Math 400
syllabus

Spring 2016

Professor:	Matthew Neal (nealm@denison.edu)
Office and phone:	Olin 2225A (6288)
Office Hours:	Thursday (to be assigned) or by appointment
Web site:	http://personal.denison.edu/~nealm/

Course

This course is a collaborative research data analytics project drawing from three real problems in industry or government. After collecting feedback on your skills and interests, you will be assigned a problem and a team on the first day of class. Your team will make a presentation each week on your progress analyzing the problem. Your grade is based on these presentations, a final video presentation and 5 page paper (max) plus Appendicees.

Grades and Expectations

The grade will be calculated thusly

- 40 % for weekly presentations
- 25 % for final video presentation
- 35 % for final paper

Grading Rubricks

The first week we will discuss your role you will play on your team. Some may be better at coding, some better at modeling, and some may be better at communication. Although you may lean toward one role, each member is expected to understand all aspects of the problem. I will verify this in our office meetings on Thursday and by rotating who gives the weekly presentation. All members of the team will be given the same grade by default **but** I may move someone down a bit if their performance has been noticeably less than the others. Your in class and final presentations will be graded on:

- 40 % Evidence of critical thinking and progress on the problem
- 20 % Clarity
- 20% Organization
- 20% Peer evaluation

In particular you must show awareness of your audience. At the end of each class I will do a group evaluation which will factor into your grade, so know your audience! Your final presentation and paper will be graded on:

- 20 % Evidence of critical thinking and progress on the problem
- 10% Overall organization
- 10 % A broadly understandable abstract/introduction
- 10 % Clear statements and justifications of assumptions
- 10 % Clear formulation of your model choices with justifications
- 10 % Clear explanation of initial data organization, cleaning and computational methods
- 10% Clear communication and interpretation of mathematical results
- 10 % Clear summary of results in English for non experts.
- 10% Peer evaluation of project

Office Hours

We will set up weekly meetings in my office on Thursdays to discuss your project

Late Work

Late assignments will receive a 20 % point penalty per day late unless there is a PRIOR written note (such as a note from Whistler) that verifies a VERY strong excuse (such as illness or important sports team events). Late quizzes are not accepted at all without a written excuse as above.

Academic Integrity

The students and faculty of Denison University and the Department of Mathematics and Computer Science are committed to academic integrity and will not tolerate any violation of this principle. Academic honesty, the cornerstone of teaching and learning, lays the foundation for lifelong integrity.

Academic dishonesty is, in most cases, 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 daily homework assignments to major exams. Students must clearly cite any sources consulted, not only for quoted phrases but also for ideas and information that are not common knowledge. Neither ignorance nor carelessness is an acceptable defense in cases of plagiarism. It is the student's responsibility to follow the appropriate format for citations.

Proposed and developed by Denison students, passed unanimously by DCGA and Denisons faculty, the Code of Academic Integrity requires that instructors notify the Associate Provost of cases of academic dishonesty, and it requires that cases be heard by the Academic Integrity Board. Further, the code makes students responsible for promoting a culture of integrity on campus and acting in instances in which integrity is violated.

For further information about the Code of Academic Integrity see http://www.denison.edu/about/integrity.html

Disabilities

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 Support and Enrichment Center in 102 Doane to verify the need for reasonable accommodations based on documentation on file in that office.

Projects

1. **Contact Barnaby Nardella-AIG** The insurance industry in the United States is over $1 trillion a year in revenue and spans products that cover your home in a hurricane to covering your pet?s medical bills with a veterinarian. All products have one thing in common in that they cover future events and that they use data to determine what to charge a customer given their risk profile. Insurers are always looking to better predict this risk utilizing new modeling techniques and data sources, which often include external data sources and even devices that track how good a driver for car insurance you are to how often you go to the gym for life and health insurance.

   The auto insurance industry has typically been the most active segment in trying to determine whether someone will get into an accident and if they do how severe it will be. Car insurers will often look at more than 40 variables to determine your rate leading to a near infinite range of possible rates. These variables include ones you would expect, like how many speeding tickets you have or your age, to ones that you would not expect, like your credit score or where you live. Your zip code is one of the most powerful determinants of your rate and if you live in a certain area, the price you pay for insurance could be up to 3 times higher than an identical customer living somewhere else.

   Your problem is to figure out what aspects of a location make it riskier than others and whether it leads to more accidents or increases severity of accidents or both. You will be utilizing data from the Highway Loss Data Institute (HLDI) that compiles data from a majority of car insurers on accidents at a zip code level. You will need to merge this data with other external datasets at the zip code level to help you figure out where certain zip codes are riskier. You will then develop an algorithm that would price insurance by zip code by these variables you determine to be predictive. There are numerous data sources you can use to do this that are publicly available and part of the challenge is determining what could make a certain area riskier than another then trying to find a dataset that contains that data.

2. **Contact Nicolaus Covey–Neilsen** As the world population approaches 8 billion , many are wondering whether we have the capacity to sustainably nourish 8 billion people and beyond. Even today, nearly 800 million individuals (1 in 9 on the planet) do not have the food they need to lead a "healthy, active life." Read into this global challenge and you?ll find many big open questions of supply: are we losing arable land? What will the impact of climate change be on crop yield? Are we making the right infrastructure investments to ensure food can reliably make it from farm to form with sufficient roads and refrigeration facilities? Less frequently discussed is the demand (and need) part of this global challenge. In order to plan on meeting food needs, don?t we first have to anticipate what the need will be? Recent demand models developed by the USDA have increased in sophistication, incorporating food quality, income level, aid and many other factors. To a lesser extend, organizations have taken on the question of what the global "need" is (a distinction that will become clearer to you).

   Question How much food would the world need to produce in order to nourish a global population of 9 billion people?

   Of course, within this question there are many underlying questions

   - What is the best way to quantify food need on global and individual levels?
   - What is the difference between food ?need? and, the more commonly discussed and modeled, food "demand??
   - How much food does an individual need to be adequately"nourished?"
   - How does food need vary by age, gender, socioeconomic status, region and other factors and how does that relate to the changing population dynamics around the world?
   - How are diets shifting in the developed world? The developing world?
   - How much of the food produced, today, is wasted and how is that likely to change between now and the point at which our global population will reach 9 billion people?
   - How much "food" needs to be produced for non-human consumption and how will that change (e.g. animal feed, biofuels)

   Challenge

   - Estimate the amount of food the world would need to adequately nourish a population of 9 billion. Derive this estimate from a new model that uses the best available data but allows for the model to be adjusted based on new information or alternative assumptions. Optionally present a range of scenarios based on the biggest unknowns related to global food needs. You may look at established demand models for inspiration or consider other predictive modeling techniques.
   - Once this accomplished, compare your results to established market demand models (or your modifications to them) to determine a true "food security gap" between what is needed and what the people are willing and able to consume through purchasing and current aid trajectories.
   - Ideally, this investigation would lead to a deep discussion of the interactions between demand and need models. Perhaps the demand models should incorporate "need" in some more sophisticated way

3. **Contact: Larry Sherman (Cambridge), Jeanette Kerr (Assistant Commissioner Crime & Specialist Services, Northern Territory Police, Australia)** Despite the large body of research on domestic abuse, too little evidence is available about fundamental patterns of violence within couples. Until very recently, for example, Police Chief Officers in the United Kingdom told the public that victims of domestic abuse suffered 34 episodes prior to reporting to police. This figure was then exposed as lacking any credible evidence for contemporary cases of domestic abuse in the United Kingdom (Strang, Neyroud, & Sherman, 2014). Similarly, unsupported assertions have long been made about "escalation", the notion that over the life of a domestic relationship, the severity and frequency of any violent events will increase with each further report. According to Pagelow (1981), escalation is something on which researchers can agree, yet a recent literature review found no consistent evidence to support that claim (Bland, 2015).

   Indeed, in England and Wales, for example, police and other organizations have put escalation at the center of their efforts to assess risk of harm in future cases. All English and Welsh forces are required to complete a risk assessment form with the victim even if no crime has been proven. That risk assessment (known as "DASH"), which to varying degrees is later reviewed by a specialist, asks specific questions designed to predict escalation in severity and frequency. Yet many questions remain about the extent of the empirical evidence that supports the validity or reliability of the DASH predictions. This project aims to build an evidence-based strategy for targeting different patterns of domestic abuse for different types of interventions at different levels of cost.

The main goal is to consider three key questions about the population of over 60,000 domestic abuse events in Australia. **First**, you must address the question of how much, if any, escalation in the severity of harm occurs in repeat cases. **Second**, you must examine whether repeat cases become more likely or frequent as cases become more chronic, independent of seriousness. **Third,** you must examine individual dyads (couples of one offender and one victim as distinct, unique units) for the distribution and concentration of serious harm for evidence of a "power few" (Sherman, 2007). **Pushing the project further**, you should attempt to find statistically significant models of the relationships between the relevant variables and validate your models with predictions.