

2025 年春季学期本科生《数据可视化》课程

第一次个人作业（使用 Python 进行数据可视化）

本文档最后修订于 2025 年 2 月 18 日。

作业描述

本作业可以在以下三个任务中任选其一；如果做了多于一个，则按照任务一的结果计分：

任务一：基于你之前课程中（如《数据科学导论》《Python 数据分析》《信息存储与检索》《机器学习》《文本挖掘》等）使用过的数据集，运用 Python 进行可视化（但也可根据数据形态与业务问题使用其他工具辅助）。请在 Jupyter Notebook 中完成可视化任务。你可以针对这一数据集进行任意内容的可视化，并在同一个 Jupyter Notebook 中对可视化结果进行描述、分析和讨论，最好产生一些故事线（storyline）。如果你使用了其他工具，那可以将全部内容打包并附说明文档。

任务二：基于在线教育数据集¹完成数据可视化。该数据集是某在线教育平台公开数据集，包含了平台课程数据、学习记录、班级数据、班级成员数据、学习日志和平时浏览记录等 9 张数据表，共计约 1200 万条数据，其中日志数据约有 15000 条。请使用可视化的方式来展示你的思考。数据、元数据见文件“在线教育.zip”。

任务三：基于光刻技术数据集完成可视化。该数据集下载自 OpenAlex 数据平台²，包括光刻（lithography）技术相关的论文信息及必要的机构数据。请使用可视化的方式，结合国际形势、技术发展等角度来展示你的思考。数据及说明见文件“光刻技术.zip”。

请选择任一数据集为主（可进一步补充需要的数据），使用 Python 程序设计语言，选择 [Matplotlib](#)、[seaborn](#)、[pyecharts](#)、[plotly](#) 等包，在 Jupyter Notebook 中完成可视化任务。你可以针对这一数据集进行任意内容的可视化，并在同一个 Jupyter Notebook 中对可视化结果进行描述、分析和讨论，最好产生一些故事线（storyline）。

下面是一些可视化实践，他们使用 Jupyter Notebook 处理数据、设计叙事、最终完成可视化的方式可供完成本作业时参考。参考 1：[探索奥运会数据](#)；参考 2：[奥运会的东道主优势](#)；参考 3：[初创企业的成败](#)。

作业要求与评分

¹ <https://tianchi.aliyun.com/dataset/172411>

² <https://docs.openalex.org/>

- 作业截止时间：北京时间 2025 年 3 月 17 日（周一）晚上 11:59:59，作业需要通过教学网指定位置提交。提交时，请将文件命名为“学号_姓名_个人作业 1.ipynb”（如“2100016601_张三_个人作业 1.ipynb”）。如果你使用了其他数据或工具，请将全部内容打包并附说明文档，将压缩包命名为“学号_姓名_个人作业 1.zip”。
- 必须独立完成，严禁合作、讨论、抄袭、套作；鼓励使用 ChatGPT 等人工智能工具辅助。
- 不得照搬或抄袭他人观点文字，需列出全部参考资料，必须遵照学术规范与诚信。
- 除遇不可抗力（不包括时间管理不善、课程冲突、数据或文档丢失等问题），如作业迟交在 24 小时以内，总分扣除 20%；迟交在 48 小时以内，总分扣除 40%；迟交在 72 小时内，总分扣除 60%；迟交在 96 小时内，总分扣除 80%；迟交 96 小时以上，该次作业不计入总分。
- 本次作业在期末总评中共占 10 分。
- 请注意，以上可视化任务不具有唯一答案。完成可视化任务后，请务必针对你做的可视化图进行描述和讨论。
- 评分细则如下：数据处理与整合（20%）、可视化图表的选择（20%）、可视化的设计（20%）、可视化结果的解读（30%）、文档的可读性（10%）。

评分细则

指标	指标的操作化	分值
数据处理与整合（20%）	数据处理的质量（数据清洗、数据描述等）	10%
	根据需求整合数据	10%
可视化图表的选择（20%）	图表选择的准确性（合适的图表类型）	15%
	图表选择的丰富度（图表类型不少于 3 种，更关注【质量】，而非【数量】，图表数量尽量不超过 10 张）	5%
可视化的设计（20%）	可视化设计的故事性	10%
	可视化设计的规范性（例：坐标轴、题注、配色等）	10%
可视化的解读（30%）	解读的准确性	15%
	解读的全面性	15%
文档的可读性（10%）	（代码等）必要的注释及 Markdown	4%
	文档的流畅性（步骤、图表间的衔接）	4%
	按规定格式提交作业	2%