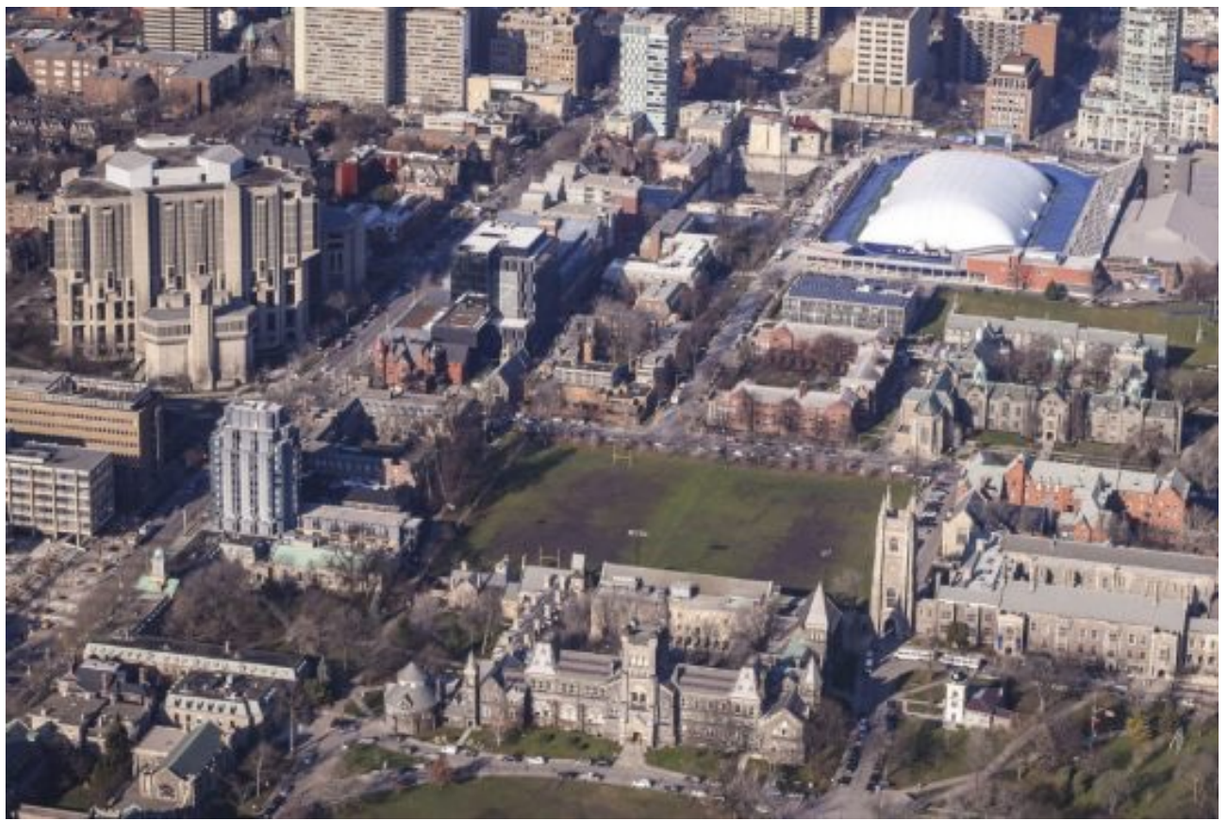
TEACHING STATISTICS IN THE ERA OF DATA SCIENCE

Alison L. Gibbs
Department of Statistical Sciences
University of Toronto

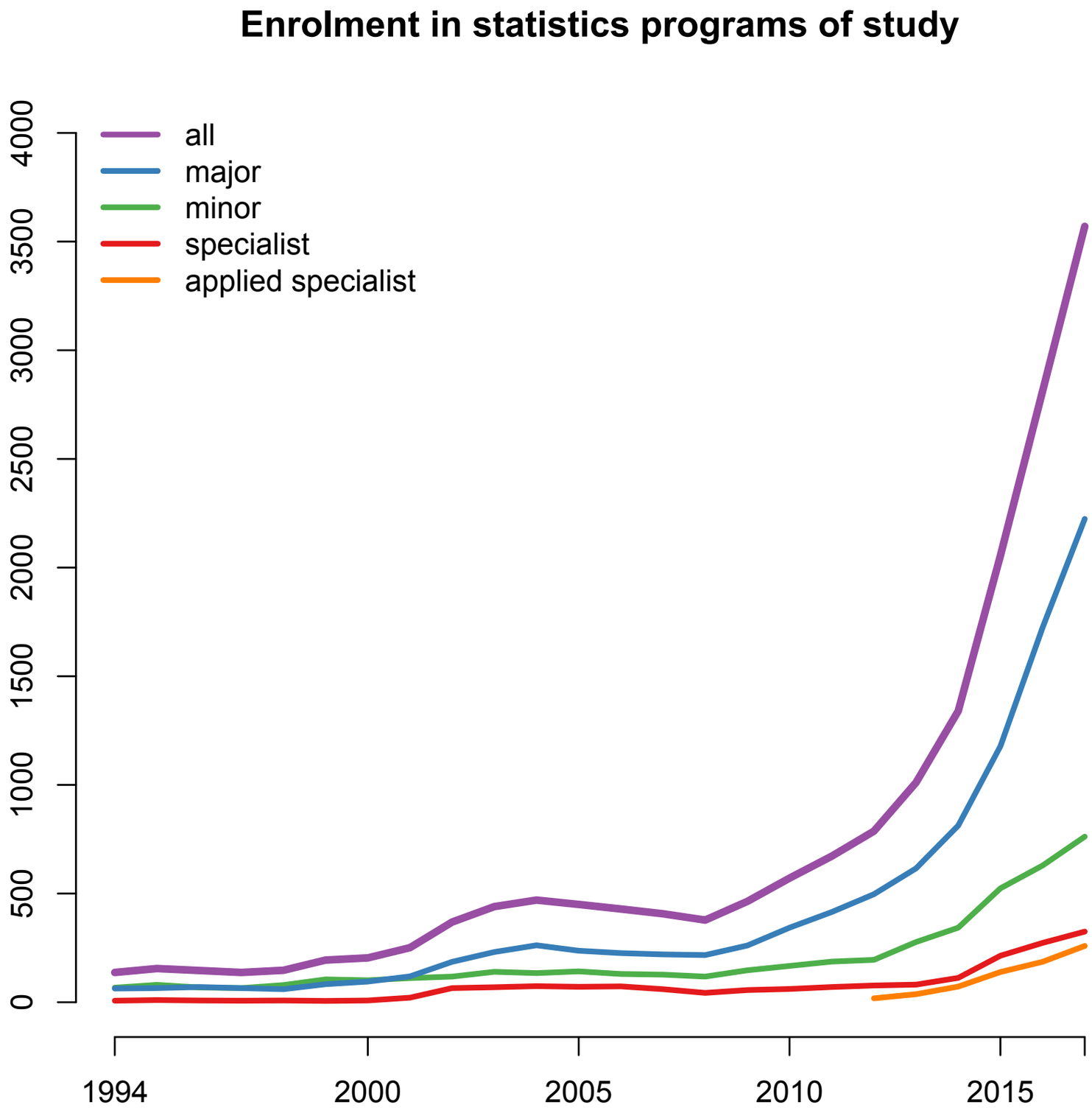
ICOTS 10

THE UNIVERSITY OF TORONTO

- **old** (founded in 1827)
- **big** (71,500 undergrads, 18,500 graduate students, 14,300 faculty)
- **public** research university in an urban setting



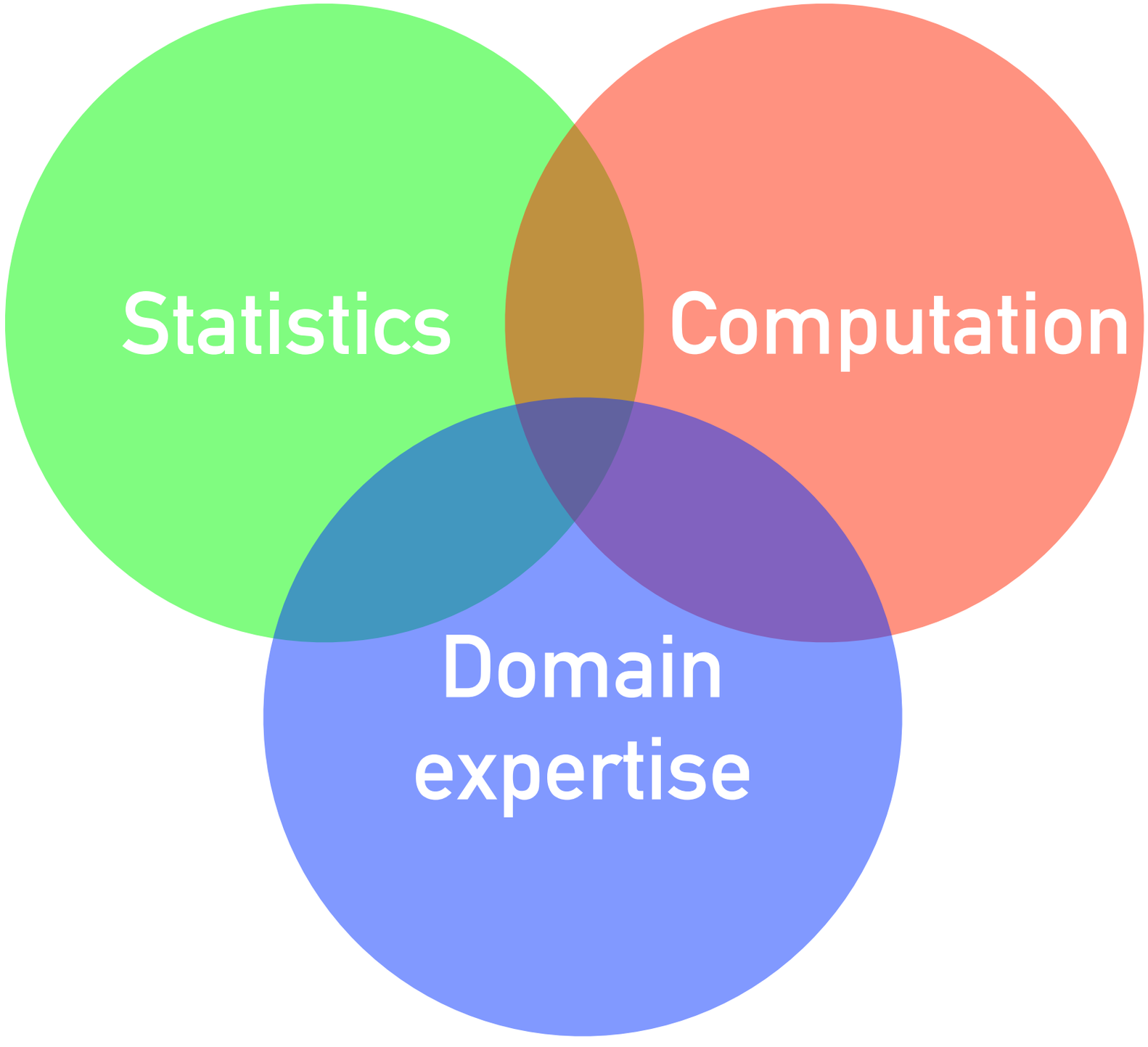
UNDERGRADUATES IN STATISTICS PROGRAMS OF STUDY AT U OF T



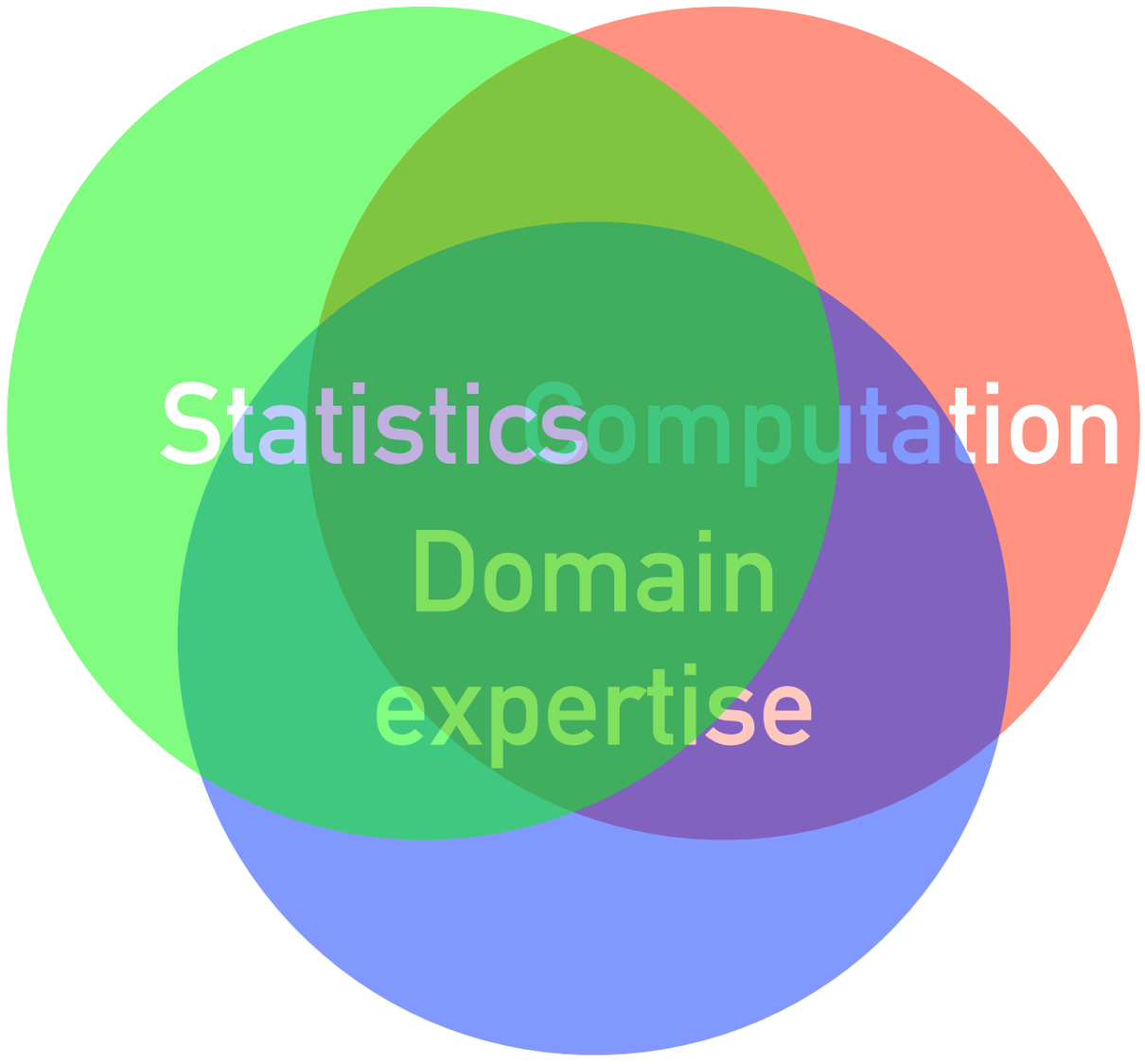
~60% international

OUR STATISTICS PROGRAMS OF STUDY?

The Data Science
Venn Diagram



OUR NEW VIEW OF OUR STATISTICS PROGRAMS OF STUDY



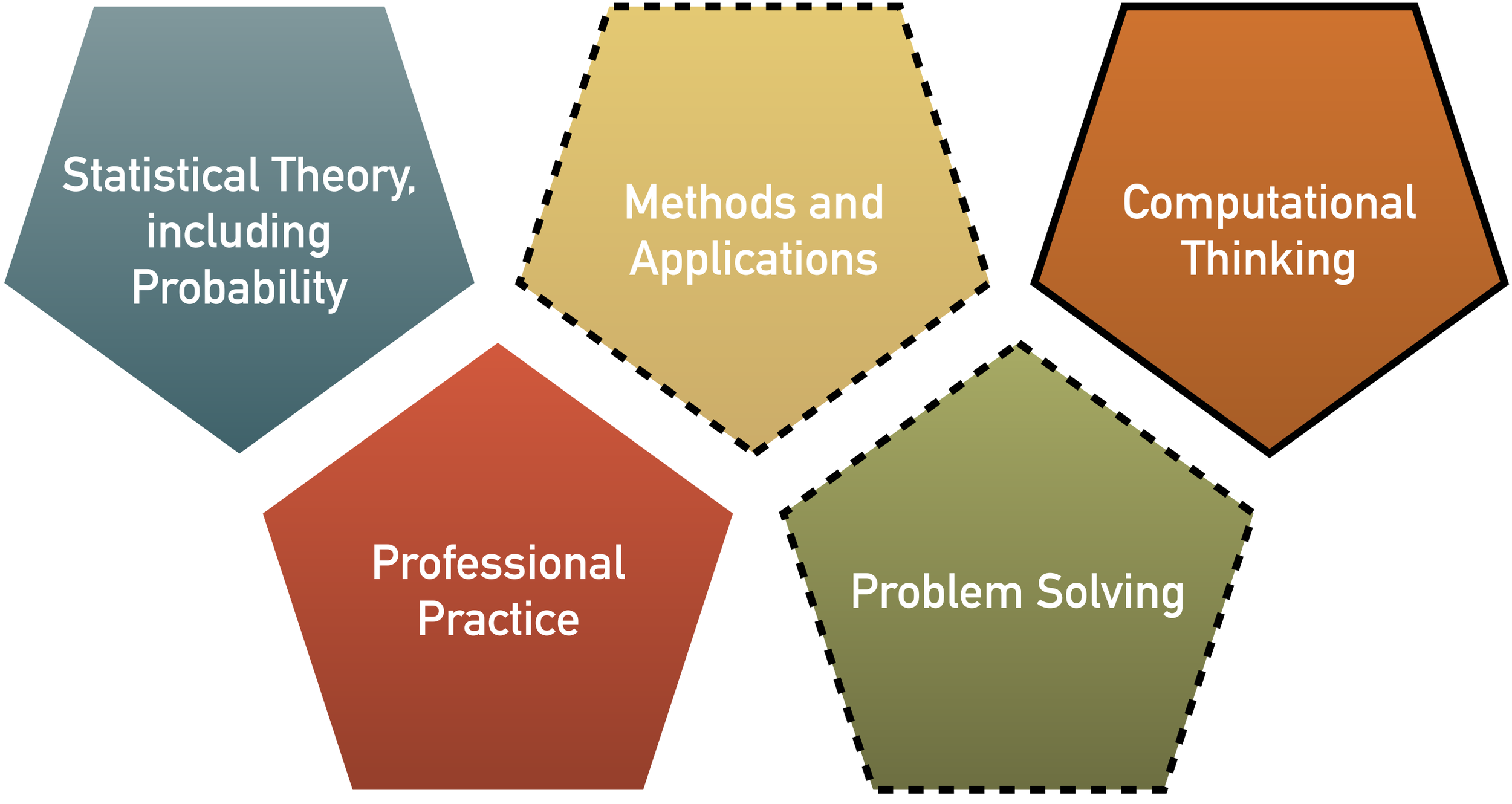
A rigorous mindset plus…

More computation and algorithmic thinking

More real problems

OUR STATISTICS PROGRAMS OF STUDY

Program Learning Outcomes organized into 5 themes:



THE MOST IMPORTANT THING WE NEED OUR GRADUATES TO TAKE WITH THEM

“

Never become so much of an expert that you stop gaining expertise. View life as a continuous learning experience.

-Denis Waitley

DEVELOPING ADAPTIVE EXPERTISE

> A subject area is ultimately about the *doing* of a subject — using the content in a disciplined way … [not] a march through content.
>
> *-Wiggins & McTighe (2005)*

DEVELOPING ADAPTIVE EXPERTISE

- **Essential questions** encourage making sense of knowledge and skills accumulated in a holistic and integrated way
- To help our students develop from routine to adaptive experts, the adaptive learning literature suggests we need to spend more time on:
  - creative problem solving in novel situations
  - learning by making errors
  - exploration and discovery
  - building knowledge, forming strategies on their own
  - socially significant contexts

How should our students start?

OUR FIRST COURSE: SOME GUIDING PRINCIPLES

- A curriculum gives opportunities to be **introduced to** / reinforce / master new concepts
- To develop adaptive expertise:
  - need learning to take place in contexts that are socially significant
  - need safer spaces to make errors and create and explore
- Because the first year of university is critically important for student success:
  - resource allocation has to reflect its importance
  - need to enhance the chances of students developing a supportive network of peers

(Yorke and Longden, 2008)

OUR FIRST COURSE: A SURVEY COURSE

- Our traditional starting point, build foundations first: calculus and linear algebra; then probability and mathematical statistics with emphasis on estimation and inference
- New starting point, survey course: inculcation of the discipline, its core ways of thinking, types of arguments used
- Developing understanding of the essential questions of a discipline requires rethinking and constant re-application
- Content can be varied, as long as all of the program learning outcome themes are introduced
- Is a statistics program of study right for me?

OUR FIRST COURSE: CONTENT

- Designed to introduce students to all of the themes in our program learning outcomes
- Engage deeper (re-think, re-apply, peek under the hood) later
- Immersion in real interesting problems to motivate
- Particular emphasis on communication
- Content:
  - Computing basics
  - Visualization (EDA)
  - Data wrangling
  - Comparing two groups (randomization test)
  - Modeling for prediction, description and inference (regression)
  - Statistical issues: multiple testing, confounding
  - Supervised learning for prediction (trees)
  - Ethical practice

OUR FIRST COURSE: THE ROLE OF CODING

- Our students have a required course in programming from Computer Science
- For this course, we want them to be able to use coding to understand data from code snippets they can adapt
- Want them to be useful for solving real problems as soon as possible
- **Develop good habits (implicitly)**
- Choice: R / R Studio / R Markdown / tidyverse

OUR FIRST COURSE: THE STRUCTURE

Mondays

- Large lecture sections
- Data stories
- A new idea or method each week
- R by example to answer data questions

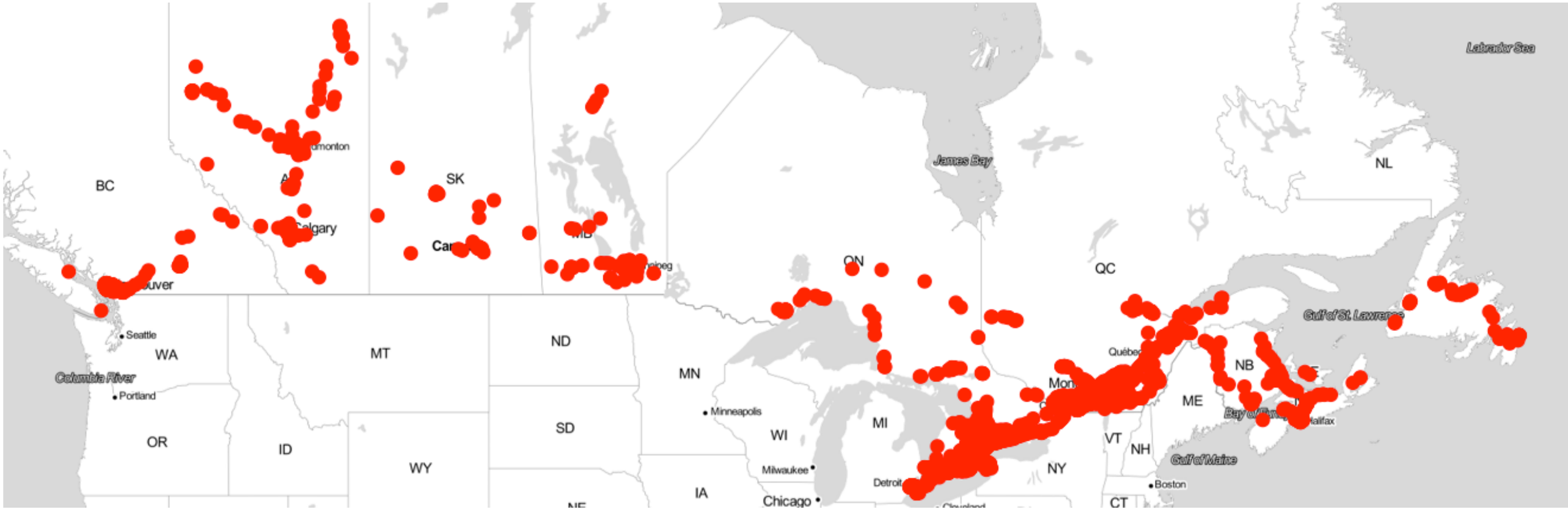
- Practice problems
- R Markdown template to start solutions
- One solution brought to tutorial for grading and discussion

Fridays

- Small group tutorials
- Random group assignments each week
- Oral or written communication exercise each week

Project

THE FINAL PROJECT 2018: HAZARDOUS DRIVING INCIDENTS



Given harsh braking, accident incidents, and an index measuring severity, location characteristics:

- What is hazardous driving?
- Where is there more hazardous driving?

Opportunity to creatively explore data, develop their own questions, in a context that is socially significant

SOME OBSERVATIONS FROM EXPERIENCE

- Students perceive it is a course about R, rather than a course about statistics or data science
- We're trying to do a lot; and needed to scale back
- **Appropriately connecting questions to methods and conclusions doesn't come naturally**
- ~70% of students self-assess as having achieved well the program learning outcomes addressed in the course

What's next for the 2nd course?

Course materials:

sta130.utstat.utoronto.ca

“

You must continue to gain expertise, but avoid thinking like a *[routine]* expert.

-Denis Waitley