



浙江大学
ZHEJIANG UNIVERSITY

2024 ZJU Summer School on Visualization and Visual Analytics



Course Project

Jianing Yin

Ph.D student

State Key Lab of CAD&CG, Zhejiang University

Email: yinjianing@zju.edu.cn

Course Assessment

- Project 1 (20%)
 - Individual work: Learn to use visual analytics tools (Tableau, ...).
 - Everyone should submit an experiment report.
 - **DDL: 7.28**
- Project 2 (50%)
 - Teamwork: Data analysis with visual analytics tools (Python, Tableau, ...) or a visual analytics system designed and developed by yourself (可视分析系统).
 - 3 members
 - Free topic (You can ask TAs for help)
 - An Analysis Report: Data preprocessing, Visualizations, Insights
- Project 2 presentation (20%)
 - 7.26 8:50-12:00 (Submit your presentation PPT before **7.25 23:59**)
 - **5 minutes/group**
 - All students should attend the project presentation.
- Attendance (10%)
- Bonus (5 points): Design and develop your visual analytics system.

Course Assessment

- 课程项目1 (20%)
 - 个人作业：学习可视化分析工具的使用 (Tableau, ...) .
 - 每人提交一份实验报告
 - DDL: 7.28
- 课程项目2 (50%)
 - 小组作业：使用可视分析工具 (Python, Tableau, ...) 做数据分析，或者自己设计并开发一个可视分析系统
 - 每组三人
 - 自由选题（可以向助教寻求帮助）
 - 数据预处理、可视化、数据洞察解读
- 课程项目2展示 (20%)
 - 7.26 8:50-12:00 (7.25 23:59之前提交展示PPT)
 - 每组5分钟，全体同学都需要到场参与答辩
- 出勤 (10%)
- Bonus (5 points): 自己设计并开发一个可视分析系统

Project 1

- Data Source
 - [2021 Global Gas Flaring Volumes](#)
- Data Description
 - https://eogdata.mines.edu/download_global_flare.html
- Tasks
 1. 请使用[Geo Heatmap](#)可视化出不同国家的气体燃烧总量分布图
 - 每个国家具体总量是多少也要在图上显示出来
 2. 对于气体燃烧总量最高的国家，请使用[Pie Chart](#)可视化不同排放类型的气体燃烧量占该国气体燃烧总量的百分比
 - 将不同类型的气体按燃烧量占比排序
 - 占比较大的气体+其百分比数字都要显示出来
 3. 设置视图联动，使得当用户点击[Pie Chart](#)中的任意扇区时，[Geo Heatmap](#)能够实时更新并突出显示该国家的气体燃烧总量数据

Project 1

- Requirement
 - Deadline: submit an experiment report before 7.28 (学在浙大)
 - 实验报告没有严格的格式要求
 - 数据清洗、制作可视化、制作仪表板的过程和结果**描述清楚，截图清楚**
 - 用到了哪些操作都需要在报告中**写明**。比如使用了什么具体的筛选器，如果做了数据清洗或者排序，也要说明等等
 - **至少需要这些截图：**
 - 数据清洗的操作配置+前后的对比图（如有）
 - Task1的界面操作配置+可视化效果图
 - Task2的界面操作配置+可视化效果图
 - Task3的界面操作配置+可视化效果图
 - 仪表板的界面操作配置+可视化效果图
 - 仪表板的可视化要**能够回答**三个tasks的问题

Project 2

- 课程项目1 (20%)
 - 个人作业：学习可视化分析工具的使用 (Tableau, ...) .
 - 每人提交一份实验报告
 - DDL: 7.28
- 课程项目2 (50%)
 - 小组作业：使用可视分析工具 (Python, Tableau, ...) 做数据分析，或者自己设计并开发一个可视分析系统
 - 每组三人
 - 自由选题（可以向助教寻求帮助）
 - 数据预处理、可视化、数据洞察解读
- 课程项目2展示 (20%)
 - 7.26 8:50-12:00 (7.25 23:59之前提交展示PPT)
 - 每组**5分钟**，小组全体成员都需要到场参与答辩
- 出勤 (10%)
- Bonus (5 points): 自己设计并开发一个可视分析系统

Project 2

- 无论是使用可视分析工具进行数据分析，还是自己设计并开发可视分析系统，**都要包括**数据预处理、可视化、数据洞察解读
- 使用可视分析工具（Python, Tableau...）进行数据分析：要求能够传达深刻的数据内涵，讲述完整的数据故事。
- 设计并开发可视分析系统：要求支持多视图（ ≥ 2 ）的可视化呈现和探索，具有跨视图的交互，能够展示完整的案例。
 - 使用可视分析工具（Python, Tableau...）进行数据预处理，设计并开发可视分析系统进行可视化，展示数据洞察
- 参考：
 - ChinaVis 2024 数据可视化竞赛
 - VAST Challenge 2024 (vast-challenge.github.io)

关于与作业相关的技术课程

- [浙江大学2024年可视化与可视分析暑期学校 \(zju.edu.cn\)](https://zju.edu.cn)
- 7.15 可视化工具的使用（可视化工具 + 数据预处理）
- 7.16 - 8.19 可视分析系统代码开发
 - 7.16 Web开发基础（HTML/CSS/JS）
 - 7.17 Web开发基础（JavaScript）
 - 7.18 数据可视化开源框架（React/Vue）
 - 7.19 可视分析系统的框架及搭建教学（会提供现成的代码框架）

关于数据

- 数据集不做限定，可以到以下网站寻找：
 - Kaggle:
 - [Find Open Datasets and Machine Learning Projects | Kaggle](#)
 - Awesome Public Datasets:
 - <https://github.com/awesomedata/awesome-public-datasets>
- 可以寻求助教帮助（每组有分配助教）
- 选好数据后要**和助教确认**

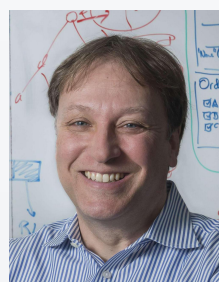
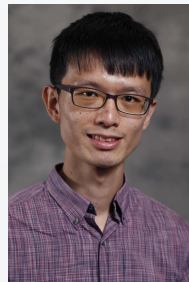
关于展示

- 线下展示 (7.26 8:50-12:00), 小组全体成员到场
- 7.25 23:59之前提交PPT
- 展示数据背景、数据预处理、可视化和数据洞察解读 (insight)
 - 如果是可视分析系统, 还需要录制系统运行的demo, 展示所有的功能
 - 如果是使用可视分析工具做数据分析, 请展示尽可能深刻的insight
- 展示小组分工, **务必**每个人都有编程/使用可视化工具的部分
 - 不要某个人只负责答辩
- 关于数据洞察解读 (insight)
 - 在哪里发现了什么规律?
 - 为什么会出现这个规律?
- 每组展示限时**5分钟**, 展示后有Q&A

Bonus Choice for Project 2

Background

- Visual Comparisons of AIs' Algorithmic Output [S. Jiang, M. Ling, H. Chao, H. Pfister, D. Haehn, J. Chen 2024]
 - The dataset comes from a data collection experiment, through a collaborative project between The Ohio State U. and Harvard U.
 - The dataset is large containing 1600 files, representing the experimental results of 1600 models trained in different conditions.
- The scholars need your help to construct a comparative visualization tool to explore these 1600 experimental results.



Bonus Choice for Project 2

- Dataset
 - 在**不同的训练条件**下训练出来的CNN，他们在完成可视化任务 (估计ratio) 时的**输出结果数据**。
 - The dataset is large containing 1600 files, representing the experimental results of 1600 models trained in different conditions.
- Design requirement: Build Visual Comparison Interface
 - 研究训练集准备上的哪些因素会影响CNN在可视化任务 (估计ratio) 上的性能。
 - Show visual details to tell us the relationship between training input (sampling values) and model output.
 - Show relationships between downsampling methods in terms of model outputs.
 - Compare between VGG19 and ResNet50 for inference accuracy.
 - Simple and elegant - but an aggregated view using one average number is not enough.
- 关于这个bonus的详细介绍请参见: [bonus详细介绍文档](#)
- **如果**你在这个数据集上针对需求设计开发了一个可视分析系统，并针对需求获得了一些有价值的insight, **那么**可以再获得**5分的加分** (计入Project 2成绩)

Timeline

- 7.15-7.16 介绍作业，确定分组
- 7.17-7.18 分配助教，确定数据
- 7.26 课程项目2展示

关于答疑

- 7.15-7.19 下午有助教开展的技术课程，可以提问
- 7.20-7.21 周末
- 7.22-7.25的每天上午9:45-12:00也都有助教在教室答疑
- 每组都有分配相应的专职助教

Related Tools

- 编辑器
 - Visual Studio Code
 - 轻量编辑器
 - 插件式编辑器
 - 适用于各种语言的代码编辑
 - WebStorm
 - 重量IDE
 - 集成对应语言的各类插件

Related Tools

- 浏览器
 - Chrome / Edge (Chromium内核)
 - 开发者工具 (F12)

Related Tools

- 环境配置
 - Node.js
 - npm包管理
 - Python 3.x
 - 如果使用Python而不是Tableau进行数据处理

Related Tools

- 代码协作
 - Git
 - 分布式版本控制系统
 - Git教程: [Git教程 - 廖雪峰的官方网站 \(liaoxuefeng.com\)](http://liaoxuefeng.com)
 - Github 一个远程代码仓库
 - <https://github.com/>



浙江大学
ZHEJIANG UNIVERSITY

2024 ZJU Summer School on Visualization and Visual Analytics

Thank you !

