

Lecture 6

May 22, 2023

1 Matplotlib

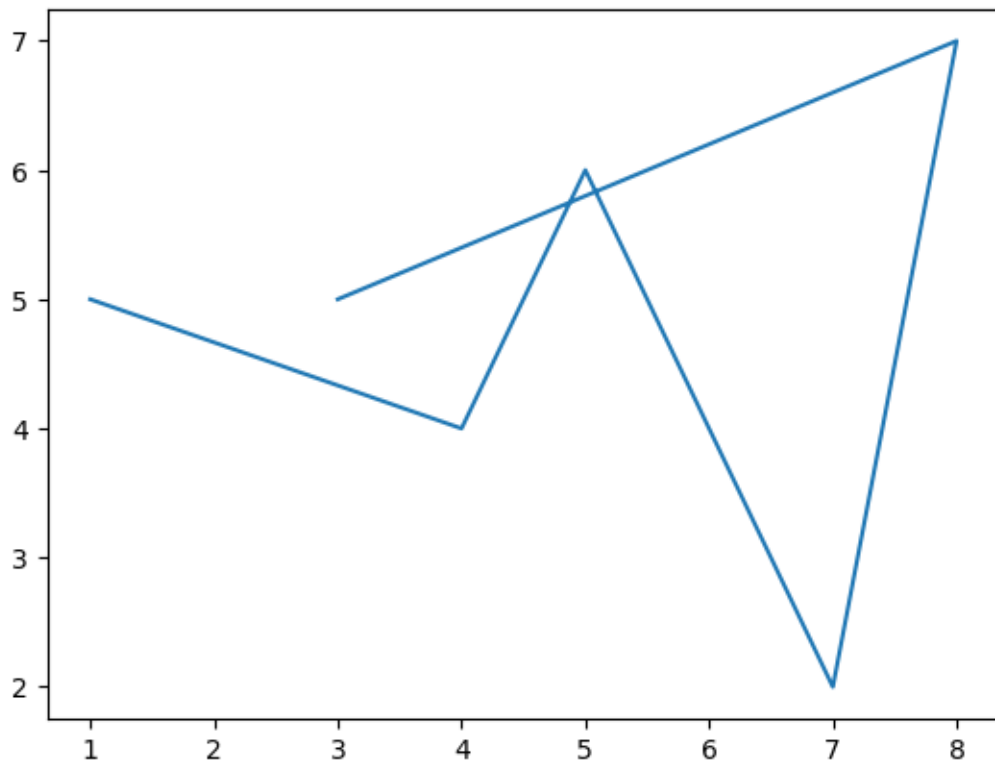
1.1 Import library

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

1.2 Line plot

```
[6]: x = [1,4,5,7,8,3]  
     y = [5,4,6,2,7,5]  
  
     plt.plot(x, y)
```

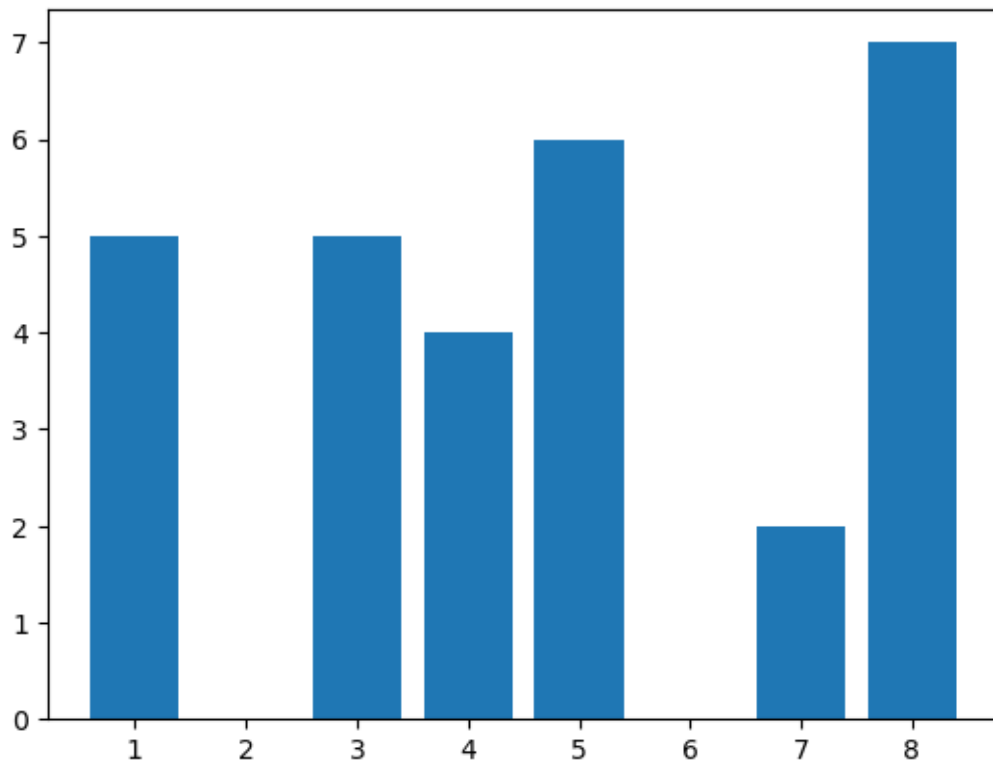
```
[6]: [<matplotlib.lines.Line2D at 0x11faaf250>]
```



1.3 Bar plot

```
[7]: plt.bar(x, y)
```

```
[7]: <BarContainer object of 6 artists>
```

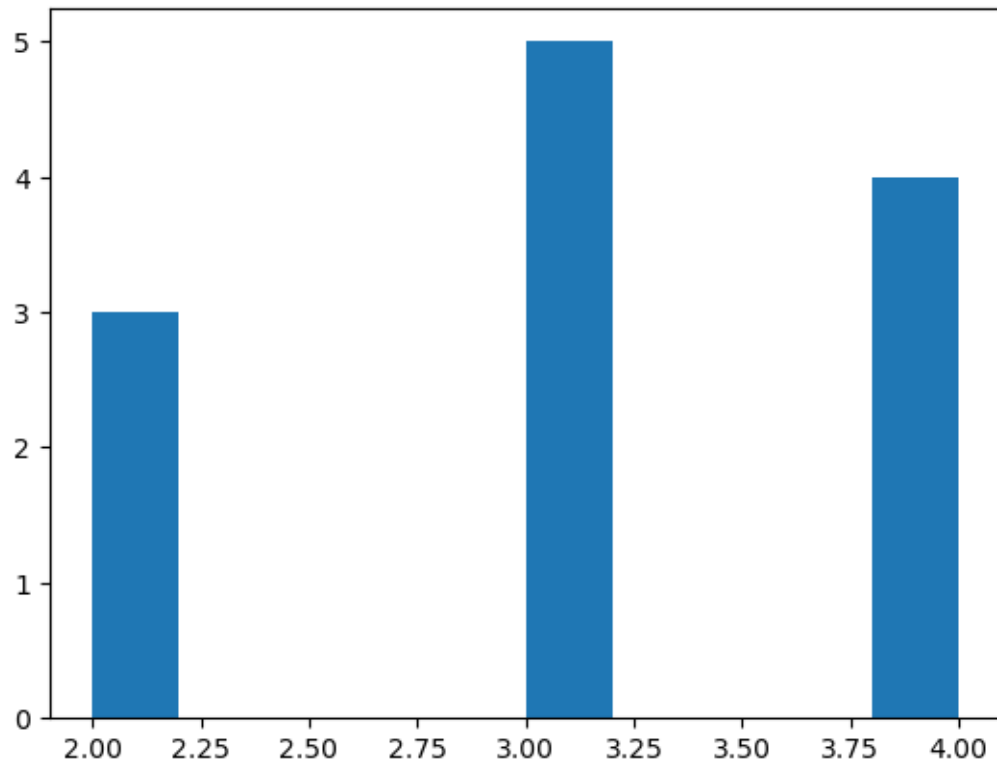


1.4 Histogram

```
[12]: x = [3,4,2,4,4,3,2,4,2,3,3,3]
```

```
plt.hist(x)    #shows frequency distribution
```

```
[12]: (array([3., 0., 0., 0., 0., 5., 0., 0., 0., 4.]),  
      array([2. , 2.2, 2.4, 2.6, 2.8, 3. , 3.2, 3.4, 3.6, 3.8, 4. ]),  
      <BarContainer object of 10 artists>)
```

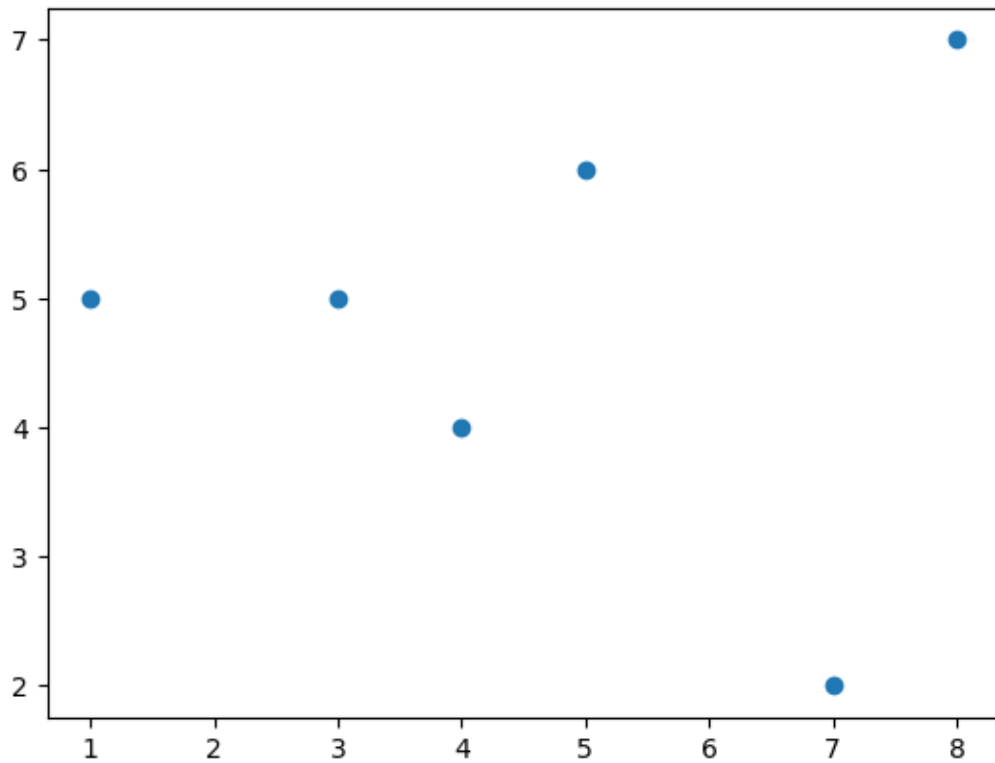


1.5 Scatter plot

```
[13]: x = [1,4,5,7,8,3]
      y = [5,4,6,2,7,5]

      plt.scatter(x, y)
```

```
[13]: <matplotlib.collections.PathCollection at 0x13fc33e50>
```

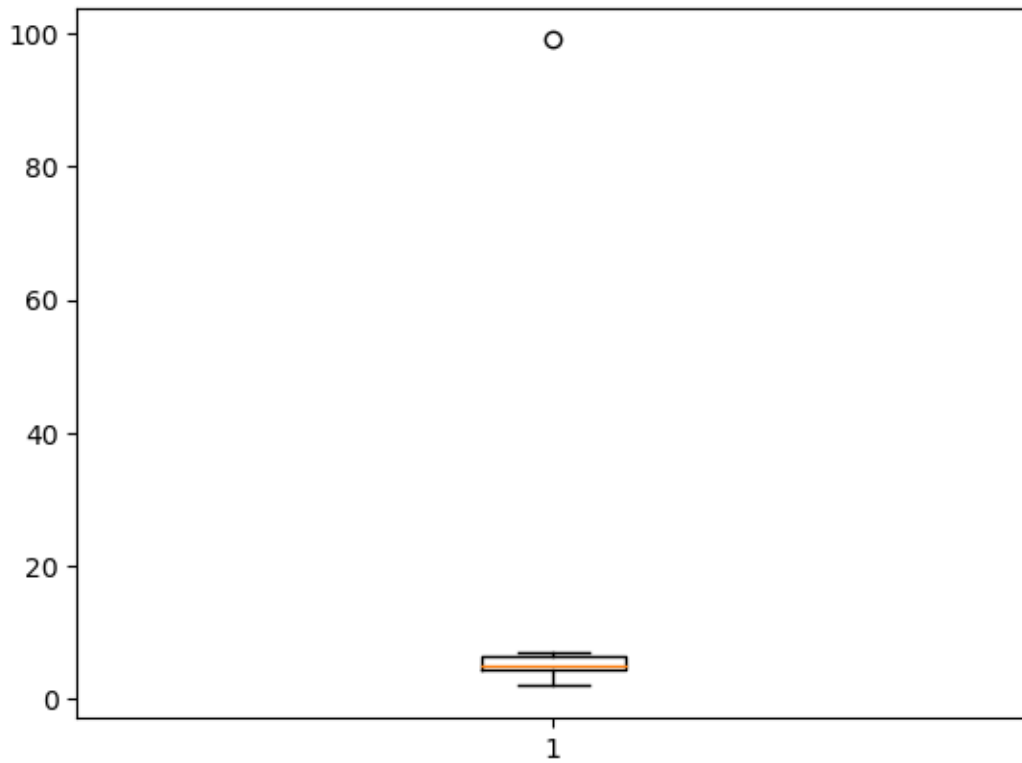


1.6 Boxplot

```
[34]: y = [5,4,6,99, 2,7,5]
```

```
plt.boxplot(y)
```

```
[34]: {'whiskers': [<matplotlib.lines.Line2D at 0x14fa92620>,
<matplotlib.lines.Line2D at 0x14fa928c0>],
'caps': [<matplotlib.lines.Line2D at 0x14ee321a0>,
<matplotlib.lines.Line2D at 0x14fa92bf0>],
'boxes': [<matplotlib.lines.Line2D at 0x14fa92380>],
'medians': [<matplotlib.lines.Line2D at 0x14fa92e90>],
'fliers': [<matplotlib.lines.Line2D at 0x14fa93130>],
'means': []}
```

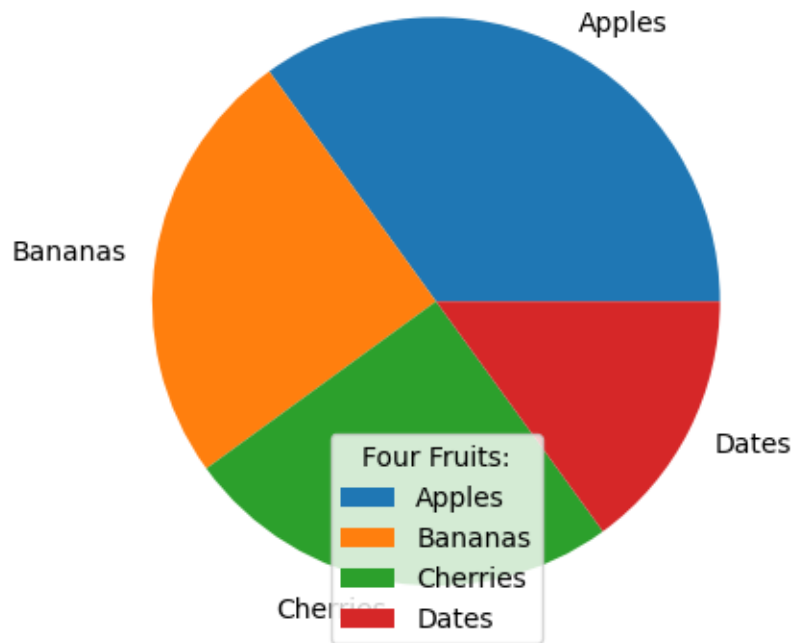


1.7 Pie Chart

```
[47]: y = [35, 25, 25, 15]
mylabels = ["Apples", "Bananas", "Cherries", "Dates"]

plt.pie(y, labels = mylabels)

plt.legend(title = "Four Fruits:")
plt.show()
```



1.8 Subplot

```
[52]: plt.figure(figsize=(10,6))

plt.subplot(2, 2, 1)
x = [1,4,5,7,8,3]
y = [5,4,6,2,7,5]

plt.plot(x, y)

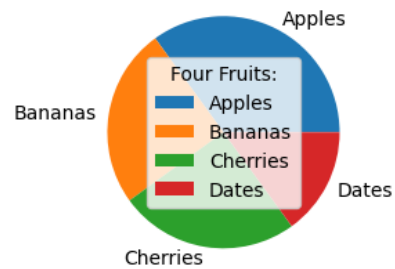
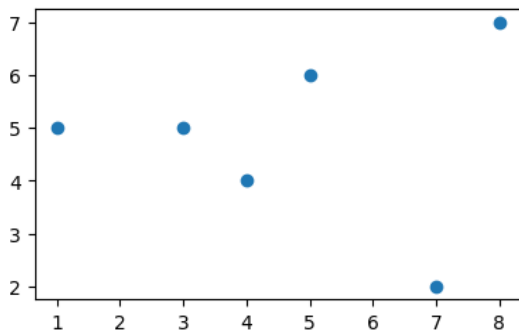
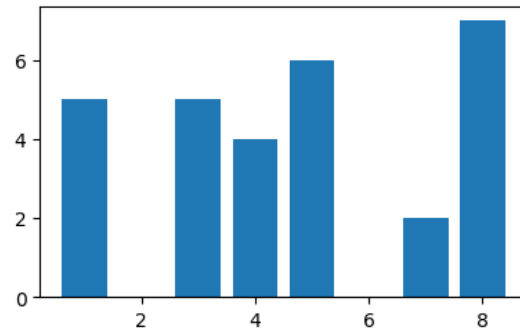
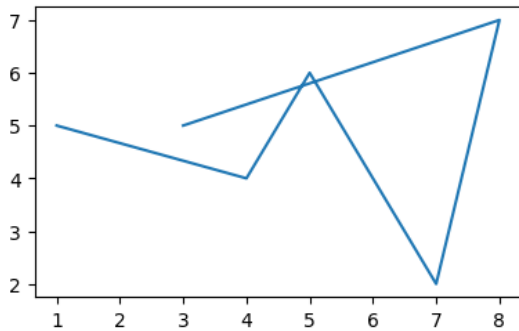
plt.subplot(2, 2, 2)
plt.bar(x, y)

plt.subplot(2, 2, 3)
x = [1,4,5,7,8,3]
y = [5,4,6,2,7,5]

plt.scatter(x, y)

plt.subplot(2, 2, 4)
y = [35, 25, 25, 15]
mylabels = ["Apples", "Bananas", "Cherries", "Dates"]
```

```
plt.pie(y, labels = mylabels)
plt.legend(title = "Four Fruits:")
plt.show()
```



```
[53]: import numpy as np
x = np.array([0, 1, 2, 3])
y = np.array([3, 8, 1, 10])

plt.subplot(2, 3, 1)
plt.plot(x,y)

x = np.array([0, 1, 2, 3])
y = np.array([10, 20, 30, 40])

plt.subplot(2, 3, 2)
plt.plot(x,y)

x = np.array([0, 1, 2, 3])
y = np.array([3, 8, 1, 10])

plt.subplot(2, 3, 3)
plt.plot(x,y)

x = np.array([0, 1, 2, 3])
```

```

y = np.array([10, 20, 30, 40])

plt.subplot(2, 3, 4)
plt.plot(x,y)

x = np.array([0, 1, 2, 3])
y = np.array([3, 8, 1, 10])

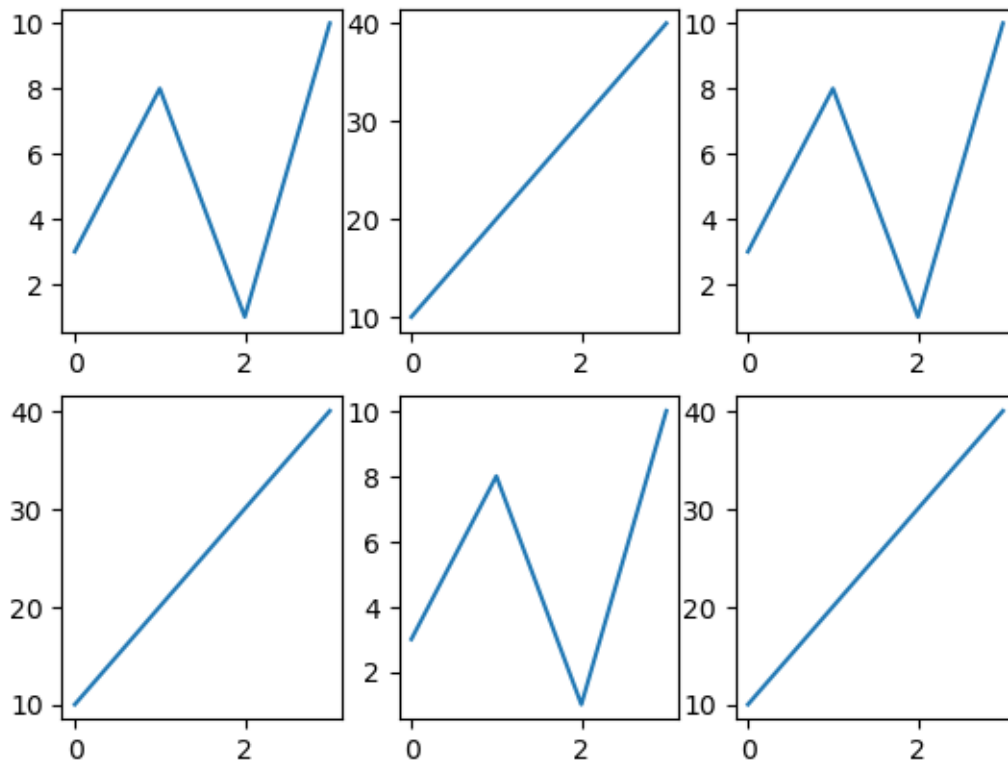
plt.subplot(2, 3, 5)
plt.plot(x,y)

x = np.array([0, 1, 2, 3])
y = np.array([10, 20, 30, 40])

plt.subplot(2, 3, 6)
plt.plot(x,y)

plt.show()

```



```

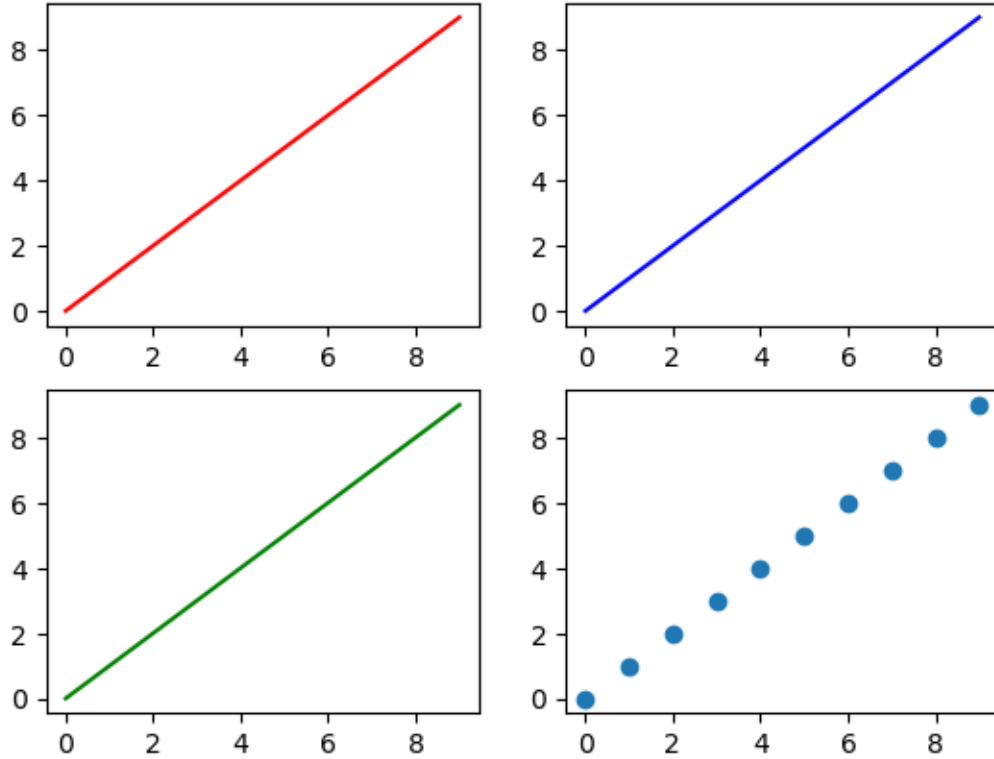
[56]: fig, ax = plt.subplots(2,2)
ax[0, 0].plot(range(10), 'r')
ax[0, 1].plot(range(10), 'b')

```



```
ax[1, 0].plot(range(10), 'g')
ax[1, 1].plot(range(10), 'o')
```

[56]: [<matplotlib.lines.Line2D at 0x169563820>]

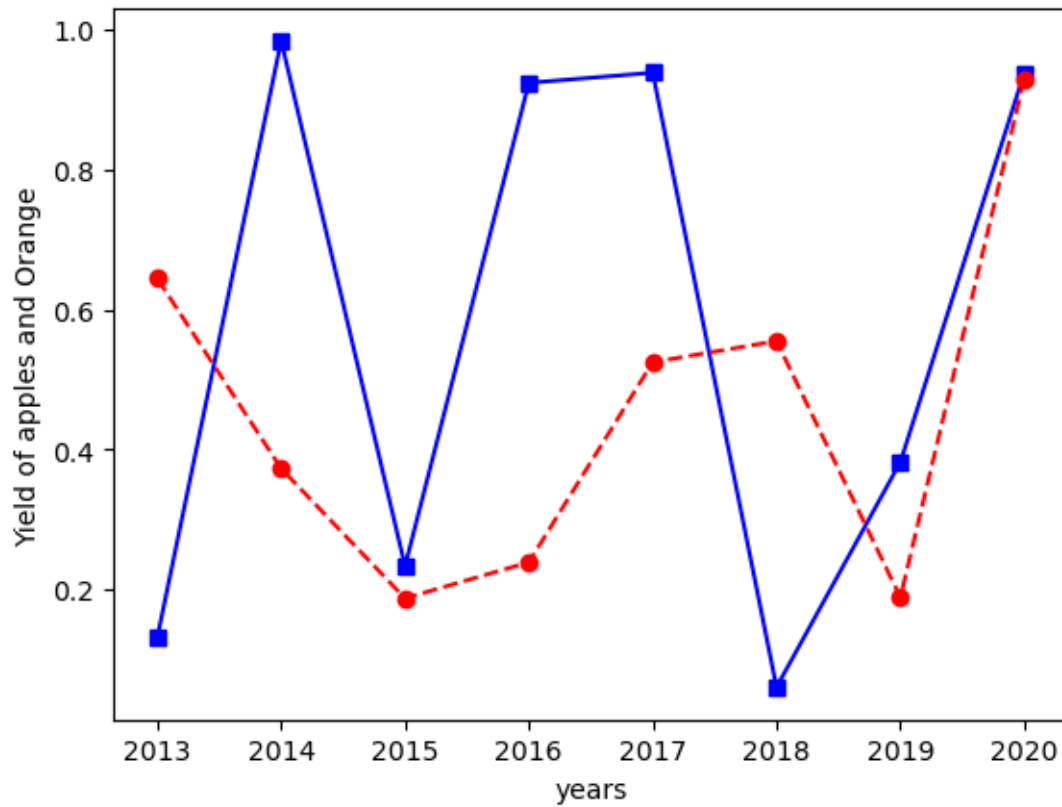


```
[70]: import numpy as np

year = range(2013, 2021)
apple_yield = np.random.rand(8)
orange_yield = np.random.rand(8)

plt.plot(year, apple_yield, 's-b')
plt.plot(year, orange_yield, 'o--r')
plt.xlabel('years')
plt.ylabel('Yield of apples and Orange')
```

[70]: Text(0, 0.5, 'Yield of apples and Orange')

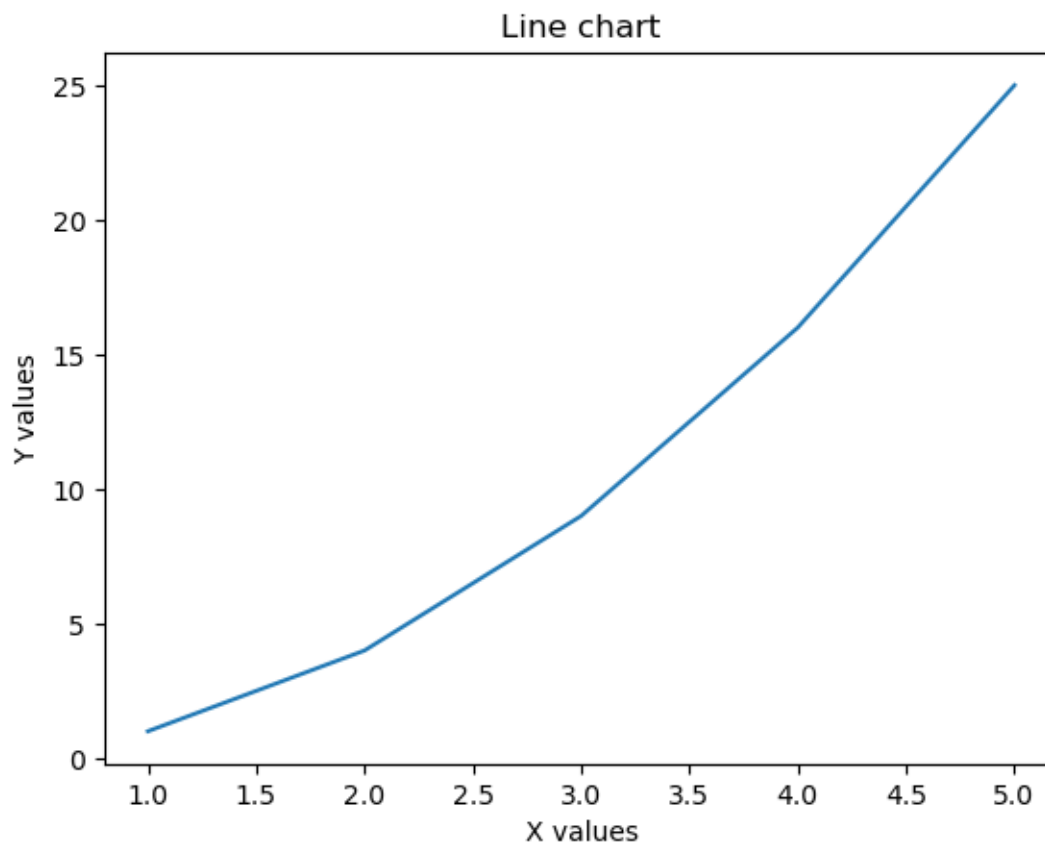


1.9 Other utility functions

```
[15]: x=[1,2,3,4,5]
      y=[1,4,9,16,25]

      plt.plot(x,y)
      plt.xlabel("X values")
      plt.ylabel("Y values")
      plt.title("Line chart")
```

```
[15]: Text(0.5, 1.0, 'Line chart')
```



1.10 Demo with insurance dataset

```
[16]: import pandas as pd
```

```
df = pd.read_csv("insurance.csv")
print(df)
```

| | age | sex | bmi | children | smoker | region | charges |
|------|-----|--------|--------|----------|--------|-----------|-------------|
| 0 | 19 | female | 27.900 | 0 | yes | southwest | 16884.92400 |
| 1 | 18 | male | 33.770 | 1 | no | southeast | 1725.55230 |
| 2 | 28 | male | 33.000 | 3 | no | southeast | 4449.46200 |
| 3 | 33 | male | 22.705 | 0 | no | northwest | 21984.47061 |
| 4 | 32 | male | 28.880 | 0 | no | northwest | 3866.85520 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 1333 | 50 | male | 30.970 | 3 | no | northwest | 10600.54830 |
| 1334 | 18 | female | 31.920 | 0 | no | northeast | 2205.98080 |
| 1335 | 18 | female | 36.850 | 0 | no | southeast | 1629.83350 |
| 1336 | 21 | female | 25.800 | 0 | no | southwest | 2007.94500 |
| 1337 | 61 | female | 29.070 | 0 | yes | northwest | 29141.36030 |

[1338 rows x 7 columns]

```
[20]: df.head()
```

```
[20]:   age    sex    bmi  children  smoker    region    charges
0    19  female  27.900         0     yes  southwest  16884.92400
1    18   male  33.770         1     no   southeast  1725.55230
2    28   male  33.000         3     no   southeast  4449.46200
3    33   male  22.705         0     no  northwest  21984.47061
4    32   male  28.880         0     no  northwest  3866.85520
```

```
[21]: df.tail()
```

```
[21]:   age    sex    bmi  children  smoker    region    charges
1333  50   male  30.97         3     no  northwest  10600.5483
1334  18  female  31.92         0     no  northeast  2205.9808
1335  18  female  36.85         0     no   southeast  1629.8335
1336  21  female  25.80         0     no  southwest  2007.9450
1337  61  female  29.07         0     yes  northwest  29141.3603
```

```
[23]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   age         1338 non-null   int64
1   sex         1338 non-null   object
2   bmi         1338 non-null   float64
3   children    1338 non-null   int64
4   smoker      1338 non-null   object
5   region      1338 non-null   object
6   charges     1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

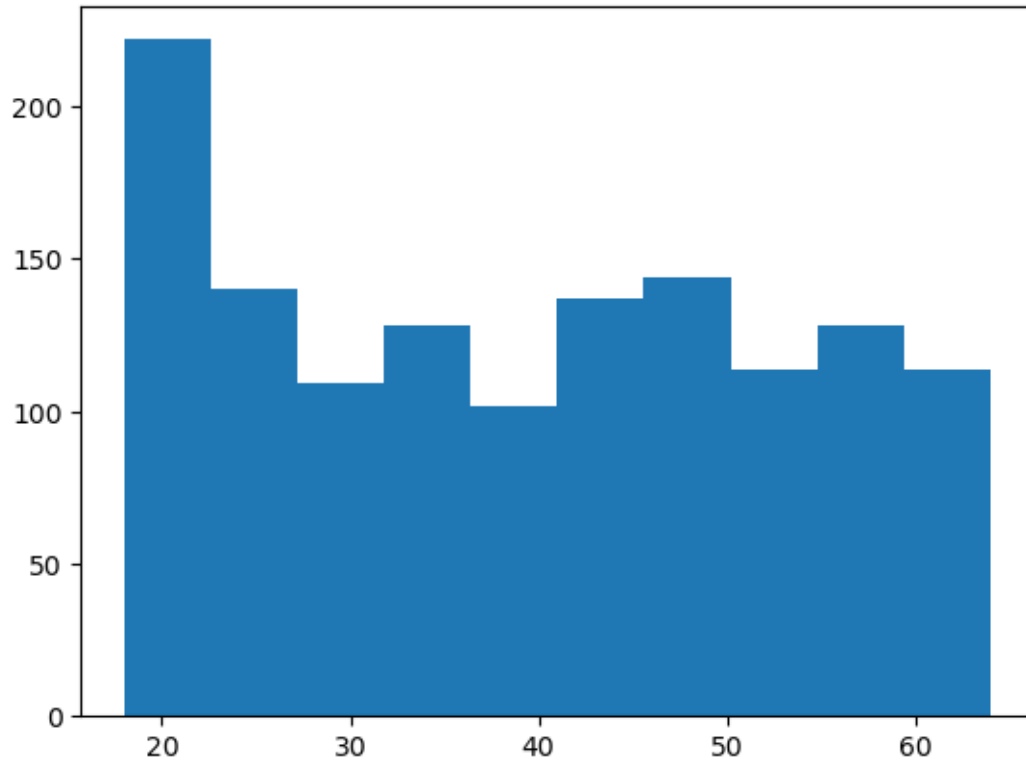
```
[25]: df['age']
```

```
[25]: 0      19
1      18
2      28
3      33
4      32
..
1333   50
1334   18
1335   18
```

```
1336    21
1337    61
Name: age, Length: 1338, dtype: int64
```

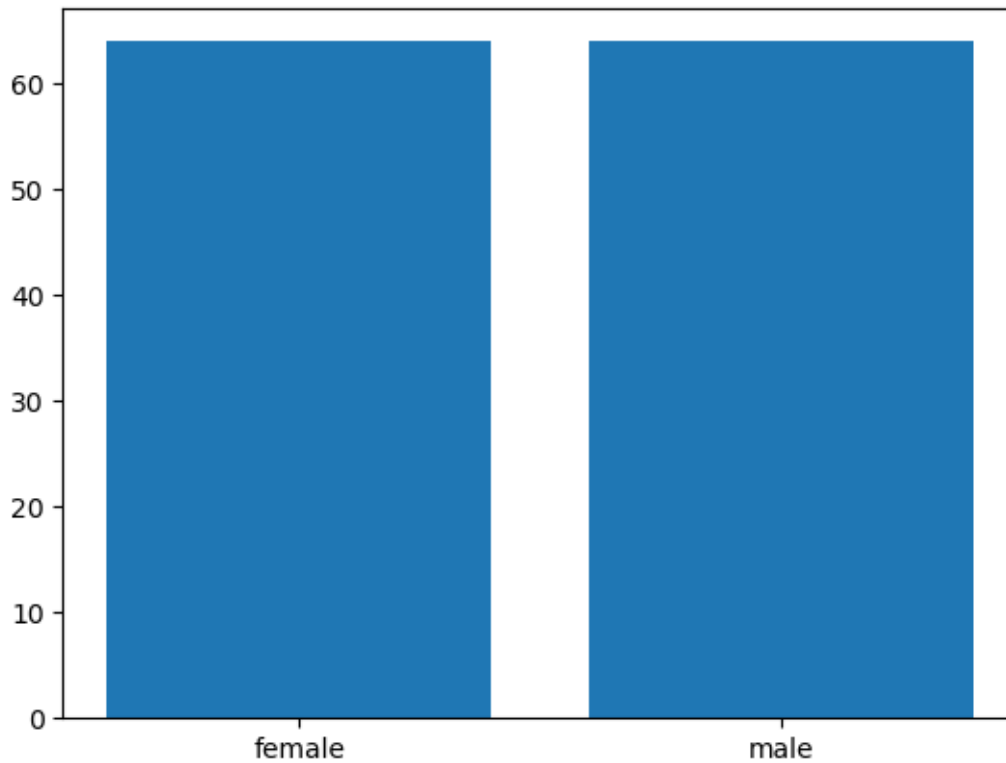
```
[26]: plt.hist(df['age'])
```

```
[26]: (array([222., 140., 109., 128., 102., 137., 144., 114., 128., 114.]),
      array([18. , 22.6, 27.2, 31.8, 36.4, 41. , 45.6, 50.2, 54.8, 59.4, 64. ]),
      <BarContainer object of 10 artists>)
```



```
[28]: plt.bar(df['sex'], df['age'])
```

```
[28]: <BarContainer object of 1338 artists>
```

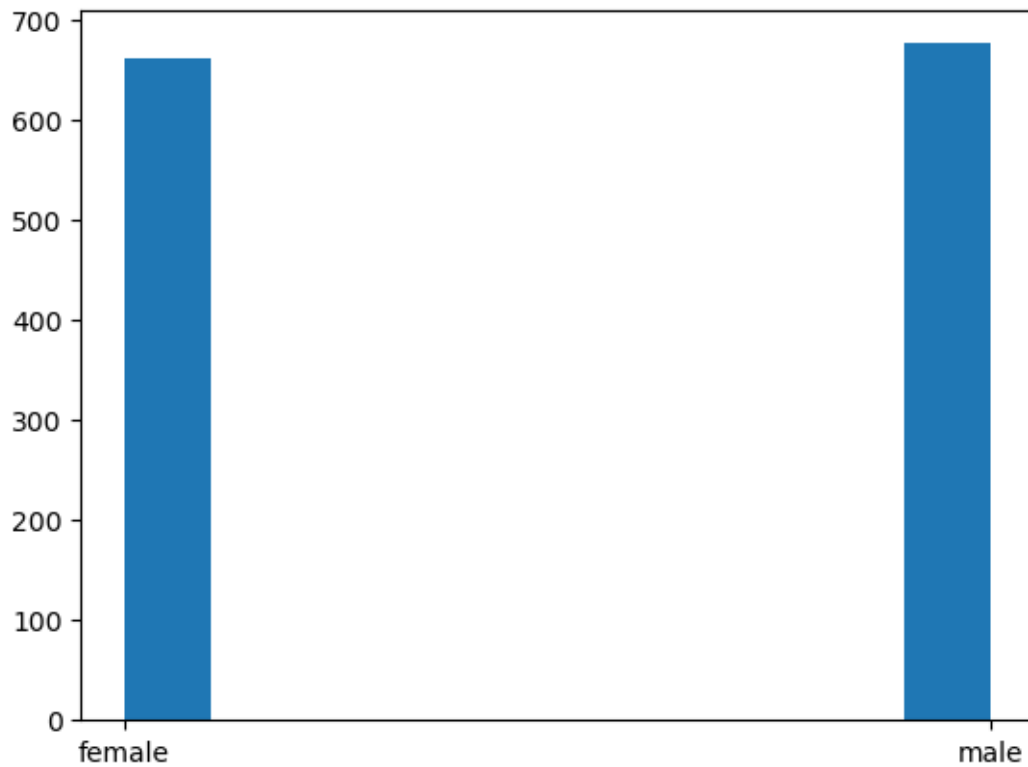


```
[31]: df['sex'].value_counts()
```

```
[31]: male      676  
      female    662  
      Name: sex, dtype: int64
```

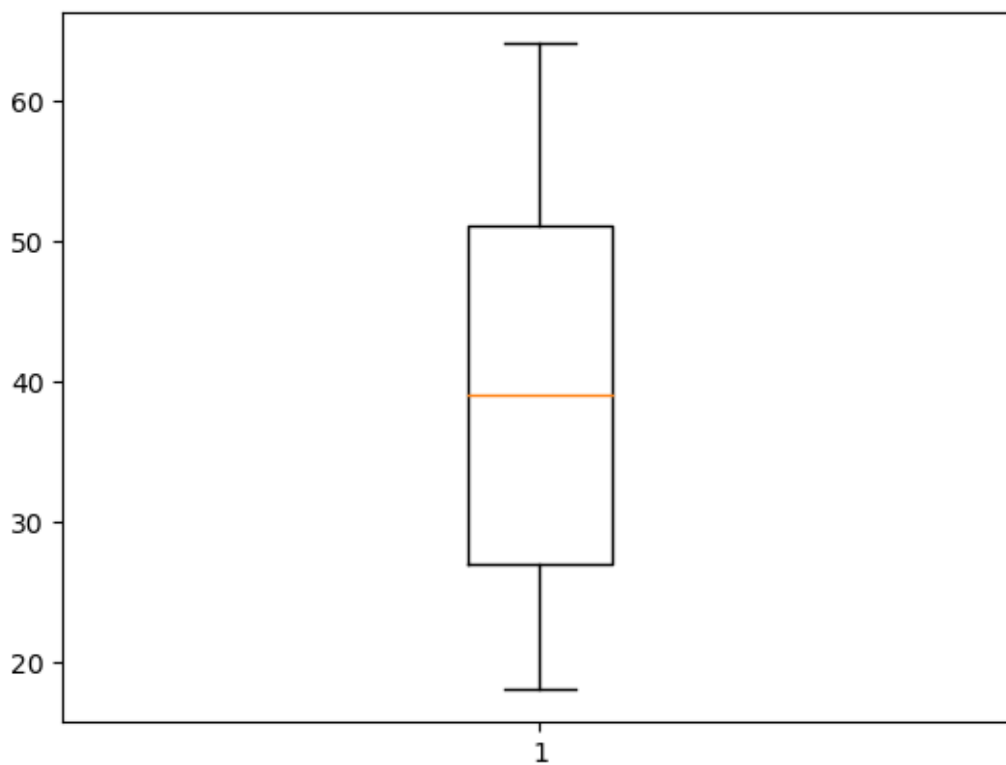
```
[32]: plt.hist(df['sex'])
```

```
[32]: (array([662.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0., 676.]),  
      array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),  
      <BarContainer object of 10 artists>)
```



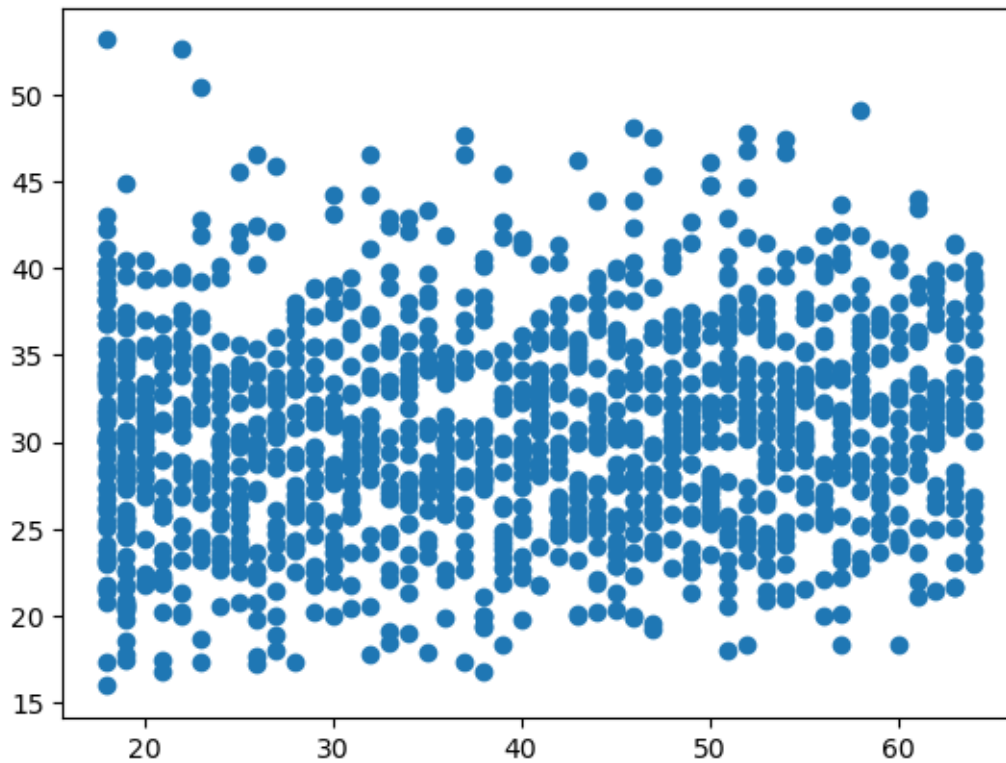
```
[33]: plt.boxplot(df['age'])
```

```
[33]: {'whiskers': [<matplotlib.lines.Line2D at 0x14fa4c1f0>,
<matplotlib.lines.Line2D at 0x14fa4c490>],
'caps': [<matplotlib.lines.Line2D at 0x14fa4c730>,
<matplotlib.lines.Line2D at 0x14fa4c9d0>],
'boxes': [<matplotlib.lines.Line2D at 0x14fa13f40>],
'medians': [<matplotlib.lines.Line2D at 0x14fa4cc70>],
'fliers': [<matplotlib.lines.Line2D at 0x14fa4cf10>],
'means': []}
```



```
[35]: plt.scatter(df['age'], df['bmi'])
```

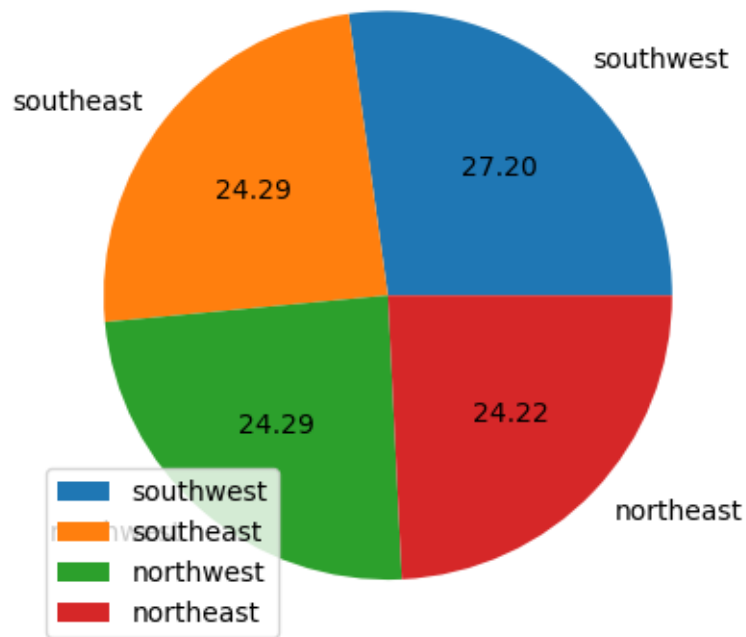
```
[35]: <matplotlib.collections.PathCollection at 0x14fb0cd90>
```

```
[45]: regions = df['region'].unique()

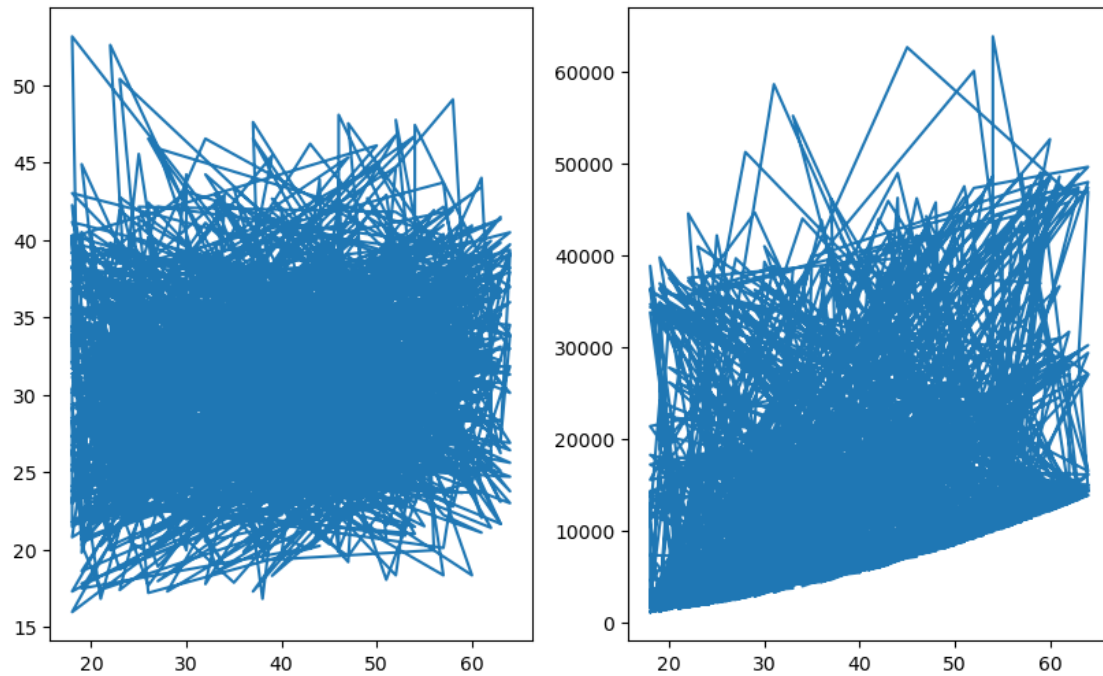
plt.pie(df['region'].value_counts(), autopct="%.2f", labels = regions)
plt.legend()
```

```
[45]: <matplotlib.legend.Legend at 0x168ad2920>
```



```
[48]: plt.figure(figsize=(10,6))  
  
plt.subplot(1, 2, 1)  
plt.plot(df['age'], df['bmi'])  
  
plt.subplot(1, 2, 2)  
plt.plot(df['age'], df['charges'])
```

```
[48]: [<matplotlib.lines.Line2D at 0x14fdaf490>]
```



```
[51]: plt.figure(figsize=(10,6))

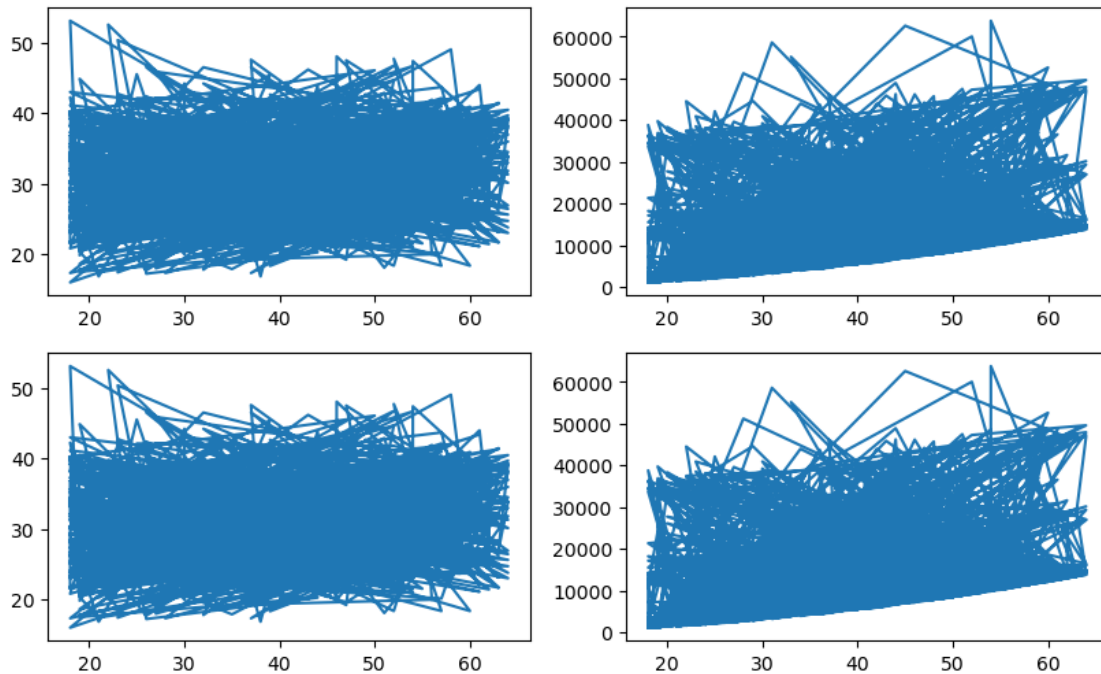
plt.subplot(2, 2, 1)
plt.plot(df['age'], df['bmi'])

plt.subplot(2, 2, 2)
plt.plot(df['age'], df['charges'])

plt.subplot(2, 2, 3)
plt.plot(df['age'], df['bmi'])

plt.subplot(2, 2, 4)
plt.plot(df['age'], df['charges'])
```

```
[51]: [<matplotlib.lines.Line2D at 0x168cdbbe0>]
```



2 Seaborn

2.1 Import library

```
[71]: import seaborn as sns
```

2.2 Get dataset names

```
[73]: sns.get_dataset_names()
```

```
[73]: ['anagrams',  
      'anscombe',  
      'attention',  
      'brain_networks',  
      'car_crashes',  
      'diamonds',  
      'dots',  
      'dowjones',  
      'exercise',  
      'flights',  
      'fmri',  
      'geyser',  
      'glue',  
      'healthexp',
```

```
'iris',
'mpg',
'penguins',
'planets',
'seaice',
'taxis',
'tips',
'titanic']
```

```
[74]: data = sns.load_dataset('tips')
```

```
[75]: data
```

```
[75]:
```

| | total_bill | tip | sex | smoker | day | time | size |
|-----|------------|------|--------|--------|------|--------|------|
| 0 | 16.99 | 1.01 | Female | No | Sun | Dinner | 2 |
| 1 | 10.34 | 1.66 | Male | No | Sun | Dinner | 3 |
| 2 | 21.01 | 3.50 | Male | No | Sun | Dinner | 3 |
| 3 | 23.68 | 3.31 | Male | No | Sun | Dinner | 2 |
| 4 | 24.59 | 3.61 | Female | No | Sun | Dinner | 4 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 239 | 29.03 | 5.92 | Male | No | Sat | Dinner | 3 |
| 240 | 27.18 | 2.00 | Female | Yes | Sat | Dinner | 2 |
| 241 | 22.67 | 2.00 | Male | Yes | Sat | Dinner | 2 |
| 242 | 17.82 | 1.75 | Male | No | Sat | Dinner | 2 |
| 243 | 18.78 | 3.00 | Female | No | Thur | Dinner | 2 |

```
[244 rows x 7 columns]
```

```
[79]: data.shape
```

```
[79]: (244, 7)
```

```
[80]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0  total_bill  244 non-null    float64
 1  tip         244 non-null    float64
 2  sex         244 non-null    category
 3  smoker      244 non-null    category
 4  day         244 non-null    category
 5  time        244 non-null    category
 6  size        244 non-null    int64
dtypes: category(4), float64(2), int64(1)
memory usage: 7.4 KB
```