Plotting and Visualization

Imports

- Matplotlib
- Seaborn

In [42]:

```
import warnings
warnings.filterwarnings('ignore')
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
plt.rcParams["figure.figsize"] = [15,5]
```

Example Dataset

In [3]:

```
cafe_df = pd.read_csv('data/cafe.csv')
cafe_df.head(10)
```

Out[3]:

	date	temperature	sold_icecream	sold_cups_coffee	sold_coke
0	2018-06-29	28	40	57	44
1	2018-06-30	25	36	61	19
2	2018-07-01	31	45	53	15
3	2018-07-02	31	47	52	26
4	2018-07-03	29	45	50	23
5	2018-07-04	29	44	55	42
6	2018-07-05	28	42	56	22
7	2018-07-06	27	40	58	31
8	2018-07-07	22	32	63	26
9	2018-07-08	24	35	61	19

Matplotlib Pyplot

- Pyplot is a Matplotlib module which provides a MATLAB-like API
- It can handle different data types:
  - Standard Python arrays
  - Numpy arrays
  - Pandas Series and DataFrames
- Gallery: https://matplotlib.org/gallery/index.html (https://matplotlib.org/gallery/index.html)

Matplotlib Example 1

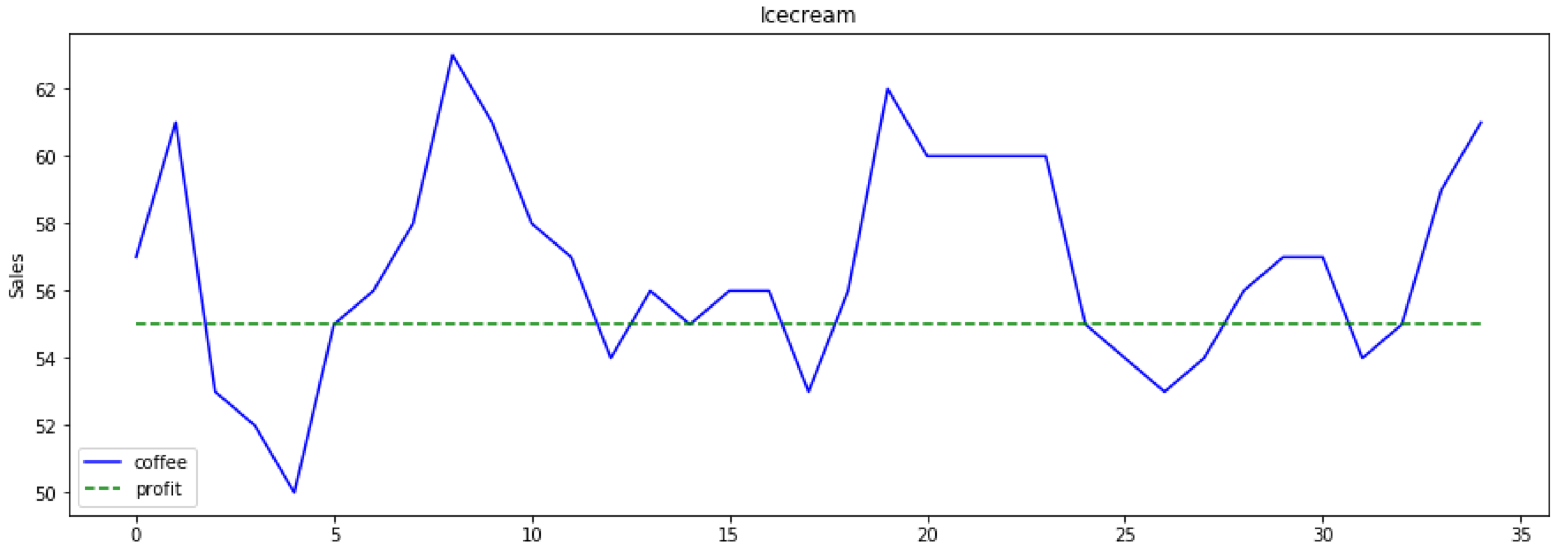
In [56]:

```
plt.title("Icecream")

plt.ylabel("Sales")

plt.plot(cafe_df['sold_cups_coffee'], color='blue', label='coffee')
plt.plot([55] * len(cafe_df.index), '--', color='green', label='profit')

plt.legend(loc='lower left');
```

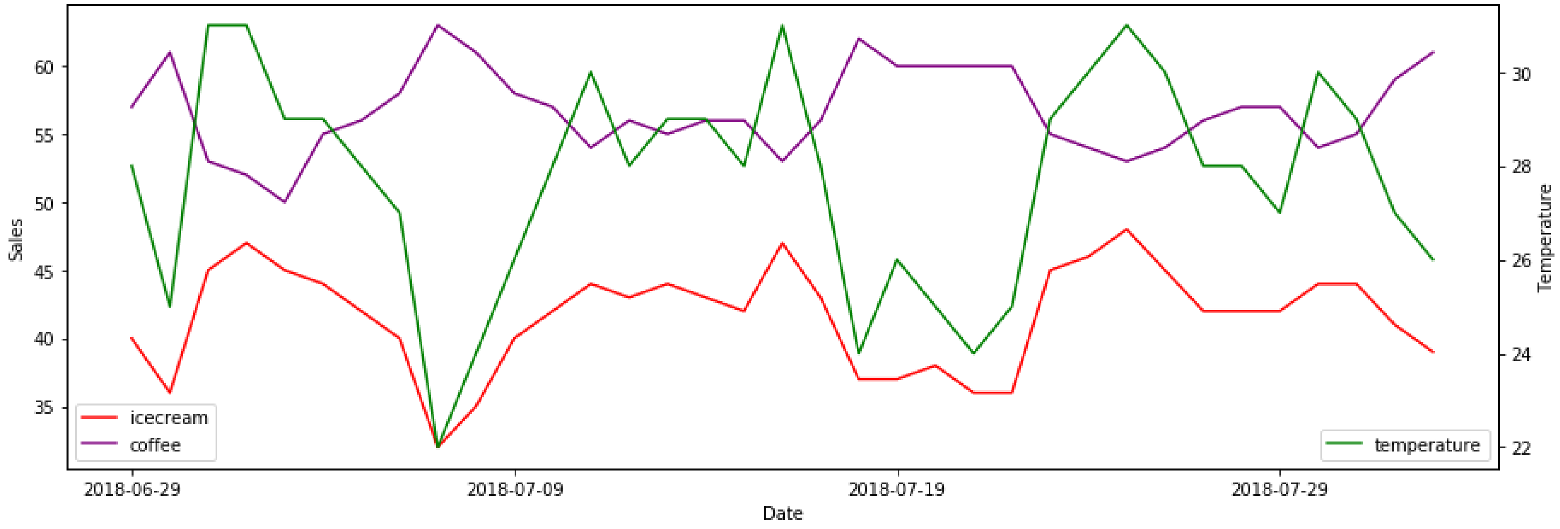


Matplotlib Example 2

In [68]:

```
plt.xlabel("Date")
plt.ylabel("Sales")
plt.plot(cafe_df['date'], cafe_df['sold_icecream'], color='red', label='icecrea
m')
plt.plot(cafe_df['date'], cafe_df['sold_cups_coffee'], color='purple', label='co
ffee')
plt.legend(loc='lower left')

ax2 = plt.twinx()
ax2.plot(cafe_df['date'], cafe_df['temperature'], color='green', label='temperat
ure')
ax2.set_ylabel('Temperature')
ax2.legend(loc='lower right');
plt.xticks([0,10,20,30]);
```



Plot with Pandas

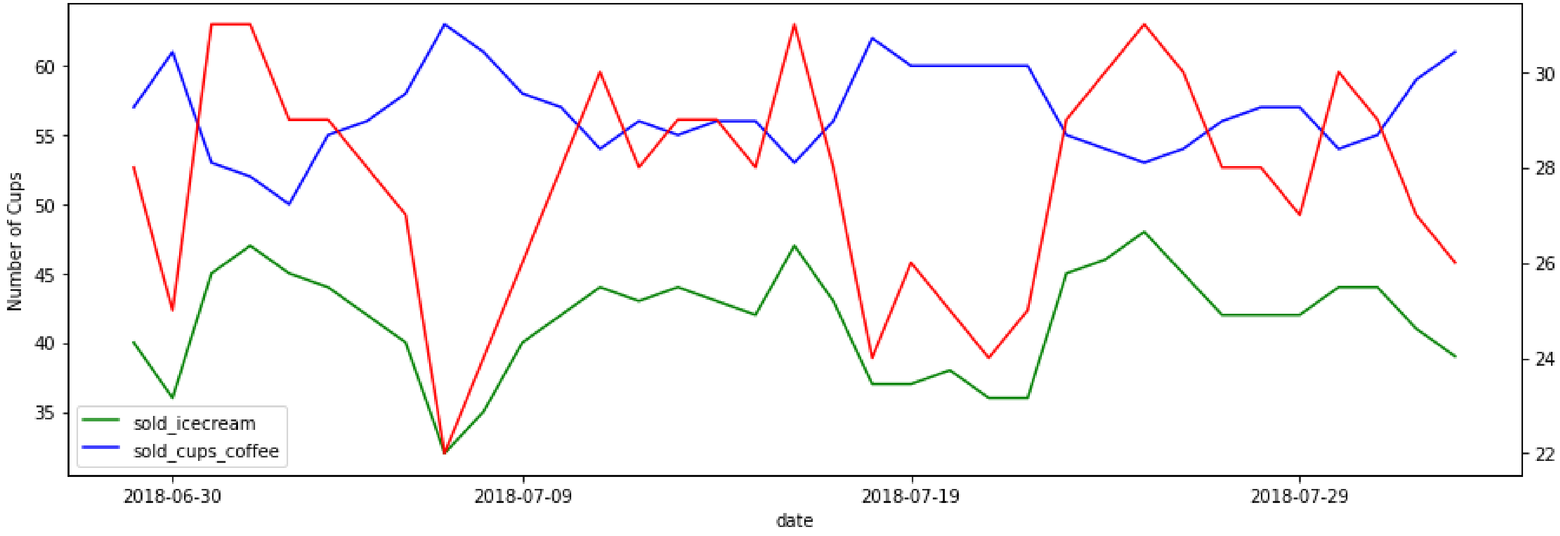
- The Series and DataFrame classes provide convenience methods to plot via Matplotlib
- By default the index is used for x axis
- Two ways:
    - `df.plot(kind='bar')`
    - `df.plot.bar()`
- Documentation: https://pandas.pydata.org/pandas-docs/stable/visualization.html (https://pandas.pydata.org/pandas-docs/stable/visualization.html)

Pandas Example

- Similar plot as before
- With less code, but not all details configured

In [51]:

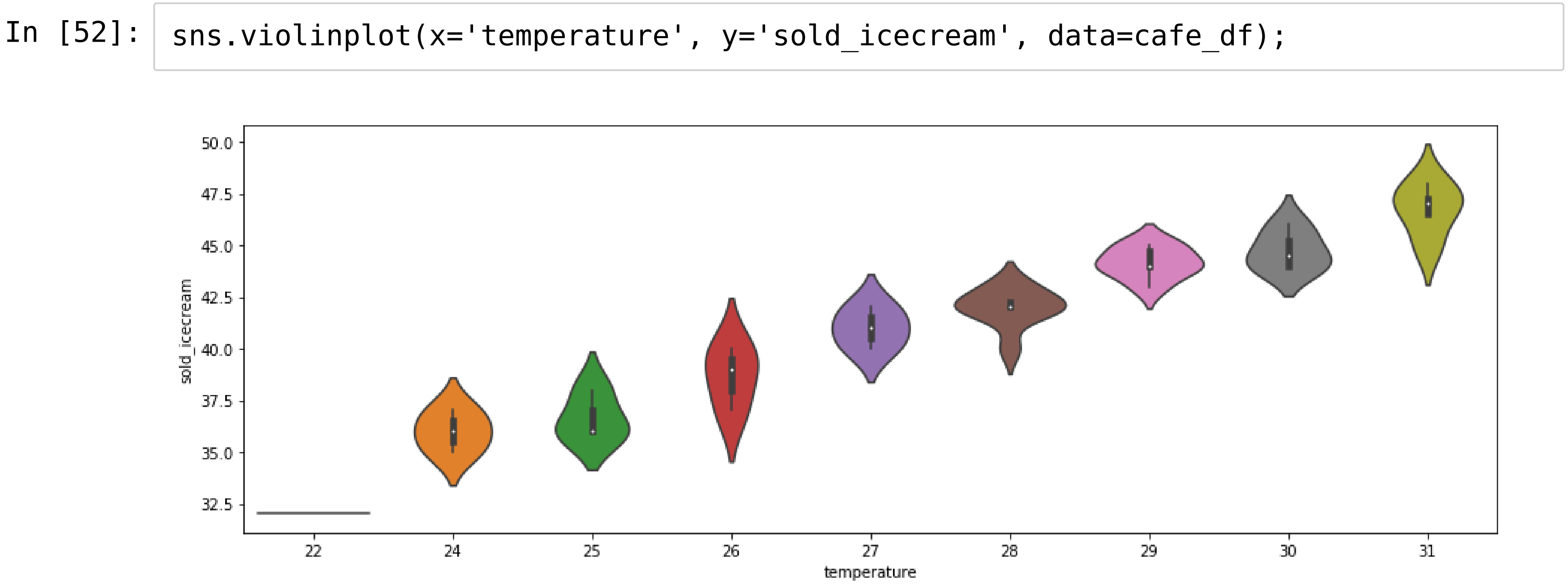
```
ax = cafe_df.plot(kind='line', x='date', y=['sold_icecream','sold_cups_coffee'],
 xticks=[1,10,20,30], style=['g', 'b']);
cafe_df['temperature'].plot(kind='line', x='date', style='r', secondary_y=True);
ax.set_ylabel("Number of Cups");
```



Seaborn

- Includes more advanced plot types: violin plot, heatmap, linear regression
- Styles and themes
- Provides example data
- Gallery: https://seaborn.pydata.org/examples/index.html (https://seaborn.pydata.org/examples/index.html)
- API: https://seaborn.pydata.org/api.html (https://seaborn.pydata.org/api.html)

Seaborn Example: Violin Plot



```
In [52]: sns.violinplot(x='temperature', y='sold_icecream', data=cafe_df);
```

Seaborn Example: Heatmap

Corrleation Coefficients

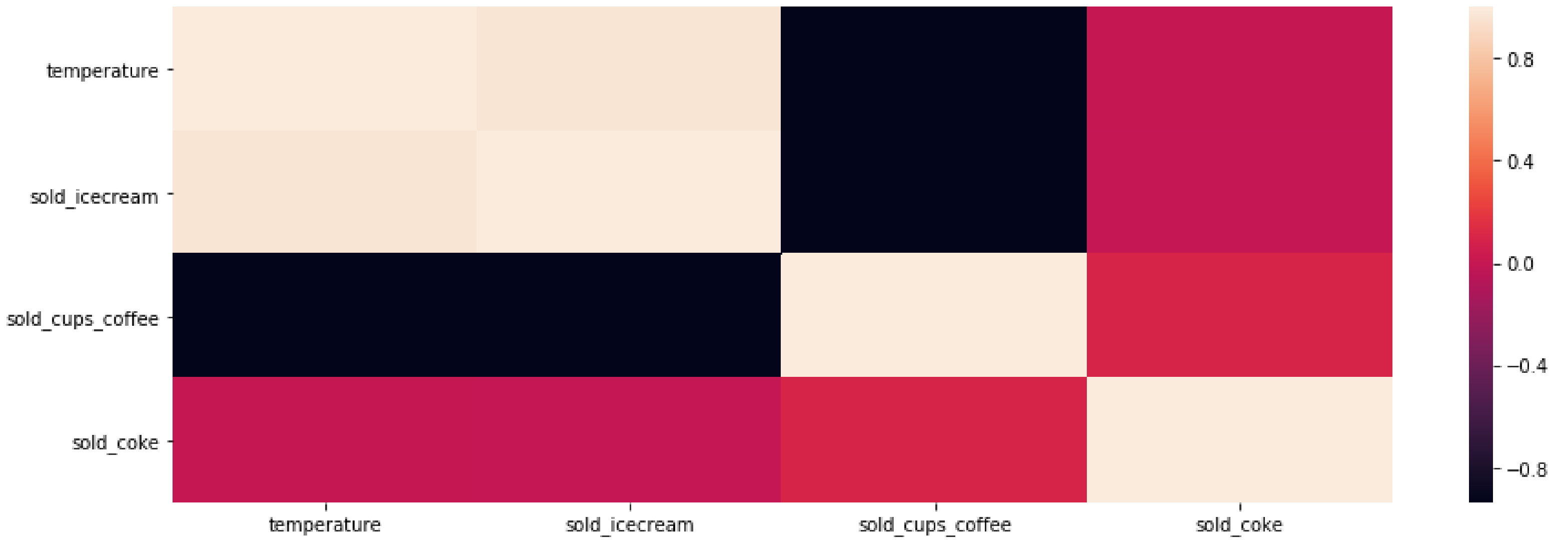
```
In [58]: corr = cafe_df.corr()
         corr
```

```
Out[58]:
```

	temperature	sold_icecream	sold_cups_coffee	sold_coke
temperature	1.000000	0.966549	-0.932512	0.002587
sold_icecream	0.966549	1.000000	-0.934670	-0.002490
sold_cups_coffee	-0.932512	-0.934670	1.000000	0.093498
sold_coke	0.002587	-0.002490	0.093498	1.000000

Heatmap

```
In [60]: sns.heatmap(corr);
```



Exercise 6

- Using matplotlib
  - Load the Rossmann sales data
  - Choose a single store and plot its customers (left y-axis) and sales (right y-axis) data, limit the time range to one month
- Load the seaborn dataset 'tips', calculate the tip percentage:
  - `tips = sns.load_dataset('tips')`
  - `tips['tip_pct'] = tips['tip'] / (tips['total_bill'] - tips['tip'])`
- Plot a bar chart with the average tip percentage per day
  - Using Pandas plot
  - Using Seaborn "barplot"
- Plot histogram of the tip percentage
  - Using Pandas plot
  - Using Seaborn "distplot" (combinded histogram and density plot)
- Draw boxplot and violin plots of the total bill per day
  - Using Pandas plot (boxplot only)
  - Using Seaborn