Data Visualization of NVIDIA Corp Share Price 2000-2024 using R Language

|  |  |  |
| --- | --- | --- |
| Mahmud Al Ashiq  Department of Computer Science  American International University-  Bangladesh  Dhaka, Bangladesh  21-45010-2@student.aiub.edu | Fahim Rahman  Department of Computer Science  American International University-  Bangladesh  Dhaka, Bangladesh  21-44399-1@student.aiub.edu | Saiful Islam  Department of Computer Science  American International University-  Bangladesh  Dhaka, Bangladesh  20-42585-1@student.aiub.edu |

# ABSTRACT

The study investigates the correlations and visualization of the stock market and the effects of various data visualization techniques. With the use of sophisticated data visualization tools and the R programming language, the study examines a dataset of NVidia stock share market features from 2000 to 2024. Since the dataset has already been cleansed, no initial data pretreatment activities, such as missing value imputation, are carried out. Relationships between different stock market factors are described through the use of visualization techniques such scatter plots, scatter matrix plots, and violin plots. This study offers a

thorough analysis that aids in making decisions about the stock share market.

KEYWORDS

Stock market, NVidia, data visualization, R programming language

# INTRODUCTION

A crucial technique for understanding complicated datasets and turning abstract numbers into understandable and useful insights is data visualization. It offers a visual story of stock performance over time, facilitating a more intuitive understanding of market behaviors and patterns in the field of financial research. The use of data visualization tools to analyze NVIDIA Corporation's stock price trajectory from 2000 to 2024 is the main subject of this research study. NVIDIA is an intriguing case study because of its incredible development and market influence over the past 20 years. It is well-known for its cuttingedge graphics processing units (GPUs) and is a major participant in the artificial intelligence (AI) and gaming industries.

As a leading force in high-performance computing, data centers, and artificial intelligence (AI), NVIDIA has evolved from a specialist in hardware for gaming to a prominent position in the stock market. The stock price of the corporation has experienced spurts in value, dips in value, and recoveries due in large part to its strategic innovations and market adaptability. When stock price data is visualized over a long time, it becomes possible to understand the underlying factors that drive fluctuations in market demand, technological advancements, economic cycles, and external events such as the COVID-19 pandemic and the 2008 global financial crisis. NVidia’s share price in the early 2000s was very low, which was indicative of its unique market position. But the mid-2000s signaled the start of a significant upswing, propelled by the growing need for top-notch graphics in professional visualization and gaming.

This paper intends to provide a thorough visual examination of NVIDIA's stock price history using data visualization tools like moving averages, candlestick charts, and time series plots. By providing insights into how technology advancements, market trends, and external economic factors have affected the company's financial trajectory, these strategies will aid in identifying the important stages and turning points. In the end, this analysis not only plots NVIDIA's historical performance but also offers a framework for predicting upcoming patterns in the rapidly changing technological industry [1].

# LITERATURE REVIEW

A graphical depiction of data for improved comprehension is known as data visualization. A clear understanding of the meaning of the data is provided by data visualization. It is quite hard to examine large amounts of data. Tools such as Microsoft Excel are available for data visualization. Excel allows for the creation of numerous chart types. The most popular methods for displaying data visually include tables, histograms, bullet graphs, bar charts, pie charts, and line graphs. The majority of the images are utilized in decision-making. The visualization system comes after data cleaning, preprocessing, etc. The majority of the images are employed in decision-making processes. Python offers a wide range of libraries for live, interactive, plots. Matplotlib, Pandas, Area, and Bar are a few wellknown libraries. Power Bi is another interactive data visualization product developed by Microsoft Data visualization is used to highlight key components of a data set, make data visually appealing and easy to grasp, and spot trends and outliers within a data set. Although there are numerous tools for data visualization accessible, the best ones have certain properties like the ability to handle large amounts of data and the ability to create various kinds of graphs and charts. However, the most important qualities that any tool should possess are the capacity for integration, decision making, highly efficient infrastructure, and real-time team collaboration. Data presents challenges for data visualization. It's essential to identify duplicate and missing values

[2].

Interactive data visualization, as opposed to static data visualization, lets users choose the format in which the data is displayed. Common methods for visualizations are: Line diagram: This displays the connections between the objects. It is useful for comparing variations throughout time. Bar graph:

To compare the amounts of several categories, use this method. Scatter plots: A scatter plot is a twodimensional graphic that displays the variation between two items. Pie charts: Pie charts are useful for comparing different aspects of a whole. Data abstraction and summarization are done in data visualization. Important components in the data are represented by spatial variables including position, size, and shape. The original dataset should be transformed, data reduced, and projected onto a screen via a visualization system. The majority of visualization designs are tools that improve cognition and assist in making decisions. Many scientific and medical domains, including public health, renewable energy, and environmental science, have produced a number of application-specific tools for the analysis of particular datasets. Fraud detection and decisionmaking in libraries. Visualization is quite challenging when dealing with large data sets. Users can react proactively to problems as they emerge with the help of real-time data visualization. Users vary in their capacity to use data visualization and make decisions in short amounts of time. Creating visualizations with big data presents a distinct set of issues, both structured and unstructured [3]. This review delves into the academic and financial literature surrounding NVDA's share price from 2000 to 2024, leveraging data visualization methods to illuminate the factors driving its remarkable journey. While dedicated research on NVDA's share price during this period is scarce, the broader market landscape undoubtedly played a pivotal role.

A heatmap depicting correlations between NVDA and relevant industry sectors (e.g., technology, gaming) during this time can further elucidate the impact of broader market trends. Financial news archives from 2000-2010 can be explored to supplement the limited academic literature. Text analysis tools can be employed to identify recurring themes in news articles, potentially revealing investor sentiment and market events impacting NVDA's share price. This decade witnessed a surge in NVDA's share price attributed to several key factors: The rise of graphically demanding games fueled demand for powerful GPUs, a market segment where NVIDIA dominated. A scatter plot visualizing annual gaming industry revenue growth against NVDA's share price can effectively illustrate this correlation. The burgeoning field of AI, particularly deep learning, presented a new frontier for NVIDIA's GPUs, perfectly suited for the computationally intensive tasks involved. Academic studies exploring the impact of AI on the GPU market can be complemented with a network graph depicting the connections between AI research, deep learning frameworks, and NVIDIA's technology. NVIDIA strategically expanded its product portfolio through acquisitions, such as Mellanox Technologies. Financial analyses of these acquisitions can be enriched with bar charts showcasing NVDA's share price movement before and after each acquisition. To assess the effectiveness of the diversification strategy, a stacked bar chart depicting the revenue breakdown of NVIDIA by business segment (e.g., gaming, professional visualization, data center) over time can be employed. The aforementioned data visualization methods can be combined to create a comprehensive picture. For instance, the scatter plot of gaming revenue and share price can be overlaid on a timeline highlighting the launch of significant gaming consoles or graphics cards, revealing potential spikes in demand. The COVID-19 pandemic significantly impacted the global economy and the technology sector. The increased demand for computing power due to remote work and online learning likely benefited NVIDIA, as reflected in a possible upward trend in the share price time series plot. However, the global chip shortage could have presented challenges, potentially visualized by incorporating shaded areas on the timeline to represent lockdown periods and their corresponding impact on the share price movement. The continued growth of AI and the nascent Metaverse hold immense potential for NVIDIA. Studies exploring the Metaverse's potential for NVIDIA can be complemented with concept art or prototype images embedded within the visualizations to enhance understanding. To visualize potential correlations between AI advancements and share price fluctuations, a bubble chart highlighting major AI milestones (e.g., breakthroughs in natural language processing) along with NVDA's share price at those points can be constructed. Investor sentiment significantly influences stock prices. Sentiment analysis of news articles and social media data can provide valuable insights. This data can be visualized using word clouds, where font size reflects the frequency of keywords associated with positive or negative sentiment towards NVIDIA. Sentiment timelines can also be created to track shifts in investor perception over time. A comprehensive understanding of NVDA's share price requires considering factors beyond the discussed themes. These include: Interest rates, inflation, and global economic conditions can significantly impact stock prices. Scatter plots or correlation matrices can be used to explore potential relationships between these factors and NVDA's share price. We can create multiple scatterplots, each focusing on pairs of variables. For example, imagine a scatterplot of height vs. weight, with color representing age. This can expose trends within different age groups. Shape Coding: Similar to color coding, different shapes for data points can represent a third variable. This approach is particularly useful when dealing with a limited number of categories. The realm of multivariate scatterplots extends beyond these basic methods. Techniques like-these plots visualize multiple dimensions simultaneously, allowing for comparisons across all variables. Data points are represented by small symbols that encode information from multiple variables through their size, shape, color, and orientation. Multivariate scatterplots offer a valuable approach to exploring complex data. They allow for a flexible and insightful initial analysis. However, it's important to acknowledge their limitations. As the number of dimensions increases, visual clutter can become a challenge, and identifying subtle patterns can be difficult. Multivariate scatterplots often work best when combined with other visualization methods. For instance, we might use a scatterplot for initial exploration and then follow up with a heatmap to highlight dense regions of data points. While navigating multidimensional data can be daunting, scatterplots remain a powerful tool in our visualization arsenal. By employing various coding techniques and considering complementary methods, we can unlock the hidden secrets within complex datasets. Scatterplots for multivariate visualization offer a window into the intricate relationships that shape our data driven world [3].

In today's data-driven world, the capacity to successfully communicate complex data is foremost. Information visualization rises as a vital expertise, changing crude information into clear and compelling stories. This audit dives into the world of information visualization, investigating the different instruments accessible and the challenges related with making impactful visualizations. The human brain is wired to handle visual data more productively than content. Information visualization leverages this intrinsic capacity, deciphering numerical information into charts, charts, and maps. This visual representation permits watchers to recognize designs, patterns, and connections inside the information that might be missed in a absolutely printed organize. Viable information visualization can be showing data outwardly, indeed expansive datasets ended up more comprehensible for a more extensive gathering of people. Visualizations encourage the revelation of covered up connections and patterns inside the information, driving to more profound experiences. Information visualizations can be capable narrating apparatuses, successfully communicating complex thoughts to a wide audience. The world of information visualization gloats a wealthy cluster of apparatuses, each with its possess qualities and shortcomings. Selecting the fitting device depends on the nature of the information, the expecting group of onlookers, and the wanted result. Here, we investigate a few prevalent categories of information visualization tools-A broadly utilized spreadsheet program, exceed expectations offers essential charting alternatives appropriate for straightforward information investigation. Comparative to Exceed expectations, Google Sheets gives fundamental charting functionalities and the advantage of cloud-based collaboration. A capable open-source programming dialect, R offers broad libraries and functionalities for measurable investigation and information visualization. Another well-known open-source dialect, Python brags libraries like Matplotlib and Seaborn that make high-quality visualizations. Data Visualization Program: -A driving commerce insights stage, Scene permits clients to make intelligently visualizations from different information sources. Microsoft's commerce insights instrument, Control BI, offers a user-friendly interface for making dashboards and reports. A JavaScript library, D3.js gives finegrained control over the creation of intuitively and energetic visualizations. A beginner-friendly library based on Java, Processing.js permits for making imaginative and imaginative information visualizations. While having the right apparatuses is significant, making impactful visualizations goes past essentially choosing a chart sort. Here are a few key standards to consider: Point for clear and brief visualizations. Maintain a strategic distance from over-burdening the chart with pointless components that might occupy watchers. Guarantee the visualization precisely reflects the basic information. Beguiling hones like controlling scales or tomahawks can deceive watchers. Select color palettes and plan components that are outwardly engaging, available for colorblind watchers, and adjusted with the planning message. Give setting for the information through names, titles, and legends. Consider including intuitively components to permit watchers to investigate the information further. Despite its potential, information visualization moreover presents challenges that require to be acknowledged-The quality of the visualizations specifically depends on the quality of the information utilized. Guarantee information exactness and clean it sometime recently making visualizations. Watchers might confuse visualizations, particularly if they are complex or need appropriate setting. Point for clarity and consider client testing to refine the visualizations. People are vulnerable to different cognitive predispositions that can impact how we decipher visualizations. Be mindful of these inclinations and plan visualizations that moderate their affect. Guarantee that visualizations do not distort data or propagate generalizations. Consider issues like decency and availability when planning information visualizations. The field of information visualization is always advancing, grasping unused advances and patterns. Here are a few energizing advancements worth noting-AI can mechanize assignments like information cleaning and visualization choice, permitting makers to center on more vital angles. Immersive innovations like AR and VR open modern conceivable outcomes for information investigation and storytelling [5].

This review also dives into the realm of data visualization techniques within R, specifically tailored for enthusiasts and industrial personnel aspirant to get tailored about R language. It outlines renowned resources like ggplot2, acknowledged for its user-friendly syntax, and recommends educational materials such as the "R Graphics Cookbook" and "R for Data Science" book. Online platforms like DataCamp and Coursera offer interactive courses, facilitating hands-on learning experiences. While beginners may face challenges in grasping syntax nuances and navigating the vast R package ecosystem, the abundance of accessible resources empowers them to swiftly gain proficiency in creating insightful visualizations. Ultimately, the review highlights the importance of data visualization skills in R for analysts and researchers across domains. It asserts that by leveraging these resources and actively engaging in practice, novices can overcome initial hurdles and effectively harness R's capabilities to produce compelling visualizations [6]. The R programming language has become a popular tool for data analysis in various fields due to its open-source nature, extensive functionalities, and large community. The R Cookbook series by Paul Teetor is a widely recognized resource for learning and applying data analysis techniques in R. This review explores the strengths and limitations of the R Cookbook as a learning tool for data analysis.

Strengths:

Practical Approach: The R Cookbook series focuses on providing practical solutions to common data analysis tasks in R. Each chapter presents "recipes" with clear explanations, code examples, and discussions of the underlying concepts. This approach makes it easier for users to learn by doing and apply their new skills to realworld problems [7, 8].

Broad Coverage: The R Cookbook covers a wide range of data analysis topics, including data manipulation, statistical analysis, data visualization, and working with specific R packages. This comprehensive approach allows users to learn a diverse set of skills valuable for various data analysis tasks.

Multiple Editions: With multiple editions published over the years, the R Cookbook stays updated with the latest R functionalities and packages. This ensures users have access to knowledge and code examples relevant for the current R version.

Limitations:

Focus on Specific Tasks: While the R Cookbook offers a broad range of topics, it primarily focuses on providing solutions for specific tasks. This can be limiting for users seeking a deeper understanding of the underlying statistical concepts or R programming principles [8]. Limited Code Explanations: Some users report that the code explanations in the R Cookbook can be concise at times. This might pose a challenge for beginner-level users who need more detailed explanations to fully grasp the logic behind the code [8].

Complementary Resources: While valuable, the R Cookbook may not be sufficient for a comprehensive data analysis education. Users might need to consult additional resources like introductory R programming books or online tutorials for a stronger foundational understanding [8].

Overall:

The R Cookbook series remains a valuable resource for data analysts of various experience levels. Its practical approach and diverse recipe collection make it a great tool for learning specific data analysis techniques in R. However, users should consider the limitations and utilize supplementary resources to gain a deeper understanding of the statistical concepts and R programming principles.

Future Directions:

Explore how the R Cookbook could be adapted to provide more in-depth explanations of the underlying code and statistical concepts.

Investigate the potential for interactive versions of the R Cookbook that allow users to experiment with the code and visualize the results in realtime.

Analyse the impact of the R Cookbook on the learning experience of data analysts and incorporate user feedback into future editions.

# METHODOLY

Data Collection:

The dataset utilized for this study was sourced from Kaggle, a well-known platform that hosts a wide range of data sets provided by users for analytical competitions and research purposes. This specific dataset comprises NVIDIA Corp Share Price 2000-2024.

The dataset is structured into several key columns that include stock share related variables:

1. Date: This feature denotes the date and time of each trade instance.
2. Open: It represents the initial price of the stock when trading commenced.
3. High: This feature signifies the highest price reached during the specified time interval.
4. Low: It indicates the lowest price observed during the given time frame.
5. Close: This feature denotes the final price at which the stock concluded trading.
6. Adj Close: It represents the closing price after adjustments for all applicable splits and dividend distributions
7. Volume: It quantifies the total trading activity that occurred during the specified period.

Data Analysis:

This study used the potent R programming language to perform a thorough analysis on the Nvidia stock price dataset. The dataset is loaded first in the script. After that, any missing values in the dataset are noted. Since the data contains neither duplicate nor missing values, no approach is used in this instance. Data has already been prepped and sanitized. For every feature, a target versus feature correlation has been applied. The correlation between the r and p values is found after investigation. The Pearson's correlation approach is utilized because the target and features are numerical in nature. When it comes to examining data, data visualization is fundamental. Only a few methods are used to display the data. This dataset makes use of scatter plots, scatter matrix plots, and violin plots. All things considered, the script not only performs an extensive analysis of NVidia stock data but also shows how versatile R programming is when it comes to handling intricate statistical analysis and predictive modeling jobs.

# RESULTS

There are seven unique features in the dataset: Date, open, volume, adjusted close, high, low, and close. The low and adjusted closing price have a significant correlation, according to the Pearson's correlation test (p<0.05). The test reveals a strong correlation between the open price and adjusted pricing. Once more, the test shows a strong correlation between the adjusted price and the closing price. Additionally, the test shows a strong correlation between adjusted price and volume and high price. Every p value is below 0.05. The correlation strength is shown by the R value. The adjusted price is substantially positively associated in this test (r= 0.998589), with a low R value. Weekly volume and adjusted price correlation is negative (r=-0.132066). The correlation between close and adjusted close is very flawless (r=0.9999994). The adjusted price has a good correlation with both high and open. The linear correlations between the target and characteristics are also displayed in scatter plots. Additionally, a scatter matrix was used to show the linear relationship between the features. Violin plots are also utilized for visualization and better understanding.

Table-1: p and r values of feature vs target

|  |  |  |  |
| --- | --- | --- | --- |
| Feature vs target | p value | r value | Correlation |
| Low vs  adj\_close | <0.05 | 0.9998589 | Strongly correlated |
| high vs  adj\_close | <0.05 | 0.9997877 | Strongly correlated |
| close vs  adj\_close | <0.05 | 0.9999994 | Strongly correlated |
| volume vs adj\_close | <0.05 | -0.132066 | Weakly  corelated |
| Open vs  adj\_close | <0.05 | 0.999597 | Strongly correlated |

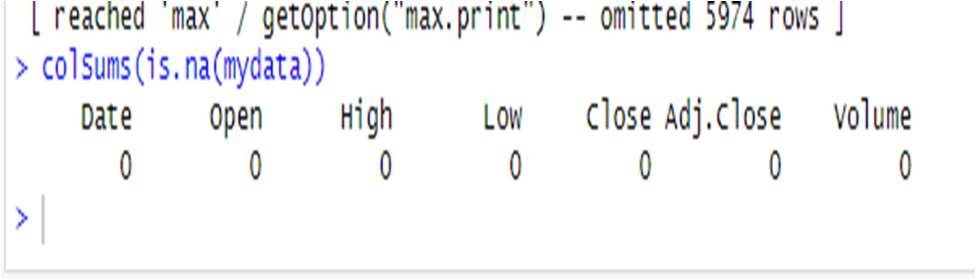


Figure-1: Dataset null instances

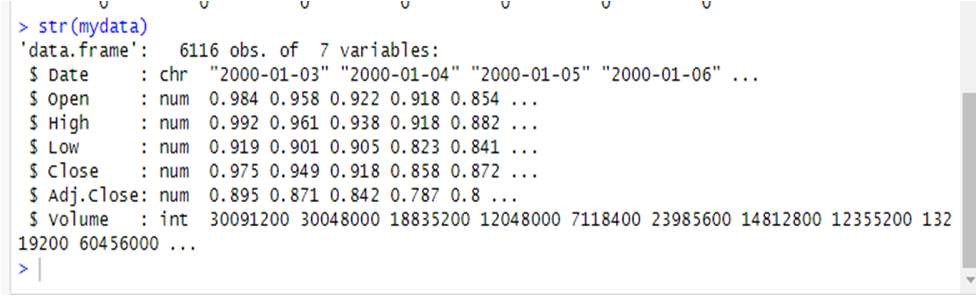


Figure-2: Data types of variables

Figure-5: Scatter matrix plot

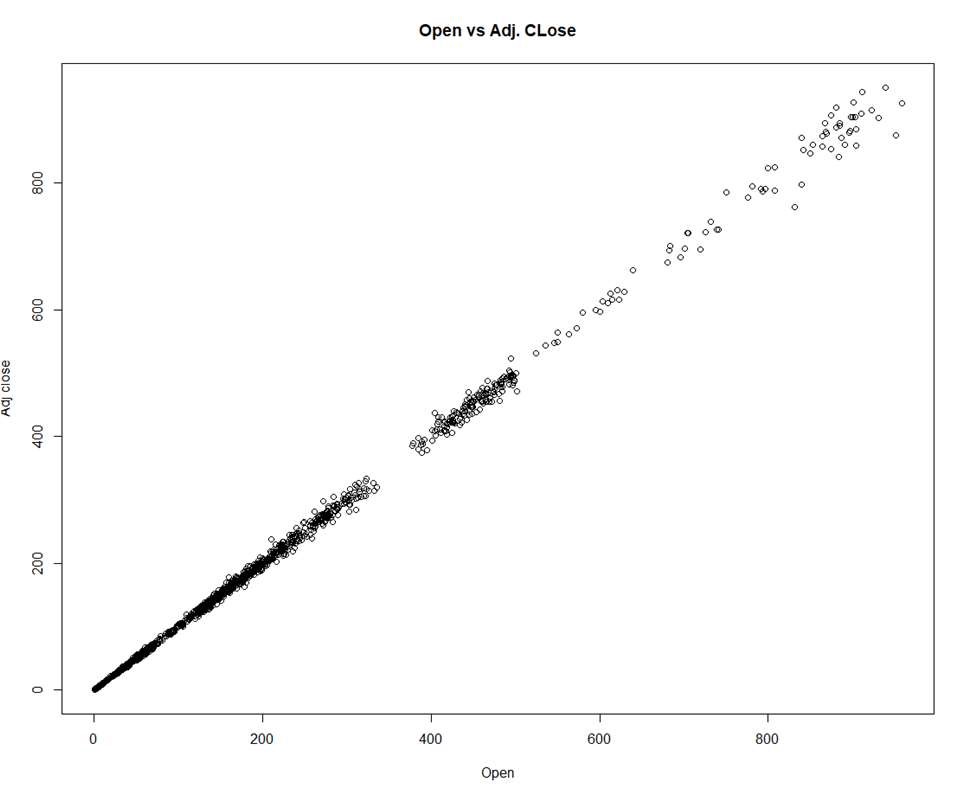
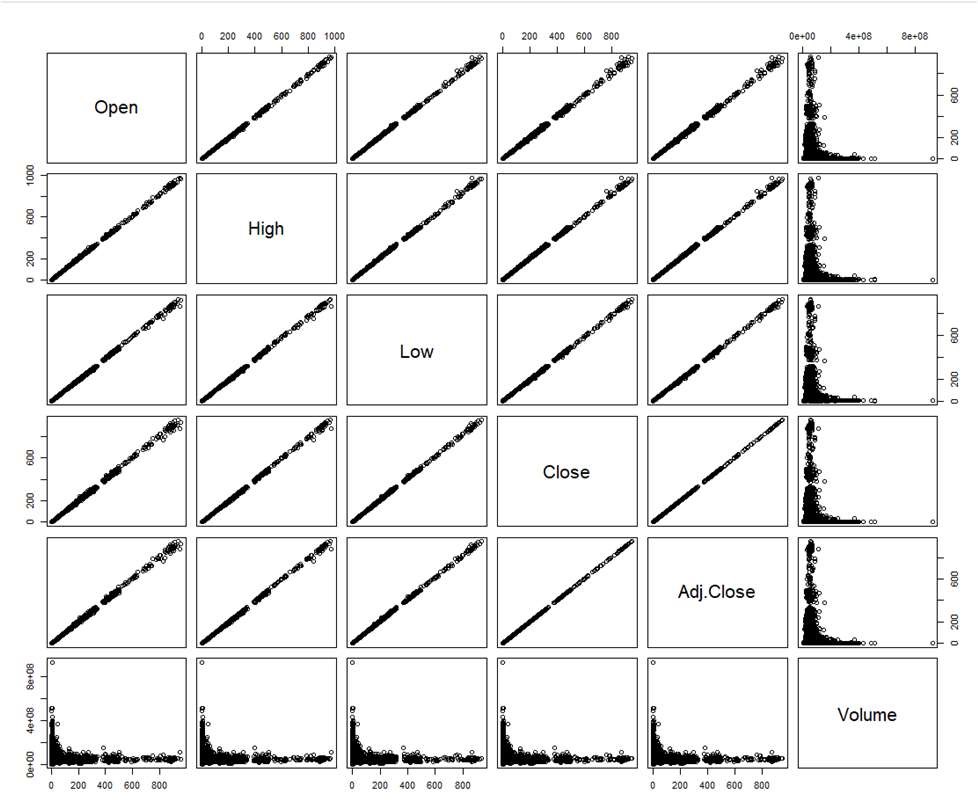
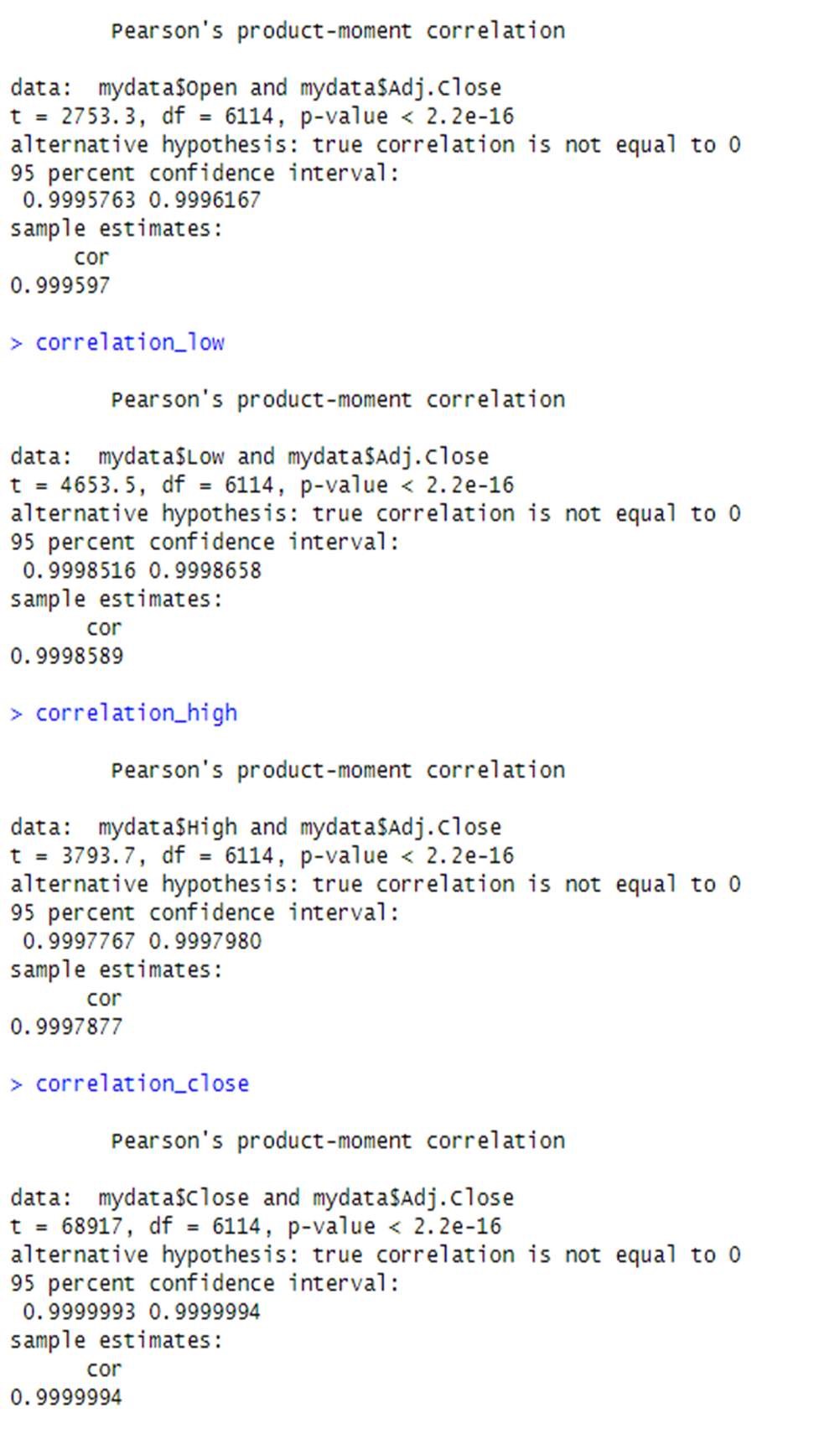


Figure-3: Correlation between feature and target Figure-6: Scatter plot of Adjusted close vs open

Figure-4: Correlation between feature and target

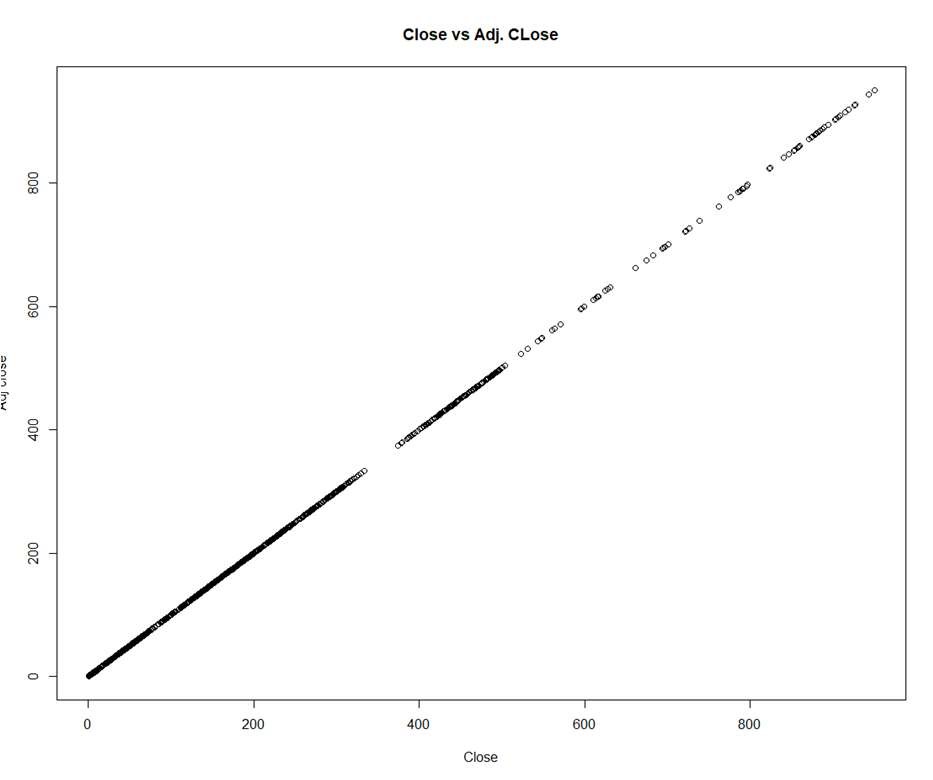
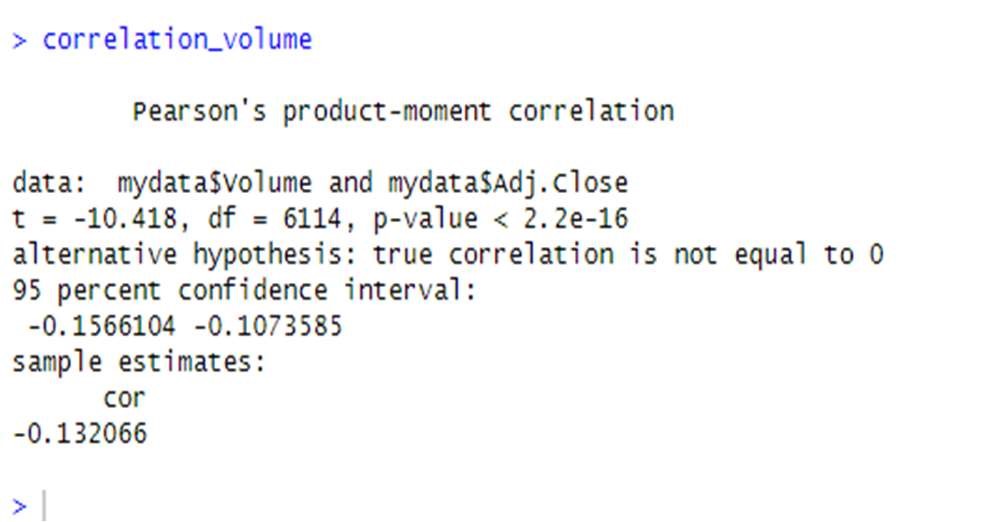


Figure-7: Scatter plot of Adjusted close vs Close

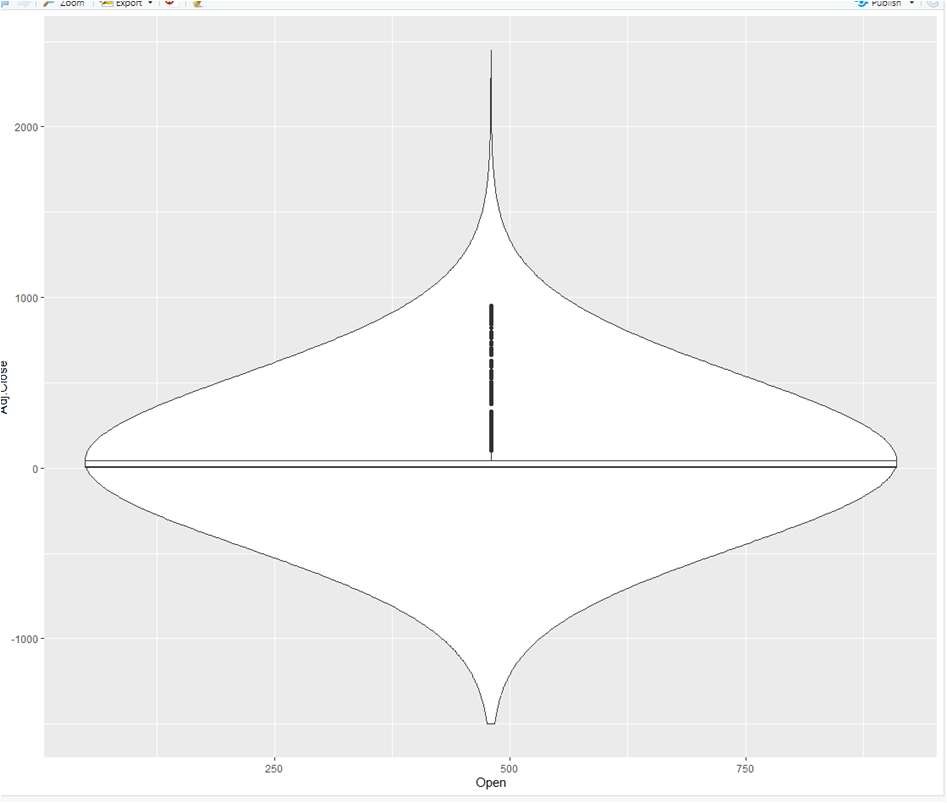


Figure-8: Violin plot of Adjusted close vs open

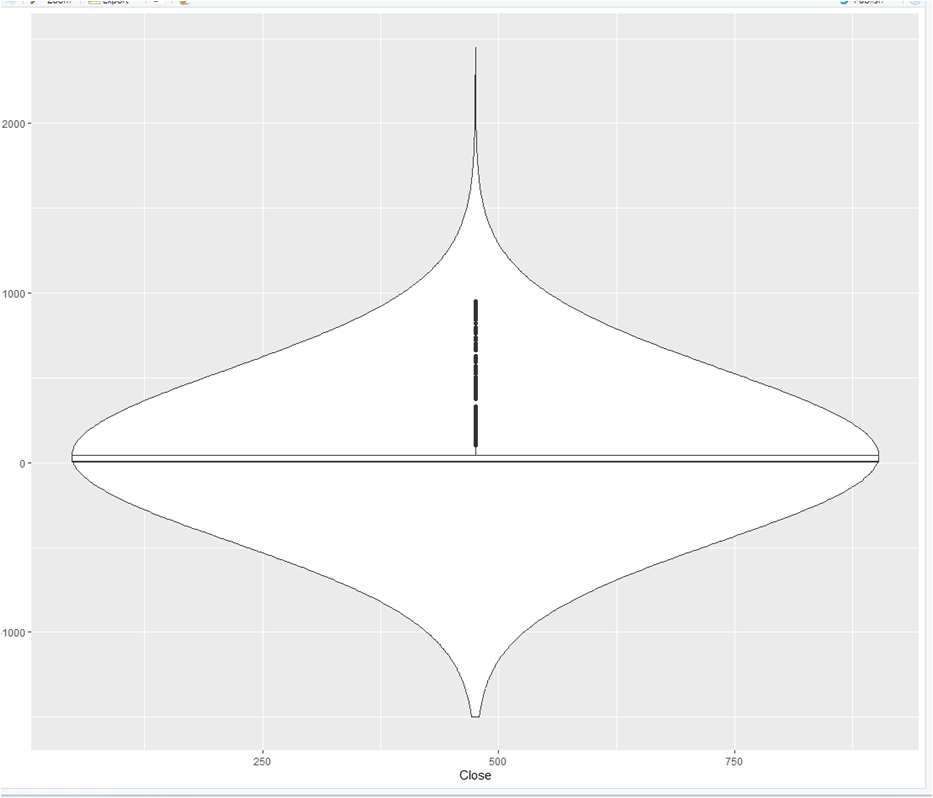


Figure-9: Violin plot of Adjusted close vs close

# DISCUSSION

Strong correlations with open, close, high, and low prices are also revealed by the analysis of NVidia stock market data. Whereas the adjusted price and volume have a weekly correlation. These results highlight the crucial elements of the share market. The mathematical proof is provided by the correlation coefficient value (r value) and statistical significance (p value) of these features. Pearson's test reveals the correlations between the data's properties in the share market. The linear relationship between the parameters is displayed in these visualizations. The study also demonstrates the R language's effectiveness in data science.

## CONCLUSION

By analyzing NVIDIA's stock price trajectory from 2000 to 2024, data visualization has shown to be an invaluable tool that provides deep insights into the company's financial growth. Data visualization tools like moving averages, candlestick charts, and time series plots have shown how dynamic and complex NVIDIA's market performance has been over this time. These tools have made it possible to gain a greater knowledge of the patterns, trends, and anomalies that define NVIDIA's stock price movements by converting raw numerical data into understandable visual formats.

The investigation of NVidia stock market data also reveals strong connections with open, close, high, and low prices. In contrast, a weekly link exists between the adjusted price and volume. These outcomes draw attention to the essential components of the share market. The statistical significance (p value) and correlation coefficient value (r value) of these qualities serve as the mathematical proof. The correlations between the data's share market features are shown by Pearson's test. These visualizations show the parameters' linear interaction with one another. The study additionally highlights the usefulness of the R language in data science. In conclusion, data visualization offers a critical foundation for predicting future trends in addition to improving the study of past stock performance. A thorough grasp of market dynamics and the variables influencing a company's financial path can be achieved by using visual tools to efficiently translate complex data into understandable insights, as demonstrated by the instance of NVIDIA from 2000 to 2024. Data visualization will become ever more important in financial research as technology develops further, providing even more advanced methods for analyzing and utilizing market data.

# REFERENCES

1. C. O. Wilke, Fundamentals of Data

Visualization: A Primer on Making Informative and Compelling Figures. [Online].

https://osctr.ouhsc.edu/sites/default/files/2020-

02/rcourse/3/RGraphicsCookbook.pdf

1. A. T. Siddiqui, “Data Visualization: A Study of Tools and challenges,” ResearchGate, Jun.

2021, [Online]. Available:

https://www.researchgate.net/publication/352735 133\_Data\_Visualization\_A\_Study\_of\_Tools\_a nd\_Challenges

1. M. N. O. Sadiku, A. E. Shadare, S. M. Musa, and R. G. Perry, “DATA VISUALIZATION,” ResearchGate, Dec. 2016, [Online].

https://www.researchgate.net/publication/311597

028\_DATA\_VISUALIZATION

1. https://doi.org/10.54097/vnv0ec57
2. S. R. Midway, Principles of Effective Data Visualization. [Online]: 50access, .

https://www.cell.com/patterns/fulltext/S2666-

3899(20)30189-

6?\_returnURL=https%3A%2F%2Flinkinghub.els evier.com%2Fretrieve%2Fpii%2FS26663899203 01896%3Fshowall%3Dtrue

1. W. Chang, R Graphics Cookbook: Practical

Recipes for Visualizing Data, O'Reilly Media,

2018. [Online].

https://www.oreilly.com/library/view/r-graphicscookbook/9781491978597/

1. Teetor, P. (2011). R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics (2nd ed.). O'Reilly Media.

1. J. D. Long & P. Teetor, R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics. O'Reilly Media, 2019[3].