Updating the DSC Curriculum

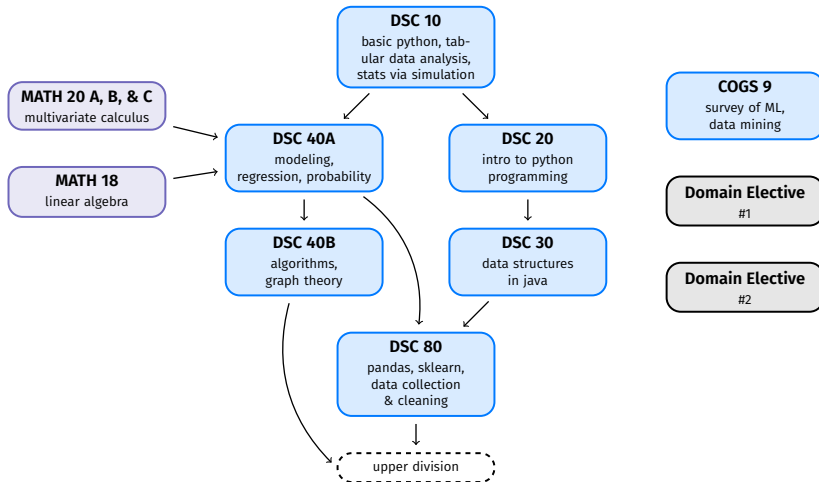# Mission

- A bachelors degree preparing students to compete with MS programs in data science

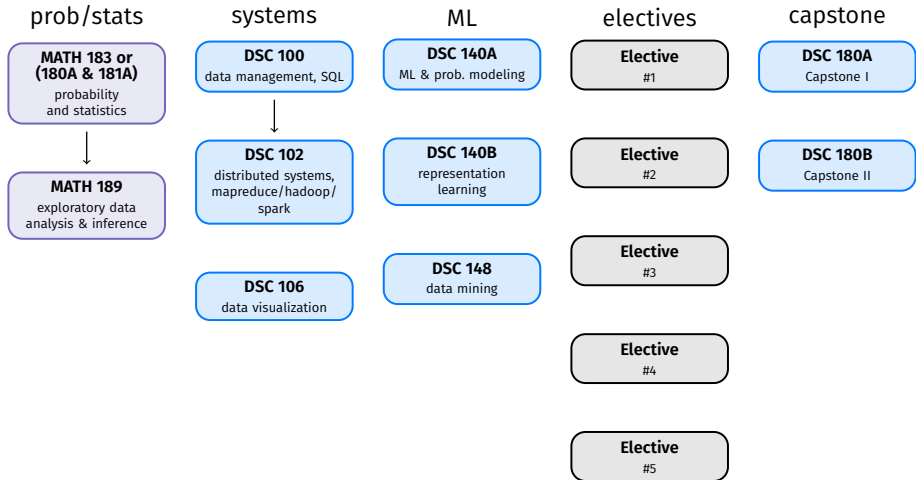# Mission

- A bachelors degree preparing students to compete with MS programs in data science

- Are we succeeding?

Lower Division

DSC 10
basic python, tab-
ular data analysis,
stats via simulation

MATH 20 A, B, & C
multivariate calculus

MATH 18
linear algebra

DSC 40A
modeling,
regression, probability

DSC 20
intro to python
programming

DSC 40B
algorithms,
graph theory

DSC 30
data structures
in java

DSC 80
pandas, sklearn,
data collection
& cleaning

upper division

COGS 9
survey of ML,
data mining

Domain Elective
#1

Domain Elective
#2

Upper Division

prob/stats

MATH 183 or (180A & 181A)
probability and statistics

MATH 189
exploratory data analysis & inference

systems

DSC 100
data management, SQL

DSC 102
distributed systems, mapreduce/hadoop/spark

DSC 106
data visualization

ML

DSC 140A
ML & prob. modeling

DSC 140B
representation learning

DSC 148
data mining

electives

Elective
#1

Elective
#2

Elective
#3

Elective
#4

Elective
#5

capstone

DSC 180A
Capstone I

DSC 180B
Capstone II

Student Feedback

▶ The data is noisy.
  ▶ **Takeaway**: interests are diverse; offer more choice.

▶ Redundancy in some areas.

▶ Long and winding road to graduation.

▶ Missing a course in data science "tools".

Instructor / Industry Feedback

- ▶ Math skills could use improvement.

- ▶ Difficulty translating mathematical ideas to code.

- ▶ Missing classes in ethics, communication, tools.

- ▶ Critical thinking, independence

Exercise

In your experience, which skills need more reinforcement?

Constraints

▶ The DSC major is already very large.

▶ Transfer students cannot graduate within 2 years.

▶ **Takeaway:** we should avoid adding units.

Units by Major

Major	Lower Div.	Upper Div.	Total
CSE	52	72	124
DSC	**52**	**60**	**112**
Math/CS	42	56	98
Cogs	44	48	92
Prob/Stats	32	56	88

Goal

- ► Without increasing number of units:
  1. Strengthen the math / theory curriculum
  2. Add communication, ethics, tools
  3. Allow students to specialize

Step 0) Remove Domain Electives Requirement

- **Last retreat:** proposed removing domain elective requirement.

- **Now:** assuming we do, what does this allow?

Step 0) Remove Domain Electives Requirement

▶ **Last retreat:** proposed removing domain elective requirement.

▶ **Now:** assuming we do, what does this allow?
  ▶ A strawman proposal

A Tools Course

- ▶ Create a 2-unit lower-div tools course
  - ▶ Git, working in the shell, LaTeX, etc.

- ▶ Concurrent with programming course?

ML / Stats Curriculum

```
DSC 40A                    DSC 80
modeling, probability  →   data science
                           in practice

MATH 183 or
(180A & 181A)          DSC 140A           DSC 140B           DSC 148
probability            ML & prob. modeling representation      data mining
and statistics                            learning

MATH 189
exploratory data
analysis & inference
```
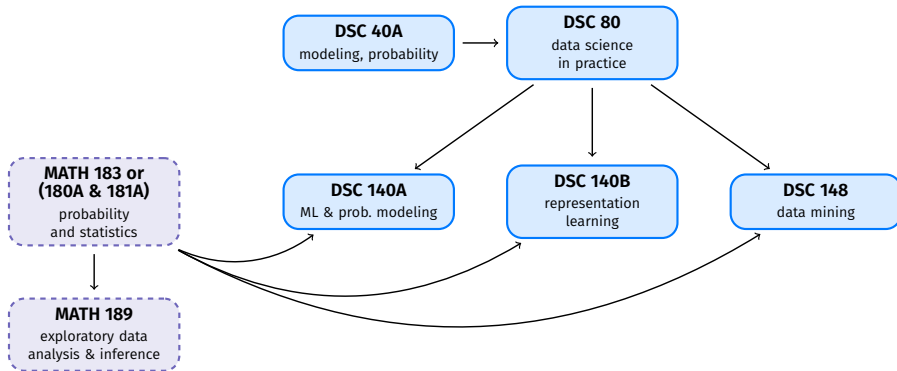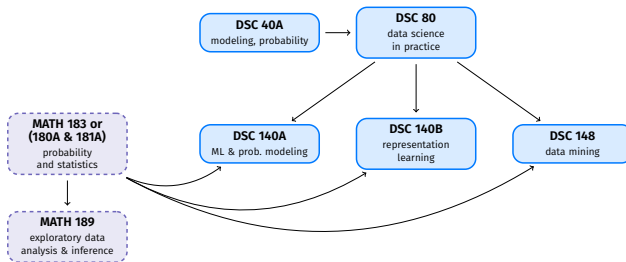
Problems

▶ Need to recap linear algebra, vector calculus in every upper-div ML course
  ▶ Preference: allow taking 140A & 140B in either order

▶ Redundancy

▶ More ML practice and implementation?

Proposals

▶ Factor out common recap into a strong lower-div "math methods for ML" course

Idea

▶ Machine learning lab, taken concurrently with ML theory courses

▶ **Problem**: scheduling

Proposal

1. Turn DSC 148 - Data Mining into third ML course, DSC 140C.

2. Spread existing ML theory in DSC 140A & B across DSC 140A, B, C.

3. Integrate lab into DSC 140A, 140B, 140C.

Proposal

4. Refocus DSC 40A to be a math methods course for 140A,B,C.
   - ▶ Remove discrete probability from DSC 40A.
   - ▶ Create a new course, DSC 40C - Probability and Statistics, replacing MATH 183 - Probability and Statistics.
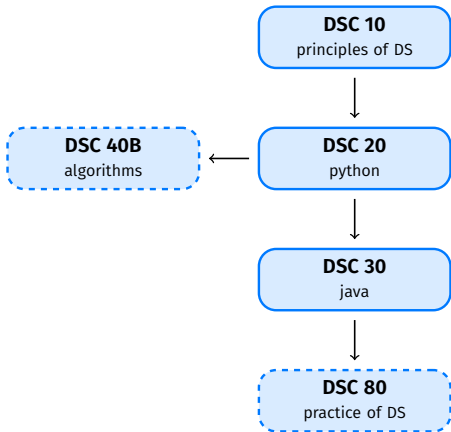
Additionally...

► Move MATH 189 into DSC.

► CSE ML courses (e.g., CSE 151A) are currently substitutable for DSC ML courses – remove them.
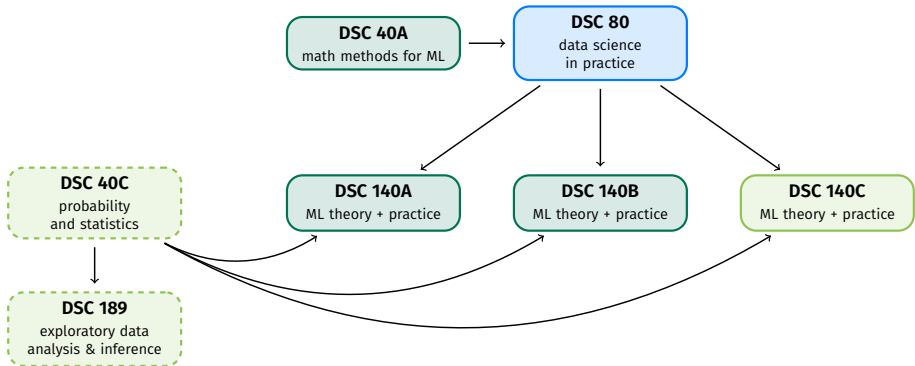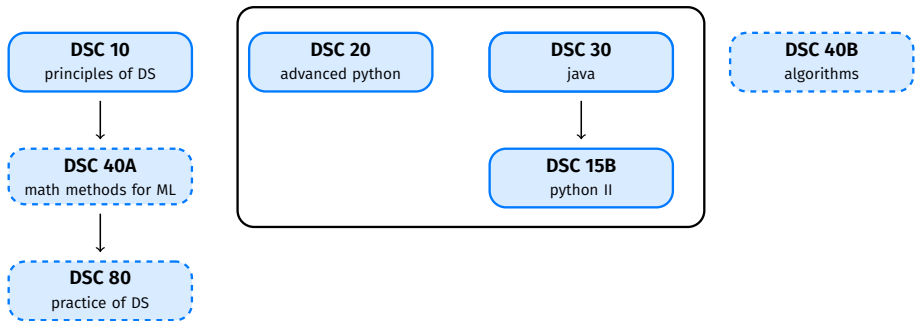
Programming Courses

Existing Curriculum

Problems

▶ Some students have prior programming experience; others don't

▶ DSC 30 (Data Structures in Java) is awkward
  ▶ Java is not used later
  ▶ But DSC 30 takes prime real estate in lower div, is a prereq for DSC 80

Proposal

1. Move DSC 30 (Java) to an upper div requirement

2. Create parallel intro programming tracks
   - ▶ DSC 20 (existing): Advanced Python; or
   - ▶ DSC 15A / 15B: Python I and II

3. Remove DSC 10 as a prereq for DSC 20 (DSC 15A)

Proposal

DSC 10
principles of DS

DSC 20
advanced python

DSC 30
java

DSC 40B
algorithms

DSC 40A
math methods for ML

DSC 15B
python II

DSC 80
practice of DS

Specializations (or "Tracks")

Problem

- ▶ Currently: 20 units (5 courses) of upper-div electives, unstructured
  - ▶ Suggested to choose courses within domain tracks (e.g., natural sciences)

Proposal

- Create specializations.
  - Data systems, theory, NLP, business analytics, …

- Re-purpose 5 electives. For example:
  - 3 should be from a specialization area
  - 1 should be data ethics/communication
  - 1 is a free DSC elective

- Different specializations could have different core requirements.

Capstone Prerequisites

- The capstone has many prerequisites:
  - an ML course, DSC 102, DSC 106, MATH 189

- Not all capstone domains may require *all* of these courses

- We might consider setting capstone prerequisites *per domain*