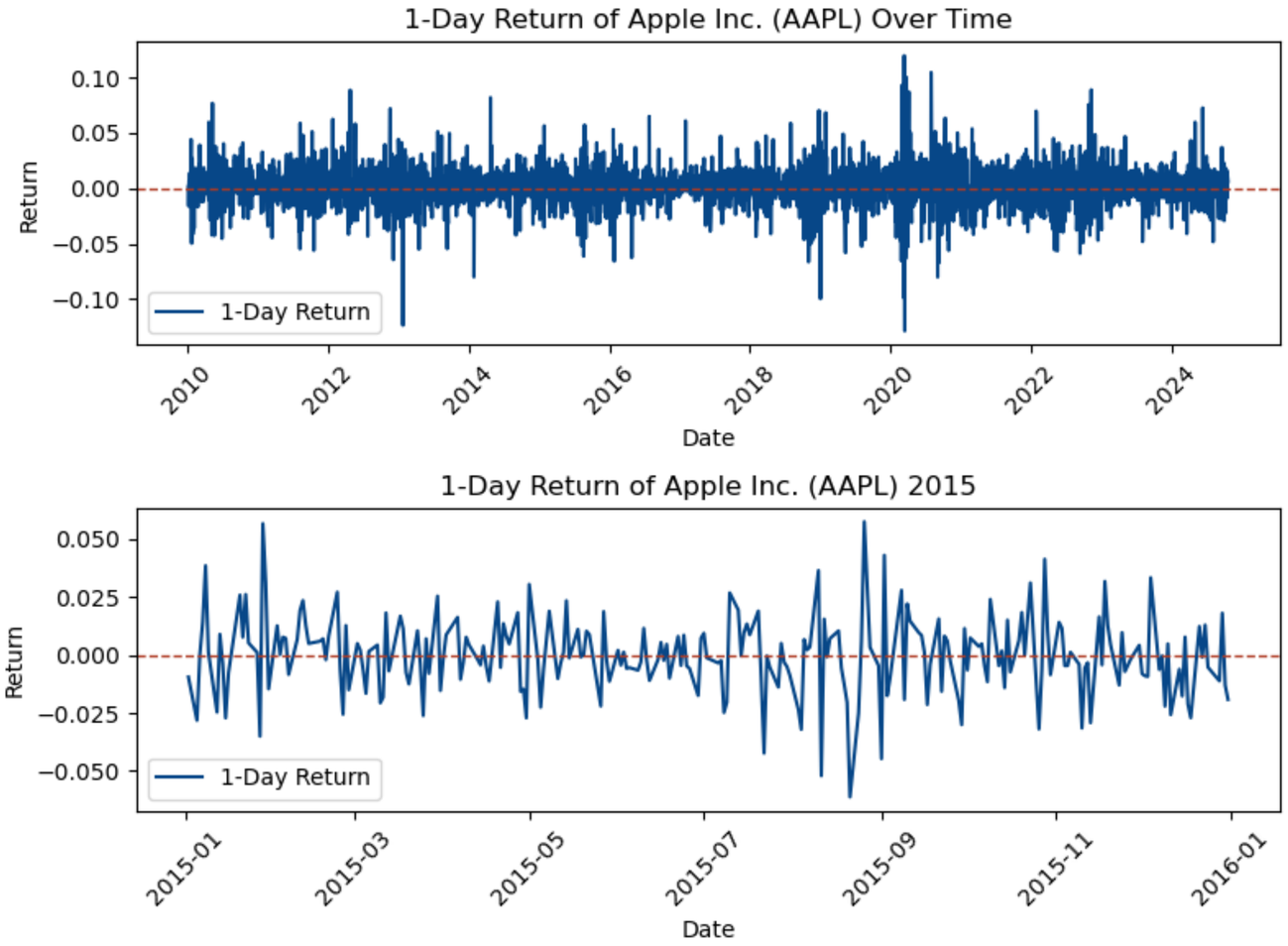
Random Matrices and the Stock Market

A preliminary EDA

Hugo C Marrochio

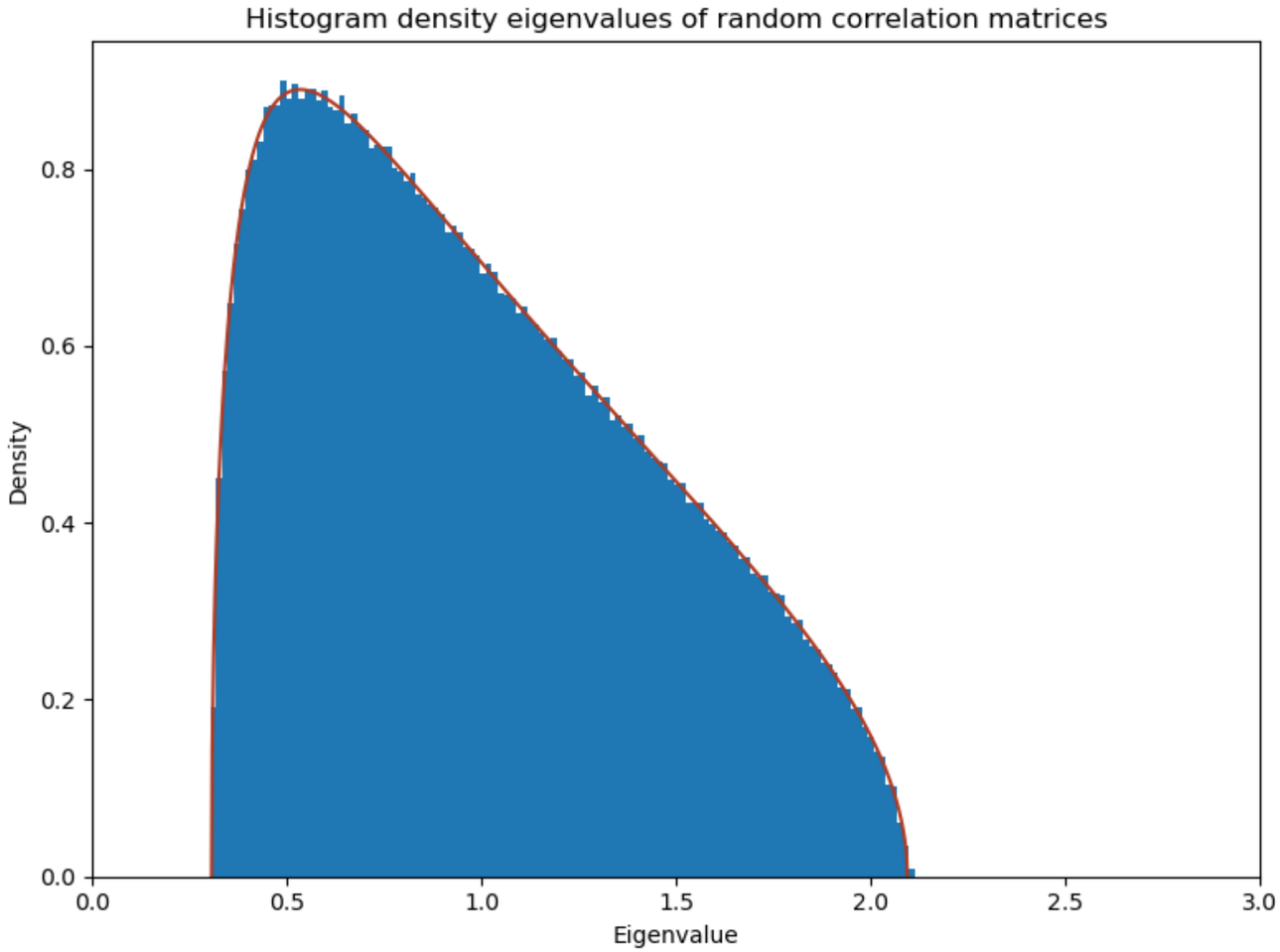
Dataset - S&P Stock Market

- S&P 500 (Standard and Poor) stock data
- https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks
- Top 503 stocks being traded is US
- Example: Daily return for Apple stock
- $$\text{Return}: \frac{P_t - P_{t-1}}{P_{t-1}}$$
- Complicated time series, hard to make predictions!



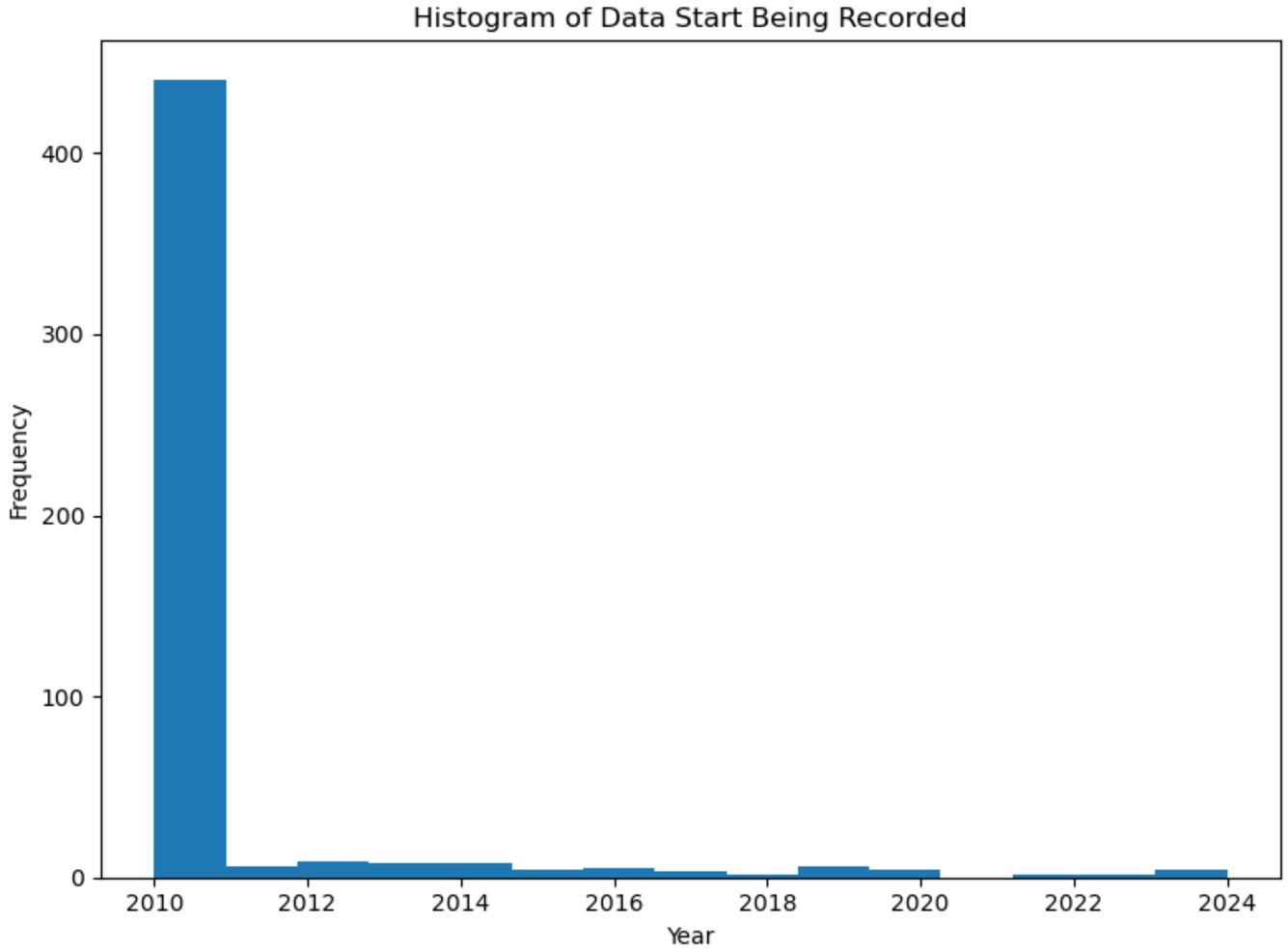
Properties of Random Matrix Theory (RMT)

- Rich subject with many applications in mathematics and physics
- A **surprising** property: In a certain limit, the eigenvalues of the correlation of random matrices follow a specific pdf (Marcenko-Pastur)
- Can use this fact to investigate ***randomness***!!!
- ***Goal:*** *"Denoise" correlation, cluster algorithms between stocks and forecast models if time permits*



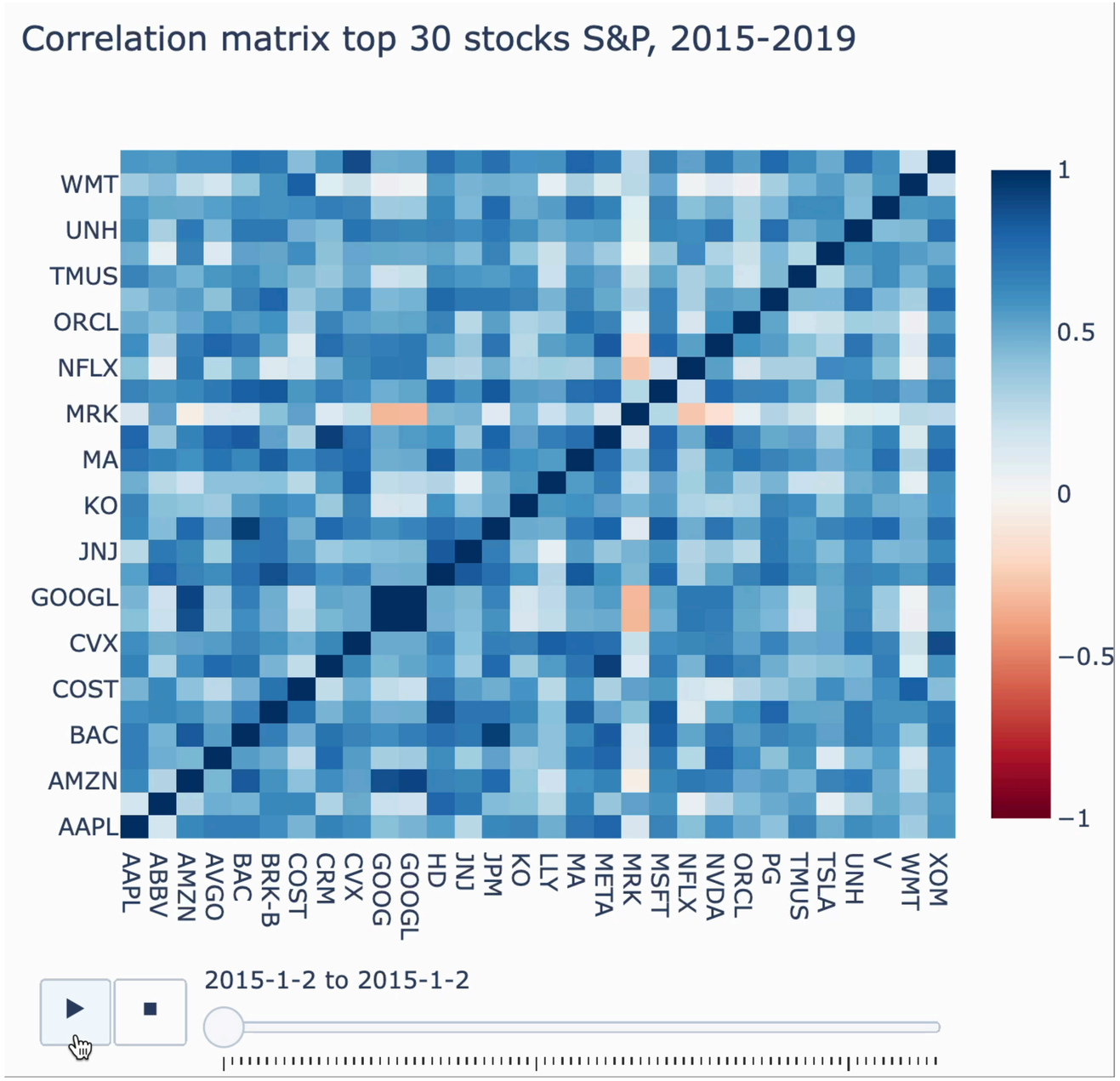
More about our data - Null Values

- Subtlety about the data: sometimes null values make sense!
- A stock is not going to have a market value before it was created!
- Since we have 503 stocks, we can always select ones (**~85% no null**)



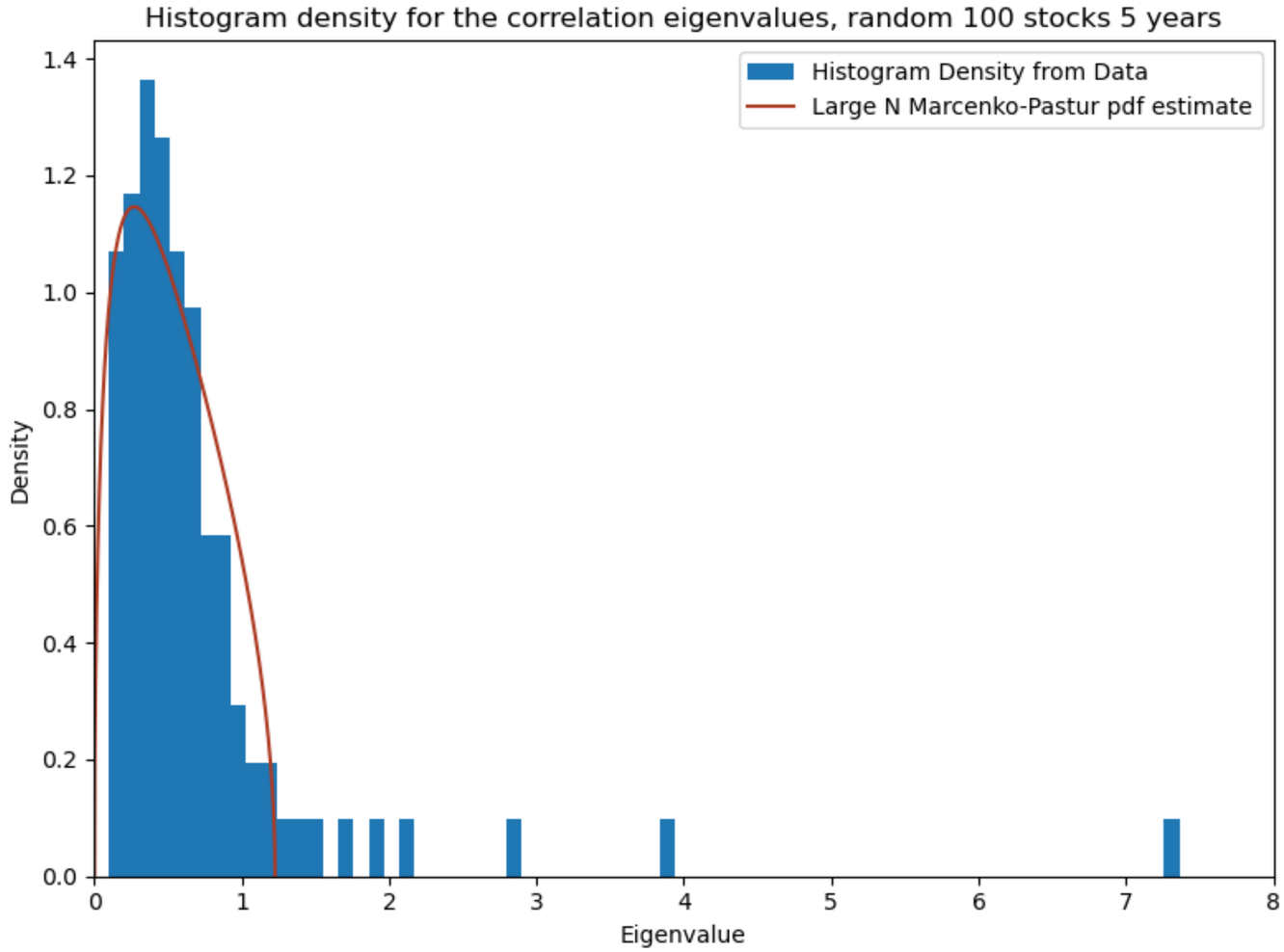
Correlation

- Our main object of study will be correlation matrices
- Difficulty with identifying trends in stock; signal and noise coexist.
- Find a good balance between averaging time window and number of stocks selected



Preliminary Eigenvalue Investigation

- Can use RMT expectation to set a range of “random” eigenvalues.
- There are some subtleties in the analysis that I plan to explore further.
- Influence of time averages, stock portfolio size, RMT fit hyper parameters.



Plan of Action

- Next step: Consolidate the time scale, investigate the optimization for the RMT fit
- Cluster algorithms taking into consideration “denoising” from the “random” eigenvalues
- If time permits, investigate simple AI models to forecast with the cluster information
- Reinforcement learning?