

Course Code	PAD21D08J	Course Name	EXPLORATORY DATA ANALYSIS	Course Category	D	Discipline Specific Elective	L	T	P	C
							4	0	4	6

Pre-requisite Courses	Nil	Co-requisite Courses	Nil	Progressive Courses	Nil
Course Offering Department	Data Science	Data Book / Codes/Standards			Nil

Course Learning Rationale (CLR):	The purpose of learning this course is to,	Learning	Program Learning Outcomes (PLO)
----------------------------------	--	----------	---------------------------------

CLR-1 :	The essential exploratory techniques for summarizing data with R	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLR-2 :	To develop more complex statistical models																		
CLR-3 :	Eliminating or sharpening potential hypotheses about the world that can be addressed by the data																		
CLR-4 :	The plotting systems in R																		
CLR-5 :	The basic principles of constructing informative data graphics																		

Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:	Level of Thinking (Bloom)	Expected Proficiency (%)	Expected Attainment (%)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLO-1 :	Making exploratory graphs	2	85	80	H	H	H	H	H	H	-	M	M	L	-	H	-	M	H
CLO-2 :	Principles of analytic graphics	3	85	80	L	H	H	H	H	H	-	M	M	L	-	H	-	M	H
CLO-3 :	Plotting systems and graphics devices in R	3	85	80	L	H	H	H	H	H	-	M	M	L	-	H	-	M	H
CLO-4 :	The base and ggplot2 plotting systems in R	3	85	80	L	H	H	H	H	H	-	M	M	L	-	H	-	M	H
CLO-5 :	Clustering methods	3	85	80	L	H	H	H	H	H	-	M	M	L	-	H	-	M	H
CLO-6 :	Dimension reduction techniques	3	85	80	L	H	H	H	H	H	-	M	M	L	-	H	-	M	H

Duration (hour)	24	24	24	24	24
S-1	SLO-1 <i>Getting Started with R</i>	<i>Principles of Analytic Graphics</i>	<i>Plotting Systems</i>	<i>Hierarchical Clustering</i>	<i>The ggplot2 Plotting System: Part 1</i>
	SLO-2 <i>Installation</i>	<i>Show comparisons</i>	<i>The Base Plotting System</i>	<i>Hierarchical clustering Explanation</i>	<i>The Basics: qplot()</i>

S-2	SLO-1	Getting started with the R interface	Show causality, mechanism, explanation, systematic structure	The Lattice System	How do we define close?	Before You Start: Label Your Data
	SLO-2	Installing R on Windows	Show multivariate data	The ggplot2 System	Example: Euclidean distance	ggplot2 "Hello, world!"
S-3	SLO-1	Managing Data Frames with the dplyr package	Integrate evidence	Graphics Devices	Example: Manhattan distance	Modifying aesthetics
	SLO-2	Data Frames	Describe and document the evidence	The Process of Making a Plot	Example: Hierarchical clustering	Adding a geom
S-4	SLO-1	The dplyr Package	Content, Content, Content	How Does a Plot Get Created?	Prettier dendrograms	Histograms
	SLO-2	Installing the dplyr package	References	Graphics File Devices	Merging points: Complete	Facets
S-5 – S-8	SLO-1	Lab 1: Working on the dplyr package	Lab 4:	Lab 7:	Lab 10:	Lab 13:
	SLO-2					
S-9	SLO-1	select()	Exploratory Graphs	Multiple Open Graphics Devices	Merging points: Average	The ggplot2 Plotting System: Part 2
	SLO-2	filter()	Characteristics of exploratory graphs	Copying Plots	Using the heatmap() function	Basic Components of a ggplot2 Plot
S-10	SLO-1	arrange()	Air Pollution in the United States	The Base Plotting System	K-Means Clustering	Example: BMI, PM2.5, Asthma
	SLO-2	rename()	Getting the Data	Base Graphics	Illustrating the K-means algorithm	Building Up in Layers
S-11	SLO-1	mutate()	Simple Summaries: One Dimension	Simple Base Graphics	Stopping the algorithm	First Plot with Point Layer
	SLO-2	group_by()	Five Number Summary	Some Important Base Graphics Parameters	Using the kmeans() function	Adding More Layers: Smooth
S-12	SLO-1	Boxplot	Boxplot	Base Plotting Functions	Building heatmaps from K-means solutions	Adding More Layers: Facets
	SLO-2	Summary	Histogram	Base Plot with Regression Line	further resources	Modifying Geom Properties
S-13 – S-16	SLO-1	Lab 2: Working on filter, mutate	Lab 5: Drawing Graphs and Histograms	Lab 8: Applying Plotting functions	Lab 11: Implementing K-Means algorithms	Lab 14: Working with plotting
	SLO-2					
S-17	SLO-1	Exploratory Data Analysis Checklist	Overlaying Features	Plotting and Color in R	Dimension Reduction	Modifying Labels
	SLO-2	Formulate your question	Barplot	Colors 1, 2, and 3	Matrix data	Customizing the Smooth
S-18	SLO-1	Read in your data	Simple Summaries: Two Dimensions and Beyond	Connecting colors with data	Patterns in rows and columns	Changing the Theme
	SLO-2	Check the packaging	Multiple Boxplots	Color Utilities in R	Related problem	More Complex Example

S-19	SLO-1	Run str()	Multiple Histograms	colorRamp(),colorRampPalette()	SVD and PCA	A Quick Aside about Axis Limits
	SLO-2	Look at the top and the bottom of your data	Scatterplots	RColorBrewer Package	Unpacking the SVD: u and v	Case Study: MAACS Cohort
S-20	SLO-1	Validate with at least one external data source	Scatterplot - Using Color	Using the RColorBrewer palettes	SVD for data compression	Example: Face data
	SLO-2	Try the easy solution first	Multiple Scatterplots	The smoothScatter() function	Components of the SVD - Variance explained	Summary of qplot()
S-21 - S-24	SLO-1 SLO-2	Lab 3:	Lab 6:Working with Scatter Plots	Lab 9:Working with Colour Palettes	Lab 12:Dimension reduction	Lab 15: working with gplot()

Learning Resources	1. Exploratory Data Analysis with R, Roger D. Peng 2. Unwin, Antony. 2015. Graphical Data Analysis with R. (Links to an external site.)Links to an external site. CRC Press. ISBN 978-1498715232
---------------------------	---

Learning Assessment											
	Bloom's Level of Thinking	Continuous Learning Assessment (50% weightage)								Final Examination (50% weightage)	
		CLA – 1 (10%)		CLA – 2 (10%)		CLA – 3 (20%)		CLA – 4 (10%)#			
		Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice
Level 1	Remember	20 %	20 %	20 %	20 %	20 %	20 %	20 %	20 %	20 %	20 %
	Understand										
Level 2	Apply	20 %	20 %	20 %	20 %	20 %	20 %	20 %	20 %	20 %	20 %
	Analyze										
Level 3	Evaluate	10 %	10 %	10 %	10 %	10 %	10 %	10 %	10 %	10 %	10 %
	Create										
	Total	100 %		100 %		100 %		100 %		100 %	

CLA – 4 can be from any combination of these: Assignments, Seminars, Tech Talks, Mini-Projects, Case-Studies, Self-Study, MOOCs, Certifications, Conf. Paper etc.,

Course Designers		
Experts from Industry	Experts from Higher Technical Institutions	Internal Experts
	Dr.Muthu, Professor, Loyola College, Chennai	1. Dr.Jayashree, SRM IST

	Dr. Vincent, Associate Professor, VIT	
--	---------------------------------------	--