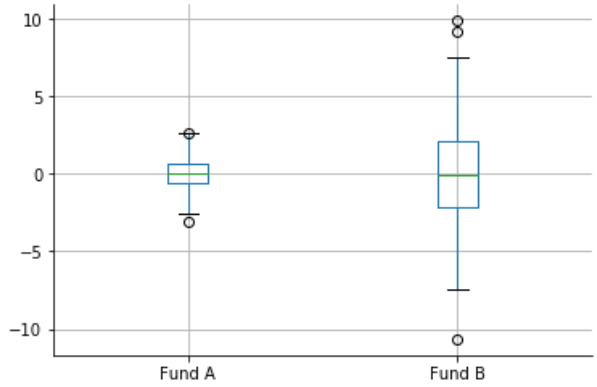
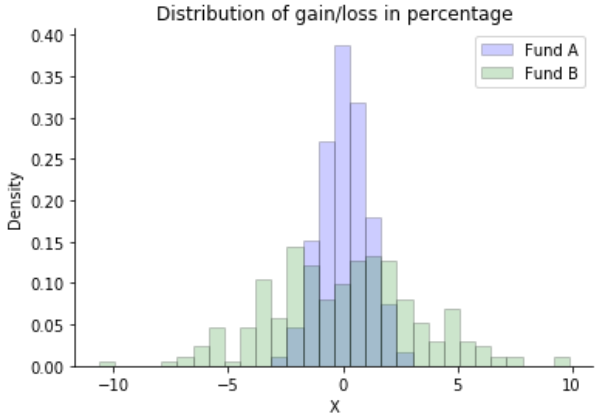
Description of the outputs and Reason of choosing

1.

(a) From the box and whiskers plot, we can see that fund B has a larger range than fund A with their mean are both close to mean 0. (.pdf may shows the top and right border but .word and jpynb have not, can be seen on Jupter notebook)



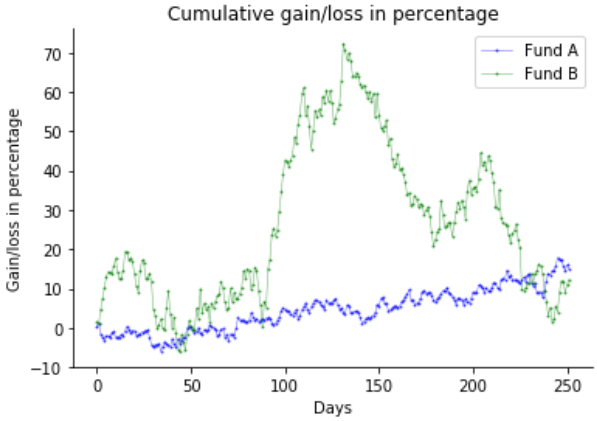
(b) From the histograms of density of funds, A and B, we can see the distributions of each fund.




```
Mean for fund A and B:
Fund A    0.059319
Fund B    0.048697
```

```
Variance for fund A and B:
Fund A     1.075422
Fund B    10.547832
```

(c) I choose to plot the cumulative gain/loss in the graph and show the data points with revising markersize and linewidth to a better visualization. As what it shows, we can conclude that fund A is more stationary than fund B; by visualizing the figure, fund B has lower mean 0.049 and much higher variance 10.55 than fund A; confirmed by calculating on Python.



2.
As the result showing on Jupyter notebooks, null hypothesis that mean of fund A in 2019 is the same as mean of fund B. Taking significance level of α=0.05 in the two tailed student's t test, we get the result t stats=0.049 and p value=0.96(two tailed).
With the definition that when p value<0.05, we reject the H0; since p value 0.96 is much higher than 0.05, we do not reject H0. Therefore, the mean of fund A is the same as the mean of fund B with the significance level of 0.05.

```
T test for H0: mean(A) = mean(B):
Ttest_indResult(statistic=0.049362132950207145, pvalue=0.9606503380102615)


Whether reject H0 or not:
Do not reject H0.
```

3. That's assume that the gain/loss in percentage is rate of return that R(t) = (P(t) – P(t-1))-1, and we are willing to find the annualized rate of return= (0.01*(R1)+1)*...*(0.01*R(252)+1)-1. Annualized rate of return of fund A is 14.56% and -1% for fund B. To sum up, fund B will lose 1% and fund A will gain 14.56% if the investor invests the fund from day 1 to day 252.

4.
Removing the top and the right line of all figures as mentioned in class, we can apparently observe the data through these figures.

For Q1:
(a) The box and whisker plot can show the spread and center of a data set with five points, including minimum, quartile Q1, median, quartile Q3 and maximum. Visualizing the boxplot of fund, A and B, we can see that fund B spreads wider and there are some extreme values.
(b) The histograms with transparent color and see the distribution of two data sets. We can see that both A and B have the mean value close to 0 and fund A has a higher density when x=0. Moreover, fund B spread more widely and has a longer tail, which means fund B's probability density function (PDF) is larger than fund A's distribution with larger values.
(c) Plot on cumulative frequency, we can see that if an investor invests fund A or B in day one, what will the price changes in a year. We can consider fund A as a stationary time series that it will continue gaining overtime; on the other hand, fund B is a non-stationary time series that it goes up and down every time step, it is hard to predict when it will go up or down.
For Q2:
Getting the t stats and p value of null hypothesis of the same mean, we can know where fund A and B have the closer mean.
For Q3:
Getting the annualized rate of return of fund, A and B, we can conclude which fund is better to invest during the days in 2019.

In my opinion, I would prefer the histograms and the boxplots. First, both of them can see the distribution of the data sets and the extreme values. Second, it is clear to see what the mean

values located. To illustrate, the histogram can clearly see the distribution of the funds, this would make me prefer to use histogram to visualize data.

5. For my observation through the data visualizing, I found out that fund A and B have really close mean values, but large difference variances and fund B has a longer tailed than fund A. This is interesting to me; long tailed investors usually look outside the norm regarding potential investments. Due to the higher variance, investing long-tailed fund B will be risker than fund A. For me, I would first choose to invest fund A for long-term and might invest fund B for a short term just to see what's really going on because I'm a risk aversion people.