1. create a normal probability plot of the sequence of yearly change for Dow index using data form the previous homework. Do you think the sequence looks like from a normal distribution?

1. What is the Shapiro-Wilk test value? Does it indicate normality?
2. Try square root and log transformation, and give the Shapiro-Wilk test values. Does these transformation improvement in normality based on these tests?

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
from scipy import stats
%matplotlib inline
```

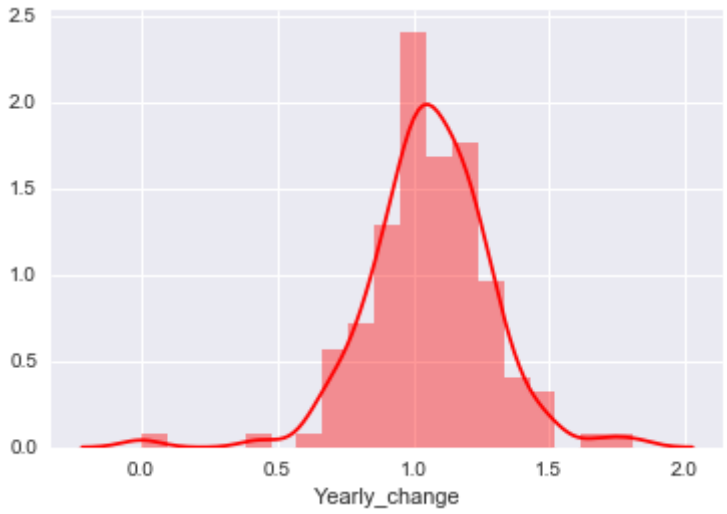
In [2]:
```
df = pd.read_csv('DJA_o.csv')
df.head(2)
```

Out[2]:

	Unnamed: 0	Date	DJIA	Year	Yearly_change
0	0	1/2/1886	39.4859	1886	1.041617
1	305	1/3/1887	41.1292	1887	0.922094

In [3]:
```
# Plotting distribution plot to see the distribution of data

x = df['Yearly_change']
sb.distplot(x,color='red')
```

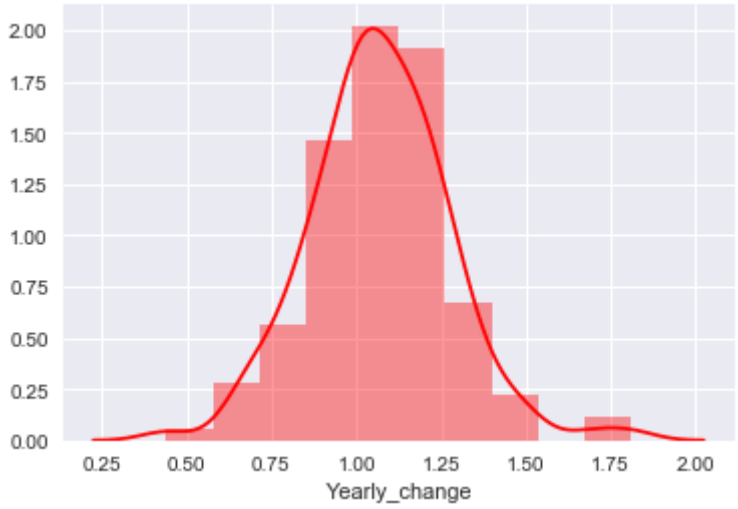
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x11bb070f0>



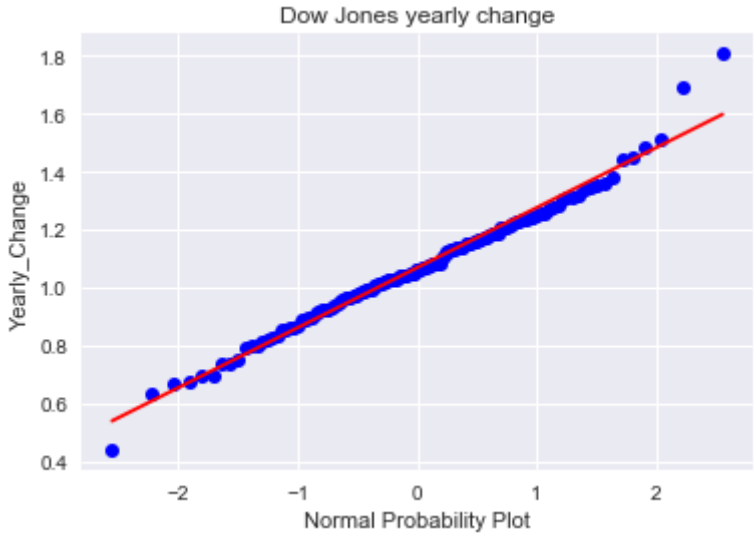
In [4]:
```
#dropping the outlier

df.drop(df.index[[-1]],inplace=True)
```

In [5]:
```
# Plot after dropping outlier

x = df['Yearly_change']
sb.distplot(x,color='red',bins=10)
```

Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x11eef67f0>



In [6]:
```
#Using stats plot
#Defining and plotting probability plot

change = df['Yearly_change']
stats.probplot(change,plot=plt)
plt.title('Dow Jones yearly change')
plt.xlabel('Normal Probability Plot')
plt.ylabel('Yearly_Change')
plt.show()
```



In [7]:
```
# Shapiro test

stats.shapiro(change)


# Result looks normal
```

Out[7]:
```
(0.9854334592819214, 0.1807788610458374)
```

In [8]:
```
# Shapiro test after Square root


change_sqrt = np.sqrt(change)
stats.shapiro(change_sqrt)



# There is not much impact on normality.
```

Out[8]:
```
(0.9856764078140259, 0.19090750813484192)
```

In [9]:
```
# Shapiro test after taking log

change_log = np.log(change)
stats.shapiro(change_log)


# Normality is impacted after taking log and it is < 0.05 now.
```

Out[9]:
```
(0.967064619064331, 0.0029979445971548557)
```

In [ ]:

In [ ]: