# Comprehensive R Programming Syllabus for Statistics Students

## Module 1: Introduction to R (Beginner Level)

* 1.1 Getting Started
* - Installing R and RStudio
* - Understanding the RStudio interface
* - Using the console vs script editor
* - Packages: Installing and loading
* 1.2 Basic Syntax and Operations
* - Data types: numeric, character, logical, factor
* - Variables and assignment
* - Arithmetic and logical operations
* - Basic functions and help system
* 1.3 Data Structures in R
* - Vectors
* - Matrices
* - Lists
* - Data frames
* - Factors
* 1.4 Importing and Exporting Data
* - Reading CSV, Excel, and text files
* - Writing data to files
* - Reading from web APIs or clipboard
* 1.5 Data Manipulation Basics
* - Indexing and subsetting
* - Filtering and sorting
* - Basic dplyr: filter(), select(), mutate(), arrange(), summarise(), group\_by()

## Module 2: Intermediate R Programming

* 2.1 Data Wrangling with tidyverse
* - Advanced dplyr
* - Data tidying with tidyr: pivot\_longer(), pivot\_wider(), separate(), unite()
* - String manipulation with stringr
* - Working with dates and times (lubridate)
* 2.2 Data Cleaning and Preparation
* - Identifying and handling missing values: is.na(), na.omit(), complete.cases()
* - Imputation techniques (mean/median/mode, forward/backward fill)
* - Outlier detection and treatment: boxplot, z-score, IQR method
* - Data type corrections: as.numeric(), as.factor(), etc.
* - Detecting and fixing data inconsistencies
* - Duplicate handling: duplicated(), distinct()
* - Standardizing names and values: janitor::clean\_names(), rename(), rename\_with()
* - Combining and splitting columns: separate(), unite(), str\_split()
* 2.3 Data Visualization with ggplot2
* - Grammar of graphics
* - Basic plots: bar, line, scatter, histogram, boxplot
* - Aesthetic mappings, facets, themes
* - Customizing plots (labels, legends, scales)
* 2.4 Programming Concepts
* - Writing functions
* - Conditional statements (if, else)
* - Loops (for, while, repeat)
* - Vectorized operations vs loops
* - Apply family: lapply(), sapply(), tapply(), mapply()
* 2.5 Working with Categorical and Continuous Data
* - Descriptive statistics
* - Frequency tables
* - Grouped summaries
* - Cross-tabulation

## Module 3: Applied Statistics with R

* 3.1 Statistical Tests
* - t-tests (one-sample, two-sample, paired)
* - ANOVA
* - Chi-square tests
* - Correlation (Pearson, Spearman)
* 3.2 Regression Analysis
* - Simple linear regression
* - Multiple linear regression
* - Logistic regression (binary)
* - Model diagnostics and assumptions
* 3.3 Non-parametric Tests
* - Wilcoxon test
* - Kruskal-Wallis
* - Mann-Whitney U test
* 3.4 Probability Distributions
* - Normal, binomial, Poisson
* - Generating random variables
* - Distribution functions: dnorm(), pnorm(), rnorm(), etc.
* - Simulation exercises
* 3.5 Sampling Techniques and Survey Analysis
* - Simple random sampling
* - Stratified sampling
* - Cluster sampling
* - Analyzing survey data with survey package

## Module 4: Advanced R for Data Science and Machine Learning

* 4.1 Advanced Data Manipulation
* - Data reshaping
* - Joins and merges
* - Working with large datasets (data.table)
* 4.2 Advanced Visualization
* - Interactive plots (plotly, shiny basics)
* - Correlation plots, heatmaps
* - Visualizing model outputs
* 4.3 Statistical Modeling and Inference
* - Model selection and stepwise regression
* - Cross-validation
* - Generalized linear models (Poisson, multinomial logistic)
* 4.4 Machine Learning with R
* - Supervised Learning: Linear/Logistic Regression, Decision Trees, Random Forests
* - Unsupervised Learning: K-means, Hierarchical Clustering, PCA
* - Model evaluation: Confusion matrix, ROC-AUC
* - ML packages: caret, mlr3, randomForest
* 4.5 Report Automation and Reproducibility
* - RMarkdown for dynamic reports
* - Creating reproducible research documents
* - Knitting to PDF, HTML, and Word

## Capstone Projects & Practice

* - Statistical data analysis project (e.g., student performance, social survey)
* - Machine learning model (e.g., predict customer churn or classify emails)
* - Shiny dashboard for interactive visualization
* - Data storytelling using RMarkdown and ggplot2

## Tools & Libraries

* - Core: tidyverse, ggplot2, dplyr, readr, lubridate
* - Statistical Analysis: car, MASS, lmtest, psych, survey
* - ML: caret, randomForest, e1071, xgboost, mlr3
* - Reporting: knitr, rmarkdown, shiny