# 项目技术说明书

豆瓣电影数据可视化

小组名称： 第5组

学员姓名： 袁浩

目录

[项目技术说明书 1](#_Toc17289747)

[1. 项目介绍 2](#_Toc17289748)

[2. 需求分析 2](#_Toc17289749)

[3. 技术选型 3](#_Toc17289750)

[4. 爬虫开发和优化 3](#_Toc17289751)

[5. 数据清洗和存储 4](#_Toc17289752)

[6. 数据查询和可视化展示 5](#_Toc17289753)

[7. 开发环境 5](#_Toc17289754)

[8. 产品原型 6](#_Toc17289755)

# 项目介绍

该项目主要用于对豆瓣网电影信息分类做可视化的数据展示

# 需求分析

需要将每一年份电影数据存储为一个json文件，再存入mongodb数据库，进行数据筛选匹配

# 技术选型

负责可视化界面的设计，以及后台数据存取的模块

Node.js：前端应用开发方法

Cheerio：实现HTML数据解析

eCharts：数据可视化展示

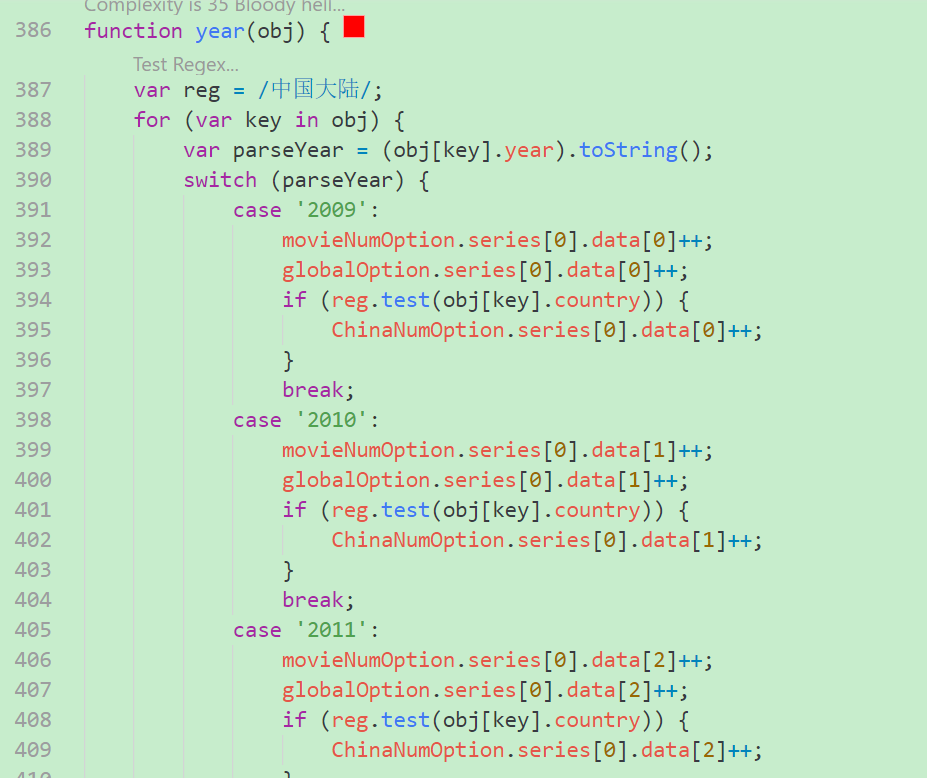
superagent：客户端请求代理模块（支持HTTPS）

Mongoose: 连接数据库

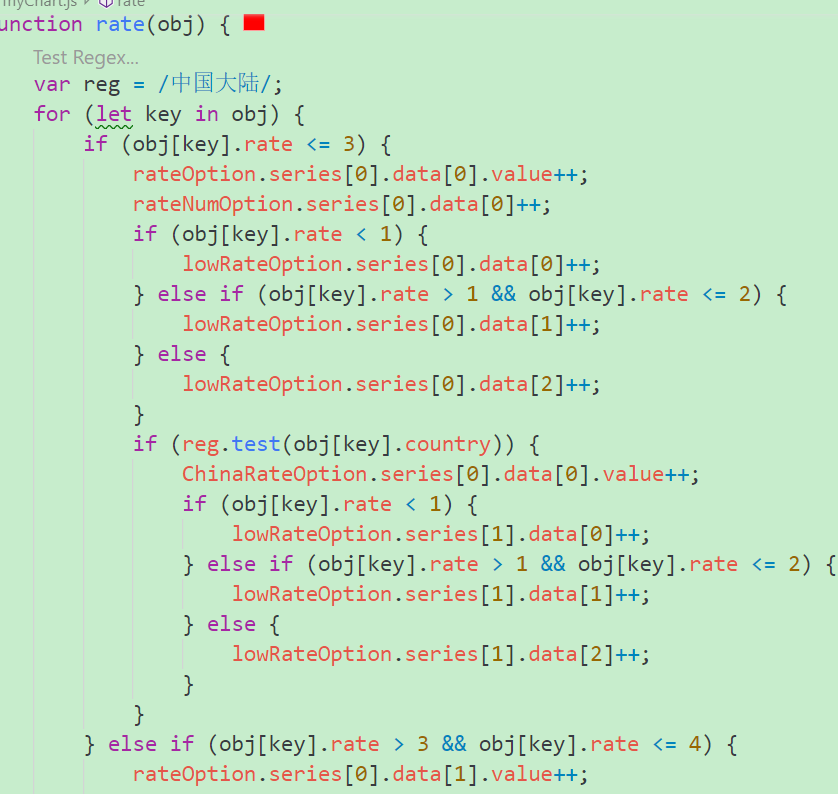
# 爬虫开发和优化

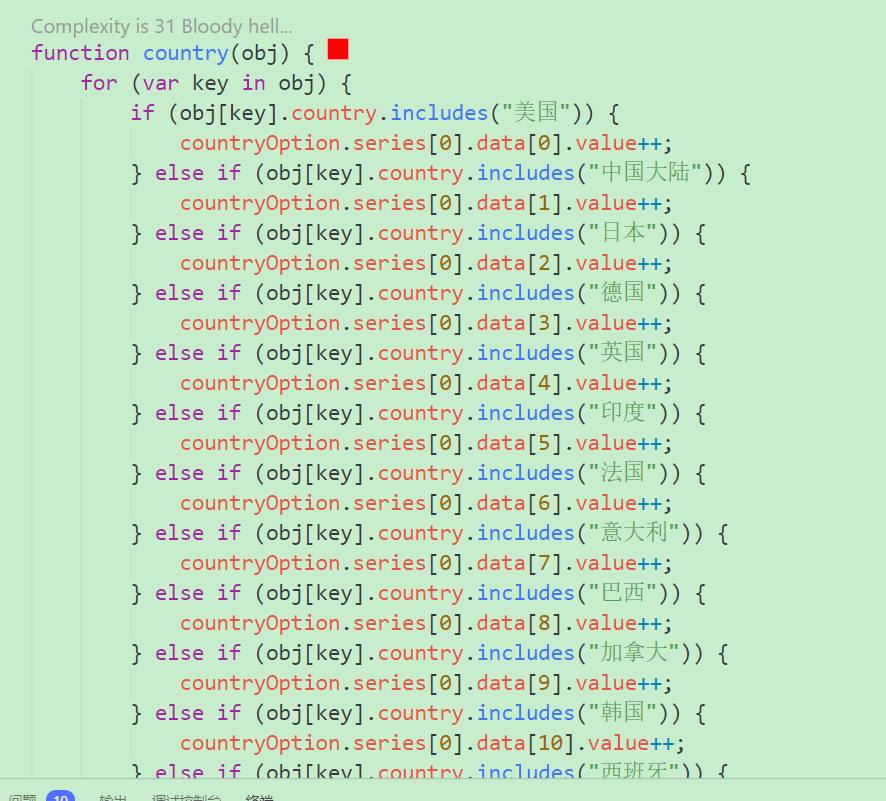
引入express,cheerio,superAgent模块，创建模块实例，使用superAgent爬取原始数据，通过cheerio（data.text）获得cheerio对象，从而方便对原文件的DOM对象定位

# 数据清洗和存储

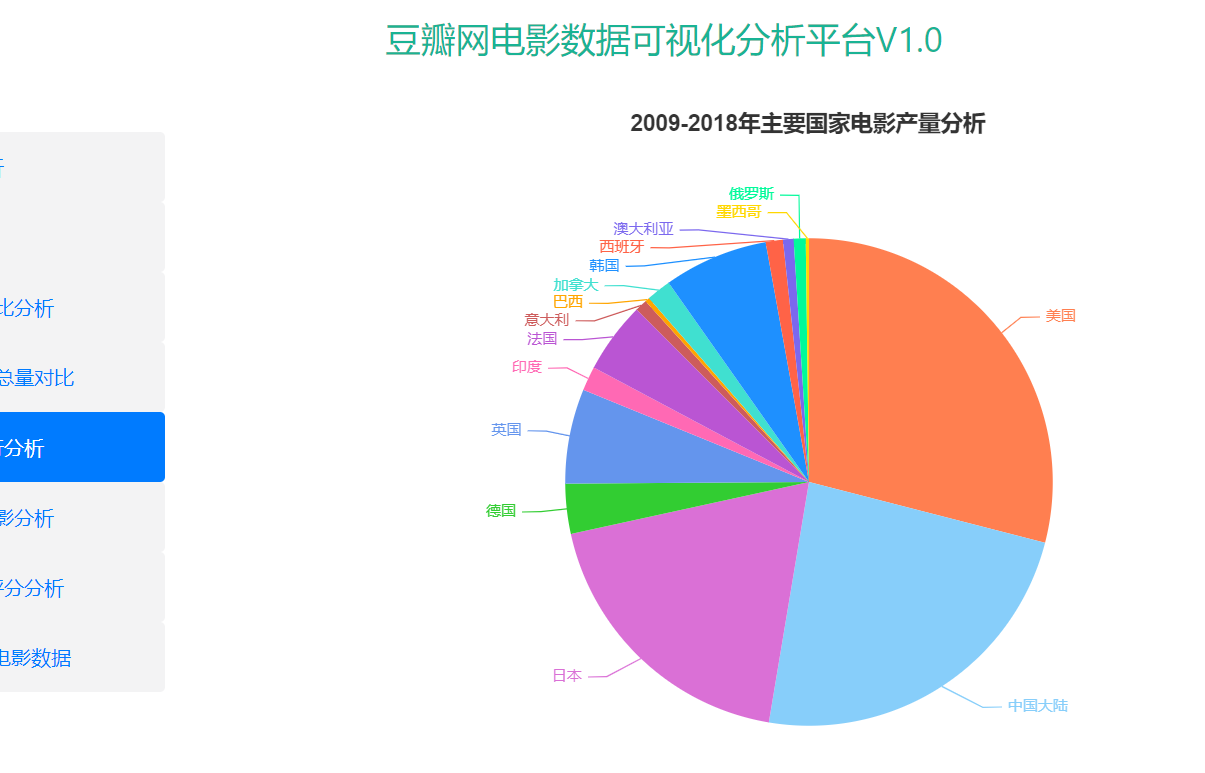


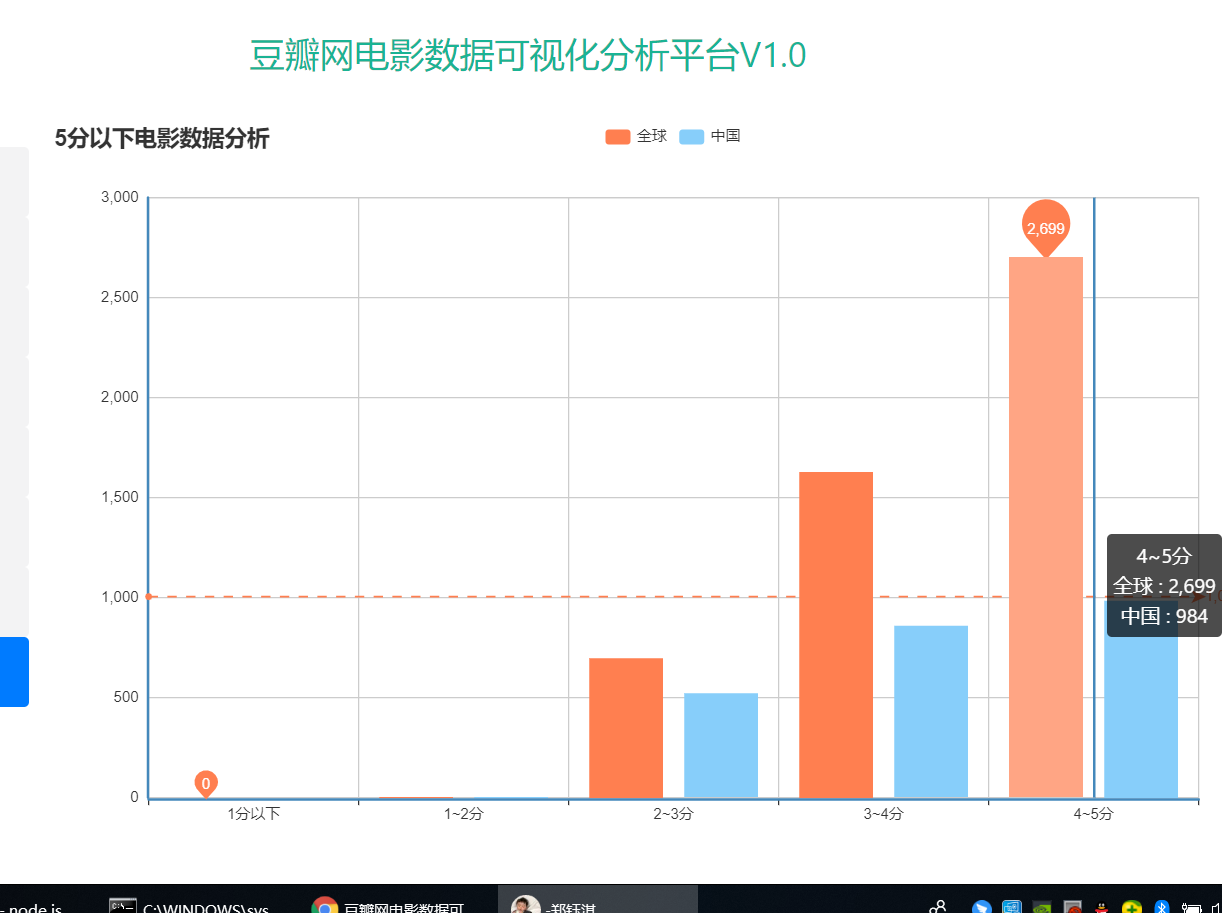






# 数据查询和可视化展示





# 开发环境

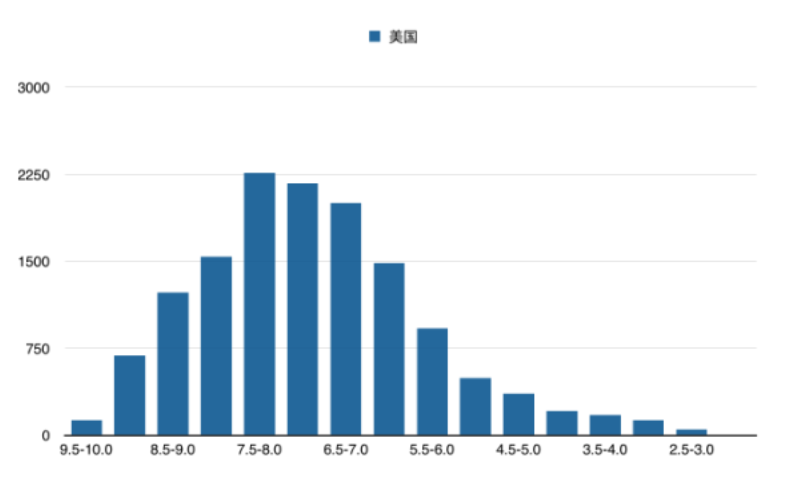
Chrome vsCode

# 产品原型

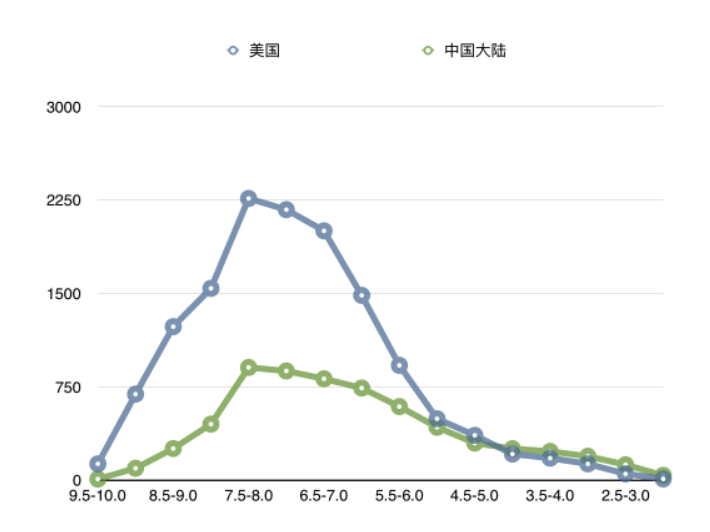
(主要说明自己所负责的的模块最终的呈现效果)

|  |  |
| --- | --- |
| 豆瓣电影数据可视化分析平台V1.0 | |
| 影评分和电影数量分析 | echarts展示 |
| 影评分和电影数量对比分析 |
| 近十年每年发行的电影分析 |
| 2019各个国家发行的电影数量分析 |
| 近十年中国发行的电影评分 |
| 近十年中国发行的电影每一年的数量分布图 |
| 2019年评分在5分以下的电影名称 |
| 自己自由发挥 |

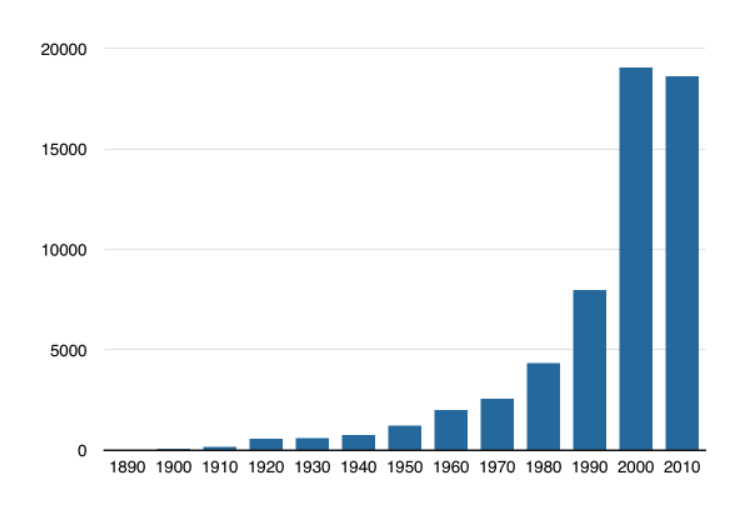
1. 根据国家生成电影评分和电影数量的柱状图



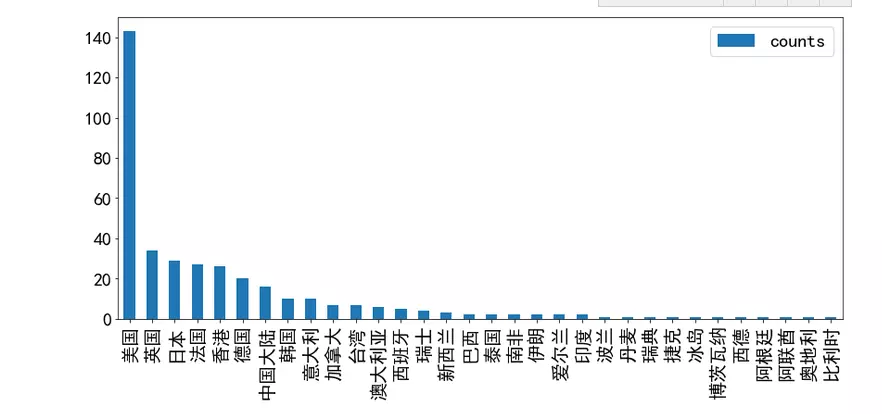
1. 对比任意两个国家电影评分和电影数量的折线图



1. 近十年每年发行的电影数量分布柱状图



1. 2019年截至到今天为止各个国家发行的电影数量分布图



1. 近十年中国发行的电影评分(0~3、3~5、5~7、7~10)的范围的数量分布图
2. 近十年中国发行的电影每一年的数量分布图
3. 2019年截至到今天为止评分在5分以下的电影名称
4. 自己自由发挥…(加分项)