HFTViz receives positive feedback and satisfaction from users, it has still some limitations. Firstly, our users mentioned dynamic font size for the visualization in order to set legends and labels by the user for convenient reading. The possible improvement for this limitation could be creating manual setting that user can change the font and theme of the visualization in order to read it easier. One other improvement that our users mentioned in the evaluation would be to add more filters to the overview section. They mentioned that filtering by industry sectors, special markets such as ETFs, S&P400, etc. could help them to find their stocks faster. Adding a specific section for filtering and provide different filters and sorting could improve further to design. In addition, our users noticed customized changes in the order of the stocks and liquidity measures can facilitate comparison between a stock and the market ETF (SPY). To solve this limitation, we could add a drag and drop interaction in order to change the stocks sequence and their liquidity metrics.

In addition, the current HFTViz system can be used to visualize other datasets. First, there are exchanges that can support high frequency trading. Also we can use our visualization in other market analyses and create a platform for trading business intelligence [97]. Second, the design in the overview panel could be extended by adding different filter and sort criteria.

3.7 Conclusion

In this paper we presented HFTViz, a tool for the visualization of High Frequency Trading data. An evaluation from our users showed that our design choices fulfilled our user needs. Based on positive feedback that we received, future research on this tool could be built based on this work. Also, for future works visual modifying of selected stocks can be added in order to improve user's performance **(Q1)**. Moreover, industry sector filters can be added to the overview visualization panel **(Q2)**. Also, there are different financial markets that have HFT mechanism for trading so we would like to expand our tool in them. Beside that, we plan to explore other methods for visualizing HFT data like animation and making maps from different markets. As some of our users recommended, we would like to add real-time feature viewing to our tool. This feature needs more investigation since HFT data are very heavy in volume and need proper storage.

CHAPTER 4 GENERAL DISCUSSION

In this chapter, we gather our important lessons from evaluation of HFTViz. After that, we will discuss HFTViz limitations and future works.

In this research, we start our evaluation by conducting interviews from each of the domain experts in order to learn about their workflow, i.e., the specific kind of events that they want to investigate using HFTViz. Our experts come from different backgrounds such as quantitative financial analysis and academic researchers in market micro-structure. Both user groups have the same user questions and needs. The difference between two groups is in their events that they would like to investigate. As HFT data analysts and researchers, their main goal is to create a model that explains impacts of some macroeconomics and microeconomics changes in the US financial market. Examples for specific tasks are impact of the announcements of a company on its stock and other stocks in the common industry section, effect of commodities shortage in stock market and policy impacts on stock market. Actually, our experts made their decision for participating in our evaluation based on the need for a solution to investigate HFT data in order to gain insights into the market and detect different factors that can affect financial market in different timescales. Our experts needs overlap with our identified user needs that we made our design choices based on (see "Data and User Questions" section). Overlapping our experts actual needs with HFTViz shows that our identified user questions fit the target users.

Our process was started by explaining the goals of HFTViz, followed by a live demo with one person in charge of answering users questions if they have any. We presented HFTViz features using that demo. After that, we started the free exploration part of the interview which helps users to get familiar with HFTViz. Then, we asked them to identify their tasks that they aimed to investigate. Subsequently, they pick a related date to the selected event and investigating overview visualization. Using overview visualization, they selected some stocks that are related to their task by searching in the search bar or clicking on its section in overview panel. After finalizing the stock selection in overview, our experts use the detailed visualization panel to investigate the details of each stock in order to fulfill their identified event.

After finishing the investigation section, we asked the experts to give us their feedback. They mentioned that they suffered from investigating HFT data without a proper visual tool since it is a large multivariate time series. However, after using HFTViz, they praised the simplicity of the tool for investigating different details of HFT data. They also mentioned that having an overview panel which shows all S&P500 together is a great idea to observe the entire market at once. The experts also appreciated the option to sort companies by market capitalization and trading volume to observe common companies evolution over time **(Q2)**. The detailed visualization panel fulfilled its design goal. The idea of visualizing number of messages by heatmap was noticed as a very interesting idea **(Q1)**. They mentioned that it helped them to identify density of highly trading moments in the market with price movements simultaneously. Moreover, they mentioned the hovering and zooming feature helpful in focusing on the moments belonging to a specific event **(Q5)**. The experts also confirmed the effectiveness of liquidity's measures besides the limit order book visualization **(Q3, Q4)**.

Beside all of the points that our experts mentioned, there are some limitations for HFTViz that we will now discuss. In the current version of HFTViz, displaying all information on S&P500 companies and selected stocks is our concentration. However, more filters of companies (i.e. filtering by companies sector) could be interesting for the experts. In addition, one of our experts mentioned that having sort by correlation would be beneficial for stock selection in overview panel. Another expert mentioned that creating a way to identify important dates in the market can be useful for date selection. Also, adding "Next" and "Previous" buttons to change dates is another point that was mentioned. Besides that, adding visual modification the order of the selected stocks in the detailed visualization panel is another limitation that our experts pointed out. Having vertical zooming is another point that our users mentioned during the open discussion after the tools evaluation. It can help users to see the movements of price in finer scales in both horizontal (by current zooming interaction) and vertical axes. Moreover, we receive a feedback that emphasizes on customization of the details panel in order to control the amount of information that users want to observe. Our expert mentioned toggling on/off the liquidity metrics can be one way to control the amount of visualization and facilitate comparison between same metrics. Our user also mentioned that currently it is hard to compare same liquidity metrics even if there is a synchronous hovering ability because other metrics exist between them that create visual clutter.

Based on our users answers to each question (Fig. 4.1), our users have dissatisfaction from sorting and readability of detailed visualization features. As we mentioned in previous section, they noticed these two as limitations for HFTViz. Our users wanted more criteria to filter and sort. Also, they would like to adjust the font size with their screen in order to have flexible readability when they are interacting with HFTViz. HFTViz visual design and features can achieve 80% of its user's satisfaction in our evaluation (Fig. 4.2).

In addition, we can apply HFTViz to other markets. Firstly, because most of the exchanges

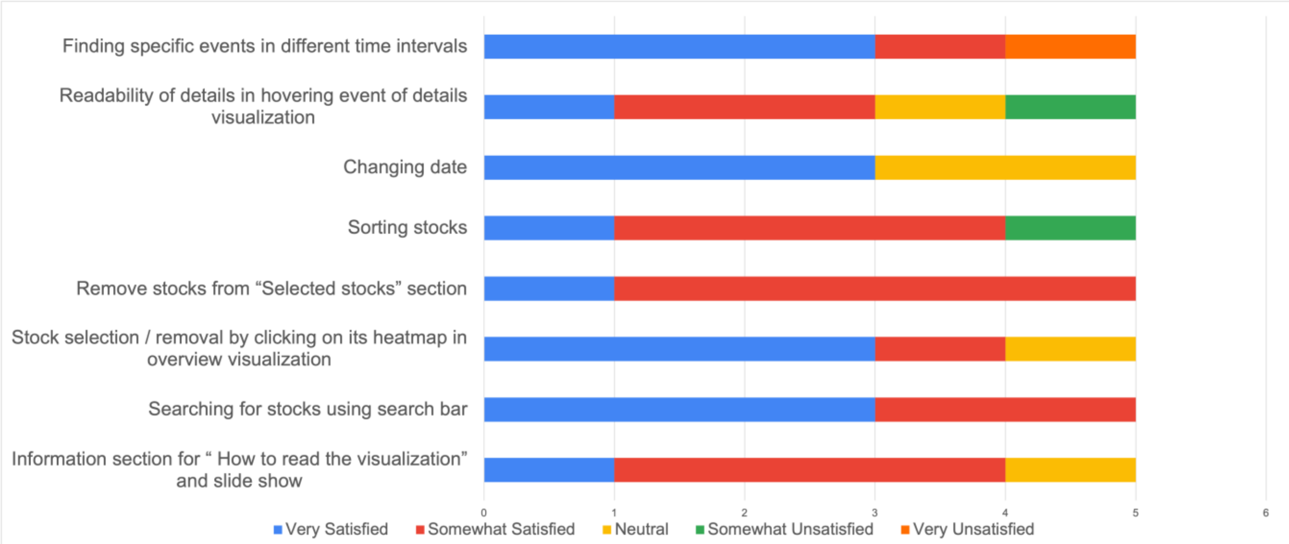

Figure 4.1 Distribution of answers to detailed questions.

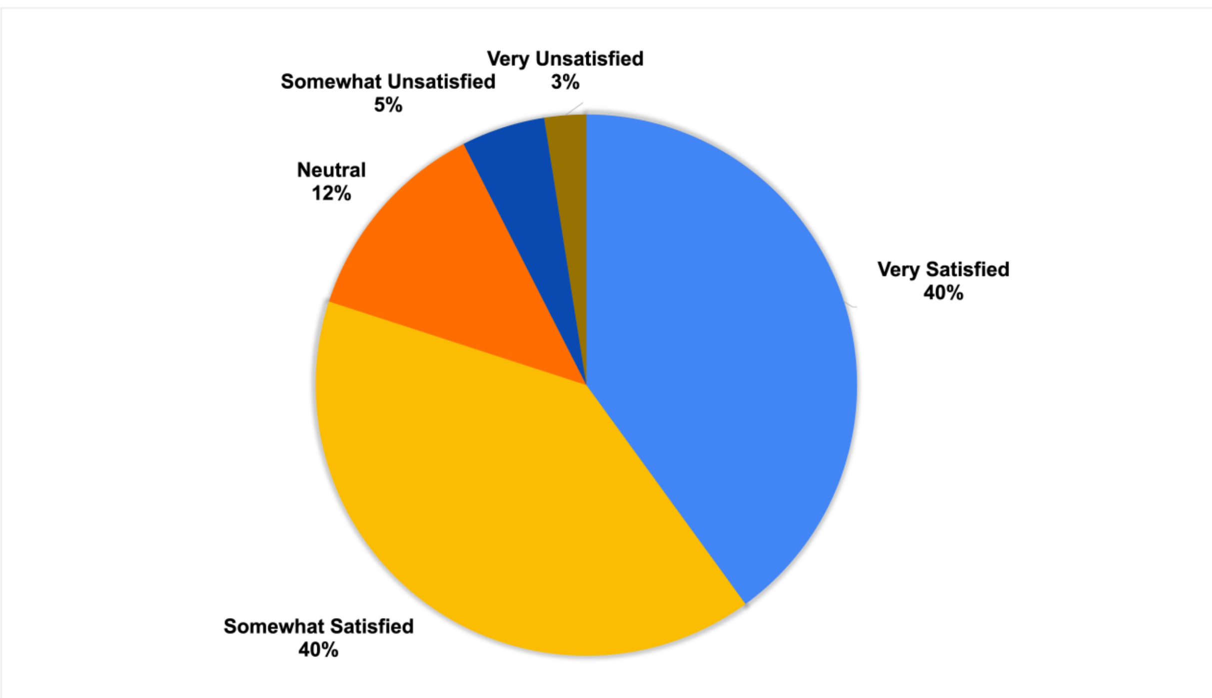

Figure 4.2 Total answer’s distribution.

support high frequency trading as a feature to their market in order to facilitate presence of big trading companies. Secondly, since our data is multivariate time series dataset, we can use this visual approach to other multivariate time series datasets. We also can apply our approach into the transaction based datasets that contain huge amount of transaction in different time scales. Moreover, we can add features for doing more quantitative analysis on detailed panels. For example drawing different indicators and metrics and overlapping them with price movement to see their correlation. Also, the design in the overview panel could be extended. We can add multiple sort criteria which encode alternative attributes of an asset [97].

CHAPTER 5 CONCLUSION

In this chapter, we summarize our work and mention HFTViz limitations briefly. After that, we wrap up the chapter by listing future research works that can be done based on current foundation.

5.1 Summary of Works

In this work, HFTViz was represented as a tool that could be used in order to investigate high frequency trading data. We started our work by conducting several interviews from two practitioners in financial industry. The goal of those interviews was to understand their needs and challenges that they currently face using HFT datasets. After understanding their needs via several interviews, we gathered all user needs and divided them into five user questions in order to answer them with our design choices. Then we started our design process based on user centered design guidelines and create different prototypes for HFTViz. After that, we filtered prototypes using review sessions that were conducted by research team. Those review sessions helped us to find the better perspective from user needs. We also added an idea to the prototype for representing message numbers leveraging heatmap in order to show density of trades in the market. After designing final version of the prototype, we conducted two case studies in order to get feedback from our potential users. We received positive feedback from them. Also, they acknowledged our idea for showing density of the trading in HFTViz. In the next section, we will briefly mention HFTViz limitations from user interviews.

5.2 Limitations

In this section, we briefly describe HFTViz limitations based on the user interviews that we organized for evaluating our design choices. In overview panel, most of our users mentioned customization in font as one of the potential improvements that we can make for next version of the HFTViz. Also, they mentioned that we can have a metric in order to mark important dates for the market in the calendar. They also commented on the filtering criteria. They mentioned that they would like to filter overview panel's stocks that are S&P500 by default. They told us filter by industry, correlation with other stocks and different groups like S&P400 are most needed filters for a financial analyst. Also, they mentioned that sorting options can be expanded. Sorting with correlation with other stocks or market itself was mentioned as a potential sorting option to add in next versions.

In the detailed panel, our users mentioned some customization comments such as adding time filtering tab in the menu side. They would like to use exact timestamp in order to investigate it in details. Regarding their comments, adding that feature helps them find the finer timescales immediately instead of investigating through multiple hours in a selected date. Beside time tab, our users like to filter the liquidity metrics in order to make a comparison between same liquidity metrics. One of our interviewees mentioned that when a user wants to compare price impact of all selected stocks, there are extra liquidity metrics present that create visual clutter. Another important limitation that one of our users mentioned is being able to change the order of selected stocks without canceling and re-selecting them. Our user told us that it could be a good idea to change the order of the selected stocks by dragging and dropping LOB section of the stocks. All those aforementioned limitations are precious for us because they lead us to create a better visualization to investigate HFT data. In the next section, we will summarize our future steps and developments that are possible based on current work.

5.3 Takeaways

In this research there are two main takeaways that can help other researchers to create more effective visual tools. First takeaway is importance of user questions. In our research, user questions are the best guide for testing our design choices and filter prototypes with them. Also, by interviewing the users, we found out that the user needs can cover wide variety of domains that we need to learn and cover them in our tools like financial liquidity in HFTViz. Second takeaway is the method of visualization that we use. In our research, we use number of messages in heatmap to show market concentration. Also, we present different price aspects like Bid, Ask and Cancel in the LOB section of the visualization. In addition, we include different liquidity metrics that present more information from data. HFTViz design choices can be applied in different markets that have high frequency infrastructure like Cryptocurrencies, commodities (oil, soybean, etc.) and Foreign Exchanges.

5.4 Future Research

Future research can be shaped around limitations of our current work. Firstly, we can add more features in order to cover limitations. Creating a menu tab to customize font size and color theme enables users to make the tool more readable based on their screen size. Also, creating a metric for showing importance of a date in the calendar helps users to observe important days in the the market using tool’s calendar. For covering filters criteria we can add a section in the menu in order to present different options for filtering the overview panel's stocks. Adding more options in sort section can solve mentioned limitations.

Beside covering the limitations, we can apply our tool for visualizing different markets that have high frequency trading option for their traders. Also, there are similarities between stocks and cryptocurrencies. Hence, we can visualize cryptocurrency trading data with our current approach and design choices. Another improvement that we can make for future versions is real-time update. This improvement needs more investigation because of the nature of the HFT data. HFT data needs large data storage to host and serve it in real-time format.

REFERENCES

[1] A. Arleo, C. Tsigkanos, C. Jia, R. A. Leite, I. Murturi, M. Klaffenböck, S. Dustdar, M. Wimmer, S. Miksch, and J. Sorger, "Sabrina: Modeling and visualization of financial data over time with incremental domain knowledge," in *2019 IEEE Visualization Conference (VIS)*. IEEE, 2019, pp. 51–55.

[2] Paul. (2021) User-centered design: the art of making things easy to use. [Online]. Available: https://www.dreamendstate.com/2021/02/19/user-centered-design-the-art-of-making-things-easy-to-use/

[3] P. Gomber and M. Haferkorn, "High frequency trading," in *Encyclopedia of Information Science and Technology, Third Edition*. IGI Global, 2015, pp. 1–9.

[4] M. Baldauf and J. Mollner, "High-frequency trading and market performance," *The Journal of Finance*, vol. 75, no. 3, pp. 1495–1526, 2020.

[5] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[6] C. Abras, D. Maloney-Krichmar, J. Preece *et al.*, "User-centered design," *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, vol. 37, no. 4, pp. 445–456, 2004.

[7] D. A. Norman, *The psychology of everyday things.* Basic books, 1988.

[8] J. D. Gould and C. Lewis, "Designing for usability: key principles and what designers think," *Communications of the ACM*, vol. 28, no. 3, pp. 300–311, 1985.

[9] J. Gulliksen, B. Göransson, I. Boivie, S. Blomkvist, J. Persson, and Å. Cajander, "Key principles for user-centred systems design," *Behaviour and Information Technology*, vol. 22, no. 6, pp. 397–409, 2003.

[10] B. Shneiderman, C. Plaisant, M. S. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos, *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.

[11] B. Shneiderman, "Direct manipulation: A step beyond programming languages," in *Proceedings of the Joint Conference on Easier and More Productive Use of Computer Systems.(Part-II): Human Interface and the User Interface-Volume 1981*, 1981, p. 143.

[12] N. Sharbatdar, “Supporting transportation decision-makers with tool design and data uncertainty visualizations,” Ph.D. dissertation, Polytechnique Montréal, 2020.

[13] R. A. Leite, T. Gschwandtner, S. Miksch, S. Kriglstein, M. Pohl, E. Gstrein, and J. Kuntner, “Eva: Visual analytics to identify fraudulent events,” *IEEE transactions on visualization and computer graphics*, vol. 24, no. 1, pp. 330–339, 2017.

[14] R. J. Bolton and D. J. Hand, “Statistical fraud detection: A review,” *Statistical science*, vol. 17, no. 3, pp. 235–255, 2002.

[15] Y. Kou, C.-T. Lu, S. Sirwongwattana, and Y.-P. Huang, “Survey of fraud detection techniques,” in *IEEE International Conference on Networking, Sensing and Control, 2004*, vol. 2. IEEE, 2004, pp. 749–754.

[16] M. Dumas, M. J. McGuffin, and V. L. Lemieux, “Financevis. net-a visual survey of financial data visualizations,” in *Poster Abstracts of IEEE Conference on Visualization*, vol. 2, 2014, p. 8.

[17] S. Ko, I. Cho, S. Afzal, C. Yau, J. Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and D. S. Ebert, “A survey on visual analysis approaches for financial data,” in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 599–617.

[18] J. D. Kirkland, T. E. Senator, J. J. Hayden, T. Dybala, H. G. Goldberg, and P. Shyr, “The nasd regulation advanced-detection system (ads),” *AI Magazine*, vol. 20, no. 1, pp. 55–55, 1999.

[19] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto, “Wirevis: Visualization of categorical, time-varying data from financial transactions,” in *2007 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2007, pp. 155–162.

[20] J. P. Steidlmayer and G. Kummel, “Financial data event flow analysis system with study conductor display,” Sep. 26 1995, uS Patent 5,454,104.

[21] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, “Temporal event sequence simplification,” *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2227–2236, 2013.

[22] M. L. Huang, J. Liang, and Q. V. Nguyen, “A visualization approach for frauds detection in financial market,” in *2009 13th International Conference Information Visualisation*. IEEE, 2009, pp. 197–202.

[23] W. N. Dilla and R. L. Raschke, "Data visualization for fraud detection: Practice implications and a call for future research," *International Journal of Accounting Information Systems*, vol. 16, pp. 1–22, 2015.

[24] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman, "Interactive information visualization to explore and query electronic health records," *Foundations and Trends in Human-Computer Interaction*, vol. 5, no. 3, pp. 207–298, 2013.

[25] M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner, "A survey of visualization systems for malware analysis," in *Eurographics Conference on Visualization (EuroVis)*, 2015, pp. 105–125.

[26] W. F. Sharpe, "The capital asset pricing model: a "multi-beta" interpretation," in *Financial Dec Making Under Uncertainty*. Elsevier, 1977, pp. 127–135.

[27] I. magazine. (2021) Total global aum 2019. [Online]. Available: https://www.ipe.com/total-global-aum-2019/10031648.article

[28] R. A. DeFusco, D. W. McLeavey, J. E. Pinto, D. E. Runkle, and M. J. Anson, *Quantitative investment analysis*. John Wiley & Sons, 2015.

[29] R. C. Grinold and R. N. Kahn, "Active portfolio management," 2000.

[30] J. Bender, R. Briand, D. Melas, and R. A. Subramanian, "Foundations of factor investing," *Available at SSRN 2543990*, 2013.

[31] M. Barra, "United states equity model handbook," *Technical Document, Berkeley*, 1998.

[32] F. Eugene and K. French, "The cross-section of expected stock returns," *Journal of Finance*, vol. 47, no. 2, pp. 427–465, 1992.

[33] X. Yue, J. Bai, Q. Liu, Y. Tang, A. Puri, K. Li, and H. Qu, "sportfolio: Stratified visual analysis of stock portfolios," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 601–610, 2019.

[34] M. Schaefer, F. Wanner, R. Kahl, L. Zhang, T. Schreck, and D. Keim, "A novel explorative visualization tool for financial time series data analysis," in *VAW2 011: The Third International UKVAC Workshop on Visual Analytics*, 2011.

[35] H. Ziegler, M. Jenny, T. Gruse, and D. A. Keim, "Visual market sector analysis for financial time series data," in *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2010, pp. 83–90.

[36] J. Alsakran, Y. Zhao, and X. Zhao, "Visual analysis of mutual fund performance," in *2009 13th International Conference Information Visualisation*. IEEE, 2009, pp. 252–259.

[37] H. Ziegler, T. Nietzschmann, and D. A. Keim, "Relevance driven visualization of financial performance measures," in *EUROVIS 2007*, 2007, pp. 19–26.

[38] ——, "Visual analytics on the financial market: Pixel-based analysis and comparison of long-term investments," in *2008 12th International Conference Information Visualisation*. IEEE, 2008, pp. 287–295.

[39] S. T. Lei and K. Zhang, "A visual analytics system for financial time-series data," in *Proceedings of the 3rd International Symposium on Visual Information Communication*, 2010, pp. 1–9.

[40] F. Xiong, E. Prakash, and K. Ho, "Er modeling and visualization of large mutual fund data," *Journal of visualization*, vol. 5, no. 2, pp. 197–204, 2002.

[41] S. Rudolph, A. Savikhin, and D. S. Ebert, "Finvis: Applied visual analytics for personal financial planning," in *2009 IEEE symposium on visual analytics science and technology*. IEEE, 2009, pp. 195–202.

[42] A. Savikhin, H. C. Lam, B. Fisher, and D. S. Ebert, "An experimental study of financial portfolio selection with visual analytics for decision support," in *2011 44th Hawaii International Conference on System Sciences*. IEEE, 2011, pp. 1–10.

[43] W.-A. Jungmeister and D. Turo, "Adapting treemaps to stock portfolio visualization," Tech. Rep., 1992.

[44] C. Csallner, M. Handte, O. Lehmann, and J. Stasko, "Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps," in *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)*. IEEE, 2003, pp. 203–208.

[45] T. Dwyer, "A scalable method for visualising changes in portfolio data," in *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24*, 2003, pp. 17–25.

[46] T. Dwyer and P. Eades, "Visualising a fund manager flow graph with columns and worms," in *Proceedings Sixth International Conference on Information Visualisation*. IEEE, 2002, pp. 147–152.

[47] T. Dwyer and D. R. Gallagher, "Visualising changes in fund manager holdings in two and a half-dimensions," *Information Visualization*, vol. 3, no. 4, pp. 227–244, 2004.

[48] V. Pham, N. Nguyen, J. Li, J. Hass, Y. Chen, and T. Dang, "Mtsad: Multivariate time series abnormality detection and visualization," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 3267–3276.

[49] V. Pham and T. Dang, "Cvexplorer: Multidimensional visualization for common vulnerabilities and exposures," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1296–1301.

[50] X. C. Chen, K. Steinhaeuser, S. Boriah, S. Chatterjee, and V. Kumar, "Contextual time series change detection," in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 503–511.

[51] V. V. Pham and T. Dang, "Mtdes: Multi-dimensional temporal data exploration system," in *2018 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2018, pp. 100–101.

[52] Google. (2021) More google cluster data. [Online]. Available: https://ai.googleblog.com/2011/11/

[53] Alibaba. (2021) Alibaba cluster trace program. [Online]. Available: https://github.com/alibaba/clusterdata

[54] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko, "Effectiveness of animation in trend visualization," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 6, pp. 1325–1332, 2008.

[55] E. R. Tufte, N. H. Goeler, and R. Benson, *Envisioning information*. Graphics press Cheshire, CT, 1990, vol. 2.

[56] M. J. Saary, "Radar plots: a useful way for presenting multivariate health care data," *Journal of clinical epidemiology*, vol. 61, no. 4, pp. 311–317, 2008.

[57] A. Dasgupta and R. Kosara, "Pargnostics: Screen-space metrics for parallel coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1017–1026, 2010.

[58] N. V. Nguyen and T. Dang, "Ant-sne: Tracking community evolution via animated t-sne," in *International Symposium on Visual Computing*. Springer, 2019, pp. 330–341.

[59] A. Inselberg and B. Dimsdale, "Parallel coordinates: a tool for visualizing multi-dimensional geometry," in *Proceedings of the First IEEE Conference on Visualization: Visualization90*. IEEE, 1990, pp. 361–378.

[60] S. T. Lei and K. Zhang, "Visual signatures for financial time series," in *Proceedings of the 2011 Visual Information Communication-International Symposium*, 2011, pp. 1–10.

[61] E. Sorenson and R. Brath, "Financial visualization case study: Correlating financial timeseries and discrete events to support investment decisions," in *2013 17th International Conference on Information Visualisation*. IEEE, 2013, pp. 232–238.

[62] D. A. Keim, "Designing pixel-oriented visualization techniques: Theory and applications," *IEEE Transactions on visualization and computer graphics*, vol. 6, no. 1, pp. 59–78, 2000.

[63] Y.-Y. Chan and H. Qu, "Finavistory: Using narrative visualization to explain social and economic relationships in financial news," in *2016 International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2016, pp. 32–39.

[64] Z. Niu, R. Li, J. Wu, D. Cheng, and J. Zhang, "iconviz: Interactive visual exploration of the default contagion risk of networked-guarantee loans," in *2020 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2020, pp. 84–94.

[65] K. W. Tsang, H. Li, F. M. Lam, Y. Mu, Y. Wang, and H. Qu, "Tradao: A visual analytics system for trading algorithm optimization," *arXiv preprint arXiv:2008.11319*, 2020.

[66] Y. Sun, H. Xiong, S. M. Yiu, and K. Y. Lam, "Bitvis: An interactive visualization system for bitcoin accounts analysis," in *2019 Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE, 2019, pp. 21–25.

[67] D. Albers, C. Dewey, and M. Gleicher, "Sequence surveyor: Leveraging overview for scalable genomic alignment visualization," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2392–2401, 2011.

[68] E. Tufte, "The visual display of quantitative information," 2001.

[69] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of time-oriented data*. Springer Science & Business Media, 2011.

[70] W. Müller and H. Schumann, "Visualization for modeling and simulation: visualization methods for time-dependent data-an overview," in *Proceedings of the 35th conference on Winter simulation: driving innovation*, 2003, pp. 737–745.

[71] E. Bertini, P. Hertzog, and D. Lalanne, "Spiralview: towards security policies assessment through visual correlation of network resources with evolution of alarms," in *2007 IEEE symposium on visual analytics science and technology*. IEEE, 2007, pp. 139–146.

[72] M. Wattenberg, "Sketching a graph to query a time-series database," in *CHI'01 Extended Abstracts on Human factors in Computing Systems*, 2001, pp. 381–382.

[73] J. J. Van Wijk and E. R. Van Selow, "Cluster and calendar based visualization of time series data," in *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99)*. IEEE, 1999, pp. 4–9.

[74] J. Zhao, F. Chevalier, and R. Balakrishnan, "Kronominer: using multi-foci navigation for the visual exploration of time-series data," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 1737–1746.

[75] J. Zhao, F. Chevalier, E. Pietriga, and R. Balakrishnan, "Exploratory analysis of time-series with chronolenses," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2422–2431, 2011.

[76] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, "Liverac: interactive visual exploration of system management time-series data," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 1483–1492.

[77] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda, "Two-tone pseudo coloring: Compact visualization for one-dimensional data," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 2005, pp. 173–180.

[78] H. Reijner *et al.*, "The development of the horizon graph," 2008.

[79] D. Albers, M. Correll, and M. Gleicher, "Task-driven evaluation of aggregation in time series visualization," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 551–560.

[80] M. Correll and M. Gleicher, "The semantics of sketch: Flexibility in visual query systems for time series data," in *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2016, pp. 131–140.

[81] D. Nadalutti and L. Chittaro, "Visual analysis of users' performance data in fitness activities," *Computers & Graphics*, vol. 31, no. 3, pp. 429–439, 2007.

[82] R. Kincaid and H. Lam, "Line graph explorer: scalable display of line graphs using focus+ context," in *Proceedings of the working conference on Advanced visual interfaces*, 2006, pp. 404–411.

[83] B. J. Swihart, B. Caffo, B. D. James, M. Strand, B. S. Schwartz, and N. M. Punjabi, "Lasagna plots: a saucy alternative to spaghetti plots," *Epidemiology (Cambridge, Mass.)*, vol. 21, no. 5, p. 621, 2010.

[84] W. Javed, B. McDonnel, and N. Elmqvist, "Graphical perception of multiple time series," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 927–934, 2010.

[85] L. Byron and M. Wattenberg, "Stacked graphs–geometry & aesthetics," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 6, pp. 1245–1252, 2008.

[86] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos, "Comparing similarity perception in time series visualizations," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 523–533, 2018.

[87] R. Brown. (2018) Nasdaq itch datafiles - market microstructure. [Online]. Available: https://rstudio-pubs-static.s3.amazonaws.com/388237_0f95ded0b8ad4026b8d43997323fccb7.html#:~:text=The%20financial%20information%20includes%20orders,speed%20makes%20a%20crucial%20difference.

[88] R. Cont and M. S. Mueller, "A stochastic pde model for limit order book dynamics," *arXiv preprint arXiv:1904.03058*, 2019.

[89] S. L. Franconeri, "The nature and status of visual resources." 2013.

[90] C. Healey and J. Enns, "Attention and visual memory in visualization and computer graphics," *IEEE transactions on visualization and computer graphics*, vol. 18, no. 7, pp. 1170–1188, 2011.

[91] G. A. Alvarez, T. Konkle, and A. Oliva, "Searching in dynamic displays: Effects of configural predictability and spatiotemporal continuity," *Journal of vision*, vol. 7, no. 14, pp. 12–12, 2007.

[92] J. M. Wolfe and T. S. Horowitz, "Five factors that guide attention in visual search," *Nature Human Behaviour*, vol. 1, no. 3, pp. 1–8, 2017.

[93] R. Arnheim, "The perception of maps," *The American Cartographer*, vol. 3, no. 1, pp. 5–10, 1976.

[94] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *The craft of information visualization*. Elsevier, 2003, pp. 364–371.

[95] A. Barsky, T. Munzner, J. Gardy, and R. Kincaid, "Cerebral: Visualizing multiple experimental conditions on a graph with biological context," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 6, pp. 1253–1260, 2008.

[96] R. García Martín, J. P. de Castro Fernández, E. Verdu Perez, M. J. Verdú Pérez, and L. M. Regueras Santos, "An ols regression model for context-aware tile prefetching in a web map cache," *International Journal of Geographical Information Science*, vol. 27, no. 3, pp. 614–632, 2013.

[97] X. Yue, X. Shu, X. Zhu, X. Du, Z. Yu, D. Papadopoulos, and S. Liu, "Bitextract: Interactive visualization for extracting bitcoin exchange intelligence," *IEEE transactions on visualization and computer graphics*, vol.25, no. 1, pp. 162–171, 2018.