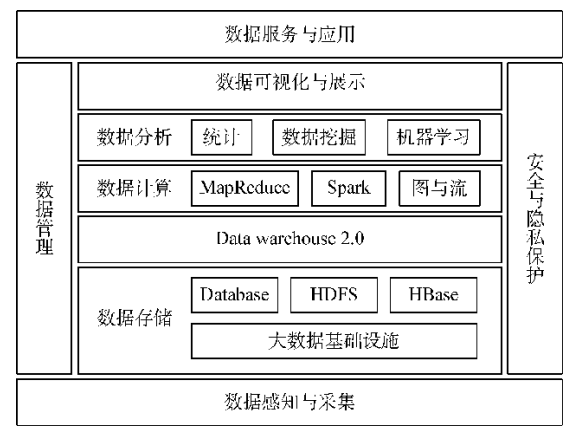


1 数据生命周期周期管理如图：



其包括了数据采集，数据存储，数据管理，数据计算，数据分析，数据展示。

2数据采集是指从真实世界对象中获得原始数据的过程。不准确的数据采集将影响  
后续的数据处理并最终得到无效的结果。目前，最常用的三种数据采集方法是：传感器，日志文件和 爬虫。

3 数据管理是利用计算机硬件和软件技术对数据进行有效的收集#存储#处理和应用  
的过程。传统的文件系统有：人工管理，文件系统，数据库系统。人工系统的数据不共享，不具有独立性，不保存；文件系统共享性差，冗余度高；数据库是统一管理数据的专门软件系统；而大数据管理，本质上是在大数据基础设施层之上向用户提供更加方便，高效，友好的人数交互界面。他帮助使用者忽视底层细节，专注数据本身。

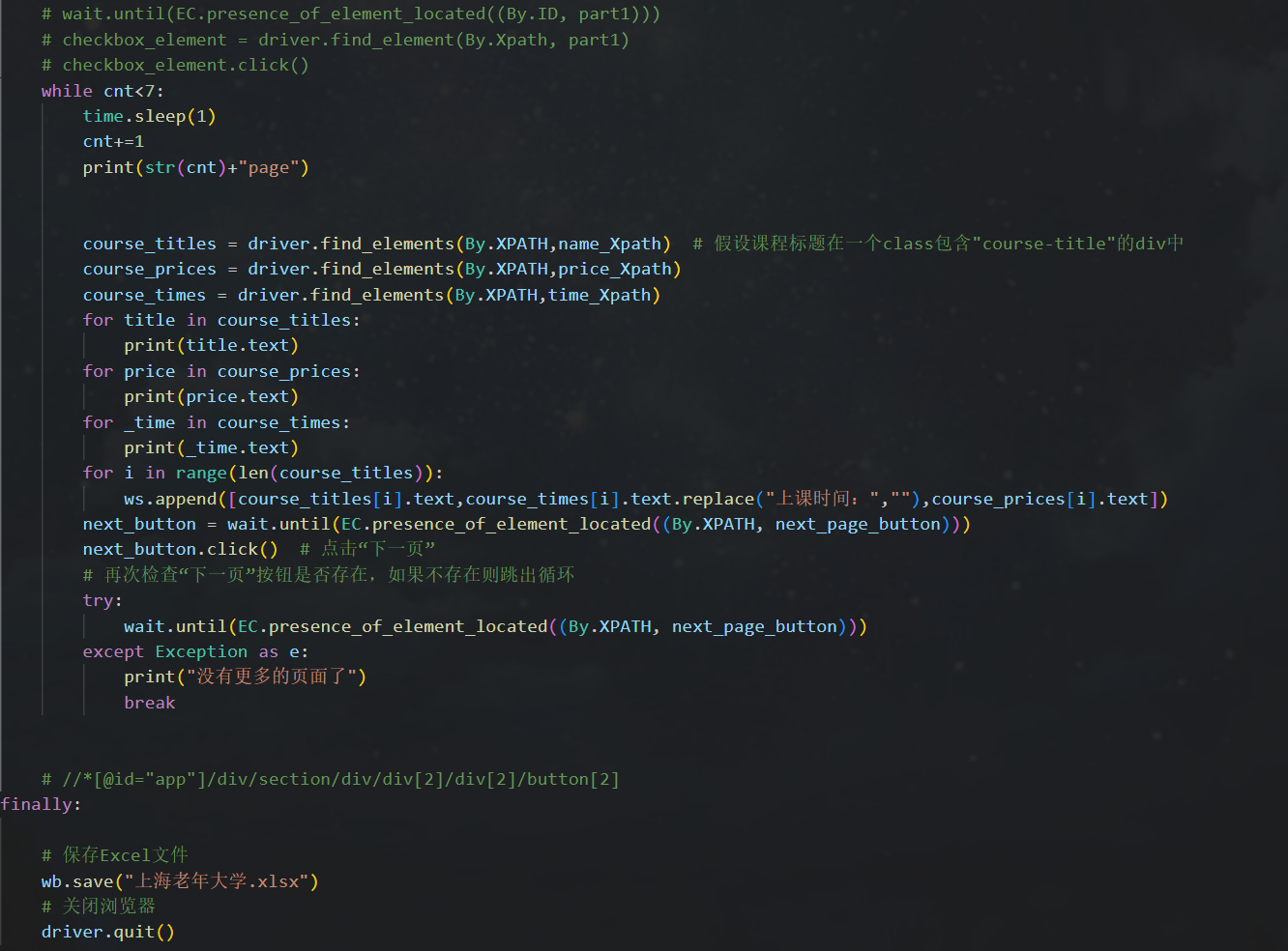
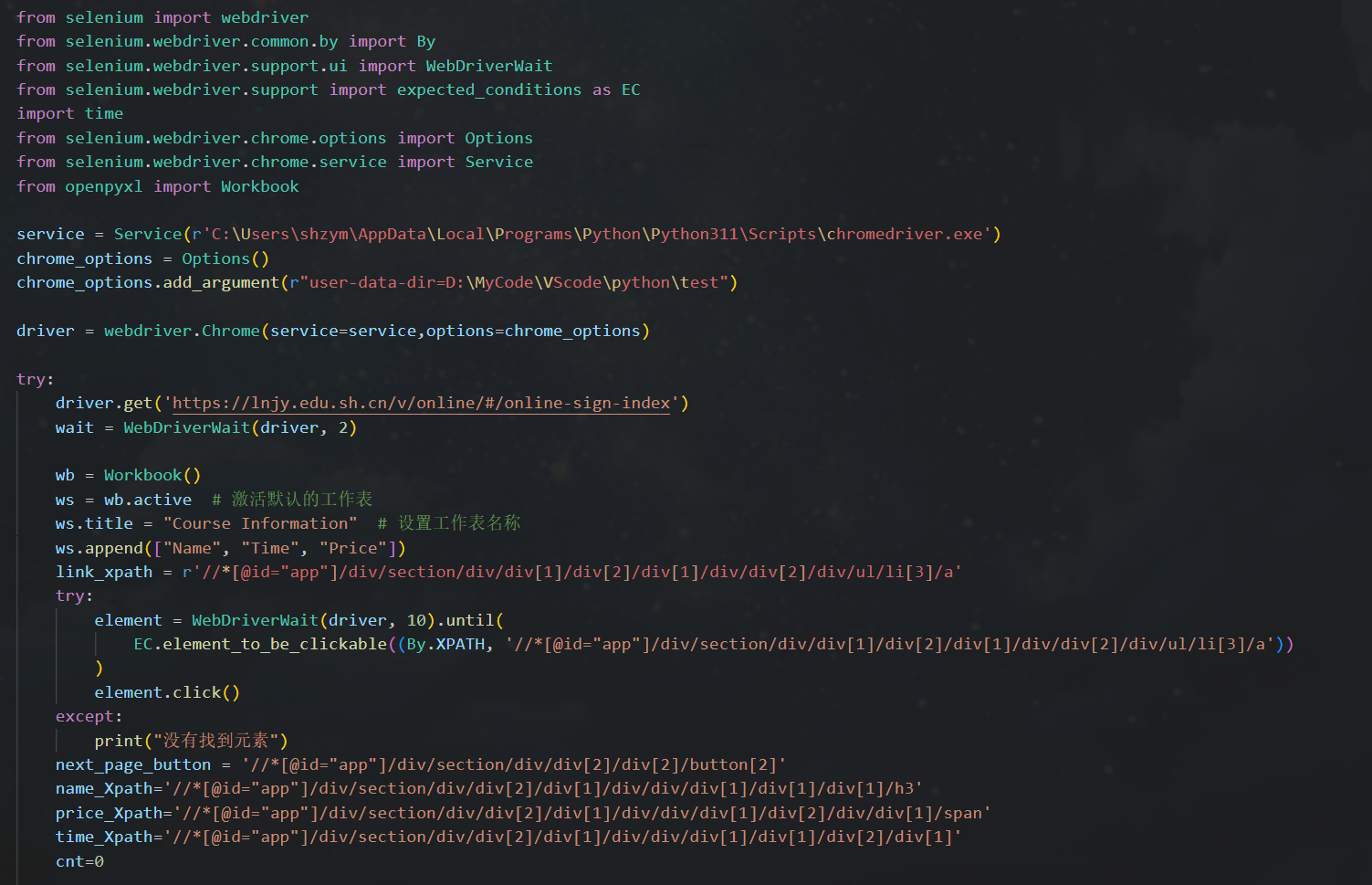
4 数据的计算模式大致分为批量计算模式，流式计算模式，交互式计算模式和图计算  
模式四类

5数据分析处理来自对某一兴趣现象的观察，测量或者实验的信息。数据分析目的  
是从和主题相关的数据中提取尽可能多的信息。

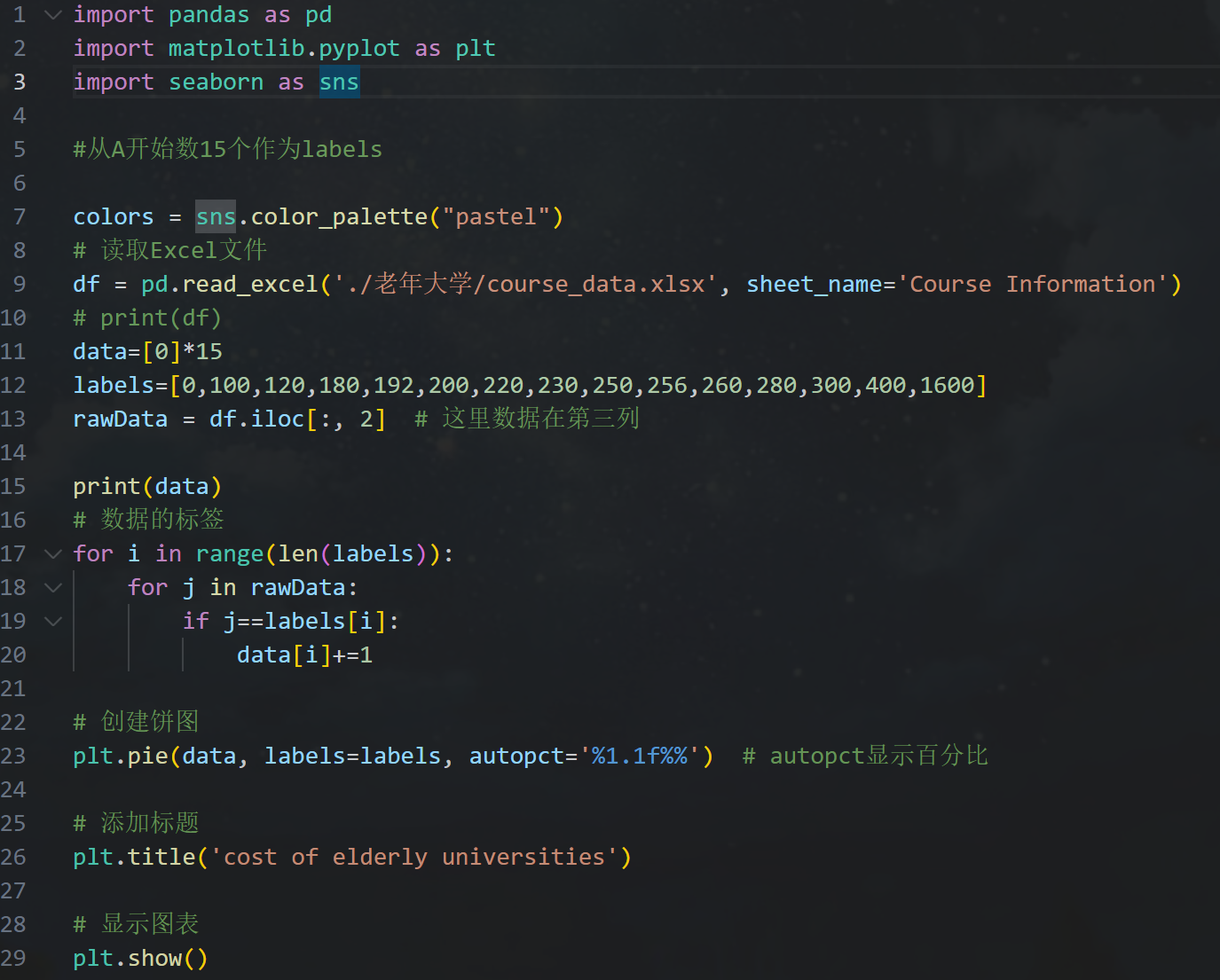
可以将根据观察和测量得到的定性或定量数据，或根据参数数量得到的一元或多元数据进行分类。根据数据分析深度将数据分析分为三个层次：描述性分析，预测性分析和规则性分析。有回归技术，回归统计，仿真等等。

6数据可视化方便我们更容易地通过视觉短时间获取大量信息，能够在小空间中展示大规模数据，能够帮助我们对数据有更加全面的认识，帮助我们更好记忆。

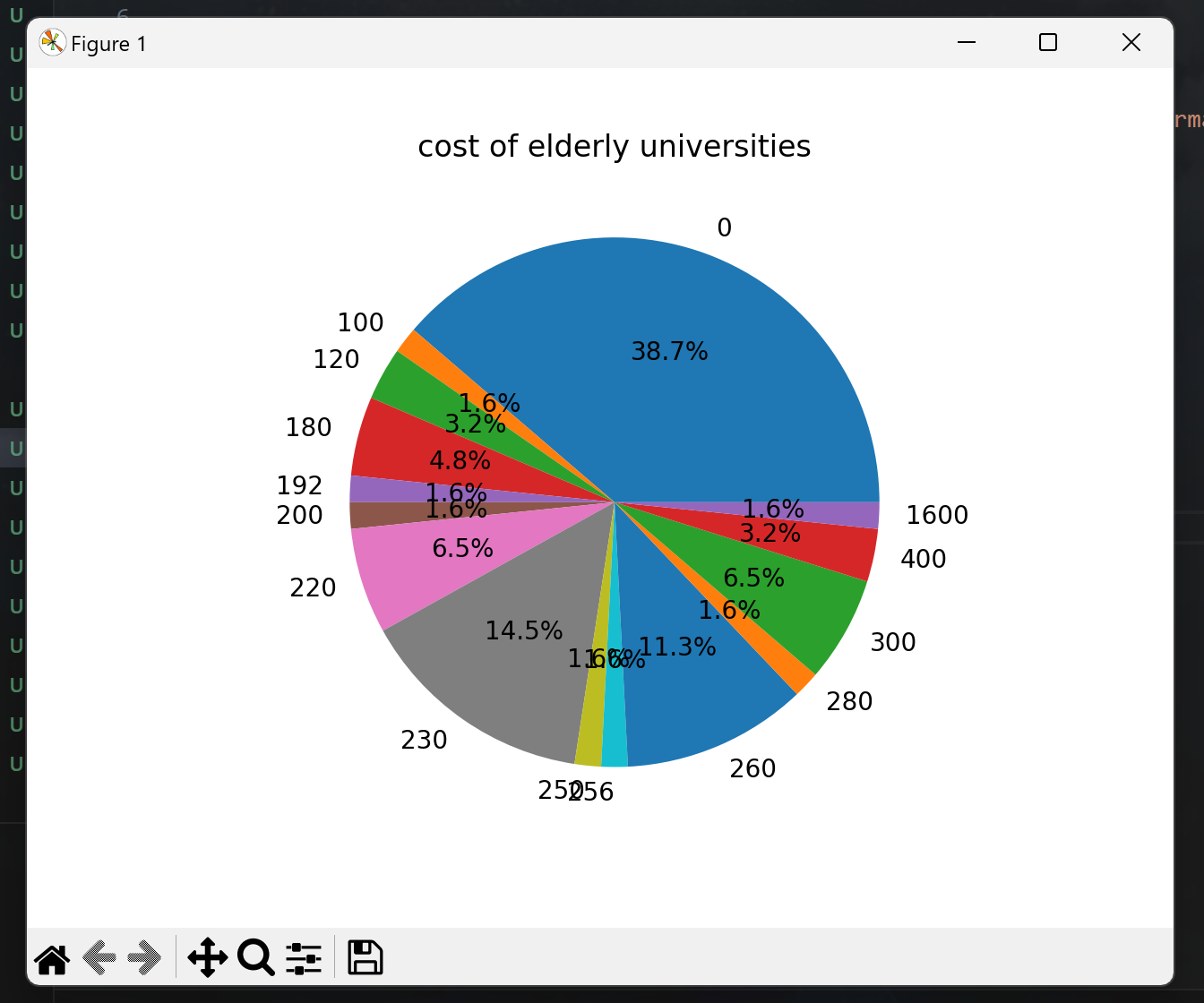
7爬虫：



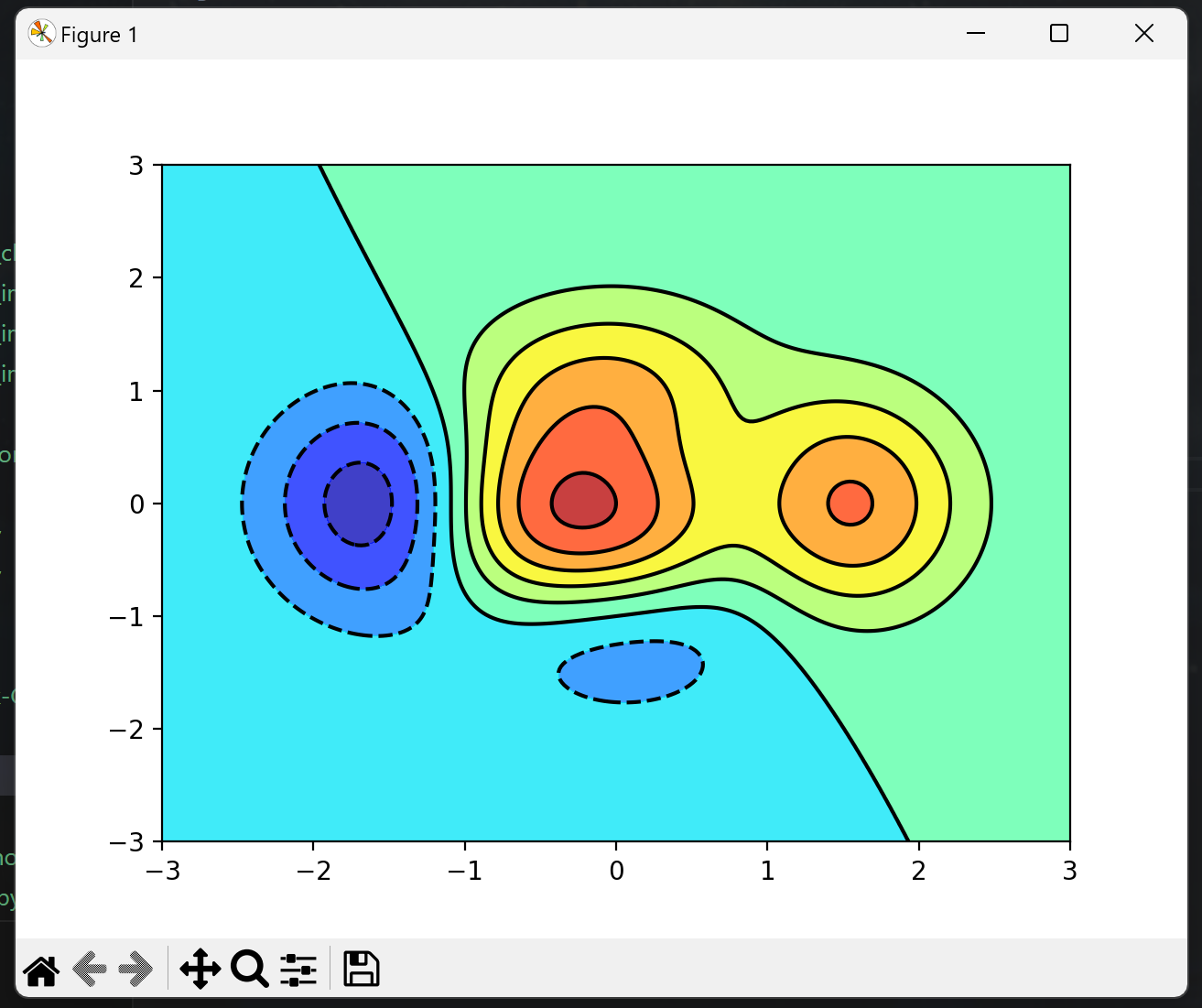
使用饼图的代码：

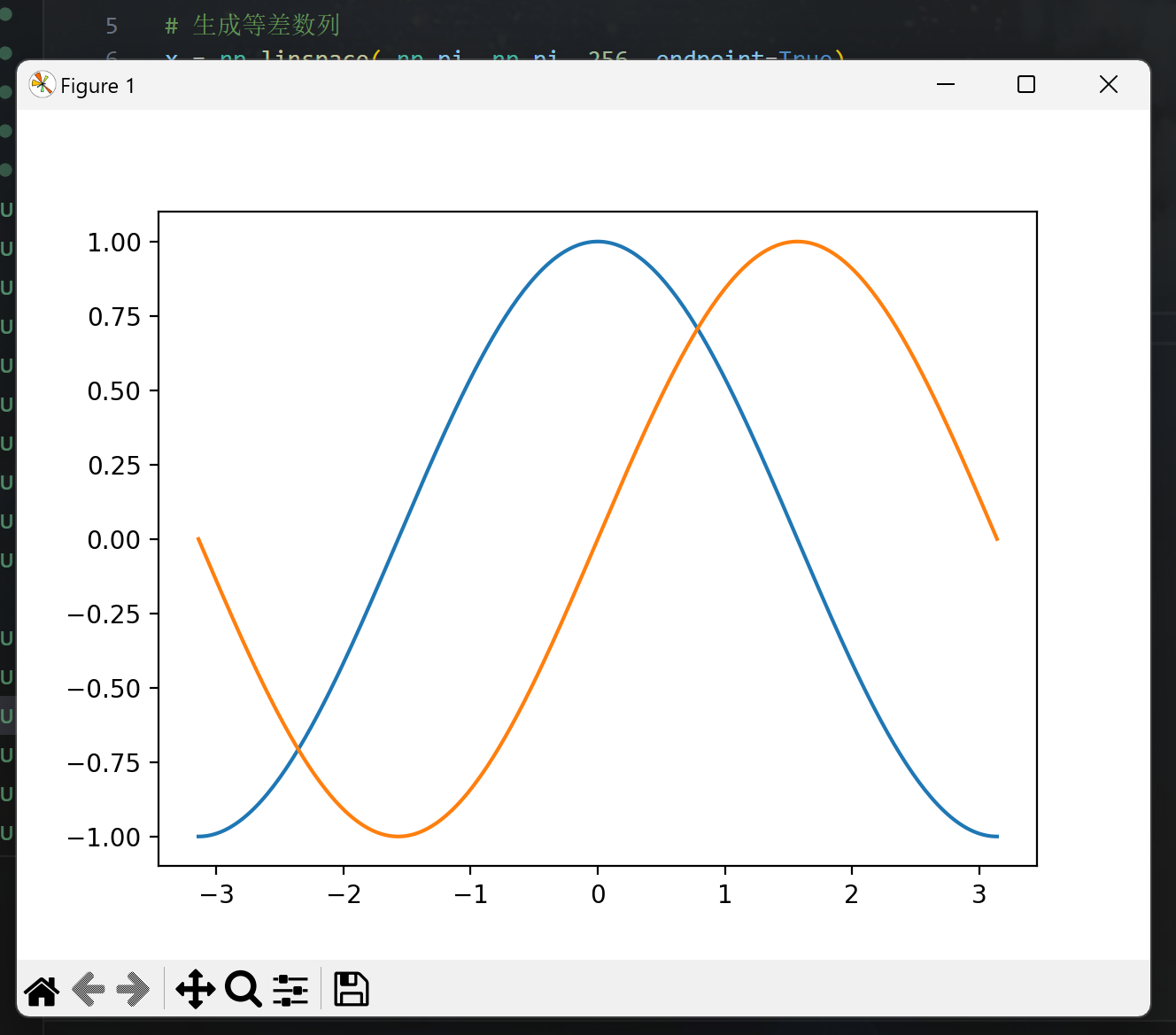


运行结果：



同理，其余两种：





使用seaborn的效果如下：

