ASSIGNMENT NUMBER 2: ANALYSIS OF YAHOO FINANCE STOCK DATA & VISUALIZATIONS

Shweta Shekhar *CS-GY 6313 B*

Introduction

In this assignment involves creating two types of static visualizations for each question: one that clearly communicates insights from the data and another that intentionally misleads the viewer. The visualizations use stock price data from the S&P 500 and Dow Jones indices, based on the yfinance dataset, which includes historical data such as stock prices (open, close, high, low), volume, dividends, and stock splits.

Question 1: What is the year-over-year stock price growth for the top 10 companies in the S&P 500?

The visualization presents the year-over-year stock price growth for the top 10 S&P 500 companies. Companies are sorted by growth percentage in descending order to highlight top and bottom performers. A neutral sky blue bar chart with subtle grid-lines ensures clarity, and exact percentages are labeled for transparency. The consistent color scheme avoids bias, keeping the focus on the data. The chart allows easy interpretation of trends and individual company performance.

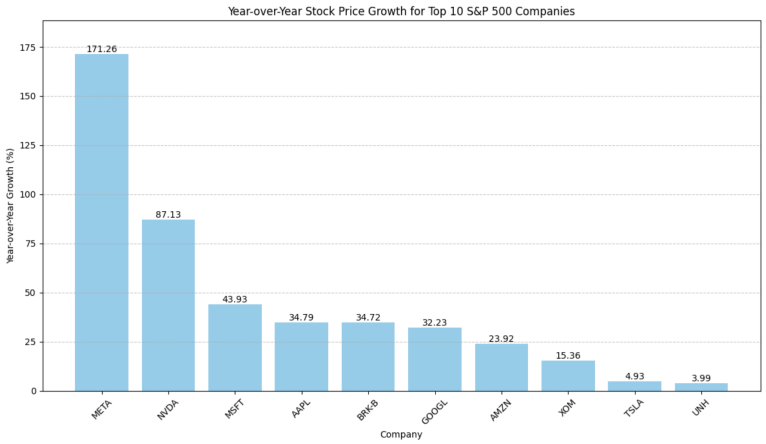

Figure 1: Year-over-Year Stock Price Growth for Top 10 Companies in S&P 500

Question 2: How did major market events affect the stock performance of the top 10 S&P 500 companies?

This visualization aims to answer the question: 'How did Apple, Microsoft, and Google perform during the COVID-19 crash?'

This visualization presents stock price changes for Apple, Microsoft, and Google during the COVID-19 crash. A dual Y-axis was used to emphasize Google's stock movements, which appear more dramatic than those of Apple and Microsoft, even though all three companies experienced similar volatility. The time frame is selectively limited to only the most volatile months of the COVID-19 crash, and titles direct attention to Google, framing its movements as more significant than they are.

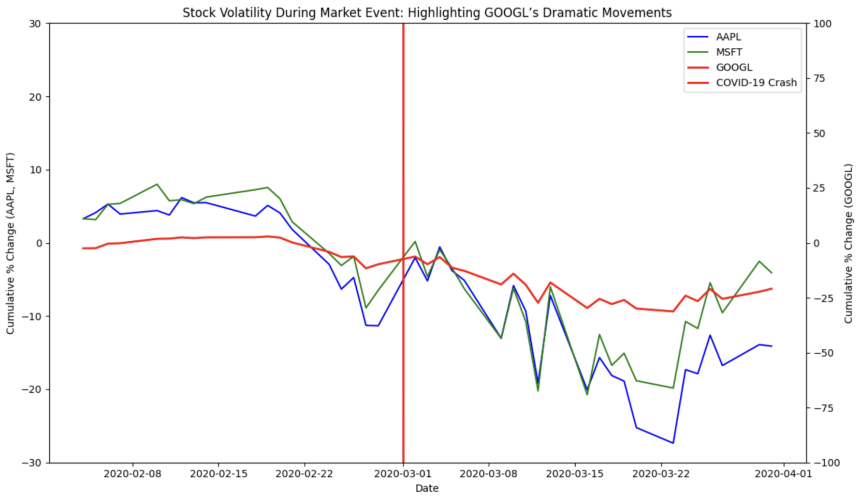

Figure 2: Out-performing Companies During Market Events

Question 3: What is the year-over-year percentage change in stock price for the selected Dow Jones companies?

The chart includes all companies from the list, and the bars show their percentage change over the past year. The Y-axis starts at zero, ensuring an accurate representation of the percentage changes across all companies, avoiding exaggeration. The grid-lines and labels ensure clarity, making it easy for viewers to interpret the data without bias or distraction.

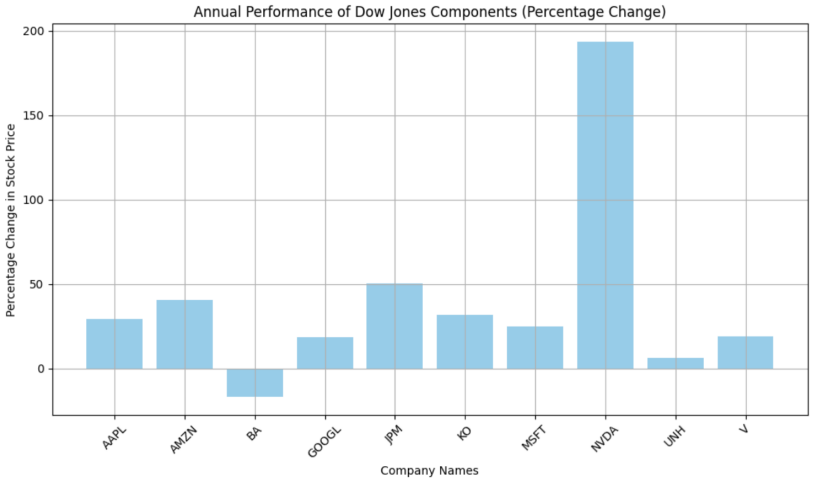

Figure 3: Year-over-Year Stock Price Change for Dow Jones Companies

Question 4: Which 3 companies have out-performed in the Dow Jones index?

This chart shows the top 3 companies in the Dow Jones index, selectively filtering the data to create a misleading impression that all Dow Jones companies have performed well. The Y-axis starts at 10%, exaggerating the differences between the top performers and making relatively small changes appear more dramatic than they are. The title and color scheme further emphasize strong performance, even though under-performers are excluded.

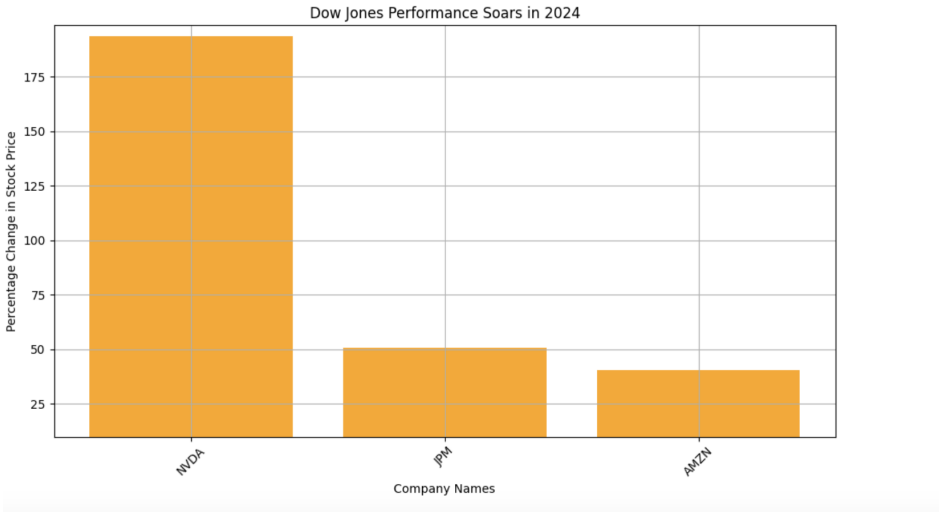

Figure 4: Top 3 Companies Performances in Dow Jones Index

Conclusion

This analysis demonstrates the difference between earnest and deceptive visualizations based on four distinct questions.

The visualizations for Question 1 (Year-over-Year Stock Price Growth) and Question 3 (Year-over-Year Percentage Change in Dow Jones Companies) are earnest visualizations. They use transparent Y-axis scaling, neutral color schemes, and provide clear, unbiased representations of the data. These charts allow viewers to easily interpret trends and company performance without introducing any misleading elements. The focus is on providing clarity to the general audience without distorting the facts.

In contrast, the visualizations for Question 2 (Stock Performance during Market Events) and Question 4 (Top 3 Dow Jones Companies) are deceptive. These visualizations manipulate the Y-axis, selectively filter data, and use exaggerated titles or color schemes to mislead viewers into perceiving larger differences or more dramatic trends than the data supports. While the data itself is technically accurate, these subtle design choices distort the viewer's understanding of the information.

These examples highlight how subtle design choices, such as Y-axis manipulation or selective filtering, can greatly influence how data is interpreted. By carefully considering these factors, it becomes clear that accurate, honest visualizations are crucial for conveying reliable insights.

Shortcomings of Deceptive Visualizations: While intentionally misleading, the deceptive visualizations could have been more subtle. For instance, the Y-axis manipulation in Question 2 could have been less obvious, and the data filtering in Question 4 might have been hidden more effectively. Enhancing the subtlety of deceptive visualizations would make it harder for viewers to immediately detect the manipulation. To further improve the deceptive visualizations:

Question 2: Including more data points from the recovery period could create an illusion of a more stable trend while still misleading the viewer into thinking the volatility was more significant during certain periods.

Question 4: The color scheme could be made less striking to avoid drawing unnecessary attention to the manipulated aspects, making the deceptive choices less obvious.

In conclusion, even minor design choices—such as Y-axis manipulation, selective filtering, or color schemes—can significantly influence how data is perceived. Maintaining accuracy and transparency in visualizations is essential to provide reliable insights. While deceptive visualizations may showcase technical skills, it is crucial to uphold ethical responsibility. Misleading visuals can lead to real-world consequences and poor decisions. Therefore, honest, accurate visualizations are vital, especially when informing key decisions in business, policy, and beyond.