Reshaping&Pivoting

August 25, 2022

```
[1]: import pandas as pd
     import numpy as np
```

```
[2]: data = pd.DataFrame(np.arange(6).reshape((2, 3)),
             index=pd.Index(['Ohio', 'Colorado'], name='state'),
             columns=pd.Index(['one', 'two', 'three'],
             name='number'))
```

```
[3]: data
```

```
[3]: number    one  two  three
     state
     Ohio        0    1      2
     Colorado    3    4      5
```

```
[4]: result = data.stack()
```

```
[5]: result
```

```
[5]: state     number
     Ohio      one       0
               two       1
               three     2
     Colorado  one       3
               two       4
               three     5
     dtype: int32
```

```
[6]: result.unstack()
```

```
[6]: number    one  two  three
     state
     Ohio        0    1      2
     Colorado    3    4      5
```

```
[8]: result.unstack(1)
```

```
[8]: number    one   two   three
     state
     Ohio        0     1      2
     Colorado    3     4      5
```

```
[9]: result.unstack(0)
```

```
[9]: state    Ohio   Colorado
     number
     one        0         3
     two        1         4
     three      2         5
```

```
[11]: result.unstack('state')
```

```
[11]: state    Ohio   Colorado
      number
      one        0         3
      two        1         4
      three      2         5
```

```
[12]: s1 = pd.Series([0, 1, 2, 3], index=['a', 'b', 'c', 'd'])
```

```
[13]: s2 = pd.Series([4, 5, 6], index=['c', 'd', 'e'])
```

```
[14]: data2 = pd.concat([s1, s2], keys=['one', 'two'])
```

```
[15]: s1
```

```
[15]: a    0
      b    1
      c    2
      d    3
      dtype: int64
```

```
[16]: s2
```

```
[16]: c    4
      d    5
      e    6
      dtype: int64
```

```
[17]: data2
```

```
[17]: one   a    0
            b    1
            c    2
            d    3
```

```
two  c    4
     d    5
     e    6
dtype: int64
```

[18]: `data2.unstack()`

[18]:
```
       a    b    c    d    e
one  0.0  1.0  2.0  3.0  NaN
two  NaN  NaN  4.0  5.0  6.0
```

[19]: `data2.unstack().stack()`

[19]:
```
one  a    0.0
     b    1.0
     c    2.0
     d    3.0
two  c    4.0
     d    5.0
     e    6.0
dtype: float64
```
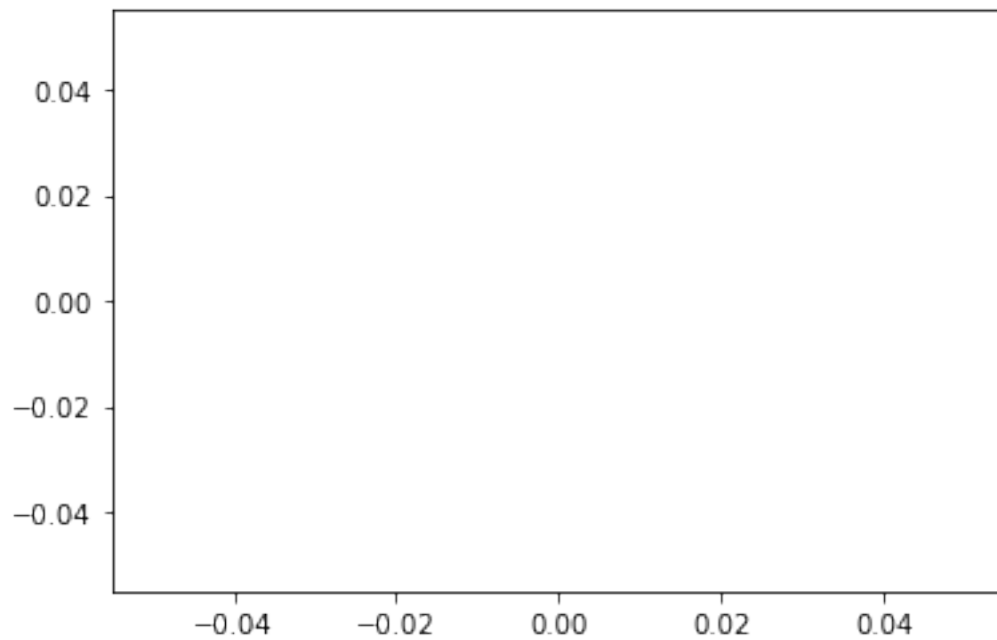
[20]: `data2.unstack().stack(dropna=False)`

[20]:
```
one  a    0.0
     b    1.0
     c    2.0
     d    3.0
     e    NaN
two  a    NaN
     b    NaN
     c    4.0
     d    5.0
     e    6.0
dtype: float64
```

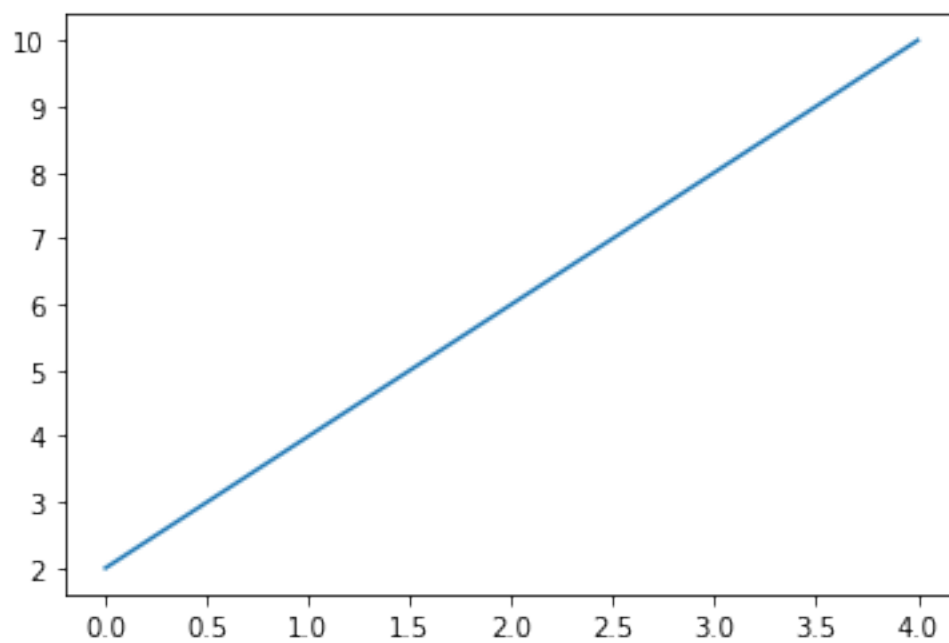# 1 Plotting & Visualization

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

[46]: `plt.plot()`

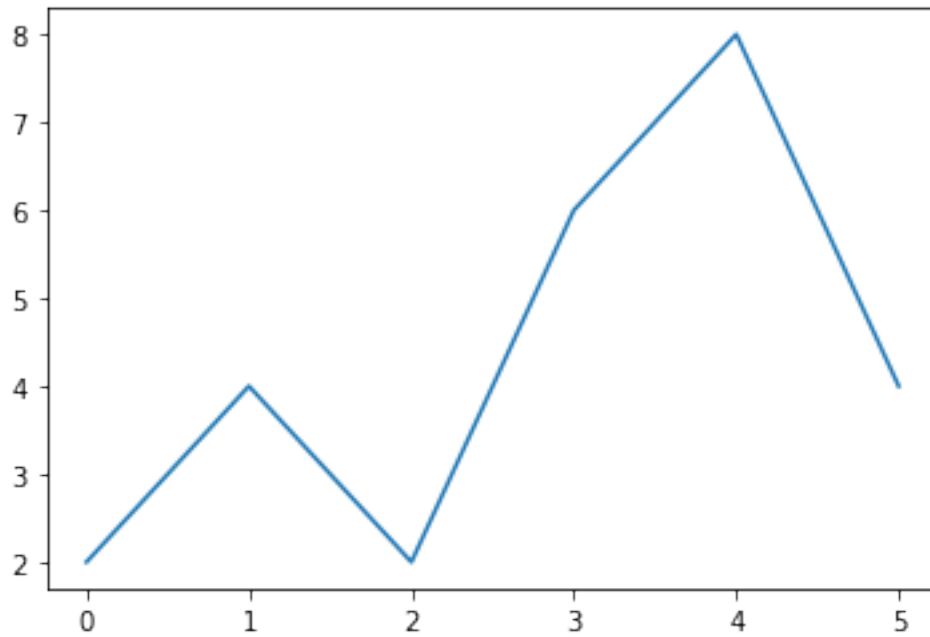[46]: `[]`

```
[48]: plt.plot([2,4,6,8, 10])
```

[48]: [<matplotlib.lines.Line2D at 0x21efe4fb2b0>]
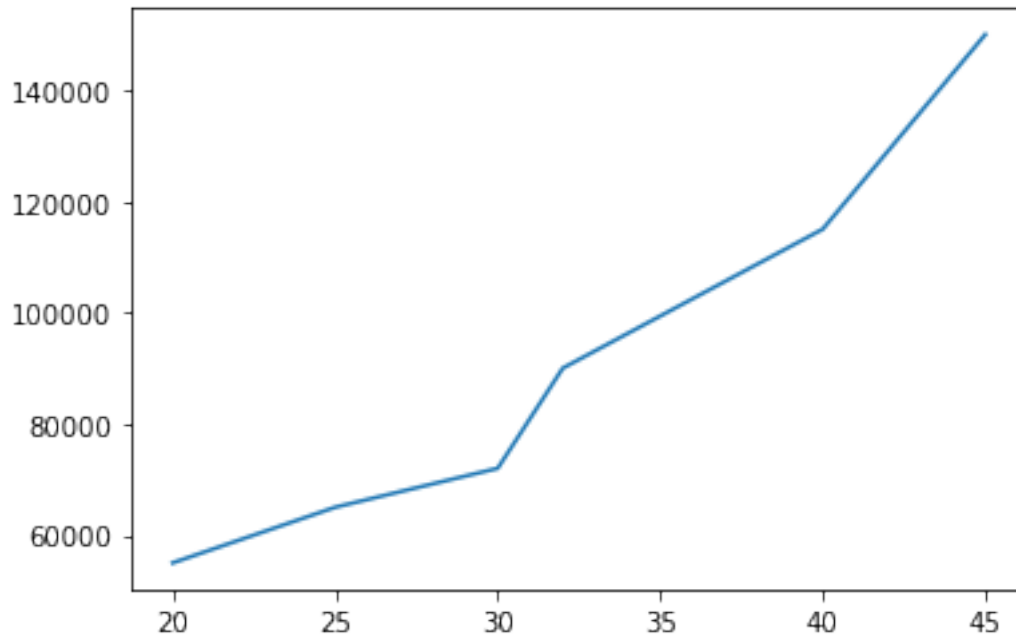
[49]: `plt.plot([2,4,2,6,8,4])`

[49]: `[<matplotlib.lines.Line2D at 0x21efe53baf0>]`
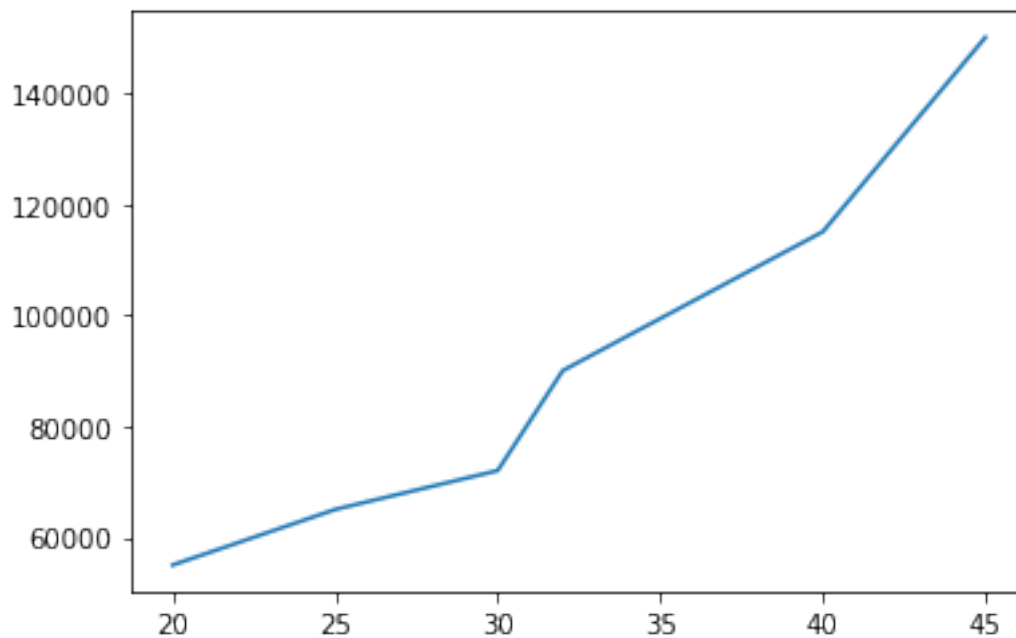

[50]:
```
salaries=[55000,65000,72000,90000,115000,150000]
ages = [20,25,30,32,40,45]
plt.plot(ages, salaries)
```

[50]: `[<matplotlib.lines.Line2D at 0x21efe583d00>]`
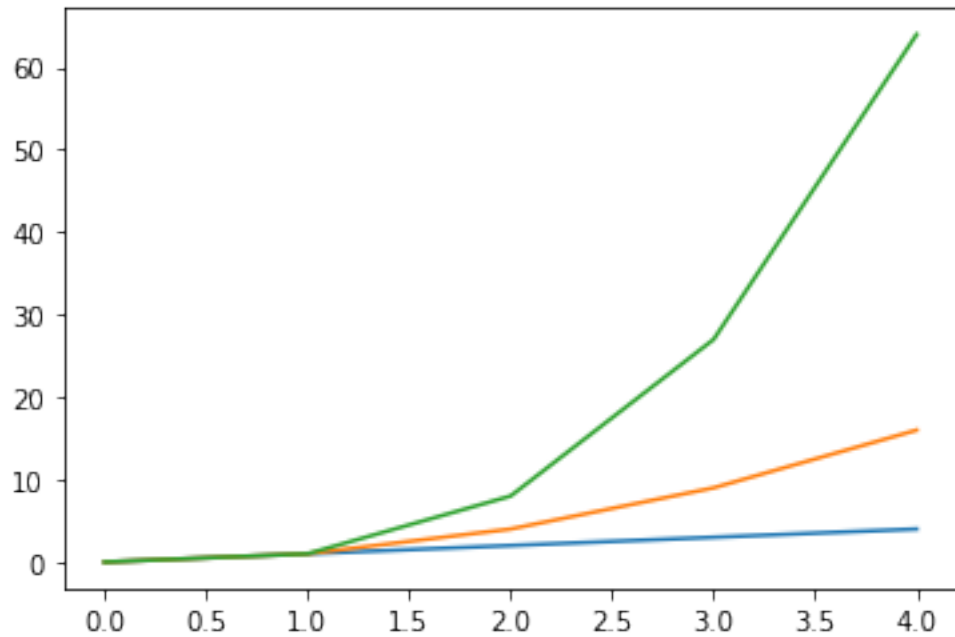
```
[52]: plt.plot(ages, salaries)
      plt.show()
```

```python
[53]: import numpy as np
      nums = np.arange(5)
```
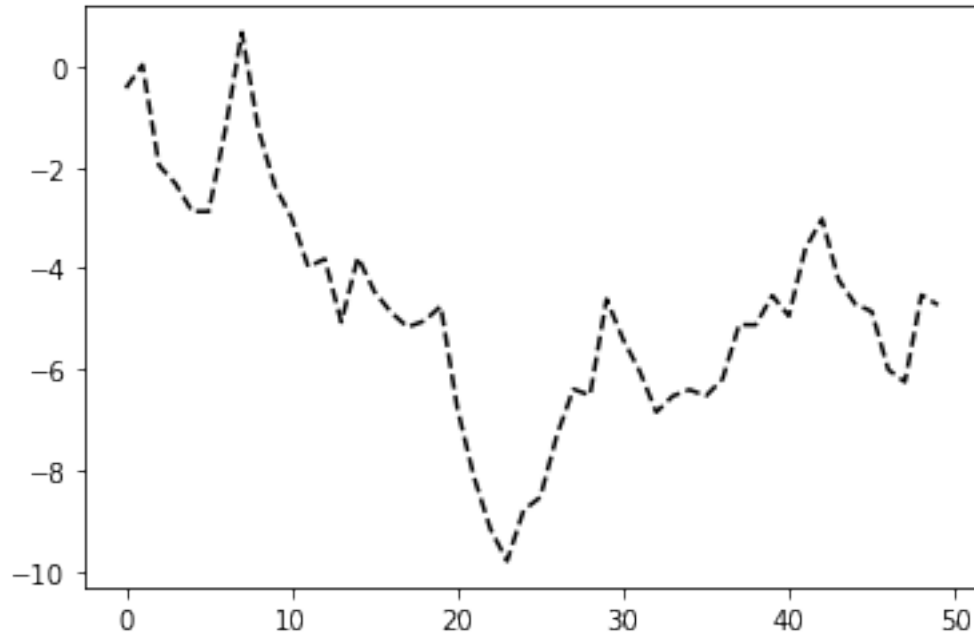
```python
[54]: plt.plot(nums,nums)
      plt.plot(nums, nums*nums)
      plt.plot(nums, nums**3)
```

[54]: [<matplotlib.lines.Line2D at 0x21efe68f760>]


```python
[55]: plt.plot(np.random.randn(50).cumsum(), 'k--')
```

[55]: [<matplotlib.lines.Line2D at 0x21efe6f2400>]

```
[56]: np.random.randn(50).cumsum()
```

```
[56]: array([ 0.88251112,  1.1210313 ,  0.03974723, -1.86491072, -1.17103473,
              0.80913692,  1.51076403,  1.30940667,  1.80332146,  2.78180673,
              4.0406618 ,  5.37580282,  6.11844244,  6.98768143,  6.09469879,
              7.43883649,  7.67368032,  6.9687057 ,  7.30581699,  7.06613848,
              6.59127327,  7.55227742,  8.15717564,  7.23913488,  6.95700623,
              7.54760751,  8.01929714,  6.20325696,  6.76004443,  5.56604273,
              5.03219262,  5.71842279,  3.2446861 ,  2.74820435,  2.33606293,
              1.70532233,  3.15349805,  1.8900307 ,  1.98474385,  0.56006597,
             -1.18473039, -1.18866348, -0.47746253,  0.41938426, -2.01158256,
             -1.08887392, -1.9116556 , -0.54090087, -1.68788468, -2.87313638])
```

```
[67]: fig = plt.figure()
      ax1 = fig.add_subplot(2, 2, 1)
      ax2 = fig.add_subplot(2, 2, 2)
      ax3 = fig.add_subplot(2, 2, 3)
      _ = ax1.hist(np.random.randn(100), bins=20, color='k', alpha=0.3)
      ax2.scatter(np.arange(30), np.arange(30) + 3 * np.random.randn(30))
      np.random.randn(50).cumsum()
```

```
[67]: array([-0.29358627,  1.80977305,  2.00111818,  1.91975405,  2.03331246,
              2.9622707 ,  3.04647658,  0.99404442,  1.06816904,  0.42493811,
              0.7290447 ,  0.0625306 ,  0.86747441,  1.07530607,  0.5453821 ,
              0.75799085,  2.27404518,  2.32872341,  0.48568792,  1.30659118,
              0.02474347, -0.57845309, -0.71932459, -1.76922755, -0.75785526,
```

8

```
       -0.2901235 ,  0.36715602,  0.00593231,  0.83090033,  0.42710711,
       -0.43478203,  0.86950238, -0.08221374,  1.15858513,  1.85256663,
        1.07992835,  0.79808668,  0.33673953, -0.75945734, -0.40395177,
        2.05744212,  0.01561595,  1.09210502,  0.55121358,  1.45293154,
       -0.04224281, -0.15805931,  1.1141645 ,  1.75120267,  1.66711956])
```
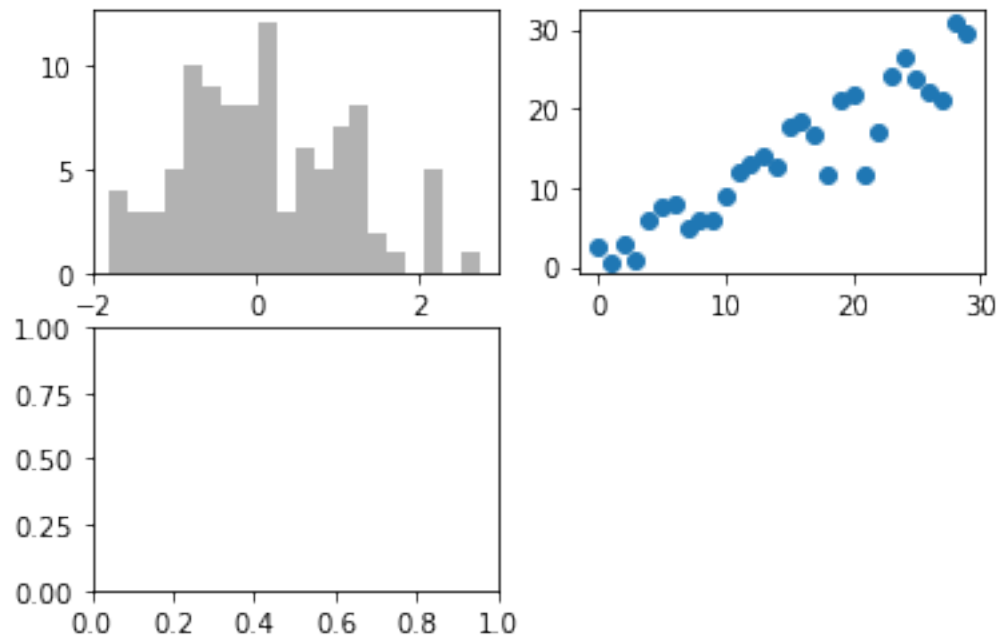

[65]:
```
_ = ax1.hist(np.random.randn(100), bins=20, color='k', alpha=0.3)
ax2.scatter(np.arange(30), np.arange(30) + 3 * np.random.randn(30))
```

[65]: <matplotlib.collections.PathCollection at 0x21efe4e1220>

[60]:

[60]: <matplotlib.collections.PathCollection at 0x21eff80c610>

[63]:
```
ax2
```

[63]: <AxesSubplot:>

[ ]: