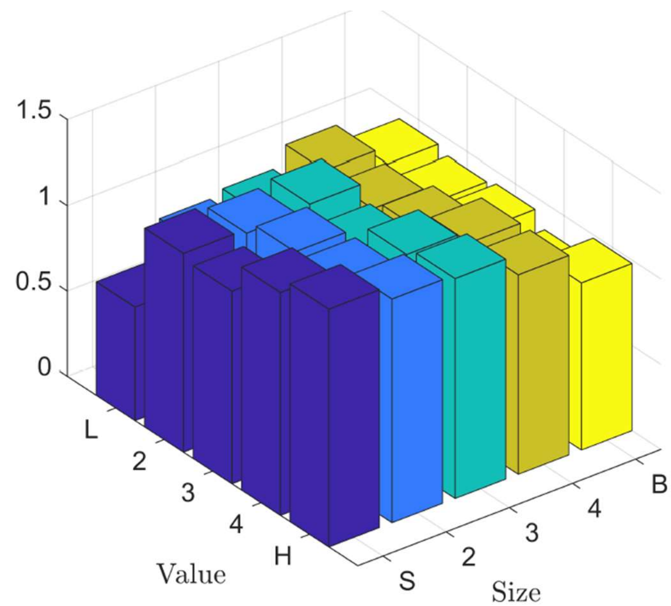
aExercise 3

1. We are going to be using the same data as last week. So we want the 25 portfolios loaded into a Pandas DataFrame again.
2. First filter the data so that the series begin in July, 1963. Then plot the raw returns for Small HiBM (Small Value Firms) and for Big LoBM (Big Growth Firms). Make sure to label your figure so someone can understand what it's showing even if given no context other than the figure itself. Do not just use variable names but use something more descriptive. If you don't know what the variables mean, do some research!
3. Find the means and variance/covariances for ALL 25 portfolios. Try to do this without using built in functions for practice. Plot the means and variance/covariances in separate 3 dimensional bar charts. One axis should be Value (loBM BM2...HiBM) and the other axis should be Size (Small, ME2, ... Big) and the height of the bar should indicate the value. Again, pay attention to labeling. Your result should look something like below:



   Which portfolios earn the lowest and highest returns? Which covary the most? Which have the highest variance?

4. The expected return for different assets is a fundamental economic object that we would like to measure in practice. One way to do this is what we've already done: found, in this case, the average <u>monthly</u> return. We are implicitly thinking about returns in the following way:

$$r_{it} = E(r_{it}) + u_{it}$$

   where r is the realized return, $E(\,)$ is the expectation operator, and u is the error, and the subscripts it mean for asset i at time t. We took the average, hoping that the error terms canceled out for all the data. But it could be difficult to tell the difference between two assets based only on this. For example, a small firm and a large firm might have different returns, but because a month is a relatively short period, the difference in average return might be small and hard to pick up, statistically speaking. If we think about cumulating monthly returns to annual returns, then the differences between average annual returns should be larger and we might be more likely to pick up these differences.

With this in mind, compound the monthly returns into annual returns. (Hint: You might consider doing this as a function, where you can change the number of total years of compounding arbitrarily). Each subsequent year should overlap for 11 months with the prior year (meaning the first return will be July to June, then August to July, then September to August, etc). We would say these are annual returns at a monthly frequency. Now plot the returns for Small HiBM and Big LoBM you did like in part 2.

5. Calculate the average returns from this series of annual returns for ALL 25 portfolios. Plot the result like you did in part 3. **Do you get the same portfolios generating the highest returns?**
6. Repeat the exercise from 4. For a 3 year cumulation period. Plot the result for the same two portfolios, and also plot the *outperformance* of Small LoBM
$$outperformance_t = r_{Small-Valu\ \ ,t} - r_{Large-Growt\ \ ,t}$$
Make sure both plots are in the same figure and that their x-axes line up for direct comparison. Often this happens by default, but double check. **Describe what happens to Small Value stocks during the recessions of 1973, 1993, 2000, 2008, and 2020.** Notice that these longer term returns are measuring the business cycle.
7. Now lets look at the actual distribution of returns instead of just a summary statistic (the mean/average). Create a *histogram* of the returns, normalizing by frequency, for the Small Value portfolio. Your histogram should have 10 bins, and you should do this for the monthly returns, the annual returns, and the 3-year compounded returns. As ALWAYS when making a figure, make sure it is clearly labeled. **What happens to the returns as we aggregate over longer horizons?**

If you don't know what a histogram is, *look it up!* As a brief explanation, a histogram is where you assign the observations (returns) in this case to "bins" and count the number of observations in each bin. If you had returns of $\{-0.02, -0.01, 0.2\}$ and your bins were the intervals $(-0.03,0)$ and $(0,0.3)$ your first bin would have 2 observations and your second bin would have 1 observation. If you *normalize* by the number of observations to get a frequency plot (which is analogous to an empirical probability density function) your first bin would be $\frac{2}{3} \approx 0.6667$ and your second bin would be $\frac{1}{3} \approx 0.3333$.

8. Now we want a better understanding of the joint density of Small Value and Big Growth. Essentially we want to create a 3 dimensional histogram where the x and y axes are the bins for each portfolio's return, and the z axis is the count where the portfolios' returns fall into that bin. Ie maybe the x-axis is for Small Growth, and has bins $(-0.03,0)$ and $(0,0.3)$, and the y axis is for Big Growth and as bins $(-0.06,-0.03)$, $(-0,0.3,0)(0,0.03)$ and $(0.03,0.6)$. Then if $r_{SV,t} = -0.02$ and $r_{BG,t} = 0.04$ there would be a count of 1 in the cell corresponding to $(-0,0.3,0)$ on the x axis and $(0.03,0.6)$ on the y axis. **Look at the result: do the two portfolios provide a hedge for each other? Meaning does one appear to provide high returns when the other has low returns?**

9. The last thing we want to investigate is if the returns follow a normal distribution. While we've visualized the empirical distribution of returns, we can be more formal about comparing with a normal distribution. First, we want to *standardize* the empirical distribution of returns for *all* portfolios. Recall, this means we are going to subtract the mean and divide by the standard deviation. Calculate the mean, variance, skewness and kurtosis for this standardized distribution. **Compare these estimates to the values for a standard normal distribution! Explain.**

10. Create and plot the standard normal pdf on the interval (-4,4). You can do this with the formula from the slides or with a built in Python function. Now add create a 2x2 array of subplots, where in the top row you are working with Small Value and in the bottom Big Growth. In the first column, plot the standard normal pdf on the interval $(-4,4)$ AND plot a histogram of monthly returns with 120 bins (on the same axes!). In the second column do the same, but with annual returns and 10 bins. **What do you notice about the histogram vs. pdf? What can you conclude about the normality of returns from this?**

11. Lastly, as additional evidence about the normality of returns, we are going to create what's called a QQ-plot (or sometimes a normal probability plot). This is where you plot the *standardized* return value calculated in 9. on the x axis and the percentile of that return in the sample on the y-axis (if you have 100 returns, and you order them from smallest to largest, the 10th smallest would be the 10th percentile, the 25 smallest would be the 25th percentile, etc.). On this plot, you should also include a

percentile plot of the standard normal distribution. Do this for the same 2x2 array from 10. (top row Small Value, bottom row Big Growth, first column monthly returns, second column annual returns). **Does the QQ plot support or refute your previous conclusion? Is there a difference for annual vs. monthly returns?**