

第五课 数据分析工具Pandas基础

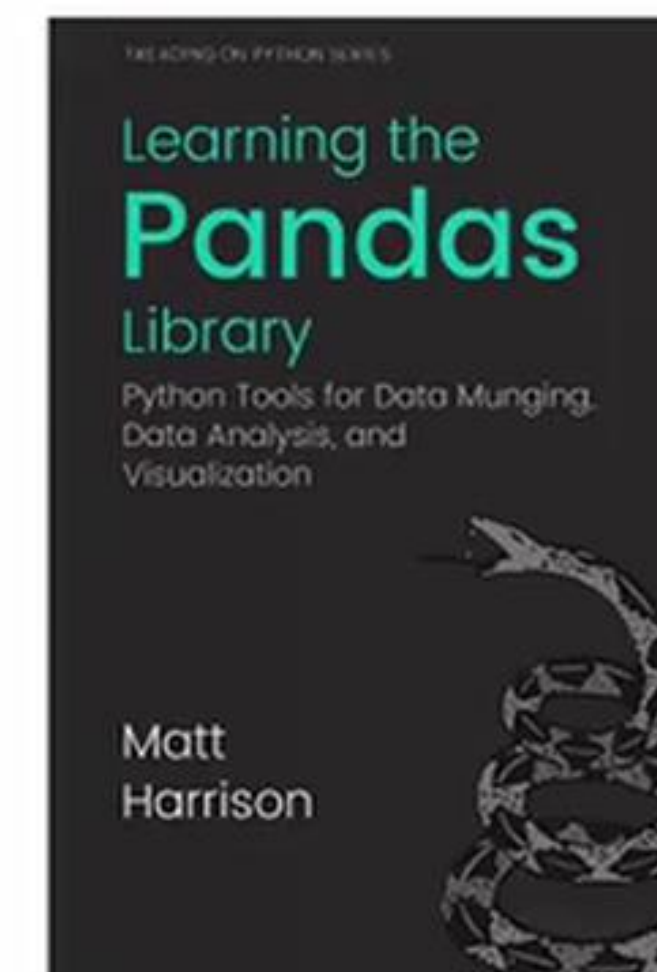
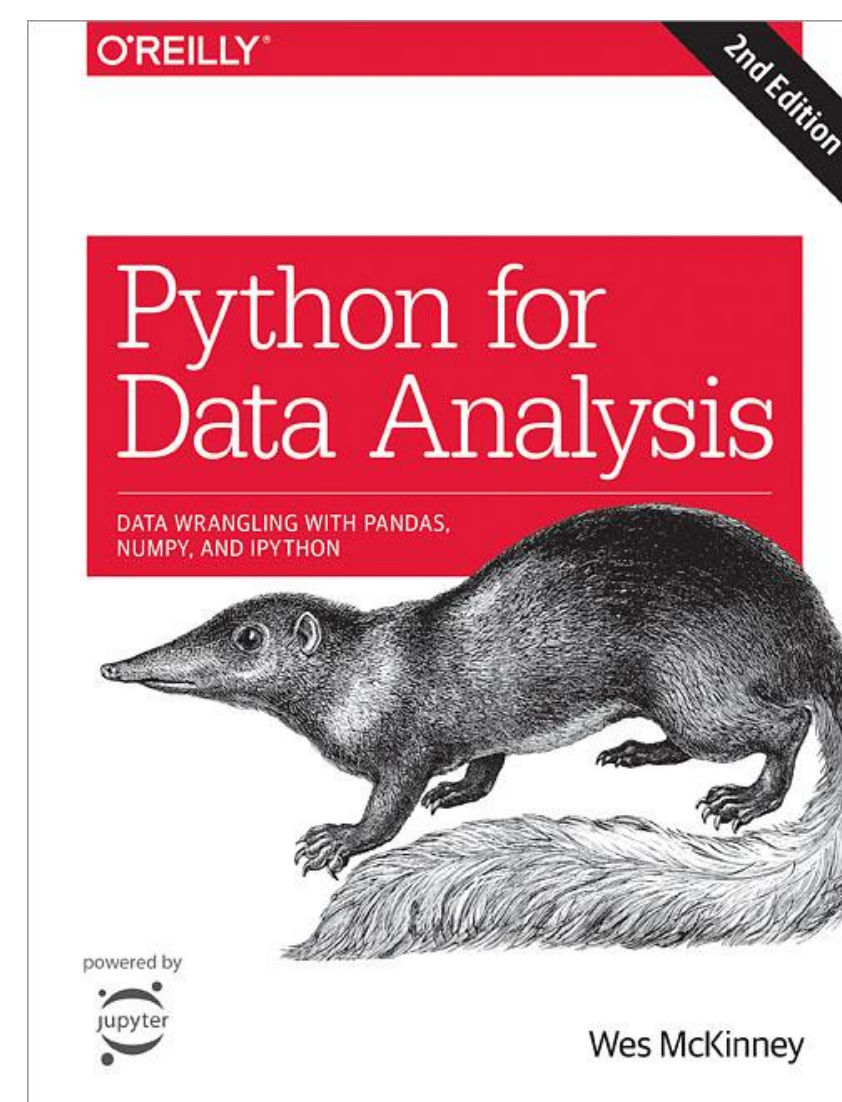
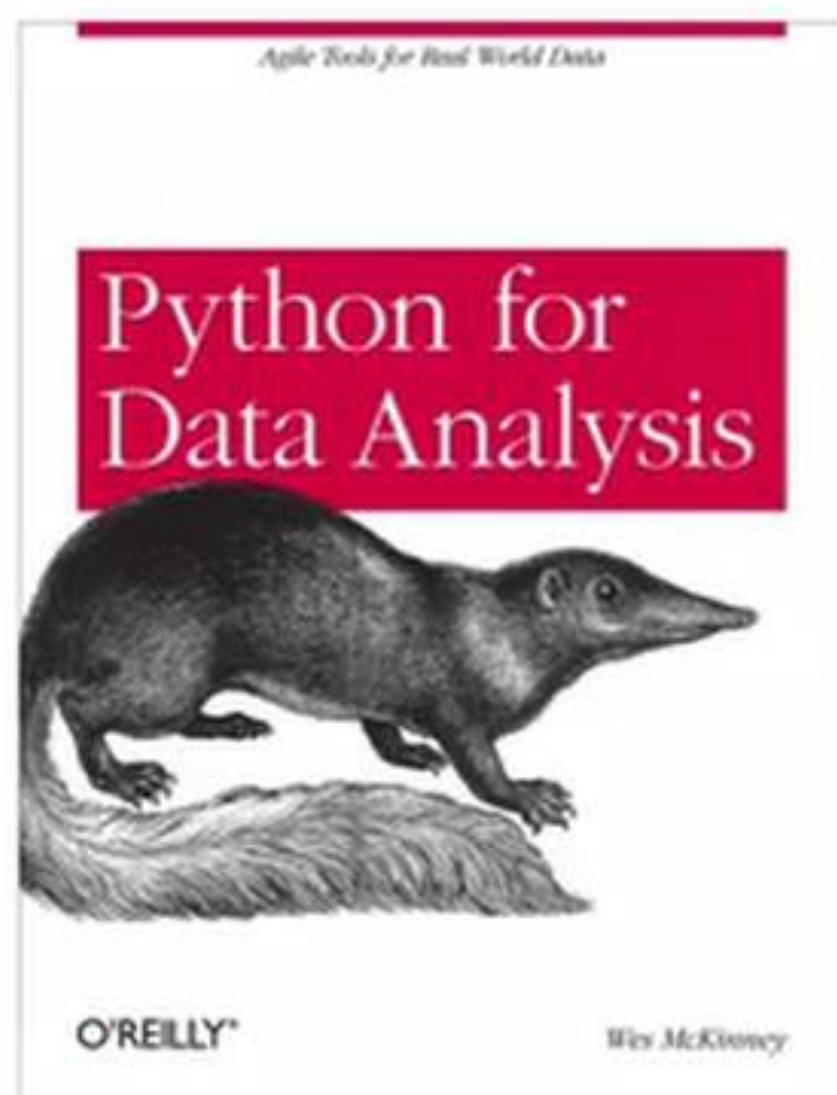
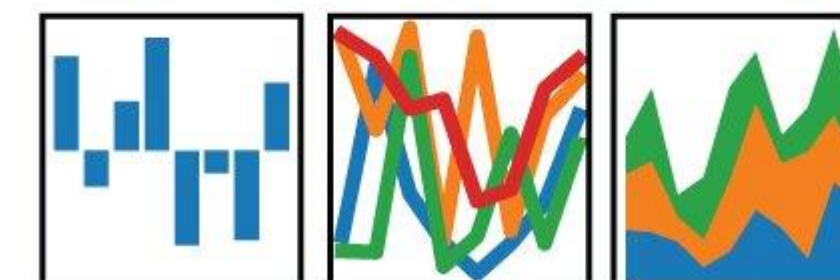
第一节 数据结构--Series

主 讲: Robin

Pandas

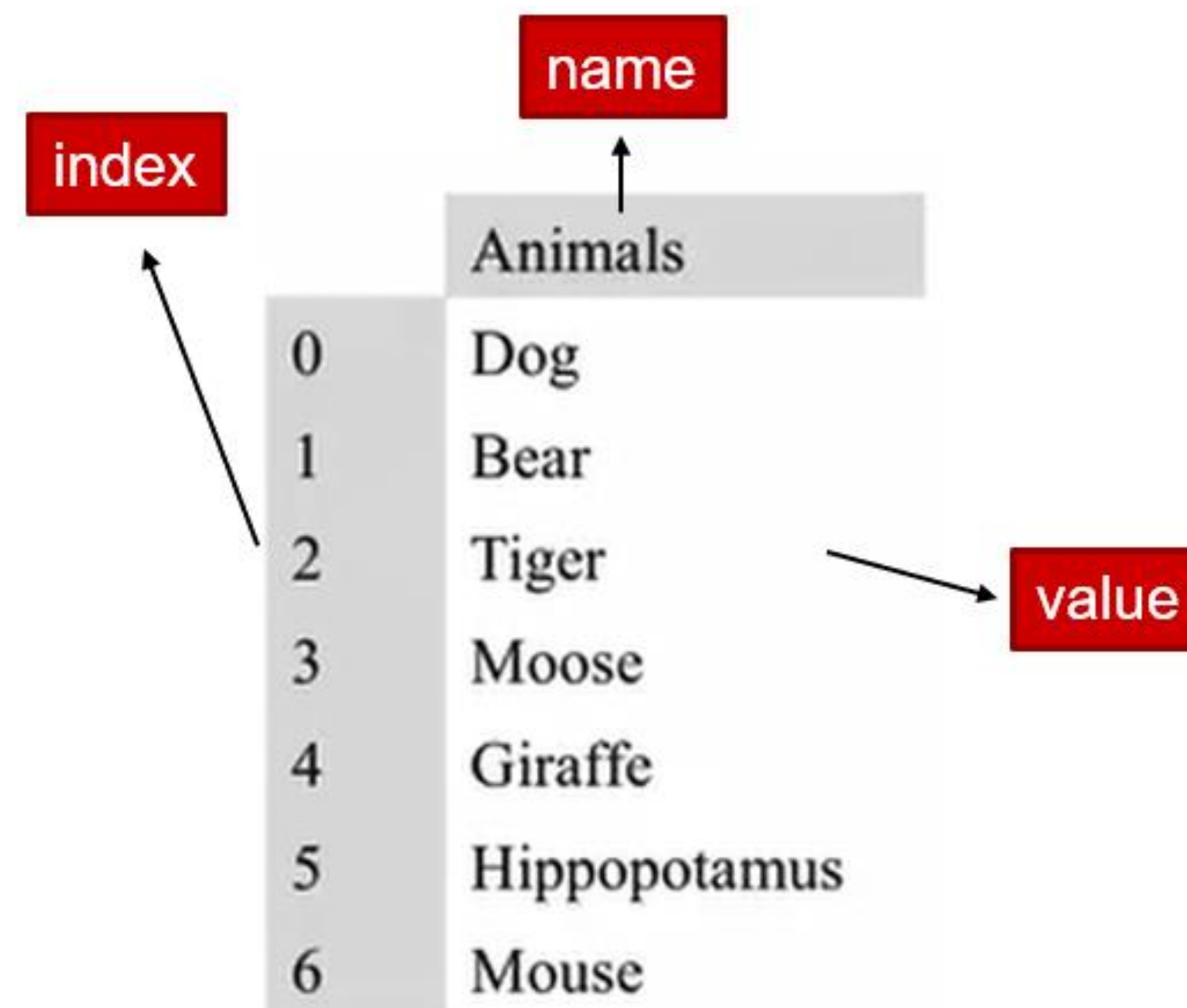
- 一个强大的分析结构化数据的工具集
- 2008年创建，最初被作为金融数据分析工具
- 基础是NumPy，提供了高性能矩阵的运算

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$


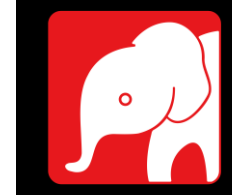
数据结构 -- Series

- 类似一维数组的对象
- 构建Series
 - 通过数组/列表
 - 通过dict
- 预览数据, `ser_obj.head()`
- 由数据和索引组成
 - 索引在左, 数据在右
 - 索引是自动创建的
- 获取数据和索引
 - `ser_obj.index`, `ser_obj.values`



数据结构 -- Series

- name属性, `ser_obj.name`, `ser_obj.index.name`
- 通过索引名(字符串)获取数据, `ser_obj['idx_name']`, 或 `ser_obj.loc['idx_name']`
- 通过in判断数据是否存在, Series也可看作定长、有序的字典
- 通过索引位置(整型)获取数据, `ser_obj.iloc[idx]`
- Pandas会根据数据类型自动处理缺失数据
 - 如object -> None, float -> NaN
- 仍然可以使用NumPy中的向量化操作



只为遇见明天更优秀的你！