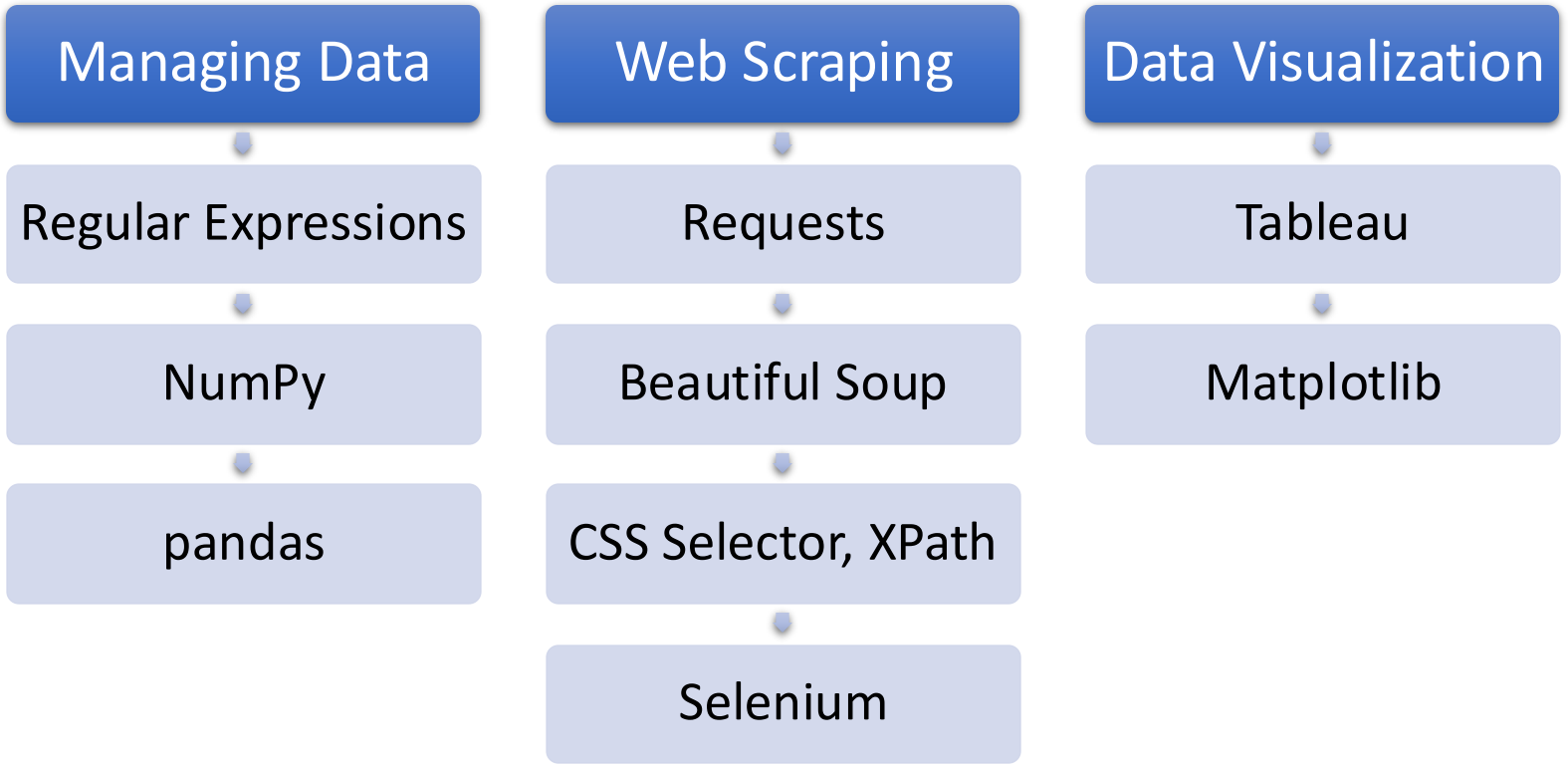
Data Visualization III

MSBA7001 Business Intelligence and Analytics
HKU Business School
The University of Hong Kong

Instructor: Dr. DING Chao

Course Roadmap

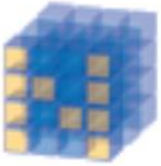


Matplotlib

Matplotlib

- Matplotlib is part of the SciPy ecosystem.
- It provides a library for basic data visualization.
- Basic module: **`matplotlib.pyplot`**

```
import matplotlib.pyplot as plt
```



NumPy
Base N-dimensional array package

SciPy library
Fundamental library for scientific computing

Matplotlib
Comprehensive 2D Plotting

IP[y]: IPython

IPython
Enhanced Interactive Console

Sympy
Symbolic mathematics

pandas
Data structures & analysis

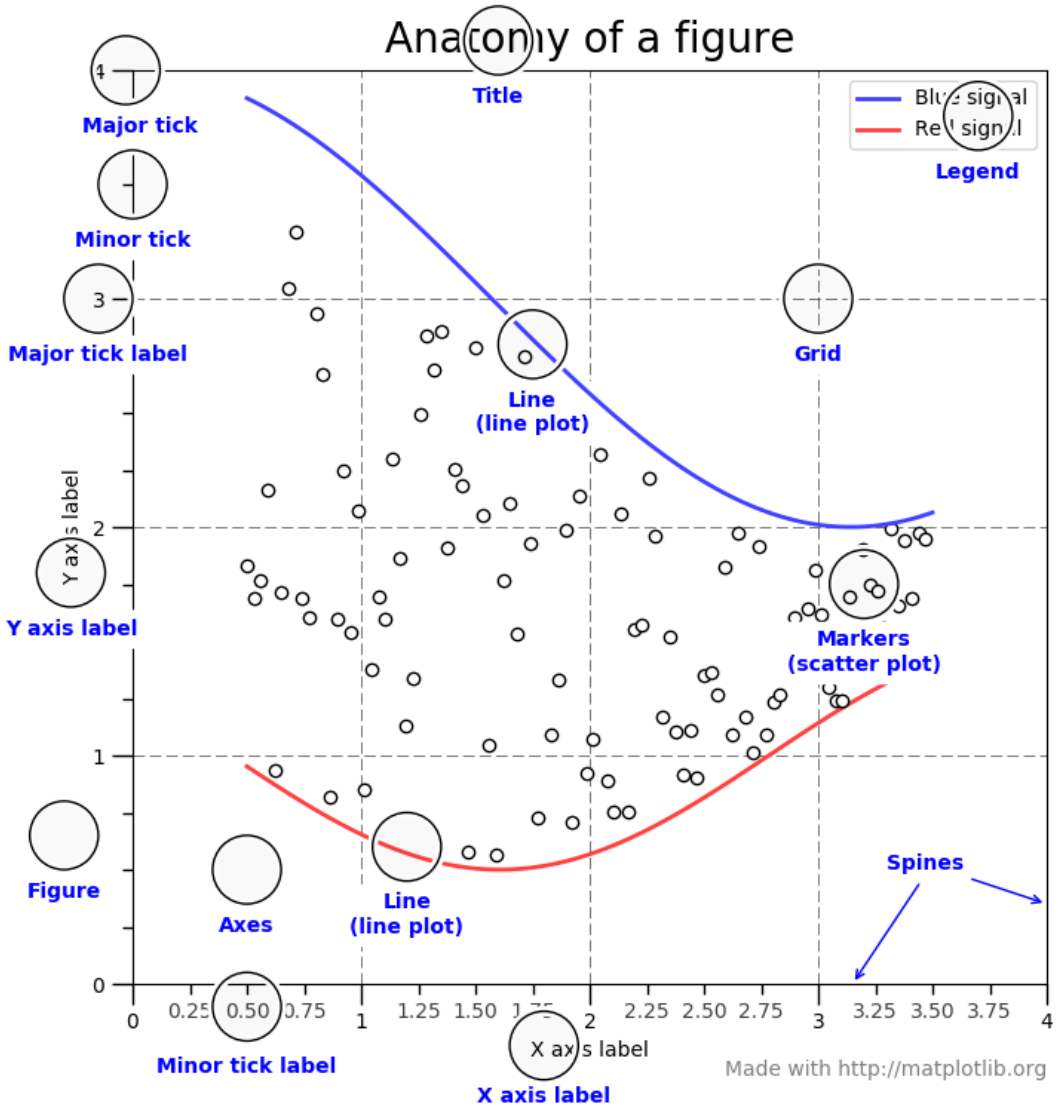
Making Basic Charts

Method	Description
plot([x], y)	Line/scatter plot
scatter(x, y)	Scatter plot
bar(x, height)	Bar plot
hist(x, [bin])	Histogram
boxplot(x)	Boxplot
pie(x)	Pie chart
pcolormesh(x)	Heat map
stackplot(x, y, labels)	Area chart

- More examples:

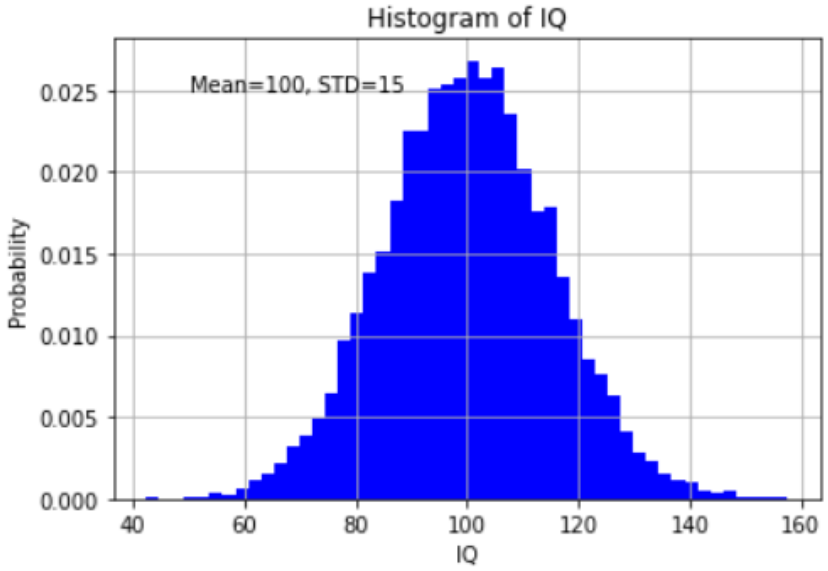
  https://matplotlib.org/stable/gallery/index.html

Formatting Charts



Method	Description
xlabel()	Adds a label to x axis
ylabel()	Adds a label to y axis
xticks()	Adds tick labels to x axis
yticks()	Adds tick labels to y axis
title()	Adds a title to the entire
legend()	Shows legend
text(x, y, text)	Adds text at location x, y
grid(True)	Shows the grid lines

Formatting Charts



Save a chart to local drive

Show then close the chart

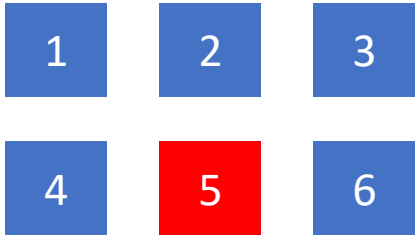
```
x = np.random.normal(100, 15, 10000)
plt.hist(x, 50, color = 'b', density = True)
plt.xlabel('IQ')
plt.ylabel('Probability')
plt.title('Histogram of IQ')
plt.text(50, .025, 'Mean=100, STD=15')
plt.grid(True)
plt.savefig('hist.png')
plt.show()
```

Making Subplots

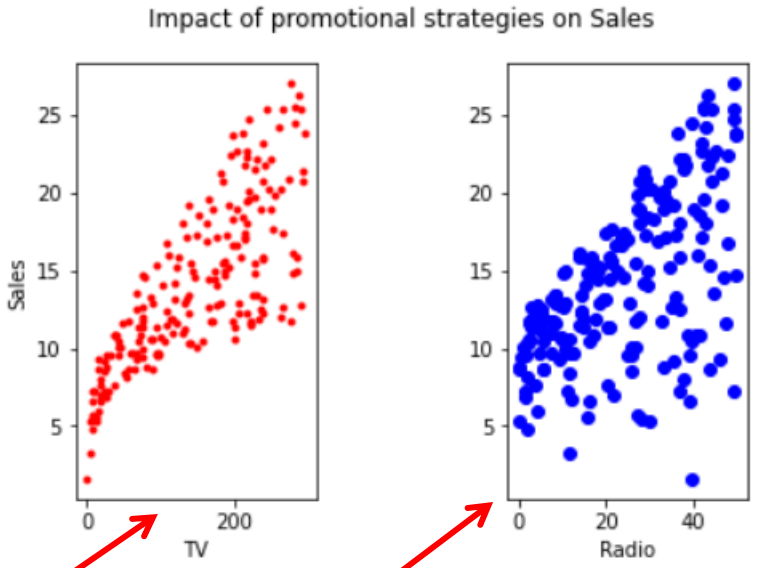
- The **`subplot`** method adds additional subplots to the same chart.

```
subplot(pos)
```

- pos is a three-digit integer, where the first digit is the number of rows, the second the number of columns, and the third the index of the subplot
- 235 means 2 rows, each row 3 subplots, so in total 6 subplots. Current index is the 5th subplot.



Making Subplots




```
plt.subplot(121)
plt.plot('TV', 'sales', 'r.', data = advertising)
plt.xlabel('TV')
plt.ylabel('Sales')

plt.subplot(122)
plt.plot('radio', 'sales', 'bo', data = advertising)
plt.xlabel('Radio')

plt.subplots_adjust(wspace = 0.8)
plt.suptitle('Impact of promotional strategies on Sales')
plt.show()
```