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

by Ted Laderas and Brian Sikora

Abstract Teaching Data Analytics in Healthcare is difficult to teach effectively, especially to Clinical Professionals. In this article, we outline our Data Analytics course, which we have delivered for 6 years, and our strategies for effective learning.

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]:

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.

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.

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

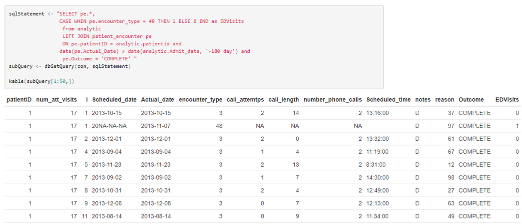
- **L**ength of Stay
- **Acuity** of Admission
- **Comorbidities**
- **Emergency Room** Visits

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

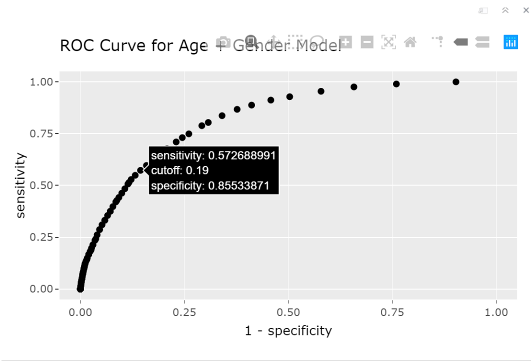
The expectation is that students not only utilize their practical skills, but make it understandable to an executive team. Our assignments center around querying and calculating the LACE metric for a simulated clinical warehouse. This clinical warehouse is stored as a SQLite Database and is accessed 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.

The assignments progress in order of difficulty:

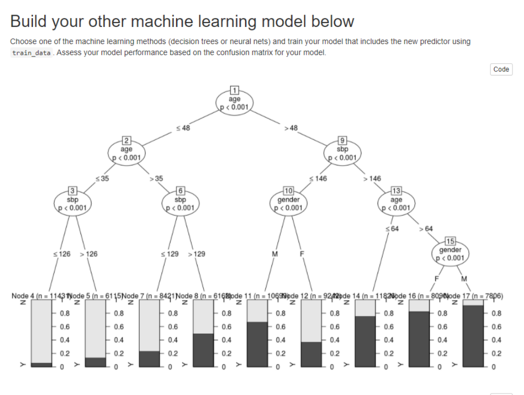
(a) Exploratory Data Analysis



(b) SQL Queries



(c) Logistic Regression



(d) Machine Learning

Figure 1: Analytics Assignments, in order of difficulty

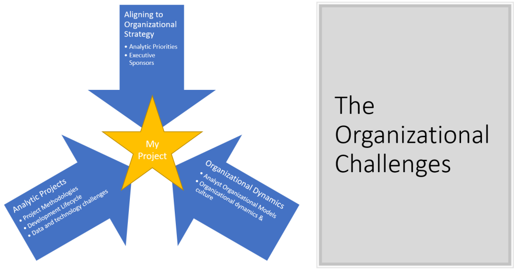


Figure 2: Final presentations of LACE

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 @ref(fig:assignments)a)

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

Calculating individual components: calculate and document how to calculate each component of the LACE Score.

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

Presenting/Interpretation: interpret a model meaningfully in a way that aligns with organizational goals in a way that is accessible to executives. In order to facilitate this, we have developed lectures both in organizational considerations in presenting data, and a section on data storytelling with visuals.

Organizational Skills and Challenges

Equally important skills for Analysts are effectively aligning analytic projects to the strategic goals within an organization (Figure @ref(fig:organization)). To this end, we provide lectures in 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 lectures are delivered by guest lecturers from Kaiser Permanente, who share their 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. The first implementation was a failure, and students get the opportunity to learn about why the implementation was not successful. The second implementation of LACE highlights the essential change management skills needed to make LACE a success within Kaiser. Students have found these lectures to be helpful in understanding the context.

```
include_graphics(here("image/org_challenges.png"))
```

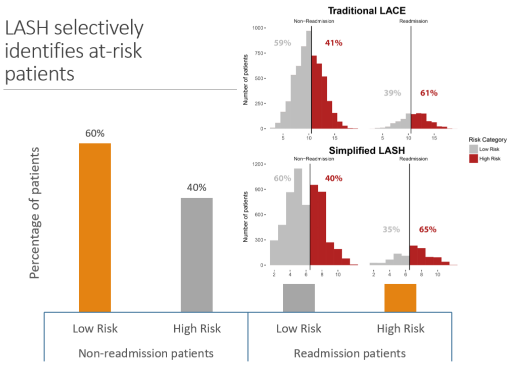
Final Presentation

We emphasize the need for presenting their analyses

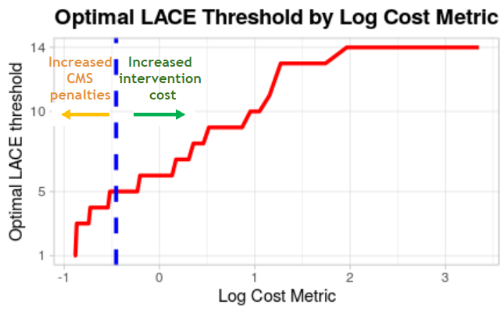
Outcomes: Creativity in Presentations

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.



(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

Conclusions

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