- 第12章、高阶pandas
  - 12.1 分类数据
    - Categorical类
      - 分类展现：按照整数展现的方式
      - 使用分类能大幅减少内存
    - 12.1.1 分类展现
      - s.take(indices, axis=0)
        - indices是整数列表
        - 返回indices在s中对应的元素的列表
        - indices = pd.Series([0, 1, 0, 0] * 2)
s = pd.Series(['apple', 'orange'])
s.take(indices)

```
0      apple
1     orange
0      apple
0      apple
0      apple
1     orange
0      apple
0      apple
dtype: object
```

    - 12.1.2 Categorical类
      - 数组转换为categrical
        - s.values # numpy.ndarray
        - s.astype('category').values
pandas.core.arrays.categorical.Categorical
      - Categorical属性
        - categories 类别唯一值
        - codes 数组元素在类别中对应的下标
      - 创建Categorical对象
        - 已知类别列表 pd.Categorical(类别列表)
        - 已知类别唯一值、codes pd.Categorical.from_codes(codes, categories)
    - 12.1.3 Categorical对象计算
      - 分箱+groupby
        - bins = pd.qcut(s, 4, labels=[])
s.groupby(bins).agg(['count', 'max']).reset_index()
    - 12.1.4 Categorical对象方法
      - obj.cat.方法名
      - obj.cat.codes / categories

方法	说明
add_categories	在已存在的分类后面添加新的（未使用的）分类
as_ordered	使分类有序
as_unordered	使分类无序
remove_categories	移除分类，设置任何被移除的值为 null
remove_unused_categories	移除任意不出现在数据中的分类值
rename_categories	用指定的新分类的名字替换分类；不能改变分类的数目
reorder_categories	与 rename_categories 很像，但是可以改变结果，使分类有序
set_categories	用指定的新分类的名字替换分类；可以添加或删除分类

      - 12.1.4.1 one-hot编码（虚拟变量）
        - 将一维的分类数据转换为DataFrame
        - pd.get_dummies(cats)
  - 12.2 高阶groupby
    - 12.2.1 展开转换
      - transform
        - 产生一个标量值，并展开到各分组的尺寸
        - 只允许在同一时间在一个Series上进行一次转换 apply可以在df上转换
    - 12.2.2 分组时间重采样
      - 对df的key进行分组，每个组按一定的频率采样
      - 选定时间列，进行重新采样 df.set_index('time')
      - 以key和采样频率构造分组键
        - 采样频率：rule = pd.Grouper(freq='5T') 索引是时间
        - by = ['key', rule]
      - 分组聚合
        - df.set_index('time').groupby(['key', rule]).sum()
        - 以key和time列为索引，其中time列重采样
  - 12.3 链式技术
    - df.assign(k=v) == df[k]=v
    - 12.3.1 pipe管道方法 方便链式调用