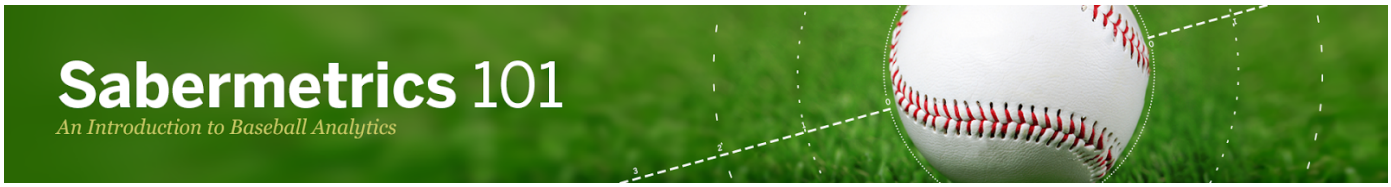


## Course Team

**Andy Andres** is a Senior Lecturer of Natural Sciences and Mathematics at Boston University College of General Studies, teaching introductory physics and biology for the past fourteen years. He is the Lead Instructor and Head Coach of the MIT Science of Baseball Program, and since 2004 has taught one of the first ever college courses in Sabermetrics at Tufts University. For seventeen years, Andres also taught a Biology Tutorial/seminar in Exercise Science and Physiology to Harvard College Biology majors. He is Chairman of the Educational Resources Committee for the Society of American Baseball Research (SABR), and winner of the 2012 *USA Today* Sports Weekly Award for his research for the Society. Andres has updated and edited the second edition of Art McGee's classic book on fantasy player valuation, *How to Value Players for Rotisserie Baseball*. A datacaster/stringer for mlb.com at Fenway Park who scored Game 6 of the 2013 World Series, he has a PhD in Nutritional Biochemistry and Physiology from Tufts University.

He has taught this course at the Experimental College at Tufts University since 2004. Many of its former students have gone on to careers writing about baseball and working in various MLB baseball operations and analytics departments. Professor Andres talks about the science of baseball in his TEDx talk: The Science of the Home Run, TedXYouth@Beacon Street.

The edX version of this course is made possible by the diligent course coordinating efforts of Gabe Gralla & Morris Greenberg, and the support of the Boston University Digital Learning Initiative.

## What Is This Course About?

Sabermetrics 101: Introduction to Baseball Analytics (#SABR101x) is an introduction to sabermetrics, baseball analytics, data science, the R Language, and SQL. It covers the theory and the fundamentals of the emerging science of Sabermetrics. We will discuss the game of baseball, not through consensus or a fan's conventional wisdom, but by searching for objective knowledge in hitting performance. Other areas of sabermetrics will be analyzed and better understood with current and historical baseball data. The course also serves as applied introduction to the basics of data science, a growing field of scholarship, that requires skills in computation, statistics, and communicating results of analyses. Using baseball data, the basics of descriptive statistics, the R Language, and SQL will be covered.

## SQL Sandbox

One of the most important features of Sabermetrics 101 is the new Boston University SQL Sandbox, a place where new SQL learners can explore the SQL language and important baseball

databases. We think it is the easiest way for digital learners to experience the hands on SQL trial and error effort it takes to become proficient in the SQL language.

## Course Requirements

This course requires no prior knowledge or experience. As described below, there are six modules of course content covering four “tracks” (sabermetrics, statistics, computer science, and history of sabermetrics).

## Course Grading

Problem Sets (SQL/R): 50%
Lecture Questions, Quizzes (Quick "comprehension" questions): 30%
Final Exam: 20%

To obtain a Course Certificate, you need to correctly answer 60% of all of the graded material.

## Course Reminders

A note about time references: Time will be reported by course staff as Eastern Daylight Time, North America (EDT). Any times listed by edX, such as due dates listed on the course site, will be reported in Coordinated Universal Time (UTC). The course staff will make every effort to make times and time zones as clear as possible. There are various Time Zone converters on the web, one good one is http://www.timeanddate.com/worldclock/converter.html.

Please review the edX Terms of Service (TOS, https://www.edx.org/edx-terms-service) especially the Honor Code which we reproduce for you here:

COLLABORATION POLICY

By enrolling in a course on edX, you are joining a special worldwide community of learners who have agreed to the Honor Code below.

EDX HONOR CODE PLEDGE

By enrolling in an edX course, you have agreed to:

- Complete all mid-terms and final exams with your own work and only your own work. You will not submit the work of any other person.
- Maintain only one user account and not let anyone else use your username and/or password.
- Not engage in any activity that would dishonestly improve your results, or improve or hurt the results of others.
- Not post answers to problems that are being used to assess student performance.

Unless otherwise indicated, learners on edX are encouraged to:

- Collaborate with others on the course videos, exercises, homework and labs.
- Discuss general concepts and materials in each course with others..
- Present ideas and written work to fellow edX learners or others for comment or criticism.

## Course Schedule

| Weekly Schedule | Content |
|---|---|
| **Module 1: Course Introduction**<br><br>Release Date:<br>29 May 2014<br>11am EDT | Defining terms: Sabermetrics, the frameworks that form the disciplines of statistics & data science; relational databases; SQL SELECT; Henry Chadwick |
| **Module 2: Measuring Hitting**<br><br>Release Date:<br>5 June 2014<br>11am EDT | Evaluating batting performance; measures of central tendency, spread & variability; the shape of data & basic distribution; basic SQL queries; Hugh Fullerton |
| **Module 3: Advanced SQL**<br><br>Release Date:<br>12 June 2014<br>11am EDT | The relationship between the number of runs a team scores, the number they allow, and the number of games that they will win; Run Estimators; basics of bivariate analysis; basics of correlation & dependence; SQL JOINS; F.C. Lane |
| **Module 4: R and RStudio**<br><br>Release Date:<br>19 June 2014<br>11am EDT | R Language & RStudio; sabermetric hitting statistics & key situational hitting metrics; "replacement level" & statistical distributions; Alan Roth and Branch Rickey |
| **Module 5: More R Code for Sabermetrics**<br><br>Release Date:<br>26 June 2014<br>11am EDT | More R Language; more important sabermetric hitting statistics & a few key advanced metrics using location and pitch data; regression to the mean; Earnshaw Cook |
| **Module 6: Putting It All Together**<br><br>Release Date: | How to adjust for park factors; how to adjust for league run scoring environments; more important sabermetric statistics & a few key advanced metrics |

| | |
|---|---|
| 10 July 2014<br>11am EDT | for overall value of baseball players; basics of sample size; George Lindsay |

All course assignments, quizzes, and problem sets are due on the end of the day on the 17th of July (at 12:00 midnight EDT). The Final Exam will be due the 21st of July (at 12:00 midnight EDT). Certificates of course completion will be distributed on the 23rd of July.