ACM Data Science Task Force Course Example

MTH 365: Introduction to Data Science
Creighton University (Omaha, NE)
Aimee Schwab-McCoy
aimeeschwab-mccoy@creighton.edu

https://github.com/aimeeschwab-mccoy/MTH365

A Creative Commons - Attribution-ShareAlike License is assumed..
For details on that license see: https://creativecommons.org/licenses/by-sa/4.0/

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	Total Number of Contact Hours
AP	9
DAMG	12
DM/ML/DP	15

Where does the course fit in your undergraduate Data Science curriculum?
This course is required for our major and minor in Data Science. Students typically take it as sophomores. This course also attracts upper-level students from other departments who need to get some data science expertise for their research.

Is this course from or used in other curricula/majors?
This course was explicitly designed to support the Data Science minor, which later expanded to a full major.

What is covered in the course?
MTH 365: Introduction to Data Science covers (1) data wrangling and formatting, (2) exploratory data analysis and data visualization, (3) data acquisition and management, (3) statistical modelling and inference, (4) unsupervised machine learning, (5) text mining, and (6) spatial data visualization. The course also introduces principles of reproducible research and dynamic programming using the R/RStudio and tidyverse ecosystem.

What is the format of the course?
The course is taught face-to-face in a half lecture/half lab format. In most weeks, Day 1 is a “lecture” day and Day 2 is a “lab” day.

How are students assessed?
Students complete weekly lab assignments and several mini-projects throughout the course of the semester.

Course tools and materials
The primary textbook is *Modern Data Science with R (https://beanumber.github.io/mdsr2e/)*. Other resources include *R for Data Science* by Hadley Wickham.

Why do you teach the course this way?
Data science in particular is impossible to learn without *doing*. Emphasizing live coding and assigning collaborative labs during class time encourages students to work together, make mistakes, and develop the resilience to keep going during a difficult task.

Body of Knowledge coverage

KA	Sub-domain	Competencies Covered	Hours
AP	AP-Foundational considerations and AP-Visualization	Creating and customizing static data visualizations using the tidyverse and interactive data visualizations using RShiny	9
DAMG	DG-Data Acquisition	Data sources, web scraping	3
DAMG	DG-Working with Various Types of Data	Text mining	3
DAMG	DG-Data Cleaning	Data wrangling and cleaning; data management and storage	6
DM/ML	DM-Cluster Analysis ML-Unsupervised Learning	Clustering algorithms, dimension reduction	6
DM/ML	DM-Classification and Regression ML-Supervised Learning	Classification methods, regression	6
DP	AS-Machine Learning Robustness and Explainability	Interpretable machine learning and algorithmic bias	3