

快速上手数据处理及绘图

数据处理

学习目标：

- 掌握Numpy与pandas
- 数据处理练习

Numpy与pandas

学习材料：[网盘下载](#) 提取码 MLHW

学习内容：

1. 细读 第 4 章 NumPy 基础： 数组和矢量计算（另外，附录A 看下广播机制即可）
2. 细读 第 5 章 pandas 入门
3. 补充细读[pandas](#)中"十分钟入门Pandas"Merge、Grouping、Reshaping

阅读说明：过程中务必敲一敲，敲过会发现理解起来更容易、思考的更广更深。

附：jupyter可以将光标移至设参数的位置，使用 Shift+tab 查看参数设置，或者使用 `help(**)` 。

数据可视化

学习目标：

- 掌握必要的绘图方法
- 感受参数设置规律，要求一搜就能画
- 理解图形要素，能选合适的图形进行数据探索和表达

学习材料：[网盘下载](#) 提取码：MLHW

matplotlib

本节需掌握：

- 导入包 as plt
- plt.plot() 及格式字符串
- plt.subplot() 用法及 plt.add_subplot() 有和区别
- plt.bar() 条形图
- numpy.histogram() 及 plt.hist()

Seaborn

本节需掌握：

- sns.boxplot() 箱线图
- sns.stripplot、sns.swarmplot 两种散点图
- sns.countplot() 记数图
- sns.catplot() 分类图
- 数值分布绘图
- sns.jointplot (双变量)
- sns.pairplot (成对关系)

进阶练习（选学）

本节可以选择感兴趣的部分敲敲，第3节学习材料：《[50个可视化图表](#)》

推荐练习：

- 连续变量的直方图 (Histogram for Continuous Variable)
- 类型变量的直方图 (Histogram for Categorical Variable)
- 使用辅助 Y 轴来绘制不同范围的图形 (Plotting with different scales using secondary Y axis)
- 日历热力图 (Calendar Heat Map)