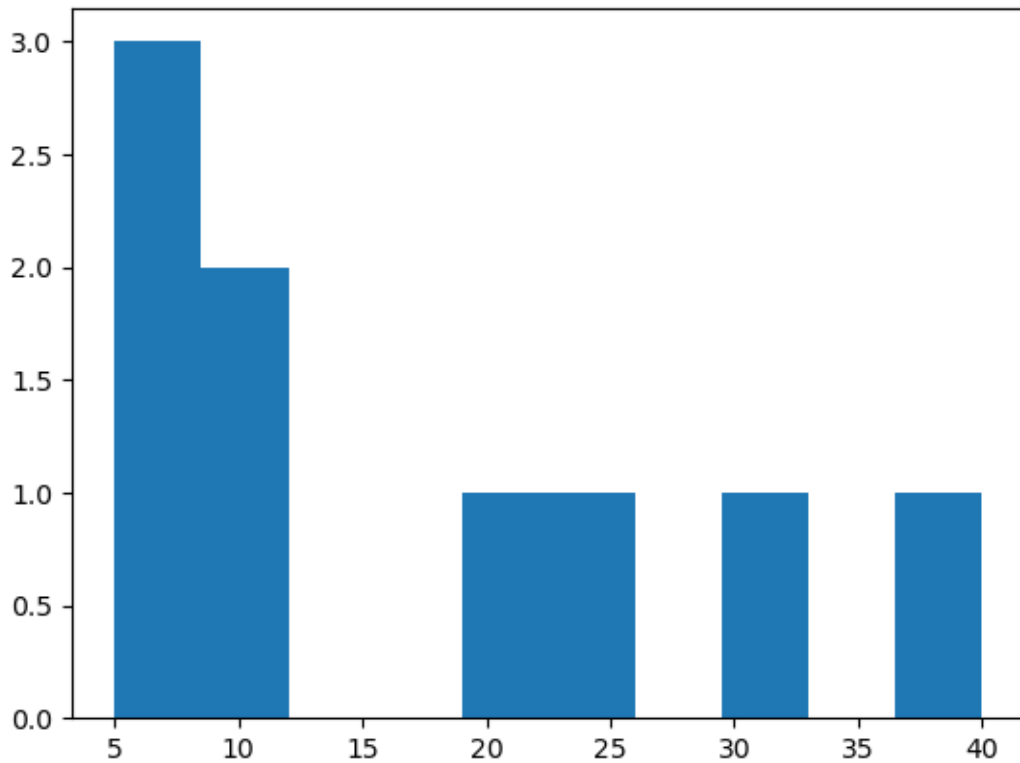


Lecture20

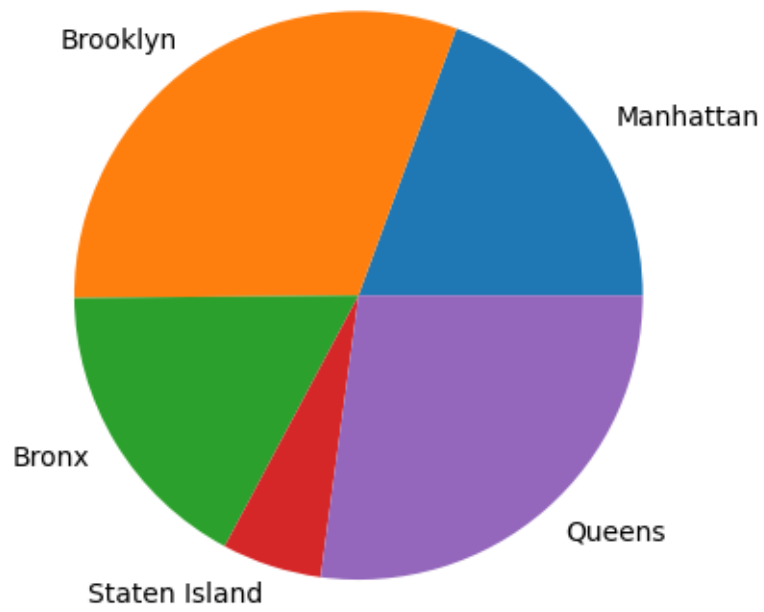
April 18, 2024

```
[2]: import matplotlib.pyplot as plt
```

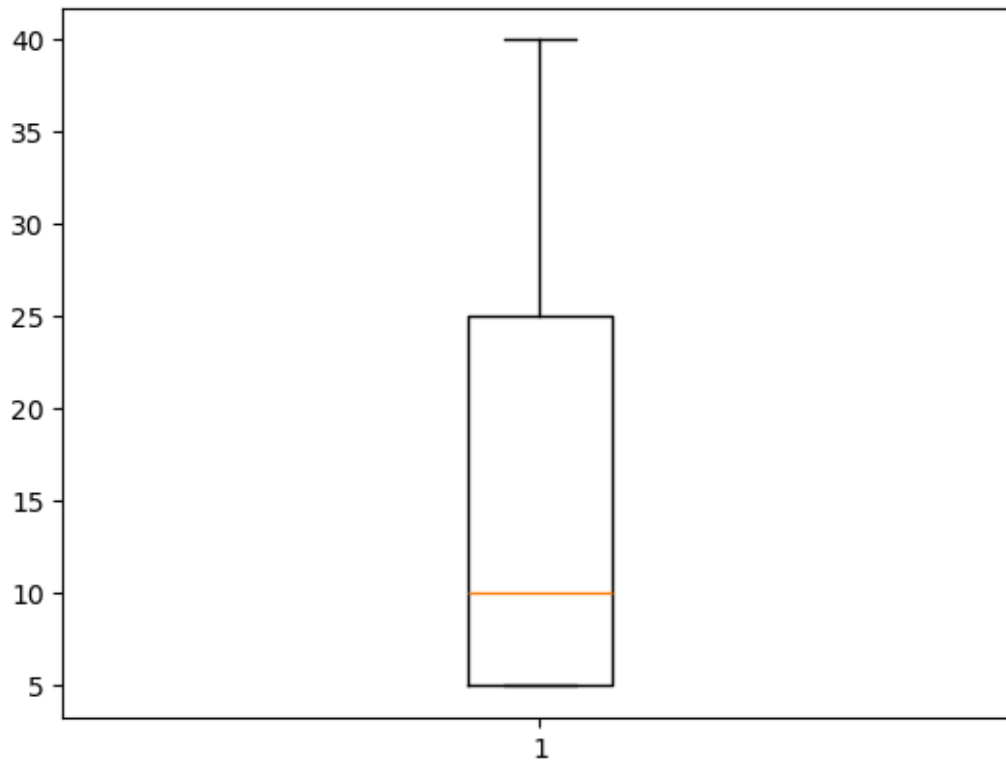
```
[5]: # Histogram  
L=[5,5,5,10,10,30,20,25,40]  
plt.hist(L);  
# plt.show()
```



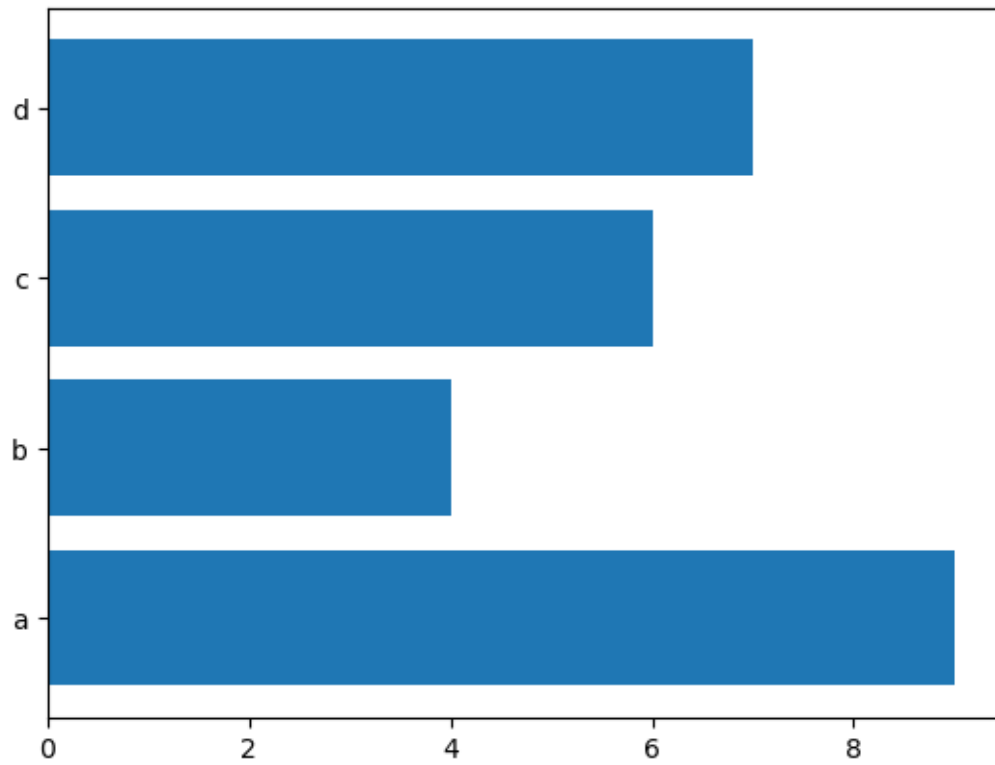
```
[9]: # pie chart  
values=[1628701, 2582830, 1432132, 476179, 2278906]  
labels=['Manhattan', 'Brooklyn', 'Bronx', 'Staten Island', 'Queens']  
plt.pie(values, labels=labels);
```



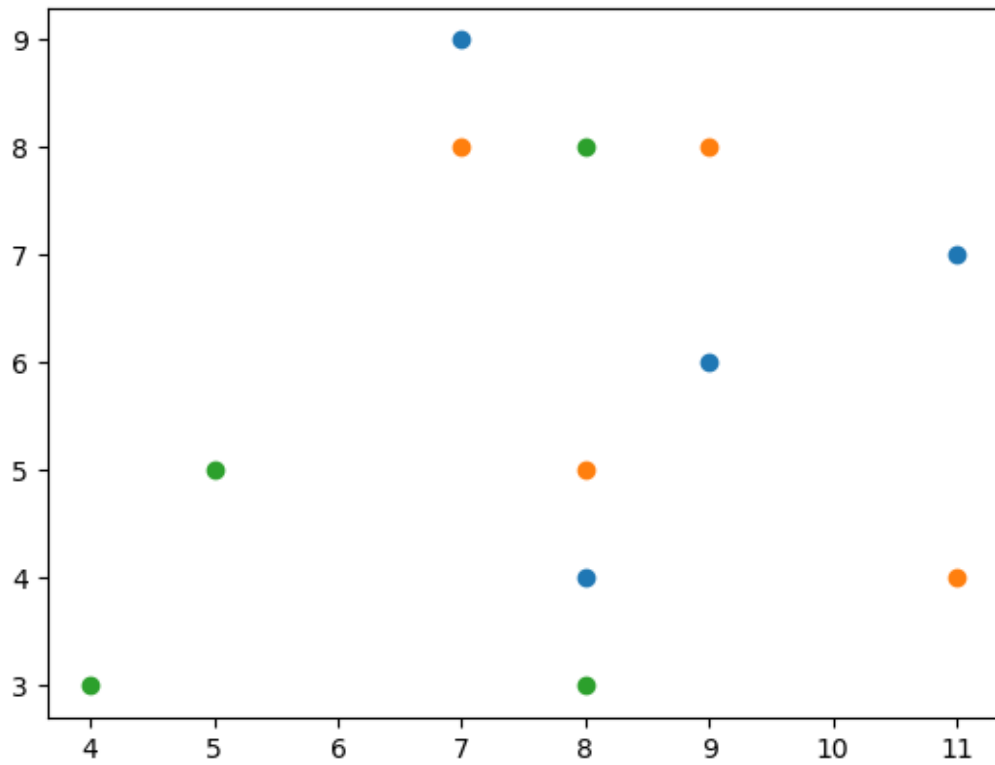
```
[11]: # boxplot chart  
plt.boxplot(L);
```



```
[14]: # bar chart
x=['a','b','c','d']
y=[9,4,6,7]
# plt.bar(x,y);
plt.barh(x,y);
```

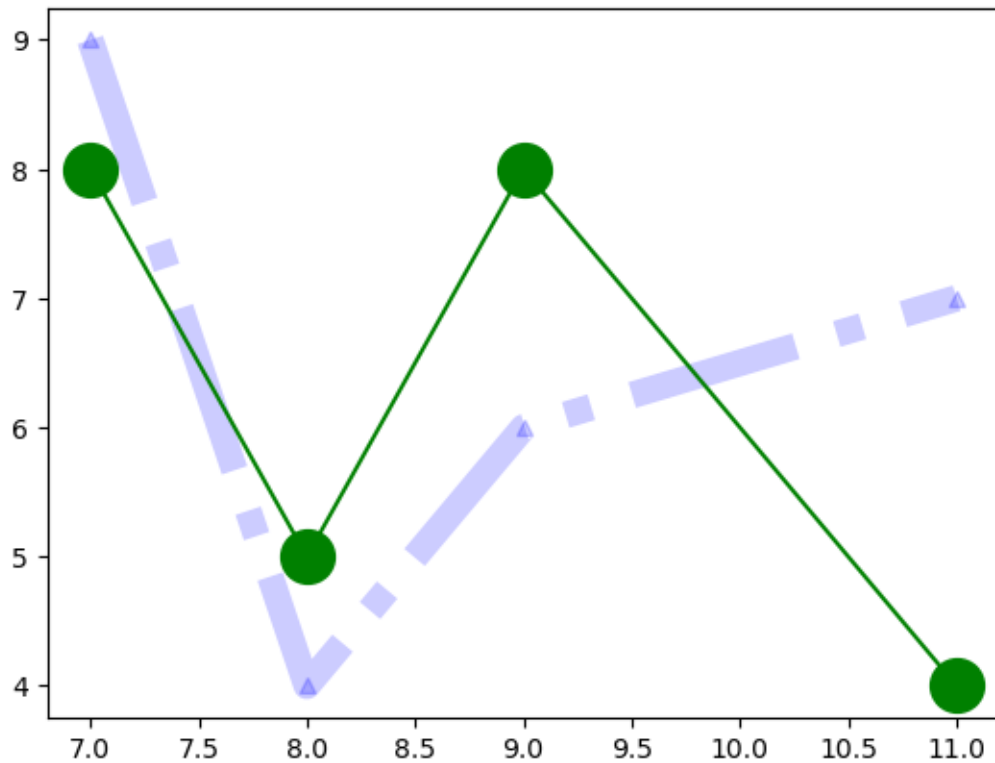


```
[18]: # Scatter Plot
x=[7,8,9,11]
y=[9,4,6,7]
y2=[8,5,8,4]
z=[3,5,8,3]
plt.scatter(x,y)
plt.scatter(x,y2)
plt.scatter(y2,z);
```

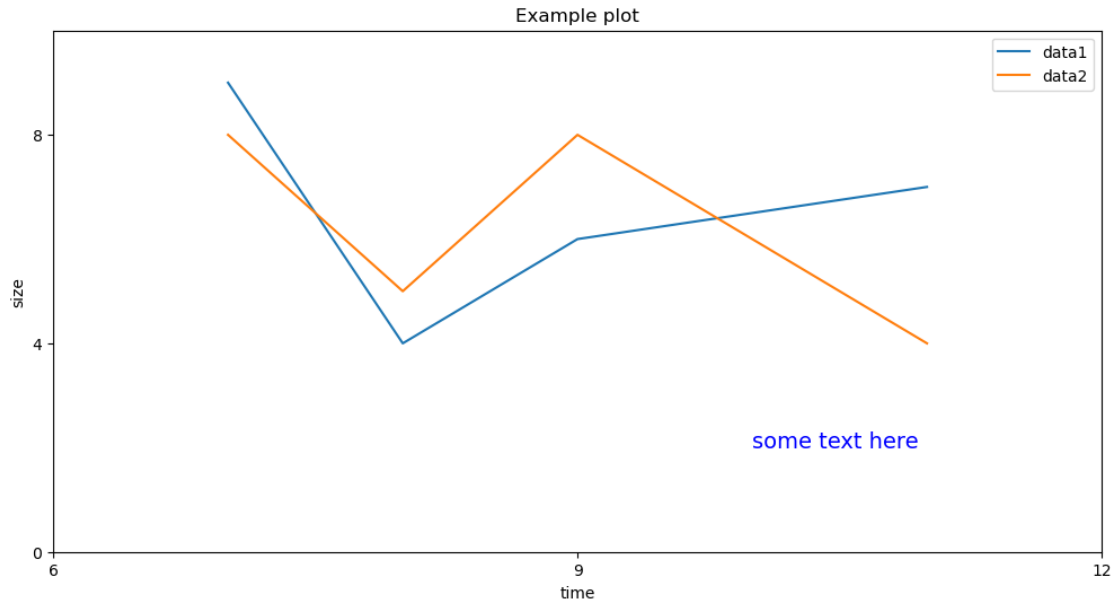


```
[27]: # Plot chart
plt.plot(x,y, 'b^-.', linewidth=10, alpha=.2)
plt.plot(x,y2, color='green', marker='o', markersize=20)
```

```
[27]: [<matplotlib.lines.Line2D at 0x7f76598f6fd0>]
```



```
[40]: # Plot Features
plt.figure(figsize=(12,6))
plt.plot(x,y, label='data1')
plt.plot(x,y2, label='data2')
plt.legend(loc=1)
plt.title('Example plot')
plt.xlabel('time')
plt.ylabel('size')
plt.xlim(6,12)
plt.ylim(0,10)
plt.xticks([6,9,12])
plt.yticks([0,4,8])
plt.text(10,2,'some text here', color='blue', size=14)
plt.show()
```



0.0.1 Exercise

```
[45]: # cat datafiles/gold_hist.txt
# - read the data as dataframe (year, price)
# - plot data with time on horizontal axis and gold price on the vertical axis
#   ↳ (provide title, and labels for each axis)
```

```
[48]: import pandas as pd
f=open('datafiles/gold_hist.txt')
data=f.read()
f.close()
```

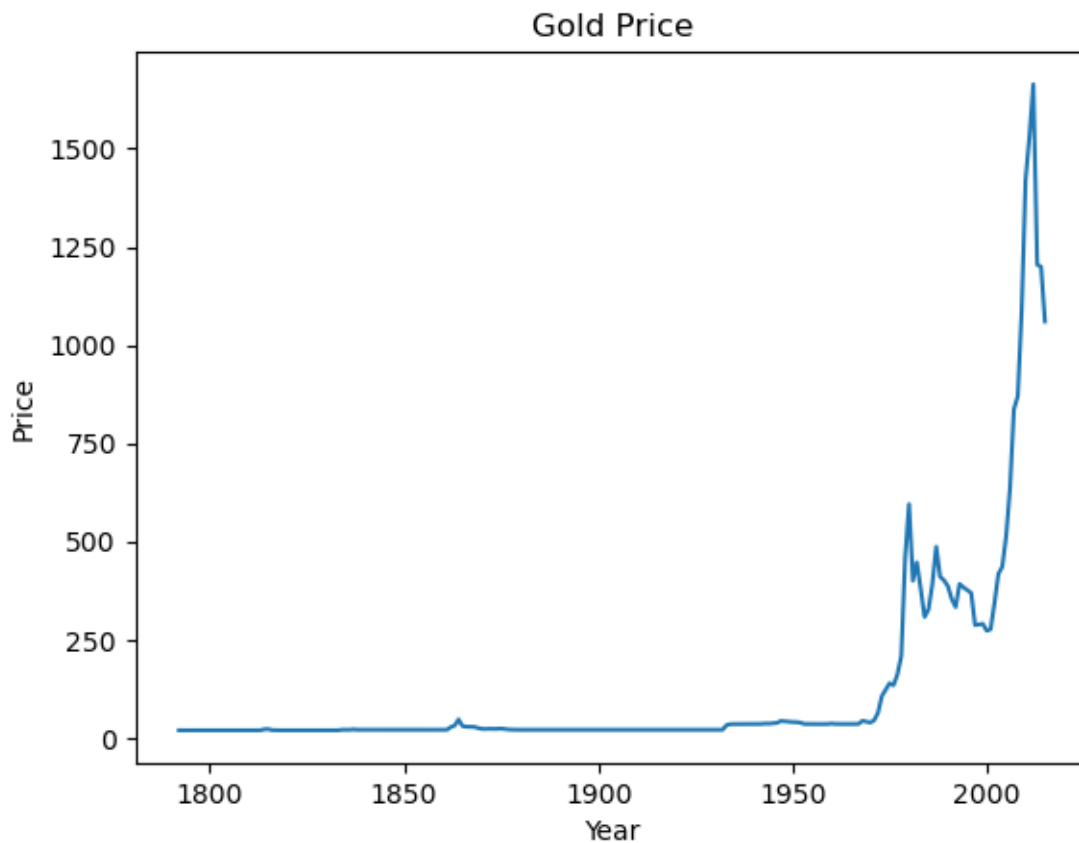
```
[66]: # [int(x) for x in data.split()[::2]]
# import numpy as np
# np.array(years).astype(int)
```

```
[71]: years=data.split()[::2]
prices=[p.replace('$','').replace(',','') for p in data.split()[1::2]]
df=pd.DataFrame({'year':years, 'price':prices})
df['year']=df['year'].astype(int)
df['price']=df['price'].astype(float)
df.head()
```

```
[71]:   year  price
0  2015  1060.00
1  1994   383.25
2  1973   106.48
```

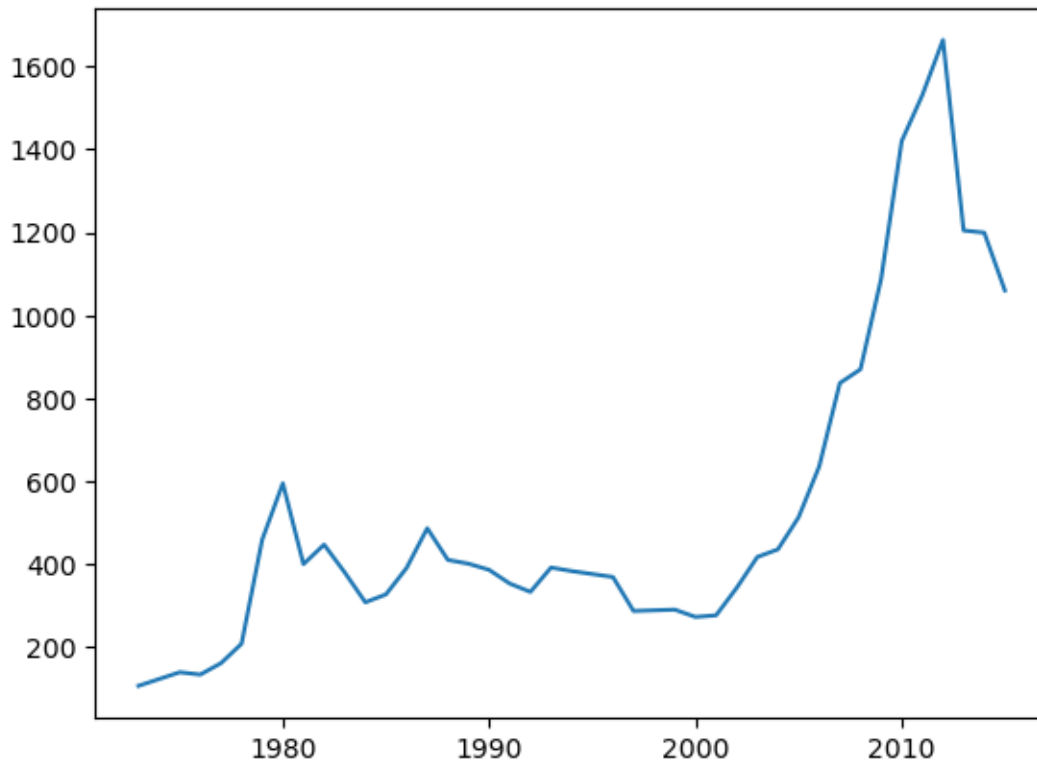
```
3 2014 1199.25
4 1993 391.75
```

```
[81]: df=df.sort_values(by='year')
plt.plot(df.year, df.price)
# plt.yscale('log')
plt.title('Gold Price')
plt.xlabel('Year')
plt.ylabel('Price');
```



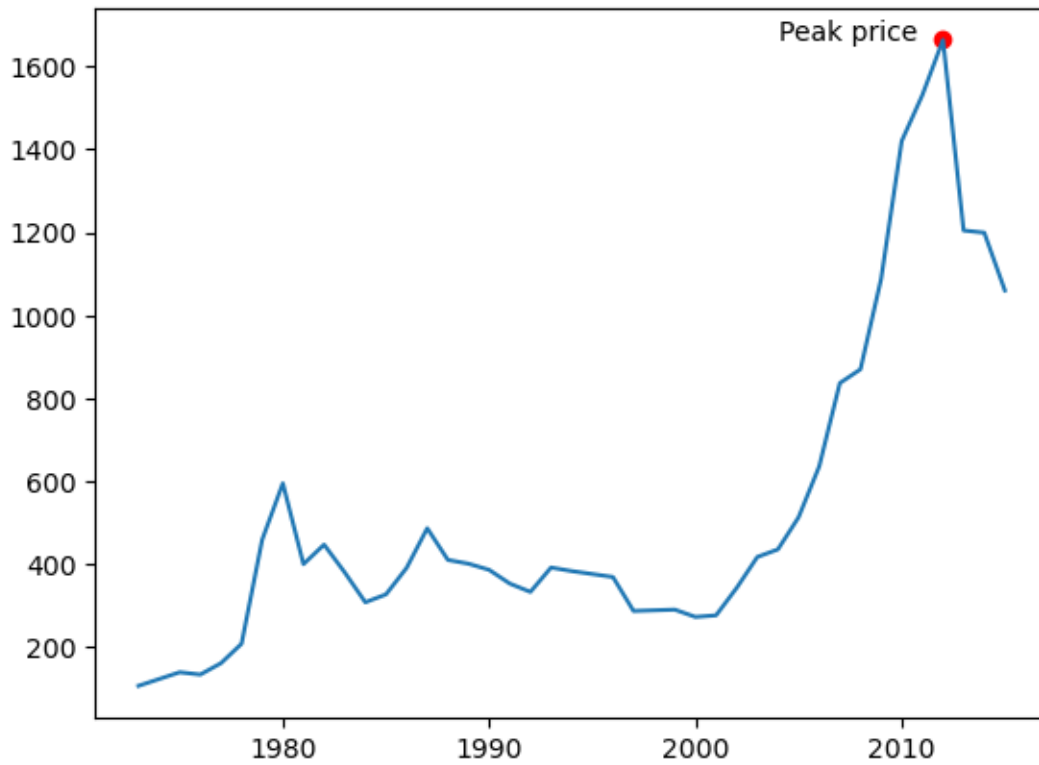
```
[82]: # - Plot the data again from 1972 and after
# - Label the point representing the peak price
```

```
[87]: tmpdf=df.loc[df.year>1972,: ]
plt.plot(tmpdf.year, tmpdf.price);
```

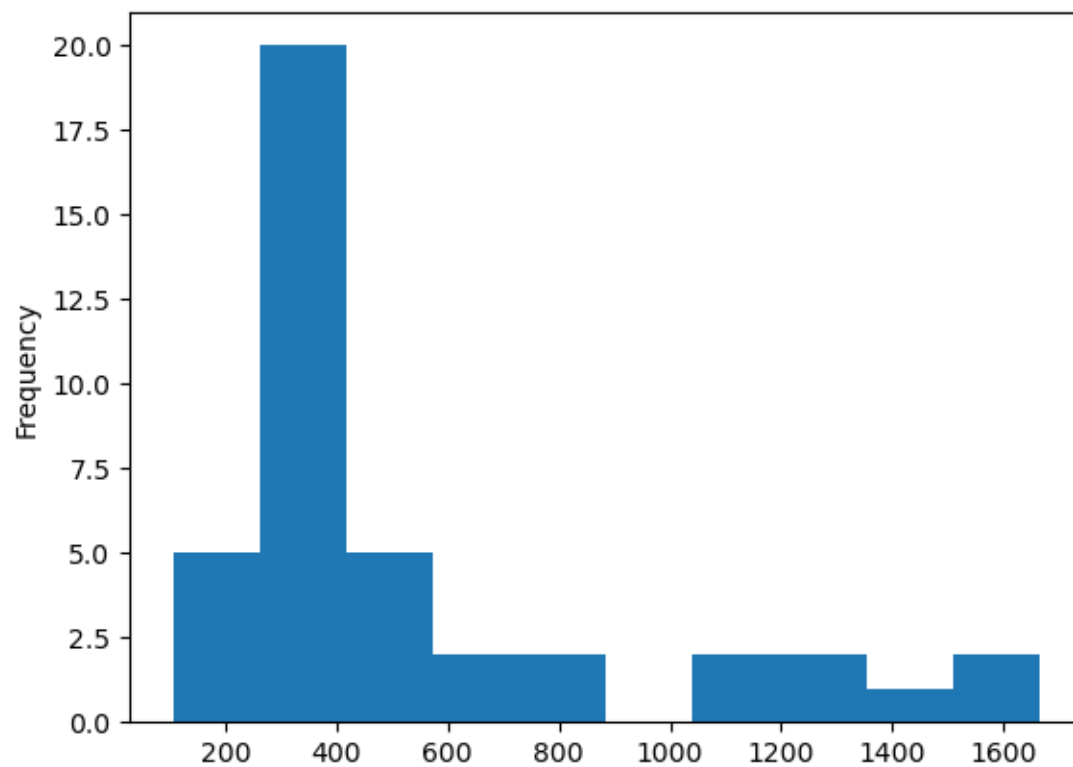
```
[94]: peak_price=tmpdf.price.max()
      peak_year=tmpdf.loc[tmpdf.price == peak_price,'year'][9]
```

```
[102]: plt.plot(tmpdf.year, tmpdf.price)
      plt.text(peak_year-8, peak_price, 'Peak price')
      plt.scatter(peak_year, peak_price, color='red');
```



```
[108]: tmpdf.price.plot(kind='hist')
# plt.hist(tmpdf.price)
```

```
[108]: <Axes: ylabel='Frequency'>
```



[]:

[]: