

Big Data, Analytics & Data Visualization –

Dispelling the Myths and Discovering New Opportunities

Professor Olivera Marjanovic, PhD
Leader of the RG "Analytics in Human Complex Systems"
Women in Engineering and IT (WiEIT) Champion

School of Professional Practice & Leadership

Faculty of Engineering and IT

University of Technology Sydney

olivera.marjanovic@uts.edu.au

Presentation Outline

- Fundamental concepts: Big data, Business Intelligence & Analytics
- Making sense of the current Big data confusion and hype
- Three different perspectives (business, data science & IT) of BI&A
- Emerging trends Data Visualization and Visual Analytics
- Our current research: Visual Atlas of Australian Cooperatives
- Beyond Business Analytics: Analytics in Human Complex Systems

Digital platforms

NFP Sector

Community Organisations

> Cooperatives and Mutuals

R&D labs

Smart Cities

Energy Sector

Sport Management

Government

agencies Environmental

Protection Agency

Cyber

security Aged-care Civil/Elec/SW

Engineering

Banking and

finance

Healthcare

(IoT) Tech start-up

Disaster Recovery

(yet to be

invented)

Humanitarian

Aid Agencies

Education

Transportation

Agribusiness

(Clean Supply-Productive Aging

Marine chains) biology

Wildlife

IT consulting

Conservation

Fashion Industry

Media industry

Societal context

Industry context

ORG CONTEXT

(Hardware and Software)

Data

(management, modeling, quality, visualization, ethics)

Processes

(Business and Other)

People

Human and Social capital (knowledge/ skills/creativity/ethics other stakeholders Change/humanity)

Services

(to customers/patients/ citizens/constituents/

Big Data Phenomenon



Jeroen Jansse

Big Data Landscape Infrastructure Analytics Applications Data Science Human Analyst Customer Service Hadoop in Analytics Sales & Marketing Hadoop Cluster Services Spark Visualization Legal Platforms Capital Platforms the Cloud Platforms RADIUS Gainsight On-Premise MEDALLIA 🙀 + a b | e a u context relevant ğıld. databricks Q Palantir RAVEL **⊜**bloomreach Zeta Microsoft CONTINUUM: R DataRobot ATTENJITY 🧠 kubernete Hortonworks Google Coud Nations. livefyre 🌽 CLARABRIDGE JUDICATA MAPR Pivotal AYASDI quavus Roambi GridGain 4 → docker IBM InfoSphere STELLAService Lattice MODE plotly ADATAD Økahuna Quid' enigma NG#DATA Preact ZOOMDATA. AZENA THEASURE Everlaw IBM InfoSphere SAILTHRU textic #dataiku @tonian Datameer persado "Infer **ósense** TACHYON Digital Researcing Core OS pepperdata (%-bluedata altiscale Qlik @ C DOMINO Sense Brevia splice entelo Du bole xplenty inter ana ORBITAL INSIGHT ALGORITHMIA. ACTIONIO gitalGenius 😭 apppuri Stack IQ CHARTIO hi□ PREMISHITION: QUANTIFIND JENGAGIC fuse(machines NewSQL Databases Social BI Platforms Statistical Log Analytics NoSQL Databases Security Vertical AI Ad Optimization Analytics Computing amazon Google Cloud Platform Power BI **amazon splunk> Clustrix Pivotal □ CYLVNCE Applications NETBASE Ssas ORACLE sumologic CounterTack cybereason paradigm4 **DATASIFT** Microsoft Azure MarkLogic Wave Analytics Threat Metrix. NUODE € GoodData 🕏 birst DATASTAX kıbana tracx bitly >rocketfuet SPSS AREA 1 ■mongoDB MariaDB VOLTDB & citusdata synthesio Becorded Future Guardian theTradeDesk* Clara Couchbase **∢EROSPIKE** Sequola DB redistabs @ Influxdata KASIST® sift science A Contested Space izaí ♥SICNIFYD MPP Graph Cloud EDW Databases Databases Finance amazon informatica ⊿ffirm IIILendingClub aiteryx NUANCE 🌑 neoų j Google Cloud Platfo Tools T METAMARKETS VERTICA Put patential to work: Microsoft Azure OnDeck> ...Kreditech TRIFACTA Outbrain Dato noro 🗗 🛕 Numenta OPENGOV 7-9 Pivotal (V) MuleSoft ---tamr Residence Lendup N NETEZZA mixpanel contical.io snowflake snowflake snapLogic FN FiscalNote tidemark. 📆 INSIKT DATATORRENT kognitio deepsenseio VISENZE OMetaMind clarifai ₩₩: MindMeld OrientDB Chartbeat enigma Z UOra B Dataminr 777 Lenddo **BedrockData** 3 € dremio data Artisans ■PredictionIO glowfish IDIBON 🐠) PREDPOL Alation A Infoworks mark43 **iSENTIUM** KENSHO Data For Business SMB/ Crowd-Search Management Security Storage App Dev "LE!" OpenDataSoft Quantopian Yieldmo M James ORACLE Services Analysts Commerce / Monitoring **TANIUM** sourcing apigee amazon Google Analytics illumio 🌃 Origamil..ogic New Relic. S EXALERO OPERA Industries Life Sciences Education/ APPDYNAMICS MANPLITUDE RJMetrics CODE42 Congle Court OP@WER eHarmony Lucidworks Learning amazon octifio Mu Sigma ClearStory C Counsul 🖢 🖟 **BLUECORE** DataGravity DATASCIENCE elastic 💽 ThoughtSpot RetailNext Numerify KNEWTON panasas/ CipherCloud XX Recombine sumAll @granify splunk Typesafe CrowdFlower CIRRO KYRUUS FLATIRON STITCH FIX VECTRA Clever M∧∧N∧ @swiftype nimblestorage Airtable . oe⇔ezymergen HealthTap® retention custora Qumulo BLUESRIVER import (6) Algólia SINEQUA Procana Anedes @eclara rachyus , Seeq FarmLogs METABIOTA ZEPHYR SwiftKey Cross-Infrastructure/Analytics PANORAMA HowGood celect @ MACHINE Ginger.io ** transcriptic Glow @entric DAiCure 🗘 Attention knowto statmuse B@XEVER Famazon Google 🏪 Microsoft 🏗 📆 SAP SSAS 🥠 🕬 VMWAYE 🕫 talend TIBC TERADATA ORACLE 🗖 NetApp Open Source Machine Learning 🚕 Stat Tools Query / Data Flow Coordination Real-Time Framework Data Access Search Security accumula Apache SINGA MAD lib. Apache Ranger •talend 👊 👊 👊 📧 Spark Spark HEASE mongoDE Aerosolve **≣**Scala Caffe CNTK TerrsorFlow Visualization























Vicinso ⊘ Human API













The reality of Big Data

Big Data is a Big Confusion

AUGUST 8, 2012 by Josh Sternberg

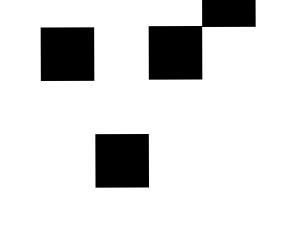
19 September, 2013 **Gartner On Big Data: Everyone's Doing It, No** We're in the era of Big Data, but we may as well call it the era of Big I Have No Clue. One Knows Why 7 May 2014 By Sharon Everitt The same enterprises that seem most confused about Big Data seem to be the ones launching Big Data projects. What gives? The gravitational pull of Big Data is now so strong that even people who u Like 0 haven't a clue as to what it's all about report that they're running Big Data projects. lack of big data know-how has created a backlog that gives way to a flood of demands for business intelligence (BI), a Strange, but true. report has predicted. As a result, data center managers could be swamped by requests for resources in the near future, it According to a recent Gartner report, 64% of enterprises surveyed indicate FINANCIAL TIMES Global Economy Lex Arts | Magazine | Food & Drink | House & Home | Lunch with the FT | Style | Books | Pursuits | Travel | Columns | How To Spend It Tools

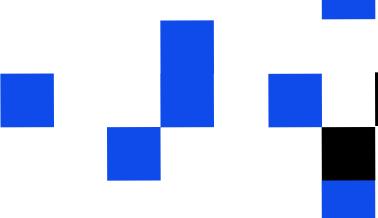
Big data: are we making a big mistake?

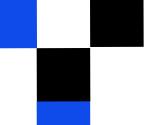
By Tim Harford

March 28, 2014 11:38 am

Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media

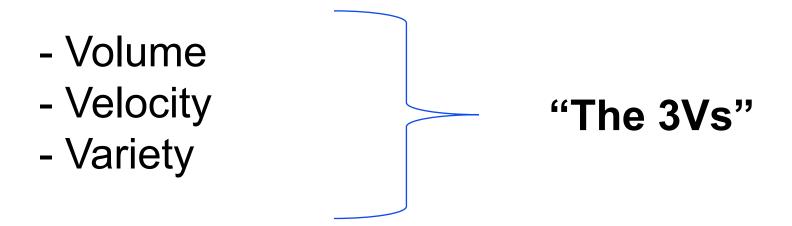






What is (are) "Big Data"?

Big Data: Key defining characteristics:



- Veracity

What is (are) "Big Data"?

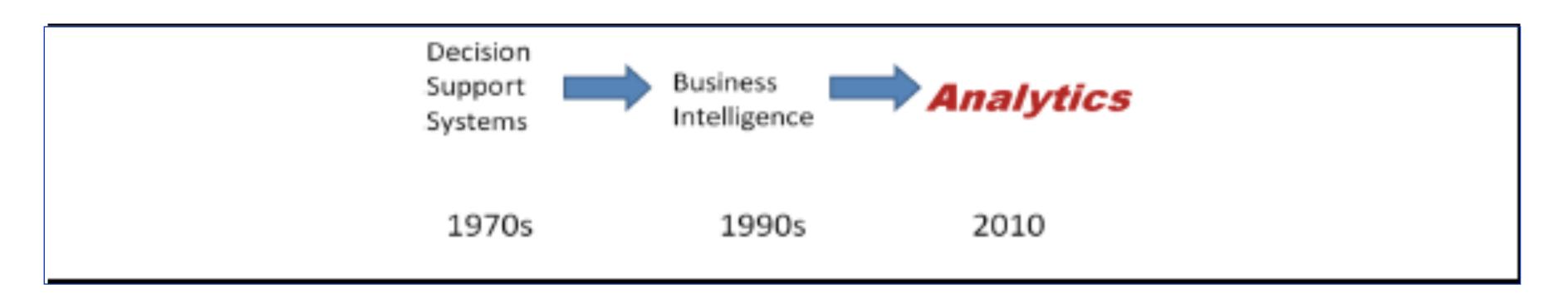
Industry practitioners keep inventing additional Vs

- Variability
- Visualisation
- Value
- Variability
- Venue
- Vocabulary
- Vagueness
- Virility
- Vendible
- Vaticination
- Voracity
- Vanity
- Etc.

"If you can sell it, and sell it as Big Data, then it 'is' Big Data."

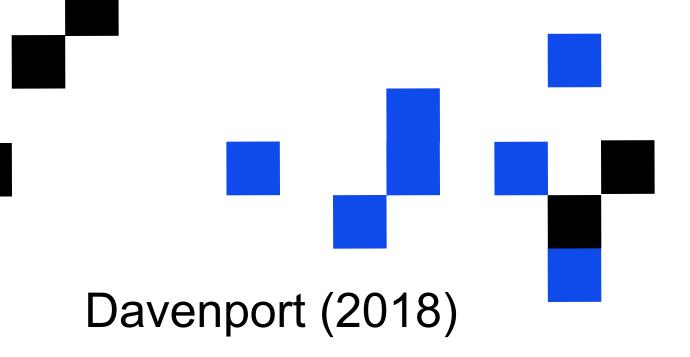
"Big Data provides the sort of visionary and predictive powers only previously obtainable through ritual sacrifice, magic potions and the casting of spells"





Watson (2014)

Evolution of Business Analytics



Analytics 1,0

- Small, structured, static data
- · Back-office analysts
- Slow, painstaking
- Internal decisions
- Descriptive analytics
- Human hypotheses

Analytics 2.0

- Big, unstructured, fastmoving data
- Rise of data scientists
- Data products in online firms
- Rise of Hadoop and open source

Visual analytics

Analytics 3.0

- Mix of all data
- Internal/external products/decisions
- Analytics is a core capability
- Move at speed and scale
- Predictive and prescriptive analytics

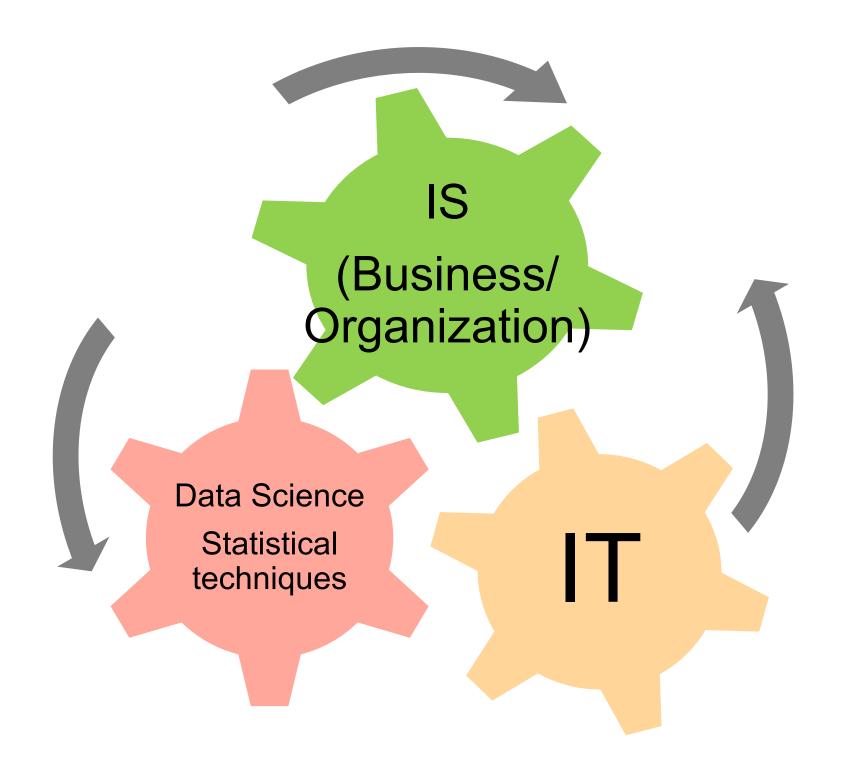
Analytics 4.0

- Analytics embedded, automated
- Cognitive technologies
- "Robotic process automation" for digital tasks
- Augmentation, not automation

Business Intelligence Business Analytics Big Data Analytics Analytics

AI & Advanced Analytics

Business Analytics: Different perspectives



MIT 's survey ("Big Data, Analytics, and the Path from Insights to Value") of 3000+ executives from 108 countries across 30 industries confirms that the biggest challenges in adopting analytics are managerial and cultural (MIT Sloan and IBM, 2014)

From Data to Insight to Action

Important Questions to ask:

What kind of data?

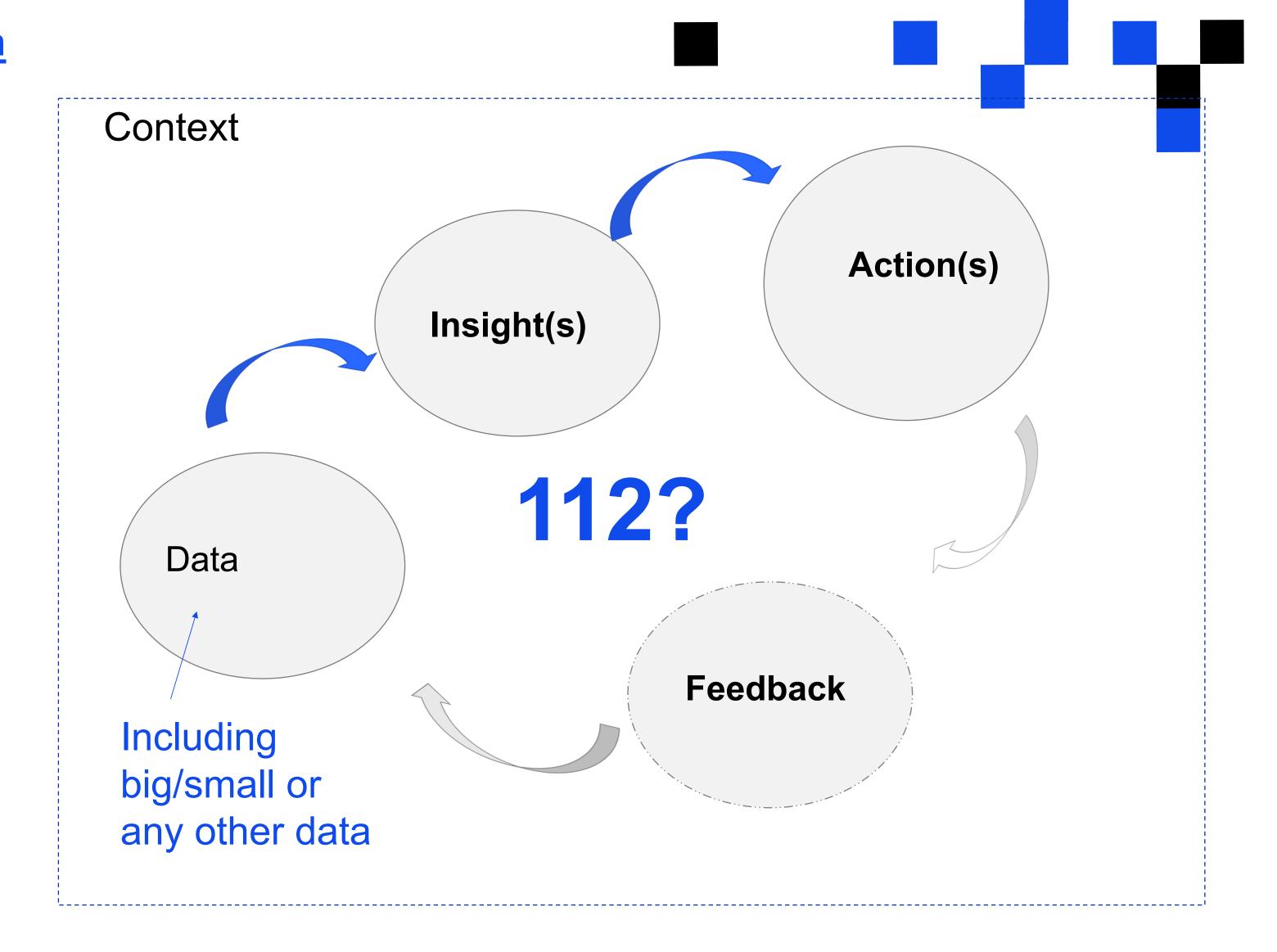
From where?

Captured how?

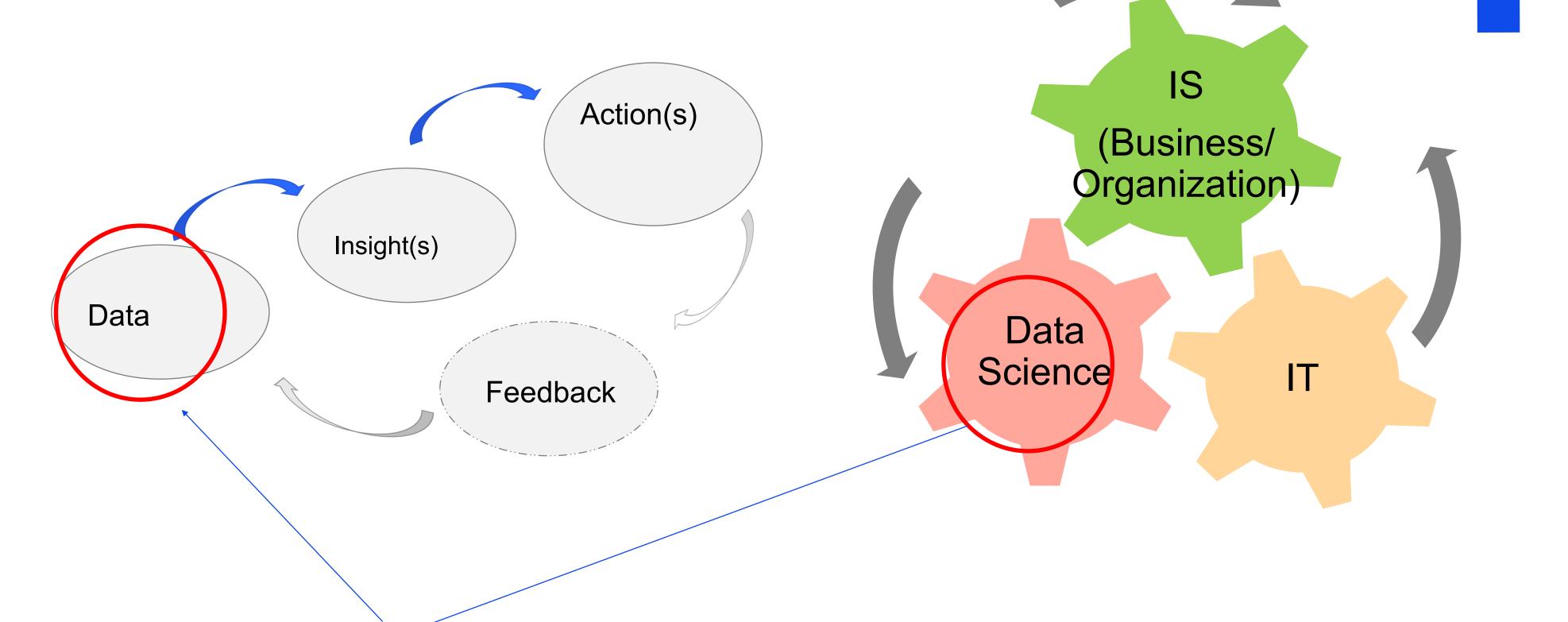
In which context?

For what purpose?

To address what problem?

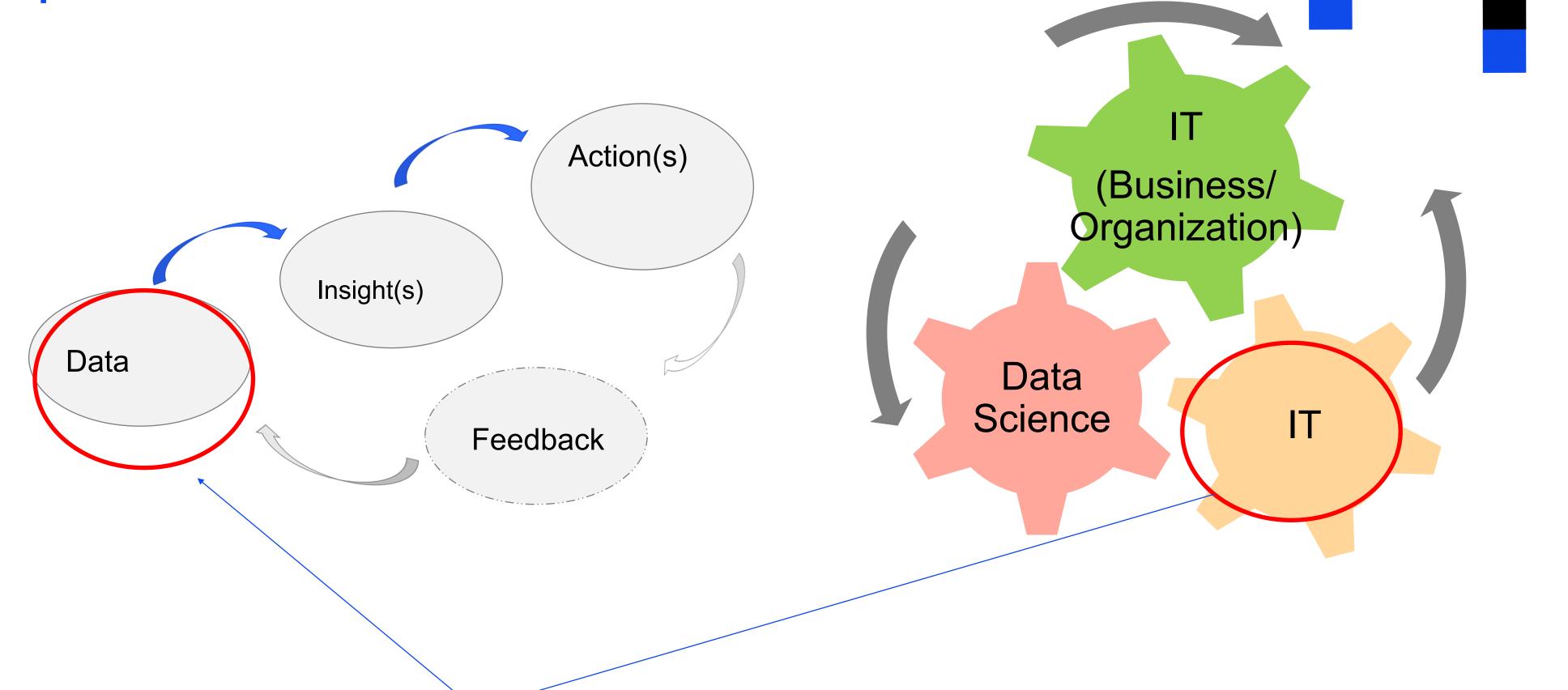


The Data Science Perspective



Data science: Data first; Business context not considered (e.g. business processes); Strong focus on numerical data; Very sophisticated knowledge of statistical methods; Limited domain knowledge; Provide information to decision makers

The IT Perspective

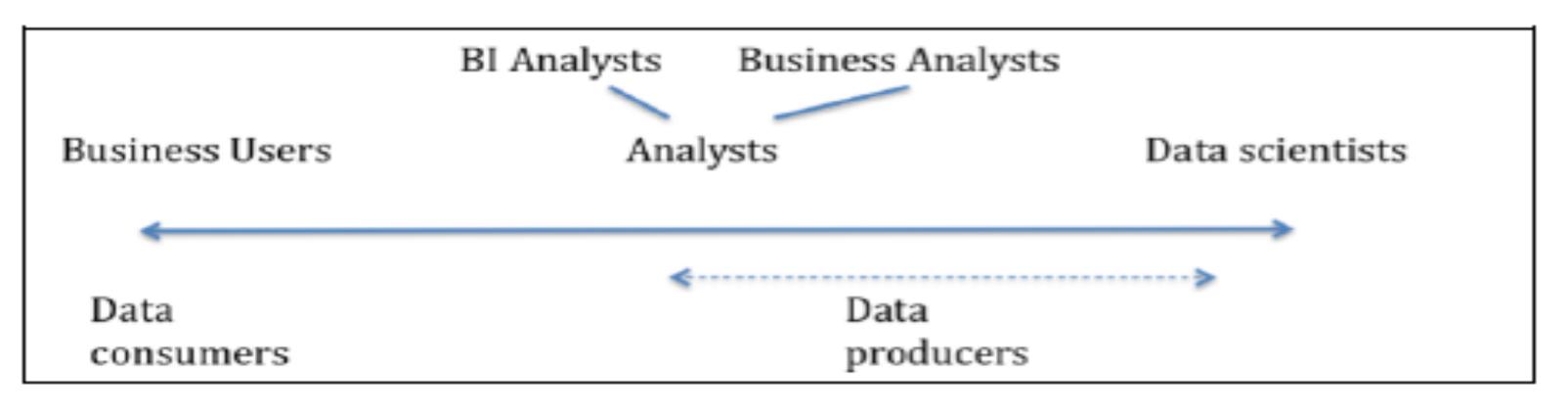


IT: Data first; Business context not considered (e.g. business processes); Strong focus on data administration, storage, (warehouses/ data marts, hadoop-based technologies), data integration, transformation & load; speed of access; algorithms; structured and unstructured data; Set environment for data science; Provide information to decision makers

The Information Systems (IS) perspective Business problem Action(s) or opportunity (Business/ Organization Insight(s) Data Data Science Feedback

Org/Business: Concern with Context; Starts from Business problem and/or opportunity; Domain knowledge; Responsible for decisions, actions and consequences; Challenges with defining and collecting effective feedback (more than numerical data; Should consider wider societal context (but often does not)

Business Analytics: Different roles



(Watson, 2014)

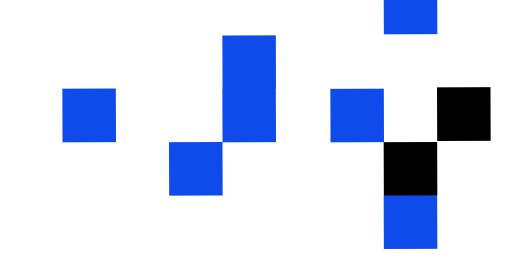
Latest update: Business Users of Visual Analytics as Data producers ("Citizen scientists")

The Emergence of Visual Analytics (VA)

- (Mainstream) Business Intelligence & Analytics:
 - numerical data,
 - focus on finding answering to given (business questions) questions
 - "Traditional" approaches to visualisation pie charts, graphs etc.
- Rapidly growing: Visual Analytics (VA)
 - Visual data of different kind (numbers, qual data, sound, pictures etc.)
 - Focus on <u>finding relevant questions</u>
 - Does not require knowledge and experience in "deep-analytics" & even statistics
 - Story-telling with data

The largest international survey conducted by TDWI: the greatest barriers to organizational adoption of visual analytics are lack of skilled personnel and training (TDWI, 2013, pp.5).

The Emergence of Data Visualisation & Visual Analytics (VA)



Data visualisation



Visual data exploration



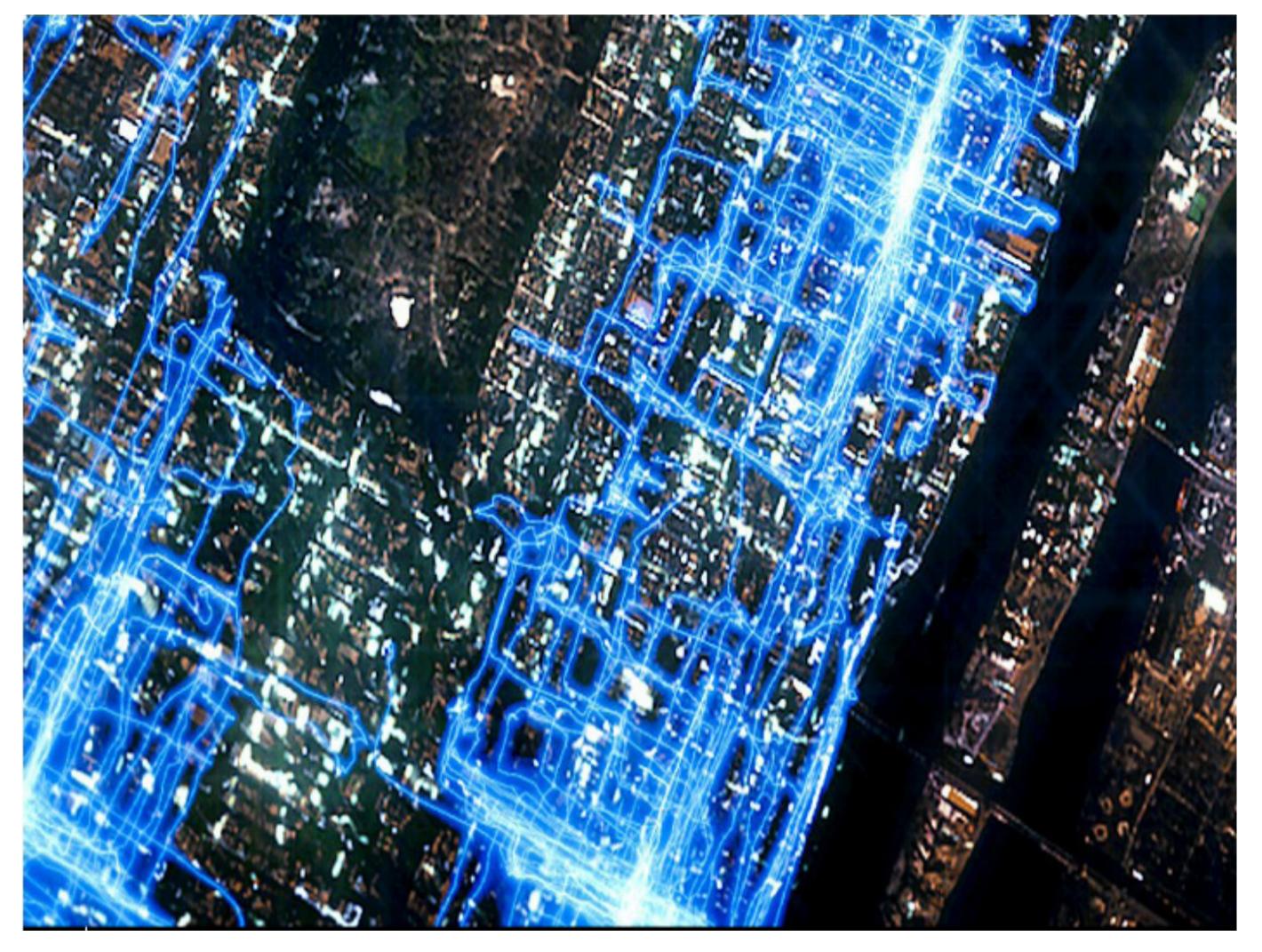
Answer the given question(s)

Formulate <u>new question(s)</u>

Data-driven story-telling

The greatest barriers to organizational adoption of visual analytics are lack of skilled personnel and training (TDWI, 2017)

VA makes it easier for humans to observe patterns in large volumes of data (HBR, 2016)

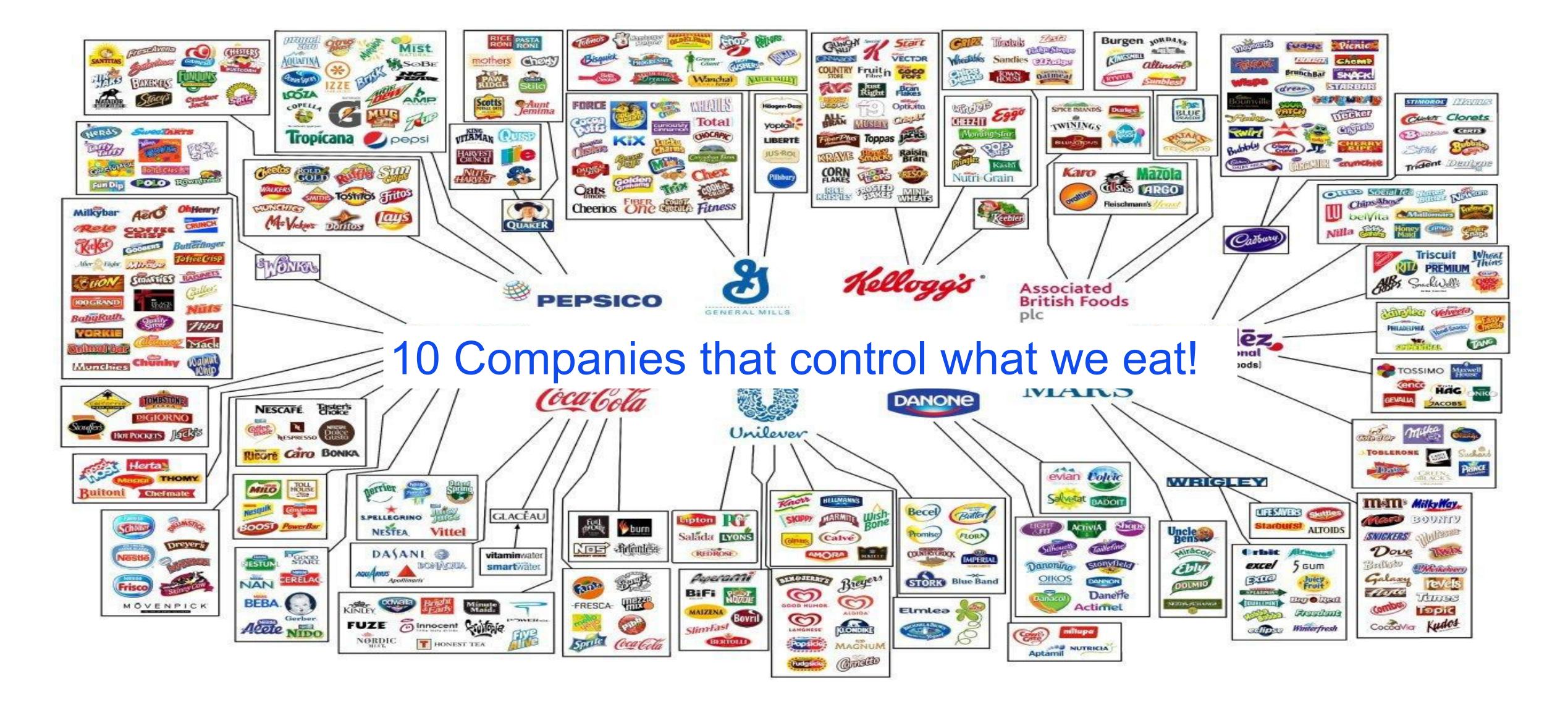


Smolan and Erwitt (2012), The Human Face of Big Data, Against All Odds Productions;

VA - ... we visualise data to ... see what could not be seen before" (HBR, 2016)



VA - ... we visualise data to ... see what could not be seen before" (HBR, 2016)



VA- we visualise data to prompt people (decision-makers) to take action

Origin	Airline	Flight No.	Scheduled	Estimated ~	Flight Listing Status
Singapore	Qantas	QF6 9W4006, AY5003, EK5006, FJ5321, MU8455 PG4537, UL3376	5, 09:50	08:59	Arrived >
Kuala Lumpur	AirAsia X	D7222	09:50	09:17	Arrived >
Auckland	Virgin Australia	VA143 AB880, AZ4211, EY6516, NZ7901, SQ6357, VA143	, 09:45	09:28	Arrived >
Singapore	Singapore Airlines	SQ231 LX4184, TK9320, VA5509, VS7231	10:25	09:37	Arrived >
Kuala Lumpur	Malaysia Airlines	MH141 KL4119, UL2341	10:00	09:54	Arrived >
Kuala Lumpur	Malaysia Airlines	MH123 KL4105, UL2323	10:00	09:54	Arrived >
Port Vila	Air Vanuatu	NF10 QF376	09:55	09:56	Arrived >



VA- we visualise data to move people to feel ..." (HBR, 2016)



VA- we visualise data to move people to feel ..." (HBR, 2016)



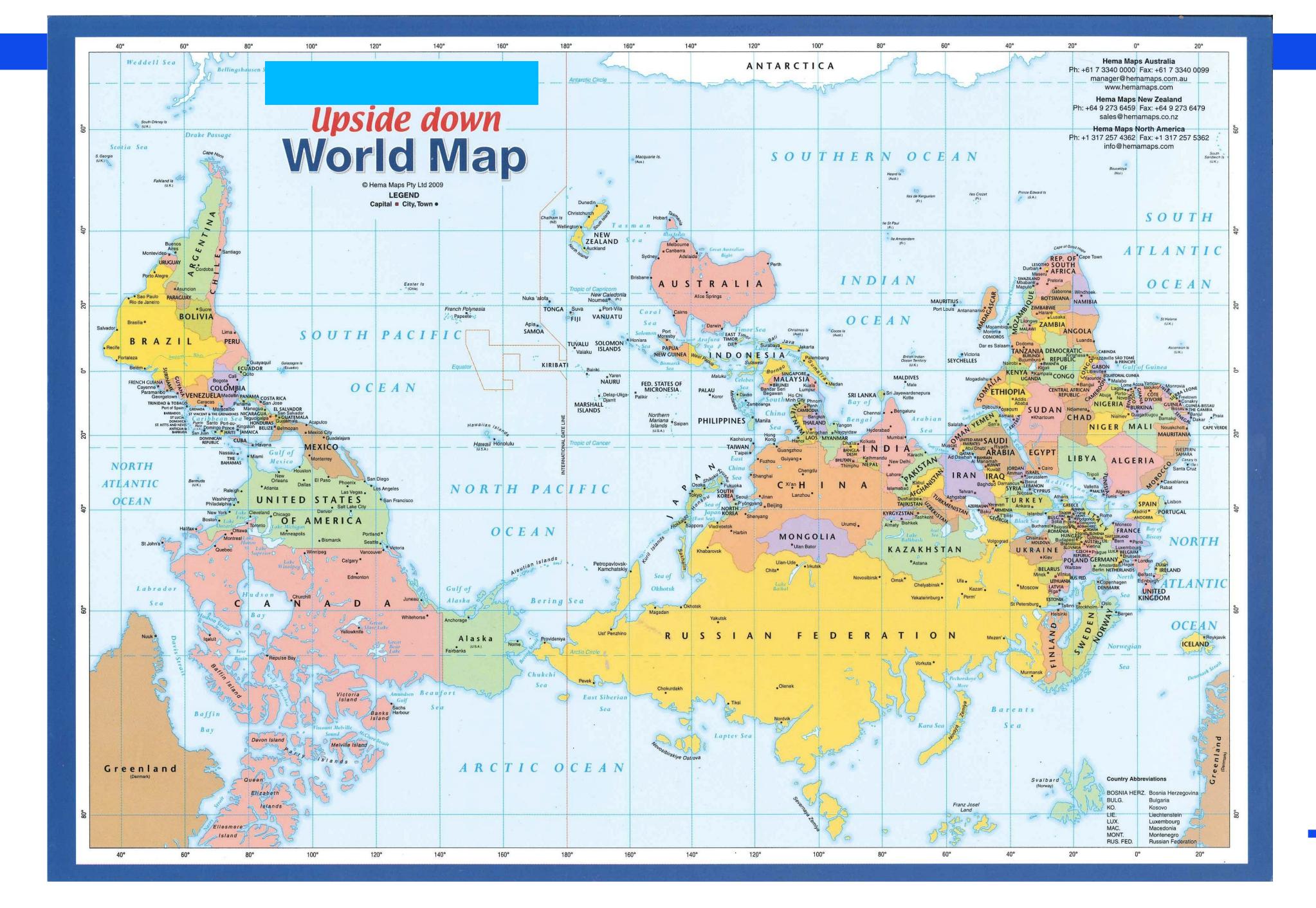
Important: What kind of data?



Visual Ethics

Visualisations influence people much more than numbers

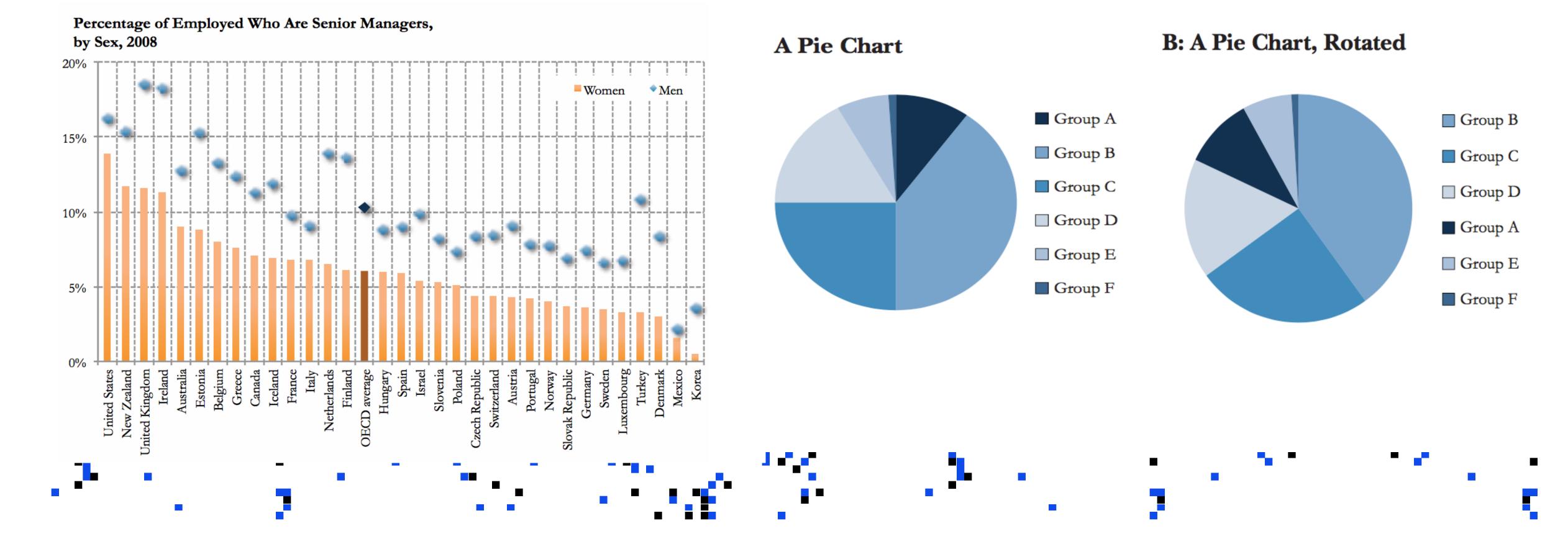




Example: Visual Ethics

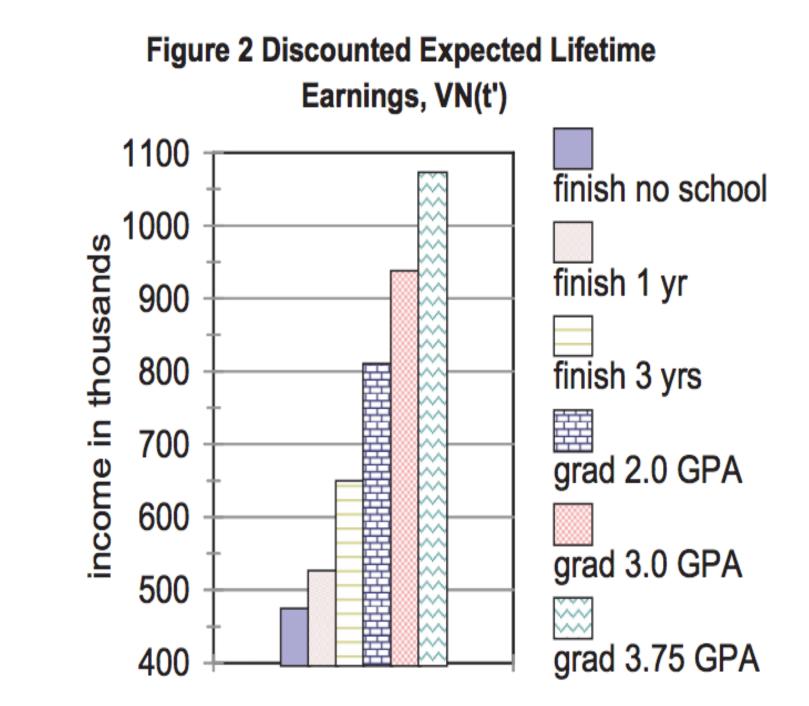
- Different forms of data visualization have the power to distort information and <u>influence</u> people much more than numbers

An Unbalanced Chart



Example: Visual Ethics

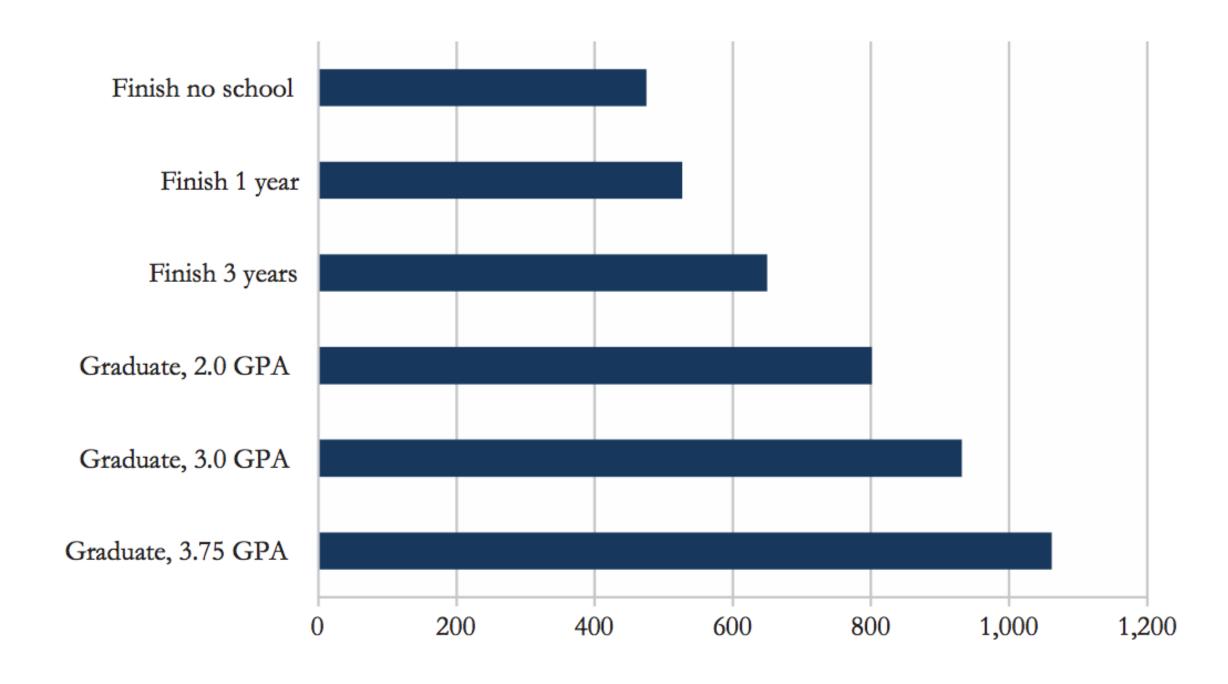
The Basic Column Chart



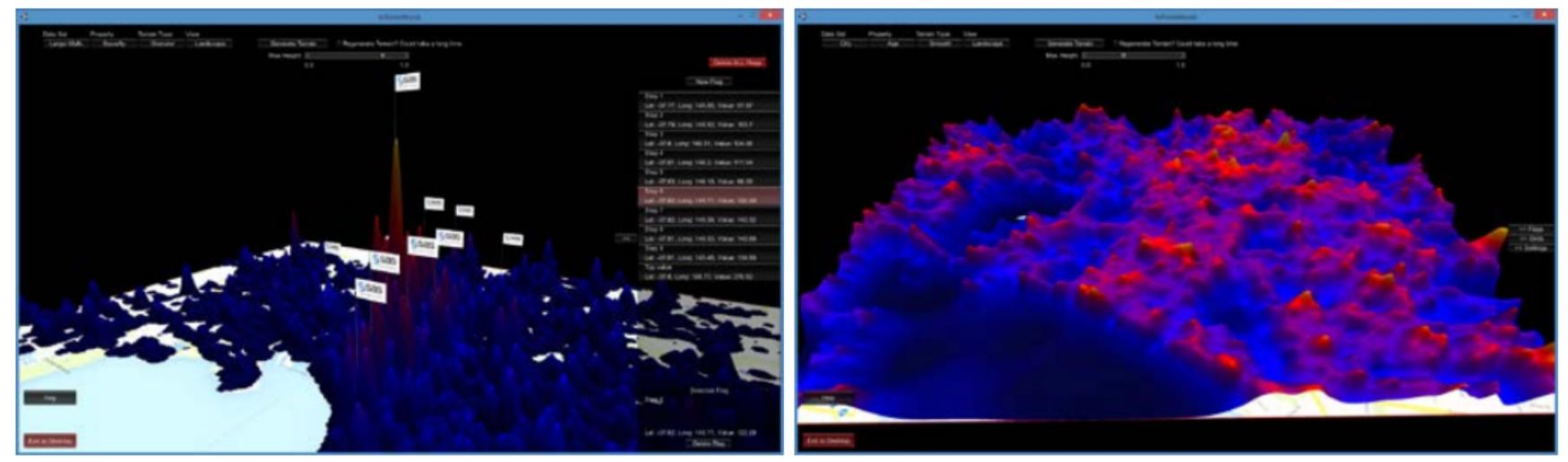
Source: Stinebrickner and Stinebrickner (2013).

The Revised Column Chart

Discounted Expected Lifetime Earnings, VN(t') (Income in thousands)



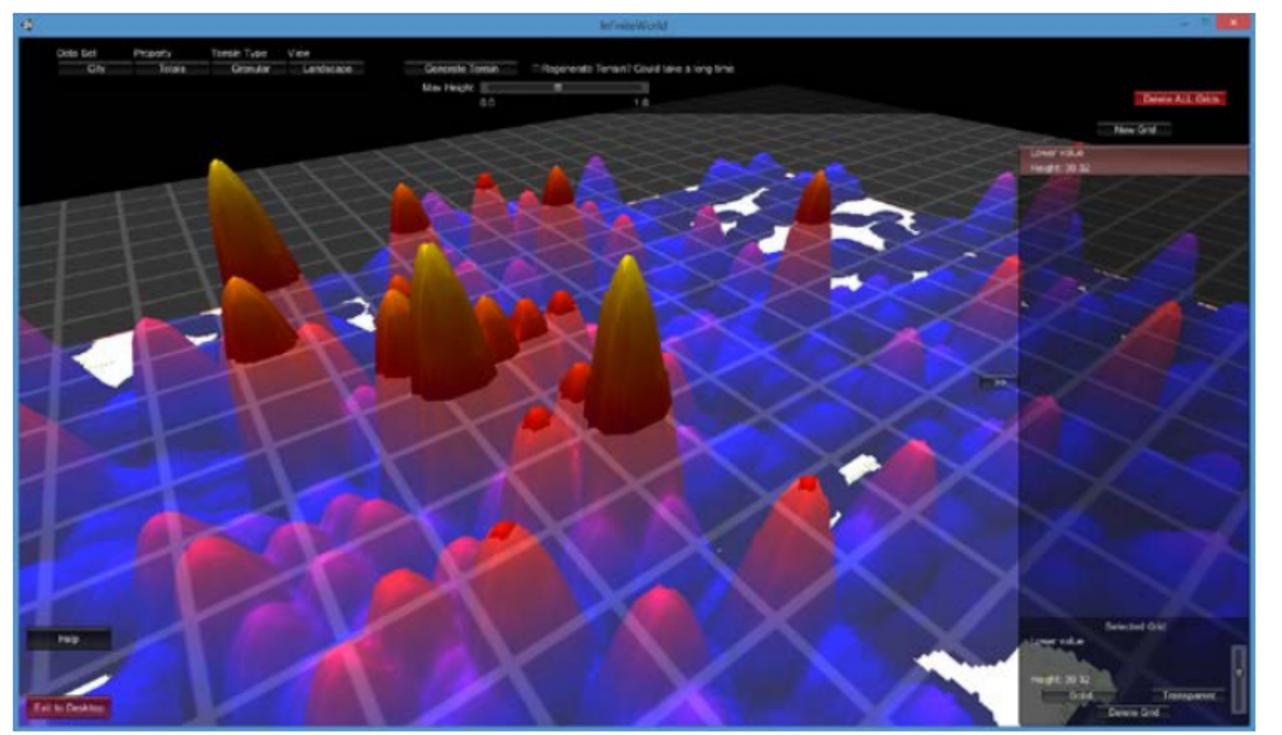
(Schwabish, 2014)



(a) Total number of accidents (granular + flags)

(b) Age of people in accidents (smooth)

3D is not necessarily better (easier to comprehend) than 2D!



(c) Navigation, measuring and comparing of data terrain features (grids)

Organisational Challenges of VA



Observations of Visual Analytics in practice



Published industry
Case studies



Interviews with industry practitioners



- -How do practitioners use VA?
- -- who is using VA?
- --for what purpose?
- -- what kind of decisions do they make?
- -What are the current challenges?
- -What are the new opportunities?
- -What are the required skills?

Industry practices

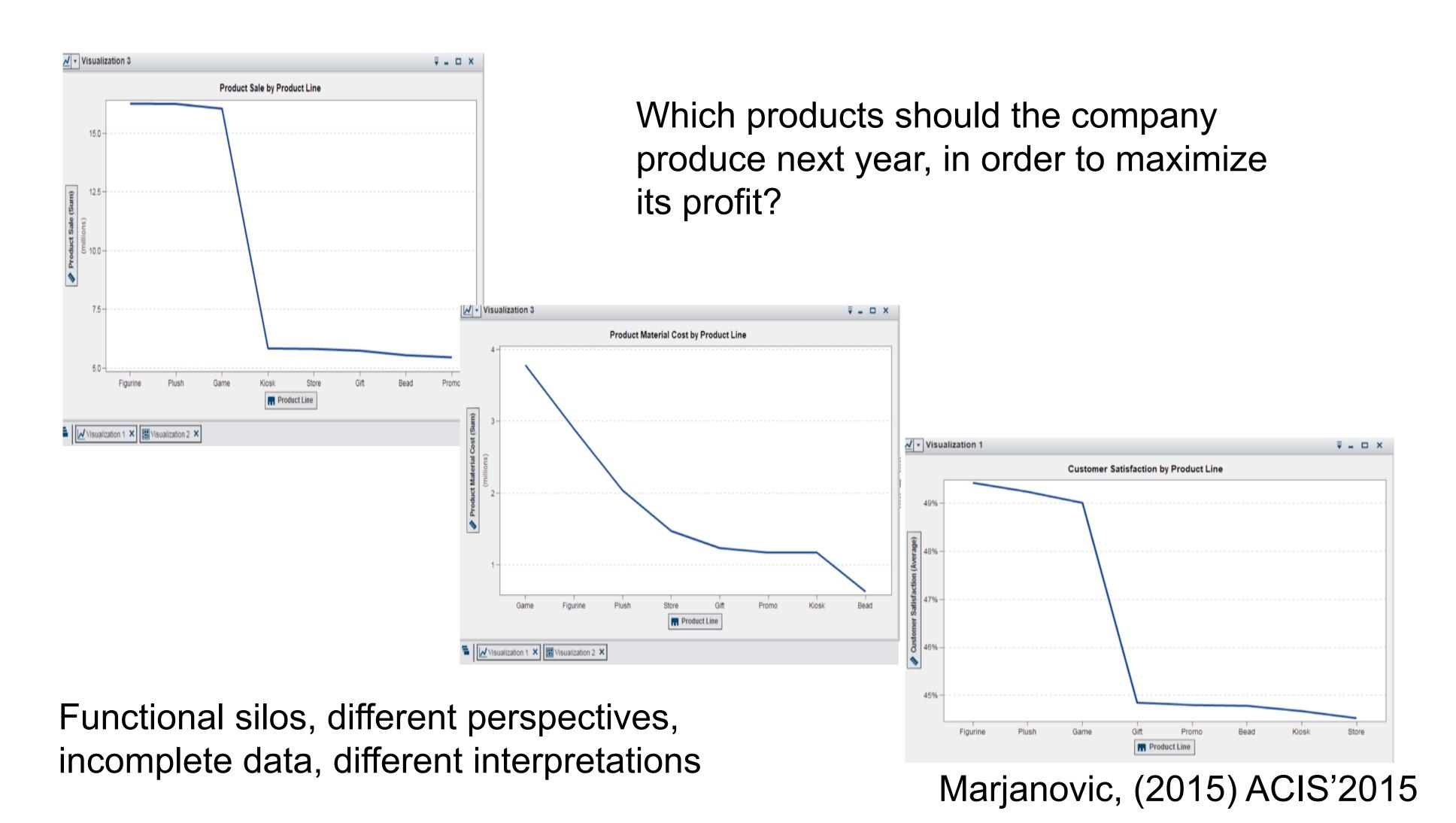
Patterns



Marjanovic: VA of the Future, Au Gov OLT grant (2014-2016)

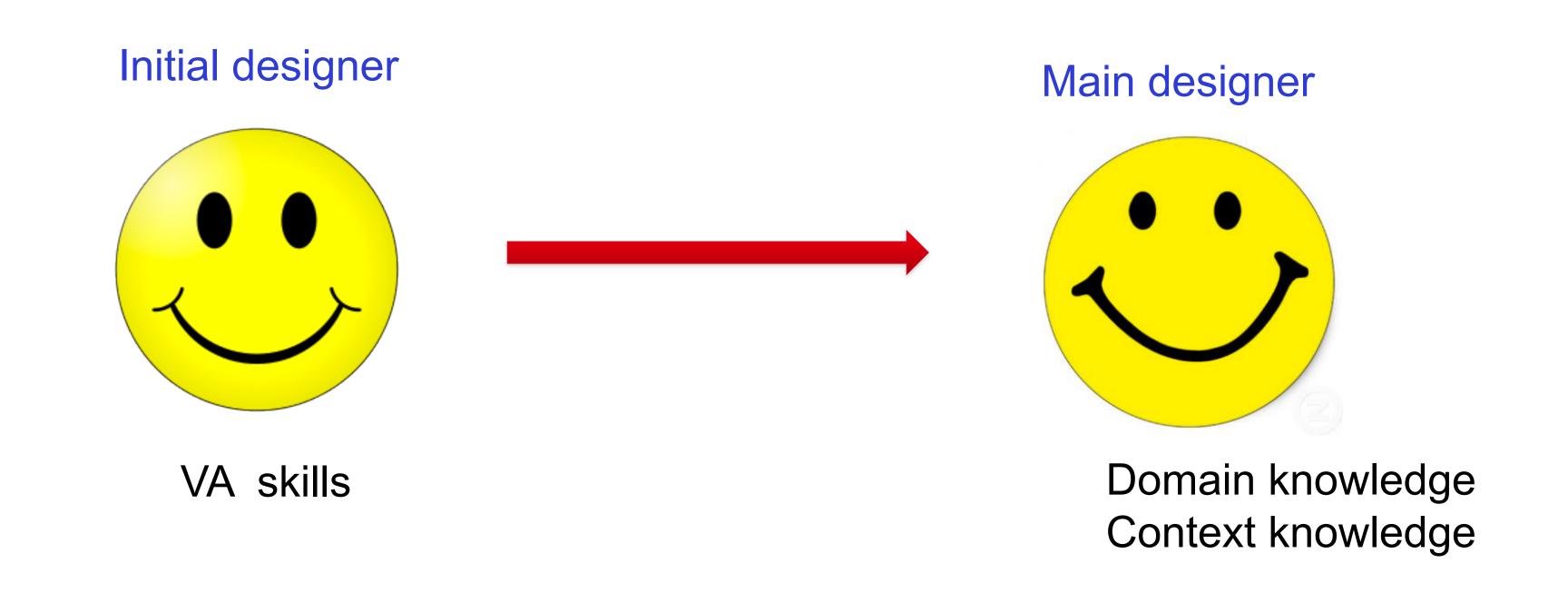
An example: Collaborative VA brings to the surface some important organisational issues...

An interesting experiment: Three groups of students: sales, finance and customer services.



A major shift:

Visualization as a process (verb) rather than an outcome (noun)



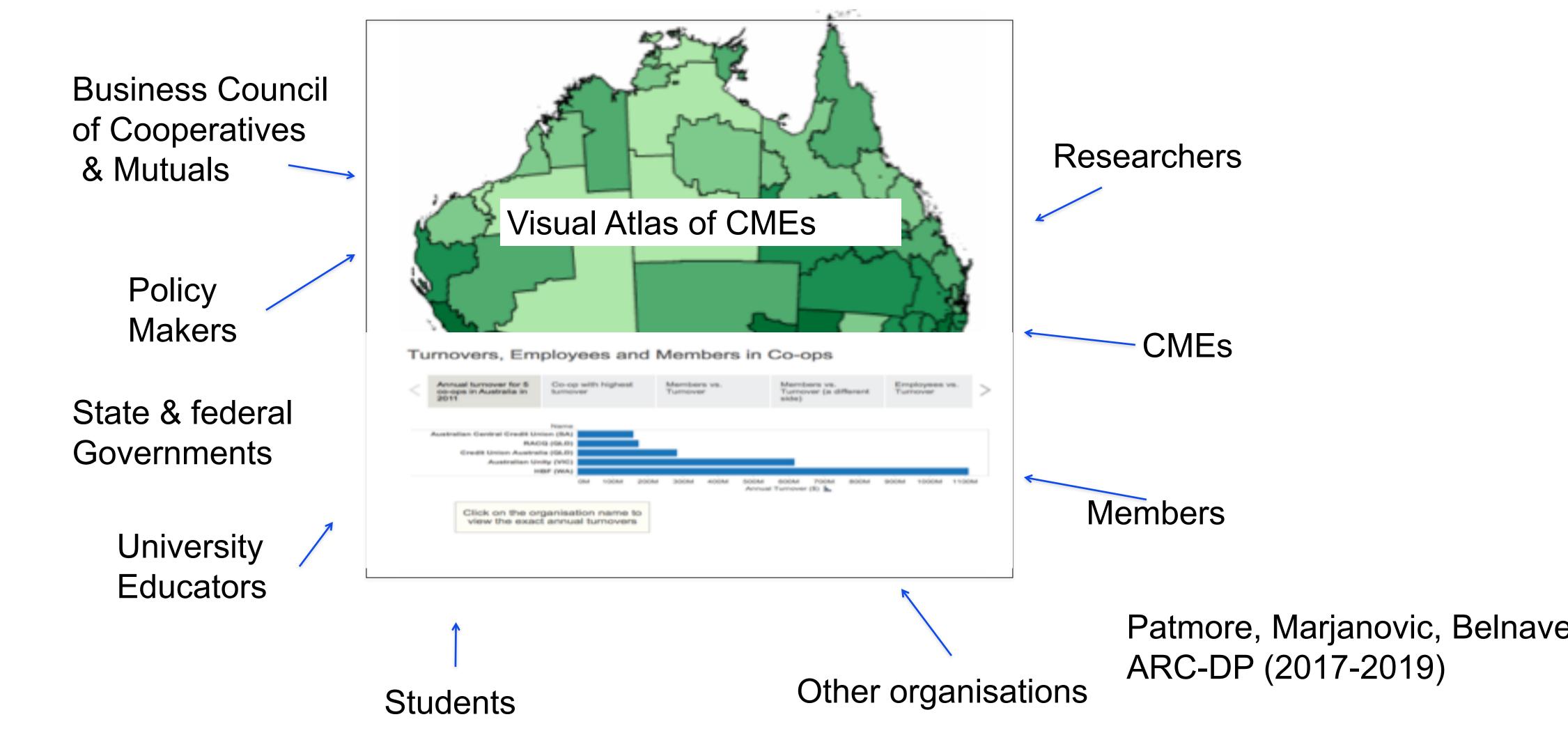
Visualising data



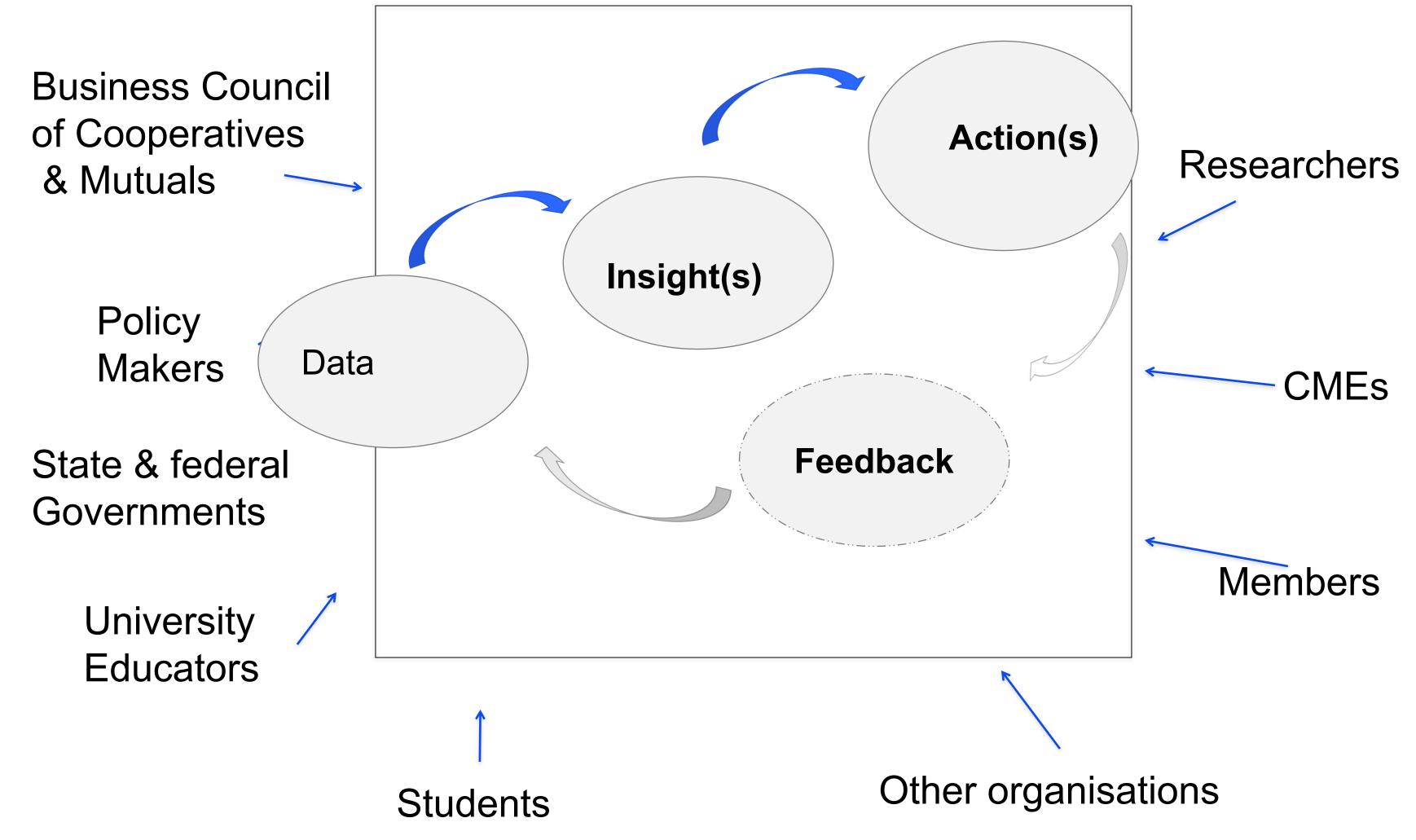
Visualising through data

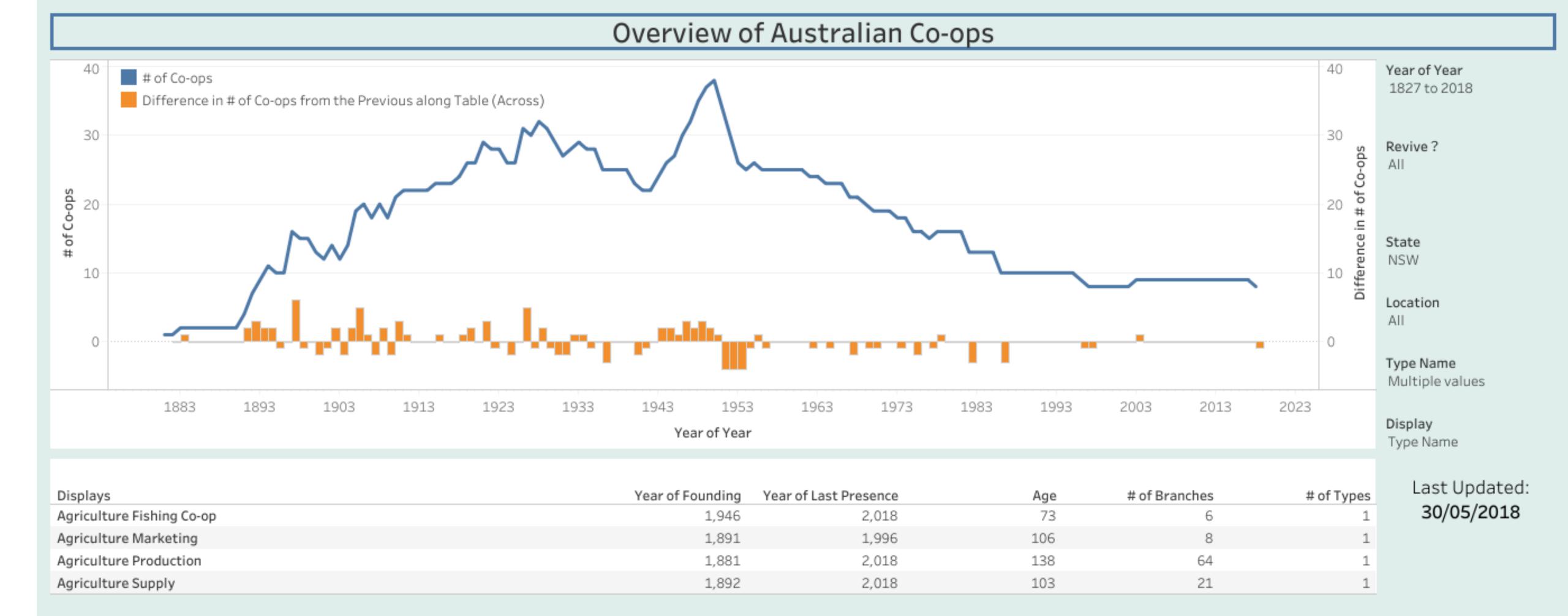
Marjanovic (2015) QC Re-Imagine Education USA

An example: Visual Atlas of Australian Cooperatives



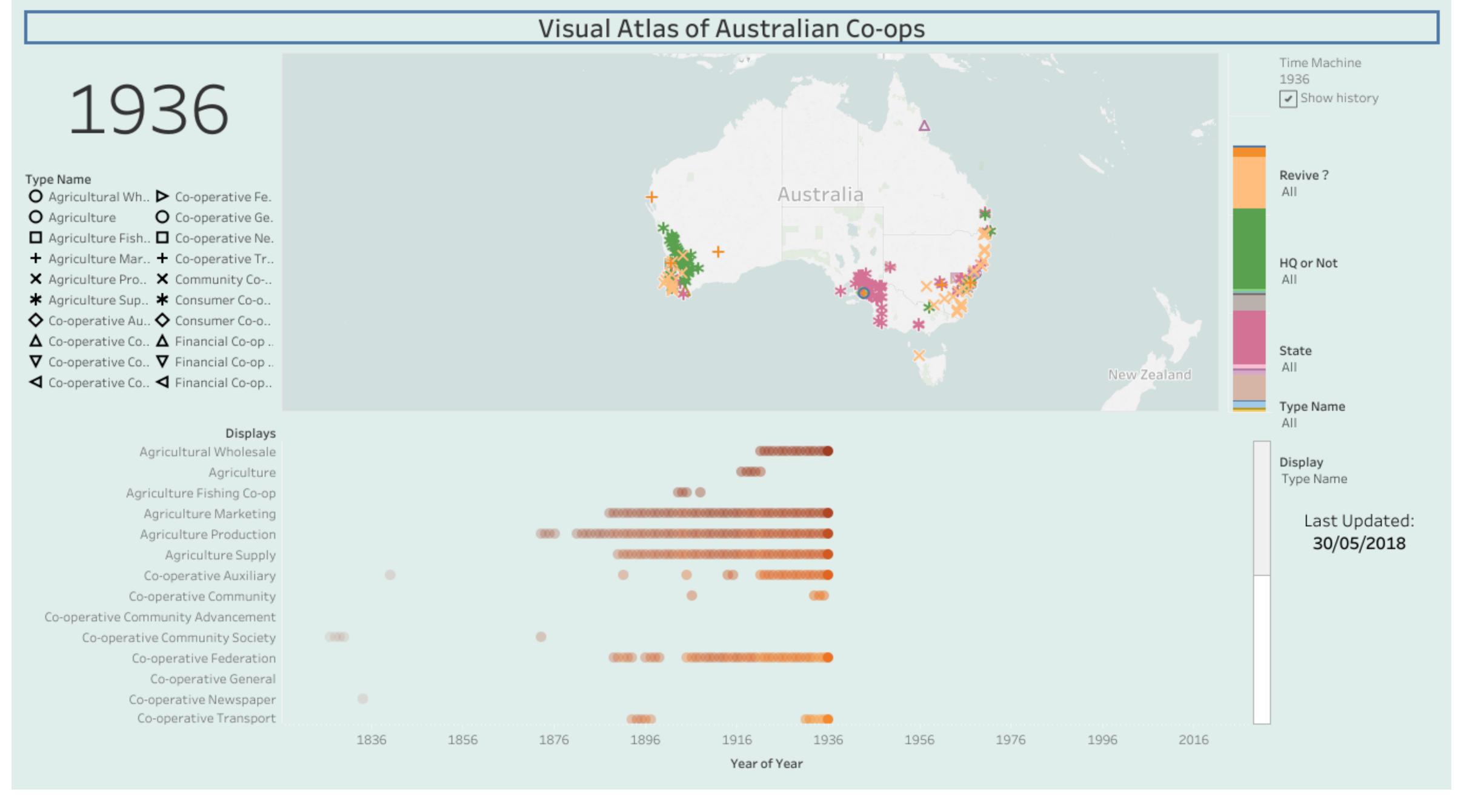
Visual Atlas of Australian Cooperatives

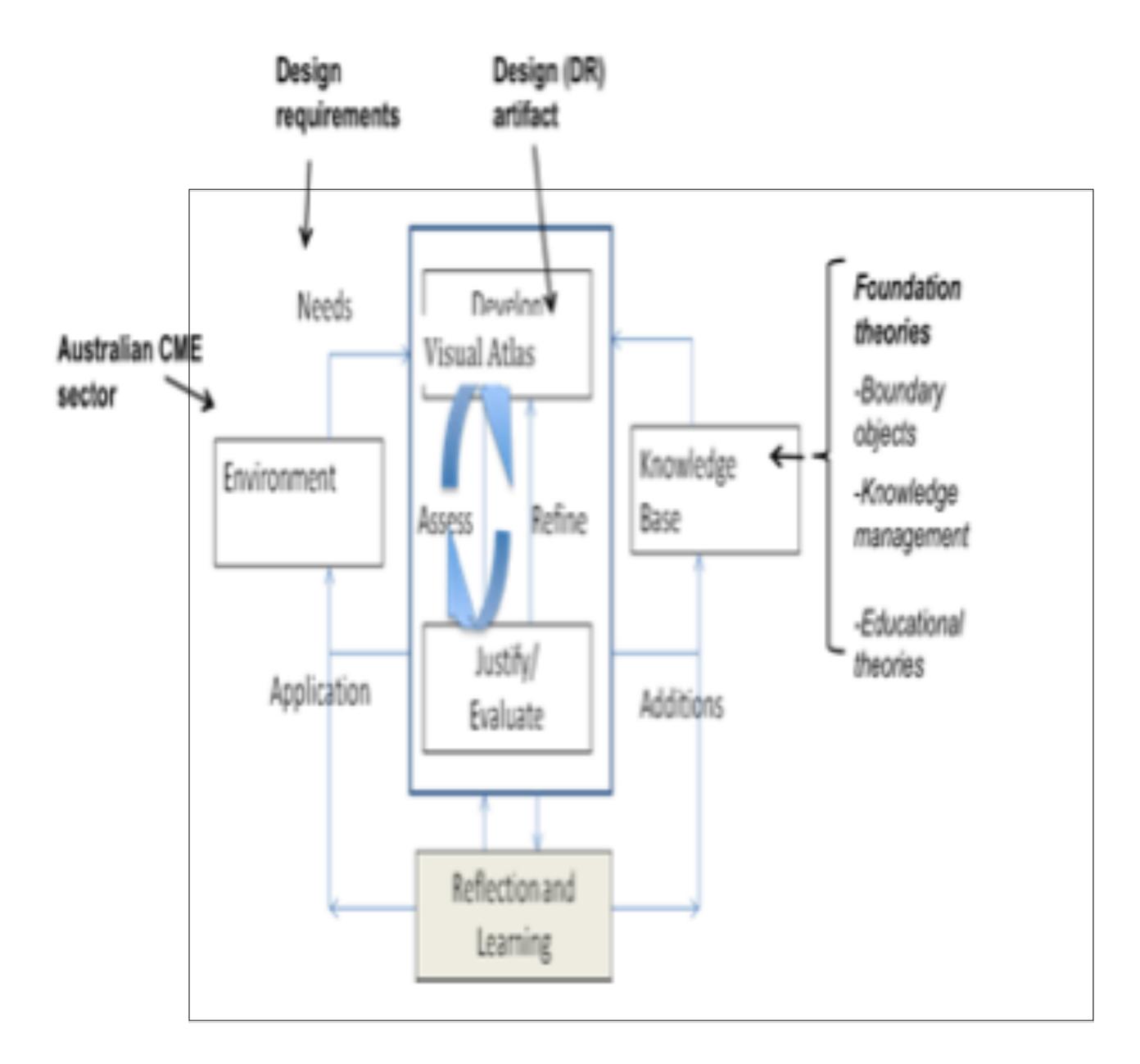




Important considerations: Visual Ethics, "Performative aspects of data"

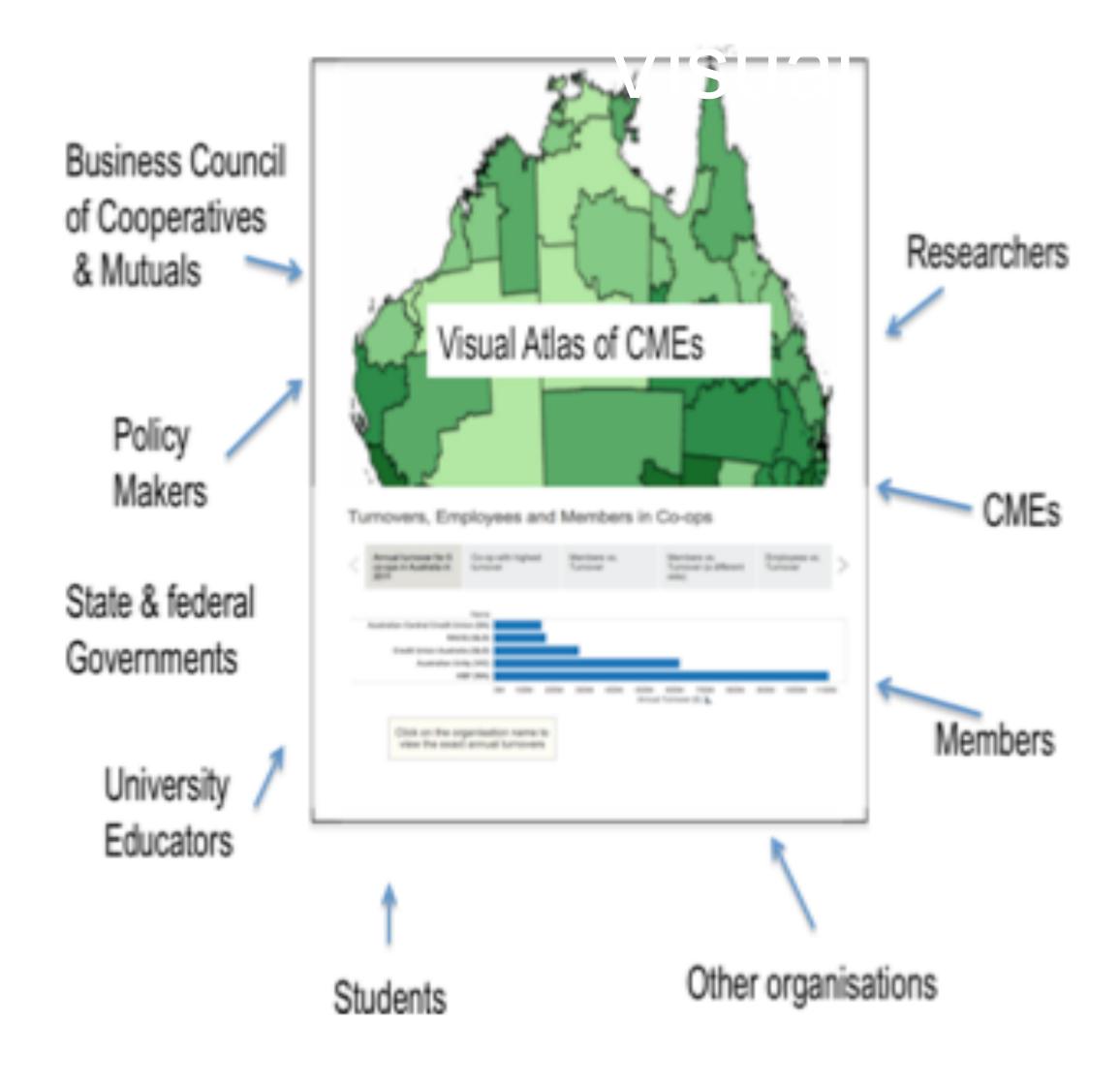
Please note: These are static screen-shots and as such do not represent dynamic aspects of Visual Atlas





Design research artifacts:

- A very comprehensive high quality data set collected by historians
- -A "Visual Atlas" infrastructure ready for future opportunities
- -- a method of empowering practitioners to engage in ongoing data exploration through secondary design
- -- a collection of datavisualization stories



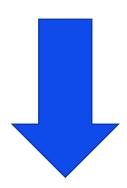
PROFESSOR O. MARJANOVIC, "BIG DATA, ANALYTICS AND DATA VISUALIZATION", NOVI SAD, 23 SEPT 2019

Interesting research challenges (so far):

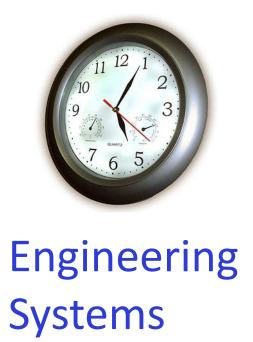
- VA Tool -> VA Infrastructure
- Participatory co-design
- Transfer of skills from Primary to "Secondary" designers
- Visual storytelling
- Knowledge-sharing processes through collaborative sensemaking
- Design of data-driven services
- Visual boundary objects
- Open data ethics and datafication effects
- Visual ethics
- Participatory data quality
- Visual Atlas as "pattern" for future VA infrastructures
- Visual Atlas as a "living data environment"
- Visualisation of conceptual phenomena (e..g. visualisation of impact)
- Visual Atlas as a Human Complex System

Beyond Visual Atlas

Business Analytics (Data-driven)



Analytics in Human Complex Systems (Requires "Thinking in Systems"*)







Living systems



Example: Datafication effects of MySchool Marjanovic and Cecez-Kecmanovic (2017) https://www.sciencedirect.com/science/article/abs/pii/S0963868717302482

Think Clouds, not Clocks

Karl Popper, the great philosopher, said all problems are either clouds or clocks..

One of the problems we have as a culture is we take clouds & pretend they are clocks.

Towards data humanism...

