

INSY 6500
Information Systems for Operations

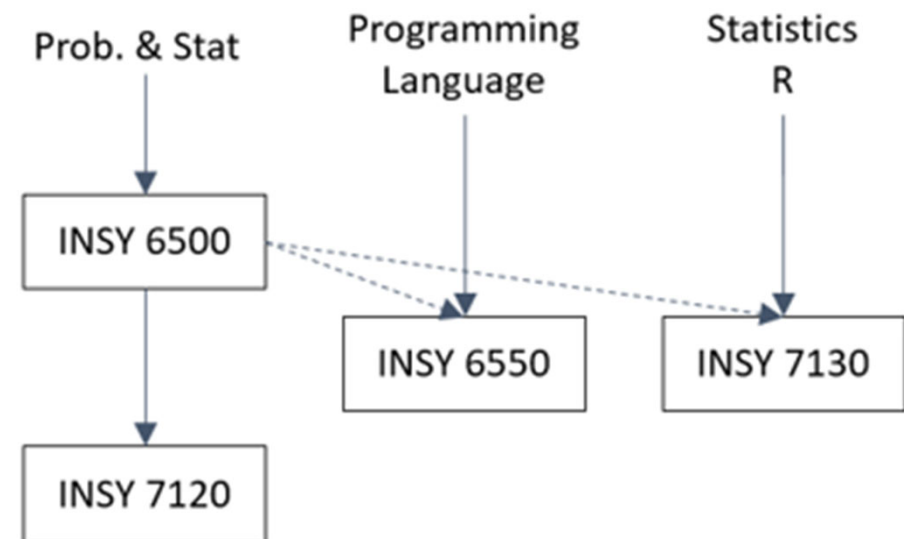
This is NOT a Programming Class

Fall 2019

Jeffrey S. Smith

Industrial and Systems Engineering Department
Auburn University
jsmith@auburn.edu

<http://eng.auburn.edu/program/modeling-and-data-analytics-for-operations-grad-cert.html>



Course Description and Objectives

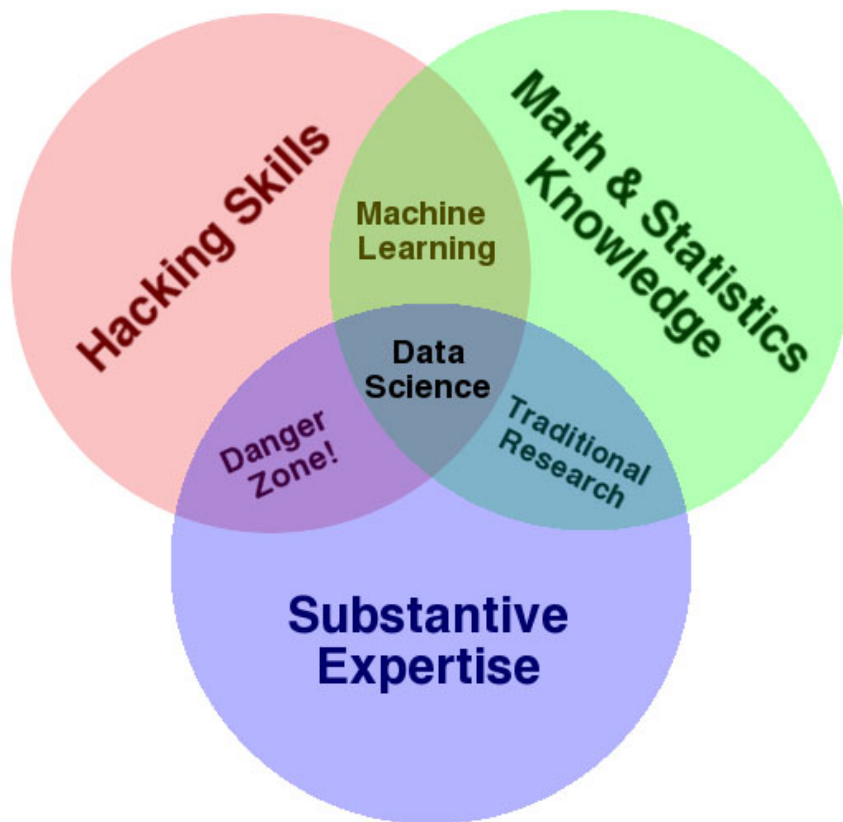
Description: Introduction to modern *data science* tools with applications in manufacturing and service industries and operations. Focus on the manipulation and use of small and large datasets. Tools include Jupyter, Python, R, and MySQL/MariDB along with the related packages that support data modeling, visualization, and analysis.

Objectives: By the completion of this class, students should understand the use of our standard toolset (Python, Jupyter, R, Matplotlib, MySQL) and have the ability to manipulate and visualize datasets of all sizes. In addition, students should have a basic understating of data science and the available tools/resources related to this topic.

This is NOT a programming course!

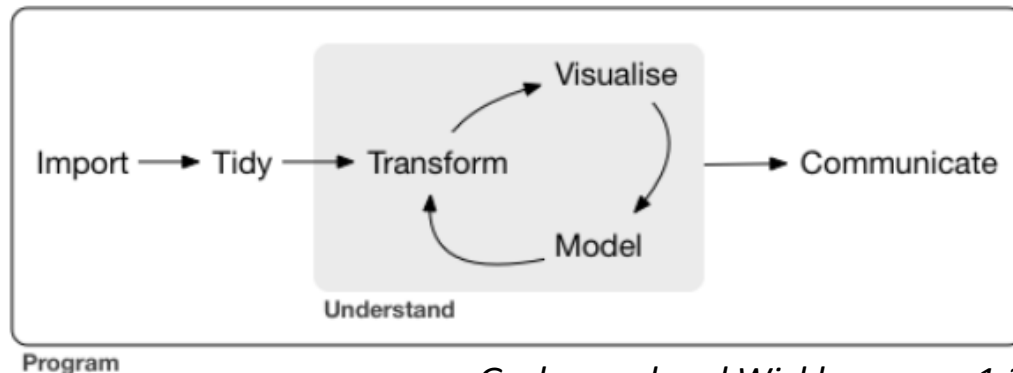
What is “Data Science”

- “Substantive Expertise” == “Domain Expertise”



[From <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> and the VanderPlas book]

Tools/Process for Typical Data Science Project



Grolemund and Wickham, sec. 1.2

- **Tidy** data – “In brief, when your data is tidy, each column is a variable, and each row is an observation.”
- **Transform** – “Transformation includes narrowing in on observations of interest (like all people in one city, or all data from the last year), creating new variables that are functions of existing variables ”

- **Visualization** – “is a fundamentally human activity. A good visualization will show you things that you did not expect, or raise new questions about the data.”
- **Models** – “are complementary tools to visualization. Once you have made your questions sufficiently precise, you can use a model to answer them. Models are a fundamentally mathematical or computational tool, so they generally scale well.”

How to Be Successful in This Class

- VanderPlas (Help and Documentation in IPython)

*When a technologically-minded person is asked to help a friend, family member, or colleague with a computer problem, most of the time it's less a matter of knowing the answer as much as **knowing how to quickly find an unknown answer**. In data science it's the same: searchable web resources such as online documentation, mailing-list threads, and StackOverflow answers contain a wealth of information, even (especially?) if it is a topic you've found yourself searching before. Being an effective practitioner of data science is less about memorizing the tool or command you should use for every possible situation, and more about learning to effectively find the information you don't know, whether through a web search engine or another means.*

The actual, up-to-date
syllabus is on Canvas!

Course Syllabus

INSY 6500/6: Information Technology for Operations Fall 2019

Instructor: Jeffrey S. Smith, Ph.D.
3306 Shelby Center
jsmith@auburn.edu
Office Hrs: M – 9:00 a.m. – 12:00 noon or
by appointment.

Credit Hours: 3

Prerequisites: None

Course Text and Reference Material:

Course Texts:

Vanderplas, J., *Python Data Science Handbook*: Essential Tools for Working with Data, O'Reilly Media, 2016. Available on Amazon and at <https://jakevdp.github.io/PythonDataScienceHandbook/>

Grolemund, G. and Wickham, H. *R for Data Science*, O'Reilly Media, 2017. Available on Amazon and at <http://r4ds.had.co.nz/>

Gitthub Repository:
<https://github.com/ausim/InfoTechforOps>

Video Modules:
<http://jsmith.co/node/244>

Course Description: Role and potential of using computer-integrated systems within manufacturing and service industries. Analysis of relevant data, synthesis of the flow of information in an operations environment, and development of databases to support the production process.

Course Objectives: By the completion of this class, students should understand the use of our standard toolset (Python, Jupyter, R, Matplotlib, MySQL) and have the ability to manipulate and visualize datasets of all sizes. In addition, students should have a basic understanding of data science and the available tools/resources related to this topic.

Course Project

- Either individually or in teams of 2 students (self-chosen)
- Want students to do some **interesting** work with data sets. Ideally, it would be something interesting to the individual students/teams. What I'm looking for are the following:
 - Mastery of the tools from our toolset (the ones that you choose to use -- you don't have to use all of them, though)
 - An appropriate “amount” of work. This is 35% of your grade -- I expect a commensurate amount of effort on your part.
 - Data + Code/Notebooks/R Projects + Report.
 - Presentation (video for the Outreach students -- unless you want to come to campus!)
- Best strategy for success: Start early, work diligently!