

# pandas

---

主讲老师：杨 洋

# 1 ipython

## ●什么是ipython?

### IPython——科学计算标准工具集的组成部分

IPython是一个免费、开源的项目，支持Linux、 Unix、 Mac OS X和Windows平台，其官方网址是<http://ipython.org/>。IPython的作者只要求你在用到IPython的科技著作中注明引用即可。

IPython中包括各种组件，其中的两个主要组件是：

基于终端方式和基于Qt的交互式Python shell

支持多媒体和绘图功能的基于Web的notebook（版本号为0.12以上的IPython支持此功能）

## ●1.1.1 ipython安装

windows: 前提是有numpy, matplotlib pandas

更新pip `python -m pip install --upgrade pip`

采用pip安装 `pip install ipython`

在Mac OS X中安装IPython:

如有必要, 请先安装苹果开发工具Xcode, 可以在Mac电脑附带的OSX DVD光盘中或者苹果应用商店中找到Xcode。使用easy\_install或pip安装IPython, 或者从源文件安装。

## ● 1.1.2 ipython壳的主要特点

提供一个更友好的界面，是一个增强的Python shell，目的是提高编写、测试、调试Python代码的速度。

提供了代码补全，对象检查，系统调用，获取输入历史等实用的功能

庞大的ipython社区努力使其成为一个高校的python科学计算环境

主要用于交互式数据并行处理，是分布式计算的基础架构。

提供了一个非常灵活的框架，可以作为其他应用的基础

提供了一个类似于Mathematica的HTML笔记本，一个基于Qt框架的GUI控制台，具有绘图、多行编辑以及语法高亮显示等功能。允许快速绘图，也支持GTK,Qt, WX,等GUI程序

## ●1.1.3 ipython壳主要内容

自动补全	Tab键
检查	? 查看对象基本信息    ?? 查看构造函数基本信息    ? *匹配对象
%run命令	使用%run调用外部Python脚本的能力    %cd E:\pycharm\python_study
魔法方法	%magic来查找所有的魔法命令
异常和错误信息	使用%run命令行的方式运行, 如果出现错误, ipython会打印错误的的路径和异常的详
和操作系统交互	可以输入shell命令, cd pwd env等
目标标签系统	创建同名目录    %bookmark TI C:\Users\user    cd TI

### ●1.1.4 ipython其他技巧

融合matplotlib库和pylab模型 `ipython --pylab plot(range(1,101),np.random.rand(100))`

输入和输出变量 `ipython` `_`单下划线指代上次的输入值 `__`双下划线指代上次的输出值

计时功能 `%time`可以进行计时 `%timeit`可以进行计时平均值

## 2 掌握 Jupyter Notebook



### ●2.1.1 Jupyter Notebook安装

windows 更新pip `python -m pip install --upgrade pip`

采用pip安装 `pip install Jupyter`

Jupyter Notebook（此前被称为 IPython notebook）是一个交互式笔记本，支持运行 40 多种编程语言。

Jupyter Notebook 的本质是一个 Web 应用程序，便于创建和共享文学化程序文档，支持实时代码，数学方程，可视化和 markdown。已迅速成为处理数据的必备工具，用途包括：数据清理和转换，数值模拟，统计建模，机器学习等等

### ●2.1.2 Jupyter Notebook简介与安装

#### Jupyter优势:

可选择语言：支持超过40种编程语言，包括Python、R、Julia、Scala等。

分享笔记本：可以使用电子邮件、Dropbox、GitHub和Jupyter Notebook Viewer与他人共享。

交互式输出：代码可以生成丰富的交互式输出，包括HTML、图像、视频、LaTeX等等。

大数据整合：通过Python、R、Scala编程语言使用Apache Spark等大数据框架工具。支持使用pandas、scikit-learn、ggplot2、TensorFlow来探索同一份数据。

### ●2.1.2 Jupyter Notebook简介与安装

#### Jupyter包含以下组件:

Jupyter Notebook 和 Notebook 文件格式

Jupyter Qt 控制台

内核消息协议 (kernel messaging protocol)

许多其他组件

### ● 2.1.2 Jupyter Notebook简介与安装

**Notebook**有两个部分组成:

**网络应用**: 基于网络的文档处理环境, 主要包括文本编辑, 数学计算, 及丰富的输出

**Notebook文档**: 网络应用中所有可见的内容, 主要包括文本, 图片等

**主要特性有:**

在浏览器编辑代码, 会自动进行语法加亮, 缩进, 补齐, 检查等

通过浏览器运行代码, 运行结果直接输出到代码的后面

有多种输出方式, 包括PNG,SVG,HTML,LaTeX

在浏览器种, 使用Markdown标记语言为代码提供注释, 不再仅限于普通文本的注释方式

可以方便的标记数学公式

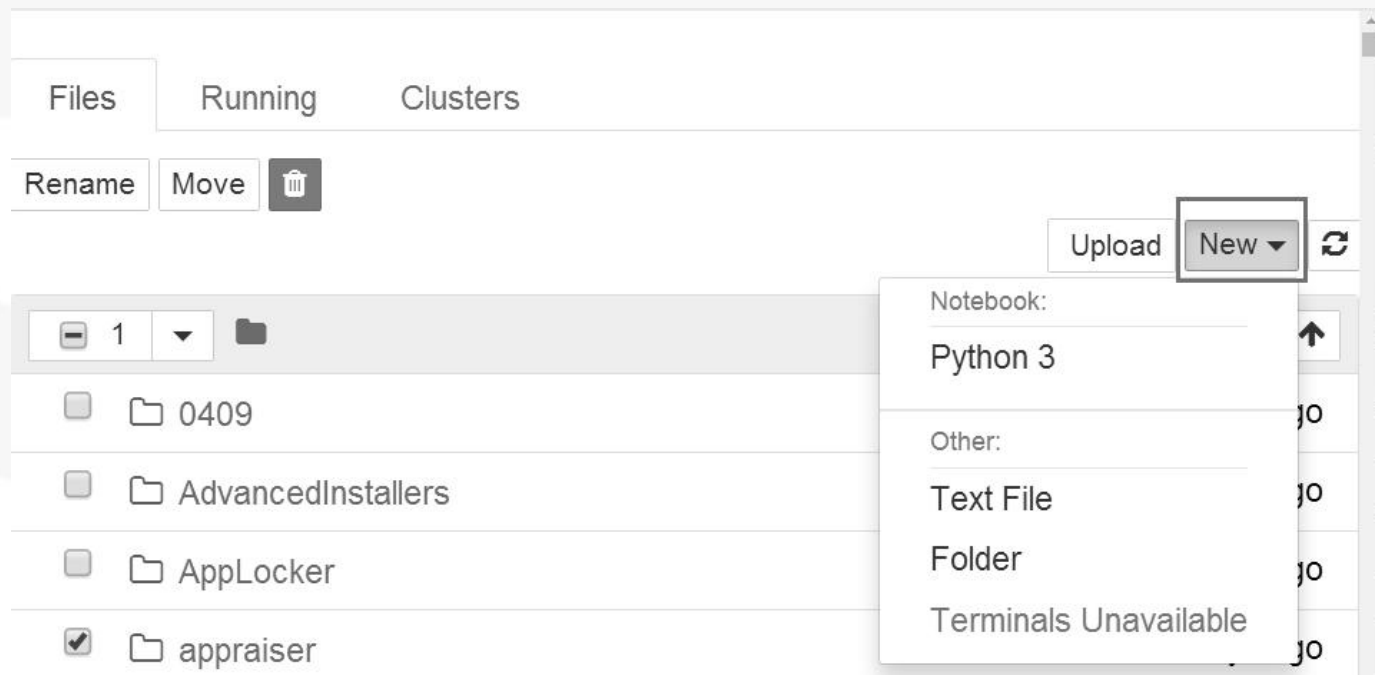
导出数据和数据分析过程

### ●2.1.2 Jupyter Notebook简介与安装

#### 打开并新建一个Notebook

打开 Jupyter Notebook

- “Text File”为纯文本型
- “Folder” 为文件夹
- “Python 3” 表示 Python 运行脚本

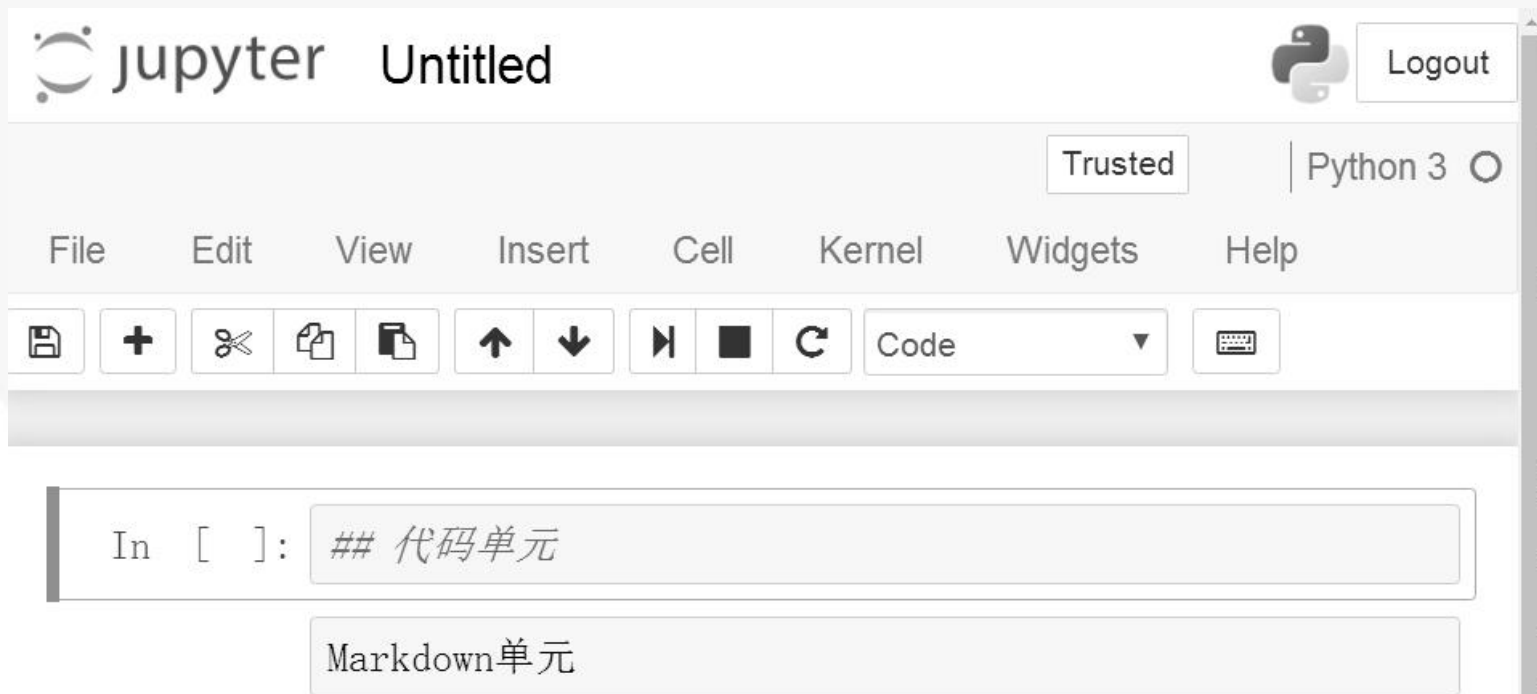


### ●2.1.2 Jupyter Notebook简介与安装

#### Jupyter Notebook 的界面及其构成

选择“Python 3”选项，进入 Python 脚本编辑界面，Notebook 文档由一系列单元（Cell）构成，主要有两种形式的单元。

- 代码单元。这里是读者编写代码的地方。
- Markdown 单元。在这里对文本进行编辑。



## ●2.1.2 Jupyter Notebook简介与安装

编辑界面：用于编辑文本和代码



命令模式：用于执行键盘输入的快捷命令。



### ●2.1.2 Jupyter Notebook简介与安装

- 快捷键

- “Esc” 键：进入命令模式
- “Y” 键：切换到代码单元
- “M” 键：切换到 Markdown 单元
- “B” 键：在本单元的下方增加一单元
- “H” 键：查看所有快捷命令
- “Shift + Enter” 组合键：运行代码



### ●2.1.3 Markdown 使用

Markdown 是一种可以使用普通文本编辑器编写的标记语言，通过简单的标记语法，它可以使普通文本内容具有一定的格式。

➤ 标题：标题是标明文章和作品等内容的简短语句。一个 “#” 字符代表一级标题，以此类推。

# 一级标题

## 二级标题

### 三级标题

#### 四级标题

##### 五级标题

##### 六级标题

一级标题

二级标题

三级标题

四级标题

五级标题

六级标题

### ●2.1.3 Markdown 使用

- 列表：列表是一种由数据项构成的有限序列，即按照一定的线性顺序排列而成的数据项的集合。
- 对于无序列表，使用星号、加号或者减号作为列表标记
- 对于有序列表，则是使用数字 “， ” “（一个空格）”。

```
* Python  
+ Python2  
- Python3  
  
1. Python  
2. Python2  
3. Python3
```

```
• Python  
• Python2  
• Python3  
  
1. Python  
2. Python2  
3. Python3
```

### ●2.1.3 Markdown 使用

- 加粗 / 斜体：前后有两个星号或下划线表示加粗，前后有 3 个星号或下划线表示斜体。

Python数据分析

**Python数据分析**

***Python数据分析***

Python数据分析

*python数据分析*

Python数据分析

**Python数据分析**

***Python数据分析***

Python数据分析

*python数据分析*

### ●2.1.3 Markdown 使用

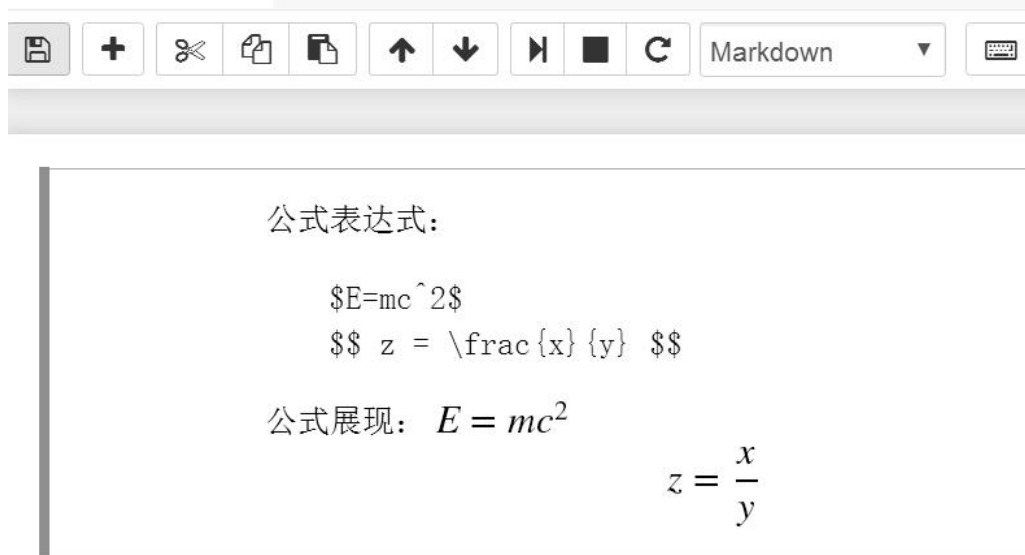
- 表格：代码的第一行表示表头，第二行分隔表头和主体部分，从第三行开始，每一行代表一个表格行；列与列之间用符号 “|” 隔开，表格每一行的两边也要有符号 “|”。

```
Python | R | MATLAB |  
-----|-----|----|  
接口统一，学习曲线平缓 | 接口众多，学习曲线陡峭 | 自由度大，学习曲线较为平缓 |  
开源免费 | 开源免费 | 商业收费 |
```

Python	R	MATLAB
接口统一，学习曲线平缓	接口众多，学习曲线陡峭	自由度大，学习曲线较为平缓
开源免费	开源免费	商业收费

### ●2.1.3 Markdown 使用

- 数学公式编辑：LaTeX 是写科研论文的必备工具，Markdown 单元中也可以使用 LaTeX 来插入数学公式。
- 在文本行中插入数学公式，应在公式前后分别加上一个 “\$” 符号
- 如果要插入一个数学区块，则在公式前后分别加上两个 “\$\$” 符号。



### ●2.1.3 Markdown 使用

### ●导出功能

- Notebook 还有一个强大的特性，就是导出功能。可以将 Notebook 导出为多种格式，如HTML、Markdown、reST、PDF（通过 LaTeX）等格式。
- 导出功能可通过选择 “File” → “Download as” 级联菜单中的命令实现。

