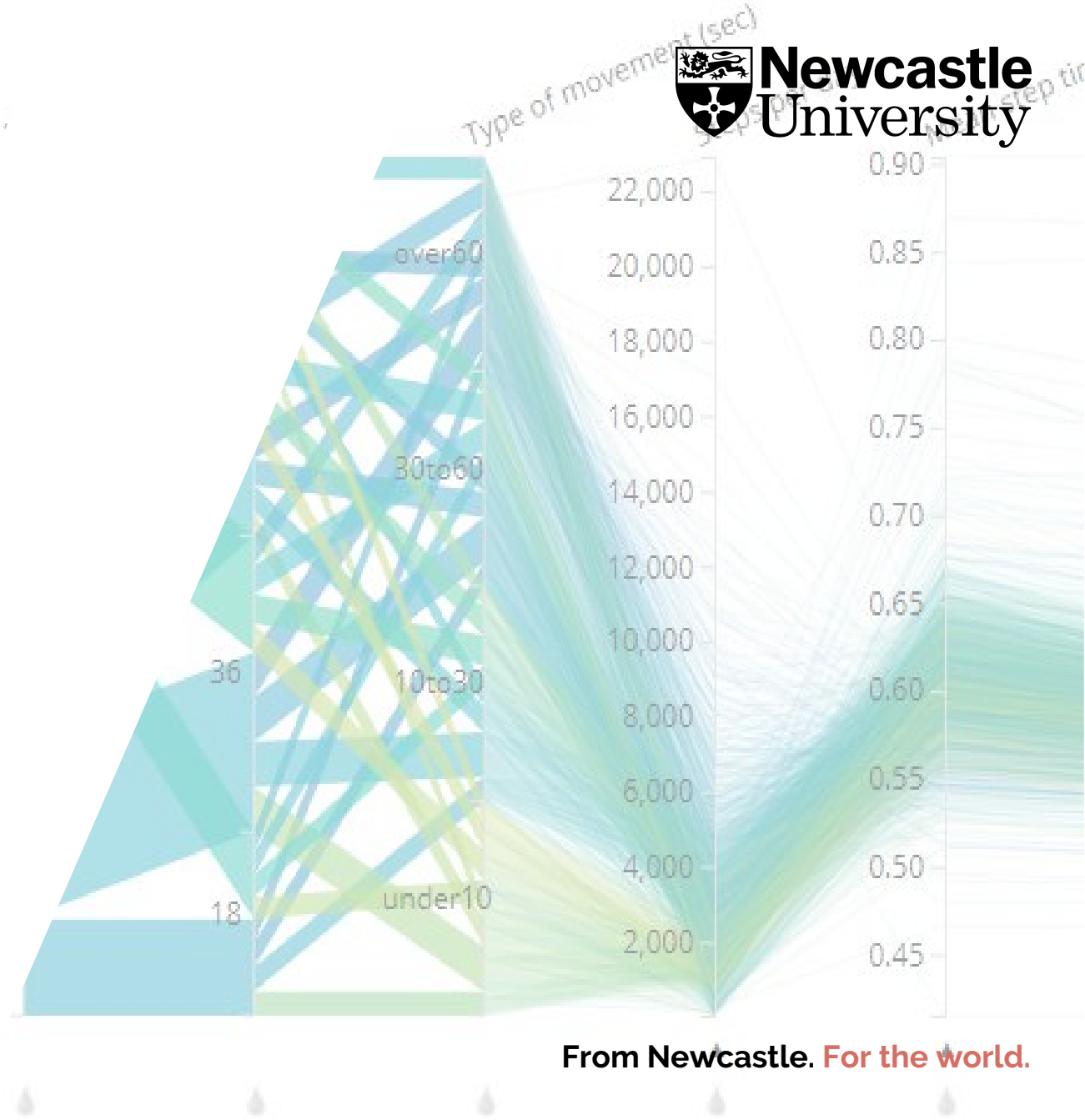
Data Visualization for Complex Data

Introduction and examples of applications

Alma Cantu
Lecturer is Data Visualization
Newcastle University



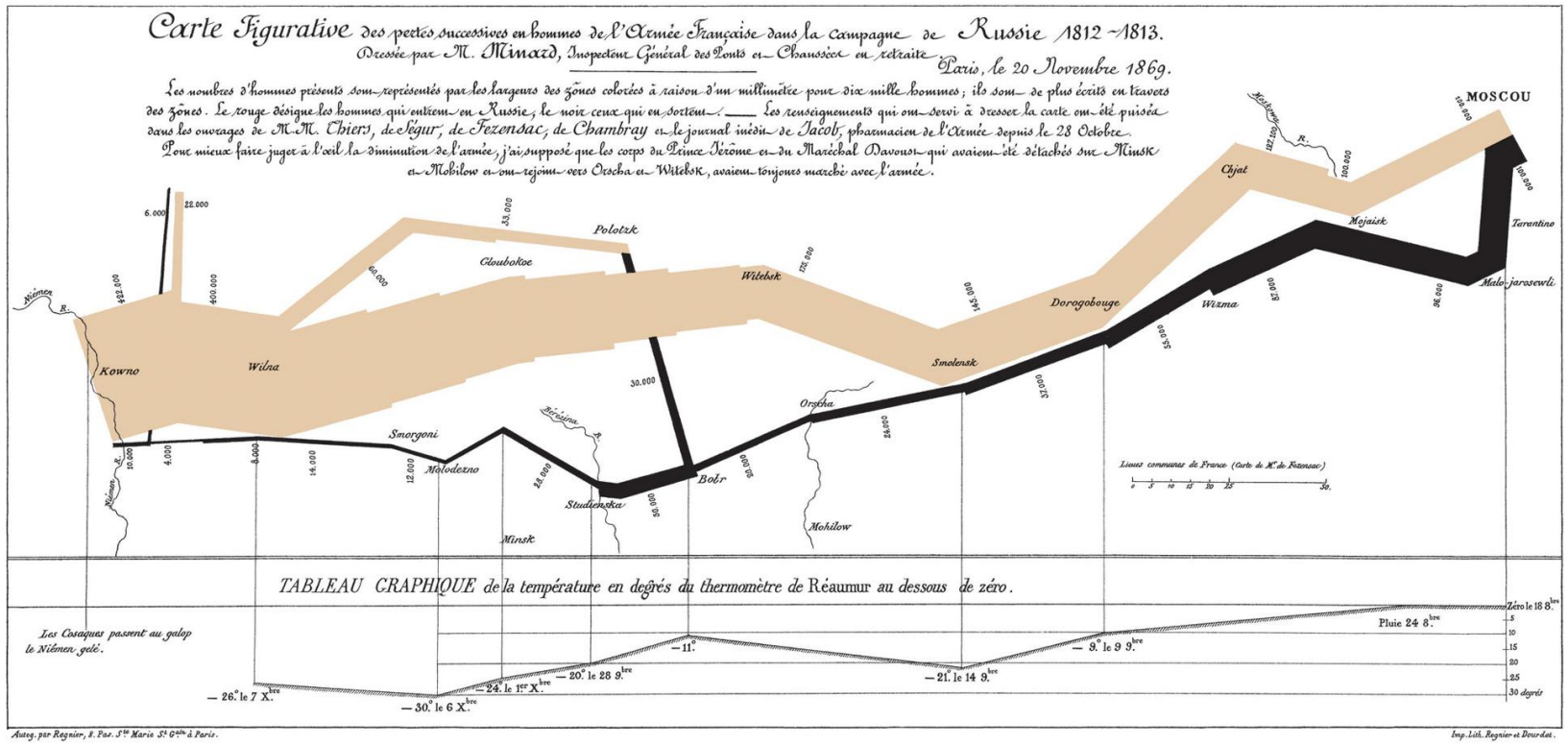
Newcastle University

From Newcastle. For the world.

Introduction

Definition

Information visualization is the study of (interactive) **visual representations** of data to reinforce **human cognition**.



Charles Minard's map of Napoleon's disastrous Russian campaign of 1812

For exploration purpose

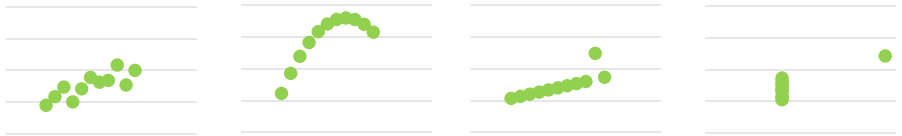
Visualization can be used for **exploration** purpose.

John Snow's map of London displaying the cholera cases (black rectangle) and pumps (red circle).

To access data behaviour

Visualizations can be designed to **access data behaviour** and **confirm information**.

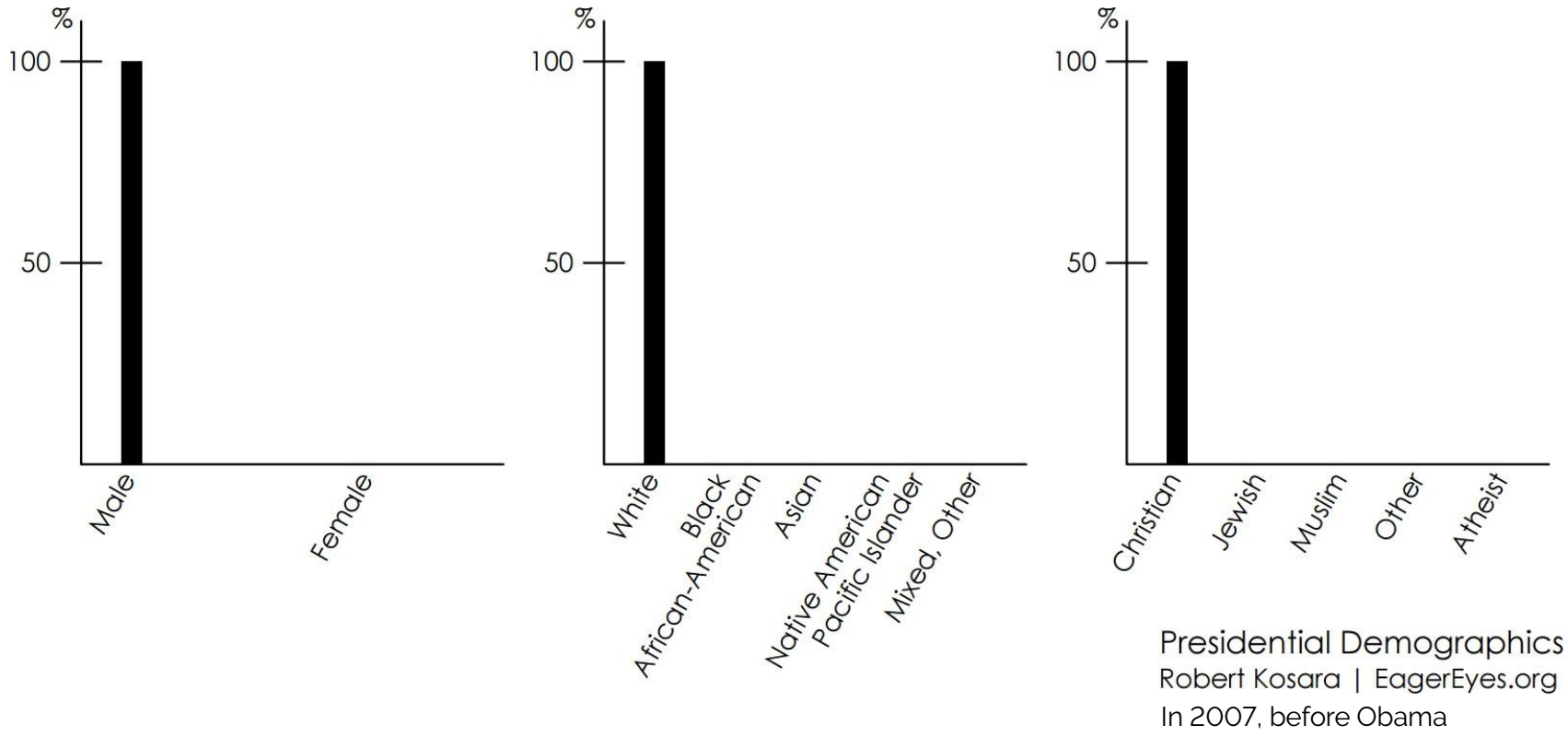
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
mean	9	11	9	11	9	11	9	11
variance	7.5	4.125	7.5	4.125	7.5	4.125	7.5	4.125
correlation	0.816		0.816		0.816		0.816	



The Anscombe's quartet

To communicate information

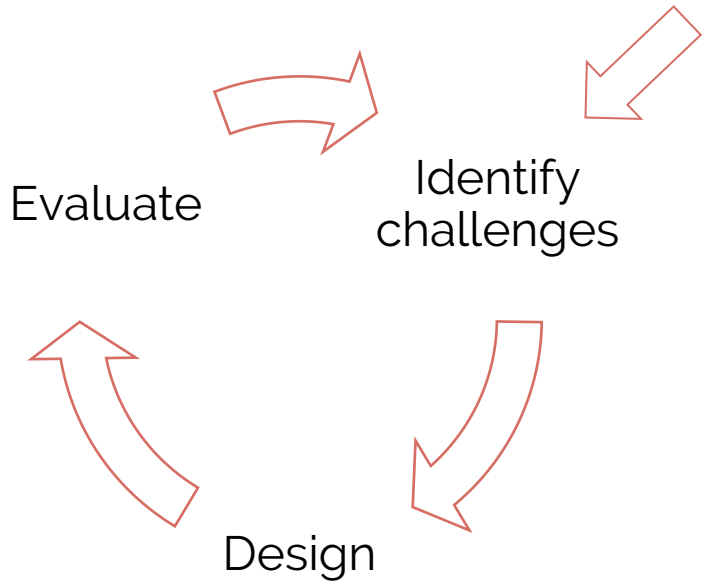
Visualizations can be designed to **communicate** describing and explaining ideas.



Design

Iterative Design Loop

Iterative design is a design methodology based on a **cyclic process** of prototyping, testing, analyzing, and refining a product or process.



Identify/refine challenges

- Identity what is the **data** that you want to be represented
- Identify who are the **users** and what is the **context**
- Identify users' **needs**

Design a solution

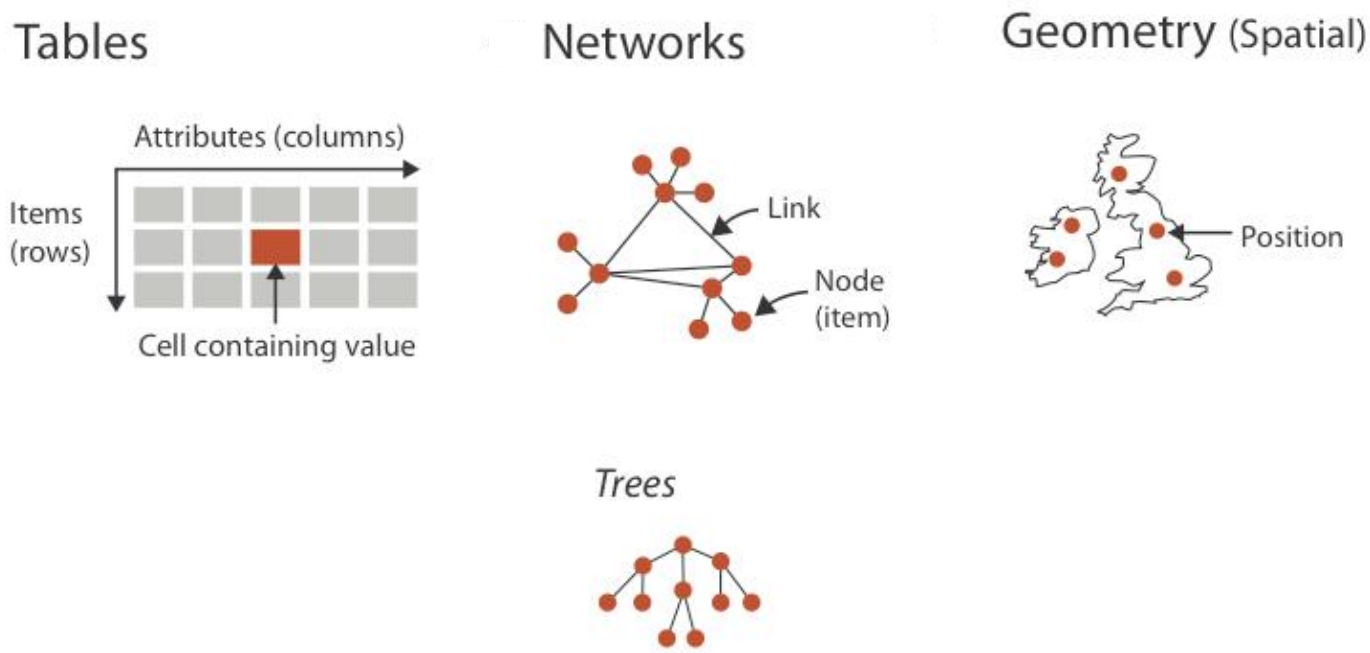
- Create or refine a solution answering the challenges identified previously

Evaluate

- Evaluate how your solution solves your challenges

Datasets

Datasets can be described based on the data it contains: **Items**, **Attributes**, **Links**, **Positions**, **Grids**. The intrinsic characteristics of these elements will heavily impact the representation.



Attributes

Attributes or **data characteristics** define the properties of an item. The representation method needs to be selected adequately to ensure that the properties are accurately restituted.

Categorical

Examples: gender, vehicle type, country, animal species

Ordered

→ *Ordinal*

Examples: education level, months, rating

→ *Quantitative*

Example: most of attributes with numerical values

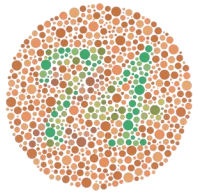
Users

Humans come in different shapes and sizes.

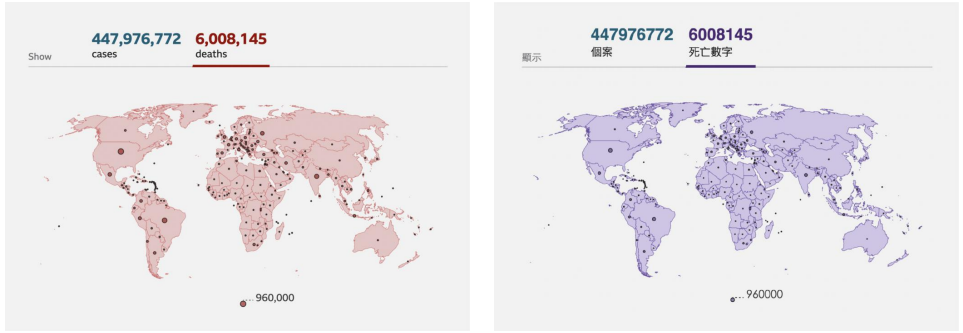
› How do they **perceive** information?
Examples: colourblind, low acuity

› How do they **interpret** information?
Examples: cultural background (particularity on icons and colours), emotional state

› How to they acquire **knowledge**?
Examples: expert/novice in the application a field, expert/novice in data visualization



1 in 12 men and 1 in 200 women cannot read the number 74 on this picture



English (in red) and Chinese (in purple) versions of the map in the BBC COVID dashboard that shows the number of deaths.

Context

Designing a system is also about factoring in the context and environment in which the system is accessed. Such aspects can have a significant impact on the end-user experience.

- **Device**

Examples: Size of the screen, type of interaction, user position

- **Psycological context**

Examples: stressful environment, high workload, steep learning curve

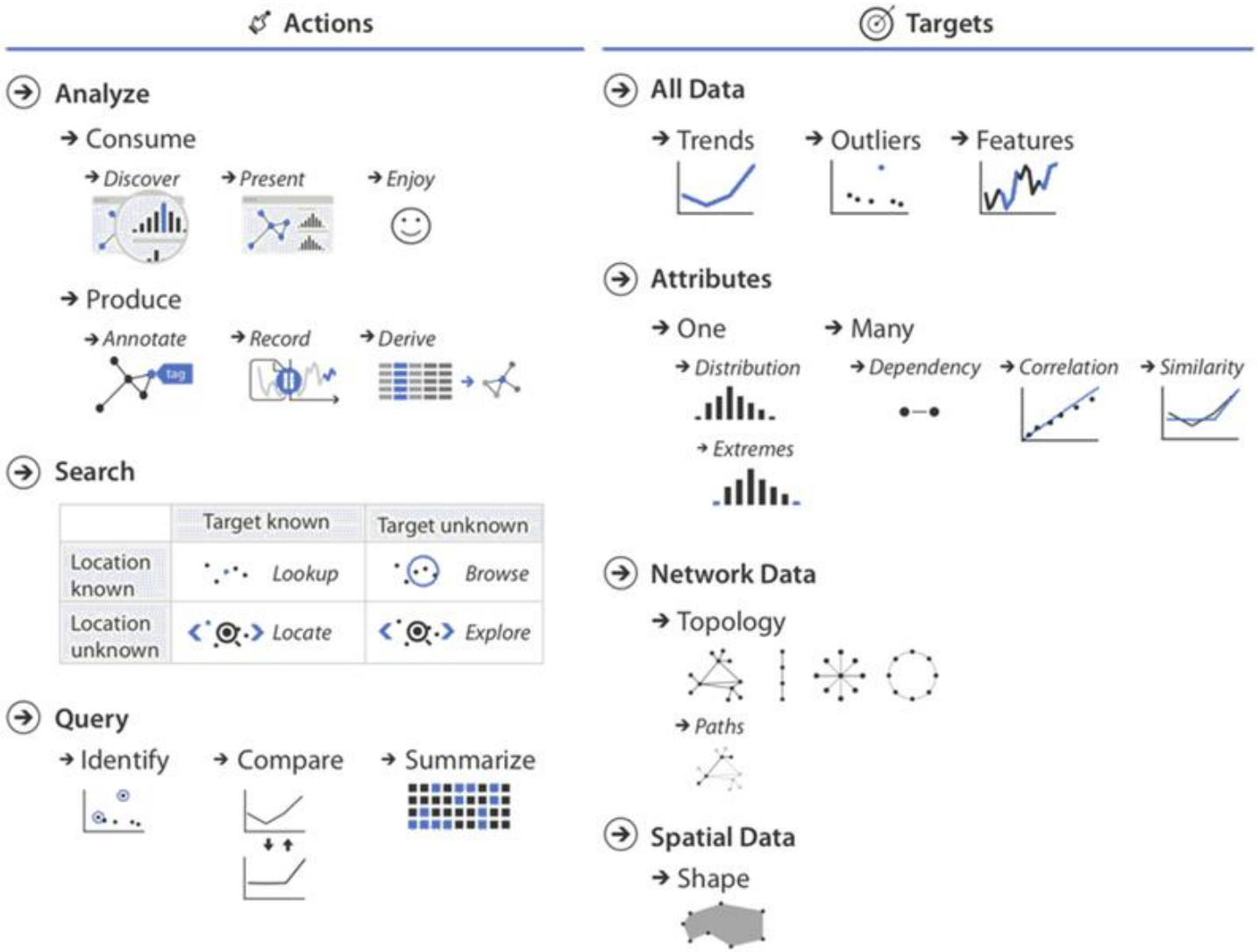
- **Surroundings**

Examples: noisy, bright, crowded

Interaction with an advertising panel in a crowded place

Military operators on a mission

Task



Munzner (2014). *Visualization analysis and design*. CRC press.

Design a solution

Once we thoughtfully identify/refine the challenge that we wish to tackle, we need to propose an adequate solution. This includes:

- Relying on **existing knowledge/heuristics**
- Creating a **prototype**
- Involving **users**

A prototype is a draft version of a product that allows you to **explore your ideas**. Depending on the state of the process, that prototype can be of various **fidelity**.

The lowest form of
prototyping: pen and paper

Encoding data

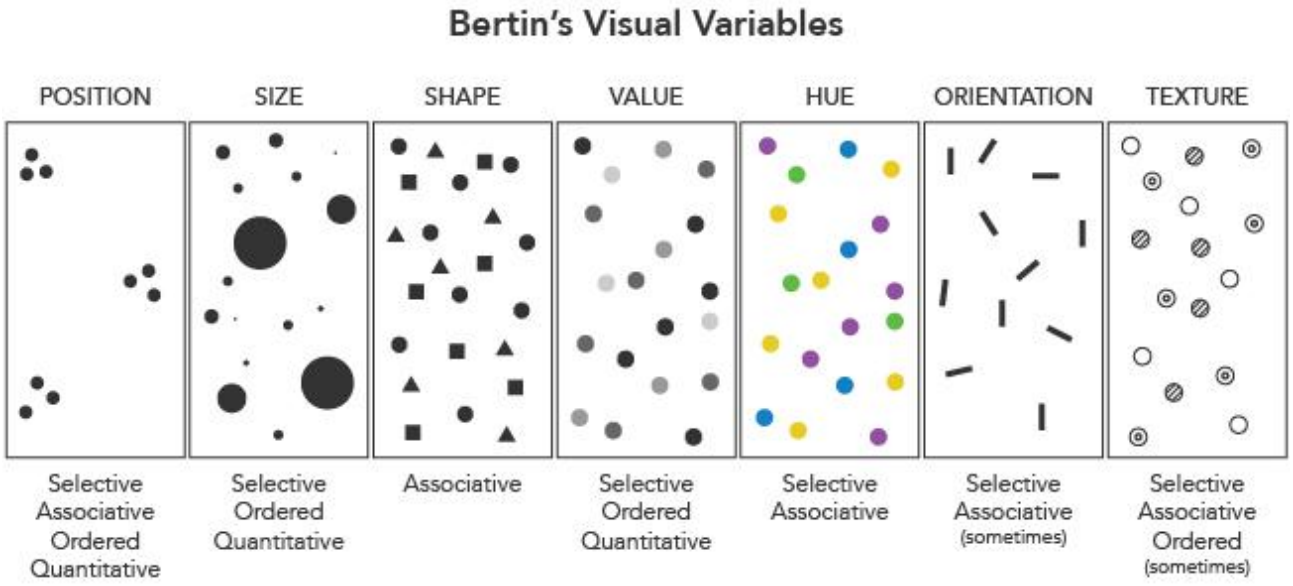
Bertin in 1968, followed up by Mackinlay in 1986, identified a set of rules that define which **visual variables** to use to relate specific **type of information**.

Visual variables

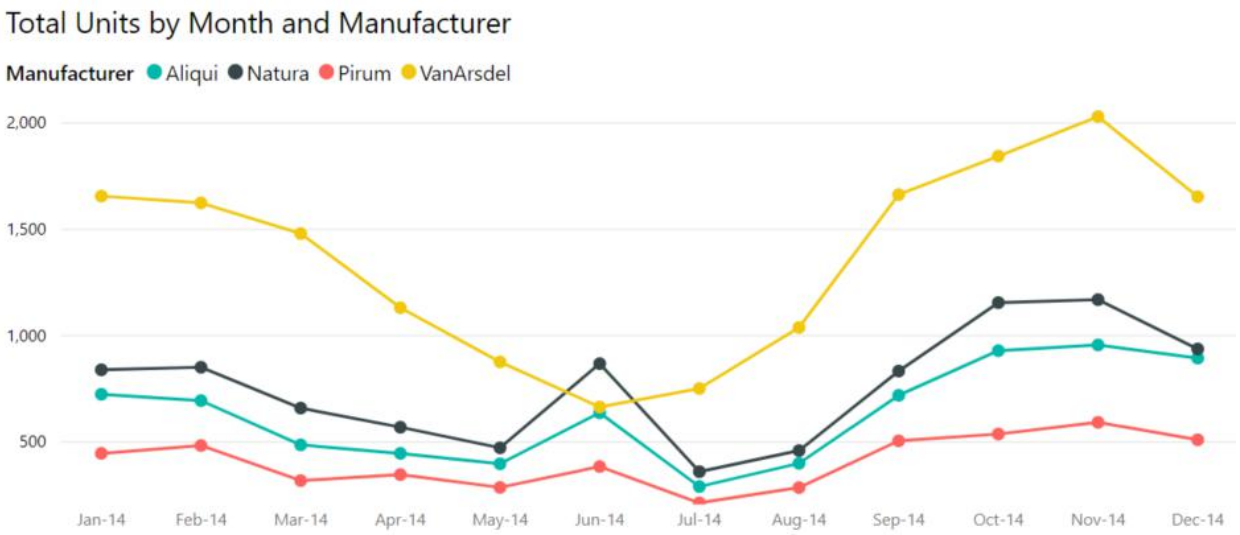
› Position, Size, Shape, Value, Hue, Orientation, Texture,…

Type of information

› Selective
› Associative
› Ordered
› Quantitative

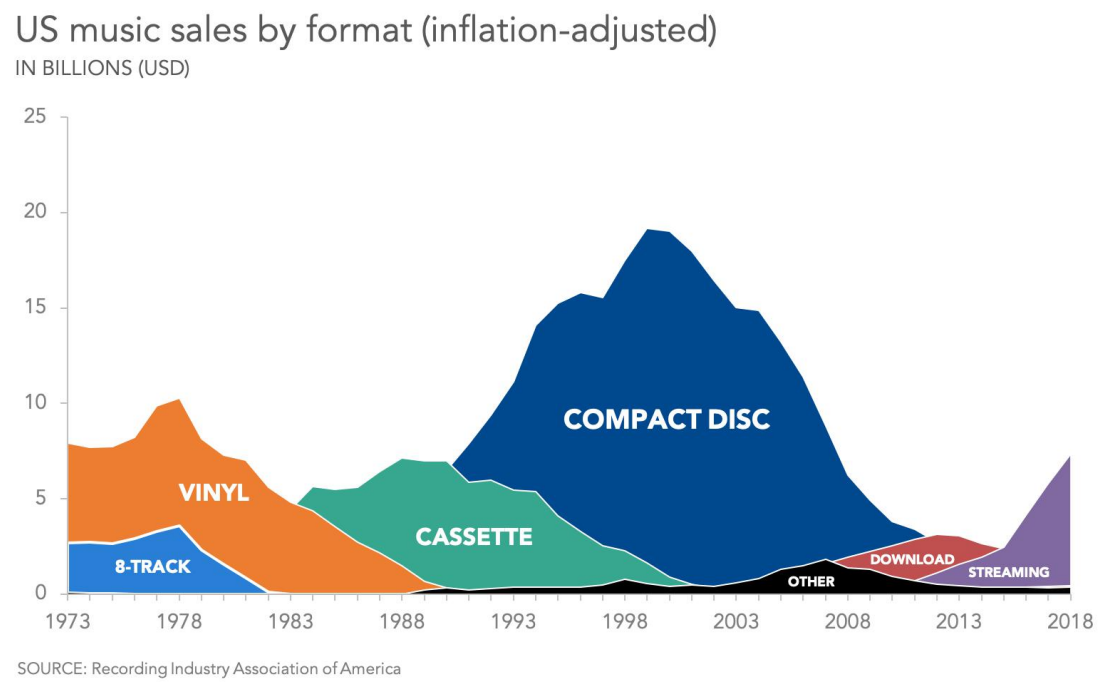


Multiple Lines Chart



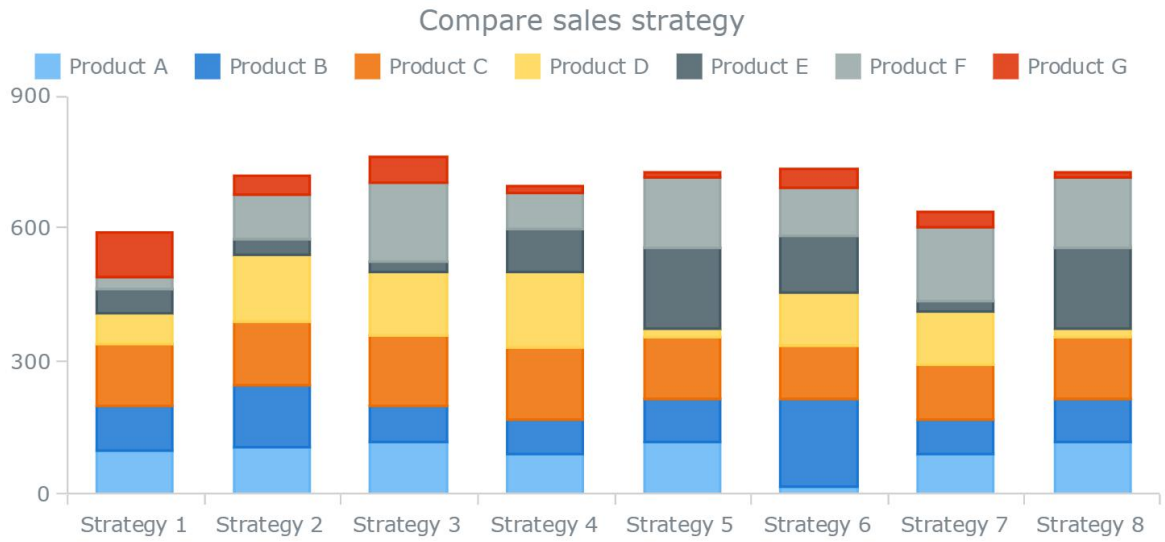
Visual Variables	Preferred Data Type
Position (x2)	Quantitative
Colour	Categorical

Stacked Area Chart



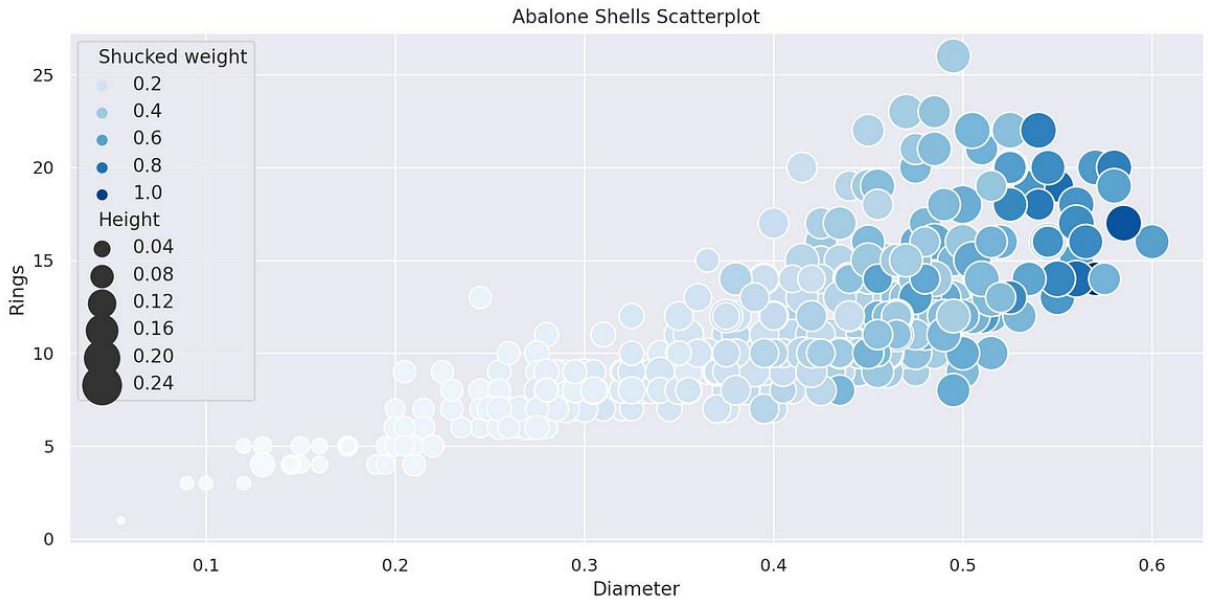
Visual Variables	Preferred Data Type
Position	Quantitative
Size	Quantitative
Colour	Categorical

Stacked Bar Chart



Visual Variables	Preferred Data Type
Position	Ordinal
Size	Quantitative
Colour	Categorical

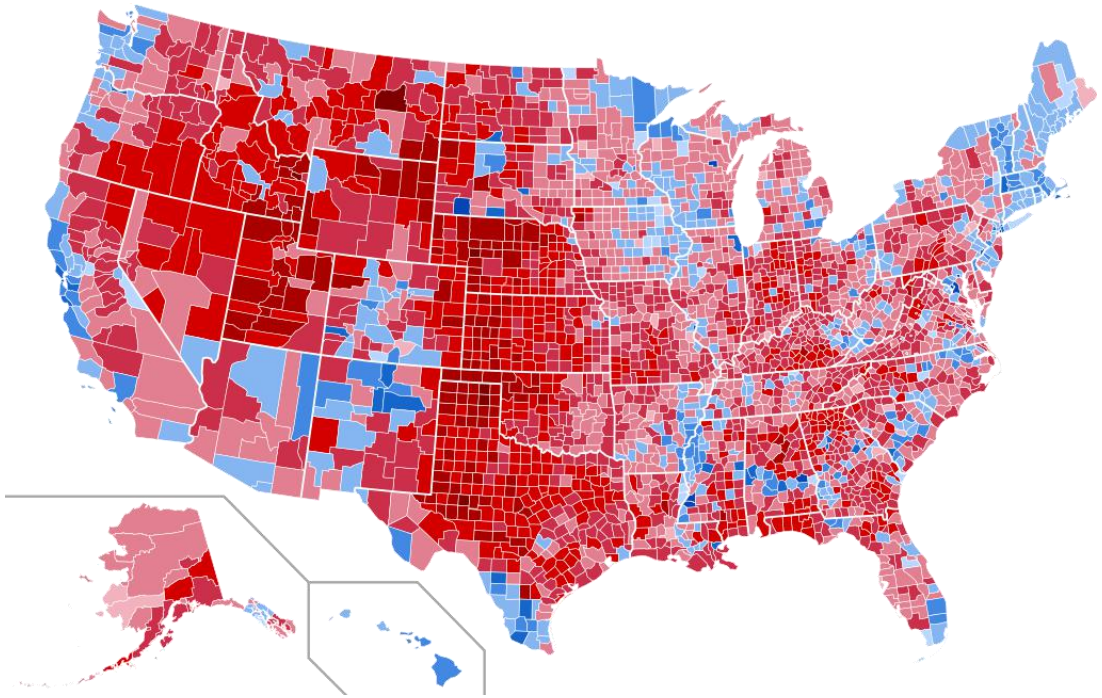
Scatterplot with colour value



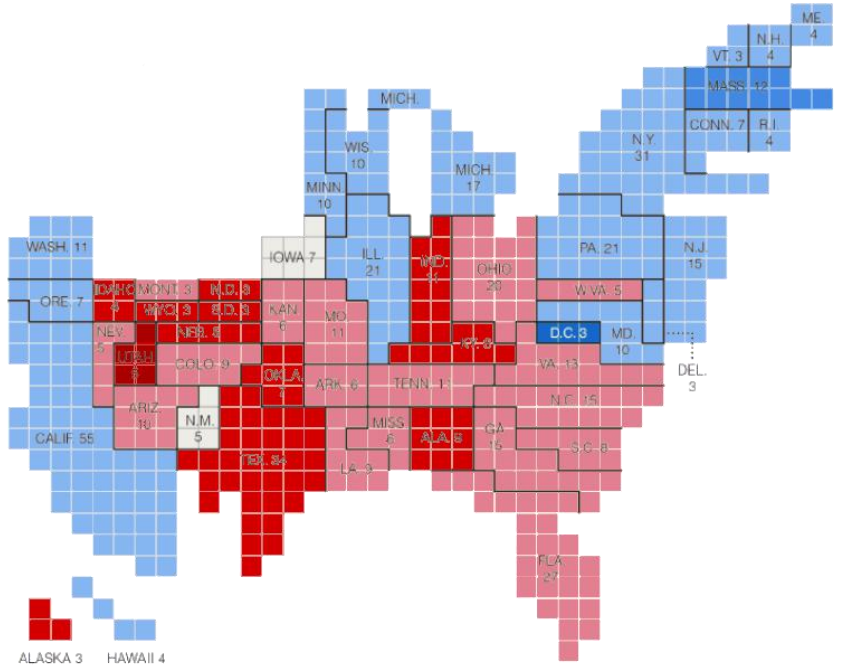
Visual Variables	Preferred Data Type
Position (x2)	Quantitative
Colour	Ordinal
Size	Quantitative

Size biases

To be displayed, colours **rely on other visual variable** like size, or texture. These additional variables have their own perceptive properties that might **bias the representation**.



2004 United States Presidential Election results.
Data: G. Bush: 50.7% J. Kerry: 48.3%



2004 United States Presidential Election results.
Size of the state weighted according of the number of electoral votes

Evaluation

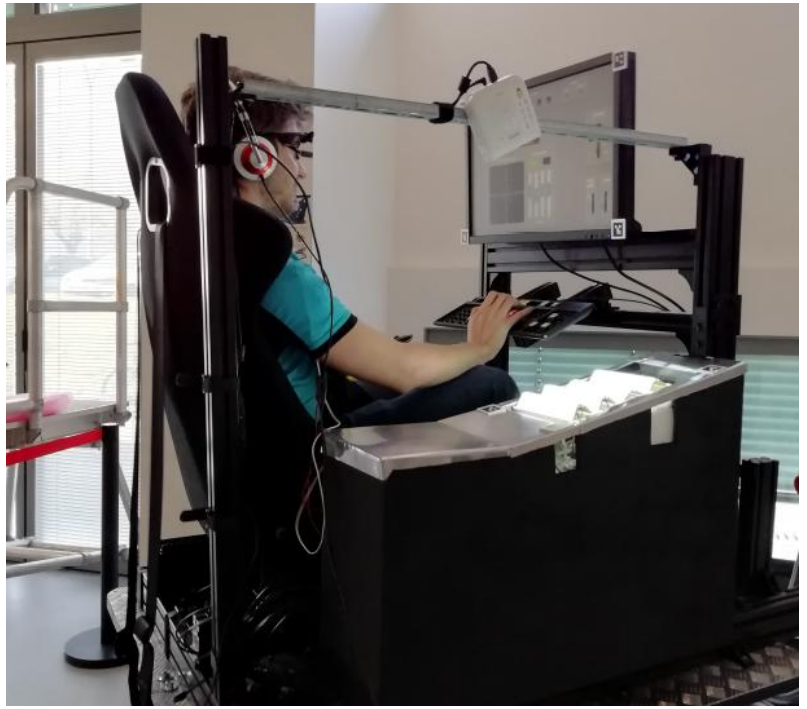
Once a prototype is finalized, it is time to evaluate it. There are different ways to evaluate a prototype.

Inspection Methods: Inspection methods rely solely on **heuristics** and do not require direct user involvement.

User Studies: User studies are a way to evaluate the usability of a tool by presenting it to **users**.

- **Quantitative** user studies seek to measure **tangible metrics**, mainly usability (effectiveness, efficiency, and satisfaction), via objective methods (time to complete, error rate).

- **Qualitative** user studies aims to evaluate if an interface meets user requirements, by gathering **subjective** users' feedback.



User study evaluation under turbulence constraints

Application to Mobility Tracking

Mobility as a Symptom

Multiple Sclerosis (MS)

- limit **walking endurance**
- cause **weakness**, **poor balance**, **numbness**, or **spasticity**

Chronic Obstructive Pulmonary Disease (COPD)

- increase **duration between steps**
- increase **step duration variability**

Parkinson's Disease (PD)

- gait **slowness**
- increased **step variability**
- **poor postural control**

Chronic Heart Failure (CHF)

- can reduce **walking capacity**

Periprosthetic Femoral Fracture (PFF)

- affect **gait** and **balance** functions

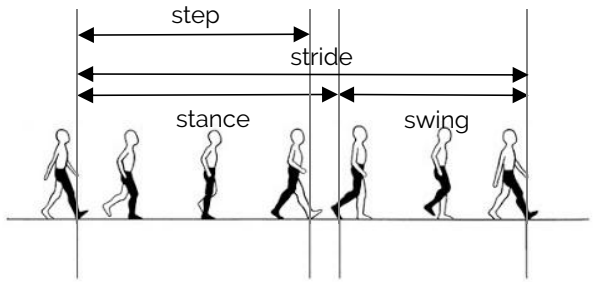
Mobility Tracking

Mobility tracking has evolved so that we now possess tools to **accurately access** every step (and more) that someone is taking.

2.12 WALKING AND BALANCE		SCORE
Over the past week, have you usually had problems with balance and walking?		
0: Normal:	Not at all (no problems).	
1: Slight:	I am slightly slow or may drag a leg. I never use a walking aid.	
2: Mild:	I occasionally use a walking aid, but I do not need any help from another person.	☐
3: Moderate:	I usually use a walking aid (cane, walker) to walk safely without falling. However, I do not usually need the support of another person.	
4: Severe:	I usually use the support of another persons to walk safely without falling.	

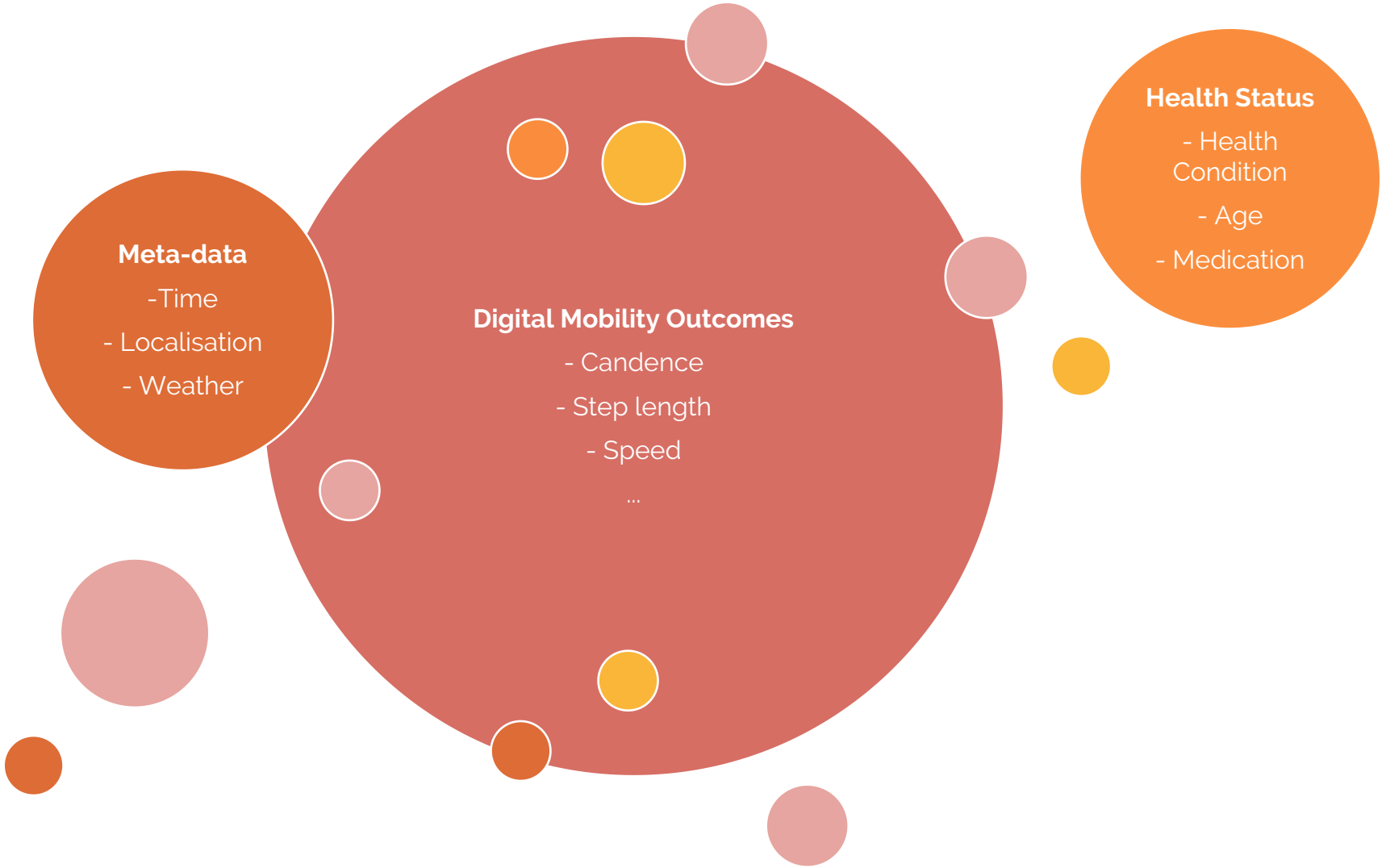
mobility measurement questionnaire

lower-back accelerometry sensor



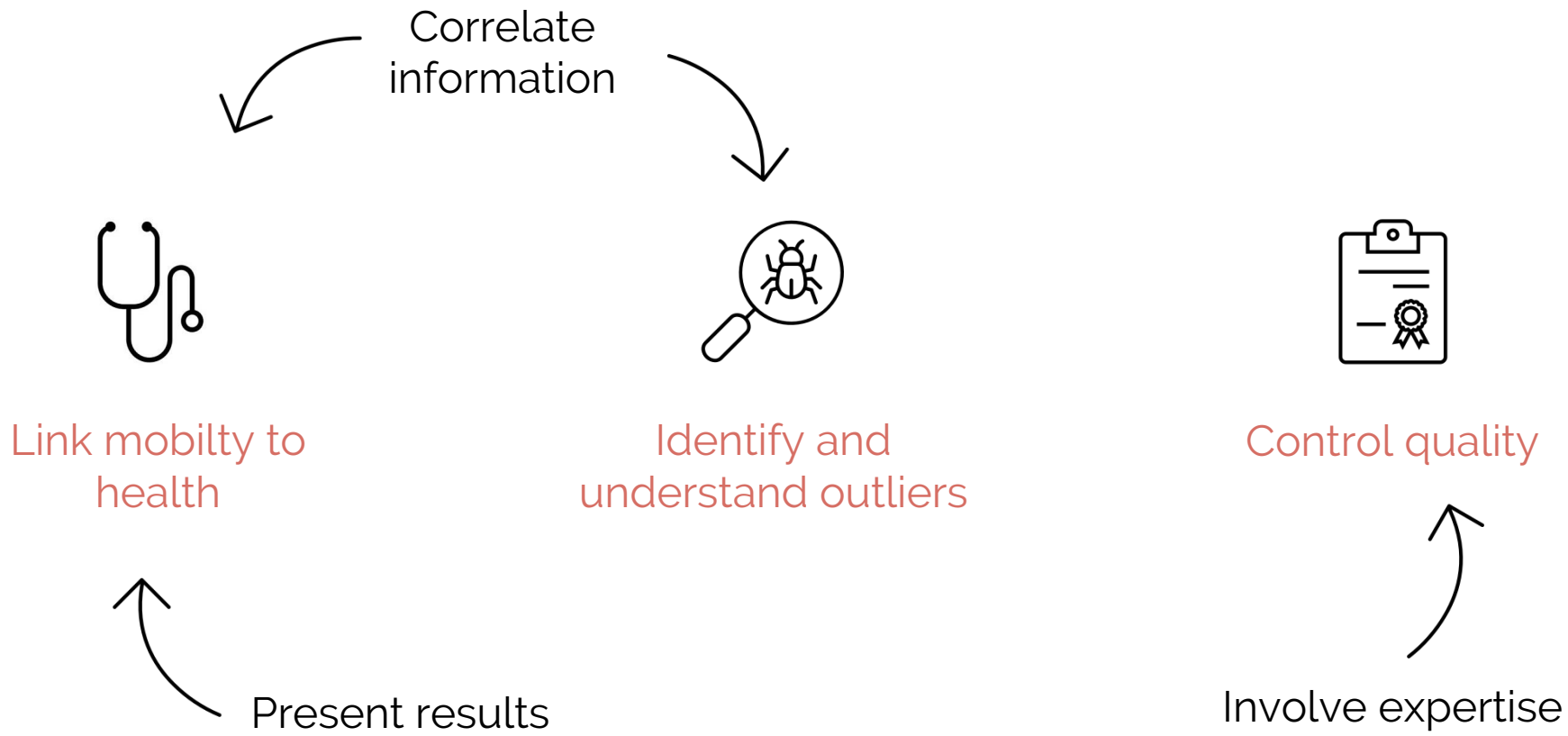
digital mobility outcome (DMO)

Data



The amount of data is not only **colossal**, data are **multidimensional**, **heterogeneous** (categorical and quantitative), subject to **missingness**, and somewhat **hierarchical**.

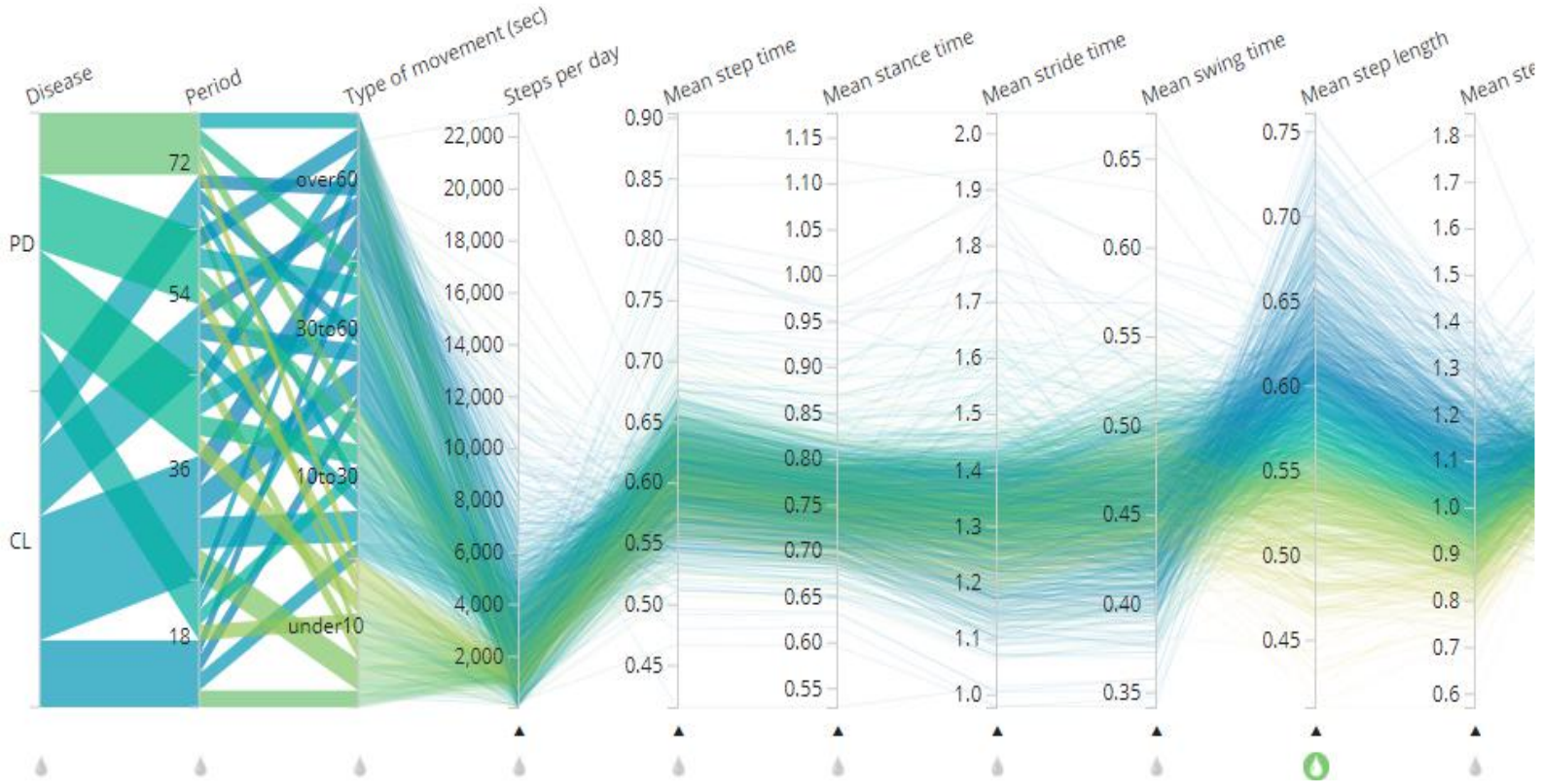
Tasks


Correlate information
Link mobilty to health
Identify and understand outliers
Control quality
Present results
Involve expertise

Our solution

Changing axis order

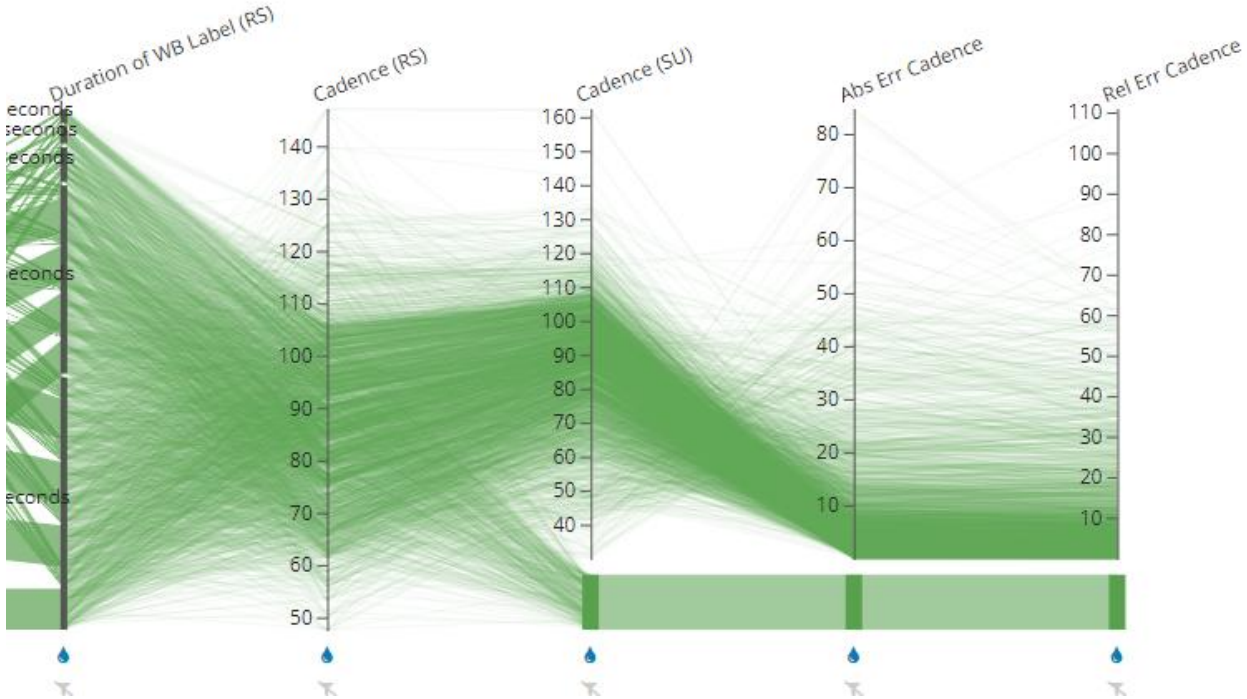
Changing axis orientation



Filtering

Colour according to an axis
(categorical of quantitative)

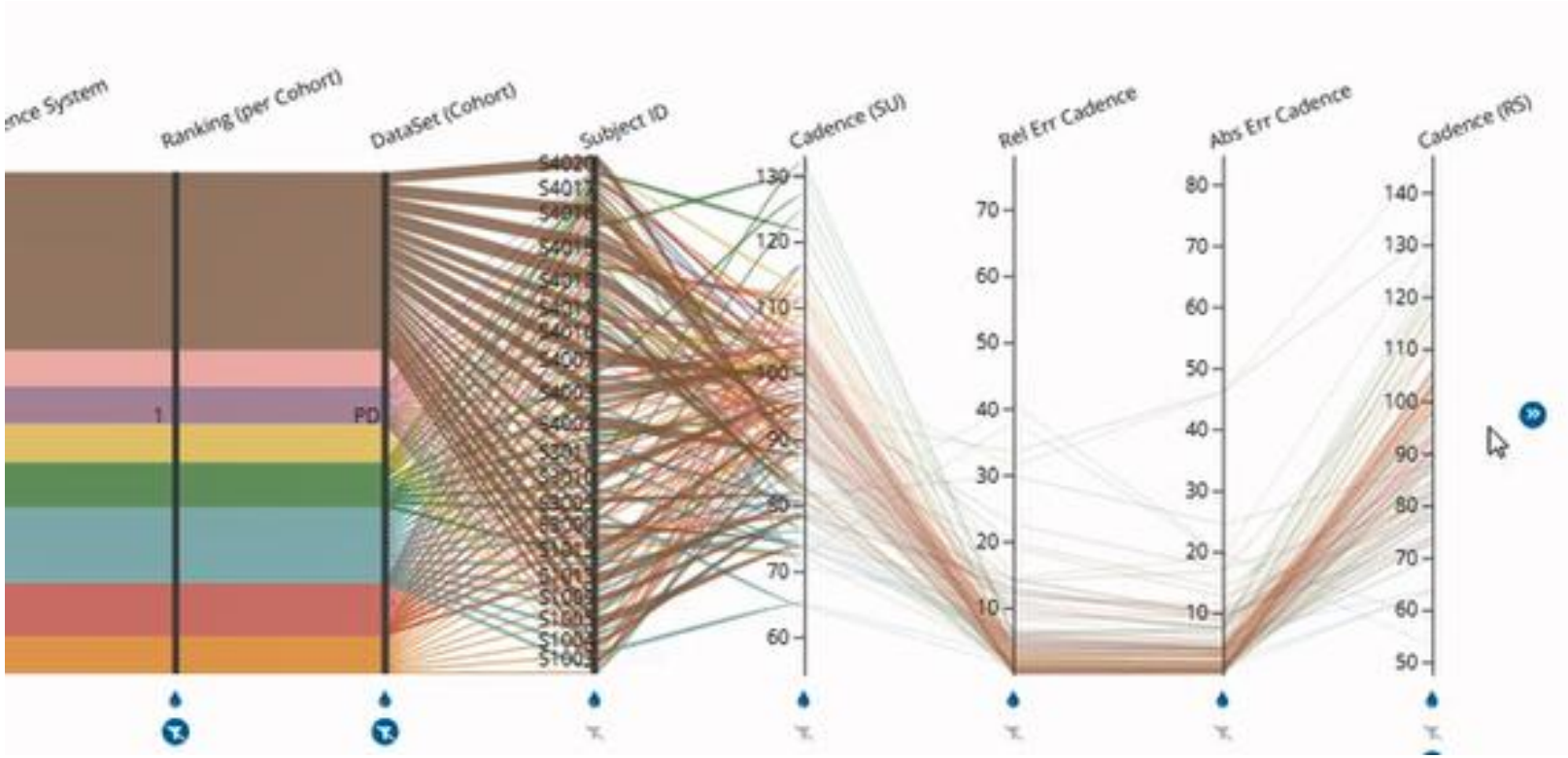
Representing missing values



Colouring the missing ribbon with mean value

Relying on a ribbon to display frequency

Aggregating data



Merging each ribbon into a line to access higher level of information

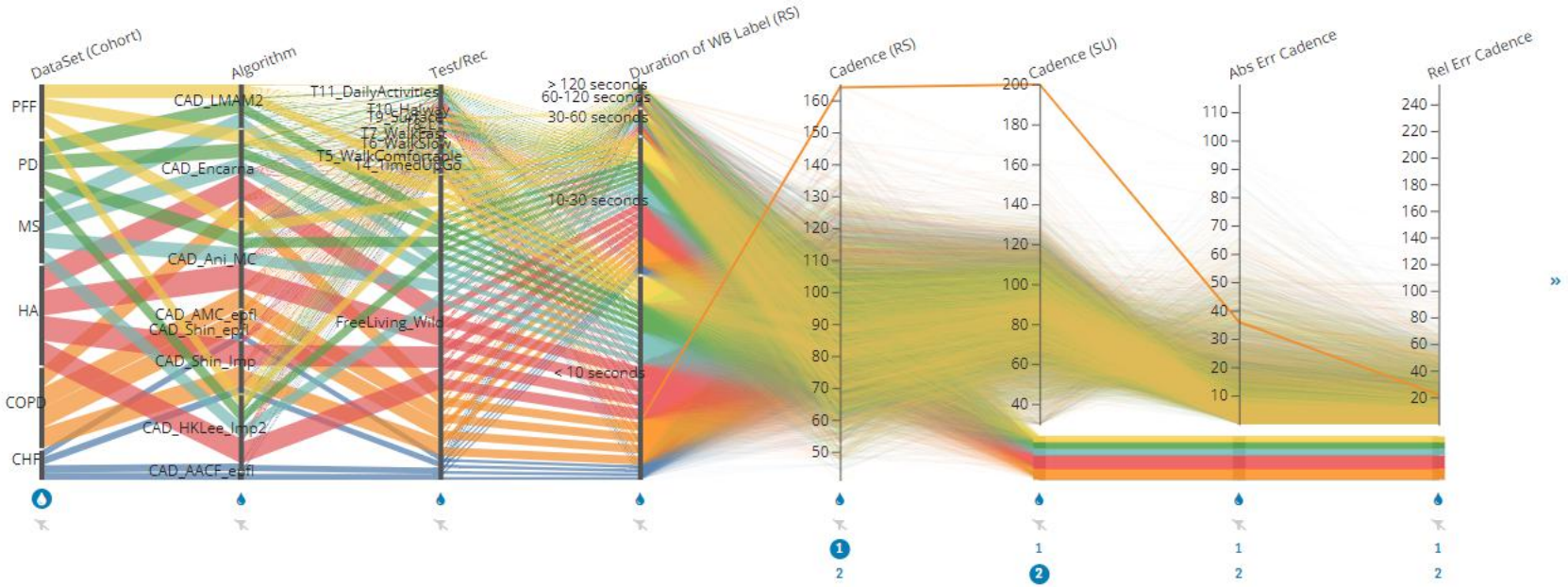
TVS Data Explorer: CAD_Matrix.csv

Legend

Coloration:

Visible data: 25968 / 25968

Overview



Visible Variables

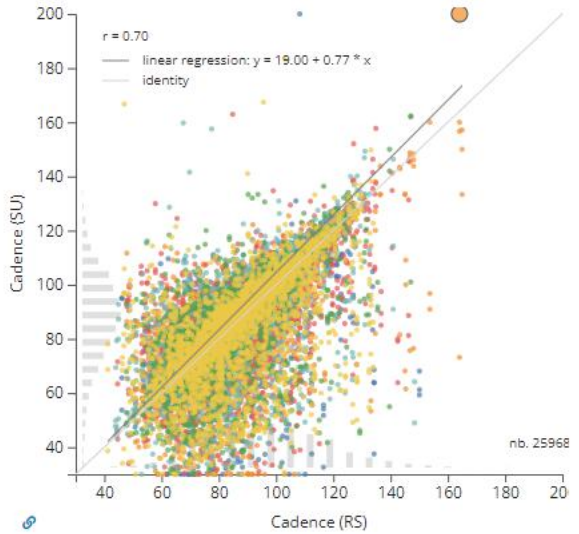
- [ ] % Perform WB (Algo)
- [ ] % Perform Subjects (Algo)
- [ ] Ranking (per Cohort)
- [x] Algorithm
- [ ] Subject ID
- [x] DataSet (Cohort)
- [ ] Location
- [ ] Reference System
- [x] Test/Rec
- [ ] Trial
- [ ] Height
- [ ] Weight
- [ ] Sensor Height
- [ ] Handedness
- [ ] Foot Size
- [ ] Walking Aid (0/1)
- [ ] Walking Aid Side
- [ ] Walking Aid Type
- [ ] Participant
- [ ] Cohort
- [ ] Site
- [ ] Age

Summary

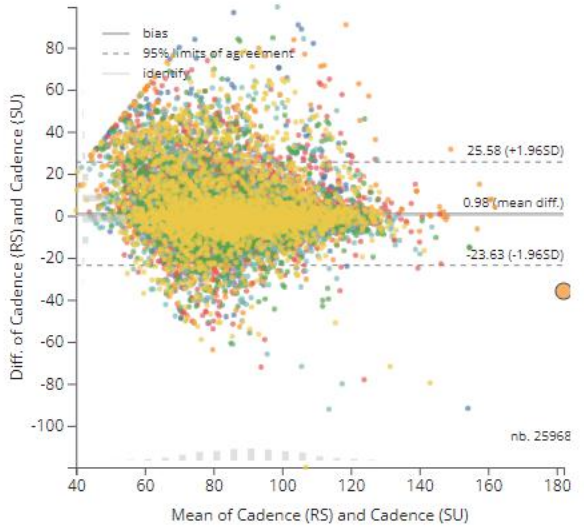
Metrics	Mean	95% CI
Cadence (RS)	87.49	[87.31, 87.68]
Cadence (SU)	86.67	[86.45, 86.88]
Abs Err Cadence	7.36	[7.23, 7.49]
Rel Err Cadence	8.92	[8.75, 9.08]

Export

Correlation Plot



Bland-Altman Plot



Selected Datum

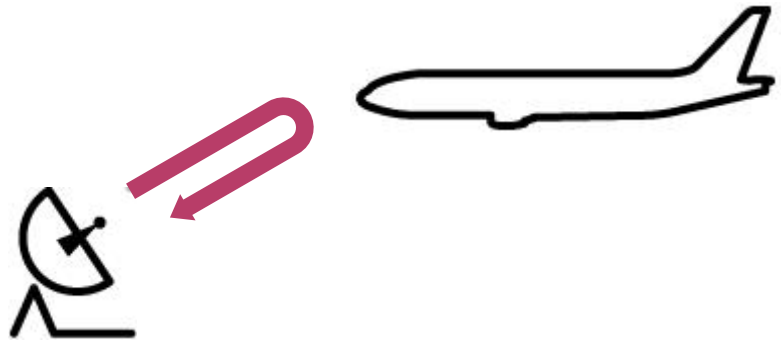
% Perform WB (Algo):	100.00
% Perform Subjects (Algo):	100.00
Ranking (per Cohort):	23
Algorithm:	CAD_Enc
Subject ID:	S1020
DataSet (Cohort):	COPD
Location:	Laborato
Reference System:	Stereoph
Test/Rec:	T7_Walk
Trial:	Trial1
Height:	157.00
Weight:	60.00
Sensor Height:	96.20
Handedness:	R
Foot Size:	24.70
Walking Aid (0/1):	0
Walking Aid Side:	none
Walking Aid Type:	none

Application to Signal Intelligence

Radar Detection

- The radar is transmitting **electromagnetique pulses**
- Pulses are reflected by a target
- The radar receives the echoes from transmitted pulses

Using various properties of the received echo, the radar can extract parameters such as the **range** and **velocity** of the target.



schema of radar detection

Nowadays radar signals are very **sophisticated** (dedicated to targeting, not detectable, ...) which makes them very complex but also very **unique**.

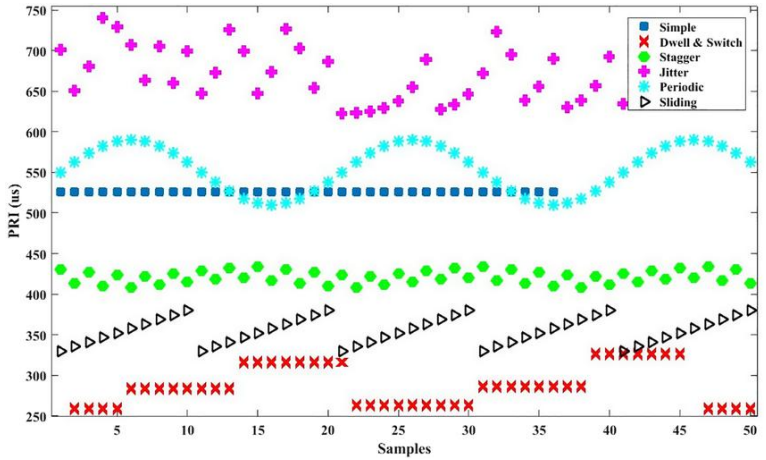
A branch of signal intelligence, called **radar intelligence**, consists of gathering intelligence about radar characteristics in order to recognise them.

Electromagnetic pulse and repetition pattern

An electromagnetic pulse is a **short surge of electromagnetic energy**. Its short duration means that it will be spread over a range of frequencies. Pulses are typically characterized by:

- The **mode of energy** transfer (radiated, electric, magnetic or conducted).
- The **range** or spectrum of frequencies present.
- Pulse waveform: **shape**, **duration** and **amplitude**.

In order to build up a discernible echo, most radar systems emit pulses continuously in **repetitive patterns**. The repetition rate of these pulses is determined by the role of the system.

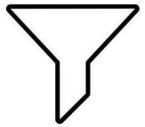


different types of pulse repetition interval modulations

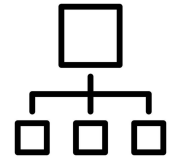
Tasks

Present results

Identify repetition patterns



Isolate signals



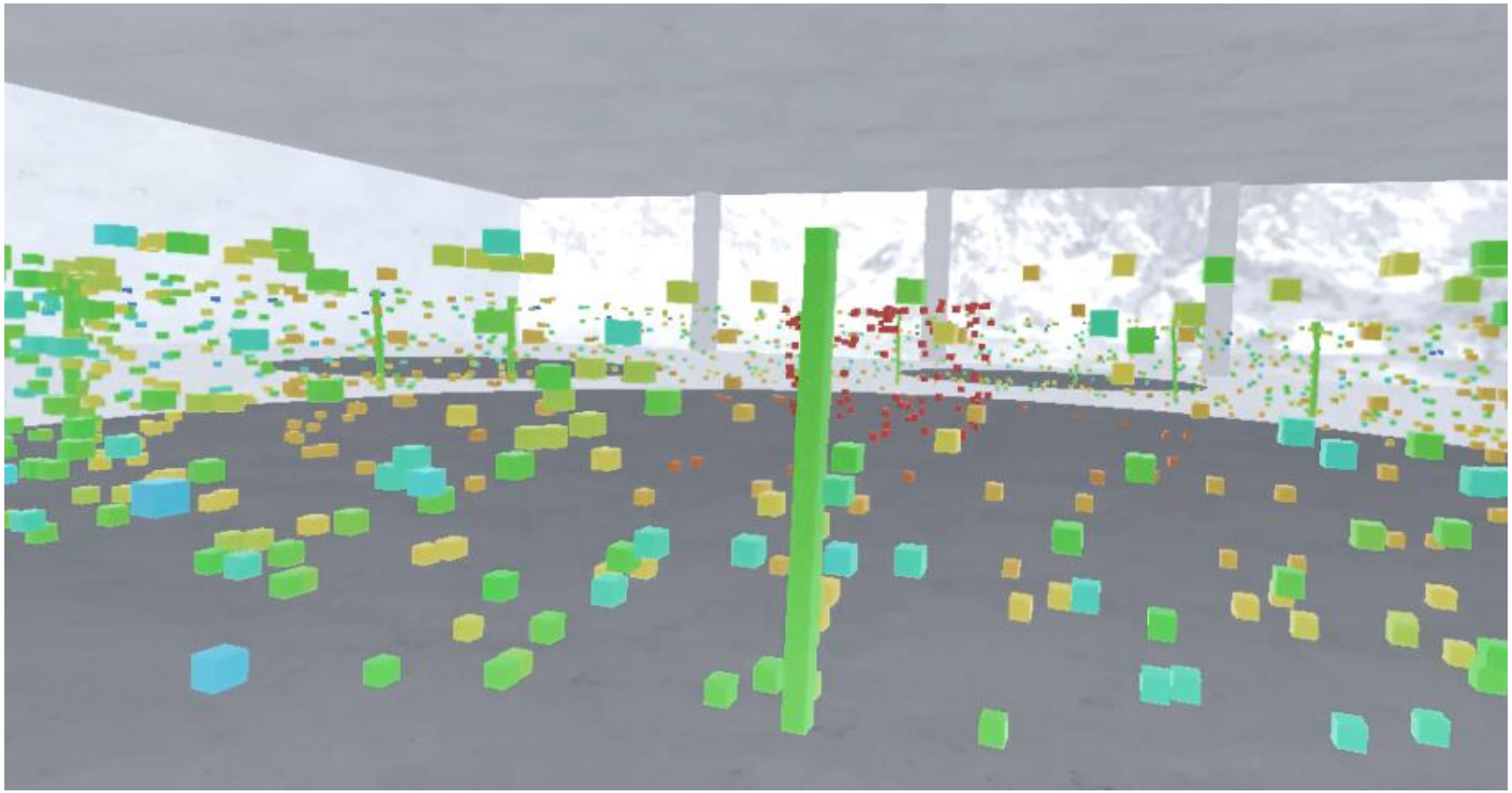
Identify radars characteristics

Correlate information

Involve expertise

Our solution

Mapping the data on a **helicoidal structure** allows users to **interact with the torsion**, relying on human perceptual capacities to **identify patterns** in an immersive environment.



Thank you for your attention

Any quesitons?