The MD in .Rmd: Teaching Clinicians Data Analytics with R

by Ted Laderas and Brian Sikora

Abstract Data Analytics in Healthcare is difficult to teach effectively, especially to clinical professionals. In this article, we outline our Data Analytics course and our strategies for effective learning of both practical data science skills (SQL and modeling in R) and strategies for effectively presenting analyses within an healthcare organization using a readmission metric. We highlight the creative ways our students present their results and the impact the course has made on them.

Introduction

Delivering analytics effectively in healthcare is a difficult concept to teach effectively. Key practical data science skills (such as querying, visualizing, and modeling real data) must be combined with organizational thinking. In this article, we outline our Data Analytics course which focuses two aspects of analytics: organizational considerations and practical R skills.

This course is unique in that it has both clinical informatics and bioinformatics students within our departments. We take advantage of this mixed audience to build a support structure for both the clinicians and the bioinformatics students.

By the end of the course, our students must synthesize both of these branches by querying and modeling the data and present their results to an Executive team.

Clinicians as Learners

In our past 6 years teaching the course, we have honed our understanding of clinicians as learners. We have built our learner persona as Mary:

> Mary is a clinician who wants to understand how analytics can be delivered in her healthcare organization. She *has little time*, and likes *learning on her own*. She has a *hard time asking for help* and can be overly self critical.

We have tailored the material in the course to accommodate the learners [in the following ways]:

Mary has little time - our assignments utilize a lot of "just in time" instruction. For instance, each assignment focuses on a key concept that builds on previous concepts, such as single table queries first, followed by table joins, and self-joins. Additionally, the assignments slowly increase in difficulty.

Mary likes learning on her own - all assignments are delivered as RMarkdown documents within Rstudio projects and distributed via RStudio.cloud. The ease of setup with RStudio.cloud gets our students up and running relatively quickly. Delivering individual assignments as projects allows us to pace the tempo of learning, and focus them on particular topics.

Mary has a hard time asking for help. We attempt to destigmatize asking for help with the following strategies: team-based learning, office hours, and individual appointments.

Simulating Data Analytics

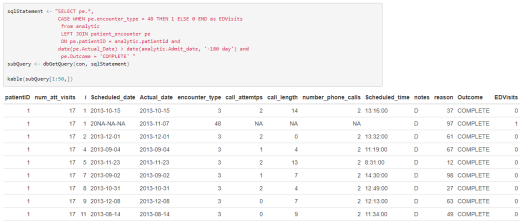
Our course centers around a single analytical task: predicting readmissions in a simulated patient population using a validated metric, LACE. [] LACE is short for

- **L**ength of Stay
- **Acuity** of Admission
- **Comorbidities**
- **Emergency Room** Visits

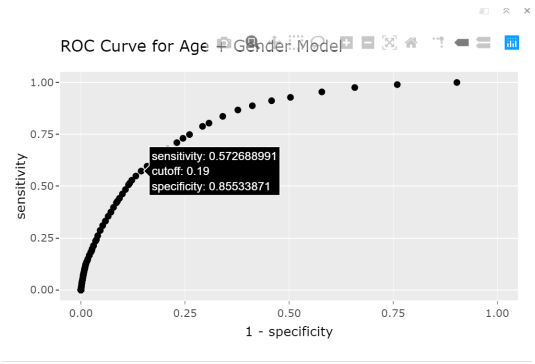
In order to calculate these, a number of assignments center around exploring, extracting, querying, summarizing, modeling, and presenting the patient information (Figure 1).

The expectation is that students not only utilize their practical skills, but make it understandable to an executive team. Our assignments center around querying and calculating the LACE metric for a simulated clinical warehouse. This clinical warehouse is stored as a SQLite Database and is accessed through the R Projects using the `DBI` and `RSQLite` packages. It represents a clinical extract of patients admitted into the hospital, along with their diagnoses.

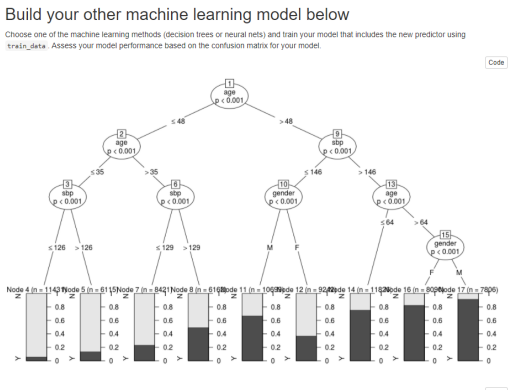
(a) Exploratory Data Analysis



(b) SQL Queries



(c) Logistic Regression



(d) Machine Learning

Figure 1: Analytics Assignments, in order of difficulty

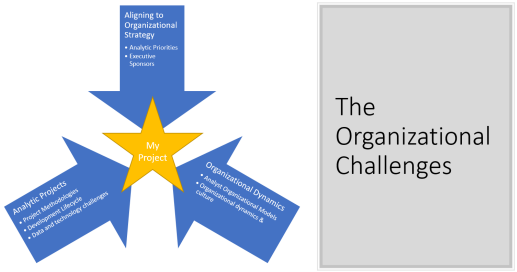


Figure 2: Organizational Challenges in delivering Data Analytics

The assignments progress in order of difficulty:

Exploratory Data Analysis: Use tools such as `visdat` and `skimr` to explore the structure of the tables in the data warehouse and missing values. (Figure 1a)

SQL Queries: start with simple one table queries, graduate to multi join queries, including self-joins and counts (Figure 1b)

Calculating individual components: calculate and document how to calculate each component of the LACE Score. These assignments require multi table queries, including querying the `patient`, `patient_encounter_hosp`, and `patient_diagnosis` tables within the database to aggregate and summarize these LACE components for each patient.

Modeling: build a logistic regression model or a machine learning model to predict readmission in the simulated patient population (Figure 1c).

Visualization: use `ggplot2` and other `tidyverse` packages to visualize outcomes and association with covariates.

Organizational Skills and Challenges

Equally important skills for Analysts are effectively aligning analytic projects to the strategic goals within an organization (Figure 2). To this end, we provide lectures in organizational behavior topics such as finding organizational sponsors, establishing analytic priorities, understanding organizational dynamics and culture, change management in an organization, evaluating the clinical utility of a metric, as well as the lifecycle of analytic projects in an organization. These lectures are supplemented by reading assignments [] and online and in-class discussion of these topics.

The organizational skills lectures are delivered by guest lecturers from Kaiser Permanente, who share their own experiences of working as an effective analytic team within Kaiser Permanente. Most importantly, the experience of implementing LACE as a useful metric within KP is highlighted.

In particular, the first implementation of LACE at KP was a failure, and students get the opportunity to learn about why the first implementation was not successful. The second successful implementation of LACE highlights essential change management skills needed to make LACE a success within KP as an organization.

Students have found these lectures to be helpful in understanding the context of delivering their analyses.

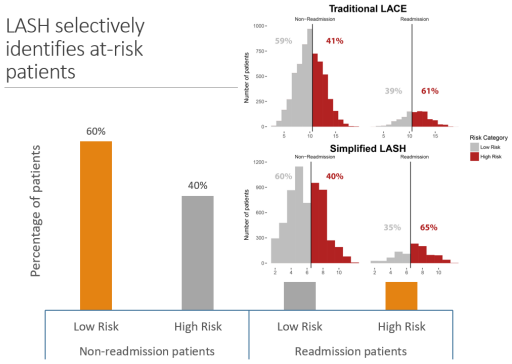
Final Presentation

The final presentations consists presenting their modeling results and findings to a healthcare organization. The instructors role play as executive sponsors within a fictional healthcare organization. The students are expected to interpret and present a model meaningfully in a way that aligns with organizational goals. The presentation needs to be pitched in a way that is accessible to executives.

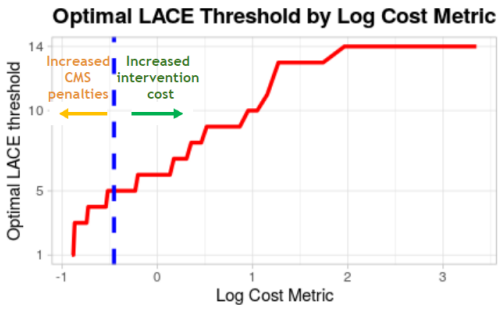
In order to facilitate this, we have developed lectures both in organizational considerations in presenting data, data storytelling with visuals, and provided a template for effectively presenting the data.

Outcomes: Creativity in Presentations

Based on our template, student teams have come up with creative ways of presenting their models to the executive team (Figure 3).



(a) Enhanced version of LACE



(b) Selecting a threshold using a Cost Metric

Potential cost savings (test data)

Sum	Number of patients	Cost/Savings per patient	Description
+ $6.7 million	608	+ $11,100	Actual Readmissions Identified by the model (~1% Medicare penalty + hospital costs)
- $2.9 million	2882	- $1000	Patients Flagged by model for Interventions
+ $3.8 million SAVINGS (6902 patients)			

(c) Summarizing the cost/savings of using LACE

Figure 3: Final presentations of LACE

Outcomes: Testimonies

Students have been highly receptive to the course. In a recent curriculum committee meeting, both the Clinical and Bioinformatics student representatives recommended that this course be taught to all students in the department.

Interviews with students about the course have featured the following themes:

Collaboration:

> Taking the Data Analytics course made me a **much more patient and effective collaborator**, especially when working with colleagues outside of science.

— Kristen Stevens, MD/PhD Candidate

Diversity:

> I highly recommend this course to anyone who wishes to get a comprehensive introduction to R and the field of data analytics. .b[The course attracts [a] very diverse set of students.] The hybrid nature of the course was ideal to get to meet and network with others.

— Meenakashi Mishra, Clinical Informatics Fellow

Comprehensive view of the field:

> I would definitely recommend this to .b[anyone who is interested in working with data in a healthcare setting], whether you're a clinician/researcher who will be gathering and using the data, or a manager who might be presenting the data or incorporating analytics into your organization's workflow.

— Pierrette Lo, Biomedical Project Manager/Data Scientist

The Future:

> As a clinician, to see how a data analysis works and how algorithms result in a recommendation was very helpful. .... informatics programs **should expand their technical analytics courses for clinicians and other clinical informatics students**. Perhaps some day there will be a formal role for clinician-data scientists in the healthcare industry outside of research, and if there will be it starts with courses like this.

— Frank Longano, MD, Clinician

Conclusions

Over six years of giving this course to both clinicians and bioinformaticians, we have honed the art of conducting analysis and making it accessible and actionable within a healthcare organization. Our two-pronged approach of teaching both practical data science skills and organizational skills

Availability

All code for the individual RStudio Projects are available here: https://github.com/laderast/AnalyticsCourse

Ted Laderas
Medical Informatics and Clinical Epidemiology, Oregon Health & Science University
3181 S.W. Sam Jackson Park Road, BICC
Portland, OR 97239, United States of America
https://laderast.github.io
ORCiD: 0000-0002-9079-593X
tedladeras@gmail.com

Brian Sikora
Kaiser Permanente Insight
line 1 affiliation 1
line 2 affiliation 1
https://journal.r-project.org
ORCiD: 0000-0002-9079-593X
author2@work