

中国社会科学院大学 数据可视化 课程大纲

课程基本信息 (Course Information)					
课程编号 (Course ID)	102113020032	*学时 (Credit Hours)	48	*学分 (Credits)	2
*课程名称 (Course Name)	数据可视化				
	Data visualization				
先修课程 (Prerequisite Courses)	大学计算机				
*课程简介 (Description)	课程为研究数据分析提供了一种非常适合于人文社科学生的方法：数据可视化分析。课程旨在提升学生计算思维、数据意识。 1. 课程通过 NumPy 讲解 N 维数据的表达、科学计算的基本概念和运算方法。 2. 通过 Pandas 讲解数据分析和处理。 3. 通过 Matplotlib 绘制坐标系、散点图、雷达图等，直观的展示数据趋势和特点。 4. 通过流行的数据可视化技术 D3 和商业 BI 工具 (Fine Report) 讲解案例。 5. 实例展示饼图、环图、词云图和地图等。 课程难点：通过编程实现数据可视化是本课程的难点。 主要教学方法：基于案例驱动的混合式教学，教师边讲边演示，学生互相演示学习。考勤、课堂参与及作业均在课堂派平台上完成。				
*课程简介 (Description)	The course provides a method of studying data analysis, such as data visualization analysis, with the aim of improving students' computational thinking and data awareness. It uses NumPy to explain the expression of N-dimensional data, the basic concepts, and operation methods of scientific computing. It also explains data analysis and processing through Pandas. Additionally, it draws coordinate systems, scatter plots, radar charts, etc. using Matplotlib to visually display data trends and characteristics. Furthermore, it explains the case of the popular data visualization technology D3 and commercial BI tool (Fine Report). Lastly, the course uses examples to display pie charts, ring charts, word cloud charts, maps, etc.				
*教材 (Textbooks)	Python 数据分析与可视化 (第 2 版), 魏伟一等, 清华大学出版社, 2021 年 7 月, 第 2 版, ISBN: 9787302577584				
参考资料 (Other References)	大数据可视化, 王珊珊等, 清华大学出版社, 2021 年 6 月, 第 1 版, ISBN: 9787302578352				
*课程类别 (Course Category)	☐ 公共基础课/全校公共必修课 ☒ 通识教育课 ☐ 专业基础课 ☐ 专业核心课/专业必修课 ☐ 专业拓展课/专业选修课 ☐ 其他_____				
*授课对象 (Target Students)	全校本科生	*授课模式 (Mode of Instruction)	☐ 线上, 教学平台_____		
			☐ 线下 ☒ 混合式 ☐ 其他		
			☐ 实践类 (70%以上学时深入基层) _		

*开课院系 (School)	计算机教研部		*授课语言 (Language of Instruction)	☒ 中文 ☐ 全外语_____ ☐ 双语：中文+_____ (外语讲授不低于 50%)			
*授课教师信息 (Teacher Information)	课程负责人 姓名及简介		朱俭，男，中国社会科学院大学计算机教研部，副教授，北京理工大学计算机博士，主要研究方向为自然语言处理、人工智能、数据分析。2020 年荣获中国社会科学院大学校级青年教学名师称号。2016 年荣获北京市青年岗位能手。				
	团队成员 姓名及简介		无				
学习目标 (Learning Outcomes)	1.掌握利用 Python 语言表达 N 维数据并结合数据特点合理展示数据的技术和方法； 2.掌握数据表示、清洗、统计和展示的能力； 3.掌握 Pandas 专业级数据分析和处理； 4.掌握 Matplotlib 绘制可视化图表； 5.了解 Python 计算生态中最优秀的数据分析和展示技术； 6.了解数据领域最优秀的可视化工具库，应用到人文社科专业领域（管理、经济、社会学专业）。						
*考核方式 (Grading)	平时成绩：30% 期末考试：70% 期末考核方式： 课下设计（数据分析程序+作品文档+可视化视频展示）						
*课程教学计划（Teaching Plan）填写规范化要求见附件							
周次 教师	周 学时	其中					教学内容摘要 (必含章节名称、讲述的内容提要、实验的名称、教学方法、课堂讨论的题目、阅读文献参考书目及作业等)
		讲 授	实 验 课	习 题 课	课 程 讨 论	其 他 环 节	
第一周	3	2	1				章节名称：第一章 数据分析与可视化概述 讲授：1.数据的本质 2.数据与信息 3.数据与可视化 4.可视化的基本理论和可视化工具 5.演示 Python 基本编程环境 6.讲解 Python 代码精要总结 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：第一周作业 Python 基础，在搭建的环境里练习 Python 的代码总结
第二周	3	2	1				章节名称：第二章 Python 编程基础 讲授：1.讲解 Anaconda 的安装 2.PyCharm 的安装 3.Python 的变量计算和类型转换 4.字符串操作 5.总结 Python 读写文件、列表和字典常用操作代码 6.介绍可视化工具：Excel、D3、ECharts、Tableau 等

						教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：练习列表和字典常用操作代码
第三周	3	2	1			章节名称：第三章 NumPy 数值计算基础 讲授：1.NumPy 概述 2.NumPy 数组(ndarray)对象 3.ufunc 函数 4.按列或行拼接 NumPy 数组 5.按列对 NumPy 数组进行排序 6.用 NumPy 处理日期 7.介绍高阶 NumPy 函数 8.获取满足条件件的索引 9.将数据导入和导出 csv 文件 10.保存和加载 NumPy 对象 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：NumPy 基础，NumPy 练习题 20 道
第四周	3	2	1			章节名称：第四章 Pandas 统计分析基础 第 1-4 节 讲授：1.Pandas 中的数据结构 (1)Series(2>DataFrame(3)索引对象(4)查看 DataFrame 的常用属性 2.Pandas 索引操作 (1)重建索引(2)更换索引 3.DataFrame 数据的查询与编辑 (1)DataFrame 数据的查询(2)DataFrame 数据的编辑 4.Pandas 数据运算 (1)算术运算(2)函数应用和映射(3)排序(4)汇总与统计 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：Pandas 基础
第五周	3	2	1			章节名称：第四章 Pandas 统计分析基础 第 5-8 节 讲授：1.数据分组与聚合 (1)数据分组(2)数据聚合(3)分组运算 2.数据透视表 (1)透视表(2)交叉表 3.Pandas 可视化 (1)线形图(2)柱状图(3)直方图和密度图 4.散点图 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：Pandas 数据载入与预处理，Pandas 进阶 120 题
第六周	3	2	1			章节名称：第六章 Matplotlib 数据可视化基础 第 1-3 节 讲授：1.Matplotlib 库入门 2.Matplotlib 绘图基础 (1)创建画布与子图 2)添加画布内容 3)绘图的保存与显示 3.Pyplot 的动态参数配置 1)全局参数定制 2)参数设置 3)绘图的填充 4)在绘图中显示公式 5)配置说明 教学方法：基于案例驱动的混合式教学，边讲边练

						实验名称：50 题 matplotlib 从入门到精通
第七周	3	2	1			章节名称：第六章 Matplotlib 数据可视化基础 第 4-5 节 讲授：1. Matplotlib 的常用绘图 (1)折线图(2)散点图(3)直方图(4)饼图(5)箱线图(6)概率图(7)雷达图(8)流向图(9)绘图中的表格设置(10)极坐标图 2. 词云 (1)安装相关的包(2)词云生成过程(3)词云生成示例 3.学习绘制进阶统计图形、堆积柱状图、分块柱状图、堆积折线图、间断条形图、内嵌环形饼图、箱线图等 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：多个统计图形的组合展示
第八周	3	2	1			章节名称：第八章 Pyecharts 可视化 第 1-2 节 讲授：1.pyecharts 简介 2.pyecharts 的使用方法 3.pyecharts 常用图表 (1)柱状图(2)饼图(3)漏斗图(4)散点图(5)K 线图(6)仪表盘(7)词云(8)组合图表(9)桑基图(10)平行坐标图和地图 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：Pyechart 可视化图表 1
第九周	3	2	1			章节名称：第八章 Pyecharts 可视化 第 3-4 节 讲授：1.学习可视化 Pyecharts 库 (1)折线图-Line(2)学习条形图-Bar(3)饼图-Pie(4)柱状图(5)漏斗图-Funnel(6)热力图-Heatmap(7)热力图-Calender(8)地理图表-Province 的绘制(9)地理热点图-Geo(10)3D 散点图-scatter3D(11)联合图等图表的绘制 2.可视化综合案例展示 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：Pyechart 可视化图表 2
第十周	3	2	1			章节名称：期中复习和总结 讲授：1.复习 Pandas 数据清洗的知识要点；2.复习总结可视化绘图知识点；3.案例演示：十万条招聘信息的数据可视化。 教学方法：教师带领回顾总结，边讲边练 实验名称：期中测试题一套
第十一周	3	2	1			章节名称：第十三章 综合案例 第 1 节 讲授：1.讲解职业人员体检数据案例的可视化步骤： (1)表格数据规范化 (2)去除实习岗位的数据信息 (3)将公司地址统一转化为所在城市 (4)对职业薪资的分类进行规范化 (5)增加一列计算薪资奖金率 (6)将血液各项目指标转化为数值形式 (7)处理极端异常值，剔除月薪小于 2000 和大于 55000 的数据 (8)将数量较少且为传统职业的数据归到其他行业。将职业来源划分为互联网、计算机软件、移动互联网、电子商务、数据服务。互联网金融、游戏、在线教育、生活服务、医疗健康、贸易进出口、企业服务、银行及其他等多个类别

						(9)通过 Python 和 Matplotlib 完成对职业危害因素对人体血液等系统的影响，用可视化图表展示 2.讲解 2014-2019 年高考数据可视化的综合案例 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：选择自己所在的省份比如湖南，完成湖南高考数据可视化案例，将课堂内容复现
第十二周	3	2	1			章节名称：第十三章 综合案例 第 2 节 讲授：1.用 BI 工具连接多种数据源（文本、csv、excel、数据库） 2.基础数据处理步骤总结 3.可视化报表设计和制作、网络发布 4.可视化报表在线浏览和分析 教学方法：用案例演示总结本学期教授的数据可视化技能 实验名称：总结复习本课程的重点内容，讲解课下设计的注意事项。
第十三周	3	2	1			章节名称：第十三章 综合案例 第 3 节 讲授：体验用 BI 工具绘制可视化图表。 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：多个统计图形的组合展示
第十四周	3	2	1			章节名称：复习和总结 1 讲授：1.数据爬取工具；2.数据库的安装；3.数据库连接工具；4.数据库导入和查询；4.用 BI 工具完成 4 万条北京二手房市场交易数据可视化分析。 教学方法：基于案例驱动的混合式教学，边讲边练 实验名称：将课堂内容复现
第十五周	3	2	1			章节名称：复习和总结 2 讲授：用案例演示本学期教授的数据可视化技能，并通过一套模拟题来加强学生技能 实验名称：复习、总结，练习一套模拟训练题
第十六周	3	2	1			章节名称：课堂展示 讲授：学生分组做展示，讲解自己做的数据可视化分析效果。由学生和教授共同打分。 教学方法：翻转课堂 实验名称：完成并演示课下设计
总计	4 8	3 2	16			
备注（Notes）						