

# Python简要介绍

中央财经大学 孟凡



- Python简介
- Python安装与配置

- 数据挖掘常用工具包



## Python的创始人为荷兰的 Guido。

1989年，Guido为了打发圣诞节的无趣，决心开发一个新的脚本解释程序，做为ABC语言的一种继承。之所以选中Python作为该编程语言的名字，是因为他是一个叫Monty Python的喜剧团体的爱好者。





# Python的特点



- 简单易学
- 开源
- 高级语言
- 可移植性
- 解释性
- 面向对象
- 可扩展
- 丰富的库
- 规范的代码



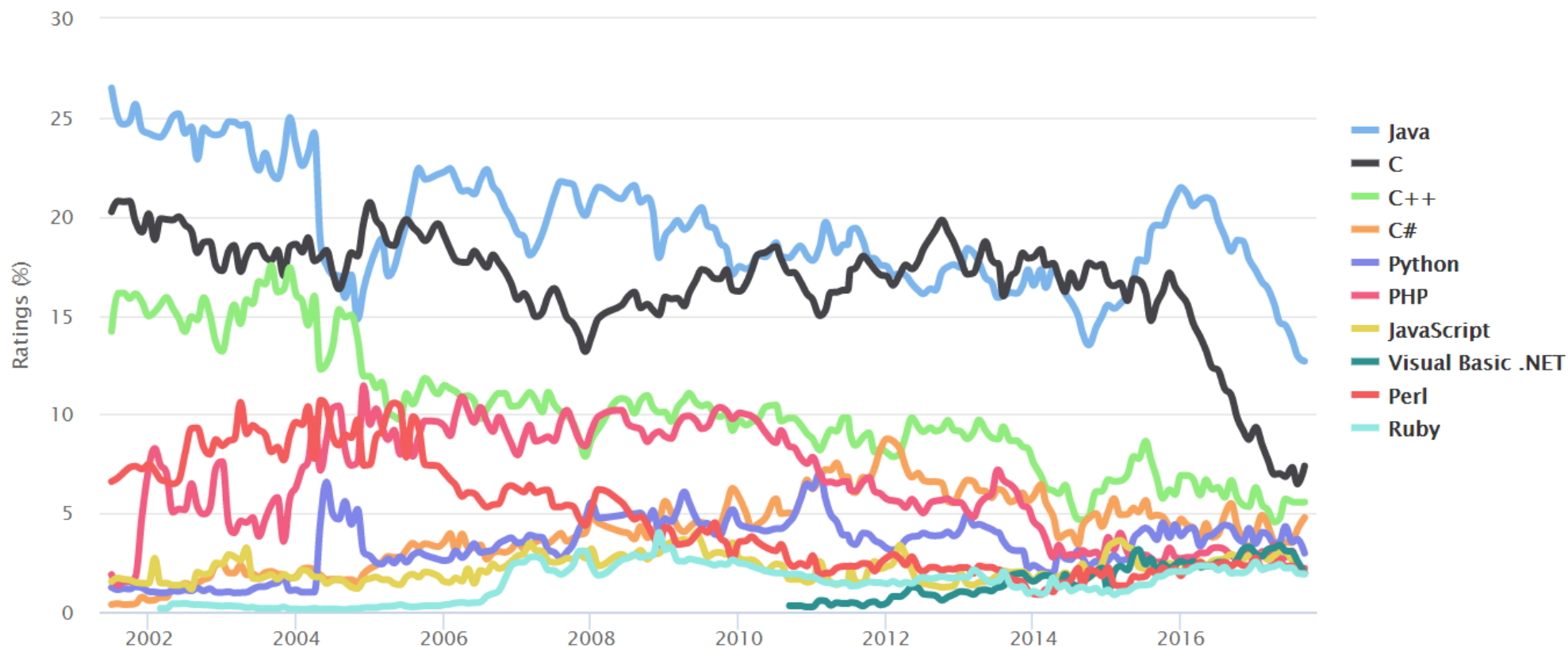


# Python的热度



TIOBE Programming Community Index

Source: [www.tiobe.com](http://www.tiobe.com)

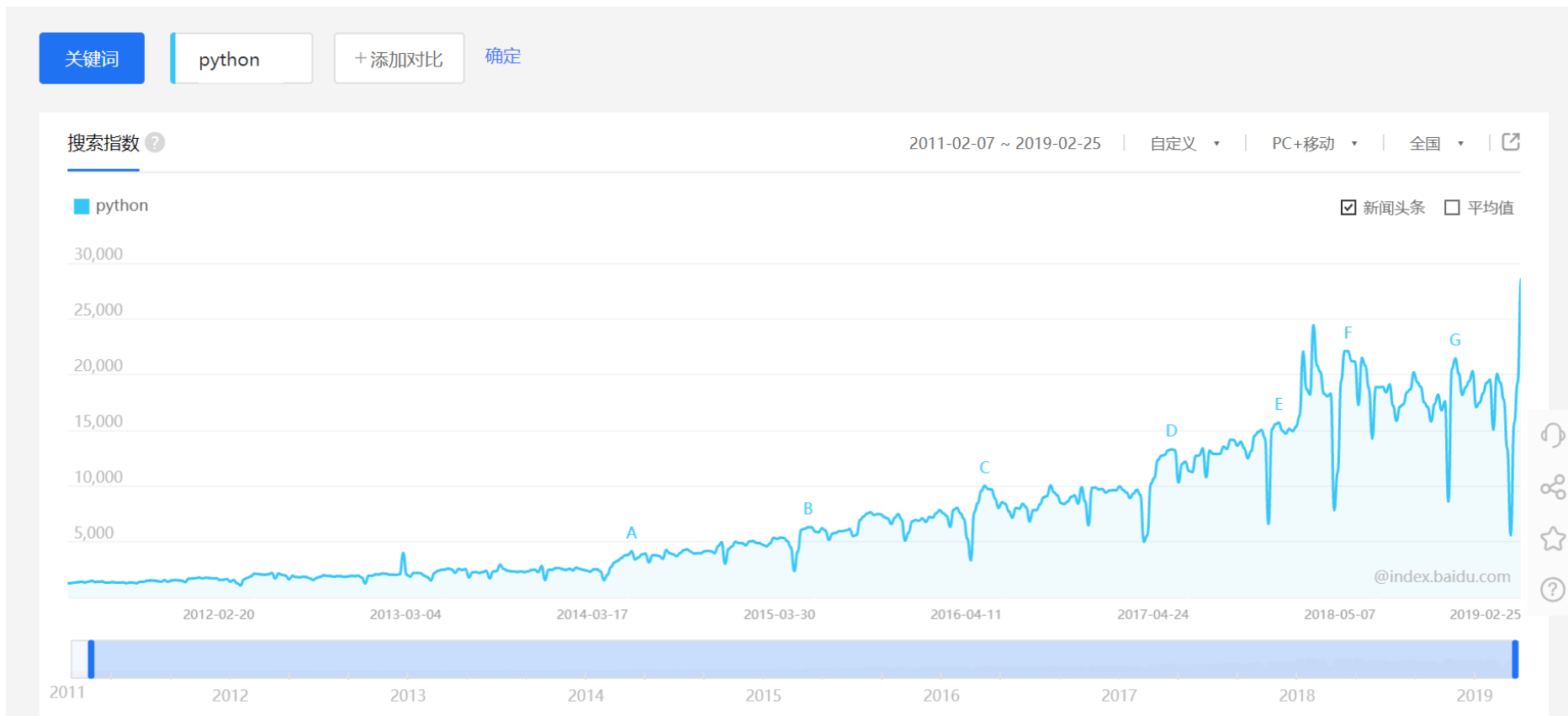




# Python的热度



## 整体热度不断提升



<http://index.baidu.com/?tpl=trend&word=python>





# Python安装与常用工具包



- Windows系统
- Linux系统
- Mac OS系统
- 通过Anaconda安装Python

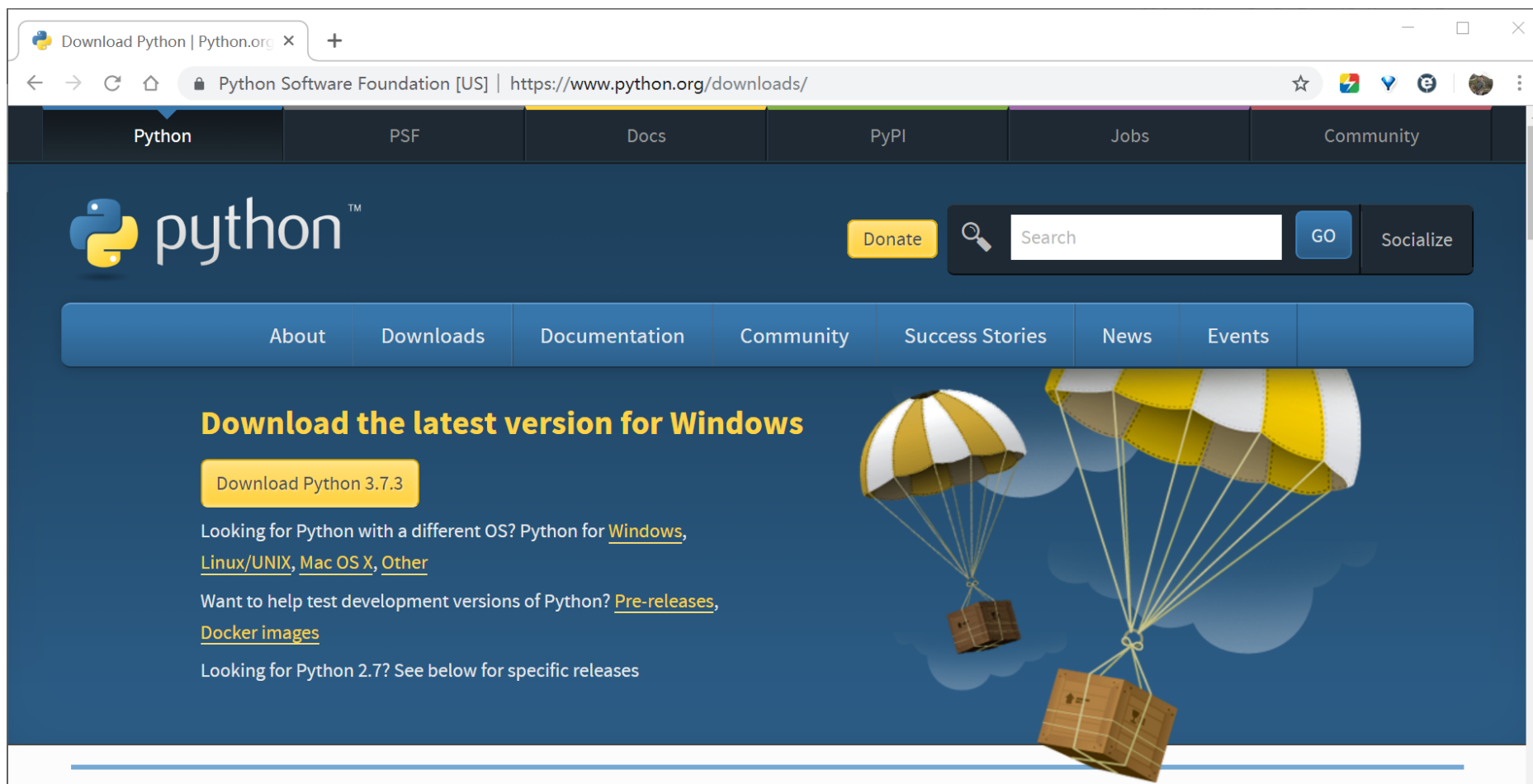




# Window安装Python



(1) 访问<http://www.python.org/download/>, 选择版本。





# Window安装Python



(1) 访问<http://www.python.org/download/>, 选择版本。

Looking for a specific release?

Python releases by version number:

Release version	Release date	Click for more	
<b>Python 3.6.9</b>	July 2, 2019	Download	Release Notes
<b>Python 3.7.3</b>	March 25, 2019	Download	Release Notes
<b>Python 3.4.10</b>	March 18, 2019	Download	Release Notes
<b>Python 3.5.7</b>	March 18, 2019	Download	Release Notes
<b>Python 2.7.16</b>	March 4, 2019	Download	Release Notes
<b>Python 3.7.2</b>	Dec. 24, 2018	Download	Release Notes
<b>Python 3.6.8</b>	Dec. 24, 2018	Download	Release Notes

[View older releases](#)



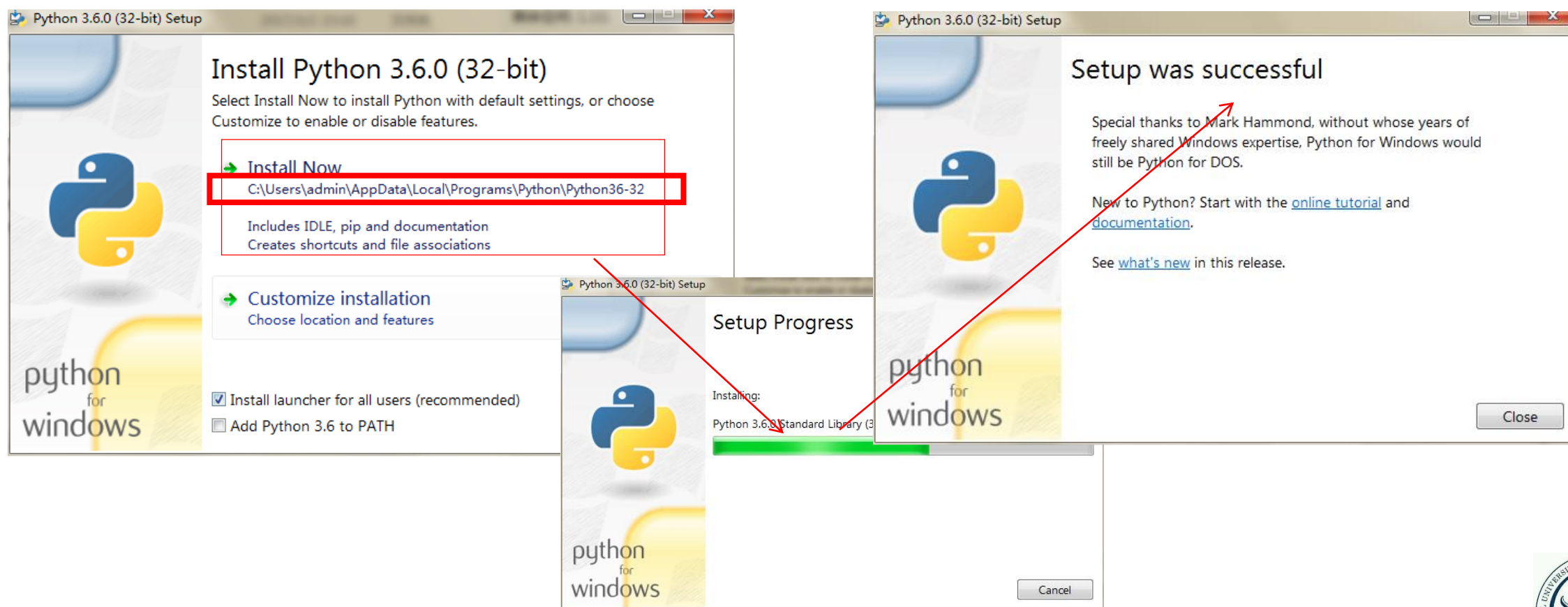




# Window安装Python



## (2) 下载并完成Python的安装。

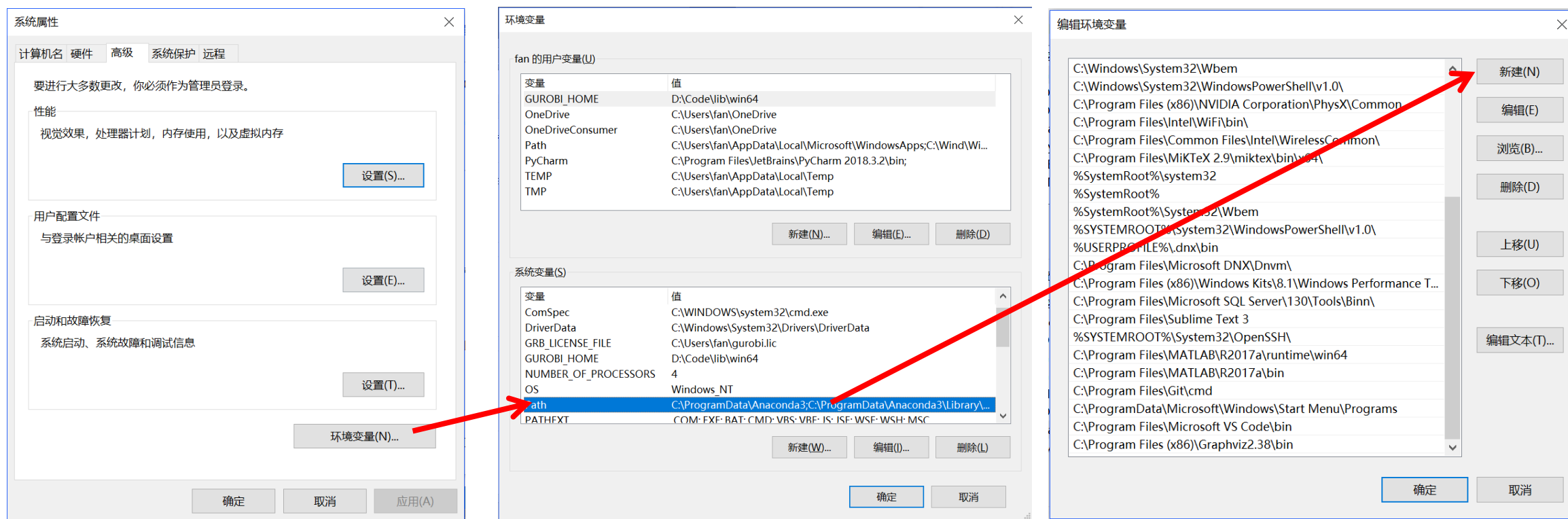




# Window安装Python



## (3) 配置环境变量。右键选择“我的电脑” -> 高级系统设置



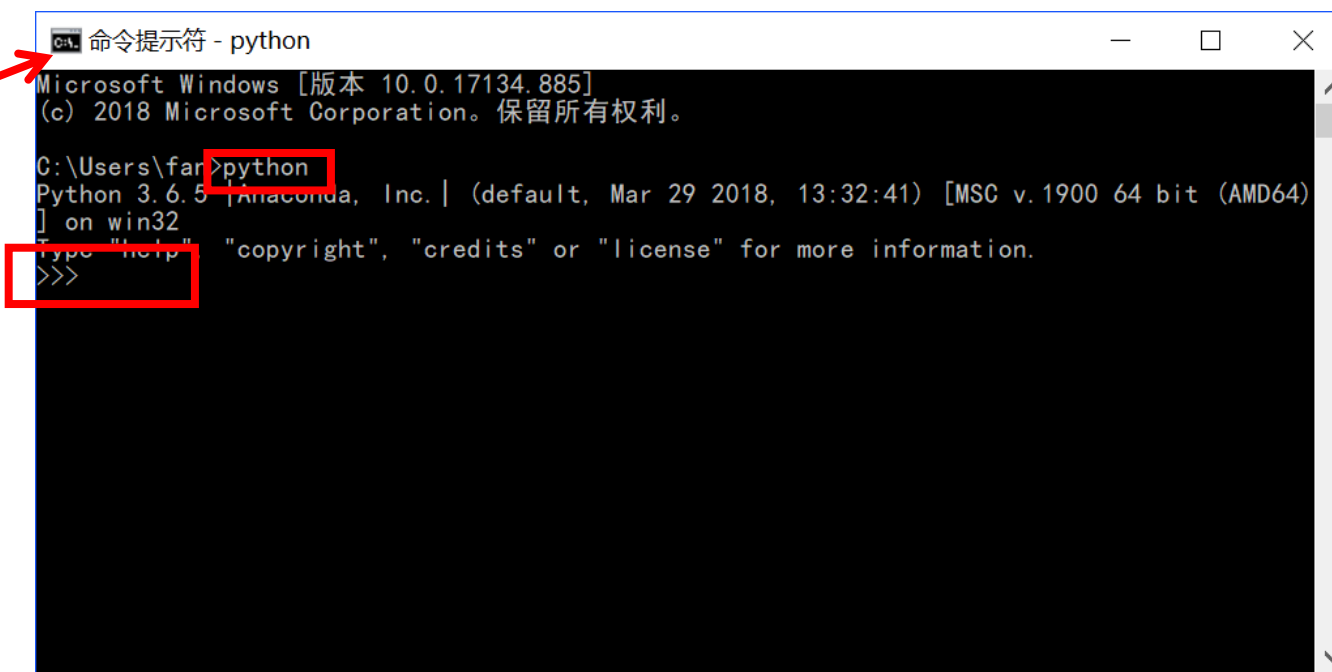
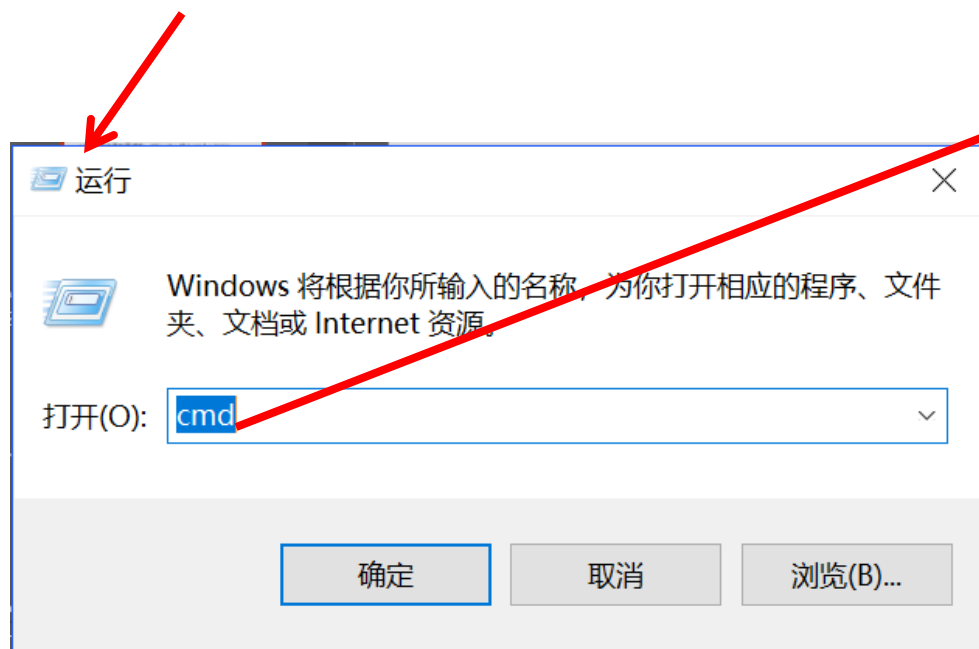


# Window安装Python



## (4) 环境变量配置成功。

Win+R->输入cmd->确定->输入python->Enter





# Window安装Python



## (5) 下载pip

下载网站: <https://pypi.org/project/pip/#files>。

The screenshot shows the PyPI page for pip 19.1.1. The page is titled 'pip 19.1.1' and includes a 'pip install pip' button. Below the title, it states 'The PyPA recommended tool for installing Python packages.' The 'Download files' section contains a table with the following data:

Filename, size & hash	File type	Python version	Upload date
<a href="#">pip-19.1.1-py2.py3-none-any.whl</a> (1.4 MB) <a href="#">SHA256</a>	Wheel	py2.py3	May 6, 2019
<a href="#">pip-19.1.1.tar.gz</a> (1.3 MB) <a href="#">SHA256</a>	Source	None	May 6, 2019





# Window安装Python



(6) 解压文件后，在Path最后面添加Scripts文件路径。再次打开控制台，输入pip list，出现Package和Version等信息表示pip安装成功。

```
C:\Windows\System32\cmd.exe
Microsoft Windows [版本 10.0.17134.829]
(c) 2018 Microsoft Corporation。保留所有权利。

C:\WINDOWS\system32>pip list
Package                                Version
-----
alabaster                              0.7.10
anaconda-clean                         1.0
anaconda-client                       1.6.3
anaconda-navigator                    1.6.2
anaconda-project                      0.6.0
argcomplete                           1.0.0
asn1crypto                            0.22.0
astroid                                1.4.9
astropy                                1.3.2
Babel                                  2.4.0
backports.shutil-get-terminal-size    1.0.0
beautifulsoup4                        4.6.0
bitarray                               0.8.1
blaze                                  0.10.1
bleach                                 1.5.0
bokeh                                  0.12.5
```





# Linux安装Python



- (1) 打开WEB浏览器访问<http://www.python.org/download/>
- (2) 选择适用于Unix/Linux的源码压缩包。
- (3) 下载及解压压缩包。
- (4) 如果你需要自定义一些选项修改Modules/Setup。
- (5) 执行 `./configure` 脚本。
- (6) `make`
- (7) `make install`





# Mac OS安装Python



Mac系统都自带有Python环境，我们可以在终端输入  
**python**命令就可以运行。

```
itcast — python — 80x11
Last login: Fri Feb 24 17:46:07 on console
[itcastdeiMac-2:~ itcast$ python
Python 2.7.10 (default, Oct 23 2015, 19:19:21)
[GCC 4.2.1 Compatible Apple LLVM 7.0.0 (clang-700.0.59.5)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> 
```







# Mac OS安装Python



如果我们希望安装其他版本的Python，只需要访问网站  
<http://www.python.org/download/>，下载最新版本的  
dmg文件，双击按照提示完成安装即可。

```
itcast — Python — 76x9
Last login: Wed Mar  8 16:03:15 on ttys000
[itcastdeiMac-2:~ itcast$ python3
Python 3.5.2 (v3.5.2:4def2a2901a5, Jun 26 2016, 10:47:25)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> █
```







# Anaconda安装Python



前面的安装方法，并没有安装丰富的Python工具包，而且由于包的依赖关系和系统兼容性问题，很多Python包的安装会很麻烦。

对于主要进行数据分析数据挖掘的同学们，强烈推荐通过Anaconda进行Python安装。下载地址：

<https://www.anaconda.com/distribution/#download-section>





# Anaconda安装Python



Windows



macOS



Linux

## Anaconda 2019.03 for macOS Installer

### Python 3.7 version

Download

64-Bit Graphical Installer (637 MB)

64-Bit Command Line Installer (542 MB)

### Python 2.7 version

Download

64-Bit Graphical Installer (624 MB)

64-Bit Command Line Installer (530 MB)

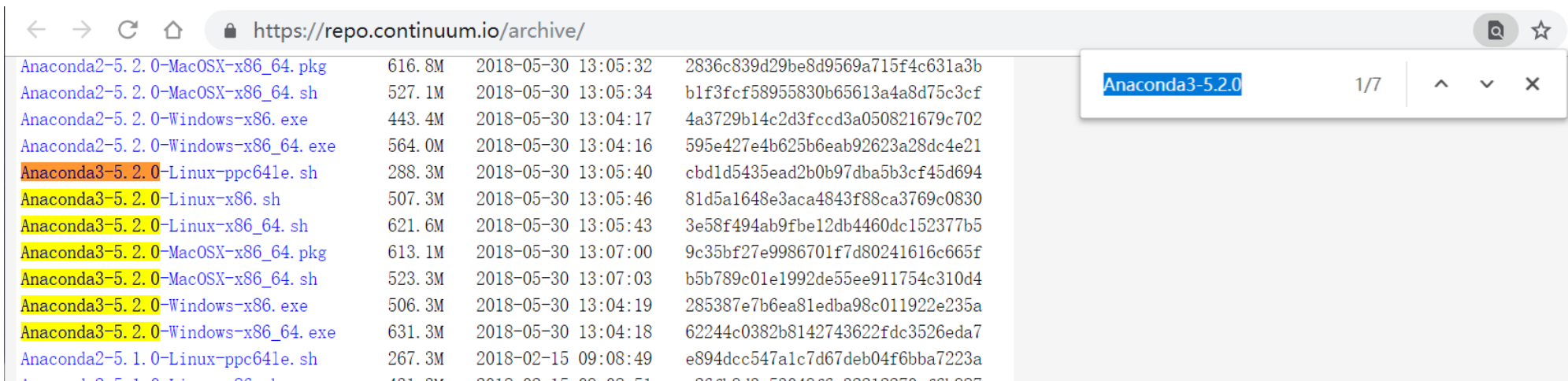




# Anaconda安装Python



官网默认提供最新版本的Ananconda，本课程我们采用Python3.6，对应Anaconda版本为Anaconda3-5.2.0，下载地址：<https://repo.continuum.io/archive/>，请根据自己系统情况选择合适版本。

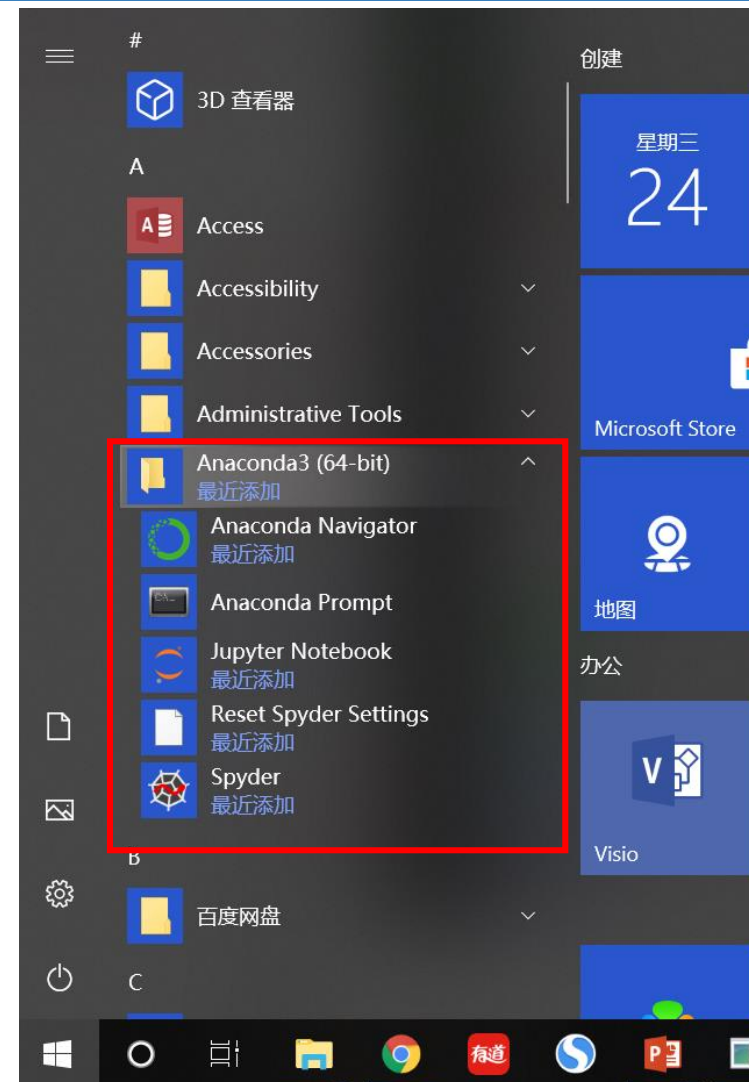
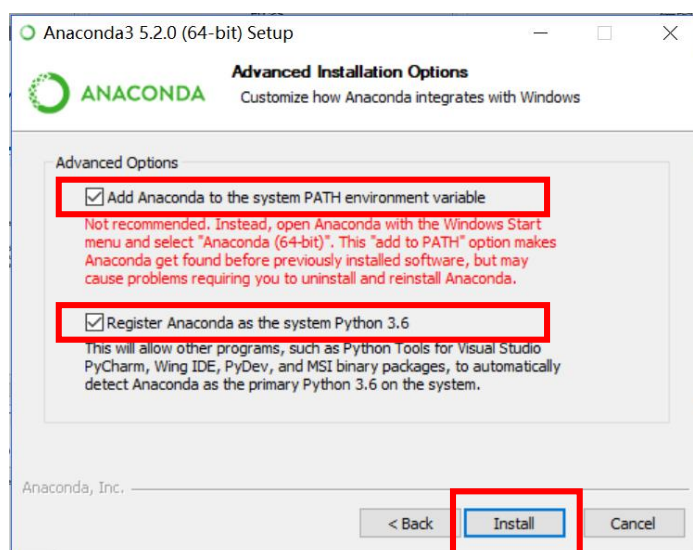
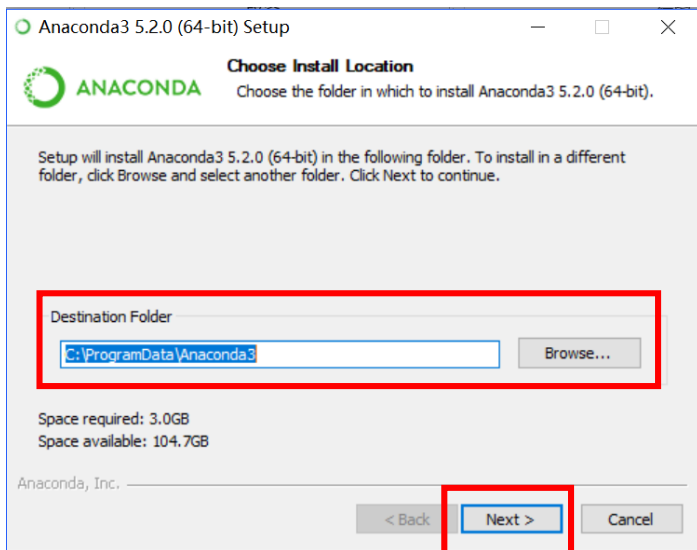
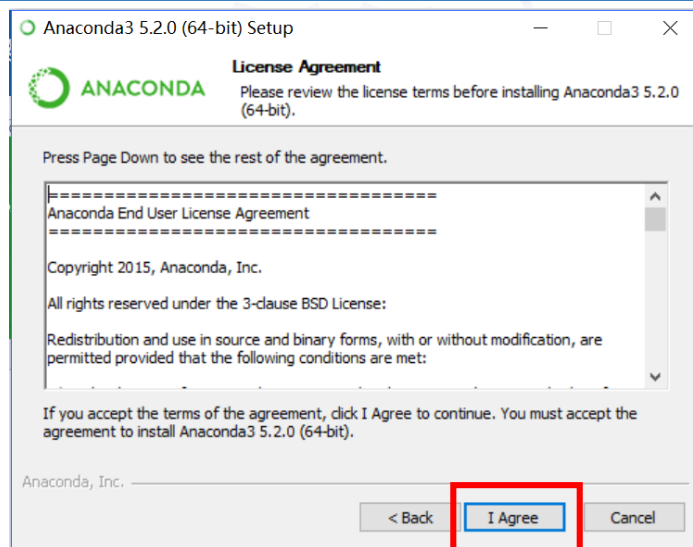
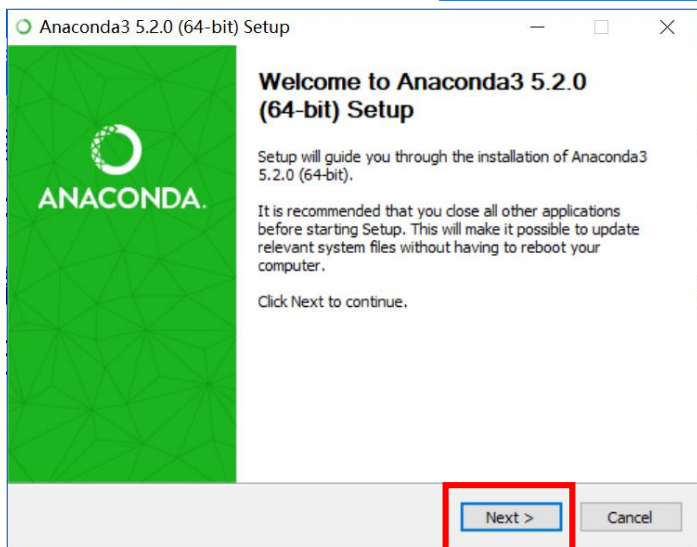


File Name	Size	Date	Time	Hash
Anaconda2-5.2.0-MacOSX-x86_64.pkg	616.8M	2018-05-30	13:05:32	2836c839d29be8d9569a715f4c631a3b
Anaconda2-5.2.0-MacOSX-x86_64.sh	527.1M	2018-05-30	13:05:34	b1f3fcf58955830b65613a4a8d75c3cf
Anaconda2-5.2.0-Windows-x86.exe	443.4M	2018-05-30	13:04:17	4a3729b14c2d3fccd3a050821679c702
Anaconda2-5.2.0-Windows-x86_64.exe	564.0M	2018-05-30	13:04:16	595e427e4b625b6eab92623a28dc4e21
Anaconda3-5.2.0-Linux-ppc64le.sh	288.3M	2018-05-30	13:05:40	cbd1d5435ead2b0b97dba5b3cf45d694
Anaconda3-5.2.0-Linux-x86.sh	507.3M	2018-05-30	13:05:46	81d5a1648e3aca4843f88ca3769c0830
Anaconda3-5.2.0-Linux-x86_64.sh	621.6M	2018-05-30	13:05:43	3e58f494ab9fbc12db4460dc152377b5
Anaconda3-5.2.0-MacOSX-x86_64.pkg	613.1M	2018-05-30	13:07:00	9c35bf27e9986701f7d80241616c665f
Anaconda3-5.2.0-MacOSX-x86_64.sh	523.3M	2018-05-30	13:07:03	b5b789c01e1992de55ee911754c310d4
Anaconda3-5.2.0-Windows-x86.exe	506.3M	2018-05-30	13:04:19	285387e7b6ea81edba98c011922e235a
Anaconda3-5.2.0-Windows-x86_64.exe	631.3M	2018-05-30	13:04:18	62244c0382b8142743622fdc3526eda7
Anaconda2-5.1.0-Linux-ppc64le.sh	267.3M	2018-02-15	09:08:49	e894dcc547a1c7d67deb04f6bba7223a





# Anaconda安装Python

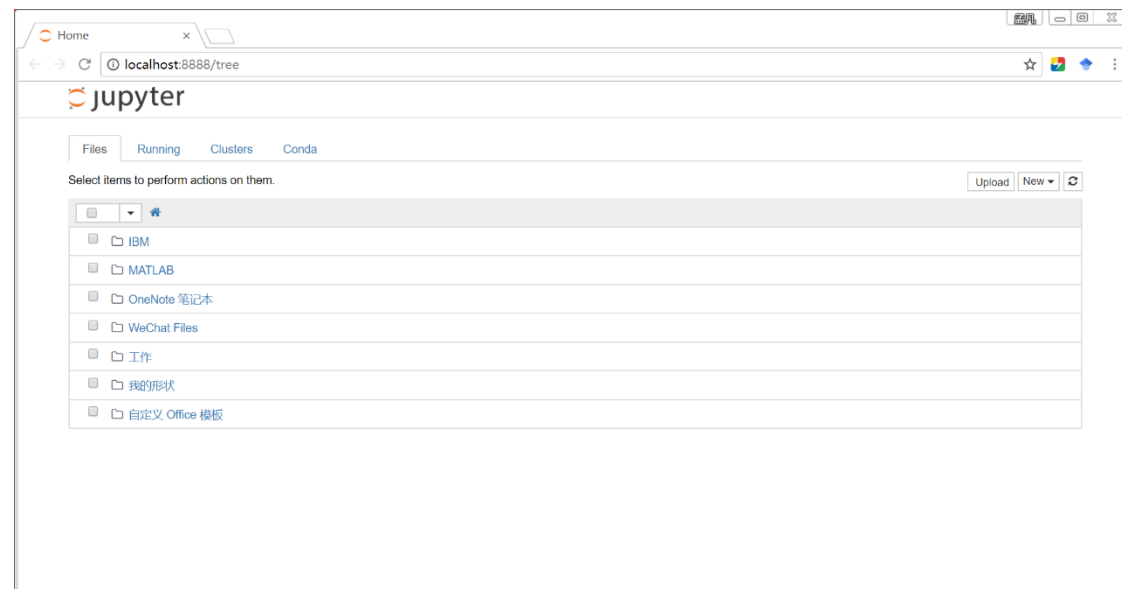
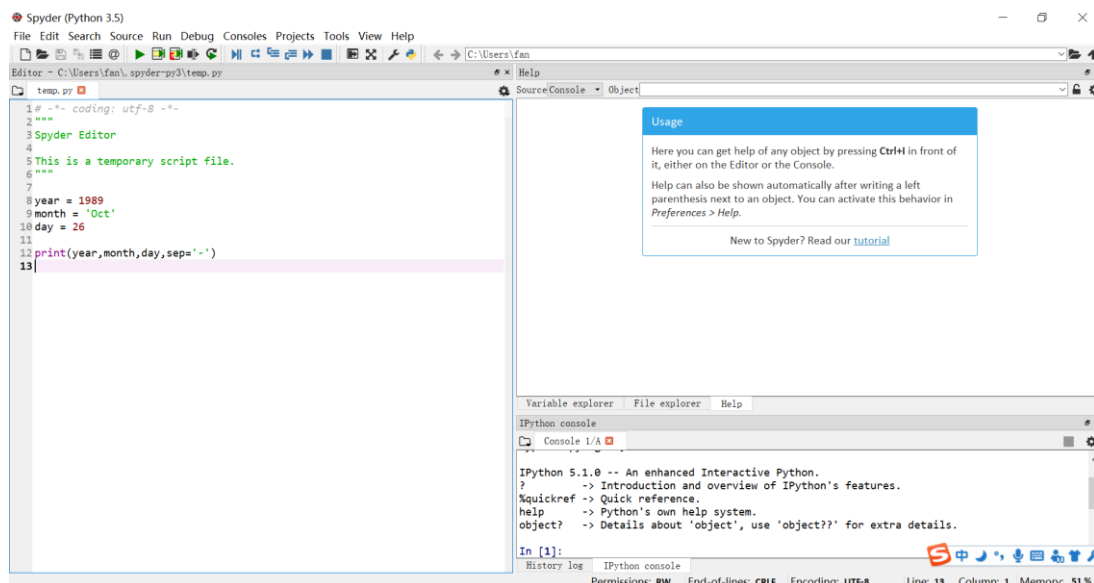




# Anaconda安装Python



除了丰富的工具包，Anaconda中还集成有丰富的开发工具，包括集成开发环境Spyder，交互式开发环境Jupyter。





# 数据挖掘常用工具包



**Python进行数据挖掘的常用工具包主要有：**

- **NumPy**
- **Pandas**
- **matplotlib**
- **statsmodels**
- **SciPy**
- **scikit-learn**
- **.....**





**NumPy**(全称为Numerical Python)是Python中科学计算的基本包。Python数据分析的绝大多数内容基于NumPy和在NumPy之上构建的库。其提供众多功能。

除了NumPy在Python中添加的快速数组处理功能之外，它在数据分析方面的主要目的之一是作为**数据在算法之间传递的主要容器**。







**Pandas**提供丰富的数据结构和功能，旨在使结构化数据快速、简单、富有表现力。它是使Python成为一个强大且高效的数据分析环境的关键因素之一。Pandas基于两种数据类型：Series与DataFrame。

- Series是一个一维的数据类型，其中每一个元素都有一个标签。Series类似于Numpy中元素带标签的数组。其中，标签可以是数字或者字符串。
- DataFrame是一个二维的表结构。Pandas的DataFrame可以存储许多种不同的数据类型，并且每一个坐标轴都有自己的标签。







## Pandas.Series

```
new_player
```

```
Fabregas      40000  
Messi          75000  
Ronaldo        85000  
Rooney         50000  
Van persie     67000  
dtype: int64
```

## Pandas.frame

```
>>> frame  
   total_bill  tip  sex  smoker  day  time  size  
1    16.99    1.01 Female    No   Sun  Dinner    2  
2    10.34    1.66  Male    No   Sun  Dinner    3  
3    21.01    3.5  Male    No   Sun  Dinner    3  
4    23.68    3.31  Male    No   Sun  Dinner    2  
5    24.59    3.61 Female    No   Sun  Dinner    4  
6    25.29    4.71  Male    No   Sun  Dinner    4  
7     8.77     2    Male    No   Sun  Dinner    2  
8    26.88    3.12  Male    No   Sun  Dinner    4  
9    15.04    1.96  Male    No   Sun  Dinner    2  
10   14.78    3.23  Male    No   Sun  Dinner    2
```





# matplotlib



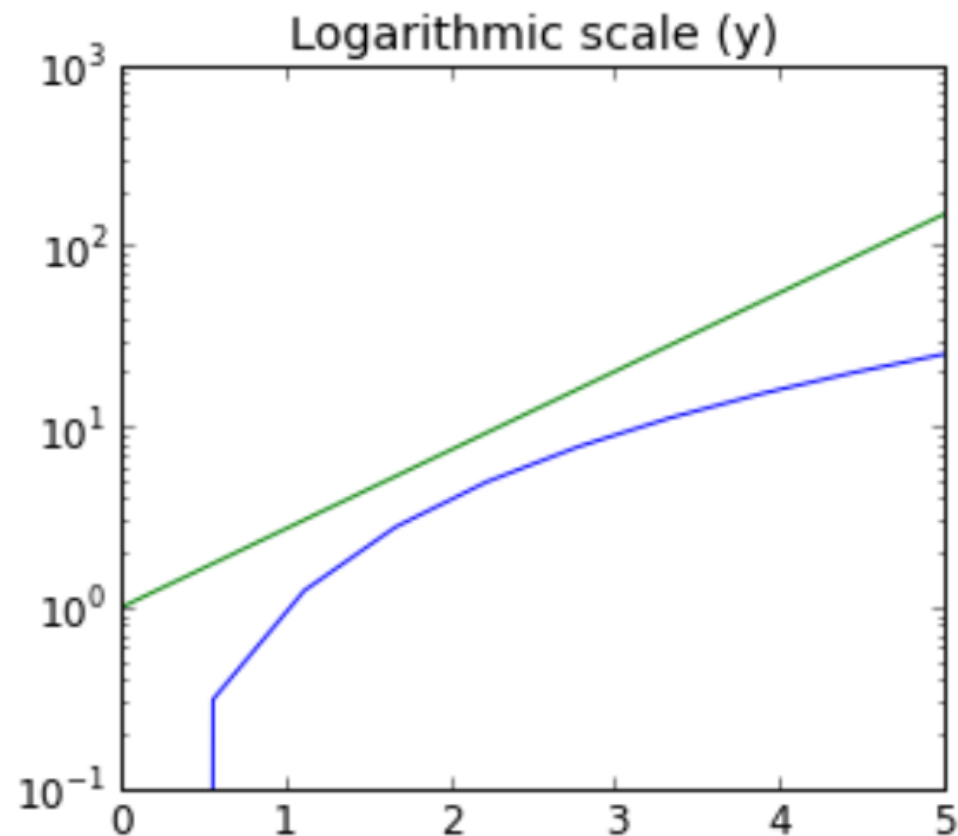
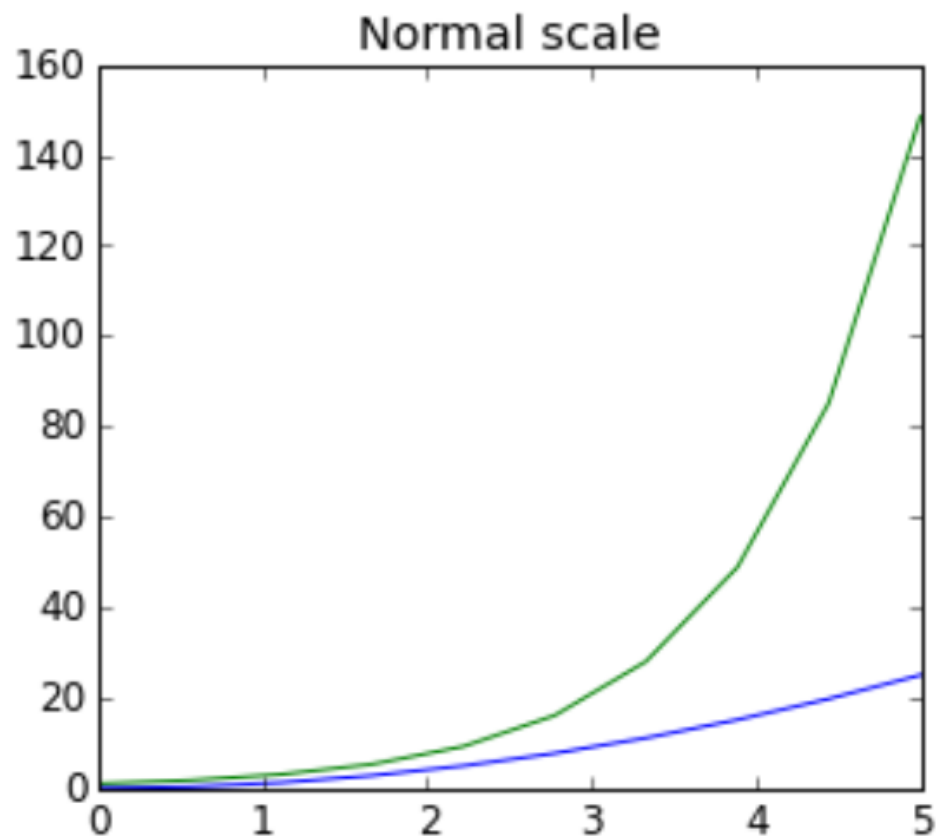
**matplotlib**是最流行的用于生成绘图和其他2D和3D数据可视化的Python库。它最初是由John d . Hunter(JDH)创建的，现在由一个大型开发团队维护。它非常适合于创建用于发布和展示的图形。

它与IPython集成得很好，从而为绘图和探测数据提供了一个舒适的交互式环境。这些图也具有互动性；可以在plot窗口中使用工具栏来放大图的一部分。



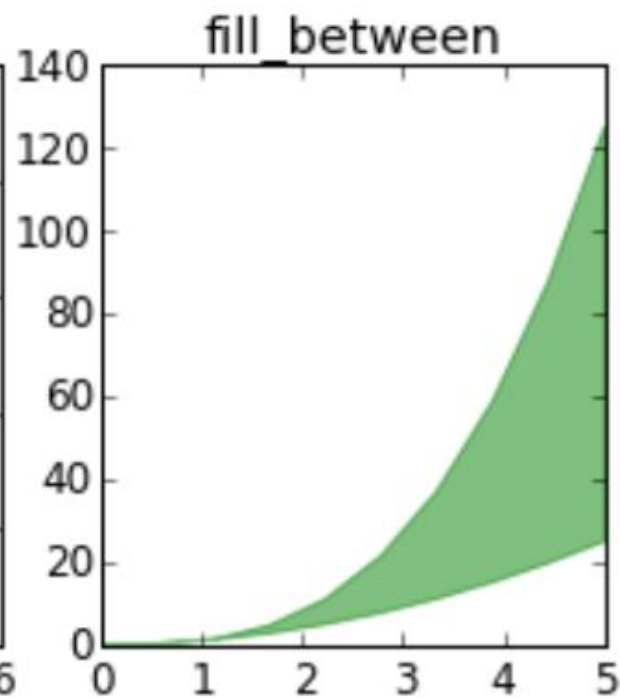
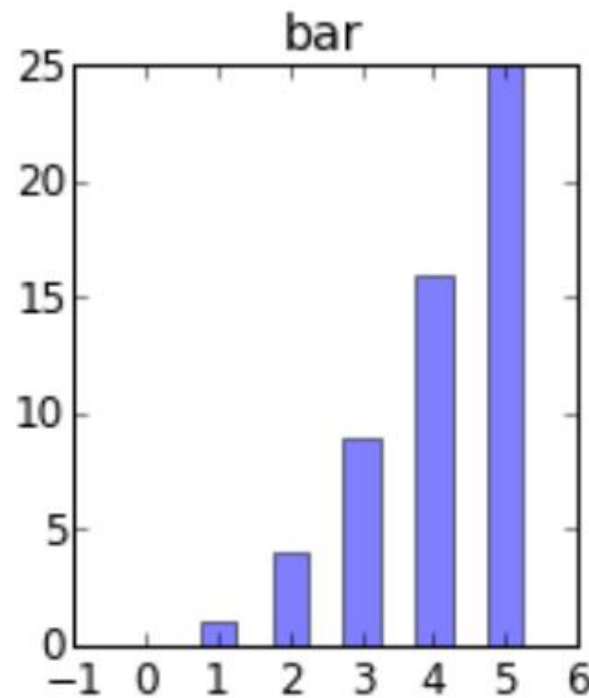
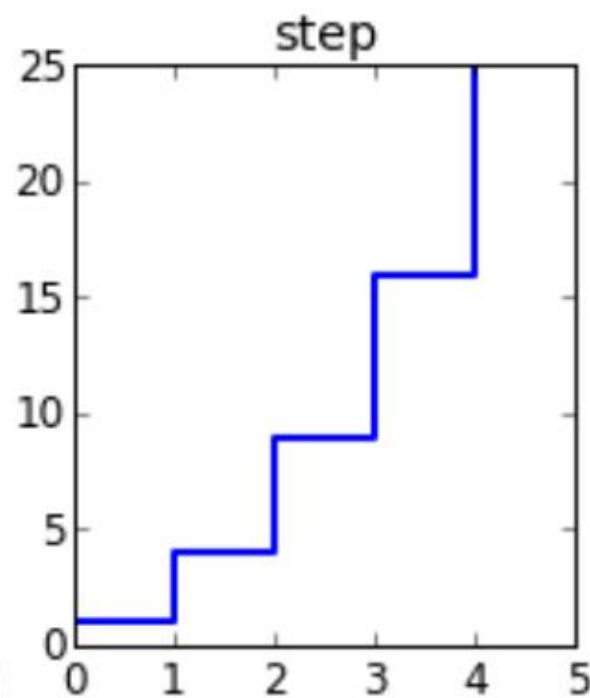
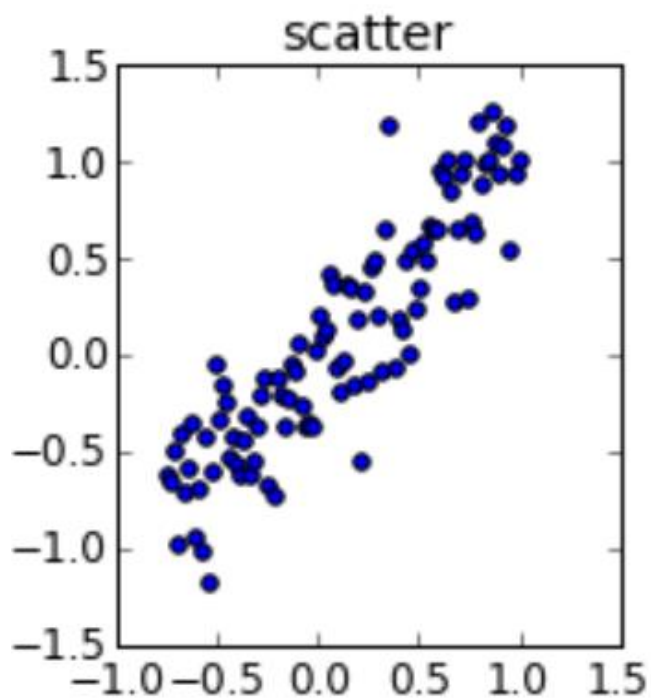


# matplotlib



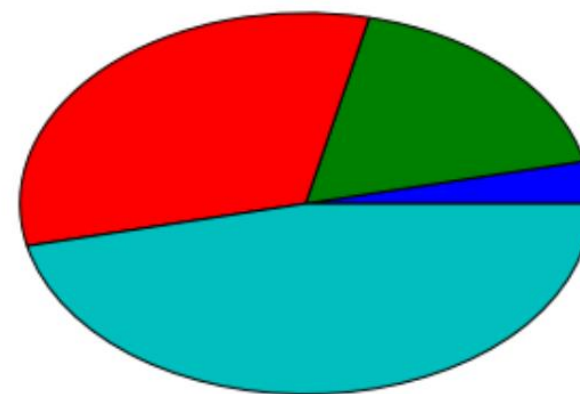
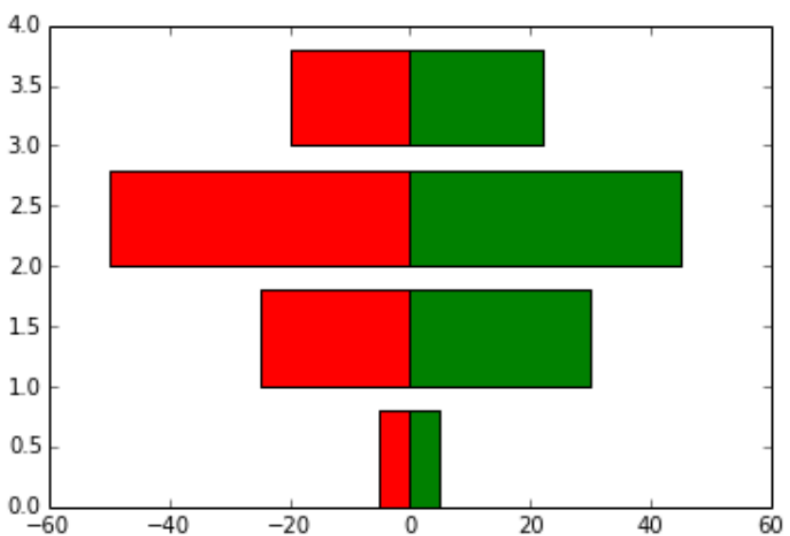
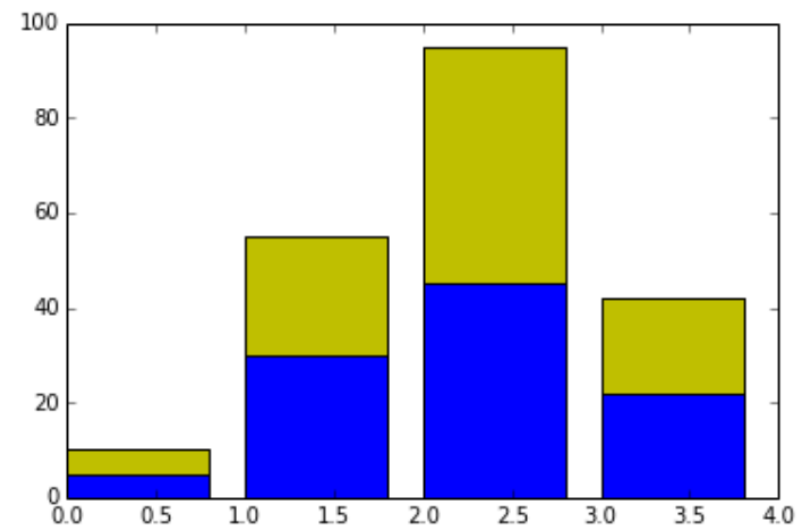
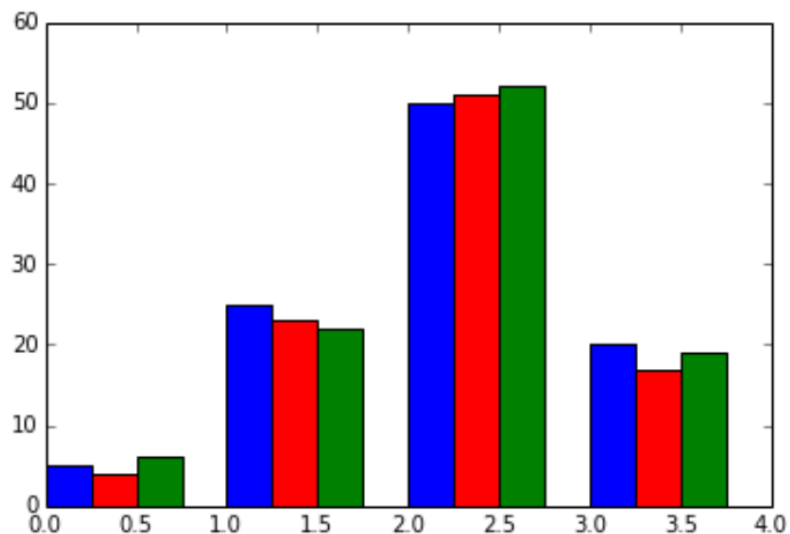


# matplotlib



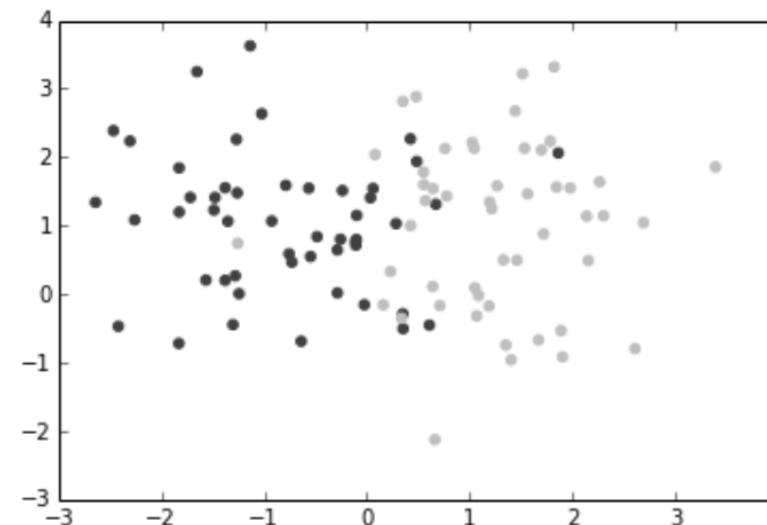
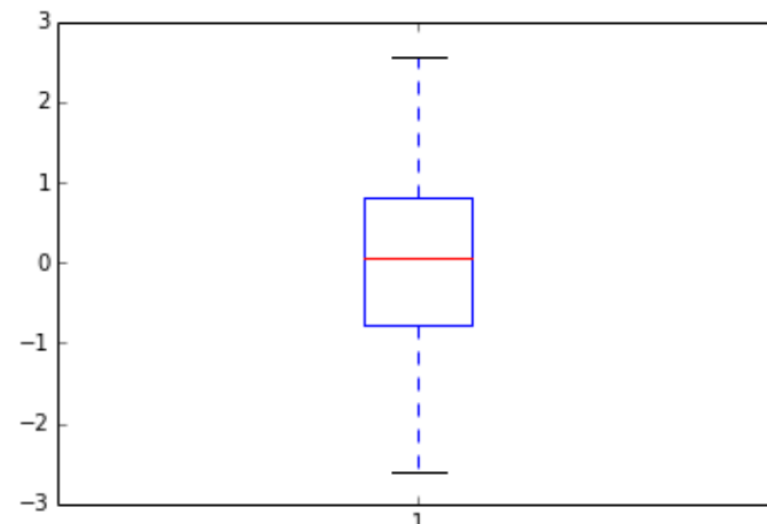
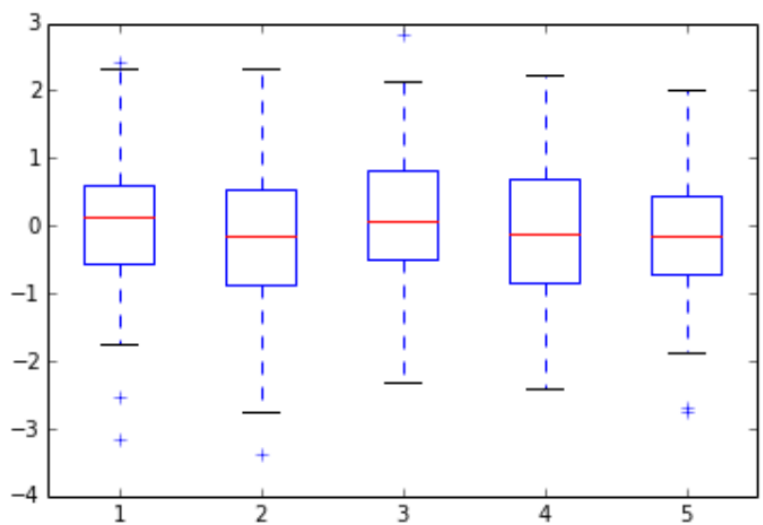
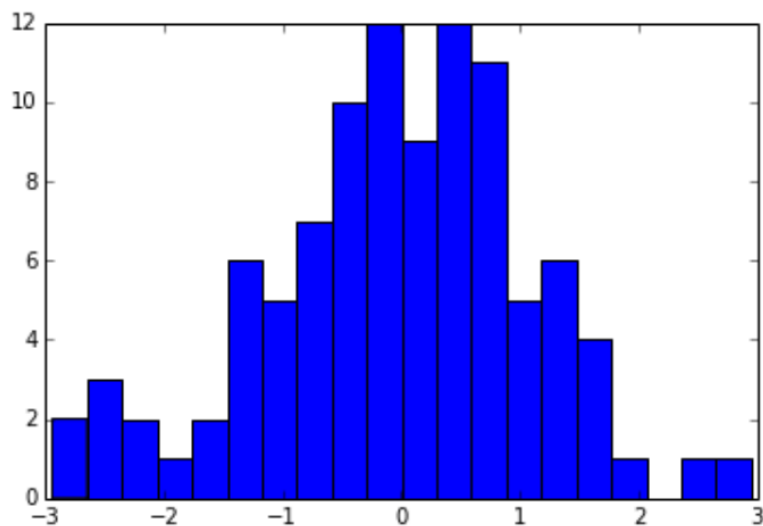


# matplotlib



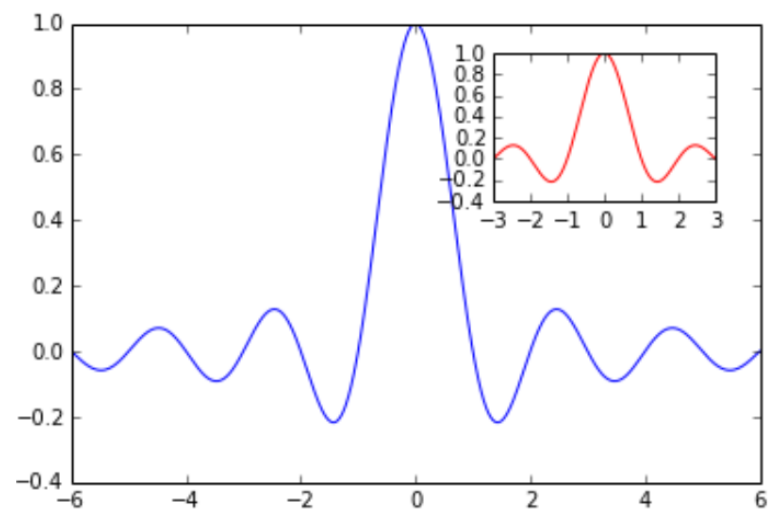
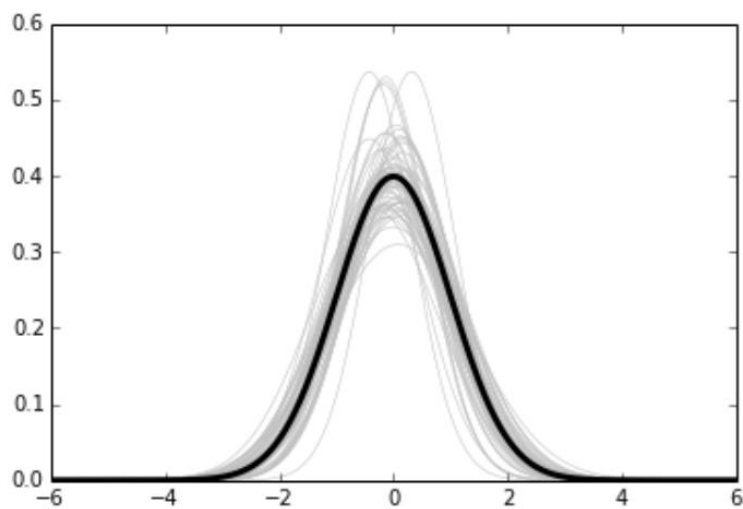
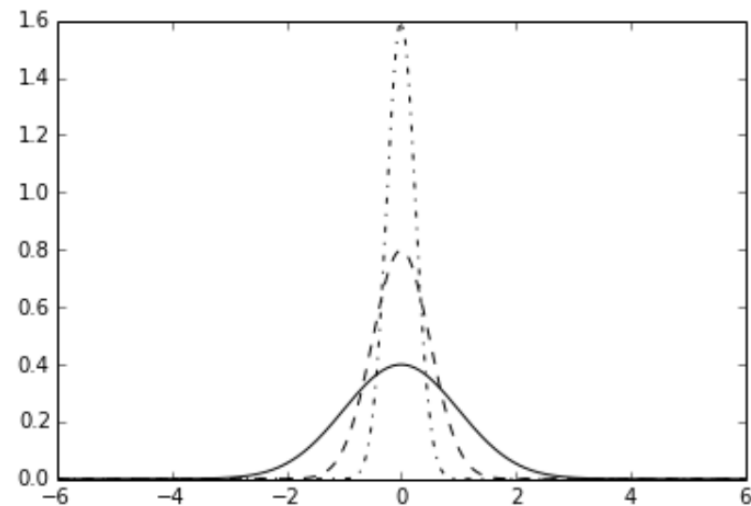
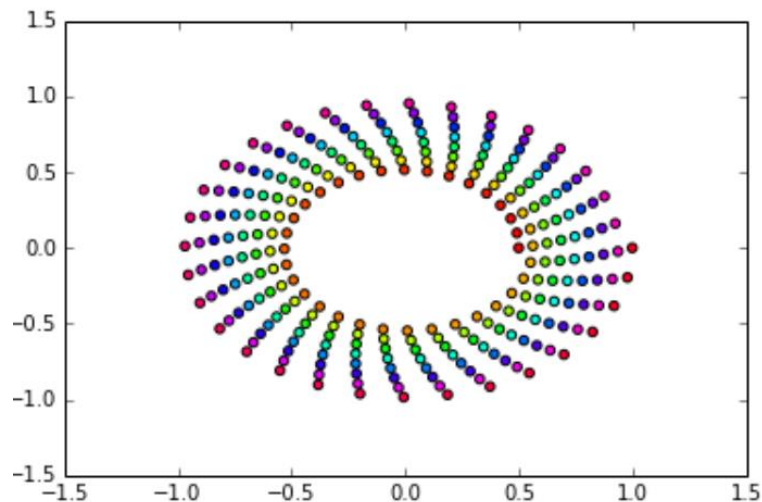


# matplotlib





# matplotlib





**Statsmodels**是Python的统计建模和计量经济学工具包，包括一些描述统计、统计模型估计和推断。主要功能包括：

- Liner regression models: 线性回归模型
- Gneralized linear models: 一般线型模型
- Robust linear models: 鲁棒线性模型
- Discrete choice models: 离散选择模型
- ANOVA: 方差分析模型
- Time series analysis: 时间序列分析
- Nonparametric estimators: 非参检验
- a wide range of statistical tests: 各种统计检验
- 以各种方式输出表格: text, latex, html; 读取各种格式的数据
- 绘图功能







**SciPy** 是基于Numpy构建在科学计算中处理多个不同标准问题域的包的集合。主要包括以下模块包括：

- `scipy.integrate`: 数值积分和微分方程求解器
- `scipy.linalg`: 拓展了`numpy.linalg`中的线性代数和矩阵分解功能
- `scipy.optimize`: 函数优化器(最小化器)和根查找算法
- `scipy.signal`: 信号处理工具
- `scipy.sparse`: 系数矩阵和线性系统求解器
- `scipy.special`: 对于SPECFUN的封装, SPECFUN库实现了许多常见的数学函数
- `scipy.stats`: 标准连续和离散概率分布(密度函数, 采样器, 连续分布函数), 各种统计检验, 和更多的描述性统计
- `scipy.weave`: 使用内联c++代码来加速数组计算的工具

通过结合使用NumPy和SciPy能够实现绝大部分matlab及其工具包的功能。

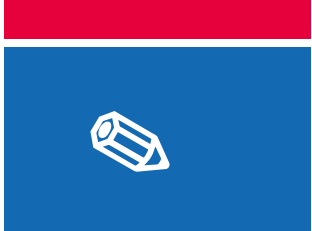




**scikit-learn**是Python的一个开源机器学习模块，它建立在NumPy, SciPy和matplotlib模块之上，实现了大量的机器学习算法。包括：

- **Classification**: 分类 - SVM, nearest neighbors, random forest, logistic regression, etc.
- **Regression**: 回归 - Lasso, ridge regression, etc.
- **Clustering**: 聚类 -  $k$ -means, spectral clustering, etc.
- **Dimensionality reduction**: 降维 - PCA, feature selection, matrix factorization, etc.
- **Model selection**: 模型选择 - Grid search, cross-validation, metrics
- **Preprocessing**: 预处理 - Feature extraction, normalization





## Classification

Identifying to which category an object belongs to.

**Applications:** Spam detection, Image recognition.

**Algorithms:** SVM, nearest neighbors, random forest, ... — Examples

## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** SVR, ridge regression, Lasso, ... — Examples

## Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** k-Means, spectral clustering, mean-shift, ... — Examples

## Dimensionality reduction

Reducing the number of random variables to consider.

**Applications:** Visualization, Increased efficiency

**Algorithms:** PCA, feature selection, non-negative matrix factorization. — Examples

## Model selection

Comparing, validating and choosing parameters and models.

**Goal:** Improved accuracy via parameter tuning

**Modules:** grid search, cross validation, metrics. — Examples

## Preprocessing

Feature extraction and normalization.

**Application:** Transforming input data such as text for use with machine learning algorithms.

**Modules:** preprocessing, feature extraction. — Examples





# Hello Python!



- 进入任意一个文件夹
- 在地址栏中输入: jupyter notebook
- 回车
- 即可进入jupyter环境
  
- [Hello Python](#)

