Syllabus Math 685 (697) Fall 19

Course Description

In this class, we will explore data science and you will learn the theory, tools, and skills to store, clean, manipulate, visualize, model, and extract information from various data sources

This is an independent study class. We will meet weekly to touch base and answer questions.

Logistics

Instructor: Dr. Robin Donatello
Office Location: Holt 202
Phone Number: 898-5767
E-mail: rdonatello@csuchico.edu
Prerequisites: Math 615 or equivelant. **Meeting Days and Times** F 1-2:30, PHYS 213

Learning Outcomes

From this course, you should achieve the following learning outcomes:

- Explain how Data Science differs from the traditional disciplines of Statistics and Computer Science
- Distinguish different data types and understand how to process them
- Understand the operations, syntax, and good practices for programming in R and apply them to solve data science problems
- Maintain documentation, versions, and revisions of their code using Markdown, Git, and GitHub
- Organize data in a "tidy" way for storage and analysis
- Manipulate and clean data to prepare it for automated analysis
- Create visualizations to accurately and effectively communicate data-driven insights
- Disseminate your work using multiple modes such as reports, tutorials, blog posts, websites, dashboards
- Building your data science profile
- Apply algorithms and statistics appropriately for exploratory data analysis
- Basic text processing, including sentiment analysis and regular expressions

This list devolved into a topic list

Required Materials

A laptop is required in each class meeting. Mac OSX and Windows 10 laptops are preferred. If you have a different system, consult with the professor, but understand that lessons and support may be different for your computer and you are responsible for meeting the requirements of the class.

This class has is considered a *zero cost class*, meaning all materials we will use in this class are free. We may sign up for web based services, but you will be able to get educational or promotional credits.

Computer software

- R
- R Studio
- LaTeX

Textbooks

Any textbook listed below that costs $$ is optional.

Programming

- R for Data Science (R4DS) http://r4ds.had.co.nz/
- Happy Git and GitHub for the useR https://happygitwithr.com/
- https://stat545.com/topics.html

Statistics

- Open Intro Statistics 3rd edition. If you need a refresher on your basic statistics. Free PDF available at https://www.openintro.org/stat/textbook.php?stat_book=os
- *Practical Multivariate Analysis, 5th ed* by Afifi, May Clark. Available cheaper at Amazon

Data Science

- Data Science and Big Data Analytics (DSBDA) https://www.amazon.com/Data-Science-Big-Analytics-Discovering/dp/111887613X

"Weekly" Outline

1. What is Data Science?
2. Basic care and feeding of data
3. Version control
4. Wrangling data
5. Exploratory data analysis project
6. Moar Data I/O through external files, possibly API's?
7. Building basic websites
8. Group project
9. Establishing your online identity by building your DS portfolio
10. Web scraping
11. Text analysis, regex hell, sentiment analysis 12 and beyond - Some of: Basic predictive modeling, clustering, bootstrapping, relational databases, dimension reduction Culminates with a final project.

Grading

A) Student made a solid effort to learn things, completed 80% of assignments and contributed to discussion
B) Student made a good attempt to learn things, completed 70% of assignments
C) Student made a half-hearted attempt to learn, completed 50% of assignments.

Less effort than this will not earn a passing grade.