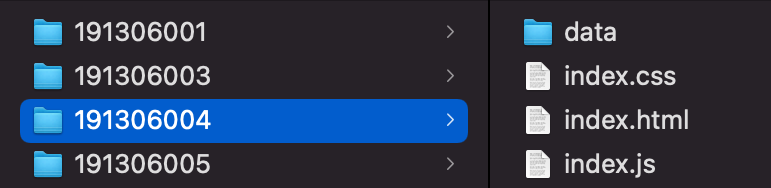
《数据可视化技术》课程报告

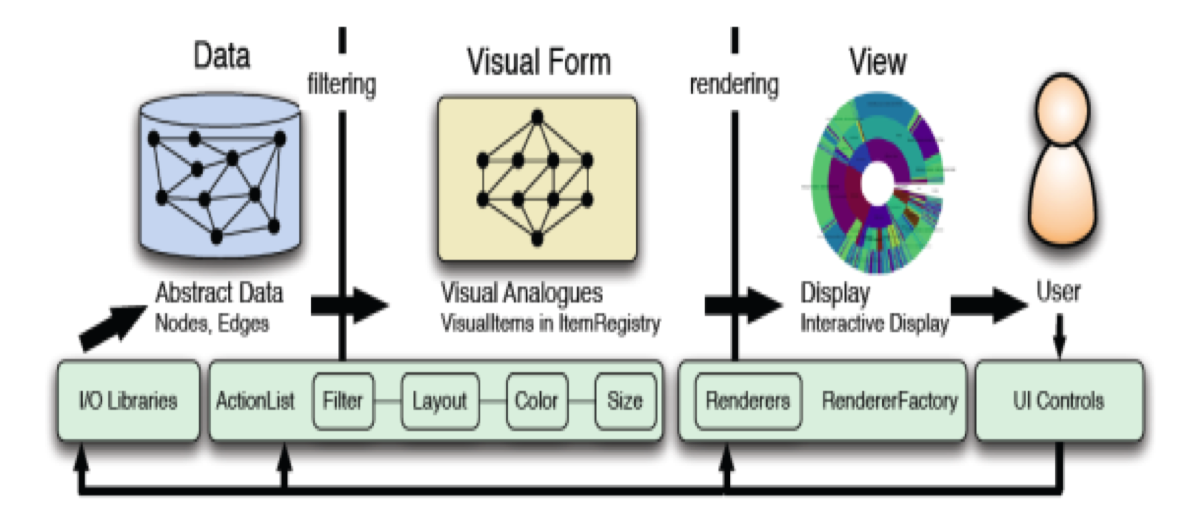
|  |  |  |  |
| --- | --- | --- | --- |
| 数 计 学院 | 软件工程 专 业 | | 21 级 班 |
| 学 生 姓 名 | 蒋天宇 | 学 号 | 211308121 |
| 指 导 教 师 | 王诗莹 | 职 称 | 硕 士 |
| 完 成 日 期 | 2024 年06月25日 | | |

**教务处 制**

**《数据可视化技术》课程设计实施与评价规则**

1. 本次课程设计由每人独立完成设计，任课老师将考量设计与他人或线上作品的相似度，**涉嫌抄袭的作品将以0分处理**，考量标准任课教师保留解释权。
2. 本次课程设计必须使用 D3.js 框架作为项目的技术手段，**不采用D3.js 框架的作品将以0分处理**，**如有特殊情况请向任课教师申请，**任课教师保留最终解释权。
3. 本次课程设计建议使用WebStorm作为项目的编辑器。
4. 本次课程设计包括9个课程实验报告，完成时间为1周。请在教室中完成资料查找与编码，务必按时上交单次报告，**逾期不接受补交**。
5. 本次课程设计最终提交内容包括如下清单：  
   1. 项目作品打包（**源代码**）命名规则：学号.rar 如 201122001.rar  
   2. 本实验报告（完整填写）命名规则：学号.docx如 201122001. docx
6. 本次课程设计项目（**源代码**）必须包含4个文件：index.html、index.js、index.css（任课教师提供模版）、所用data文件、其他相关项目文件，项目结构可以参考如下：  
   
7. 本次课程设计的**评分比例**如下：  
   1. 项目评分 50 %  
   2. 实验报告评分 40 %  
   3. 课堂表现 10 %  
   **以上评分若有一项为0分，课程设计总分将直接划为0分。**
8. 本实验报告的模板原内容仅作为提醒，并非硬性要求，根据自身项目完成即可。报告内容需逻辑清晰、注意格式、作品正确且美观。
9. 本次课程设计最终评价阶段将通过课程答辩方式，个人**制作PPT**进行演讲答辩、展示项目，限时**3分钟**以内。务必参加答辩，评判标准任课教师保留解释权。

**数据可视化流程：**

  
**数据可视化设计规范：**

[数据可视化设计规范指南](https://m2.material.io/design/communication/data-visualization.html)

课程实验报告 1 ： 数据采集

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 | 蒋天宇 | 学 号 | 211308121 |
| 1. 数据来源：Kaggle   [world-economic-data.csv](https://www.kaggle.com/datasets/mittvin/world-economic-indicators-1960-2022-dataset)  [world-terrorist-data.csv](https://www.kaggle.com/datasets/rafsunahmad/terrorist-attacks-data-since-1970-2023)  [world-unemployment-data.csv](https://www.kaggle.com/datasets/sazidthe1/global-unemployment-data)  [world-population-data.csv](https://www.kaggle.com/datasets/sazidthe1/world-population-data)   1. 原始数据读取代码截图：      1. 原始数据各项属性分析：   **world\_economic\_data.csv**   |  |  |  |  | | --- | --- | --- | --- | | 属性名 | 字段类型 | 取值范围 | 是否使用 | | Country Name | string | 国家英文名称 | 是 | | Country Code | string | 唯一的三个字母 | 是 | | Year | int | 1960-2022 | 是 | | Personal remittances, received (% of GDP) | double | 0-235.9279913 | 是 | | Unemployment, total (% of total labor force) | double | 0.095-38.8 | 是 | | GDP (current US$)\_x | double | 8824746.238-100562000000000 | 是 | | GDP growth (annual %)\_x | double | -64.04710697-149.9729635 | 是 | | GDP (current US$)\_y | double | 8824746.238-100562000000000 | 否 | | GDP growth (annual %)\_y | double | -64.04710697-149.9729635 | 否 |   **world\_unemployment\_data.csv**   |  |  |  |  | | --- | --- | --- | --- | | 属性名 | 字段类型 | 取值范围 | 是否使用 | | country\_name | string | 国家英文名称 | 是 | | indicator\_name | string |  | 是 | | sex | string | Female / Male | 是 | | age\_group | string |  | 是 | | age\_categories | string | Youth / Adults / Children | 是 | | 2014 | double | 0.027-74.485 | 是 | | 2015 | double | 0.034-74.655 | 是 | | 2016 | double | 0.038-74.72 | 是 | | 2017 | double | 0.035-75.416 | 是 | | 2018 | double | 0.044-76.395 | 是 | | 2019 | double | 0.036-77.173 | 是 | | 2020 | double | 0.056-83.99 | 是 | | 2021 | double | 0.064-82.135 | 是 | | 2022 | double | 0.067-78.776 | 是 | | 2023 | double | 0.063-78.541 | 是 | | 2024 | double | 0.06-78.644 | 是 |   **world\_terrorist-data.csv**   |  |  |  |  | | --- | --- | --- | --- | | 属性名 | 字段类型 | 取值范围 | 是否使用 | | Entity | string | 国家英文名称 | 是 | | Code | string | 唯一的三个字母 | 否 | | Year | int | 1970-2021 | 是 | | Terrorist attacks | int | 0-16820 | 是 | | Terrorism deaths | int | 0-44576 | 是 | | Attack method: Hijacking | int | 0-58 | 是 | | Attack method: Hostage Taking (Barricade Incident) | int | 0-85 | 是 | | Attack method: Unarmed Assault | int | 0-106 | 是 | | Attack method: Facility/Infrastructure Attack | int | 0-776 | 是 | | Attack method: Hostage Taking (Kidnapping) | int | 0-1409 | 是 | | Attack method: Assassination | int | 0-1007 | 是 | | Attack method: Armed Assault | int | 0-4037 | 是 | | Attack method: Bombing/Explosion | int | 0-8739 | 是 | | Death Age 100+ | int | 0-29 | 是 | | Death Age: 51-99 | int | 0-52 | 是 | | Death Age : 21-50 | int | 0-275 | 是 | | Death Age : 11-20 | int | 0-522 | 是 | | Death Age : 6-10 | int | 0-853 | 是 | | Death Age : 1-5 | int | 0-6612 | 是 | | Terrorist Death Type : Suicide | int | 0-10337 | 是 | | Terrorist Death Type : Killed | int | 0-38627 | 是 |   **world\_population\_data.csv**   |  |  |  |  | | --- | --- | --- | --- | | 属性名 | 字段类型 | 取值范围 | 是否使用 | | rank | int | 1-234 | 否 | | cca3 | string | 唯一的三个字母 | 否 | | country | string | 国家英文名称 | 是 | | continent | string | 大洲英文名称 | 否 | | 2023 population | int | 518-1428627663 | 是 | | 2022 population | int | 510-1425887337 | 是 | | 2020 population | int | 520-1424929781 | 是 | | 2015 population | int | 564-1393715448 | 是 | | 2010 population | int | 596-1348191368 | 是 | | 2000 population | int | 651-1264099069 | 是 | | 1990 population | int | 700-1153704252 | 是 | | 1980 population | int | 733-982372466 | 是 | | 1970 population | int | 752-822534450 | 是 | | area (km虏) | double | 0.44-17098242 | 是 | | density (km虏) | int | 0-21403 | 是 | | growth rate | double | -7.45%-4.98% | 是 | | world percentage | double | 0.00%-17.85% | 是 |  1. 可视化数据筛选相关代码截图： | | | |

课程实验报告 2 ： 可视化方案构建

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 1. 设计草图（多个）： 2. 设计方案描述： | | | |

课程实验报告 3 ： 数据清理

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 前期的数据采集得到的数据，不可避免的含有噪声和误差，数据质量较低数据清洗和规范是必不可少的步骤。首先把脏数据、敏感数据过滤掉，其次再剔除和目标无关的冗余数据，最后调整数据结构到系统能接受的方式。  常见的数据质量问题包括：   1. 数据收集错误，遗漏了数据对象，或者包含了本不应包含的其他数据对象。 2. 数据中的离群点，即不同于数据集中其他大部分数据对象特征的数据对象。 3. 存在遗漏值，数据对象的一个或多个属性值缺失，导致数据收集不全。 4. 数据不一致，收集到的数据明显不合常理，或者多个属性值之间互相矛盾。例如，体重是负数，或者所填的邮政编码和城市之间并没有对应关系。 5. 重复值的存在，数据集中包含完全重复或几乎重复的数据。   数据清理过程及方法描述：（文字+代码截图+图例） | | | |

课程实验报告 4 ： 数据过滤和筛选

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 可视化结果中，不可能把所有的数据统统展示出来，因此会涉及到包括标准化（归一化）、采样、离散化、降维、聚类等数据处理的方法，本次课程设计D3工具中涉及到方法如： filter，map，sort等。  数据过滤和筛选方法/工具类描述：（文字+代码截图+图例） | | | |

课程实验报告 5 ： 视觉编码

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 视觉编码的设计是指如何使用位置、尺寸、灰度值、纹理、色彩、方向、形状等视觉通道，以映射我们要展示的每个数据维度。背景知识 – 视觉编码：<https://zhuanlan.zhihu.com/p/24835448> 视觉编码总体描述：（文字+代码截图） | | | |

课程实验报告 6 ： 图例&标签

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 简介：在简单图表中，可以使用直接标签。在密集的图表（或更大的图表组的一部分）中，可以用图例。    项目中使用图例&标签场景描述：（文字+实现截图） | | | |

课程实验报告 7 ： 布局

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 数据可视化中，好的布局方式，往往能带来舒服的视觉效果。项目中，通过改变可视化部件的位置、大小、角度来获得我们需要的效果  背景知识：<https://blog.csdn.net/github_36904248/article/details/73342373>  布局的设计描述：（文字+实现截图） | | | |

课程实验报告 8 ： 交互

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 常见的交互方式包括：  1.滚动和缩放  2.颜色映射的控制  3.数据映射方式的控制：这个是指用户对数据可视化映射元素的选择，一般一个数据集，是具有多组特征的，方便用户按照自己感兴趣的维度去探索数据背后的信息。  4.数据细节层次控制：比如隐藏数据细节，hover或点击才出现。  背景知识：<https://zhuanlan.zhihu.com/p/49528045> <http://www.qiutianaimeili.com/html/page/2020/03/8rqxzs4fnl.html>  <https://www.jianshu.com/p/c49dfb38fc0d>  项目中使用交互场景描述：（文字+代码截图+实现截图） | | | |

课程实验报告 9 ： 项目评价和改进

|  |  |  |  |
| --- | --- | --- | --- |
| 姓 名 |  | 学 号 |  |
| 项目最终效果呈现：（文字+实现截图）项目优化描述： | | | |