How to Evaluate (your) Visualizations PA214 — Visualization II



Talk Outline

- Methodologies for visualization design
- Evaluation Categories
 - Understanding the tool vs. understanding the processes
 - Evaluation without users vs. with users
- Some tips and tricks for doing the evaluation

Why do we evaluate the visualizations?



Five Design Sheets

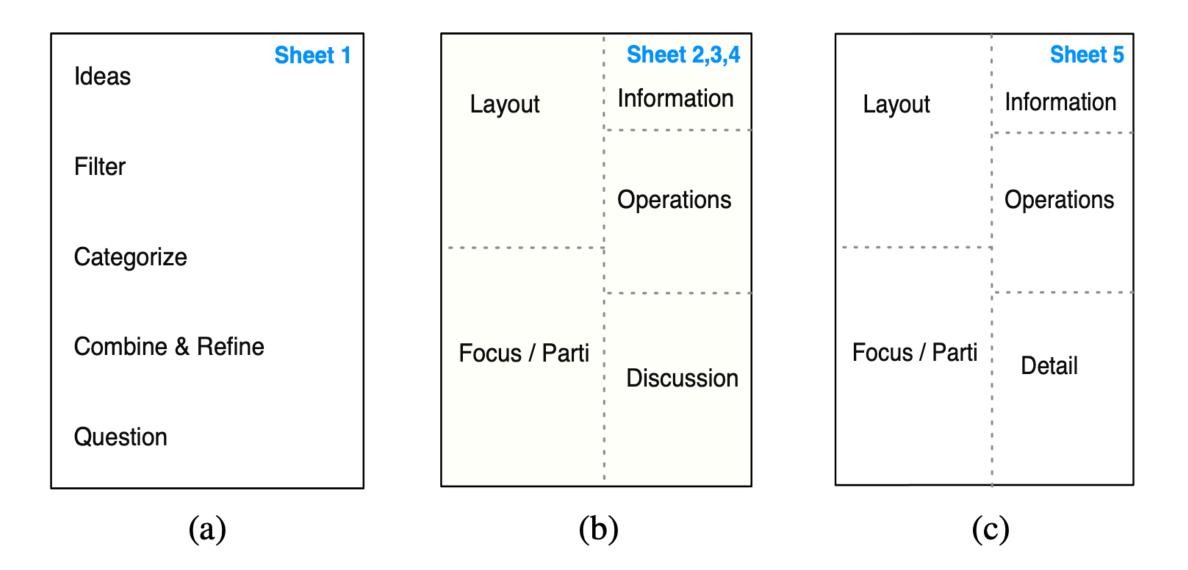
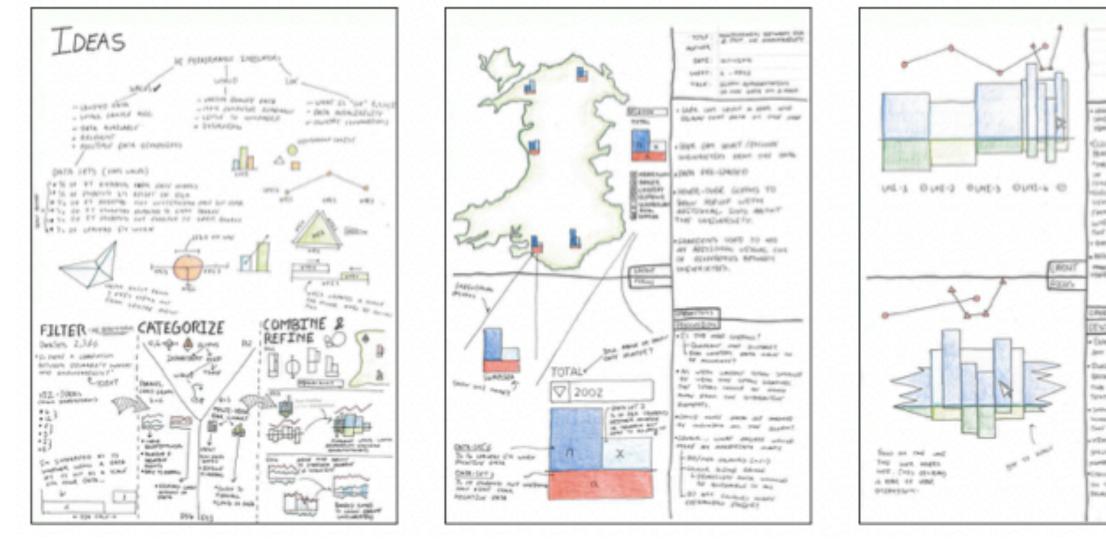
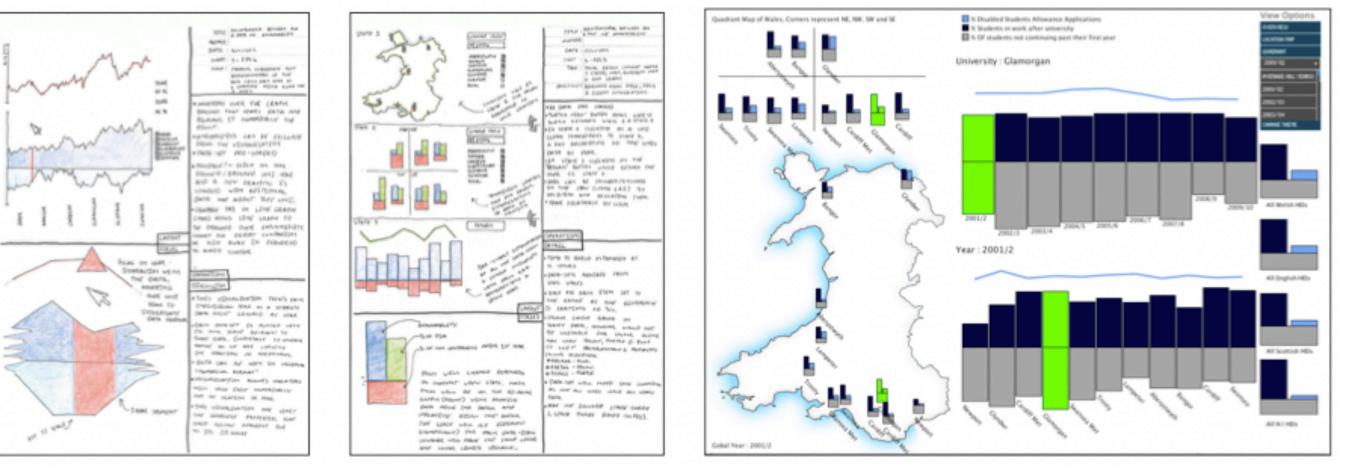


Fig. 3: The FdS sheets. (a) Sheet 1: Generate Ideas, filter, categorize, combine & refine then question. (b) Sheets 2,3,4 with the five sections in the 2-row 3-row format; (c) Sheet 5, the realization sheet where *Detail* is included instead of Discussion.

https://pdritsos.com/projects/FDS/





Пто различи и на протоколого амент ред наличи ред наличи у на протоколого у на протоколого у на протоколого у на протоколого и на протоколого на про	
beet per servers, serve & 4200 Star an come componential person an impresentation person an impresentation person and impresentation person and impresentation person and impresentation person and an implementation of any year despectives, attent of an implementation of any server of an implementation and an of any server of an implementation and an of any server of an implementation and and of any server of any server	stress managements makeling that
per univers per univers per 1 - 2011 the per of the personal per constant personal per constant personal per constant per consta	
Matt 1 4481 Matt 1 4481 Matt 1 448 1 2007 Anterester g control control and the g control control and the p control control and the p control control and the state of your control and the control and and the control and the control and and the control and	
All and and a second	
y con the procession of the second series of the series of	
guarde and provide part, part and provide part, part and part provide part, part and part of the york descending of the york descending of the york descending of the provide state of the part of the state of the part of the part of the part of the state of the part of the part of the state of the part of	THE LOCAL COMPANY OF THE PARTY OF
ALTER AND DECEMPTOR AND	A COMPANY OF THE OWNER
Destrictly sum organization of by your organization attention an analytic description of by your any production of by your any production of products of any sum and products and and any products of any sum any of a sum of any sum any sum of a sum any sum of a sum any sum of any sum any sum any sum of any sum any sum	
<pre>d br your expendition, appe an an registration, br back part and temperature br back part and temperature and your and and and antalas, appearing according antalas, appearing according and according according to an according according to an according according to an according according to an according according and according acco</pre>	state part part parter?
HING OF AN THEIR CARAGE E which set their seadown (b) (AD) 10 that siles seadown (b) (AD) 10 that siles (ATTON AT THE ATTON (ATTON AT THE ATTON (ATTON AT THE ATTON (ATTON AT THE ATTON (ATTON (ATTON) (
K. Visuld Jaff Mill Statution By (440, 52) their Allianti must presid asseed and metabolic approximation and including metabolic and units of all asseements to and units must discussion to protection (2000) and and asseements in the classifier and and and asseements and and and asseements and president asseed and and an and asseed to an automatic theory and an an automatic theory and an and and an analy and an and and an analy and an an automatic theory and an an automatic theory and a state of the and and a see an automatic theory of an automatic theory and a see an automatic theory of a sector of the automatic an automatic theory and an automatic of a sector of a sector of a sector of a sector of the automatic an automatic and automatic theory automatic an automatic theory automatic an automatic theory automatic automatic and automatic theory automatic and automatic theory automatic an automatic theory automatic automatic and automatic theory automatic and automatic theory automatic and automatic automatic and automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automat	4 St. Apr. Seharburg.
K. Visuld Jaff Mill Statution By (440, 52) their Allianti must presid asseed and metabolic approximation and including metabolic and units of all asseements to and units must discussion to protection (2000) and and asseements in the classifier and and and asseements and and and asseements and president asseed and and an and asseed to an automatic theory and an an automatic theory and an and and an analy and an and and an analy and an an automatic theory and an an automatic theory and a state of the and and a see an automatic theory of an automatic theory and a see an automatic theory of a sector of the automatic an automatic theory and an automatic of a sector of a sector of a sector of a sector of the automatic an automatic and automatic theory automatic an automatic theory automatic an automatic theory automatic automatic and automatic theory automatic and automatic theory automatic an automatic theory automatic automatic and automatic theory automatic and automatic theory automatic and automatic automatic and automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automatic automat	wine on an reflected.
nucleon of peel and and and matchin afree on traj million with unsup universally a and unive out annually a set balance of annually a set balance (1990) (19	a would let all considers
nucleon of peel and and and matchin afree on traj million with unsup universally a and unive out annually a set balance of annually a set balance (1990) (19	or pair to the stands
MELIANS INFORM ANT THE PROPAGE INFORMATION I	must years, sorred shuff
 Millionski unter mel szimelege a ank uster mel szimelege b polenistru ja sila. Million d szimelege inter, And Stanstern inter, And Stanstern inter, And Stanstern inter, And Stanstern inter, And Stanstern b descenter, 1993 Millionski (riti) app stanster, And Stanstern site descenter, 1993 Millionski (riti) app stanster, 1994 Millionski (riti) app stanster, 1994 Millionski (riti) stanster, 1994 Millionski (riti) stanste	PR04.04011
 Millionski unter mel szimelege a ank uster mel szimelege b polenistru ja sila. Million d szimelege inter, And Stanstern inter, And Stanstern inter, And Stanstern inter, And Stanstern inter, And Stanstern b descenter, 1993 Millionski (riti) app stanster, And Stanstern site descenter, 1993 Millionski (riti) app stanster, 1994 Millionski (riti) app stanster, 1994 Millionski (riti) stanster, 1994 Millionski (riti) stanste	APR ALTS ATM
 and, active and spreaders probabilities (a status, and an establishing) probabilities (a status, and an establishing) probabilities (and status, and an establishing) probabilities (an establishing) probabilities (an establishing) 	whether the state
D primiting in miles Million in consideration of the considerati	a may my a set the standard
a particulari control and distant control and distant control and distant control and distant and distant and distant and distant control provided and and distant and	to guilendifus in state.
Alter to read dialer reads An Electron b descentions Styles	5.585.J
A MANDARIAN METORY M	in par-cash(1)
A MANDARIAN METORY M	ON to these distant
A MANDARIAN METORY M	reinholds and themptoper-
ALEDNA SKULDA	6 inverservors?
SELECT 6 TRODUCTION UND ADD TROUGHT (PTERAIPSINT TROUGHT (PTERAIPSINT TROUGHT OF DOLLARS) T TROUGHT OF DOLLARS) T TROUGHT OF DOLLARS TROUGHT (PTERAIPSINT TROUGHT OF ADDRESS TROUGHT OF TROUGHT OF TROUGHT OF TROUGHT OF T TROUGHT T TROUGHT OF T T TROUGHT OF T	
SELECT 6 TRODUCTION UND ADD TROUGHT (PTERAIPSINT TROUGHT (PTERAIPSINT TROUGHT OF DOLLARS) T TROUGHT OF DOLLARS) T TROUGHT OF DOLLARS TROUGHT (PTERAIPSINT TROUGHT OF ADDRESS TROUGHT OF TROUGHT OF TROUGHT OF TROUGHT OF T TROUGHT T TROUGHT OF T T TROUGHT OF T	
DELADON G TELEFOLDIN' LINES AND TELEFOLT SATERATION TO ANY PARAMETER TO ANY PARA	
SELECT 6 TRODUCTION UND ADD TROUGHT (PTERAIPSINT TROUGHT (PTERAIPSINT TROUGHT OF DOLLARS) T TROUGHT OF DOLLARS) T TROUGHT OF DOLLARS TROUGHT (PTERAIPSINT TROUGHT OF ADDRESS TROUGHT OF TROUGHT OF TROUGHT OF TROUGHT OF T TROUGHT T TROUGHT OF T T TROUGHT OF T	00.2042
6 Productions with app received (protecting) Protecting (protecting)	Profilered.
auchei Grittlandennen Holen an ceite somet Holen an ceite somet Holen an ceite somet Holen an ceite somet Holen and the ceite Holen Hol	a land strategy and a set
Des ber solen-suge Hilforn af deta sonnalt Vite de ditterang py r and der ditterang py r and der in berten deut af der in gesten deut af gern, mighetite deute far ber r andertite deute far der i angest den ar dittigen ar gelit februghen an	0 herefore and who
Helphon and their source, p helphon and strend and phy r and physic and physical and or an authors should dar for a source of the source of the genes, majorettion descends that is present dates are strettioned are physical belowing and are physical belowing and are physical belowing the source of the source of the source of the source of the are physical belowing the source of the source of the are physical belowing the source of the are physical belowing the source of the	
Helphon and their source, p helphon and strend and phy r and physic and physical and or an authors should dar for a source of the source of the genes, majorettion descends that is present dates are strettioned are physical belowing and are physical belowing and are physical belowing the source of the source of the source of the source of the are physical belowing the source of the source of the are physical belowing the source of the are physical belowing the source of the	State (1997) and and and
Vite of Eliteracy py r and their on Despective in an porter churt by decs. mplactive deserves for r and/attive deserves for r and/attive s proof dese or Elititate or pair deserves an pairs	A DATE OF CASE AND A
red Dear ne Despectation in als parties that by loss dya't (2) we'r awroll dear, regberttin Anneth dya' r chillettin Anneth dya' r chillettin an ar fibiliaen ar geltt Belanghain cu	views for structures in
er als perfet shuar by bes sport processing glass, mapleton desembly buy r calletonic when by collection, a prest beau or thebaux or gaint beaughtuin co	rat they an Desperate
ne get processor gets, nepation annessor r calabilities with the effort, a creek ben or thebase or get theorytheir co	
decs. mplatitive decembs that i constations while der eDas, a creat dens ar stationen ar gebit februghean co	
ngbellin konto bar 7 oktorioni -da, Gr elin, 8 met kes ar Ehlinan ar gelit februhan os	deca.
r callettors with the eDias, a ment sea or ethilows or patch theory that or	and streets and
elling, a proved dans air ethillings air getst terangilingen og	
a preef data or ED-Doot.	constrained with the
to create pair in the create and the creater pair in the creater pair of the creater p	
n createring and station	a prese new or strikings
to created some to place	real theory and the property of the
	or completeness at Starly

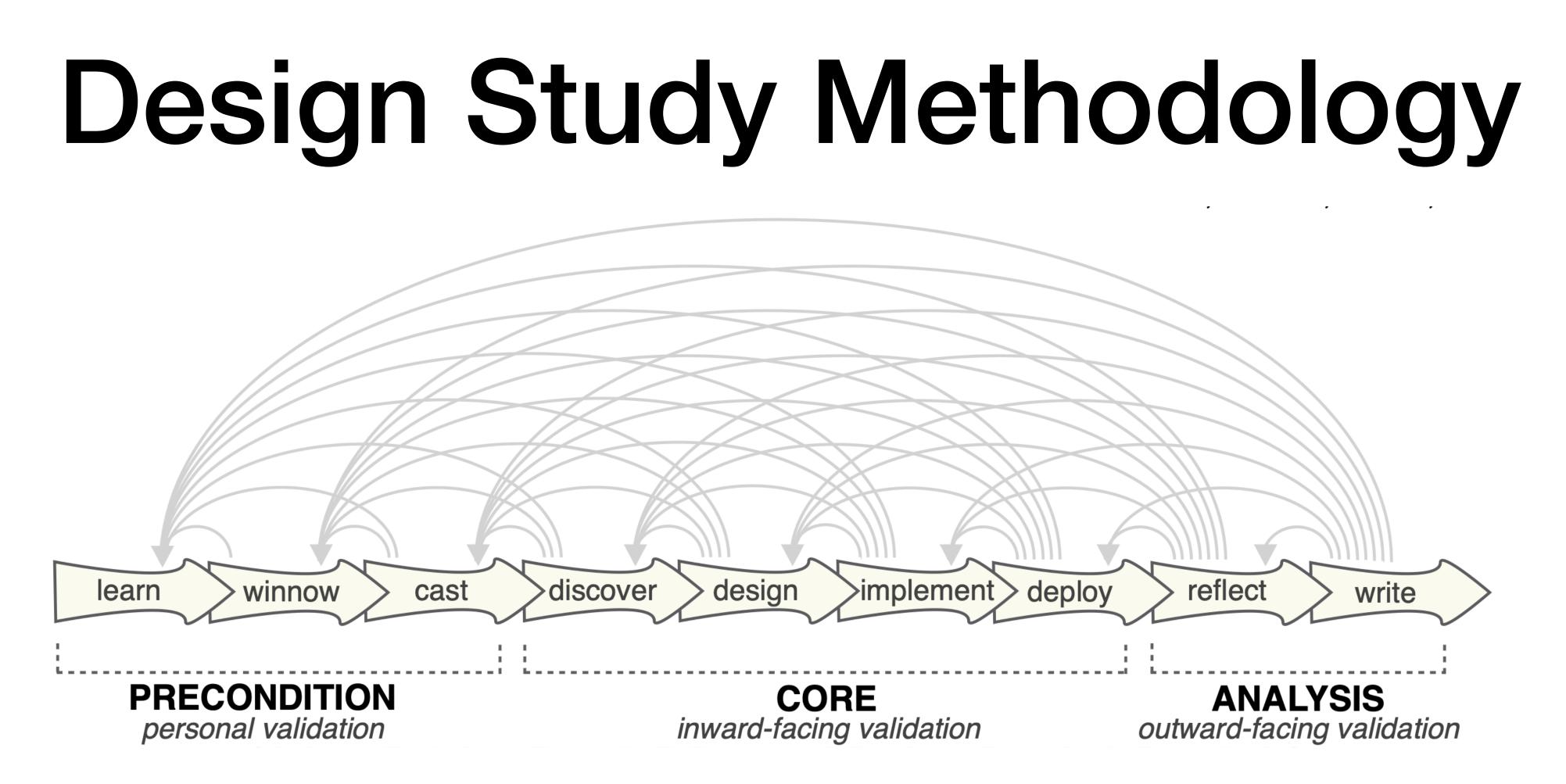
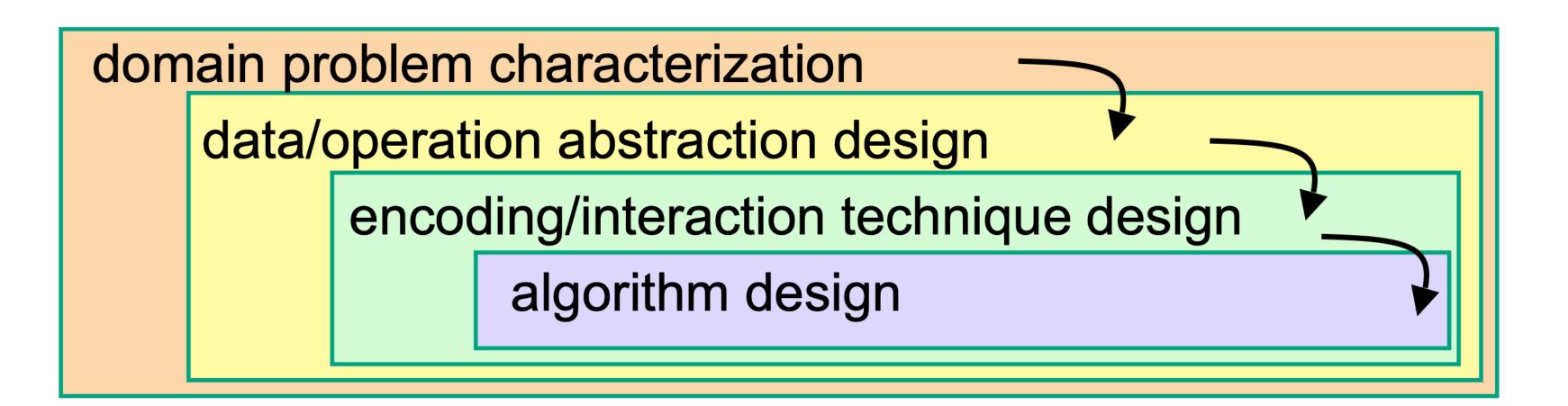


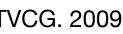
Fig. 2. Nine-stage design study methodology framework classified into three top-level categories. While outlined as a linear process, the overlapping stages and gray arrows imply the iterative dynamics of this process.

Source: M. SedImair, M. Meyer and T. Munzner, Design Study Methodology: Reflections from the Trenches and the Stacks. IEEE TVCG. 2012.

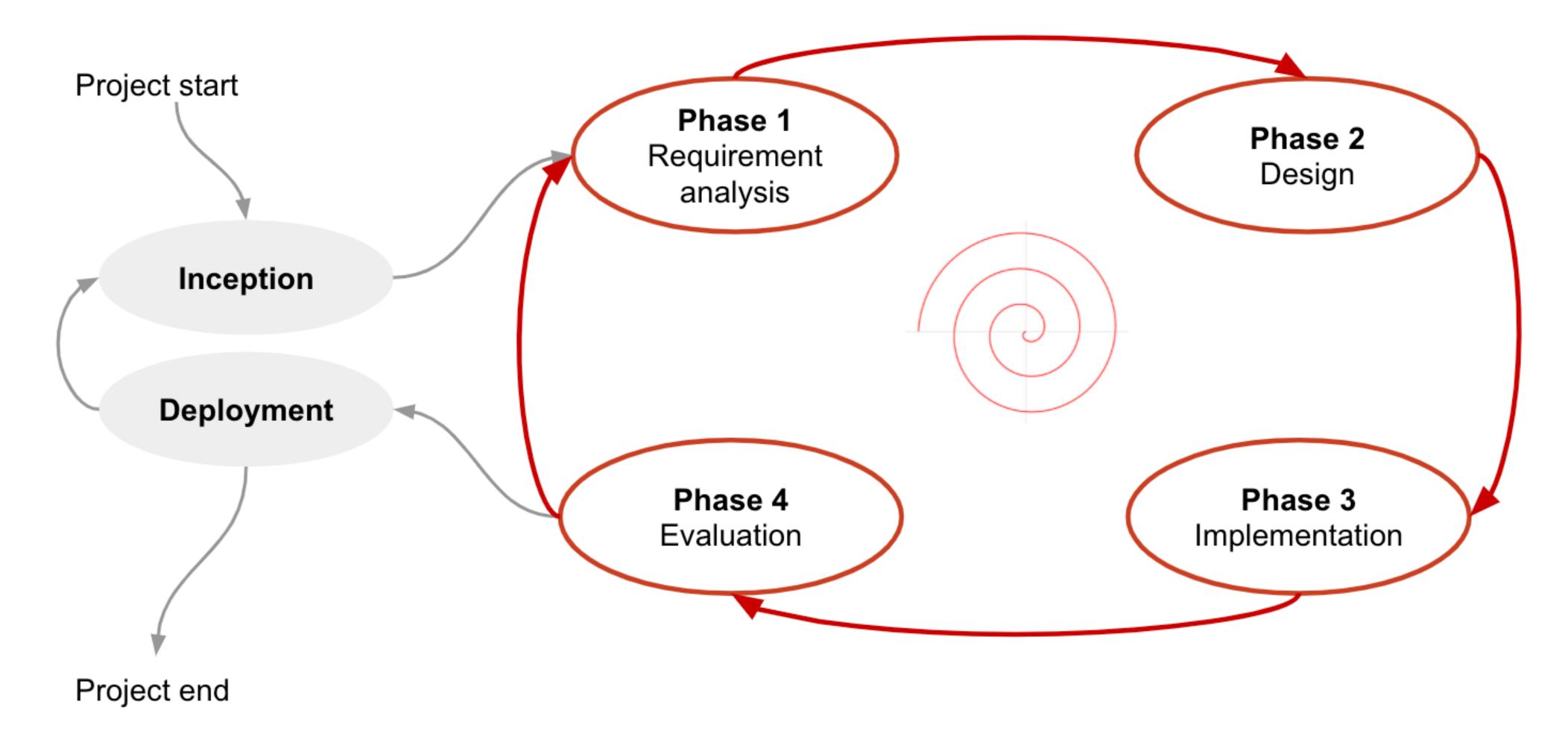


Nested Model Methodology





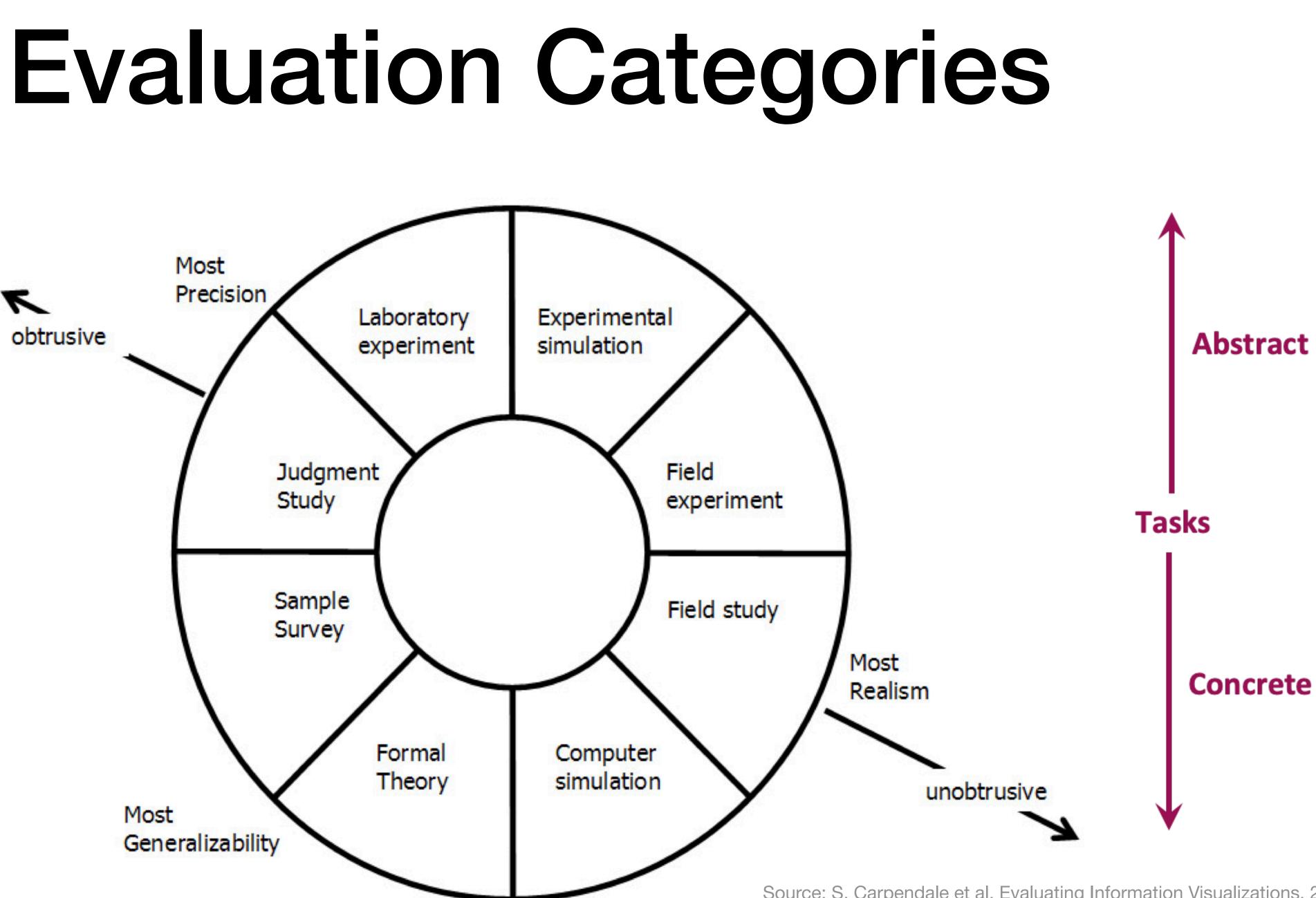
User-Centered Design





Which evaluation methods do you know?





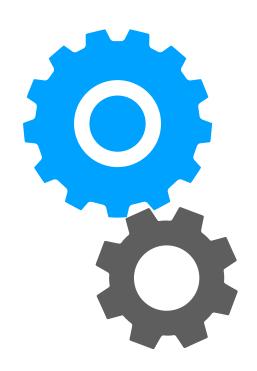
Source: S. Carpendale et al. Evaluating Information Visualizations. 2008.



Understanding the tool

- Algorithm/Technique performance benchmarking
- Qualitative result inspection expert evaluation, heuristics*

Understanding the processes (with users)



- Understanding environments and work practices questionnaires, surveys, ...*
- User evaluation methods user experience and performance
- Reasoning case studies

The aim

*later on, talk by Simone Kriglstein



Benchmarking

- Quantitative
- Performance comparison of a (novel) algorithm or technique
- Computation time, rendering speed (fps), memory footprint, ...
- The importance of test datasets and their availability
- Reported using descriptive statistics

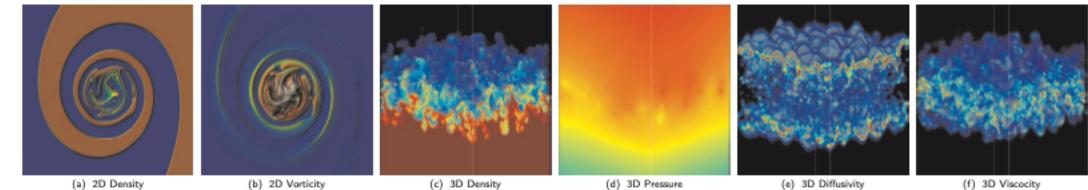
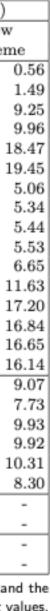


Fig. 1. Visualizations of 2D data (as pseudocolored height fields) and 3D data (volume rendered) used in our experiments.

																	_
data set compressed size (MB) and compression time (second										seconds	s)						
name	-	entropy	range	min	max	size	time	al	ib	IRKI	32006]	[EFF	72000]	[ILS2	20051	ne	w
name	(%)	(bits)	(bits)	mm	max	(MB)	(sec)	21	10		52000]		2000]	[11.52	.000	sche	
m2d density	3.89	3.49	21.83	8.7E - 01	1.2E+00	19.6	0.71	1.6	0.86	4.3	0.49	4.4	0.56	1.3	1.08	1.3	
m2d vorticity	99.20	22.25	31.05	-1.4E+02	2.5E+01	19.6	0.71	18.4	2.14	11.8	1.21	15.5	1.29	12.9	2.22	13.8)
m3d density	7.67	5.16	23.60	1.0E+00	3.0E + 00	364.5	12.81	50.4	17.55	100.5	9.06	96.3	8.48	35.7	19.03	35.5	J
m3d pressure	27.29	23.91	31.06	-3.7E+00	2.3E+03	364.5	12.80	229.2	99.76	95.6	9.31	87.9	8.87	40.1	18.79	40.4	
m3d diffusivity	36.87	23.19	30.02	0.0E + 00	6.8E + 00	364.5	12.68	297.6	42.90	250.8	19.09	239.3	15.02	198.8	31.92	203.0	1
m3d viscocity	50.07	24.86	28.59	8.6E - 15	2.9E+00	364.5	12.62	314.0	46.09	249.4	18.95	246.1	14.68	209.2	32.66	207.5	1
h3d temp	65.70	23.54	31.56	-7.7E+01	1.0E + 35	95.4	3.77	75.8	14.56	59.3	4.64	53.0	4.27	44.1	8.04	44.1	
h3d pressure	81.82	24.13	31.58	-3.4E+03	1.0E + 35	95.4	3.78	82.3	12.00	64.3	5.14	52.9	4.87	45.0	7.78	45.2	J
h3d x velocity	84.18	24.18	31.55	-5.3E+01	1.0E + 35	95.4	3.89	86.1	11.27	67.4	6.22	63.3	4.59	54.5	8.86	55.4	
h3d y velocity	84.32	24.18	31.55	-4.6E+01	1.0E + 35	95.4	3.83	84.5	11.42	67.1	5.74	62.3	5.04	53.5	8.64	53.8	ł
h3d z velocity	86.82	24.24	31.54	-3.2E+00	1.0E + 35	95.4	3.87	88.4	10.76	85.6	8.50	76.9	5.29	68.9	9.83	69.1	
M3d density	40.14	18.84	52.59	1.0E+00	3.0E + 00	288.0	11.28	136.8	41.91	160.3	11.63	121.6	10.94	-		105.2	1
M3d pressure	100.00	25.17	63.00	-2.2E+00	2.2E+00	288.0	11.20	272.6	35.18	237.3	14.91	225.1	16.59	-		208.4	1
M3d x velocity	100.00	25.17	63.00	-2.2E+00	2.3E+00	288.0	10.83	275.6	32.30	230.4	14.73	215.1	15.91	-		197.7	1
M3d y velocity	100.00	25.17	63.00	-2.1E+00	2.3E+00	288.0	10.54	275.1	32.19	223.1	14.27	215.2	15.16	-		197.7	1
M3d z velocity	100.00	25.17	63.00	-5.2E+00	9.0E + 00	288.0	10.32	275.5	32.62	226.6	14.74	213.7	16.05	-		196.8	1
atom x position	61.10	23.82	31.01	-4.8E - 02	4.6E+02	107.7	7.07	84.3	21.18	76.0	7.88	78.8	7.61	67.3	12.88	68.6	
atom y position	45.90	23.32	26.99	3.7E - 02	2.1E+03	107.7	7.08	65.9	30.76	60.4	6.97	56.4	6.31	47.0	10.49	46.9	J
atom z position	61.68	23.84	27.48	9.1E - 05	4.6E + 02	107.7	7.46	94.6	19.86	82.6	9.00	86.1	8.25	75.7	13.80	78.2	į.
atom y velocity	64.65	23.87	30.96	-1.5E - 01	1.4E - 01	107.7	7.30	95.7	19.88	93.8	10.07	99.1	9.65	84.3	14.93	87.6	Į.
atom temp	64.91	23.94	27.41	3.0E - 03	7.1E+03	107.7	6.69	95.7	19.76	91.6	10.27	95.9	8.34	84.6	15.02	84.6	1
atom energy	3.45	18.57	21.79	-3.6E+00	-2.7E + 00	107.7	7.15	77.9	38.59	74.1	7.98	71.8	7.01	60.8	12.66	60.5	J
lucy	61.39	24.38	31.09	-6.1E+02	1.2E+03	160.5	-	137.8	-	99.5	-	90.0	-	73.6	-	77.8	
$david_{1mm}$	25.23	17.08	31.11	-4.4E+03	1.8E+03	322.5	-	144.9	-	155.7	-	163.4	-	108.6	-	131.9	J
torso	84.72	18.48	31.08	-2.7E+02	5.8E + 02	1.9	-	1.7	-	1.5	-	1.5	-	1.3	-	1.3	_
rbl	71.90	20.14	25.99	1.5E+00	3.6E + 02	8.4	-	7.1	-	5.8	-	5.6	-	4.7	-	4.8	Į
																	_

Table 1. Compression results for the Miranda (m2d, m3d, M3d) and hurricane (h3d) structured grids, the atom point set, the lucy and david triangle meshes, and the torso and rbl tetrahedral meshes. All data but M3d is represented in single precision. The [ILS2005] scheme operates on single precision only, hence the missing values For the meshes we report only the compressed size of vertex coordinates; timings are dominated by connectivity coding, and are hence excluded. The range measures (the logarithm of) the number of floating-point values between min and max. Note that the first-order entropy is limited by the number of samples in a data set.

> Source: P. Lindstrom and M. Isenburg, "Fast and Efficient Compression of Floating-Point Data," in IEEE Transactions on Visualization and Computer Graphics, vol. 12, no. 5, pp. 1245-1250, Sept.-Oct. 2006.



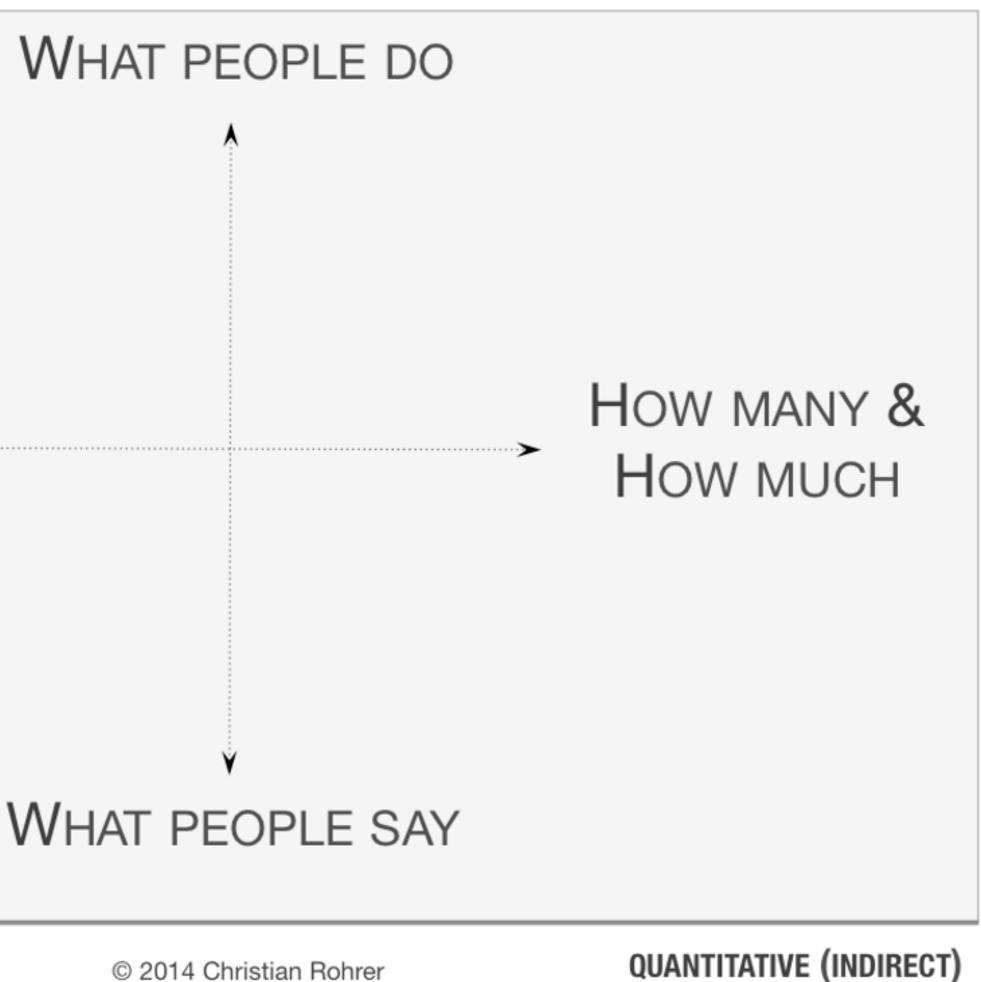
User Experience and Performance

BEHAVIORAL

WHY & HOW TO FIX

ATTITUDINAL

QUALITATIVE (DIRECT)



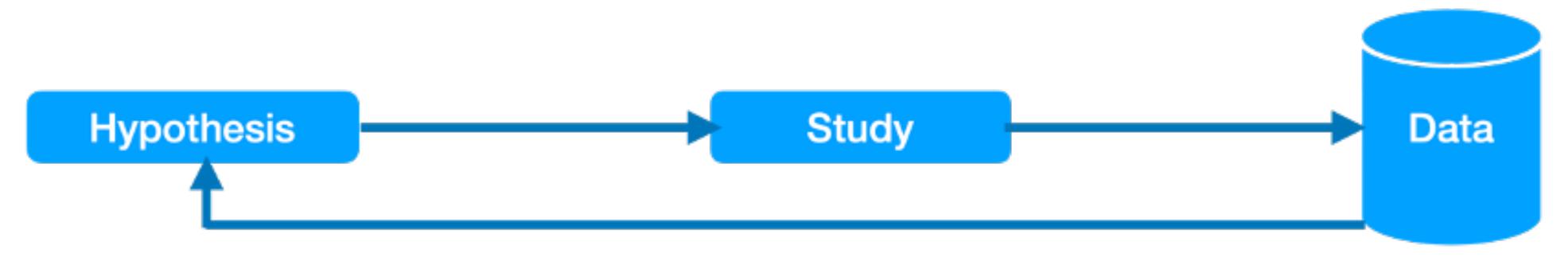
Source: <u>https://www.nngroup.com/articles/which-ux-research-methods/</u>



Evaluating User Performance

- Gathering evidence, not proving things (mathematicians do)
- Focus on users' effectiveness while using the system
- Used to obtain quantitative data about test participants' performance when they perform the tasks during usability tests
- Results are compared against baseline, automatic or competing techniques
- Hypotheses testing
- Reported using both inferential and descriptive statistics

Hypothesis



- A precise problem statement that can be directly tested through an empirical investigation
- More focused statement that can be examined by an experiment
- The goal is to find statistical evidence to confirm or reject null hypothesis in a reliable fashion
- Null hypothesis (H_0): There is <u>no difference</u> between the OLD and NEW methods.
- Alternative hypothesis (H_1): the NEW method will perform better than the OLD method.

Tasks

- Low-level tasks \bullet
 - those attributes.
 - Usually described by some task taxonomy (check <u>Task Taxonomy for Graph Visualization</u>)
- **High-level tasks** lacksquare
 - same person.
- **Explorative tasks** \bullet

 - E.g., replicability study: Re-create this example visualization using our (shiny) tool.

• Example: Given a set of data cases and two attributes, determine useful relationships between the values of

• Example: Due to errors in the data, several nodes may represent the same entity. For example, the coauthorship graphs often have duplicate author nodes. Identify whether two or more nodes represent the

• More general assignment, tasks implicitly include sub-tasks and higher cognitive load (creativity)

But what if you want to ... evaluate users' experience in terms of user satisfaction, system usability, learnability or others?



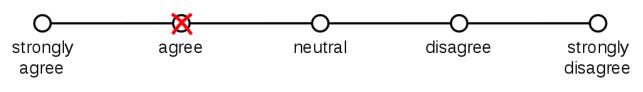


Likert Scales

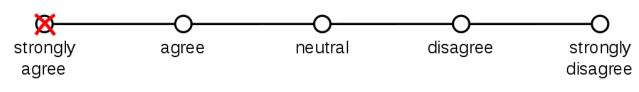
- Statement soliciting level of agreement
- Gradations between responses are (more or less) equal
- Ordinal data => Be careful with averaging (median is often better)
- Even vs. odd number of options

Website User Survey

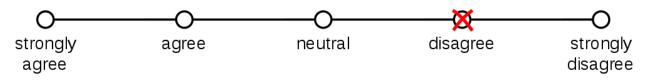
1. The website has a user friendly interface.



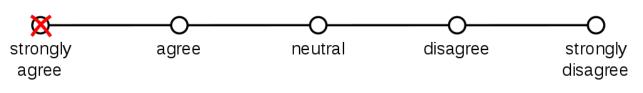
2. The website is easy to navigate.



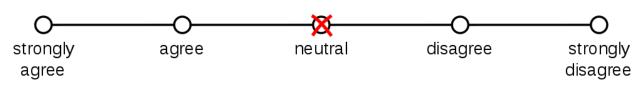
3. The website's pages generally have good images.



4. The website allows users to upload pictures easily.



5. The website has a pleasing color scheme.



Source: https://en.wikipedia.org/wiki/Likert_scale

Standardized Usability Questionnaires

Questionnaires designed for the assessment of perceived usability, typically with a specific set of questions presented in a specified order using a specified format with specific rules for producing scores based on the answers of respondents.

- Post-task: SEQ, SMEQ, ER, NASA-TLX, ...
- Post-study: SUS, UMUX, SUMI, PSSUQ, ...
- Benefits:

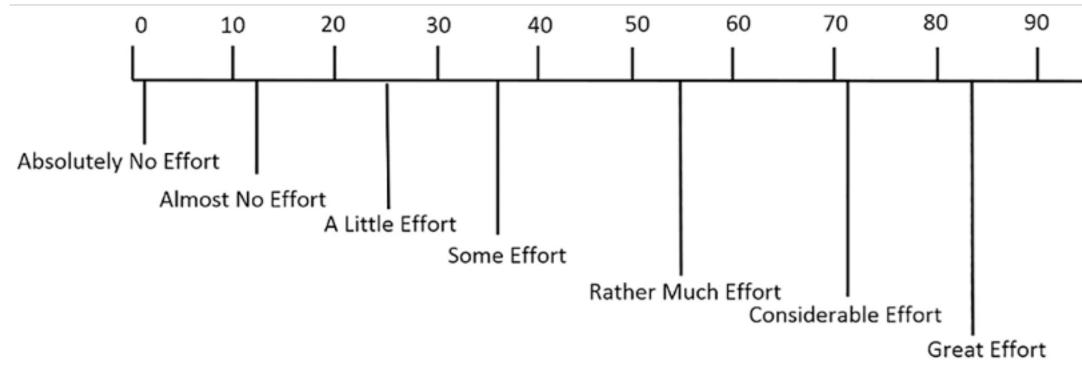
J. Sauro, J. R. Lewis, Quantifying the User Experience, 2016

objectivity, replicability, quantification, economy, generalization, communication

Post-task: Examples

Single Ease Question (SEQ) Overall, this task was : \bigcirc \bigcirc 0 \bigcirc \bigcirc \bigcirc Very Easy \bigcirc

Subjective Mental Effort Question (SMEQ)



Very Difficult

Source: Sauro J. and Dumas J. S. Comparison of Three One-Question, Post-Task Usability Questionnaires.

10	00 1	.10 1	.20	130	140	150
	at Effort					
ery Gre	eat Effort Extre	l eme Effor	t			

Source: So, et al. Subjective mental effort questionnaire.



Post-study: Examples

System Usability Scale (SUS)

Questions:

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

Strongly Disagree 1	2	3	4	Strongly Agree 5
0	0	0	0	0

Usability Metric for User Experience (UMUX) and UMUX-Lite

1.	[Thi	is syst	em's]	capa	bilitie	s me	et my	requirements.		
	1	2	3	4	5	6	7			
	Stro	ongly						Strongly		
	Disa	agree						Agree		
2.	Usi	ng [thi	is sys	tem] i	is a fr	ustra	ting e	kperience.		
	1	2	3	4	5	6	7			
	Stro	ongly						Strongly		
	Disa	agree						Agree		
3.	[Thi	is syst	em] i	s easy	to u	se.				
	1	2	3	4	5	6	7			
	Stro	ongly						Strongly		
	Disa	agree						Agree		
4.	I ha	ve to	spend	l too i	much	time	corre	cting things with		
	[this system].									
	1	2	3	4	5	6	7			
	Stro	ongly						Strongly		
	Disa	agree						Agree		

Source: Finstad K. The Usability Metric for User Experience

UMUX-Lite – same 7-point likert scale, only two questions

- This system's capabilities meet my requirements.
- This system is easy to use.

Source: Sauro J. Measuring Usability: From the SUS to the UMUX-Lite. MeasuringU

- environment"*
 - Case study != Usage scenario
- Four key aspects:
 - in-depth investigation of a small number of cases (often up to 5)

 - multiple data sources
 - emphasis on qualitative data an analysis (results in validity and reliability concerns)
- applicability in their work)

Case Studies

• "A detailed reporting about a small number of individuals working on their own problems in their normal

• examination in context (how the participant use the tool in his/her natural setting, not a lab-study)

Summarized feedback (feature requests, opinion of participants on the tool functions and limits and its

★ B. Shneiderman and C. Plaisant. 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. BELIV '06



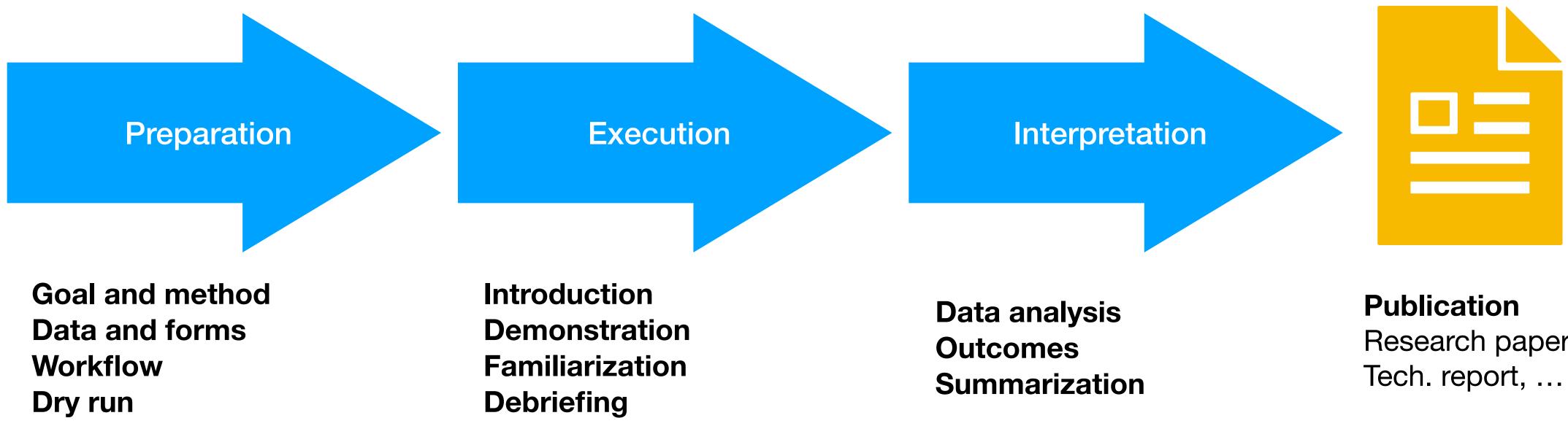
Goals of Case Studies

- Exploration understanding novel problems or situations
- Explanation developing models that can be used to understand a context of use
- Description documenting a system, technology use (in context) or the process
- Demonstration showing how the tool was successfully used

Case Study Design

- There are four main components of a case study design:
 - (research) questions What are you interested in?
 - hypotheses or propositions What you expect to find?
 - units of analysis What are you focusing on?
 - data analysis plan Which data we collect and how to process them?

Evaluation Workflow



Research paper, Tech. report, ...

Preparation

- Set the goal, then choose the method (with/without participants)
- Prepare data and related documents, datasets, consent forms, questionnaires (pre-, post-)
- Always do the pilot or dry run => identification of unexpected problems
- Make a checklist always follow the same steps
- Get the participants

Participants

- People participating in the experiment (don't use subjects)
- How many?
 - Short answer: use the same number as used in similar research
 - Too many: unnecessary work

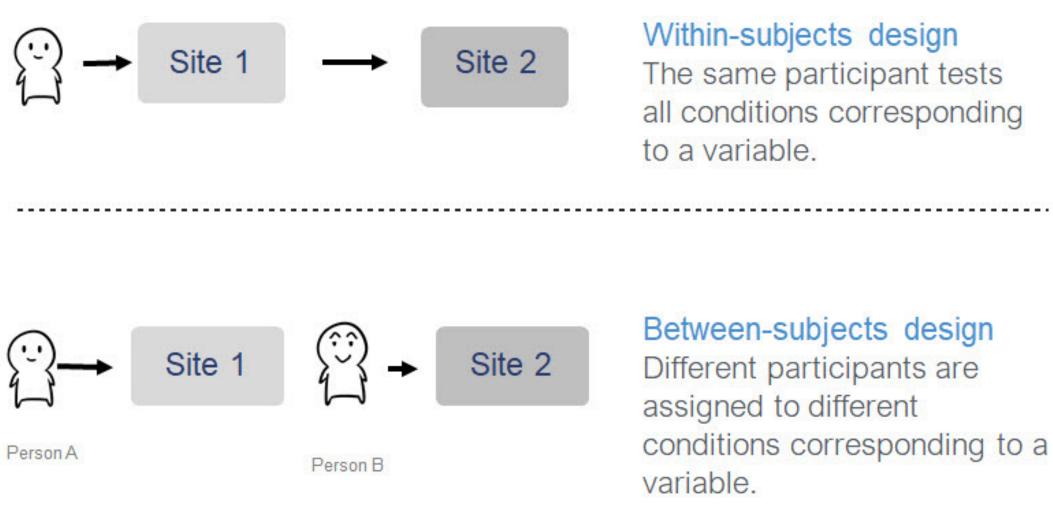
• Too few: fail to get statistically significant results => paper reject

Execution

- Follow the checklist
- Do not change experiment design or conditions after starting it
- Use different dataset for practice trials and main experiment
- With participants:
 - Get consent first, debrief participants afterwards
 - Record: audio/video, mouse traces, make notes

Evaluator's Toolbox

Within vs. Between Subjects



Source: https://www.nngroup.com/articles/between-within-subjects/

NNGROUP.COM NN/g

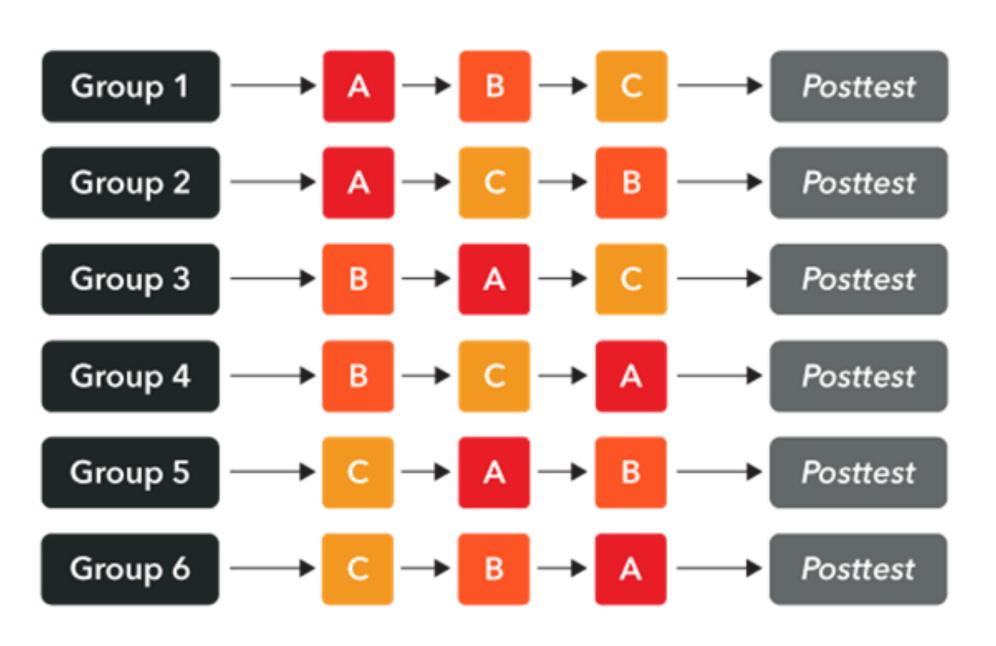
Advantages & Limitations

- + Smaller sample size
- + Effective isolation of individual differences
- + More powerful tests
- Hard to control learning effect
- Large impact of fatigue
- + Avoids learning effect
- + Better control of confounding factors (e.g., fatigue)
- Requires more people
- Harder to get statistically significant results
- Large impact of individual differences



Counterbalancing

- The effect of one condition "carries over" into the next one
- Common in within-subjects designs, e.g., learning effect
- Counterbalancing = compensation of carryover effects
- The order of tasks or datasets used in the experiment
- (Pseudo)Randomized order one for each participant
- **Latin Square:** n×n array filled with n different symbols, occurring exactly once in each row and column (=Sudoku).
 - Problem with the odd ones (from order 3)
 - Solution (for even-ordered only) is **Balanced Latin Square**
 - <u>Online generator</u>



Source: https://tophat.com/marketplace/social-science/psychology/-/research-methods-in-psychological-science-laura-freberg/736/68472/



Consent Form

- Who you are
- What you are asking the participants to do
- What kind of data you will be collecting and how it will be used
- What rights the participant has
- If they will be compensated
- The participant must explicitly say "yes" to the consent form

SIMON FRASER UNIVERSITY

INFORMED CONSENT BY SUBJECTS TO PARTICIPATE IN **EVALUATION OF AN INTERACTIVE COMPUTER SYSTEM FOR DATA** VISUALIZATION

The University and those conducting this project subscribe to the ethical conduct of research and to the protection at all times of the interests, comfort, and safety of subjects. This form and the information it contains are given to you for your own protection and full understanding of the procedures. Your signature on this form will signify that you have received a document which describes the procedures, possible risks, and benefits of this research project, that you have received an adequate opportunity to consider the information in the document, and that you voluntarily agree to participate in the project.

Knowledge of your identity is not required. You will not be required to write your name or any other identifying information on the research questionnaires. An audio recording of your voice and a video recording of the computer screen only will be made during the session. The video and audio recordings of the session will be reviewed only by the Principal Investigator. All research materials will be held confidential by the Principal Investigator and kept in a secure location. These research materials will be destroyed after the completion of the study.

Having been asked by Daryl H. Hepting of the School of Computing Science of Simon Fraser University to participate in a research project study, I have read the procedures specified in the accompanying information sheet. I understand the procedures to be used in this study and the personal risks and benefits to me in taking part. I understand that I may withdraw my participation in this study at any time.

I understand that my decision to participate in this study, and my subsequent involvement in it, will have absolutely no bearing on any other dealings I have with Mr. Hepting. This includes, but is not limited to, the case that I am a student in the CMPT 361 course taught by Mr. Hepting, offered at SFU during the 99-2 semester.

I understand that I may register any complaint I might have about the study with the researcher named above or with Dr. Jim Delgrande, Director, School of Computing Science of Simon Fraser University, Burnaby, BC, V5A 1S6, telephone 604-291-4277.

I may obtain copies of the results of this study, upon its completion, by contacting Mr. Daryl Hepting, in care of the School of Computing Science at Simon Fraser University.

I understand that my supervisor or employer may require me to obtain his or her permission prior to my participation in a study such as this.

I agree to participate by completing: a pre-task questionnaire; a training session on the prototype software system; a task with the prototype software system; and a post-task questionnaire. I understand that these activities will require approximately one hour at a time scheduled with Mr. Hepting. I understand that the experiment will be conducted in Room 9836 in the Applied Science Building of Simon Fraser University.

NAME (please type or print legibly):

ADDRESS:

SIGNATURE:

WITNESS:

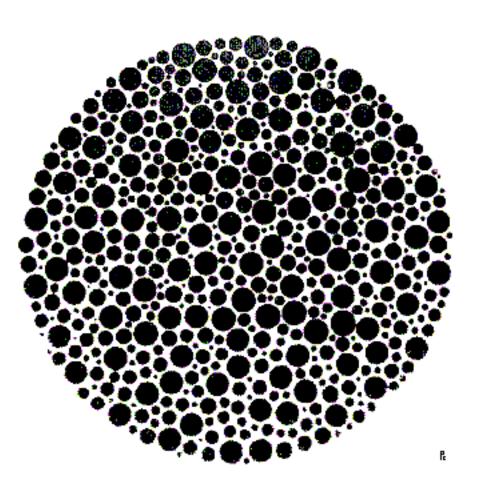
DATE:

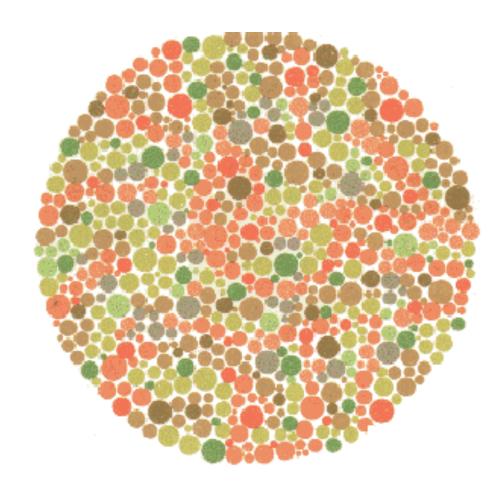
A COPY OF THIS SIGNED CONSENT FORM AND A SUBJECT FEEDBACK FORM WILL BE PROVIDED TO YOU AT YOUR EXPERIMENT SESSION.

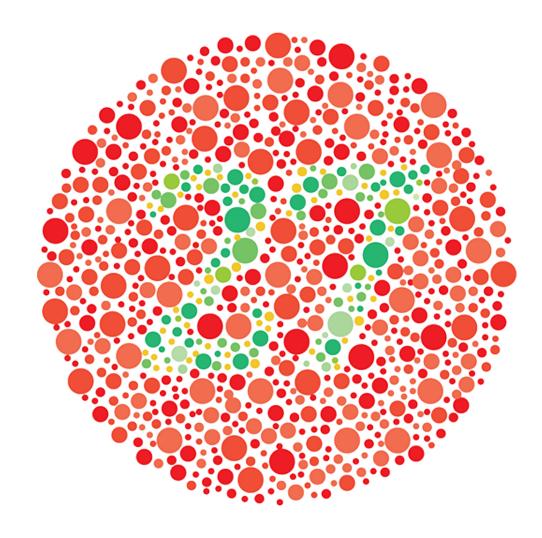
Source: D. Hepting. "A New Paradigm for Exploration in Computer-Aided Visualization". Dissertation thesis. Simon Fraser University. 1999.

Color Perception Test

- Shinobu Ishihara, 1917
- Ishihara plates
- Diagnostic test for color perception deficiencies
- 38 plates (full set)
- Variants with 10, 12 or 24







Statistical Evaluation

Descriptive statistics

- Summary of a data set characteristics
- Mean, median, mode, standard deviation, spread, central tendency, ...

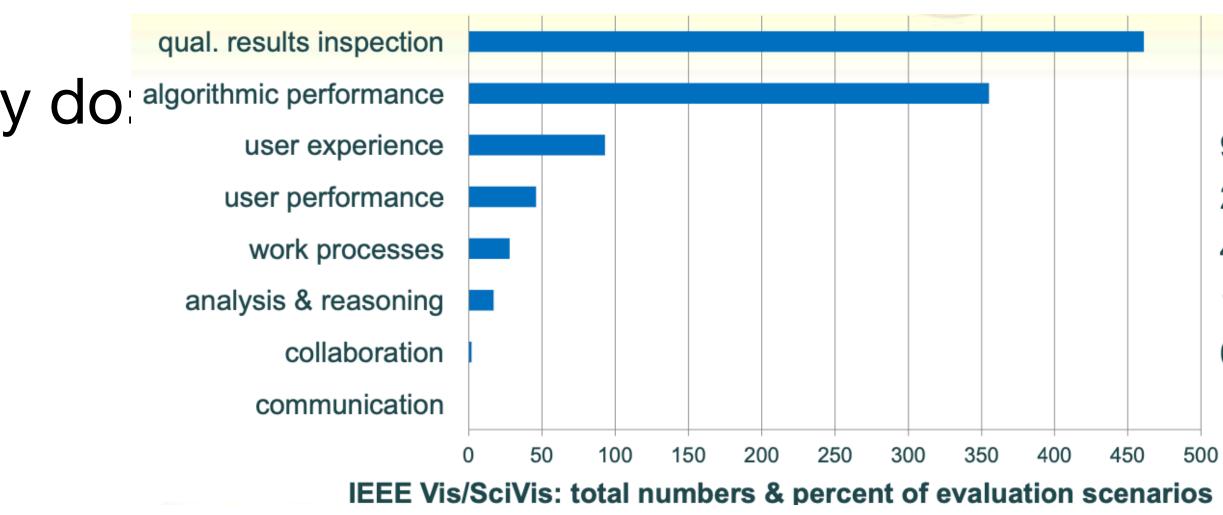
Inferential statistics

- Infers properties of a population based on a sample data
- Testing hypotheses and deriving estimates
- Parametric (t-test, ANOVA) and non-parametric tests
- Concrete methods are out of scope of this talk • Further reading: <u>Statistical Methods for HCI Research</u>

- In SciVis, InfoVis, VAST, we mostly do:
 - algorithm benchmarking, ullet**user performance** (quantitative)
 - case studies, qualitative inspection, \bullet **User experience** (qualitative)

- Evaluation methods are similar (same) to those in HCI

Take away...



 Contribution of real users is invaluable but also painful (involve them ASAP) Use methodologies and best practices from the field (learn from papers)

Chart source: http://tobias.isenberg.cc/personal/papers/lsenberg_2013_SRP_Slides.pdf

46% 35% 9.3% 2.8% 4.6% 1.7% 0.2% 0%

References

- Wiley Publishing.
- Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- for User Research". 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- 4.J. W. Creswell, Ch. N. Poth, "Qualitative Inquiry and Research Design: Choosing Among Five Approaches", 4th Ed. SAGE Publishing, 2018.
- 5.T. Isenberg et al., "A Systematic Review on the Practice of Evaluating Visualization," in *IEEE*
- 6.B. Shneiderman, et al. "Designing the User Interface: Strategies for Effective Human-Computer Interaction". 6th ed. Pearson Global Edition. 2018.

1. J. Lazar, J. H. Feng, and H. Hochheiser. 2010. Research Methods in Human-Computer Interaction.

2. I. S. MacKenzie. 2013. Human-Computer Interaction: An Empirical Research Perspective (1st. ed.).

3. J. Sauro and J. R. Lewis. 2016. "Quantifying the User Experience, Second Edition: Practical Statistics

Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2818-2827, Dec. 2013.