MONASH University

ETC3250

Business Analytics

Introduction

27 February 2018

Who are we?

Souhaib Ben Taieb
(Chief examiner)

Mojdeh Shirazi-Manesh
(Teaching Associate)

Zina Quryakos
(Teaching Associate)

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

Outline

Week	Topic	Chapter	Lecturer
1	Introduction	1	Souhaib
2	Statistical learning	2	Souhaib
3	Regression	3	Souhaib
4	Classification	4	Souhaib
5	Clustering	10	Souhaib
	Semester break		
6	Model selection and resampling methods	5	Souhaib
7	Dimension reduction	6,10	Souhaib
8	Advanced regression	6	Souhaib
9	Advanced regression	6	Souhaib
10	Advanced classification	9	Souhaib
11	Tree-based methods	8	Souhaib
12	Project presentation		Souhaib

Assessment

- Final exam (*Open book*) (2 hours): **60%**
- One project due at the end of the semester (group of 5 students): **20%**
- Five fortnightly assignments (group of 3 students): **20%** (4% each)

Task	**Due Date**	**Value**
Final exam	Official exam period	60%
Project	Fri 18 May	20%
Assignments 1–5	Thursday 11:55pm	20%

Assignments

No	Release Date (Thursday)	Due Date (Sunday next week)
1	8 March	18 March, 11:55pm
2	22 March	1st April, 11:55pm
3	12 April	22 April, 11:55pm
4	26 April	6 May, 11:55pm
5	10 May	20 May, 11:55pm

Communication

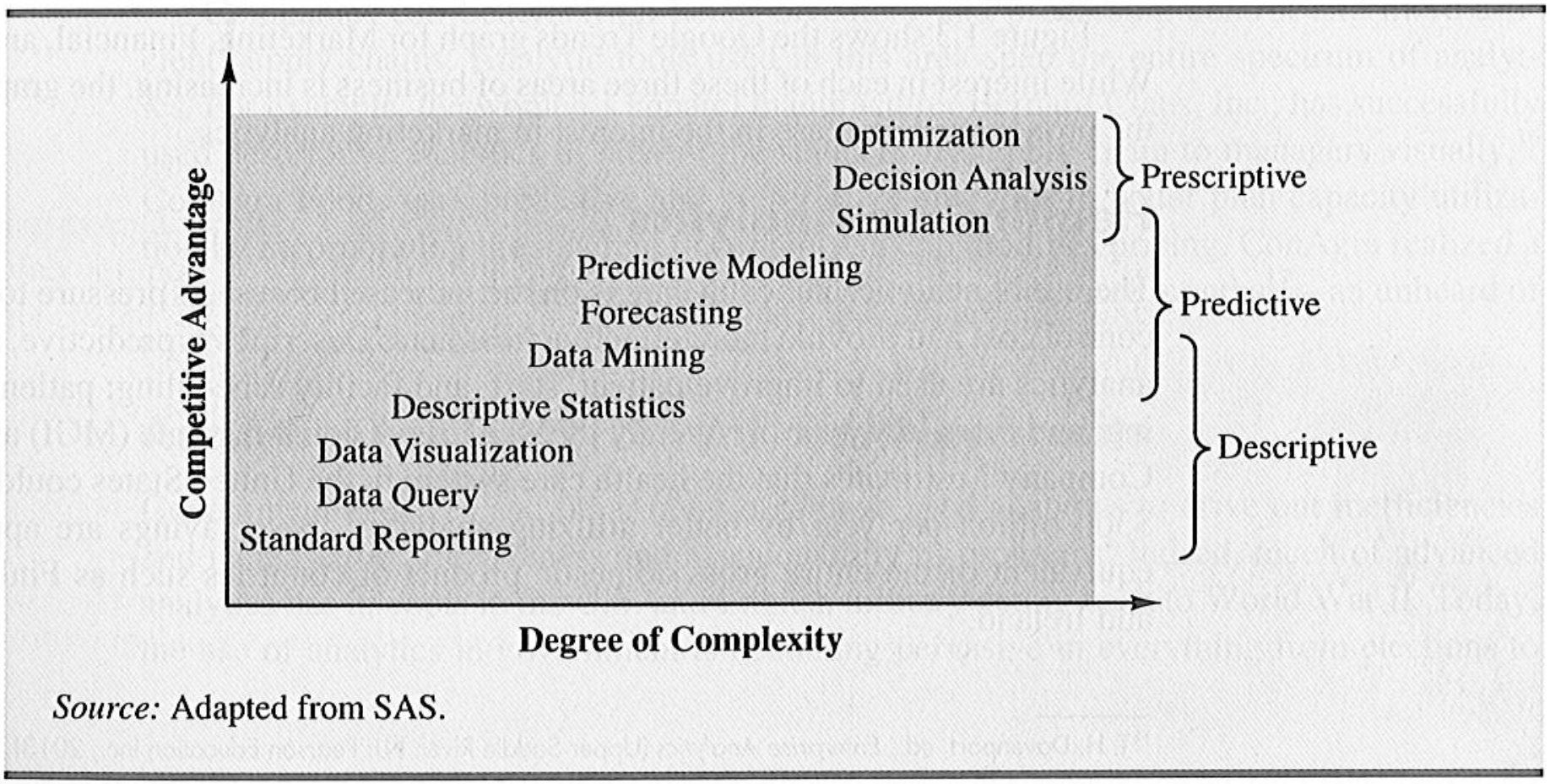
- **Website**
  - `https://github.com/bsouhaib/BA2018`
  - Lecture notes
- **Moodle**
  - `https://moodle.vle.monash.edu/course/view.php?id=42004`
  - Forum for asking questions, etc.
  - Assignment submissions
  - **No email please — use the forum**

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.

Examples



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**".

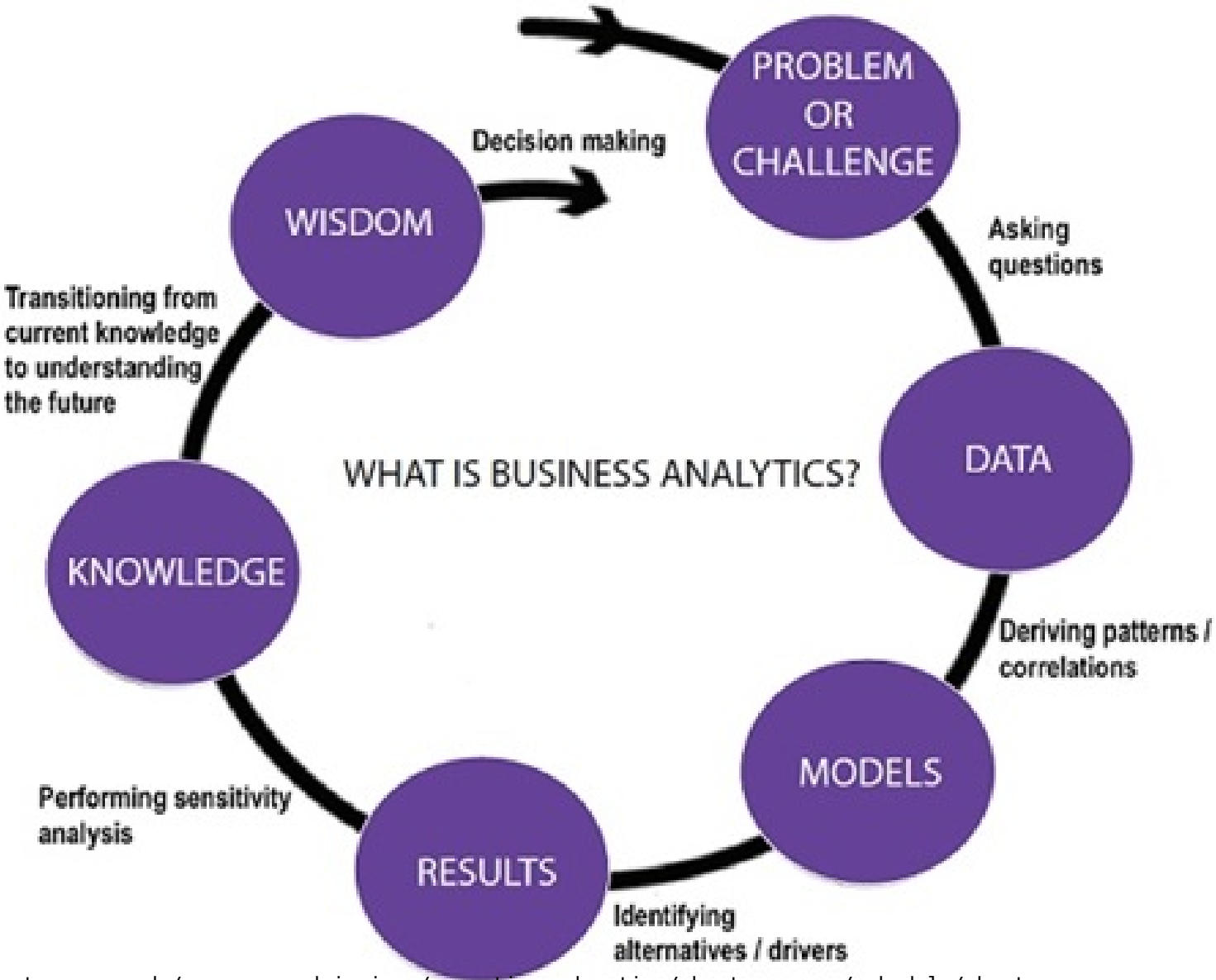
"Data mining, [...], is the **computational process** of **discovering patterns** in **large data sets** involving methods at the intersection of **artificial intelligence**, **machine learning**, **statistics**, and **database systems**".

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

Why study business analytics?

Play/Pause

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.

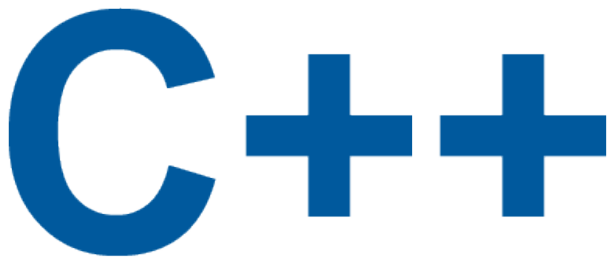
Learning goals

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.

Programming languages



R

RStudio

ETC3555 - Statistical Machine Learning

This unit covers the methods and practice of statistical machine learning for modern data analysis problems. Topics covered will include ***recommender systems***, ***social networks***, ***text mining***, ***matrix decomposition and completion***, *and* ***sparse multivariate methods***. *All computing will be conducted using the* ***R programming language***.

Prerequisites: ETC3250 or FIT3154