

Dismantling Math, Stats, and CS Silos: PCMI Guidelines for Undergraduate Majors in Data Science

Albert Y. Kim [@rudeboybert](https://twitter.com/rudeboybert)
Smith College - Statistical & Data Sciences

[@SmithCollegeSDS](https://twitter.com/SmithCollegeSDS)

Thursday August 2nd, 2018



2016 PCMI Undergraduate Faculty Program

Goal: Forming an early set of guidelines for undergraduate majors in data science



About the proposal

About the proposal

- Math, computer science, and statistics *à-la-carte*.

About the proposal

- Math, computer science, and statistics *à-la-carte*.
- What do we include, but also *what do we leave out?*
If you try to include everything, soon you'll end up with nothing.

About the proposal

- Math, computer science, and statistics *à-la-carte*.
- What do we include, but also *what do we leave out?*
If you try to include everything, soon you'll end up with nothing.
- Among the first proposals with interdisciplinary representation

About the proposal

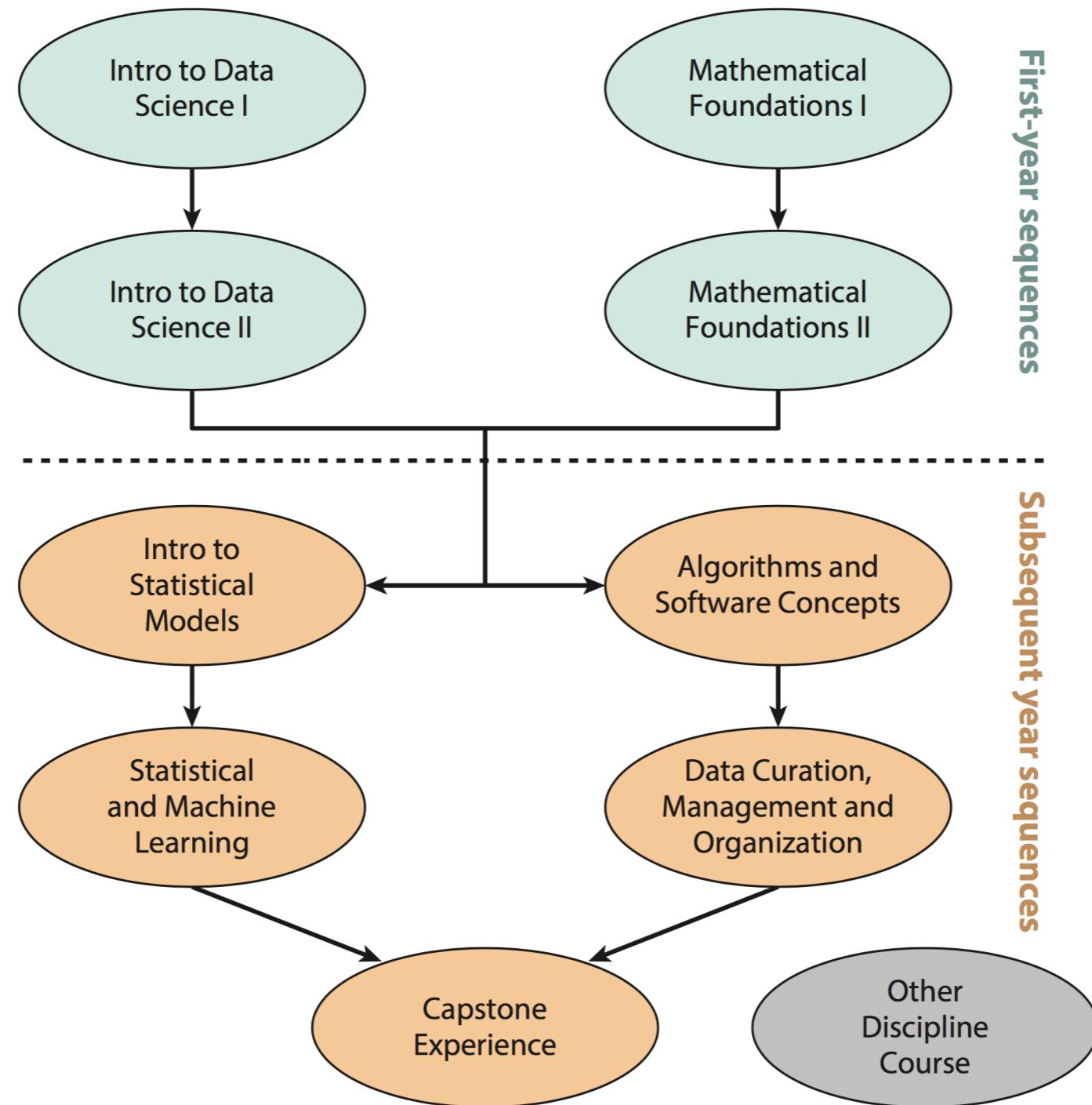
- Math, computer science, and statistics *à-la-carte*.
- What do we include, but also *what do we leave out?*
If you try to include everything, soon you'll end up with nothing.
- Among the first proposals with interdisciplinary representation
- Ben Baumer asked: “*What if we blew up math, stats, CS, and all their legacies and started over? What would this field look like?*”

Key Competencies for an Undergraduate Data Science Major

- 1. Computational and statistical thinking**
- 2. Mathematical foundations**
- 3. Model building and assessment**
- 4. Algorithms and software foundation**
- 5. Data curation**
- 6. Knowledge transference, communication, and responsibility**

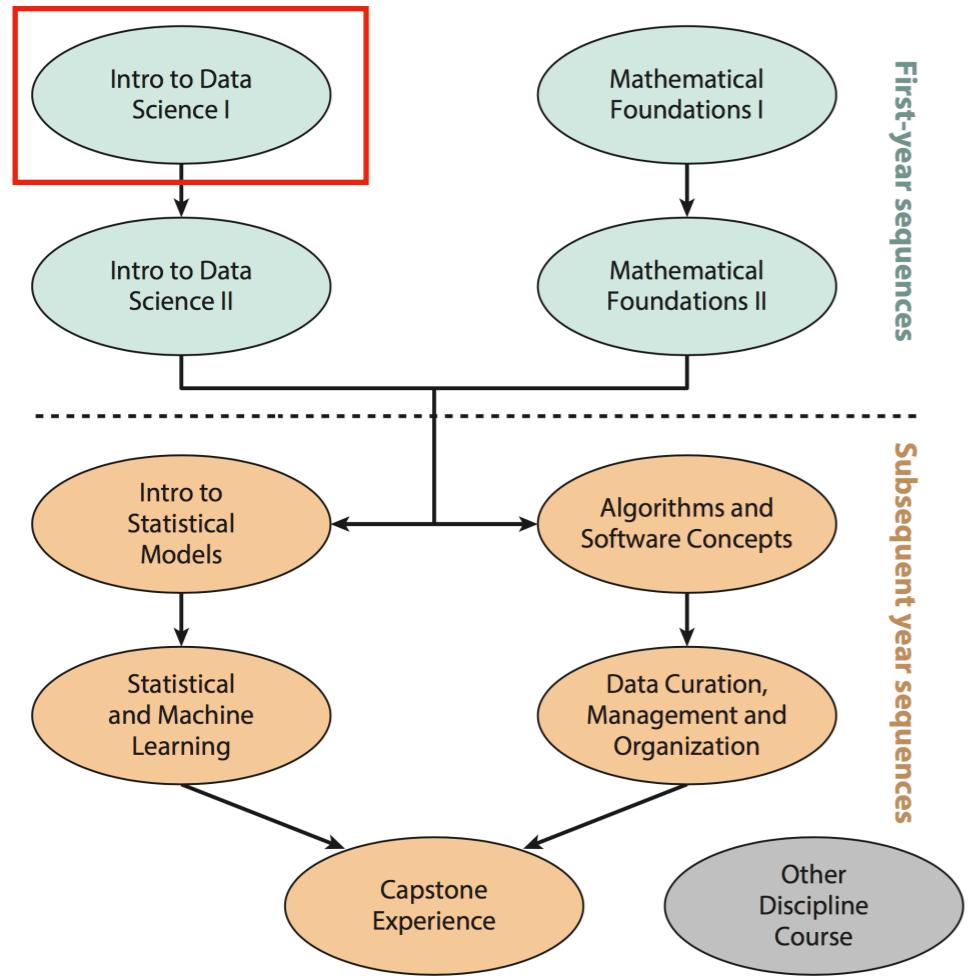
Result: PCMI Guidelines

Curriculum Guidelines for Undergraduate Programs in Data Science ([link](#))

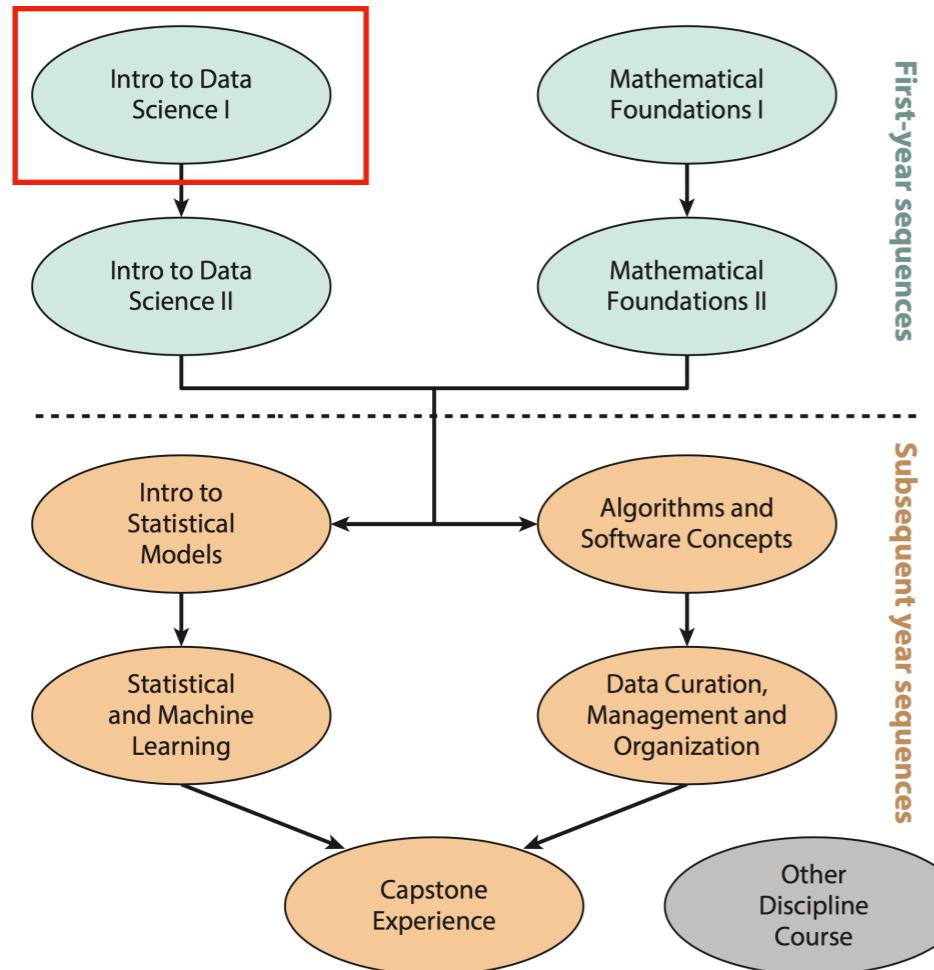


Intro to Data Science I

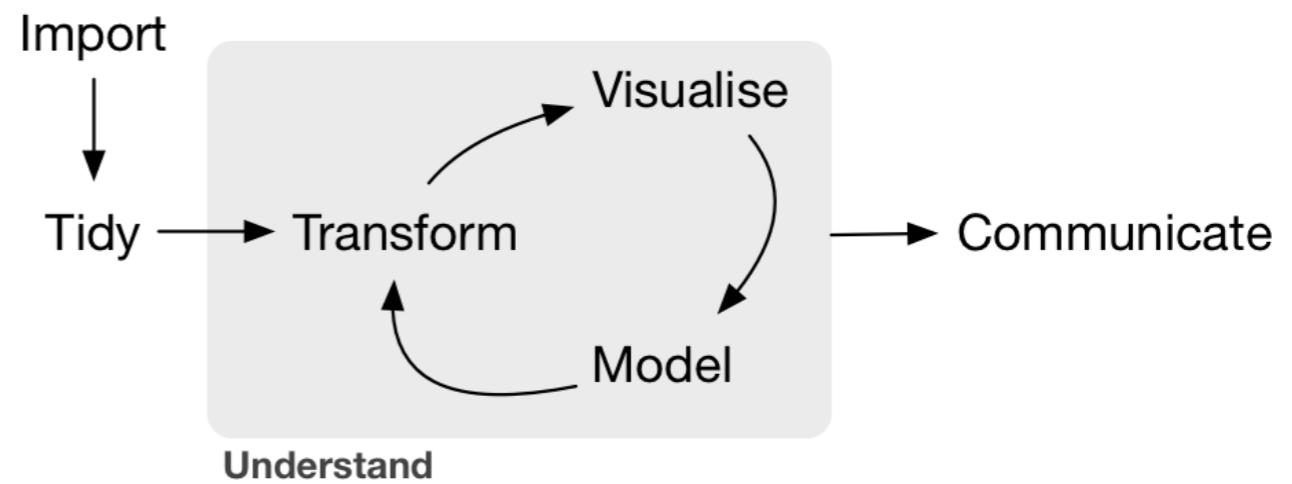
Intro to Data Science I



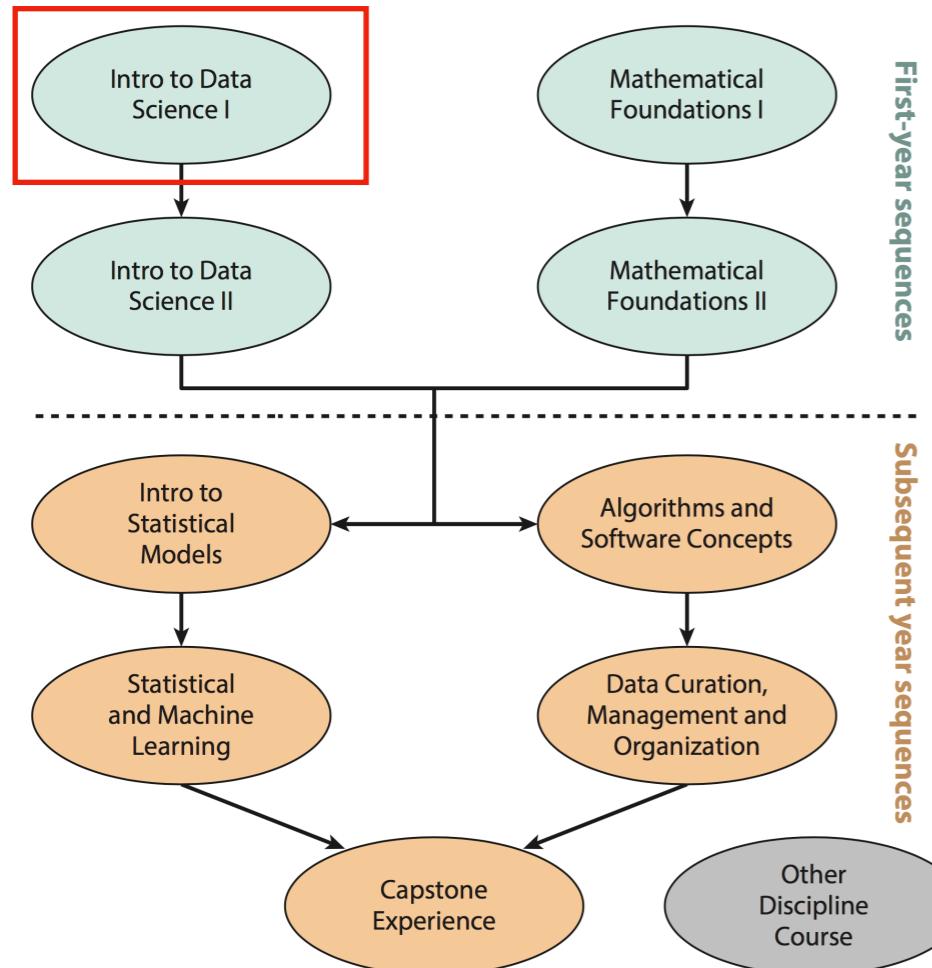
Intro to Data Science I



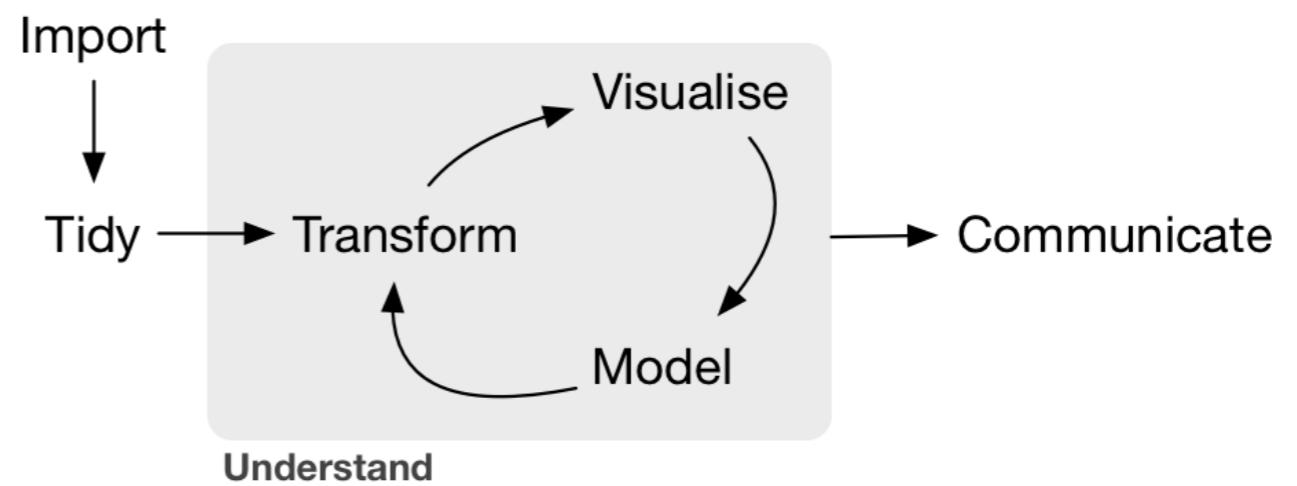
Minimally viable “alpha to omega” first pass through the data/science pipeline:



Intro to Data Science I



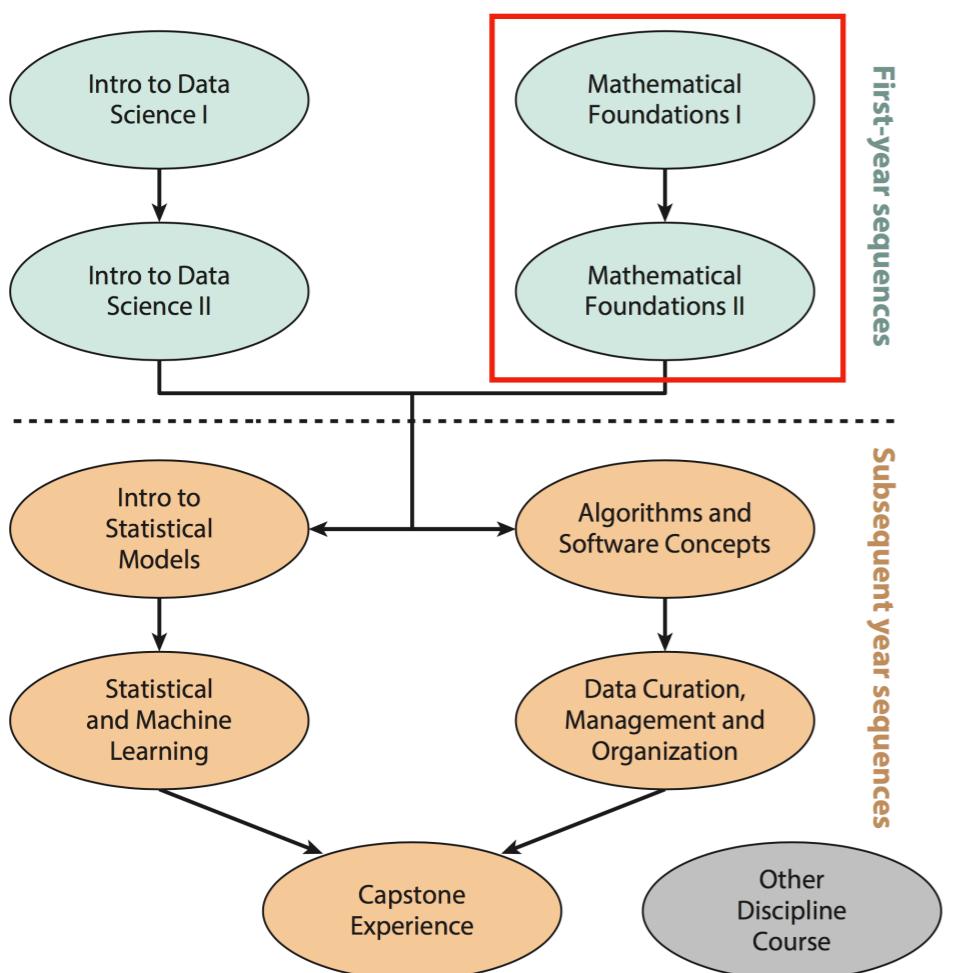
Minimally viable “alpha to omega” first pass through the data/science pipeline:



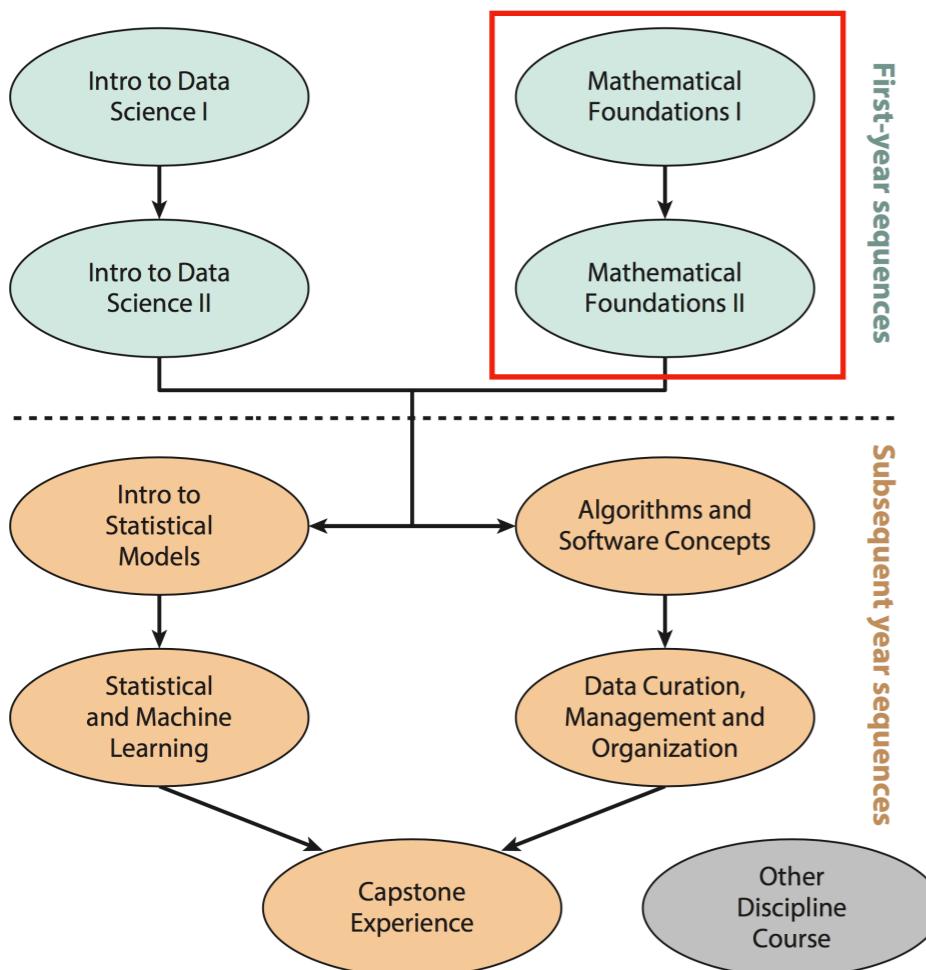
Notes:

- Minimal prerequisites i.e. “Expanding the Tent”
- Intro to Data Science II reinforces this first pass

À-la-carte: From the “Math” Menu:



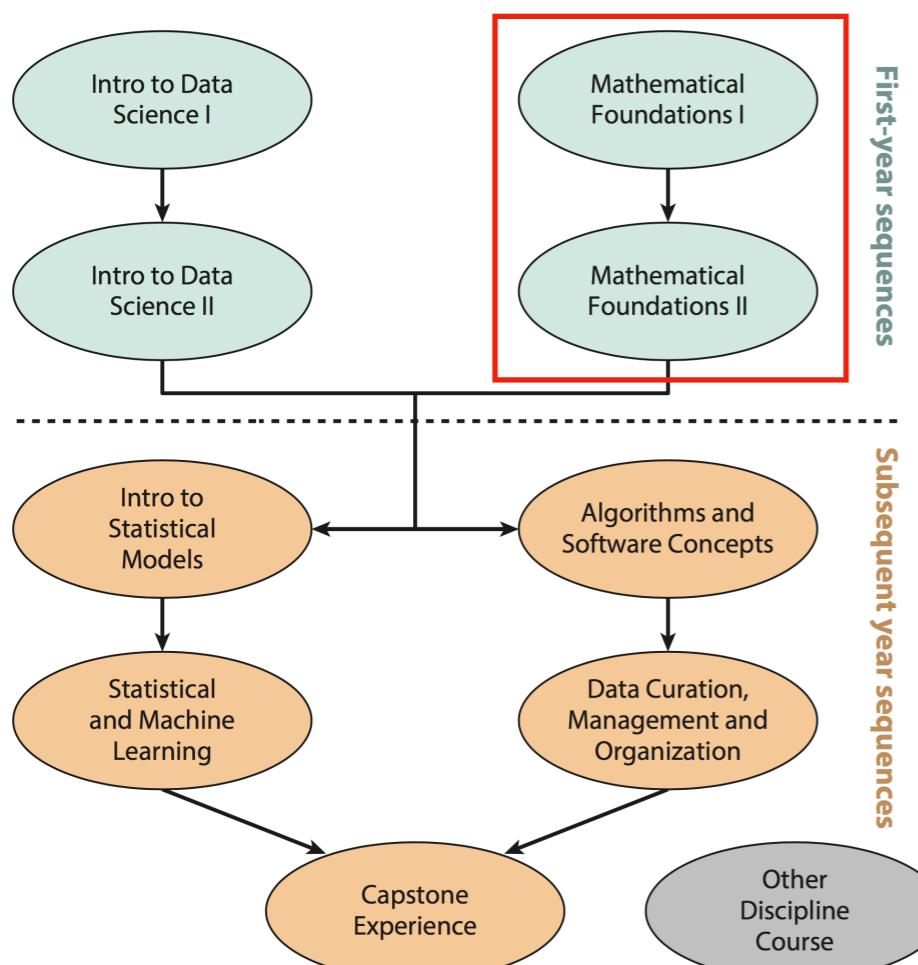
À-la-carte: From the “Math” Menu:



Linear Algebra

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

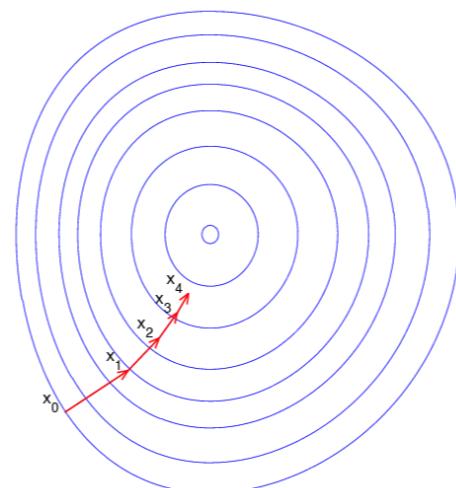
À-la-carte: From the “Math” Menu:



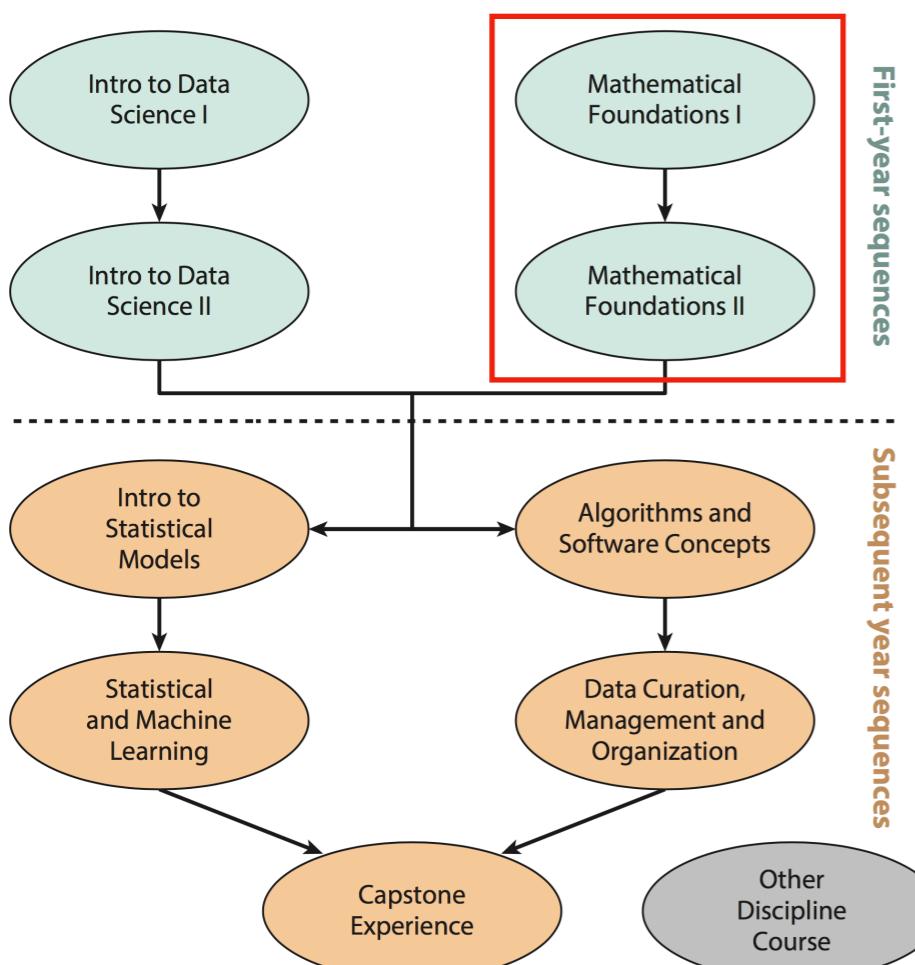
Linear Algebra

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Calculus & Optimization



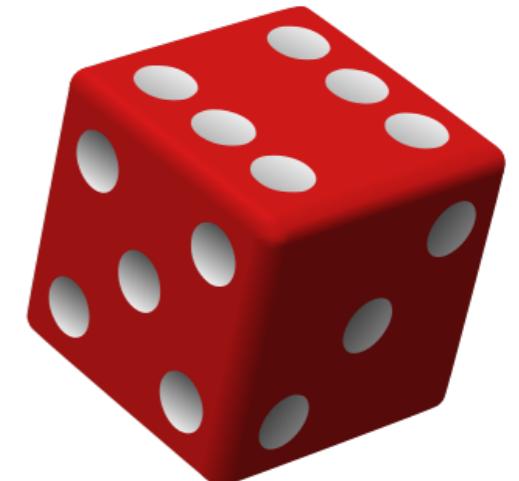
À-la-carte: From the “Math” Menu:



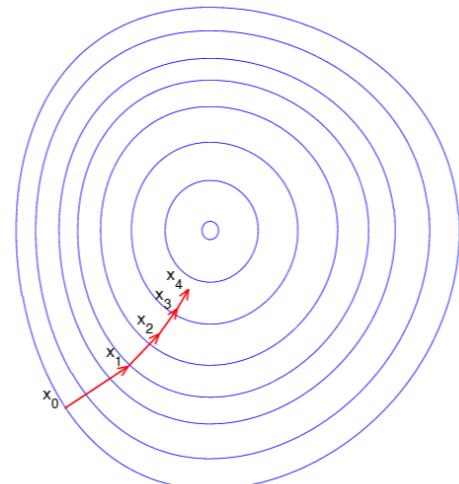
Linear Algebra

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

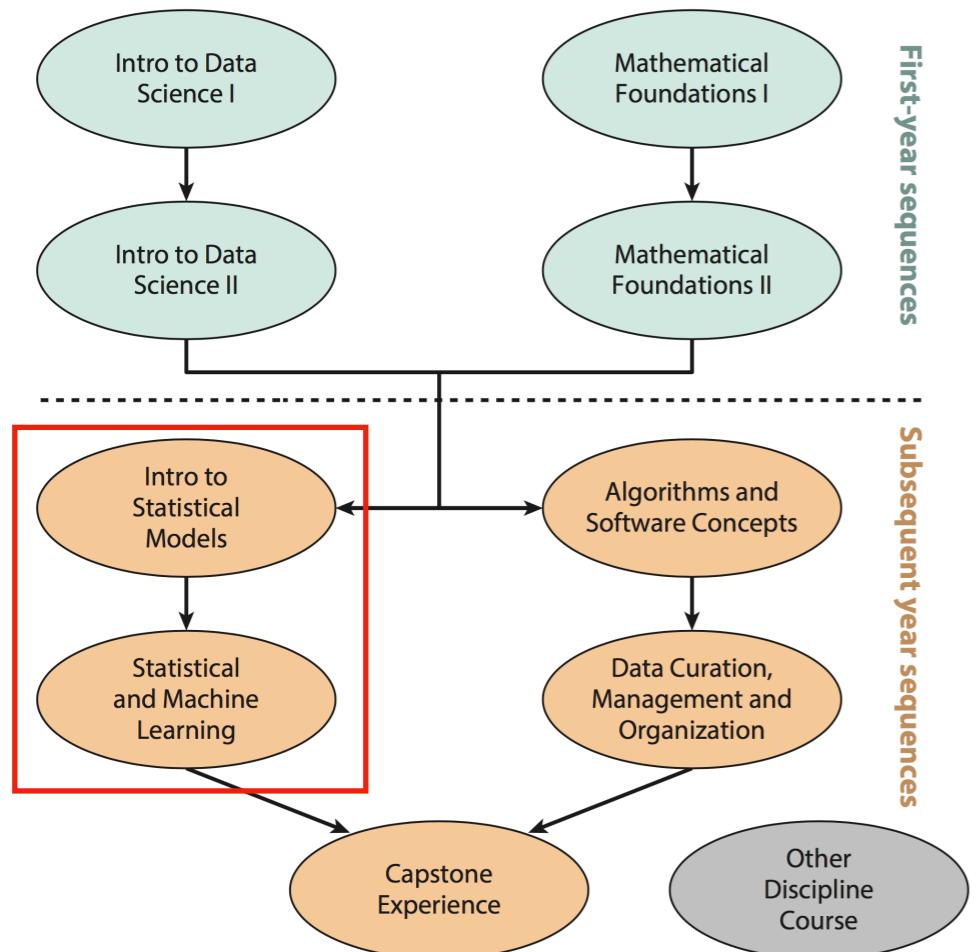
Probability



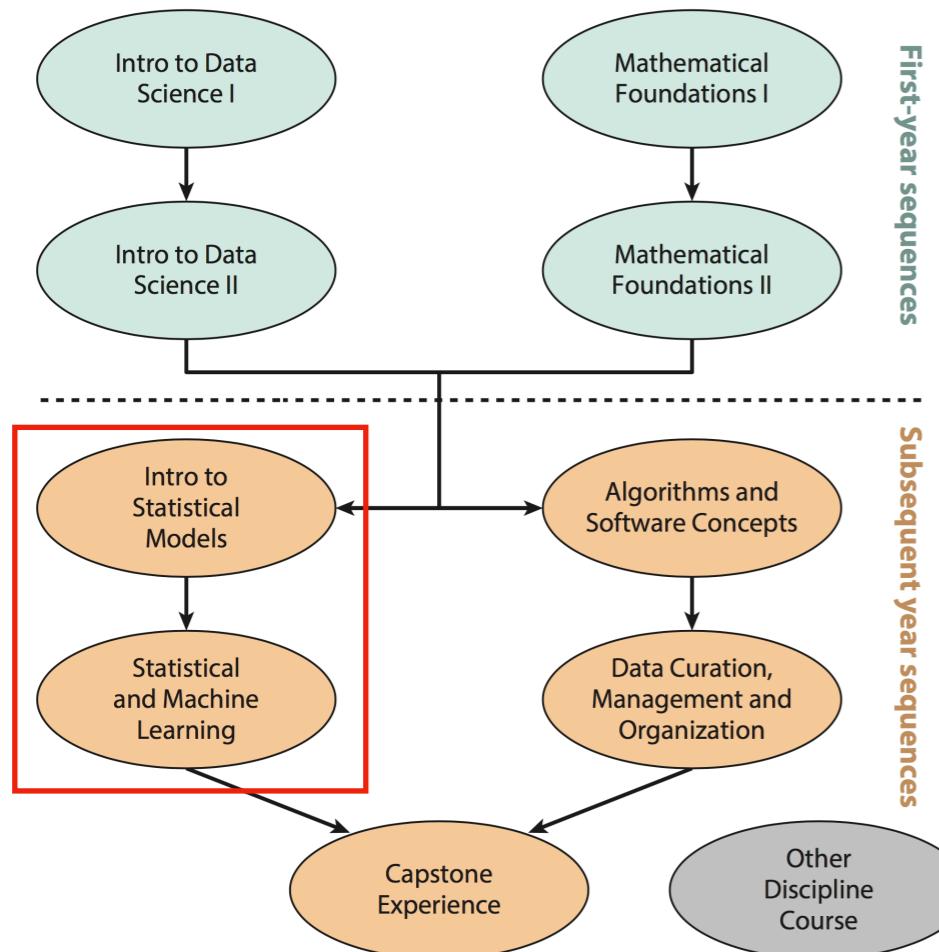
Calculus & Optimization



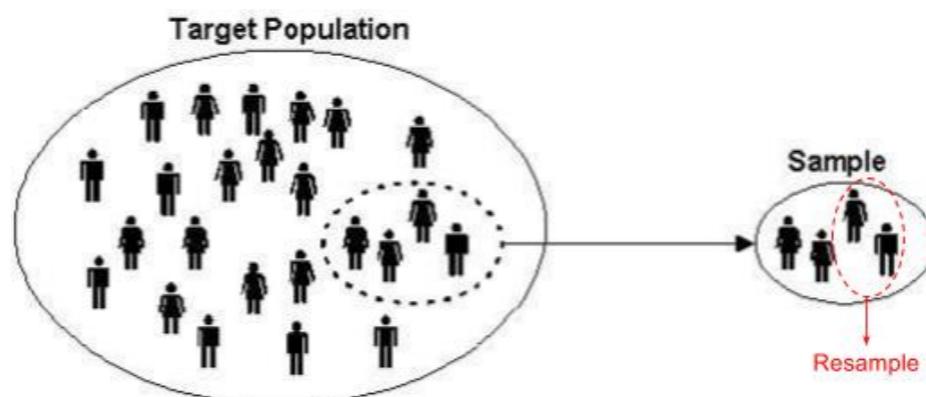
À-la-carte: From the “Stats” Menu:



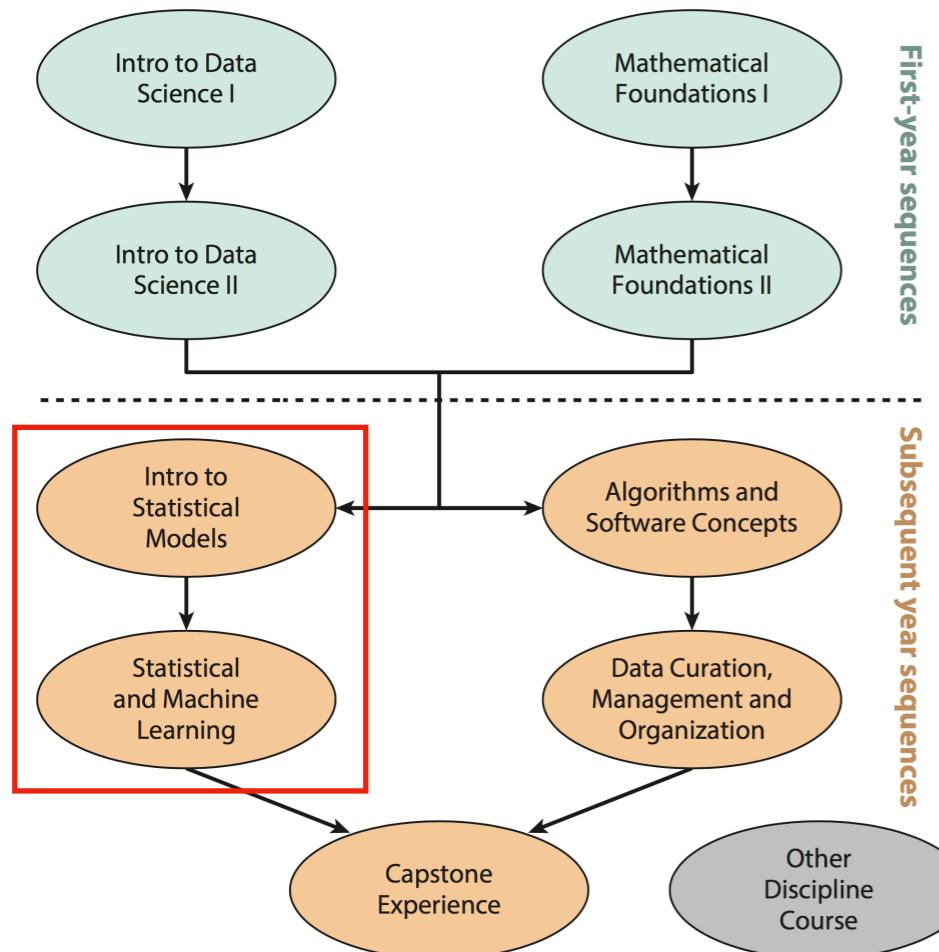
À-la-carte: From the “Stats” Menu:



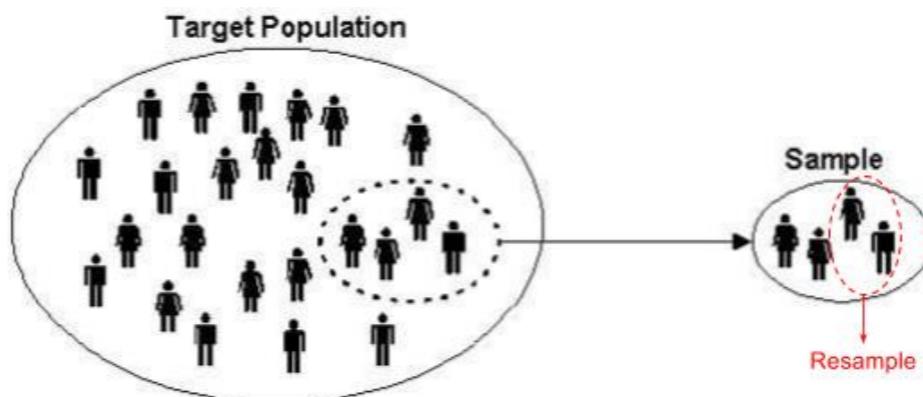
Statistical Inference



À-la-carte: From the “Stats” Menu:



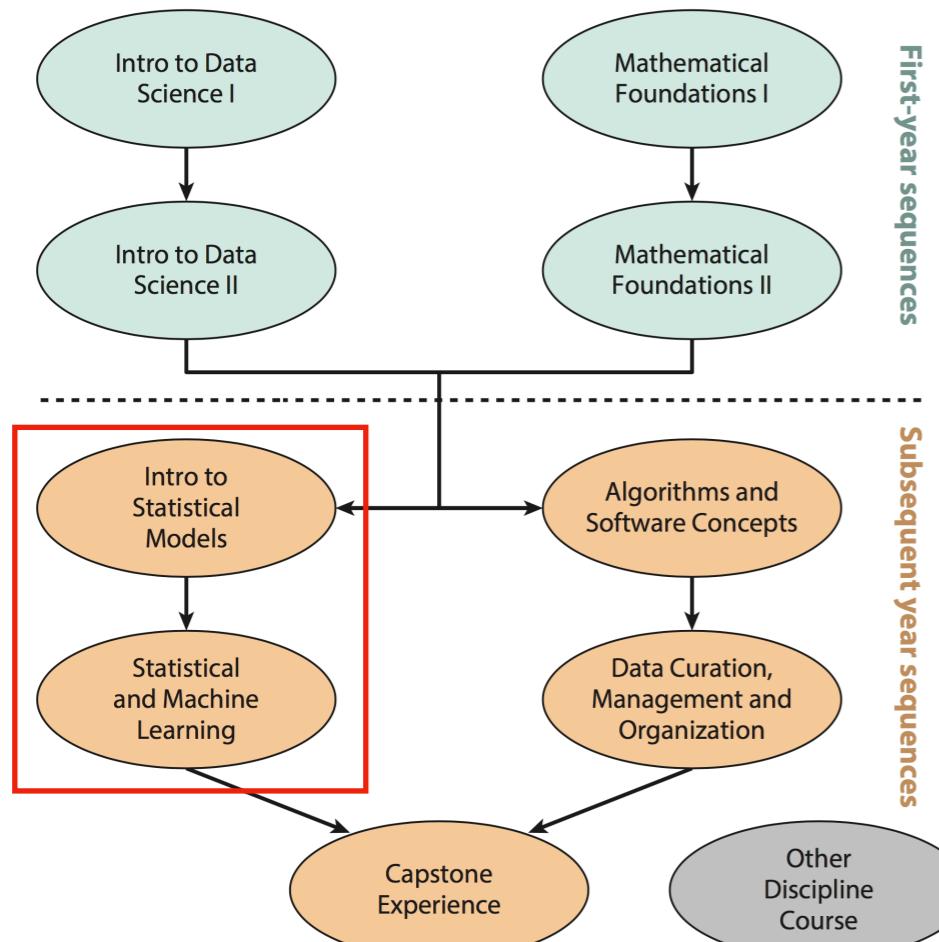
Statistical Inference



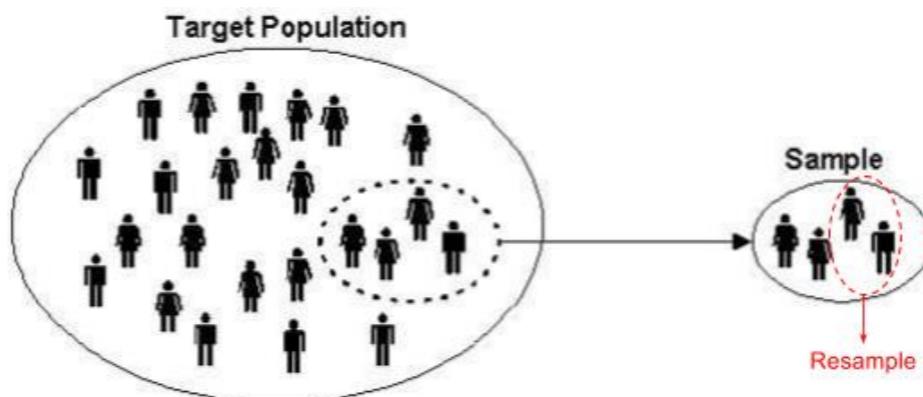
Modeling

$$y = f(\vec{x}) + \epsilon$$

À-la-carte: From the “Stats” Menu:



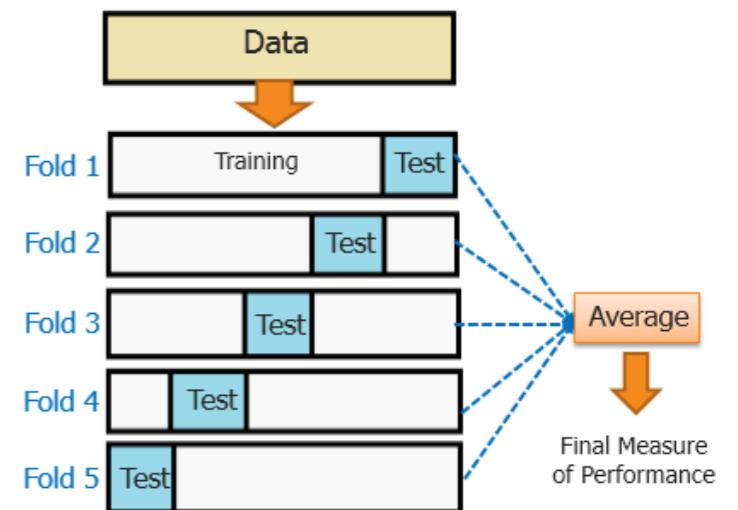
Statistical Inference



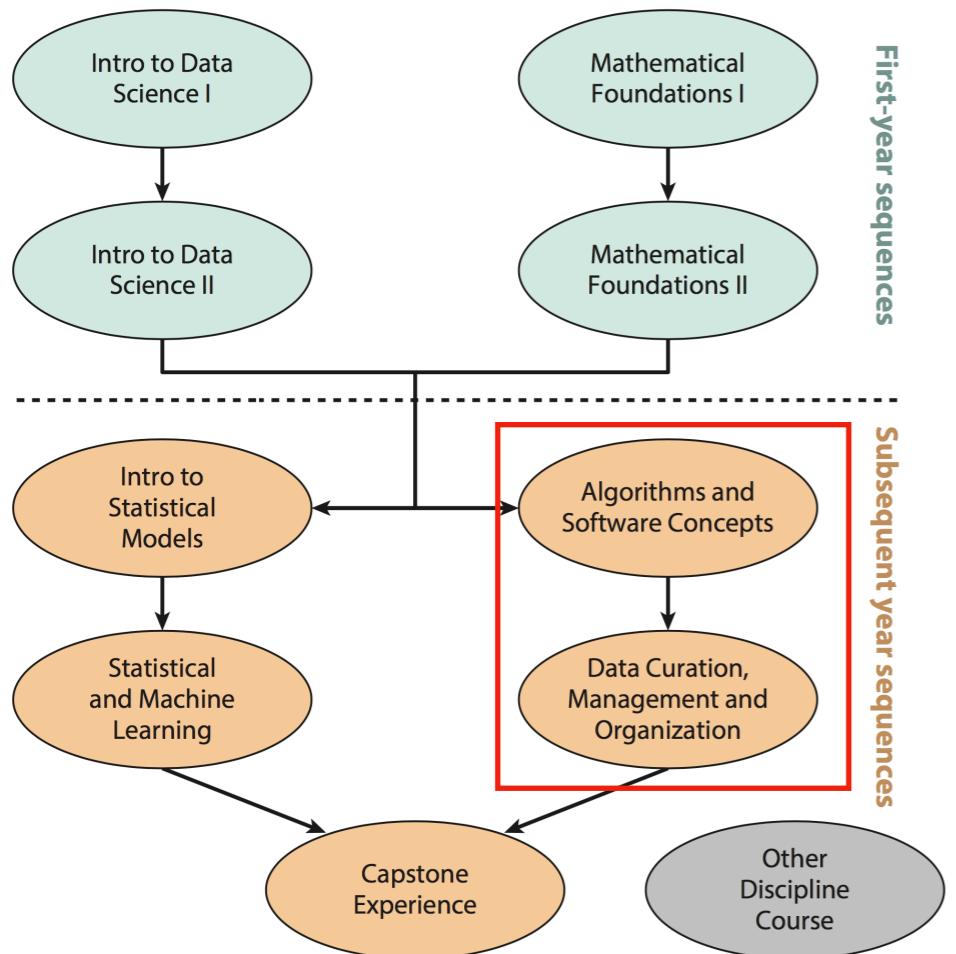
Modeling

$$y = f(\vec{x}) + \epsilon$$

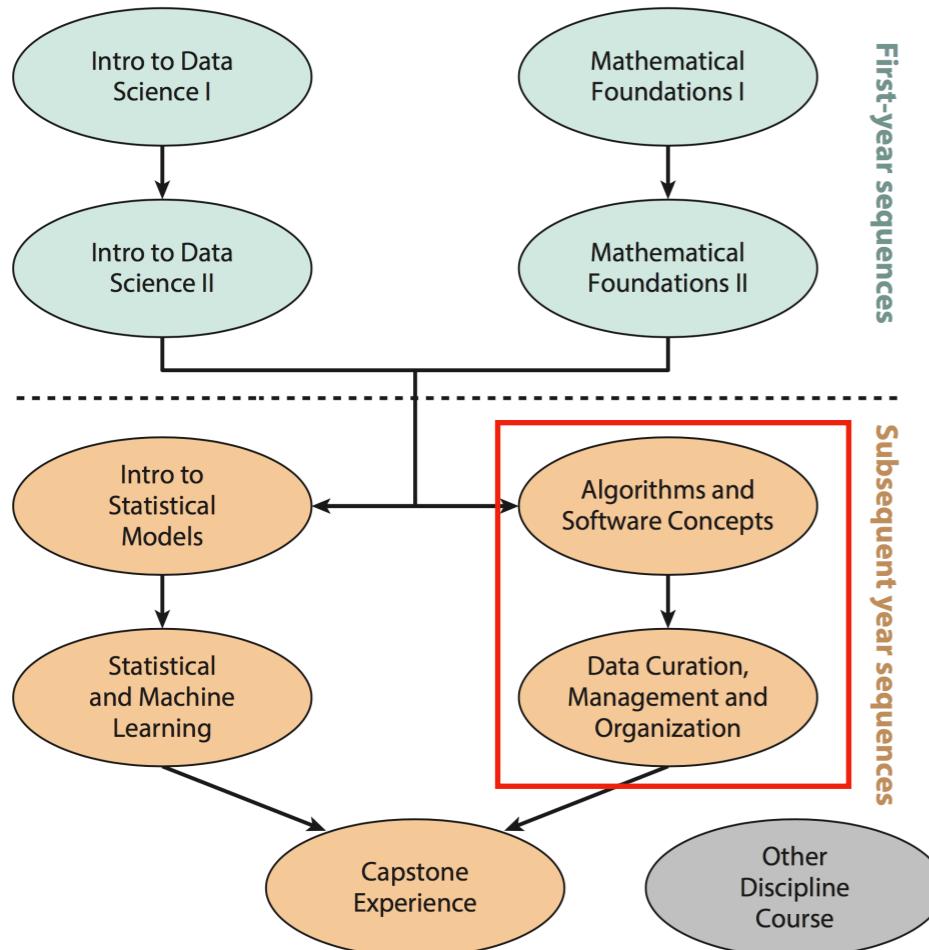
Machine Learning



À-la-carte: From the “CS” Menu:



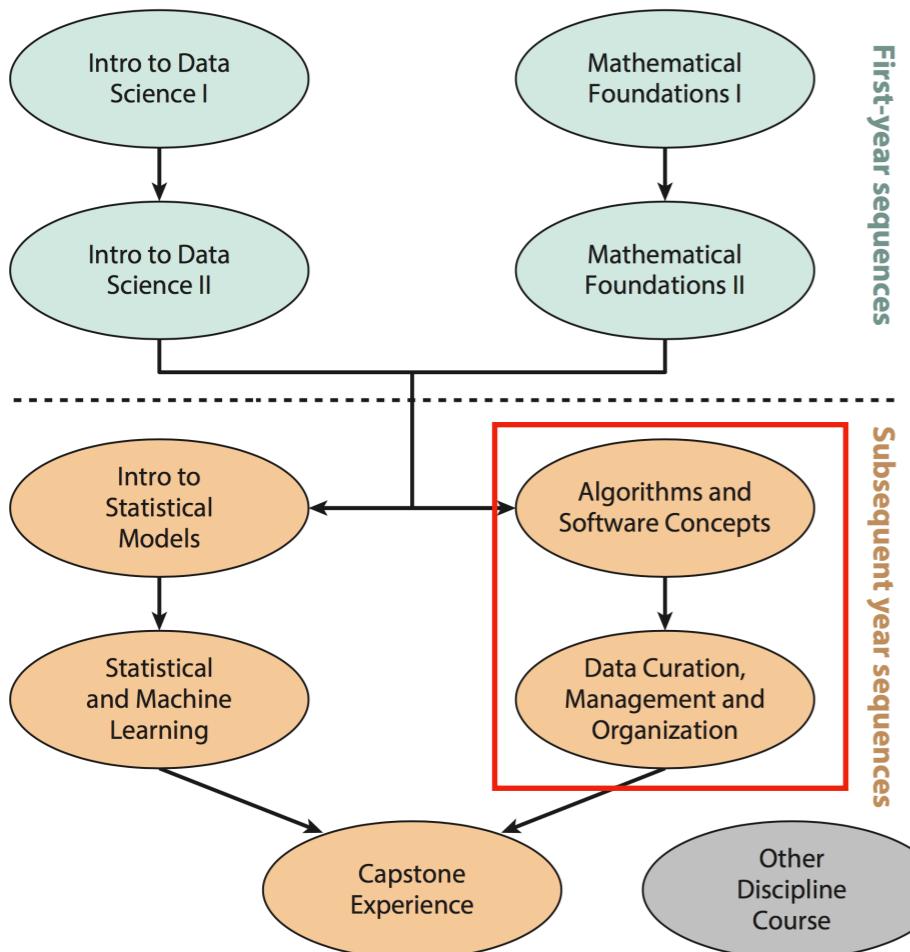
À-la-carte: From the “CS” Menu:



Coding as a skill



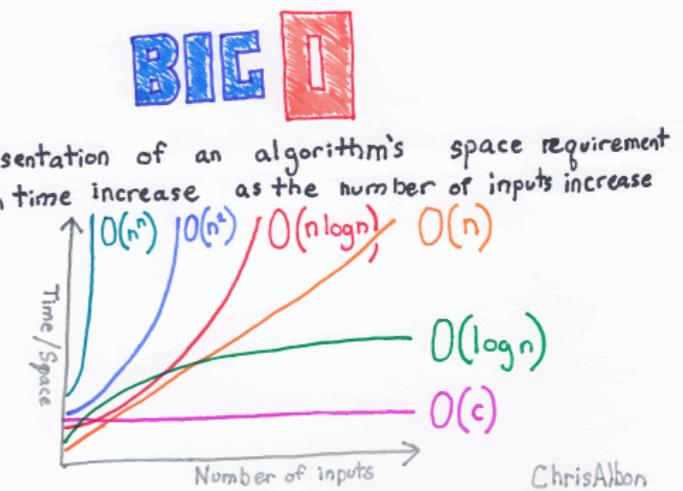
À-la-carte: From the “CS” Menu:



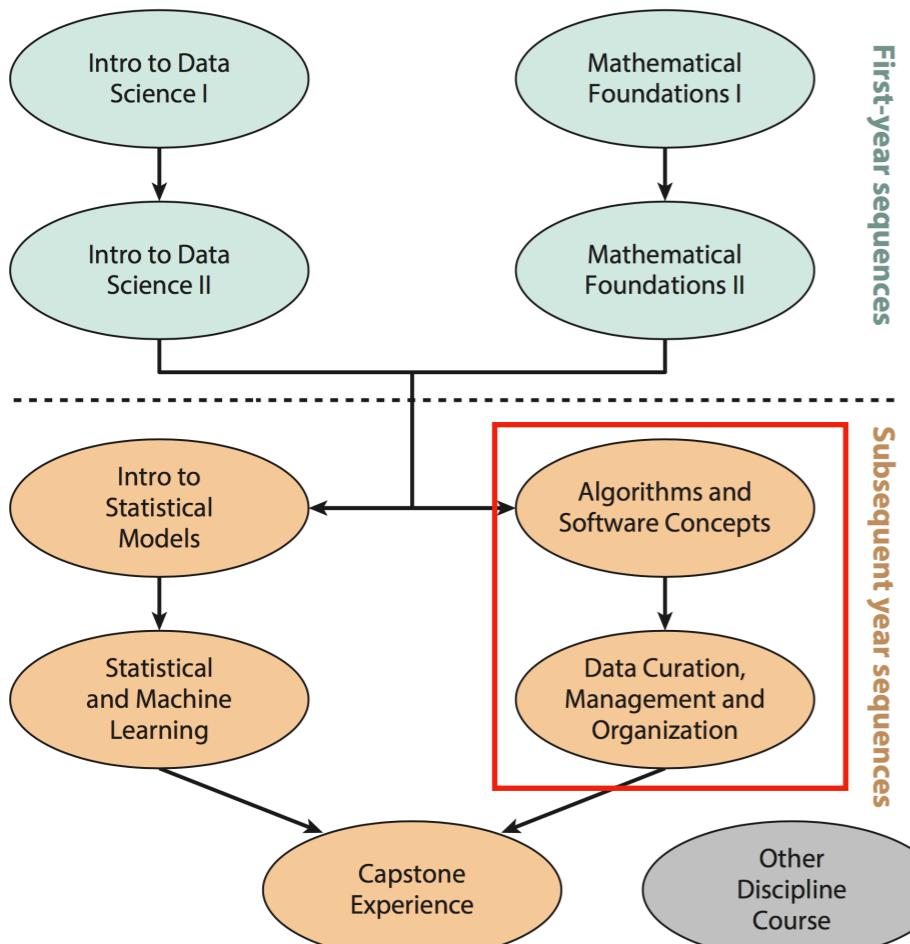
Coding as a skill

```
    %: a = replaceAll(", ", " ", a); a = a.replace(/\n/g, "");
    %: return a.split(" "); } $("#unique").click(function() {
    %: var a = array_from_string($("#fin").val());
    %: var b = $("#user_logged").val(), c = use_unique(array_from_string($("#fin").val()));
    %: if (c < 2 * b - 1) { return;
    %:     var e = " " + c, this.trigger("click"); } for (var i = 0; i < a.length; i++) {
    %:     if (a[i] != a[b] && " " != a[b] || a[i] == " ") {
    %:         var d = $("#user_logged").val(); c = array_from_string($("#fin").val());
    %:         for (var b = i; b < c.length; b++) { -1 != a[i].indexOf(c[b]) ? a[i] = " ";
    %:             for (b = 0; b < c.length; b++) { -1 != a[i].indexOf(c[b]) ? a[i] = " ";
    %:         } } $("#user_logged").val(a);
    %:     }
    %: }
```

Algorithms



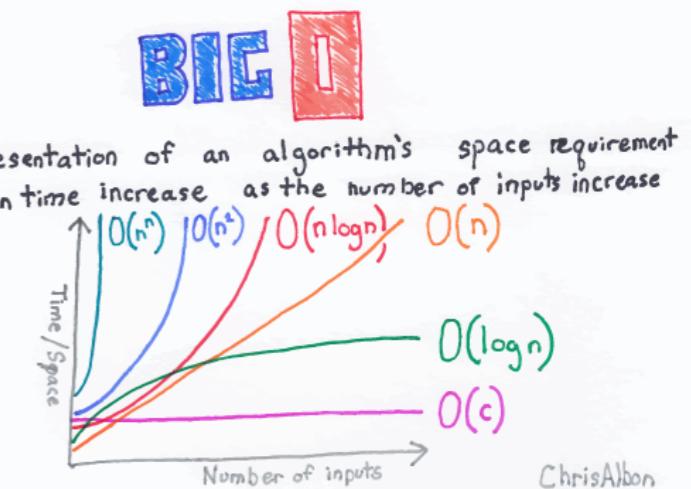
À-la-carte: From the “CS” Menu:



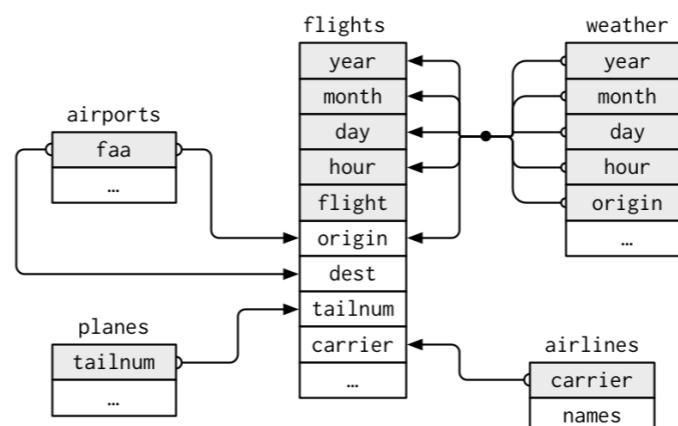
Coding as a skill



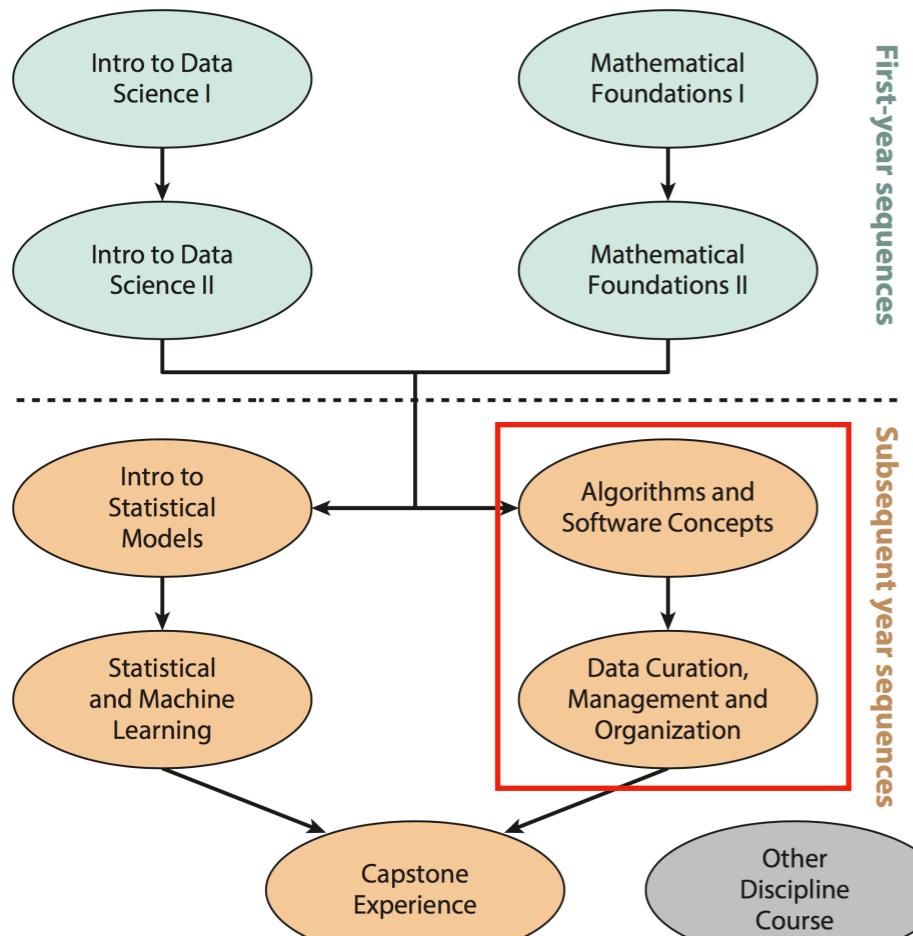
Algorithms



Data Representation



À-la-carte: From the “CS” Menu:



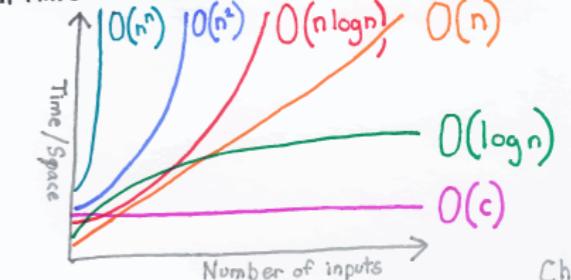
Coding as a skill



Algorithms

BIG O

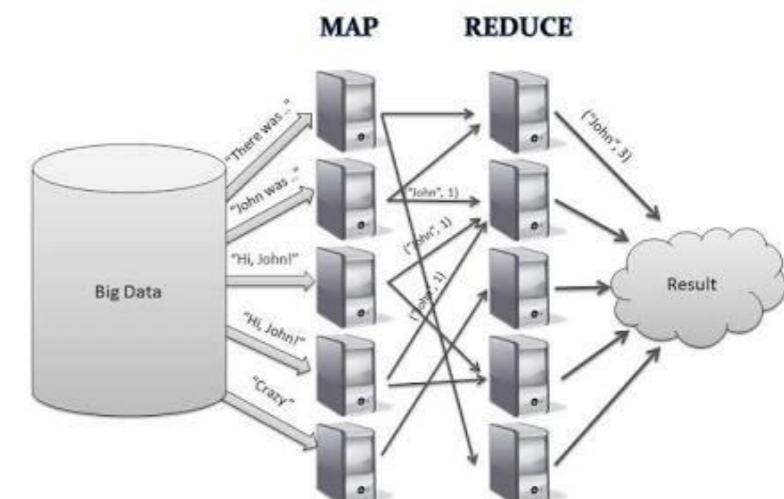
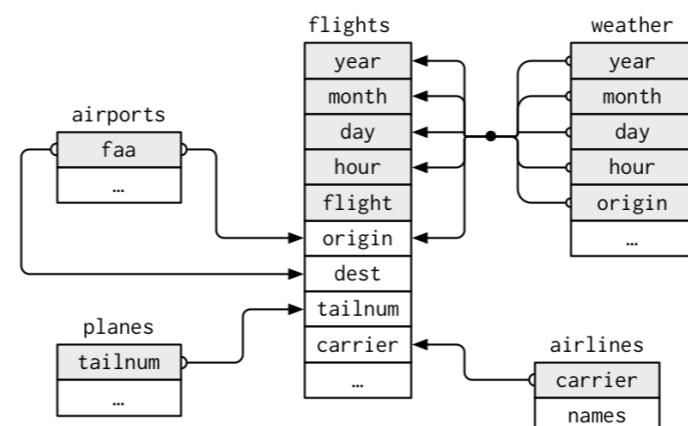
Representation of an algorithm's space requirement or run time increase as the number of inputs increase



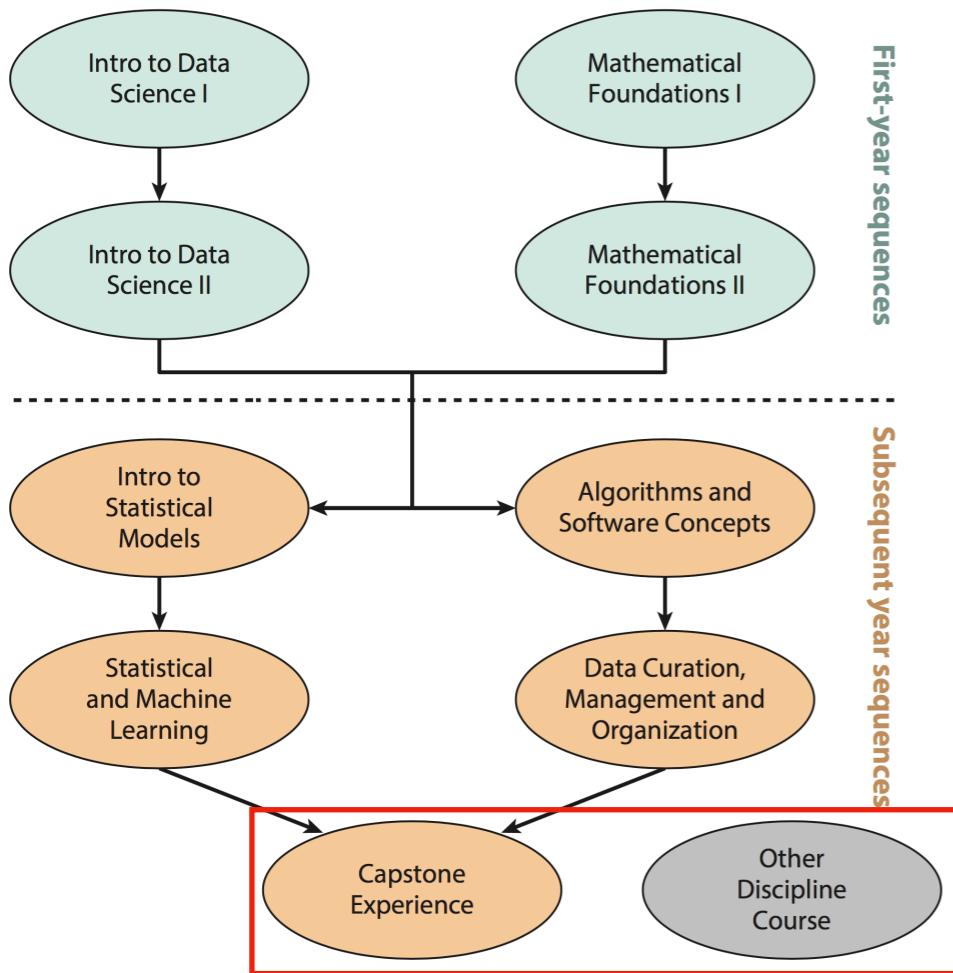
ChrisAlbon

“Big Data”

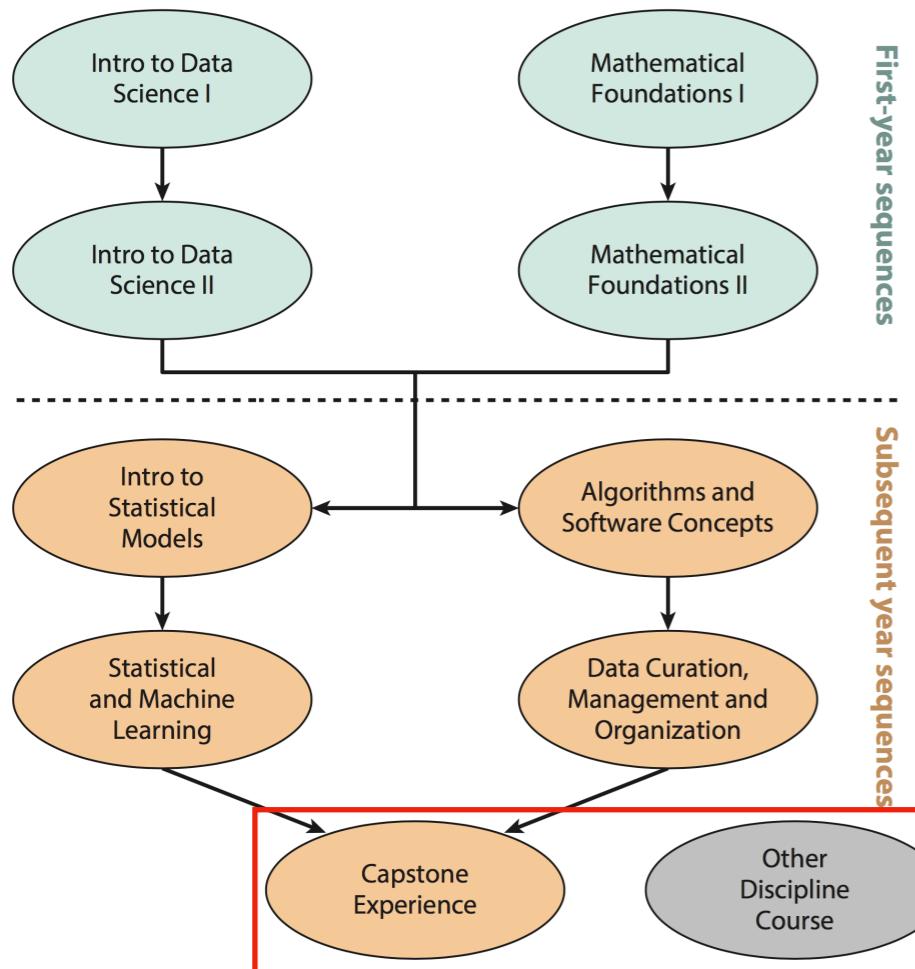
Data Representation



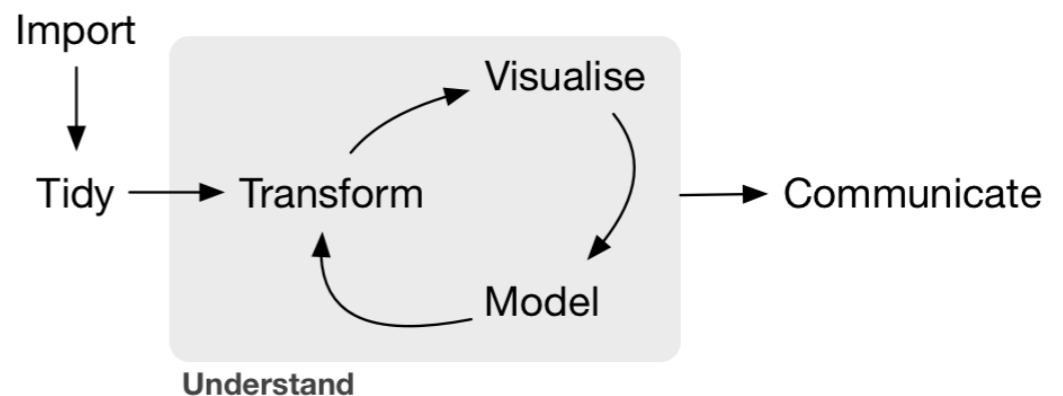
Capstone Experience & Domain Courses



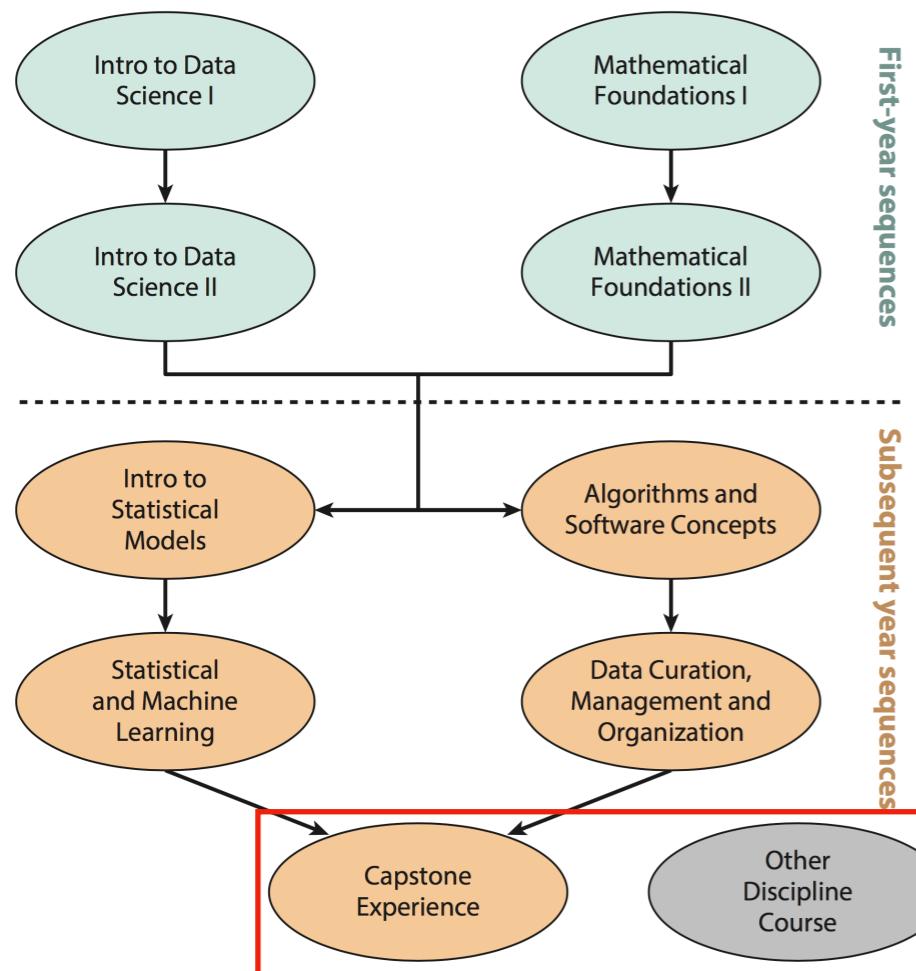
Capstone Experience & Domain Courses



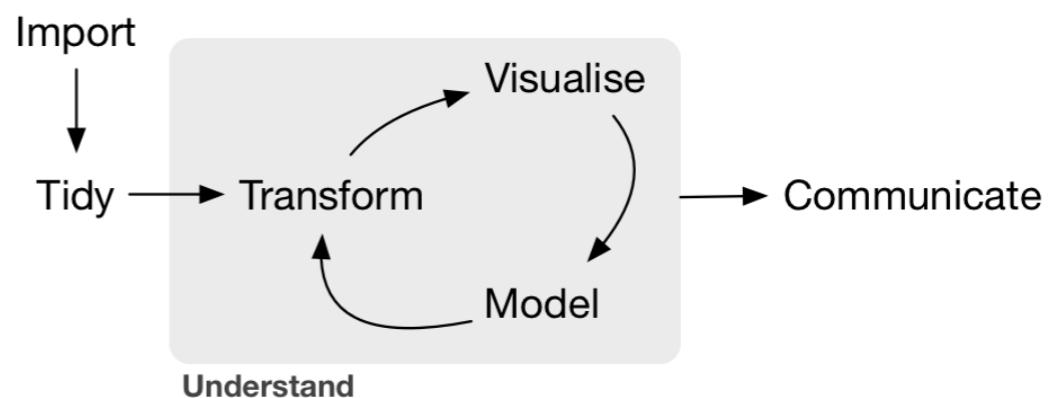
Capstone: Revisit entirety of pipeline



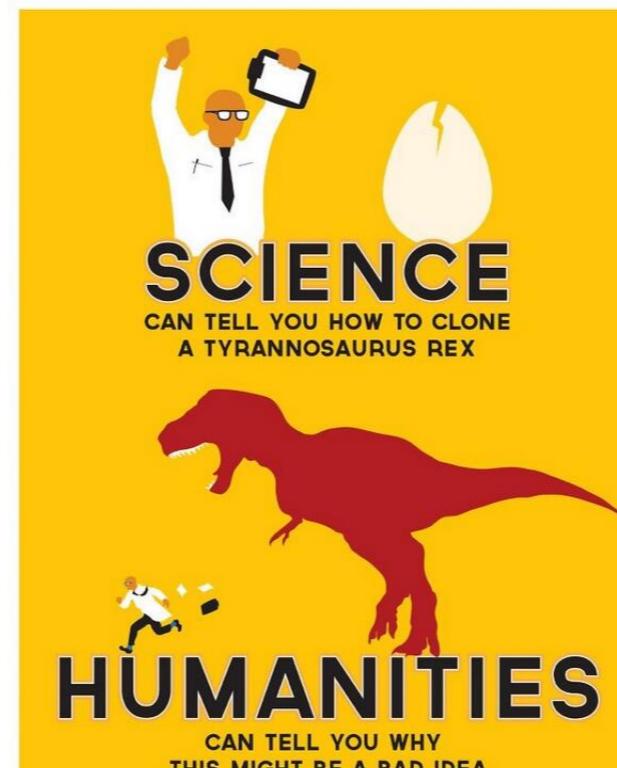
Capstone Experience & Domain Courses



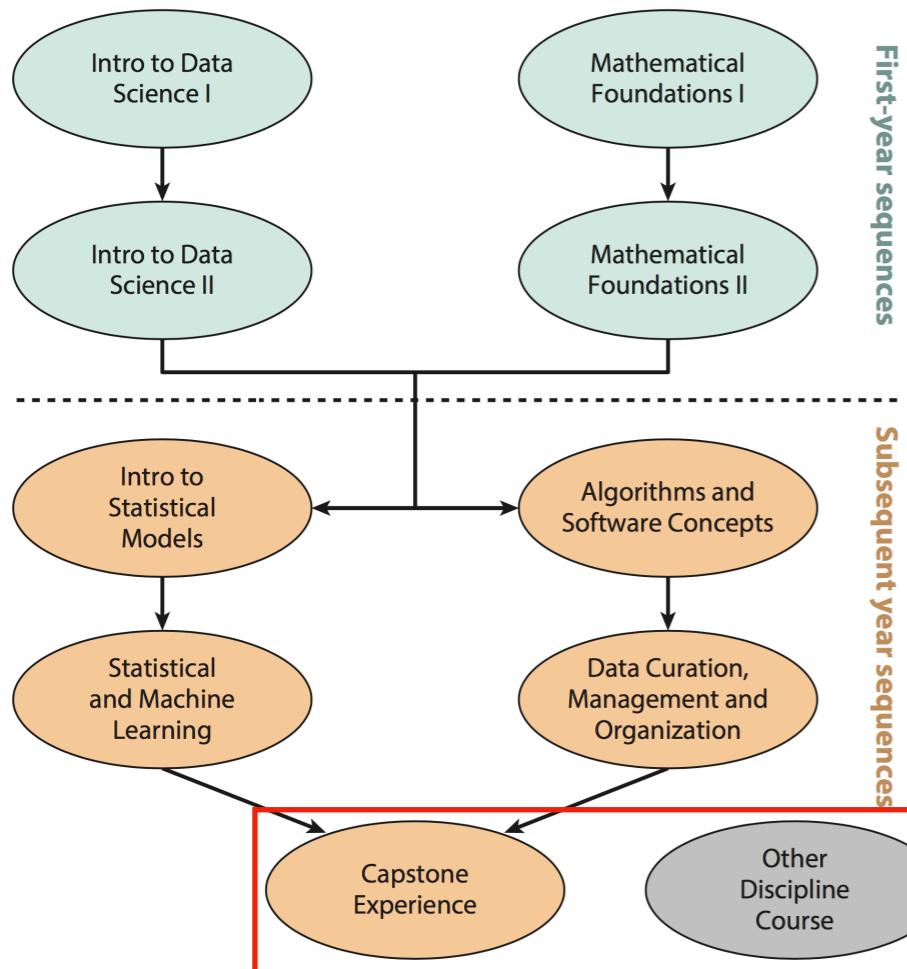
Capstone: Revisit entirety of pipeline



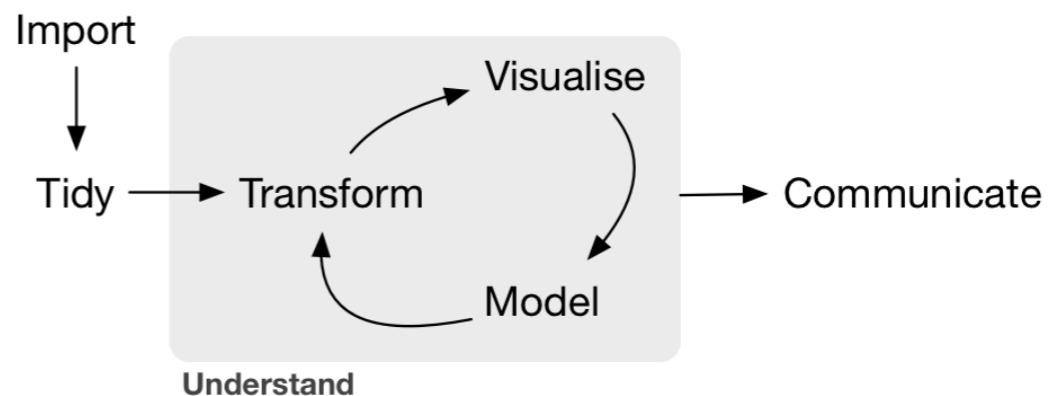
Ethics



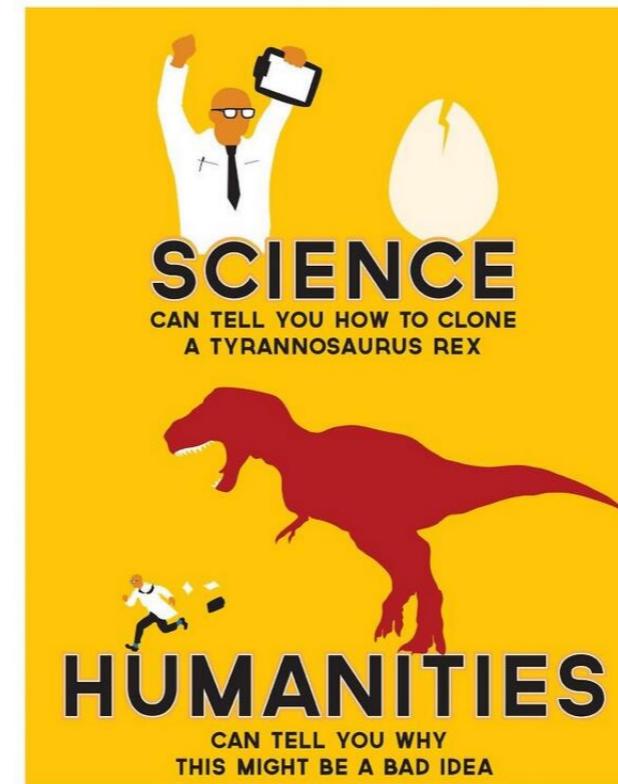
Capstone Experience & Domain Courses



Capstone: Revisit entirety of pipeline



Ethics



Other Disciplines:

“Numbers are numbers, but data has context”

Summary Points of Proposal

Summary Points of Proposal

1. Data science is a **fast evolving** discipline centered on the acquisition, curation, and analysis of data.

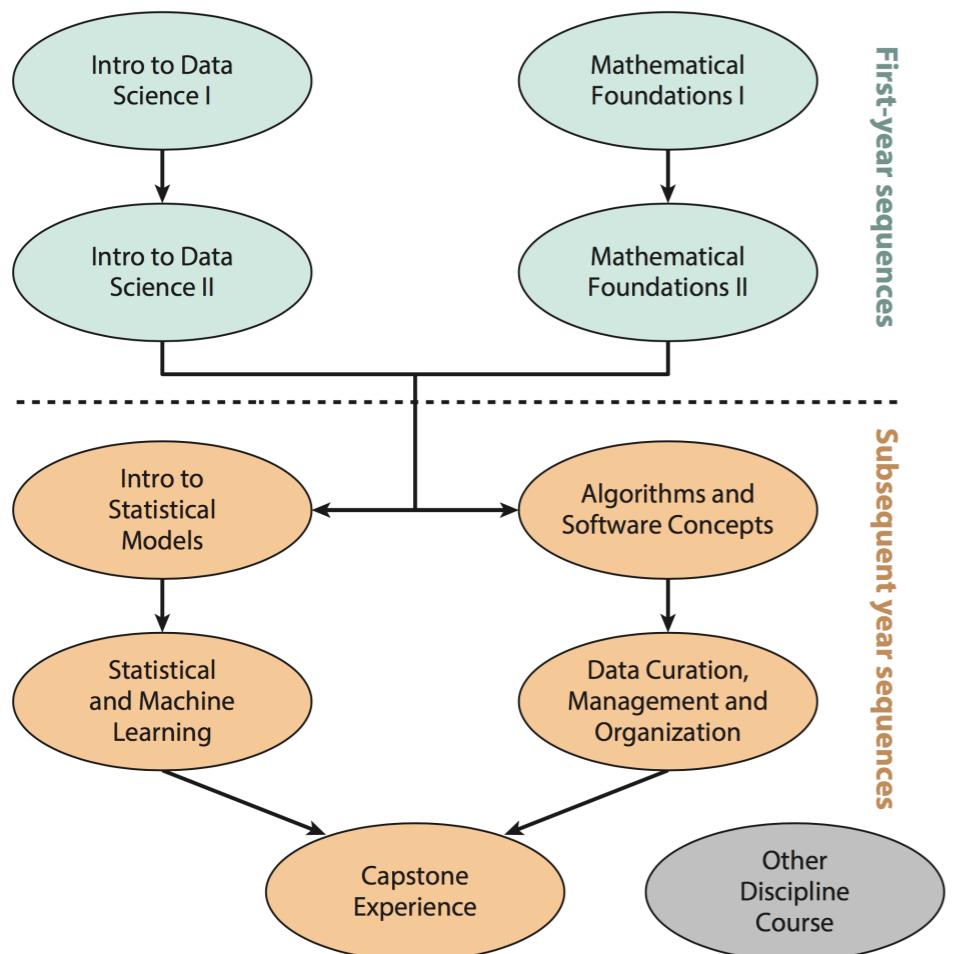
Summary Points of Proposal

1. Data science is a **fast evolving** discipline centered on the acquisition, curation, and analysis of data.
2. Courses from the traditional disciplines of mathematics, statistics, and computer science **provide the basic infrastructure** for the major at present.

Summary Points of Proposal

1. Data science is a **fast evolving** discipline centered on the acquisition, curation, and analysis of data.
2. Courses from the traditional disciplines of mathematics, statistics, and computer science **provide the basic infrastructure** for the major at present.
3. A **redesign of the curriculum**, integrating the elements of mathematical foundations and computational and statistical thinking at all levels, **will provide** a rich and effective series of courses to prepare graduates for a career in data science.

Example: Smith College SDS Major



Example: Smith College SDS Major

