TEACHING STATISTICS IN A DATA SCIENCE WORLD

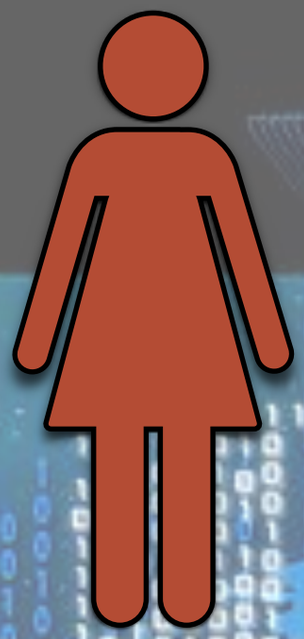
Alison L. Gibbs
Department of Statistical Sciences
University of Toronto

University of Manitoba
Department of Statistics
March 21, 2019

A DATA SCIENCE WORLD?

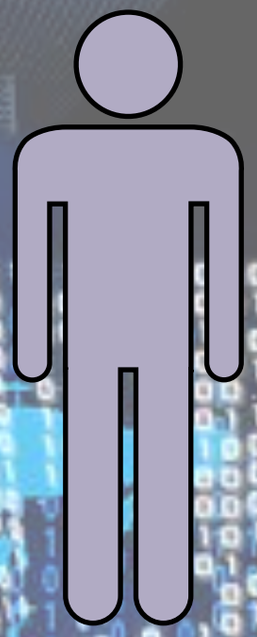
A DATA SCIENCE WORLD?

Some of the inhabitants



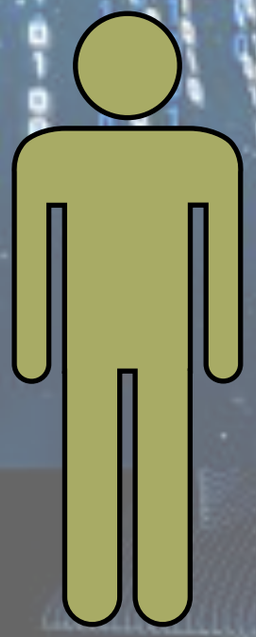
Computer scientist
2014

Data capitalist
2016



A "DATA SCIENCE" WORLD?

Some of the inhabitants



Laboratory scientist
2017

GOOGLE SEARCHES FOR "DATA SCIENCE" SINCE 2004

trends.google.com

Google Trends | Explore

Sign in

data science
Search term

+ Compare

United States | 2004 - present | All categories | Web Search

Interest over time

GOOGLE SEARCHES FOR "DATA SCIENCE" SINCE 2004

trends.google.com

Google Trends | Explore

Sign in

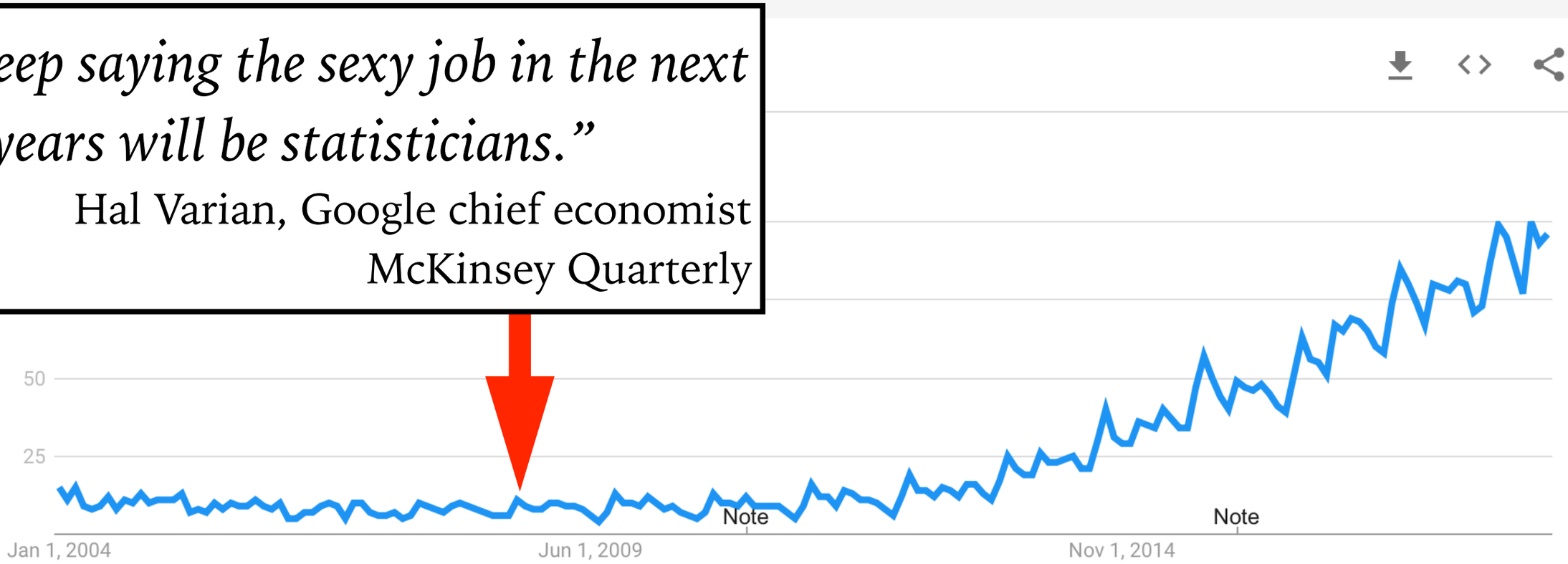
data science
Search term

+ Compare

United States | 2004 - present | All categories | Web Search

"I keep saying the sexy job in the next ten years will be statisticians."

Hal Varian, Google chief economist
McKinsey Quarterly



GOOGLE SEARCHES FOR "DATA SCIENCE" SINCE 2004

trends.google.com

Google Trends | Explore

Sign in

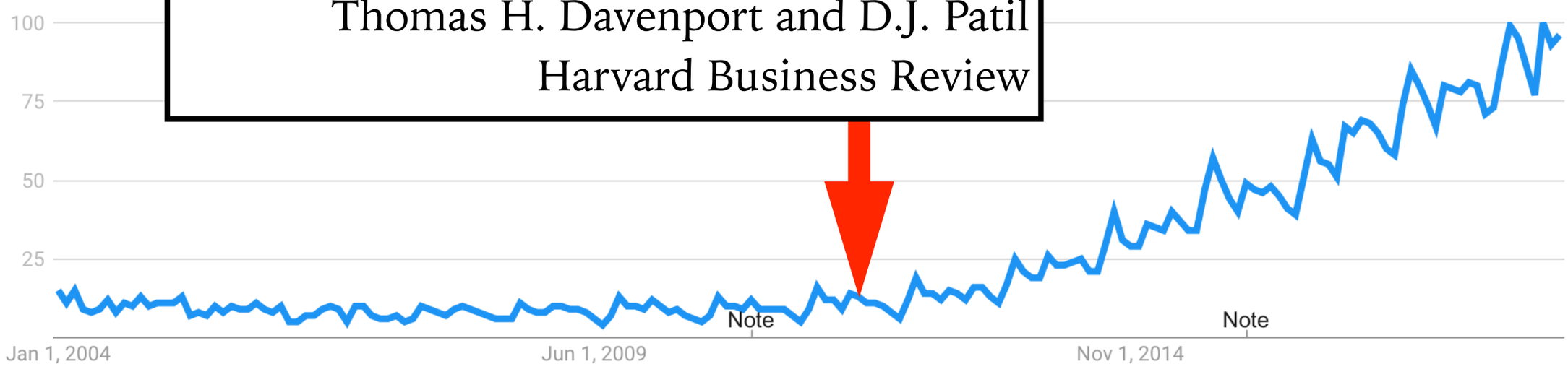
data science
Search term

+ Compare

United States | 2004 - present | All categories | Web Search

Interest over t

"Data Scientist: The Sexiest Job of the 21st Century"
Thomas H. Davenport and D.J. Patil
Harvard Business Review



GOOGLE SEARCHES FOR "DATA SCIENCE" & "STATISTICS" SINCE 2004

trends.google.com

Google Trends | Compare

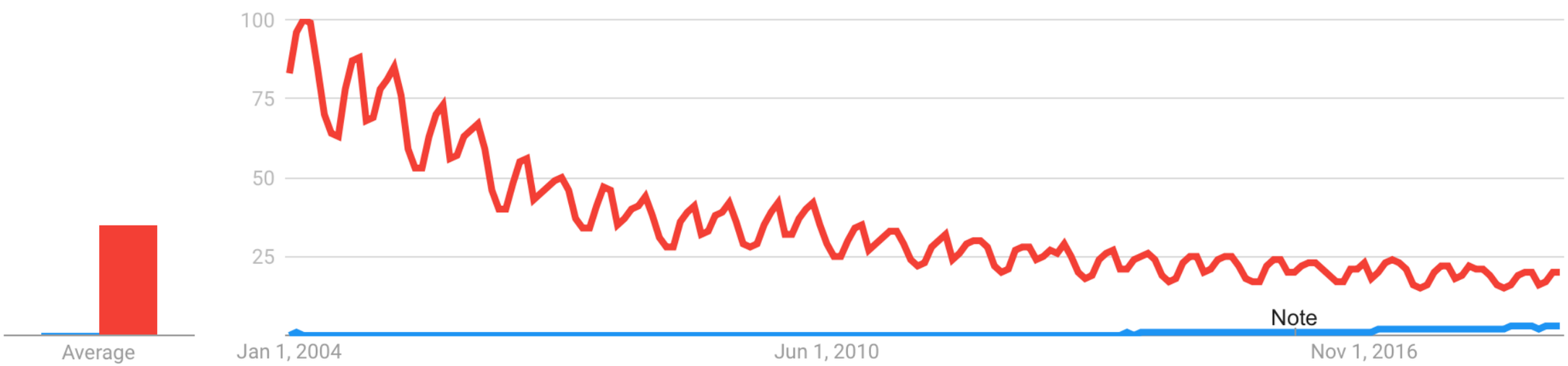
data science
Search term

statistics
Search term

+ Add comparison

Worldwide | 2004 - present | All categories | Web Search

Interest over time



1.

Tukey's new field (1962)

"For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt... All in all I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data"

1.

Tukey's new field (1962)

Four major influences act on data analysis today:

1. The formal theories of statistics

2. Accelerating developments in computers and display devices

3. The challenge, in many fields, of more and ever larger bodies of data

4. The emphasis on quantification in an ever wider variety of disciplines

2.

Donoho's science of data (2015)

"The imminent revolution in Data Science is not about mere 'scaling up', but instead the emergence of scientific studies of data analysis science-wide. In the future, we will be able to predict how a proposal to change data analysis workflows would impact the validity of data analysis across all of science, even predicting the impacts field-by-field."

3.

A huge tent

"We believe that Data Science should not be confined to the obvious disciplines of computer engineering or mathematics, but should be broadly distributed across campus and in the Faculty of Arts and Science, in order to bring in a diversity of ideas and people into the problems it faces."

CHARTER OF THE AMERICAN STATISTICAL ASSOCIATION
SECTION ON STATISTICS AND DATA SCIENCE EDUCATION~~STATISTICAL EDUCATION~~

Charter revision draft of October 15, 2018
Charter last revised 2001
Section chartered in 1948

ARTICLE I. NAME

The name of this organization is the Section on Statistics and Data Science Education~~Statistical Education~~ (hereafter called the "Section") of the American Statistical Association (hereafter called the "Association").

ARTICLE II. MISSION STATEMENT AND OBJECTIVES

A. Mission Statement:

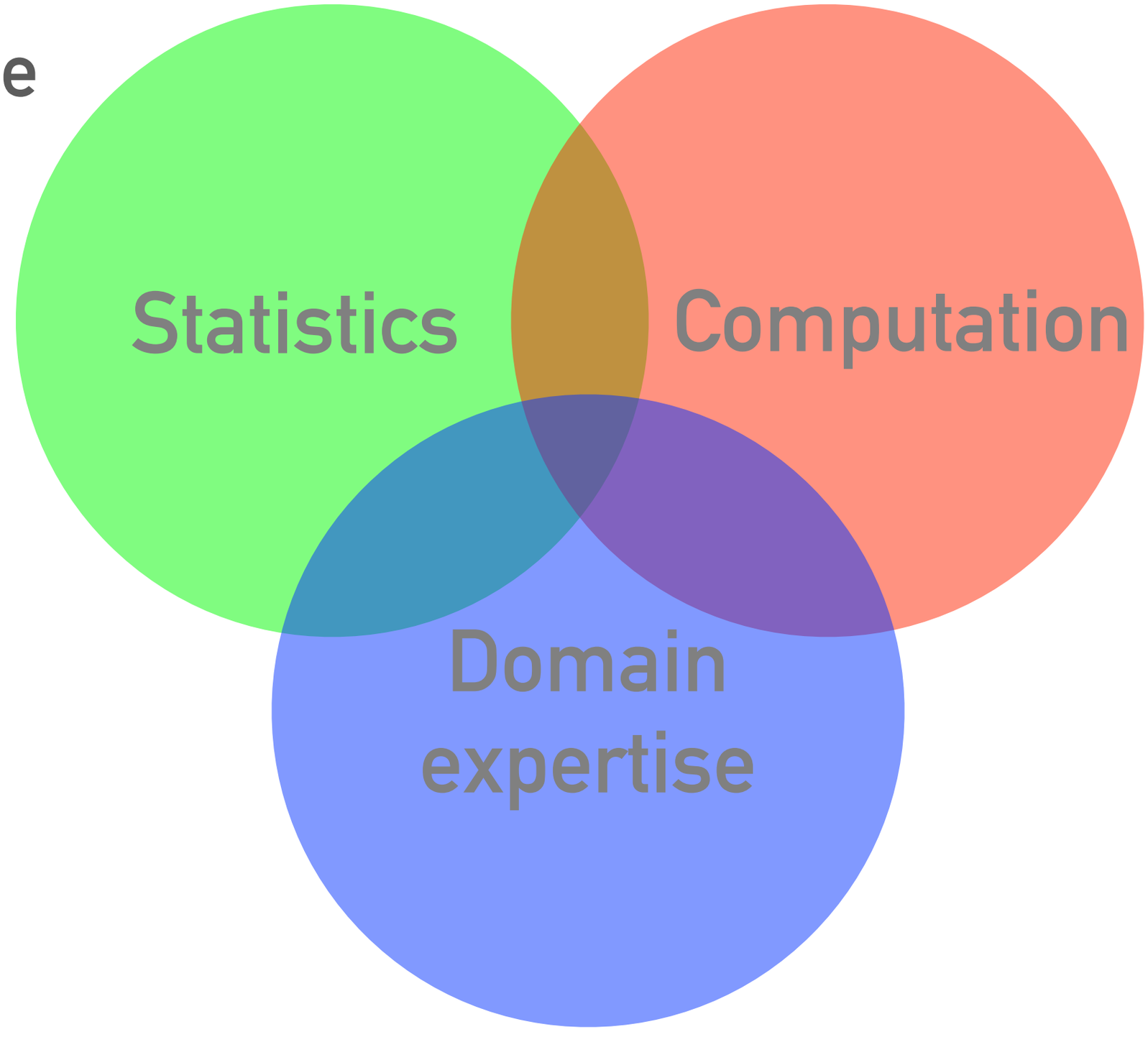
Statistics ~~and Data Science~~ is the science of learning from data, and we consider data science a partner of statistics. Our mission is to promote the teaching and learning of data-driven decision making ~~statistics and data science~~ at any level for any audience. Specifically we will:

1. Advise the Association on the educational elements of the Association strategic plan, including the need for education on the role and value of statisticians and data scientists to the following audiences:

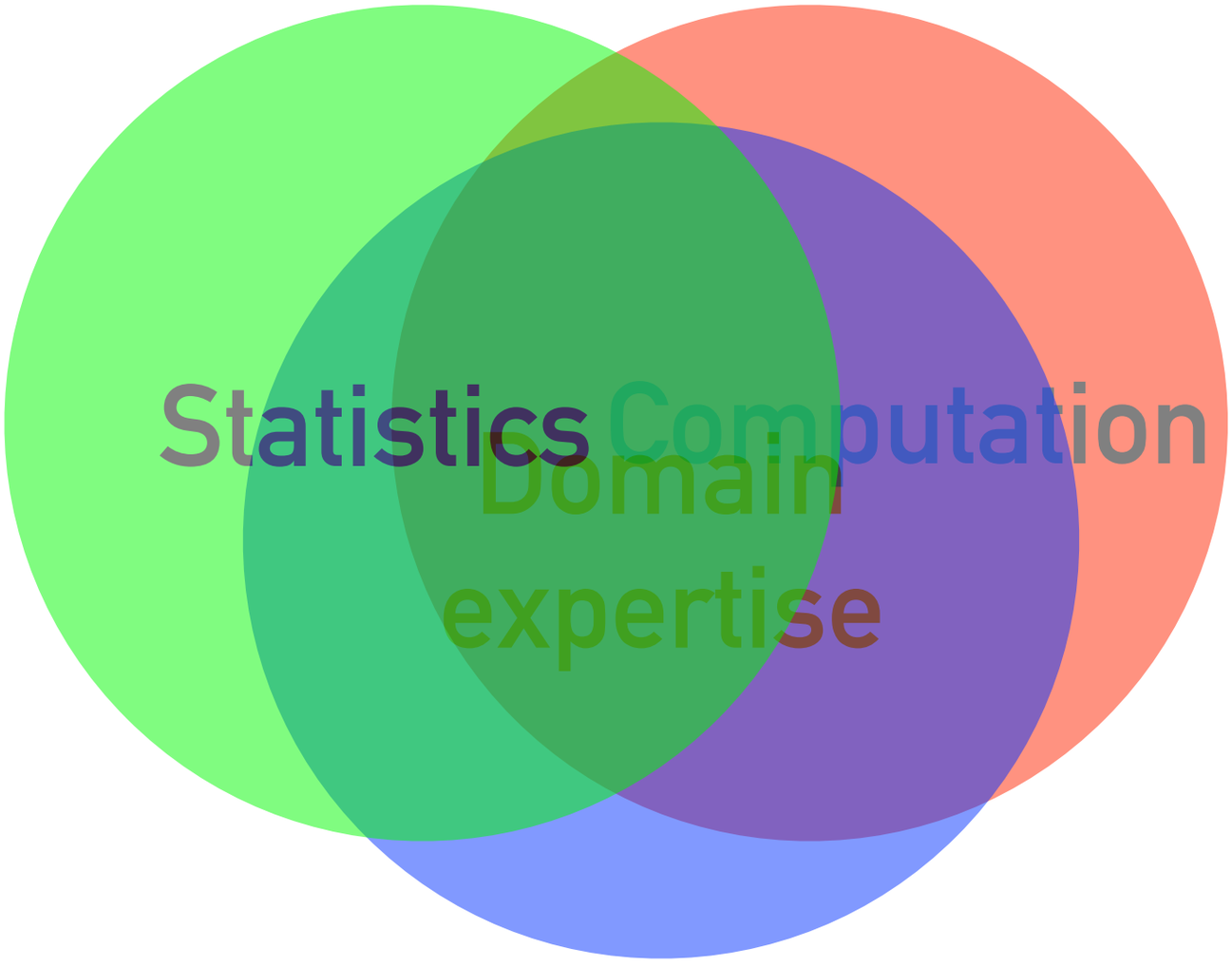
a. Management in business, technology, and government
b. Professionals such as journalists, lawyers, computer scientists, and medical personnel

INFLUENCE ON OUR STATISTICS PROGRAMS OF STUDY

The Data Science Venn Diagram



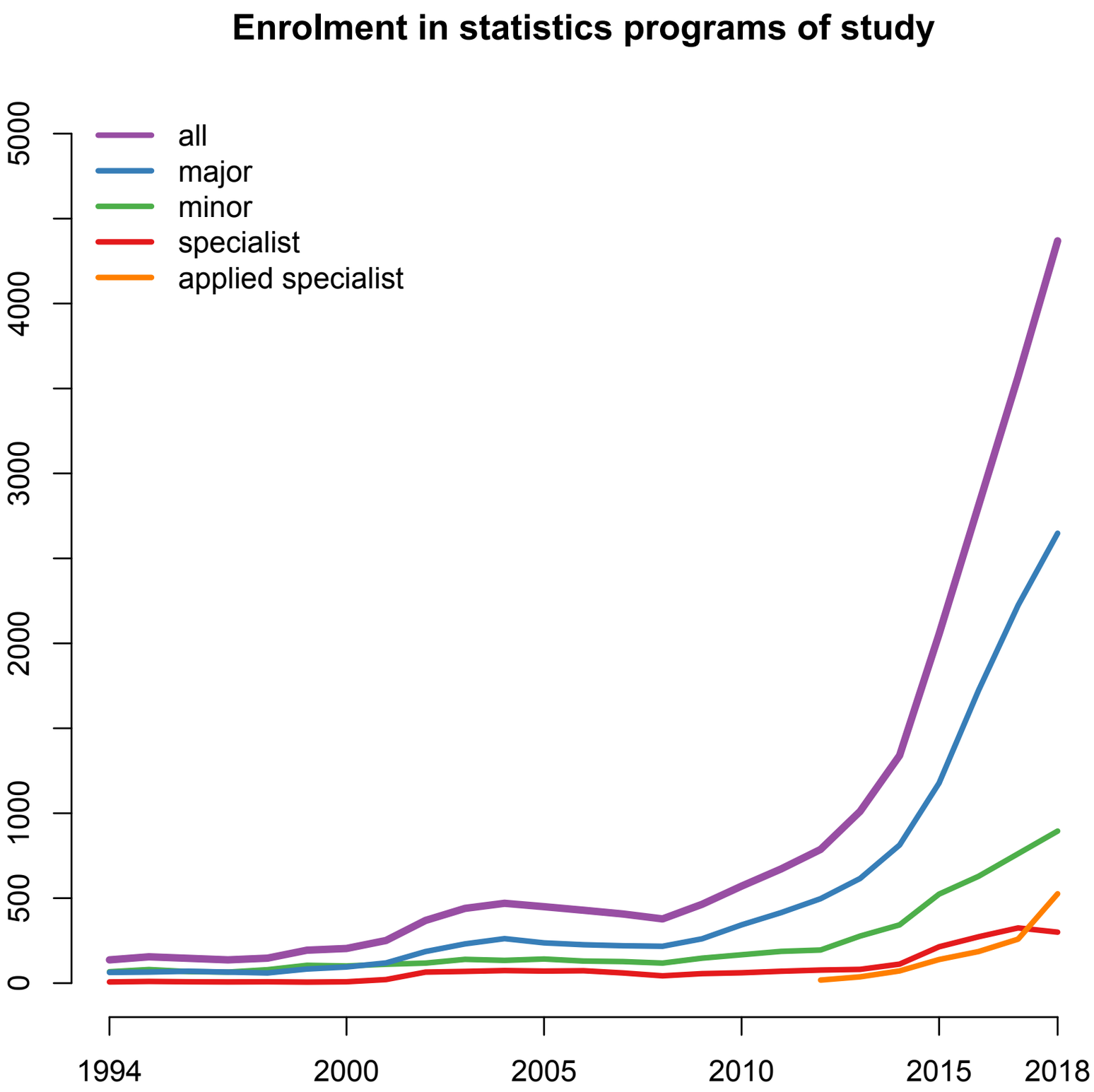
A NEW VIEW OF STATISTICS PROGRAMS OF STUDY



A rigorous mindset plus…

- More computation and algorithmic thinking
- More real problems
- Able to talk and write about it

UNDERGRADUATES IN STATISTICS PROGRAMS OF STUDY AT U OF T



~60% international

New in 2018:

Data Science specialist

(joint with CS)

25 students

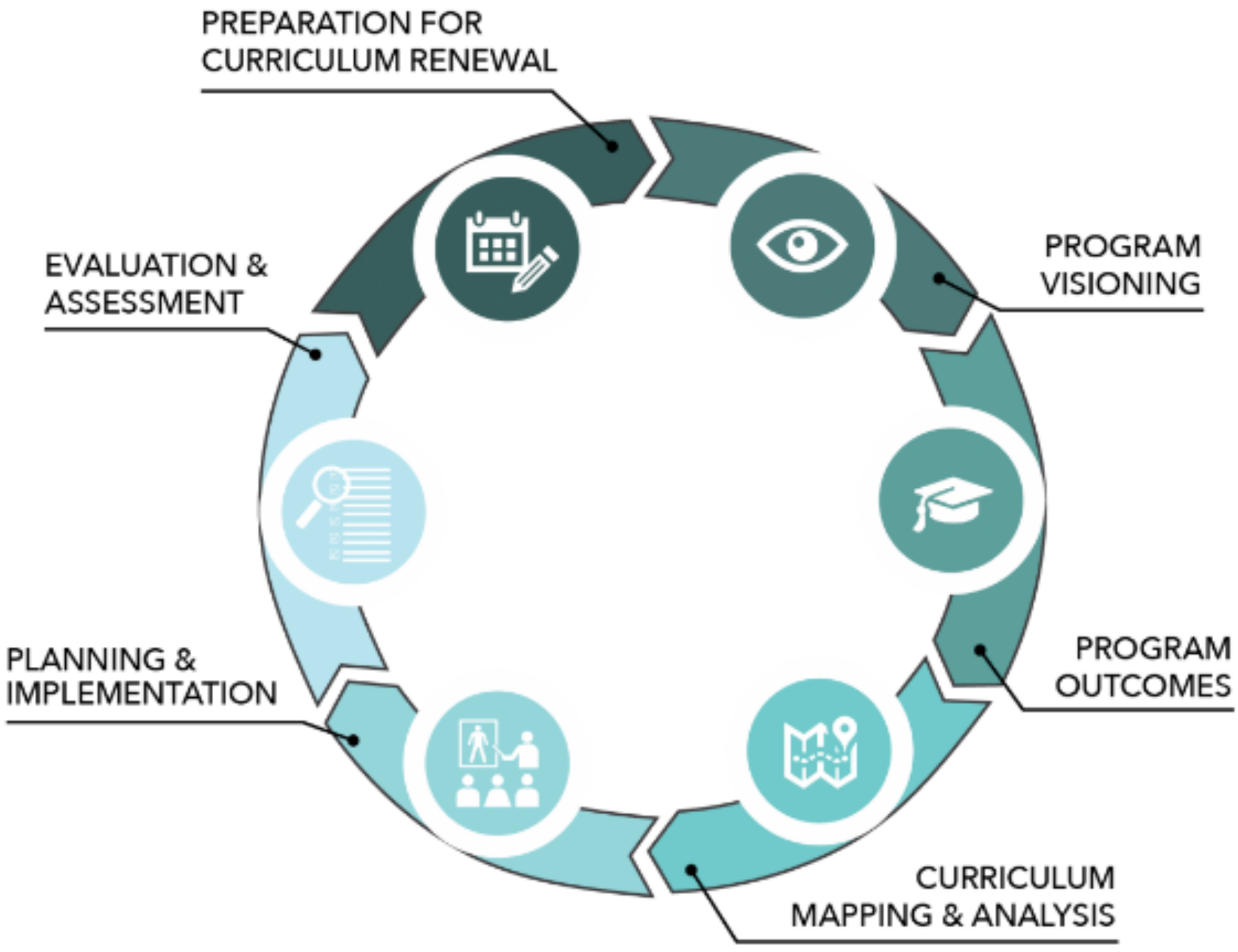
“

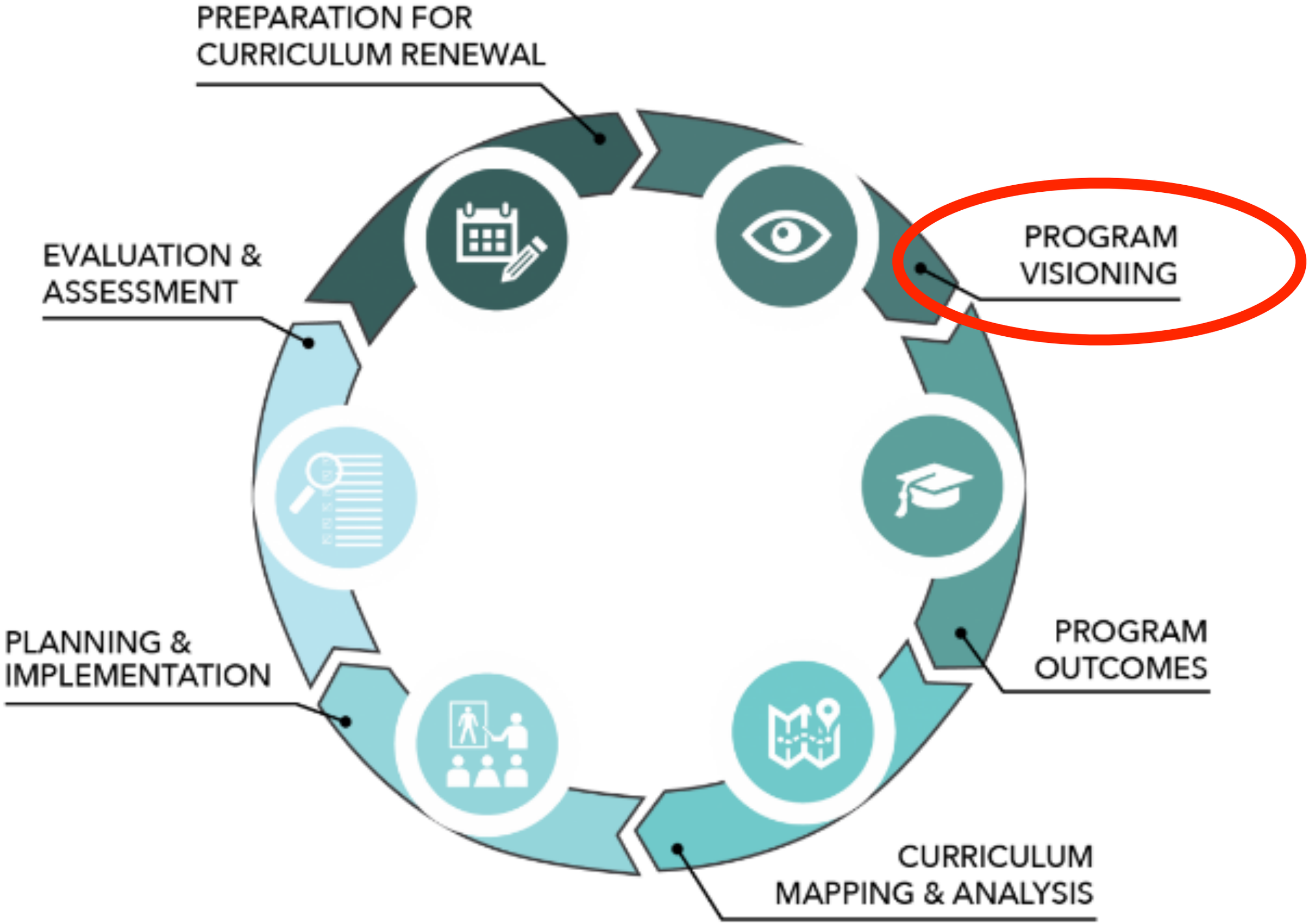
You never let a serious crisis go to waste.

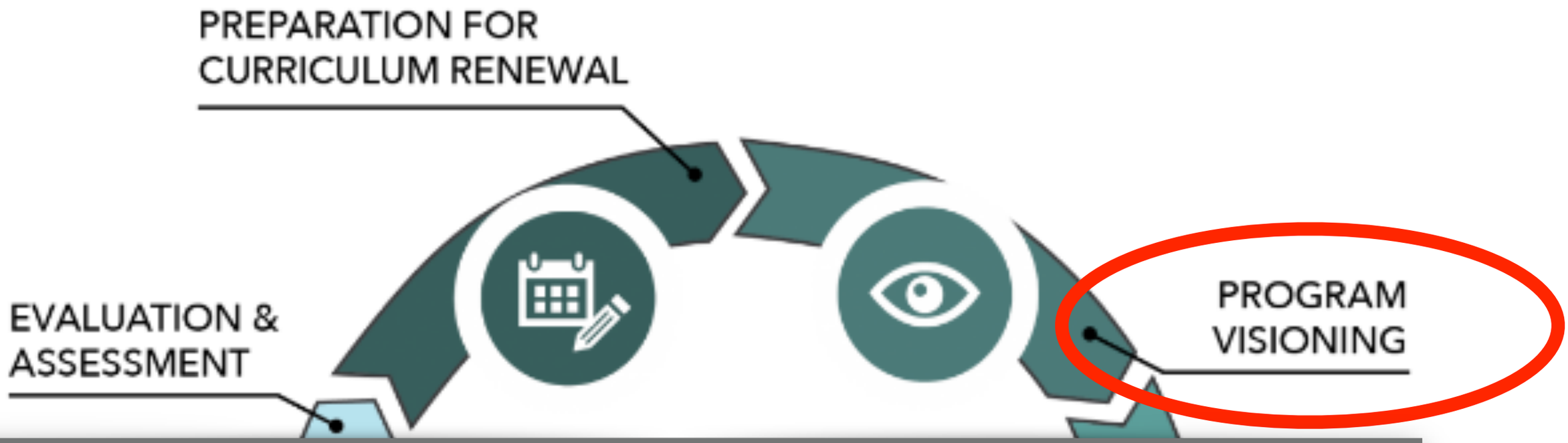
-Rahm Emanuel (2008)

STATISTICS CURRICULUM RENEWAL

CURRICULUM RENEWAL PROCESS




PREPARATION FOR CURRICULUM RENEWAL
PROGRAM VISIONING
PROGRAM OUTCOMES
CURRICULUM MAPPING & ANALYSIS
PLANNING & IMPLEMENTATION
EVALUATION & ASSESSMENT



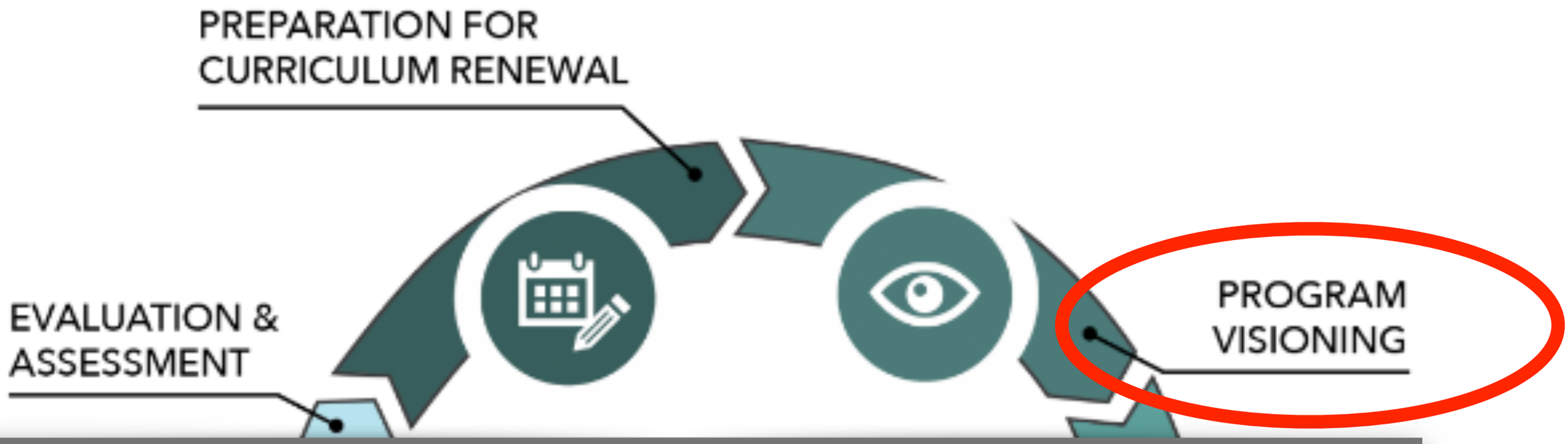
Statistical Science encompasses methods and tools for obtaining knowledge from data and for understanding the uncertainty associated with this knowledge.

The purpose of our undergraduate programs are to:

(1) equip students with a general framework for obtaining knowledge from data;

(2) give students skills that they are able to flexibly apply to a variety of problems; and

(3) provide students with the ability to learn new methods as needs, data sources, and technology change

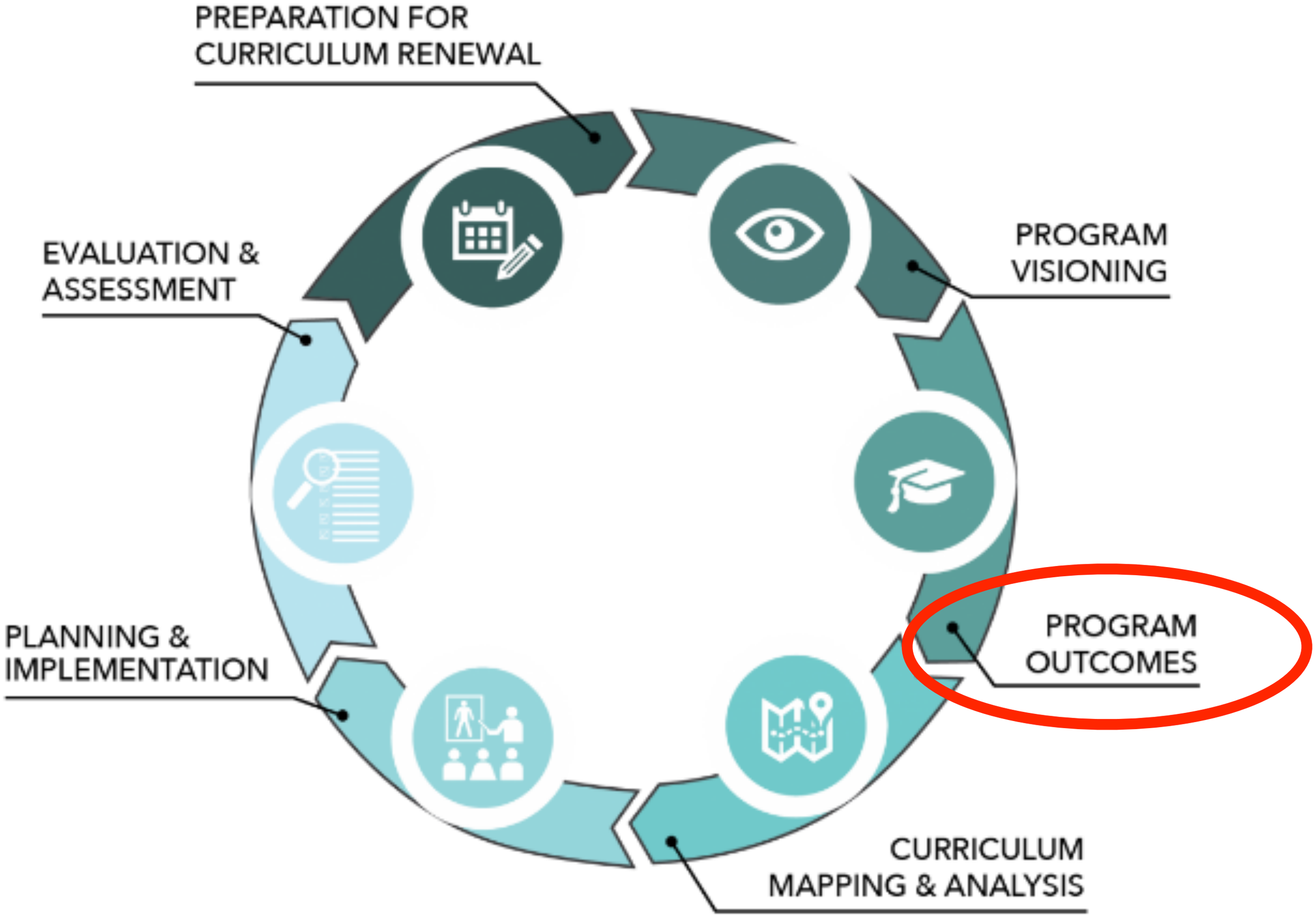


Key words:

- obtain knowledge from data
- understand associated uncertainty
- flexible skills
- ability to adapt to change

More of…	Introduce…
Computation	New forms of data
Communication	Program requirements that send students down different paths
Experiential learning	
Predictive modelling	

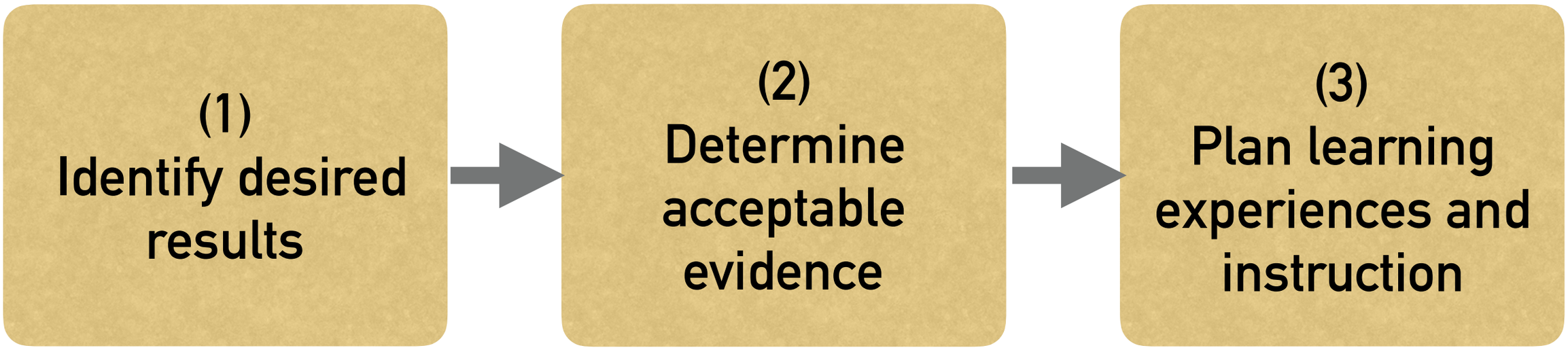
Keep…		
Design in data collection	Uncertainty	Inference, when appropriate
Overfitting	Multiple testing	Confounding
Bias	Causation	etc…


PREPARATION FOR CURRICULUM RENEWAL
PROGRAM VISIONING
PROGRAM OUTCOMES
CURRICULUM MAPPING & ANALYSIS
PLANNING & IMPLEMENTATION
EVALUATION & ASSESSMENT

CURRICULUM DEVELOPMENT GUIDING PRINCIPLE #1

Backwards Design

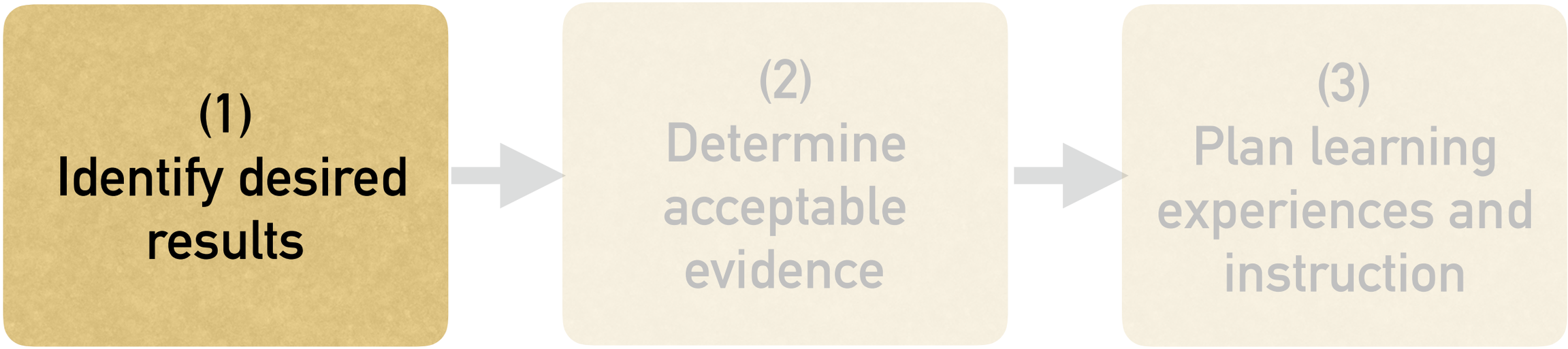
Wiggins & McTighe (2005)



CURRICULUM DEVELOPMENT GUIDING PRINCIPLE #1

Backwards Design

Wiggins & McTighe (2005)



“

A subject area is ultimately about the *doing* of a subject — using the content in a disciplined way … [not] a march through content.

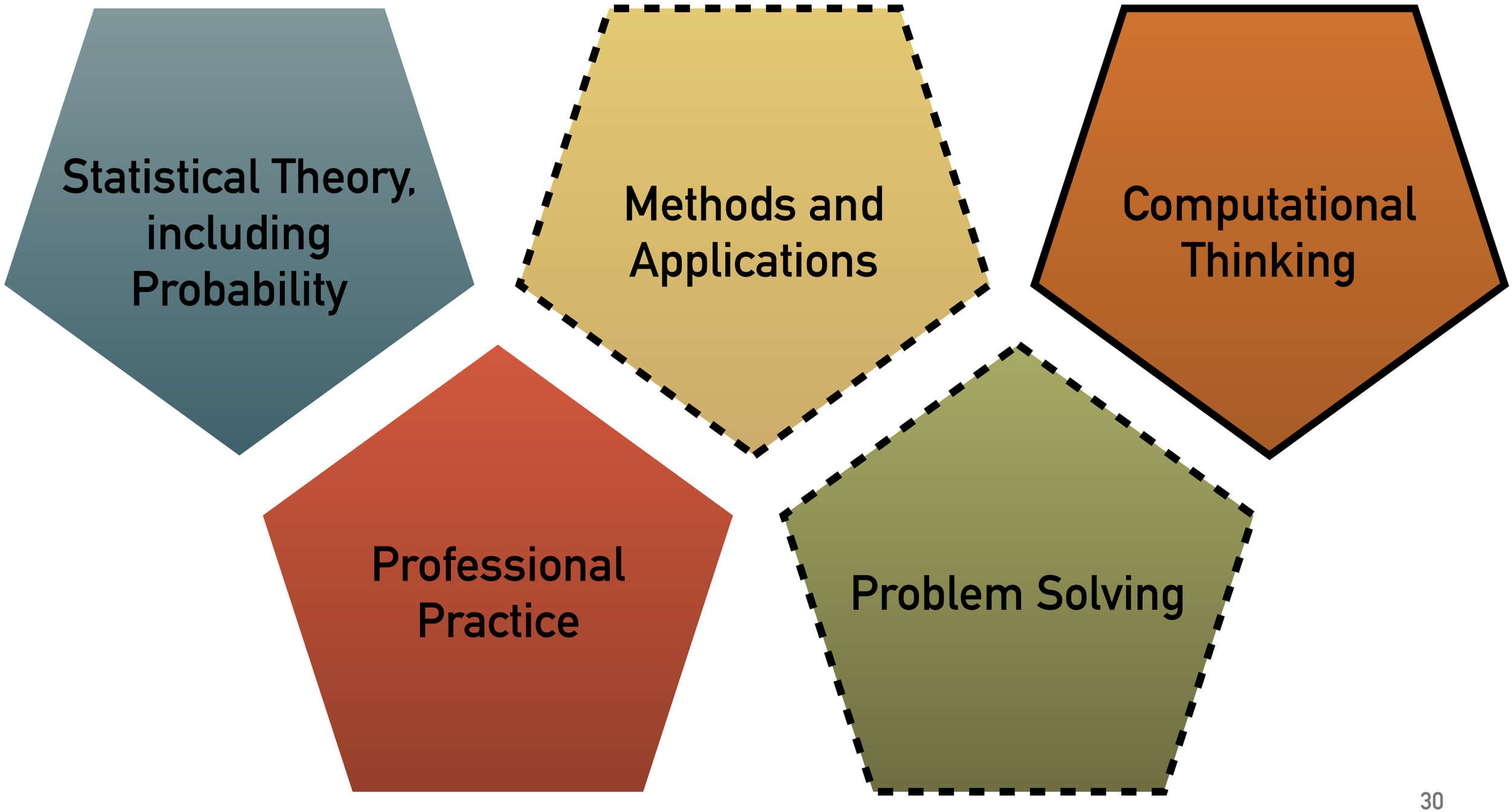
-Wiggins & McTighe (2005)

What is important for our graduating students

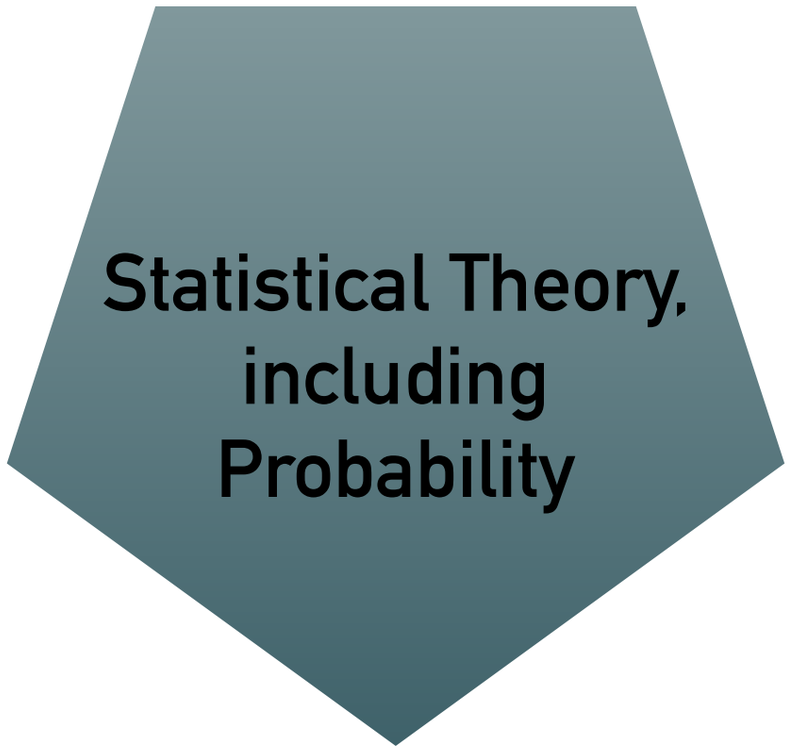
- to know
- to know how to do
- to value
- to have as habits of mind?

END GOALS FOR STUDENTS IN OUR STATISTICS PROGRAMS OF STUDY

20 Program Learning Outcomes organized into 5 themes:



STATISTICS PROGRAM LEARNING OUTCOMES

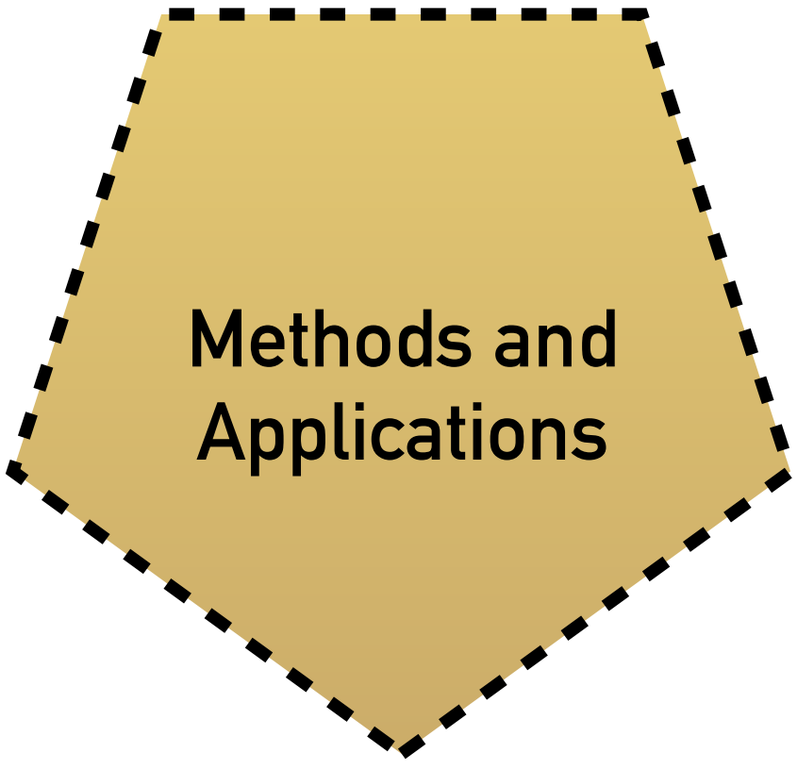


Includes

- understanding and application of the role of probability as a mathematical framework to represent uncertainty
- paradigms for statistical inference
- theoretical results that support the rationale of statistical methods

More connection to computing

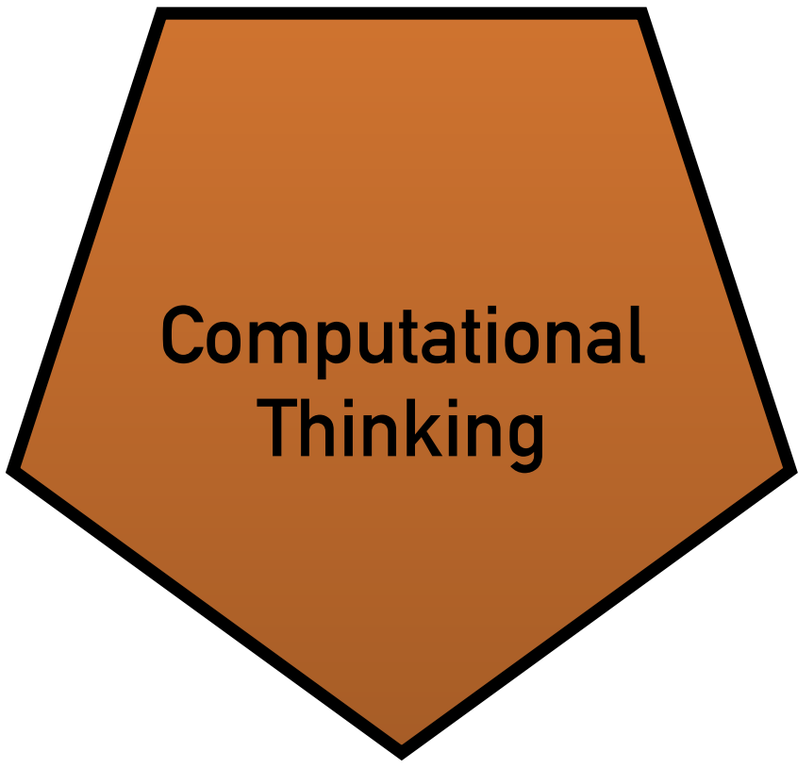
STATISTICS PROGRAM LEARNING OUTCOMES



Includes

- appropriate application of a variety of statistical methods for description, prediction, explanation
- study design and data collection strategies
- data visualization

STATISTICS PROGRAM LEARNING OUTCOMES



Includes

- using simulation for a variety of purposes
- data manipulation
- computing for data analysis in a literate format that supports reproducibility

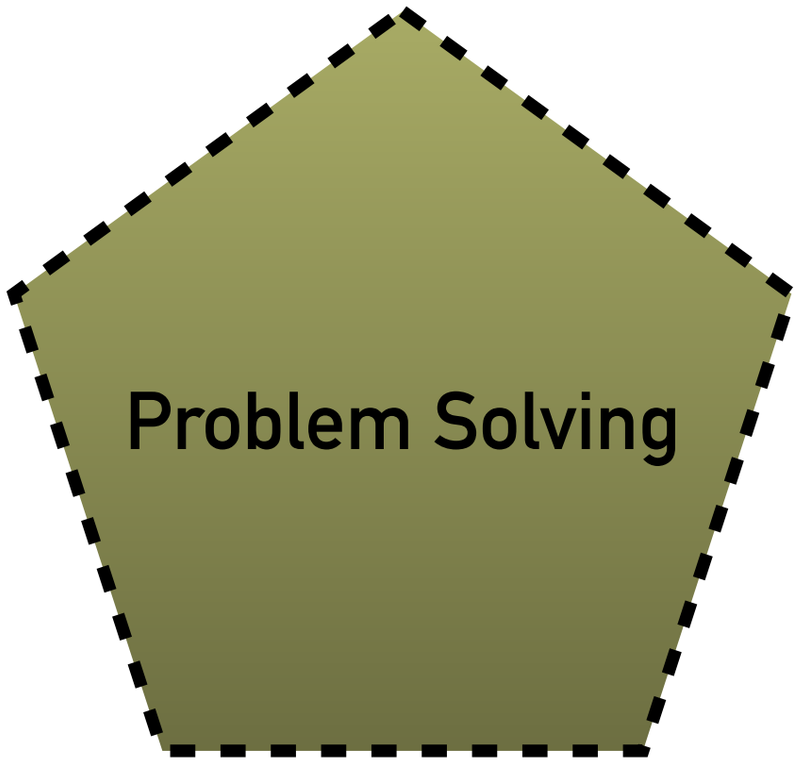
STATISTICS PROGRAM LEARNING OUTCOMES

Includes ability to

- speak and write clearly for both technical and non-technical audiences
- apply ethical principles
- collaborate with both statisticians and non-statisticians
- recognize the role of context in using statistics to devise solutions

More….

STATISTICS PROGRAM LEARNING OUTCOMES



Includes broad, higher order learning outcomes that are applicable to both theory and application

- apply and adapt existing knowledge to learn new methodologies
- apply knowledge in new areas of application
- evaluate relative merits of competing approaches
- recognition of strengths and limitations of methods

Program: **Statistics**
Report: Proposed program map
Standard: B.Sc. (Major, Specialist, Applied Statistics Specialist)
Date/Time: 2017-09-25 (updated with new course titles 2018-03-20)

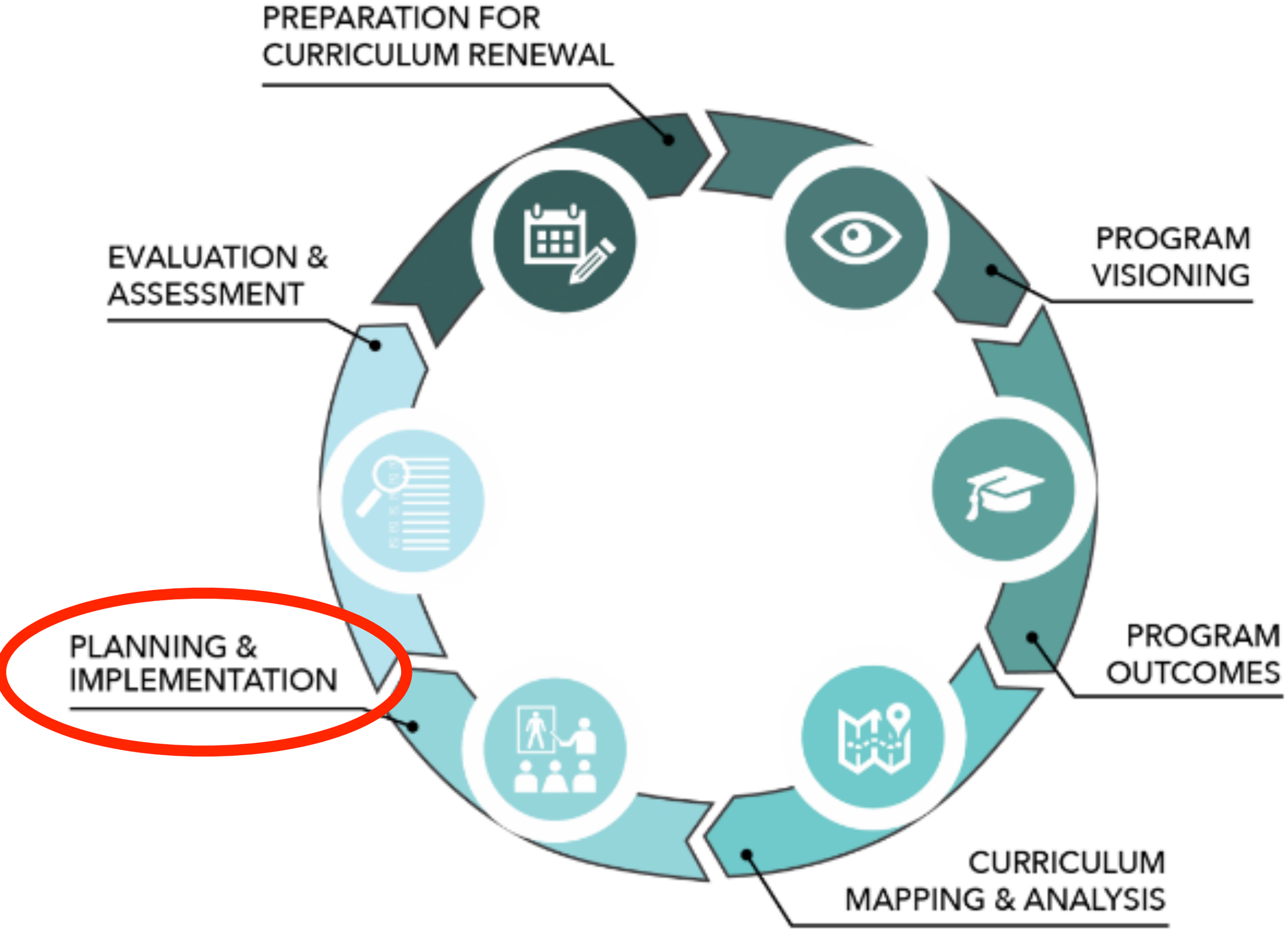
Required?	Year	Course Code	Course Name	THEORY:									METHODS:									COMPUTATIONAL THINKING:												PROFESSIONAL PRACTICE:										
				PLO 1			PLO 2			PLO 3			PLO 4			PLO 5			PLO 6			PLO 7			PLO 8			PLO 9			PLO 10			PLO 12			PLO 13			PLO 14			PLO	
				I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	D	P	I	
Core	1	STA130	Introduction to Statistical Reasoning and Data Science	X			X									X			X			X			X						X			X						X			X	
Core	1	CSC108/120/148	Introduction to Computer Programming																										X															
Core	1	MAT135+135/137/157	Univariate Calculus																																									
Core	1	MAT223	Linear Algebra																																									
Core option 1	2	STA237	Probability, Statistics and Data Analysis I		X																		X																					
Core option 1	2	STA238	Probability, Statistics and Data Analysis II					X			X		X										X									X												
Core option 2	2	STA257	Probability and Statistics I		X																		X																					
Core option 2	2	STA261	Probability and Statistics II					X			X		X										X									X												
Core	2	MAT235/237/257	Multivariate Calculus																																									
Both specialists	2	STA210	Statistical Computation																								X			X		X												
Core	3	STA302/3	Methods of Data Analysis (Linear models and GLM)					X			X			X																		X			X			X			X			X
Both specialists	3	STA304/5	Data Collection					X				X			X			X+						X										X		X				X			X		X
One of for specialists	3	STA314	Statistical Methods for Machine Learning I		X				X						X									X			X			X														
One of for specialists	3	STA310	Topics in Statistical Computation											X										X						X														
Statistics specialist	3	STA347	Probability			X					X			X																					X									
Both specialists	3	STA355	Theory of Statistical Practice			X			X+			X+								X				X											X									
	4	STA410	Statistical Computation						X+						X									X+						X														
One of for core	4	STA414	Statistical Methods for Machine Learning II			X			X						X+									X+			X+			X+						X								
One of for core	4	STA437	Methods for Multivariate Data						X						X																		X			X								
One of for core	4	STA442	Methods of Applied Statistics						X+						X+			X+						X+									X+			X+			X+					
One of for stats spec	4	STA447	Stochastic Processes			X+																																	X					
One of for stats spec	4	STA452,453	Mathematical Statistics			X+			X+			X+																											X					
One of for core	4	STA457	Time Series Analysis			X		X				X			X														X			X			X									X
One of for core	4	STA465	Theory and Methods for Complex Spatial Data												X																					X								
One of for core	4	STA480	Fundamentals of Statistical Genetics												X																					X								
Applied specialist	4	STA490	Statistical Consultation, Communication, and Collaboration																														X			X+			X+			X+		
Statistics specialist	4	STA495	Research Seminar in Statistics			X+			X+			X+			X+			X+																		X+								
			Core Total	1	1	1	1	2	1		2		1	1	1	1			1			1	2		1				1		1	2		1	1	1		1		1	1		1	
			Statistics Specialist Total			4			3		1	3		1	1			2		1				1			1			1		1			2	1			2					
			Applied Statistics Specialist Total			1			1			1						1		1				1			1			1		1	1		1	1			1			1		

I - Introduce: This course explicitly introduces this learning outcome
D - Develop: This course explicitly develops this learning outcome
P - Proficient: This course explicitly reinforces this learning outcome to the level of proficiency expected of a graduating stude

Notes:

PLO11 (more advanced computational skills) was written only for the Data Science program

PLO16 (interdisciplinary teamwork) was written only for the Applied Statistics and Data Science programs


PREPARATION FOR CURRICULUM RENEWAL
PROGRAM VISIONING
PROGRAM OUTCOMES
CURRICULUM MAPPING & ANALYSIS
PLANNING & IMPLEMENTATION
EVALUATION & ASSESSMENT

CURRICULUM GUIDING PRINCIPLE #2

A curriculum gives opportunities to be introduced to / reinforce / **master** concepts

How should our students end their program?

“

The best thing about being a statistician is that you get to play in everyone’s backyard.

-John Tukey

“

The

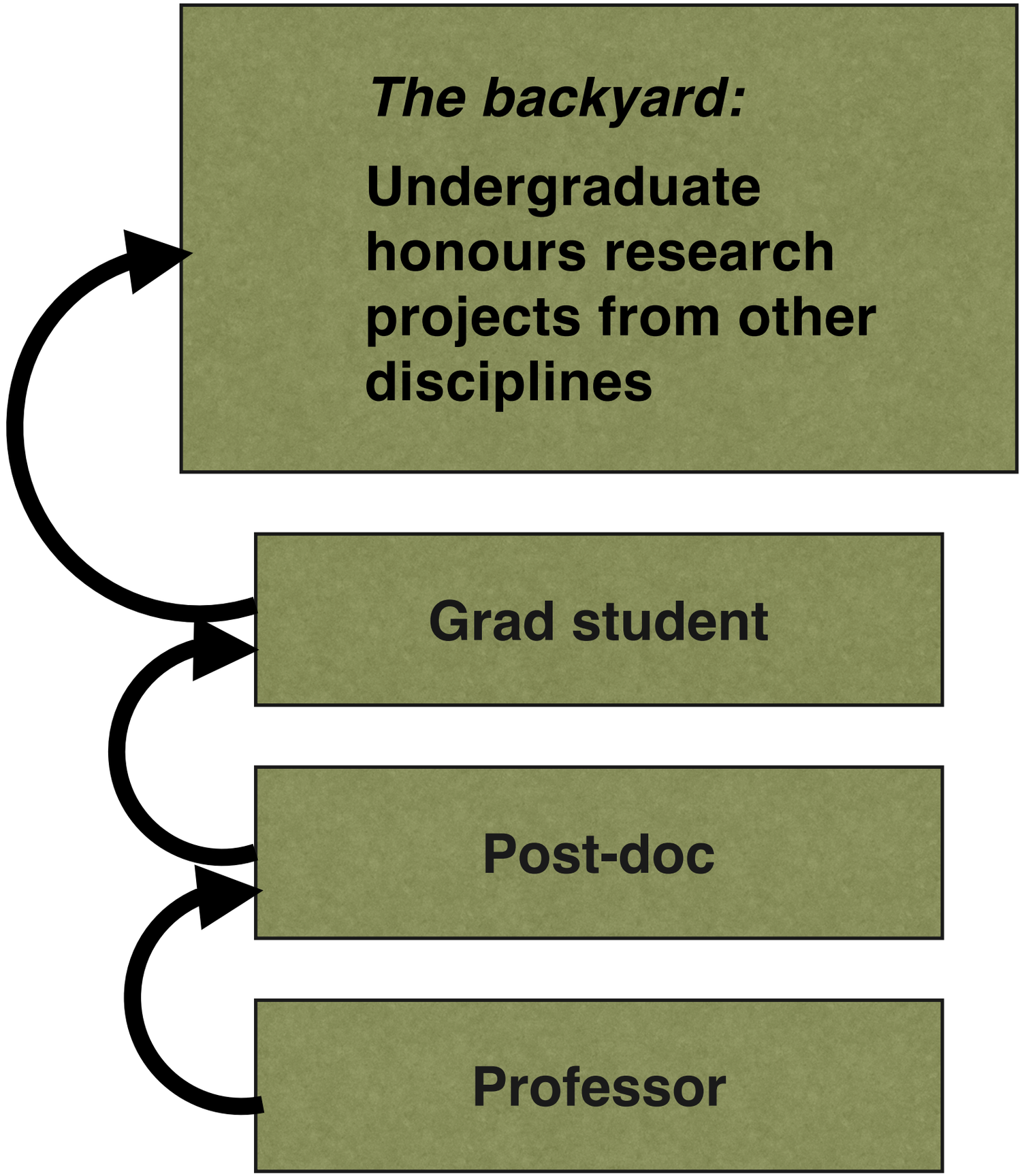
stat

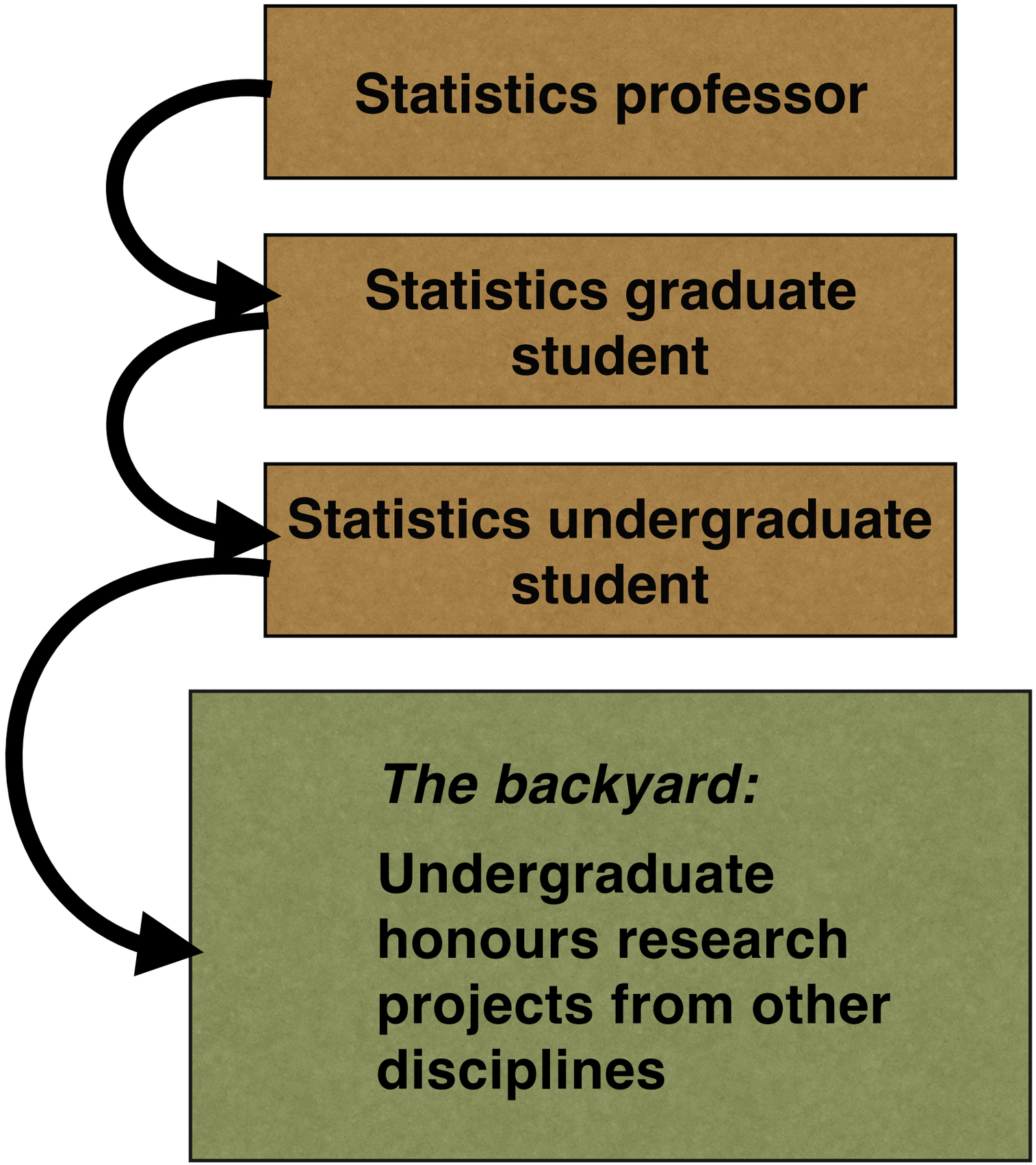
ever

play in

ohn Tukey

PL013:
Interpret a **non-statistician's** description of a **problem** and formulate a question, study design, and analysis in statistical terms


The backyard:
Undergraduate honours research projects from other disciplines
Grad student
Post-doc
Professor


Statistics professor
Statistics graduate student
Statistics undergraduate student
The backyard:
Undergraduate honours research projects from other disciplines

STA490: STATISTICAL CONSULTATION, COMMUNICATION, & COLLABORATION

- Capstone course for undergraduate statistics students
- Centred around a collaborative project with a research student from another discipline
- Stakeholders:
  - graduating students in statistics programs of study
  - graduate students in statistics
  - research students from other disciplines
  - research labs

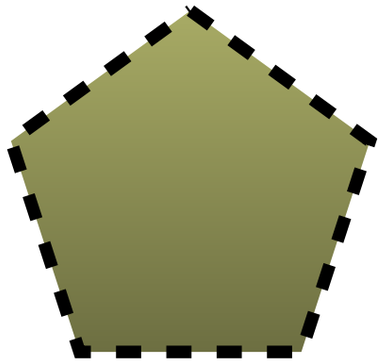
STA490: CONTENT

- I don't teach any methodology
- But the students learn methodology

> A good waiter in a good restaurant is around enough to help you when you need things but leaves you alone enough to enjoy yourself.
>
> *-Murray (1998)*

STA490: CONTENT

- I don't tea
- But the st



"

A g rant is
arc en you
ne ne
en

-Murray (1998)

PL018:
Apply and adapt prior knowledge and recent scholarly work to learning and evaluating new methods and solving problems in new areas of application

STA490: STRUCTURE

Seminars

- Presentations
- Data presentation principles
- Principles of communication
- Replication vs pseudo-replication
- Researcher degrees of freedom
- Reproducibility, replicability, and false discovery
- Ethics

Project group meetings

- Run by statistics graduate student advisor
- Collaborator comes sometimes

STA490 SEMINAR: THE FIRST SEMINAR

- Process produces a product that must be between 0.09 and 0.11 mm in diameter to be useable
- Sample of 40 products has been taken
- 99% confidence interval for the mean is $0.102 \pm 2.71 * 0.01 / \sqrt{40} = (0.0977, 0.1063)$

1. Is the analysis correct?
2. Is the analysis useful?
3. What questions would you ask?

[adapted from a Sanford Weisberg exercise]

STA490 SEMINAR, THE FIRST SEMINAR

- Process
  0.09 and
- Sample
- 99% confi
  0.102 ± 2

1. Is the an
2. Is the an
3. What qu

PLO13:
Interpret a non-statistician's description of a problem and formulate a question, study design, and analysis in statistical terms

[adapted from a Sanford Weisberg exercise]

STA490: SEMINAR READINGS

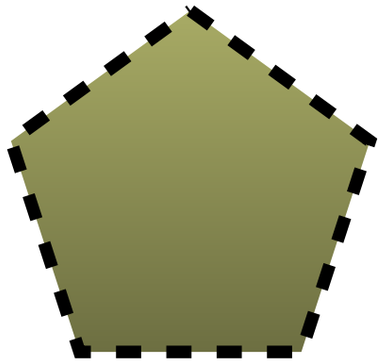
- Kass, Caffo, Davidian, Meng, Yu, and Reid (2016). Ten Simple Rules for Effective Statistical Practice. PLOS Computational Biology 12(6),
- S.E. Lazic (2010). The problem of pseudo replication in neuroscientific studies: is it affecting your analysis? BMC Neuroscience 2010 11:5
- Lefevre et al (2010). Beer consumption increases human attractiveness to malaria mosquitoes. PLoS One 5:3
- J. Simmons, L. Nelson, and U. Simonsohn (2011) False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science 22(11), 1359-66.
- Statistical Society of Canada Code of Ethical Statistical Practice and American Statistical Association Ethical Guidelines for Statistical Practice
- Upchurch, Tom (2018). To work for society, data scientists need a hippocratic oath with teeth. Wired, 8 April 2018.
- Silberzahn et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science 1(3), 337-356
- BMJ Statistics Notes, Endgames, and Nature Methods Points of Significance
- JT Leek and LR Jager (2017). Is Most Published Research Really False? Annual Review of Statistics and Its Application, Vol. 4:109-122

STA490: SEMINAR READINGS

- Kass, Caffo, Davidian, Meng, Yu, and Reid (2016). Ten Simple Rules for Effective Statistical Practice. PLOS Computational Biology 12(6),

STA490: SEMINAR READINGS

- Kass, Ca… …id (2016). Ten Simp… …l Practice. PLOS Co…



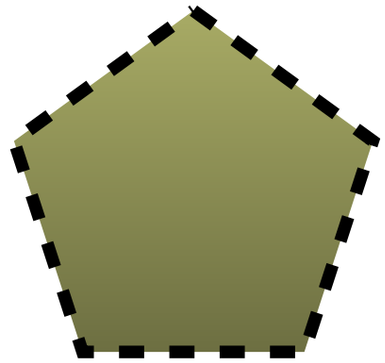
PL017:
Understand statistical analysis as a unified framework involving an iterative process of question formulation, data collection and / or evaluation of the suitability of available data, data cleaning and preparation, exploratory data analysis, modelling, interpretation and communication of results.

STA490: SEMINAR READINGS

- JT Leek and LR Jager (2017). Is Most Published Research Really False? Annual Review of Statistics and Its Application, Vol. 4:109-122

STA490: READING

- JT Leek ublished
  Researcl of Statistics
  and Its A



PLO20:
Assess results of statistical analysis for inconsistencies and sources of error, recognize the limitations of the analysis and when more complex analysis is warranted, recognize the implications of issues such as bias, causality, model assumptions, measurement error, confounding, multiple comparisons

STA490 IN A DATA SCIENCE WORLD

What is the same?	What has changed?
Asking questions	Emphasis on reproducibility
Communication	Ethics
Collaboration	More projects that don't require formal inference

CURRICULUM GUIDING PRINCIPLE #2

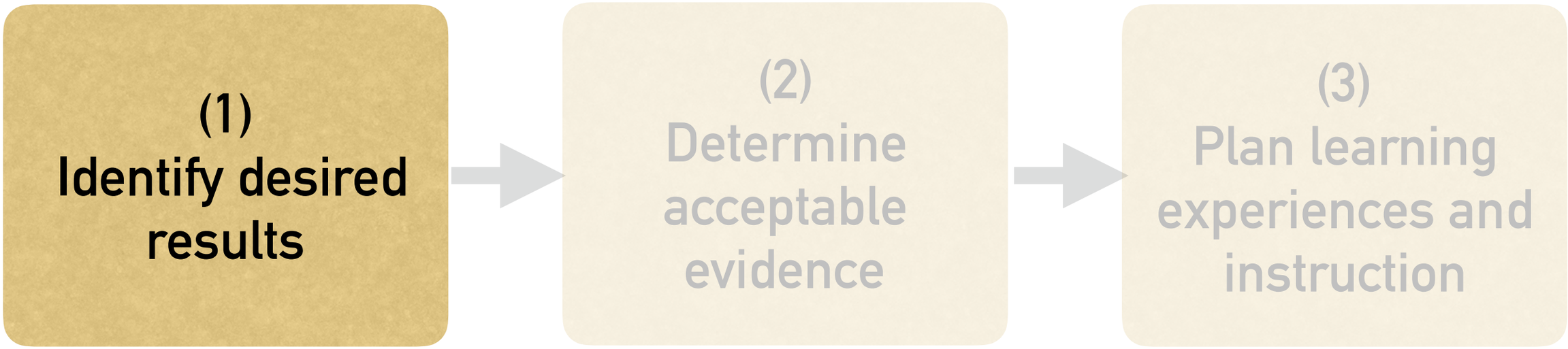
A curriculum gives opportunities to be **introduced to** / reinforce / master concepts

How should our students start?

CURRICULUM DEVELOPMENT GUIDING PRINCIPLE #1

Backwards Design

Wiggins & McTighe (2005)



A goal:

We want our students to be able to do something creative with data, while demonstrating appropriate statistical thinking **and** be able to talk about it

How do we get here?

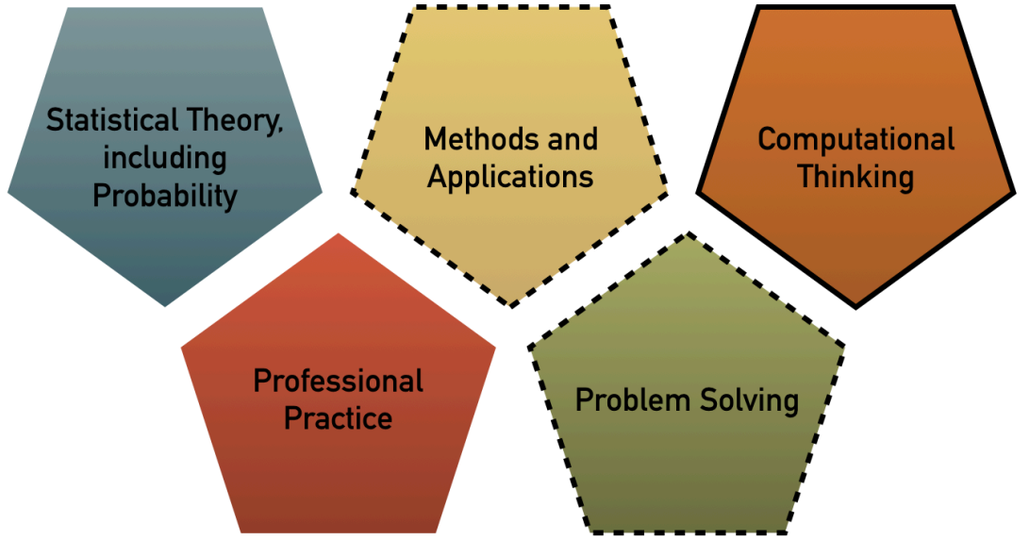
A FIRST COURSE?

- Traditional starting point:

  Mathematical foundations

- New starting point:

  Survey course



- Immersion in real interesting problems to motivate

- Particular emphasis on communication

A FIRST COURSE: CONTENT

- Computing basics
- Visualization (EDA)
- Data wrangling
- Comparing two groups (randomization test)
- Modelling for prediction, description, and inference (regression)
- Statistical issues: multiple testing, confounding
- Supervised learning for prediction (classification trees)
- Ethical practice

A FIRST COURSE: 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

A FIRST COURSE

Mondays

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

- Introduction to statistical methods for
  - summarizing
  - testing
  - estimation
  - prediction

Example week:

Are couples more likely to lean left or right when kissing?

Introduction to testing using model of coin flips (then simulations in R) for inference

A FIRST COURSE

M

- Larg
  secti
- Data
- A ne
  meth
- R by
  answ
  ques

roduction

summari

testing

estimatio

predictio

ple week:

ouples mo

kissing?

duction to

simulatio

for

ght

oin flips

PLO2:
Demonstrate an understanding of the philosophy and purpose of statistical inferential reasoning

PLO7:
Use simulation to evaluate statistical methods, support theoretical solutions, and as an approach to inference

A FIRST COURSE

<u>Example week:</u>

Detecting Parkinson's disease by smell

- *Vary the number of simulations*
- *Calculating the P-value when observed proportion is 1*

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

A FIRST COURSE

Example

Detectin
Parkinso
by smell

- *Vary ti*
  of simu
- *Calcul*
  value v
  observ
  propor

ctice prob

larkdown
plate to s
utions

e solution
ught to tut
grading a
cussion

PLO10:
Carry out data analysis for common methods using statistical software in a reproducible way

PLO18:
Apply prior knowledge to learning and evaluating the appropriate application of new methods and to solving problems in new areas of application

A FIRST COURSE

- In small groups, students reflect and articulate on concepts (for example, comparing, combining results from simulations)

Example week:

Given vocabulary, discussion of

- *Dealing with range of possible answers for different simulations*
- *What happens when we simulate more?*
- *Interpreting P-value when there is nothing "more extreme"*

Fridays

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

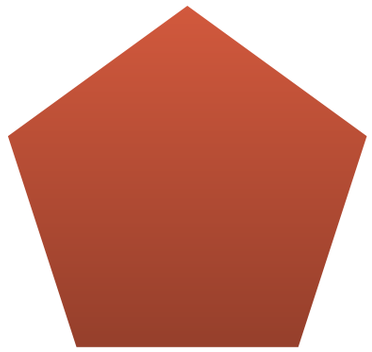
A FIRST COURSE

- In small groups, articulate on con comparing, com simulations)

Example week:

Given vocabulary, di

- *Dealing with range different simulatio*
- *What happens whe*
- *Interpreting P-valu "more extreme"*



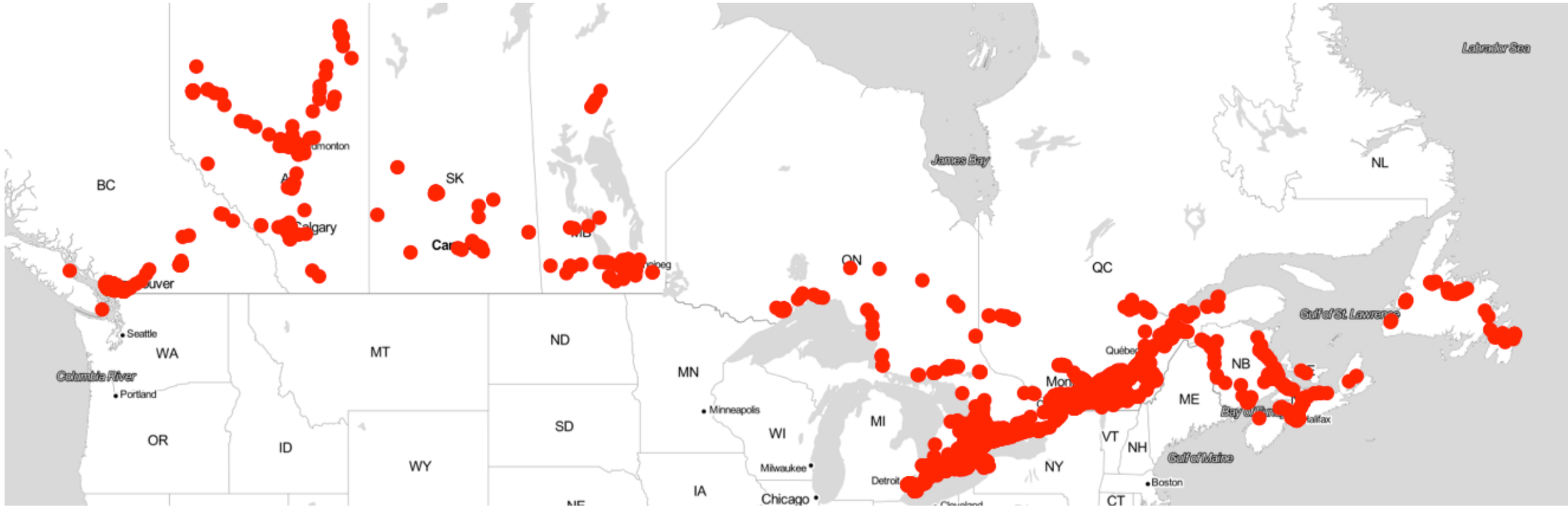
PLO12:
Present accurate, clear, concise descriptions of statistical methods and the results of analysis to statisticians and non-statisticians orally, in writing, and through appropriate data visualizations

Fridays

Small group tutorials

Random group assignments each week

Oral or written communication exercise each week

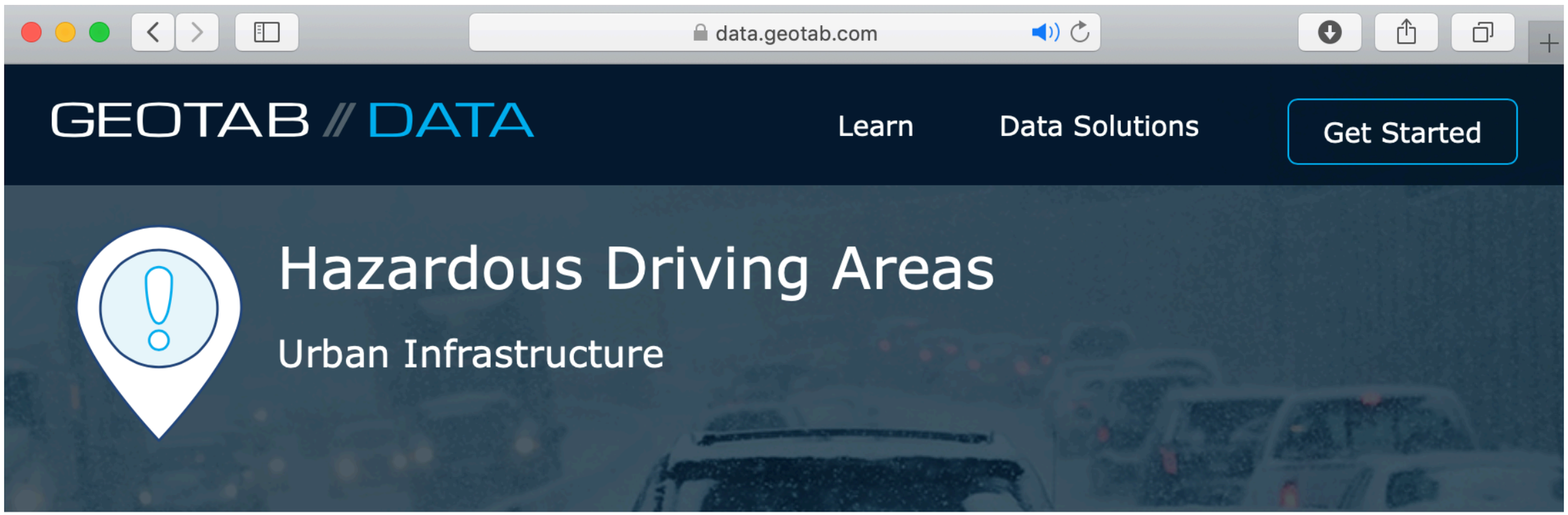


Given harsh braking, accident incidents, and an index measuring severity, location characteristics:

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

2018 Project

FIRST YEAR COURSE: FINAL PROJECT DATA 2018



- Areas Of Idling
- Cell Coverage Dark Spots
- Hazardous Driving Areas
- Intersection Metrics
- Road Impediments
- Searching For Parking

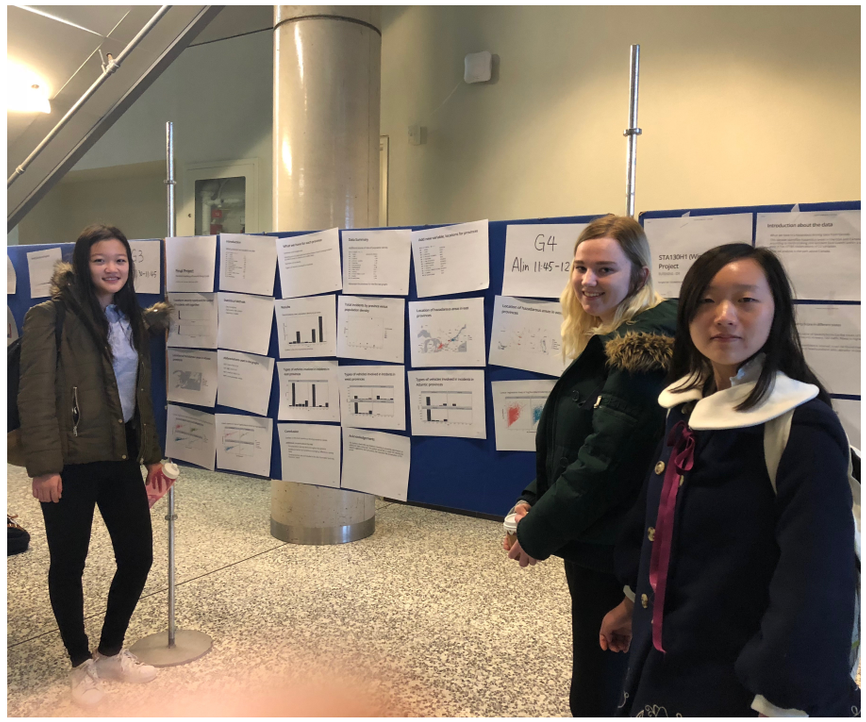
Description

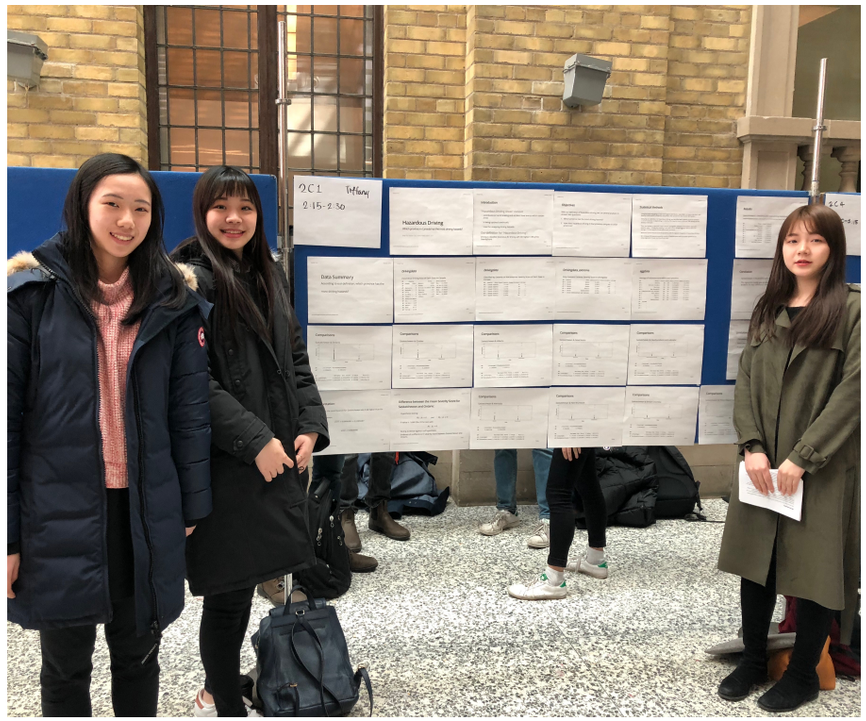
This dataset identifies hazardous areas for driving according to harsh braking and accident level events within a specific area. Each month a new set of dangerous driving areas is produced and encapsulates one year of rolling data (i.e. from the previous month back 1 year).

Associated with each area is a severity score that is based on the frequency of occurrences in the area and the severity of said occurrences.

Real-Time and Historical Incident Data for Accidents and Near-Miss Events

Around the world, over one million people die every year as a result of road traffic crashes, according to a study conducted by the World Health Organization (WHO). This number does not include the massive amount of non-fatal injuries and vehicle damage.





2018 Project Poster Fair

Current Project

https://app.riipen.com/companies/3209

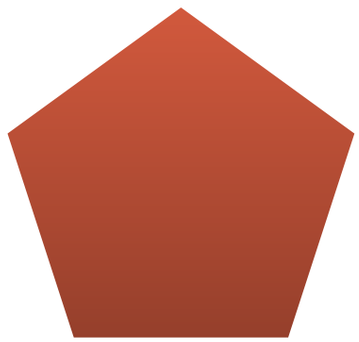
How has use of "Requests" on Riipen changed over time?

Riipen
Vancouver, BC, Canada

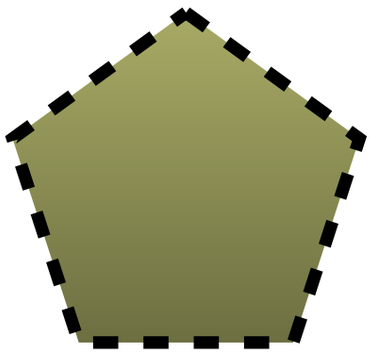
We would be interested in having students look into how requests have evolved over time.

- Number of requests per user, by user type
- Volume of requests (of both types) over time
- Request response time
- Request acceptance rate
- Request expiry rate

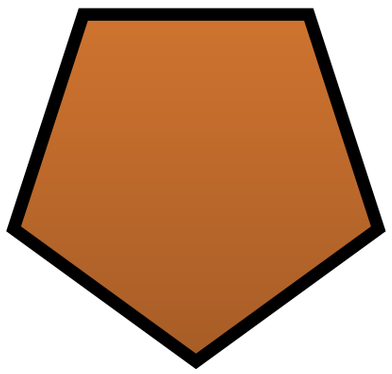
Do users now respond to requests more quickly? Send more requests? Less? Our hypothesis was that they would respond more quickly, and send more requests due to the improved response rate.



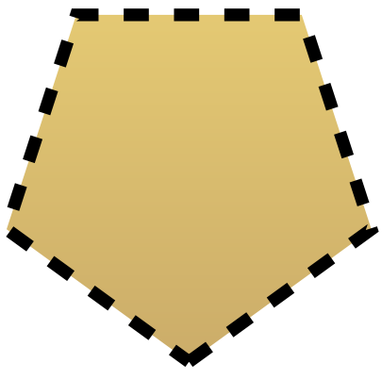
- Oral and written communication
- Interpret practical problem in statistical terms
- Importance of context in devising solutions



- Multiple approaches to solving problems
- Assessing analyses for inconsistencies, bias, etc.



- Reproducible analysis



- Correct application of a variety of statistical methods

quickly, and send more requests due to the improved response rate.

A FIRST COURSE: COMPUTATION

- Our students take 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 students to have skills and approaches for solving real problems as soon as possible
- Develop good habits (implicitly)
  - readable code
  - practices that encourage reproducibility
- Choice: R / R Studio / R Markdown / tidyverse
- Platform: R Studio Cloud

A FIRST COURSE: COMPUTATION

- Ways that computation is used in the course:
  - Simulation as an approach to inference
  - Clean and organize data
  - Carry out data analysis using common methods (summary statistics, visualizations, regression, classification trees)

PLO7:
Use simulation to evaluate statistical methods, support theoretical solutions, and as an approach to inference

PLO8:
Employ technological tools to access, store, clean, and organize for analysis data in a variety of representations and a variety of sizes

PLO10:
Carry out data analysis for common methods using statistical software in a reproducible way

SOME OBSERVATIONS OF THE FIRST COURSE FROM EXPERIENCE

Some challenges:

- Students perceive it is a course about R, rather than a course about statistics or data science
- Appropriately connecting questions to methods and conclusions doesn't come naturally

Some success:

- ~70% of students self-assess as having achieved well the program learning outcomes addressed in the course

Course materials: sta130.utstat.utoronto.ca

ALSO HAPPENING AND IN THE WORKS…

- A computational approach to introductory mathematical statistics
- Communication across the curriculum
- Fewer “traditional” courses
- New 3rd year courses
  - machine learning
  - statistical computation
  - data visualization
- More experiential opportunities
- Assessment of effects of changes

PROGRAM IN A NUTSHELL

First year:

Calculus

Linear algebra

Computer programming

Introduction to Statistical Reasoning and Data Science

PROGRAM IN A NUTSHELL

Second year:

Probability…
Statistical theory…
…with computation

PROGRAM IN A NUTSHELL

Third year:		
Courses in traditional methods	Courses in modern methods / with a computational focus	Courses in theory
Linear models Data collection	Machine Learning I Bayesian statistics Statistical computing Data visualization	Probability Statistical Theory

PROGRAM IN A NUTSHELL

Fourth year:		
Theory	Courses with a project	Capstone
Stochastic processes Mathematical statistics	Statistical computation Machine Learning II Multivariate methods Time Series Spatial Statistics Statistical Genetics Internship	Research seminar Statistical collaboration

REFERENCES

Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics* **26**(4), 745-766.

GAISE (2016). Guidelines for assessment and instruction in statistics education: College report. Alexandria, VA: American Statistical Association.????

American Statistical Association. (2014). "2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science."

Chance, B., & Peck, R. (2015). From Curriculum Guidelines to Learning Outcomes: Assessment at the Program Level. *The American Statistician* **69**(4), 409-416.

Gibbs, A.L. (2018). Building a Foundation in Statistics in the Era of Data Science. In *Proceedings of the Tenth International Conference on Teaching Statistics* (ICOTS10, July, 2018), Kyoto, Japan. Voorburg, The Netherlands: International Statistical Institute.

National Academy of Sciences, Engineering, and Medicine. (2018). Data Science for Undergraduates: Opportunities and Options. Washington, DC: The National Academies Press. https://doi.org/10.17226/25104.

Wiggins, G., & McTighe, J. (2005). *Understanding by Design* (expanded 2nd edition). Alexandria, VA: ASCD.

Zheng, T. (2017). Teaching Data Science in a Statistical Curriculum: Can We Teach More by Teaching Less? *Journal of Computational and Graphical Statistics* **26**(4), 772-774.