FIT1043 Introduction to Data Science

Module 2:
Data Models in Organisations
Lecture 3: Data Business Models

Monash University

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!

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”

Case Study: City Science

City X

- new research group under the Institute of Transport Studies at Monash University
- 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"*

Unit Schedule: Modules

Module	Week	Content
1.	1	overview and look at projects
	2	(job) roles, and the impact
2.	**3**	data business models
	4	**application areas and case studies**
3.	5	characterising data and "big" data
	6	data sources and case studies
4.	7	resources and standards
	8	resources case studies
5.	9	data analysis theory
	10	data analysis process
6.	11	issues in data management
	12	data management frameworks

Data and Decision Models (ePub section 2.1)

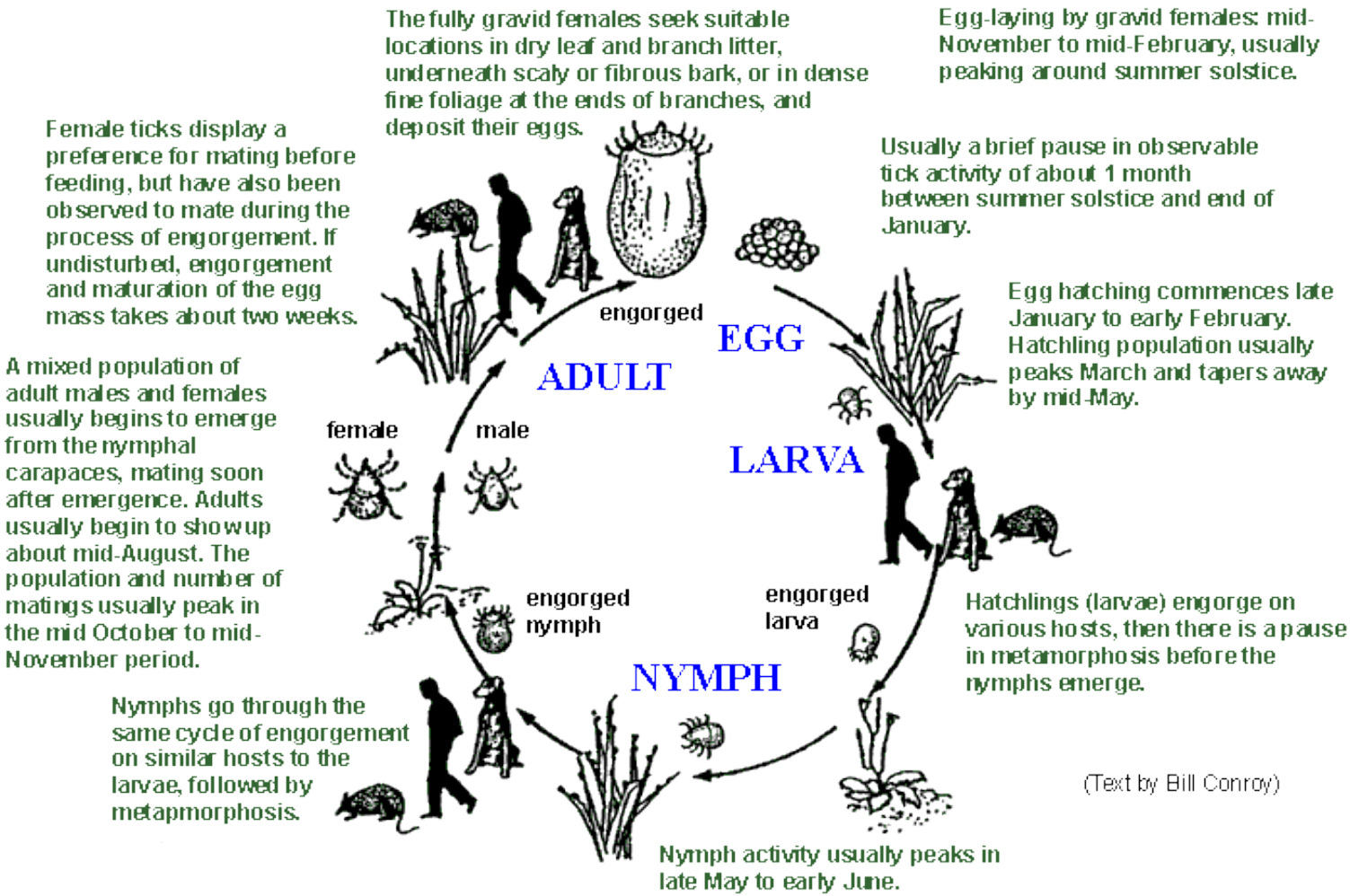
some general models for understanding businesses and decision making:

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

Data and Decision Models: Life cycles and value chains

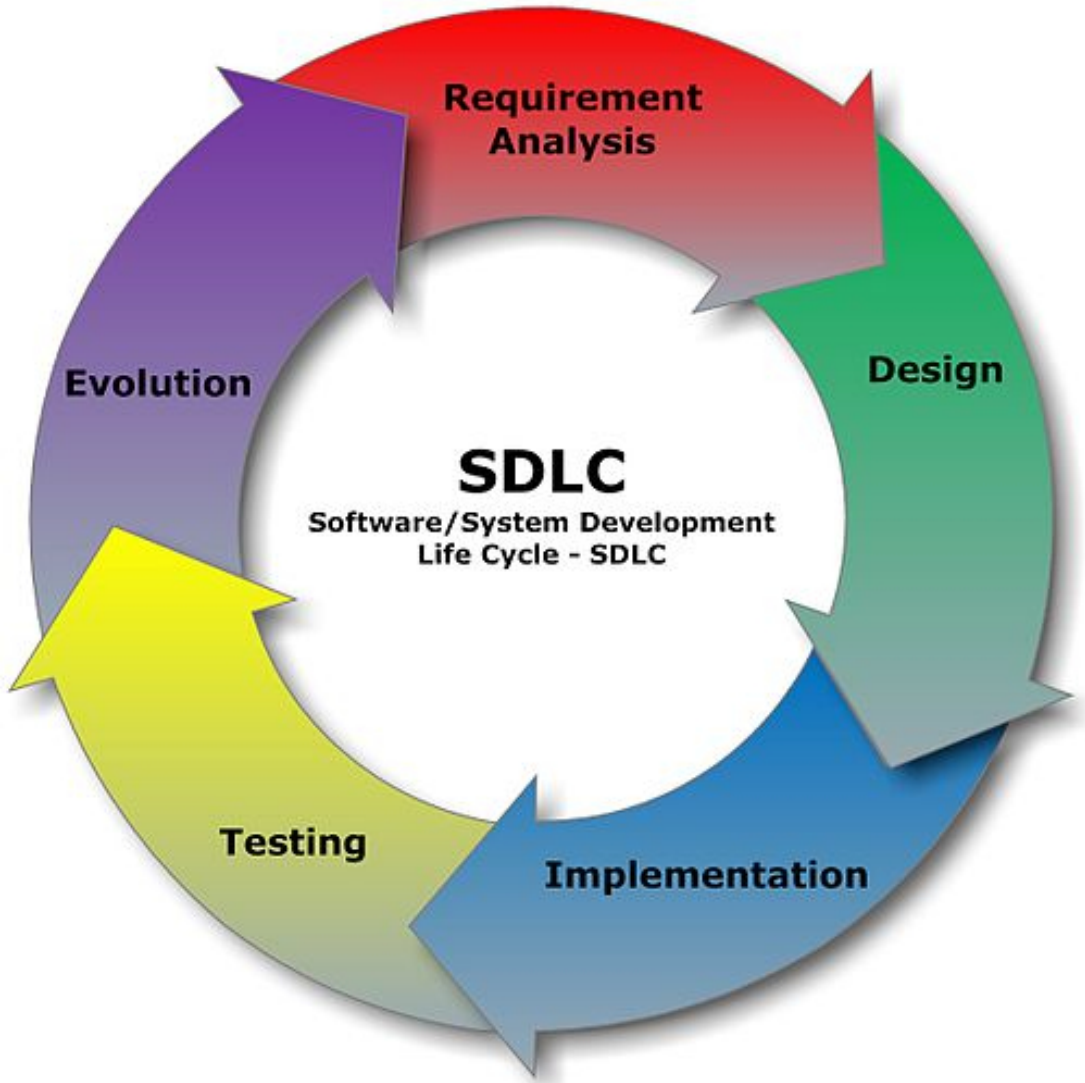
sequence of processes done to create value

A Biological Life Cycle



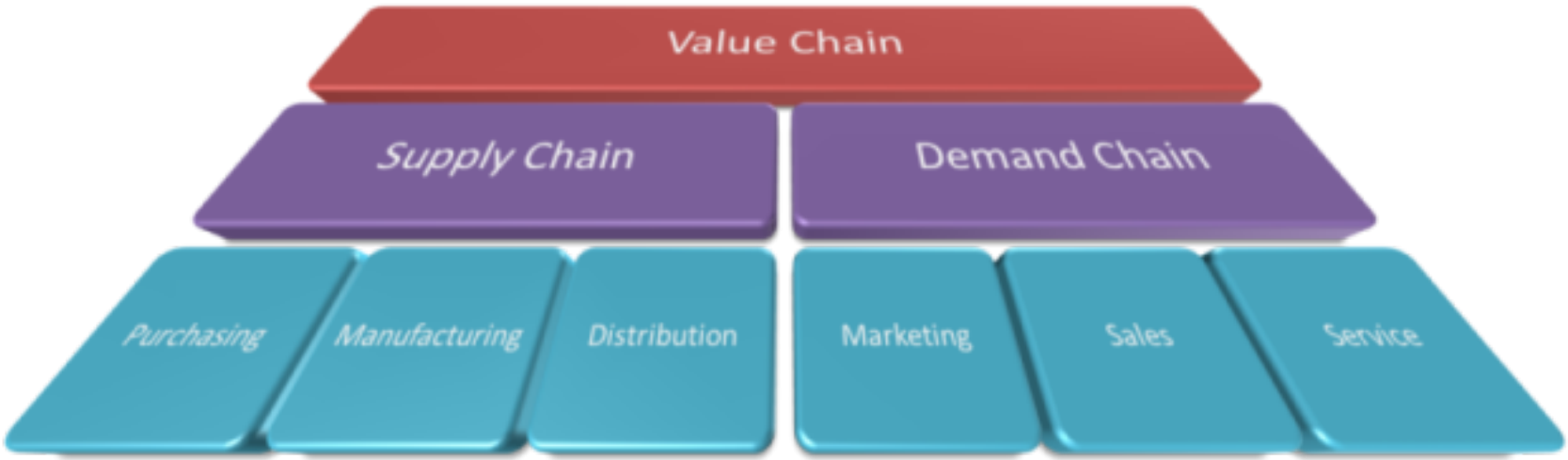
by Commonsource (Own work) *CC BY 3.0*, via Wikimedia Commons

A Software Life Cycle



by Cliffydcw (Own work) *CC BY-SA 3.0*, via Wikimedia Commons

The Retail Value Chain



by diduncculus99 (My artwork) *CC0 1.0*, via Wikimedia Commons

The Engineering Value Chain

by Yufeng Zhang (Own work) *CC BY-SA 3.0*, via Wikimedia Commons

Life Cycles versus Value Chains

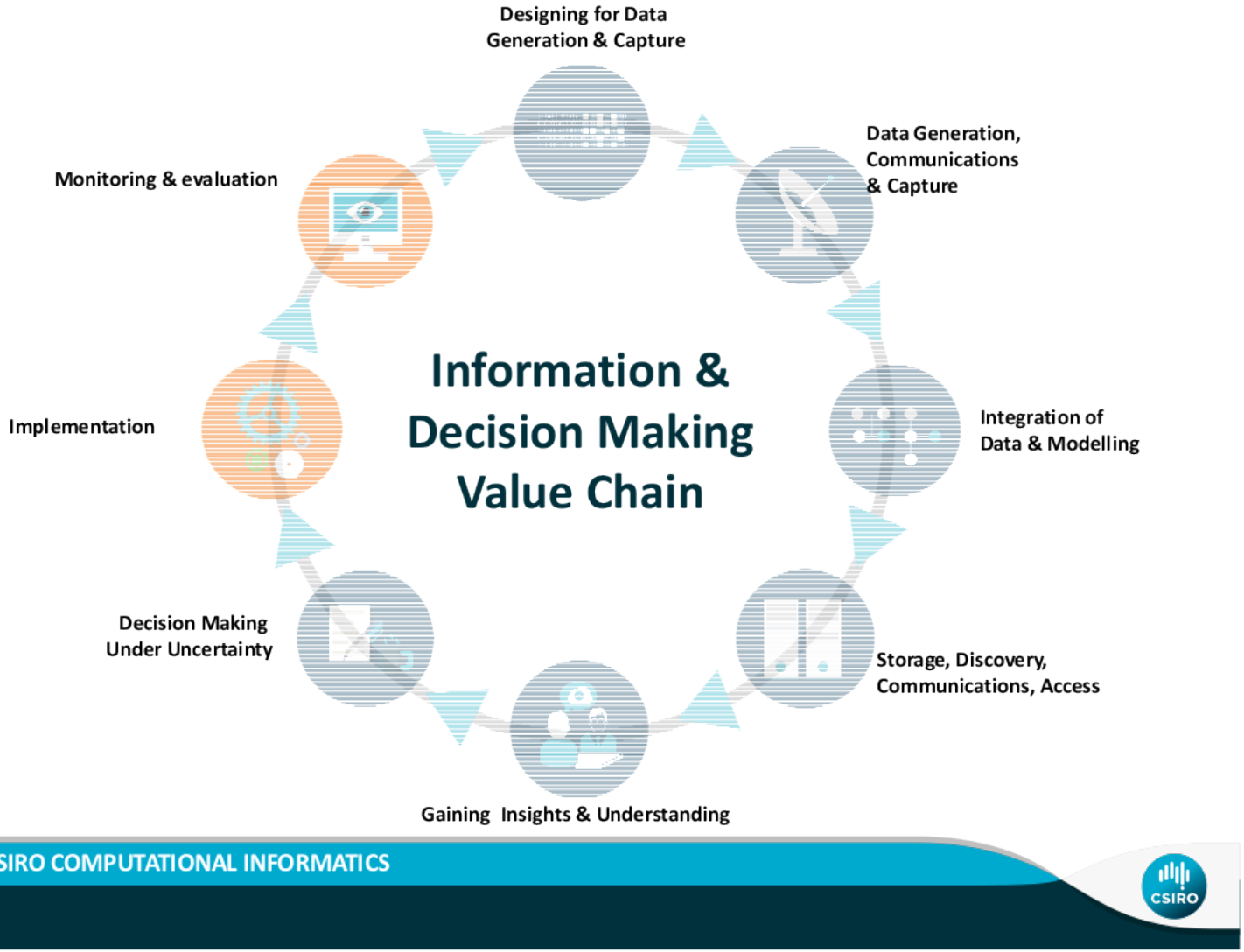
Life Cycle:

- used in biology to model the cycle of life:
  - from birth / conception / ideation
  - through to death / completion / retirement
- term has been adapted for use in computer science & business:
  - product lifecycle, software development lifecycle, ...

Value Chain:

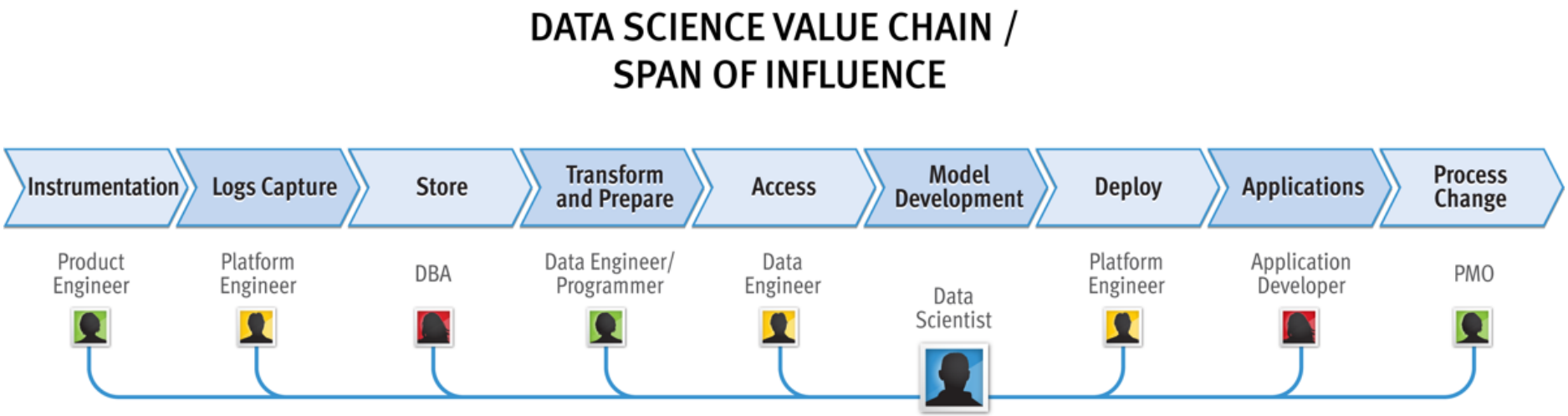
- business term used to describe the series of activities done to create an item of value

CSIRO's Information Value Chain



From Iain Collings, CSIRO, 2014

Pivotal's Data Value Chain



slide 12 in their *"Data Science + Data Engineering"*

Miller and Mork's Data Value Chain

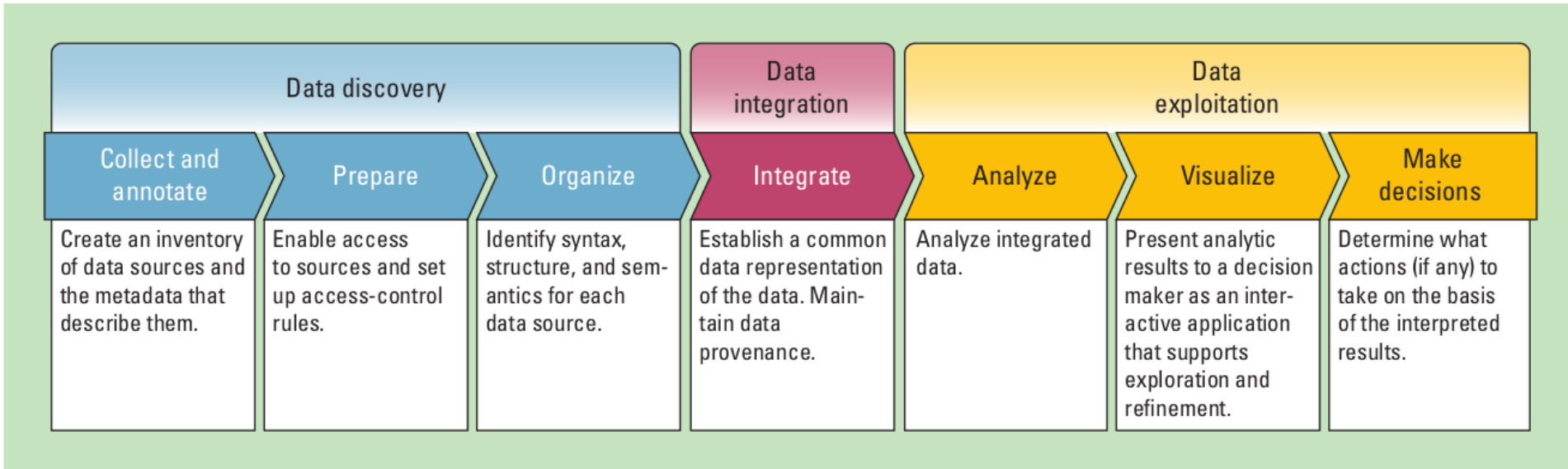
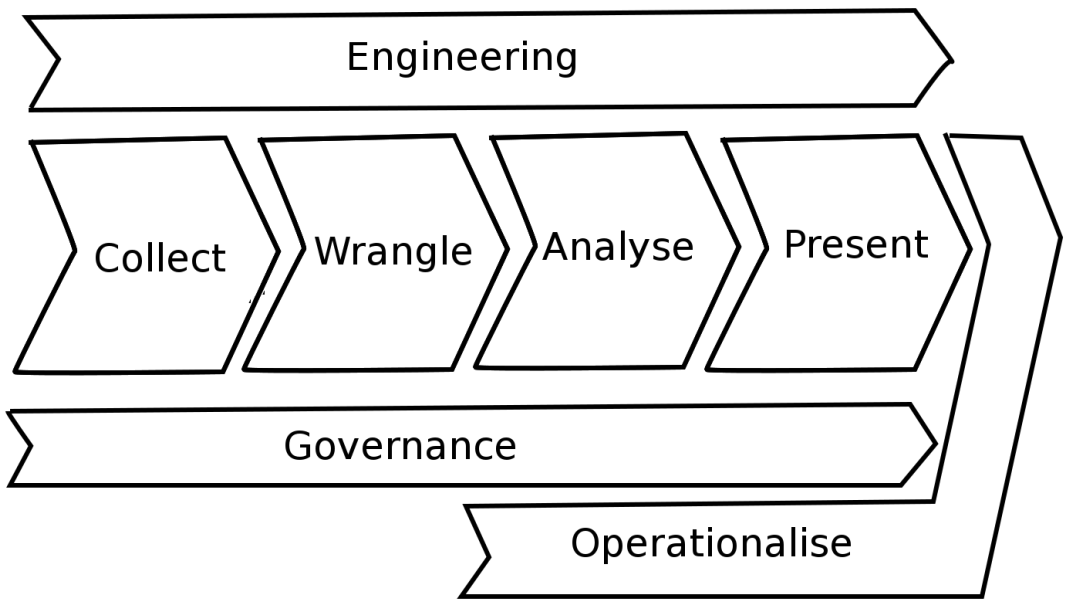


Figure 1. The data value chain. The chain provides a framework with which to examine how to bring disparate data together in an organized fashion and create valuable information that can inform decision making at the enterprise level.

see *"From Data to Decisions: A Value Chain for Big Data"*

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 different kinds of analysis

SAS Analytic Levels

See *"Eight Levels of Analytics"*, a SAS report from 2008.

Roughly correspond to the standard levels of statistical analysis.

Analytic Levels

Descriptive analytics: gain insight from historical data

- e.g. plot sales results by region and product category
- correlate with advertising revenue per region

Predictive analytics: make predictions using statistical and machine learning techniques

- e.g. predict next quarter's sales results using economic projections and advertising targets

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

- e.g. recommend which regions to advertise in given a fixed budget

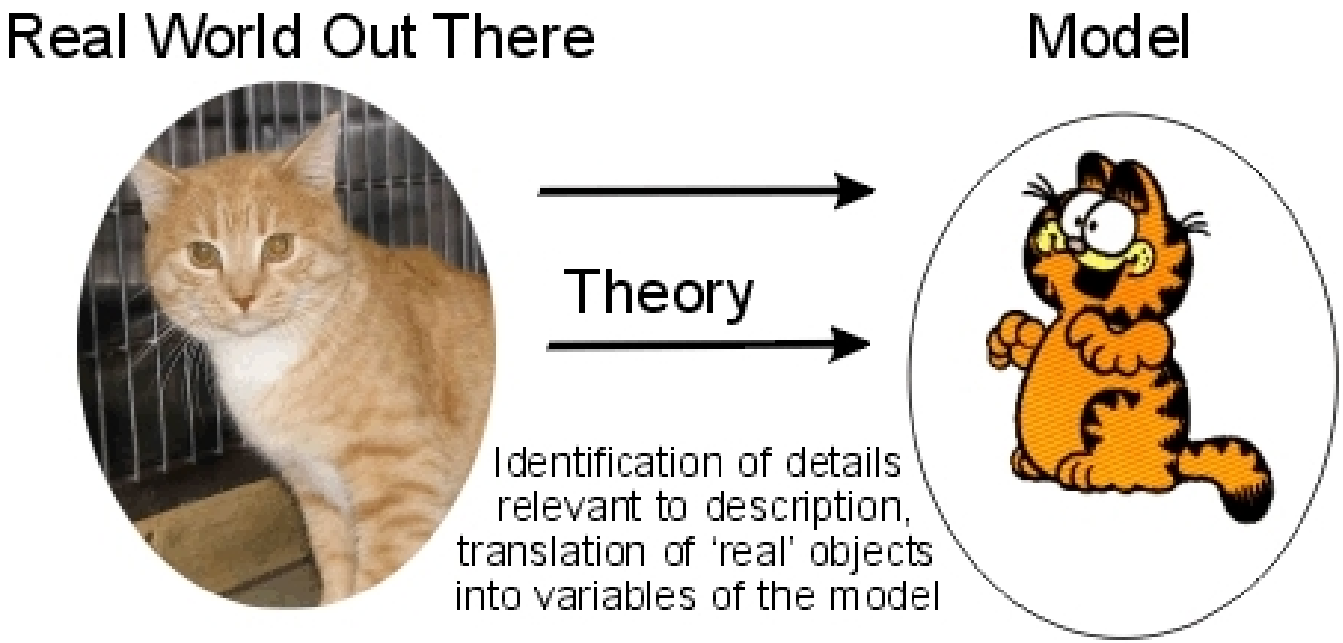
primarily a descriptive classification for general discussions

Data and Decision Models: Influence diagrams

method for modelling decision making

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

Modelling

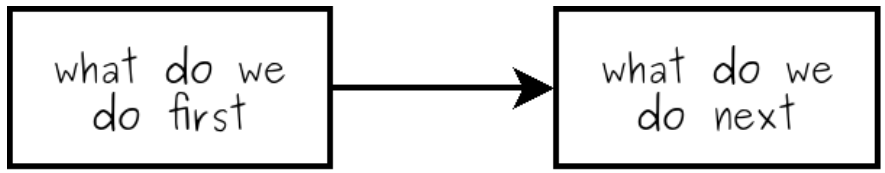


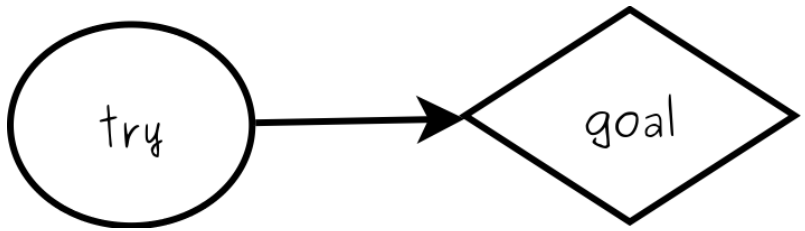
from *the BackReaction blog by Sabine Hossenfelder*

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





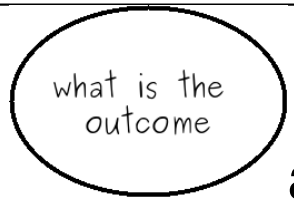
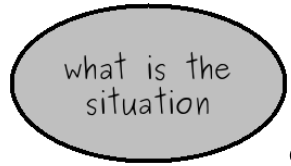
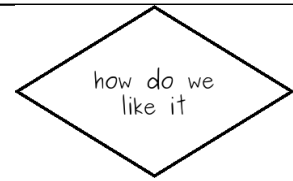
Motivating Influence Diagrams

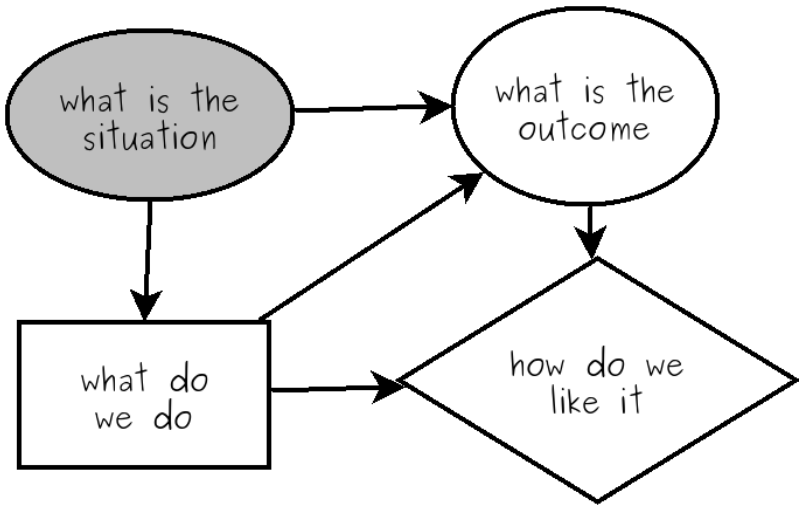
Influence Diagrams (a.k.a Decision Graphs) are

- 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

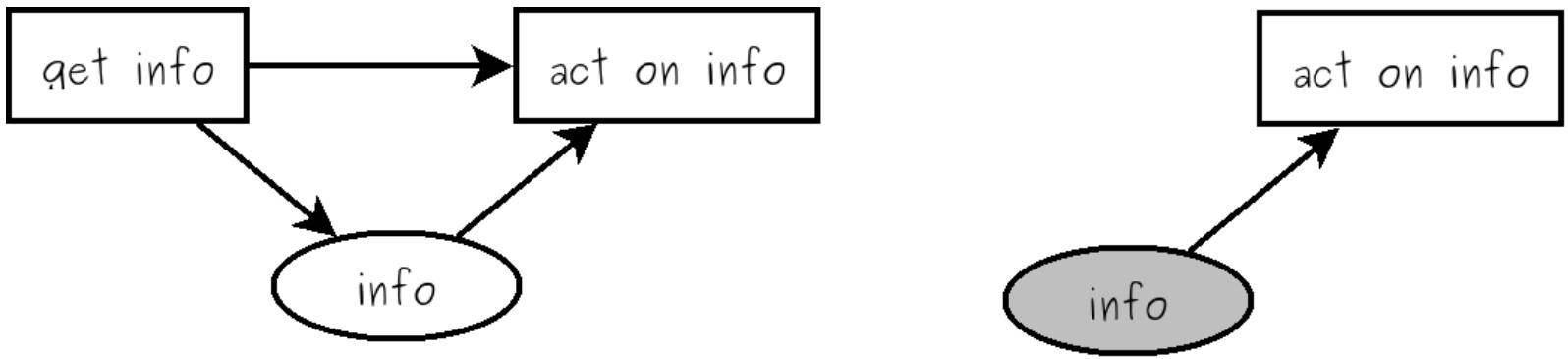
Influence Diagrams

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

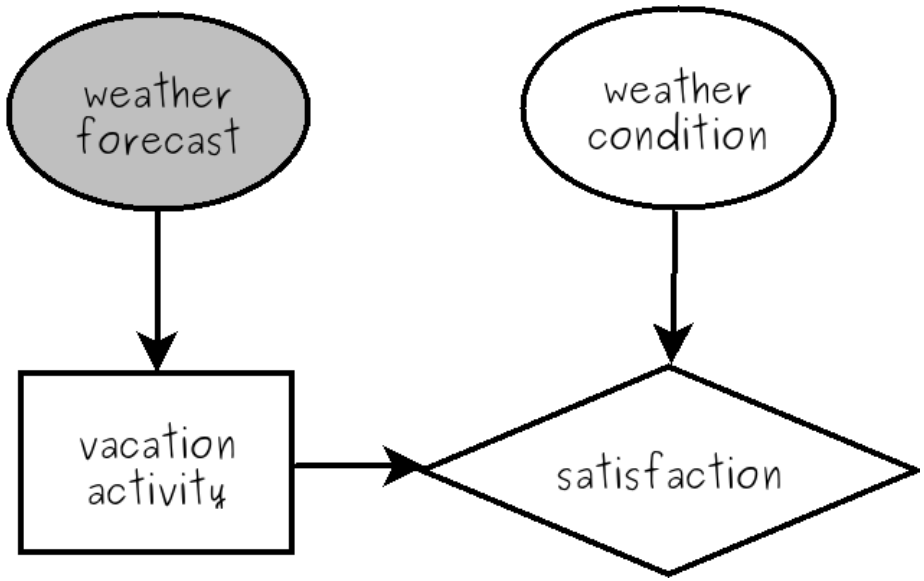


nodes are strung together in a directed graph to convey influence

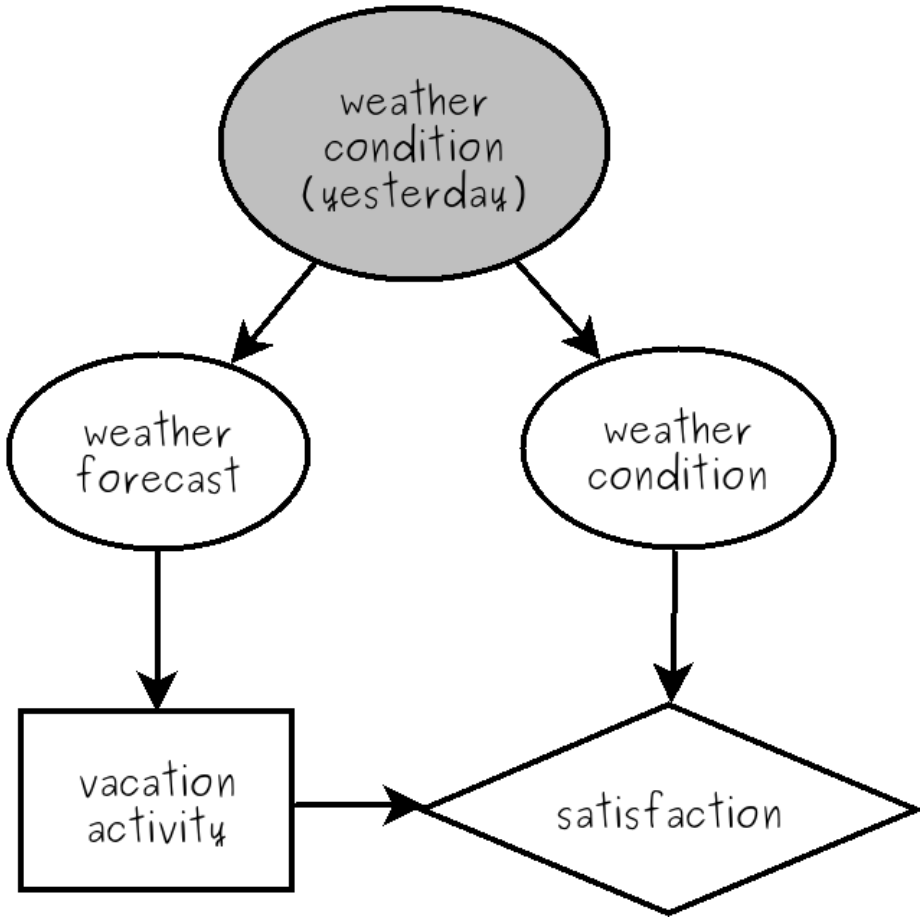
Influence Diagrams – Arcs



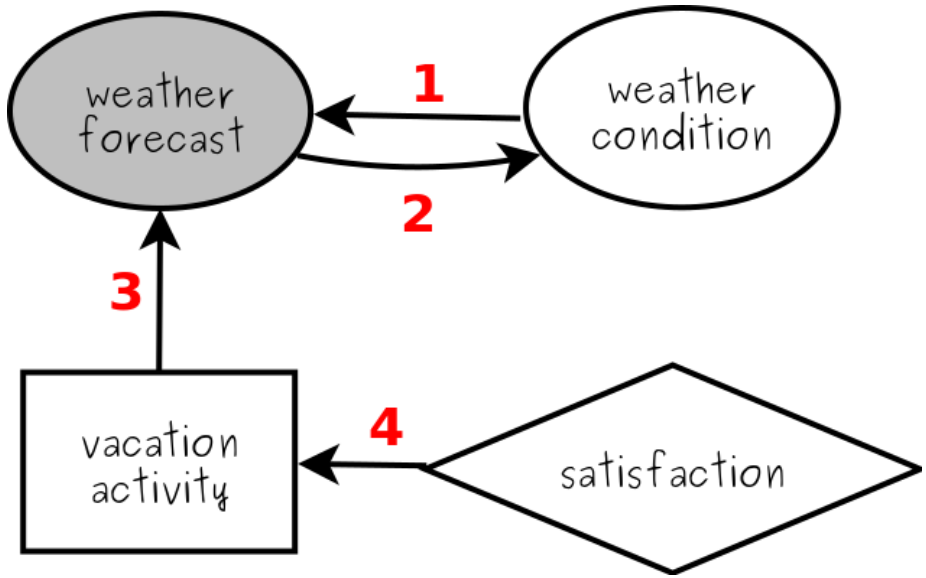
Last Minute Vacation



Last Minute Vacation (cont)

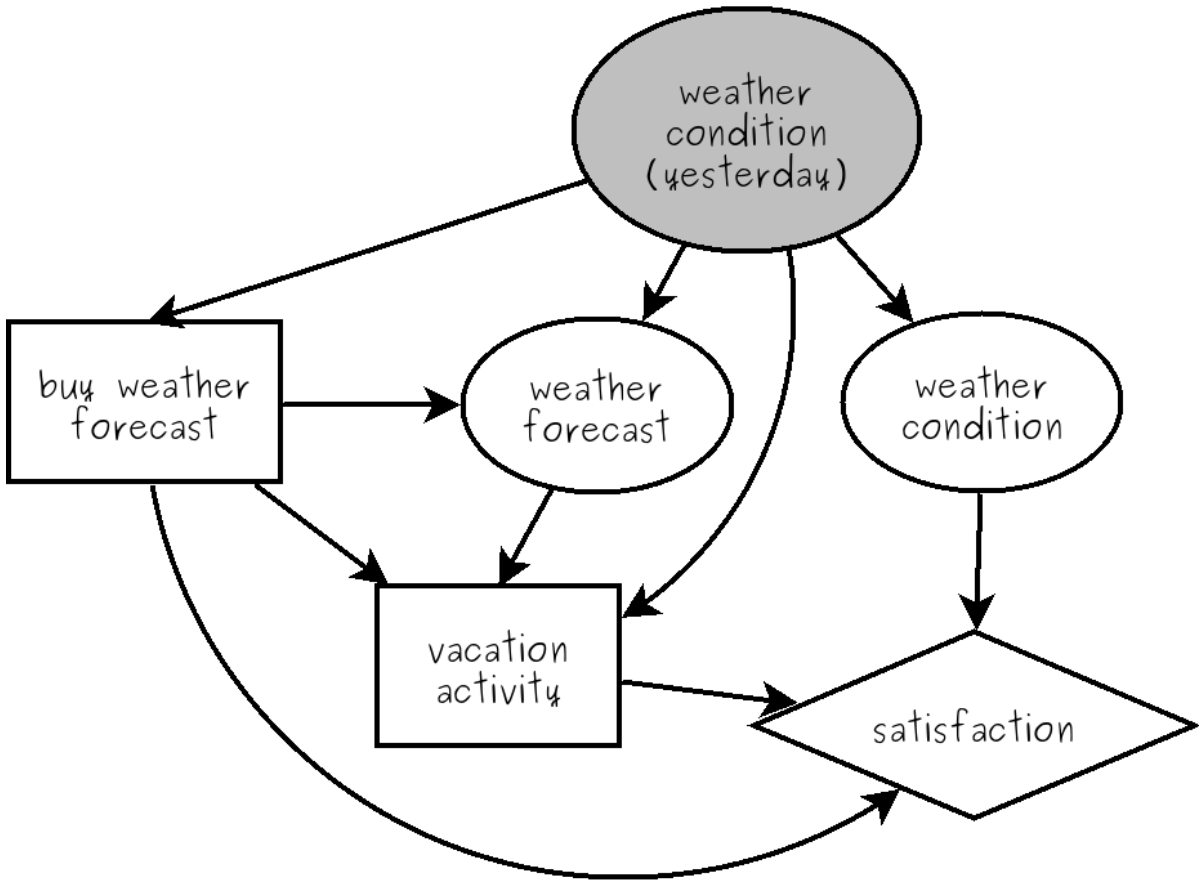


Bad Arcs for Last Minute Vacation

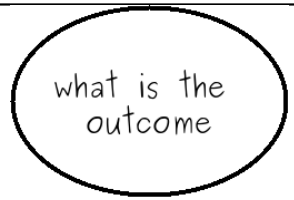
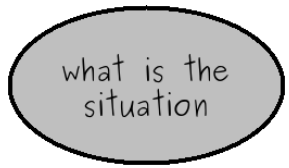
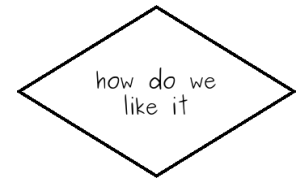


1. Weather *cannot cause* its forecast!
2. The forecast *cannot cause* the weather!
3. You're decision to go on vacation *follows in time* after you have obtained forecast.
4. The success (failure) of the vacation *follows in time* after your decision.

Last Minute Vacation with Forecast



Node Types

chance variable	known variable	decision or action	objective
			

When do we connect an arc to a node?

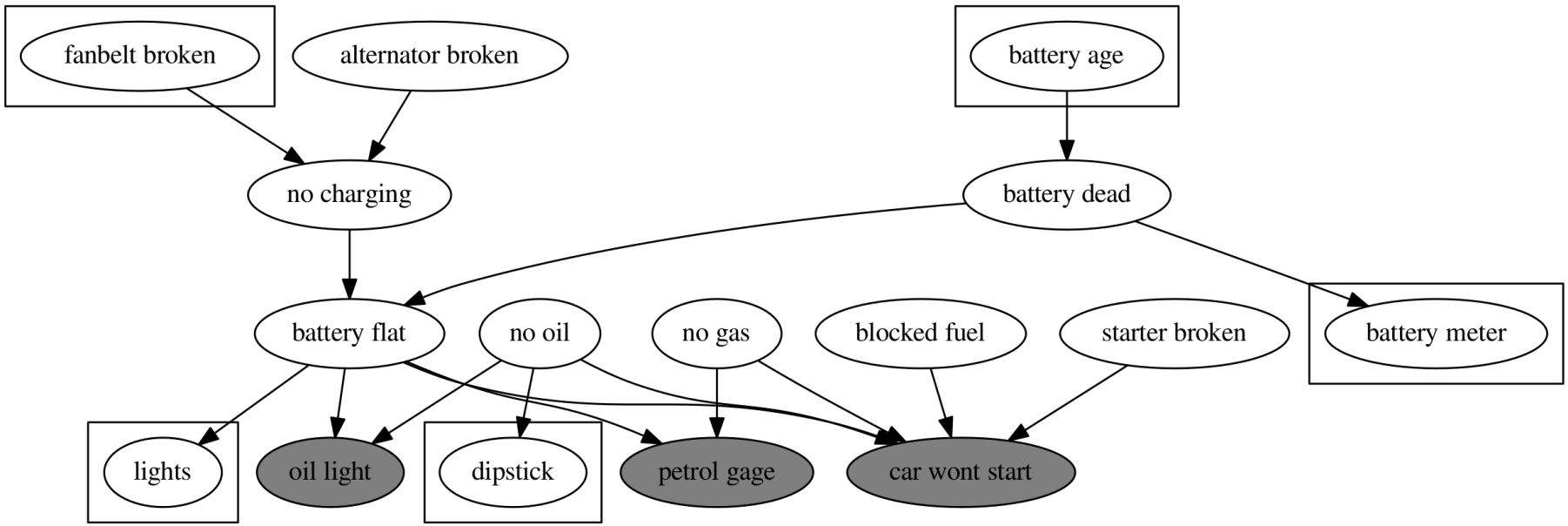
Chance variable: connect node A to chance node B if changes to the value of A can "cause" changes in B;

Known variable: same as chance node

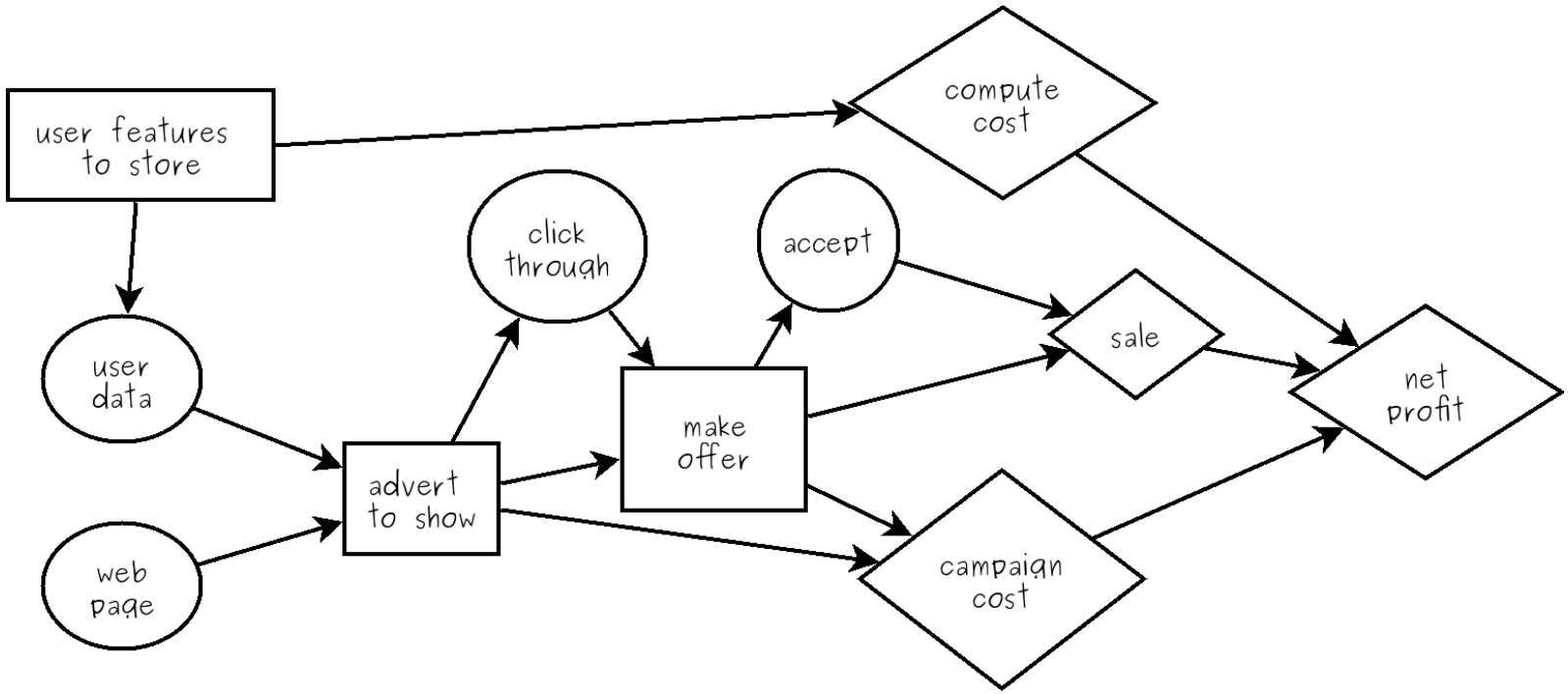
Decision: connect node A to decision node B, if variable A is used when making decision B;

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

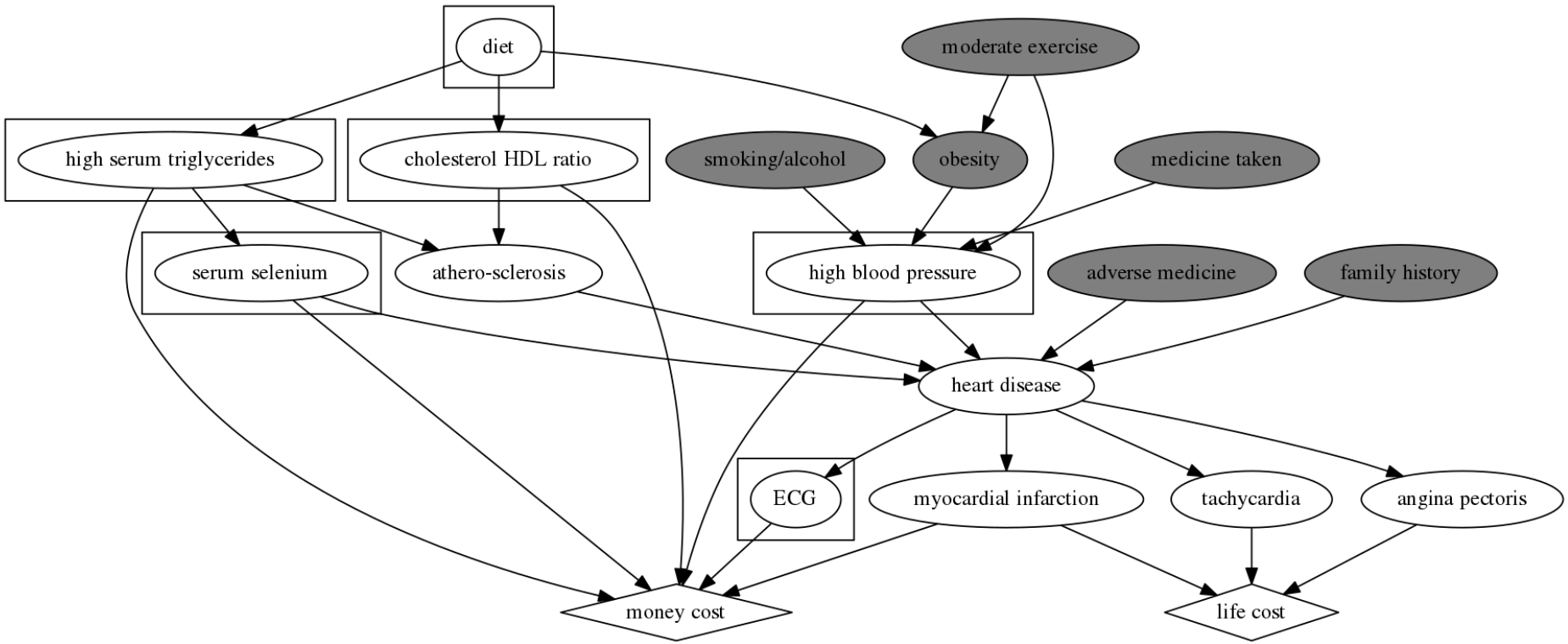
Your Car Wont Start



Internet Advertising



Heart Disease



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?

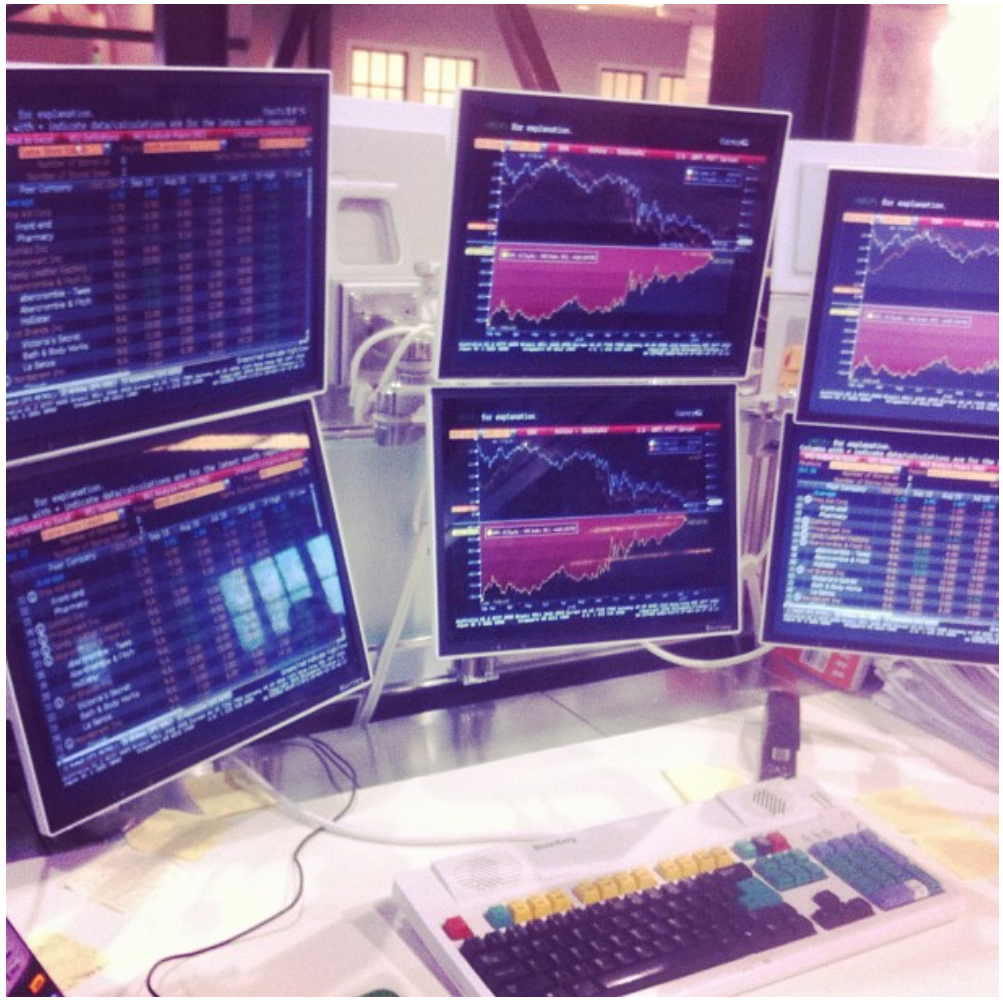
Business Models with Data

- data business models:
  - distinct business models based around data
- intelligent systems business models:
  - distinct business models for “smarter” systems
- big data value chain:
  - big data involves the broader field, so lets look at this
- big data landscape:
  - mapping the ecosystem of tools, services, *etc.*

Business Models with Data:
Data business models

what are some data-based businesses?

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:

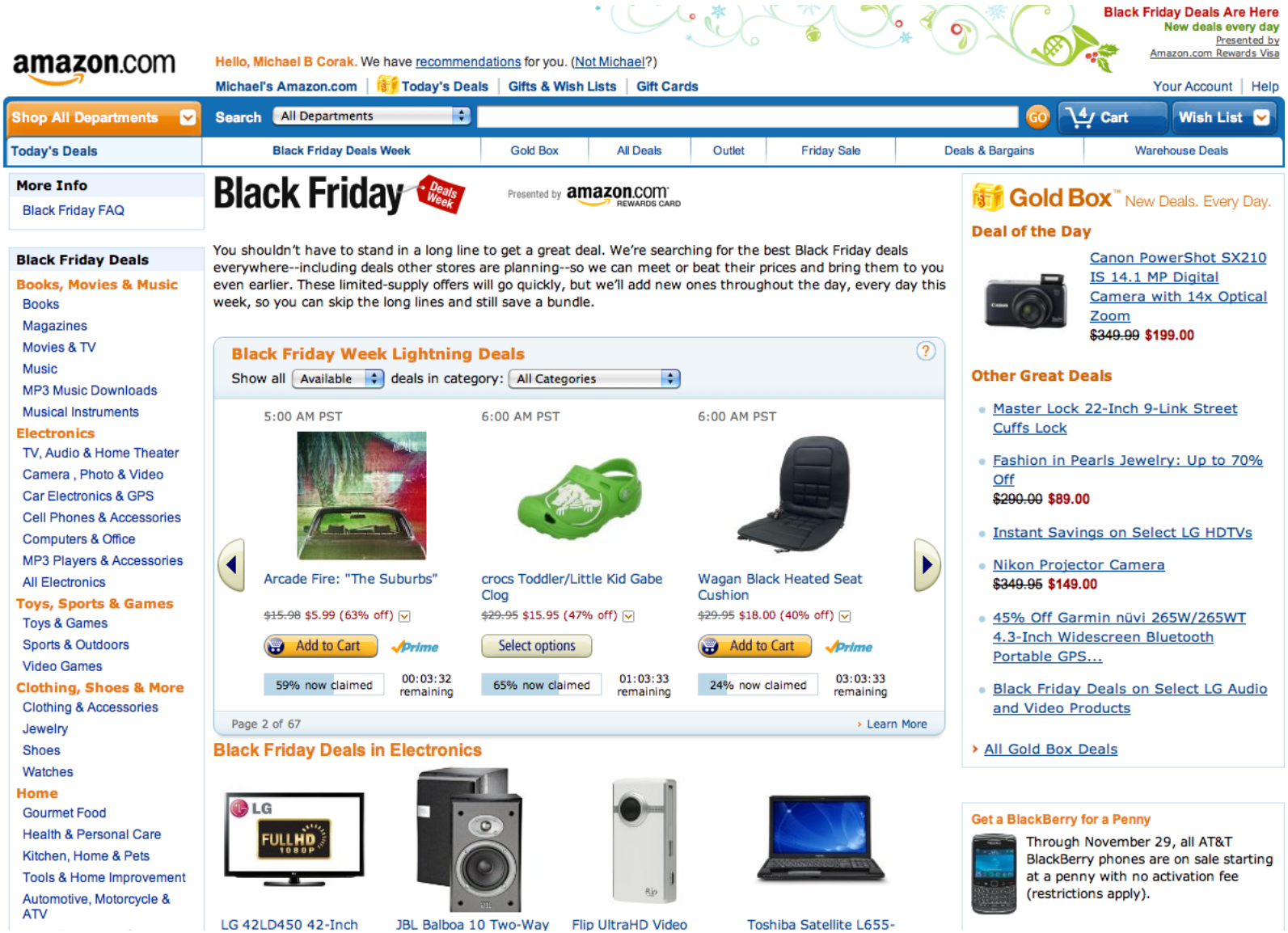
- where does the data originally come from?
- why don't users of the terminals get their data from the original source?
- why wouldn't people who sell the data to Bloomberg set up a similar service themselves?

Bloomberg Terminal (cont)

Bloomberg provides an information brokering service.

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

Amazon.com


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
Black Friday Deals Are Here
New deals every day
Presented by Amazon.com Rewards Visa
Shop All Departments
Search
All Departments
GO
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
Deals Week
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
LG 42LD450 42-Inch
JBL Balboa 10 Two-Way
Flip UltraHD Video
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


P-1 B52 500-670
P-1 B53 500-676
P-1 B54 500-676
P-1 B55 500-676
P-1 B56 500-676
P-1 B57 500-676
P-1 B58 500-676
P-1 B59 500-676
P-1 B60 500-676
P-1 B61 500-676
P-1 B62 500-670
P-1 B63 500-670
P-1 B64 500-670
P-1 B65
BODMAN 91

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

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.

WARNING: many Data Science companies pursue other business models, software as a service, consulting, CRM, *etc.*

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

Data Providers?

data provider ::= business selling the "data" it collects,
e.g., *LexisNexis*

- this is a traditional business model, selling data not widgets
- so does not fit into Wang's categories (though is borderline "data broker")
- fastest growing segment of the IT industry post 2000 (see Evan Quinn's blog post on `Infochimps.com` April 2013 *"Is Big Data the Tail Wagging the Data Economy Dog?"*)
- some call this the data economy

Business Models with Data: Intelligent systems business models

what are some strategies to build smarter systems?

Intelligent System Bus. Models

From *Machine Intelligence In The Real World* by Shivon Zilos, Nov. 2015.

Business Models used by Intelligent Systems companies:

Panopticons	collect a broad dataset (e.g. satellite imagery)
Lasers	collect a focused dataset (e.g. 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)
Pioneers	the clever innovators (e.g. deep learning researchers)

Business Models with Data:
Big data value chain

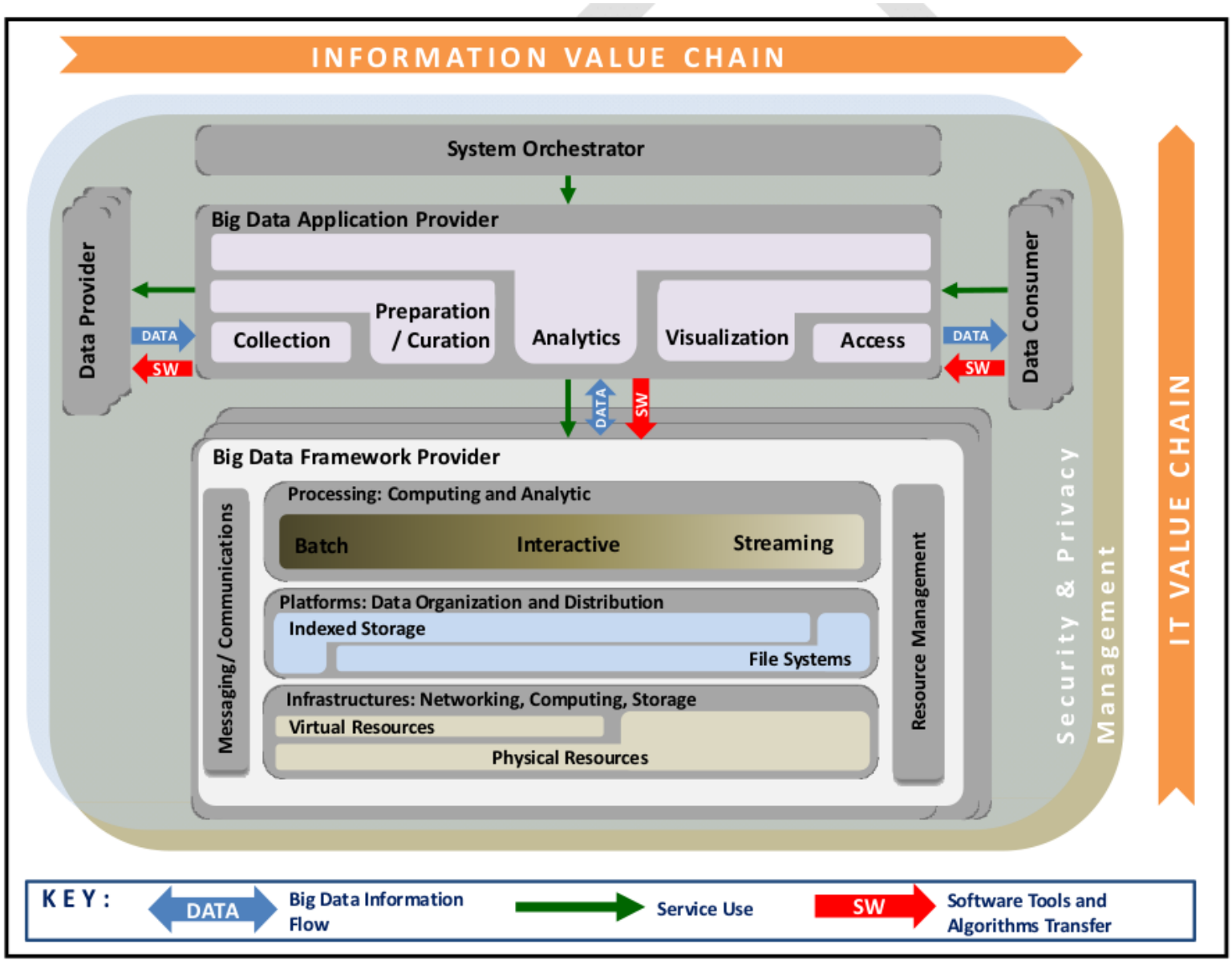
looking at the full big data space (not just Data Science)

Big Data Value Chain

The next slide shows

- the NIST Big Data Reference Architecture (from *"Volume 6: Reference Architecture"*)
- with our Standard (Data Science) value chain embedded at the top as a "big data application provider"

Value Chains



Business Models with Data:
Big data landscape

mapping the ecosystem of tools, services, *etc.*

The Big Data Landscape

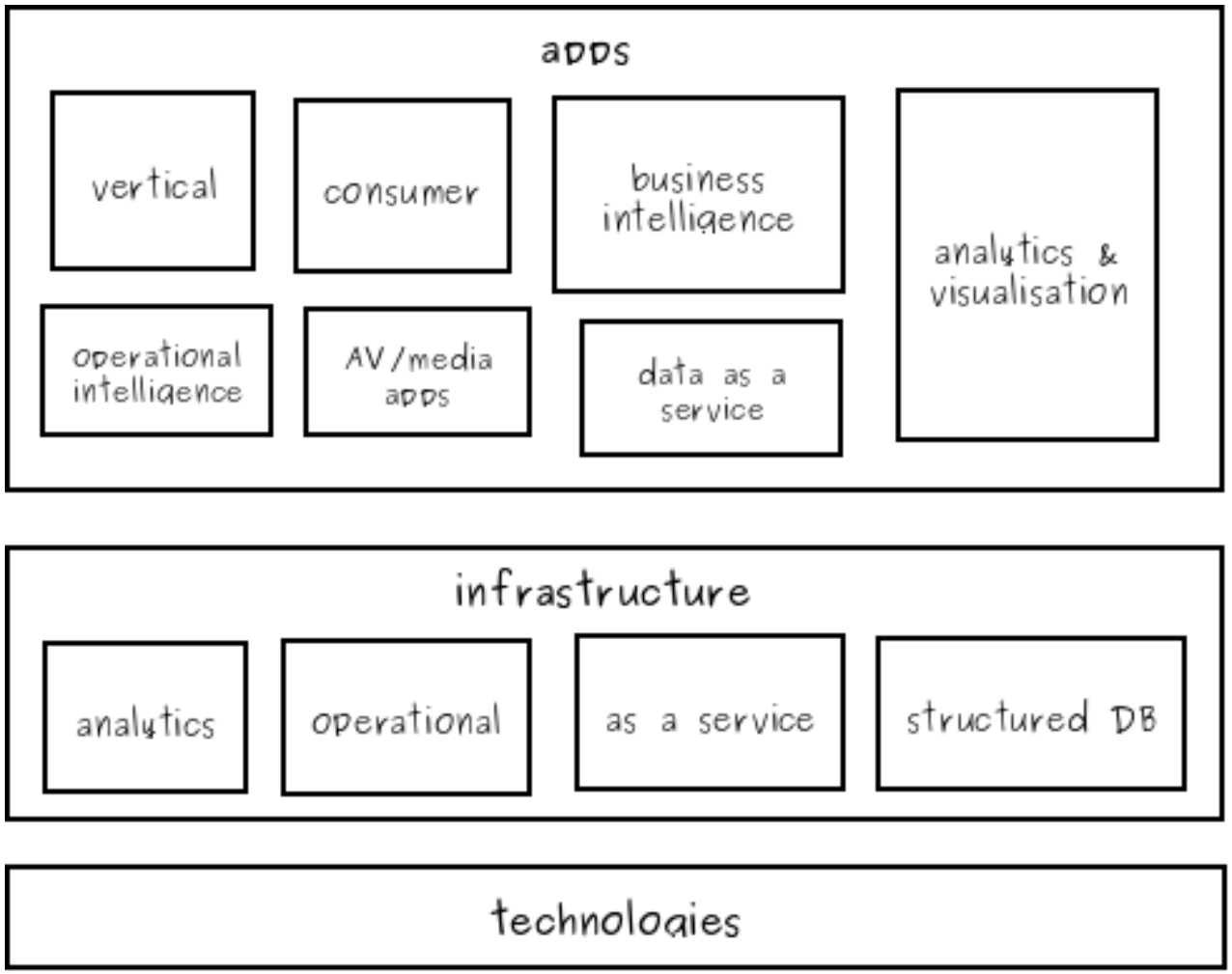
Landscapes categorise companies in the data science area:

- *Big Data Landscape* (2012) by Dave Feinleib on a Forbes blog
- *Big Data Landscape 2016* by Matt Turck of FirstMark Capital

WARNING: don't get lost in the detail of these, just briefly review and look at the major headings.

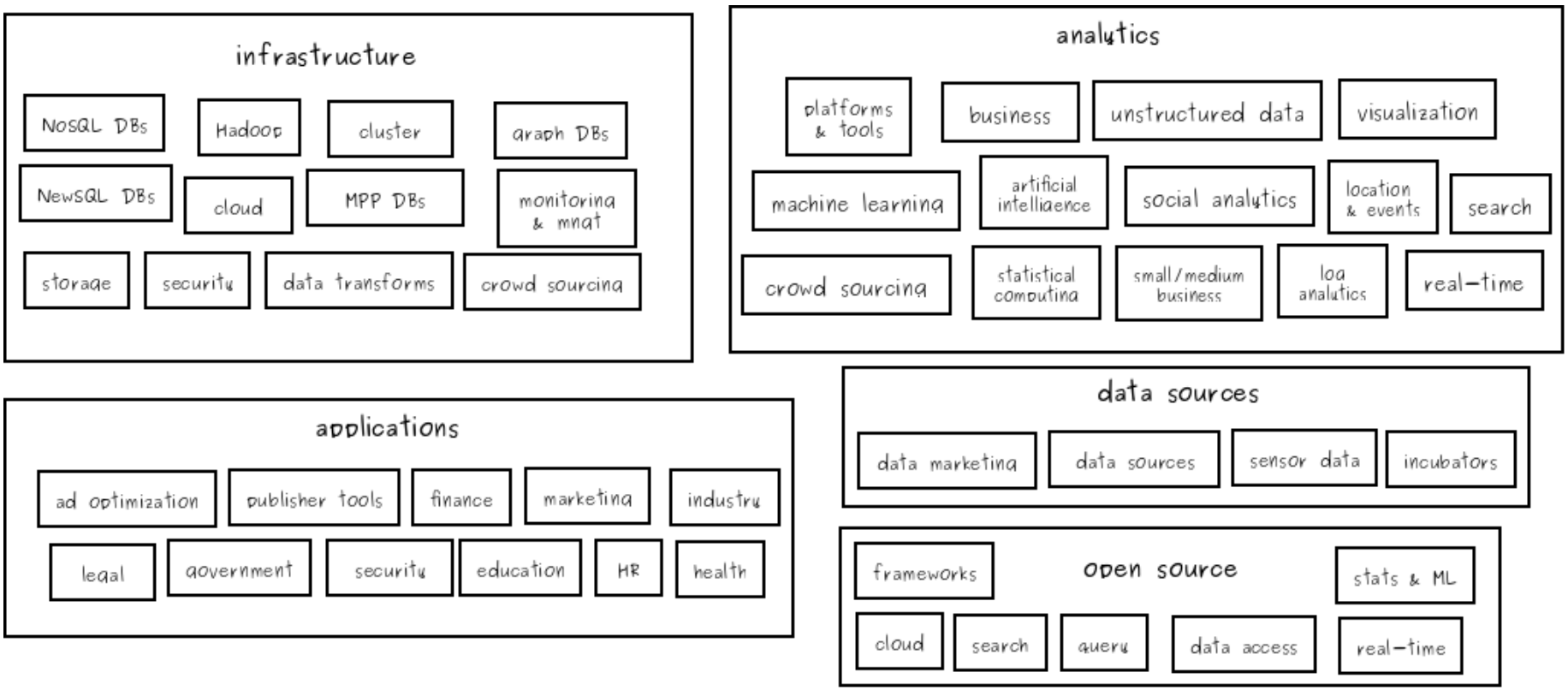
Feinleib's Big Data Landscape

Categorisations from the *Big Data Landscape*



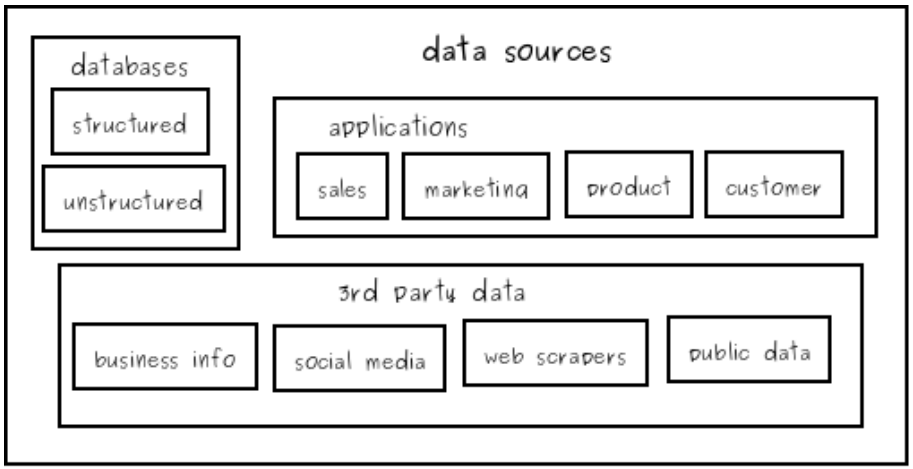
Turck's Big Data Landscape

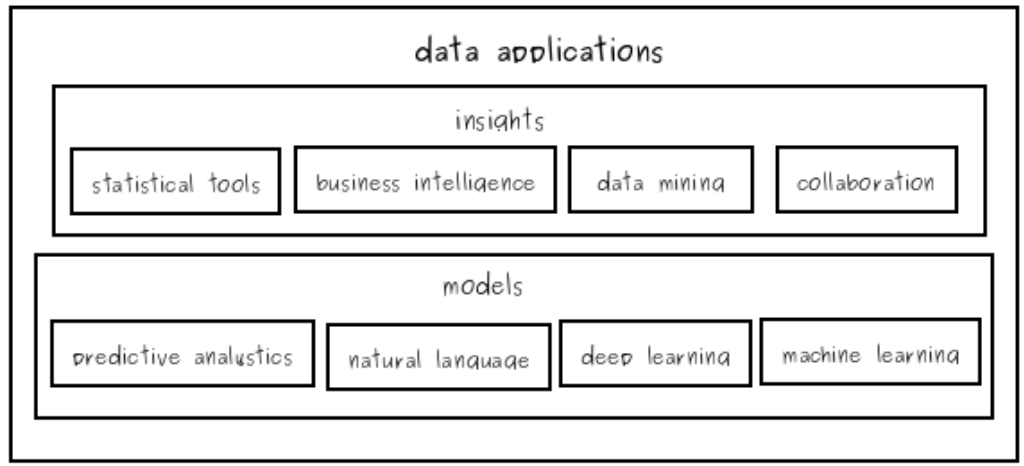
Categorisations from the *Big Data Landscape 2016*

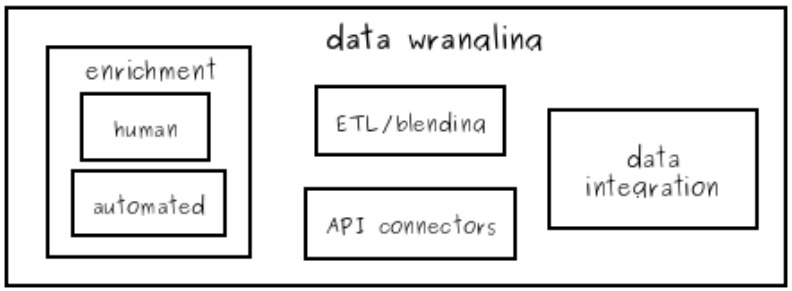


CrowdFlower Ecosystem

Categorisations from *The Data Science Ecosystem*







Next Week: Applications Areas (ePub section 2.5)

Homework:

- read *"The Mayor's Geek Squad"* from New York Times
- view the infographic: *"How People Spend Their Time Online"* by GO-Gulf
- read the article *"Data Science and its Relationship to Big Data and Data-Driven Decision Making,"* by Foster Provost and Tom Fawcett.
  Big Data. March 2013, 1(1): 51-59 (8 page PDF).