

数据可视化学习计划

目录

1	学习资料	1
2	内容安排	2
3	学习成果	3

1 学习资料

创造营使用数据酷客平台的数据可视化本科版本课程作为本次计划的学习资源，包含课程视频、视频习题以及案例等。平台课程的内容包含

- ☐ 数据可视化的原理、工具与元素
- ☐ 结构数据的可视化
- ☐ 机器学习算法模型的可视化
- ☐ 非结构数据的可视化
- ☐ 视觉信息传递的 Infographics

五个方面的内容。本次计划选择前四部分重点学习，并使用 Python3.x 作为主要的编程语言。

- ☐ 数据酷客网站：
<http://www.cookdata.cn>
- ☐ 数据可视化课程链接：
http://cookdata.cn/course/course_introduction/37/
- ☐ 参考资源：
 - 《现代统计图形》
 - 《Python 数据可视化编程实战》
 - 《The Visual Display of Quantitative Information》第 2 版
 - 《Information visualization: perception for design》第 3 版

2 内容安排

根据学习进度安排表，创造营的小伙伴们可以登录数据酷客平台，学习数据可视化课程的相关章节内容。我们鼓励小伙伴们撰写学习笔记（Scribe Note）。在课程结束后，我们会整理汇集学习笔记制作成“小册子”在数据酷客平台发布并和广大的数据酷客用户交流分享。

□ 课程辅导：高扬（yangg@cookdata.cn）；吴佳佳（jiajiaw@cookdata.cn）

□ 学习进度安排：

时间	学习内容	学习章节
第一周	数据可视化基础	1. 数据可视化概述 2. 可视化工具与元素
第二周	结构数据的基本图形	3.1 基本数据图形 3.2 描述分布图形
第三周	结构数据的复杂图形	3.3 多维数据图形
第四周	算法模型的可视化分析 I	4.1 线性回归模型可视化分析 4.2 决策树可视化分析
第五周	算法模型的可视化分析 II	4.3 聚类算法可视化分析 4.4 降维算法可视化分析
第六周	非结构数据的可视化 I	5.1 时间数据可视化 5.2 网络数据可视化
第七周	非结构数据的可视化 II	5.3 空间数据可视化 5.4 文本数据可视化

3 学习成果

为了激发小伙伴儿们的学习兴趣、巩固学习的效果，创造营为大家规划了 CapStone 项目。根据兴趣选择其中一个作为主题，利用所学知识去完成这个 CapStone 项目。

□【必选】撰写课程章节的学习笔记，具体要求如下：

- 易读性。在观看学习课程视频的时候，请认真记录老师讲解的内容。在此基础上，组织完整的语句形成学习笔记，杜绝简单的摘抄和罗列。笔记的内容对于初学者是友好的，容易理解；
- 完整性。笔记内容应包含本章节的主要知识点，需要加入自己的阐述和理解，可以做适当的延伸，杜绝机械的重复。使用多种展现形式（如代码、图片）来丰富学习笔记是欢迎的；
- 规范性。学习笔记使用外部的资料需要注明引用来源，学习笔记以 Word 格式/Markdown 格式文档提交。课程助教会评审提交的学习笔记，反馈评审意见。学习笔记在数据酷客平台发布前，需要根据评审意见进行优化。

□【必选】CapStone 主题（二选一）：

- ✓研究类主题：请选择一种图形，借助学习的理论和实践知识撰写一份报告探索图形的历史起源、组成要素，分析图形的使用场景以及代码实现细节。报告采用 Jupyter Notebook 的格式，须包含文字和相关代码。（示例：选择热力图，分析热力图的来龙去脉和组成结构，聚焦使用热力图展现机器学习模型评价中的混淆矩阵的使用场景，分析混淆矩阵在 Scikit-learn 和 Yellowbrick 两种库中的实现方式，比较绘图源代码的异同。）
- ✓分析类主题：请使用数据可视化手段对数据集进行探索性分析，并撰写一份分析报告。报告采用 Jupyter Notebook 的格式，须包含文字和相关代码。报告的具体要求如下：
 - 包含结构化数据和非结构数据的可视化内容；

- 可视化图形种类不少于 4 种，为了便于分析，可引用外部数据集；
- 阐明使用某种类型图形进行可视化的原因，并对可视化的结果进行分析描述。

请从下列数据集中选择一个开展分析：

- Netflix 影视节目数据集；
- 酒店预订需求数据集；
- 食物偏好调查数据集。

报告样例链接：

http://cookdata.cn/note/view_static_note/197c2cb8f702ad86374a290374f7b2b2/