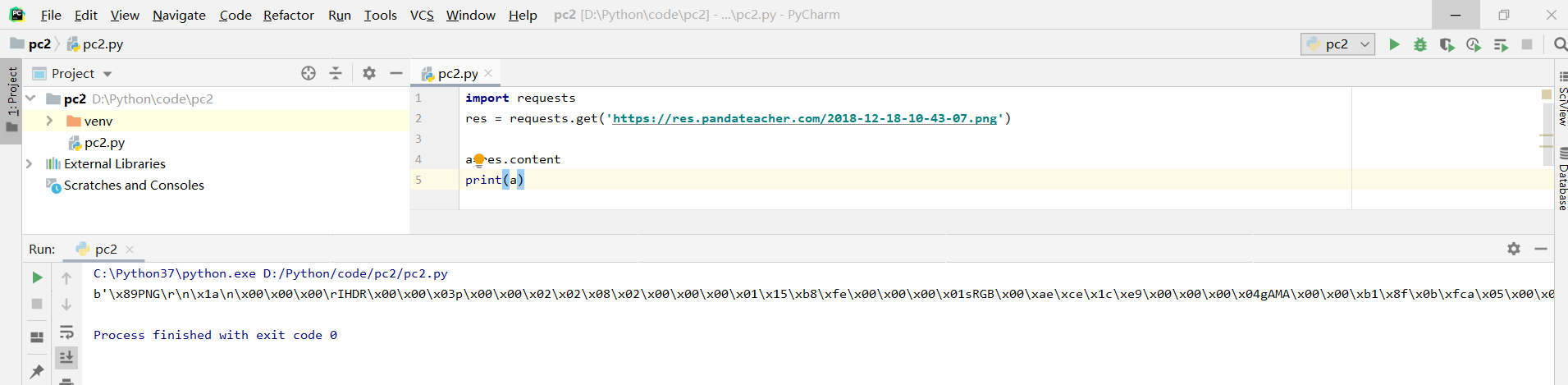
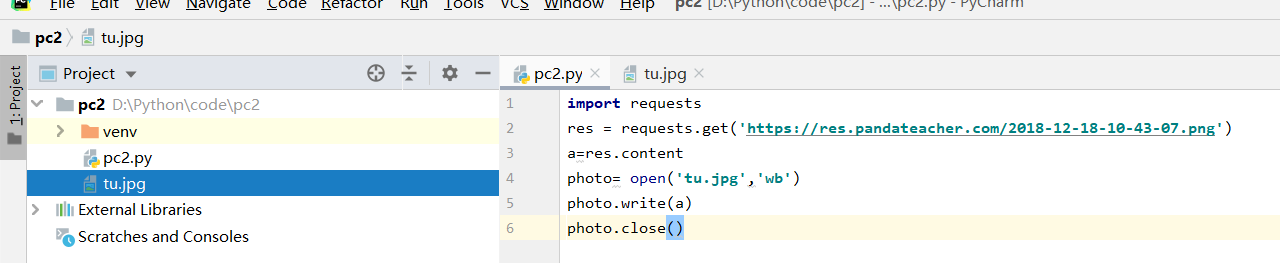
**第0关**

# 问题1：

下载图片的时候 为什么我这么写 最后打印出来的是二进制

为什么自己写出来的结果是这样的？ 应该是属于自己就显示二进制的那种





这么写就写入一个jpg的文件 就能正常显示

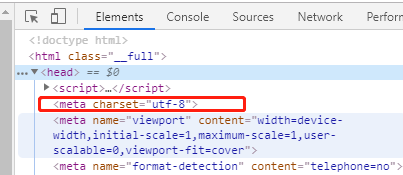
A：图片和音乐都属于字节流，读写模式都需要用rb+或者是wb+

# 问题2：

课程内容里 有这么一部分

遇到乱码时 用 res.encoding定义编码时，课上说要这么去实现： requests库会对数据的编码类型判断不准确，就会出现一堆乱码，那我们就可以去查看目标数据的编码，然后再用res.encoding把编码定义成和目标数据一致的类型即可。 那么这里目标数据的编码怎么去查询呢？

A：打开检查，在html源代码中，如下图



**第一关**

# 问题1：

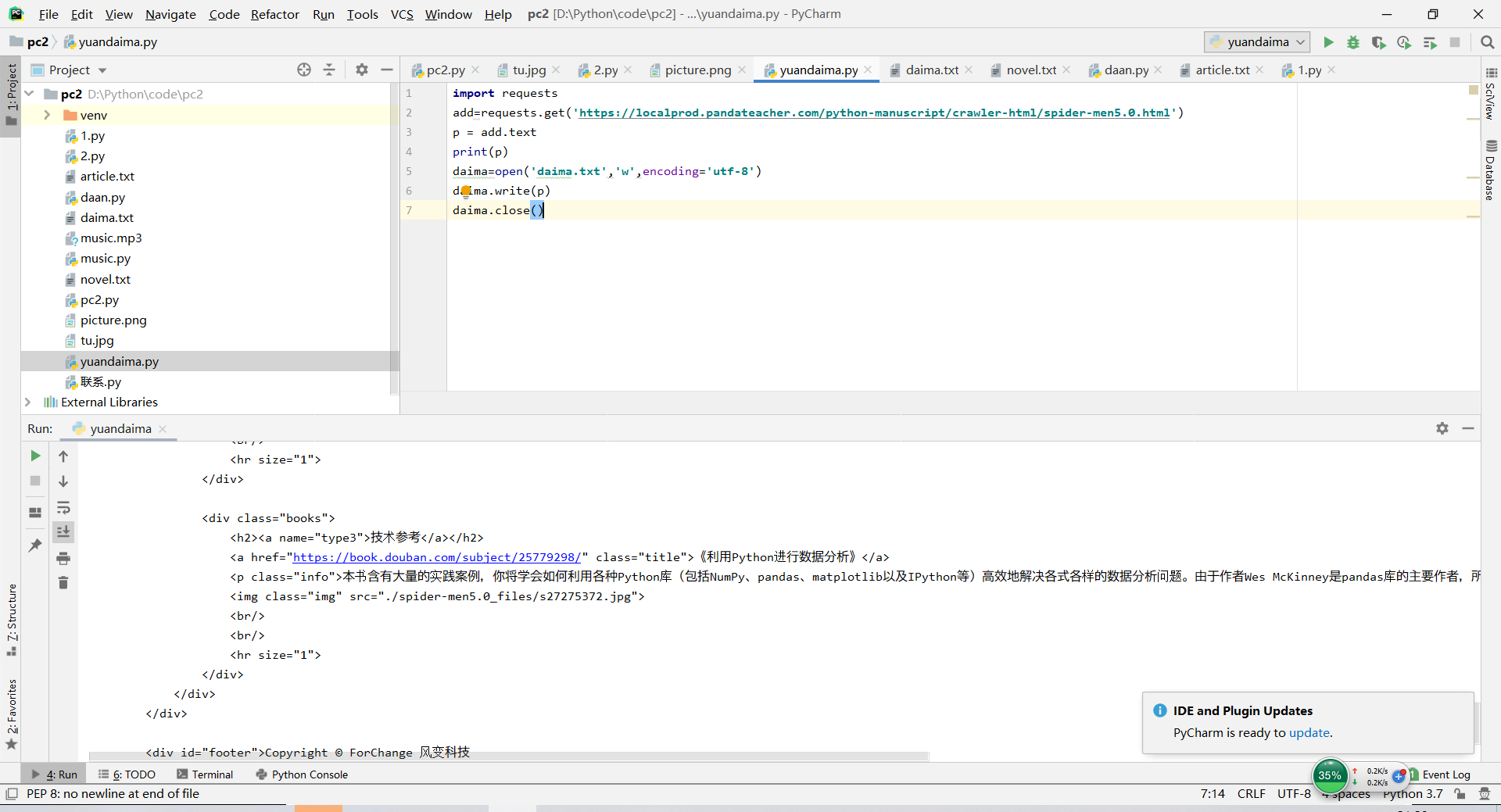
最后的练习，自己在课程的环境里可以修改，但是怎么样在真正的网页上去修改添加一本书的描述呢？添加一个div啥的 选了开发者模式 感觉双击某一个要修改的区域 只能改名字啥的 添加不了东西呀？

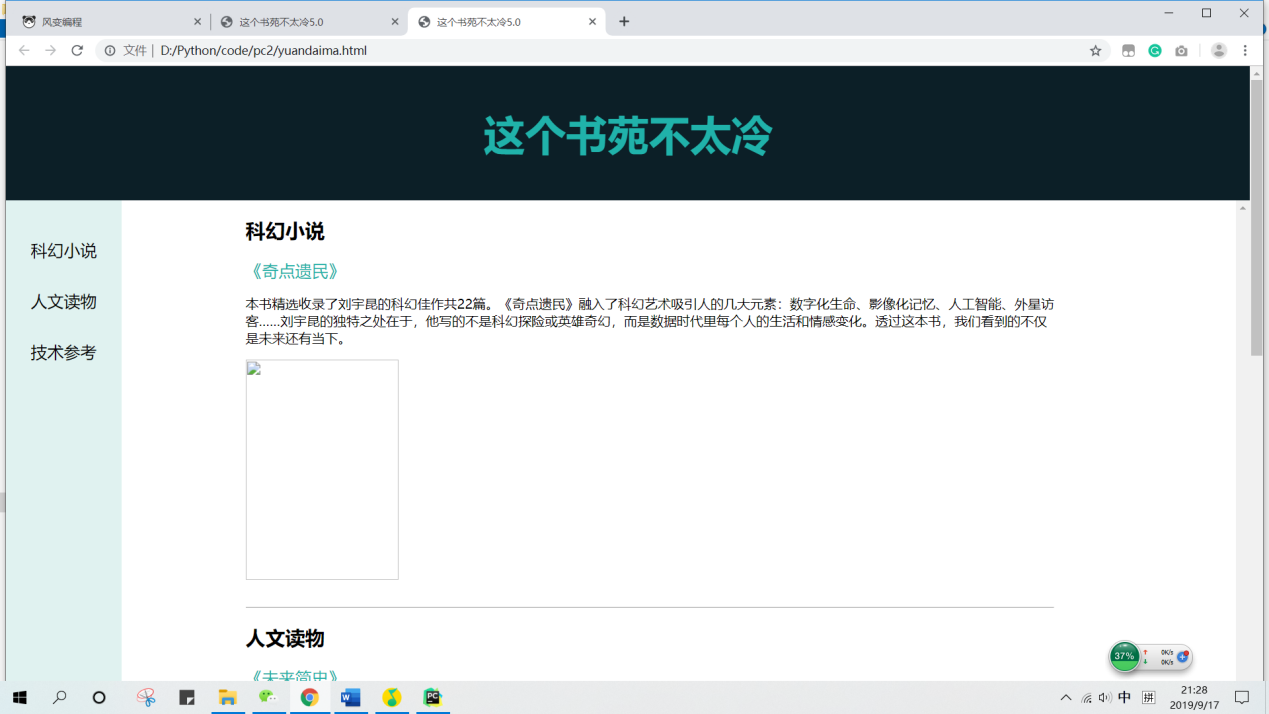
A：具体问题具体分析，课程上的网页是为了教学而准备的，所以可以进行修改，但一般其它的正常网站，是不能修改的，毕竟服务是别人的，修改之后也是自行能看而已，修改的方式可以将html代码都爬取到一个txt文档，通过添加里面的元素块就可以了

# 问题2：

把刚才请求到的HTML源文件复制粘贴，在vscode中保存为后缀为.html 的文档，它就是一个保存在你本地中的网页了。

但是我这里发现写入本地的html 只有文本 怎么样可以同时储存网页的文本和图片？





A：这里是无法一起爬取的，因为爬虫的本身就是对某一数据的单独提取，有些时候爬取所有本文，在运行html代码时，发现图片是无法显示的，这也是因为图片是属于服务器上的图片，并不是自己本地的电脑图片，所以导致无法显示

# 问题3：

在这里，5.0对比4.0版本变化的部分是：（第9、19、29行）给h2元素标注了**name属性**，< section id='nav'>中的超链接标签(第3、4、5行)以这个name属性为标识，设置了跳转到这个标题的**锚点**；（第10、20、30行）给书名添加了超链接，可以链接到这本书的豆瓣主页；以及，（第12、22、32行）用<img>标签添加了书的封面图片。

<https://localprod.pandateacher.com/python-manuscript/crawler-html/spider-men5.0.html>

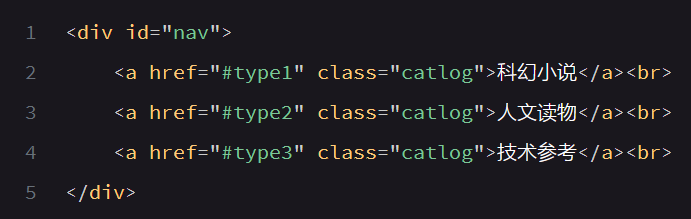
这里的< section id='nav'>中的超链接标签是什么？我并没有找到这个东西



A：< section id='nav'>只是一个属性和属性值，并没有跳转的用意，这相当于一个标题的定位

# 问题4：

想知道这里的<br>是什么？？



A：<br>是换行的意思