

レシピ93 pandasのプロットの基本 (p326)

- pandasのDataFrame, Seriesのプロットは内部的にはmatplotlibで行われる
- plot メソッドでアクセスできる
 - 重要 :
 - ここでの plot は matplotlib.pyplot の plt ではなく、pandasの plot
 - たぶん df.plot(kind=~) の形？そのことか？
- pandasでは、線グラフ、棒グラフ、箱ひげ図、散布図、カーネル密度推定（KDE）、ヒストグラムといった
- matplotlibの一部のプロットしかできない ←一部とは？
 - 機能が限定されるということか、ではdf.plot,s.plot以外のやり方とは？
- pandasでプロットするメリットは、1行ほどで簡単かつ効率的にEDAできること

- pandasのプロットを理解のコツは、メソッドでプロットするのに必要な**変数が1または2つかを把握する**こと
- 例えば、線グラフや散布図、棒グラフには2変数が必要。箱ひげ図やヒストグラム、カーネル密度推定は1変数
- 2変数が必要なグラフでは、デフォルトでインデックスをX軸、カラムの値をY軸にとる
- 1変数では、インデックスを無視して、各変数に変換または集約を行いプロットする
- このレシピでは、pandasにおける2変数と1変数のプロットの違いを確認する

(1) DF作成

```
In [1]: import pandas as pd
df = pd.DataFrame(index=['Atiya', 'Abbas', 'Cornelia', 'Stephanie', 'Monte'],
                  data = {'Apples': [20,10,40,20,50],
                          'Orange': [35,40,25,19,33]})
```

In [2]:

df

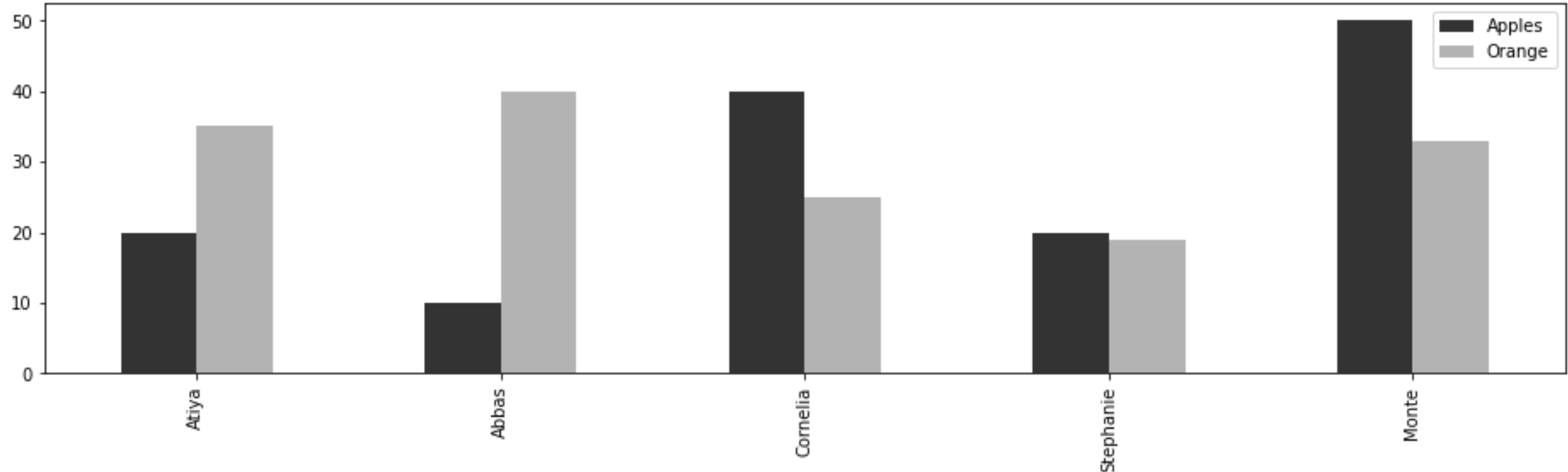
Out[2]:

	Apples	Orange
Atiya	20	35
Abbas	10	40
Cornelia	40	25
Stephanie	20	19
Monte	50	33

(2) 棒グラフ

- ラベル（カラム）のインデックスをX軸、カラム要素値をY軸の高さとする
- kindパラメータはbar

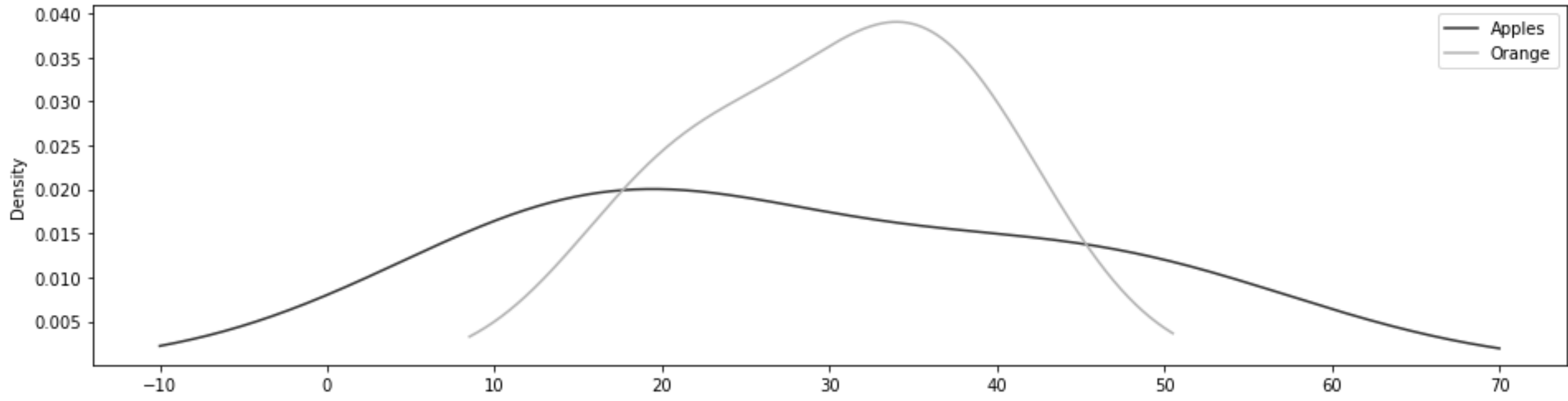
```
In [3]: color = ['.2', '.7']
df.plot(kind='bar', color=color, figsize=(16,4));
```



(3) KDE

- KDEプロットはインデックスを無視して各カラム値をX軸、確率密度をY値に取る

```
In [5]: df.plot(kind='kde', color=color, figsize=(16,4));
```



(4)

- 2変数プロットすべてを1つのFigureに作成する
- 散布図だけがx値y値のカラムを指定する必要がある
- 散布図にインデックスを使いたい場合は、reset_indexを使いカラムにする
- 他の2つのプロットでは、インデックスをX軸に使い、数値カラムで線・棒を引く

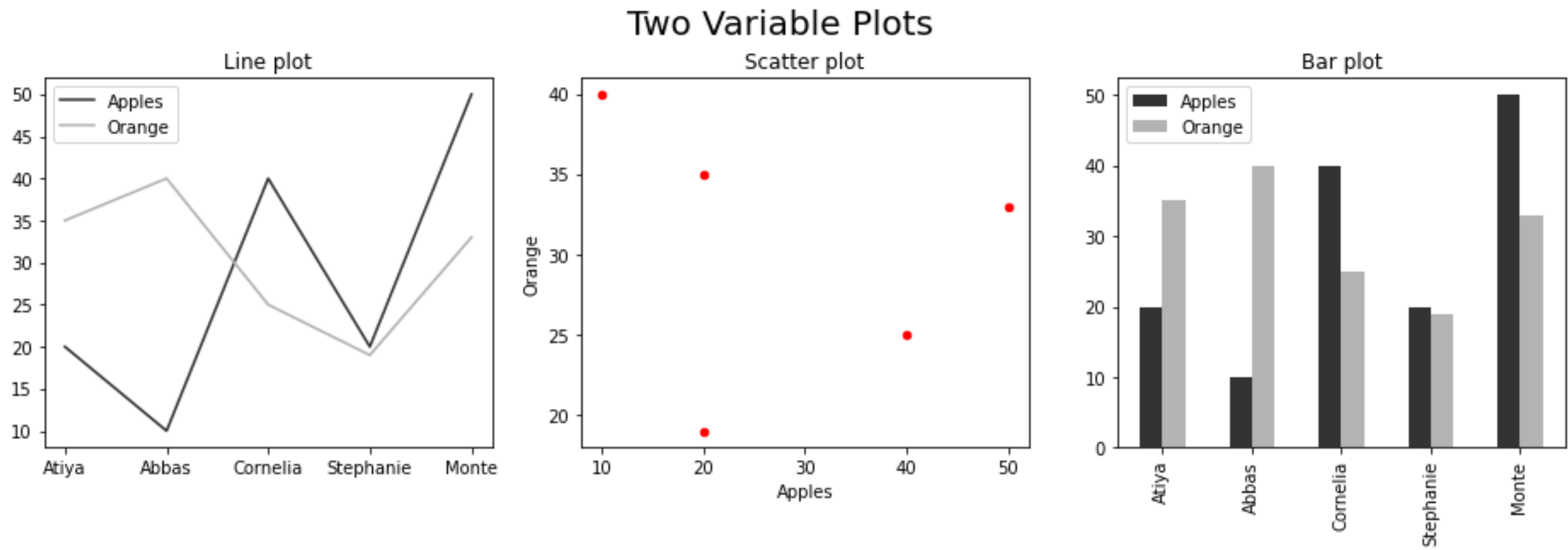
→散布図を除き、プロットではカラムを指定しなくてよい（もちろん指定も可能）

```
In [16]: import matplotlib.pyplot as plt

# 下部の各グラフにplotする際、ax=ax*としている
fig, (ax1,ax2,ax3) = plt.subplots(1,3,figsize=(16,4))

fig.suptitle('Two Variable Plots', size=20, y=1.02)

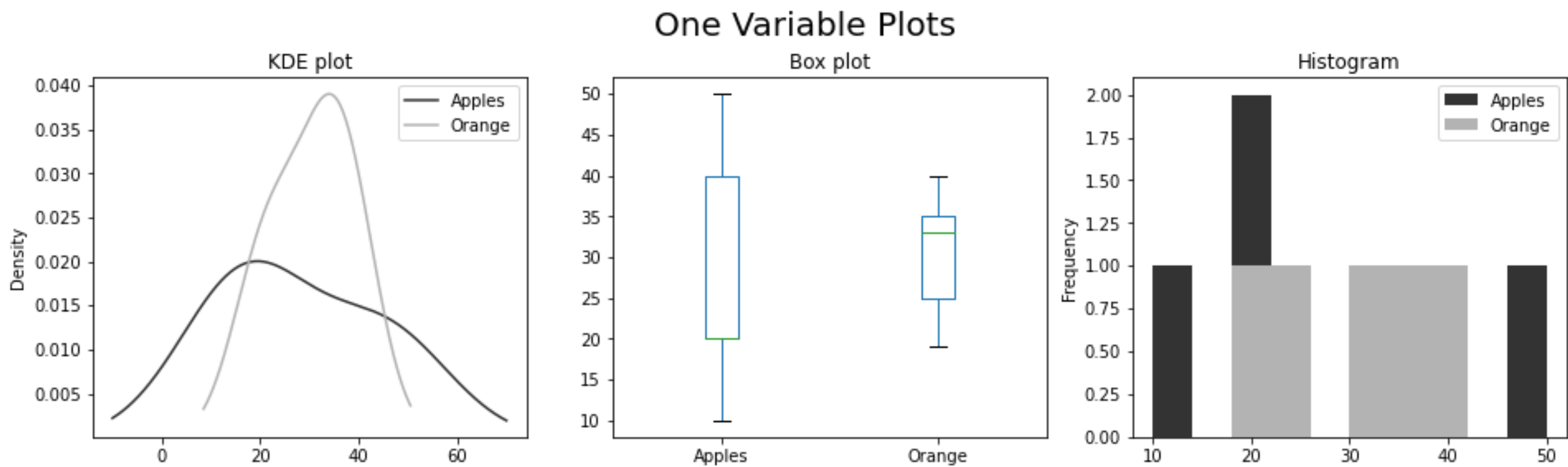
# bar
df.plot(kind='line', color=color, ax=ax1, title='Line plot')
# scatter (散布図)
df.plot(x='Apples', y='Orange', kind='scatter', c='red', ax=ax2, title='Scatter plot')
# bar
df.plot(kind='bar', color=color, ax=ax3, title='Bar plot');
```



(5)

- すべての1変数プロットも同じFigureにする

```
In [14]: fig, (ax1,ax2,ax3) = plt.subplots(1,3,figsize=(16,4))
fig.suptitle('One Variable Plots', size=20, y=1.02)
df.plot(kind='kde', color=color, ax=ax1, title='KDE plot')
df.plot(kind='box', ax=ax2, title='Box plot')
df.plot(kind='hist', color=color, ax=ax3, title='Histogram');
```



参照

- 公式・可視化：https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html