

## Syllabus

**Instructor:** Brian Steele      Office: Math 314  
Phone: 243-5396      email: brian.steele@umontana.edu

**Course Format:** One meeting per week: Wednesday 2:10-5:00 PM, Math 305.

**Course website:** <https://moodle.umt.edu/>

**Office hours:** Tuesdays 1-2:30PM, Thursdays 12-2PM (subject to change). Also by appointment.

**Course Objectives:** Learn about methods for the analysis of big data. Gain experience in applying these methods.

**Course Content:** The focus of the course is on three topic areas that dominate big data analytics: high volume or massively large data sets, streaming data, and predictive analytics.

The first portion of the course is aimed at introducing mathematical and programming foundations. The foundations cover computational methods for processing big data and the programming foundation is `Python`. Later, we will work with `R` and the `R` data visualization package `ggplot2`.

The second portion of the course develops specific tools for big data; specifically, `hadoop/map reduce` for massively large data sets, and `Python` for streaming data. The last portion of the course focuses on predictive analytics.

**Learning Outcomes:**

1. Learn mathematical, statistical and computational principles used for solving big data problems.
2. Gain knowledge of `R`, `python`, and `hadoop/map reduce` and the Linux operating environment.
3. Gain practical experience in working with big data.

**Prerequisites:** STAT 341, and one of M 221 or M 273, or consent of instructor.

**Textbook:** None. Electronic resources will be posted on the course webpage.

**Organization:** This course is lab course, and course grades are based entirely on lab work. The three-hour time block will be split between lecture and lab time. During the lab time students will be engaged in tutorials and homework exercises.

**Attendance:** Class attendance is critical. Please inform me by email if you will or have missed a class and provide an explanation of your absence.

**Grading:** Your course grade will be based on the successful completion of the tutorials and homework assignments. There are 12 mandatory tutorials, each with a point value of 60 for a total of 720. The remainder of the points, 280, is homework. Tutorials are due on the last day of each month. You are to complete

4 each month from the material covered during the month. There are roughly 15 possible tutorials each month. There will be exercises due every week or two, usually short questions. These must be completed for full points.

The grading scale for this course is tentative; however, the table below shows the probable scale for the assignment of course letter grades:

Grade	A- to A+	B- to B+	C- to C+	D- to D+	F	CR (Credit)
Point total	950 or more	825-949	680-824	570-679	Less than 570	570 or more

### **Schedule of Topics and Labs:**

1. Week 1: Data Mappings and Data Dictionaries.
2. Week 2: Data Mappings and Data Dictionaries, con't.
3. Week 3: Computational Methods for Massive Data.
4. Week 4: Computational Methods for Massive Data, con't.
5. Week 5: Data Analytics and Visualization
6. Week 6: Data Analytics and Visualization, con't.
7. Week 7: Distributed computing.
8. Week 8: Distributed computing, con't.
9. Week 9: Streaming Data.
10. Week 10: Streaming Data, cont'd.
11. Week 11: Predictive analytics:  $k$ -nearest neighbor methods.
12. Week 12: Predictive analytics: Naive Bayes.
13. Week 13: Cluster Analysis.
14. Week 14: Project.

**Exams:** There will be no exams during the semester. The final exam 1-3, December 17 (Thursday).

**Incomplete (I) Grades:** Incompletes (I's) are given at the discretion of the instructor. See the 2009-2010 UM catalog for the conditions under which an incomplete may be given. Incompletes will not be given under any other circumstances.

**Adding/dropping the course:** Deadlines for withdrawing from course are listed at <http://www.umt.edu/registrar/calendar.aspx>.

**Academic Misconduct:** All students must practice academic honesty. Academic misconduct is subject to an academic penalty by the course instructor and/or a disciplinary sanction by the University. All students need to be familiar with the Student Conduct Code. The Code is available for review online at <http://life.umt.edu/sa/documents/fromWeb/StudentConductCode1.pdf>.