Understanding Asset Correlations Through Visualisation

A Beginner-Oriented Quantitative Finance Project

Cameron Pritchard

February 2, 2026

Abstract

This project explores asset return correlations using correlation heatmaps applied to major technology stocks. The motivation behind this work was personal: at the outset, I had only a vague understanding of what correlation heatmaps were and why they were useful in quantitative finance. Rather than approaching the topic purely theoretically, the goal was to develop intuition through visualisation and incremental quantitative analysis. This paper documents the methodology, mathematical foundations, and insights gained from building and extending a correlation heatmap framework in Python.

1 Introduction

Correlation is one of the most frequently referenced concepts in finance, particularly in discussions around diversification, portfolio risk, and market behaviour. Despite its importance, it can feel abstract and unintuitive to newcomers, especially when introduced purely through formulas.

At the start of this project, I did not have a clear understanding of how correlation matrices or correlation heatmaps were constructed or interpreted. My primary goal was therefore not to optimise a trading strategy, but to answer a simpler question: *"What does correlation actually look like in real market data?"*

By focusing on visual tools and progressively adding structure, this project aims to bridge the gap between mathematical definition and practical intuition.

2 Data Selection and Preprocessing

The analysis focuses on a small universe of large-cap technology stocks:

$$\{\mathrm{AAPL}, \mathrm{MSFT}, \mathrm{GOOGL}, \mathrm{AMZN}, \mathrm{META}, \mathrm{NVDA}, \mathrm{AMD}\}.$$

Daily adjusted closing prices were downloaded and transformed into daily log returns. Returns are used instead of prices to avoid spurious correlations caused by long-term price trends.

The log return for asset i at time t is defined as:

$$r_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right),$$

where $P_{i,t}$ is the adjusted closing price.

3 Correlation Measurement

The relationship between two assets is measured using the Pearson correlation coefficient. For two return series X and Y, the correlation is given by:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y},$$

where σ_X and σ_Y are the standard deviations of X and Y respectively.

This value lies in the interval $[-1, 1]$:

- $\rho = 1$: perfect positive co-movement,
- $\rho = 0$: no linear relationship,
- $\rho = -1$: perfect negative co-movement.

A correlation matrix is constructed by computing $\rho_{i,j}$ for every pair of assets.

4 Heatmap Visualisation

To make the correlation matrix interpretable at a glance, it is visualised as a heatmap. Each cell represents the correlation between two assets, with colour intensity encoding the magnitude and sign of the relationship.

This visual approach was central to the learning objective of the project. Rather than focusing on exact numerical values, patterns such as clustering, redundancy, and diversification opportunities become immediately apparent.

To reduce visual clutter:

- The upper triangle of the matrix is masked, since the matrix is symmetric.
- Assets are reordered using hierarchical clustering so that similar assets appear adjacent.

5 Hierarchical Clustering

To identify natural groupings among assets, hierarchical clustering is applied to the correlation matrix. The clustering procedure groups assets based on similarity, producing a dendrogram that visually represents the structure of correlations.

Although clustering is not used for prediction in this project, it provides a powerful explanatory tool. For example, semiconductor stocks tend to cluster together, while platform-based technology firms form a separate group.

6 Time-Varying Correlations

A single correlation matrix hides the fact that correlations change over time. To address this, correlations are recomputed using rolling windows:

- 60 trading days (short-term),
- 252 trading days (long-term).

Comparing these windows highlights how asset relationships strengthen or weaken across different market regimes, reinforcing the idea that correlation is not a fixed property.

7 Quantitative Summary Metrics

To complement visual analysis, two simple quantitative summaries are computed:

- **Most and least correlated pairs**, identifying redundant and diversifying asset combinations.
- **Average absolute correlation**, defined as

  $$\bar{\rho}_{\text{abs}} = \frac{1}{N(N-1)} \sum_{i \neq j} |\rho_{i,j}|,$$

  which serves as a simple diversification score for the asset set.

Lower values of $\bar{\rho}_{\text{abs}}$ indicate greater diversification.

8 Discussion

The most important outcome of this project was not numerical precision, but conceptual clarity. By building the analysis step by step and visualising each result, correlation moved from being an abstract statistical concept to a concrete and interpretable market property.

This process mirrors how quantitative finance is often practised in reality: exploratory analysis, visual inspection, and incremental refinement precede formal modelling.

9 Conclusion

This project demonstrates how a relatively simple mathematical concept can become significantly more intuitive through careful visualisation and structured experimentation. Starting from a position of limited understanding, the goal was to develop insight rather than sophistication, and the correlation heatmap proved to be an ideal tool for this purpose.

Future extensions could include broader asset universes, alternative dependence measures, or direct integration into portfolio optimisation frameworks.