02-Data Visualization by Python

课件：Data Visualization by Python\_PPT\_LiuHong\_20220511.pdf

试验环境链接：https://pan.baidu.com/s/1C43EqMM9oV66GLT-ikkFKg

提取码：n24t

培训录制链接：https://cisco.webex.com/recordingservice/sites/cisco/recording/playback/d10e2e73b2f7103ab7df0050568150c6

密码：iHYrApn2

菜鸟网站关于python3相关的语法，建议有空时根据进度自学，边学边写Python程序练习

<https://www.runoob.com/python3/python3-tutorial.html>

请回答下面题目

1. Python当前非常流行，特别是在大数据分析方面，主要原因有二方面。一是因为Python的语法比其它程序语言简单，二是因为Python有大量方便易用的第三方工具库。那么在Python中导入第三方工具库有几种方式？并列举出它们的不同点。

|  |  |  |
| --- | --- | --- |
|  | **语法** | **不同点** |
| 方式一 | import 包.包…模块 as 别名  import matplotlib.pyplot as plt | 只能导到模块这一级  调用时需要加别名为前缀 |
| 方式二 | from包.包…模块 import 类或函数  from sklearn.metrics import r2\_score | 可以导到模块里的类或者函数  可以直接调用 |

1. Python运行时有时候会有一些警告信息，如何消除这些警告信息？

增加下面二行代码，就可以忽略掉警告信息的输出。

import warnings

warnings.filterwarnings("ignore")

1. 在Python中如何定义一个类，以及类的成员变量和方法？可以用一个例子来说明。

在Python中定义一个类时，要注意如下几点：

* 类定义用关键字 class，且顶格开始
* 类名符合“驼峰“命名规则，类名最后用冒号
* 类中的方法必须缩进一个tab距离，且第一个参数必须为self
* 类中的成员变量必须用self.为前缀
* 类使用时必须申请一个对象，相当于实例化一个类
* 类的成员变量和方法的使用，必须用对象作为前导，（其实这个对象就是类中的self）

1. 在问题3中我们提到类的方法，本质上也是一个函数，那么普通的函数和类的方法有什么相同点和不同点，请对比说明。

|  |  |  |
| --- | --- | --- |
|  | **函数** | **类方法** |
| 相同点 | 本质上都是函数，必须符合函数定义   * 用def 进行函数的定义 * 函数名最好统一用小写 * 函数参数位于函数名后，用小括号()括起来 * 最后以冒号: 结尾 * 函数体比函数名缩进一个tab距离 | |
| 不同点 | 函数顶格  参数全部是形参，全由外部传入  调用时用函数名+参数  函数只能引用传入的参数 | 类方法在类里面，比类缩进一个tab  第一参数必须是self，表示调用的类对象  调用时必须用类对象作为前导，参数不带self  方法可以使用类变量 |

1. 在Python的可视化培训中，老师提到过三个可视化工具库，它们分别是什么？请列出它们之间的联系和区别。

|  |  |
| --- | --- |
| 可视化工具库 |  |
| matplotlib | 官网：https://matplotlib.org/  教程：https://www.runoob.com/matplotlib/matplotlib-tutorial.html  Python中最基础的可视化库  非常灵活，参数也非常多，比较底层，可以基于它开发出新的可视化库 |
| seaborn | 官网：https://seaborn.pydata.org/  说明：http://www.techweb.com.cn/cloud/2021-08-05/2852336.shtml  在matplotlib的基础上进行了更高级的API封装  matplotlib相当于自由点菜，seaborn相当于套餐 |
| pyecharts | 官网：https://pyecharts.org/#/  pyecharts更高级，更动态（生成是一个页面，里面嵌入JavsScript）  可以保存为静态的图片  是基于类，而不是函数  上述二个可视化工具库是Python内部集成的，可以直接调用，而pyecharts需要单独安装（pip install pyecharts） |

1. 在可视化时，我们需要根据具体的情况展示不同类型的图形，请列举出4-5种可视化图形，以及matplotlib中对应的函数名。

我们可以在matplotlib官网 – Users guide - Plot types找到所有的图形。

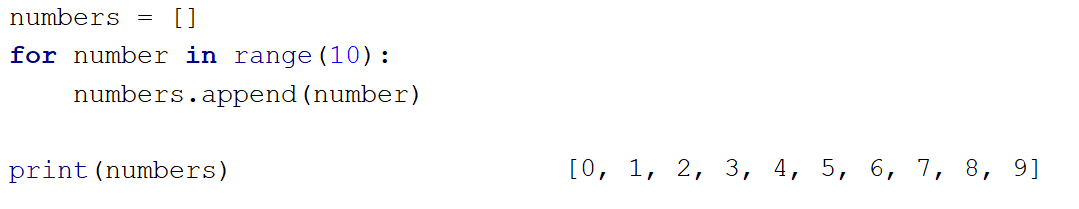
https://matplotlib.org/stable/plot\_types/index.html

常见的图形有以下几种：

|  |  |  |  |
| --- | --- | --- | --- |
| **图形** | **名称** | **示例** | **函数名或模块** |
| 折线图 | plot |  | plot |
| 柱状图 | bar |  | bar |
| 饼图 | pie |  | pie |
| 散点图 | scatter |  | scatter |
| 词云图 | wordcloud |  | jieba  wordcloud  imshow |

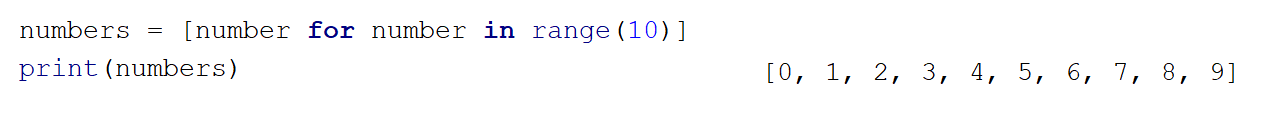
1. 什么是列表解析式？如何理解列表解析式？

在讲列表解析式之前，我们来看一个常规的循环的例子：



在此示例中，实例化了一个空列表 numbers。然后使用 for 循环迭代 range(10) 并使用 append() 方法将每个数字附加到列表的末尾

上面的逻辑可以简化为下面的形式，这就是列表解析式：



1. 在Python类中如何定义public、protected、private的方法？

Python是一种极简的语言，它摈弃了许多传统语言的繁琐语法，这种思想同样体现在对访问权限的控制上，它是通过名称前面的下划线\_来表示的

|  |  |  |
| --- | --- | --- |
| **访问权限** | **规则** | **示例** |
| public | 字母开头 | add |
| protected | 以一个下划线\_开头 | \_add |
| private | 以二个下划线\_\_开头 | \_\_add |
| system | 以二个下划线\_\_开头，同时以二个下划线\_\_结尾 | \_\_init\_\_ |

1. 在用Python进行数据分析时，经常遇到对数据进行切片处理，所谓切片就是从大量数据中提取部分行或列数据，一般会用到loc，iloc等，请用例子演示如何切片？

数据切片是为了对数据进行提取，一般是用二个函数loc，iloc：

Graphical user interface, text, application, email

Description automatically generated

|  |  |
| --- | --- |
| **函数或操作** | **数据切片** |
| iloc | df.iloc [row index, column index] |
| loc | df.loc [row name, column name] |
| 下标 | df[column name] |

对数据来说，一般是二维的，行为样本，多少行就表示多少个样本数，列为特征，多少列就表示多少个特征值。样本数只能说明数据的多和少，特征值才是真正的数据分析对象，所以取列的情况要远远大于取行的情况，故Python对列的操作进行了简化，直接采用下标来访问。

1. 如何用布尔值对数据进行过滤和选择？

我们已经知道DataFrame，Series和ndarray之间的关系，同时也知道ndarray相当于一个二维的矩阵，可以进行运算，包括比较运算和逻辑运行，这样就会生成布尔值，也就是True和False。

然后我们可以利用布尔值为True来对数据进行过滤和选择：对应的值为True的就被选中，对应的值为False就被过滤掉。

1. 对数据进行可视化的意义是什么

回忆一下，当我们编程的时候，特别是对大型程序，一定会进行单步调试，或者将中间的值输出，以检查结果是否正确。

同样道理，当我们对大量数据进行处理时，也很关心中间的处理是否正确，我们不可能将每一个结果都输出，这对大数据处理是不现实的，所以我们可能会用可视化进行检查。

故数据可视化的意义不外乎下面二种：

1，把最后的结果直观地呈现给用户，比如向领导汇报

2，进行数据分析时，通过可视化排除不相关的特征，比如鸢尾花示例

1. 可视化编程题1：做一个词云图

要求：

1，以text.txt为数据来源

2，做出一个词云图（效果参考培训教材）

3，完成代码，并上传到git上

1. 可视化编程题2：做一个销量趋势图

要求：

1，以byd\_sales.xlsx为数据来源

2，做一个2021-2022年销量数据同比（参考byd\_sales.png）

3，完成代码，并上传到git上