|  |  |
| --- | --- |
| 学号 | 10212818122 |



课 程 设 计

课程名称 python程序设计课程设计

|  |  |
| --- | --- |
| 题 目 | 数据分析的学习与实践-地震的数据分析 |
| 专 业 | 软件工程 |
| 班 级 | 软件1181 |
| 姓 名 | 黄 婷 |
| 成 绩 |  |
| 指 导 老 师 | 徐 勇 |

2021 年 1 月 11 日至 2021 年 1 月 22 日

武汉华夏理工学院信息工程学院

**课 程 设 计 任 务 书**

课程名称：python程序设计课程设计 指导教师：徐勇

班级名称： 软件1181-1182 开课院、系：计算机与网络工程系

**一、课程设计目的与任务**

“python程序设计课程设计”是一个综合性的学习实践型实验教学环节，将在“python程序设计”课程的授课基础上，对python的基础语法、python 对文件的操作、python 对数据的操作、数据处理基础、数据可视化等若干个知识点进行综合运用。

python是一种具有天然开源基因的编程语言，了解开源社区和广泛的使用开源工具，也是Python学习实践的重要环节。因此在本次课程设计中，要求掌握jupyter notebook、Git等常用工具，以及github等重要开源社区的使用。

**二、课程设计的内容与基本要求**

数据分析的基本技术和方法在“python程序设计”课程中已经进行完整的讲授，由于数据分析本身是基于业务场景的，因此本次课程设计环节更加偏重于实际的业务场景的实践。通过对近期互联网热点的调查，准备了15个具体的业务场景，用于本次课程设计的具体任务场景。具体包括如下：

1. 北上广深租房状况分析；
2. 蔡某坤粉丝数及转发数据真假状况分析；
3. 地震的数据分析；
4. 英文名字的数据分析；
5. 外籍英文老师收入虚高情况数据分析；
6. 我国城市空气污染和烟花燃放的关系分析；
7. 针对996工作，程序员群体的看法的分析；
8. 吴某凡微博热点的分析；
9. 节假日长假景点人满为患的数据分析；
10. 针对荔枝的品种、销售地等维度，进行价格数据分析；
11. 分析芒果TV《我是大侦探》的观众评论数据；
12. 针对当前儿科医生的缺乏，对相关数据进行分析；
13. 著名网游《绝地求生》的数据分析；
14. 实习岗位状况的数据分析；
15. 电影《流浪地球》的观众评价的数据分析。

每三个学生组建一个课程设计小组，最后的任务输出包括代码每小组一份、课程设计报告每人一份、答辩ppt每小组一份，并进行课程设计成果答辩。小组成员均参与前述工作，但是每个人的侧重点不同。

每个课程设计小组可以从上述15个场景中选取一个作为课程设计的选题，选题中提供了待分析的数据，和现有的分析方法。各小组，通过学习和实践现有的分析方法，理解实战分析的思维过程并锻炼实际动手能力，再此基础上可以扩展更多维度的分析和数据展现形式。

每个班每个选题最多只能被两个小组选中，先选先得。同时，如果各小组发现更有意思的场景，并能够获取到相关待分析的数据，也可以申请作为选题方向。

本次课程设计的目标是培养学生的团队协作能力、对python知识点的综合运用、对实际场景的理解和适应能力、针对答辩的表达能力等。注重过程，期待成果，但不强求结果的尽善尽美。

**三、学时分配进度安排**

|  |  |  |
| --- | --- | --- |
| 序号 | 设计内容 | 所用时间 |
| 1 | 下发任务书，学生查阅资料 | 1天 |
| 2 | 组建团队，并进行选题和团队匹配 | 1天 |
| 3 | 各小组对自己的选题开始进行研究 | 2天 |
| 4 | 各小组开始准备数据并形成初步处理意见 | 1天 |
| 5 | 代码和实现分析的动手实践 | 2天 |
| 6 | 形成初步的报告书和ppt | 1天 |
| 7 | 答辩并完成报告书 | 2天 |
| 合 计 | | 2周 |

**四、课程设计考核及评分标准**

**1.设计报告要求**

课程设计报告要求逻辑清晰、层次分明、书写整洁。课程设计报告为每人一份，同一个小组的各成员的整体报告内容结构一致，但个人侧重点不同，个人着重撰写自己的工作内容，其他人的内容只要体现文档结构的完整性即可。

课程设计考核将综合考虑学生考勤和参与度、团队协作能力，过程管理能力、成果达成情况等。

**2.过程要求**

整个过程要求通过github来进行过程化发布，即阶段性地在github上提交过程结果。

**3.评分标准**

|  |  |
| --- | --- |
| **评分依据** | **评分成绩** |
| 1．团队协作能力 | 25分 |
| 2．python综合运用能力 | 25分 |
| 3．态度认真、刻苦钻研、遵守纪律 | 10分 |
| 4．过程完成、对工具的使用、对github的运用 | 20分 |
| 5．课程设计答辩逻辑清晰，内容正确 | 10分 |
| 6. 课程设计期间的课堂考勤、创新能力 | 10分 |
| 总分 | 100分 |

注：按上述六项分别记分后求和，根据小组成员贡献率综合评定，记载个人最后成绩。

成绩等级：优（90分—100分）、良（80分—89分）、中（70分—79分）、及格（60分—69分）、60分以下为不及格。

**五、指导地点与时间**

本课程设计将安排在第20-21周，采用腾讯会议和QQ群的方式，以在线形式进行。具体安排如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 周次 | 星期一 | 星期二 | 星期三 | 星期四 | 星期五 |
| 第20周 | 第5-8节 | 第5-8节 |  | 第5-8节 |  |
| 第21周 |  | 第5-8节 |  | 第5-8节 |  |

执笔:徐 勇 日期：2021-1-8

审阅:钱小红 日期：2021-1-8

目 录

[1.项目目的 1](#_Toc61701282)

[2.项目内容 1](#_Toc61701287)

[2.1项目选题 1](#_Toc61701288)

[2.2项目介绍 1](#_Toc61701291)

[3.编程环境及准备 1](#_Toc61701294)

[3.1硬件平台 1](#_Toc61701295)

3.[2项目上传平台GitHub 1](#_Toc61701296)

[3.3开发平台jupyter 2](#_Toc61701297)

[3.4编程语言python 2](#_Toc61701298)

[4.数据采集 2](#_Toc61701299)

[4.1数据库引入和查看 3](#_Toc61701300)

[4.2 数据库表注释 3](#_Toc61701301)

[4.3数据分析库注解 4](#_Toc61701302)

[5. 数据分析 5](#_Toc61701310)

[5.1问题探索 5](#_Toc61701311)

[5.1.1引发地震的因素有哪些？ 5](#_Toc61701312)

[5.1.2全世界地震频发的地区有哪些？ 6](#_Toc61701313)

[5.1.3有哪些引发全世界舆情关注的大地震？ 7](#_Toc61701314)

[5.1.4为什么好像2008年之后四川经常发生地震？ 8](#_Toc61701315)

[5.2 任务分配 9](#_Toc61701317)

[6.总结与心得 1](#_Toc61701316)0

[课程设计成绩评定表 1](#_Toc61701317)1

**1.项目目的**

1. 通过综合实践练习，能够掌握Python基础语法，能够应用Python语言编程解决某一具体领域的应用问题。
2. 通过综合实践练习，能够理论联系实际，提升分析和解决问题的能力。
3. 学会使用jupyter notebook、Git等常用工具，以及github等重要开源社区的使用。

**2.项目设计内容**

**2.1项目选题**

地震的数据分析

**2.2项目介绍**

地震（英文名称:earthquake）又称地动、地振动，是地壳快速释放能量过程中造成的振动，期间会产生地震波的一种自然现象。地球上板块与板块之间相互挤压碰撞，造成板块边沿及板块内部产生错动和破裂，是引起地震的主要原因。 地震开始发生的地点称为震源，震源正上方的地面称为震中。破坏性地震的地面振动最烈处称为极震区，极震区往往也就是震中所在的地区。地震常常造成严重人员伤亡，能引起火灾、水灾、有毒气体泄漏、细菌及放射性物质扩散，还可能造成海啸、滑坡、崩塌、地裂缝等次生灾害。

**3.编程环境及设备**

**3.1硬件平台**

个人计算机：Windows 32/64

内存：16G

硬盘空间：SSD-120G

## 3.2项目上传管理平台--GitHub

GitHub可以托管各种git库，并提供一个web界面，但它与外国的[SourceForge](https://baike.baidu.com/item/SourceForge/6562141)、[Google Code](https://baike.baidu.com/item/Google%20Code)或中国的[coding](https://baike.baidu.com/item/coding/8921246)的服务不同，GitHub的独特卖点在于从另外一个项目进行分支的简易性。在GitHub，用户可以十分轻易地找到海量的[开源](https://baike.baidu.com/item/%E5%BC%80%E6%BA%90/20720669)代码。为一个项目贡献代码非常简单：首先点击项目站点的“fork”的按钮，然后将代码检出并将修改加入到刚才分出的代码库中，最后通过内建的“pull request”机制向项目负责人申请代码合并。已经有人将GitHub称为代码玩家的MySpace。

在GitHub进行分支就像在[Myspace](https://baike.baidu.com/item/Myspace)（或[Facebook](https://baike.baidu.com/item/Facebook)…）进行交友一样，在社会关系图的节点中不断的连线。

GitHub项目本身自然而然的也在GitHub上进行托管，只不过在一个私有的，公共视图不可见的库中。[开源项目](https://baike.baidu.com/item/%E5%BC%80%E6%BA%90%E9%A1%B9%E7%9B%AE)可以免费托管，但私有库则并不如此。Chris Wanstrath，GitHub的开发者之一，肯定了通过付费的私有库来在财务上支持免费库的托管这一计划。

## 3.3开发平台--jupyter

Jupyter Notebook（此前被称为 IPython notebook）是一个交互式笔记本，支持运行 40 多种编程语言。

Jupyter Notebook 的本质是一个 Web[应用程序](https://baike.baidu.com/item/%20%E5%BA%94%E7%94%A8%E7%A8%8B%E5%BA%8F/5985445)，便于创建和共享文学化程序文档，支持实时代码，数学方程，可视化和[markdown](https://baike.baidu.com/item/markdown/3245829)。 用途包括：数据清理和转换，数值模拟，统计建模，机器学习等等

用户可以通过电子邮件，Dropbox，GitHub 和 Jupyter Notebook Viewer，将 Jupyter Notebook 分享给其他人。

在Jupyter Notebook 中，代码可以实时的生成图像，视频，LaTeX和JavaScript。

## 3.4编程语言--python

Python是一种面向对象、解释型、弱类型的脚本语言，它也是一种功能强大而完善的通用型语言。相比其他编程语言（比如Java）,Python代码非常简单，上手非常容易。比如我们要完成某个功能，如果用 Java 需要 100 行代码，但用Python可能只需要20行代码，这是Python具有巨大吸引力的一大特点。

Python的两大特色是清晰的语法和可扩展性：  
Python的语法非常清晰，它甚至不是一种格式自由的语言。例如，它要求if语句的下一行必须向右缩进，否则不能通过编译。Python的可扩展性体现为它的模块，Python具有脚本语言中最丰富和强大的类库（这些类库被形象地称为“batteries included ，内置电池”），这些类库覆盖了文件 I/O、GUI、网络编程、数据库访问、文本操作等绝大部分应用场景。

**4.数据采集**

数据集有csv、json、xlsx等格式，可以储存在本地或者服务器上，本项目多采用的是csv。在分析数据之前需要将数据集导入到Jupyter中。  
 **4.1数据库引入和查看**  
#数据库引入

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

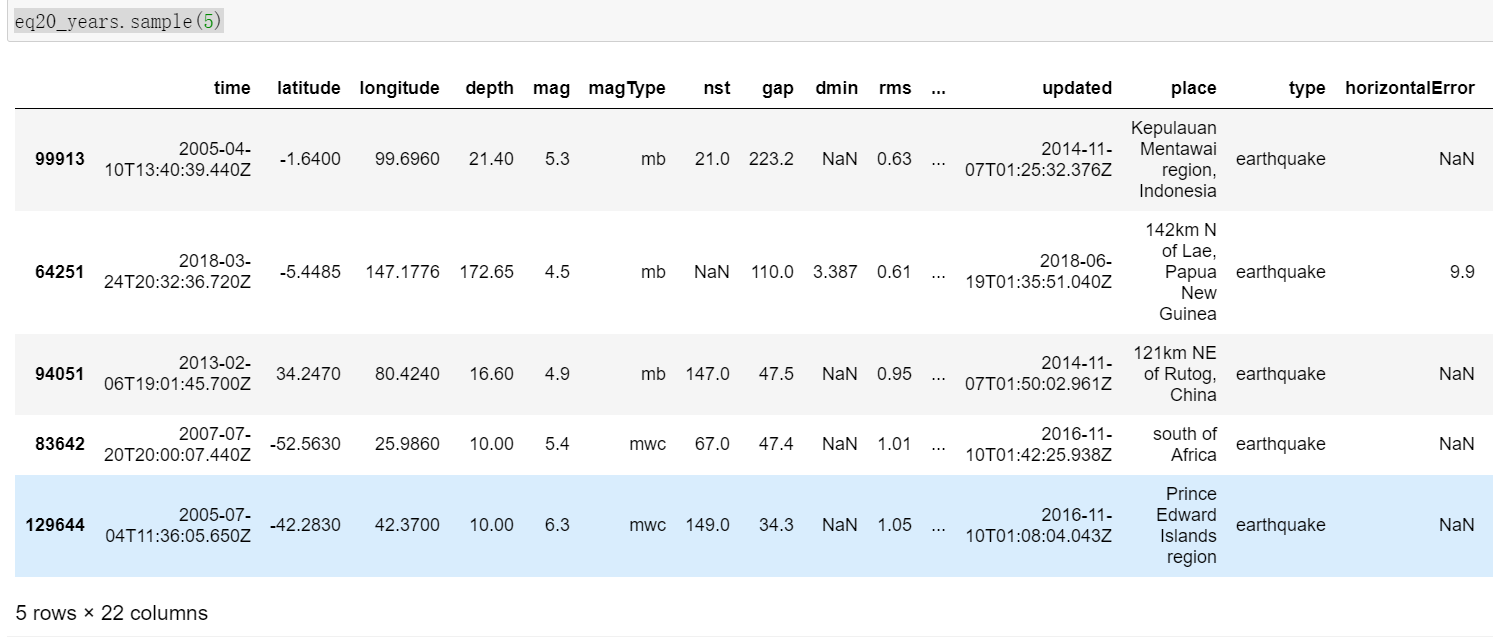
#引用函数pd.read\_csv读取csv文件  
eq20\_years = pd.read\_csv('1999-2019.csv')

eq1\_year = pd.read\_csv('earthquake.csv')

eq\_sichuan = pd.read\_csv('sichuan.csv')

#查看前'1999-2019.csv'数据，sample函数

eq20\_years.sample(5)



**4.2数据表注释：**

（1）1999-2019.csv数据文件的数据是1999年-2019年全球的地震信息数据

（2）eartquake.csv数据文件的数据代表的是全球热点地区的地震信息数据

（3）sichuan.csv数据文件的数据代表的是中国四川近年来的地震信息数据

**4.3数据分析库注解**

1.numpy的全称是Numerical Python，是Python的一个扩展程序库，它不仅针对数组运算提供了大量的函数库，而且它还能够支持维度数组与矩阵运算。重要的是，numpy内部解除了CPython中的全局解释器锁（GIL），运行效率非常好，是处理大量数组类结构和机器学习框架的基础库！

2.pandas是一个强大的分析结构化数据的工具集；它的使用基础是Numpy（提供高性能的矩阵运算）；用于数据挖掘和数据分析，同时也提供数据清洗功能。

3.Seaborn其实是在matplotlib的基础上进行了更高级的API封装，从而使得作图更加容易，在大多数情况下使用seaborn就能做出很具有吸引力的图，而使matplotlib就能制作具有更多特色的图。

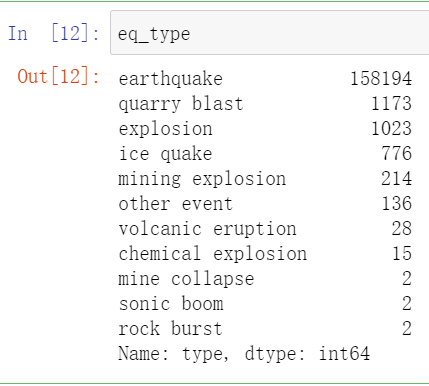
4.使用%matplotlib命令可以将matplotlib的图表直接嵌入到Notebook之中，或者使用指定的界面库显示图表，它有一个参数指定matplotlib图表的显示方式。inline表示将图表嵌入到Notebook中。

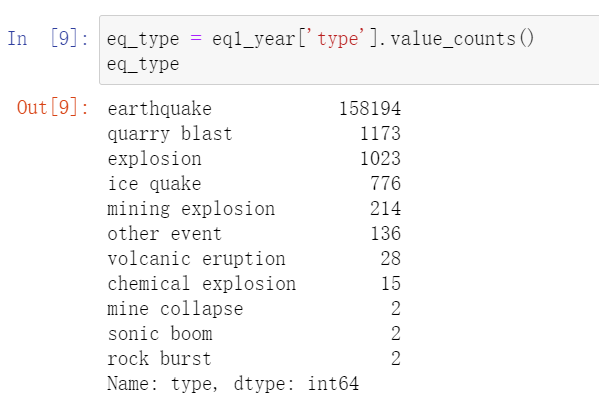
**5.数据分析**

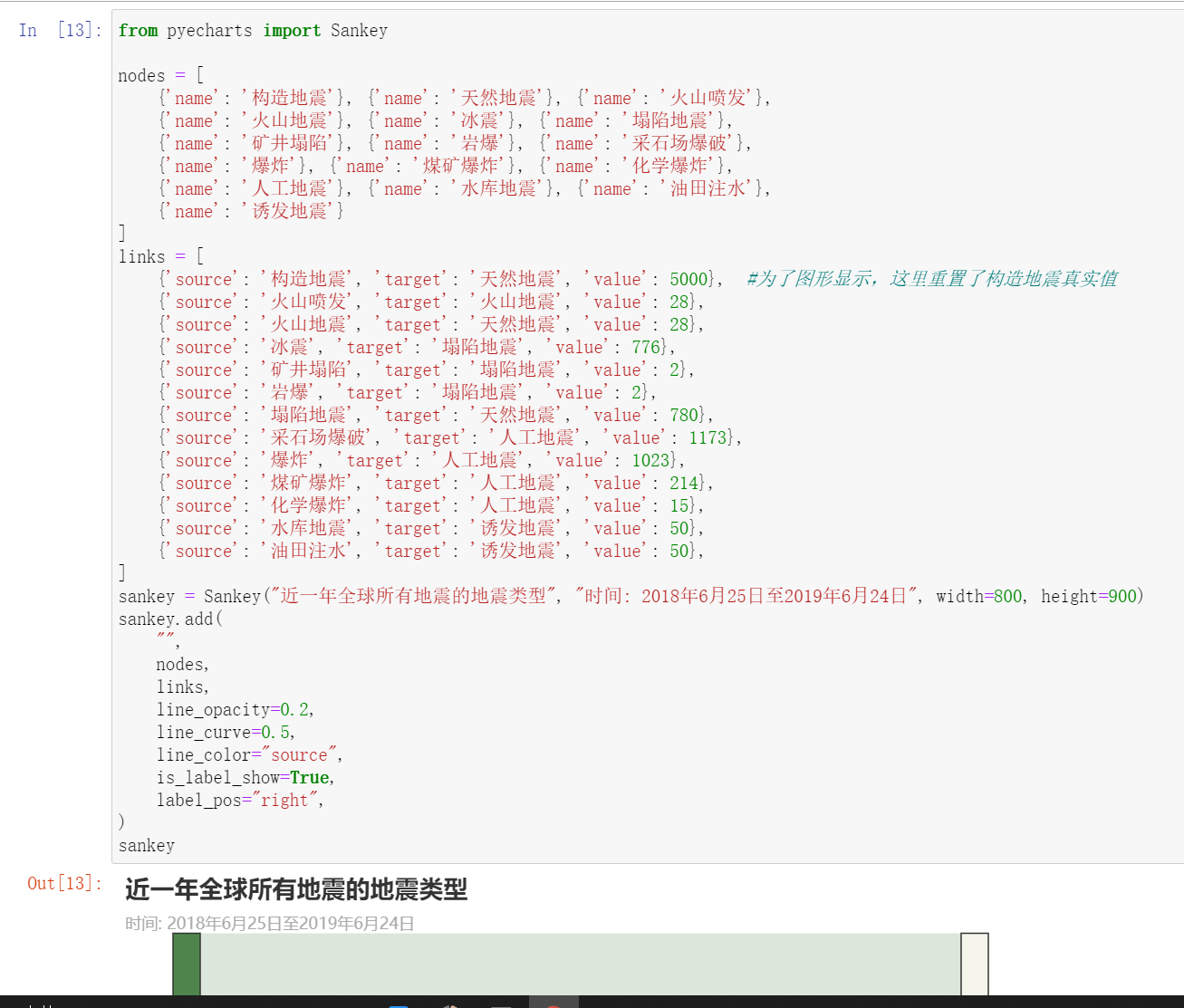
据统计，地球上每年约发生500多万次地震，即每天要发生上万次的地震。其中绝大多数太小或太远，以至于人们感觉不到；真正能对人类造成严重危害的地震大约有十几二十次；能造成特别严重灾害的地震大约有一两次。

**5.1问题探索**

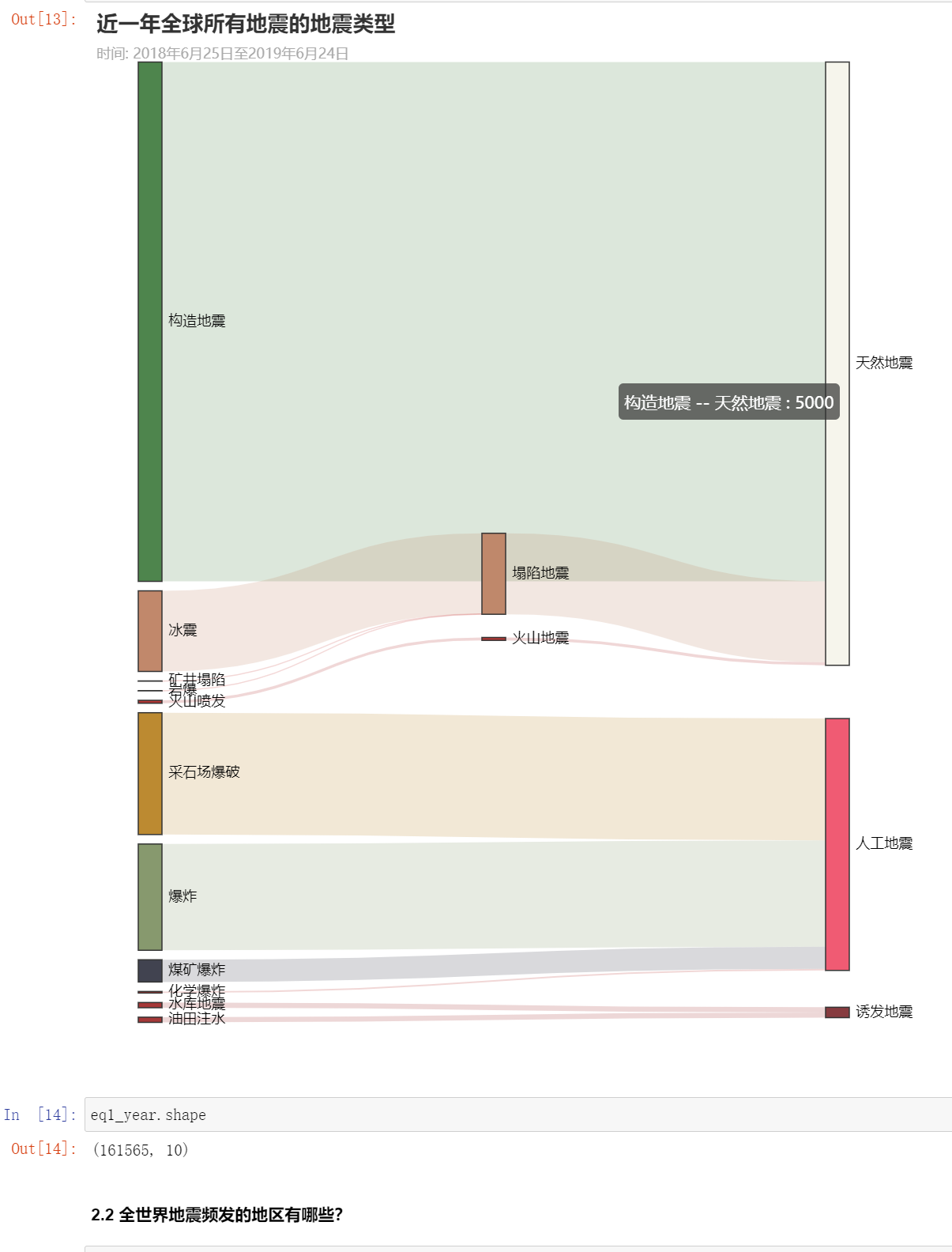
**5.1.1引发地震的因素有哪些？**

部分代码：





图表结果显示：



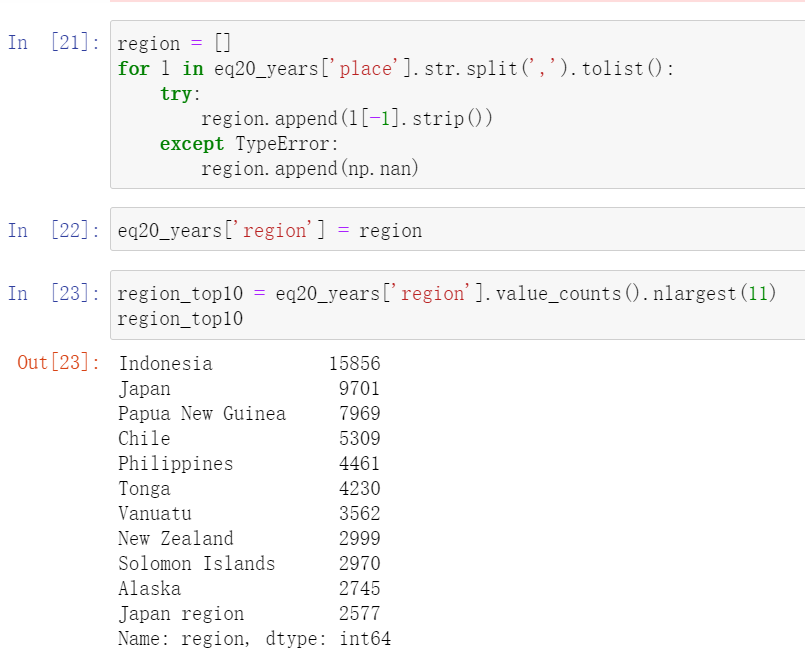
分析及结论：

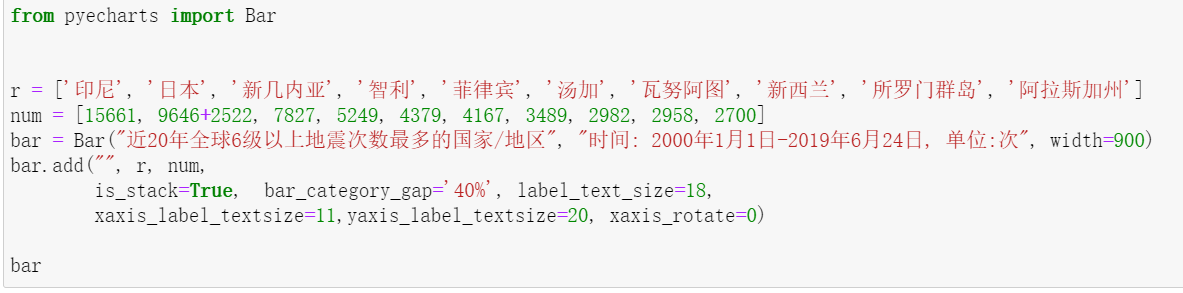
近一年里，地震主要包括天然地震和人工地震，天然地震是自然因素引发的地震，主要有构造地震和塌陷地震,构造地震约占地震总数的90%以上。其次是由塌陷发引起的地震,称为塌陷地震,约占地震总数的7%。此外,某些特殊情况下了也会产生地震,如火山爆发(火山地震) 等。它是由于地下深处岩石破裂、错动把长期积累起来的能量急剧释放出来,以地震波的形式向四面八方传播出去,到地面引起的房摇地动。

人工地震是由人为活动引起的地震，主要是采石场爆破和各种东西引起的爆炸，其外，就是诱发地震，相对来说占得比例较小，主要是些化学爆炸和水库地震，油田注水现在发生的较为少一些。

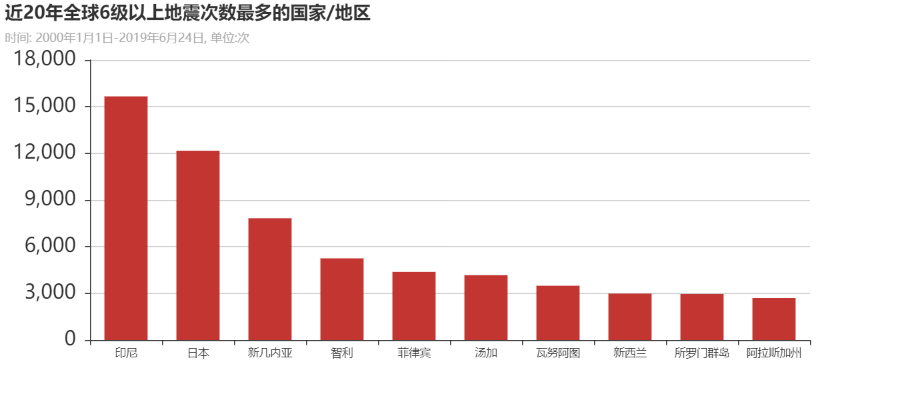
**5.1.2全世界地震频发的地区有哪些？**

通过数据分析筛选目前最多的10个地震频发的地区，部分代码：





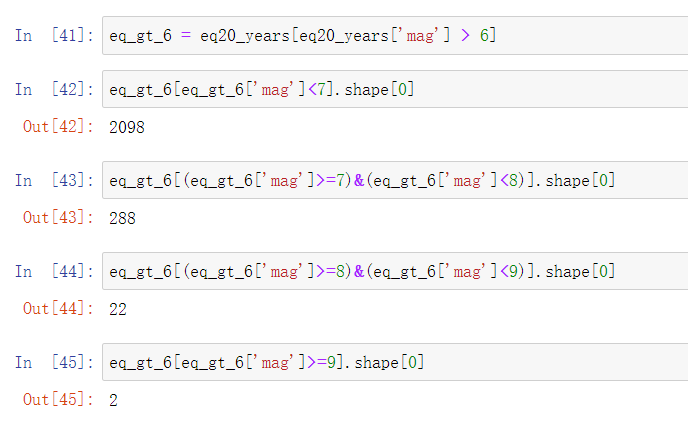
柱状图结果显示：

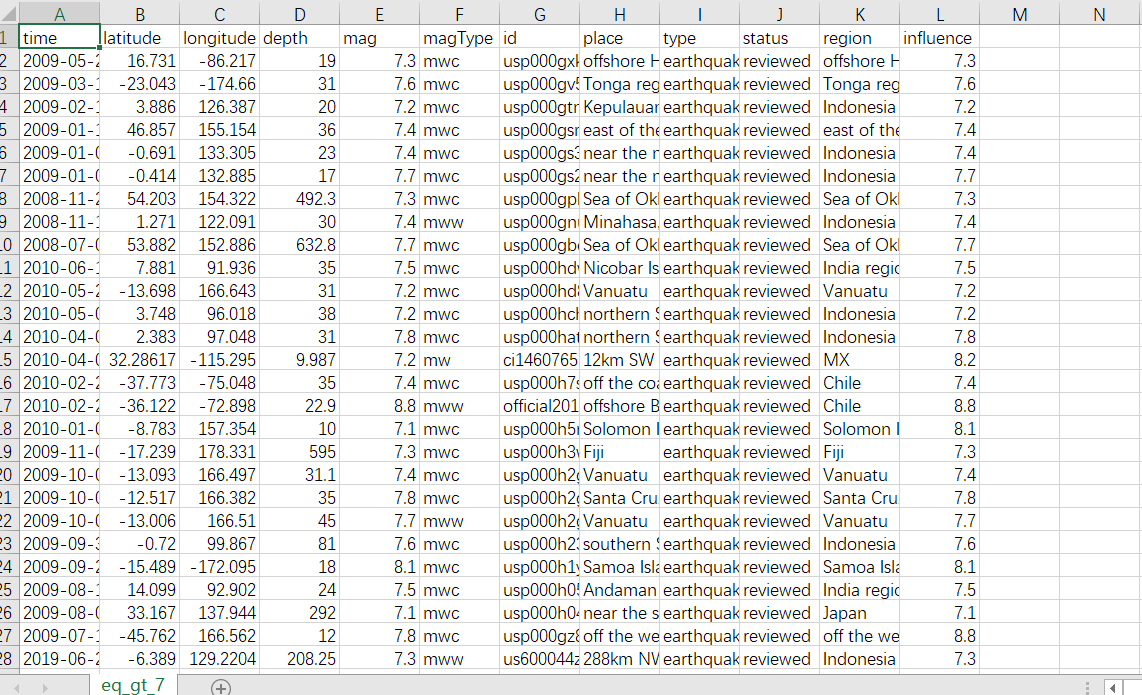


分析与总结:在近二十年来全球6级以上地震次数排名前十的国家，最多的国家就是印尼高达15000次；排名第二的就是日本高达12000；排名第三的是新几内亚高达7827次；排名第四的是智利高达5249次；排名前四的都达到了5000次以上，后六位比较平均，都在3000次左右。

**5.1.3有哪些引发全世界舆情关注的大地震？**

部分代码：

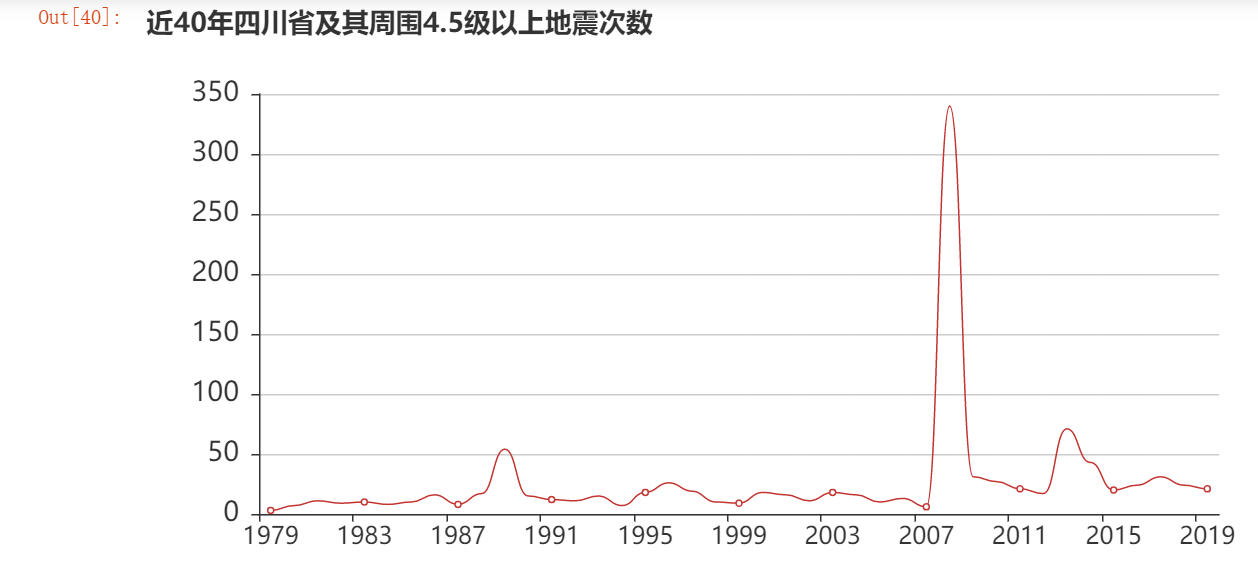




**5.4为什么四川省及其周围好像从2008年开始就地震频发？**

有很多朋友可能有这样一些疑问：怎么四川省自从2008年汶川地震之后就好像经常发生地震啊？2008年之前好像都没有听说过四川省有地震呢？

真的是这样吗？我们获取了1979年以来近40年四川省及其周围4.5级以上地震的所有数据，并且绘制出每年四川省及其周围地区发生4.5级以上地震次数图。



分析与总结：在这个问题探究中，我们小组只得出了这个地震次数折线图，可以知道从2007年开始快速上升最主要的原因就是四川省及其周围都处于地震带，而且四川全省都是山区，容易频发大地震，从四川地理位置来看，四川省尤其是四川西部位于四川龙门山地震带上，因此地震多且强度大。从整体位置来看，四川地处青藏高原东缘，地处青藏高原与四川盆地交汇处、云贵高原与四川盆地交汇处。印度洋板块与欧亚板块的挤压、碰撞导致地壳隆起、变形并且抬升，形成了青藏高原，又因为地壳的挤压和碰撞使得地壳不稳定，形成了断裂带，这就是为何青藏地区多地震的原因。板块内部并不是完整的一块，而是由小板块组成的，中国南方大部分均位于扬子准地台上，四川盆地也位于其上。

2008年以后四川地震很频繁呢？我们总结了以下一些原因：

1. 因为四川省位于地震带，四周全是山，发生地震的几率非常高；

2. 科技的发达，出现朋友圈夸大造谣等恶劣的现象；

3. 受2008年汶川地震的影响，人们对地震有了更多的关注，四川大地震导致的部分地区地壳松动会更容易导致地震的发生。

**5.2任务分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **小组分工明细表** | | | | | |
| **项目名称** | **地震的数据分析** | | | | |
| **小组成员** | **黄婷** | **陈廷峰** | | **李思非** | |
| **任务名称** | **主负责人** | | **协助人** | | **协助人** |
| **选题** | **黄婷** | | **陈廷峰** | | **李思非** |
| **小组讨论分析** | **黄婷** | | **陈廷峰** | | **李思非** |
| **建立github** | **李思非** | | **陈廷峰** | | **黄婷** |
| **数据采集** | **黄婷** | | **无** | | **无** |
| **环境搭建** | **三人共同完成** | | | | |
| **编码** | **三人共同参与** | | | | |
| **数据图形化** | **陈廷峰** | | **黄婷** | | **李思非** |
| **数据分析** | **黄婷** | | **陈廷峰** | | **李思非** |
| **代码检查** | **陈廷峰** | | **黄婷** | | **李思非** |
| **注释文档编写** | **陈廷峰** | | **无** | | **无** |
| **PPT报告编写** | **黄婷** | | **陈廷峰** | | **无** |
| **Word报告编写** | **黄婷** | | **陈廷峰** | | **李思非** |
| **上传项目文件** | **李思非** | | **无** | | **无** |
| **Github管理** | **李思非** | | **陈廷峰** | | **无** |
| **项目汇报** | **陈廷峰** | | **无** | | **无** |

**6.总结与心得**

这次Python的课程设计让我学会了GitHub和jupyther的使用，使用之后认为juputer 相比pycharm、ipython在画图和数据展示方面更有优势；同时对Python的实际应用又进一步地了解，也感受到了团队的力量，正如“三个臭皮匠顶个诸葛亮”我们共同参与到每个地方，但是每个人的侧重点不同，组长黄婷，也就是我主要负责组内成员上课考勤及纪律，项目上主要负责前期的数据收集以及PPT和报告的制作；陈霆锋同学主要负责代码的运行及测试并完成答辩；李思非同学主要负责数据清洗以及GitHub的上传。

在设计过程中虽然遇到了一些问题，但经过一次又一次的思考，一遍又一遍的检查终于找出了原因所在，也暴露出了前期我在这方面的知识欠缺和经验不足。实践出真知，通过亲自动手制作，使我们掌握的知识不再是纸上谈兵。在课程设计过程中，我们不断发现错误，不断改正，不断领悟，不断获取。最终的检验修改环节，本身就是在践行“过而能改，善莫大焉”的知行观。这次课程设计终于顺利完成了，在设计中遇到了很多问题，最后在不懈的努力下，终于迎刃而解。在今后社会的发展和学习实践过程中，一定要不懈努力，不能遇到问题就想到要退缩，一定要不厌其烦的发现问题所在，然后一一进行解决，只有这样，才能成功的做成想做的事，才能在今后的道路上劈荆斩棘，而不是知难而退，那样永远不可能收获成功，收获喜悦，也永远不可能得到社会及他人对你的认可！课程设计不仅是一门专业课，给我很多专业知识以及专业技能上的提升，同时又是一门辩思课，给了我很多思路，给了我莫大的空间。

地震是我们每个人都不愿意听到的一个话题，地震是全球上所有自然灾害中给人类社会造成损失最大的一种地质灾害。破坏性地震，往往在没有什么预兆的情况下突然来临，大地震撼、地裂房塌，甚至摧毁整座城市，并且在地震之后，火灾、水灾、瘟疫等严重次生灾害更是雪上加霜，给人类带来了极大的灾难。我们应该早预防早防控多演练，让我们在遇到地震时可以保持冷静，尽量让伤害缩到最小，让我们拥有更美好的家园。

**课程设计成绩评定表**

|  |  |  |  |
| --- | --- | --- | --- |
| 课程设计题目 | 地震的数据分析 | | |
| 课程设计学生答辩或质疑记录：  1. 编码过程中遇到什么问题？ 答：刚开始跑代码的时候，图表出不来，一直报错，后面发现是pyecharts的版本太低，我们换成了0.5.5的版本就可以正常运行。  2. 可视化扩展库matplotlib的模块pyplot中哪个函数可以用来设置同一个画布中多个子图之间的水平间距和垂直间距？ 答： subplots\_adjust()函数。  3. %matplotlib是干嘛的？ 答：使用%matplotlib命令可以将matplotlib的图表直接嵌入到Notebook之中，或者使用指定的界面库显示图表，它有一个参数指定matplotlib图表的显示方式。inline表示将图表嵌入到Notebook中。 | | | |
| **评 分 依 据** | | **分 值** | **评分成绩** |
| 1．团队协作能力 | | 25分 |  |
| 2．python综合运用能力 | | 25分 |  |
| 3．态度认真、刻苦钻研、创新能力 | | 10分 |  |
| 4．过程完成、对工具的使用、对github的运用 | | 20分 |  |
| 5．课程设计答辩逻辑清晰，内容正确 | | 10分 |  |
| 6. 课程设计期间的课堂考勤、遵守纪律 | | 10分 |  |
| 总 分 | | 100分 |  |
| 最终评定等级为：  指导老师签字：  2021 年 1 月 15日 | | | |