

# Pandas

2018年8月14日 11:17

## 1. Pandas 介绍

- a. Pandas 可以处理时间序列数据分析
- b. Pandas 是基于 numpy 的
- c. 有两个新的高级数据结构：系列 Series 和数据框 DataFrame

## 2. 数据框

- a. 数据框每列表示一个变量，每行表示一次观测（类似SPSS的结构）
- b. 常用字典（key=变量，value=值）或者numpy数组来创建数据框，注意创建时会动排序，导致创建结构与预想可能不符，但数据的对应关系不会出现错位
- c. 代码：

## 3. 系列

- a. 系列就是某一列变量的序列，实际上是对同一个属性（变量）进行多次观测后得到的结果的序列。
- b. 从统计学上看，系列表达的是某个变量的分布
- c. 代码：

## 4. 代码：本节代码见 code\_rewrite