MONASH University

ETC3250

Business Analytics

Week 1.
Introduction to Business Analytics & R

24 July 2017

Who are we?

Souhaib Ben Taieb
(Chief examiner)

Mojdeh Shirazi-Manesh
(Teaching Associate)

Zina Quryakos
(Teaching Associate)

Outline

Week	Topic	Chapter	Lecturer
1	Introduction to business analytics & R	1	Souhaib
2	Statistical learning	2	Souhaib
3	Regression for prediction	3	**Anastasios**
	Regression for prediction	3	**David**
4	Classification	4,8	Souhaib
5	Classification	4,8	Souhaib
6	Visualization		Souhaib
7	Data wrangling		Souhaib
8	Resampling	5	Souhaib
9	Dimension reduction	6,10	Souhaib
10	Clustering	10	Souhaib
	Semester break		
11	Advanced regression	6	Souhaib
12	Advanced learning methods	8	Souhaib

Assessment

- Final exam (Open book) (2 hours): 60%
- One project due at the end of the semester: 20%
- Ten short weekly assignments: 20% (2% each)

Task	Due Date	Value
Final exam	Official exam period	60%
Project	Fri 27 October	20%
Assignments 1–10	Sunday 11:55pm	20%

Moodle site

- Includes all lecture notes, handouts, assignments
  $\rightarrow$ `https://github.com/bsouhaib/BA2017`
- Forum for asking questions, etc.
- **No email please — use the forum**.
- Assignment submissions

Key reference

James, Witten, Hastie and Tibshirani (2012) *An Introduction to Statistical Learning*. Springer.

www.statlearning.com

- Free pdf online
- Data sets in associated R package **ISLR**
- R code for examples

What is business analytics?

"*Business analytics* is the **scientific process** of transforming **data** into **insight** for making better **decisions**"

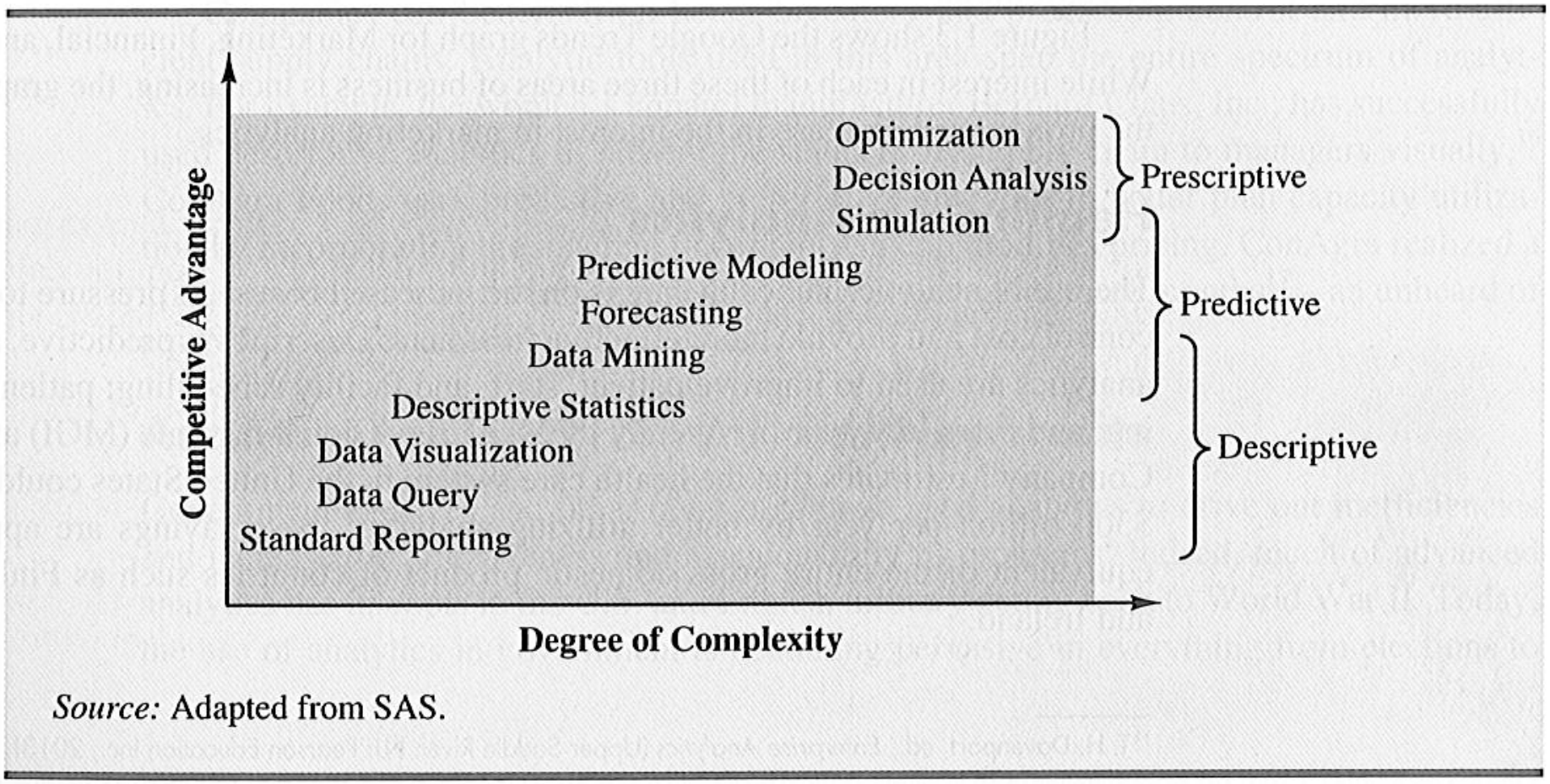
- **Broader than business intelligence** which focuses on describing and predicting performance.
- **Broader than econometrics** as we are interested in more than economics and finance.
- **Narrower than data science** as we are focusing on business issues.

What is business analytics?

"*Business analytics* is the **scientific process** of transforming **data** into **insight** for making better **decisions**"

- **Broader than business intelligence** which focuses on describing and predicting performance.
- **Broader than econometrics** as we are interested in more than economics and finance.
- **Narrower than data science** as we are focusing on business issues.

What is business analytics?



Source: Adapted from SAS.

What is business analytics?

- Financial Analytics
- Human Resource Analytics
- Marketing Analytics
- Health Care Analytics
- Supply Chain Analytics
- Analytics for Government and Nonprofits
- Sport Analytics
- Web Analytics

Related fields

"Statistics is the **science of learning from data**, and of **measuring**, **controlling**, and **communicating uncertainty**; [...]".

"Machine learning is a **scientific discipline** that explores the **construction and study of algorithms** that can **learn from data**".

"Data mining, [...], is the **computational process** of **discovering patterns** in **large data sets** involving methods at the intersection of **artificial intelligence**, **machine learning**, **statistics**, and **database systems**".

"Data Science means the **scientific study** of the **creation**, **validation** and **transformation of data** to **create meaning**".

Related fields

"Statistics is the **science of learning from data**, and of **measuring**, **controlling**, and **communicating uncertainty**; [...]".

"Machine learning is a **scientific discipline** that explores the **construction and study of algorithms** that can **learn from data**".

"Data mining, [...], is the **computational process** of **discovering patterns** in **large data sets** involving methods at the intersection of **artificial intelligence**, **machine learning**, **statistics**, and **database systems**".

"Data Science means the **scientific study** of the **creation**, **validation** and **transformation of data** to **create meaning**".

Related fields

"Statistics is the **science of learning from data**, and of **measuring**, **controlling**, and **communicating uncertainty**; [...]".

"Machine learning is a **scientific discipline** that explores the **construction and study of algorithms** that can **learn from data**".

"Data mining, [...], is the **computational process** of **discovering patterns** in **large data sets** involving methods at the intersection of **artificial intelligence**, **machine learning**, **statistics**, and **database systems**".

"Data Science means the **scientific study** of the **creation**, **validation** and **transformation of data** to **create meaning**".

Related fields

"Statistics is the **science of learning from data**, and of **measuring**, **controlling**, and **communicating uncertainty**; [...]".

"Machine learning is a **scientific discipline** that explores the **construction and study of algorithms** that can **learn from data**".

"Data mining, [...], is the **computational process** of **discovering patterns** in **large data sets** involving methods at the intersection of **artificial intelligence**, **machine learning**, **statistics**, and **database systems**".

"Data Science means the **scientific study** of the **creation**, **validation** and **transformation of data** to **create meaning**".

Why study business analytics?

Play/Pause

Why study business analytics?

Play/Pause

Some quotes

By 2018, the US could face a shortage of up to 190.000 workers with analytical skills —McKinsey

Top-ranked jobs 2016: CareerCast

1. Data Scientist
2. Statistician
3. Information Security Analyst
4. Audiologist
5. Diagnostic Medical Sonographer
6. Mathematician
7. Software Engineer
8. Computer Systems Analyst
9. Speech Pathologist
10. Actuary

Some quotes

***Data Scientist*: The Sexiest Job of the 21st Century** — Thomas H. Davenport and D. J. Patil, Harvard Business Review, October 2012.

To have any hope of extracting anything useful from big data, . . . effective inferential skills are vital. That is, at the heart of extracting value from big data lies statistics — *David J. Hand, 2014.*

Most of my life I went to parties and heard a little groan when people heard what I did. Now they are all excited to meet me — *Robert Tibshirani, a statistics professor at Stanford University, New York Times, January 26, 2012.*

Am I a data scientist?

April 2013: Larry Wasserman blog

Data science: the end of statistics?

If you're analyzing data, you're doing statistics. You can call it data science or informatics or analytics or whatever, but it's still statistics. normaldeviate.wordpress.com/2013/04/13/data-science-the-end-of-statistics/

Am I a data scientist?

April 2013: Larry Wasserman blog

Data science: the end of statistics?

If you're analyzing data, you're doing statistics. You can call it data science or informatics or analytics or whatever, but it's still statistics.

normaldeviate.wordpress.com/2013/04/13/data-science-the-end-of-statistics/

April 2013: Karl Broman blog

Data science is statistics

You may not like what some statisticians do. You may feel they don't share your values. They may embarrass you. But that shouldn't lead us to abandon the term "statistics".

kbroman.wordpress.com/2013/04/05/data-science-is-statistics/

Am I a data scientist?

April 2013: Larry Wasserman blog

Data science: the end of statistics?

If you're analyzing data, you're doing statistics. You can call it data science or informatics or analytics or whatever, but it's still statistics.

normaldeviate.wordpress.com/2013/04/13/data-science-the-end-of-statistics/

April 2013: Karl Broman blog

Data science is statistics

You may not like what some statisticians do. You may feel they don't share your values. They may embarrass you. But that shouldn't lead us to abandon the term "statistics".

kbroman.wordpress.com/2013/04/05/data-science-is-statistics/

July 2013: ASA President blog

Davidian: **Aren't *we* data science?**

Am I a data scientist?

November 2013: Andrew Gelman blog

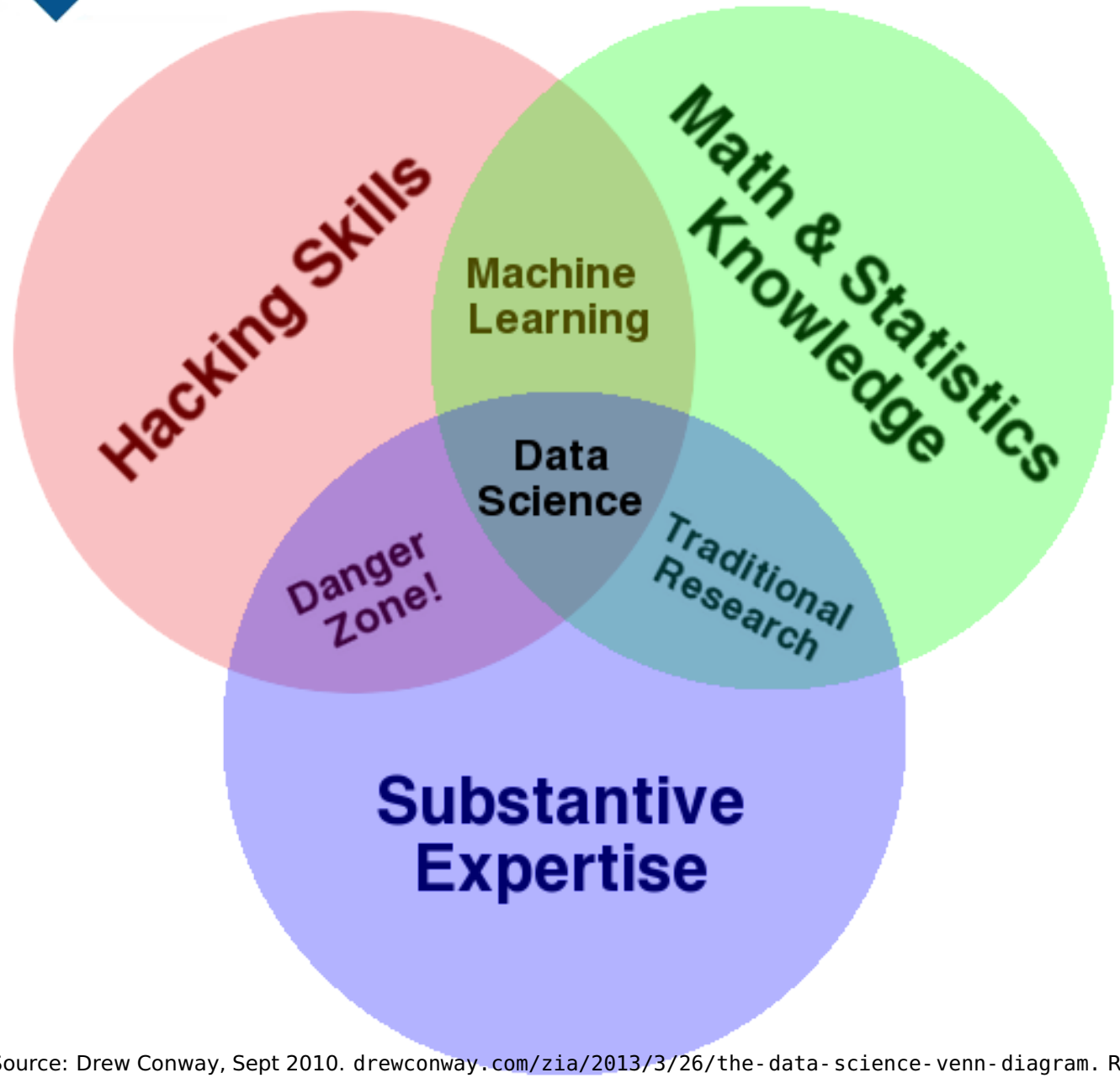
Statistics is the *least* important part of data science

There's so much that goes on with data that is about computing, not statistics. I do think it would be fair to consider statistics as a subset of data science . . .

Statistics is important—don't get me wrong—statistics helps us correct biases . . . estimate causal effects . . . regularize so that we're not overwhelmed by noise . . . fit models . . . visualize data . . . I love statistics! But it's not the most important part of data science, or even close.

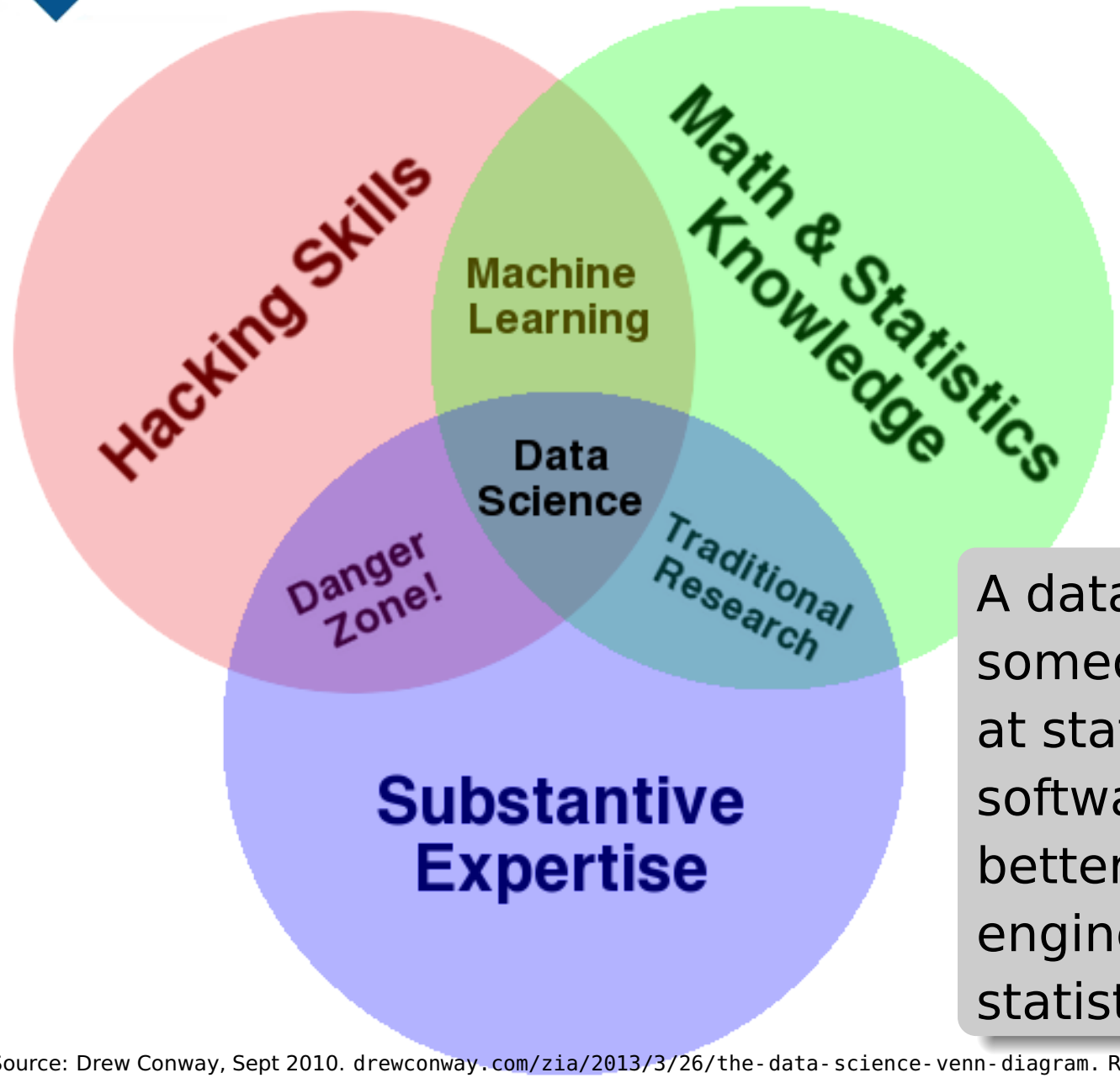
andrewgelman.com/2013/11/14/statistics-least-important-part-data-science/

Am I a data scientist?



Source: Drew Conway, Sept 2010. drewconway.com/zia/2013/3/26/the-data-science-venn-diagram. Reproduced under a Creative Commons License.

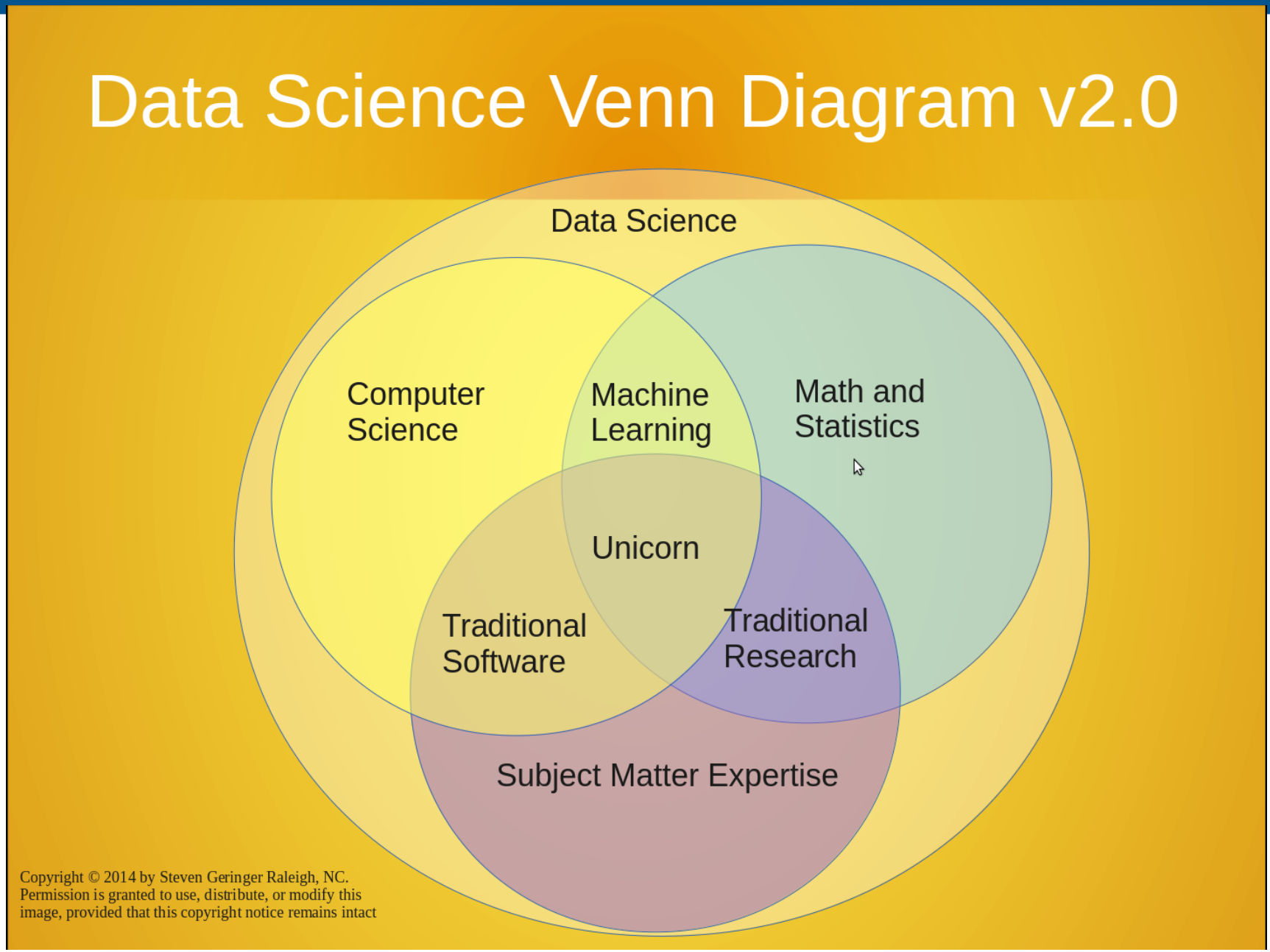
Am I a data scientist?



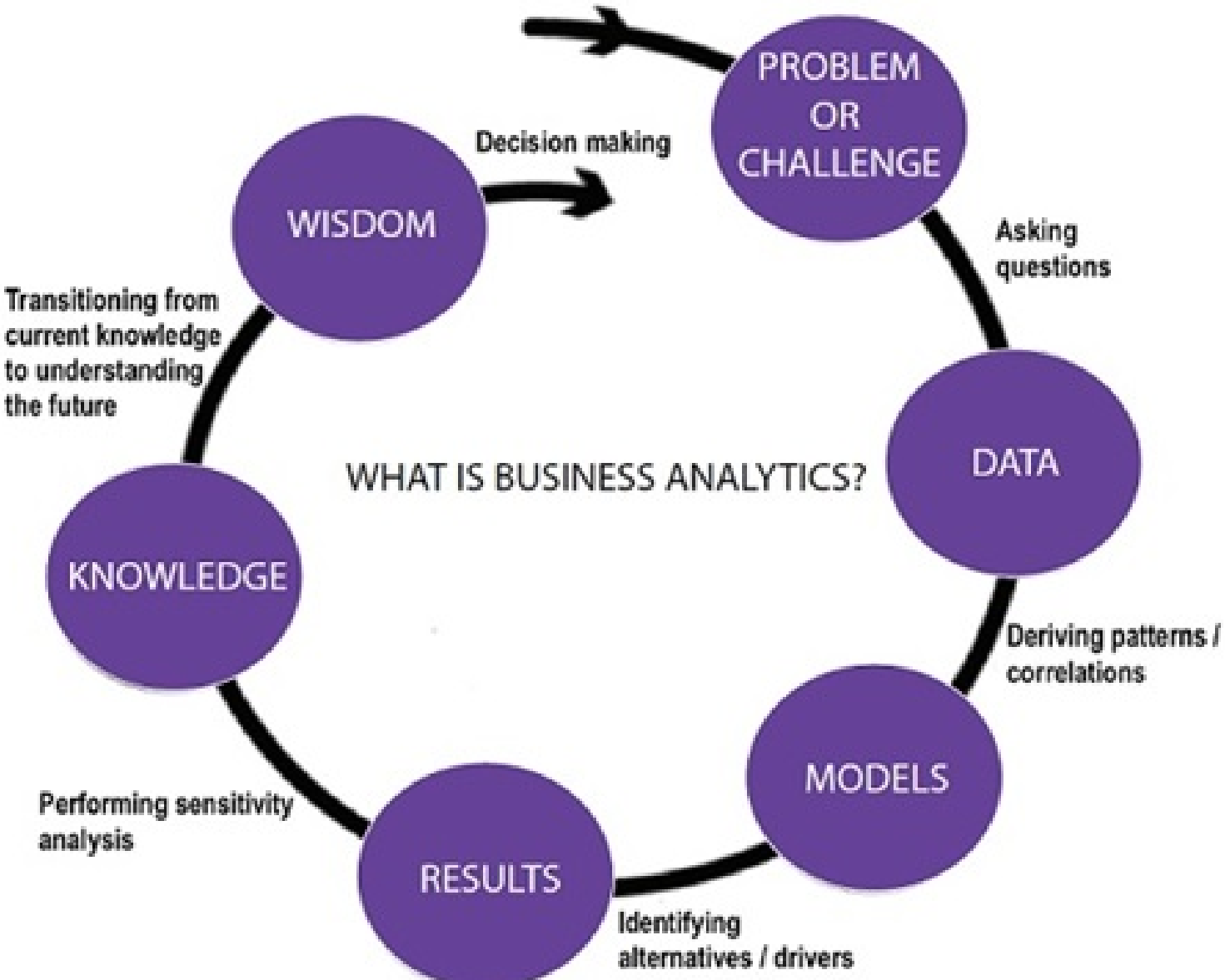
A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.

Source: Drew Conway, Sept 2010. drewconway.com/zia/2013/3/26/the-data-science-venn-diagram. Reproduced under a Creative Commons License.

Am I a data scientist?



The business analytics process



Source: http://www.stern.nyu.edu/programs-admissions/executive-education/short-courses/schedule/short-course-program-7

The business analytics tools

- Pulling together and **cleaning data**
- Exploring and **visualizing data**
- **Fitting, comparing and assessing models**
- Tools for **fitting models**: optimization, training and testing
- Tools for **understanding randomness**: simulation, resampling, permutation, cross-validation
- Tools for **handling large data sets**: dimension reduction, regularization, distributed computing.

Learning goals

1. Select and develop appropriate models for regression, classification or clustering
2. Estimate and simulate from a variety of statistical models, and measure the uncertainty of a prediction using resampling methods
3. Manage large data sets in a modern software environment, and explain and interpret the analyses undertaken clearly and effectively
4. Apply business analytic tools to produce innovative solutions in finance, marketing, economics and related areas

Teaching and learning approach

Two 1-hour lectures and a one 1.5 hour lab class each week for 12 weeks.

Learning goals

1. Select and develop appropriate models for regression, classification or clustering
2. Estimate and simulate from a variety of statistical models, and measure the uncertainty of a prediction using resampling methods
3. Manage large data sets in a modern software environment, and explain and interpret the analyses undertaken clearly and effectively
4. Apply business analytic tools to produce innovative solutions in finance, marketing, economics and related areas

Teaching and learning approach

Two 1-hour lectures and a one 1.5 hour lab class each week for 12 weeks.

R

R

RStudio