

DATA HANDLING AND VISUALIZATION WITH R			
Course Code:	AD1602-1	Course Type	PCC Lab
Teaching Hours/Week (L: T: P)	0:0:2	Credits	01
Total Teaching Hours	0+0+26	CIE + SEE Marks	50+50
Prerequisite	CS1001-1, AD1102-1		
Teaching Department: Artificial Intelligence & Data Science			
Course Objectives:			
1.	To understand methods for data preprocessing.		
2.	To apply appropriate models for different types of data.		
3.	To visualize and present data.		
4.	To perform complex data manipulations and automate data analysis tasks.		
5.	To develop programs for data manipulation, data cleaning, and implementing algorithms.		
List of Experiments			
1	Program for Data Analysis:		
2	<ul style="list-style-type: none"><li>Read a dataset from a CSV file and perform exploratory data analysis, including summary statistics, data visualization, and identifying missing values.</li><li>Clean the data by removing duplicates, handling missing values, and transforming variables if necessary.</li><li>Perform data manipulation operations such as filtering, sorting, and merging datasets based on certain criteria.</li><li>Conduct statistical analysis, such as hypothesis testing or correlation analysis, to derive insights from the data.</li><li>Generate reports or visualizations to present the analysis results.</li></ul>		
3	Program for Linear Regression:		
4	<ul style="list-style-type: none"><li>Read a dataset from a CSV file that contains variables for independent and dependent variables.</li><li>Perform linear regression analysis to model the relationship between the variables.</li><li>Calculate the coefficients and intercept of the regression model.</li><li>Evaluate the model's goodness-of-fit using metrics like R-squared and adjusted R-squared.</li><li>Plot the regression line and residuals to visualize the relationship between the variables.</li></ul>		
5	Program for Web Scraping and Data Extraction: Use R packages like rvest or httr to scrape data from a specific website or API. <ul style="list-style-type: none"><li>Define the target website or API endpoints and specify the data to be extracted.</li><li>Retrieve the HTML content or JSON response from the website or API.</li><li>Parse and extract the desired data using CSS selectors, XPath, or JSON parsing techniques.</li><li>Save the extracted data to a file or perform further analysis on it.</li></ul>		
6	Program for Web Scraping and Data Extraction:		

	<p>Use R packages like rvest or httr to scrape data from a specific website or API.</p> <ul style="list-style-type: none"> <li>• Define the target website or API endpoints and specify the data to be extracted.</li> <li>• Retrieve the HTML content or JSON response from the website or API.</li> <li>• Parse and extract the desired data using CSS selectors, XPath, or JSON parsing techniques.</li> <li>• Save the extracted data to a file or perform further analysis on it.</li> </ul>
7	Program for Web Scraping and Data Extraction:
8	<ul style="list-style-type: none"> <li>• Install and Load Required Packages:           <ul style="list-style-type: none"> <li>▪ Install the necessary packages for web scraping, such as rvest, xml2, and httr.</li> <li>▪ Load the packages into your R environment using the library() function.</li> </ul> </li> <li>• Specify the Target Website and URL:           <ul style="list-style-type: none"> <li>▪ Identify the website you want to scrape data from.</li> <li>▪ Define the URL of the specific webpage or API endpoint containing the data you need.</li> </ul> </li> <li>• Send HTTP Requests and Handle Authentication (if required):           <ul style="list-style-type: none"> <li>▪ Use the GET() or POST() functions from the httr package to send HTTP requests to the website.</li> <li>▪ Set headers, parameters, or authentication credentials as needed.</li> </ul> </li> <li>• Retrieve HTML Content and Parse XML/HTML:           <ul style="list-style-type: none"> <li>▪ Use the GET() function to retrieve the HTML content of the webpage.</li> <li>▪ Parse the HTML content using the read_html() function from the rvest package.</li> </ul> </li> <li>• Extract Data from HTML:           <ul style="list-style-type: none"> <li>▪ Inspect the HTML structure of the webpage to identify the elements and attributes containing the desired data.</li> <li>▪ Use functions such as html_nodes() and html_text() from the rvest package to extract specific elements or text from the HTML content.</li> <li>▪ Apply CSS selectors or XPath expressions to target specific elements.</li> </ul> </li> <li>• Perform Data Cleaning and Transformation:           <ul style="list-style-type: none"> <li>▪ Clean the extracted data by removing unwanted characters, handling missing values, or applying regular expressions.</li> <li>▪ Convert the extracted data into appropriate data structures (e.g., data frames, lists, or vectors) for further analysis or storage.</li> </ul> </li> <li>• Save or Analyse the Extracted Data:           <ul style="list-style-type: none"> <li>▪ Save the extracted data to a file (e.g., CSV or Excel) using R functions like write.csv() or write.xlsx().</li> <li>▪ Perform further data analysis or visualization on the extracted data using appropriate R packages and techniques.</li> </ul> </li> </ul>
9	Program for Data Visualization:
10	<p>Use packages like ggplot2 or plot to create various types of charts, such as bar charts, line plots, scatter plots, or heatmaps.</p> <ul style="list-style-type: none"> <li>• Read a dataset from a CSV file or other data sources.</li> <li>• Customise the charts by adding labels, titles, legends, and adjusting the axis scales.</li> </ul>

	<ul style="list-style-type: none"> <li>Create interactive visualizations with tooltips, zooming, or filtering options.</li> <li>Export the visualizations to different file formats or display them within an R notebook or Shiny application.</li> </ul>
11	Program for Data Manipulation:
12	<ul style="list-style-type: none"> <li>Read multiple datasets from different files or sources.</li> <li>Merge or join the datasets based on common variables or keys.</li> <li>Perform aggregation operations, such as calculating sums, means, or counts, by groups or categories.</li> <li>Filter the data based on specific conditions or criteria.</li> <li>Create new variables or transform existing variables using functions or mathematical operations.</li> </ul>
13.	Lab Exam

**Course Outcomes:** At the end of the course student will be able to

1.	Apply methods for data preprocessing, exploratory data analysis, and derive meaningful insights from the data.
2.	Select and apply appropriate models for different types of data and problem domains, interpret model outputs, and evaluate model performance.
3.	Effectively visualize and present data using charts, graphs, and other visual representations.
4.	Perform complex data manipulations, handle large datasets, and automate data analysis tasks.
5.	Apply the programming skills in R programming languages for data manipulation, data cleaning, and implementing algorithms.

#### Course Outcomes Mapping with Program Outcomes & PSO

Program Outcomes→	1	2	3	4	5	6	7	8	9	10	11	12	PSO↓		
↓ Course Outcomes													1	2	3
<b>AD1602-1.1</b>	2	1	-	-	2	-	-	-	-	-	-	1	3	2	-
<b>AD1602-1.2</b>	2	1	-	-	2	-	-	-	-	-	-	1	3	2	-
<b>AD1602-1.3</b>	2	2	1	-	2	-	-	-	-	-	-	1	3	2	-
<b>AD1602-1.4</b>	2	1	1	-	2	-	-	-	-	-	-	1	3	2	-
<b>AD1602-1.5</b>	2	1	-	-	2	-	-	-	-	-	-	1	3	2	-

**1: Low 2: Medium 3: High**

#### REFERENCE BOOKS:

1.	Hadley Wickham, Garrett Grolemund, "R for Data Science", OREILLY Publication, 2017.
2.	Steven Keller, "R Programming for Beginners", CreateSpace Independent Publishing Platform 2016.
3.	Roger D. Peng, "R Programming for Data Science" Lean Publishing, 2016.