Homework1: Coker

Probability Practice

Part A

RC = Random Clicker

TC = Truthful Clicker

$P(RC) = 0.3$

$P(Y|RC) = 0.5$

$P(Y) = P(Y|TC) * P(TC) + P(Y|RC) * P(RC)$

$0.65 = P(Y|TC) * 0.7 + 0.15$

$P(Y|TC) = (0.65 - 0.15)/0.7 = 0.714$

The fraction of people who are truthful clickers answered yes is 71.4%.

Part B

D = Disease

Pos = Testing Positive

$P(D) = 0.00025$

$P(Pos|D) = 0.993$

$P(Pos) = P(Pos|D) * P(D) + P(Pos|notD) * P(notD)$

$P(Pos) = 0.993 * 0.000025 + 0.0001 * 0.999975$

$P(Pos) = 0.0001248225$

$P(D|Pos) = \frac{P(Pos|D)*P(D)}{P(Pos)} = \frac{0.993*0.000025}{0.0001248225} = 0.198$

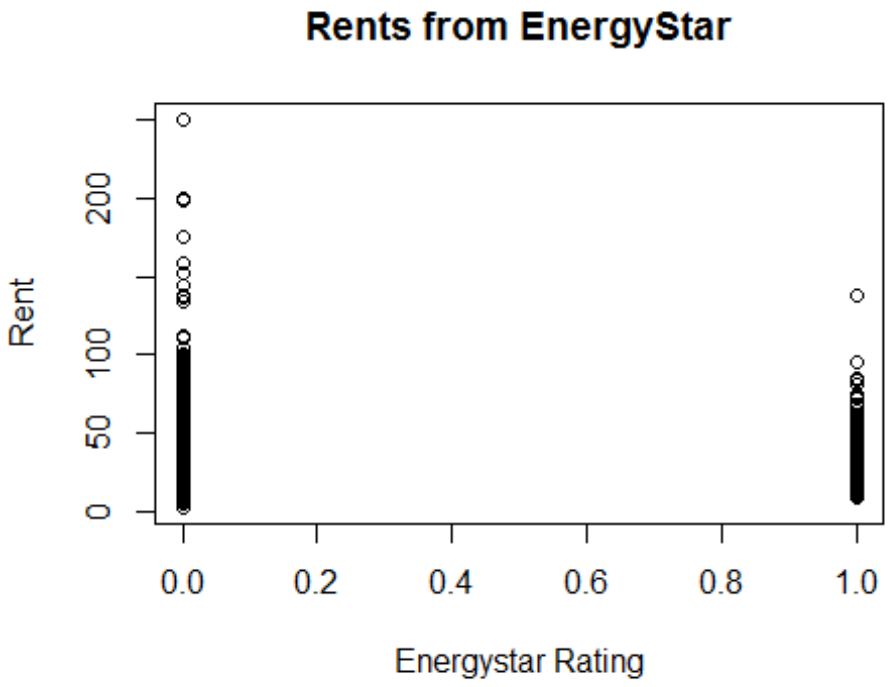
The probability that someone has the disease given that they tested positive is 20%.

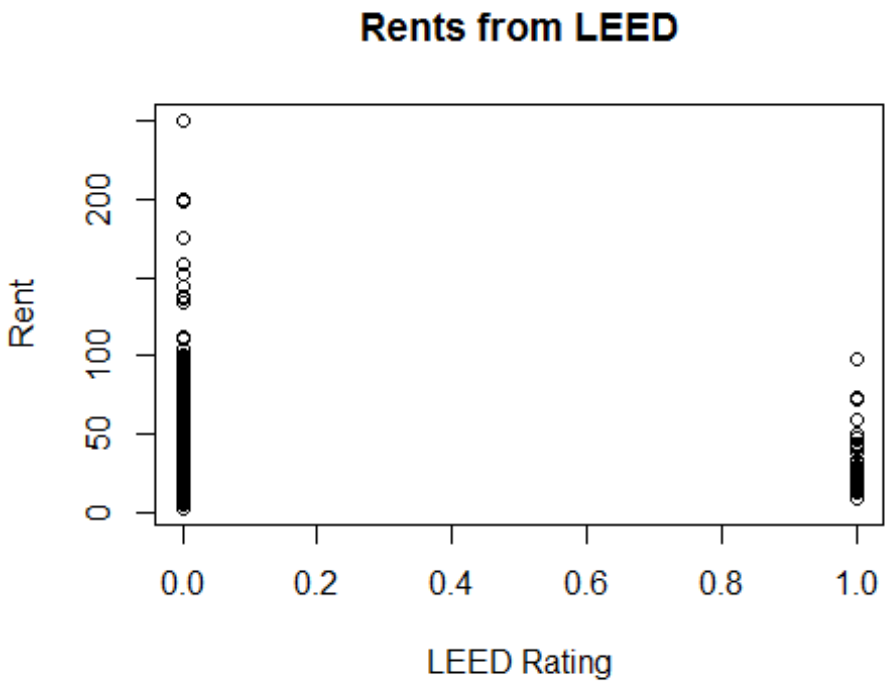
Given this low percentage, I don't think this test would be implemented in a universal testing policy because 80% of the people who tested positive would needlessly be required to undergo further testing only to learn that they did not have the disease.

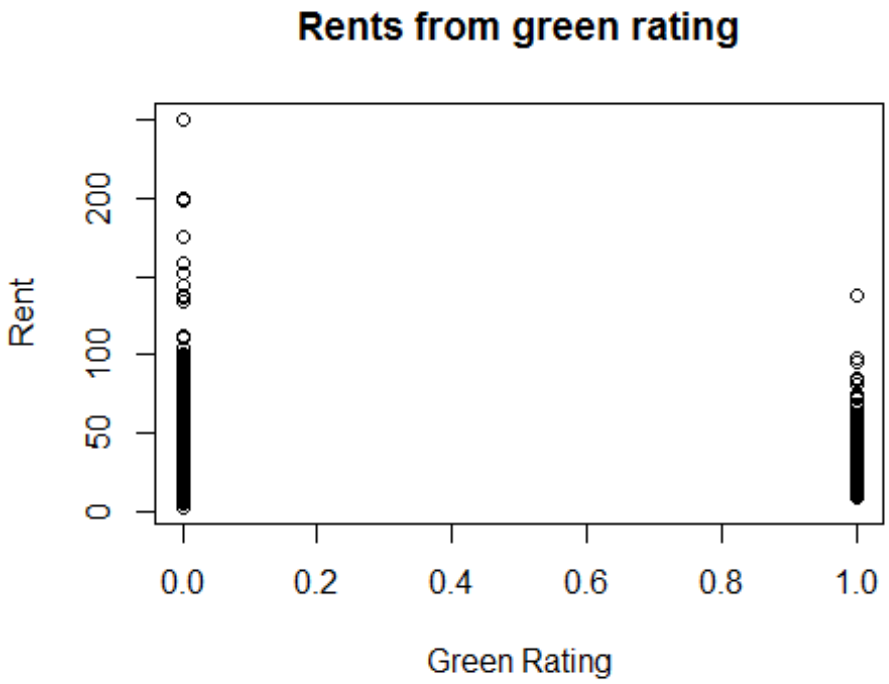
Exploratory analysis: green buildings

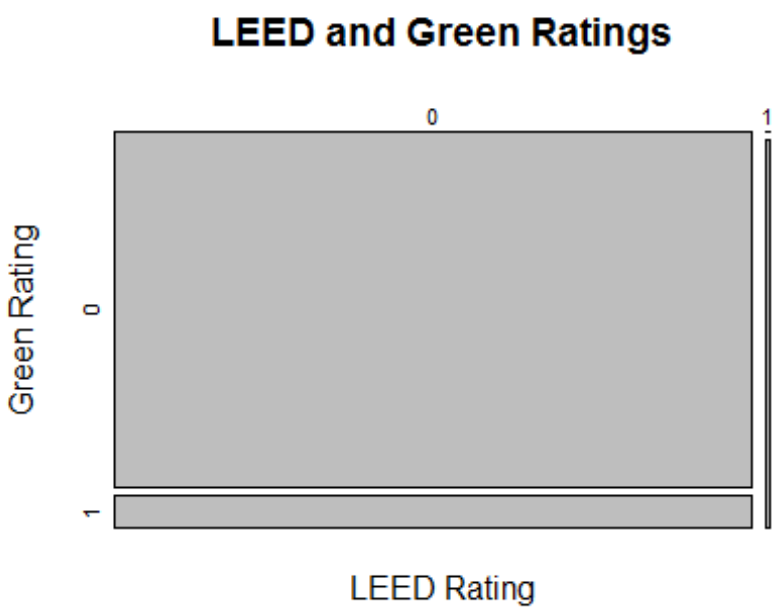
The previous analysis indiscriminately removed buildings that had occupancy rates below 10% without doing further research to determine whether these should be removed or if they have an impact on the final analysis of the rent amounts. The previous analysis also looked at the building types separately (green vs non green buildings), but did not consider that these two building types should be considered together due to the possibility that they may have an impact on each other when analyzed together.

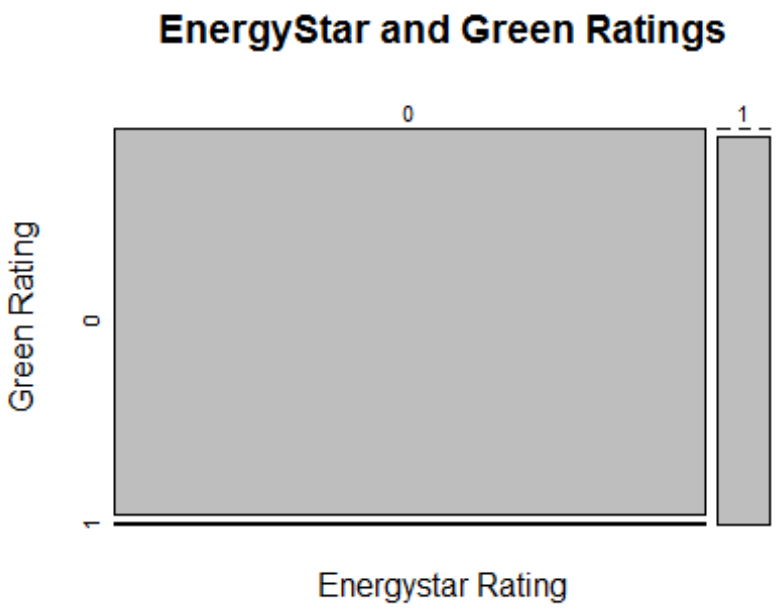
```
##
## Call:
## lm(formula = Rent ~ . - renovated - cd_total_07 - total_dd_07,
##     data = Buildings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.893  -3.625  -0.511   2.493 173.692
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -8.497e+00  1.000e+00  -8.494  < 2e-16 ***
## CS_PropertyID      3.028e-07  1.561e-07   1.940 0.052432 .
## cluster            7.613e-04  2.838e-04   2.682 0.007328 **
## size               6.702e-06  6.544e-07  10.241  < 2e-16 ***
## empl_gr            5.889e-02  1.546e-02   3.809 0.000140 ***
## leasing_rate       9.550e-03  5.323e-03   1.794 0.072855 .
## stories           -3.518e-02  1.616e-02  -2.177 0.029474 *
## age               -1.316e-02  4.179e-03  -3.149 0.001646 **
## class_a            2.890e+00  4.336e-01   6.664 2.85e-11 ***
## class_b            1.179e+00  3.409e-01   3.459 0.000546 ***
## LEED               1.918e+00  3.581e+00   0.536 0.592177
## Energystar        -2.004e-01  3.817e+00  -0.053 0.958125
## green_rating       6.789e-01  3.838e+00   0.177 0.859610
## net               -2.594e+00  5.915e-01  -4.386 1.17e-05 ***
## amenities          6.512e-01  2.511e-01   2.593 0.009529 **
## hd_total07         5.708e-04  8.172e-05   6.985 3.08e-12 ***
## Precipitation      4.675e-02  1.602e-02   2.918 0.003534 **
## Gas_Costs         -3.805e+02  7.318e+01  -5.200 2.05e-07 ***
## Electricity_Costs  1.937e+02  2.435e+01   7.957 2.01e-15 ***
## cluster_rent       1.011e+00  1.393e-02  72.612  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.412 on 7800 degrees of freedom
##   (74 observations deleted due to missingness)
## Multiple R-squared:  0.6126, Adjusted R-squared:  0.6116
## F-statistic:   649 on 19 and 7800 DF,  p-value: < 2.2e-16
```


Rents from EnergyStar
Rent
0
50
100
200
0.0
0.2
0.4
0.6
0.8
1.0
Energystar Rating


Rents from LEED
Rent
0
50
100
200
0.0
0.2
0.4
0.6
0.8
1.0
LEED Rating


Rents from green rating
Rent
0
50
100
200
0.0
0.2
0.4
0.6
0.8
1.0
Green Rating


LEED and Green Ratings
0
1
Green Rating
0
1
LEED Rating



The summary output above shows that holding the other variables constant, neither Energystar, LEED, nor green_rating appear to have a significant impact on the rent of the building.

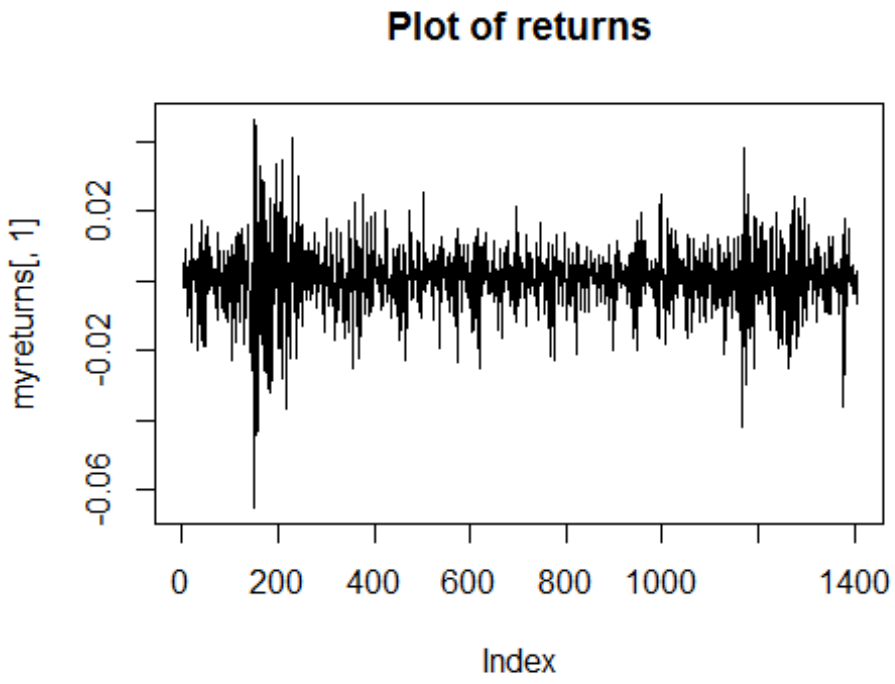
The plots above also show that having any of the these three ratings do not bring in a higher rent, and may bring in a lower rent, when analyzed separately.

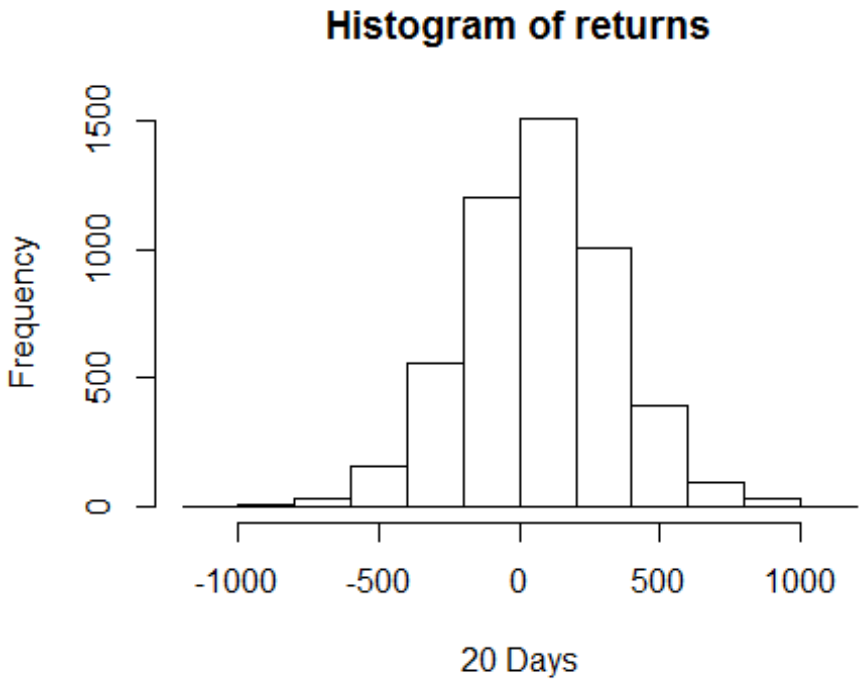
The mosaic plots above show that very few buildings have a green rating and most buildings are not LEED or Energystar rated. However, when a building has the LEED or Energstar ratings, it is also has a green rating.

It is my conclusion that rent is not currently affected by having a green building, but will need to be further researched when more buildings have been built with the LEED and Energyystar ratings.

Bootstrapping

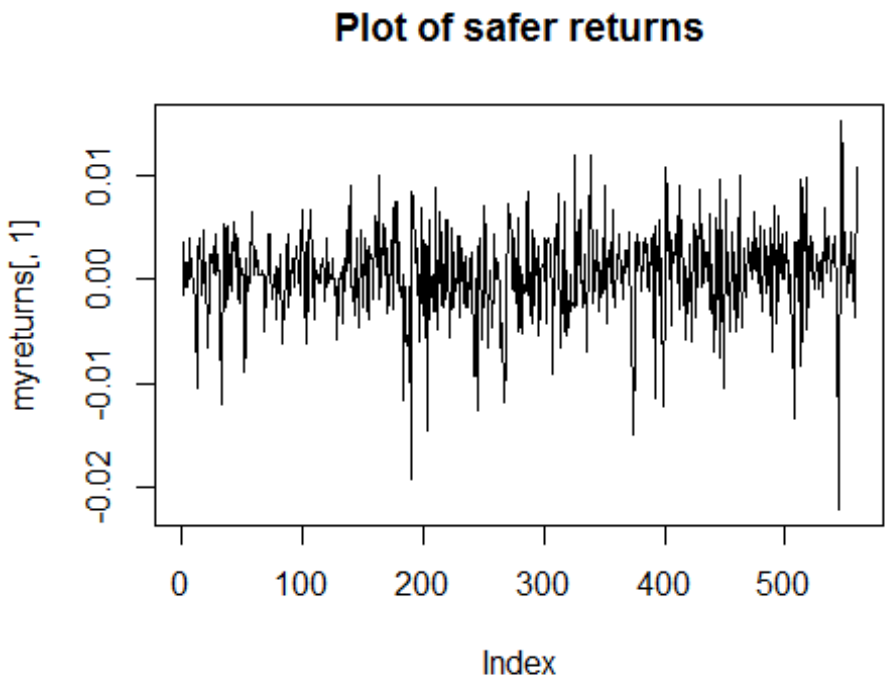
For an even split among 5 stocks: SPY, TLT, LQD, EEM, VNQ.

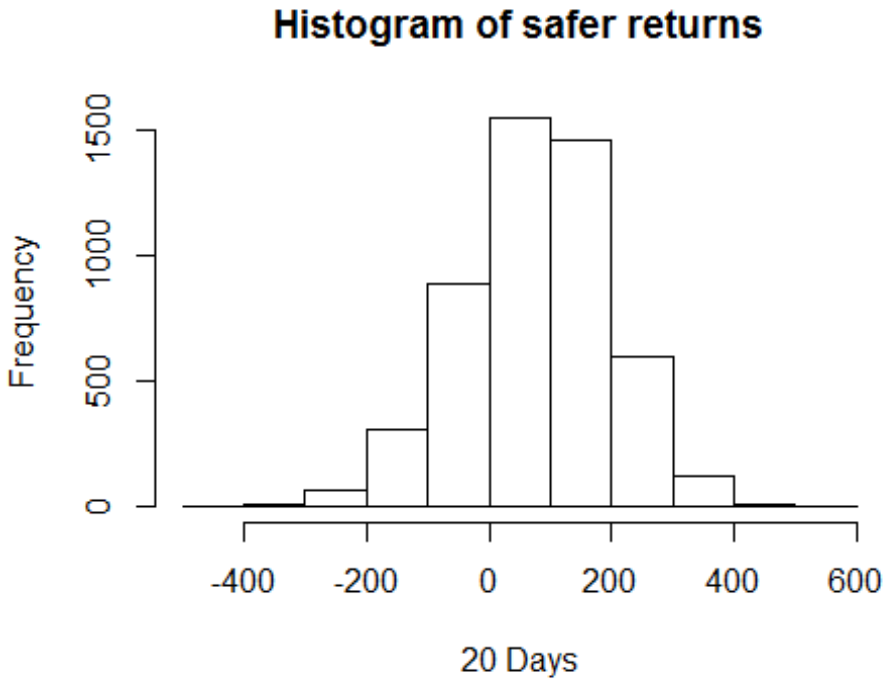

Plot of returns
myreturns[, 1]
0.02
-0.02
-0.06
0
200
400
600
800
1000
1400
Index


Histogram of returns
Frequency
0
500
1000
1500
-1000
-500
0
500
1000
20 Days

```
##         5%
## -370.1526
```

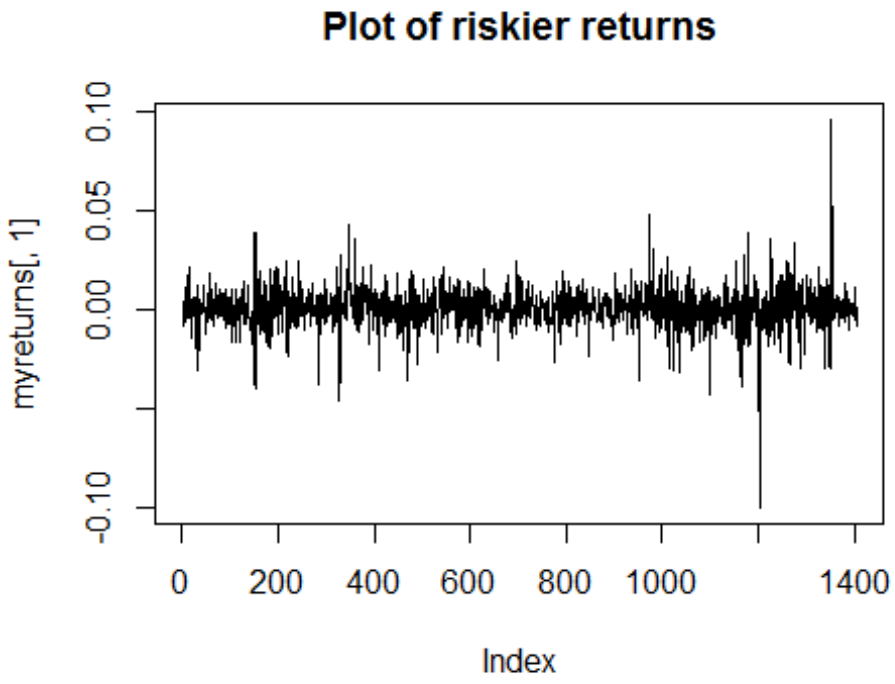
For a safer choice than the even split above. I chose to enevly invest in top rated bonds and fixed-income EFTs. Historically the bond market has been less vulnerable to price swings or volatility than the stock market.

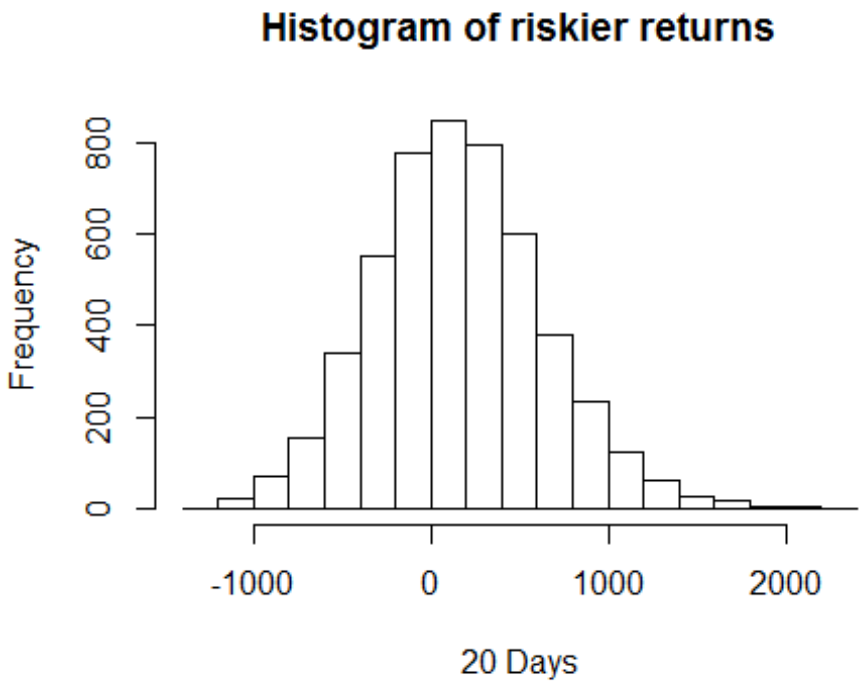




```
##       5% 
## -123.968
```

For a risker position that the even split above. These are high-yeild stocks, which provide the opportunity have much higher returns, but they are also have the potential to loss money as well


Plot of riskier returns
myreturns[, 1]
0.10
0.05
0.00
-0.10
0
200
400
600
800
1000
1400
Index

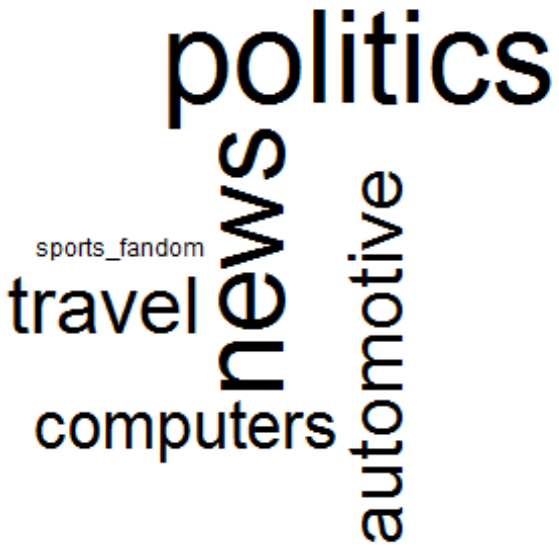

Histogram of riskier returns
Frequency
0
200
400
600
800
-1000
0
1000
2000
20 Days

```
##        5%
## -598.905
```

The histogram of the original, even spilt portfolio and the riskier porfolio both have a mean of gains/losses that are typically centered around zero. Whereas The safer portfolio has a men that is usually a bit higher than zero.

The relative risk of these three portfolios can also be seen in the differences in the value at risk of each portfolio at the 5% level. The original portfolio has a higher risk than the safer portfolio, and both af these have a lower risk than the risker portfolio.

Market segmentation




```
##      1                  2            3                4
## [1,] "health_nutrition" "politics"   "photo_sharing" "religion"
## [2,] "personal_fitness" "news"       "chatter"       "parenting"
## [3,] "outdoors"         "travel"     "shopping"      "sports_fandom"
## [4,] "cooking"          "automotive" "college_uni"   "food"
## [5,] "eco"              "computers"  "fashion"       "school"

## [1] 1421 1206 3707 1548
```

Above are the four clusters that this market can be segmented into, including the top interests for each segment. As well as a printout of the sizes of the four clusters.

Also included is a word cloud for one of the clusters.