Introduction to Data Analytics

DA-101

Fall 2018

M – W- F 11:30am – 12:20pm

W 1:30pm – 4:20pm (lab)

Burton Morgan 219

Course Description

Many of the most pressing problems in the world can be addressed with data. We are awash in data and modern citizenship demands that we become literate in how to interpret data, what assumptions and processes are necessary to analyze data, as well as how we might participate in generating our own analyses and presentations of data. Consequently, data analytics is an emerging field with skills applicable to a wide variety of disciplines. This course introduces analysis, computation, and presentation concerns through the investigation of data driven puzzles in wide array of fields - political, economic, historical, social, biological, and others. No previous experience is required.

Instructor

Anthony Bonifonte

Office: Burton Morgan 410

Office Hours: Monday 1:30-4:00, Thursday 1:30: - 4:00

If you are unable to attend these hours, you are encouraged to email questions or schedule an appointment.

Email: bonifontea@denison.edu

TA: Ming Chen, chen_m1@denison.edu

TA Office Hours: Sunday 6:00pm – 8:00pm, Burton Morgan 405

Course Goals

At the end of the course, students should be able to:

- Understand the different types of data and the value of data

- Collect, clean, store, and extract relevant data

- Create data visualizations and summary statistics

- Statistically analyze data and draw conclusions

- Create basic computer programs for analyzing data

- Communicate findings to a diverse, non-technical audience in both written and oral format

- Reflect on ethical, social, and legal issues in data collection and security

"Information is the oil of the 21st century, and analytics is the combustion engine.

-Peter Sondergaard

Course Logistics

Prerequisites: None

Textbook: Introduction to Statistical Investigations, by Tintle et. al.

Other supplemental readings:
R for Data Science, by Grolemund and Wickham http://r4ds.had.co.nz/index.html

"Data is one of the most powerful mechanisms for telling stories. I take a huge pile of data and I try to get it to tell stories.

- Steven Levitt

Technology Policy: Please be respectful with your use of laptops and technology in class. I request you only use them for class related purposes, as I and others may find them distracting. Cell phones should be kept silent and away, and you can expect the same from me.

Software: All projects in this course will be analyzed using R, an open source data analysis language and environment. R is a powerful free statistics program that you will use throughout the major and your career. A small time investment to learn R will greatly benefit you in the future. No previous experience with R, statistical software packages, or computer programming is assumed.

Class forums: We will use NoteBowl. If you have conceptual questions that may benefit others, please post them there.

Expectations

Academic Honesty: Academic honesty, the cornerstone of teaching and learning, lays the foundation for lifelong integrity. Academic dishonesty is intellectual theft. It includes, but is not limited to, providing or receiving assistance in a manner not authorized by the instructor in the creation of work to be submitted for evaluation. This standard applies to all work ranging from homework assignments to major exams. I will assume that you are familiar with the Code of Academic Integrity. To learn more about it, please go to https://denison.edu/academics/curriculum/integrity

Class communication: All electronic communication will be through announcements using NoteBowl and delivered to your Denison email. You are responsible for checking these messages periodically to stay informed of important dates and potential changes to the syllabus.

I am pleased to reply to questions via email at bonifontea@denison.edu. Please include 'DA-101' in the subject line. I check my email frequently, but I reserve the right to a 48-hour response period. This means questions immediately before an assignment due date may not receive a timely response. Please send all emails through your Denison account so it does not get blocked by spam filters.

Disability Accomodations: Any student who feels he or she may need an accommodation based on the impact of a disability should contact me privately as soon as possible to discuss his or her specific needs. I rely on the Academic Resource Center (ARC) in 020 Higley to verify the need for reasonable accommodations based on the documentation on file in that office.

Assignments and Grading

Final Project 35%
Labs 55%
Reading Responses 10%
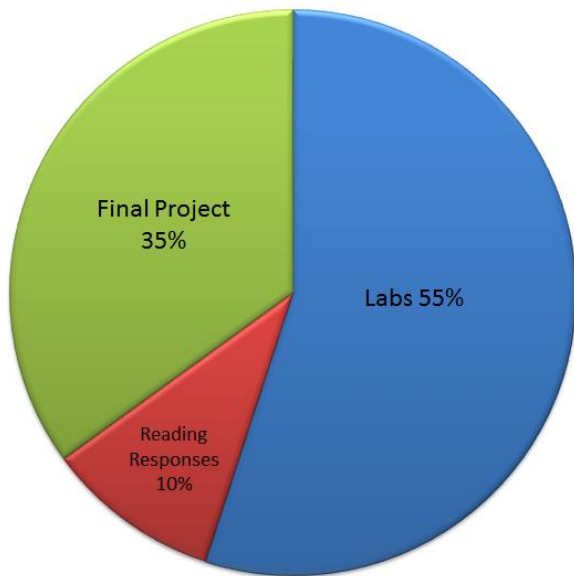
Final Course Grade:

A+:	98%	A:	92%	A-:	90%
B+:	88%	B:	82%	B-:	80%
C+:	78%	C:	72%	C-:	70%
D+:	68%	D:	62%	D-:	60%

Labs 55%

Weekly labs will challenge you to practice, apply, and extend the concepts and skills that you learn in the course.   There will be six topics from different subject areas where data analytics skills can provide insight into interesting questions, approaches, and hopefully some answers. The topics will increase in complexity, skills, breadth of material used, and expectations. Consequently, later labs will be worth more points toward your final grade than earlier labs.

Some labs will be short individual assignments to practice skills introduced that week. Some labs will be larger group work projects that require you to apply the skills from previous weeks to creatively investigate the overarching questions and write a group report.

Topic 1: Biology – Garlic mustard plants – 4%

Topic 2: Infrastructure Planning – Parking data collection - 6%

Topic 3: Sociology - Poverty Simulation Survey – 8%

Topic 3: Psychology – Drug use – 10%

Topic 4: Political Science – Congressional Polarization – 11%

Topic 5: Ethics - 4%

Topic 6: Entertainment Business - Movie Analytics – 12%

Reading Responses 10%

Friday class time will be used for discussion of selected readings. To ensure you are prepared for the discussion, you will be required to turn in a short (1-2 page) response to specific questions. Each response will be graded on a 0 – 2 point scale.

Final Project 35%

While labs are more structured and focused more intensely on skills introduced in the weeks prior, the final project will be more comprehensive and offer more flexibility in terms of your design, data choices, and approach while demonstrating mastery of the breadth of skills discussed in the course.

Potential topics for the project will be discussed with the instructor for suitability and suggestions. Topics can arise from any of the data sets we use in class, or from outside sources. Final reports will discuss the data, analytical approach, and visual presentation of results.

"I never guess. It is a capital mistake to theorize before one has data. Insensibly

one begins to twist facts to suit theories, instead of theories to suit facts."

"Data! Data! Data! I can't make bricks without clay!"

Sherlock Holmes, as written by Sir Arthur Conan Doyle
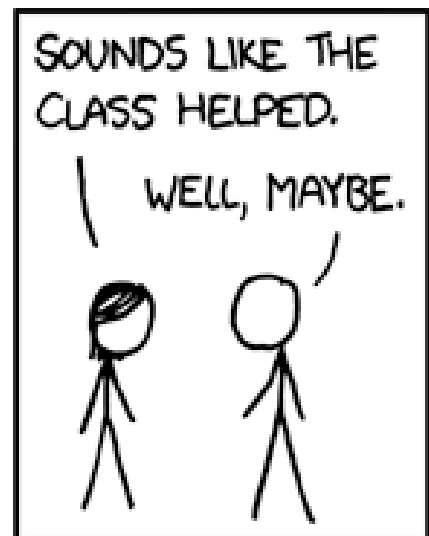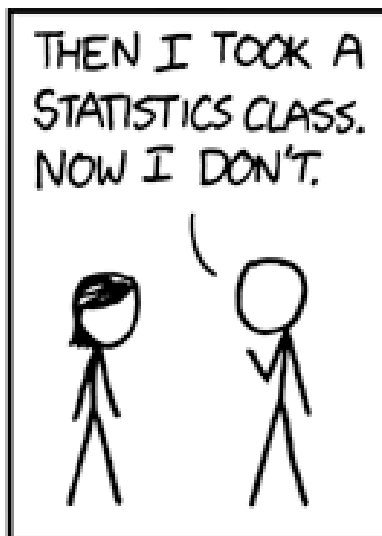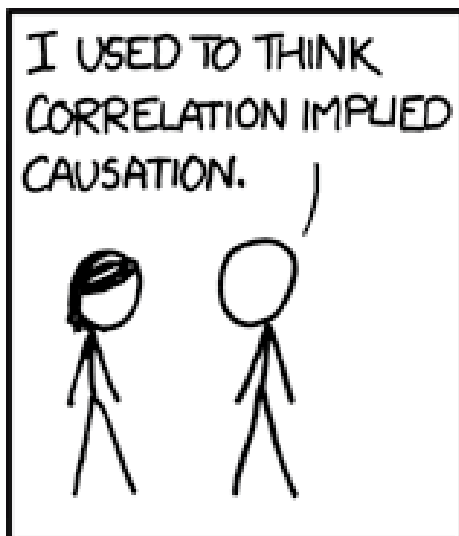
Course Schedule

Each course topic will prepare you to answer the quoted question. Readings for each week, due Fridays, are in green.

Dates	Topics	Labs
8/31 – 9/07	Introduction to Data Analytics "What is data, what are the types of data, why are they important, and what can I do with them?" "The difference between Data, Analytics, and Insight", section p.1 of text	Biology
9/10 – 9/14	Challenges in Handling Data and R software "How do I acquire, store, and access data? How can I clean data and put it into a usable format? How should I handle missing data? What should I do about outliers?" "Making data analytics work for you-instead of the other way around"	Infrastructure Planning
9/17 – 9/21	Descriptive Analytics: Data visualization "How can I visualize data for better understanding? "7 Elements of Good Data Visualization", "What makes a good data visualization"	Sociology
9/24 – 9/28	Descriptive Analytics: Summary statistics "How do I summarize and report historical data?" "The Fallacy of Interpreting Deaths and Driving Distances", Textbook 1.3	Psychology
10/01 – 10/05	Descriptive Analytics: Hypothesis Testing "How can I statistically test two competing claims? "A Dirty Dozen: Twelve P-Value Misconceptions", Textbook 1.4	Psychology
10/08 – 10/12	Predictive Analytics: Linear Regression "How can I extrapolate historical data into predictions about the future?" "What is the difference between missing completely at random and missing at random?"	Political Science
10/15 – 10/17	Predictive Analytics: Multivariate Regression "How can I extrapolate many variables at once?" Textbook 4.1 – 4.2	Political Science
10/22 - 10/26	Correlation vs. Causation "What are the limits to the conclusions that can be drawn from data? How can we design studies to establish causation?" "Types of Sampling Designs Used in Sociology"	Business
10/29 – 11/02	Prescriptive Analytics: Optimization "How can I use my predictions to make decisions and account for constraints?" "Principles and Applications of Operations Research"	Business
11/05 – 11/09	Ethics and Issues in Data Analytics "What are the ethical, legal, and social considerations of data acquisition, storage, and analysis?" Excerpt from: *How to Lie with Statistics*	Ethics
11/12 – 11/16	Good Practices in Coding and Science "What are good standards for sharing code, maintaining data, and reporting results?" "Learn any of these 16 programming languages and you'll always have a job"	Project Introduction
11/26 – 11/30	Other topics in Data Analytics – may include big data, machine learning, coding issues, or other topics of student interest "Beyond the hype: Big data concepts, methods, and analytics"	Project Working Day
12/03 – 12/07	Presentation Skills "What are best practices for presenting quantitative results to audiences?" "Oral Presentations", "General Guidelines for Writing Reports"	Project Working Day
12/10 – 12/14	Other topics in Data Analytics Student choice of reading	Project Presentations