

## Syllabus: Clustering and Finding Patterns

This course is designed for students with a basic familiarity with R and some experience with data analysis and data manipulation. In just 100 minutes of instructional time, students learn how to think about finding patterns, perform different types of clustering analyses, and evaluate the quality of the results.

By the end of this course, students will be able to:

1. Find latent patterns and groups in different types of data
2. Evaluate the accuracy and effectiveness of clustering analyses
3. Understand the purpose and implications of what clustering methods can and cannot achieve

Assessment:

1. **Concept reviews:** these are comprised of short five question quizzes that cover the most important concepts and ideas in each lesson. They encourage holistic understanding and are multi-faceted question types (i.e. drag and drop, fill-in-the-blanks, matching, etc).
2. **Exercises:** these are additional videos that cover the coding functions in the instructional video in more depth. They are project-based and include coding templates for students to strengthen their skills outside of the course.

Materials provided:

1. Accompanying PDFs to use as reference materials
2. R code templates from the instructional videos and exercises
3. Data sets used in the instructional videos and exercises

## Course Outline

1. Introduction to clustering: 30 min
  - a) The impact of data
  - b) Where is data today?
  - c) What is clustering?
  
2. Clustering politicians: 30 min
  - a) Clustering Congress
  - b) Visualizing Congress
  - c) Evaluating accuracy
  
3. Clustering athletes: 15 min
  - a) Clustering basketball salaries
  - b) Evaluating results
  
4. Clustering customers: 27 min
  - a) Finding customer patterns
  - b) Digging into our data
  - c) Calculating the best number of clusters
  - d) Additional tips and resources

*Total instructional time:* 1 hr, 42 min