

- Pandas: 导入数据

```
1 import pandas as pd
2
3 names = ['mpg', 'cylinders', 'displacement', 'horsepower',
4          'weight', 'acceleration', 'model year', 'origin', 'car name']
5
6 df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/'+
7                  'auto-mpg/auto-mpg.data',
8                  header=None, delim_whitespace=True, names=names, na_values='?')
9
10 # 丢弃无效行
11 df = df.drop('Unnamed: 0', axis=1)
12 # 查看前六行
13 df.head(6)
14
15 # 一些属性
16 # 返回 (行数, 列数)
17 df.shape
18 # 返回列名
19 df.columns
20 df.columns.tolist()
21 # 返回行数索引
22 df.index
23 # 返回每行的数据
24 df.values
25
26 # 选取其中某些列
27 df2 = df[['cylinders', 'horsepower']]
28 # 转化成数组
29 y = np.array(df['mpg'])
30 # 转化成多维数组, 注意这里要用两层 [[]] 包起来! 不知道为啥
31 y = np.array(df[['mpg', 'xxx']])
32
33 # 将列名转化成数组
34 names = df.columns.tolist()
35 names_x = names[0:7]
36 name_y = names[8]
```