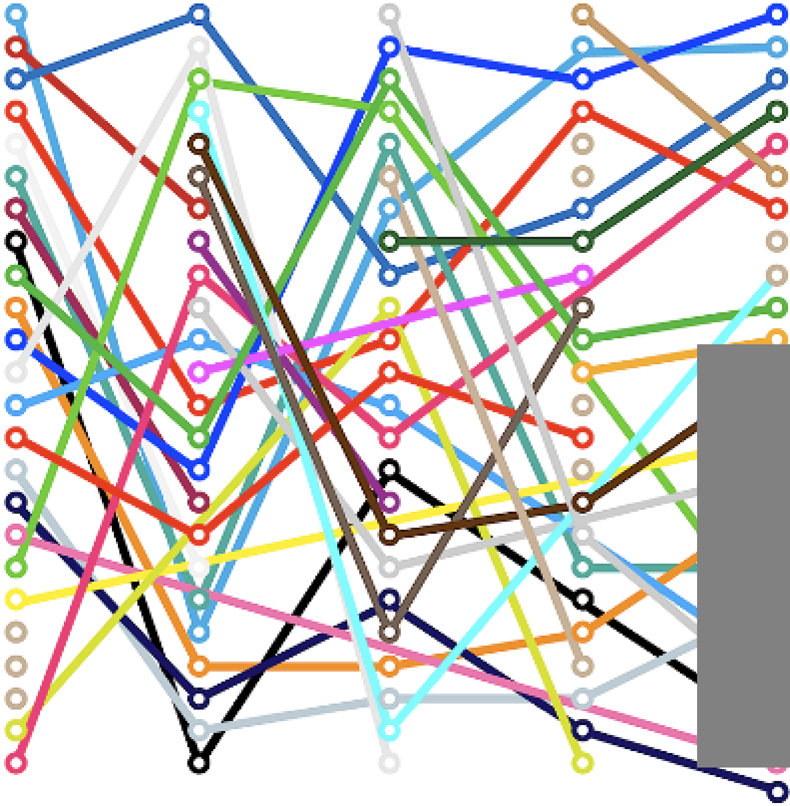


VU 188.305 – VO: 2 hrs. / 3 ECTS

Information Visualization

Human-Centered Visualization Design & Evaluation

Markus Bögl

Markus.boegl@tuwien.ac.at

www.cvast.tuwien.ac.at/~boegl

Vienna University of Technology

Institute of Visual Computing & Human-Centered Technology

FAKULTÄT FÜR INFORMATIK

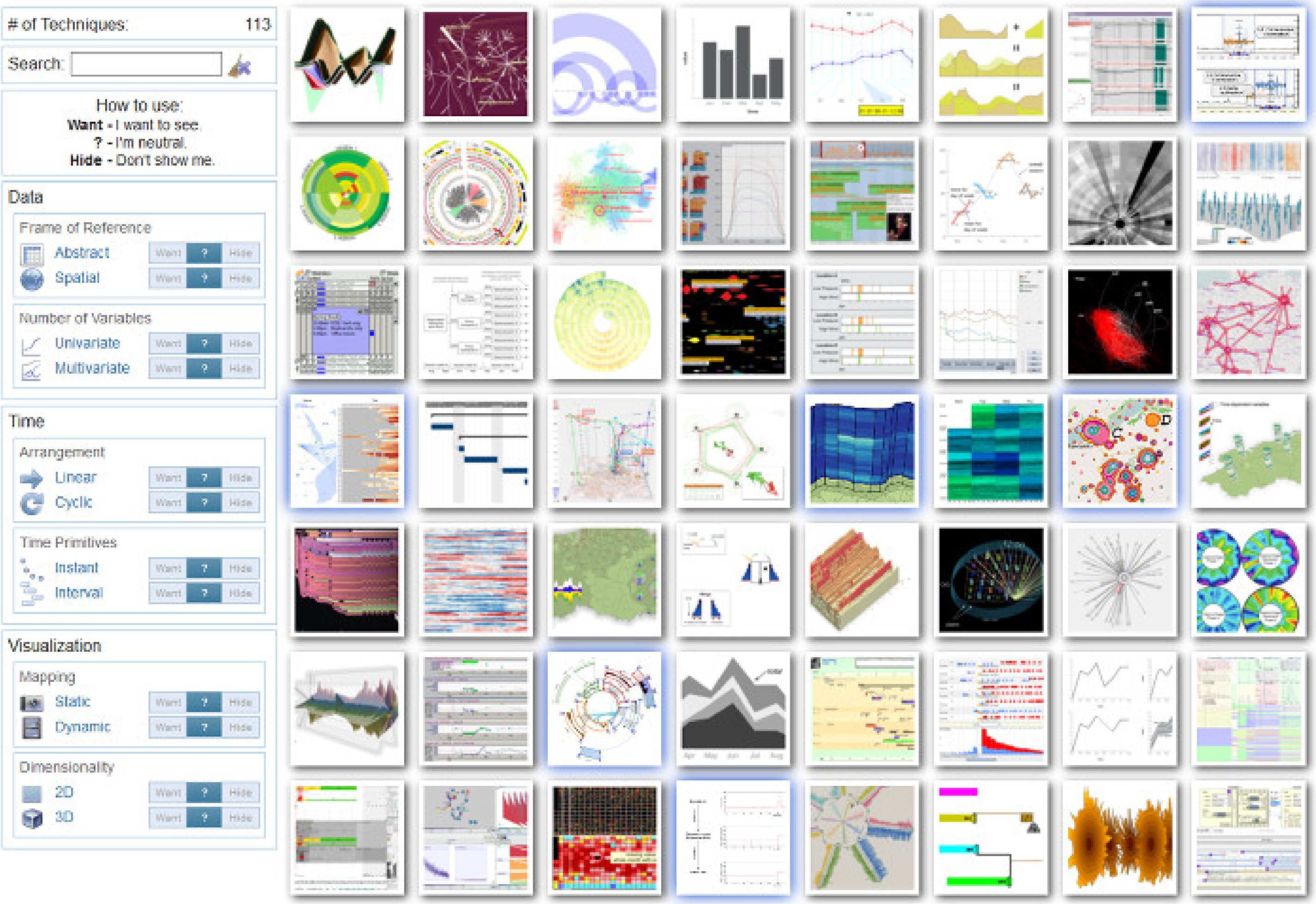
Faculty of Informatics

Laura Bassi Centre of Expertise

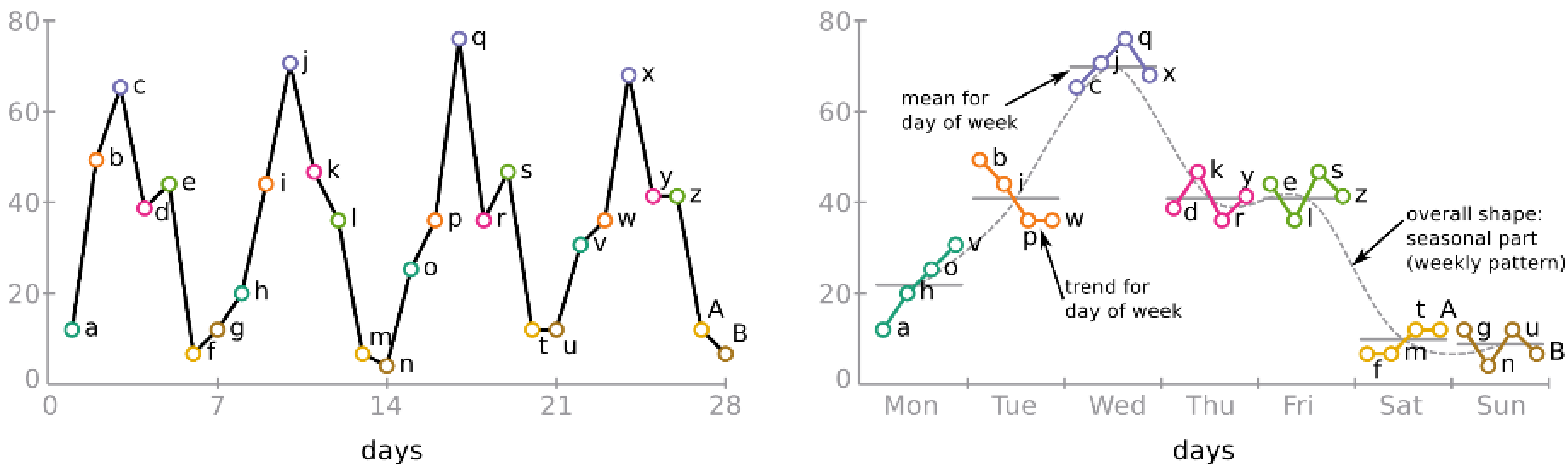
Centre for Visual Analytics Science & Technology

Motivating Example – Time Series Data

Which Time Series Representation to use?



Cleveland's Cycle Plot



Cleveland (1993)

Aigner et al. (2011)

Defining visualization (vis)

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why?...

Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

don’t need vis when fully automatic solution exists and is trusted

many analysis problems ill-specified

- don’t know exactly what questions to ask in advance

possibilities

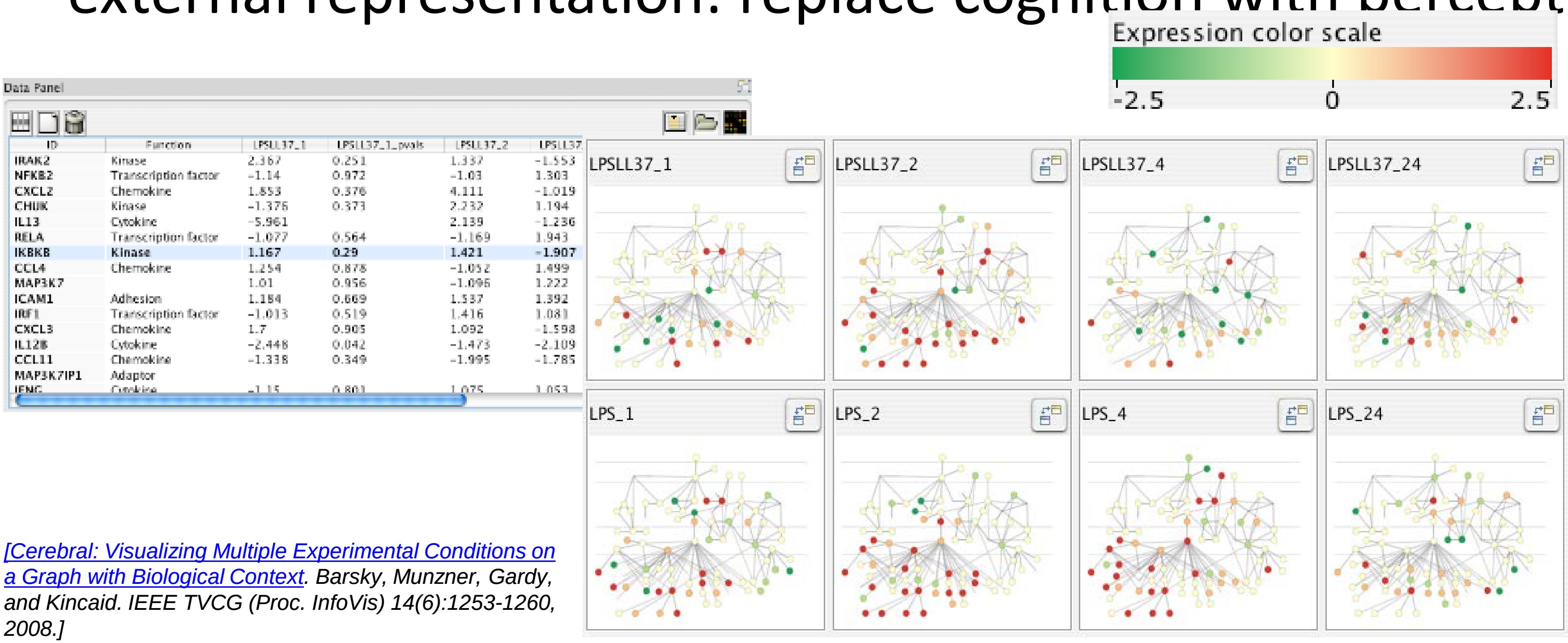
- long-term use for end users (e.g. exploratory analysis of scientific data)
- presentation of known results
- stepping stone to better understanding of requirements before developing models
- help developers of automatic solution refine/debug, determine parameters
- help end users of automatic solutions verify, build trust

Slides from Tamara Munzner’s Mini Course
“Visualization Analysis & Design” 2014

Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

external representation: replace cognition with perception



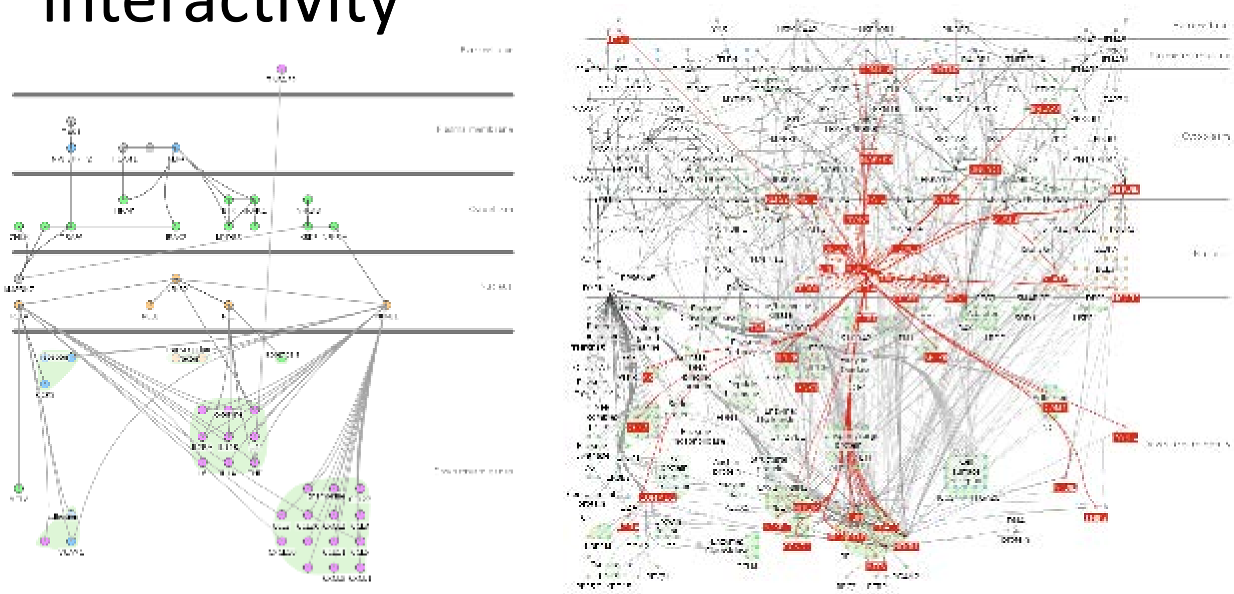
ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37
IRAK2	Kinase	2.367	0.251	1.337	-1.553
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303
CXCL2	Chemokine	1.853	0.376	4.111	-1.019
CHUK	Kinase	-1.376	0.373	2.232	1.194
IL13	Cytokine	-5.961		2.139	-1.236
RELA	Transcription factor	-1.077	0.564	-1.169	1.943
IKBKB	Kinase	1.167	0.29	1.421	-1.907
CCL4	Chemokine	1.254	0.878	-1.052	1.499
MAP3K7		1.01	0.956	-1.096	1.222
ICAM1	Adhesion	1.184	0.669	1.537	1.392
IRF1	Transcription factor	-1.013	0.519	1.416	1.081
CXCL3	Chemokine	1.7	0.905	1.092	-1.598
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785
MAP3K7IP1	Adaptor				
IFNG	Cytokine	-1.15	0.801	1.075	1.053

[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why have a computer in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

beyond human patience: scale to large datasets, support interactivity



[[Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation](#). *Barsky, Gardy, Hancock, and Munzner. Bioinformatics 23(8):1040-1042, 2007.]*

Why depend on vision?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- human visual system is high-bandwidth channel to brain
  - overview possible due to background processing
    - subjective experience of seeing everything simultaneously
    - significant processing occurs in parallel and pre-attentively
- sound: lower bandwidth and different semantics
  - overview not supported
    - subjective experience of sequential stream
- touch/haptics: impoverished record/replay capacity
  - only very low-bandwidth communication thus far
- taste, smell: no viable record/replay devices

Slides from Tamara Munzner's Mini Course "Visualization Analysis & Design" 2014

Why show the data in detail?

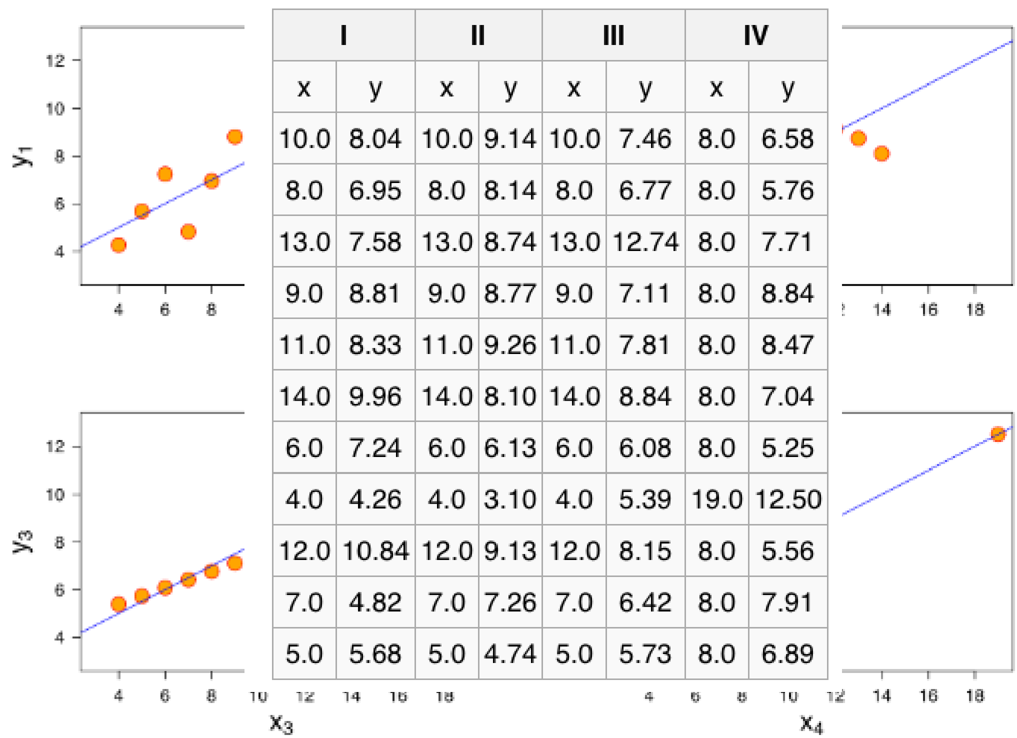
summaries lose information

confirm expected and find unexpected patterns

assess validity of statistical model

Anscombe's Quartet

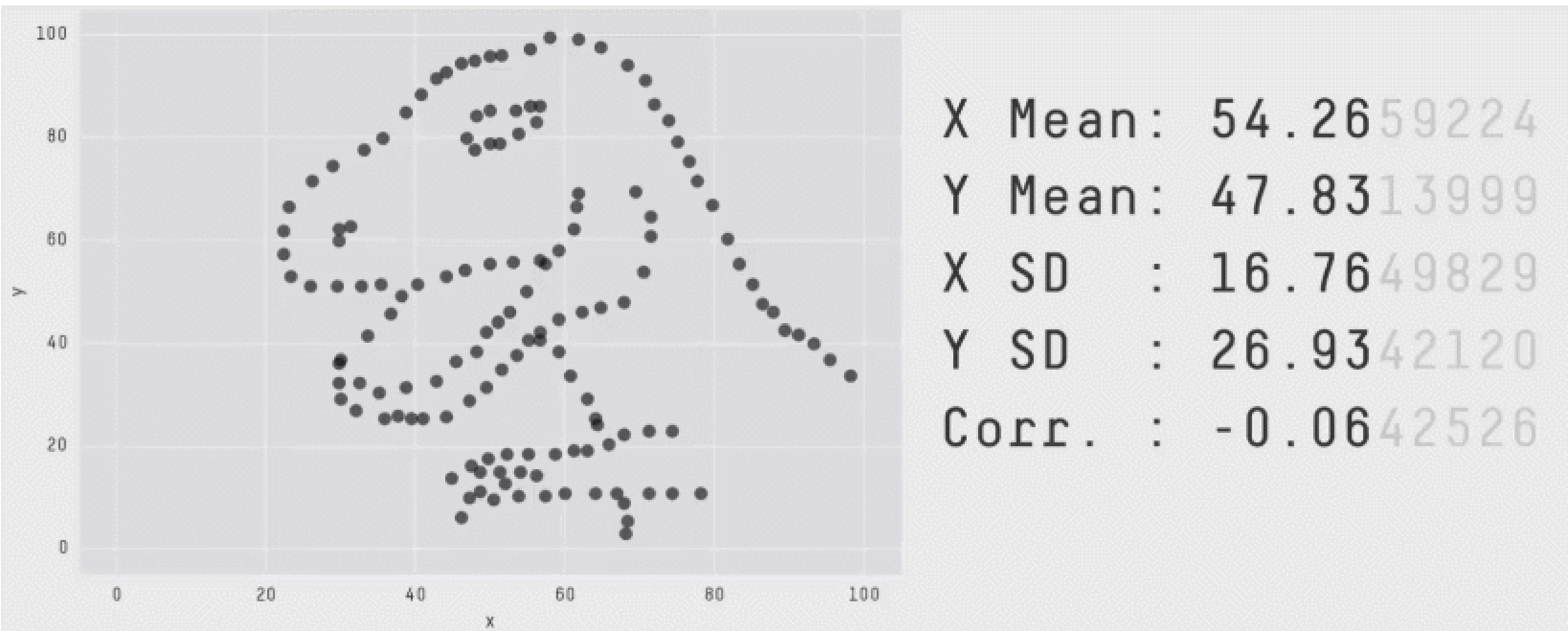
Identical statistics	
x mean	9
x variance	10
y mean	8
y variance	4
x/y correlation	1



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

Slides from Tamara Munzner's Mini Course
"Visualization Analysis & Design" 2014

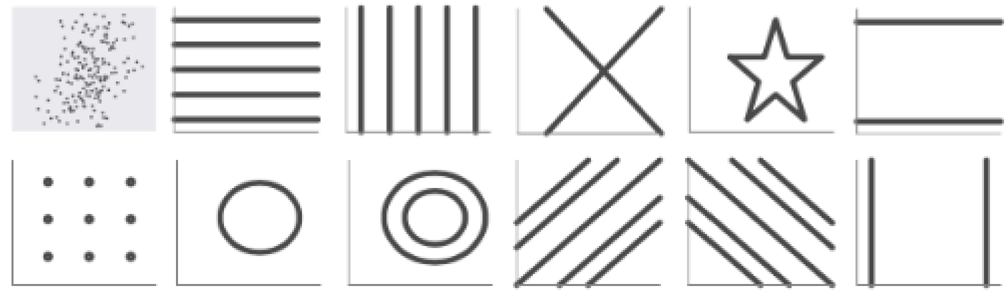
Datasaurus - Same Stats, Different Graphs

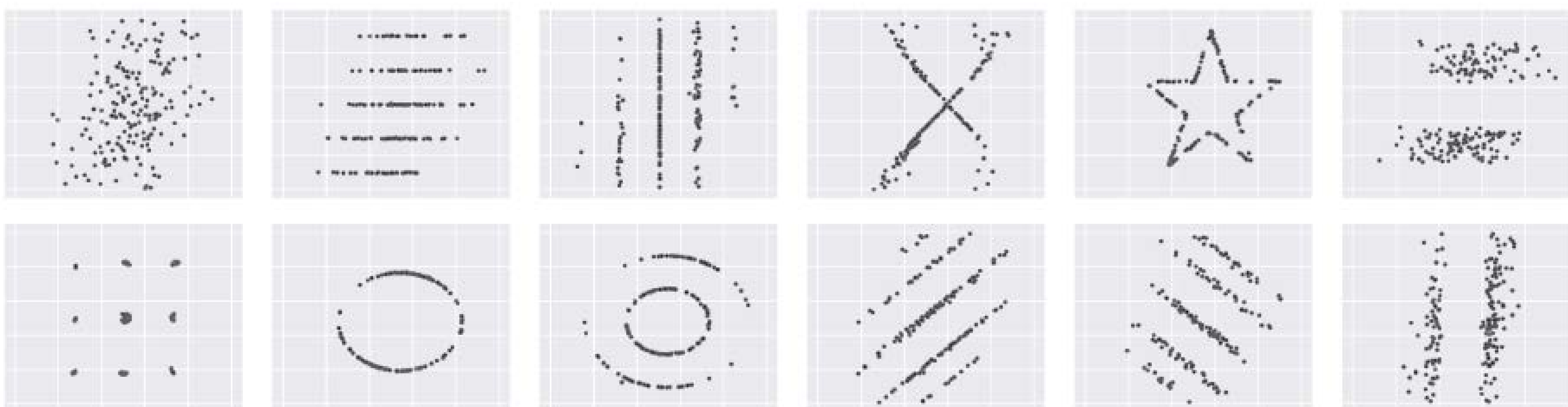


[Matejka and Fitzmaurice, CHI 2017]

Same Stats, Different Graphs

$\overline{x}$ =54.26
$\overline{y}$ = 47.83
sdx = 16.76
sdy = 26.93
Pearson’s Correlation
r = -0.06



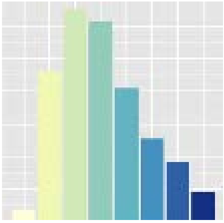


[Matejka and Fitzmaurice, CHI 2017]

Idiom design space

The design space of possible vis idioms is huge, and includes the considerations of both how to create and how to interact with visual representations.

idiom: distinct approach to creating or manipulating visual representation

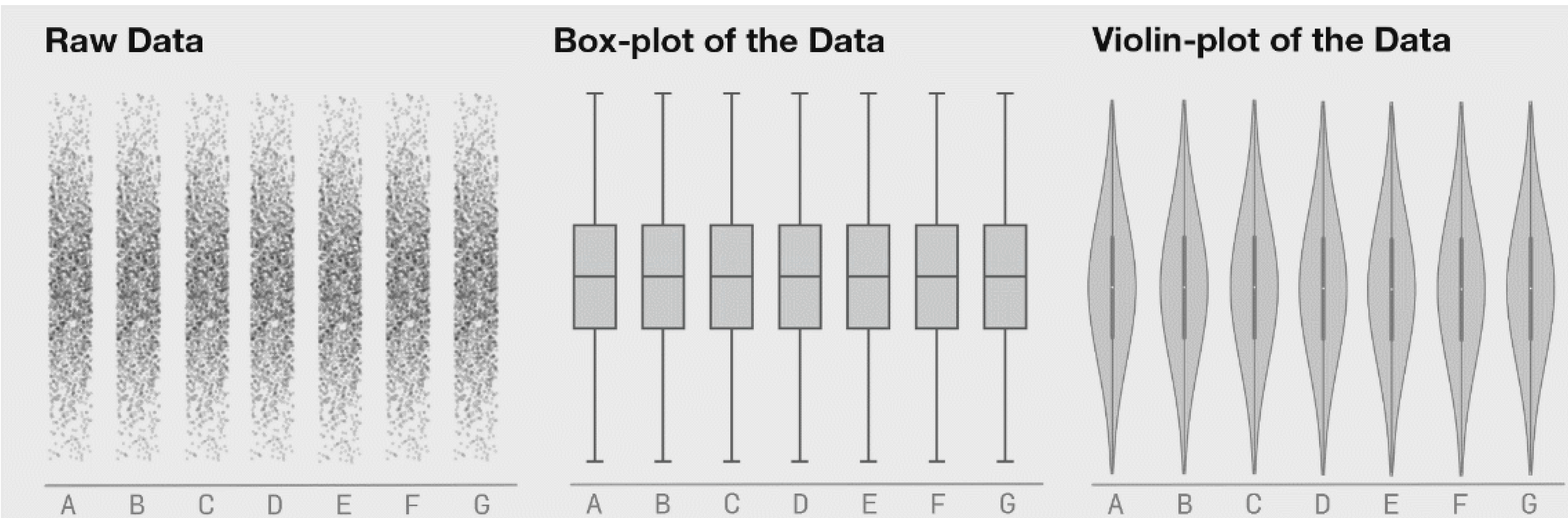


- how to draw it: **visual encoding** idiom
  - many possibilities for how to create
- how to manipulate it: **interaction** idiom
  - even more possibilities
    - make single idiom dynamic
    - link multiple idioms together through interaction

[A layered grammar of graphics. Wickham. Journal of Computational and Graphical Statistics 19:1 (2010), 3–28.]

[Interactive Visualization of Large Graphs and Networks. Munzner. Ph.D. thesis, Stanford University Department of Computer Science, 2000.]

More on Same Stats, Different Graphs



[Matejka and Fitzmaurice, CHI 2017]

Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- tasks serve as constraint on design (as does data)
  - idioms do not serve all tasks equally!
  - challenge: recast tasks from domain-specific vocabulary to abstract forms
- most possibilities ineffective
  - validation is necessary, but tricky
  - increases chance of finding good solutions if you understand full space of possibilities
- what counts as effective?
  - novel: enable entirely new kinds of analysis
  - faster: speed up existing workflows

Slides from Tamara Munzner’s Mini Course
“Visualization Analysis & Design” 2014

Resource limitations

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

- computational limits
  - processing time
  - system memory
- human limits
  - human attention and memory
- display limits
  - pixels are precious resource, the most constrained resource
  - **information density**: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space, find sweet spot between dense and sparse

Slides from Tamara Munzner's Mini Course
"Visualization Analysis & Design" 2014

User-Centered Design/Human-Centred Aspects

Stone, et al. 2005, pp 628 in Kerren, et al. 2007:

“An approach to user interface design and development that views the ***knowledge about intended users*** *of a system as a* ***central concern****, including, for example, knowledge about user’s abilities and needs, their task(s), and the environment(s) within which they work. These* ***users*** *would also be* ***actively involved*** *in the design process.”*

General idea

- Adapt to the user’s needs, skills, and limitations
- Engage users
- Adapt to the context
- Work in real life

User-Centered Design/Human-Centred Aspects

[Kerren, et al. 2007]

International Organization for Standardization (ISO) produced the standard ISO 13407 Human-Centered Design Processes for Interactive Systems

1. The active involvement of users in the design process and a clear understanding of them, their tasks, and their requirements.
2. An appropriate allocation of functions between users and technology, specifying which functions can be carried out by users.
3. An iteration of design solutions in which feedback from users becomes a critical source of information.
4. A multidisciplinary design perspective that requires a variety of skills. Multidisciplinary design teams should be involved in the human-centered design process. The teams should consist of end users, purchasers, business analysts, application domain specialists, systems analysts, programmers, as well as marketing and sales personnel.

User-Centered Design/Human-Centred Aspects

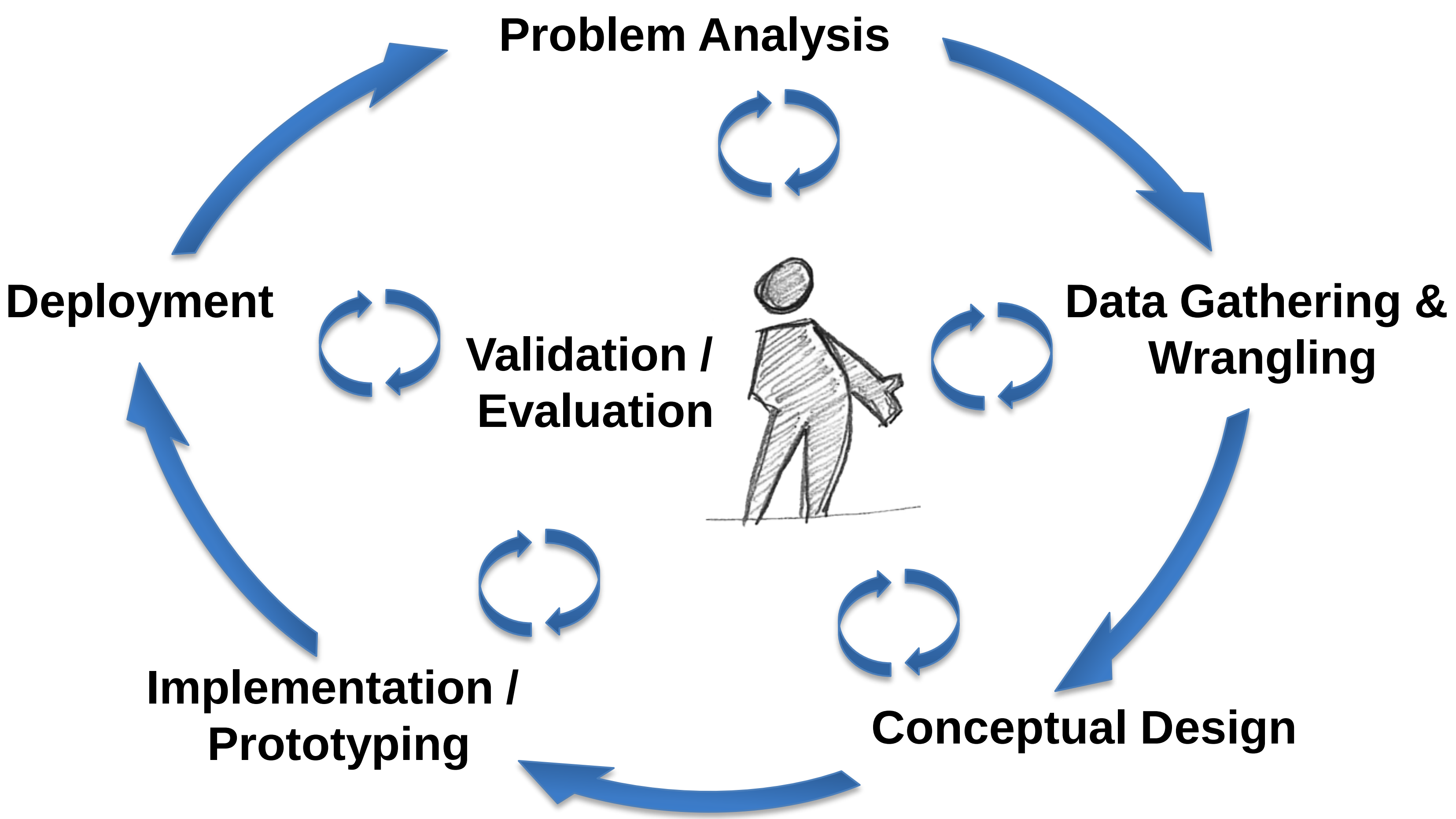
[Kerren, et al. 2007]

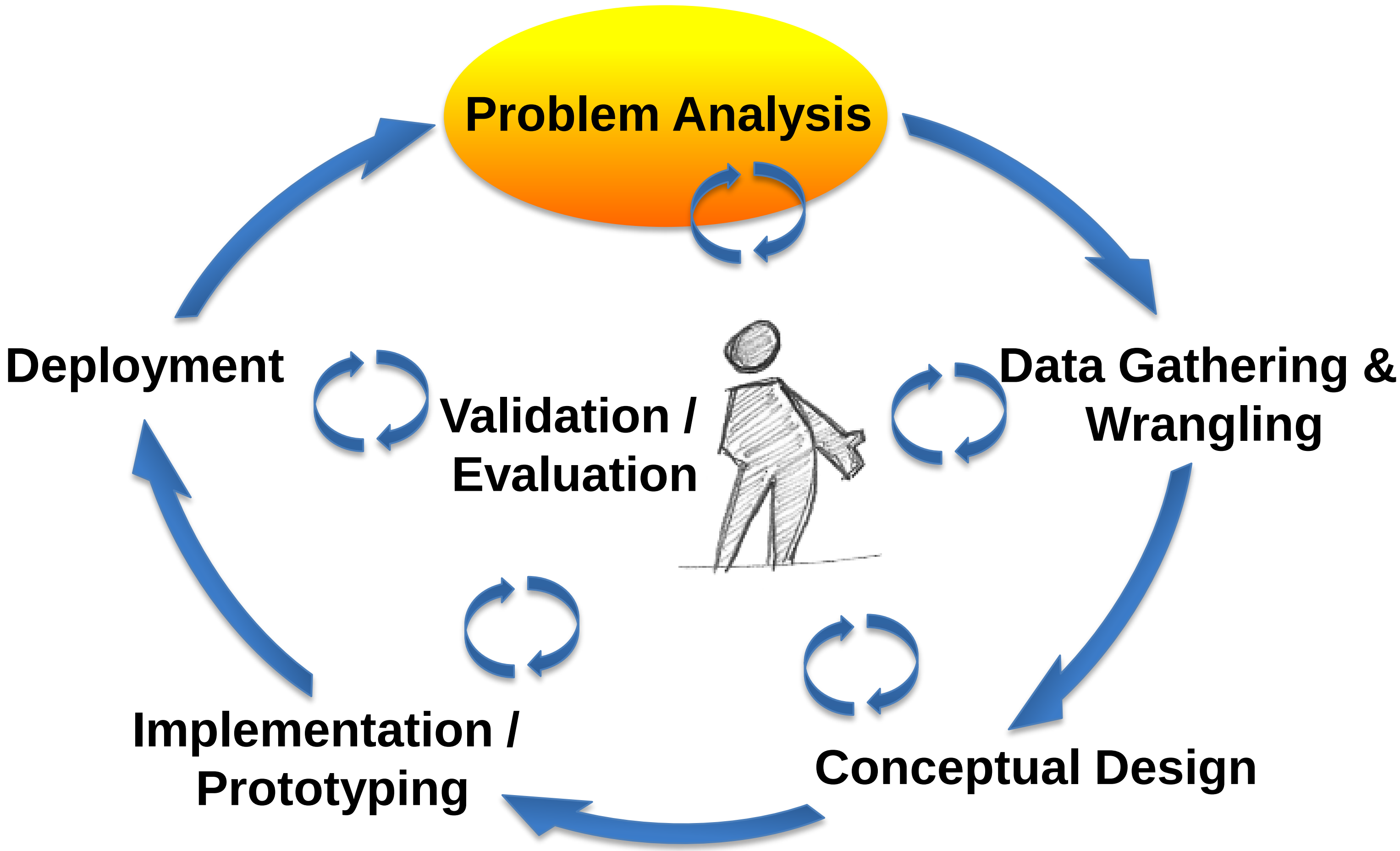
Benefits

1. Systems are easy to understand and use, thus reducing training and support costs.
2. Discomfort and stress are reduced, therefore the user's satisfaction is improved.
3. The productivity of users and the operational efficiency of organizations is improved.
4. Product quality, aesthetics, and impact are improved, therefore a competitive advantage can be achieved.

DESIGN CYCLE

Human-Centered Design Cycle





Problem Analysis

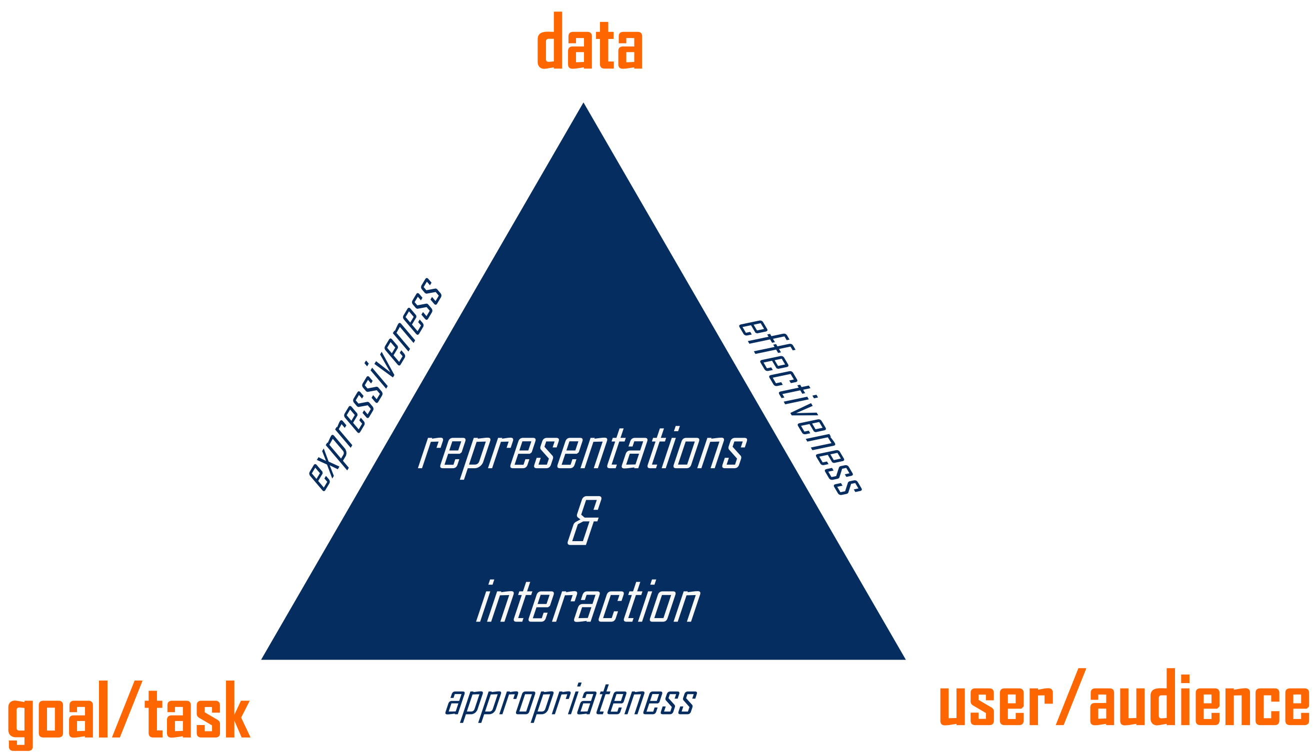
Users & Context

Data

Tasks & Goals

Requirements

Three central questions



Who are the users of the systems? (Users)
What kind of data are they working with? (Data)
What are the general tasks of the users? (Tasks)

Users & Context: Who are your users?

[Börner slides, Kulyk et al., 2007]

it is important to know

- who the users are,
- what their capabilities are,
- what kind of activities they do and
- context in which they work

Users

- Who is the intended **audience** (profession, location, age, or lifestyle preferences)?
  - e.g., administrator, physician, child, etc.
- What is their level of **technical & subject expertise**?
  - Visual language used has to match the user's understanding of its function and/or content. e.g., known visualization techniques, SW users are familiar with
- Do users have **information preferences**?
  - Which pieces of information do users want first, second, third, and so on?
- Are there **metaphors / mental models** that are used?
- What are the user's **information needs/tasks**?
- Users with **disabilities**?
  - e.g., color-blindness, physical disabilities

Users & Context: Who are your users?

[Börner slides, Kulyk et al., 2007]

Context

- domain
  - vocuabulary - speak the users' language
  - e.g., medicine vs. petroleum industry
- physical
  - e.g., poor lighting, noise, sitting or standing in front of peripheral, kinds of interaction users are experienced with
- social
  - collaboration
    - Do users work in groups? (using their own computers? in front of a large screen? who is the one who has control?)
  - cultural and international diversity
- technical
  - e.g., hardware, number of colors, browser software , monitors & screen resolution

Data: What kind of data are users working with?

Which parameters / variables?

What scale types?

- e.g., nominal, ordinal, discrete, continuous, binary, etc.

What frame(s) of reference?

- e.g., space, time

Which structure(s)?

- e.g., multidimensional, tree/hierarchy, network/graph, etc.

Any specifics?

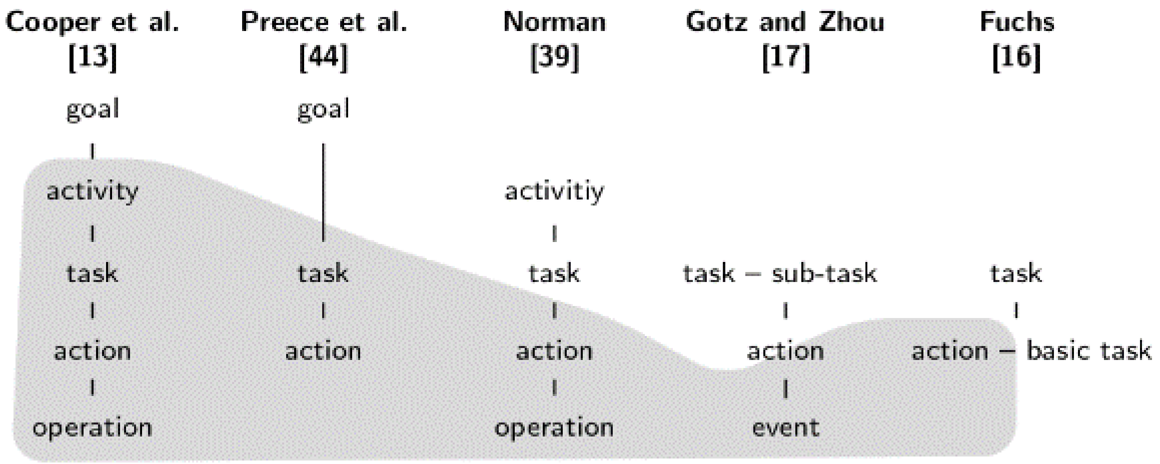
Amount of data

- How many data sets?
- Size of data sets?
- Number of elements?

Tasks & Goals: What do your users do?

*"Goals are not the same as tasks or activities. A **goal** is an **expectation of** an **end condition**, whereas both **activities and tasks** are **intermediate steps** (at different levels of organization) that help someone to reach a goal or a set of goals."*
(Cooper, Reimann, and Cronin, About Face 3, 2007)

Goals --> Activities --> Tasks --> Actions --> Operations (Norman & Cooper)



[Rind et al. 2015]

Task Analysis

research process of identifying which activities are performed by the user groups

How are tasks conducted currently?

- advantages
- disadvantages
- problems
    - what kind of problems do people currently have?
    - how can it be done better?
    - how do practitioners overcome these problems?
    - --> potential for improvements

Purpose

Exploration / Explorative Analysis

undirected search

no a priori hypotheses

get insight into the data

begin extracting relevant information

come up with hypotheses

Confirmation / Confirmative Analysis

directed search

verify or reject hypotheses

Presentation

communicate and disseminate analysis results

Requirements: How should it be?

[Kulyk et al., 2007]

identify the users' needs of the design

make users doing their job more efficient and enjoyable

functional

- what kind of activities do people want to do
- what kind of functionality the system should have or what the application should be able to do

technical

- embedding into an existing system
- data interface (e.g., files vs. database)
- basic architecture (e.g., server/client/client-server, online/live system)
- used technology (e.g., web-based, programming language, etc.)

usability

- quality measures like efficiency, effectiveness, safety, utility, learnability and memorability
- satisfaction goals like enjoyability, pleasurable, aesthetically pleasing, and motivation

Priorization: MoSCoW rules

[Benyon et al., 2005]

Must have—fundamental to the projects success

o

Should have—important but the projects success does not rely on these

Could have—can easily be left out without having impact on the project

o

Won't have (this time round)—can be left out in the current state, but can be added in later project increments

--> Validation / discussion of requirements with users

How?

User, Data & Tasks Analysis Methods

- Interviewing
- Questionnaire
- Ethnographic observation
- Participatory workshops / focus groups
- Task demonstration
- Document analysis

Useful resource:
http://www.usabilitynet.org/tools/methods.htm

User, Data & Tasks Analysis Methods

[Kulyk et al., 2007]

Interviewing

- time-consuming
- small selection of users
- important: high diversity of users (representative proportion of the target group); average users & expert users
- interviews rely on recall rather than direct capturing tasks

Questionnaire

- get statistical information or to get a public opinion
- can consist of open or closed questions
- disadvantages in contrast to interviews
  - not possible to ask for explanations
  - questions may be misunderstood
- careful design is necessary
- testing with a small pilot groups

Ethnographic observation

- observing the users' working environment in practice
- goals of the observation are to understand the users' subject, the visualization and interaction styles they are currently using, and how to improve the working environment
- can be very useful
- observer is not allowed to ask the target group to explain something since this will disrupt the practice
- possible problems
  - easy to misinterpret the observations
  - observation can disturb the actions that the target group is performing because they know that they are being observed
  - observer can overlook important information

User, Data & Tasks Analysis Methods

[Kulyk et al., 2007, Börner slides]

Participatory workshops / focus groups

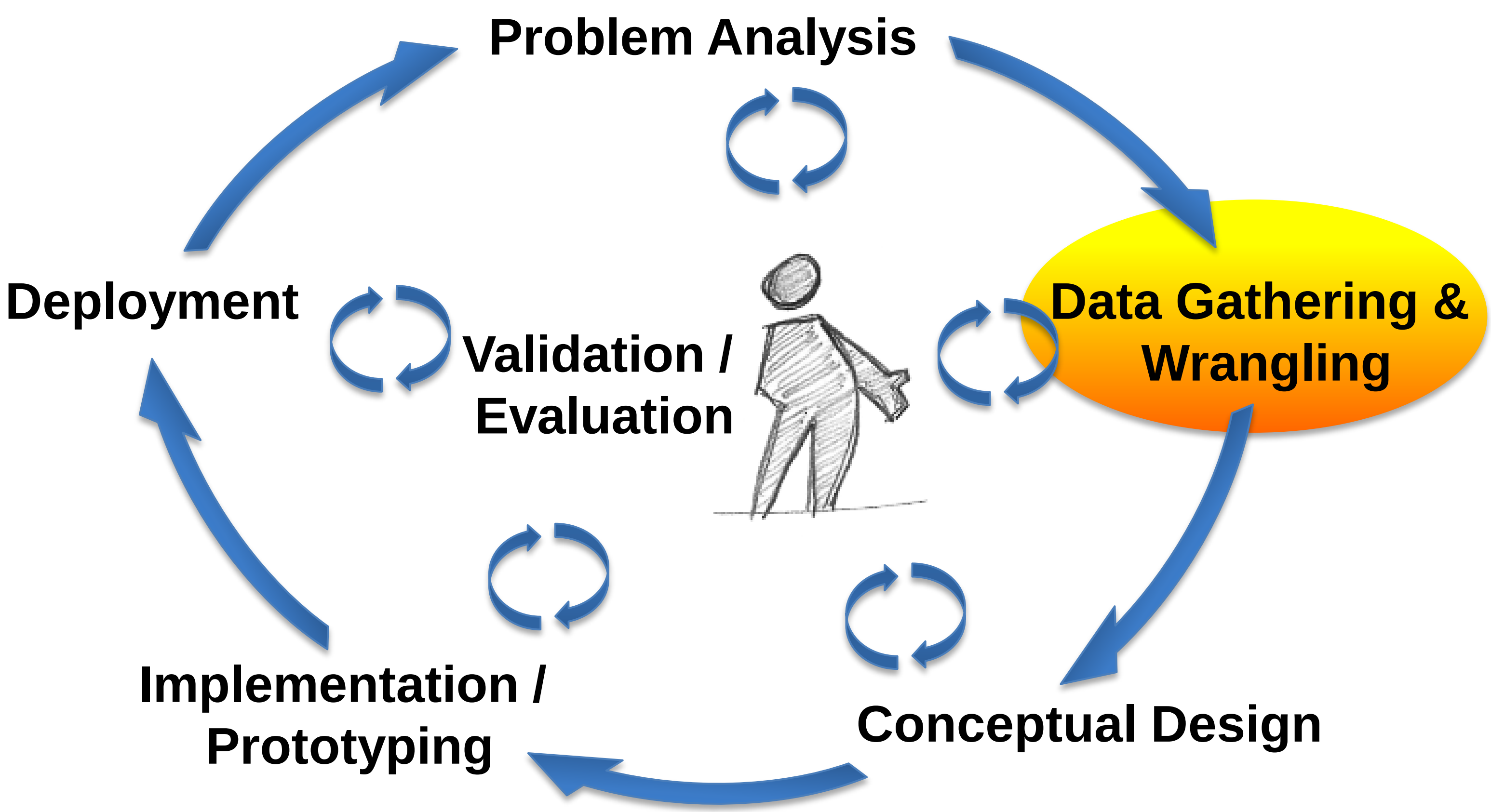
- organized for a specific focus group
- Clients, users, and designers meet each other and discuss issues and requirements of the system
- workshop can be structured or unstructured
- advantage: multiple viewpoints
- careful selection of the participants is essential

Task demonstration

- users demonstrating the task to the observer
- allows the observer to explain some actions in order to gain more insight
- task is described from the perspective of the observed users
- disadvantage
  - existing problems may not become visible during the observation, since most experienced users are not aware of these problems (anymore)
  - feedback may be very limited
  - could be the case that the tool to be demonstrated is discussed rather than demonstrated
- alternative possibility: giving the user a set of predefined tasks

Document analysis

- reviewing documentation of existing systems and processes, State-of-the-Art Research, Scientific literature, Commercial products



Data Gathering & Wrangling

Elephant in the room

http://www.fuelyourblogging.com/blogging-about-the-elephants-in-the-room/

Data Quality Problems

Missing data

no measurements, redacted, ...?

Duplicates

e.g., same person twice

Implausible values

e.g., age: 130

Wrong data

e.g., wrong format, misspellings, outliers

Ambiguous data

e.g., 05/04/2012 -- May 4 or April 5?

Data Integration

combining multiple sources

Data Usability, Credibility, Usefulness

Can I work with the data? (Is it usable)

- can be parsed and manipulated by computational tools

Do I trust the data? (Is it credible)

- suitably representative of a phenomenon to enable productive analysis

Can I learn from it? (Is it useful)

- usable, credible, and responsive to one's inquiry

[Heer lecture slides, 2011]

Data Transformation

Reformatting

e.g., date formats

Extraction

e.g., first name & last name out of string field

Erroneous value correction

e.g., removing outliers

Type conversion

e.g., zip code to lat-lon

Schema mapping

e.g., mapping schemata of different sources

[Kandel et al., 2011]

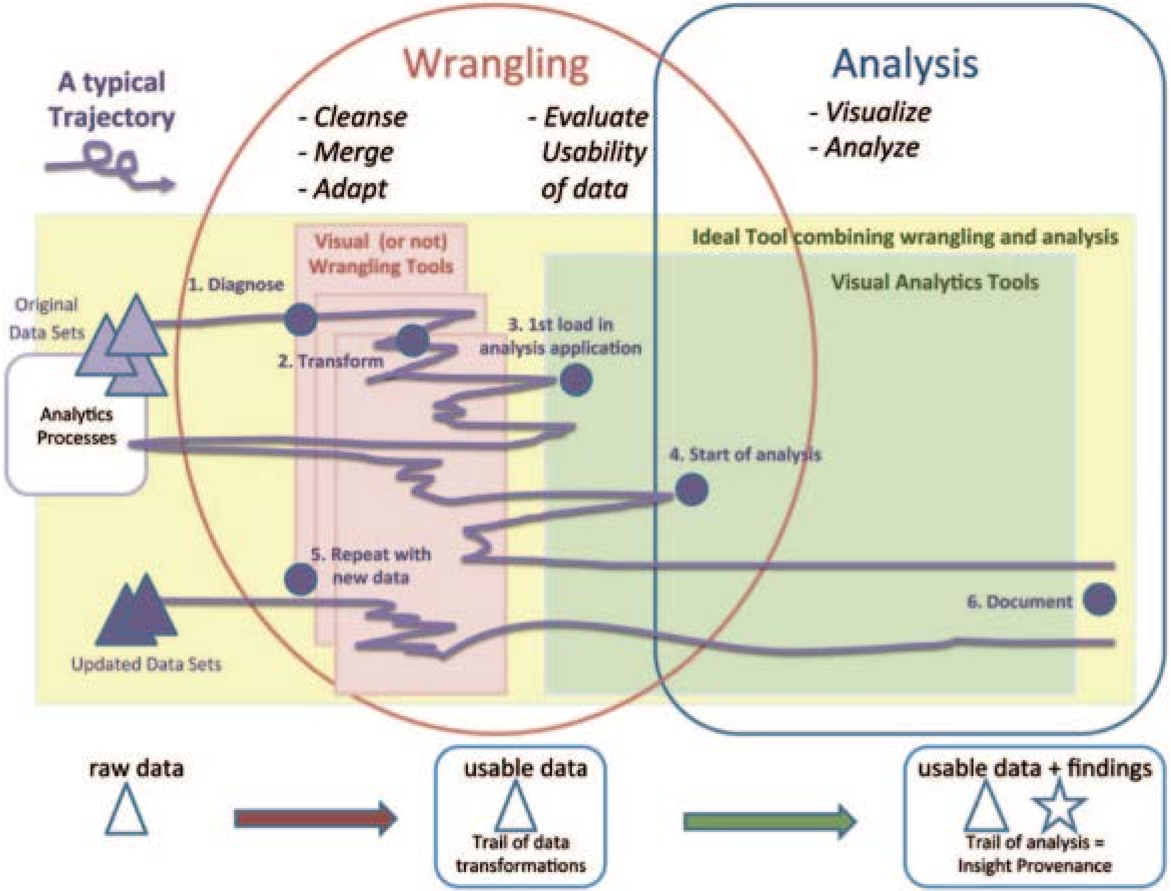
Data Wrangling

A process of iterative data exploration and transformation that enables analysis.

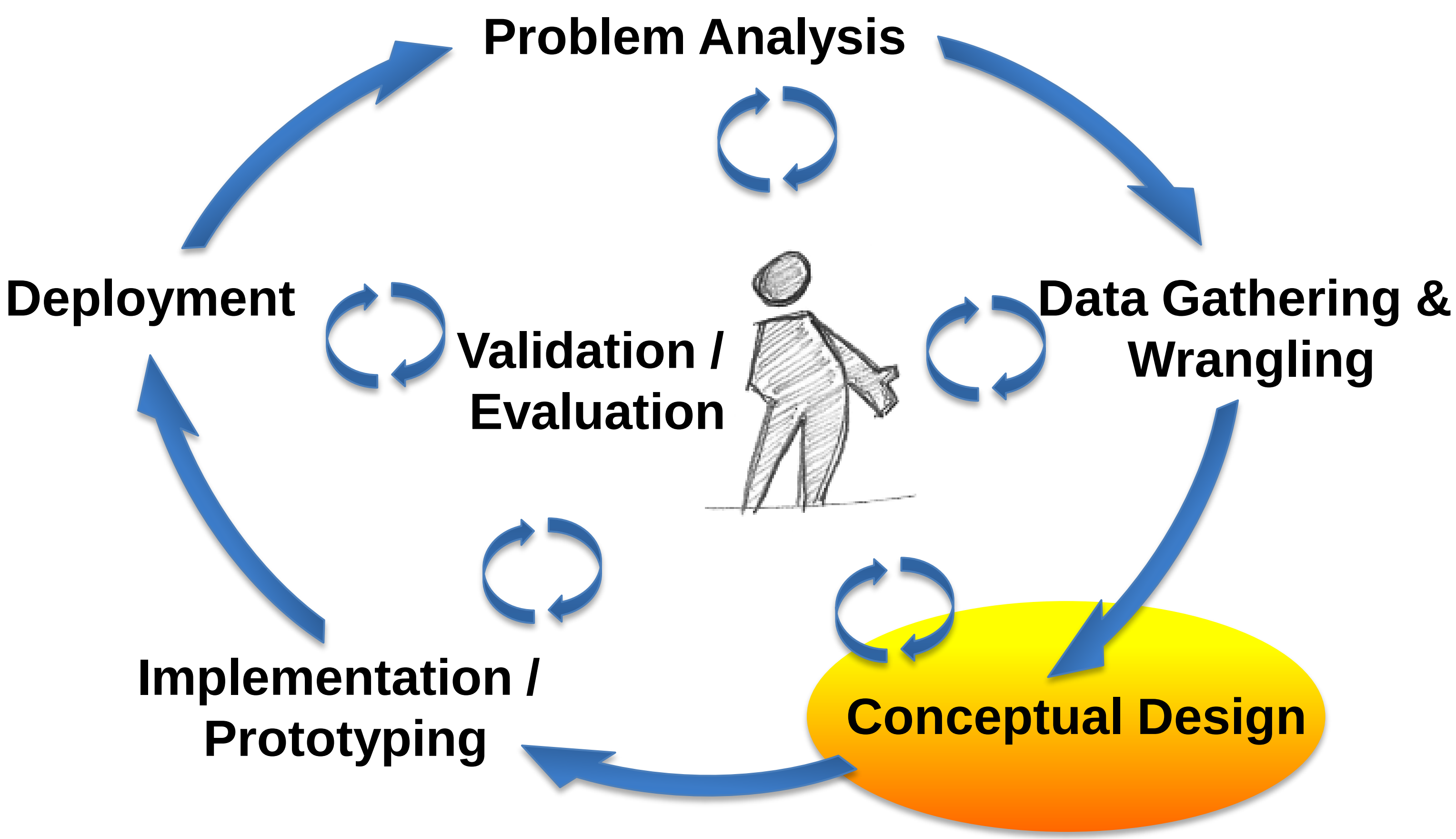
The goal of wrangling is to make data useful:

Map data to a form readable by downstream tools (database, stats, visualization, ...)

Identify, document, and (where possible) address data quality issues.



[Kandel et al., 2011]



Conceptual Design

Main components

User Interface

Visualization Techniques

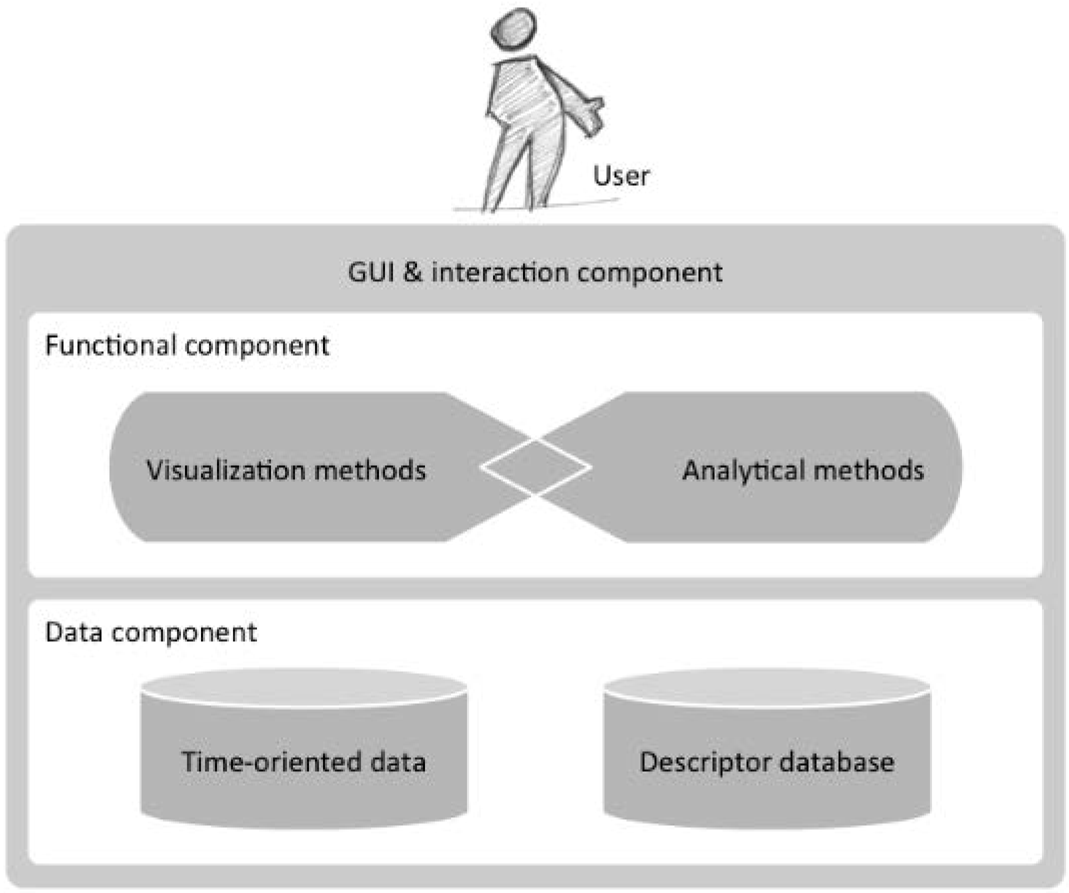
- Visual Mappings
- Metaphors
- Multiple Coordinated Views
- ...

Interaction Methods

- Navigation & Zooming
- Selection & Brushing
- Details-on-Demand
- Dynamic Querying
- ...

Analytical Methods

Data Model



Conceptual Frameworks

Visual Information Seeking Mantra (Shneiderman, 1996)

Task typologies

- Task list (Wehrend & Lewis, 1990)
- User Intents (Yi et al., 2007)

Visual Information Seeking Mantra

Overview: gain an overview of the entire dataset

Zoom: zoom in on data of interest

Filter: filter out uninteresting information

Details-on-demand: select data of interest and get details when needed

Relate: view relationships among data items

History: keep a history of actions to support undo and redo

Extract: allow extraction of data and of query parameters

[Shneiderman, 1996]

Task list

Locate (search for a known object)
Identify (object is not necessarily known previously)
Distinguish
Categorize
Cluster
Distribution
Rank
Compare within entities
Compare between relations
Associate
Correlate

[Wehrend & Lewis, 1990]

User intents

show me something else (**explore**)

show me a different arrangement (**reconfigure**)

show me a different representation (**encode**)

show me more or less detail (**abstract/elaborate**)

show me something conditionally (**filter**)

show me related items (**connect**)

mark something as interesting (**select**)

let me go to where I have already been (**undo/redo**)

let me adjust the interface (**change configuration**)

[Yi et al., 2007]

Methods

Sketches

- range from hand-drawn paintings to electronic drawings using painting/vector graphics tools, from cardboard modeling to 3D computer graphics
- advantages
  - complete control about the visualization
  - require very little technical support
  - designer is not limited by technical tools
  - designer does not have to focus on the tools for creating visualizations, but can put his focus completely on his inspiration

Screen prototypes

[Kulyk et al., 2007]

- usually created with tools
- consist of real software components, but it is not possible to interact with the prototype
- prototypes are more like screenshots of the products

Functional prototypes

- look like real products
- user can interact with the product
- horizontal: as much functionality as possible in the prototypes, with a limited set of options
- vertical: little functionality in, but the functionality is highly configurable

Evaluation of conceptual design

Expert review

Cognitive walkthrough

Heuristic evaluation

Participatory workshop / focus group

User testing (Wizard of Oz)

Good design …

Is **thorough** to the last detail, nothing is arbitrary or left to chance

Is primarily about **usability**

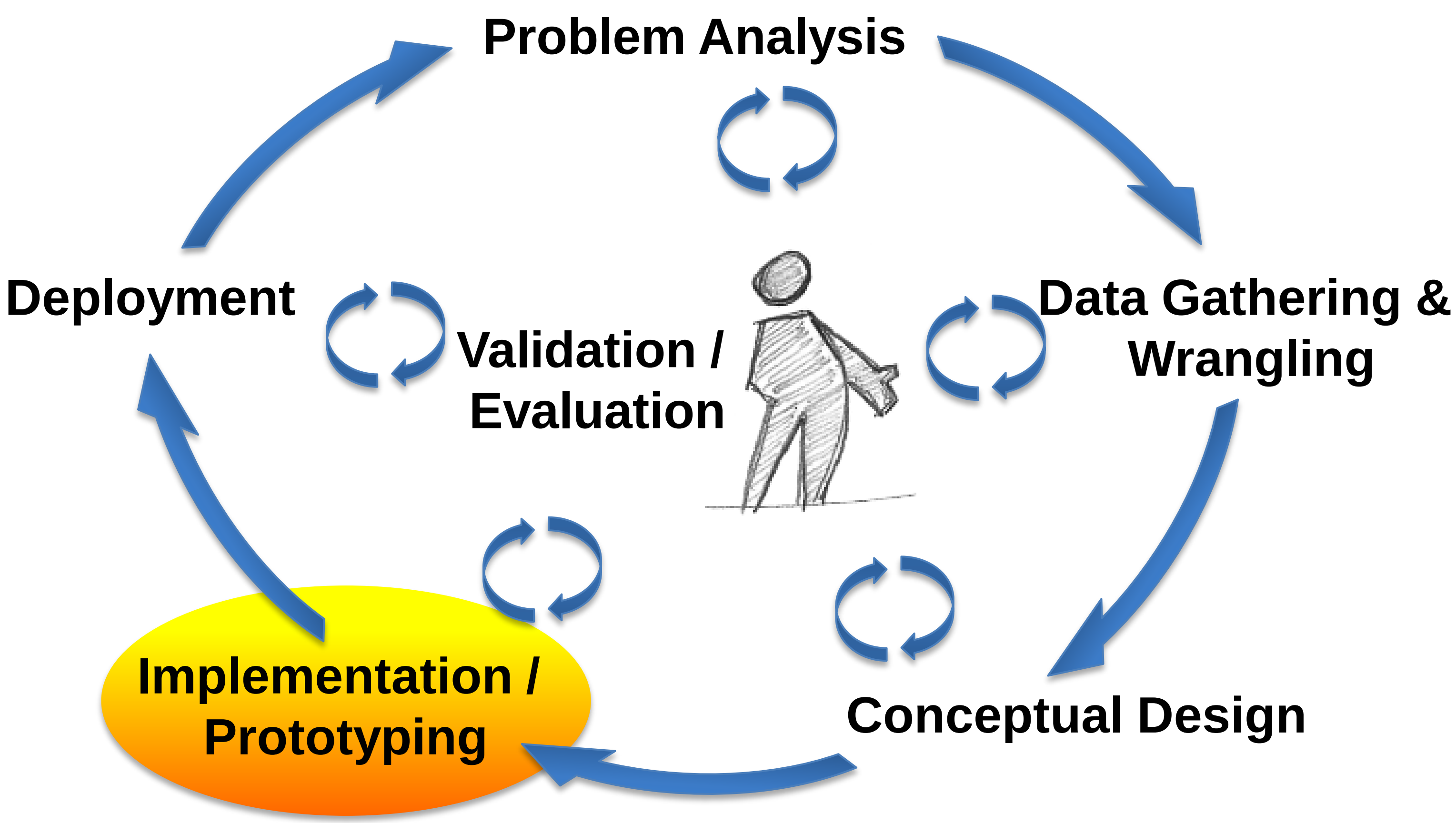
Is minimalistic: we start with an "xmas tree" but keep removing details

Involves taste, creativity, talent, aesthetic, inspiration: **you can learn it**!

Is **user-centered:**

Think as a **user** Act as a **user** Be a **user**

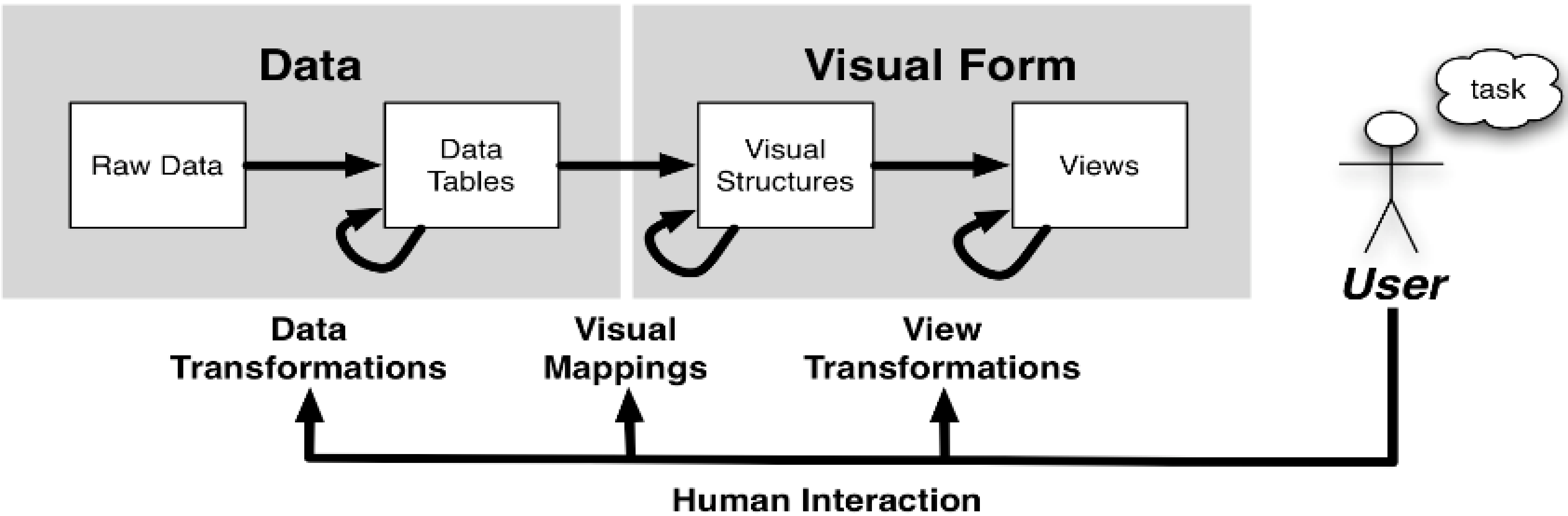
J. Van Wijk IEEE VIS 2013 Keynote
http://vimeo.com/80334651



Implementation / Prototyping

InfoVis Reference Model

[Card et al., 1999]



Raw Data: idiosyncratic formats
Data Transformations: Mapping raw data into an organization appropriate for visualization
Data Tables: relations (cases by variables) + metadata
Visual Mappings: Encoding abstract data into a visual representation
Visual Structures: spatial substrates + marks + graphical properties
View Transformations: Changing the view or perspective onto the visual presentation
Views: graphical parameters (position, scaling, clipping, ...)
Human Interaction: User influence at any level

Off-the-shelf Software vs. Implementation

[Börner slides]

Effort vs. flexibility

BUT: Don’t base your decisions on

- Availability of software tools.
- Personal interest/preferences for tools.

Libraries / Toolkits

Various JavaScript Libraries

- d3.js, jQuery, Node.js, Bootstrap, React, ...

Improvise

- Java; http://www.cs.ou.edu/~weaver/improvise/index.html

Vega

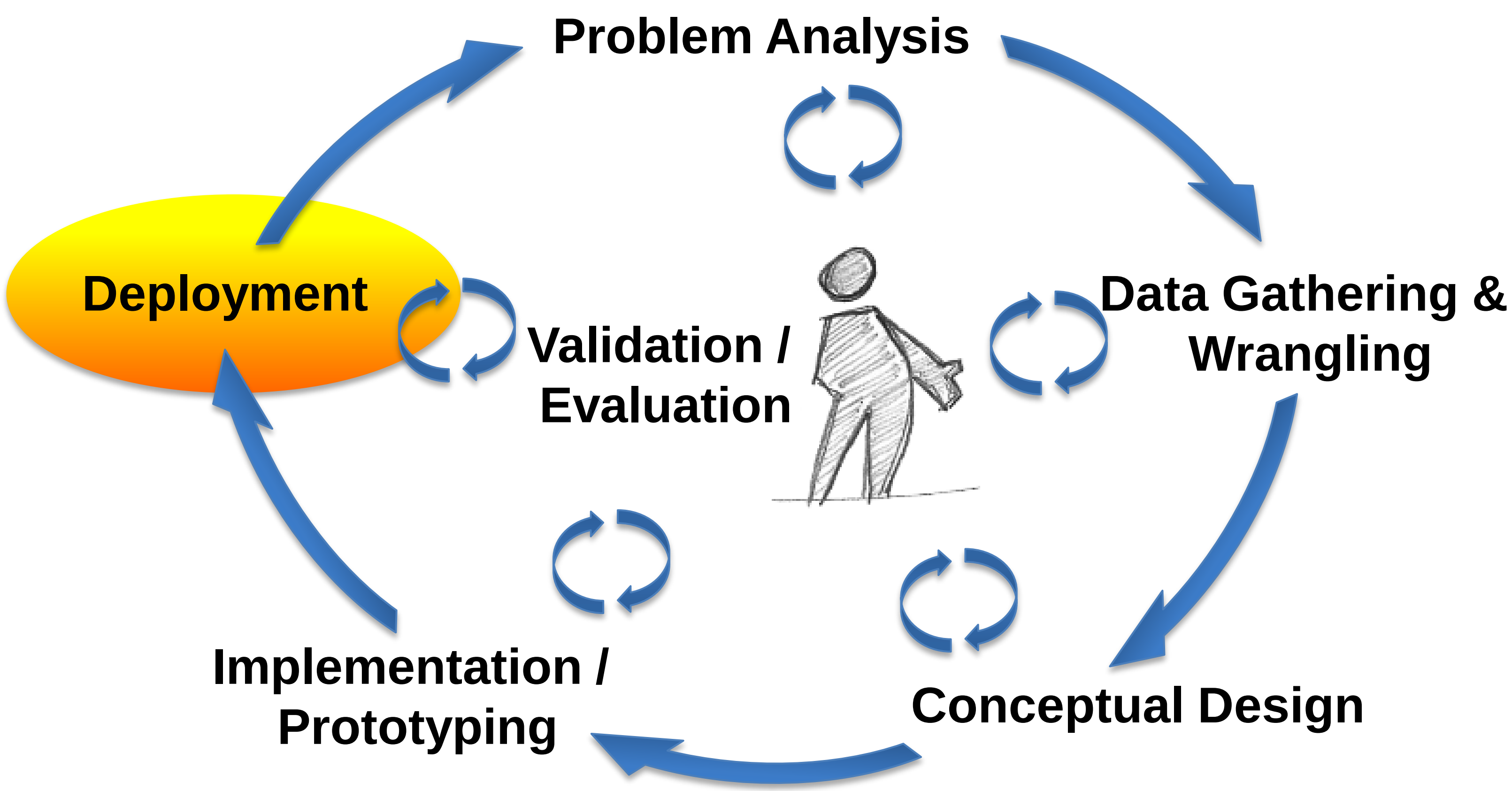
- visualization grammar, a declarative language for creating, saving, and sharing interactive visualization designs; https://vega.github.io/vega/

Vega-Lite

- high-level grammar of interactive graphics, for rapidly generating visualizations; https://vega.github.io/vega-lite/

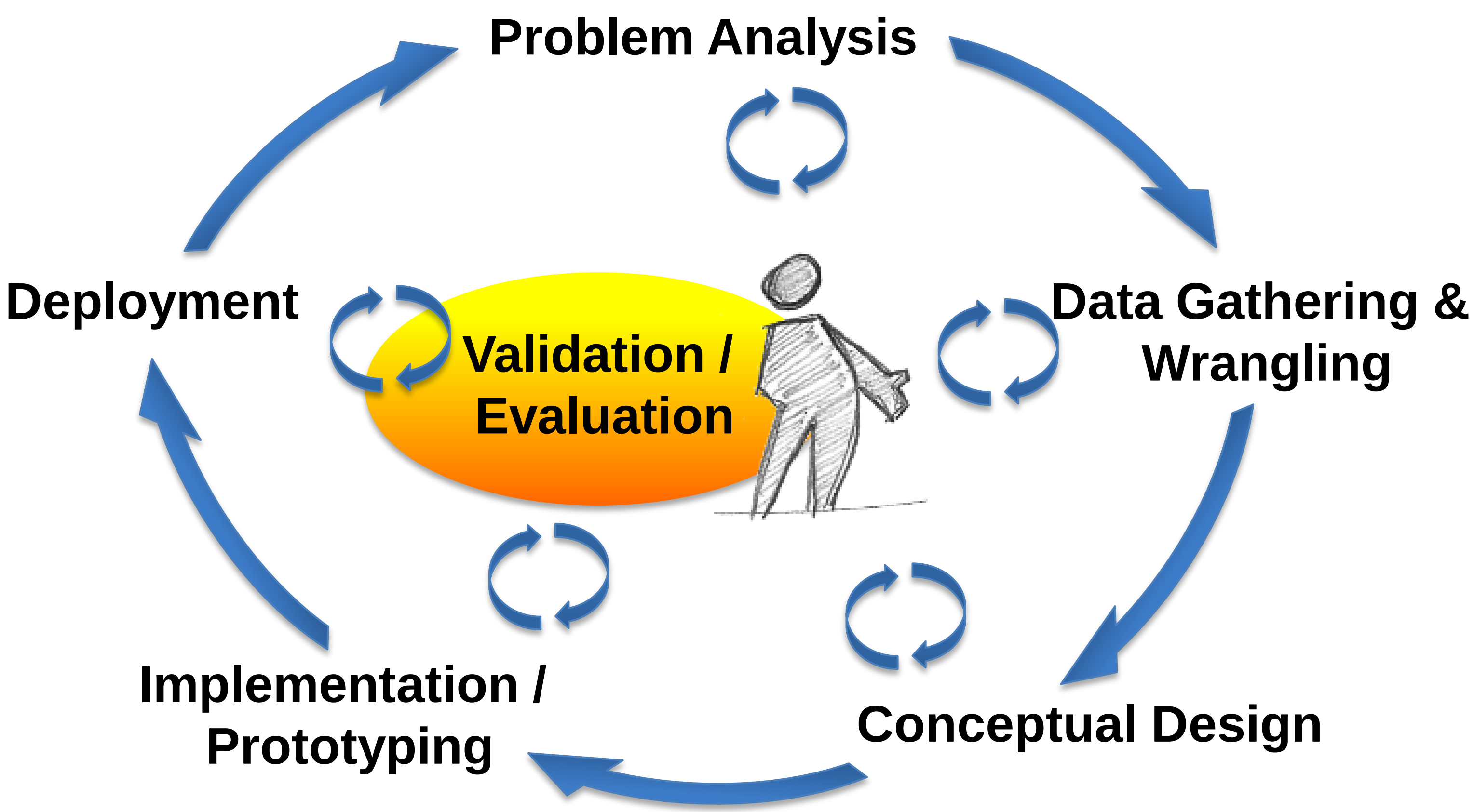
More @ InfoVis:Wiki

- http://www.infovis-wiki.net/index.php?title=Toolkit_Links
- http://www.infovis-wiki.net/index.php?title=Software_Links_%28InfoVis_Applications%29



Deployment

deploying a tool and gathering feedback about its use in the wild



Validation / Evaluation

Why?

[Kulyk et al., 2007]

Every design needs to be tested to determine how well the visualization fits its intended purpose and meets user requirements.

help to diagnose the usability problems and errors that can be an input for optimization of visualization

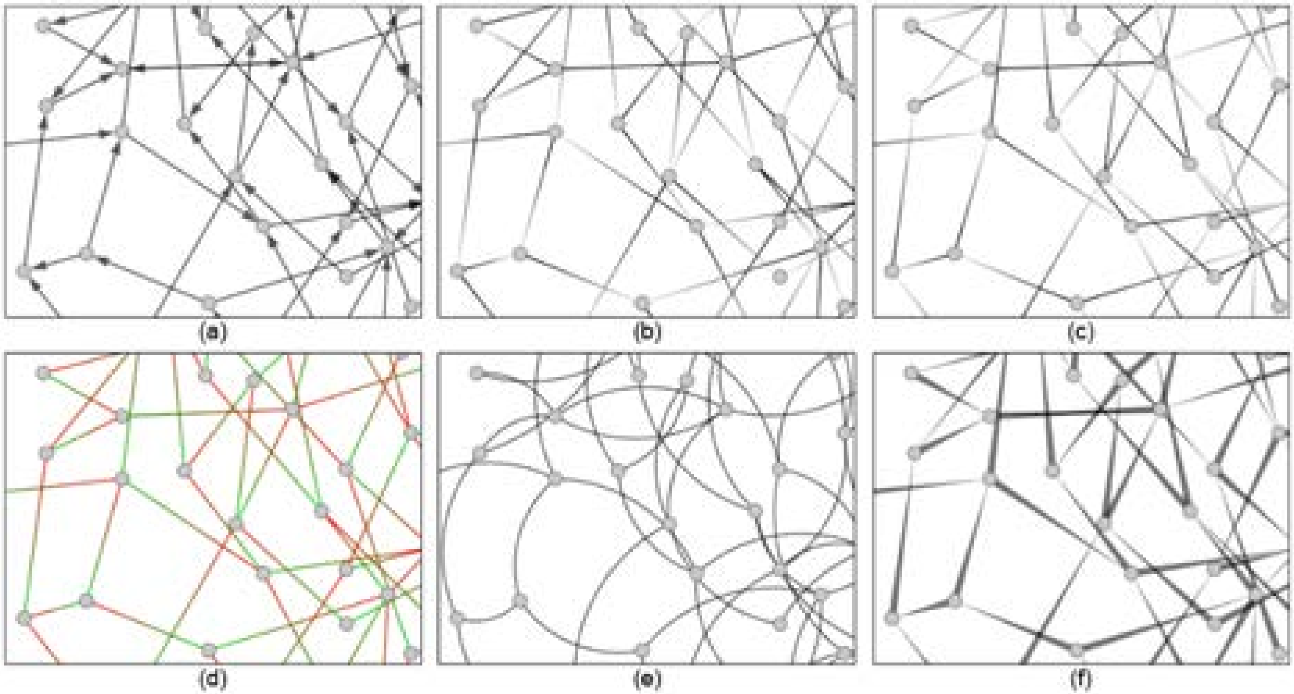
valuable for testing the efficiency of interactions with visualization

valuable input for improvement of the data representation

check whether a future visualization product will be adopted by the target audience

Example

Visualizing directed edges in graphs



Which one works best?

[Holten & van Wijk, CHI '09]

Elements of a successful visualization system

[Kulyk et al., 2007]

support the tasks the user wants to perform
- functional with respect to the tasks a user wants to perform

acceptance of the application among the whole target group
- easy to use
- easy to learn

data should be easy to explore
- frequent user should be able to explore the data, make certain details visible or hide some information

effectiveness
- Did the user extract the information he was searching for?

expressiveness
- consistency of the representation

subjective satisfaction
- user likes to use the application to solve his/her research problem and thinks that this application is helpful to him/her

It is often not possible to design a single visualization that scores high on all factors --> trade-offs; multiple views

User-Centered Design

[Kulyk et al., 2007]

„The main motivation to do evaluation before actual implementation is: **the earlier the evaluation takes place, the more successful is the design**. This increases the chances of the visualization to be adopted by the target users. Usability evaluation should be carried out throughout the whole process of an interactive application design. Therefore, the design and development of any system should be done ideally in an ***iterative way***. This means that the design process should **include several iterations of analysis, design, and evaluation**.“

Resource limitations – Human Limits Revisit

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

human limits: attention and memory

Examples for limits of human attention:

Perceptual blindness

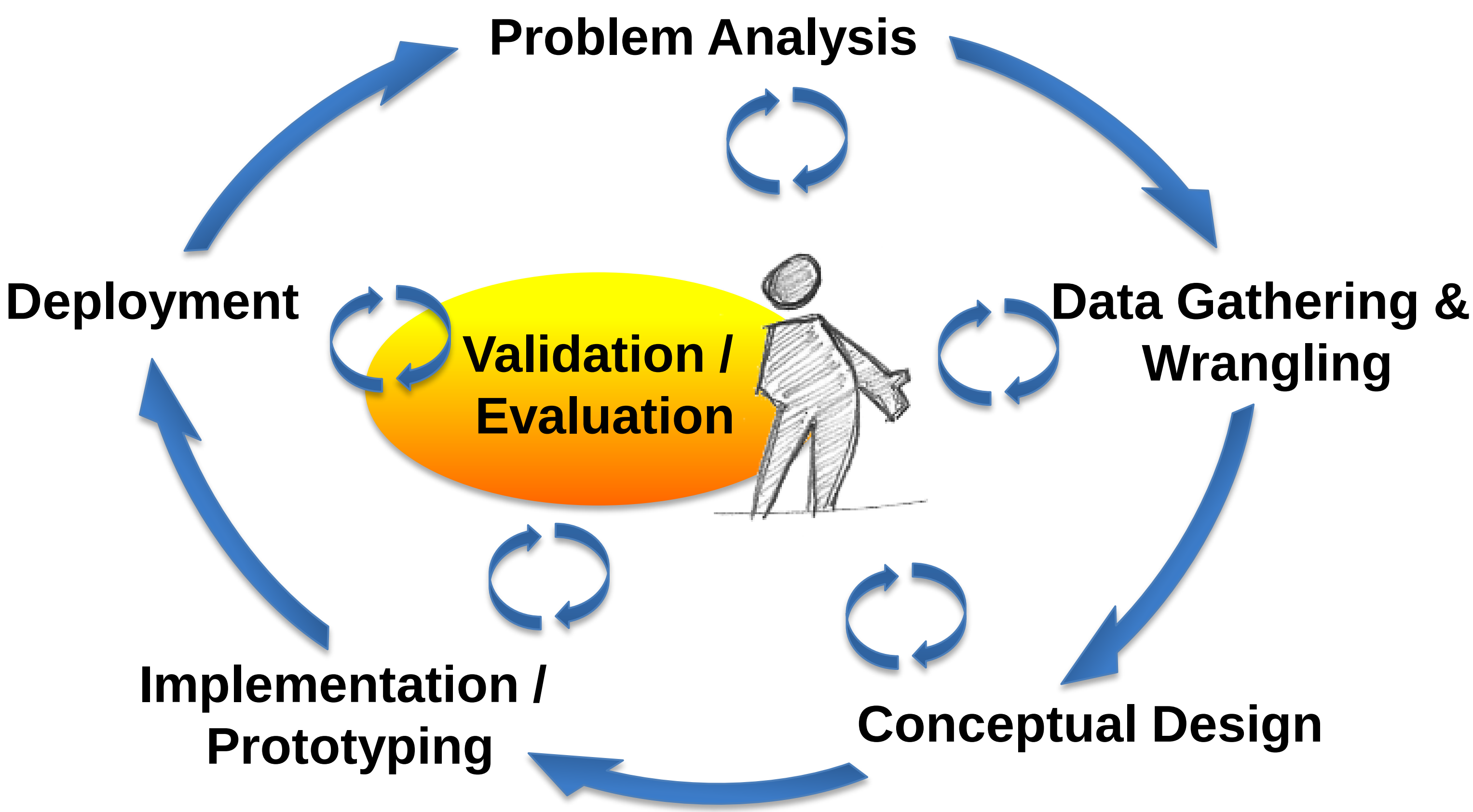
https://www.youtube.com/watch?v=IGQmdoK_ZfY

https://www.youtube.com/watch?v=hhXZng6o6Dk

Resource limitations – Human Limits Revisit

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

[Illusion by Burt Anderson used in Dan Simons TEDx talk]



Validation / Evaluation

Types of Evaluation

[Robert Stakes]

Formative evaluation

evaluation and development are done in parallel

(iterative development process)

feedback about usability and utility

results cause improvement of the tool

Summative evaluation

development of the tool is finished

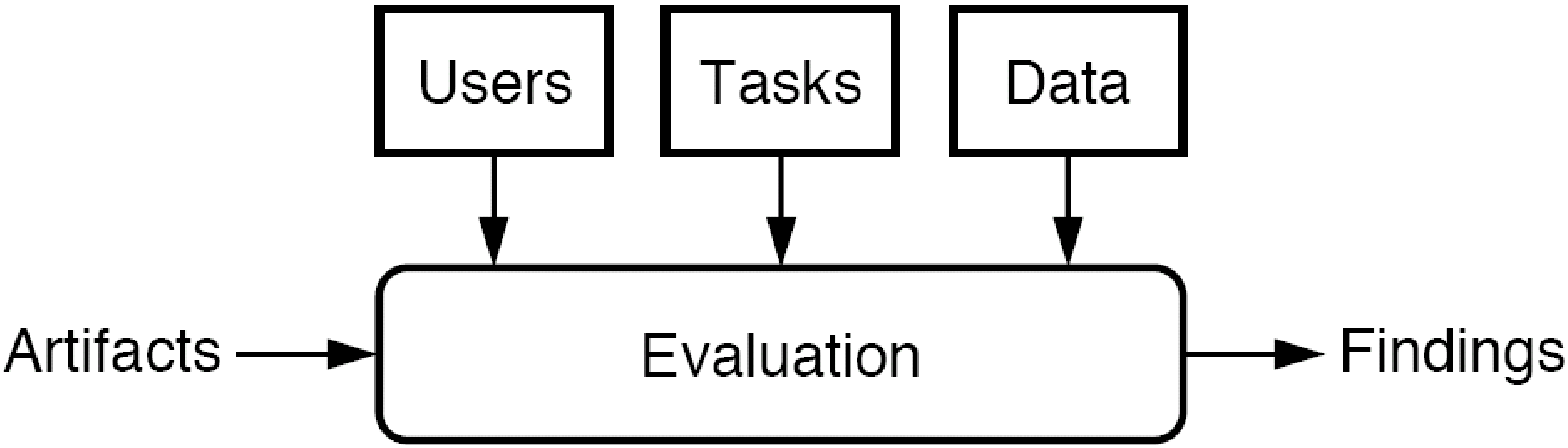
assessment of efficacy and features (e.g., comparative evaluation)

results may support buyers' decisions

'When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative.'

The Main Ingredients of Evaluation

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]



For Example,

Artifact :: scatterplots

Users :: training in the proper interpretation

Task :: helpful to find clusters

Data :: a limited number of real valued attributes

Evaluation

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

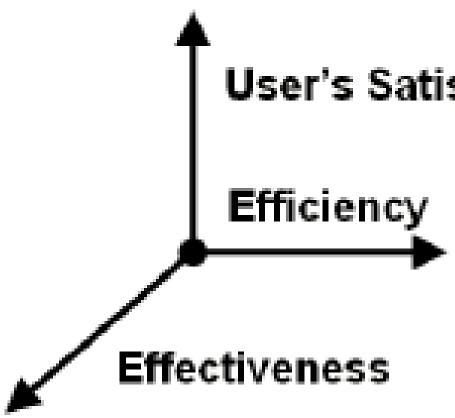
Quality of artifacts

Artifacts are not limited to software tools:

- Techniques, methods, models, theories, software tools

Quality

- Effectiveness
- Efficiency
- User's satisfaction



Evaluation Criteria

Functionality - to what extend the system provides the functionalities required by the users?

Effectiveness - do the visualization provide value? Do they provide new insight? How? Why?

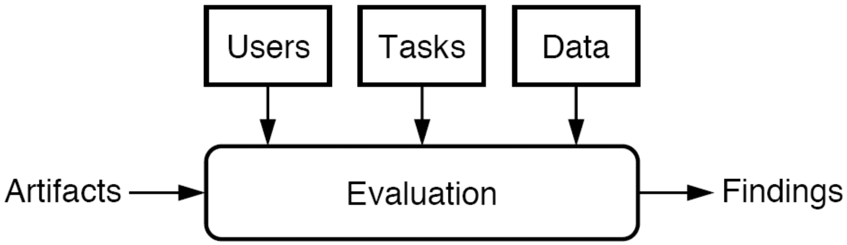
Efficiency - to what extend the visualization may help the users in achieve a better performance?

Usability - how easily the users interact with the system? Are the information provided in clear and understandable format?

Usefulness - are the visualization useful? How may benefit from it?

Artifacts

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]



Several Levels

- **Low-Level Encodings**
  - e.g., grey value vs. size
- **Component Level**
  - e.g., visualization/interaction technique
- **System Level**
  - e.g., system X vs. system Y
- **Environment Level**
  - e.g., integration of system X in environment Z

Users

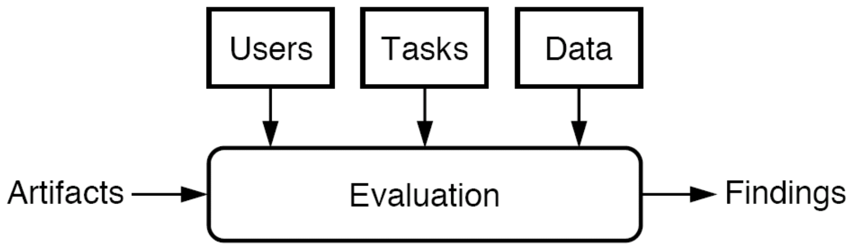
Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

Can be professional well trained or lay persons

Can be proficient with computers or not

Can be young or old

…



Difficult issues

Expert are well trained and know the tasks but their time is precious and they are scarce resources

Students as found in our labs will not exhibit the same kinds of performance as experts for real tasks

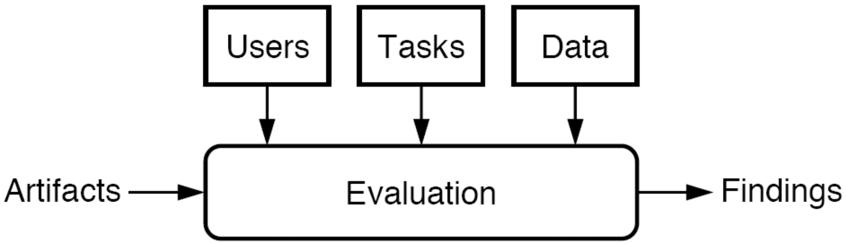
Tasks

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

Several Levels

Low level: important but not "ecologically valid" and not sufficient

Can be done in clean lab settings



Evaluation - Specification of Goals

What to investigate? What are the research questions?

How to investigate in order to get answers?

Domain knowledge helps to identify relevant research questions

Example: E-learning system

Question 1: Did the participants learn the content?
Method: Exam

Question 2: Did the participants like to use the system?
Method: Interviews

Question 3: Is the system easy to use?
Methods: Observation, Software logs

Evaluation - Implementation of a Study

Select and find participants for the study (subjects)

Laboratory setting

+ clear conditions allow for good identification of causality
– simulated and restricted setting could yield irrelevant statements

Field study

+ lifelike and informative
– identification of valid statements is difficult because of the complexity (high number of variables)

Types of Evaluation (2)

[Robert Stakes]

Quick-and-dirty

informal and non-systematic

small number (2 to 10) subjects use the product and

tell what they think about it

usually conducted during product development

low cost

Scientific evaluation

elaborated process

definition and validation of scientific hypotheses

minimum of 20 subjects for quantitative studies

standardized evaluation methods: quantitative or qualitative

conducted to investigate core questions of a product or research topic, e.g., command-line interaction versus direct manipulation of objects

Evaluation Methods

[Mazza 2009]

Analytic Methods

based on formal analysis models and conducted by experts

Heuristic/expert evaluation

Cognitive walkthroughs

Empirical Methods

realized through experiments with user test

Quantitative studies

Controlled experiments (also called experimental studies)

Software Logs

Qualitative studies

Observations

Thinking Aloud

Longitudinal Studies (MILCS)

Field Studies

Insight-based Method

LESSONS LEARNED

Lessons learned

Each Project Has Unique Requirements

A visualization should convey the unique properties of the data set it represents.

Know Your Audience

who is your audience? What are their goals when approaching a visualiza- tion? What do they stand to learn? Unless it's accessible to your audience, why are you doing it?

Be prepared to spend a lot of time data munging

famous 80/20 rule

Work with real data

getting good data is often difficult and annoying (e.g., legal issues, data scraping)

first data, then visualization design

Automate what you can

quick & dirty vs. reusability

Visualize early and often—but know when to say when

working iteratively is important

Avoid the All-You-Can-Eat Buffet

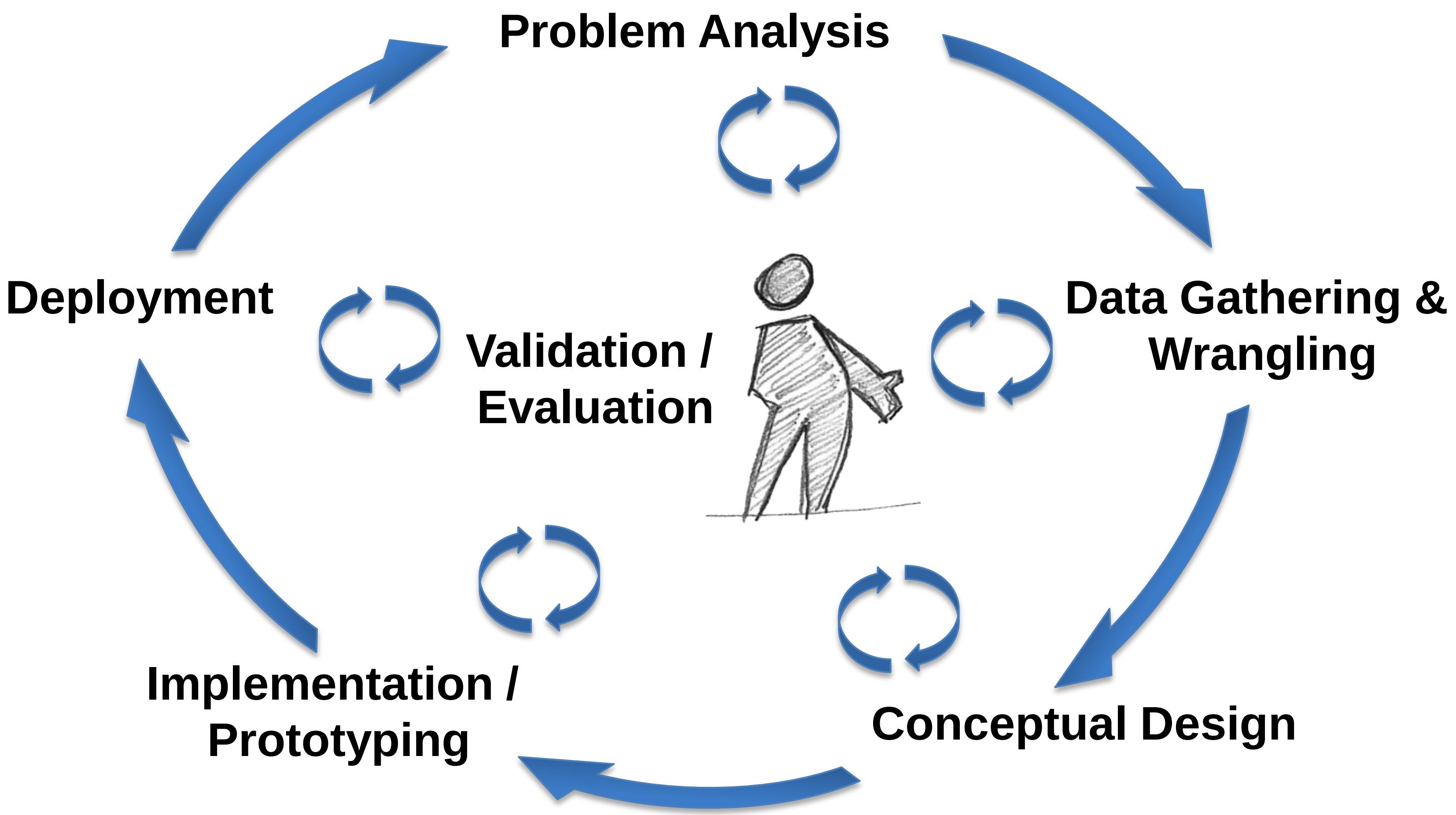
more data is not implicitly better, and often serves to confuse the situation.

Be aware of the larger process

visualization is just one step in a larger chain of analysis

[Odewahn, 2010; Wattenberg & Viegas, 2010; Fry, 2008]

Human-Centered Design Cycle



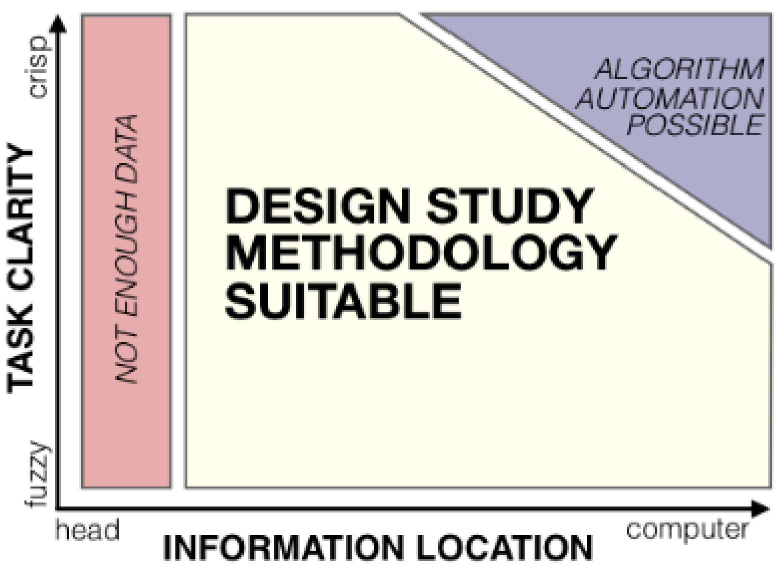
Design Study Methodology

[Sedlmair et al., 2012]

Is a research-oriented extension to the
Human-Centered Design Cycle discussed before.

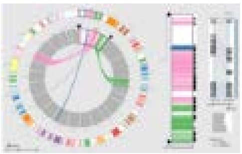
Definition of "design study":

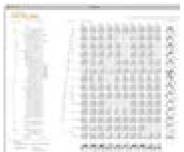
"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines."

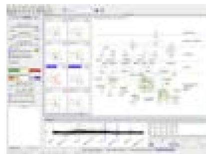


M. Sedlmair, M. Meyer, and T. Munzner, "Design Study Methodology: Reflections from the Trenches and the Stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

Lessons learned after 21 of them


MizBee
genomics


Pathline
genomics

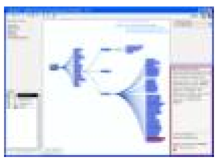

Cerebral
genomics


MulteeSum
genomics

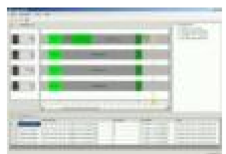

Vismon
fisheries management

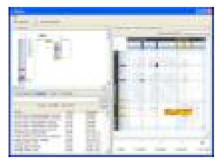

QuestVis
sustainability

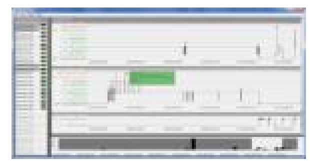

WiKeVis
in-car networks


MostVis
in-car networks

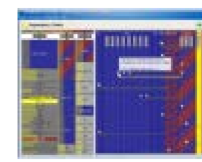

Car-X-Ray
in-car networks


ProgSpy2010
in-car networks

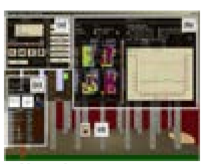

RelEx
in-car networks

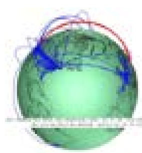

Cardiogram
in-car networks

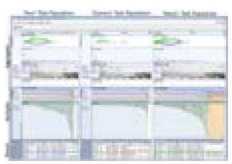

AutobahnVis
in-car networks

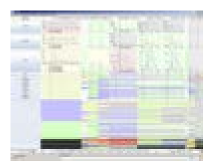

VisTra
in-car networks

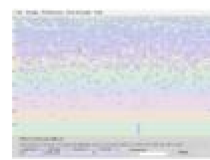

Constellation
linguistics


LibVis
cultural heritage


Caidants
multicast


SessionViewer
web log analysis


LiveRAC
server hosting


PowerSetViewer
data mining


LastHistory
music listening

commonality of representations cross-cuts domains!

Slides from Tamara Munzner's Mini Course
"Visualization Analysis & Design" 2014

Design studies: problem-driven vis research

a specific real-world problem

- real users and real data,
- collaboration is (often) fundamental

design a visualization system

- implications: requirements, multiple ideas

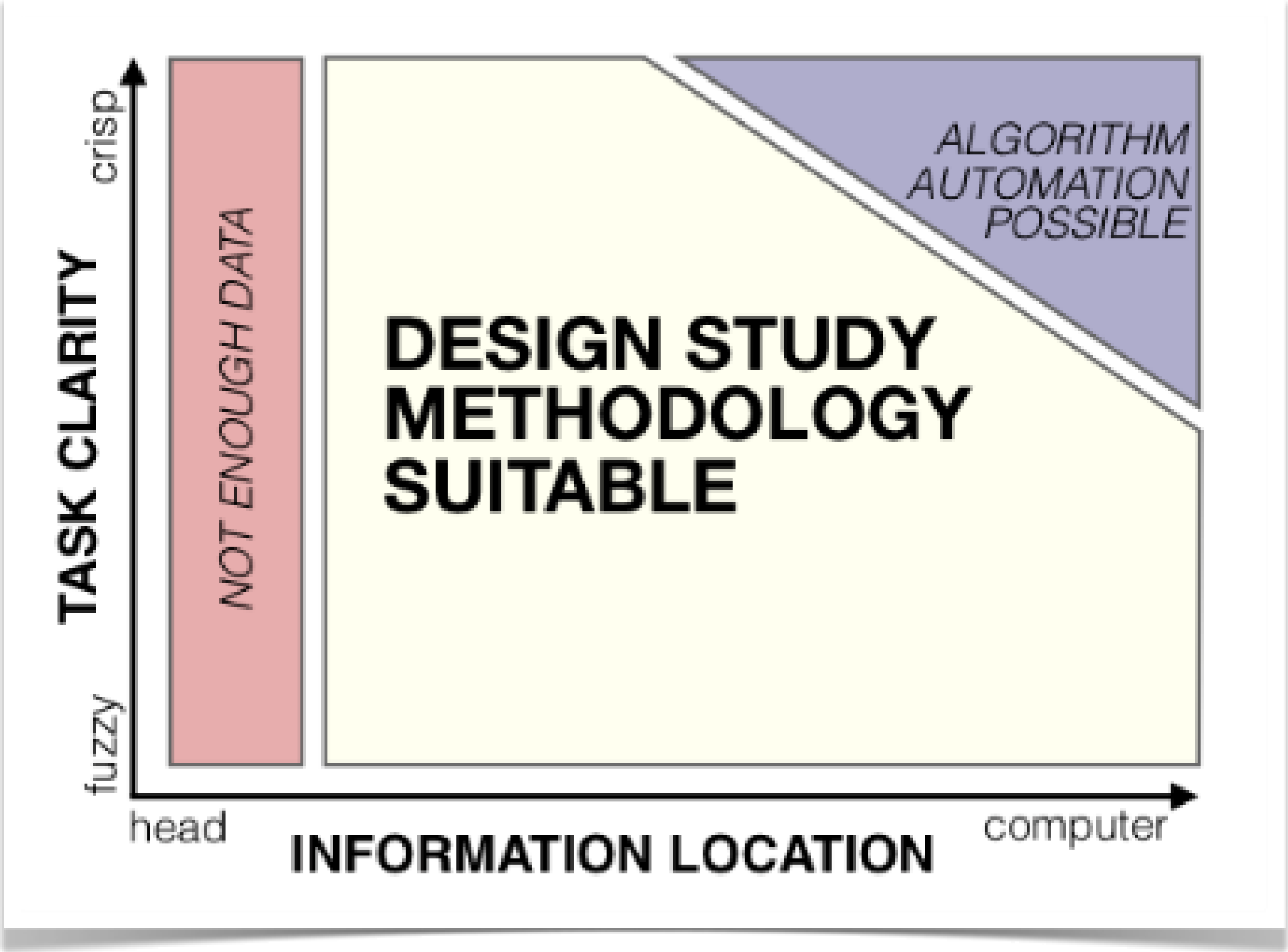
validate the design

- at appropriate levels

reflect about lessons learned

- transferable research: improve design guidelines for vis in general
  - confirm, refine, reject, propose

When To Do Design Studies

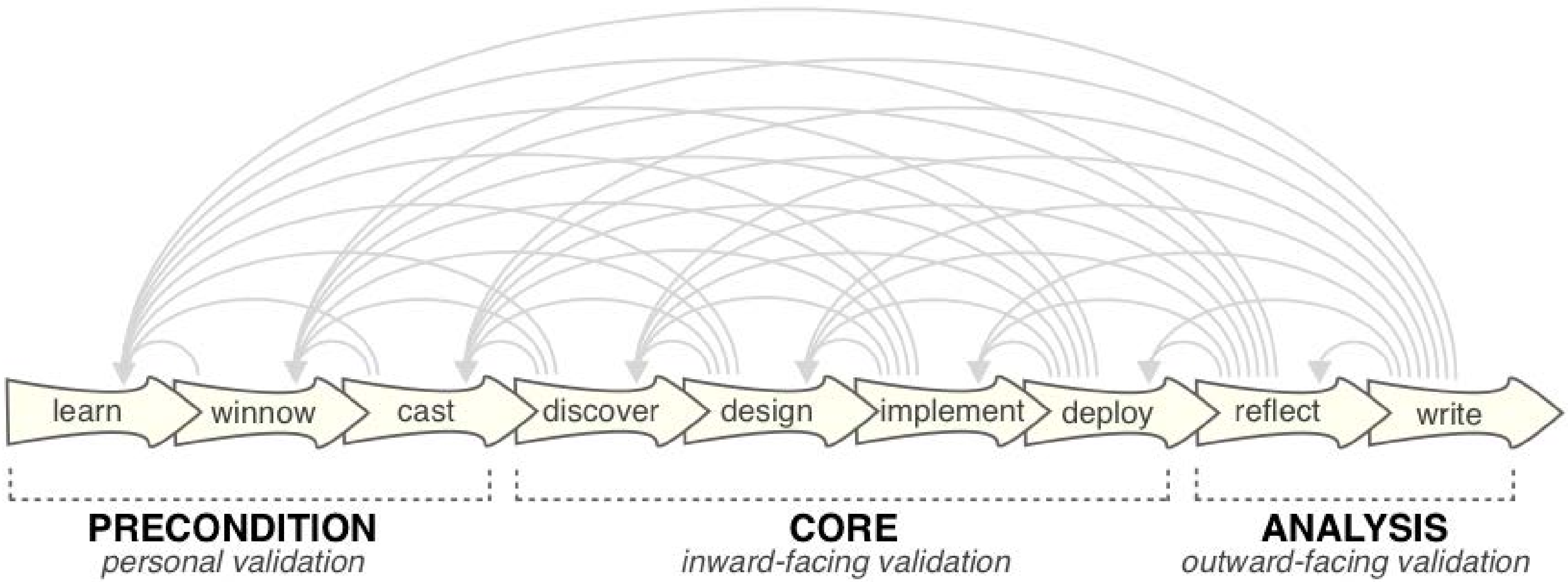


Slides from Tamara Munzner's Mini Course
"Visualization Analysis & Design" 2014

Design Study Methodology: 9-stage framework

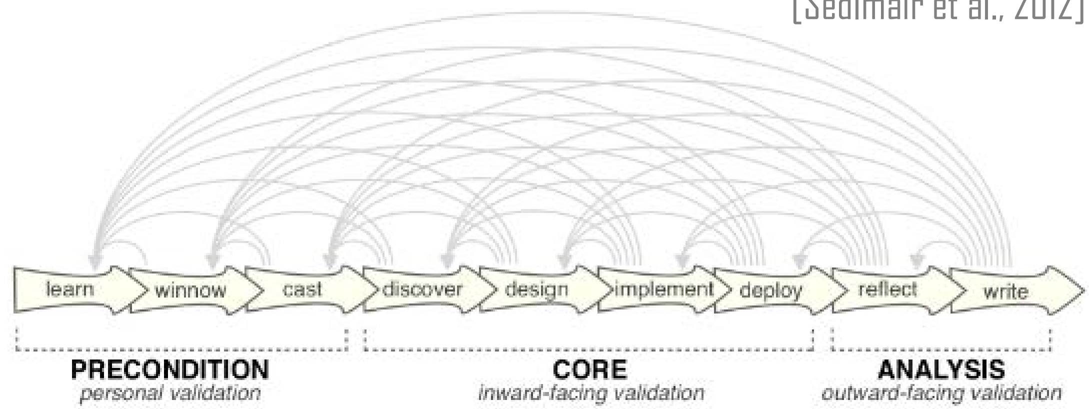
[Sedlmair et al., 2012]

Nine-stage framework



Design Study Methodology: 9-stage framework

[Sedlmair et al., 2012]



Precondition phase

1) Learn: Visualization Literature

2) Winnow: Select Promising Collaborations

3) Cast: Identify Collaborator Roles

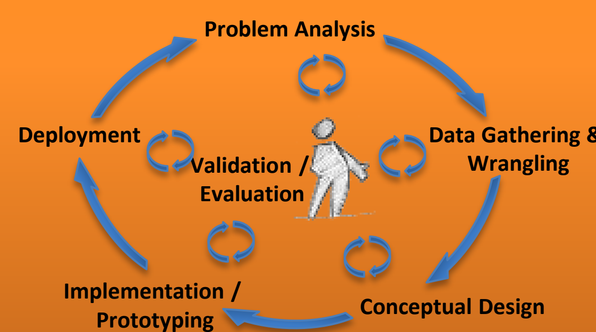
Core phase

4) Discover: Problem Characterization & Abstraction

5) Design: Data Abstraction, Visual Encoding & Interaction

6) Implement: Prototypes, Tool & Usability

7) Deploy: Release & Gather Feedback



Analysis phase

8) Reflect: Confirm, Refine, Reject, Propose Guidelines

9) Write: Design Study Paper

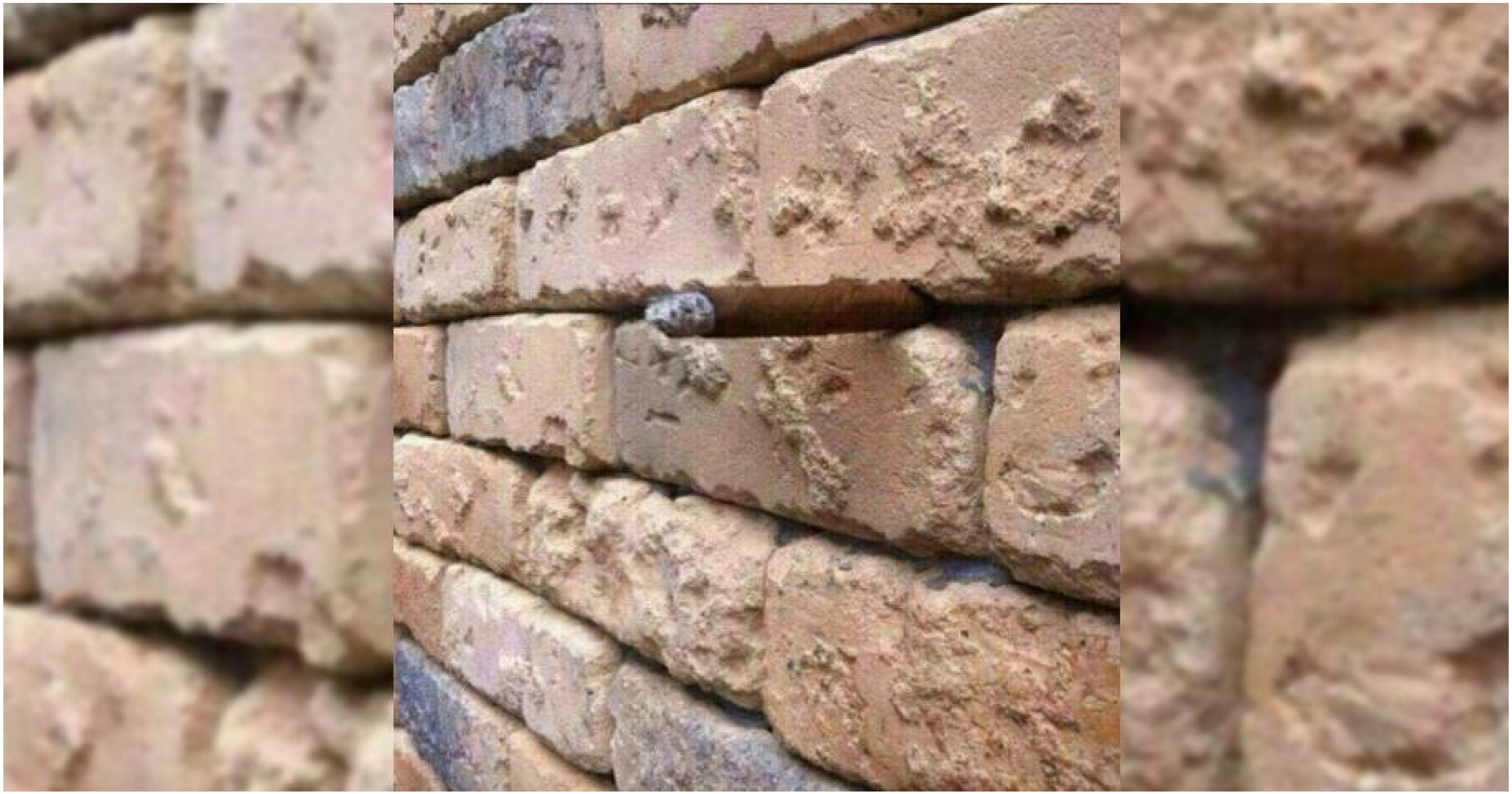
Resource limitations – Human Limits Revisit

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

[Illusion by Burt Anderson used in Dan Simons TEDx talk]

Resource limitations – Human Limits Revisit

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.



Literature

Kulyk et al., Human-Centered Aspects, in: Human-Centered Visualization Environments, Kerren et al. (eds), Springer, 2007.

Tamara Munzner. A Nested Model for Visualization Design and Validation. IEEE TVCG 15(6):921-928 (Proc. InfoVis 2009), 2009.

M. Sedlmair, M. Meyer, and T. Munzner, “Design Study Methodology: Reflections from the Trenches and the Stacks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

Glenn J. Myatt and Wayne P. Johnson. Designing Visual Interactions (Chapter 4), in: Making Sense of Data III: A Practical Guide to Designing Interactive Data Visualizations, John Wiley & Sons, 2011, p. 104-145.

A. Johannes Pretorius and Jarke J. van Wijk. What does the user want to see? What do the data want to be? Information Visualization 8(3):153–166, Palgrave Macmillan, 2009.

Ben Fry, Chapter 1: The Seven Stages of Visualizing Data, in: Visualizing Data, O'Reilly, 2008.

Benyon, D., Turner, P., Turner, S.: Designing Interactive Systems: People, Activities, Contexts, Technologies. Addison-Wesley, Reading, MA (2005)

Cooper, A. The Inmates Are Running The Asylum: Why High Tech Products Drive Us Crazy and How To Restore The Sanity. SAMS Publishing, 1999.

Shneiderman, B. 1996. The eyes have it: a task by data type taxonomy for information visualizations. Proceedings of IEEE Symposium on Visual Languages, Boulder, CO, September 3-6, 336-343

Wehrend, S. and C. Lewis. 1990. A problem-oriented classification of visualization techniques Proceedings IEEE Visualization '90, October, pp.139 - 143, IEEE Computer Society Press

Yi, J. S., ah Kang, Y., Stasko, J., and Jacko, J. (2007). Toward a Deeper Understanding of the Role of Interaction in Information Visualization. IEEE Transactions on Visualization and Computer Graphics, 13(6):1224–1231.

Andrew Odewahn, Visualizing the U.S. Senate Social Graph (1991–2009), in: Steele, J. and Iliinsky, N. (Eds.): Beautiful Visualization, O'Reilly, Chapter 8, 2010.

Martin Wattenberg and Fernanda Viégas, Beautiful History: Visualizing Wikipedia, in: Steele, J. and Iliinsky, N. (Eds.): Beautiful Visualization, O'Reilly, Chapter 11, 2010.

Literature

Data Wrangling

Kandel S, Paepcke A, Hellerstein J and Heer J. Wrangler: Interactive visual specification of data transformation scripts. ACM Human Factors in Computing Systems (CHI) 2011.

Sean Kandel, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck and Paolo Buono: Research directions in data wrangling: Visualizations and transformations for usable and credible data. Information Visualization, 10(4) 271–288, 2011.

Huynh, D. and Mazzocchi, S.: Freebase GridWorks. http://code.google.com/p/google-refine/

Implementation

Jeffrey Heer, Stuart K. Card, James Landay. Prefuse: A Toolkit for Interactive Information Visualization. Proc ACM CHI, 421-430, 2005.

Michael Bostock and Jeffrey Heer. Protovis: A Graphical Toolkit for Visualization. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2009.

Michael Bostock, Vadim Ogievetsky, Jeffrey Heer. D3: Data-Driven Documents. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2011.

Literature

Evaluation

T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller, “A Systematic Review on the Practice of Evaluating Visualization,” IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2818–2827, Dec. 2013.

Sheelagh Carpendale. Evaluating Information Visualizations. Chapter in "Information Visualization: Human-Centered Issues and Perspectives", Springer LNCS 4950, 2008, p 19-45.

Catherine Plaisant. The challenge of information visualization evaluation. Proc. Advanced Visual Interfaces (AVI) 2004

Ben Shneiderman and Catherine Plaisant. Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. Proc. AVI Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV), 2006, p 38--43.

Michael Sedlmair, Petra Isenberg, Dominikus Baur, Andreas Butz. Information Visualization Evaluation in Large Companies: Challenges, Experiences and Recommendations. Journal of Information Visualization, Special Issue on Evaluation (BELIV 10), Volume 10, Number 3, July 2011.

Purvi Saraiya, Chris North, Karen Duca. An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. IEEE Trans. Vis. Comput. Graph. 11(4):443-456 (2005)

C. North. Toward measuring visualization insight. IEEE Computer Graphics and Applications, 26(3):6–9, 2006.

Acknowledgements

Thanks to

Wolfgang Aigner, Silvia Miksch, Jeff Heer, Katy Börner, Tamara Munzner

… for making nice slides of previous classes available

MORE DETAILS ABOUT EVALUATION (OPTIONAL FURTHER INFORMATION)

Heuristic Evaluation (1)

[Nielsen 1994]

A small number of trained evaluators (typically 3 to 5) separately inspect a user interface by applying a set of **'heuristics'**, **broad guidelines** that are generally relevant

Use more evaluators if usability is critical or evaluators aren't domain experts

Go through interface **at least twice**:

1. Get a feeling for the flow of the interaction
2. Focus on specific interface elements

Write reports

Reference rules, describe problem, one report for each problem.

Don't communicate before all evaluations are completed!

Observer assists evaluators

Use additional usability principles

Provide **typical usage scenario** for domain-dependent systems

Conduct a debriefing session (provides design advice)

Phases:
pre-evaluation training / evaluation / debriefing / severity rating

Heuristic Evaluation (2)

[Nielsen 1994]

Visibility of system status

The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

Match between system and the real world

The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

User control and freedom

Users often chose system functions by mistake and will need a clearly marked „emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

Error prevention

Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

Heuristic Evaluation (3)

[Nielsen 1994]

Recognition rather than recall

Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Flexibility and efficiency of use

Accelerators — unseen by the novice user — may often speed up the interaction for the expert user to such an extent that the system can carter to both inexperienced and experienced users. Allow users to tailor frequent actions.

Aesthetic and minimalist design

Dialogues should not contain information which is irrelevant od rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

Help users recognize, diagnose, and recover from errors

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Help and documentation

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Heuristic Usability Evaluation (1)

[Forsell & Johansson, 2010]

A new set of 10 heuristics out of 63 heuristics (from 6 earlier published heuristic sets)

Especially tailored to the evaluation of common and important usability problems in *Information Visualization techniques*

Heuristic Usability Evaluation (2)

[Forsell & Johansson, 2010]

1. **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.

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

3. **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.

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

5. **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.

Heuristic Usability Evaluation (3)

[Forsell & Johansson, 2010]

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

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

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

9. **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.

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

Controlled Experiment (Experimental Study)

… a methodical procedure carried out with the goal of verifying, falsifying, or establishing the validity of a hypothesis.

“controlled” environment

- **Independent** variables
- **Dependent** variables
- Representative sample of users (test users/subjects)
- Tasks
- Measurements/Metrics: e.g., completion time, correctness)

Qualit. & Quant. Evaluation Methods

Interviews / focus groups

Questionnaire

Observation

Software logs

Thinking Aloud

Interviews / Focus Groups

Interviews

can give a differentiated idea of the usability and efficacy of a tool

subjects cannot always report their behavior,
since some cognitive processes are automatic and unconscious

subjects' intentions can provide reasons
for measurements and objective data

allows for in-depth analysis

based on guidelines

Focus groups

discussions with groups

sometimes a problem to ensure equal participation

group situation could influence topics

based on guidelines for discussion and moderation

Questionnaire

In contrast to interviews questionnaires allow for studying large groups of people (quantitative evaluation)

Can yield representative data

Should avoid bias

Difficult to prevent misunderstandings because of different interpretations

Simple questions

Closed questions: given answer categories

Open questions: free answers, etc.

Observation

Collection of information does not depend on subjects' reports (sometimes subjects can give no information about their activities)

Subjective falsifications are impossible

Problem to understand why persons set certain actions.

No guarantee that the observed person behaves naturally (Hawthorne effect)

Observations can take place in laboratories or in real-world situations

Yields an abundance of data

Difficult to select relevant data

Based on guidelines (what to observe)

Software logs

Monitoring tool collects data about computer and user activities, e.g., about number and location of clicks or type of keyboard input

Observes only a limited number of activities

Delivers high amount of data

Procedure is not visible for user

Does not intervene user's activities

Activity sequences yield more information than single step

Analysis of activity sequences is difficult

Software logs do not register the intentions or goals of the users

Thinking Aloud

Mixes observation and questioning

Subjects are asked to describe their thoughts while using the product

Gives more details than interviews, because information filtering is reduced

Thinking aloud could impede the interaction processes

It is difficult to express the thoughts if interaction with the tool requires attention

Sometimes crucial situations are not reported

Provides with highly relevant and interesting data

Empirical Evaluation Methodologies

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

- User-Centered Design Methods
- Usability studies
- Quantitative methods
- Qualitative methods
- Mixed methods
- Informal evaluations (to inform designers or reviewers)
- Longitudinal studies (MILCS)
- Insight-based methods
- Contests & Repositories
  - Graph Drawing Contest
  - InfoVis Contest
  - SoftVisContest
  - VAST Contest
    - Generation of plausible scenario with ground truth
  - KDD Cup
  - Netflix contest

Empirical Evaluation Methodologies

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

User-Centered Design Methods
Usability studies
Quantitative methods
Qualitative methods
Mixed methods
Informal evaluations (to inform designers or reviewers)
Longitudinal studies (MILCS)
Insight-based methods
Contests & Repositories

- Graph Drawing Contest
- InfoVis Contest
- SoftVisContest
- VAST Contest
    - Generation of plausible scenario with ground truth
- KDD Cup
- Netflix contest

Longitudinal Studies (MILCS)

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

B. Shneidermanand C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In Proc. AVI BELIV workshop, pages 1–7, ACM, 2006.

- Multi-dimensional in-depth case studies
- Select motivated experts (1 or 2)
- Present them the tool with their data
- Organize weekly sessions (2h or more) to work on their problem
- Continue for months
- Record their findings and issues

Insight-Based Method

Jean-Daniel Fekete → [Keim, et al. 2010 - RoadMap]

C. North. Toward measuring visualization insight. IEEE Computer Graphics and Applications, 26(3):6–9, 2006.

Work with experts

Give them the tools

Ask them to write down each time they find an “insight”

Count and classify the insights

[Kerren, et al. 2007]

Method	Design Stage	Advantages	Disadvantages
Heuristic Evaluation	Concept design	Investigates individual usability problems; can address expert user issues; fast	Does not involve real users; can not reveal "surprises" relating to their needs
Cognitive Walkthrough	Concept design	Puts the focus on the user; recognition of user goals	Bias because of task selection; tries to make the designer the user
Focus Groups	Analysis, Concept design	Takes a short period of time; produces variety of design choices and new features; spontaneous reaction and group dynamics	Records what users think they want; doesn't test actual interaction
Interviews	Analysis, Concept design	Flexible; in-depth attitude and experience probing	Time consuming; hard to analyze and compare
Observations	Analysis, Detailed design, Implementation	Made in real-use environment	Very costly; difficult to analyze and reason the behavior
Empirical Experiments	Detailed design, Implementation	Allows to design hypotheses or alternatives; close approximation to actual usage	Experts knowledge required; limited generalization of results due to controlled settings
Field Studies	Implementation	Made in real use environment; reveals real-life problems	Costs a lot of time; impossible to control the settings