

FIT1043 Introduction to Data Science

Module 2

Data Models in Organisations

2018 Lecture 3

Monash University

Discussion: Motion Charts

Advantages:

- ▶ time dimension allows deeper insights & observing trends
- ▶ good for exploratory work
- ▶ motion allows identification for this out of common “rhythm”
- ▶ “appeal to the brain at a more instinctual intuitive level”

Disadvantages:

- ▶ not suited for static media
- ▶ display can be overwhelming, and controls are complex
- ▶ not suited for representing all types of data, e.g. other graphics might be suitable for business data
- ▶ “data scientists who branch into visualization must be aware of the limitations of uses”

Discussion: Data Science Jobs

Data Science Job Market in Australia

- ▶ smaller (per capita) market compared to USA & UK, where giant industry players are making better use of Data Science
- ▶ currently lacks proper synchronization between academia, practice and policy

Job Adverts

- ▶ **communication skills** and **domain expertise** are rated highly
 - ▶ the Metromap doesn't mention either!
 - ▶ different jobs require different toolset skills
- ▶ see Adzuna's [CV upload page](#) for an interesting application!

MARS Question

Data scientists are primarily people who develop insights with data.

A. TRUE

B. FALSE



Unit Schedule: Modules

Module	Week	Content
1.	1 2	Overview and look at projects (job) roles, and the impact
2.	3	Data business models / Application areas
3.	4 5	Characterising data and "big" data Data sources and case studies
4.	6 7	Resources and standards Resources case studies
5.	8 9 10	Data analysis theory Regression and decision theory Data analysis process
6.	11 12	issues in data management GUEST SPEAKER & EXAM INFO

Case Study: City Science

City X

- ▶ is a research lab at UNSW, formerly at Monash
- ▶ focuses on improving scientific understanding of cities and providing new insights into developing a data-driven approach to design, plan, and operate future cities
- ▶ see video on [Pedestrian traffic](#) in Melbourne

Case Study: Is Big Data Better?

See this Strata-Hadoop video from 2013:

- ▶ [“Is Bigger Really Better? Predictive Analytics with Fine-grained Behavior Data”](#)
- ▶ by [Foster Provost](#), author of the book [“Data Science for Business”](#)

Data and Decision Models

(ePub section 2.1)

some general models for understanding businesses and decision making:

- ▶ value chains:
 - ▶ sequence of processes done to create value
- ▶ analytic levels:
 - ▶ to broadly classify kinds of analysis
- ▶ influence diagrams:
 - ▶ method for modelling decision making
 - ▶ will extend later to model learning

Data and Decision Models: value chains

sequence or processes for data science

Our Standard Value Chain

Collection: getting the data

Engineering: storage and computational resources

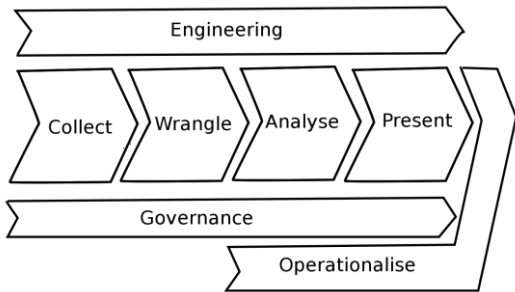
Governance: overall management of data

Wrangling: data preprocessing, cleaning

Analysis: discovery (learning, visualisation, etc.)

Presentation: arguing that results are significant and useful

Operationalisation: putting the results to work

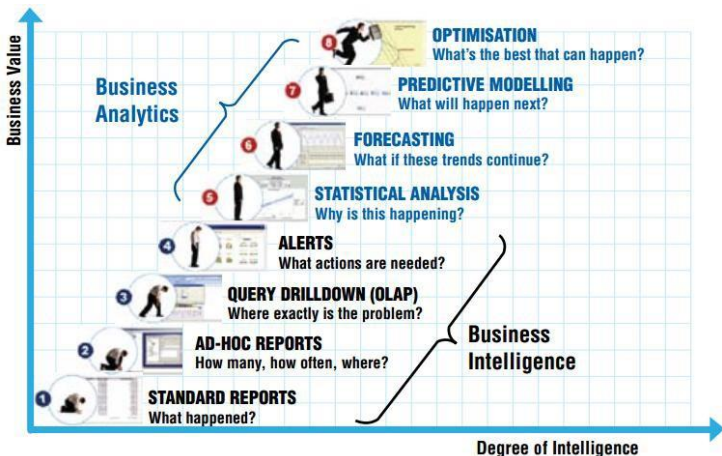


we will refer to
this throughout!

Data and Decision Models: Analytic levels

descriptive terms used to broadly classify kinds of analysis

SAS Analytic Levels



See ["Eight Levels of Analytics"](#).

Analytic Levels

Descriptive Analytics: **gain insight** from historical data

- ▶ plot sales results by region and product category
- ▶ correlate with advertising revenue per region

Predictive analytics: **make prediction** using statistical and machine learning techniques

- ▶ predict next quarter's sales results using economic projections and advertising targets

Prescriptive analytics: **recommend decisions** using optimization, simulation, *etc.*

- ▶ recommend which regions to advertise in given a fixed budget

primarily a descriptive classification for general discussions

MARS Question



Which of the following is an optimization or prescriptive analytics task (as opposed to a predictive analytics task)?

- A. Recommending a traffic route based on prior data for the time of data and incident reports.
- B. Recommend a credit card transaction be checked for fraud.
- C. Recommending a web advertisement to show a website visitor.
- D. Recommending that a doctor check the patient's nursing notes for evidence of fungal infection.

Data and Decision Models: Influence diagrams

method for modelling data and decision making

- ▶ what do we know?
- ▶ what don't we know?
- ▶ what value can we get?
- ▶ what decisions do we have to make?

Modelling

Real World Out There



Theory



Identification of details
relevant to description,
translation of 'real' objects
into variables of the model

Model

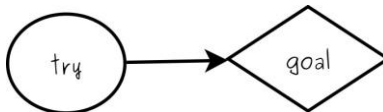
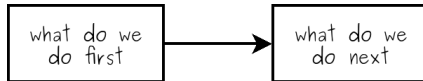


from [*the BackReaction blog by Sabine Hossenfelder*](#)

Modelling

- ▶ “all models are wrong, but some are useful”
... George Box
- ▶ “the approximate nature of the model must always be borne in mind”
... George Box
- ▶ “the purpose of models is not to fit the data but to sharpen the questions”
... Samuel Karlin

Influence Diagrams



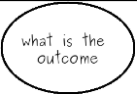
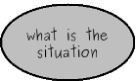


Motivating Influence Diagrams

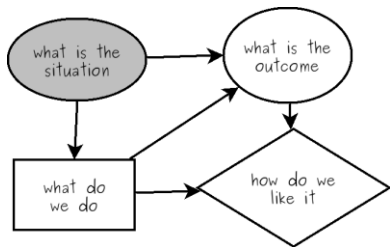
Influence Diagrams (a.k.a Decision Graphs) are

- ▶ directed graphical model with 4 types of nodes:
 - ▶ chance nodes, known variable nodes, action/decision nodes and objective/utility nodes
- ▶ model the “influences”, “causes”, random (“chance”) outcomes, “actions”, “goals” involved in a decision problem
- ▶ provide a coarse abstraction, a conceptual model

a conceptualisation aid to get you thinking about actions, values, and unknowns

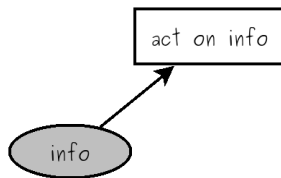
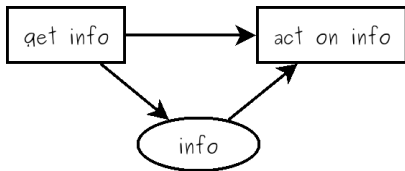
Influence Diagrams

 <p>a chance variable, an uncertain quantity</p>	 <p>a known variable, an certain quantity</p>
 <p>a decision or action</p>	 <p>an objective is a measure of value</p>

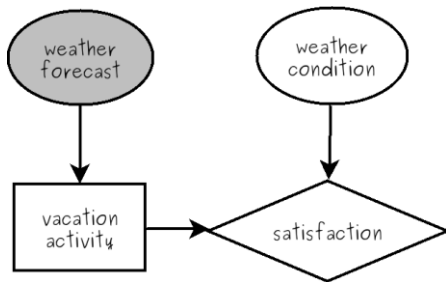


nodes are strung together in a directed graph to convey influence

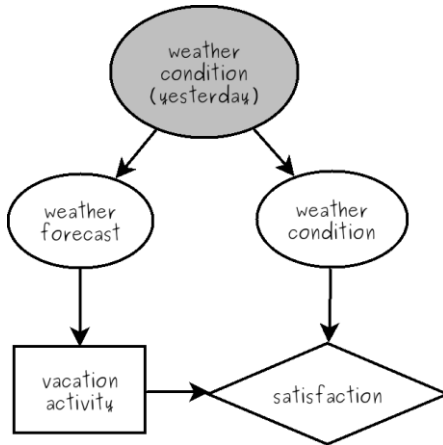
Influence Diagrams – Arcs



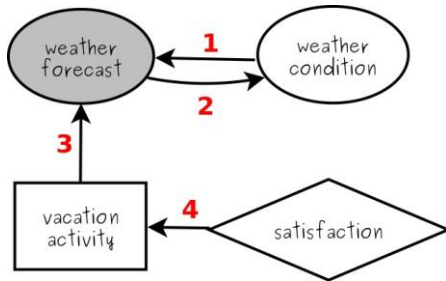
Last Minute Vacation



Last Minute Vacation (cont)

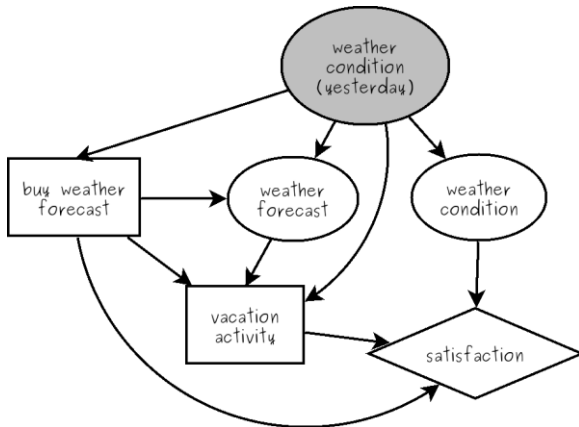


Bad Arcs for Last Minute Vacation


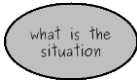




1. Weather *cannot cause* its forecast!
2. The forecast *cannot cause* the weather!
3. You're decision to go on vacation *follows in time* after you have obtained forecast.
4. The success (failure) of the vacation *follows in time* after your decision.

Last Minute Vacation with Forecast



Node Types

chance variable	known variable	decision or action	objective
			

When do we connect an arc to a node?

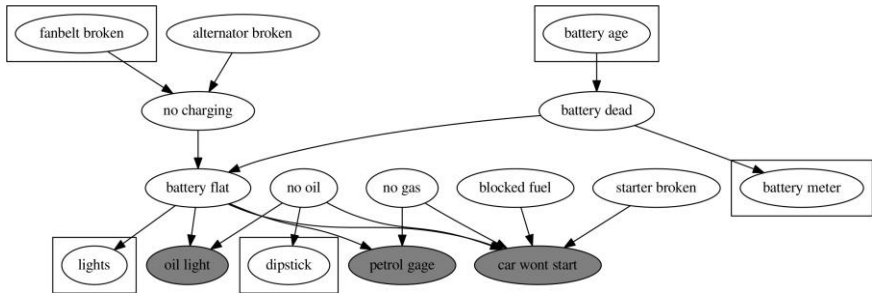
Chance variable: connect node A to chance node B if changes to the value of A can “cause” changes in B;

Known variable: same as chance node

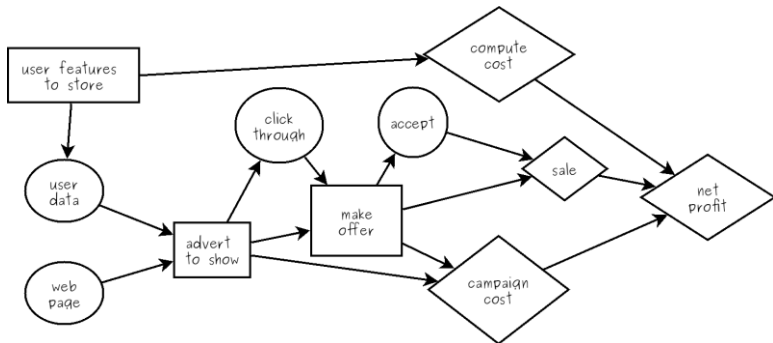
Decision: connect node A to decision node B, if variable A is used when making decision B;

Objectivity: connect node A to objective node B if variable A is used when evaluating the value of the objective (e.g. quality or cost)

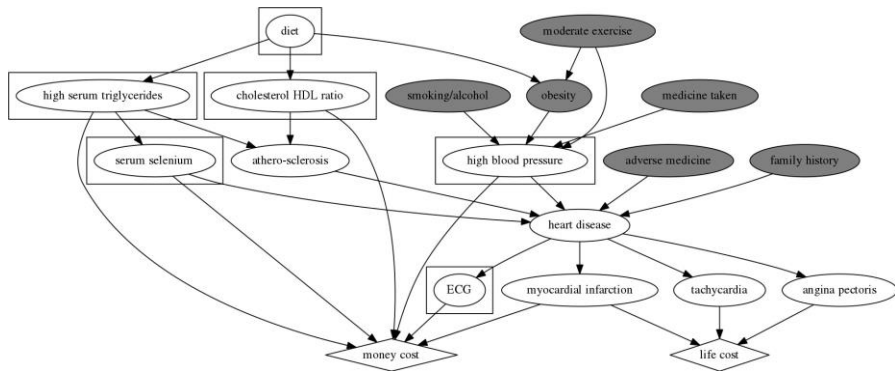
Your Car Wont Start



Internet Advertising



Heart Disease



MARS Question



An Influence Diagram:

- A. is a model giving possible situations or outcomes.
- B. consists of nodes and arcs.
- C. is an alternative to decision tree.
- D. consists of nodes and arcs and is an alternative to decision tree.

Business Models with Data (ePub section 2.3)

what kinds of businesses do we have operating
in the Data Science world?

Business Models

From Wikipedia:

A [business model](#) describes the rationale of how an organization creates, delivers, and captures value, in economic, social, cultural or other contexts.

Examples of general classes:

- ▶ retailer versus wholesaler
- ▶ luxury consumer products
- ▶ software vendor
- ▶ service provider

What kinds of businesses do we have operating in the Data Science world?

Business Models for Data Science

Many Data Science companies fit into traditional IT business models.

- ▶ software as a service (SaaS)
- ▶ consulting
- ▶ customer relationship management
- ▶ *etc.*

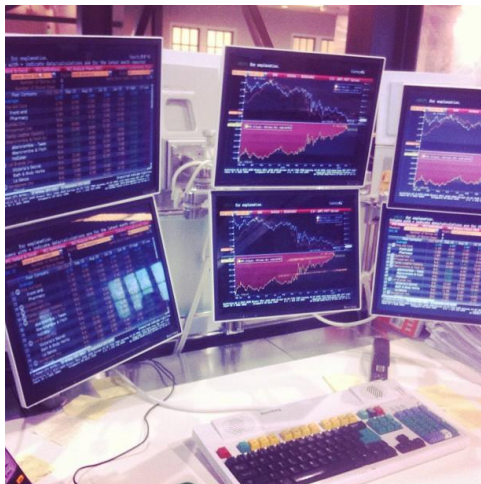
e.g. SAS is both a **software vendor** and a **consultancy**, both traditional IT business models

But there are business models somewhat unique to data-based businesses like data science.

Business Models with Data: Data business models

what are some business models specific to data science?

Bloomberg Terminal



by Jm3 [CC BY-SA 3.0](#), via Wikimedia Commons

Bloomberg Terminal (cont)

the Bloomberg Terminal:

- ▶ a computer system provided by Bloomberg L.P
- ▶ enables professionals to monitor and analyze real-time financial market data
- ▶ also place trades on the electronic trading platform
- ▶ is a proprietary secure network

Questions:

- ▶ where does the data originally come from?
- ▶ why don't users of the terminals get their data from the original source?
- ▶ why wouldn't people who sell the data to Bloomberg set up a similar service themselves?

Bloomberg Terminal (cont)

Bloomberg provides an **information brokering service**.

broker ::= a person who buys and sells goods or assets for others

Amazon.com

Black Friday Deals Are Here

New deals every day

Presented by

Amazon.com Rewards Visa

amazon.com

Hello, Michael B Corak. We have [recommendations](#) for you. (Not Michael?)

[Michael's Amazon.com](#)



[Today's Deals](#)

[Gifts & Wish Lists](#)

[Gift Cards](#)

[Your Account](#) | [Help](#)

[Shop All Departments](#)

Search



[Cart](#)



[Wish List](#)

Today's Deals

[Black Friday Deals Week](#)

[Gold Box](#)

[All Deals](#)

[Outlet](#)

[Friday Sale](#)

[Deals & Bargains](#)

[Warehouse Deals](#)

More Info

[Black Friday FAQ](#)

Black Friday Deals

Books, Movies & Music

[Books](#)

[Magazines](#)

[Movies & TV](#)

[Music](#)

[MP3 Music Downloads](#)

[Musical Instruments](#)

Electronics

[TV, Audio & Home Theater](#)

[Camera , Photo & Video](#)

[Car Electronics & GPS](#)

[Cell Phones & Accessories](#)

[Computers & Office](#)

[MP3 Players & Accessories](#)

[All Electronics](#)

Toys, Sports & Games

[Toys & Games](#)

[Sports & Outdoors](#)

[Video Games](#)

Clothing, Shoes & More

[Clothing & Accessories](#)

[Jewelry](#)

[Shoes](#)

[Watches](#)

Home

[Gourmet Food](#)

[Health & Personal Care](#)

[Kitchen, Home & Pets](#)

[Tools & Home Improvement](#)

[Automotive, Motorcycle & ATV](#)

Black Friday

Presented by [amazon.com](#) REWARDS CARD

You shouldn't have to stand in a long line to get a great deal. We're searching for the best Black Friday deals everywhere—including deals other stores are planning—so we can meet or beat their prices and bring them to you even earlier. These limited-supply offers will go quickly, but we'll add new ones throughout the day, every day this week, so you can skip the long lines and still save a bundle.

Black Friday Week Lightning Deals

Show all [Available](#) deals in category: [All Categories](#)

5:00 AM PST



Arcade Fire: "The Suburbs"

\$15.98 \$5.99 (63% off)

[Add to Cart](#) [Prime](#)

59% now claimed 00:03:32 remaining

6:00 AM PST



crocs Toddler/Little Kid Gabe Clog

\$29.95 \$15.95 (47% off)

[Select options](#)

65% now claimed 01:03:33 remaining

6:00 AM PST



Wagan Black Heated Seat Cushion

\$29.95 \$18.00 (40% off)

[Add to Cart](#) [Prime](#)

24% now claimed 03:03:33 remaining

Page 2 of 67

[Learn More](#)

Black Friday Deals in Electronics



Intro. to Data Science, (c) Wray, Bertone, 3016 Series II

Flip UltraHD 1080p

Flip UltraHD 1080p

Toshiba Satellite L655-

Gold Box™ New Deals. Every Day.

Deal of the Day



[Canon PowerShot SX210](#)

[IS 14.1 MP Digital](#)

[Camera with 14x Optical](#)

[Zoom](#)

~~\$349.99~~ **\$199.00**

Other Great Deals

- [Master Lock 22-Inch 9-Link Street Cuffs Lock](#)
- [Fashion in Pearls Jewelry: Up to 70% Off](#)
~~\$290.00~~ **\$89.00**
- [Instant Savings on Select LG HDTVs](#)
- [Nikon Projector Camera](#)
~~\$349.95~~ **\$149.00**
- [45% Off Garmin nüvi 265W/265WT 4.3-Inch Widescreen Bluetooth Portable GPS...](#)
- [Black Friday Deals on Select LG Audio and Video Products](#)
- [All Gold Box Deals](#)

Get a BlackBerry for a Penny



Through November 29, all AT&T BlackBerry phones are on sale starting at a penny with no activation fee (restrictions apply).

Amazon.com



Amazon.com



Amazon.com (cont.)



- ▶ an assembly line for the retail industry, with support for embedded online retailers
- ▶ huge stock of books, DVDs, CDs, *etc.* easily searchable
- ▶ extensive customer reviews

Amazon.com (cont.)

Information-based differentiation: satisfies customers by providing a differentiated service:

- ▶ superior information including reviews about products
- ▶ superior range

Information-based delivery network: they deliver information for others; retailers in the Amazon marketplace get:

- ▶ customers directed to them
- ▶ other retailers' support

LexisNexis

See [LexisNexis](#), provides world's largest electronic database for legal and public-records related information.

Information provider: business selling the data it collects

- ▶ like a traditional business model, selling data not widgets
- ▶ fast growing segment of the IT industry post 2000 (cited by Evan Quinn's blog post on [Infochimps.com](#) April 2013 "Is Big Data the Tail Wagging the Data Economy Dog?", now offline)
- ▶ some call this the [data economy](#)

e.g. [data brokers sell consumer data](#) to major retailers or internet companies

Data Business Models

information brokering service: buys and sells data/information for others.

information-based differentiation: satisfies customers by providing a differentiated service built on the data/information.

information-based delivery network: deliver data/information for others.

information provider: business selling the data/information it collects.

“What a Big-Data Business Model Looks Like” by Ray Wang in the Harvard Business Review claims these are unique in the data world.

Business Models with Data: Intelligent systems business models

more generally, what are some intelligent
systems business models?

Intelligent Systems Bus.Models

From [Machine Intelligence in the real world](#) by Shivon Zilos, 26th Nov. 2015.

Business Models used by Intelligent Systems companies:

Data providers collect dataset (e.g. satellite imagery, vineyard irrigation)

Alchemists promise to turn your data into gold (e.g. self-service APIs)

Gateways create new use cases from specific data types
(e.g. image, audio, video, genomic data)

Magic Wands fix a workflow using software as a service (SaaS) tools
(e.g. help recruiters write better job descriptions)

Navigators autonomous systems for the physical world
(e.g. self-driving cars)

Agents create cyborgs and bots to help with virtual tasks
(e.g. customer service realtime chat)

A brief Introduction to Python for Data Science

MARS Question



What is .ipynb?

- A. An illegal file extension.
- B. Interactive Python NoteBook.
- C. Intelligent Python Nota Bene.
- D. Typo, it should be 'pinyin'

MARS Question



What is a dataframe?

- A. An array.
- B. A list.
- C. A theory about data.
- D. A structure that stores tabular data

Next: Module 2

Data Types and Storage