# IT-2090: DATA ANALYTICS PROGRAMMING

## Cuyahoga Community College

### Viewing: IT-2090 : Data Analytics Programming

**Board of Trustees:**

May 2023

**Academic Term:**

Fall 2023

**Subject Code**

IT - Information Technology

**Course Number:**

2090

**Title:**

Data Analytics Programming

**Catalog Description:**

This course covers the fundamental concepts of R and the use of R for effective data analysis. Students will develop skills to develop solutions to complex problems across a variety of disciplines using data and real-world case studies.

**Credit Hour(s):**

4

**Lecture Hour(s):**

3

**Lab Hour(s):**

2

## Requisites

**Prerequisite and Corequisite**

IT-1025 Concepts for Programmers, IT-1050 Programming Logic, and IT-2070 Introduction to Data Science and Analytics.

## Outcomes

**Course Outcome(s):**

Install and demonstrate use of data analytics tool to import, graph, wrangle and tidy data for effective data transformation.

**Objective(s):**

1. Define the taxonomy for understanding well-designed data graphics and the importance of patterns in visualizing data.
2. Explain and demonstrate the process of wrangling data.
3. Explain and demonstrate the process of data tidying.

**Course Outcome(s):**

Use statistical methods and models to quantify patterns and their strength and provide meaning from data.

**Objective(s):**

1. Practice key statistical methodologies to connect samples, data and populations.
2. Define and apply supervised and unsupervised learning models to explain the relationship between variables.
3. Create simulation models to create data from speculation.
4. Explain the concept of interactive data graphics as a means to data understanding.

**Course Outcome(s):**

Apply data modeling techniques to heterogeneous data.

**Essential Learning Outcome Mapping:**

Critical/Creative Thinking: Analyze, evaluate, and synthesize information in order to consider problems/ideas and transform them in innovative or imaginative ways.

**Objective(s):**

1. Discuss and demonstrate how regular expressions can be used to process data.
2. Demonstrate methods to create efficient databases to optimize query speeds.
3. Build and interpret models that include spatial data.

---

**Methods of Evaluation:**

a. Participation and discussion
b. Online research
c. Written reports
d. Demonstrations
e. Projects
f. Tests
g. Quizzes

**Course Content Outline:**

a. Data Science overview
    i. Datasets and course example files
    ii. Professional ethics
    iii. Installation and overview of statistical tool
    iv. Objects and packages
    v. Importing and exporting data
    vi. Data frames
b. Data visualization
    i. Variable relationships
    ii. Taxonomy of data graphics
    iii. Scatter plots, histograms and cartesian plots
    iv. Coordinate systems
    v. Scale, guides, facets, layers
    vi. Canonical data graphics in R
    vii. Univariate and Multivariate displays
c. Data wrangling
    i. Functions (select), (filter), (mutate), (arrange), (summarize)
    ii. Joins
d. Tidy data
    i. Cases and variables
    ii. Categorical and quantitative variables
    iii. gather(), spread(), and apply() functions
    iv. Iteration over subgroups
    v. Data cleansing
e. Statistics and modeling
    i. Vectors, matricies, and arrays
    i. Lists
    ii. Sampling distribution concepts
    iii. Bootstrap method
    iv. Outliers
    v. p-values
    vi. Predective analytics
    vii. ANOVA
    viii. Regression and multiple regression
    ix. Bayes theorem
    x. Artifical neural networks
    xi. Evaluating models

      xii.  Measuring prediction error
     xiii.  Unsupervised learning
          1.  Clustering
          2.  k-means
          3.  Dimension reduction
 f.  Simulation
      i.  Randomizing functions
     ii.  Simulating variability
    iii.  Random networks
g.  Report generation
      i.  PDF
     ii.  HTML
h.  Creating Web applications

## Resources

Mccoy, Scott. *R for Data Analytics*. Mike Murach & Associates Inc., 2023.

---

Juretic, Francisco. *R Statistics Cookbook: Over 100 recipies for performing complex statistical operations with R 3.5*. Packt Publishing, 2019.

---

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., Lichtendahl, K. C. *Data mining for business analytics: concepts, techniques, and applications in R.* Hoboken, NJ: John Wiley Sons, 2018.

---

Saltz, J. S., Stanton, J. M. (2018) *An introduction to data science,* Thousand Oaks, CA: SAGE Publications, Inc.

---

Lander, J. P. (2017) *R for everyone: advanced analytics and graphics,* Boston: Addison-Wesley.

---

## Resources Other

 a.  R tutorial at w3schools: https://www.w3schools.com/r/
 b.  R Tutorial for Beginners: Learn R Programming Language: https://www.guru99.com/r-tutorial.html

Top of page
Key: 2497