

综合素质选修课课程教学大纲

【R 与 Python 数据分析基础】

【Fundamental to R and Python Programming Data Analysis】

一、基本信息

课程代码：【】

课程学分：【2】

面向专业：【不限】

课程性质：【综合素质选修课】

开课院系：【信息技术学院计算机科学与技术系】

使用教材：

教材：【】

参考书目

【数据可视化分析——基于 R 语言（第 2 版），贾俊平，中国人民大学出版社，2021 年 5 月】

【Python 程序设计（第 3 版 微课版），董付国，清华大学出版社，2020 年 6 月】

课程网站网址：

先修课程：无

二、课程简介

R 语言与 Python 语言都是经典数据科学方法处理语言，通过 R 语言与 Python 语言的学习，可以极大优化理工科，人文社科类学生知识结构，重构数据科学方法思维。同时通过学习 R 语言与 Python 语言在数据科学方法中的数据处理技术，为学生后续学习提供一项数据处理必备技能。

本课程在注重数据科学方法的基础上，精选国际教材、国外著名高校的案例数据，采用开源软件包的 R 语言与 Python 语言编写程序，而练习案例则尽可能结合中国实际，采用国家、地方统计局、高校数据库的数据进行数据科学建模，R 语言与 Python 语言程序编写，对实际问题进行探索性发现。

在教学中，根据非统计专业选修课的特点，课程突出数据科学在大数据时代发展趋势，强调案例演示的数据可视化技术，增强课程的趣味性，可阅读性。

三、选课建议

本课程面向全校所有专业和年级。

四、课程目标/课程预期学习成果

R 语言与 Python 语言的语法规则、数据结构、数据清洗、数据变换、数据整合、数据可视

化、数据建模的学习，较好地训练学生利用计算机解决对数据进行分析 and 展现，使学生具有数据分析和数据展现的能力，未培养学生有初步数据开发能力打下良好基础。R 语言与 Python 语言的语法规则、数据结构、数据清洗、数据变换、数据整合、数据可视化、数据建模的学习，较好地训练学生利用计算机解决对数据进行分析 and 展现，使学生具有数据分析和数据展现的能力，未培养学生有初步数据开发能力打下良好基础。

序号	课程预期学习成果	课程目标 (细化的预期学习成果)	教与学方式	评价方式
1	L0311 编程实现： 使用 R 语言与 Python 语言环境，掌握程序示例，理解 R 语言 Python 语言的基本语法。	1. 掌握 R 语言与 Python 开发必备的基础知识。包括环境搭建、数据结构、数据输入与导出 (TXT, CSV)。 2. 掌握基本描述统计和简单假设检验。 3. 掌握基本图形、中级图形的理解与 R 语言与 Python 语言上编程作图。 4. 不同领域库的介绍及使用	案例教学、任务引领、练习实践	实作评价、测试等
2	L0514 了解在不同应用场景下如何使用该语言。	1. R 语言与 Python 安装及图表制作开发 2. 综合运用 R 语言与 Python 语言编程技巧，根据个人兴趣点，自行收集数据、整理数据、展示数据、分析数据，挖掘数据价值。能够正确合理使用 R 语言实现数据分析，并将此案例整理成数据分析报告。	自主学习、实践	资料汇总
			自主学习、实践	大作业、作品展示

五、课程内容

第一单元 R 语言与 Python 语言基础知识

(一) 课程主要内容

本单元主要介绍 R 语言与 Python 语言的基础知识。要求掌握如何搭建不同操作系统下的 R 语言与 Python 语言不同版本的编程环境，了解安装过程中的陷阱和关键点。掌握 R 语言与 Python 语言变量、简单数据类型的定义，着重理解并掌握数组、向量、矩阵、数据框、因子、列表等重

要的数据结构。掌握对这些 R 语言与 Python 语言数据结构的操作和处理方法。了解并掌握 R 语言与 Python 语言编程规范、用户输入和程序设计结构、函数编程技巧（参数传递各类方法、将函数存储在模块中等）和函数编写指南。

（二）教学建议

本单元重点：R 语言与 Python 语言编程开发环境搭建和重要数据结构。

本单元难点：数组、向量、矩阵、数据框、因子、列表等重要数据结构，R 语言与 Python 语言函数参数的传递（传递实参、列表、任意数量实参），以及对数组、向量、矩阵、数据框、因子、列表的数据操作技巧。

教学建议：本单元操作内容较多、实践性较强，建议结合课中实例和课后习题，讲练结合，互动交流，使学生通过多动手编写代码，建立感性认识，从而更熟练和深入地理解 R 语言与 Python 语言编程规范和基本方法。

第二单元 R 语言与 Python 语言绘图工具的使用方法

（一）课程主要内容

本单元主要介绍 R 语言与 Python 语言的统计分析功能、绘图的语言和操作环境。要求掌握成分关系图表（柱图、饼图）、对比关系图（时序图、折线图、面积图）、相关关系图表（散点图、气泡图）、多维比较图表（雷达图、圆环图）。熟练掌握散点点图、折线图、直方图、饼图、柱状图、箱线图、密度图、直线的绘制方法。理解 ggplot2 图层绘图原理，并熟练使用 ggplot2 绘图。

（二）教学建议

本单元重点：理解并掌握 R 语言与 Python 语言的统计分析功能以及各种图表的绘制方式。理解 ggplot2 图层绘图原理，并熟练使用 ggplot2 绘图。

本单元难点：理解 ggplot2 图层绘图原理，并熟练使用 ggplot2 绘图。

教学建议：本单元理论知识较多，建议适当采用课堂讨论，对于较抽象的理论知识结合课中及课后习题进行讨论、分析、进一步加深理解。课堂和课后多以教学实例阐释概念，多通过讲解代码印证理论。

第三单元 项目和专题

（一）课程主要内容

本单元主要通过几个示例项目介绍使用 R 语言与 Python 语言编程基础知识进行统计分析、数据可视化处理、应用开发的基本方法，另外通过一个网络爬虫的专题，介绍 R 语言与 Python 语言进行网络数据爬取和数据分析的基本实现路径。要求同学们在掌握前续第一、第二单元知识

的前提下，实现几个统计专题的开发，在学有余力的情况下，实现简单数据可视化项目。在最后的 R 语言与 Python 语言爬虫算法专题中，介绍当前热门的爬虫算法和数据存储、数据分析方法，作为知识的扩展与延伸。

(二) 教学建议

本单元重点: 完整统计专题/数据可视化/应用项目的开发, R 语言与 Python 语言网络爬虫算法和数据分析方法。

本单元难点: 完整项目的构思、版本控制和实现, 数据可视化中 ggplot2 包的使用, R 语言与 Python 语言爬虫框架的熟悉, 网络爬虫算法及数据分析模型。

教学建议: 本单元动手内容较多、实践性较强, 建议结合具体项目, 讲练结合, 互动交流, 以 Seminar 方式调动学生的积极性, 以分享和讨论的形式, 综合运用前续课堂教授的知识, 从整体上切实掌握 R 语言与 Python 语言在某些领域的开发方法, 提升同学们创造性的编程能力。

六、课内实验名称及基本要求 (适用于课内实验)

序号	实验名称	主要内容	实验时数	实验类型	备注
1	R 语言与 Python 语言开发环境搭建	学会在不同操作系统下面搭建和配置 R 语言与 Python 语言开发环境, 编写第一个简单的 R 语言与 Python 语言程序, 体验 R 语言与 Python 语言代码的运行。	2	设计型	PC 机每人 1 台; 各类 OS 环境下的 R 语言与 Python 语言软件和开发工具。
2	R 语言与 Python 语言程序实例解析	R 语言与 Python 语言编程初体验, 通过温度转换实例、R 语言与 Python 语言程序的结构框架、相关语句和编程结构。	4	设计型	PC 机每人 1 台; R 语言与 Python 语言开发环境。
3	R 语言与 Python 语言基本数据类型和复杂数据结构	掌握 R 语言与 Python 语言变量命名和基本数据类型, 掌握字符串、数字类型的处理技巧, 尤其要掌握数组、向量、矩阵、数据框、因子、列表等数据结构的操作。	4	设计型	PC 机每人 1 台; R 语言与 Python 语言开发环境。

4	R 语言与 Python 语言图表设计与输出	掌握 R 语言与 Python 语言中条形图（简单的条形图、堆砌条形图和分组条形图、均值条形图、条形图的微调、棘状图）、饼图、直方图、核密度图、箱线图、小提琴、点图，以及散点图（散点图矩阵、高密度散点图、三维散点图、气泡图、相关图、马赛克图）、折线图、相关图、马赛克图。	4	设计型	PC 机每人 1 台； R 语言与 Python 语言开发环境。
5	R 语言与 Python 语言应用场景知识拓展	学会通过利用第三方库，编写统计图、科学计算和可视化、网络爬虫和自动化等应用（可根据个人爱好，选择一至二个实现，学有余力者可完成多个）	4	设计型	PC 机每人 1 台； R 语言与 Python 语言开发环境。

七、评价方式与成绩

总评构成 (1+X)	评价方式	占比
1	论文	50%
X1	课堂学习	10%
X2	章节测验	20%
X3	课外扩展阅读	20%

撰写人：彭青松

系主任审核签名：戴智明

审核时间：2023.2.18