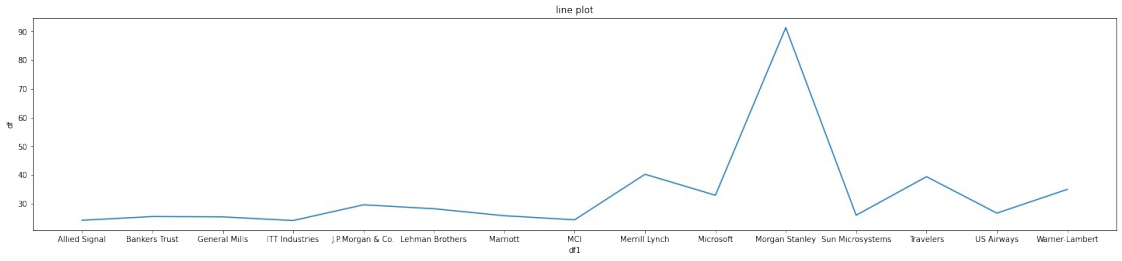
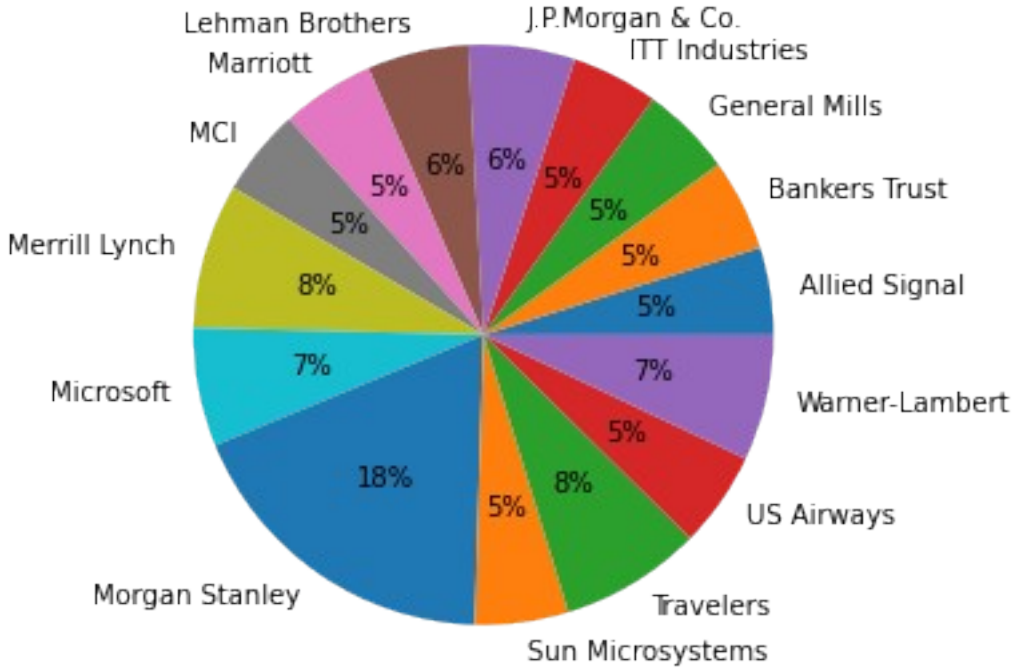
Q1

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32
.95,91.36,25.99,39.42,26.71,35.00])
df1=['Allied Signal','Bankers Trust','General Mills','ITT
Industries','J.P.Morgan & Co.','Lehman Brothers',
      'Marriott','MCI','Merrill Lynch','Microsoft','Morgan
Stanley','Sun Microsystems','Travelers','US Airways',
      'Warner-Lambert']

fig=plt.figure(figsize=(25,5))
plt.plot(df1,df)
plt.xlabel('df1')
plt.ylabel('df')
plt.title('line plot')

Text(0.5, 1.0, 'line plot')
```

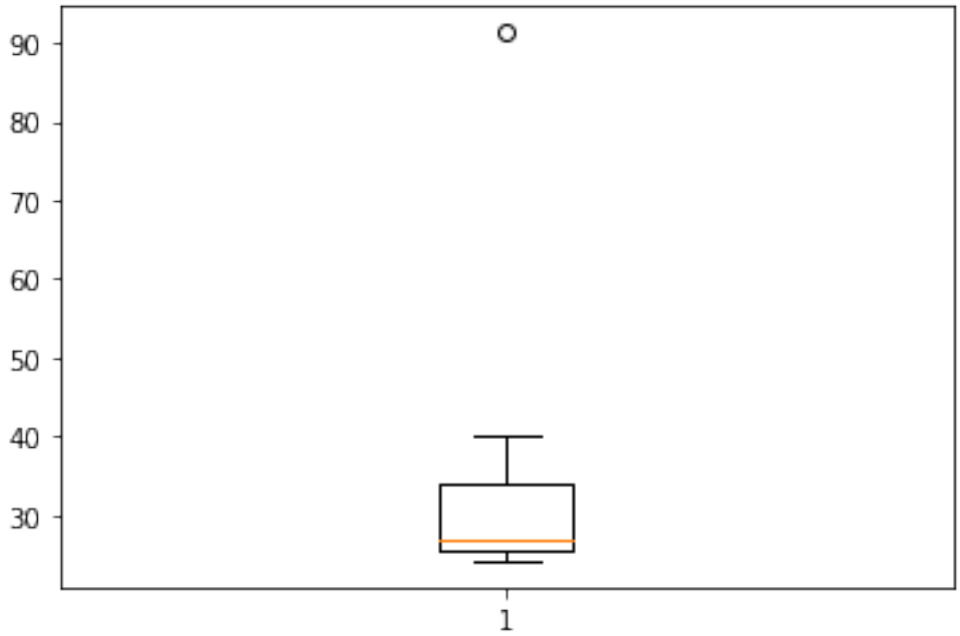



```
plt.figure(figsize=(10,5))
plt.pie(df,labels=df1,autopct='%1.0f%%')
plt.show()
```




```
plt.boxplot(df)
#sns.boxplot(df)
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x2473b075c70>,
  <matplotlib.lines.Line2D at 0x2473b075fd0>],
 'caps': [<matplotlib.lines.Line2D at 0x2473b086370>,
  <matplotlib.lines.Line2D at 0x2473b0866d0>],
 'boxes': [<matplotlib.lines.Line2D at 0x2473b075910>],
 'medians': [<matplotlib.lines.Line2D at 0x2473b086a30>],
 'fliers': [<matplotlib.lines.Line2D at 0x2473b086d90>],
 'means': []}
```




```
df.mean()
```

```
33.27133333333333
```

```
df.std()
```

```
16.945400921222028
```

```
df.var()
```

```
287.1466123809524
```

Q2. Answer the following three questions based on the box-plot above.

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

(ii) What can we say about the skewness of this dataset?

(iii) If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

ANS

i). IQR = Q3-Q1(upper quartile – lower quartile) 13 – 5 = 9 It means that 9 is the median which is called Q2

ii). The skewness of this dataset is right skewed which is positively skewed in other words. Here mean>median>mode

iii). If it was found that the data point with the value 25 is actually 2.5, then our new box-plot will definitely be affected because in that case there will be no outliers found in the box plot. It will affect the mean also.

Q3. Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?

(ii) Comment on the skewness of the dataset.

(iii) Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

ANS

i). The mode of this dataset lies in between 4 – 8 values of y whose frequency is above 20.

ii). This histogram represents the right skewed distribution of data which has a peak towards the left. Here mean is greater than median and median is greater than the mode. So in other words this data is positively skewed. It has 2 peaks of the same length which means the distribution is bimodal.

iii). As both the plots helps to visualize the numerical data. Here we can clearly see that in both the plots the tail is towards the right side, data value at 25 represents the outliers. Between 5 – 11 we observe more data. Both the plots showing right skewed data distribution. 50% data lies near 7(approx) value where there is our median. On the right of the median there is our mean and on the left the median there is mode.

Q4

AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that “could happen.” Suppose that one in 200 long-distance telephone calls is misdirected. What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

ANS

As we know, one in 200 long-distance telephone calls is misdirected. Probability of call misdirecting = 1/200 Probability of call not misdirecting =1-1/200=199/200 n=5 p=1/200 q=199/200

probability that at least one in five attempted telephone calls reaches the wrong number

Q5

Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

(i) What is the most likely monetary outcome of the business venture?

(ii) Is the venture likely to be successful? Explain

(iii) What is the long-term average earning of business ventures of this kind? Explain

(iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure

ANS.

i). The most likely monetary outcome of the business venture is 2000 because it has the maximum probability which is 0.3

ii). Yes, the venture is likely to be successful because the Expected value is in positive which is $800.

iii)(x) = -2000,-1000,0,1000,2000,3000

P(x) = 0.1,0.1,0.2,0.2,0.3,0.1

E(X)=(x)*P(x) = -200,-100,0,200,600,300

Total=800

So, the long-term average earning of business ventures of this kind is $ 800

iv). $E(X^2) = x^2 * P(x)$= 4000000,100000,0,200000,1200000,9000000= 2800000

Var (X) = $E(X^2) - \{ E(X) \}^2$

= $2800000 - 800^2$

= 2160000 (very High)

Because the Variability is very high as we can see, So the Risk is high.