Portfolio_optimization

Nguyen Trung Duc

#1.Installing the packages

For the construction of this report in portfolio optimization, we will be using several packages for the following reasons:

quantmod

- **Purpose:** Used for quantitative financial modeling and analysis.

rvest

- **Purpose:** Primarily used for web scraping in R.

TTR (Technical Trading Rules)

- **Purpose:** Designed for technical analysis and building trading rules.

dplyr

- **Purpose:* data manipulation.

GA (Genetic Algorithms)

- **Purpose:** Used for optimization problems with genetic algorithms.

#2.Download the data

##2.1.Download the list of S&P100 ticker

We will be using stocks data from the S&P100 basket, the main reason for this choice is because:

1. **Liquidity**: Stocks included in the S&P 100 index are mostly regarded as blue chips, meaning they have significant trading volume. High liquidity is essential for efficient portfolio management.
2. **Diversification**: S&P 100 is a diversified set of large-cap stocks across various sectors, this help reduce risk through diversification, spreading risk across different sectors.
3. **Stability**: S&P 100 companies are mostly well-established, large-cap companies with stable performance.
4. **Information Availability**: These companies are extensively covered by analysts and financial media.

5. **Risk Management**: As large-cap stocks, S&P 100 tend to be more stable and less volatile than smaller-cap stocks.

To get access to the list of current stocks in the index, we based our selection on the wikipedia page of S&P100

##2.2.Crawling data of the selected stocks

As time-series analysis usually based on multiple assumptions, including absence of outliers, so we would avoid using data between the period of 2020 until now (02/2024) as there are many generational events (Covid-19, Russia invasion of Ukraine), and for this we will be working with data between 2016 to end of 2019.

##2.3.Splitting the data into train and set

As there may exist inconsistent between the current stocks in S&P100 and the time period chosen, so we will only be working with stocks that exist through the period. As for splitting the training and testing dataset, we will use the data within the year 2019 as test set and 3 years from 2016 to 2018 as training set

#3.Portfolio weight optimization

##3.1. Create a random combination base on sector percentage

So for a benchmark, we will create a portfolio that take stocks from each sectors based on the percentage of that sector present on the S&P 100, and from t

```
## # A tibble: 11 x 2
##    Sector                 Count
##    <chr>                  <int>
##  1 Communication Services    10
##  2 Consumer Discretionary    11
##  3 Consumer Staples          10
##  4 Energy                     3
##  5 Financials                15
##  6 Health Care               14
##  7 Industrials               13
##  8 Information Technology    17
##  9 Materials                  2
## 10 Real Estate                2
## 11 Utilities                  4
```

```
##  [1] "CHTR" "HD"   "CL"   "GS"   "JPM"  "BMY"  "CAT"  "V"    "TXN"  "NEE"
```

And each stocks will have a balance weight of 10% witin the portfolio

##3.2.Portfolio weight balancing

###3.2.1.Function for portfolio weight calculation

To incorporate both the risk and return factor into the calculation, we will use Sharpe ratio for the fitness function

$$S = \frac{R_p - R_f}{\sigma_p}$$

Where: - S is the Sharpe ratio
- R_p is the average return of the investment or portfolio
- R_f is the risk-free rate
- σ_p is the standard deviation of the investment's returns

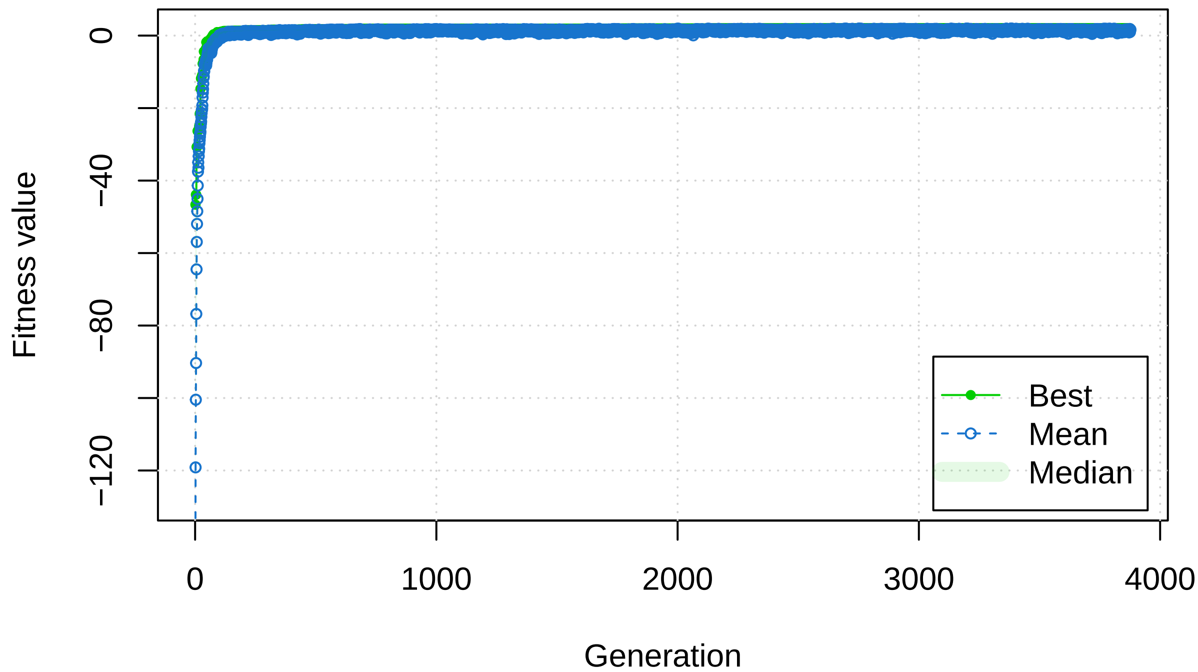
Understanding: - Sharpe ratio > 1: The investment is generating excess returns relative to its risk. This is generally considered good.
- 1 >= Sharpe ratio >= 0: The investment's returns are acceptable, but may not sufficiently compensating for the risk taken.
- Sharpe ratio < 0 : The investment is not providing adequate returns given the level of risk.

In theory, investors would want to maximize Sharpe ratio by seeking better investment or portfolios. However this should not be the sole purpose and other factors also need to be taken into consideration, including, investment objectives, time horizon, and risk tolerance.

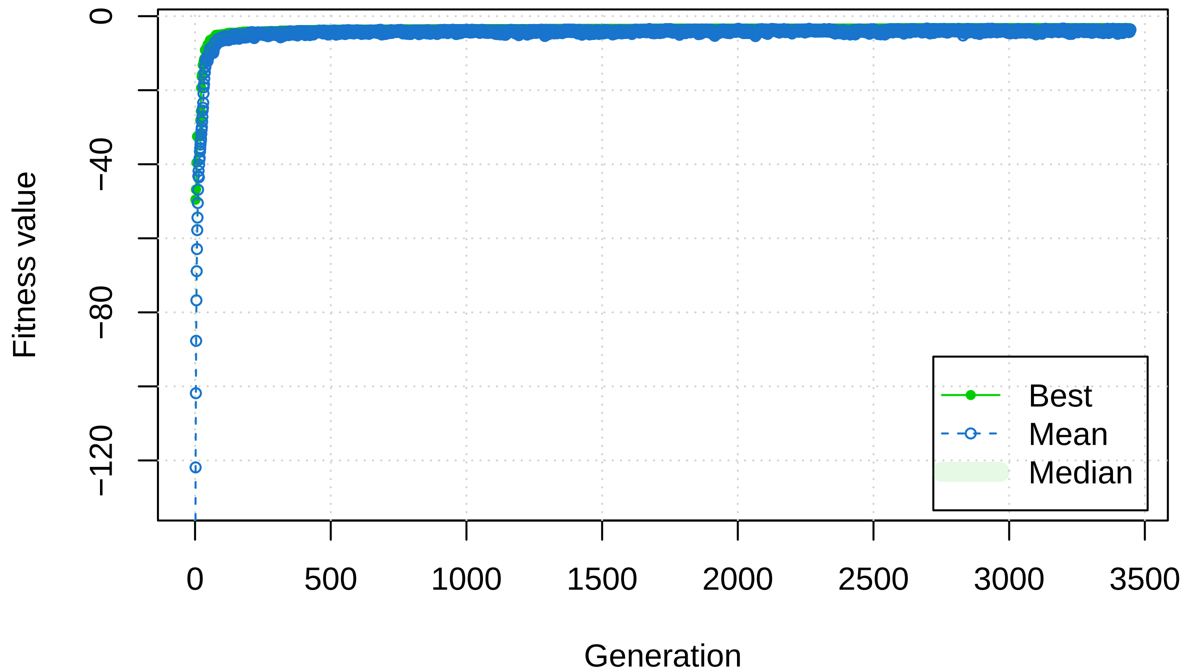
###3.2.2. Optimising the weight within the given portfolio

With the current selected portfolio, we will be apply 3 different approach, first is to maximise the Sharpe ratio (without including the risk free rate as this information is limited access by yahoo), second is to maximise the return, and third is to minimise the risk of the portfolio

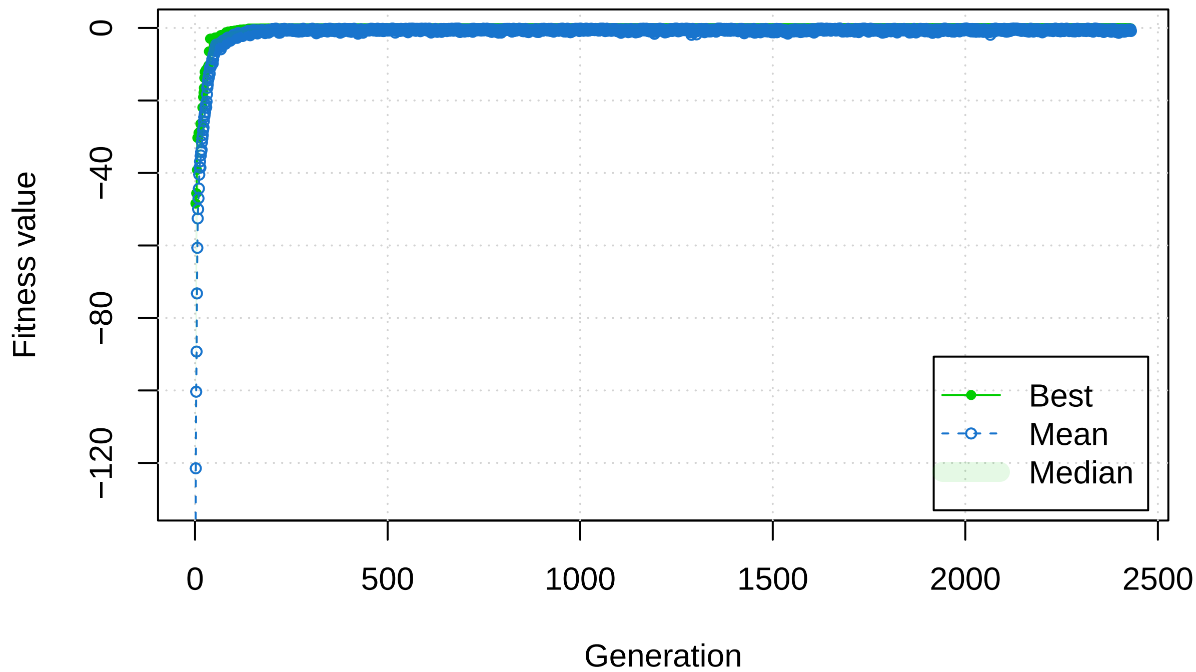
```
## -- Genetic Algorithm -------------------
##
## GA settings:
## Type                  =  real-valued
## Population size       =  50
## Number of generations =  50000
## Elitism               =  2
## Crossover probability =  0.8
## Mutation probability  =  0.1
## Search domain =
##       x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
## lower  0  0  0  0  0  0  0  0  0   0
## upper  1  1  1  1  1  1  1  1  1   1
##
## GA results:
## Iterations             = 3877
## Fitness function value = 1.995998
## Solution =
##              x1          x2          x3          x4        x5          x6
## [1,] 0.06125565 0.005768865 0.004038602 0.004481375 0.1027622 0.008703083
##             x7         x8        x9       x10
## [1,] 0.1275955 0.07137497 0.1044578 0.5152653
```




```
## -- Genetic Algorithm -------------------
##
## GA settings:
## Type                  =  real-valued
## Population size       =  50
## Number of generations =  50000
## Elitism               =  2
## Crossover probability =  0.8
## Mutation probability  =  0.1
## Search domain =
##       x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
## lower  0  0  0  0  0  0  0  0  0   0
## upper  1  1  1  1  1  1  1  1  1   1
##
## GA results:
## Iterations             = 3447
## Fitness function value = -3.299677
## Solution =
##              x1         x2         x3         x4         x5          x6
## [1,] 0.08129822 0.01651922 0.01396915 0.01367503 0.03249189 0.007424146
##             x7         x8       x9        x10
## [1,] 0.6251925 0.02501331 0.272756 0.05808906
```




```
## -- Genetic Algorithm -------------------
##
## GA settings:
## Type                  =  real-valued
## Population size       =  50
## Number of generations =  50000
## Elitism               =  2
## Crossover probability =  0.8
## Mutation probability  =  0.1
## Search domain =
##       x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
## lower  0  0  0  0  0  0  0  0  0   0
## upper  1  1  1  1  1  1  1  1  1   1
##
## GA results:
## Iterations             = 2430
## Fitness function value = -0.1111342
## Solution =
##             x1        x2        x3         x4         x5         x6         x7
## [1,] 0.0383705 0.1373935 0.1672483 0.03869013 0.09332669 0.08738577 0.02545126
##              x8        x9       x10
## [1,] 0.01269747 0.0467016 0.3474057
```



#4.Choosing optimal portfolio

##4.1. Function for finding optimal stock combination

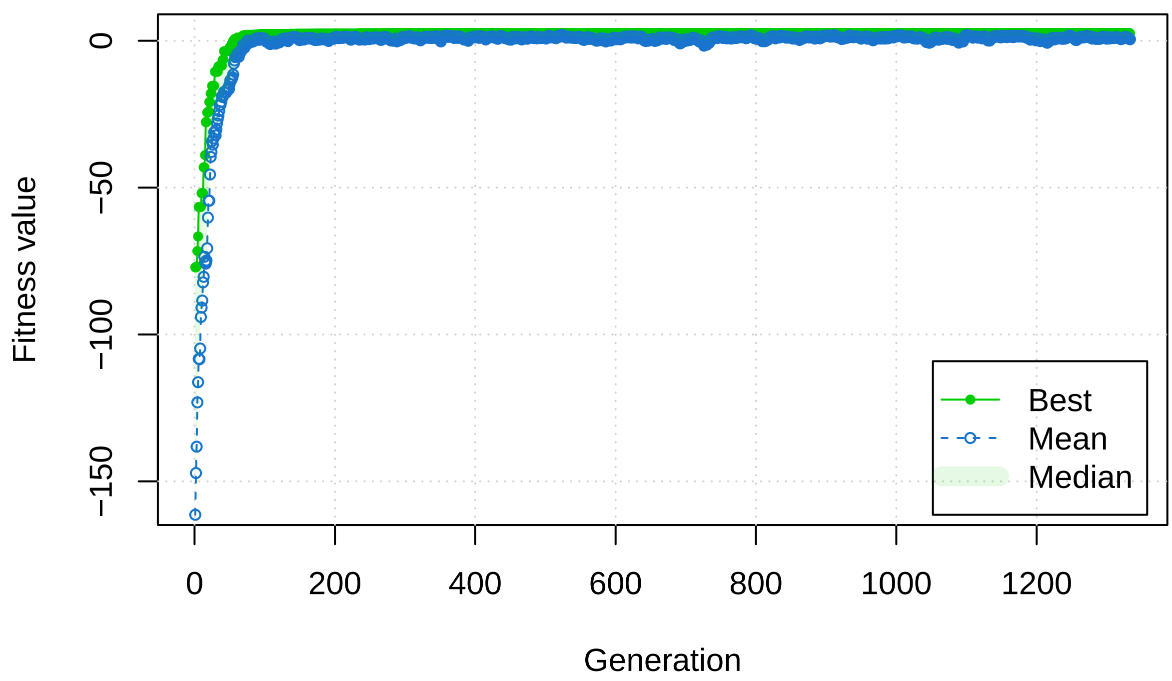
To find an optimal combination of stock, we will use the GA function and switch to type = "binary". In this method, we will continuously create binary lists of values the length of all stocks available on the S&P 100. To limit the choice to just 10 stocks, we divide the binary values by number of stock we want (in this case 10), and use the same sum = 100% as we were using on the weight optimisation.

As we have a very large number of combination to test out so we will be using a high level of mutation and elitism to increase our chance of avoiding any local maxima.

##4.2. Finding the optimal combination

```
## -- Genetic Algorithm -------------------
##
## GA settings:
## Type                  =  binary
## Population size       =  80
## Number of generations =  50000
## Elitism               =  5
## Crossover probability =  0.5
## Mutation probability  =  0.5
##
## GA results:
## Iterations             = 1333
## Fitness function value = 2.579568
## Solution =
```

```
##      x1 x2 x3 x4 x5 x6 x7 x8 x9 x10  ...  x99 x100
## [1,]  0  0  0  0  0  0  1  0  1   0         1    0
```



##4.3. Optimising the portfolio weight

###4.3.1. Creating a subset of the chosen stocks

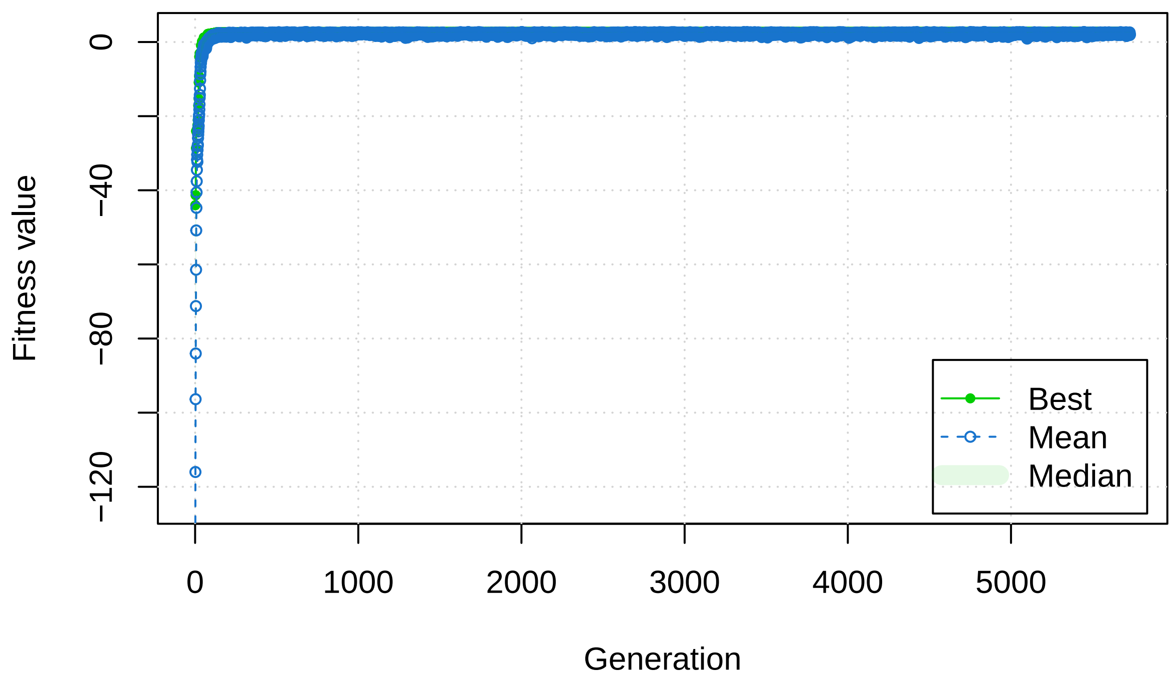
```
##     Ticker.symbol       Company.name                 Sector
## 7             AMD                AMD Information Technology
## 9             AMT     American Tower            Real Estate
## 13             BA             Boeing            Industrials
## 32             DE     Deere & Company           Industrials
## 38            EXC             Exelon              Utilities
## 61            MCD         McDonald's Consumer Discretionary
## 71            NEE      NextEra Energy             Utilities
## 74           NVDA             Nvidia Information Technology
## 93            UNH UnitedHealth Group            Health Care
## 100           WMT            Walmart       Consumer Staples
```

The selected portfolio visually show a diverse portfolio between high yield sector and stable stocks so we can expect diversification within the portfolio.

###4.3.2. Optimise portfolio weight

```
## -- Genetic Algorithm -------------------
##
## GA settings:
```

```
## Type                     =  real-valued
## Population size          =  50
## Number of generations =  50000
## Elitism                  =  2
## Crossover probability =  0.8
## Mutation probability  =  0.1
## Search domain =
##       x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
## lower  0  0  0  0  0  0  0  0  0   0
## upper  1  1  1  1  1  1  1  1  1   1
##
## GA results:
## Iterations             = 5729
## Fitness function value = 2.716358
## Solution =
##              x1          x2         x3         x4         x5         x6
## [1,] 0.06990665 0.001269683 0.09267367 0.05288671 0.07137217 0.05160937
##             x7         x8        x9       x10
## [1,] 0.2836458 0.07938342 0.1903466 0.1300773
```



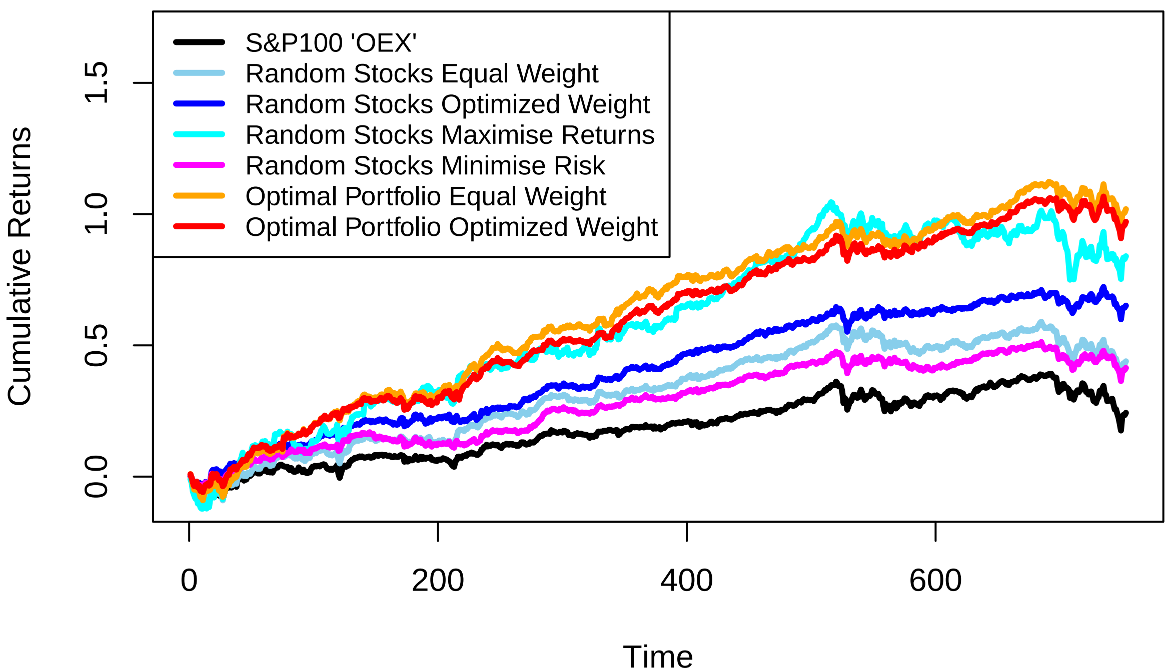
#5. Result
##5.1. On the train set ### Checking the result on the train set

We examinate the Sharpe ratio, risk and return of the portfolios on the traning dataset between 2016 and 2018

```
##                                           Name     Return       Risk Sharpe_ratio
```

```
## 1         Random Stocks Equal Weight - train 0.1582683 0.1382903     1.144465
## 2     Random Stocks Optimized Weight - train 0.2434135 0.1219309     1.996323
## 3     Random Stocks Maximise Returns - train 0.3241214 0.2305103     1.406104
## 4        Random Stocks Minimise Risk - train 0.1485681 0.1108502     1.340261
## 5     Optimal Portfolio Equal Weight - train 0.4059334 0.1573649     2.579568
## 6 Optimal Portfolio Optimized Weight - train 0.3835379 0.1409171     2.721727
```

###Graphing the cummulative returns of the portfolios



First of all, we can see all the portfolios outperform the S&P 100 index. On the training set we see that all of the objective functions perform well within expectation, and showing the expected result where applying optimisation on the weight of a portfolio improve the sharpe ratio result, and an optimal picked portfolio of 10 is better than a randomly picked porftolio.

The results for the maximise return and minimise risk portfolios also show the clear linear relationship, with the max return also result in a high risk, and a low risk result in low return

##5.2. On the test set

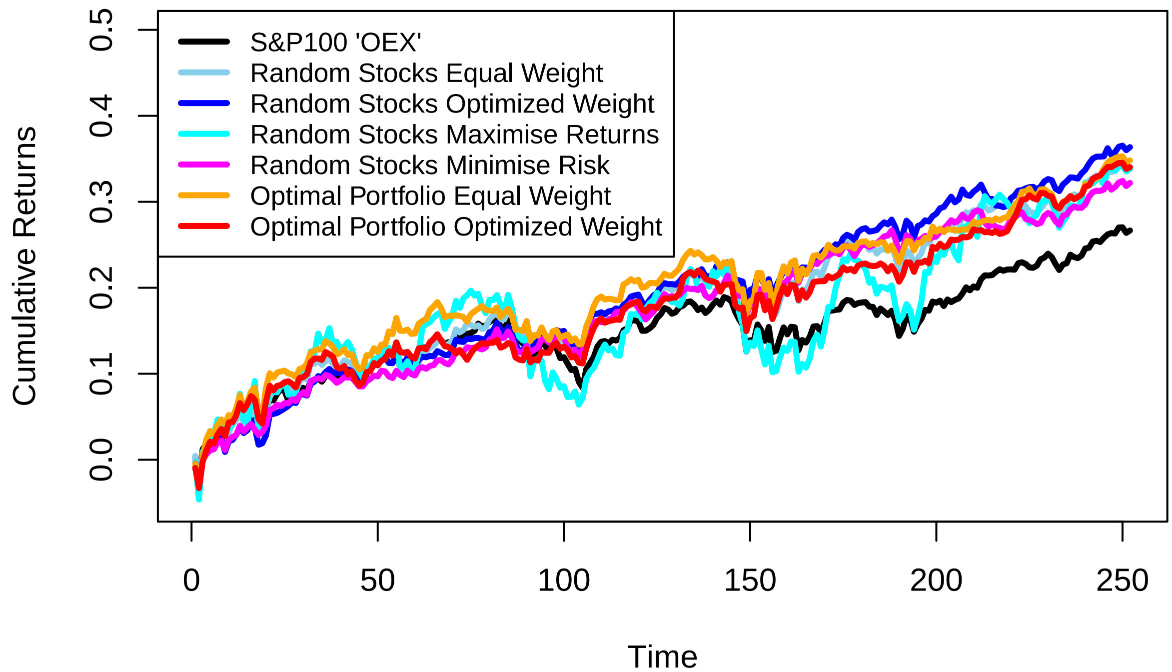
Checking the result on the test set

We examine the Sharpe ratio, risk and return of the portfolios on the test dataset in the year 2019

```
##                                          Name    Return      Risk Sharpe_ratio
## 1         Random Stocks Equal Weight - test 0.4046279 0.1275456     3.172418
## 2     Random Stocks Optimized Weight - test 0.4380209 0.1081283     4.050937
## 3     Random Stocks Maximise Returns - test 0.4021619 0.2256729     1.782056
```

```
## 4         Random Stocks Minimise Risk - test 0.3795066 0.1032384     3.676021
## 5     Optimal Portfolio Equal Weight - test 0.4161061 0.1394532     2.983840
## 6 Optimal Portfolio Optimized Weight - test 0.4050185 0.1267193     3.196186
```

###Graphing the cummulative returns of the portfolios



Once again, we see all the portfolio outperform the S&P100 index. On the testing set, the result of the risk minimised function also show a similarly low risk, and while the weight optimisation still maintain a good improvement compare to balanced weights, the previously optimal chosen portfolios perform worse, this may indicate the portfolio has not been as diversified as we would like, or maybe a group of 10 is not the optimal number to diversify in the market.