

# DEMOS - MEETING 03/11

Thi Truong

2023-03-11

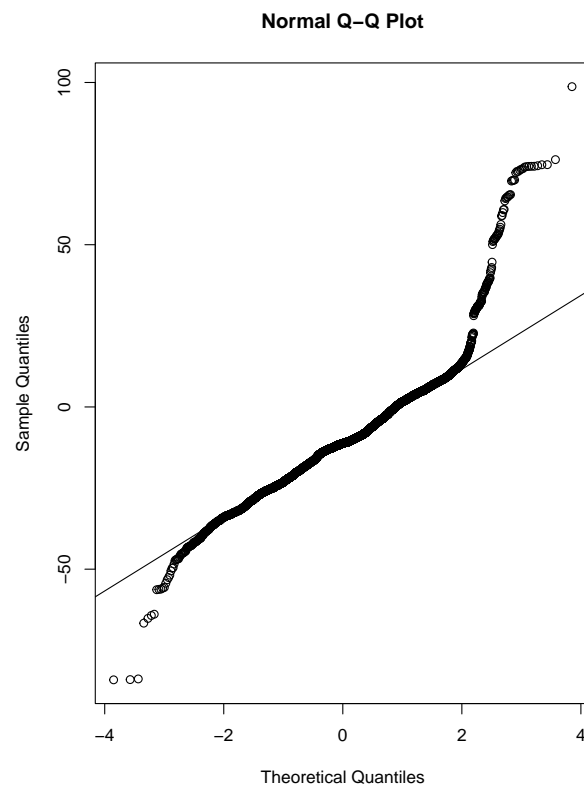
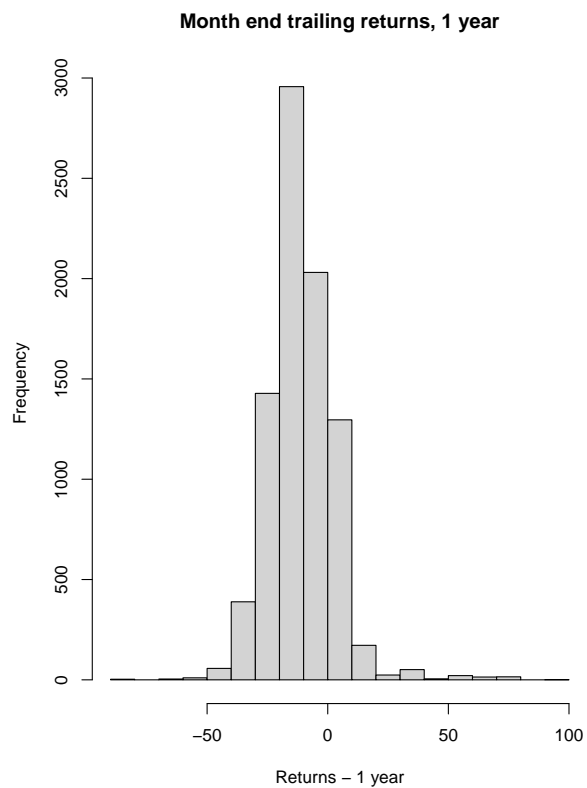
## Some information about the data

This demo used and analyzed from the *Invest+Your+Values+shareclass+results+20221216.xlsx* data in Google Drive's folder/Data

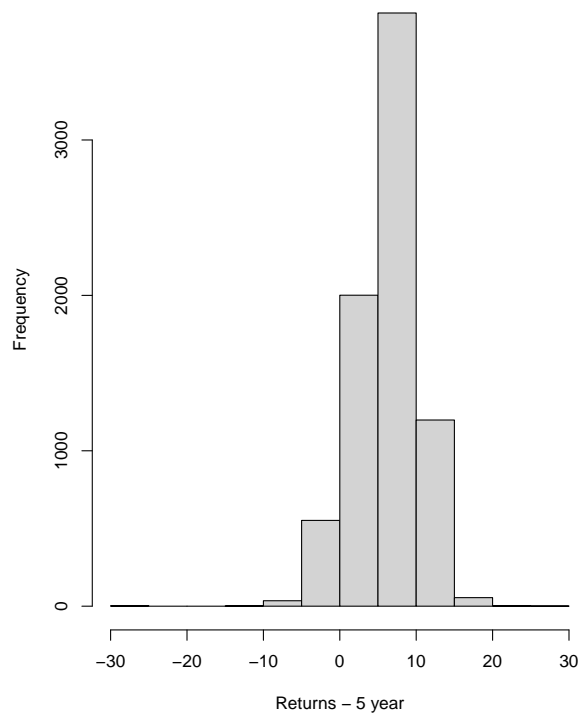
- The **Invest Your Value** data is based on updates to mutual funds holding, which is a large data with 8665 rows and 137 columns indicating factors of ESG (or ESG scores) and returns
- ESG factors are considered in the data includes: Fossil Free Funds, Deforestation Free Funds, Gender Equality Funds, Gun Free Funds, Prison Free Funds, Weapon Free Funds, Tobacco Free Funds.
  - Within each funds: Variables are evaluated including grade, weight, asset, and count.
    - \* Grade: Based on grade of each particular fund - A, B, C, D, and F
    - \* Weight: The percent of fund assets invested
    - \* Asset: The amount in US of fund assets invested
    - \* Count: The number of holding in a fund's portfolio
- The data consisted of different period of time for returns such as 1 month, 3 months, 6 months, 1 year, 3 years, 5 years, 10 years, 15 years, 20 years, and since inception.

# 1. Hypothesis: Returns are normally distributed

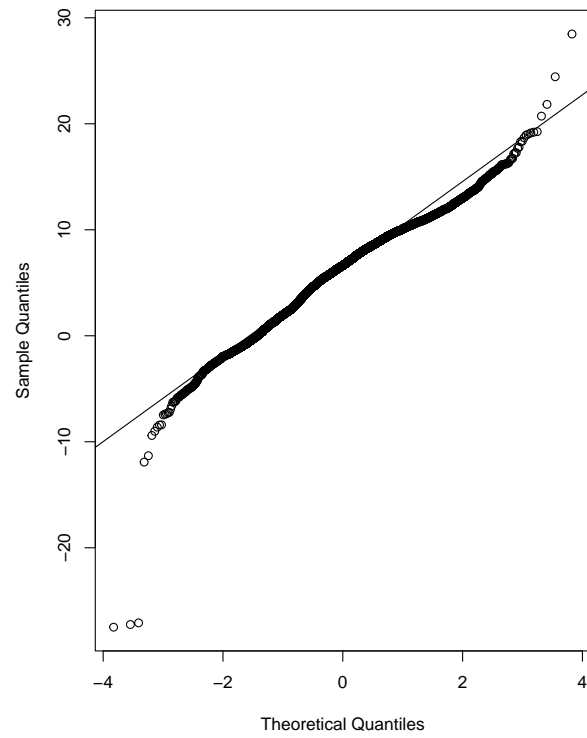
## (i). Visual by histogram



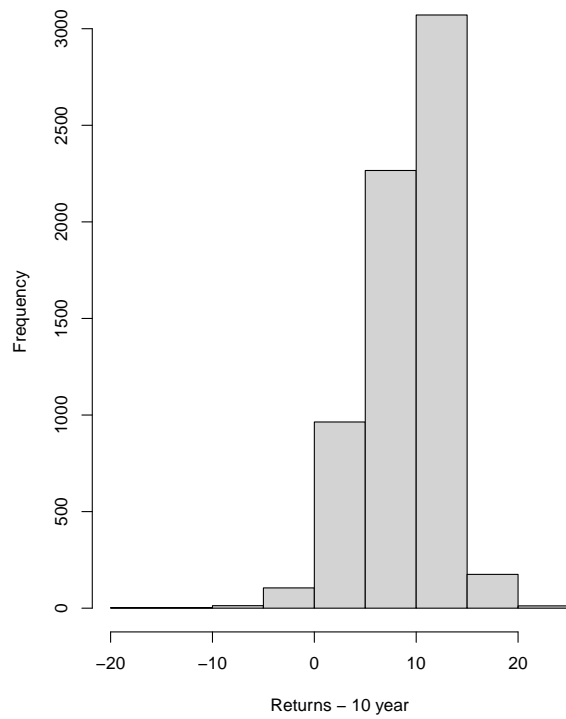
Month end trailing returns, 5 year



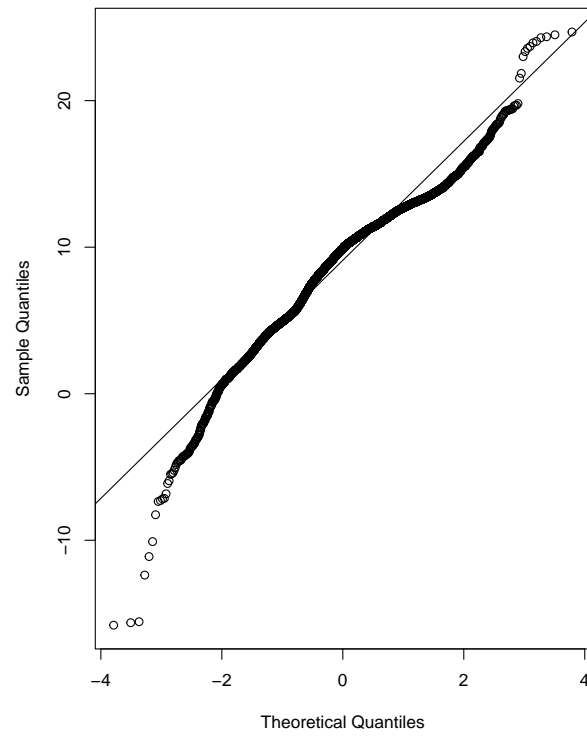
Normal Q-Q Plot

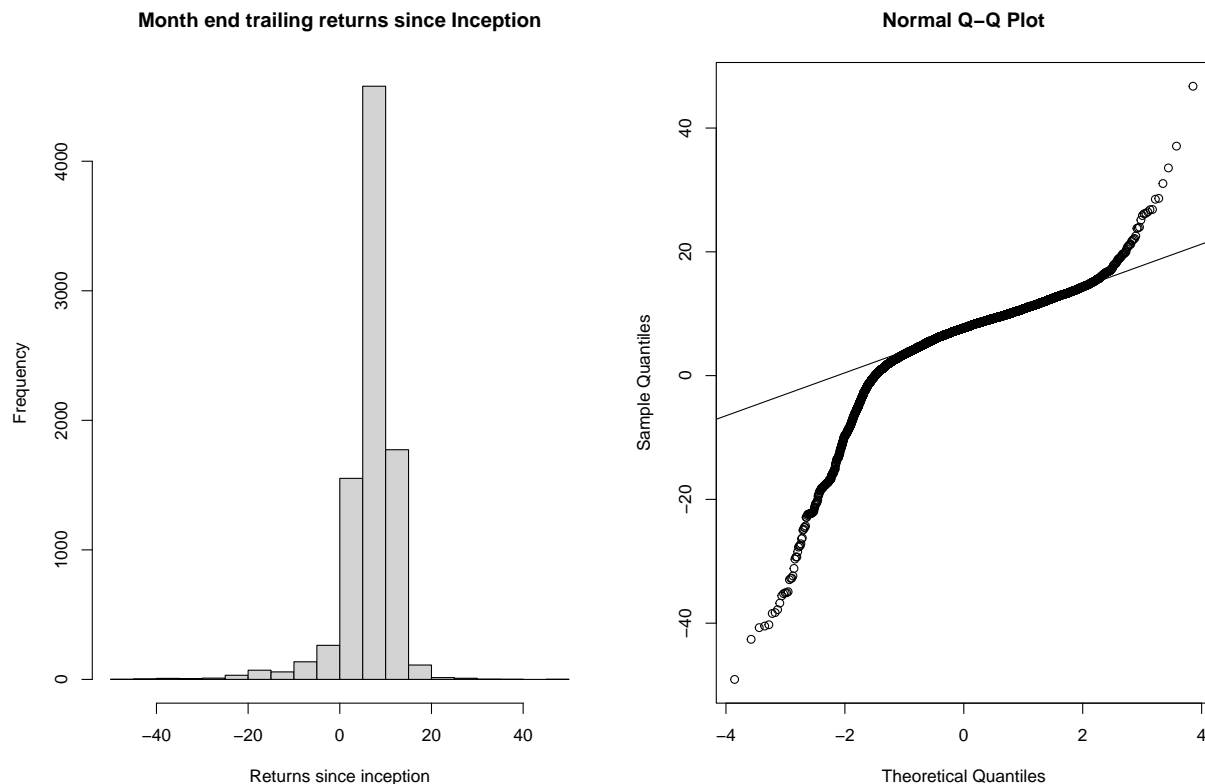


Month end trailing returns, 10 year



Normal Q-Q Plot





Based on the graph, returns are mostly likely appear to be normally distributed for all graphs or returns

## (ii). A statistic test of returns are normal.

The statistic test is performed by a Shapiro-Wilk Test as following:

- (1). Perform Shapiro-Wilk Test by using R
- (2). Consider the significantly level  $\alpha = 0.05$ :
  - If p-value is not less than  $\alpha$  (i.e  $p\text{-value} > \alpha$ ), then, returns are normally distributed
  - If p-value is less than  $\alpha$  (i.e  $p\text{-value} < \alpha$ ), then, returns are not normally distributed.

**Note:** Since our sample size is large (8664), R only allows a Shapiro-Wilk Test with a maximum sample size of 5000. Thus, I'll construct 2 tests for each returns with sample size of 4332 for each variable.

- Month end trailing returns 1 year:

```
##
##  Shapiro-Wilk normality test
##
## data:  return1y[1:4332]
## W = 0.93314, p-value < 2.2e-16
```

```
##
## Shapiro-Wilk normality test
##
## data:  return1y[4333:length(return1y)]
## W = 0.92194, p-value < 2.2e-16
```

Based on the two test, both of them have  $p\text{-value} < \alpha = 0.05$ , thus month end trailing returns 1 year is not normally distributed.

- Month end trailing returns 5 years:

```
##
## Shapiro-Wilk normality test
##
## data:  return5y[1:4332]
## W = 0.98878, p-value < 2.2e-16
```

```
##
## Shapiro-Wilk normality test
##
## data:  return5y[4333:length(return5y)]
## W = 0.962, p-value < 2.2e-16
```

Based on the two test, both of them have  $p\text{-value} < \alpha = 0.05$ , thus month end trailing returns 5 years is not normally distributed.

- Month end trailing returns 10 years:

```
##
## Shapiro-Wilk normality test
##
## data:  return10y[1:4332]
## W = 0.97287, p-value < 2.2e-16
```

```
##
## Shapiro-Wilk normality test
##
## data:  return10y[4333:length(return10y)]
## W = 0.94541, p-value < 2.2e-16
```

Based on the two test, both of them have  $p\text{-value} < \alpha = 0.05$ , thus month end trailing returns 10 years is not normally distributed.

- Month end trailing returns since inception:

```
##
## Shapiro-Wilk normality test
##
## data:  returnInception[1:4332]
## W = 0.79725, p-value < 2.2e-16
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: returnInception[4333:length(returnInception)]  
## W = 0.77797, p-value < 2.2e-16
```

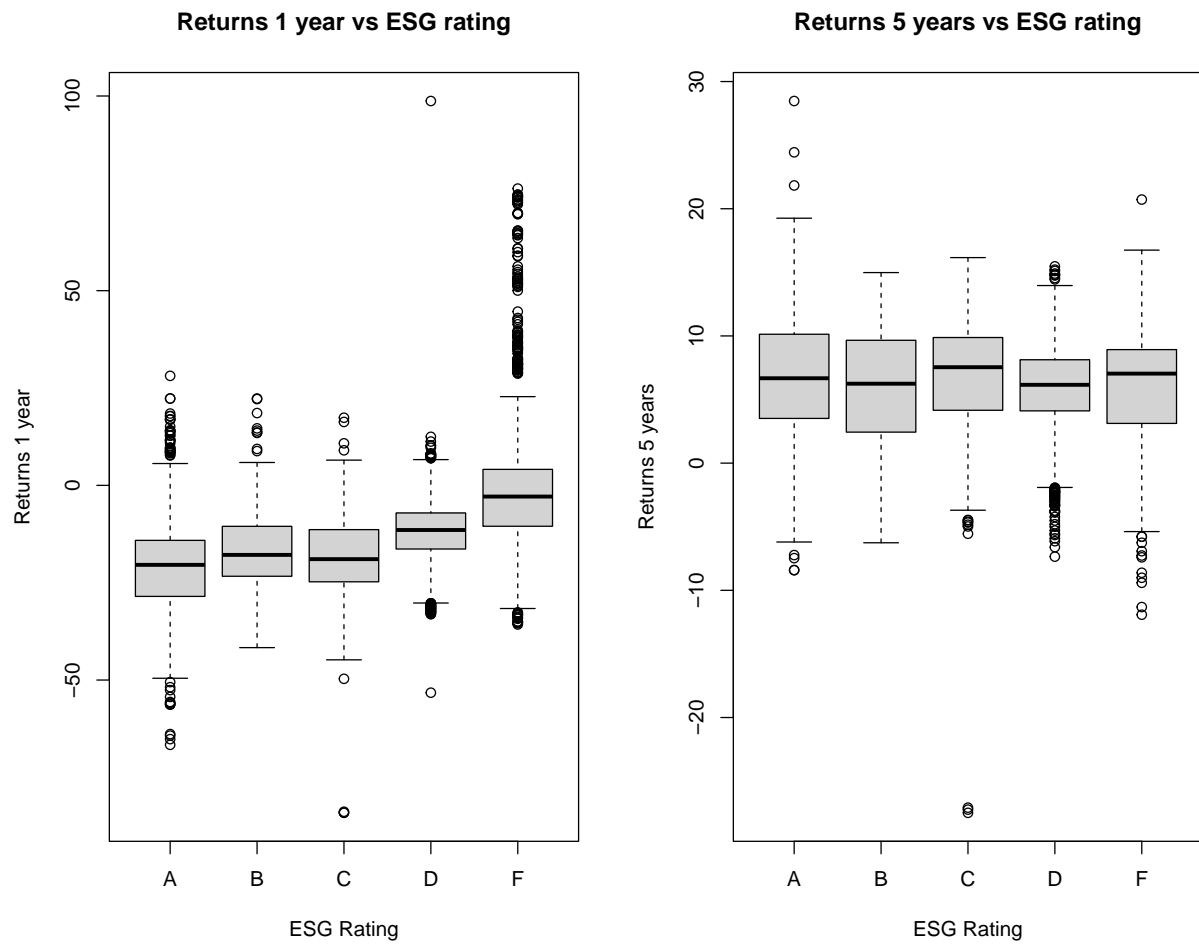
Based on the two test, both of them have  $p\text{-value} < \alpha = 0.05$ , thus month end trailing returns since inception is not normally distributed.

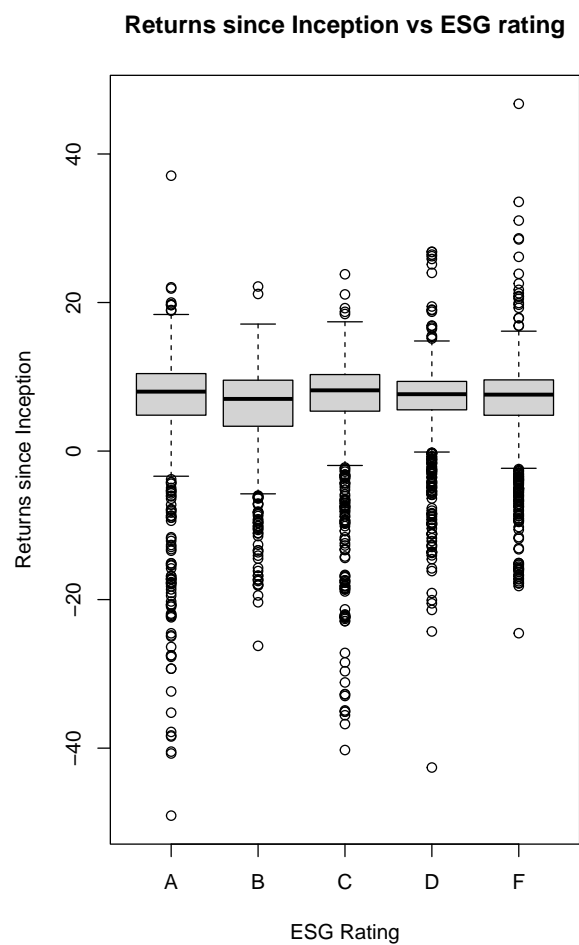
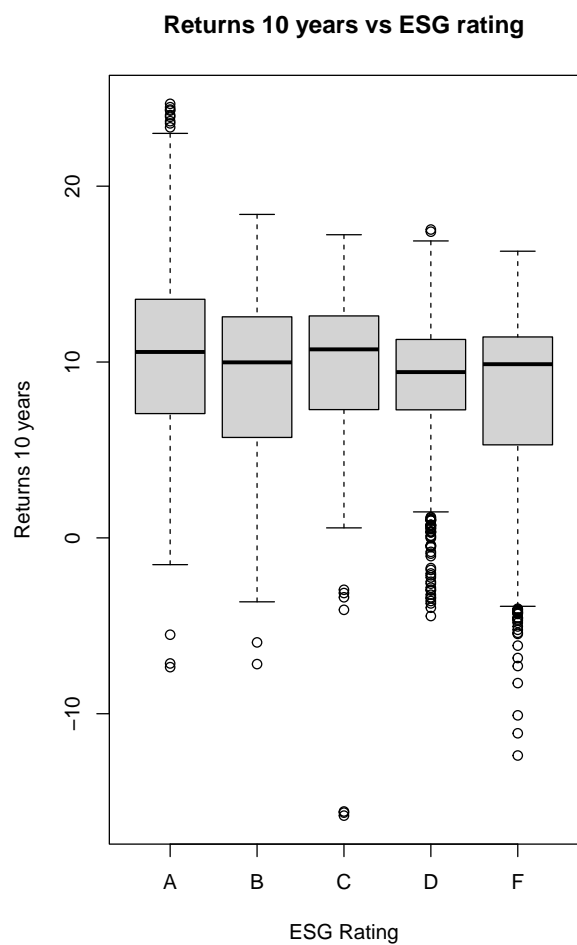
Overall, based on the tests, all returns distribution are not normal. Indeed, there is a limitation of Shapiron-Wilk Test, because the Shapiron-Wilk Test is sensitive to a small deviation, our sample size is large that could lead to a high probability of rejecting normal distribution. For this reason, it may have an impact of statistical methods we'd want to apply if we would like to predict returns, since some of the methods might assume normality.

## **2. Hypothesis: ESG rating have linear or polynomial relationship with returns**

For this hypothesis, we only consider “Fossil Free Funds: Fossil fuel grade” as the ESG rating and returns since inception as dependence variable.

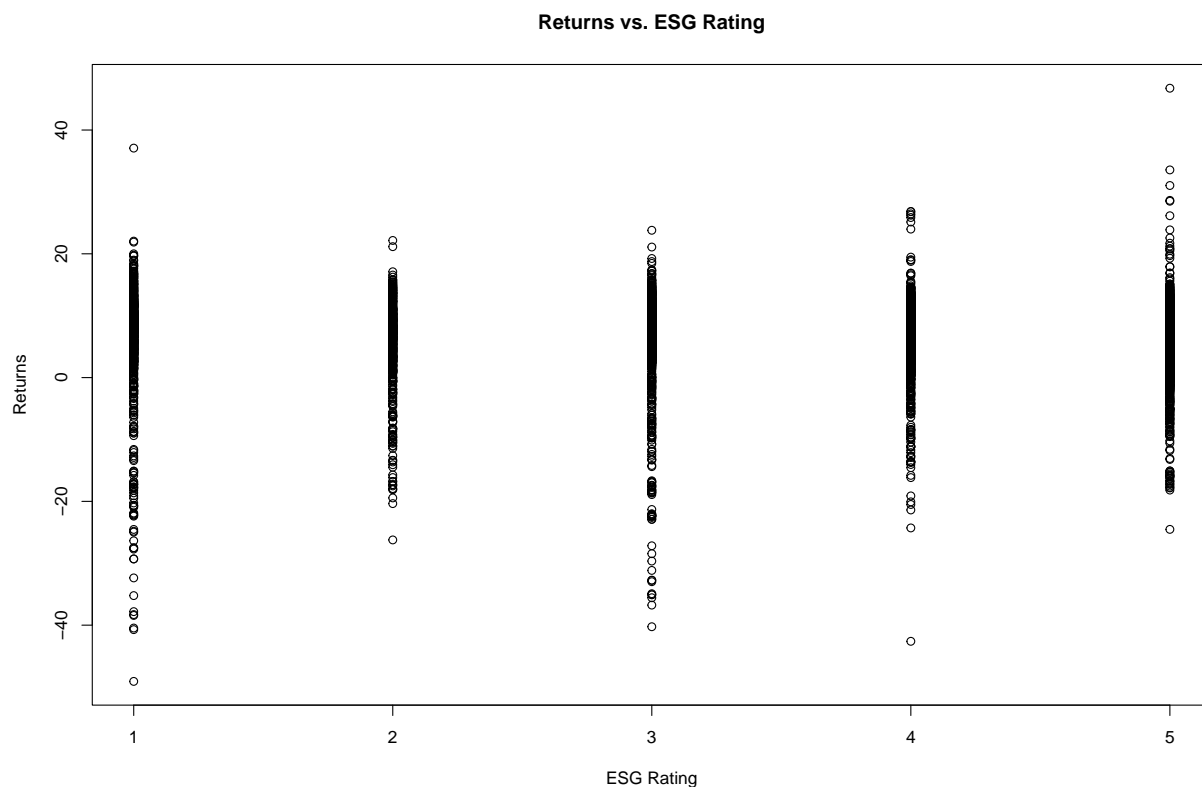
(i). Boxplot of ESG ratings vs. Returns







(ii). Plot of ESG rating vs. Returns



**Note:** 1 indicates grade A, 2 indicates grade B, 3 indicates grade C, 4 indicates D, and 5 indicates F.

(iii). Correlation coefficient between ESG rating and Returns

Compute their correlation coefficient, we obtain 0.042.

With a correlation coefficient  $r = 0.042$ , the association of ESG rating and returns are very weak or no association, so there is non-linear relationship. Thus, we can conclude that the ESG rating have a polynomial relationship with returns since inception.