



就业班系列课程

*Python*数据可视化

第一阶段：Python数据可视化入门

第二阶段：Matplotlib绘图入门

第三阶段：Matplotlib绘图高级参数

第四阶段：高级绘图工具

什么是数据可视化？

1. 信息的视觉表达是一种古老的思维和经验的分享方式。
 - 将数据以图片或图形的方式展现的科学。
 - 以连贯和简短的形式展现大量的信息。
2. 重点是用图片和图像的形式向观众传递信息。
3. 信息多样性——可以传递各种信息（理念和假设）
4. 技术多样性——新的技术和选择交互式的可视化方法。

什么是数据可视化?

Anscombe's quartet		X				Y				(X,Y)	
		count	mean	std	var	count	mean	std	var	corr(X,Y)	cov(X,Y)
I		11	9	3.3	11	11	7.5	2.0	4.1	0.82	5.5
II		11	9	3.3	11	11	7.5	2.0	4.1	0.82	5.5
III		11	9	3.3	11	11	7.5	2.0	4.1	0.82	5.5
IV		11	9	3.3	11	11	7.5	2.0	4.1	0.82	5.5
Linear regression equation						$y = 0.5x + 3$					

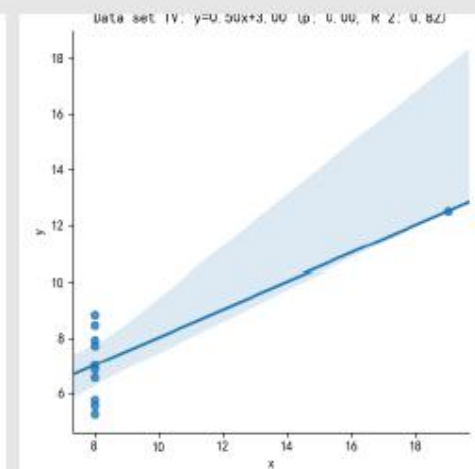
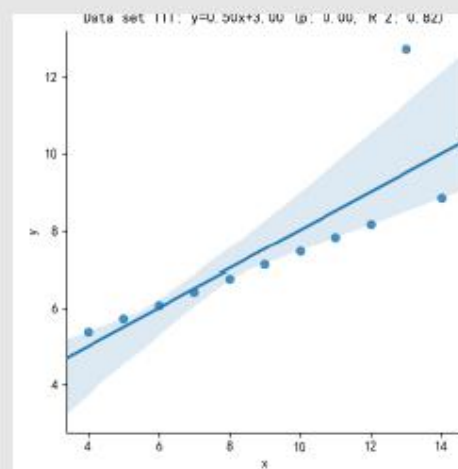
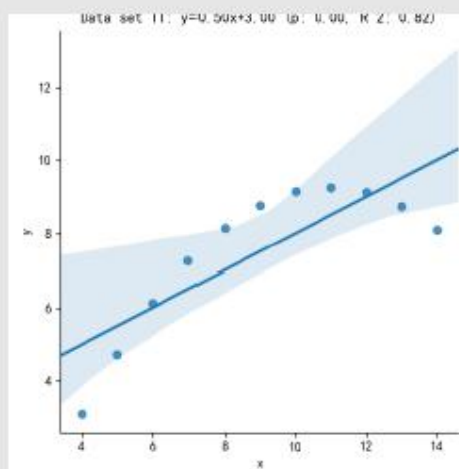
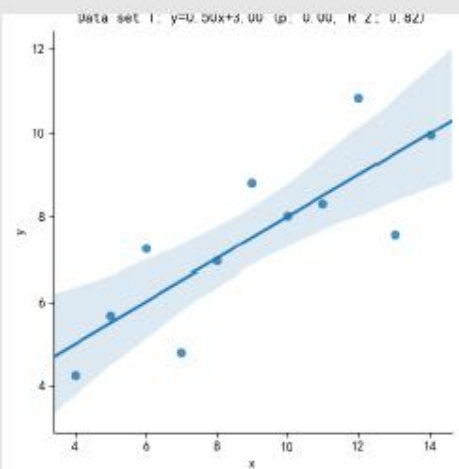
什么是数据可视化?

Anscombe's quartet

dataset	count	x				count	y				(x,y)	
		mean	std	var			mean	std	var		corr(x,y)	cov(x,y)
I	11	9	3.3	11		11	7.5	2.0	4.1		0.82	5.5
II	11	9	3.3	11		11	7.5	2.0	4.1		0.82	5.5
III	11	9	3.3	11		11	7.5	2.0	4.1		0.82	5.5
IV	11	9	3.3	11		11	7.5	2.0	4.1		0.82	5.5

Linear regression equation

$$y = 0.5x + 3$$



为什么数据可视化重要

1. 图像和图表是交流和学习新信息的有效方法。
 - 80%的人能记住他们看到的，但只有20%的人记得怎么读！
 - 在接受信息3天以后，人们只能记住10%他们所听到的，但是如果利用图表展示信息，所记住的信息可以提升到65%。
2. 数据可视化能够帮助人们更快地抓住信息关键点。
3. 一张图表可能会突出显示多个方面，人们就能对数据形成不同的观点。
4. 更容易地从大量的数据和信息中寻找关联

常用可视化第三方库介绍

Python Data Viz

matplotlib

ggplot

Plotly

Pyecharts

Glean

pandas

Pygal

Bokeh

Missingno

Leather

Seaborn

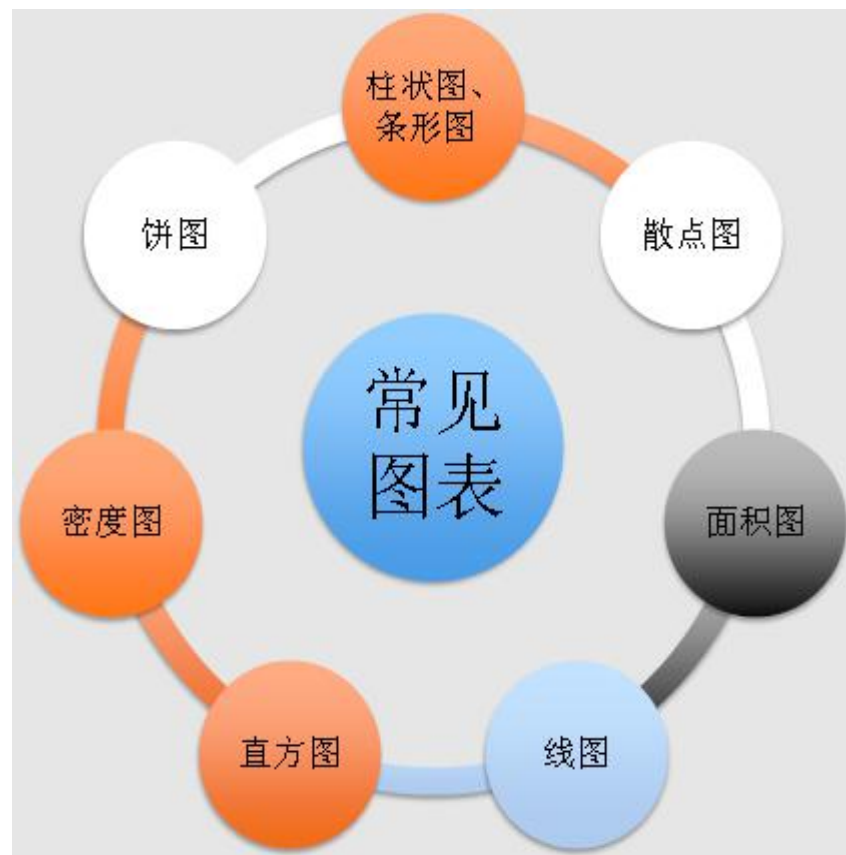
Geoplotlib

Dash

pastalog

GazeParser

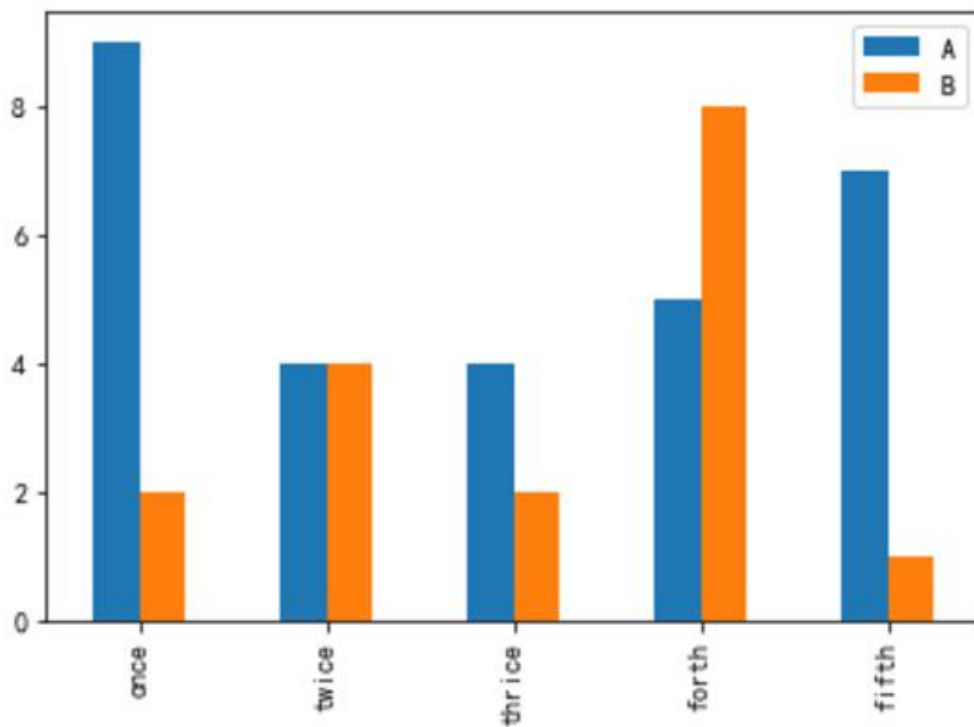
常用可视化图形介绍



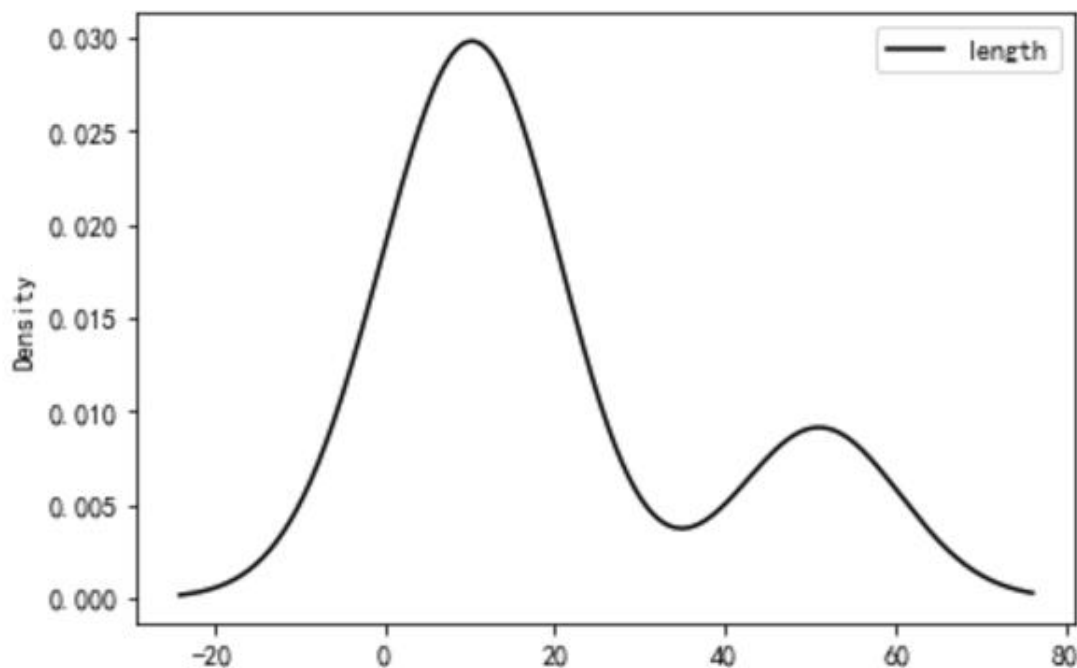
Pandas绘图方法

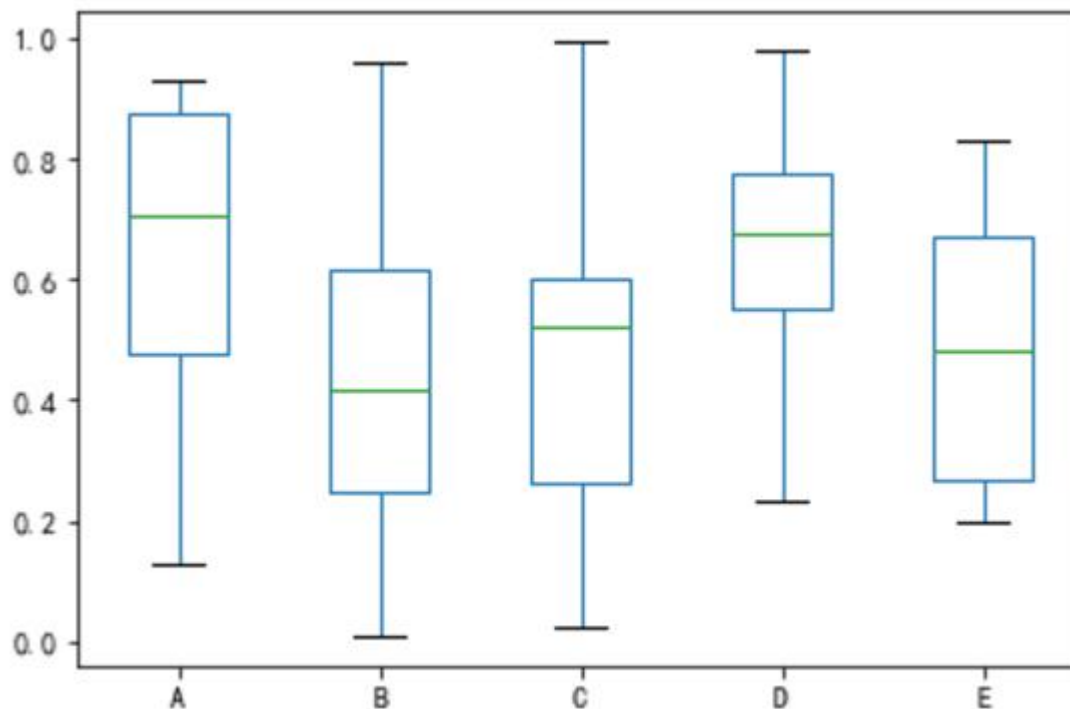
Pandas绘图方法

```
1 dataframe=pd.DataFrame({'A':[9, 4, 4, 5, 7], 'B':[2, 4, 2, 8, 1]})  
2 dataframe.index=['once', 'twice', 'thrice', 'forth', 'fifth']  
3  
4 dataframe.plot.bar();
```



```
1 dataframe.plot.density(color='k');
```



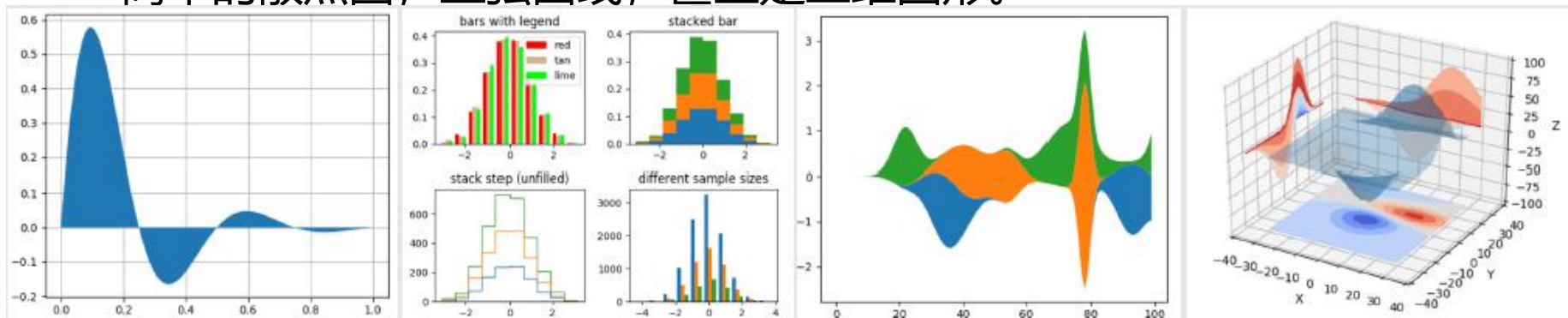


Matplotlib绘图

Matplotlib介绍

1. matplotlib是一个用于创建出版质量图表的桌面绘图包。
2. 包含大量的工具，可以使用这些工具创建各种图形（主要是2D方面）。

简单的散点图，正弦曲线，甚至是三维图形。



3. 它提供了一整套和MATLAB相似的命令API，十分适合交互式地进行制图。

4. 可以方便地将它作为绘图控件，嵌入GUI应用程序中。

Matplotlib三种不同的绘图方式

绘图方式	特点	优劣
pyplot	提供了一套和MATLAB类似的绘图API，将众多绘图对象所构成的复杂结构隐藏在这套API内部， 推荐使用	简单易用，方便快捷，交互使用时方便，可以根据命令实时作图，但底层定制能力不足
pylab	包括了许多NumPy和pyplot模块中常用的函数，方便用户快速进行计算和绘图，十分适合在IPython交互式环境中使用。	完全封装，环境最接近MATLAB，方便用户快速进行计算和绘图，十分适合在IPython交互式环境中使用，不推荐使用
面向对象	Matplotlib的精髓，更基础和底层的方式	接近Matplotlib基础和底层，难度稍大，但定制能力强

总结：实战中推荐更具需求综合使用pyplot和面向对象的方式，显示导入numpy

常用模块导入代码：

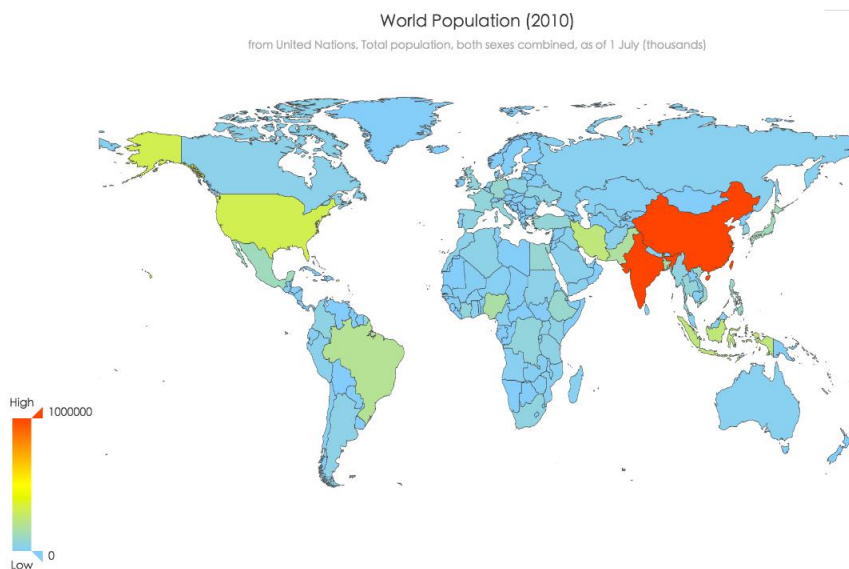
```
import matplotlib.pyplot as plt  
import numpy as np
```

Pyecharts介绍与使用

- Echarts 是由百度开发的一款开源免费，覆盖各行业图表的纯 JavaScript 的可视化库。
- Echarts可以提供直观、生动、可交互和可高度个性化定制的数据可视化图表。
- Pyecharts 是一个用于生成 Echarts 图表的类库，能利用几行代码轻松生成Echarts 风格的图表。
- pyecharts包含的图表：
 - Bar (柱状图/条形图)
 - Bar3D (3D 柱状图)
 - Boxplot (箱形图)
 - EffectScatter (带有涟漪特效动画的散点图)
 - Funnel (漏斗图)
 - Gauge (仪表盘)
 - Geo (地理坐标系)
 - ...

Pyecharts的使用

1. `pip install pyecharts`
2. 下面在 CMD 命令提示行中输入命令，安装 Pyecharts，命令代码如下。
`conda install pyecharts`
3. Pyecharts 自动把 echarts 脚本文件装在了 jupyter nbextensions 文件夹下面。



Talk is cheap
Show me the
CODE