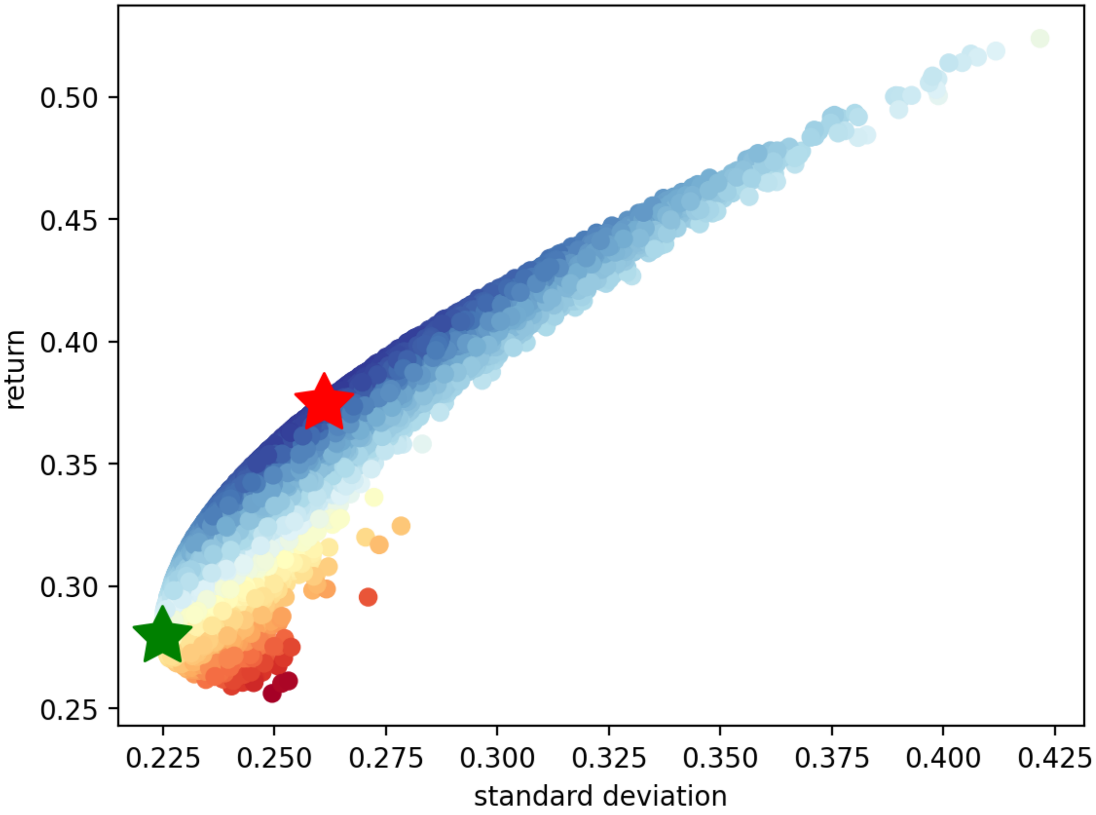
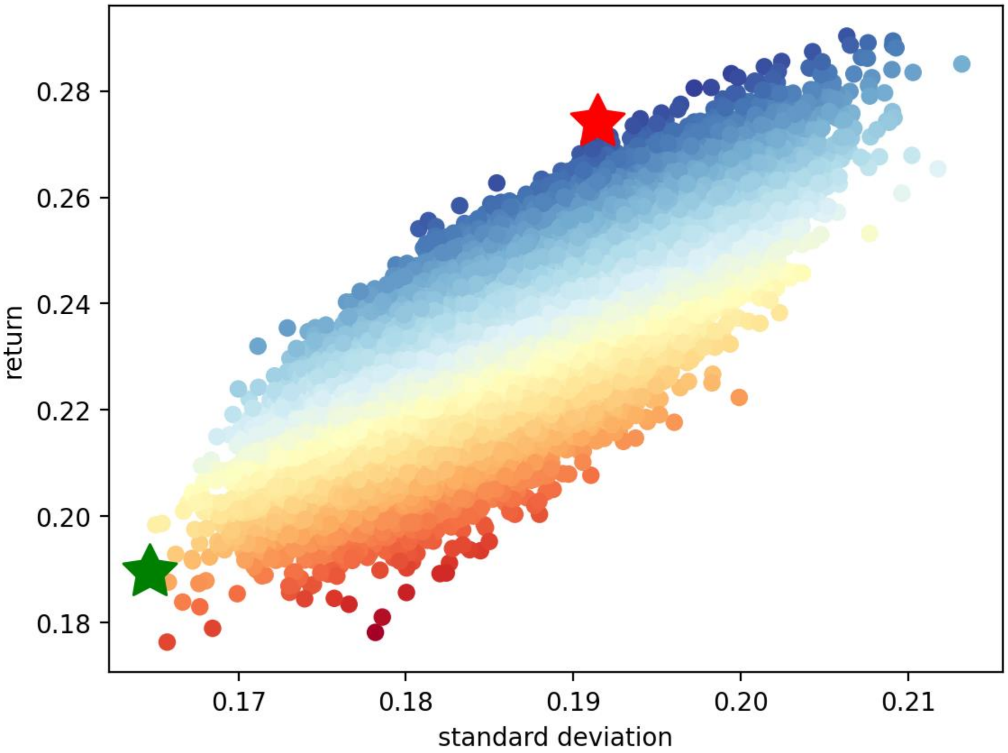
This experiment is built upon the Efficient Frontier app (link here). Although the app works as accurate testing tool the experimentation was conducted using this code base (link here). The only difference between the codebase and the app is that the codebase for the app is outfitted with the proper API to work with web-facing GUI. Both codebases use the same calculation and functions.

The efficient frontier tests use 10Y of historical adjusted close data pulled from yahoo finance. They also use the mean of daily percentage changes in adjusted close for the returns method and covariance of those returns for the risk method. Each test uses 100,000 simulations which are really 100,000 randomly generated weights and applies the mean and covariance to them to make the plots.

My initial interest in this topic was when I was trying to implement the efficient frontier to a series of stocks. Normal efficient frontiers look like this, which is an efficient frontier ran with around 7 stocks.



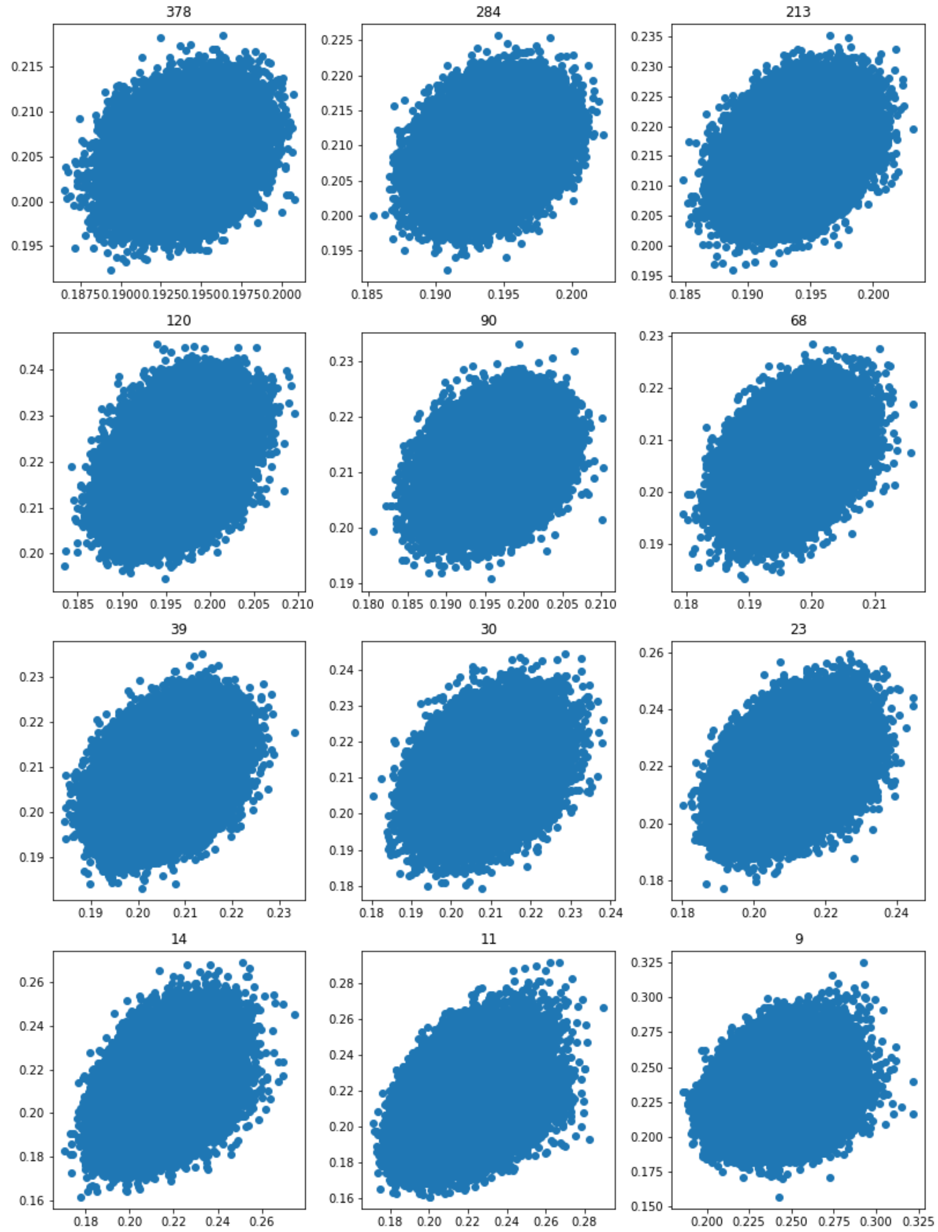
But when I tried to make an efficient frontier with around 30 stocks, I would get a graph that looks like this.



At first, I was confused and though it had something to do with the underlying securities. But after some experimentation I had a feeling, it had something to do with the number of stocks used. I started writing a program that would randomly drop stocks and then run the efficient frontier. The way that the program works is that picks several stocks and then drops a random amount (usually 25%). It then successively does that over and over, and never resamples therefore the smallest number of stocks efficient frontier has the same stocks as the largest.

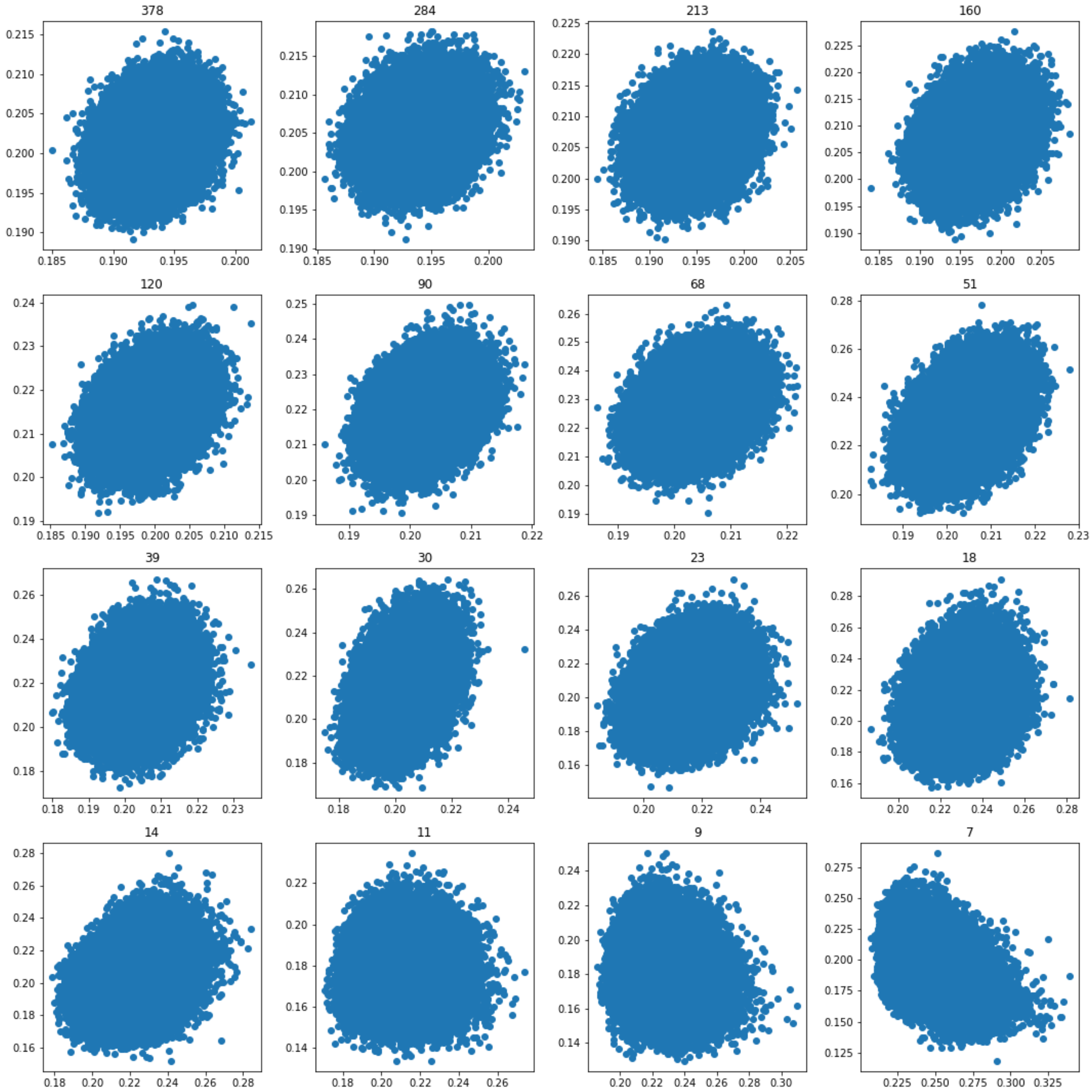
I ran about 10 tests (link here). I have put 3 tests below. The title of each plot is the number of stocks used. The program also keeps track of all of the dataframes used to make the plots but for data reasons I wrote another program that keeps track of the stocks used in each test. All tests all have an associated timestamp with them.

Timestamp: 1620411524.873581



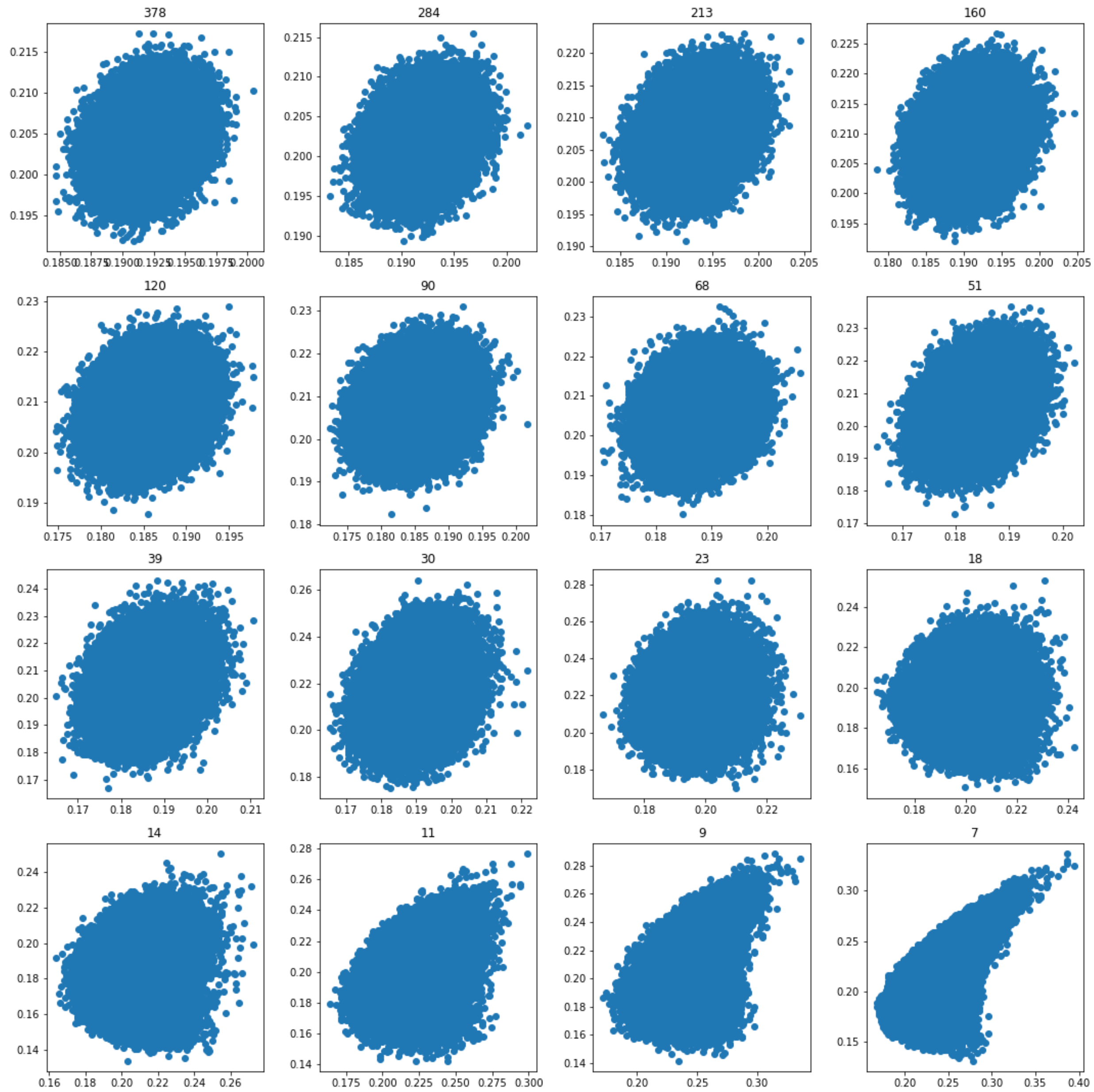
Information found in this document
does not constitute as investment advice

Timestamp: 1620400255.83549



Information found in this document
does not constitute as investment advice

Timestamp: 1620376163.978867



It seems that as we drop the number of stocks used, we drop the number of outlier points and the efficient frontier starts to take shape. As of right now it seems evident that the number of stocks used determines the shape. But there are many more questions that I would like to investigate.

Information found in this document
does not constitute as investment advice

Future Questions & research topics:

1. Mathematical explanation for this phenomenon
2. What outlier reduction / anomaly detection methods can we use to get rid of the outliers
    a. When outliers are removed is the mean-markowitz and minimum-variance portfolio still "valid"
    b. Does the outlier reduction method work consistently across different groups of stocks and returns and risk methods?
3. Does the number of simulations play a role in getting outliers?
4. Does the proportion of outlier stay the same or does it increase / decrease as we drop the number of stocks?
5. Do different returns and risk methods yield different results?
    a. Log returns
    b. Cholesky decomposition
6. What is an acceptable level to consider a point as an outlier.

Diego Alvarez
Diego.alvarez@colorado.edu
LinkedIn | GitHub | Medium

Information found in this document
does not constitute as investment advice