

Course Outline



General Information

COURSE ID (CB01A AND CB01B)

CIS 44H

COURSE TITLE (CB02)

R Programming

COURSE CREDIT STATUS

Credit - Degree Applicable

EFFECTIVE TERM

Fall 2023

COURSE DESCRIPTION

This course is an introduction to the R programming language and its utility in big data analytics. Topics covered include data objects, data cleansing, merging and sorting, statistical analysis of data, data graphics, and visualization, and working with R-Studio.

FACULTY REQUIREMENTS

COURSE FAMILY

Not Applicable

Course Justification

This course is part of the Database Development Practitioner Certificate and is CSU transferable. An introduction to the R programming language will be provided, which is a free software environment for statistical computing and graphics and is widely used among statisticians and data miners for developing statistical software and data analysis.

Foothill Equivalency

DOES THE COURSE HAVE A FOOTHILL EQUIVALENT?

No

FOOTHILL COURSE ID

Formerly Statement

FORMERLY STATEMENT (Formerly CIS 064H.)

Course Development Options

BASIC SKILL STATUS (CB08)

Course is not a basic skills course.

GRADE OPTIONS

- Letter Grade
- Pass/No Pass

REPEAT LIMIT

0

Transferability & Gen. Ed. Options

Information below is subject to change. For the official listing of courses, their approval dates, and transfer credit limitations, check the De Anza catalog (by academic year), [ASSIST.ORG \(https://assist.org/\)](https://assist.org/) and [C-ID.NET \(https://c-id.net/\)](https://c-id.net/).

TRANSFERABILITY

Transferable to both UC and CSU

Units and Hours

Summary

MINIMUM CREDIT UNITS	4.5
MAXIMUM CREDIT UNITS	4.5

Weekly Student Hours

Type	In Class	Out of Class
Lecture Hours	4.0	8.0
Laboratory Hours	1.5	0.0

Course Student Hours

COURSE DURATION (WEEKS)

12.0

HOURS PER UNIT DIVISOR

36.0

Course In-Class (Contact) Hours

LECTURE

48.0

LABORATORY

18.0

TOTAL

66.0

Course Out-of-Class Hours

LECTURE

96.0

LABORATORY

0.0

NA

0.0

TOTAL

96.0

Prerequisite(s)

Corequisite(s)

Advisory(ies)

ESL 272 and ESL 273, or ESL 472 and ESL 473, or eligibility for EWRT 1A or EWRT 1AH or ESL 5
CIS 22A or CIS 36A or CIS 40

Limitation(s) on Enrollment

Entrance Skill(s)

General Course Statement(s)

Methods of Instruction

Lecture and visual aids
Discussion of assigned reading Discussion and problem solving performed in class Collaborative learning and
small group exercises Collaborative projects

Assignments

- A. Reading: Required reading from the textbook and class notes
- B. Programs: 7-10 programming homework assignments.
- C. Group Project: Data exploration and visualization of assigned datasets.

Methods of Evaluation

- A. One or two midterm examinations requiring some programming, concepts clarification and exhibiting mastery of R programming constructs presented in the course.
- B. A final examination requiring concepts clarification and exhibiting mastery of data exploration, analysis and visualization principles.
- C. Evaluation of programming assignments and group project, based on correctness, documentation, code quality, and test plan executions.

Essential Student Materials/Essential College Facilities

Essential Student Materials:

- None.

Essential College Facilities:

- Access to a computer lab with RStudio

Examples of Primary Texts and References

Author	Title	Publisher	Date/ Edition	ISBN
Wickham, Hadley and Grolemund, Garrett: R for Data Science: Import, Tidy, Transform, Visualize, and Model Data 1st Edition. O'Reilly. ISBN-13: 978-1491910399, 2017.				
Campbell, Matthew: Learn RStudio IDE: Quick, Effective, and Productive Data Science 1st Edition. Apress. ISBN-13: 978-1484245101, 2019.				

Examples of Supporting Texts and References

Author	Title	Publisher
Matloff, Norman: The Art of R Programming: A Tour of Statistical Software Design 1st Edition. William-Pollock. ISBN-13: 978-1593273842, 2011.		
Teetor, Paul: R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 1st Edition. O'Reilly. ISBN-13: 978-0596809157, 2011.		

Learning Outcomes and Objectives

Course Objectives

- Describe R basics
- Exhibit understanding of R data objects.
- Illustrate basic data transformation concepts.
- Demonstrate extracting data from various sources.
- Perform data manipulations to enable analysis.
- Analyze data to derive patterns and hypotheses.
- Design data visualizations to demonstrate analyses.

CSLOs

- Design, implement and debug R programs to process data from various sources for data analysis.
- Use R-graphics to display and visualize data.

Outline

- A. Describe R basics
 - 1. What is R?
 - 2. Introduction to R and RStudio
 - 3. Installing and using R packages
 - 4. Working with R workspaces
- B. Exhibit understanding of R data objects.
 - 1. Vectors
 - 2. Matrices
 - 3. Data Frames
 - 4. Lists
 - 5. Local data import/export
- C. Illustrate basic data transformation concepts.
 - 1. Variables

2. Character and String Manipulation
 3. Dates and Timestamps
 4. Regular Expressions
 5. Control Statements
 6. Functions
- D. Demonstrate extracting data from various sources.
1. Web data capture
 2. API data sources
 3. Connecting to external data sources
 4. Data in single and distributed environments
- E. Perform data manipulations to enable analysis.
1. Using 'dplyr'
 2. Reshaping data
 3. Cleansing data
 4. Merging data
 5. Splitting data
 6. Conversion of data
- F. Analyze data to derive patterns and hypotheses.
1. Data architecture patterns
 2. Correlation clustering
 3. Predictive analysis
 4. Groupwise operations
 5. Data redundancy
 6. Descriptive statistics
 7. Regression
 8. Hypothesis testing
- G. Design data visualizations to demonstrate analyses.
1. Core concepts of data graphics and visualization
 2. R graphics engines
 - a. Base
 - b. Grid
 - c. Lattice
 - d. ggplot2
 3. Customizing graphics with 'ggplot2'
 - a. Titles
 - b. Coordinate systems
 - c. Scales
 - d. Themes
 - e. Axis labels
 - f. Legends

Lab Topics

- A. Data types and data structures
- B. Flow control and looping
- C. Writing and calling functions
- D. Split/apply/combine pattern
- E. Working with character data and regular expressions
- F. Regular expressions and web scraping
- G. Reshaping data and database access
- H. Simulation
- I. Optimization
- J. Data and predictive analysis