Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Evaluation in Visualization

Beatriz Sousa Santos, Universidade de Aveiro, 2021

How can we produce a Visualization?

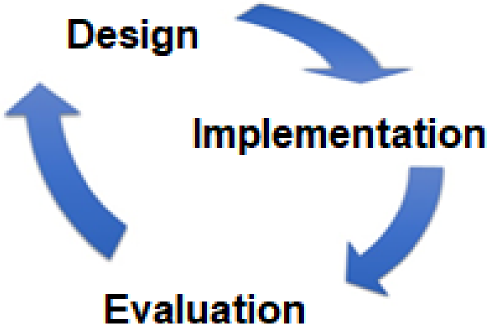
- There are **principles** (derived form human perception and cognition)

  paradigms (examples resulting form past experience)

  and many **methods**

- To obtain **efficacy** it is fundamental:
  - a correct definition of **goal** and **user tasks**
  - apply adequate methods and **evaluate**

  in **several iterations** until the goals are satisfied …



How can we evaluate?

- **Many methods** can be used to evaluate a Visualization application (some specifically developed, others adapted)

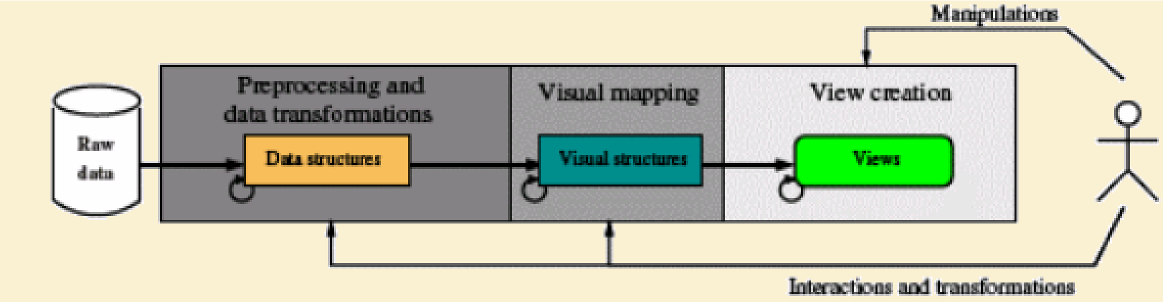
- Evaluation methods from other disciplines may and have been adapted and used to evaluate Visualization applications, as methods from:
  - Human- Computer Interaction
  - Image Processing
  - S/W Engineering

- Applications to visually explore data are interactive and should be usable
- Usability is, according to ISO 9241-11:

"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"

- How to measure it??
- We can use methods used in Human-Computer Interaction

- Evaluating a visualization technique should involve evaluation of all phases:



- data transformation

  e.g.

  low level: accuracy of methods (errors, artifacts)

  high level: efficacy and efficiency in supporting users tasks

- visual mapping

  e.g.

  high level, efficacy and efficiency in supporting users tasks

- view creation

  low level: accuracy of methods (errors, artifacts)

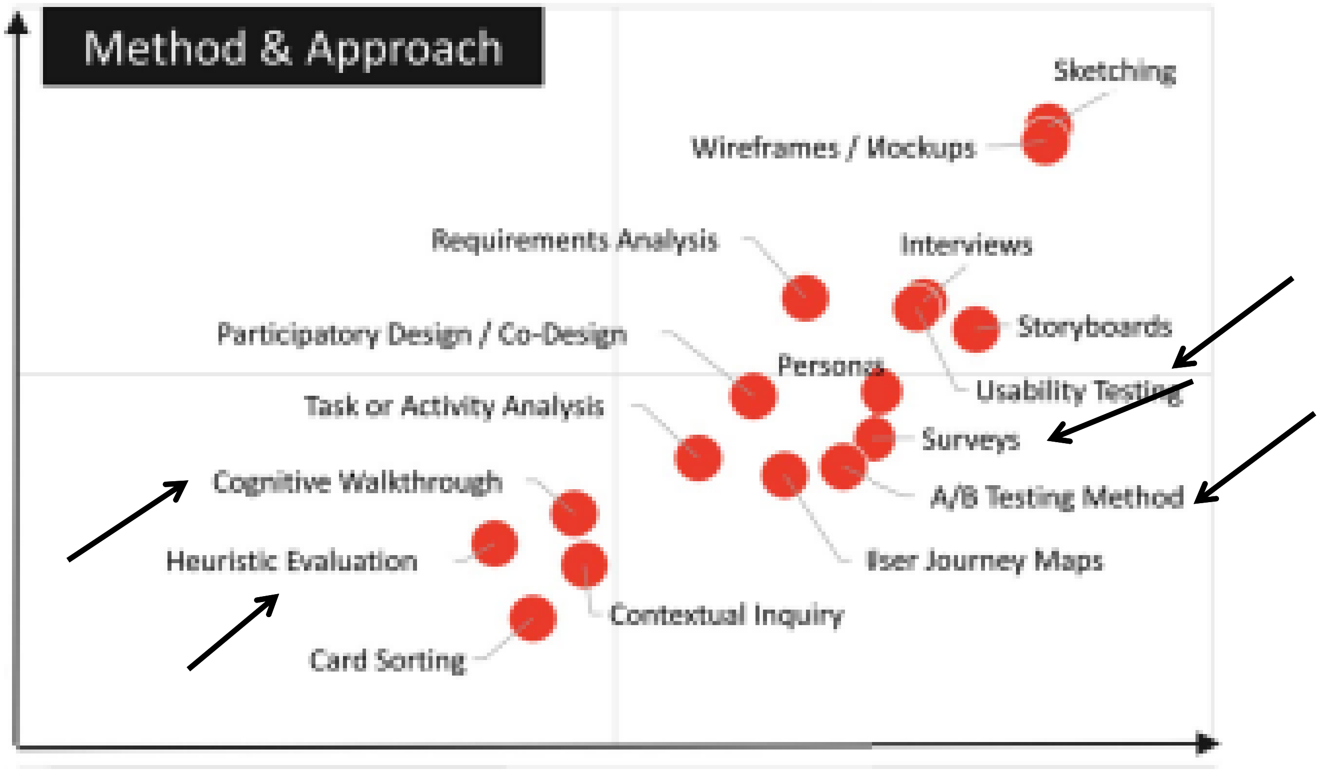
  high level: efficacy and efficiency in supporting users tasks

- Cannot forget the interaction (not only visual) aspects!

Usability evaluation Methods

- **Analytical** (without users)
  - Heuristic Evaluation
  - Cognitive Walkthrough
  - Model based methods
  - Review methods
  - ...
- **Empirical** (involving users)
  - Observation
  - Query
  - Controlled Experiments
  - ...

usability tests



P. Parsons, "Understanding Data Visualization Design Practice," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2021.3114959.

Heuristic Evaluation (Nielsen and Molich 1990)

- A “**discount usability engineering method**” for quick, cheap, and easy evaluation of a UI design
- Most popular usability inspection method; yet is **subjective**
- It is a **systematic inspection** of a design for usability
- Meant to find the usability problems in the design so that they can be attended to as part of an iterative design process
- Involves a small set of analysts judging the UI against a list of usability principles ("**heuristics**")

How to perform HE

- Should be performed by **several evaluators** (one person will never be able to find all the problems)

- Evaluators should work independently:
  - First get a general idea of the UI
  - Then perform a detailed inspection using a set of heuristics
  - Listing usability problems (heuristics not followed and severity degree)

- Findings of all evaluators should be **integrated in the same report**

- The report should **help the development teem to prioritize problem fixing**

- Nielsen proposed **10 general usability heuristics**, yet **there are other sets** (e.g., visualization, web, mobile applications for seniors or children…)

- More details on how to conduct a heuristic evaluation at: http://www.nngroup.com/articles/how-to-conduct-a-heuristic evaluation

- And how to rate severity of the usability problems found: http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

- **The list of problems and severity rates should help the development team to prioritys problem fixing**

List of recognized usability principles (“the heuristics”)

1-Visibility of system status

2-Match between system and the real world

3-User control and freedom

4-Consistency and standards

5-Error prevention

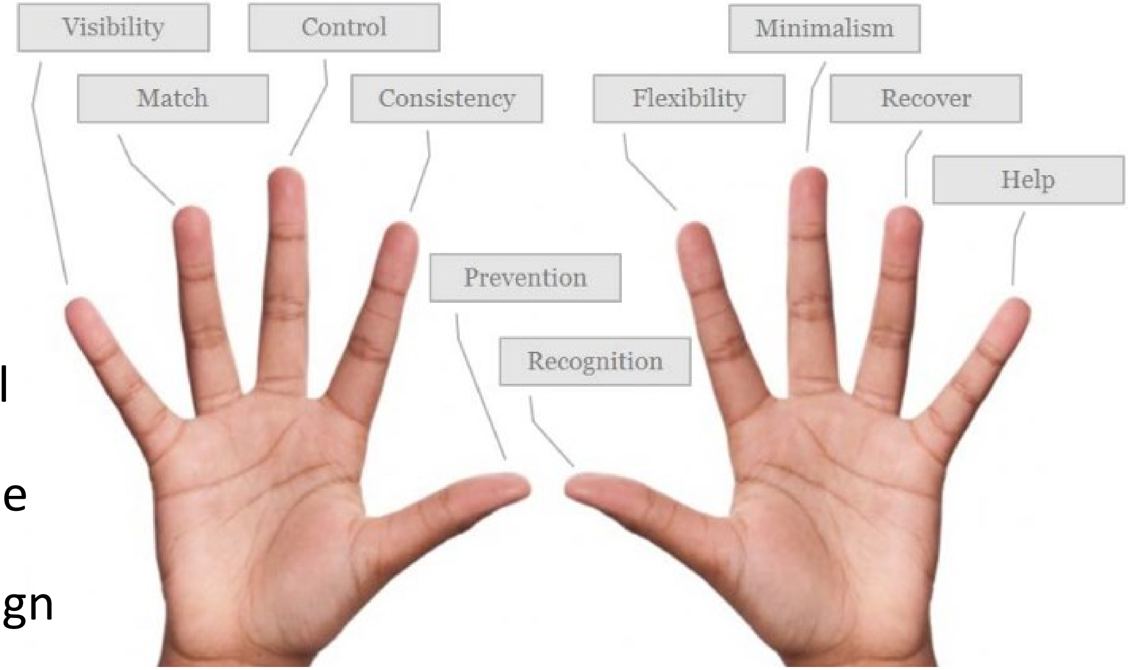
6-Recognition rather than recall

7-Flexibility and efficiency of use

8-Aesthetic and minimalist design

9-Help users recognize, diagnose, and recover from errors

10-Help and documentation



https://www.nngroup.com/articles/ten-usability-heuristics/

Specific Heuristics for Visualization

- Zuk's Perceptual and Cognitive heuristics (Zuk *et al*., 2006)
- Forsell's. heuristic set for evaluation in InfoVis (Forsell and Johanson, 2010)
- Shneiderman's "Visual Information-Seeking Mantra"
- Freitas's *et al*. Ergonomic Criteria for Hierarchical Information Visualization Techniques (Freitas et al., 2009)
- Amar and Stasko's Knowledge and task-based framework
- ...

Zuk and Carpendale’s (2006) heuristics

1-Ensure visual variable has sufficient length

2-Don’t expect reading order from color

3-Color perception varies with size of colored item

4-Local contrast affects color & gray perception

5-Consider people with color blindness

6-Preattentive benefits increase with field of view

7-Quantitative assessment requires position or size variation

8-Preserve data to graphics dimensionality

9-Put the most data in the least space

10-Remove the extraneous (ink)

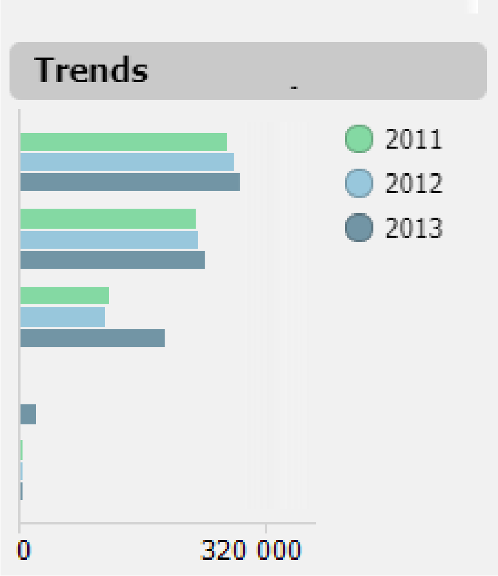
11-Consider Gestalt Laws

12-Provide multiple levels of detail

13-Integrate text whenever relevant

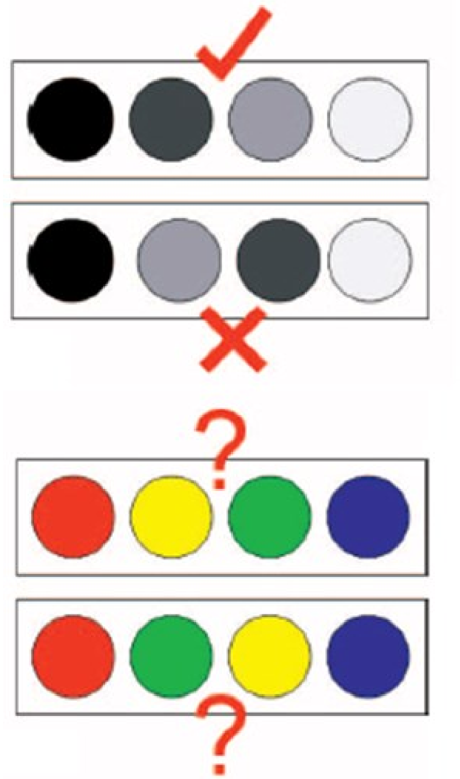
Explaining some of the Specific Heuristics for Visualization

Visual variables must have sufficient length

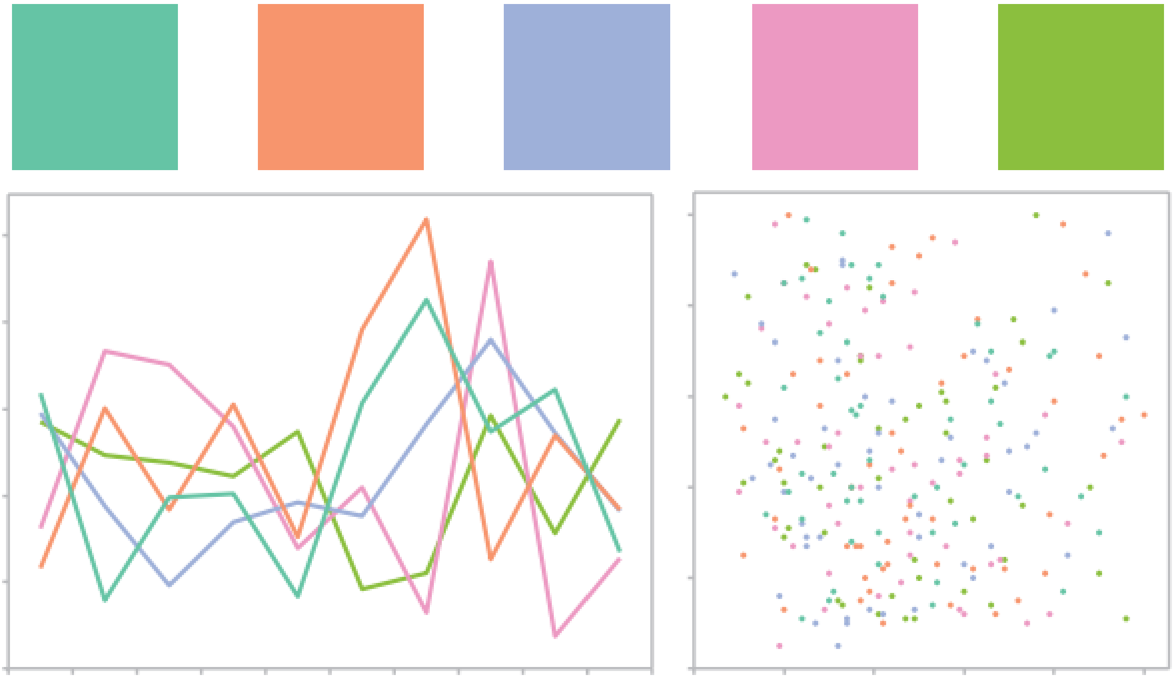


Too small

Do not expect to easily perceive order from color



Color perception varies with size of colored item



A set of colors with different hues but equal luminance (L* = 72). In the squares, these colors are distinct and visually balanced (www.colorbrewer.org).
They are harder to distinguish in smaller items

Stone, M., “In color perception, size matters”, IEEE *Computer Graphics & Applications*. 32, 2, 2012, pp. 8-13

Consider people with color blindness

The most common form of color blindness is deuteranopia (“daltonism”)

There are color blindness simulators

Drag and drop or paste your file in the area below or: Escolher ficheiro Nenhum ficheiro selecionado

Trichromatic view:
● Normal

Anomalous Trichromacy:
○ Red-Weak/Protanomaly
○ Green-Weak/Deuteranomaly
○ Blue-Weak/Tritanomaly

Dichromatic view:
○ Red-Blind/Protanopia
○ Green-Blind/Deuteranopia
○ Blue-Blind/Tritanopia

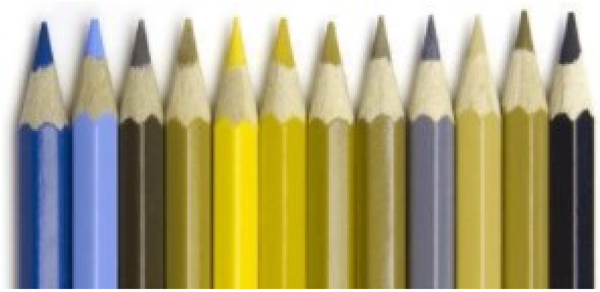
Monochromatic view:
○ Monochromacy/Achromatopsia
○ Blue Cone Monochromacy

Use lens to compare with normal view: ● No Lens ○ Normal Lens ○ Inverse Lens
Reset View

Zoom, move and lens functionality only with your own images available.

http://www.color-blindness.com/coblis-color-blindness-simulator

Normal vision

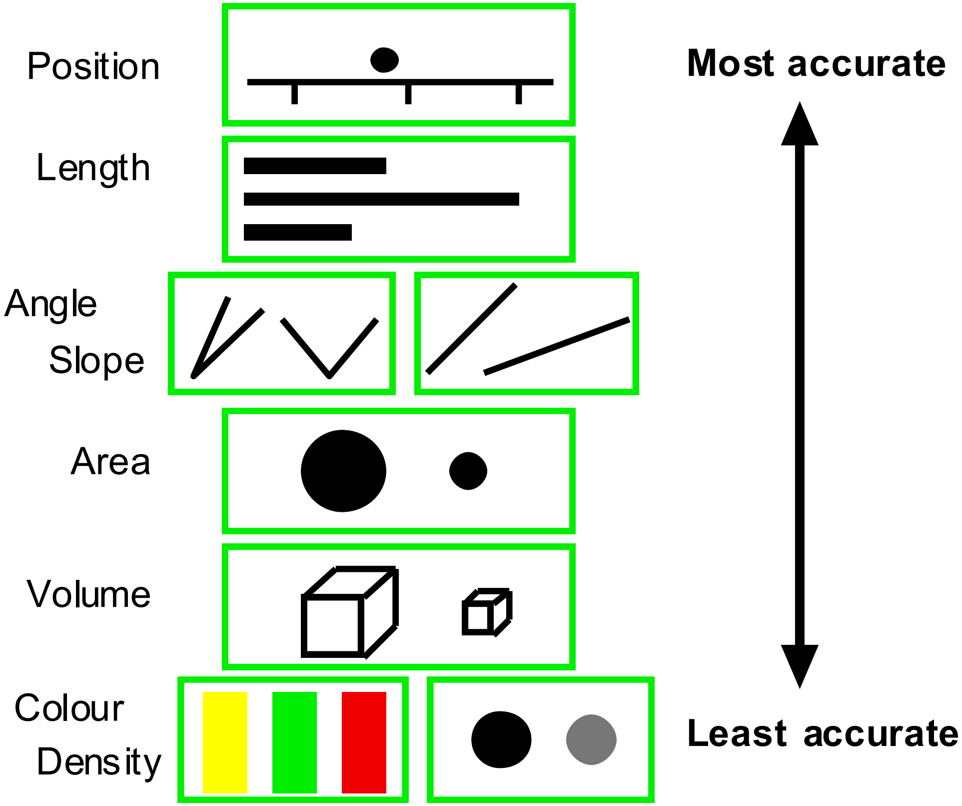


Deuteranopia

Tritanopia

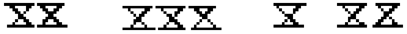
http://www.colourblindawareness.org/

Quantitative assessment requires position or size variation



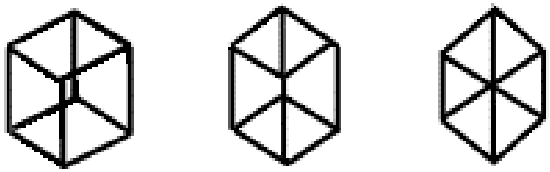
The relative difficulty of assessing quantitative value as a function of encoding mechanism, as established by Cleveland and McGill (Spence, 2007)

Gestalt Laws



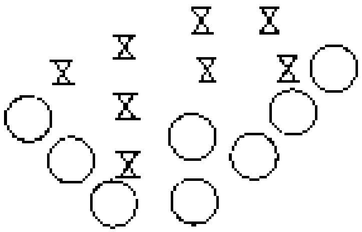
Proximity

Near stimuli are perceived as a group



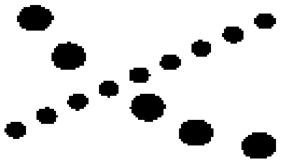
Simplicity

Ambiguous stimuli tend to be resolved
Using the simplest explanation



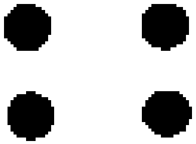
Similarity

Similar stimuli tend to be grouped
(may override proximity)



Good continuation

Stimuli tend to be grouped as to minimize
variations or discontinuities



Closure

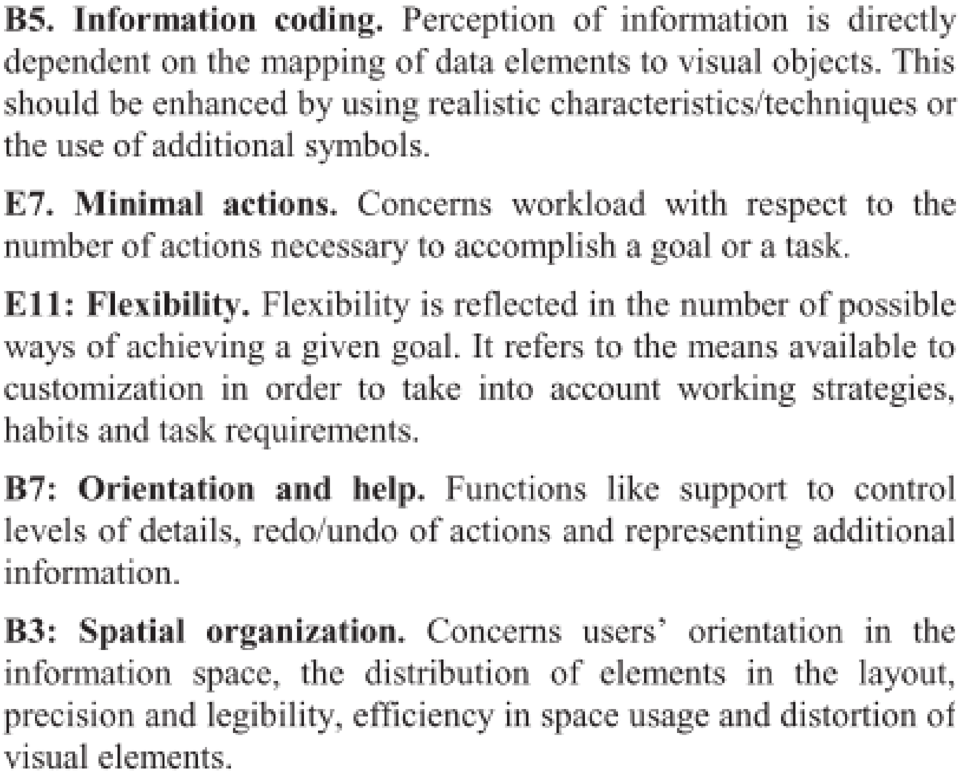
Stimuli tend to be grouped
in complete figures

Symmetry

Regions delimited by symmetric tend
to be perceived as coherent figures

Forsell´s et al. (2010) heuristics

Uses heuristics from other sets:
B- Freitas et al.
C - Nielsen
D- Zuck and Carpendale
E- Bastien & Scapin



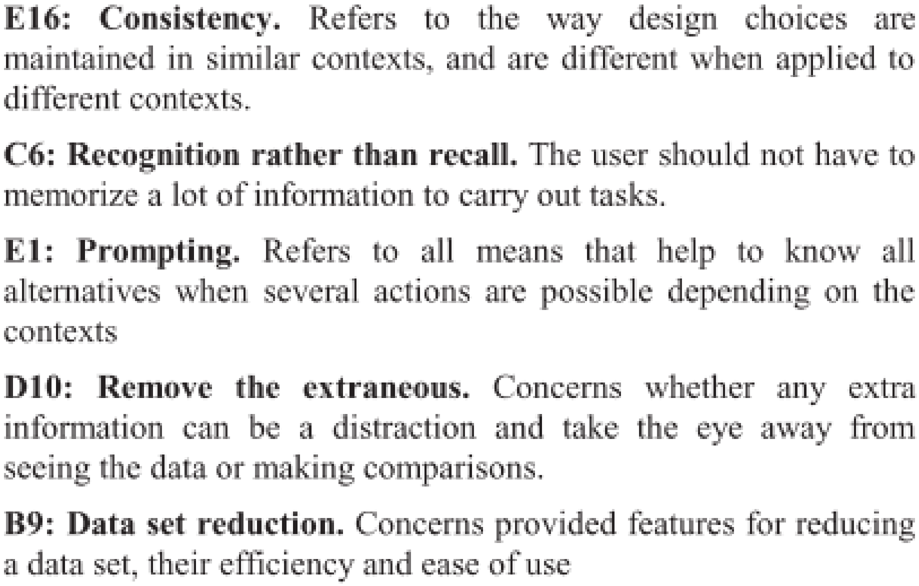
B5. Information coding. Perception of information is directly dependent on the mapping of data elements to visual objects. This should be enhanced by using realistic characteristics/techniques or the use of additional symbols.

E7. Minimal actions. Concerns workload with respect to the number of actions necessary to accomplish a goal or a task.

E11: Flexibility. Flexibility is reflected in the number of possible ways of achieving a given goal. It refers to the means available to customization in order to take into account working strategies, habits and task requirements.

B7: Orientation and help. Functions like support to control levels of details, redo/undo of actions and representing additional information.

B3: Spatial organization. Concerns users' orientation in the information space, the distribution of elements in the layout, precision and legibility, efficiency in space usage and distortion of visual elements.



E16: Consistency. Refers to the way design choices are maintained in similar contexts, and are different when applied to different contexts.

C6: Recognition rather than recall. The user should not have to memorize a lot of information to carry out tasks.

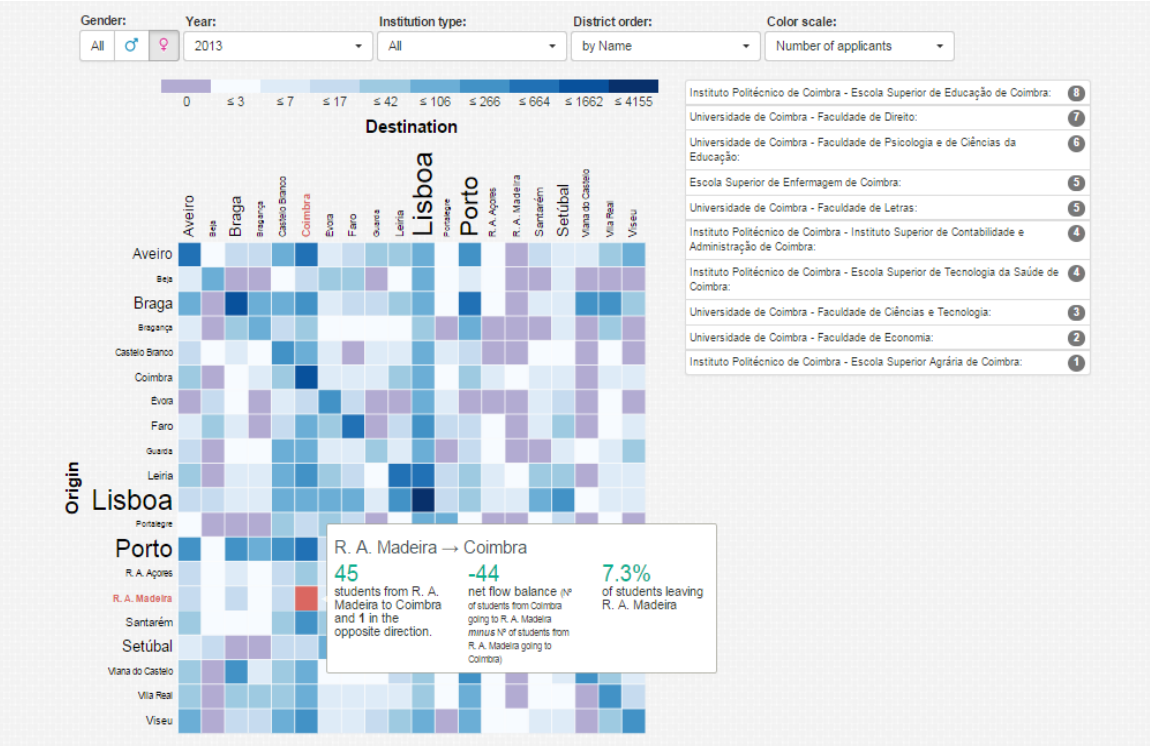
E1: Prompting. Refers to all means that help to know all alternatives when several actions are possible depending on the contexts

D10: Remove the extraneous. Concerns whether any extra information can be a distraction and take the eye away from seeing the data or making comparisons.

B9: Data set reduction. Concerns provided features for reducing a data set, their efficiency and ease of use

Example:

https://migration-flow.herokuapp.com/



Gender: All

Year: 2013

Institution type: All

District order: by Name

Color scale: Number of applicants

Instituto Politécnico de Coimbra - Escola Superior de Educação de Coimbra:	8
Universidade de Coimbra - Faculdade de Direito:	7
Universidade de Coimbra - Faculdade de Psicologia e de Ciências da Educação:	6
Escola Superior de Enfermagem de Coimbra:	5
Universidade de Coimbra - Faculdade de Letras:	5
Instituto Politécnico de Coimbra - Instituto Superior de Contabilidade e Administração de Coimbra:	4
Instituto Politécnico de Coimbra - Escola Superior de Tecnologia da Saúde de Coimbra:	4
Universidade de Coimbra - Faculdade de Ciências e Tecnologia:	3
Universidade de Coimbra - Faculdade de Economia:	2
Instituto Politécnico de Coimbra - Escola Superior Agrária de Coimbra:	1

R. A. Madeira → Coimbra

45 students from R. A. Madeira to Coimbra and **1** in the opposite direction.

-44 net flow balance (Nº of students from Coimbra going to R. A. Madeira *minus* Nº of students from R. A. Madeira going to Coimbra)

7.3% of students leaving R. A. Madeira

This visualization is compliant with Zuk and Carpendale' heuristics number 3 and 12, "*Color perception varies with size of colored item*" and "*Provide multiple levels of detail*", respectively.

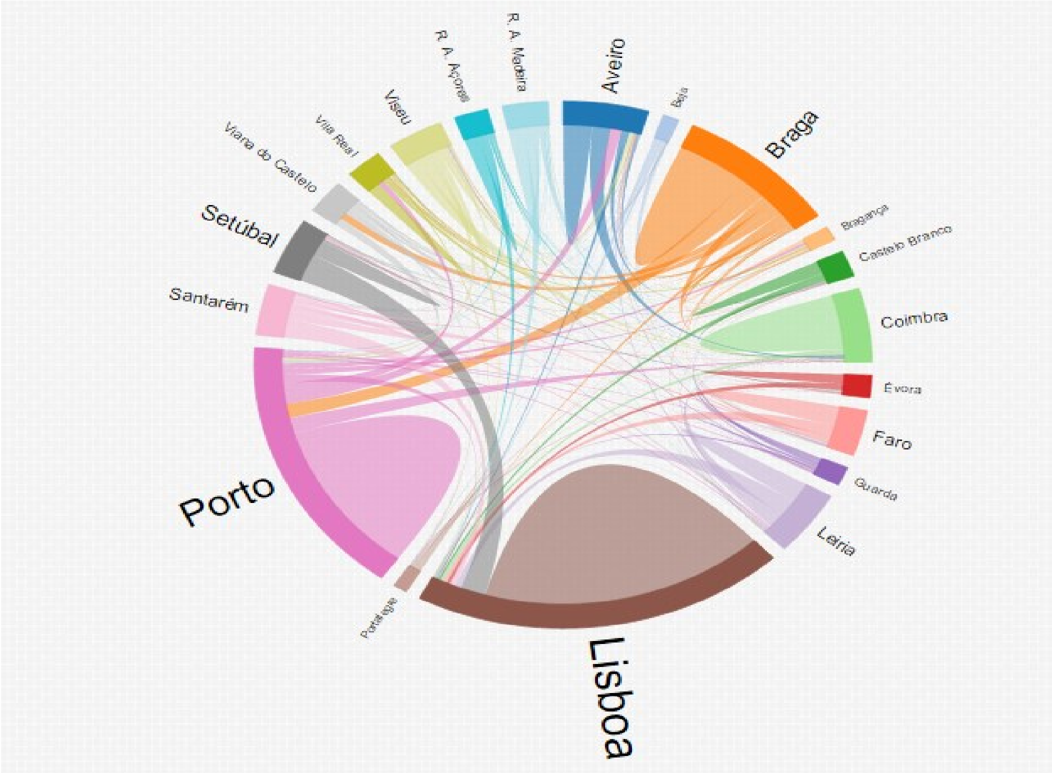
Nielsen´s heuristic number 4 "Consistency" is not complied with, as there are different types of filter selection and the color scale varies depending on the filters applied.

Example (cont.)

Zuk and Carpendale's heuristic number 11 "Consider Gestalt Laws" is complied with,

heuristic number 1 ”Ensure visual variable has sufficient length” is not satisfied.

Nielsen´s heuristic number 2 "Match between system and the real world", and Forsell and Johansson's heuristic number 1 "Information Coding" are not satisfied



Practical activity on evaluation
(groups of 3 students)

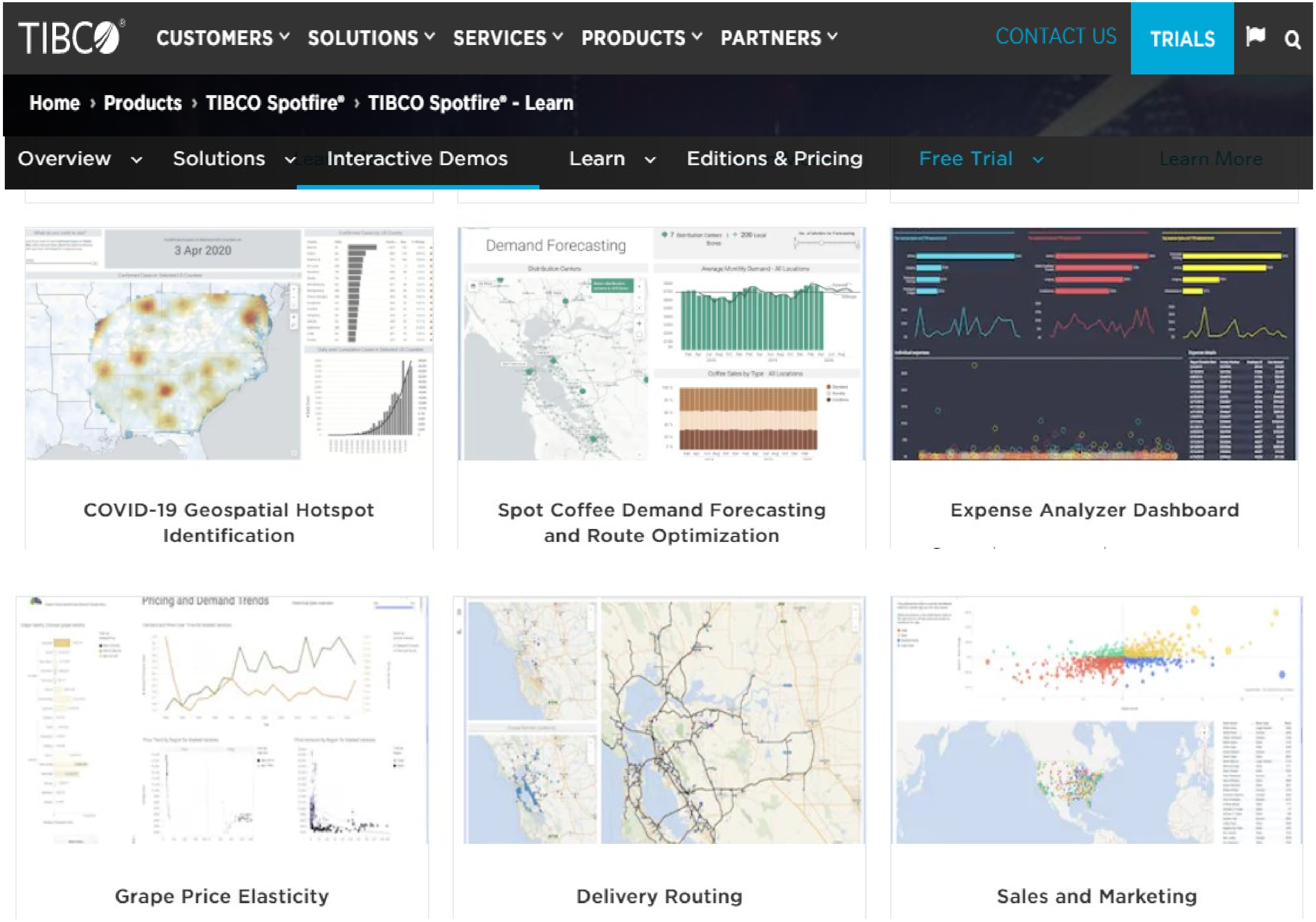
- Select a (preferably not very usable) Visualization/Visual Data Mining application and evaluate it using heuristic evaluation with one of these heuristics sets:
  - Nielsen
  - Zuk et al.
  - Other …

- You may find interesting examples at:
  - Tableau public – gallery
  - Spotfire – gallery

    …
- Each evaluator should read carefully and try to understand each heuristic and analyze independently the application registering the potential problems and their classification

- Discuss the problems with the other group members and consolidate a list of problems

- Prepare a 5 min presentation with the main potential problems you found and send it to bss@ua.pt



COVID-19 Geospatial Hotspot Identification

Spot Coffee Demand Forecasting and Route Optimization

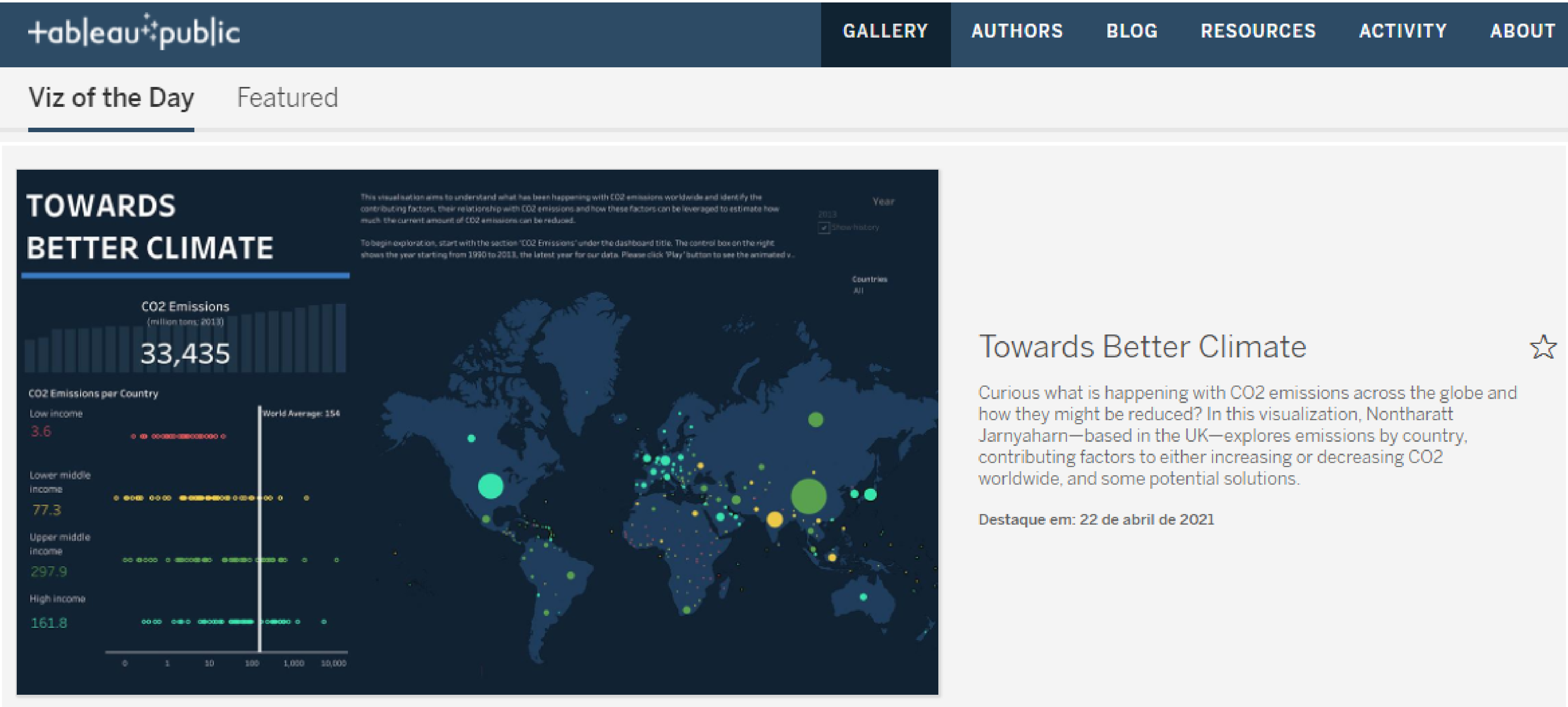
Expense Analyzer Dashboard

Grape Price Elasticity

Delivery Routing

Sales and Marketing

https://www.tibco.com/products/tibco-spotfire/learn/demos



This visualisation aims to understand what has been happening with CO2 emissions worldwide and identify the contributing factors, their relationship with CO2 emissions and how these factors can be leveraged to estimate how much the current amount of CO2 emissions can be reduced.

To begin exploration, start with the section 'CO2 Emissions' under the dashboard title. The control box on the right shows the year starting from 1990 to 2013, the latest year for our data. Please click 'Play' button to see the animated v...

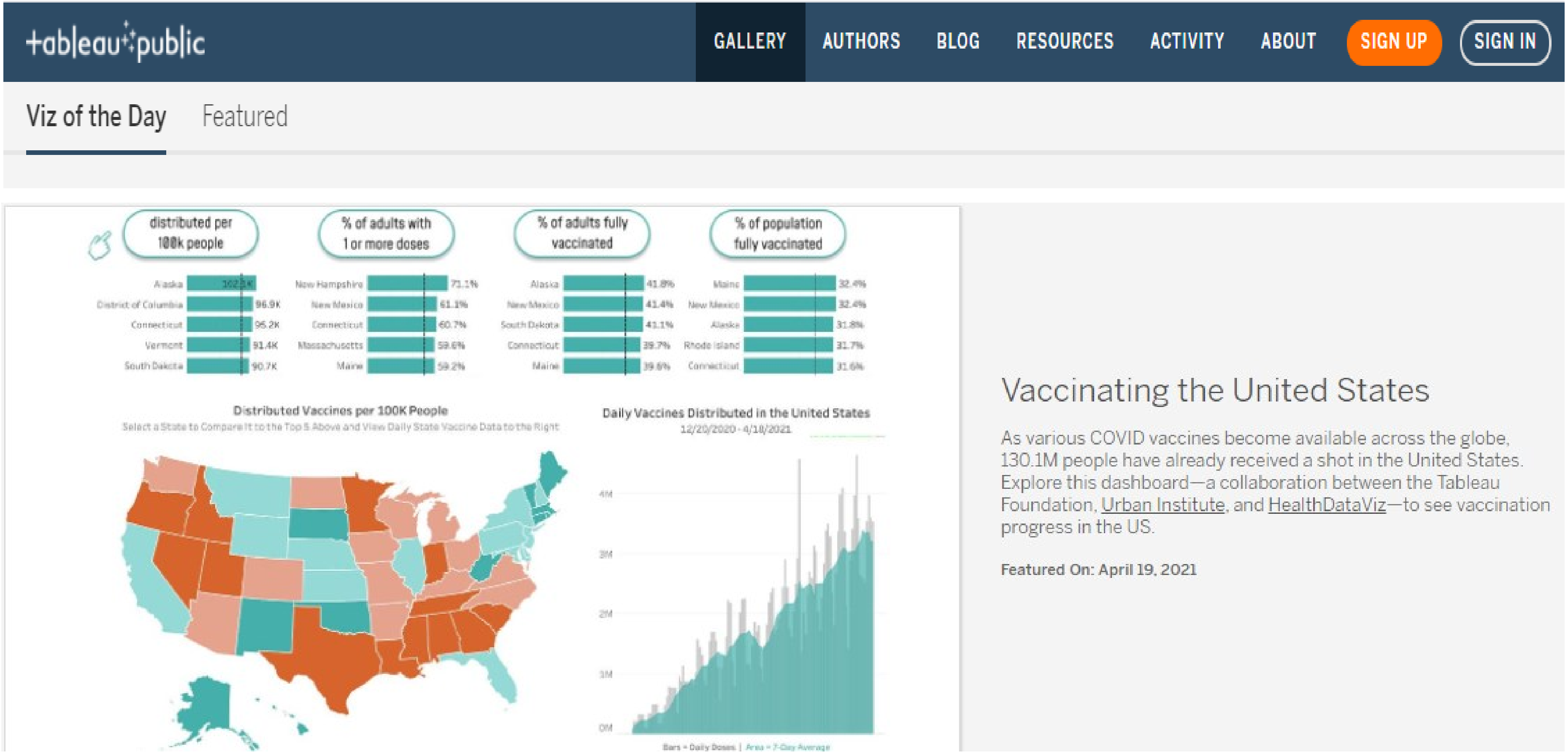
Towards Better Climate ☆

Curious what is happening with CO2 emissions across the globe and how they might be reduced? In this visualization, Nontharatt Jarnyaharn—based in the UK—explores emissions by country, contributing factors to either increasing or decreasing CO2 worldwide, and some potential solutions.

Destaque em: 22 de abril de 2021

https://public.tableau.com/en-us/s/gallery

Example: vaccinating the united states



Vaccinating the United States

As various COVID vaccines become available across the globe, 130.1M people have already received a shot in the United States. Explore this dashboard—a collaboration between the Tableau Foundation, Urban Institute, and HealthDataViz—to see vaccination progress in the US.

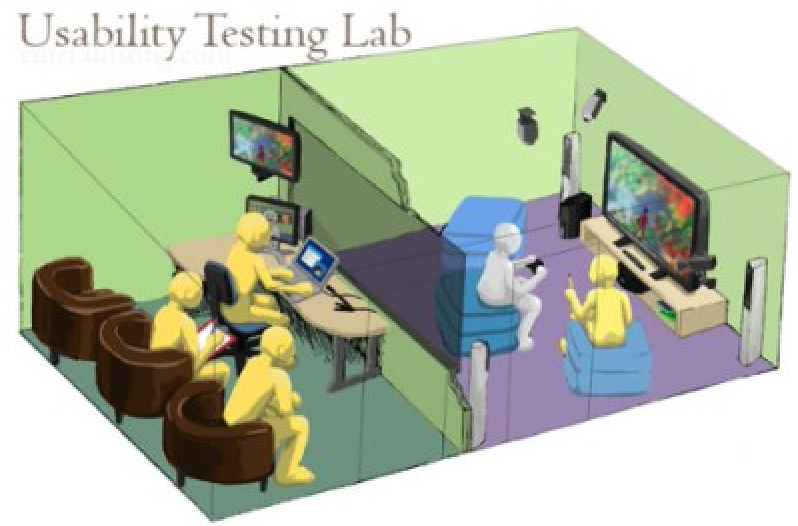
Featured On: April 19, 2021

https://public.tableau.com/en-us/gallery/vaccinating-united-states?tab=viz-of-the-day&type=viz-of-the-day

Usability tests

- “Engineering approach”
- Involve observation and query
- Main aspects:
    - Participants
    - Tasks
    - Test facilities and systems
    - Experimental design
    - Usability measures
    - Data analysis
- May have a complex logistics

Observation



Has many variants from very simple
to very complex and expensive:

- **Direct**: observer takes notes
- **Indirect**: through audio/ video – more complex and time consuming
- **Think Aloud**: users are asked to explain what they are doing
- **Logging**: users activity is logged by the system
- Combinations of the previous, etc.

Query

- Two main variants:
  - Questionnaire (reach more people; less flexible)
  - Interview
- **Should always be carefully prepared and tested**
- Collected data should be carefully analyzed

https://www.interaction-design.org/literature/article/useful-survey-questions-for-user-feedback-surveys
https://www.interaction-design.org/literature/article/how-to-conduct-user-interviews

Well-known usability questionnaires

- System Usability Scale (**SUS**)

- Questionnaire for User Interface Satisfaction (**QUIS**)

- SUS provides a “quick and dirty”, reliable tool for measuring the usability
- It includes 10 questions with five response options

- QUIS is designed to assess a user's subjective satisfaction with the UI

- It is designed to be configured according to the needs of each UI analysis by including only the sections that are of interest to the user

- Both questionnaires should be completed following use of the UI in question

System Usability Scale (SUS)

- Provides a “quick and dirty”, reliable tool for measuring the usability
- It includes 10 questions with five response options
- It allows to evaluate a wide variety of products and services

  (H/W, S/W, mobile devices, websites and applications)
- Has become an industry standard, with references in over 1300 publications

Benefits of using a SUS

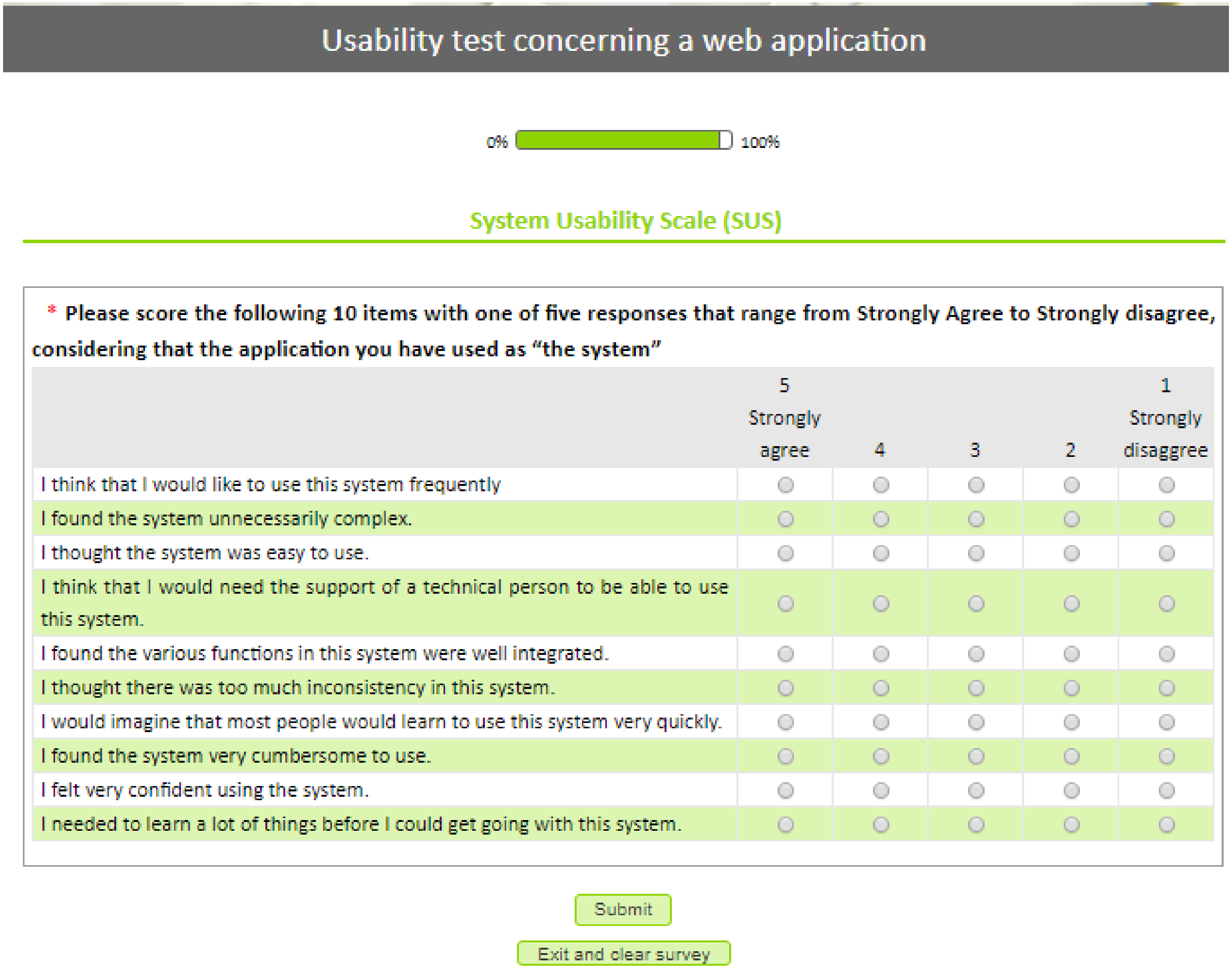
- Is a very easy scale to administer to participants
- Can be used on small sample sizes with reliable results
- Is valid – it can differentiate between usable and unusable systems

https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

SUS Questions

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

https://www.usability.gov/how-to-and-tools/resources/templates/system-usability-scale-sus.html



Usability test concerning a web application

0% 100%

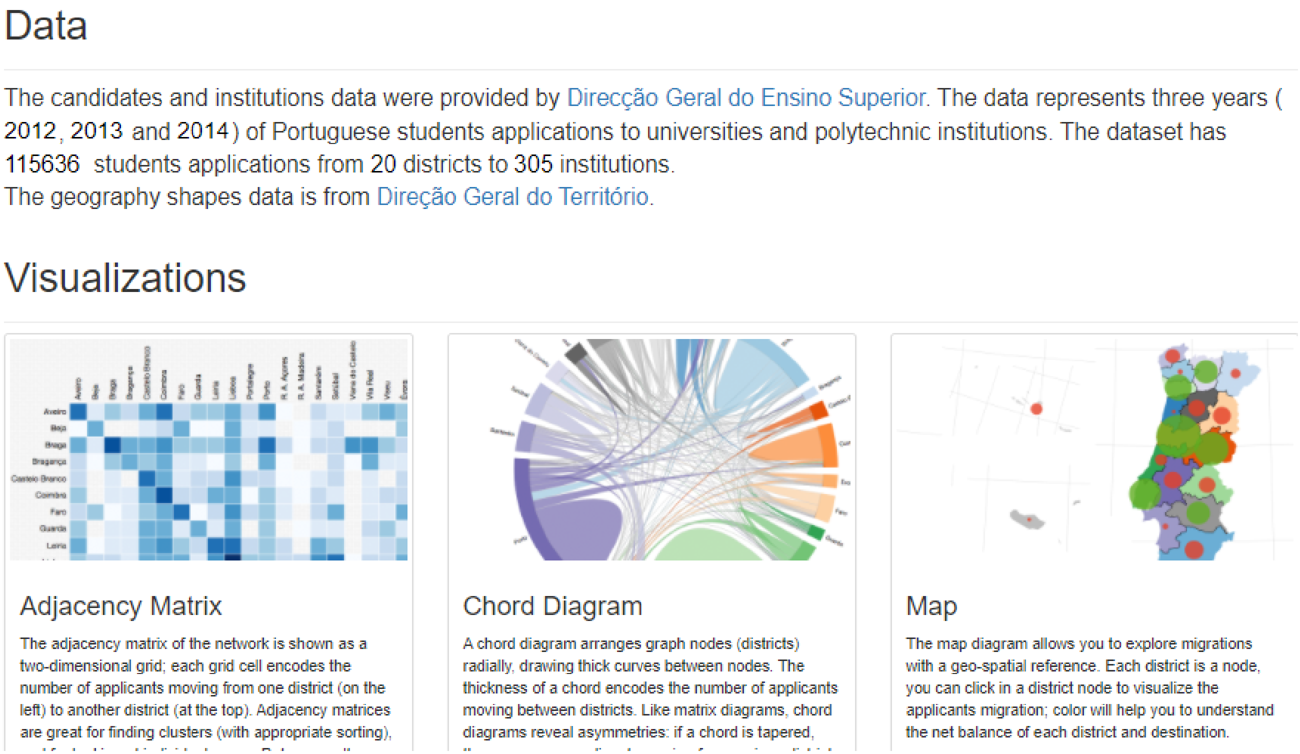
System Usability Scale (SUS)

* Please score the following 10 items with one of five responses that range from Strongly Agree to Strongly disagree, considering that the application you have used as “the system”

	5 Strongly agree	4	3	2	1 Strongly disaggree
I think that I would like to use this system frequently	○	○	○	○	○
I found the system unnecessarily complex.	○	○	○	○	○
I thought the system was easy to use.	○	○	○	○	○
I think that I would need the support of a technical person to be able to use this system.	○	○	○	○	○
I found the various functions in this system were well integrated.	○	○	○	○	○
I thought there was too much inconsistency in this system.	○	○	○	○	○
I would imagine that most people would learn to use this system very quickly.	○	○	○	○	○
I found the system very cumbersome to use.	○	○	○	○	○
I felt very confident using the system.	○	○	○	○	○
I needed to learn a lot of things before I could get going with this system.	○	○	○	○	○

Example of a usability test of a visual exploration app

based on a web questionnaire to be answered by a user while observed by an experimenter



Data

The candidates and institutions data were provided by Direcção Geral do Ensino Superior. The data represents three years (2012, 2013 and 2014) of Portuguese students applications to universities and polytechnic institutions. The dataset has 115636 students applications from 20 districts to 305 institutions.
The geography shapes data is from Direção Geral do Território.

Visualizations

Adjacency Matrix

The adjacency matrix of the network is shown as a two-dimensional grid; each grid cell encodes the number of applicants moving from one district (on the left) to another district (at the top). Adjacency matrices are great for finding clusters (with appropriate sorting),

Chord Diagram

A chord diagram arranges graph nodes (districts) radially, drawing thick curves between nodes. The thickness of a chord encodes the number of applicants moving between districts. Like matrix diagrams, chord diagrams reveal asymmetries: if a chord is tapered,

Map

The map diagram allows you to explore migrations with a geo-spatial reference. Each district is a node, you can click in a district node to visualize the applicants migration; color will help you to understand the net balance of each district and destination.

https://forms.ua.pt/index.php?r=survey/index&sid=489227

Controlled experiments

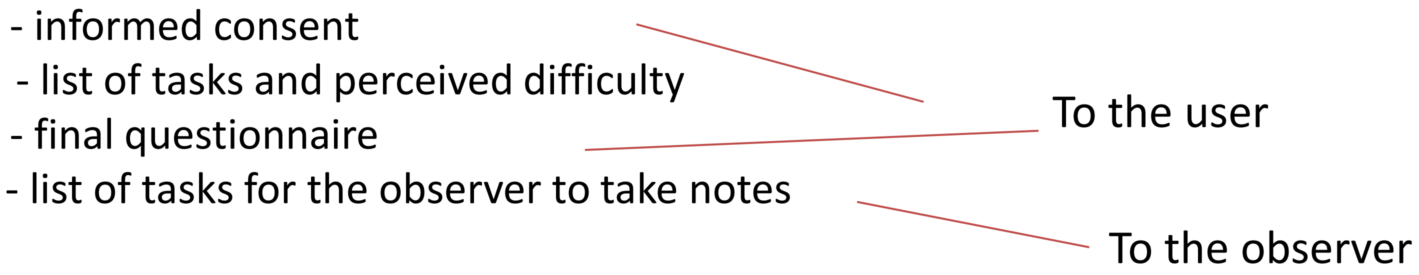
- The **work horse of experimental science** ...

- Important issues to consider:

  - **Hypothesis**
  - **Variables** (input or independent; output or dependent)
  - **Secondary variables**
  - **Experimental design** (within groups; between groups)
  - **Participants** (number, profile)
  - **Statistics**

Controlled experiment

- Define an hypothesis
- Define input (independent), output (dependent) and secondary variables
- Define experimental design (within-groups / between groups)
- Select the participants
- Prepare all the documentation:



- informed consent
- list of tasks and perceived difficulty
- final questionnaire
- list of tasks for the observer to take notes

To the user

To the observer

- Run a pilot test
- Take care of the logistics … and after the experiment analyze data

Participants

Important issues in usability tests and controlled experiments:

- The total number of participants to be tested
  (a valid statistical analysis implies a sufficient number of subjects)

- Segmentation of user groups tested, if more than one

- Key characteristics and capabilities of user group
  (user profile: age, gender, computing experience, product experience, etc.)

- How to select participants

- Differences between the participant sample and the user population
  (e.g. actual users might have training whereas test subjects were untrained)

Tasks

- The task scenarios for testing (or experiments)

- Why these tasks were selected

  (e.g. the most frequent tasks, the most troublesome tasks)

- The source of these tasks

  (e.g. observation of users using similar products, product specifications)

- Any task data given to the participants

- Completion or performance criteria established for each task

  (e.g. n. of clicks < N, time limit)

Test Facilities and equipment

- The setting and type of space in which the evaluation will be done

  (e.g. usability lab, cubicle office, meeting room, home office, home family room, manufacturing floor, etc.)

- Any relevant features or circumstances that can affect the results

  (e.g. video and audio recording equipment, one-way mirrors, or automatic data collection equipment)

- Participant’s Computing Environment

  (e.g. computer configuration, including model, OS version, required libraries or settings, browser name and version; relevant plug-in, etc.)

- Display and input devices characteristics

- Any questionnaires to be used

Experimental design

- Procedure/ protocol: the logical design of the test/experiment
- Participant general instructions and task instructions
- The independent variables and control variables
- The usability measures to be used:
  - a) for effectiveness (completeness rate, errors, assists…)
  - b) for efficiency (times)
  - c) for satisfaction

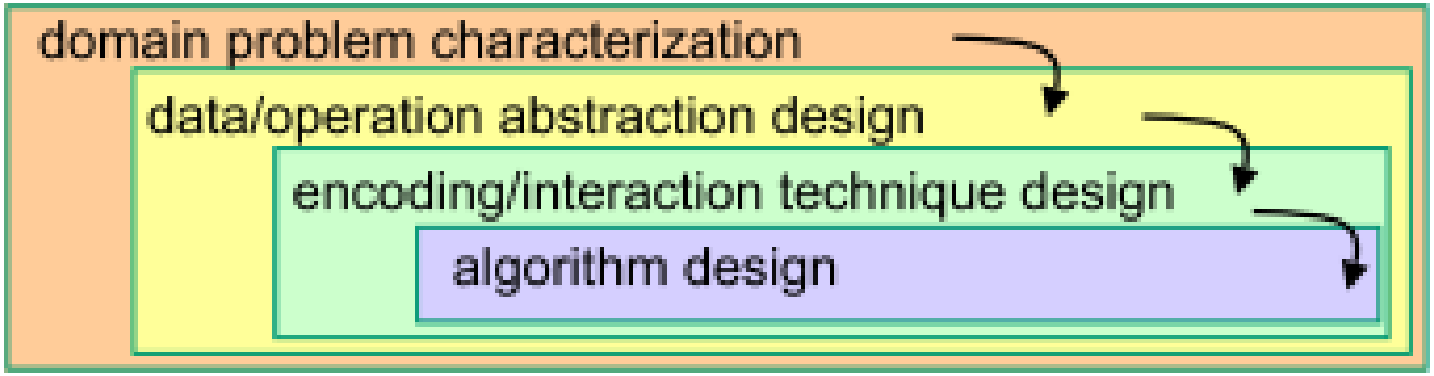
We know issues and methods, but how to use them?

A nested model for visualization design and validation:

- This model can be used :
  - to **analyze existing systems** or papers,

  or
  - to **guide the design** process

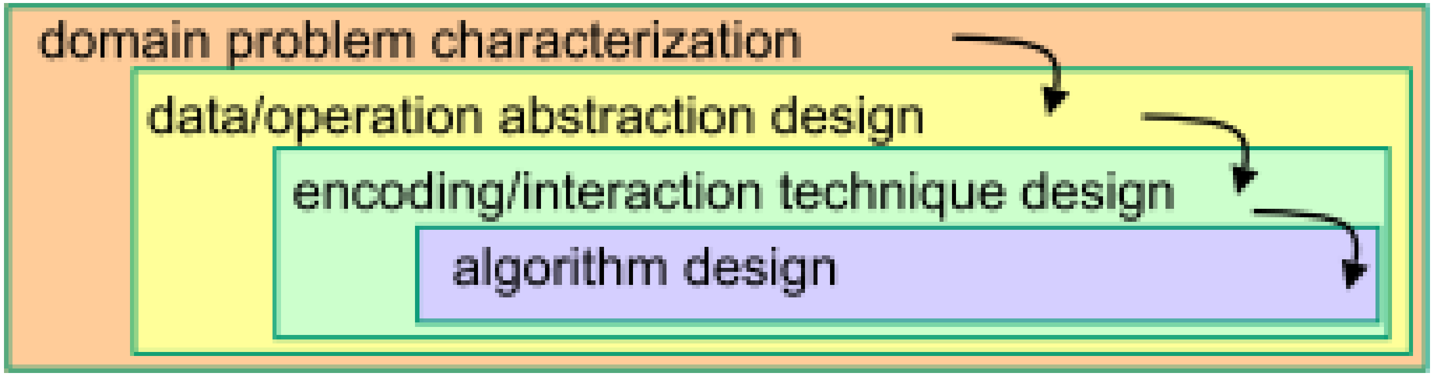
- Provides explicit guidance on what **evaluation methodology** is appropriate and identifies **threats to validity at each level**



Munzner, T.. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*. 15, 6, 2009, pp. 921–8

Four levels for visualization design and validation

- **characterize** the tasks and data in the vocabulary of the **problem domain**,
- **abstract into operations and data types**,
- **design visual encoding and interaction techniques**,
- **create algorithms** to execute these techniques with efficiency and efficacy



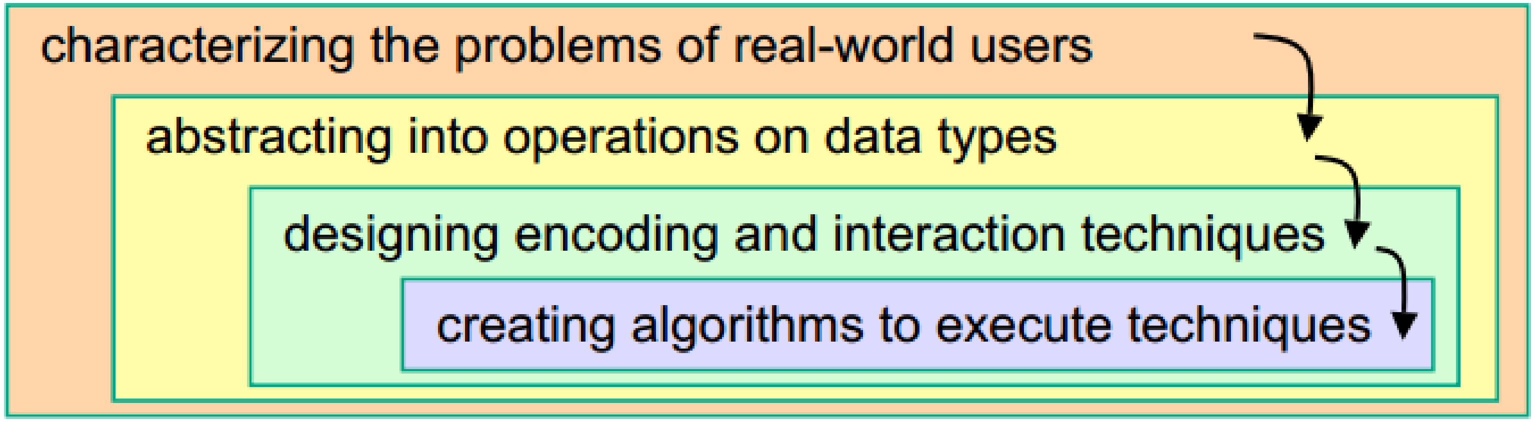
3 design levels

In each of the four levels it is necessary to :

- 1- learn about the tasks and data of target users in some particular target **domain**,
- 2- map problems and data from the vocabulary of the specific domain into a more **abstract description** in the vocabulary of computer science (e.g. filter, retrieve value, sort, find extrema, etc.)
- 3- design the **visual encoding**, presentation and interaction,
- 4- create an **algorithm** to carry out the visual encoding and interaction designs automatically. The issues of algorithm design are not unique to visualization .

This split into levels is motivated by shared **threats to validity** at each one:

- Wrong problem: they don't do that;
- Wrong abstraction: you're showing them the wrong thing;
- Wrong encoding/interaction: the way you're showing the thing doesn't work;
- Wrong algorithm: your code is not adequate (e.g. too slow).



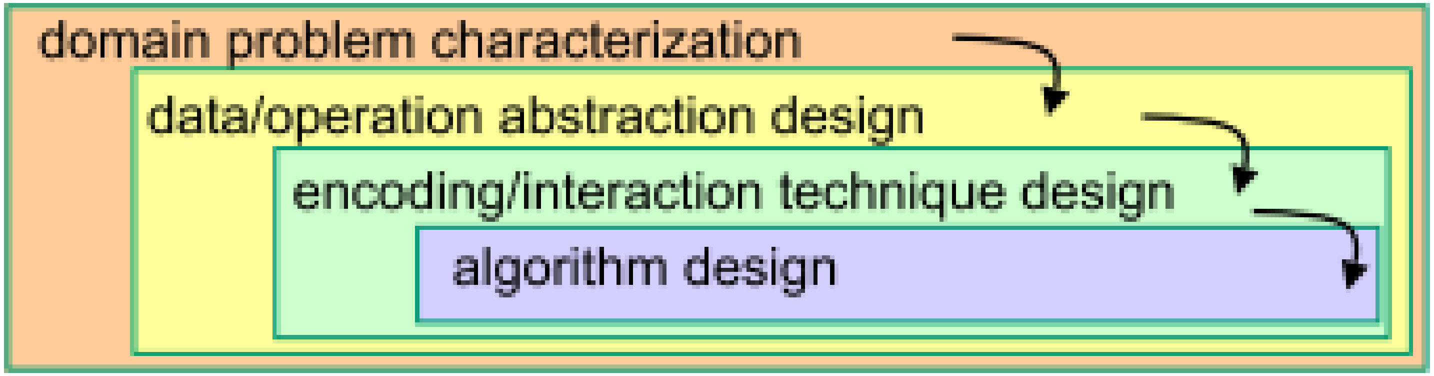
characterizing the problems of real-world users

abstracting into operations on data types

designing encoding and interaction techniques

creating algorithms to execute techniques

- Output from a level above is input to the level below
- An **upstream error inevitably cascades to all downstream levels**:
- E.g. a poor choice in the abstraction stage will not create a visualization system that solves the intended problem, even with perfect visual encoding and algorithm design .



- Evaluating Visualizations is **challenging**

- It will become more challenging as Visualization evolves to be more interactive, collaborative, distributed, multi-sensorial, mobile …

- It is **fundamental** to:
  - evaluate solutions to specific cases
  - develop new visualization methods / systems
  - establish guidelines

- i.e**. to make Visualization more useful, more usable, and more used**

Usability Evaluation Bibliography - Papers

- Carpendale, S., “Evaluating Information Visualizations,” in *Information Visualization , Human-centered issues and perspectives*, A. et al. Karren, Ed. Springer, 2008, pp. 19–45.
- Cockton, G. Usability Evaluation. In: Soegaard, Mads and Dam, Rikke Friis (eds.), *The Encyclopedia of Human-Computer Interaction*, 2nd Ed, 2013, Aarhus, Denmark: The Interaction Design Foundation, 2013

  http://www.interaction-design.org/encyclopedia/usability_evaluation.html
- Zuk, T., L. Schlesier, Neumann, P., Hancock, M. and Carpendale, S. Heuristics for Information Visualization Evaluation,” in *First Workshop on Beyond Time and Errors Novel Evaluation Methods for Visualization BELIV’06*, 2006, pp. 1–6.