

DATA-LA 485

Statistical Modeling in R

Spring 2021

Tuesdays and Thursdays 12-1:15pm
Bronco Gym RM 218

Lantz McGinnis-Brown

Office: Online

Office Hours: After class or by appointment

Email: LantzBrown@boisestate.edu

Phone: (208) 386-0056

Welcome!

Welcome to DATA-LA 485! As your professor, I value your opinions and feedback on how to make this course better for you and others, so please feel free to communicate with me if there are ways I can make adjustments to allow for more an effective learning environment.

Course Description: Focuses on statistical methods for practical data analysis, including parametric and non-parametric analyses, ANOVA, multiple and logistic regression, generalized linear models, and dimension reduction methods using R to examine and understand human behavior. Students will conduct a research project designed in partnership with a professional stakeholder that delivers actionable outcomes. A research methods course is strongly recommended.

Course Learning Outcomes:

After successful completion of this course, students will be able to:

- 1 Utilize R statistical software to analyze data
- 2 Manage, manipulate and “wrangle” data
- 3 Develop deliverables (e.g. report or presentation) for research project(s)
- 4 Describe the ethical concerns related to the misuse of data, data privacy, and inclusivity in data science algorithms.
- 5 Execute the data analysis process to answer a research question and communicate results.

Course Format:

This class provides a *service-learning opportunity* which gives students hands-on experience applying what they learn in the classroom to issues in the community, while gaining skills in the data analysis process. Students will work with data collected from the Idaho Policy Institute’s statewide survey to develop insights into practical problems facing the state of Idaho. Furthermore, students will have the opportunity to participate in a “data science for social good” competition, through the DrivenData organization.

Service learning activities will include: reading background literature to understand the context of the problem (literature review), reviewing the data to identify questions related to the project (data

mapping), developing a data analysis plan (scope of work), executing the plan, and communicating results.

We will engage in two project-related reflection activities – one pre-project, and one post-project, in addition to regular in-class discussions about challenges and learning opportunities. If you feel that you cannot participate in this service learning opportunity, please let me know and we can arrange for an alternative assignment.

This course may include other group projects, guest lectures, lectures, and homework assignments.

Hybrid Class Format

This class is in a hybrid format, which typically means that we are going to be alternating between in-person and online meetings and learning methods. However, given the unique circumstances that the coronavirus pandemic presents, we will discuss everyone's comfort with in-person meetings as a class, and make decisions accordingly. The first two weeks of class and the weeks after spring break will be entirely online.

Accountability to community:

Please remember that, in service-learning, you represent yourself, our class and Boise State University, and people are counting on you and appreciating your presence. Integrity, at all levels, is demanded. If, for any reason, you cannot meet a particular commitment for your service-learning, please communicate that to all involved parties immediately.

Prerequisites:

This course requires DATA-LA 322 as a prerequisite. This previous experience gives you an overview of coding in R and an introduction to data visualization, statistical data analysis, and machine learning algorithms. If you'd like to review the textbook used in this course, it can be found here:

Grolemund, G. & Wickham, H. (2017) *R for Data Science: Visualize, Model, Transform, Tidy, and Import Data*. Available at <https://r4ds.had.co.nz/>

Texts:

James, G., D. Witten, T. Hastie, & R. Tibshirani (2017) [An Introduction to Statistical Learning: With Applications in R](#).

Navarro, Danielle (2016) Learning statistics with R: A tutorial for psychology students and other beginners. <https://learningstatisticswithr.com/lsr-0.6.pdf>

Johnson, Alicia, Ott, Miles, & Dogucu, Mine (2020). [Bayes Rules! An Introduction to Bayesian Modeling with R](#)

Other Resources Available on-line:

Irizarry, Rafael A. (2019) *Introduction to Data Science: Data Analysis and Prediction Algorithms with R*. Available at: Other <https://rafalab.github.io/dsbook/>

Software:

R Statistical Software: <https://www.r-project.org/>

R Studio: <https://www.rstudio.com/products/rstudio/download/>

Supplemental Resources:

RStudio's 'cheatsheets': <https://rstudio.com/resources/cheatsheets/>

Course Website:

We will be using Blackboard, a course management system which provides a website for me to post lecture materials, homeworks, readings, etc. All assignments will be posted to blackboard and you can submit your assignments via blackboard as well. You can find access to the site at: <https://blackboard.boisestate.edu>

Course Requirements: The course will consist of a mix of lectures by the instructor and in class activities. Grades will be based on a combination of: attendance/in-class activities, homework, and exams.

- 1 Attendance & in-class activities:** You are required to attend class.
- 2 Homework:** You will have homework assignments. Details will be provided in class. You are welcome to come to office hours or email me with specific questions on the homework, but I will not *check* your homework before it is due.
- 3 Projects:** Projects will consist of a data set that you need to examine to produce insight into some kind of data question.

Grading:

Attendance:	10%	*One absence allowed without penalty
Homeworks / Reflections:	20%	
Project 1 (DrivenData Practice Competition)	15%	
Project 2 (IPI Statewide Survey Analysis)	20%	
Project 3 (Food Bank Usage Analysis)	15%	
Project 4 (Choose your own)	20%	
	100%	

Students needing accommodation:

Students needing accommodations to fully participate in this class should contact the Educational Access Center (EAC). All accommodations must be approved through the EAC prior to being implemented. To learn more about the accommodation process, visit the EAC's website at <https://eac.boisestate.edu/new-eac-students/>.

Resources for Harassment:

Title IX makes it clear that violence and harassment based on sex and gender (including violence and harassment based on sexual orientation and gender identity) are civil rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, etc. If you or someone you know has been harassed or assaulted, you have resources on campus ready to help. Please refer to Boise State's Title IX compliance policy here: <https://www.boisestate.edu/compliance/title-ix/>.

You have a right to feel safe on campus!

Many at Boise State are deeply committed to the safety, health, and happiness of every student. If you feel you need help, please reach out. Additional support can be found through the Gender Equity Center: <https://www.boisestate.edu/genderequity/>.

Student Conduct and Academic Integrity:

In order to create a safe space for learning, I expect all of us to exhibit behavior that reflects Boise State's Statement of Shared Values (<http://deanofstudents.boisestate.edu/statement-of-shared-values/>) and is characterized by

- Academic Excellence
- Caring
- Citizenship
- Fairness
- Respect
- Responsibility
- Trustworthiness

In addition, students in this course are expected to uphold standards outlined in the Boise State University Student Code of Conduct (<http://deanofstudents.boisestate.edu/student-code-of-conduct/>). Any work submitted by a student in this course for academic credit will be the student's own work. Note: While I encourage students to help each other on homework, if you submit the same document as another student, you will receive a 0%. Some projects may be group projects. In this case, I expect only 1 submission to be made on behalf of the group.

Tentative Course Outline & Reading List:

Please note that this list will be updated throughout the term depending on the amount of material we complete each class.

Week	Date	Class Topic	Reading	Homework Due (will provide more details in class)
Week 1	Jan 12	Introduction to Course; Overview of Service-Learning		Nothing due for class #1
	Jan 14	Ethics in Data Science	Watch Weapons of Math Destruction & Rethinking Research Data videos	Find (and read) an article related to Ethics in Data Science to discuss
Week 2	Jan 19	How to complete data analysis projects?, Overview of basic statistics in R	Ch.12-13 in Learning Statistics with R	
	Jan 21	Overview of basic statistics in R, cont'd.	Ch. 14 & 16 in Learning Statistics with R	
Week 3	Jan 26	Review project #1 data; Develop Analysis Plan (Data mapping)		HW Exercises (#1)
	Jan 28	Executing an Analysis Plan		Reflection Questions 1; Scope of Work for Project #1
Week 4	Feb 2	Classification Models	Ch. 4 - ISLR	Work on Project #1, Find (and read) a clustering article to discuss
	Feb 4	Classification Models Cont'd.		HW Exercises (#2)
Week 5	Feb 9	Writing up research		HW Exercises (#3)
	Feb 11	Presenting research	Work on write-up and presentation	

Week 6	Feb 16	Present results from project #1	Ch. 39 Intro to Data Science	Deliverable (Project #1)
	Feb 18	IPI Statewide Survey Introduction	Read “Problems in the Analysis of Survey Data”, Sections A-D, by Morgan & Sonquist.	
Friday, February 21 st Abstracts due for Undergraduate Research Symposium				
Week 7	Feb 23	Linear Regression	Ch. 3 - ISLR	Work on Project #2
	Feb 25	Linear Model Selection & Regularization	Ch. 6 - ISLR	Work on Project #2
Week 8	Mar 2	Nonlinear Regression Models	Ch. 7 - ISLR	HW Exercises (#4)
	Mar 4	Regression Review		Submit Project #2, , Find (and read) a regression article to discuss
Week 9	Mar 9	Time-Series Analysis	<u>A Little Book of R for Time Series</u>	Start Project #3
	Mar 11	Decision Trees	Ch. 8 - ISLR	HW Exercises (#5)
Week 10	Mar 16	Decision Trees Cont’d.		Find (and read) a decision tree article to discuss
	Mar 18	Catch - up		HW Exercises (#6)
Week 11	Mar 23	Support Vector Machines	Ch. 9 - ISLR	Reflection 2; Identify final project for class
	Apr 2	Project Presentations		Project #3 Due
Abstract and Poster Submission Deadline for Undergraduate Research Showcase – Monday, April 5 th				
Week 12	Apr 7	Principal Components/Multivariate Analysis	Ch. 10 – ISLR	HW Exercises (#7)
	Apr 9	TBD		
Week 13	Apr 14	Work on final project, Factors that Can Affect Model Performance	BR! - Ch. 1-2	
	Apr 16	Work on final project	BR! - Ch. 3-4	
Undergraduate Research Showcase – Friday, April 23 rd 9-1:30 (Poster sessions are 9-10:30 and 12-1:30)				
Week 14	Apr 21	Bayesian Modeling	BR! - Ch. 5-6	Find (and read) a Bayesian modeling article to discuss
	Apr 23	Bayesian Modeling	BR! - Ch. 7	
Week 15	Apr 28	TBD	BR! - Ch. 8	
	Apr 30	TBD		Final Project Due (Project #4)