

Public Health Datathon, Insight North East

Power BI Workshop
Fundamentals of
Data Visualisation

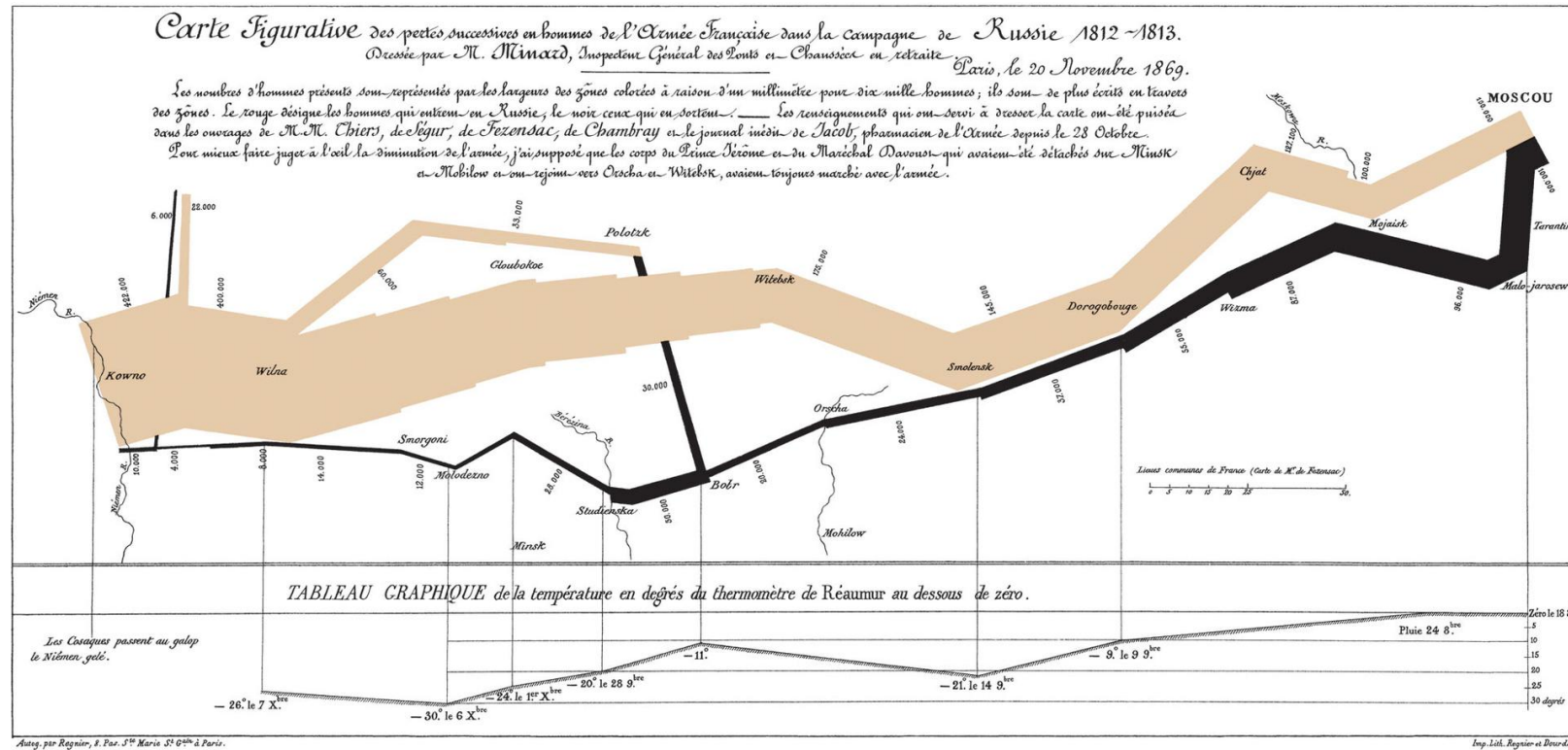
Dr Alma Cantu, Newcastle University
19th of April 2024, Farrell Center



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.



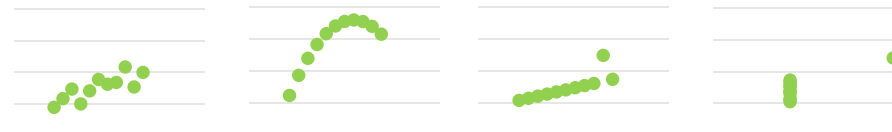
John Snow's map of London displaying the cholera cases (black rectangle) and pumps (red circle).

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

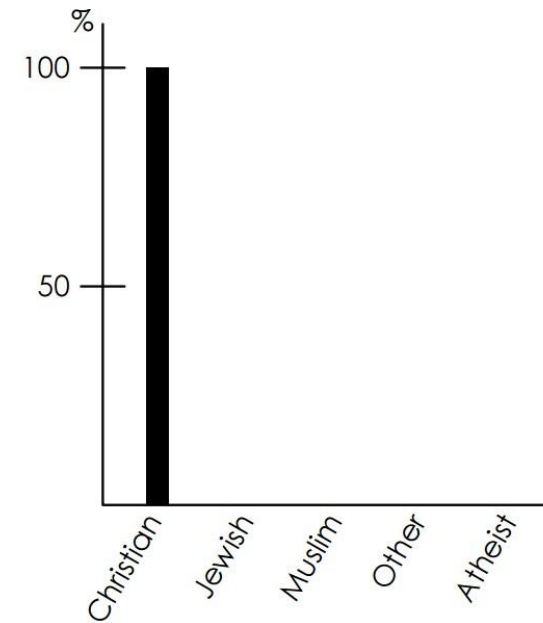
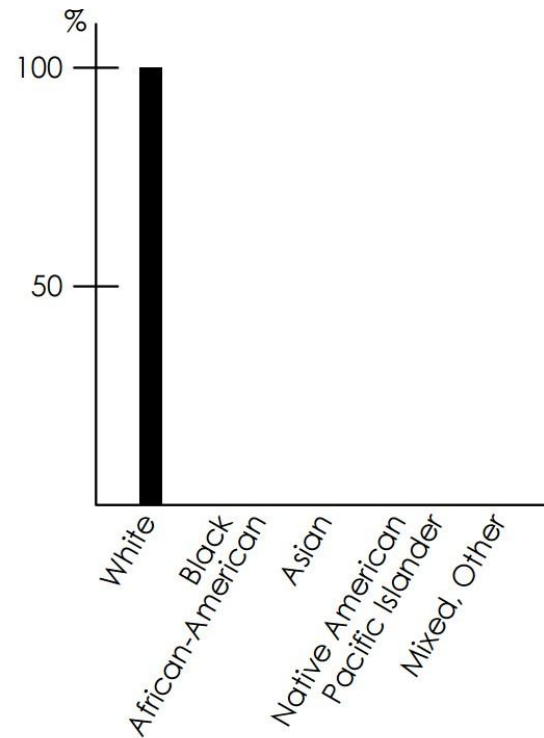
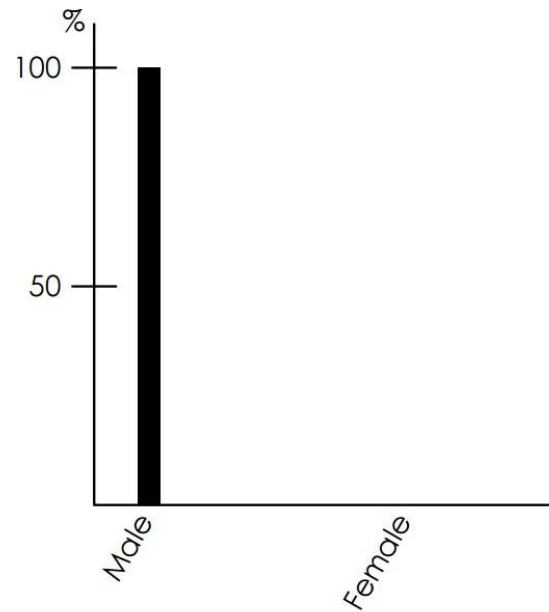
mean	9	11	9	11	9	11	9	11
variance	7.5	4.125	7.5	4.125	7.5	4.125	7.5	4.125
correlation	0.816		0.816		0.816		0.816	



The Anscombe's quartet

To communicate information

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

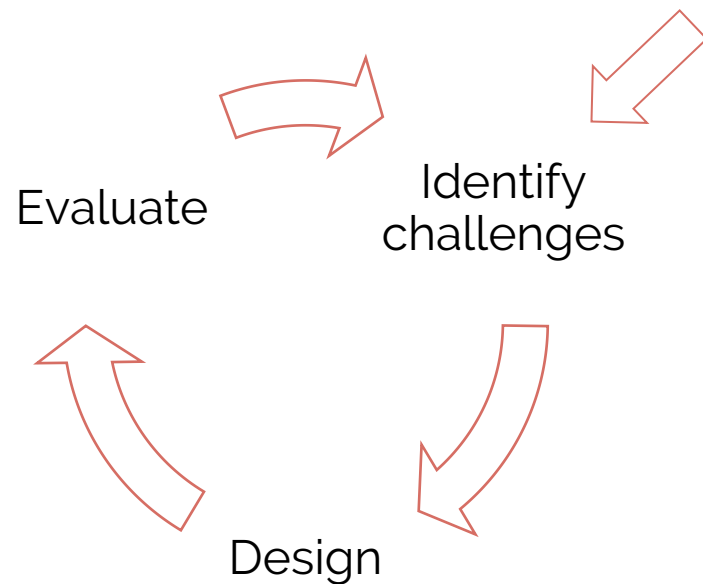


Presidential Demographics
Robert Kosara | EagerEyes.org
In 2007, before Obama

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

- > Identify 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

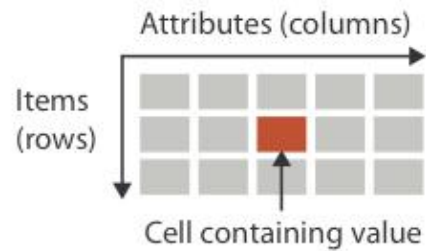
Evaluate

- > Evaluate how your solution solves your challenges

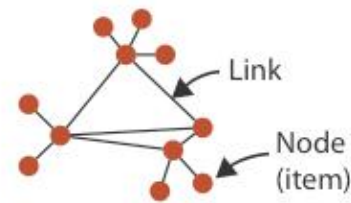
Datasets

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.

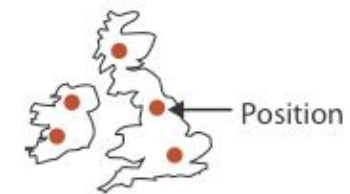
Tables



Networks



Geometry (Spatial)



Trees



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?

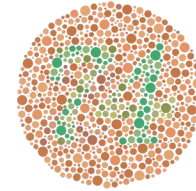
Examples: colourblind, low acuity

> How do they **interpret** information?

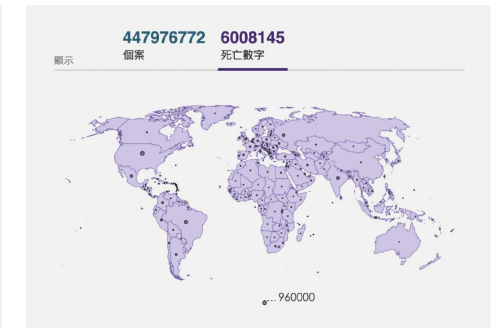
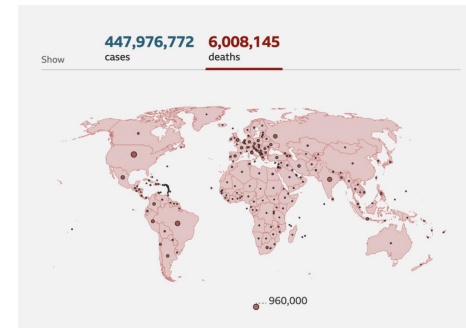
Examples: cultural background (particularity on icons and colours), emotional state

> How do 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

- > **Psychological context**

Examples: stressful environment, high workload, steep learning curve

- > **Surroundings**

Examples: noisy, bright, crowded

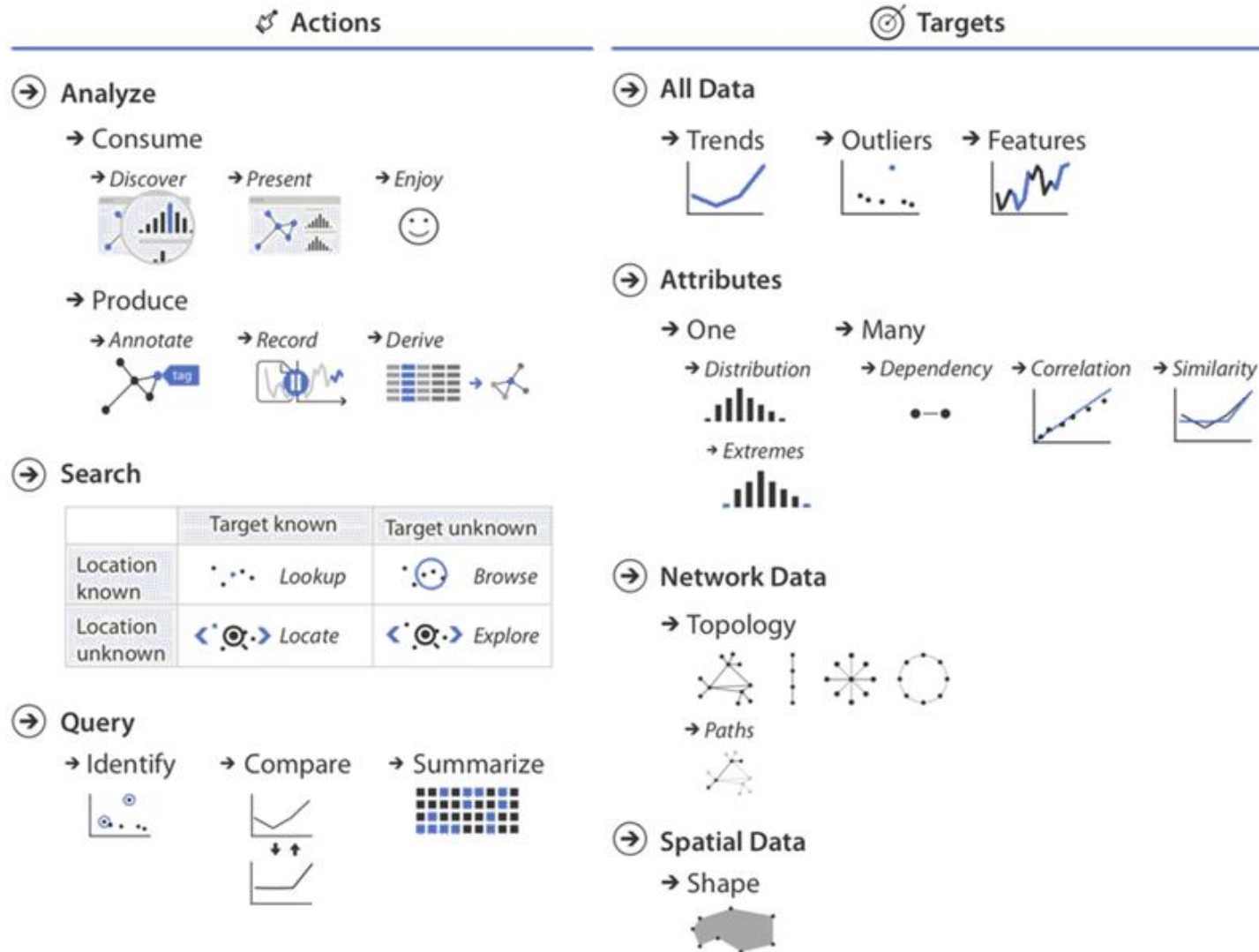


Interaction with an advertising panel in a crowded place



Military operators on a mission

Task



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

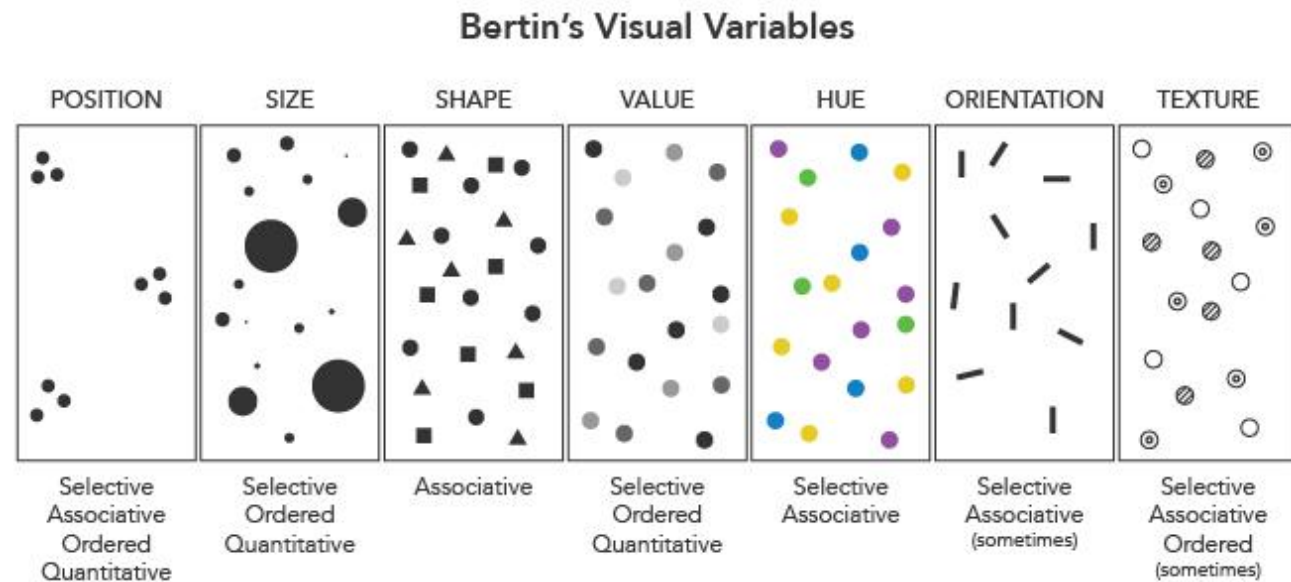
Bertin in 1968, followed up by Mackinlay in 1986, identified a set of rules that define which **visual variables** to use to relate specific **type of information**.

Visual variables

> Position, Size, Shape, Value, Hue, Orientation, Texture,...

Type of information

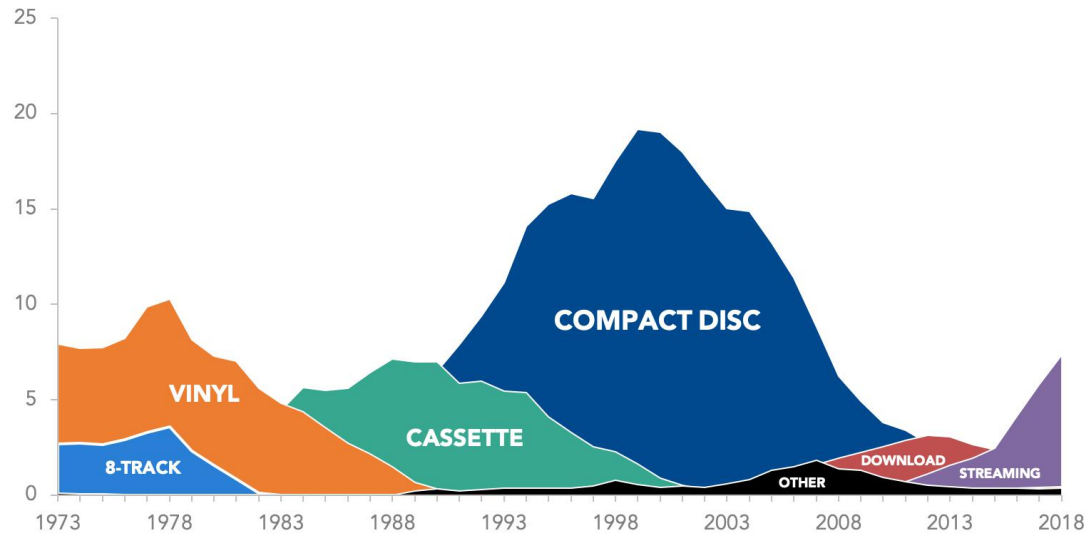
- > Selective
- > Associative
- > Ordered
- > Quantitative



Examples

Area Chart

US music sales by format (inflation-adjusted)
IN BILLIONS (USD)



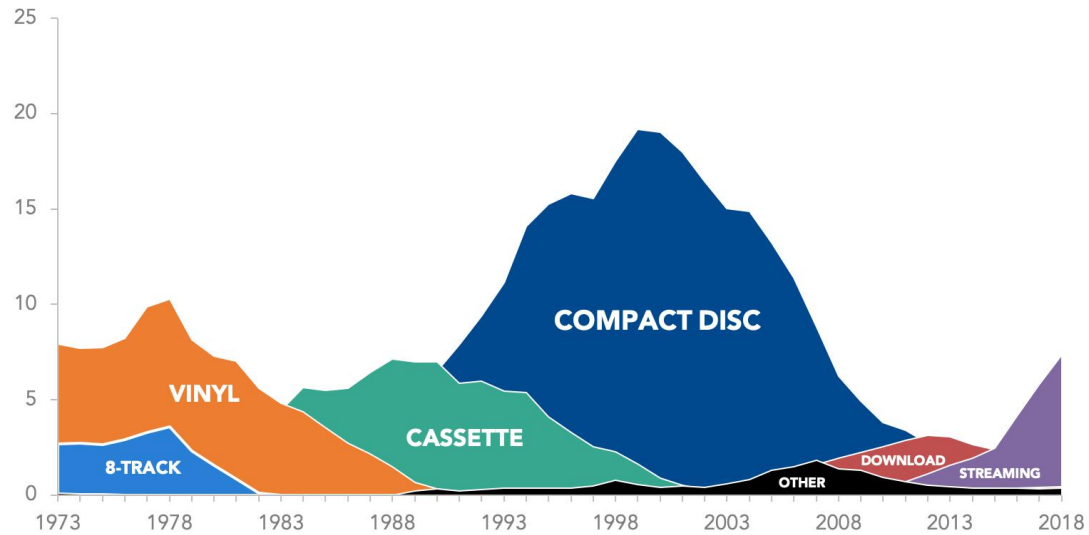
SOURCE: Recording Industry Association of America

Visual Variables	Preferred Data Type
Position	Quantitative
Size	Quantitative
Colour	Categorical

Area Chart

US music sales by format (inflation-adjusted)

IN BILLIONS (USD)

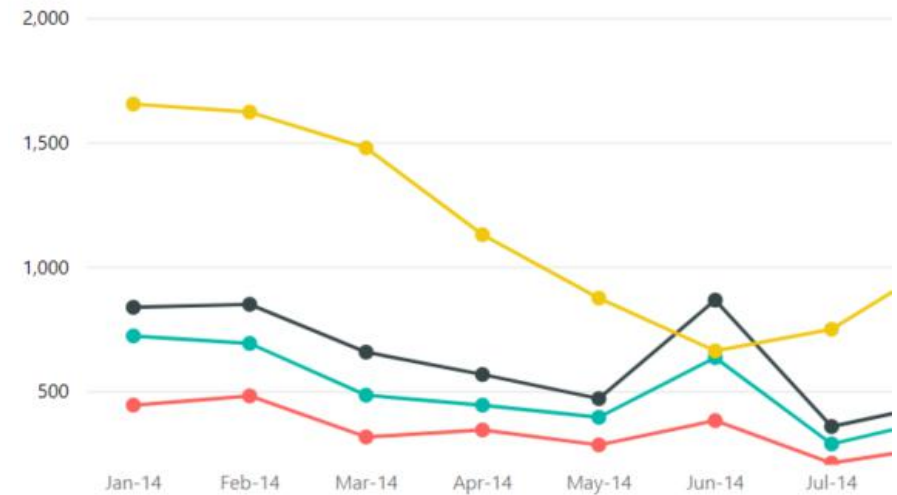


SOURCE: Recording Industry Association of America

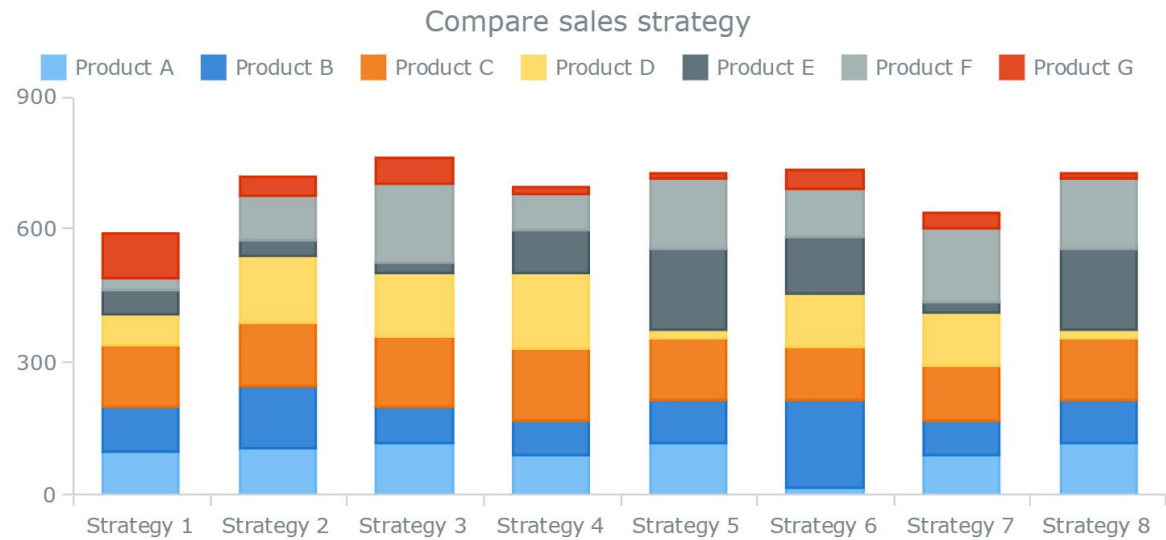


Line Chart

Manufacturer Aliqui Natura Pirum VanArsdel

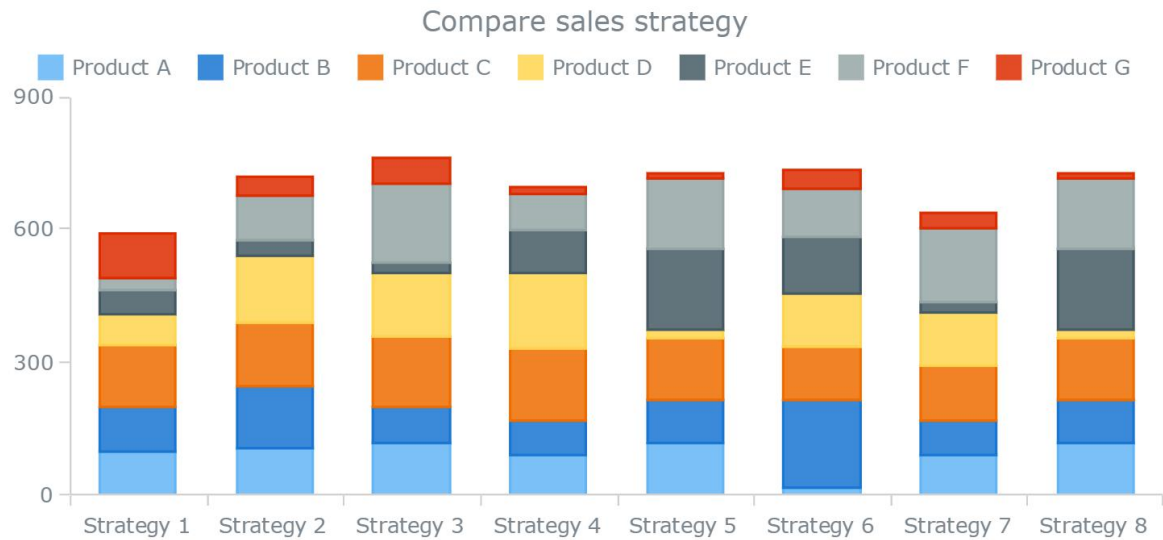


Stacked Bar Chart



Visual Variables	Preferred Data Type
Position	Ordinal
Size	Quantitative
Colour	Categorical

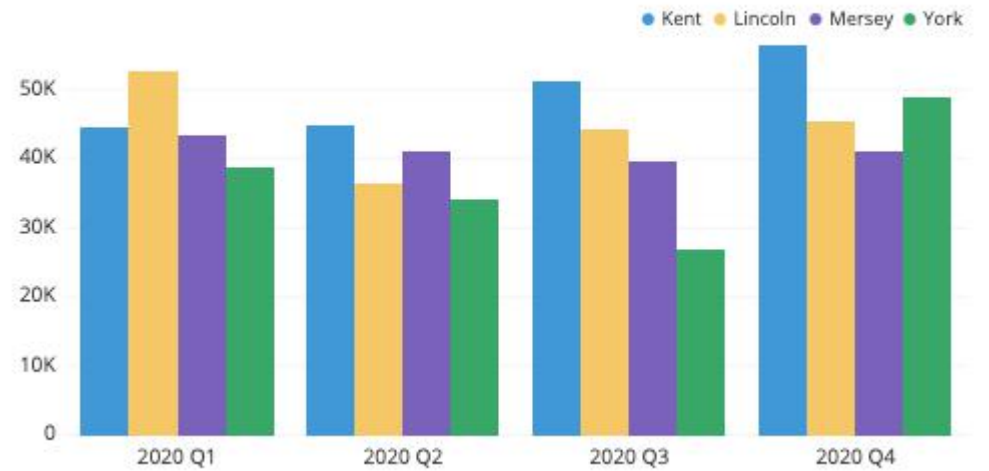
Stacked Bar Chart



VS

Multiple Bar Chart

New Revenue

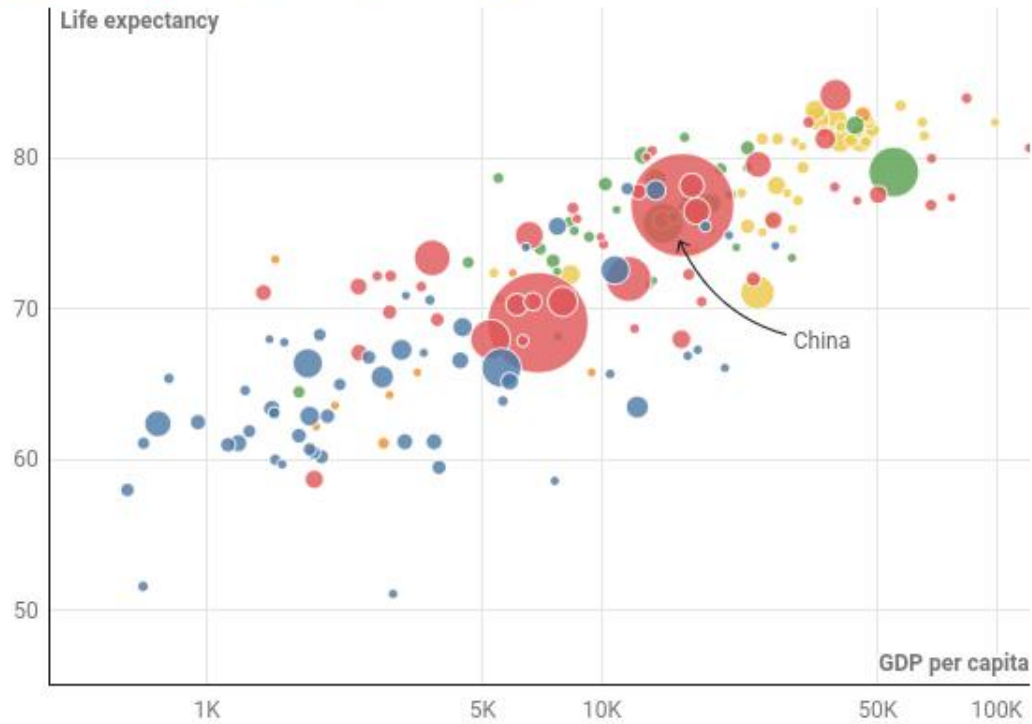


Scatterplot

GDP per person and life expectancy (2018)

The bigger a circle, the more people live in a country.

● Australia ● Asia ● Africa ● America ● Europe



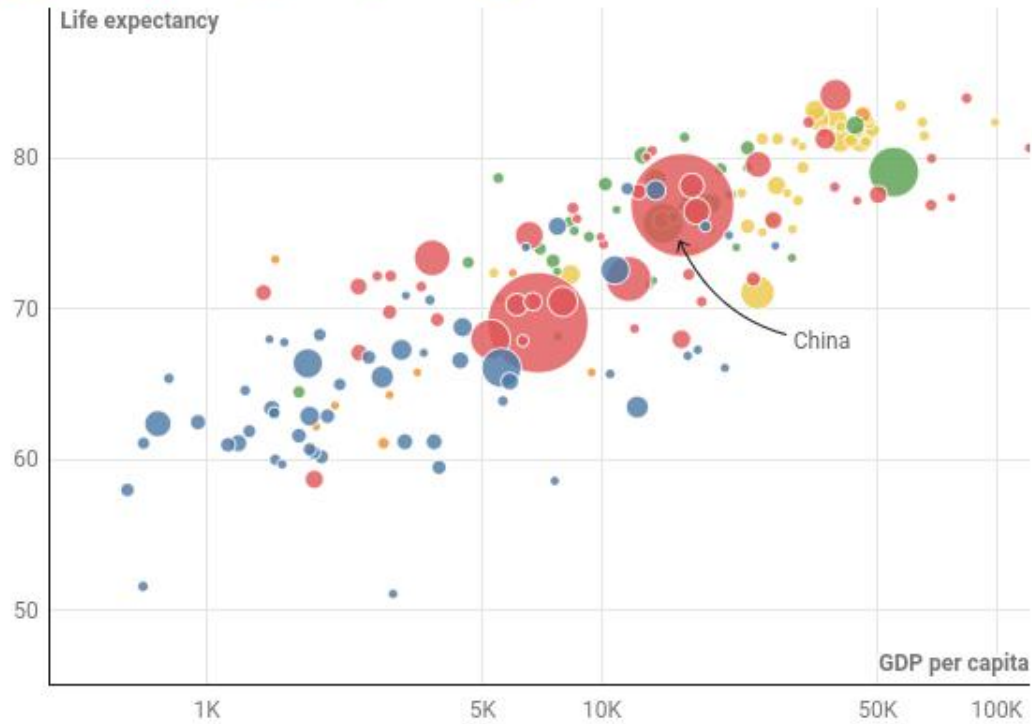
Visual Variables	Preferred Data Type
Position (x2)	Quantitative
Size	Quantitative
Colour	Categorical

Colour Hue

GDP per person and life expectancy (2018)

The bigger a circle, the more people live in a country.

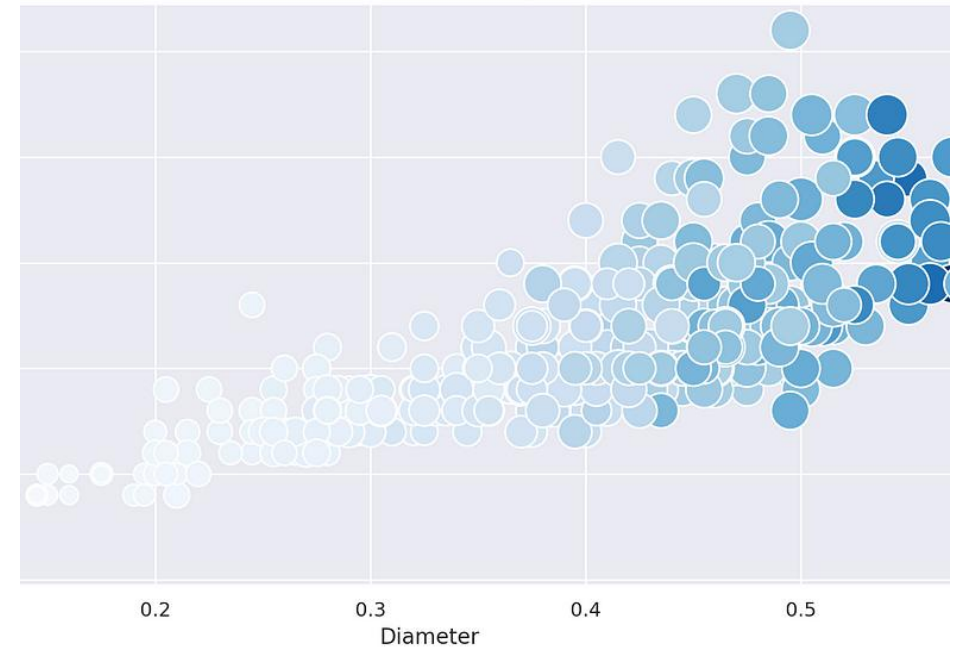
● Australia ● Asia ● Africa ● America ● Europe



VS

Colour Value

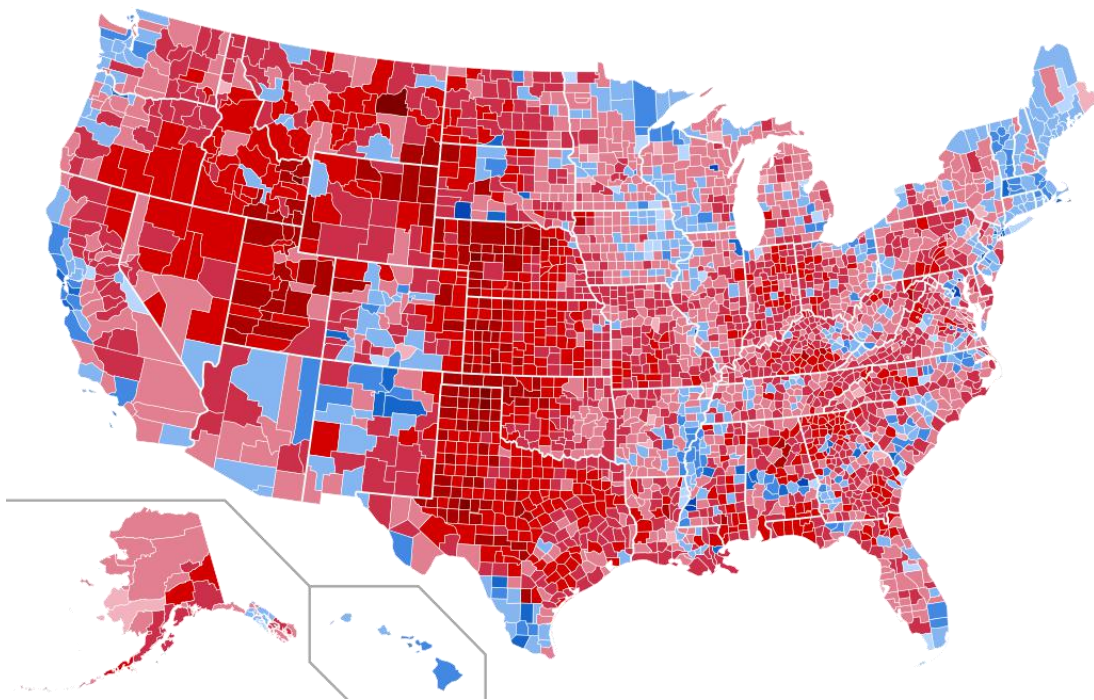
Abalone Shells Scatterplot



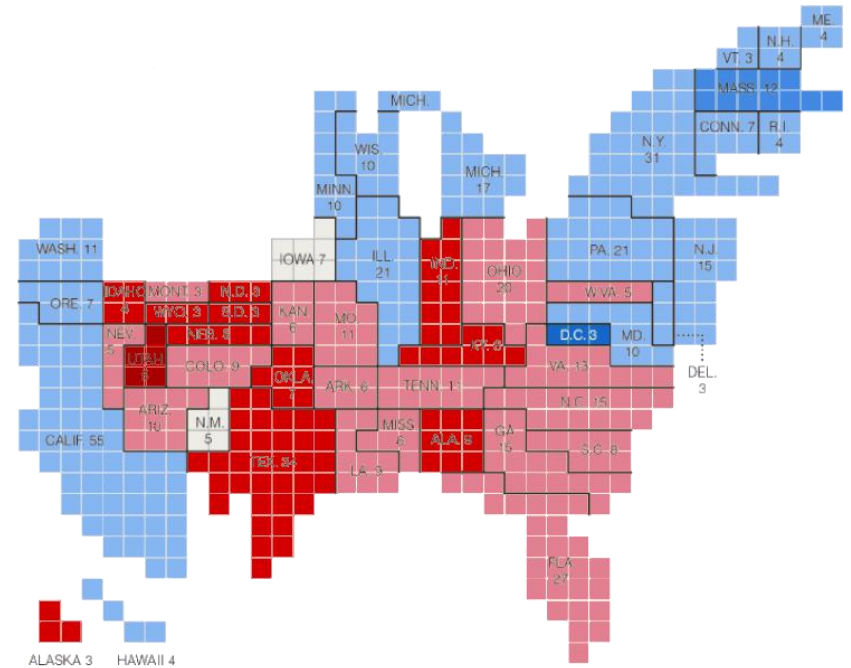
Biases

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%



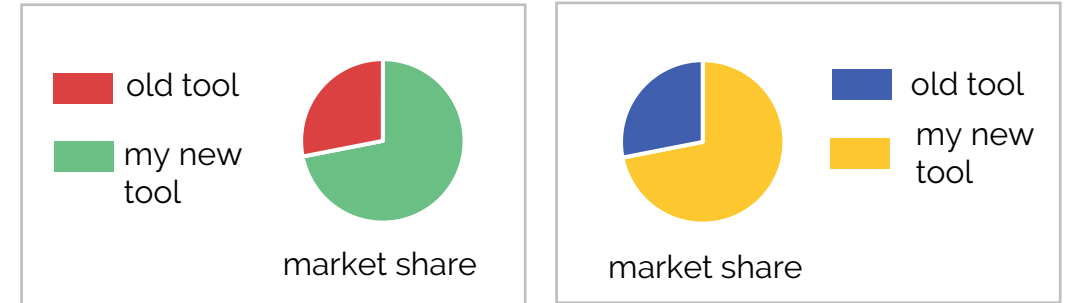
2004 United States Presidential Election results.
Size of the state weighted according of the number of electoral votes

Significance of colour

Colour meanings can facilitate information acquisition but can, conversely, affect data interpretation.

- Match colours when the correlation is universally accepted (blue for water, green for forest). **Do not match colours with clichés** (gender, ethnicity, ...) or with emotional connotations.
- Do not impose order (by using a gradient scale) when there is none. Instead, use a categorical colour scale.

Let the data speak for itself.



The use of red and green on the left could indicate a difference in value between the two items. If they are none, neutral colours are preferred.



The use of a colour scale on the left could indicate a value order between these people. If they are none, colour hue is preferred.

Scale Biases

Playing with the scale of a representation (zooming or distorting) can be very handy when one wish to display a certain level of detail. But such design choice will remove access to the context, potentially detrimental to the overall understanding.

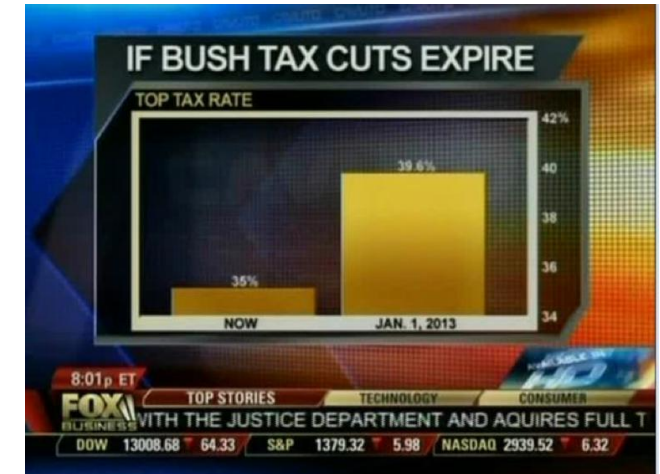
How to avoid: Display an overview view or use interactivity to access overview



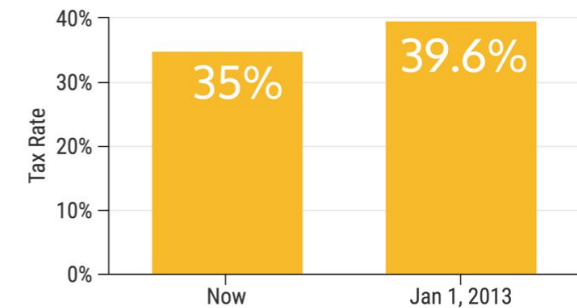
Al Gore distorting the scale on the top of his graph to accentuate the effect of climate change



Steve Jobs tilting his pie chart the accentuate Apple's sales



If Bush Tax Cuts Expire



Fox news playing with the legend to accentuate the effect of removing Bush tax

Evaluation

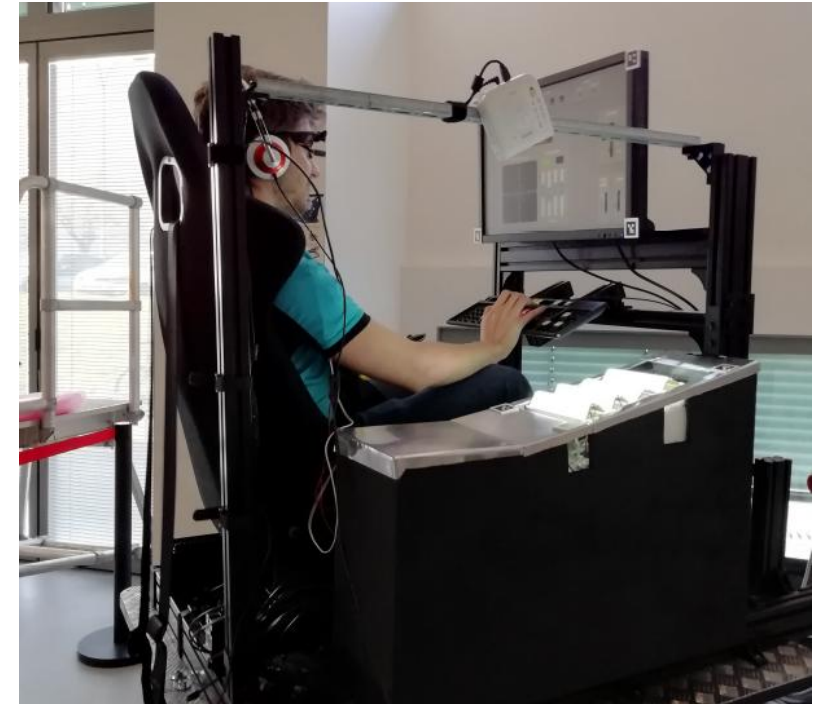
Evaluation

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

Use of Demographic

Finally, be mindful of your sample when conducting a user study.

Recording participant demographics allows us to identify the population sample that provided the data and, as such, reference the outcomes.

For example, there was a significant bias in automated face detection algorithms for a long time as they were tested only on specific ethnicities and genders.



Random picture from the internet about face detection. Can you tell what is wrong?

Thank you for your attention

Any quesitons?

Public Health Datathon, Insight North East

Power BI Workshop

Practical

Exercise

Dr Alma Cantu, Newcastle University
19th of April 2024, Farrell Center

