



项目名称 濒危野生动植物分类及相关信息数据可视化

学 院 计算机学院（软件学院）

项目成员 郭智溢 万珂蓝 王枫 杨凯

指导教师 朱 敏

目录

[1.选题背景 1](#_Toc20681_WPSOffice_Level1)

[2.项目成员与分工 1](#_Toc31769_WPSOffice_Level1)

[3.数据说明 1](#_Toc9609_WPSOffice_Level1)

[3.1数据来源 1](#_Toc31769_WPSOffice_Level2)

[3.2数据预处理 1](#_Toc9609_WPSOffice_Level2)

[4.研究内容描述 1](#_Toc5073_WPSOffice_Level1)

[4.1调查同类型的数据可视化分析案例 1](#_Toc5073_WPSOffice_Level2)

[4.2数据的前期分析 2](#_Toc10210_WPSOffice_Level2)

[4.3数据的清洗 2](#_Toc13030_WPSOffice_Level2)

[4.4技术前期验证 4](#_Toc20579_WPSOffice_Level2)

[5.研究方法与技术路线 4](#_Toc10210_WPSOffice_Level1)

[5.1数据选取与视图构建 4](#_Toc16319_WPSOffice_Level2)

[5.2框架、工具选择 7](#_Toc20454_WPSOffice_Level2)

[6.分析结果与效果呈现 7](#_Toc13030_WPSOffice_Level1)

[6.1分析结果 7](#_Toc9136_WPSOffice_Level2)

[6.2呈现效果 7](#_Toc32453_WPSOffice_Level2)

[7.存在问题 7](#_Toc20579_WPSOffice_Level1)

[8.项目收获 8](#_Toc16319_WPSOffice_Level1)

1.选题背景

通过国际濒危物种贸易公约cites濒危动物进出口数据来分析濒危动物进出口流向、以及交易的数量变化，以及用途的变化

适用对象：

2.项目成员与分工

组名：visual-punk，组员分工如表1所示。

表1 项目成员与分工

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 组员 | 学号 | 导师 | 学院 | 任务 |
| 郭智溢 | 2018226045016 | 刘艳丽 | 视觉合成图形图像技术国防重点学科实验室 | 数据清洗、服务端后台实现 |
| 万珂蓝 | 2018223045160 | 冯伟森 | 计算机学院（软件学院） | 界面设计、前端实现、数据项选取、数据收集 |
| 王枫 | 2018223040103 | 吕泽均 | 计算机学院（软件学院） | 界面设计、数据对照表制作 |
| 杨凯 | 2018223040014 | 王明辉 | 计算机学院（软件学院） | 前端技术实验、文档、PPT制作 |

3.数据说明

3.1数据来源

cites濒危物种信息表（来源：https://www.cites.org ）：数据量共35455条，每条数据项共计15项（处理后）。

cites2000-2017年濒危物种交易数据：数据量共1498853条，每条数据项共计10条（处理后）。

ISO国家代码数据表：数据量共261条，每条数据项共计3项。

国家代码与国家经纬度信息对照表：数据量共243条，每条数据项共计6项。

3.2预处理

a.切割拼接cites交易数据的csv文件，去除非分析重点的数据项；

b.将数据导入数据库；

c.将焦点数据导出为json格式；

d.预处理json数据，将所需数据项再次拼接。

4.研究内容描述

4.1 调查同类型的数据可视化分析案例

从数据可视分析分享网站kaggle上，查询对于cites交易数据的相关研究，同时也参考了国家地理对于这部分数据所进行的研究，确定了在交易数据中物种的等级动植物分类、濒危等级的分类上做一定程度的简化，转而把首要的研究重点放在进出口的国家以及这些国家的进口来源、出口目的地的分析上，在此基础上去结合这些国家进出口动植物的数量、交易案件数量、目的等进行分析。

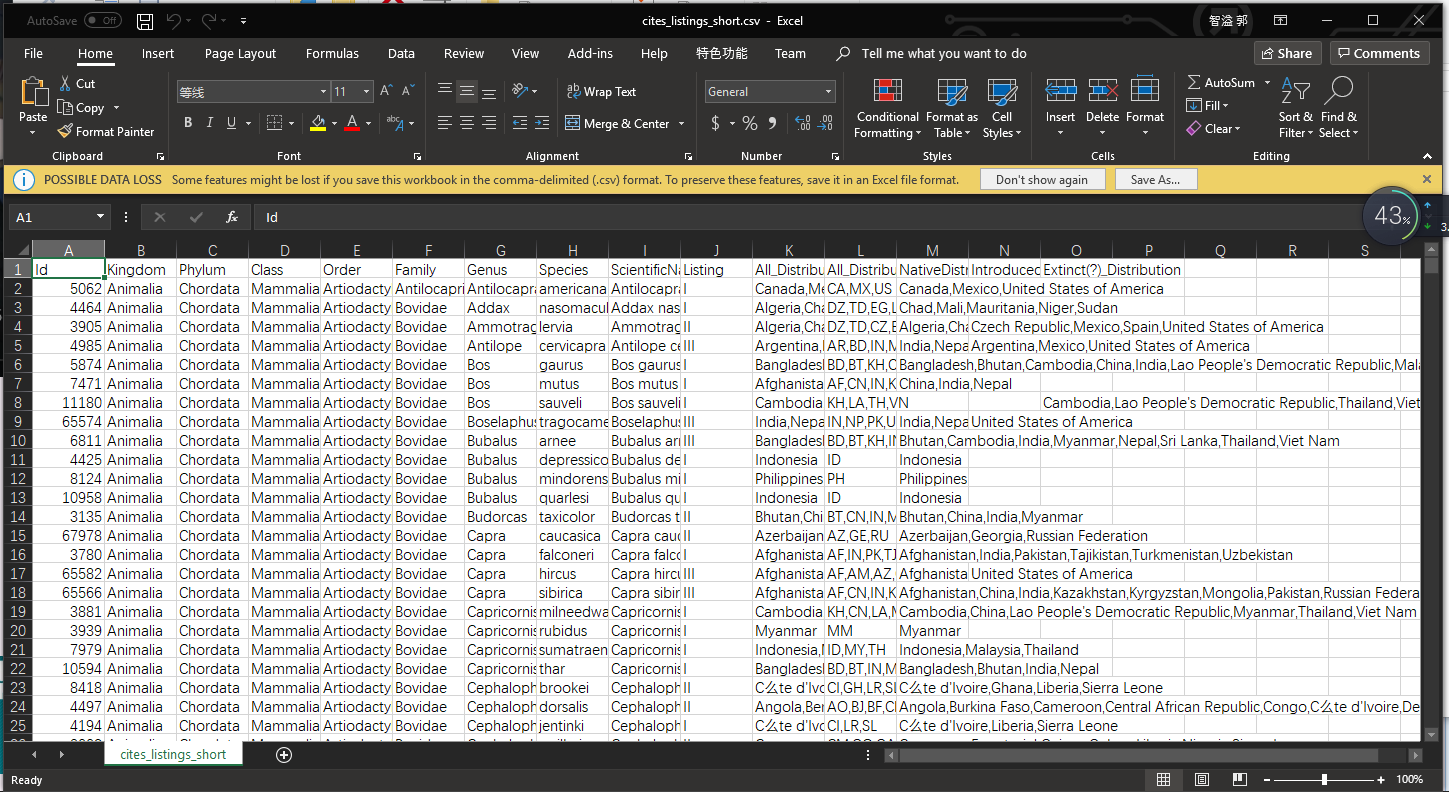
4.2 数据的前期分析

cites濒危动植物分类目录和交易目录中有很多数据项，需要组员对于这些数据项进行一定的了解，在查阅了官方的指导资料后，建立了中英文对照的说明书来指导数据的分析。

这次可视分析使用了4个有关联但又相互独立的数据表，需要确定每个数据表是如何关联的，例如通过交易数据表中每一项交易的动植物的名称，可以用于索引濒危动植物信息表，而其中的国家代码可以用于索引国家的经纬度信息（可用于地图视图定位）。

4.3 数据的清洗

因为交易数据量非常大，并且文件是按年份以csv来存储，不方便读取分析，因此将所有的数据导入到数据库，但测试发现数据库在索引这种直接导入的百万级数据，有明显的延迟，因此选择再次精简数据，同时将感兴趣的数据以json形式存储读取（例如预先存入进口来源国的地理数据、中英文名称），这样避免前端的操作因为数据库的读取而产生较大的延迟而影响了交互。

图4.1 csv数据表格

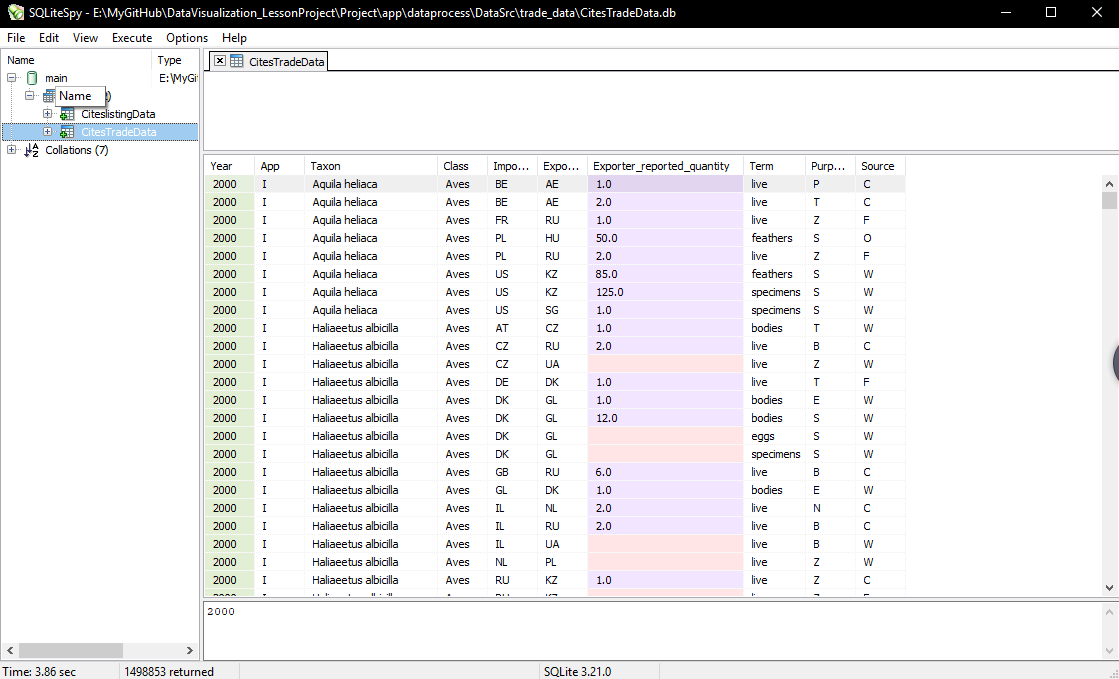
图4.2数据库导入

图4.3 json数据

国际地理项目的网址如下：

<https://www.nationalgeographic.com/animal-trade/#Actinopterygii-Sarcopterygii>

4.4 技术前期验证

4.4.1 前端技术验证

项目从立项到正式开发之前，对于所需要的技术做了很多研究与试验，在先验性经验不足的情况下，可视化框架经历了d3到pyecharts再到echarts的转变，因为echarts在使用上更符合团队对于前端知识的掌握情况，而pyecharts虽然使用更加方便，但实际提供的组件无法完成我们所希望的交互效果。

4.4.2 后端技术验证

后端上面使用了flask，为此负责前后端的组员学习了ajax技术，这也让本来预计是运行在本地的前端界面能够使用flask框架中使用的的python语言，在后台做更灵活的数据处理，也让整个工程更方便的部署在远程的服务器上。

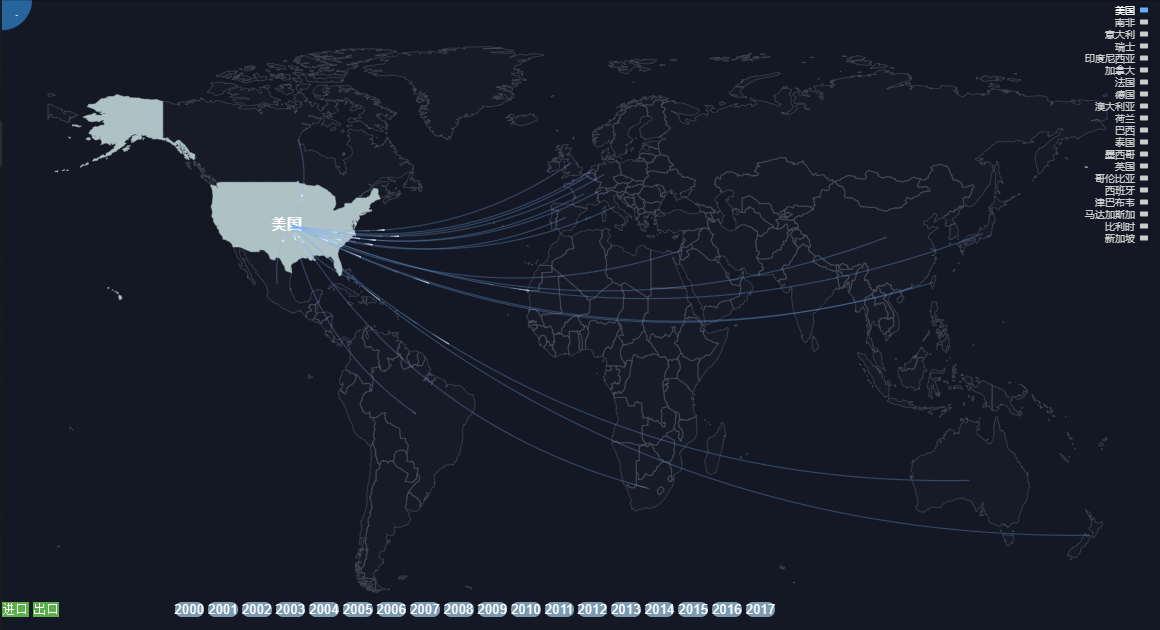
4.4.3 服务器部署

服务器使用了vultr提供的云主机，组员有在ubuntu系统下搭建简易服务器的框架，因此在选用了ubuntu 18.04为主机的系统，为最后快速的将工程部署到服务器上打下了基础。

5.研究方法与技术路线

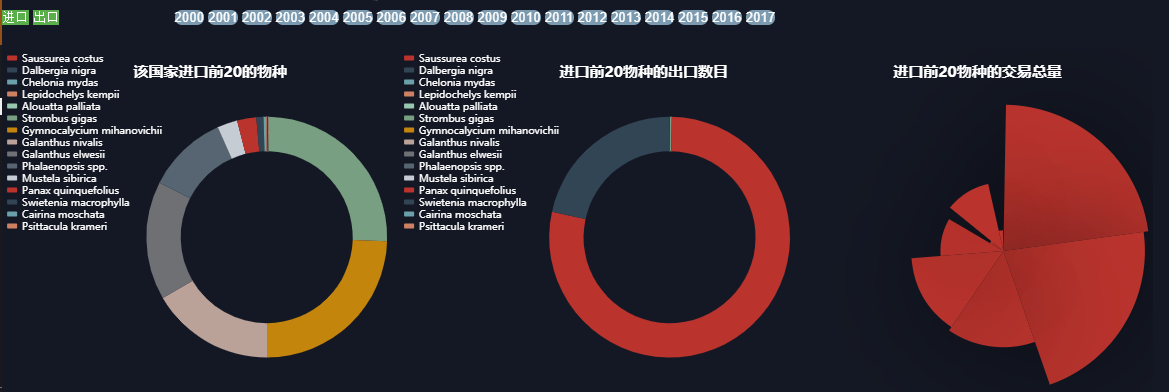
5.1数据选取与视图构建

2000年至2017年进出口前20的国家：使用世界地图标定这些国家，并使用航线图来标定这些国家的进口来源国或者出口目的地国家。

图5.1国家进口来源国和出口目的国

特点：可以直观地通过航线轨迹图，清除的知道每个国家的动植物进出口流入国或者进口国。

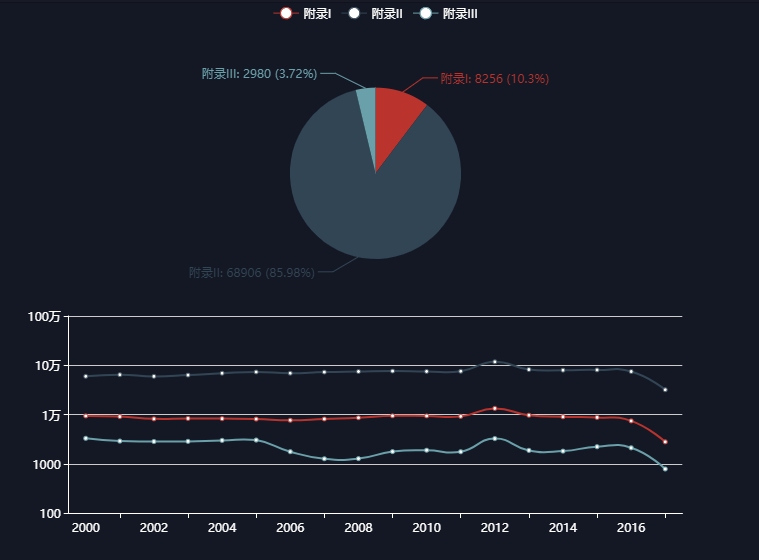
对于进口国家，会考察每年它们进口的来源、进口的物种种类、交易量以及目的：使用饼图视图来考察这些国家的进口数量前20的物种在全年所占的比重，并使用矩形树图来展现进口的目的。

图5.2进口国家详情

特点：饼图可以直观的体现数据的占比的情况.

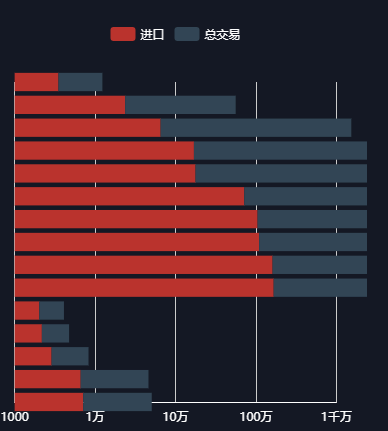
对于出口国家，会考察每年它们出口的目的地，出口的物种种类、交易量以及目的：与出口国类似。

全球濒危动植物物的交易总量：以时间线，配合饼图与折线图，来考察全球濒危动植物交易总数的趋势，以及I、II、III级濒危动植物在交易数量中占比变化。

图5.3全球濒危动植物物的交易总量

特点：以时间线为筛选条件的可视化图组，可以很明显的看到趋势的变化。

辅助视图：我们在工程中还使用了一些辅助视图来协助了解我们的研究点，在此选用矩形树图与横向柱状图。

图5.4辅助视图-矩形树图

矩形树图特点：通过数据大小排序，能够给出相关数据的占比情况，能够突出重要的数据。

图5.5辅助视图-柱状图

横向柱状图特点：用与对比不同筛选条件下的数据（例如进出口案件数量与进出口案件中所记录的总量进行对比），一定程度上可以很好的反映了我们想看到的交易活跃度。

5.2 框架、工具选择

可视化开源框架：echarts

团队协作、版本控制工具：github

后端服务器框架：flask

数据库：sqlite3

6.分析结果与呈现效果

6.1 分析结果

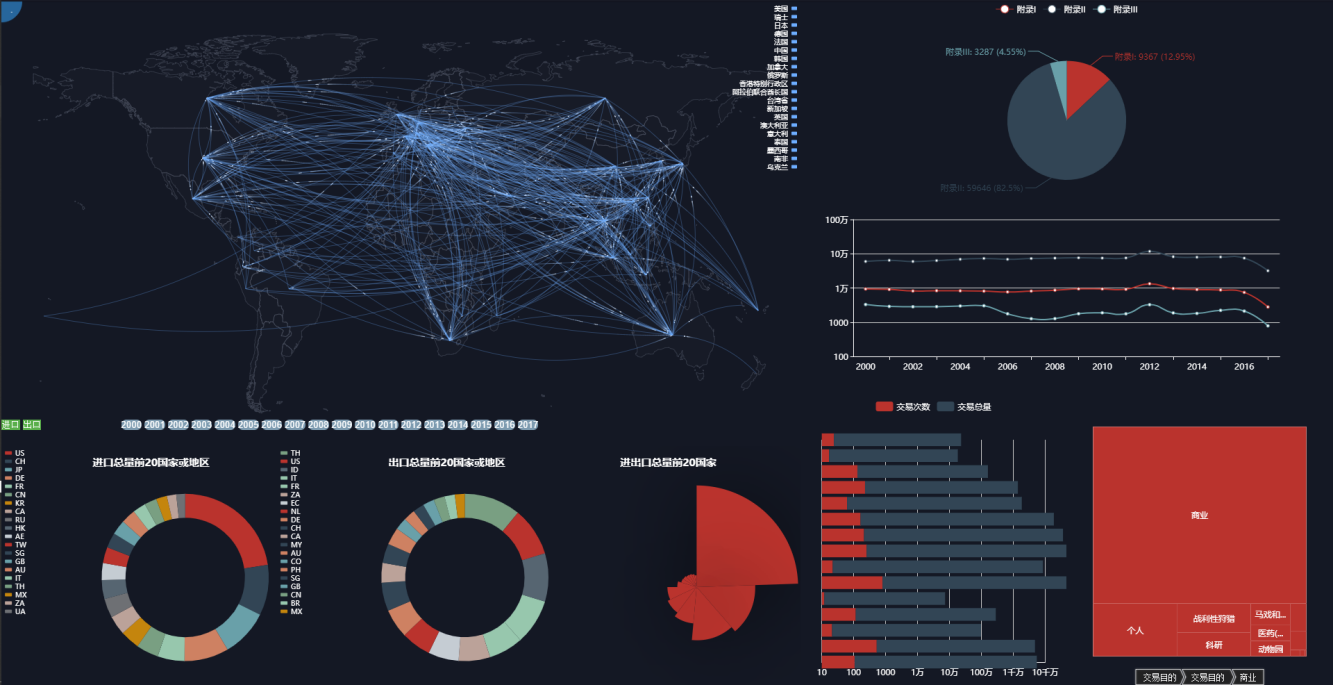
a.濒危动物交易量有所下降。

b.商业用途与个人用途始终占据主流 ,商业性狩猎与科研比重有所增加。

c.进口国家大多为欧美发达国家或者人口大国，中国印度排名上升较快。

6.2 呈现效果

网络地址：<http://66.42.44.7:5000>

图6.1最终效果

7.存在的问题

a.对动植物知识存在不足，无法从物种分类等方面进行更全面的分析。

动植物中英文名称对照表数据量较大，无法直观的将整个视图的所有信息翻译成中文，对于使用视图进行分析时，需要额外的工作。

b.所使用的数据条数很多，但数据维度比较少，可供分析的面难以做到更宽。

c.对比国家地理对于交易数据的分析，我们的可视化分析还存在交互不够流畅、板块信息不够明确的问题，在程序设计上也存在一些bug没有完美的解决，特别是在交易的动植物濒危等级分类上还没有做到更直观，影响了使用者更直观的观察分析结果。

8.项目收获

项目开发流程总时间虽短，但放眼到开始至结束，期间经历了近两个月，从git提交列表来看，全组时间安排比较合理，避免了最后时刻的赶工，项目基本在这两个的每周都有进展，这种协作经历是非常有意义与宝贵的。在研一课程比较紧张且多个课程有作业要求的状态下，组员也适当的调整了自己时间规划，高效保质的完成本课程项目每个时间节点所需要完成的工作。

本课程项目使用了Github来进行团队协作，虽然让工作效率得到了提高，但对于工作分配、文件修改冲突方面面临了一些挑战，通过全组的努力得到了解决，增加了全组成员在团队协作方面的经验。在通过课程本身与课程大作业的学习经历之后，对可视化有了更深入的了解，开拓了小组成员的视野，在所有成员都并非可视分析相关方向的情况下，提供了一种新的思路去思考自身专业方向中的数据处理业务。而可视分析同样也是技术要求范围比较广的技术门类，开发过程中对于数据库、html、json、javascript、python等技术的应用较多，一定程度了上锻炼了小组中负责开发的人员的技术能力。

本课程强调在文档上的规范性，也因为协作的原因，整个项目自身也使用了很多文档来协助每个人的分管的工作，这种对文档的规范性要求也渗透到了平日的工程开发中，使得最后积累了足够数量的文档来协助全组完成最后的项目收尾工作，对于全组成员以后参与实验室项目或者工作有了一定程度的训练，这也是这次课程学习的额外收获。