Introduction to Data Analytics - Course Outline

Ahmed Eleish

Data Analytics ITWS-4600/ITWS-6600/CSCI-4600 BCBP- 4600/ MGM-4600/MGMT-6600
Group 1 Module 1, August 29th, 2025

Tetherless World Constellation
Rensselaer Polytechnic Institute

Admin information

- Class: ITWS-4600/ 6600/CSCI-4600/MGMT 4600/6600/BCBP 4600
- Hours:
  Section 01: 10:00am ET - 11:50am ET on Tues/Fri – Sage 2701
  Section 02: 02:00pm ET - 03:50pm ET on Tues/Fri – Lally 02
- Instructor: Ahmed Eleish
- Instructor contact: eleisa2@rpi.edu
- Instructor office hours: Thursday from 02:00 PM - 4:00 PM ET or by appointment/email
- Instructor office location: Amos Eaton 134
- TA: Manish Bhat - bhatm@rpi.edu
- TA office hours: TBA
- TA office: TBA
- Web site: https://tw.rpi.edu/classes/data-analytics-fall-2025
- LMS (http://lms.rpi.edu/)

Contents

- Intro – about this course
- Learning objectives
- What is expected
- What skills are needed
- Outline of the course
- Definitions and why Analytics means more than Analysis
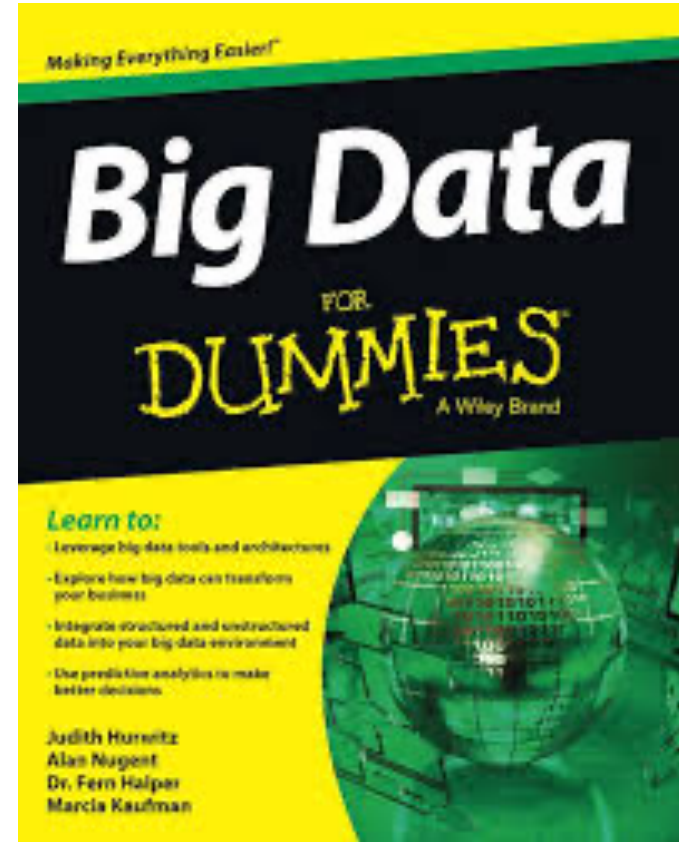
Objectives

• Introduce students to relevant methods to recognize and apply quantitative algorithms, techniques and interpretation.

• To develop students' strategic thinking skills, combined with a solid technical foundation in data and model-driven decision-making.

• Develop ability to apply critical and analytical methods to formulate and solve science, engineering, medical, and business problems.

Objectives

• Students will examine real-world examples to place data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science.

• By the end of the course, students can effectively communicate analytic findings to non-specialists.

- [6600 level] Students must develop and demonstrate a working knowledge of decision making under uncertainty, be able to optimize models that incorporate parameters.

4600 versus 6600

- 6600 students are assessed at:
– Higher level of demonstration
– Additional questions or tasks in assignments

- 4400 students are welcome to complete these higher requirements for extra credit

Assessment and Assignments

• Via written assignments with specific percentage of grade allocation provided with each assignment.

• Via individual oral presentations with specific percentage of grade allocation provided.

• Via participation and attendance in labs (not to exceed 5% of total, start with 5% and lose % by not attending labs).

• Late submission policy: first time with valid reason – no penalty, otherwise 20% of score deducted each late day. Talk to me EARLY if you are having schedule problems in completing assignments.

Assessment and Assignments

• Readings are given when needed to support key topics or to complete assignments

• Will not be discussed in class unless there are questions

• You will mostly perform individual work but you are encouraged to work with others in the lab sessions

Current assignment structure (no final exam)

- Assignment 1: Review of a DA Case Study. - 5% (written)
- Assignment 2: Distributions, Linear Models, Classification & Clustering – 10% (written + figures)
- Assignment 3: Term project proposal - 5% (oral/written)
- Assignment 4: Term project - 30% (25% written, 5% presentation - oral)
- Assignment 5: Preliminary and Statistical Analysis - 15% (written + figures)
- Assignment 6: Patterns, trends, relations: model development and evaluation - 15% (written + figures)
- Assignment 7: Predictive and Prescriptive Analytics - 15% (15% written + figures)
- 5% participation (labs)

Project options (examples)

- Social networks
- Financial
- Social-economic, marketing
- Geo/ space science data
- Network/ security data
- Linked data
- Movie databases
- Transportation
- Competitions (Web and local)*
- Research Projects*

Research Projects & Competitions* : Need the Instructor's approval for the datasets

Academic Integrity

• Student-teacher relationships are built on trust. For example, students must trust that teachers have made appropriate decisions about the structure and content of the courses they teach, and teachers must trust that the assignments that students turn in are their own. Acts, which violate this trust, undermine the educational process. The Rensselaer Handbook of Student Rights and Responsibilities defines various forms of Academic Dishonesty and you should make yourself familiar with these. In this class, all assignments that are turned in for a grade must represent the student's own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration.

• Submission of any assignment that is in violation of this policy will result in a penalty. If found in violation of the academic dishonesty policy, students may be subject to two types of penalties. The instructor administers an academic (grade) penalty of full loss of grade for the work in violation, and the student may also enter the Institute judicial process and be subject to such additional sanctions as: warning, probation, suspension, expulsion, and alternative actions as defined in the current Handbook of Student Rights and Responsibilities.

• Second violation will result in failure of the course.

• **If you have any question concerning this policy before submitting an assignment, please ask for clarification.**

What is expected

- Attend class, complete assignments, participate

- Ask questions, offer answers in class

- Work individually on assignments

- Work in a group on labs, learn from each other, help each other especially with software

- Work constructively in class labs

Skills needed

- Basic knowledge of data structures, computer programming
- Literacy with computers and applications that can handle the data we will use
- Ability to access internet, servers and retrieve/ acquire data, **install/ configure software**
- **Pick up R programming**, terminology and syntax, and some refinement
- Presentation of proposal, project and assignment results

Current Syllabus/Schedule

- Web site: https://tw.rpi.edu/classes/data-analytics-fall-2025

- Note: in general lectures are on Tuesdays, labs on Fridays

- Attendance is taken on lab day; lab attendance is part of participation grades
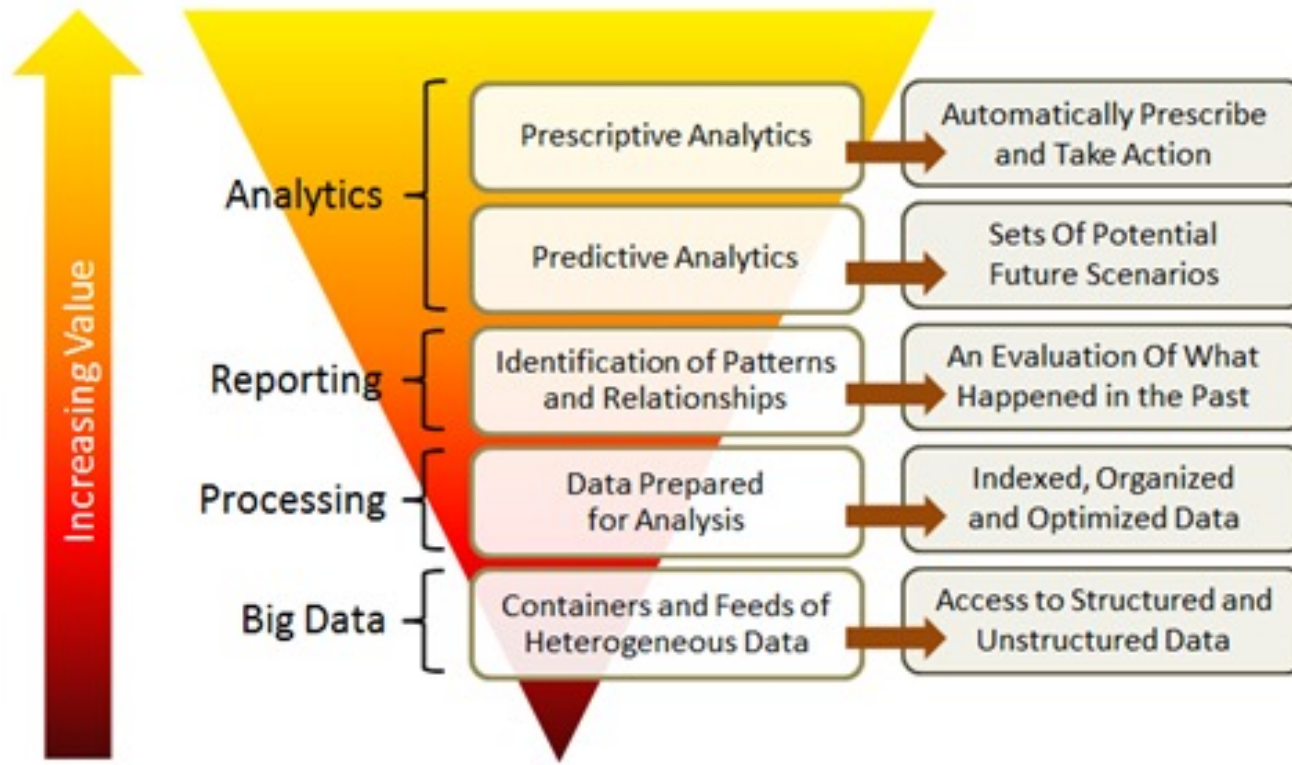
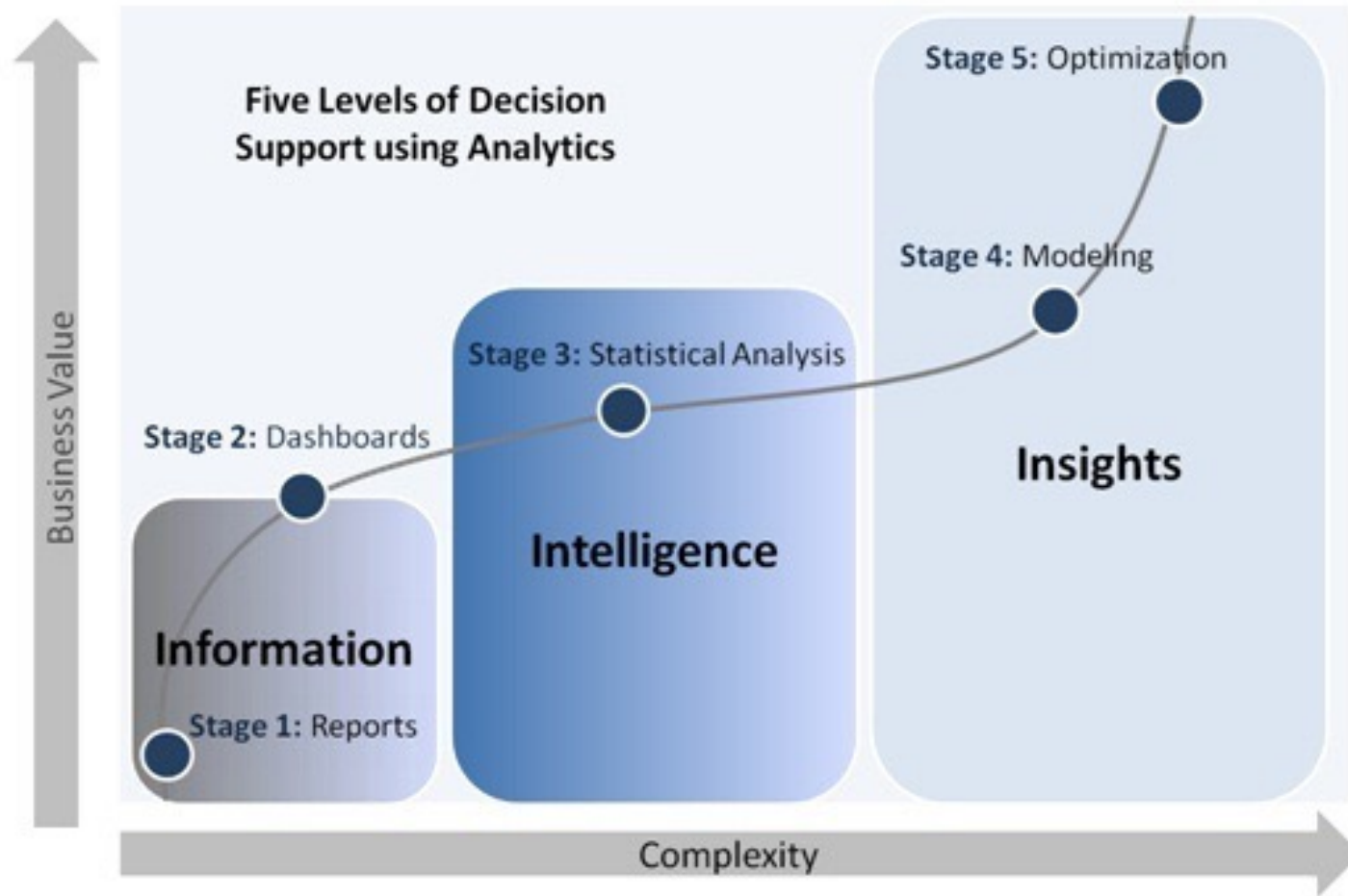# Questions so far?

Peter Arthur Fox, Ph.D.
1959 - 2021

Introductions

- Who you are

- Why you are here

- What you expect to learn

- Your interests/hobbies

The nature of the challenge

Five Levels of Decision Support using Analytics

Business Value

Complexity

Stage 5: Optimization

Stage 4: Modeling

Stage 3: Statistical Analysis

Stage 2: Dashboards

Stage 1: Reports

Information

Intelligence

Insights

High Level Perspective

• People make decisions every day and they are increasingly using resources/ services (that run on computers) to assist or decide for them.

• Knowledge can translate to "power":
– Or accurate/ reliable knowledge is actionable

• Knowledge is acquired from (often multiple sources of) information and data sources

# So what does that mean in practice?

Definitions (at least for this course)

• Data - are encodings that represent the qualitative or quantitative attributes of a variable or set of variables.

• Data (plural of "datum", which is seldom used) - are typically the results of measurements, computations, or observations and can be the basis of graphs, images of a set of variables.

• Data - are *often* viewed as the lowest level of abstraction from which information and knowledge are derived.

• Model = formula/equation that approximates the phenomenon being studied.

Data-Information-Knowledge Ecosystem
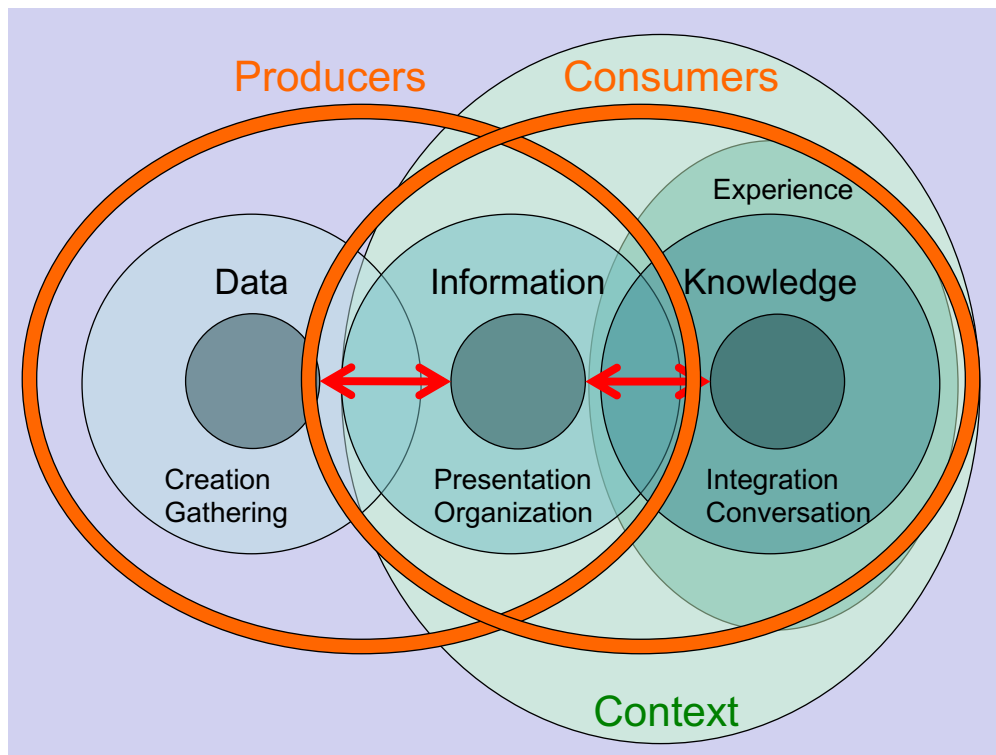
And then there ~~is~~ was Big Data

5 V's: volume, variety, veracity, velocity, value
https://en.wikipedia.org/wiki/Big_data

Journals/ conferences: IEEE, http://www.liebertpub.com/big

In short: crawl before you walk, before you run, before you become famous ;-)

A view from IBM ...

• "Anyone who wants to learn something about data analytics should take a road trip. Myriad real-time decisions must be made based on analysis of static information as well as ever-changing conditions. Data about traffic, weather, road construction, fuel, time, current location and available funds are just a few of the factors."

• This information and much more are needed to answer questions like:
  – If I skip this gas station, will I run out of gas before the next one?
  – Is it worth driving 50 miles out of the way to see the Corn Palace? How late will that side trip make us?
  – Can I make it to Billings, Montana., by sunset or should I look for a place to stop?

Case Studies

• Sports Analytics–Moneyball (http://www.imdb.com/title/tt1210166/), Nate Silver (http://en.wikipedia.org/wiki/Nate_Silver)

• Marketing Analytics – Amazon, Walmart/Target, etc.

• Netflix, etc. recommenders – "If you liked,..."

http://www.slideshare.net/lsakoda/case-studies-utilizing-real-time-data-analytics

Analysis

- Software packages / environments:
  – Gnu R
  – Rstudio
    - Extensive libraries
  – <Jupyter Notebook/ Lab>

- Going from preliminary to exploratory analysis...

- Parametric (assumes or asserts a probability distribution) and non-parametric statistics

What are "statistics"?

• The term "statistics" has **two common meanings**, which we want to clearly separate: **descriptive** and **inferential** statistics.

• But to understand the difference between descriptive and inferential statistics, we must first be clear on the difference between **populations** and **samples**.

• Coming up in later lectures!

Summary

• We'll work our way through the stages of analytics.

• We'll use both laptop installed software and potentially some server data infrastructures for analytics to give you practical experience.

• We'll cover algorithms, parameter choices, models, results, interpretation, and the software.

* This is a fast-paced course *

Reading/ watching

- SportsAnalytics–Moneyball
- (http://www.imdb.com/title/tt1210166/),
- NateSilver(http://en.wikipedia.org/wiki/Nate_Silver)
- http://www.slideshare.net/lsakoda/case-studies- utilizing-real-time-data-analytics
- http://www.marketquotient.com/case-studies.html
- http://www.ibm.com/analytics/us/en/case-studies/

- More in the Assignment …

Reference Material

- On course website

Files

- We will use Box, please make sure your accounts are set up

- Links will be shared in slides

Assignment 1

Next class: Tuesday Sep 2nd – quick refresher on statistics and intro to labs

Head start for lab - R

- http://lib.stat.cmu.edu/R/CRAN/ - install this first
- http://cran.r-project.org/doc/manuals/
- http://cran.r-project.org/doc/manuals/R-lang.html
- R Studio

(https://www.rstudio.com/products/rstudio/)
(desktop version)

Thanks!

* Have a good weekend!! (and work on the assignment!)