



# Python 数据科学与统计分析

作者：陈震

组织：西南大学

时间：January 22, 2021

版本：1.00



*Victory won't come to us unless we go to it. — M. Moore*

# 目 录

<b>1</b>	<b>Python 概述与安装</b>	<b>1</b>
1.1	Python 的诞生与应用 . . . . .	1
1.2	安装 Python . . . . .	2
1.2.1	Anaconda + Spyder . . . . .	2
1.2.2	Python + Pycharm . . . . .	5
1.3	数据分析常用 Python 工具包的安装 . . . . .	11
1.3.1	用 Anaconda 安装 . . . . .	11
1.3.2	用 pip 安装 . . . . .	12

# 第 1 章 Python 概述与安装

## 学习目标

- ❑ 1. 了解 Python 的历史与发展
- ❑ 2. 掌握 Python 的安装
- ❑ 3. 掌握 Python 包的安装

## 1.1 Python 的诞生与应用

Python 的诞生时间非常早，第一版 Python 发行于 1991 年，甚至比另一个编程语言 Java 的历史都早。在大部分时间内，Python 一直作为一个小众的编程语言，并没有大规模流行起来。近些年来，随着大数据技术、人工智能技术的发展和普及，“简洁、具有良好扩展性”的 Python 非常契合大数据与人工智能技术对编程语言的要求，超越其他编程语言迅速崛起。Python 分别在 2010 年，2018 年，2020 年被全球知名的编程语言流行度排行榜网站 TIOBE 评为“年度最佳编程语言”，并长期位于各年或各月流行编程语言排行榜的前三名。中国的网络上甚至产生了一句流行语：“人生苦短，我用 Python”。

Python 的创始人为荷兰人吉多·范罗苏姆（Guido van Rossum），在 1989 年的圣诞节期间，吉多·范罗苏姆为了打发时间，决心开发一个新的脚本解释程序，作为 ABC 语言的继承，于是 Python 诞生了。之所以选择 Python 作为名字，是由于吉多·范罗苏姆非常喜欢一部 BBC 电视剧—Monty Python’s Flying Circus（中文译名：蒙提·派森的飞行马戏团）。Python 的英文意思为蟒蛇，这也是为什么一些 Python 相关的软件或书籍用蟒蛇作为图标的原因。

Python 的设计哲学是“优雅”、“明确”、“简单”，Python 可以说是语法功能最简单的编程语言之一。Python 具有良好的可扩展性，Python 提供了丰富的接口和工具，以便程序员能够轻松地使用 C、C++、Cython 来编写扩展模块，因此，有很多人把 Python 作为一种“胶水语言”使用。Python 是完全面向对象的语言，函数、模块、数字、字符串都是对象，并且完全支持继承、重载、派生、多重继承等，Python 还支持重载运算符。并且随着计算机速度运算速度越来越快，Python 运算速度慢的缺点逐渐可以忽略，因此，许多网站开始使用 Python 开发，许多专业软件也开始支持 Python 调用。

Python 的应用范围包括 Web 开发网站与网络爬虫，GUI 开发桌面软件，科学计算等。使用 Python 编写的著名应用包括：

- Youtube — 视频分享网站
- Dropbox — 文件分享软件
- Instagram — 图片分享软件

- 豆瓣网 — 国内著名的图书、唱片、电影评论网站
- 知乎网 — 国内著名的问答网站
- 果壳网 — 国内著名的泛科技主题网站

Python 2.0 于 2000 年 10 月 16 日发布, Python 3.0 于 2008 年 12 月 3 日发布, 此版不完全兼容之前的 Python 源代码。目前的主要 Python 版本为 Python3, 本书的所有 Python 程序都是 Python3 版本。

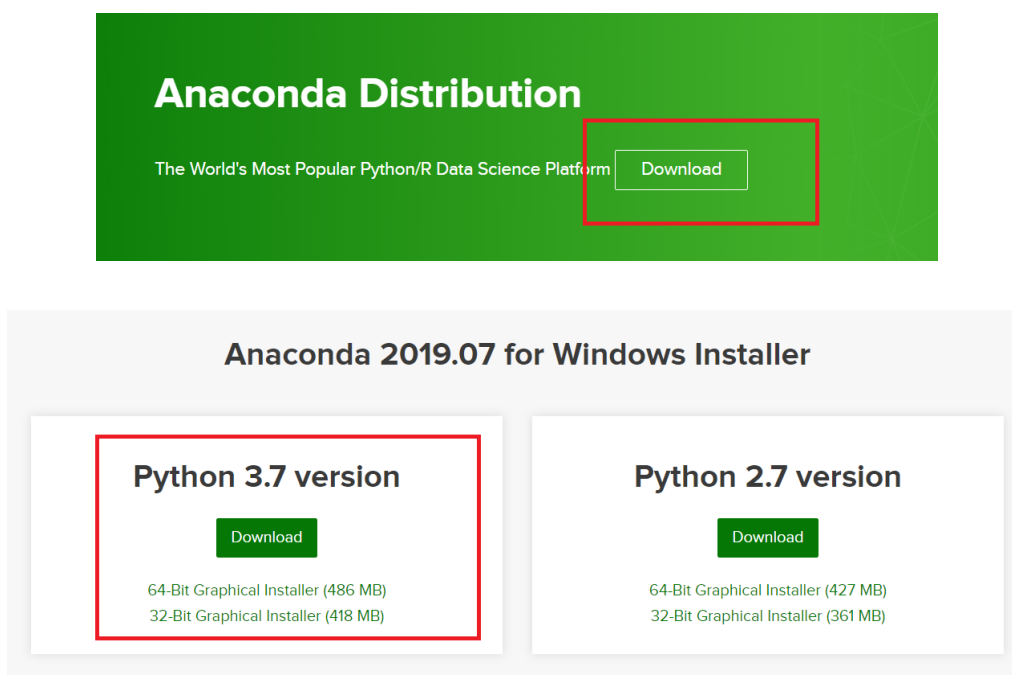
## 1.2 安装 Python

使用 Python 时一般有两种方式: 一是下载 Anaconda (自带 Python, 编辑器 Spyder 以及许多常用包), 二是下载 Python 和编辑器 Pycharm, 从实践经验来看, Pycharm 功能比 Spyder 更多, 对程序员来说更加便捷, 但 Anaconda 对新手比较友好, 很多常用包已经自动安装好了, 不用自己专门再下载。若对 Python 有一定经验, 需要经常用 Python 编写程序, 推荐使用 Pycharm; 若是刚刚接触 Python, 则推荐使用 Anaconda。

### 1.2.1 Anaconda + Spyder

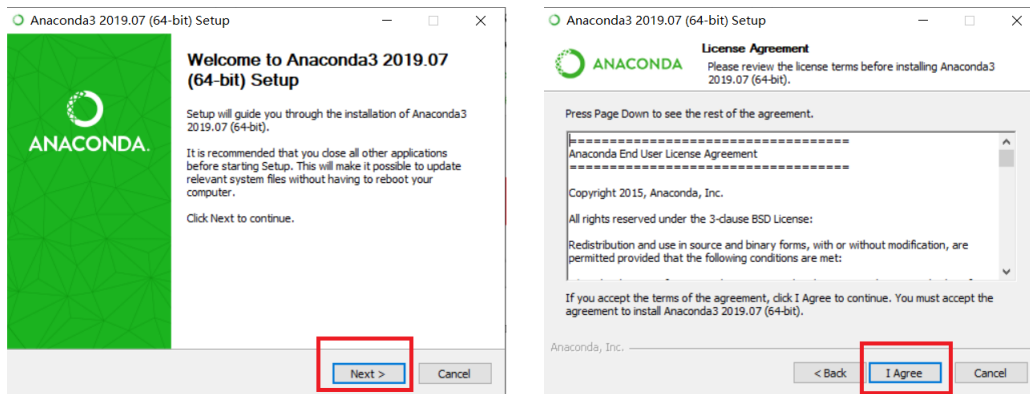
Anaconda 包括 Spyder 编译器以及一大堆安装好的工具包, 比如: numpy、pandas、matplotlib、scipy, 在 Anaconda 里面还可以使用 R 语言、Notebook 等。

Anaconda 下载的官方网址为: <https://www.anaconda.com/distribution>。点击 Download 选择自己操作系统的 Python 3 版本下载 (本书的软件安装以 Windows 系统为例)。

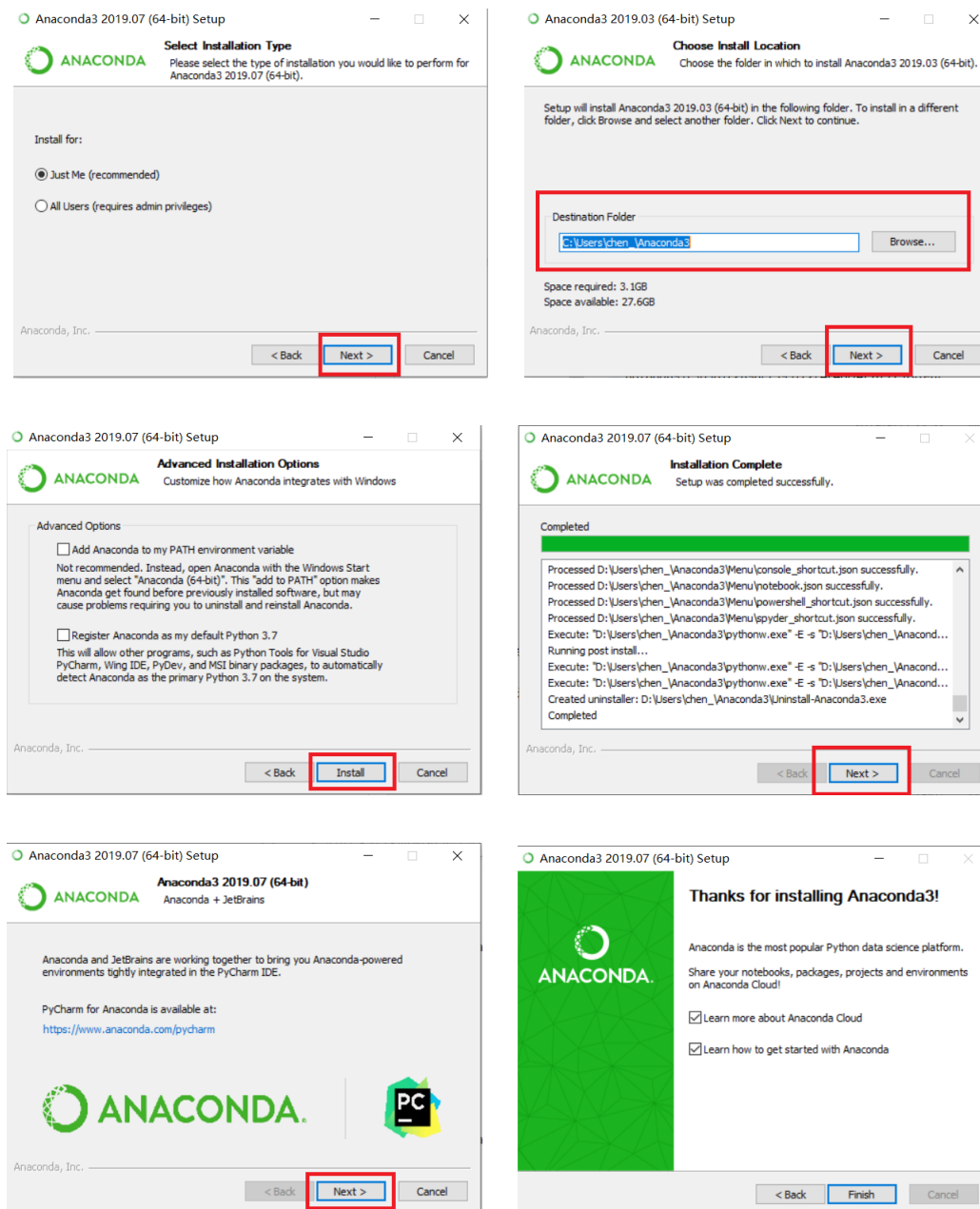


下载完毕后, 双击 exe 文件安装, 安装过程参照下图:



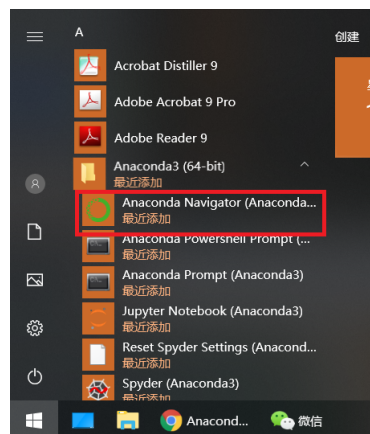


安装目录默认的是 C 盘，可以通过点击 Browse 更改为自定义位置。

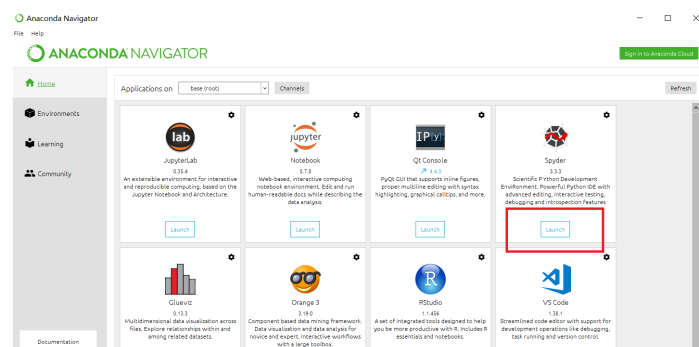


安装完毕后，在 Windows 菜单栏里找到名为“Anaconda-Navigator”的图标，点击打

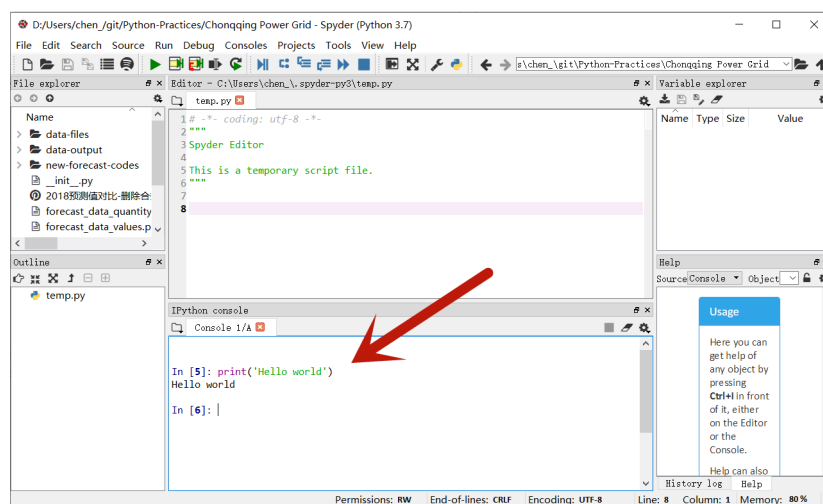
开 Anaconda。



启动（Launch）Spyder，就能够编译 Python 了。



Spyder 的界面非常类似 Matlab，交互式工具 IPython console 就在窗口界面中。我们在 IPython 中输入 `print('Hello world')`，目的是让 Python 把 Hello World 打印到屏幕上，测试 Spyder 的运行。



## 1.2.2 Python + Pycharm

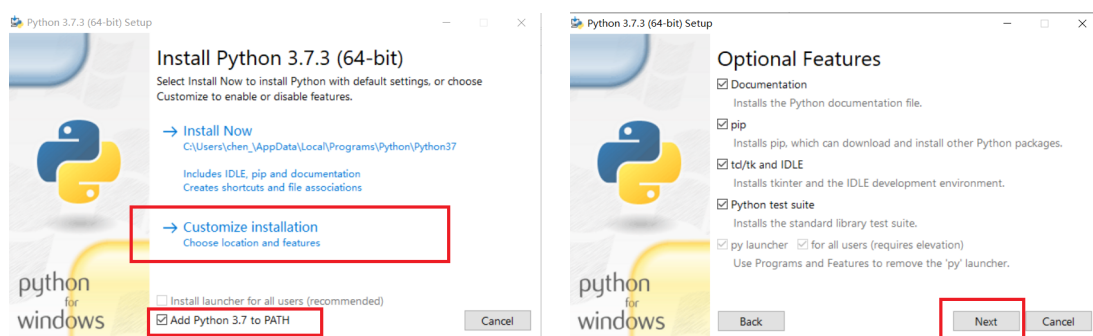
另外一种方式是使用 Pycharm 编写 Python 程序，但由于 Pycharm 并不自带 Python，我们还需要专门把 Python 下载下来。

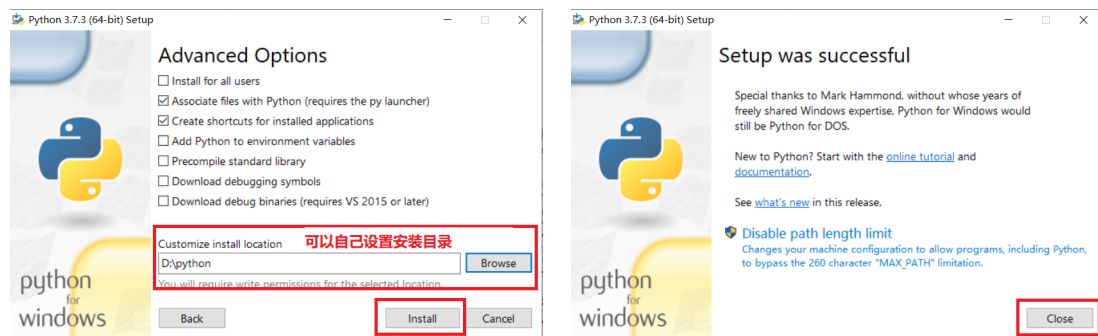
### 1.2.2.1 下载 Python

Python 的官方网址为：<https://www.Python.org>。打开网址后，从 Download 里面可以找到不同操作系统的 Python 版本下载。下面我们以 Windows 系统为例讲述 Python 的安装。



下载完成后，双击 exe 文件安装，切记要勾选 Add Python 3.7 to PATH，然后点击 Install Now（默认安装目录，默认安装 pip 等）或 Customize installation（可以自己设置安装目录等），我们选择 Customize installation。





为了测试 Python 是否安装成功，打开命令行窗口（win10 系统中敲 win 键，再输入 cmd，然后回车），输入 `Python --version`（注意是两个短横线），若能显示 Python 的版本信息，则代表 Python 已正确安装。

```
命令提示符
Microsoft Windows [版本 10.0.17763.737]
(c) 2018 Microsoft Corporation。保留所有权利。

C:\Users\chen_>python --version
Python 3.7.3

C:\Users\chen_>
```

### 1.2.2.2 下载 Pycharm

Pycharm 的官方网址：<http://www.jetbrains.com/pycharm>，点击 DOWNLOAD 开始安装。

## Download PyCharm

Windows

macOS

Linux

### Professional

For both Scientific and Web Python development. With HTML, JS, and SQL support.

DOWNLOAD

Free trial

### Community

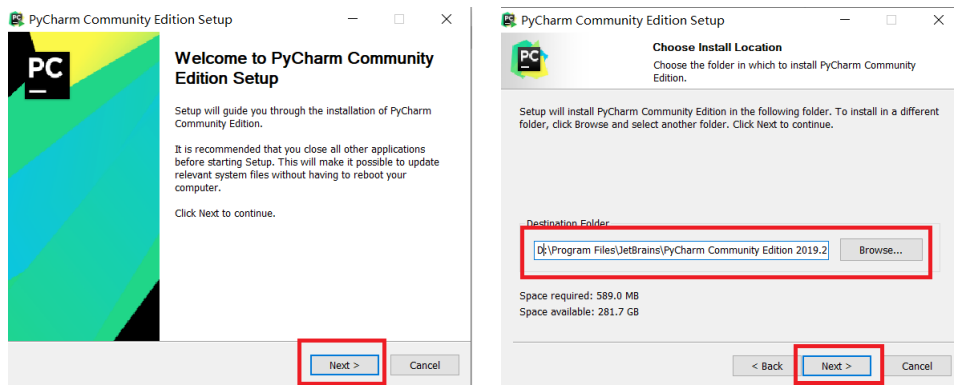
For pure Python development

DOWNLOAD

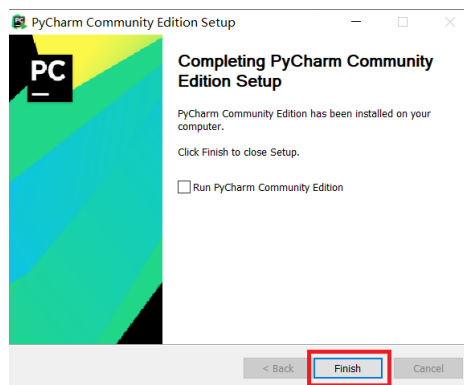
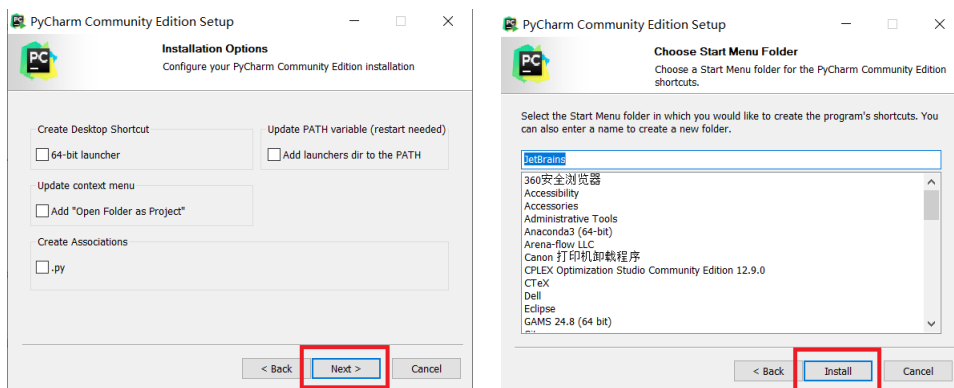
Free, open-source

一般选择免费版本下载，下载完成后双击 exe 进行安装。

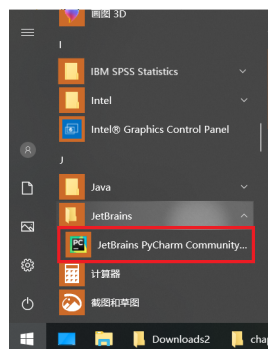




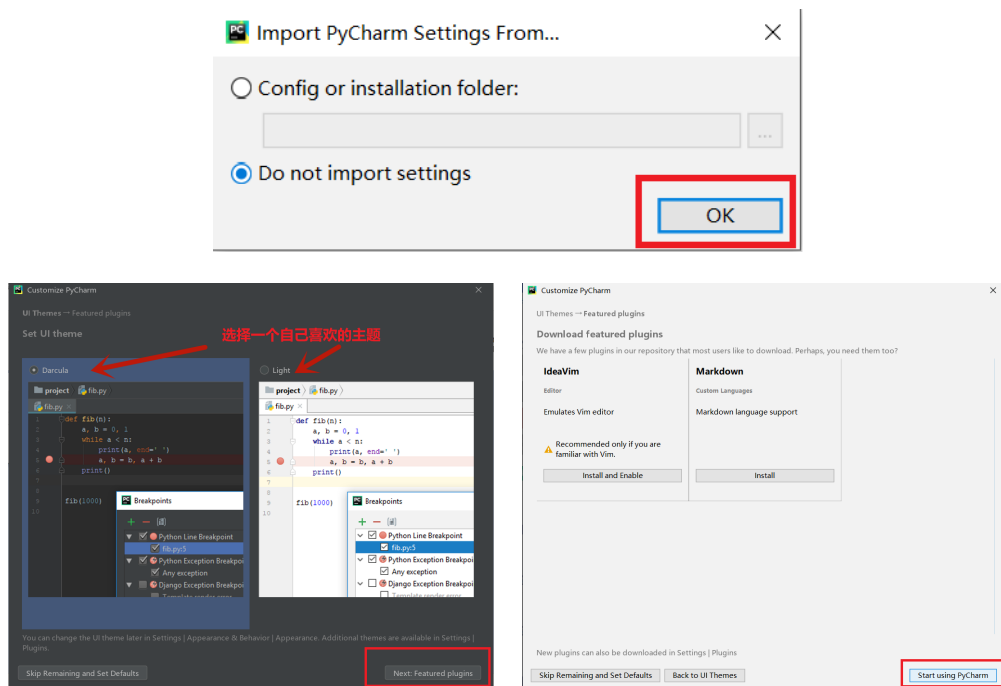
可以通过图中的 **Browse** 更改安装目录，其他步骤按照默认设置即完成安装。



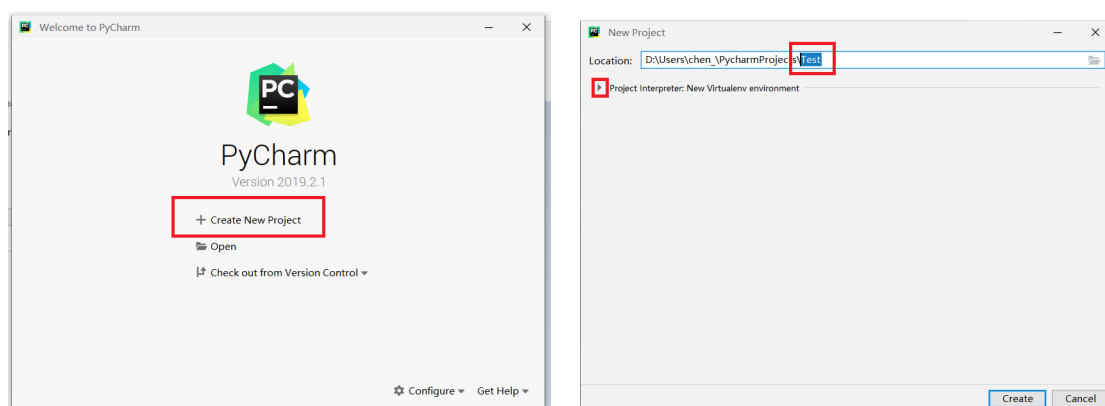
完成安装后，从 Windows 菜单栏里找到 jetbrains Pycharm，点击打开 Pycharm。



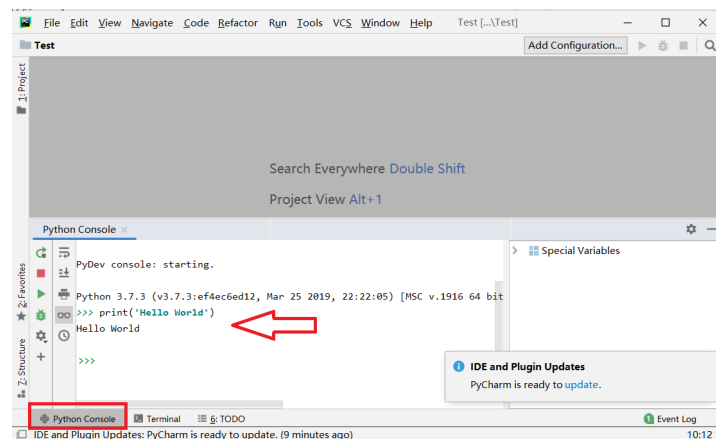
之后的设置如下面的图所示：



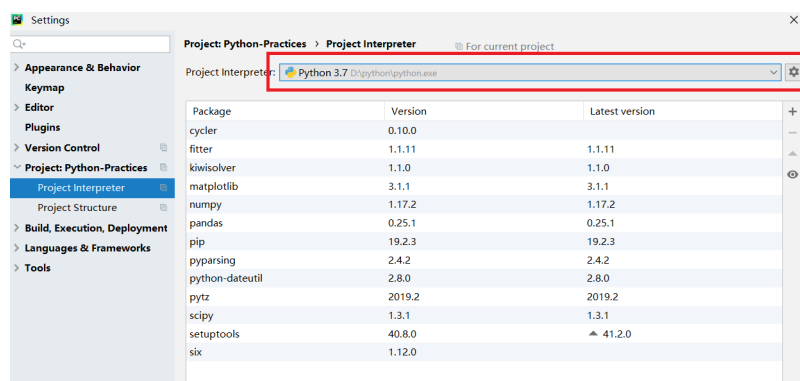
点击 Create New Project，创建一个新项目，项目的地址以及名字可以自己替换。



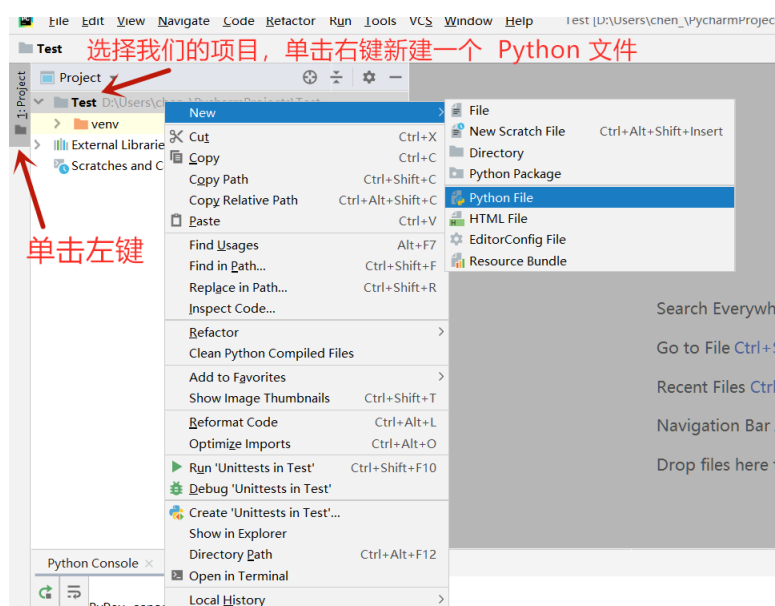
为了测试 Pycharm 是能安装成功，可以点击 Pycharm 界面最下面的 Python Console，打开控制台（Python Console 是一个交互式的 Python 编辑器，可以快速运行一些指令，一般调试时用），输入一条 Python 指令：`print('Hello World')`，若能正确运行，则表示 Pycharm 安装成功了。



有时候可能提醒我们设置 Pycharm 中的 Python 解释器。设置方式如下：依次点击 file—settings—project interpreter，将 interpreter 设置为我们 Python 安装路径中的 Python.exe（笔者电脑的位置为 D:\Python\Python.exe）。



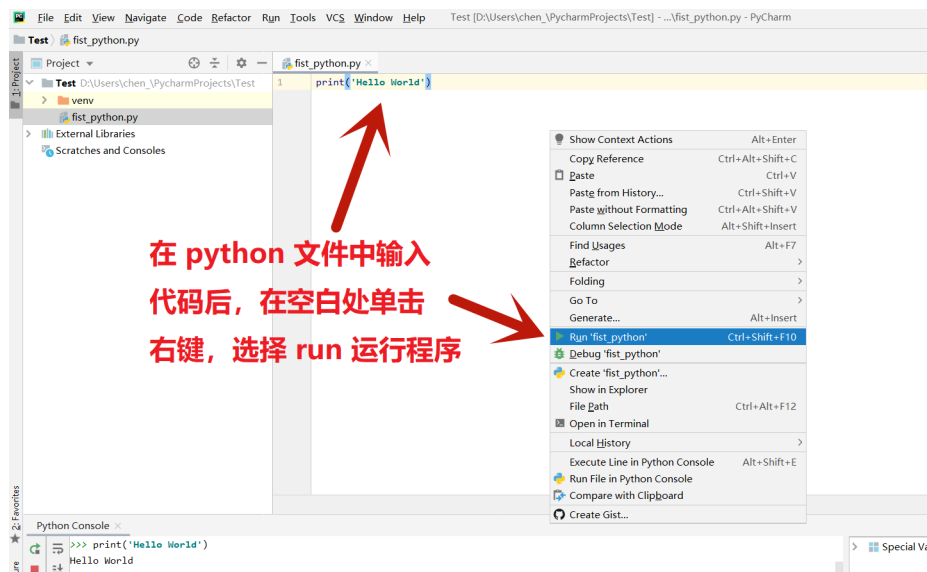
我们可以在刚刚新建的项目（project）中新建一个 Python 文件，操作步骤如下图所示：



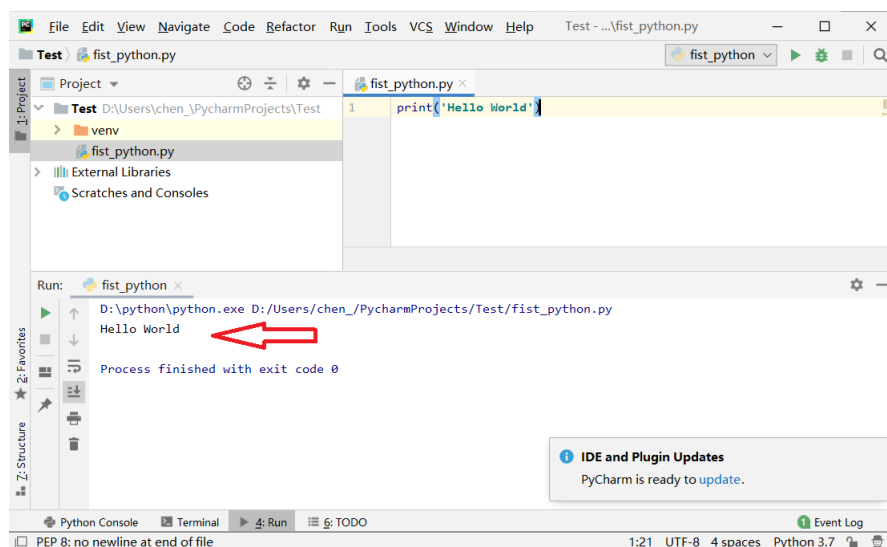
命名 Python 文件的名字：



输入代码后，在空白处单击右键，选择 run 运行程序：



在 Pycharm 下面的窗口中，就能看到运行输出的文字：



## 1.3 数据分析常用 Python 工具包的安装

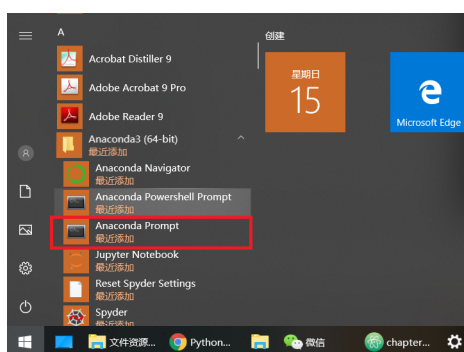
Python 的最大优势在于它有很多第三方开发的包，能够实现各种各样的功能。若我们使用 Anaconda，一般可以通过 conda 工具安装。若我们使用 Pycharm，一般用 pip 工具进行安装。

数据分析的常用包有：

- numpy — 用于数组和矩阵操作
- pandas — 用于处理表格数据，数据分析的重要工具
- matplotlib — 常用的基础绘图包
- seaborn — 另外一个绘图包
- scipy — 科学计算包，包括一些基础的统计学
- statsmodels — 统计建模与分析包
- sklearn — 机器学习包

### 1.3.1 用 Anaconda 安装

Anaconda 已经默认帮我们安装了很多常用的包，很多时候并不需要我们再安装新包。若实在有必要，可以通过 Anaconda 自带的 Conda 工具安装。方法是：在 Windows 菜单栏里，点击 Anaconda Prompt，打开 Anaconda 的命令行窗口，如下图所示：



常用的 conda 语法如下：

#### conda 的常用语法

安装包：	<code>conda install [package_name]</code>
搜索包：	<code>conda search [package_name]</code>
卸载包：	<code>conda uninstall [package_name]</code>
更新包：	<code>conda install [package_name] -U</code>
显示已安装包列表：	<code>conda list</code>



例如，安装 `numpy` 包的命令输入为：`conda install numpy`。不过，国内用 `conda` 安装包时的网速有时候比较慢。

### 1.3.2 用 `pip` 安装

`pip` 是 Python 包管理工具，该工具提供了对 Python 包的查找、下载、安装、卸载的功能。在我们安装 Python 时，`pip` 也是自动安装的，并且 `pip` 的地址已经添加到了计算机的环境变量里了。打开命令行窗口（win10 系统中敲 `win` 键，再输入 `cmd`，然后回车），输入 `pip`（有时候若电脑里其他软件也内置 `pip`，则需要输入 `pip.exe` 才能正确显示）回车，若能显示如下图所示的很多内容，则表示 `pip` 已经成功安装了。

```
C:\Users\chen_>pip
Usage:
  pip <command> [options]

Commands:
  install          Install packages.
  download         Download packages.
  uninstall        Uninstall packages.
  freeze           Output installed packages in require
  list             List installed packages.
  show             Show information about installed pac
  check            Verify installed packages have compa
  config           Manage local and global configuratio
```

常用的 `pip` 语法如下：

#### `pip` 的常用语法

安装包：	<b><code>pip install [package_name]</code></b>
搜索包：	<b><code>pip search [package_name]</code></b>
卸载包：	<b><code>pip uninstall [package_name]</code></b>
更新包：	<b><code>pip install [package_name] -U</code></b>
显示已安装包列表：	<b><code>pip list</code></b>

例如，安装 `numpy` 包时在命令行窗口输入 `pip install numpy`，操作命令与运行结果如下图所示。

```
C:\Users\chen_>pip install numpy
Collecting numpy
  Using cached https://files.pythonhosted.org/packages/bd/51/7dfla3858ff0
9/numpy-1.17.2-cp37-cp37m-win_amd64.whl
Installing collected packages: numpy
Successfully installed numpy-1.17.2

C:\Users\chen_>
```

因此，我们可以在命令行窗口，利用 `pip` 的安装命令将其他包安装到 Python 中。

```
pip install pandas
pip install xlrd
pip install openpyxl
```

```
pip install matplotlib  
pip instal scipy  
pip install statsmodels  
pip install seaborn
```