

# InfoVis Individual Task [A1] - Doc

3130000073 杨加锐

---

## I. 系统概述

本系统被实现为一个主要由三个数据布局构成、存在多关联视图、且支持多种交互模式的网页端信息可视化系统，其中：

### • 源数据

基于 *Epinion* 在2000前作为电影评价网站的数据集，主要包括3952条movies、6040条users和1000209条ratings，在项目的 `data/` 文件夹下可找到；数据格式为：

- **movies** -> MovieID::Title::Genres(genre1|genre2|...)
- **users** -> UserID::Gender::Age::Occupation::Zip-code
- **ratings** -> UserID::MovieID::Rate::Timestamp

### • 布局

- 参考自 [Stacked Bars](#)，横轴主坐标表示数据集中的18个电影流派（Genre），细分的副坐标颜色从浅到深分别表示不同评价等级（1-2差评，3中评，4-5好评）；而纵坐标表示特定的流派、评分组合分类中，给出评价的用户数量。
- 参考自 [Animated Donut Chart with Labels](#)：表达特定的流派、评分组合分类中，给出评价的用户的年龄段构成的统计信息。
- 参考自 [Cluster Purity Visualizer](#)：表达每个年龄段的人（共七个pie chart）喜好（给出好评）的电影年代的比例统计信息，从50年代之前到90年代，以十年为区间划分；饼图可以用扇形的半径表示比例，也可以用中心角角度表示比例。

### • 交互模式

- Stacked Bars支持使用右上角的Radio Button在堆栈式和分组式之间切换同一组数据（基于同一Movie Genre）的显示效果，并伴有平滑的动画效果；
- 光标划过Stacked Bars中的任何一矩形时，上方label都会显示对应的Genre和Rate-class；点击时，右侧的Donut Chart会更改数据源为对应的流派、评分组合，并伴有平滑的过渡效果；
- 点击Title跳转到 [Loved Movie Eras Of Each Age Periods](#)，可以通过Radio Button切换显示方式（用扇形的半径、或用中心角角度表示比例）；也可以通过点击右侧的Movie Era列表前面的圆圈，选中该颜色对应的Era为显示或不显示，更方便对比两个数量相近的扇形量级。

## II. 运行环境及库依赖关系

### • 运行环境

- 在最新稳定版本（截至2015.12.28）的Firefox浏览器下测试通过，浏览器必须支持运行 `Javascript` ；
- 如果已经安装了Python，在终端下进入本项目主目录并键入命令 `python -m SimpleHTTPServer` 即可在浏览器中使用 `http://localhost:8000/` 体验效果；
- 若默认端口被本地其他程序占用，则需要在命令中额外指定一个空闲端口： `python -m SimpleHTTPServer [port-number]` ；
- 若未设置Python在系统中的环境变量，则需要用python程序的全路径来代替 `"python"` ；
- 在Mac OS X下，也可以使用前端开发工具Coda 2直接打开 `index.html` 预览实时效果。

### • 引用的外部库

- 以 `d3.js` 作为支持数据可视化效果呈现的主要Javascript库，即项目目录中的 `lib/d3.min.js` ；
- 引用了 `lib/clusterpurityChart.js` 来实现多个饼图的效果；
- 引用了 `Bootstrap 3` 的外部库来构建大方简洁的UI效果，即 `lib/bootstrap.min.js` 和 `lib/bootstrap.min.css` ；
- 预处理数据格式及获取统计数据时引用了Python的 `codecs` 库以便读取 `latin-1` 编码的部分源数据。

## III. 项目文件结构和详解

•

### • `jiarui_vis.html`

---

主要的数据展示和交互接口，包含本人撰写的大部分Javascript脚本，即Stacked Bar Chart和Donut Chart部分；

### • `jiarui_cluster_chart.html`

---

为辅的数据展示和交互页面，包含部分Javascript脚本，即Cluster Purity Chart；

## • ***data/***

---

包含各数据视图需要加载的序列化数据 `*.json` 和源数据：

- **stacked\_bar\_jiarui.json**

由Stacked Bars视图加载的支持数据；

- **donut\_jiarui.json**

由Donut Chart视图加载的支持数据；

- **cluster\_jiarui.json**

由Cluster Purity Chart视图加载的支持数据；

- **movies.txt & ratings.txt & users.txt**

Epinion数据集的源数据；

- **README**

Epinion数据集的数据格式和分类说明。

## • **data\_cleaning.py**

---

用于源数据清洗和读取、构建三个支持视图的序列化json数据的脚本。

- `##lib/` 包含下载到本地的部分外部库。

## IV. 关键算法、代码和技术难点

- 为了保证多个数据视图所绑定的函数都能够按正确的顺序获取到支持显示所需的序列化数据对象，解决Javascript的异步加载问题所带来的困惑，采取了嵌套 `d3.json()` 函数的写法：

```
d3.json("data/chart1.json",function(error,root){
    //...
    d3.json("data/chart2.json",function(error,graph)
        //...
    });
});
```

- 用 `d3.json()` 一次性加载任何一个源数据集都会导致得到“undefined”，通过采用统计数据、缩小数据规模才能够得到数据后正确显示。
- 在Cluster Purity Chart的构建过程中，发现d3无法读取 `' '` 包裹关键字的Json，必须要用双引号包裹。

## V. 参考资料

### 1. D3JS官方文档：

<https://github.com/mbostock/d3/wiki>

### 2. D3JS官方示例代码库：

<https://github.com/mbostock/d3/wiki/Gallery>

### 3. W3CSchool Javascript教程：

<http://www.w3school.com.cn/js/index.asp>

### 4. 数据可视化专题站：

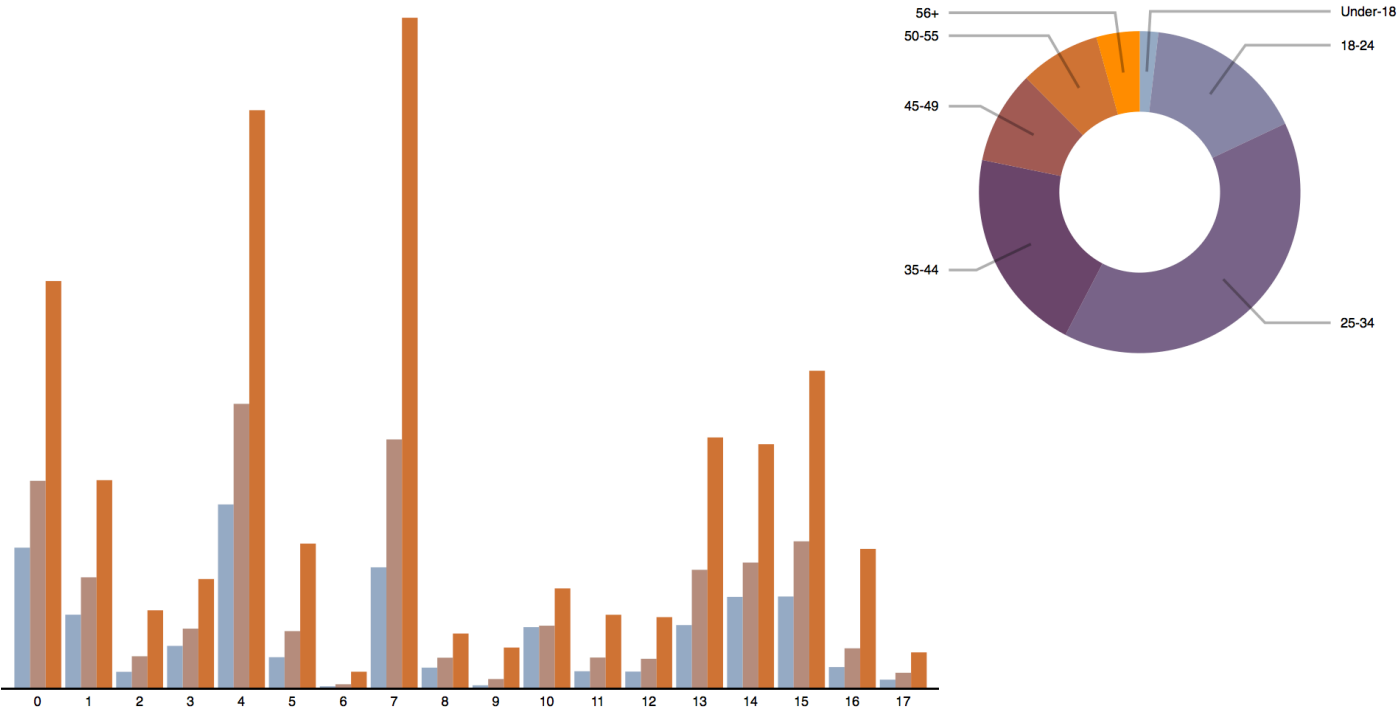
<http://www.ourd3js.com/wordpress/?cat=2>

## VI. 系统运行截图

### 整体效果

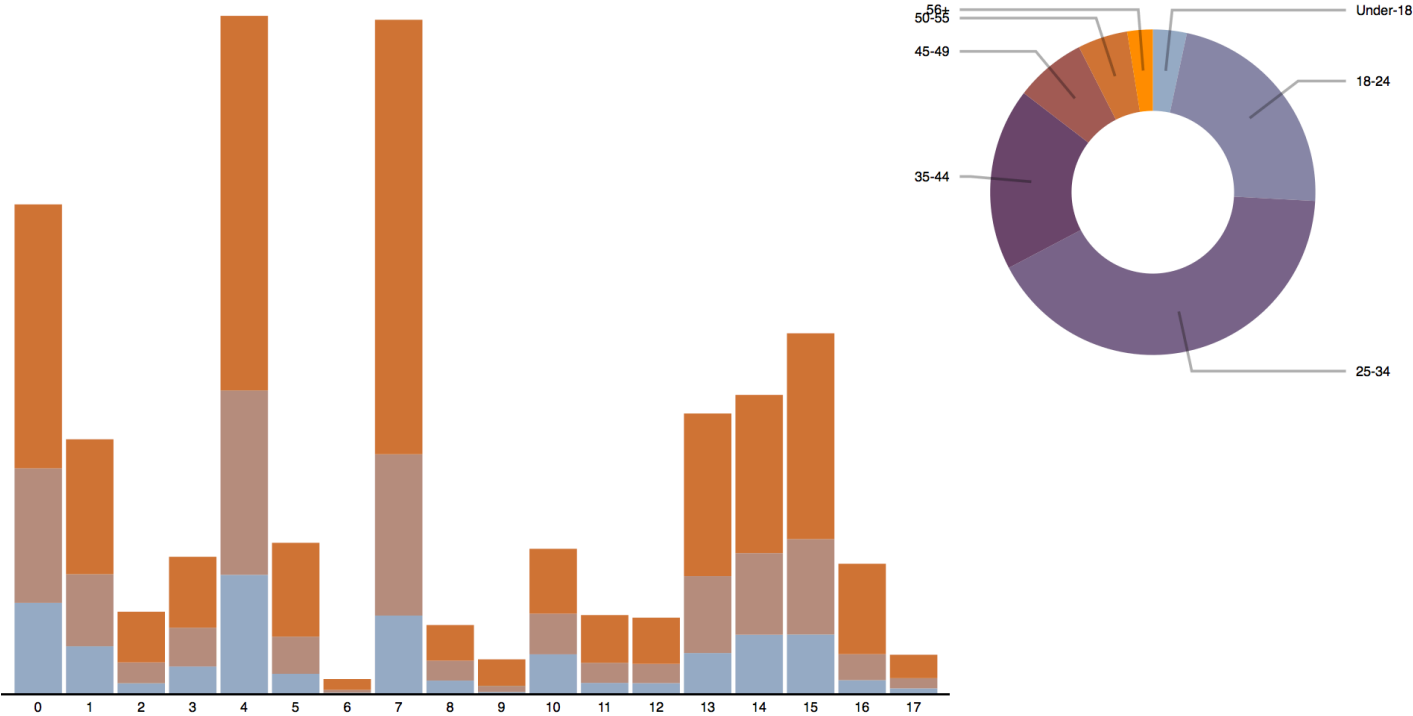
---

Drama like-it



切换到堆叠视图并选择不同矩形后，两图相应地同步改变

Adventure dislike

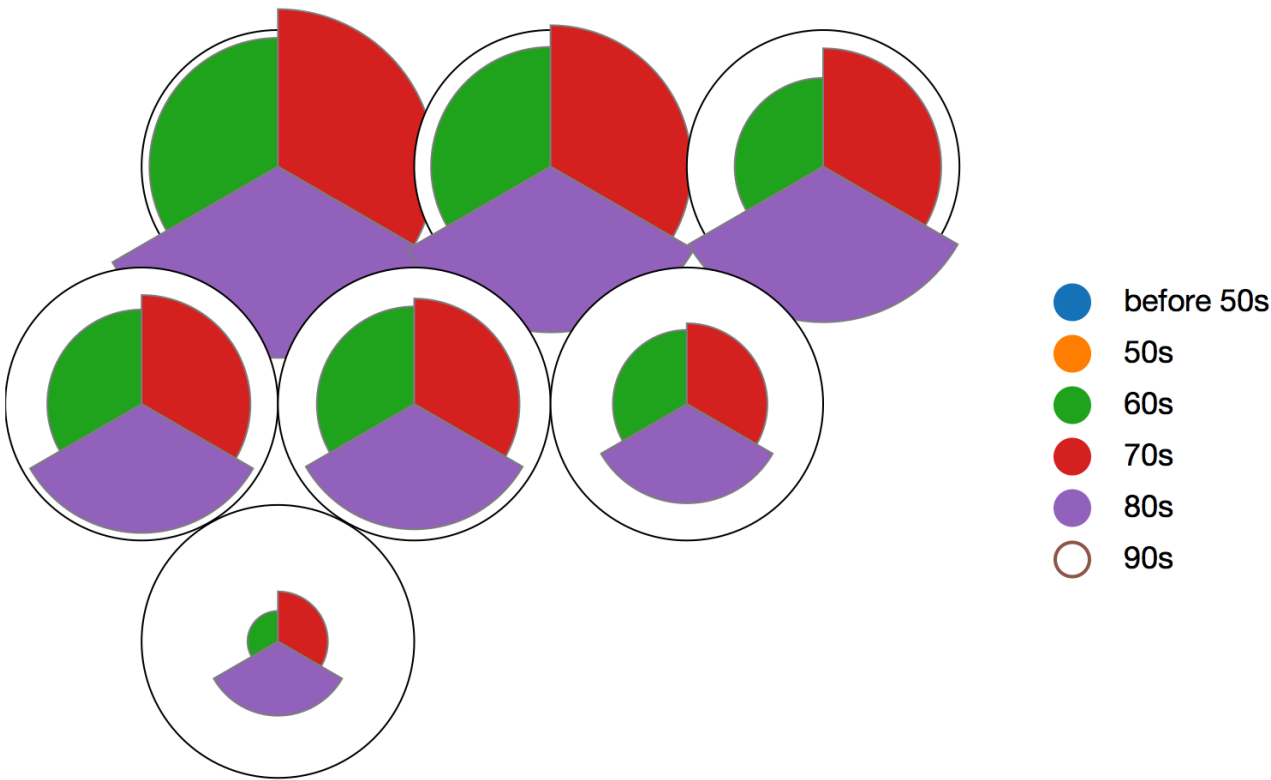


用半径表示的状态下第三视图的效果

# Loved Movie Eras Of Each Age Periods by Jiarui Yang

Length representation

Angle representation



用角度表示的状态下取消两个age-class后的效果

# Loved Movie Eras Of Each Age Periods by Jiarui Yang

Length representation

Angle representation

