Course Syllabus Part I
DSC 520 Statistics for Data Science

3 Credit Hours

Course Description
The R programming language and software environment is commonly used to explore all types of data. Using R, students perform statistical tests on the data. Report writing and presentation of data are introduced.

Course Prerequisites
Recommend DSC 500

Course Objectives
Students who successful complete this course should be able to:

1. Write R code to perform statistical tests on data sets.
2. Investigate R packages and libraries for research needs.
3. Evaluate and create data visualizations that are used for communicate stories from data.
4. Develop a workflow that supports reproducible results.
5. Research algorithms used for machine learning.
6. Discuss the ethical considerations with managing and reporting of data.

Grading Scale

93 – 100% = A	87 – 89% = B+	77 – 79% = C+	67 – 69% = D+
90 – 92% = A-	83 – 86% = B	73 – 76% = C	63 – 66% = D
	80 – 82% = B-	70 – 72% = C-	60 – 62% = D-
			0 – 59% = F

Topic Outline

I. R Programming Environment
   A. Integrated Development Environment (IDE)
   B. Setting up R studio environment
   C. Essential R programming terminology and syntax

II. Statistics for Data Science
   A. Examples of everyday uses
   B. Domain specific uses
   C. Emerging and future potential
   D. Report writing and presentation of data

III. R Programming
   A. Data types
   B. Control structures
   C. Operators
   D. Debugging and code quality assurance

IV. Collecting Data for Statistical Uses
   A. Importing and using data
   B. Handling missing data
   C. Developing a workflow
   D. Reshaping data

V. Built-In and Open Source Packages & Datasets
   A. Importing and using R packages
   B. Using built-in datasets
   C. Documenting and managing R environment

VI. Statistical Tests on Data
   A. Descriptive Statistics
   B. Correlation
   C. Distributions
   D. Frequencies
   E. Clustering
   F. Time-series
   G. Linear Regression

VII. Machine Learning Algorithms
   A. Cluster analysis
   B. Unsupervised learning
   C. Supervised learning