

➤ 代码审核中的疑问

▪ 对超链接的使用和可视化

超链接地址在图书版和电子版本的处理方法是否需要加以区分

不太确定，出版时，是否编辑人员会修改超链接，为超链接提供脚注

}
如需下载数据，请求参数是需要申请访问密码'ak'（申请地址，注意，需要注册登录）。首先
Values，逗号（字符）分隔值，以纯文本形式存储表格数据）和".json"（JavaScript Object No
式的文件。因为要通过网页地址来检索数据，因此调入urllib的HTTP库实现相应的URL（Uniform

▪ 对正文里，提及 python 库时，是否需要采用代码格式

}
如需下载数据，请求参数是需要申请访问密码'ak'（申请地址，注意，需要注册登录）。首先来配置基本的参数，最终要实现的目的将下载后的数据分别存储为".csv"（Comma-Separated Values，逗号（字符）分隔值，以纯文本形式存储表格数据）和".json"（JavaScript Object Notation，轻量级数据交换格式，层次结构简洁清晰）数据格式，因此调入json和csv库，辅助读写对应格式的文件。因为要通过网页地址来检索数据，因此调入urllib的HTTP库实现相应的URL（Uniform Resource Locator，统一资源定位符，即网络地址）处理。对文件路径的管理则可以使用os以及pathlib库。
• 矩形区域检索（百度地图开放平台规定：目前矩形区域检索为高级权限，如有需求需提交【工单】咨询。）
| 如果无法通过矩形区域检索，可以从本书仓库中下载已经检索的POI数据，继续后续的实验

▪ 部分已经注释的代码，是否需要保留

```
poi_json=open(poi_save_path,'w',encoding='utf-8')  
  
# url=urlRoot+urllib.parse.urlencode(query_dic)  
# print(url)  
jsonDS=[] # 存储读取的数据，用于.json格式数据的保存  
for p in tqdm(page_num_range):  
    # 更新请求参数  
    query_dic.update({'page_num':str(p)})  
    url=urlRoot+urllib.parse.urlencode(query_dic)  
    data=urllib.request.urlopen(url)  
    responseOrLoad=json.loads(data.read())  
    # print(url,responseOrLoad.get("message"))  
    if responseOrLoad.get("message")== "ok":  
        results=responseOrLoad.get("results")  
        for row in range(len(results)):  
            subData=results[row]  
            baidu_coordinateSystem=[subData.get('location').get('lng'),subData.get('location').get('lat')]  
            Mars_coordinateSystem=cc.bd09togcj02(baidu_coordinateSystem[0], baidu_coordinateSystem[1])  
            WGS84_coordinateSystem=cc.gcj02towgs84(Mars_coordinateSystem[0],Mars_coordinateSystem[1])  
  
            # 更新坐标  
            subData['location']['lat']=WGS84_coordinateSystem[1]  
            subData['detail_info']['lat']=WGS84_coordinateSystem[1]  
            subData['location']['lng']=WGS84_coordinateSystem[0]  
            subData['detail_info']['lng']=WGS84_coordinateSystem[0]  
            jsonDS.append(subData)  
  
    # break  
if poi_json:  
    |  
    json.dump(jsonDS,poi_json)  
    poi_json.write('\n')  
    poi_json.close()  
print("The download is complete.")
```

- 注释的格式

一般现在普遍采用 pep8 编码风格,

如果单行注释, 注释从一行的最前面开始, 井号后面空一格

如果是注释位于一行代码后面, 井号前面空两格, 井号后空一格

Python PEP-8 编码风格指南中文版

<https://alvin.red/2017/10/07/python-pep-8/>

Python 风格规范 — Google 开源项目风格指南

https://zh-google-styleguide.readthedocs.io/en/latest/google-python-styleguide/python_style_rules/#section-2

python 注释规范 - 简书

<https://www.jianshu.com/p/4facd9ff2fcd>

- 路径取名问题,

最好是不要带空格的文件夹和文件名, 有时候会出问题

```
xi an_regi on_gdf=gpd.read_file('./data/xi an _regi on/xi an _regi on. shp' )
```

- 长代码问题

有的代码, 太长, 会影响可读性, 建议分行

比如改成这样

```
for p in page_num_range:
    # 更新请求参数
    query_dic.update({'page_num':str(p),
                      'bounds':str(leftBottomCoordi[1]) + ',' + str(leftBottomCoordi[0]) + ',' + str(rightTopCoordi[1]) + ',' + str(rightTopCoordi[0]),
                      'output':'json',
                      })
```

```
rightTopCoordi=[bound_coordinate[ leftBottom ][0]+(1+1)*xDis,bound_coordinate[ leftBottom ][1]+(1+1)*yDis]
for p in page_num_range:
    # 更新请求参数
    query_dic.update({'page_num':str(p),
                      'bounds':str(leftBottomCoordi[1]) + ',' + str(leftBottomCoordi[0]) + ',' +
                      str(rightTopCoordi[1]) + ',' + str(rightTopCoordi[0]),
                      'output':'json',
                      })
```

▪ Json 数据格式化显示

此处，其实是文本格式的 json 数据，可以参考以下链接，对数据进行结构化查看

2) 将 csv 格式的 POI 数据转换为 pandas 的 DataFrame

读取已经保存的 poi_csv.csv 文件。因为在文件保存时，使用的是 csv 和 json 库，因此读取时仍旧使用对应的库。最为经常使用的 pandas 库中也有 read_csv() 和 read_json() 等方法，在 csv 文件具体存储时，保存的方式也是多样的，因此 pandas 读取 csv 或者 json 文件最好是其自身所存储的文件，数据格式能够对应。而如果直接读取上述保存的 POI 的 csv 文件，则会出现错误。只有知道数据格式，才能够有目的的提取数据，读取每一行的数据格式如下：

```
[{"name": "昆明池遗址", "location": {"lat": 34.210991, "lng": 108.779778}, "address": "西安市长安区昆明池七夕公园内", "province": "陕西省", "city": "西安市", "area": "长安区", "detail": 1, "uid": "c7332cd7fbcc0d82ebe582d9", "detail_info": {"tag": "旅游景点;景点", "navi_location": {"lng": 108.7812626866, "lat": 34.217484892966}, "type": "scope", "detail_url": "http://api.map.baidu.com/place/detail?uid=c7332cd7fbcc0d82ebe582d9&output=html&source=placeapi_v2", "overall_rating": "4.3", "comment_num": "77", "children": []}]
```

[在线 JSON 校验格式化工具 \(Be JSON\)](#)

```
1 {
2   'name': '昆明池遗址',
3   'location': {
4     'lat': 34.210991,
5     'lng': 108.779778
6   },
7   'address': '西安市长安区昆明池七夕公园内',
8   'province': '陕西省',
9   'city': '西安市',
10  'area': '长安区',
11  'detail': 1,
12  'uid': 'c7332cd7fbcc0d82ebe582d9',
13  'detail_info': {
14    'tag': '旅游景点;景点',
15    'navi_location': {
16      'lng': 108.7812626866,
17      'lat': 34.217484892966
18    },
19    'type': 'scope',
20    'detail_url': 'http://api.map.baidu.com/place/detail?uid=c7332cd7fbcc0d82ebe582d9&output=html&source=placeapi_v2',
21    'overall_rating': '4.3',
22    'comment_num': '77',
23    'children': []
24  }
25 }
```

▪ 参考文献的设定

位置： 每章末尾还是在第一次引用出，或者当前小节末尾，有的书，好像被提到多次
以下两个截图是以前的格式，这两个截图中，参考文献的引用格式不同

SQLite数据库应用途径，引用[1]中的数据，结合代码实现阐释。同时使用DB Browser for SQLite(DB4S)辅助查看、管理SQLite数据库。

参考文献

1. (日)高桥麻奈著,崔建锋译,株式会社TREND-PRO漫画制作.漫画数据库[M].科学出版社.北京.2010.5.
2. Miguel Grinberg.Flask Web Development: Developing Web Applications with Python[M]. O'Reilly Media; 2nd edition, April 3, 2018. (中文版: Miguel Grinberg.安道译.Flask Web开发: 基于Python的Web应用开发实战[M].人民邮电出版社.2018.)

总分布假设所计算的概率结果将不再有效。

对于统计学的知识，通常是依据已出版的著作或者教材为依据进行解说，并根据阐述的内容，以python语言为工具分析数据的变化，从而可以直接应用代码来解决对应的问题，并能够通过具体的数据分析更深入的理解统计学相关的知识。本节参考：Timothy C.Urdan.Statistics in Plain English(白话统计学)[M].中国人民大学出版社.2013.12.第3版。

现在我把正文里的参考文献都采用了统一的格式，放在了提及的小节最后，

3) 描述性统计

本小节内容主要参考《漫画数据库》[1]、《漫画统计学》[2]及维基百科(Wikipedia)。以枯燥的知识以漫画的方式讲出来，并结合实际的案例由简入繁使枯燥的学习变得有趣起来，欧姆社学习漫画系列和众多以漫画和图示的方式讲解知识的优秀图书都值得推荐。但是有利有弊，大部分的漫画图书往往以基础知识为主，深入的研究还是要搜索科学文献和相关论著。同时漫画形式可以引起读者的兴趣，但是因为穿插故事情节，知识点不易定位，核心的知识内容相对分散，阅读上也要花费更多的时间。因此想学习一门知识，想以哪种形式入手，需要根据个人的情况确定。

• 频数(次数)分布表和直方图

建立简单数据示例，数据来源于《漫画统计学》[2]中'美味拉面畅销前50'上刊载的拉面馆的拉面价格。虽然一组数据通常使用pandas.Series()建立Series格式数据，但是后续分析会加入新的数据，因此仍旧建立DataFrame格式数据。使用df.describe()可以粗略查看主要统计值。

参考文献

1. (日)高桥麻奈著,崔建锋译,株式会社TREND-PRO漫画制作.漫画数据库[M].科学出版社.北京.
2. (日)高桥信著,陈刚译,株式会社TREND-PRO漫画制作.漫画统计学[M].科学出版社.北京.

但是由于 markdown 的设置问题，编号不从 1 开始，会丢失换行格式

情分布假设概率。同时，需要注意，正态分布是统计学中所谓的理论分布，即使少有数据，也服从正态分布，而是近似于，也有可能相差较远。违背正态分布假设，则依据正态分布假设所计算的概率结果将不再有效。

> 对于统计学的知识，通常是依据已出版的著作或者教材为依据进行解说，并根据阐述的内容，以python语言为工具分析数据的变化，从而可以直接应用代码来解决对应的问题，并能够通过具体的数据分析更深入的理解统计学相关的知识[3]。

> 参考文献

- > 3. Timothy C.Urdan.Statistics in Plain English(白话统计学)[M].中国人民大学出版社.2013,12.第3版。

对于统计学的知识，通常是依据已出版的著作或者教材为依据进行解说，并根据阐述的内容，以python语言为工具分析数据的变化，从而可以直接应用代码来解决对应的问题，并能够通过具体的数据分析更深入的理解统计学相关的知识[3]。

参考文献 3. Timothy C.Urdan.Statistics in Plain English(白话统计学)[M].中国人民大学出版社.2013.12.第3版。